Detection and mapping of artillery craters with very high spatial resolution satellite imagery and deep learning

Duncan, Erik C.; Skakun, Sergii; Kariryaa, Ankit; Prishchepov, Alexander V.

Published in:
Science of Remote Sensing

DOI:
10.1016/j.srs.2023.100092

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-ND

Citation for published version (APA):
Detection and mapping of artillery craters with very high spatial resolution satellite imagery and deep learning

Erik C. Duncan\textsuperscript{a,b,1}, Sergii Skakun\textsuperscript{a,c,7,1}, Ankit Kariryaa\textsuperscript{b,d}, Alexander V. Prishchepov\textsuperscript{b,1}

\textsuperscript{a} Department of Geographical Sciences, University of Maryland, College Park, MD, USA
\textsuperscript{b} Department of Geosciences and Natural Resource Management (IGN), University of Copenhagen, Copenhagen, Denmark
\textsuperscript{c} College of Information Studies (iSchool), University of Maryland, College Park, MD, USA
\textsuperscript{d} Department of Computer Science (DIKU), University of Copenhagen, Copenhagen, Denmark

**A R T I C L E   I N F O**

**Keywords:**
- Artillery crater
- Very high spatial resolution
- Satellite imagery
- WorldView-2
- Deep learning
- U-Net
- Agriculture
- Ukraine
- War
- Escalate

**A B S T R A C T**

Unexploded munitions are some of the most enduring remnants of conflicts around the world. Their effects on the economy, health, environment, and post-conflict rehabilitation are long reaching and devastating for the areas they plague. With the advancements in very high spatial resolution (VHR) satellite multispectral imaging at sub-meter resolution, it becomes possible to detect object attributes at the scale of individual impacts (craters) of heavy weapon shelling. Manual identification and delineation of artillery craters in satellite imagery is time and resource consuming, especially when large territories and volumes of VHR data are considered. Therefore, automatic image processing methods should be explored. Here, we evaluate the application of a deep learning approach for identifying and mapping artillery craters in agricultural fields in Eastern Ukraine during the onset of armed conflict in 2014. The model was applied to pansharpened multispectral VHR imagery acquired by the WorldView-2 satellite at 0.5-m spatial resolution. The model can detect artillery craters with producer’s accuracy (PA) or recall of 0.671 and user’s accuracy (UA) (or precision) of 0.392 in terms of crater area and shape, and PA of 0.559 and UA of 0.427 in terms of binary crater identification. The model’s performance is dependent on crater size. Reliability of crater detection and mapping improves as the size of craters increases. For example, for craters larger than 60 m\textsuperscript{2} PA is 0.803 and UA is 0.449 (per-pixel), and PA is 0.891 and UA is 0.721 (per-object).

1. Introduction

Armed conflicts have a paramount impact on land systems and human well-being (Daskin and Pringle, 2018; Kuemmerle and Baumann, 2021). Armed conflict footprints result in human losses, destruction of infrastructure and negative environmental implications (Pereira et al., 2022). Satellite remote sensing serves as a vital and feasible technology to monitor the consequences of armed conflicts on the environment on a regular basis (Mueller et al., 2021). For example, multispectral 30-m Landsat imagery (Wulder et al., 2022) was used to monitor cropland abandonment in armed conflict areas of Eastern Ukraine (Skakun et al., 2019, 2022), Nagorno-Karabakh (Baumann et al., 2015), Bosnia and Herzegovina (Witmer, 2008), and Syria (Li et al., 2022). Recent progress in multisource image fusion has made it possible to identify and map a conflict-driven cropland abandonment in smallholder farming areas of South Sudan (Olsen et al., 2021), and 30-m Sentinel-1 Synthetic Aperture Radar (SAR) imagery (Torres et al., 2012) enabled detection of declining paddy-rice agriculture in Rakhine, Myanmar (Huang et al., 2022). These authors share first authorship.

https://doi.org/10.1016/j.srs.2023.100092

Received 17 March 2023; Received in revised form 1 May 2023; Accepted 17 May 2023

Available online 25 May 2023

2666-0172/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
present when the shell hits the ground. A proxy for UXO locations must be used. The rationale of using shell craters as a proxy stems from the spatial characteristics of artillery fire, can be used as a proxy for highly dangerous when agriculture-related activities, e.g., plowing, are resumed on the fields. Therefore, the identification and mapping of artillery craters becomes an extremely important problem for post-conflict policies of restoration of the natural and human landscapes. Specifically, accurate and timely geoinformation products on crater density and location can be used in relief (de-mining) efforts to reduce the risk of explosions, which is critical for the rehabilitation of Ukraine in accordance with international humanitarian efforts. UXO mapping can also be used to identify damaged agricultural fields to estimate the impact of artillery shelling on agriculture production and in calculating the war reparations. It is important to note that the UXO themselves are likely impossible to detect even with VHR imagery, as there will be no crater present when the shell hits the ground. A proxy for UXO locations must be used. The rationale of using shell craters as a proxy stems from the patterns of unguided artillery usage, notably in Soviet and post-Soviet doctrine. Unguided artillery or mortar shells fired from a single weapon source salvo (that is multiple shots fired at the same target at the same time) will land in a concentrated landing zone. This zone is known as a Circular Error Probable (CEP). The size of this CEP depends on multiple factors, including the weapon type, barrel and munition quality, and weather conditions (Elder, 1986; Fann, 2006). In the case of Soviet unguided projectiles employed by both sides in Ukraine, it is usually between several hundred to several thousand square meters. Thus, the detection of the visible craters, combined with the known spatial characteristics of artillery fire, can be used as a proxy for highly accurate estimation of UXO locations.

The size of artillery craters can range from 10 to 30 m² to over 500 m² depending on the munition type, barrel, munition quality and weather conditions (Bellingcat Report, 2015). The detection and mapping of the full distribution of crater sizes requires the use of very high spatial resolution (VHR) satellite imagery. Manual identification and delineation of craters is time and resource consuming, especially when considering a full conflict area and time scale with numerous corresponding VHR images. Consequently, automatic image processing methods should be applied. The problem of automatic crater detection and mapping in remote sensing imagery has been addressed in previous studies (Cheng and Han, 2016), which can be divided into two categories: (i) detection of craters on foreign celestial bodies, e.g., Moon, Mars and asteroids; (ii) detection of artillery or bomb craters on the Earth that occurred in the past in historical or modern images. Most the works fall into the first category and utilize various imaging technologies and image processing methods for identifying craters (Salamunicar et al., 2011), with recent works focusing on deep learning methods and models (DeLatte et al., 2019; Lee, 2019; Silburt et al., 2019). For example, Lee (2019) trained a U-Net model for detecting craters on Mars using a digital terrain model (DTM) and reported ~75% accuracy of crater detection.

However, artillery and bomb craters in multispectral VHR imagery differ from craters on celestial objects. These differences include irregularities in crater shape, temporal changes, land cover variation, and obstruction from vegetation. Lin et al. (2020) explored various machine learning models and VHR imagery acquired by WorldView-2 satellite at 0.5-m spatial resolution for detecting and mapping craters in Cambodia as a result of bombings that occurred during the Vietnam War in 1970–1973. They found that specifically designed features using classical image processing methods such as histogram of oriented gradient (HOG) and scale-invariant feature transform (SIFT) outperformed models that used convolutional neural network (CNN)-derived features and yielded an overall F-score of 0.79. CNN is a type of artificial neural networks that consist of multiple layers of convolutional units with trainable weights that applied to input images along with max-pooling and activations (Fukushima, 1979; LeCun et al., 1989). However, no previous studies focused on detecting and mapping craters resulting from the use of contemporary artillery during ongoing armed conflicts.

Here, we explore the suitability of VHR satellite imagery at sub-meter spatial resolution and deep learning classification models for detecting and mapping artillery craters in Eastern Ukraine in 2014. We aim to address the following research questions: (i) What accuracy can be achieved when detecting and mapping artillery craters on an individual level with VHR satellite imagery? (ii) What is the efficiency of the deep-learning model compared to the baseline model based on random forest approach? (iii) How is the detection and mapping of craters dependent on crater size?

The rest of the paper is organized as follows. We first describe the study area, artillery crater characteristics, data used and methods in Section 2. We next describe the results on detecting craters in terms of numbers, area, and implications (Section 3 and Section 4). Lastly, we summarize our findings in the conclusion section and outline steps forward.

2. Materials and methods

2.1. Study area

The conflict chosen as an area of study are in Eastern Donetsk and Southern Luhansk oblasts of Ukraine, where battles occurred in the summer of 2014 (Fig. 1). Since 2014 (until the full-scale invasion in February 2022) these regions of the country experienced most of the armed conflict following the 2013 Revolution of Dignity (Davis, 2016; Ivanov, 2015). In July 2014, a large government offensive to retake seized territory brought about widespread heavy fighting. This heavy fighting and use of heavy weapons systems led to crater fields throughout the region (Bellingcat Report, 2015).

The characteristics of the conflict in Eastern Donetsk and Southern Luhansk oblasts – in particular its position as one of the first examples of contemporary heavy weaponry damage accessible to study with VHR satellite imagery – create a robust zone on which to focus this study. Study areas were selected based on known battle areas and VHR image availability for the summer and fall of 2014. Two regions were selected for imagery acquisition. Both regions cover the southwestern area of Luhansk through the easternmost part of Donetsk oblast (Fig. 1). The first area includes five known Ukrainian military encampments: one in Avmsosiivka and four others in the vicinity of Svyystuny, which were targets of heavy weapon attacks. The main battleground within this zone is centered around the village of Avmsosiivka. The second zone, located 4 km east of the first area, also contained two known Ukrainian military positions, specifically in Oleksivka and Marynivka. The main battleground in this zone was centered around the village of Stepanivka and extends south towards Marynivka and north to Snizhne. All villages
2.2. Weapon and craters

The target objects of detection are the craters caused by high-explosive rocket- or ranged artillery systems. The craters are visible in the true-color band combination from VHR imagery as dark brown, usually circular shapes often surrounded by ‘wings’ of displaced earth from the crater (Fig. 2). The craters take on a different appearance depending on the landscape type and season. Craters are more visible when they are present in actively growing crops and grasses, as the dirt contrasts the green or tan crops. In the winter and spring, craters often fill with snow and ice, creating an inverse effect of a bright shape surrounded by dark tilled earth surface. The most difficult conditions for crater detection in open areas without tree cover are in areas of muddy, churned or tilled ground, where the crater has a limited contrast with the surrounding area. Craters will also be difficult to detect if vegetation has grown from the churned earth surface or if snow and ice have occluded them. Similarly, the passing of time degrades craters, causing them to fade into the background landscape over months or years.

There are three classes of ranged artillery systems commonly used during the armed conflict which cause craters that are detectable in sub-meter satellite VHR imagery. They are multiple-launch rocket systems (MLRS) and field artillery consisting of static and self-propelled artillery. The MLRS consist of three main systems, though the first is by far the most commonly used: The BM-21 “Grad” which fires 122 mm warheads with a range of up to 40 km (more commonly 20–35 km); the BM-27 “Uragan” which fires 220 mm unguided rockets with a range of 35 km; and the BM-30 “Smerch” which fires 300 mm artillery rockets to a range of 70 or 90 km depending on the munition (An Examination of Arms, 2014). Each of these MLRS can be used to fire high-explosive, chemical, and cluster munitions. In the conflict, both high-explosive and cluster munitions have been widely used by all sides (Morley, 2014). The most common artillery systems fire 120, 122, 152, and rarely 203 mm unguided projectiles (An Examination of Arms, 2014; Explosive weapon effects, 2017). And in the third class, mortar systems have also been extensively used, though typically their craters are of a much smaller size than artillery, with the exception of the 240 mm heavy mortar (Organization for Security and Co, 2015).

2.3. Satellite imagery

Through the NASA Commercial Smallsat Data Acquisition (CSDA) Program (Neigh et al., 2013) we collected Maxar’s WorldView-2 (WV-2) images over the study area. WV-2 acquires multispectral (eight spectral bands) and panchromatic images of the earth surface at 1.85-m and 0.46-m nadir spatial resolution, respectively. The VHR images were acquired on September 5th, 2014. Two strips of satellite imagery covering 1,279 km² were used in the study (Fig. 1): the first consisted of 17 image tiles covering 421 km² (acquired at 08:16 UTC, viewing angle of 13.0°) and the second consisted of 36 tiles covering 858 km² (08:17 UTC, 18.8°). We used red (630–690 nm), green (510–589 nm), and blue (450–510 nm) spectral bands as the main input to detect craters. The acquired images were ortho-rectified using rational polynomial coefficients (RPC) provided by the data provider and digital elevation model (DEM) SRTM, reprojected to local UTM projection, and pixel values were converted to top-of-atmosphere (TOA) reflectance. We next applied a pan-sharpening procedure using a Brovey transformation (Gillespie et al., 1987) to convert multispectral images to the resolution of the panchromatic band, i.e., 0.46-m.

We used a 2015 tree canopy cover (TCC) product (version 4) derived from 30-m Landsat data (Sexton et al., 2013) to mask out areas with more than 15% tree cover. This was done to reduce detection to tree-free areas. Craters are undetectable with the proposed approach and optical imagery under heavy tree canopy cover.

2.4. Classification models

We explored and compared two classification models for detecting and mapping artillery craters from sub-meter satellite imagery. The first one is the U-Net model, which is based on the architecture used for mapping tree crowns in the VHR satellite imagery (Brandt et al., 2020). This classification was compared in a small experiment to a baseline model of image segmentation and subsequent classification with
random forest (RF) (Csillik, 2017).

2.4.1. U-Net model

U-Net is a type of CNN architecture which is popular for the semantic segmentation of input images (Ronneberger et al., 2015). It is specifically designed to produce a pixel-wise classification of an input image into a specified number of classes. U-Net consists of two main components: encoder and decoder. The encoder, through convolution and downsampling (max-pooling), non-linearity and normalization, transforms an input image into a set of robust low-level features, sometimes referred to as feature maps. The direct use of these features for classification (classical CNN) would result in the mismatch between the size and spatial resolution of the input and output images, as the latter would appear “blurred” because of multi-level downsampling (Kussul et al., 2017; Maggiori et al., 2016). Therefore, the goal of the decoder is to upscale feature maps to provide the output that would correspond spatially to the input image. This is performed through a series of upsampling, normalization, non-linearity and convolution operations. In addition, the decoder would concatenate feature map outputs from the corresponding level from the encoder.

The U-Net architecture used in this work is similar to the one used to detect tree crowns in the VHR satellite imagery (Brandt et al., 2020; Tucker et al., 2023). Each downsampling block of the encoder consists of two 2d convolution layers with the ReLu activation function followed by a batch normalization layer (Ioffe and Szegedy, 2015) and 2d max pooling layer. There were five downsampling blocks. Each upsampling block of the decoder consists of 2d upsampling layer with the nearest neighborhood algorithm followed by 2d convolution, batch normalization, concatenation of the outputs from the corresponding block from the encoder and two layers of 2d convolution with the ReLU activation function. 2d convolution and batch normalization layers have trainable parameters. Decoder output is fully connected to the output image map of the same size as the input one using the sigmoid activation function (for two classes: crater and no -crater) and L2 regularization (coefficient of regularization was set to $\lambda = 0.0001$). U-Net was trained using a Tversky loss function (Eq. (2)) (Salehi et al., 2017) and the Adadelta optimization algorithm (Zeller, 2012), which is based on the adaptive learning rate $\eta$ (Eq. (3)):

$$T_{\alpha}(y, \hat{y}) = \frac{TP}{TP + \alpha FP + \beta FN},$$

$$L_{\text{Tversky}}(y, \hat{y}) = 1 - T_{\alpha}(y, \hat{y}),$$

where $y$ and $\hat{y}$ are a reference and estimated outputs; $TP$, $FP$ and $FN$ are the number of true positives, false positives and false negatives, respectively; $\alpha$ and $\beta$ are Tversky loss parameters, and

$$\eta_{t+1} = \eta_t (f)^{\beta},$$

where $\eta_t$ is the learning rate at epoch $t$ and $f$ is the scaling factor for updating the learning rate. We used the initial value for $\eta_0 = 1.0$ and $f = 0.33$.

We set parameters $\alpha = 0.6$ and $\beta = 0.4$ in the Tversky loss function (Eqs. (1) and (2)) after multiple trainings and visual inspection to make sure that omission of craters was low.

The input image size of U-Net was 256x256 px at 0.5-m spatial resolution. We applied data augmentation techniques to avoid U-Net overfitting by randomly modifying input imagery at each iteration. Data augmentation techniques included horizontal and vertical flips, cropping and zooming, piecewise affine and perspective transformations, and linear contrast enhancement.

U-Net was trained with a batch size of 16 images for 100 epochs. Labeled images from the training area (Fig. 1) were split into training, validation (33%, for fine-tuning parameters) and testing (33%, for model robustness testing) subsets. We must emphasize that evaluation of the model performance on a test subset does not constitute a statistically rigorous validation of the model-derived maps. This is because the test subset is acquired from the same location and distribution as a training subset and represents only a small portion of the area of interest. Such an approach is suitable only for assessing the classification capabilities of the classification model and is not appropriate for evaluating mapping capabilities, i.e., how accurate the map is. And finally, manual delineation of all craters for large areas is not feasible and sampling-based approaches should be used. Therefore, we selected a separate region for validation employing a statistically rigorous approach with sampling (Fig. 1 and section 2.5.2).
2.5. Evaluation and validation

We trained the U-Net model on the fully labeled WV-2 images located in the eastern part of the study region (Fig. 1, section 2.5.1). We selected a small subregion for comparing U-Net (section 2.4.1) against the baseline model which is based on segmentation and RF (section 2.4.2). We followed recommended practices for land-cover product validation (Olofsson et al., 2014) to validate the U-Net-generated maps of craters in the western part of the study area (Fig. 1, section 2.5.2).

2.5.1. Training data generation

For training classification models, we selected WV-2 images if they contained evidence of artillery craters. This selection process was done with a visual inspection of VHR satellite imagery and by comparison with the crater fields which were previously delineated by Bellingcat (Case, 2016). Bellingcat broadly delineated crater fields (not individual craters) at a regional level manually for their report. Furthermore, this report included only those crater fields that could be directly connected with potential firing positions within the Russian Federation. These crater field maps, though confined to the areas near the international border of Ukraine and the Russian Federation, comprise the only mapping efforts focused on crater fields and were therefore used as an initial resource for determining the study area. Overall, we labeled 18,472 craters that distributed over 61,056 patches of 256x256 px size that were used for training the U-Net model.

Before performing large-area crater detection and mapping, we compared U-Net (section 2.4.1) and RF (section 2.4.2) classification capabilities to distinguish craters for the subset region which included two images each 2.2 x 2.7 km². The first image contained 1,374 labeled craters and was used for training the models, and the second image contained 873 craters and was used for model inter-comparison. Both models were trained and tested on the same crater locations. Both models were run with the parameter settings described in section 2.4.

2.5.2. Validation

We performed validation of the U-Net model for the western strip (Fig. 1) area of 858 km², which was independent from the area used for training. Labeling the entire region is not feasible, and therefore a sample-based approach with an area frame was adopted, following recommended practices for land cover land use accuracy assessment and area estimation (Olofsson et al., 2014). The area was split into 1 x 1 km² blocks, each representing a sampling unit. The population size was N = 858 and the total number of samples was 45. The reason for selecting a frame size of 1 x 1 km² comes from previous studies (Gallego et al., 2014; Vermote et al., 2020) with a trade-off of having area large enough to incorporate multiple craters (10’s and 100’s) and small enough so labeling would be still feasible. We ran the U-Net model and calculated the number of detected craters in each 1 x 1 km² block. Sampling was stratified to account for a varying number of craters. The following three strata were used (Fig. 3).

- Stratum 1 (“low”): with the number of detected craters less than 20 (accounted for 33% of population with the number of units N₁ = 283).
- Stratum 2 (“medium”): with the number of detected craters more than 20 and less than 40 (37%, N₂ = 315).
- Stratum 3 (“high”): with the number of detected craters more than 40 (30%, N₃ = 260).

For estimate of the parameter in stratified sampling, we used standard equations (Cochran, 1977; Ch. 5):

\[
\overline{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi},
\]

\[
\bar{y} = \frac{1}{n} \sum_{h=1}^{H} n_h \overline{y}_h,
\]

where \( H \) is the number of strata \( (H = 3) \); \( W_h = N_h/N \) is the stratum weight; \( y_{hi} \) is the value of sample \( i \) in stratum \( h \); \( \overline{y}_h \) is the sample mean and \( n_h \) is the number of samples in stratum \( h \) (in our case \( n_h = 15 \)); \( \bar{y} \) is the estimate of stratified sampling. Examples of parameter \( \overline{y}_h \) include the number of correctly predicted craters (true positives) or the area of incorrectly mapped craters (area of false positives or commission error).

An estimate of the variance of \( \overline{y}_h \) within stratified random sampling (Cochran, 1977):

\[
v(\overline{y}_h) = \frac{1}{N^2} \sum_{i=1}^{N} N_i (N_i - n_i) s_i^2 / n_i,
\]

where \( s_i^2 = \frac{1}{s_i - 2} \sum_{i=1}^{n_i} (y_{hi} - \overline{y}_h)^2 \) is an unbiased estimate of variance within stratum \( h \).

Fig. 3 shows geographical distribution of the number of craters detected by the U-Net model, strata location and samples for the validation area.

We used standard accuracy metrics for assessing U-Net performance derived from the confusion matrix (Table 1):

Producer’s accuracy (or recall): \( PA = TP / (TP + FN) \),

User’s accuracy (or precision): \( UA = TP / (TP + FP) \),

F-score = \( 2 \times PA \times UA / (PA + UA) \),

where \( TP, FN \) and \( FP \) were estimated in terms of the number of craters and the area of craters with stratified random sampling.

3. Results

Table 2 provides metrics when comparing the performance of the SLIC and RF-based model against the U-Net models based on various spectral bands inputs using two specifically allocated regions of the study area (section 2.5.1). The RF-based model failed to properly identify craters, while U-Net had enough discriminative power to identify craters in the satellite imagery.

Table 3 and Table 4 provide confusion matrices along with performance metrics (Eqs. (7)-(9)) when validating the U-Net model for the validation area (section 2.5.2) in terms of crater area and counts, respectively.

The average size of the detected craters (true positives) and missed craters (false negatives) in the validation area was 73.8 ± 11.6 m² and...
49.4 ± 7.0 m$^2$, respectively. (Uncertainties are expressed as one standard error, SE.) The estimated area of craters that were overdetected (false positives) by U-Net was 41.0 ± 5.6 m$^2$.

Fig. 4 shows examples of detected craters by the U-Net model and comparison against the reference (labeled) craters. Fig. 5 shows how performance metrics vary depending on the reference crater size. Distributions of the area of labeled and detected craters are shown in Fig. 6.

Within the validation area of 858 km$^2$, we estimated that craters occupied 1.20 ± 0.43 km$^2$ (p-value = 0.008) or 0.14% of the area, which corresponds to approximately 170 soccer fields. The estimated number

Table 1
Confusion matrix. TP, FP, FN and TN are the number or area of true positives, false positives, false negatives and true negatives craters.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Crater</th>
<th>TP</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No crater</td>
<td>FN</td>
<td>TN</td>
</tr>
</tbody>
</table>

Table 2
Performance metrics (Eqs. (7)–(9)) when comparing the performance of the SLIC and RF-based and U-Net models.

<table>
<thead>
<tr>
<th>Metric</th>
<th>SLIC + RF</th>
<th>U-Net (RGB)</th>
<th>U-Net (RGB + NIR)</th>
<th>U-Net (eight bands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA</td>
<td>0.02</td>
<td>0.83</td>
<td>0.84</td>
<td>0.83</td>
</tr>
<tr>
<td>UA</td>
<td>0.29</td>
<td>0.64</td>
<td>0.64</td>
<td>0.46</td>
</tr>
<tr>
<td>F-score</td>
<td>0.04</td>
<td>0.72</td>
<td>0.73</td>
<td>0.59</td>
</tr>
</tbody>
</table>

Table 3
Confusion matrix and performance metrics of the U-Net validation in terms of average crater size (in m$^2$) per sample unit.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Crater</th>
<th>No crater</th>
<th>PA = 0.67</th>
<th>UA = 0.39</th>
<th>F = 0.50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Crater</td>
<td>936</td>
<td>1453</td>
<td>0.39</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>No crater</td>
<td>459</td>
<td>997,151</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. Distribution of the number of detected craters (left image) that were used for stratification and selection of samples (right image) within the validation area (Fig. 1).
of craters was $22,072 \pm 8,044$ ($p$-value = 0.009).

4. Discussion

The main objective of the study was to investigate the efficiency of sub-meter satellite imagery to detect and map artillery craters in agricultural fields without substantive obstruction from vegetation cover such as trees. We selected an area in Eastern Ukraine (parts of Donetsk and Luhansk oblasts) as a case study during the Russian aggression against Ukraine in 2014 (section 2.1), which is characterized by the heavy use of artillery systems (section 2.2). We used images acquired by the WorldView-2 system at 0.5-m spatial resolution and made available within the NASA CSDA Program (section 2.3).

First, we compared two methodologies: (i) a combination of segmentation using the SLIC algorithm and subsequent classification of objects using the RF classifier (section 2.4.2); (ii) semantic segmentation with the deep learning model, U-Net (section 2.4.1). Our results showed that the RF-based model did not provide enough discriminative power to identify craters in the imagery. This was probably due to the fact that the spatial context of the imagery was used only during segmentation and the RF model used spectral properties only (RGB bands) when classifying segments. RF is known to be dependent on the input features (Rodriguez-Galiano et al., 2012), and we did not provide any specifically designed features to target craters, e.g., spatial filters. In turn, U-Net was able to automatically extract meaningful features from the input imagery, as this is one of the advantages of neural networks, and yielded much better performance ($F$-score = 0.72) compared to the RF-based model ($F$-score = 0.04) for the small testing area (section 2.5.1). Also, we did not find substantial improvements in performance when more spectral bands were added, probably because the model is driven by spatial context rather than spectral context. We did not use multi-date imagery and likely the spatial context is deemed more important than the spectral one. Therefore, the U-Net-derived crater map (with WV-2 RGB inputs) was used for the large-scale validation (section 2.5.2).

We found that U-Net was able to identify artillery craters with the following performance metrics (Tables 3 and 4): $PA = 0.671$, $UA = 0.392$ and $F$-score $= 0.495$ in terms of crater area and $PA = 0.559$, $UA = 0.427$ and $F$-score $= 0.484$ in terms of crater counts. The model’s performance depended on the crater size (Fig. 5)—as the size of reference

![Fig. 4. Examples of the reference (labeled) and detected craters by the U-Net model. In this example, the detection success between multiple groundcover variations can be seen, with correct detections in both bare soil and green grass fields. (c) 2014, Maxar. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)](image)

![Fig. 5. Dependence of the U-Net PA, UA and F-score on the crater size.](image)
craters increased all performance metrics improved. For example, for craters larger than 60 m², F-score of crater area mapping increased 16% (F-score = 0.576) and F-score of detecting the number of craters increased 65% (F-score = 0.797) compared to the situation when all craters were considered. For crater size between 10 and 30 m², the U-Net model was likely to miss more craters, while for crater size less than 10 m² the likelihood of false positives increased (Fig. 6). Overall, the average area of missed craters was 49.4 ± 7.0 m² and the majority of those were likely coming from mortars.

The U-Net model favored PA over UA. This is the desired behavior in crater mapping as we prioritize the minimization of omission errors even at the expense of the increased commission errors. This dynamic of the model is intrinsically connected with its real-world application. In favoring PA over UA in the context of UXO hazard mapping, it is preferable to produce liberal estimates of danger, which we can subsequently investigate and decide to lower. If the model was too conservative, we would be indicating that non-safe areas are in fact safe. The risks of this scenario are obvious, and are therefore a critical consideration when designing the automation methodology.

Our U-Net model was better at detecting the number of craters compared to the crater boundaries and shapes as the number-wise metrics were consistently higher than the area-wise metrics (Fig. 5, Tables 3 and 4). This is once again the desired behavior of the model, considering that detecting crater locations is the priority over detecting the exact crater shape. In real-world applications of UXO detecting, it is more important to derive ‘yes’/’no’ relationships of crater location, as opposed to idealized crater shapes or sizes.

The approach for validating the U-Net-derived crater map followed recommended practices for land cover validation (Olofsson et al., 2014) and allowed us not only to estimate accuracy metrics (Tables 3 and 4), but to provide unbiased estimates of the number and area of craters along with uncertainties. We shall emphasize that the approach when all labeled data are split into train and validation sets is only appropriate for evaluating the performance of the classification model but not necessarily the generated land-cover maps. The model should be run on previously unseen data and locations and either the entire population should be labeled, which is not feasible in most real-world situations, or a sampling-based approach (section 2.5.2) should be employed. The advantages of the sampling-based approach are its feasibility for large areas, well-formulated statistical framework, and ability to provide unbiased estimates along with uncertainties. Using a stratified random sampling approach, we estimated that more than 22,000 craters occupied 0.14% of the validation area of 858 km².

The proposed approach has a number of applications. Among them are the identification of hazardous areas for the purpose of demining testing. Both exploded and unexploded ammunition may also result in heavy metal contamination of soils, with short- and long-term consequences (Williams and Rintoul-Hynes, 2022; Jaishankar et al., 2014). The presence of artillery and bomb craters changes the functioning of landscapes (Lin, 2022), may limit agricultural production, pollute crops, and in the aggregate may bring extra pressure on regional and global food security (Pereira et al., 2022; Deininger et al., 2023; Sytar and Taran, 2022) (Fig. 7). Artillery crater maps can guide environmental accounting, decontamination, and rehabilitation efforts for restoration of degraded lands and vegetation.

The presented study has some intrinsic limitations. The biggest challenge is the lack of truly independent validation data on crater shape and location. Ideally, that would be ground-based surveys or surveys aided with drone or aerial imaging at much better spatial resolution (<0.5 m) than our satellite imagery. Unfortunately, such data do not exist for 2014 and collection of such data during the recent full-scale invasion in February 2022 is impractical because of safety and security concerns. Therefore, we had to rely on visual photointerpretation of satellite imagery that were also used to generate crater maps. The disadvantage of such an approach is that we are limited by the spatial resolution of the imagery (~0.5 m) and therefore very small craters, likely from mortars, might not be visible both within photointerpretation by the analyst or the U-Net model. Consequently, we might underestimate the area and number of craters. However, where artillery craters occur, we may expect the appearance of mortar craters too. Thus, developed crater products may serve as a guide for deactivation and decontamination purposes from other types of lethal weapons.

Another limitation is the difficulty in identifying craters in areas with tree cover (forest) and/or heterogeneous surfaces. Fig. 8 provides various examples of the landscape where the identification of craters, either by an expert or the U-Net model, becomes extremely challenging. For now, we are limiting our detection and analysis for agricultural fields with short vegetation (<3–5 m).

The presented study offers multiple avenues for further research. Among them is the detection of the direction from which artillery was fired and the identification of weapon type based on the shape properties of craters. That would require substantial efforts in providing fine-grained classification nomenclatures when labeling craters. Sub-meter satellite imagery is currently acquired irregularly, which limits the identification of the timing of craters’ appearance. At the same time, coarser spatial resolution imagery such as commercial Planet Dove (3 m) and non-commercial Sentinel-2 (10 m) are acquired at a daily or near-daily scale and, therefore, data fusion approaches should be explored to assess the capabilities of the 3-10 m satellites to map craters (Bennett et al., 2022). And, finally, our model for crater detection and mapping was based on RGB spectral bands and we did not find improvements in performance when other spectral bands were added. However, we limited our study to single-date imagery, and for multi-date multi-sensor models the impact of other spectral bands, e.g., infrared spectrum or active microwave sensors, should be further explored.

5. Concluding remarks

We presented one of the first studies on detecting and mapping craters from contemporary artillery weapons using sub-meter satellite imagery and the state-of-the-art deep-learning image processing technique. Specifically, we demonstrated the efficiency of the U-Net semantic segmentation model to identify craters in the WorldView-2 imagery at 0.5-m spatial resolution during the Russia’s war of aggression against Ukraine in Donetsk and Luhansk regions in 2014. The developed model yielded PA = 0.671, UA = 0.392 and F-score = 0.495 in terms of crater area and PA = 0.559, UA = 0.427 and F-score = 0.484 in terms of...
crater counts. That performance varied with crater size and improved to $PA = 0.803$, $UA = 0.449$ and $F$-score = 0.576 (area-wise) and $PA = 0.891$, $UA = 0.721$ and $F$-score = 0.797 (number-wise), when identifying craters with area $>60$ m$^2$ (68%-percentile of all labeled craters). We estimated $22,072 \pm 8,044$ craters occupying $1.20 \pm 0.43$ km$^2$ (or 0.14%) in a subregion of Donetsk oblast of 858 km$^2$ showing the impact of warfare, and the use of artillery, on agricultural fields. We identified limitations of the proposed approach to detect craters under substantial tree cover and highly heterogeneous landscapes and outlined avenues for further research.

**Declaration of competing interest**

The authors declare that they have no known competing financial
Sexton, J.O., Song, X.P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.H.,
Pereira, P., Barceló, D., 2022. Russian-Ukrainian war impacts the
scitotenv.2022.155865.
Rodríguez-Galiano, V.F., Chica-Olmo, M., Arbaña-Hernández, F., Atkinson, P.M.,
biomedical image segmentation. In: Proc. 18th International Conference Medical
catalogue of Martian impact craters and advanced evaluation of crater detection
algorithms using diverse topography and image datasets. Planet. Space Sci 59,
segmentation using 3D fully convolutional deep networks. In: Proc. 8th International
Workshop Machine Learning in Medical Imaging (MILMI 2017). Springer
Collins, K.M., Channan, S., DiMiceli, C., Townshend, J.R., 2013. Global, 30-m
resolution continuous fields of tree cover: Landsat-based rescaling of MODIS
vegetation continuous fields with lidar-based estimates of error. Int. J. Digital Earth
https://doi.org/10.1016/j.icarus.2018.06.022.
reveal cropland losses in South-Eastern Ukraine under military conflict. Front. Earth
Sytar, O., Taran, N., 2022. Effect of heavy metals on soil and crop pollution in Ukraine –
23.4.3603.
Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P.,
Rommen, B., Floury, N., Brown, M., Traver, J.N., Deghaye, P., Duesmann, B.,
Rosich, B., Miranda, N., Bruno, C., I’Abbate, M., Croci, R., Pietropaolo, A.,
Tucker, C., Brandt, M., Henaux, P., Karityya, A., Rasmussen, K., Small, J., Igel, C.,
Reiner, F., Meloicz, K., Meyer, J., Sinno, S., Romero, E., Glennie, E., Fitzs, Y.,
Morin, A., Pinzon, J., McClain, D., Morin, F., Porter, C., Loeffler, S., Kergoat, L.,
Insoufou, B.A., Savadogo, F., Wigneron, J.-P., Poulter, B., Ciais, P., Kaufmann, R.,
Myneni, R., Saatchi, S., Fensholt, R., 2023. Sub-continental-scale carbon stocks of
individual trees in African drylands. Nat 615, 80–86. https://doi.org/10.1038/s41586-
022-05653-6.
trees in Tonga using very high spatial resolution WorldView-3 data. Rem. Sens.
8, 3113. https://doi.org/10.3390/rs12193113.
Williams, O.H., Rintoul-Hynes, N.L.L., 2022. Legacy of War: Pedogenesis divergence and
heavy metal contamination on the WWI front line a century after battle. Eur. J. Soil
Sci. 73, e13297 https://doi.org/10.1111/13297.
Witmer, F.D., 2008. Detecting war-induced abandoned agricultural land in northeast
Wulder, M.A., Roy, D.P., Radoloff, V.C., Loveland, T.R., Anderson, M.C., Johnson, D.M.,
Healey, S., Zhu, Z., Scambos, T.A., Pahlevani, N., Hansen, M., Gorelick, N.,
Crawford, C.J., Masek, J.G., Hermosilla, T., White, J.C., Belward, A.S., Schaaf, C.,
Woodcock, C.E., Huntington, J.L., Lymburner, L. Hostert, P., Gao, F., Lyapustin, A.,
Skakun, S., Ahys, C., Adegbenro, M., Becker-Reshef, I., Duncan, E., Eun, J., Hall, J.,
High-Impact Hot Spots of Land Cover Land Use Change in Ukraine. In: 12th
International Conference on Dependable Systems, Services and Technologies