



**DGK** Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

---

Reihe C

Dissertationen

Heft Nr. 895

**Christian Kruse**

**Impact maps from bomb craters  
detected in aerial wartime images  
using marked point processes**

**München 2023**

**Bayerische Akademie der Wissenschaften**

**ISSN 0065-5325**

**ISBN 978-3-7696-5307-6**

---

**Diese Arbeit ist gleichzeitig veröffentlicht in:**

**Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover**

**ISSN 0174-1454, Nr. 383, Hannover 2023**





DGK Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

---

Reihe C

Dissertationen

Heft Nr. 895

Impact maps from bomb craters  
detected in aerial wartime images  
using marked point processes

Von der Fakultät für Bauingenieurwesen und Geodäsie  
der Gottfried Wilhelm Leibniz Universität Hannover  
zur Erlangung des Grades  
Doktor-Ingenieur (Dr.-Ing.)  
genehmigte Dissertation

von

Christian Kruse, M. Sc.

München 2023

Bayerische Akademie der Wissenschaften

ISSN 0065-5325

ISBN 978-3-7696-5307-6

---

Diese Arbeit ist gleichzeitig veröffentlicht in:  
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover  
ISSN 0174-1454, Nr. 383, Hannover 2023

Adresse der DGK:



Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften (DGK)

Alfons-Goppel-Straße 11 • D – 80 539 München

Telefon +49 – 89 – 23 031 1113 • Telefax +49 – 89 – 23 031 - 1283 / - 1100

e-mail [post@dgk.badw.de](mailto:post@dgk.badw.de) • <http://www.dgk.badw.de>

Prüfungskommission:

Vorsitzender: Prof. Dr. Philipp Otto

Referent: Prof. Dr.-Ing. habil. Christian Heipke

Korreferenten: Prof. Dr.-Ing. habil. Monika Sester

Prof. Dr.-Ing. Uwe Sörgel (Uni Stuttgart)

apl. Prof. Dr. techn. Franz Rottensteiner

Gutachter: Dr.-Ing. habil. Fabio Remondino (Foundation Bruno Kessler, Trentino)

Tag der mündlichen Prüfung: 16.12.2022

---

© 2023 Bayerische Akademie der Wissenschaften, München

Alle Rechte vorbehalten. Ohne Genehmigung der Herausgeber ist es auch nicht gestattet,  
die Veröffentlichung oder Teile daraus auf photomechanischem Wege (Photokopie, Mikrokopie) zu vervielfältigen

ISSN 0065-5325

ISBN 978-3-7696-5307-6

# Abstract

Even more than 75 years after the Second World War ended, numerous unexploded bombs (duds) linger in the ground and pose a considerable hazard to society. The areas that may contain duds are documented in so-called impact maps, which are based on locations of exploded bombs; these locations can be found in aerial wartime images taken after bombing. Often, several surveillance flights were carried out and, thus, areas may be covered by multiple images. Such images are being used today by experts to manually identify suspicious locations concerning duds, which entails an immense processing effort. Consequently, for the cost-efficient creation of an impact map, in this work flagging areas to be probed, an automatic generation is indispensable.

To generate impact maps, in this thesis a novel probabilistic approach based on marked point processes (MPPs) for the automatic detection of bomb craters in aerial wartime images is investigated. The object model for the craters is represented by circles and is embedded in the MPP-framework. By means of stochastic sampling the most likely configuration of objects within the scene is determined. Randomly adding objects to and removing them from the current configuration, or changing their positions and modifying the circle parameters, creates new object configurations. Each configuration is evaluated using an energy function that describes its consistency with the predefined model. High gradient magnitudes along the object border, homogeneous grey values inside the object as well as a high contrast between the object and a concentric annulus around it are favoured while overlaps between objects are penalized. In connection with the stochastic sampling, the Reversible Jump Markov Chain Monte Carlo method in combination with simulated annealing is used to search for the global optimum of the energy function in an iterative way. To take advantage of the multi-image coverage, the procedure allows the combination of individual detection results covering the same location. Afterwards, a probability map for duds is generated from the detections via kernel density estimation and areas around the detections are classified as contaminated, resulting in an impact map.

The method is tested on aerial wartime images acquired by the Allied forces during World War II and taken over different areas in Europe. Within the images, content and appearance differ, the latter e.g. due to blurring, uneven illumination, lack of contrast or changes due to time. For evaluation purposes, reference data manually annotated by experts are used. The results show the potential of the procedure, also indicating that complex image content pushes the method to its limits. By using redundant image information, compared to the single image approach, a clear improvement is achieved; experiments show an increased F1-score in almost every case. Finally, experiments in connection with the precision of the generated impact map reveal that 45 % of the areas

that actually need to be probed can be detected with a precision of 90 %, rendering the procedure attractive for supporting the manual inspection of the images.

**Keywords:** Marked point processes, Reversible Jump Markov Chain Monte Carlo sampling, aerial wartime images, bomb craters, duds, impact maps

# Kurzfassung

Selbst mehr als 75 Jahre nach Ende des Zweiten Weltkriegs verweilen noch zahlreiche nicht explodierte Bomben (Blindgänger) im Boden und stellen eine erhebliche Gefahr für die Gesellschaft dar. Die Gebiete, in denen sich Blindgänger befinden können, werden in so genannten Belastungskarten dokumentiert, die auf den Positionen der explodierten Bomben beruhen; diese Positionen sind auf Kriegsluftbildern zu finden, die nach der Bombardierung aufgenommen wurden. Oftmals wurden mehrere Aufklärungsflüge durchgeführt, sodass Gebiete durch mehrere Bilder abgedeckt sein können. Solche Bilder werden heutzutage von Experten genutzt, um manuell verdächtige Stellen im Hinblick auf Blindgänger zu identifizieren, was mit einem immensen Bearbeitungsaufwand verbunden ist. Folglich ist für die kosteneffiziente Erstellung einer Belastungskarte, welche in dieser Arbeit zu sondierende Gebiete ausweist, eine automatische Generierung unerlässlich.

Zur Erstellung von Belastungskarten wird in dieser Arbeit ein neuartiger probabilistischer Ansatz auf Basis von markierten Punktprozessen (MPPs) zur automatischen Detektion von Bombenkratern in Kriegsluftbildern untersucht. Das Objektmodell für die Krater wird durch Kreise repräsentiert und in das Verfahren der MPPs eingebettet. Mittels stochastischem Sampling wird die wahrscheinlichste Konfiguration der Objekte innerhalb der Szene ermittelt. Durch das zufällige Hinzufügen und Entfernen von Objekten zu bzw. aus der aktuellen Konfiguration, die Änderung ihrer Positionen und die Modifikation der Kreisparameter entstehen neue Objektkonfigurationen. Jede Konfiguration wird anhand einer Energiefunktion bewertet, die ihre Übereinstimmung mit dem vordefinierten Modell beschreibt. Hohe Gradientenmagnituden entlang des Objektrandes, homogene Grauwerte innerhalb des Objekts sowie ein starker Kontrast zwischen dem Objekt und einem konzentrischen Umring um dieses werden favorisiert, während Überlappungen von Objekten bestraft werden. Im Zusammenhang mit dem stochastischen Sampling wird die Reversible-Jump-Markov-Chain-Monte-Carlo-Methode in Kombination mit Simulated Annealing genutzt, um das globale Optimum der Energiefunktion iterativ aufzufinden. Um die Vorteile der Mehrbildabdeckung zu nutzen, erlaubt das Verfahren die Kombination einzelner Detektionsergebnisse desselben Ortes. Anschließend wird aus den Detektionen mittels Kerndichteschätzung eine Wahrscheinlichkeitskarte für Blindgänger erzeugt und Bereiche um die Detektionen werden als belastet klassifiziert, wodurch eine Belastungskarte entsteht.

Das Verfahren wird an Luftbildern aus dem Zweiten Weltkrieg getestet, die von den alliierten Streitkräften über verschiedenen Gebieten in Europa aufgenommen wurden. Innerhalb der Bilder unterscheiden sich Inhalt und Aussehen, letzteres z. B. aufgrund von Unschärfe, ungleichmäßiger Beleuchtung, fehlendem Kontrast oder zeitlich bedingter Veränderungen. Zur Evaluierung werden

von Experten manuell annotierte Referenzdaten herangezogen. Die Ergebnisse zeigen das Potential der Methode auf, geben aber auch zu erkennen, dass komplexe Bildinhalte das Verfahren an seine Grenzen bringen. Durch die Verwendung redundanter Bildinformationen wird im Vergleich zum Einzelbildansatz eine deutliche Verbesserung erzielt; Experimente weisen in fast allen Fällen eine höhere F1-Score auf. Schließlich zeigen Experimente im Zusammenhang mit der Korrektheit der erzeugten Belastungskarte, dass 45 % der tatsächlich zu sondierenden Gebiete mit einer Korrektheit von 90 % gefunden werden können, was das Verfahren für die Unterstützung der manuellen Inspektion der Bilder attraktiv macht.

**Schlagworte:** Markierte Punktprozesse, Reversible-Jump-Markov-Chain-Monte-Carlo-Sampling, Kriegsluftbilder, Bombenkrater, Blindgänger, Belastungskarten

---

# Table of contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation.....	1
1.2	Objective and scientific contribution .....	4
1.3	Thesis outline .....	5
<b>2</b>	<b>Related work</b>	<b>7</b>
2.1	Methods for object detection.....	7
2.2	Object detection based on marked point processes.....	9
2.2.1	Simple geometric primitives.....	10
2.2.2	More complex object shapes .....	12
2.2.3	Favouring sampling of objects at certain locations .....	13
2.3	Bomb crater detection .....	14
2.4	Spatial data analysis .....	16
2.4.1	Classification of spatial data analysis techniques.....	16
2.4.2	Impact map generation from detected bomb craters.....	18
2.5	Methods for planetary crater detection .....	18
2.6	Discussion .....	21
<b>3</b>	<b>Basics</b>	<b>25</b>
3.1	Marked point processes.....	25
3.1.1	Poisson point processes .....	27
3.1.2	Gibbs energy .....	28
3.2	Monte Carlo methods.....	29
3.2.1	Markov Chain Monte Carlo sampling .....	29
3.2.2	Reversible Jump Markov Chain Monte Carlo sampling .....	32
3.2.3	Simulated annealing.....	34
3.3	Kernel density estimation.....	35

---

<b>4</b>	<b>Generation of impact maps from bomb craters</b>	<b>39</b>
4.1	Bomb crater detection using marked point processes.....	39
4.1.1	Object model .....	40
4.1.2	Energy function .....	41
4.1.3	Changes in the object configuration.....	47
4.1.4	Limitation of the search space.....	51
4.1.5	Pre-processing .....	53
4.1.6	Conceptual workflow .....	54
4.2	Fusing the results from multiple images.....	55
4.3	Impact map .....	57
4.4	Discussion.....	58
<b>5</b>	<b>Experimental setup</b>	<b>61</b>
5.1	Objectives .....	61
5.2	Test data.....	63
5.3	Parameter settings.....	66
5.4	Evaluation procedure .....	72
5.5	Comparison with a deep learning based object detector .....	74
<b>6</b>	<b>Results and Discussion</b>	<b>77</b>
6.1	Analysis of the model .....	77
6.1.1	Random numbers.....	78
6.1.2	Object model .....	79
6.1.3	Energy terms .....	80
6.2	Detection of bomb craters and impact map generation .....	82
6.2.1	Performance for different aerial wartime images.....	82
6.2.2	Redundant image information.....	89
6.2.3	Focus on precision.....	93
6.3	Comparison to a state-of-the-art object detector based on CNNs .....	96
6.4	Object detection for counting purposes in images.....	100
<b>7</b>	<b>Conclusions and Outlook</b>	<b>105</b>
<b>A</b>	<b>Evaluation of the assignment process in the context of redundant image information</b>	<b>109</b>
	<b>References</b>	<b>113</b>

# 1 Introduction

## 1.1 Motivation

Although the last combat operations of World War II took place more than three quarters of a century ago, their aftermath is still present today. Unexploded ordnance, such as grenades or bombs, remain hidden in the ground. The latter, in particular, are dangerous due to their high explosive force. Experts of Lower Saxony's Explosive Ordnance Disposal Service estimate that about 10 % - 15 % of all dropped bombs did not detonate. During construction works there is still a real danger today that these duds explode.

Surveillance flights were often carried out shortly before and after an air strike, but also on other occasions; more information regarding military aerial surveillance can be found in (Vogler, 2020). Thus, while multiple coverage over time may exist, these images are often blurry, contain image errors, have only little contrast and are unevenly illuminated. Furthermore, they differ in scale and the earth surface changed quickly at occasions, for instance due to weather (e.g. craters were filled with water) or human intervention (e.g. craters were filled with soil). Consequently, the appearance of wartime images of the same location taken at different times may vary considerably. In addition, the depth of the crater as well as the geological and hydrological conditions of the soil body affects the appearance of craters in the image (Waga and Fajer, 2021). All this makes the automatic bomb crater detection a challenging task (Figure 1.1).

The wartime images are being used today by experts to find potentially dangerous sites. In this context, a central task of Lower Saxony's Explosive Ordnance Disposal Service is the manual inspection of aerial wartime images to identify suspicious locations concerning duds. Despite the restriction to particularly endangered or otherwise relevant areas, it entails an immense processing effort, as the number of aerial wartime images taken over Lower Saxony between 1939 and 1945 exceeds 150.000; the number of images for the whole of Germany is around 1.3 million, according to a query by Lower Saxony's Explosive Ordnance Disposal Service. For many applications, it is sufficient to have comprehensive information on the basic occurrence of warlike impacts. Such knowledge can be represented in an "impact map" indicating whether areas are likely to be contaminated or not. In this context, contaminated areas are expected to contain duds with a high probability, whereas uncontaminated areas should not contain any.

Compared to bomb craters, duds (Figure 1.1 g) are far more difficult to detect in aerial images. They are only visible if they are not covered by the material ejected by bomb explosions in their

vicinity or by the detonation hole itself. Moreover, a large scale and good image quality are essential. In addition, duds can easily be confused with other structures, such as small image errors or one-man holes.

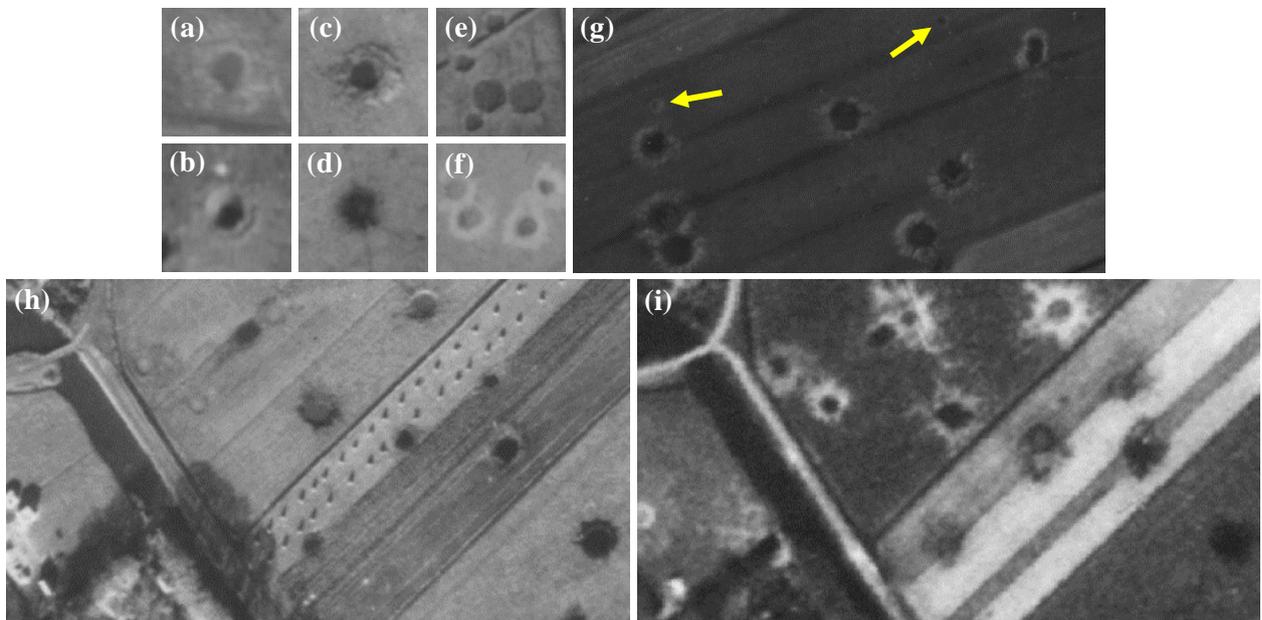
For the cost-efficient creation of an impact map, an automatic generation is indispensable. This work focuses on bomb craters to deduce the probability for the occurrence of duds as craters indicate areas where duds may be located. Hence, this probability can be used to identify contaminated areas to be represented in the impact map. The thesis focuses on a scenario in which the precision of the impact map is most important: One is interested in detecting areas that have a high likelihood of containing a dud so that it makes sense to send a team of experts to that area to probe it using geophysical detectors. As this is expensive, false detections should be avoided. In general, probing is recommended for the areas of individual bomb craters as well as clusters of craters, given that the former already need to be probed due to a possible occlusion of duds by bomb craters together with their ejecta. The main benefit for the Explosive Ordnance Disposal Service is that the aerial images would then no longer have to be inspected manually in areas flagged as to be probed. Another scenario could be to exclude areas in advance in which there is a high probability that no dud exists. Here, bomb craters must not be missed, because an area falsely classified as uncontaminated could contain a dud, which might lead to a detonation, e.g. in a building project. Thus, this type of application is to be judged as very critical and will not be pursued further in this work.

For the automatic detection of objects, various techniques from the field of image analysis have been suggested. In this context, prior knowledge expressed in terms of probabilities is frequently integrated into the object model. Markov Random Fields (Geman and Geman, 1984) and Conditional Random Fields (Kumar and Hebert, 2006), integrating probabilistic models of context, are restricted to local interactions, e.g. between neighbouring pixels. More global constraints about the objects, for instance with respect to their shape, are difficult to integrate. This limitation can be addressed using marked point processes (MPPs; Descombes and Zerubia, 2002; Daley and Vere-Jones, 2003), a model-based probabilistic method, which is especially suitable if a strong object model is needed, e.g. to obtain useful results even for challenging data. The basic concept of MPPs is to model objects by a stochastic process. This involves modelling the number and the distribution of the objects in the image as well as the parameters describing their geometry as random variables. Based on simulations, different object configurations can be created to find the globally optimal configuration that best fits the input data and the integrated prior knowledge. In this way, knowledge about objects is expressed on a holistic level and characteristics of objects can be integrated beyond pixel-based relations. In addition, MPPs benefit from their flexibility in integrating knowledge about the objects and their mutual relationships. Furthermore, the number of objects in the scene during sampling is variable, i.e. it does not have to be known beforehand.

Starting with pioneering works (Baddeley and van Lieshout, 1993; Rue and Hurn, 1999), MPPs have achieved good results in several object detection problems (e.g. Lafarge et al., 2010; Tournaire et al., 2010; Börcs and Benedek, 2015; Favreau et al., 2019). Often, the Bhattacharyya distance (Bhattacharyya, 1943), measuring the contrast between the object and a concentric annulus

around it, is used within the energy function of the MPPs for the detection of circular or elliptical objects (e.g. Gamal-Eldin et al., 2010; Descamps et al., 2011; Dubosclard et al., 2014). However, the frequently present annulus of bomb craters, caused by the ejected material, may differ considerably from image to image, e.g. in terms of its width and appearance (see Figure 1.1). Therefore, this measure does not appear suitable on its own. In this context, the additional consideration of gradient magnitudes at the object's border (e.g. Ortner et al., 2007; Brédif et al., 2013) seems reasonable, though these two approaches are designed for the detection of buildings in digital surface models.

As usually multiple but coarsely georeferenced images of the same area exist, exploiting redundancy by merging detection results from overlapping image parts seems beneficial. In order to overcome the coarse georeferencing, it would be possible to improve the co-registration of the respective images based on global approaches, e.g. automatic aerial triangulation (Ackermann and Krzystek, 1997). However, this may become challenging, particularly due to the different appearances (e.g. caused by seasonal changes, noise) and missing camera information. Given this, a local approach, such as in (Brenner et al., 2018), the only work the author of the thesis is aware of on that topic that makes use of redundant image information in an automatic manner, may be considered. However, Brenner et al. (2018) do not provide detailed information on their procedure and double detections from different images are eliminated, which is not suitable for the proposed application scenario.



**Figure 1.1:** Different appearances of bomb craters and aerial wartime images. (a-f) Bomb craters appear differently due to varying scale, size or appearance. (g) Two potential duds are indicated by yellow arrows. (h-i) Two subsets of images of the same location taken in April and May 1944, respectively.

Having in mind the application scenario, individual positions of bomb craters are not very informative to indicate areas that contain duds with a high probability. In this context, a useful tool to deduce area-based statements is statistical modelling. Here, the existence of a dud is represented by a probability density function (pdf; Parzen, 1962). In contrast to proposed tile-based approaches (Tavakkoli Sabour et al., 2014; Juhász and Neuberger, 2018), this allows for a more precise delimitation of potentially contaminated areas. In general, parametric or non-parametric approaches may be used. While the former assume an analytical model for the pdf, with a parameterisation derived from training data, the latter estimates the pdf directly from the data. This second approach avoids having to select a pdf model and to estimate its distribution parameters (Bishop, 2006). As a popular non-parametric technique, kernel density estimation is used (e.g. Scott, 2015).

## 1.2 Objective and scientific contribution

The aim of the work is to develop a method for the automatic detection of bomb craters in aerial wartime images and the subsequent generation of an impact map using the detected objects. The scientific contribution can be summarised as follows.

- **The development of a novel approach for the detection of bomb craters in aerial wartime images based on marked point processes.** MPPs have not yet been used for the detection of bomb craters. Inspired by other works with the aim of detecting round (e.g. Zhou et al., 2010; Aval et al., 2018) or elliptical (e.g. Descamps et al., 2011; Craciun and Zerubia, 2013) isolated objects, which applies to bomb craters, in this thesis bomb craters are modelled as circles. In this context, it can be shown that the object model of an ellipse, which may seem more obvious due to the sometimes elliptical shape of bomb craters, leads to similar results as the circle. Using a circle, an object is defined by fewer parameters, keeping the search space in the optimization process small and making the model generally more stable. In connection with the modelling of the energy function of the MPP for circular objects, the Bhattacharyya distance is often used in the literature. However, this measure alone does not appear to be suitable for the detection of bomb craters, as their frequently present annulus may differ considerably from image to image, e.g. in terms of its width and appearance. For this reason, a novel energy function is introduced in this work, which takes into account – among other characteristics – the homogeneity of grey values inside the object, a frequently occurring property of bomb craters.
- **The consideration of redundant image information.** Considering that aerial wartime images vary considerably in their appearance, unlike most existing work on that topic, this approach exploits redundancy by merging detection results from multiple overlapping and coarsely georeferenced images. In order to relax the requirements for precise georeferencing of aerial wartime images, which is difficult to obtain, similar to (Brenner et al., 2018),

matching of the individual detections referring to the same object is taken into account. However, in their work double detections from different images are eliminated in order not to miss any crater in an investigated area. Thus, the idea of supporting detections, i.e. that several detections of the same object in different images are an indication for indeed having found a correct object, is not pursued in the way needed in the proposed application scenario. Consequently, this is addressed in the newly developed approach. By exploiting redundant image information, the impact map derived from the combined detection results can be improved considerably.

- **The development of an approach transferable to a variety of aerial wartime images and to further images concerning population counting.** The MPP-procedure is expected to be applicable not only for the detection of bomb craters in aerial wartime images. Indeed, it is primarily designed for this purpose and can be used for different areas with varying characteristics (e.g. in terms of appearance, content, lighting, ground sampling distance, number of bomb craters). At the same time, however, it should also allow population counting in images.

## 1.3 Thesis outline

The contents of this thesis are structured as follows. In Chapter 2, related work on methods for object detection tasks, including the procedure of MPPs, is given. Moreover, previous works on bomb and planetary crater detection are surveyed. Finally, possible ways to derive area-based information from the locations of individual objects are illustrated. The mathematical basics of stochastic optimization based on MPPs using Reversible Jump Markov Chain Monte Carlo sampling coupled with simulated annealing and of kernel density estimation are described in Chapter 3. Chapter 4 presents the novel framework developed in this thesis for the detection of bomb craters, the combination of individual detection results in connection with redundant image information and the derivation of the impact map. The experimental setting, including the setup of the experiments with their associated objectives, the data sets used, the parameters chosen for the experiments and the evaluation procedure for the results, are given in Chapter 5. Chapter 6 shows and discusses the results of the conducted experiments to reveal strengths and weaknesses of the method. Finally, conclusions are drawn and future directions of research are sketched in Chapter 7.



## 2 Related work

This chapter reviews and discusses literature relevant to this thesis. Methods for object detection are surveyed on a more general level in Section 2.1. Afterwards, related work on object detection with marked point processes (MPPs) is discussed (Section 2.2). Existing approaches associated with the task of bomb crater detection are addressed in Section 2.3. As this thesis is concerned with the generation of an impact map, i.e. the deduction of area-based information from objects, Section 2.4 deals with spatial data analysis and its techniques. Given that crater detection plays an important role in planetary science too, and because several interesting methods have been developed in the context of planetary data, previous work on planetary crater detection is also surveyed (Section 2.5). The chapter closes with a comprehensive discussion of the reviewed literature in Section 2.6.

### 2.1 Methods for object detection

Within the last decades, due to its important role for a wide range of applications, several methods have been developed for the automatic detection of objects in optical remote sensing images, the kind of data to which the aerial wartime images used in this thesis can also be attributed. As a consequence of the widespread interest regarding the detection of objects, a number of publications already exist that provide an overview of the topic. In this context, while the works of Zhao et al. (2019) and Dhillon and Verma (2020) solely survey deep learning based approaches, Cheng and Han (2016) as well as Zou et al. (2019) also include more traditional detection methods. Generally, following Cheng and Han (2016), methods for object detection can be divided into four main categories.

One category includes template matching-based approaches, which can be considered as one of the earliest methods for object detection. In a first step, templates for the objects to be detected are generated, which are afterwards used to scan an image at each possible position to find, according to a similarity measure, the best matches. Early research in the area of template matching concentrated on rigid templates (e.g. McKeown and Denlinger, 1988) that have been designed for detecting specific objects with simple appearance and small variations such as roads (e.g. Kim et al., 2004). However, due to their scale and rotation dependency and the usually existent large intra-class variations among the objects, the use of rigid templates is limited. In contrast, deformable template matching (Fischler and Elschlager, 1973), making use of free-form deformable templates

or parametric templates (Jain et al., 1998), is more powerful and flexible in dealing with shape deformations and intra-class variations. Nevertheless, for template design, prior information regarding possible manifestations of objects in the data and parameters of the geometrical shape are needed.

Another category are knowledge-based object detection methods, a review related to optical remote sensing images is given in (Baltsavias, 2004). These approaches typically translate the task of object detection into a hypotheses testing problem by making use of knowledge and rules on the objects to be detected. In this context, geometric knowledge and context knowledge is widely employed. The geometric information of the object encodes prior knowledge by using parametric or generic shape models. For example, Weidner and Förstner (1995) developed an approach to extract the 3D shape of buildings from digital elevation models making use of domain knowledge, specifically parametric and prismatic building models. Context knowledge includes information on how an object interacts with its neighbourhood or involves spatial constraints or relationships between objects and background. An example can be found in (Peng and Liu, 2005), who developed a shadow-context model to extract buildings in dense urban aerial images. It is worth noting that the core of knowledge-based object detection methods is how to effectively transform implicit knowledge about the objects into explicit detection rules. If the defined rules are too strict, objects may be missed; conversely, rules that are too loose will probably lead to false positives.

With the growing availability of sub-metre imagery around the turn of the millennium, an increase in contributions of object-based image analysis (Blaschke, 2010) was observed, consisting of two steps. First, the image is segmented into regions, each representing a relatively homogeneous group of pixels. In a second step, those segments, also referred to as objects, are classified. For the segmentation, numerous techniques have been developed; a review of algorithms and challenges can be found in (Hossain and Chen, 2019). However, Hay et al. (2005) pointed out that the real challenge lies in the definition of appropriate segmentation parameters for the varying size, shape, and spatial distribution of objects composing a scene, as the subsequent feature extraction and classification is highly dependent on the quality of the image segmentation.

Finally, there are machine learning-based approaches, in which object detection can be performed by learning a classifier that captures the variation in object appearances from training data in a supervised framework. The input of the classifier is a set of image regions for which their corresponding feature vectors and class labels are known. In traditional machine learning, for the extraction of the so-called hand-crafted features, e.g. spectral, textural, structural and 3D information can be considered; a review on feature extraction is available in (Li et al., 2015). In this context, it is important to cleverly design the features, as the performance of the classifier will be limited by their quality. After feature extraction, the classifier can be trained with the objective of minimizing the misclassification error on such training samples, i.e. a model of the appearance of the objects in the data is learned. Finally, during classification, the classifier predicts the class label for unseen samples based on the learned features. For the classification, probabilistic approaches,

such as logistic regression (Bishop, 2006), and non-probabilistic approaches, as, for instance, support vector machines (Cortes and Vapnik, 1995) or random forests (Breiman, 2001), can be used in connection with traditional machine learning. In contrast, deep learning-based approaches making use of convolutional neural networks (Krizhevsky et al., 2012) and having been very popular for several years, do not require the separation into feature extraction and classification. In fact, features and model are learned together, i.e. both tasks support each other. In the context of CNN-based object detection, there are one-stage and two stage detectors. The latter follows the traditional object detection pipeline by generating region proposals at first and then classifying each proposal into object categories. Girshick et al. (2014) took the lead in proposing regions with CNN features (R-CNN), followed by networks improved with regard to certain aspects. Well-known works are, for instance, Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2015) and Feature Pyramid Networks (Lin et al., 2017). On the contrary, one-stage detectors, such as You Only Look Once (Redmon et al., 2016), adopt a unified framework to extract the object classes and locations simultaneously, making them fast. Both traditional machine learning-based as well as deep learning-based approaches need labelled training data to work at all, also making the quality of the results dependent on the samples considered for training. Due to the huge number of parameters to be learned, approaches based on CNNs need a comparatively larger set of training samples to yield good results and, thus, their creation is rather time-consuming.

## 2.2 Object detection based on marked point processes

MPPs, a model-based probabilistic approach, provide an alternative to the previously mentioned methods. They allow the use of a strong object model, an important asset if data quality might be poor. Based on simulations, different object configurations are created in order to find the globally optimal configuration that best fits the input data and the integrated prior knowledge. To evaluate the quality of the object configuration, an energy function is employed, typically consisting of a data and a prior energy. The former measures the consistency of the configuration with the input data, the latter integrates prior knowledge about the arrangement of the objects in the scene. As a result, the object configuration is found for which the energy is minimal. For the detection of objects using MPPs, simple geometric primitives are used (Section 2.2.1); more complex models are also considered (Section 2.2.2). In this context, favouring the sampling of objects at certain locations is commonly applied (Section 2.2.3).

## 2.2.1 Simple geometric primitives

### Single geometric primitives

Often, simple geometric primitives (e.g. circles, ellipses and rectangles) are employed to represent the objects to be detected. This frequently involves the use of one individual primitive, e.g. a circle for the detection of tree crowns.

Circles have been applied to a wide variety of data types, e.g. to detect tree crowns in airborne (Zhang et al., 2013; Aval et al., 2018) and mobile (Yu et al., 2012; Li et al., 2016) laser scanning data sets or remotely sensed images (Perrin et al., 2004), combined with canopy altimetry from airborne Light Detection and Ranging (LiDAR; Zhou et al., 2010). Furthermore, they have been used for nuclei detection in microscopic images (Kowal and Korbicz, 2018) and oil tank detection in synthetic aperture radar (SAR) data (Arslan et al., 2009). In quite a few of these works and, e.g. (Descombes et al., 2009), overlaps between objects are penalized. Arslan et al. (2009), Descombes et al. (2009), Zhou et al. (2010) and Aval et al. (2018) compute the radiometric (grey level) distance between the pixels in the circle and the pixels inside a concentric annulus around it. Likewise, but in the context of elliptical objects, numerous works based on the Bhattacharyya distance have been proposed, e.g. for the detection of flamingos (Gamal-Eldin et al., 2010; Descamps et al., 2011), seed products (Dubosclard et al., 2014), boats in harbours (Craciun and Zerubia, 2013) or cell nuclei (Gadgil et al., 2016). Moreover, Verdié and Lafarge (2014) make use of ellipses for population counting from images (e.g. birds and bees). In the context of tree crown detection from optical aerial images and airborne LiDAR, Perrin et al. (2005) and Andersen et al. (2012) also employ ellipses and ellipsoids, respectively. Besides the penalization of overlap, the works of (Perrin et al., 2005; Craciun and Zerubia, 2013; Aval et al., 2018) make use of objects being locally aligned. In clusters of bomb craters, no specific patterns exist, which is why such prior information cannot be exploited in the scope of this thesis.

Rectangles are frequently used to extract buildings or other human-made objects. In this context, MPPs have been applied to digital surface models (DSMs; e.g. Ortner et al., 2007; Tournaire et al., 2010; Brédif et al., 2013) and remotely sensed images (e.g. Benedek et al., 2012; Chai et al., 2012). In all of these works, except (Chai et al., 2012), a rectangle is included in the object configuration during sampling if high gradient magnitudes at the rectangle border are present while configurations with overlapping rectangles are penalized. Wenzel and Förstner (2016) use rectangles to interpret facades of buildings based on rectified images, where statistics of typical configurations of facade objects (windows, entrances) are learned from training data. Börcs and Benedek (2015) employ rectangles for the extraction of vehicle groups in LiDAR point clouds. In order to improve their results, prior knowledge with respect to the expected vehicle configuration is incorporated. Similarly, Ortner et al. (2007) and Brédif et al. (2013) include assumptions regarding the alignment of rectangles in connection with the extraction of buildings. However, as mentioned

earlier, the integration of such (learned) prior information is not feasible within this work. Rectangles have also been used, for instance, to detect and reconstruct rectangular road marks in high resolution aerial images (Tournaire and Paparoditis, 2009) and oil spills in SAR intensity imagery (Li and Li, 2010).

Cylinders are more rarely employed and have been used, for example, to detect people in multiple calibrated camera views (Utasi and Benedek, 2011) and cosmic filaments from three-dimensional data sets of galaxies (Stoica et al., 2007). While in the works listed so far the objects to be detected are modelled by isolated primitives, in the case of linear networks, individual objects (the line segments with a certain width) are connected. Examples using line segments for facial wrinkle detection, line network, road, river or vascular tree extraction and crack detection can be found in (Jeong et al., 2014), (Lacoste et al., 2005), (Stoica et al., 2004), (Verdié and Lafarge, 2014), (Sun et al., 2007) and (Vandoni et al., 2016), respectively. Chai et al. (2016) represent a linear feature by a string of points instead of a geometric mark. By choosing a graph as object model (Chai et al., 2013; Schmidt et al., 2017), the neighbourhood between individual objects is directly taken into account and thus, for example, no gaps occur at line network intersections.

The description of objects by means of a single geometric primitive has the disadvantage that all objects to be detected in the scene should have similar characteristics, e.g. that tree crowns always appear circular in images. If the assumption with respect to the object's characteristics does not apply and, for example, tree crowns also have an elliptical shape, such objects can hardly be detected. On the other hand, the use of one single geometric primitive may be preferable if there are comparatively few objects of other manifestation, meaning that, for instance, a circle is to be preferred if only a few tree crowns of elliptical shape exist, as this could reduce potential false detections of objects that also appear elliptical in the data. Furthermore, when using a circle, an object is defined by only a few parameters, keeping the search space in the optimization process small and making the model generally more stable.

## Combinations of geometric primitives

In addition to using individual object types, combinations of more than one geometric primitive have also been suggested. In this way, different objects and their possibly different manifestations in the data can be modelled.

Benedek (2017) proposes a method for extracting complex hierarchical object structures using different types of primitives, namely ellipses, rectangles and isosceles triangles. Within this MPP framework, object-subobject ensembles in parent-child relationships are admitted. In (Ortner et al., 2008), a model based on a process of rectangles and segments, respectively, is used for the extraction of buildings from a DSM. The former aims at segmenting homogeneous areas, whereas the latter is dedicated to the detection of linear networks of discontinuities. An energy is defined, favouring alignments of rectangles, connections of segments, and a relevant interaction between both types of objects. Other contributions that also use multiple object types include those related

to SAR image sequences (Benedek and Martorella, 2014), remotely sensed optical imagery (Perrin et al., 2006; Ben Hadj et al., 2010), microscopic images (Soubies et al., 2013; 2015), the latter two (Li et al., 2018) or others (Ben Salah et al., 2018). Based on a library of three geometric primitives (circle, line and rectangle), Lafarge et al. (2010) allow the object type to be changed in the sampling process. Thus, detailed representations of a variety of scenes in terms of their linear and areal components can be produced.

By using different geometric primitives, the detection of objects is not limited to only one particular characteristic. However, due to the larger number of parameters, the dimension of the search space increases, which may lead to a higher number of iterations in the process of finding the optimal object configuration. Furthermore, the complexity of the model is increased. Finally, the use of combined primitives can lead to problems with regard to correct detections, as potentially more disturbing objects (objects that appear similar to the objects to be found in the data) may be detected.

### 2.2.2 More complex object shapes

In some cases, objects cannot be approximated by simple geometric primitives or it is important to describe them more precisely, e.g. in terms of their shape. In this context, Descombes (2017) uses circles and ellipses for cell detection. In order to improve the resulting poor approximation of the cell shapes, the object space may be defined as a dictionary of precomputed shapes. Such a dictionary can be obtained from previous segmentation methods (Poulain et al., 2015) or by constructing an exhaustive description of convex shapes inside a small region (e.g. bounded by 5 x 5 pixels; Cedilnik et al., 2018). Further examples making use of more complex models based on a predefined library of desired objects can, for instance, be found in (Mallet et al., 2010; Hervieu et al., 2015; Ghanta et al., 2018; Zhao et al., 2020). Hervieu et al. (2015) detect road-markings in intensity images derived from laser scanning data from Mobile Mapping Systems. Based on predefined templates for road-markings (e.g. dashed-lines, arrows, characters), the correlation between the distribution of the model's intensity values with those in the data is used to evaluate a proposal in the sampling process. Mallet et al. (2010) and Zhao et al. (2020) use libraries including different parametric functions or road elements to model LiDAR waveforms or extract road networks, respectively. Kim et al. (2019) model narrow gaps as geometric structures called channels. In total, three different channel models are used for its detection in microscopic images. A so-called Delaunay Point Process for the extraction of geometric structures is introduced by Favreau et al. (2019). Their approach simultaneously locates and groups geometric primitives (line segments, triangles) to form extended structures (line networks, polygons) for a variety of image analysis tasks such as line network extraction or object contouring. Another way to achieve a more accurate delineation of the desired objects is presented by Kulikova et al. (2009; 2012). They present an MPP approach for multiple arbitrarily-shaped object extraction (here cell nuclei) by modelling individual object boundaries as closed planar curves. Zhao and Comer (2016) combine

MRFs and MPPs to take advantage of the strengths of each approach (MRFs are useful for imposing local constraints while, on the contrary, it is convenient to model global constraints within the MPP framework). In this way, a precise boundary localization, here of materials, is likewise possible.

The benefit of the methods described above is that, in contrast to modelling objects using simple geometric primitives, a more accurate delineation of the desired objects is possible. However, one drawback is that usually the library of respective objects has to be defined in advance (e.g. Zhao et al., 2020) or, in some cases, is derived by another method (e.g. Poulain et al., 2015). This typically involves considerable effort, as it may not be easy to find appropriate descriptions for the objects to be detected. In addition, the scene specific modelling, such as in (Mallet et al., 2010), limits the transferability to other applications.

### 2.2.3 Favouring sampling of objects at certain locations

In connection with MPPs, knowledge is often integrated such that the generation of objects being focused to specific locations. Respective examples can, for instance, be found in (Utasi and Benedek, 2011; Benedek and Martorella, 2014; Kim et al., 2019) or (Descamps et al., 2008; Arslan et al., 2009; Benedek et al., 2012; Schmidt et al., 2017). The first group makes use of a foreground-background classification, i.e. in the sampling process, objects are only created on foreground pixels. In the work of (Utasi and Benedek, 2011), cylinders are used to model people in 3D space and objects are only generated on a 2D ground plane, derived from multiple calibrated camera views. Benedek and Martorella (2014) segment inverse SAR images into foreground and background classes using a binary Markov Random Field model (Benedek and Szirányi, 2009). The other group derives a probability for the existence of an object for every pixel. In this context, Benedek et al. (2012) use low-level image features, for example local gradient orientations or the homogeneity of grey values. Schmidt et al. (2017), in connection with network extraction, set a threshold for simple raster data features, e.g. the heights in the digital terrain model (DTM) or the grey values in the image. However, for the data used in this thesis there is no height information available and objects such as trees or buildings as well as their shadows make an appropriate threshold selection rather difficult. In the context of planetary crater detection, Solarna et al. (2017; 2020) create a so-called birth map from the available contour map by finding possible ellipse centres midway between contour points via a generalized Hough transform, which are then spread through Gaussian filtering. Bomb craters are generally not characterised by a distinct pair of highlight and shadow regions (Figure 1.1), as planetary craters often are.

If the method for limiting the search space is not chosen appropriately, it is possible that the sampling of objects no longer takes place at locations where objects to be detected are located. This, in turn, may lead to the fact that these objects are not detected at all. On the other hand, a sampling of objects at certain locations adapted to the particular task can also lead to the avoidance of, for example, more complex image contents that can potentially lead to more false detections.

Furthermore, a limitation of the search space to relevant areas in the data results in an accelerated detection of the optimal object configuration and, thus, the computational effort for the MPPs in the overall sampling process is reduced. In addition to the search space limitation, it seems reasonable to also derive information regarding the initialisation of the object model used in the MPP framework in order to further reduce the sampling effort. However, the proposed methods do not allow this.

## 2.3 Bomb crater detection

There is only a limited number of contributions dealing with the detection of bomb craters available in the literature, and none – other than own works (Kruse et al., 2018; 2019; 2020; 2022) – that use MPPs for this task. Generally, aerial wartime images are used, but there have also been works based on contemporary high resolution satellite images and LiDAR.

In connection with aerial wartime images, Lücke et al. (2007) developed a method that directly detects duds semi-automatically using a neural network. Understanding the detection of duds as a binary classification problem, the selection of suitable training samples for the classes dud and background was carried out together with experts in aerial photo interpretation. However, this approach suffers from a rather high misclassification rate due to the easy confusion of duds with other structures, such as one-man holes, cattle watering tanks, chimneys or small image errors. Jensen et al. (2010) use a two-step approach for the detection of bomb craters. First, candidates are extracted, which are afterwards classified by linear discriminant analysis. In connection with the former task, the circle centre detector of Förstner and Gülch (1987) and the SIFT operator (Lowe, 2004) are tested; assuming the shape of bomb craters to be circular. However, the two methods are rejected due to the high number of falsely identified candidates. Instead, an approach based on cross correlation with representative templates for craters is used, which considerably reduces the amount of false candidates. In this context, however, it should be noted that the selection of crater templates must be conducted carefully in order not to exclude craters from the subsequent classification. Especially in view of the fact that bomb craters differ not only in size and spatial intensity patterns, but also in the appearance of the ejecta, such an approach is to be considered disadvantageous. Likewise, the wide range of variations in the appearance of craters in the existing data complicates the use of templates – Jensen et al. (2010) conducted the investigations on fewer than 20 aerial wartime images. Building on (Furlanello et al., 2003), Merler et al. (2005) use a sliding window approach for the extraction of crater proposals from aerial wartime images. For the classification of the proposals, a promising variant of AdaBoost is applied, which aims to reduce the number of false alarms. The result is a map of the spatial density of craters, an indicator for the risk of finding duds. However, their method focuses only at detecting clusters of craters and, thus, the detection of individual craters not belonging to a cluster, the areas of which must also be probed due to the possible occlusion of duds, is omitted.

Brenner et al. (2018) and Clermont et al. (2019), also in connection with aerial wartime imagery, propose methods based on convolutional neural networks (CNNs). While the former authors make use of a sliding window approach for crater candidate extraction similarly to (Merler et al., 2005), the latter employ a blob detector-based approach described in (Mallick, 2022) to lower the number of extracted proposals; the candidates are then classified using a CNN. In this context, Brenner et al. (2018) employ the DenseNet architecture (Huang et al., 2017) in a binary classification setup using the same number of samples for the two classes, crater and background, during training. Testing their approach on data with the same class distribution, a precision of approx. 91 % was achieved. However, in the case of a more realistic class distribution, in which the number of samples for background is much higher than for bomb craters, the precision drops to about 4 %. To counteract this problem, spatial information and a-priori assumptions regarding bomb crater distributions are exploited to filter the output of the CNN. In contrast, the approach of (Clermont et al., 2019), using a pre-trained version of Inception ResNet V2 (Szegedy et al., 2017) for feature extraction, directly considers the imbalance of the class distribution of training samples by assigning higher weights to false detections during training. They show that the use of a weight corresponding to the ratio of positive and negative samples is beneficial for the classification of samples with a realistic distribution, even though the false positive weight is generally not known when applying a trained CNN to new data. However, CNNs usually require a large set of training data to achieve good results. This fact, in addition to the given high appearance variation of bomb craters, is to be seen as a drawback regarding the employment of CNNs (see also Clermont et al., 2019).

In connection with contemporary satellite data of a Cambodian village, Lacroix and Vanhuyse (2015) make use of the circular shape of the craters by applying a circle detection algorithm based on the gradients of the intensity image that has to fulfil compatibility conditions with respect to the gradient angles and their local variations. On the other hand, Lin et al. (2020) address the crater detection task within a two-stage learning-based framework applied to a multispectral WorldView-2 image. A loose statistical classifier based on the histogram of oriented gradients and spectral information is used for a first pass of crater recognition. In a second stage, a patch-dependent spatial feature is developed through dynamic mean-shift segmentation and scale-invariant feature transform descriptors. Finally, LiDAR data was utilized for the creation of highly accurate elevation data; the derived DTMs are then used for crater detection purposes (Foley, 2008; Juhász and Neuberger, 2018; Kazimi et al., 2019; Dolejš et al., 2020). Nowadays, for countries where there has been no warfare for many years, such as Germany, bomb craters will predominantly exist, if at all, in areas unaffected by humans or in forested areas. Thus, with respect to the identification of areas that have to be probed, the utility of contemporary data is limited. Particularly in forested areas, LiDAR data could help because it can partly penetrate vegetation, as in these areas bomb craters are usually not visible in images showing the respective locations. However, the risks associated with duds are mostly related to construction projects, which are planned less frequently in forest areas, so that the latter are of lesser importance.

The methods described above are applied to individual images. Typically, redundancy is not exploited, i.e. information from overlapping image parts is not combined. However, given the fact that aerial wartime images vary considerably in their appearance, using redundant image information seems reasonable. In this regard, to the best of the author's knowledge, only the works of (Lücke et al., 2007) and (Brenner et al., 2018) form exceptions. In the former, as a result of the neural network, the object classes dud and background are differentiated, and the patches classified as duds are subsequently checked interactively by human experts. By using overlapping images, image errors can be eliminated, as they will typically only be present in one image. However, this is a manual step that is not integrated into the framework. This issue is handled by Brenner et al. (2018), who combine the individual detections automatically (the outputs of the CNNs were converted to individual crater positions of each image). For that purpose, the authors apply a neighbourhood-based clustering method and double detections from different images are eliminated in order not to miss any crater in an investigated area. Finally, the clusters are replaced by their centroids. Further details, e.g. the georeferencing accuracy of the images, are not provided. However, given the application scenario proposed within this thesis, it is not reasonable to focus on the detection of each crater, meaning that same objects not being detected in several images are eliminated due to their higher chance of being falsely detected objects (cf. objectives in Section 1.2).

## **2.4 Spatial data analysis**

Spatial data analysis refers to a set of techniques having the ability to manipulate spatial data in different forms and extract additional information as a result. More rigorously, for a technique to be spatial, its results have to potentially differ when the objects of analysis are moved, i.e. location matters. Thus, the data that are subject to spatial data analysis must record the locations of phenomena within some space, which is often the geographic domain (Goodchild, 2008). Section 2.4.1 is concerned with the question how to classify spatial data analysis techniques with respect to the scope of this thesis. The literature concerning the use case of generating an impact map derived from the locations of bomb craters is discussed in Section 2.4.2.

### **2.4.1 Classification of spatial data analysis techniques**

As in (Bailey, 1994), spatial data analysis techniques (Oyana and Margai, 2015) can be distinguished by their data structure. Depending on the data structure, different techniques may be applied. First, locational data, often referred to as object data, consists purely of the locations at which events occur. An example is the locations of cases of some disease within a study area. In the case of attribute data, their values are associated with locations; soil property at certain point locations in a field may serve as an example. Thus, interest is in analysing spatial variation of

attribute values, depending on the locations. Finally, there is interaction data consisting of quantitative measurements, each of which is associated with a pair of locations. An example are the flows of individuals from residences to retail shopping outlets. Consequently, given the above separation, the thesis is concerned with the first group of data, i.e. data in which the point locations of objects are of prime interest. This is often referred to as point pattern analysis (e.g. Gatrell et al., 1996) studying the spatial distribution of points. In addition to (Bailey, 1994), further information on spatial data analysis and its techniques can, for instance, be found in (Anselin et al., 2010).

To study the properties of the point patterns, density-based and distance-based approaches have been developed. In general, these properties can be divided into the two categories of first-order and second-order properties. The former focuses on the characteristics of individual locations and their variations across space, whereas the latter focuses on properties that concern not only individual points, but also the interactions between points and their influences on one another. While distance-based methods consider the distance between point pairs, e.g. the degree of dispersion of a point pattern, and therefore measure second-order properties, density-based methods mostly address first-order properties. As this thesis is concerned with the deduction of area-based information from the locations of individual objects in terms of flagging probing areas and not with how the objects influence each other, techniques based on densities are of relevance; a survey of existing methods regarding density estimation is given in (Silverman, 1986).

A basic measure of a pattern's density is its global density referring to the ratio between the observed number of points relative to the size of the study area. Consequently, such an approach is not capable of capturing local variations. In contrast, local density shows varying point densities at different locations; in the literature commonly used density-based measures are square density and kernel density. In the former, the study area is divided into smaller sub-regions, typically squares or tiles, but also other shapes may be used. Afterwards, the point density is determined in the same way as for the global density, but for each region. However, the result is highly sensitive to the selection of the region shape and size. If small regions are used, one risks having regions with no points, which may prove uninformative, whereas the choice of large regions may lead to missed subtle changes (Anderson and Marcus, 1993). In addition, for the same set of observed data, the densities could vary substantially depending on the delineation of the regions used, which leads to a difficulty in the interpretation of the distribution pattern. This is also known as the modifiable areal unit problem in spatial data analysis (Fotheringham and Wong, 1991). Moreover, each region has a constant density and, thus, local variations within a region are neglected. Finally, such a density surface generally is not continuous as density may change abruptly at the region boundaries. Unlike square density analysis, kernel density estimation (Silverman, 1986) can handle the above issues by constructing a continuous density surface from the observed discrete data based on a kernel function. Kernel density estimation has been applied in various fields, such as archaeology (Baxter et al., 1997; Bonnier et al., 2019), medicine (Rushton et al., 2004; Yang et al., 2006; Lin et al., 2011), transport (Thakali et al., 2015), criminology (Hu et al., 2018) and ecology (Kuter et al., 2011), often for the purpose of hotspot detection and risk mapping. Therefore, kernel density

estimation is particularly well suited for the task pursued in this work, namely the generation of an impact map from detected bomb craters.

### **2.4.2 Impact map generation from detected bomb craters**

In the literature, there are only a few works dealing with the generation of impact maps based on bomb craters. As in the thesis, automatic bomb crater detection may be carried out in advance (e.g. Juhász and Neuberger, 2018); the work of (Tavakkoli Sabour et al., 2014) makes use of manually mapped craters. The development of approaches for the derivation of the impact map is always based on the assumption that the more bomb craters there are in the vicinity, the higher the probability of duds. In this context, Juhász and Neuberger (2018) propose an impact map based on tiles, which are colour-coded according to the level of impact, i.e. the number of detections per tile, following a traffic light system. Especially with regard to flag areas to be probed, approaches based on tiles are not considered to be suitable for several reasons; see also the more general discussion above (Section 2.4.1). On the one hand, the choice of the tile size directly effects the resulting impact map; it may therefore not represent the actual extent of the contamination. More precisely, tiles that are too small result in a situation in which some areas that need to be probed will not be included, while tiles that are too large will result in overly large areas to be probed. In practice, the latter leads to unnecessary costs. In addition, designating probing areas correctly is hindered by the fact that detections can be located in the border areas of tiles, requiring a specific handling for neighbouring tiles. Another tile-based approach is given by Tavakkoli Sabour et al. (2014), who present a method for calculating a per-tile probability of duds according to the distribution and density of exploded bombs. The drawbacks mentioned above, basically the need for an appropriate choice of the tile size and the fact that the designation of probing areas by tile-based approaches is only quite coarse, also apply to their work.

## **2.5 Methods for planetary crater detection**

Planetary craters are among the most important topographic features on planetary surfaces, often formed by the impact of meteoroids. That is why various methods for crater detection have been extensively exploited to study the history of planets (Salamuniccar and Loncaric, 2008; Woicke et al., 2018; Emami et al., 2019). In this regard, catalogues of extracted craters were made publicly available (e.g. Robbins and Hynek, 2012; Wang and Wu, 2020) and are used for planetary science studies, e.g. to analyse the craters characteristics, such as diameter, slope or depth (Liu et al., 2017; Savage et al., 2018) or for planetary age determination based on counted craters (DeLatte et al., 2019; Benedix et al., 2020).

Methods for planetary crater detection can be broadly divided into unsupervised, supervised and hybrid approaches. The former often assume a circular or elliptical shape for the craters and

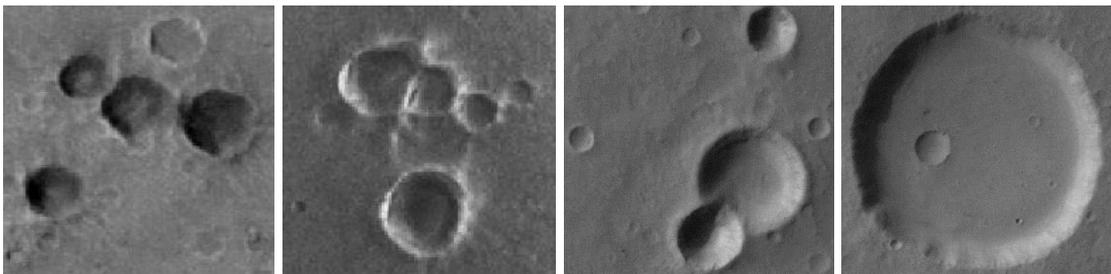
edge information is used for their detection. To detect such regions, the Hough transform (HT) and its variations have been among the most popular approaches. Cross (1988) applies common edge filters to highlight the edges in the image followed by a HT to reconstruct the circular shape of the craters. In a work based on a Martian digital elevation model (DEM), Salamuniccar and Loncaric (2010) also make use of HT and edge detectors. To detect the centres of crater regions, a chord midpoint or generalized HT is used by Meng et al. (2009) and Troglia et al. (2012), respectively. The effect of different pre-processing techniques on the performance of planetary crater detection based on the HT was also evaluated (Honda et al., 2000). Other unsupervised approaches use highlight and shadow regions of the craters as the main features for their detection (Smirnov, 2002; Liu et al., 2015). While the former detects such regions based on thresholding and shadow shape analysis, the latter map salient crater edge features and combine them with ellipse fitting. The works of Bandeira et al. (2007), Pedrosa et al. (2017) and Salih et al. (2017) are based on template-matching algorithms. In the former, a probability volume, determined from the template matching of circular crater templates in a certain range related to the extracted edge images, is analysed to detect the craters. Pedrosa et al. (2017) precede the template matching by morphological image processing. Yue et al. (2013) and, more recently, Chen et al. (2018) make use of topographic analysis and mathematical morphology methods to DEM data of the moon. Also in the context of 3D-data, Huang et al. (2018) employ an alternative parametric active contour model for the detection of lunar craters. To the best of the author's knowledge, there are only three contributions dealing with MPPs in the context of planetary crater detection (Troglia et al., 2010; Solarna et al., 2017; Solarna et al., 2020). All three works model craters as ellipses and pursue the same idea: First, contours are extracted by using the Canny edge detector (Canny, 1986), whereby noise is reduced beforehand by means of smoothing filters. To find the arrangement of ellipses that best fit the edge map in the sampling process, the spatial correlation between each ellipse and the extracted edges is measured and their mutual distance is taken into account. Furthermore, overlapping objects are penalized.

In the context of supervised learning methods used for crater detection, a continuously scalable template-matching algorithm is applied to Martian regions (Vinogradova et al., 2002). Wetzler et al. (2005) compare different supervised approaches for planetary crater detection, namely the previously mentioned template-matching method, support vector machines (SVMs), ensemble methods as well as an unsupervised HT. The approach based on SVMs with normalized image patches provided the best performance and was found to be more accurate than boundary-based approaches such as the HT. SVMs have also been used in (Ding et al., 2013; Chen et al., 2014; Kang et al., 2019). Further supervised methods include boosting (Martins et al., 2009; Bandeira et al., 2012; Jin and Zhang, 2014; Wang and Wu, 2019) and, especially in recent years, CNNs (Cohen et al., 2016; Silburt et al., 2018; Wang et al., 2018; Lee, 2019; Downes et al., 2021). Lee and Hogan (2021) extend the work of Lee (2019) by improving the crater detection algorithm in terms of precision and recall, mainly with the help of post-processing filters. Their algorithm uses optical imagery and DTMs and produces competitive results compared to those from human experts while

being hundreds of times faster. Another very recent work by Tewari et al. (2022) simultaneously utilizes slope maps besides optical images and elevation data.

Hybrid approaches, i.e. combinations of unsupervised and supervised techniques, for planetary crater detection have also been proposed. The procedure of Kim et al. (2005) consist of three stages. First, crater edges are extracted and, secondly, optimal ellipses for craters are evaluated based on a fitness function and refined and verified by template matching. Finally, false detections are removed by a supervised neural network. Urbach and Stepinski (2009), similar to e.g. Chen et al. (2018), use different mathematical morphology-based filters to detect crescent-like regions, which are subsequently classified into crater or non-crater regions by a supervised machine learning technique. Unlike many other works in this field, Emami et al. (2015) assume crater regions to have a nearly convex shape (instead of being almost circular / elliptic or follow the highlight and shadow region assumption). The candidate regions extracted by convex grouping are then verified using CNNs. A recent hybrid approach can be found in (Li et al., 2021), where a random structured forest is trained to detect edges first. Secondly, considering the detected edge information, morphological methods are employed for the determination of crater candidate areas, which are finally classified as either crater or background via a CNN.

Even though the task of planetary crater detection in images seems to be similar to the one of detecting bomb craters, there are some major differences. One key aspect is the high appearance variation, or intra-class variation, of bomb craters, also due to possible fast changes in short time interval (e.g. water in craters) or human intervention (e.g. filling craters with soil). Furthermore, unlike on planets, man-made objects or trees and their shadows often lead to confusion with bomb craters. The poor quality of many aerial wartime images compared to planetary imagery is also worth mentioning. While planetary craters differ strongly in size, bomb craters are small and, thus, blend well into the surrounding terrain. Although planetary craters often have (partly) dark interiors, their appearance is usually different from that of bomb craters. In this context, planetary craters can also exhibit a distinct pair of highlight and shadow regions, which is generally not the case for bomb craters. Finally, larger planetary craters may contain smaller craters. Examples of images illustrating some of the characteristics of planetary craters just mentioned are shown in Figure 2.1; for a comparison with bomb craters see Figure 1.1. Thus, it can be argued that the above approaches are not directly applicable to the task of bomb crater detection.



**Figure 2.1:** Exemplary appearances of craters in planetary imagery.

## 2.6 Discussion

As illustrated in Section 2.1, there are plenty of methods in the literature dealing with the detection of objects in images. Although template matching-based approaches can be implemented fast, they are of limited use if the objects to be detected differ considerably in the data. Even though knowledge-based approaches offer the possibility to address the detection of objects through a coarse-to-fine hierarchical structure, how to define prior knowledge and detection rules is still subjective; too loose rules will cause false positives and vice versa. Another approach is object-based image analysis, whereby the process of segmenting the image into objects can be seen as critical, as the subsequent feature extraction and classification is highly dependent on the quality of the image segmentation. Finally, machine learning-based approaches learn a model of the appearance of the objects in the images from training data, making the detection system scalable and compatible. In contrast to traditional machine learning-based approaches, in which objects are represented by hand-crafted features, in deep learning-based approaches features are learned autonomously from the data, which is beneficial, as the usage of appropriate features is crucial for success. This is why methods based on CNNs having been very popular for several years. However, deep learning-based approaches usually need a large set of labelled training data to yield good results and the generation of such training samples is very time-consuming, especially given the high appearance variation of bomb craters (see also Clermont et al., 2019).

In contrast, the procedure of MPPs does not require any training data at all and allows for the introduction of a strong object model, an important asset when data quality is poor. As illustrated in Section 2.2, there are many ways to describe the objects to be detected. Typically, simple geometric primitives are chosen, while higher flexibility can be achieved by a combination of primitives. However, bomb craters are usually round, sometimes slightly elliptical, and do not exhibit any object-subobject relationships. By using more complex models based on a library of desired objects, a more precise delineation becomes possible. However, this is of less importance, as area-based information to be derived in a subsequent step, i.e. an impact map should be generated from the centres of detected bomb craters (see Section 1). Moreover, such a library would be difficult to obtain, since bomb craters together with their ejecta can have very different appearances, but the assumption of a roundish inner area with darker, homogeneous grey values compared to the surroundings is often justified. Thus, given the aim of detecting bomb craters in aerial wartime images, the use of a simple circular object model seems to be reasonable in the scope of this work. In this way, additionally, objects are defined by only a few parameters, keeping the search space in the optimization process small and potential false detections stemming from objects that appear elliptical in the data may be reduced. Regarding the modelling of the MPP's energy function to be optimized, the Bhattacharyya distance is often used in the literature for the detection of elliptical or circular objects (e.g. Descamps et al., 2011). However, this measure alone does not appear to be suitable for the detection of bomb craters, as their frequently present annulus may differ considerably from image to image, e.g. in terms of its width and appearance. For this reason, a novel

energy function is introduced in this thesis, which further takes into account – among others characteristics – gradient magnitudes along the object border. As discussed in Section 2.2.3, in the context of MPPs it has proven beneficial to favour the sampling of objects at certain locations. Therefore, that idea is pursued within this work, but in contrast to existing research, a blob detector is used to not only limit the search space during sampling but also exploit the blob-like appearance of the bomb craters in the images. In addition to the position, the detector provides information on the size of the associated crater, which is used for the initialisation of the object radius within the MPP framework (see Section 4.1.4).

As discussed in Section 2.3, there is only a limited number of contributions available in the literature dealing with the automatic detection of bomb craters. In (Lacroix and Vanhuyse 2015; Lin et al. 2020) and (Foley, 2008; Juhász and Neuberger, 2018; Kazimi et al., 2019; Dolejš et al., 2020) contemporary satellite and LiDAR data is used, respectively. In the proposed application scenario, due to the large amount of change of land cover and local terrain shape, contemporary data is only useful for very special cases. In this context, e.g. Valjavec et al. (2018) remarked that due to intensive land use, war-related traces such as bomb craters are often not recognizable in present-day LiDAR topography. Particularly in wooded areas, LiDAR data could help because it can partly penetrate vegetation; however, the risks associated with unexploded bombs are mostly related to construction projects, which are planned less frequently in forest areas, characterising the latter as less relevant. Therefore, aerial wartime images are used that are available in abundance for large areas bombed in World War II. In particular, to overcome limitations due to the poor image quality and the complexity of the problem, redundancy is exploited, i.e. the fact that many of the bombed regions were photographed multiple times. Brenner et al. (2018) also make use of redundant image information by combining detection results from individual images. However, unlike in the thesis case, double detections from different images are eliminated in order not to miss any crater in an investigated area. Thus, the idea of supporting detections, i.e. that several detections of the same object in different images are an indication for indeed having found a correct object, is not pursued in the way needed in the proposed application scenario. Consequently, this is addressed in the newly developed procedure.

As pointed out in Section 2.4, the goal pursued in the thesis is to generate an impact map from the locations of the detected bomb craters. In this context, approaches based on tiles in general, including the ones from the literature related to an impact map generation, entail several disadvantages. Besides others, they do not allow for a precise delineation of potentially contaminated areas, which is of great concern due to the costs involved in probing. Therefore, in the thesis, a method is used that allows to represent probing areas more precisely. Moreover, in contrast to (Merler et al., 2005), the primary objective is not only the identification of clusters of craters, i.e. high-risk zones for duds, but there is also a considerable interest in the identification of individual reliable craters that are not part of a cluster, as these areas likewise need to be probed (see Section 1.1). Such less heavily bombed sites are mainly found in rural areas, which have been built on

comparatively rarely since the end of the war when compared to cities, and are thus of great relevance regarding the designation of probing areas for future construction projects. Furthermore, it is usually necessary to completely probe bombed densely built-up areas, as a reliable interpretation (both automatically and manually by experts) of aerial wartime images of such areas is not possible, especially due to coverage by the debris of destroyed buildings. Against this background, the focus of the investigations within this thesis is on rural sites.



## 3 Basics

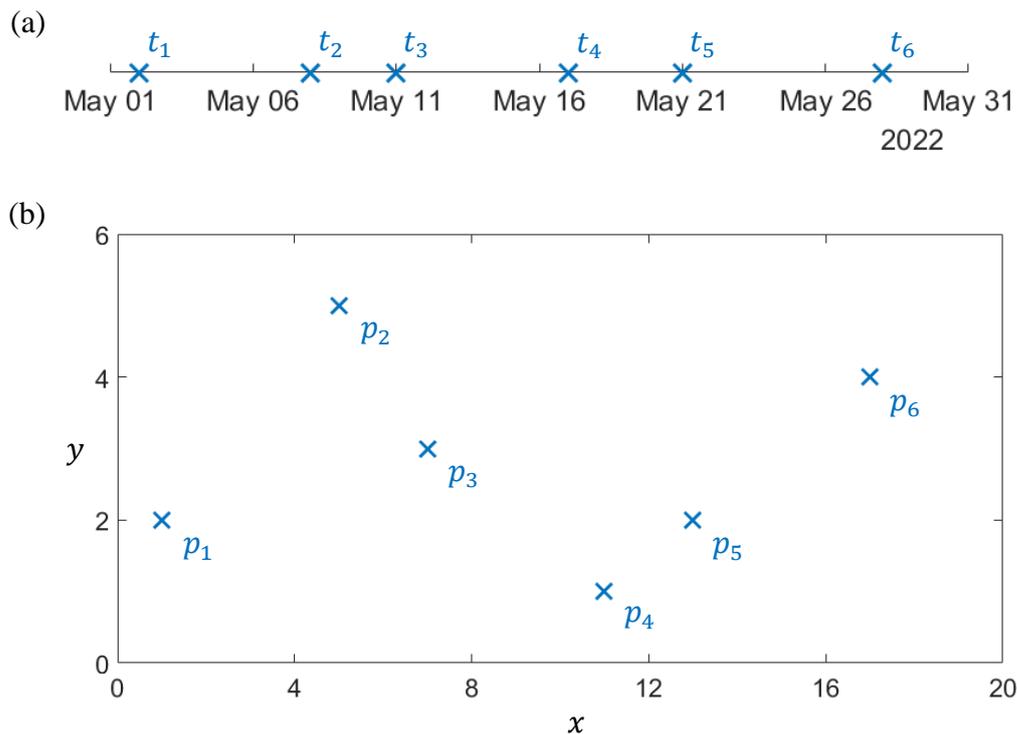
This thesis is concerned with the detection of objects in raster data using marked point processes (MPPs). In this chapter, the mathematical basics of point processes are presented and, in addition, the Poisson point processes as well as the Gibbs energy are discussed (Section 3.1). From the object configurations generated by the MPP, the most likely configuration has to be determined; this optimization problem is solved by Monte Carlo methods, which are described in Section 3.2. Besides Markov Chain Monte Carlo techniques, two extensions, the Reversible Jump Markov Chain Monte Carlo methods and simulated annealing, are discussed. The former allow to model scenes with an unknown number of objects and changes in the number of parameters. Simulated annealing ensures that the procedure converges to the globally optimal object configuration and accelerates the convergence behaviour. As the generation of the impact map from the detected objects is done by kernel density estimation, its basics are described in Section 3.3.

### 3.1 Marked point processes

A point process is a special stochastic process (also referred to as random process) allowing the mathematical description of random events; it is studied in the field of probability theory. In contrast to deterministic approaches, modeling is carried out based on probabilities. For a detailed description of stochastic processes, the reader is referred to Ross (1996), Beichelt and Montgomery (2003) and Florescu (2014); details regarding point processes are given in (van Lieshout, 2000), (Daley and Vere-Jones, 2003) and (Baddeley, 2007).

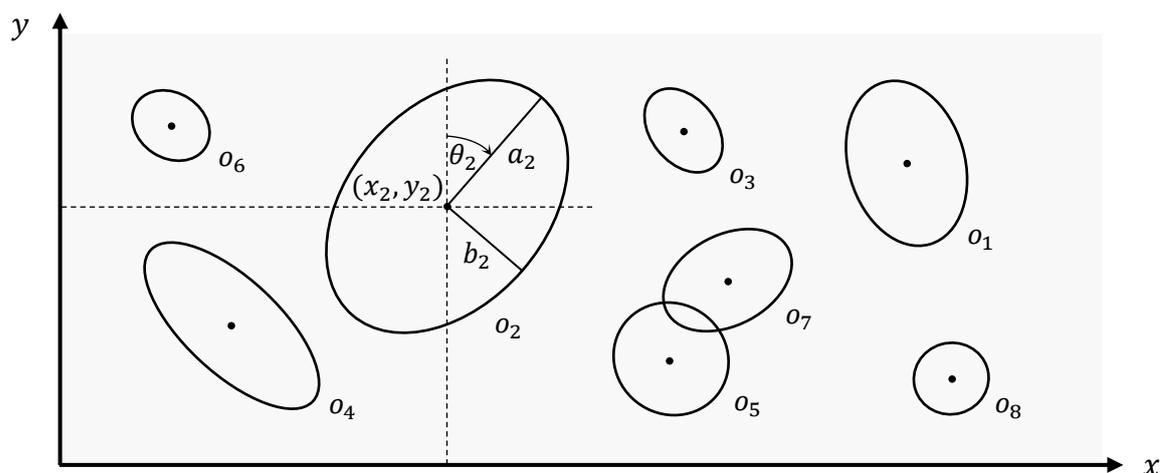
Following probability theory, a stochastic process is a collection of random variables defined on a probability space  $(\Omega, \mathcal{F}, P)$ . Here,  $\Omega$  is the sample space of an experiment, i.e. the set of all possible outcomes of that experiment.  $\mathcal{F}$  represents subsets of  $\Omega$ , also known as events. The probability measure  $P$  is assigned to the events and corresponds to a number between 0 and 1 that expresses how likely the event is to occur. A simple example is the tossing of an ideal coin (Dekking et al., 2005). Assuming that the coin will never land on its rim, there are two outcomes: heads  $H$  and tails  $T$ . Consequently, the sample space is the set  $\Omega = \{H, T\}$ , possible events are  $\mathcal{F} = \{\{H\}, \{T\}\}$  and  $P(\{H\}) = P(\{T\}) = 0.5$ . The random variables all take values in a state space  $S$  that has to be measurable with respect to some  $\sigma$ -algebra (Lamperti, 1977). In the case that the state space is finite or countably infinite, the process is called a point process.

In general, point processes  $X$  can be classified as temporal point processes and spatial point processes. The temporal point process  $X = \{t_1, t_2, \dots, t_n\}$  is one-dimensional and captures the time points  $t_i$  ( $i = 1, 2, \dots, n$ ) of occurrence of events that consist of the times of isolated events scattered in time. As an example may serve the random instants of time when a hospital receives emergency calls (Figure 3.1 a). Temporal point processes have been applied in fields such as seismology, finance and the web to model, for instance, earthquake occurrences, trades and orders and user activities, respectively (Yan et al., 2019; Shchur et al., 2021). In contrast to temporal point processes, a spatial point process  $X = \{p_1, p_2, \dots, p_n\}$  captures the positions  $p_i$  of events in a  $d$ -dimensional space, where  $d \geq 2$ . For example, if a map of the locations of all the people who called the emergency service is created, this map constitutes a random pattern of points in two dimensions (Figure 3.1 b). Among others, applications of spatial point processes can be found in astrostatistics (Babu and Feigelson, 1996) and statistical ecology (Ludwig and Reynolds, 1988), for example to model the position of stars or trees; a general survey of applications is given in (Stoyan et al., 1995). Last but not least, a spatial point process forms the basis of the procedures for object detection in imagery and is used in this thesis.



**Figure 3.1:** To describe a random event, point processes  $X$  may be used. While the temporal point process specifies the times  $t_i$  at which events occur, the spatial point process provides the positions  $p_i$  of occurred events. (a) An exemplary temporal point process indicating the days in May 2022, when a hospital received emergency calls concerning a particular disease, here  $X = \{\text{May } 02, \text{May } 08, \text{May } 11, \text{May } 17, \text{May } 21, \text{May } 28\}$ . (b) An exemplary spatial point process representing the locations of all the emergency calls received, here  $X = \{(1,2), (5,5), (7,3), (11,1), (13,2), (17,4)\}$ .

Within the procedure of spatial point processes, various configurations of points are generated in some space. As the data used in this work are digital images, in the following, the state space  $S$  is given by  $S = \mathbb{R}^2$ , where the positions  $p_i = (x_i, y_i)$  of generated points are restricted by the image extent. As it is often not sufficient to describe a scene solely by points, approaches have been developed to extend each point with a vector of additional parameters  $m_i$  (marks). In this way, the point can be associated with a geometric object  $o_i = (p_i, m_i)$ : The object  $o_i$  is fixed by the position  $p_i$  (often its centre of mass) and the mark  $m_i$ , which can differ from point to point, containing further information about the object. By adding marks to each point, a MPP is created, which can be understood as a stochastic process of configurations of an unknown number of objects of type  $o_i = (p_i, m_i)$  in a bounded region  $F \subset \mathbb{R}^2$ . As the vector  $m_i$  can be used to model any properties of the object, its detailed description is possible, which in turn can be exploited for the detection of objects in images. As already pointed out in Section 2.2.1, simple geometric primitives are often used to represent the objects to be detected. For example, ellipses can be described by their centre, their major and minor semi-axis and their orientation; a possible realisation of a point process  $X$  using ellipses can be found in Figure 3.2.



**Figure 3.2:** A realisation of a marked point process  $X$  using an ellipse as object model, whereby the exemplary configuration consists of a total of eight objects. Each object  $o_i = (p_i, m_i)$  is characterised by its position  $p_i = (x_i, y_i)$  in the image as well as the vector  $m_i = (a_i, b_i, \theta_i)$ , where the semi-major and the semi-minor axis are denoted by  $a$  and  $b$ , respectively, and  $\theta$  is the orientation of the ellipse; here indicated for object  $o_2$ .

### 3.1.1 Poisson point processes

Different assumptions may be made about the distribution of points or objects in space and their relation to each other. To model this distribution, the spatial Poisson point process, i.e. a Poisson point process defined in the plane  $\mathbb{R}^2$  (Merzbach and Nualart, 1986) and, thus, subject of this

thesis, may be used. Considering a bounded region  $F$  of the plane (e.g. an image), the number  $N$  of objects of a point process  $X$  in this region  $F \subset \mathbb{R}^2$  is modelled to be a Poisson-distributed random variable, denoted by  $N_X(F)$ . Then, for a homogeneous Poisson point process with parameter  $\lambda > 0$ , the probability  $p$  of  $n$  objects existing in  $F$  is given by

$$p(N_X(F) = n) = \frac{\lambda(F)^n}{n!} \cdot e^{-\lambda(F)}, \quad (3.1)$$

where  $\lambda$ , often referred to as intensity parameter, describes the expected number of objects within  $F$ . In other words,  $\lambda(F)$  corresponds to the mean of the Poisson distribution for region  $F$ . For a homogeneous Poisson point process, the intensity  $\lambda$  is constant in space, whereas for an inhomogeneous Poisson point process,  $\lambda$  is defined by some location-dependent function.

In conclusion, Poisson point processes induce a complete spatial randomness, given by the fact that the object positions are uniformly and independently distributed. However, this characteristic often does not apply, because dependencies exist between objects. Consequently, more complex models are required to measure the quality of the object configurations.

### 3.1.2 Gibbs energy

In order to allow for a quantitative evaluation of different object configurations, a probability density  $h(\cdot)$  of the MPP can be formulated with respect to a reference point process, which is usually defined as the Poisson point process. In this thesis,  $h(\cdot)$  is expressed through a Gibbs energy  $U(\cdot)$  in the form of  $h \propto \exp - U(\cdot)$ , consisting of two parts, the data energy  $U_D(\cdot)$  and the prior energy  $U_P(\cdot)$ ; their relative influence is controlled by a weight parameter  $\beta \in [0, 1]$ . Thus, the Gibbs energy may be modelled by

$$U(\cdot) = \beta \cdot U_D(\cdot) + (1 - \beta) \cdot U_P(\cdot). \quad (3.2)$$

The conformity of the object configuration with the input data is measured by  $U_D(\cdot)$ . Interactions between the objects are taken into account by  $U_P(\cdot)$ , allowing certain configurations to be favoured or penalized based on prior knowledge. In principle, both the data and the prior energy can consist of an arbitrary number of terms; typical examples of measures related to image analysis that may be incorporated into the terms can be found in Section 2.2, where object detection using MPPs has been discussed.

The aim of the MPP is to find the most likely configuration, i.e. the one having the highest conformity with a predefined model. Such an optimal object configuration  $\hat{X}$  can be determined by maximizing the probability density  $h(\cdot)$ , i.e.  $\hat{X} = \arg \max h(\cdot)$  or, equivalently, by minimizing the Gibbs energy  $U(\cdot)$ , i.e.  $\hat{X} = \arg \min U(\cdot)$ . However, depending on the definition of the

underlying model of the MPP, typically complex processes in high-dimensional spaces evolve. This results in a probability density  $h(\cdot)$ , which is multi-modal and defined in a configuration space with variable dimension, because the number of objects can change. For this reason, simulation methods are typically employed to estimate the global minimum of  $U(\cdot)$ . In this context, Monte Carlo techniques are a widely used and are also applied in this work.

## 3.2 Monte Carlo methods

Monte Carlo (MC) methods are part of probability theory and rely on random experiments being performed a large number of times. Problems that are difficult or impossible to solve with deterministic algorithms are determined numerically with the help of MC simulations based on the law of large numbers. This law states that the average of the results from a large number of trials under the same conditions should be close to the expected value and tends to approach the expected value as more trials are conducted. Consequently, MC techniques provide an approximate inference based on numerical sampling. There are also exact MC sampling methods (Propp and Wilson, 1996). However, for most probabilistic models that are of practical interest, exact inference is intractable. For a detailed description of MC methods, the reader is referred to Neal (1993), Gilks et al. (1996), MacKay (2003) and Bishop (2006).

MC methods are applied in various fields, such as in physical sciences or in finance and business, to simulate fluids or to model phenomena with uncertainties in inputs, e.g. the calculation of risk. This work is concerned with optimization tasks, more precisely with determining the optimum of an arbitrary target distribution  $P(\cdot)$ . Given the law of large numbers, its optimum can be achieved by considering a sufficient number of samples of  $P(\cdot)$ . For drawing the samples, fundamental techniques include importance sampling (Kahn and Harris, 1951) and rejection sampling (von Neumann, 1951). These approaches follow the idea of making use of a simpler distribution, also referred to as proposal distribution  $Q(\cdot)$ , which is easy to sample from. However, importance sampling and rejection sampling only work well if  $Q(\cdot)$  is a good approximation of  $P(\cdot)$ , but in large and complex problems it is difficult to create a single proposal distribution having this property. Here, Markov Chain Monte Carlo methods are more flexible and allow to draw samples from the high-dimensional and complex distributions that usually occur in practical problems.

### 3.2.1 Markov Chain Monte Carlo sampling

Markov Chain Monte Carlo (MCMC; Andrieu et al., 2003) is a technique for generating samples while exploring the state space  $S$  based on a Markov Chain mechanism. This mechanism is constructed in a way that the chain spends more time in the most important regions while satisfying the Markov property. This states that the future of a stochastic process is independent of the past

given the current state. Thus, knowing solely the current state, predictions about the future can be made just as well as if the complete history of the process was already known. In other words, the mechanism of the Markov chain ensures that the drawn samples follow the target distribution  $P(\cdot)$ .

## Markov Chains

A Markov chain consists of random variables  $X_t$ , where  $t$  denotes the time or index in the Markov chain. The variables  $X_t$ , also referred to as states, all take values in the state space  $S$ , whereby the states of the chain are governed by transition probabilities, typically represented within a matrix  $T$ . A transition probability specifies for an arbitrary state the probability of changing to any new state. More formally, when introducing Markov chains in a finite state space  $S$ , where  $X_t$  can only take a certain number  $n$  of discrete values  $X_t \in S = \{x_1, x_2, \dots, x_n\}$ , then the stochastic process is called a Markov chain if

$$P(X_{t+1}|X_t, X_{t-1}, \dots, X_1) = T(X_{t+1}|X_t) \quad (3.3)$$

holds. Consequently, the evolution of the chain in a space  $S$  only depends on the current state  $X_t$  and not on its history  $\{X_1, X_2, \dots, X_{t-1}\}$  (Markov property). In case the transition probabilities remain constant for all  $t$ , it is a homogeneous Markov chain.

An example of a discrete and homogeneous Markov Chain with state space  $S = \{x_1, x_2, x_3\}$  is given in Figure 3.3. For this example, let the transition probabilities be given by

$$T = \begin{bmatrix} 0 & 0.6 & 0 \\ 0.4 & 0.3 & 0.1 \\ 0.6 & 0.1 & 0.9 \end{bmatrix} \quad (3.4)$$

and an initial state  $X_0$  by  $P(X_0) = (0.7, 0.2, 0.1)^T$ , where the column sums add up to one in each case. Then, with the transition between two states of the Markov Chain given by  $P(X_{t+1}) = T \cdot P(X_t)$ , the process evolves to  $P(X_1) = T \cdot P(X_0) = (0.12, 0.35, 0.53)^T$ . When multiplying the respective new state with the transition matrix for several iterations, the chain will stabilize at  $P(\cdot) = (0.10, 0.16, 0.74)$ , regardless of the initial state. The fact that for any initial distribution the chain will converge to the target distribution  $P(\cdot)$ , then called stationary (or invariant) distribution, plays a fundamental role in MCMC simulations. However, prerequisite for convergence towards the stationary distribution, for which  $P(\cdot) = T \cdot P(\cdot)$  holds, is, that the transition matrix  $T$  obeys certain properties.

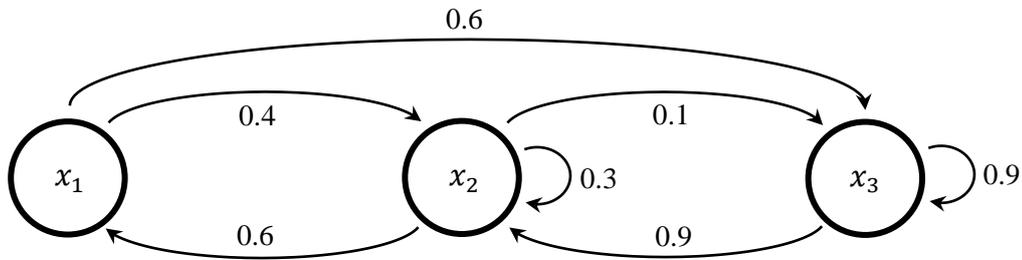
These properties, which can also be used to analyse the behaviour of Markov chains, include aperiodicity, irreducibility and recurrence. The former implies that the Markov chain does not exhibit cyclic behaviour, i.e. the process may return to a specific state at irregular time but, importantly, not at fixed intervals. Irreducibility means that for any state of the chain, there is a

positive probability of visiting all other states. A Markov chain is recurrent if all states can almost surely be reached infinitely often and, thus, recurrence characterises the long-term behaviour of a Markov chain.

In order to construct a Markov Chain for which the distribution one wishes to sample from is stationary, time reversible homogeneous Markov Chains may be used. In this context, the reversibility condition, also referred to as detailed balance, must hold. It states that the transition to a new state  $X_{t+1}$  while the process is in state  $X_t$  must be as likely as the transition to state  $X_t$  when the process is in state  $X_{t+1}$ . In other words,

$$P(X_t) \cdot T(X_{t+1}|X_t) = P(X_{t+1}) \cdot T(X_t|X_{t+1}) \quad (3.5)$$

must hold.



**Figure 3.3:** An example of a discrete Markov Chain given in form of a transition graph. Here, the nodes represent the state space  $S = \{x_1, x_2, x_3\}$  and respective transition probabilities are given by directed edges. For instance, for state  $x_1$ , the probability of a transition to  $x_2$  and  $x_3$  is 0.4 and 0.6, respectively, whereas the probability of maintaining the state is zero.

## Metropolis-Hastings

There are different MCMC samplers, the most popular being the Metropolis-Hastings (Hastings, 1970) algorithm, which is an extension of the Metropolis et al. (1953) algorithm. Indeed, most practical MCMC algorithms can be interpreted as special cases or extensions of this algorithm and, thus, it also provides the basis for Reversible Jump Markov Chain Monte Carlo sampling being employed in this thesis.

The goal of the Metropolis-Hastings algorithm is to generate a collection of states according to the target distribution  $P(\cdot)$ . To accomplish this, the algorithm uses a Markov Chain that fulfils the above given properties (aperiodicity, irreducibility, recurrence) and the condition of detailed balance, such that its stationary distribution is chosen to be  $P(\cdot)$ . As with importance sampling and rejection sampling, samples are drawn from a proposal distribution  $Q(\cdot)$ . In the Metropolis-Hastings algorithm, however,  $Q(\cdot)$  depends only on the current state  $X_t = X$  and not on the previous

states (Markov property). Given a potential new state  $X_{t+1} = X'$ , the proposal distribution  $Q(X'|X)$  may be any fixed distribution (e.g. Gaussian) from which samples can be drawn. In contrast to importance sampling and rejection sampling, it is not necessary for  $Q(X'|X)$  to look similar to the target distribution in order for the algorithm to be practically useful. In addition, a measure for the acceptance of the proposed state  $X'$ , the acceptance probability or acceptance ratio  $\alpha$ , is considered. Providing an initial state  $X_0$ , the Metropolis-Hastings algorithm can be summarised as follows, whereby the following steps are repeated iteratively until a convergence criterion (e.g. a maximum number of iterations) is reached:

- Generate a new state  $X'$  based on the current state  $X$  according to the proposal distribution  $Q(X'|X)$ .
- Compute the acceptance probability  $\alpha$  for the new state  $X'$  via

$$\alpha(X'|X) = \min\left(1, \frac{P(X')}{P(X)} \cdot \frac{Q(X|X')}{Q(X'|X)}\right) \quad (3.6)$$

and sample a uniform random number  $\delta \in [0, 1]$ . Accept the new state  $X'$  if  $\delta < \alpha$ , otherwise reject it, i.e. maintain the current state  $X$ .

Note that the Metropolis-Hastings algorithm allows for drawing samples from non-symmetric proposal distributions. In contrast, the Metropolis algorithm assumes the proposal to be symmetric, i.e.  $Q(X'|X) = Q(X|X')$ . Thus, the acceptance probability  $\alpha$  is simply given by  $\alpha(X'|X) = \min(1, P(X')/P(X))$ . Hence, the transition probabilities for the Metropolis and the Metropolis-Hastings algorithm are given by

$$T(X'|X) = Q(X'|X) \cdot \alpha(X'|X), \quad (3.7)$$

whereas the computation of the acceptance probability depends on the selected algorithm as stated before.

Given the problem of object detection using MPPs, MCMC sampling would only allow determining the optimal configuration of objects if their number is known and constant. However, this is usually not the case, as, for example, when detecting tree crowns within an image (e.g. Zhang et al., 2014), the number of trees is not known in advance. Thus, making use of MCMC would be basically pointless. Here, Reversible Jump Markov Chain Monte Carlo methods provide a solution regarding the outlined drawback.

### 3.2.2 Reversible Jump Markov Chain Monte Carlo sampling

Reversible Jump Markov Chain Monte Carlo (RJMCMC) methods, introduced by Green (1995), form an extension of MCMC and can model scenes with a differing number of objects as well as

changes of the number of parameters between two sampling steps. In other words, RJMCMC sampling allows to construct a Markov Chain of varying dimension and, importantly, to compare the states that may be present in such state spaces of differing dimension. The ability to change the dimension of the state space and to map one space to another is achieved by a move known as reversible (dimension) jump. Here, as the name implies, it must always be possible to return to the previous state. In principle, any reversible type of move can be defined, which makes RJMCMC approaches very flexible. More information on RJMCMC techniques are available in (Richardson and Green, 1997) and (Smith, 2007).

As pointed out before, the mapping of the samples of two different state spaces to a common dimension is achieved by reversible jumps. This move defines a transition between a current state  $X$  and a proposed new state  $X'$  by means of a deterministic, differentiable and invertible function, the transition function  $\psi$ . Moreover, for dimension-matching purposes, auxiliary variables  $u$  and  $u'$  are introduced. Consequently, a transition from  $X$  to  $X'$ , which may involve a change of dimension between the two states, can be described by  $(X', u') = \psi(X, u)$ . To ensure reversibility, the corresponding reverse move, i.e. the function  $\psi^r$  leading from  $X'$  to  $X$ , has to be defined also. In practice, the possibility to return to the previous state is achieved by modelling move types being either a reversible pair of moves or reversible themselves. The latter group comprises the so-called non-jumping transformations that randomly select an object from the current configuration and randomly alter its parameters. For instance, the reverse move for “translation” (shift an object from its current position by a random transition vector) is “translation” itself. Thus, this group includes move types that do not involve a change in dimension. This is different for the other group, the dimensional jumping transformations. Here, to allow a return to the previous state, reversible move pairs are defined appropriately. For instance, one can think of move types being defined by adding objects to and removing them from the configuration, or splitting and merging objects. The move types used within this thesis are detailed in Section 4.1.3.

Starting from an initial state  $X_0$  (as with MCMC), in each iteration, the RJMCMC sampler proposes a change to the current object configuration  $X$  from a set of predefined reversible move types  $m$ . Each move type is associated with a proposal distribution  $Q_m$ , in the following also called kernel. Then, similar to the MCMC algorithm, a new object configuration  $X'$  is generated according to the kernel  $Q_m$  and the acceptance probability  $\alpha$  of the new state is calculated, which for RJMCMC is given by

$$\alpha(X'|X) = \min\left(1, \frac{P(X')}{P(X)} \cdot \frac{q_m^r}{q_m} \cdot \frac{Q_m^r(X, u|X', u')}{Q_m(X', u'|X, u)} \cdot \left| \frac{\partial\psi(X, u)}{\partial(X, u)} \right| \right). \quad (3.8)$$

Note that in the literature the fraction in Equation 3.8 is often referred to as the Green ratio  $G$  (Green, 1995), i.e.  $\alpha(\cdot | \cdot) = \min(1, G)$ . Here,  $P(X)$  and  $P(X')$  correspond to the values of the target distribution in the current state  $X$  and in the new state  $X'$ .  $Q_m(X', u'|X, u)$  is the proposal for a move from  $X$  to  $X'$  and, in analogy,  $Q_m^r(X, u|X', u')$  the proposal for the reverse move from  $X'$  to

$X$ , where  $u$  and  $u'$  are auxiliary dimension-matching variables. The probabilities for choosing kernel  $Q_m$  and its reverse kernel  $Q_m^r$  are given by  $q_m$  and  $q_m^r$ , respectively. The last term  $|\partial\psi(X, u)/\partial(X, u)|$  corresponds to the Jacobian matrix of the transition function  $\psi$  from  $X$  to  $X'$ . Here, it has to be ensured that the derivatives in the matrix exist and that the Jacobian determinant is non-zero. Given that a transition from state  $X$  to a new state  $X'$  can be described by  $(X', u') = \psi(X, u)$ , where  $X$  and  $X'$  contain  $n$  objects, the Jacobian determinant can be determined via

$$\det \left[ \frac{\partial\psi(X, u)}{\partial(X, u)} \right] = \det \left[ \frac{\partial(X', u')}{\partial(X, u)} \right] = \det \begin{bmatrix} \frac{\partial X'_1}{\partial X_1} & \cdots & \frac{\partial X'_1}{\partial X_n} & \frac{\partial X'_1}{\partial u} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial X'_n}{\partial X_1} & \cdots & \frac{\partial X'_n}{\partial X_n} & \frac{\partial X'_n}{\partial u} \\ \frac{\partial u'}{\partial X_1} & \cdots & \frac{\partial u'}{\partial X_n} & \frac{\partial u'}{\partial u} \end{bmatrix}. \quad (3.9)$$

In the process of determining optimal object configurations with RJMCMC, or MCMC methods in general, long simulations can occur, caused by the so-called random walk behaviour. An example of how slowly a random walk can explore the state space can be found in (MacKay, 2003), chapter 29.4. In order to reduce the random walk behaviour and thus to accelerate convergence, different possibilities exist (see e.g. MacKay (2003), chapter 30); one of them is simulated annealing.

### 3.2.3 Simulated annealing

The algorithm of simulated annealing (Kirkpatrick et al., 1983) originates from the observation that solving combinatorial optimization problems is analogous to the annealing (heat treatment that alters the physical properties of a material) of solids in condensed matter physics; its aim is to achieve a minimum energy state of the solid. First, the solid is melted in a heat bath. The following slow decrease in temperature (cooling) allows the solids' particles to rearrange in a crystalline lattice structure that corresponds to the desired minimum energy state for the solid (Varanelli, 1996). Cerny (1985) and Kirkpatrick et al. (1983) independently made use of this principle and proposed the combination of an MCMC algorithm with a temperature parameter to find the global minimum of an energy function. Here, the temperature can be interpreted as a probability with which an intermediate result of the optimization is allowed to get worse. An important advantage of this is that the algorithm would then not get stuck in a local optimum, but rather converge to the global optimum. Since the analogy was discovered, it has been used in numerous optimization problems, which also include applications in image analysis (see e.g. Winkler, 2003).

To be more explicit, simulated annealing involves simulating a non-homogeneous Markov Chain in the form of  $P_t(\cdot) \propto P^{1/T_t}(\cdot)$ , where  $T_t$  is the parameter at iteration  $t$  and also referred to

as temperature. Here, the sequence of temperatures  $T_t$  tends towards zero while  $t \rightarrow \infty$ , i.e.  $\lim_{t \rightarrow \infty} T_t = 0$ . In this way, as already indicated above, the state space can be widely explored at the beginning of the sampling procedure, while with an increasing number of iterations the process becomes more selective, i.e. it operates around the global optimum. Consequently, given the temperature  $T_t$ , also referred to as a decreasing cooling schedule, the calculation of the acceptance probability  $\alpha$  for the RJMCMC algorithm (Equation 3.8) extends to

$$\alpha(X'|X) = \min \left( 1, \frac{P^{1/T_t}(X')}{P^{1/T_t}(X)} \cdot \frac{q_m^r}{q_m} \cdot \frac{Q_m^r(X, u|X', u')}{Q_m^r(X', u'|X, u)} \cdot \left| \frac{\partial \psi(X, u)}{\partial (X, u)} \right| \right). \quad (3.10)$$

There are different cooling schedules, one of which is a logarithmic scheme that guarantees convergence to the global optimum. However, it leads to high computation times and is therefore too slow for a practical application. This is the reason why a number of cooling schemes were proposed that are able to reach low temperatures in finite time (van Laarhoven and Aarts, 1987). Although here the convergence to the global optimum is not guaranteed, these techniques are faster and still provide a good approximate solution (Salamon et al., 2002). The most common cooling schedule in the context of MPPs is based on a geometric sequence (cf. e.g. Mallet et al., 2010; Zhang et al., 2014) that is given by

$$T_t = T_0 \cdot cc_T^t, \quad (3.11)$$

where  $T_0$  is the starting temperature and  $cc_T$  a cooling coefficient for which  $cc_T < 1$  applies. These two parameters have to be chosen according to the given task, whereby two aspects have to be taken into consideration. Firstly, to reduce computation time, the algorithm should not take too long to reach the final temperature close to zero. On the other hand, a wrong choice of parameters can lead to the process not converging to the global optimum, but getting stuck in a local minimum instead. Moreover, a stop criterion *stop\_crit* must be defined, which determines when the execution of the simulation terminates (van Laarhoven and Aarts, 1987). In this context, often a maximum number of iterations is set (e.g. Schmidt et al., 2017), but other criteria can equally be used, such as stopping the algorithm if the last configurations of consecutive Markov chains are identical for a number of chains (e.g. Sechen and Sangiovanni-Vincentelli, 1985).

### 3.3 Kernel density estimation

Kernel density estimation, introduced by Rosenblatt (1956) and Parzen (1962), allows estimating the probability density function of a random variable in a non-parametric way. Unlike the histogram as another non-parametric method for density estimation, the kernel technique produces a smooth estimate of the probability density function and uses all the sample points' locations (see

also discussion in Section 2.4.1). Therefore, kernel density estimates provide an effective way of data representation. For more details regarding non-parametric density estimation and particularly kernel density estimation, the reader is referred to Silverman (1986), Izenman (1991), Simonoff (1996) and Scott (2015).

Given a sample  $(x_1, x_2, \dots, x_n)$  drawn from a distribution with an unknown density  $p$ , an estimate  $\hat{p}$  of this density can be calculated via

$$\hat{p}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right). \quad (3.12)$$

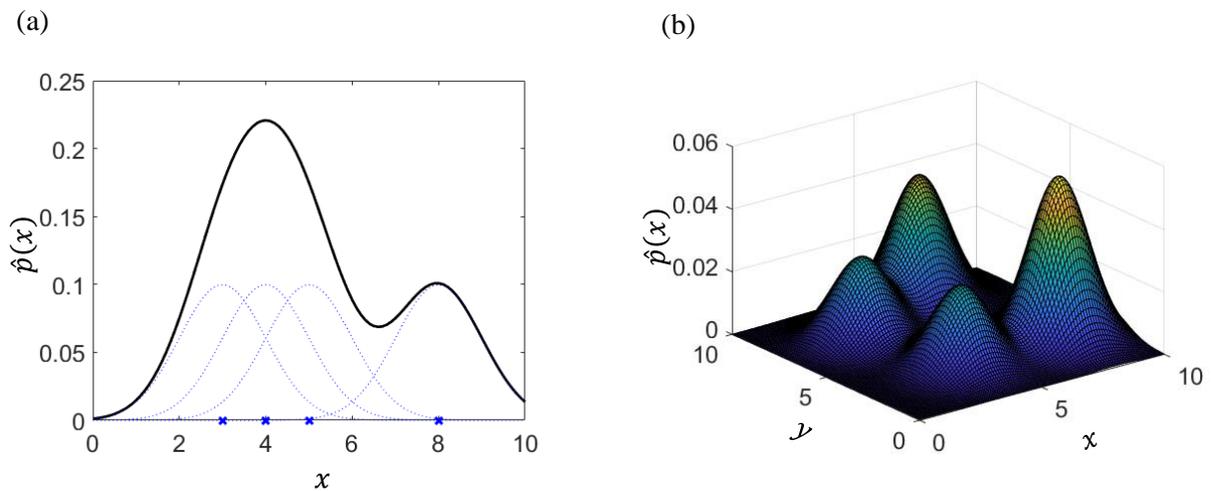
Here,  $h > 0$  is the smoothing parameter, called the bandwidth, and  $K(k)$  is a kernel function (not to be confused with the kernels for RJMCMC sampling from Section 3.2.2) that has to be a non-negative function ( $K(k) \geq 0$ ) that integrates to one ( $\int K(k) dt = 1$ ). In other words, the kernel estimator can be considered as a sum of “bumps” placed on each data point, where the kernel function  $K$  determines the bumps’ shape and the bandwidth  $h$  their width, i.e.  $h$  governs the amount of smoothing applied to the sample. An example showing the individual bumps and the density estimate  $\hat{p}$  constructed by adding them up is given in Figure 3.4 a.

For density estimation, a range of kernel functions may be used. In this context, the Epanechnikov kernel is optimal in terms of minimizing the discrepancy of the density estimator from the true density function (Epanechnikov, 1969). However, the loss in efficiency for other commonly used kernels, such as the Gaussian kernel, the triangular kernel or the rectangular kernel, is comparatively small (Wand and Jones, 1995). Therefore, it is appropriate to base the choice of the kernel function on other considerations, such as computational aspects or domain knowledge of the application. Note that in contrast to the symmetric kernel functions just mentioned, asymmetric functions can also be employed; the work of Chen (2000), who makes use of asymmetric gamma kernels, may serve as an example.

On the other hand, the choice of the bandwidth is much more crucial to the accuracy of kernel density estimators (Hall and Marron, 1987). Here, too large values for the bandwidth may result in oversmoothed information contained in the sample and, thus, important characteristics of the true density (e.g. multimodality) may be hidden. In contrast, too small bandwidth values may cause the estimator to create spurious data artefacts, i.e. valleys and peaks that actually do not exist. For the selection of the bandwidth, a variety of approaches can be used. A natural method is to subjectively choose the bandwidth based on prior ideas with respect to the true density. Here, it can often be convenient to consider kernel density estimates for multiple values of the bandwidth. More objective or data-driven choices of the bandwidth can also be made. To this end, a wide range of methods have been proposed; a description of such techniques can be found, for instance, in Marron (1988). Moreover, there are ways to vary the bandwidth in different regions of the sample space, depending upon the location of either the estimate (balloon estimator) or the samples

(pointwise estimator). For details on such techniques and on the so-called adaptive bandwidth kernel density estimation in general, the reader is referred to Terrell and Scott (1992).

Spatial data analysis (cf. Section 2.4) usually deals with two-dimensional point data, which also applies to this work. Thus, in analogy with the previous discussion of univariate kernel density estimation, one can think in terms of  $n$  points in a plane (e.g. a digital image) defined by coordinates  $x_{(i)} = (x_i, y_i)$  for  $i = 1, 2, \dots, n$ . Here, not a bump is placed at each point, but a “hill” (three-dimensional bump). Then, in analogy, at each point in the plane, the height of the hills is summed up. An example of a kernel density estimate of bivariate data is given in Figure 3.4 b.



**Figure 3.4:** Kernel density estimates  $\hat{p}(x)$  using a normal (Gaussian) kernel and a bandwidth of 1.0 (dotted blue lines in (a), not shown in (b)). (a) An example of kernel density estimation (continuous black line) for univariate data, here consisting of four data points (blue crosses). (b) An example of kernel density estimation for bivariate data, here consisting of six two-dimensional points with its coordinates  $(3, 7)$ ,  $(4, 3)$ ,  $(6, 8)$ ,  $(7, 9)$ ,  $(8, 3)$  and  $(8, 3)$ .



---

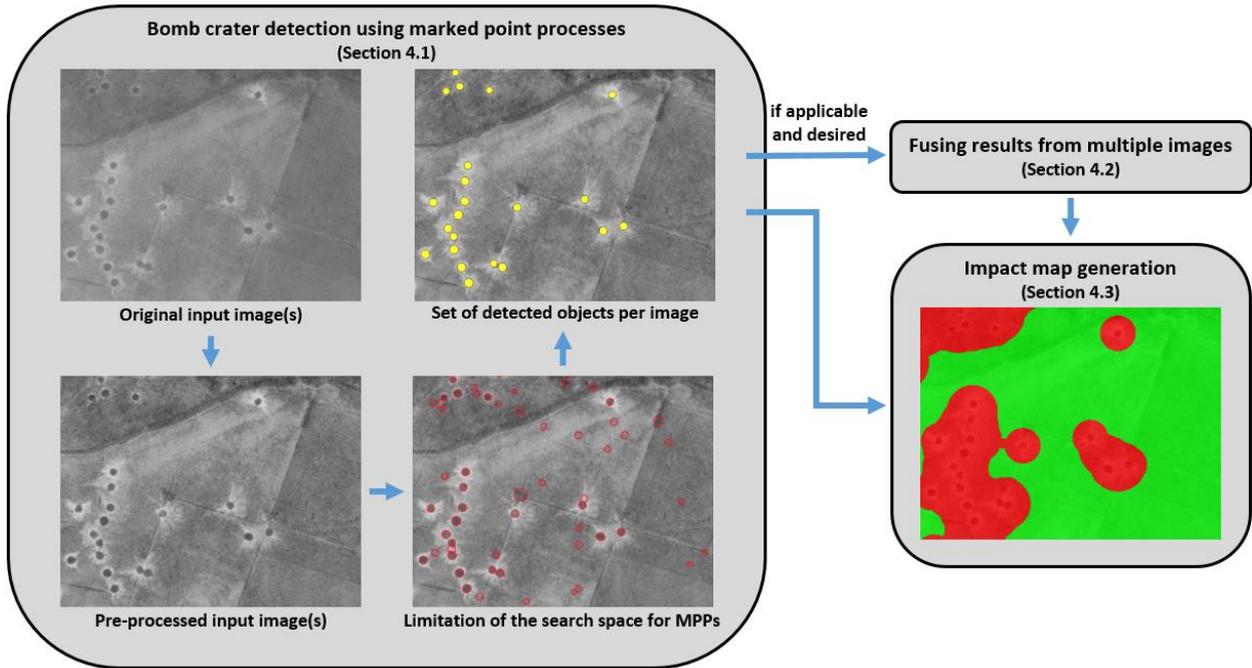
## 4 Generation of impact maps from bomb craters

This chapter introduces a new approach for the automatic detection of bomb craters in aerial wartime images and the subsequent impact map generation. Motivated by the available literature, this work uses marked point processes (MPPs), a model-based probabilistic approach, to deal with the bomb crater detection. Moreover, pursuing one of the research objectives within this thesis, the approach should also exploit redundant image information by merging detection results from multiple overlapping images; the overall workflow is illustrated in Figure 4.1 (note that only a subset of an image is shown). On the basis of pre-processed aerial wartime images (Figure 4.1 bottom left; note that the image shown on the top left is of comparatively good quality), bomb craters are detected using MPPs (Section 4.1). Here, craters are represented as circles and during sampling, the search space in the image is restricted using the blob detector described in Section 4.1.4, providing the coordinate centres of each valid blob as well as its size (Figure 4.1 bottom right, red circles). In the sampling process, high gradient magnitudes along the object borders, homogeneous grey values inside the objects as well as high contrast between the objects and a concentric annulus around them are favoured, while overlapping objects are penalized; the result is a set of detected objects per image (Figure 4.1 top right, filled yellow circles). If applicable, redundancy may be used, i.e. MPP results of multiple images covering the investigated area are combined (Section 4.2). Finally, a probability map for duds based on the centres of the detected (and combined) bomb craters is created by kernel density estimation. By applying a threshold, areas around the detections are classified as contaminated or uncontaminated sites, respectively, resulting in an impact map (Section 4.3). Section 4.4 closes this chapter with a discussion of the proposed approach for the impact map generation from bomb craters.

### 4.1 Bomb crater detection using marked point processes

The MPP procedure is specified by three key components. First, the object model, describing the objects to be detected in the data (Section 4.1.1). Second, the energy, validating the quality of an object configuration based on the image content and the spatial interaction of neighbouring objects (Section 4.1.2). Third, the optimization method, allowing to determine the globally optimal object configuration, i.e. the configuration for which the energy function is minimal (Section 4.1.3). These three components and the way in which they are modelled in the context of bomb crater

detection are addressed in more detail in the referred subsections. In connection with MPPs, knowledge is typically integrated such that the creation of objects is focused to specific locations (cf. Section 2.2.3); the procedure for the limitation of the search space is presented in Section 4.1.4. Afterwards, the pre-processing of the aerial wartime images is detailed, also including the description of a manual masking procedure (Section 4.1.5). The overall conceptual workflow of the bomb crater detection method is outlined in Section 4.1.6.



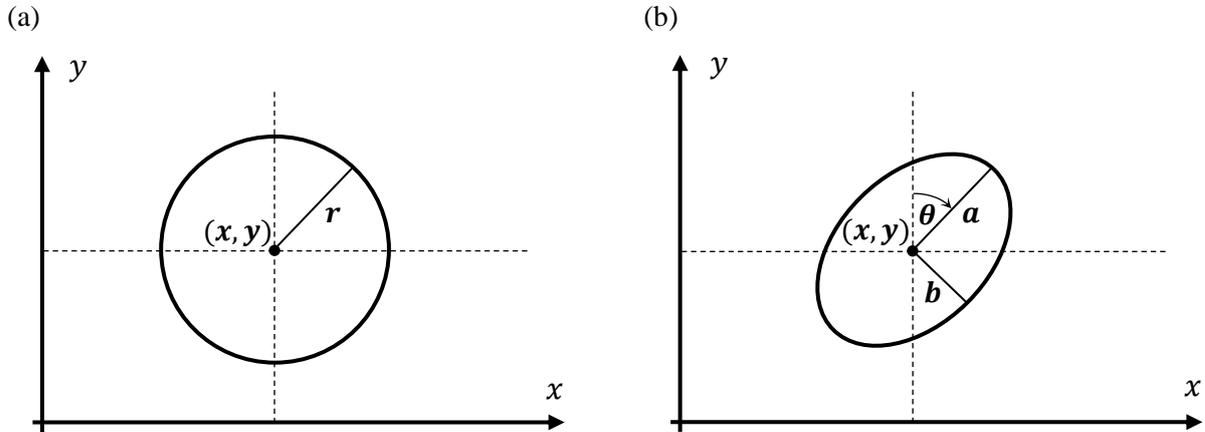
**Figure 4.1:** Workflow of the proposed method for the generation of impact maps from detected bomb craters (for details see main text).

### 4.1.1 Object model

The object model used for the detection of bomb craters is a circle. Each circle is described by its two-dimensional position  $(x, y)$  in the image and a one-dimensional mark, the radius  $r \in [r_m, r_M]$ , where  $r_m$  and  $r_M$  are the minimum and maximum value, respectively (Figure 4.2 a). The circular model enables to work on a small state space  $S \subset \mathbb{R}^3$ , which is advantageous for the simulations.

This aspect is investigated in Section 6.1.2 by experiments comparing the results achieved using a circular and an elliptical model. As illustrated in Figure 4.2 b, an ellipse is described by its two-dimensional position  $(x, y)$  and a three-dimensional mark  $(a, b, \theta)$ . The semi-major and the semi-minor axes are denoted by  $a \in [a_m, a_M]$  and  $b \in [b_m, b_M]$ , respectively, where  $a_m$  and  $b_m$  are the minimum values while  $a_M$  and  $b_M$  are the maximum values. Here,  $a \geq b$  as well as  $b > a/1.5$  are required, because bomb craters usually are almost circular. Moreover, if there is no constraint

on the value of  $b$ , elongated objects, such as roads, may be detected. The orientation of the ellipse is described by the angle  $\theta \in [0, \pi[$ .



**Figure 4.2:** Two different object models, a circle and an ellipse, used for the detection of bomb craters in aerial wartime images. (a) The circle, being the default object model used within this thesis, is described by its position  $(x, y)$  and the radius  $r$ . (b) The ellipse is described by its position  $(x, y)$  as well as the semi-major axis  $a$ , the semi-minor axis  $b$  and the orientation of the longer semi-axis  $\theta$ .

### 4.1.2 Energy function

In order to evaluate each object configuration, the current configuration is compared with the new one based on the Gibbs energy (see Equation 3.2), describing the consistency of the object configuration  $X$  with a predefined model for the objects to be detected. The energy function  $U(X)$  consists of two parts, the data energy and the prior energy. While the former measures the conformity of the objects with respect to the given image, in the latter spatial interactions between the objects are taken into account, allowing certain configurations to be favoured or penalized based on prior knowledge. Basically, the energy function can consist of an arbitrary number of terms. In this thesis, it is expressed as the sum of four terms. Each term will be described in the next subsections.

Given the overall aim of determining the global minimum of the energy function, object configurations that appropriately describe the bomb craters to be detected in the images should reduce the value of the energy function. In contrast, unfavourable configurations should increase its value, meaning that such object configurations are less likely to be accepted in the sampling process.

#### Data energy

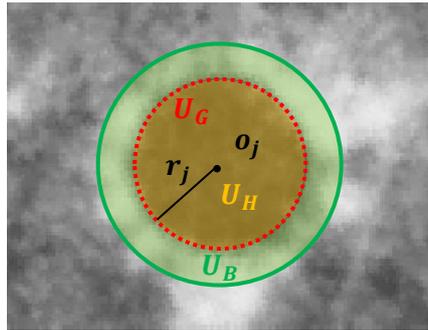
The data energy  $U_D(X)$  checks the consistency of the object configuration with the input data. Bomb craters are predominantly characterized by locally darker grey values in comparison with

the surrounding area. This is mainly due to the frequently circular shadow within the bomb craters cast by the sun. Moreover, due to the earth often ejected by the detonation, there may be a band of typically brighter pixels around the detonation hole, even though that annulus may differ considerably from image to image, e.g. in terms of its width and appearance (Figure 1.1). Still, it is expected that this will help to differentiate bomb craters from other objects. Consequently, bomb craters are assumed to have high gradient magnitudes in the transition region from the detonation hole to its surroundings, more or less homogenous grey values inside the detonation hole as well as a high contrast between their inside and a concentric annulus around it. In the thesis, the data energy is thus modelled by

$$U_D(X) = U_G(X) + U_H(X) + U_B(X), \quad (4.1)$$

where each term represents one of the three assumptions regarding bomb craters stated above, i.e.  $U_G$  considers the gradient magnitudes at the object borders,  $U_H$  the homogeneity of grey values inside the objects and  $U_B$  the contrast between the objects and a concentric annulus around them (see Figure 4.3).

Note that the grey values in a square window centred over the object's centre and having a side length of  $2 \cdot (r + n_m)$  are normalized locally to the interval  $[0, 255]$ , because experiments have shown that a normalization is beneficial for the quality of the detection results. The parameter  $n_m$  is empirically set to  $n_m = 5$  m and should ensure that the complete part of the inside of the crater as well as some surrounding pixels (typically including the craters annulus and some additional pixels) are contained in the window.



**Figure 4.3:** Illustration of the three terms of the data energy for an object  $o_j$  with radius  $r_j$ , where  $U_G$  considers the gradient magnitudes at the object border,  $U_H$  the homogeneity of grey values inside the object and  $U_B$  the contrast between the object and a concentric annulus around it.

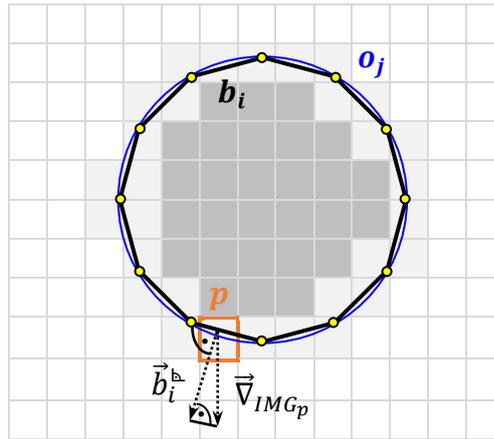
**High gradient magnitudes:** According to the previously made assumptions, a newly created or modified object leads to a reduction of the data energy if high gradient magnitudes occur along the

edges of the object, i.e. the shape of the circle fits to the border of the shadow. The gradients along the border of the circle are determined and the corresponding data term is modelled by

$$U_G(X) = f_G \cdot \sum_{o_j \in X} \left( c - \sum_{i=1}^{n_v} \frac{1}{n_{b_i}} \sum_{p=1}^{n_{b_i}} \nabla_{IMG_p}^b \right). \quad (4.2)$$

Here,  $\nabla_{IMG_p}^b = \langle \vec{\nabla}_{IMG_p}, \vec{b}_i^b \rangle$  is the component of the grey value gradient at the border pixel  $p$  in the direction of the normal vector  $\vec{b}_i^b$  of object  $o_j$  pointing outside (Figure 4.4). The border of the object is approximated by a polygon with a constant number of  $n_v$  vertices and the gradient magnitudes are summed up for all pixels  $p$  along the object edges  $b_i$ , with  $n_{b_i}$  representing the corresponding number of pixels; the gradient magnitude of each pixel is equally weighted. The whole term is weighted by a positive factor  $f_G$  and the constant  $c \geq 0$  ensures that the energy only decreases if the sum in Equation 4.2 is larger than  $c$ . Without considering  $c$ , objects with very small gradient magnitudes at the object border would already reduce the energy, so that the optimal configuration would consist of an extremely large number of (mostly false positive) objects.

Obviously, the data energy would decrease infinitely if a superimposition of objects on attractive locations, i.e. locations that lead to an energy reduction, was possible. Consequently, if this accumulation of objects is not prevented, the algorithm will not converge due to the energy constantly decreasing. To avoid this problem, a term that penalizes the overlap of objects is included in the prior energy (see below).

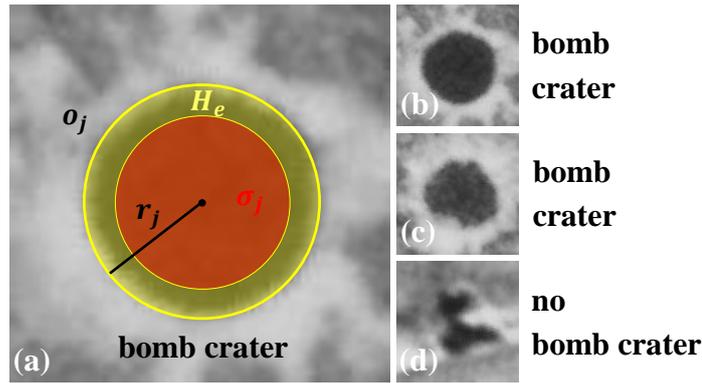


**Figure 4.4:** To calculate the first data term, the border of the object  $o_j$  is approximated by a polygon with a constant number of vertices (here twelve; yellow dots) and the gradient magnitudes in direction of the normal vector of the object edges  $b_i$  pointing outside are summed up for all pixels  $p$  along the border (light grey). Dark grey pixels indicate the expected darker area within a bomb crater.

**Homogeneous grey values:** The second term of the data energy from Equation 4.1 requires the grey values inside the object to be homogeneous. Homogeneity is measured by the standard deviation  $\sigma$  of the grey values within the object, which is assumed to be higher for a falsely detected object than for a bomb crater (Figure 4.5 b-d). A new or modified object  $o_j$  increases the energy if  $\sigma_j$  is higher than a predefined threshold  $H_t$ , which results in

$$U_H(X) = f_H \cdot \sum_{o_j \in X} \max(0, \sigma_j - H_t). \quad (4.3)$$

In Equation 4.3, the energy term is weighted by a positive constant  $f_H$ . For the computation of  $\sigma_j$  the grey values of all pixels inside the object  $o_j$  are considered. In this context, a certain percentage of border pixels are excluded to allow the shapes of bomb craters to deviate slightly from the geometrical model of a circle; this possibility is controlled via a parameter  $H_e$  (Figure 4.5 a).



**Figure 4.5:** (a) Homogeneity of an object  $o_j$  with radius  $r_j$  is measured by the standard deviation  $\sigma_j$  of the grey values in the red area; pixels in  $H_e$  (yellow) are excluded. (b) Bomb craters with a circular and (c) not quite circular shape; (d) shows a background object with a high standard deviation.

**High contrast:** The last term of the data energy requires high contrasts between the object  $o_j$  and their concentric annulus  $C(o_j)$ . For this purpose, i.e. to highlight areas with distinct contrast in the data, the Bhattacharyya distance  $d_B$  is used, which is given by

$$d_B(o_j, C(o_j)) = \frac{(\mu_{obj} - \mu_{annu})^2}{4 \cdot \sqrt{\sigma_{obj}^2 + \sigma_{annu}^2}} - \frac{1}{2} \cdot \log \left( \frac{2 \cdot \sigma_{obj} \cdot \sigma_{annu}}{\sigma_{obj}^2 + \sigma_{annu}^2} \right), \quad (4.4)$$

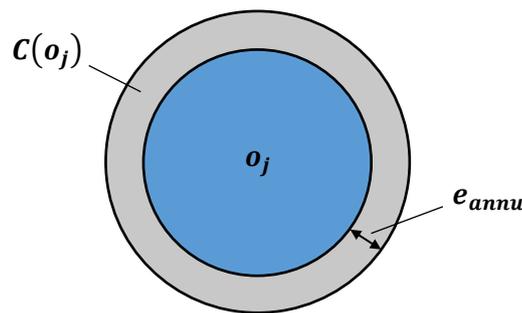
where  $\mu_{obj}$  and  $\sigma_{obj}$  are the mean and the standard deviation computed based on the grey values of all pixels being part of the object. Accordingly,  $\mu_{annu}$  and  $\sigma_{annu}$  are the mean and the standard deviation of the grey values within the annulus. In connection with the annulus, a parameter  $e_{annu}$  describing its extent, is introduced (see Figure 4.6). For instance, a value of  $e_{annu} = 5$  means that all pixels within the concentric annulus of 5 pixels in diameter are considered for the computation of  $\mu_{annu}$  and  $\sigma_{annu}$ . The data energy term, weighted by the positive factor  $f_B$ , is defined via

$$U_B(X) = f_B \cdot \sum_{o_j \in X} Q(d_B(o_j, C(o_j))), \quad (4.5)$$

with  $Q(d_B) \in [-1, 1]$  a quality function that favours or penalizes the object configuration according to

$$Q(d_B) = \begin{cases} 1 - \frac{d_B}{d_0} & \text{if } d_B < d_0 \\ \exp\left(\frac{d_0 - d_B}{100}\right) - 1 & \text{if } d_B \geq d_0 \end{cases}, \quad (4.6)$$

where  $d_0$  is a threshold. Consequently, objects with a distinct contrast between their inside and their annulus, i.e.  $d_B > d_0$ , are favoured and the associated data energy becomes negative. In this context, higher values for the parameter  $d_0$  will of course lead to a more selective object fitting, resulting in fewer objects being included in the final object configuration.



**Figure 4.6:** For the computation of the Bhattacharyya distance the grey values of all pixels inside the object  $o_j$  (blue area) and its annulus  $C(o_j)$  (grey area) are considered; the extent of the annulus is controlled via  $e_{annu}$ .

## Prior energy

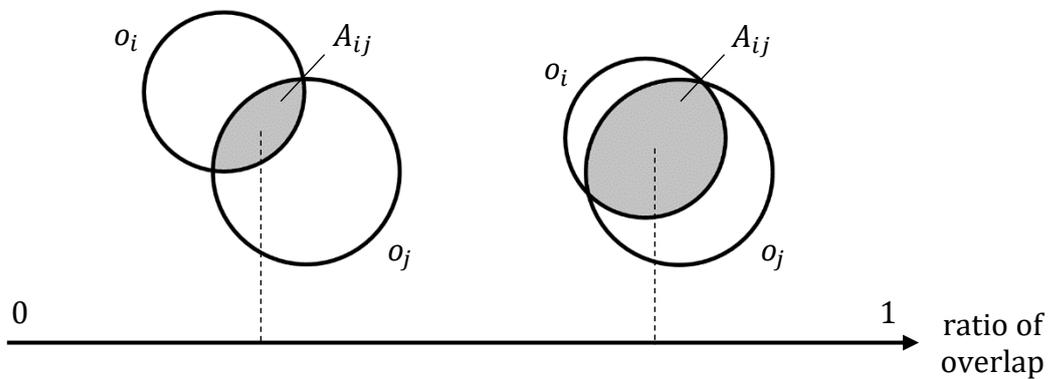
The prior energy  $U_p(X)$  integrates prior knowledge into the model, allowing to favour or penalize certain object configurations. In this context, certain aspects of the configuration, e.g. with respect to the geometric relation of neighbouring objects, may be evaluated by different terms of  $U_p$ . For MPPs, it is in general reasonable to penalize the overlap of individual objects (see Section 2.2). As bomb craters or their clusters usually do not exhibit any particular patterns, the prior energy consists of one term only, resulting in

$$U_p(X) = U_o(X), \quad (4.7)$$

where  $U_o$  represents the prior energy term that penalizes the overlap of objects.

**Non-overlapping objects:** To avoid the accumulation of objects in regions where the data terms lead to a reduction of the energy, configurations with overlapping objects are not prohibited but penalized, as, obviously, bomb craters can in principle overlap to any extent (see also the discussion in Section 4.4 regarding the separate detection of each crater). For this purpose, an overlapping object pair  $o_i, o_j$  is considered (Figure 4.7). The overlapping areas  $A_{ij}$  of the objects  $o_i$  and  $o_j$  as well as the respective relative overlapping areas  $A_{ij}/A(o_i)$  and  $A_{ij}/A(o_j)$  are computed. Here,  $A(o_i)$  and  $A(o_j)$  are the areas of the objects  $o_i$  and  $o_j$ , respectively. The prior energy term with positive weight  $f_o$  becomes

$$U_o(X) = f_o \cdot \sum_{o_i, o_j \in X} \max\left(\frac{A_{ij}}{A(o_i)}, \frac{A_{ij}}{A(o_j)}\right). \quad (4.8)$$



**Figure 4.7:** The prior energy term penalizes overlapping object configurations depending on their degree of overlap. The larger the common area, the higher the prior energy will be.

### 4.1.3 Changes in the object configuration

The task is to find the object configuration that minimizes the energy function  $U$ , the components of which have been introduced in Section 4.1.2. As this function is of high complexity and its dimensionality is not known due to the unknown number of objects, to find the global optimum, optimization is carried out by means of Reversible Jump Markov Chain Monte Carlo sampling (Section 3.2.2) coupled with simulated annealing (Section 3.2.3). During sampling, for the transition from an object configuration to a new one, potential changes and corresponding kernels  $Q_m$  have to be defined. In this work, transformations are used that involve and do not involve a change in dimension, respectively. More precisely, in each iteration, one of the two following move types is randomly selected:

- *Birth-and-death*: An object is added to or removed from the configuration; this involves a change in dimension. This move type comprises the moves *birth* and *death*, forming a reversible pair, in which one move is the reverse of the other.
- *Modification*: The object's parameters are modified; this does not involve a change in dimension. This move type comprises the moves *translation* and *mark-variation*.

The move types used in this thesis are described in more detail in the subsections below. Theoretically, adding objects to or removing them from the configuration would be sufficient to simulate a MPP. However, to increase the convergence speed, it is reasonable to use also non-jumping transformations, as, for instance, a single translation is more efficient than a death followed by a birth, leading to the same result. Then, the new configuration  $X'$  is evaluated depending on the current configuration  $X$  using the Green ratio (cf. Section 3.2.2). Subsequently, the new state is accepted with acceptance probability  $\alpha$  (Equation 3.10), which, accounting for the energy function  $U$ , is given by

$$\alpha(X'|X) = \min \left( 1, \exp \left( -\frac{U(X') - U(X)}{T_t} \right) \cdot \frac{q_m^r}{q_m} \cdot \frac{Q_m^r(X, u|X', u')}{Q_m(X', u'|X, u)} \cdot \left| \frac{\partial \psi(X, u)}{\partial (X, u)} \right| \right). \quad (4.9)$$

In connection with the Jacobian matrix in Equation 4.9, in the following subsections it will also be shown that the absolute value of the Jacobian determinant (Equation 3.9) reduces to unity for all moves.

#### Birth-and-death

During sampling, an object configuration is iteratively created by adding objects to or removing them from the configuration. In this context, while the *birth*-move adds an object, the *death*-move allows to return to the previous state by removing it. The corresponding changes to the object configuration induced by this reversible move pair are modelled by the kernel  $Q_{BD}$ .

Within the *birth*-move, the position of a new object is sampled from likely positions for bomb craters detected using the blob detector described in (Mallick, 2022; see also Section 4.1.4; Figure 4.8). As pointed out before, this procedure provides information about the size of the associated crater, which is used for the initialisation of the circle radius. Given Equation 3.1 related to the Poisson point process and the fact that one object is added to the current configuration  $X$ , resulting in a new configuration  $X'$  of  $n' = n + 1$  objects, the kernel ratio from Equation 4.9 reduces to the probability  $p$  for the number of objects in the respective configuration and is therefore

$$\frac{Q_D(X, u|X', u')}{Q_B(X', u'|X, u)} = \frac{p(n|n')}{p(n'|n)} = \frac{\lambda^{n'} \cdot e^{-\lambda}}{n'!} \cdot \frac{n!}{\lambda^n \cdot e^{-\lambda}} = \frac{\lambda^{(n+1)} \cdot n!}{n! \cdot (n+1) \cdot \lambda^n} = \frac{\lambda}{n'} , \quad (4.10)$$

where  $\lambda$  corresponds to the expected number of objects in the image. Similarly, for the *death*-move, a randomly selected object is removed (Figure 4.8), i.e. the number of objects is reduced by one ( $n' = n - 1$ ), and the kernel ratio results in

$$\frac{Q_B(X, u|X', u')}{Q_D(X', u'|X, u)} = \dots = \frac{\lambda^{(n-1)} \cdot n!}{(n-1)! \cdot \lambda^n} = \frac{\lambda^n \cdot n! \cdot n}{\lambda \cdot n! \cdot \lambda^n} = \frac{n}{\lambda} , \quad (4.11)$$

where  $n$  is the number of objects in the current object configuration  $X$ .

To recap, the transition from  $X$  to  $X'$  is modelled by a transition function  $\psi$  that includes auxiliary variables  $u$  and  $u'$ . Thereby, the Jacobian determinant of  $\psi$  is of relevance, as it affects the calculation of the acceptance probability. In the following, using Figure 4.8, it is exemplarily shown how the value of the Jacobian determinant is computed in the case of the *birth*- and *death*-move, respectively. Note that all parameters indicated with a prime belong to the proposed new configuration  $X'$ . In the *birth*-move, object  $o'_3$  is added to the current object configuration  $X = \{o_1, o_2\}$ , forming a proposed new state  $X' = \{o'_1, o'_2, o'_3\}$ , while all other objects (here  $o_1$  and  $o_2$ ) are kept fixed. Thus, with the auxiliary variable  $u$  acting as a placeholder for the missing dimension in  $X$ , the transition from  $X$  to  $X'$  can be expressed via

$$X = \begin{Bmatrix} o_1 \\ o_2 \\ u \end{Bmatrix} \rightarrow X' = \begin{Bmatrix} o'_1 = o_1 \\ o'_2 = o_2 \\ o'_3 = u \end{Bmatrix} \quad (4.12)$$

and the Jacobian determinant reduces to

$$\det \left[ \frac{\partial(X')}{\partial(X, u)} \right] = \det \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} = 1. \quad (4.13)$$

The *death*-move, being the reverse of the *birth*-move, is designed in a way that it may revert the state back to the previous configuration. This is achieved by applying the *death*-move to the proposed configuration  $X'$  generated by the *birth*-move. However, for the sake of consistency, the following notation treats the *death*-move as applied to an initial configuration  $X$ . Thus, in analogy to the *birth*-move, the transition can be represented by

$$X = \begin{pmatrix} o_1 \\ o_2 \\ o_3 \end{pmatrix} \rightarrow X' = \begin{pmatrix} o'_1 = o_1 \\ o'_2 = o_2 \\ u' = o_3 \end{pmatrix}. \quad (4.14)$$

Note that as with Equation 4.12, Equation 4.14 refers to the exemplary object configuration from Figure 4.8, where object  $o_3$  is removed using the auxiliary variable  $u' = o_3$  and objects  $o'_1$  and  $o'_2$  remain unchanged from their previous state  $X$  (note that the notation in Figure 4.8 is given according to the *birth*-move; for the *death*-move it would be reversed). Again, the Jacobian determinant reduces to unity, as it is determined via

$$\det \left[ \frac{\partial(X', u')}{\partial(X)} \right] = \det \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} = 1. \quad (4.15)$$

Consequently, with the kernel ratios from Equations 4.10 and 4.11 as well as the Jacobian determinants equal to unity in both cases, the acceptance probability for the *birth*-event  $\alpha_{birth}$  is given by

$$\alpha_{birth}(X'|X) = \min \left( 1, \exp \left( -\frac{U(X') - U(X)}{T_t} \right) \cdot \frac{q_D}{q_B} \cdot \frac{\lambda}{n'} \right) \quad (4.16)$$

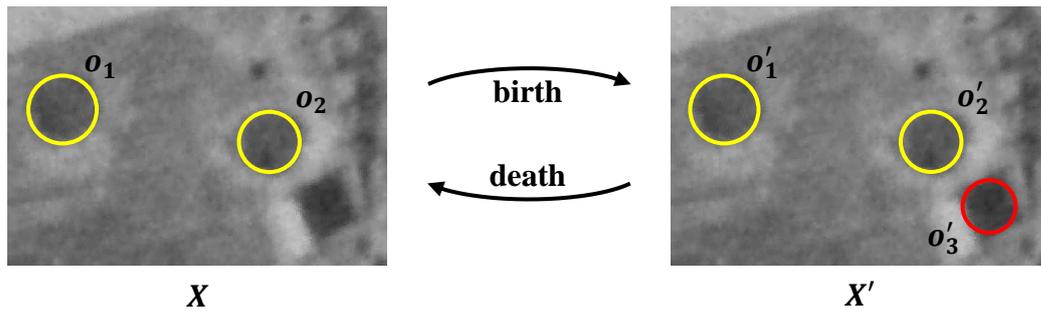
and for the *death*-event  $\alpha_{death}$  by

$$\alpha_{death}(X'|X) = \min \left( 1, \exp \left( -\frac{U(X') - U(X)}{T_t} \right) \cdot \frac{q_B}{q_D} \cdot \frac{n}{\lambda} \right), \quad (4.17)$$

where  $q_B$  and  $q_D$  are the probabilities for choosing the *birth*- and the *death*-move, respectively.

## Modification

Besides adding objects to or removing them from the configuration, the parameters of objects can be modified. Here, an object is allowed to change with respect to its position (move: *translation*)



**Figure 4.8:** Move type *birth-and-death* based on an exemplary configuration  $X$  consisting of two objects  $\{o_1, o_2\}$ . In the *birth*-move, object  $o'_3$  is added to  $X$ , forming a new state  $X' = \{o'_1, o'_2, o'_3\}$ . To return to the previous state  $X$ , the *death*-move is applied to  $X'$ , removing object  $o'_3$  again.

or mark (move: *mark-variation*), while all other objects are kept fixed. These two modifications, illustrated in Figure 4.9, are assumed to be equally probable, and one of them is chosen randomly. For a *translation*-move (Figure 4.9 upper row), a randomly chosen object is shifted from its current position by a random transition vector  $\vec{t}$ . The movement is realized in a given interval based on a uniform distribution. For a *mark-variation*-move (Figure 4.9 lower row), a circle of the current configuration is randomly selected and its radius is changed to a new value drawn from a uniform distribution within a predefined interval. Impossible proposals, e.g. a *translation*-move that moves the object (partly) out of the image, are not used. In this case, the respective move is rejected.

Again, the determination of the Jacobian determinant is illustrated by using Figure 4.9. In this example, the parameters of object  $o'_2$  (either the position within the *translation*-move or the radius within the *mark-variation*-move) are modified using the auxiliary variable  $u$ . Given that the move type *modification* is self-reversible,  $u'$  is the reverse counterpart of  $u$  and the transition from  $X$  to  $X'$  can be expressed via

$$X = \begin{Bmatrix} o_1 \\ o_2 \\ u \end{Bmatrix} \rightarrow X' = \begin{Bmatrix} o'_1 = o_1 \\ o'_2 = u \\ u' = o_2 \end{Bmatrix} \quad (4.18)$$

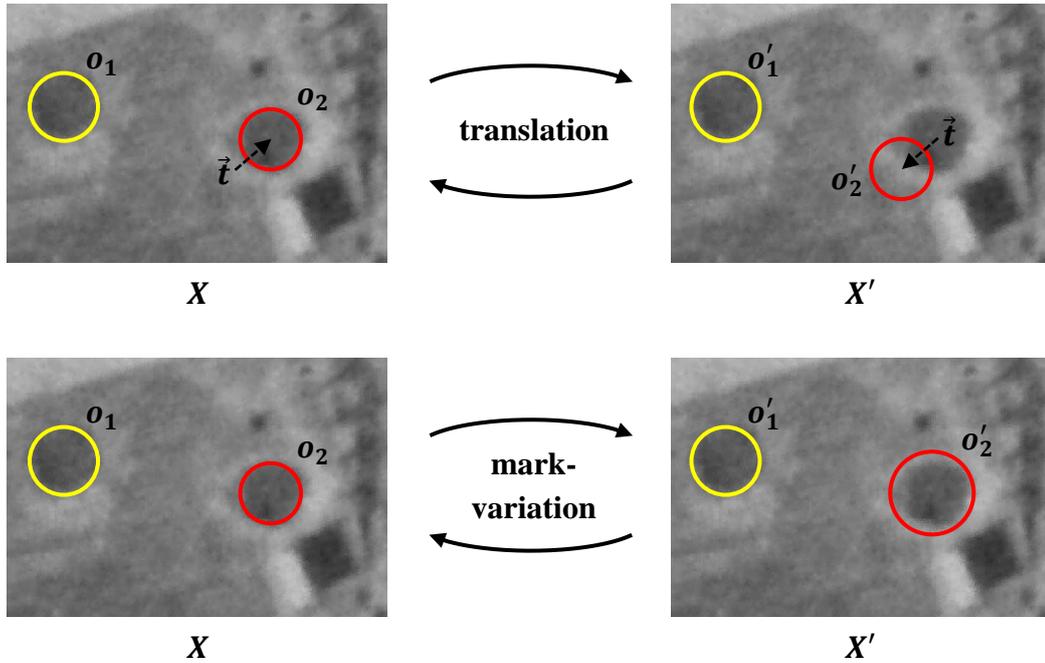
and the Jacobian determinant reduces to

$$\det \left[ \frac{\partial(X', u')}{\partial(X, u)} \right] = \det \begin{bmatrix} 1 & & \\ & 0 & 1 \\ & 1 & 0 \end{bmatrix} = -1. \quad (4.19)$$

Note that the absolute value of the Jacobian determinant is of interest (cf. Equation 4.9). Considering this and the fact that the number of objects does not change, the acceptance probability for *modification*  $\alpha_{Mod}$  is simply given by

$$\alpha_{Mod}(X'|X) = \min\left(1, \exp\left(-\frac{U(X') - U(X)}{T_t}\right) \cdot \frac{q_{Mo}^{inv}}{q_{Mo}}\right), \quad (4.20)$$

where  $q_{Mo}$  and  $q_{Mo}^{inv}$  are the probabilities for choosing the respective modification and its inverse.



**Figure 4.9:** The move type *modification* induces changes to the object parameters, here illustrated using the same exemplary object configuration as in Figure 4.8. Upper row: In the *translation*-move, the coordinates of object  $o_2$  are modified by a randomly generated transition vector  $\vec{t}$ . The reverse move from the proposed state  $X'$  to the previous one  $X$  is possible by applying the inverse modification, i.e. shifting  $o'_2$  in the opposite direction. Lower row: In the *mark-variation*-move, the mark of object  $o_2$ , i.e. its radius, is randomly changed within a predefined interval.

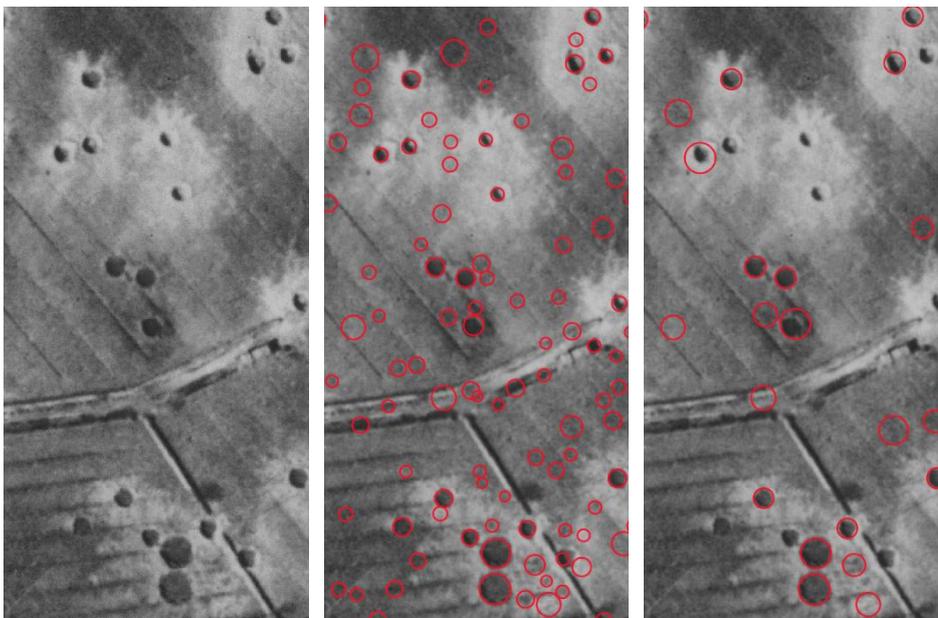
#### 4.1.4 Limitation of the search space

As outlined in Section 2.2.3, in connection with MPPs, the generation of objects is often favoured at certain locations. In this way, the computational effort for the MPPs in the sampling process is reduced. That is why the general idea of restricting the search space in the image is also taken up in this thesis. However, it seems reasonable to also derive information with respect to the initialization of the object model used in the MPP framework, which the discussed approaches from the literature do not address. For that purpose, in this work, the blob detector described in (Mallick, 2022) is employed. A blob is a group of connected pixels that share common properties (e.g. similar grey values). The aim of the blob detection is to find and mark these pixel regions in the image. Hence, the detector provides the coordinate centres for each valid blob as well as its size. That is,

during sampling, the birth of an object is only possible on foreground pixels, i.e. a blob centre and the associated size information is used for the initialisation of the object radius.

In the first step of blob detection, the image is converted into several binary images by applying different thresholds. Starting at a minimum threshold  $B_{T\_min}$ , this threshold is increased by step size  $B_{T\_step}$ , up to a maximum  $B_{T\_max}$ . After extracting connected components from each binary image, they are grouped according to the distance of their centres to form blobs. Blobs located closer than  $B_D$  are merged. The algorithm also provides filter options allowing to detect only dark blobs, bright blobs or both types of blobs. Moreover, blobs can be filtered according to their circularity  $B_{Circ}$ , convexity  $B_{Conv}$ , inertia ratio  $B_{InRat}$  or size  $B_S$ , defined in the respective intervals  $[B_{Circ\_min}, B_{Circ\_max}]$ ,  $[B_{Conv\_min}, B_{Conv\_max}]$ ,  $[B_{InRat\_min}, B_{InRat\_max}]$  and  $[B_S\_min, B_S\_max]$ , meaning that e.g. the circularity  $B_{Circ}$  has to be in the range of  $B_{Circ\_min}$  and  $B_{Circ\_max}$  to not be filtered out, where  $B_{Circ\_min}$  and  $B_{Circ\_max}$  are the minimal and maximal circularity, respectively.

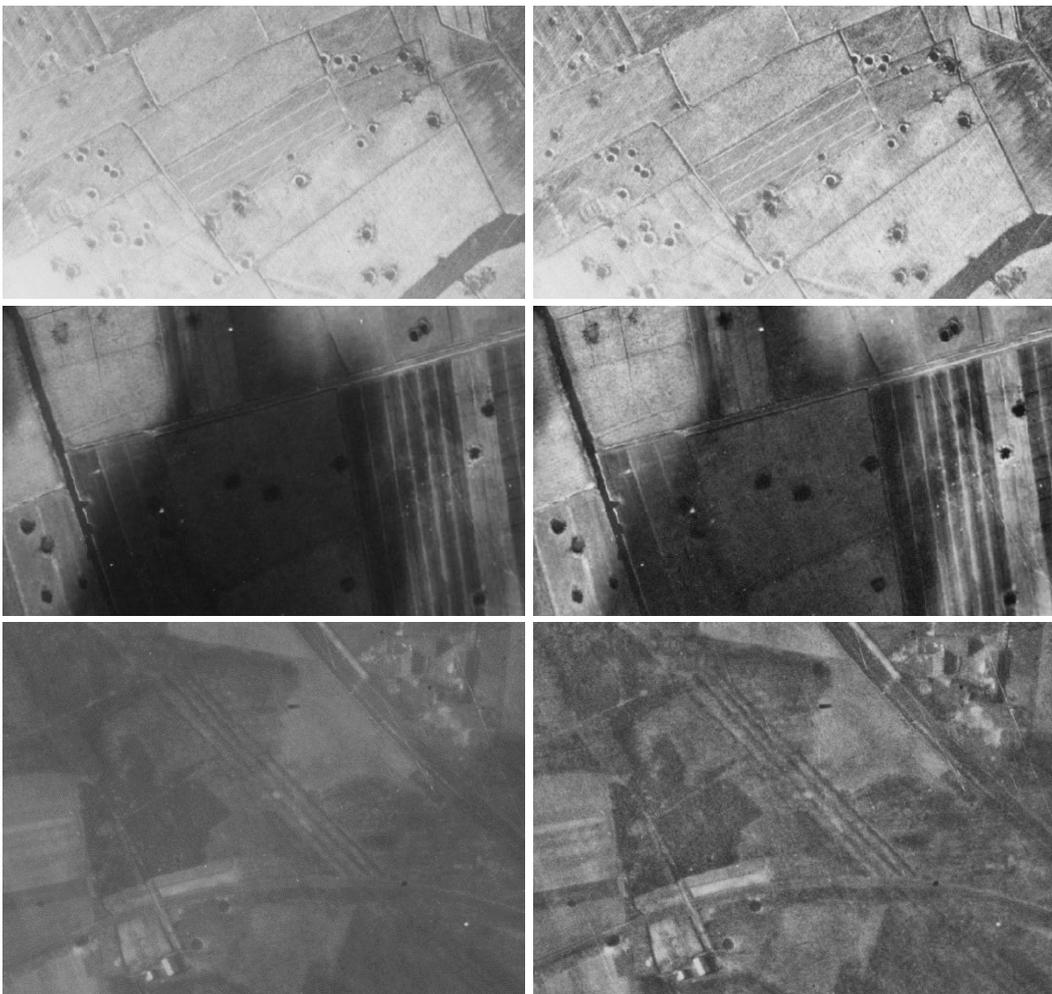
For the subset of an aerial wartime image, Figure 4.10 exemplarily shows the results of the blob detection for a rather suitable (Figure 4.10 middle) and rather unsuitable (Figure 4.10 right) selection of the parameters of the blob detector. In the example, in the latter case, smaller craters and those that are less circular are not detected. Although the number of false detections is reduced by a more restrictive choice of parameters, the MPP result will be less complete, because births of objects can only occur at blob centres, as stated above. In order not to miss any craters that could possibly be detected by the proposed MPP model, the parameters of the blob detector are selected accordingly (Section 5.3).



**Figure 4.10:** Subset of an aerial wartime image (left) and the result of the blob detector (red circles) for a rather suitable (middle) and rather unsuitable (right) selection of its parameters, the latter resulting in smaller and less circular craters not being detected.

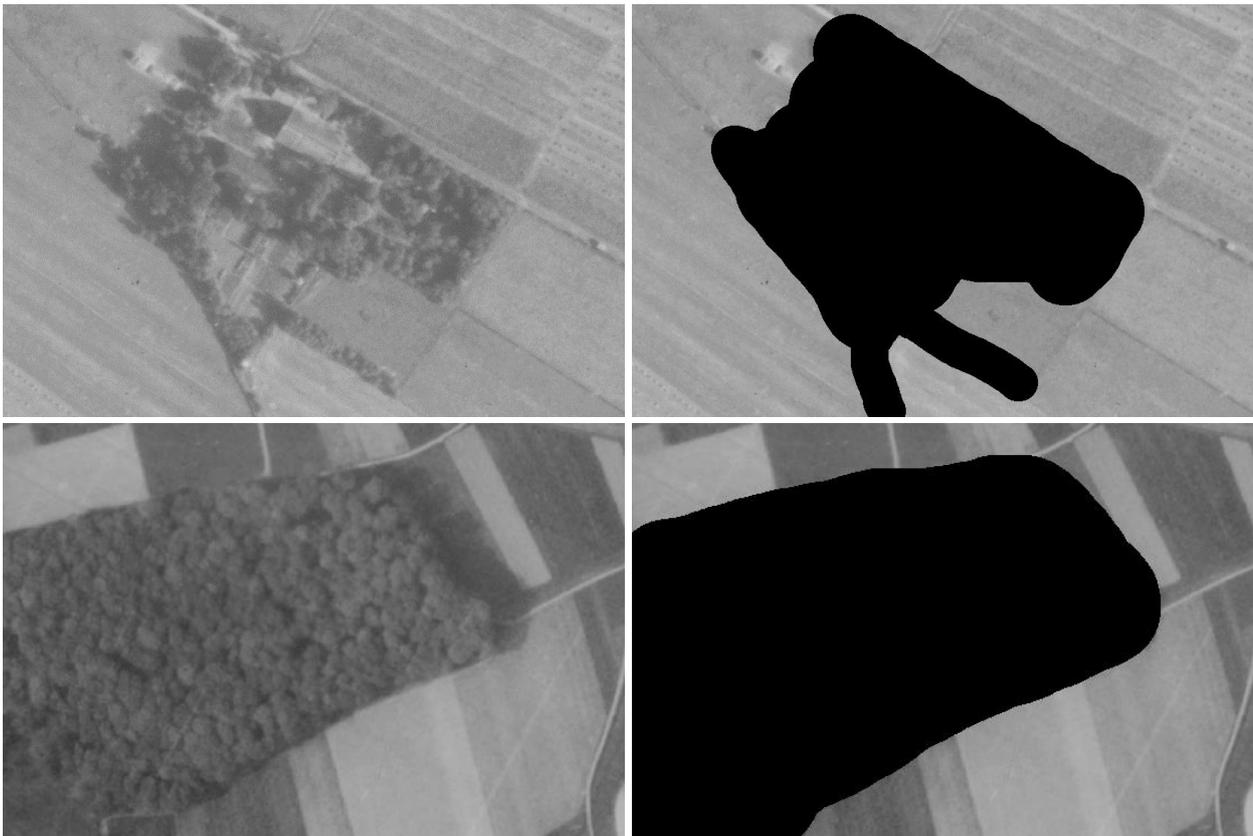
### 4.1.5 Pre-processing

The quality of the aerial wartime images differs considerably due to their age as well as the circumstances during acquisition. In particular, as also mentioned by Farella et al. (2022), there are both, underexposed and overexposed areas as well as poor brightness and contrast levels. To counteract these challenges, Contrast Limited Adaptive Histogram Equalization (CLAHE; Pizer et al., 1987 – in this work, the OpenCV implementation is used) is applied to each image as a pre-processing step. In this context, the image is initially divided into small blocks with a size of  $C_b$  and histogram equalization is applied to every block. To avoid the amplification of noise, contrast limiting is applied before the equalization: Pixels in a histogram bin above the specified contrast limit  $C_l$  are distributed uniformly to other bins. Finally, bilinear interpolation is applied at the borders of the blocks. In Figure 4.11, some exemplary image subsets (Figure 4.11 left column) of images to which CLAHE is applied (Figure 4.11 right column) can be found.



**Figure 4.11:** Subsets of wartime images (left column) and the adapted subsets after applying Contrast Limited Adaptive Histogram Equalization (right column) resulting in more highlighted salient image features.

There are objects that have a similar appearance as bomb craters. Among others, shadow casts by buildings and trees can lead to false detections. If stereoscopic imagery is available, one way of counteract these limitations is to integrate 3D information. Another possibility is to consider additional information, such as historical maps and GIS data, to find and subsequently exclude such areas from further processing. For simulation purposes, high static objects (buildings, forests, trees, others) including their shadows were manually masked. Pixels inside such masked areas are considered neither for the detection nor for the evaluation. Figure 4.12 shows two examples of subsets of aerial wartime images with the corresponding masks.



**Figure 4.12:** Exemplary subsets of aerial wartime images (left) and the respective manually created masks of static objects with a certain height including their shadows (black); in these examples the masked areas include buildings and trees (top right) and forest (bottom right).

### 4.1.6 Conceptual workflow

In this section, for the sake of clarity, the overall conceptual workflow of the proposed method for bomb crater detection is briefly summarized. The individual steps can be summarized as follows, with all parameters involved in the procedure having to be initialized beforehand:

- I. Pre-processing:** CLAHE is applied to the input data (see Section 4.1.5).
- II. Blob detection:** Likely positions and size information for bomb craters are determined using the blob detector (see Section 4.1.4).
- III. MPPs:** The following MPP-related steps are repeated until the number of objects has not changed for a certain number of iterations (stop criterion):
  - a. One of the available move types (*birth-and-death* or *modification*) is randomly chosen.
  - b. All required parameters for the specific move to be applied (*birth*, *death*, *translation* or *mark-variation*) are generated and a new object configuration  $X'$  is proposed.
  - c. The energy difference  $\Delta U = U(X') - U(X)$  of the proposed configuration  $X'$  and the current one  $X$  is determined.
  - d. The corresponding acceptance probability  $\alpha$  (Equation 4.16, 4.17 or 4.20) is computed.
  - e. A uniform random number  $\delta \in [0, 1]$  is sampled. The proposed object configuration  $X'$  is accepted if  $\delta < \alpha$ , otherwise it is rejected, i.e. the current configuration  $X$  is maintained.

The result of the MPP is the object configuration that best describes the bomb craters in the given data, considering the predefined model.

## 4.2 Fusing the results from multiple images

In the presence of multiple overlapping images, the MPP process is applied independently to all images (Section 4.1). In a subsequent step, the resulting detections are combined before generating an impact map (Section 4.3). This means that detections from multiple overlapping images that refer to the same object (e.g. a certain bomb crater) have to be matched. A particular problem is that the georeferencing information of the aerial wartime images is not very accurate. For instance, the images used in the experiments have a georeferencing accuracy of about 5 m - 40 m (see Section 5.2). Furthermore, as already indicated in Section 2.3, it should be exploited that several detections of the same object in different images are an indication for indeed having found a correct object (redundancy).

In order to merge the detections of every image that refer to the same object, in this thesis, a local solution is opted for. First, a master image (the detections within this image will be referred to as master detections MDs) is selected and detections from the remaining images (non-MDs) are assigned to the respective MDs. However, due to the coarse georeferencing, the positions of identical objects in object space can differ by up to 40 m. For this reason, for each MD, a search radius

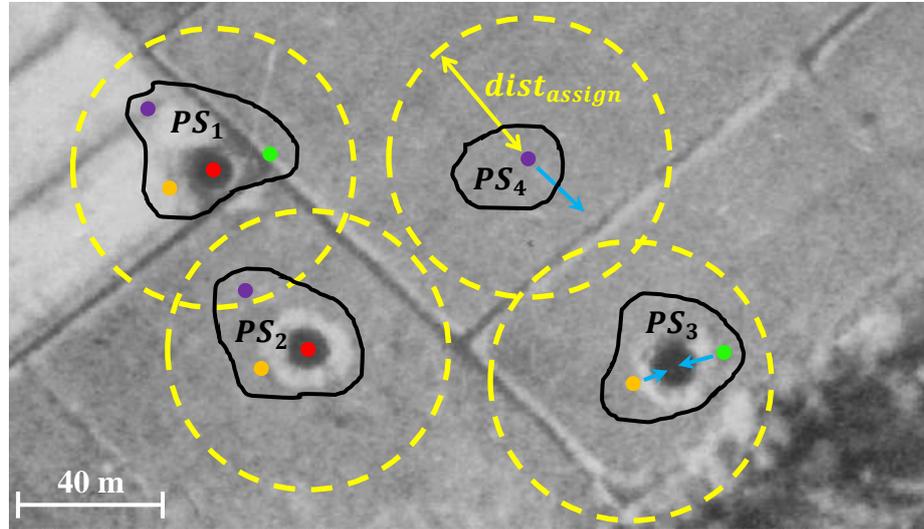
$dist_{assign}$  is defined in order to be able to cover all possible detections that belong to the same object. For each MD, all detections from the remaining images within  $dist_{assign}$  may form a point set  $PS$  with this MD. Note that when searching for neighbours, only the closest neighbour per image is considered, and each detection can only be used once. After all MDs have been used to define  $PS$ , the remaining non-MDs are treated in the same way. It should be noted that due to the fact that the georeferencing is only coarse, incorrect assignments are possible in this local approach. This issue is briefly discussed in Section 4.4; an evaluation of the approach presented in this section, addressing the question to what extent the inaccurate georeferencing affects the quality of the generated impact map, can be found in Appendix A.

Non-MDs are usually not located at the positions of the craters in the master image due to the inaccuracy of the georeferencing. To counteract this problem, the fact that detections from the same image will have (almost) identical offsets to the respective MDs in the local neighbourhood is exploited: Corresponding mean displacement vectors  $\vec{d}_n$  are determined for each detection  $n$  of  $PS$  without MD from  $PS$  containing a MD and being in a circular neighbourhood (600 m, based on empirical investigations). These vectors are then used to move the respective detections according to the particular offset. If only a few  $PS$  with MD are available for the determination of the displacement vectors  $\vec{d}_n$ , this method is of course less reliable.

Finally, for  $PS$  containing a MD, the MD coordinates are used for the subsequent generation of the impact map, while for the other  $PS$  the centre of gravity after translation is utilized. Despite the local translations, the centres of gravity from  $PS$  not including a MD are usually still not located at the exact centres of the craters in the master image, which has an impact on the evaluation of the results and should be counteracted (see Section 5.4). Given the fact that  $PS$  with one (e.g. image errors) or only a few (e.g. shadows) detections may be erroneous,  $PS$  not consisting of at least  $MinNumDet_{PS}$  detections are deleted.

Figure 4.13 illustrates the procedure based on a subset of four overlapping aerial wartime images. Here, the centres of the respective detections are shown in red, orange, green and purple, respectively. First, one MD (red) is randomly selected (here the top left one) and all neighbouring detections within the radius  $dist_{assign}$  (dashed yellow circle) may form  $PS_i$  with this MD. According to the previously mentioned rules, the purple detection in the upper left corner is assigned to this  $PS$  (next to the orange and green one), forming  $PS_1$  (delineated in black). In analogy,  $PS_2$  is formed by the red, orange and purple detections. All MDs having been used to define  $PS$ , the detections from the remaining images are considered. It is proceeded with the detections of a randomly selected image, in this example the orange ones; the remaining orange detection forms  $PS_3$ , to which the green detection is assigned. Only the purple detection in the upper right corner has not yet been considered; it forms  $PS_4$ . Subsequently,  $PS$  with a MD (here  $PS_1$  and  $PS_2$ ) are used to determine the respective displacement vectors  $\vec{d}_n$  (light blue) to move all detections from  $PS$  without MD (here  $PS_3$  and  $PS_4$ ) accordingly. After computing the new centres of gravity for  $PS_3$

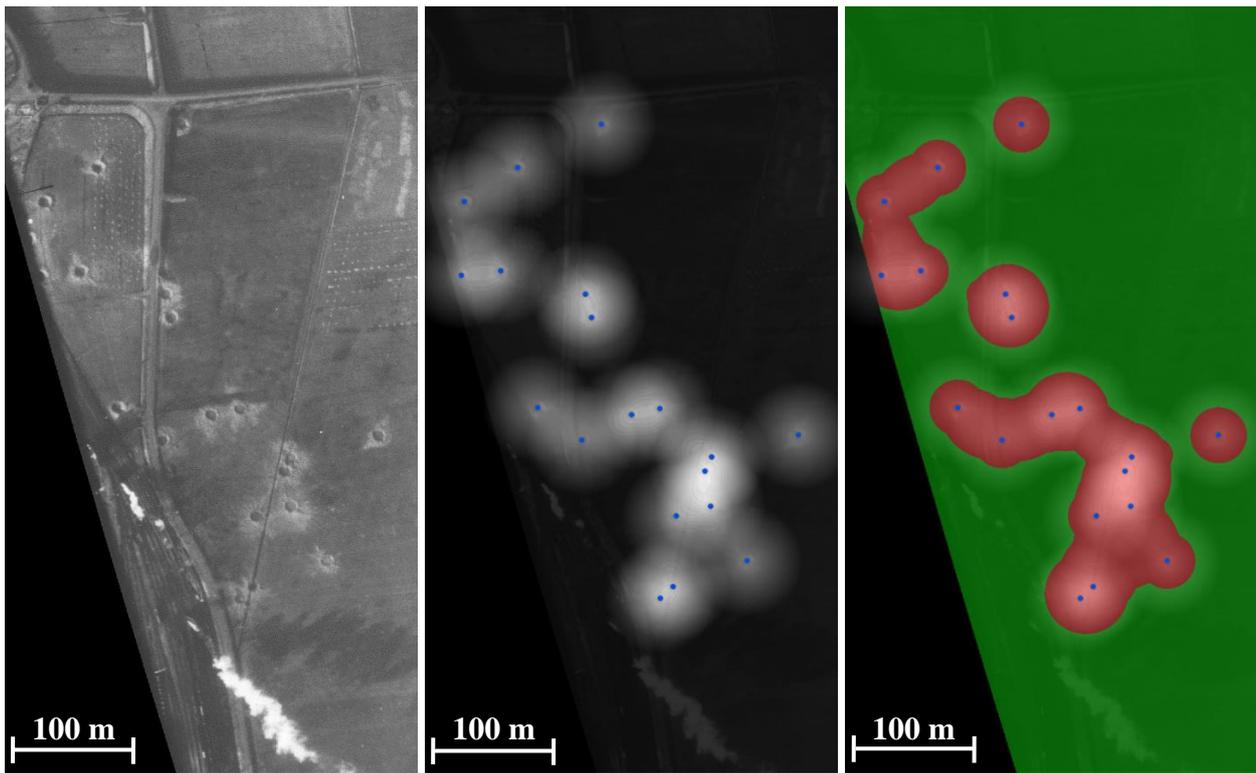
and  $PS_4$ ,  $PS$  with too few detections are deleted (for instance, for  $MinNumDet_{PS} = 2$ , this is the case for  $PS_4$ ).



**Figure 4.13:** Subset of an aerial wartime image covered by a total of four images with the detection results in red, orange, green and purple, respectively; these detections are assigned to point sets  $PS_i$  (outlined in black) according to criteria described in the main text. After the assignment process, certain detections are translated (light blue arrows) and  $PS$  have to fulfil a certain criterion (see main text).

### 4.3 Impact map

The detections, obtained from either combined (Section 4.2) or single image (Section 4.1) detection results, are used to derive a probability for each location that there are duds nearby. The associated probability map is generated from the centres of the detections by means of kernel density estimation with the conic kernel function  $K(k) = (1 - |k|)$ . In this context, the bandwidth  $h$  in Equation 3.12 indicates how large the area of influence of a detection is. Using the probability map, the entire scene is classified into potentially contaminated and uncontaminated areas. For that purpose, a threshold  $p$  is applied to the probabilities, resulting in an impact map that indicates areas with a high probability of containing duds. Figure 4.14 illustrates the overall process based on a subset of an aerial wartime image (Figure 4.14 left), including the probability map generation from the detections by kernel density estimation (Figure 4.14 middle) and the subsequent derivation of the impact map by setting a threshold for the probabilities (Figure 4.14 right).



**Figure 4.14:** Subset of an aerial wartime image (left). Superimposition of the subset of the aerial wartime image and a probability map derived from the centres of detected bomb craters shown in blue, where brighter colours represent higher probabilities for the occurrence of duds and vice versa; the area of influence of a detection is controlled via the bandwidth-parameter, here set to 50 m (middle). Superimpositions of the subset of the aerial wartime image, the probability map and an impact map, obtained by setting a threshold for the probabilities, with potentially contaminated and uncontaminated areas shown in red and green, respectively (right).

## 4.4 Discussion

In this section, some aspects of the proposed method are discussed; these mainly include limitations related to the detection of craters by means of the MPP procedure and its transferability to a different object detection task. By selecting appropriate parameters (Section 5.3), the MPP method can be adapted accordingly.

The proposed model for the MPPs only allows the detection of dark bomb craters, i.e. craters that are characterized by locally darker grey values in comparison to the surrounding area. Although craters typically appear dark in the images, this does not always hold true. For instance, if craters have been filled with earth before the image was taken, or if sandy soil is present, they usually appear brighter than their surroundings. To detect this type of craters, it would be possible to use the complementary image as input (implying that only bright craters exist) or to reverse the direction of the gradients calculated based on Equation 4.2. However, experiments have shown

that while bright craters can be found, the number of false detections is significantly increased and therefore – especially with regard to the proposed application scenario – the detection of bright craters is not pursued further.

For certain applications, such as counting trees (e.g. Perrin et al., 2004) or birds (e.g. Descamps et al., 2008), the detection of each object is important, e.g. to derive information about the number and density of objects. In connection with the impact map generation, however, especially in highly bombed areas, the separate detection of each crater is of secondary importance, as this area has to be probed anyway (see discussion in Section 1.1). In this context, it should be noted that, obviously, bomb craters can overlap to any extent. Particularly when they cover each other (almost) completely or only fragments are visible, the separate detection of all craters becomes impossible for the proposed bomb crater detection method. However, in order not to completely prohibit the detection of overlapping objects, such configurations are penalized as specified in Equation 4.8. In this way, one object in the image will not be represented by several circles, but the detection of overlapping craters is still possible in principle, although in practice usually only for craters that overlap slightly.

In the presence of multiple overlapping aerial wartime images, the MPP procedure is applied independently to all images (Section 4.1). Subsequently, the results are combined locally (Section 4.2). However, due to the coarse georeferencing accuracy of the images, incorrect assignments, especially in densely bombed areas, are possible in this local approach. As already pointed out in the motivation of this work (Section 1.1), to overcome this problem, it would be possible to first improve the co-registration of the respective images based on global approaches such as automatic aerial triangulation. However, this may become challenging, particularly due to the different appearances (e.g. caused by seasonal changes or noise) and missing camera information. Given these problems, the approach presented in Section 4.2 has been developed. It has been evaluated and found to work satisfactorily, although, not surprisingly, the quality of the impact maps would be improved – at least to a small extent – if there was a more accurate co-registration of the respective images (see Appendix A).

Although the proposed MPP method is designed for the detection of bomb craters in aerial wartime images, it can be used on other images as well, as long as the objects to be detected basically follow the MPP model and the parameters of the method are adapted to the new data. Certain objects in microscopic images, such as cells, show crater-like characteristics. More precisely, cells possibly overlap and appear circular as well as have a homogeneous area. This basically also applies to stomata (pores on the bottom side of a leaf) in microscopic images and birds or tree crowns in aerial images. Usually, such imagery is used for population counting purposes. If the objects appear brighter than their surroundings, the complementary image can be used as input, as discussed at the beginning of this section.



## 5 Experimental setup

This chapter presents the experimental setup used to evaluate the methodology proposed in the previous chapter. For this purpose, the evaluation objectives are introduced in Section 5.1, before describing the data used for testing (Section 5.2). Section 5.3 reports the parameter settings for the experiments, while the evaluation procedure and criteria, allowing for a quantitative assessment of the results, are presented in Section 5.4. This chapter closes with a description of the test setup, including methodological aspects, regarding a comparison with a deep learning based object detector (Section 5.5).

### 5.1 Objectives

The overall objective of this work is to develop a procedure that generates an impact map from bomb craters automatically detected in aerial wartime images; the crater detection task is addressed by marked point processes (MPPs). The quality of such an impact map and that of the bomb crater detection itself can be evaluated using reference data. For this purpose, the results are evaluated quantitatively based on the criteria presented in Section 5.4. The experiments are designed as follows: First, the proposed MPP model is examined by evaluating the influence of different aspects on the results. Moreover, the applicability of the overall approach to differing aerial wartime images is investigated and results are compared with a state-of-the-art object detector based on CNNs. Additionally, the influence of using redundant image information on the results and its contribution with respect to the generation of an improved impact map in terms of its precision, is investigated. Finally, the extent to which the MPP model is suitable for population counting purposes is examined. The following questions are addressed by the experiments:

*(1) How stable are the results in the proposed MPP method for the detection of bomb craters despite the use of random numbers? How does the circle as default object model compare to an ellipse? How do the energy terms affect the quality of the results?*

In order to verify whether and how well the MPP model used is suitable for the detection of bomb craters in aerial wartime images, individual parts of the model are examined in more detail. By repeating the same experiments several times with identical parameter settings, the influence of the random numbers generated during sampling on the results is investigated. Furthermore, the

influence of the chosen object model as well as that of the proposed energy terms on the quality measures of the results is analysed.

*(2) How well does the proposed approach perform on different aerial wartime imagery? To what extent does redundant image information help to improve the quality of the impact map? Can the method produce useful results with respect to the proposed application scenario? What would be the general benefit of height information? How does the MPP-based method for detecting bomb craters compare to a state-of-the-art object detector for this task?*

The proposed method is tested on a variety of aerial wartime images acquired by the Allied forces during the Second World War and taken over Lower Saxony, Salzburg and Italy. In this way, the general validity of the approach is assessed and potential limitations are identified. As there are typically multiple images of the same area, in another set of experiments, the influence of using redundant image information, in this case of combining the MPP results of coarsely georeferenced panchromatic images of Lower Saxony, on the quality of the results is studied. Lastly, it is investigated whether the generated impact map can be used in an appropriate way with respect to the proposed application scenario. For this purpose, while in the experiments discussed so far, the focus is on achieving results with a high F1-score, in these experiments the procedure is tuned to achieve a high precision, i.e. the areas falsely classified as contaminated should be as small as possible, which is important for the application scenario. Furthermore, to simulate the existence of 3D and/or further information, such as historical maps and GIS data, a manual masking of high static objects plus their shadows has been carried out (cf. Section 4.1.5). By applying the proposed method to the masked data as well, the overall benefit of such additional information can be assessed for both the single (Section 6.2.1) and the multi-image case (Sections 6.2.2 and 6.2.3). Finally, the results of the MPP for the detection of bomb craters are compared with those of a state-of-the-art object detector based on CNNs.

*(3) To what extent is the MPP method suitable for the purpose of population counting in images? How does it perform compared to other algorithms for this task?*

The proposed MPP method is designed for the detection of bomb craters in aerial wartime imagery. At the same time, obviously, the MPP model can in principle also be used for other applications (cf. discussion in Section 4.4 regarding the transferability of the approach). With the objective of population counting, this work examines the applicability of the model to panchromatic microscopic images for the detection of cells or stomata, and to a panchromatic aerial image for the detection of flamingos. For the selected scenes, the results of other methods from the literature are available, with which a quantitative and qualitative comparison is made.

## 5.2 Test data

To allow for a comprehensive assessment, the proposed method is evaluated on different data and with respect to different applications; the particular test data is described in the following.

Regarding the generation of impact maps from bomb craters automatically detected in aerial wartime images, the provided data (Table 5.1) stem from three different sources, namely the Explosive Ordnance Disposal Service of Lower Saxony, Germany, the 3D Optical Metrology research unit of the Bruno Kessler Foundation (FBK) in Trento, Italy, and the Salzburger Geographisches Informationssystem (SAGIS) of the Federal State of Salzburg, Austria, showing images of the respective territories. Reference information (position and radius for each bomb crater) for all three sources was generated by Lower Saxony's Explosive Ordnance Disposal Service by manual annotation. The number of images ( $n_I$ ) per source with respective ranges of ground sampling distance (GSD), covered area per image ( $A$ ) and number of pixels ( $n_P$ ) are also given in Table 5.1. Note that for the data of Lower Saxony, a reference only exists for 55 images (Ref). Furthermore, all images have a radiometric resolution of 8 bit and their degree of impact, i.e. the number of bomb craters per image, varies between zero and more than 1000. In general, the focus of the investigations is more on rural sites. In densely built-up areas, it is not possible to clearly identify craters in the images because they are largely covered by the debris of destroyed buildings. Moreover, the images are representative for certain cases, e.g. different lighting situations exist and the image content varies. Overall, the content is less challenging for images of Lower Saxony than for those of Italy and, in particular, of Salzburg (Figure 5.1): while the former images include comparatively few disturbing objects (e.g. forests, trees, buildings and their shadows or craters not representative for the MPP model; Figure 5.1 top row), this is different for images from Italy. Here, in addition, single trees and craters deviating from the model (e.g. bright craters due to sandy soils) are more often present (Figure 5.1 middle row). The last set of images shows the city of Salzburg and its immediate surroundings, where apart from non-representative craters there are more crater-like objects (e.g. numerous detached buildings with striking shadows; Figure 5.1 bottom row). Note that the images from Salzburg – in contrast to all other wartime images used in the experiments – already have homogeneous exposure conditions, which is why Contrast Limited Adaptive Histogram Equalization (CLAHE; Section 4.1.5) is not applied.

The images from Lower Saxony are coarsely georeferenced (approx. 5 m - 40 m), often affected by a considerable amount of image distortion, and were scanned at a resolution of 1200 dpi. In connection with the analysis of the MPP model, all 55 images from Lower Saxony having a reference are used, except for the experiments to assess the impact of the random components of the method, in which a representative selection of 10 images is used instead.

For the investigations regarding the general applicability of the approach, besides the data from Lower Saxony, 10 images from Italy and 9 images from Salzburg are considered. The images of

**Table 5.1:** Characteristics of the data used for the evaluation. For more information and an explanation of the abbreviations, see main text.

Source		nI	GSD [m]	A [km <sup>2</sup> ]	nP · 10 <sup>8</sup> [px]
Lower Saxony	Ref	55	0.13 - 0.36	1.8 - 19.3	0.8 - 1.6
	noRef	272	0.11 - 0.84	1.7 - 21.3	0.3 - 1.9
	DS A	52 (RG 1)	0.16 - 0.52	2.5 - 21.0	0.4 - 1.9
		48 (RG 2)			
		23 (RG 3)			
	DS B	51 (RG 1)	0.11 - 0.84	1.7 - 21.3	0.3 - 1.6
		67 (RG 2)			
58 (RG 3)					
Italy		10	0.28 - 0.64	4.7 - 9.8	0.2 - 1.2
Salzburg		9	0.20	6.25	1.56

Italy are not georeferenced, but a rough image scale, in combination with the known scanning resolution, allows to determine the approximate GSDs. The Salzburg images are part of a digital orthophoto mosaic with known GSD.

For the investigations on the influence of the use of the redundant image information, 272 images (noRef) of three different regions (RG) of Lower Saxony are available in addition to 27 of the 55 images with reference. Thus, in total,  $27 + 272 = 299$  images are used for this set of experiments, with 103, 115 and 81 of the images showing parts of one of the three regions, respectively; the three regions are Emden, Lehrte and Osnabrück. The remaining 28 images with reference have no overlap with these three test regions and thus cannot be used for these experiments. The image content gets more complex from region 1 to region 3, because of an increasing number of disturbing objects, while the number of craters decreases. For every region, the images are separated into two data sets (DS) acquired at different time periods, called DS A and DS B. Note that the images within one time period stem from several separate surveillance flights that have been carried out, for instance, between March and November 1944. The reason for the split is that the numbers of bomb craters per image between the two time periods per region differ significantly, because there had been bomb raids in between. Hence, the images in DS B usually have more craters. The split of the image set into DS A and DS B is based on a visual inspection. A joint evaluation faces the problem that some bomb craters correctly detected in DS B would correspond to false positive detections when evaluated using the reference of DS A, which would lead to a bias in the evaluation metrics. The investigations are carried out as follows: For each of the three regions, 6 images with a reference serve as master images for DS A and 3 images with a reference are selected as master images for DS B, hence  $3 \cdot 6 + 3 \cdot 3 = 27$  images with a reference are used for evaluation, as mentioned earlier. These 27 images are selected in such a way that in each region and each data



**Figure 5.1:** Exemplary subsets of aerial wartime images from Lower Saxony (top row), Italy (middle row) and Salzburg (bottom row).

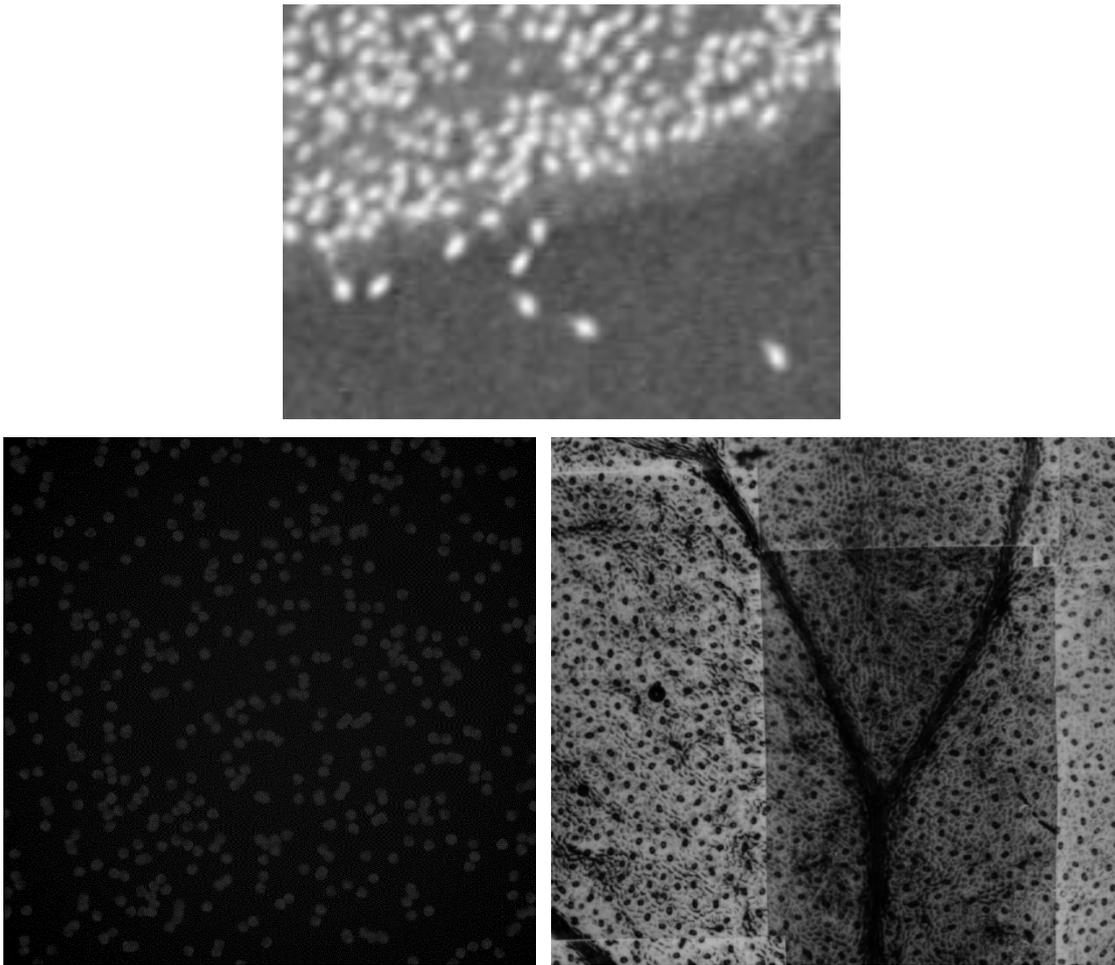
set, there is an equal number of images of what is considered to correspond to a difficult, a moderate and an easy situation for the algorithm, respectively. When combining the detection results (Section 4.2), all the images from the same DS having an overlap with the respective master image are additionally considered. The amount of coverage varies locally, because the footprints of the images inside a region are not aligned. Hence, each *PS* (Section 4.2) of a master image is covered by a different number of images (the respective mean number of images considered as well as the number of bomb craters of each master image are listed in Table 6.4).

For the investigations concerning the proposed application scenario in which the parameters of the algorithm are varied to achieve a higher precision at the cost of a lower recall, the thesis concentrates on the multiple image case. Thus, the same data is considered as with the experiments regarding the influence of the use of redundant image information on the results (see previous paragraph).

In addition to generating impact maps, the suitability of the MPP method for the problem of counting and locating specific objects from images is examined (see Figure 5.2). In the scenes being considered, the shape of the objects is relatively simple, whereas the objects partially occlude each other. More specifically, there is one aerial image showing flamingos and two microscopic images showing cells as well as stomata (all images are panchromatic); the three scenes are hereafter referred to as *flamingos*, *cells* and *stomata*. The first scene, *flamingos*, is a sample of an aerial image representing a colony of migrating flamingos (Figure 5.2 top). In the second scene, *cells*, a simulated microscopic image of cells is given (Figure 5.2 bottom left), while the last scene, *stomata*, shows a real microscopic image of a leaf with stomata (Figure 5.2 bottom right). Further information about the images, such as their size and the number of objects they contain, can be found in Table 5.2. Moreover, reference data is available for all three scenes in the form of binary images, where for each object its position and radius (identical for all objects of a scene) in the image are given. The images and reference data just described are part of an online data set (Benchmark, 2013) provided by Verdié and Lafarge (2014), where results of methods from the literature can also be found. More precisely, besides numerical results, images showing the boundaries of the detected objects superimposed on the respective image of a scene are given.

### 5.3 Parameter settings

The developed method contains a number of free parameters to be selected (see Table 5.3). Following own work (Kruse et al., 2022), the parameters in the experiments are set to values that have been determined empirically. If not specifically indicated, they are set to identical values for all images and tests, making the procedure more relevant for a potential use case. An explanation and justification of the parameters is given in the following.



**Figure 5.2:** The marked point process procedure is also evaluated on other panchromatic data, namely the scenes *flamingos* (top), *cells* (bottom left) and *stomata* (bottom right).

**Table 5.2:** The transferability of the marked point process approach to the application of counting populations based on images is examined by using different images from (Benchmark, 2013).

Scene	Size [px]	Number of objects
Flamingos	160 x 120	148
Cells	1000 x 1000	500
Stomata	657 x 617	676

For CLAHE (Section 4.1.5), the parameter selection for the block size  $C_b$  and the contrast limit  $C_l$  is performed based on visual inspection in a way that underexposed and overexposed images, respectively, are enhanced appropriately. Moreover, experiments have shown that this choice of parameters leads to better results than the default values.

The parameters of the blob detector (Section 4.1.4) are selected as follows: The thresholds in connection with the binary image creation  $B_{T\_min}$  and  $B_{T\_max}$  are set in a way that in each case at least all grey values in the images being related to bomb craters are considered, i.e. there are (almost) no images where craters exhibit grey values lower than  $B_{T\_min}$ . The step size  $B_{T\_step}$  for the binary image creation is set to  $B_{T\_step} = 2$ . Even though a value of one would marginally increase the number of detected craters, this also leads to comparatively many more detections of non-crater objects. Blobs may be merged if they are closer than  $B_D$ ; this parameter is set to a low value ensuring that blobs are basically not merged. This is important because for high values of  $B_D$ , for instance, two craters close to each other are no longer represented individually, but only by one blob centred between them. Given the facts that the additional detection of bright craters leads to many false detections and bomb craters or their shadows, respectively, are generally characterized by darker grey values than those in their surroundings (see discussion in Section 4.4), the procedure should only detect dark blobs. In connection with the filter parameters, the lower boundaries  $B_{Circ\_min}$ ,  $B_{Conv\_min}$  and  $B_{InRat\_min}$  are set in a loose way and the upper ones are not considered (i.e. set to one). This allows craters to deviate from a circle by a large extent. The selection of  $B_S$  in the interval  $[B_{S\_min}, B_{S\_max}]$  makes it possible to detect objects and thus bomb craters with different sizes. Depending on the GSD of the respective image,  $B_{S\_min}$  and  $B_{S\_max}$  are set in a way that blobs with a diameter between 6 m and 18 m can be detected (note that the GSDs are only roughly known), which corresponds to the diameters of craters appearing in the data. Although selecting such loose filter restrictions results in many false detections, experiments have shown that a more restrictive choice can exclude the detection of many bomb craters in advance.

Regarding the object model used by the MPP, the lower and upper limits of the radius (circle) or semi-major ( $a_m, a_M$ ) and semi-minor ( $b_m, b_M$ ) axes (ellipse; only used for the comparative experiments in Section 6.1.2), respectively (Section 4.1.1), are derived from the minimum and maximum blob radius  $B_r \in \{B_{r\_min}, B_{r\_max}\}$  occurring in the image after blob detection, i.e.  $B_{r\_min} = r_m = a_m = b_m$  and  $B_{r\_max} = r_M = a_M = b_M$ .

In connection with the energy function (Section 4.1.2), the data and prior energy are equally weighted based on  $\beta$  (Equation 3.2). The parameter  $c$  of the first data term ( $U_G(X)$ , Equation 4.2) is not set identical for the images of Lower Saxony, Italy and Salzburg, respectively. In general, the different values are due to the varying crater appearance in the images. Moreover, it is reasonable to set the value of  $c$  higher for images that contain a comparatively larger amount of disturbing objects, which is the case for the data set from Italy and especially for the one from Salzburg. Such objects lead to a potentially higher number of false detections and by increasing  $c$ , less objects with smaller gradients at the object border will be detected, i.e. the number of false detections is potentially decreased. Obviously, correct ones may also be affected, though usually not to the same extent. Besides  $c$ , the factor  $f_G$ , weighting  $U_G(X)$ , and  $n_v$ , the number of vertices of the polygon used to approximate the border of an object, have to be selected. The parameters of the second term  $U_H(X)$  of the data energy (Equation 4.3) are the factor  $f_H$ , weighting  $U_H(X)$ ,  $H_t$ , the threshold

related to the grey value standard deviations within an object and  $H_e$ , the percentage of the border pixels of an object that are excluded from the computations. As craters are often not completely homogeneous inside,  $H_t$  is set to  $H_t = 10$ , meaning that there is no increase in energy for standard deviations lower than  $H_t$ . Similarly, as craters are not always exactly circular,  $H_e$  is set to  $H_e = 20\%$ . For the last data term ( $U_B(X)$ , Equation 4.5), there are three parameters to be selected: First, the factor  $f_B$ , weighting  $U_B(X)$  and, second, the threshold  $d_0$  (Equation 4.6), controlling when the associated energy becomes negative (depending on the contrast between the object and its annulus). Third, the parameter  $e_{annu}$  (Figure 4.6), describing the extent of the annulus considered with respect to the computation of the Bhattacharyya distance, which is set to  $e_{annu} = 2$  m. As discussed, the annulus may differ from image to image, e.g. in terms of its width. In this context, even though the annulus can be larger, higher values for  $e_{annu}$  could lead to the fact that in addition to pixels of the annulus (often bright), other, often comparatively darker pixels of the immediate surroundings would also be considered for the computations. Finally, in connection with the prior energy (Equation 4.7), an overlap of objects is possible with  $f_o = 10^4$  (Equation 4.8).

Regarding the changes in the object configuration (Section 4.1.3), the probability for choosing move type *birth-and-death* is set four times higher than the one for move type *modification*. This is reasonable, because often a circle no longer has to be shifted nor altered significantly due to the position and size information provided by the blob detector. The probability of selecting a particular movement and its inverse is considered to be equal, i.e.  $q_B = q_D = q_{Mo} = q_{Mo}^{inv} = 0.5$  (Equations 4.16, 4.17 and 4.20), and therefore the corresponding ratios in the equations are one. In order to avoid manual intervention, the intensity parameter  $\lambda$  in the Poisson point process (Equation 3.1), being involved in the calculation of the acceptance probabilities with respect to the *birth-* and *death-*events (Equations 4.16 and 4.17), is set automatically based on the number of blobs in the image. The initial configuration for the sampling process is an empty set of objects. Simulated annealing uses a geometric cooling scheme by reducing the temperature  $T_t$  (Equation 3.10) using a start temperature  $T_0$  and a cooling coefficient  $cc_T$  (Equation 3.11); the latter is set to  $cc_T = 0.9994$ . Higher values for  $cc_T$  do not affect the quality of the results, but lead to an increased runtime. The optimization stops as soon as the number of objects has not changed for a certain number of iterations, which is defined by *stop\_crit*. In this way, the choice of a maximum number of iterations as a stop criterion can be circumvented. Note that still a maximum number of iterations exists to ensure that the algorithm will eventually stop, but this number was never reached in the experiments.

With regard to the investigations concerning redundant image information (Section 4.2), the value of  $dist_{assign}$  is adjusted to the given coarse georeferencing accuracy, ensuring that theoretically all possible detections belonging to the same object can be found. Moreover,  $MinNumDet_{PS}$  is set to 4, because a small number of detections for the same object indicates that detections are more likely to be incorrect.

**Table 5.3:** Parameters in the proposed approach for the generation of an impact map from detected bomb craters. \*: The parameter  $c$  of the first data term is set to 1000, 1200, and 1500 for the images of Lower Saxony, Italy and Salzburg, respectively (for a discussion, see main text).

		Parameter	Value	Unit	Description
<b>Bomb crater detection</b>	CLAHE	$C_b$	60 x 60	pixel	block size
		$C_l$	2.0	-	contrast limit
	Blob detection	$B_{T\_min}$	10	-	minimum threshold for binary image creation
		$B_{T\_max}$	245	-	maximum threshold for binary image creation
		$B_{T\_step}$	2	-	step size for binary image creation
		$B_D$	5	pixel	blobs located closer than $B_D$ are merged
		$B_{Circ\_min}, B_{Circ\_max}$	[0.1, 1.0]	-	min. resp. max. value for a blobs' circularity
		$B_{Conv\_min}, B_{Conv\_max}$	[0.4, 1.0]	-	min. resp. max. value for a blobs' convexity
		$B_{InRat\_min}, B_{InRat\_max}$	[0.1, 1.0]	-	min. resp. max. value for a blobs' inertia ratio
		$B_{S\_min}, B_{S\_max}$	[3.0, 9.0]	metre	min. resp. max. value for a blobs' radius
		Energy function	$\beta$	0.5	-
	$f_G$		1	-	factor weighting first data term
	$n_v$		32	-	number of vertices of the approximated object border
	$c$		1000, 1200, 1500	-	constant related to gradient magnitudes; set identical for each image source*
	$f_H$		5	-	factor weighting second data term
	$H_t$		10	-	threshold related to grey value standard deviations
	$H_e$		20	percent	percentage of border pixels excluded from the computations
	$f_B$		2000	-	factor weighting third data term
	$d_0$		25	-	threshold related to the contrast between an object and its annulus
	$e_{annu}$		2	metre	extent of the annulus
	$f_O$		$10^4$	-	factor weighting the prior term
	Changes in the object configuration	$\lambda$	$\#blobs / 20$	-	expected number of objects in the image; $\#blobs$ : number of blobs
		$T_0$	100	-	starting temperature for simulated annealing
		$cc_T$	0.9994	-	cooling coefficient for simulated annealing
		$stop\_crit$	$10^4$	-	stop criterion: algorithm stops when number of objects has not changed for $stop\_crit$ iterations
	<b>Fusing results from multiple images</b>	$dist_{assign}$	40	metre	search radius in the assignment process
		$MinNumDet_{PS}$	4	-	point sets not consisting of at least $MinNumDet_{PS}$ detections are erased
<b>Impact map</b>	$h$	40	metre	bandwidth for kernel density estimation	
	$p$	0.5	-	threshold related to the probabilities in the probability map	

Associated with the impact map generation (Section 4.3), the bandwidth  $h$  (Equation 3.12) for kernel density estimation is set to 40 m. In this context, the threshold  $p$  for the probabilities in the probability map is set in a way that for single detections the area around the centre of an object within a radius of 20 m is classified as contaminated. This value, which is the result of discussions with experts in this field, is set relatively small in order to only flag those areas that should subsequently be probed.

For the investigations regarding population counting in images, due to the different data considered, some parameters have to be adapted. In this context, for CLAHE,  $C_b$  is set to  $C_b = 8 \times 8$  (default value) because the objects to be detected are smaller than the bomb craters in the aerial wartime images. The parameter  $n_m$  related to the local normalization of grey values (see Section 4.1.2) is adapted to the unit of pixel, i.e.  $n_m = 5$  pixels (instead of 5 m). Likewise, the parameter  $e_{annu}$  describing the extent of the concentric annulus around the objects (see Figure 4.6) that is considered for the computation of the Bhattacharyya distance is adapted, here to  $e_{annu} = 2$  pixels (instead of 2 m). For the scene *flamingos*, it turned out beneficial neither to perform CLAHE nor the local normalization; supposedly due to various reasons, such as the already comparatively good contrast, the small and fairly overlapping objects, and the noisy image content. The reference radii of the objects of the three scenes are used to adjust  $B_{S_{max}}$  accordingly. However, as not all objects of the same scene have identical sizes (Figure 5.2),  $B_{S_{min}}$  is simply set to  $B_{S_{min}} = 0.5 \cdot B_{S_{max}}$  per scene in order to be able to also detect smaller objects (Table 5.4). In addition, the parameters  $c$  and  $d_0$  are varied due to the different gradient magnitudes and contrasts, respectively, occurring in the data. Finally,  $f_0$  is modified to allow for partial overlap of objects, but at the same time prevent their accumulation. All other parameters are identical to those in Table 5.3 and, obviously, the parameters associated with the fusion of detections and the generation of the impact map are not involved here.

**Table 5.4:** Parameter settings in the experiments with respect to population counting in images. Note that only those parameters whose values differ from those associated with the impact map generation (Table 5.3) are listed here.

Parameter	Flamingos	Cells	Stomata
$B_{S_{min}}, B_{S_{max}}$ [pixel]	[1.5, 3.0]	[6.0, 12.0]	[3.5, 7.0]
$c$	50	300	600
$d_0$	5	5	15
$f_0$	1500	4000	8000

In summary, it should be noted that in the context of the impact map generation, almost identical parameter values are used for all experiments on the aerial wartime images. Only one parameter

has been adjusted for the different sources (Lower Saxony, Italy and Salzburg). More or less similar parameter settings are also chosen for the three scenes concerning the population counting from images. The values for the parameters  $c$ ,  $d_0$  and  $f_0$  are to be chosen comparatively smallest for the scene *flamingos* due to not having performed CLAHE and local grey value normalization resulting in low gradient magnitudes and contrasts occurring in the image. Similarly, as can be concluded from the visual comparison of the scenes *cells* and *stomata* (Figure 5.2), the contrast is higher in the latter, which is why the three parameters are set to comparatively high values here.

## 5.4 Evaluation procedure

This section describes the strategy and criteria applied for the purpose of evaluating the proposed approach and its components. To obtain quantitative results in connection with both the detection of objects in images and the generation of an impact map, the evaluation process is carried out accordingly, i.e. in an object-based and a pixel-based way, respectively, and illustrated in Figure 5.3.

In the object-based evaluation (Sections 6.1, 6.2.1, 6.3 and 6.4), an automatically detected object is defined to be correct (True Positive, TP) if the distance from the centre of the detected object to a reference centre is smaller than the reference radius of this crater, otherwise it is defined as being incorrect (False Positive, FP). The precision  $P$  is the percentage of detected object centres fulfilling this criterion, i.e.

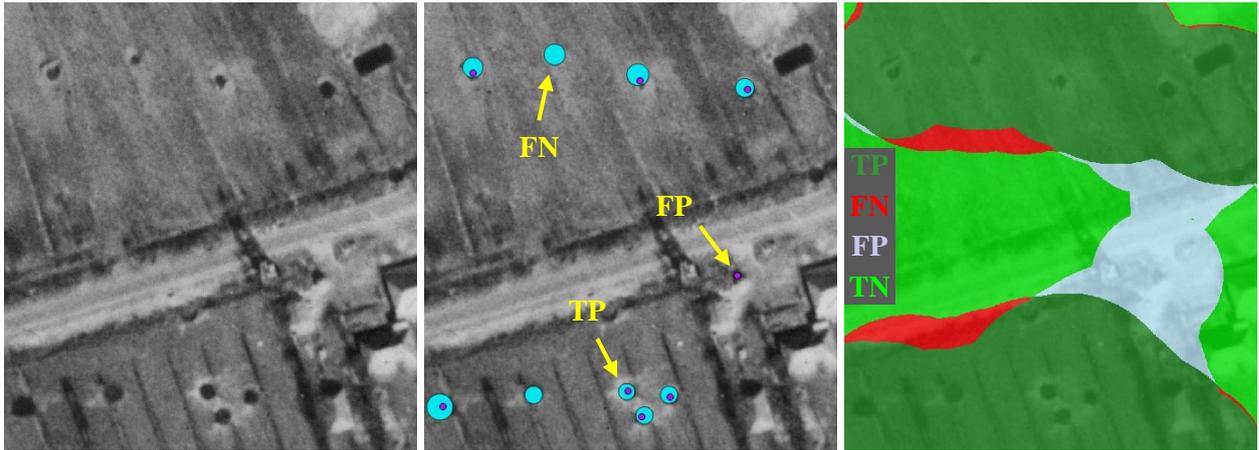
$$P = \#TP / (\#TP + \#FP) \quad (5.1)$$

where  $\#TP$  and  $\#FP$  are the numbers of TPs and FPs, respectively. If there are several detections for one reference (e.g. this would be the case if there is an accumulation of objects on, for instance, a crater), obviously, only one of them is considered as a TP, the rest as FPs. If more than one reference is eligible for a detection, the reference is assigned for which the distance of its centre to the detection centre is the smallest. A crater not detected by the proposed procedure (identified as a reference object without a detection within a circle of the reference radius) is a False Negative (FN). The recall  $R$  of the object configuration is the percentage of reference bomb craters found by the method, i.e.

$$R = \#TP / (\#TP + \#FN) \quad (5.2)$$

with  $\#FN$  being the number of FNs. The F1-score  $F1$  is the harmonic mean of  $P$  and  $R$ , i.e.

$$F1 = 2 \cdot (R \cdot P) / (R + P). \quad (5.3)$$



**Figure 5.3:** The evaluation of the results, illustrated for a subset of an aerial wartime image (left), is carried out both in an object- and a pixel-based way. In the object-based evaluation, a detection is a TP if the distance from the centre of the detected object (purple) to a reference centre is smaller than the reference radius of this crater (reference objects are shown in cyan), otherwise it is a FP, and a crater not detected by the procedure (identified as a reference object without a detection within a circle of the reference radius) is a FN; one example for each case is indicated by yellow arrows (middle). In the pixel-based evaluation based on impact maps (generated from the centres of the reference and the detections, respectively), each pixel of the two impact maps is compared and classified as described in the main text; the corresponding impact map with TP areas in dark green, FP areas in pale blue, FN areas in red, and TN areas in lime green, superimposed on the subset of the aerial wartime image (right).

The pixel-based evaluation of the results in Sections 6.2.1, 6.2.2 and 6.2.3 is based on impact maps. The reference centres of the bomb craters are used for the generation of the reference impact map (same parameter setting as for the generation of the impact map from the detection centres). In connection with redundancy (see Section 4.2), due to the inaccuracies of the georeferencing information, the centres of gravity of point sets not containing a master detection usually do not coincide with the centres of the respective reference, which would make the evaluation erroneous. To counteract this problem, these detections are moved to the closest reference of the master image within 40 m. Detections whose distance to a reference centre is larger than 40 m are not affected. The corresponding impact maps are then compared and each pixel is classified as being either a TP, FN, FP or TN. A TP is a pixel that has been correctly classified as contaminated in both, the reference and the automatic detection. FN pixels have been classified as uncontaminated by the automatic detection although they are in fact contaminated. FP pixels have been falsely classified as contaminated. Finally, a TN is a pixel that has been correctly classified as uncontaminated in both cases. The recall  $R$  is the percentage of the actually contaminated area found by the proposed method, i.e.

$$R = \frac{\text{area correctly classified as contaminated}}{\text{overall contaminated area}} ; \quad (5.4)$$

the precision  $P$  is the percentage of areas classified as contaminated in the automatic detection that lie in areas which are actually contaminated, i.e.

$$P = \frac{\text{area correctly classified as contaminated}}{\text{overall area classified as contaminated}}. \quad (5.5)$$

The F1-score  $F1$  is also determined via Equation 5.3.

Of course, recall and precision depend on each other, meaning that it is always possible to tune one at the expense of the other. In the case presented in Section 6.2.3, the main concern is the precision of the results, because areas falsely classified as having to be probed would result in high costs. Nevertheless, in all other experiments, the more neutral case is illustrated, with the goal of achieving an optimal F1-score.

## 5.5 Comparison with a deep learning based object detector

Note that, to the best of the author’s knowledge, no labelled data set for bomb crater detection in aerial wartime imagery is currently publicly available. As a result, the data used by different authors vary, making it difficult to perform meaningful comparisons. A comparison is further hindered by the fact that the results are strongly dependent on the images used, as will be shown in this thesis, but has also been reported, for example, in (Brenner et al., 2018). Therefore, in this work, the comparison is based on data used in other experiments within this thesis.

For a comparison with the proposed MPP approach for bomb crater detection, the author makes use of a state-of-the-art object detector from deep learning, namely the Faster R-CNN (Ren et al., 2015; the implementation provided by Yuxin et al. (2019) is used). The respective architecture consists of a feature extraction network, which takes an image patch as input and generates a feature map, and a region proposal network that takes the feature map as input and predicts objectness scores for predefined anchor points relative to each pixel position. Unlike Ren et al. (2015), a classifier is not applied to the regions, because the aim is not to differentiate multiple object types. Instead, the sigmoid function is applied to the objectness score of each region  $r$  to obtain a probabilistic score  $p_r$ . Proposed regions are considered as craters if  $p_r$  is larger than a threshold  $p_t$ .

This work uses ResNet-50 (He et al., 2016) as feature extraction network, pre-trained on the ImageNet data set (Deng et al., 2009; the pre-trained parameters and hyper-parameters are provided by Yuxin et al., 2019). The region proposal network is trained from scratch on image patches of size 640 x 640 pixels which are randomly cropped from the images of the training set used in this thesis. For data augmentation, random horizontal and vertical flipping and a random change of brightness within a range of  $\pm 10\%$  are performed. In total, 90.000 iterations during training

---

are conducted and the model is evaluated on the validation set every 10.000 iterations; the parameter set with the best validation performance is used for the test set.

In order to train, validate and test the model, the 55 images from Lower Saxony with known reference are used. More precisely, eight images are considered as test set; the respective test images are not used during training. Their selection is carried out in such a way that the content of the images differs, various crater manifestations are present and the number of craters varies between the images. The remaining 47 images are randomly split into 32 for training and 15 for validation. Note that four images having a spatial overlap with one of the test images and additionally stemming from the same surveillance flight are excluded, as the appearance of the objects in such overlapping images is (almost) identical. The experiment is carried out three times, each time with a different random split of training and validation data. To allow for a comparison regarding the performance stability, three runs of the MPP procedure on the eight test images mentioned above are also carried out.

Regarding the probability threshold  $p_t$ , two approaches are reported. In the first one,  $p_t$  is set to 0.5, which is the most intuitive choice. In the second approach,  $p_t$  is optimized to maximise the F1-score on the validation set in a grid search. The evaluation criteria are identical to the ones described in Section 5.4, except that while for the MPP the object centre is used, for the CNN it is the centre of the bounding box.



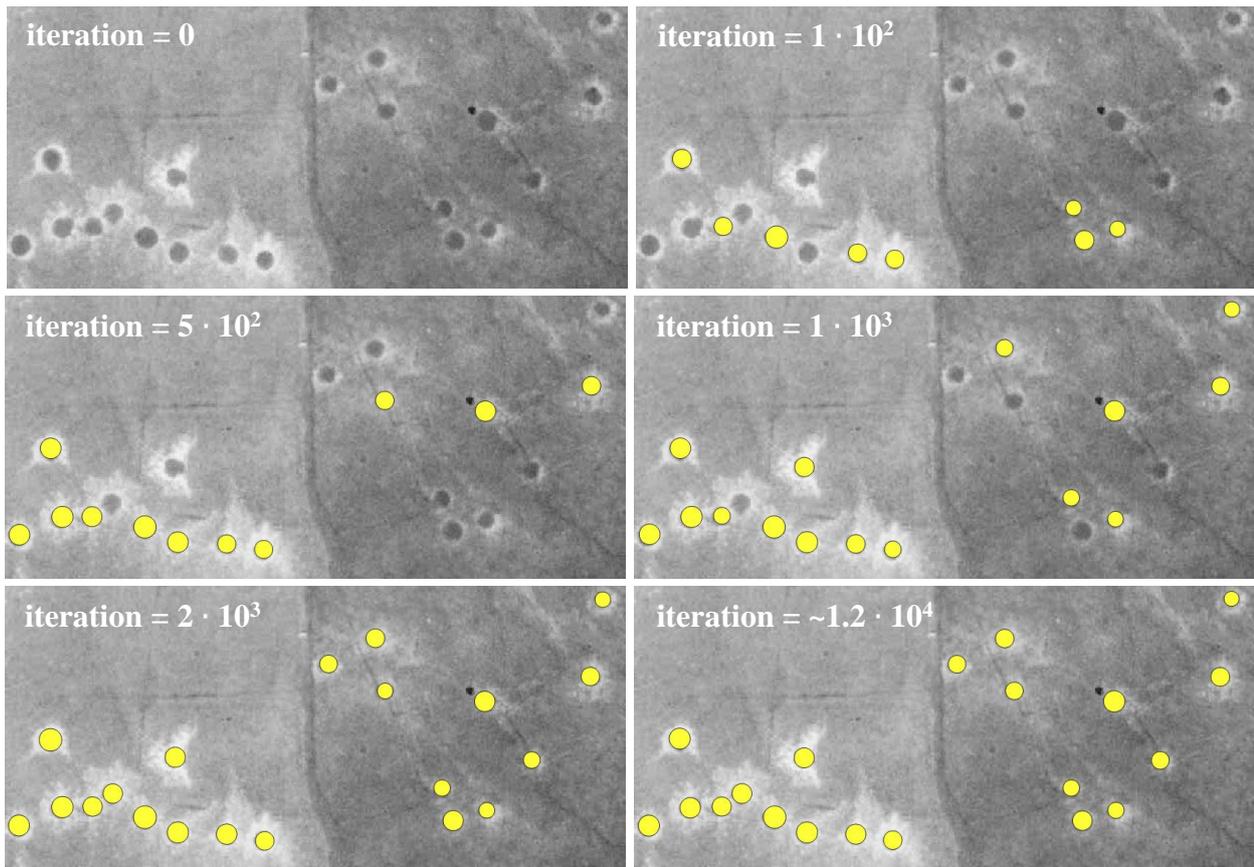
## 6 Results and Discussion

In this chapter, the experiments conducted in the context of this thesis are described and their results are presented and discussed. For this purpose, in accordance to the structure of the evaluation objectives (Section 5.1), first, the proposed marked point process (MPP) model used for bomb crater detection is analysed (Section 6.1). Section 6.2 examines the applicability of the overall approach to different aerial wartime images and the influence of using redundant image information on the results. In the final set of experiments described in this section, it is investigated whether the generated impact map can be appropriately used with respect to the proposed application scenario. For this purpose, the method is tuned to achieve a high precision. In Section 6.3, the results of the MPP are compared with those of a state-of-the-art object detector based on convolutional neural networks (CNNs). The chapter closes with investigations concerned with the question to what extent the MPP model is suitable for population counting purposes from images (Section 6.4).

### 6.1 Analysis of the model

In this section, the developed MPP model for the detection of bomb craters in aerial wartime images is analysed. For this purpose, individual parts of the model are examined in more detail based on images from Lower Saxony. More precisely, the influence on the results of the generated random numbers during sampling (Section 6.1.1), the choice of the object model, in particular comparing the circular and elliptical models (Section 6.1.2), and each term in the energy function (Section 6.1.3), are investigated.

Before analysing the MPP model as described above, the sampling process in which an object configuration is iteratively created by adding objects to or removing them from the configuration or by modifying their parameters, is illustrated using Figure 6.1. Here, for a part of an image from Lower Saxony with comparatively very easy image content (e.g. craters can be clearly differentiated from background objects and there are basically no disturbing objects such as trees and shadows that could lead to false detections), it is shown how the object configuration evolves in the iteration process. After about  $2 \cdot 10^3$  iterations, the number of objects does not change anymore and every bomb crater is represented by one object. Moreover, there are no false detections and the object borders correspond well to the borders of the craters.



**Figure 6.1:** Results for a part of an image from Lower Saxony after  $1 \cdot 10^2$ ,  $5 \cdot 10^2$ ,  $1 \cdot 10^3$ ,  $2 \cdot 10^3$  and  $\sim 1.2 \cdot 10^4$  (stop criterion takes effect) iterations, with the respective object configuration represented by yellow circles.

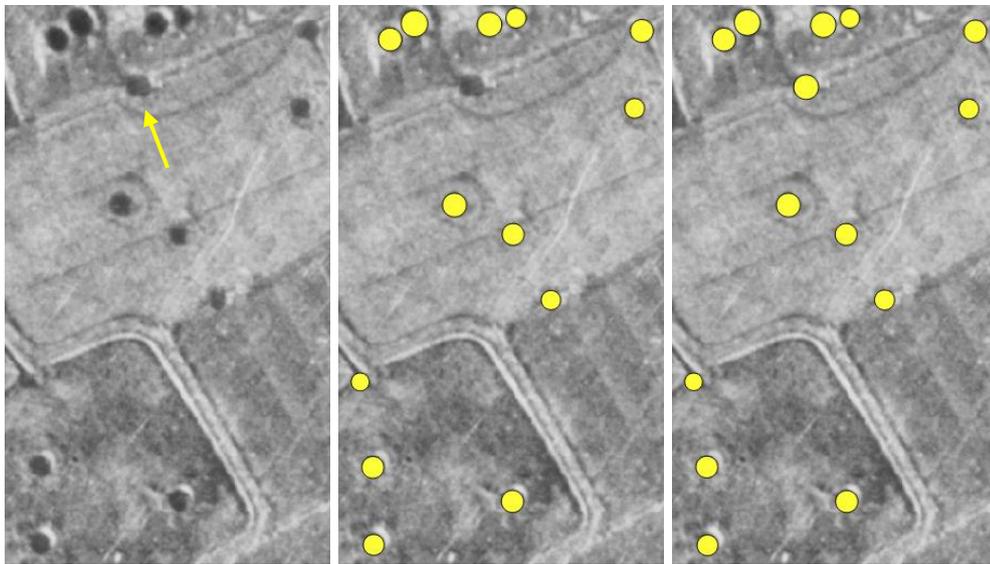
### 6.1.1 Random numbers

A MPP relies on random numbers being drawn at different stages of the process, such as when deciding whether or not to accept a new state or which move should be selected to propose a new object configuration. Consequently, the results will vary between different runs. In order to analyse the overall influence of the random number generation on the results, the MPP procedure is run 50 times with identical parameter settings on a representative selection of 10 images from Lower Saxony, meaning that in these images, for instance, different lighting situations and crater appearances exist and the image content varies. In this context, the standard deviations of the number of objects in the final results are calculated, i.e. there is one standard deviation for each of the 10 images stemming from the 50 runs on the same image. Moreover, the standard deviations of the F1-scores (object-based evaluation) are analysed.

If different random numbers are generated in each run, the number of objects in the final object configuration is similar; the standard deviation varies by 3.9 % on average. An example is given in Figure 6.2, where the final object configuration in one test run consists of one object more

(Figure 6.2 right) compared to another run (Figure 6.2 middle). Here, one crater is not detected. With regard to the position and size of the created objects, almost no differences can be observed for this example. The small differences in the number of objects are also reflected in the standard deviations of the F1-scores, which shows that the final F1-scores only vary slightly around their mean ( $\pm 0.2\%$ ).

The results show that the final object configuration between multiple runs differs only slightly, if at all, when the same experiment is carried out several times. Thus, in general, the performance can be considered as stable, as the influence of the random number generator on the quality of the results is marginal.



**Figure 6.2:** The final object configuration may differ slightly due to the random components of the algorithm. Subset of an image from Lower Saxony with a bomb crater highlighted by an arrow (left). A result of the MPP in which that crater is not detected (middle). A result of the MPP in which that crater is detected (right).

### 6.1.2 Object model

To represent the objects to be detected in the data, various object models may be used. As bomb craters in aerial wartime images do not always appear as being circular, but sometimes as being slightly elliptical, an ellipse seems to be more obvious than a circle as an object model. In order to investigate whether this holds true, 55 images with a reference from Lower Saxony are considered and the results achieved are compared when using a circle vs. an ellipse as object model. For this purpose, the average values of recall, precision and F1-score are computed.

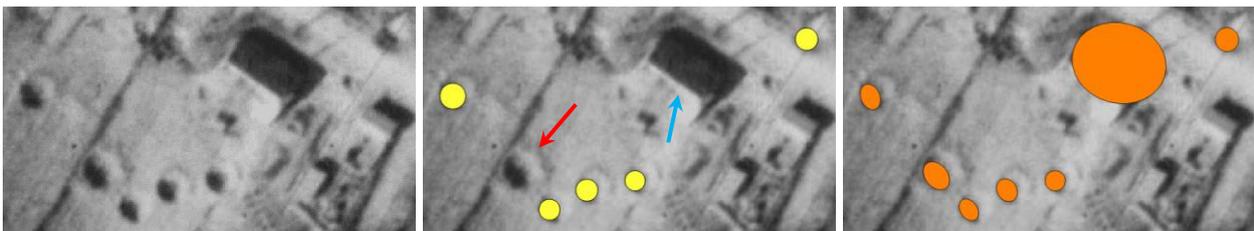
The resulting quality measures are given in Table 6.1. It can be seen that the F1-score is almost the same for both object models. For the elliptical model, the recall is about 1 % higher than for

the circular model, whereas its precision is about 1 % lower than for the object model circle. This can be explained by the fact that some bomb craters with a more elliptical shape are detected to a larger extent, while additional false positive objects that appear similar to elliptical bomb craters are also found. A related example for a subset of an image from Lower Saxony is shown in Figure 6.3.

Even though using a circle as an object model slightly reduces the number of correct detections, compared to an ellipse, a circle keeps the search space in the optimization process smaller due to its lower number of parameters, making the model generally more stable. Furthermore, as can be concluded from the quality measures, there are comparatively few craters with a stronger elliptical shape, but at least as many objects that are not falsely detected when the circle is used as object model. Finally, in terms of the proposed application scenario, the precision of the results is of primary concern, which is higher for the circular model. Recall that the focus of the experiment here is on obtaining results with an optimal F1-score, which is similar for both models. Therefore, in all further experiments within this thesis, the circle is used as object model.

**Table 6.1:** Evaluation results (recall  $R$ , precision  $P$ , F1-score  $F1$ ) based on 55 images from Lower Saxony for two different object models, namely a circle and an ellipse.

	Circle			Ellipse		
	R [%]	P [%]	F1 [%]	R [%]	P [%]	F1 [%]
<b>Results</b>	47.0	64.3	54.3	47.6	63.7	54.5



**Figure 6.3:** Comparison of the object models circle and ellipse based on a subset of an image from Lower Saxony (left). While for the circular model a bomb crater (red arrow) and a shadow cast by a building roof (cyan arrow) are not detected (middle), the opposite is true for the elliptical model (right).

### 6.1.3 Energy terms

Besides the object model to be chosen, for a MPP the energy function has to be defined as well. In this work, it consists of four energy terms (cf. Section 4.1.2); the analysis of the importance of each term is carried out by successively excluding one of them. For this set of experiments, all 55

images with a reference from Lower Saxony are used. In addition to determining the particular quality measures, the respective effects on the object configuration are illustrated qualitatively by means of one example each.

The quantitative results as well as the qualitative examples can be found in Table 6.2 and Figure 6.4, respectively. In Figure 6.4, the left column shows subsets of different aerial wartime images, the middle column the respective results of the MPP based on the proposed model and the right column the situation when one of the energy terms is not considered; Table 6.2 is structured in the same order. By applying the proposed model, all bomb craters on the subsets shown are detected and no false detections exist (Figure 6.4 middle column). When not considering the term of the data energy addressing high gradient magnitudes ( $f_G = 0$ , Equation 4.2), there are more false detections and the number of detected craters is reduced; the average F1-score is decreased by about 3 % (cf. Table 6.2). One reason for a lower recall is that not all bomb craters have a distinct contrast between their inside and the annulus, e.g. due to no ejecta or an ejecta containing besides brighter also darker pixels (Figure 6.4 right column, top). On the other hand, the consideration of  $U_G$  (Equation 4.2) can help to eliminate false detections if there is a distinct contrast, but the gradient magnitudes at the object border are rather low (e.g. to be found in forests). When not taking into account the homogeneity term ( $f_H = 0$ , Equation 4.3), the number of false detections usually increases (Figure 6.4 right column, second from top). However, when applying the full model, some bomb craters might also not be detected. This may occur because bomb craters do not always have homogeneous dark grey values inside, or have an elliptic shape, which can also lead to higher standard deviations for the grey values. All these observations are reflected in the quantitative evaluation metrics: when the data term  $U_H$  (Equation 4.3) is excluded, the average F1-score decreases only slightly, but the average precision is comparatively reduced more (approx. 4 %) than the average recall is increased (approx. 2 %; cf. Table 6.2). When excluding the data energy term related to the contrast between the objects and their annuli ( $f_B = 0$ , Equation 4.5), while the recall stays basically the same, there are considerably more false detections, resulting in a precision that is on average reduced by ca. 9 % (cf. Table 6.2). Consequently, as already expected, this term helps in distinguishing between bomb craters and other objects (Figure 6.4 right column, second from bottom). Finally, if the prior energy term that penalizes the overlap of objects is excluded ( $f_o = 0$ , Equation 4.8), detections accumulate with their centres and radii being almost identical for the same object (Figure 6.4 right column, bottom; close-up view for the bomb crater on the right). As a result, the algorithm does not converge due to the constantly decreasing energy and stops only after reaching a specified maximum number of iterations (here  $10^7$ ). Therefore, reporting the precision is refrained from; the recall of 46.9 % is slightly lower in comparison to the full model (cf. Table 6.2).

The results show that the proposed model is reasonable as it performs best among the alternatives evaluated. If certain terms are not considered, the quality of the results decreases or an appropriate detection of bomb crater in aerial wartime images is no longer possible.

**Table 6.2:** Evaluation results (recall  $R$ , precision  $P$ , F1-score  $F1$ ) based on 55 images from Lower Saxony when excluding successively energy terms; (n/d: not defined; \*: no convergence, for more details see main text).

Method	R [%]	P [%]	F1 [%]
all terms	47.0	64.3	54.3
without gradient term ( $f_G = 0$ )	44.2	60.7	51.1
without homogeneity term ( $f_H = 0$ )	48.9	60.4	54.0
without contrast term ( $f_B = 0$ )	47.2	55.2	50.9
without overlap term ( $f_O = 0$ )	46.9*	-*	-*

## 6.2 Detection of bomb craters and impact map generation

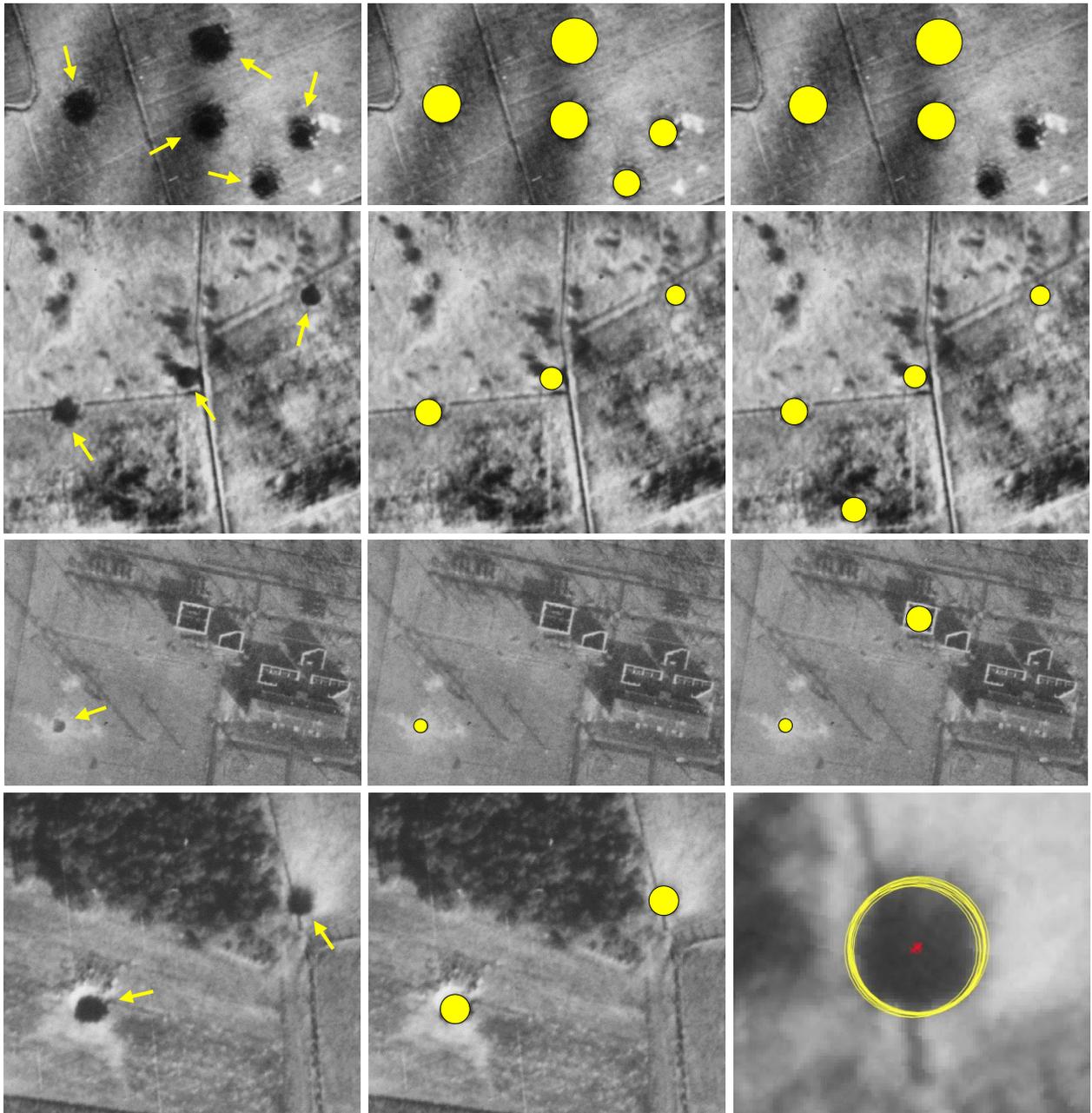
The MPP model is designed for detecting bomb craters in aerial wartime imagery. In this context, the first set of experiments in this section contains investigations regarding the general applicability of the approach to differing aerial wartime images from Lower Saxony, Italy and Salzburg (Section 6.2.1). In another set of experiments, the influence of using redundant image information on the quality of the results is examined (Section 6.2.2). Lastly, it is investigated whether the impact map derived from the detection results can be appropriately used with respect to the proposed application scenario. For this purpose, the method is tuned to achieve a high precision (Section 6.2.3).

### 6.2.1 Performance for different aerial wartime images

In the following, it is investigated how well the proposed approach performs on different aerial wartime images. For this purpose, images from Lower Saxony (LS), Italy (I) and Salzburg (S) are available. These images differ, for instance, in terms of their content, not only between the three sources, but also within a source (cf. Section 5.2), allowing to assess the general validity of the approach and to identify its potential limitations. In this context, the numerical values of recall, precision and the F1-score for the object- and pixel-based evaluation for each image, including the respective averages, are computed and differences in the quality measures of the two evaluation principles are also discussed. Finally, in addition to the average results per source, the averages for the masked images (Section 4.1.5) are also shown and the influence of the masking procedure on the results is discussed. Note that the images of LS are split into two subsets LS1 (28 images) and LS2 (27 images), because only the latter subset is used in the investigations reported in Sections 6.2.2 and 6.2.3.

The numerical values for recall, precision and the F1-score can be found in Table 6.3. It can be seen that for the images from Lower Saxony an average recall and precision of 48 % and 64 %, respectively, are achieved.

respectively, is achieved for the pixel-based evaluation (LS, columns 7-9), and 47 % and 64 % for the object-based evaluation (columns 4-6); the F1-score increases slightly from 54 % to 55 %. The higher recall for the pixel-based evaluation can be explained by the fact that in areas with more craters not every crater needs to be found, as the detection of surrounding craters already leads to



**Figure 6.4:** Comparison of the final object configurations for subsets of images from Lower Saxony (left column; bomb craters are indicated by yellow arrows) when successively excluding one of the terms of the energy function (right column); the respective results of the MPP based on the proposed model are also shown (middle column). Results achieved when excluding the term  $U_G$  (right column, top),  $U_H$  (right column, second from top),  $U_B$  (right column, second from bottom) and  $U_O$  (right column, bottom; close-up view for the bomb crater on the right).

**Table 6.3:** Evaluation results (recall  $R$ , precision  $P$ , F1-score  $F1$ ), sorted by the number of craters per image, for the images from Lower Saxony (LS), Italy (I) and Salzburg (S); (n/d: not defined). Besides the average results per source, the averages for the masked images (masked) are also shown. More details on the corresponding number of craters as well as the respective quality measures of the images marked with an asterisk can be found in Table 6.4.

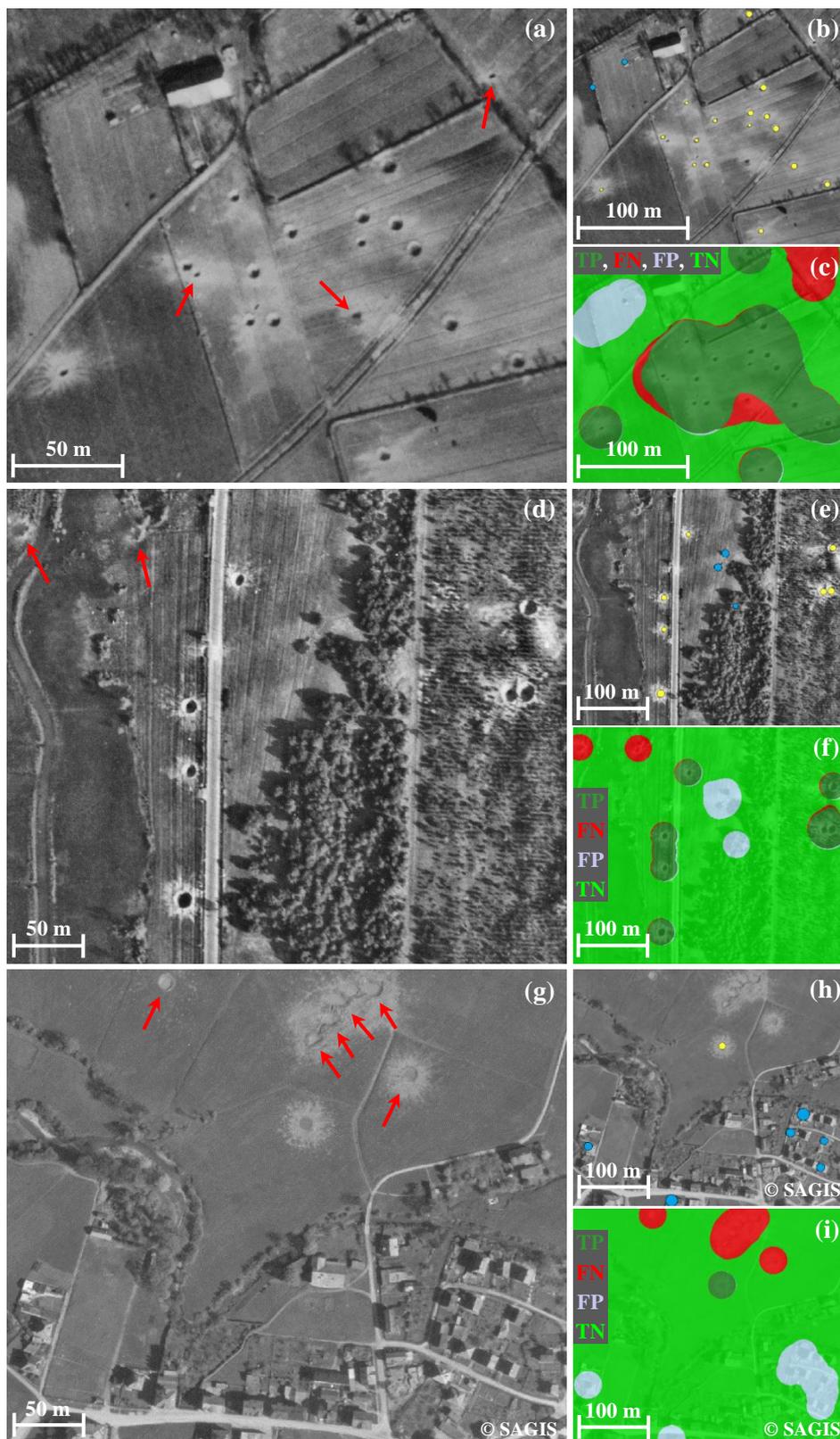
	Image ID	Number of craters	object-based			pixel-based		
			R [%]	P [%]	F1 [%]	R [%]	P [%]	F1 [%]
LS	1 <sub>LS1</sub>	0	n/d	0	n/d	n/d	0	n/d
	2 <sub>LS1</sub>	0	n/d	0	n/d	n/d	0	n/d
	3 <sub>LS1</sub>	0	n/d	0	n/d	n/d	0	n/d
	4 <sub>LS1</sub>	0	n/d	0	n/d	n/d	0	n/d
	5 <sub>LS1</sub>	9	11	11	11	10	11	10
	6 <sub>LS1</sub>	14	43	16	23	57	18	27
	7 <sub>LS1</sub>	16	13	40	19	12	40	19
	8 <sub>LS1</sub>	17	12	5	7	12	5	7
	9 <sub>LS1</sub>	18	39	44	41	35	40	38
	10 <sub>LS1</sub>	24	13	12	12	12	11	12
	11 <sub>LS1</sub>	26	19	36	25	18	35	24
	12 <sub>LS1</sub>	37	49	75	59	47	73	57
	13 <sub>LS1</sub>	53	47	81	60	45	80	57
	14 <sub>LS1</sub>	57	44	47	45	46	51	48
	15 <sub>LS1</sub>	114	25	31	27	31	33	32
	16 <sub>LS1</sub>	132	26	61	36	29	60	39
	17 <sub>LS1</sub>	140	45	62	52	48	64	55
	18 <sub>LS1</sub>	140	44	21	29	41	20	27
	19 <sub>LS1</sub>	172	33	51	40	33	50	40
	20 <sub>LS1</sub>	239	59	67	63	58	67	62
	21 <sub>LS1</sub>	300	35	47	40	39	46	42
	22 <sub>LS1</sub>	301	53	73	62	58	73	64
	23 <sub>LS1</sub>	345	30	76	43	29	76	42
	24 <sub>LS1</sub>	425	68	98	80	74	98	84
	25 <sub>LS1</sub>	495	61	64	62	67	66	66
	26 <sub>LS1</sub>	589	71	95	81	74	94	83
	27 <sub>LS1</sub>	813	29	96	45	32	96	48
	28 <sub>LS1</sub>	938	70	76	73	78	75	76
1 - 28 <sub>LS1</sub>	5414	51	68	58	53	68	60	
1 - 27 <sub>LS2</sub> *	3467	41	58	48	41	59	48	
<b>Average results LS</b>			<b>47</b>	<b>64</b>	<b>54</b>	<b>48</b>	<b>64</b>	<b>55</b>
<b>Average results LS masked</b>			<b>47</b>	<b>82</b>	<b>60</b>	<b>49</b>	<b>83</b>	<b>61</b>
I	1 <sub>I</sub>	60	2	2	2	2	2	2
	2 <sub>I</sub>	105	22	49	30	23	48	31
	3 <sub>I</sub>	119	23	41	29	29	52	37
	4 <sub>I</sub>	146	36	56	44	39	57	46
	5 <sub>I</sub>	250	10	19	13	11	22	15
	6 <sub>I</sub>	321	50	61	55	54	64	59
	7 <sub>I</sub>	411	11	49	18	11	49	18
	8 <sub>I</sub>	434	16	11	13	24	14	18
	9 <sub>I</sub>	753	49	92	64	59	94	73
	10 <sub>I</sub>	1069	57	89	69	60	88	71
<b>Average results I</b>			<b>38</b>	<b>56</b>	<b>45</b>	<b>42</b>	<b>58</b>	<b>49</b>
<b>Average results I masked</b>			<b>37</b>	<b>83</b>	<b>51</b>	<b>43</b>	<b>85</b>	<b>57</b>
S	1 <sub>S</sub>	5	0	0	n/d	0	0	n/d
	2 <sub>S</sub>	15	20	27	23	17	24	20
	3 <sub>S</sub>	19	0	0	n/d	0	0	n/d
	4 <sub>S</sub>	51	2	2	2	2	2	2
	5 <sub>S</sub>	75	31	20	24	31	20	24
	6 <sub>S</sub>	80	6	12	8	6	13	8
	7 <sub>S</sub>	220	25	39	31	26	39	32
	8 <sub>S</sub>	236	31	52	39	33	54	41
	9 <sub>S</sub>	244	25	59	35	25	60	35
<b>Average results S</b>			<b>23</b>	<b>34</b>	<b>28</b>	<b>24</b>	<b>35</b>	<b>29</b>
<b>Average results S masked</b>			<b>23</b>	<b>95</b>	<b>37</b>	<b>24</b>	<b>95</b>	<b>39</b>

the area being classified as contaminated (see also Figure 6.5 a-c). This can be observed particularly in images with a high number of craters (e.g. 21<sub>LS1</sub>, 25<sub>LS1</sub>, 28<sub>LS1</sub>), as the probability of crater clusters is higher. The precision is similar in both cases. Compared to the object-based results, the slightly lower values in precision in the pixel-based case may be due to small geometrical differences between the centres of the detected craters and the centres of the reference; thus, areas classified as contaminated differ somewhat for reference and detection. This is possible, for instance, because the shadow in the crater is not always circular, so the detected centre is shifted towards the side of the shadow cast in the crater. Furthermore, it has been observed that the reference centres do not always coincide with the centre of the craters. On the other hand, slightly higher values in precision can be explained by FPs very close to crater centres, as then, by coincidence, a certain amount of the respective crater surroundings is correctly classified as contaminated.

Regarding the comparison between object- and pixel-based results, these observations also hold true for the images of Italy and Salzburg. For the former, recall and precision are both moderately increased for the pixel-based evaluation, achieving 42 % and 58 % (columns 7-9) compared to 38 % and 56 % for the object-based evaluation (columns 4-6), respectively. The two sets of quality measures for the Salzburg images are very similar also, with a recall and precision of 23 % and 34 %, respectively, for the object-based evaluation and 24 % and 35 %, respectively, for the pixel-based evaluation.

Comparing the results related to the three different sources, clear differences can be observed: For the object-based evaluation, the mean F1-score for the images from Lower Saxony is 54 % while it is considerably lower for the data from Italy and Salzburg; this trend is similar for the pixel-based evaluation. Thus, the results confirm the observation made earlier that overall the image content of Lower Saxony is not as challenging as the one of Italy and especially Salzburg. In this context, Figure 6.5 shows exemplary results for subsets of the images 25<sub>LS1</sub>, 2<sub>I</sub> and 8<sub>S</sub> (Figure 6.5 a, d and g). The corresponding object configurations as well as the corresponding impact maps are given in Figure 6.5 b-c, Figure 6.5 e-f and Figure 6.5 h-i. For the example of Lower Saxony, there are only two FPs, while there are more FPs for the subsets of Italy and Salzburg (Figure 6.5 b, e, h; cyan circles). In the examples, the detection of all craters becomes challenging due to low contrast and the occurrence of strong small elliptical shadows as well as bright craters (Figure 6.5 b, e, h; red arrows). These observations are reflected in the superimposition of the two impact maps generated from the reference centres and the centres of the automatic detection (Figure 6.5 c, f, i). That means, for instance in the example of Lower Saxony, comparatively few areas are falsely classified as contaminated (Figure 6.5 c, pale blue) and uncontaminated (Figure 6.5 c, red), which is different for the example of Salzburg. Here, only a small area is correctly detected as contaminated (Figure 6.5 i, dark green), whereas the areas falsely classified as contaminated (Figure 6.5 i, pale blue) or uncontaminated (Figure 6.5 i, red) are larger.

When choosing the same parameter setting for all three sources, i.e. the parameter  $c$  of the first data term is not varied (Table 5.3) but set to  $c = 1200$  (optimal in terms of the F1-score) for each

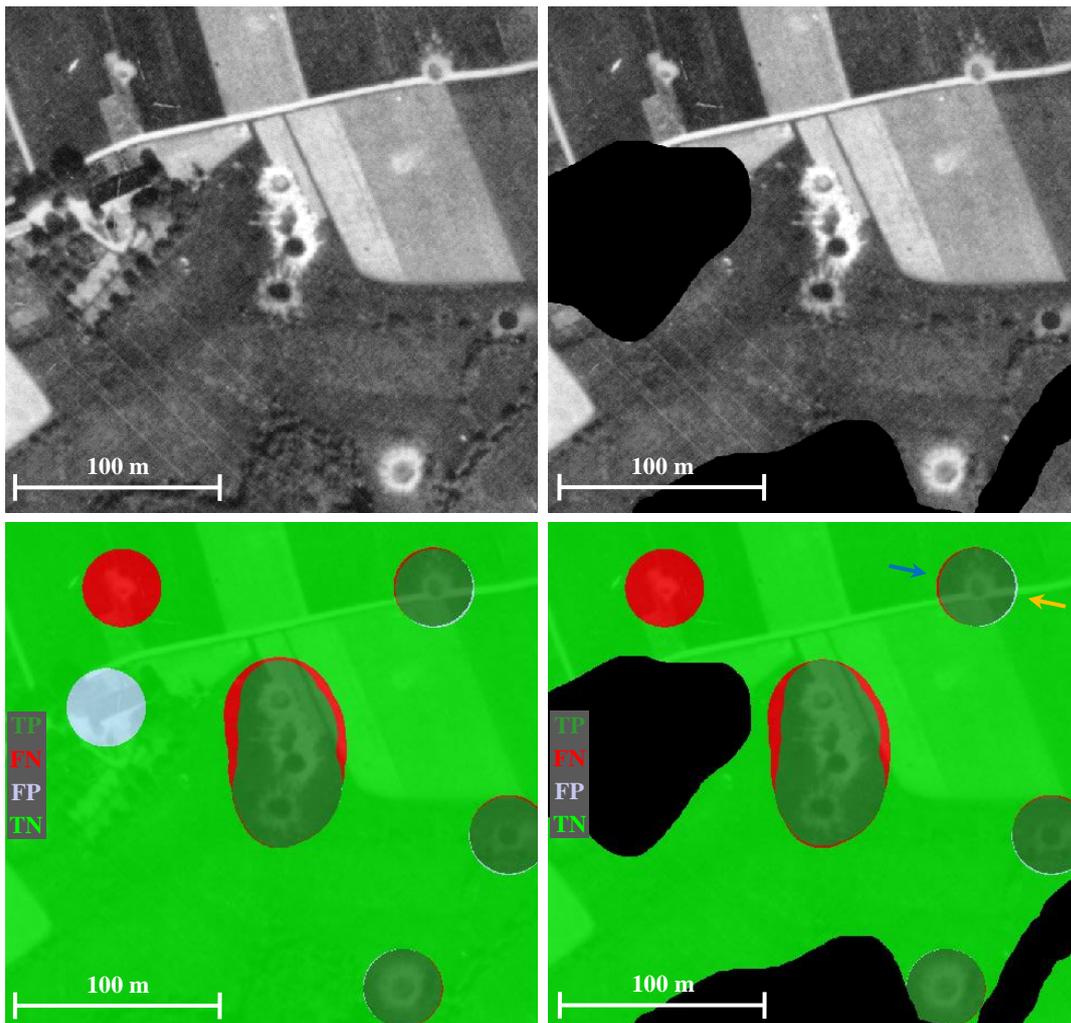


**Figure 6.5:** (a, d, g) Subsets of images 25<sub>LS1</sub>, 2<sub>I</sub> and 8<sub>S</sub>, respectively. (b, e, h) The object configurations with TPs in yellow and FPs in cyan; FNs are tagged by red arrows in (a), (d) and (g). (c, f, i) Superimpositions of the corresponding impact map and evaluation with TP areas in dark green, FN areas in red, FP areas in pale blue and TN areas in lime green.

source, the resultant average F1-score (pixel-based evaluation) over the total of 74 images is 50 %, compared to 51 % when the optimal values per source are chosen (not shown in Table 6.3). Thus, there are no major differences regarding the quality of the generated impact map, meaning that selecting an identical value for  $c$  for the different sources would be conceivable. On the other hand, in general, the results for the individual images vary considerably. In some cases, such as for images  $24_{LS1}$ ,  $26_{LS1}$ ,  $28_{LS1}$ ,  $9_I$  and  $10_I$  the F1-scores are acceptable with values larger than 70 %. However, the impact maps for other images (e.g.  $5_{LS1}$  -  $8_{LS1}$ ,  $10_{LS1}$ ,  $11_{LS1}$ ,  $1_I$ ,  $5_I$ ,  $7_I$ ,  $8_I$  and basically all images from Salzburg) are far worse. It should be noted that if there is a small number of craters, the evaluation results should be considered with caution, as a small number of errors already has a significant impact on the quality measures.

In connection with the masked data it can be seen from Table 6.3 that, as to be expected, the average recall stays almost the same for all three sources for both types of evaluation. Slightly lower values in the recall can be attributed to the stochastic behaviour of the method as well as to the fact that during the manual masking, one or the other crater was eventually masked after all, at least partially, so that the method can no longer detect them. On the other hand, it is evident that high static objects and especially their shadows are indeed responsible for numerous false detections. This is reflected in the average values of the precision, which are 82 %, 83 % and 95 % for LS, I and S, respectively, in the object-based evaluation. Thus, the average precision is increased by 18 %, 27 % and 61 %, respectively, compared to the non-masked images. This is very similar in the pixel-based evaluation where, due to the masking, the average precision is increased by 19 %, 27 % and 60 %, respectively, resulting in average values of 83 %, 85 % and 95 %. The comparatively high increase in precision for the images from Salzburg can be attributed to the fact that the parameter  $c$  is set to the highest value for these images (Table 5.3) and thus in total fewer objects are detected. If  $c$  is set, for example, to  $c = 1000$ , as for the images from Lower Saxony, the average values for recall and precision become more similar (29 % and 79 %, respectively), leading to an average F1-score of 43 % for the object-based evaluation. Obviously, the other F1-scores (Table 6.3) are also improved due to the increased precision caused by the masking; the average values for the data from LS, I and S are 60 %, 51 % and 37 % for the object-based evaluation and 61 %, 57 % and 39 % for the pixel-based evaluation.

A qualitative example illustrating the effect of the masks on the results based on a subset of image  $24_{LS2}$  is given in Figure 6.6. In the exemplary subset, for the masked data (Figure 6.6 top right), there are no false detections and, thus, the corresponding impact map basically does not show areas falsely classified as contaminated (Figure 6.6 bottom right). Note that there are usually minor areas falsely classified at certain boundaries of the classified areas due to the small locational deviations between the detection and reference centres; an example related to such an FP or FN area is indicated in the same figure by an orange and blue arrow, respectively. This is different for the non-masked data (Figure 6.6 top left), where in the example there is additionally a larger area falsely classified as contaminated stemming from one false detection (Figure 6.6 bottom left). Apart from that, the two impact maps are almost identical. Having a look to the quantitative results



**Figure 6.6:** Qualitative comparison of the results for non-masked (top left) and masked (black pixels; top right) data based on a subset from image 24<sub>LS2</sub> (Table 6.4). Corresponding impact maps superimposed on the subset of the aerial wartime image and evaluation with TP areas in dark green, FN areas in red, FP areas in pale blue and TN areas in lime green. While there is an area falsely classified as contaminated (stemming from one false detection) in the left part of the image (pale blue, bottom left), it does not occur in the other case due to the masking of the static objects with a certain height including their shadows (bottom right).

of the whole image, the recall is almost unchanged, but the precision increases from 11 % to 59 % when the method is run on the masked data.

The results show that the presented MPP method can in principle be successfully applied for the detection of bomb craters in different aerial wartime images. However, the mean quality measures for all three sources (LS, I and S) are on a rather low level, especially for the images of Salzburg, and the quality of the impact maps derived from the detected objects is only slightly better. In this context, investigations on the masked data show that objects with a certain height (e.g. forests, trees and buildings) and especially their shadows are responsible for a large number

of false detections. Thus, if stereoscopic imagery is available, one way of counteracting these limitations would be to integrate 3D information. Another one is to consider information from other sources, such as historical maps and GIS data, to find and subsequently exclude such areas from further processing. Note that in the remainder of this thesis the information just mentioned (3D information as well as information from other sources) will be referred to as additional information. Furthermore, the height information could be embedded into the energy function of the MPP in order to improve the detection of bomb craters due to their characteristic surface shape.

Finally, the results strongly depend on the images used, as has been also reported, for example, by Brenner et al. (2018). In principle, this does not only apply to images that show different scenes. For instance, it is also possible that images of different surveillance flights showing the same scene are of comparatively good and poor quality, respectively, resulting in varying outcomes. To account for these aspects in an automatic way, at least partly, and as there are usually multiple images covering the same area, the use of redundant image information seems to be beneficial.

### 6.2.2 Redundant image information

In this section, the influence of using redundant image information, i.e. of combining the MPP results of multiple coarsely georeferenced panchromatic images covering the investigated area, on the quality of the results is studied. In this context, the pixel-based quality measures for the comparison of the single and the multiple image approach based on the two data sets from Lower Saxony (Table 5.1) are computed and analysed. Similarly to the previous section, also the masked data is taken into account.

The results of the investigations can be found in Table 6.4. Here, in addition to the quality measures for each of the 27 images from the three different regions, the average results for the two data sets (DS A and DS B) as well as for the combination of both data sets (DS A + DS B; non-masked and masked data) are shown. Furthermore, the mean number of images considered in the case of the multiple image approach (NI) and the number of bomb craters (NC) are given.

First, the results without masks are presented. They show that an average recall and precision of 64 % and 77 %, respectively, can be achieved for the multiple image approach (columns 9-11). Compared to the single image results (columns 6-8) with a recall and precision of 41 % and 59 %, respectively, the F1-score increases strongly from 48 % to 70 %. For DS A, the increase in F1-score is 23 % (69 % compared to 46 %, average results DS A) whereas for DS B the increase is slightly lower with 18 % (71 % compared to 53 %, average results DS B).

The F1-score of each image improves in 25 out of 27 cases, the only exceptions being images 14<sub>LS2</sub> and 23<sub>LS2</sub>, both from DS A (regions 2 and 3, respectively). Note that the loss in F1-score for these two images is small with 2 % and 3 %, respectively (cf. Table 6.4). Additionally, the improvement in F1-score within region 3 of DS A is small compared to the other two regions. This can be explained by the comparatively low number of images taken into account (about half as

many as in regions 1 and 2). Furthermore, some image areas are only covered by one or very few other images, which is also the case for image 23<sub>LS2</sub>, and/or the additional images contain lots of clouds (applies also to image 14<sub>LS2</sub>, i.e. some areas are mainly covered by images with clouds). In consequence, bomb craters found at first are removed again, which has a direct negative effect on the recall. This is due to the fact that the parameter  $MinNumDet_{PS}$  from Section 4.2 is set to 4; see also discussion at the end of this section. Similarly, when only a relatively small number of additional images is available, objects actually representing craters are also eliminated if they are not (or no longer) representative for the predefined MPP model (e.g. if craters have been filled up with soil in the meantime). Moreover, for DS B in region 2, the increase in the F1-score is relatively low. One reason for this is that the F1-scores are already comparatively high for the single image approach in this region.

On the other hand, as already mentioned, the F1-score increases almost in all images, in many cases significantly (e.g. 1<sub>LS2</sub>, 7<sub>LS2</sub>, 10<sub>LS2</sub>, 11<sub>LS2</sub>, 13<sub>LS2</sub>, 21<sub>LS2</sub>, 26<sub>LS2</sub>) if redundant image information is taken into account. Figure 6.7 illustrates an example for which the recall of the results is increased by exploiting redundant image information (Figure 6.7 a-c). In Figure 6.7 a, a subset of image 1<sub>LS2</sub> with reference centres of the bomb craters (in turquoise) is shown. The shadow cast by clouds (mainly the lower and right part in Figure 6.7 a) leads to poor contrast and furthermore some craters do not appear dark, possibly due to a previous filling with soil (e.g. two of the three craters close together to the north of the river, Figure 6.7 a). The resulting impact maps (centres are shown in yellow) can be found in Figure 6.7 b and Figure 6.7 c. By combining the detection results, the areas falsely classified as uncontaminated almost completely vanish (Figure 6.7 b-c, red / dark green). The areas falsely classified as contaminated (Figure 6.7 b-c, pale blue) are marginal in both cases (such an area is indicated by a black arrow in Figure 6.7 b and Figure 6.7 c, respectively). Two exemplary cases showing an improvement in precision can be found in Figure 6.7 d-i. Here, high-contrast shadows of trees (Figure 6.7 d) or buildings (Figure 6.7 g) lead to FPs, resulting in areas being falsely classified as contaminated in the impact maps (Figure 6.7e, h; pale blue). After the combination of the detection results, there are no FPs and, thus, no falsely classified areas, left (Figure 6.7 f, i).

In connection with the masked data, a strong improvement in precision can be observed when considering multiple images (Table 6.4). More precisely, it increases from 77 % to 88 %, i.e. by 11 %. The recall stays almost the same with 63 % compared to 64 % for the non-masked data, resulting in a F1-score of 74 % (70 % in the non-masked case). Compared to the single image approach, where the precision increases from 59 % to 81 % and, thus, by 22 %, the increase in precision is only half as large. This can be attributed to the fact that by combining the results in the multiple image approach, a large number of false detections caused by high objects and their shadows, such as those shown in Figure 6.5 b, e, h and Figure 6.7 e, h, could already be eliminated.

Overall, the experimental results analysed in this section demonstrate the benefit of using redundant image information. By combining the individual detection results, the quality of the

**Table 6.4:** Evaluation results (recall  $R$ , precision  $P$ , F1-score  $F1$ ) for the single (SIA) and multiple (MIA) image approach for the two data sets DS A and DS B from Lower Saxony (LS), each part of the three regions (RG). For each master image (Image ID), the mean number of images considered within the multiple image approach (NI) and the number of bomb craters (NC) is shown.

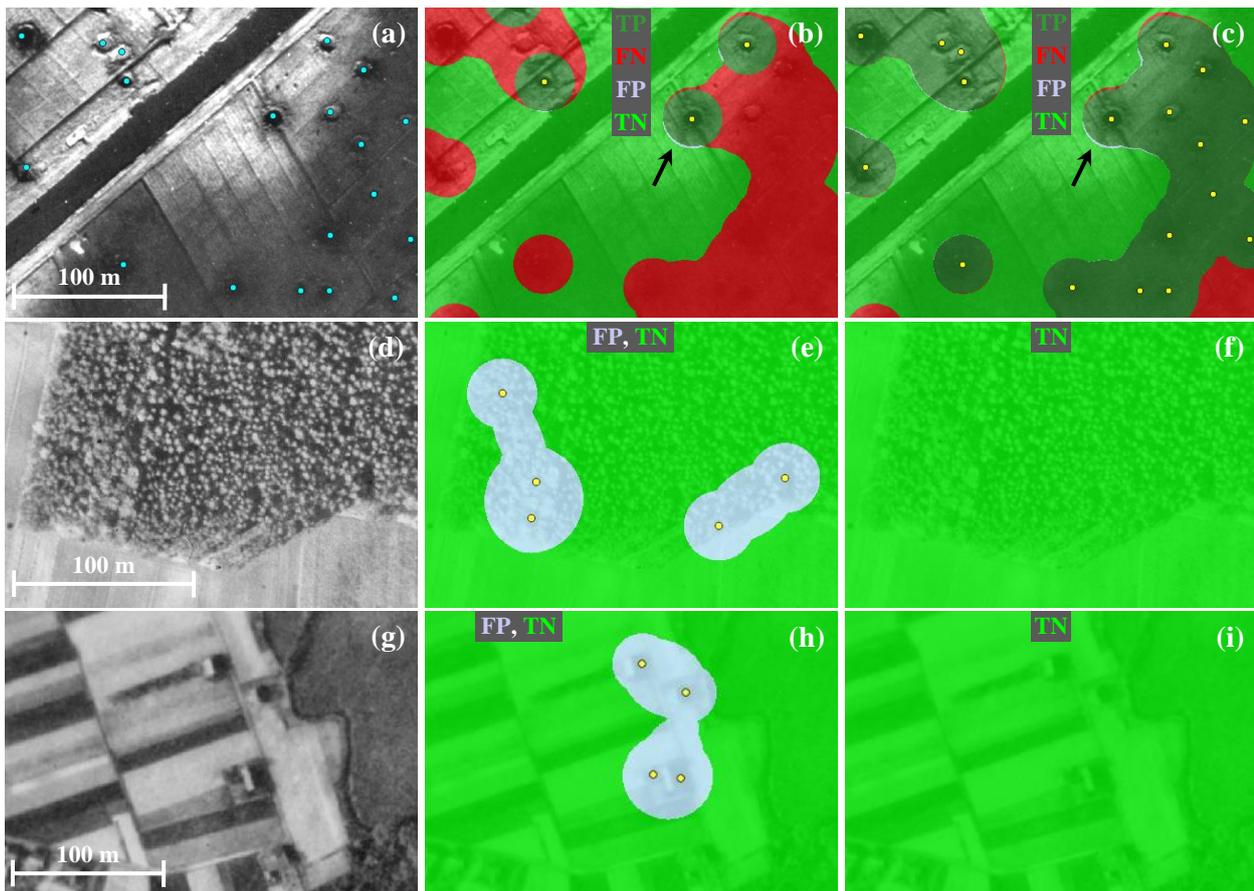
RG	DS	Image ID	NI	NC	SIA			MIA		
					R [%]	P [%]	F1 [%]	R [%]	P [%]	F1 [%]
1	A	1 <sub>LS2</sub>	22	443	23	80	35	72	87	79
		2 <sub>LS2</sub>	15	114	34	33	34	58	74	65
		3 <sub>LS2</sub>	20	238	55	79	65	76	87	81
		4 <sub>LS2</sub>	18	191	34	81	48	62	83	71
		5 <sub>LS2</sub>	22	424	71	93	80	80	85	82
		6 <sub>LS2</sub>	19	43	49	68	57	56	69	62
	B	7 <sub>LS2</sub>	19	220	12	82	21	61	87	71
		8 <sub>LS2</sub>	18	522	46	80	59	64	86	73
		9 <sub>LS2</sub>	16	129	40	87	55	59	97	74
2	A	10 <sub>LS2</sub>	16	137	17	27	21	61	58	59
		11 <sub>LS2</sub>	15	33	11	30	16	43	50	47
		12 <sub>LS2</sub>	17	17	48	54	51	78	41	54
		13 <sub>LS2</sub>	17	19	54	9	15	66	84	74
		14 <sub>LS2</sub>	17	26	77	34	47	35	64	45
		15 <sub>LS2</sub>	15	248	50	56	53	56	70	63
	B	16 <sub>LS2</sub>	23	50	53	69	60	62	84	72
		17 <sub>LS2</sub>	24	67	49	55	52	64	68	66
		18 <sub>LS2</sub>	15	213	65	90	76	81	90	85
3	A	19 <sub>LS2</sub>	8	62	12	12	12	20	27	23
		20 <sub>LS2</sub>	8	24	4	2	3	9	70	16
		21 <sub>LS2</sub>	8	52	8	6	7	24	48	32
		22 <sub>LS2</sub>	11	50	21	17	19	26	53	35
		23 <sub>LS2</sub>	8	11	69	20	31	21	43	28
		24 <sub>LS2</sub>	9	37	26	11	15	32	28	30
	B	25 <sub>LS2</sub>	14	24	20	26	23	28	47	35
		26 <sub>LS2</sub>	20	10	18	9	12	58	20	30
		27 <sub>LS2</sub>	20	63	45	29	35	59	41	48
<b>Average results DS A</b>					<b>41</b>	<b>53</b>	<b>46</b>	<b>64</b>	<b>76</b>	<b>69</b>
<b>Average results DS B</b>					<b>42</b>	<b>70</b>	<b>53</b>	<b>64</b>	<b>80</b>	<b>71</b>
<b>Average results DS A + DS B</b>					<b>41</b>	<b>59</b>	<b>48</b>	<b>64</b>	<b>77</b>	<b>70</b>
<b>Average results DS A + DS B masked</b>					<b>42</b>	<b>81</b>	<b>55</b>	<b>63</b>	<b>88</b>	<b>74</b>

generated impact map is strongly improved for both precision and recall (cf. numerical values in Table 6.4). In this context, however, investigations on the masked data reveal that, despite the use of redundant image information, there still is a considerable number of high objects leading to false detections; this can be inferred from the higher values in precision for the masked data. Consequently, a major benefit remains in integrating additional information.

Regarding the local approach that combines the individual detection results, it does not seem reasonable at first glance that point sets ( $PS$ ; Section 4.2) having fewer detections than

$MinNumDet_{PS}$  are also eliminated if the number of overlapping images for the respective  $PS$  is less than  $MinNumDet_{PS}$ . In principle, this parameter could be reduced or omitted in such a case, but this would contradict the basic idea that the detections should support each other (cf. objectives in Section 1.2), i.e. that several detections of the same object in different images are an indication for indeed having found a correct object. Consequently, this would lead to more correct detections, but the number of false ones would also increase.

As shown in the previous section, the results strongly depend on the images used. To reduce this impact, it would be feasible to conduct a visual selection of the images in advance to exclude those with an inappropriate appearance (e.g. due to blur, low contrast and uneven illumination) and severe cloud cover from the calculations. This concept is used by Lower Saxony's Explosive Ordnance Disposal Service for the manual analysis of aerial wartime images, e.g. when processing applications for building projects. Furthermore, the impact map could be improved by a more accurate co-registration of the overlapping images, though only marginally (cf. Appendix A).



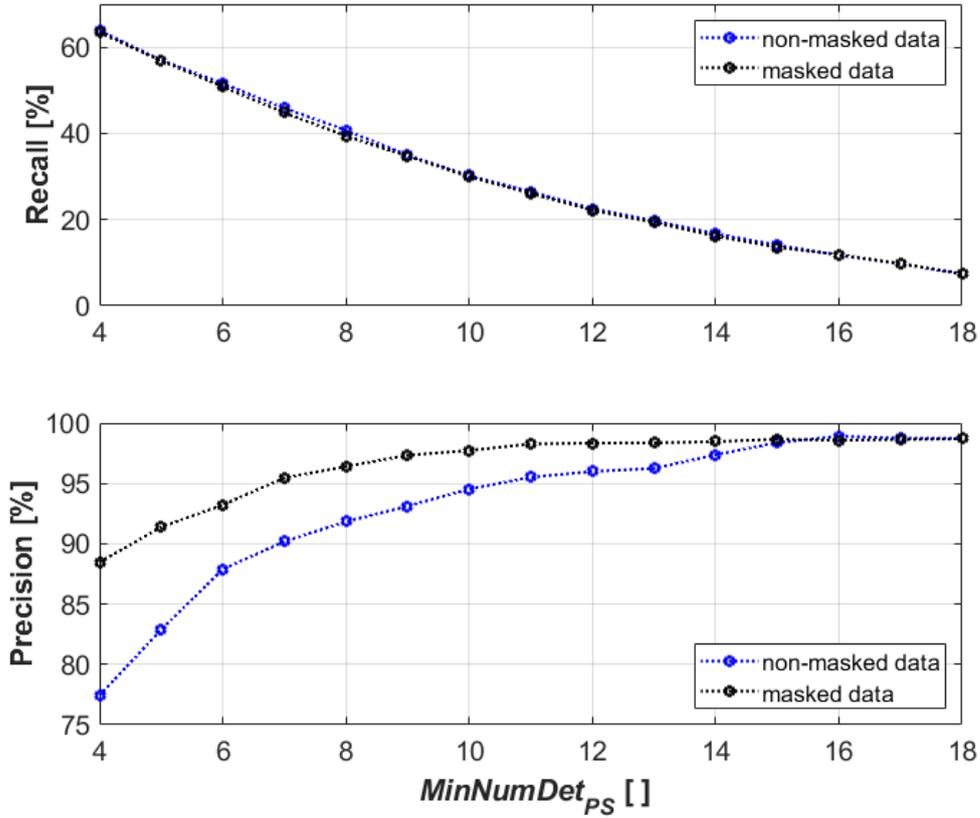
**Figure 6.7:** Subsets of images  $1_{LS2}$ ,  $13_{LS2}$  and  $21_{LS2}$  with reference centres of the bomb craters in turquoise (only the subset of  $1_{LS2}$  contains craters). Corresponding impact maps superimposed on the image subsets and their evaluation with TP areas in dark green, FN areas in red, FP areas in pale blue and TN areas in lime green for detection results based on (b, e, h) single and (c, f, i) multiple images.

### 6.2.3 Focus on precision

Given the proposed application scenario, the precision of the results is most important, i.e. the areas falsely classified as contaminated should be as small as possible. Using an optimum F1-score, the quality measures reported so far are too low to directly integrate the developed procedure into the workflow of the Explosive Ordnance Disposal Service, also when using redundant imagery. With a precision of 77 %, too many areas would still be unnecessarily probed, resulting in enormous costs. Thus, in order to achieve a higher precision at the cost of a lower recall, the parameters of the algorithm are varied. In the context of the MPP-based bomb crater detection on single images (Section 6.2.1), it is for instance possible to vary the parameter  $c$  (Equation 4.2) of the first data term  $U_G$ . Increasing  $c$  will result in more objects with smaller gradients at the object border being removed from the object configuration. In a similar way, the parameters of the second data term  $U_H$  (Equation 4.3) or the third data term  $U_B$  (Equation 4.5) could be adapted, but these terms have more free parameters. As there are usually multiple images covering the investigated area, the thesis concentrates on the multiple image case, i.e. only results related to such data are shown and analysed. In this context, to achieve a higher precision, the parameter  $MinNumDet_{PS}$  is varied; increasing  $MinNumDet_{PS}$  leads to more  $PS$  with a smaller number of detections being removed. Hence,  $MinNumDet_{PS}$  is increased starting with  $MinNumDet_{PS} = 4$ . Again, the quality measures are based on the area-weighted averages of the respective images and the masked data is also considered within the experiments.

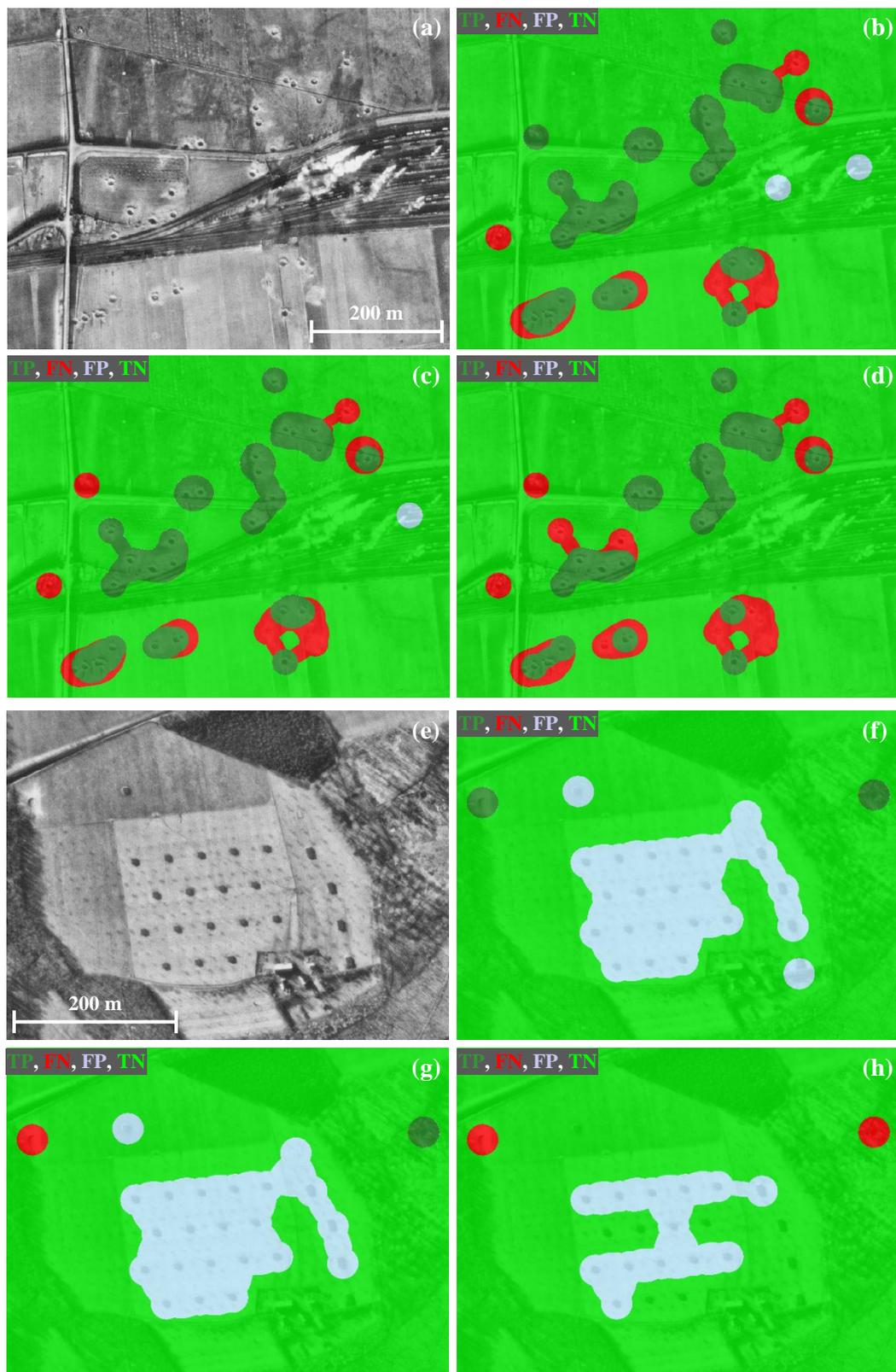
Figure 6.8 shows the dependence of recall and precision on  $MinNumDet_{PS}$  for the non-masked and masked images of DS A + DS B. It can be seen that for cases, the recall decreases more or less linearly until about  $MinNumDet_{PS} = 9$ , and subsequently the decrease becomes slightly smaller. At the same time, there is an also linear increase in precision until  $MinNumDet_{PS} = 6$  (for the masked data until  $MinNumDet_{PS} = 7$ ) that becomes smaller afterwards, especially for the masked images. Furthermore, it should be noted that, as to be expected, both curves for recall overlap. Moreover, the curves for precision converge with increasing  $MinNumDet_{PS}$ , showing that additional information is particularly meaningful for smaller values of  $MinNumDet_{PS}$ . In the non-masked case, from approx.  $MinNumDet_{PS} = 6$  onward, the loss in recall is higher than the gain in precision; for the masked data this holds true for  $MinNumDet_{PS} = 4$ . However, for the non-masked data, if  $MinNumDet_{PS}$  is set to 7, a precision of 90 % and a recall of approx. 45 % can still be achieved; for the masked data, the precision is increased by 5 %. For the non-masked images, the precision of about 95 % results for  $MinNumDet_{PS} = 10$ , where the remaining recall is 30 %. From about  $MinNumDet_{PS} = 16$  onwards, in both cases, the precision is approx. 99 %, whereas the recall still decreases.

There are different reasons why some falsely classified areas may remain despite the combination of the detection results. Substantial in this context are objects that very much follow the proposed MPP model, i.e. round objects with high gradients at the border, homogeneous grey values



**Figure 6.8:** Recall and precision as a function of  $MinNumDet_{PS}$  based on the area-weighted average of DS A + DS B for the non-masked and masked images.

inside and a high contrast with their annulus. This applies for example to round shadows on a bright background or also industrial chimneys (dark interior with a bright border, the brickwork, around it). Consequently, such objects will be preserved longer than bomb craters that show the above mentioned properties to a lesser extent. The discrepancies between reference and detection centres can also be mentioned here (see discussion in Section 6.2.1). Two examples, one in which higher values for  $MinNumDet_{PS}$  appropriately eliminate false detections and one in which this does not work, are illustrated in Figure 6.9, showing a subset of the image 17<sub>LS2</sub>. The derived impact maps from the detection results for  $MinNumDet_{PS} = 4$ ,  $MinNumDet_{PS} = 6$  and  $MinNumDet_{PS} = 8$  are given in Figure 6.9 b, c and d, respectively. It can be seen that by increasing  $MinNumDet_{PS}$  from  $MinNumDet_{PS} = 4$  (value that leads to the best F1-score for all the data considered) to  $MinNumDet_{PS} = 8$ , the areas that are wrongly classified as contaminated vanish (Figure 6.9 b-d, pale blue). However, the number of pixels falsely not classified as contaminated increases as well (Figure 6.9 b-d, red). Consequently, for the proposed application scenario, the procedure provides very good results for  $MinNumDet_{PS} = 8$  in this example: all areas classified as contaminated (Figure 6.9 d, dark green) actually have to be probed. This is different for the results related to the Figure 6.9 e-h, where, similarly to Figure 6.9 a-d, Figure 6.9 e shows a subset



**Figure 6.9:** (a, e) Subsets of images  $17_{LS2}$  and  $27_{LS2}$ , respectively. Corresponding impact maps superimposed on the image subsets and their evaluation with TP areas in dark green, FN areas in red, FP areas in pale blue and TN areas in lime green for detection results resulting from (b, f)  $MinNumDet_{PS} = 4$ , (c, g)  $MinNumDet_{PS} = 6$  and (d, h)  $MinNumDet_{PS} = 8$ , respectively.

of image 27<sub>LS2</sub> and Figure 6.9 f-h shows the corresponding impact maps superimposed on the subset, again for  $MinNumDet_{PS} = 4$ ,  $MinNumDet_{PS} = 6$  and  $MinNumDet_{PS} = 8$ , respectively. In this example, a larger number of false detections and, thus, areas falsely classified as contaminated, remain longer than the two areas at first correctly classified as contaminated (Figure 6.9 f-h). This can be explained by the fact that, as just noted, in all the images considered, the false positive objects follow the MPP model to a greater extent than the two bomb craters. Indeed, the crater on the left in the master image shown (Figure 6.9 e) exhibits a crescent-shaped shadow and the crater on the right in the same image is difficult to identify in-between the trees, whereas the roundish shadow casts in the centre of the image (presumably caused by hay bales and small huts) are striking.

In summary, the results are useful with respect to the application scenario as a rather large percentage of the areas that actually need to be probed can be detected with a high degree of precision. In this context, experts of Lower Saxony's Explosive Ordnance Disposal Service have not made any concrete statements regarding a "suitable" value for precision – values of at least 90 % seem to be desirable. Anyway, the procedure can be adapted with respect to the achieved precision by means of just one parameter, whereby the higher the values for  $MinNumDet_{PS}$ , the more strongly the recall decreases in comparison to the increase in precision. Here, the fact of convergence of the curves of the masked and non-masked data for higher values of  $MinNumDet_{PS}$  should be pointed out again. This is of interest if it is not possible to automatically derive useful height information and no other information concerning objects with a certain height is available, as higher values for  $MinNumDet_{PS}$  would then be comparatively more preferable to lower values (cf. Figure 6.8). On the other hand, it has been shown that additional information is particularly useful for smaller values of  $MinNumDet_{PS}$ .

Thus, especially in the case where additional information is available, it would be conceivable to select smaller values for  $MinNumDet_{PS}$ , e.g.  $MinNumDet_{PS} = 5$ , which leads to a precision of more than 90 % and a recall of almost 60 %. Subsequently, the automatically generated impact map, being already correct to a large extent, could be manually checked and improved by human experts before usage. Hence, e.g. in the case of construction projects, the information regarding a contamination could be used such that images do not have to be inspected manually any longer in the areas flagged as to be probed, saving costs and resources.

### 6.3 Comparison to a state-of-the-art object detector based on CNNs

In this section, the results of the comparison of the MPP method for bomb crater detection with the Faster R-CNN object detector are reported. In this context, in addition to the average numerical values of recall, precision and the F1-score for the object-based evaluation for the MPP and the

CNN approach, the respective object-based means and standard deviations (each experiment is carried out three times) are computed. Examples allowing for a qualitative comparison of the results are also given.

The quantitative results, sorted by the number of craters, can be found in Table 6.5, whereas the examples are illustrated in Figure 6.10. From Table 6.5, it can be seen that if the threshold  $p_t$  for the CNN (Section 5.5) is set to 0.5 ( $CNN_{p_t=0.5}$ ), the results are the worst, with a mean F1-score of 47.3 %. Although the mean recall of 72.2 % is the highest, there are many false detections, resulting in the lowest mean precision of 35.3 %. Thus, the choice of  $p_t = 0.5$ , i.e. the natural choice for a binary classifier based on selecting the class having the highest class score, proves not to be most suitable regarding an optimal F1-score. However, if the value of  $p_t$  is optimized during training ( $CNN_{p_t=0.9}$ ), the highest mean F1-score is achieved (63.8 %); with 61.2 % the one based on the MPP is worse by a margin of less than 3 %. Moreover, the MPP approach yields a higher mean precision than  $CNN_{p_t=0.9}$  (71.7 % compared to 64.2 %), though, obviously, the recall is lower (53.4 % compared to 63.6 %).

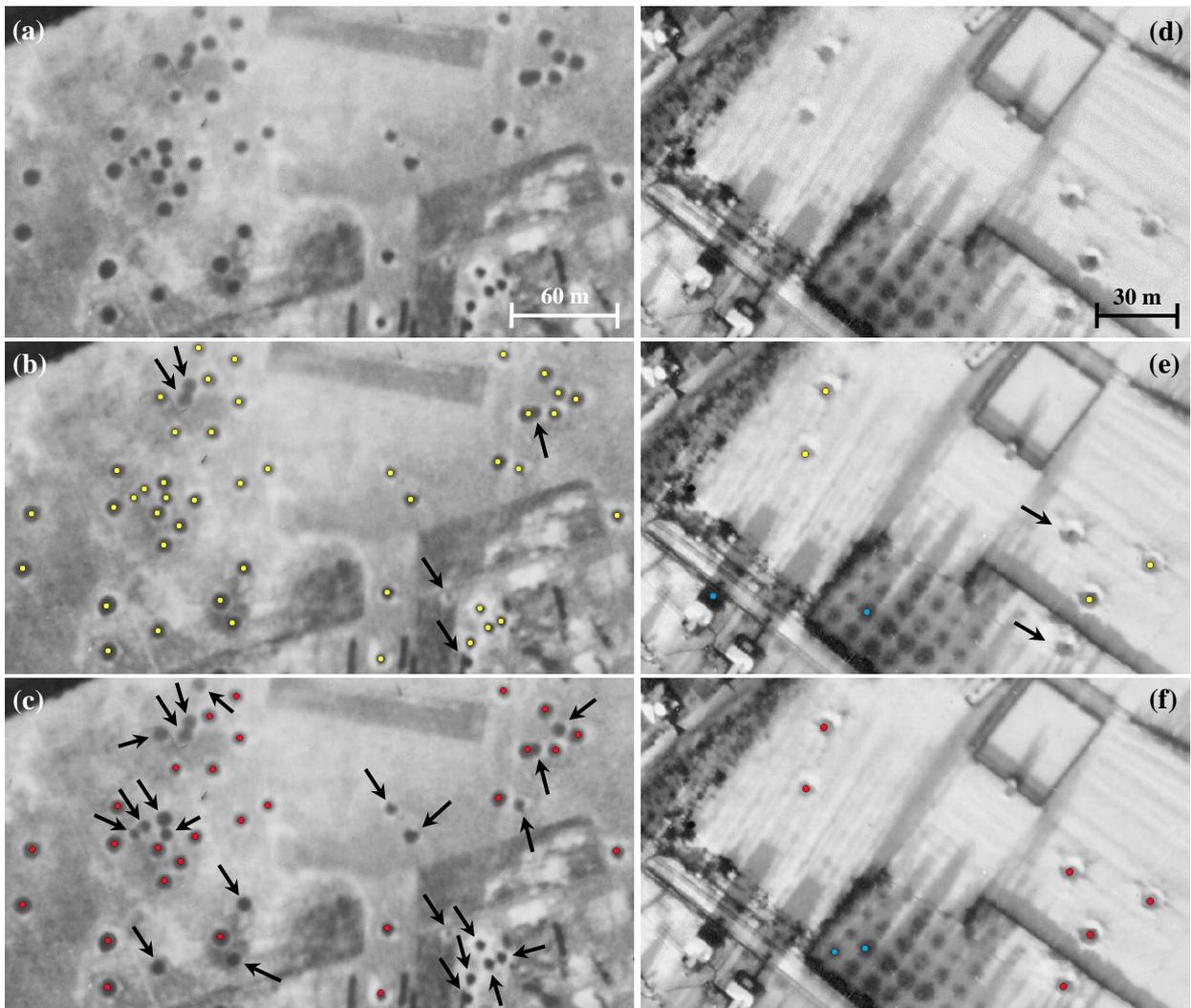
It can be seen from Table 6.5 that compared to  $CNN_{p_t=0.9}$ , the MPP only gives better results – in terms of the F1-score – for two images, namely 16<sub>LS2</sub> and 22<sub>LS1</sub>. In the latter, where the difference with respect to the F1-scores is higher, a larger number of craters follow the model used in the MPP procedure to a very high extent. The CNN-based approach also detects many of these craters, but not to the same extent (Figure 6.10 a-c). Moreover, in image 22<sub>LS1</sub> there are some craters with a quite small diameter, which are not detected by  $CNN_{p_t=0.9}$ , though some of them are detected by the MPP. The overall rather low value of recall for the MPP related to image 22<sub>LS1</sub> can be explained by additional bright craters in the image and some regions with low contrast. As already discussed in Section 4.4, the proposed MPP procedure cannot detect such bright craters, whereas the CNN-based approach can. This is one reason why the mean recall for  $CNN_{p_t=0.9}$  is considerably higher than the one for the MPP. Furthermore, in addition to the low-contrast craters already mentioned, the craters that deviate too much from the MPP model also lead to lower values in recall. For the other six images, the F1-scores for  $CNN_{p_t=0.9}$  are higher than those based on the MPP method, in some cases by a large margin. In this context, for instance, the F1-score based on image 16<sub>LS1</sub> is more than 20 % lower (approx. 36 % compared to 59 %). Within this image, there is a relatively large number of craters that appear bright and also craters that follow the MPP model only to a limited extent. Consequently, such craters are not detected at all or detected less frequently by the MPP, resulting in a quite low recall of approx. 26 %. In contrast, the precision of about 61 % is slightly higher compared to the one based on  $CNN_{p_t=0.9}$ ; for the MPP, false detections can be partly attributed to shadow cast from houses (Figure 6.10 d-f). Such false detections and others caused by shadows of, e.g., trees, trains or buildings without roofs, usually occur less frequently with the CNN-based approach. This is the case because it learned, at least partly, the context in the local surroundings from the training data and is thus able to differentiate between crater and background in a better way than the MPP.

**Table 6.5:** Evaluation results, sorted by the number of craters (NC), for the MPP and the CNN approach based on 8 images (IMG) from Lower Saxony; each experiment is carried out three times. Besides the average values of recall  $R$ , precision  $P$  and the F1-score  $F1$  for each image, the respective means and standard deviations over all images are given.

IMG	NC	MPP			CNN					
					threshold $p_t = 0.5$			threshold $p_t = 0.9$		
		R [%]	P [%]	F1 [%]	R [%]	P [%]	F1 [%]	R [%]	P [%]	F1 [%]
23 <sub>LS2</sub>	11	78.8	23.0	35.6	93.9	8.7	15.9	90.9	24.5	<b>38.5</b>
12 <sub>LS1</sub>	37	46.9	74.3	57.5	89.2	30.1	44.8	84.7	54.7	<b>66.4</b>
16 <sub>LS2</sub>	50	52.7	68.7	<b>59.6</b>	80.0	16.0	26.7	70.7	45.0	54.8
14 <sub>LS1</sub>	57	44.4	46.9	45.6	80.7	17.9	29.3	75.4	44.7	<b>56.1</b>
15 <sub>LS1</sub>	114	24.3	31.2	27.3	49.7	17.8	26.1	38.9	36.6	<b>37.4</b>
16 <sub>LS1</sub>	132	25.5	60.8	35.9	73.6	27.5	39.9	62.4	56.2	<b>59.0</b>
22 <sub>LS1</sub>	301	53.9	72.9	<b>62.0</b>	49.1	41.3	44.6	38.0	66.5	48.1
5 <sub>LS2</sub>	424	70.7	94.0	80.7	90.0	72.8	80.4	84.0	88.9	<b>86.4</b>
<b>Object-based mean</b>		53.4 ± 0.1	71.7 ± 0.2	61.2 ± 0.1	72.2 ± 2.9	35.3 ± 1.6	47.3 ± 0.8	63.6 ± 2.6	64.2 ± 1.8	<b>63.8 ± 0.6</b>

In terms of performance stability, it can be seen in Table 6.5 that the standard deviations for the CNN-based approach are higher than those of the MPP for all three quality measures. As with the experiments on the influence of random numbers on MPP results in Section 6.1.1, again, the final results vary only slightly around the mean for the F1-score (here  $\pm 0.1$  %). This value is six ( $CNN_{p_t=0.9}$ ) and eight ( $CNN_{p_t=0.5}$ ) times higher for the CNN-based variants, respectively. The differences in recall and precision are even more distinct. While the performance of the MPP-based approach is again stable (recall  $\pm 0.1$  %, precision  $\pm 0.2$  %), the values for  $CNN_{p_t=0.9}$  vary considerably more with  $\pm 2.6$  % for the recall and  $\pm 1.8$  % for the precision; this is similar for  $CNN_{p_t=0.5}$ . This finding may be an indication that the neural network has not yet fully learned the high appearance variation of both, bomb craters and disturbing objects. In this context, besides further training data, to circumvent the need for large amounts of manually labelled training data, semi-supervised and self-supervised methods, e.g. (Zhai et al., 2019), could be employed.

Even though the results based on  $CNN_{p_t=0.9}$  are better than those of the MPP, there are some important advantages of the model-based approach. This solution does not require any training data at all, thus avoiding the time-consuming generation of training samples. It has to be noted that the CNN-based method is only better than the MPP-based approach if the threshold for selecting the class label is included in the learning procedure, which further increases the number of labelled samples required for training; if the standard approach for a classifier, i.e., the selection of the most likely class label, is followed, the results of the CNN are considerably worse than those of the MPP. Furthermore, the high intra-class variability of bomb craters likely implies another increase



**Figure 6.10:** (a, d) Subsets of images  $22_{LS1}$  and  $16_{LS1}$ , respectively. (b, e) Detection results of the MPP approach and (c, f) detection results of the CNN approach with  $p_t = 0.9$ . The centres of TPs are shown in yellow (MPP) and red (CNN), respectively. Centres of FPs are shown in cyan (both in e and f there are two FPs, b and c do not contain any). FNs are tagged by black arrows in b, c and e; there are none in f.

in the training data needed, which can be conjectured from the experiments regarding the performance stability. Nevertheless, if the just mentioned limitations concerning a sufficient amount of representative training data are not present, i.e. in the long run, a CNN-based approach is considered to be more suitable, also because recent literature shows that such approaches then usually outperform model-based ones.

## 6.4 Object detection for counting purposes in images

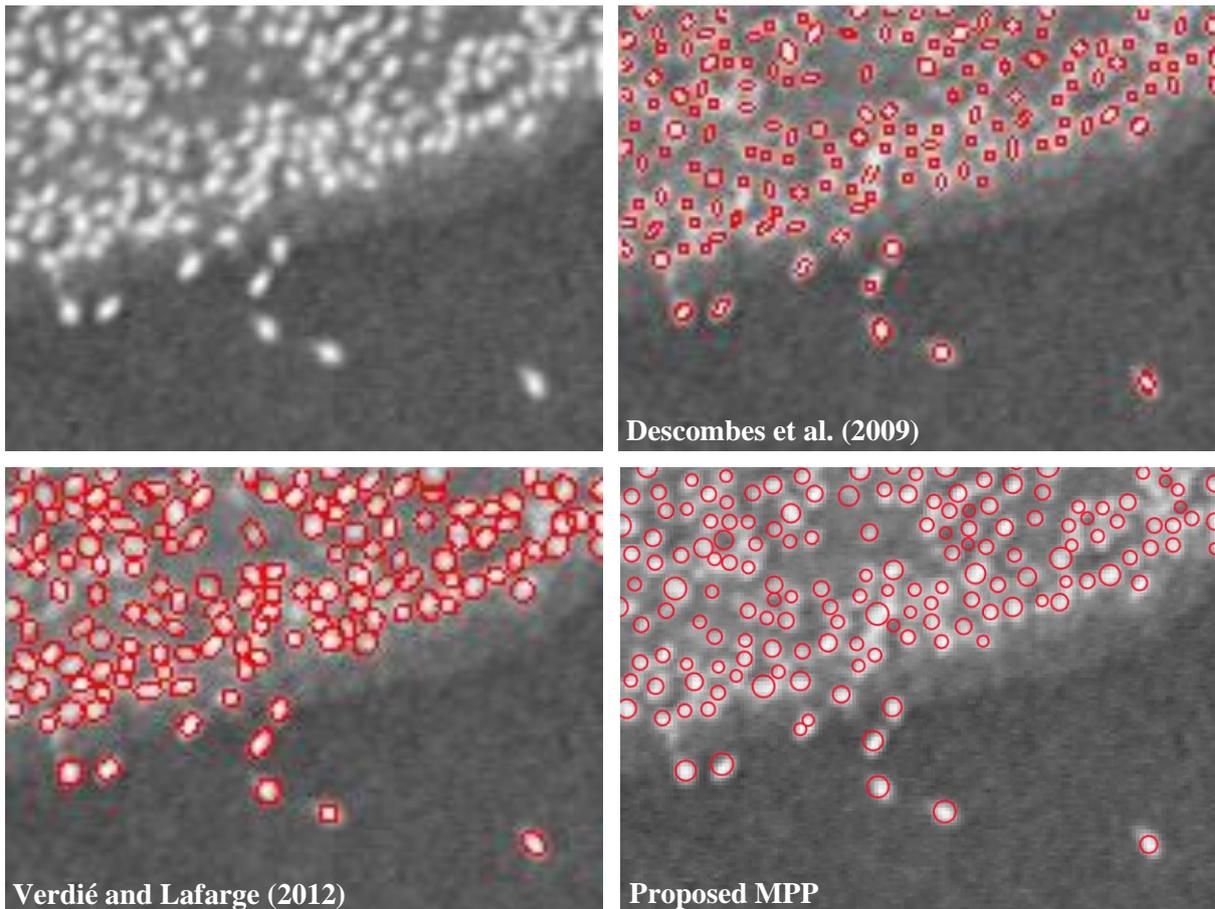
In this section, the transferability of the proposed MPP approach designed for bomb crater detection (Section 4.1) to the task of population counting from images is examined. For these experiments, panchromatic microscopic images containing cells or stomata, and a panchromatic aerial image showing flamingos are considered. For the three scenes *flamingos*, *cells* and *stomata* (Figure 5.2), using the given reference information, the numerical values of recall, precision and the F1-score for the object-based evaluation are computed (Section 5.4) and compared with the corresponding existing quality measures from the literature (which were calculated in the same way). Moreover, a qualitative comparison is made.

The quantitative results of the proposed MPP and the methods of Descombes et al. (2009) as well as Verdié and Lafarge (2012) for the scenes *flamingos*, *cells* and *stomata* can be found in Table 6.6. Both approaches are also based on MPPs but use an ellipse as object model. Their energy is specified by only one data energy term that is based on the Bhattacharyya distance, and, similar to the proposed approach, a strong overlap of objects is penalized within the prior energy. A visual comparison of the results for the three scenes is possible based on the Figures 6.11, 6.12 and 6.13, respectively. Note that for the method proposed in this thesis, for the scenes *flamingos* and *cells*, the calculations were performed on the complementary image, as in the original images the objects appear brighter than their surroundings; the original image is shown only for the better comparison of the individual detection results.

Having a look to Figure 6.11, it can be seen that the proposed MPP detects almost all birds in the image despite the low image quality and the partial overlap of the flamingos (Figure 6.11 bottom right). However, the borders of the detected objects do not always represent the actual object borders, especially in the case of flamingos that appear more elliptical in the image. This can be attributed to the use of a circle as an object model in the proposed MPP. Nevertheless, experiments have shown that the quality measures are very similar when an ellipse is used instead. Comparing the qualitative detection results with the ones of the methods from Descombes et al. (2009) and Verdié and Lafarge (2012), illustrated in Figure 6.11 top right and Figure 6.11 bottom left, no major differences are visible. One aspect with respect to the results of the proposed MPP is its comparatively higher number of false detections. Furthermore, as already mentioned, in the other methods the borders of the detected objects represent the actual object borders in a better way, especially in the case of Verdié and Lafarge (2012). However, on the other hand, a comparatively large number of birds is not detected, for example, due to the fact that only one of two overlapping flamingos is found. These findings are also reflected in the quantitative detection results (see Table 6.6), which are basically on a similar high level. In this context, the procedure of Descombes et al. (2009) performs comparatively best with a F1-score of 96.6 %, followed by 95.1 % for the proposed MPP and 93.4 % for the method of Verdié and Lafarge (2012). Moreover, the proposed MPP achieves the highest recall (98.0 %), though the precision is the lowest with 92.4 %.

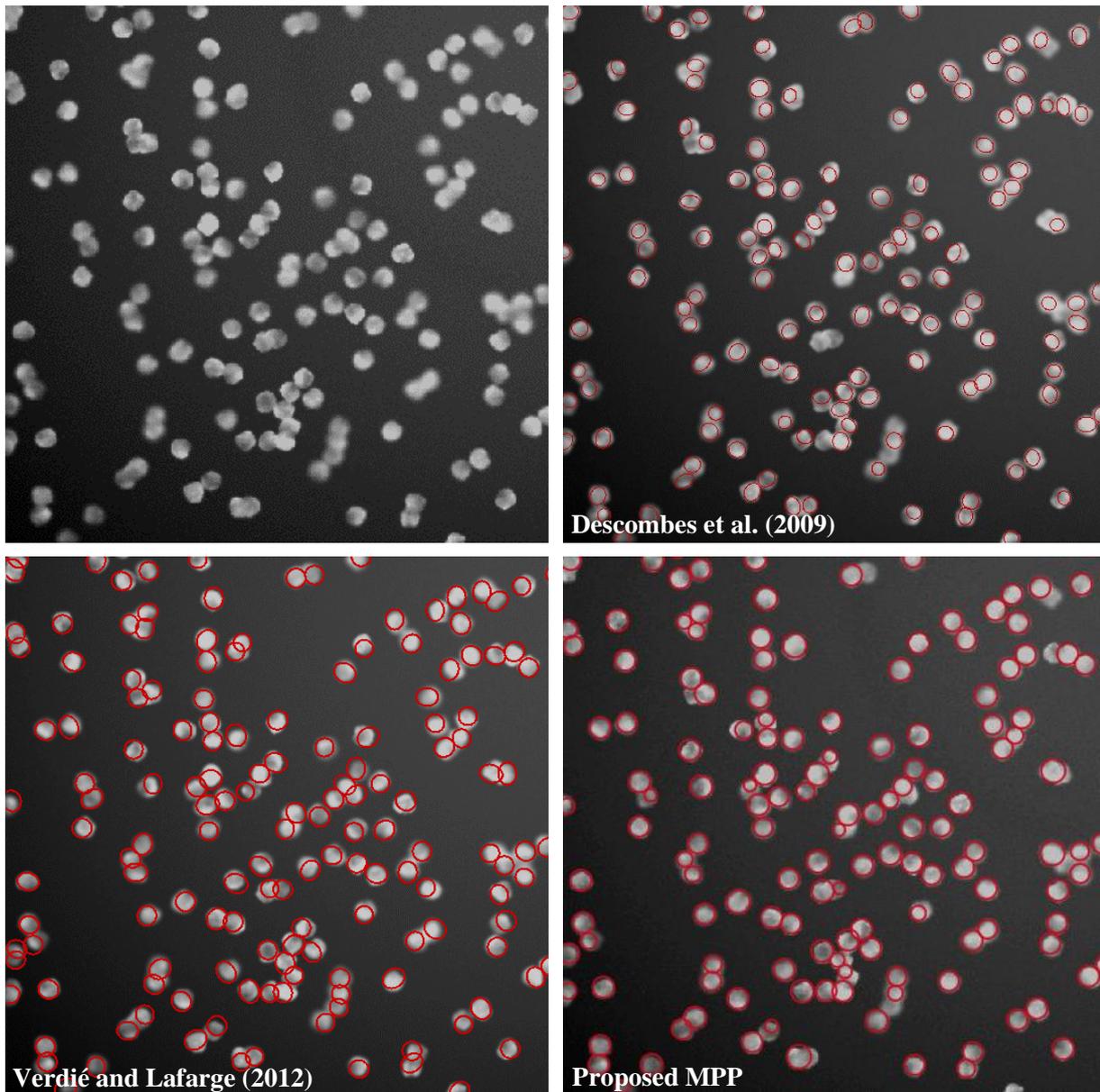
**Table 6.6:** Quantitative comparison of the proposed MPP with the methods of Descombes et al. (2009) and Verdié and Lafarge (2012) for the scenes *flamingos*, *cells* and *stomata* based on the quality measures recall  $R$ , precision  $P$  and the F1-score  $F1$ . Furthermore, the respective numbers of TPs  $\#TP$ , FPs  $\#FP$  and FNs  $\#FN$  are given. Note that for the scene *stomata*, only results for the approach of Verdié and Lafarge (2012) are available.

Scene	Method	#TP	#FP	#FN	P [%]	R [%]	F1 [%]
<i>Flamingos</i>	Descombes et al. (2009)	143	5	5	96.6	96.6	<b>96.6</b>
	Verdié and Lafarge (2012)	133	4	15	<b>97.1</b>	89.9	93.4
	Proposed MPP	145	12	3	92.4	<b>98.0</b>	95.1
<i>Cells</i>	Descombes et al. (2009)	447	0	53	<b>100</b>	89.4	94.4
	Verdié and Lafarge (2012)	480	2	20	99.6	<b>96.0</b>	<b>97.8</b>
	Proposed MPP	450	1	50	99.8	90.0	94.6
<i>Stomata</i>	Descombes et al. (2009)	-	-	-	-	-	-
	Verdié and Lafarge (2012)	560	156	116	78.2	82.4	80.2
	Proposed MPP	619	112	57	<b>84.7</b>	<b>91.6</b>	<b>88.0</b>

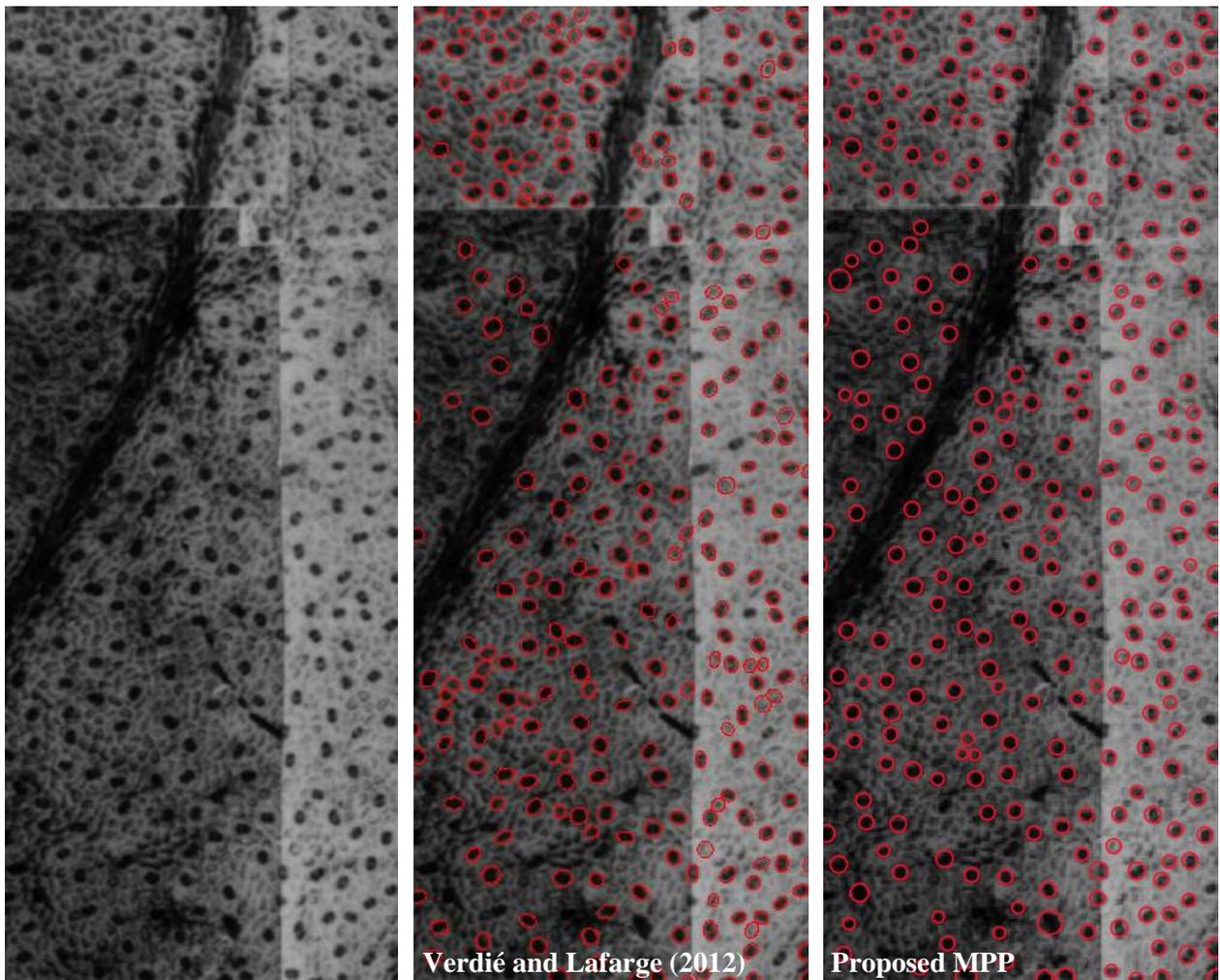


**Figure 6.11:** Comparison of the results of the proposed MPP with the results of the MPP-based methods of Descombes et al. (2009) and Verdié and Lafarge (2012) for the scene *flamingos*; the boundaries of the detected objects are shown in red.

The quality measures for the scene *cells* are also in a similarly high range (see Table 6.6). With a F1-score of 97.8 %, the method of Verdié and Lafarge (2012) provides the best results; the F1-scores for the proposed method and that of Descombes et al. (2009) are about 3 % lower, specifically 94.6 % and 94.4 %. The number of false detections is very low in all cases ( $< 3$ ), although comparatively many cells are not detected, which can be attributed to the sometimes strong overlap of the objects and thus the more difficult detection. More precisely, in the case of Verdié and Lafarge (2012), the proposed MPP and Descombes et al. (2009), there are 20, 50 and 53 non-detected cells, respectively. This can also be seen in Figure 6.12. The subset of the scene *cells* with



**Figure 6.12:** Comparison of the results of the proposed MPP with the results of the MPP-based methods of Descombes et al. (2009) and Verdié and Lafarge (2012) for a subset of the scene *cells*; the boundaries of the detected objects are shown in red.



**Figure 6.13:** Comparison of the results of the proposed MPP with the results of the MPP-based method of Verdié and Lafarge (2012) for a subset of the scene *stomata*; the boundaries of the detected objects are shown in red.

the superimposed detection results shows that the method of Verdié and Lafarge (2012) detects the highest number of cells, whereas this is lower in the other two cases. Unlike the flamingos, the cells can be well approximated by circles. Indeed, for the proposed MPP, most of the detected object borders represent the actual shape of the cells (Figure 6.12 bottom right), which is less the case for the other two results, especially for those of Descombes et al. (2009; Figure 6.12 top right).

According to the results of the experiments analysed so far, it can be concluded that similar quality measures can be obtained with the three methods for the scenes *flamingos* and *cells*. This is not the case for the scene *stomata*, where the difference in F1-score for the proposed MPP and the method of Verdié and Lafarge (2012) is larger; the respective F1-scores are 88.0 % and 80.2 %, i.e. the increase in F1-score is about 8 % (see Table 6.6). In this context, with the procedure of Verdié and Lafarge (2012), the number of FNs is about twice as high (116 vs. 57) and the number

of FPs is also increased (156 vs. 112). A possible reason for this result is the fact that in the proposed MPP Contrast Limited Adaptive Histogram Equalization (CLAHE, Section 4.1.5) is carried out in advance, which should improve the detection in low-contrast areas (e.g. bottom left in Figure 6.13 left) and reduce over-detection in already high-contrast areas. It can also be assumed that the additional data terms in the proposed MPP support the detection quality, because the stomata have high gradient magnitudes at the object border and homogeneous grey values in the interior. The better quality measures for the proposed MPP (Figure 6.13 right) compared to the method of Verdié and Lafarge (2012; Figure 6.13 middle) can also be confirmed visually (note that Figure 6.13 right shows the original image and not the image after applying CLAHE). For instance, the method of Verdié and Lafarge (2012) incorrectly detects comparatively more of the objects that are slightly smaller compared to stomata, are not as dark and usually exhibit a less homogeneous inside as well (e.g. top left in Figure 6.13 middle). In addition, it can be seen that in total fewer stomata are detected.

Overall, the MPP proposed in this thesis, which has been originally designed for the detection of bomb craters in aerial wartime images, can be adapted and used very well for the objective of population counting from images. Furthermore, compared to other work dealing with this task, similar or even moderately better results can be achieved for different scenes.

---

## 7 Conclusions and Outlook

In this final chapter, conclusions are drawn with respect to the presented approach including the experimental results, and remaining open issues are identified, providing an outlook on promising future research topics.

Addressing the task of bomb crater detection in aerial wartime images and the subsequent generation of an impact map using the detected objects, in this thesis, a novel stochastic approach is proposed. It is based on the method of marked point processes (MPPs), in which object configurations are iteratively created during the optimization process, each of which is evaluated with an energy function that describes the consistency with the predefined model; the bomb craters are modelled as circles. The detections are used to generate an impact map that provides a quick overview of contaminated areas to detect sectors that have a high likelihood of containing a dud, and therefore makes it necessary to have them probed on the ground. The approach is evaluated on a total of 74 panchromatic images, for which experts of Lower Saxony's Explosive Ordnance Disposal Service generated reference information by manual annotation.

In connection with the single image approach, the experiments show that the results based on the impact maps for the 55 coarsely georeferenced images from Lower Saxony yield a recall and a precision of about 48 % and 64 %, respectively. The impact maps derived from the images of Italy and particularly of Salzburg are less reliable due to the comparatively more complex image content. When considering additional information, which has been simulated by manually masking high static objects including their shadows, the average precision increases strongly, e.g. for the images from Lower Saxony from 64 % to 83 %, showing the overall benefit and reveals that, indeed, the just mentioned objects are responsible for a large number of false detections.

Comparing the MPP method to a state-of-the-art convolutional neural network (CNN) approach, it is shown on 8 images from Lower Saxony that the performance of the CNN critically depends on selecting the threshold  $p_t$  for a region to be considered as a crater. Particularly, the natural choice of  $p_t = 0.5$  leads to considerably worse results compared to the MPP (mean F1-scores: 47.3 % vs. 61.2 %). However, when tuning  $p_t$  using labelled validation data, the CNN outperforms the MPP approach for most images with respect to the F1-scores, though the difference in the mean F1-scores is less than 3 %. It can be concluded that the CNN has the potential to outperform the MPPs in a scenario where a sufficient amount of labelled training data is available. However, in a scenario where labelled training data is sparse or non-existent, the MPP delivers better results, which is seen as an important advantage of the model-based approach.

Typically, multiple images of the same area exist. To make use of this, an approach that combines the individual detection results of the MPP procedure is proposed. In this context, based on a set of 27 images from Lower Saxony, it is shown that the F1-score of the pixel-based evaluation is increased considerably, namely from 48 % to 70 %. Despite the fact that errors can be compensated to a great extent by using redundant image information, there still remains a considerable, albeit smaller, number of high objects or their shadows leading to false detections compared to the single image approach, which can be inferred from the higher average precision of 88 % for the masked data compared to 77 % for the non-masked data.

Even if redundant image information is considered, however, the results are not good enough for a direct integration of the procedure into the workflow of the Explosive Ordnance Disposal Service if the aim is to have an optimal balance between precision and recall. Too many areas would have to be probed unnecessarily, resulting in high costs. In this context, the experiments show that the precision can be increased at the expense of the recall by varying only one parameter within the approach, rendering the procedure attractive for the proposed use case. Based on the 27 images from Lower Saxony, for instance, a precision of 90 % with a recall of approx. 45 % can be achieved, showing the benefit for the discussed application scenario. Again, additional information increases the precision, especially for lower values of the parameter mentioned before. In this context, for a precision of 90 %, a significantly higher recall of approx. 60 % is achieved. Thus, the method proposed in this thesis allows in a suitable way to automatically flag probing areas, as, e.g. in the case of construction projects, the information regarding the contamination can be used, and consequently images of these areas (for precision values close to 100 %) no longer have to be inspected manually, saving cost and resources.

The proposed MPP can be adapted to different scenes and for different applications by parameter variation. In this work, with the objective of population counting from images, its transferability to panchromatic microscopic images for the detection of cells or stomata, and to a panchromatic aerial image for the detection of flamingos, is demonstrated. The results of the experiments show that, compared to other work from the literature dealing with the same task, similar or even better results can be achieved.

Although the procedure provides satisfying results, there is a number of limitations that offer promising starting points for possible future work: A problem in connection with false detections arises often from objects with a certain height (e.g. forests, trees and buildings) and especially their shadows, as they may appear similar to bomb craters in the image. As shown by simulations, one way to counteract these limitations is to integrate 3D information to find and subsequently exclude such areas from further processing. For the estimation of 3D information, stereo matching could be applied in the frequent case in which stereo information is available, though the poor quality of many aerial wartime images is a limiting factor. Indeed, large radiometric differences, different sensors and scenes as well as heterogeneous acquisition conditions, pose a considerable challenge in terms of finding dense and robust feature correspondences across such data. In this context, recently, Zhang et al. (2021) proposed a promising fully automatic approach to recover the 3D

land-cover information solely based on historical images. In their approach, first of all, digital surface models (DSMs) are derived within individual epochs (in the thesis case, an epoch contains all images of a particular surveillance flight) by means of a standard structure-from-motion pipeline. Afterwards, the DSMs are incorporated in a so-called rough-to-precise matching, basically consisting of a DSM matching between different epochs to roughly co-register the orientations and DSMs (in this step the authors exploit the fact that the 3D landscape usually stays globally constant over time), which is followed by a precise feature matching using the original RGB images, enabled by narrowing down the search space using the co-registered data. Nevertheless, it remains to be examined whether this approach can also be applied to panchromatic aerial wartime images.

Moreover, to reduce false detections without additional information, the surrounding area in the vicinity of a detection could be analysed beyond the Bhattacharyya distance. Within an image, the shadow of objects is always on the same side, which means, for example, that in connection with shadows cast by houses, a bright object (roof) is always found on the same side of the shadow (see e.g. Figure 6.7 g); for bomb craters, the typically differently appearing bright area is then on the other side due to their depth. However, it should be noted that it is of course not known which detections correspond to a crater or another object, so 3D information would also help here. Finally, another more general idea to increase the precision of the impact map would be to classify an area as contaminated only after several detections (and not already for individual ones). For this purpose, the threshold  $p$  (Section 4.3) applied to the probability map created with kernel density estimation could be adapted. This, obviously, reduces the recall of the results as well (except in strongly bombed areas).

Furthermore, an extension of the energy function of the MPP is conceivable. In this context, to improve the detection of bomb craters, an energy term could be considered that takes into account their characteristic surface shape, i.e. that craters are usually deepest in their centre and become shallower towards the rim. Craters fulfilling this characteristic would lead to a decrease of the energy. Of course, such an extension requires quite detailed height information, as craters, while occasionally up to 5 m deep, are usually more shallow (about 1 m to 3 m), and it must also be taken into account that the shadow cast within craters may not correspond to the actual crater profile.

In connection with image pre-processing, Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to the aerial wartime images before further processing to address issues such as poor contrast and brightness levels. However, radiometric inconsistencies can differ largely from image to image and, thus, the same parameter selection for all images, as applied in this thesis, does not seem to be ideal. Moreover, generally an improper selection of the parameters may heavily decrease the image quality. In this context, Campos et al. (2019) have developed a learning-based method to automatically determine the two essential parameters of CLAHE (cf. Section 4.1.5). Although the results of their experiments based on different images from so-called image quality assessment data sets seem promising, the transferability of their method to aerial wartime imagery and the general benefit for the subsequent automatic detection of bomb craters need to be

examined. Of course, there are various other methods for the purpose of image enhancement besides CLAHE. In this context, Maurya et al. (2022) compare a number of different approaches, with the method presented by the authors showing enhanced quality evaluation metrics compared to the other conventional techniques. Consequently, it seems reasonable to investigate the influence of this method on the quality of the detection results in future work.

The results of the comparison of the MPP method for bomb crater detection with the Faster R-CNN object detector show the moderate superiority of the latter approach based on CNNs. Given the only slightly worse mean F1-score and the higher mean precision of the MPP, a combination of the results of both approaches might be conceivable. Moreover, the probability predicted by the CNN for a region proposal to correspond to a crater could be integrated into another prior energy term of the MPP in such a way that lower probabilities lead to an increase in energy, which in turn should further improve the quality of the results. However, according to the more recent literature, CNN-based approaches usually outperform model-based ones, especially if a sufficient amount of representative training data is available. In that case, i.e. in the long run, CNNs are considered to be the more suitable choice. In this context and in view of the proposed application scenario, instead of first applying a CNN for object detection and then deduce the impact map based on the detection results, a reasonable alternative would be a direct pixelwise classification with fully convolutional neural networks, such as U-Net (Ronneberger et al., 2015). U-Net and its variants have shown great success in related tasks of remote sensing such as the extraction of roads (Zhang et al., 2018) and, thus, it is supposed that they may perform well for the automated prediction of the impact maps, too. As this approach also requires a sufficient amount of labelled training data, future research should additionally investigate methods that try to reduce the amount of required labelled samples. One example of such methods is semi-supervised learning (e.g. Ling et al., 2018), where one tries to incorporate unlabelled samples to the training process to improve the classifier.

It can be concluded that with the procedure presented in this work, the stated objectives could be achieved. The developed concept for generating an impact map from bomb craters automatically detected in aerial wartime images can be seen as a further step towards the thorough understanding of this topic and can serve as a solid basis for promising future research addressing the discussed possibilities.

## A Evaluation of the assignment process in the context of redundant image information

In the presence of multiple overlapping aerial wartime images, the MPP procedure is applied independently to all images (Section 4.1). Subsequently, the results are combined, i.e. detections from multiple overlapping images that refer to the same object (e.g. a certain bomb crater) have to be matched. Due to the coarse georeferencing accuracy of the images of about 5 m - 40 m, incorrect assignments are possible in the local approach employed in this thesis (cf. Section 4.2). Given the overall objective to generate impact maps from bomb craters automatically detected in such aerial images, this appendix is concerned with the question to what extent the inaccurate georeferencing affects the quality of the impact maps. For this purpose, a simulation has been performed based on 18 aerial wartime images with a reference (master images; detections from a master image will be referred to as master detections MDs and detections from other images as non-MDs) from data set A (see Section 5.2), which is described below.

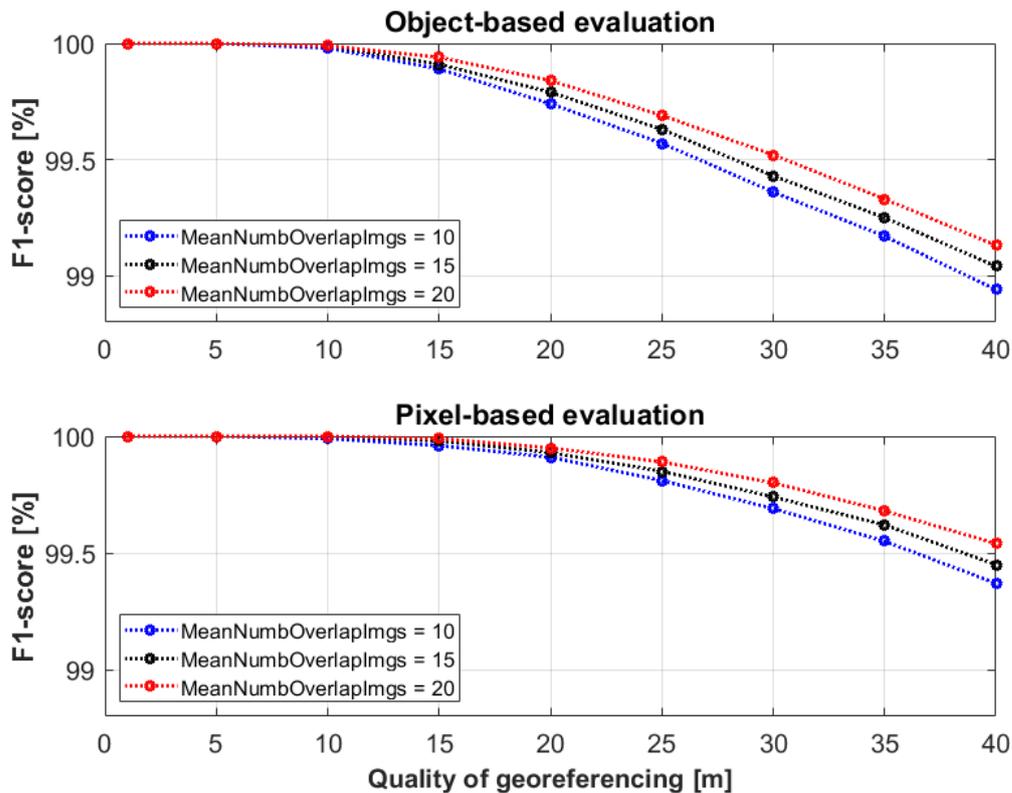
For the simulation, two parameters are of relevance. First,  $Georef_{Acc}$ , representing the quality of the georeferencing, i.e. in the images the positions of the same object can differ by up to  $Georef_{Acc}$  metres, where  $Georef_{Acc} = \{1 \text{ m}, 5 \text{ m}, 10 \text{ m}, 15 \text{ m}, \dots, 40 \text{ m}\}$ . This parameter is gradually increased to allow the simulation of different georeferencing accuracies that may appear in the data. Moreover, by increasing it until 40 m, the influence of the georeferencing accuracy on the results appearing in the data used in this thesis can be investigated. Second, the mean number of images  $MeanNumbOverlapImgs$  that overlap a master image. This parameter follows a normal distribution with  $\mu = 10, 15, 20$  and  $\sigma = 2.5$  to consider the fact that each MD may be covered by a different number of images  $NumbOverlapImgs$ . The values for  $\mu$  and  $\sigma$  are based on experience from own works (e.g. Kruse et al., 2022).

The simulation is carried out as follows: First of all, a MD is generated at each centre of a crater in the master image, where the crater centres are known due to the provided reference. Then, according to the normal distribution, the number of overlapping images  $NumbOverlapImgs$  per MD is determined. Thus, there can be a maximum of  $NumbOverlapImgs + 1$  (MD) detections per crater. Subsequently, the respective number of non-MDs is randomly created within a radius of  $Georef_{Acc}$  meters around the MDs. Here, it is considered that non-MDs from the same image will have (almost) identical offsets to the respective MDs in the local neighbourhood. Finally, it is simulated that the MPP procedure (Section 4.1) detects only a limited number of craters per image. For this purpose, a random number of detections is deleted in each image (this also applies to

MDs) according to a uniform distribution in the interval  $[0, NumbOfCrat]$ , where  $NumbOfCrat$  corresponds to the number of craters in the image.

For each possible parameter constellation, e.g.  $MeanNumbOverlapImgs = 15$  and  $Georef_{Acc} = 30$  m, the previously described simulation procedure plus the developed fusion process (Section 4.2) is repeated 100 times on each of the 18 aerial wartime images, resulting in  $3 \cdot 9 \cdot 1800 = 48600$  computations. Corresponding results with the mean F1-scores (object- and pixel-based; Section 5.4) as a function of the georeferencing quality for different mean numbers of overlapping images are illustrated in Figure A.1.

In Figure A.1, it can be seen for both the object- and pixel-based evaluation that for georeferencing accuracies up to 10 m the F1-score is (almost) 100 %, independent of the mean number of overlapping images. Then, in the object-based evaluation, the F1-score decreases slowly until a georeferencing accuracy of about 30 m and subsequently more or less linearly. This is similar for the evaluation results of the impact map (pixel-based evaluation), where the F1-score decreases slowly until about  $Georef_{Acc} = 40$  m and afterwards decreases linearly likewise (not shown). The curves for different values of  $MeanNumbOverlapImgs$  exhibit a very similar behaviour, with



**Figure A.1:** Mean F1-scores (object- and pixel-based) as a function of the georeferencing quality for different mean numbers of overlapping images  $MeanNumbOverlapImgs$  based on 18 aerial wartime images. For each parameter constellation, the same experiment is repeated 100 times.

marginally higher F1-scores for larger values of *MeanNumbOverlapImgs*. Having a look at  $Georef_{Acc} = 40$  m, which basically reflects the georeferencing accuracy of the data used in this work, shows F1-scores of about 99.0 % and 99.5 % for the object- and pixel-based evaluations, respectively.

Consequently, it can be summarized that, not surprisingly, there would be a benefit of a more accurate co-registration of the respective images. As Figure A.1 shows and as described earlier, the procedure presented in Section 4.2 would work completely correctly if the positions of corresponding objects in the images are not more than approx. 10 m apart. On the other hand, the simulation has shown that the inaccurate georeferencing of the data used in this thesis leads to a very small quality loss of the generated impact maps (F1-score = 99.5 %). Also, considering that the quality of the results for the impact maps in connection with redundant image information (Section 6.2.2) is in the range of 70 % for the F1-score, the discrepancies can be regarded as negligible. Moreover, it should be noted that generally it is not very critical whether the area designated as to be probed is marginally larger or smaller, which would be the case in clusters of detections. For detections that have a sufficient distance to other detections, there are no problems in the assignment process. Finally, given that the assignment process operates locally, misclassified areas would occur exclusively in the proximity of areas designated as to be probed.



---

## References

- Ackermann F, Krzystek P (1997) Complete automation of digital aerial triangulation. *The Photogrammetric Record* 15(89):645–656.
- Andersen H-E, Reutebuch S, Schreuder G (2012) Bayesian object recognition for the analysis of complex forest scenes in airborne laser scanner data. *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXIV-3A*:35–41.
- Anderson S, Marcus LF (1993) Effect of quadrat size on measurements of species density. *Journal of Biogeography* 20(4):421–428.
- Andrieu C, de Freitas N, Doucet A, Jordan MI (2003) An introduction to MCMC for machine learning. *Machine Learning* 50(1):5–43.
- Anselin L, Syabri I, Kho Y (2010) GeoDa: An introduction to spatial data analysis. In: Fischer MM, Getis A (eds) *Handbook of Applied Spatial Analysis: Software Tools, Methods and Application*. Springer, Berlin, Heidelberg, pp 73–89.
- Arslan F, Descombes X, Zerubia J (2009) Object extraction from high resolution SAR images using a birth and death dynamics. In: *Proceedings of the IEEE International Conference on Image Processing*, pp 561–564.
- Aval J, Demuyneck J, Zenou E, Fabre S, Sheeren D, Fauvel M, Adeline K, Briottet X (2018) Detection of individual trees in urban alignment from airborne data and contextual information: A marked point process approach. *ISPRS Journal of Photogrammetry and Remote Sensing* 146:197–210.
- Babu GJ, Feigelson ED (1996) *Astrostatistics*. Vol 3, Chapman & Hall, London, UK.
- Baddeley A (2007) Spatial point processes and their applications. In: Baddeley A, Bárány I, Schneider R, Weil W (eds) *Stochastic Geometry: Lecture Notes in Mathematics*. Vol 1892, Springer, Berlin, Heidelberg, pp 1–75.
- Baddeley AJ, van Lieshout MNM (1993) Stochastic geometry models in high-level vision. *Journal of Applied Statistics* 20(5-6):231–256.
- Bailey TC (1994) A review of statistical spatial analysis in geographical information systems. In: Fotheringham AS, Rogerson P (eds) *Spatial Analysis and GIS*. Taylor & Francis, London, UK, pp 13–44.
- Baltsavias EP (2004) Object extraction and revision by image analysis using existing geodata and knowledge: Current status and steps towards operational systems. *ISPRS Journal of Photogrammetry and Remote Sensing* 58(3-4):129–151.
- Bandeira L, Ding W, Stepinski TF (2012) Detection of sub-kilometer craters in high resolution planetary images using shape and texture features. *Advances in Space Research* 49(1):64–74.

- Bandeira L, Saraiva J, Pina P (2007) Impact crater recognition on Mars based on a probability volume created by template matching. *IEEE Transactions on Geoscience and Remote Sensing* 45(12):4008–4015.
- Baxter MJ, Beardah CC, Wright RVS (1997) Some archaeological applications of kernel density estimates. *Journal of Archaeological Science* 24(4):347–354.
- Beichelt FE, Montgomery DC (2003) *Teubner-Taschenbuch der Stochastik: Wahrscheinlichkeitstheorie, Stochastische Prozesse, Mathematische Statistik*. Teubner, Stuttgart, Leipzig, Wiesbaden.
- Ben Hadj S, Chatelain F, Descombes X, Zerubia J (2010) Parameter estimation for a marked point process within a framework of multidimensional shape extraction from remote sensing images. In: *Proceedings of the ISPRS Technical Commission III Symposium on Photogrammetry Computer Vision and Image Analysis*, pp 1–6.
- Ben Salah R, Alata O, Tremblais B, Thomas L, David L (2018) Tomographic reconstruction of 3D objects using marked point process framework. *Journal of Mathematical Imaging and Vision* 60(7):1132–1149.
- Benchmark (2013) *Population counting*. [http://www-sop.inria.fr/members/Florent.Lafarge/benchmark/population\\_counting/counting.html](http://www-sop.inria.fr/members/Florent.Lafarge/benchmark/population_counting/counting.html). Accessed 18 July 2022.
- Benedek C (2017) An embedded marked point process framework for three-level object population analysis. *IEEE Transactions on Image Processing* 26(9):4430–4445.
- Benedek C, Descombes X, Zerubia J (2012) Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(1):33–50.
- Benedek C, Martorella M (2014) Moving target analysis in ISAR image sequences with a multiframe marked point process model. *IEEE Transactions on Geoscience and Remote Sensing* 52(4):2234–2246.
- Benedek C, Sziranyi T (2009) Change detection in optical aerial images by a multilayer conditional mixed Markov model. *IEEE Transactions on Geoscience and Remote Sensing* 47(10):3416–3430.
- Benedix GK, Lagain A, Chai K, Meka S, Anderson S, Norman C, Bland PA, Paxman J, Towner MC, Tan T (2020) Deriving surface ages on Mars using automated crater counting. *Earth and Space Science* 7(3):e2019EA001005.
- Bhattacharyya A (1943) On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society* 35:99–109.
- Bishop CM (2006) *Pattern recognition and machine learning*. Springer, New York, USA.
- Blaschke T (2010) Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65(1):2–16.
- Bonnier A, Finné M, Weiberg E (2019) Examining land-use through GIS-based kernel density estimation: A re-evaluation of legacy data from the Berbati-Limnes survey. *Journal of Field Archaeology* 44(2):70–83.
- Börcs A, Benedek C (2015) Extraction of vehicle groups in airborne lidar point clouds with two-level point processes. *IEEE Transactions on Geoscience and Remote Sensing* 53(3):1475–1489.

- Brédif M, Tournaire O, Vallet B, Champion N (2013) Extracting polygonal building footprints from digital surface models: A fully-automatic global optimization framework. *ISPRS Journal of Photogrammetry and Remote Sensing* 77:57–65.
- Breiman L (2001) Random forests. *Machine learning* 45(1):5–32.
- Brenner S, Zambanini S, Sablatnig R (2018) Detection of bomb craters in WWII aerial images. In: *Proceedings of the Austrian Association for Pattern Recognition Workshop*, pp 94–97.
- Campos GFC, Mastelini SM, Aguiar GJ, Mantovani RG, Melo LF de, Barbon S (2019) Machine learning hyperparameter selection for contrast limited adaptive histogram equalization. *EURASIP Journal on Image Video Processing* 1:1–18.
- Canny J (1986) A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8(6):679–698.
- Cedilnik N, Debreuve E, de Graeve F, Besse F, Descombes X (2018) SPADE: A small particle detection method using a dictionary of shapes within the marked point process framework. In: *Proceedings of the IEEE International Symposium on Biomedical Imaging*.
- Cerny V (1985) Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications* 45(1):41–51.
- Chai D, Förstner W, Lafarge F (2013) Recovering line-networks in images by junction-point processes. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 1894–1901.
- Chai D, Förstner W, Ying Yang M (2012) Combine Markov random fields and marked point processes to extract building from remotely sensed images. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* I-3:365–370.
- Chai D, Schmidt A, Heipke C (2016) Detecting linear features by spatial point processes. *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B3:841–848.
- Chen H-Z, Jing N, Wang J, Chen Y-G, Chen L (2014) A novel saliency detection method for lunar remote sensing images. *IEEE Geoscience and Remote Sensing Letters* 11(1):24–28.
- Chen M, Liu D, Qian K, Li J, Lei M, Zhou Y (2018) Lunar crater detection based on terrain analysis and mathematical morphology methods using digital elevation models. *IEEE Transactions on Geoscience and Remote Sensing* 56(7):3681–3692.
- Chen SX (2000) Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics* 52(3):471–480.
- Cheng G, Han J (2016) A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing* 117:11–28.
- Clermont D, Kruse C, Rottensteiner F, Heipke C (2019) Supervised detection of bomb craters in historical aerial images using convolutional neural networks. *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2/W16:67–74.
- Cohen JP, Lo Z H, Lu T, Ding W (2016) Crater detection via convolutional neural networks. *arXiv preprint arXiv:1601.00978*.
- Cortes C, Vapnik VN (1995) Support-vector networks. *Machine Learning* 20(3):273–297.

- Craciun P, Zerubia J (2013) Unsupervised marked point process model for boat extraction in harbors from high resolution optical remotely sensed images. In: *Proceedings of the IEEE International Conference on Image Processing*, pp 4122–4125.
- Cross AM (1988) Detection of circular geological features using the Hough transform. *International Journal of Remote Sensing* 9(9):1519–1528.
- Daley D, Vere-Jones D (2003) *An introduction to the theory of point processes: Volume I: Elementary theory and methods*. Springer, New York, USA.
- Dekking FM, Kraaikamp C, Lopuhaä HP, Meester LE (2005) *A modern introduction to probability and statistics: Understanding why and how*. Vol 488, Springer, London, UK.
- DeLatte DM, Crites ST, Guttenberg N, Tasker EJ, Yairi T (2019) Segmentation convolutional neural networks for automatic crater detection on Mars. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12(8):2944–2957.
- Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) ImageNet: A large-scale hierarchical image database. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 248–255.
- Descamps S, Béchet A, Descombes X, Arnaud A, Zerubia J (2011) An automatic counter for aerial images of aggregations of large birds. *Bird Study* 58(3):302–308.
- Descamps S, Descombes X, Béchet A, Zerubia J (2008) Automatic flamingo detection using a multiple birth and death process. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp 1113–1116.
- Descombes X (2017) Multiple objects detection in biological images using a marked point process framework. *Methods* 115:2–8.
- Descombes X, Minlos R, Zhizhina E (2009) Object extraction using a stochastic birth-and-death dynamics in continuum. *Journal of Mathematical Imaging and Vision* 33(3):347–359.
- Descombes X, Zerubia J (2002) Marked point process in image analysis. *IEEE Signal Processing Magazine* 19(5):77–84.
- Dhillon A, Verma GK (2020) Convolutional neural network: A review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence* 9(2):85–112.
- Ding M, Cao Y, Wu Q (2013) Novel approach of crater detection by crater candidate region selection and matrix-pattern-oriented least squares support vector machine. *Chinese Journal of Aeronautics* 26(2):385–393.
- Dolejš M, Pacina J, Veselý M, Brétt D (2020) Aerial bombing crater identification: Exploitation of precise digital terrain models. *ISPRS International Journal of Geo-Information* 9(12):713.
- Downes LM, Steiner TJ, How JP (2021) Neural network approach to crater detection for lunar terrain relative navigation. *Journal of Aerospace Information Systems* 18(7):391–403.
- Dubosclard P, Larnier S, Konik H, Herbulot A, Devy M (2014) Automatic method for visual grading of seed food products. In: *Proceedings of the International Conference on Image Analysis and Recognition*, pp 485–495.

- Emami E, Ahmad T, Bebis G, Nefian A, Fong T (2019) Crater detection using unsupervised algorithms and convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing* 57(8):5373–5383.
- Emami E, Bebis G, Nefian A, Fong T (2015) Automatic crater detection using convex grouping and convolutional neural networks. In: *Proceedings of the International Symposium on Visual Computing*, pp 213–224.
- Epanechnikov VA (1969) Non-parametric estimation of a multivariate probability density. *Theory of Probability and Its Applications* 14(1):153–158.
- Farella EM, Morelli L, Remondino F, Mills JP, Haala N, Cromptvoets J (2022) The EuroSDR time benchmark for historical aerial images. *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B2-2022*:1175–1182.
- Favreau J-D, Lafarge F, Bousseau A, Auvolet A (2019) Extracting geometric structures in images with Delaunay point processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42(4):837–850.
- Fischler MA, Elschlager RA (1973) The representation and matching of pictorial structures. *IEEE Transactions on Computers* 22(1):67–92.
- Florescu I (2014) *Probability and stochastic processes*. John Wiley & Sons, Hoboken, New Jersey, USA.
- Foley J (2008) Demonstration of LiDAR and orthophotography for wide area assessment at Pueblo Precision Bombing Range #2, Colorado. *Final report*, Sky Research, Ashland, Oregon, USA.
- Förstner W, Gülch E (1987) A fast operator for detection and precise location of distinct points, corners and centers of circular features. In: *Proceedings of the ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pp 281–305.
- Fotheringham AS, Wong DWS (1991) The modifiable areal unit problem in multivariate statistical analysis. *Environment and planning A* 23(7):1025–1044.
- Furlanello C, Merler S, Menegon S, Paoli E, Fontanari S (2003) Mapping the risk of unexploded bombs from World War Two. *Institute for Scientific and Technology Research*, Trento, Italy.
- Gadgil N, Salama P, Dunn K, Delp E (2016) Nuclei segmentation of fluorescence microscopy images based on midpoint analysis and marked point process. In: *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation*, pp 37–40.
- Gamal-Eldin A, Descombes X, Charpiat G, Zerubia J (2010) Multiple birth and cut algorithm for multiple object detection. *Journal of Multimedia Processing and Technologies* 1(4):260–276.
- Gatrell AC, Bailey TC, Diggle PJ, Rowlingson BS (1996) Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers* 21(1):256–274.
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6(6):721–741.
- Ghanta S, Dy JG, Niu D, Jordan MI (2018) Latent marked Poisson process with applications to object segmentation. *Bayesian Analysis* 13(1):85–113.

- Gilks WR, Richardson S, Spiegelhalter DJ (1996) *Markov chain Monte Carlo in practice*. Chapman & Hall, London, UK.
- Girshick R (2015) Fast R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp 1440–1448.
- Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 580–587.
- Goodchild MF (2008) Data analysis, spatial. In: Shekhar S, Xiong H (eds) *Encyclopedia of GIS*. Springer, New York, USA, pp 200–203.
- Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82(4):711–732.
- Hall P, Marron JS (1987) On the amount of noise inherent in bandwidth selection for a kernel density estimator. *The Annals of Statistics* 15(1):163–181.
- Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1):97–109.
- Hay GJ, Castilla G, Wulder MA, Ruiz JR (2005) An automated object-based approach for the multiscale image segmentation of forest scenes. *International Journal of Applied Earth Observation and Geoinformation* 7(4):339–359.
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 770–778.
- Hervieu A, Soheilian B, Brédif M (2015) Road marking extraction using a model & data-driven RJ-MCMC. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W4:47–54.
- Honda R, Konishi O, Azuma R, Yokogawa H, Yamanaka S, Iijima Y (2000) Data mining system for planetary images—crater detection and categorization. In: *Proceedings of the International Workshop on Machine Learning of Spatial Knowledge in conjunction with ICML*, pp 103–108.
- Hossain MD, Chen D (2019) Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing* 150:115–134.
- Hu Y, Wang F, Guin C, Zhu H (2018) A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation. *Applied Geography* 99:89–97.
- Huang G, Liu Z, van der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 2261–2269.
- Huang S, Yin J, Zhu H, Cao Z (2018) Gaussian attractive force-based alternative parametric active contour model for 3D lunar crater detection. In: *IEEE International Symposium on Geoscience and Remote Sensing*, pp 7082–7085.
- Izenman AJ (1991) Review papers: Recent developments in nonparametric density estimation. *Journal of the American Statistical Association* 86(413):205–224.

- Jain AK, Zhong Y, Dubuisson-Jolly MP (1998) Deformable template models: A review. *Signal Processing* 71(2):109–129.
- Jensen L, Drauschke M, Förstner W (2010) Automatische Detektion von Bombentrichtern in digitalisierten Luftbildern des 2. Weltkriegs. In: *DGPF Tagungsband 19/2010*, pp 228–235.
- Jeong S-G, Tarabalka Y, Zerubia J (2014) Marked point process model for facial wrinkle detection. In: *Proceedings of the IEEE International Conference on Image Processing*, pp 1391–1394.
- Jin S, Zhang T (2014) Automatic detection of impact craters on Mars using a modified adaboosting method. *Planetary and Space Science* 99:112–117.
- Juhász A, Neuberger H (2018) Automatic identification of bomb craters and their potential location. *Mitteilungen der Österreichischen Geographischen Gesellschaft* 160(1):241–258.
- Kahn H, Harris TE (1951) Estimation of particle transmission by random sampling. *National Bureau of Standards: Applied Mathematics Series* 12:27–30.
- Kang Z, Wang X, Hu T, Yang J (2019) Coarse-to-fine extraction of small-scale lunar impact craters from the CCD images of the Chang'E lunar orbiters. *IEEE Transactions on Geoscience and Remote Sensing* 57(1):181–193.
- Kazimi B, Thiemann F, Sester M (2019) Semantic segmentation of manmade landscape structures in digital terrain models. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-2/W7:87–94.
- Kim DW, Aguilar C, Zhao H, Comer ML (2019) Narrow gap detection in microscope images using marked point process modeling. *IEEE Transactions on Image Processing* 28(10):5064–5076.
- Kim J-R, Muller J-P, Gasselt S, Morley J, Neukum G (2005) Automated crater detection, a new tool for Mars cartography and chronology. *Photogrammetric Engineering and Remote Sensing*, 71(10):1205–1217.
- Kim T, Park S-R, Kim M-G, Jeong S, Kim K-O (2004) Tracking road centerlines from high resolution remote sensing images by least squares correlation matching. *Photogrammetric Engineering and Remote Sensing* 70(12):1417–1422.
- Kirkpatrick S, Gelatt C, Vecchi M (1983) Optimization by simulated annealing. *Science* 220(4598):671–680.
- Kowal M, Korbicz J (2018) Marked point process for nuclei detection in breast cancer microscopic images. In: *Proceedings of the Polish Conference on Biocybernetics and Biomedical Engineering*, pp 230–241.
- Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: *Proceedings of the International Conference on Neural Information Processing Systems*, pp 1097–1105.
- Kruse C, Rottensteiner F, Heipke C (2019) Marked point processes for the automatic detection of bomb craters in aerial wartime images. *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2/W13:51–60.

- Kruse C, Rottensteiner F, Heipke C (2020) Using redundant image information from multiple aerial images for the detection of bomb craters based on marked point processes. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* V-2-2020:861–870.
- Kruse C, Rottensteiner F, Hoberg T, Ziemis M, Rebke J, Heipke C (2018) Generating impact maps from automatically detected bomb craters in aerial wartime images using marked point processes. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-3:127–134.
- Kruse C, Wittich D, Rottensteiner F, Heipke C (2022) Generating impact maps from bomb craters automatically detected in aerial wartime images using marked point processes. *ISPRS Open Journal of Photogrammetry and Remote Sensing* 5:100017.
- Kulikova M, Jermyn I, Descombes X, Zerubia J, Zhizhina E (2009) A marked point process model with strong prior shape information for the extraction of multiple, arbitrarily-shaped objects. In: *Proceedings of the IEEE International Conference on Signal-Image Technologies and Internet-Based System*, pp 180–186.
- Kulikova M, Veillard A, Roux L, Racoceanu D (2012) Nuclei extraction from histopathological images using a marked point process approach. In: *Medical Imaging: Image Processing, Proceedings of SPIE*, Vol 8314, pp 831428-1–831428-8.
- Kumar S, Hebert M (2006) Discriminative random fields. *International Journal of Computer Vision* 68(2):179–201.
- Kuter N, Yenilmez F, Kuter S (2011) Forest fire risk mapping by kernel density estimation. *Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering* 32(2):599–610.
- Lacoste C, Descombes X, Zerubia J (2005) Point processes for unsupervised line network extraction in remote sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(10):1568–1579.
- Lacroix V, Vanhuysse S (2015) Crater detection using CGC - a new circle detection method. In: *Proceedings of the International Conference on Pattern Recognition Applications and Methods*, Vol 1, pp 320–327.
- Lafarge F, Gimel'farb G, Descombes X (2010) Geometric feature extraction by a multimarked point process. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(9):1597–1609.
- Lamperti J (1977) *Stochastic processes: A survey of the mathematical theory*. Applied Mathematical Sciences, Vol 23, Springer, New York, Heidelberg, Berlin.
- Lee C (2019) Automated crater detection on Mars using deep learning. *Planetary and Space Science* 170:16–28.
- Lee C, Hogan J (2021) Automated crater detection with human level performance. *Computers and Geosciences* 147:104645.
- Li H, Jiang B, Li Y, Le Cao (2021) A combined method of crater detection and recognition based on deep learning. *Systems Science and Control Engineering* 9(2):132–140.
- Li J, Yu Y, Guan H, Gong Z (2016) Extraction of tree crowns from mobile laser scanning data using a marked point process model. In: *ISPRS International Conference on Computer Vision in Remote Sensing, Proceedings of SPIE*, Vol 9901, pp 990107-1–990107-6.

- Li T, Comer M, Zerubia J (2018) A connected-tube MPP model for object detection with application to materials and remotely-sensed images. In: *Proceedings of the IEEE International Conference on Image Processing*, pp 1323–1327.
- Li Y, Li J (2010) Oil spill detection from SAR intensity imagery using a marked point process. *Remote Sensing of Environment* 114(7):1590–1601.
- Li Y, Wang S, Tian Q, Ding X (2015) Feature representation for statistical-learning-based object detection: A review. *Pattern Recognition* 48(11):3542–3559.
- Lin E, Qin R, Edgerton J, Kong D (2020) Crater detection from commercial satellite imagery to estimate unexploded ordnance in Cambodian agricultural land. *PLoS One* 15(3):e0229826.
- Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 2117–2125.
- Lin Y-P, Chu H-J, Wu C-F, Chang T-K, Chen C-Y (2011) Hotspot analysis of spatial environmental pollutants using kernel density estimation and geostatistical techniques. *International Journal of Environmental Research and Public Health* 8(1):75–88.
- Ling Z, Li X, Zou W, Guo S (2018) Semi-supervised learning via convolutional neural network for hyperspectral image classification. In: *Proceedings of the IEEE International Conference on Pattern Recognition*, pp 1–6.
- Liu A, Zhou D, Chen L, Chen M (2015) Saliency detection and edge feature matching approach for crater extraction. *Journal of Systems Engineering and Electronics* 26(6):1291–1300.
- Liu D, Chen M, Qian K, Lei M, Zhou Y (2017) Boundary detection of dispersal impact craters based on morphological characteristics using lunar digital elevation model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10(12):5632–5646.
- Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2):91–110.
- Lücke C, Cruse C, Mütterthies A (2007) Entwicklung und operationelle Anwendung der halbautomatisierten Kriegsflugbildanalyse zur Detektion von Blindgängern. In: *Photogrammetrie Fernerkundung Geoinformation (PFG)*, Heft 7, pp 495–502.
- Ludwig JA, Reynolds JF (1988) *Statistical ecology: A primer on methods and computing*. John Wiley & Sons, New York, USA.
- MacKay DJC (2003) *Information theory, inference and learning algorithms*. Cambridge University Press, Cambridge, UK.
- Mallet C, Lafarge F, Roux M, Soergel U, Bretar F, Heipke C (2010) A marked point process for modeling lidar waveforms. *IEEE Transactions on Image Processing* 19(12):3204–3221.
- Mallick S (2022) *Blob detection using OpenCV (Python, C++)*. <https://www.learnopencv.com/blob-detection-using-opencv-python-c/>. Accessed 10 April 2022.
- Marron JS (1988) Automatic smoothing parameter selection: A survey. *Empirical Economics* 13(3):187–208.

- Martins R, Pina P, Marques J, Silveira M (2009) Crater detection by a boosting approach. *IEEE Geoscience and Remote Sensing Letters* 6(1):127–131.
- Maurya L, Lohchab V, Kumar Mahapatra P, Abonyi J (2022) Contrast and brightness balance in image enhancement using Cuckoo Search-optimized image fusion. *Journal of King Saud University - Computer and Information Sciences* 34(9):7247–7258.
- McKeown DM, Denlinger JL (1988) Cooperative methods for road tracking in aerial imagery. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 662–672.
- Meng D, Yunfeng C, Qingxian W (2009) Method of passive image based crater autonomous detection. *Chinese Journal of Aeronautics* 22(3):301–306.
- Merler S, Furlanello C, Jurman G (2005) Machine learning on historic air photographs for mapping risk of unexploded bombs. In: Roli F, Vitulano S (eds) *Image Analysis and Processing – ICIAP 2005. Lecture Notes in Computer Science*, Vol 3617, Springer, Berlin, Heidelberg, pp 735–742.
- Merzbach E, Nualart D (1986) A characterization of the spatial Poisson process and changing time. *The Annals of Probability* 14(4):1380–1390.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21(6):1087–1092.
- Neal RM (1993) Probabilistic inference using Markov chain Monte Carlo methods. *Technical Report CRG-TR-93-1*, Department of Computer Science, University of Toronto, Canada.
- Ortner M, Descombes X, Zerubia J (2008) A marked point process of rectangles and segments for automatic analysis of digital elevation models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(1):105–119.
- Ortner M, Descombes X, Zerubia J (2007) Building outline extraction from digital elevation models using marked point processes. *International Journal of Computer Vision* 72(2):107–132.
- Oyana TJ, Margai F (2015) *Spatial analysis: Statistics, visualization, and computational methods*. First Edition, CRC Press, Boca Raton, Florida, USA.
- Parzen E (1962) On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33(3):1065–1076.
- Pedrosa MM, Azevedo SC de, Da Silva EA, Dias MA (2017) Improved automatic impact crater detection on Mars based on morphological image processing and template matching. *Geomatics, Natural Hazards and Risk* 8(2):1306–1319.
- Peng J, Liu YC (2005) Model and context-driven building extraction in dense urban aerial images. *International Journal of Remote Sensing* 26(7):1289–1307.
- Perrin G, Descombes X, Zerubia J (2004) Tree crown extraction using marked point processes. In: *Proceedings of the IEEE European Signal Processing Conference*, pp 2127–2130.
- Perrin G, Descombes X, Zerubia J (2005) A marked point process model for tree crown extraction in plantations. In: *Proceedings of the IEEE International Conference on Image Processing*, pp 661–664.
- Perrin G, Descombes X, Zerubia J (2006) 2D and 3D vegetation resource parameters assessment using marked point processes. In: *Proceedings of the IEEE International Conference on Pattern Recognition*, pp 1–4.

- Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, ter Haar Romeny BM, Zimmerman JB, Zuiderveld K (1987) Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing* 39(3):355–368.
- Poulain E, Prigent S, Soubies E, Descombes X (2015) Cells detection using segmentation competition. In: *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp 1208–1211.
- Propp JG, Wilson DB (1996) Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* 9(1-2):223–252.
- Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 779–788.
- Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: Towards real-time object detection with region proposal networks. *Proceedings of the International Conference on Neural Information Processing Systems*, Vol 1, pp 91–99.
- Richardson S, Green PJ (1997) On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society: Series B* 59(4):731–792.
- Robbins SJ, Hynek BM (2012) A new global database of Mars impact craters  $\geq 1$  km: 1. Database creation, properties, and parameters. *Journal of Geophysical Research* 117:E05004.
- Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds) *Medical Image Computing and Computer-Assisted Intervention*, Vol 9351, Springer, Cham, pp 234–241.
- Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics* 27(3):832–837.
- Ross SM (1996) *Stochastic processes*. Second Edition, John Wiley & Sons, New York, USA.
- Rue H, Hurn MA (1999) Bayesian object identification. *Biometrika* 86(3):649–660.
- Rushton G, Peleg I, Banerjee A, Smith G, West M (2004) Analyzing geographic patterns of disease incidence: Rates of late-stage colorectal cancer in Iowa. *Journal of Medical Systems* 28(3):223–236.
- Salamon P, Sibani P, Frost R (2002) *Facts, conjectures, and improvements for simulated annealing*. Society for Industrial and Applied Mathematics, Philadelphia, USA.
- Salamuniccar G, Loncaric S (2008) Open framework for objective evaluation of crater detection algorithms with first testfield subsystem based on MOLA data. *Advances in Space Research* 42(1):6–19.
- Salamuniccar G, Loncaric S (2010) Method for crater detection from Martian digital topography data using gradient value/orientation, morphometry, vote analysis, slip tuning, and calibration. *IEEE Transactions on Geoscience and Remote Sensing* 48(5):2317–2329.
- Salih A, Schulte P, Grumpe A, Wohler C, Hiesinger H (2017) Automatic crater detection and age estimation for mare regions on the lunar surface. In: *Proceedings of the IEEE European Signal Processing Conference*, pp 518–522.
- Savage R, Palafox LF, Morrison CT, Rodriguez JJ, Barnard K, Byrne S, Hamilton CW (2018) A Bayesian approach to subkilometer crater shape analysis using individual HiRISE images. *IEEE Transactions on Geoscience and Remote Sensing* 56(10):5802–5812.

- Schmidt A, Lafarge F, Brenner C, Rottensteiner F, Heipke C (2017) Forest point processes for the automatic extraction of networks in raster data. *ISPRS Journal of Photogrammetry and Remote Sensing* 126:38–55.
- Scott DW (2015) Kernel density estimators. In: Scott DW (ed) *Multivariate density estimation: Theory, practice, and visualization*. Wiley, Hoboken, New Jersey, USA, pp 137–216.
- Sechen C, Sangiovanni-Vincentelli A (1985) The TimberWolf placement and routing package. *IEEE Journal of Solid-State Circuits* 20(2):510–522.
- Shchur O, Türkmen AC, Januschowski T, Günnemann S (2021) Neural temporal point processes: A review. In: *Proceedings of the International Joint Conference on Artificial Intelligence*, pp 4585–4593.
- Silburt A, Ali-Dib M, Zhu C, Jackson A, Valencia D, Kissin Y, Tamayo D, Menou K (2018) Lunar crater identification via deep learning. *Icarus* 317(6):27–38.
- Silverman BW (1986) *Density estimation for statistics and data analysis*. Chapman & Hall, London, UK.
- Simonoff JS (1996) *Smoothing methods in statistics*. Springer, New York, USA.
- Smirnov AA (2002) Exploratory study of automated crater detection algorithm. *Technical report*, Boulder, Colorado, USA.
- Smith K (2007) *Bayesian methods for visual multi-object tracking with applications to human activity recognition*. PhD thesis, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.
- Solarna D, Gotelli A, Le Moigne J, Moser G, Serpico SB (2020) Crater detection and registration of planetary images through marked point processes, multiscale decomposition, and region-based analysis. *IEEE Transactions on Geoscience and Remote Sensing* 58(9):6039–6058.
- Solarna D, Moser G, Moigne J, Serpico S (2017) Planetary crater detection and registration using marked point processes, multiple birth and death algorithms, and region-based analysis. In: *IEEE International Symposium on Geoscience and Remote Sensing*, pp 2337–2340.
- Soubies E, Weiss P, Descombes X (2013) A 3D segmentation algorithm for ellipsoidal shapes. Application to nuclei extraction. In: *Proceedings of the International Conference on Pattern Recognition Applications and Methods*, pp 97–105.
- Stoica R, Descombes X, Zerubia J (2004) A Gibbs point process for road extraction from remotely sensed images. *International Journal of Computer Vision* 57(2):121–136.
- Stoica RS, Martínez VJ, Saar E (2007) A three-dimensional object point process for detection of cosmic filaments. *Journal of the Royal Statistical Society. Series C Applied Statistics* 56(4):459–477.
- Stoyan D, Kendall WS, Mecke J (1995) *Stochastic geometry and its applications*. Second Edition, Wiley, Chichester, UK.
- Sun K, Sang N, Zhang T (2007) Marked point process for vascular tree extraction on angiogram. In: *Proceedings of the International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp 467–478.
- Szegedy C, Ioffe S, Vanhoucke V, Alemi A (2017) Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp 4278–4284.

- Tavakkoli Sabour SM, Agarius J, Sadidi J (2014) Calculation of per parcel probability for dud bombs in Germany. *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XL-2/W3*:261–266.
- Terrell GR, Scott DW (1992) Variable kernel density estimation. *The Annals of Statistics* 20(3):1236–1265.
- Tewari A, Verma V, Srivastava P, Jain V, Khanna N (2022) Automated crater detection from co-registered optical images, elevation maps and slope maps using deep learning. *Planetary and Space Science* 218:105500.
- Thakali L, Kwon TJ, Fu L (2015) Identification of crash hotspots using kernel density estimation and kriging methods: A comparison. *Journal of Modern Transportation* 23(2):93–106.
- Tournaire O, Brédif M, Boldo D, Durupt M (2010) An efficient stochastic approach for building footprint extraction from digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing* 65(4):317–327.
- Tournaire O, Papanoditis N (2009) A geometric stochastic approach based on marked point processes for road mark detection from high resolution aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing* 64(6):621–631.
- Troglio G, Benediktsson J, Moser G, Serpico S (2010) Crater detection based on marked point processes. In: *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, pp 1378–1381.
- Troglio G, Le Moigne J, Benediktsson JA, Moser G, Serpico SB (2012) Automatic extraction of ellipsoidal features for planetary image registration. *IEEE Geoscience and Remote Sensing Letters* 9(1):95–99.
- Urbach ER, Stepinski TF (2009) Automatic detection of sub-km craters in high resolution planetary images. *Planetary and Space Science* 57(7):880–887.
- Utasi Á, Benedek C (2011) A 3-D marked point process model for multi-view people detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 3385–3392.
- Valjavec MB, Zorn M, Ribeiro D (2018) Mapping war geoheritage: Recognising geomorphological traces of war. *Open Geosciences* 10(1):385–394.
- Van Laarhoven PJM, Aarts EHL (1987) Simulated annealing. In: Van Laarhoven PJM, Aarts EHL (eds) *Simulated annealing: Theory and applications*. Springer, Dordrecht, Netherlands, pp 7–15.
- Van Lieshout MNM (2000) *Markov point processes and their applications*. Imperial College Press, London, UK.
- Vandoni J, Le Hégarat-Masclé S, Aldea E (2016). Crack detection based on a marked point process model. In: *Proceedings of the IEEE International Conference on Pattern Recognition*, pp 3933–3938.
- Varanelli JM (1996) *On the acceleration of simulated annealing*. PhD thesis, University of Virginia, Charlottesville, USA.
- Verdié Y, Lafarge F (2012) Efficient Monte Carlo sampler for detecting parametric objects in large scenes. In: Fitzgibbon A, Lazebnik S, Perona P, Sato Y, Schmid C (eds) *Computer Vision – ECCV 2012. Lecture Notes in Computer Science, Vol 7574*, Springer, Berlin, Heidelberg, pp 539–552.

- Verdié Y, Lafarge F (2014) Detecting parametric objects in large scenes by Monte Carlo sampling. *International Journal of Computer Vision* 106(1):57–75.
- Vinogradova T, Burl M, Mjolsness E (2002) Training of a crater detection algorithm for Mars crater imagery. In: *Proceedings of the IEEE Conference on Aerospace*, Vol 7, pp 3201–3210.
- Vogler P (2020) *Die deutsche militärische Luftbildaufklärung. Von den Anfängen bis 1945*. Karlsruher Institut für Technologie (KIT), Scientific Publishing.
- Von Neumann J (1951) Various techniques used in connection with random digits. *Journal of Research of the National Bureau of Standards* 3:36–38.
- Waga JM, Fajer M (2021) The heritage of the Second World War: Bombing in the forests and wetlands of the Kožle Basin. *Antiquity* 95(380):417–434.
- Wand MP, Jones MC (1995) *Kernel smoothing*. Chapman & Hall, London, UK.
- Wang H, Jiang J, Zhang G (2018) CraterIDNet: An end-to-end fully convolutional neural network for crater detection and identification in remotely sensed planetary images. *Remote Sensing* 10(7):1067.
- Wang Y, Wu B (2019) Active machine learning approach for crater detection from planetary imagery and digital elevation models. *IEEE Transactions on Geoscience and Remote Sensing* 57(8):5777–5789.
- Wang Y, Wu B (2020) A new global catalogue of lunar craters ( $\geq 1$  km) with 3D information and preliminary results of global analysis. *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B3-2020*:1171–1176.
- Weidner U, Förstner W (1995) Towards automatic building extraction from high-resolution digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing* 50(4):38–49.
- Wenzel S, Förstner W (2016) Facade interpretation using a marked point process. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences III-3*:363–370.
- Wetzler P, Honda R, Enke B, Merline W, Chapman C, Burl M (2005) Learning to detect small impact craters. In: *Proceedings of the IEEE Workshop on Applications of Computer Vision*, Vol 1, pp 178–184.
- Winkler G (2003) *Image analysis, random fields and Markov chain Monte Carlo methods: A mathematical introduction*. Stochastic Modelling and Applied Probability, Vol 27, Springer, Berlin, Heidelberg.
- Woicke S, Moreno Gonzalez A, El-Hajj I, Mes J, Henkel M, Klavers R (2018) Comparison of crater-detection algorithms for terrain-relative navigation. In: *AIAA Guidance, Navigation, and Control Conference*, Kissimmee, Florida.
- Yan J, Xu H, Li L (2019) Modeling and applications for temporal point processes. In: *Proceedings of the International Conference on Knowledge Discovery and Data Mining*, pp 3227–3228.
- Yang D-H, Goerge R, Mullner R (2006) Comparing GIS-based methods of measuring spatial accessibility to health services. *Journal of Medical Systems* 30(1):23–32.
- Yu Y, Li J, Guan H, Wang C, Cheng M (2012) A marked point process for automated tree detection from mobile laser scanning point cloud data. In: *Proceedings of the IEEE International Conference on Computer Vision in Remote Sensing*, pp 140–145.

- Yue S, He L, Wen Y, Lu G, Lin H (2013) Shape characteristics-based extraction of lunar impact craters: Using DEM from the Chang'E-1 satellite as a data source. *Annals of GIS* 19(1):53–62.
- Yuxin W, Kirillov A, Massa F, Lo W-Y, Girshick R (2019) *Detectron2*. <https://github.com/facebookresearch/detectron2>. Accessed 02 August 2022.
- Zhai X, Beyer L, Oliver A, Kolesnikov A (2019) S4L: Self-supervised semi-supervised learning. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp 1476–1485.
- Zhang J, Sohn G, Brédif M (2013) Single tree detection from airborne laser scanning data using a marked point process based method. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W1:41–46.
- Zhang J, Sohn G, Brédif M (2014) A hybrid framework for single tree detection from airborne laser scanning data: A case study in temperate mature coniferous forests in Ontario, Canada. *ISPRS Journal of Photogrammetry and Remote Sensing* 98:44–57.
- Zhang L, Rupnik E, Pierrot-Deseilligny M (2021) Feature matching for multi-epoch historical aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing* 182:176–189.
- Zhang Z, Liu Q, Wang Y (2018) Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters* 15(5):749–753.
- Zhao H, Comer M (2016) A hybrid Markov random field / marked point process model for analysis of materials images. *IEEE Transactions on Computational Imaging* 2(4):395–407.
- Zhao Q, Wu Y, Wang H, Li Y (2020) Road extraction from remote sensing image based on marked point process with a structure mark library. *International Journal of Remote Sensing* 41(16):6183–6208.
- Zhao Z-Q, Zheng P, Xu S-T, Wu X (2019) Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems* 30(11):3212–3232.
- Zhou J, Proisy C, Descombes X, Hedhli I, Barbier N, Zerubia J, Gastellu-Etchegorry J-P, Coueron P (2010) Tree crown detection in high resolution optical and lidar images of tropical forest. *Remote Sensing for Agriculture, Ecosystems, and Hydrology, Proceedings of SPIE*, Vol 7824, pp 78240Q-1–78240Q-6.
- Zou Z, Shi Z, Guo Y, Ye J (2019) Object detection in 20 years: A survey. *arXiv preprint arXiv:1905.05055*.



## Acknowledgements

I would like to take this opportunity to thank all the people who have supported me in the preparation of this dissertation over the last years and, therefore, have contributed to its success.

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Dr.-Ing. habil. Christian Heipke. I thank him for his excellent professional guidance and for always taking time, not only for scientific questions. His support, along with his trust and the very pleasant collaboration, have contributed significantly to the outcome of this work. Furthermore, I would like to thank apl. Prof. Dr. techn. Franz Rottensteiner for co-reviewing this dissertation, including the valuable comments and suggestions he provided. I also appreciate his expertise and the constructive discussions we had. Besides, I would like to thank Prof. Dr.-Ing. habil. Monika Sester, Prof. Dr.-Ing. Uwe Sörgel and Dr.-Ing. habil. Fabio Remondino, who additionally acted as co-reviewers for this thesis.

Parts of this dissertation were financially supported by the State Office for Geoinformation and Surveying of Lower Saxony, Germany, and its Explosive Ordnance Disposal Service (KBD) as a department of the regional directorate Hamelin-Hanover, as well as the EU-project ‘VOLTA – innoVation in geOspatiaL and 3D daTA’, for which I would like to express my gratitude. Many thanks to the KBD for the expert input and the fruitful discussions during the project meetings. In addition, I thank the company Vermessung AVT in Imst, Austria for hosting me as part of the VOLTA project – I enjoyed the seven-week stay very much. Moreover, I would like to thank the KBD, the 3D Optical Metrology research unit of the Bruno Kessler Foundation (FBK) in Trento, Italy and the Salzburger Geographisches Informationssystem (SAGIS) of the Federal State of Salzburg, Austria, for providing the data as well as Fabio Remondino and his colleagues for the great week at FBK.

To all my colleagues and friends at the Institute of Photogrammetry and GeoInformation (IPI) I would like to say thank you for the great time and the relaxed working atmosphere. Thanks a lot for making it always a pleasure to come to work. Among other things, I will certainly miss the informal conversations during coffee breaks and joint events such as Christmas parties.

Finally, I would like to thank my family who have always supported me over the years and motivated me to keep going. To close, a particularly big thank you to my fiancé Karoline for her love, support, words of encouragement and understanding during stressful times.



# Curriculum Vitae

## Personal information

Name	Christian Kruse
Date and place of birth	20.07.1990 in Hanover, Germany

## Work experience

since 01/2023	<b>Head of department</b> State Office for Geoinformation and Surveying of Lower Saxony, Regional directorate Hamelin-Hanover, Department 2 Geodata management, Germany
01/2022 – 12/2022	<b>Guest researcher</b> Leibniz University Hannover, Institute of Photogrammetry and GeoInformation (IPI), Germany
02/2016 – 12/2021	<b>Research assistant</b> Leibniz University Hannover, IPI, Germany
01/2016, 04/2013 – 06/2013, 10/2012 – 12/2012	<b>Student assistant</b> Leibniz University Hannover, IPI, Germany
06/2010 – 08/2010	<b>Compulsory work placement for the Bachelor's course in Geodesy and Geoinformatics</b> Surveying office Dipl.-Ing. Fiedler, Barsinghausen, Germany

## Education

10/2013 – 12/2015	<b>Course Geodesy and Geoinformatics</b> Leibniz University Hannover, Germany <i>Master of Science</i>
10/2010 – 09/2013	<b>Course Geodesy and Geoinformatics</b> Leibniz University Hannover, Germany <i>Bachelor of Science</i>
07/2009 – 03/2010	<b>Basic military service</b>
2002 – 2009	<b>High School</b> <i>A-levels</i>



# Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover

*(Eine vollständige Liste der Wiss. Arb. ist beim Geodätischen Institut, Nienburger Str. 1, 30167 Hannover erhältlich.)*

- Nr. 356 BLOTT, Gregor: Multi-View Person Re-Identification (Diss. 2020)  
Nr. 357 MAAS, Alina Elisabeth: Klassifikation multitemporaler Fernerkundungsdaten unter Verwendung fehlerbehafteter topographischer Daten (Diss. 2020)  
Nr. 358 NGUYEN, Uyen: 3D Pedestrian Tracking Using Neighbourhood Constraints (Diss. 2020)  
Nr. 359 KIELER, Birgit: Schema-Matching in räumlichen Datensätzen durch Zuordnung von Objektinstanzen (Diss. 2020)  
Nr. 360 PAUL, Andreas: Domänenadaption zur Klassifikation von Luftbildern (Diss. 2020)  
Nr. 361 UNGER, Jakob: Integrated Estimation of UAV Image Orientation with a Generalised Building Model (Diss. 2020)  
Nr. 362 COENEN, Max: Probabilistic Pose Estimation and 3D Reconstruction of Vehicles from Stereo Images (Diss. 2020)  
Nr. 363 GARCIAFERNANDEZ, Nicolas: Simulation Framework for Collaborative Navigation: Development - Analysis - Optimization (Diss. 2020)  
Nr. 364 VOGEL, Sören: Kalman Filtering with State Constraints Applied to Multi-sensor Systems and Georeferencing (Diss. 2020)  
Nr. 365 BOSTELMANN, Jonas: Systematische Bündelausgleichung großer photogrammetrischer Blöcke einer Zeilenkamera am Beispiel der HRSC-Daten (Diss. 2020)  
Nr. 366 OMIDALIZARANDI, Mohammad: Robust Deformation Monitoring of Bridge Structures Using MEMS Accelerometers and Image-Assisted Total Stations (Diss. 2020)  
Nr. 367 ALKHATIB, Hamza: Fortgeschrittene Methoden und Algorithmen für die computergestützte geodätische Datenanalyse (Habil. 2020)  
Nr. 368 DARUGNA, Francesco: Improving Smartphone-Based GNSS Positioning Using State Space Augmentation Techniques (Diss. 2021)  
Nr. 369 CHEN, Lin: Deep learning for feature based image matching (Diss. 2021)  
Nr. 370 DBOUK, Hani: Alternative Integrity Measures Based on Interval Analysis and Set Theory (Diss. 2021)  
Nr. 371 CHENG, Hao: Deep Learning of User Behavior in Shared Spaces (Diss. 2021)  
Nr. 372 MUNDT, Reinhard Walter: Schätzung von Boden- und Gebäudewertanteilen aus Kaufpreisen bebauter Grundstücke (Diss. 2021)  
Nr. 373 WANG, Xin: Robust and Fast Global Image Orientation (Diss. 2021)  
Nr. 374 REN, Le: GPS-based Precise Absolute and Relative Kinematic Orbit Determination of Swarm Satellites under Challenging Ionospheric Conditions (Diss. 2021)  
Nr. 375 XU, Wei: Automatic Calibration of Finite Element Analysis Based on Geometric Boundary Models from Terrestrial Laser Scanning (Diss. 2021)  
Nr. 376 FENG, Yu: Extraction of Flood and Precipitation Observations from opportunistic Volunteered Geographic Information (Diss. 2021)  
Nr. 377 YANG, Chun: A hierarchical deep learning framework for the verification of geospatial databases (Diss. 2021)  
Nr. 378 MEHLTRETTER, Max: Uncertainty Estimation for Dense Stereo Matching using Bayesian Deep Learning (Diss. 2021)  
Nr. 379 KAZIMI, Bashir: Self Supervised Learning for Detection of Archaeological Monuments in LiDAR Data (Diss. 2021)  
Nr. 380 PETERS, Torben: Learning Multi-View 2D to 3D Label Transfer for Semi-Supervised Semantic Segmentation of Point Clouds (Diss. 2022)  
Nr. 381 WASSINK, Martin: Kommunal- und Regionalentwicklung durch Kooperation und Teilung von Verantwortung in ländlichen Räumen - eine multiperspektivische Untersuchung an Beispielen aus dem Raum Steinwald/Fichtelgebirge (Diss. 2022)  
Nr. 382 GOLDSCHMIDT, Jürgen: Die Berücksichtigung künftiger Entwicklungen bei der Verkehrswertermittlung (Diss. 2022)  
Nr. 383 KRUSE, Christian: Impact maps from bomb craters detected in aerial wartime images using marked point processes (Diss. 2023)

