Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 20 No. 5 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v20i05.47171

PAPER

A Proposed Approach for Object Detection and Recognition by Deep Learning Models Using Data Augmentation

Ismael M. Abdulkareem[1], Faris K. AL-Shammri[2], Noor Aldeen A. Khalid[3](✉), Natiq A. Omran[2]

[1]Information Technology Engineering Department, College of Engineering, University of Qom, Qom, Iran

[2]Biomedical Engineering Department, College of Engineering, University of Warith Al-Anbiyaa, Karbala, Iraq

[3]Department of Medical Instruments Engineering Techniques, Bilad Alrafidain University College, Diyala, Iraq

dr.nooraldeen@bauc14.edu.iq

ABSTRACT

Object detection and recognition play a crucial role in computer vision applications, ranging from security systems to autonomous vehicles. Deep learning algorithms have shown remarkable performance in these tasks, but they often require large, annotated datasets for training. However, collecting such datasets can be time-consuming and costly. Data augmentation techniques provide a solution to this problem by artificially expanding the training dataset. In this study, we propose a deep learning approach for object detection and recognition that leverages data augmentation techniques. We use deep convolutional neural networks (CNNs) as the underlying architecture, specifically focusing on popular models such as You Only Look Once version 3 (YOLOv3). By augmenting the training data with various transformations, such as rotation, scaling, and flipping, we can effectively increase the diversity and size of the dataset. Our approach not only improves the robustness and generalization of the models but also reduces the risk of overfitting. By training on augmented data, the models can learn to recognize objects from different viewpoints, scales, and orientations, leading to improved accuracy and performance. We conduct extensive experiments on benchmark datasets and evaluate the performance of our approach using standard metrics such as precision, recall, and mean average precision (mAP). The experimental results demonstrate that our data augmentation-based deep learning approach achieves superior object detection and recognition accuracy compared to traditional training methods without data augmentation. We compare the average accuracy of the YOLOv3-SPP model with two other variants of the YOLOv3 algorithm: one with a feature extraction network consisting of 53 convolutional layers and the other with 13 convolutional layers. The average accuracy of the proposed model (YOLOv3-SPP) is reported as accuracy of 97%, F1-score of 96%, precision of 94%, and average Intersection over Union (IoU) of 78.04%.

KEYWORDS

object detection, object recognition, deep learning, data augmentation, convolutional neural networks (CNNs), You Only Look Once version 3 (YOLOv3)

Abdulkareem, I.M., AL-Shammri, F.K., Khalid, N.A.A., Omran, N.A. (2024). A Proposed Approach for Object Detection and Recognition by Deep Learning Models Using Data Augmentation. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(5), pp. 31–43. https://doi.org/10.3991/ijoe.v20i05.47171

Article submitted 2023-12-04. Revision uploaded 2024-01-23. Final acceptance 2024-01-23.

1 INTRODUCTION

Deep learning is a subfield of machine learning that focuses on training and building artificial neural networks with multiple layers, also known as deep neural networks. It draws inspiration from the structure and function of the human brain, particularly the interconnectedness of neurons in neural networks. Deep learning algorithms are designed to automatically learn hierarchical representations of data through multiple layers of computation. Each layer of neurons in a deep neural network processes and transforms the input data, passing it forward to the next layer. The hidden layers in deep networks enable the models to learn increasingly abstract and complex features as they progress deeper into the network [1].

Object detection is a fundamental area within computer vision that has significant implications for both scientific research and practical industrial applications. It involves the task of identifying and localizing objects of interest within images or videos. The applications of object detection are diverse and impactful. For instance, face detection plays a crucial role in various domains, including facial recognition systems used for security and access control, emotion analysis for understanding human behavior, biometric authentication, and video surveillance for monitoring public spaces [2]. Text detection is essential in tasks such as optical character recognition (OCR) for digitizing printed or handwritten text, document analysis for information extraction, automatic license plate recognition (ALPR) systems, and text extraction from images for content analysis or translation purposes. Pedestrian detection is of utmost importance in the development of autonomous vehicles, where accurate detection of pedestrians in real-time is vital for ensuring their safety. Additionally, pedestrian detection is employed in surveillance systems for crowd analysis and tracking. Logo detection finds applications in brand monitoring, where it enables the detection and analysis of logos in images or videos to gauge brand presence, measure advertising effectiveness, and protect copyright [3]. Video detection involves the continuous analysis of video streams to detect and track objects of interest, enabling applications such as video surveillance for security purposes, action recognition in sports or human-computer interaction, and object tracking in visual tracking systems. Vehicle detection is utilized in various traffic monitoring systems, parking systems for space management, intelligent transportation systems for traffic flow analysis, and in the development of autonomous driving technologies. In the medical field, object detection plays a vital role in the analysis of medical images. It helps in identifying and localizing abnormalities or specific structures within medical images, assisting in the diagnosis of diseases, and planning for surgical procedures [4].

Overall, object detection has proven to be a versatile and essential tool in computer vision, with a wide range of applications that contribute to scientific advancements and enhance industrial processes across multiple domains [5]. The goal of object detection is to identify and localize the presence of various objects of interest, such as people, cars, animals, or specific objects like stop signs or traffic lights [6], as shown in Figure 1.

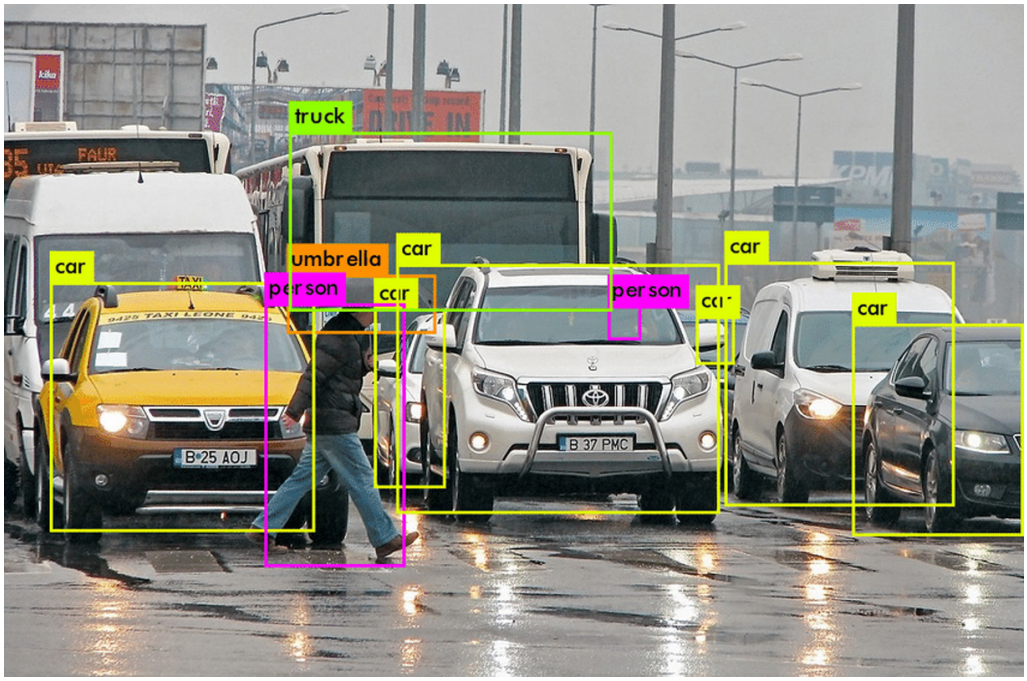

Fig. 1. Object detection [7]

Object detection goes beyond simple image classification, which focuses on assigning a single label to an entire image. Instead, object detection algorithms provide more detailed information by identifying the individual objects and their respective locations within the image or video frame. The output typically consists of bounding boxes that enclose the detected objects, along with corresponding class labels or categories [7].

1.1 Data augmentation

Data augmentation refers to the process of artificially increasing the diversity and size of a given dataset by applying a set of predefined transformations or modifications to the original data samples. It is commonly used in machine learning and deep learning tasks, including computer vision, natural language processing, and audio processing [8]. The primary objective of data augmentation is to enhance the performance, generalization, and robustness of machine learning models by exposing them to a wider range of variations and scenarios. By applying various transformations to the existing data, data augmentation creates new training samples that are similar but not identical to the original ones. This helps to overcome the limitations of limited training data and prevent overfitting, as shown in Figure 2. In the context of computer vision, data augmentation techniques often involve geometric transformations such as random cropping, rotation, flipping, scaling, and translation. These transformations simulate real-world variations and augment the dataset with variations in object position, size, orientation, and appearance. Additionally, data augmentation may include color and contrast [9], noise injection, occlusion, or other modifications specific to the task or domain [10].

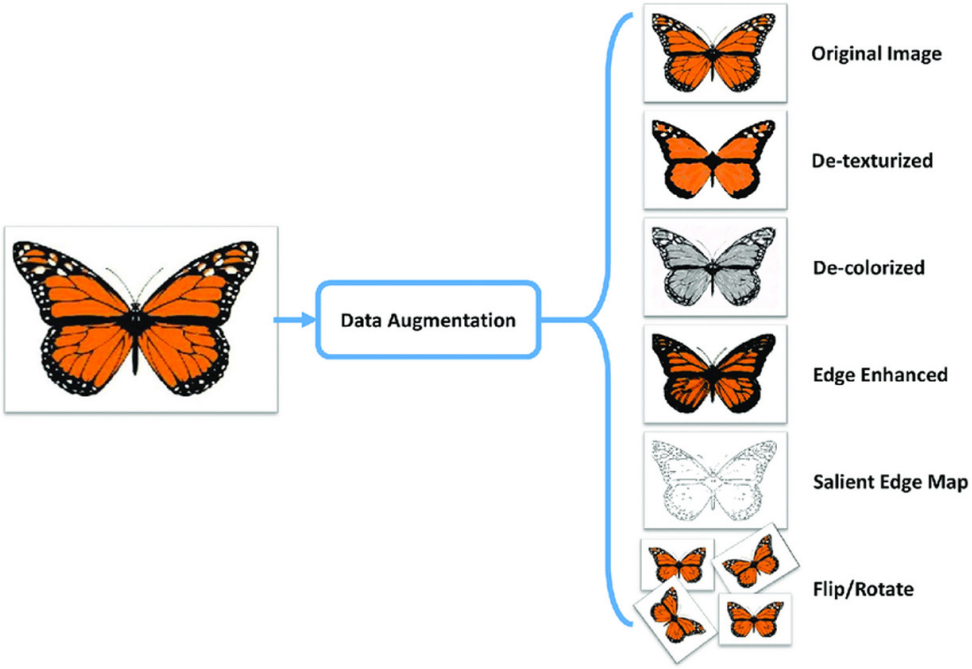

Fig. 2. Data augmentation [9]

1.2 Convolutional neural networks

Convolutional neural networks (CNNs) are a type of deep neural network specifically designed for processing and analyzing structured grid-like data, such as images, videos, and sequential data. CNNs have become a cornerstone of computer vision and image processing due to their ability to effectively capture spatial and hierarchical patterns in visual data [11]. The key feature of CNNs is the convolutional layer, which performs a convolution operation on the input data using a set of learnable filters or kernels. This operation involves sliding the filters across the input, computing element-wise multiplications, and summing the results to produce a feature map. By learning different filters, CNNs can automatically extract meaningful features from the input data at various spatial locations. The architecture of a typical CNN consists of multiple layers, including convolutional layers, pooling layers, and fully connected layers [12], as shown in Figure 3. Deep learning-based object detection methods have achieved significant advancements and have become state-of-the-art in the field. These methods use CNN architectures such as R-CNN (region-based CNN), Fast R-CNN, Faster R-CNN, You Only Look Once (YOLO), Single Shot Multibook Detector (SSD), and others. They leverage the hierarchical representations learned by CNNs to detect objects at different scales and locations within an image.

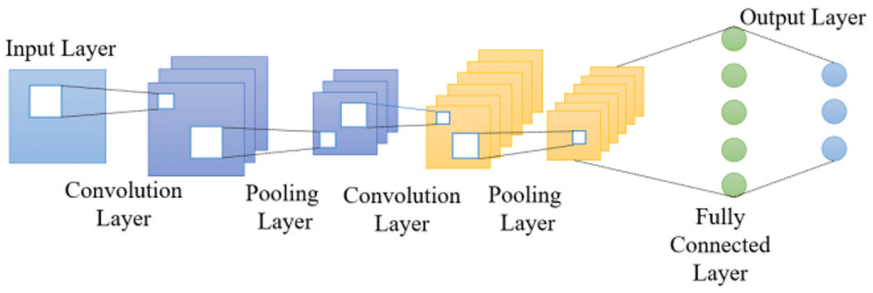

Fig. 3. General scheme of a convolutional neural network [13]

2 RELATED WORK

A literature review is conducted to gather existing knowledge, methodologies, and advancements related to object detection, recognition, deep learning, and data augmentation. This helps in establishing the current state of the field, identifying research gaps, and informing the research design.

Shaoqing Ren et al. [14] presented a regional proposal network (RPN) in the Faster RCNN. This produces the region proposals instead of the selective search algorithm. Similar whole image convolution characteristics can be found in the integration of "the RPN region proposal extraction into the DCNNs." Given that the feature mapping and RPN are implemented on the GPU, the operation is essentially free. Faris K. Al-Shammri et al. [15] proposed a technique for removing rain streaks to achieve high detection process reliability and reduce the mistake rate in normal settings and various rain conditions (light, medium, and heavy). After increasing the image quality, the literature review revealed that YOLOv3 and tiny YOLOv3 are the most effective and acceptable algorithms for recognizing objects in real time. To assess the efficacy of the developed strategy (Deraining + YOLOv3) using the proposed method combining DDN with the YOLOv3 technology. The precision was 92.92%. Zheng et al. [16] suggested a training technique to make deep neural networks stable against distortions in the input images, i.e., stability training makes the deep neural network more powerful by teaching the model to be stable on the input images with various disturbances, and this technique performs well in the presence of noise. On the ImageNet dataset, precision for stability training on the original image was 75.1%. The authors of [17] [18] studied the effects of disturbances in the input data and made an effort to improve the CNNs' accuracy by purposefully including noisy images throughout the training process. Zhihao Chen et al. [19] gave a comparison of the YOLOv3 and SSD algorithms for object tracking and detection. And had developed the SSD and YOLOv3 object detection algorithms to determine which approach was better suited for the application. The comparison shows that the YOLOv3 is superior to the SSD algorithm. Determining if a pedestrian or vehicle's trajectory could result in a risky situation was the goal. Table 1 shows details of all related work.

Table 1. Summary related work

Ref.	Year	Method	Dataset	Results
Shaoqing et al. [14]	2015	Region Proposal Network (RPN) and Fast R-CNN	PASCAL VOC 2007	Accuracy 73.2% mAP
Faris et al. [15]	2022	Deep Detail Network (DDN) and YOLO	COCO, Rainy image and YTVOS201	Precision 92.92%
Zheng et al. [16]	2016	DNN	ImageNet	75.1%
Tian at al. [18]		DeepTest	Synthetic images	Improved up to 46%
Zhihao Chen et al. [19]	2019	YOLOv3 and SSD	Cityscape and Kitty	

3 RESEARCH METHODOLOGY

In all models, pre-trained weights from the ImageNet dataset are utilized, as depicted in Figure 4. These weights are originally trained for classification on the ImageNet dataset and represent the feature extraction network's weights. This approach employs knowledge transfer, similar to transfer learning, to facilitate learning. Additionally, data augmentation is employed to train deep learning models, enhancing their accuracy and generalization capabilities even with limited training examples. The correlation, variance, and quantity of training samples directly impact the training model's efficiency. In this paper, due to the limited number of training samples, various methods are

employed to augment the data, including saturation adjustment, contrast adjustment, brightness adjustment, and color adjustment. Across all examined models, the input image size is set to 416, and Python is the programming language utilized. The models are trained on the Google Colab system with 6000 iterations. During the model training process, the following hyperparameter values are used: a batch size of 64, a learning rate of 0.001, and the SGD optimizer with a momentum of 0.9 to adjust network parameters. Additionally, a weight decay of 0.0005 is applied to prevent model overfitting.

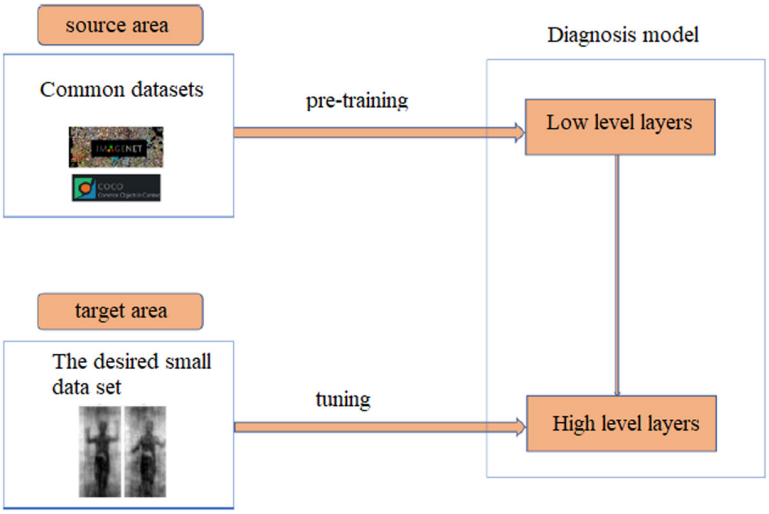

Fig. 4. The general rule of transfer learning used in network training

3.1 Data collection

The tested data is a dataset of images obtained from sensors at Peking University, China [20]. This collection includes images of people with guns, which is actually a two-class dataset. The total number of images in this dataset is 1620. An XML file is generated for each image, which contains label data and the coordinates of the bounding boxes for each class in the image. The dataset is to be annotated using XML files, which contain label data and the coordinates of bounding boxes for each class (person with gun) in the images. This type of annotation is commonly used in object detection tasks, where bounding boxes are used to indicate the location and extent of objects within an image.

3.2 YOLOv3

YOLOv3 is a popular object detection algorithm that builds upon the previous versions of YOLO. It is known for its real-time object detection capabilities, achieving a good balance between accuracy and speed. Here are some key features and improvements in YOLOv3:

- Backbone network: YOLOv3 utilizes a deep neural network as its backbone. The backbone network is responsible for extracting high-level features from the input image [21].
- Multi-scale detection: YOLOv3 operates at three different scales or resolutions, which enables it to detect objects of various sizes.
- Feature pyramid network (FPN): YOLOv3 incorporates a feature pyramid network to capture features at multiple scales. This helps in detecting objects at both small and large scales [22].

- Anchor boxes: YOLOv3 employs anchor boxes, which are predefined boxes of different aspect ratios and sizes. The use of anchor boxes allows YOLOv3 to handle objects of different shapes and sizes effectively.

3.3 SPP network

The spatial pyramid pooling (SPP) network is not connected after the last convolutional layer, but instead, it is typically connected before the fully connected layers in a CNN. The purpose of the SPP layer is to allow CNN to accept input images of different sizes and produce fixed-length feature vectors regardless of the input size. Dividing the input feature maps into sub-regions and pooling them at multiple scales achieves a fixed-size representation that encodes spatial information at various levels. The SPP layer operates as follows:

- Input feature maps: The SPP layer takes the input feature maps (output of the last convolutional layer) as its input.
- Spatial pyramid pooling: The SPP layer divides the input feature maps into a set of fixed-size grids or bins, each with a specific scale. The number of scales and bins can be adjusted to 64 based on the requirements of the network. Common choices include 1×1, 2×2, 3×3, etc. [23].
- Pooling operation: Within each grid or bin, pooling is applied to aggregate information. The pooling operation, often max pooling, extracts the most important features within each grid.
- Concatenation: The pooled features from each grid or bin at different scales are concatenated together into a single vector.
- Output: The resulting concatenated vector is fed into the subsequent fully connected layers for further processing and classification.

By employing the SPP layer, CNN can handle variable-sized inputs, allowing it to be more flexible in processing images of different resolutions. This is particularly useful in applications such as object detection and image classification, where input images may have varying sizes [24], as seen in Figure 5. It's worth noting that the SPP layer has been used in various CNN architectures, including the popular R-CNN (region-based CNN) series, to handle input images of different sizes efficiently.

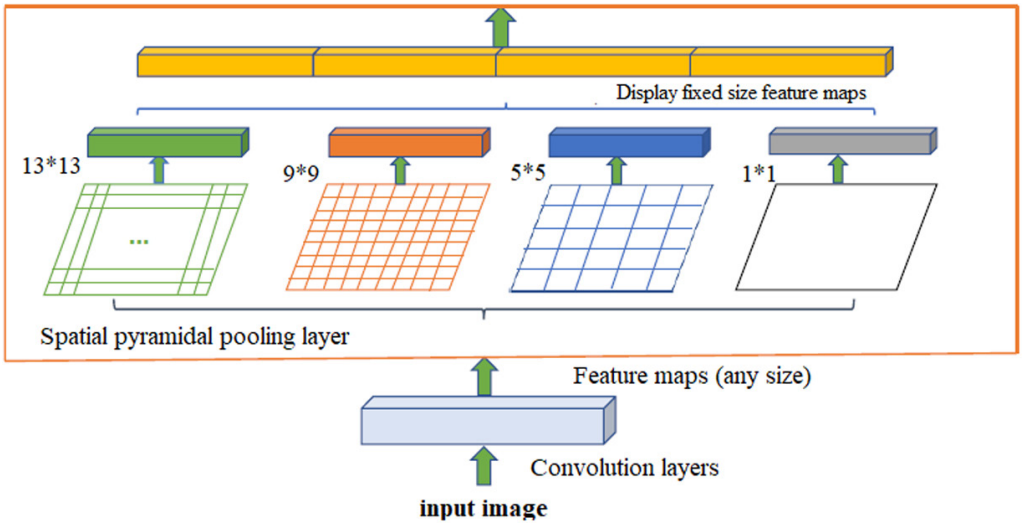

Fig. 5. Structure of the SPP module used in the YOLOv3 algorithm [24]

3.4 YOLOv3-SPP

Adding the SPP module to the YOLOv3 algorithm structure, using appropriate anchor boxes. It seems as if you are describing a modified version of the YOLOv3 algorithm that incorporates the SPP module and uses anchor boxes obtained through k-means clustering. This modified model has a total of 113 layers, with layers 89, 101, and 113 serving as detection layers for detecting different object sizes. The SPP module is added after the last convolutional layer of the feature extractor network. Its purpose is to generate feature maps at multiple scales by pooling regions of different sizes. In this case, the output of the SPP module is obtained as $13 \times 13 \times 512$ for each box. The final output of this module consists of four boxes, resulting in a final output size of $13 \times 13 \times 2048$. This output represents the predicted bounding boxes and associated class probabilities for objects in the input image. Figure 6 shows the structure of the YOLO and SPP networks.

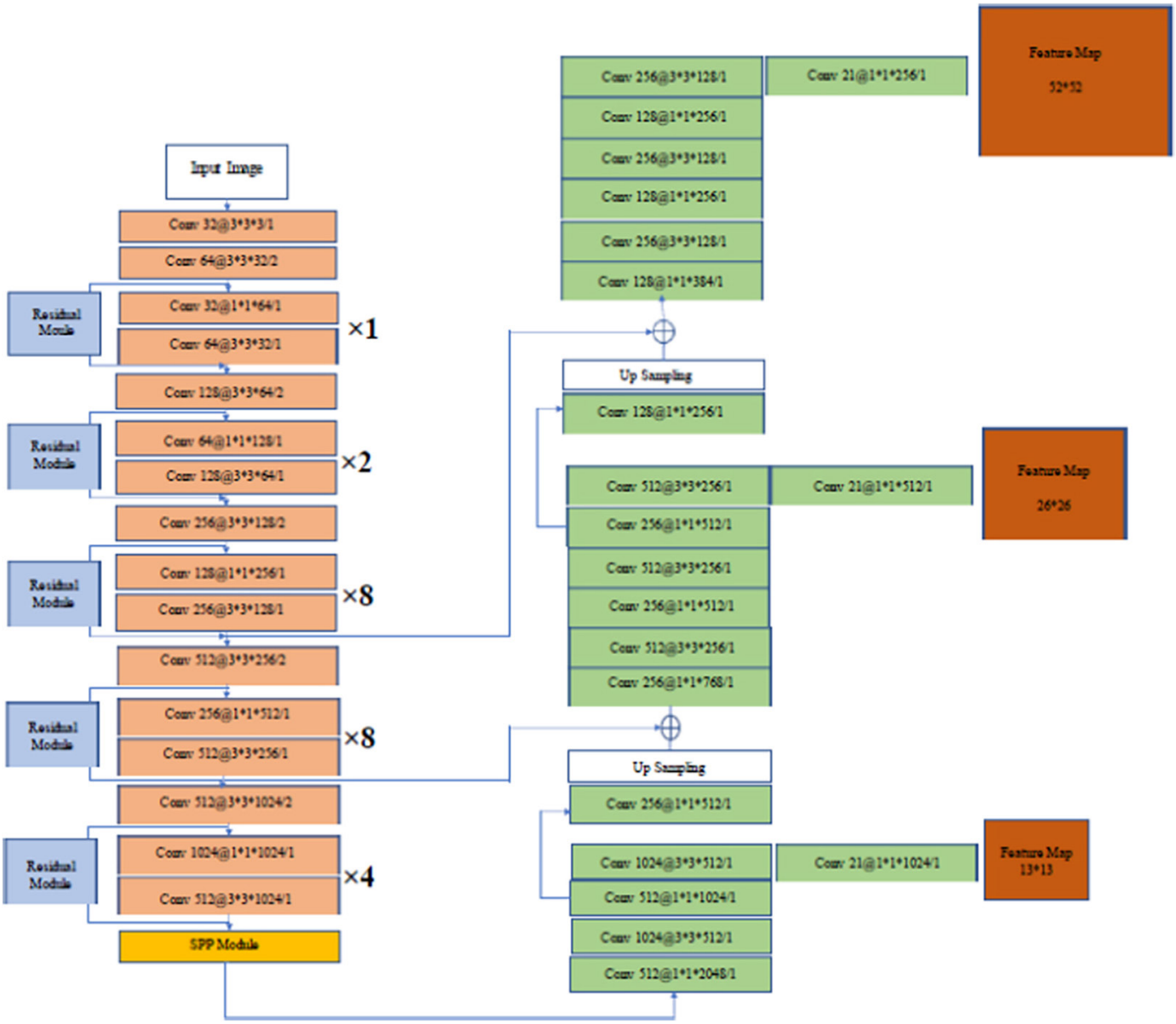

Fig. 6. The structure of the YOLOv3-SPP

4 EXPERIMENTS AND RESULT

The results were presented using the YOLOv3 model separately, and then YOLOv3-SPP on a specific dataset or set of test images includes performance metrics such as accuracy, precision, recall, IoU, mAP, or other relevant evaluation measures for the proposed method, as well as a comparison with the YOLOv3 model. The basic parameters used to evaluate these metrics are explained below: True positives (TP): number of correctly identified instances. False positives (FP): number of incorrectly identified instances. True negatives (TN): number of correctly rejected instances. False negatives (FN): number of incorrectly rejected instances. Evaluation of models to calculate accuracy, precision, recall, and F1 score is as follows:

$$\text{Precision} = \frac{(\text{TP})}{(\text{TP} + \text{FP})} \tag{1}$$

$$\text{Recall} = \frac{(\text{TP})}{(\text{TP} + \text{FN})} \tag{2}$$

$$\text{F1 Score} = \frac{\text{TP}}{\text{TP} + 1/2(\text{FP} + \text{FN})} \tag{3}$$

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})} \tag{4}$$

Figure 7 provides the performance evaluation results of the proposed algorithm compared to the YOLOv3 algorithm described in reference [19]. The evaluation metrics used in this comparison are IoU (intersection over union) and mAP (mean average precision). IoU is a metric that measures the degree of overlap between the predicted bounding boxes and the ground truth bounding boxes. It is commonly used to evaluate the accuracy of object localization. In the evaluation stage, a threshold of 0.5 IoU is typically considered, meaning that a predicted bounding box is considered a correct detection if its IoU with a ground truth box is above 0.5.

Mean average precision is a measure that combines precision and recall values across different IoU thresholds. Average precision (AP) is calculated for each class individually, and mAP is the average AP across all classes. AP quantifies the precision-recall trade-off and provides a single value to evaluate the overall performance of an object detection algorithm.

These values are likely obtained by evaluating the algorithms on a specific dataset or set of test images. By comparing the AP and mAP values, you can assess the performance of the proposed algorithm relative to the reference YOLOv3 algorithm. Figure 7 presents the AP and mAP values for the first model algorithm as well as the YOLOv3 algorithm from reference [19]. The average accuracy of the weapon class is 4.95%, and the average accuracy of the algorithm is 98.24%, which is improved compared to the YOLOv3 and YOLOv3-SPP algorithms.

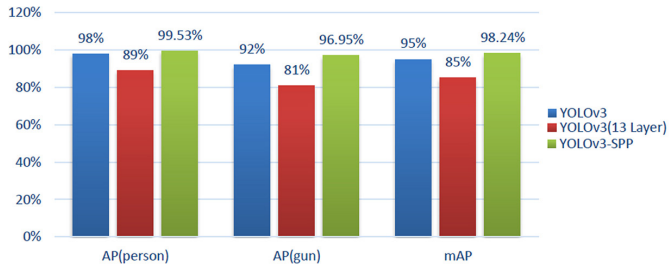

Fig. 7. Algorithm prediction comparison YOLOv3-SPP with algorithm YOLOv3

However, based on the information provided, it appears that the average accuracy and average accuracy of the first model (YOLOv3-SPP) are compared with two other variants of the YOLOv3 algorithm: one with 53 to 76 convolutional layers in its feature extraction network and the other with 13 convolutional layers. In our method (YOLOv3-SPP), the average accuracy of the weapon class has increased by 4.95% compared to the other variants. This suggests that the addition of the SPP module and the use of appropriate anchor boxes have positively impacted the detection accuracy for the weapon class.

Additionally, the first model achieves an average accuracy of 98.24%, indicating strong performance in detecting objects. This high accuracy score suggests that the algorithm effectively identifies and localizes objects within the dataset. Figure 8 displays the detection results of the YOLOv3-SPP algorithm on various test images.

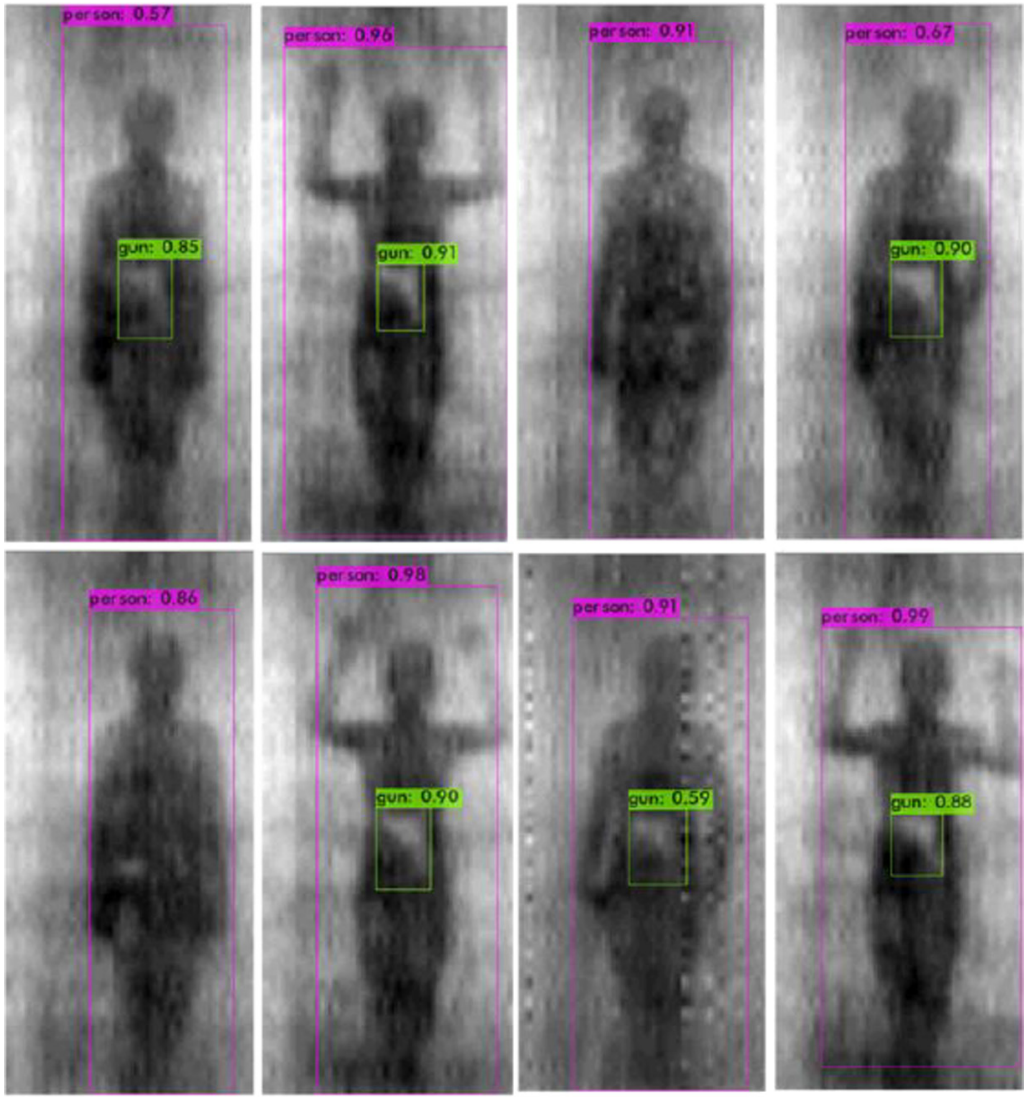

Fig. 8. Prediction results of the first proposed algorithm

Table 2 contains performance metrics such as accuracy, precision, recall, IoU, mAP, or other relevant evaluation measures for the proposed algorithms, along with a comparison to the YOLOv3 algorithm. Additionally, the table includes statistical measures such as standard deviation or p-values to evaluate the significance of

observed differences. This information offers insights into the statistical significance and reliability of the reported results.

Table 2. Evaluation results of proposed method and YOLOv3 algorithm

Method	Accuracy	Precision	F1 Score	Avg IoU
YOLOv3 [26]	94%	96%	95%	74.69%
YOLOv3-SPP	97%	94%	96%	78.04%

5 CONCLUSION

The experimental results obtained from the evaluation of the proposed algorithms for object detection and recognition have provided valuable insights and observations. These findings contribute to the understanding of the effectiveness and efficiency of the algorithms and their potential applications. Firstly, the addition of the SPP module to the YOLOv3 algorithm structure has shown significant improvements in object detection and recognition. By adding the SPP module, the algorithms become more scale-invariant, accurately detecting objects of different sizes in the input image. This addresses one of the limitations of the original YOLOv3 algorithm and improves its ability to handle objects of different sizes. Furthermore, the use of appropriate anchor boxes, obtained through the k-means clustering algorithm, has also played a crucial role in improving the algorithms' performance. The experimental results have shown that the proposed algorithms outperform the original YOLOv3 algorithm in terms of average accuracy. Specifically, the proposed model, which incorporates the SPP module, achieved a significant increase in accuracy for the weapon class, indicating its effectiveness in detecting specific objects of interest. This improvement highlights the relevance of the proposed algorithms for applications requiring precise object recognition, such as security or surveillance systems. The proposed algorithms, which incorporate the SPP module and appropriate anchor boxes, have shown significant improvements in object detection and recognition. The experimental results indicate enhanced accuracy, precision, and recall rates compared to the original YOLOv3 algorithms [25]. The addition of the SPP module enables better scale invariance, while the appropriate anchor boxes aid in accurate localization. Although there are limitations and further improvements that can be explored, the proposed algorithms present a valuable contribution to the field of object detection and recognition. Their effectiveness in detecting specific objects, such as weapons, holds promise for security-related applications.

6 REFERENCES

[1] G. Menghani, "Efficient deep learning: A survey on making deep learning models smaller, faster, and better," *ACM Comput. Surv.*, vol. 55, no. 12, pp. 1–37, 2023. https://doi.org/10.1145/3578938

[2] H. Huang, H. Zhou, X. Yang, L. Zhang, L. Qi, and A.-Y. Zang, "Faster R-CNN for marine organisms detection and recognition using data augmentation," *Neurocomputing*, vol. 337, pp. 372–384, 2019. https://doi.org/10.1016/j.neucom.2019.01.084

[3] R. E. González, R. P. Munoz, and C. A. Hernández, "Galaxy detection and identification using deep learning and data augmentation," *Astronomy and Computing*, vol. 25, pp. 103–109, 2018. https://doi.org/10.1016/j.ascom.2018.09.004

[4] P. Oza, V. A. Sindagi, V. V. Sharmini, and V. M. Patel, "Unsupervised domain adaptation of object detectors: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2023. https://doi.org/10.1109/TPAMI.2022.3217046

[5] P. Kaur, B. S. Khehra, and E. B. S. Mavi, "Data augmentation for object detection: A review," in *IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, IEEE, 2021, pp. 537–543. https://doi.org/10.1109/MWSCAS47672.2021.9531849

[6] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023. https://doi.org/10.1109/JPROC.2023.3238524

[7] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *Multimed. Tools Appl.*, vol. 82, no. 6, pp. 9243–9275, 2023. https://doi.org/10.1007/s11042-022-13644-y

[8] H. Dai *et al.*, "Chataug: Leveraging chatgpt for text data augmentation," *arXiv preprint arXiv:2302.13007*, 2023.

[9] "Data augmentation," smartly.ai. https://docs.smartly.ai/docs/typos-robustness

[10] H. Yu *et al.*, "A multi-stage data augmentation and AD-ResNet-based method for EPB utilization factor prediction," *Autom. Constr.*, vol. 147, p. 104734, 2023. https://doi.org/10.1016/j.autcon.2022.104734

[11] J. Zhang, C. Li, Y. Yin, J. Zhang, and M. Grzegorzek, "Applications of artificial neural networks in microorganism image analysis: A comprehensive review from conventional multilayer perceptron to popular convolutional neural network and potential visual transformer," *Artif. Intell. Rev.*, vol. 56, no. 2, pp. 1013–1070, 2023. https://doi.org/10.1007/s10462-022-10192-7

[12] Y. Wang *et al.*, "Multi-modal 3d object detection in autonomous driving: A survey," *Int. J. Comput. Vis.*, pp. 1–31, 2023. https://doi.org/10.2139/ssrn.4398254

[13] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," *Digit Signal Process*, vol. 126, p. 103514, 2022. https://doi.org/10.1016/j.dsp.2022.103514

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Adv. Neural Inf. Process Syst.*, vol. 28, 2015.

[15] F. K. Al-Shammri, A. S. Mohammed, and F. V. Çelebı, "A combined method for object detection under rain conditions using deep learning," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, IEEE, 2022, pp. 1–8. https://doi.org/10.1109/HORA55278.2022.9799899

[16] S. Zheng, Y. Song, T. Leung, and I. Goodfellow, "Improving the robustness of deep neural networks via stability training," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4480–4488. https://doi.org/10.1109/CVPR.2016.485

[17] J. Yim and K.-A. Sohn, "Enhancing the performance of convolutional neural networks on quality degraded datasets," in *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, IEEE, 2017, pp. 1–8. https://doi.org/10.1109/DICTA.2017.8227427

[18] Y. Tian, K. Pei, S. Jana, and B. Ray, "Deeptest: Automated testing of deep-neural-network-driven autonomous cars," in *Proceedings of the 40th International Conference on Software Engineering*, 2018, pp. 303–314. https://doi.org/10.1145/3180155.3180220

[19] Z. Chen, R. Khemmar, B. Decoux, A. Atahouet, and J.-Y. Ertaud, "Real time object detection, tracking, and distance and motion estimation based on deep learning: Application to smart mobility," in *2019 Eighth International Conference on Emerging Security Technologies (EST)*, IEEE, 2019, pp. 1–6. https://doi.org/10.1109/EST.2019.8806222

[20] L. Pang, H. Liu, Y. Chen, and J. Miao, "Real-time concealed object detection from passive millimeter wave images based on the YOLOv3 algorithm," *Sensors*, vol. 20, no. 6, p. 1678, 2020. https://doi.org/10.3390/s20061678

[21] L. Deng, H. Li, H. Liu, and J. Gu, "A lightweight YOLOv3 algorithm used for safety helmet detection," *Sci. Rep.*, vol. 12, no. 1, p. 10981, 2022. https://doi.org/10.1038/s41598-022-15272-w

[22] C. Gong, A. Li, Y. Song, N. Xu, and W. He, "Traffic sign recognition based on the YOLOv3 algorithm," *Sensors*, vol. 22, no. 23, p. 9345, 2022. https://doi.org/10.3390/s22239345

[23] S. Liu *et al.*, "Dab-detr: Dynamic anchor boxes are better queries for detr," *arXiv preprint arXiv:2201.12329*, 2022.

[24] H. Zhang *et al.*, "Dino: Detr with improved denoising anchor boxes for end-to-end object detection. arXiv 2022," *arXiv preprint arXiv:2203.03605*, 2022.

[25] P. Ma *et al.*, "A state-of-the-art survey of object detection techniques in microorganism image analysis: From classical methods to deep learning approaches," *Artif. Intell. Rev.*, vol. 56, no. 2, pp. 1627–1698, 2023. https://doi.org/10.1007/s10462-022-10209-1

[26] Pang, Lei *et al.*, "Real-time concealed object detection from passive millimeter wave images based on the YOLOv3 algorithm," *Sensors*, vol. 20, no. 6, p. 1678, 2020. https://doi.org/10.3390/s20061678

7 AUTHORS

Ismael M. Abdulkareem, currently a Phd candidate, Information Technology Engineering Department, College of Engineering, University of Qom, Qom, Iran.

Faris K. AL-Shammri, Graduated with a bachelor's degree from Department of Computer Networks Engineering, Iraqia University, Baghdad, Iraq. He obtained on the degrees of M.Sc in Computer Engineering, Karabuk University, Turkey. Currently working as an assistant lecturer at Biomedical Engineering Department, College of Engineering, University of Warith Al-Anbiyaa, Karbala, Iraq. His main research interests and subjects are the Internet of Things (IoT), Artificial intelligence and Information Technologies (E-mail: faris.kar@uowa.edu.iq).

Noor Aldeen A. Khalid, received his Ph.D. in Computer Engineering from Universiti Malaysia Perlis (UniMAP), School of Computer and Communication Engineering since in 2022. Studied his M.Sc. at UniMAP, Malaysia, from 2016 until 2017. He is a senior lecturer at Medical Instruments Techniques Engineering Department, Bilad Alrafidain University College. His main research interests are pattern recognition and image processing (E-mail: dr.nooraldeen@bauc14.edu.iq).

Natiq A. Omran, holds a bachelor's degree in biomedical engineering from the University of Baghdad in 2006 and a master's degree in the same specialty from Al-Nahrain University in 2015. PhD student in biomedical engineering at the research stage at Al-Nahrain University (E-mail: natikaziz81@gmail.com).