

Received 20 July 2022, accepted 29 August 2022, date of publication 6 September 2022, date of current version 15 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3204876

RESEARCH ARTICLE

Drone Navigation Using Region and Edge Exploitation-Based Deep CNN

MUHAMMAD ARIF ARSHAD^{1,2,3}, SADDAM HUSSAIN KHAN^{1,4}, SULEMAN QAMAR^{1,3},
MUHAMMAD WALEED KHAN^{1,3,5,6}, IQBAL MURTZA⁷, JEONGHWAN GWAK^{8,9,10,11},
AND ASIFULLAH KHAN^{1,3,12}

¹Pattern Recognition Lab, Department of Computer and Information Sciences, Pakistan Institute of Engineering & Applied Sciences, Nilore, Islamabad 45650, Pakistan

²Centres of Excellence in Science & Applied Technologies, Islamabad 44000, Pakistan

³PIEAS Artificial Intelligence Center, Pakistan Institute of Engineering & Applied Sciences, Nilore, Islamabad 45650, Pakistan

⁴Department of Computer Systems Engineering, University of Engineering and Applied Sciences, Swat 19060, Pakistan

⁵Department of Mechanical and Aerospace Engineering, Columbus, OH 43210, USA

⁶Center for Automotive Research, The Ohio State University, Columbus, OH 43212, USA

⁷Department of Creative Technologies, Faculty of Computing and AI, Air University, Islamabad 44230, Pakistan

⁸Department of Software, Korea National University of Transportation, Chungju 27469, South Korea

⁹Department of Biomedical Engineering, Korea National University of Transportation, Chungju 27469, South Korea

¹⁰Department of AI Robotics Engineering, Korea National University of Transportation, Chungju 27469, South Korea

¹¹Department of IT and Energy Convergence (BK21 FOUR), Korea National University of Transportation, Chungju 27469, South Korea

¹²Deep Learning Laboratory, Center for Mathematical Sciences, Pakistan Institute of Engineering & Applied Sciences, Nilore, Islamabad 45650, Pakistan

Corresponding authors: Asifullah Khan (asif@pieas.edu.pk) and Jeonghwan Gwak (jgwak@ut.ac.kr)

This work was supported by PAIC, PIEAS, IT Endowment under Pakistan Higher Education Commission (HEC) and “Regional Innovation Strategy (RIS)” through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2021RIS-001 (1345341783)) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant No. NRF-2020R111A3074141).

ABSTRACT Drones are unmanned aerial vehicles (UAV) utilized for a broad range of functions, including delivery, aerial surveillance, traffic monitoring, architecture monitoring, and even War-field. Drones confront significant obstacles while navigating independently in complex and highly dynamic environments. Moreover, the targeted objects within a dynamic environment have irregular morphology, occlusion, and minor contrast variation with the background. In this regard, a novel deep Convolutional Neural Network(CNN) based data-driven strategy is proposed for drone navigation in the complex and dynamic environment. The proposed Drone Split-Transform-and-Merge Region-and-Edge (Drone-STM-RENet) CNN is comprised of convolutional blocks where each block methodically implements region and edge operations to preserve a diverse set of targeted properties at multi-levels, especially in the congested environment. In each block, the systematic implementation of the average and max-pooling operations can deal with the region homogeneity and edge properties. Additionally, these convolutional blocks are merged at a multi-level to learn texture variation that efficiently discriminates the target from the background and helps obstacle avoidance. Finally, the Drone-STM-RENet generates steering angle and collision probability for each input image to control the drone moving while avoiding hindrances and allowing the UAV to spot risky situations and respond quickly, respectively. The proposed Drone-STM-RENet has been validated on two urban cars and bicycles datasets: udacity and collision-sequence, and achieved considerable performance in terms of explained variance (0.99), recall (95.47%), accuracy (96.26%), and F-score (91.95%). The promising performance of Drone-STM-RENet on urban road datasets suggests that the proposed model is generalizable and can be deployed for real-time autonomous drones navigation and real-world flights.

INDEX TERMS Residual network, drone, convolutional neural network, perception and autonomy, drone split transform merge.

The associate editor coordinating the review of this manuscript and approving it for publication was Chuan Li.

I. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) are one of the most significant disciplines in recent technology, with autonomous

drones being a key study focus. Self-flying, also called self-piloting, refers to a drone's capacity to conduct aerial movements without the assistance of a person. In this case, autonomy is defined as the drone's decision to run the aforementioned self-flying activities without the need for human intervention. UAVs operated manually face many functional and operational challenges. As a result, it is planned to build drones that a front and rear camera will control, from which the drone would get real-time visual information and act independently. A significant open challenge in robotics is the safe and dependable outside navigation of autonomous systems, such as unmanned aerial vehicles (UAVs). The autonomous agent must not only operate while avoiding accidents, but also interact with other agents in the environment, such as people or automobiles, in a safe manner.

Significant advancements have been achieved in the field of UAVs in the last decade, owing to the fast development of low-cost off-the-shelf drones. But it is challenging for autonomous systems [1], i.e. unmanned aerial vehicles (UAVs), to navigate safely and reliably. The ability to travel while avoiding obstacles is critical for applications such as traffic monitoring, surveillance, and construction purposes [2], [3] in urban areas. Due to the complexity of the environment, it becomes quite a challenging task [4]. The autonomous agent should interact with other agents and navigate while avoiding obstacles in these scenarios.

Two steps process are used to solve such problems using traditional approaches that includes (i) In a given map, automated localization is performed (using visual, GPS or/and any other range sensor) (ii) controlling the drones manually to avoid hindrances while accomplishing its goal [2], [5], [6].

Recently, new machine learning [7], [8] and deep learning techniques [9], [10] have been producing excellent results in various domains [11], [12], [13], [14], [15], [16]. It gives significant results in cyber security [17], [18], [19], [20] as well. Reinforcement learning (RL)-based techniques, in particular, suffer from a significant rise in sample complexity, making them unsuitable for usage by UAVs in safety-critical settings. Successful flying policies, on the other hand, may be learned using supervised-learning approaches [10], [11], [21], [22] However, it has not worked out how to collect enough expert trajectories from replicating yet. In addition, as mentioned by [21] drones must learn how to react in dangerous circumstances just like human pilots.

Due to its usage and suitability on commercially feasible drones that are often implemented with a front-looking camera and lack additional sensors that are power-hungry or obese. At the same time, advances in machine learning have improved visual navigation capabilities. Deep Neural Networks (DNNs) have enabled the creation of tail-to-tail learning methods [23]. Contrary to previous method, which has limited generalisation capabilities, DNNs offer visual navigation in real-world contexts where visual appearances are inevitably diverse [24].

In this work, we propose a drone navigation method using Region and Edge Exploitation-Based Deep CNN. Our main

focus is to provide a CNN model through which drone navigation can be done using 1 single mounted camera.

A UAV effectively flying in the streets must follow the road and react to dangerous circumstances in the same manner that any other manned ground vehicle would. As a result, we introduce employing the information acquired from ground automobiles incorporated in the above-mentioned settings. The Drone-STM-RENet architecture is compatible with the input feature map dimensions and output multi-class challenge by changing the initial and final layers (2 classes). Comprehensively, contributions made by this work are as follows:

- 1) We propose a novel Drone Split Transform Merge Region and Edged based convolutional neural network (Drone-STM-RENet) that can undertake a safe UAV flying in urban areas by predicting the probability of collision and steering angle.
- 2) For training, an outside dataset collected from vehicles and bicycles was used. To allow a UAV to detect potentially harmful scenarios, an outside collision sequences dataset is used.
- 3) In every block of the proposed Drone-STM-RENet, STM-based CNN blocks concept is developed, which leverages the concept of Region and Edge-based (RE) feature extraction systematically. Effective usage of RE-based operations at every branch of the Drone-STM-RENet block captures a wide range of characteristics on numerous levels, most notably those including obstacles.

Despite our system's impressive outcomes, we do not want to change the standard "map-localize-plan" drone navigation approaches; instead, we want to explore if a likewise task can be performed with a single shallow neural network. Traditional and learning-based techniques, we believe, will eventually complement one another.

II. LITERATURE REVIEW

In this section, a detailed review of the available literature is given, which is not only the inspiration for this research but provides insights on how Convolution Neural Network emerged as one of the most researched areas in artificial intelligence.

The obstacle identification and avoidance tasks [25] are closely linked with those of autonomous navigation. Object detection methods are based on either machine learning algorithms or computer vision techniques to identify obstacles.

The GPS range and optical sensors of an unmanned aerial vehicle (UAV) that operates outside are usually used to assess the device status, detect the presence of obstacles, and determine the flight route [2], [5]. However, these kinds of work are still likely to suffer in urban areas because of the building, huge rushes, and dynamic states. This results in critical unobserved errors in the estimation of system state. In such cases, SLAM is a typical approach in which the robot develops a map of the environment while also self-locating within it [26]. Although it may be beneficial for global navigation

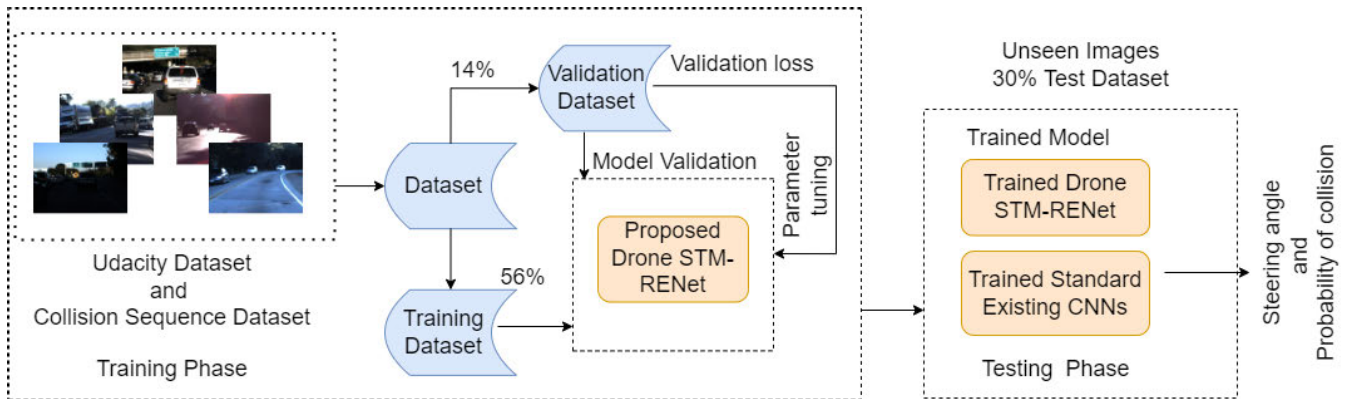


FIGURE 1. Dataset distribution for Drone-STM-RENet.

and localization, it is uncertain how to extract control commands for a secure and stable flight from an expressive 3D reconstruction of the surroundings.

A self-supervised deep CNN-based indoor navigation system was created by Alexandros Kouris [27]. Using a regression CNN to evaluate the agent's distance to collision based on raw visual input data from the inbuilt monocular camera addresses the issue of real-time obstacle avoidance. Using an external sensor placed on a drone, they trained the model on their indoor-flight dataset, which they generated with real-distance labels using an external sensor. CNN extracts Spatio-temporal features that capture both static presence and motion information by simultaneously processing the consecutive input frames, including the current and previous frames, to estimate the robot's distance from the closest hindrance in several directions, including the current and previous. Using these predictions, the linear velocity and yaw of the unmanned aerial vehicle (UAV) are adjusted to ensure safe navigation. When applied to real-world indoor flights, it produced state-of-the-art results, leaving behind previously reported techniques from the literature. They utilized a two-stream CNN architecture for Spatio-temporal feature extraction to estimate the distance between robots and the environment in various directions.

Using machine learning tools for object identification and classification, the authors of [28] conducted a thorough evaluation of the literature that addressed the idea of drone detection. In essence, the use of machine learning makes it easier to identify drones using a binary classification model such as "drone" or "no drone." However, some study in the literature uses a multiclass classification to identify different types of drones in addition to the conventional categorization. The article's first section outlines many goals including drone detection, verification, classification, and characterisation as well as a multi-drone detection method based on radar signals. The important study on drone detection is recognised as using high-end 3-D holographic radar coupled with machine learning of time-domain properties.

The D-CNN based-model developed by Karim Amer [29] as it performs well in different computer vision [30] tasks

such as detection [31], [32], localization [33] and segmentation [29], [34]. A CNN in conjunction with a regressor is utilized to generate the drone steering instructions. The information was enhanced to create a 'navigation envelope.' In applications such as surveillance, package delivery, or humanitarian assistance distribution, the technique may be used to automate drone navigation to reduce the number of excursions or visits to the same location. A D-CNN was built to produce drone steering instructions based on observed images and to achieve autonomous drone navigation. The suggested approach uses video acquired by a camera mounted on drone. When a CNN and fully connected regression are used together, it has been shown that it is possible to forecast the steering angles required to fly the drone on its planned path with high accuracy.

Wang *et al.* [35] utilized some generic CNN-based object detectors for the computer vision challenge using Stanford Drone Dataset. The employment of a focal loss dense detector RetinaNet-based method for fast and efficient object identification from a drone has yielded results that are at the cutting edge of the field. In their research, Saripalli *et al.* [36] looked into vision-based autonomous control techniques for UAV landings. Scaramuzza *et al.* [37] integrated vision approaches into their Unmanned aerial Vehicle design to improve navigation accuracy. PIXHAWK [38] is a well-known flight control device that uses machine vision algorithms to detect obstacles during UAV operations. Kendoul *et al.* [39] used machine vision methods focused on optical flow to incorporate self-manoeuvring operations for aerial vehicles. He successfully investigated the use of CNN-based object detectors, especially the recently reported RetinaNet focal loss dense detector for Unmanned aerial vehicles object detection.

Ashraf *et al.* [40] reported a two-staged network. Targets drones' irregular movement, tiny size, arbitrary form, significant intensity fluctuations, and occlusion makes drone identification a difficult task [28]. Drones' low costs will increase the number of unmanned aerial vehicles (UAVs) in the sky. Researchers for a variety of purposes have tackled the issue of drone detection. The method described in this

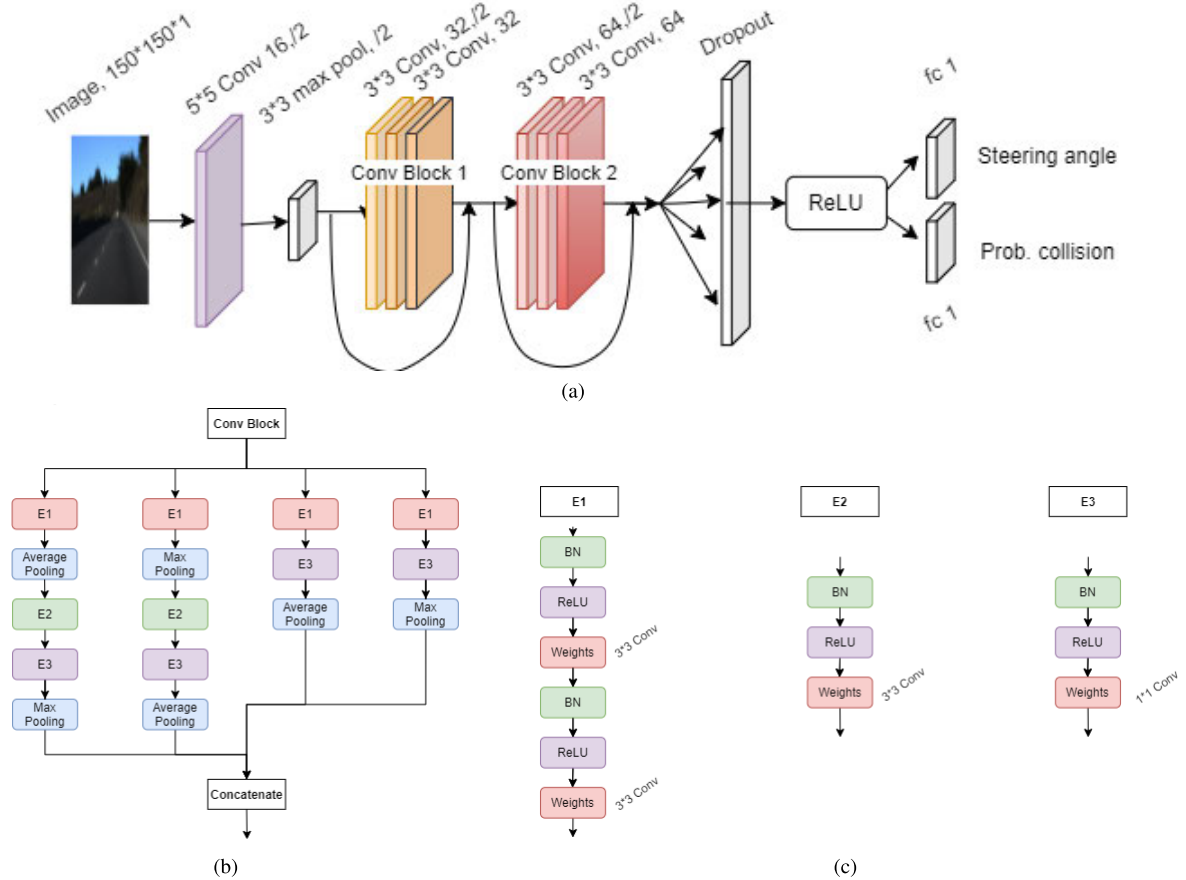


FIGURE 2. (a) Drone-STM-RENet is a branched CNN that anticipates the steering angle and collision probability from a single 150 x 150 grey-scale frame. A Drone-STM-RENet with two residual blocks (b) forms the common component of the architecture, which is accompanied by 50% dropout and ReLU non-linearity. Following that, the architecture is divided into two different fully connected layers, one for predicting steering angle and the other for inferring collision probability. In the figure above, we first give the kernel size, and after that the number of filters, and finally, the stride if it's not 1. Residual blocks consist (c) a combination of Batch normalization, ReLU activation function, and convolution layer.

article does not need precisely centered cuboids and instead uses I3D to learn rich Spatio-temporal information. Rather than relying on region-proposal-based techniques, this article suggested a two-stage segmentation-based strategy using Spatio-temporal attention cues. This article aims to identify and locate drones in multiple video frames recorded by other drones. It reported a two-stage segmentation-based detection approach for drones in heavily populated areas. The first stage is solely visual, while the second stage is spatialtemporal.

Xie *et al.* [41] leveraged aggregated residual transformations for image classification. This article's methodology is based on three observations: (i) A bottom-up segmentation-based technique for drone identification that classifies each pixel is superior to a region proposal-based approach.

(ii) Optical flow data has been successfully used in a number of research, including action recognition.

(iii) Because large motions of the target and source drones may be insufficient, we must depend on optical flow data.

This starts with appearance-based pixel classification to accurately find drones and then adapts ResNet50 [42] to maintain local information as the network becomes more

significant. It utilized pixel- and channel-level attentions to concentrate feature maps on the foreground. This article described a two-stage method for detecting flying drones using Spatio-temporal cues. It utilized a segmentation-based approach rather than depending on region-based techniques for effective drone identification.

III. PROPOSED TECHNIQUE

A. CNN WITH DEEP CHANNEL BOOSTING FOR DRONE AUTONOMY

Accurate analysis of real-time images [43] in a changing environment is complex due to the following factors: (i) low contrast variation between foreground and background boundaries, (ii) high texture variation (iii) significant variation in the size, shape, and position of the foreground in images and (iv) low illumination. Additionally, these photos are heavily deformed due to the shifting environment's noise level.

This work reported approach for automating drones based on CB-CNN. The suggested approach is used to forecast the steering angle and probability of collision. In this context,



FIGURE 3. (a) Images from Udacity were utilized for learning the steering angles. (b) The images were gathered to evaluate the possibility of a collision. Green box indicates no-collision frames, whereas red box indicates collision frames.

we develop a new CNN classifier based on innovative Drone-STM-RENet blocked [44] that incorporates RE-based procedures for steering angle and collision probability learning. The workflow for the drone navigation is illustrated in fig. 1.

1) PROPOSED DRONE-STM-RENet

D-CNNs have a lot of pattern mining capabilities, which is why it's so prevalent in image processing [45]. Because of its automated feature extraction capability, CNN has outperformed traditional Machine Learning algorithms on visual observation tasks. It uses a convolution technique to leverage the image's structural information and extracts feature hierarchies dynamically based on the intended application. To investigate the background-foreground issue in Drone-STM-RENet, unique convolution blocks based on split-transform-merge (STM) techniques are devised and implemented. It is the first time an individual block has been developed that consistently performs region and edge-based operations at each branch to capture a comprehensive collection of properties at many levels, particularly those related to region homogeneity, textural changes, and backdrop borders. Many advancements in CNN architecture have increased its application in the robotics sector for robot vision. Before this, the best of our knowledge practice was to squeeze and excite using 1×1 dimension reduction or expansion, but this is a novel squeeze and excitation technique idea that uses squeezing (channel concatenation) and excitation (channel concatenation). By squeezing and concatenating, we choose salient feature maps.

In this work, a new CNN architecture based on the innovative Drone-STM-RENet based feature extraction has been developed. This new architecture for drones is referred to as the Drone STM-based RENet (Drone-STM-RENet) architecture shown in figure 2. Drone-STM-RENet creates unique convolution blocks based on split-transform-merge (STM) to investigate the background-foreground dilemma. This innovative block systematically executes region and edge-based operations at each branch to capture the broad range of characteristics at many levels, particularly those related to region homogeneity, textural changes, and background borders. The

proposed block is made up of four sub-branches, as shown in the diagram. The principle of Region and Edge-based feature extraction is applied thoroughly at every branch, with maximum and average-pooling along with convolution and ReLU-activation to capture discriminating features in considerable detail. To extract patterns from an image dataset, the Drone-STM-RENet separates the input into four branches, uses the RE-based operator to learn region-specific variations and their distinct boundaries, and then uses the concatenation operation to merge the output from numerous paths. Drone-STM-RENet extracts a different set of abstract level features by stacking two blocks of STM with the identical topology in series. At the end of the process, Dropout is used, followed by ReLU activation and two fully connected layers in parallel for steering angle prediction and for calculating the probability of colliding with another vehicle. Incorporating this concept into the Drone-STM-RENet allows it to extract various variants from the input feature maps.

2) PROPOSED DRONE-STM-RENet CHANNEL BOOSTING (CDSTM-RENet-CB)

Vehicle data has a lot of variance in the images that's why a strong CNN is essential for excellent discrimination. Using Channel Boosting [46], [47], the proposed Drone-STM-RENet's discriminating ability is improved. It proposed the concept of Channel Boosting to solve complex problems. Extraction of significant characteristics from distorted pictures is made possible by average smoothing of the image contents inside the distorted images recorded, and outliers are also managed using the suggested approach. The region and edge operation assists in managing region homogeneity and smoothing and the systematic exploitation of resources inside a given block. It helps delineate discriminating boundary or edge characteristics.

B. IMPORTANCE OF USING AUXILIARY CHANNELS

CNNs with various architectural designs have varying capabilities for feature learning. Multilevel information can be seen in many channels learned from distinct deep CNNs. These channels reflect different patterns that might help in

TABLE 1. Quantitative data on problems involving regression and classification: When doing the steering regression task, EVA and RMSE are measured, and when performing the collision prediction task, Average accuracy and F1-score are evaluated. Also we have calculated confidence interval for recall at 95%. In comparison to the baselines that were evaluated, our model performed wonderfully. Even though Drone-STM-RENet does not have many parameters, it performs brilliantly on both tasks.

Model	EVA	RMSE	Avg. Accuracy	F-1 Score	Recall+-C.I.	Num. Parameters	Num. Layers
Random Baseline	-1+-0.022	-0.3+-0.001	50+-1%	0.3+-0.001	-	-	-
Constant Baseline	-	0.34	75.6%	-	-	-	-
Resnet-50 [42]	0.795	0.097	96.6%	0.92	-	$2.6 * 10^7$	50
VGG-16 [50]	0.712	0.119	92.7%	0.85	-	$7.5 * 10^6$	16
Dronet [23]	0.737	0.109	95.4%	0.90	-	$3.2 * 10^5$	8
R. Chew et al. [52]	-	-	86.0%	0.86	-	-	-
HD+LO Networks [53]	-	-	96.1%	-	-	-	12
Drone-STM-RENet	0.99	0.48	96.26%	0.92	0.9547+-0.064	$5.6 * 10^5$	8

the exact explanation of class-specific features. The local and global representation of the image may be improved by combining diverse-level abstractions learned from multiple sources. The term “intelligent feature-space ensemble” refers to the concatenation of auxiliary and primary channels [48] in which a single learner makes the ultimate choice by assessing numerous image-specific patterns [49].

C. IMPLEMENTATION OF THE EXISTING CNNs

For the evaluation of proposed architecture, several known deep CNNs (ResNet50 [42], VGG16 [50], Dronet [23]) have been implemented. To provide a fair comparison with the suggested technique, all models are first trained on the same data set from scratch.

IV. EXPERIMENTAL SETUP

A. DATASET

The deep CNN models are trained and evaluated using a holdout cross validation technique. The data set was split into three parts: a training set, a validation set, and a testing set. The testing set was used to evaluate the model, which was maintained distinct from the training and validation datasets.

We utilize one of the freely accessible datasets from Udacity’s project [51] to learn steering angles. Nearly 60,000 images of car driving are split across five experiments, 1 for testing and 4 for training. Every experiment saves images from three cameras (left, center, and right) and data from the GPS, IMU, brake, steering angles, throttle, gear, and speed. To learn the probability of collision from images, We utilize a dataset that is freely available online from Dronet’s project [23]. They collected the dataset by mounting the GoPro camera on the bicycle’s handlebar. And drive the bicycle in multiple city areas, attempting to mix up the barriers (vehicles, people, plants, construction sites) and appearance of surroundings. As a result, the drone can generalize in a variety of settings. This dataset has 32000 images spread across 137 sequences. In these sequences, frames are labeled as 0 (no collision) if the distance from the obstacle is enough, which means there is less chance of collision, and those frames in which there are some obstacles are labeled as 1 (collision). Frames labeled as 1 (i.e. collision frames) are that kind of data that is very difficult for a drone to gather yet required for the development of a reliable as well as safe

TABLE 2. Training hyper-parameters.

Sr. No.	Hyper-parameter	Value
1	Epochs	100
2	Batch size	16
3	Learning Rate	10^{-5}
4	Optimizer	Adam

technique. We divided the dataset into training validation and testing. For training 51520 and for validation 12880 images are used while for testing we uses 30% of the total data i.e. 27600 Dataset examples are shown in fig. 3.

B. DRONE CONTROL

The UAV is instructed to fly with a forward velocity of v_k and a steering angle of θ_k using the outputs. The network uses the probability of collision to regulate the forward velocity: when the collision probability is zero, vehicles are commanded to move with the maximum velocity i.e. V_{max} , and it stops when the probability of collision is close to 1. The forward velocity is filtered using a low pass filter ($1 > \alpha > 0$) as shown in (1).

$$v_k = (1 - \alpha)v_{k-1} + \alpha(1 - p_t)V_{max} \quad (1)$$

Similarly, The predicted steering angle is also converted into a yaw angle (rotation around the z-axis). We transform s_k from the $[-1, 1]$ range to the required yaw angle θ_k in the $[\frac{-\pi}{2}, \frac{\pi}{2}]$ range and low-pass-filter it as shown in (2).

$$\theta_k = (1 - \beta)\theta_{k-1} + \beta(\frac{\pi}{2})s_k \quad (2)$$

Lastly, a novice dynamic navigation strategy that will operate a drone accurately with only a single forward-looking camera is developed. Our method has the advantage of calculating a collision probability using one image, excluding any prior knowing the platform’s velocity. We believe that the proposed architecture will be making decisions based on the range between noticed items in the sphere of vision [42].

C. HYPER-PARAMETER

Hyper-parameters that are used for training is illustrated in the table 2.

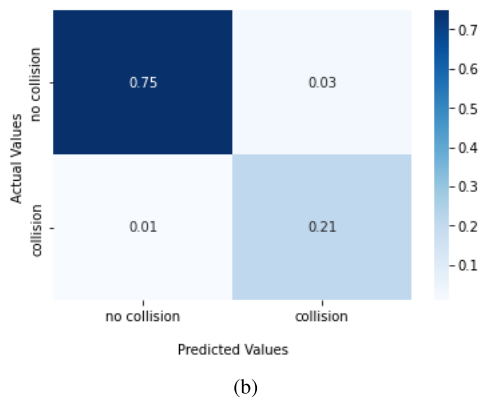
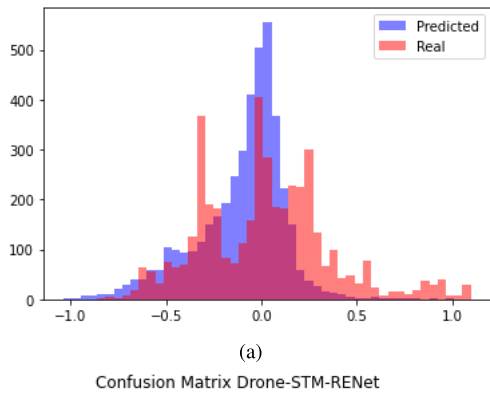


FIGURE 4. Model performance: (a) Actual vs anticipated steerings of the Udacity dataset testing sequence using the Probability Density Function (PDF). (b) Confusion matrix for collision categorization using test images from the dataset. The percentage of samples in each category is represented by the numbers in this matrix.

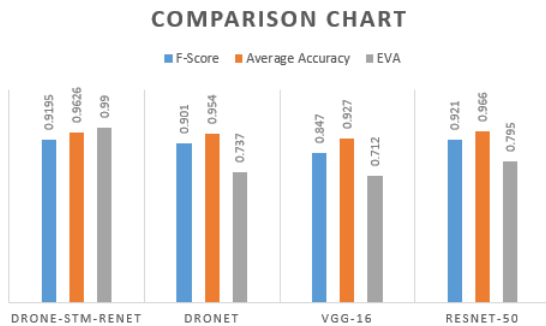


FIGURE 5. Comparison of various known architecture with Drone-STM-RENet.

V. RESULTS

A. REGRESSION AND CLASSIFICATION RESULTS

We initially analyze our model’s regression performance using the Udacity dataset’s testing sequence [54]. We utilize two measures to measure steering prediction performance: Root Mean Square Error (RSME) and Explained Variance Ratio (EVA). We use F-score and average classification accuracy to evaluate collision prediction performance.

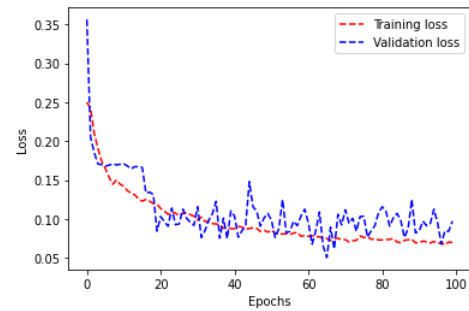


FIGURE 6. Training loss of the proposed Drone-STM-RENet.

Table 1 compares Drone-STM-RENet architecture with several known architecture from the literature [10], [42], [50]. Weak baselines include a constant estimator that anticipates 0 for steering angle always and “no collision,” as well as a random estimator. Our method outperforms it in terms of prediction accuracy. In Figure 5 comparison of various known architectures can be seen along with the Drone-STM-RENet and it can be seen that our architecture is performing very well in comparison to other known architectures in literature and also number of parameters and number of layers are less as shown in 1. Additionally, a favorable contrast to the VGG-16 network [50] demonstrates a utilization for the residual learning scheme in terms of generalization. As demonstrated in Table 1 and fig. 4 our design achieves a great performance as compared to other models in the literature.

B. PERFORMANCE METRICS

Various common performance indicators are used to assess the performance of the implemented models. Accuracy, recall, and F-score is examples of these measures. (3) is used to assess accuracy by counting the total number of accurate assignments. Recall is a metric that measures the fraction of accurate collision probability estimates (4). The F-score is specified by (5). Explained Variance is a metric that is used to measure the quality of a regressor (6). The major goal of Equation (7) is to enhance the true positive rate while lowering False-Negative for foreground or region of interest detection. As a result, the Standard Error (S.E.) at 95 percent Confidence Interval (CI) is presented for recall/sensitivity/detection rate to identify the uncertainty of the proposed Drone-STM-RENet [46]. In this instance, $z = 1.96$ for S.E. at the 95% CI. The mistake is expressed as $(100-98)/2$, or 20%. Images or the size of the dataset are both considered to be total samples. Figure 6 shows learning plots for the proposed Drone-STM-RENet, which illustrate loss values for the training set and the validation set.

$$Accuracy = \frac{TruePositive + TrueNegative}{Totalnumbersamples} \quad (3)$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (4)$$

$$F - Score = 2 * \frac{precision * recall}{precision + recall} \quad (5)$$

$$EVA = \frac{\text{Var}[Y_{true} - Y_{pred}]}{\text{Var}[Y_{true}]} \quad (6)$$

$$CI = z * \sqrt{\frac{\text{error}(1 - \text{error})}{\text{TotalSamples}}} \quad (7)$$

VI. DISCUSSION

Both traditional and learning-based methodologies have benefits and drawbacks, and our system is no different. The advantages include the ability for a drone to securely explore previously uncharted places by applying our easy learning and control approach. Unlike previous systems, no online environment map or pre-defined collision-free locations is required. Furthermore, we demonstrated remarkable generalizability across a wide range of situations. It might be a valuable adjunct to normal “map-localize-plan” navigation procedures in tasks like search and rescue and aerial deliveries. Furthermore, our method may be useful for platforms with limited resources because of the relatively simple and efficient network architecture. One disadvantage is that the agile dynamics of drones aren’t fully used. As a result, unlike prior CNN-based controllers [10], [55], [56], It is not feasible to give the robot a specific aim to pursue. There are several approaches to coping with the limits outlined above. When there is a high probability of collision, 3D collision-free pathways can be built to take use of the drone’s agility, as demonstrated in [56]. The distance to the destination [57] may be approximated using a 2D map and utilized [58] to generalize goal-driven challenges. Furthermore, as stated in [59], a measure of uncertainty might be introduced into our system to improve its resilience. The system could then switch back to a safety mode whenever it was required.

VII. CONCLUSION

Drones can confront various difficulties when navigating unstable and highly dynamic environments. This paper presented a new architecture called Drone-STM-RENet that can safely pilot a drone across city streets. This architecture is based on the Split transform merge concept. As a result, utilizing an unmanned aerial vehicle (UAV) to collect data in an uncertain environment is both unsafe and time-consuming. As a result, our proposed techniques learn to fly by imitating automobiles and bikers, which already follow traffic regulations. When compared to other well-known architectures in the literature, this model gives promising results by predicting collision probability and steering angle with accuracy (96.26%), recall (95.47%), F-score (91.95%) and explained variance (0.99) by Drone-STM-RENet enabling a UAV to respond swiftly to unforeseen occurrences and obstacles. The reason behind the performance of Drone-STM-RENet is that it captures the texture variations prominent features are boosted. Apart from this, edge operations and region homogeneity are dealt with through RENet block, which helps the drone to differentiate between background and foreground in challenging scenarios. In comparison to previous strategies utilized in the literature, the proposed Drone-STM-

RENet converges quite quickly. Extensive experimentation has demonstrated that a drone can be trained for urban navigation by emulating manned autos. It might be beneficial to typical “map-localize-plan” procedures in navigation-related operations like aerial deliveries.

REFERENCES

- [1] A. Mushtaq, I. U. Haq, M. U. Imtiaz, A. Khan, and O. Shafiq, “Traffic flow management of autonomous vehicles using deep reinforcement learning and smart rerouting,” *IEEE Access*, vol. 9, pp. 51005–51019, 2021.
- [2] S. Scherer, J. Rehder, S. Achar, H. Cover, A. Chambers, S. Nuske, and S. Singh, “River mapping from a flying robot: State estimation, river detection, and obstacle mapping,” *Auto. Robots*, vol. 33, no. 1, pp. 189–214, 2012.
- [3] M. Faessler, F. Fontana, C. Forster, E. Mueggler, M. Pizzoli, and D. Scaramuzza, “Autonomous, vision-based flight and live dense 3D mapping with a quadrotor micro aerial vehicle,” *J. Field Robot.*, vol. 33, no. 4, pp. 431–450, 2016.
- [4] S. Qamar, S. H. Khan, M. A. Arshad, M. Qamar, and A. Khan, “Autonomous drone swarm navigation and multi-target tracking in 3D environments with dynamic obstacles,” 2022, *arXiv:2202.06253*.
- [5] S. Shen, Y. Mulgaonkar, N. Michael, and V. Kumar, “Multi-sensor fusion for robust autonomous flight in indoor and outdoor environments with a rotorcraft MAV,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 4974–4981.
- [6] T. Lee, S. McKeever, and J. Courtney, “Flying free: A research overview of deep learning in drone navigation autonomy,” *Drones*, vol. 5, no. 2, p. 52, Jun. 2021.
- [7] A. Khan, M. H. Shamsi, and T.-S. Choi, “Correlating dynamical mechanical properties with temperature and clay composition of polymer-clay nanocomposites,” *Comput. Mater. Sci.*, vol. 45, no. 2, pp. 257–265, Apr. 2009.
- [8] M. Tahir, A. Khan, and A. Majid, “Protein subcellular localization of fluorescence imagery using spatial and transform domain features,” *Bioinformatics*, vol. 28, no. 1, pp. 91–97, Jan. 2012.
- [9] S. H. Khan, A. Sohail, A. Khan, M. Hassan, Y. S. Lee, J. Alam, A. Basit, and S. Zubair, “COVID-19 detection in chest X-ray images using deep boosted hybrid learning,” *Comput. Biol. Med.*, vol. 137, Oct. 2021, Art. no. 104816.
- [10] A. Giusti, J. Guzzi, D. C. Ciresan, F. L. He, J. P. Rodríguez, F. Fontana, M. Faessler, C. Forster, J. Schmidhuber, G. D. Caro, and D. Scaramuzza, “A machine learning approach to visual perception of forest trails for mobile robots,” *IEEE Robot. Automat. Lett.*, vol. 1, no. 2, pp. 661–667, Dec. 2015.
- [11] S. Ross, N. Melik-Barkhudarov, K. S. Shankar, A. Wendel, D. Dey, J. A. Bagnell, and M. Hebert, “Learning monocular reactive UAV control in cluttered natural environments,” in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2013, pp. 1765–1772.
- [12] A. Khan, A. Sohail, and A. Ali, “A new channel boosted convolutional neural network using transfer learning,” 2018, *arXiv:1804.08528*.
- [13] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning,” 2015, *arXiv:1509.02971*.
- [14] S. H. Khan, A. Sohail, M. M. Zafar, and A. Khan, “Coronavirus disease analysis using chest X-ray images and a novel deep convolutional neural network,” *Photodiagnosis Photodyn. Therapy*, vol. 35, Sep. 2021, Art. no. 102473.
- [15] A. Khan, S. H. Khan, M. Saif, A. Batool, A. Sohail, and M. W. Khan, “A survey of deep learning techniques for the analysis of COVID-19 and their usability for detecting omicron,” 2022, *arXiv:2202.06372*.
- [16] J. Long et al., “A novel self-training semi-supervised deep learning approach for machinery fault diagnosis,” *Int. J. Prod. Res.*, pp. 1–14, 2022.
- [17] M. Asam, S. J. Hussain, M. Mohatram, S. H. Khan, T. Jamal, A. Zafar, A. Khan, M. U. Ali, and U. Zahoor, “Detection of exceptional malware variants using deep boosted feature spaces and machine learning,” *Appl. Sci.*, vol. 11, no. 21, p. 10464, Nov. 2021.
- [18] M. Asam, S. H. Khan, T. Jamal, U. Zahoor, and A. Khan, “Malware classification using deep boosted learning,” 2021, *arXiv:2107.04008*.
- [19] R. Chamlawi, A. Khan, and A. Idris, “Wavelet based image authentication and recovery,” *J. Comput. Sci. Technol.*, vol. 22, no. 6, pp. 795–804, Nov. 2007.

- [20] M. Asam, S. H. Khan, T. Jamal, and A. Khan, "IoT malware detection architecture using a novel channel boosted and squeezed CNN," 2022, *arXiv:2202.04121*.
- [21] D. Gandhi, L. Pinto, and A. Gupta, "Learning to fly by crashing," in *Proc. IEEE/RJS Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 3948–3955.
- [22] Z. Yang, J. Long, Y. Zi, S. Zhang, and C. Li, "Incremental novelty identification from initially one-class learning to unknown abnormality classification," *IEEE Trans. Ind. Electron.*, vol. 69, no. 7, pp. 7394–7404, Jul. 2022.
- [23] A. Loquercio, A. I. Maqueda, C. R. D. Blanco, and D. Scaramuzza, "DroNet: Learning to fly by driving," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 1088–1095, Apr. 2018.
- [24] S. Javadi, M. Dahl, and M. I. Pettersson, "Vehicle detection in aerial images based on 3D depth maps and deep neural networks," *IEEE Access*, vol. 9, pp. 8381–8391, 2021.
- [25] H.-Y. Lin and X.-Z. Peng, "Autonomous quadrotor navigation with vision based obstacle avoidance and path planning," *IEEE Access*, vol. 9, pp. 102450–102459, 2021.
- [26] S. Lynen, T. Sattler, M. Bosse, J. A. Hesch, M. Pollefeys, and R. Siegwart, "Get out of my lab: Large-scale, real-time visual-inertial localization," in *Robotics: Science and Systems*, vol. 1, Jul. 2015, p. 1.
- [27] A. Kouris and C.-S. Bouganis, "Learning to fly by MySelf: A self-supervised CNN-based approach for autonomous navigation," in *Proc. IEEE/RJS Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 1–9.
- [28] B. Taha and A. Shoufan, "Machine learning-based drone detection and classification: State-of-the-art in research," *IEEE Access*, vol. 7, pp. 138669–138682, 2019.
- [29] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [30] M. Q. Butt and A. U. Rahman, "Audiovisual saliency prediction in uncategorized video sequences based on audio–video correlation," 2021, *arXiv:2101.03966*.
- [31] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, May 2016, pp. 779–788.
- [32] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
- [33] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," 2013, *arXiv:1312.6229*.
- [34] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. Cham, Switzerland: Springer*, Oct. 2015, pp. 234–241.
- [35] X. Wang, P. Cheng, X. Liu, and B. Uzochukwu, "Fast and accurate, convolutional neural network based approach for object detection from UAV," in *Proc. 44th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2018, pp. 3171–3175.
- [36] S. Saripalli, J. F. Montgomery, and G. S. Sukhatme, "Vision-based autonomous landing of an unmanned aerial vehicle," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2002, pp. 2799–2804.
- [37] D. Scaramuzza, M. C. Achtelik, L. Doitsidis, F. Friedrich, E. Kosmatopoulos, A. Martinelli, M. W. Achtelik, M. Chli, S. Chatzichristofis, L. Kneip, and D. Gurdan, "Vision-controlled micro flying robots: From system design to autonomous navigation and mapping in GPS-denied environments," *IEEE Robot. Automat. Mag.*, vol. 21, no. 3, pp. 26–40, Sep. 2014.
- [38] L. Meier, P. Tanskanen, L. Heng, G. H. Lee, F. Fraundorfer, and M. Pollefeys, "PIXHAWK: A micro aerial vehicle design for autonomous flight using onboard computer vision," *Auto. Robots*, vol. 33, nos. 1–2, pp. 21–39, Aug. 2012.
- [39] F. Kendoul, I. Fantoni, and K. Nonami, "Optic flow-based vision system for autonomous 3D localization and control of small aerial vehicles," *Robot. Auto. Syst.*, vol. 57, nos. 6–7, pp. 591–602, Jun. 2009.
- [40] M. W. Ashraf, W. Sultani, and M. Shah, "Dogfight: Detecting drones from drones videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7067–7076.
- [41] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1492–1500.
- [42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [43] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks," *IEEE Access*, vol. 7, pp. 59069–59080, 2019.
- [44] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [45] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng, "Recent progress on generative adversarial networks (GANs): A survey," *IEEE Access*, vol. 7, pp. 36322–36333, 2019.
- [46] S. H. Khan, A. Sohail, A. Khan, and Y.-S. Lee, "COVID-19 detection in chest X-ray images using a new channel boosted CNN," *Diagnostics*, vol. 12, no. 2, p. 267, Jan. 2022.
- [47] S. H. Khan, N. S. Shah, R. Nuzhat, A. Majid, H. Alquhayz, and A. Khan, "Malaria parasite classification framework using a novel channel squeezed and boosted CNN," *Microscopy*, May 2022.
- [48] G. Tian, Y. Sun, Y. Liu, X. Zeng, M. Wang, Y. Liu, J. Zhang, and J. Chen, "Adding before pruning: Sparse filter fusion for deep convolutional neural networks via auxiliary attention," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 6, 2021, doi: 10.1109/TNNLS.2021.3106917.
- [49] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe, and S. Mougiakakou, "Multisource transfer learning with convolutional neural networks for lung pattern analysis," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 76–84, Jan. 2017.
- [50] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [51] M. Mancini, G. Costante, P. Valigi, T. A. Ciarfuglia, J. Delmerico, and D. Scaramuzza, "Toward domain independence for learning-based monocular depth estimation," *IEEE Robot. Autom. Lett.*, vol. 2, no. 3, pp. 1778–1785, Jul. 2017.
- [52] R. Chew, J. Rineer, R. Beach, M. O'Neil, N. Ujeneza, D. Lapidus, T. Miano, M. Hegarty-Craver, J. Polly, and D. S. Temple, "Deep neural networks and transfer learning for food crop identification in UAV images," *Drones*, vol. 4, no. 1, p. 7, Feb. 2020.
- [53] S. Back, G. Cho, J. Oh, X.-T. Tran, and H. Oh, "Autonomous UAV trail navigation with obstacle avoidance using deep neural networks," *J. Intell. Robot. Syst.*, vol. 100, nos. 3–4, pp. 1195–1211, Dec. 2020.
- [54] A. Kazakova, "Parallelization of autonomous driving tasks for safe high-speed vehicle control," Ph.D. dissertation, College Eng./Eng. Pract. School, Dept. Comput. Sci. Eng.
- [55] N. Smolyanskiy, A. Kamenev, J. Smith, and S. Birchfield, "Toward low-flying autonomous MAV trail navigation using deep neural networks for environmental awareness," in *Proc. IEEE/RJS Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 4241–4247.
- [56] S. Yang, S. Konam, C. Ma, S. Rosenthal, M. Veloso, and S. Scherer, "Obstacle avoidance through deep networks based intermediate perception," 2017, *arXiv:1704.08759*.
- [57] M. Pfeiffer, M. Schaeuble, J. Nieto, R. Siegwart, and C. Cadena, "From perception to decision: A data-driven approach to end-to-end motion planning for autonomous ground robots," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 1527–1533.
- [58] W. Gao, D. Hsu, W. S. Lee, S. Shen, and K. Subramanian, "Intention-Net: Integrating planning and deep learning for goal-directed autonomous navigation," in *Proc. Conf. Robot Learn.*, 2017, pp. 185–194.
- [59] C. Richter and N. Roy, "Safe visual navigation via deep learning and novelty detection," in *Proc. Robot., Sci. Syst. XIII*, Jul. 2017.



MUHAMMAD ARIF ARSHAD received the bachelor's degree in computer systems engineering from the Islamia University of Bahawalpur (IUB), in 2019, and the master's degree in computer science from the Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad, Pakistan, in 2021. He is currently doing his research at CESAT. His research interests include artificial intelligence, computer vision, deep learning, and machine learning. He received the bronze medal for bachelor's degree and also received the silver medal for master's degree.



SADDAM HUSSAIN KHAN received the bachelor's degree from the University of Engineering and Technology (UET) Peshawar, the master's degree from the University of Engineering and Technology (UET) Taxila, and the Ph.D. degree from the Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad, Pakistan. He is currently an Assistant Professor with the Department of Computer System Engineering, University of Engineering and Applied Science (UEAS) Swat, Pakistan. His research interests include computer vision, deep neural networks, machine learning, medical image analysis, deep learning in cyber security, and deep reinforcement learning.



SULEMAN QAMAR received the B.S. degree (Hons.) in computer science from the University of Kotli, Azad Jammu and Kashmir (UoKAJK), and the M.S. degree in computer science from the Pakistan Institute of Engineering and Applied Sciences (PIEAS). His research interests include deep reinforcement learning, deep neural networks, autonomous navigation and tracking, swarm intelligence, biomedical informatics, and medical image analysis utilizing machine learning techniques.



MUHAMMAD WALEED KHAN received the B.E. degree (Hons.) in mechatronics engineering, in 2017, and the M.S. degree (Hons.) in systems engineering, in 2019. He is currently a Lecturer with the Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad, Pakistan. He is currently an Active Researcher in mechatronics system design control systems and artificial intelligence, as his areas of interest.



IQBAL MURTZA received the B.Sc. degree in mathematics and physics from GC University, Faisalabad, in 2006, and the M.Sc. and M.Phil. degrees from the Department of Electronics, Quaid-i-Azam University, Islamabad, in 2006 and 2011, respectively. He is currently pursuing the Ph.D. degree in computer vision with the Department of Computer and Information Sciences (DCIS), Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad, Pakistan. He is also an Assistant Professor with the Department of Creative Technologies, Faculty of Computing & Artificial Intelligence, Air University, Islamabad.



JEONGHWAN GWAK received the Ph.D. degree in machine learning and artificial intelligence from the Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2014. From 2002 to 2007, he worked for several companies and research institutes as a Researcher and a Chief Technician. From 2014 to 2016, he worked as a Postdoctoral Researcher at GIST, and from 2016 to 2017 as a Research Professor. From 2017 to 2019, he was a Research Professor with the Biomedical Research Institute and the Department of Radiology, Seoul National University Hospital, Seoul, South Korea. In 2019, he joined the Korea National University of Transportation (KNUT) as an Assistant Professor and since 2021, he has been an Associate Professor. He is currently the Director of the Applied Machine Intelligence Laboratory. His current research interests include deep learning, computer vision, image and video processing, AIoT, fuzzy sets and systems, evolutionary algorithms, optimization, and relevant applications of medical and visual surveillance systems.



ASIFULLAH KHAN has been the Head of the Department of Computer and Information Sciences, PIEAS, from 2016 to 2019, and currently the Head of the PIEAS AI Center. He has more than 21 years of research experience and is working as a Professor at PIEAS. He became a Full Professor, in 2016. In the field of machine learning and pattern recognition, he has 110 international journals, 53 conferences, and nine book-chapter publications with more than 6000 citations to his credit. He has successfully supervised 20 Ph.D. scholars so far and is on the panel of reviewers of 53 ISI international journals. He has won eight research grants as a Principal Investigator. His research interests include machine learning, deep neural networks, image processing, and pattern recognition.

He has been awarded the President's Award for Pride of Performance, in 2018. In addition, he has received four HEC's Outstanding Research Awards and one Best University Teachers Award. He has also received PAS-COMSTECH Prize 2011 in Computer Science & IT. He has received Research Productivity Awards from Pakistan Council for Science and Technology (PCST), in 2012, 2013, 2014, 2015, and 2016.

...