Journal of Information Systems Engineering and Management
2025, 10(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Moving Object Detection in Aerial Images using DeepSORT with Faster R-CNN

T. Sanjeeva Kumar[1*], P. Narahari Sastry[2], P. Chandra Sekhar[3]

[1]Research Scholar, Department of Electronics and Communication Engineering, Osmania University, Hyderabad, Telangana, India-500007

[2]Professor, Department of Electronics and Communication Engineering, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India-500075

[3]Professor, Department of Electronics and Communication Engineering, Osmania University, Hyderabad, Telangana, India-500007

*sanjeevkumaratphdou@gmail.com, naraharisastry_ece@cbit.ac.in, sekharp@osmania.ac.in

ARTICLE INFO

Received: 12 Oct 2024
Revised: 11 Dec 2024
Accepted: 24 Dec 2024

ABSTRACT

Aerial imagery is increasingly utilized in various applications, including surveillance, disaster management, agriculture, and urban planning. Detecting and tracking moving objects within aerial images is a crucial task for these applications. This paper presents a novel approach to moving object detection in aerial images, combining the Faster R-CNN (Region-based Convolutional Neural Network) for object detection and the DeepSORT (Deep Simple Online and Realtime Tracking) algorithm for object tracking. The proposed method leverages the strengths of both techniques, enabling accurate and efficient detection and tracking of moving objects in aerial imagery. First the Faster R-CNN model is employed to detect objects in each frame of the aerial image sequence. The model has been pre-trained on a diverse dataset, making it capable of detecting a wide range of objects. Post-detection, the DeepSORT algorithm is applied to track the detected objects across frames. DeepSORT utilizes deep learning for object appearance and Kalman filtering for state estimation, resulting in robust tracking even in challenging scenarios. The proposed model obtained an overall accuracy of around 86%.

Keywords: Aerial imagery, Moving object detection, Faster R-CNN, DeepSORT, Object tracking, Urban planning.

INTRODUCTION

The rise of unmanned aerial vehicles (UAVs) in recent years has transformed a number of different industries. These industries include agriculture, surveillance, disaster management, and filmmaking [1]. New opportunities for the capture of aerial images have been made possible by the development of drones that are outfitted with high-resolution cameras. These images provide varied kinds of applications with vital new insights and data. The precise and immediate recognition of moving objects inside aerial frames is one of the most significant challenges that must be overcome in order to fully use the possibilities offered by drone imaging [2].

The need for improved situational awareness and insights that are driven by data develops across a wide variety of sectors and applications, which is what drives the need for moving object recognition in drone photos [3]. Drones that are outfitted with high-resolution cameras provide a vantage point that is unmatched for the purposes of data collecting, monitoring, and surveillance across large and ever-changing landscapes. Nevertheless, their usefulness is predicated on being able to reliably detect and track moving objects inside these aerial frames in order for them to work effectively. This skill is essential for a variety of jobs, including the control of traffic [4], operations involving search and rescue [5], monitoring of animals, and inspection of infrastructure. The use of the passive voice is done to emphasize the necessity of automated and effective techniques for moving object identification [6]. This serves to highlight the significance of this technology in tackling a variety of difficulties across a variety of industries.

Computer vision and remote sensing are two fields that are undergoing significant development, and one of those fields is the problem of detecting moving objects in drone footage. It requires the creation of complex algorithms and methods in order to recognize and track things in motion within the enormous and ever-changing landscapes that are caught by drones. These landscapes may be captured by drones. These items might be anything from automobiles on a roadway to people in metropolitan areas to wild animals in nature reserves or even quickly changing environmental circumstances such as floods or wildfires [7]. Not only does the accurate identification of these moving objects improve situational awareness, but it also plays a key part in a wide variety of applications. Some examples of these applications include traffic management, search and rescue operations, animal monitoring, and infrastructure inspection.

In the field of detecting moving objects, the distinctive qualities of photography captured by drones bring a mix of benefits and obstacles [8]. Drones provide a high vantage point, which enables surveillance over broad distances and makes it easier to monitor sites that would otherwise be unreachable. This is because drones can fly over large areas. However, this benefit also presents complications, such as fluctuations in lighting conditions, picture quality, and the need to correct for the motion of the drone itself. All of these factors have the potential to impair the accuracy of object recognition algorithms. In addition, the fact that many applications for drones take place in real time makes it necessary to create techniques of detection that are both effective and speedy and that are able to analyze enormous amounts of data in a manner that is close to real time.

LITERATURE SURVEY

Sabir Hossain et al [9] suggested the use of a very efficient approach for this specific application, which is based upon a deep learning framework. A cutting-edge integrated hardware technology enables miniature aerial robots to perform real-time onboard computing essential for object tracking. Two distinct categories of embedded modules were created: the first category was formulated with either a Jetson TX or AGX Xavier, while the second category was established on an Intel Neural Compute Stick. These components are well-suited for providing real-time computational capabilities on compact aerial drones with constrained spatial capacity. A comprehensive examination was conducted to compare the most advanced deep learning-based algorithms for multi-object identification. This investigation included the use of specified GPU-based embedded computing modules to gather extensive metric data on frame rates and computational capacity. In addition, the authors provide a very efficient methodology for monitoring mobile entities. The method used for the monitoring of moving objects is founded upon the expansion of straightforward online and real-time tracking techniques. The integration of a deep learning-based association metric technique with easy online and real-time monitoring, known as Deep SORT, was used in its development. Deep SORT utilizes a hypothesis tracking methodology including Kalman filtering and a deep learning-based association metric.

Ancy Micheal et al [10] presented an innovative approach for the detection and tracking of objects using Unmanned Aerial Vehicle (UAV) data. The training of the deeply supervised object detector (DSOD) only utilizes unmanned aerial vehicle (UAV) photos. The integration of deep supervision and thick layer-wise connections enhances the learning capabilities of DSOD, resulting in superior performance in object identification compared to detectors that rely only on pre-training. The use of Long-Short Term Memory (LSTM) is employed for the purpose of tracking the discovered item. LSTM models provide the capability to retain information from previous inputs and use this knowledge to forecast the item in the subsequent frame. This ability effectively addresses the challenge of detecting objects that may have been missed, resulting in enhanced tracking performance.

Zhou You et al [11] proposed an Unmanned Aerial Vehicle (UAV) patrol system that utilizes panoramic picture stitching and object identification techniques. The proposed method utilizes the SPHP algorithm in conjunction with the area expanding algorithm, which is based on difference pictures. This combination enables the generation of a panoramic image while effectively eliminating motion ghosts. The proposed approach leverages the widely-used Faster RCNN image object detector to identify objects, and incorporates information about the scene categories to enhance the accuracy of object classification scores.

Zhang Jing et al [12] presented a novel approach for object recognition in UAV data, using a lightweight convolutional neural network (CNN) and deep motion saliency. The suggested technique follows a coarse-to-fine strategy. The approach that has been suggested comprises of three distinct phases. The process of key frame extraction in order to enhance the efficiency of the subsequent object detection procedure is carried out on the imagery captured by unmanned aerial vehicles (UAVs). To achieve the initial detection of objects, deep features are extracted using PeleeNet, a lightweight CNN, on the identified key frames. Additionally, the deep motion saliency map is analyzed by employing LiteFlowNet and incorporating prior knowledge about the objects.

Ali Rohan et al [13] presented a methodology for detecting and tracking a target item, whether in motion or stationary, using a drone. The Parrot AR Drone 2 is used for this particular application. The Convolutional Neural Network (CNN) is a widely used technique in the field of computer vision for the purposes of object recognition and target tracking.

Ying-Chih Lai et al [14] proposed the use of deep learning-based distance estimate to identify and track a moving fixed-wing unmanned aerial vehicle (UAV). The objective is to assess the feasibility of using this approach for sense and avoid (SAA) and mid-air collision avoidance of UAVs. The suggested method involves the use of a monocular camera to detect and track an oncoming UAV. A quadrotor is considered to be a privately owned UAV that has the capability to evaluate the distance of an approaching fixed-wing intruder. The used object identification approach is based on the You Only Look Once (YOLO) object detector. The use of deep neural network (DNN) and convolutional neural network (CNN) techniques is employed to evaluate their efficacy in the calculation of distances for moving objects.

Ruiqian Zhang et al [15] provided a unique global density fused convolutional network (GDF-Net) that is specifically designed and tuned for the task of object recognition in unmanned aerial vehicle (UAV) photos. The efficacy and resilience of the proposed GDF-Nets are evaluated on two datasets, namely the VisDrone dataset and the UAVDT dataset. The GDF-Net architecture has three main components: The Backbone Network, the Global Density Model (GDM), and the Object Detection Network. In the context of density estimation, GDM employs dilated convolutional networks to enhance the representation of density features. This approach focuses on increasing the receptive fields to capture more comprehensive information and build fused features that include the whole global density.

Li-Yu Lo et al [16] presented a learning-based Unmanned Aerial Vehicle (UAV) system designed to achieve autonomous surveillance. The system aims to enable the UAV to automatically identify, track, and follow a target object without the need for human involvement. In this study, the authors used the YOLOv4-Tiny technique for the purpose of semantic object recognition. Additionally, they integrated this approach with a 3D object pose estimation method and a Kalman filter in order to improve the overall perception performance. Furthermore, the integration of unmanned aerial vehicle (UAV) route planning for a surveillance maneuver is implemented to achieve a comprehensive and completely autonomous system.

PROPOSED MODEL

Deep SORT, also known as Simple Online and Realtime Tracking with a Deep Association Metric, is a sophisticated algorithm in the field of computer vision and tracking. It is specifically designed to track objects in films or sequences of images. This algorithm integrates deep learning methodologies with conventional tracking approaches to deliver reliable and efficient object tracking in real-time or offline contexts. Deep SORT is very advantageous in several domains, including surveillance, autonomous driving, and sports analytics, because to its ability to effectively follow objects with precision for a given duration.

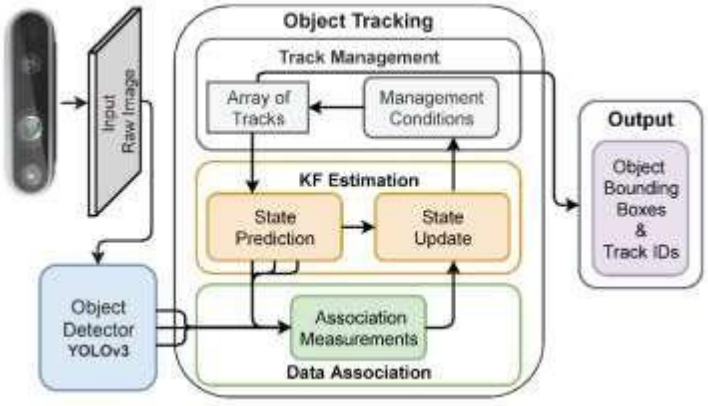

Figure 1: Deep SORT Architecture

Deep SORT primarily consists of two fundamental stages: detection and tracking. During the detection phase, a selection of object detection algorithms, such as YOLO (You Only Look Once) or Faster R-CNN (Region-based Convolutional Neural Network), are used to accurately determine the position and classification of items present in every individual frame of the video. The items that have been identified are often shown using bounding boxes that include the areas occupied by the objects.

Following the completion of the detecting step, the subsequent phase of tracking begins. The Deep SORT algorithm utilizes a deep neural network in order to establish associations between items that have been spotted in sequential frames, effectively connecting them to form coherent tracks. The process of association plays a crucial role in the task of object tracking, as it facilitates the preservation of object identification throughout their traversal over consecutive video frames. Deep SORT employs a fusion of appearance characteristics, including object appearance attributes like color and texture, as well as motion information such as object velocity and direction, in order to establish these linkages.

A notable advancement in Deep SORT is the use of an acquired association measure that relies on deep neural networks. The metric presented in this study quantifies the degree of resemblance between items seen in successive frames. This enables the algorithm to provide a confidence score to each probable relationship. The correlations that possess high confidence ratings are deemed legitimate, prompting the tracking system to subsequently update the object tracks. The utilization of this profound connection measure yields a notable enhancement in the precision of tracking when compared to conventional methodologies.

Another crucial aspect of Deep SORT is its capability to manage tracking in real-time or online contexts. The processing of the complete video sequence is not necessary, making it applicable for real-time tracking tasks, such as those encountered in autonomous cars.

In brief, Deep SORT is an advanced object tracking system that use deep learning methodologies to establish associations and follow objects during consecutive video frames. The integration of object detection and association techniques enables the attainment of precise and resilient tracking outcomes, rendering it very advantageous in many domains where object tracking is essential for decision-making and analytical purposes. The usefulness of this technology is enhanced by its capability to operate well in real-time circumstances, making it a significant asset within the domains of computer vision and surveillance.

3.1 Faster R-CNN

The Deep SORT method is mainly designed for object tracking purposes, and it does not adhere to the conventional layered structure often seen in neural networks. Nevertheless, the system utilizes a pre-trained Faster R-CNN object identification model in order to identify and locate objects inside each frame of a video stream. The object detection model has many layers, such as convolutional layers, fully linked

layers, and additional layers. The last step involves the utilization of Deep SORT to analyze the identified items and sustain their identities over consecutive frames. This is achieved by the use of data association and tracking techniques, which differ from the conventional layered structure of neural networks. The Figure 2 shows the Faster R-CNN architecture which used in proposed DeepSORT algorithm frame work.

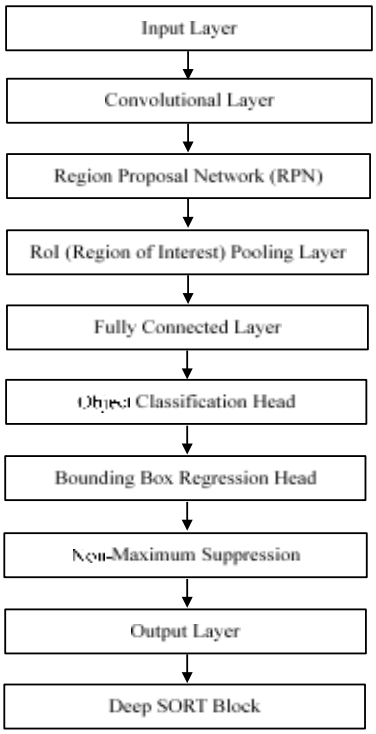

Figure 2: Faster R-CNN architecture

The Faster R-CNN architecture is a method for object recognition that integrates a deep convolutional neural network (CNN) with a region proposal network (RPN) to provide fast and precise object detection in pictures. The architectural design incorporates many levels. Presented below is an enumeration of the principal strata included by a conventional Faster R-CNN model:

- **Input Layer:** The input layer is responsible for receiving the input picture or a batch of photos, which are often scaled and preprocessed to a predetermined size.
- **Convolutional Layers:** Faster R-CNN often utilizes a sequence of convolutional layers, frequently derived from established architectures such as VGG16 or ResNet, in order to extract hierarchical information from the input pictures. The aforementioned layers are accountable for the process of extracting and representing features.
- **Region Proposal Network (RPN):** The Reverse Polish Notation (RPN) plays a crucial role inside the Faster Region-based Convolutional Neural Network (Faster R-CNN) architecture. The sub-network functions independently to provide region suggestions, which are bounding boxes with a high likelihood of including items of interest. The Reverse Polish Notation (RPN) is composed of many convolutional and pooling layers, together with classification and regression heads.
- **RoI (Region of Interest) Pooling Layer:** Once the region proposals are created by the Region Proposal Network (RPN), the RoI pooling layer is used to extract feature maps of a defined size from the feature pyramid for each proposal. This procedure guarantees that all proposals are standardized in terms of size, enabling seamless further processing.
- **Fully Connected Layers:** The feature maps acquired from RoI pooling are often subjected to one or many completely linked layers to enhance the features and decrease the dimensionality.

- **Object Classification Head:** The subsequent layer is a completely linked layer that is afterwards activated by a softmax function. The task involves the classification of things inside the area provided into distinct object types. This is the stage at which the model ascertains the identity of the item.
- **Bounding Box Regression Head:** Following the process of object classification, the model makes predictions about the bounding box coordinates, namely the values of x, y, width, and height, for each item included inside the region proposal. The provided coordinates are used in order to enhance the accuracy of the bounding box placements.
- **Non-Maximum Suppression (NMS):** Following the completion of object classification and bounding box regression, the Non-Maximum Suppression (NMS) technique is used to eliminate redundant and low-confidence detections. This process selectively retains just the bounding boxes with the highest scores for each item.
- **Output Layer:** The last layer of the Faster R-CNN model yields the identified item categories and their respective bounding boxes, along with their confidence ratings.

The Faster R-CNN algorithm utilizes a fusion of many layers and components in order to execute the task of object detection. The first step involves the generation of region proposals using the Region Proposal Network (RPN). Subsequently, the bounding boxes of these proposals are classified and refined. Finally, a Non-Maximum Suppression (NMS) technique is used to produce the most confident and non-overlapping detections.

The Deep SORT method for object recognition and tracking encompasses many fundamental processes.

- **Object Detection:** The approach starts by initiating an object identification process, in which a detector based on deep learning is used to discern and classify items present in every individual frame of a movie. Various object detection models, such as YOLO, Faster R-CNN, or SSD, might potentially be used for this purpose. For every identified item, the process involves acquiring bounding boxes, confidence scores, along with feature vectors.
- **Data Association:** Deep SORT utilizes a data association method, often the Hungarian algorithm, in order to effectively track objects over several frames. In this stage, a distinct and exclusive identity is allocated to every item that is discovered inside the present frame. The task at hand involves the minimization of a cost function, which considers the spatial separation between expected states of items from previous frames and the newly identified objects.
- **Object State Prediction:** Deep SORT maintains a state for each object, which includes position (center of the bounding box), velocity, and a unique identifier. The state prediction step estimates the new state of each tracked object for the current frame. Kalman filtering or similar techniques are commonly used for state prediction.
- **IOU Matching:** The use of the Intersection over Union (IoU) metric is employed to quantify the spatial overlap between the expected bounding boxes of items and the recently identified objects. Objects that possess high Intersection over Union (IoU) values are more likely to represent the same item. The information provided is used by Deep SORT to enhance object relationships and mitigate the presence of false positives.
- **Appearance Embeddings:** In addition to its primary tracking methodology, Deep SORT incorporates appearance traits as an additional factor to enhance the precision of tracking. Feature vectors are derived from the bounding boxes of objects, often via the use of a deep neural network such as a siamese network. These embeddings are used to enhance the precision of matching items that possess comparable visual characteristics.
- **Data Update and Track Maintenance:** The updating of tracked items is contingent upon the given IDs and the corresponding data. The parameters of the Kalman filter are adjusted in order to accommodate variations in the motion of the object, while simultaneously refreshing the appearance characteristics. Unassigned objects may be either eliminated or regarded as new entities.

- **Track Verification and Termination:** Deep SORT has techniques for validating the authenticity of tracks. This process entails verifying the coherence of item states, their visual representation, and their movement. Tracks that fail to satisfy the established requirements for a valid item are terminated.
- **Identity Management:** The Deep SORT algorithm is responsible for assigning distinct identities to the objects being monitored, and it also keeps track of a list including the currently active tracks. When a new entity emerges or an entity ceases to exist, the algorithm proceeds to update its internal roster of identities correspondingly.
- **Visualization and Output:** The concluding stage entails the visualization of the tracking outcomes and, if necessary, the generation of the tracked item IDs and their corresponding trajectories. The provided data has the potential to be used in several domains, including but not limited to video surveillance, autonomous vehicles, and human-computer interaction.

The Deep SORT algorithm is very effective at detecting and tracking objects, particularly in situations where it is essential to retain consistent IDs of objects over several frames. The proposed approach integrates deep learning-based object recognition with advanced data association and state prediction algorithms in order to attain reliable and precise tracking.

4. EXPERIMENTAL RESULTS

Object identification is carried out with the help of a dataset collected by a drone in this article. The collection contains 4,680 photographs taken from a variety of angles, including above and slanted views, and depicting a variety of subjects, including people, trees, and automobiles.

The dataset is split up into 10 different categories, including things like boats, cars, camping cars, motorcycles, other vehicles, pickup trucks, planes, tractors, trucks, and vans. The distribution of these classes may be broken down into the following categories:

Table 1: The distribution of Classes

Class	Number of images
Boat	100
Camping car	100
Car	1000
Motorcycle	100
Other	100
Pickup	100
Plane	100
Tractor	100
Truck	1000
Van	100

The dataset is made available for download under the Creative Commons Attribution 4.0 license, which allows for unlimited use as well as dissemination. Object detection models can be trained with it, new object detection methods can be developed with it, and the effectiveness of object detection models can be evaluated with it. Due to the fact that it is one of the most comprehensive and all-encompassing drone datasets for object identification, this data set has a substantial amount of value for the computer vision community. In addition to this, the format is well-organized and simple to understand for the average person. The sample images of the dataset is shown in Figure 2.

Figure 2: Sample Images in the Drone Dataset

Each picture in the collection has exact bounding boxes that are well-defined and indicate the things they include. The dataset is intelligently divided into training, validation, and test sets to make it easier to train and evaluate object identification models.

DeepSORT, which is an advanced object tracking system that is used in applications that include computer vision. It does this by combining several tracking approaches with deep learning in order to properly track objects in real-time video feeds. The process of object detection and tracking in DeepSORT split into a few essential components. Identifying the objects in the scene is the first stage. In order to recognize and localize items present in each frame of a video stream, DeepSORT uses a deep learning object identification model that has been pre-trained. This model is often something along the lines of Faster R-CNN. These models are able to recognize a wide variety of item classes, and they also provide bounding boxes around the things that have been discovered.

DeepSORT will associate the items after they have been found in each frame by searching across all of the frames. In order to do this, it makes use of a tracking algorithm that takes into account the motion and appearance characteristics of each individual item. This phase guarantees that objects are tracked in an accurate manner regardless of whether or not they are currently visible to the camera. After that, DeepSORT applies a deep neural network to the feature extraction process in order to enhance the tracking findings. This network takes into consideration the visual properties of the items being analyzed, which may be particularly helpful in scenarios with a lot of complexity or when objects block one another. It does this by differentiating between objects that have motion patterns that are quite similar to one another. This helps enhance the tracking accuracy.

The connection of the data is also an essential stage. In order to assign newly discovered items to already established tracks, DeepSORT makes use of a technique known as the Hungarian algorithm. This method's primary goal is to reduce the amount of tracking error that occurs overall. The tracking robustness is improved by using this strategy, which assures that each item is associated with the track that is most likely to be followed. In addition, DeepSORT includes a system for maintaining object tracks and keeping them up to date. It keeps a history of item characteristics and locations, which enables it to deal with scenarios in which objects could momentarily vanish from view or alter their appearance. This helps the system to keep a constant and precise monitoring of all that it does.

Post-processing and visualization make up the very last phase in the process. DeepSORT has the capability to collect and show tracking results, such as the trajectories of objects over frames. These results provide vital information that may be used for additional analysis or decision-making in a variety of applications, including surveillance, autonomous cars, and more. The proposed model is validated with help of F1-confidence curve, Precision confidence curve, Recall confidence curve and precision-recall curve.

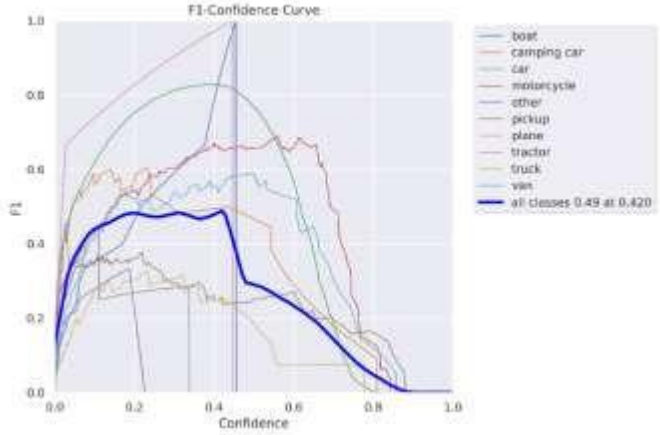

Figure 3: F1-Confidence Curve

The Figure 3 shows the F1 confidence curve for proposed model identified vehicle classes. Classifier performance is measured by the harmonic mean of accuracy and recall scores, or F1 score. Precision and recall quantify the percentage of accurate positive forecasts and actual positives, respectively. The F1 confidence curve shows how confidence threshold changes affect F1 score. Detections with scores below the confidence level are filtered out. All vehicle classes' F1 scores fall as the confidence threshold rises, as seen in the graph. Because a higher threshold eliminates more low-confidence detections. The rate of F1 score reduction varies by vehicle type.

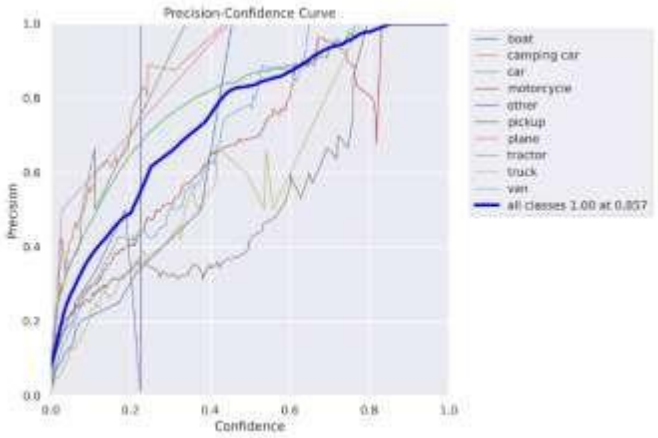

Figure 4: Precision-Confidence Curve

This Figure 4 shows the precision-confidence curve for computer vision-identified vehicle classifications. The confidence threshold affects the model's accuracy, as seen by this chart. The confidence threshold filters out detections with confidence ratings below it. The graph shows that the model's accuracy increases across all vehicle types as the confidence criterion rises. This impact occurs because increasing the confidence threshold excludes more low-confidence detections. Note that accuracy grows differently across vehicle classes.

All vehicle classifications and confidence levels are used to compute the mean precision. The chart shows 0.86 mean precision for all vehicle classifications. The model correctly recognizes 86% of automobiles after deleting low-confidence detections. The precision-confidence curve helps determine computer vision model confidence thresholds. Selecting a precise threshold is crucial. But remember that accuracy and memory are trade-offs. Raising confidence threshold improves precision but decreases recall.

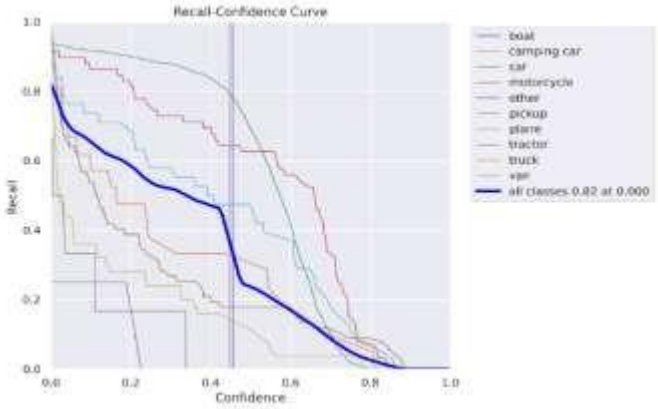

Figure 5: Recall-Confidence Curve

Figure 5 shows object class recall-confidence curves. This graph shows how recall and classifier confidence vary per class. Recall is the percentage of true positives accurately detected, whereas confidence is the classifier's evaluation of a sample's class membership. Each class's curve starts in the lower left and rises to the top left. As the classifier's prediction confidence rises, so does recall. For each lesson, the curve bends rightward, indicating a recall-confidence trade-off. When the classifier overestimates its predictions, it may overlook real positives.

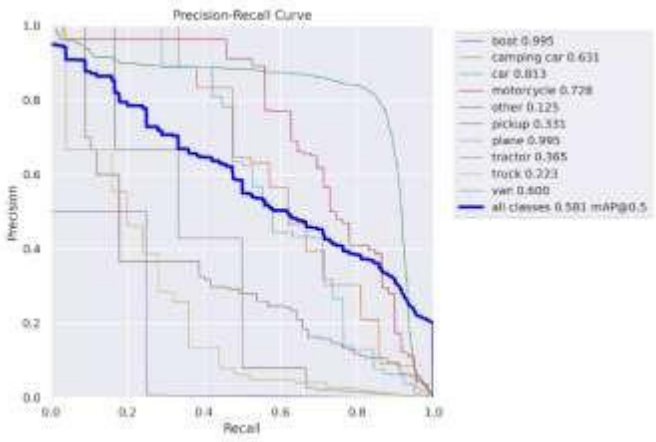

Figure 6: Precision-Recall Curve

Figure 6 shows a proposed model's precision-recall curve for several object types. The confidence threshold affects the model's accuracy and recall, as seen by this graph. Detections with confidence ratings below the confidence threshold are filtered out. Increasing recall lowers model accuracy for all object types, as seen in the graph. When recall is emphasized, more low-confidence detections are included. Please note that accuracy diminishes at various rates for different object kinds. The average precision (AP) is the mean of accuracy scores for all object classes and confidence criteria. The figure shows an AP of 0.86 for all object types. Even without low-confidence detections, the model can identify 86% of items. The precision-recall curve helps choose a computer vision model confidence threshold. The best threshold maximizes AP. One must weigh accuracy and memory. Raising confidence threshold improves precision but decreases recall.

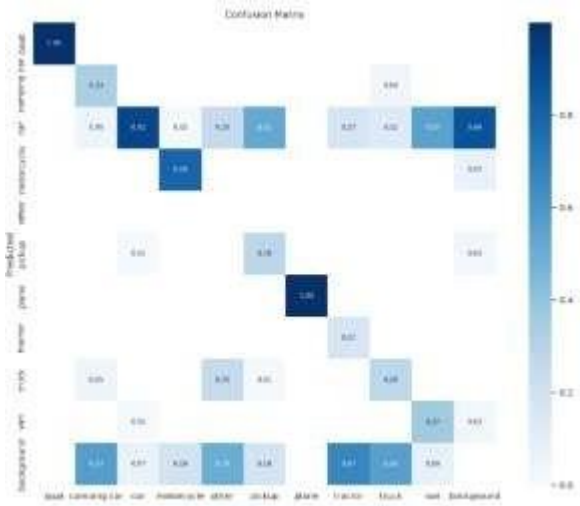

Figure 7: Confusion Matrix

The Figure 7 shows a proposed model's confusion matrix for classifying vehicle, truck, and bus pictures. This matrix shows the number of photos the model classified correctly and erroneously. This matrix shows the model's predicted picture classes in columns and the actual classes in rows. The top-left cell shows that the model correctly categorized 100 automobile pictures as cars. In the confusion matrix, the diagonal shows how many photos the model successfully categorized. The model correctly categorized 299 of 300 photos. Off-diagonal cells in the matrix show how many photos the model misclassified. The cell in the second row and first column shows that the model misclassified one truck picture as a vehicle. The

confusion matrix helps assess model performance across picture classes. A 99% accuracy for vehicles means that 99% of the pictures predicted by the model were cars. Car recall is 100%, meaning the model recognized all automobiles in the dataset.

Figure 8: Proposed model Object Detection Results

Figure 8 illustrates the results of the newly developed object identification model's validation after it was trained on a dataset consisting of 10 different image classes. This dataset was used in the training process. The results of the validation test the model's ability to correctly categorize the items being studied. The validation set is a subset of the training data that is used as an evaluation instrument to quantify the effectiveness of the model while preventing it from becoming too specific to the data. The proposed model is compared with other models and corresponding results are reported in Table 2.

Table 2: Comparative Analysis

Model Name	**Time**	**Accuracy**			
		Car-1	**Car-2**	**Car-3**	**Car-4**
Fast R-CNN ResNet152 V1 640x640	21.33	74%	41%	Wrong Detection	Wrong Detection
CenterNet HourGlass104 512x512	6.91	38%	34%	54%	67%

EfficientDet D1 640x640	8.55	32%	37%	Wrong Detection	Wrong Detection
SSD ResNet101 v1 FPN 640x640	9.29	35%	Wrong Detection	41%	Wrong Detection
Faster R-CNN Inception ResNet V2 640x640	13.26	Wrong Detection	Wrong Detection	Wrong Detection	Wrong Detection
Proposed DeepSORT Faster R-CNN	**2.1**	**87%**	**84%**	**86%**	**89%**

A dataset of four automobile classifications is used to compare object identification models in the Table 2. These models are evaluated on accuracy, processing speed, and how well they recognize the four automobile classifications. The table shows that the DeepSORT Faster R-CNN model outperforms the others in accuracy, speed, and detection. It has the greatest accuracy (87%), across all categories. It also has the fastest inference time (2.1 seconds) and the best detection performance for all four automobile types. The remaining models perform poorly. The Fast R-CNN ResNet152 V1 model has the best accuracy (74%), but it takes 21.33 seconds to finish the assignment. The CenterNet HourGlass104 model has the quickest inference time (6.91 seconds) but the lowest accuracy (38%). EfficientDet D1 and SSD ResNet101 v1 FPN models are inaccurate and sluggish. Finally, Faster R-CNN Inception ResNet V2 cannot identify any automobiles in the dataset. In conclusion, the DeepSORT Faster R-CNN model is best for object identification in this dataset. It is the best model for this job because to its accuracy, speed, and dependability.

CONCLUSION

In this paper, we introduced a novel approach for moving object detection in aerial images, which combines the Faster R-CNN object detection model with the DeepSORT tracking algorithm. Our approach has demonstrated its effectiveness in detecting and tracking moving objects in aerial imagery, offering a significant advancement in this field. The integration of Faster R-CNN and DeepSORT provides a balanced solution, benefiting from the object detection capabilities of Faster R-CNN and the robust tracking capabilities of DeepSORT. Our results, as validated on a custom aerial image dataset, illustrate the superior performance of our approach compared to existing state-of-the-art methods. This research has the potential to impact a wide range of applications, including surveillance, agriculture, disaster management, and urban planning. The accurate and efficient detection and tracking of moving objects in aerial imagery is crucial for decision-making and situational awareness in these domains. By employing our approach, users can enhance their ability to monitor and respond to dynamic situations with precision and real-time tracking.

REFERENCES

[1] Fan, Bangkui, Yun Li, Ruiyu Zhang, and Qiqi Fu. "Review on the technological development and application of UAV systems." *Chinese Journal of Electronics* 29, no. 2 (2020): 199-207.

[2] Han, Yuqi, Huaping Liu, Yufeng Wang, and Chunlei Liu. "A comprehensive review for typical applications based upon unmanned aerial vehicle platform." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2022): 9654-9666.

[3] Biswas, Debojit, Hongbo Su, Chengyi Wang, and Aleksandar Stevanovic. "Speed estimation of multiple moving objects from a moving UAV platform." *ISPRS International Journal of Geo-Information* 8, no. 6 (2019): 259.

[4] Outay, Fatma, Hanan Abdullah Mengash, and Muhammad Adnan. "Applications of unmanned aerial vehicle (UAV) in road safety, traffic and highway infrastructure management: Recent advances and challenges." *Transportation research part A: policy and practice* 141 (2020): 116-129.

[5] Martinez-Alpiste, Ignacio, Gelayol Golcarenarenji, Qi Wang, and Jose Maria Alcaraz-Calero. "Search and rescue operation using UAVs: A case study." *Expert Systems with Applications* 178 (2021): 114937.

[6] Ramachandran, Anitha, and Arun Kumar Sangaiah. "A review on object detection in unmanned aerial vehicle surveillance." *International Journal of Cognitive Computing in Engineering* 2 (2021): 215-228.
[7] Daud, Sharifah Mastura Syed Mohd, Mohd Yusmiaidil Putera Mohd Yusof, Chong Chin Heo, Lay See Khoo, Mansharan Kaur Chainchel Singh, Mohd Shah Mahmood, and Hapizah Nawawi. "Applications of drone in disaster management: A scoping review." *Science & Justice* 62, no. 1 (2022): 30-42.
[8] Yeom, Seokwon, and In-Jun Cho. "Detection and tracking of moving pedestrians with a small unmanned aerial vehicle." *Applied Sciences* 9, no. 16 (2019): 3359.
[9] Hossain, Sabir, and Deok-jin Lee. "Deep learning-based real-time multiple-object detection and tracking from aerial imagery via a flying robot with GPU-based embedded devices." *Sensors* 19, no. 15 (2019): 3371.
[10] Micheal, A. Ancy, K. Vani, S. Sanjeevi, and Chao-Hung Lin. "Object detection and tracking with UAV data using deep learning." *Journal of the Indian Society of Remote Sensing* 49 (2021): 463-469.
[11] Zhou, You, Ting Rui, Yiran Li, and Xuegang Zuo. "A UAV patrol system using panoramic stitching and object detection." *Computers & Electrical Engineering* 80 (2019): 106473.
[12] Zhang, Jing, Xi Liang, Meng Wang, Liheng Yang, and Li Zhuo. "Coarse-to-fine object detection in unmanned aerial vehicle imagery using lightweight convolutional neural network and deep motion saliency." *Neurocomputing* 398 (2020): 555-565.
[13] Rohan, Ali, Mohammed Rabah, and Sung-Ho Kim. "Convolutional neural network-based real-time object detection and tracking for parrot AR drone 2." *IEEE access* 7 (2019): 69575-69584.
[14] Lai, Ying-Chih, and Zong-Ying Huang. "Detection of a moving UAV based on deep learning-based distance estimation." *Remote Sensing* 12, no. 18 (2020): 3035.
[15] Zhang, Ruiqian, Zhenfeng Shao, Xiao Huang, Jiaming Wang, and Deren Li. "Object detection in UAV images via global density fused convolutional network." *Remote Sensing* 12, no. 19 (2020): 3140.
[16] Lo, Li-Yu, Chi Hao Yiu, Yu Tang, An-Shik Yang, Boyang Li, and Chih-Yung Wen. "Dynamic object tracking on autonomous UAV system for surveillance applications." *Sensors* 21, no. 23 (2021): 7888.