A Survey of Deep Learning Techniques and Computer Vision in Robotic and Drone with Applications

Maysoon Khazaal Abbas Maaroof[1,2*] and *Med* Salim BOUHLEL[3]

[1]National School of Electronics and Telecommunications, Information Technology, University of Sfax, Tunisia
[2]M. SC Information Technology, University of Babylon, Babil, Iraq
[3]Professor in Higher Institute of Biotechnologies of Sfax, ISBS, University of Sfax, Tunisia

Abstract. The methods of deep learning have lately demonstrated outstanding outcomes of robotic objects such as imagination, localization and striping. Its exceptional abilities in order to learn idealizations from complicated data gathered in the real world ambiance make it perfect for a high range of independent applications of robot. Simultaneously, unmanned aerial vehicles are becoming more used for a variety of civilian stints ranging from security, superintending, and disaster relief, extraditing of package and repository arrangement. A thorough exegesis one of the primary deep learning techniques is also supplied. A set of the main difficulties in using deep learning with UAV-based solutions. Even still, autonomous navigation remains a challenge where computer vision technologies can shine. As a result, development the forecast made by the network and the ground-truth attention distribution, increased the use of vision systems and algorithms, been a major focus of studies conducted recently. An organized mapping investigation is conducted to gain a broad perception of subject. Some studies provide a thorough examination of addressing computer vision in relation to the following independent unmanned aerial vehicles vision establish chores such as navigation, control, back trace and sense.

1 Introduction

Deep learning methodologies the research community is very interested in the recent breakthroughs in solving more complicated perception problems through learning from raw data from sensors. A modern model was not deep learning, convolutional neural networks (CNNs), a computational technique that was established in the 1980s but trained effectively in the 1990s, gained popularity in the 2000s when they demonstrated strong performance in tasks involving the recognition of visual objects. At the time, performance was sometimes worse than that of conventional computer vision techniques because to tiny datasets and

* Corresponding author: basic.maysoon.maroof@uobabylon.edu.iq

underpowered processors. The popularity of CNNs with several layers was significantly influenced and the release of the image net dataset, which comprised 1.2 million photos categorized into 1000 categories (109 to 1010 connections and 107 to 109 parameters). These deep models excel at science application distinguishing, signal analytical, and natural language processing in addition to computer vision applications [1].

The subject of study and research was categorized using the taxonomy put forth in Aerostack [2], an aerial robotics architecture compatible with standard parts. The purpose of denoting to architecture, which is shown in figure 1, is to better comprehend the makeup of the parts of the under consideration aerial robotic systems. This taxonomy can be used to find areas of a system where deep learning has not yet been implemented. The following systems and interfaces are used to categorize the parts of an unmanned aerial robotic system, according to Aerostack:

- Hardware interfaces: interfaces with sensors and actuators are included in this category.
- Motor system: motion controllers are its components., which exemplarily extradites commands specifying desired real values for a variable. Actuators get these necessary Low-level commands, which are derived from values.
- Characteristic extraction: The system that handles idealizing and traits from data sensors is called peculiarity extraction. Deep learning techniques come with feature extraction systems built-in because of the processes used to learn data representations.
- Situational awareness system: to understand the environment, this system consists of parts that aggregate sensor input into state variables pertaining to the robot's perception of its surroundings. One element of a situational awareness system is the SLAM algorithm. The manager The framework creates intricate behavior sequences based on high-level symbolic actions that are received.

Planning systems: these systems facilitate worldwide problem-solving for intricate activities like mission and path planning.

- The supervision system consists of parts that simulate self-consciousness in order to oversee other integrated systems. An example of this kind of component is an algorithm that monitors the robot's progress toward its objective and reacts with recovery actions to issues (unexpected barriers, malfunctions, et).

- Communication system: The elements of this system are responsible for facilitating efficient communication between robots and/or people.

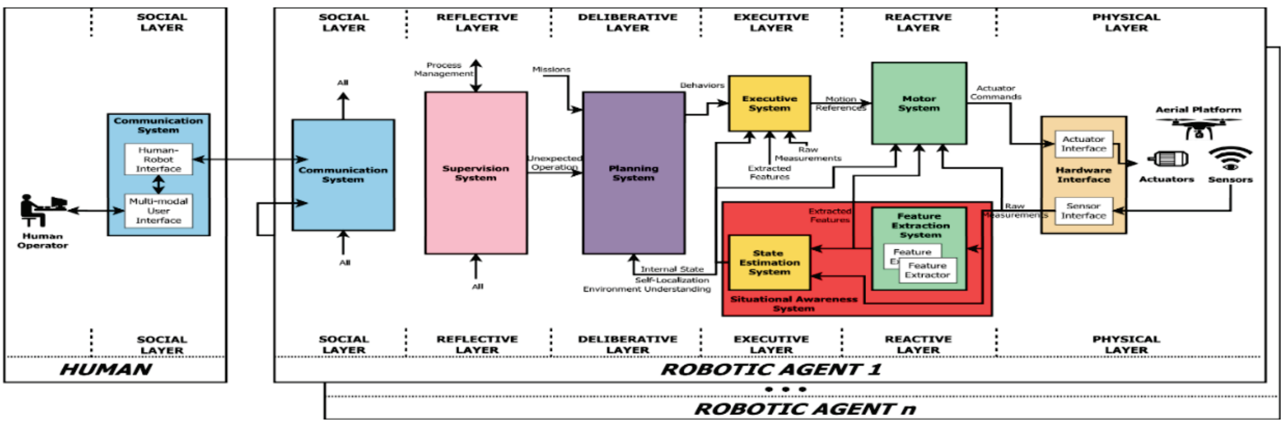

Fig.1. Aerostack system architecture, [2].

2 Deep Learning

2.1 The Application of Deep Learning to Machine Learning

Developed an object detection solution based on deep learning that can be run locally on an autonomous drone. The goal of the drone in the International Aerial Robotics Competition is to correctly heard a group of ground robots (IARC). The main challenge is figuring out how to reliably detect these ground robots using the drone's cameras. With limited computational power, all of this must be done in at the same time, in four distinct directions, in real time. The recommended course of action is to gather a sizable dataset of detection examples from photographs taken during the IARC 2016 competition, then utilize the dataset to train the quickest and most precise detection neural networks available, such is YOLO "You Only Look Once" and SSD "Single Shot Multibox Detection". The result is an assortment of 6100 detection images with labels for each of the three competing robot types as well as a number of detection networks that can operate in real-time and have differing degrees of accuracy on the drone [3], as shown in figure 2.

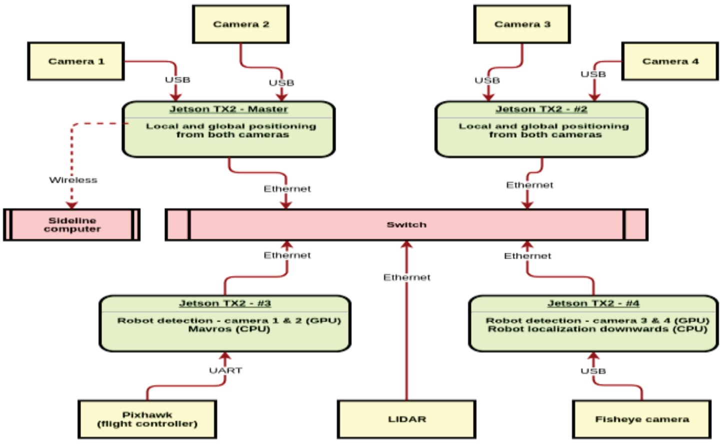

Fig. 2. The different computational components and sensors onboard.

Unmanned Aerial Vehicles (UAVs) provided a flexible means of landing in unique circumstances, like those where There was no global positioning system accessible. When more than one UAV touches down on the same surface, there is a possibility that they will collide, without communicating. With the help of this effort, Vision-based self-landed vehicle is achieved, and deep learning is used to prevent collisions when landing. There are two stages for landing UAVs: I and II. Level II UAVs will use the deep learning approach. The automatic collision avoidance while landing system of Level II UAVs approach is depicted in Figure 3. Level II UAV stay in place until the Level I UAV completes its landing if It recognizes when a Level I UAV is nearby. If not, it moves forward to the landing area and keeps following the repeated pictures [4].

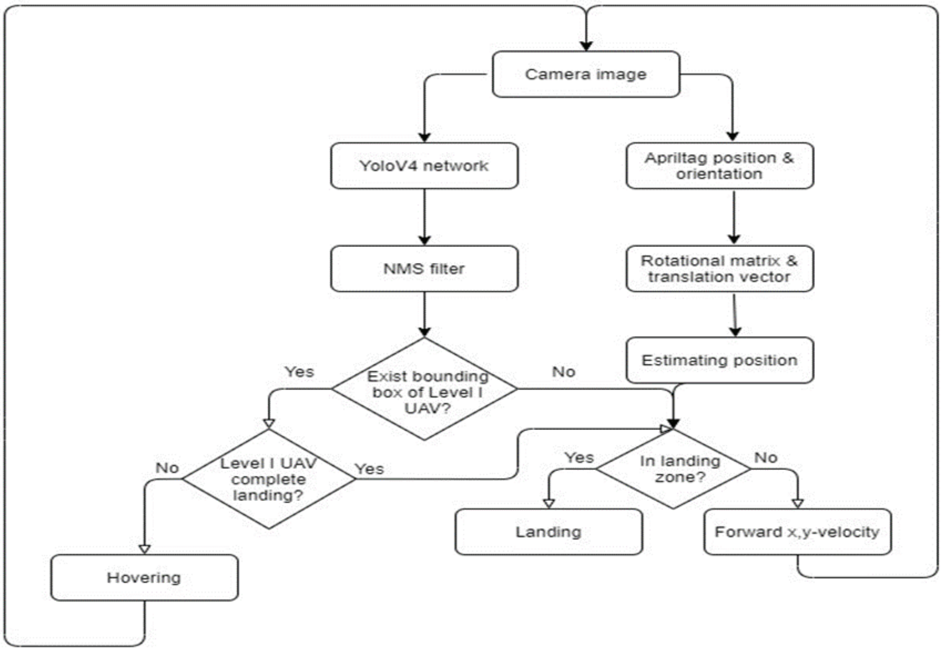

Fig. 3. Autonomous landing collision avoidance procedure proposed, [4]

Research on airborne sensing has been dominated by introduced unmanned airborne vehicle (UAV)-based applications. Nevertheless, no evaluation of the literature integrating. making use of "deep learning" and "UAV remote sensing" Providing a thorough overview of the principles of Deep Learning as they relate to UAV-based photography was the aim of this research. discussed the regression and classification methods applied in recent UAV-acquired data applications. Talk about the potential of deep learning for processing tasks requiring picture data from UAVs and how it produces promising outcomes. Finally, the project's future perspectives are discussed, including prominent routes to be investigated in the realm of remote sensing with UAVs. As this update includes a method for introducing, commenting on, and summarizing the cutting edge of DNNs algorithms for UAV-based image applications in different remote sensing subfields, such as environmental, urban, and agricultural [5], as shown in figure 4.

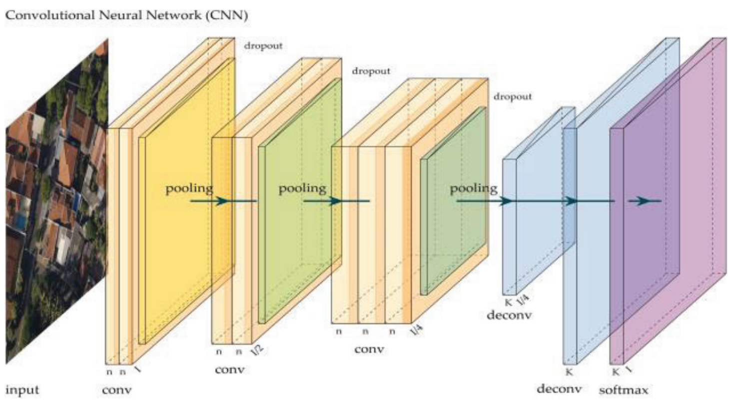

Fig. 4. layers of convolution and deconvolution in a CNN architecture, [5].

Showed unmanned aerial vehicles (UAV), which are versatile, effective, and affordable, are crucial in the security industry. Urban intelligent traffic monitoring can be made much more effective by having a UAV's capacity to detect, track, and position cars. Deep SORT, a deep learning approach for tracking and detecting vehicles automatically, urban environments, has been designed in this study by combining the target detection network-based detection-based multi target tracking technique, or YOLO v4. An interactive particle filter prototype multimodal was utilized to achieve the moving target's state estimate to resolve the in the nonlinear system issue placing a vehicle target with a UAV. Equations for the system's measurement and condition were produced. The simulation's outcomes show that the recommended algorithm is capable of automatically detecting and tracking cars in urban settings. Additionally, the UAV's performance in maneuvering target positioning is considerably improved by the interactive particle filter technique multimodal. which has good engineering application value [6], as shown in figure 5.

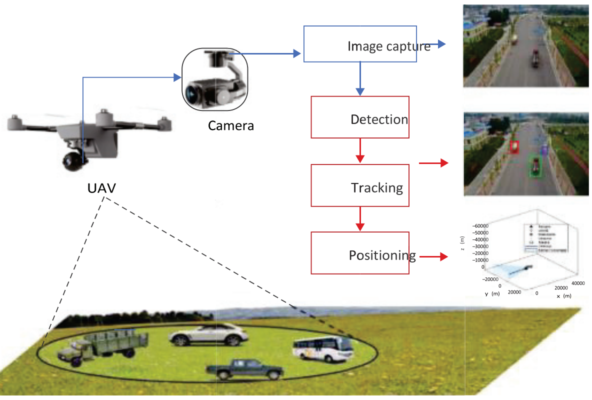

Fig. 5. Structure of an algorithm system [6].

Proposed a novel method depend upon the using a deep learning technique to recognize and detect two different kinds of birds and drones. The four primary steps are shown in Figure 6. First, the data must be appropriately prepared for use as the input in the suggested architecture. In order to identify and detect two different types of drones, the second stage involves implementing the network training phase [7].

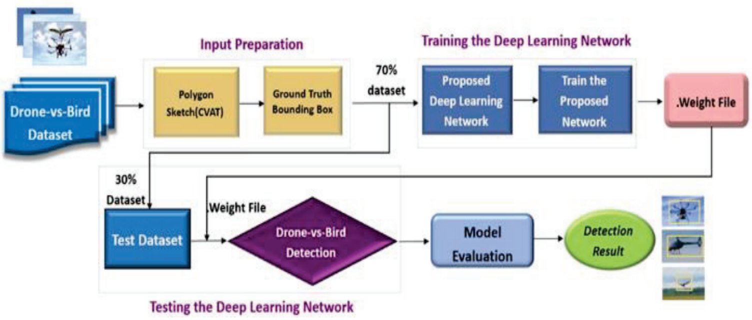

Fig. 6. Diagram illustrating the suggested utilization of convolutional neural networks in identification and classification, [7].

Provided an in-depth examination of recent processes for the identification and classification of traffic lights, roads, traffic signs, and three-dimensional objects. Tasks of object detection in autonomous driving using deep learning are carried out if an image (s) or point cloud is used as the input. Additionally, it offers comparison data to compare perspectives and stimulate more research [8], as shown in figure 7.

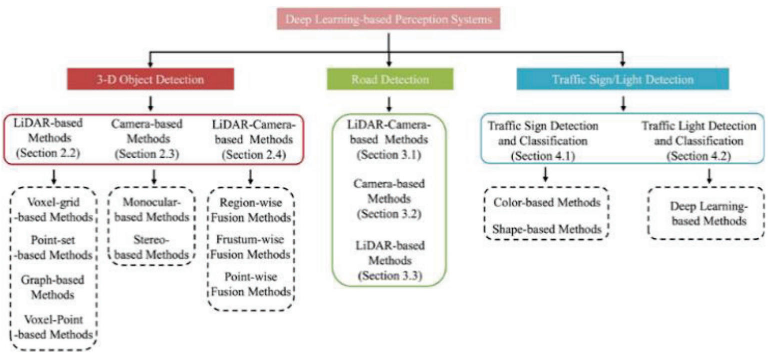

Fig. 7. A summary of techniques for deep learning-based perception systems, [8].

Applied the aalgorithms of machine learning into three groups listed below: Learning methods include Reward learning, unsupervised learning, and supervision:

- Algorithms are given a dataset, which is an assortment of characteristics, in supervised learning. Marks or desired values for each sample are additionally supplied. This mapping of characteristics to target value labels contains the knowledge.
- The objective Unsupervised learning seeks to identify significant representations and elucidate important properties of the data. In this instance, learning from the data does not require labels or target values.
- An AI agent interacting with a virtual or real-world environment is the basis of reinforcement learning approaches. By means of this link, the educational program and the social engagement provide feedback that can be used to enhance performance in the activity being learnt. In order to facilitate the creation of complex concepts from smaller ones, deep learning architecture [9].

2.2. Extraction of Features Using Deep Learning

Feature extraction models were considered main The goal is to extract typical qualities from the unprocessed measurements that UAV sensors offer.

1. Image Sensors are used applications leveraging diverse imaging technologies have used deep learning algorithms image. A number of layered convolution and pooling layers make up CNN models. By applying learnt filters to the input image, convolution layers' extract features from data, whereas pooling layers reduce dimensionality across prior convolution layers. systems for extracting features from Robotics has mostly used CNN models for scene classification and object recognition. Recent developments in object identification have merged bounding box regression-based item detection techniques with object classification tools inside the same CNN model. For object recognition, unsupervised feature learning eliminates the requirement for training data with manual labeling, which can be laborious and inefficient money-consuming to acquire. In order to get representative picture characteristics, recent developments in scene classification have concentrated on learning, that learnt features derived from pertained even in domains that were greatly unlike from those where they were taught, such as aerial image classification, CNN models could generalize appropriately. An image of a forest trail as input was classify into three classes using a 10-layered CNN on a Parrot AR. Drone quadrotor to classify scenes. [1]

https://doi.org/10.1051/bioconf/20249700008

Convolutional networks are used to extract features from each area proposal. The proposal's sub image is first distorted to fit the necessary CNN input size before being fed into the convolution neural network. Subsequently, the extracted characteristics are fed into support vector machines (SVMs), which produce the final classification. The R-CNN steps are shown in Figure 8. R-CNN has a number of drawbacks, including its training being extremely sluggish due to its multi-stage pipeline training that starts with training a Conv Net for feature extraction, following that, training SVMs for classification is done, and lastly, a bounding box regression is performed. Slow object detection speed because each object suggestion in the test image has its features extracted during the test period. As seen in figure 8, feature extraction from each object proposal in each image required a large amount of storage space [10].

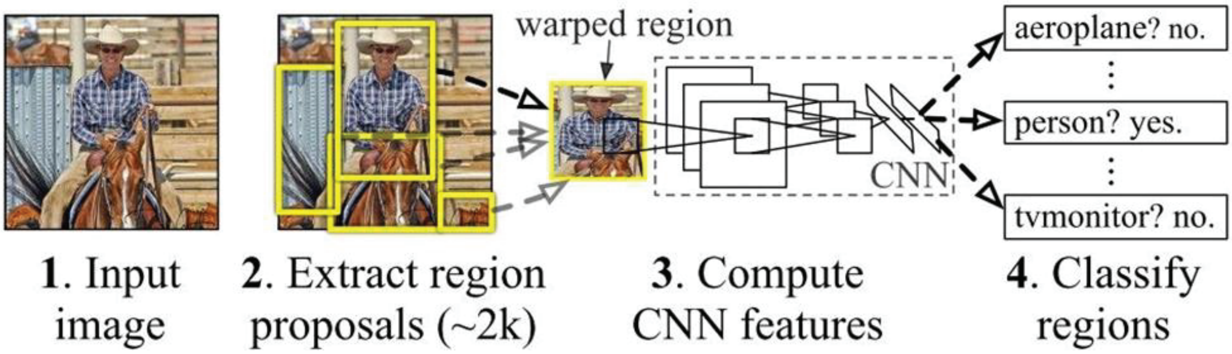

Fig. 8. R-CNN Stages, [10].

In order to automatically locate and recognize significant architectural flaws (such as efflorescence, spalling, cracking, and vandalism) using images, they suggested A neural network model that is convolutional. Because object localization, this model used class activation mapping, which was according to a pertained CNN classification model VGG-16. assessed the robustness and accuracy of the model's detection and localization flaws in building outside tiles for walls after recognizing its limitations in practical applications. Drones and mobile devices are used to use this model for localization and detection in real time. The findings demonstrate that combining deep learning with unmanned aerial vehicles (UAVs) can enhance the effectiveness of external wall fault identification and identify a wider range of flaws [11].

2. Including Extra Sensors The majority utilizing deep learning to reduce workload that has been published in the literature has been used to process data that has been captured by image sensors as a result of the aggregated results generated using CNN models. Contrarily, deep learning algorithms can be applied to a variety of applications and combined other sensors besides cameras.

The suggested utilizing deep convolutional architecture to extract the necessary characteristics to separate different the pixels in an aerial photo, is presented in this section at different phases. The architecture leverages the advantages of both significant architectures to create a fully connected, end-to-end CNN. Test the proposed model on two datasets for segmenting aerial images using UAVs, NITR Drone dataset1 with the Urban Drone Dataset (UDD), to demonstrate its effectiveness and accuracy. The outcomes are contrasted with those of other cutting-edge techniques, such as U-Net [12].

They tried the SFE method and used spectral band reduction prior to CNN processing in this subsection. Using this strategy, obtain a 2D image at the network's entry and quote the results. From the testing, were able to determine the class details for Salinas A and Indian Pines datasets, which ranged from 82.31 to 100.00% and 50.48 to 91.30%, respectively. For the first dataset, the VA is 89.18%, and for the second, it is 87.51%. In addition to increasing classification accuracy, the SFE reduced computing time; only needed to spend 2.22 x 102 and 5.39 x 102 seconds, respectively, for the sequential classification of Indian Pines and Salinas A. In fact, with x being the number of bands in the image, the band reduction method saves a significant amount of time T1 ≃ T0 x [13].

2.3 Deep learning methods for situational awareness and planning

Robots can grasp their own condition as well as the state of their surroundings through situational awareness activities. These include self-localization, mapping, and robot state estimate.

1. Preparation Planning a route for group search and rescue operations. The goal of this effort is to reduce deployment time overall by using a UAV to explore and map the finding a ground robot's path that can be taken through the environment. The mapping of the landscape and identify a path that can be followed, a CNN is suggested for terrain categorization. The classifier may be trained instantly with the landscape at the disaster site by using on-the-fly training rather than a pertained CNN. However, it takes around 15 minutes to train the model.
2. Situational awareness. The localization of an image across views is accomplished via deep learning. The trials were carried out using just ground-level photos, despite the fact that the work is promoted as a UAV localization solution. The process involves extracting nearest neighbor visual features (such as landmarks) from a collection of unprocessed image data and contrasting them with characteristics from an input query image.

Presented A situational analysis using a UAV awareness an apparatus known as Person-Action-Locator that can be helpful for activities like surveillance and search and rescue. The three primary parts of the PAL system are the Pixel2GPS converter, which uses image processing to estimate a person's GPS position, the PAL interface, which shows the detected activities and people on a map, and the Deep Learning component, which uses automatic face and activity recognition. In the field the test of the integration of all the components was successful. Additionally, the Deep Learning models in our lab have undergone extensive testing [14].

Proposed the Intelligent Network for Video Surveillance (INUS), an effective and adaptable, multifunctional platform system for situational awareness that can assist people in the analysis of video feeds and instantly notify them about items that have been recognized and monitored in a target region. In addition to supporting combining thermal and optical sensors, as well as utilizing cutting-edge UAVs and land-based technology include on-board processing, the system makes use of reliable together with effective machine learning, computer vision, and image processing approaches. To provide localized events, a module for heading, speed, and localization was also created. In addition to offering a better comprehension of the scene, the data retrieved based on the regional occurrences Computation 2021, can also benefit external systems by supplying rule-based schemes that allow the data to be connected with other properties. Their approach's main limitation is that for ground-level queries, they must at test time, estimate scale, orientation, and dominant depth [15].

3 Computer Vision

Presented A new geolocation framework in order to locate Unmanned aerial vehicles that solely rely on their internal camera in a satellite imagery map, the framework makes use of a large amount of satellite imagery in addition to well- reputable deep learning and computer vision techniques. Two distinct datasets that each reflect a different region of the world are used to test the system. Figure 9 illustrates how the on-board camera is the only source of precise UAV geolocation. It consists of the following essential elements, as indicated. S and rare are provided to the SIFT registration if the requirement of being the first or third sequence frame is satisfied. The SIFT registration then creates a homograph, which is utilized with respect to r and changes the position of the UAV [16], presented in figure 10.

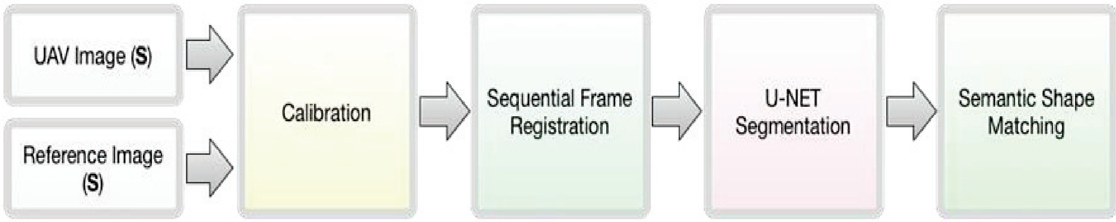

Fig.9. The primary elements of the proposed UAV geolocation framework, [16].

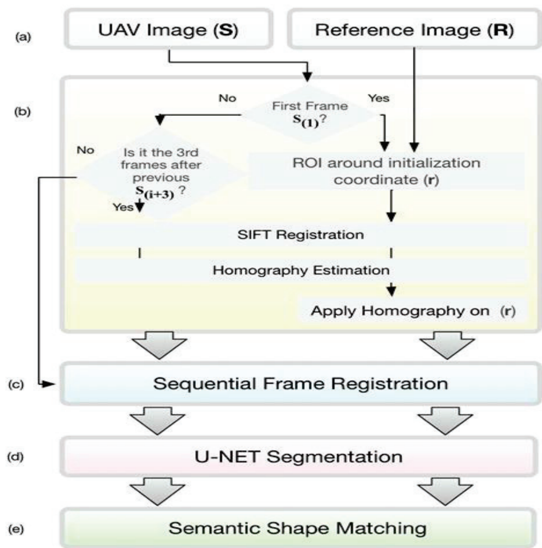

Fig. 10. Frame-by-frame sequence, [16]

Created a set of five test scenarios that make up an improved answer through iterative testing. The developed prototype uses Convolutional Neural Network (CNN)-based Computer Vision (CV) techniques to enhance GPS. These algorithms operate on with a Pi NoIR Camera-equipped Raspberry Pi 3, which does not have an IR filter. The trials were carried out using the SSDLite-MobileNet-V2 and Single Shot Detector (SSD) MobileNet-V2 devices [17].

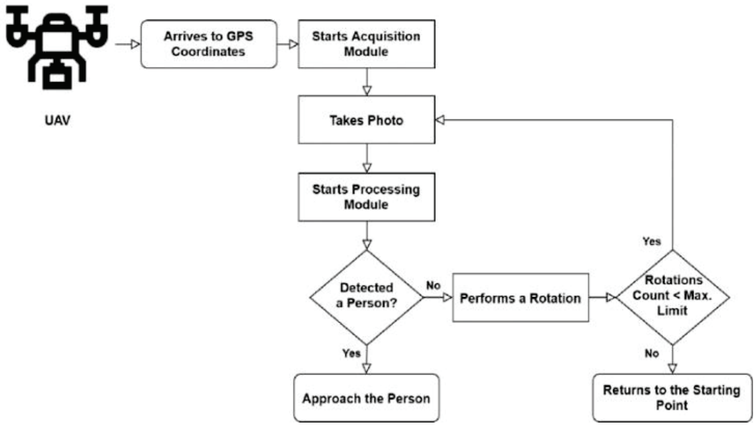

Fig. 11. The solution's progression from the beginning to the next stage is depicted in the architecture proposal block diagram, [17].

In the autonomous functionalities, computer vision is critical. While the primary functions of unmanned aerial vehicles technologies for instance, localization and mapping, target detection and tracking, navigation control, landing control, and path planning are already developed and extensively studied, factors like crowd safety, avoiding flying over people, and emergency landing capabilities are typically considered after the fact when designing these systems independent unmanned aerial vehicles platforms for unstructured environments [18].

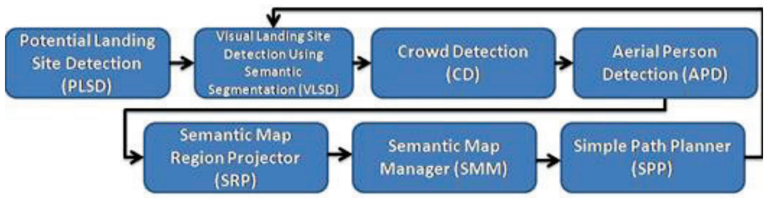

Fig. 12. Pipeline for UAV landing safety, [18].

Presented a novel cooperative computer vision-based method for monitoring targets depending on the image's precise area of interest Any object can be tracked using the suggested method, regardless of its size, shape, color, or pattern. The target must to remain in the line of sight and visible during tracking. Any target in the image can be tracked by the user, who can then construct a formation around it [19].

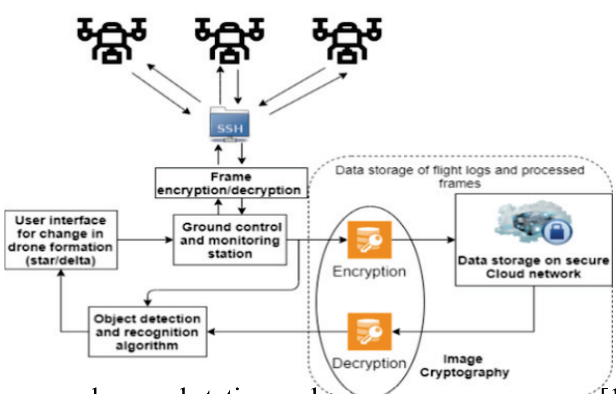

Fig. 13. Individual drone and ground station end-to-end communication, [19].

4 Robotic

Solved difficult issue facing the field of robotics research in order to tackle this problem, earlier researchers combined various camera systems with multiple sensors, including GPS receivers and inertial measurement units. While these techniques are effective in estimating the location of an unmanned aerial vehicle during landing, high detection accuracy requires a number of calibration steps. In addition, scenarios where drones function in a non-GPS signaled diverse environment must to be considered. To ensure that resolve these issues, created a monitoring algorithm that makes advantage of remote markers and just one camera sensor that detects visible light securely place a drone in a location where GPS is not available. In order to forecast the location of a marker using an observable light camera sensor on a drone, the system uses a convolutional neural network called light Dense YOLO to take out learned characteristics of an input picture rather than hand-crafted features. outcomes of an experiment demonstrate that the technique does much better than cutting-edge product trackers that employ neural networks with convolution, and those that do not regarding precision and processing duration. The entire flowchart of the suggested method is shown in Figure 14. Reduced the initial input picture captured by the drone with a 1280 x 720 pixels' resolution to 320 320 pixels in order in order to match the CNN light Dense YOLO input [20], as shown in figure (15).

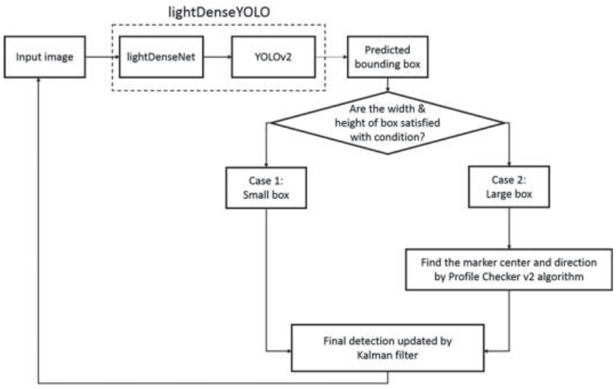

Fig. 14. Flowchart illustrating the technique for tracking markers at a significant distance.

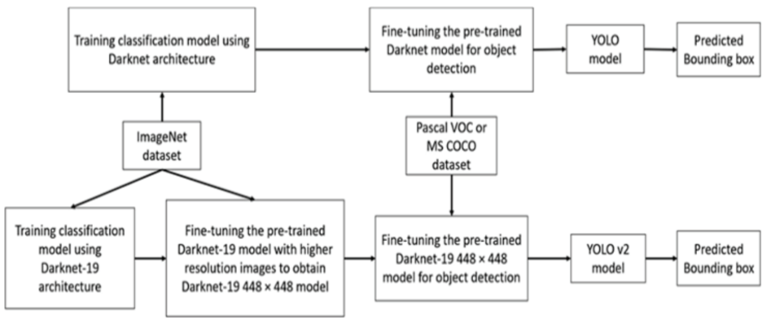

Fig. 15. Flowchart of YOLO and YOLO v2's overall object detection training.

Proposed an object tracking algorithm that utilizes data obtained from the visual system. The proposed solution was executed and evaluated using an unmanned quadrotor. The detailed presentation of the proposed vision-based object tracking technique was provided. The primary objective of this study was to conduct experimental tests on the proposed visual servoing approach. The algorithm has undergone rigorous testing in order to evaluate its performance across a range of scenarios and determine its level of quality. This will allow the operator to efficiently monitor the thing being tracked by the aerial robot.

The block diagram shown in figure 16 (a) represents all of the operations mentioned. The concept of regulating the position and movement of a drone can be illustrated as a block diagram based on the tracking mechanism depicted in figure 16 (b), [21].

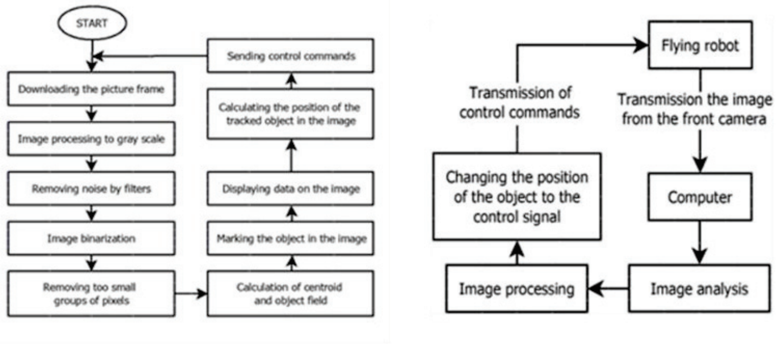

Fig. 16. (a) A diagram of algorithm's operations; (b) A diagram of the tracking process, [21].

Dedicated to cutting-edge study on the integration of deep understanding along with current robotic vision models to produce more potent remedies for practical robotics applications, like drones and driverless cars an application area that blends robotics, machine learning, and computer vision is called robotic vision knowledge. Seeing how deep neural networks improve the performance of several computer vision challenges. The main focus of autonomous vehicle 3D perception using LiDAR. In the panoptic segmentation task, each point's semantic label is predicted, and points that belong to the same object are given an instance id. The suggested hybrid method combines a conventional point cloud cluster algorithm with a semantic segmentation neural network as a post-processing step. In terms of performance, this hybrid approach beats all recently announced end-to-end neural network techniques. When it comes to drones or more versatile UAVs [22].

The machine learning techniques, it is possible to demonstrate the dependence of the saturation temperature on the robot positioning accuracy. The robot location inaccuracy was reduced from 0.07 mm to 0.02 mm by the deep learning system, demonstrating its effectiveness. This outcome emerged after the training process had gone through about 800 iterations. the just made technique shortens the time required for the machine learning process while increasing the precision and reliability of robot positioning. A unique neural network is used in the developed methodology for simulation with numerous input parameters. Without having access to the internal robot control system, the external compensation for the robot's positional drift in the target point proved effective [23].

5 Drone

An SUAV's ability to recognize and follow moving pedestrians was looked into. With consideration for the true size of targets, Limiting and anatomical filtering, and the decrease of false alarms come after frame removal. The next tracking step receives the center of the observed area. Multiple Kalman filtering modes are used in applying filtering to an interacting multiple model (IMM), which estimates the states between covariance matrices and vectors. Throughout the trials, a stationary drone over the road records 12 individuals and 1 car. The accuracy of state estimation is evaluated between the two- or three-mode IMM filter and the Kalman filter. Each target's location and velocity's root-mean-squared errors (RMSE) indicate high precision in identifying and monitoring the target's location, with an 96.5% on average for detection rates. The minimum average position and velocity root mean square error (RMSE) obtained with The two-input multimode filter approximately 0.8 and 0.59 meters per second, in that order. The block diagram for tracking moving objects in

 https://doi.org/10.1051/bioconf/20249700008

Figure 16 shows that a track is halted after a predefined amount of frames if it does not update its state with correct data. The track can update if there are no measurements in the validation region. If there aren't enough updates that include validated measurements, the terminated track is also believed to be false because the genuine target is expected to produce at least a specified number of verified measurements [24].

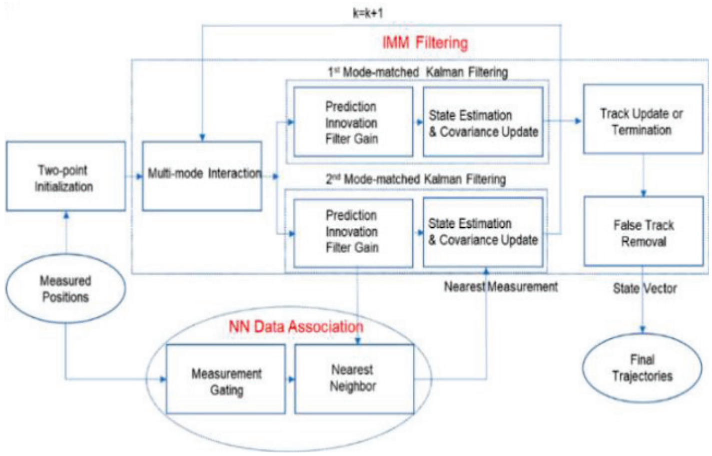

Fig. 17. Moving object tracking block diagram, [24].

.

Fig. 18. Drone with a camera attached.

A DJI Phantom 4 Advanced drone was employed to record mobile entities. The drone, equipped with a gimbal and camera, as depicted in Figure 18. The gimbal has the ability to adjust the camera's angle within a range of 120 degrees, from -90 degrees to 30 degrees. The camera was inclined at an angle of -30 degrees for the testing.

Both the generation of electricity and the decrease of drag increase the longevity of drones. The propagation (i.e., reflection, transmission, absorption) of electromagnetic waves can be changed by an artificial periodic structure in two or three dimensions. Thermal spraying is a method for applying coatings on small- to large-scale surfaces. The process of repeatedly colliding fully or partially molten particles of a substance exposed to varying process conditions produces the sprayed layer [25].

A stock, unmodified quadrotor drone was demonstrated to be capable of autonomous flight in hallway environments. When these structural elements of the hallway are detected, control planning algorithms can react appropriately, braking and turning at junctions or stopping before hitting dead ends like walls and doors [26].

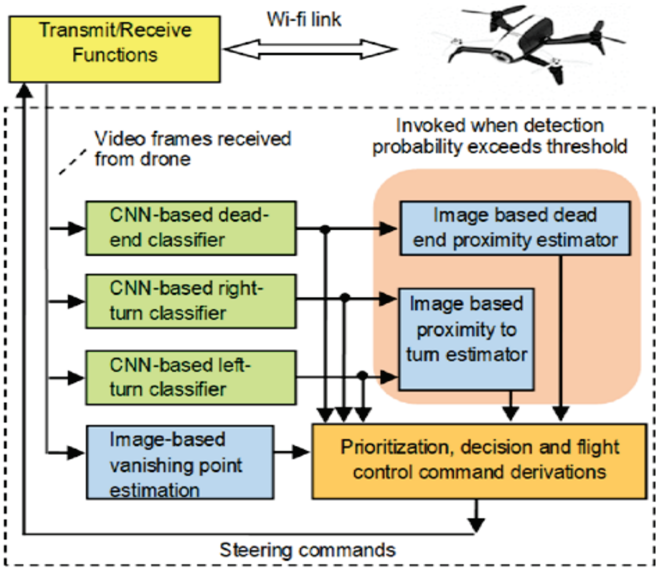

Fig. 19. The base station's system overview and main modules (dashed box), [26].

Proposed a simple method of solution for the drone game competition in NeurIPS (2019), with an emphasis on the perception tier Create a workable modular mechanism using three layers: planning, mapping, and creativity based on machine learning. Both a monocular depth estimate and a strong gate detection are necessary for target direction and collision avoidance. The gate positions in three dimensions are generated and updated on a map using estimated targets. Finally, for optimal flying, rule-based trajectory planning is used. The approach runs in real-time on a cutting-edge GPU and can traverse a variety of virtual racetracks while navigating challenging obstacles like fast speeds, perplexing gate placement, and asymmetrical forms. Upon the ultimate leaderboard, approach is ranked third, presented a detailed system design as well as additional experimental results [27], as shown in figure 20.

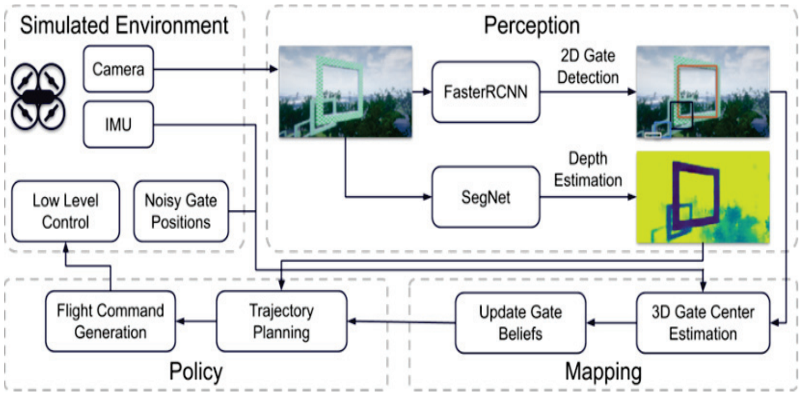

Fig. 20. Architecture of the System, [27].

Presented Drones are employed as auxiliary instruments in civil applications for conducting extensive aerial surveys of buildings, a challenging undertaking for surveyors because to the inaccessibility of certain areas, the time-consuming nature of the task, and the high costs associated with restricted resources and equipment. In order to resolve this problem, this study presents the concept of utilizing Unmanned Aerial Vehicles (UAVs) for the purpose of mapping. Moreover, when operating on an alternative flight trajectory, the unmanned aerial vehicle (UAV) will record and gather visual images. Subsequently, the drone-captured image underwent processing in the Airsoft Meta form programed, resulting in the creation of a three-dimensional model of the building. This technique will undergo multiple stages to analyze the efficacy of different picture capture methods in generating high-fidelity 3D mapping. The objective of this project is to ascertain the photogrammetry technique that can produce a 3D mapping of superior quality with precision and efficiency [28], as shown in figure 21.

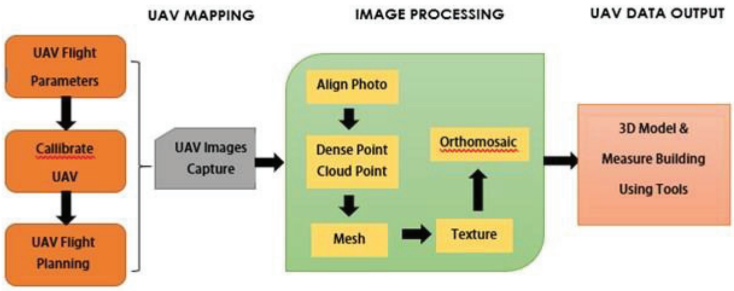

Fig. 21. Flowchart depicting UAV-assisted building generation, [28].

The introduced drone is a new technique for figuring out the Mine's stockpile volume. UAV data collecting technology is faster, safer, and requires less work. Thus, in order to estimate the stockpile amounts from a mine, UAV technology was deployed in an open pit in this study. Data from UAVs and GPS were gathered for this research to measure stockpile volumes of materials mined [29].

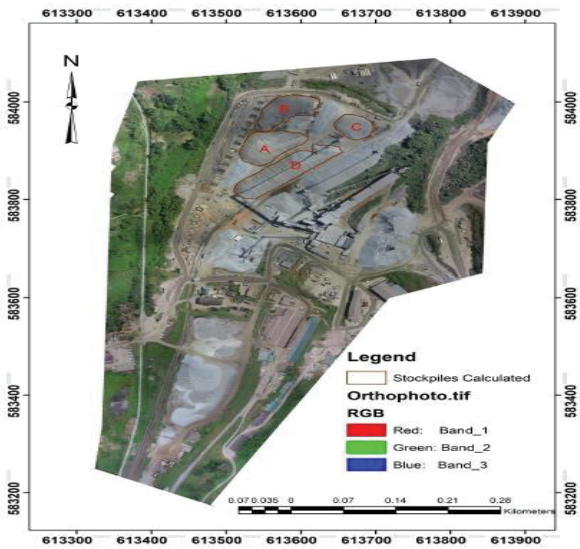

Fig. 22. The A, B, C, and D stocks are depicted in orthophotos, [29].

presented a modular drone racing navigation system that is vision-based and combines the advantages of database- and model-based approaches by using Modern planners and controllers provide low-level control commands, while a specialized convolutional neural network (CNN) intended for the module by perception generates elevated navigation directives. The perceptual module with CNN support has been taught to mimic a professional policy that, in contrast to the existing approach, only uses the image from the camera right now as the CNN input and, using previously calculated global trajectories, creates ground truth navigation commands automatically because of significant randomization and changed dataset aggregation [30].

Investigated a learning-based unmanned aerial vehicle autonomous surveillance system, wherein the UAV uses for semantic object detection, the YOLOv4-Tiny method, combines it using a Kalman filter and a 3D object position estimation technique to automatically detect, track, and follow an objective object without the help of humans improve perception performance. The source code has been made available to be used in the future by the scholarly community. A success plot's x-axis is comparable to that of precision plots represents the IoU value threshold, while the y-axis represents the proportion of frames that surpass this threshold. The area under the curve in Figure 23 indicates that one system fared better than the other [31].

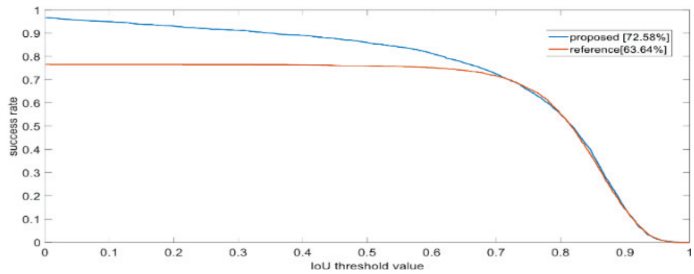

Fig. 23. The success plots of the comparing system and the suggested tracking mechanism, [31].

The goal of The goal of autonomous drone racing is to get as fast as possible through a series of waypoints. One of the main challenges Designing the time-optimal trajectory, which is often accomplished by presuming complete knowledge of the waypoints to pass, is the focus of this work. Beforehand. The resultant The choice is either not ideal because of, or extremely specialized for a single-track configuration. oversimplification of platform dynamics assumptions. Offer a new method for generating quadrotor trajectories in close to real time. By utilizing significant observations of gates and deep reinforcement learning, the method is able to calculate the near-time-optimal paths and adjust the path toward modifications in the environment. In the case of complex track arrangements, the method outperforms approaches based on trajectory optimization in terms of computational efficiency. The suggested technique is tested in the real world and in simulation on a range of racetracks, with a physical quadrotor attaining speeds of up to 60 km/h [32].

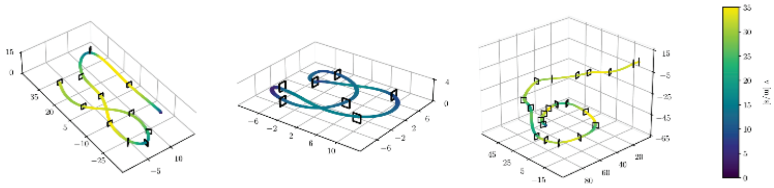

Fig. 24. Race tracks and approach trajectories were used for the baseline comparison, [32].

Presented and proposed a hybrid method that models quadrotors and their aerodynamic effects with previously unheard-of precision by combining machine learning and basic principles. Traditional methods for modeling or controller tuning are imprecise because first principles cannot account for such aerodynamic effects. In data-driven methods, black box modeling is used. such as neural networks, to capture aerodynamic effects; They find it difficult to robustly generalize to any set of flying conditions, though. The hybrid technique is outperformed by both learnt residual dynamics and first-principles blade element momentum theory. Tested on one of the biggest motion-capture systems in the world, which uses data from Quadrotor flying autonomously at up to 65 km/h. The resultant model outperforms previous models that demonstrate 50% lower prediction errors significant ability to generalize outside of the training set. It accurately captures aerodynamic thrust, torques, and parasitic effects [33].

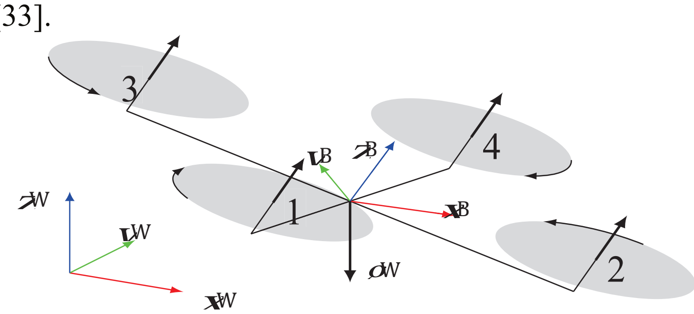

Fig. 25. Quadrotor model diagram displaying the world and body frames as well as the propeller numbering convention, [33].

After examining the relationship between parameters and performance, it was found that a controller's sensitivity to its parameters increased with the speed of the movement. Test Auto Tune thoroughly in the real world and in simulation. In trials, achieve up to 90% better trajectory completion results than current tuning techniques. Test the controllers that emerged from the AirSim Game of Drones contest, where achieve up to a 25% lap time advantage over the winner [34].

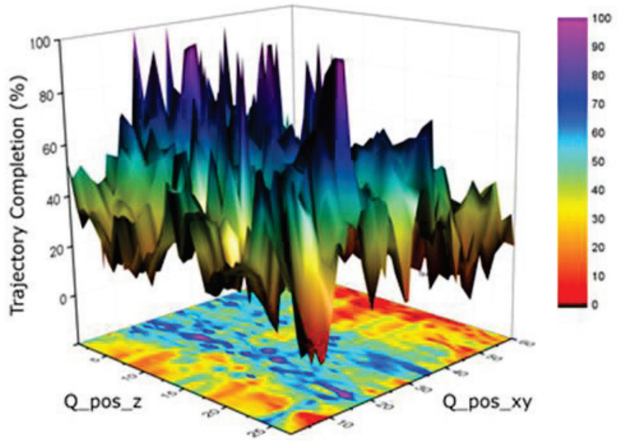

Fig. 26. The relationship between two model-predictive controller parameters and trajectory completeness (%) [34].

Presented a proposed technique for recognizing including Euclidean Distance Smart Geometric Analysis is invariant owing to picture perspective projection and may effectively and reliably identify the platform, orientation, rotation, or scaling The proposed algorithm, Computation of Hu moments, is based on human reasoning and adheres to the architecture depicted in Figure 26. Scaling, translation, or rotation have no effect on these. These descriptors are strong enough to handle marker distortions resulting from perspective projection of the image and/or from the UAV deviating from a vertical incline, in addition to platform zenithal position variations, which are all normal during the landing. A novel acknowledgment method that combines Hu moments with Discrimination based on color L*a*b* [35].

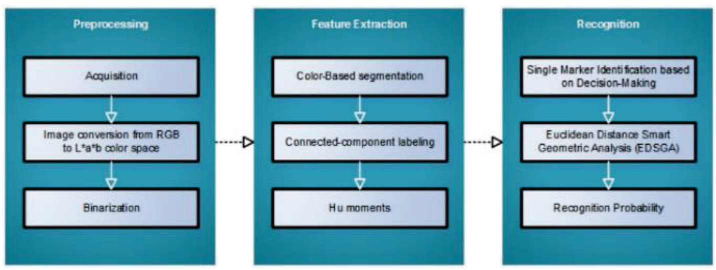

Fig. 27. The proposed image recognition algorithm's design and architecture are visible in the image, [35].

Investigated the performance of neural networks in the demanding position of racing autonomous drones using vision can be enhanced by neural networks that can imitate human eye gaze behavior and attentiveness. The hypothesis that gaze-dependent attention forecasting serve as an effective method for choosing and making decisions based on visual input in a drone racing simulator activity. The hypothesis to the test by training a visual attention prediction model with 18 The flight and eye gaze of human drone pilots trajectory data. Then, using imitation learning, develop an autonomous drone racing end-to-end controller based on vision employing this visual attention prediction model. As seen in Figure 27, The Visual Attention Prediction Network Building Block Architecture is intended to forecast dispersion of visual attention across pixels in an image. The individual frames are used to train the network and employs ResNet18 layers that have been pre-trained on ImageNet. It employs the ResNet-18 architecture's first four residual blocks, consisting of pooling and strides convolution [36].

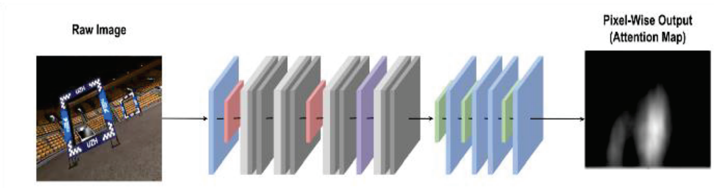

Fig. 28. ResNet-18-based attention-prediction network architecture, [36].

$$D_{K\,L}(A \| \hat{A}) = \sum_{x,y} A(x,y) log\left(\frac{A(x,y)}{\hat{A}(x,y)}\right) \quad (1)$$

where The picture coordinates are x and y. A is network's forecast, and The attention distribution in ground truth is denoted by A. Finally, a represents both the network's forecast and the ground-truth attention allocation, "Deep Drone Acrobatics" (DDA) architecture.

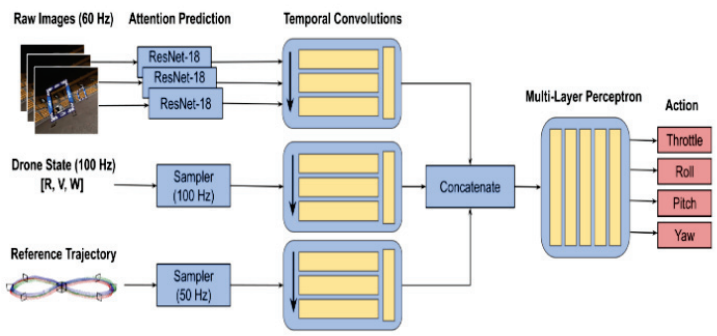

Fig. 29. Attention prediction-driven end-to-end management system architecture, [36].

6 Discussion and Conclusion

Power consumption of image sensors, images obtained from UAVs are now the most popular sort of information used by deep learning, according to the studies evaluated. Taking CNNs' This remarkable fact explains their popularity among deep learning algorithms are employed in UAV applications because of their exceptional capacity to extract valuable data from images. UAV technology, deep learning techniques, and their combination continue to pose a number of difficulties that impede this subject from developing more quickly.

Deep learning methods still have a lot of obstacles to overcome, starting with their own theoretical comprehension. Additionally, a lot of work is currently being invested into developing effective methods for performing unsupervised learning as the cost of gathering substantial volumes of unlabeled data is falling both technologically and economically. If this objective is accomplished, algorithms will be able to learn about the environment through simple observation, just like people can. These issues still lack a practical solution, which is a hurdle. UAV Challenges of Autonomy the size, weight, and power consumption of payloads are still constrained by the severe flight endurance restrictions of UAV aircraft. The capabilities necessary for autonomous operations are constrained by these limits, which are principally caused by the level of sensor and battery technology today. In the upcoming years, there will surely be developments in these fields preventing the creation of practical applications.

Given deep learning's outstanding capacity to extract data representations from unprocessed sensor data, this makes sense. The research community hasn't paid much attention to systems incorporating. The implementation of supervised learning (making tagged datasets) is challenging when dealing with these systems' complex behavior, which must be learnt. Feature extraction systems, need a lot of processing capacity to operate because they operate at a lower level of abstraction. Because they require strong both off-board processing and communication capability, these resources are presently challenging to incorporate on board UAVs. Furthermore, the applications that need reactive behaviors are restricted by the limited computational resources that are often incompatible with online processing. This provides the previously noted the difficulty of creating embedded hardware technologies, but it ought to furthermore motivate academics to create deeper learning systems with greater efficiency. The outcomes even demonstrate that the accuracy/speed trade-off may be modified by only altering the input picture size of the detection networks to achieve the ideal balance. The methods under test used SSDs with various input sizes. Additionally, examined a YOLOv2 network with a bigger input size was discovered to be quicker but less dependable than SSD. The outcome of an SSD network with an input size of 420x420, allowing for fast computation as well as very accurate robot detections.

References

1. Adrian Carrio, Carlos Sampedro, Alejandro Rodriguez-Ramos, and Pascual Campoy, “A Review of Deep Learning Methods and Applications for Unmanned Aerial Vehicles ", Hindawi Journal of Sensors Volume 2017, Article ID 3296874, 13 pages, (2017).
2. J. L. Sanchez-Lopez, M. Molina, H. Bavle, Carlos Sampedro, Ramón A. Suárez Fernandez, “A multilayered component-based approach for the development of aerial robotic systems: The aerostack framework,” Journal of Intelligent & Robotic Systems, pp. 1–27, (2017).
3. Markus Teigen Pike, " Computer Vision and Deep Learning in Autonomous Drones", Norwegian University of Science and Technology (NTNU), June (2017).
4. Tianpei Liao, Amal Haridevan, Yibo Liu, Jinjun Shan,"Autonomous Vision-based UAV Landing with Collision Avoidance using Deep Learning", arXiv: 2109.08628v1[cs.LG],17 Sep (2021).
5. Lucas Prado Osco a, *, Jos´e Marcato Junior b, Ana Paula Marques Ramos c,d, Lúcio Andr´e de Castro Jorge e, Sarah Narges Fatholahi f, Jonathan de Andrade Silva g, Edson Takashi Matsubara g, Hemerson Pistori g,h, Wesley Nunes Gonçalves b,g, Jonathan Li f," A review on deep learning in UAV remote sensing",International Journal of Applied Earth Observations and Geoinformation 102 (2021) 102456, 27 July (2021).

6. Xin Liu and Zhanyue Zhang," A Vision-Based Target Detection, Tracking, and Positioning Algorithm for Unmanned Aerial Vehicle", Wireless Communications and Mobile Computing Volume 2021, Article ID 5565589, 12 pages, (2021).
7. Farhad Samadzadegan, Farzaneh Dadrass Javan 1,2, *, Farnaz Ashtari Mahini 1 and Mehrnaz Gholamshahi 3," Detection and Recognition of Drones Based on a Deep Convolutional Neural Network Using Visible Imagery", journal of aerospace, 9, 31, 10 January, (2022).
8. Li-Hua Wen and Kang-Hyun Jo," Deep learning-based perception systems for autonomous driving: A comprehensi survey", Neurocomputing 489 (2022) 255–270, 17 March (2022).
9. Ahmed Abdulhameed and Qurban A. Memon, "Support Vector Machine Based Design and Simulation of Air Traffic Management for Prioritized Landing of Large Number of UAVs", European Journal of Artificial Intelligent and Machine Learning,Vol 1|I s s u e 2|A p r i l, (2022).
10. Subrahmanyam Vaddi1, Dongyoun Kim1, Chandan Kumar1, Shafqat Shad1 & Ali Jannesari1," Efficient object detection model for real-time UAV applications" Computer and Information Science; Vol. 14, No. 1; 2021 ISSN 1913-8989 E-ISSN 1913-8997 Published by Canadian Center of Science and Education Online Published: January22, 2021, doi:10.5539/cis.v1 4n1p45.
11. Ren-Yi Kung,et al]," Application of Deep Learning and Unmanned Aerial Vehicle Building Maintenance", Advances in Civil EngineeringVolume 2021, Article ID 5598690, 12 pages https://doi.org/10.1155/2021/5598690.
12. Tanmay Kumar Behera, Sambit Bakshi, Pankaj Kumar Sa, "Vegetation Extraction from UAV-based Aerial Images through Deep Learning" Computers and Electronics in Agriculture, https://doi.org/10.1016/j.compag.2022.107094, Volume 198, July (2022).
13. Maissa Hamouda1, Karim Saheb Ettabaa2, Med Salim Bouhlel3, "Smart feature extraction and classification of hyperspectral images based on convolutional neural networks" IET Image Processing the Institution of Engineering and Technology 2020, Vol. 14 Iss. 10, pp. 1999-2005, (2020).
14. Ruben Geraldes et al, "UAV-Based Situational Awareness System Using Deep Learning", IEEE access Digital Object Identifier 10.1109/ACCESS.2019.2938249, volume 7,29, (2019).
15. Evangelos Maltezos et al, The INUS Platform: A Modular Solution for Object Detection and Tracking from UAVs and Terrestrial Surveillance Assets, www.mdpi.com/journal/computation, Computation 2021, 9, 12. https://doi.org/10.3390/computation9020012(2021).
16. Ahmed Nassar, Karim Amer, Reda ElHakim, Mohamed ElHelw " A Deep CNN-Based Framework for Enhanced Aerial Imagery Registration with Applications to UAV Geolocalization", IEEE Xplore, CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)18-22 June 2018, (2018).
17. David Safadinho, João Ramos, Roberto Ribeiro, Vítor Filipe, João Barroso and António Pereira," UAV Landing Using Computer Vision Techniques for Human Detection" journal/sensors (MDPI), 20, 613, (2020).
18. Efstratios Kakaletsis, Charalampos Symeonidis, Maria Tzelepi, Ioannis Mademlis, Anastasios Tefas, Nikos Nikolaidis, and Ioannis Pitas" Computer Vision for Autonomous UAV Flight Safety: An Overview and a Vision-based Safe Landing Pipeline Example", Vol. 1, No. 1, Article. Publication date:(2020).

19. Jatin Upadhyay, Abhishek Rawat and Dipankar Deb *," Multiple Drone Navigation and Formation Using Selective Target Tracking-Based Computer Vision", journal/electronics (MDPI), 10, 2125, (2021).
20. Phong Ha Nguyen, Muhammad Arsalan, Ja Hyung Koo, Rizwan Ali Naqvi, Noi Quang Truong and Kang Ryoung Park" LightDenseYOLO: A Fast and Accurate Marker Tracker for Autonomous UAV Landing by Visible Light Camera Sensor on Drone", Journal of Applied Sciences(MDPI), 18, 1703, (2018).
21. Mateusz Apoń1, a), Arkadiusz Nikonowicz1, b), Leszek Ambroziak1, c), Mirosław Kondratiuk1, d), Piotr Burzyński1, e) and Adam Kuczyński1, f)," Vision-Based Autonomous Object Tracking for Unmanned Aerial Vehicles", AIP Conference Proceedings, 29 October (2018).
22. Yiming Zhao," Empowering Computer Vision Models with Deep Learning for Robotic Perception and Drone Geolocation", Worcester Polytechnic Institute Phd Electrical and Computer Engineering, (2021).
23. Jihad Kadhim AbdAli and Maysoon Khazaal Abbas Maaroof," Phase Change Analysis of Robot Laser Drilling with Accuracy Enhancement by Deep Learning", International Journal of Analytical, Experimental and Finite Element Analysis Volume 9, Issue 3, pp 50 – 61, Sept (2022).
24. Seokwon Yeom * and In-Jun Cho," Detection and Tracking of Moving Pedestrians with aSmall Unmanned Aerial Vehicle", Journal of Applied Sciences(MDPI) 9, 3359, (2019).
25. Jihad Kadhim AbdAli and Maysoon Khazaal Abbas Maaroof, "Theoretical Investigations of Wing Coating on the Performance of UAV", International Journal of Advances in Engineering and Emerging Technology (IJAEET) Vol. 13, No.2, pp88-100, September (2022).
26. Adriano Garcia, Sandeep S. Mittal, Edward Kiewra and Kanad Ghose," A Convolutional Neural Network Vision System Approach to Indoor Autonomous Quadrotor Navigation ", The 23rd Conference on Process Integration, Modelling and Optimisation for Energy Saving and Pollution Reduction - PRES'20, (2019).
27. Florian Olsner, Stefan Milz," Catch Me, If You Can! A Mediated Perception Approach Towards Fully Autonomous Drone Racing " Proceedings of Machine Learning Research 123:90–99, (2020).
28. M. Saifizi1, *, N. Syahirah1, Wan Azani Mustafa1, Hasliza A Rahim2, Mohd Wafi Nasrudin2," Using Unmanned Aerial Vehicle in 3D Modelling of UniCITI Campus to Estimate Building Size", The 1st International Conference on Engineering and Technology (ICoEngTech) 2021, Journal of Physics: Conference Series, (2021).
29. 1S. Mantey and 1M. S. Aduah," Comparative Analysis of Stockpile Volume Estimation using UAV and GPS Techniques", GMJ Vol. 21, No.1, June., (2021).
30. Tianqi Wang1 and Dong Eui Chang1," Robust Navigation for Racing Drones based on Imitation Learning and Modularization", arXiv:2105.12923v1[cs.RO],27May (2021).
31. Li-Yu Lo 1, Chi Hao Yiu 1, Yu Tang 1, An-Shik Yang 2, Boyang Li 1, * and Chih-Yung Wen 1," Dynamic Object Tracking on Autonomous UAV System for Surveillance Applications", journal/sensors, 21, 7888, 27 November (2021).
32. Yunlong Song∗, Mats Steinweg∗, Elia Kaufmann, and Davide Scaramuzza," Autonomous Drone Racing with Deep Reinforcement Learning" International Conference on Intelligent Robots and Systems (IROS), Prague, 2021. ©IEEE, (2021).

33. Leonard Bauersfeld∗, Elia Kaufmann∗, Philipp Foehn, Sihao Sun, Davide Scaramuzza," NeuroBEM: Hybrid Aerodynamic Quadrotor Model", Science and Systems 2021 conference, (2021).
34. Antonio Loquercio*, Alessandro Saviolo*, and Davide Scaramuzza," AutoTune: Controller Tuning for High-Speed Flight", arXiv:2103.10698v2 [cs.RO] 8 Feb (2022).
35. J. A. García-Pulido1 · G. Pajares2 · S. Dormido1," UAV Landing Platform Recognition Using Cognitive Computation Combining Geometric Analysis and Computer Vision Techniques", published online8 June, (2022).
36. Christian Pfeiffer1, Simon Wengeler1, Antonio Loquercio1, Davide Scaramuzza1," Visual Attention Prediction Improves Performance of Autonomous Drone Racing Agents", journal. pone.0264471, Research Article: PLOS ONE Vol. 17, Issue 3, (2022).