# Helmet Wearing Detection Using Deep Learning: A Comparative Performance Evaluation of YOLOv12 and YOLOv11

Amirhussein safarzadeh

a.h.safarzadeh@gmail.com

**ABSTRACT**

*Object detection in images using deep learning methods has made significant progress in recent years. One of the most efficient approaches is YOLO (You Only Look Once), which, in addition to providing good accuracy, also offers high speed, making it practical in many applications. YOLO version 12 is the first version to extensively integrate the attention mechanism into its architecture as a method to improve the detection of small objects in images. To evaluate the performance of YOLOv12, a helmet-wearing detection model was developed and its accuracy and speed were compared with a model implemented using YOLOv11.*

***Keywords:*** *Image Processing, Deep Learning, Artificial Neural Networks, Object Detection, YOLO, Attention Mechanism, Helmet Wearing Detection*

## 1. *INTRODUCTION*

*In recent years, image processing models based on deep learning have made significant progress and have been able to provide efficient solutions for a wide range of challenges and issues in the field of image processing [1]. YOLO (You Only Look Once) is one of the deep learning-based models, first introduced in 2015 [2]. This model is widely used for object detection and has demonstrated satisfactory performance in both detection accuracy and speed compared to other models [3].*

*So far, multiple versions of YOLO have been introduced, each improving upon the previous one in terms of speed, accuracy, and the incorporation of new architectural structures [4]. Moreover, some specialized architectures have been developed by extending YOLO as a base model with new modules, aiming at specific applications [5].*

*Among the different versions of YOLO, version 12 stands out due to a major difference from its predecessors—its extensive use of the attention mechanism to enhance detection accuracy and speed. In YOLOv12, attention modules are embedded in different parts of the architecture, which marks a substantial distinction from previous versions. In contrast, YOLOv11 was the first to introduce the attention mechanism in the C2PSA block, while most other components of the architecture remained largely similar to earlier designs. In YOLOv12, however, the attention mechanism plays a central role in the design of the building blocks, to the extent that YOLOv12 can be considered a redesign of YOLO based on the attention mechanism [6].*

*The attention mechanism itself is inspired by biological vision, where the visual system tends to focus on the most informative parts of an image. While there are different ways to implement attention in artificial neural networks, YOLOv12 employs the Area Attention Mechanism [7]. In this mechanism, a horizontal range and a vertical range are defined, and the overlap of these ranges specifies the attention region [7,8].*

*The aim of this paper is to compare YOLOv12 with YOLOv11. To this end, we developed a model for helmet-wearing detection in an industrial environment. Helmet-wearing detection using YOLO has already been the focus of several research works, where different YOLO versions and datasets were employed.*

*For instance, in [9], an extended YOLOv8 architecture was proposed for helmet detection, where a mechanism called Dilation-wise Residual Attention was incorporated to improve performance. Similarly, in [10], YOLOv5 was combined with an attention module called Efficient Channel Attention (ECA), which was specifically designed to introduce minimal computational overhead so as not to compromise the network's speed. In [11], YOLOv5 served as the base architecture while the CBAM module (a combination of channel-wise and spatial attention) was employed to enhance performance.*

*Overall, the use of attention mechanisms in these works primarily aims to improve the detection of small objects in images. In this paper, we construct a helmet-wearing detection model using YOLOv12—which extensively applies the attention mechanism—and compare it with a similar model based on YOLOv11. The goal is to investigate whether the major architectural changes and redesigned modules in YOLOv12 result in a tangible improvement in accuracy and speed over YOLOv11.*

## 2. *DATA & METHOD*

### *2.1 Dataset*

*The dataset used for training the model was collected by the author in an industrial environment. During data collection, efforts were made to ensure diversity in terms of lighting conditions, viewing angles, and distances, so that the dataset could better reflect real-world scenarios. The dataset consists of 510 images, of which 450 images were used for training and 60 images were reserved for evaluation. The dataset contains two classes: class 0 represents "without helmet," and class 1 represents "with helmet." Figure 1 shows a sample of the images included in the dataset.*




*Figure 1: Sample images from the dataset.*

### *2.2 Evaluation method*

*As described in the dataset section, the collected data was divided into two subsets: one used for training the model and the other applied for evaluating its performance.*
*The primary metric used to assess the effectiveness of an object detection model is Mean Average Precision (mAP). Figure 2 illustrates the factors required for its calculation.*

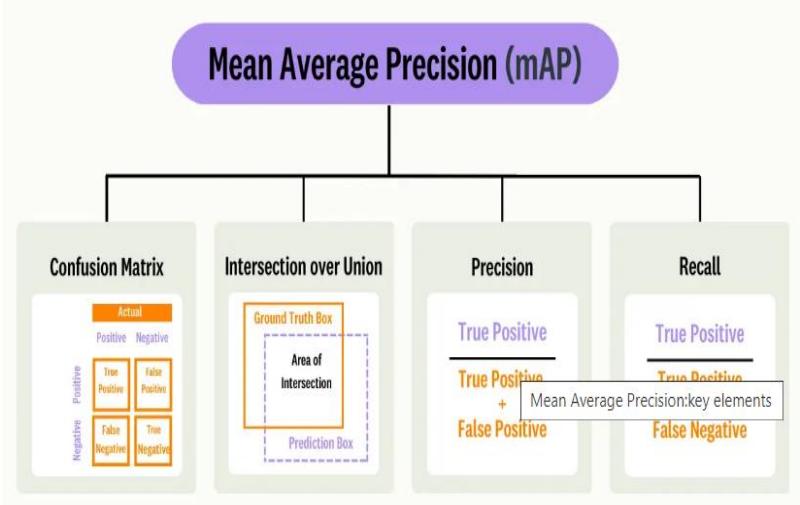



*Figure 2: Mean Average Precision (mAP)*

*This metric is derived from four fundamental factors commonly used in evaluating object detection models:*

***Confusion Matrix:*** *A standard method to visualize the distribution of the model's predictions across different classes.*

***Intersection over Union (IoU):*** *A factor used to measure the accuracy of the model in localizing objects.*

***Precision:*** *Indicates the model's ability to correctly classify positive predictions among all positive predictions made. It is calculated using Equation (1).*

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

***Recall:*** *Represents the model's ability to correctly identify positive samples relative to all actual positives. It is calculated using Equation (2).*

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

*finally, after calculating Precision and Recall, the average precision (Api) can be obtained, followed by the mean average precision (mAP). Equation (3) shows the calculation of APi , and Equation (4) presents for mAP.*

$$AP = \frac{1}{n} \sum_{i=1}^{n} \big(Recalls(k) - Recalls(k+1)\big) \times Precision \quad (3)$$

$$mAP = \frac{1}{n} \sum_{i=1}^{n} APi \quad (4)$$

### *2.3 YOLOv11 Architecture*

*Figure 3 illustrates the block diagram of the YOLOv11 architecture. All YOLO versions are composed of three main parts: the Backbone, the Neck, and the Head.*

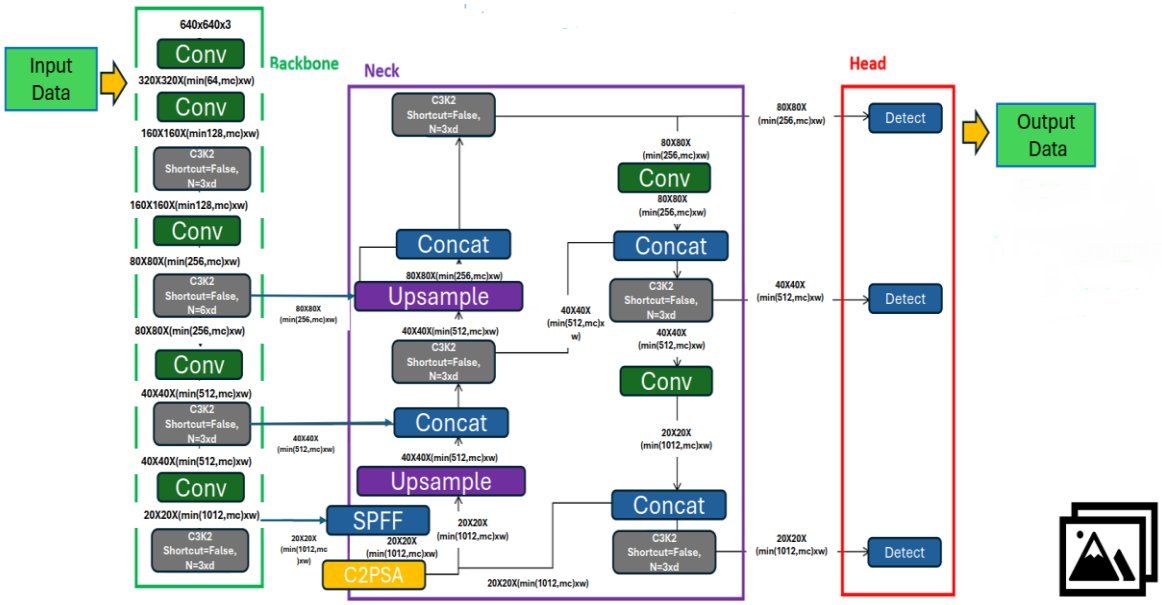



*Figure 3: YOLOv11 architecture.*

*the Backbone is the first part that processes the input image. It applies multiple convolutional layers to transform the input image into different channels, which are then used for feature extraction.*

*The second part, the Neck, receives the processed information from the Backbone. Its role is to extract meaningful features from the obtained channels and prepare them for the Head.*

*The final part, the Head, uses the extracted information to determine the location of the target object in the image as well as its class.*

*As illustrated in Figure 3, each part of the architecture is composed of several blocks:*

- ***Conv Block:** Consists of multiple convolutional layers that filter the input image and create different channels to enable feature extraction.*
- ***SPPF Block (Spatial Pyramid Pooling Fast):** Contains several Max Pooling layers, each performing matrix operations to remove redundant information and reduce the computational load.*
- ***Concat Block:** Merges two matrices of different dimensions.*
- ***Upsample Block:** Increases the size of a matrix, allowing matrices to be aligned for matrix operations and for use in the Concat block.*
- ***C3K2 Block:** Responsible for feature extraction. It is an improved version of the C3 block used in previous YOLO versions. C3K2 offers higher accuracy for small object detection while reducing additional computations.*
- ***C2PSA Block:** In YOLOv11, this block enhances class detection accuracy by focusing on regions of the image where the object is most likely to appear. It consists of two parts: C2, a 2D convolutional network, and PSA (Partial Spatial Attention), which employs a spatial attention mechanism to focus on regions with a higher probability of containing the target object.*

- ***Detect Block:*** *Located in the Head, this block performs the final calculations to determine the object's location and its class in the image.*

### *2.4 YOLOv12 Architecture*

*The YOLOv12 architecture is illustrated in Figure 4. As shown, it also consists of three main components: Backbone, Neck, and Head, each performing the same general functions as in previous YOLO versions.*

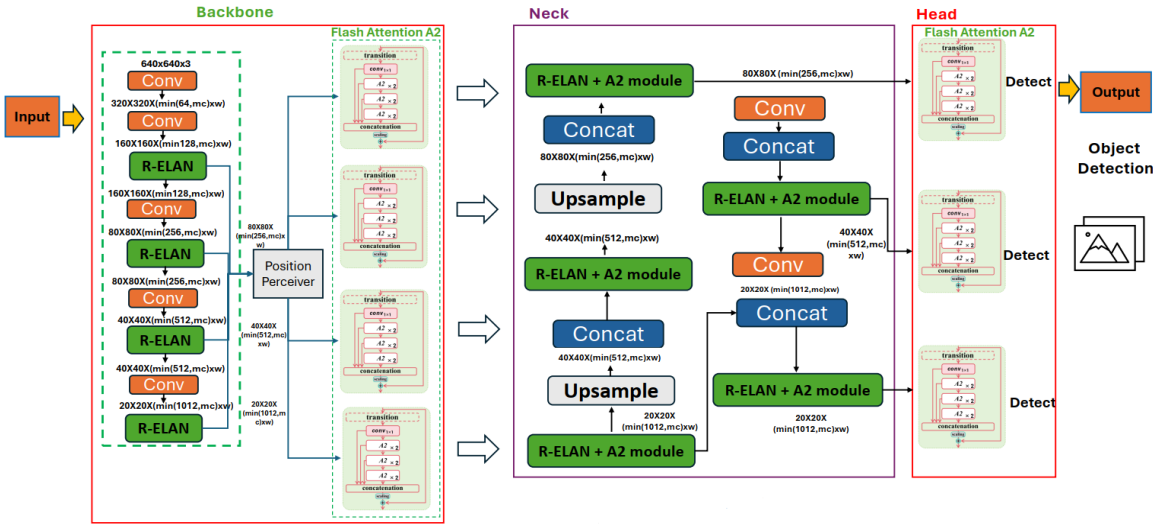



*Figure 4: YOLOv12 architecture*

*Some of the blocks in YOLOv12 are similar to those in YOLOv11. However, YOLOv12 introduces two new blocks: R-ELAN and the A2 module.*

- ***R-ELAN Block:*** *R-ELAN stands for Residual Efficient Layer Aggregation Network. This block is designed to efficiently extract detailed features from the image, enabling better classification. It combines two key deep learning concepts: Residual Connection and Efficient Layer Aggregation.*
  - ***Residual Connection:*** *In deep neural networks with many layers, important information from early layers can be lost during forward propagation, negatively affecting performance. Residual connections address this by adding the input of previous layers to the output of subsequent layers, preserving critical information and facilitating training of deep networks.*
  - ***Efficient Layer Aggregation:*** *This concept aggregates features extracted by all layers of the network to produce the output. By combining information from all layers, the network retains richer and more comprehensive features without losing important details.*
- ***A2 Module:*** *The Area Attention (A2) Module is used for the first time in YOLOv12. Its purpose is to leverage the benefits of the attention mechanism while maintaining lower computational complexity compared to other attention-based structures. The module divides the feature map into several regions and processes them individually, selecting the regions with the highest probability of containing relevant object information for further processing. This strategy focuses computations on the most informative parts of the image, improving detection accuracy. Unlike conventional attention mechanisms that often operate on fixed-size cells (e.g., 3×3), the A2 module can process variable-sized groups of cells, providing more flexibility and efficiency.*

## *3. RESULT & DISCUSSION*

*To implement and compare the models, both YOLOv11 and YOLOv12 were trained for 100 epochs. The models' accuracy was then evaluated using the validation dataset. Table 1 summarizes the hardware and environment used for implementation, execution, and evaluation.*

| ***CPU*** | ***GPU*** | ***RAM*** | ***CAMERA*** | ***ENVIRONMENT*** |
|---|---|---|---|---|
| *Intel cori5*<br><br>*10210U* | *Geforce MX 130* | *8GB* | *Lenovo HD Webcam 720p* | *Jupyter Lab* |

Table 1: Hardware and implementation environment

*Figure 5 shows the Precision-Recall curve for the evaluation of YOLOv11. Based on the Mean Average Precision (mAP) metric, which is commonly used to assess the accuracy of object detection models, the model achieved an accuracy of 0.809.*

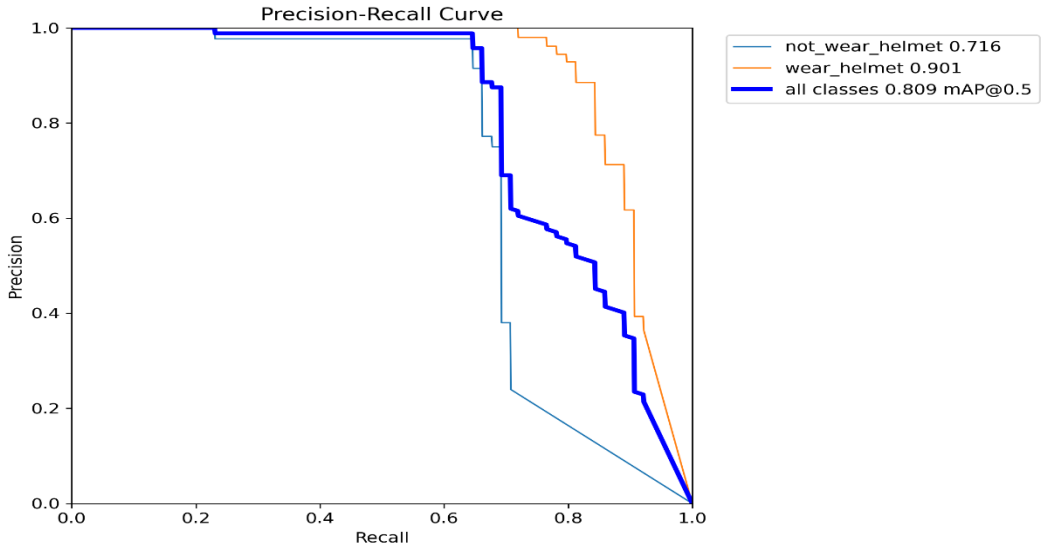



*Figure 5: Precision-Recall curve for YOLOv11 evaluation*

*Figure 6 shows the Precision-Recall curve for the evaluation of YOLOv12, which indicates that the model achieved an accuracy of 0.833.*

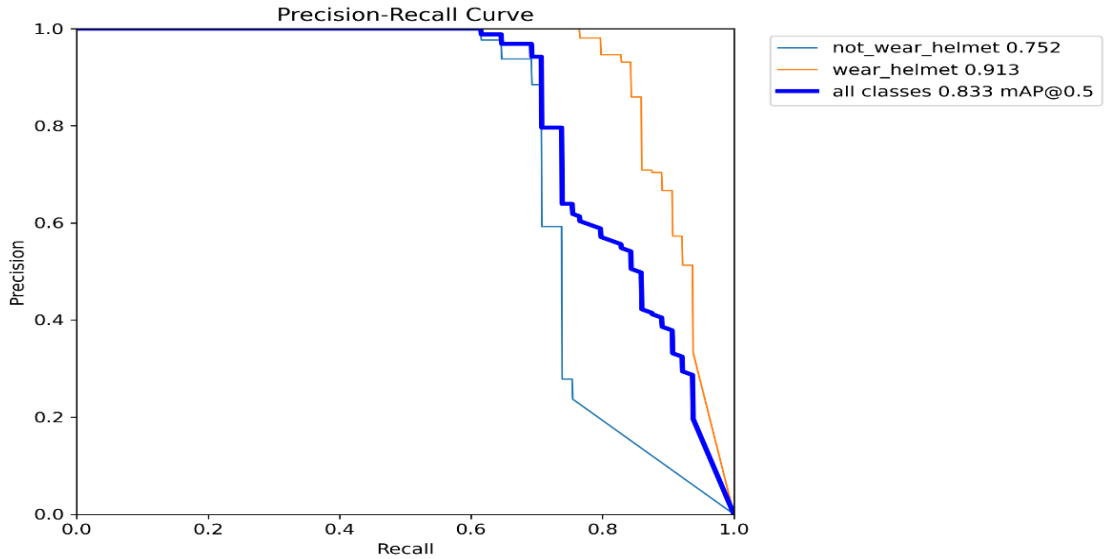



*Figure 6: Precision-Recall curve for YOLOv12 evaluation*

*Additionally, the average object detection speed for the YOLOv11 and YOLOv12 based models was 61.05 ms and 58.55 ms, respectively. Table 2 presents the evaluation results for both models.*

| ***model*** | ***map 50%*** | ***average*** | ***FPS*** |
|---|---|---|---|
| *Yolov11s* | *0.809* | *61.5ms* | *16/fps* |
| *Yolov12s* | *0.833* | *58.55ms* | *17/fps* |

*Table 2: Evaluation results*

*Based on the obtained results, it can be concluded that YOLOv12 outperforms YOLOv11 in both accuracy and detection speed for helmet-wearing detection. This demonstrates the effectiveness of the new blocks and algorithms incorporated in YOLOv12. The performance improvement of YOLOv12 over YOLOv11 is primarily attributed to the R-ELAN and A2 modules, which enhance both the model's accuracy and speed compared to previous versions.*

## REFERENCES

1. *Altarawneh A. Deep Learning Applications in Computer Vision: Experimental and Comparative Study. 2022 Jan 1.*
2. *Joseph R., Santosh D., Ross G., Ali F. You Only Look Once: Unified, Real-Time Object Detection. arXiv, 2016 Jan 1.*
3. *Redmon J., Farhadi A. YOLOv3: An Incremental Improvement. arXiv:1804.02767, 2018 Apr 8.*
4. *Khanam R., Hussain M. A Review of YOLOv12: Attention-Based Enhancements vs. Previous Versions. 2025 Apr 17.*
5. *Sapkota R., Flores-Calero M., Qureshi R., et al. YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series. Artificial Intelligence Review, 58(9), 2025 Jun 11.*
6. *Rabbani A., Hussain M. YOLOv12: A Breakdown of the Key Architectural Features. arXiv, 2025 Feb 20.*

7. *Guo M.H., Xu T.X., Liu J.J., et al. Attention mechanisms in computer vision: A survey. Computational Visual Media, 2022 Mar 15.*
8. *Tian Y., Ye Q., Doermann D. YOLOv12: Attention-Centric Real-Time Object Detectors. arXiv, 2025 Feb 17.*
9. *Song X., Zhang T., Yi W. An improved YOLOv8 safety helmet wearing detection network. Scientific Reports, 14(1), 2024 Jul 30.*
10. *Li S., Lv Y., Liu X., Li M. Detection of safety helmet and mask wearing using improved YOLOv5s. Scientific Reports, 13(1):21417, 2023 Dec 5. Available: https://www.nature.com/articles/s41598-023-48943-3*
11. *Liu Y., Jiang B., He H., et al. Helmet wearing detection algorithm based on improved YOLOv5. Scientific Reports, 14(1):8768, 2024 Apr 16. Available: https://www.nature.com/articles/s41598-024-58800-6*
12. *Wang L., Zhang X., Yang H. Safety Helmet Wearing Detection Model Based on Improved YOLO-M. IEEE Access, 11:26247–57, 2023. Available: https://ieeexplore.ieee.org/document/10068495*
13. *Vaswani A., Shazeer N., Parmar N., et al. Attention Is All You Need. arXiv:1706.03762, 2017. Available: https://arxiv.org/abs/1706.03762*