# Development of Near Real Time Target Detection and Identification System for Ground Combat Vehicles via Artificial Intelligence

Reşat Ali Tütüncüoğlu [1], Selda Güney [2,*]

## Abstract

Modern military intelligence systems are advancing rapidly, particularly in the classification of combat vehicles in battlefield environments. Accurately identifying the capabilities and vulnerabilities of these vehicles is crucial for developing effective tactics and operational strategies. The integration of artificial intelligence, specifically through image processing, pattern recognition, and deep learning techniques, has greatly enhanced the feasibility of military vehicle classification. This study employs the You Only Look Once (YOLO) algorithm, particularly the YOLOv8 model, which is optimized for mobile applications and has shown high performance in evaluations. Enhancements include the incorporation of a Squeeze and Excitation (SE) block, improving detection accuracy. The research emphasizes the importance of detecting detailed target characteristics using images collected from electro-optical systems and aims to bolster decision support processes in target management systems. The development of a deep learning-based Target Detection and Identification (TDI) system is outlined, encompassing specialized targeted data acquisition, pre-processing, image partitioning, feature extraction, classification, and recommendation generation through decision support matrices. This model significantly improves image processing and object detection capabilities. Notably, it facilitates passive identification of combat vehicles using existing electro-optical devices without the need for additional hardware, while estimating distances discreetly. The system delivers real-time predictions, achieving a mean Average Precision (mAP) of 88.38%, thereby facilitating informed decision-making and risk reduction in military operations.

**Keywords:** combat military vehicles, image processing, classification, detection, assessment, attention mechanism, decision support matrix.

Received:03.11.2025 Accepted: 20.12.2025 Research Article

[1] HAVELSAN Inc., Ankara Türkiye
[2] Department of Electrical-Electronics Engineering, Başkent University, Ankara, Türkiye
* Corresponding author.
✉ seldaguney@baskent.edu.tr

**Cite this article as:** Tütüncüoğlu, R. A. & Güney, S. (2026). Development of near real time target detection and identification system for ground combat vehicles via artificial intelligence. *Journal of Defence and Security Industry: Strategy and Technology 1*, 96-115.

## 1. Introduction

Combat vehicles are specialized military assets designed for operational effectiveness in armed conflicts. These vehicles exhibit a wide range of designs and dimensions and are classified according to their primary functions and characteristics. A thorough comprehension of the advantages and disadvantages of different combat vehicle types is critical for the development of effective military strategies and tactics. This understanding enables military leaders to leverage the advantages of each vehicle type while mitigating their inherent limitations, thereby enhancing the probability of favourable outcomes in combat scenarios.

The classification of enemy combat vehicles, which involves the identification and assessment of adversary capabilities on the battlefield, is essential for several reasons. Key considerations in this process include target prioritization, informed tactical decision-making, effective allocation of firepower, troop protection, intelligence gathering, reconnaissance missions, and the maintenance of situational awareness.

The integration of Artificial Intelligence (AI) models for the classification of enemy combat vehicles substantially enhances military operational capabilities. AI-driven classification systems provide rapid and accurate intelligence, which improves decision-making processes and reduces risks to personnel. These systems act as force multipliers, significantly impacting the efficacy of military operations in contemporary combat environments. By augmenting situational awareness and operational planning, AI technologies can profoundly influence the outcomes of military engagements.

Object classification represents a significant area of study within computer vision and machine learning, focused on categorizing objects found in images into distinct and meaningful groups (Szeliski, 2010). Convolutional Neural Networks (CNNs) can be used at the core of applications ranging from photo annotation to popular self-driving studies. They are working intensively behind the scenes in everything from agriculture to military security. There is an increasing demand of object detection and automatic vehicle type recognition systems in the last 10 years as a result of the development of the cheaper sensor and camera technologies (Berwo et al., 2023). Because human operators alone cannot efficiently analyse large volumes of visual data and this data has to be classified very fast and correlated to evaluation.

A disparity exists both object detection and classification algorithms in the field of computer vision. Object detection algorithms are designed to identify and delineate specific objects within an image by creating bounding boxes around the objects of interest. In this context, these algorithms are tasked with accurately positioning the objects within the image frame. Notably, object detection algorithms may generate multiple bounding boxes, each representing a distinct object within the image. It is crucial to highlight that the exact number of objects and, consequently, bounding boxes are unknown in advance, adding complexity to the detection process.

Addressing this type of issue involves extracting various regions of interest from the image and employing a Convolutional Neural Network (CNN) to determine the presence of objects within those regions. However, a significant challenge associated with this approach is that the objects of interest may be located at different spatial coordinates within the image and possess varying aspect ratios (Gandhi, 2025).

Several algorithm techniques have been suggested and investigated by different researchers to detect and classify vehicle types. There are one-stage and two-stage target detection algorithms. Two-stage CNN algorithms are Region-based CNN, Fast R-CNN and Faster R-CNN (Hu, 2014). The one-stage algorithm is the You Only Look Once (YOLO) algorithm. It has ability to calculate the classification and regression (Liu, 2017).

Except for YOLO algorithm, all previously described object detection algorithms use regions to localize the object in the image. The network does not analyse the entire image; rather, it focuses on regions of the image that exhibit a high probability of containing the object of interest. In the YOLO framework, a single convolutional network is responsible for predicting both the bounding boxes and the class probabilities associated with these boxes (Redmon et al., 2016). YOLO offers inherent advantages in terms of processing speed, superior intersection over union for bounding boxes, and enhanced prediction accuracy when compared to other real-time object detection systems.

YOLO is significantly faster than other object detection algorithms; however, it has a limitation in that it struggles to accurately detect small objects within the image (Redmon and Angelova, 2014). This is because of the algorithm's spatial limitations. YOLOv8 provides a variety of models tailored for specific tasks in computer vision. These models address a range of needs, including object detection and more complex tasks such as instance segmentation, pose and key point detection, oriented object detection, and classification (Jocher et al., 2023). Experimental results indicate that the optimized YOLOv8 model enhances the detection of small objects with greater accuracy (Huadong et al., 2024).

Challenging environmental conditions—variable vehicle sizes, heterogeneous backgrounds, and adverse weather—complicate the detection of ground combat vehicles, motivating continual improvements in deep learning methods. Building on the YOLOv8 framework, Wu and Dong (2023) propose YOLO-SE to address multi-scale and small-object detection: the architecture incorporates a lightweight convolutional module (SEConv) that reduces parameter count and accelerates inference, replaces the original detection head with a redesigned prediction head, and adds an auxiliary detection head specialized for different object scales, collectively yielding improved detection accuracy. Additionally, the advantages of deep learning-based object detection algorithms, particularly YOLOv8, in real-time applications should be considered. In the study by Gang Cheng and colleagues, a lightweight model named SGST-YOLOv8 was developed, achieving an accuracy rate of 88.6% in detecting pedestrians and vehicles in complex environments (Cheng et al., 2024). Such accuracy is critical for detecting ground combat vehicles, as rapid and precise detection can directly impact the success of military operations.

The quality of the datasets used for detecting ground combat vehicles is also crucial. Datasets that account for various types, sizes, and environmental conditions of vehicles serve a crucial function in training deep learning models. For instance, in a study by Xin Zhang and colleagues, the YOLOv8 algorithm was utilized for vehicle and pedestrian detection in complex traffic environments, integrating a Coordinate Attention (CA) module to enhance feature extraction (Zhang et al., 2024). Such innovations contribute to achieving higher success rates in detecting ground combat vehicles.

Cui et al. (2024) developed an optimized YOLOv7 algorithm that utilizes a hybrid convolutional fusion attention mechanism to tackle complex background challenges in optical remote sensing images. This algorithm effectively addresses issues such as the dense clustering of small targets, substantial variations in target scales, and intricate backgrounds. These improvements in target detection enable the integration of distance measurement algorithms, allowing for the development of more comprehensive systems.

Hu at al. (2018) proposed a new architectural unit called the Squeeze-and-Excitation (SE) block to enhance the representational power of networks. This block adaptively adjusts the responses of features across channels by explicitly modelling the relationships between them. Their research demonstrates that stacking these SE blocks can lead to the creation of SENet architectures that generalize exceptionally well on challenging datasets. Notably, it has been stated that SE blocks enhance the performance of existing deep architectures while incurring very little additional computational cost.

Jocher et al. (2022) conducted a study using the YOLOv5 algorithm to estimate the distance of certain objects. The system is based on their box ratios. This study aims to accurately estimate the distances of objects in the image based on their size and ratios using the object detection capabilities of YOLOv5.

Near real-time object distance estimation and measurement using deep learning Canny Edge Algorithm and OpenCV has been achieved for small objects and small distance (1m≤ X) (Basavaraj et al., 2023).

Lan et al. (2023) developed the YOLOv8 network model for distance measurement operations and parsed the optimized data, and incorporated it into the binocular distance system, proposing a fusion sensory distance model. In addition, the model provides high portability on executable embedded development boards by edge detection.

As a summary, over the past decade, research has shifted from two-stage region-proposal detectors (e.g., R-CNN variants) - which offer strong localization accuracy at the cost of throughput -toward single-stage, YOLO-type architectures that unify localization and classification for real-time operation (Liu, 2017; Redmon et al., 2016). To compensate for YOLO's known small-object and multi-scale limitations, recent studies integrate lightweight attention and multi-scale modules: SE and Coordinate Attention modules improve channel and positional feature discrimination with minimal overhead (Hu et al., 2018; Zhang et al., 2024),

and YOLO-SE augments YOLOv8 with SEConv plus an auxiliary detection head to boost small-object performance (Wu & Dong, 2023). Lightweight variants (e.g., SGST-YOLOv8) prioritize edge deployment and maintain high accuracy in complex scenes but often omit full latency and dataset details (Cheng et al., 2024), while hybrid convolutional-fusion designs target dense, scale-variant scenarios such as remote sensing (Cui et al., 2024). For metric ranging, monocular box-ratio methods are simple but depend on prior size models and have limited absolute accuracy (Jocher et al., 2022); edge-based geometry pipelines perform well at very short ranges (~1 m) (Basavaraj et al., 2023); and binocular/stereo fusion yields better metric estimates at the expense of hardware/calibration complexity (Lan et al., 2023). Key trade-offs across these works are accuracy versus latency, small-object sensitivity versus computational cost, and metric fidelity versus sensor/calibration requirements; remaining gaps include limited tactical datasets, underreported domain-shift evaluations, and sparse end-to-end embedded system validations.

In our study, we extract the features of objects using the improved by SE module added to YOLO algorithm. By employing the feature transfer learning method, the newly obtained features are utilized as input for distance calculations. This method utilizes the knowledge and experiences of models trained on large datasets to be applied in new tasks that work with less data. The trigonometric calculation method we employed enables us to achieve measurements that are both very simple and rapid. The need for this type of technology within this particular segment is to identify opposing force vehicles without revealing our presence. Moreover, a thorough review of the literature reveals that there are no studies of this kind specifically focused on the military tactical domain.

## 2. Methodology

This section describes the methodology used for the targeted system. It thoroughly investigates the architecture (Figure 1) and aims to identify the most effective approach to achieve the desired outcome.

Initially, detailed information about the targets is obtained using a single-stage detector algorithm powered by SE module. Subsequently, the target's distance is calculated using an algorithm that employs the features extracted through the feature transfer learning method and the prepared data tables. Following the estimation of the target distance, this information is interpreted in conjunction with the generated decision support matrices, which provide recommendations to the user. This approach allows the model to demonstrate a high level of generalization while requiring less data.

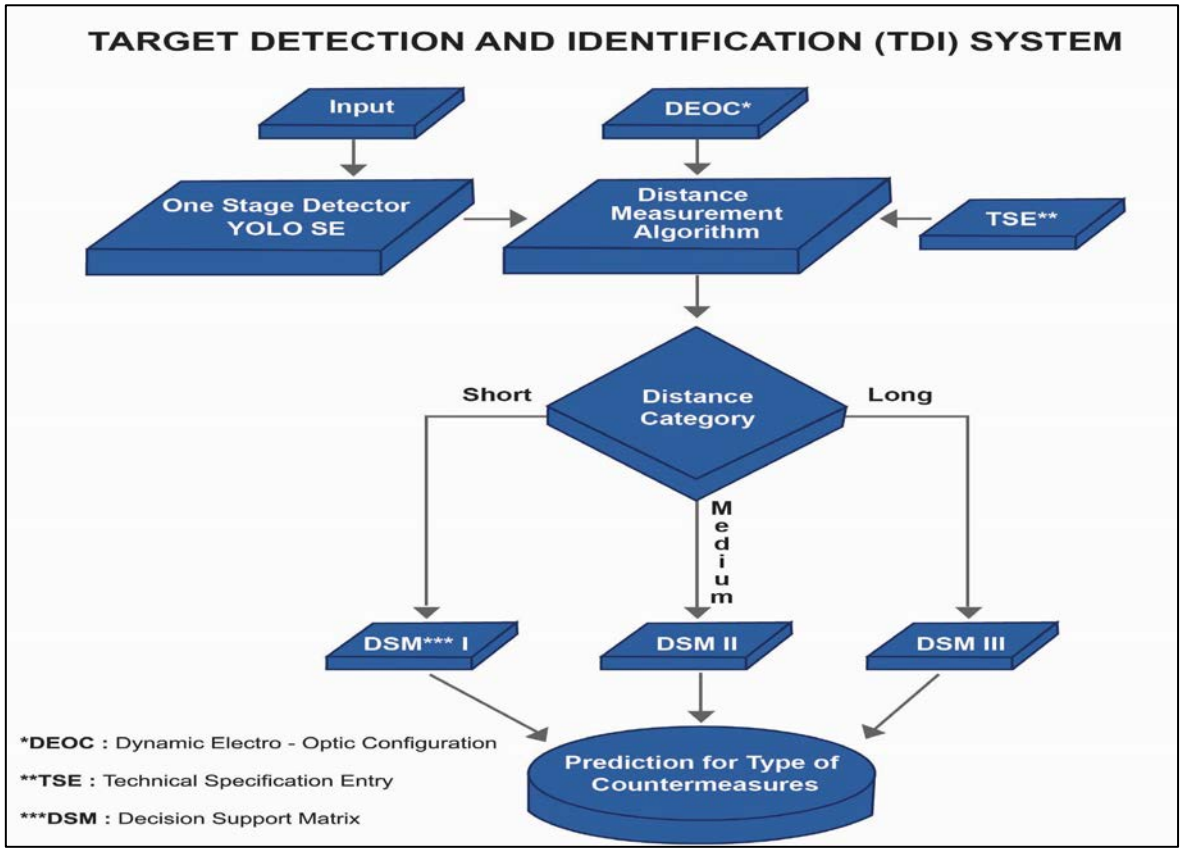


**Figure 1.** Main algorithm and steps.

### 2.1. Dataset

For the experimentations part, the dataset was prepared using Roboflow Universe (Roboflow, 2024). The initial dataset comprised 1,730 images and was first expanded through geometric augmentations (horizontal/vertical flips and image rotations); image rotations were applied in the range −20° to +20°, while associated bounding-box rotations were constrained to −15° to +15°, yielding 2,450 images. To maintain class balance, the set was then uniformly reduced to 2,400 images. Subsequently, a second augmentation stage introduced additional transformations - image rotation (−15° to +15°), shearing (analogous angular perturbations), random cropping, and grayscale conversion applied to 25% of the images (data augmentation techniques as seen at Figure 2) - producing a final augmented dataset of 6,000 images. These controlled augmentations were selected to emulate realistic viewpoint, scale, and illumination variations and thereby increase intra-class variability, with the explicit aim of improving model generalization and the robustness of detection and classification results.

(a) (b)

**Figure 2.** (a) Shows a sample light armoured wheeled vehicle, and (b) shows the same picture after applying rotate, shear, and grayscale augmentation techniques (MCV2).

By applying the rotation, bounding box rotation, shear, and grayscale, the new augmented data set consists of 6.000 images (see data split below Table 1).

**Table 1.** Data split.

| Train Set | Valid Set | Test Set |
|---|---|---|
| %75 | %15 | %10 |
| 4,500 | 900 | 600 |

Average annotations per image is 1.2 per image, average image sizes are 0,41mp, median image ratios are 640x640). Also see for the training classes and number of samples that is used for training at Table 2.

**Table 2.** Train set data classes and number of samples.

| Train Set Classes | Number of Samples |
|---|---|
| Medium and Heavy Armoured Wheeled (MHAW) Carriers | 670 |
| Medium and Heavy Armoured Tracked (MHAT) Carriers | 652 |
| Military Trucks (MT) | 594 |
| Heavy (Main Battle) Tanks (HT) | 660 |
| Light Armoured Wheeled (LAW) Carriers | 653 |
| Armoured Artillery (AA) | 625 |
| Mine Resistant Ambush Protected Vehicles (MRAP) | 646 |

### 2.2. Operational Dataset Types

The classification of combat vehicles is a complex system that varies significantly across different countries and military organizations. Furthermore, advancements in technology can lead to the development of new categories or modifications to existing classifications. Combat vehicles are primarily categorized based on their mobility systems, which can be classified as

either wheeled or tracked. Subsequently, military equipment is classified according to its weight category, specifically into heavy, medium, and light classifications (NATO, 2017).

In the context of surveillance reports, the type and number of combat vehicles present a critical aspect of data collection and analysis. True identification of combat vehicles is very essential for assessing the operational landscape and understanding potential engagements. Studies indicate that the effectiveness of certain combat vehicles against friendly forces on the battlefield significantly influences their classification and strategic deployment (U.S. Army, 2019) (U.S. Army, 2018).

Moreover, some vehicles are designed with versatility in mind, enabling them to adapt to various roles depending on mission requirements and configurations. This multipurpose capability further complicates the classification process but also enhances operational flexibility in diverse combat scenarios. Thus, the chosen classification framework not only reflects the physical characteristics of the vehicles but also their functional applications in different mission sets.

This data set consist of 7 (seven) types that are mostly used in the armies. These are Medium and Heavy Armoured Wheeled (MHAW) Carriers, Medium and Heavy Armoured Tracked (MHAT) Carriers, Military Trucks (MT), Heavy (Main Battle) Tanks (HT), Light Armoured Wheeled (LAW) Carriers, Armoured Artillery (AA), and Mine Resistant Ambush Protected Vehicles (MRAP). Also, the test data composed free from the trained datasets.

### 2.3. Main Algorithm

YOLOv8, specifically, is an improvement and refinement of the YOLO object detection algorithm (Jocher et al., 2022). A comparative analysis of the widely used YOLOv5 and YOLOv7 is carried out by Olorunshola et al. (2023) shows significant contribution compared to other works earlier mentioned in the literature review. The experiment demonstrates the effectiveness of the YOLOv5 model compared with YOLOv7. YOLOv8 is a computer vision model architecture developed by Ultralytics, the creators of YOLOv5.

#### 2.3.1. Performance Comparison of YOLOv8

The results of the YOLOv8 to v5, v10 and v11 algorithms with the same dataset were obtained to check the performance of the model. For precision and inference speed, YOLOv8 outperforms the rest of the algorithms' results. YOLOv8 is offering cutting-edge performance in terms of accuracy and speed.

#### 2.3.2. Advanced Features of YOLOv8

Building upon the advancements of previous YOLO versions, YOLOv8 introduces new features and optimizations. These new specifications make it an ideal choice for various object detection tasks as given below:

i. Upgraded backbone and neck architectures: YOLOv8 utilizes cutting-edge backbone and neck architectures, enhancing feature extraction and improving object detection performance.
ii. Anchor-Free split head: YOLOv8 features an anchor-free split head, which enhances accuracy and streamlines the detection process compared to traditional anchor-based methods.
iii. Optimized accuracy-speed trade-off: YOLOv8 maintains a balanced approach to accuracy and speed, making it suitable for real-time object detection across various applications.
iv. Variety of pre-trained models: YOLOv8 provides a selection of pre-trained models tailored to different tasks and performance needs, facilitating the selection of the appropriate model for specific applications.

### 2.4. Model Structure

The architecture is composed of three main components (Jocher et al., 2023):

i. Backbone: The backbone is the CNN that extracts features from the given image. YOLOv8 uses a custom CSPDarknet53 backbone, which utilizes cross-stage partial connections to enhance information flow between layers and improve accuracy.
ii. Neck: The neck combines feature maps from different stages of the backbone to capture information at various scales. Instead of the traditional Feature Pyramid Network (FPN), YOLOv8 utilizes a novel C2f module. This module integrates high-level semantic features with low-level spatial information, resulting in improved detection accuracy, particularly for small objects.
iii. Head: The head is responsible for generating predictions. YOLOv8 incorporates multiple detection modules that predict bounding boxes, objectness scores, and class probabilities for each grid cell in the feature map. These predictions are then combined to produce the final detection results.

### 2.5. The Squeeze-and-Excitation Module

The SE module is a structure developed to enhance channel attention and improve the representational power of models in deep learning. This module is effective specifically in CNNs. It is primarily used in object recognition and image classification.

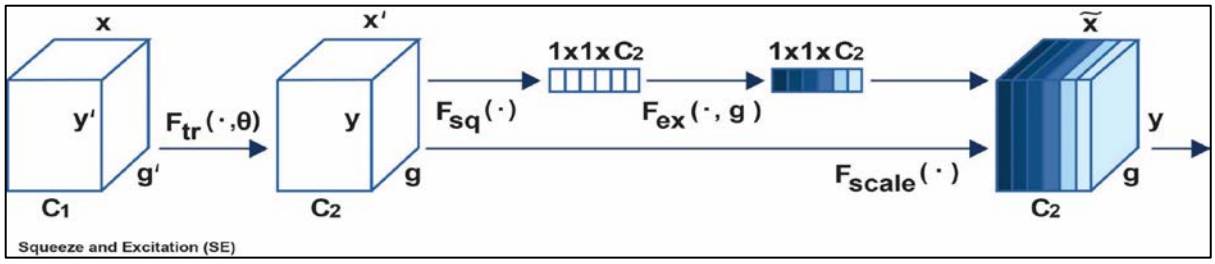


**Figure 3.** The SE module (Wu et al., 2023).

As shown in Figure 3, it has a simpler structure compared to other attention mechanisms. In the context of this design study, the (SE module has been integrated into each convolution (Conv) block, as indicated in Figure 4 and Figure 5. This allows the network to emphasize more critical channels, reduce the contribution of irrelevant channels, and acquire more distinctive features. It has been assessed that for military vehicles, which have similar subtypes, the inclusion of the SE layer positively influences the model's performance.

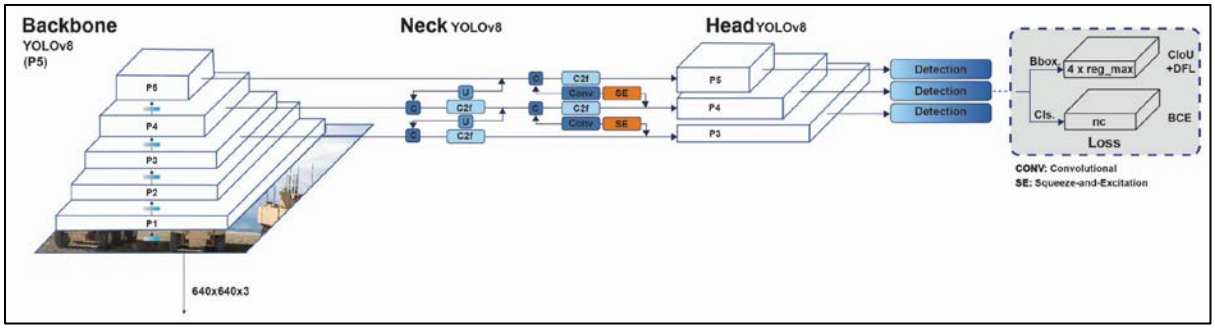


**Figure 4.** YOLOv8 SE algorithm.

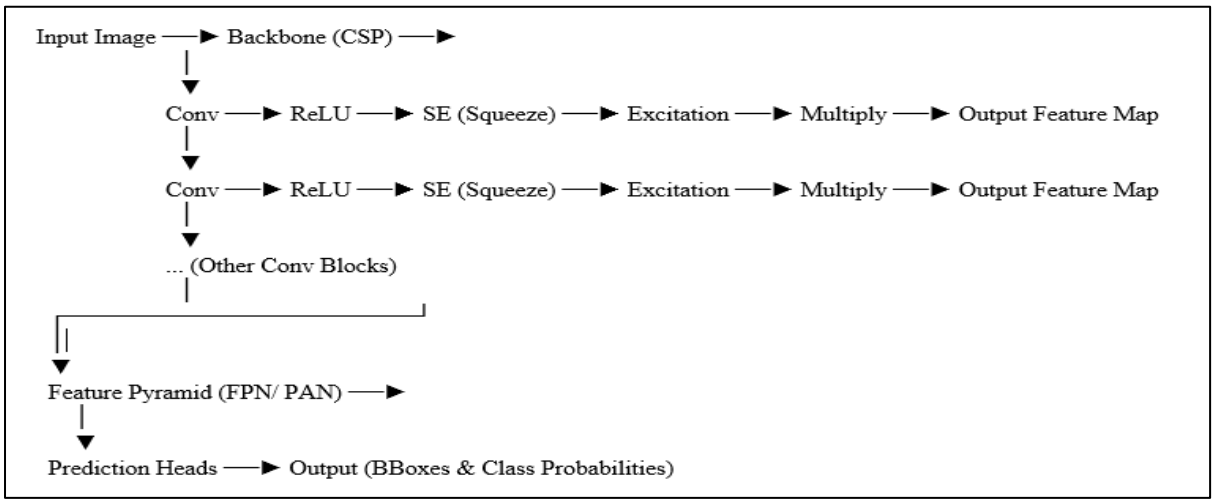


**Figure 5.** Adding the SE module to the YOLOv8 algorithm.

The SE module learns the importance of each channel. Then it enables the re-scaling of these channels with weights. This process helps the model understand which features are important and create a more effective representation. Module consists of three phases:

i. Squeeze phase: The input feature map, which is a tensor with height, width, and channel count, undergoes global average pooling for each channel to create a "feature summary" representing the overall characteristics of each channel.
ii. Excitation phase: The squeezed feature vector is rescaled through connected (dense) two layers. The first layer uses the activation function of “ReLU”. The second layer uses the activation function of “Sigmoid”. At this stage, weights indicating the importance of each channel are calculated.

iii. Scaling phase: The original feature map to rescale it multiplies the computed channel importance weights. This process adjusts each channel according to its importance value.

The SE module enhances the model's ability to learn better and more meaningful features by emphasizing channels, thereby improving performance without incurring extra computational load. It generally performs well even in systems with limited resources. By utilizing the SE module, the generalization capability of networks is increased, resulting in better performance on more challenging datasets.

The SE module is an effective attention mechanism that can be used in deep learning architectures. By enhancing channel attention in convolutional neural networks, it significantly improves model performance. The applications of the SE module are extensive, making it a critical component in many modern deep learning models.

### 2.6. Additional Feature Extraction by Using Distance Measurement Algorithm

A picture can be used to localize the target in an image and associate it with the distance. Four numbers can be obtained in the bounding box. These are width and height variables. These variables are used in the formula for calculating the distance of object and describing the details of the detected object/objects. Depending on the distance of the object to the camera, width and height of the objects will change.

The values detected by the YOLOv8 algorithm and marked with a bounding box are used to estimate the distance based on the type of target detected, utilizing the calculation method described in Equation (1) and the logic of the calculation method can be understand from Figure 6. To determine the distance, revised formulas have been adapted to the algorithm (see distance measuring formulas in Equations (2) and (3). The focal length (f) depends on the specific type of electro-optic device used.

Using an empirical scale derived from a calibration target of known physical size (0.25 m) that subtends 1471.875 pixels in the image, we computed a linear conversion factor of 1471.875 / 0.25 = 5887.5 pixels per meter; consequently, any focal length given in meters can be converted to pixel units by multiplying by this factor.

$$f/d=r/R \quad (1)$$

$$f=d\times r/R \quad (2)$$

$$d=f\times R/r \quad (3)$$

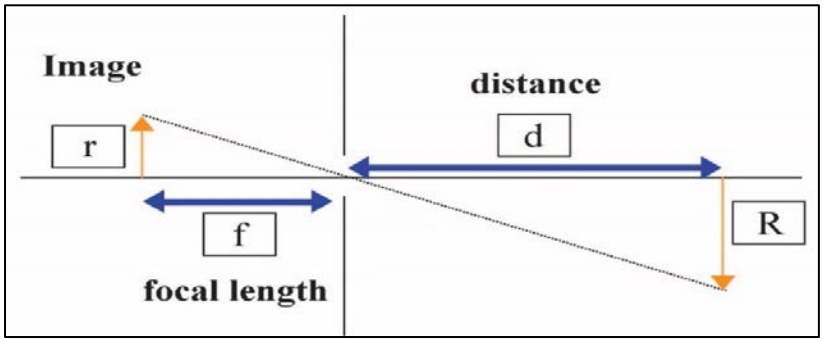


**Figure 6.** Trigonometric logic of distance measuring.

To achieve this, the width and height matrices (R) of the most commonly used vehicle types were created. When vehicle lengths are used, the false prediction rate due to the vehicle's stance position tends to increase. However, height generally remains consistent; therefore, average height values were adopted for different vehicle types (Tutuncuoglu and Guney, 2024). Additionally, when these values are adjusted according to the specific regions in which they will be applied, the estimation error rate decreases. The image height (r) of the identified vehicle is calculated by the YOLO algorithm.

### 2.7. Prediction by Using Decision Support Matrix (DSM)

After the combat vehicles are detected and identified, estimating their distance will force the decision maker to decide what precautions to take or what type of weapon system to use to neutralize the target.

The decision support matrix, which is created by examining the Land Combat Vehicles and the measures to be taken against these vehicles, makes recommendations (Table 3) to the user according to the information obtained from system.

**Table 3.** DSM I recommendation.

| Type of Military Combat Vehicles | Minimum Ammunition /Weapon System Type |
|---|---|
| Short Distance (0-X00 m.) | Recommendation |
| Light Armoured Wheeled (LAW) Carriers | Medium Armour Piercing Ammunition, RPG |
| Medium and Heavy Armoured Wheeled (MHAW) Carriers | A/T Ammunition with Short Distance Weapon System |
| Medium and Heavy Armoured Tracked (MHAT) Carriers | A/T Ammunition with Short Distance Weapon System |
| Military Trucks (MT) | A/T Ammunition with Short Distance Weapon System |
| Heavy (Main Battle) Tanks (HT) | RPG |
| Armoured Artillery (AA) | RPG |
| Mine Resistant Ambush Protected Vehicles (MRAP) | RPG |

# 3. Results and Discussion

## 3.1. Model Training Results

The YOLOv8 model was used for the training and test purpose. The YAML file was defined to configure the paths for the dataset and the number of classes to train. The model was trained for 241 epochs. A batch size of 32 was used for all epochs. The mean average precision is reached 83.6% mAP as seen Figure 7.

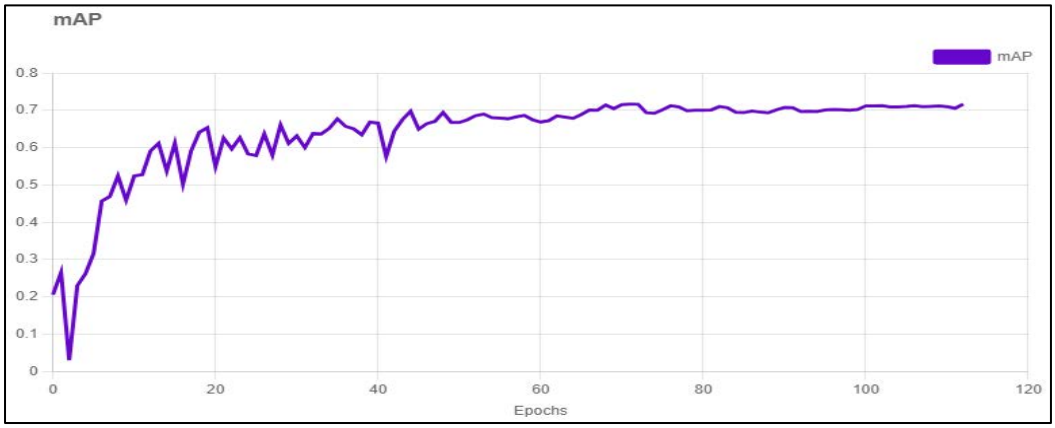


**Figure 7.** Data set mAP results.

For precision and inference speed, Table 4 also, provides the comparison of YOLO v10 and v11 results with YOLOv8. It can be seen that YOLOv8 outperforms the rest of the algorithms' results.

**Table 4.** YOLO v8 versus YOLO v10 and v11 at our dataset.

| Class | Images | Mean Average Precision | | |
|---|---|---|---|---|
| | | YOLOv8 | YOLOv10 | YOLOv11 |
| All | 6,000 | 83.6% | 68.7% | 70.5% |
| Speed Of Inference | | 117 FPS | 80 FPS | 100 FPS |

When the confusion matrix (Figure 8) is examined, it is observed that the lowest results are in MHAW, LAW and MT type classifications. It is considered that the similarity of LAW and MRAP type vehicles to civilian vehicles in some cases increases the error rates.

The detection and classification of a single observed combat vehicle is carried out very quickly by the YOLOv8 algorithm.

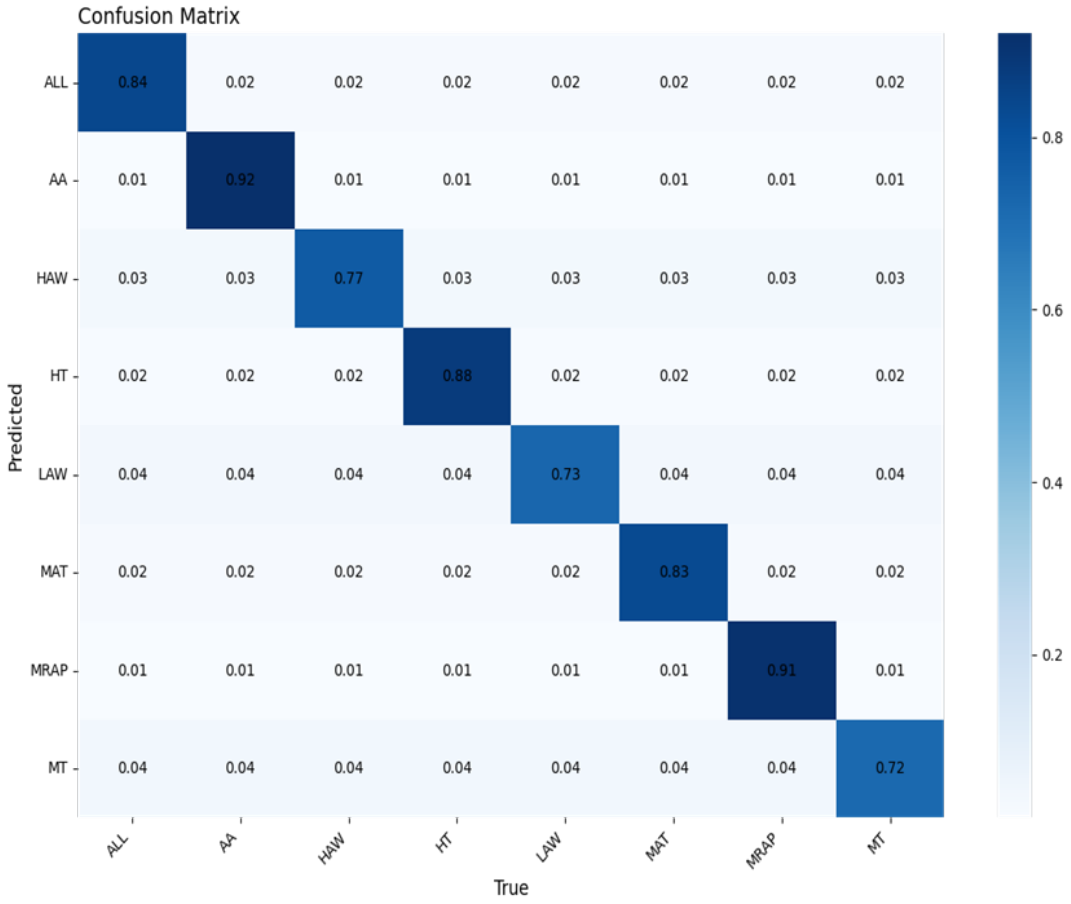


**Figure 8.** Confusion matrix.

### 3.2. Improved YOLOv8 by SE Module

A general evaluation of the application results of YOLOv8 SE has been conducted, along with a comparison to the classic YOLOv8 results Tütüncüoglu and Güney (2024). It has been determined that YOLOv8 SE, with the addition of SE modules, achieves mAP values exceeding 88.38% as seen at Table 5. This represents a significant improvement (approximately 5.7%), particularly for small objects and challenging scenes.

In our study, the addition of SE modules introduced no extra cost. The speed of YOLOv8 SE increase (approximately 17%) when we compare the classic YOLOv8.

**Table 5.** YOLOv8 and YOLOv8 SE test results comparison.

| Class | Images | Mean Average Precision | |
|---|---|---|---|
| | | YOLOv8 | YOLOv8 SE |
| All | 6,000 | 83.6% | 88.38% |
| Speed of Inference | | 117 FPS | 137 FPS |

In addition, when we compared the lowest results of the YOLOv8 (LAW, MHAW and MT), YOLOv8 SE increased the performance of the lowest sub-types as seen at Table 6.

**Table 6.** Comparison of sub-types.

| Lowest Classes Comparison | Mean Average Precision | |
|---|---|---|
| | YOLOv8 | YOLOv8 SE |
| LAW | 73% | 82% |
| MHAW | 77% | 97% |
| MT | 72% | 86% |

Following the main classification, the estimation of the distance using additional data obtained from the relevant tables, and the use of the appropriate decision support matrices, with the presentation of the results are calculated in milliseconds (Example for single target - Figure 9 and Table 7).


**Figure 9.** Single target detection.

**Table 7.** Single target detection and identification with recommendation.

| Class | MAT (80.35%) |
|---|---|
| Detected Height (h) | 490.35 px |
| Distance | 74.98 meters |
| Recommendation | Medium Armour Piercing Ammunition, RPG |

In cases where there are multiple targets, the required computation times are negligibly low (Example for multi target is given in Figure 10 and Table 8).

**Table 8.** Multi target identification with recommendation.

| Class | LAW (92.89%) |
|---|---|
| Detected Height (h) | 411.84 px |
| Distance | 75.05 meters |
| Recommendation | Medium Armour Piercing Ammunition, RPG |

| Class | LAW (77.89%) |
|---|---|
| Detected Height (h) | 83.87 px |
| Distance | 368.54 meters |
| Recommendation | Medium Armour Piercing Ammunition, RPG |


**Figure 10.** Multi target detection.

### 3.3. Operational Test Results and Discussion

The trained YOLOv8 SE model (algorithm, measuring distance and decision support matrix) is uploaded to the military type of computer (Panasonic Toughbook CF33). Next, field tests were conducted to evaluate the success of the TDI System utilizing both the tablet's camera and the vehicle's camera.

The type of the vehicles was MHAW type vehicles. The success of detection is 100% and classification precision 90%, the false precision is 10%. The false classification is only for 1 (one) LAW as seen at Table 9.

**Table 9.** Operational test results by YOLOv8 SE at field.

| Class/T&F | MHAW |
|---|---|
| True | 9 |
| False | 1 (LAW) |

For validation, ten printed exemplars of each vehicle class were photographed with a tablet camera and processed by the YOLOv8-SE model deployed on the Panasonic Toughbook CF-33; the measured classification accuracy on these printed-image captures was approximately 6 percentage points lower than the accuracy observed during the controlled training test evaluation as seen at Table 10. This decline is consistent with a domain shift between the training distribution and the printed-image test distribution: printing and re-photographing alter image statistics, introduce halftone and paper-texture artifacts, and often reduce effective resolution and dynamic range; additionally, uncontrolled capture conditions (illumination, viewing angle, distance, tablet auto-exposure/white-balance, and lens distortions) and compression/noise from the tablet camera further diverge the test images from the original training data. Together, these

factors reduce the discriminative cues the model learned, producing the observed performance drop.

**Table 10.** Printed picture test results by YOLOv8 SE.

| Class/T&F | LAW | MHAW | MHAT | MT | HT | AA | MRAP |
|---|---|---|---|---|---|---|---|
| True | 9 | 9 | 8 | 7 | 8 | 8 | 8 |
| False | 1 | 1 | 2 | 3 | 2 | 2 | 2 |

### 3.4. Operational Understanding of Training

MT, MHAW and MRAP type vehicle identification confusion results are investigated. It is observed that their sizes, heights from ground, and number of the wheels are different. Remember that the selected pictures are very distinctive examples (Figure 11).

(a) (b) (c)

**Figure 11**. (a) MT, (b) MHAW, and (c) MRAP examples (MCV2).

MHAT and HT type of vehicle identification confusion results are investigated. It is observed that their bases are almost same. Specifically, the towers are resembling each other. Only, the diameter of the tank barrels is bigger than the diameter of MHAT barrels (Figure 12).

(a) (b)

**Figure 12.** (a) MHAT, and (b) HT example (MCV2).

## 4. Conclusion

In this study, the YOLOv8 object detection model has been improved by SE and used to detect combat vehicles and to categorize their types efficiently. Also, the distances of the vehicles have been estimated by sub-algorithm, and a supporting decision matrix has been created with the obtained data to develop countermeasure recommendations. The model's detection and identification rates have been calculated as 88.38% in terms of mAP and 93% in terms of precision. When evaluating these results, it is evident that the model provides high accuracy and effectiveness in vehicle detection.

The ability to accurately identify the type and distance of the vehicle passively indicates that the model could be easily integrated with electro-optical systems or military vehicles in the future. These findings suggest that the model may offer a practical solution for military applications.

The results clearly highlight the potential of object recognition and distance estimation applications in the military domain. Future studies are recommended to additional improve the model's performance and test it under different operational scenarios. Suggestions for the development of the model can guide future research efforts. Due to the complex structure and dynamic variability of the real battlefield, there is a need to detect and to classify targets in this environment in near real time. This paper proposes an improved image recognition method. This method is formed by the combination of the image recognition algorithm, distance estimation algorithm (by additional feature extraction), and decision support matrices.

In conclusion, the enhancement of the capabilities of the developed model holds significant potential for supporting faster and more effective decision-making processes on the battlefield. Future studies may explore strategies to further improve the integration and performance of the model in this regard.

## Author Contributions

The first author contributed to the study through research design, implementation, and manuscript development. The second author contributed significantly to the study by providing guidance and insights in areas of consultation and mentoring. Their expertise played a vital role in shaping the research direction and enhancing the overall quality of the work.

## Data and Code Availability

The data and code used in this study are available upon reasonable request from the corresponding author. Access to the datasets and software will be provided to researchers who seek to replicate or build upon the findings of this work.

## Supplementary Materials

The data utilized in this study are openly accessible in the Roboflow database. Researchers can freely access and download the datasets to facilitate further analysis and replication of the results.

## References

Basavaraj, U., Raghuram, H., & Mohana. (2023). Real-time object distance and dimension measurement using deep learning and OpenCV. *Proceedings of the Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, 929–932. https://doi.org/10.1109/ICAIS53537.2023.9762892

Berwo, M. A., Khan, A., Fang, Y., Fahim, H., Javaid, S., Mahmood, J., & Abideen, Z. U. (2023). Deep learning techniques for vehicle detection and classification from images/videos: A survey. *Sensors, 23*, 4832.

Cheng, G., Chao, P., Yang, J., & Ding, H. (2024). SGST-YOLOv8: An improved lightweight YOLOv8 for real-time target detection for campus surveillance. *Applied Sciences, 14*(12), 5341.

Cui, C., Wang, R., Wang, Y., Zhou, F., Bian, X., & Chen, J. (2024). Research on optical remote sensing image target detection technique based on DCH-YOLOv7 algorithm. *IEEE Transactions, 12*, 34741–34751.

Gandhi, R. (n.d.). *R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object detection algorithms.* Medium. https://medium.com/towards-data-science/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e

Hu, J. (2014). *Research on fast tracking algorithm of moving target based on optical flow method* (Master's thesis). Xidian University, Xi'an, China.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 713–722).

Huadong, H., Binyu, W., Jiannan, X., & Tianyu, Z. (2024). Improved small-object detection using YOLOv8: A comparative study. *Applied and Computational Engineering, 41*, 80–88.

Jocher, G., Chaurasia, A., & Qiu, J. (2023). *Ultralytics YOLOv8*. https://github.com/ultralytics/ultralytics

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Michael, K., Fang, J., Yifu, Z., Wong, C., Montes, D., Wang, Z., Fati, C., Nadar, J., Sonck, V., Skalski, P., Hogan, A., Nair, D., Strobel, M., & Mrinal. (2022). *Ultralytics/yolov5: v7.0 – YOLOv5 SOTA realtime instance segmentation* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7035401

Lan, M., Wang, J., & Zhu, L. (2023). Perception and range measurement of sweeping machinery based on enhanced YOLOv8 and binocular vision. *IEEE Transactions, 11*, 126398–126408.

Liu, X. (2017). *End-to-end target detection and attribute analysis algorithm based on deep learning and its application* (Master's thesis). South China University of Technology, Guangzhou, China.

NATO. (2017). *Land symbols in APP-6 (D).* https://nisp.nw3.dk/standard/nato-app-06-ed.d-v1.html

Olorunshola, O., Irhebhude, M., & Evwiekpaefe, A. (2023). A comparative study of YOLOv5 and YOLOv7 object detection algorithms. *Journal of Computing and Social Informatics, 2*, 1–12. https://doi.org/10.33736/jcsi.5070.2023

Redmon, J., & Angelova, A. (2015). Real-time grasp detection using convolutional neural networks. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA).*

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 779–788).

Roboflow. (n.d.). *Roboflow annotation tool.* https://roboflow.com/annotate and MCV2 Dataset; https://app.roboflow.com/baskent-univercity-5cybx/mcv2/browse?queryText =&page Size = 50&startingIndex=0&browseQuery=true

Szeliski, R. (2010). *Computer vision: Algorithms and applications* (1st ed.). Springer.

Tutuncuoglu, R. A., & Guney, S. (2024). Development of target detection and identification system for combat military vehicles via artificial intelligence and decision support matrix. *International Journal for Multidisciplinary Research, 6*(6).

U.S. Army. (2018). *ATP 3-21.10 Infantry rifle company: Appendix F—Armored, Stryker, and mounted employment*. U.S. Army Publishing Directorate.

U.S. Army. (2019). *ATP 2-01.3 Intelligence preparation of the battlefield: Chapter 5—Evaluate the threat*. U.S. Army Publishing Directorate.

Wu, T., & Dong, Y. (2023). YOLO-SE: Improved YOLOv8 for remote sensing object detection and recognition. *Applied Sciences, 13*(24), 12977.

Zhang, X., Huang, H., Yang, J., & Jiang, S. (2024). Research on deep learning-based vehicle and pedestrian object detection algorithms. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-4/W10-2024*, 213–220. https://doi.org/10.5194/isprs-archives-XLVIII