11th International Conference on Electrical, Electronics and Computer Engineering (IcETRAN), Nis, 3-6 June 2024

Subsurface Object Recognition using YOLOv8

1st Primož Smogavec
Faculty of Electrical Engineering and Computer Science,
University of Maribor,
Maribor, Slovenia
primoz.smogavec@um.si

2nd Božidar Potočnik
Faculty of Electrical Engineering and Computer Science,
University of Maribor,
Maribor, Slovenia
bozidar.potocnik@um.

3rd Dušan Gleich
Faculty of Electrical Engineering and Computer Science,
University of Maribor,
Maribor, Slovenia
dusan.gleich@um.si

***Abstract*—This paper presents the implementation and evaluation of the YOLOv8 detection method within the domain of ground-penetrating radar (GPR) imagery analysis. YOLOv8, an evolution of previous YOLO models, is employed for object detection tasks, providing efficient and accurate results. The study focuses on detecting and classifying three distinct objects: corner reflector, anti-tank (AT) mine, and glass bottle. Training data preparation involves annotating radar images and splitting them into training, validation and testing sets. The YOLOv8 model is trained on labeled data, and its performance is evaluated based on precision, recall, and F1-score metrics. Results demonstrate high accuracy in detecting and classifying objects, with minimal false positives. Further testing in real-world scenarios is suggested to assess model robustness under varied environmental conditions and object densities.**

Keywords—ground-penetrating radar, object classification, YOLOv8, mine detection

I. Introduction

In today's times, one of the still prevalent inhumane military tactics is the laying of minefields. According to the latest information [1], mines or other explosive devices are still present in 59 countries, including the former Yugoslav republics: Croatia, Serbia, and Bosnia and Herzegovina. According to the regulations of the United Nations Department of Humanitarian Affairs (UNDHA), 99.6% of mines and unexploded ordnance must be safely removed for the civilian area to be considered mine-free. This significantly differs from military demining, where only a few paths need to be cleared for vehicle passage, with little attention paid to surrounding areas. Humanitarian demining is typically done manually, making this operation extremely risky and time-consuming. Various solutions have been proposed in the literature to facilitate demining operations and increase safety. This paper focuses on detection and classification of objects once a B-scan of the scene has been obtained.

Several methods have been proposed in the literature. Authors in [2] introduced a new method for detecting and extracting parabolas from ground-penetrating radar images. They introduced a data filtering method using wavelet analysis and Gabor filtering. They proposed an automated method for detecting and classifying hyperbolas. The classification methodology was based on the analytic hierarchy process method, introducing a classification coefficient and considering the weights of proposed conditions and coefficient weights. The proposed method enabled a reduction in object detection time. Authors in [3] present an automated approach for detecting parabolas and their peaks in ground-penetrating radar images. The article introduced a two-step method. Firstly, the Fast-RCNN model was used for parabola detection, and secondly, image processing methods were employed for extracting the parabola peaks. The method was evaluated on both real and simulated images. The main obstacle for successful parabola detection using the proposed method on real images was the presence of noise. Challenges of detecting and locating parabolic objects in GPR images were addressed in [4]. The YOLOv7 neural network was used, enabling real-time object detection. It was employed in the first step of the proposed method. In the second step, a two-stage curve fitting method was utilized. Evaluation of the YOLOv7 model showed an 11.1% improvement in detection accuracy and an 18.2% improvement in detection speed compared to the YOLOv5 model, using the same hardware and data.

This paper presents the implemented offline method for object detection called YOLOv8, successor to the methods YOLOv7 and YOLOv5. YOLOv8 is the latest version of the object detection method named YOLO (You Only Look Once), first released in May 2023. The YOLOv8 method is developed based on YOLOv5 and represents an update to this approach. YOLOv8 is based on the architecture of a deep convolutional neural network (CNN). In section II creation of a custom dataset is presented utilizing a modular Stepped-Frequency Continuous Wave (SFCW) Radar. Section III presents training of the YOLOv8 model and analyses metrics such as precision, recall, and F1-score. In section IV YOLOv8n model results are presented on a test image set. The paper concludes with Section V, providing suggestions for further testing.

II. Creating a Custom Dataset

For the purpose of training YOLOv8 model we have gathered our own data using a modular SFCW radar described in [5], which operates within the 500MHz to 2.5GHz frequency band with a fixed frequency step of 10MHz. All measurements were performed in a controlled environment shown in Figure 1. We have limited the number of objects of which radar signatures have been obtained to three: corner reflector, AT (anti-tank) landmine, and full glass bottle depicted in Figure 2.

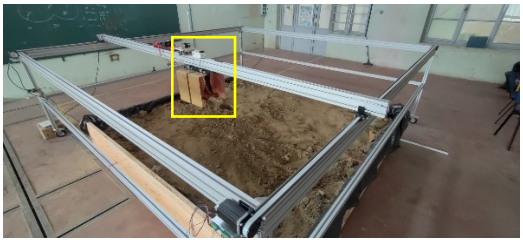

Figure 1: Test environment with 2D rail system and soil box underneath for data gathering. Radar position is marked in yellow.

Supported by NATO SPS under Grant No. G5953

(a) (b) (c)

Figure 2: Objects from which radar signatures were acquired: Corner reflector (a), AT mine (b), and full glass bottle (c)

Radar images were captured with a single object on the surface or buried beneath the surface at depths ranging from 0 to 30 cm. We captured 200 radar images of AT mine, 100 images of corner reflector, and 100 images of glass bottle. The images were divided into three groups: training images, validation images, and test images, which were used for the final evaluation of the neural network's learning performance after it had already been successfully trained. The latter group of images is separate from the training set and is not used during the learning process. The training set consisted of 165 images of AT mines, 75 images of corner reflectors, and 75 images of full glass bottles, validation set consisted of 30 AT images, 20 corner reflector images and 20 bottle images, while the test set consisted of 5 images for each class. All images were 640x480 pixels in size and were annotated using CVAT tool [6].

III. YOLOv8 MODEL TRAINING AND EVALUATION

YOLOv8 operates by first dividing the input image into a grid of cells. For each cell, YOLOv8 predicts a set of bounding boxes along with class probabilities for each bounding box. YOLOv8 then utilizes the non-maximum suppression (NMS) algorithm, which filters overlapping bounding boxes and selects the most probable bounding box for each object in the image. YOLOv8 supports various operations such as object detection, image segmentation, position estimation, tracking, and object classification. This model is available in five versions: YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large), and YOLOv8x (extra-large). We have chosen the YOLOv8n model for our application.

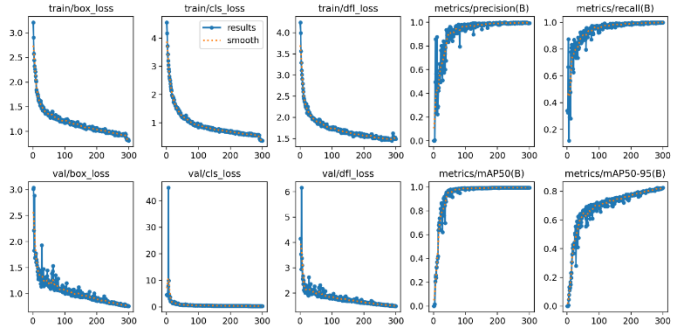

Figure 3: Training results of YOLOv8n model after 300 epochs with custom data.

In Figure 3, we observe the training results of the YOLOv8n model after 300 epochs. Graphs labeled with (B) illustrate bounding box metrics. In the first three loss functions, we notice that there is still some room for improvement, which could be achieved by increasing the number of epochs or by selecting a larger model or by finely tuning the hyperparameters of the network. Nevertheless, we observe that the precision of predictions and the recall are close to 1, so we opt not to retrain.

Figures 4-6 depict precision-recall curves and F1 metrics at various confidence score thresholds. In our application, detecting and distinguishing mines from other objects is crucial, with minimizing false negatives being paramount, while false positives are not as critical from a safety perspective. The precision-confidence curve seen in Figure 4 illustrates how precision varies with different confidence thresholds. Ideally, we aim for high precision at all confidence levels. It gives us the ratio of correctly predicted positives to all predicted positives. The equation for calculating precision is as follows:

$$precision = \frac{true\ positives}{true\ positives + false\ positives} \quad (1)$$

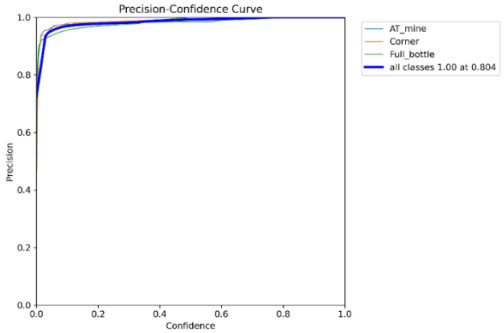

Figure 4: Precision-Confidence curve

The recall-confidence curve in Figure 5 illustrates recall at various confidence thresholds. A high recall across the entire range is desirable. Recall is the ratio of correctly predicted positives to all actual positives. The equation for calculating recall is:

$$recall = \frac{true\ positives}{true\ positives + false\ negatives} \quad (2)$$

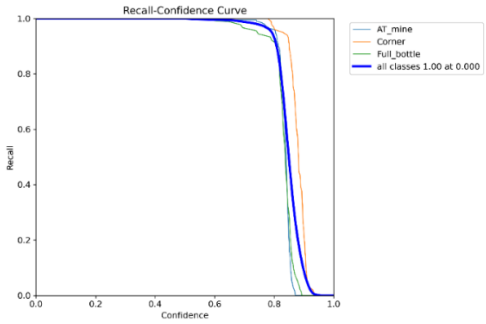

Figure 5: Recall-Confidence Curve

The F1-confidence threshold curve depicts the F1 score (the harmonic mean of precision calculated by equation (1) and recall (2)) at various confidence thresholds (Figure 6). A higher peak signifies better model performance. The F1 score is calculated using the following equation:

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall} \quad (3)$$

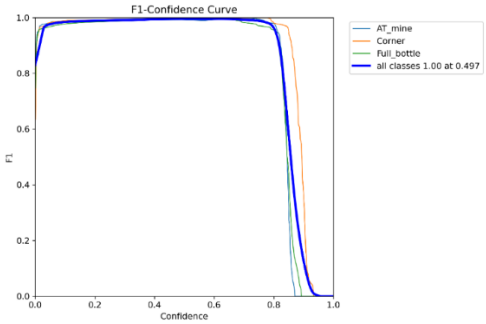

Figure 6: F1-Confidence Curve

IV. Results

Figures 7-10 show the results of the YOLOv8n model on the test set of images. The bounding boxes in green along with their corresponding labels and confidence scores in red indicate the detected objects. All predicted objects in the test image set are identified correctly and have confidence scores between 0.8 and 0.95. However, there are instances where a detected object in the image is not actually present in the scene, as seen in Figure 10, indicating a false positive. It's noteworthy that the confidence score for such detections is quite low, specifically 0.31 for instance shown in Figure 10. Such detections can be filtered out by introducing a minimum confidence score threshold.

Table I shows results of YOLOv8n on the test set images. As mentioned before each class consists of 5 test images. Images indexed from 1-5 represent AT mine, images 6-10 depict corner reflector, and images 11-15 showcase full glass bottle results. Each image contained only one object. We can identify two instances where an object was detected incorrectly, at image index numbers 12 and 13. The mean confidence scores for detecting an AT mine, corner reflector, and full glass bottle were 0.858, 0.902, and 0.824 respectively.

TABLE I: Results of YOLOv8n with test dataset

Test Image Index	No. Of Detected Objects	Confidence Score		
		AT mine	*Corner Reflector*	*Full Bottle*
1	1	0.82	/	/
2	1	0.86	/	/
3	1	0.84	/	/
4	1	0.87	/	/
5	1	0.9	/	/
6	1	/	0.9	/
7	1	/	0.88	/
8	1	/	0.91	/
9	1	/	0.95	/
10	1	/	0.87	/
11	1	/	/	0.82
12	2	/	0.27	0.81
13	2	/	0.31	0.87
14	1	/	/	0.8
15	1	/	/	0.82

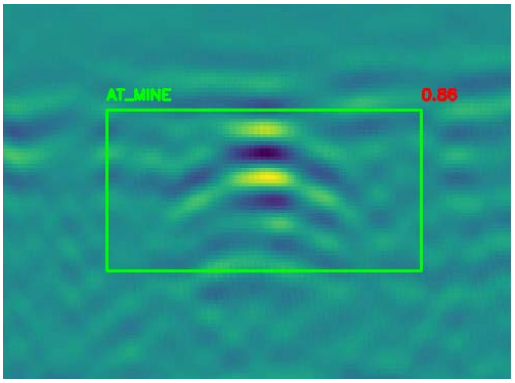

Figure 7: Correct classification of AT landmine in the test image.

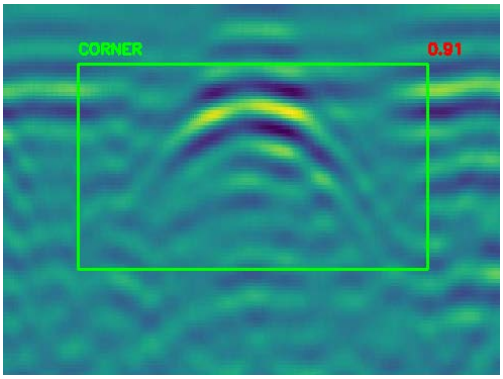

Figure 8: Correct classification of corner reflector in the test image.

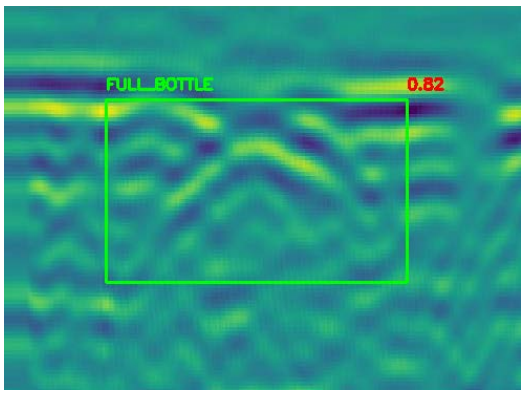

Figure 9: Correct classification of full glass bottle in the test image.

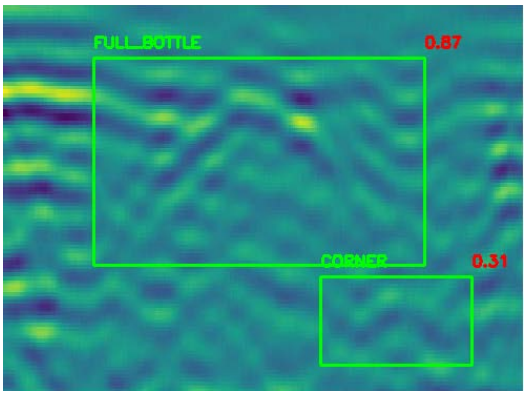

Figure 10: Correct classification of full glass bottle but incorrect detection of corner reflector in the test image.

V. Conclusions

It was demonstrated within this study that YOLOv8n model performs relatively well and provides accurate predictions. One factor contributing to the model's high effectiveness is the controlled environment in which the data were captured. Furthermore, only one object was consistently present at a specific location in GPR images. Further testing of the system would be necessary in real-world scenarios where multiple objects are present in the image, some of which may be buried at unknown locations. Additionally, the influence of soil type and soil moisture percentage on the prediction accuracy needs to be examined. The images were captured in soil composed of a sand-clay mixture in a ratio of 9:1 with a relative humidity of 0%. Favourable results can also be attributed to the small number of object classes and to objects themself which significantly differ in shape and have substantial differences in dielectric constant which makes them more distinguishable. In order to facilitate reproducibility and further research in this area, the training dataset used in this study can be made available upon request. Interested parties are encouraged to contact the authors for access to the dataset.

References

[1] "Landmine Monitor 2020." [Online]. Available: https://www.the-monitor.org/en-gb/reports/2020/landmine-monitor-2020.aspx

[2] K. Onyszko and A. Fryśkowska-Skibniewska, "A New Methodology for the Detection and Extraction of Hyperbolas in GPR Images," *Remote Sensing*, vol. 13, no. 23, Art. no. 23, Jan. 2021, doi: 10.3390/rs13234892.

[3] D. Dewantara and W. W. Parnadi, "Automatic Hyperbola Detection and Apex Extraction Using Convolutional Neural Network on GPR Data," *J. Phys.: Conf. Ser.*, vol. 2243, no. 1, p. 012027, Jun. 2022, doi: 10.1088/1742-6596/2243/1/012027.

[4] C. Zhu and H. Ye, "A Modular Method for GPR Hyperbolic Feature Detection and Quantitative Parameter Inversion of Underground Pipelines," *Remote Sensing*, vol. 15, no. 8, Art. no. 8, Jan. 2023, doi: 10.3390/rs15082114.

[5] P. Smogavec, B. Pongrac, and D. Gleich, "Evaluation of Compact and Modular SFCW GPR Systems for Detecting Buried Objects," in *2023 30th International Conference on Systems, Signals and Image Processing (IWSSIP)*, Jun. 2023, pp. 1–5. doi: 10.1109/IWSSIP58668.2023.10180244.

[6] "CVAT annotation software." [Online]. Available: https://www.cvat.ai/