Research Article GU J Sci, Part A, 11(3): 546-562 (2024) 10.54287/gujsa.1526979

Gazi University
Journal of Science
PART A: ENGINEERING AND INNOVATION
http://dergipark.org.tr/gujsa

Drone Detection Performance Evaluation via Real Experiments with Additional Synthetic Darkness

Furkan ORUÇ[1] H. Birkan YILMAZ[1*]

[1] Boğaziçi University, Department of Computer Engineering, İstanbul, Türkiye

Keywords

Drone Detection
Yolo
LSTM
Vision Transformers
Darkness

Abstract

Detecting drones is increasingly challenging, particularly when developing passive and low-cost defense systems capable of countering malicious attacks in environments with high levels of darkness and severe weather conditions. This research addresses the problem of drone detection under varying darkness levels by conducting an extensive study using deep learning models. Specifically, the study evaluates the performance of three advanced models: Yolov8, Vision Transformers (ViT), and Long Short-Term Memory (LSTM) networks. The primary focus is on how these models perform under synthetic darkness conditions, ranging from 20% to 80%, using a composite dataset (CONNECT-M) that simulates nighttime scenarios. The methodology involves applying transfer learning to enhance the base models, creating Yolov8-T, ViT-T, and LSTM-T variants. These models are then tested across multiple datasets with varying darkness levels. The results reveal that all models experience a decline in performance as darkness increases, as measured by Precision-Recall and ROC Curves. However, the transfer learning-enhanced models consistently outperform their original counterparts. Notably, Yolov8-T demonstrates the most robust performance, maintaining higher accuracy across all darkness levels. Despite the general decline in performance with increasing darkness, each model achieves an accuracy above 0.6 for data subjected to 60% or greater darkness. The findings highlight the challenges of drone detection under low-light conditions and emphasize the effectiveness of transfer learning in improving model resilience. The research suggests further exploration into multi-modal systems that combine audio and optical methods to enhance detection capabilities in diverse environmental settings.

Cite

Oruç, F., & Yılmaz, H. B. (2024). Drone Detection Performance Evaluation via Real Experiments with Additional Synthetic Darkness. *GU J Sci, Part A*, *11*(3), 546-562. doi:10.54287/gujsa.1526979

Author ID (ORCID Number)		Article Process	
0000-0001-6866-5285	Furkan ORUÇ	**Submission Date**	02.08.2024
0000-0002-4773-2028	H. Birkan YILMAZ	**Revision Date**	13.08.2024
		Accepted Date	19.09.2024
		Published Date	30.09.2024

1. INTRODUCTION

Drones are widely used unmanned aerial vehicles for defense systems and object detection technologies. Adam (2020) claim that the use of drone network applications has significantly increased. Drones serve various roles, including acting as sensors, providing network services, and facilitating delivery services, among other applications. We can observe and utilize drones in various ways, such as size and shape. Thus, there is always a requirement for such a technological system for drone detection (Moustafa et al., 2017). Traditional detection methods, such as radar and RF-based systems, often struggle with small, low-flying drones, particularly in cluttered or dark environments. This study's focus on enhancing drone detection models with transfer learning under varying levels of synthetic darkness is thus both timely and relevant.

Detecting drones in an accurate manner is crucial for preventing threats and ensuring public safety. Traditional radar and radio frequency (RF) detection methods, while effective in some scenarios, often struggle with the small size, low altitude, and low radar cross-section of modern drones (Moustafa et al., 2017). These limitations

*Corresponding Author, e-mail: birkan.yilmaz@bogazici.edu.tr

have contributed to the development of alternative detection techniques, particularly those based on computer vision and machine learning, which offer the potential for more precise and adaptable detection systems.

On the other hand, defense systems may observe malicious drone attacks that might cause vulnerabilities that might even lead to a war. To be able to handle this problem, the detection and surveillance of small drones under severe weather conditions are quite necessary. Drone detection systems are highly popular nowadays. Thus, analyzing and understanding when and how an unmanned aerial vehicle is approaching any target is crucial. As Moustafa and Jolfaei (2020) state, there might be developed a drone intrusion system with information sources, communication links, and autonomous control basis components. For this work, we would like to focus on information sources on the vision dataset among these components.

In recent years, models such as YOLO (You Only Look Once), Vision Transformers (ViT), and Long Short-Term Memory (LSTM) networks have shown significant impact in object detection tasks, including drone detection. These models leverage large datasets and complex architectures to learn and recognize patterns in visual data, making them well-suited for detecting drones in diverse conditions. However, the performance of these models can still be significantly impacted by factors such as darkness, which can obscure visual features and reduce detection accuracy.

This work will focus on the results and methods of drone detection using different models on a certain level of darkness. We also conducted our methodology by gathering real data as an experimental setup. After the collection part, we applied darkness procedures and compared the results in a precise manner.

The structure contains an introduction, literature review, and related work, our setup for experiments, results, and discussion. The literature review is about previous work done on drone detection and model performance comparison on dark images. In addition, the experimental setup talks about what equipment we used during our experiments, such as cameras, drones, and other electronic devices.

The Methodology section briefly is about model comparison and how we added darkness. After that, Yolov8, LSTM (long-short-term memory), and vision transformers are discussed. The results section explains the detection performance comparison and decrease-increase among different levels of darkness using Precision and Recall Curves and ROC (Receiver operating characteristic) Curves.

Experimental Results section presents the results of the experiments, comparing the performance of the original models against their transfer learning-enhanced versions across different levels of darkness. It includes detailed discussions of the models' Precision-Recall and ROC curves and the impact of darkness on detection accuracy.

Discussion section interprets the results, discussing the implications of the findings for drone detection in real-world scenarios. It also explores the strengths and limitations of the models tested, with particular attention to how transfer learning improves performance under challenging conditions.

Finally, our paper concludes by summarizing the key findings, emphasizing the advantages of transfer learning for drone detection under varying levels of darkness, and suggesting directions for future research, including the potential for multi-modal detection systems.

One of the novelties of the approach is the impact varying levels of synthetic darkness (20%, 40%, 60%, 80%) on drone detection performance, using models such as Yolov8-T, ViT-T (Vision Transformer with Transfer Learning), and LSTM-T (Long-Short Term Memory with Transfer Learning). Those methods will be mentioned in the methodology section. This approach will address a significant gap in the literature where the gradual impact of darkness on detection models is less explored. Additionally, we have a comprehensive dataset, which includes a mix of darkness levels providing more robust and challenging test environments as compared to other datasets used in previous studies. Transfer learning enhancements also provide an optimized pre-trained network for drone detection under dark conditions, which improves performance in scenarios where training data is limited or highly variable. Lastly, an insight to our model behavior such as CONNECT-M (which will be defined in methodology section later) challenges models more comprehensively than single-condition datasets, leading to a better understanding of model generalizability and robustness.

Lastly, This study fills this research gap by systematically evaluating the performance of state-of-the-art drone detection models—YOLOv8, Vision Transformers (ViT), and Long Short-Term Memory (LSTM) networks—under varying levels of synthetic darkness. By using experimental data and applying transfer learning to these models (resulting in YOLOv8-T, ViT-T, and LSTM-T), the study investigates whether pre-trained models can be fine-tuned to improve performance in low-light conditions, which will be addressing the current limitations of drone detection systems.

2. LITERATURE REVIEW

Khan et al. (2023) proposes a drone detection system, which is GAANet (Ghost auto anchor network) Using Yolov5. The images are under a high level of darkness at night and low visibility. They also implemented an auto anchor calculation for their model architecture for adding and removing convolutions. As compared to Yolov5, GAANet experiences overall precision and recall around 2.3 and 1.4 percent more respectively.

As Misbah et al. (2023) mentioned, we can observe that TF-net which is defined as TensorFlow network by authors is another approach that focuses on drone detection under nighttime. They propose a tiny feature network, which contains detection of unmanned aerial vehicles under night vision based on infrared images. TF-Net also focuses on detection with complex background images. Misbah et al. (2023) work with four different Yolo algorithms with different hyperparameters. Their conclusion shows that 95.7 precision results with 84 percent of mean absolute precision. Lastly, their detection threshold for intersection over union is 44.8 percent to consider the frame as true positive.

Not only darkness addition, but also severe weather conditions such as rainy, foggy, and stormy weather are important to make analysis for drone detection, which will provide challenges to the dataset. Methods of the authors of Munir et al. (2024) mention about the drone dataset with a complex and severe condition. Methods for benchmarks are Yolov8, Yolov5, and Faster R-CNN (Region based convolutional neural network) respectively. It is shown that Faster-RCNN and Yolov5 perform better when it comes to detecting unmanned aerial vehicle image frames with rainy conditions. In conclusion of their research, they claim average precision of their models for Yolov5, Yolov8, and Faster R-CNN are 69.3, 67.2, and 59.2 respectively

Yi et al. (2019) propose a drone detection and classification approach which will separate drones from birds or different backgrounds. Yolov2 is used as an object detection model. However, they also tried transfer learning from ImageNet under low level of light with additional data augmentation. They called this approach Yolov2 DarkNet. Lastly, average intersection and over union and recall are analyzed and compared as evaluation metrics. Andraši et al. (2017) defined their detection system as Unmanned Aerial Vehicles (UAVs) detection for safety and security. They challenged their image dataset with low visibility, urban environments, and the night. Their approach proposes a thermal infrared camera object detection. At the end, they claim that they tested the applicability of low cost long-wave infrared cameras for various examples of swarm Unmanned Aerial Vehicles in flight.

Svanström et al. (2022) developed an automatic detection of flying drones. However, they combine thermal infrared cameras and microphone sensors for multi-sensor drone detection. The solution of Svanström et al. (2022) also integrates an implementation of ADS-B receiver and a GPS receiver. However, the detection range is limited for certain distances. Their claim shows that the detection system can be more efficient in low-level of light and more robust to mitigate false detection with the fusion of audio and vision modalities.

Another tiny object detection method is proposed by Zhai et al. (2023). They developed the detection algorithm by Yolov8 similar to ours. First, a high-resolution detection head is added for small targets. Then, redundant network layers are cut to mitigate network parameters and improve detection performance. Thirdly, preprocessing steps such as multi-scale extraction were applied with SPD-Convolution. As compared to the baseline model of Yolov8, their method improved by 11.9 percent, 15.2 percent, and 9 percent in terms of precision, recall, and mean average precision. In addition, the deployment of their system indicates that the number of parameters decreased by 59.9 percent.

Ramadan et al. (2021) propose a novel approach of drone detection technology called Ad Hoc Network (FANET). It involves frameworks such as Recurrent Neural Networks (RNN) as a base. It includes gathering network data and using big data analytics to detect drone anomalies. The data collection part of this procedure is working inside of each Ad Hoc Network for intrusion drone detection. They claim that they set extensive ways of experiments based on various datasets to examine the efficiency of the proposed framework. Their results show that FANET outperforms RNN as a baseline detection model.

Finally, Jamil et al. (2022) built a drone detection system for malicious drone attacks using vision transformers, which is one of our models for this work. They proposed a vision transformer based on a framework with drone image dataset into fixed-size patches. In addition, linear embeddings and position embeddings are applied. Their proposed work is compared to deep convolutional neural networks (D-CNN) which revealed that their proposed model has experienced an accuracy of 98.3 percent overall.

3. MATERIAL AND METHOD

3.1. Setup Architecture

Figure 1 shows the rectangular area we used for our experiments. In addition, locations are calculated. The distance between first and last lines is 2800 cm, and each line is 700 cm. In addition, a line (drain) is detected through the right corner of the area as an inner side. The setup takes place as approximately 410cm with the drone flight experiments.

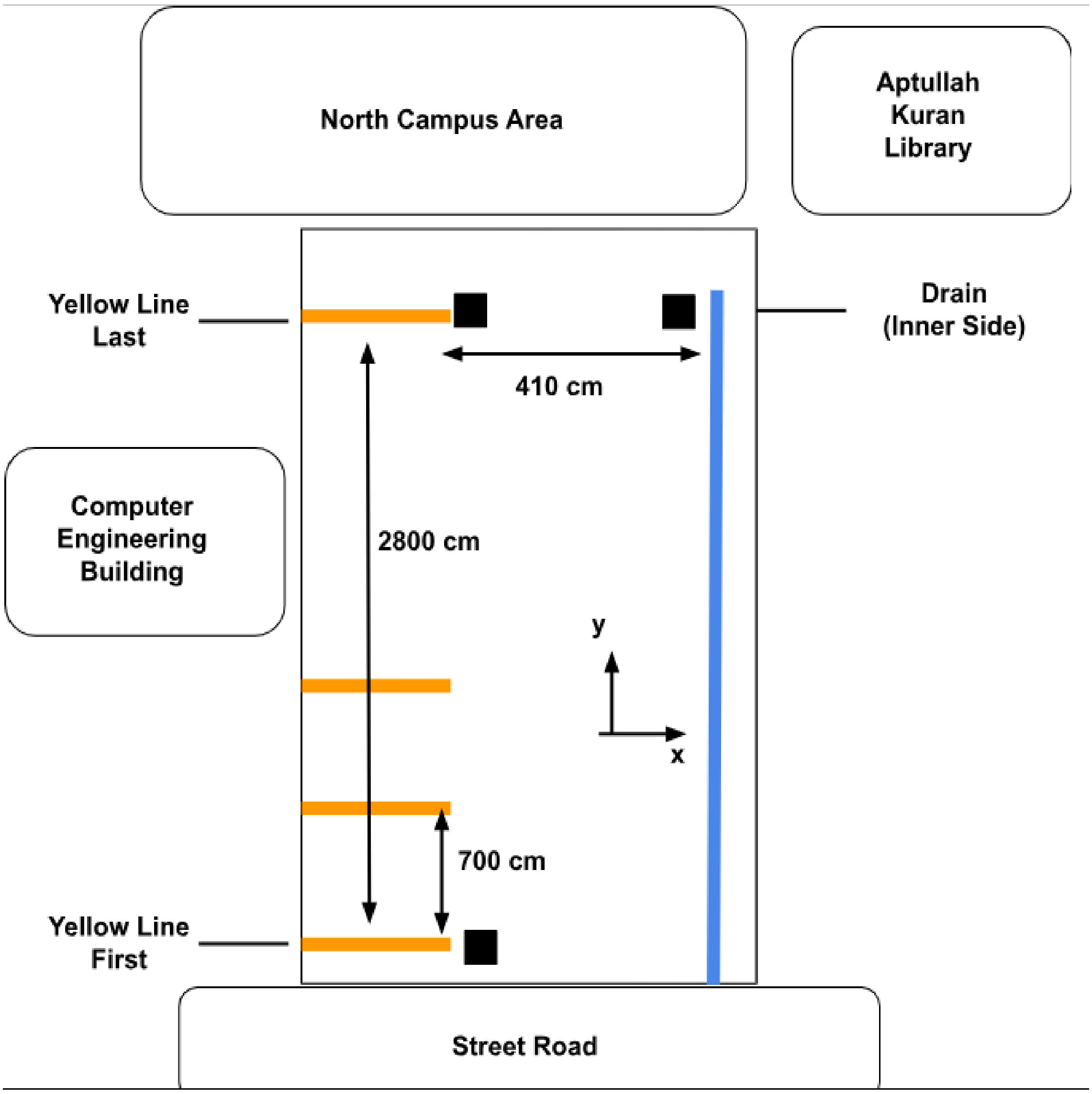


Figure 1. *Schematic map of the experimental area*

The first and last yellow line coordinates as latitudes and longitudes are demonstrated in Table 1. Also, latitude and longitude coordinates of the Inner side are given.

Figure 2 indicates our setup with different modalities. However, we will only focus on the vision modality part in this approach. There are parallel-located cameras for vision modality. One computer is designed as a controller. Two others are showing the footage for left and right cameras simultaneously. The closest point between the left and right tripods is 87 cm. In addition to setup and data collection, Dell Precision t3600, 64 GiB, NVIDIA GA102GL RTX A5000 is used for data analysis, learning, visualization and testing. The following features of the setup are as follows:

- Processor: 12th Gen Intel Core i9-12900 with 24 cores, providing the necessary computational power for real-time data processing and model training.
- Disk Capacity: 4.5TB of disk capacity, ensuring sufficient space for storing high-resolution images, video data, and trained model weights.
- OS Name and Type: Ubuntu 22.04.3 LTS, 64-bit,
- GNOME Version: 42.9,
- GPU: NVIDIA GA102GL RTX A5000, a high-performance graphics card essential for deep learning tasks, enabling the acceleration of model training and inference processes.

Table 1. *Inner Side and Yellow Line Coordinates*

Label	**Latitude**	**Longitude**
Yellow Line - Last	41.0857753	29.0437393
Yellow Line - First	41.08570841	29.04412709
Drain (Inner Side)	41.08581119	29.04373023

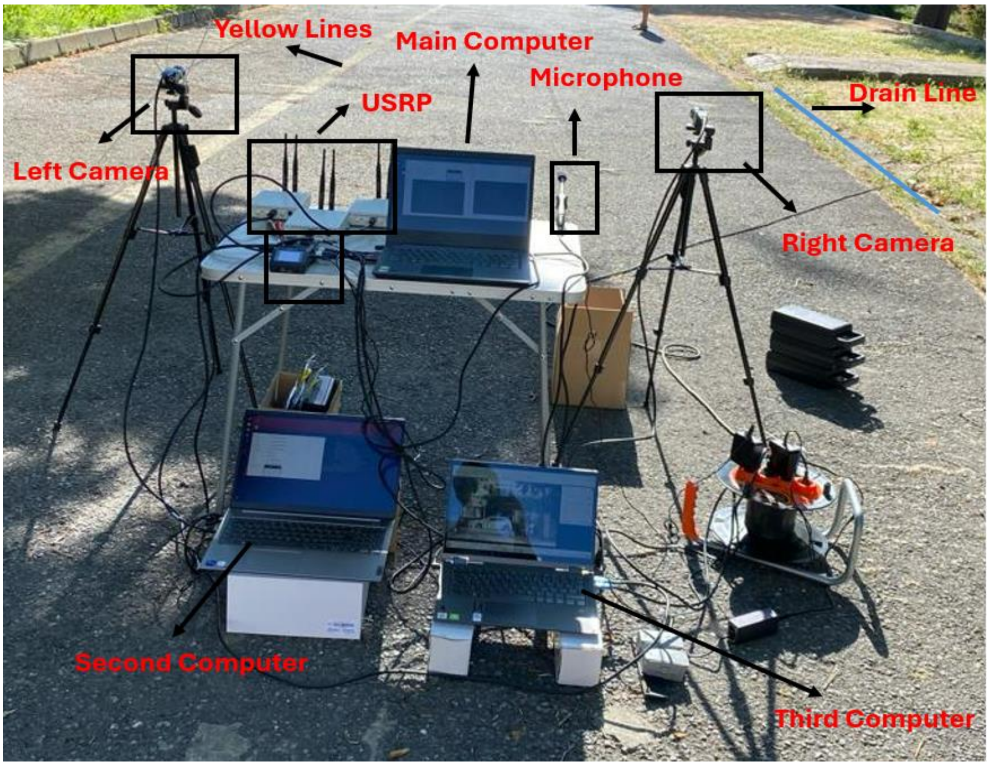
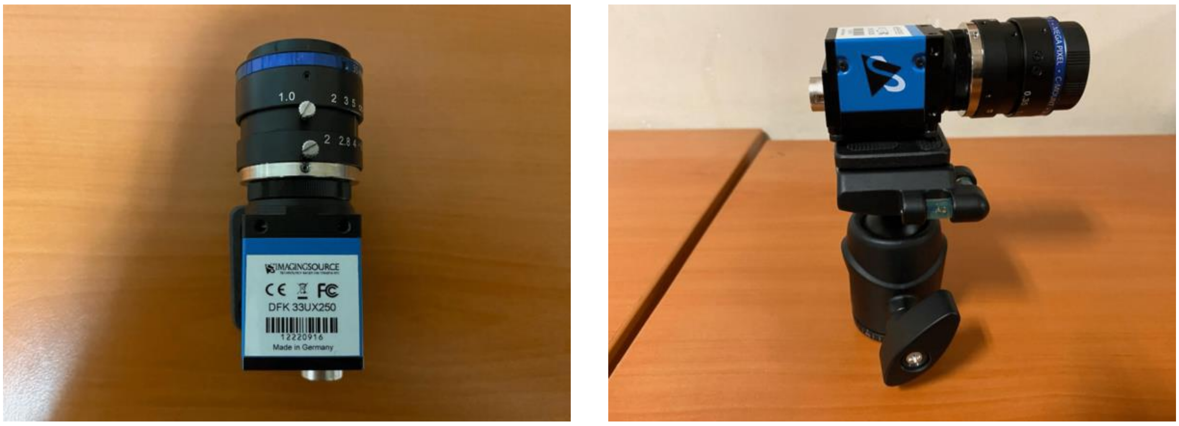


Figure 2. *Designed setup outside*

Figure 3 shows the industrial cameras used to obtain drone footage for image dataset. These cameras are placed on the right and the left sides of the table of the setup. Further features of our cameras are as follows: USB 3.0 Interface, 2/3-inch Sony CMOS Pregius IMX250 Sensor, 2,448×2,048 (5 MP) Resolution, Global Shutter, Trigger and I/O Inputs, 29x29x43 mm width, height, and length.

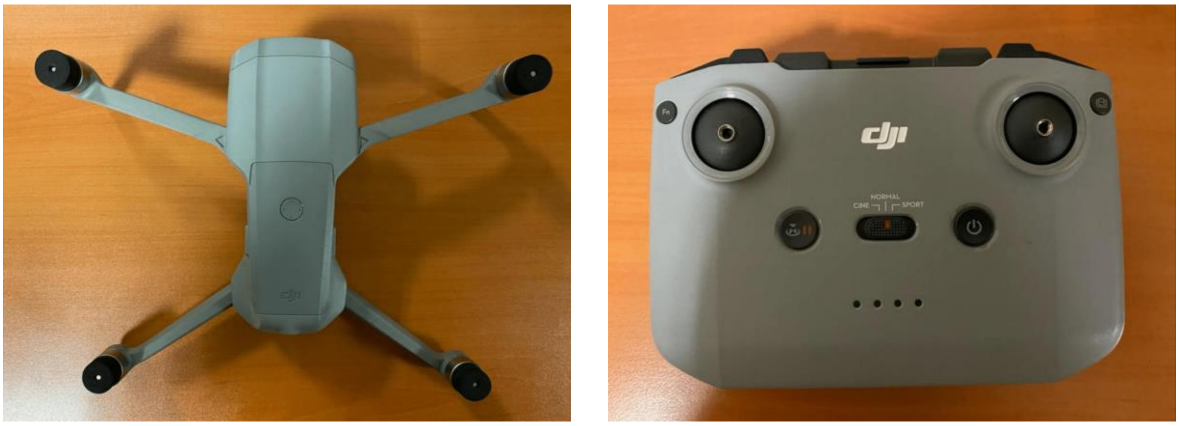


(a) (b)

Figure 3. a) *Image source camera,* ***b)*** *Image source camera with fixing device*

The model and several of the drone flew during the experiment is shown in Figure 4. DJI Mavic Air 2 is used with its control device and four propellers. Device controller has three options: Cine, Normal and Sport 5 modes. We mostly completed our experiments in normal mode. In addition, the control device is used for up, down, left, right, back, and forward moves. Weight of the drone is 570g and the diagonal distance is 302mm. There is also a gimbal with 3-axis (tilt, roll, and pan) which we did not use and record any frame from the drone's point of view. Further features of our drone for experiments are as follows:

- Battery Capacity: 2500 mAh,
- Max Flight Time: 34 minutes,
- Control Modes: The drone was primarily operated in Normal Mode for the experiments, although Cine and Sport modes were also available on the controller.
- Angular Velocity: 250 ◦/s (N Mode),
- Propellers: Quick release, low noise, folding,
- Gimbal: A 3-axis (tilt, roll, and pan) gimbal was included, although it was not utilized in this study as the focus was on the drone's external detection rather than its onboard camera.



(a) (b)

Figure 4. a) *DJI Mavic Air 2 drone,* ***b)*** *Controller of DJI Mavic Air 2 drone*

3.2. Dataset and Structure

The dataset consists of 28 .avi videos having the duration ranging between 5-20 minutes with 30 fps and in total 518 700 frames. For each experiment, two videos were created at the same time, both left and right cameras were recording during the process from different angles. While half of them belong to the left, the others belong to the right camera.

The images usually contain one drone in each frame with a complex background with external objects such as buildings, a gate, and trees. Figure 5 shows a randomly selected frame among the other frames of the dataset. In Figure 5, there is a drone on the left part of the gate. The dataset consists of 28 videos of which 14 of them are recorded from left, and the remaining are recorded from right camera. The video format has been recorded as AVI format then, we obtained 30 frames per second since the cameras have 30 fps settings.

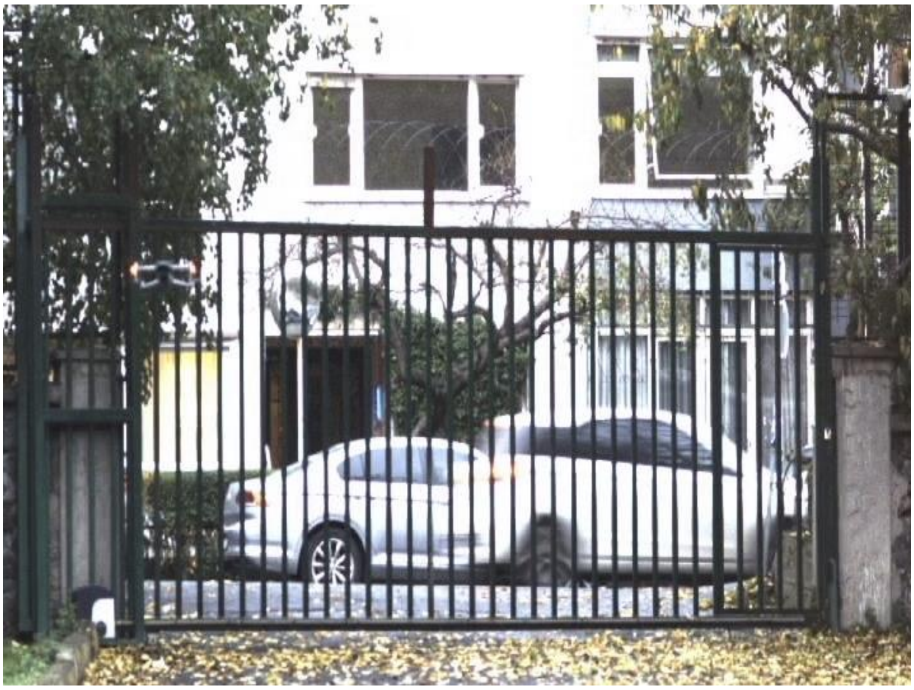


Figure 5. *Sample image from the dataset*

Our methodology uses this data architecture with three different models: Yolov8, Vision Transformer (ViT), and Long-short term memory (LSTM). However, the detection procedure is not only evaluated in one dataset. There are other datasets with combined or single darkness addition. Entire datasets used for this approach are defined as follows:

- CONNECT: Original data without darkness,
- CONNECT-D20 Dataset: 20 percent darkness addition of entire data,
- CONNECT-D40 Dataset: 40 percent darkness addition of entire data,
- CONNECT-D60 Dataset: 60 percent darkness addition of entire data,
- CONNECT-D80 Dataset: 80 percent darkness addition of entire data.

Lastly, we defined a mixture of darkness levels as the CONNECT-M dataset. CONNECT-M dataset contains the followings:

- 30 percent of the Data: 20 percent darkness addition,
- 30 percent of the Data: 40 percent darkness addition,
- 20 percent of the Data: 60 percent darkness addition,
- 20 percent of the Data: 80 percent darkness addition.

3.3. Yolov8-T

Authors of Reis et al. (2023) propose a generalized method for real-time object detection including flying objects such as drones. This state-of-art method is called Yolov8, which we also conducted for our research. The authors provide an in-depth explanation of the new architecture and functionality that YOLOv8 has adapted. According to evaluation metrics mAP50-95 with the 50 fps 1080p videos. The algorithm gets an average of 0.835 mAP50-95 as a result.

Furthermore, Yolov8 uses Feature Pyramid Network (FPN) and Path Aggregation Network (PAN). It is important to note that they make our annotation part easier. The first parameter of Yolo architecture is input image (Reis et al., 2023). However, PAN architecture utilized different layers from different levels as part of the network architecture (Li et al., 2023).

Li et al. (2023) also states the backbone of YOLOv8 is a deep Convolutional Neural Network (CNN) designed to extract hierarchical features from the input image. YOLOv8 typically uses CSPDarknet (Cross Stage Partial Darknet) as its backbone, which is known for its efficient feature extraction and reduction of computational complexity. YOLOv8 also utilizes a Feature Pyramid Network (FPN) combined with a Path Aggregation Network (PAN). This neck architecture is crucial for fusing feature maps from different layers, allowing the model to detect objects at multiple scales (Minderer et al., 2022).

Those approaches make the Yolov8 approach capture tiny and small objects. In addition, since we work with additional layers with parameter arrangement, our model structures for Yolov8 are defined as Yolov8-T. The term shows the yolov8 model with transfer learning applied. Following parameters have been used for model testing and parameter tuning: Input Image Size, Batch Size, Number of Epochs, Learning Rate, Anchor Boxes, Confidence Threshold. Various image sizes were tested, ranging from 320x320 to 640x640 pixels. The final choice of 416x416 was selected as it provided a balance between detection accuracy and computational efficiency, ensuring that small drones were detectable without overwhelming the GPU. Batch sizes of 8, 16, and 32 were tested. The model was trained with a batch size of 16, which was found to offer a good trade-off between stability and the ability to utilize the GPU effectively without memory overflow issues. As a tuning process, the model was initially trained for 50 epochs, with performance monitored on the validation set. Then, Learning rates were experimented with in the range of 0.001 to 0.0001. A learning rate of 0.0005 was chosen as it allowed the model to converge properly.

3.4. Vision Transformer-T (ViT-T)

Minderer et al. (2022) introduced a novel technique for detecting objects on multiple scales by integrating both textual and visual elements using Vision Transformers. This technique is officially termed "Open-Vocabulary Object Detection." It employs a conventional vision transformer that encompasses both an image encoder and a text encoder. A token pooling and projection layer is typically used in this model to derive image classification embeddings The core of ViT consists of multiple transformer encoder layers, each containing multi-head self-attention mechanisms and feed-forward neural networks.. However, in our specific application, where we focus solely on single-object scenarios, we have omitted this aspect of the model. The process begins with the utilization of a pre-trained COCO model in tandem with a text encoder, resulting in a combined image-text output. This is subsequently adapted for open-vocabulary object detection by bypassing the token pooling step in the image-encoding phase. For object detection, we have employed specific textual inputs. The final classification head is a simple Multilayer Perceptron with a softmax output, responsible for determining the class probabilities and the presence of a drone in each patch.

The output for object detection can be achieved using any one of the four options provided, as this enhances the likelihood of successfully detecting an object in a single attempt. The specific hyper parameters implemented in this method are as follows:

- focal loss with α γ= 0.3 and γ= 2.0,
- Adam optimizer with $\beta1$= 0.9 and $\beta2$ = 0.999,which improved the model’s ability to adapt to different levels of darkness,

- Cosine learning rate, initial learning rates were set between 0.0001 and 0.001. A learning rate of 0.0002 was chosen,
- Text encoder input length to 16 characters,
- Dropout rates of 0.1, 0.3, and 0.5 were tested. A dropout rate of 0.3 was selected, providing a good level of regularization while maintaining model performance, especially under darker conditions where overfitting could lead to significant performance drops.

And this specific model with hyperparameters tuned and transfer learning applied is called ViT-T.

3.5. Long-short Term Memory-T (LSTM-T)

The concept of Long Short-Term Memory (LSTM) was first introduced by Hochreiter and Schmidhuber (1997). They described LSTM as a method to truncate gradients, enabling the model to minimally bridge time intervals using specialized units. Over time, contemporary LSTM models have evolved, incorporating additional features such as dropout layers, dense layers, and various parameter enhancements. In our drone detection model, we have integrated both dropout and dense layers.

The dropout layer, as explained by Goodfellow et al. (2016), serves as a regularization strategy in artificial neural networks to address the problem of overfitting. It works by randomly omitting certain output features during the training phase. Additionally, they also describe a dense layer as a fully connected layer, where every input is linked to each output through a learnable weight (Goodfellow et al., 2016). In addition, Hochreiter and Schmidhuber (1997) also propose that an input vector described as x, and the output described as y of a dense layer can be calculated as:

$$y = \phi(Wx + b) \tag{1}$$

where W is the weight matrix, b is the bias vector, φ is the activation function. The equation shows that the matrix multiplication ‘Wx’ calculates the weighted sum of the inputs, and then biases ‘b’. In addition, since we work with additional layers with parameter arrangement, our model structure for LSTM is defined as LSTM-T. The term shows that long-short term memory model with transfer learning applied.

Our LSTM-T model uses the following parameters for tuning and performance development:

- Sequence length: 9, sequence lengths of 5, 9, and 15 were tested. A sequence length of 9 was chosen, as it provided enough temporal context to detect drones moving across frames without overwhelming the model with too much data.
- Number of layers: 5, LSTM - Dropout - Dense - Dropout - Dense, models with 2, 3, and 5 layers were tested. A 3-layer LSTM network was selected,
- Activation function: ReLu for LSTM and sigmoid for Dense layer,
- Optimizer: Adam, Loss: binary cross-entropy,
- Hidden Size: 256,hidden sizes of 128, 256, and 512 units were tested. A hidden size of 256 was optimal, allowing the model to capture sufficient information,
- Number of Classes: 1,
- Learning Rate 0.0001,
- Batch Size: 10,
- Number of Epochs: 150.

3.6. Key Evaluation Metrics

We used several key metrics to evaluate the performance of drone detection models under varying conditions of darkness.

Firstly, precision is the ratio of true positive detections to the sum of true positives and false positives, which measures the accuracy of positive predictions made by the model and Hochreiter and Schmidhuber (1997) uses the precision as

$$Precision\ = True\ Positives\ (TP)\ /\ (True\ Positives\ (TP)\ +\ False\ Positives\ (FP)) \tag{2}$$

Also, they used the metric Recall as

$$Recall\ = True\ Positives\ (TP)\ /\ (True\ Positives\ (TP)\ +\ False\ Negatives\ (FN)) \tag{3}$$

In terms of ROC (Receiver Operator Characteristic) curve calculations, we used True Positive Rate (TPR) or sensitivity and False Positive Rate (FPR). TPR is defined as

$$TPR\ = True\ Positives\ (TP)\ /\ (True\ Positives\ (TP)\ +\ False\ Negatives\ (FN)) \tag{4}$$

Similarly, False Positive Rate (FPR) is defined as

$$FPR\ = False\ Positives\ (FP)\ /\ (False\ Positives\ (FP)\ +\ True\ Negatives\ (TN)) \tag{5}$$

which provides a single value representing the overall performance of the model across all classification thresholds. AUC-ROC value indicates better model performance, with being perfect and 0.5 being equivalent to random guessing. In terms of justification, precision is crucial in applications where false positives can lead to significant consequences, such as unnecessary defensive actions or false alarms in security systems. Precision also measures true positive detections among all positive detections made by the model, indicating how many of the detected drones are actual drones rather than background noise or other objects. Recall is also chosen because a metric to be balanced to ensure results with accuracy and AUC tables. Lastly, AUC-ROC is useful when comparing different models or configurations and that is quite important in understanding the trade-offs between precision and recall.

4. EXPERIMENTAL RESULTS

The collected experimental data is expanded with synthetic darkness, and we tested our trained network on different datasets. For evaluation, we consider precision-recall curves and ROC curves. First, we need to see the differences between standard Yolo, Vit, and LSTM algorithms as compared to Yolov8-T, ViT-T, and LSTM-T in order to measure the development.

Table 2 shows the difference between three original algorithms and three transfer learning applied algorithms on CONNECT dataset.

Table 2. *AUC scores of original and transfer learning applied models on CONNECT dataset*

Dataset	Method	AUC (Precision-Recall)	AUC (ROC Curve)
CONNECT	Yolov8-T	0.80	0.78
CONNECT	ViT-T	0.73	0.74
CONNECT	LSTM-T	0.72	0.71
CONNECT	Yolov8	0.56	0.52
CONNECT	ViT	0.55	0.53
CONNECT	LSTM	0.59	0.56

Table 2 shows that The AUC score for Yolov8-T is 0.80, significantly higher than Yolov8’s 0.56. Similarly, the AUC (ROC Curve) for Yolov8-T is 0.78 compared to Yolov8’s 0.52. This clearly indicates that the application of transfer learning dramatically improves the model’s ability to accurately detect drones, enhancing both precision and recall, as well as overall classification performance. In addition, ViT-T also shows a substantial improvement over ViT, with a Precision-Recall AUC of 0.73 versus 0.55 for the original model. The ROC AUC improves from 0.53 to 0.74. These results suggest that transfer learning has also a strong positive effect on the performance of Vision Transformers in drone detection tasks. Lastly, The LSTM-T model outperforms the original LSTM model, with an AUC (Precision-Recall) of 0.72 compared to LSTM's 0.59. The AUC (ROC Curve) shows a similar improvement, with LSTM-T scoring 0.71 versus LSTM's 0.56. This improvement underscores the benefit of transfer learning in recurrent neural networks like LSTM for this application. There is an improvement on the CONNECT Dataset with all three different methods. Therefore, we are going to apply the same logic to different levels of darkness with added datasets.

4.1. Results of CONNECT Dataset

Figure 6 shows the results of Yolov8-T, ViT-T, and LSTM-T according to Precision-Recall Curves and ROC Curves. The results are 0.80, 0.73, and 0.72 for the Precision Recall area under the curve.

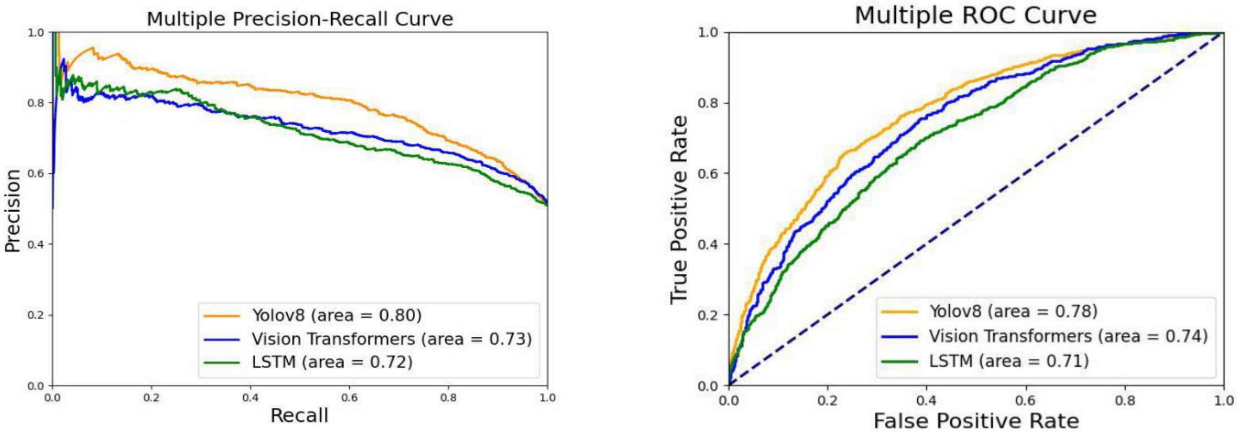


Figure 6. *Precision-Recall and ROC curves under no additional darkness.*

4.2. Results of CONNECT-D20 Dataset

Figure 7 indicates the ROC Curves and Precision-Recall Curve of the following models with 20 percent darkness added. Yolov8-T, ViT-T, and LSTM-T's performances are shown with AUC scores (Area under the curve) which are 0.69, 0.68, and 0.63, respectively. We observe there is a performance reduction as compared to the CONNECT Dataset. Similarly, while precision-recall curves of Yolov8-T, ViT-T, and LSTM-T are similarly approached in terms of AUC, the AUC score of ViT-T is closer to Yolov8-T.

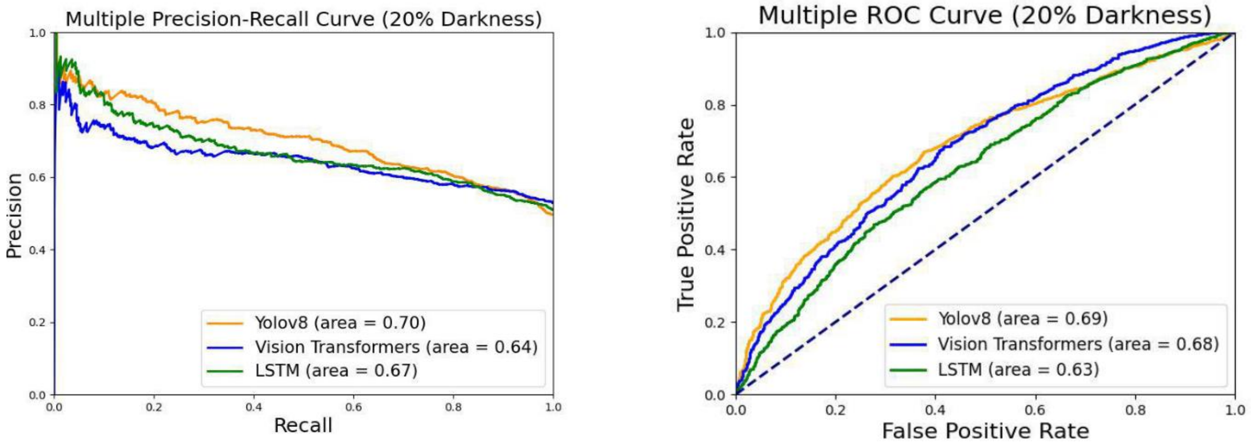


Figure 7. *Precision-Recall and ROC curves on the CONNECT-D20 dataset*

4.3. Results of CONNECT-D40 Dataset

Figure 8 indicates the ROC Curve and Precision-Recall Curves of the following models with 40 percent darkness added. The areas under the curve are 0.68, 0.66, and 0.64, respectively. We can observe that ViT-T models experienced a decrease in terms of AUC score, unlike LSTM-T and Yolov8-T. In addition, the AUC (Area under curve) of the ROC curve for LSTM-T has increased by 0.1 compared to 20 percent darkness.

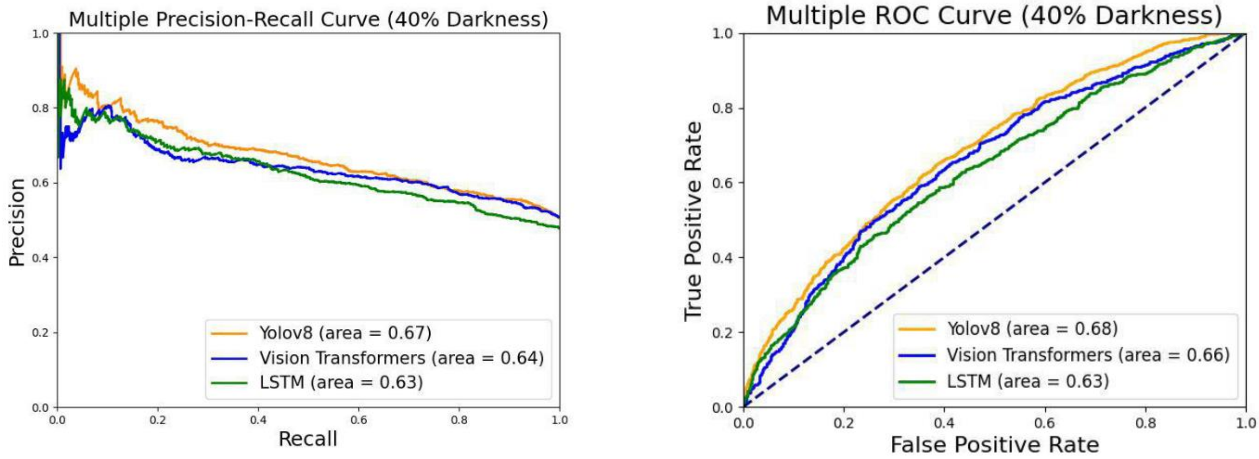
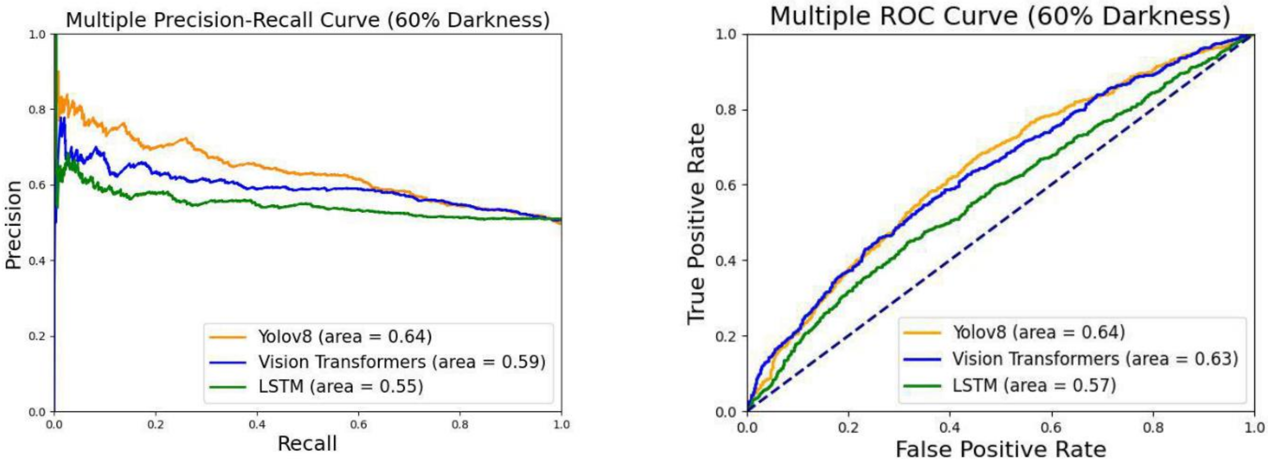


Figure 8. *Precision-Recall and ROC curves on the CONNECT-40 dataset*

4.4. Results of CONNECT-D60 Dataset

Figure 9 displays the calculated areas under curves, using True Positive Rate and False Positive Rates. These rates are 0.64 for Yolov8-T, 0.63 for ViT-T, and 0.57 for LSTM-T. It's noted that the LSTM-T exhibited a more consistent decrease than both Yolov8-T and Transformers when subjected to an increase in darkness by 20 percent. Furthermore, Figure 9 reveals that in terms of precision and recall values, ViT-T performed less effectively compared to their accurate positive and false favorable rates.



(a) Precision Recall Curves (b) ROC Curves

Figure 9. *Precision-Recall and ROC curves on the CONNECT-60 dataset*

4.5. Results of CONNECT-D80 Dataset

At this increased level of darkness, The ROC Curve outcomes depicted in Figure 10 indicate a minor decline in performance across all three deep learning models. Specifically, the areas under the curve for Yolov8-T, ViT-T, and LSTM-T are 0.56, 0.53, and 0.51, respectively. Moreover, in the Precision-Recall Curve, the areas under the curves for these models are 0.58, 0.54, and 0.51, respectively. These findings suggest that as darkness intensifies, the effectiveness of the models' accurate favorable rates deteriorates. It is observed that the detection results are nearing those of a randomized model with 80 percent darkness addition.

In summary, the addition of darkness leads to a decline in the AUC performance for all three models, although LSTM-T shows some exceptions at specific values of the constant darkness level. The darkest samples show the most significant drops in performance metrics.

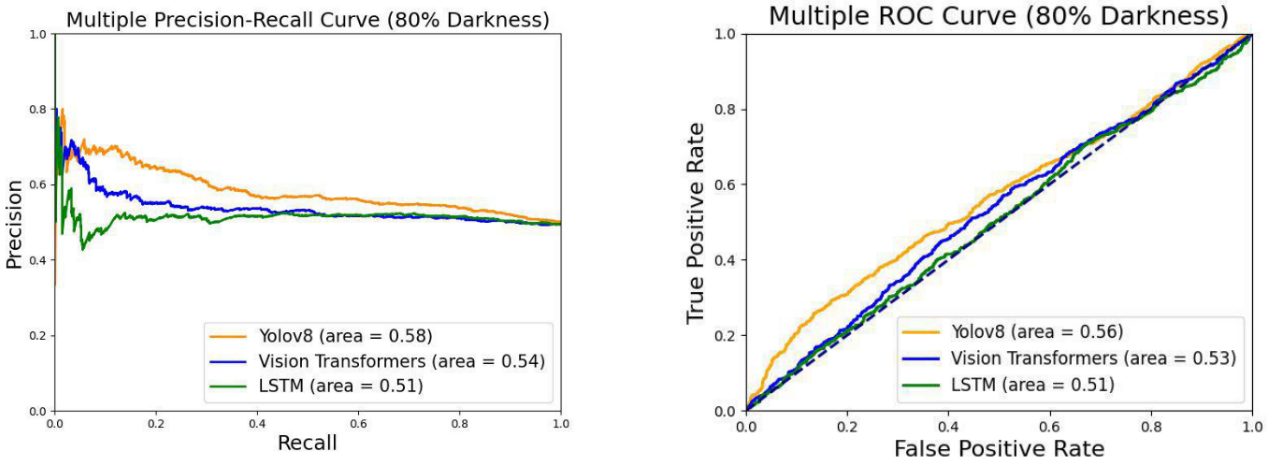


***Figure 10.** Precision-Recall and ROC curves on the CONNECT-80 dataset*

4.6. Results of CONNECT-M Dataset

The CONNECT-M dataset presents more significant challenges than the CONNECT dataset, as it incorporates images with varying degrees of darkening to mimic nighttime conditions. Specifically, 30 percent of its data is darkened by 20 percent, another 30 percent by 40 percent, 20 percent by 60 percent, and the remaining by 80 percent. Figure 11 illustrates that different darkness levels impact the performance of the models. Furthermore, there might be a decline in model performance when encountering new image frames with unfamiliar darkness levels that are not seen during training. To elaborate, when the dataset includes a darkened frame from one video, another frame with the same level of darkness is selected from a different video subset to address potential issues like over-fitting.

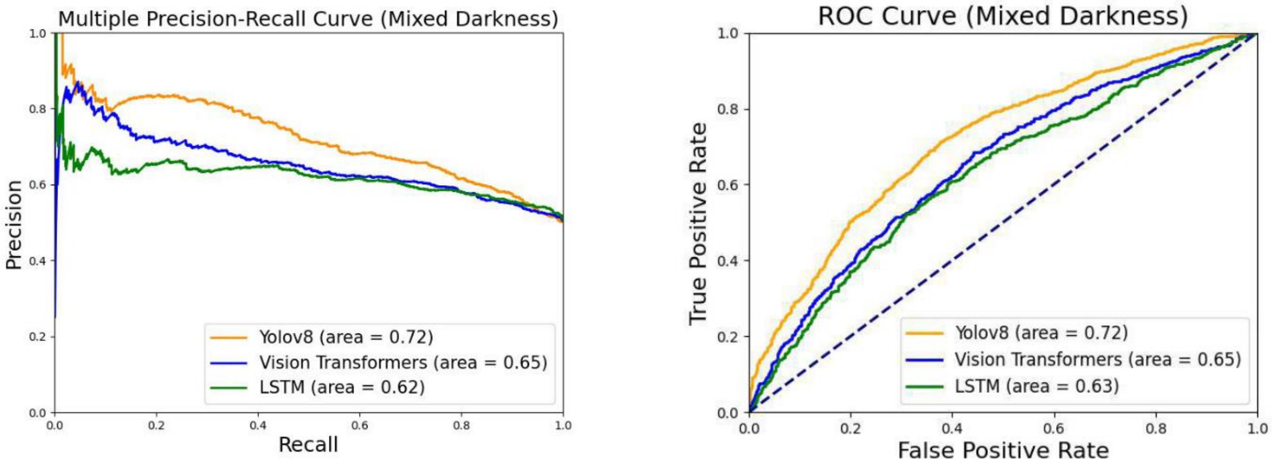


***Figure 11.** Precision-Recall and ROC curves on the CONNECT-M dataset*

As seen in Table 3, all-AUC scores of the Precision-Recall Curve and ROC Curve of Yolov8-T, ViT-T, and LSTM-T vision models. They are shown for each darkness-added variation of the CONNECT dataset and the CONNECT dataset itself.

Table 3. *AUC scores of vision models on all datasets*

Dataset	Method	AUC (Precision-Recall)	AUC (ROC Curve)
CONNECT	Yolov8-T	0.80	0.78
CONNECT	ViT-T	0.73	0.74
CONNECT	LSTM-T	0.72	0.71
CONNECT-D20	Yolov8-T	0.70	0.69
CONNECT-D20	ViT-T	0.64	0.68
CONNECT-D20	LSTM-T	0.67	0.63
CONNECT-D40	Yolov8-T	0.67	0.68
CONNECT-D40	ViT-T	0.64	0.66
CONNECT-D40	LSTM-T	0.63	0.63
CONNECT-D60	Yolov8-T	0.64	0.64
CONNECT-D60	ViT-T	0.59	0.63
CONNECT-D60	LSTM-T	0.55	0.57
CONNECT-D80	Yolov8-T	0.58	0.56
CONNECT-D80	ViT-T	0.54	0.53
CONNECT-D80	LSTM-T	0.51	0.51
CONNECT-M	Yolov8-T	0.72	0.72
CONNECT-M	ViT-T	0.65	0.65
CONNECT-M	LSTM-T	0.62	0.63

We can say that across all models (Yolov8-T, ViT-T, LSTM-T), there is a consistent decrease in the AUC scores as the level of darkness increases. For instance, the AUC (Precision-Recall) for Yolov8-T drops from 0.80 in the CONNECT dataset (no additional darkness) to 0.58 in the CONNECT-D80 dataset (80% darkness). This trend is similarly observed in both the Precision-Recall and ROC Curve AUC scores for all models. However, Yolov8-T generally outperforms ViT-T and LSTM-T across all datasets. For example, in the CONNECT dataset, Yolov8-T achieves an AUC (Precision-Recall) of 0.80, while ViT-T and LSTM-T achieve 0.73 and 0.72, respectively. This suggests that Yolov8-T is more robust in maintaining performance under varying levels of darkness compared to the other models. Lastly, The LSTM-T model shows more variability and less robustness in its performance compared to Yolov8-T and ViT-T. For instance, the AUC (Precision-Recall) for LSTM-T in the CONNECT-D60 dataset is 0.55, which is significantly lower than the 0.64 achieved by Yolov8-T in the same dataset. This suggests that LSTM-T may require more fine-tuning or additional modifications to handle varying darkness levels effectively.

5. DISCUSSION

One of the most striking findings is the consistent decline in model performance as the level of darkness increases. This trend is evident across all three models—Yolov8, Vision Transformer (ViT), and Long Short-Term Memory (LSTM)—and is particularly pronounced at the highest levels of darkness (80%). The decrease in AUC (Precision-Recall and ROC) scores suggests that these models, while effective in well-lit conditions, struggle to maintain accuracy in low-light environments. This performance degradation likely stems from the reduced visibility of drones in darker images, which challenges the models' ability to discern relevant features and make accurate predictions.

Another notable observation in our analysis is the varying performance of specific deep learning models at different levels of darkness. For instance, the LSTM-T model exhibited lower performance under 20 percent darkness compared to 40 percent darkness. Specifically, the areas under the curves were 0.67 for CONNECT-D20 and 0.63 for CONNECT-D40 datasets. This suggests that changes may influence the tuning of model parameters in the pixel composition of the frames. However, it is interesting to note that LSTM-T's performance on the ROC curve remained consistent at 0.63 for both 20 percent and 40 percent darkness levels. Since recall and actual positive rate are considered identical in this context, the calculation of the false positive rate could be impacting the area under the curve to a certain extent.s.

Among the models tested, Yolov8-T consistently outperforms the others, demonstrating superior accuracy and robustness across all darkness levels. This can be attributed to several factors, including the advanced architecture of Yolov8, which is optimized for real-time object detection, and the benefits of transfer learning, which allows the model to build on existing knowledge rather than starting from scratch. Yolov8-T’s ability to maintain high AUC scores even as darkness increases indicates its strong potential for practical deployment in scenarios where lighting conditions are variable and unpredictable.

5.1. Limitations

The primary focus on synthetic darkness, while useful for controlled experiments, may not fully capture the complexity of real-world low-light environments, where factors such as shadows, reflections, and varying light sources can further complicate detection tasks. Future studies should consider incorporating more diverse datasets that include naturally dark conditions to better simulate real-world scenarios.

5.2. Future Research

While the transfer learning has been shown to significantly improve model performance, it is also important to explore other advanced techniques, such as self-supervised learning or generative adversarial networks (GANs), which could provide additional benefits by enabling models to learn more effectively from limited data. These techniques, combined with ongoing improvements in model architectures, could lead to even more robust and reliable drone detection systems. We can also say that the variability observed in the LSTM-T model suggests that further research is needed to optimize recurrent neural networks for static image detection tasks. Investigating hybrid models that combine the strengths of different architectures (e.g., CNN-LSTM or Transformer-LSTM hybrids) could lead to new insights for effective solutions for drone detection in challenging conditions.

6. CONCLUSION

In this study, we have undertaken a thorough comparison of model performance, focusing on object detection through optical methods, particularly under conditions of added darkness. The results demonstrate that models with transfer learning, specifically Yolov8-T, ViT-T, and LSTM-T, consistently outperform their non-enhanced counterparts (Yolov8, ViT, LSTM) across all conditions. For instance, Yolov8-T achieves an AUC (Precision-Recall) of 0.80 on the CONNECT dataset, compared to 0.56 for Yolov8. This highlights the substantial benefit of transfer learning in improving model robustness and accuracy.. For drone detection in images, we utilized Yolov8-T, vision transformers (ViT-T), and LSTM-T. The findings offer valuable insights into the advantages and constraints of vision-based modalities when darkness is introduced. In addition, the research highlighted the effects of environmental factors, such as darkness, on detection performance in multi-modal systems under real-world conditions By means of observed trends, across all models, there was a consistent decline in performance metrics as the level of synthetic darkness increased. This trend was most pronounced in the CONNECT-D80 dataset because of high level of darkness. In addition to that, YOLOv8-T consistently outperformed ViT-T and LSTM-T across all levels of darkness. Because transfer-learning method was more robust with pre-trained processes and fine-tuning parameters Also, we can conclude that Yolov8-T is the most robust model. Among the models tested, Yolov8-T consistently shows the highest performance across all datasets, including under varying levels of darkness. It not only maintains higher AUC scores in both Precision-Recall and ROC curves but also exhibits better generalization capabilities, making it the most reliable model for drone detection in challenging conditions. On the other hand, The LSTM-T model, while improved by transfer learning, shows more variability in performance compared to Yolov8-T and ViT-T. This indicates that LSTM architectures might be more sensitive to changes in darkness levels, requiring further fine-tuning or potentially alternative approaches to handle extreme conditions effectively. In our final approach, even though models with 60 percent and 80 percent added darkness showed accuracy below 0.6 in precision-recall and ROC curve results, an accuracy above 0.6 was observed for each model when considering the total 40 percent of data subjected to 60 percent or greater darkness.

AUTHOR CONTRIBUTIONS

F. Oruç and H. B. Yılmaz wrote the manuscript, F. Oruç and H. B. Yılmaz developed the approach, F. Oruç conducted the experiments, and F. Oruç and H. B. Yılmaz interpreted the results.

ACKNOWLEDGEMENT

This study has been supported by Boğaziçi University Scientific Research Projects Under Grant 17862 (BAP-SUP-17862). We also want to thank our research team, who helped a lot during the experiments in NetLab of the Computer Engineering Department of Bogazici University. Special thanks will go to Yusuf Özben, Kerem Avcı, and Ömer Şahin Albayram at this point.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Adam, E. Y. (2020). Connectivity considerations for mission planning of a search and rescue drone team. *Turkish Journal of Electrical Engineering and Computer Sciences*, *28*(4), 2228-2243. https://doi.org/10.3906/elk-1912-46

Andraši, P., Radišić, T., Muštra, M., & Ivošević, J. (2017). Night-time detection of UAVs using thermal infrared camera. *Transportation Research Procedia*, *28*, 183-190. https://doi.org/10.1016/j.trpro.2017.12.184

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jamil, S., Abbas, M. S., & Roy, A. M. (2022). Distinguishing malicious drones using vision transformer. *AI*, *3*(2), 260-273. https://doi.org/10.3390/ai3020016

Khan, M. U., Misbah, M., Kaleem, Z., Deng, Y., & Jamalipour, A. (2023, June 20-23). *GAANet: Ghost auto anchor network for detecting varying size drones in dark*. In: Proceedings of the 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring) (pp. 1-5). Florence, Italy. https://doi.org/10.1109/VTC2023-Spring57618.2023.10200720

Li, Y., Fan, Q., Huang, H., Han, Z., & Gu, Q. (2023). A modified YOLOv8 detection network for UAV aerial image recognition. *Drones*, *7*(5), 304. https://doi.org/10.3390/drones7050304

Minderer, M., Gritsenko, A., Stone, A., Neumann, M., Weissenborn, D., Dosovitskiy, A., Mahendran, A., Arnab, A., Dehghani, M., Shen, Z., Wang, X., Zhai, X., Kipf, T., & Houlsby, N. (2022, October 23-27). *Simple Open-Vocabulary Object Detection*. In: S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, & T. Hassner (Eds.), Proceedings of the 17th European Conference on Computer Vision (ECCV 2022) (pp. 728-755). Tel Aviv, Israel. https://doi.org/10.1007/978-3-031-20080-9_42

Misbah, M., Khan, M. U., Yang, Z., & Kaleem, Z. (2023, March 12-13). *Tf-net: Deep learning empowered tiny feature networks for night-time UAV detection*. In: J. Zhao (Eds.), Proceedings of the 13th EAI International Conference on Wireless and Satellite Systems (pp. 3-18). Virtual Event, Singapore. https://doi.org/10.1007/978-3-031-34851-8_1

Moustafa, N., & Jolfaei, A. (2020). Autonomous detection of malicious events using machine learning models in drone networks. In: Proceedings of the 2nd ACM MobiCom Workshop on Drone Assisted Wireless Communications for 5G and Beyond (pp. 61-66). https://doi.org/10.1145/3414045.3415951

Moustafa, N., Slay, J., & Creech, G. (2017). Novel geometric area analysis technique for anomaly detection using trapezoidal area estimation on large-scale networks. *IEEE Transactions on Big Data*, *5*(4), 481-494. https://doi.org/10.1109/TBDATA.2017.2715166

Munir, A., Siddiqui, A. J., & Anwar, S. (2024, January 01-06). *Investigation of UAV Detection in Images with Complex Backgrounds and Rainy Artifacts*. In: Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW) (pp. 221-230). Waikoloa, HI, USA. http://doi.org/10.1109/WACVW60836.2024.00031

Ramadan, R. A., Emara, A. H., Al-Sarem, M., & Elhamahmy, M. (2021). Internet of drones intrusion detection using deep learning. *Electronics*, *10*(21), 2633. https://doi.org/10.3390/electronics10212633

Reis, D., Kupec, J., Hong, J., & Daoudi, A. (2023). Real-time flying object detection with YOLOv8. https://doi.org/10.48550/arXiv.2305.09972

Svanström, F., Alonso-Fernandez, F., & Englund, C. (2022). Drone detection and tracking in real-time by fusion of different sensing modalities. *Drones*, *6*(11), 317. https://doi.org/10.3390/drones6110317

Yi, K. Y., Kyeong, D., & Seo, K. (2019). Deep learning-based drone detection and classification. *The Transactions of the Korean Institute of Electrical Engineers*, *68*(2), 359-363. http://doi.org/10.5370/KIEE.2019.68.2.359

Zhai, X., Huang, Z., Li, T., Liu, H., & Wang, S. (2023). YOLO-Drone: an optimized YOLOv8 network for tiny UAV object detection. *Electronics*, *12*(17), 3664. https://doi.org/10.3390/electronics12173664