Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 7111248, 13 pages
https://doi.org/10.1155/2022/7111248

Research Article

Generative Adversarial Networks for Unmanned Aerial Vehicle Object Detection with Fusion Technology

Nageswara Guptha M,[1] Y. K. Guruprasad,[1] Yuvaraja Teekaraman,[2] Ramya Kuppusamy,[3] and Amruth Ramesh Thelkar[4]

[1]*Department of Computer Science and Engineering, Sri Venkateshwara College of Engineering, Bangalore 562157, India*
[2]*Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield, UK*
[3]*Department of Electrical and Electronics Engineering, Sri Sairam College of Engineering, Bangalore 562106, India*
[4]*Faculty of Electrical and Computer Engineering, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia*

Correspondence should be addressed to Yuvaraja Teekaraman; y.teekaraman@sheffield.ac.uk and Amruth Ramesh Thelkar; amruth.rt@gmail.com

Received 6 January 2022; Revised 11 February 2022; Accepted 11 March 2022; Published 4 April 2022

Academic Editor: Muhammad Arif

Unmanned aerial vehicles (UAVs) also called as a drone comprises of a controller from the base station along with a communications system with the UAV. The UAV plane can be precisely controlled by a machine operator, similar to remotely directed aircraft, or with increasing grades of autonomy, as like autopilot assistance, up to completely self-directed aircraft that require no human input. Obstacle detection and avoidance is important for UAVs, particularly lightweight micro aerial vehicles, but it is a difficult problem to solve because pay load restrictions limit the number of sensors that can be mounted onto the vehicle. Lidar uses Laser for finding the distance between objects and vehicle. The speed and direction of the moving objects are detected and tracked with the help of radar. When many sensors are deployed, both thermal and electro-optro cameras have great clustering capabilities as well as accurate localization and ranging. The purpose of the proposed architecture is to create a fusion system that is cost-effective, lightweight, modular, and robust as well. Also, for tiny object detection, we recommend a novel Perceptual Generative Adversarial Network method that bridges the representation gap between small and large objects. It employs the Generative Adversarial Networks (GAN) algorithm, which iimproves object detection accuracy above benchmark models at the same time maintaining real-time efficiency in an embedded computer for UAVs. Its generator, in particular, learns to turn unsatisfactory tiny object representations into super-resolved items that are similar to large objects to deceive a rival discriminator. At the same time, its discriminator contests with the generator to classify the engendered representation, imposing a perceptual restriction on the generator: created representations of tiny objects must be helpful for detection. With three different obstacles, we were able to successfully identify and determine the magnitude of the barriers in the first trial. The accuracy of proposed models is 83.65% and recall is 81% which is higher than the existing models.

1. Introduction

Object identification, as a central task in computer vision, has come a long way, but it remains a difficult task, especially from the standpoint of an unmanned aerial vehicle (UAV), due to the small scale of the target. Due to their low resolution and chaotic depiction, detecting small objects is notoriously difficult. Small items are typically detected using existing object detection pipelines by learning symbols of all items at several scales. The recital advantage of such ad hoc structures, on the other hand, is usually limited to cover the computational cost.

These preprogrammed or remotely piloted aircraft are being developed for a range of civil applications, counting manufacturing monitoring, technical data collecting, agricultural, public safety, and pursuit and rescue. Many more uses will undoubtedly arise, some of which are now unknown [1]. The deployment of unmanned aerial systems

(UAS) naturally creates safety concerns, which has shortest in situations for the regulator and nonpayload message systems that must be utilized to function UAS. Likewise, the precision and reliability of navigation and surveillance skills must be increased. A fundamental problem in multisensory UAV requests is that the data from the many sensors are not bonded to generate an output, but rather alert signals are used individually from each system constituent to deliver numerous early notices that are subsequently validated by a manual operator.

The excessively high cost of HR imaging across large regions is another important difficulty with small-object identification. To achieve their objectives, many organizations rely on exceptionally high-resolution satellite photography. Purchasing HR photos on a regular basis for nonstop monitoring of a huge region for the purposes of instruction or traffic is costly. There are two problems associated with this. Owing to sensor noise, geometric misrepresentation and atmospheric effects, small-object detection accuracy is poorer than large-object detection, even with HR images [2]. Second, we require HR imaging, which is too costly for such a broad region that requires regular updates. As a result, we need a way to increase the precision of microscopic item recognition in LR images.

Object detection is classified into two kinds: traditional feature-based object detection and machine-learning-based object detection [3]. It focuses on the creation of target-feature extraction methods for handcrafted feature-based object identification; nevertheless, varied conditions are still difficult to satisfy; therefore, most of these methods are employed in limited environments. Deep-learning-based approaches, on the other hand, can not only improve accuracy but also achieve real-time detection with the advancement of processing hardware.

Although deep-learning-based techniques have made significant progress in object detection, miss-detection concerns still exist in UAVs [4]. The following factors are primarily to blame for these problems: (i) the network's receptive field is not robust enough to handle small objects; (ii) the training dataset is restricted to UAVs. In general, object feature representation and the accompanying training dataset are required to improve object detection performance. In addition, the accuracy versus. processing time trade-off is critical in real-world applications. Figure 1 describes how UAV planes transmits and receives sensor data from various required locations.

Single object recognition in each picture with an axis-aligned bounding box representing the object's location and size is referred to as the object detection task. The sensor selection is a very important task because the challenge is aimed at UAV applications, and there are a few points that should be highlighted. First, rather than objects from a training category, the object detection job is to locate a specific object from the training dataset [5]. Second, the object detection operation must be completed with the high throughput and precision that UAV applications need. Algorithms must be created and applied if a higher-resolution light detection and ranging (LiDAR) sensor is employed for this task.

A machine-learning algorithm, which is considered as the subset of artificial intelligence, helps to build a mathematical model based on the samples or features extracted during training phase and further help in prediction or classification of new test vector based on the learning that has happened. Generative adversarial networks use statistical parameters as the training set to produce a new set of data [6]. Though it is considered to be completely an unsupervised learning algorithm, they are useful for semi-supervised or reinforcement learning. In our proposed work contribution, the generator will make use of the features or statistics to generate results based on the camera sensor image, whereas the discriminator will try to filter the results and feedback the output to a generator again to do the process in a loop until a satisfactory condition is reached. This along with the proper choice of loss function helps to make the system more efficient and can also be seen as an alternative to relevance feedback mechanism but without the end-user intervention in the process.

This study is organized into six sections. Section 2 is related work of the research performed. Section 3 discusses about fusion technique, and Section 4 discusses about object detection using GAN. Section 5 has experimental analysis with outcome, and Section 6 has a conclusion of the work with future possibilities.

2. Related Work

The development in UAV manufacturing has resulted in an upsurge in investigate publications relevant to UAV finding and classification during the last several years. Over 100 publications have been published since 2017, compared to less than 20 in prior years.

Convolutional neural networks (CNNs) have evolved into a strong class of representations for distinguishing visual content, and they are extensively recognized as the de facto normal answer for most computer vision difficulties. Object identification with CNNs, on the other hand, is computationally intensive, necessitating high-end graphics processing units that are too powerful and heavy for a trivial and low-cost drone. Lee et al. have proposed a mechanism for offloading computation to a cloud through keeping low-level object identification and temporary navigation onboard [7]. Faster sections with CNNs, a cutting-edge method, is used to perceive hundreds of different object classes in near to real-time.

UAVs are utilized to power a range of critical computer vision applications, providing more competence and ease than traditional security cameras with secure camera angles, sizes, and perspectives. Only a few UAV datasets have been offered, and they are all constrained to a precise task, such as visual trailing or object recognition in limited environments. As a result, progressing in the related research requires the creation of an unconstrained UAV benchmark. Qi et al. developed a new UAV benchmark that focuses on complicated environments and includes new level difficulties [8]. Approximately 80,000 example settings are completely marked with bounding boxes and up to 14 characteristics for three machine vision responsibilities: object recognition,

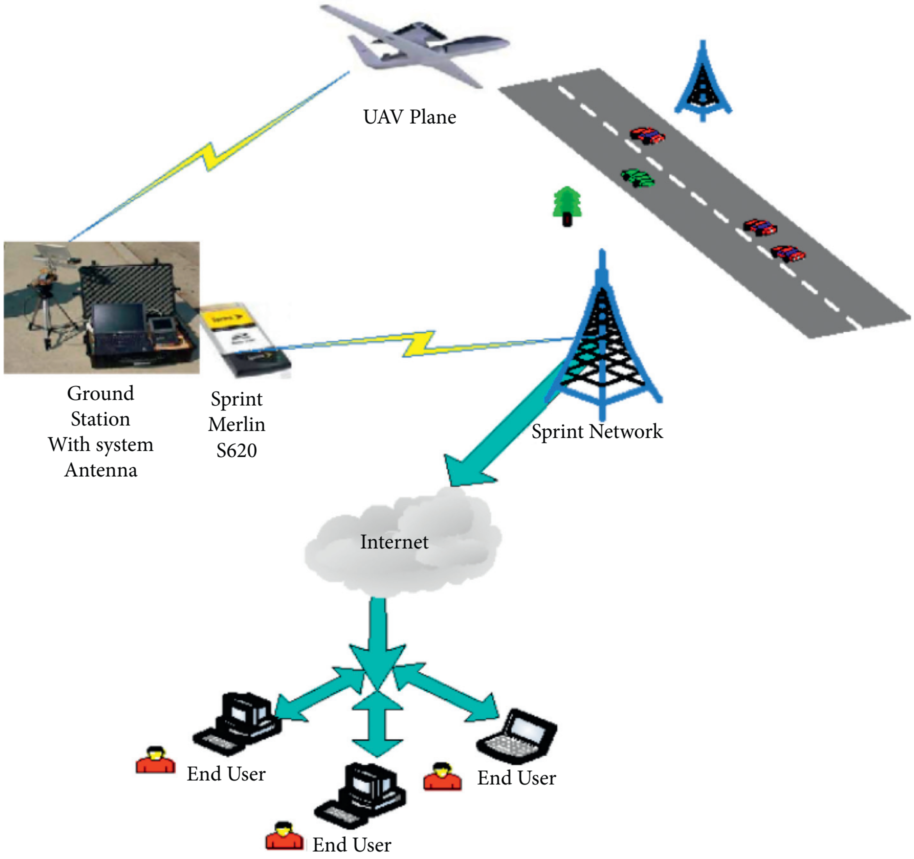


Figure 1: A typical UAV traffic surveillance system.

single object chasing, and different objects tracking (some of the examples include: flying distance, different weather conditions, the view of the camera, the category of the vehicle, and also the occlusion). Then, for each job, a comprehensive quantitative analysis is carried out using the most up-to-date advanced algorithms. Investigational results show that existing state-of-the-art algorithms perform relatively worse on the authors dataset because of extra obstacles in UAV-based actual situations, such as high density, tiny objects, and camera motion.

Reconnaissance and investigation, search-and-rescue, and substructure examination are just a few of the UAV applications that need real-time object detection. Goodfellow et al. have presented a novel approach that uses adversarial nets to estimate generative models [9]. They train two models simultaneously in this proposed system: (a) a reproductive model that seizures data distributions; and (b) a judicial model that estimates the probability of model presence from exercise data. During training or sample generation, this technique eliminates the need for any unrolled approximate inference networks or Markov chains. The experimental results provided a major breakthrough in this field and prompted the researchers to make use of GAN networks for other related problems.

Intellectual UAV video examination has grabbed the interest of a rising number of investigators, thanks to the growing use of UAV in machine vision-related requests. To facilitate study in the sector, Yu et al. offer a UAV database with 100 videos exhibiting around 2700 automobiles recorded under unimpeded settings and 840k hand-marked bounding boxes [10]. These different UAV videos were shot in challenging practical situations that provide considerable new tests for traditional object identification and tracking systems, including complex sceneries, high concentration, tiny matters, and huge camera signal.

The UAV dataset were used to establish a standard for three important machine vision tasks: object finding, tracking one single object (SOT), and tracking multiple objects. On the specified UAV dataset, their UAV benchmark, in particular, makes it easy to test and analyze advanced discovery and tracking systems. An innovative technique based on the CMSN model was provided, which can be used for SOT and MOT and analyses of new signals in UAV footage by assessing the grade of consistency between different objects and situations.

For any machine learning model to perform better, sufficient supervised information is essential. Kong et al. have discussed about using active generative adversarial networks for the purpose of image classification [11]. As labeling data is both expensive as well as hard to obtain, active learning approach is preferred as that helps to obtain annotations by selecting samples that have high probability for performance enhancement. In this study, GAN networks are used along with active learning approach for the purpose of generating good candidates. For each sample, a fresh reward is designed to quantify the uncertainty, which then drives the CGAN to produce revealing samples for a specific label.

Rezaei et al. have proposed a novel adversarial network that helps to learn multiple tasks at the same time [12]. A weighted loss method is discussed that takes care of mitigating imbalance data problem frequently encountered in the field of medical imaging. Generator and discriminator are the two components, as like any other GAN solution and in this application, the generator takes care of sequential magnetic resonance images training, whereas the discriminator helps to classify whether the result is fake or real. Two player mini-max game approach is followed for training along with the selective weighted loss algorithm. This proposal is implemented and tested on the ACDC-2017 benchmark, and the results were promising when compared to the literature methods.

The usage of Indian navigation satellite (INS) and global positioning system (GPS) to calculate location and velocity is discussed by Hartana and Sasiadek [13]. To merge the readings from the INS and GPS sensors, the Kalman filter is employed. To pick a mixture of satellites to be utilized as extent data, the Dilution of Precision (DOP) approach is used. Feedforward and feedback are two Kalman filter methods that are employed. The experiment demonstrates that the satellites chosen have an impact on the measurements. For the autonomous UAV, the technique and tests outlined in this research were created and tested.

One of the major problems during the development of UAVs has been how to enhance the accuracy, coverage, and dependability of autonomous navigation systems while keeping weight and cost in mind. In typical aerial navigation systems, the Inertial Measurement Unit (IMU) and GPS either independently or in combination are widely utilized. International interferences can cause the GPS signal in aerial vehicles to become unreliable, obstructed, or jammed. A standalone IMU, on the other hand, wanders with time and becomes unusable after a few seconds. Samadzadegan and Abdi describe a system for multisensor-based aerial vehicle navigation that determines the vehicle's accurate posture parameters in real time [14]. A Vision-Based Navigation (VBN) system transmits attitude and position data to an extended Kalman filter (EKF) algorithm, which uses an IMU motion model to precisely estimate the vehicle's pose parameters. The suggested approach has been tested in a number of different places, and the results demonstrate that it is both practicable and robust.

Remote-sensing image interpretation application requires aircraft type recognition. Machine learning methods will require huge sets of data for training along with class labels which is expensive and time-consuming. On the other side, the conventional methods also have bad generalization performance. To overcome these issues, Zuo et al. have proposed an aircraft type recognition framework that is based on the working principle of the conditional GAN [15]. The aircraft's key points are first identified, after which the aircraft's masks are produced, and the aircraft's positions are tracked. After that, a conditional GAN is trained on unlabeled aircraft images using a ROI-weighted loss function and its corresponding masks on unlabeled aircraft images. Finally, to excerpt multiscale structures in the aircraft's areas, a region of interest-based feature extraction technique is used. A SVM-based classifier is used to categorize samples based on their attributes. The ROI-weighted loss job, in combination with the GAN technique, contributes in creating features more connected to aircrafts, thereby enhancing feature quality and recognition accuracy. The data also reveal that the proposed technique outperforms existing frameworks.

3. Multisensor Fusion Approach

Object identification, classification, and multiobject sensor information tracking are all challenges for the UAVs. Recent advances in counter-UAV technology give systems a multisensory arsenal for maintaining situational awareness and defending crucial infrastructure or a large event. In these applications, many integrated sensors, generally radar or electrooptical/thermal instruments, as well as less frequently used RF and acoustic sensors, are generally utilized to identify the threats [16].

Thermal as well as electro-optro sensors both provide excellent clustering skills as well as precise localization and range when a large number of sensors are installed. Despite the fact that electro-optro sensors are less costly than thermal sensors, both are sensitive to the environment. Acoustic sensors, on the other hand, are typically immune to the elements, but their inadequate effective series makes them a less popular option. Finally, due to their precise localization and extended range, as well as outstanding categorization capabilities that work in any environment, radar sensors are the most prevalent detection approach.

A fundamental problem in multisensory UAV requests or usage is that the data from many sensors are not bonded together in order to generate a result, but rather alert data are used individually from each system constituent to deliver numerous early notices that are subsequently validated by a machine operator. Due to the desire to integrate data from multiple types of sensors for a variety of applications, data fusion techniques have received a lot of attention in recent years [17]. Data fusion's purpose is to deliver more accurate findings than single-sensor results while simultaneously adjusting for their shortcomings. Nowadays, intelligent drones are in UAV applications.

Machine-learning algorithms are used to handle a wide range of data from a number of UAV bases due to their capacity to find high-level and nonconcrete qualities that traditional feature extraction techniques cannot. Deep learning techniques in data fusion elements may be critical in addressing UAV challenge of multisensory data collection. Obstacle distance information is crucial for obstacle evasion in many requests, and it may also be used to gauge the danger of object collision.

Communication technology used in commercial flights, including the traffic warning and collision avoidance system and the autonomous reliance on surveillance transmission, must be able to interact with aircraft in the same airspace for cooperative sensors to work. In contrast to cooperative sensors, noncooperative sensors do not require the same communication equipment to share data with other aircraft in order to use the same territory. Noncooperative instruments like RADAR, light detection and ranging, and optical instruments like cameras can detect both air and ground targets.

One of the disadvantages of small-scale UAVs is their payload capability [18]. As a result, the camera becomes an excellent object and target detection sensor. The camera's light weight, low cost, and ease of use are just a few of its benefits. It is also widely used in a wide range of applications. To develop crucial technologies for mid-air collision avoidance for UAVs, this research effort creates a vision-based object recognition technique employing deep-learning-based distance estimation processing. A fixed-wing intruder may be detected, and the distance amid the invader and the owner can be calculated using the developed approach.

A monocular camera was chosen as the single sensor for identifying the target item in the airspace as the camera is an unreceptive noncooperative sensor. A multistage object discovery technique is used to estimate the distance of stirring entities on the picture plane in both short and long coverages. Based on the methodology, the contextual subtraction method is used to perceive the long-range target and the stirring item with a moving backdrop on the picture level. As the final object approaches the possessed UAV, a GAN model is accomplished to calculate the distance [19]. Then, based on the detected object's distance approximation on the image plane and its dynamic signal, a risk valuation of a mid-air impact might be undertaken to avoid a collision.

Visual sensors, such as cameras, have numerous advantages, the most notable of which is that they are lighter in weight and use less power, allowing them to be easily placed on UAVs. In a UAV, the camera can serve two tasks at the same time. It can be used for two purposes: first, for visual monitoring of the relevant geographical areas, and second, for estimating the pose of a UAV in GPS-deficient conditions [20]. A monocular camera has better scalability and accuracy than a stereo camera and requires less calculation.

Before fusing, the measurement gating module will reject any incorrect measurements from sensors that are above a predetermined threshold value as shown in Figure 2. The inputs from camera and GPS are send to position analysis. The analyzed data is send for data fusion process. Due to interference from other variables, each sensor can only extract a portion of the obstacle-avoiding navigation information, and the extracted part of the information cannot precisely reflect the accurate information of the target or can only extract a specific part of the target information. Two sets of coordinate frames (body frame and navigation frame) for coordinate transformation are defined to derive the system update equations and measurement equations for the filter.

As the GPS-based receiver is susceptible to jamming in a dynamic environment, and GPS velocity measurements are noisy due to signal strength variations, changing multipath effects, and user lock instability, it is necessary to integrate the INS into the navigation system to reap benefits over using the GPS alone. The position error between the light sensor and the GPS is the value to be assessed when GPS is active. The filter is used to combine data from GPS and light sensors based on expected patterns, and then the critical parameters of the light sensors are updated to reduce the dimension error of placement between the visible light sensors and the GPS. When GPS is unavailable, the value to be calculated is the positional error among the INS and light sensors.

The working frequency of light sensors is greater than that of GPS. In the situation, when just INS is available, the UAV's exactness and real-time presentation can be amended in this way. In principle, the simulated annealing process is a broad optimization approach with good global optimization results. The aforementioned filter is sensitive to the initial value; an incorrect initial value can result in a bigger filter error and a slower convergence rate; hence, a good initial value is required. The initial value of the mentioned filter is derived from the simulated annealing technique's optimal solution. The combination of the two techniques can reduce data acquisition error due to Gaussian noise, enhance estimate convergence rate, and improve correctness and real-time performance. The algorithm steps for data processing are shown in Figure 3: initially values are selected and preprocessed to remove noises. Variation and vectors are calculated. If the following measurement is corrected, then the image is filtered else again it goes to preprocessing stage.

The model of a drone and the cargo carriers where visualized using a camera. Position and speed are determined by combining data from radio sensors and radar. As different types of sensors have varied detection ranges, the position generated from the fusion of data from long-range sensors can be transmitted to short-range sensors, which can then lock on the approaching drone instantaneously. Measuring the same thing with a group of identical sensors has a lot of advantages. The comparison of sensor data enables for the detection of a sensor deficit. The redundant sensors will take continuous measurements without causing any visible service degradation. As a result of fusing the sensor data, the measured parameter's accuracy improves.

To accurately establish the state of the target item, multisource information fusion relies on the comprehensive processing of several types of data. We also looked into different ways for improving super-resolution network's performance in order to help with the small object detection challenge. Before turning it into a cycle model, we first integrate the sensor fusion output into the Wasserstein generative adversarial network. Then, to complete the answer, we add an auxiliary network to our architecture, which is also the study's core proposition. This method of GAN for object detection is discussed in detail in the next section.

4. GAN for Object Detection

A GAN is a machine learning paradigm that pits two neural networks against one other in a game [21]. The GAN architecture's two networks will compete to produce new data built on the training set figures, producing better outcomes than the originals. With the help of labels, GAN modeling can be improved, which can assist with the discrimination process. Conditional generative adversarial network (C-GANs) are used to accomplish this. We use C-GANs because

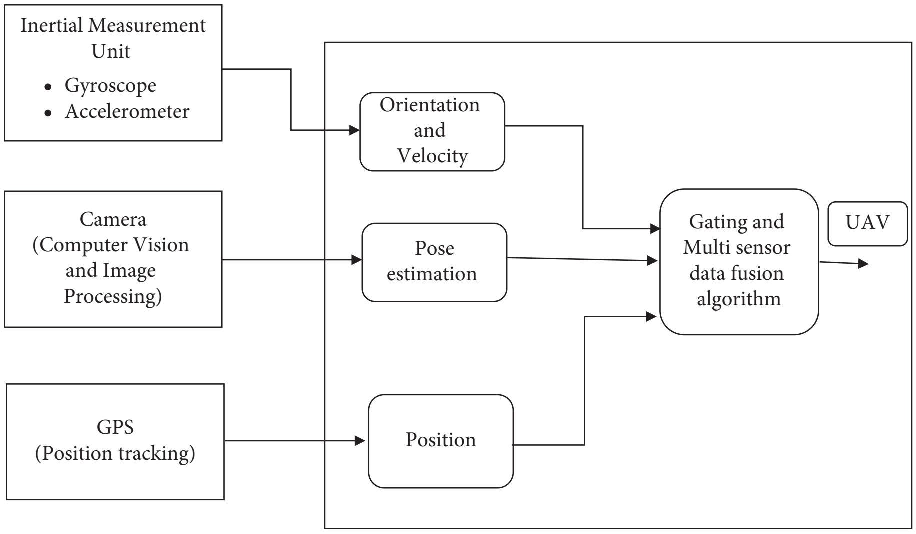

Inertial Measurement Unit
Gyroscope
Accelerometer
Camera (Computer Vision and Image Processing)
GPS (Position tracking)
Orientation and Velocity
Pose estimation
Position
Gating and Multi sensor data fusion algorithm
UAV

Figure 2: Multisensor data fusion architecture pipeline.

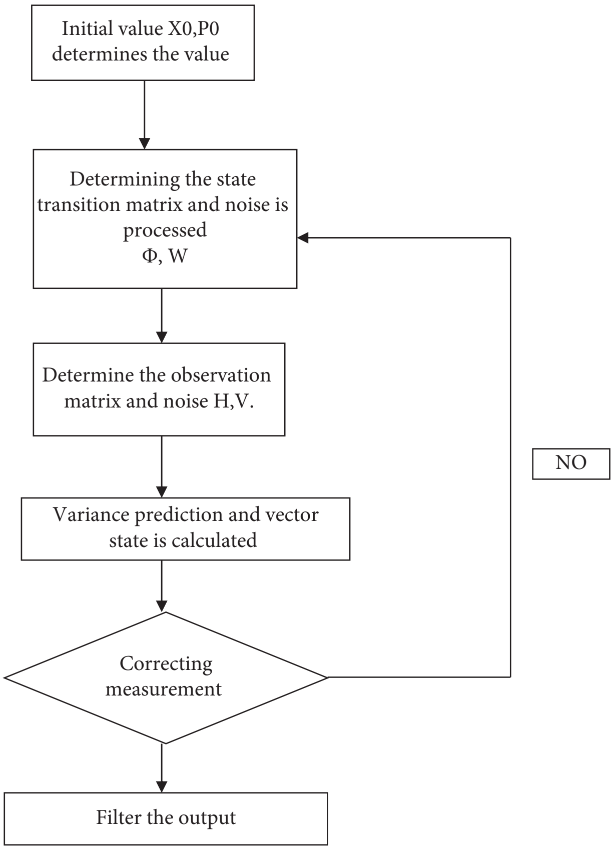

Initial value X0,P0 determines the value
Determining the state transition matrix and noise is processed Φ, W
Determine the observation matrix and noise H,V.
Variance prediction and vector state is calculated
Correcting measurement
NO
Filter the output

Figure 3: Flow chart for sensor data processing.

conventional GANs do not allow us to control the sample types that are generated. C-GANs use certain conditions to produce the output samples. Different class labels can be encoded and integrated into discriminator and generator models in a variety of ways. The discriminator is then fed the newly generated sample set to see if the yield is true or false, and the model is selected accordingly. Figure 4 depicts the proposed procedure for having multiple conditions.

When the objects appear to be identical and practical, the generator receives positive feedback. When the merged object is empty, the feedback will be negative, indicating that the objects are different. In comparison to traditional GAN

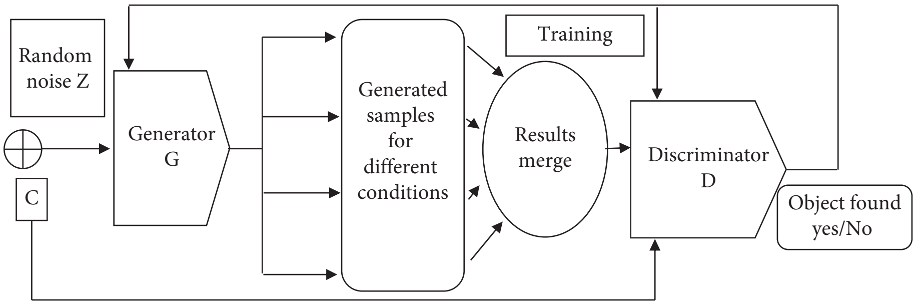


Figure 4: Conditional GAN model architecture for object detection.

models, we can generate and classify a larger number of samples this way.

4.1. Methods. Input: Given an input image (I) from the camera sensor output.

4.1.1. Process

(a) Generator will excerpt the following structures from the compressed image. The information is mined directly from the Discrete Cosine Transform (DCT) coefficients, rather than taking their inverse and decoding them completely thereby saving time and efficiency in retrieval. The feature set include the following:

   (i) From Color planes

      (1) Color histogram
      (2) Color moments
      (3) Edge histogram

   (ii) Binary texture descriptor
   (iii) Orientation information
   (iv) Energy information

By removing redundant data and keeping only the necessary ones, the Discrete Cosine Transformation provides a large number of object detection features. This approach seeks to extract the most useful data that help to construct our object detection system.

(1) From Color Planes. In the object detection method, color is a primitive function that cannot be avoided, and this work mainly talks in one form or another about image match. In our work, the color coefficients are derived from the compressed image's partial decoding. Cb and Cr represent the chrominance data in the YCbCr color space. The median values of the subblocks are obtained and all of them represent the image's color information. Through this, we receive the color histogram. Although standard deviation and skewness are widely used to extract color moments, we prefer to define images using hue and saturation because they are more compatible with human perception. It is given by the following:

$$Hue = \frac{2 + (\mathrm{Cb} - \mathrm{Cr})}{(\max - \min)}. \tag{1}$$

This hue value is multiplied by 60 and then converted in to degrees on the color wheel. If the maximum and minimum values are the same, there is no saturation. Then, it is calculated as follows:

$$\text{Saturation value} = \frac{(\max\text{value} - \min\text{value})}{(\max\text{value} + \min\text{value})}. \tag{2}$$

While more characteristics may be introduced in the future, we limit color to avoid false positives when various objects have similar colors. Additionally, having fewer items allows for faster similarity matching. Vector quantization (VQ) has a lot of advantages for image coding with high compression ratios. Classified Vector Quantization (CVQ) is a VQ variant that considers the significance of the image block in the process of classification [22]. It divides the image into blocks of edges and nonedges and classifies the blocks of edges. Ultimately, the histogram of edge classes is created as the image index. In the main database, these three characteristics (color histogram, color moments, and edge histogram) are then stored. This will be further used in the method of image indexing and object detection.

(2) Binary Texture Descriptor. Due to its ability to derive well-known feature values, texture analysis has been extensively used in computer vision and pattern identification use cases. The coefficients in the DCT domain reflect the image's directionality and higher energy level. In addition, for similarity matching, there is no calculation involved in choosing the correct set of values [23]. To this, the statistical texture characteristics are applied as an additional collection for performance enhancement. This include

$$\text{Skew} = \left(\left(\frac{1}{\sigma}\right)^3\right) \sum (b - \text{mean})^3 p(b), \tag{3}$$

where b changes from 1 to L, and the probability distribution value of bin b in the Luma plane is represented by p (b).

$$p(b) = \frac{H(b)}{M}, \tag{4}$$

where the number M in the input image I denotes the number of blocks in the image. In the same way, the mean is calculated as follows:

$$\text{Mean} = \sum bp(b). \tag{5}$$

GAN

$$\sigma = \sqrt{\sum (b - \text{mean})^2 p(b)}. \tag{6}$$

The third order moment is represented by the skew, and the fourth order moment is represented by the kurtosis, which is also included in our feature set as

$$\text{Kurtosis} = \left(\frac{1}{(\sigma)^4}\right) \sum (b - \text{mean})^4 p(b). \tag{7}$$

The entropy is the last parameter, and it represents the randomness of the distribution of coefficient values over the intensity and is given by

$$\text{Entropy} = -\sum p(b) \log_2 [p(b)]. \tag{8}$$

All of these features are combined to form our texture feature set, which is then used for object detection.

(3) Orientation and Energy Information. One of the difficulties in constructing an effective object detection framework is the use of the object form. The shape of the object is essential in the database's for searching similar images. The shape definition is invariant to the object translation, object scaling, and object rotation in the object detection system. Based on the object, it is expected to be either two dimensional (2D) or three dimensional (3D). Owing to the inherent difficulty of describing the forms, descriptors where compared to the color and texture with the growth of feature shapes [24]. The image regions where occupied by an entity, and it must be found in order to clarify its form. A low-level characteristic has been used in split-and-merge or region growing process approach in a variety of well-known segmentation techniques due to the diversity of the promising projections of a 3D object into 2D shapes, the intricacy of each distinct object shape, nonuniform illumination, the existence of shadows, occlusions, changing surface reflectance, and so on. It is difficult to classify accurately an image into important areas using low-level characteristics.

Their shapes have to be defined, indexed, and compared after segmenting the image objects. However, in the object detection system, no detailed definition can completely capture all aspects of visually assumed shapes, and shape comparison is also a very difficult problem. Not only does the restricted nature of the shape inhibit the systematic study of the trade-off between the ambiguity of the shape and the definition but also its ability to define the shape of the picture as opposed to the shapes of interest. Two large groups of 2D shape descriptors, namely contour-based and area-based, are currently exploited by the object detection method, representing either an outer boundary (or contour) or an entire region.

In the sense that each one can be used as a basis to compute the other, both boundary-based and region-based definitions are perceptually meaningful and interchangeable. But in each form of description, the shape features directly available are very different, so that all boundaries and regions should be included in an ideal description in order to achieve a more efficient object detection system. For the detection of horizontal and vertical edges from DCT blocks, horizontal and vertical coefficients can be used [25]. Two edge feature sets can be calculated by these edges in an 8× 8 block:

$$\begin{aligned} \text{Horizontal features: } H &= H_i\!: i = 0, 1, 2, \ldots, 7, \\ \text{Vertical features: } V &= V_j\!: j = 0, 1, 2, \ldots, 7, \end{aligned} \tag{9}$$

w GAN re H_i and V_j correspond to the DCT coefficients $F_{u,v}$, for $u, v = 0, 1, 2, \ldots, 7$, which describes the 2D DCT.

(b) After that, the generator will attempt to calculate the (Euclidean) distance between the input image and the database images.

(c) The retrieved images are ranked accordingly and based on the distance measurement.

(d) The discriminator will distinguish between the input image and the retrieved images.

(e) The discriminator will then feedback the results to the discriminator to eliminate the false positives.

(f) The generator uses this feedback to retrieve a new set of samples from the database.

(g) The iteration stops when the discriminator could not feedback any results to the generator.

4.1.2. Output.

Set of retrieved images that matches the input image. An image distance scale contrasts the resemblance of two images in different dimensions, such as color, texture, form, and others. For example, with respect to the dimensions that were considered, a distance of 0 implies an exact match to the question. A value greater than 0 implies different degrees of similarity between the images, as one can intuitively compile.

4.2. Major Contribution.

(1) Different important features are extracted from the images in the compressed domain rather than the decoding and extracting domain

(2) Generator training

(3) Discriminator training

GANs employ a loss function that depicts the distance among the GAN-generated data points and the input data values.

In classic GAN algorithms that employ the mini-max loss function, the generator seeks to reduce the below function and the discriminator helps to maximize it as follows:

$$E_x[\log(D(x))] + E_z[\log(1 - D(G(z)))]. \tag{10}$$

In this study,

The discriminator's approximation of the probability that real data case x is real is $D\ (x)$.

E_x is the average of all practical data cases.

When given noise z, the generator's output is $G\ (z)$.

The discriminator's estimation of the likelihood that a fake occurrence is real is $D\ (G\ (z))$.

E_z is the probable value of all the generator's random inputs (in effect, the probable value of all produced fake instances $G\ (z)$). This can further be represented as follows:

$$\min_{\theta_g}\max_{\theta_d}\left[E_{x\sim P\mathrm{data}}\log D_{\theta_d}(x) + E_{Z\sim P(Z)}\log\left(1 - D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)\right]. \tag{11}$$

Instead of using the traditional mini-max loss function, we propose to use Wasserstein loss function in our work. This function helps to estimate better the data distribution pragmatic in a given training data samples than the mini-max loss function. In this loss function, Earth Mover's distance is used. This is calculated as follows:

$$W(pr, pg) = \inf\gamma \sim \prod(pr, pg)E(x, y) \sim \gamma[\|x - y\|]. \tag{12}$$

Wasserstein distance can help to deliver a meaningful and smooth illustration of the distance among two distributions even when they are located in lower dimensional manifolds without overlaps. We maximize the probability assigned to the samples by the discriminator as follows:

$$J_G = -\frac{1}{n}\sum_{i=1}^{n}\log D(G(z_i, y_i)). \tag{13}$$

An ideal training process for the proposed GAN system involves

(a) First, the generator will generate random images with simple distance measurements between the sensor image and the database images.

(b) The discriminator network will learn with the help of basic filters to distinguish between the real images and the random noise.

(c) The generator will update the variable parameters present in the system including bias, threshold, etc., to produce more images and confuses the discriminator.

(d) The discriminator now becomes more attuned to real images matching to the sensor image and other noisy images and provides the feedback.

(e) The process continues until the discriminator is maximally confused and no further feedback can be provided.

4.3. Novelty with this Proposed Work.

(1) GAN networks are not tried much for object detection systems. So this will be a new attempt to see the behavior across different datasets.

(2) Most GAN networks use only mini-max loss function to replicate a probability distribution. So the idea of using Wasserstein loss function in GAN network for an object detection system can bring in better results than the existing systems.

(3) There will be two loss functions in a GAN: one for generator and the other for discriminator exercise. So we can try different combinations of loss functions (mini-max vs. Wasserstein) to see which one behaves better for a given dataset.

4.4. Loss Functions. The three objective function responsible for the GAN learning process include the reconstruction loss, GAN loss and the Metric loss. The root mean square error between the input and output functions refers to the reconstruction loss which penalizes the generator network for the differences introduced in it. It is formulated as follows:

$$L_{\mathrm{recon}} = \|\underline{x} - x\|^2, \tag{14}$$

where x corresponds to the original image and $\underline{x}$ refers to the generated output. The second loss is the GAN loss which is created for the better controlled training process. It is represented as follows:

$$LGAN_G = E_{\underline{x}} \sim p(\underline{x}|x)[(D(\underline{x}) - 1)2], \tag{15}$$

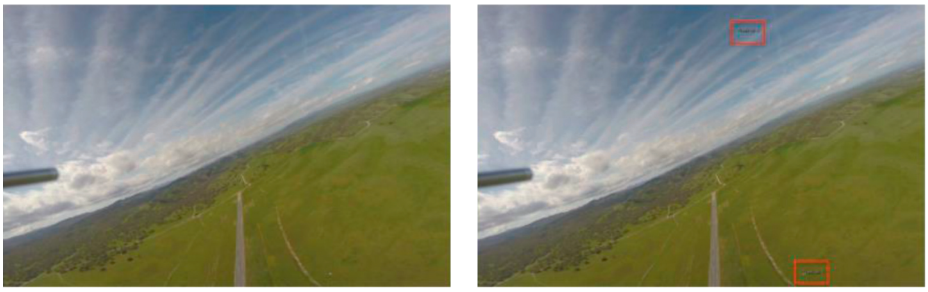
where $p\ (x|x)$ refers to the conditional distribution of the adversarial examples. The last loss function is the Metric loss which helps to push the examples away from the actual image along with its neighbors in the feature space. It is given by

$$L_{\mathrm{metric}} = \max(d(fx, fx') + m - d(fx, f \sim x), 0). \tag{16}$$

The complete loss function corresponds to the combination of three loss functions along with the proposed Wasserstein loss function.

5. Results and Discussion

The experimental findings of the suggested detection technique are described in this section. First, we shall go through the acquired UAV-viewed dataset [26] as well as the implementation specifics. After that, we present the experimental results comparisons with several training approach optimizations. The dataset collection consists of 50 video sequences totaling 70,250 frames at a frame rate of 30 fps. The resources are GoPro 3 camera sensor (that has the HD resolution size: 1920 × 080 or 1280 × 60) installed on a



(a)

(b)

Figure 5: Sample video frame (a) and object detection (b) from the experimental dataset.

Table 1: Performance assessment of proposed model with the literature models.

Model	Mean average precision	Inference time	Skew	Kurtosis
Retina net	23	8 fps	Not available	Not available
HAL model	31.8	Not available	Not available	Not available
Proposed model	30.5	12 fps	5m NS	>3

modified aircraft captures the action. There are many target UAVs (up to 8) in each movie, each with a different appearance and shape.

Our train/test data split was in the ratio of 80 : 20. The training data samples are divided in to equal portions first. To train our end-to-end system, we employed an expanded training dataset containing arbitrary horizontal flip-flops and ninety-degree revolutions. An example video frame and the item detection result are shown in Figure 5.

Bounding boxes with corresponding classes were the result of our detection. We computed intersection over union (IoU), precision, and recall to generate average precision (AP) and utilized it to assess our outcomes. True positives (TP) are a collection of successfully identified things, whereas false positives (FP) are a group of wrongly discovered objects [27] (FP). The accuracy is now calculated as the ratio of all predicted objects to the number of TPs:

$$\text{precision} = \frac{|\text{TP}|}{|\text{TP}| + |\text{FP}|}. \tag{17}$$

The set of items that the detector does not detect is referred to as false negatives (FN). The value of recall is then calculated by dividing the number of detected items (TP) by the total number of dataset objects as follows:

$$\text{Recall} = \frac{|\text{TP}|}{|\text{TP}| + |\text{FN}|}. \tag{18}$$

To quantify the localization fault of foretold bounding boxes, IoU first measures the overlap among two bounding boxes: the ground truth box and the detected box. We picked AP as our assessment metric because both of our datasets in this study only contained one class. By considering all boxes with an IoU as TP and the all other discoveries as FP, we may achieve accuracy at IoU.

Table 1 compares the performance of two existing literature models to our suggested system. The outcomes are assessed based on the mean Average Precision on corroboration and test data airborne pictures. Our system is suited for real-time object identification since it operates at a frame rate of 12 frames per second.

UAVs gathered primary node position and velocity, as well as high-precision sensor data, prior to field testing, and the dataset utilized in this work came from sensors mounted to a vehicle to get simulation information via a real flight. The mixture process was done in MATLAB, and the offline synthesis state data were likened with data from high-precision instruments to test the efficiency of the proposed approach. Table 2 captures the performance characteristics of the suggested object detection model as well as sensor accuracy.

Experiments on GPS and IMU sensors were conducted to test the performance of the filtering presented in this research. Based on the timestamp category to inform filter, this work offers a multisliding space organization adaptive unscented filter. The sensor error setting value is shown in Table 3.

For the purpose of assessing the system model, an adaptive filter built on multisensor synthesis is used on the UAV aligning system. Multisensor synthesis UAV aligning solutions require embedded environments. In the context of hardware, the flight stage is built on a quadrotor drone [28]. Noise estimation for the accelerating faults 0.30 to 0.34 m/sec, which is considered as minimum potential. Fault in measuring positions are 0.01–0.20 m, which is minimum. Range of fault is measured to be 0.1%. this makes our proposed data fusion approach very efficient.

Table 4 compares our method to various advanced algorithms for the purpose of object detection in terms of accuracy and recall. Our technique outperforms the Faster R-CNN in relation to detection performance, as evidenced by the data. Furthermore, the Fast R-CNN approach, which is better at recognizing small objects, offers no discernible advantage over the detection result. This is evident from Figure 6 as well. The accuracy of the FastR-CNN is 60.9%,

Table 2: Different sensor accuracy performance comparison.

Sensor	State	Accuracy
Inertial measurement Unit	Roll	0.1°
GPS position	Horizontal position	1 cm
GPS position	Vertical position	2 cm
Camera	Pose estimation	<0.5 mm

Table 3: Different sensor accuracy performance comparison.

Type of error	Drift	Noise estimation
Acceleration faults	X-axis value: 0.10 m/sec Y-axis value: 0.10 m/sec Z-axis value: 0.11 m/sec	X-axis value: 0.33 m/sec Y-axis value: 0.34 m/sec Z-axis value: 0.30 m/sec
Positional faults	X-axis value: 0.10 m Y-axis value: 0.11 m Z-axis value: 0.10 m	X-axis value: 0.20 m Y-axis value: 0.20 m Z-axis value: 0.01 m
Ranging faults	0	0.1% of the range

Table 4: Comparison of test results of different methods.

Methods	Accuracy	Recall
Fast R-CNN	60.91	78.53
Faster R-CNN	70.18	49.39
Proposed GAN method	83.65	81.27

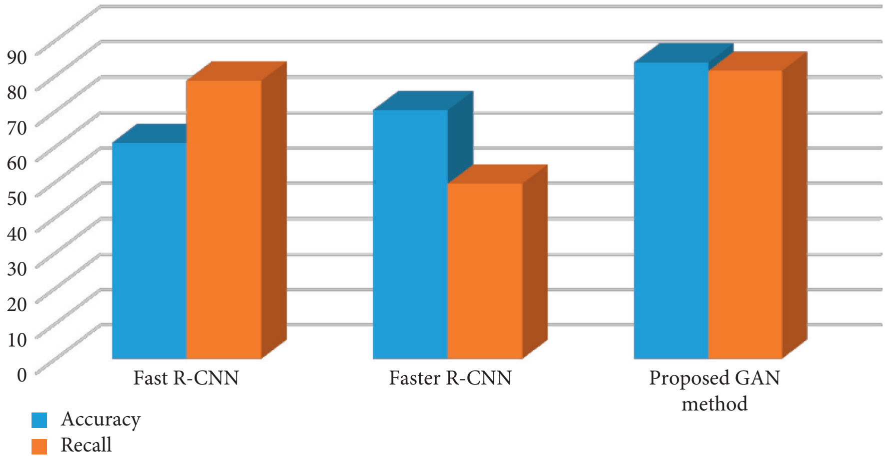


Figure 6: Accuracy and recall comparison across methods.

whereas our proposed technique achieves 83.65% which is higher than traditional techniques.

6. Conclusion

The goal of multisensor data fusion is to combine remarks from a variety of sensors to deliver a thorough and complete explanation of a setting or a region of interest. Object recognition, environment mapping, and localization are just a few of the robotics applications that employ data fusion. Only drone detection systems using many types of sensor technology are capable of detecting, tracking, and locating all types of drones. However, without intelligent sensor data fusion, all types of sensor-mix are toothless. Multiple low-quality sensors are generally less expensive than a single high-end sensor but can produce identical findings with the help of proposed data fusion method. Along with multisensor fusion, we have also proposed to use GAN-based object identification framework which is not limited to any specific architecture and may be used to a variety of DNN-based architectures. When compared to existing techniques, the performance results obtained utilizing GAN-object detection on various datasets show improved robustness to fluctuating image quality and a higher mean average precision. We investigated the object recognition objective in the small data area with reproductive modeling, learning to produce new pictures with bounding boxes. We showed that just training a prevailing generative model does not yield enough results because it prioritizes visual realism above object detection accurateness. To do this, we established a new model based on a revolutionary opening method that concurrently augments a generative system and a detector, resulting in produced pictures that improve the detector's performance. Our result proves that this method outdoes the current advanced methods on a variety of datasets. As conventional UAVs have resource constraints, future directions include optimizing memory footprint and compute power, analyzing camera input data with low expectancy and performing quicker to execute grave functions such as object detection, avoidance, and path navigation in real time. The limitation is overfitting issues in training models. Future AI techniques can be used to improve the accuracy of the system.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] D. W. Matolak, "Unmanned aerial vehicles: communications challenges and future aerial networking," in *Proceedings of the International Conference on Computing, Networking and Communications (ICNC)*, pp. 567–572, Garden Grove, CA, USA, February 2015.

[2] C. Dumitrescu, M. Minea, I. M. Costea, I. Cosmin Chiva, and A. Semenescu, "Development of an acoustic system for UAV detection," *Sensors*, vol. 20, no. 17, p. 4870, 2020.

[3] P. Mittal, R. Singh, and A. Sharma, "Deep learning-based object detection in low-altitude UAV datasets: a survey," *Journal of Image and Vision Computing*, vol. 104, 2020.

[4] S. Luo, Y. Xiao, R. Lin et al., "Opportunistic spectrum access for UAV communications towards ultra dense networks," *IEEE Access*, vol. 7, pp. 175021–175032, 2019.

[5] C. Mitash, K. E. Bekris, and A. Boularias, "A self-supervised learning system for object detection using physics simulation and multi-view pose estimation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 545–551, Vancouver, BC, Canada, September 2017.

[6] A. Domínguez-Rodrigo, A. Fernández-Jaúregui, G. Cifuentes-Alcobendas, and E. Baquedano, "Use of generative adversarial networks (GAN) for taphonomic image augmentation and model protocol for the deep learning analysis of bone surface modifications," *Applied Science journal*, vol. 11, no. 11, p. 5237, 2021.

[7] J. Lee, J. Wang, D. Crandall, Š. Selma, and G. Fox, "Real-time, cloud-based object detection for unmanned aerial vehicles," in *Proceedings of the First IEEE International Conference on Robotic Computing (IRC)*, pp. 36–43, Taichung, Taiwan, April 2017.

[8] D. Du, Y. Qi, H. Yu et al., "The unmanned aerial vehicle benchmark: object detection and tracking," *Journal of Computer Vision and Pattern Recognition*, vol. 23, pp. 370–386, 2018.

[9] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems (ICNIPS)*, Montreal, Canada, December 2014.

[10] H. Yu, G. Li, W. Zhang, D. Du, Q. Tian, and S. Nicu, "The unmanned aerial vehicle benchmark: object detection, tracking and baseline," *International Journal of Computer Vision*, vol. 14, 2020.

[11] Q. Kong, B. Tong, M. Klinkigt, Y. Watanabe, N. Akira, and T. Murakami, "Active generative adversarial network for image classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 4090–4097, Honolulu, HI, USA, 2019.

[12] M. Rezaei, H. Yang, and C. Meinel, "Generative adversarial framework for learning multiple clinical tasks," in *Proceedings of the Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–8, Canberra, ACT, Australia, December 2018.

[13] J. Z. Sasiadek and P. Hartana, "GPS/INS sensor fusion for accurate positioning and navigation based on Kalman filtering," *IFAC Proceedings Volumes*, vol. 37, no. 5, pp. 115–120, 2004.

[14] F. Samadzadegan and G. Abdi, "A decision-based multi-sensor classification system using thermal hyperspectral and visible data in urban area," *European Journal of Remote Sensing*, vol. 50, no. 1, pp. 414–427, 2017.

[15] J. Zuo, Y. Zhang, G. Xu, H. Sun, and X. S. H. Wang, "Aircraft type recognition in remote sensing images based on feature learning with conditional generative adversarial networks," *Journal of Remote sensing*, vol. 10, no. 7, p. 1123, 2018.

[16] F. Costa, S. Genovesi, M. Borgese, A. Michel, F. A. Dicandia, and G. Manara, "A review of RFID sensors, the new frontier of internet of things," *Sensors*, vol. 21, no. 9, p. 3138, 2021.

[17] F. Castanedo, "A review of data fusion techniques," *Scientific World Journal*, vol. 2013, 2013.

[18] A. Mohammadi, Y. Feng, C. Zhang, S. Rawashdeh, and S. Baek, "Vision-based autonomous landing using an MPC-controlled micro UAV on a moving platform," in *Proceedings of the International Conference on Unmanned Aircraft Systems (ICUAS)*, Athens, Greece, September 2020.

[19] S.-W. Park, J.-S. Ko, J.-H. Huh, and J.-C. Kim, "Review on generative adversarial networks: focusing on computer vision and its applications," *Electronics*, vol. 10, no. 10, p. 1216, 2021.

[20] G. Balamurugan, V. Jayaraman, and V. P. S. Naidu, "Survey on UAV navigation in GPS denied environments," in *Proceedings of the International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, pp. 198–204, Paralakhemundi, India, October 2016.

[21] C. Zhang, C. Xiong, and L. Wang, "A research on generative adversarial networks applied to text generation," in *Proceedings of the 14th International Conference on Computer Science & Education (ICCSE)*, pp. 913–917, Toronto, ON, Canada, August 2019.

[22] K. N. Ngan and H. C. Koh, "Predictive classified vector quantization," *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp. 269–280, 1992.

[23] A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, "Cosine similarity to determine similarity measure: study case in online essay assessment," in *Proceedings of the 4th International Conference on Cyber and IT Service Management*, pp. 1–6, Bandung, Indonesia, April 2016.

[24] E. Unlu, E. Zenou, and N. Riviere, "Using shape descriptors for UAV detection," *Electronic Imaging*, vol. 30, no. 9, pp. 128–1, 2018.

[25] W. Wan, J. Wu, X. Xie, and G. Shi, "A novel just noticeable difference model via orientation regularity in DCT domain," *IEEE Access*, vol. 5, pp. 22953–22964, 2017.

[26] J. Li, D. H. Ye, T. Chung, M. Kolsch, J. Wachs, and C. Bouman, "Multi-target detection and tracking from a single

camera in Unmanned Aerial Vehicles (UAVs)," in *Proceedings of the Intelligent Robots and Systems (IROS)*, pp. 4992–4997, Daejeon, South Korea, October 2016.

[27] R. Rădescu and M. Dragu, "Automatic analysis of potential hazard events using unmanned aerial vehicles," in *Proceedings of the 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pp. 1–6, Pitesti, Romania, June 2019.

[28] T. C. Mallick, M. A. I. Bhuyan, and M. S. Munna, "Design & implementation of an UAV (Drone) with flight data record," in *Proceedings of the International Conference on Innovations in Science, Engineering and Technology (ICISET)*, pp. 1–6, Dhaka, Bangladesh, October 2016.