

# Deep Domain Adaptation for Detecting Bomb Craters in Aerial Images

Marco Geiger  
 Karlsruhe Institute  
 of Technology (KIT)  
[marco.geiger@alumni.kit.edu](mailto:marco.geiger@alumni.kit.edu)

Dominik Martin  
 Karlsruhe Institute  
 of Technology (KIT)  
[dominik.martin@kit.edu](mailto:dominik.martin@kit.edu)

Niklas Kühl  
 Karlsruhe Institute  
 of Technology (KIT)  
[niklas.kuehl@kit.edu](mailto:niklas.kuehl@kit.edu)

## Abstract

*The aftermath of air raids can still be seen for decades after the devastating events. Unexploded ordnance (UXO) is an immense danger to human life and the environment. Through the assessment of wartime images, experts can infer the occurrence of a dud. The current manual analysis process is expensive and time-consuming, thus automated detection of bomb craters by using deep learning is a promising way to improve the UXO disposal process. However, these methods require a large amount of manually labeled training data. This work leverages domain adaptation with moon surface images to address the problem of automated bomb crater detection with deep learning under the constraint of limited training data. This paper contributes to both academia and practice (1) by providing a solution approach for automated bomb crater detection with limited training data and (2) by demonstrating the usability and associated challenges of using synthetic images for domain adaptation.*

## 1. Introduction

For Germany in particular, unexploded bombs from World War II air raids are a problem that is still present. It is estimated that 10-20 % of allied aerial bombs have failed to detonate (Harrabi, 2019). Today, tens of thousands of duds lie in the ground in Germany, posing a significant threat to the environment, buildings, animals and ultimately human lives (Dittrich, 2021). However, unexploded ordnance (UXO) is not systematically searched for but only on a case-by-case basis. Each time a building application is submitted, a visual inspection of historical war images is carried out. By looking at aerial images taken from surveillance planes after an air raid, skilled experts assess whether a bomb is potentially unexploded in the area at hand (Dittrich, 2021). Since the analysis of the historical imagery can only be done by skilled staff and is mentally demanding, i.e., it cannot be carried out for multiple consecutive hours, the entire process is costly (Mayer,

2021).

Prior research suggests using deep learning (DL), specifically convolutional neural networks (CNNs) to detect bomb craters on historical aerial imagery to partially automate or assist the task (Clermont et al., 2019). CNNs trained on a large number of labeled images produce visual representations that may be utilized for the automatic detection of bomb craters. However, obtaining labeled data is a time-consuming and expensive procedure (Murez et al., 2018). Qualified experts must identify and label the craters to create suitable training samples. Therefore, this paper challenges an approach to detect bomb craters in aerial wartime images with mitigating the limited training data issue by domain adaptation (DA) techniques creating synthetic training data through input data from other domains.

Three research questions are derived from the problem statement and motivation:

- RQ1:** To what extent can DL support the UXO disposal process?
- RQ2:** How can the problem of missing training data in bomb crater detection be solved?
- RQ3:** Can DA be used to improve the detection of craters in the UXO disposal process?

Generally, the present work aims to evaluate to what extent DL can support the UXO disposal process and whether the problem of limited training data can be mitigated by applying DA techniques.

## 2. Research Design

We follow the Design Science Research (DSR) paradigm (Gregor & Jones, 2007; Peffers et al., 2007), which aims to guide the creation of innovative artifacts solving practical problems with a sound methodological framework. Existing theory can serve as justificatory knowledge supporting the design and rigorous construction and evaluation of the artifact (Gregor & Jones, 2007). To ensure this, we base our research on the three-cycle view proposed by Hevner

(2007). Thus, the goal of this article is to create an artifact (design cycle) that solves the introduced problem of bomb crater detection with limited training data (relevance cycle) while drawing on justificatory knowledge from theory (rigor cycle).

To structure the application of DSR, we guide our design through the six-step process of Peffers et al. (2006). The process starts with the *identification of the problem*, which is outlined in Section 1 and is part of the relevance cycle. Subsequently, the *objective definition* is based on relevant scientific foundations from literature (cf. Section 3) and thus is part of the rigor cycle. We start the *design & development* (design cycle) step with a conceptual design suggestion leveraging synthetic images of bomb craters based on moon craters. Next, the design suggestion is instantiated by means of an artifact in the form of a model (Peffers et al., 2012).

The artifact's evaluation is described in two evaluation episodes according to Venable et al. (2016). In a first artificial episode, a technical experiment is carried out (cf. Section 5.3), using four different datasets for training the crater detection model. In a second episode (cf. Section 6), a more naturalistic evaluation is conducted considering image quality issues and a combination of different models for better inspection quality in practice.

On the one hand, this research contributes to the application domain by evaluating the potential of DL and DA to improve crater detection in the UXO disposal process under the constraint of limited training data availability. In particular, the use of synthetic data based on similar geological artifacts (i.e., lunar craters) using DA is a novelty. On the other hand, the present work provides insights into the challenges and caveats of DA for synthetic image creation, thus, contributing to DL and DA research. Today, the practical use of DA for real-world problems is still insufficient due to the novelty of the research area—but promises enormous potential. Thus, this article represents a specific instance that aims to contribute to the applicability of DA in general.

### 3. Theoretical Foundations

A *Generative Adversarial Network (GAN)* is a learning algorithm that was initially proposed by Goodfellow et al. (2014) and is characterized by training a pair of neural networks (generator and discriminator) in an adversarial fashion. The goal of the generator is to learn the distribution of real data, whereas the discriminator's goal is to accurately identify whether the input data is drawn from real or artificial data produced by the generator. Generator  $G$  generates samples  $G(z)$ ,

where  $z$  is a random variable. Discriminator  $D$  receives real samples  $x$  as well as transformed data  $G(z)$  and maps the given received data to a probability that the input is from real data  $x$  ( $P = 1$ ) or from  $G(z)$  ( $P = 0$ ). Thus the purpose of  $D$  is to make correct classifications on the given input ( $D(x, G(z)) \rightarrow (0, 1)$ ), while  $G$  tries to make the performance on the generated data ( $G(z)$ ) consistent with the performance on real data  $D(x)$  (Creswell et al., 2018). The discriminator will be maximally confused if the generator distribution  $p_g(x)$  can completely match the real data distribution  $p_{data}$ .

*Image-to-Image Translation (I2I)* is the task of transferring images from a particular source domain  $X$  to a target domain  $Y$  while preserving the content representation (M. Y. Liu et al., 2017; Pang et al., 2021). In the unsupervised case, two independent sets of images exist and no paired examples. Consider the input images  $x$  from source domain  $X$  and the images  $y$  from target domain  $Y$ . To carry out the I2I translation, a mapping  $G_{X \rightarrow Y}$  is needed to generate image  $x \in X$  that cannot be distinguished from a target domain image  $y \in Y$  given a source image  $x \in X$  (Pang et al., 2021).

*CycleGAN* is an architecture that extends the idea of using generative models for unsupervised I2I. The mapping  $G : X \rightarrow Y$  for the I2I translation does not imply that an individual input  $x$  and output  $\hat{y}$  can be associated (Goodfellow et al., 2014; Murez et al., 2018; Zhu et al., 2017). For an image, this would mean that the I2I translation does not preserve the visual structure of the original image (Creswell et al., 2018). Therefore *cycle consistency* is introduced to the translation process. This means that an input image  $x$  is transformed to the target domain and generates an image  $\hat{y}$ , which generates the original image  $x$  when transformed back into the source domain. To do this, CycleGAN uses two translators  $G : X \rightarrow Y$  and  $F : Y \rightarrow X$ , which are inverse of each other and their mappings are bijections. To ensure that an individual input  $x_i$  is mapped to a desired output  $y_i$ , the space of possible mapping functions is reduced by introducing a *cycle consistency loss* (Zhou et al., 2016). The loss encourages  $F(G(x)) \approx x$  and  $G(F(y)) \approx y$ .

*You only look once (YOLO)* is a supervised DL algorithm for object detection. YOLO is a one-stage object detection approach that treats the detection problem as a continuous regression from the input image to the final detections. YOLO uses a CNN to extract features from the input image. The resulting feature maps are used in the final layers to determine the positions (bounding boxes) and classes of objects. Since YOLO is a supervised approach, the training requires labeled data (Redmon et al., 2016).

## 4. Related Work

Related research work can be divided into two types. On the one hand, related research pursues the same goal of crater detection, and on the other hand, related work uses the same DA approach.

Concerning the goal of crater detection, the most common applications and algorithms exist in geoscience and are concerned with exploring the geology of the moon and mars. Many image data approaches use circular pattern recognition (Salamunićcar et al., 2011), while others try to exploit shadow regions of craters (Urbach & Stepinski, 2009). In addition, Jung et al. (2005) approach the detection of craters with a multi-stage template matching approach. Other research tries to detect craters directly from Digital Elevation Model (DEM) information (Y. Wang & Wu, 2019). The criticism of the presented methods is that high accuracy of crater detection can only be achieved on specific data, certain planetary surfaces and for certain craters (Stepinski et al., 2012).

Recently Machine Learning approaches, especially CNNs, have shown outstanding results on the task at hand. The research work based on CNN use the networks for either classification of craters (Cohen et al., 2016), crater detection (Benedix et al., 2018; Emami et al., 2018) or crater segmentation (Delatte et al., 2019; Lee, 2019; Silburt et al., 2019). Finkelstein et al. (2019) and Silburt et al. (2019) show that a transfer of a trained model on lunar craters also yields good results when applied to martian craters.

There is far less research on detecting bomb craters in aerial war images. Merler et al. (2005) and Brenner et al. (2018) use a sliding window approach to create search windows for crater candidates, while the latter use a CNN to classify crater candidates and the first use a variant of AdaBoost. In contrast to the research presented above, Kruse et al. (2018) do not aim to provide a complete detection of bomb craters but rather impact maps, indicating the contamination of a specific area. Clermont et al. (2019) use a blob detector to select crater candidates classified by a CNN. The results show an incompleteness in automated crater detection for several datasets.

Regarding Domain Adaptation, a distinction can be made between shallow and deep DA methods (M. Wang & Deng, 2018). The most prevalent shallow DA algorithms can be divided into two groups: instance-based DA (Bruzzone & Marconcini, 2010; Chu et al., 2017) and feature-based DA (Gheisari & Baghshah, 2015; Long et al., 2013). The first class reduces the disparity by reweighing the source samples, whereas the second class learns a common shared

space in which the distributions of the two datasets are matched (M. Wang & Deng, 2018). Recently, neural networks for DA have shown significant improvements.

These so-called deep DA methods use neural networks and can be categorized into methods using feature adoption and methods using generative models (You et al., 2019). Feature adoption techniques use several approaches, for example, minimizing the maximum mean discrepancy of deep features across domains (Long et al., 2015). The research works of Bousmalis et al. (2016), Ganin et al. (2017), and Hoffman et al. (2017) all use an adversarial learning paradigm to train a domain classifier that differentiates features from source and target domains and pushes the feature extractor to confuse that domain classifier. Besides the feature adoption methods presented, generative methods based on GANs show impressive results by creating synthetic training samples (Bousmalis et al., 2016; Huang et al., 2018; Y. C. Liu et al., 2018; Murez et al., 2018; Sankaranarayanan et al., 2018). These generative methods build on I2I translation networks such as CycleGAN (Zhu et al., 2017), AugGAN (Huang et al., 2018) and DiscoGAN (Kim et al., 2017).

The CycleGAN approach, in particular, serves as a basis for the further development of DA techniques, such as CyCada (Hoffman et al., 2018) or SBADA-GAN (Russo et al., 2018). The research of Schutera et al. (Schutera et al., 2021), Lin (Lin, 2019) and Arruda et al. (Arruda et al., 2019) shows improved object detection results for autonomous driving based on generative models that translate daylight to night images.

## 5. Design Cycle

As a first design step, we suggest a tentative design in the form of a conceptual framework based on identified requirements (cf. relevance cycle) and foundational methods (cf. rigor cycle). Figure 1 depicts the conceptualized artifact steps. The presented artifact creates synthetic images of bomb craters based on moon crater images. It adapts the source domain moon crater to the target domain bomb crater in order to create training data, respectively improving a bomb crater detection model.

The concept envisages training an *unpaired I2I translation* model that is able to transform images of the source domain moon crater to the visual appearance of the target domain bomb crater. The output is a synthetic image dataset whose images have the same contours and shapes as the source lunar crater images. Therefore, in the *annotation transfer* step, the labels are projected onto the artificially generated images. Subsequently, in

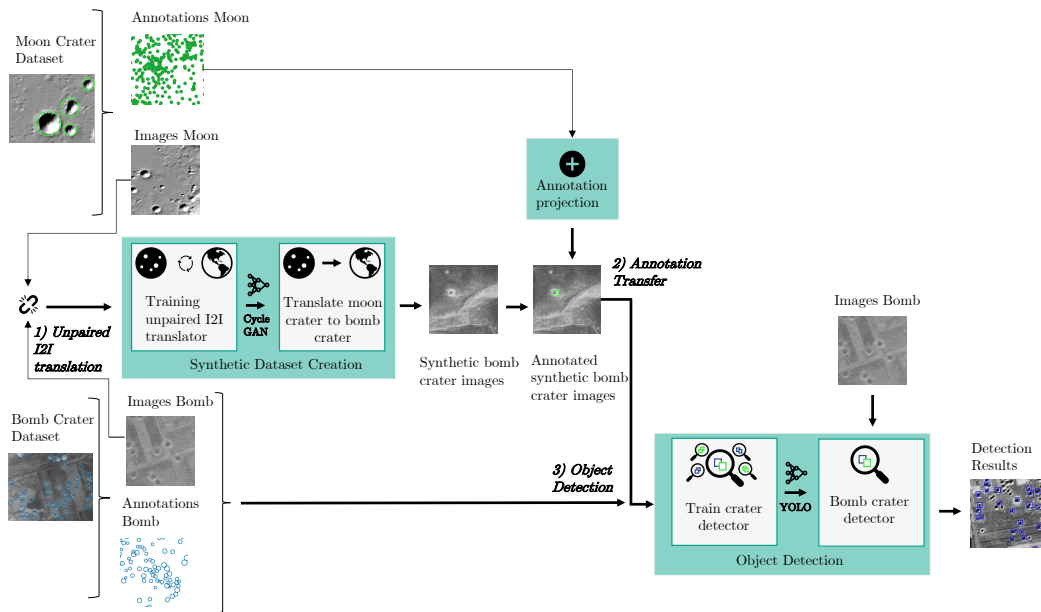


Figure 1. Concept of the design artifact

the last step *object detection*, the actual bomb crater dataset and the synthetic dataset are used to evaluate the artifact. This is carried out by training several crater detection models and comparing them based on their detection quality on unseen test images.

### 5.1. Dataset Analysis and Preparation

To instantiate and evaluate the suggested artifact, we use real-world bomb crater images (cf. Figure 2; dataset  $T$  representing the target domain), and images of the lunar surface (cf. Figure 3; dataset  $S$  representing the source domain). The aerial images  $t \in T$  were taken during surveillance flights after allied air raids during WWII. Domain experts from Lower Saxony’s explosive ordnance disposal service labeled the images manually to create the annotation set  $C$ .

Dataset  $S$  is created from a single, geo-referenced image of the moon surface (NASA, 2021b). As annotations for  $S$ , we introduce set  $M$ , derived from the moon crater database v1 Robbins (NASA, 2021a), which contains around 1.3 million lunar crater positions (Robbins, 2019).

Several preparation steps are necessary to use the datasets mentioned above as input for the I2I translation and the training of an object detection model. These steps are shown in Table 1. For the image data, this includes a Region of Interest (ROI) cropping, mitigation of noise (e.g., clouds, grey and black shadows) through contrast limited adaptive histogram equalization (Pizer et al., 1987) and a tiling algorithm (Unel et al., 2019)



Figure 2. Exemplary bomb crater image  $t_i$

to generate suitable input sizes for feeding the data into a neural network. For the label data, the processing includes transforming the global coordinates onto the generated image tiles and selecting the lunar crater labels based on diameter and visual appearance using a pixel-based heuristic. The final results are the processed datasets  $S'$ ,  $T'$ ,  $M'$  and  $C'$ , which are used to create synthetic images of the target domain bomb crater.

### 5.2. Synthetic Images Creation

We use the cycleGAN (Zhu et al., 2017), to carry out an I2I translation. The goal is to create a cycle consistent mapping from the moon crater source domain to the target bomb crater domain  $G : S' \rightarrow T'$  such that the synthetic output image  $\hat{t} = G(s')$ ,  $s' \in S'$  cannot

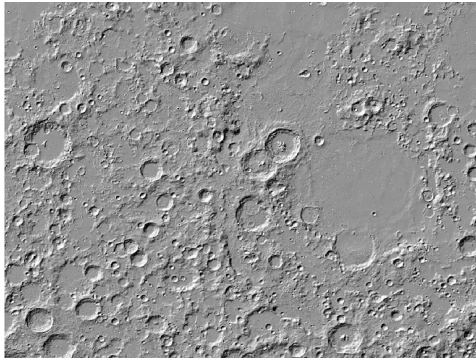


Figure 3. Exemplary moon crater image  $s_i$

Input	Processing Steps	Output
$S$	Tiling	$S'$
$T$	ROI Crop; Quality Improvement; Tiling	$T'$
$M$	Transform coordinates; Filter craters based on diameter; Prune crater set pixel-based; Project labels	$M'$
$C$	Transform coordinates; Project labels	$C'$

Table 1. Data Preparation

be distinguished from the original bomb crater image  $t' \in T'$ .

According to the DSR framework, we develop the artefact in an iterative approach. We analyze our findings during the development process and try to adapt the process to improve our results. As there are no meaningful quantitative metrics for assessing a trained cycleGAN model (Zhu et al., 2017), we visually analyze the produced synthetic images. We achieve significantly better results by adapting hyperparameters (e.g., batch size, epochs, image size), but especially by filtering the input data based on image similarity and present distortions. Moreover, it became apparent that the craters of the transformed images  $\hat{t}' \in \hat{T}'$  look better if the target domain images constantly show craters, not only landscape.

After generating the artificial dataset  $\hat{T}'$ , the annotations  $M'$  are projected onto the synthetic images. Unfortunately, despite the preprocessing of set  $M$ , the annotations still have inaccuracies. This is primarily due to the fact that previously visible craters of the original images in  $S'$  are no longer recognizable as craters after the I2I translation. Figure 5 shows good and bad examples from set  $\hat{T}'$  with corresponding annotations from  $M'$ .

### 5.3. Technical Experiments

To evaluate the artifact, a technical experiment (Peppers et al., 2012) is carried out, which assesses the quality improvement of automated crater detection by

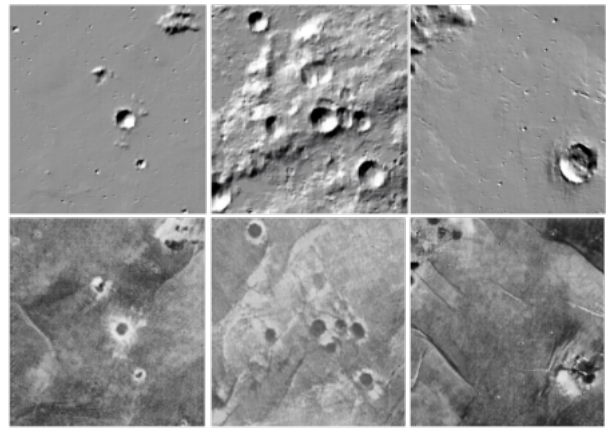


Figure 4. I2I translations from moon (upper row) to bomb crater domain (lower row)

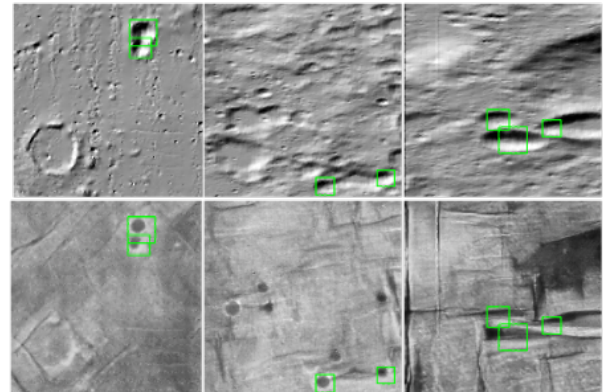


Figure 5. Annotations of  $M'$  projected on tiles of  $\hat{T}'$ . First column shows good examples; second column craters without annotations; third column the opposite case.

leveraging synthetic image data. The object detection network ScaledYolov4 (C.-Y. Wang et al., 2021) is trained on the bomb crater data dataset  $T'$  with annotations  $C'$  to serve as a reference. To compare this base case scenario against our proposed approach, we pose three hypotheses to show the effect of DA:

- H1:** A model trained on moon crater images leads to equally good detection results as a model trained on real bomb crater images.
- H2:** A model trained on synthetic bomb crater images leads to equally good detection results as a model trained on real bomb crater images.
- H3:** Adding synthetic bomb crater images to the original bomb crater training data leads to improved detection results.

The dataset  $T'$  is split into three subsets, a training, a validation and a test set. Table 2 shows an overview of

the models based on hypotheses H1-3.

Model	Training set	Validation set	Test set
<i>bomb</i>	$T'_{train}$		
<i>moon</i>	$S'$	$T'_{val}$	$T'_{test}$
<i>synthetic</i>	$\hat{T}'$		
<i>combined</i>	$T'_{train} \& \hat{T}'$		

**Table 2. Overview of models and used datasets**

We use common metrics for evaluating the models, particularly the *precision*, *recall* and the  $mAP_{0.5}$  (Padilla et al., 2020) metrics. The  $mAP_{0.5}$  uses the average precision calculated for the 0.5 IoU threshold. For the crater detection task at hand,  $mAP_{0.5}$  is considered sufficient because a detection with  $IoU > 0.5$  is accurate enough to indicate the presence of a crater, which human experts can use in the UXO disposal process to classify duds (Kruse et al., 2018).

## 5.4. Evaluation

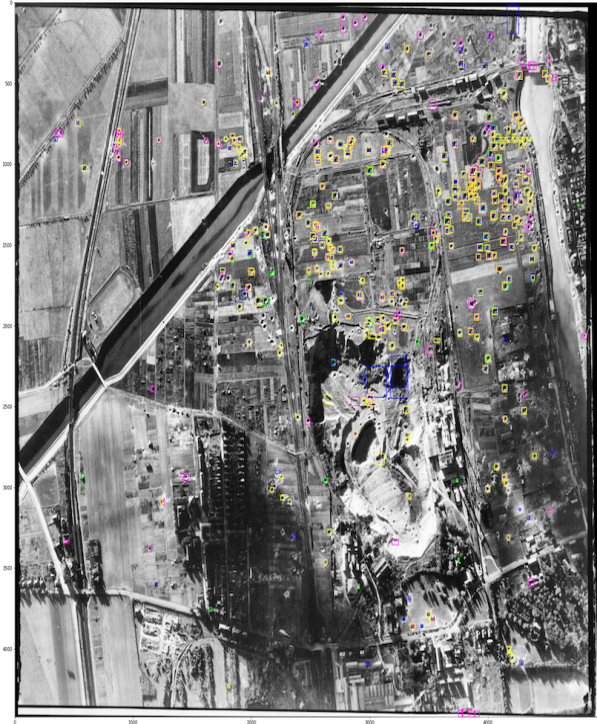
Table 3 illustrates the introduced metrics for the different models depicted in Table 2. Looking at the different train losses as well as the  $mAP_{train}$  scores, it can be seen that all models achieve solid performance on their respective training data. However, the validation losses and the  $mAP_{val}$  scores show a significantly worse performance on the validation set  $T'_{val}$ . Looking at the metrics  $mAP_{val}$  and  $mAP_{test}$ , it can be seen that there is a poor performance on the bomb crater data. The model detects no craters in aerial wartime images, thus, hypothesis H1 can be falsified. Moreover, hypothesis H2 can be rejected since the synthetically generated dataset  $\hat{T}'$  leads to a worse  $mAP_{test}$  compared to the  $model_{bomb}$ . Hypothesis H3 cannot be verified or falsified with certainty based on the quantitative metrics. The  $mAP_{test}$  is higher for the  $model_{combined}$  than for the  $model_{bomb}$ , but it is rather a slight improvement.

Metric	<i>bomb</i>	<i>moon</i>	<i>synthetic</i>	<i>combined</i>
<i>train loss</i>	1.645	3.098	2.312	1.411
<i>validation loss</i>	5.589	6.767	5.698	5.136
$mAP_{train}$	0.717	0.93	0.837	0.876
$mAP_{val}$	0.245	0	0.046	0.239
$mAP_{test}$	0.099	0	0.043	0.103
<i>precision<sub>test</sub></i>	0.188	0	0.083	0.191
<i>recall<sub>test</sub></i>	0.288	0	0.164	0.278

**Table 3. Performance metrics**

## 6. Discussion

Overall, the quantitative metrics in Table 3 indicate a poor detection quality of all models on the available bomb crater images. There seems to be strong overfitting for each model, which could be due to the



**Figure 6. Detections of different models as colored bounding boxes**

high complexity, i.e., the high number of layers of ScaledYOLOv4 (Bejani & Ghatee, 2021) or the fact that the amount of data generally seems sufficient, but the images often show the same geographical areas, thus, the same impact craters. However, these quantitative metrics cannot be considered a single source of truth.

Visual inspection of the test dataset  $T'_{test}$  reveals that some recognizable bomb craters are not labeled. Moreover, the test set, in particular contains images of poor quality, which stems from geographical clusters of the bomb crater imagery, limiting the training, validation, and test data selection. These constraints are limitations of the present work and impair the reliability of the quantitative evaluation. Therefore, a visual analysis of the detection results is carried out to generate conclusive findings. Figure 6 shows an image that is not included in any of the introduced datasets and can serve as a basis for qualitative analysis. Crater detections for each trained model are drawn on the image and indicated by colored bounding boxes. Green stands for  $model_{moon}$ , blue for  $model_{bomb}$ , pink for  $model_{synthetic}$  and yellow for  $model_{combined}$ .

The  $model_{moon}$  does not show good detection results both qualitatively and quantitatively. This means that a direct transfer of a crater detection model from the *source domain* moon craters to the *target domain* bomb

craters is not possible and there is a need for applying DA.

Concerning the synthetic data, visual observation reveals a high number of False-Positive (FP) detections from the  $model_{synthetic}$ , indicated by the significantly lower  $precision_{test}$  value compared to the  $model_{bomb}$ . The  $model_{synthetic}$  often detects trees, houses or other objects as craters since these represent a collection of darker pixels surrounded by brighter pixels. This likely happens because the synthetic images used for training mainly show craters and no other landscape objects. It seems that the model does not explicitly learn the difference between these objects in the training process. Generally, it can be observed that the performance of the  $model_{synthetic}$  is significantly worse compared to the  $model_{bomb}$  even in the  $recall_{test}$  metric, which means that no equivalent detection quality is achieved even if the FP detections are ignored. However, FP detections are not as problematic in the overall context of the UXO disposal process, since the human experts could quickly identify erroneously detected craters visually.

The barely improved performance of the  $model_{combined}$  compared to the  $model_{bomb}$  seems unusual at first since the visual similarity of the synthetic images to the original bomb crater images is given, as shown in Figure 4. Due to the more extensive training set and higher variance of craters, improved results were expected. Reasons for the lack of this improvement could be an insufficient annotation quality in  $M'$  or still an inadequate quality of the fake images in  $\hat{T}'$  as the quality could only be assessed visually.

Although synthetically generated training data does not significantly affect the quality of object detection, it is evident that the models perform differently in different image regions. Looking at Figure 6, the upper edge of the image is particularly striking, as the  $model_{synthetic}$  (indicated in pink) detects craters that the other models do not recognize. In addition, the sandpit in the middle of the image is conspicuous, as only the  $model_{synthetic}$  and the  $model_{combined}$  produce suitable detections. This suggests that models trained with synthetic images are superior in detecting craters with a brighter background. Moreover, the partially low metrics originate from the existence of crater fields on aerial wartime images. It frequently happens that the trained models do not detect all craters of a crater field. However, this is not problematic in practical use since humans can quickly recognize the entire crater field based on the detected craters of the automated detection. The detection of each individual crater is not necessarily decisive to improve the process (Kruse et al., 2018).

Furthermore, other advantages arise despite the low detection quality. Detecting outlier craters within the

UXO disposal process is especially difficult for human experts. Especially for these crater types, automatic crater detection may offer added value, although the overall detection performance is relatively low. While the quality of crater detection, even with the addition of synthetically generated data, seems insufficient for the problem of automatic crater detection presented in Section 1, the combination of detection results from the different models might be a promising approach to provide overall good detection results. Especially because images  $t'_i \in T'_{test}$  partially overlap and show the same geographical area, so that multiple models can use different images for detecting a specific crater. This is a meaningful finding of the conducted experiment and suggests the application of ensemble techniques (Murphy, 2012) to generate a higher overall crater detection quality. Therefore, even a moderate detection quality may have an economic benefit in terms of accelerating the process.

## 7. Conclusion and Outlook

This research used the DSR methodology to address the problem of automated bomb crater detection with limited training data in the UXO disposal process. For this purpose, a DA approach was conceptualized to use publicly available moon craters for detecting bomb craters. More specifically, an artifact based on state-of-the-art GAN research was created to generate synthetic training data for an object detection network. This artifact was evaluated by an experiment investigating the detection quality of several bomb crater detection models trained on different datasets.

Regarding  $RQ1$ , the experiment showed that a fully automated detection of bomb craters is not possible due to insufficient detection quality. This finding is in line with the recommendations of previous research (Clermont et al., 2019). However, even moderate detection quality can support the manual UXO detection task. Concerning  $RQ2$ , it can be stated that missing training data of the bomb crater domain can be generated synthetically with visually good quality. However, this involves considerable effort. Furthermore, it is shown that synthetically generated bomb crater data cannot equally replace original images. Finally, with regard to  $RQ3$ , it can be said that the use of DA does not directly improve a single bomb crater detection model. However, combining the detections from several models achieves better results.

This work highlights challenges and requirements when using DA based on a practical use case. It should be emphasized that I2I translation in combination with object detection has difficulties regarding the consistent

use of labels, because small objects may disappear. This issue is accompanied by increased efforts in preparing datasets, as source datasets usually cannot be used for DA without further adjustments.

Since the application of DA through synthetic data generation is still at an early stage of research, more practical cases should be contributed to generalize the generated findings. Moreover, it is crucial to develop criteria to assess when datasets or practical use cases are suitable for DA. In general, indicators need to be developed for assessing which type of DA is suitable for a particular task or domain.

## References

- Arruda, V. F., Paixao, T. M., Berriel, R. F., De Souza, A. F., Badue, C., Sebe, N., & Oliveira-Santos, T. (2019). Cross-Domain Car Detection Using Unsupervised Image-to-Image Translation: From Day to Night. *Proceedings of the International Joint Conference on Neural Networks*.
- Bejani, M. M., & Ghatee, M. (2021). A systematic review on overfitting control in shallow and deep neural networks. *Artificial Intelligence Review*, 54(8), 6391–6438.
- Benedix, N. G. K., Norman, C. J., Bland, P. A., Towner, M. C., Paxman, J., & Tan, T. (2018). Automated Detection of Martian Craters using a Convolutional Neural Network. *49th Lunar and Planetary Science Conference 2018*.
- Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D., & Erhan, D. (2016). Domain separation networks. *Advances in Neural Information Processing Systems*, 343–351.
- Brenner, S., Zambanini, S., & Sablatnig, R. (2018). Detection of Bomb Craters in WWII Aerial Images. *Proceedings of the OAGMWorkshop 2018*, 94–97.
- Bruzzone, L., & Marconcini, M. (2010). Domain adaptation problems: A DASVM classification technique and a circular validation strategy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5), 770–787.
- Chu, W. S., De La Torre, F., & Cohn, J. F. (2017). Selective transfer machine for personalized facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(3), 529–545.
- Clermont, D., Kruse, C., Rottensteiner, F., & Heipke, C. (2019). Supervised detection of bomb craters in historical aerial images using convolutional neural networks. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(2/W16), 67–74.
- Cohen, J. P., Lo, H. Z., Lu, T., & Ding, W. (2016). Crater Detection via Convolutional Neural Networks. *47th Lunar and Planetary Science Conference (2016)*.
- Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, 35(1), 53–65.
- Delatte, D. M., Crites, S. T., Guttenberg, N., Tasker, E. J., & 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.
- Dittrich, M. (2021). Weltkriegserbe - Munition und Blindgänger in Meer und Boden [last accessed: 2021-05-24]. <https://www.deutschlandfunk.de/weltkriegserbe-munition-und-blindgaenger-in-meer-und-boden.724.de>
- Emami, E., Ahmad, T., Bebis, G., Nefian, A., & Fong, T. (2018). Lunar Crater Detection via Region-based Convolutional Neural Networks. *49th Lunar and Planetary Science Conference 2018*.
- Finkelstein, M., Baby, S. M., & Kitano, H. (2019). *Automatic Lunar Crater Detection from Optical Images and Elevation Maps*.
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., & Lempitsky, V. (2017). Domain-adversarial training of neural networks. *Advances in computer vision and pattern recognition* (pp. 189–209). Springer London.
- Gheisari, M., & Baghshah, M. S. (2015). Unsupervised domain adaptation via representation learning and adaptive classifier learning. *Neurocomputing*, 165, 300–311.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 3(January), 2672–2680.
- Gregor, S., & Jones, D. (2007). The Anatomy of a Design Theory. *Journal of the Association of Information Systems*, 8, 312–335.
- Harrabi, K. (2019). Wissenschaft vs. Tatort: Wie viele Blindgänger liegen noch im Boden? [last accessed: 2021-05-24]. <https://www.faz.net/aktuell/karriere-hochschule/wissenschaft-vs->

- tatort - wie - viele - blindgaenger - liegen - noch - im - boden - 16112973.html
- Hevner, A. R. (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, 19(2).
- Hoffman, J., Tzeng, E., Darrell, T., & Saenko, K. (2017). Simultaneous deep transfer across domains and tasks. *Advances in computer vision and pattern recognition* (pp. 173–187). Springer London.
- Hoffman, J., Tzeng, E., Park, T., Zhu, J. Y., Isola, P., Saenko, K., Efros, A. A., & Darrell, T. (2018). CyCADA: Cycle-Consistent Adversarial Domain adaptation. *35th International Conference on Machine Learning, ICML 2018*, 5, 3162–3174.
- Huang, S. W., Lin, C. T., Chen, S. P., Wu, Y. Y., Hsu, P. H., & Lai, S. H. (2018). AugGAN: Cross domain adaptation with GAN-based data augmentation. *Lecture Notes in Computer Science*, 11213, 731–744.
- Jung, R. K., Muller, J. P., Van Gassel, S., Morley, J. G., & 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., Cha, M., Kim, H., Lee, J. K., & Kim, J. (2017). Learning to discover cross-domain relations with generative adversarial networks. *34th International Conference on Machine Learning, ICML 2017*, 4, 2941–2949.
- Kruse, C., Rottensteiner, F., Hoberg, T., Ziem, 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 the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4(3), 127–134.
- Lee, C. (2019). Automated crater detection on Mars using deep learning. *Planetary and Space Science*, 170, 16–28.
- Lin, C. T. (2019). Cross Domain Adaptation for on-Road Object Detection Using Multimodal Structure-Consistent Image-to-Image Translation. *Proceedings - International Conference on Image Processing, ICIP*, 3029–3030.
- Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in Neural Information Processing Systems*, 701–709.
- Liu, Y. C., Yeh, Y. Y., Fu, T. C., Wang, S. D., Chiu, W. C., & Wang, Y. C. F. (2018). Detach and Adapt: Learning Cross-Domain Disentangled Deep Representation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8867–8876.
- Long, M., Cao, Y., Wang, J., & Jordan, M. I. (2015). Learning transferable features with deep adaptation networks. *32nd International Conference on Machine Learning, ICML 2015*, 1, 97–105.
- Long, M., Wang, J., Ding, G., Sun, J., & Yu, P. S. (2013). Transfer feature learning with joint distribution adaptation. *Proceedings of the IEEE International Conference on Computer Vision*, 2200–2207.
- Mayer, Y. (2021). So funktioniert's: Eine Bombe entschärfen [last accessed: 2021-05-24]. <https://www.br.de/wissen/bombe-fliegerbombe-entschaerfen-weltkrieg-100.html>
- Merler, S., Furlanello, C., & Jurman, G. (2005). Machine learning on historic air photographs for mapping risk of unexploded bombs. *Lecture Notes in Computer Science*, 3617, 735–742.
- Murez, Z., Kolouri, S., Kriegman, D., Ramamoorthi, R., & Kim, K. (2018). Image to Image Translation for Domain Adaptation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4500–4509.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press.
- NASA. (2021a). Moon Crater Database v1 Robbins [last accessed: 2021-12-02]. [https://astrogeology.usgs.gov/search/map/Moon/Research/Craters/lunar\\_crater\\_database\\_robbins\\_2018](https://astrogeology.usgs.gov/search/map/Moon/Research/Craters/lunar_crater_database_robbins_2018)
- NASA. (2021b). Moon LRO LOLA Shaded Relief 237m v4 [last accessed: 2021-12-02]. [https://astrogeology.usgs.gov/search/map/Moon/LMMP/LOLA-derived/Lunar\\_LRO\\_LOLA\\_Shade\\_Global\\_128ppd\\_v0](https://astrogeology.usgs.gov/search/map/Moon/LMMP/LOLA-derived/Lunar_LRO_LOLA_Shade_Global_128ppd_v0)
- Padilla, R., Netto, S., & da Silva, E. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 237–242.
- Pang, Y., Lin, J., Qin, T., & Chen, Z. (2021). Image-to-Image Translation: Methods and Applications. *IEEE Transactions on Multimedia*.
- Peffers, K., Rothenberger, M., Tuunanen, T., & Vaezi, R. (2012). Design science research evaluation. *Lecture Notes in Computer Science*, 7286, 398–410.

- Peffer, K., Tuunanen, T., Gengler, C., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). The design science research process: A model for producing and presenting information systems research. *Proceedings of First International Conference on Design Science Research in Information Systems and Technology DESRIST*.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77.
- Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J. B., & Zuiderveld, K. (1987). Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3), 355–368.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 779–788.
- Robbins, S. J. (2019). A New Global Database of Lunar Impact Craters over 1 to 2 km: 1. Crater Locations and Sizes, Comparisons With Published Databases, and Global Analysis. *Journal of Geophysical Research: Planets*, 124(4), 871–892.
- Russo, P., Carlucci, F. M., Tommasi, T., & Caputo, B. (2018). From Source to Target and Back: Symmetric Bi-Directional Adaptive GAN. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8099–8108.
- Salamunićar, G., Lonarić, S., Pina, P., Bandeira, L., & Saraiva, J. (2011). MA130301GT catalogue of Martian impact craters and advanced evaluation of crater detection algorithms using diverse topography and image datasets. *Planetary and Space Science*, 59(1), 111–131.
- Sankaranarayanan, S., Balaji, Y., Castillo, C. D., & Chellappa, R. (2018). Generate to Adapt: Aligning Domains Using Generative Adversarial Networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8503–8512.
- Schutera, M., Hussein, M., Abhau, J., Mikut, R., & Reischl, M. (2021). Night-to-Day: Online Image-to-Image Translation for Object Detection Within Autonomous Driving by Night. *IEEE Transactions on Intelligent Vehicles*, 6(3), 480–489.
- Silburt, A., Ali-Dib, M., Zhu, C., Jackson, A., Valencia, D., Kissin, Y., Tamayo, D., & Menou, K. (2019). Lunar crater identification via deep learning. *Icarus*, 317, 27–38.
- Stepinski, T. F., Ding, W., & Vilalta, R. (2012). Detecting impact craters in planetary images using machine learning. *Intelligent data analysis for real-life applications: Theory and practice* (pp. 146–159). IGI Global.
- Unel, F. O., Ozkalayci, B. O., & Cigla, C. (2019). The power of tiling for small object detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 582–591.
- Urbach, E. R., & Stepinski, T. F. (2009). Automatic detection of sub-km craters in high resolution planetary images. *Planetary and Space Science*, 57(7), 880–887.
- Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, 25(1), 77–89.
- Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2021). Scaled-YOLOv4: Scaling Cross Stage Partial Network. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13024–13033.
- Wang, M., & Deng, W. (2018). Deep visual domain adaptation: A survey. *Neurocomputing*, 312, 135–153.
- 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.
- You, K., Long, M., Cao, Z., Wang, J., & Jordan, M. I. (2019). Universal domain adaptation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2715–2724.
- Zhou, T., Krähenbühl, P., Aubry, M., Huang, Q., & Efros, A. A. (2016). Learning dense correspondence via 3D-guided cycle consistency. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (1), 117–126.
- Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision*, 2242–2251.