Enhancing Small Target Detection in Aerial Imagery with Bi-PAN-FPN and EDMOA-Optimized YOLOv8-s

[1]Sarojini Yarramsetti, [2]Rasi D, [3]Srikanth Mylapalli P, [4]Pushpa S, [5]Gopala Krishna P and [6]Gowri G
[1]Nehru Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India.
[2]Department of Computer Science and Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, Tamil Nadu, India.
[3]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.
[4]Department of Computer Science and Engineering, St.Peter's Institute of Higher Education and Research, Avadi, Chennai, Tamil Nadu, India.
[5]Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, Bachupally, Hyderabad, Telangana, India.
[6]Department of Artificial Intelligence and Data Science, Karpagam Institute of Technology, Coimbatore, Tamil Nadu, India.
[1]nietdrsarojiniyarramsetti@nehrucolleges.com, [2]priyamudanrasi@gmail.com, [3]msrikanth@klunivesity.in, [4]pushpasangar96@gmail.com, [5]gopalakrishna@griet.ac.in, [6]ggowri762@gmail.com

Correspondence should be addressed to Sarojini Yarramsetti : nietdrsarojiniyarramsetti@nehrucolleges.com

Article Info
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi : https://doi.org/10.53759/7669/jmc202404084
Received 12 March 2024; Revised from 22 April 2024; Accepted 18 July 2024.
Available online 05 October 2024.

Abstract – Across the globe, people are working to build "smart cities" that will employ technology to make people's lives better and safer. Installing cameras at strategic spots across the city to monitor public spaces besides provide real-time footage to law enforcement besides other local authorities is a crucial part of smart city infrastructure, which includes video surveillance. A more effective answer is provided by deep learning algorithms, however research in this area still faces significant problems from changes in target size, form change, occlusion, and illumination circumstances as seen from the drone's perspective. In light of the aforementioned issues, this study presents a highly effective and resilient approach for aerial picture identification. To begin, the concept of Bi-PAN-FPN is presented to enhance the neck component of YOLOv8-s, taking into consideration the prevalent issue of small targets being easily misdetected or ignored in aerial photos. We achieve a more advanced and thorough feature fusion procedure much as feasible by completely considering and reusing multiscale features. To further reduce the amount of parameters in the model and prevent info loss during long-distance feature transfer, the benchmark model's backbone incorporates the GhostblockV2 structure in lieu of a portion of the C2f module. With the help of the Enhanced Dwarf Mongoose Optimization Algorithm (EDMOA), the suggested model's hyper-parameters are optimised. Lastly, a dynamic nonmonotonic focusing mechanism is employed in conjunction with WiseIoU loss as bounding box regression loss. The detector accounts for varying anchor box quality by utilizing "outlier" evaluations, thus improving the complete presentation of the detection task.

Keywords – Video Surveillance, Enhanced Dwarf Mongoose Optimization Algorithm, GhostblockV2, Smart Cities, YOLOv8.

I. INTRODUCTION

Drones are airplanes that do not have a pilot on board. They go by the acronym "UAV" for "unmanned aerial vehicle" [1]. They run without the need for a human pilot. Their degrees of independence vary. Drones can avoid collisions by using in-built schemes that include Light Detection detectors and other sensitive sensors to determine their location and the airspace around them [2]. Autopilot and enhanced autonomy are terms that describe this. Additionally, it is capable of what is called remotely piloted autonomy, which allows a human to control its motions. The size and design of a drone might vary according to its intended use [3]. They can go different distances and at different speeds. A drone with a very short range can cover a distance of up to three miles. Hobbyists frequently utilize these. Drones designed for shorter

flights can reach distances of up to 30 miles [4]. A drone's maximum altitude can be more than three thousand feet, or four hundred kilometers.

From delivering groceries to rescuing individuals buried by avalanches or rubble to gathering data for scientists, their uses are practically endless. Dangerous tasks often involve these drones [5]. The aerospace industry and the military were the initial users of these tools. Common safety applications for drones included checking if an area was safe for troops to move into or if anyone was buried under rubble entered. Drones equipped with weapons, such bombs, might even replace humans in some situations [6, 7]. Drones have gained popularity and are making their way into civilian hands as a result of the efficiency and effective safety they provide. Civilian drones have several potential applications, such as delivering groceries to a customer's door or taking aerial photographs of any desired location [8].

In addition to human guards, drones may survey sites, film resources from above, secure perimeters, and deter intruders. Worksite patrols can also make use of them. Because they are integrated with AI technology, they can deliver round-the-clock real-time data streaming [9]. The surveillance drone, on the other hand, heralds new uses for "tomorrow's security" thanks to its remarkable mobility and observation capabilities, which are enhanced by cutting-edge tech [10]. In complicated or high-stakes situations, whether involving public or private security, it is a vital air asset due to its wide variety of applications and great operational value [11]. Operations and maintenance capabilities, application intelligence, besides professional adoption are thus more important than technological aspects in determining a drone safety system's success.

Through the use of data-driven solutions and cutting-edge information technology, cities around the globe are becoming into "smart" communities [12]. A smart city is an urban area that has invested in its infrastructure to the point where it can maximize urban services like transportation, energy distribution, communications, and public safety while also improving the quality of life for its residents and making the city more sustainable [13]. Public safety, crime prevention, traffic management, and environmental monitoring are just a few of the smart city applications that can be detected and identified by a video surveillance system (VSS). The installation of video cameras in public spaces allows for the round-the-clock surveillance of citizen behavior for the purpose of public safety. These cameras can detect suspicious movements within crowds, as well as criminal acts like vandalism, theft, and public disturbances [14]. This is accomplished by utilizing motion-based approaches, such as frame distinguishing, optical action recognition. These methods are applied in real-time on a nearby server. Further, encoder and classifier tasks can be handled by deep learning algorithms like CNN, LSTM, RNN, and DNN, which allow for the detection and classification of restricted human movements in a given environment [15]. The safety of the public depends on the regulation of cold weapons in public spaces. Recently, fine-grained algorithms based on deep learning have shown promise for handheld cold weapons because to their resemblance to mobile phones, wallets, and cards. Furthermore, computer vision algorithms based on color, form, and texture are mainly responsible for the development of urban greenery and building deployments [16]. As an example, in order to keep an eye on plant diseases, color-based classification methods like SVM and k-nearest neighbor (kNN) are utilized. On the other hand, city buildings are analyzed using texture-based approaches like Gabor filtering besides local binary patterns (LBP) histograms.

The primary benefits of this study are as follows.

- By focusing on large-scale feature maps besides proposing the concept of Bi-PAN-FPN, this study enhances the model's detection capability for small targets while simultaneously increasing the likelihood and duration of fusion, leading to improved feature engineering.
- One typical issue with UAV photos is that small targets are often overlooked or misdetected. This is now resolved. Enhances the model's loss function and backbone network.
- Improve the model's generalization performance from feature diversity, long-distance feature info collection, and avoidance of geometric factors by integrating the Ghostblock unit and regression loss.
- Reduces the model's parameter count without sacrificing accuracy. The problems of long-range information loss and anchor prediction balance are both resolved by this.
- We show that our proposed model is better by comparing it to other models already in use. To further demonstrate why this strategy is preferable, we compare the interpretability of three top-notch models.

The rest of the paper is prearranged as follows: Section 2 mentions the related works; Section 3 presents the projected procedure; Section 4 discuss the results examination and finally, the conclusion of the research work is given in Section 5.

II. RELATED WORKS

The innovative firefighting drone introduced by Jahan et al. [17] can put out fires while simultaneously monitoring gas concentration, fire location, and giving real-time images, all with the goal of reducing dangers to firefighters. The suggested intelligent quadcopter processes data using the Pixhawk Telemetry system and uses the Pixhawk PX4 microprocessor for accurate control. Several gas sensors, a servo motor to put out the fire, a camera to capture fire events in real time, frame, NodeMCU, and an Arduino Nano make up the suggested gadget. It facilitates effective navigation with the Flysky I6X controller by transmitting a live video feed to the ground via its FPV camera and video transmitter. Using an adaptive optimization practise called fuzzy-based backstepping control, the drone is guided to fly at the desired height and location. Data on gas emissions from controlled burns of different materials are collected and analyzed in this

article to show how successful the device is. The drone collected useful data for firefighting operations by measuring CO, CO2, O3, SO2, and NO2 concentrations in the impacted locations. Depending on the concentration, various quantities of gasses have been measured when burning various things such as alcohol, clothing, plastic, paper, leaves, and more. This work is unique because it builds and analyzes a firefighting drone that uses the internet of things (IoT) to conduct extensive trials in real time.

Airport operations, such as protecting runways from wildlife, maintaining infrastructure, and detecting foreign object debris (FOD), can benefit from the approaches developed by Kovács et al., [18]. When it comes to surface diagnostics, drones with remote sensing devices are invaluable for evaluating taxiways, runways, and aprons. Furthermore, drones can improve airport security by providing data to bolster current air traffic control models and systems and by effectively monitoring and detecting threats. Our goal in writing this article is to share what we've learned about using UAVs equipped with high-resolution RGB, thermal, and LiDAR sensors in possible airport applications. Our study seeks to transform airport operations, safety, and security procedures through interdisciplinary collaboration and novel techniques. Our goal is to provide a route toward a safer and more efficient airport environment.

A categorization method for the delivery drone dropping procedure at a preset target has been proposed by Alsawy et al., [19] using image processing. With the use of GPS data and a single onboard camera, it uses live streaming. On the basis of picture segmentation and classification, a two-step processing method is suggested. Camera and lighting settings, drop zone size, drone height from the ground, and other relevant parameters are considered in the classification. Based on the experimental results, the suggested method offers a quick way with trustworthy accuracy using low-order calculations.

To dropping off delivery drones at specified locations, Abdelhak et al. [20] suggests a categorization method based on image analysis. It makes use of GPS data and a single onboard camera to broadcast live video. An approach with two steps is suggested for processing. using picture categorization and segmentation. Camera and lighting settings, drop zone size, drone height from the ground, and other relevant parameters are considered in the classification. As far as we are aware, this the initial effort to tackle the issue of drone delivery. Based on low-demand computations, the testing findings show that the suggested method gives dependable accuracy.

Using the YOLOv8 algorithm, Chen et al. [21] introduce a method for risk assessment using drones. By iteratively improving the algorithm, the performance has been quadrupled, allowing low-altitude law enforcement to detect more accurately and surpassing the limitations of current surveillance technology. Moreover, five big marathons were subjected to the algorithm's security risk assessments using EWM-TOPSIS and FCM clustering. The events were ranked in order of risk: Zhanjiang, Shanghai, Qingdao, Jiangsu, and Tianjin. The Zhanjiang Marathon had the highest people risk index of 72.12, suggesting that more security measures are needed. By bringing AI to the field of public safety, this study provides a scalable answer to many security problems.

By using a constructive research technique, Jacobsen et al. [22] have connected innovation needs with ideas, designs, and validations, which involve showing and simulating important design components. We take into account the UAS's complexity in our design process and offer a range of technological components for control software and hardware, including algorithms for autonomous service interface. Using UAS communication technologies, an AI-powered drone perception system with accelerated onboard computing, swarm membership, fault detection algorithms are presented in the study. With today's cutting-edge hardware, software, and communication capabilities, we have determined that it is very feasible to create a swarm of cooperative drones and incorporate them into a purpose-built UAS for infrastructure inspection.

III. PROPOSED METHODOLOGY

In this section, the thorough explanation of the predictable methodology is given in **Fig 1**.

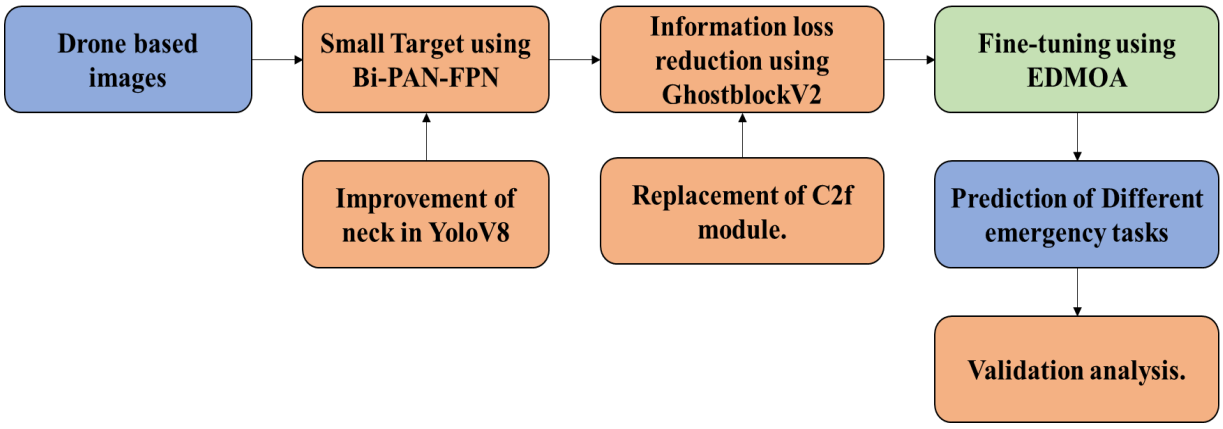

Fig 1. Workflow of The Projected Model.

Dataset Collection

If you want to train a deep learning system to classify UAV images for use in disaster management and emergency response, you'll need to gather a good dataset. We were unable to find any publicly available dataset that is frequently utilized for response applications. Consequently, AIDER (Aerial Image Dataset for Emergency Response applications) is built as a specialized database for this purpose. Each of the five disaster types—fire/smoke, flood, collapsed building/rubble, and traffic accidents—and the normal case (a single class) were manually collected during the dataset preparation [23]. Pictures with common visual elements are combined, for instance, pictures of smoke and raging flames. Classification at a finer level is certainly doable, but will have to wait for another project.

Search terms like "Aerial View," "UAV," or "Drone" plus an event name like "Fire," "Earthquake," "Highway accident," etc., were used to gather the aerial snaps for the catastrophic occurrences from a variety of online sources (e.g., Google photographs, Bing Images, YouTube, news agency web sites, etc.). Before training, images are normalized from varying sizes. First, each image was carefully reviewed by hand to ensure it contained the relevant event. Then, to prevent any geometric alterations used for augmentation from erasing it, the event was centered inside the image. Various catastrophic occurrences were recorded during the data gathering procedure using different resolutions and in varying lighting and perspective conditions. Lastly, the dataset is skewed toward the Normal class in order to reflect real-world situations. Naturally, this can make training more difficult; nevertheless, a suitable approach is used to overcome this throughout training, as will be shown later on.

It is crucial that the dataset consist of "clean" and "clear" photographs because the UAV's operational conditions can differ based on the environment. Data collection also has the potential to be costly and time-consuming. Therefore, each image is probabilistically enhanced with a number of random enhancements before being included to the training batch, in order to further improve the dataset. In particular, these encompass both the geometric transformations and the image alterations (such as changes in illumination, color shifting, blurring, sharpening, and shadowing) that take place along the horizontal axis. To further improve the training set, we additionally use sample pairing to combine images [24]. So that the network doesn't pick up on the augmentation qualities inherent to the dataset, we apply transformations with a random probability that is configured such that only a subset of photos in a training batch are altered. The goal of these changes is to improve generalizability by reducing the likelihood of overfitting and increasing the degree of variability in the training set. In general, over 17 times more data was collected compared to related publications that address multiclass problems, such as [25]. Therefore, augmentations approaches were employed to further improve the baseline dataset. We believe this establishes AIDER as a reliable supplementary data source for establishing besides evaluating data-driven approaches in catastrophe monitoring and emergency response domains. The dataset is described in full in **Table 1**.

Table 1. Summary of the Aider Applications

Class	**Set**			**Whole Pre class**
	Train	**Validation**	**Testing**	
Traffic Accidents	400.	100	200	700
Normal	2700	1000	2000	5700
Collapsed Rubble	420	100	210	700
Smoke	400	110	210	740
Flood	420	100	200	700
Overall Per Set	4320	1410	2810	Inclusive: 8540

Improved Aerial Image Detection Model

The three components that make up this paper's approach for UAV aerial image identification that are both fast and accurate are: To begin with, an upsampling progression is incorporated to concentrate on features, pyramid networks (Bi-PAN-FPN) are substituted for the PAN-FPN in YOLOv8. This is done to address the common issue of small targets in aerial images being misdetected or missed. We attain more sophisticated and thorough feature fusion while minimizing parameter costs by thoroughly analyzing and reusing features across scales. Secondly, since only a handful of parameters have been added to FPN for improvement, it is suggested to substitute some C2f modules in the backbone with the GhostblockV2 structure. This would greatly reduce the number of model parameters while simultaneously suppressing info transmission. Last but not least, WiseIoU loss is switched out for bounding box regression loss from CIoU loss. The detector takes into consideration anchor boxes of varying qualities and enhances the overall task by including a dynamic nonmonotonic mechanism to assess anchor box quality utilizing "outlier.

Improvement of the Neck

According to YOLOv8, feature maps are categorized into five scale feature types in descending order: B1-B5, P3-P5, and N4-N5 in the case of backbone, FPN, and PAN structures, respectively. The original YOLOv8 made use of a PAN-FPN structure, which is an extension of the conventional FPN that transfers deep semantic characteristics in a top-down fashion. There is some loss of location information due to the semantic enhancement of the feature pyramid achieved by combining B3-P3 and B4-P4. The bottom-up structure of the FPN is supplemented by PAN-FPN, and the learning of

localization features is strengthened by fusing P4-N4 and P5-N5, resulting in a complementing effect. Nevertheless, there is potential for enhancement when this structure is used for small target object detection. Firstly, the detection model might miss out on some valuable features and lower detection quality because large-scale feature maps aren't taken into account. Secondly, feature reuse is low and original info after path, even when B, P, and N features are considered and fused and supplemented. The UAV aerial photography dataset underwent the following neck structural adjustments:

Our initial emphasis was on feature maps on a grand scale. The detection effect for small targets was enhanced by adding an upsampling technique to the FPN and fusing it with the features in the B2 layer of the backbone. In a manner analogous to the prior FPN upsampling procedure, the C2f module was employed to augment the post-fusion feature extraction's quality. A significant enhancement over its predecessor, the C3 module, the C2f takes advantage of YOLOv7's ELAN structure and its better gradient information. In order to get more detailed information about the gradient flow while keeping the module lightweight, this one uses the bottleneck module to enlarge the gradient branch and removes one typical convolutional layer.

Next, we presented the concept of Bi-PAN-FPN [26]. Improving the likelihood and timeframes of multiscale feature fusion to achieve improved detection accuracy is the central notion of this structure. Here are the steps to implement it: Feature maps with a single input path do not undergo any additional processing. In a typical scenario, these features don't add much to feature engineering. If the feature maps' sizes are equal, an extra path is added from the backbone features and the features in PAN input paths. Additional parameter cost is not introduced by such a processing method. As a last step, improve blending by treating each top-bottom and bottom-top bidirectional path as a unit and reusing this. Extra routes of B3-N3 and B4-N4 were unit was employed, considering the modest weight of the model. Here is an expression for this process:

$$N_5^{out} = C2f(Concate(Conv(N_5^{in}), B_5^{out}), n) \tag{1}$$

$$N_i^{out} = C2f(concate(Conv(N_i^{in}), B_i^{out}, P_i^{out}), n) \tag{2}$$

in which C2f and Conv are the matching module actions, and B, P, and N are the feature maps at the backbone, FPN, and PAN levels, simultaneously; The value of i can be three or four, and n is the number of times C2f has been used.

Improvement of the Backbone

Thanks to the standard convolution unit and the C2f module, YOLOv8 was able to downsample images and extract features with great quality. The combination of Bi-PAN-FPN and an upsampling method in the neck section did, however, enhance the model's complexity and parameter count. This post will present the Ghostblock concept in backbone and demonstrate how to replace certain C2f modules using this framework. An optimization method for lightweight convolution called GhostNet is Ghostblock [27]. There are two key areas where its benefits are most apparent. Ghostblock, on the one hand, is based on the core principles of GhostNet. The basic feature map is generated using traditional convolution. To improve the feature map's information, it combines many linear transformation processes. This guarantees that characteristics are diverse while being effectively extracted. A decoupled fully connected (DFC) attention mechanism, however, is suggested. This mechanism is unique in that it may capture feature information across great distances while avoiding the computational complexity limits of conventional attention techniques. The overall structure's feature engineering is improved by the structure's advantages. In particular, GhostNet employs a convolutional form known as the cheap operation. These are the steps involved in putting it into action:

$$\Upsilon' = X * F_{1*1} \tag{3}$$

$$\Upsilon = Concat([\Upsilon', \Upsilon' * F_{dp}]) \tag{4}$$

where $X \in R^{C,H,W}, Y \in R^{C\frac{C'}{out},H,W}$; F_{1*1} characterizes pointwise convolution; F_{dp} represents depth-wise convolution; and $C'_{out} \leq C_{out}$. To start the cheap operation, instead of using conventional convolution, we use pointwise convolution to get a feature map that is half the size of the actual output standard. Then, we use depth-wise convolution to get a linear transformation. The output is obtained by splicing the feature maps of the two phases. By recycling features and discarding redundant information that may present in conventional convolutions, this processing method greatly lessens the parameter cost and computing cost. The downsides of this approach are equally apparent: feature maps created using depth-wise convolution capture spatial information, but pointwise convolution misses the interaction process with other pixels in space. This will have an effect on the model's detection accuracy because spatial information will be drastically underrepresented. The self-attention mechanism can quickly raise the model's complexity, in contrast to the convolutional structure's limited ability to concentrate on local input.

The aforementioned issues can be effectively addressed by the DFC attention mechanism. Directly obtaining the attention map with global information by means of a deeply separable structure with a simple construction is the fundamental idea. The particular method of calculation is illustrated in Equations (5) and (6).

$$\alpha'_{hw} = \sum_{h'=1}^{H} F_{h,h'\omega}^{H} \odot X_{h'w'} h = 1,2,\dots,H, \omega = 1,2,\dots,W \quad (5)$$

$$x_{hw} = \sum_{w'=1}^{W} F_{h,h'\omega}^{H} \odot a'_{hw'} h = 1,2,..,H, \omega = 1,2,\dots,W \quad (6)$$

where $X \in R^{C,H,W}$, where the input is given by Equation (3); F is a convolution process that is separable along depth and is separated into horizontal ($K_W * 1$) and vertical ($1 * K_H$) in two different directions: a' represents the direction and an is the attention map in the horizontal direction based on a'. Removing the link between the two ways makes extracting feature global information much easier. The usage of deep separable structures, too, like $1 * K_H$ and $K_W * 1$, the intricacy of the DFC is greatly summary (full connection: $O(H^2W + HW^2)$; DFC: $O(K_H HW + K_W HW)$). By integrating DFC with Ghostblock, we can drastically simplify the model while still accounting for feature global information, all while keeping the cost of operation low.

Improvement of the Loss Function

The loss function of YOLOv8 is significantly different from the YOLOv5 series since the anchor-free idea is used. Classification and regression make up its optimization direction. The binary cross entropy loss (BCEL) is still used for the classification loss, and DFL and BBRL are used for the regression component. One way to express the entire loss function is:

$$floss = \lambda_1 f_{BCEL} + \lambda_2 f_{DFL} + \lambda_1 f_{BBRL} \quad (7)$$

Among them, the prediction sort loss is fundamentally the loss, besides the appearance is:

$$f_{BCEL} = weight[class](-x[class] = log(\sum_j \exp(x[j])) \quad (8)$$

when x is the probability value following sigmoid activation, class is the number of categories, and weight[class] is the set of weights for each class. Optimizing the focal loss function, which integrates the discrete classification results into unceasing ones, is what DFL is all about. This word means:

$$f_{DFL}(S_i, S_{i+1}) = -\big((y_{i+1} - y)\log(S_i) + (y - y_i)\log(S_{i+1})\big) \quad (9)$$

where y_{i}, $y_{\mathrm{i+1}}$ characterizes the values from the left besides consecutive labels y, satisfying $y_i < y < y_{i+1}, y = \sum_{i=0}^{n} P(y_i) y_i$; among the equation, P can be applied finished a softmax layer, $P(y_i)$, that is, S_{i}.

Here, the bounding box regression loss is the Wise-IoU function, which differs from the CIoU loss used in YoloV8. When the training data labeling quality is poor, the loss function uses a "outlier" to assess the anchor frame quality and prevent the model from being penalized too much for geometric factors like distance and aspect ratio. When there is a high degree of coincidence between the prediction box and function decreases the penalty of geometric variables, allowing the model to achieve superior generalization ability with less training intervention. This research utilizes Wise-IoU v3, which incorporates a two-layer attention nonmonotonic FM mechanism, based on this framework. Here is its expression:

$$f_{BBRL} = \left(1 - \frac{W_i H_i}{S_u}\right) exp\left(\frac{(x_p - x_{gt})^2 + (y_p - y_{gt})^2}{(W_g^2 + H_g^2)}\right)\gamma \quad (10)$$

$$\gamma = \beta / \delta \alpha^{\beta - \delta} \quad (11)$$

where β signifies the level of discordance with the anticipated box; a lower level of discordance suggests a better anchor box quality. Therefore, using β In order to build a nonmonotonic focus number, it is possible to reduce the detrimental gradients of low-quality training tasters by assigning tiny gradient with big outliers; a and δ are hyperparameters. The meanings of the supplementary restrictions are publicised in Fig 2. x_{p} and y_{p} represent the coordinate standards of the forecast box, while x_{gt} and y_{gt} serve as the coordinate standards representing the Ground truth. The two boxes' width and height are denoted by the appropriate H and W numbers, correspondingly. It can be seen that $S_u = wh + w_{gt}h_{gt} - W_i H_i$. The loss, backbone, and neck functions are better than the original YOLOv8. The figures graphic labels have the details of the revisions.

Fine-tuning using Enhanced DMOA

There are three distinct social strata in the dwarf mongoose (DM) population in the DMOA: the alpha group, the babysitters, and the scouts. In a family foraging team, the alpha female takes the lead and decides where to sleep, how far

to go, and what direction to go. Babysitters are provided by a subset of the DM population that is typically a mix of male and female kinds. They accompany the kids till the remainder of the group shows up in the afternoon. First, during the exploitation period, the babysitters are switched so that they can start foraging with the group. In order to find a fresh area to forage, the DM group never builds a nest; instead, they constantly move the sleeping mound. The DMs have settled into a seminomadic lifestyle in an expansive region that can house all of the party members (the exploration phase). Nomadic conduct is characterized by a lack of attachment to any one area. Additionally, it checks that no previously visited sleeping mounds are revisited by ensuring that the entire terrain is surveyed.

For each NDM individual's potential solution, the DMOA randomly generates an initial DM population in the following ways:

$$D_{j,d}(0) = D_{min,d} = rand(0,1).\left[D_{max,d} - D_{min,d}\right], j = 1{:}\,N_{DM}, d = 1{:}\,Dim \tag{12}$$

where $D_{j,d}$ represents the location as a searching individual concerning each $DM(j)$ and each control variable (d); $D_{min,d}$ and $D_{max,d}$ denote the bounds of each control variable (d); Dim refers to the sum of the decision variables task.

After the population is initialized, the solution is determined. The probable worth of each group's fitness is determined by Equation (13) besides the alpha female (α) on this likelihood.

$$a = \frac{F_j}{\sum_{j=1}^{N_{DM}} F_j} \tag{13}$$

The alpha group's DM count is proportional to the discrepancy between the total group size (NDM) and the babysitter count (Bst). The symbol bs represents the total number of babysitters. Peep is the DM family's guiding vocalization, which comes from the alpha female. The first sleeping mound, designated for ˇ, is where every DM goes to sleep. Equation (14) is the formula that the DMOA uses to create a possible food position.

$$D_{k,d}(i+1) = D_{k,d}(i) + rand(0,1) \times peep, k = 1{:}\,N_{DM} - Bse, d = 1{:}\,Dim \tag{14}$$

where "i" denotes the current iteration. Following each cycle, the sleeping mound can be constructed in the following way:

$$SM_i = \frac{F_{j+1} - F_j}{max(|F_{j+1} - F_j|)} \quad j = 1{:}\,N_{DM} - Bs \tag{15}$$

where j mentions to each DM in group which are the alteration among the whole group number (NDM) and the sum of babysitters (Bst).

According to that, Equation (16) gives the mean value (y) of the discovered.

$$\psi_j = \frac{\sum_{j=1}^{N_{DM}} SM_j}{N_{DM}} \quad j = 1{:}\,N_{DM} - Bst \tag{16}$$

Once the childcare exchange threshold is satisfied, the DMOA approach moves to the scouting stage, when a subsequent found. While foraging, scouts in DMOA look for a different sleeping mound to make sure they've explored everything. Based on the complete mongooses, the resultant motion is shown as an evaluation of whether or not to create a new mound.

$$D_{k,d}(i+1) = \begin{cases} D_{k,d}(i) - CF \times rand(0,1) \times \left(D_{k,d}(i) - M\right) if\ \psi_{j+1} > \psi_j \\ D_{k,d}(i) + CF \times rand(0,1) \times \left(D_{k,d}(i) - M\right) \quad Else \end{cases} \\ k = 1{:}\,N_{DM}, d = 1{:}\,Dim \tag{17}$$

where CF decreases linearly with the number of repeats, as shown in Equation (18), and M is a vector that determines where the mongoose will sleep, as shown in Equation (19). The CF limit controls the collective volitional movements of the mongoose group.

$$CF = \left(1 - \frac{i}{Iter_{max}}\right)^{\left(\frac{2 \times i}{Iter_{max}}\right)} \tag{18}$$

$$M = \sum_{j=1}^{N_{DM}} \frac{D_j \times SM_j}{D_j} \tag{19}$$

where $Iter_{\max}$ refers to the highest sum of repetitions.

Proposed EDMOA

Here we provide a novel EDMOA that incorporates an Learning Strategy (LS) to address various mathematical benchmarking tasks and engineering obstacles. The revised alpha partially guides the updating process of the improved LS, which is part of the unique proposed solver and improves the searching capabilities. In an attempt to improve the searching capabilities, the alpha-directed LS is combined with the formula given in Equation (14) to produce a possible food site. Consequently, each search solution's location inside the search space is enhanced in the subsequent way:

$$D_{k,d}(i+1) = \begin{cases} BestDM_d(i) + rand(0,1) \times \left(D_{k,d}(i) - D_{R,d}(i)\right) \; if \; rand > CP \\ D_{k,d}(i) + rand(0,1) \times peep \;\; Else \end{cases} \quad (20)$$
$$k = 1{:}N_{DM} - Bst, d = 1{:}Dim$$

DR,d is a randomly designated searching separate from the DM populace; CP is the choice probability; and BestDMd is the site as the seeking separate with the best fitness score. Equation (20) mentions greater exploitation qualities, and Equation (14) mentions exploratory features; to maintain a balance between the two, CP is set to 50%. Utilizing the aforementioned structure results in substantial and powerful exploitation characteristics while preserving and obtaining exploratory seeking qualities using the conventional way simultaneously.

IV. RESULTS AND DISCUSSION

The trials are directed on a PC with an $Intel\ Core\ i5-7200\ CPU$, 8 GB of RAM, and a processing speed of 2.7 GHz. The processes are executed using a specialized User Interface (UI) and Jupyter Notebook on Windows 10, a 64-bit operating scheme (Python 3.7) The setting.

Learning Rate Analysis

The presentation of the projected model is tested with different learning rate in terms of different metrics that is exposed in **Fig 2**.

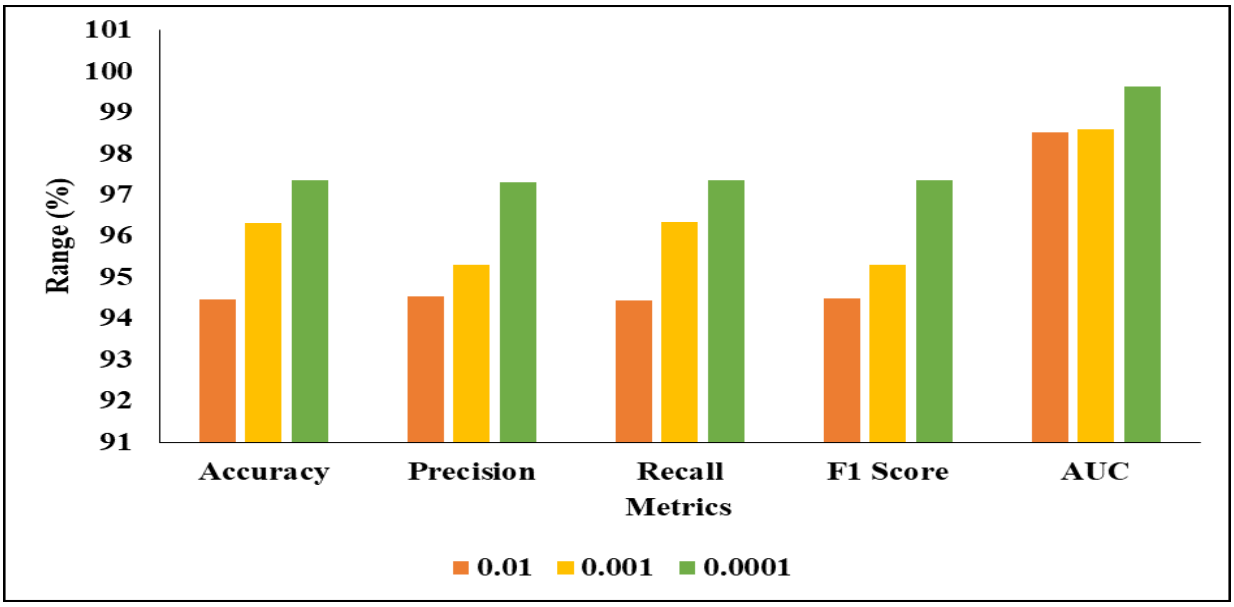

Fig 2. Investigation of Proposed Classical on Learning Rate.

In **Fig 2** represent as proposed model on learning rate. In the analysis of 0.01 learning rate condition, the accuracy as 94.47 then precision ranges of 94.55 and also recall as 94.45 and then f1-score as 94.48 and also AUC rate as 98.53 congruently. Then the 0.001 learning rate condition, the accuracy as 96.33 then precision ranges of 95.30 and also recall as 96.35 and then f1-score as 95.30 and also AUC rate as 98.60 congruently. Then the 0.0001 learning rate condition, the accuracy as 97.35 then precision ranges of 97.32 and also recall as 97.35 and then f1-score as 97.37 and also AUC rate as 99.64 similarly.

Comparative Analysis of Proposed model

The existing model is tested with proposed model's effectiveness in terms of different metrics is exposed in **Fig 3**.

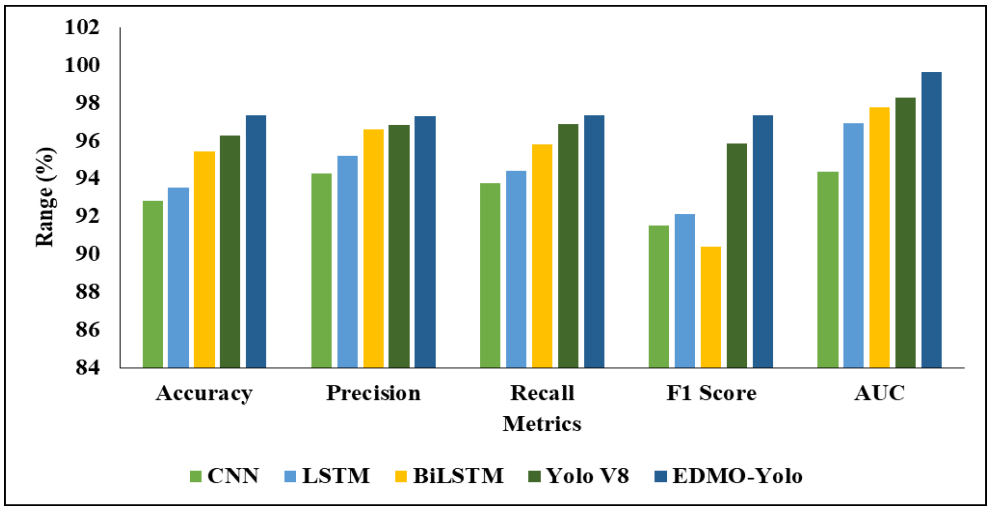

Fig 3. Visual Representation of The Projected Model with Existing Actions.

In **Fig 3** represent the Visual Illustration of the proposed model with existing techniques. In the analysis of CNN technique attained the accuracy as 92.82 then precision ranges of 94.29 and also recall as 93.76 and then f1-score as 91.52 and also AUC rate as 94.38 similarly. Then the LSTM technique attained the accuracy as 93.52 then precision ranges of 95.21 and also recall as 94.43 and then f1-score as 92.11 and also AUC rate as 96.96 correspondingly. Then the BiLSTM technique attained the accuracy as 95.45 then precision ranges of 96.61 and then f1-score as 95.80 and then f1-score as 90.40 and also AUC rate as 97.80 harmoniously. Then the Yolo V8 technique attained the accuracy as 96.29 then precision ranges of 96.83 and also recall as 96.90 and then f1-score as 95.87 and also AUC rate as 98.28 congruently. Then the EDMO-Yolo technique attained the accuracy as 97.35 then precision ranges of 97.32 97.35 and then f1-score as 97.37 and AUC rate as 99.64 congruently.

V. CONCLUSION

In order to automatically detect and categorize catastrophic situations in real-time from a UAV, a study was conducted on the design besides application of an effective deep learning system. The suggested approach strikes a good balance between accuracy, implication speed, and complexity; moreover, it can serve as a foundation for future use-cases that have comparable requirements. The goal of this study is to present a YOLOv8-s based aerial image identification algorithm that, when deployed, can reliably identify targets in aerial images in real time. This model eliminates the detrimental impacts of several elements on the identification task, such as camera angle, lighting, and backdrop. First, the concept of Bi-PAN-FPN is proposed to enhance the neck part in YOLOv8-s, addressing the prevalent issue of small targets being misdetected or missing in aerial photos. We achieve a more progressive and thorough feature fusion procedure while upholding the parameter cost as much as feasible by completely seeing and reusing multiscale features. Secondly, in order to improve the overall presentation of task, the sensor takes into consideration different quality anchor boxes "outlier" to evaluate their quality. Lastly, the GhostblockV2 structure is used in perfect to substitute part of the C2f module. This structure effectively reduces the number of model limits while suppressing transmission. To improve classification accuracy, EDMOA optimally selects the model's fine-tuning. Unfortunately, another issue surfaced throughout the experiment: the ablation tests showed that the model presented in this article does not outperform other structures in any of the small categories. We intend to address these issues in our future studies by combining them with individualized detection tasks and investigating how to adaptively alter the representation's structure from the viewpoints of the representation's hyperparameters and the network's composition.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests

References

[1]. A. A. Adegun, S. Viriri, and J.-R. Tapamo, "Review of deep learning methods for remote sensing satellite images classification: experimental survey and comparative analysis," Journal of Big Data, vol. 10, no. 1, Jun. 2023, doi: 10.1186/s40537-023-00772-x.

[2]. X. Liu, K. H. Ghazali, F. Han, and I. I. Mohamed, "Review of CNN in aerial image processing," The Imaging Science Journal, vol. 71, no. 1, pp. 1–13, Jan. 2023, doi: 10.1080/13682199.2023.2174651.

[3]. A. Thirumalraj, V. Asha, and B. P. Kavin, "An Improved Hunter-Prey Optimizer-Based DenseNet Model for Classification of Hyper-Spectral Images," AI and IoT-Based Technologies for Precision Medicine, pp. 76–96, Oct. 2023, doi: 10.4018/979-8-3693-0876-9.ch005.

[4]. H. Ouchra, A. Belangour, and A. Erraissi, "Machine Learning Algorithms for Satellite Image Classification Using Google Earth Engine and Landsat Satellite Data: Morocco Case Study," IEEE Access, vol. 11, pp. 71127–71142, 2023, doi: 10.1109/access.2023.3293828.

[5]. Ö. İNİK, "Havasal Görüntülerdeki Sahnelerin Derin Öğrenme Modelleri ile Sınıflandırılması," Türk Doğa ve Fen Dergisi, vol. 12, no. 1, pp. 37–43, Mar. 2023, doi: 10.46810/tdfd.1225756.

[6]. I. Papoutsis, N. I. Bountos, A. Zavras, D. Michail, and C. Tryfonopoulos, "Benchmarking and scaling of deep learning models for land cover image classification," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 195, pp. 250–268, Jan. 2023, doi: 10.1016/j.isprsjprs.2022.11.012.

[7]. V. Revathi, B. P. Kavin, A. Thirumalraj, E. Gangadevi, B. Balusamy, and S. Gite, "Image Based Feature Separation Using RBM Tech with ADBN Tech for Accurate Fruit Classification," 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Feb. 2024, doi: 10.1109/ic2pct60090.2024.10486564.

[8]. P. Gadhave, P. Chaturvedi, S. Bera, A. Singh, and R. Joseph, "Post-Disaster Aerial Image Analysis Using Deep Learning and Image Processing," Soft Computing for Security Applications, pp. 345–362, 2023, doi: 10.1007/978-981-99-3608-3_24.

[9]. A. Bouguettaya, H. Zarzour, A. Kechida, and A. M. Taberkit, "A survey on deep learning-based identification of plant and crop diseases from UAV-based aerial images," Cluster Computing, vol. 26, no. 2, pp. 1297–1317, Aug. 2022, doi: 10.1007/s10586-022-03627-x.

[10]. F. Kateb and M. Ragab, "Archimedes Optimization with Deep Learning Based Aerial Image Classification for Cybersecurity Enabled UAV Networks," Computer Systems Science and Engineering, vol. 47, no. 2, pp. 2171–2185, 2023, doi: 10.32604/csse.2023.039931.

[11]. A. Clark, S. Phinn, and P. Scarth, "Optimised U-Net for Land Use–Land Cover Classification Using Aerial Photography," PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, vol. 91, no. 2, pp. 125–147, Feb. 2023, doi: 10.1007/s41064-023-00233-3.

[12]. R. F. Mansour and E. Alabdulkreem, "Disaster Monitoring of Satellite Image Processing Using Progressive Image Classification," Computer Systems Science and Engineering, vol. 44, no. 2, pp. 1161–1169, 2023, doi: 10.32604/csse.2023.023307.

[13]. R. Jain et al., "Internet of Things-based smart vehicles design of bio-inspired algorithms using artificial intelligence charging system," Nonlinear Engineering, vol. 11, no. 1, pp. 582–589, Jan. 2022, doi: 10.1515/nleng-2022-0242.

[14]. Y. Li, Q. Fan, H. Huang, Z. Han, and Q. Gu, "A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition," Drones, vol. 7, no. 5, p. 304, May 2023, doi: 10.3390/drones7050304.

[15]. T. Mollick, M. G. Azam, and S. Karim, "Geospatial-based machine learning techniques for land use and land cover mapping using a high-resolution unmanned aerial vehicle image," Remote Sensing Applications: Society and Environment, vol. 29, p. 100859, Jan. 2023, doi: 10.1016/j.rsase.2022.100859.

[16]. Mahaveerakannan R, C. Anitha, Aby K Thomas, S. Rajan, T. Muthukumar, and G. Govinda Rajulu, "An IoT based forest fire detection system using integration of cat swarm with LSTM model," Computer Communications, vol. 211, pp. 37–45, Nov. 2023, doi: 10.1016/j.comcom.2023.08.020.

[17]. N. Jahan, T. B. M. Niloy, J. F. Silvi, M. Hasan, I. J. Nashia, and R. Khan, "Development of an IoT-based firefighting drone for enhanced safety and efficiency in fire suppression," Measurement and Control, Apr. 2024, doi: 10.1177/00202940241238674.

[18]. B. Kovács, F. Vörös, T. Vas, K. Károly, M. Gajdos, and Z. Varga, "Safety and Security-Specific Application of Multiple Drone Sensors at Movement Areas of an Aerodrome," Drones, vol. 8, no. 6, p. 231, May 2024, doi: 10.3390/drones8060231.

[19]. A. Alsawy, D. Moss, A. Hicks, and S. McKeever, "An Image Processing Approach for Real-Time Safety Assessment of Autonomous Drone Delivery," Drones, vol. 8, no. 1, p. 21, Jan. 2024, doi: 10.3390/drones8010021.

[20]. A. ABDELHAK, D. Moss, A. Hicks, and S. mckeever, "An Image Processing Approach for Real-Time Safety Assessment of Autonomous Drone Delivery," SSRN Electronic Journal, 2022, doi: 10.2139/ssrn.4192669.

[21]. G. Chen, W. Du, T. Xu, S. Wang, X. Qi, and Y. Wang, "Investigating Enhanced YOLOv8 Model Applications for Large-Scale Security Risk Management and Drone-Based Low-Altitude Law Enforcement," Highlights in Science, Engineering and Technology, vol. 98, pp. 390–396, May 2024, doi: 10.54097/fbxtqk26.

[22]. R. H. Jacobsen et al., "Design of an Autonomous Cooperative Drone Swarm for Inspections of Safety Critical Infrastructure," Applied Sciences, vol. 13, no. 3, p. 1256, Jan. 2023, doi: 10.3390/app13031256.

[23]. C. Kyrkou and T. Theocharides, "EmergencyNet: Efficient Aerial Image Classification for Drone-Based Emergency Monitoring Using Atrous Convolutional Feature Fusion," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 1687–1699, 2020, doi: 10.1109/jstars.2020.2969809.

[24]. K. Kamaraj, B. Lanitha, S. Karthic, P. N. Senthil Prakash, and R. Mahaveerakannan, "A Hybridized Artificial Neural Network for Automated Software Test Oracle," Computer Systems Science and Engineering, vol. 45, no. 2, pp. 1837–1850, 2023, doi: 10.32604/csse.2023.029703.

[25]. M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2020, doi: 10.1109/cvpr42600.2020.01079.

[26]. K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More Features From Cheap Operations," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2020, doi: 10.1109/cvpr42600.2020.00165.

[27]. J. O. Agushaka, A. E. Ezugwu, and L. Abualigah, "Dwarf Mongoose Optimization Algorithm," Computer Methods in Applied Mechanics and Engineering, vol. 391, p. 114570, Mar. 2022, doi: 10.1016/j.cma.2022.114570.