drones

MDPI

Article

DRONET: Multi-Tasking Framework for Real-Time Industrial Facility Aerial Surveillance and Safety

Simeon Okechukwu Ajakwe, Vivian Ukamaka Ihekoronye, Dong-Seong Kim and Jae Min Lee *

Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi 39253, Korea; simeon.ajakwe@kumoh.ac.kr (S.O.A.); vivian.ihekoronye@kumoh.ac.kr (V.U.I.); dskim@kumoh.ac.kr (D.-S.K.)
* Correspondence: ljmpaul@kumoh.ac.kr

Abstract: The security of key and critical infrastructures is crucial for uninterrupted industrial process flow needed in strategic management as these facilities are major targets of invaders. The emergence of non-military use of drones especially for logistics comes with the challenge of redefining the anti-drone approach in determining a drone's harmful status in the airspace based on certain metrics before countering it. In this work, a vision-based multi-tasking anti-drone framework is proposed to detect drones, identifies the airborne objects, determines its harmful status through perceived threat analysis, and checks its proximity in real-time prior to taking an action. The model is validated using manually generated 5460 drone samples from six (6) drone models under sunny, cloudy, and evening scenarios and 1709 airborne objects samples of seven (7) classes under different environments, scenarios (blur, scales, low illumination), and heights. The proposed model was compared with seven (7) other object detection models in terms of accuracy, sensitivity, F1-score, latency, throughput, reliability, and efficiency. The simulation result reveals that, overall, the proposed model achieved superior multi-drone detection accuracy of 99.6%, attached object identification of sensitivity of 99.80%, and F1-score of 99.69%, with minimal error, low latency, and less computational complexity needed for effective industrial facility aerial surveillance. A benchmark dataset is also provided for subsequent performance evaluation of other object detection models.

Keywords: aerial surveillance; anti-drone communication; drone detection; deep learning; facility; security; weapons

Citation: Ajakwe, S.O.; Ihekoronye, V.U.; Kim, D.-S.; Lee, J.M. DRONET: Multi-Tasking Framework for Real-Time Industrial Facility Aerial Surveillance and Safety. *Drones* **2022**, *6*, 46. https://doi.org/10.3390/drones6020046

Academic Editor: Abdessattar Abdelkefi

Received: 13 January 2022
Accepted: 10 February 2022
Published: 15 February 2022

1. Introduction

In recent times, the influx of drones and its derivatives in the airspace due to the emergence of drone transportation system (DTS) for logistics and other civilian-military purposes [1] has sparked global concerns on the viability and verity of the existing anti-drone designs that focus mainly on detecting drones rather than the object attached to or conveyed by the drone. This surging influx has resulted in illegal usage and unsolicited intrusion of drones into private properties, as well as protected areas [2]. Drones are used in carrying out series of remotely coordinated sophisticated attacks especially on critical infrastructures such as nuclear plants, thermal and hydro-electric stations, industrial equipment, telecommunication gadgets, monumental buildings, tourist sites, etc.—thereby disrupting socio-economic activities and causing untold hardship [3,4]. The perceived threat of a drone in the airspace is a function of the object attached to it [5]. Therefore, the object attached to or conveyed by a drone is key to determining the harmful status of a drone in the airspace—hence the need to re-invent the approach to anti-drone system design not just to detect drones but the object airborne by the drone.

The increase in the challenge to quickly detect and identify such unmanned aerial vehicles (UAV) to which drones belong is occasioned by the dynamism in its underlying technologies [6,7]. Despite sophistication in security technologies, the incidences of attacks on industrial facilities and key infrastructures with the aid of drones and its derivatives are

on the increase. This is demonstrated in the premeditated crashing of a drone into a French nuclear plant [8] and the flying of drones over thirteen nuclear plants in Paris [9] and Saudi Arabia [10] with little or no effort to counter them. Up to now, there has only been a successful interception of a drone attack once, and it was military grade [5]. Inaccurate and non-real-time detection, lack of simultaneous adaptive multi-drone detection, and safe-channel neutralization, as well as poor feedback and response procedures, contributed largely to the occurrence of these incidences [5]. Developing and deploying an advanced detection approach in high priority areas with key and critical infrastructures for seamless and preemptive aerial communication in curtailing threats cannot be over-emphasized.

Drone detection entails estimating a drone's location for defense and detecting the drone before it flies into a sensitive area. On the other hand, drone identification is concerned with determining the legality or illegality, harmfulness or otherwise of a drone in sight which invariably determines the neutralization strategy to adopt (jamming, hunting, or re-assembling to overwrite control) through flexible secured authentications to counter and keep the illegal drone or its derivatives within the authorized area [11,12] as depicted in Figure 1. These concepts are the integral components of an anti-drone system, which is a multi-tasking, multi-modal, and complex hard real-time critical mission network-controlled system used in engaging drones and other aerial vehicles in the airspace [13]. This suggests that attempting to address the detection concerns in isolation without considering the other components of the anti-drone system is counterproductive—hence a multi-tasking approach.

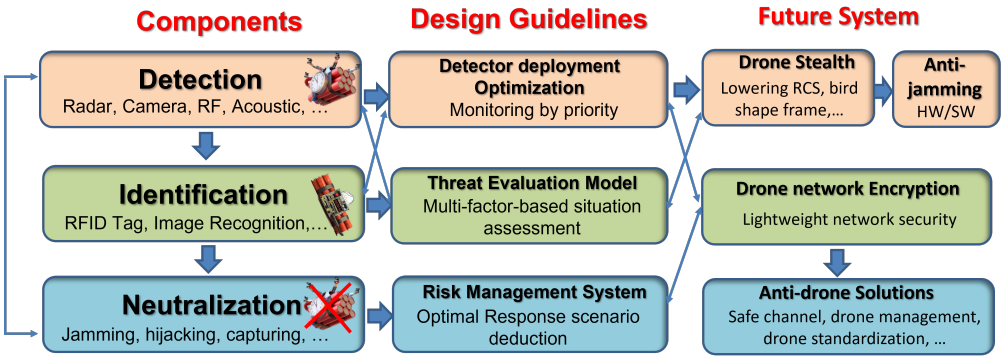

Figure 1. Components of an anti-drone system (Adapted from [5]).

Previous research efforts have been directed towards industrial facility surveillance and safety. The authors [14] deployed drones for mapping raised bog vegetation communities using machine learning and deep learning classifiers. In another related work, Sergio et al. [15] proposed a remote management architecture solar power plants surveillance using a fleet of UAVs. Though these works provided an invaluable insight into surveillance, it focused on using drones to detect objects rather than detecting drones and cannot perform automatic simultaneous detection and neutralization, as it requires a human expert's assistance. Other attempts to detect drones in the airspace have focused mainly on drone detection with little emphasis on the attached objects/payload, which is critical in guaranteeing industrial facility surveillance and safety [13,16–19].

With the dynamic changes in intelligent transportation systems, the use of drones and its derivatives for logistics purposes and transportation has changed the landscape of anti-drone system designs. This disruption in technology not only heightens the complexity of drone detection approach but also stresses the need for timely attached object/payload visual identification. With payload identification comes the problem of tiny, distant, and obscure object identification especially in a hazy/cloudy environment. In addition, there is a lack of simultaneous adaptive and scenario-based drone detection and neutralization in the existing solutions [5]. Inaccessibility and unavailability of a reliable and robust vision-based drone dataset for performance evaluation are a challenge. Finally, the need to

de-militarize the existing anti-drone system design to meet with the increase in demands is the motivation for this study.

Several state-of-the-art drone detection approaches are deployed to match up with the ever-dynamic drone technologies. These approaches blend in radar technology [20], vision technology [12], radio frequency (RF) technology [11], thermal technology [21], and acoustic technology [22] convergence to achieve better results (availability, mobility, installation flexibility, and detection precision/scope), thereby cushioning the inherent limitations of each technique. Examples include but are not limited to kinetic means, drone vs. drone, electronic warfare, cyber-warfare techniques, directed energy weapons using high-powered microwave or lasers, etc. [3,23]. However, these approaches are highly militarized, have legislative bottlenecks in its implementation for civilian deployment, and lacked simultaneous multi-drone detection and an identification mechanism. Only a vision-based detection technique provides exact and clear visual representation of a detected object [24], which is of critical importance for an adaptive countering strategy. Unarguably, timely, early, and accurate detection and safe-channel neutralization strategy is needed to minimize incidences of drone attacks on key infrastructures.

The danger posed by a drone in the airspace is a function of its illegality and/or harmfulness. Illegality is a function of license issuance/permit, navigation boundary, route, or flight path. Harmfulness, on the other hand, is a function of the object attached, payload being conveyed, and the source of the drone [25]. Therefore, a modern drone detection and identification system (as an event-triggered critical-mission based real-time system) must have specialized environment-sensitive multi-drone detection capability, intuitive multi-drone visual identification ability, multi-drone defensibility, co-operative inter-connectivity and security, system portability and mobility, as well as non-militarized situation-based response strategy to handle emerging UAVs. Such task of real-time detection and adaptive neutralization of illegal drones using civilian approaches demands an investigative and innovative alternative that can meet up with emerging threats to cushion the technical competition between anti-drone and drone industries. Artificial intelligence, especially a deep Convolution Neural Network (CNN), has proven to be a veritable tool for solving these aforementioned complex and dynamic-driven challenges with high precision and accuracy [26].

This paper therefore seeks to provide such novel approach for real-time multi-drone detection, payload identification, and adaptive neutralization leveraging on the potential of CNN for safety of key industrial facilities with the following contributions:

1. To develop a deep learning model to detect tiny drones and as well as recognize the attached objects in real-time;
2. To formulate a perceived threat analysis model for determining the harmful status of a detected drone in the airspace;
3. To formulate a framework for determining the legality status of a detected drone in the airspace;
4. To establish a scientific basis for adaptive neutralization response to counter a harmful drone in the airspace;
5. To provide a benchmark dataset for subsequent drone detection performance evaluation in the public domain.

The rest of this paper is divided into Related Works presented in Section 2, Proposed Method and System Design in Section 3, Result Discussion and Performance Evaluation in Section 4, and finally Conclusions in Section 5.

2. Related Works on Drone Detection Techniques and Technologies

A vision-based drone detection technique is an object detection and pattern recognition technique that use infra-red or electro-optical camera sensors (as seen in Figure 2) to automatically identify a moving object against its background [27,28]. Just like other drone detection techniques, it has inherent issues of occlusion, inability to distinguish smaller objects, and detection range [12].

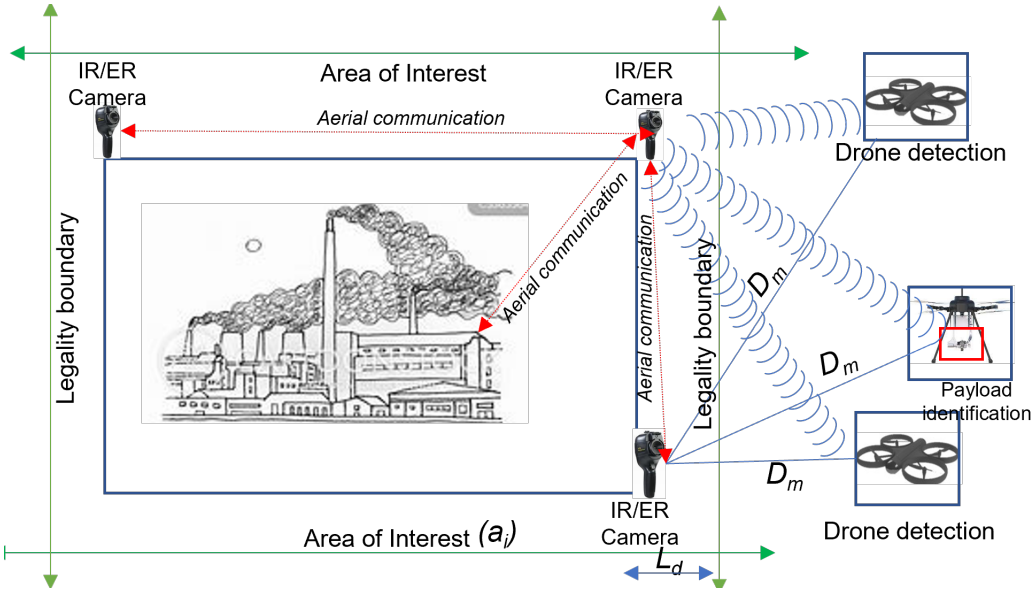

Figure 2. Vision-based drone detection and aerial communication system illustrating airborne object detection around an industrial facility.

Visual-based object detection is usually achieved by combining three different computer vision approaches, namely traditional feature engineering (motion features and appearance features) [27], Machine Learning (ML) approach [28], and several Deep Learning (CNN) approaches to achieve faster detection with higher accuracy [29]. A typical anti-drone system for an industrial facility inspection and surveillance adopts these approaches as its underlying drone detection operational paradigm.

A traditional feature engineering approach to object detection makes use of an object's appearance properties such as colors, contour lines, geometric forms or edges, etc. and its motion features such as analyzing the consecutive frames of the captured image) [30]. Examples include appearance-based, template-based, part-based, region-based, and contour-based methods. The drawback of traditional feature engineering lies in its great difficulty in distinguishing drones from other similar small objects like birds and/or other flying objects in obscure/camouflage backgrounds without motion information. In addition, there is difficulty in differentiating a moving drone from a gliding bird [31].

Similarly, the ML approach fuses these traditional computer vision methods to achieve better object classification and detection results [32]. Examples of ML Approaches include Haar Wavelet features [33]; Haar-like features and motion information, Implicit Shape Models, Histogram of Oriented Gradients (HOG) [34], Covariance descriptor, and Extended Histogram of Gradients (ExHOG) [35]. The drawbacks of the ML approach include time-cost in image extraction, limited range of image detection, low resolution of extracted image feature, and scalar/static image detection, which makes it unfit for real-time drone detection.

Modern drone detection methods based on CNN are grouped as One-stage Detectors, Two-stage detectors, or Hybrid detectors. One-stage detectors, otherwise called Single Stage Detectors (SSD), perform sequential image detection and processing, which makes them very fast but with low detection accuracy. SSDs include Single Shot Multibox Detectors, RetinaNet, MobileNet, and You Only Look Once (YOLO) family [36]. Two-stage Detectors also known as Region-based Convolution Neural Networks (R-CNN) [37] perform more accurate image detection and feature extraction than SSD but require more memory and power to run, use complex pipeline, and adopts a sliding window method that is expensive—thus rendering two stage detectors unsuitable and impracticable for real-time detection necessary for a time sensitive and mission critical anti-drone system. Common examples include R-CNN and Faster R-CNN. Hybrid detectors fuse SSD and R-CNN to blend speed

and accuracy in performance [38]—for example, VGGNet, ShuffleNetV2, COCO Model, ResNet-101, and Yolov5.

Several studies have made attempts at improving the detection accuracy of YOLO architecture for distant and tiny objects. Authors [39] proposed a strip bottleneck (SPB)-YOLO model to solve the problem of scale variations and dense distributions of objects. The model is an end-to-end detector that has the SPB module that used the attention mechanism approach to solve the dependency of scalar variations of UAV images. In addition, by the up-sampling of the detection head of YOLOv5 in the addition of a detection head based on the Path Aggregation Network, the challenge of dense object distribution was mitigated. Authors [40] in their work demonstrated that the YOLOV5 architecture is not only a fast object detector but also achieved accuracy that is plausible to that of Faster R-CNN. In a similar study, authors [41] proved that the YOLO model can adequately classify and detect multiple objects in real-time, which makes it suitable as an underlying detector for the anti-drone system. However, the disparity experienced in the detection and classification of flying objects in different weather conditions and altitudes is still a research gap.

The surveillance of an industrial facility site or key infrastructure using a static surveillance system is counter-productive for efficient wide-area coverage of a distant impending danger, especially illegal drones with harmful objects. In an anti-drone system, determining the legality/illegality and harmfulness or otherwise of a detected flying drone can be achieved via active or passive identification. With the use of Radio Frequency Identification (RFID) tags, the legality of an approaching drone can be periodically and passively identified [42]. However, for prompt and proactive counter drone decision, an active identification strategy is the key in determining a hazard level through proper threat analysis of the detected drone via drone tracking/flight estimation [43]. Hence, exact visual representation of acquired surveillance information is a critical resource.

Therefore, a robust anti-drone model should have the capacity to visually identify a wide range of flying objects irrespective of their features, and its identification system should be able to accurately recognize the payloads attached to the drone, access its proximity to a defined perimeter, determine its threat level based on defined metrics, and take responsive and adaptive decision about it [44]. This is the motivation of this proposed approach and design.

3. Materials and Methods

3.1. Proposed System Design

An ideal drone detection and aerial communication system (otherwise known as anti-drone system) is a complex, multi-tasking, and multi-modal system that fuses several technologies such as heterogeneous sensors, networks, security protocols, data acquisition and synchronization mechanism, feature extraction and prediction technology, tracking controllers, databases, etc., while engaging a drone in the airspace. These interwoven technologies help an anti-drone system to achieve its main task of drone detection, localization or tracking, and decision-making. Hence, it is imperative for an anti-drone system to have automatic and high detection accuracy, high density deployment management strategy, ultra-fast network feedback mechanism, prompt, and adaptive safe-channel neutralization mechanics, and operate in a cost-efficient manner.

This study focuses mainly on the drone detection component and partly on the neutralization scheme of an anti-drone system with emphasis on the deep learning model for accurate determination of the status a sighted drone within a field of view. However, such an important concept cannot be presented in isolation considering its relationship with other critical components of a typical anti-drone system. Hence, the neutralization component is partly discussed based on the changing landscape in the anti-drone system development sector.

The proposed drone detection design as condensed in Figure 3 provides the surveillance information flow from the input, detection process, and neutralization response strategy, respectively.

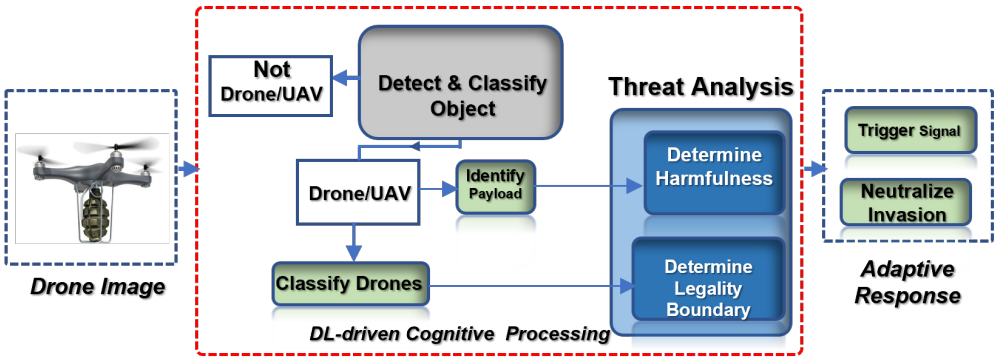

Figure 3. Proposed aerial drone detection design for industrial facility inspection highlighting the flow of input, process, and output components.

During surveillance, an anti-drone system acquires images/signals within its defined perimeter (field of view) using different underlying data acquisition technologies (cameras, radars, LIDARS, etc.) and sends such data to the ground station via a secured network where further processing is carried out. With a vision-based anti-drone system, visual images of drones and other aerial objects are captured within the peripheral and central vision as input using infra-red or electro-optical camera sensors at different distance, altitude, and climatic conditions (sunny, cloudy, evening) as illustrated in Figure 2. These visual data are transmitted to a ground station via a secured network for the processing phase to commence. Drone capturing and transmission operation is carried out in real time as long as there is an object in the airspace. Hence, a reliable and fast communication network is key because of the cost of a split second's delay in data transmission.

Then, the acquired input is fed into the proposed underlying deep learning model for processing via feature extraction as discussed in Section 3.2. The functionality of the proposed object detector model is to accurately detect and distinctively classify not just the different drones in sight but also to identify the potential attached object's harmfulness or otherwise, without being hindered by obscure climatic conditions or the object's altitude. Based on the proposed detector framework, outputs are generated according to the class of probability for the classification and identification task; *drone* or *not drone*, $harmful$ or *not* $harmful$, as discussed later.

The output from the feature extraction process is thereafter transferred to the neutralization/decision-making component where it is utilized by the neutralization superpositioning strategy as presented in Section 3.3 to determine the best alternative approach to adopt in countering the drone from invading an industrial facility area such as a warehouse. In addition, the neutralization component carries out a drone's legality status determination using underlying technologies and metrics in conjunction with the feedback from the detection module prior to taking a cognitive action.

This flow suggests an intuitive, real-time, multi-tasking, and synergistic approach to checkmating targeted attacks and other illicit activities around an industrial environment without disrupting or undermining the flow of the airspace activities such as logistics and transportation that are carried out using drones and its derivatives as captured in Figure 3 and elaborated in subsequent sections.

3.2. Proposed Drone Detection Model

In this study, the proposed drone detection model, DRONET, is designed to overcome the problem of detection of tiny distant objects under harsh or hazy weather conditions that exist in YOLOv5. As an object detector framework, YOLO architecture is a robust model that has undergone several evolutions to enhance object detection by integrating several breakthroughs in computer vision for ease of use, model ensembling, precision, and hyperparameter characterization. Its series are namely small (YOLOv5s), medium (YOLOv5m), large (YOLOv5l), and extra-large (YOLOv5x) network architecture with

the same structure as summarized by Figure 4, but with different depth and width size, which consequently contributes to the model's overall performance in terms of size, speed, and accuracy.

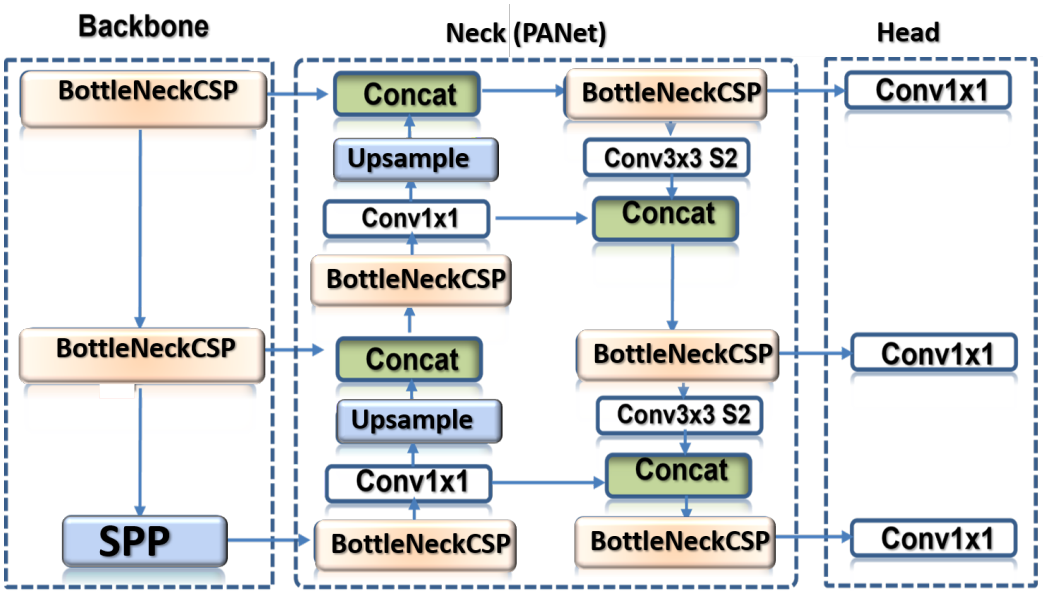

Figure 4. Compact drone detector model showing backbone, neck, and head as its components, networks, and convolutions.

Although YOLO as an object detector has been proven to have the capability of extracting 140 frames per second (fps) [45,46] in real time, which is very crucial considering the quantity of data needed for the surveillance operation, its drawback is feature extraction from very tiny objects with high accuracy [47]. Therefore, this proposed model is designed to tackle this problem inherent in YOLOv5, which is pivotal in an anti-drone system design while retaining its robust operational capability. In addition, the similarity and disparity, and tininess of objects in the airspace created the necessity for instance segmentation; that is, the need for explicit detection, exact classification, and precise localization of various object instances in an image—hence the addition of Path Aggregation Network to the proposed custom drone detector to enhance the propagation of low-level features for improved performance as captured in Figure 4.

The proposed detector model has three distinct blocks (as seen in Figure 4); *Backbone*, *Neck*, and *Head*. The *Backbone* uses Cross Stage Partial Network (CSPDarknet) for image feature extraction. The *Neck* uses feature pyramid network (FPN) and Path Aggregation Network (PANet) to perform feature aggregation. The *Head* does the predictions using anchor boxes for eventual object detection.

The captured and transmitted image (via the secured network by the electro-optical camera) as an input goes into the CSPDarknet (designed based on DenseNet architecture) to upgrade the learning ability of the CNN by reducing the complexity of large gradient information, thereby truncating gradient flow of the optimized network while preserving accuracy. CSPDarknet separates the feature map of the base layers into two; one goes through the dense block while the other part is fused with a feature map and relayed to the next stage as shown in the feed-forward propagation and weight update equation expressed herewith in Equation (1):

$$\begin{aligned} \Rightarrow I_j &= W_j \times [I_{0''}, I_1, \ldots, I_{j-1}] \\ I_t &= W_t \times [I_{0''}, I_1, \ldots, I_j] \\ I_u &= W_u \times [I_{0'}, I_t,] \end{aligned} \tag{1}$$

$$\begin{aligned} \Rightarrow W_{j'} &= f(W_j, g_{0''}, g_2, g_3, \ldots, g_{j-1}) \\ W_{j'} &= f(W_j, g_{0''}, g_2, g_3, \ldots, g_j) \\ W_{u'} &= f(W_u, g_{0'}, g_t); \end{aligned}$$

where I_j is the input of the $(j+1)$th dense layer; w is the weight; and g is the gradient information of the network. With CSPDarknet, the model size is scaled by reducing the image size, number of layers, and number of channels. After the feature extraction is carried out, the output is fed into the *Neck* for feature fusion or aggregation using the *PANet*.

PANet (as an image instance segmentation network) shortens the information path of the output from the backbone, improves the feature pyramid process, and enhances accuracy of image localization. It adopts bottom-up path augmentation, adaptive feature pooling, fully connected fusion, and mask prediction as shown in Figure 5.

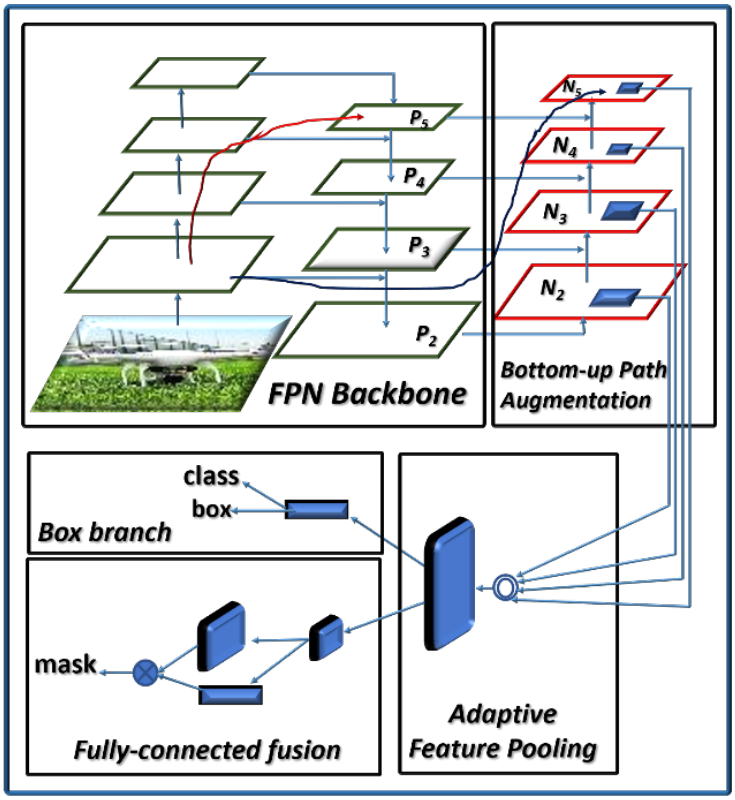

Figure 5. Path Aggregation Network showing the underlying components and process; FPN Backbone, Bottom-up Path Augmentation, Adaptive feature pooling, Box branch, and Fully connected fusion.

This framework tries to boost the localization capability of feature extraction hierarchy by propagating the low-level patterns. With bottom-up augmentation, feature levels are created with the same spatial size denoted as P_2 to P_5. Then, the spatial size is reduced to generate new feature maps; N_2 to N_5. Each N_i goes through convolution layers with strides for spatial size reduction. Next, each of the P_i and reduced map are laterally summed to produce a fused feature map, which goes through another convolution layer to create new N_{i+1} for the next sub-network. To perform adaptive feature pooling, each proposal is mapped to different feature levels. Thereafter, a fusion operation is carried out

where the feature grids are fused together at various levels using Region of Interest (ROI) alignment. The fully connected layer does the actual prediction by allowing parameters in its layers to train more samples, thereby increasing the generality of distinct differentiation of the foreground from the background mask characteristics. Unlike the standard YOLOv5 architecture, PANet is added to YOLOv5 architecture to improve the image feature process by taking a single object of an arbitrary size as input and outputs it in proportional sized feature maps of multiple levels in a fully convolution manner. The outputs from the *Neck* are fed into *Head*.

Finally, the *Head* carries the eventual detection and prediction of different drones and the attached objects by applying anchor boxes on the features and generates output vectors containing class probabilities, object scores, and bounding boxes with a sort of characterization to determine the overlap expressed in Equation (2):

$$\Rightarrow I_{ou} = \frac{A_i}{A_u}; \tag{2}$$

where A_i is Area of Intersection; A_u is Area of Union; and I_{ou} is Intersection Over Union, which returns a value between 0 and 1 with a common threshold of $>$0.5. To ensure nonlinearity in the network, leaky ReLU and Softmax activation function is adopted to classify multi-class outputs. Softmax function returns the probability of each class expressed by the Equation (3):

$$\sigma(\overline{S}_i) = \frac{e^{s_i}}{\sum_{j=i}^{k} e^{s_j}}, \tag{3}$$

where σ is softmax, $(\overline{S}_i)$ is the input vector, e^{s_i} is the standard exponential function for input vector, k is the number of classes in the multi-class classifier, and e^{s_j} is the standard exponential function for output vector. In addition, to perform optimization, the ADAM and stochastic gradient descent optimization function is used.

3.3. Airborne Object Identification and Safe-Channel Response Strategy

As a drone approaches an industrial environment, it is pertinent to determine in real-time the source of the drone, the proximity of the drone to the facility, the drone's intention, and the potential objects airborne by the drone (such as explosives, weapons, guns, radio-actives, etc.). An accurate determination of these parameters is a precursor to the response strategy to be taken intently. Having detected and classified the object in the airspace as a drone (see Section 3.2), the model proceeds to check the drone's harmful and legal status. In order to achieve this, perceived danger and threat analysis of the drone in the airspace is carried out using Equation (4):

$$\Rightarrow \delta D_t = \Sigma(P_t, P_a), \tag{4}$$

where δD_t is overall danger posed by a drone in sight, P_t is a drone's physical feature threat quantifier, and P_a is the facility perimeter/Area of Interest (a_i) classifier based on priority level. While P_a value determines the legality/illegality of a detected drone, the P_t value defines the harmfulness or otherwise of such drone.

3.3.1. Perceived Threat Analysis

The perceived physical threat or damage (P_t) likely to occur by a drone in the airspace is achieved through a super-positioning strategy that involves categorization of the available drone according to its model, its technology characteristics, and the airborne object properties. The sum of these parameters suggests either a high or low priority level. Thus, the value of P is further expressed in Equation (5):

$$P_t = (P_{object} + D_{path})^{D_{time}}, \tag{5}$$

where P_{object} is physical threat level induced by drone physical features, D_{path} is flight path, and D_{time} is response time with their maximum value set at α, β, and 1, respectively. However, the value of P_{object} is computed based on several factors (such as drone's weight/kinetic energy, size, noise, loaded object, and scanability) depending on the technology deployed. This is summarized in Equations (6)–(9). Assessing the drone's kinetic energy, which conveys useful information such as drone's weight/mass and speed, is given by Equation (6).

◯ **Kinetic Energy ($W_{kinetic}$)**:

$$p_{object,kinetic} = (observed\ kinetic\ energy) \div 1400J \times W_{kinetic}, \tag{6}$$

$\Delta d_{object,kinetic}$ is a clear indication that the drone in sight is carrying an object but such an alert is insufficient for a proper airborne response scenario.

The sound generated by the drone's rotor can be determined as:

◯ **Noise Level (W_{noise})**:

$$p_{object,noise} = (1 - (observed\ noise\ level) \div 80dB) \times W_{size}, \tag{7}$$

An increase in $d_{object,noise}$ value suggests that an object in air is disrupting the airspace. However, in a noisy environment, the $d_{object,noise}$ value alone can be of little help for a proactive counter response.

The object attached to drone is expressed as:

◯ **Loaded Objects (W_{loaded})**:

$$p_{object,loaded} = (level) \div 4 \times W_{loaded}, \tag{8}$$

If the identified loaded object is harmful (such as gun, missile, explosives, radio-actives, etc.), then swift reaction and extra caution are taken to re-route the drone to an area with less crowds and facilities to minimize casualty.

Where radar technology is available,

◯ **RF Scannable ($W_{scannable}$)**:

$$p_{object,scannable} = \begin{cases} 0, \ if\ scannable \\ 1.0\ W_{scannable}, \ if\ unscannable, \end{cases} \tag{9}$$

Hence, the overall or total value of p_{object} is expressed as:

$$\therefore p_{total} = max[\alpha(p_{(object,kinetic)} + p_{(object,noise)} + p_{(object,loaded)} + p_{object,scannable}) \times N_{drone}]; \tag{10}$$

with N_{drones} is the number of swarming drones. Although these factors are worthwhile, only visual representation of a loaded object, $p_{object,\,loaded}$, can provide the precise information needed for adaptive response based on an airborne object. Hence, the proposed scheme factored into consideration of the value of P_{object} to be synonymous with the value of $p_{object,\,loaded}$. Since this study focused on a vision-based technique, d_{path} and d_{time} are not covered, yet the combination of these with accurate determination of p_{object} provides a precise value for P_t. Hence, the drone's physical feature threat, P_t, is considered to be the value of the loaded/attached object, $p_{object,\,loaded}$.

3.3.2. Drone Area Mapping

To decide the priority level of the industrial facility perimeter (P_a) otherwise known as *Area of Interest* (a_i), the legality of a detected drone in the airspace is determined by calculating the estimated distance between the approaching drone and the *Threat zone* as seen in Figure 6 and expressed in Equation (11):

$$\Rightarrow L_d(\overline{AB}) = \frac{1}{1 + exp(d - \frac{D_m}{2})}, \tag{11}$$

where L_d is the legality boundary/detection range which has a high value for a priority area such as facility environment, d is the distance between area of interest and detected drone, and D_m is a maximum detection range of the system. From Figure 6, the maximum detection range, D_m is the sum of *Threat zone* $(\overline{AB})$ and the *Allowable zone*, $(\overline{BC}, \overline{BD}, or \overline{BE})$ as shown in Equation (12):

$$D_m(\overline{AC} = \overline{AD} = \overline{AE}) = a_d + t_d, \tag{12}$$

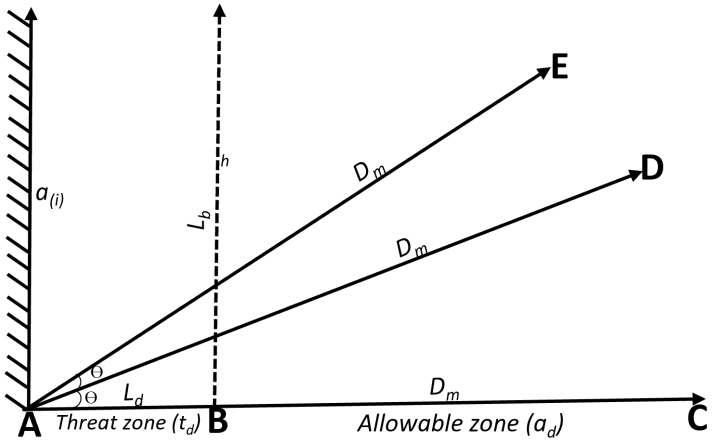

Figure 6. Area Mapping for adaptive drone neutralization depicting detection measurement logic with point A as a position of a mounted electro-optical camera, point B is airspace boundary, while point C, D, and E represent the detected drones in the airspace.

Hence, the value of d is expressed as:

$$d = a_d - t_d \tag{13}$$

However, for accurate estimation of d, the height (h) and the angle (θ) of detected drone are factored into consideration:

$$\therefore d = (a_d - t_d) \times sin(\frac{h}{D_m}) \tag{14}$$

As the value of d shrinks and tends towards 0 at point B (Lb), referred to as the legal boundary, an automatic control mission event is triggered to take appropriate neutralization action. However, when the drone is within the allowable/legal zone (a_d), the proposed anti-drone scheme stays on standby mode.

Therefore, the proposed adaptive neutralization scheme utilizes the value of the detection output (from Section 3.2), perceived threat analysis state, $p_{object,loaded}$ (from Section 3.3.1), and the value of legality status, (L_d) to determine the best appropriate response as captured by Algorithm 1.

Algorithm 1: Proposed DRONET Algorithm.

Data: $Capture_{image}$
Result: Danger Level, Drone-det, Obj-det
initialization;
Perform Drone Detection and Classification and Payload Recognition;
if $Capture_{image}$ *equals* $Drone_{det}$ **then**
  Perform Perceived Threat Analysis;
  Check Drone Harmful Status ($P_t = p_{object,loaded}$);
  if Obj_{status} *is* $Harmful$ **then**
    Check Drone Legality Status (P_a);
    if *If* $Legal_{status}$ *is Valid* **then**
      Trigger $Stand - byMode$;
      Continue *legalityboundarycheck*;
    else
      Trigger Automatic $DefenseMode$;
    end
  else
    if Obj_{status} *not* $Harmful$ *and* $Legal_{status}$ *is Valid* **then**
      Trigger *Drone Disarm and Rerouting Mode*;
    else
      Trigger *Harmless Mode for Hobby Drones*;
    end
  end
else
  Ignore $Capture_{image}$ in the airspace;
end

This is to ensure that unnecessary alarm is prevented when a legal or harmless drone is in the airspace unlike previous anti-drone design models that trigger an alarm once a drone is in sight.

3.4. Dataset Capturing, Description, and Pre-Processing

Two separate datasets for drone detection and payload identification were manually generated. For detection and classification, 5460 drone samples were manually captured at different altitudes (30 m–100 m) and scenarios (cloudy, sunny/normal, and evening) and under different environments to form our test bed. Seven (7) drone models, namely; Anafi extended, DJIFPV, DJIPhantom, Mavic2-Air, Mavic2-Air enterprise, Mavic2-Enterprise Zoom, and EFT-E410S, were used for this study.

Each of these drone models were flown into the air at different locations, at different heights and distance from the controllers, at different times of the day and days of the week, and under different climatic conditions to reflect the intended scenarios. The video sequences of each of these flight operations were captured and recorded in different time frames. This exercise was carried out in months to ensure that all scenarios were adequately covered to improve data quality, thereby avoiding errors, insufficiency of training data, and non-representative of data that can cause model overfitting. Next, each of the captured video's sequences were converted into a sequence of data frames using an appropriate software, Free Video to JPG Converter. These data frames contain the raw drone scenario images/samples that make up each sample size as shown in Table 1 and Figure 7 and stored accordingly.

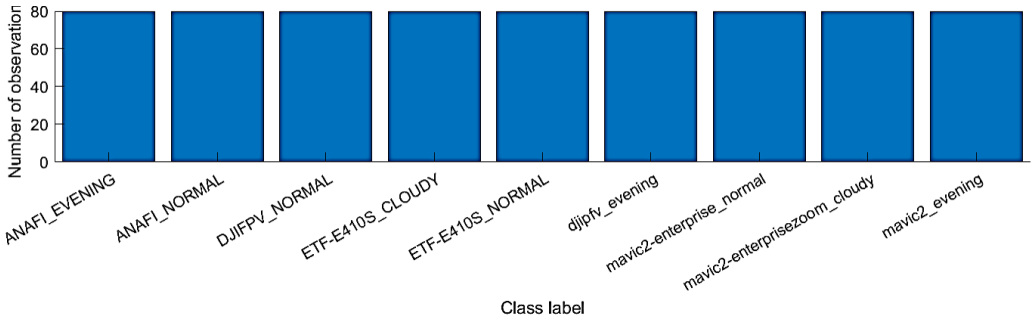

Figure 7. Drone detection dataset distribution for the models' training showing various drones models and climatic scenarios.

Furthermore, the data frames were sorted and cleaned by removing all images that neither had drones but only background nor met the expected requirements. Thereafter, each scenario-based data frame was labelled to generate normalized ground truth values ($x_{center}, y_{center}, width, height$) in pixels by creating bounding boxes around the objects using a *MakeSense* application. These transformed datasets are stored as text that represents the labelled dataset used for model training, validation, and testing.

For a payload identification dataset, 1709 samples of drones with different attached objects were manually captured, cleaned, and labelled using the same procedure as aforementioned. The attached payloads include guns, missile, bomb, hidden cameras, explosives, etc. as captured in Table 1 and Figure 8.

Table 1. Datasets' characterization.

Datasets Description and Distribution			
Attached Objects	**Sample Size**	**Drone Models/Scenario**	**Sample Size**
Bomb	90	Anafi-Extended/Evening	660
Explosives	100	Anafi-Extended/Sunny	600
Gun	340	DJIFPV/Sunny	600
Missile	35	ETF-E4410S/Cloudy	600
Secret camera	180	ETF-E4410S/Sunny	600
Sealed packages	180	DJIFPV/Evening	600
UAV	15	Mavic2-Enterprise/sunny	600
drones	950	Mavic2-Enterprise evening	600
Total	1790	Mavic2-Enterprise zoom/cloudy	600
		Total	5460

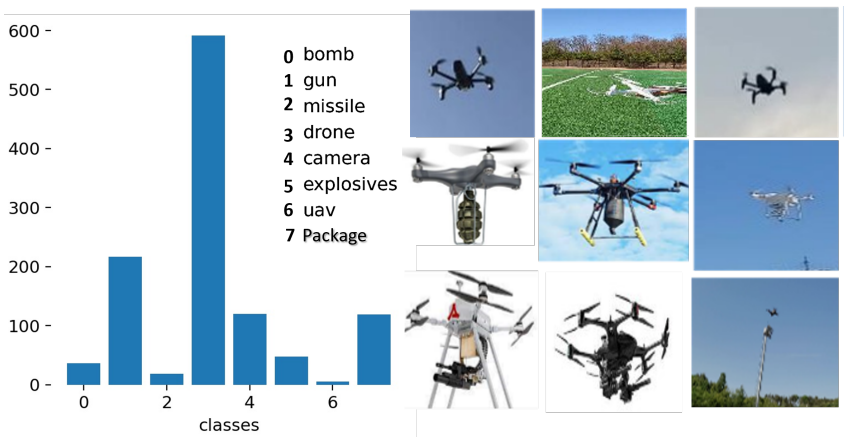

Figure 8. Payload recognition dataset distribution showing the different drone models and attached objects under different climatic conditions and altitudes.

Finally, data augmentation was carried out on both datasets by transforming the training data so as to expose the models to a wider range of semantic variation rather than isolated training. The dataset distribution for both drone detection and payload identification is shown in Figures 7 and 8, respectively.

Robust on-site capturing of an attached payload dataset was, however, hampered by government policy issues as regards flying of drones with harmful objects on-board—hence the 1709 sample size for attached objects as against 5460 drone samples.

3.5. Simulation and Experimental Setup

The Pareto principle (also known as the 80/20 rule which stipulates those 80% of effects usually comes from 20% cause) is applied for training and testing sets to ensure optimal performance and avoid model over-fitting. Simulation was carried out in a Python environment using a PyTorch 1.10 library on a system with specification; Intel(R) Core(TM) i5-8500 CPU @ 3.00 GHz, NVIDIA GeForce GT 1030, GPU CUDA:0 (Tesla K80, 11,441.1875 MB), 24 GB RAM, and Windows 10 operating system software. The hyper-parameters used for the proposed models' simulation are as summarized in Table 2.

Table 2. Models' hyper-parameters.

NO.	PARAMETERS	VALUES
1	Number of Epoch	100
2	Batch size	16
3	Image input size	416 × 416×3
4	Learning rate	0.01
5	Weight-decay	0.0005
6	Warmup-epochs	3.0
7	Warm-momentum	0.8
8	Box loss threshold	0.05
9	Optimization Function	Stochastic Gradient Descent

The same hyper-parameter was used across the selected models for training of the deep neural network. Hyper-parameter tuning was performed on the model. In addition, transfer learning was carried out by training the last layers of convolution blocks in the models to improve the model's learning ability on a new task through existing knowledge. Moreover, to reduce misrepresentation of detected objects and enhance the robustness of the model in detecting new outcomes with variety of characteristics, different data augmentations, such as horizontal and vertical flips (hsv_h = 0.015, hsv_s = 0.7, hsv_v = 0.4, degrees = 0.0), translation (translate = 0.1), scaling (scale = 0.5), shearing (shear = 0.0), etc., were carried out on the image prior to training. Lastly, to test the model's inference, best trained weights' values were used on the test sets.

4. Results and Discussion

This section presents a detailed discussion of the simulation results for the drone detection (see Section 4.1), weapons, and payload identification (see Section 4.2), performance evaluation with other models (see Section 4.3), and the neutralization response approach (see Section 4.4). The performance metrics used to evaluate the model's performance are mean average precision (mAP), accuracy, sensitivity (recall), F1-score, throughput (measured by floating point operations per second, FLOPS), latency/response time (measured in frame per second), and reliability. Thereafter, the model's performance efficiency was compared with other state-of-the-art models.

4.1. Multi-Drone Detection and Classification

The results from Table 3 represent the drone detection by DRONET in different climatic conditions.

Table 3. Multi-drone Detection and Classification.

Drone Detection and Classification Result			
Drones	**Scenarios**	**Recall (%)**	**mAP @0.5 (%)**
Anafi-Extended	Cloudy	100	99.1
DJIFPV		100	99.7
ETF-E410S		99.8	99.9
Mavic Air		99.8	93.5
Mavic Zoom		100	99.9
Anafi-Extended	Evening/Gloomy	99.2	99.1
DJIFPV		99.6	99.7
ETF-E410S		100	99.9
Mavic Air		93.4	93.5
Mavic Zoom		100	99.9

In a cloudy or hazy environment, the proposed model achieved the highest precise detection of 99.9% and sensitivity of 100% for *Mavic Zoom* drone and least detection accuracy of 93.5% and sensitivity of 99.8% for *Mavic Air* drone. In a gloomy environment or evening scenario, DRONET achieved an optimal detection accuracy value of 99.9% and sensitivity of 100% for *ETF-E410S* and *Mavic Zoom* drones and least detection accuracy of 93.5% and sensitivity of 93.4% for *Mavic Air* drone. In addition, with a 99.1% detection accuracy and 100% sensitivity value for $Anafi\ Extended$ drone recorded by DRONET, it attests to the ability of the proposed model for precise detection of tiny and distant drones since most of the drone models in this study are miniature and flown to a high altitude. Furthermore, the model also achieved similar prediction results for tiny drones in a gloomy condition and night scenarios for $Anafi\ Extended$.

To find a balance between the robustness of stochastic gradient descent and the efficiency of batch gradient descent of the proposed model, parameter tuning is carried out by splitting the training set into different batch sizes (as shown in Table 4), which are used to calculate the proposed model's error and update the model co-coefficients.

Table 4. Parameter tuning of DRONET.

Parameter Tuning of Proposed Model					
Batch Size	**Epoch**	**mAP @ 0.5 (%)**	**Precision (%)**	**Recall (%)**	**Box_Loss**
8	100	99.5	97.5	100	0.046
16	100	99.5	91.7	100	0.040
32	100	99.6	99.8	100	0.055
64	100	99.5	95.3	100	0.061
128	100	99.5	82.2	100	0.063

At a smaller batch size of 8, the learning process of DRONET converged quickly at the cost of noise (Box_{Loss}) of 0.046 in the training process. With an increase in the batch size, for instance 128, the proposed model converged slowly with an accurate estimation of the error gradient of 0.063 as highlighted in Figures 9 and 10.

This attests to the stability of the model in learning from the training data. In addition, the values from the confusion matrices in Figures 11–13 assert that the model exhibited a high degree of exact classification of various drones types under different climatic conditions and altitudes, which is crucial in an anti-drone system design in guaranteeing adequate surveillance and safety of an industrial facility.

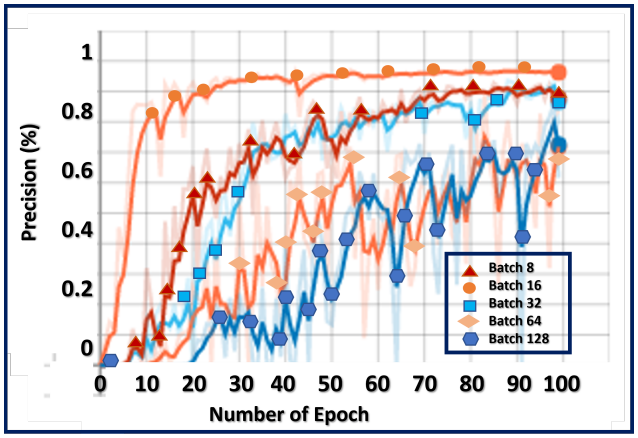

Figure 9. Precision graph of the proposed model across different batch sizes.

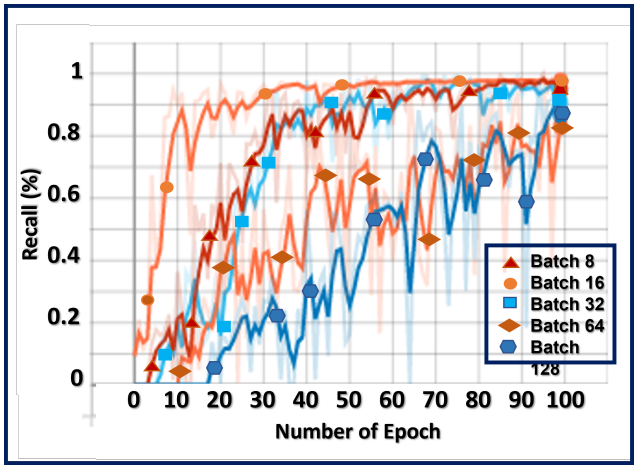

Figure 10. Recall graph of the proposed model across different batch sizes.

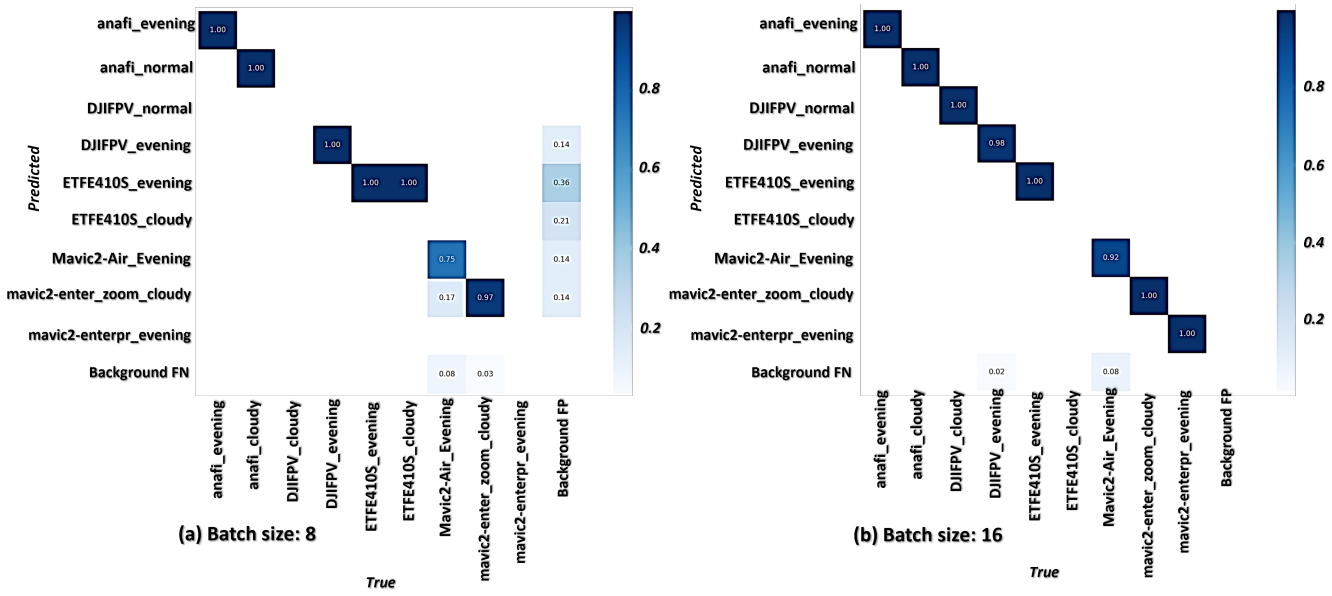

Figure 11. Confusion Matrix Graphs of showing drone classification by DRONET under different climatic conditions with different hyper-parameters; (**a**) drone classification at batch size of 8, (**b**) drone classification at a batch size of 16.

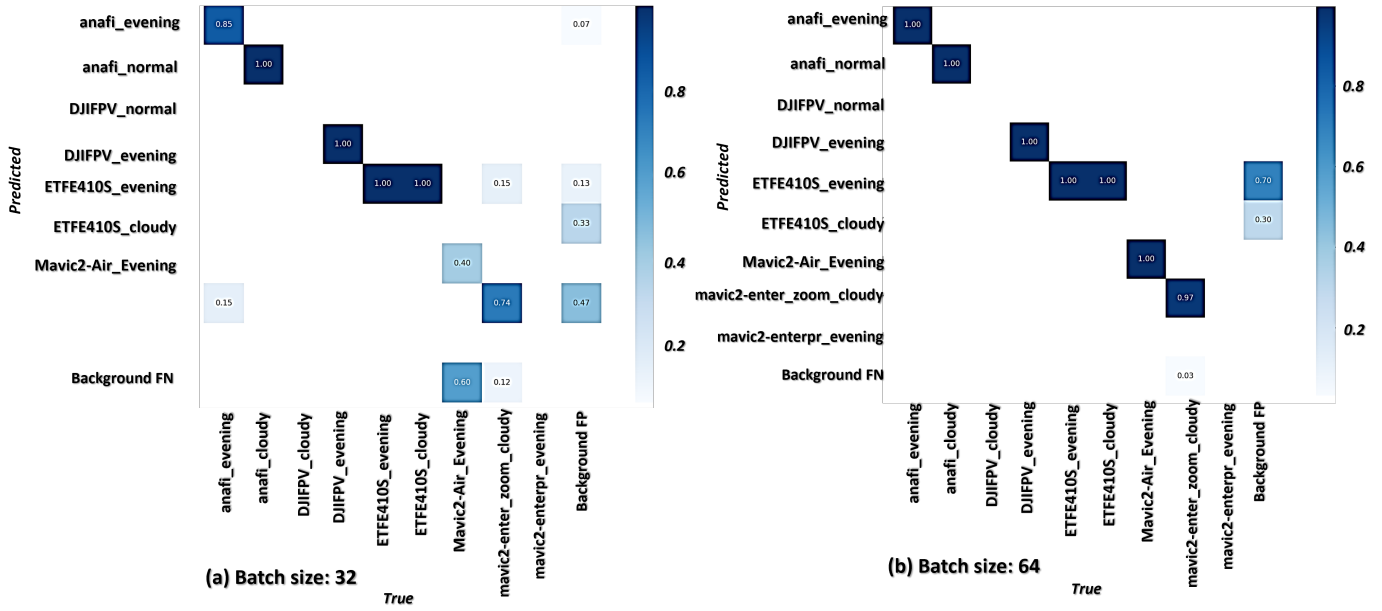

Figure 12. Confusion Matrix Graphs of showing drone classification by DRONET under different climatic conditions with different hyper-parameters; (**a**) drone classification at a batch size of 32, (**b**) drone classification at a batch size of 64.

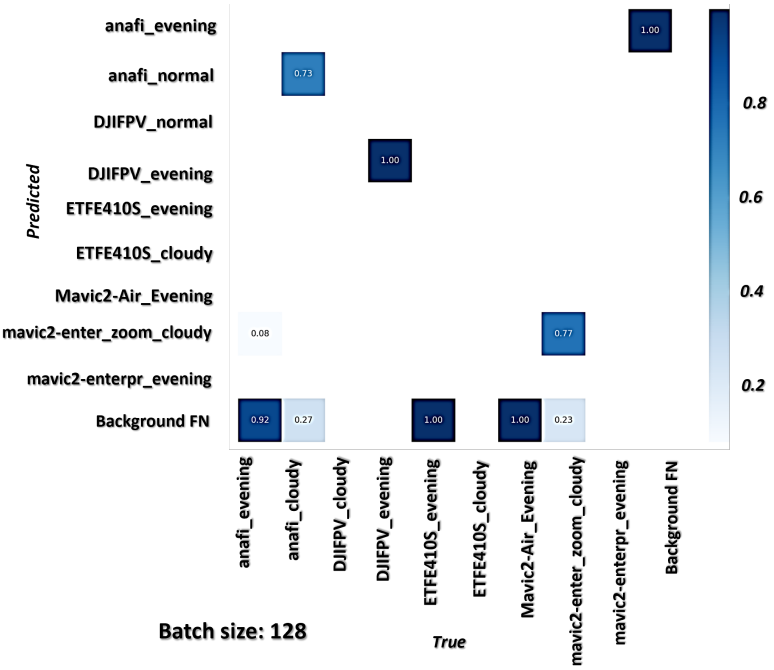

Figure 13. Confusion Matrix Graphs showing drone classification by DRONET under different climatic conditions with different hyper-parameters; drone classification at batch size of 128.

These results further confirm that there is a minimal false prediction rate by the model across the selected scenarios.

4.2. Weapons and Payload Identification

The danger potential posed by a drone in the airspace is relative to the object definition status airborne by the drone, the legalization, and the intention of its source. Since intention cannot be quantified and legalization is still a global issue, the results in Table 5 highlight

the weapons/payload identification at different distances, heights, and environments by the proposed model based on the object's physical features.

Table 5. Payload identification results of models.

Weapons/Payload Identification		
Attached Objects	**Recall (%)**	**mAP @0.5 (%)**
Bomb	50	56.5
Explosives	50	71.5
Gun	22.9	35.7
Missile	71.5	100
Secret camera	59.1	46.6
Sealed packages	52.6	52.3
UAV	75.8	76.3

From Table 5, the proposed model exhibited the capability of recognizing various sizes of objects attached to or conveyed by the drones under different scenarios with certain levels of disparity. DRONET achieved optimal precision value of 100% in identifying relatively small objects like *missile*, 75% for *explosives* and UAV, and average value for recognizing smaller objects like *secret camera*, *bomb* and *sealed packages*. Similarly, the model achieved a sensitivity value of 75.8% and 71.5%, which is necessary for proper physical threat level analysis, and effective aerial communication required in adaptive safe-channel response to counter a drone in sight. However, the low recognition values of 35.7% for tiniest distant objects like *Gun* is attributed to the insufficient number of samples for robust model training and not from the model's inadequacy. With this visual information, the proposed anti-drone model then proceeds to determine the recognized object's harmfulness or otherwise and utilize the information for appropriate neutralization response in countering an approaching drone in the airspace.

4.3. Performance Evaluation

Rationality in decision-making, timeliness in response, and seamless feedback communication are imperative in an anti-drone design as a hard real-time critical mission system. Therefore, F1-score, latency, and throughout performance metrics are used to validate the proposed model's efficiency in comparison with other state-of-the-art models with similar characteristics as shown in the result in Table 6.

Table 6. Comparing DRONET with variants of YOLO models.

DRONET vs. Variants of YOLO					
Model	**Precision (%)**	**Recall (%)**	**F1 Score (%)**	**Time (fps)**	**GFLOPS**
DRONET	**99.8**	**100**	**99.9**	**0.021 s**	**16.1**
YOLOv5s	96.4	100	98.1	0.021 s	16.4
YOLOv5fpn	97.6	100	98.7	0.022 s	16.4
YOLOv5p2	99.5	99.5	99.5	0.023 s	19.2

4.3.1. F1-Score Measurement

F1-score measures the rational behavior of a model with changes in its precision and sensitivity expressed as

$$\Rightarrow F1 - score = 2 \times \frac{(P_r * R_c)}{(P_r + R_c)}, \tag{15}$$

while detecting objects. P_r is the precision representing the positive predictive values, and R_c is the sensitivity representing the true positive detection rate by the models.

In this study, P_r is the proportion of positive predictions and recognition of drones and attached objects made by the models. This is given as expressed in Equation (16):

$$\Rightarrow P_r = \frac{(T_p)}{(T_p + F_p)}, \tag{16}$$

with T_p representing T_p true positive predictions, and F_p is the number of false positive predictions (otherwise known as type I error). As a rule in ML, the model with lesser false positive and consequently higher precision value is considered a good model. However, sensitivity (R_c) is the proportion of actual positive recognition that the models identified correctly. From Equation (17), sensitivity is given as:

$$\Rightarrow R_c = \frac{(T_p)}{(T_p + F_n)}, \tag{17}$$

where F_n is the number of false negative predictions (otherwise known as type II error) by the models. A model with lesser false negatives and consequently a higher sensitivity value is a better model than others because its predictions are more reliable.

From Table 6, DRONET had a superior F1-score value of 99.9% in comparison to other variants of YOLO models with same network configuration, showing an improved detection with a minimal false detection rate needed in an anti-drone system as captured in Figure 14.

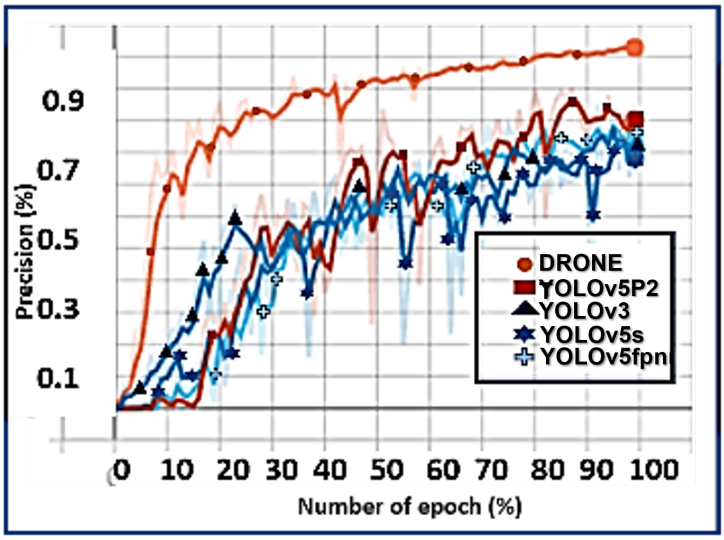

Figure 14. DRONET vs. variants of the YOLO model.

4.3.2. Throughput and Latency Determination

Throughput analyses the rate at which requests are serviced by a model. The actual throughput of a model is derived by computing the maximum number of instances/requests the model can process in a unit time frame expressed by Equation (18):

$$\Rightarrow Throughput\ (T_p) = \frac{(N_b \times b_s)}{(t_s)}, \tag{18}$$

where

N_b is number of batches;
b_s is batch size;
and t_s is total time in seconds. The results in Table 6 indicate that DRONET outperformed other YOLO variants with a throughput value of 16.1 measured in floating points operations per seconds, GFLOPS.

Finally, latency measures the timeliness (inference time) of a neural network in detecting a drone frame per second. It is a measure of the delay experienced by the model in performing inference on the server during training. Using the asynchronous execution mechanism, DRONET achieved a timeliness of 0.021 s, which is considerably better than other YOLO models based on the overall model performance rather than viewing only from the negligible difference in response time across the models.

Furthermore, the results in Table 7 provide a detailed performance comparison between DRONET and other state-of the-art object detection models for both drone detection and payload identification on a batch size of 16, 100 epochs, and learning rate of 0.0001. With a precision value of 99.60%, the proposed model displayed a superior ability to identify only the relevant data points (attached objects) on the drones and attached objects better than its counterparts. In addition, DRONET achieved a sensitivity and F1-score value of 99.80% and 99.69%, respectively, which is about 10% higher in performance than the closest model, GoogleNet, which has a value of 89.10%, 88.52%, and 88.71% for precision, sensitivity, and F1-score as seen in Figure 15. Overall, the results in Figure 15 attest to the superior performance of the proposed model in precision, sensitivity, and F1-score.

Table 7. Comparing DRONET with other object detection models.

DRONET vs. State-of-the-Art Models				
Models	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Loss**
DRONET	99.60	99.80	99.69	0.0407
VGG-16	88.52	81.50	84.86	0.0753
SqueezeNet	82.59	84.21	83.38	0.0891
GoogleNet	89.10	88.52	88.71	0.0654
MobileNet	25.56	26.35	25.94	0.0983

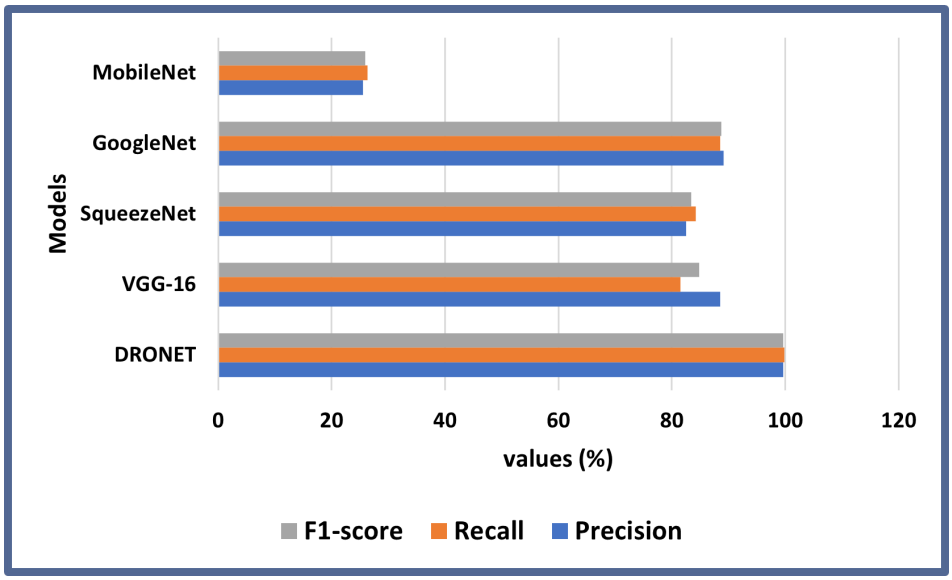

Figure 15. Performance Evaluation of DRONET vs. other models showing Precision, Recall, and F1-score values.

4.3.3. Reliability Measurement

In addition, a good model's performance is judged based on empirical risk minimization, otherwise called loss. The loss value indicates how bad or well a model behaves in making its prediction after each optimization iteration. A perfect prediction is synonymous to a zero-loss value while a higher loss is an indication of bad prediction as presented in Figure 16.

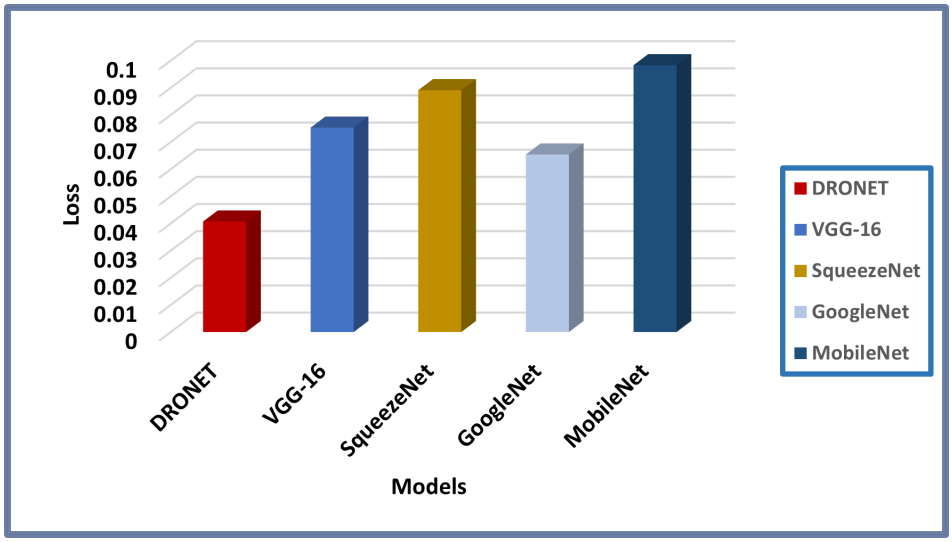

Figure 16. Loss Graph Result of DRONET vs. other models.

The results from the bar chart in Figure 16 indicate that DRONET achieved a very low loss value of 0.040, which is below the box loss threshold of 0.050 (see Table 2 in Section 3.4) when compared with other models. This further proves that the proposed model's prediction is reliable with negligible errors or false predictions. Therefore, the reliability of the proposed model is undoubtedly assertive based on the foregoing empirical analysis and available results as summarized by the reliability graph in Figure 17.

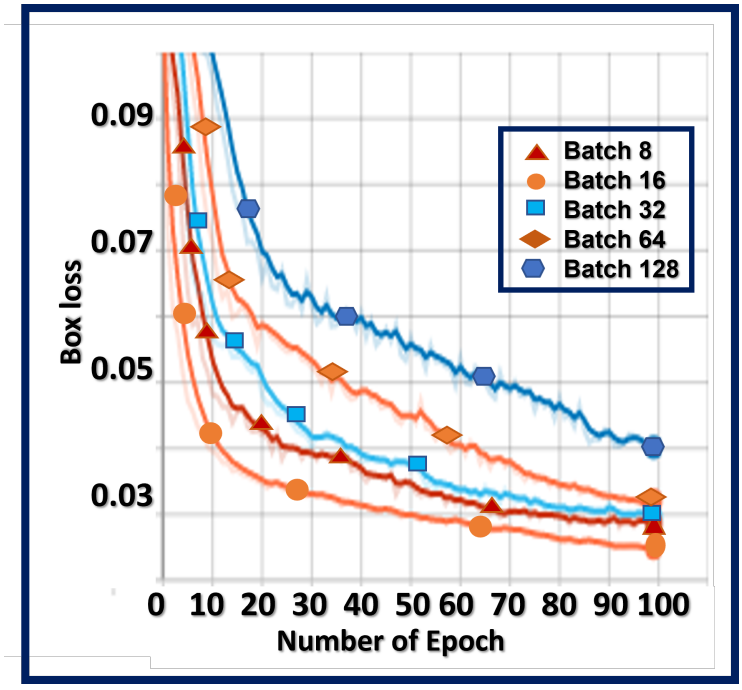

Figure 17. Reliability Graph of DRONET across different batch size showing minimal errors in prediction.

4.3.4. Efficiency Measurement

Finally, the efficiency of a neural network is determined by the intersection between its ability to identify only the relevant predictions (precision) and its ability to find all the relevant cases in the given dataset (sensitivity). Hence, the graph in Figure 18 shows the performance efficiency of DRONET in comparison with other object detection models.

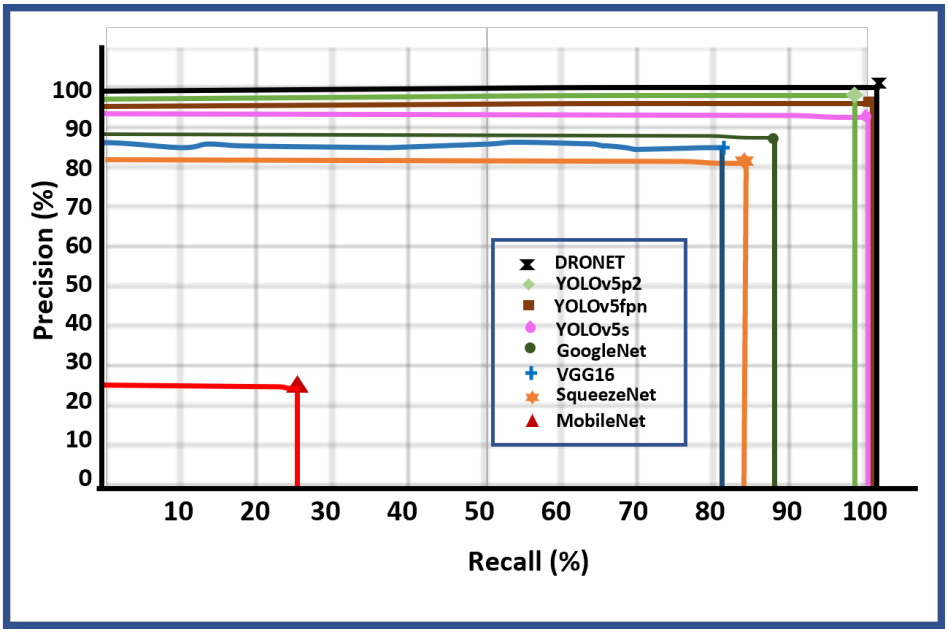

Figure 18. Efficiency Graph of DRONET and other object detection models.

These results validate the superiority of DRONET's performance as a reliable and efficient model for adaptive multi-scale anti-drone system design suitable for real time visual multi-drone detection and classification, and timely payload identification with low computational complexity and latency. With these claims, effective aerial surveillance against illegal and legal drones around an industrial facility premise is guaranteed vis-à-vis a secured network protocol for seamless communication. Hence, the predicted outputs from the CNN model becomes the input/feedback which the neutralization component of the anti-drone system utilizes to carry out the required adaptive response for safe-channel neutralization.

4.4. Safe-Channel Neutralization Result

The results from the drone detection and payload identification form the basis for the neutralization stage; that is, countering the observed drone within the field of view in the airspace. However, as discussed in Section 3.3, adaptive neutralization response requires timely perceived threat analysis and drone area mapping, which runs simultaneously with the detection process to enhance response time in split seconds.

With a 99.8% drone detection and attached object identification prediction by the underlying proposed CNN model, the anti-drone system picks the outputs from the detection process, determines the harmful status of the identified attached object, $P_{object,\,loaded}$ through perceived threat analysis (as highlighted in Algorithm 1), and finds the proximity of the drone in the airspace to the mapped area, P_t (as described in Section 3.3.2).

The outcome of P_a and P_t informs the decision to either disarm, destroy, or ignore the drone in the airspace. If the detected object in the airspace is a drone, the identified conveyed object is a bomb (which is classified as harmful), and the proximity of the drone to the mapped area is determined as far, and the anti-drone system triggers a *Stand-by* mode in readiness to engage the drone in close contact since the identified conveyed object is harmful though the drone is outside the restricted area. However, if the identified object is harmless and legality is confirmed as no intrusion, the anti-drone system simply ignores the process signaling *Harmless* mode except when there is an attempt to intrude into the mapped area.

This adaptive approach to detecting, localizing, and neutralizing illegal and harmful drones guarantees minimal resource usage on the part of the system since actions are only triggered as the need arises. In addition, it ensures that only harmful and illegal drones are detected and neutralized, thereby reducing unnecessary interference to the emerging

DTS, which allows UAVs to be used for logistics and other non-military purposes such as such as search and rescue, emergency response, aerial photography, etc. Finally, this approach guarantees automatic simultaneous detection and neutralization without the aid of a human expert since the system learns and acts based on acquired knowledge.

These claims not withstanding are constrained by the increased difficulty in identifying concealed attached objects to the drones accurately as witnessed in the payload identification simulation results. In addition, the detection of objects in a dark environment or where the drone's color blends with its surrounding environment is still an on-going research issue in object detection. Lastly, intruders leverage on loopholes in an anti-drone's system communication and network protocol to engage and defeat an anti-drone system response time of detecting and neutralizing them effectively. Therefore, a careful consideration of these issues will further improve the anti-drone system development industry.

5. Conclusions

In this study, a multi-tasking multi-drone detection, attached object identification, and safe-channel neutralization model is proposed to bridge the gap in the existing anti-drone system designs that focus only on drone detection with little effort to examine its harmful status based on the airborne object. To ensure superior detection and wide-range identification coverage, different drone models of varying sizes were captured under cloudy, sunny, and evening scenarios. In addition, harmful and harmless objects were attached to drones and flown at different altitudes and environments to capture real-life scenarios howbeit government restrictions.

The proposed deep learning model for drone detection and payload identification exhibited high detection performance of tiny drones in all scenarios in comparison with other state-of-the models. In addition, the perceived threat analysis scheme was formulated to determine the harmful status of an identified airborne object. Then, the proposed legality status model determines the rule of engagement strategy for neutralizing the drone based on the proximity of the drone in the airspace to the area of interest.

These results are invaluable milestones to the design and development of anti-drone systems as a defense mechanism to guarantee safety and sanity in the airspace in curtailing the use of drones for heinous activities without undermining the emerging DTS. However, issues such as inaccurate identification of concealed airborne objects attached to drones, recognizing drones and other aerial objects that are camouflaged in their surrounding environment, and identifying drones in dark scenarios, still need to be addressed for robust performance of vision-based anti-drone systems. Leveraging on the possibilities of a hybrid model by combining more than one detection technique as well as increasing the payload dataset to include more objects will help in addressing some, if not all, of these observed issues in this study. Future research efforts will be channeled in this direction.

Author Contributions: Conceptualization, S.O.A.; Data curation, S.O.A. and V.U.I.; Formal analysis, S.O.A.; Funding acquisition, D.-S.K. and J.M.L.; Investigation, S.O.A.; Methodology, S.O.A. and J.M.L.; Project administration, D.-S.K. and J.M.L.; Resources, D.-S.K.; Supervision, D.-S.K. and J.M.L.; Validation, S.O.A.; Visualization, S.O.A.; Writing—original draft, S.O.A.; Writing—review and editing, S.O.A. and V.U.I. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (2018R1A6A1A03024003) and by the National Research Foundation of Korea (NRF) grant funded by Korea Government (MSIT) (2019R1F1A1064055) and the Grand Information Technology Research Center support program (IITP-2022-2020-0-01612) supervised by the IITP (Institute for Information communications Technology Planning Evaluation).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work was supported by the Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (2018R1A6A1A03024003) and by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (2019R1F1A1064055) and Grand Information Technology Research Center support program (IITP-2022-2020-0-01612) supervised by the IITP (Institute for Information communications Technology Planning Evaluation).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolution Neural Network
COCO	Common Object in Context
CSP	Cross Stage Partial Network
DTS	Drone Transportation system
FLOPS	Floating Point Operation per seconds
FPN	Feature Pyramid Network
ML	Machine Learning
PANet	Path Aggregation Network
SSD	Single Shot Detectors
UAV	Unmanned Aerial Vehicles
VGG	Very Deep Convolutional Network
YOLO	You Only Look Once

References

1. Tao, J.; Han, T.; Li, R. Deep-Reinforcement-Learning-Based Intrusion Detection in Aerial Computing Networks. *IEEE Netw.* **2021**, *35*, 66–72. [CrossRef]
2. Taha, B.; Shoufan, A. Machine Learning-Based Drone Detection and Classification: State-of-the-Art in Research. *IEEE Access* **2019**, *7*, 138669–138682. [CrossRef]
3. Shi, X.; Yang, C.; Xie, W.; Liang, C.; Shi, Z.; Chen, J. Anti-Drone System with Multiple Surveillance Technologies: Architecture, Implementation, and Challenges. *IEEE Commun. Mag.* **2018**, *56*, 68–74. [CrossRef]
4. Floreano, D.; Wood, R.J. Science, Technology and The Future of Small Autonomous Drones. *Nature* **2015**, *521*, 460–466. [CrossRef] [PubMed]
5. Park, S.; Kim, H.T.; Lee, S.; Joo, H.; Kim, H. Survey on Anti-Drone Systems: Components, Designs, and Challenges. *IEEE Access* **2021**, *9*, 42635–42659. [CrossRef]
6. Haviv, H.; Elbit, E. Drone Threat In addition, CUAS Technology: White Pape. *Elbit Syst.* **2019**, *1*, 1–20.
7. Ripley, W. Drone with Radioactive Material found on Japanese Prime Minister's Roof. CNN.com. 2015. Available online: https://edition.cnn.com/2015/04/22/asia/japan-prime-minister-rooftop-drone/index.html (accessed on 31 January 2022).
8. Reuters. Greenpeace Slams Superman-Shaped Drone into Nuclear Plant. *New York Post*, 3 July 2018; Volume 1, pp. 1–2.
9. Phillips, C.; Gaffey, C. Most French Nuclear Plants 'should be shutdown' over Drone Threat. *Newsweek Magazine*, 24 February 2015; Volume 1, pp. 1–3.
10. Hubbard, B.; Karasz, P.; Reed, S. Two Major Saudi Oil Installations Hit by Drone Strike, and US Blames Iran. *The New York Times*, 14 September 2019.
11. Akter, R.; Doan, V.S.; Lee, J.M.; Kim, D.S. CNN-SSDI: Convolution Neural Network Inspired Surveillance System for UAVs Detection and Identification. *Comput. Netw.* **2021**, *201*, 108519. [CrossRef]
12. Gopal, V. Developing an Effective Anti-Drone System for India's Armed Forces. *Obs. Res. Found. Issue Brief* **2020**, *1*, 1–20.
13. Çetin, E.; Barrado, C.; Pastor, E. Counter a Drone in a Complex Neighborhood Area by Deep Reinforcement Learning. *Sensors* **2020**, *20*, 2320. [CrossRef]
14. Bhatnagar, S.; Gill, L.; Ghosh, B. Drone Image Segmentation Using Machine and Deep Learning for Mapping Raised Bog Vegetation Communities. *Remote Sens.* **2020**, *12*, 2602. [CrossRef]
15. Bemposta Rosende, S.; Sánchez-Soriano, J.; Gómez Muñoz, C.Q.; Fernández Andrés, J. Remote Management Architecture of UAV Fleets for Maintenance, Surveillance, and Security Tasks in Solar Power Plants. *Energies* **2020**, *13*, 5712. [CrossRef]
16. Sun, H.; Yang, J.; Shen, J.; Liang, D.; Ning-Zhong, L.; Zhou, H. TIB-Net: Drone Detection Network With Tiny Iterative Backbone. *IEEE Access* **2020**, *8*, 130697–130707. [CrossRef]
17. Carrio, A.; Tordesillas, J.; Vemprala, S.; Saripalli, S.; Campoy, P.; How, J.P. Onboard Detection and Localization of Drones Using Depth Maps. *IEEE Access* **2020**, *8*, 30480–30490. [CrossRef]
18. Shi, Z.; Chang, X.; Yang, C.; Wu, Z.; Wu, J. An Acoustic-Based Surveillance System for Amateur Drones Detection and Localization. *IEEE Trans. Veh. Technol.* **2020**, *69*, 2731–2739. [CrossRef]

19. Dogru, S.; Marques, L. Pursuing Drones With Drones Using Millimeter Wave Radar. *IEEE Robot. Autom. Lett.* **2020**, *5*, 4156–4163. [CrossRef]
20. Alnuaim, T.; Mubashir, A.; Aldowesh, A. Low-Cost Implementation of a Multiple-Input Multiple-Output Radar Prototype for Drone Detection. In Proceedings of the 2019 International Symposium ELMAR, Zadar, Croatia, 23–25 September 2019; pp. 183–186.
21. Guvenc, I.; Koohifar, F.; Singh, S.; Sichitiu, M.L.; Matolak, D. Detection, Tracking, and Interdiction for Amateur Drones. *IEEE Commun. Mag.* **2018**, *56*, 75–81. [CrossRef]
22. Kim, J.; Park, C.; Ahn, J.; Ko, Y.; Park, J.; Gallagher, J.C. Real-time UAV Sound Detection and Analysis System. In Proceedings of the 2017 IEEE Sensors Applications Symposium (SAS), Glassboro, NJ, USA, 13–15 March 2017; pp. 1–5.
23. Dongkyu, R.L.; Woong, G.L.; Hwangnam, K. Drone Detection and Identification System using Artificial Intelligence. In Proceedings of the 9th International Conference on ICT Convergence (ICTC), Jeju Island, Korea, 17–19 October 2018; pp. 1131–1133. ISBN 9781538650400.
24. Tunze, G.B.; Huynh-The, T.; Lee, J.M.; Kim, D.S. Sparsely Connected CNN for Efficient Automatic Modulation Recognition. *IEEE Trans. Veh. Technol.* **2020**, *69*, 15557–15568. [CrossRef]
25. Ajakwe, S.; Arkter, R.; Kim, D.; Kim, D.; Lee, J.M. Lightweight CNN Model for Detection of Unauthorized UAV in Military Reconnaissance Operations. In Proceedings of the 2021 Korean Institute of Communication and Sciences Fall Conference, Yeosu, Korea, 17–19 November 2021; pp. 113–115. Available online: https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE11022567 (accessed on 20 December 2021).
26. Anwar, M.Z.; Kaleem, Z.; Jamalipour, A. Machine Learning Inspired Sound-Based Amateur Drone Detection for Public Safety Applications. *IEEE Trans. Veh. Technol.* **2019**, *68*, 2526–2534. [CrossRef]
27. Nalamati, M.; Kapoor, A.; Saqib, M.; Sharma, N.; Blumenstein, M. Drone Detection in Long-Range Surveillance Videos. In Proceedings of the 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Taipei, Taiwan, 18–19 September 2019; pp. 1–6. [CrossRef]
28. Yong, S.P.; Chung, A.L.W.; Yeap, W.K.; Sallis, P. Motion Detection Using Drone's Vision. In Proceedings of the 2017 Asia Modelling Symposium (AMS), Kota Kinabalu, Malaysia, 4–6 December 2017; pp. 108–112. [CrossRef]
29. Zhang, Z.; Cao, Y.; Ding, M.; Zhuang, L.; Yao, W. An Intruder Detection Algorithm for Vision-Based Sense and Avoid System. In Proceedings of the International Conference on Unmanned Aircraft Systems (ICUAS), Arlington, VA, USA, 7–10 June 2016; Volume 1, pp. 550–556. [CrossRef]
30. Bay, H.; Tuytelaars, T.; Van, G.L. SURF: Speeded Up Robust Features. In *Computer Vision—ECCV 2006. ECCV 2006*; Lecture Notes in Computer Science; Leonardis, A., Bischof, H., Pinz, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; Volume 3951, pp. 879–892. [CrossRef]
31. Yong, S.P.; Yeong, Y.C. Human Object Detection in Forest with Deep Learning based on Drone's Vision. In Proceedings of the 2018 4th International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, Malaysia, 13–14 August 2018; pp. 1–5. [CrossRef]
32. Rozantsev, A.; Lepetit, V.; Fua, P. Detecting Flying Objects Using a Single Moving Camera. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 879–892. [CrossRef]
33. Thompson, A. The Cascading Haar Wavelet Algorithm for Computing the Walsh–Hadamard Transform. *IEEE Signal Process. Lett.* **2017**, *24*, 1020–1023. [CrossRef]
34. Zhang, Z.; Zou, C.; Han, P.; Lu, X. A Runway Detection Method Based on Classification Using Optimized Polarimetric Features and HOG Features for PolSAR Images. *IEEE Access* **2020**, *8*, 49160–49168. [CrossRef]
35. Awais, M.; Iqbal, M.J.; Ahmad, I.; Alassafi, M.O.; Alghamdi, R.; Basheri, M.; Waqas, M. Real-Time Surveillance through Face Recognition Using HOG and Feedforward Neural Networks. *IEEE Access* **2019**, *7*, 121236–121244. [CrossRef]
36. Sapkota, K.R.; Roelofsen, S.; Rozantsev, A.; Lepetit, V.; Gillet, D.; Fua, P.; Martinoli, A. Vision-based Unmanned Aerial Vehicle Detection and Tracking for Sense and Avoid Systems. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 1556–1561. [CrossRef]
37. Jung, J.; Yoo, S.; La, W.G.; Lee, D.R.; Bae, M.; Kim, H. AVSS: Airborne Video Surveillance System. *Sensors* **2018**, *18*, 1939. [CrossRef] [PubMed]
38. Doan, V.S.; Huynh-The, T.; Kim, D.S. Underwater Acoustic Target Classification Based on Dense Convolutional Neural Network. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1500905. [CrossRef]
39. Wang, X.; Li, W.; Guo, W.; Cao, K. SPB-YOLO: An Efficient Real-Time Detector For Unmanned Aerial Vehicle Images. In Proceedings of the 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Jeju Island, Korea, 13–16 April 2021; pp. 99–104. [CrossRef]
40. Pestana, D.; Miranda, P.R.; Lopes, J.D.; Duarte, R.P.; Véstias, M.P.; Neto, H.C.; De Sousa, J.T. A Full Featured Configurable Accelerator for Object Detection With YOLO. *IEEE Access* **2021**, *9*, 75864–75877. [CrossRef]
41. Li, S.; Li, Y.; Li, Y.; Li, M.; Xu, X. YOLO-FIRI: Improved YOLOv5 for Infrared Image Object Detection. *IEEE Access* **2021**, *9*, 141861–141875. [CrossRef]
42. Buffi, A.; Nepa, P.; Cioni, R. SARFID on Drone: Drone-based UHF-RFID tag localization. In Proceedings of the 2017 IEEE International Conference on RFID Technology Application (RFID-TA), Warsaw, Poland, 20–22 September 2017; pp. 40–44. [CrossRef]

43. Xie, W.; Wang, L.; Bai, B.; Peng, B.; Feng, Z. An Improved Algorithm Based on Particle Filter for 3D UAV Target Tracking. In Proceedings of the ICC 2019—2019 IEEE International Conference on Communications (ICC), Shanghai, China, 20–24 May 2019; pp. 1–6. [CrossRef]
44. Ajakwe, S.O.; Akter, R.; Kim, D.H.; Mohatsin, G.; Kim, D.S.; Lee, J.M. Anti-Drone Systems Design: Safeguarding Airspace through Real-Time Trustworthy AI Paradigm. In Proceedings of the 2nd Korea Artificial Intelligence Conference (KAIC 2021), Da Nang, Vietnam, 27–28 October 2021; pp. 7–9.
45. Jacob, S. YOLOv5 v6.0 is Here—New Nano Model at 1666 FPS. 2021. Available online: https://blog.roboflow.com/yolov5-v6-0-is-here/ (accessed on 31 January 2022).
46. Joseph, N.; Jacob, S. YOLOv5 is Here: State-of-the-Art Object Detection at 140 FPS. 2020. Available online: https://blog.roboflow.com/yolov5-is-here/ (accessed on 31 January 2022).
47. Versaci, M.; Zeng, Y.; Zhang, L.; Zhao, J.; Lan, J.; Li, B. JRL-YOLO: A Novel Jump-Join Repetitious Learning Structure for Real-Time Dangerous Object Detection. *Comput. Intell. Neurosci.* **2021**, *2021*, 5536152. [CrossRef]