# Detection and Identification of Unexploded Ordnance using a Two-Step Deep Learning Methodology

Marian G. Craioveanu and Grigore Stamatescu

*Abstract*— **Localisation and disposal of unexploded ordnance (UXO) is a crucial task that can save the lives of both military personnel and civilians. In comparison to immediate post-war intervention situations, EOD (Explosive Ordnance Disposal) teams can now leverage emerging technologies based on computer vision architectures, mitigating the perceived risks associated with hands-on inspection of ammunition. The paper analyzes the use of new convolutional neural network architectures, in detection and identification of unexploded ordnance by combining specialised domain knowledge with computer vision models and methods. Additionally, it presents image preprocessing methods, research techniques, results, and conclusions. Moreover, we propose a complementary approach to previous research, often based on the interpretation of external sensor signals, representing the missing link in a comprehensive and extensive identification. Standardized metrics such as mean average precision, precision, recall, and F1-score are reported to evaluate the outcomes. Results, using the YOLOv8 architecture, achieve up to 80.8% mAP for the binary classification task (detection problem) and up to 90.6% mAP performance for the subsequent identification task. The work aims to offer a baseline study and new perspective on addressing a highly significant issue: mitigating by computer vision the risks associated with unexploded ordnance.**

## I. INTRODUCTION

According to US NOAA, unexploded ordnance (UXO) are "explosive weapons such as bombs, bullets, shells, grenades, mines, etc. that did not explode when they were employed and still pose a risk of detonation". These ominous remnants of past conflicts conflict pose a continuing threat to human health and safety and can also inflict significant economic and environmental damage. Classical UXO identification, removal and disposal methods have been typically marked by painstaking, often hazardous, manual labor and old technological approaches [1], influenced by the subjectivity and, sometimes limited, knowledge of human operators.

The emergence of computer vision and image processing and the accelerated adoption of high performance convolutional neural networks (CNNs), in diverse fields such as medicine [2] and industry [3], contributes to the development of intelligent systems for UXO detection and identification. Given the limited datasets of unexploded ordnances images, researchers have consistently sought to address this deficiency by resorting to the interpretation of signals from sensors, such as the analysis of GPR signals [4], or utilizing magnetometry [5]. Recent technological advancements, leveraging images captured by single-spectrum or multispectral cameras [6], or based on state-of-the-art explosive detection technologies within unexploded ordnance [7], present a novel, efficient yet costly approach as viable working systems, albeit time-consuming.

As UXOs are often present in challenging terrains, including mountain passes with constructed trenches, leading to unreliable internet connectivity. Essentially, bolstering resilience requires an approach that addresses the identification of unexploded ordnance both in a general context for immediate risk assessment using limited data from portable devices and in detail by class, enabling combat engineers to make informed decisions for neutralization or destruction. An additional complexity factor concerns the UXO that suffer modifications and improvisations with regard to their original design. Simultaneously, a binary classification model will be developed, serving as a benchmark for comparison with other architectures. Also, UXOs are frequently encountered in incomplete states, often found in challenging and semi-buried locations, including storage depots where multiple unexploded ordnances are clustered, a first task was to collect and curate a representative dataset for CNN model training and validation. This dataset encompasses both real unexploded ordnances and replicas and has been expertly labeled using image segmentation techniques by a specialist in the field of UXOs, supported by pretrained model.

The main contributions of the paper are two-fold:

- System architecture for UXO detection and identification which includes dataset collection, curation and image augmentation as preprocessing stage;
- Deep learning two-step CNN classification implementation and evaluation for baseline performance evaluation using the state-of-the-art YOLOv8 architecture.

The rest of the article is structured as follows. Section II introduces the scientific and technical background through an examination of prior related approaches using deep learning architectures for UXO classification. Section III describes the primary methods employed to enhance detection, encompassing the presentation of the dataset, acquisition module, image analysis, and implementation details. Section IV conducts an in-depth analysis of the achieved results, providing a detailed discussion of the implementation, emphasizing its novelty and offering standardized metrics. Section V concludes the paper by providing insights into the potential practical significance and prospects for expanding the approach in the field.

The authors are with the Department of Automation and Industrial Informatics, National University of Science and Technology Politehnica of Bucharest, Splaiul Independentei 313, 060042 Bucharest, Romania. marian.craioveanu@stud.acs.pub.ro, grigore.stamatescu@upb.ro. Financial publication support from the University POLITEHNICA of Bucharest, through the PubArt program is gratefully acknowledged.

## II. RELATED WORK

The classification of unexploded ordnance (UXO) left over from armed conflicts has a particular importance and has drawn the attention of nations worldwide and researchers [8]. While detection technologies have advanced, researchers face a challenge in the field of UXOs detection: low quality datasets. To address the lack of images, researchers have attempted to use additional data from external sensors, such as the analysis of signals from active external sensors. In some cases, researchers have tried to analyze signals from metal detectors or Ground Penetration Radar [9] [4]. In simple terms, signals known to originate from munitions are used, and researchers search for a pattern in the formation of these signals. Although promising, the results depend on the material of objects, their distance from the sensors, and do not take into account one of the most crucial aspects in the world of detection: working environments are non-uniform, mineralized, and urban areas may have "parasitic" signals that can alter the data. Also, it must be considered that in conflict zones, there are other objects alongside unexploded ordnance, objects that can mislead the sensors: military equipment, remnants from armored vehicles, shrapnel from explosions, and canned goods. Another problem is the cost, which is high, and the intervention method is slow, which is not favorable when dealing with a large quantity of munitions.

The challenge of implementing computer vision models for the detection of unexploded ordnance has been addressed by other researchers as well. In creating the dataset, some researchers [10] often used UXOs replicas that were photographed from various angles. A practical issue is that the variety of UXOs types, brightness, and the introduction of a diverse background were not considered, making it difficult to account for false-negative results. However, a significant positive aspect which they discovered is that by removing certain parts of the ammunition (e.g., fuse, stabilizer), CNN models manage to capture more relevant details and generalize much better even though the dataset is limited - a fact that is highlighted in their paper, even with a limited dataset efficient results can be achieved in the classification of ammunitions. In another study [11], researchers attempted to implement single-spectrum or multi-spectral cameras on drones for ammunition detection. Inference occurred on a server, with the drone acting as the client. The possibility of the server-client connection being unavailable was not considered, but still highlighting the advantage of using edge devices: risk and time reduction. Therefore, our paper will emphasize the relevance of edge devices, considering the server's unavailability for immediate inference. Furthermore, according to the mentioned article, the use of a multi-spectral camera did not significantly contribute in terms of costs and time. Hence, in our case, we prefer using RGB smartphone cameras. In another related work [12], authors tried to use open-source datasets to train models. They demonstrated that dataset classes do not necessarily need to have an equal number of images, as the opposite would not reflect reality. They also differentiated dataset classes well, based on ammunition types, providing concise descriptions and features. A problem in creating the dataset is that they used only horizontally positioned and cropped objects without background or with easily differentiable backgrounds. Also, the unexploded ordnance from the dataset are fully equipped, not reflecting a real-world scenario of initial identification. In the paper, the fuse is introduced as a class of ammunition, although it is more of a component of every unexploded ordnance. This leads to errors in accurately determining the class. Additionally, submunitions often represent mines part of cluster bombs, causing confusion for the model classification of submunitions, aviation bombs, or landmines. All related works offer important information and hypotheses that will be implemented in creating the dataset, selecting the working methodology, training, and implementation, taking into account the expertise of the authors of this novel paper and other publications addressing the optimization of object classification and localization in a general context, as will be discussed in Section III.

To provide a new avenue for research and to save lives, the introduction of a dataset with annotated images under varied conditions is proposed. The images were collected from both real missions with real UXOs, originating from armed conflicts in Romania, and with replicas - UXOs without active charge. Based on our knowledge, unexploded ordnance found in the dataset originates from multiple sources of manufacturing: Russia/Soviet Union, Germany, Great Britain, United States of America, France, Italy, Austro-Hungarian Empire, and Romania.

## III. METHODS AND DATASETS

### A. Methods

Methods must, above all, take into account the practical aspects of the field. Using field expertise, it can be stated that munitions are often found in altered, semi-buried conditions, in hard-to-reach areas - old battlefields. A diagram which emphasizes the client-server aspects is found in Figure 1. To address the detection of Unexploded Ordnance, multiple methods will be employed, ensuring there is a benchmark for comparison. The research process unfolds through iterative and meticulous steps, beginning with the development of a lightweight model designed for general detection. This model is engineered to operate seamlessly on edge devices, requiring minimal data consumption and functioning independently of client-server connections. The dataset undergoes augmentation to ensure uniform labeling as "unexploded ordnance," with a concurrent emphasis on integrating secondary performance metrics to optimize efficiency. Continuing the progression, a model is devised to leverage general recognition capabilities, now incorporating a client-server architecture. This model enhances detection capabilities, specifically targeting UXOs with manufacturing anomalies. It also facilitates the validation of initial results upon client-server reconnection. Simultaneously, it serves as a benchmark for evaluating the trade-offs between performance metrics variation and improved categorization accu-

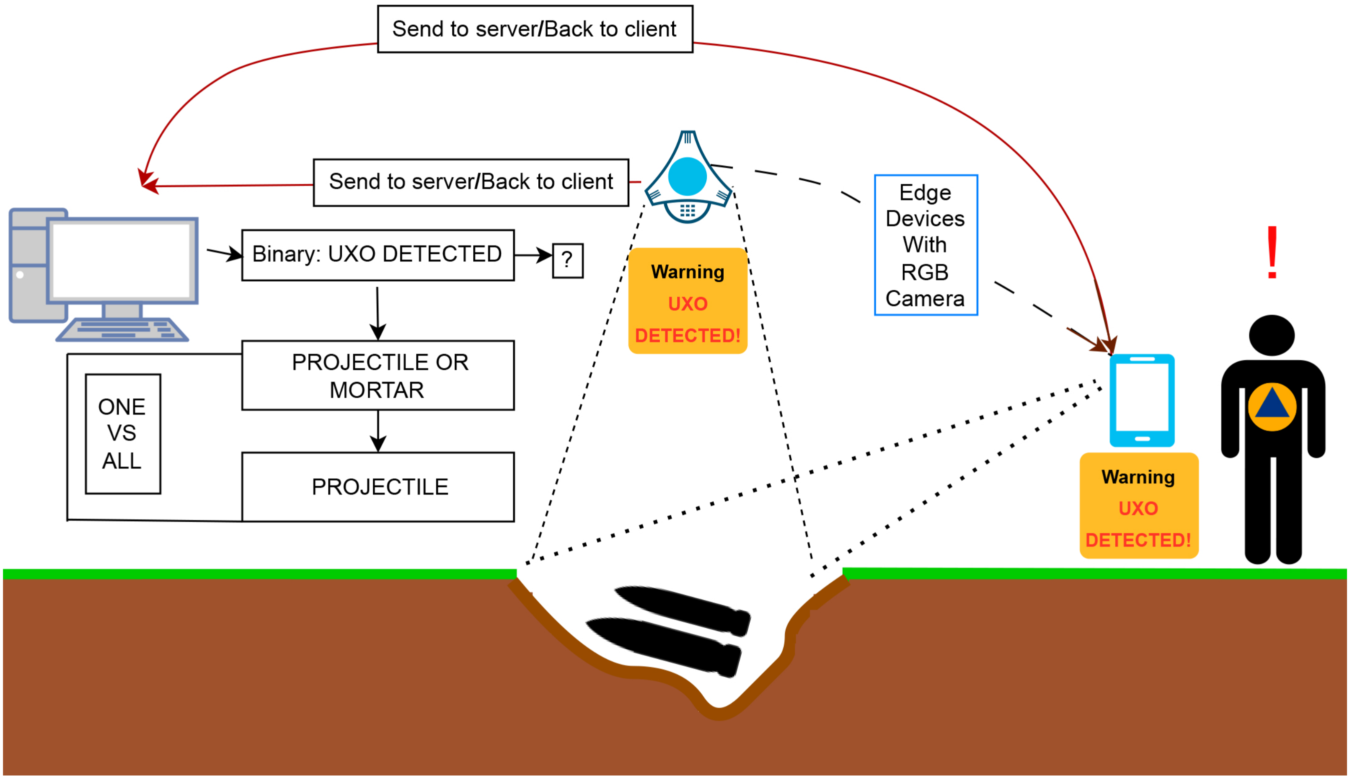


Fig. 1. System Architecture for the Detection and Identification of Unexploded Ordnance

racy for subsequent models. These models require access to high-computing servers and operate on datasets categorized into multiple classes. Special attention is given to exploring the effectiveness of ONE vs ALL techniques in addressing undersampling to enhance performance metrics. Throughout the research endeavor, all models are evaluated relative to a baseline binary classification model (UXO Detected/UXO Not Detected). This comparative analysis enables the tracking and quantification of progress percentage-wise, providing valuable insights into the advancement of UXO detection methodologies. As specified above, all models will be reported relative to a baseline model, which is based on binary classification (UXO Detected/UXO Not Detected), so that progress can be tracked and quantified percentage-wise. A diagram illustrating the approaches to the problem and how to implement our models in an end-to-end solution is shown in Figure 1. The process begins with the user, typically a combat engineer, employing a portable device, such as a phone or drone, referred to as an edge device, to capture an image of a suspected unexploded ordnance. In an offline setting, utilizing an optimized model, an immediate analysis is performed to evaluate the likelihood of the presence of ordnance. The results are then relayed back to the edge device interface. Upon establishing a connection between the edge client and server, the captured image is transmitted to the server. Here, a more sophisticated version of the previous model assesses the probability of ordnance presence. Concurrently, the server employs a model trained on the labeled dataset, categorized according to ordnance classes. This model addresses the fragmentation of the problem into multiple classes and specifications, enabling precise identification of the ordnance class. The server's response, confirming the prediction, is sent back to the client. If the response indicates potential ordnance presence, the image captured in step 1 is incorporated into the existing dataset for future enhancements and updates. The problem will be divided into several pieces, conducting an ablation study [13] to iteratively and analytically establish the ideal balance for a real-life situation. Therefore, based on the previously presented information, this article will focus on the optimization approach to the problem itself, a performance vs. compromise vs. practical application optimization. As the base arhitecture, You Only Look Once (YOLO) V8, a popular open-source architecture optimized for fast instance segmentation, will be utilized, considering the need for fast and accurate results in emergency situations [14] [15]. More exactly, it represents the fine-tuned YOLO V8X-segment model, with 71.8M parameters, convolutional neural network, cross-stage partial connections, Leaky ReLU activation function for hidden layers, 344.5 GFLOPs, with initialization hyperparameters established based on our domain knowledge, with subsequent tuning for 100 epochs. For edge device inference, a post-training float16 quantization on UXO/NON-UXO model is applied. The YOLO models are trained on the MS-COCO dataset, consisting of over 330,000 images, with classes number, $n_c$, of 80 (plus background), containing common objects in contexts, as standard best practice for these types of applications.

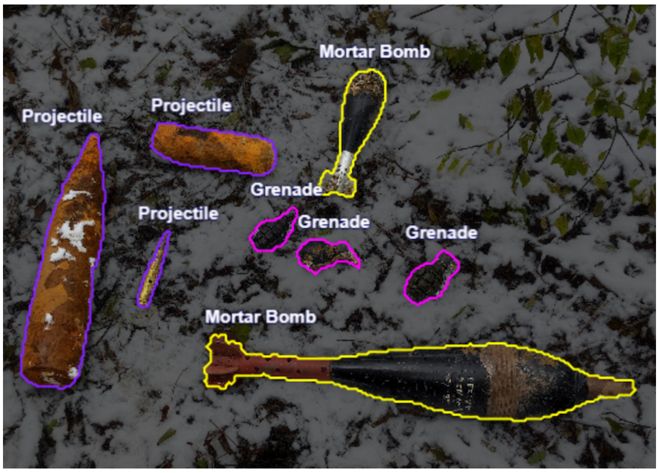


Fig. 2. Unexploded ordnances from our dataset

The initial layers of an architecture are responsible for learning general features e.g., lines, edges [16]. Fine-tuning represents a suitable approach in weight initialization, aiming to reduce time and computational resources consumption, providing a robust final model. Moreover, YOLO offers the advantage of being a one-shot detection architecture, thus reducing the need for computational resources both during training and inference. One-shot detection architectures are suitable for deployment on edge devices. [17] A representation of the pipeline architecture is shown in Figure 3.

### B. Dataset Description

The images from our proposed dataset were collected from pyrotechnic interventions, both with real unexploded ordnance from World Wars and with new inert munitions that adhere to all visual characteristics of real ordnance. Considering that munitions are often found in multiple quantities, stored in depots, image segmentation techniques were employed. Where the ordnance's condition allowed safe handling, it was placed in different environments to increase the training data volume. Based on other researchers' results [10], the decision was made to disassemble models where possible, removing certain components such as fuzes, to diversify the dataset. Labeling was performed by an expert in unexploded ordnance, ensuring high accuracy and consistency. Images were captured in various environmental conditions, lighting, and angles. In Figure 2, a sample with unexploded ordnances from the dataset is displayed. The dataset facilitates the advancement of machine learning algorithms dedicated to UXO detection and enhances sensor-based methods. It also explores applications in detecting objects camouflaged in the visible spectrum and serves as a platform for evaluating neural convolutional architectures in UXO detection. In Table I, the munition classes in the dataset are listed. The selected munition classes were chosen as they represent, in practice, the majority of types encountered in interventions. Additionally, in certain cases, these classes exhibit similar constructive features and can be easily confused. Given that munitions are often found in camouflaged environments, and considering that a false-negative error could result in physical harm to the Explosive Ordnance Disposal operator, a fourth category of images has been introduced—namely, the background class—to limit risks. From practical experience, it is more beneficial to eliminate false negatives in such situations, meaning to have false-positive cases, ensuring that no one is exposed to risks in a false-detection case.

TABLE I

MAIN CHARACTERISTICS OF OUR DATASET

| | |
|---|---|
| Classes | Mortar Bomb, Projectile, Grenade |
| Number of UXOs instances | 7880 |
| Medium file size | 2.5MB |
| Input size | 800px x 800px |
| Background images | 1330 |

In the case of the lite method, based on the general identification of ordnance without specifying the exact type, all classes of UXOs have been reassigned simply as "unexploded ordnance". It is essential to highlight the advantage provided by this method: the capability to detect improvised or incorrectly manufactured munitions. Toward the end of wars, when external pressure is high, there is a possibility that munitions were improvised, combining elements from one munition class with another. In Figure 4, a 60mm caliber explosive projectile from the Second World War is depicted. According to welding traces, the bomb stabilizer has been replaced with a tail from an aviation bomb, and the fuse has been changed, improvising the mortar bomb into an aviation bomb. It is clear that this explosive projectile is in an oxidized state, making it challenging to identify accurately.

A "copy-paste on new backgrounds" method was utilized as the basic augmentation technique, resizing labeled objects to optimal dimensions in various backgrounds, as recommended by the specialized literature [18]. Through authors' knowledge, the fact that munitions can be incomplete, analysis of instances and backgrounds, augmentation methods such as HSV variations, rotation and translation were established to capture even unfavorable cases. Additional augmentations such as rotate, zoom, shear, and flip random will be employed. Based on the fact that UXOs can be found mixed as classes, overlapped, or incomplete, a mosaic-type augmentation is also necessary, combining multiple images and instances to obtain new samples.

Fig. 4. Oxidized mortar bomb with aviation bomb elements

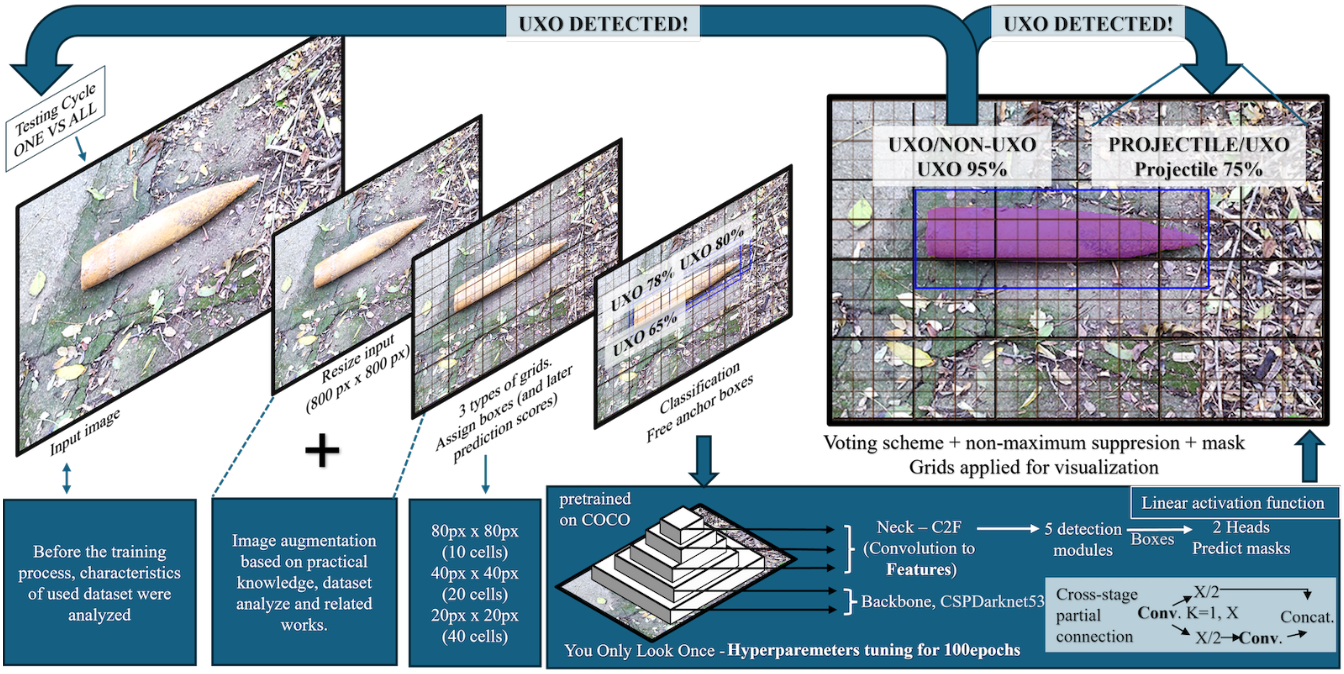


Fig. 3. CNN-based Deep Learning Pipeline for UXO Identification

The dataset and specific annotations have been formatted to adhere to the specific input of YOLO, while also providing the capability to be easily adaptable to other architectures.

## IV. RESULTS

For result analysis, the implementation has been carried out in Python, using as GPU a NVIDIA RTX 4090 24GB VRAM. The dataset distribution is 80% training data, 20% validation data. The metrics used for evaluation are of two types: primary and secondary. Primary metrics are mean average precision, precision, recall and F1-score. Secondary metrics related to computational aspects of the models.

Mean Average Precision is a metric in computer vision that evaluates algorithms by computing the average precision for each object class and then taking the mean of these values across all classes i.e. the area under the precision-recall curve:

$$mAP = \sum_{n}(Recall_n - Recall_{n-1})Precision_n \quad (1)$$

Precision measures the quality of the positive prediction of used model:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall measures the ability of the model to correctly identify all positive instances in the dataset:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

The F1-score is a metric commonly used in machine learning to balance precision and recall. It is the harmonic mean of precision and recall, providing a single score that combines both measures:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

Results of standardized metrics for the six types of models are presented in Table II. In Figure 6, more detailed results for top-performing model are presented.

TABLE II

PERFORMANCE OF FINE-TUNED YOLOV8X-SEG. AN ADDITIONAL BACKGROUND CLASS WAS INTRODUCED TO HANDLE FALSE-NEGATIVES.

| Classification | mAP | Precision | Recall | F1 |
|---|---|---|---|---|
| Binary | 80.8% | 85.9% | 87% | 86.44% |
| Binary-edge | 79.2% | 84.5% | 78.2% | 81.2% |
| Grenade/UXO | 90.6 % | 89.9% | 84.6 % | 87.16% |
| Mortar/UXO | 74.7% | 85.5% | 84.3% | 84.8% |
| Projectile/UXO | 73.6% | 89% | 82.2% | 85.4% |
| All classes | 75.5% | 89.1% | 84.3% | 86.63% |

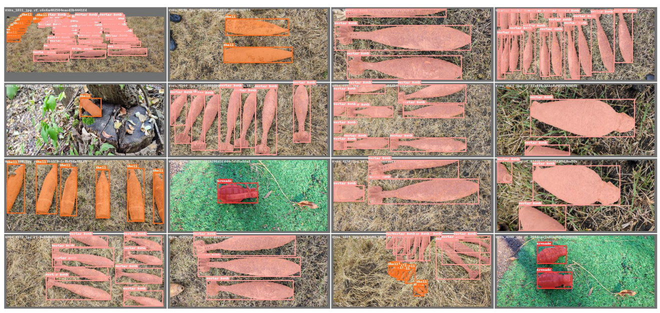

Fig. 5. Validation batch of multi-class model

The least performing model, with a classification mAP of 73.6%, is ”Projectile/UXO”, while the top-performing model, with a mAP of 90.6%, is ”Grenade/UXO”. The ”Projectile/UXO” model performs 7.2 percentage points lower than the ”Binary” model, which has a mAP of 80.8%. Conversely, the ”Grenade/UXO” model outperforms the ”Binary” model by 9.8 percentage points. An example of validation batch with unexploded ordnances, based on multi-class model (grenade, mortar bomb, shell, background) is shown in Figure 6.

Secondary metrics are represented by preprocess, inference and postprocess times per image (Table III). The disk usage for each model is 137MB. Models were tested on Nvidia RTX 4090. For Binary-Edge model, test was done on Qualcomm Snapdragon 780G (Post-training float16 quantization). Considering that the initial inference of a model requires time for loading the model, and preprocessing and post-processing steps depend on the input image and the model's output results, the inference was applied to all images in the test dataset, and subsequently, the arithmetic mean of the resulting times was calculated. The Binary (UXO/NON-UXO) model is the fastest, while the edge model is 4.24 times slower than the fastest model. The results show that the models demonstrate good inference speed even on edge devices, being conditioned by the computational resources available.

TABLE III
PERFORMANCE IN TERMS OF SECONDARY METRICS [MS]. THE PERCENTAGE FROM TOTAL TIME IS CALCULATED RELATIVE TO THE FASTEST MODEL.

| Model | Preprocess | Inference | Postprocess | Total |
|---|---|---|---|---|
| Binary | 2.1 | 11 | 2 | 15.1 |
| Binary-edge | 4.3 | 53.41 | 6.4 | 64.11 |
| Grenade/UXO | 2.8 | 11.5 | 2.7 | 17 |
| Mortar/UXO | 3 | 12.5 | 3 | 18.5 |
| Projectile/UXO | 2.4 | 11.5 | 2.4 | 16.3 |
| All classes | 2.3 | 15 | 1.1 | 18.4 |

## V. CONCLUSIONS

This study presented the precise and objective identification of unexploded ordnance (UXO) using computer vision. By harnessing advanced technologies, particularly through the utilization of neural network architectures and edge computing devices for emergency scenarios, combined with a new and curated dataset, this research demonstrates a promising leap forward in UXO detection and mitigation strategies. The models exhibit highly encouraging results, indicating their potential for real-world deployment and future studies.

The GRENADE+UXO model's superior performance in general munitions identification compared to binary UXO/NON-UXO classification highlights the efficacy of nuanced approaches in addressing complex scenarios. Furthermore, the adoption of standardized evaluation metrics such as mean average precision, precision, recall, and F1-score provides robust quantification of model performance, facilitating informed decision-making in UXO management efforts.

Two-step deep-learning methodology developed, the dataset, and the results of the models open a new path in Unexploded Ordnance identification, while also complementing existing methods. In essence, this study presents a transformative paradigm shift in UXO detection methodologies, offering a holistic approach that integrates cutting-edge technologies with domain-specific expertise, following future research, including by the authors, regarding optimization methods, as well as the analysis and dissemination of the dataset for the scientific community.

## REFERENCES


[1] C. Bruschini, "Commercial systems for the direct detection of explosives for explosive ordnance disposal tasks," *Subsurface Sensing Technologies and Applications*, vol. 2, pp. 299–336, 2001.

[2] D. Popescu, M. Deaconu, L. Ichim, and G. Stamatescu, "Retinal blood vessel segmentation using pix2pix gan," in *2021 29th Mediterranean Conference on Control and Automation (MED)*, 2021, pp. 1173–1178.

[3] S. Rosioru, G. Stamatescu, I. Stamatescu, I. Fagarasan, and D. Popescu, "Deep learning based parts classification in a cognitive robotic cell system," in *2022 26th International Conference on System Theory, Control and Computing (ICSTCC)*, 2022, pp. 403–408.

[4] F. Lombardi, M. Lualdi, F. Picetti, P. Bestagini, G. Janszen, and L. A. Di Landro, "Ballistic ground penetrating radar equipment for blast-exposed security applications," *Remote Sensing*, no. 4, p. 717, 2020.

[5] S. D. Billings, "Discrimination and classification of buried unexploded ordnance using magnetometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 6, pp. 1241–1251, 2004.

[6] S. Cho, "Ai-based uxo detection using suas equipped with a single-or multi-spectrum eo sensor," Ph.D. dissertation, Monterey, CA; Naval Postgraduate School, 2021.

[7] J. Kapoor and G. Kannan, "Landmine detection technologies to traceexplosive vapour detection techniques," *Defence Science Journal*, vol. 57, no. 6, p. 797, 2007.

[8] "Horizon 2023: Identify, inspect, neutralise unexploded ordnance (uxo) at sea," https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-cl3-2023-bm-01-02, [Accessed 10-11-2023].

[9] X. Núñez-Nieto, M. Solla, P. Gómez-Pérez, and H. Lorenzo, "Gpr signal characterization for automated landmine and uxo detection based on machine learning techniques," *Remote sensing*, no. 10, 2014.

[10] A. Lebbad, G. Clayton, and C. Nataraj, "Classification of uxo using convolutional networks trained on a limited dataset," in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2017, pp. 1098–1101.

[11] S. Cho, J. Ma, and O. A. Yakimenko, "Aerial multi-spectral ai-based detection system for unexploded ordnance," *Defence Technology*, vol. 27, pp. 24–37, 2023.

[12] G. Begkas, P. Giannakeris, K. Ioannidis, G. Kalpakis, T. Tsikrika, S. Vrochidis, and I. Kompatsiaris, "Automatic visual recognition of unexploded ordnances using supervised deep learning," in *Proc. of the 2022 Intl. Conference on Multimedia Retrieval*, 2022, pp. 286–294.

[13] R. Meyes, M. Lu, C. W. de Puiseau, and T. Meisen, "Ablation studies in artificial neural networks," *arXiv preprint arXiv:1901.08644*, 2019.

[14] H.-J. Jeong, K.-S. Park, and Y.-G. Ha, "Image preprocessing for efficient training of yolo deep learning networks," in *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, 2018, pp. 635–637.

[15] R. Huang, J. Pedoeem, and C. Chen, "Yolo-lite: a real-time object detection algorithm optimized for non-gpu computers," in *2018 IEEE international conference on big data (big data)*. IEEE, 2018.

[16] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*. Springer, 2014, pp. 818–833.

[17] J. Dai, "Real-time and accurate object detection on edge device with tensorflow lite," in *Journal of physics: conference series*, vol. 1651, no. 1. IOP Publishing, 2020, p. 012114.

[18] G. Ghiasi, Y. Cui, A. Srinivas, R. Qian, T.-Y. Lin, E. D. Cubuk, Q. V. Le, and B. Zoph, "Simple copy-paste is a strong data augmentation method for instance segmentation," in *Proc. of IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2918–2928.