Unexploded Ordnance Detection in Hyperspectral Images by Using Deep Neural Networks

Milan Bajić,[1] Božidar Potočnik[2]

[1] Zagreb University of Applied Sciences, Department of IT and Computer Sciences, Zagreb, Croatia
mbajic@tvz.hr
[2] University of Maribor, Faculty of Electrical Engineering and Computer Science, Institute of Computer Science, Maribor, Slovenia
bozidar.potocnik@um.si

Unexploded Ordnance (UXO) is a major threat affecting the lives of people in more than 60 countries. This work tests deep neural networks to automatically detect UXO in Hyperspectral Images (HSI). Initially, we constructed our own dataset of 134 HSI cubes divided into three folds: two for training and one for validation. U-Net was selected through preliminary experiments as the most promising detection method among those compared. Customised loss functions were designed for the U-Net, resulting in 3 different models. These models were trained and validated in a supervised manner on our data. The results obtained are very promising with a UXO detection rate of around 70% and an F1 score above 0.8.

DOI
https://doi.org/10.18690/um.feri.2.2025.7

ISBN
978-961-286-960-1

Keywords:
nexploded ordnance, hyperspectral images, image segmentation, convolutional neural networks, dataset construction

University of Maribor Press

DOI
https://doi.org/10.18690/um.feri.2.2025.7

ISBN
978-961-286-960-1

DETEKTIRANJE NEEKSPLODIRANIH UBOJNIH SREDSTEV NA HIPERSPEKTRALNIH SLIKAH Z UPORABO GLOBOKIH NEVRONSKIH MREŽ

MILAN BAJIĆ,[1] BOŽIDAR POTOČNIK[2]

[1] Tehničko Veleučilište u Zagrebu, Katedra za IT in računalništvo, Zagreb, Hrvaška
mbajic@tvz.hr

[2] Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko, Inštitut za računalništvo, Maribor, Slovenija
bozidar.potocnik@um.si

Ključne besede:
neeksplodirana ubojna sredstva,
hiperspektralne slike,
segmentacija slik,
konvolucijske nevronske mreže,
konstrukcija podatkovne zbirke

Neeksplodirana ubojna sredstva (UXO) so velika grožnja, ki ogroža življenja ljudi v več kot 60 državah. To delo preizkuša globoke nevronske mreže za samodejno zaznavanje UXO v hiperspektralnih slikah (HSI). Na začetku smo izdelali lasten nabor podatkov iz 134 kock HSI, razdeljenih v tri dele: dva za učenje in en za validacijo. U-Net je bil s predhodnimi poskusi izbran kot najbolj obetavna detekcijska metoda med primerjanimi. Prilagojene funkcije izgube so bile zasnovane za U-Net, s čimer smo dobili 3 različne modele. Ti modeli so bili naučeni in preizkušeni na nadzorovan način na naših podatkih. Dobljeni rezultati so zelo obetavni z uspešnostjo detektiranja UXO okoli 70 % in oceno F1 nad 0,8.

Univerzitetna založba Univerze v Mariboru

1 Introduction

Hyperspectral Imaging (HSI) is a camera-based technique that captures a dense image series, characterized by its wavelength range and number of channels. The Specim IQ camera utilized in this study possesses a spatial resolution of 512x512 pixels and 204 spectral channels. This camera functions on the principle of an integrated hyperspectral push-broom line scanner, with internal processing occurring during the acquisition process. The results of this process are calibrated reflectance values within the range of 400nm to 1000nm, with a mean spectral resolution of 7nm within this range. However, higher FWHM (i.e., full width at half maximum) values at higher wavelengths generally result in data that is not useful (from 900nm to 1000nm). Hyperspectral images are defined as a set of pixels containing spatial dimensions (m rows and n columns) and spectral information (K wavelength channels). This is commonly referred to as a three-dimensional hyperspectral cube (hypercube), data cube, spectral cube, spectral volume, or data volume. Hyperspectral imaging provides valuable insights into the physical and chemical properties of analysed materials, with the gathered information including geometric and physical characteristics such as size, orientation, shape, texture, and colour. A raw hyperspectral image consists of multiple interconnected sub-images, with each sub-image representing an object's spatial distribution and intensity at a specific wavelength. It is possible to extract an individual spatial image from the hypercube at any point within the system's spectral sensitivity range. Consequently, a hyperspectral image represented as I(x, y, l) can be interpreted either as a standalone spatial image I(x, y) at a given wavelength (l) or as a spectral profile I(l) at each pixel (x, y). Each pixel within the hyperspectral image contains a unique spectral signature corresponding to its specific location, effectively acting as a fingerprint for identifying its composition.

The proposed research is concerned with the detection of explosive objects. It is estimated that over sixty countries currently possess remnants of Unexploded Ordnance (UXO), (Bajić & Potočnik, 2024). For instance, UXOs from the First World War have been unearthed during archaeological excavations and construction (National University of Public Service, Hungary & Ember, 2021), (Roberts & Williams, 1995). Furthermore, over 100 countries have been affected by explosions at ammunition storage sites, resulting in casualties, environmental damage, mass displacement, and economic disruption (iMMAP-IHF, Humanitarian Access

Response - Monthly Security Incidents Situation Report, November 2018, 2019). It is estimated that there have been tens of thousands of accidental explosions at ammunition depots (Ammunition Storage Area Explosions – EOD Clearance, 2021). In contrast to buried UXOs, surface UXOs can be effectively detected with hyperspectral imaging (Bajić et al., 2013), (Bajić & Bajić, 2021). International efforts to address UXO contamination include initiatives such as the Mine Ban Treaty, which aims to eliminate landmines and promote demining activities worldwide. Organizations such as the United Nations Mine Action Service (UNMAS) collaborate with affected countries to clear contaminated areas and educate on the dangers of UXOs. Furthermore, international funding and partnerships have been established to support technological advancements and training for more effective UXO detection and removal. The presence of explosive remnants that have not been removed thus far continues to pose a significant threat.

Deep neural networks have been employed for many years with great success in the segmentation and classification of multidimensional data (e.g. images, volumes) or for object detection. One such early and extensively deployed neural network is the U-Net network (Ronneberger et al., 2015). U-Net is a well-established and recognized Convolutional Neural Network (CNN) architecture with a left-side contracting phase (i.e., encoder), a bottleneck phase, and an expansive right path (i.e., decoder). This configuration, in conjunction with skip connections from the encoder to the corresponding decoder layer, facilitates the preservation of both spatial and spectral characteristics. U-Net was originally developed for the classification and segmentation of biomedical images. Biomedical and hyperspectral imaging are similar in that both rely on multi-channel imaging, which is why it is so difficult to collect large datasets in either area. U-Net was designed for data with typically 3 channels, whereas hyperspectral images typically have more than 100 channels. The integration of U-Net architectures with HSI has demonstrated efficacy in detecting UXOs. According to (Tuohy et al., 2023), UAV-based HSI can effectively detect surface-level Explosive Remnants of War (ERW), suggesting that deep learning models like U-Net could enhance detection accuracy. This study combined U-Net models with HSI data to detect surficial explosive ordnance using UAV-mounted HSI systems. Authors employed a U-Net to detect artillery and rocket craters in Ukraine, achieving an 89% accuracy rate compared to human marking, thereby demonstrating the model's ability to detect UXO. When applied to

hyperspectral data, U-Net-based segmentation offers a more efficient and safer alternative to conventional methods for UXO and ERW detection.

In this study, a computational method is proposed for the identification of UXOs through the utilization of a U-Net neural network in conjunction with HSI data. The subsequent sections will provide a detailed exposition of the acquisition of the dataset, its annotation, and the division of the data into training and testing sets. Section 3 will provide a concise overview of the computational methods employed and the salient characteristics of the convolutional networks utilized in this study. The ensuing section, i.e. Section 4, will present the experiments, implementation details, metrics used, details about model training and hyperparameters' fine-tuning. The results section will present the quantitative and qualitative results obtained. Finally, the work is concluded with a discussion of potential future directions for research.

2 Dataset

The dataset was created using an experimental Graeco Latin square design (Guthrie, 2020). This was done to minimise the specific spatial sensor position, the influence of light change, and the influence of the surface environment. The surface is characterised by gravel, grass, grassland, grass in sunlight, and bush leaves in shadow.

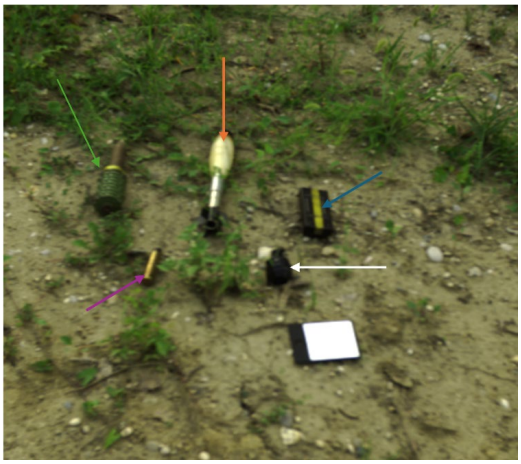

Figure 1: Four UXO, plastic pressure plate, and calibration card
Source: Own.

Four different types of UXO were identified in this study: a mortar mine (marked with an orange arrow in Figure 1), an anti-personnel mine (green arrow), a hand grenade (white arrow), a large bullet (purple arrow) and a plastic pressure plate used to activate explosive ordnance (blue arrow). It is important to note that a white reference card is incorporated into all images, as it is utilised for post-recording calibration in the camera to calculate reflectance values from digital numbers. The approximate image acquisition time ranges between 60 and 120 seconds, depending on exposure and integration settings. The dataset under consideration consists of 134 hyperspectral cubes, with the presence of one to five of the aforementioned objects in every HSI image.

Figure 2: Acquisition of dataset
Source: Own.

The objects were supplied by HCR-CTRO, an educational and research company specialising in land mine clearance education and land mine clearance technology certification. Each of the five objects is present in at least 100 HSI images. The recorded images (see Figure 2) were subjected to a visual control process following the calibration process. In instances where the data proved inadequate, the image capture was repeated with the same position and using the same exposure and integration values. Each cube is stored in the ENVI HDR + data file format. The

data type is 4, interleave BIL, while channels 70, 53 and 19 were utilised for the purposes of visualization and annotation of RGB images.

The images in the database were annotated manually, using visualised RGB images. Each object was denoted by polygons. It was ensured that the same object (i.e., the target) was assigned the same class label in all HSI images. We used the Supervisely platform for annotation. Initial annotations were automatically determined using the 'segment anything' option, which we visually inspected and manually corrected. The resulting mask was used for all wavelengths in the HSI cube.

The dataset under consideration thus contains 134 HSI images. The data was randomly divided into three parts (i.e., subgroups, folds) to enable 3-fold validation. The first two folds comprise 45 images each, while the third fold consists of 44 images. The division of the dataset into subgroups is illustrated in Table 1 for each fold.

Table 1: Dataset division aimed for 3-fold cross validation. The numerical values represent the image ID within the database.

	Image ID
Fold 1	180, 181, 182, 184, 186, 187, 189, 196, 197, 201, 202, 203, 204, 206, 208, 215, 221, 225, 241, 244, 249, 250, 252, 256, 258, 259, 266, 271, 273, 275, 276, 281, 284, 285, 286, 288, 295, 297, 298, 301, 304, 305, 307, 309, 310
Fold 2	178, 179, 183, 190, 195, 198, 211, 212, 214, 219, 220, 222, 223, 226, 227, 232, 237, 238, 239, 240, 243, 246, 251, 254, 255, 260, 265, 268, 269, 270, 272, 274, 280, 282, 283, 287, 291, 293, 296, 299, 300, 302, 311, 312, 313
Fold 3	185, 188, 191, 193, 194, 199, 200, 205, 207, 210, 213, 216, 217, 218, 224, 228, 229, 230, 231, 233, 234, 235, 236, 242, 245, 247, 248, 253, 257, 261, 262, 263, 264, 267, 277, 278, 279, 289, 290, 292, 294, 303, 306, 308

It should be noted that the file size of each HSI image is 209 MB, whereas the mask (i.e., annotations) is 1.36 KB.

3 Computational methods

In the preliminary phase of this research, experimentation was conducted with various CNN architectures incorporating distinct attention mechanisms, utilising a binary cross entropy loss function and an Adam optimiser. The following architectures were assessed: U-Net and U-Net with attention, DeepLabV3+ and

FCN. The aim was to identify the most promising architecture for further experimentation.

U-Net is an exemplary, well-established convolutional neural network that is founded on an encoder-decoder architecture (Ronneberger et al., 2015). Attention U-Net is a modification of U-Net that uses attention gates by upsampling operation at each decoder layer. In skip connections, the gates are used to highlight the important regions and, thus, allowing the network to focus on the more relevant features (Yan et al., 2018). DeepLabV3+ employs Atrous Spatial Pyramid Pooling (ASPP), a technique that captures features at multiple scales. This approach facilitates the model's comprehension of the context. Atrous convolution, a pioneering innovation, introduces gaps between the values in a convolutional kernel, thereby expanding the filter's capacity to encompass larger areas of the input image without increasing the number of parameters (Chen et al., 2018). The replacement of fully connected layers with convolutional layers in a Fully Convolutional Network (FCN) (Long et al., 2015) results in a network capable of accommodating inputs of any size. FCN incorporates skip connections for feature map combination; however, these are generally simpler and less structured in comparison to U-Net.

Table 2 summarizes the trainable parameters and number of layers for the four tested models.

Table 2: Four tested CNNs: Number of layers and trainable parameters

	Model	Trainable parameters	Number of layers
1.	FCN	33.311.928	156
2.	U-Net	34.590.913	81
3.	Attention U-Net	31.977.317	97
4.	DeepLabV3+	10.689.537	434

All tested models were downloaded from the Keras and Pytorch GitHub, respectively. The architectures of the models were not modified, only the input and output layers were adapted to our problem (the expected input size was 512x512x102 and the size of the predicted binary mask was 512x512x1). The Binary Cross Entropy (BCE) Loss function has been utilized. All experiments were performed on an optimised Google Colab pay-per-use environment, in combination with Google Drive for large data storage. The GPU used was an A8 with 40 GB of

RAM, with 81 GB of system RAM. The code was written in Python, and the models were saved in the h5 format.

In preliminary tests, the models were evaluated on only 5 HSI test images, the rest were used for training. Table 3 summarises the metrics obtained, calculated at the pixel level. It can be seen that the most balanced results were obtained using the classic U-Net, and this model was, therefore, selected for all further experiments. In the sequel of this research, we thus experimented with the original U-Net architecture, focusing on the design of a custom loss function.

Table 3: Results of a preliminary model selection experiment. The best model is in bold.

Model	Dice	IoU	Precision	Recall
Attention U-Net	0,639	0,523	0,710	0,640
DeepLabv3+	0,663	0,519	0,851	0,568
U-Net	**0,764**	**0,648**	**0,917**	**0,689**
FCN	0,992	0,008	0,875	0,632

4 Results

This section describes the experimental design and implementation details, followed by a presentation of the quantitative and qualitative results obtained.

4.1 Experiments and implementation details

We dealt with UXO detection from HSI images in this research. All UXOs were treated as a common class (i.e., we did not distinguish between different types of UXOs). We therefore dealt with the so-called binary segmentation problem: pixel belongs to UXOs or pixel is part of the background. The computational method used was the U-Net neural network (see previous section). Three different modifications of the U-Net are proposed in this work (models M1 to M3). All of these models are based on the same 20-layer architecture, they just use different loss functions.

In this research, we therefore focused on investigating the effect of the loss function on segmentation effectiveness. A custom loss function was designed by combining the following loss functions in the form of a linear combination: Binary Cross Entropy Loss, Dice Loss, Focal Loss, Tversky Loss and Edge Loss. Edge loss is also

a custom loss function that we use to estimate areas near the edges of objects where there should be differences in the spectral signature. The weights of the loss functions have been chosen so that their sum equals 1. All three models had the same weights for BCE loss (set to 0.25), Dice loss (0.4), and Focal loss (0.1). Model M1 had the weights for the Tversky loss set to 0.1 (with alpha=0.2) and the Edge loss set to 0.15, model M2 had these weights set to 0.15 (Tversky loss) and 0.1 (Edge loss), while model M3 used the weight of 0.1 (with alpha=0.05) for the Tversky loss and 0.15 for the Edge loss.

Let's give some more details about the training and evaluation of the models. The implementation of the methods and the hardware used are practically the same as those presented for the preliminary tests (see previous section). No regularisation or normalisation was used in the training. The models were trained for 200 epochs using the Adam optimiser, with an initial adaptive learning rate of 0.00001 and cosine decay. Our own dataset was used for training. A full 3-fold validation was not performed in our study due to time constraints. All three models were trained only once, with folds 1 and 2, and tested on fold 3. The results obtained in this way are reported in the sequel.

We evaluated the effectiveness of the methods using established metrics such as Precision, Recall, and F1 score. The methods were evaluated both in terms of their effectiveness in detecting UXO (we used a threshold of 0.5 for the overlap between prediction and ground truth) and in terms of how well each UXO was detected (undetected UXO were of course excluded from these statistics). In both cases, we will present the results in two ways: i) per-dataset statistics and ii) per-image statistics. In the first case, we will calculate the statistics for the test set as a whole (i.e. as if all the test images were combined into one large common image). In the second case, we calculate the statistics for the test set as the average of the statistics calculated for the individual images in the test set.

4.2 Quantitative and qualitative results

Firstly, the effectiveness of the UXO detection is indicated by the so-called 'per-dataset' statistics. There was a total of 151 UXOs (i.e., ground truth regions) in the entire test set, while methods M1 to M3 predicted between 211 and 295 regions, with matches between 99 and 110 regions. Table 4 summarises these statistics for all

three methods. The M3 model proved to be the best, with the highest F1 score and precision.

Table 4: Effectiveness of UXO detection: 'per-dataset' statistics. Best results are in bold.

Model	F1 score	Precision	Recall
M1	0,488	0,381	0,682
M2	0,493	0,373	**0,728**
M3	**0,547**	**0,469**	0,656

We also provide the effectiveness of UXO detection in the 'per-image' way, where the calculated metrics for each image are averaged over the entire test set. Such calculated metrics are shown in Table 5 (mean and standard deviation). Again, the M3 method proves to be the best. However, it should be noted that method M1 did not segment any UXO in 2 images. Methods M2 and M3 did not detect any UXO in 4 images. Table 5 also takes into account undetected UXO.

Table 5: Effectiveness of UXO detection: 'per-image' statistics. Best results are in bold.

Model	F1 score	Precision	Recall
M1	0.549 ± 0.305	0.515 ± 0.324	0.655 ± 0.313
M2	0.556 ± 0.289	0.508 ± 0.321	**0.711 ± 0.265**
M3	**0.583 ± 0.285**	**0.582 ± 0.336**	0.634 ± 0.279

We also wanted to answer the following question: If the UXO was detected, how well was it detected? Again, we calculated statistics using both ways, but since the differences are extremely small, we present only the statistics calculated using the 'per-dataset' way. Table 6 shows these results (mean and standard deviation). Even according to these statistics, the M3 method was the best performing method.

Table 6: Quality of UXO detection: 'per-dataset' statistics. Best results are in bold.

Model	F1 score	Precision	Recall
M1	0.817 ± 0.108	0.853 ± 0.140	0.813 ± 0.145
M2	0,815 ± 0.111	0,816 ± 0.144	**0,850 ± 0.152**
M3	**0,832 ± 0.009**	**0,882 ± 0.117**	0,813 ± 0.142

Finally, we show an example of successful and less successful UXO detection using the most promising model, M3. An example of successful detection is shown in Figure 3. The quality of UXO detection was extremely high (average F1 score was 0.945) for this example image.

Figure 3: Example of successful UXO detection with the M3 model: original RGB image (left) and model prediction overlaid on the RGB image (right).
Source: Own.

Figure 4 shows an example of unsuccessful detection with the M3 model. For this example image, the quality of the detected UXO was very low (average F1 score was 0.550), with two UXO not detected at all.

Figure 4: Example of unsuccessful UXO detection with the M3 model: original RGB image (left) and model prediction overlaid on the RGB image (right).
Source: Own.

5 Conclusion

Our preliminary research has confirmed the feasibility of UXO detection from HSI images using deep learning. With a detection rate between 60 and 70% and the F1 score above 0.8, the approach is certainly encouraging, but considering all other metrics, it is still not efficient enough for wider applicability. The main problem lies in the structure of UXO, which consists of different parts made of different materials and colours. All of this results in different spectral signatures, which were not properly accounted for in the CNN.

In further research, we will focus on improving our computational model, where we will try to account for the diversity of materials in the method. We will also investigate the implementation of a multi-class approach (i.e., including the UXO class). We will perform a full 3-fold cross-validation by testing the models.

References

Ammunition storage area explosions – EOD clearance. (2021). United Nations Office for Disarmament Affairs.

Bajić, M., & Bajić, M. (2021). Modeling and Simulation of Very High Spatial Resolution UXOs and Landmines in a Hyperspectral Scene for UAV Survey. *Remote Sensing*, *13*(5), 837. https://doi.org/10.3390/rs13050837

Bajić, M., Ivelja, T., Krtalić, A., Tomić, M., & Vuletić, D. (2013). The multisensor and hyper spectral survey of the UXO around the exploded ammunition depot, of the land mines test site vegetation. *Proceedings 10th International Symposium HUDEM, ISSN*, *9206*, 91–96.

Bajić, M., & Potočnik, B. (2024). Spectral Response of Two Hyperspectral Cameras for UXO Endmember Selection. *ROSUS 2024 - Računalniška Obdelava Slik in Njena Uporaba v Sloveniji 2024: Zbornik 18. Strokovne Konference*, 111–120. https://doi.org/10.18690/um.feri.1.2024.9

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018* (Vol. 11211, pp. 833–851). Springer International Publishing. https://doi.org/10.1007/978-3-030-01234-2_49

Guthrie, W. F. (2020). *NIST/SEMATECH e-Handbook of Statistical Methods (NIST Handbook 151)* [Dataset]. National Institute of Standards and Technology. https://doi.org/10.18434/M32189

iMMAP-IHF, Humanitarian Access Response—Monthly security incidents situation report, November 2018. (2019). iMMAP.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965

National University of Public Service, Hungary, & Ember, I. (2021). The role and the risks of explosive ordnance decontamination in Hungary. *Science & Military*, *16*(1), 32–42. https://doi.org/10.52651/sam.a.2021.1.32-42

Roberts, S., & Williams, J. (1995). *After the guns fall silent: The enduring legacy of landmines*. Veterans of America Foundation, Washington D.C., USA.

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1505.04597

Tuohy, M., Baur, J., Steinberg, G., Pirro, J., Mitchell, T., Nikulin, A., Frucci, J., & De Smet, T. S. (2023). Utilizing UAV-based hyperspectral imaging to detect surficial explosive ordnance. *The Leading Edge*, *42*(2), 98–102. https://doi.org/10.1190/tle42020098.1

Yan, K., Wang, X., Lu, L., Zhang, L., Harrison, A. P., Bagheri, M., & Summers, R. M. (2018). Deep Lesion Graphs in the Wild: Relationship Learning and Organization of Significant Radiology Image Findings in a Diverse Large-Scale Lesion Database. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9261–9270. https://doi.org/10.1109/CVPR.2018.00965