IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 24 May 2024, accepted 21 June 2024, date of publication 28 June 2024, date of current version 9 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3420730

RESEARCH ARTICLE

Enhanced Small Drone Detection Using Optimized YOLOv8 With Attention Mechanisms

FATIN NAJIHAH MUHAMAD ZAMRI[1], (Student Member, IEEE), TEDDY SURYA GUNAWAN[1], (Senior Member, IEEE), SITI HAJAR YUSOFF[1], (Member, IEEE), AHMAD A. ALZAHRANI[2], (Member, IEEE), ARIF BRAMANTORO[3], (Member, IEEE), AND MIRA KARTIWI[4], (Member, IEEE)

[1]Department of Electrical and Computer Engineering, International Islamic University Malaysia, Kuala Lumpur 53100, Malaysia
[2]Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[3]School of Computing and Informatics, Universiti Teknologi Brunei, Bandar Seri Begawan 1410, Brunei Darussalam
[4]Department of Information Systems, International Islamic University Malaysia, Kuala Lumpur 53100, Malaysia

Corresponding author: Teddy Surya Gunawan (tsgunawan@iium.edu.my)

This research work was funded by Institutional Fund Projects under grant no. (IFPIP: 1175-611-1443). The authors gratefully acknowledge technical and financial support from the Ministry of Education and King Abdulaziz University, Deanship of Scientific Research (DSR), Jeddah, Saudi Arabia.

ABSTRACT The increasing misuse of drones poses significant safety and security risks, including illegal transportation of prohibited goods, interference with manned aircraft, and threats to public safety. This has raised concerns about the increased use of unmanned aerial vehicles (UAVs) due to their small size. Addressing these concerns has sparked significant research into developing effective drone detection systems. Deep learning, especially YOLO, is known as a lightweight model that offers real-time detection capabilities. Attention mechanisms have proven effective in many studies for detecting objects. This research focused on optimizing the YOLOv8n-based model by incorporating the Attention Module into the neck and improving the detection head by adding a tiny detection head, making the model work efficiently in detecting objects of tiny size. To obtain the most effective model, multiple training sets have been experimented with involving different types of attention modules, such as the Convolutional Block Attention Module (CBAM), ResBlock CBAM, Global Attention Mechanism (GAM), and Efficient Channel Attention (ECA). Therefore, based on the results, YOLOv8n + ResCBAM + high-resolution detection head, called P2-YOLOv8n-ResCBAM significantly improves the mean Average Accuracy (mAP) from 90.3% to 92.6%. Although the increased model complexity reduced frames per second (fps) from 263 to 166, the detection speed remains suitable for real-time applications. The proposed model effectively distinguishes drones from birds and recognizes them at long distances, demonstrating its potential for enhancing aerial surveillance and security measures.

INDEX TERMS Artificial intelligence, deep learning, convolutional neural networks, small drone detection, YOLOv8, attention mechanism, visual object detection, real-time detection, aerial surveillance, autonomous systems.

I. INTRODUCTION

The rapid spread of drones has led to significant progress in multiple industries, such as delivery, agriculture, and surveillance. Nevertheless, the rise in drone utilization has resulted in heightened apprehensions regarding security and privacy, including the illicit conveyance of prohibited items, disruption of manned aircraft operations, and jeopardizing public well-being. To tackle these challenges, efficient drone detection systems must be used to accurately differentiate drones from other objects in real-time.

Drones come in many sizes, from small to large, and are often used in many industries such as monitoring, transportation, communication, and photography [1], [2], [3]. Hence, the proliferation of drones proves the advantages of improving our daily lives [4]. Nevertheless, while drones provide

The associate editor coordinating the review of this manuscript and approving it for publication was Shashikant Patil.

numerous advantages to society, their misuse can pose significant risks to safety, privacy, and security [5], [6]. Threats include privacy invasion, target attacks, breach of the No-Fly Zone (NFZ), and illegal transportation for smuggling, such as explosive things or drugs [3], [5], [7], [8], [9], [10], [11], [12]. The increase in the use of drones has sparked public concerns, and if this trend continues unabated, we may face a future where the sky is drowned by drones [3]. Therefore, implementing a drone detection system is an important step in reducing and dealing with this issue. This has increased researchers' awareness of the importance of developing an effective drone detection system. Various techniques have been introduced in the development of this system.

Compared to traditional methods, deep learning offers superior capabilities in automatically extracting and learning target features directly from data [13]. Deep learning is a branch of artificial intelligence (AI), using neural networks to process data. Through machine learning, these networks can be trained on vast data sets, allowing them to learn and recognize patterns autonomously. In essence, AI systems can emulate the functioning of the human brain by predicting outcomes based on observed patterns. Due to the advanced technology, its increasing popularity can be attributed to the accessibility of training data, advanced hardware, and computational resources [14], which have significantly expanded the use of deep learning techniques. As a result of these technological advances, the use of deep learning is increasing in various industries, especially in object detection [15]. Various techniques are available within the realm of Convolutional Neural Networks (CNNs). Object detection stands out as a superior choice over conventional radar and infrared in developing drone detection system [6], [16]. In general, object detection involves two main tasks: localization and classification. This task aims to determine the exact location of a target object in an image or video and identify the category of the object.

Object detection has two main approaches: one-stage and two-stage detectors. One-stage detectors, such as You Only Look Once (YOLO) [17] and Single Shot MultiBox Detector (SSD) [18], directly predict bounding boxes and class labels at the same stage. In contrast, two-stage detectors, such as Region-based Convolutional Neural Networks (R-CNN) [19], Fast R-CNN [20] and Faster R-CNN [21], involves two stages to identify the potential area and classify the targeted objects within this area. A study conducted by [22] aimed to determine the optimal model between Faster R-CNN, YOLO, and SSD to detect drones in various environments, focusing speed and accuracy. The results demonstrates that although SSD better in detection ability, Faster R-CNN and YOLO exhibit superior recognition abilities. However, according to [23], among various algorithms under object detection, YOLO offers a balanced combination of speed and accuracy, making it as a fast and reliable detection model. YOLO was designed expressly to overcome problems involving speed of inference while conserving competitive accuracy [6]. This is achieved by simultaneously performing bounding box determination and classification in the same stage. Researchers have continued to improve YOLO since its launch in 2015, leading to multiple versions. Figure 1 shows a timeline for the many variants of YOLO. Figure 2 gives an overview of its general functionality.

When choosing an efficient model to create a real-time drone detection system, YOLO-based models have performed well in many studies. As evidence, many studies have proven its effectiveness by demonstrating good results in terms of accuracy and speed. The YOLOv4 model was chosen for developing a drone detection system in a study by [9]. Using the transfer learning method, this model has outperformed Faster R-CNN in terms of mean Average Precision (mAP) after tuning the model. A study by [24] that has trained a fine-tuned YOLOv5 model, has achieved the highest precision, 97.4 % and at the same time surpassed YOLOv3, YOLOv4 and Mask R-CNN. In addition, this model also demonstrated a good performance in detecting small drones. To detect small objects and balance between accuracy and speed, [25] has modified the YOLOv8m-based model by adding a P2 Layer and Multi-Scale Image Fusion (MSIF). The highest frame per second (fps), 45.7 fps, shows that this method can detect drones very fast, even if they are small objects.

The capability of the attention mechanism to improve the model performance during training is proven by many studies, such as [6], [26], [27], [28], [29], [30], [31], and [32]. Selective hearing refers to the cognitive process in which individuals prioritize and concentrate on significant auditory stimuli amidst a distracting or noisy environment, rather than attempting to process all auditory information simultaneously. The attention module in deep learning operates on the same principle, enabling the model to concentrate on significant elements while disregarding irrelevant information selectively. Due to various elements like edges, textures, and background in the input image, the model faces challenges in accurately identifying the specific object in the image. By incorporating an attention module into the original architecture, the model can effectively concentrate on capturing the specific details of the targeted object, resulting in improved performance. Despite technological progress, the task of detecting small drones remains difficult because of their compact size and the intricate process of differentiating them from similar entities such as birds. The objective of this study is to improve the performance of the YOLOv8 model by integrating different attention modules, thereby increasing its precision in detecting small drones.

This paper proposes an effective method to develop drone detection that can distinguish between birds and drones and detect them even at long distances. The works done, including modifying the YOLOv8n model, are as follows:

1. Builds a new dataset by gathering drones and birds of a small size from numerous available datasets.

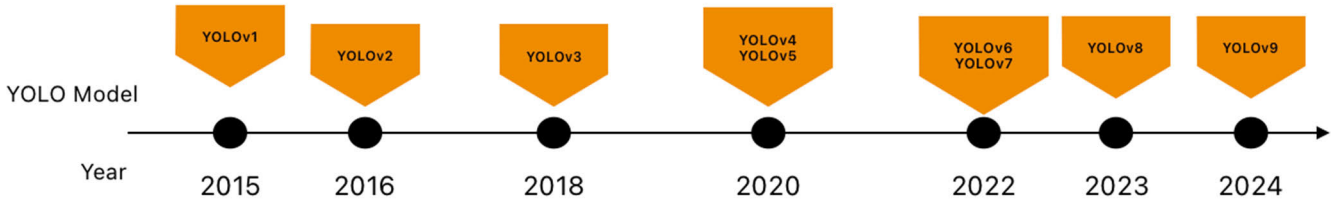

FIGURE 1. The evolution of the YOLO family.

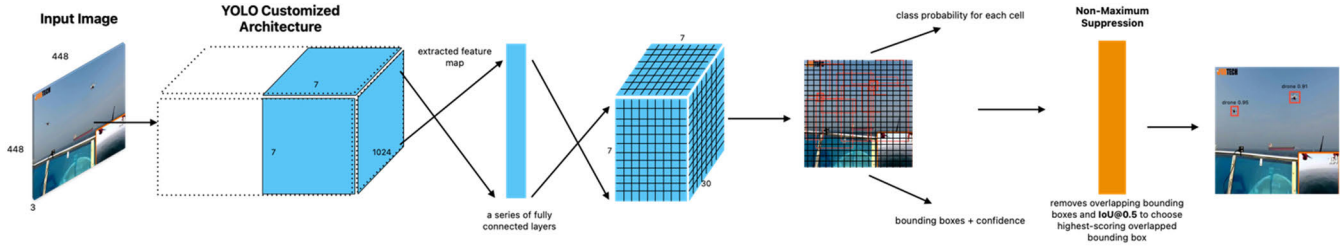

FIGURE 2. A simplified diagram of the YOLO mechanism.

2. Employs four different attention mechanisms to the neck part of the YOLOv8n model where it involves Convolutional Block Attention Module (CBAM), Global Attention Mechanism (GAM), Efficient Channel Attention (ECA), and ResBlock CBAM (ResCBAM) during training.
3. Adds high-resolution head to the head part, where it increases the model capability in detecting small targets.
4. Tunes the hyperparameters during training.
5. Ablation tests are carried out for every attention module, with and without a high-resolution head, utilizing different hyperparameter sets. The goal of these experiments is to find the best model, and ResCBAM + high-resolution head + tune hyperparameters achieve the best mAP.

This paper is organized as follows: Section II presents and explains the original architecture of YOLOv8 and the proposed version. Section III displays the training platform, including the software and hardware requirements and the training setup. This section also describes the dataset used and how the model is trained. Several experiments were conducted to compare each of the models with other YOLO versions. Next, section IV analyses the results based on the experiments. Section V presents the detection result based on the P2-YOLOv8n-ResCBAM model, and the last section concludes the overall work.

II. VISUAL DRONE DETECTION SYSTEM

Visual Drone Detection System is a system designed to identify and categorize objects of interest, including drones, by visual means. The presence of the drones is then recognized by extracting their characteristics from the captured image. The optimized drone detection model will be built on the foundation of the YOLOv8 model.

A. YOLOv8 MODEL

YOLOv8 offers five sizes, and in this research, the smallest model, YOLOv8n, is selected. Three main parts can be divided to represent YOLOv8 architecture: backbone, neck, and head. The backbone is responsible for extracting meaningful features from input images at various scales, the neck is known as multi-feature fusion, where all extracted features from different layers will be combined to get meaningful information, and the head works to make predictions. In the development of YOLOv8, three important elements can be highlighted based on [33]:

1. New convolution, C2f module is replaced C3 block as main building block in YOLOv8. To build C2f, the concept of ELAN (Efficient Layer Aggregation Network) [34] is used to improve resonance speed [35]. Unlike C3, which only uses the last bottleneck output, all bottleneck outputs in C2f will be combined. The idea is comparable to the ResNet Block [36].
2. Anchor-free detection is utilized instead of predicting using an offset bounding box known as an anchor box like other models. This means it will predict directly from the center of an object. This innovation has reduced the number of overlapped prediction boxes, which can accelerate the Non-Maximum Suppression (NMS).
3. Closing the mosaic augmentation for the last 10 epochs during training. During training, by applying mosaic augmentation, the model can learn objects in different locations as four images will be gathered in one image together. However, this augmentation can somehow decrease performance, and it is believed that turning

off this augmentation for the last 10 epochs can prevent the deterioration.

Several innovations in the development of YOLOv8 have contributed to an increase in accuracy and beat YOLOv5 and YOLOv7 when tested using dataset Microsoft COCO and Roboflow 100.

B. IMPROVED YOLOv8 MODEL

Two methods are proposed to improve the YOLOv8 model, and they involve the neck and head in the architecture. The architecture of the proposed model is displayed in Fig. 5.

1) IMPROVEMENT OF THE FEATURE FUSION MODULE

Since the dataset used has a lot of small objects with just a few pixels representing them, the chance of losing the feature information is high. This might result in false or miss detection. Therefore, this work proposed to improve the feature fusion, which will operate on the neck of the algorithm. The attention mechanism offers an effective way to improve the feature fusion model where it can improve object detection [37]. Then, many studies have produced good results when they integrate attention mechanism [26], [27], [30], [32], [38].

This research proposes to combine the ResBlock and Convolutional Block Attention Model (CBAM), which is called ResCBAM. Unlike the Squeeze-and-Excitation module (SENet) [39] that only uses Global Average Pooling (GAP) for channel-wise attention, CBAM is a lightweight model that involved the combination of operation channel attention (what to focus) and spatial attention modules (where to focus) [32]. Not only Global Average Pooling (GAP), Global Max Pooling (GMP) that is missing in SENet is also employed in CBAM to work together on the feature maps. Eq. (1) and (2) represents a mechanism of Channel Attention and Spatial Attention, as attached in Fig. 3(a) and 3(b), respectively, and an overview of CBAM is illustrated in Fig. 3(c).

$$F' = M_C(F) \otimes F \tag{1}$$

$$F'' = M_S(F') \otimes F' \tag{2}$$

Residual Block [36] or ResBlock is a main block in Residual Networks (ResNet) where it is one of the components in deep neural network architecture. ResBlock is designed to tackle vanishing gradient problems while training deep neural networks. Eq. (3) represents the mechanism of ResBlock as illustrated in Fig. 4(a).

$$F(x) + x = Output \tag{3}$$

Eq. (4) represents how ResCBAM works in Fig. 4(b). An overview of the ResCBAM mechanism is portrayed in Fig. 4(c).

$$Foutput = F + F'' \tag{4}$$

TABLE 1. Python library specifications.

No.	Open Libraries/Modules	Version
1	Ultralytics	8.0.20
2	PyTorch	2.1.0
3	matplotlib	3.7.2
4	numpy	1.22.2
5	Opencv-python	4.6.0
6	Pillow	10.0.1
7	Pyyaml	6.0.1
8	Requests	2.31.0
9	Scipy	1.11.3
10	Torch	2.1.0
11	Torchvision	0.16.0
12	Tqdm	4.66.1

2) IMPROVEMENT OF THE DETECTION HEAD

A new detection head is added at the head part of the YOLOv8 architecture to achieve a tiny target. Initially, there were three detection heads, and this proposed model has four detection heads and is known as the p2-YOLOv8 model. P2 there means the prediction will use p2 layers. The addition of the detection head has a resolution of 160×160 pixels, making it suitable to work with low-level features and high-resolution feature maps. Low-level features encompass edges, corners, colors, and textures of objects. High-resolution feature maps are packed with abundant information that can show the image at a very fine level and make it more sensitive to small targets. Having multiple detection heads can combine low-level spatial features and high-level semantic features to improve the feature information of each layer where this helps identify small items and increases feature information [40]. Fig. 5 shows how the additional head (light green box) is added to the proposed architecture.

III. EXPERIMENT SETUP AND DATA ANALYSIS

The experimental environment, training setup, and dataset used have been explained in this section.

A. TRAINING PLATFORM

1) HARDWARE AND SOFTWARE REQUIREMENTS

This research has been trained using Intel i9-14900K with 64GB memory. It allows multitasking and efficient handling of big and complex data. A large storage device, a 1.8 TB SSD drive, and a high-powered GPU, NVIDIA GeForce RTX 4090, are used to accelerate the training process. Several software specifications are required to perform this research, and the details are displayed in Table 1.

2) TRAINING SETUP

The training setup includes hyperparameters for model training are listed in Table 2.

B. DATASET CONSTRUCTION

This research aims to develop a drone detective system that is able to differentiate between drones and drone-like objects, such as birds, and track them even at long distances.

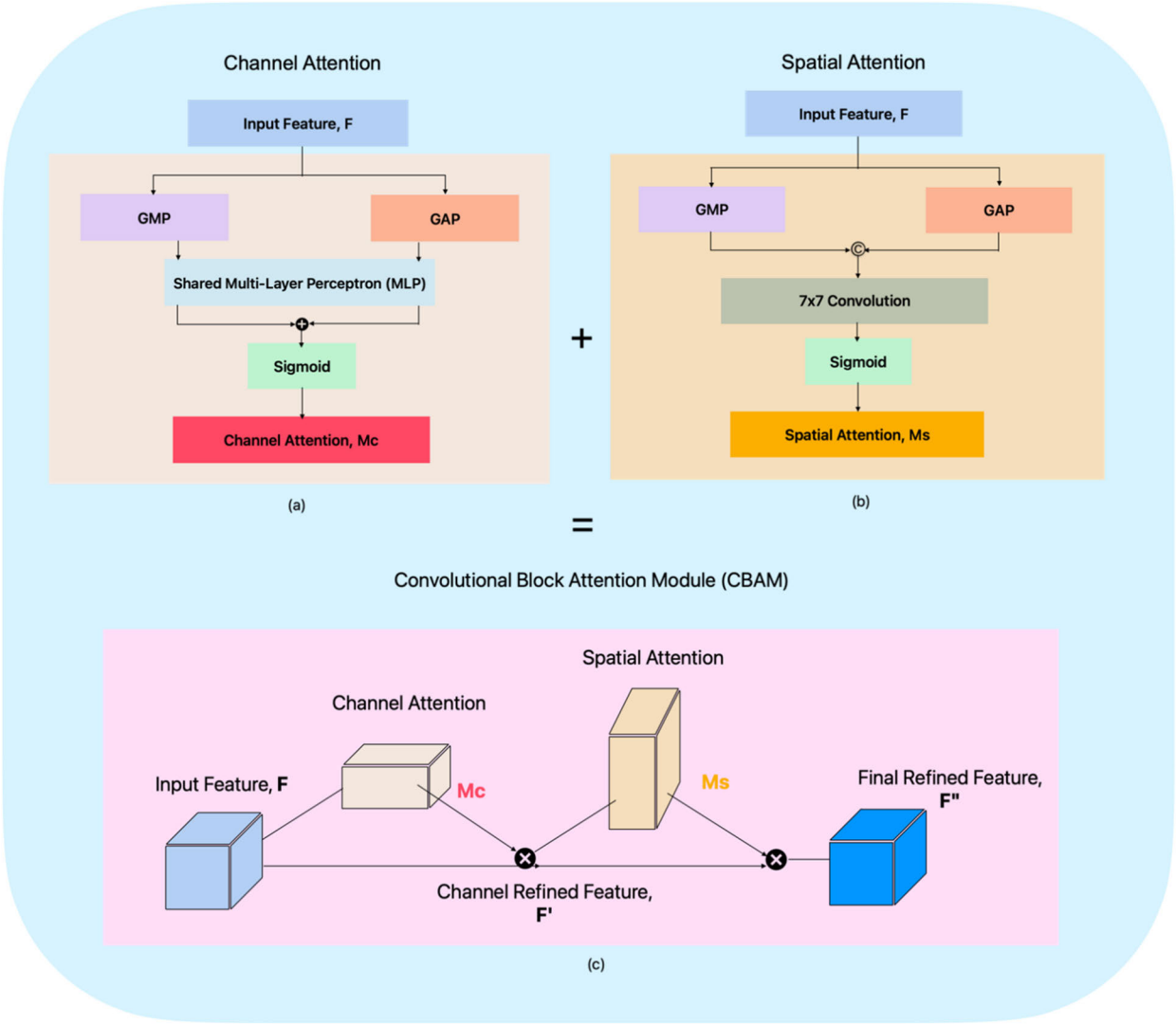

FIGURE 3. Overview of CBAM.

TABLE 2. Hyperparameters setup.

No.	Parameters	Value
1	Epochs	100
2	Warmup epochs	3
3	Warmup momentum	0.8
4	Batch size	16
5	Image size	640
6	Initial learning rate	0.01
7	Final learning rate	0.01
8	Patience	100
9	Optimizer	SGD
10	NMSIoU	0.7
11	Momentum	0.94
12	Mask ratio	4
13	Weight Decay	0.00012

Regardless of the actual distance of the object from the source, its small size in the frame, resulting in small pixels, can represent how far the drone or bird is from the camera. Therefore, it is crucial to provide a dataset that meets those criteria to allow the model to learn effectively. Hence, a new dataset, BirDrone [41] was prepared by collecting images of small drones, including multirotor types, such as quadcopters, hexacopters, and octocopters, as well as birds from different datasets [42], [43], [44], [45], [46], [47]. We have included images with multiple drones or birds in one image for model training. The YOLO framework itself is designed to detect multiple objects in one frame. The proof of detection will be shown in Section V. This dataset also includes different types of backgrounds and lighting. Fig. 6 and 7 show examples of raw images in our dataset. Before proceeding to training, the dataset needed to be annotated first, and it was done manually using Roboflow. The smallest annotation bounding box, which represents the size of the targeted object, is 7×14 pixels, and the largest one is 65×182 pixels. Then, the dataset went through several pre-processing methods, such as auto-orient, which can standardize overall orientation, improve

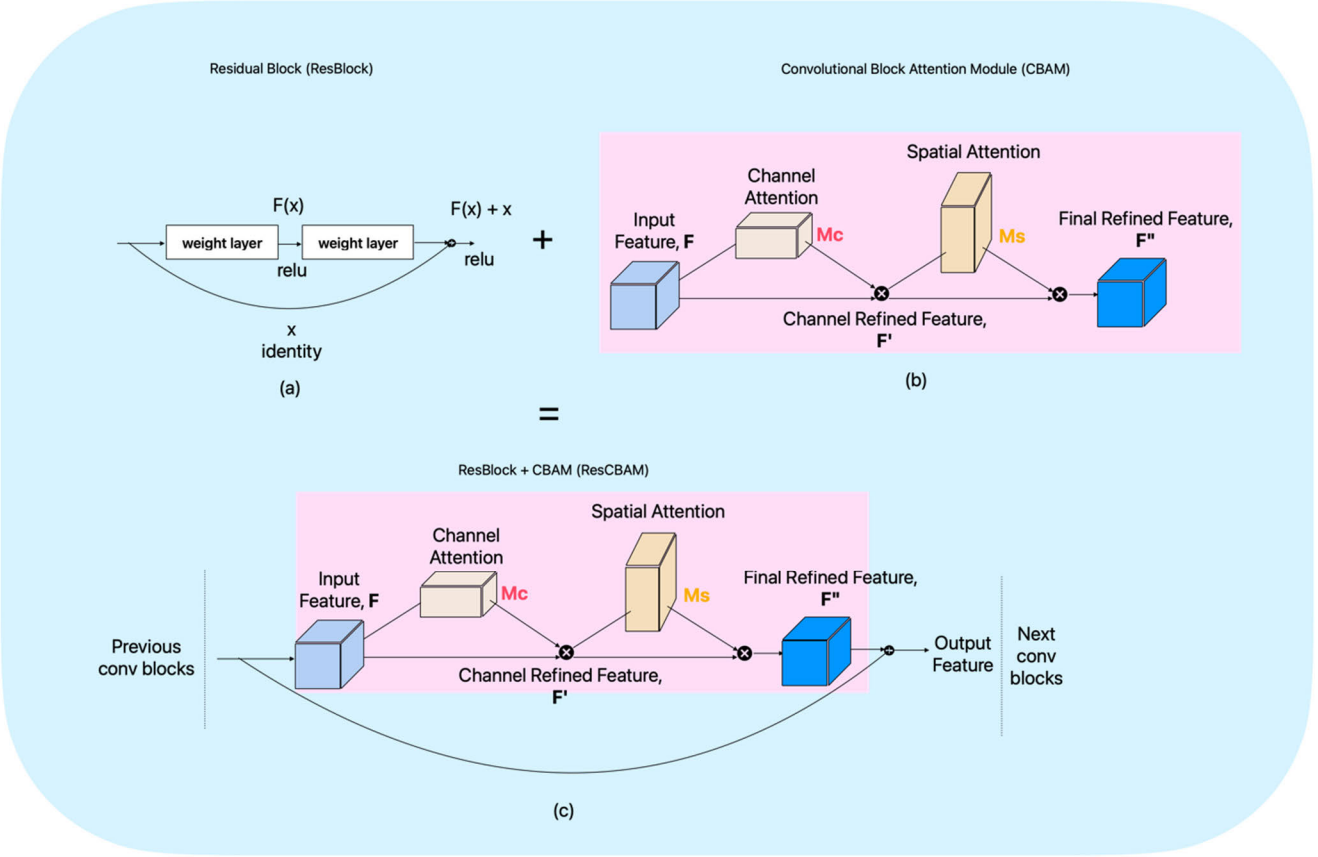

FIGURE 4. Overview of ResCBAM.

overall analysis, and be suitable for real-time applications. Lastly, auto-adjust contrast makes the details about birds and drones easier to see. Next, a data augmentation approach is used for the model to identify drones and birds in various scenarios. For each image, geometric transformations such as rotation and exposure are used. By displaying multiple perspectives, rotation can increase model robustness and help avoid overfitting. Applying exposure to the images gives more variability to the dataset regarding lighting and environment. After that, the total of images 2970 was divided into 80% for training and 20% for validation.

C. MODEL TRAINING AND RESULTS

1) ABLATION EXPERIMENTS

Several training sessions were carried out in this section using a designated dataset to verify various aspects of the proposed model. Initially, the effectiveness of the supplementary detection head and the P2-YOLOv8n model, specifically developed for identifying minuscule entities, was evaluated. Furthermore, an assessment was conducted to determine the effect of incorporating attention modules into the YOLOv8-based model. Furthermore, an analysis was conducted to evaluate the efficacy of optimizing hyperparameters. The comparison was evaluated based on precision, recall, and mAP, and the formula is shown in Table 3 with reference to Fig. 8.

TABLE 3. List of metrics used for evaluation.

Metrics	Definition	Calculation
Precision (P)	The accuracy of positive predictions made by the model among the identified positive samples.	$\frac{TP}{TP+FP}$
Recall (R)	The accuracy of positive predictions made by the model among all the positive samples.	$\frac{TP}{TP+FN}$
AP (Average Precision)	Summary of trade-off between precision and recall value per class	$\int_0^1 P(R)\,dR$
mAP (mean Average Precision)	Average of AP values across all classes	$\frac{1}{N}\sum_{i=1}^{N} AP_i$, N=total classes

Fig. 9 depicts a training sample utilized in these experiments, demonstrating the model's ability to manage diverse training scenarios and configurations effectively. Table 4 has demonstrated the training results of several YOLOv8n-based models with several attention modules but using default hyperparameters. Table 5 presents the training results of the YOLOv8n-based model but with a tuning hyperparameter for the optimizer, which SGD [48] is used, and the value

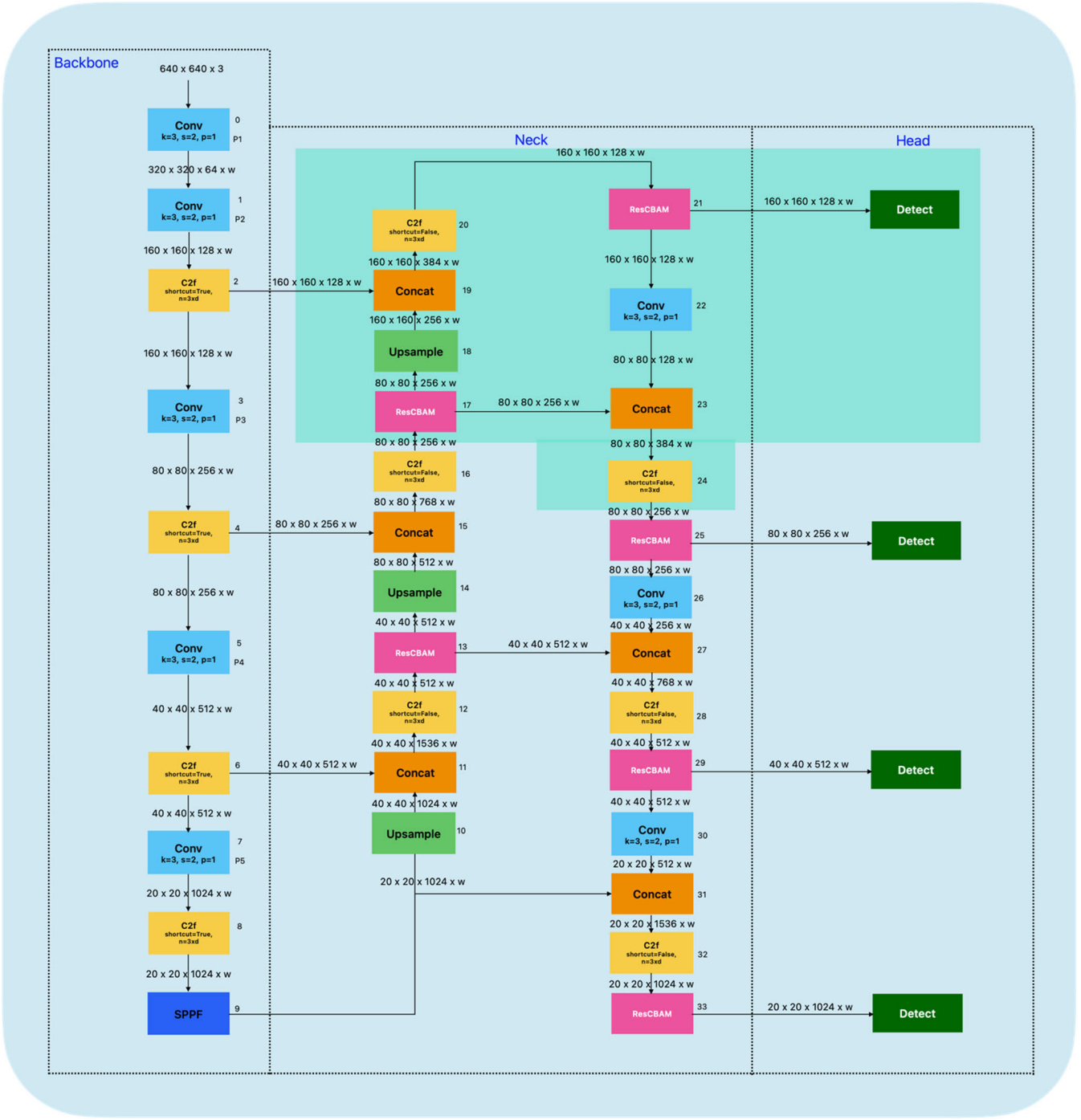

FIGURE 5. Architecture of the proposed model.

of weight decay was set to 0.00015 as a recommendation from Ultralytics [49]. Table 6 shows the training results when using 0.73375 for momentum value as a recommendation from Ultralytics [49], 0.00015 for weight decay and SGD for the optimizer. Table 7 displays the training results of several YOLOv8n-based models when tuning the hyperparameters, with the momentum value set to 0.94, a slight increment from the previous table. Also, weight decay was set lower than before, at 0.00012. Table 8 and Table 9 showcase the training result with tuning hyperparameters. The momentum value was set to 0.942, a slight increase from Table 5 and Table 6, which used 0.0005 for weight decay, and the rest of the hyperparameters were the same as Table 5. Finally, Table 10 shows the training results for the proposed model for all classes, including both drone and bird, drone only, and bird only.

TABLE 4. experimental results using different attention modules, different detection head with default hyperparameters, optimizer = auto, momentum = 0.937, weight decay = 0.0005.

Model	Precision	Recall	mAP	Parameters (M)	Training Time (H)
Yolov8n	0.901	0.893	0.915	3.006038	0.232
P2- Yolov8n	0.917	0.874	0.92	2.921304	0.565
Yolov8n-GAM	0.903	0.860	0.911	3.687318	0.368
P2-YOLOv8n-GAM	0.877	0.834	0.888	3.645176	0.656
Yolov8n-ECA	0.903	0.874	0.912	3.00605	0.312
P2-YOLOv8n-ECA	0.901	0.866	0.912	2.929514	0.616
Yolov8n-CBAM	0.903	0.874	0.915	3.019846	0.488
P2-YOLOv8n-CBAM	0.903	0.876	0.918	2.944244	0.350
Yolov8n-ResCBAM	0.900	0.868	0.912	4.239262	0.313
P2-YOLOv8n-ResCBAM	0.904	0.883	0.909	4.216836	0.369

TABLE 5. Experimental results using hyperparameters, optimizer = SGD, momentum = 0.937, weight decay = 0.00015.

Model	Precision	Recall	mAP	Parameters (M)	Training Time (H)
Yolov8n	0.917	0.882	0.916	3.006038	0.245
P2- Yolov8n	0.914	0.863	0.917	2.921304	0.488
P2-YOLOv8n-GAM	0.784	0.757	0.814	3.645176	0.658
P2-YOLOv8n-ECA	0.924	0.862	0.909	2.929514	0.544
P2-YOLOv8n-CBAM	0.150	0.867	0.911	2.944244	0.347
P2-YOLOv8n-ResCBAM	0.923	0.875	0.919	4.239262	0.422

TABLE 6. Experimental results using hyperparameters, optimizer = SGD, momentum = 0.73375, weight decay = 0.00015.

Model	Precision	Recall	mAP	Parameters (M)	Training Time (H)
Yolov8n	0.895	0.870	0.906	3.006038	0.279
P2- Yolov8n	0.904	0.855	0.908	2.921304	0.394
P2-YOLOv8n-GAM	0.833	0.780	0.835	3.645176	0.465
P2-YOLOv8n-ECA	0.905	0.867	0.907	2.929514	0.402
P2-YOLOv8n-CBAM	0.903	0.853	0.899	2.944244	0.347
P2-YOLOv8n-ResCBAM	0.893	0.829	0.899	4.225028	0.368

TABLE 7. Experimental results using hyperparameters, optimizer = SGD, momentum = 0.94, weight decay = 0.00012.

Model	Precision	Recall	mAP	Parameters (M)	Training Time (H)
Yolov8n	0.909	0.864	0.903	3.006038	0.253
P2- Yolov8n	0.917	0.868	0.909	2.921304	0.335
P2-YOLOv8n-GAM	0.867	0.827	0.873	3.645176	0.393
P2-YOLOv8n-ECA	0.887	0.852	0.904	2.929514	0.371
P2-YOLOv8n-CBAM	0.906	0.886	0.918	2.944244	0.347
P2-YOLOv8n-ResCBAM	0.915	0.879	0.926	4.216836	0.367

TABLE 8. Experimental results using hyperparameters, optimizer = SGD, momentum = 0.942, weight decay = 0.00012.

Model	Precision	Recall	mAP	Parameters (M)	Training Time (H)
Yolov8n	0.909	0.881	0.915	3.006038	0.229
P2- Yolov8n	0.908	0.863	0.912	2.921304	0.331
P2-YOLOv8n-GAM	0.853	0.852	0.883	3.645176	0.391
P2-YOLOv8n-ECA	0.903	0.878	0.908	2.929514	0.321
P2-YOLOv8n-CBAM	0.913	0.876	0.911	2.944244	0.349
P2-YOLOv8n-ResCBAM	0.924	0.883	0.926	4.216836	0.370

2) MODEL COMPARISON

Proposed model, P2-YOLOv8n-ResCBAM was compared with different models which include, YOLOv5n, YOLOv6, YOLOv7, YOLOv8n and YOLO-Drone [6]The comparison results are displayed in Table 11, which shows the trend of the model comparison results. To make it fair, all models were trained from scratch, which means no transfer learning was used, and the results are shown in Table 11.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. ABLATION EXPERIMENTS

Assessing the impact of different factors on the model's performance is essential in drone detection. The

TABLE 9. Experimental results using hyperparameters, optimizer = SGD, momentum = 0.94, weight decay = 0.0005.

Model	Precision	Recall	mAP	Parameters (M)	Training Time (H)
Yolov8n	0.911	0.896	0.921	3.006038	0.253
P2- Yolov8n	0.914	0.876	0.909	2.921304	0.335
P2-YOLOv8n-GAM	0.861	0.829	0.874	3.645176	0.395
P2-YOLOv8n-ECA	0.921	0.882	0.921	2.929514	0.319
P2-YOLOv8n-CBAM	0.905	0.873	0.912	2.944244	0.352
P2-YOLOv8n-ResCBAM	0.912	0.883	0.917	4.216836	0.371

FIGURE 6. Small bird.

FIGURE 7. Tiny drone.

	Actual Positive	Actual Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

FIGURE 8. Confusion matrix.

YOLOv8n-based model was utilized to conduct multiple training sessions and evaluate them based on various attention modules, detection head configurations, and sets of hyperparameters. The results of the models using default hyperparameters are shown in Table 4. Among these models, the P2-YOLOv8n model achieved the highest mean Average Precision (mAP), indicating the effectiveness of the additional detection head. Nevertheless, the inclusion of attention modules in the models did not enhance mean average precision (mAP) when compared to the base model using the default settings.

TABLE 10. Training result for the proposed model.

Class	Precision	Recall	mAP
All	0.915	0.879	0.926
Drone	**0.956**	**0.953**	**0.975**
Bird	0.875	0.806	0.876

Table 5 presents the outcomes obtained using a fresh set of hyperparameters. Among the models mentioned, only the P2-YOLOv8n-ResCBAM model demonstrated an increase in mAP (mean average precision). In contrast, the other models did not exhibit noteworthy enhancements compared to the results in Table 4. This implies that the choice of momentum and weight decay values is of utmost importance, especially when utilizing the SGD optimizer, as these parameters substantially influence the outcome of the optimization process.

The data in Table 6 shows that performance did not improve when using momentum values lower than the default, indicating that these values are incompatible with other hyperparameter values. By subsequently increasing the momentum value to 0.94 and setting the weight decay slightly lower than the default at 0.00012, a significant improvement in mean average precision (mAP) was observed for the P2-YOLOv8n-ResCBAM model, as indicated in Table 7. The mean average precision (mAP) experienced a boost of around 2.3% compared to the base model using its default configuration. The combination of increased momentum and reduced weight decay was effective, as supported by Tables 8 and 9. Despite achieving the highest mean Average Precision (mAP) among various hyperparameter sets, the P2-YOLOv8n-ECA module did not surpass the highest mAP achieved by the P2-YOLOv8n-ResCBAM model in Table 7.

In summary, the findings suggest that employing the SGD optimizer can improve the model's performance during training. However, it is crucial to carefully choose the values of momentum and weight decay to achieve the best possible outcomes.

For the attention module, a combination of P2-YOLOv8n with GAM did not perform well, whereas the map obtained in Table 4 was the lowest among other models. This indicates

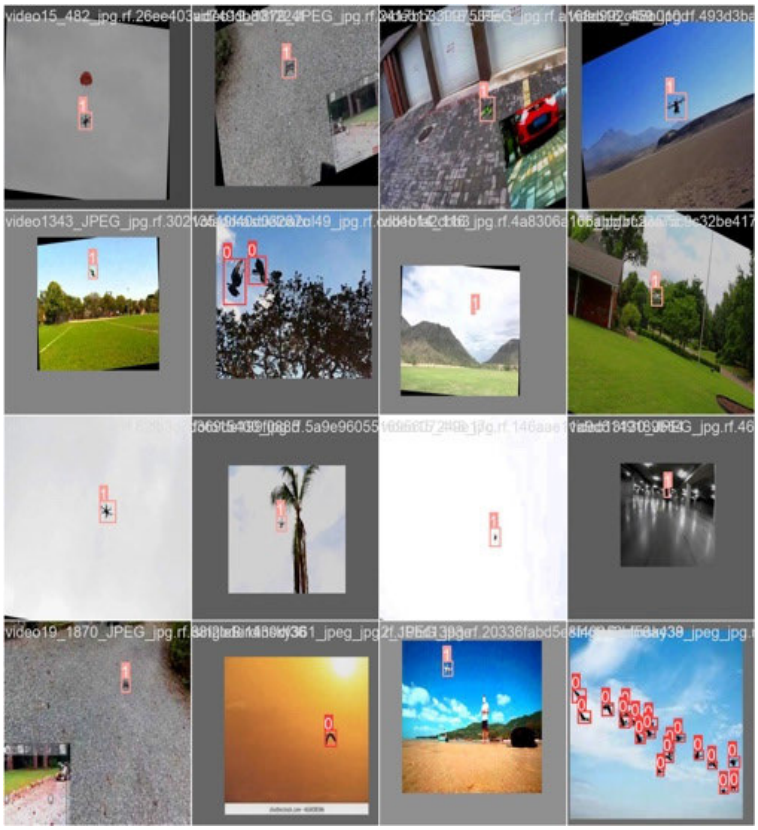

FIGURE 9. Training samples.

TABLE 11. Training result for model comparison.

Model	Precision	Recall	mAP	Training Time (H)
YOLOv5n	0.892	0.868	0.898	0.673
YOLOv6	0.883	0.866	0.901	0.343
YOLOv7	0.859	0.856	0.857	0.915
YOLOv9c	0.914	**0.892**	0.906	0.932
YOLOv8n	0.909	0.864	0.903	**0.253**
p2-YOLOv8	0.917	0.868	0.909	0.335
YOLO-Drone [6]	0.903	0.862	0.88	4.732
Proposed P2-YOLOv8n-ResCBAM	**0.915**	0.879	**0.926**	0.367

that the GAM may not be well-suited for combining with p2-YOLOv8 when training on the dataset with fine-tuned hyperparameters. As mentioned in [26], GAM has abandoned pooling to extract features for each channel. Therefore, it is believed that such a decision has influenced the results obtained. The paper [26] has also shown that GAM can outperform CBAM. However, due to the limited size of the data set used in this research, which is not as wide as they are, the results achieved cannot follow the proposed theory.

From Table 4, CBAM produced similar or lower mAP than ResCBAM. Only in Table 4, P2-YOLOv8n-CBAM outperformed P2-YOLOv8n-ResCBAM. However, the result was not the best mAP as P2-YOLOv8n can achieve higher mAP than both models. After we tuned the hyperparameters and the results shown in Table 7, P2-YOLOv8n-CBAM and P2-YOLOv8n-ResCBAM surpassed the mAP of P2-YOLOv8n and in these hyperparameter settings as well, P2-YOLOv8n-ResCBAM achieved the highest mAP compared to other models in Table 4 -IX. This shows that with the value chosen, the model can efficiently detect small objects with an extra detection head.

Overall, it can be seen that the combination of ResCBAM with an extra detection head showed improvements in Table 5, 7, and 8. This is believed to be due to the mechanisms applied in ResBlock, where shortcut connections are involved. The output of spatial attention is combined with the input feature map using element-wise addition, a method known as "identity mapping." Therefore, this residual connection can help in the gradient flow during training and can help prevent vanishing gradient problems, which leads to better detection

results. Lastly, Table 10 shows the proposed model's accuracy for each class. The overall bird detection result is lower than the overall model detection result in both classes. This is because BirDrone dataset contains significantly fewer bird images than drones.

B. MODEL COMPARISON

Several metrics can be used to evaluate the model's performance throughout the training and validation phases to determine its quality. As described in Table 11, three measurable metrics—precision, recall, and mean average precision—were utilized to assess trained models.

For precision, P2-YOLOv8n, YOLOv9c, and the proposed model, P2-YOLOv8n-ResCBAM, are the top three models that obtained high precision. High precision indicates the models' effectiveness in detecting objects and reducing false alarms, particularly when drones and bird-like objects are easily confused. Therefore, this suggests that these models are consistently accurate in predicting objects, especially in distinguishing between actual drones and birds and minimizing incorrect classifications.

None of the models achieved recall values as high as their precision values. However, the proposed model, YOLOv9c and P2-YOLOv8n, ranked highest in the recall, with values of 87.9%, 89.2%, and 86.8%, respectively. These three models captured the majority of positive samples in the dataset, demonstrating their high ability to identify the targeted objects correctly. This sensitivity in object detection is important for applications, as missing the targets can have serious consequences.

YOLOv9c and P2-YOLOv8n obtained 90.6% and 90.9% mean Average Precision (mAP), respectively, while the proposed model obtained the greatest mAP, at 92.6%. This suggests that the proposed model, which strikes a great balance between precision and recall, can detect objects while minimizing false positives and false negatives. The highest mAP demonstrates superior performance during training and validation, showing its robustness and consistency in predictions across various object detection scenarios. Therefore, it proves its suitability for real-world applications where accuracy is crucial.

Apart from those factors, the training time for the proposed model was not significantly longer compared to YOLOv9c. Since YOLOv9c has more parameters than the proposed model, the substantial difference in training time is believed to be due to the complexity of the YOLOv9c model. When comparing the proposed model with P2-YOLOv8n, despite the proposed model having more parameters due to the addition of attention modules, its training time was slightly increased compared to P2-YOLOv8n. For the same reason, it's possible due to the complexity of the proposed model that contributed to longer training time. However, the proposed model is well-optimized during training, as indicated by the small difference in training time despite the large gap in the number of parameters between these two models. This highlights the efficiency and effectiveness of the proposed model during the learning process.

Based on Table 11, the recall value of YOLO-Drone is among the lowest when compared to other models. This indicates that YOLO-Drone still struggles to identify and locate the targeted objects accurately. Consequently, it contributes to a notable difference in mAP, with a 2.3% difference from the base model, YOLOv8n, and a 4.6% difference from the proposed model, P2-YOLOv8n-ResCBAM. While YOLO-Drone showed promising results in a previous study [6], these favorable outcomes were not replicated in this study, possibly due to the use of different datasets. YOLO-Drone is trained using TIB-Net, which only has one class that consists of different types of UAVs. However, there are two classes in the dataset used for this research. So, the performance of each class affects the whole performance. Other than that, different training setups between this research and the YOLO-Drone paper may also contribute to the differences in model behavior.

V. MODEL VALIDATION AND DEPLOYMENT RESULT

A. MODEL VALIDATION

To address the overfitting issue, the dataset was partitioned into training and validation sets, with a ratio of 80:20. This division guarantees that the model is trained on a significant portion of the data while being validated on a distinct subset, facilitating improved generalization to unfamiliar data. In addition, the dataset underwent various data augmentation techniques to introduce variability and improve the model's robustness. The model was exposed to a diverse range of scenarios using techniques such as rotation, scaling, and color adjustments to reduce overfitting.

The validation process was performed using the specified validation set, which offered insights into the model's capacity to generalize. Figure 10 depicts a validation outcome that showcases the ability of the proposed model, P2-YOLOv8n-ResCBAM, to accurately distinguish between drones and birds, even when they are of small dimensions. Accurately distinguishing between similar objects is vital for the model's effectiveness in real-world situations.

B. MODEL DEPLOYMENT

To ensure successful implementation, it is imperative to thoroughly evaluate the model by subjecting it to real-life scenarios that accurately mimic the actual conditions. A video was obtained from YouTube with a resolution of 406×720 to evaluate the model's performance. Fig. 11 depicts a specific moment captured in the video footage that showcases the implementation of the P2-YOLOv8n-ResCBAM model. The model attained a frame rate of 166 frames per second, effectively differentiating between diminutive unmanned aerial vehicles and avian creatures. Nevertheless, it is crucial to acknowledge that the base YOLOv8n model surpassed the proposed model's inference speed, achieving an impressive 263 frames per second. The difference can be ascribed to

FIGURE 10. P2-YOLOv8n-ResCBAM model validation.

FIGURE 11. Scene from model deployment footage.

the heightened intricacy of the suggested model, which might necessitate additional processing time for each frame.

Fig. 12 illustrates a scenario where the bird, indicated by a red box, remains undetected when its body is hidden by tree branches. The bird's diminutive stature and the overlapping

FIGURE 12. Scene of bird detection failure.

of tree branches pose a challenge for detection, as the bird's body seamlessly merges with the background. This situation emphasizes a possible constraint of the model in identifying minuscule entities in crowded surroundings.

Data augmentation techniques, such as geometric transformations and exposure adjustments, enhance the model's robustness by simulating various real-world conditions, helping the model generalize better to unseen data. The validation results indicate that the model can accurately distinguish between drones and birds, which is essential for practical uses. While the P2-YOLOv8n-ResCBAM model achieved a frame rate of 166 fps, the base YOLOv8n model's higher frame rate of 263 fps indicates a trade-off between model complexity and processing speed. The proposed model's additional attention mechanisms and detection heads enhance accuracy but increase computational demand. The deployment of the model in real-world video footage demonstrates its practical applicability, although it faces challenges when dealing with objects occluded by other elements. This indicates the necessity for additional improvements and fine-tuning. The analysis highlights the significance of achieving a balance between the complexity of the model and its performance, emphasizing the necessity of optimization to uphold high accuracy while minimizing computational burden. Future work should optimize the model for real-time applications, develop sophisticated data augmentation strategies, and validate the model on diverse datasets to ensure generalizability across different environments and scenarios.

VI. CONCLUSION AND FUTURE WORKS

Detecting a drone with dynamic movements, small size, and even a shape similar to a bird is indeed challenging. Therefore, building an accurate model that can detect in real-time is crucial. However, speed and accuracy always have a trade-off between them.

By using YOLOv8n as a base model, this research proposed integrating an attention module and adding an extra high-resolution detection head. To support this proposed model to reduce false detection, the new dataset has been created to provide effective learning for the model to learn how to differentiate between drones and birds as well as recognize them even at long distances. Powerful hardware is utilized to ensure the inference speed is aligned with real-time detection. Fine-tuning the hyperparameters is also one of the methods used to optimize training performance, which can lead to better detection.

Based on training results, the P2-YOLOv8n-ResCBAM model has demonstrated improvement in mAP, which is from 90.3% to 92.6%, showing a 2.3% increment. However, due to the noticeable increment in model parameters, it can be noticed that the fps is decreased from base model during deployment, which is from 263 fps to 166 fps, but the fps achieved remains suitable in real-time detection. The addition of the attention module and detection head has certainly led to an increase in both the number of parameters and the complexity of the model. It is also believed to be why the inference speed in fps decreased. Apart from that, the model deployment result also portrayed a good result, where the model was able to differentiate between drones and birds even at long distances by using video and images as input. However, the model struggles to detect objects when they overlap with other objects in the same frame.

While the proposed model shows promising results, this research needs to highlight several aspects. Firstly, considering the model's complexity, the inference speed is fast due to the powerful hardware used during training and deployment. Therefore, the performance may differ depending on the type of hardware used, especially if low-end hardware is used. While the model may not be ideal for implementation on low-end hardware, the complexity of the model could be justified by its capability to detect a wide range of targets, especially tiny ones, and differentiate between drones and birds. These capabilities offer significant benefits regarding accuracy, reliability, and applicability in various real-world scenarios. Therefore, the decision to use the proposed model should be based on weighing these potential benefits against the hardware requirements. Next, they have built advanced drone technology that mimics birds' looks and behavior, such as the eagles. Although the proposed model is being trained to differentiate between drones and birds, the model may not detect this type of technology as the detection is only based on visuals. Therefore, this research could further explore the application as there is room for further improvement, such as model size reduction, and optimize this method to make it suitable for industry needs.

ACKNOWLEDGMENT

This research work was funded by Institutional Fund Projects under grant no. (IFPIP: 1175-611-1443). The authors gratefully acknowledge technical and financial support from the Ministry of Education and King Abdulaziz University, Deanship of Scientific Research (DSR), Jeddah, Saudi Arabia. The authors would like to sincerely thank the Kulliyyah of Engineering at the International Islamic University Malaysia for awarding the first author the IIUM Engineering Merit Scholarship and for providing a supportive and conducive environment for their research activities.

REFERENCES

[1] E. Yanmaz, S. Yahyanejad, B. Rinner, H. Hellwagner, and C. Bettstetter, "Drone networks: Communications, coordination, and sensing," *Ad Hoc Netw.*, vol. 68, pp. 1–15, Jan. 2018, doi: 10.1016/j.adhoc.2017.09.001.

[2] B. Aydin and S. Singha, "Drone detection using YOLOv5," *Eng*, vol. 4, no. 1, pp. 416–433, Feb. 2023, doi: 10.3390/eng4010025.

[3] M. A. Khan, H. Menouar, A. Eldeeb, A. Abu-Dayya, and F. D. Salim, "On the detection of unauthorized drones—Techniques and future perspectives: A review," *IEEE Sensors J.*, vol. 22, no. 12, pp. 11439–11455, Jun. 2022, doi: 10.1109/JSEN.2022.3171293.

[4] Y. Sun, X. Zhi, H. Han, S. Jiang, T. Shi, J. Gong, and W. Zhang, "Enhancing UAV detection in surveillance camera videos through spatiotemporal information and optical flow," *Sensors*, vol. 23, no. 13, p. 6037, Jun. 2023, doi: 10.3390/s23136037.

[5] A. Coluccia, A. Fascista, A. Schumann, L. Sommer, A. Dimou, D. Zarpalas, M. Méndez, D. de la Iglesia, I. González, J.-P. Mercier, G. Gagné, A. Mitra, and S. Rajashekar, "Drone vs. Bird detection: Deep learning algorithms and results from a grand challenge," *Sensors*, vol. 21, no. 8, p. 2824, Apr. 2021, doi: 10.3390/s21082824.

[6] X. Zhai, Z. Huang, T. Li, H. Liu, and S. Wang, "YOLO-Drone: An optimized YOLOv8 network for tiny UAV object detection," *Electronics*, vol. 12, no. 17, p. 3664, Aug. 2023, doi: 10.3390/electronics12173664.

[7] A. Coluccia, A. Fascista, A. Schumann, L. Sommer, A. Dimou, D. Zarpalas, F. C. Akyon, O. Eryuksel, K. A. Ozfuttu, S. O. Altinuc, F. Dadboud, V. Patel, V. Mehta, M. Bolic, and I. Mantegh, "Drone-vs-bird detection challenge at IEEE AVSS2021," in *Proc. 17th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Nov. 2021, pp. 1–8, doi: 10.1109/AVSS52988.2021.9663844.

[8] H. Liu, F. Qu, Y. Liu, W. Zhao, and Y. Chen, "A drone detection with aircraft classification based on a camera array," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 322, Mar. 2018, Art. no. 052005.

[9] S. Singha and B. Aydin, "Automated drone detection using YOLOv4," *Drones*, vol. 5, no. 3, p. 95, Sep. 2021, doi: 10.3390/drones5030095.

[10] G. E. M. Abro, S. A. B. M. Zulkifli, R. J. Masood, V. S. Asirvadam, and A. Laouti, "Comprehensive review of UAV detection, security, and communication advancements to prevent threats," *Drones*, vol. 6, no. 10, p. 284, 2022.

[11] R. J. Bunker and J. P. Sullivan. (2021). *Additional Weaponized Consumer Drone Incidents in Michoacan and Puebla*. [Online]. Available: https://smallwarsjournal.com/jrnl/art/mexican-cartel-tactical-note-50-additional-weaponized-consumer-drone-incidents-michoacan

[12] Z. Abdullah, "Unauthorised drones around Changi Airport delay 37 flights, affect operations of one runway," The Straits Times, SPH Media Limited, Singapore, 2019. Accessed: May 9, 2024. [Online]. Available: https://www.straitstimes.com/singapore/transport/37-flights-delayed-one-runway-closed-for-10-hours-due-to-unauthorised-drones

[13] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261–318, Feb. 2020, doi: 10.1007/s11263-019-01247-4.

[14] R. Elshawi, A. Wahab, A. Barnawi, and S. Sakr, "DLBench: A comprehensive experimental evaluation of deep learning frameworks," *Cluster Comput.*, vol. 24, no. 3, pp. 2017–2038, Sep. 2021, doi: 10.1007/s10586-021-03240-4.

[15] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.

[16] H. J. A. Dawasari, M. Bilal, M. Moinuddin, K. Arshad, and K. Assaleh, "DeepVision: Enhanced drone detection and recognition in visible imagery through deep learning networks," *Sensors*, vol. 23, no. 21, p. 8711, Oct. 2023, doi: 10.3390/s23218711.
[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.
[18] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot MultiBox detector," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2016, pp. 21–37.
[19] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.
[20] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448, doi: 10.1109/ICCV.2015.169.
[21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.
[22] S. M. Alkentar, B. Alsahwa, A. Assalem, and D. Karakolla, "Practical comparation of the accuracy and speed of YOLO, SSD and faster RCNN for drone detection," *J. Eng.*, vol. 27, no. 8, pp. 19–31, Aug. 2021, doi: 10.31026/j.eng.2021.08.02.
[23] J. Terven, D. M. Cordova-Esparza, and J. A. Romero-Gonzalez, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," in *Machine Learning and Knowledge Extraction*, vol. 5, 2023.
[24] N. Al-Qubaydhi, A. Alenezi, T. Alanazi, A. Senyor, N. Alanezi, B. Alotaibi, M. Alotaibi, A. Razaque, A. A. Abdelhamid, and A. Alotaibi, "Detection of unauthorized unmanned aerial vehicles using YOLOv5 and transfer learning," *Electronics*, vol. 11, no. 17, p. 2669, Aug. 2022, doi: 10.3390/electronics11172669.
[25] J.-H. Kim, N. Kim, and C. S. Won, "High-speed drone detection based on YOLO-v8," in *Proc. IEEE Int. Conf. Acoust.*, Jun. 2023, pp. 1–2.
[26] Y. Liu, Z. Shao, and N. Hoffmann, "Global attention mechanism: Retain information to enhance channel-spatial interactions," 2021, *arXiv:2112.05561*.
[27] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539, doi: 10.1109/CVPR42600.2020.01155.
[28] M. Kim, J. Jeong, and S. Kim, "ECAP-YOLO: Efficient channel attention pyramid YOLO for small object detection in aerial image," *Remote Sens.*, vol. 13, no. 23, p. 4851, Nov. 2021, doi: 10.3390/RS13234851.
[29] T. Wu and Y. Dong, "YOLO-SE: Improved YOLOv8 for remote sensing object detection and recognition," *Appl. Sci.*, vol. 13, no. 24, p. 12977, Dec. 2023, doi: 10.3390/APP132412977.
[30] C. Shen, C. Ma, and W. Gao, "Multiple attention mechanism enhanced YOLOX for remote sensing object detection," *Sensors*, vol. 23, no. 3, p. 1261, Jan. 2023, doi: 10.3390/S23031261.
[31] H. Sun, J. Yang, J. Shen, D. Liang, L. Ning-Zhong, and H. Zhou, "TIB-net: Drone detection network with tiny iterative backbone," *IEEE Access*, vol. 8, pp. 130697–130707, 2020, doi: 10.1109/ACCESS.2020.3009518.
[32] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2018, pp. 3–19.
[33] J. Solawetz and Francesco. (2023). *What is YOLOv8? The Ultimate Guide*. [Online]. Available: https://blog.roboflow.com/whats-new-in-yolov8/
[34] C. Y. Wang, H. Y. M. Liao, and I. H. Yeh, "Designing network design strategies through gradient path analysis," *J. Inf. Sci. Eng.*, vol. 39, no. 2, pp. 1–12, 2023.
[35] Q. Luo, C. Wu, G. Wu, and W. Li, "A small target strawberry recognition method based on improved YOLOv8n model," *IEEE Access*, vol. 12, pp. 14987–14995, 2024, doi: 10.1109/ACCESS.2024.3356869.
[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[37] C. Wang and H. Wang, "Cascaded feature fusion with multi-level self-attention mechanism for object detection," *Pattern Recognit.*, vol. 138, Jun. 2023, Art. no. 109377, doi: 10.1016/J.PATCOG.2023.109377.
[38] Z. Zhang, "Drone-YOLO: An efficient neural network method for target detection in drone images," *Drones*, vol. 7, no. 8, p. 526, Aug. 2023, doi: 10.3390/DRONES7080526.
[39] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020, doi: 10.1109/TPAMI.2019.2913372.
[40] X. Wang, D. Zhu, and Y. Yan, "Towards efficient detection for small objects via attention-guided detection network and data augmentation," *Sensors*, vol. 22, no. 19, p. 7663, Oct. 2022, doi: 10.3390/s22197663.
[41] F. N. M. Zamri and T. S. Gunawan, "BirDrone," *IEEE DataPort*, May 2024, doi: 10.21227/ettb-0w28. [Online]. Available: https://ieee-dataport.org/documents/birdrone
[42] H. Walia. *Bird vs Drone New*. Accessed: May 9, 2024. [Online]. Available: https://www.kaggle.com/datasets/harshwalia/bird-vs-drone-new
[43] E. Zhang. *Drone Object Detection*. Accessed: May 9, 2024. [Online]. Available: https://www.kaggle.com/datasets/sshikamaru/drone-yolo-detection
[44] J. Minaya. *Birds Detector Dataset*. Accessed: May 9, 2024. [Online]. Available: https://universe.roboflow.com/jess-minaya/birds-detector-tis9s
[45] *Drone. Drone Dataset*. Accessed: May 9, 2024. [Online]. Available: https://universe.roboflow.com/drone-rwsrk/drone-cmxwz
[46] S. S. T. D. Detection. *Drone_Detect Dataset*. Accessed: May 9, 2024. [Online]. Available: https://universe.roboflow.com/sst-drone-detection/drone_detect-kby2p
[47] Shskjcxds. *Drone Detection Dataset*. Accessed: May 9, 2024. [Online]. Available: https://universe.roboflow.com/shskjcxds/drone-detection-umvuu
[48] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*.
[49] J. Glenn. *Ultralytics YOLO Hyperparameter Tuning Guide*. Accessed: May 9, 2024. [Online]. Available: https://docs.ultralytics.com/guides/hyperparameter-tuning/

FATIN NAJIHAH MUHAMAD ZAMRI (Student Member, IEEE) received the bachelor's degree in electronics-computer and information from International Islamic University Malaysia (IIUM), Kuala Lumpur, in 2023, where she is currently pursuing the Master of Science degree in computer and information engineering. Her research interests include artificial intelligence, deep learning, and object detection.

TEDDY SURYA GUNAWAN (Senior Member, IEEE) received the B.Eng. degree (cum laude) in electrical engineering from the Institut Teknologi Bandung (ITB), Indonesia, in 1998, the M.Eng. degree from the School of Computer Engineering, Nanyang Technological University, Singapore, in 2001, and the Ph.D. degree from the School of Electrical Engineering and Telecommunications, The University of New South Wales, Australia, in 2007. His research interests include speech and audio processing, biomedical signal processing and instrumentation, image and video processing, and parallel computing. He was awarded the Best Researcher Award from IIUM, in 2018. He was the Chairperson of the IEEE Instrumentation and Measurement Society–Malaysia Section, in 2013, 2014, and 2020, a Professor, since 2019, the Head of the Department of Electrical and Computer Engineering, from 2015 to 2016, the Head of Programme Accreditation, and the Quality Assurance for Faculty of Engineering, International Islamic University Malaysia, from 2017 to 2018. He has been a Chartered Engineer (IET, U.K.) and Insinyur Profesional Madya (PII, Indonesia), since 2016, a registered ASEAN Engineer, since 2018, and an ASEAN Chartered Professional Engineer, since 2020.

SITI HAJAR YUSOFF (Member, IEEE) received the M.Eng. degree (Hons.) in electrical engineering and the Ph.D. degree in electrical engineering from the University of Nottingham, U.K., in 2009 and 2014, respectively. In 2015, she became an Assistant Professor with the Department of Electrical and Computer Engineering, International Islamic University Malaysia, Gombak. She is currently a Lecturer in control of power electronics and electrical power systems. Her research interests include controlling power converters and drives, matrix and multilevel converters, the IoT, smart meter, wireless power transfer for dynamic charging in electric vehicles (EVs), and renewable energy.

AHMAD A. ALZAHRANI (Member, IEEE) received the bachelor's degree (Hons.) in computer science from King Abdulaziz University, in 2003, and the master's degree in information technology and the Ph.D. degree in computer science from La Trobe University, in 2008 and 2014, respectively. He is currently an Associate Professor with the Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University. Since 2015, he has been a member of a House of Experience that provides consultation and IT technical solution. He provides several consultation services for both government and private sector. He has been a General Supervisor of Educational Affairs with the Faculty of Computing and Information Technology, since 2015. From August 2016 to September 2018, he was the Head of the Information System Department, Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Saudi Arabia. He took cloud and virtualization training from Global Knowledge Training in Washington, DC, USA, in 2014. In 2004, he took a CCNA course from Saudi Computer Society in Riyadh, Saudi Arabia. His research interests include pervasive computing, context-aware mobile applications, augmented reality, mobile computing, crowdsensing, crowdsourcing, and human–computer interaction. He is the author of three journal articles and three conference proceedings.

ARIF BRAMANTORO (Member, IEEE) received the bachelor's degree from the Department of Informatics, Institute Technology of Bandung, Indonesia, in 2001, the master's degree from the Faculty of Information Technology, Monash University, Melbourne, Australia, in 2006, and the Ph.D. degree from the Department of Social Informatics, Kyoto University, Japan, in 2011. He is currently a Senior Assistant Professor with the School of Computing and Informatics, Universiti Teknologi Brunei, Brunei Darussalam. Previously, he was an Associate Professor with the Information Systems Department, Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Saudi Arabia. From 2014 to 2016, he was an Assistant Professor with the Information Systems Department, College of Computer Sciences and Information, Al-Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia. He was an Expert Researcher with the Information Services Platform Laboratory, National Institute of Information and Communication Technology, Japan, from 2011 to 2012. He is the author of more than 50 articles. His research interests include service systems, business process workflow, and business intelligence. He was a recipient of the Research Excellence Award, in 2016, from the Deanship of Scientific Research, Al-Imam University, Saudi Arabia; and the Best Paper Award from the IEEE International Conference on Cloud Computing, in 2015.

MIRA KARTIWI (Member, IEEE) is currently a Professor with the Department of Information Systems, Kulliyyah of Information and Communication Technology, and also the Director of Professional Development with International Islamic University Malaysia (IIUM). She was one of the recipients of Australia Postgraduate Award (APA), in 2004. For her achievement in research, she was awarded the Higher Degree Research Award for Excellence in 2007. She has also been appointed as an editorial board member in local and international journals to acknowledge her expertise. She is also an experienced consultant specializing in the health, financial, and manufacturing sectors. Her areas of expertise include health informatics, e-commerce, data mining, information systems strategy, business process improvement, product development, marketing, delivery strategy, workshop facilitation, training, and communications.