

Real Time Drone Detection Using YOLO and Coordinate Attention Integrated GELAN-E Models

Kapil Kumar Pandey¹, Dr. S. K. Singh², Dr. J. N. Singh³

¹ Research Scholar, Amity Institute of Information Technology, Amity University, Lucknow, India

² Professor, Amity Institute of Information Technology, Amity University, Lucknow, India

³ Professor, Computer Science and Engineering Department, Galgotia University, India

Abstract:- In the recent scenario drones have play important role in war, aerial surveillance, illegal spying, maintain public safety in sensitive areas, logistics and environmental observation. To prevent the unauthorized drone, use the drone detection in real time. In real time one of the difficult tasks in object detection is drone detection due to visibility conditions, complex backgrounds, weather condition, tiny visible things and similar objects that are difficult to identify. Our study proposes a real-time drone identification object using YOLOv8, YOLOv11, and GELAN-E integrated with Coordinate Attention deep learning models, that enhanced processes and interpretability of object detection techniques. By integrating the GELAN-E integrated with Coordinate Attention YOLO is known as YOLOv8n-CA-GE YOLOv11n-CA-GE, the study significantly improves performance, particularly for small and hidden drones. The proposed drone detection model YOLOv8n-CA-GE with an accuracy of precision 0.962, a recall of 0.917, mAP@0.5 score of 0.956 and mAP@0.5:0.95 score of 0.755, YOLOv11n-CA-GE with an accuracy of precision 0.962, a recall of 0.930, mAP@0.5 score of 0.966 and mAP@0.5:0.95 score of 0.773. For all weather and day conditions. In addition to visual interpretability, enhances drone detection performance in real-time.

Keywords: Real-time, Deep learning, object detection, Generative AI Models.

1. Introduction

The rapid use of unmanned aerial vehicles (UAVs), also known as the drone has extremely impacted industries like logistics, aerial surveillance, environmental monitoring, military, security risks, as drones can be misused for illicit surveillance, smuggling, or carrying dangerous materials. In the drone identification main barrier is the vast array of UAVs is now in use, each with different sizes, speeds, and altitudes of flight, so the drone detection in real time is a challenging task [1]. The cutting-edge drone detection technologies that can accurately recognize and remove potential threats are required to ensure the security of susceptible sites, such as essential infrastructure, commercial buildings, and public events [1]. Unmanned Aerial Vehicles (UAVs) can be either multi-rotor or fixed-wing aircraft, controlled by remote operator or autonomously. They vary widely in shape and size, from models resembling insects to large aircraft weighing several tons. Major organizations such as NATO, the Department of Defense (DoD), NASA, and regulations bodies have established primary categories for UAV systems [15]. The wide variety of UAVs currently in use, each with unique dimensions, and different velocities, flight altitudes, is a major barrier to drone identification. Additionally, the COVID-19 epidemic has led to a sharp rise in the employment of drones for social distance as well as autonomous delivery of essential medical supplies and groceries [2].

As the latest first drone vs. bird identification challenge was outlined in 2017 in connection with the very first edition of the International Workshop on Small-Drone Surveillance, Detection and Counteraction Techniques (WOSDETC) due to the recent and quick interest in drone use. The more technical issues in differentiating between the drones vs. birds were the focus of this task [3]. In reality, drones can be mistaken for birds, especially

when they are far away, which makes the surveillance duty even more difficult. Since the challenge's inception, drone detection a branch of object detection has received a lot of attention. Although there were many other approaches used in the solutions put forth for drone tracking and identification, deep learning is currently the most popular approach. Additionally, other works use methods like SVMs and AdaBoost that are not dependent on deep learning [3] [4]. The deep learning process involves figuring out the problem, picking the right tool, gathering and cleaning your data, teaching the computer using examples, and finally checking if it can make smart guesses on its own [35]. The presence of similar objects in the sky, such airplanes, birds, and so on, may make it challenging to identify drones. The dataset used by the scientists included both birds and drones [6].

A number of specialized datasets and benchmarks have been established to aid in UAV identification studies. While drone-versus-bird datasets and the Anti-UAV benchmark offer domain-specific samples encompassing a variety of weather conditions and cluttered surroundings [3][10], UAVSwarm offers multi-target UAV tracking data for large-area surveillance [8]. By exposing models to uncommon drone categories and harsh environments, synthetic datasets and augmentation techniques further improve generalization [2][8]. Confusion between classes is another issue that drone detection systems must handle, particularly when attempting to distinguish drones from birds or other flying objects. Time and motion pattern information can significantly reduce misclassification, as demonstrated by the Anti-UAV benchmark and sequence-model-driven drone-bird categorization [3][10]. Additionally, by increasing the diversity of training data and improving time consistency, synthetic data creation and track boosting techniques improve detection stability over long distances and when objects are hidden [2][4]. Currently majority of the anti-drone systems use military-grade parts to guarantee that hostile drones are consistently neutralized. However, there are a number of difficulties when implementing these military-grade systems in civilian settings. [16].

Because drones often look like small targets with few pixels, finding small things in the air is still a big problem for UAV surveillance. Attention mechanism helps model to handle long sequences in data [34]. Several studies have attempted to tackle this issue by including attention mechanisms into YOLO-based models, augmenting feature pyramids, and refining multi-scale fusion. Two techniques, DroBoost and VisioDECT that use attention and boosting to increase feature quality, making it simpler to locate objects in complex scenes [1] [4]. In other, models that used for on small objects, YOLO-Drone and TF-Net, strengthen and fine-grained feature extraction to make it simpler to distinguish between drones and distant or background objects [5] [9].

One of the greatest frameworks for real-time object recognition for UAV is the YOLO family. YOLOv5-based long-range small-drone identification and drone detection work well when speed and accuracy are critical [6][7]. When there are plants, structures, varying light levels, and a lot of background noise, YOLO-Drone and YOLOv8-based detectors are examples of optimized variants that increase resilience [5] [7]. Recent developments such as YOLOv10 and YOLOv11 improve long-range surveillance by increasing its sensitivity to small, fast-moving aerial targets and making it easier to extract characteristics at different scales [11] [5]. By these challenges, this study proposes an enhanced YOLO11n-based framework in conjunction with Coordinate Attention (CA) to detect small flying objects in complicated backgrounds and inclement weather. By including positional information into channel attention, Coordinate Attention enables the network to preserve fine spatial details that are normally lost during down sampling [12]. The goal of adding CA to YOLO11n is to improve long-range sensitivity, decrease false detections, and improve feature localization while preserving a lightweight architecture. Accuracy loss in lightweight YOLO designs can be successfully reduced by CA, according to earlier research [10] [14]. The proposed study approach to create a fast, accurate, and real-time UAV recognition system that may be applied in various multi-weather surveillance scenarios.

2. Related Work

Islam et al. [1] developed the VisioDECT framework, which incorporates a CBAM-enhanced YOLO model (AttnYOLO) to boost drone detection performance in cloudy and low-light evening scenarios. Their approach is designed to better identify small or partially hidden UAVs by applying both spatial and channel attention, utilizing a dataset of 20,924 labeled images. The model achieved mAP50-95 score of 0.72 under cloudy condition and

0.651 during evening scenes. Additionally, EigenCAM was used to provide heatmap visualizations, offering greater insight into how the model makes its predictions.

Akyon et al. [2] developed a comprehensive framework for drone object recognition and tracking, which got in the pl 2021. Drone vs. Bird Challenge by combining YOLOv5 with a Kalman filter-based object detection. To tackle the difficulty of identifying distant UAVs in challenging conditions, such as low visibility and poor contrast, they enhanced the model by training it on a mix of real-world and synthetically generated images. Their findings showed that carefully augmenting the training dataset with select synthetic images can greatly improve detection reliability and overall system performance.

Hou et al. [12] introduced Coordinate Attention (CA), an efficient attention mechanism designed to enhance neural network performance by integrating both spatial and channel information. Unlike traditional attention modules that may overlook positional details, CA allows the network to retain long-range dependencies while simultaneously zeroing in on fine-grained spatial features. This dual capability makes it especially beneficial for tasks like tiny-object detection, where recognizing subtle differences and maintaining awareness of object locations are both critical for accurate results.

Yuanyuan, et al. [17] introduce Effective anti-UAV measures rely heavily on UAV object detection technology. YOLO v3, a leading single-stage detector, offers an optimal ratio between accuracy and speed by effectively capturing both high-level and deep features. The method utilizes the last four feature map scales, rather than the traditional three, to predict object bounding boxes, enabling the capture of more texture and contour details for improved detection of small objects.

Ren et al [19] introduce the Region Proposal Network (RPN) share the convolutional features of whole image with the detection network, along with region proposals to be generated very minimal computational cost. An RPN is the object ness scores are predicted using a fully convolutional network and object boundaries at each and every position in the image. It is trained end-to-end and produce high-quality region proposals, which are then used by Fast R-CNN for object detection [18]. He et al. [20] Mask R-CNN enhances Faster R-CNN by introducing a branch that uses the current bounding box prediction in conjunction with object mask predictions train, and operates at 5 frames per second.

Redmon et al. [21] and Liu et al. [22] are achieved real-time speed by using entire image in a single pass YOLO novel approach to object recognition. Unlike previous studies classifiers use for detection tasks, while YOLO treats object recognition as a regression problem and directly predicting spatially distinct bounding boxes and their associated probability in the classes [21]. Lin et al. [23] proposed that the during the training of dense detector is the severe imbalance between foreground and background classes. They suggest tackling this class imbalance by modifying the standard cross entropy loss to reduces the loss contribution in well-defined classified example.

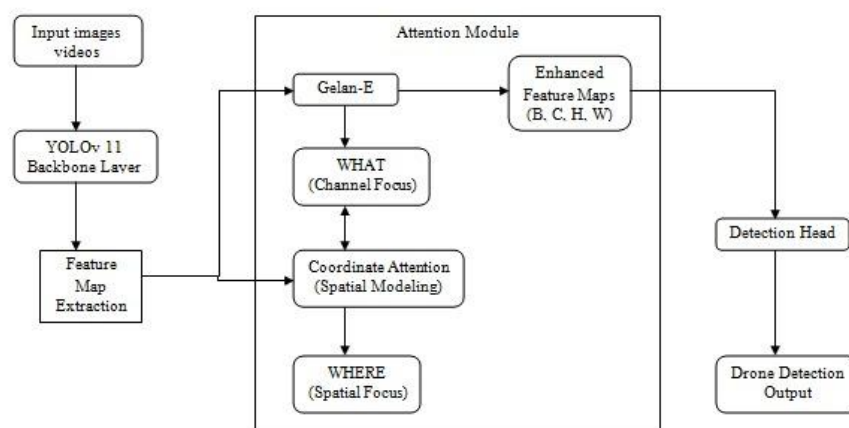


Figure 1. Proposed flow of work

3. Proposed Methodology

Fig. 1. Shows our methodology is to have two type real data set image and video with the use of generative AI model stable diffusion together to generate samples, where extract the frame and pass through YOLOv11 model for detection with Grelan E coordinate attention module for data processing.

3.1 Dataset Description

Det-Fly dataset used to images feature a target micro-UAV (DJI Mavic), with the resolution of 3840×2160 pixels. These source images are taken from both video frames captured at 5 FPS and specific relative positions of images. All images have been manually annotated by experts. The dataset can be accessed at <https://github.com/Jake-WU/Det-Fly> [25].

Det-Fly encompasses a diversify set of scenarios, capturing various viewing different angles, backgrounds, distances, and varying lighting conditions (fig 2) [24]. The dataset features four main different types of environmental conditions backgrounds sky, urban, field, and mountain with each background type representing approximately 20% to 30% of the total images. Regarding viewing perspectives, the dataset is categorized into front, top, and bottom views, which make up 36.4%, 32.5%, and 31.1% of the images, respectively. Based on statistical research, the size of the UAV in the photos shows that many UAV Nearly half of the UAV instances appear to be rather small, taking up less than 5% of the whole image area. When a UAV's height and breadth are less than 10% of the whole image, it is recognized as a tiny object., making detection particularly challenging. Additionally, the dataset takes into consideration variable lighting conditions by incorporating at various times of the day., from morning to evening. and, it includes difficult situations like motion blur (11.2%), strong or weak lighting (10.8% of photos), and situations where the UAV is partially obscured (0.8%). Figure 2 sample image of drone in the dataset.

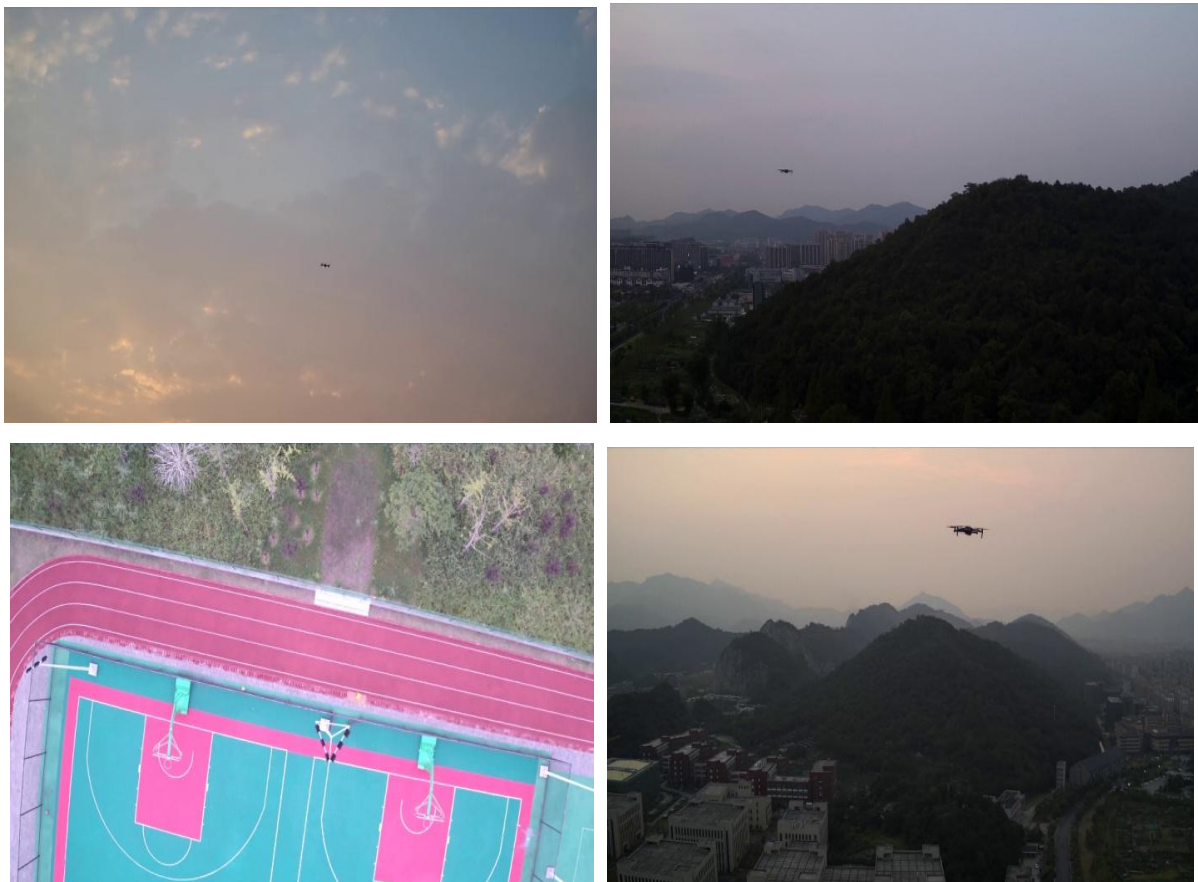


Figure 2. sample image of the Det-fly dataset

3.2 Dataset Preprocessing

The dataset was preprocessed using a consistent methodology for consistency throughout in the training process. Image was resized to 960×960 pixels to align with the input dimensions as required by the YOLO framework. The pixel values were normalized to lie inside the $[0,1]$ range., XML or CSV formats, annotation files were transformed into TXT label files that are compatible with YOLO framework. and, class labels were carefully design to distinguish between single-class and multi-class dataset., a comprehensive manual assessment was also carried out to remove any corrupted images and fix mislabeled samples to ensuring the quality and dependability on training data.

The Latent Diffusion Model (LDM), as explained by Rombach et al., was used to create over 3,000 synthetic UAV photos plus a number of negative samples and to increase the datasets strengthen and diversity of the model's robustness [27]. These synthetic images were included to represent all scenarios of weather condition, lighting conditions that are not present in the original dataset, thereby data is more comprehensive. and model can more effectively generalize to a greater variety of real-world scenarios by using these synthetic images also used data augmentation to enhance data quality and reliability of data.

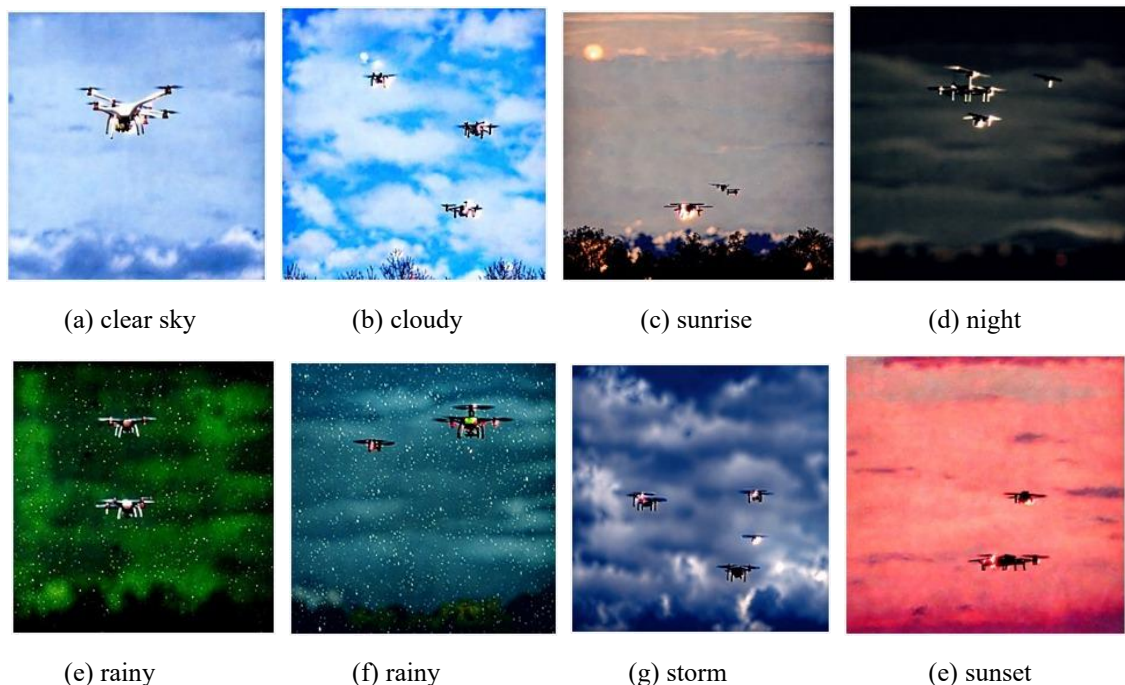


Figure 3. Sample synthetic drone images generated under diverse background conditions.

Figure 3. the samples of the Synthetic data which is generated by with the help of Generative AI Models stable diffusion. We include different weather conditions like clear sky, cloudy, sunrise, night, rainy, storm and sunset so the model detect the drone in different weather condition.

3.3 Model Selection

Stable Diffusion Model

Stable Diffusion is the generative pertained AI model which is based on Latent Diffusion (LDMs) and mainly used for Text-to-Image generation [29]. When it comes to creating images, it can produce better and enhanced results than the GAN model. Random Gaussian noise can be progressively de-noised following training in applications including unconditional synthesis of pictures, image restoration (image inpainting), super-resolution, generation of text-to-image, etc. The steady diffusion model consists of three parts fundamental: First is, Variational autoencoders (VAE) contain both encoder and decoder. The encoder converts the image into a low-dimensional latent space representation for U-Net while preserving important deep picture properties, and the

decoder generates images from latent space representations. Second is, U-Net is a residual module-based encoder and decoder that, following image compression, decodes and enhanced lower quality-resolution images to higher quality -resolution images. Third is, the input text is converted into a meaningful text that U-Net can understand and use to guide the model as it de-noises the embedding by the Text-Encoder, which converts into the tagged sequence to sequence for embedding text [28][29].

YOLOv11 Model

With YOLOv11 [30] [31], the YOLO algorithm's development has hit a new milestone and a major breakthrough in object identification in real time. In addition to improving the characteristics of earlier models, this most recent edition adds cutting-edge capabilities that increase its utility in a variety of computer vision applications. YOLOv11 stands out for being more versatile, enabling it to perform a variety of computer vision tasks beyond object detection. Pose estimation and instance segmentation are important enhancements that increase the model's applicability in various domains. Its architecture is carefully crafted to balance usability and performance, addressing industry-specific issues more accurately and effectively. The release of this model expands the potential for computer vision applications and reflects the ongoing advancements in real-time object detection. YOLOv11 stands out as a significant innovation with its increased performance and versatility, opening the door for wider real-world use across numerous industries [30]. Three fundamental components comprise the YOLOv11 architecture. First, the primary feature extractor is the backbone, which uses convolutional neural networks to transforming unprocessed data of images into multi-scale feature mapping. Secondly, the components of neck acts as an intermediate processing by utilizing customized layers to gather and enhance feature representation at multiple scales [32].

Galen-E Channel Attention Module

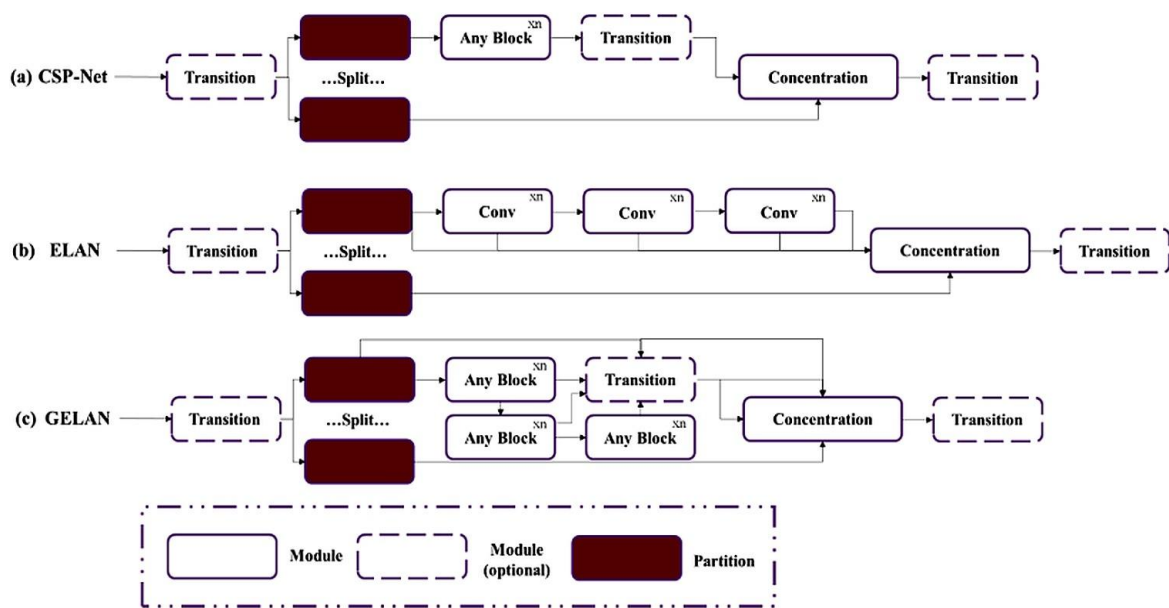


Figure 4. The architecture and configuration of GELAN [1].

The architecture of the GELAN model is shown in Figure. 4. The GELAN architecture is divided into two main part one is the backbone and another is the head. The backbone functioning is to take the input image and extract its features. CSP-ELAN blocks, which carry out complex procedures to extract hierarchical features at different sizes. The combined features are processed by the GELAN model's head for final detection. In order to capturing multi-scale contextual information and its essential for identifying the object of various size, it incorporates spatial pyramid pooling in the SPPELAN block. CSP-ELAN blocks facilitate improved gradient flow and are in charge of effective feature aggregation from various backbone levels.[33]

Coordinate Attention

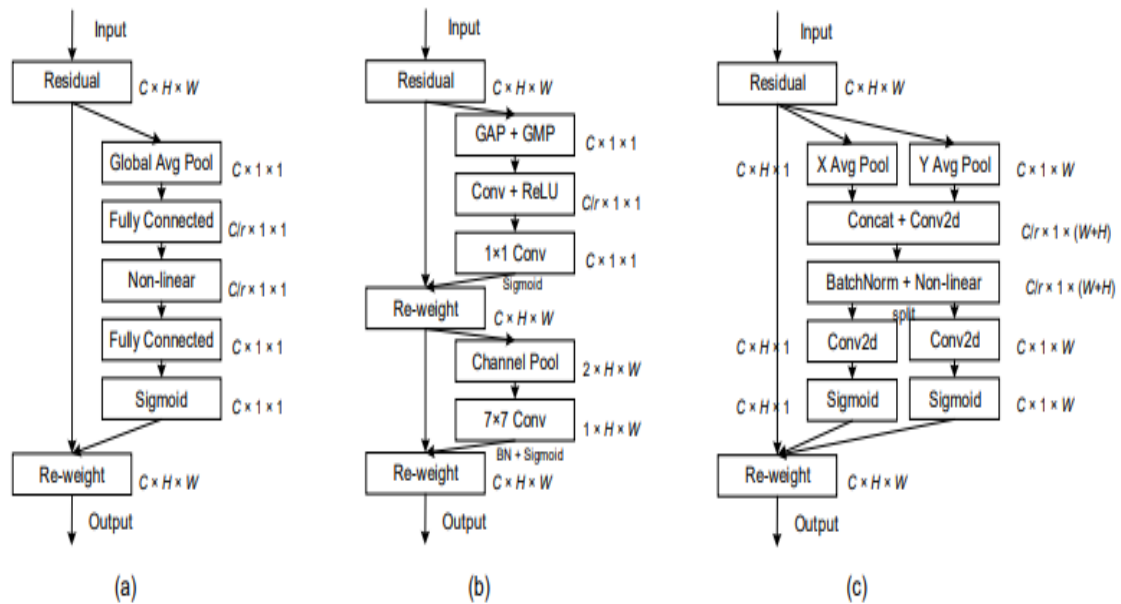


Figure 5. Coordinate Attention [12].

Figure 5. The diagrammatic comparison between the traditional SE channel attention block (a) and CBAM (b) and the suggested coordinate attention block (c). In this context, "GAP" and "GMP" stand for global average pooling and global maximum pooling, respectively. The terms "X Avg Pool" and "Y Avg Pool" denote 1D vertical global pooling and 1D horizontal global pooling, respectively [12]. A coordinate attention block is a computational module intended to increase the representational capacity of learned information. The result is a converted tensor with enhanced $Y = [y_1, y_2, \dots, y_C]$, with the same dimensions as X [12]. It can process any intermediate feature of tensor $X = [x_1, x_2, \dots, x_C] \cdot RC \times H \times W$.

4. Experimental Setup

Drone detection experiments were performed on a Kaggle GPU T4x2 environment using Python, PyTorch [26], OpenCV, Ultralytics, Pathlib, sys, and os. Both the image-based and video-based models were trained & test using the same experimental settings for consistency. For training, we used the following hyper parameters:

Batch size: 16

Image size: 960×960 (input)

Learning rate: 1e-4

Epochs: 100, 200

Workers: 4

4.1 Train, Validation, and Test Split

The dataset is divided into three parts, 70% of the dataset is used for training, 20% for validation, and the remaining 10% for testing. To ensure consistency, the split was performed such that the distribution of UAV sizes, flight distances, and background complexities was preserved across all subsets. This method ensures that difficult samples are included in both the validation and test sets., incorporating situations with complex backgrounds and small drones at farther ranges., relevant drone features were learned using the training set, and model selection and hyper parameter used for the validation, and the test set was kept completely apart for the last assessment of the model's performance. Table 1 shows the image description for experiment.

Table 1: DETFY and Synthetic Dataset images Train, Validation, and Test

Class	Dataset	Train	Validation,	Test	Total
Real	DETFLY UAV	11067	3162	1582	15811
Synthetic	StyleGAN2	2100	600	300	3000
Total		13167	3762	1882	18811

5. Results and Discussion

Table 2: Detection Performance on DETFLY+ Synthetic Dataset

Model	Epochs	Recall	Precision	mAP@0.5	mAP@0.5-0.95
YOLOv8n-CA-GE	200	0.917	0.962	0.956	0.755
YOLOv11n-CA-GE	100	0.930	0.962	0.966	0.773

Table 3: Comparison of the suggested methodology with previous research

Reference	Dataset	Name of Model	mAP.5	mAP.5-.95
[13]	Det-Fly	YOLOv3	0.9756	0.7044
[13]	Det-Fly	YOLOv4	0.8341	0.5211
[13]	Det-Fly	YOLOv5	0.9784	0.7230
[13]	Det-Fly	YOLOv6	0.9048	0.627
[2]	Drone vs. Bird Detection Challenge + Synthetic Dataset	YOLOv5+ Tracker+ Track Boosting	0.794	
[6]	Self-built dataset	YOLOv5	0.904	
[4]	Drone-vs-Bird+ Real-world UAV+ Det-Fly+ MAV-VID+ Synthetic & Negative	YOLOv5 + CB + TB	0.866	
[4]	Drone-vs-Bird+ Real-world UAV+ Det-Fly+ MAV-VID	YOLOv5 + CB + TB	0.863	
[4]	Drone-vs-Bird+ Real-world UAV+ Det-Fly+ MAV-VID+ Synthetic & Negative	YOLOv5 + CB	0.855	
[4]	Drone-vs-Bird+ Real-world UAV+ Det-Fly+ MAV-VID	YOLOv5 + CB	0.854	
Proposed studies.	Det-Fly+Synthetic Dataset	Proposed model YOLOv8n	0.956	0.755
Proposed studies.	Det-Fly+Synthetic Dataset	Proposed model YOLOv11n	0.966	0.773

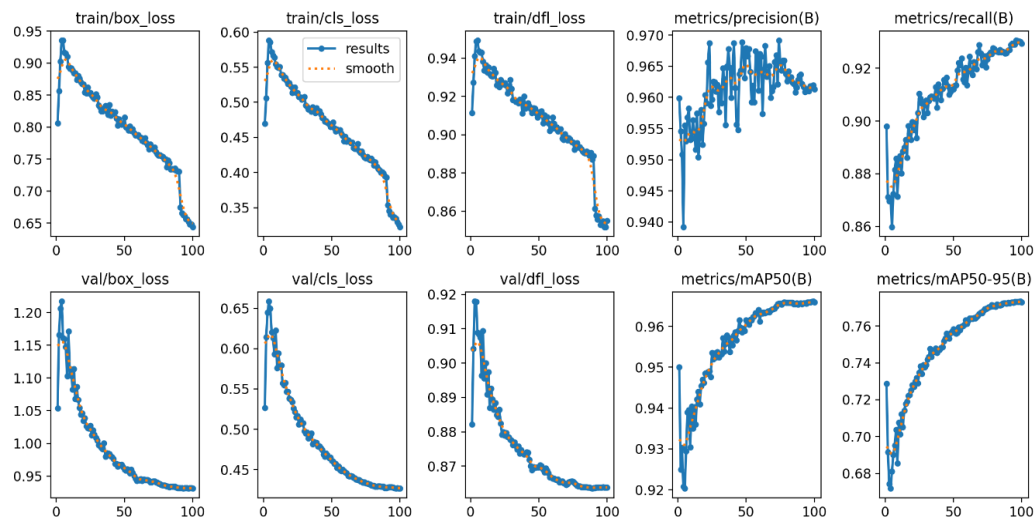


Figure 6. Yolov11n-CA-GE's training and validation performance curves demonstrate improvement in detection metrics and convergence of loss.

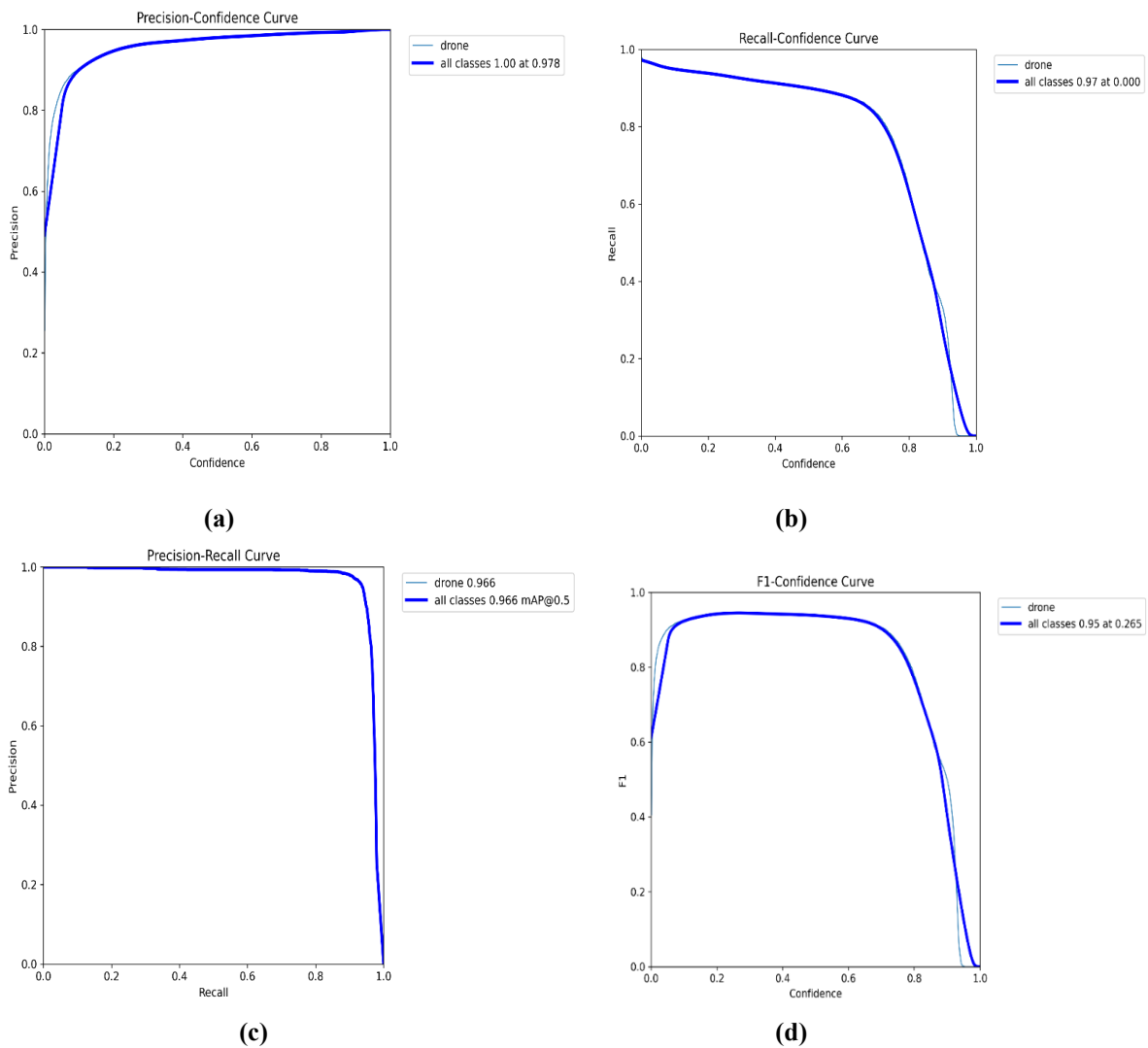


Figure 7. YOLOv11n-CA-GE's performance curves



Figure 8. Testing Images

According to the result, YOLOv11n model improved with Coordinate Attention and Gelan-E performs better than the YOLOv8n model using the same attention processes. *The YOLOv11n-Coordinate Attention-Gelan-E model improves $mAP@0.5$ (0.966 vs. 0.956), increases recall (0.930 vs. 0.917), and matches the baseline's precision (0.962).* These results shown in table 2. demonstrate how successful and efficient the suggested method is for real-time UAV identification and. table 3 compare with the previous studies. Figure 8 shows the result of testing is to identifying drone in various conditions.

6. Conclusion & Future Scope

Galen-E channel attention and Coordinate Attention techniques were included into an attention-enhanced UAV detection framework that uses YOLOv11n to detect small and distant airborne targets in difficult conditions. The resulting CA-YOLO model improves both spatial localization and channel-wise feature extraction; however, real-time applications might benefit more from maintaining a lightweight structure. YOLOv11n-CA-GE performs better than models based on YOLOv8n, according to experiments performed using the DETFLY dataset. However, certain challenges persist. It's still challenging to identify tiny drones or ones that blend in with complex backgrounds conditions in future studies, to reduce false negatives and improve detection in dynamic environments by utilizing temporal information from video sequences. Further improvements, like transformer-based attention mechanisms, sophisticated data augmentation methods, and multi-scale feature fusion with higher-resolution inputs, could further improve long-range detection capabilities. To gain further insight into the model's real-world performance, test it on embedded devices and in real-world situations.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics Statement

This study did not involve human or animal subjects and, therefore, did not required ethical approval.

Statement of Conflict of Interests

The authors declare no conflicts of interest related to in this study.

References

- [1] M. Sakib Bin Islam, M. E. Chowdhury, M. Hasan-Zia, S. B. A. Kashem, M. E. Majid, A. K. Ansaruddin Kunju et al., "VisioDECT: A novel approach to drone detection using CBAM-integrated YOLO and GELAN-E models," *Neural Computing and Applications*, vol. 37, no. 24, pp. 20181–20204, 2025.
- [2] F. C. Akyon et al., "Track boosting and synthetic data aided drone detection," in *Proc. 17th IEEE Int. Conf. Advanced Video and Signal Based Surveillance (AVSS)*, 2021.
- [3] F. C. Akyon, E. Akagunduz, S. O. Altinuc, and A. Temizel, "Sequence models for drone vs bird classification," *arXiv preprint arXiv: 2207*, 2022.
- [4] O. Eryuksel et al., "DroBoost: An intelligent score and model boosting method for drone detection," in *Int. Conf. Image Analysis and Processing*, Cham, Switzerland: Springer, 2022.
- [5] X. Zhai et al., "YOLO-Drone: An optimized YOLOv8 network for tiny UAV object detection," *Electronics*, vol. 12, no. 17, p. 3664, 2023.
- [6] B. Aydin and S. Singha, "Drone detection using YOLOv5," *Eng*, vol. 4, no. 1, pp. 416–433, 2023.
- [7] S. Zhou et al., "Improved YOLO for long range detection of small drones," *Scientific Reports*, vol. 15, no. 1, p. 12280, 2025.
- [8] C. Wang et al., "UAVSwarm dataset: An unmanned aerial vehicle swarm dataset for multiple object tracking," *Remote Sensing*, vol. 14, no. 11, p. 2601, 2022.
- [9] M. Misbah et al., "TF-Net: Deep learning empowered tiny feature network for night-time UAV detection," in *Int. Conf. Wireless and Satellite Systems*, Cham, Switzerland: Springer Nature, 2023.
- [10] A. Coluccia et al., "Drone vs. bird detection: Deep learning algorithms and results from a grand challenge," *Sensors*, vol. 21, no. 8, p. 2824, 2021.
- [11] M. L. Ali and Z. Zhang, "The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in object detection," *Computers*, vol. 13, no. 12, p. 336, 2024.
- [12] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [13] D. Kaur et al., "Analysis of vision-based air-to-air UAV detection using YOLO based architectures," *Robotics and Autonomous Systems*, vol. 193, p. 105074, 2025.
- [14] J. Wu et al., "A lightweight YOLOv5 optimization of coordinate attention," *Applied Sciences*, vol. 13, no. 3, p. 1746, 2023.
- [15] G. Lykou, D. Moustakas, and D. Gritzalis, "Defending airports from UAS: A survey on cyber-attacks and counter-drone sensing technologies," *Sensors*, vol. 20, no. 12, p. 3537, 2020.
- [16] S. Park et al., "Survey on anti-drone systems: Components, designs, and challenges," *IEEE Access*, vol. 9, pp. 42635–42659, 2021.
- [17] Y. Hu et al., "Object detection of UAV for anti-UAV based on improved YOLOv3," in *2019 Chinese Control Conf. (CCC)*, 2019.
- [18] R. Girshick, "Fast R-CNN," *arXiv preprint arXiv:1504.08083*, 2015.
- [19] S. Ren et al., "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [20] K. He et al., "Mask R-CNN," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017.
- [21] J. Redmon et al., "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [22] W. Liu et al., "SSD: Single shot multibox detector," in *European Conf. Computer Vision (ECCV)*, Cham, Switzerland: Springer, 2016.

-
- [23] T.-Y. Lin *et al.*, “Focal loss for dense object detection,” in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017.
 - [24] Y. Zheng *et al.*, “Air-to-air visual detection of micro-UAVs: An experimental evaluation of deep learning,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1020–1027, 2021.
 - [25] “Det-Fly Dataset.” GitHub - Det-Fly Dataset [Accessed: May 9, 2026].
 - [26] A. Paszke *et al.*, “Automatic differentiation in PyTorch,” 2017.
 - [27] R. Rombach *et al.*, “High-resolution image synthesis with latent diffusion models,” in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2022.
 - [28] M. Wen *et al.*, “Architectural facade design with style and structural features using stable diffusion model,” *Journal of Intelligent Construction*, vol. 2, no. 4, pp. 1–12, 2024.
 - [29] H. Ma and H. Zheng, “Text semantics to image generation: A method of building facades design based on stable diffusion model,” in *Int. Conf. Computational Design and Robotic Fabrication*, Singapore: Springer Nature, 2023.
 - [30] R. Khanam and M. Hussain, “YOLOv11: An overview of the key architectural enhancements,” *arXiv preprint arXiv:2410.17725*, 2024.
 - [31] P. Hidayatullah *et al.*, “YOLOv8 to YOLO11: A comprehensive architecture in-depth comparative review,” *arXiv preprint arXiv:2501.13400*, 2025.
 - [32] K. K. Pandey, S. K. Singh, and J. N. Singh, “Real-time helmet detection using synthetic,” in *Proc. Data Analytics and Management: ICDAM 2025*, vol. 3, p. 451, 2025.
 - [33] P. Basak, A. Mushtak, M. Ouda, S. F. Nobi, A. Hasan, and M. E. Chowdhury, “Improving pediatric trauma care: An automated system for wrist trauma detection using GELAN,” *Neural Computing and Applications*, vol. 37, no. 30, pp. 25095–25121, 2025.
 - [34] A. Agnihotri and N. Kohli, “XAI-AGUWEM: Explainable artificial intelligence-based attention guided uncertainty weighting ensemble model for the classification of COVID-19 and pneumonia in X-ray medical images,” *Recent Advances in Electrical & Electronic Engineering*, vol. 18, no. 7, pp. 862–884, 2025.
 - [35] A. Agnihotri and N. Kohli, “A Hybrid Deep Neural approach for multi-class Classification of novel Corona Virus (COVID-19) using X-ray images,” *2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, Gharuan, India, 2023, pp. 1-5, doi: 10.1109/InCACCT57535.2023.10141782.