Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 7 (2025) 2292

Journal of the Nigerian Society of Physical Sciences

A blind navigation guide model for obstacle avoidance using distance vision estimation based YOLO-V8n

Ebere Uzoka Chidi[a], Edward Anoliefo[a,*], Collins Udanor[b], Asogwa Tochukwu Chijindu[c], Lois Onyejere Nwobodo[d]

[a]*Department of Electronics Engineering, University of Nigeria, Nsukka, Enugu State Nigeria*
[b]*Department of Computer Science, University of Nigeria, Nsukka, Enugu State, Nigeria*
[c]*Department of Computer Science, Enugu State University of Science and Technology*
[d]*Department of Computer Engineering, Enugu State University of Science and Technology*

Abstract

Obstacle is an object positioned along a path of propagation with the potential to cause a collision and hence, an accident. While there are many computer vision models which can detect objects, there is gap in their ability to different between actual objects and obstacles. The aim of this paper is to develop a blind navigation guide model for obstacle avoidance using distance vision estimation-based YOLO-V8n. To achieve this, an improved data model was developed using the MS COCO dataset and primary data collected from several indoor environments. Then, the YOLO-V8n architecture was improved by adding a Weighted Feature Enhancement (WFE) model to the backbone for improved feature extraction, and Bi-directional Feature Pyramid Network (Bi-FPN) was applied to the neck to improve multi-scale feature representation. In addition, a Distance Vision Estimation (DVE) algorithm was developed and applied to the Bi-FPN before connecting it to the head of the YOLO-V8n to facilitate simultaneous object detection and distance measurement in real-time video. Furthermore, the issue of bounding box overlap in the model was addressed by applying a Wise Intersection over Unit (WIoU) loss function. Collectively, these formulated the new transfer learning algorithm called YOLO-V8n+WFE+Bi-FPN+DVE+WIoU used in this work for high-level obstacle detection and distance estimation. The model was trained considering different experimental architectures of the YOLO-V8 and loss functions, respectively, and then evaluated with precision, recall, mean absolute precision, and average precision, respectively, before validation through comparative analysis. Upon selection of the best model, it was further validated through comparison with other state-of-the art algorithms before deployment for obstacle avoidance in an indoor environment, having satisfied the condition of reliability. Real world testing of the model was performed at four different indoor sites, and the results showed that while the model was able to correctly classify objects, it could also measure their distance accurately, thereby making it suitable for deployment as a blind vision guide navigation system.

DOI:10.46481/jnsps.2025.2292

Keywords: YOLO-V8n, COCO dataset, Blind guide, DVE, WFE

Article History :
Received: 08 August 2024
Received in revised form: 21 October 2024
Accepted for publication: 23 October 2024
Available online: 15 December 2024

Communicated by: O. Akande

1. Introduction

"In the gaze of one's eyes, the beauty of the world unfolds, reflecting the interconnected dance of nature, where every leaf,

*Corresponding author: Tel.: +234-806-363-0992.
Email address: edward.anoliefo@unn.edu.ng (Edward Anoliefo)

every sunrise, and every object holds a unique verse in the symphony of existence. According to Ref. [1], the human eyes are gateways to the soul," and when they are closed, "the beauty of the outside world drops dead." The ability to see is a gift bestowed upon man to bear witness to the outside world's visual perception. To see, the eyes collect light from the environment or object through the lens and cornea. These lights are converted to electrical signals using photoreceptors, which are cells within the retina, and then transmitted to the part of the brain called the visual cortex for interpretation into visual information [2]. Vision facilitates lots of human activities such as navigation, mental alertness, interaction, sport, and entertainment, to mention a few [3]. However, Ref. [4] revealed that over 240 million people worldwide do not enjoy this grace of vision due to visual impairment-related diseases. Hence, it is necessary to create alternative innovations that help visually impaired people, particularly the blind, interact with objects around them and navigate freely without collision or accident.

Traditionally, the approach for navigation by the blind involves the use of walking sticks and physical touches to detect obstacles around their trajectory path and avoid them [3]. Today, the process of navigation for the blind has been transformed with the invention of Computer Vision Technology (CVT). CVTs are Artificial intelligence (AI)-based systems that allow computers to see, localize the position of objects, and recognize the contents of digital images or videos [5].AI systems such as deep learning [6] are specialized algorithms carefully designed to facilitate the real-time application of CVTs for the detection and localization of objects in video images. In the context of blind navigation, the CVTs are applied for object detection as the initial step, before conversion to sound for the person's hearing and navigation.

According to Ref. [3], while computer vision has greatly improved through innovations such as 3D models, robust feature detection, and real-time image matching [7], one of the biggest challenges remained how to detect objects of interest in clustered scenes [8], detection of objects with dynamic behavior [9], detection of actionable objects [10], and real-time object detection [11]. To address these problems, research has leveraged the application of deep learning algorithms, specifically Convolutional Neural Networks (CNN) [11], to develop state-of-the-art models for enhanced object detection and classification applications. These models are classified into 1-stage detectors such as short single multi-box detectors [12], You-Can-Only-Look-Once (YOLO) series [12, 13], and then 2-stage detectors such as Recurrent-CNN (R-CNN) [14], Fast-R-CNN [15], Mask-CNN [16], and Faster-R-CNN [17].

Overall, CNN models are characterized by high accuracy of classification in object detection problems; however, in the context of reliability [18], it was revealed that the 1-stage object detection model is more reliable for real-time object detection tasks because of its speed of detection and classification when compared with its 2-stage counterparts. Among the 1-stage detectors, the You Can Only Look Once (YOLO) series has continued to gain increased research attention when solving real-time classification problems [19]. The application of the YOLO series in object detection was based on several advantages, which include speed of classification, very high accuracy potential, integrated output made of classification results, confidence scores, and labeling capabilities. However, while the YOLOV series offers these benefits, occlusion has remained one of the major constraints that has hindered the successful realization of these goals by YOLOV [20]. Occlusion occurs when an object of interest is hidden in the background by another object [20]. According to Ref. [21], it is rare for an object to exist in isolation, hence posing a significant challenge in various applications of computer vision, including tracking, 3D reconstruction, and object recognition. Occlusion can occur in many forms [21], each informed by a particular area of application. For instance, same object occlusion arises in applications such as autonomous vehicles, surveillance, health care, agriculture, blind navigation guide systems, etc. In this applications, objects appears similar and also in a clustered background, hence making it very challenging for accurate segmentation. While the requirements to solve these problems differ in different applications, in this paper, the occlusion problem is defined considering a blind guide navigation system for obstacle avoidance as shown in the Figure 1.

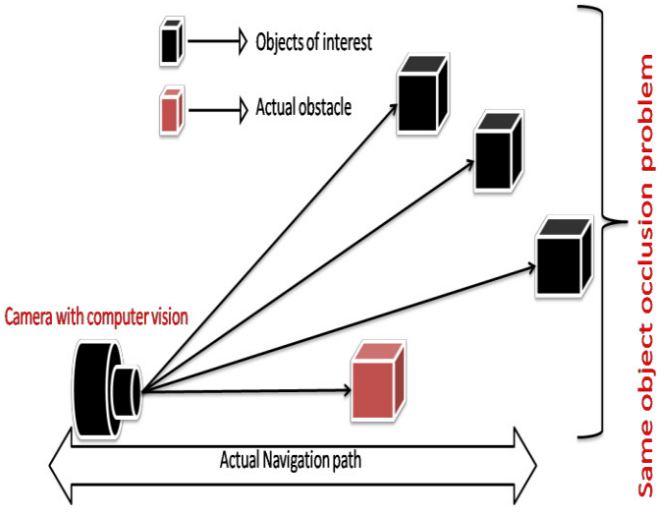


Figure 1: Description of the occlusion problem.

Figure 1 demonstrates a specialized kind of occlusion problem which occurs for a blind guide navigation system. During blind people's movement with assumed wearable computer vision-based devices for obstacle detection and avoidance, many times, the multiple same object of instance is captured as shown with the boxes; while the red box is the actual obstacle along the path of propagation, the user receives multiple notifications on the other classified objects, and this presents several issues, including confusion, indecision on the actual object, a model obstacle problem, false alarms, and even accidents. Hence, there is a need for a model that, in this kind of situation, is able to identify the OOI and ignore other multiple objects until they actually model obstacles in the propagation path.

Some of the innovations tailored towards solving this problem include [22], who proposed a self-supervised method that

recovers the ordering of occlusion and completes the invisible part of the occluded object using R-CNN, while Ref. [23] integrated deep convolutional neural networks and compositional models to address the partial occlusion problem. More so, Ref. [24] proposed new loss functions to enforce predicted boxes to locate compactly the ground truth objects that are far away from other objects. While these studies addressed occlusion problems, they did not solve the same object occlusion. More recent innovation with YOLOV algorithms applied attention mechanisms [25–29], optimization of non-maximum suppression [30–34], while others applied instant segmentation [35–42] approaches to solve this problem of same object occlusion, but despite the success, there is a need for a model that can identify the right obstacle even in a clustered scene and correctly classify it so as to facilitate free navigation. Such model will be vital to facilitate autonomous navigation by the blind without collision. The paper contributions are as follows:

i. A systematic literature review on real-time occlusion management using transfer learning algorithm and identify new research gap tailored towards reliable obstacle avoidance system for blind guide navigation application.
ii. A new YOLO-V8n model with improve multiscale feature representation and extraction process using weighted feature enhancement algorithm and bi-directional feature pyramid network.
iii. A simple but effective distance vision estimation algorithm capable of measuring objects distance from computer vision image in real-time
iv. A reliable model for obstacle avoidance specific to blind guide navigation application systems

2. Related works

Recently, the issue of object occlusion has continued to gain research attention because the reliability of every real-time object detection model is dependent on it. To achieve this, Ref. [25] used the convolutional block attention module to optimize the extraction process of YOLOV-7 and then improve the classification of object recognition. In another study Ref. [26] applied hard example mining using hard positive and negative group techniques to extract diverse features of images and then trained YOLOV-4 to address occlusion. The improved YOLOV-4 was trained using the GOPRO dataset of self-driving autonomous vehicles, and the results reported an F1-score value of 90% and a mean average precision (mAP) value of 90.49%. In another study [27], a hard-switch layer, batch normalization, and convolutional block were used to develop an attention mechanism and integrate it as a YOLO-Attention Convolutional Neural Network (YOLO-CAN) for improved object classification. The results when evaluated showed that the addition of the attention mechanism improved the accuracy when tested on tiny objects, from 4% to 55%, and mAP reached 18.2%. Ref. [28] combined Receptive Field Enhancement (RFE) module, Normalized Gaussian Wasserstein Distance (NWD), and Separated and Enhancement Attention Module (SEAM) to improve face recognition in real time using YOLOV-2. The RFE was used to improve feature extraction of smaller face particles, while weight disparity between the features was addressed using a weight function slide. NWD was used for loss evaluation, while occlusion was addressed using SEAM. After training the model with the WilderFace dataset and evaluating it, the detection rate for faces was reported at 87.7%. While Ref. [29] addressed the issues of occlusion using multiple datasets and a distributed loss function to train YOLOV-5 and improve classification performance with an accuracy of 6% when compared to the traditional YOLOV-5, Overall, it was observed that while these studies focused on improving attention mechanisms to solve occlusion problems, issues of same-object occlusion were not addressed.

The identification of the same object occlusion has been considered by Ref. [30], who applied soft-NMS to address occlusion of the same object. This was achieved using a predefined threshold that suppresses bounding boxes that do not satisfy the threshold detection score. While this approach is good, it may not be reliable to classify multiple overlapping objects. To solve this problem, an adaptive NMS was proposed by Ref. [31], using the density of objects as a variable to differentiate and classify OOI. In the same vein, Ref. [32] modeled the relationships between the learning pair-wise of multiple OOI within a clustered scene to address issues of false alarm and improve discrimination between nearby objects in real-time object classification. In Ref. [33], the application of Deterministic Point Process (DPD) was applied as an alternative to NMS. The DPD, through the selection of diverse detection sub-sets, was able to address object overlap, while Ref. [34] combined NMS and IoU-based methods to address the issues of redundancy in object detection tasks. This was achieved through the selection of boxes closest to other boxes within a cluster and the deletion of highly interfering bounding boxes.

Instant segmentation is another popular approach used by researchers to address occlusion problem in object detection. It involves the separation of objects and predicting dense areas [35], and to achieve this, a dynamic and sparse Related Semantic Perceived Attention mechanism (RSPA) for adaptive perception of different semantic information of various targets during feature extraction and also search for adjacency matrix in regions with high semantic correlation was presented [35]. In addition, GSConv [36], which contains two symmetric kernel-1 convolutions, a simple attention mechanism [37], and an inverted bottleneck, was applied to address issues of redundancy and strengthen the concatenation of features during feature selection. Finally, a Mixed Receptive Field Context Perception Module (MRFCPM) was applied for multi-scale feature representation and trained on COCO dataset to generate an effective sparse attention model. The evaluation results reported 22.1 detection times and a mAP of 45.2, which is good but needs room for improvement. In another study Ref. [38], the instance segmentation of tomato plants was improved by replacing C2f [39] of the YOLOV-8 with RepBlock module [40], while Sim convolution with a rectified linear unit activation function was added instead of the conventional sigmoid convolution to boost feature extraction. Ref. [41] presents a new framework called SMFF-YOLO and applies it, for instance segmentation problem management. To achieve this, a swin transformer network and

convolution were fused inside the head of the SMFF-YOLO, and then an Adaptive Atrous Spatial Pyramid Pooling (AASPP) module and a Bi-directional feature fusion pyramid were applied for feature classification and enhancement of multiscale classification. The experimental results of the model reported a mAP of 42.4% when tested with UAVDT dataset, while Ref. [42] applied a semantic segmentation dense U-shaped Network (UNet) network to improve YOLOV-4. The YOLOV-4 was used for the classification of salient objects in the image, while the UNet was applied to address the occlusion problem in the object background. Overall, while these studies have been able to address occlusion through segmenting OOI, there is still a general gap in the need to classify objects from multiple OOI.

3. Materials and method

The methodology used for this paper began with the data collection of selected indoor objects, which are common obstacles. These objects include chairs and tables and were prepared through augmentation, annotation, and labeling before being integrated with the MS COCO dataset. YOLO-V8n was adopted as the transfer learning algorithm of choice and then improved using a proposed weight enhancement algorithm that is connected to the backbone to enhance the feature extraction process. In addition, a bi-directional feature pyramid network was applied to improve multiscale representation of features and speed up the model operation process. To address issues of bounding boxoverlap in the model output, a wise intersection over unit loss function was applied, while a distance vision estimation algorithm was also added to the head to measure object distance from the camera focal point and then apply to differentiate between objects and obstacles. The model was trained to generate the real-time obstacle avoidance model, which was comparatively analyzed considering other YOLO-V8n architectures, loss functions, and other state-of-the art obstacle avoidance algorithms. Finally, the selected model was deployed for real-time obstacle avoidance and validated through practical experiments.

3.1. Data collection

The primary dataset used for the work was an improved MS COCO dataset [43]. The COCO dataset is made of 328,000 images of everyday objects, representing 91 different objects with a total of 2.5 million labeled instances. This dataset was improved by a secondary dataset made of two classes (chairs and tables) collected from several indoor environments using a ZED2i camera for 20 days, with 50 pictures taken per day, given a total sample size of 1000 images. The number of chairs collected was 556 images, while the number of tables collected was 444 images. These two objects were selected because of their popularity in most indoor environments, and are often positioned in open space, thereby indirectly posing as obstaclse. The data diversity was artificially applied to the images using data augmentation processes such as contrast and brightness adjustment techniques. Further pre-processing steps, such as annotation and labeling, were also applied to the dataset before integration with the MS COCO database.

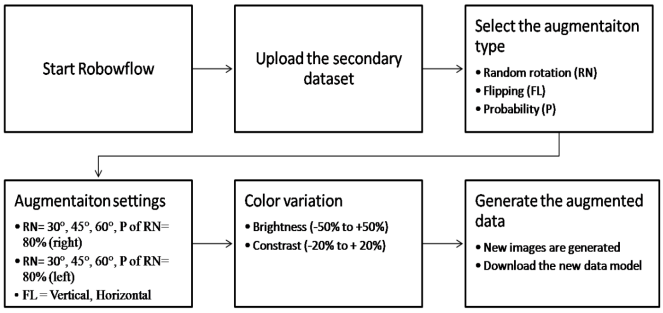


Figure 2: Block diagram of the data augmentation process.

3.2. Data augmentation and preparation

The process of augmentation was necessary to address the issues of over-fitting during training of models. To achieve this augmentation process, Robowflow tool was used, while considering various data augmentation techniques such as random rotation, contrast adjustment, and flipping. The step used was presented in Figure 2.

Figure 2 presents the steps applied to augment the new dataset of tables and chairs classes created using Robowflow tool. This data were imported into the tool and then the respective augmentation approach such as rotation and flipping were applied. Upon the techniques selection, the operation settings were configured considering diverse rotation angles and position such as left and right respectively. In addition, the data diversity was captured after each steps using contrast and brightness adjustment within the defined threshold, before generating the new dataset.

3.3. Data annotation

The annotation of the new data model was also performed using Robowflow tool. The sequence for the annotation step involves loading the processed downloaded dataset back into the Robowflow and then creates annotation task and type which in this case is bounding box and classification. This precedes the manual annotation of the image in the environment by assigning class labels and bounding as needed. After the process, the images were reviewed and then exported into COCO format, to allow seamless integration into the existing COCO dataset and then create the improve COCO data version.

3.4. YOLOV-8 model

This work applied the YOLOV-8n [44] as the transfer learning model of choice for the object classification process. The YOLOV-8n consists of four sections: the input layer, which is responsible for the enhancement of data; the backbone, which performs the feature extraction process; the feature maps, which are fed to the neck for fusioning; and the head, which decouples the output into class probability, label, and bounding box [45]. Figure 3 presents the architecture of the existing YOLO-V8n.

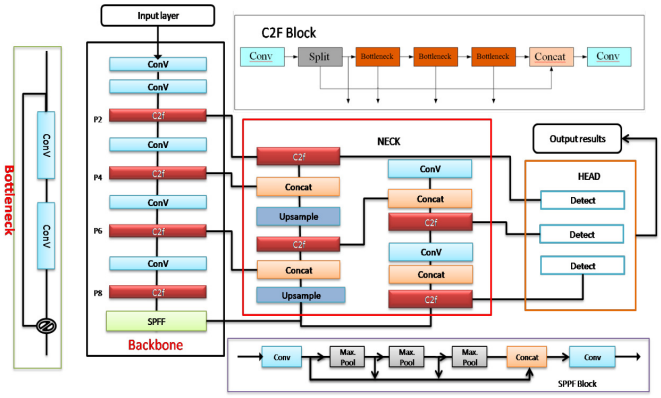


Figure 3: Existing YOLOV-8n model [46].

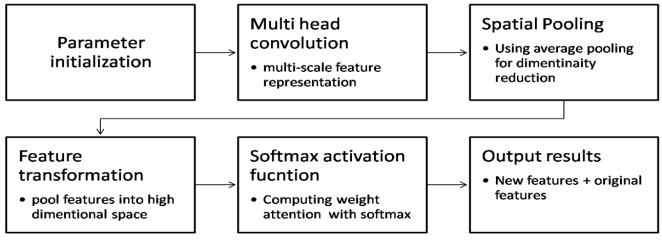


Figure 4: Block diagram of the weighted feature enhancement.

3.4.1. *Backbone with weighted feature enhancement*

The backbone of this YOLOV-8 applied Spatial Pyramid Pooling Fast (SPPF) for the optimization of receptive field, but Ref. [45] revealed that its inability for multi scale feature identification makes it weak for classification of occluded objects. This is becauseocclusion objects are characterized with low pixel information, because some of the parts are hidden in the background of the image and hence need attention mechanism to help the model identifies these features. To achieve this, there is need for an attention mechanism which prioritized the available visible part of the image and then extracts the best features to facilitate improved classification success. This paper propose a WFE approach which utilizes the Global Average Pooling (GAP) to condense feature maps along channel dimension and provide a compact representation of features.

3.4.2. *Weighted feature enhancement*

The WFE is made of multi-head convolution which introduces multiscale feature representation, then the features are extracted using average pooling techniques to allow a global extraction of the features on the strides while preserving channel-wise contents. The features are transformed using rectified linear unit activation function, with the attention weight computed across spatial dimensions using softmax activation function, while the weights are summed and combined with the original features to form the new output as shown in the Figure 4.

To mathematically explain the WFE in Figure 4, let the input X represent the feature as with $H \times W \times C$ as the height, weight and channel of the image input; F_x donates the number of filters and strides presented as S, with the output features defined as Y, Z is the feature map, and K is the number of attention heads.

The WFE algorithm (Algorithm 1).

1. Start
2. Parameter initialization
3. Initialize weights (W_k) and bias $B_k$for each k head.
4. $Z_k = \text{Conv2}D\ (X,\ W_k) + b_k$ for each k %% for multi head convolution
5. $Z_k = Z\ (H * W * K)$
6. Spatial pooling with average technique on Z with strides (Pool (Z) = spatial pooling (Z, stride)
7. Outcome of pooling = $\frac{H_{\text{image}}}{\text{stride}} * \frac{W_{\text{image}}}{\text{stride}} *$ K * filter
8. Feature transformation = F_t = ReLU (Conv2D (pool (Z), W_f)+b_f) :
9. Where W_f is the weight for feature transformation b_f is the bias.
10. To compute the weight of the attention, the softmax was applied as $A = softmax\ (Conv2D\ (F,\ W_a) +\ b_a)$ while a is the attention mechanism.
11. The weighted sum of the features output = $Y_w = Y + X$
12. End

3.4.3. *Bi-directional feature pyramid network with distance vision estimator*

The neck of the model utilized the Cross-Stage Partial Network Fusion (C2f) module and a combination of the Path Aggregation Network (PAN) and Feature Pyramid Network (FPN). The FPN adapts channel-deep features to shallow layers to allow high-level insight in feature map identification. The PAN, on the other hand, allows for precise data positioning in an upward direction from superficial layers to the deep, rich feature strata [46]. The combination of PAN and FPN in the conventional YOLOV8 as a PANET module allows for a mastery amalgamation of shallow deep features and also detection of deep features, thereby bolstering the quality of feature extraction [47]. However, in Ref. [46], it was revealed that the conventional PANET suffers some setbacks, which include poor information flow between different feature levels and not very efficient multiscale feature fusion. To solve this problem, the Bi-directional FPN innovated by Ref. [46] was adapted and used to improve the neck of the YOLOV-8. The Bi-directional FPN was achieved by applying an extra two lateral connection paths to the existing PANET structure. The aim was to allow for adaptive preservation and identification of raw features that were extracted from the network backbone into the detection feature map [47, 48]. More so, the p2 layer and additional detection head were integrated into the architecture neck to allow for a more expansive feature map size and fast convolutional process and were connected to the first head. In the context of obstacle avoidance for blind guide navigation systems, applying this model the way it is will raise issues of reliability because, while the model will accurately classify objects, it will struggle to differentiate between objects and actual obstacles. To address

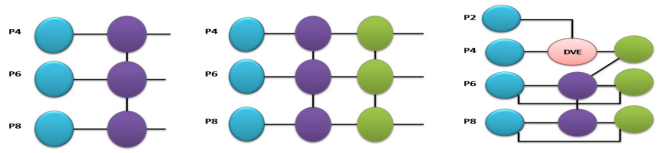


Figure 5: (left) FPN, (middle) PANET, and (right) Bi-FPN with DE.

these issues, there is a need for a distance vision estimation process, which was proposed to be connected with the p2 layer to the head of the YOLO-V8. Figure 5 presents the FPN, PANET, and Bi-FPN with the Distance Vision Estimator (DVE).

3.4.4. *Distance vision estimation (DVE)*

To achieve this a distance computation algorithm was developed, which used key spatial information from the camera such as bounding box sizes, camera information like pixel distance, coordinates of the bounding box centriod, and distance calculation algorithm to develop an object distance measurement model and integrated in the model head to allow for real-time measurement and hence identification of obstacles.The mathematical model for the distance calculation of objects is defined using the relationship between camera calibration such as focal point defined as F, principle point of coordinated in the captured image defined as P_x and P_y. The object size is defined as $(x_1,\ y_1)$ for the top left and $(x_2\ ,\ y_2)$ for the bottom right corner of the detected image bounding box. Where $W_{\text{image}} = (x_2 - x_1)$ is the weight and $H_{\text{image}} = (y_2 - y_1)$ is the height of the bounding box, where D_{real}define the actual object weigh1t and height. To compute the distance calculation, the pinhole camera model was applied as in equation (1) [49];

$$D = \frac{F \times D_{\text{real}}}{W_{\text{image}} \times H_{\text{image}}}, \tag{1}$$

where $W_{\text{image}}$is the weight of the image bounding box, while the focal length is given as equation (2) [50]:

$$F = \frac{\text{Object size} \times \text{bounding box size}}{\text{object size}}. \tag{2}$$

The DVE algorithm (Algorithm 2).

1. Input
2. P_{in} Predicted input from YOLO-V8
3. Load camera information such as F, H_{image}, W_{image}
4. Output
5. Estimate D
6. Load video
7. If $P_{\text{in}} = true$,
8. Convert H_{image}, *and* W_{image} to D
9. Else
10. Return error
11. End if
12. End

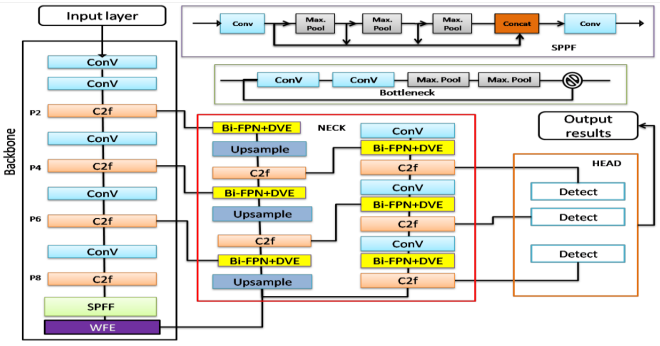


Figure 6: Architecture of the improved YOLOV-8n with WFR and Bi-FPN + DVE.

3.4.5. *System integration of the new YOLOV-8n with WFR and Bi-FPN + DVE*

This section presented the system integration of the new YOLO-V8n presented in this study using the WFE in algorithm 1 was proposed for facilitate multi-scale feature representation of objects, while the Bi-FPN allows a bi-directional extraction of features while the DVE computes the object distance from the camera focal point. The Figure 6 presents the improved YOLOV-8n architecture.

In the Figure 6, the backbone of the model was improved by connected a WFE to the output of the SPPF to improve multiscale feature extraction process. This module gives distinct weights to the characteristics features extracted from multi scales, making sure that the more significant aspects are emphasized, thus bolstering the model's capacity for more accurately objects detection across varying scales. To further allows for more expansive feature representation and improve speed of convolution, the Bi-FPN which is made of the FPN and PANET were applied in the neck. This structure improves the convolution performance, enabling more comprehensive feature representation for better object detection. In addition, a DVE was also integrated with the Bi-FPN to measure the distance of objects classified and label simultaneous with the image bounding box of the identified object.

3.4.6. *Wise-IoU loss function*

Over the years, several loss function for bounding box regression such as the Intersection over Union (IoU) [50], Distance-IoU [51], Generalized IoU [52], Complete IoU [52], Wise-IoU [50] have been presented. Among these IoUs, the Wise-IoU was adopted because [50] revealed that apart from light bounding box support, it also allows easy distance computation. In addition, Ref. [53] compared the traditional IoU techniques and identify WIoU as the best to correctly predict bounding box for obstacle avoidance study, and also focuses on distance between the center points of bounding box and targets when two overlaps [53]. The WIoU loss function in equation (5) is defined using the relationship between IoU in equation (3) and weighted (R) IoUin equation (4).

$$L_{\text{IoU}} = 1 - \text{IoU}, \tag{3}$$

$$R_{\mathrm{WIOU}} = \exp\left(\frac{\left(x - x_{\mathrm{gt}}\right)^2 + \left(y - y_{\mathrm{gt}}\right)^2}{\left(W_g^2 + H_g^2\right)^*}\right), \tag{4}$$

$$L_{WIoUv1} = R_{\mathrm{WIoU}} L_{\mathrm{IoU}}, \tag{5}$$

where [x,y]are the bounding box coordinates and [w, h]are the ground truth coordinates. The ground truth box defined as $\left[x_{\mathrm{gt}} y_{\mathrm{gt}} w_{\mathrm{gt}} h_{\mathrm{gt}}\right]$.the exponential term is a wise factor which prioritizes the importance of center location of the bounding box. In equations (3) to (5), the WIoU minimizes the impact of geometry factor, and also amplify penalty for poor anchor boxes.

3.4.7. Training of the model

The experiments in this paper were carried out in NVIDIA GeForce RTX 3030 Ti Laptop graphical Processing Unit (GPU), operating with Windows 11, and equipped with Intel ® Core™ i9-12900 processor, with 16GB graphics memory. In addition, the CUDA® parallel computing toolbox was also installed in the computer to improve computation power of the GPU and accelerate training speed of the YOLOV-8n model. The training parameter settings for image size is 640 by 640, epochs are 300, the initial learning rate is 0.01, IoU thresholds setting is 0.5 and the batch size is 32. To train our new YOLOV-8n, the improved COCO dataset was splitted into training, test and validation set in the ratio of 80:10:10 and then apply stochastic gradient descent optimization algorithm to train the model, minimizing bounding box classes and object loss through back-propagation process. Finally, test and validation set were used to evaluate the model performance, considering the performance metric in equations (6)-(9), and upon convergence, the model was generated.

3.4.8. Evaluation metrics

To evaluate the performance of the new YOLO-V8n model, several metrics such as Precision (P),this measures the relationship between successful true detection and false detection during the classification process. Recall (R) measures the model's ability to detect all relevant instances. High recall means that the model has a low false negative rate. Average precision (AP) It combines precision and recall into a single metric by plotting precision against recall at various threshold settings, and Mean average precision (mAP) provides a single number to summarize the performance of the model across all classes, making it a widely used metric for object detection tasks. The metrics are mathematically defined in equations (6)-(9), respectively.

$$P = \frac{\mathrm{TP}}{TP + FP}, \tag{6}$$

$$R = \frac{\mathrm{TP}}{TP + FN}, \tag{7}$$

$$AP = \frac{\sum P_{\mathrm{ri}}}{\sum r}, \tag{8}$$

$$mAP = \frac{\mathrm{AP}}{num\_class}, \tag{9}$$

where TP is true positive, FP is false positive, FN is false negative, r is ranks of each instance, and Pri is the precision sum for all instances.

Table 1: Experimental results of YOLOV-8n with IoU loss.

Models	Precision	Recall	mAP
YOLOV-8n + IoU	89.9	80.6	86.4
YOLOV-8n + WFE+ IoU	90.4	81.5	87.9
YOLOV-8n + WFE+ Bi-FPN+ IoU	94.7	84.3	91.5
YOLOV-8n + Bi-FPN+ DVE+ IoU	94.1	83.5	90.7

Table 2: Experimental results of YOLOV-8N with R_{WIoU}.

Models	Precision	Recall	mAP
YOLOV-8n + R_{WIoU}	90.1	81.3	86.8
YOLOV-8n + WFE+ R_{WIoU}	91.7	82.9	89.3
YOLOV-8n + WFE+ Bi-FPN+ R_{WIoU}	96.2	84.6	92.8
YOLOV-8n + Bi-FPN+ DVE+ R_{WIoU}	95.2	83.7	91.5

4. Results and discussions

To evaluate the model developed, several experiments were performed considering the different loss function in equations (3) to (5) and also various formations of the YOLOV-8N models during the development of the proposed YOLOV-8n + Bi-FPN+ DVE using our dataset. Table 1 presents the model performance during experiment with diverse architecture with IoU loss function.

Table 1 presents the comparative results of the YOLO-V8n architectures with IoU as the loss function. From the results it was observed that the alteration in the different components of the model has impact on the performance. For instance it was observed that with WFE was added to the SPPF and trained with our dataset, the performance was better than the conventional standalone YOLOV8n. More so when Bi-FPN was added to the YOLO-V8 + WFE, it was observed that the model performance further improved. The reason was because while the WFE improves the feature extraction process, the Bi-FPN allows for more expansive feature representation in multiscale and hence improves training data quality. The other result showed the integration of DVE on the improved YOLO-V8n with Bi+FPN and WFE and it was observed that the results slightly reduced from YOLOV-8n + WFE+ Bi-FPN+ IoU, but still reported very good results. Overall, the YOLOV-8n + WFE+ Bi-FPN+ IoU which reported the best results with P of 94.7, thus suggesting the models ability to correctly predict object positively, which is good. The recall reported 84.3 as a measure to detect all relevant instances, which is also good, but leaves great room for improvement, while the mAP which summarizes the model performance across diverse object classes reported 91.5 which is also good but leave room for improvement. This improvement was explored by applying R_{WIoU} as the loss function and the retraining the model comparatively, with results collected reported in Table 2. Table 3 presented other results which evaluates the different YOLO-V8n architecture considering L_{WIoU}. Table 4 presents experimental comparative analysis of our model.

Table 2 present the comparative analysis of the four experimental YOLO-V8n models considering R_{WIoU} as the loss function. From the data presented, it was observed that the YOLOV-

Table 3: Experimental results of YOLOV-8n with and $\mathbf{L}_{\mathrm{WIoU}}$.

Models	Precision	Recall	mAP
YOLOV-8n + L_{WIoU}	92.2	83.7	88.2
YOLOV-8n + WFE+ L_{WIoU}	93.9	85.4	90.1
YOLOV-8n + WFE+ Bi-FPN+ L_{WIoU}	98.6	85.5	93.8
YOLOV-8n + Bi-FPN+ DVE+ L_{WIoU}	97.4	84.3	92.9

Table 4: Comparative results of YOLO-V8n with different loss functions.

Models	Precision	Recall	mAP
YOLOV-8n + IoU	89.9	80.6	86.4
YOLOV-8n + R_{WIoU}	90.1	81.3	86.8
YOLOV-8n + L_{WIoU}	92.2	83.7	88.2
YOLOV-8n + WFE+ IoU	90.4	81.5	87.9
YOLOV-8n + WFE+ R_{WIoU}	91.7	82.9	89.3
YOLOV-8n + WFE+ L_{WIoU}	93.9	85.4	90.1
YOLOV-8n + WFE+ Bi-FPN+ IoU	94.7	84.3	91.5
YOLOV-8n + WFE+ Bi-FPN+ R_{WIoU}	96.2	84.6	92.8
YOLOV-8n + WFE+ Bi-FPN+ L_{WIoU}	98.6	85.5	93.8
YOLOV-8n + Bi-FPN+ DVE+ IoU	94.1	83.5	90.7
YOLOV-8n + Bi-FPN+ DVE+ R_{WIoU}	95.2	83.7	91.5
YOLOV-8n + Bi-FPN+ DVE+ L_{WIoU}	97.4	84.3	92.9

8n + WFE+ Bi-FPN+ R_{WIoU} reported the best performance with P of 96.2, R of 84.6 and mAP of 92.8 respectively. This results demonstrates the effectiveness of the R_{WIoU} in predicting bounding box and overlapping ground truth, which collectively improves the model performance.

In the Table 3, the comparative YOLO–V8n results considering L_{WIoU} which combines the R_{WIoU} and IoU to improve bounding box prediction. From the results it was observed that YOLOV-8n + WFE+ Bi-FPN+ L_{WIoU} recoded the best performance for P with 98.6, R with 85.5 and mAP with 93.8 respectively, when compared with other models. This results signifies that the integration of WFE, Bi-FPN was able to improve feature exaction and representation during the training of the model, while the L_{WIoU} leverage the advantages of IoU and R_{WIoU} to improve bounding box and ground truth prediction performance. In the next results, the three loss function and various YOLO-V8n experiments were compared in Table 4, to determine the best model for system integration. Table 4 compared the performance of the experimented YOLO-V8n models considering different loss and YOLO-V8n architectures. From the results, it was observed that YOLOV-8n + WFE+ Bi-FPN+ L_{WIoU} recorded the overall best performance for all the three metrics considered, however in the context of obstacle avoidance for the blind, the model will not be reliable because while it will correctly identify every objects in the computer vision, it will give false alarm and will not be able to different objects from obstacles. In solving this problem, the DVE algorithm connected to the Bi-FPN and integrated in the YOLOV head as a YOLOV-8n + Bi-FPN+ DVE+ L_{WIoU} model for the classification objects and simultaneous distance detection in real time. The results reported 97.4 for P, 84.3 for R and mAP for 92.9, respectively. There results even though measurement wise it is not better then YOLOV-8n + WFE+ Bi-FPN+ L_{WIoU} , however

Table 5: Comparative analysis with other obstacle detection models.

Author	Models	Precision	Recall	mAP
Ref. [54]	SSD	72.8	66.5	70.1
Ref. [55]	Faster RCNN	72.8	66.5	76.4
Ref. [56]	YOLOv5n	86.7	86.8	88.8
Ref. [57]	TPH-YOLO-V5	92.7	82.3	88.4
Ref. [58]	YOLO-V7tiny	73.5	81.2	81.8
Ref. [44]	YOLO-V8n	90.3	80.7	86.3
Ref. [59]	Golf-YOLO	83	95	87.9
Ref. [46]	YOLO-V8 + BI-FPN + SimAM	97.9	91.2	95.8
Our work	YOLOV-8n + Bi-FPN+ DVE+ L_{WIoU}	97.4	84.3	92.9

in terms of reliability is the holy grail of performance evaluation model, the YOLOV-8n + Bi-FPN+ DVE+ L_{WIoU} is the best because it will correctly classify object, differentiate objects from obstacle and facilitate navigation without collision with obstacles. Finally the YOLOV-8n + Bi-FPN+ DVE+ L_{WIoU} was compared with other state of the art algorithms as reported in the Table 5.

(a)

(b)

Figure 7: (a) Test site 1 indoor environment, (b) Test site 2 indoor environment.

(a)

(b)

Figure 8: (a) Test site 3 indoor environment, (b) Test site 4 indoor environment.

(a)

(b)

Figure 9: Result of the obstacle classification model with YOLOV-8n + Bi-FPN+ DVE+ L_{WIoU}. (a) Result at test 2 indoor, (b) Result at test 2 indoor.

Table 5 presents a comparative results of the new model developed for same object occlusion management with YOLOV-8n + Bi-FPN+ DVE+ L_{WIoU} and other state of the art algorithm. From the results, it was observed that new model is the second best in term of P, R and mAP, but overall in the context of reliability for blind vision navigation system, the new model supersedes the other because it has the ability to different obstacle of different occlusions using DVE algorithm. Therefore this model was deployed as a computer vision system for obstacle avoidance to facilitate blind vision navigation system.

4.1. Validation of the new YOLO-V8n model at indoor environment with obstacles

The software was validated at different indoor environments with obstacles scattered across area as shown in the Figures 7 and 8, respectively.

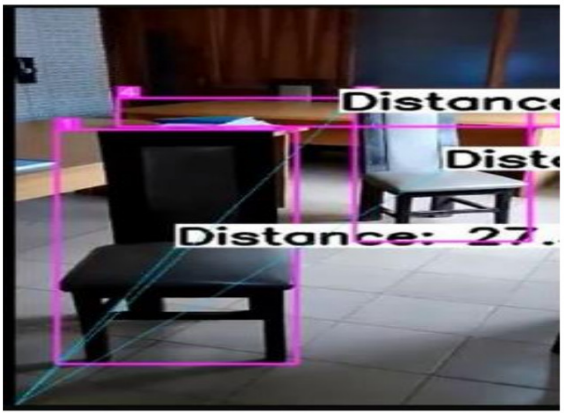
Figures 7 and 8, respectively, showed the different case study indoor test sites where the model was validated through live experiment. The results of the four validate scenarios was presented in Figures 9 and 10, respectively.

Figure 9 presents the result of the model deployment for obstacle avoidance in two different indoor environments. The Figure 9(a) showed the result of the model when tested at site 1, while the Figure 9(b) showed the results of the model when tested at site 2. In another validation scenario considering site 3 and 4, the results are reported in Figure 10.

Figure 10 presents the result of the model deployment for obstacle avoidance in two different indoor environments. Figure 10a showed the result of the model when tested at site 3, while Figure 10b showed the results of the model when tested at site 4. Overall these result haves shown demonstrated success in correctly classifying objects and using distance to differentiate obstacles. This solution will address issues of same multiple object occlusion which has remained an open problem in computer vision.

4.2. Discussion considering success and weakness of our model

Without doubt YOLOV-8n has provided in literature to be successful in solving object detection problems, however chal-



(a)

(b)

Figure 10: Result of the obstacle classification model with YOLOV-8n + Bi-FPN+ DVE+ L_{WIoU}. (a) Result at indoor three, (b) Result at test four indoor.

lenges poses by objects varies with areas of application. In the context of obstacle avoidance, issues such as variability in obstacle size, occlusion, and similarity are a common challenge which affects performance of traditional YOLOV-8n. This problems were addressed through improved feature extraction using WFE, more feature representation with Bi-FPN, obstacle detection with DVE and L_{WIoU}. This unique introduced modules allows our YOLOV-8n to be reliable in correctly classifying obstacles within indoor environments, and to the best of our knowledge is the first to be applied to facilitate autonomous movement y a blind person without collision with obstacles.

4.3. Weakness of our work

The weakness of this work in the context of a blind guide navigation system is that the object identified as an obstacle cannot be communicated to the user because that will require a text-to-speech library which converted the classified object label and convert to speech to notify the user about the obstacle and hence prevent collision. However, this part is recommended for further studies.

5. Conclusion

An obstacle avoidance model has been developed in this paper using YOLOV-8n + WFE + Bi-FPN + DVE+L_{WIoU}. While the existing YOLO-V8n, among other techniques, has successfully presented models capable of object detection and classification in real-time, there is weakness in their application for blind guide vision navigation systems due to issues of multiple object occlusions and clustered backgrounds. This paper addresses the issues of fine-tuning YOLO-V8n with an improved COCO dataset, using data collected from several indoor environments. To improve the performance of the YOLO-V8, a WFE was applied to the SPPF for enhanced feature extraction, while to improve multiscale feature representation, the Bi-FPN was applied in the model neck. In addition, a DVE algorithm was developed and integrated with the Bi-FPN to allow measurement of objects captured by the camera using focal length and bounding box information. The model after training was evaluated through several comparative analyses and then validated in real-world indoor environments. The results showed the ability of the model to correctly classify multiple objects and also measure their distance from the camera, thus making it the perfect obstacle avoidance system for blind guide navigation.

Data availability

The data used for this work are available on https://www.kaggle.com/datasets/clkmuhammed/microsoft-coco-2017-common-objects-in-context.

References

[1] S. Alagarsamy, D. Rajkumar, L. Syamala & L. Niharika, "A real time object detection method for visually impaired using machine learning", International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-6. [Online]. https://doi.org/10.1109/ICCCI56745.2023.10128388.

[2] L. Russel, "How does the eye work?", Optimetrists, 2020. [Online]. https://www.optometrists.org/general-practice-optometry/guide-to-eye-health/how-does-the-eye-work/.

[3] C. Sagana, P. Keerthika, R. Manjula Devi, M. Sangeetha, R. Abhilash, M. Dinesh Kumar & M. Hariharasudhan, *Object recognition system for visually impaired people*, 2021 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), Nitte, India, 2021, pp. 318-321. https://doi.org/10.1109/DISCOVER52564.2021.9663608.

[4] S. Shubham, M. Sushruta, S. Kshira & N. Anand, "Research article vision navigator: a smart and intelligent obstacle recognition model for visually impaired users", Hindawi Mobile Information Systems **2022** (2022), 9715891. https://doi.org/10.1155/2022/9715891

[5] P. E. Kekong, A. Ajah & U. C. Ebere, "Real time drowsy driver monitoring and detection system using deep learning based behavioural approach", International Journal of Computer Sciences and Engineering **9** (2019) 11. http://dx.doi.org/10.14569/IJACSA.2021.0120794.

[6] P. O. Ugwoke, C. N. Udanor & F. S.Bakpo, "Deep learning algorithms for predicting the geographical locations of Pandemic disease patients from Global Positioning System (GPS) trajectory datasets research square", 2023. [Online]. https://doi.org/10.21203/rs.3.rs-2770308/v1.

[7] T. C. Asogwa & C. B. Onah, "Improving the performance of obstacle detection and avoidance autonomous mobile robot using transfer learning technique", International Journal of Real-Time Application and Computing Systems **1** (2022) 85. https://ijortacs.com/paper?id=8op3ixmbzs.

[8] Y. Chen, P. Yang, N. Zhang & J. Hou, "Edge-Assisted Lightweight Region-of-Interest Extraction and Transmission for Vehicle Perception", 2023. [Online]. https://arxiv.org/abs/2308.16417.

[9] S. Reddy, P. Khatravath, N. Surineni & R. Mulinti, *Object detection and action recognition using computer vision*, 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 874-879. https://doi.org/10.1109/ICSCSS57650.2023.10169620.

[10] J. Au, D. Reidand & A. Bill, "Challenges and opportunities of computer vision applications in aircraft landing gear", 2022 IEEE Aerospace Conference (AERO), Big Sky, MT, USA, 2022, pp. 1-10. https://doi.org/10.1109/AERO53065.2022.9843684.

[11] I. C. Oliver, I. J. Odegwo, F. C. Obodoeze & L. O. Nwobodo,"Internet-of-things based real-time accident alert and reporting system for Nigeria", International Journal of Trend in Scientific Research and Development (IJTSRD)**6** (2022) 1171. https://www.researchgate.net/publication/364333800_Internet-Of-Things_Based_Real-time_Accident_Alert_and_Reporting_System_for_Nigeria.

[12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy,S. Reed, C. Y. Fu & A. C. Berg, "SSD: Single shot multibox detector", in *Computer Vision-ECCV 2016. Lecture Notes in Computer Science*, B. Leibe, J. Matas, N.Sebe, M. Welling (Ed.), Springer, Cham, Berlin/Heidelberg, 2016, pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.

[13] E. U. Chidi, C. N. Udanor & E.Anoliefo, "Exploring the depths of visual understanding: a comprehensive review on real-time object of interest detection techniques", 2024. Preprints. [Online]. https://doi.org/10.20944/preprints202402.0583.v1.

[14] R. Girshick, J. Donahue, T. Darrell & J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014, pp. 580–587. https://ieeexplore.ieee.org/document/6909475.

[15] R. Girshick, *Fast R-CNN*, Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015, pp. 1440–1448. https://ieeexplore.ieee.org/document/7410526.

[16] K. He, G. Gkioxari, P. Dollár & R. Girshick, *Mask R-CNN*, Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 2980–2988. https://arxiv.org/abs/1703.06870.

[17] S. Ren, K. He, R .Girshick & J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks", IEEE Trans. Pattern Anal. Mach. Intell. **39** (2017) 1137. https://ieeexplore.ieee.org/document/7485869.

[18] A. Bochkovskiy, C. Y. Wang & H. Y. M. Liao, "YOLOv4: optimal speed and accuracy of object detection". [Online]. https://doi.org/10.48550/arXiv.2004.10934.

[19] E. Oluwaseyi, E. Martins & E. Abraham, "A comparative analysis of YOLOV-5 and YOLOV-7 object detection algorithms", Journal of Computing and Social informatics **2** (2023) 1. http://dx.doi.org/10.33736/jcsi.5070.2023.

[20] T. C. Asogwa, "Vision based behavioural approach to drowsy detection using viola jones algorithm and artificial neural network", International Journal of Trend in Research and Development **9** (2019) 202. https://www.ijtrd.com/papers/IJTRD25202.pdf.

[21] J. Qi, Y. Gao & Y. Hu, "Occluded video instance segmentation: a benchmark", Int J Comput Vis **130** (2022) 2022. https://doi.org/10.1007/s11263-022-01629-1.

[22] X. Zhan, X. Pan, B. Dai, Z. Liu, D. Lin, & C. C. Loy, "Self-supervised scene de-occlusion", 2020. [Online]. https://arxiv.org/abs/2004.02788.

[23] A. Kortylewski, Q. Liu, A. Wang, Y. Sun & A. Yuille, "Compositional convolutional neural networks: a robust and interpretable model for object recognition under occlusion", Int J Comput Vis. **129** (2021) 736. https://doi.org/10.1007/s11263-020-01401-3.

[24] S. Zhang, L. Wen, X. Bian, Z. Lei, & S. Z. Li, "Occlusion-aware R-CNN: Detecting pedestrians in a crowd", 2018. [Online]. https://doi.org/10.48550/arXiv.1807.08407.

[25] J. Wang, J. Wu, J. Wu, J. Wang & J. Wang, "YOLOv7 optimization model based on attention mechanism applied in dense scenes", Appl. Sci **13** (2023) 9173. https://doi.org/10.3390/app13169173.

[26] S. Ryu & K. Chung,"Detection model of occluded object based on yolo using hard-example mining and augmentation policy optimization", Appl. Sci. **11** (2021) 7093. https://doi.org/10.3390/app11157093.

[27] Y. Li, S. Li, H. Du, L. Chen, D. Zhang & Y. Li, "YOLO-ACN: focusing on small target and occluded object detection", IEEE Access Digital Object Identifier **10** (2020) 1109. https://doi.org/10.1109/ACCESS.2020.3046515.

[28] Z. Yu,H. Huang, W. Chen, Y. Su, Y. Liu, & X. Wang, "YOLO-FaceV2: a scale and occlusion aware face detector", 2022. [Online]. https://doi.org/10.48550/arXiv.2208.02019.

[29] Y. Zhao & S. Geng, "Face occlusion detection algorithm based on YOLOV5", Journal of Physics: Conference Series **2031** (2021) 012053. https://doi.org/10.1088/1742-6596/2031/1/012053.

[30] N. Bodla, B. Singh, R. Chellappa & L. S. Davis, *Soft-NMS–improving object detection with one line of code*, Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017. pp. 5561–5569. https://arxiv.org/abs/1704.04503.

[31] S. Liu, D. Huang & Y. Wang, *Adaptive NMS: refining pedestrian detection in a crowd*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019. pp. 6459–6468. https://doi.org/10.48550/arXiv.1904.03629.

[32] Y. Liu, L. Liu, H.Rezatofighi, T. T. Do, Q. Shi & I. Reid, "Learning pairwise relationship for multi-object detection in crowded scenes", 2019. [Online]. https://doi.org/10.48550/arXiv.1901.03796.

[33] S. Some, M. D. Gupta & V. P. Namboodiri, "Determinantal point process as an alternative to NMS", 2020. [Online]. https://doi.org/10.48550/arXiv.2008.11451.

[34] A. J. Shepley, G. Falzon, P. Kwan & L. Brankovic, "Confluence: a robust non-iou alternative to non-maxima suppression in object detection", IEEE Trans. Pattern Anal. Mach. Intell. **45** (2023) 11561–11574. https://ieeexplore.ieee.org/document/10119209.

[35] Q. Zhang, L. Chen, M. Shao, H. Liang & J. Ren, "ESAMask: real-time instance segmentation fused with efficient sparse attention", Sensors **23** (2023) 6446. https://doi.org/10.3390/s23146446

[36] H. Li, J. Li, H. Wei, Z. Liu, Z. Zhan & Q. Ren, "Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles", J Real-Time Image Proc **21** (2022) 62. https://doi.org/10.1007/s11554-024-01436-6.

[37] L. Yang, R. Y. Zhang, L. Li & X. Xie, *Simam: a simple, parameter-free attention module forconvolutional neural networks*, Proceedings of the International Conference on Machine Learning, [Online], 18–24 July 2021. pp. 11863–11874. https://proceedings.mlr.press/v139/yang21o.html.

[38] X. Yue, K. Qi, X. Na, Y. Zhang, Y. Liu & C. Liu, "Improved YOLOv8-Seg network for instance segmentation of healthy and diseased tomato plants in the growth stage", Agriculture **13** (2023) 1643. https://doi.org/10.3390/agriculture13081643.

[39] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng & W. Nie, "YOLOv6: a single-stage object detection framework for industrial applications", 2022. [Online]. https://ar5iv.labs.arxiv.org/html/2209.02976.

[40] K. Weng, X. Chu, X. Xu, J. Huang & X. Wei, "EfficientRep: an efficient repvgg-style convnets with hardware-aware neural network design", 2023. [Online]. https://arxiv.org/abs/2302.00386.

[41] Y. Wang, H. Zou, M. Yin & X. Zhang, "SMFF-YOLO: a scale-adaptive YOLO algorithm with multi-level feature fusion for object detection in uav scenes", Remote Sens **15** (2023) 4580. https://doi.org/10.3390/rs15184580.

[42] J. Hua, T. Hao, L. Zeng & G. Yu, "YOLOMask: an instance segmentation algorithm based oncomplementary fusion network", Mathematics" **9** (2021) 1766. https://doi.org/10.3390/math9151766.

[43] T. J. Alahmadi, A. U. Rahman, H. K. Alkahtani & H. Kholidy, "Enhancing object detection for vips using yolov4_resnet101 and text-to-speech conversion model", Multimodal Technol. Interact **7** (2023) 77. https://doi.org/10.3390/mti7080077.

[44] D. Reis, J. Kupec, J. Hong & A. Daoudi, "Real-time flying object detection with Yolov8", 2023. [Online]. https://doi.org/10.48550/arXiv.2305.09972/

[45] Z. Huangfu & S. Li, "Lightweight you only look once V8: an upgraded you only look once v8 algorithm for small object identification in unmanned aerial vehicle images", Appl. Sci. **13** (2023) 12369. https://doi.org/10.3390/app132212369.

[46] N. Li, T. Ye, Z. Zhou, C. Gao & P. Zhang, "Enhanced YOLOv8 with BiFPN-SimAM for precise defect detection in miniature capacitors", Appl. Sci. **14** (2024) 429. https://doi.org/10.3390/app14010429.

[47] Z. Qu, L. Y. Gao, S. Y. Wang, H. N. Yin & T. M. Yi, "An improved YOLOv5 method for large objects detection with multi-scale feature cross-layer fusion network", Image Vis. Comput. **125** (2022) 104518. https://doi.org/10.1016/j.imavis.2022.104518.

[48] V. Chiley, V.Thangarasa, A. Gupta, A. Samar, J. Hestness & D. De Coste, "RevBiFPN: the fully reversible bidirectional feature pyramid network", 2023. https://doi.org/10.48550/arXiv.2206.14098.

[49] Y. Hao, V. C. Tai & Y. C. Tan, "A systematic stereo camera calibration strategy: leveraging latin hypercube sampling and 2k full-factorial design of experiment methods", Sensors **23** (2023) 8240. https://doi.org/10.3390/s23198240.

[50] Z. J. Khow, Y. F. Tan, H. A. Karim & H. A. A. Rashid, "Improved YOLOv8 model for a comprehensive approach to object detection and distance estimation", IEEE Access **12** (2024) 63754 https://ieeexplore.ieee.org/document/10517525.

[51] J. Yu, Y. Jiang, Z. Wang, Z. Cao & T. Huang, *UnitBox: an advanced object detection network*, Proceedings of the 24th ACM International Conference on Multimedia, pp. 516–520. https://doi.org/10.1145/2964284.2967274.

[52] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid & S. Savarese, *Generalized intersection over union: a metric and a loss for bounding box regression*, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 658-666 (2019). https://doi.ieeecomputersociety.org/10.1109/CVPR.2019.00075.

[53] Q. Zhao, H. Wei, & X. Zhai, "Improving tire specification character recognition in the Yolov5 network", Applied Sciences **13** (2023) 12. https://doi.org/10.3390/app13127310.

[54] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C-Y. Fu & A. C. Berg, "SSD: single shot multibox detector", in *Computer Vision—ECCV,* Lecture Notes in Computer Science, B. Leibe, J. Matas, N.Sebe, M. Welling, (Ed.), Springer, Cham, Switzerland, 2016, pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.

[55] S. Ren, K. He, R.Girshick & J. Sun, *Faster R-CNN: Towards real-time object detection with region proposal networks*, Proceedings of the Advances in Neural Information Processing Systems 28 (NIPS 2015), Montreal, QC, Canada, 7–12 December 2015, pp. 91–99. https://doi.org/10.48550/arXiv.1506.01497.

[56] J. Wang, Y. Chen, Z. Dong & M. Gao, "Improved YOLOv5 network for real-time multi-scale traffic sign detection", Neural Comput. Appl. **35** (2022) 7853. https://doi.org/10.1007/s00521-022-08077-5.

[57] C. Wang, W. He, Y. Nie, J. Guo, C. Liu, K. Han & Y. Wang, "Gold-YOLO: efficient object detector via gather-and-distribute mechanism", 2023. [Online]. https://doi.org/10.48550/arXiv.2309.11331.

[58] C-Y. Wang, A. Bochkovskiy & H-Y. M. Liao, *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023. pp. 7464–7475. https://arxiv.org/abs/2207.02696.

[59] X. Zhu, S. Lyu, X. Wang & Q. Zhao, "TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios", 2021. [Online]. https://doi.org/10.48550/arXiv.2108.11539.