Enhancing IR Drone Detection through Synthetic Drone Imagery for Counter-UAV Technology: An Investigation

Student Name: Mr. Jacob Ismail Curtis

Supervisor Name: Hubert P. H. Shum

Submitted as part of the degree of BSc Computer Science to the

Board of Examiners in the Department of Computer Sciences, Durham University

Abstract — This paper presents a novel approach for enhancing infrared (IR) drone detection using synthetic IR drone imagery. The proposed approach involves creating a diverse, synthetic dataset which is used to augment a training dataset for a YOLOv7 model, capable of detecting drones in real-time. The paper introduces two data augmentation techniques for generating IR drone imagery. The first technique involves rendering 3D drones in Blender by randomising simulation parameters, applying post-processing and pasting them onto a 2D IR scene, while the second technique uses an adversarial data augmentation methodology to reduce the synthetic-real domain gap for these 3D drone renders. The proposed approach addresses the challenge of limited IR drone imagery for training and offers promising results for improving drone monitoring systems for counter-UAV applications. The model shows an average of 93.8%% accuracy when training data is augmented by 20% using synthetic data, which is an 2.8% improvement on the model trained using only 'real' drone dataset.

Index Terms— Unmanned Aerial Vehicles, Object Detection, Convolutional Neural Networks, Synthetic IR Imagery

1 Introduction

In recent years, the use of Unmanned Aerial Vehicles (UAVs), or drones, has significantly increased, with companies such as DJI, Parrot, and Skydio leading the industry in innovation [1]. The commercial drone market size in 2020 was estimated at USD 13.44 billion, with a projected compound annual growth rate of 57.5% from 2021 to 2028 (USD 501.4 billion) [1]. As UAVs become increasingly accessible, cost-effective, and easier to operate, their utility for commercial, non-commercial, and military applications has expanded to fields such as remote sensing [2], surveillance [3], and delivery [4]. However, the widespread use of UAVs has also led to a surge in their misuse, both intentional and accidental, ranging from illegal surveillance and data capture to weaponization [5]. Such misuse poses a threat to public and private safety, as well as causing economic and political damage.

The proliferation of UAVs has raised significant concerns about security. Media outlets have reported numerous instances of UAVs flying over restricted areas, critical infrastructure, and public events, presenting significant risks to safety and security. The most notable example of this was an incident in December 2018, where a UAV flying near Gatwick Airport led to the closure of Britain's second-largest airport for 36 hours and disrupted 1000 flights [6]. The cost of this incident was estimated to be £1.4 million, and many passengers were affected. Sensitive sites such as prisons, airports, sporting venues, and public buildings are particularly vulnerable to the challenges presented by UAVs. To ensure the safety and security of critical infrastructure and citizens, the development of effective drone monitoring systems, a crucial element of Counter-UAV Systems (CUSs), is necessary. Drone monitoring systems play a crucial role in detecting and identifying the features of a drone, including its model, size, and payload, which can support the development of mitigation strategies to mitigate potential risks. By providing real-time information about a drone's position, drone monitoring systems can help prevent security breaches and minimise the negative impacts of UAVs on public safety, financial stability, and the economy. Therefore, the improvement of these detection systems is of utmost importance for the protection of critical infrastructure and the mitigation of potential socio-economic risks posed by UAVs.

CUSs can be broadly decomposed into three elements: prevention, detection, and mitigation [7]. Prevention aims to prevent users from mistakenly causing disruptions by imposing restrictions on drone operations. For example, geo-fencing restricts drone operators from flying into restricted airspaces based on the GPS position of the drone [7]. However, these systems are vulnerable to circumvention by malicious users and may not be effective against amateur/hobbyist drone users. Therefore, detection and mitigation systems are particularly essential.

Commercial detection systems rely on various technologies such as acoustic, radio frequency (RF), radar, and visual to detect and track unauthorised drone activities. According to Lykou et al. [8], visual detection systems are the most commonly used commercial systems, accounting for approximately 40% of the total, followed by radar (28%), RF (26%), and acoustic (6%) systems. However, traditional sensor-based drone detection methods, mentioned above, often only utilise a single sensor, leading to a lack of accuracy in drone detection and tracking [8]. To address this issue, data fusion techniques can be employed by combining sensing data from multiple sensors of the same or different types, resulting in more comprehensive and accurate data for UAV detection and tracking. Deep learning algorithms can then be used to process this data, significantly improving accuracy over single-sensor detection systems [9]. Despite these improvements, sensor-based detection remains largely inaccurate in detecting drones at long distances and in high-noise environments, which are often areas where these systems are needed the most, such as airports and military compounds [9]. Given the challenges posed by environmental noise in traditional sensor-based CUSs, there has been a recent trend towards vision-based detection systems for security. Vision-based methods are often preferred due to the affordability of camera sensors, their ability to cover a large area, zoom in on objects of interest, and the resulting imagery is directly interpretable by human operators. Visual cues and information about a drone's model and payload can provide more descriptive information about the drone's utility and if it is being used for malicious intent. This study will employ the visual-based detection approach, leveraging recent breakthroughs in the computer vision and deep learning fields.

The objective of visual-based object detection and tracking is to automatically detect and track instances of a target object from image sequences. Traditional techniques for object detection and tracking involved the extraction of discriminant features from the image, such as the scale-invariant feature transform (SIFT) [10] and histograms of oriented gradients (HOG) [11]. The SIFT feature vector has several attractive properties, including its invariance to an object's translation, orientation, and uniform scaling. It is also less sensitive to projective distortions and illumination changes, making it a robust feature set for object recognition. Similarly, the HOG feature vector is obtained by computing normalised local histograms of image gradient directions or edge orientations in a dense grid. This method provides another powerful feature set for object recognition and has been widely used in various applications such as pedestrian detection [11]. Despite their success, traditional object detection and tracking techniques have some limitations, including the need for manual feature engineering and their limited generalization ability to handle variations in object appearance, pose, and background clutter which are essential for the task of drone detection.

Krizhevsky et al. [12] demonstrated the effectiveness of Convolutional Neural Networks (CNNs) in large-scale object classification tasks through their winning entry in the ImageNet challenge. This breakthrough paved the way for further research and applications of deep learning methods in computer vision. CNNs are composed of multiple convolutional and fully connected layers, with each layer followed by a non-linear activation function. These networks can be trained end-to-end using backpropagation. By generating highly discriminant features, CNN-based methods have significantly outperformed traditional object detection techniques in terms of detection accuracy and efficiency [12].

CNN-based object detection techniques can be broadly categorised into one-stage and two-stage detectors. One-stage detectors aim to detect objects in a single shot by generating dense bounding boxes across the image and classifying them using convolutional neural networks. YOLO (You Only Look Once) [13] and SSD (Single Shot Detector) [14] are examples of one-stage detectors. In contrast, two-stage detectors first propose a set of region proposals, which are potential locations of objects, before classifying them. This approach is computationally more expensive, but it is typically more accurate than one-stage detectors. R-CNN [15] and Faster R-CNN [16] are examples of two-stage detectors. One-stage detectors are typically faster and more efficient, while two-stage detectors offer better accuracy and are more robust to occlusion and cluttered backgrounds. This paper will utilise the YOLOv7 [17] model for drone detection, due to its reduced computational requirements, faster training time and processing speed, and proven accuracy in literature for drone detection. Existing research into vision-based deep learning drone detection will be discussed in more detail in Section 2.

Although deep learning techniques mentioned above have achieved significant success in vision-based drone detection, a critical challenge remains the detection of small objects in cluttered backgrounds and low light conditions, such as dawn, dusk, or nighttime [18]. To address these limitations, utilising infrared (IR) or thermal imagery is an alternative method to normal, visible imagery in an image-based drone detection system. Despite their different visual characteristics, the structure of the input data to object detection is similar in both cases allowing us to utilise the same deep learning techniques that are used on visible imagery. Due to the paucity and diversity of publicly available IR drone datasets, this paper will propose methods to generate synthetic IR drone imagery for the purpose of training YOLOv7 model for IR drone detection. This approach will rely on only using 3D drone models. The use of a synthetic dataset offers several advantages over real-world imagery, such as known ground truth labels, eliminating the time cost and human error associated with labelling input data. Moreover, it enables specific control over what drone scenarios are generated, which can be tailored for model training to address any under-representation in a training dataset.

The rest of this paper is organized as follows. Related work is reviewed in Section 2. The methodology is introduced in Section 3 which includes the methods used to generate the synthetic images, the process of training the convolutional neural network using the synthetic images and the choice of hyperparameters and model architecture. The synthetically generated imagery is presented and the results of the YOLOv7 experiments are presented in Section 4. An evaluation of the results is in Section V where these results are discussed regarding existing literature. Concluding remarks and suggestions for future work are given in Section 5.

2 Related Work

Numerous studies have been conducted into the UAV detection, particularly for use in CUSs. In this section, research found in literature pertaining to vision-based drone detection is reviewed. Related works are examined for both RGB and IR drone detection, primarily deep learning-based, that are both essential for understanding the various challenges encountered in drone detection and provide context for the overall scope of this study.

2.1 Visible Drone Detection

In the field of computer vision, drone detection can be conceptualized as an object detection problem. Object detection involves identifying the presence of predefined objects in an image and pinpointing their location using the smallest possible bounding box. As outlined in Section 1, early approaches to this task relied on hand-crafted features to describe objects [19], [20], [21], [22], while current state-of-the-art algorithms employ deep learning. The advancements in deep learning have led to significant improvements in object detection algorithms, and Unmanned Aerial Vehicle (UAV) detection is no exception. UAV detection methods utilize Convolutional Neural Networks (CNN) to autonomously detect drones using standard daylight cameras. Deep learning has enabled considerable progress over traditional object detection methods, as it facilitates learning highly expressive features. This has shifted the focus towards developing generic object detectors that perform well across various classes and datasets, as opposed to designing detectors for specific objects, which can be a laborious and error-prone process [23]. Two main categories of deep-learning architectures have been widely explored: two-stage detectors and one-stage detectors.

Two-stage detectors, such as RCNN [15], Faster RCNN [16], and RFCN [24], involve a proposal generation stage to identify candidate regions of interest that may contain objects. These regions are then classified, and their bounding box parameters are regressed in a second stage. Although two-stage detectors generally achieve high detection accuracy, they are generally more computationally intensive than one-stage detectors.

In contrast, one-step detectors, including OverFeat [25], YOLO [13], SSD [14], and CornerNet [26], avoid the need for a separate proposal generation step. Instead, they directly predict bounding boxes and class probabilities. OverFeat employs a deep multiscale and sliding window approach [25], while YOLO [13] processes the entire image to produce predictions in a single evaluation. SSD [14] utilizes default boxes of varying aspect ratios and scales, associating them with each location on the feature map. The goal of one-step detectors is to achieve faster training and testing times, albeit at the expense of slightly reduced accuracy.

The detection of unmanned aerial vehicles (UAVs) introduces unique challenges. UAVs often appear at varying sizes within images, which poses difficulties for deep networks in extracting object feature information [18]. Particurly, at extremely small scales, where drones are represented by far fewer pixels often with motion blur, accuracy tends to suffer significantly. The deepest layers of the convolutional backbone struggle to effectively represent small objects, while subsequent max-pooling layers can inadvertently suppress detection responses [18]. Small-object detection algorithms can be mainly divided into two kinds. One is to improve the detection performance of small objects with multiple scales in a video or image sequence and the other is to improve the detection performance of small objects with only one scale in an image. The former is typically addressed using feature pyramid networks, to fuse features from multiple layers, thereby enabling detection at different scales [27]. Apart from increasing the frequency of small objects in detection model training images, approaches for only detecting smalls scale drones tend to be designing network [28], using context information [29] and generating super-resolution imagery [30].

Distinguishing UAVs from visually similar objects, such as birds, represents a notable challenge, as exemplified by the Drone-vs-Bird detection challenge [31]. The presence of small, similarly represented objects in the sky, including but not limited to birds, necessitates the development of sophisticated techniques to filter out false positives. Akyon et al. [32] emerged as previous winners by employing a YOLOv5 detection model fine-tuned with synthetically generated imagery, combined with a Kalman-based object tracker to enhance detection confidence [32]. Temporal information is also often utilized in two-stage detectors, which incorporate temporal-aware input channels and standard tracking algorithms to address this issue and improve discrimination between UAVs and other objects [33].

Complex backgrounds because of adverse weather conditions, poor visibility and low light scenarios or diverse, cluttered backgrounds all pose additional complications for UAV detection [18]. Techniques such as dehazing, de-raining, and low-light enhancement have been explored to improve visibility in detection tasks [34]. For low-light and foggy scenarios, Liu et al. [35] uses an Image-adaptive YOLO framework that shows great promise for image such scenarios. However, there is an inherent limit to the extent to which these techniques can improve visibility, especially in low-light conditions, indicating a need for alternative modalities for drone data capture.

2.2 IR Drone Detection

In contrast to the existing body of research focused on the identification of drones in the visible spectrum, limited attention has been given to the detection of drones using thermal imaging. Thermal imaging offers distinct advantages over the visible spectrum, such as the ability to overcome challenges related to low-light conditions, complex backgrounds, and adverse weather. Thermal energy emitted by objects differentiates them from their backgrounds, enabling visibility during nighttime or in challenging environments. Drones, with their various physical components including motors, batteries, and internal hardware, emit significant amounts of heat, which can be effectively detected by thermal cameras [36]. Numerous studies have proposed the detection of target drones based on their thermal signatures. For instance, Andraši et al. [37] proposed a drone detection scheme that relies on identifying the thermal energy emitted by drones during flight. Carrio et al. [38] presented a system for detecting obstacles in flight, demonstrating that thermal images are more suitable than color images in extreme lighting conditions when it comes to UAV detection. Despite the advantages of thermal detection in terms of weather resilience and identification availability, it is worth noting that the practical detection range achieved by Andraši et al. [37] is significantly shorter (51 m) compared to other approaches. Consequently, improving the granularity of the detection scheme or enhancing the resolution of thermal imaging cameras are critical challenges. Svanström et al. [39] investigated the influence of distance on detection results and developed a multi-sensor drone detection system using YOLOv2 [40], incorporating thermal IR data. They also released a public drone detection (DD) dataset containing 650 annotated IR and visible videos of drones, birds, airplanes, and helicopters, although it is important to acknowledge the strong bias of these images, as they were recorded solely in a single location [39] with a simple background and clear weather conditions. In addition to the DD dataset, the Anti-UAV [41] dataset is a notable dataset that offers unaligned, paired visible-IR, labelled drone imagery in a variety of locations. This dataset is particularly useful as it allows us a unique opportunity to analyse how a variety of visible light scenarios manifest in an IR spectrum. This dataset will be utilized in this paper and elaborated on in Section 3.2.2. Finally, the USC dataset [42] comprises 10 thermal video clips, each lasting one minute, captured exclusively on the USC campus using a single drone. While this dataset offers a specific context for IR imagery, its size and limited diversity likely present strong bias.

The primary challenge in IR drone detection lies in the scarcity of thermal drone detection datasets. Only three thermal datasets for drone detection currently exist, namely Anti-UAV [41], USC [42], and DD [39] as mentioned above. The limited availability of these datasets restricts the development and validation of new algorithms and approaches in this field. To further advance the study of IR drone detection, there is a desparate need for new, diverse, and representative datasets, covering various geographical locations, environmental conditions, and drone types.

2.3 Synthetic Data Generation

The great need for more IR drone datasets in addition to the large amount of effort associated in compiling comprehensive data collection with accurate annotation opens doors for synthetic IR data generation that aims to address this shortage. One potential solution to address this is data augmentation, a technique that involves creating new training samples by applying various transformations, such as rotation, flipping, and scaling, to the existing data. This approach increases the size and diversity of a dataset, which helps improve a model's generalization capabilities. However, traditional data augmentation techniques may not be sufficient for the drone detection task, as the image transformations are often applied to existing images in the dataset that often reinforce existing content rather than introducing novel variations.

Additionally, pre-processing techniques that emulate IR imagery through RGB image transformations, have been utilized to address the scarcity of annotated IR drone imagery. The abundance of labeled RGB drone imagery allows for leveraging existing annotations in the detection of IR imagery. One approach, as demonstrated by Sommer et al. [43], involves transforming RGB images into the HSV color space, subsequently reducing saturation and inverting the colors to achieve images that closely resemble the IR domain. These transformed images were then employed to augment the Anti-UAV dataset for training a drone detector based on Faster R-CNN [16] and FPN architectures, resulting in significant performance improvements compared to utilizing only real IR training data.

Another method of generating synthetic data is the use of 3D graphics engines. This enables the generation of diverse datasets but also poses the challenge of the domain gap. Yang et al. [44] employed an open-source simulation environment Aerial Informatics and Robotics Simulation (AIRSim) [45], [46] to synthesize photorealistic 2D RGB aerial images of pedestrians and cars in both day and nightime scenarios. The researchers then applied CycleGAN [47], an unpaired generative adversarial network for image-image translation, to convert the RGB images to the IR domain. While this method seemed to provide visually realistic results and provided a solution end-to-end synthetically generated IR imagery, it exhibited poor performance when tested with an actual IR dataset. Despite this, the use of 3D graphic engines, to improve object detection and pose estimation in the visual spectrum domain continues to receive a lot of attention from the research community, particularly in the autonomous driving field [48].

Other uses of GAN’s for generating synthetic IR imagery include ThermalGAN [49] and InfraGAN [50] which build on the Pix2Pix [51] architecture that take in paired visible-IR imagery. Although, use of these techniques is impractical for generating IR drone imagery as their exist no aligned, paired visible-IR drone imagery. Wang et al., has used a modified CycleGAN approach that exploits adaptations to the relative importance of the cycle-consistency loss so that only that the network only learns the texture translation between the two domains.
Based on the literature into synthetic IR imagery for object detection, it is evident that a recommended approach for incorporating synthetic IR imagery alongside real IR imagery is to employ a 10-20% augmentation of synthetic images to enhance object detection. This recommendation will be adhered to in the methodology described in Section 3.

3 METHODOLOGY

This section details the chosen object detection model (YOLOv7) [17] and the methodology for the creation of a synthetic IR drone dataset, using Blender [52] a freely available 3D modelling program and CycleGAN [47].

3.1 YOLOv7

For IR drone detection, the YOLOv7 [17] was selected, a widely popular one-stage object detection method that employs a single convolutional neural network (CNN) to concurrently predict bounding boxes and class probabilities for all objects within an image. YOLOv7 boasts superior precision and speed compared to its predecessors and achieved a peak average precision of 56.8% [17] on the COCO dataset.

The YOLOv7 architecture comprises a head, neck, and backbone, as depicted in Fig 1, with the model's predicted outputs situated in the head. Owing to its inspiration from Deep Supervision [53], YOLOv7 is not restricted to a single head; the lead head generates the final output, while the auxiliary head facilitates middle-layer training. To enhance model training further, a Label Assigner method was devised, which assigns soft labels based on ground truth and network prediction results. Unlike its YOLO predecessors, YOLOv7 moves away from using the darknet backbone, instead opting for The Extended Efficient Layer Aggregation Network (E-ELAN) backbone for its core computational block [17]. By incorporating "expand, shuffle, merge cardinality," the YOLOv7 E-ELAN architecture allows the network to continually augment its learning capacity without disrupting the original gradient path, ultimately promoting more effective learning.

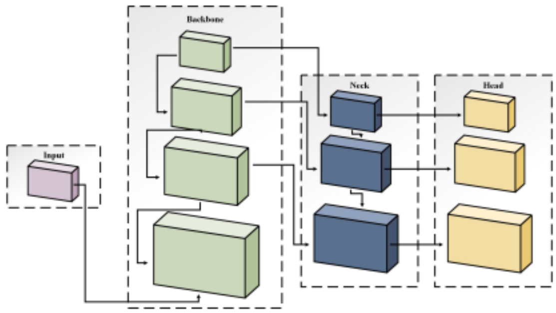
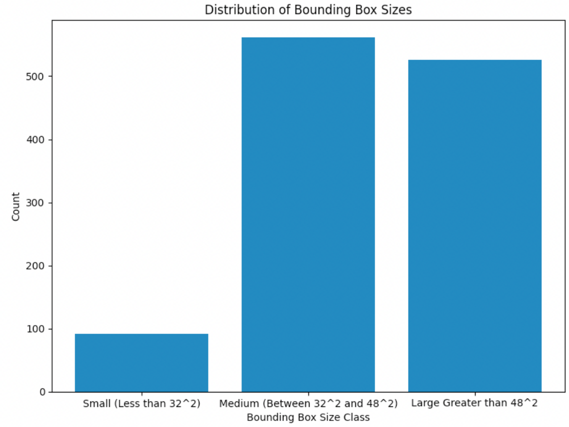


Fig 1. Block Diagram of Single Stage Object Detector Algorithms.

3.2 Anti-UAV Dataset

The Anti UAV Challenge [41] dataset is an extensive and publicly available resource for detecting drones in IR imagery. The dataset contains 100 pairs of RGB-IR unaligned video sequences, each of which includes annotated bounding boxes. These videos were recorded in an urban environment with the backgrounds primarily consisting of buildings and clouds and has a resolution of 640 × 512 pixels. The dataset contains three drone models, including an octocopter and two quadcopters, which present at varying scales in the dataset, ranging in size from 8 × 11 to 57 × 198 pixels, with an average pixel area of 2786 pixels. The distribution of drone bounding box sizes can be seen in Fig. 2 below.

Fig 2. Drone size distribution in Anti-UAV dataset.



For model training, the video frames are randomly divided into a training and a validation set, using an 80:20 split. Due to the high frame rate, the adjacent frames often contain redundant image content, which is addressed by selecting every a random sample of frames of each video, resulting in a total of 3136 training images and 784 validation images, and significantly reducing model training time. Example images from the Anti-UAV dataset can be seen In Fig. 3 BELOW.

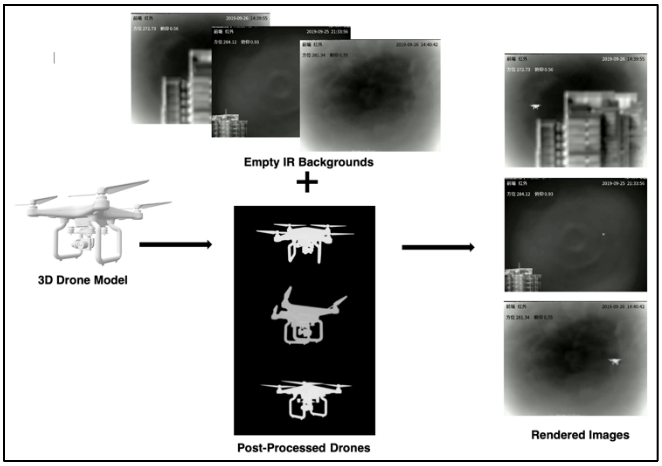
Fig 3. Frames from Anti-UAV dataset demonstrating different backgrounds, drone models and object scales.

The Anti-UAV dataset includes a JSON file containing bounding box annotations for each frame in its associated video. To make the dataset compatible with YOLOv7 training, several preprocessing steps are performed. Separate annotation files are generated for each frame, assigned names corresponding to their respective frames, and converted the bounding box coordinates to represent the center rather than the top-left corner. Lastly, these values are normalized in relation to the image dimensions.

3.2 Generating Synthetic IR Imagery

Training a neural network for object detection requires a comprehensive and high-quality dataset, which poses a significant challenge. Collecting actual images from real-life drone scenarios involves a substantial amount of time and effort to outline bounding boxes around the objects of interest. While the Anti-UAV dataset has been proposed as a potential solution, evaluating the performance of a YOLOv7 model using this dataset led to explore alternative options. To address the scarcity of IR drone datasets, this paper investigates synthetic IR image generation, randomising the drone's lighting, pose and scale, which could then be used to train a YOLOv7 model, without the need for the manual labelling of drone images. This approach also allowed us to create datasets for drone models that were not physically accessible to us and to address under-representations in real-life datasets. To this end, 3D models are used, obtained from free online sources, namely [54, 55], to mimic the drones found in the Anti-UAV dataset. Both data augmentation and CycleGAN techniques detailed below involve rendering and processing a drone image and a background image separately and pasting the drone image on top of the background image. This technique is depicted below in Fig. 4.

Fig 4. Process of rendering a synthetic drone. Random conditions are applied to the drone which are then post-processed and pasted onto IR backgrounds.



3.2.1 Data Augmentation from 3D Models

Similar to the implementation described in [56], to generate a diverse range of drone images, a script was developed to randomly generate drone images. A python script can randomise the view-distance, viewing-angle, and lighting conditions of the camera in relation to the 3D drone meshes in blender, allowing for flexible rendering of 2D images. The aim of this approach is to produce an extensive set of synthetic images that replicate the intricate background images and foreground drone models found in a real-world setting. Fig. 5 below provides examples of drone instances in various appearances found in the Anti-UAV Dataset.

Fig 5. IR drone instances from the Anti-UAV dataset [41]

The specific transformations and post-processing applied to the drone renders are detailed below.

1) Lighting Variations

To simulate lighting variations, a point light source is set-up, and its energy is adjusted between 200 and 500 units. This randomization of light intensity ensured that the scene would have different levels of illumination, thereby creating diverse lighting conditions and drone shadows across the rendered images.

2) Drone Orientation

To simulate various orientations and poses of the drone, the script generated a random camera position for each rendered image by assigning random coordinates to the camera object at 350 units away from the drone object. The camera was positioned at elevation angles ranging between approximately -25 and 5 degrees relative to the drone and set to face the center of the drone object, ensuring that the drone was always in view. The drone propellers were rotated such that their position was perpendicular to the camera to simulate their appearance in a 2D image due to motion blur.

3) Drone Texture and Colour

The code utilized a custom IR material, which was applied to the drone object within the scene. The material was created using a combination of emission and principled BSDF shaders, configured with attributes such as roughness, specular, base color, and metallic properties. The drone's color was primarily determined by color mapping, which mapped IR intensity ranges based on light intensity to a grayscale color output. This design generated a unique texture for each rendered image, mimicking the appearance of drones in IR imagery.

4) Post-processing

Observations of the Anti-UAV dataset had revealed that the quality of drone images could vary significantly due to factors such as camera motion and focus, as illustrated in Fig. 3. To replicate this variation in drone appearance, Gaussian blur and motion blur were employed to simulate the blurring effects commonly observed in the dataset. Furthermore, the images were proportionally scaled within a range of 5 to 100 pixels before being superimposed onto IR backgrounds. Fig. 6 provided examples of the final, synthesized image with the render pasted on an IR background. These IR backgrounds were collected from empty frames from the Anti-UAV dataset that did not contain drones. Random horizontal flips were applied to these background images to increase the number of available empty backgrounds.

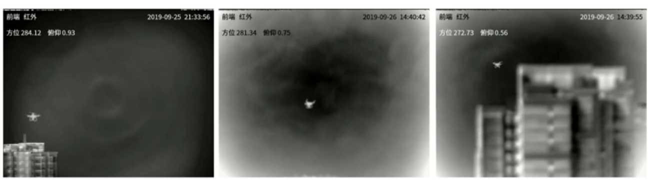


Fig 6. Synthesized IR drone images with randomised illumination, scales image qualities, and complex backgrounds.

3.2.2 CycleGAN

While the data augmentation method (Section 3.2.1) for generating 2D IR drone renders shows promise in creating a diverse dataset, it faces challenges that may hinder its ability to accurately capture real-world complexities. These challenges include a possible domain gap between synthetic and real images due to limited realism, an incomplete representation of IR properties from various drone materials and surfaces, and the constraints of post-processing techniques like Gaussian blur and motion blur in mimicking real-world blurring effects. Moreover, the method depends on manual parameter selection, which might not yield images that best resemble real-world IR drone imagery. To address these concerns, implementing CycleGAN [47] as an alternative approach to refine the 2D IR drone renders could help overcome these limitations and generate a more realistic dataset for training the YOLOv7 detection model.

CycleGAN is a generative adversarial network (GAN) architecture designed for unsupervised image-to-image translation [47]. This method could be used to enable the transformation of synthetic drone renders to appear more like real drone images without the need for paired training examples. By learning the underlying distribution of the real drone images and mapping it to the synthetic drone images, CycleGAN can generate more realistic IR drone images, effectively reducing the domain gap between the two domains.

CycleGAN operates by training two generator networks and two discriminator networks simultaneously. The generator networks learn to transform images from one domain to another (i.e., synthetic drone renders to real drone images and vice versa), while the discriminator networks assess the authenticity of the generated images [47]. Through adversarial training, the generators improve the quality of the translated images, ultimately converging to produce images that resemble the target domain. The incorporation of cycle consistency loss ensures that the mapping between the domains is consistent, resulting in more accurate translations [47]. The loss function for CycleGAN consists of two main components: the adversarial loss and the cycle consistency loss. The full loss function can be expressed as:

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda * L_{cycle}(G, F, X, Y)$$

Here, G and F are the two generator functions ($G: X \rightarrow Y$, $F: Y \rightarrow X$), and DX and DY are the discriminators for domains X and Y, respectively. λ is a hyperparameter that controls the relative importance of the cycle consistency loss. The CycleGAN loss function aims to optimise the generators and discriminators simultaneously by minimizing the adversarial loss and the cycle consistency loss [47].

By combining described script for generating diverse drone images with CycleGAN to refine the 2D IR drone renders, improving the dataset's generalizability to real-world situations. Training data for domain Y (real IR domain) will be collected from the Anti-UAV dataset. Drone images will be cropped using their YOLO bounding boxes and scaled to 128x128 pixels. To paste the output drone images onto an IR background, transparent drone images are required. To achieve this, a background subtraction algorithm is employed to eliminate the background from the drone image, utilizing the original drone render image to determine which pixels should be retained. These images are then pasted onto IR background like in Section 3.2.1. Fig. 7 shows the results of applying CycleGAN to pre-processed 3D model renders and the results of the background subtraction algorithm. Fig. 8 shows the final training data with the drone renders pasted onto IR backgrounds.

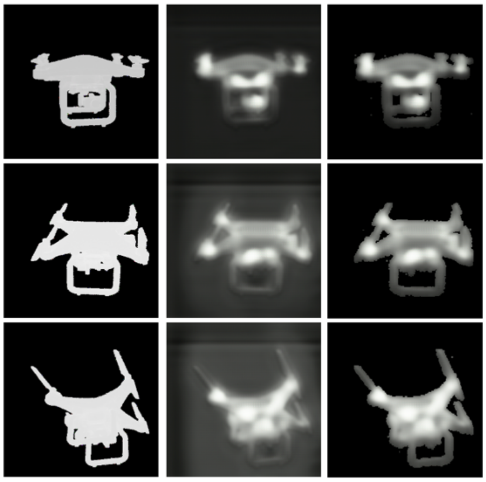


Fig 7. Pre-processed 3D model renders (left), CycleGAN results (middle), subtracted background (right)

Fig 8. Final synthetic imagery using method Cycle-GAN method described in Section 3.2.2

As anticipated, the visual analysis demonstrated that the CycleGAN enhanced the realism of the drones. The original shape of the drone in the 2D render was maintained, while the texture and color were adjusted to align more closely with authentic IR drone imagery. Notably, the CycleGAN effectively highlighted the contact points of the propeller, motor, battery, and camera, which would all be expected to possess higher thermal energy due to friction and their mechanical nature. Aside from the background removal, no additional blur was applied to the renders before or after the application of CycleGAN. This CycleGAN approach successfully able to capture the anticipated blur typically associated with aerial objects in motion.

3.3 Experimental Setup

In this section, the methodology for training the neural network utilizing the synthetic image dataset, as discussed in Section 3.2, is outlined. To evaluate the efficacy of different IR drone synthetic data generation methods for training object detection models, YOLOv7 models were trained using five different datasets. The primary objective of these experiments is to determine the efficacy of different synthetic data generation techniques for replicating real IR data and the impact of incorporating synthetic and augmented data on the model's performance in detecting real IR drone objects. The five training datasets are described below:

1) **100% Real IR Data**

In the first experiment, real IR drone data sourced from the Anti-UAV Dataset was utilized. This experiment served as a baseline for comparison with other experiments that involved augmented or synthetic data.

2) **Real IR Data + 20% 3D Model**

The second experiment employed real IR drone data from the Anti-UAV Dataset combined with a 20% augmentation dataset, which was generated using a 3D drone model as described in Section 3.2.1. This experiment aimed to assess the benefits of augmenting the original dataset with 3D-model-based synthetic data.

3) **100% 3D Models**

In the third experiment, solely relies on a 100% augmented dataset using 3D drone models. This experiment aimed to determine the feasibility of using entirely synthetic data for training object detection models.

4) **Real IR Data + 20% CycleGAN**

The fourth experiment combined real IR drone data from the Anti-UAV Dataset with a 20% augmentation dataset generated using CycleGAN applied to the 3D model dataset as described in Section 3.2.2. This experiment sought to explore the advantages of incorporating CycleGAN-generated synthetic data in training object detection models.

5) **100% CycleGAN**

In the fifth experiment, the detector was trained with 100% augmented dataset generated with a 3D drone model and CycleGAN. This experiment investigated the potential of an entirely synthetic dataset, created using advanced generative techniques, in training object detection models for IR drone detection.

The YOLOv7 experiments are all trained for 20 epochs, with a learning rate of 0.02 for the first, a batch size of 64 a momentum of 0.9. The input image size is set to 640 × 640 pixels for all training data. By conducting these five experiments, comprehensive analysis of the impact of various synthetic data generation methods on the performance of object detection models in the context of IR drone detection hopes to be provided. The results of these five experiments are discussed in Section 4.

4 RESULTS

Having completed the training of the YOLOv7 models detailed in Section 3.3, each model's performance was evaluated using a consistent test set derived from the Anti-UAV dataset. To comprehensively evaluate the impact of varying training data on detection performance, an array of metrics was employed. In this study, the $mAP_{0.5}$ metric was adopted, which classify detections as true positives if their Intersection Over Union (IOU) with a ground truth bounding box exceeds 50%. Additionally, the $mAP_{0.5:0.95}$ metric was utilised, which calculates the mean average precision at IOU thresholds ranging from 50% to 95%, at increments of 5%. This metric is reported for entire frames ($mAP_{0.5:0.95}$) as well as for three distinct object scales: small (mAP_S), medium (mAP_M), and large (mAP_L). These object scales represent bounding box areas of less than 1024 pixels, between 1024 and 2034 pixels and more than 2023 pixels respectively. The results for these five metrics are listed below.

TABLE 1
MAP METRICS FOR ALL YOLOV7 DETECTION MODELS

Experiment	Training Data	$mAP_{0.5}$	$mAP_{0.5:0.95}$	mAP_S	mAP_M	mAP_L
1	Real Anti-UAV	0.910	0.515	0.793	0.934	**0.989**
2	Real Anti-UAV + Processed 3D Models	0.906	0.503	0.814	0.923	0.962
3	Processed 3D Models	0.621	0.293	0.427	0.742	0.834
4	Real Anti-UAV + CycleGAN	**0.938**	**0.554**	**0.842**	**0.962**	0.985
5	CycleGAN	0.778	0.425	0.571	0.679	0.812

5 EVALUATION

As seen in Table 1, the best training data appeared to be a combination of real IR data from the Anti-UAV dataset and synthetic IR data generated using post-processing and CycleGAN. This resulted in a 2.8% increase in performance over the baseline model (Experiment 1). This method produced improved results in every metric apart from the detection of large UAVs. This could likely be attributed to the fact that more pixel gives more opportunity to increase the domain gap for synthetic images. The results for purely synthetic IR drone detection (experiment 3 and 5) show reasonable results for IR drone detection. Using CycleGAN appears to increase the $mAP_{0.5}$ by a significant 15.7%, validating the generative adversarial method of reducing the domain gap.

The focus of the qualitative detection results will predominantely be centered on results from training from Experiment 4 (Real Anti-UAV + CycleGAN augmentation) which yielded the best $mAP_{0.5}$ results out of all the experiments. Fig 8. demonstrates successful detection results in random Anti-UAV test data.

5.1 Reviewing the Detections

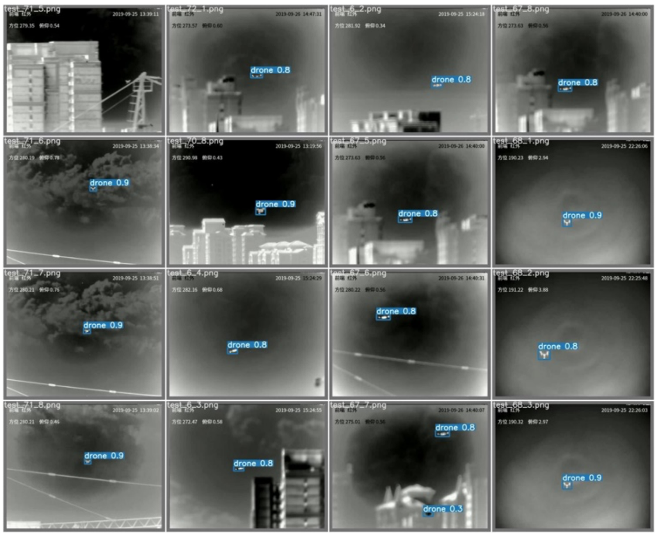


Fig 8. YOLOv7 detection test results for Experiment 4 detecting 15/16 present drones

5.2 Fail Cases

In Fig. 9, failure cases for our detection model trained on Experiment 4 are presented. From my test results, it seems failure cases can be categorized into three scenarios, extreme motion blur (first row), small scale (second row), and complex or blending into background (third row).

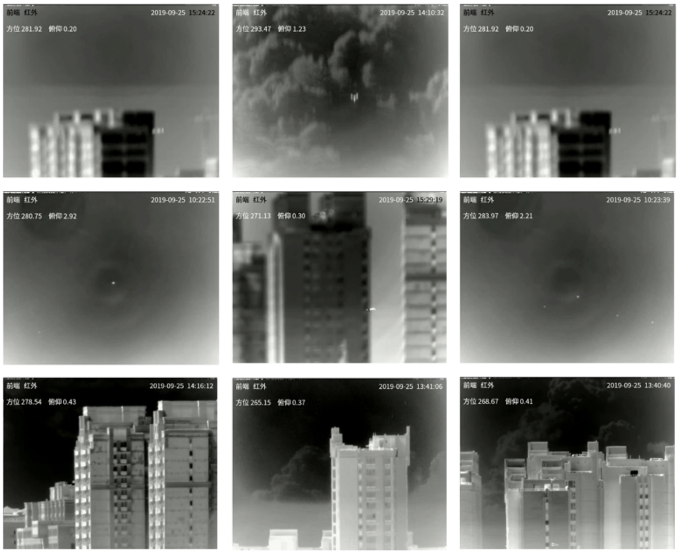


Fig 9. Failure cases for YOLOv7 model trained in Experiment 4

Despite the failure cases listed above, the model trained using Real Anti-UAV + CycleGAN augmentation in Experiment 4 proved to be more robust to motion blur than the model trained purely on real IR anti-uav imagery in Experiment 1. This can likely be attributed to the emphasis placed on gaussian and motion blur in the pre-processing stage. Although small-scale UAV detection did show improvements in Experiment 4 over the base model in Experiment 1 as demonstrated by a 4.9% increase in mAP_S, there is still a lot to be desired for small object detection. This is likely a result of a class imbalance in the training dataset given only 7% of training data is considered small scale. While fine-tuning with more small-scale synthetic drone images will likely improve this metric, this could also just be a limitation of the IR medium of detection, as at large distances, IR information gets lost as discussed in Section 3.2.2. Finally, the detection system ocassionally encountered difficulties in complex backgrounds, where the colour appearance of the drone blended with the background, often buildings. While this too could be addressed with fine-tuning, it is also worth mentioning that should this detection system be integrated into a drone monitoring system, temporal and trajectory information of the drone may help to localise the area in which the drone could be located.

5.3 Comparing to Other Results

Evaluating the efficacy of our infrared (IR) drone detection solution using synthetic imagery generation is challenging, as there is no established state-of-the-art detection benchmark that specifically utilizes synthetic training data. Most research in the field of synthetic training data generation primarily focuses on visible detection. Table 2 provides a compilation of significant outcomes in drone detection and tracking. While a direct comparison with the results presented in this study may not be appropriate, some notable outcomes in this category will be discussed.

Isaac-Medina et al. [57], boasts the best detection results for the Anti-UAV dataset, comparing four different detection architectures and achieving a $mAP_{0.5}$ of 98.0% using the YOLOv3 [58] model for the IR Anti-UAV dataset. While these results are impressive, it is worth mentioning that this model was trained on the entire Anti-UAV dataset as opposed to a small subset as used in this study.

Wisniewski et al. [56] introduced a comparable method for synthetic training data generation, employing a DenseNet201 [59] classification model trained exclusively on synthetic visible drone data. Their approach involved the use of 3D drone models and domain randomization techniques such as Gaussian noise and texture randomization to generate diverse training data. With a training dataset size of 1,000 images, they achieved a 92.4% classification accuracy, demonstrating improved performance compared to dataset sizes of 5,000 (82.6%) and 10,000 (86.9%). This suggests that perhaps the training dataset sizes for experiments 3 and 5 in our study were too large.

Sommer et al. [43] examined the impact of augmenting real IR training data with pre-processed visible data to simulate IR imagery. They employed various processing techniques and real-synthetic training data ratios to train a Faster RCNN [16], which was then tested on their own

TABLE 2
COMPARISON OF DRONE DETECTION AND TRACKING RESULTS IN THE LITERATURE.

Paper	Training Dataset	Method	Test Dataset	Task	Classification Accuracy	Detection mAP
This Paper	Mixture Synthetic and Real	YOLOv7	Subset of Anti-UAV Dataset	Detection	N/A	93.8%
Isaac-Medina et al., 2021	Real IR	YOLOv3	Anti-UAV	Both	N/A	98.0%
Wisniewski et al., 2022	Synthetic Visible	DensetNet201	Subset of Anti-UAV	Classification	92.4%	N/A
Sommer et al., 2022	Mixture Synthetic and Real	Faster RCNN	Own dataset	Detection	N/A	87.9%

IR dataset.

The optimal detection results were achieved with a 1:1 ratio of real Anti-UAV data and HSV-inverted synthetic data, yielding a mAP of 87.9%. This study provides the most relevant comparison to our work, as it also utilized a combination of real and synthetic training data, with the real training data originating from a subset of the Anti-UAV dataset. If their test dataset were publicly accessible, it would present an intriguing opportunity to compare detection performance and assess the effectiveness of our synthetic IR data generation techniques.

6 Conclusions

In this work, the task of deep learning-based drone detection in IR imagery with limited training data is examined. Due to the scarcity amount of publicly available IR imagery, this paper analyzed the impact of synthetic IR drone imagery generation using 3D models on the performance of a YOLOv7 drone detection model. For data generation, a variety of pre-processing techniques is used in addition to a generative adversarial apporach to reduce the domain gap. While training the detector with synthetic IR imagery with only basic post-processing garnered poor detection results of 62.1% ($mAP_{0.5}$), this performance was shown to increase by 15.7% when the Cycle-GAN was applied to the synthetic data. The best training data was shown to be an augmentation of the real IR dataset by 20% using the CycleGAN on post-processed 3D renders, increasing $mAP_{0.5\ \text{from}}$ 91% to 93.8%. These final detection results are respectable results metrics in similar existing literature [43], [57]. This paper has demonstrated that an end-to-end IR data generation method can be used to exclusively train a YOLOv7 detection model for IR drone detection yielding adequate results and that synthetic data augmentation can be used to improve IR drone detection by 2.8%.

With the increasing prevalence of drone incursions in restricted airspaces, it is essential for security measures to accurately detect aerial threats [5]. This study contributes to the development of effective detection strategies as part of CUSs. Numerous avenues exist that could be explored to build on this research. In this paper, only the task of drone detection is investigated, and the task of drone classification remains to be considered. This was primarily because the Anti-UAV dataset was not designed for the purpose of classification. Nevertheless, undesrstanding a drone's model and payload is useful information when determining a drone's utility in the context of determining malicious intent and as such a dataset designed for drone classification would significantly enhance the state of this research. In addition, classification can be used when discriminating against birds that often cause false positive in drone detection systems. Additionally, this work solely focuses on YOLOv7 detection. To conduct a more robust and comprehensive investigation, it would be beneficial to examine the performance of a two-stage detector like Faster R-CNN when employing the proposed techniques for generating synthetic IR data. Moreover, it is worth investigating the expansion of the research to include other drone model types, particularly fixed-wing drones, as there appears to be a noticeable gap in the literature in this regard. Furthermore, exploring other types of flying objects such as helicopters and planes would be beneficial. Regarding the method of synthetic IR imagery generation, this paper combines multiple post-procsesing techniques in addition to generative adversarial techniques to generate photorealistic IR imagery. However, the specific effects of including or exclusing individual processing methods have not been studied. Conducting an ablation study would be beneficial in quantifying the effects of each method utilized for synthetic data generation. Additionally, rendering and processing 3D models requires a powerful graphics card which might not often be accessible. Validating the effectiveness of synthetic IR training data for drone detection on unseen IR datasets is a crucial next step to assess the system's robustness and determine the extent of biases introduced by the limited locations covered by the Anti-UAV dataset. The omission of validating on unseen data is attributed to the sheer scarcity of IR drone datasets. Finally, the scope of this paper focused on drone detection. Future work could concentrate on integrating this detection system into a comprehensive drone monitoring system. This integration could involve merging the detection module with a generic object tracker such as Multi-Domain Network (MDNet) [60] serving as the tracking backbone as seen in [42].

References

[1] Grand View Research. (2021). Commercial Drone Market Size, Share & Trends Analysis Report By Product (Fixed-wing, Rotary Blade, Hybrid), By Application, By End-use, By Region, And Segment Forecasts, 2021 - 2028 (978-1-68038-482-6).

[2] Hu, Y., Wu, X., Zheng, G. and Liu, X., 2019, July. Object detection of UAV for anti-UAV based on improved YOLO v3. In *2019 Chinese Control Conference (CCC)* (pp. 8386-8390). IEEE.

[3] Jung, S. and Kim, H., 2017. Analysis of amazon prime air uav delivery service. *Journal of Knowledge Information Technology and Systems*, *12*(2), pp.253-266.

[4] Zhou, G., Ambrosia, V., Gasiewski, A.J. and Bland, G., 2009. Foreword to the special issue on unmanned airborne vehicle (UAV) sensing systems for earth observations. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(3), pp.687-689.

[5] Kang, H., Joung, J., Kim, J., Kang, J. and Cho, Y.S., 2020. Protect your sky: A survey of counter unmanned aerial vehicle systems. *IEEE Access*, *8*, pp.168671-168710.

[6] Guardian, T. Gatwick Drone Disruption Cost Airport Just £1.4 m. 2018. Available online: https://www. theguardian.com/uk-news/2019/jun/18/gatwick-drone-disruption-cost-airport-just-14m (accessed on 6 May 2019).

[7] Faughnan, M.S., Hourican, B.J., MacDonald, G.C., Srivastava, M., Wright, J.P.A., Haimes, Y.Y., Andrijcic, E., Guo, Z. and White, J.C., 2013, April. Risk analysis of unmanned aerial vehicle hijacking and methods of its detection. In *2013 IEEE Systems and Information Engineering De-*

sign Symposium (pp. 145-150). IEEE.

[8] Lykou, G., Moustakas, D. and Gritzalis, D., 2020. Defending airports from UAS: A survey on cyber-attacks and counter-drone sensing technologies. *Sensors*, *20*(12), p.3537.

[9] Hauzenberger, L. and Holmberg Ohlsson, E., 2015. Drone detection using audio analysis.

[10] Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, *60*, pp.91-110.

[11] Dalal, N. and Triggs, B., 2005, June. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). Ieee.

[12] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Communications of the ACM* 60, no. 6 (2017): 84-90.

[13] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

[14] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C., 2016. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14* (pp. 21-37). Springer International Publishing.

[15] Girshick, R., Donahue, J., Darrell, T. and Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).

[16] Ren, S., He, K., Girshick, R. and Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, *28*.

[17] Wang, C.Y., Bochkovskiy, A. and Liao, H.Y.M., 2022. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*.

[18] Rozantsev, A., Lepetit, V. and Fua, P., 2015. Flying objects detection from a single moving camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4128-4136).

[19] Mejias, L., McNamara, S., Lai, J. and Ford, J., 2010, October. Vision-based detection and tracking of aerial targets for UAV collision avoidance. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 87-92). IEEE.

[20] Gökçe, F., Üçoluk, G., Şahin, E. and Kalkan, S., 2015. Vision-based detection and distance estimation of micro unmanned aerial vehicles. *Sensors*, *15*(9), pp.23805-23846.

[21] Ganti, S.R. and Kim, Y., 2016, June. Implementation of detection and tracking mechanism for small UAS. In *2016 International Conference on Unmanned Aircraft Systems (ICUAS)* (pp. 1254-1260). IEEE.

[22] Unlu, E., Zenou, E. and Riviere, N., 2018. Using shape descriptors for UAV detection.

[23] Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X. and Pietikäinen, M., 2020. Deep learning for generic object detection: A survey. *International journal of computer vision*, *128*, pp.261-318.

[24] Dai, J., Li, Y., He, K. and Sun, J., 2016. R-fcn: Object detection via region-based fully convolutional networks. *Advances in neural information processing systems*, *29*.

[25] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R. and LeCun, Y., 2013. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*.

[26] Law, H. and Deng, J., 2018. Cornernet: Detecting objects as paired keypoints. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 734-750).

[27] Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S., 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).

[28] Schumann, A., Sommer, L., Klatte, J., Schuchert, T. and Beyerer, J., 2017, August. Deep cross-domain flying object classification for robust UAV detection. In *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (pp. 1-6). IEEE.

[29] Han, W., Li, J., Wang, S., Wang, Y., Yan, J., Fan, R., Zhang, X. and Wang, L., 2022. A context-scale-aware detector and a new benchmark for remote sensing small weak object detection in unmanned aerial vehicle images. *International Journal of Applied Earth Observation and Geoinformation*, *112*, p.102966.

[30] Magoulianitis, V., Ataloglou, D., Dimou, A., Zarpalas, D. and Daras, P., 2019, September. Does deep super-resolution enhance uav detection?. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (pp. 1-6). IEEE.

[31] Coluccia, A., Fascista, A., Schumann, A., Sommer, L., Dimou, A., Zarpalas, D., Akyon, F.C., Eryuksel, O., Ozfuttu, K.A., Altinuc, S.O. and Dadboud, F., 2021, November. Drone-vs-bird detection challenge at IEEE AVSS2021. In *2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (pp. 1-8). IEEE.

[32] Akyon, F.C., Eryuksel, O., Ozfuttu, K.A. and Altinuc, S.O., 2021, November. Track boosting and synthetic data aided drone detection. In *2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (pp. 1-5). IEEE.

[33] Craye, C. and Salem Ardjoune. Spatio-temporal semantic segmentation for drone detection. In 2019 16th IEEE International conference on advanced video and signal based surveillance (AVSS), pages 1–5. IEEE, 2019

[34] Yang, L., Ma, R. and Zakhor, A., 2022. Drone object detection using rgb/ir fusion. *arXiv preprint arXiv:2201.03786*.

[35] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C., 2016. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14* (pp. 21-37). Springer International Publishing.

[36] Guvenc, I., Koohifar, F., Singh, S., Sichitiu, M.L. and Matolak, D., 2018. Detection, tracking, and interdiction for amateur drones. *IEE*

[37] Andraši, Petar, Tomislav Radišić, Mario Muštra, and Jurica

Ivošević. "Night-time detection of uavs using thermal infrared camera." *Transportation Research Procedia* 28 (2017): 183-190.*E Communications Magazine*, *56*(4), pp.75-81.

[38] Carrio, Adrian, Yucong Lin, Srikanth Saripalli, and Pascual Campoy. "Obstacle detection system for small UAVs using ADS-B and thermal imaging." *Journal of intelligent & robotic systems* 88 (2017): 583-595.

[39] Svanström, F., Englund, C. and Alonso-Fernandez, F., 2021, January. Real-time drone detection and tracking with visible, thermal and acoustic sensors. In *2020 25th International Conference on Pattern Recognition (ICPR)* (pp. 7265-7272). IEEE.

[40] Redmon, J. and Farhadi, A., 2017. YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7263-7271).

[41] Jiang, N., Wang, K., Peng, X., Yu, X., Wang, Q., Xing, J., Li, G., Zhao, J., Guo, G. and Han, Z., 2021. Anti-UAV: A large multi-modal benchmark for UAV tracking. *arXiv preprint arXiv:2101.08466*.

[42] Wang, Y., Chen, Y., Choi, J. and Kuo, C.C.J., 2019. Towards visible and thermal drone monitoring with convolutional neural networks. *APSIPA Transactions on Signal and Information Processing*, *8*, p.e5.

[43] Sommer, L. and Schumann, A., 2020, September. Deep learning-based drone detection in infrared imagery with limited training data. In *Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies IV* (Vol. 11542, p. 1154204). SPIE.

[44] Yang, L., Ma, R. and Zakhor, A., 2022. Drone object detection using rgb/ir fusion. *arXiv preprint arXiv:2201.03786*.

[45] Shah, S., Dey, D., Lovett, C. and Kapoor, A., 2018. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and Service Robotics: Results of the 11th International Conference* (pp. 621-635). Springer International Publishing.

[46] Bondi, E., Dey, D., Kapoor, A., Piavis, J., Shah, S., Fang, F., Dilkina, B., Hannaford, R., Iyer, A., Joppa, L. and Tambe, M., 2018, June. Airsim-w: A simulation environment for wildlife conservation with uavs. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies* (pp. 1-12).

[47] Zhu, J.Y., Park, T., Isola, P. and Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232).

[48] Gaidon, A., Lopez, A. and Perronnin, F., 2018. The reasonable effectiveness of synthetic visual data. *International Journal of Computer Vision*, *126*(9), pp.899-901.

[49] Kniaz, V.V., Knyaz, V.A., Hladuvka, J., Kropatsch, W.G. and Mizginov, V., 2018. Thermalgan: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops* (pp. 0-0).

[50] Özkanoğlu, M.A. and Ozer, S., 2022. InfraGAN: A GAN architecture to transfer visible images to infrared domain. *Pattern Recognition Letters*, *155*, pp.69-76.

[51] Isola, P., Zhu, J.Y., Zhou, T. and Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

[52] https://www.blender.org/

[53] Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z. and Tu, Z., 2015, February. Deeply-supervised nets. In *Artificial intelligence and statistics* (pp. 562-570). PMLR.

[54] www.cgtrader.com. (n.d.). Drone | CGTrader. [online] Available at: https://www.cgtrader.com/items/3653841/download-page [Accessed 8 May 2023].

[55] sketchfab.com. (n.d.). DJI INSPIRE 2 with ZENMUSE X5S - Download Free 3D model by geoffreycouppey. [online] Available at: https://sketchfab.com/3d-models/dji-inspire-2-with-zenmuse-x5s-3979efe28b3a4221bdd462638582d0a6 [Accessed 8 May 2023].

[56] Wisniewski, M., Rana, Z.A. and Petrunin, I., 2021, October. Drone Model Identification by Convolutional Neural Network from Video Stream. In *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)* (pp. 1-8). IEEE.

[57] Isaac-Medina, Brian KS, Matt Poyser, Daniel Organisciak, Chris G. Willcocks, Toby P. Breckon, and Hubert PH Shum. "Unmanned aerial vehicle visual detection and tracking using deep neural networks: A performance benchmark." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1223-1232. 2021.

[58] Redmon, J. and Farhadi, A., 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

[59] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[60] Zhang, Z., Xie, Y., Xing, F., McGough, M. and Yang, L., 2017. Mdnet: A semantically and visually interpretable medical image diagnosis network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6428-6436).