Object Detection in Aerial Imagery

Advanced Machine Learning

(MIS-64061-001)

Submitted by

Venkata Naga Siddartha Gutha

Table of contents:

Summary:

Aerial object detection involves the use of computer vision and machine learning techniques to detect and classify objects in aerial imagery or video captured by drones, satellites, or other aerial platforms. Aerial object detection has challenges such as variations in object scale, altitude, view angles, and lighting/weather conditions. It has numerous applications, including agriculture, disaster response, surveillance, urban planning, and environmental monitoring. Datasets play a major role in the performance of any machine learning algorithm and it is same for aerial object detection as well. There are various publicly available datasets for aerial object detection, such as DOTA, VisDrone. These datasets provide a large number of annotated aerial images, which facilitate the development and evaluation of object detection algorithms. Faster R-CNN, YOLO, and SSD are just a few of the deep-learning models that have been suggested for aerial object detection. These methods extract information from aerial photos using convolutional neural networks (CNNs), which they then use to conduct object detection and categorization. To increase precision and effectiveness, some models also use methods like feature pyramid networks (FPNs), Single Shot Detection (SSD) and Faster R-CNN. There is ongoing research on improving the accuracy, efficiency, and robustness of aerial object detection methods. To improve detection performance, researchers are testing new architectures, loss functions, and training techniques. New datasets and evaluation criteria are also being created to better reflect the particular difficulties and demands of aerial object detection. Overall, the study of aerial object identification is a topic that continually evolves, and researchers are constantly looking for new approaches and technologies to overcome its challenges and expand its capabilities.

Introduction:

Detecting objects in aerial images is a crucial and difficult task that involves identifying and labeling each visible object in the image with a corresponding category. This task is essential for many applications such as managing land resources, monitoring ecosystems, and evaluating land ecosystems [1]. Object recognition has been extensively studied and is effective in spotting objects in sharp images captured by ground-based cameras. Unmanned aerial aircraft with cameras are being used more often in a variety of uses, including security monitoring, agricultural, and disaster relief, which has opened up new possibilities for computer vision applications [2].

Numerous real-time applications, including surveillance, delivery services, traffic monitoring, agricultural, disaster management, and marine surveillance, make extensive use of drones. Amazon's use of drone delivery has gotten federal approval. According to a study by Hii, Courtney, and Royall (2019), it is viable to use drones to distribute medications. Drone use in precision agriculture is anticipated to develop significantly and is becoming an essential component of managing farm tasks [3].

Due to differences in size within the same category and the spatial resolution of the sensors, object instances in aerial photographs show enormous scale fluctuations. Aerial photos sometimes include dense, little objects in random orientations, as well as cases with incredibly

huge aspect ratios, such as bridges [4]. Object detection in aerial photos is a popular application, although datasets like UCAS-AOD and NWPU VHR-10 frequently use idealized examples that fail to accurately capture the complexity of real-world applications. To further research in Earth Vision, a large-scale dataset for object detection in aerial photographs like UAVDT [2], DOTA[4], and VIVID[ ] has been established to address this problem.

There are several popular frameworks for aerial object detection. Single-shot detectors and two-stage detectors are two common frameworks for airborne item detection. Single-shot detectors, such YOLO (You Only Look Once) and SSD (Single Shot Detector) [3], are quicker and more effective since they only need to pass through the neural network once in the forward direction. The two-stage method and classification used by two-stage detectors like Faster R-CNN (Region-based Convolutional Neural Network) and R-FCN (Region-based Fully Convolutional Network) [3], however, often results in higher accuracy. A lot of frameworks also use feature pyramid networks (FPN [3]) to manage objects of different scales and sizes. These frameworks have greatly enhanced the performance of aerial object recognition and are utilized in many different applications, including disaster management, ecological monitoring, and land resource management.

Challenges in aerial object detection:

Due to the distinctive qualities of aerial photos, object detection in aerial photographs presents a number of challenges compared to typical object detection. When compared to conventional ground-based cameras, the usage of UAV-mounted cameras provides more difficulties in the detection of aerial objects[2]. Variations in object scale and altitude provide a problem since the size of objects in an image is affected by the UAV's flying altitude. There may be differences in object scale throughout the film due to the camera's ability to record differing levels of object detail at various altitudes [2]. The different view angles that UAVs may record photos from, including bird's-eye views that are uncommon in ground-based object recognition, present another difficulty. The UAV-based detection model must be able to manage these variations in visual appearance because this can lead to objects having arbitrary orientations and aspect ratios. Finally, in outdoor settings where UAVs are frequently used, variations in lighting and weather can have a significant impact on object visibility and appearance [2].

Applications of Aerial object detection

Unmanned aerial vehicle (UAV)-based aerial object detection, which uses cameras and sensors to detect and track objects on the ground, has grown in importance and has a wide range of applications. Due to their ability to fly at high altitudes and quickly cover large areas, UAVs have shown to be an effective tool for applications including search and rescue, surveillance, agriculture, environmental monitoring, and more. Aerial object identification is crucial in this scenario for providing vital information to decision-makers in real-time so they may make informed decisions and take the required actions. Following are the few areas with a wide range of applications of Aerial object detection

Traffic surveillance:

Traffic monitoring using aerial object detection offers a lot of potential, especially in big, busy cities [3]. UAVs with high-resolution cameras can record real-time traffic conditions, including vehicle density, traffic flow, and congestion levels since they can see a larger area from an elevated viewpoint This data can be analyzed to improve traffic management techniques like modifying traffic lights, rerouting cars, and giving drivers real-time traffic updates. [5] Additionally, by supplying high-quality video as proof, airborne object detection can help in identifying and following traffic infractions like speeding or reckless driving. Aerial object detection offers an effective solution for traffic surveillance and can also help traffic authorities to make wise judgments and enhance the management and safety of traffic [6].

Civil industry:

There are several civil industry uses for aerial object identification. It can be useful for maintaining and monitoring infrastructure like electricity lines, pipelines, and bridges. Large building sites might benefit from surveys to monitor progress and spot potential dangers. Aerial object identification also offers high-resolution photos for analysis of land usage and development, which can help with urban planning. By analyzing aerial images, the system can detect changes in the river banks caused by erosion, landslides, and floods, providing early warning signs for potential disasters[7]. In order to locate damaged locations and gauge the severity of the damage, it can also be utilized in disaster response and relief activities. Overall, aerial object identification offers a quick and economical technique to obtain crucial information for making decisions in the civil business.

Military:

There are numerous uses for aerial object detection in military operations. It can be used for many different things, such locating targets and conducting surveillance and reconnaissance. Unmanned aerial vehicles (UAVs) equipped with object detection systems can provide real-time intelligence regarding the positions, movements, and tools of adversaries. They can also be used to monitor border areas and alert people to potential threats in advance. Tanks, planes, and ships can all be identified and monitored by aerial object detection systems in addition to military personnel. [8]. By enhancing situational awareness and facilitating more effective decision-making, the adoption of such systems can increase mission success rates while lowering dangers to military personnel.

Disaster response and recovery

In disaster response and recovery, aerial object detection is helpful because it can offer a quick and effective means to gauge the level of damage brought on by natural disasters like hurricanes, earthquakes, and floods.[9] Rescue teams can properly prioritize their work and distribute resources by using aerial pictures to discover blocked highways, collapsed buildings, and flooded areas.[3] Aerial object detection can also help in search and rescue operations by detecting persons in need of help and identifying survivors. The technology can also be used to track how recovery operations are going and spot any places that still need work.

Environmental monitoring

Environmental monitoring and recovery have found aerial object detection to be highly helpful. Natural disasters like forest fires, oil spills, and floods can be detected and monitored with its

help, allowing for a more efficient and targeted response [10]. In addition, it can be used to study vegetation and land use, monitor wildlife movement, and assess the effects of climate change. In disaster recovery activities, aerial object detection can be used to map and assess the damage, monitor the progress of restoration and recovery efforts, and identify possible problem areas.

Aerial object detection has demonstrated to be an efficient and effective way to acquire vast volumes of data, which can then be reviewed and used to make decisions. There are many additional industries in which it can play a significant role. Given its ability to quickly cover huge areas, provide high-resolution images and videos, and identify objects that are challenging to see with the human eye, it is the ideal tool to decrease human labor and enhance decision-making processes. Therefore, aerial object identification holds the promise of revolutionizing a wide range of sectors as well as how we work and interact with our environment.

Datasets available for training Aerial object detection:

The creation of accurate and reliable aerial object detection models depends heavily on data sets. They offer a wide variety of annotated images, allowing the algorithms to accurately learn and detect items. Researchers and developers may train their models with a variety of object types, sizes, forms, and orientations thanks to access to big and extensive data sets. The availability of annotated data sets additionally enables the evaluation and comparison of various detection methods, assisting in the determination of the best strategy for a particular application. The creation and testing of transfer learning approaches are also made possible by data sets, allowing models to be trained on one data set and then applied to another with similar object classes. Following are few benchmark datasets used for training Aerial object detection models

DOTA:

DOTA (Dataset for Object Detection in Aerial Images) is a large dataset for object detection in aerial photographs that includes 188,282 annotated object instances over 2806 high-resolution images. [4]It covers a variety of object types, including buildings, automobiles, ships, and aircraft. Due to its size and diversity, the DOTA dataset has emerged as a benchmark in the field of aerial object detection.

The evolution of object detection techniques for aerial photos has greatly benefited from the availability of the DOTA dataset [17]. On the basis of this dataset, researchers can train their models and assess the performance of those models using common assessment measures.

The DOTA dataset offers a wide range of real-world uses, including transportation, disaster response, and urban planning. Decision-making across a range of industries can benefit from the precise recognition and tracking of objects in aerial photographs [4]. For instance, in urban planning, the identification of buildings and road networks might reveal information about the rates of habitation and the movement of traffic. In the field of transportation, the identification of automobiles and ships can help with traffic control and waterway surveillance. Finding

damaged structures and infrastructure during a disaster response can help determine the degree of the damage and organize recovery operations.

UAVDT:

Unmanned Aerial Vehicle Detection and Tracking (UAVDT) is a publicly available dataset that was made specifically for developing and testing aerial object recognition and tracking algorithms. The dataset consists of 10,209 video clips with an average length of 30 seconds that were captured by a drone flying over various urban and rural locations while equipped with a high-resolution camera. [11]. The objects included in the videos are diverse and include people, cars, animals, and other items. The dataset includes annotations for both object detection and tracking, making it appropriate for the creation and testing of algorithms for each job. Significant progress has been made in the field as a result of the UAVDT dataset, which has grown in popularity as a benchmark for assessing the effectiveness of airborne object detection and tracking algorithms.[11]

VisDrone

VisDrone is a large-scale benchmark dataset for drone-based vision research. It is intended to assess how well several computer vision algorithms for drone-based surveillance applications, including object detection, tracking, and counting, among others, perform. The collection consists of more than 2 million photos and more than 10 hours of video that were taken by drones at various altitudes and with various camera configurations, weather, and lighting conditions.[12] As it encompasses a wide range of items, such as pedestrians, automobiles, and other objects of interest, as well as occlusions, scale differences, and cluttered backdrops, VisDrone presents a diverse set of problems for academics in the field.

Applications for the VisDrone dataset are numerous and include crowd management, disaster response, crowd surveillance, and environmental monitoring. For instance, it can be used in traffic monitoring and management systems to track vehicles and pedestrians, to identify and track environmental changes like deforestation and coastal erosion, and to support search and rescue operations in disasters. The VisDrone dataset is continuously updated and expanded and it is an essential tool for aerial vision researchers and developers since it offers a trustworthy standard for assessing object detection and tracking algorithms.

There are several other datasets that are being developed and these are essential for the creation and testing of aerial object detection algorithms. They assist discover the strengths and shortcomings of various models and algorithms by providing a standardized and dependable benchmark for performance comparison. For the field and its applications to advance, a robust data set is necessary. It can significantly improve the precision and efficiency of airborne object identification systems.

Object detection models in aerial object detection

The automatic identification of items of interest in aerial photography is made possible by object detection models, which are essential for aerial object detection. These models provide predictions about the location and class of items in new photos by using machine learning methods and deep neural networks to learn the traits and attributes of objects in the imagery. Aerial vision systems' capabilities can be considerably improved by accurate and effective object identification models, making them more useful and efficient for a variety of tasks like monitoring, surveillance, and disaster response [6,8,9,10]. The development and improvement of object detection models are essential for the advancement of aerial object detection technology.

YOLO:

YOLO (You Only Look Once) is a real-time object recognition system created for quick and precise object detection in still photographs and moving pictures. In 2015, Joseph Redmon offered the idea for the first time. A separate region proposal phase is not required by the end-to-end system YOLO, which can detect objects in a single stage [13]. It is consequently far quicker than other object detection techniques, which call for many passes over an image or video.

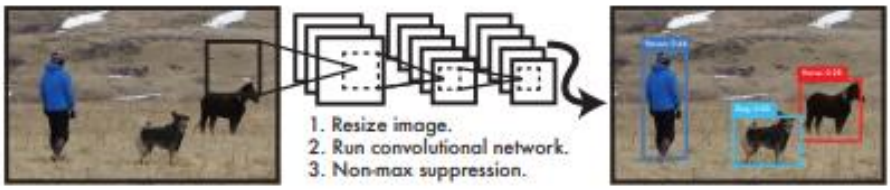

Figure 1: The YOLO Detection System. Processing images with YOLO is simple and straightforward. YOLO (1) resizes the input image to 448 × 448, (2) runs a single convolutional network on the image, and (3) thresholds the resulting detections by the model's confidence [13]

There are 24 convolutional layers in the YOLO detection network, followed by 2 fully linked layers. The feature space from previous layers is reduced by alternating 1x1 convolutional layers. YOLO is pre-trained on the ImageNet classification and the convolutional layers at half the resolution (224X224 input picture), and then double the resolution for detection.

YOLO is used in many different industries, including robotics, self-driving automobiles, and surveillance. Due to its real-time capabilities, it is especially beneficial in applications that call for quick object recognition, including following people or vehicles in congested areas. YOLO has also been used to track and find endangered species in their native habitats as part of wildlife conservation efforts. YOLO has also been modified for use in medical imaging, where it has demonstrated potential in identifying cancers and other anomalies in medical pictures.

SSD

Single Shot Detection (SSD) is a state-of-the-art object detection algorithm introduced by Wei Liu, Dragomir Anguelov, and others in 2016 [14]. It is a single-stage object detection framework that is able to simultaneously perform object localization and classification in real-time on a single input image.

The SSD framework uses a fully convolutional neural network (CNN) that is trained to directly predict bounding boxes and class scores for objects of different sizes and aspect ratios in the input image. The network is composed of a base network that extracts features from the input image, followed by several convolutional layers that produce predictions for object bounding boxes and class scores.

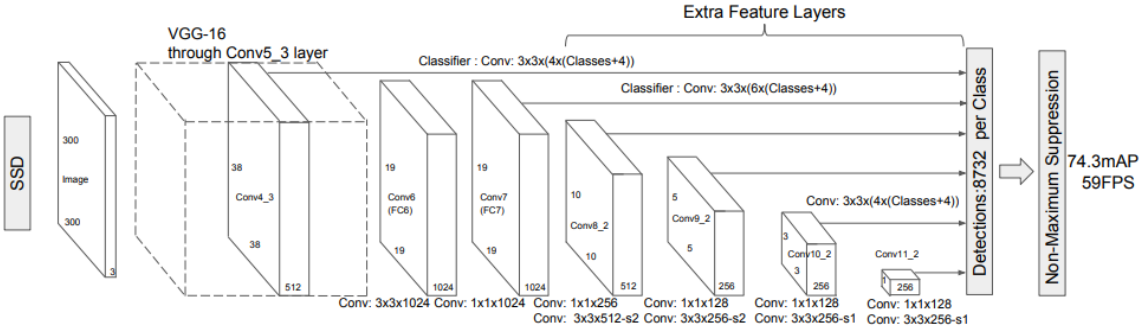

Figure 2: The SSD model adds several feature layers to the end of a base network, which predicts the offsets to default boxes of different scales and aspect ratios and their associated confidences [14].

The accuracy and efficiency of the SSD structure are its key benefits. SSD provides object identification in a single forward pass through the network, which makes it quick and effective compared to conventional approaches that involve numerous steps and complex calculations. SSD has a wide range of applications in various fields from autonomous vehicles to robotics.

Region-based object detection

Region-based object detection is a type of object detection algorithm that was first introduced by Girshick et al. in 2014. It is a two-stage technique to object detection that first locates things using region proposals, then classifies the objects found there [15]. A selective search technique is used to generate the region suggestions and suggests regions that are likely to contain objects. The features needed for object classification and localization are then extracted from the proposed regions using a convolutional neural network (CNN).

Faster R-CNN is a popular region-based object detection algorithm that was introduced by Ren et al. in 2015. By adopting a region proposal network (RPN), a fully convolutional network that forecasts item bounding boxes and objectness scores, for the selective search algorithm in earlier methods, faster RCNN outperformed them [16]. Faster R-CNN is well known for its accuracy and speed, and it has been extensively employed in a variety of applications including

face, vehicle, and pedestrian recognition. Faster R-CNN has the advantage of being able to handle objects with various scales and aspect ratios.

Region-based object detection has various variants, including Faster R-CNN, R-FCN, and Mask R-CNN [3]. Similar to Faster R-CNN, R-FCN (Region-based Fully Convolutional Networks) detects objects using position-sensitive feature maps. By including a branch to predict object masks in addition to bounding boxes and class labels, Mask R-CNN expands Faster R-CNN. These changes give object detection jobs more flexibility and precision.

Current research on aerial object detection:

Clustered object detection:

Research by Yang, F., Fan, H and ther team proposed a novel approach called Clustered Detection (ClusDet) for detecting objects in aerial images. The small size of the targets in pixels, as well as their sparsity and asymmetric distribution, present the key challenges in object detection for aerial photos. ClusDet proposes a network that combines object clustering and detection in an end-to-end framework to address both of these issues. The suggested approach achieves good running time efficiency while significantly reducing the number of chips required for final object detection. The detection of small items is improved by the cluster-based scale estimation over the single-object-based methods that were previously utilized. The last DetecNet is used to locate clustered areas and implicitly models prior context data to improve detection accuracy. On three well-known aerial image datasets, ClusDet is evaluated, and in comparison, to cutting-edge detectors, it shows promising results. The suggested approach takes into account the scale and target sparsity issues in aerial picture object detection and makes use of the clustering of items in certain regions to increase effectiveness and accuracy.

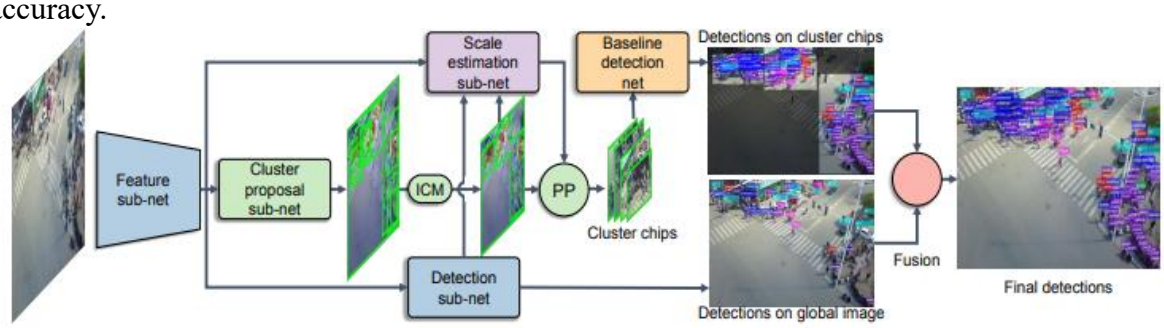

Figure 3: Clustered object Detection (ClusDet) network[18]

The cluster proposal subnet (CPNet), scale estimation subnet (ScaleNet), and specialized detection network (DetecNet) are the three main parts of the ClusDet network. The cluster regions are predicted using CPNet. To determine the object scale in the clusters, ScaleNet is used. On cluster chips, DetecNet carries out detection. The final detections are produced by combining global image and cluster chip detections [18].

The proposed method is compared to state-of-the-art detectors such as Faster RCNN and RetinaNet on three public datasets: VisDrone, UAVDT, and DOTA. Experimental results show that ClusDet outperforms the state-of-the-art methods by a large margin over various backbone settings on the VisDrone dataset. On the UAVDT dataset, applying EIP on test data does not improve the performance, but ClusDet is superior to other methods due to its different image crop operations. On the DOTA dataset, ClusDet achieves similar performance with state-of-the-art methods but processes dramatically fewer image chips.

3D Object Detection

The capacity of drones to operate in three dimensions opens up a world of possibilities for interpreting three-dimensional scenes. However, the object detection capabilities of existing drones are solely restricted to the 2D image space and the associated 2D boxes, which lack any 3D physical significance.

Yue Hu and Shaoheng Fang has done the research and proposed a 3D object identification method for drone-shot photos. It is a dual-view, aerial monocular 3D object detection system termed DVDET However, creating such a system confronts three major obstacles: a well-designed detection algorithm; a well-organized dataset; and an appropriate 3D representation for drones. The authors suggest a comprehensive dataset that combines simulation and real-world data with 2D and 3D annotations in order to address these issues. The suggested system introduces a new 3D representation method that is appropriate for drones to handle the deformation issue of aerial view variation and distant imaging in 3D detection [19].

There are three methods of monocular 3D object detection: direct, depth-based, and grid-based. Current approaches, which rely on depth learning, are created for driving scenarios. A unique geo-deformable transformation is developed to address the variation and deformation of drone views, utilizing geometric priors and learning capabilities for exact transformation. View transformation is a common task that needs to be done in this model. There are two types of view transformation: geometric and parametric. Geometrics is stable but unable to estimate unseen areas, while parametric can adjust to deformation but hard to fit diverse views.

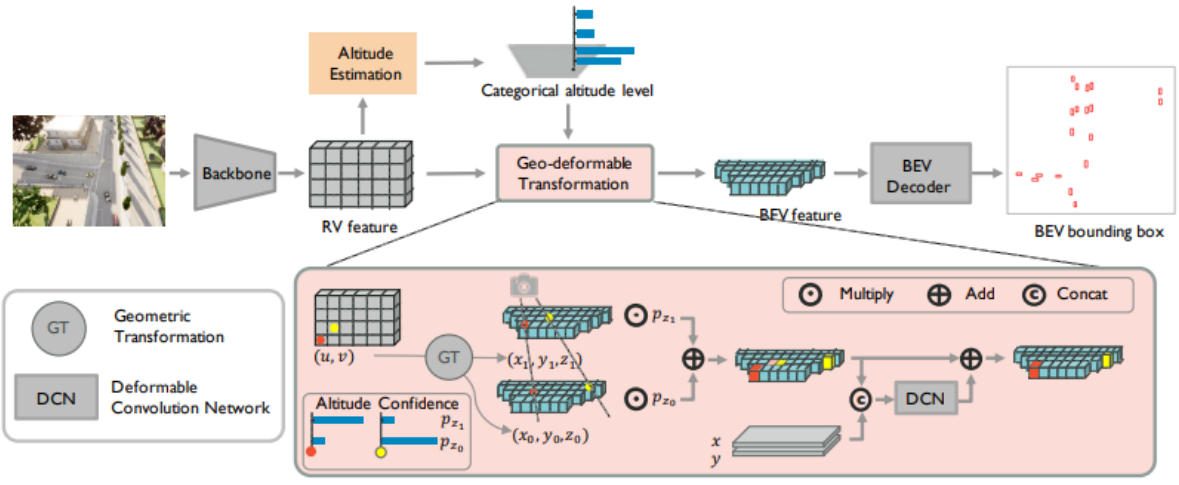

Figure 4. The overall framework of aerial monocular 3D object detection[19]

The results of the model show that DVDET outperforms baseline methods on both the simulated and real-world datasets, demonstrating the effectiveness of 3D object detection from an aerial perspective. Additionally, the model pre-trained on the simulation dataset shows improved performance on the real-world dataset and achieves leading performance on the KITTI benchmark [19].

Moving object detection

Moving object detection is an essential task in aerial object detection, which aims to identify objects that are moving in the scene. The complexity and dynamic nature of the aerial environment, which includes camera motion, dense backgrounds, and small object sizes, make this work difficult.[20]

Moving object recognition is currently being researched using a variety of strategies, including as optical flow-based methods, background subtraction methods, deep learning methods, and object tracking. While background removal techniques identify moving objects by eliminating the backdrop from the current frame, optical flow-based techniques estimate the motion vector between successive frames. Convolutional neural networks are used in deep learning-based techniques to extract features and recognize moving objects. To find moving objects, object tracking techniques monitor the same object over multiple frames.

Recent studies have concentrated on creating more reliable and precise algorithms for Aerial object detection. For larger training sets and better model performance, these approaches frequently include data augmentation techniques. In order to increase the precision of moving object recognition, some techniques also make use of temporal information by integrating numerous frames.

The accuracy and robustness of the approaches utilized in moving object detection for aerial object detection have been the subject of recent study. These developments have significant ramifications for a number of applications, including aerial surveillance, traffic monitoring, and disaster response.

Conclusion:

Aerial object detection is a vital field with a lot of application potential in surveillance, agriculture, mapping, and urban planning, among other fields. Researchers can precisely and reliably identify and categorize items in aerial pictures and videos with the help of deep learning systems. Due to the accessibility of massive datasets like DOTA and VisDrone, researchers have access to a lot of annotated data which helps in the development of better algorithms and models.

Aerial object recognition will become increasingly important in addressing practical issues as computer vision and machine learning technologies develop. The creation of fresh algorithms that can recognize things more precisely even in difficult situations like occlusion, lighting, and background clutter is something we can anticipate. This will improve public safety, boost efficiency, and assist industries in making better-informed decisions.

Aerial object detectoin is becoming increasingly important, and current developments in computer vision and machine learning make this field of study an interesting one with limitless potential. We may anticipate even more remarkable discoveries in the future that will change industries and improve the safety and effectiveness of our lives with ongoing research and investment.

References:

[1] Wei, Z., Liang, D., Zhang, D., Zhang, L., Geng, Q., Wei, M., & Zhou, H. (2022). Learning calibrated guidance for object detection in aerial images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, 2721-2733.

[2] Wu, Z., Suresh, K., Narayanan, P., Xu, H., Kwon, H., & Wang, Z. (2019). Delving into robust object detection from unmanned aerial vehicles: A deep nuisance disentanglement approach. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1201-1210).

[3] Ramachandran, A., & Sangaiah, A. K. (2021). A review on object detection in unmanned aerial vehicle surveillance. International Journal of Cognitive Computing in Engineering, 2, 215-228.

[4] Xia, G. S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., ... & Zhang, L. (2018). DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3974-3983).

[5] Bozcan, I., & Kayacan, E. (2020, May). Au-air: A multi-modal unmanned aerial vehicle dataset for low-altitude traffic surveillance. In 2020 IEEE International Conference on Robotics and Automation (ICRA) (pp. 8504-8510). IEEE.

[6] C. Kyrkou, G. Plastiras, T. Theocharides, S. I. Venieris and C. Bouganis, "DroNet: Efficient convolutional neural network detector for real-time UAV applications," 2018 Design, Automation & Test in Europe Conference & Exhibition (DATE), Dresden, Germany, 2018, pp. 967-972, doi: 10.23919/DATE.2018.8342149

[7] Boonpook W, Tan Y, Ye Y, Torteeka P, Torsri K, Dong S. A Deep Learning Approach on Building Detection from Unmanned Aerial Vehicle-Based Images in Riverbank Monitoring. Sensors. 2018; 18(11):3921.

[8] Kamran, F., Shahzad, M., & Shafait, F. (2018, December). Automated military vehicle detection from low-altitude aerial images. In 2018 Digital Image Computing: Techniques and Applications (DICTA) (pp. 1-8). IEEE.

[9]Lygouras E, Santavas N, Taitzoglou A, Tarchanidis K, Mitropoulos A, Gasteratos A. Unsupervised Human Detection with an Embedded Vision System on a Fully Autonomous UAV for Search and Rescue Operations. Sensors. 2019; 19(16):3542. https://doi.org/10.3390/s19163542

[10] Koo, V., Chan, Y. K., Vetharatnam, G., Chua, M. Y., Lim, C. H., Lim, C. S., ... & Sew, B. C. (2012). A new unmanned aerial vehicle synthetic aperture radar for environmental monitoring. Progress In Electromagnetics Research, 122, 245-268.

[11] Du, D., Qi, Y., Yu, H., Yang, Y., Duan, K., Li, G., ... & Tian, Q. (2018). The unmanned aerial vehicle benchmark: Object detection and tracking. In Proceedings of the European conference on computer vision (ECCV) (pp. 370-386).

[12] Cao, Y., He, Z., Wang, L., Wang, W., Yuan, Y., Zhang, D., ... & Liu, M. (2021). VisDrone-DET2021: The vision meets drone object detection challenge results. In Proceedings of the IEEE/CVF International conference on computer vision (pp. 2847-2854).

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

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

[15] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2015). Region-based convolutional networks for accurate object detection and segmentation. IEEE transactions on pattern analysis and machine intelligence, 38(1), 142-158.

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

[17] Ding, J., Xue, N., Xia, G. S., Bai, X., Yang, W., Yang, M. Y., ... & Zhang, L. (2021). Object detection in aerial images: A large-scale benchmark and challenges. IEEE transactions on pattern analysis and machine intelligence, 44(11), 7778-7796.

[18] Yang, F., Fan, H., Chu, P., Blasch, E., & Ling, H. (2019). Clustered object detection in aerial images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 8311-8320).

[19] Hu, Y., Fang, S., Xie, W., & Chen, S. (2023). Aerial monocular 3d object detection. IEEE Robotics and Automation Letters, 8(4), 1959-1966.

[20] Shen, H., Li, S., Zhu, C., Chang, H., & Zhang, J. (2013). Moving object detection in aerial video based on spatiotemporal saliency. Chinese Journal of Aeronautics, 26(5), 1211-1217.