Engineering and Technology Journal e-ISSN: 2456-3358
Volume 09 Issue 04 April-2024, Page No.- 3829-3837
DOI: 10.47191/etj/v9i04.23, I.F. – 8.227

Optimizing Affordable Drone Surveillance with Advanced Image Processing Techniques

Abhinav Sivakumar[1], Dr. Sridhar Ranganathan[2]
[1] Student, SCOPE Vellore Institute of Technology, Chennai, Tamil Nadu
[2]Associate Professor, SCOPE, Vellore Institute of Technology, Chennai Tamil Nadu

ABSTRACT: The widespread adoption of drones can be attributed to their low cost and convenience which led to a growth in their use for surveillance reasons leading to their extensive use in other areas too. In spite of this, the problem of maximizing their production while simultaneously minimizing their expenses is still one that they face. This paper provides a comprehensive methodology that can improve the efficiency of drone surveillance at a cheaper cost. This methodology is accomplished through the application of contemporary image processing technology. The employment of RRDB ESRGAN for the purpose of image enhancement, the utilization of face recognition for the purpose of authentication, the utilization of YOLO for the purpose of object detection, and the streamlining of data collecting and processing are all components of our plan. By enhancing image quality, implementing secure access through facial recognition, and facilitating real-time object detection, our system seeks to maximize drone surveillance, thereby improving both efficiency and accuracy. It has been demonstrated by the findings of this study that drone systems that are less expensive and have been improved by more advanced image processing algorithms have the potential to improve security and surveillance capabilities in a variety of different fields.

KEYWORDS: Drone Surveillance, Image Enhancement, object detection, Facial recognition, ESRGAN, RRDB, YOLOv8

I. INTRODUCTION

The proliferation of drone technology has led to a radicalization of surveillance activities in a wide range of businesses, including agriculture, law enforcement, security, and infrastructure monitoring, to name just a few of these areas. Aerial reconnaissance can be achieved at a lesser cost by using drones since they offer various perspective points and the potential to record data in real time. This allows drones to be used for aerial reconnaissance. However, the effectiveness of drone surveillance systems is dependent on the use of advanced image processing techniques. These approaches are designed to improve the quality of the data that is acquired as well as the interpretation of that data. The fundamental purpose of this research is to use modern image processing techniques in order to increase the cost-effectiveness of drone surveillance. This research is currently being carried out. As a result of the increasing prevalence of unmanned aerial vehicles (UAVs) across a wide range of industries, there is an immediate requirement to work toward the optimization of the performance of unmanned aerial vehicles (UAVs) while concurrently lowering the costs of their operations. Our goal is to overcome the inherent drawbacks of low-cost drone systems and increase their accuracy in surveillance operations by applying cutting-edge algorithms for object detection, face improvement, and picture enhancement. This will allow us to eliminate the inherent limitations of these systems. Utilizing these methods will allow for the successful completion of this task. There are four primary components that make up the approach that is outlined in the paper. These components are facial recognition for authorization, object detection via YOLO, image enhancement through the deployment of RRDB ESRGAN, and efficient data collection and processing. Increasing the capability of drone surveillance systems that are both cost-effective and efficient is our goal, and we plan to do this by combining these technologies in a way that is complementary to one another. Because of this, these systems will be able to provide insights that can be put into action and will be able to increase situational awareness in a variety of different scenarios. In the following parts, which provide a comprehensive study of each component, detailed explanations of the fundamental principles, approaches, and implementation strategies that underpin each component are provided. These sections include a comprehensive examination of each component. Establishing a complete framework that may be utilized to enhance drone surveillance operations is the objective that we have set for ourselves. offering empirical proof and offering a clear explanation of the thought that drives our methodology will be the means by which this particular objective will be attained. We are able to establish the practicability and efficacy of our suggested method in boosting the functionalities of drone surveillance systems that are both cost-effective and efficient by making use of empirical evaluations and case studies. This allows us to demonstrate that our method in question is both practical and effective. To facilitate the development of robust yet economical solutions

that are capable of effectively addressing the ever-changing security and monitoring demands of contemporary society, the purpose of this study is to establish a connection between the affordability of drone surveillance equipment and its functionality. This will be accomplished by establishing a connection between the functionalities of the equipment and its affordability. The purpose of the study is to determine whether or not there is a connection between the two. In an environment that is getting more dynamic and complicated, our goal is to make sure that institutions and organizations have access to the resources they need to safeguard their assets, lessen their vulnerabilities, and enhance their situational awareness. The utilization of a multidisciplinary approach that integrates cutting-edge image processing techniques with technology that is founded on unmanned aerial vehicles will be the means by which this objective will be realized.

II. RELATED WORKS

In their paper, *Guo, G., Wang, H., Yan, Y., Zheng, J., & Li, B et al* [1] presented a quick face detection technique based on Discriminative Correlation Filters (DCFs) that are retrieved by a Convolutional Neural Network (CNN) that has been specially constructed. This method directly classifies DCFs, which leads to significant gains in detection efficiency over existing CNN-based face identification approaches. By using a sliding window strategy on DCFs, it successfully recognizes small faces, surpassing the shortcomings of existing CNN-based techniques that have trouble with small-sized face identification because they depend too much on object proposal generating techniques. The experimental results show that the suggested DCFs-based face detection method performs well on a variety of face detection datasets.

In their paper, *Kong, T., Sun, F et al* [2] presented a simple yet effective framework for generic object identification called FoveaBox, which may be used without the use of anchor boxes that have already been set. FoveaBox provides a simplified approach to object detection by anticipating the object's position and boundary at the same time, doing away with the need for candidate boxes. Experiments using industry standard benchmarks verify the efficacy of FoveaBox, which is further corroborated by extensive analytical outcomes. The scientists hope that by using this straightforward and efficient method, object detection research will have a solid foundation, allowing for future developments in the area.

In their paper, *Zhu, Y., Cai, H et al* [3] suggested that techniques developed for generic object identification can be used for face detection applications, highlighting face detection as a one-class general object detection problem. The authors present TinaFace, a baseline technique based on generic object detection concepts that is straightforward but dependable. TinaFace is made with simplicity in mind, utilizing new developments and methods that are simple to use. TinaFace outperforms many contemporary face detection techniques, like ASFD-D6, even without Test-Time Augmentation (TTA), particularly on difficult subsets of datasets like WIDER FACE. Furthermore, the finished TinaFace model performs in the forefront of face detection.

In their paper, *Qi, D., Tan, W et al* [4] presented a YOLO5Face, based on the YOLOv5 object detection framework. From the largest (YOLOv5l6) to the smallest (YOLOv5n), the authors employ eight models. These models perform on the Easy, Medium, and Hard subsets of the WiderFace dataset in a way that is comparable to or better than the State-of-the-Art (SOTA) findings. This illustrates how well YOLO5Face works to achieve both high performance and quick inference speed. The authors have enabled the construction of other programs and mobile apps that showcase remarkable performance by making the source publicly available.

In their paper, *Rashid, S. I., Shakibapour, E et al* [5] discussed the task of creating and verifying single MR image super-resolution with a Generative Adversarial Network (GAN). To create high-resolution (HR) MR pictures, the authors use the Real-ESRGAN model on 2D MR images from the BraTS dataset. They show that Real-ESRGAN improves resolution by a factor of 4 compared to conventional interpolation techniques like bilinear and bicubic interpolation. With crisper edges and fewer artifacts, the resulting images outperform interpolated ones both qualitatively and perceptually. The findings show that interpolated images frequently include suppressed edge information including ghosting, shadowing, and blurriness around boundaries. In the future, the model will be expanded to produce super-resolution photos with arbitrary zooming factors, which should lead to even greater resolution and quality gains.

In their paper, *Dilshad, N., Hwang, J. et al* [6] emphasised on numerous applications such object detection, video summarization, persistent monitoring, traffic management, SAR operations, and disaster management, this article offers a succinct summary of the literature on drone video surveillance. Real-time object detection in the presence of shakiness and distortion presents challenges. Finding important moments in lengthy recordings is the first step in video summarizing. When object detection is dependent only on backdrop subtraction, persistent tracking encounters problems. SAR activities bring to light the difficulty of striking a balance between UAV capabilities and human labor in terms of time and coverage area. The challenge of vehicle counting at crossings without occlusion is a topic of discussion in traffic management. There is an outline of future research directions that include resource-friendly modules for permanent tracking, CNN models for multi-angle video summarization, embedded pre-processing techniques for improved object detection, and higher resolution cameras with effective algorithms for SAR missions. The objective is to provide comprehensive unmanned aerial vehicle (UAV) detection systems for prompt alarms in traffic and catastrophe management scenarios. The paper proposes to use UAVs for smart city monitoring by utilizing forthcoming technologies such as 5G and IoT. It highlights the potential for ongoing AI-

5G, IoT, and UAV initiatives and suggests study utilizing smart chipsets for context-aware learning and human-like scene perception in smart city environments.

In their paper, *Jiang, Y., & Li, J.* [7] presented TSRGAN, a super-resolution model built on the generative adversarial network framework. A number of improvements are incorporated into TSRGAN: Batch Normalization layers are eliminated, residual dense blocks are added, Wasserstein GAN with Gradient Penalty is used for better adversarial loss, perceptual loss is improved, and texture loss is incorporated. The outcomes of the experiment suggest that TSRGAN may produce images that are more realistic by doing better in both subjective and objective evaluation indices. Subsequent investigations will concentrate on employing super-resolution reconstruction to improve image quality in certain scenes or fields.

In their paper, *Hasan, M. K., Ahsan, M. S et al* [9] provided a thorough analysis of face detection techniques, dividing them into feature-based and image-based methods. Image-based techniques work well with grayscale images, whereas feature-based techniques are superior for real-time detection. Advances in hardware are reflected in the transition from older ASM models to more modern NN models, where SA models offer computational efficiency. These algorithms have uses in pattern recognition, EEG data analysis, and problem diagnostics in addition to face detection. Although problems like occlusion, complicated backgrounds, and lighting still exist, more recent techniques, such as neural networks, show promise in resolving these problems. False positives, however, continue to be a problem for algorithms and provide difficulties for vital applications like security and healthcare. Unlocking the full potential of face detection technology in domains like criminal identification and payment verification requires improving its accuracy.

In their paper, *Joseph, K. J., Khan, S. et al* [10] discussed the constraints of closed-set datasets and evaluation methodologies in classical object detection techniques. Open World Object Detection is introduced, in which the detector may recognize objects it has never seen before and gradually pick up new labels over time. Two notable contributions are a contrastive clustering strategy for open world learning and an energy-based classifier for unknown detection. The authors hope that by encouraging more study in this important and unrestricted area, their technique will help enhance object detection beyond the limitations of closed-set datasets.

In their paper, *Larsen, G. D., & Johnston, D. W.* [11] discussed drone integration with wildlife biology offers revolutionary possibilities for research methodology and theory, especially in the field of pinniped studies. Although many drone uses are still limited to "proofs of concept" or replacing more conventional research methods, researchers can now take advantage of the unique benefits that drones offer, like vast spatial coverage, high resolution, rapid deployment, and customization, thanks to technological improvements in the field. Drone imagery enriches data with a wealth of detail and metadata, opening up downstream prospects in photogrammetry, computer vision, and other as-yet-unimagined applications. These developments make it possible to combine previously separate data streams, which encourages the creation of new syntheses and shifts the focus of research toward integrated, multiscalar issues. Similar integration has been shown in the study of cetaceans, where biologging, biomechanical models, and drone data are integrated to reveal novel ecological and evolutionary insights. In the future, researchers will need to resolve constraints, evaluate and improve drone methodologies, and expand the scope of pinniped study to meet management goals and statistical rigor standards.

In their paper, *Rakotonirina, N. C., & Rasoanaivo, A* [12] suggested ESRGAN+ and nESRGAN+, two improved versions of ESRGAN. These variations achieve higher perceptual quality than other approaches. To boost network capacity, they add a new basic block, and they include noise inputs to add stochastic variation. Images with finer details, sharper edges, and more realistic textures are produced as a result of these advancements.

In their paper, *Wang, X., Yu, K. et al* [13] discussed how in terms of perceptual quality, their Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) model continuously beats out earlier super-resolution techniques. Based on perceptual index, their solution won first place in the PIRM-SR Challenge. Several Residual in Residual Dense Blocks (RDDB) without Batch Normalization (BN) layers are included in the unique architecture, coupled with training aids such as residual scaling and reduced initialization. To assist the generator in restoring intricate textures, they make use of a relativistic GAN discriminator. They also improve perceptual loss by using features prior to activation in order to provide more robust supervision. There are several grants and financial sources supporting the effort.

In their paper, *Wang, P., Bayram, B., & Sertel, E.* [14] presented a new Densely Residual Channel Attention-based framework for Remote Sensing Image Super-resolution. For deep feature extraction, the system uses residual in residual blocks with dual skip connections. Furthermore, for additional information extraction, novel techniques for densely residual spatial attention and densely connected channel attention blocks are devised. The objective of a deep learning architecture is to discover intricate relationships between channels in a highly packed network by including attention techniques. In order to enhance performance, transfer learning algorithms are presented, especially for acquiring finer edge data like roads and bridges. Based on experimental results, the suggested method surpasses other deep learning techniques both statistically and qualitatively, reconstructing finer texture details. Future research will examine large-scale super-resolution methods and investigate spectral and spatial attention mechanisms to improve the quality of satellite pictures. Funding and assistance sources are acknowledged.

In their paper, *Wu, Y., Chen, Y. et al* [15] do a thorough analysis that shows that the convolution head and fully connected head, two popular head architectures in object identification, have different preferences for tasks related to classification and localization. It turns out that the convolution head (conv-head) performs better at localization, while the fully connected head (fc-head) is better at classification. Furthermore, by looking at the output feature maps, they find that the fc-head has more spatial sensitivity than the conv-head, which allows it to discriminate between whole and partial objects but is not as robust in whole object regression. Expanding upon these realizations, the authors suggest a Double-Head approach that combines a convolution head for bounding box regression with a fully connected head for classification. Their approach outperforms FPN baselines with ResNet-50 and ResNet-101 backbones by +3.5 and +2.8 AP, respectively, on the MS COCO dataset without the need for further complexity.

III. METHODOLOGY

MODULES

•Image enhancement

•Object detection

•Face recognition

A. Image Enhancement

Within the surveillance system, the Image Enhancement Module serves as the primary level of processing that is carried out. In order to enhance the quality of aerial imagery received from drones, the system makes use of sophisticated algorithms such as ESRGAN and RRDB. Through the utilization of ESRGAN and RRDB algorithms, the module is able to process raw picture data, which ultimately results in enhanced photos that display a higher level of visual fidelity. The objective of photo enhancement is to enhance the clarity of features, sharpen the edges, and reduce noise, which ultimately results in improved object detection and facial recognition in future processes. Photographs are enhanced in order to achieve these goals. By improving image quality, the module improves the accuracy and reliability of future analysis and interpretation activities.

B. Object Detection

It is the responsibility of the Object Detection Module to recognize and follow the various items that are captured in the enhanced photographs that are taken by the drones. The YOLO algorithm, which is well-known for its great efficiency and its capacity to handle data in real-time, is utilized by the system. Real-time analysis of augmented photographs is carried out by the YOLO method in order to identify and categorize things of interest, such as autos, people, and buildings. The purpose of object detection is to enable the system to identify potential threats, keep an eye on things of interest, and provide essential information that can be used for decision-making. Advantages: The real-time capabilities of YOLO enhance the system's capacity to comprehend and react to its surroundings, making it an excellent choice for surveillance scenarios that call for monitoring that is both swift and flexible.

C. Face Recognition

The Face Recognition Module's major purpose is to reliably identify and authenticate individuals based on the high-quality photographs obtained from the drones. This is accomplished by using the modules to analyse the images. It makes use of the face recognition library in addition to advanced facial recognition techniques. The major objective of the module is to verify the identities of persons by identifying faces in the enhanced photographs and then comparing those faces to a database of known people in order to establish their identity. Face recognition technology is being implemented with the intention of enhancing the security and authentication capabilities of the surveillance system. This will be accomplished by providing the system with the ability to discern between authorized individuals and potential threats. Accurate facial recognition has a number of benefits, including the enhancement of security procedures, the simplification of worker monitoring, and the facilitation of restricted area access management.

D. Algorithms and Models

1. RRDB ESRGAN:

For the purpose of constructing a deep neural network, the generator network of RRDB ESRGAN is made up of a large number of RRDBs that are stacked in a particular configuration. The RRDBs are made up of a number of residual connections, which enables the model to gain complex and non-linear correlations between images with low resolution and those with high resolution. In addition, the generator network adds a sub-pixel convolutional layer in order to improve the resolution of the image that is produced. This image is then improved even further by another convolutional layer. It is the purpose of the discriminator network in RRDB ESRGAN to discern between high-resolution ground truth images and high-resolution images that have been created. The network is made up of a number of convolutional layers that have leaky ReLU activation functions. It undergoes training in order to optimize the adversarial loss, which is a measurement of the disparity between the images that are created and the images that are based on the ground truth. The feature extraction network is applied in order to compute the feature loss. This is accomplished by comparing the high-resolution ground truth photographs with the high-resolution images that were generated. A number of convolutional layers make up the network, which is utilized for the purpose of extracting feature maps from both the generated images and the ground truth images. As a result of the architecture of RRDB ESRGAN being carefully constructed to maximize on the advantages of both the RRDB and ESRGAN frameworks, a robust image enhancement model has been produced. The deep feature extraction capabilities of RRDB are utilized by RRDB ESRGAN, which then combines these capabilities with the adversarial training and feature loss calculation methods of ESRGAN. Through the utilization of this combination, RRDB ESRGAN is able to produce high-resolution images of extraordinary quality, which are distinguished by details that appear genuine and almost lifelike.

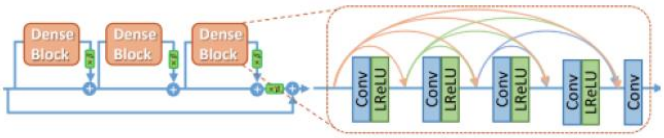

Fig 1. RRDB

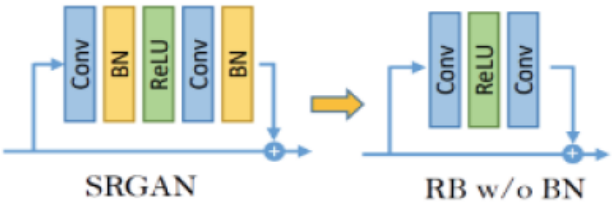

Fig 2. SRGAN to ESRGAN

2. YOLOv8:

Data Preparation:

The dataset, which consists of images of weapons and the annotations that accompany them, is prepared in a way that is compatible with YOLO. This is typically accomplished through the utilization of Roboflow. Each weapon object in the collection is annotated with its bounding box coordinates and class label, and the collection is comprised of labelled photographs representing the weapon.

Model Architecture:

YOLOv8, sometimes referred to as You Only Look Once version 8, is a sophisticated object detection algorithm that is famous for its excellent speed and precision. A sophisticated neural network architecture lies at the heart of the system, which enables it to precisely and rapidly recognize a wide variety of objects inside a picture. A backbone network, a feature pyramid network (FPN), and a detecting head are the components that make up the architecture. The backbone network, which often makes use of the Darknet architecture, is responsible for gathering features from the input image at a number of different scales. Through the incorporation of properties from a wide range of scales, the FPN significantly enhances the network's capability to recognize objects of varying quantities. Additionally, the detection head makes use of anchor boxes and convolutional layers in order to make predictions regarding bounding boxes, confidence scores, and class probabilities for each object that has been discovered. The third step in the training process involves the YOLOv8 model being trained with the dataset that is provided in order to acquire the capability of recognizing guns in photographs. In the course of training, the model makes adjustments to its parameters by employing methods such as stochastic gradient descent (SGD) or Adam optimization. The goal of these adjustments is to minimize the disparity between the bounding boxes and class labels that were anticipated and those that were actually observed. It is usual practice to add components for object localization (bounding box regression), object categorization, and confidence score prediction into the loss function.

The process of prediction:

Once the YOLOv8 model has been trained, it can be used to make predictions on new, previously unseen photographs or videos in order to identify weapons. The model analyses the input data by utilizing its structure, and it generates bounding boxes, confidence scores, and class probabilities for the objects that have been discovered. Typically, these predictions are screened using a confidence threshold in order to keep only the detections that have high levels of confidence.

Visualization of the Output:

The items that have been identified, along with the enclosed regions of those items and the certainty ratings, are presented on the video frames or images that have been input. A better understanding of the performance of the model can be gained through visualization, which also helps with the analysis of the results of the detection.

Post-Processing:

In order to get rid of redundant or overlapping detections, you can utilize post-processing techniques like non-maximum suppression (NMS), which is an option. The NMS method ensures that only the detections that are at the highest level of reliability for each object are retained, hence eliminating any predictions that are redundant.

The interpretation of the results:

The YOLOv8 model generates the final result, which includes the precise positions of weapons inside the input photographs or videos, together with the confidence scores that correlate to those locations. With the use of metrics like accuracy, recall, and average precision, which are computed on the basis of a ground-truth dataset, it is possible to evaluate the performance of the model.

3. Face Recognition:

The Haar cascade classifier is utilized in order to identify faces in both still images and video streams. This feature is referred to as "face detection." The input image is retrieved and converted into grayscale while the image processing pipeline is in operation. Identifying probable faces within a grayscale image is accomplished by the Haar cascade classifier through the utilization of the detectMultiScale() function. This is accomplished by identifying regions of interest that correlate to specific facial characteristics. Using the cv2.rectangle() function, bounding boxes are drawn around the faces that have been recognized in order to provide a visual representation of them.

Frames from the incoming video are read in sequential order within the video processing pipeline. In a manner analogous to that of image processing, each frame is converted into grayscale, and then it is fed into the stage that is responsible for face detection. In order to visually indicate the faces that have been recognized, rectangular boxes have been drawn on the frames. The following presentation of these frames is accomplished by utilizing the cv2_imshow() method that is included in the Google Colab patches. Face Recognition makes use of the face_recognition library in order to carry out tasks related to face recognition . For the sake of comparison, two photographs of faces serve as samples. These photographs are referred to as

image_path_1 and image_path_2. Obtaining face encodings from each image is accomplished through the utilization of the face_recognition.face_encodings() method. Under the influence of this function, the face is transformed into a numerical representation that can be used for comparison. Through the usage of the face_recognition.compare_faces() method, it is possible to compare the face encodings of two different photos and determine whether or not they depict the same individual. Depending on the findings of the comparison, the outcome of the comparison is displayed to indicate whether the faces are the same or different from one another.

The system represents the faces that have been identified, along with the bounding boxes, in both still images and moving video frames. Users are able to better appreciate the results of the face detection and identification procedures with the assistance of this visualization, which also enables them to evaluate the correctness of the system's performance.

During the setup process, pip install commands are utilized in order to successfully install the necessary dependencies. These dependencies include opencv-python-headless and face_recognition. The Python environment is checked to ensure that the relevant libraries are present in order to carry out face detection and identification operations. Error handling technologies are deployed in order to manage exceptions in a gracious manner. Issues with the file not being located, incorrect input formats, and errors in face detection or recognition are all examples of potential problems that can arise. By ensuring the stability and resilience of the face recognition module through the implementation of proper error handling methods, it is possible to eliminate any unexpected crashes or disruptions that may occur while the module is in operation.

E. Data Collection

The collection of datasets that are both of high quality and diverse is absolutely necessary in order to achieve the objective of developing an efficient weapon identification system. The successive steps of model training and evaluation are built upon the foundation of data collection as their primary foundation. The purpose of this part is to provide an explanation of the methods that were utilized to collect data. We took advantage of the Roboflow platform in order to streamline the process of obtaining data and ensure that substantial datasets were available for the purpose of training our weapon recognition model. Roboflow is a versatile and effective program that assists with the organization and enhancement of image collections for projects using machine learning. By utilizing the integration capabilities of Roboflow, we are able to employ automation to expedite the process of data collection and ensure the availability of datasets that have been carefully picked to match our particular requirements. By integrating Roboflow's application programming interface (API) into our development environment, we are able to simply connect and interact with the extensive variety of features offered by the platform. We are able to take use of a broad variety of capabilities by integrating with Roboflow through the use of programming code. These features include the capacity to handle several versions of datasets, to apply data augmentation techniques, and to download specific datasets. Characteristics of the Project: Within the Roboflow environment, we locate the particular project that we are interested in. We are able to concentrate our efforts on gathering data on a specific topic thanks to this thorough project-level definition, which ensures that the composition of the dataset is relevant and consistent. Following the specification of the project and workspace, we retrieve the dataset by utilizing the versioning features of Roboflow in order to gain access to the most recent version of the dataset. By doing so, we ensure that our model training process operates with the most recent data, incorporating any new additions or modifications that have been made to the dataset over the course of various time periods. After that, the dataset is downloaded and incorporated into our development environment so that it can undergo further processing and analysis.

F. Fine-Tuning Yolov8 for Weapon Detection

The process of refining a pre-trained YOLOv8 model involves modifying the existing model weights to better suit a specific task, such as identifying weapons, through additional training on a dataset that has been customized to meet the requirements of the task.

"epochs=" is the option that sets the number of training epochs, which is a phrase that describes the number of times the model will process the entire dataset while it is being trained. Each and every epoch is comprised of both forward and backward sweeps through the network. During these sweeps, the model parameters, also known as weights, are adjusted in order to guarantee the least amount of loss possible. We have set the number of epochs to 35.

When the "imgsz=" option is selected, the size of the input picture that is used for training is specified. The photographs that are stored in YOLOv8 are reduced to a size that has been established in advance before they are uploaded to the network. Increasing the size of the input has the potential to result in higher detection accuracy; however, doing so may need a greater amount of processing resources to be utilized. As part of the process of fine-tuning, the YOLOv8 model undergoes iterative training with a dataset that is specifically designed for it. Through the examination of the photographs and annotations that are provided, it acquires the ability to recognize and classify firearms.

Calculation of Loss: After the forward pass, the predictions of the model are evaluated in comparison to the annotations of the ground truth in order to compute the loss. When it comes to determining the degree of disparity between the expected and real characteristics of an object, YOLOv8 often makes use of a combination of localization loss, classification loss, and confidence loss. Backpropagation, which is another name for the backward pass, is a technique that includes altering the weights of the model by employing gradient descent optimization techniques. This technique requires propagating the estimated loss through the network. In order to limit the amount of weight

that is lost, the gradients provide information on how to alter each weight adjustment. The optimizer makes adjustments to the weights of the model by changing them in a manner that decreases the loss. These adjustments are made using the gradients that are obtained by backpropagation. Through the use of this method, which requires iterating across a large number of epochs, the performance of the model on the weapon detection task might finally be improved.

Evaluation: The performance of the model is evaluated at regular intervals using a separate validation dataset in order to monitor its progress and prevent it from being overfit. When evaluating the effectiveness of detection, it is usual practice to make use of metrics such as precision, recall, and mean average precision (mAP). Using a bespoke dataset, we optimize the performance of the YOLOv8 model in order to accurately detect weapons. This is accomplished through the process of fine-tuning the model. Because of this, its total effectiveness and its capacity to be utilized in applications involving surveillance and security in the actual world are improved.

IV. RESULTS AND DISCUSSION

A. Image Enhancement

During the image enhancement stage of the project, the goal was to enhance the quality and sharpness of aerial footage obtained from drones. This section showcases the outcomes achieved through the image improvement procedure, providing a comprehensive account of the conversion from initial photographs, which were characterized by blurriness and low resolution, to final images that exhibit improved sharpness and clarity. Input photographs: The input photographs consisted of a diverse collection of aerial footage collected by drones, showcasing different environmental situations. The photographs displayed typical difficulties encountered in aerial photography, such as blurriness, noise, and diminished visibility caused by factors such as distance, lighting conditions, and motion blur. Blurred photographs: In order to replicate real-life situations and showcase the effectiveness of image enhancement techniques, a specific group of input photographs was deliberately deteriorated to induce blurriness and decrease image quality. The blurred photographs were used as a reference for comparison, emphasizing the level of deterioration found in regular footage acquired by drones. Output Images: After applying sophisticated image enhancement techniques, such as ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks), the low-quality input images underwent a significant improvement. The resulting photos showed substantial enhancements in clarity, sharpness, and detail, substantially reducing blurriness and improving the overall visual quality of the footage. Visual Comparison: Comparing the input (blurred) photographs with the equivalent output (improved) images clearly demonstrates the effect of the image enhancement method. The resulting photographs exhibit enhanced texture, enhanced edge definition, and heightened overall clarity in comparison to their blurred counterparts. The enhanced photos display clear and distinguishable fine details, including object outlines, texture patterns, and structural characteristics. This contributes to a more informative and visually appealing portrayal of the scene.

Fig 3. ESRGAN Input

Fig 4. ESRGAN Output

B. Object Detection

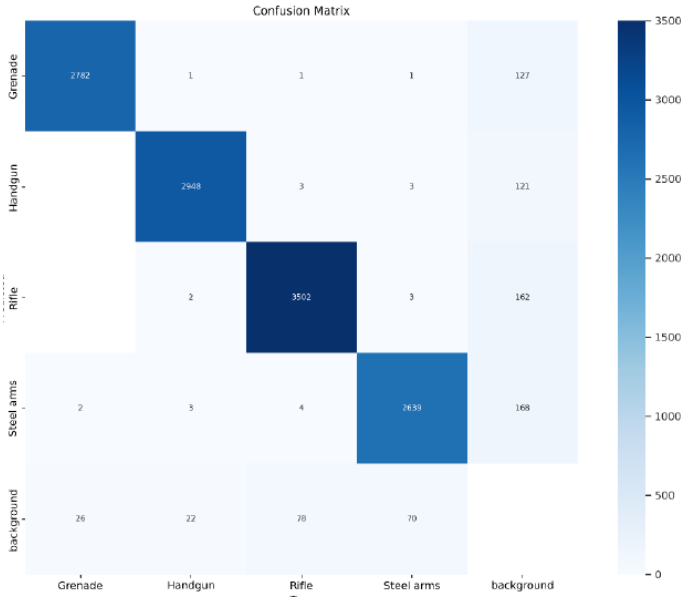

Fig 5. Confusion Matrix of fine-tuned YOLOv8 Model

The objective of the weapon detection phase of the project was to create and assess a reliable system that could precisely

identify and locate weapons in aerial images taken by drones. The following section displays the outcomes of the weapon identification procedure, demonstrating the efficiency of the implemented YOLO (You Only Look Once) algorithm in identifying firearms under different environmental circumstances. In order to make evaluation and benchmarking easier, the input photos were tagged manually with ground truth bounding boxes that indicate the precise placement and size of weaponry within the scene. The annotations functioned as the benchmark against which the effectiveness of the weapon detection system was evaluated. The detection findings consisted of bounding boxes delineating the identified weapons, together with confidence scores reflecting the algorithm's certainty in the accuracy of each identification.

C. Face Recognition

The aim of the project's face detection and authorization phase was to create a system with the ability to precisely identify faces and determine the authorization status of persons. This part showcases the outcomes derived from the face detection and authentication procedure, emphasizing the effectiveness of the implemented face recognition algorithm. The input photos comprised a variety of datasets containing facial images of humans. The datasets included a range of scenarios and environmental variables, accurately reflecting the real-world settings experienced in surveillance applications. A repository of authorized personnel was established, consisting of photographs and accompanying biometric data. This repository served as the basis for facial recognition and authorization. Every record in the database had facial photos and metadata that indicated the person's authorized status. The face detection and recognition algorithm were utilized on the input photographs, leading to the identification of faces and their subsequent acknowledgment. Identified faces were marked by rectangular boxes, while acknowledged persons were compared to the database of authorized personnel to ascertain their authorization status. The system assessed the authorization of each identified face by analysing the recognition outcomes and comparing them with the data recorded in the database. Designated persons were categorized as authorized, whilst unauthorized individuals were identified and marked accordingly.

Fig 6. Face recognition Output

V. CONCLUSION

Overall, the research has effectively proven the viability and efficiency of enhancing cost-effective drone monitoring by incorporating sophisticated image processing methods. The project has successfully tackled important difficulties in surveillance and security applications by improving image quality, implementing real-time object identification, and integrating facial recognition capabilities. The utilization of advanced techniques like ESRGAN in the image enhancement module has greatly enhanced the quality and level of detail in aerial footage obtained from drones. This improvement has resulted in enhanced situational awareness and decision-making abilities. Moreover, the utilization of the YOLO algorithm in the object detection module has facilitated instantaneous recognition and monitoring of diverse items, hence augmenting the system's capacity to identify and address possible hazards. The addition of facial recognition capabilities has boosted the system's powers by allowing for the identification and verification of persons, thereby boosting security and facilitating access control measures. In summary, the project has showcased the capability of cost-effective drone surveillance systems to improve security, oversee vital infrastructure, and assist diverse businesses. Through the utilization of sophisticated image processing methods, the project has created opportunities for the creation of surveillance solutions that are both affordable and capable of adapting to the changing demands of contemporary society.

REFERENCES

1. Guo, G., Wang, H., Yan, Y., Zheng, J., & Li, B. (2020). A fast face detection method via convolutional neural network. Neurocomputing, 395, 128-137.
2. Kong, T., Sun, F., Liu, H., Jiang, Y., Li, L., & Shi, J. (2020). Foveabox: Beyound anchor-based object detection. IEEE Transactions on Image Processing, 29, 7389-7398.
3. Zhu, Y., Cai, H., Zhang, S., Wang, C., & Xiong, Y. (2020). Tinaface: Strong but simple baseline for face detection. arXiv preprint arXiv:2011.13183.
4. Qi, D., Tan, W., Yao, Q., & Liu, J. (2022, October). YOLO5Face: why reinventing a face detector. In European Conference on Computer Vision (pp. 228-244). Cham: Springer Nature Switzerland.
5. Rashid, S. I., Shakibapour, E., & Ebrahimi, M. (2022). Single MR image super-resolution using generative adversarial network. arXiv preprint arXiv:2207.08036.
6. Dilshad, N., Hwang, J., Song, J., & Sung, N. (2020, October). Applications and challenges in video surveillance via drone: A brief survey. In 2020 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 728-732). IEEE.
7. Jiang, Y., & Li, J. (2020). Generative adversarial network for image super-resolution combining texture loss. Applied Sciences, 10(5), 1729.

8. Amit, Y., Felzenszwalb, P., & Girshick, R. (2021). Object detection. In Computer Vision: A Reference Guide (pp. 875-883). Cham: Springer International Publishing.
9. Hasan, M. K., Ahsan, M. S., Newaz, S. S., & Lee, G. M. (2021). Human face detection techniques: A comprehensive review and future research directions. Electronics, 10(19), 2354.
10. Joseph, K. J., Khan, S., Khan, F. S., & Balasubramanian, V. N. (2021). Towards open world object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5830-5840).
11. Larsen, G. D., & Johnston, D. W. (2023). Growth and opportunities for drone surveillance in pinniped research. Mammal Review.
12. Rakotonirina, N. C., & Rasoanaivo, A. (2020, May). ESRGAN+: Further improving enhanced super-resolution generative adversarial network. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 3637-3641). IEEE.
13. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... & Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European conference on computer vision (ECCV) workshops (pp. 0-0).
14. Wang, P., Bayram, B., & Sertel, E. (2021). Super-resolution of remotely sensed data using channel attention based deep learning approach. International Journal of Remote Sensing, 42(16), 6048-6065.
15. Wu, Y., Chen, Y., Yuan, L., Liu, Z., Wang, L., Li, H., & Fu, Y. (2020). Rethinking classification and localization for object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10186-10195).