IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 12 July 2024, accepted 26 July 2024, date of publication 7 August 2024, date of current version 23 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3440064

SURVEY

Exploring Deep Learning-Based Visual Localization Techniques for UAVs in GPS-Denied Environments

OMAR Y. AL-JARRAH[1], AHMED S. SHATNAWI[1], MOHAMMAD M. SHURMAN[1], OMAR A. RAMADAN[2], AND SAMI MUHAIDAT[3], (Senior Member, IEEE)

[1]Department of Network Engineering and Security, Faculty of Computer and Information Technology, Jordan University of Science and Technology, Irbid 22110, Jordan
[2]Department of Mechanical Engineering, Faculty of Engineering, Jordan University of Science and Technology, Irbid 22110, Jordan
[3]KU 6G Research Center, Department of Computer and Information Engineering, Khalifa University, Abu Dhabi, United Arab Emirates

Corresponding author: Omar Y. Al-Jarrah (oyaljarrah1@just.edu.jo)

This work was supported by Jordan University of Science and Technology, Deanship of Research, under Grant 20220556.

ABSTRACT Unmanned Aerial Vehicles (UAVs) have proliferated across diverse domains. However, optimal UAV operations necessitate precise and reliable navigation systems. UAVs predominantly rely on the Global Navigation Satellite System (GNSS), such as the Global Positioning System (GPS), for navigation. Nevertheless, GNSS signals are susceptible to blockage, reflection, and spoofing, introducing significant risks, including navigation loss and potential UAV loss. This research investigates cutting-edge navigation solutions, emphasizing deep learning-based visual localization approaches tailored for UAVs. Our focus is on scenarios characterized by GPS-denied environments where GPS signals may be absent or unreliable. We provide a comprehensive review of contemporary deep learning-based visual localization approaches and compare them to traditional aerial visual localization methods, such as template matching and feature matching. This comparison highlights both the potential benefits and challenges associated with these approaches. Furthermore, we systematically evaluate and classify recent deep learning-based methods based on main criteria, including model type/architecture, reference imagery, operational context, and resultant accuracy levels. Our findings underscore the substantial promise inherent in various approaches while also shedding light on their unique deployment challenges. Finally, we discuss potential research directions, to inspire further innovations and progress in this domain. The ultimate goal is to develop more accurate, dependable, and secure navigation solutions for UAVs.

INDEX TERMS Deep learning, GPS-denied, localization, visual localization, navigation, UAVs.

I. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) have found wide-ranging applications in various fields, such as surveying and mapping, search and rescue, exploration and surveillance, inspection, payload transportation, firefighting, public security, border security, and object tracking. These applications require autonomous or semi-autonomous operations, which entail accurate navigation capability [1], [2], [3], [4].

Navigation solutions or localization systems can generally be categorized into two main types: absolute localization and relative localization. Absolute localization provides information about the position using a global or standard coordinate system, answering the question: "Where am I about a fixed, universally recognized point or system?". On the contrary, relative localization provides information about the position of a specific starting point or a previously known position, answering the question: "Where am I in comparison to where I was before or to a specific reference point?" [5].

The majority of existing navigation techniques rely on the integration of the Global Navigation Satellite System (GNSS) and Inertial Navigation Systems (INS) for position estimation [6], [7], [8]. GNSS is considered an absolute

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

navigation system, as it provides position data about a global coordinate system [9]. On the contrary, INS is considered a relative navigation system, as it tracks changes in position, velocity, and orientation from a known starting point. GNSS typically offers reliable long-term accuracy, although its short-term precision leaves room for improvement [10]. INS excels in providing high short-term accuracy; however, it struggles to maintain this accuracy over the long term due to the accumulation of errors. These errors originate from the numerical integration process used to convert acceleration measurements into displacement estimations. Consequently, the accuracy of The INS tends to degrade over time [11].

The combination of GNSS and INS systems yields an effective solution for position estimation. This integrated approach results in precise data, produced by cost-effective and lightweight sensors.

However, this approach can suffer from serious reliability issues such as blocked signals, multipath reception, and Non-Line-Of-Sight (NLOS) conditions [12], which can significantly degrade the system's accuracy as long as the signal is compromised or not received correctly.

Multipath reception refers to the phenomenon where one or more reflected signals from the satellite interfere with the direct signal. On the other hand, NLOS conditions occur when the direct signal from the satellite is blocked and the GNSS system receives a reflected version of the signal instead. These challenges underline the complexity of maintaining accuracy and reliability in GNSS-INS integrated navigation systems. Figure 1 illustrates the common GNSS reception issues such as blocked signals, NLOS signals, and reflected signals. Also, Figure 2 shows that GNSS-INS navigation systems may also be susceptible to signal spoofing, where an attacker can manipulate the system by usurping the signal. This manipulation may deceive the compromised system into believing it is in a different location, effectively allowing the attacker to control the UAV [13], [14]. Based on the above, it's clear that maintaining the integrity of signals in GNSS-INS systems is crucial to prevent such security breaches and ensure the safe and accurate operation of UAVs in sensitive applications.

Given the aforementioned challenges, researchers have started exploring vision-based and vision-aided localization techniques as alternative solutions to GNSS-INS systems. Vision-based localization in UAVs takes images captured by the onboard camera and extracts distinctive features used to estimate the position of the UAV. Vision-based localization systems for UAVs can be divided into Relative Visual Localization (RVL) systems and Absolute Visual Localization (AVL) [10]. RVL encompasses popular methods such as Visual Odometry (VO) and Simultaneous Localization and Mapping (SLAM). However, the main challenge with RVL is still the same as in the traditional methods discussed previously which is the accumulation of error, also known as drift, over time.

In their early stages of development, traditional RVL methods relied on handcrafted features and geometric techniques

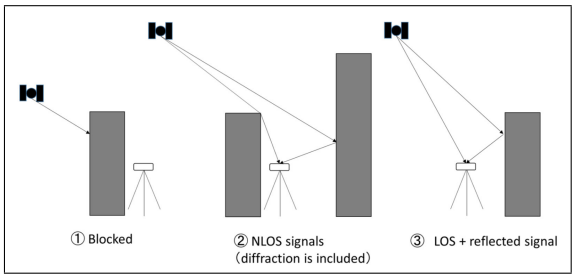

FIGURE 1. Examples of common GNSS reception issues [11].

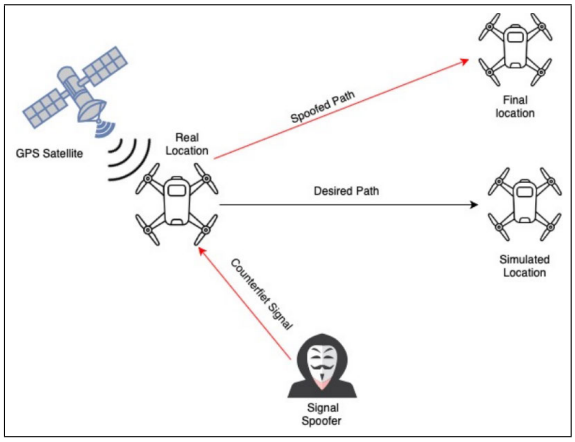

FIGURE 2. Example of GNSS reception issues: Signal spoofing [13].

to estimate the position and orientation of aerial vehicles using visual data. While these methods showed promise in some scenarios, they often struggled to handle diverse environments, occlusions, and varying lighting conditions effectively [15].

Therefore, the momentum of the research effort has shifted towards AVL methods, which are more immune to drifting errors observed in RVL methods [16]. Such methods utilize previously collected data (e.g., images from a past flight or satellite data) that are assumed to be precisely georeferenced to localize the UAV. The accuracy of the georeferenced data affects the accuracy of the localization [11]. Popular AVL methods include template matching, feature point matching, and most recently, deep learning-based methods. Often, these methods are integrated with VO or SLAM to enhance results by providing continuous position estimation.

Several reviews have been conducted in the field of visual localization, each contributing to the body of knowledge but with unique focal points that distinguish them from our current work. For instance, in their work, Lu et al. [17] published a review paper that does not concentrate on deep learning-based methods, creating a disparity between their scope and ours. Similarly, Leung and Shamwell [18] presented a comprehensive review of outdoor visual localization. Despite their thorough investigation, their research did not specifically cater to UAV localization. This lack of focus on UAV-specific challenges and perspectives limits its applicability to UAV-based applications, setting it apart

from our study. Couturier et al. published two notable papers in this field [10], [11]. The first paper focused primarily on relative localization methods, thus differing from our emphasis on absolute visual localization. Their second paper, which is closely aligned with ours, offered an exhaustive review of AVL methods, covering both traditional and deep learning-based approaches. Despite the similarity, this review significantly contributes to the literature by including more recent works and advancements in the field, extending the scope beyond 2020. Therefore, our review can be considered a complementary extension of Couturier et al.'s work, offering updated insights into the rapidly evolving domain of absolute aerial visual localization for UAVs.

The contributions of this work are as follows:

1) We provide an overview of traditional aerial visual localization methods used in GPS-denied environments, discussing their strengths and limitations.
2) We review and categorize state-of-the-art deep learning-based approaches used for various localization problems, offering a comprehensive summary of contemporary research in this field. This review covers the most recent work up to the date of publication.
3) We discuss and identify the challenges and current gaps in the domain of deep learning-based visual navigation in GPS-denied environments.
4) We suggest potential future research directions based on the current landscape of the work.

The remaining of this paper is organized as follows: Section II discusses traditional aerial visual localization methods for UAVs in Global Positioning System (GPS)-denied environments, focusing on template matching-based and feature matching-based methods, while Section III reviews and classifies deep learning-based aerial visual localization techniques for UAVs in GPS-denied environments. In Section V, we discuss the challenges facing the real-world deployment of these techniques. In Section VI, we present, identify, and discuss potential research directions in this field, and conclude this paper in Section VII.

II. TRADITIONAL METHODS FOR AERIAL VISUAL LOCALIZATION IN GPS-DENIED AREAS

This section reviews traditional aerial visual localization methods shown in Figure 3, focusing on template matching and feature matching methods, given their widespread applications and robust performance. It also thoroughly presents and examines the underlying principles, algorithms, characteristics, and challenges associated with these methods.

A. TEMPLATE-MATCHING-BASED METHODS

Template-matching, a.k.a, direct matching, is a method used in computer vision and image processing that relies on matching a sub-image within an input image with a reference template [8]. The process of template matching typically involves finding the similarity between the reference template and different regions within the input image, using mathematical formulas such as cross-correlation [19], [20], Sum of Squared Differences (SSD) [21], or normalized cross-correlation [22]. The region with the highest similarity score is considered the best match and its position is identified as the position of the matching area within the template. In aerial visual localization, the current view from the camera attached to the UAV is matched with a previously saved map or image, relying on the image patch comparison operator to obtain a measure of similarity among two image patches. Once a good match is found, the current absolute location will be estimated from the reference image.

Dalen et al. [23] presented a novel approach that estimates the absolute position of a drone in indoor and outdoor environments, using a Particle Filter (PF) in conjunction with a vision-based SLAM system. A key aspect of their methodology is the use of Normalized Cross-Correlation (NCC) for template matching, crucial for comparing onboard drone images with photographic maps. The NCC, defined by Lewis [24], is expressed as:

$$\gamma(u,v) = \frac{\sum_{x,y}[f(x,y)-\bar{f}_{x,y}][t(x-u,y-v)-\bar{t}]}{\sqrt{\sum_{x,y}[f(x,y)-\bar{f}_{x,y}]^2[t(x-u,y-v)-\bar{t}]^2}}$$

Here, $\gamma(u,v)$ is the NCC result, with 1 indicating a full positive correlation and -1 indicating a negative correlation. The pair (u,v) represents the location on the map f where the template t is matched. $\bar{t}$ and $\bar{f}$ denote the average intensities of the image patches that are being matched. Lewis suggests optimizing the NCC computation using the Fourier transform for the entire map. The NCC result is then converted into a Probability Density Function (PDF) for particle filtering, focusing on positive correlations.

The PF, preferred over the Extended Kalman Filter (EKF) due to its adaptability to non-Gaussian noise, involves steps such as initialization, propagation, measurement update, resampling, and statistical measures. The final position estimate and variance obtained from the PF are integrated into the EKF for navigation updates.

The effectiveness of this method was tested both in simulations and real-world flight tests. For outdoor testing, a Yamaha RMax helicopter (GTMax) was used, flying at 30.5 m altitude and 3 m/s speed, with Bing Maps as the global map and Google Maps images for simulation. Despite challenges due to mismatches between simulated and actual imagery, the system exhibited robust performance.

Real-world tests in Ft. Benning, Georgia, involved flights at 61 m altitude, using differential GPS for navigation and validation. The system showed an average position error of 3.6 m and a maximum error of 12.5 m compared to GPS positioning. Additionally, in a closed-loop test, the map alignment algorithm converged to the correct location within 50 seconds, with the PF updating the EKF's absolute location after 60 seconds. The error plots confirmed the position error within the 2σ bounds of the EKF-where σ refers to the standard deviation in the statistical analysis-.

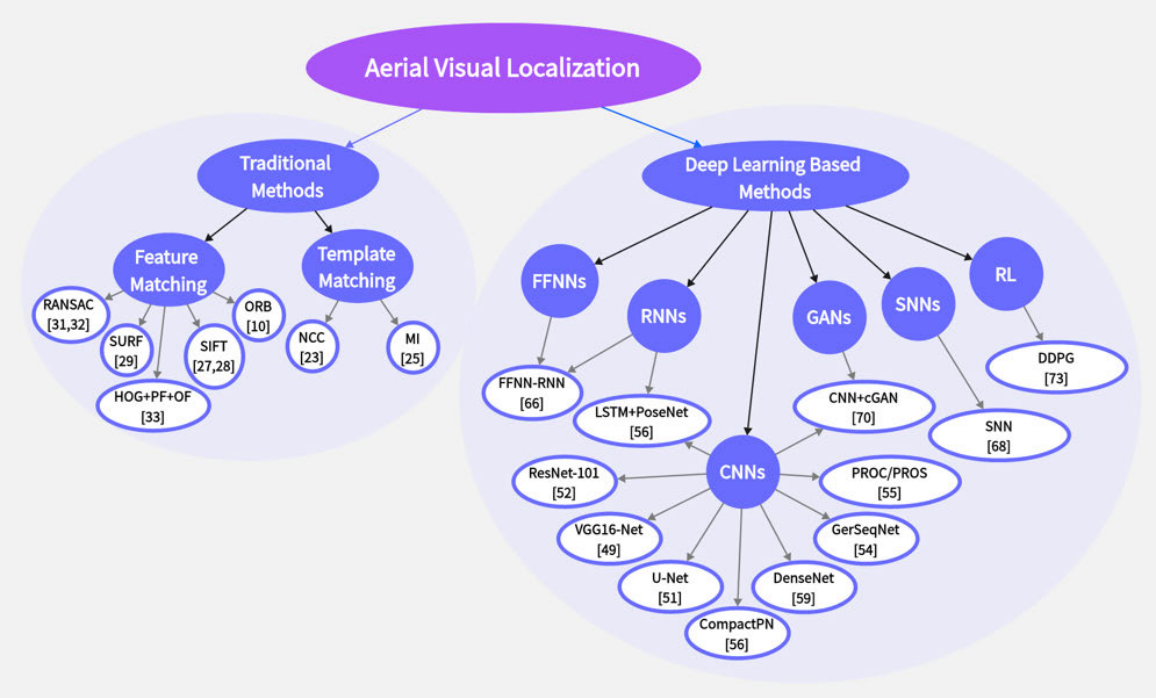

FIGURE 3. Classification of visual localization methods.

Overall, this system provides accurate absolute position estimates for drone navigation in both indoor and outdoor environments, demonstrating its robustness and accuracy. It is designed to function without absolute visual localization but can leverage it when available, enhancing the state estimate.

Yol et al. [25] introduced a novel method for the localization of UAVs using vision processing only. They developed an image registration technique based on Mutual Information (MI), which effectively determines the two-dimensional motion parameters of the UAV between a current image and a reference image. This approach ensures robustness against variations in lighting conditions and environmental changes. It also enhances the algorithm's efficiency as the derivatives related to the displacement parameters are calculated using the reference image's gradient, allowing these derivatives to be precomputed. The testing environment included flight-test data acquired with a GoPro camera mounted on a hexacopter UAV. The flight data was collected over the Université de Rennes 1 campus, with the UAV positioned approximately 150 meters high. The UAV's geographical localization and altitude were logged through an embedded GPS, which was used solely as a ''ground truth'' to validate the computed localization. The researchers used a georeferenced mosaic extracted from Google Earth as a reference for image matching, simulating conditions of the real world. Additionally, a motorized Nadir system was utilized, wherein the UAV's pitch and roll motions were counterbalanced via a brushless gimbal. The study demonstrated impressive accuracy, with the estimated trajectory closely aligning with the ground-truth data gathered via GPS. Specifically, the results exhibited a Root Mean Square Error (RMSE) of 6.56m in latitude, 8.02m in longitude, and 7.44m in altitude. Although the method demonstrated robustness in challenging conditions, including seasonal changes and significant lighting variations, it exhibited errors stemming from perspective effects caused by the UAV's altitude. These errors have the potential to impact the navigation process. Yol et al.'s research notably contributes to UAV localization using vision-based techniques, showing promising results in real-world testing. Subsequent efforts may involve integrating this approach into a comprehensive global estimation framework incorporating inertial measurement units (IMU) data. This would enable real-time onboard localization in case adequate CPU power is available on the UAV.

B. FEATURE-MATCHING-BASED METHODS

Feature point matching is a computer vision technique that refers to recognizing points of interest (features) of an object across images with slightly different viewpoints. These points of interest are extracted from images and represented using a descriptor, which describes the local appearance of the feature. The goal of feature point matching is to match these features between images, or within a single image and use the matched features to estimate the relative position and orientation of the images, or to perform image stitching, object recognition, or other computer vision tasks [26].

There are various algorithms for feature extraction, description, and matching, including Scale-Invariant Feature Transform (SIFT) [27], [28], Speeded Up Robust Features (SURF) [29], Oriented FAST and Rotated BRIEF (ORB)

[30], Harris corner detection, and Random Sample Consensus (RANSAC) [31], [32]. Feature point matching is considered an efficient alternative method to template matching as it is faster and requires less data storage capacity, allowing a larger area to be covered in the localization process. In aerial visual localization, the feature points are extracted from a real flight image and compared with features extracted from images stored in a database. This database is made by processing the referenced images that usually have differences in illumination, scale, rotation, and viewpoint.

Couturier and Akhloufi [10] presented an RVL system, that leverages feature point detection during in-flight missions, activated when the GNSS signal is unavailable. The system employs a monocular camera to detect and track visual features in the surrounding environment, allowing estimation of the UAV's motion relative to these features. The system underwent testing in diverse scenarios, considering various feature extraction methods. Among these methods, the ORB (Oriented FAST and Rotated BRIEF) demonstrated superior performance with the lowest mean absolute error. The estimated average distance between the best match and ground-truth localization was around 70 meters at an altitude of 150 meters. Although originally presented as an RVL solution, this approach could also be regarded as an AVL solution since it utilizes pre-existing maps for referencing.

Shan et al. [33] developed a framework that combines Histogram of Oriented Gradient (HOG), PF, and Optical Flow (OF). The framework was presented for Google Maps-aided UAV navigation in a GPS-denied environment. HOG features were used for registration on Google Maps. PF was used in the matching process to avoid sliding window searches. For efficiency, the search is confined around the UAV location predicted by OF. Experiments were performed with real flight data from a UAV operating in a 40 m $\times$ 225 m environment. An RMSE of 6.77 m was obtained in the results. Feature matching methods are robust to environmental factors and can handle large changes in scale and viewpoint, making them well-suited for aerial visual localization tasks. Template matching methods are simple and efficient, but they are sensitive to changes in scale, rotation, and illumination, which can limit their performance in real-world applications. Depending on the specific requirements of a given task, one method may be more suitable than the other, or a combination of both methods may be used to achieve the best results.

III. DEEP LEARNING-BASED METHODS IN AERIAL VISUAL LOCALIZATION

Deep learning is a subset of machine learning that uses neural networks, inspired by the human brain, to analyze and process vast amounts of data. Unlike traditional machine learning which works best with structured data, deep learning excels with both structured and unstructured data [34]. Neural networks consist of an input layer, multiple hidden layers, and an output layer. Data flows through these layers, getting processed using weights, biases, and activation functions. The primary goal is to adjust these weights and biases to minimize the error in predictions, known as the cost function. This adjustment is achieved through a process called backpropagation, where the system learns from its mistakes. Deep learning has various applications, from recognizing images and speech to advanced predictive analytics [35]. It's the backbone of many modern AI-driven solutions, enhancing accuracy and efficiency in numerous domains.

Deep learning has become a powerful tool in various fields, including aerial visual localization for UAVs. Traditional methods rely on hand-crafted features and models that are pre-designed to extract certain characteristics from images. Deep learning offers an end-to-end approach that can automatically learn the relevant features directly from the raw images which makes it increasingly popular in this field. The advantage of deep learning-based AVL is that it can provide a more robust and accurate solution compared to traditional visual localization techniques, especially in challenging environments, such as low-light or cluttered scenes [36]. However, the performance of deep learning-based AVL depends heavily on the quality and variety of the training data. Optimally, the dataset should be collected from the environment where the solutions are intended to be deployed. Otherwise, if testing data significantly differs from the training data, which is known as data drift, the performance of the solution will deteriorate [37].

In the context of AVL, various deep learning techniques, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and their variants, have been explored to achieve state-of-the-art results in aerial visual localization for UAVs [38]. In this section, we will review the recent deep learning-based methods and approaches adopted in the aerial visual localization field.

A. CNNs-BASED METHODS

CNNs are a type of deep learning neural network that is commonly applied to analyze visual imagery, detect objects, and separate different parts of an input image. Typical CNNs use special filters to identify specific patterns and shapes within an image, that are used to recognize objects in the input image [39]. CNNs have proven to be very effective in computer vision applications and are used in a variety of tasks, including facial recognition [40], [41], [42], self-driving cars [43], [44], [45], and medical image analysis [46], [47], [48]. They are particularly well-suited for tasks like object detection, classification, and localization because they can recognize patterns and features in images.

In line with the utility of CNNs as an AVL approach, Amer et al. [49] proposed an aerial visual localization method, particularly in GPS-denied urban environments for UAV applications. They innovatively employed a CNN model, leveraging the foundational architecture of a pre-trained VGG16-Net [50]. This approach was pivotal in demonstrating the practical application of CNNs for complex image analysis tasks necessary for drone localization. They gathered an extensive collection of geotagged images from

urban areas and trained the model to recognize distinct features of the environment, such as building facades, roads, and intersections. The innovative aspect of their approach was in training the CNN model to not only recognize these features but also to extract features from these images to create what they called ''deep urban signatures'' which assist the drones in navigation and precise self-localization within complex urban landscapes. The deep urban signatures are then used in a matching mechanism to match the signatures of images captured in real-time by a drone's onboard camera with the signatures of those in the pre-trained model database, enabling accurate estimation of the drone's location. Several experiments were conducted to evaluate and compare its performance with traditional feature-based approaches, using a dataset of images captured by a drone flying over a large urban area. A commercially available drone, specifically the Phantom DJI model, equipped with an integrated camera featuring a 94-degree field of view was used to conduct several experiments at various altitudes, with a maximum altitude of 333m. Experiments results show that the proposed approach outperforms the traditional approach in terms of accuracy and computational efficiency, demonstrating the potential of deep learning techniques for improving drone localization in urban environments. The model was able to discriminate between 7 different districts with an average accuracy of 91.2%. For neighborhood retrieval, an overall localization error of 200.75 meters in 6 districts has been achieved.

Advancing the application of CNNs even further, Nassar et al. [51] presented a framework that uses conventional computer vision methodologies with CNNs to ensure precise localization in GPS-denied environments, a significant challenge in urban UAV navigation. The framework begins with a calibration stage where a SIFT detects the key points in the input image, while RANSAC estimates a 'homography matrix', which illustrates the transformation between the drone's view and a reference satellite map. This stage is crucial as it aligns the drone's perspective with the map. Following calibration, the framework engages in sequential frame registration. Here, it uses the ORB algorithm to process frames at regular intervals. This stage continuously updates the calibrated reference map, keeping the drone's perspective in sync with the map. A critical advancement is the utilization of the U-Net algorithm to perform 'semantic segmentation'. This phase is designed to extract meaningful shape information from both UAV-captured and satellite images, focusing on identifiable urban features like buildings and roads. The segmented elements are further refined through morphological operations, ensuring cleaner and more defined feature extraction. The final and most critical phase involves Semantic Shape Matching (SSM). This step matches the semantically segmented shapes from the UAV and satellite images, employing a scoring system based on shape features like area, location, and orientation. The framework uses these matched shapes to compute a refined homography, adjusting the UAV's current location with improved accuracy.

The proposed framework was tested on two datasets, achieving an impressive average geolocation error of just over 10.4m and 6.3m, respectively. The incorporation of deep learning for semantic segmentation further reduced the error rate to around 5.1m and 3.6m. One limitation of the system, though, is its reliance on known urban areas for optimal performance. However, it's worth noting that the calibration process assumes that the transformation between the two images can be adequately described by a homography. This is a reasonable assumption when the scene being imaged is flat, or the drone is high above the ground so that any depth differences in the scene can be ignored. However, if there are significant depth differences (like in mountainous terrain), the transformation may be more complex and not adequately described by a single homography matrix. Overall, Nassar et al.'s approach represents a substantial advancement in the field of UAV localization in GPS-denied environments, effectively combining the precision of deep learning with the robustness of traditional computer vision techniques.

Mugal et al. [52] introduced a visual localization method designed for UAVs in GPS-denied environments. Their system integrates onboard camera footage with archived georeferenced imagery, leveraging deep learning to pinpoint the UAV's location within a pre-stored ortho-mosaic map with remarkable accuracy. The core of their approach lies in the development of a network that employs a neighborhood consensus technique. This method enhances feature point matches between the aerial image and the stored orthomosaic. Utilizing ResNet-101, a deep CNN known for its robust feature extraction capabilities, the network extracts convolutional features from the images. The next step in their process involves constructing a correlation matrix that meticulously captures feature matches for each pinpointed feature point. The ingenuity of their model is further exemplified in the subsequent step, where a trainable network refines the reliability of correspondences. This is achieved by applying probabilistic constraints, and meticulously aligning each feature point from the source image with corresponding points in the orthomosaic. A novel aspect of their system is the integration of a soft-argmax layer, which intelligently identifies the best match indices. These indices are then processed through a fully connected network, culminating in the precise determination of point correlations. Mughal et al. also incorporates RANSAC to establish direct correspondences between images, thereby bolstering the durability and precision of their vision-based UAV localization model. This approach allows for the accurate identification of the UAV's position on the map through its camera feed. The training of their model is conducted on a set of template images, each meticulously labeled to signify correspondences with the relevant orthomosaic. Data augmentation techniques, including image transformations, are employed to enhance the model's robustness and generalizability. The training process, spanning 185 epochs with early stopping at a learning rate of 0.0008, leverages stochastic gradient descent with momentum for optimization.

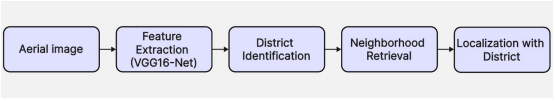

FIGURE 4. Step-wise procedure of the methodology proposed by Amer et al. [49].

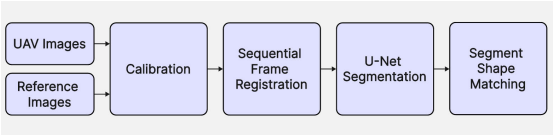

FIGURE 5. The procedure of the presented methodology by Nassar et al. [51].

Notably, their system exhibited a minor decrease in matching accuracy when tested with a different UAV from the one used for data collection, indicating a high degree of specificity to the training conditions. The mean error recorded was 3.594 m, with a maximum error of 31.281 m, demonstrating the system's overall effectiveness in real-world scenarios. In conclusion, Mughal et al.'s research presents an approach to visual-based UAV localization in GPS-denied areas. By harnessing the power of deep learning and integrating efficient outlier detection methods like RANSAC, they have devised a system capable of accurately determining UAV positions using aerial images. Despite a slight variation in accuracy under different operational conditions, the system's performance is remarkable, highlighting its potential for real-world application. Furthermore, the team's contribution extends beyond their innovative methodology; they have generously made available their source code and an original dataset, consisting of 2052 high-definition aerial images collected from diverse flights across Pakistan, covering an area of approximately 2 km^2.

In their study, Goforth and Lucey [53] developed a method that utilizes a monocular RGB camera mounted on the UAV and leverages pre-existing satellite imagery for localization. The method stands out for its use of CNN representations, trained on readily available satellite data, to address the challenges posed by differences in image-capturing conditions such as seasonal and perspective changes. A key aspect of their methodology is the development of an optimization process that jointly minimizes errors between adjacent UAV frames and the satellite map. This optimization significantly increases the localization accuracy, especially in environments with few landmarks. The method, therefore, demonstrates improved performance over recent systems, achieving an average localization error of less than 8 meters in a GPS-denied flight at an altitude of 0.2km and over a distance of 0.85km. The approach involves three primary steps; The First step is Visual Odometry, where the Initial motion parameter estimates are derived using visual odometry, a crucial step for determining the UAV's trajectory and orientation. The second step is the satellite map comparison where a subset of recent frames is compared to the satellite map to geolocalize those frames. The CNN representations play a vital role in aligning the UAV imagery with the satellite images under varying conditions. The last step is joint optimization which is the most important step where the geolocalized pose of all UAV frames is refined through a joint optimization process. This process considers both frame odometry and map alignment, crucial for accurate localization across all UAV frames, even those not directly compared with the map. By integrating deep learning with satellite imagery and employing an innovative optimization process, they have crafted a method that is not only accurate but also capable of generalizing across different environments, from urban to rural settings. Their work underscores the potential of advanced machine learning techniques in enhancing UAV navigation and localization where conventional GPS systems fall short.

Arturo and Martinez [54] presents a groundbreaking method for UAV localization in the challenging environment of autonomous drone racing, an area that requires high-speed and precise maneuvering. Their novel approach, named 'GreySeqNet,' is a compact CNN that significantly differs from traditional methods by using a sequence of grey-scale images rather than color images as input. The core innovation of their method lies in the network architecture, which is based on the Inception model. This architecture is adept at handling different views of the same object and effectively extracting necessary features for localization. GreySeqNet processes a stack of three consecutive grey-scale images, providing a more comprehensive view of the scene's dynamics while reducing computational load. One of the key technical advancements of their approach is the separation of the final layer of feature extraction for each axis (x, y, z), allowing the network to specialize in capturing specific displacement features for each axis. This architectural decision enhances the network's ability to generalize pose estimation across different racing scenarios. The training of GreySeqNet involves a novel dataset generated within the Gazebo simulator, offering a rich and varied set of aerial images that mimic real-world racing conditions. This extensive training ensures the model's robustness and accuracy in predicting the UAV's pose. Arturo and Martinez's system was rigorously tested in a simulated racing environment. The experiments demonstrated that GreySeqNet achieves a high-frequency operation (up to 83 Hz on GPU) with an average camera pose error of around 31 cm, making it highly suitable for real-time applications in drone racing. In addition to its core functionality, the model employs a Kalman Filter with a constant velocity model to smooth out the predictions, contributing to more stable and accurate drone flight trajectories. This integration of advanced deep learning techniques with practical filtering methods represents a significant step forward in the field of autonomous drone localization, particularly in GPS-denied scenarios like drone racing. While the study by Arturo and Martinez introduces a technically advanced and practical approach for UAV localization in the context of autonomous

drone racing, it is important to note that their focus primarily lies on processing grey-scale image sequences for high-speed navigation, rather than aligning visual data with a reference for global positioning. This distinction means that, although innovative, their methodology does not directly address the specific challenges of global UAV localization in GPS-denied environments as per our primary topic of interest. However, the underlying principles and techniques they've developed could be adapted or expanded upon to make their approach more relevant to our area of focus, particularly by incorporating mechanisms for reference-based global positioning.

Relying on the coordinates or sizes of reference objects, Pi et al. [55] introduced two mapping approaches namely, projection from Perspective to orthogonal based on Reference Objects' Coordinates (PROC) and projection from Perspective to orthogonal based on Reference Objects' Size (PROS). These approaches utilize two CNN models, Model-P and Model-O, which handle different views, Perspective View (PV) and Orthogonal View (OV), respectively. The views are derived from input images captured by an onboard RGB camera, respectively. Testing of the approaches demonstrated that Model-P achieved a 97% mean Average Precision (mAP), showing superior performance compared to Model-O, which achieved a 51% mAP. This difference can be attributed to the similarity of the Targets of Interest (ToIs) and objects' appearances in the orthogonal view compared to the perspective view. The projection results obtained from the PROC and PROS approaches showed that both methods achieved Average Projection Errors (APEs) as small as 11.85 inches for PROS and 13.86 inches for PROC. These results indicate that the proposed methods are feasible solutions for real-time localization and mapping using only RGB camera inputs. However, they rely on single UAV video input and need reference objects on the ground, which might not be available in unknown locations or territories. Future research may explore using other visual inputs (e.g., thermal imagery) and creating ad-hoc reference points using multiple cooperative UAVs. It is worth noting that the proposed methods allow for the localization and mapping of ToIs from the RGB camera's inputs, without explicitly estimating the UAV's position.

Cabrera-Ponce et al. [56] introduced an innovative approach for geolocation using CNNs on aerial images captured by drones, aiming to accurately deduce GPS positions from these visuals. This method stands out for its focus on efficiency and accuracy, employing streamlined CNN architectures that balance rapid inference speeds with predictive reliability. The study rigorously compares this new approach against leading models such as PoseNet [57], PoseNet + LSTM, VGG16, and ResNet-50 [58], demonstrating significant advancements in processing speed and accuracy. By optimizing the CNN architecture, unnecessary layers are discarded, and the Fully Connected (FC) sections are minimized, enhancing the model's geolocation capabilities without compromising performance. A pivotal aspect of their research involves the experimental design and the use of diverse datasets gathered at different elevations. This methodological choice not only tests the model's robustness across various terrain types but also its adaptability to changing altitudes, a critical factor for real-world UAV navigation. The detailed experimentation, conducted on Ubuntu 16.06 with PyTorch 1.1.0, OpenCV 3, and CUDA 9.0 on an Nvidia GeForce GTX 960M and managed through ROS, offers profound insights into the model's performance. The results of the study, featuring a prediction error range of 2.8 to 6.1 meters and an impressive processing speed of 103 frames per second, significantly outperform the fastest contemporary models, which achieve 69 fps. This showcases the model's potential to revolutionize drone-based geolocation tasks by offering a quicker and more accurate alternative to existing solutions. Furthermore, the study underscores the practical applications and future directions of drone technology, especially in scenarios where GPS signals are unreliable or unavailable. However, the research also prompts a critical evaluation of potential limitations, such as the model's reliance on clear visual cues and computational demands, which could challenge deployment on drones with limited processing capabilities. These considerations are crucial for the practical application and scalability of UAV solutions. In summary, the work of Cabrera-Ponce et al. marks a significant contribution to UAV navigation and autonomous operations, demonstrating the feasibility and benefits of using CNNs for drone-based geolocation. It emphasizes the importance of efficiency, adaptability, and practical applicability in drone technology advancements, paving the way for future innovations that could transform UAV operations across various sectors by enabling more reliable and independent navigation capabilities.

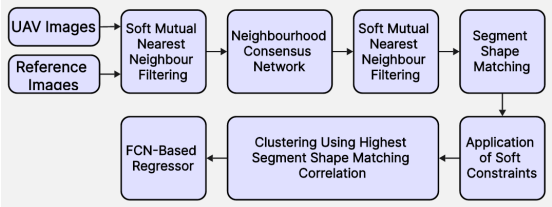

FIGURE 6. The components of Mugal et al. [52] proposed framework.

Cao et al. [59] presented an innovative approach to the visual localization of UAVs in satellite remote sensing images. Their method leverages an enhanced DenseNet [60] CNN model and a quality-aware template matching technique. In terms of methodology, the researchers utilized the University-1652 dataset [61], which uniquely contains both UAV and satellite view images of various scenes. To further improve the training and adaptability of the CNN to UAV localization, they collected additional UAV images using a DJI Phantom 4 Pro V2.0 from Nanjing University of Science and Technology. The feature extraction was carried out via four CNNs: AlexNet [62], VGG [50], ResNet [58], and DenseNet [60]. From comparing these

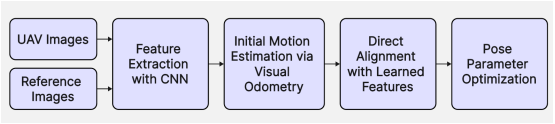

FIGURE 7. The components of Goforth and Lucey [53] proposed framework.

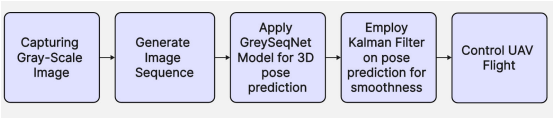

FIGURE 8. The components of Arturo and Martinez [54] proposed framework where it's based on pose estimation performed by a compact CNN.

networks, DenseNet proved superior in feature extraction and localization and thus was selected for further enhancements. The authors opted for DenseNet121 to mitigate overfitting, which presents features at different levels. The transition layer of DenseNet121 was used for feature extraction due to its ability to compress the model and its suitability for matching small images. The researchers introduced a multi-scale feature fusion technique, fusing features from Transition Layers 2 and 3 of DenseNet121. This fusion was shown to improve the overall UAV localization performance. The team trained their improved DenseNet121 network on their dataset for 300 epochs, using a cross-entropy loss function and a Stochastic Gradient Descent (SGD) optimizer with a momentum value of 0.9. The experimental results showcased that the improved method significantly increased the localization accuracy of the UAVs, with the error largely within 5 meters. Compared to traditional template matching methods like NCC, TM_SQDIFF, and TM_SQDIFF_NORMED, their method demonstrated similar or superior accuracy but significantly faster detection speed. The authors concluded that their approach of using an enhanced DenseNet model and a quality-aware template matching technique effectively allows for UAV localization in large-scale satellite remote sensing images, offering notable advantages for feature extraction and localization accuracy.

Notably, certain studies do not primarily concentrate on aerial visual localization to determine the UAV's position directly. Instead, these studies adopt analogous methods to trace specific trajectories or fulfill distinct applications that hinge upon the precise positioning of an object, such as deliveries using UAVs. One such study was made by Amer et al. [63]. They focus on using visual inputs to navigate a UAV along a specific trajectory, particularly in GPS-denied situations. The purpose of their research is not directly about localizing the UAV in standard coordinates but rather about maintaining its desired path using visual cues when GPS data is unavailable. However, tracking a trajectory does involve a form of localization, as the UAV needs to know where it is relative to its intended path. However, this type of localization is usually more about the relative position (i.e., am I on the path or not?) rather than the absolute position (i.e., what are my exact coordinates?). However, it can be modified to be a complete visual localization solution.

The core of their methodology lies in the amalgamation of deep CNNs and regression models to guide drone steering commands. This approach relies exclusively on visual data from an onboard camera. In one of their key works, the team utilized a pre-trained VGG-16 network to extract features from the visual inputs. These features were then fed into a Fully Connected Neural Network (FCNN) or a Recurrent Gated Neural Network (GRU), which functioned as the regressor, predicting the drone's yaw angle for navigation.

Amer et al. enhanced their model's resilience to drift and variability in starting points by introducing the concept of flight path augmentation. This technique involved the creation of multiple auxiliary navigation paths, which were slightly deviated versions of the optimal path. These paths formed a 'navigation envelope' for training the model. For training data collection, the team employed synthetic environments generated using the Unreal Engine with the AirSim plugin. They created and trained on multiple paths in abstract and complex realistic scenarios. Each path was augmented with noise for robustness. The training was conditioned on path starting points and visual feedback, leading to improved model performance and minimizing conflicting decisions during autonomous navigation.

The team's experiments showcased the superior performance of the FCNN as a regressor compared to the GRU, achieving lower error rates in both abstract and realistic environments. Specifically, their approach realized an average of 1.37 meters cross-track distance across four paths in simulated environments, highlighting their methodology's robustness and potential real-world applicability. Despite the promising results, Amer's research also indicates some areas for future work. As the current model was trained and tested only in simulated environments, real-world testing and validation are essential next steps. Further enhancements could also include the use of Generative Adversarial Networks (GANs) for style transfer between synthetic and real training images, which could enhance system performance and reduce the size of the training dataset.

On the other hand, Luo et al. [64] delves deeper into precise localization in GPS-denied or GPS-unreliable environments. They propose a framework, KeepEdge, which leverages visual information and edge computing to enhance the localization accuracy of a UAV for parcel delivery. This involves following a given trajectory and identifying the exact delivery location. This is a form of absolute positioning, where the UAV can determine its location in a standard coordinate system. In the proposed system, the UAV takes photos and sends them to an edge server. The edge server, equipped with a deep learning model, interprets the images, identifies the target delivery position, and sends back the coordinates to the UAV. This allows the UAV to adjust its flight path and accurately pinpoint the delivery location. To overcome the issue of resource constraints on UAVs,

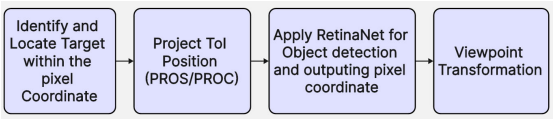

FIGURE 9. The components of Pi et al. [55] proposed framework.

they adopt a teacher-student paradigm for model training. A larger deep learning model (the teacher) is trained in the cloud, and knowledge distillation is used to create a smaller, more efficient model (the student) that is deployed on the edge server. The results indicate that the proposed solution significantly improves the accuracy of UAV deliveries. This is especially noticeable in complex environments with weak or unreliable GPS signals. The authors provide a comparative analysis of their solution with other photo-matching methods, illustrating the robustness and efficiency of their system. However, they have yet to mention the accuracy of the estimated position directly, which makes it incomparable with the other work mentioned in the survey.

Zhuofan Cui et al. [65] introduces a method for geo-localizing UAVs. The core of their approach is the use of a Vision Transformer, a type of neural network that excels at processing image data. This model is particularly adept at capturing both local and global information, which is essential for the accurate identification and matching necessary in geo-localization tasks. A central feature of their method is the adoption of cross-view consistent attention. This technique ensures the model prioritizes similar features in images taken from different viewpoints (UAV and satellite), which is crucial for aligning and accurately geo-localizing features across diverse images. The authors streamline model training with a single-stage approach, a departure from the conventional multi-stage training methods. This simplification reduces the computational resources required and enhances the model's efficiency in training and deployment. Another key element of their methodology is the implementation of a piecewise soft-margin triplet loss function. This function improves the model's capability to differentiate between positive and negative examples within the dataset, which enhances the accuracy of image matching. Additionally, the methodology employs a color transfer technique to mitigate color inconsistencies between UAV and satellite images. By adjusting for these color differences, the model is better able to concentrate on structural features rather than being misled by color variations, which is essential for accurate image comparison and matching. This methodology enhances the accuracy and efficiency of geo-localizing images from varied sources and perspectives, leveraging an advanced neural network architecture and innovative techniques in attention mechanisms, loss functions, and color consistency.

The study reports high accuracy levels, quantified by metrics such as Recall (R1) and Average Precision (AP). Using Euclidean distance, the method achieves an R1 score of 91.5% and an AP score of 93.31%. With cosine similarity,

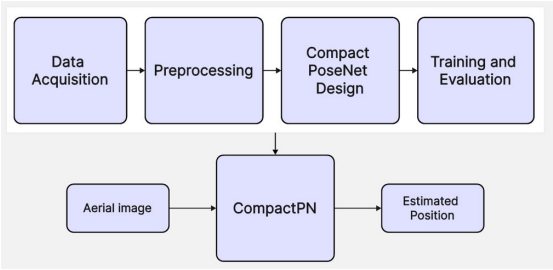

FIGURE 10. The components of Cabrera-Ponce et al. [56] proposed framework.

the R1 score slightly decreases to 86.60%, and the AP to 88.78%. These results underscore the method's efficacy in image retrieval and geo-localization tasks but unfortunately, it does not provide error measurements in meters, focusing instead on the precision of image retrieval and matching. As Cao et al. [59], they used the benchmark University-1652 dataset [61] to utilize and test the performance of their methodology.

The diverse approaches to CNN-based methods for aerial visual localization as detailed in this section underscore the dynamic intersection of deep learning with UAV technology, highlighting a robust frontier for innovation in navigating and understanding complex environments. These methods, ranging from leveraging pre-trained networks for urban drone navigation to the innovative use of vision transformers for geo-localization, collectively illustrate the depth and breadth of current research endeavors aimed at enhancing UAV capabilities in GPS-denied settings.

A central theme across these methodologies is the pursuit of accuracy and computational efficiency in localization tasks. This is evident in the efforts to integrate CNNs with traditional computer vision techniques, as seen in Nassar et al. framework [51], which employs semantic segmentation and shape matching for precise geolocation. Similarly, the development of compact CNN architectures, such as Arturo and Martinez's [54] GreySeqNet, highlights a focus on optimizing performance for real-time applications, addressing the critical need for speed and efficiency in scenarios like autonomous drone racing.

Moreover, these studies emphasize the importance of adaptability and robustness in UAV localization systems. Mughal et al.'s approach [52], which uses a neighborhood consensus technique, and Cabrera-Ponce et al.'s streamlined CNN models [56] for drone-based geolocation, demonstrate innovative strategies for enhancing model generalizability and reliability across varied operational conditions and environments. These advancements suggest a growing recognition of the complex, often unpredictable nature of real-world scenarios faced by UAVs, driving research toward more versatile and resilient solutions.

However, the discussion also highlights significant challenges and areas for future research. One recurring issue is the dependency on specific environmental features or datasets, which may limit the applicability of certain models in diverse

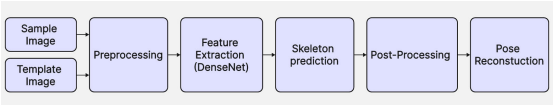

FIGURE 11. The components of Cao et al. [59] proposed framework.

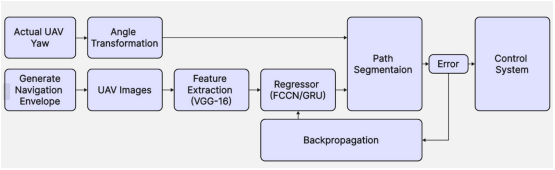

FIGURE 12. The components of Amer et al. [63] proposed framework.

or previously unmapped areas. This underscores the need for further advancements in feature extraction, transfer learning, and unsupervised learning techniques to broaden the scope of CNN-based localization methods.

Additionally, the exploration of novel neural network architectures, such as the Vision Transformer used by Cui et al. [65], points to an ongoing evolution in the tools and techniques available for aerial image analysis. These developments hold promise for addressing the limitations of current models, potentially leading to breakthroughs in how UAVs navigate and interact with their surroundings.

In conclusion, the exploration of CNN-based methods for aerial visual localization presents a vibrant area of research with significant implications for UAV technology. The aforementioned advancements not only demonstrate the current state of the art but also pave the way for future innovations. As these methodologies continue to evolve, they offer the potential to transform UAV operations, enhancing the autonomy, precision, and safety of drones across a myriad of applications, from urban navigation and autonomous delivery to disaster response and environmental monitoring. The ongoing fusion of deep learning with UAV technology heralds a future where drones can navigate and understand the world with unprecedented accuracy and efficiency.

B. RNNs-BASED METHODS

RNNs have emerged as powerful tools for processing time-varying inputs such as image sequences. The capability of RNNs to capture sequential dependencies and context information in input data makes them particularly useful for tasks that rely on specific patterns between images in a sequence, such as localization. RNNs have been employed in various localization tasks, including drone navigation, object tracking, and creating 3D models (e.g., 3D maps of the environment). However, RNNs can be slow and challenging to train when working with long sequences of images. To enhance performance, researchers have developed hybrid models that combine RNNs with CNNs or other algorithms. In this context, Zahedi et al. [66] proposed a solution to the problem of moving target localization and tracking using historical information generated by UAVs. The proposed solution combines two different types of Neural Networks (NNs) in a two-stage processing framework. In the first stage, a Feed-Forward Neural Network (FFNN) is utilized to learn a mapping function from the UAV, gimbal, and camera information to the targets' x- and y-locations. This network comprises two hidden layers with Rectified Linear Unit (ReLU) activation functions, which a piecewise linear functions that will output the input directly if it is positive, otherwise, it will output zero, and a linear output layer. The input to this model is a vector of 11 features and the output is the mapped x-location and y-location of a target. Generated data from a MATLAB simulation environment were used to train a model with 50 epochs, using MAE as a loss function. Adam is the optimizer and has a learning rate of 0.001. In the second stage, a sequence-to-sequence RNN is used to predict the targets' location while they are out of the FOV of the UAV. The RNN model consists of two layers, each with 200 LSTM cells, and a dense layer in between. The input to this model includes the mapped location from the previous phase, along with historical location data. The output is the predicted future location of the target. The research collected data for 400 targets over 200 seconds of simulation time. They collected 1000 data points for each target, as the camera sensor updates every 0.2 seconds. They randomly selected 300 targets for training their NN models and 100 targets for testing. A 4-folds cross-validation strategy was used to validate their models. The result showed that the proposed FFNN model outperformed the traditional Kalman filter for location mapping in terms of both accuracy and stability. Furthermore, the RNN model maintained relatively stable performance across different time steps for location prediction. At the same time, the error of the baselines continued to increase as the number of future locations to predict increased.

In the previous study Cabrera-Ponce et al. [56] mentioned before. The authors used an RNN, specifically a Long Short-Term Memory (LSTM) model, in combination with PoseNet, a CNN architecture. This combination was used as a benchmark to compare against their proposed CompactPN models. PoseNet is a CNN architecture that was originally designed to estimate the 3D position and orientation of the camera in a scene from a single image. The LSTM model was used to account for temporal dependencies in the sequences of images, with the idea being that the estimated pose for a given image is likely to be similar to the estimated pose for the previous image in the sequence. Despite the theoretical advantage of using an LSTM to account for temporal dependencies, Cabrera-Ponce et al. found that their CompactPN models, which are simplified versions of PoseNet and do not use any form of RNN, were able to perform similarly in terms of accuracy while providing significantly faster inference speeds. In particular, their CompactPN-2 model achieved the best balance between accuracy and speed among their proposed models. In summary, while the PoseNet + LSTM model might slightly improve accuracy, the CompactPN

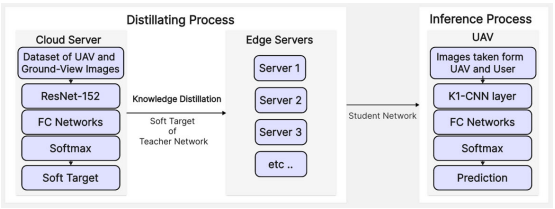

FIGURE 13. The components of Luo et al. [64] proposed framework.

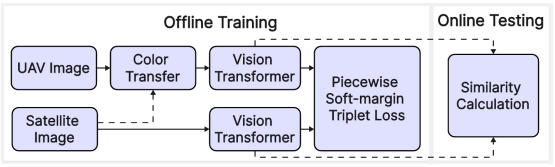

FIGURE 14. The components of Cui et al. [65] proposed framework.

models are much faster and thus more suitable for real-time applications. The study demonstrated that a simpler, more compact architecture could still deliver comparable results, which is beneficial when considering computational constraints in real-world applications such as UAVs.

C. SNNs-BASED METHODS

A Siamese Neural Network (SNN) is a unique class of neural network architectures that comprises two or more identical sub-networks. The term ''identical'' signifies that they share the same configuration, parameters, and weights. Updates to these parameters are reflected across all sub-networks [67]. SNNs function by comparing feature vectors to identify similarities between inputs. Notably, Siamese networks require only a few images to make accurate predictions. This ability to learn from minimal data has increased popularity recently.

Seongha Ahn and colleagues [68] introduced a system that employs an SNN combined with contrastive learning, targeting two primary functions: gauging image similarity for retrieval purposes and image matching to ascertain central coordinates. Their approach uses a triplet loss function, aiming to reduce the gap between the anchor and positive samples while increasing the one between the anchor and negative samples. For the training phase, they begin by selecting image pairs with overlapping regions as positive samples and those with minimal overlaps as negative samples. This method instructs the CNN and the image retrieval component to discern basic features and image similarity criteria, in that order. After training the CNN and FC layers dedicated to image retrieval, they then focus on the FC layers associated with the image-matching segment. This is achieved using a supervised technique that determines the central pixel in each segment, ensuring other weights remain unchanged. The image retrieval segment's role is to forecast potential patches with significant overlaps, while the image matching part identifies the central pixel connecting aerial and satellite imagery's candidate segment. The global UAV position is then inferred from the satellite image's related coordinate, which carries geolocation data. Their dataset includes 1.3k images, each with a 720 $\times$ 720 resolution, capturing a terrestrial area of 200m x 200m. For every coordinate, six images are captured to produce both positive and negative samples with overlapping regions. However, due to the areas' overlaps, the dataset's scope is limited. The authors advocate for additional studies to enhance learning technique representations and to evaluate performance in broader settings. The efficacy of the suggested system is gauged using the RMSE measure, yielding an RMSE of 36.4 meters.

SNNs are particularly effective at learning to differentiate between similar and dissimilar pairs of inputs, which is useful in this case for distinguishing between positive and negative samples (i.e., image pairs with large vs. small mutual areas). The use of contrastive learning in the SNN enables the model to effectively learn low-level features from the images, which are important for accurately identifying mutual areas and predicting center coordinates. The separation of the learning process into two stages can help the model to more effectively specialize in the two distinct tasks. However, there are also some disadvantages or challenges associated with using an SNN in this work such as the limited dataset used in this study which is relatively small and contains a lot of overlap between images, which could potentially limit the diversity of features that the SNN can learn from. This might impact the generalizability of the model to new, unseen areas. Also, while the SNN is effective for learning image similarity metrics, the image matching module requires supervised learning to predict the center pixel in each patch. This means that the model's performance is dependent on the availability and quality of labeled training data. The model's performance on the other hand might be affected by the presence or absence of mutual areas in the images. If an image lacks sufficient mutual areas, it might be harder for the model to accurately predict the center coordinates. The authors suggest the need for further research for efficient representation of learning methods and performance evaluation in general environments, indicating that the current model's performance might not be fully understood or optimized.

D. GANs-BASED METHODS

A Generative Adversarial Network (GAN) is a deep learning model that generates new synthetic data, similar to a given dataset, by simultaneously training two neural networks, a generator, and a discriminator, in an adversarial manner. The generator produces synthetic data that is intended to be similar to the training data, while the discriminator distinguishes between the real and generated data. Through this iterative process, the generator learns to produce data that is increasingly similar to real data. GANs have been used for a wide range of applications, including image synthesis, text-to-image generation, and video prediction [69].

In the study by Scheiss [70], a visual localization technique was introduced that leverages a camera along with Open Street Maps (OSM) information as an alternative to GNSS. This approach is versatile and capable of adapting to various environments and incorporating different kinds of landmarks, such as structures and roadways. The system produces location estimations at a higher rate and with increased reliability, making it suitable for elevations common in commercial drone operations. The localization system is broken down into three key phases: capturing a photo, converting this photo into a map-based format via a conditional GAN (cGAN), and aligning the altered image with a pre-existing map of the targeted area. The cGAN excels in tasks involving image segmentation, as its discriminator is trained to optimize a loss function that takes into account overall structural consistency, rather than just pixel-level accuracy. To position the segmented image within the mission-area map, it's assumed that the vehicle's altimeter and compass provide the necessary scale and rotational information. The segmented image is then systematically shifted over the reference map, and a mathematical operation–the sum of the normalized squared disparities between pixel intensities–is performed at each potential position. The position that minimizes this sum is deemed the best match, which then allows for an accurate estimation of the vehicle's current location. For the purpose of training the image segmentation component, a comprehensive dataset composed of aerial photos and corresponding OSM data was used. The training and test datasets featured images from Bonn and the surrounding regions. OSM annotations for roadways and building outlines served as labels. After training for 100 epochs, the system achieved an Intersection over Union (IoU) score of 69% for building outlines and 58% for roads in the validation dataset.

In an experiment focusing on localization, data was gathered from a plane journey over a region located to Bonn's south. Initially, the localization efficiency was assessed using a segment of the data before presenting findings for the complete dataset. The median discrepancy for this segment of the flight route was recorded at 22.7m, while it stood at roughly 40m for the entire dataset. When juxtaposed against the precision of standard GNSS receivers used by consumers – usually falling between 5 and 10m in optimal settings – the results were viewed positively.

Prior research has spotlighted the potential benefits of GANs in heightening the precision of aerial visual localization. This is achieved by crafting synthetic data, thus enhancing the caliber of datasets used for training. Nonetheless, a deeper dive into GANs' potential in this sphere warrants further investigation.

E. REINFORCEMENT LEARNING

Reinforcement Learning (RL) is a type of machine learning paradigm that enables an agent to learn how to behave in an environment by performing actions and experiencing the results of these actions. It operates on the principle of reward and punishment: actions that lead to positive outcomes are reinforced, encouraging the agent to repeat them in the future, while actions that result in negative outcomes are discouraged. This learning process is iterative, with the agent continuously refining its strategy to maximize cumulative rewards over time [71], [72].

Incorporating Reinforcement Learning (RL) into navigation systems, especially for autonomous vehicles such as Unmanned Aerial Vehicles (UAVs), presents a paradigm shift in how machines adapt and respond to their environment. Unlike conventional navigation systems that rely heavily on pre-programmed instructions and external signals such as GPS, RL enables these systems to learn optimal navigation paths through trial and error, making decisions based on real-time feedback from their surroundings.

Bodi et al. [73] introduced an innovative approach for UAV formation control in GPS Denied environments. Their method leverages a Lidar-based localization system paired with a sophisticated reinforcement learning algorithm, specifically the Deep Deterministic Policy Gradient (DDPG) [72], to manage UAV formations with unprecedented precision and adaptability. This technique is particularly crucial in dense urban areas or heavily obstructed natural environments where GPS signals are weak or non-existent, posing significant challenges for traditional navigation systems.

The utilization of Lidar technology for localization enables UAVs to accurately determine their relative positions within a formation without the need for GPS. This not only ensures reliable navigation and positioning but also opens new avenues for UAV applications in complex environments. The integration of Lidar technology provides a robust solution for autonomous UAV navigation in overcoming the limitations posed by GPS-denied areas.

The DDPG algorithm plays a central role in this methodology, enhancing the UAVs' formation maintenance capabilities through dynamic adjustments in response to environmental feedback. What sets DDPG apart is its ability to learn and adapt from experience, optimized further by a dynamically prioritized experience replay mechanism. This mechanism focuses on significant learning episodes, improving the learning process's efficiency and the effectiveness of the formation control process. By prioritizing critical learning instances, the DDPG algorithm ensures a more focused and effective adaptation to changing environmental conditions, thereby maintaining optimal formation integrity and navigational accuracy without relying on GPS signals.

This comprehensive approach, combining Lidar-based localization with the DDPG algorithm, represents a significant contribution to the field of UAV navigation. It not only addresses the critical challenge of operating in GPS-denied environments but also enhances the reliability, efficiency, and adaptability of UAV formations, paving the way for broader applications and innovations in UAV technology.

Experimental results showcased in the paper highlight the system's robustness and accuracy. The UAVs were able to maintain precise formations, adapt to dynamic obstacles, and reconfigure based on the changing environment, all without

the aid of GPS. This outcome not only demonstrates the feasibility of the proposed method but also its superiority over traditional GPS-dependent control systems in certain contexts. The results indicate a significant reduction in formation errors and an improvement in the adaptability and responsiveness of the UAVs to unforeseen changes in their operating environment.

However, the reliance on Lidar and sophisticated computational algorithms brings forth considerations regarding the practical deployment of this technology. The hardware requirements for Lidar and the computational demands for processing the reinforcement learning algorithm may pose challenges, particularly in terms of power consumption and the computational capacity of smaller UAVs.

In conclusion, this paper makes a significant contribution to the field of UAV navigation and control. It provides a viable solution for UAV formation control in GPS-denied environments, paving the way for broader applications of UAV technology in complex and challenging scenarios. Despite its advantages, the practical implementation of this system will require addressing the limitations related to hardware and computational resources, ensuring the technology's adaptability to a wide range of UAV platforms.

IV. DISCUSSION

This section delves into the insights derived from the reviewed deep learning-based aerial visual localization approaches, summarized in the comparison table 1. The methodologies predominantly harness the power of CNNs due to their exceptional ability in feature extraction. The emergence of hybrid models, which integrate CNNs with RNNs and FFNNs, marks a significant advancement toward addressing the complex challenges in aerial visual localization. These models have shown promise in enhancing performance, particularly in environments where GPS signals are compromised or absent, and in transforming aerial images into actionable map-like representations for real-time localization.

The diversity in datasets, ranging from high-definition aerial images to meticulously crafted simulations, highlights the versatility of these approaches. Tools such as MATLAB/Simulink and Google Earth™ play a pivotal role in generating simulated environments that closely mirror real-world scenarios, providing a robust foundation for model training and validation. Reference imagery, primarily sourced from Google Earth™ and Open Street Maps, serves as a cornerstone for these studies, offering a ground truth that is essential for the calibration and verification of localization models.

Despite these advancements in technology, challenges persist. The quest for comprehensive datasets that cover a wider range of environmental conditions continues to be a crucial focus for future research. Moreover, enhancing model Understanding and minimizing computational demands are critical considerations for the deployment of these technologies in UAV systems, which often struggle with limited resources.

The validation environments detailed in Table 2 underscore the varied scales and operational altitudes across different studies. This variability is crucial for assessing the applicability and effectiveness of each UAV localization method under diverse conditions. The need for extensive and varied datasets is evident, as is the importance of model adaptability to different landscapes and operational scenarios.

The performance metrics summarized in the subsequent tables provide valuable insights into the accuracy and reliability of different UAV localization methods. Metrics such as Euclidean distance, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) offer a quantifiable measure of each approach's efficacy. These results not only showcase the potential of neural network-based approaches in surpassing traditional methods like Kalman filters in accuracy and stability but also highlight the importance of innovative model architectures and learning strategies in pushing the boundaries of UAV localization technology.

The reviewed studies collectively contribute to the evolving landscape of UAV localization research. The integration of advanced neural network architectures, along with the strategic use of diverse datasets and reference imagery, lays the way for significant improvements in localization accuracy and efficiency. As the field continues to advance, addressing the outlined challenges will be paramount in realizing the full potential of UAV localization technologies for a wide array of applications in complex and dynamic environments.

V. CHALLENGES OF DEEP LEARNING APPROACHES

Integrating cutting-edge deep learning techniques into the realm of aerial visual localization holds enormous promise, presenting unprecedented potential for reshaping several industries and applications (e.g., aerial mapping, surveying, inspection, delivery applications, environment monitoring, and precision agriculture) profoundly. Nonetheless, this domain suffers from intricate challenges:

- Evaluating and comparing different methods performance: Though the research on aerial visual localization using deep learning approaches has grown rapidly, comparing existing works in this domain can be challenging due to several factors, ranging from the absence of a standardized methodology and variations in data collection to differences in evaluation metrics and missing details of implementations. Remarkably, there is no standardized approach for conducting experiments on aerial visual localization using deep learning. This manifests in terms of the data collection process, sensor utilization, and image annotation methods. Unlike other domains (e.g., computer vision), up to our knowledge, a publicly available benchmark dataset for evaluating and comparing deep learning-based aerial visual localization methods remains elusive. To address this problem, researchers opt to collect data under

TABLE 1. Comprehensive comparison of the reviewed methods for aerial visual localization.

Reference	Model Architecture	Methodology	Dataset	Performance Metrics	Reference Imagery	Key Findings
Amer *et al.* [49].	CNN	Pre-trained VGG16-Net for creating deep urban signatures	Urban geotagged images	Accuracy: 91.2%, Localization error: 200.75m in 6 districts	Google Earth™	Demonstrated superior performance in urban localization over traditional methods.
Nassar *et al.* [51].	CNN	Framework integrating SIFT, RANSAC, and U-Net for enhanced aerial imagery registration	Simulated aerial flight video with a bird-eye view using Google Earth™ + YouTube video of a UAV flight	Average geolocation error reduced to 5.1m and 3.6m with deep learning	Google Earth™ & Bing™	Showcased precise localization in GPS-denied environments.
Mughal *et al.* [52]	CNN	Neighborhood consensus technique with ResNet-101 for feature matching	Dataset of 2052 HD aerial images	Mean error: 3.594m, Max error: 31.281m	Pre-stored georeferenced orthomosaic images	Achieved remarkable accuracy within prestored orthomosaic maps.
Goforth and Lucey [53]	CNN	Monocular RGB camera with CNN representations for satellite imagery comparison	Geographical Data from United States Geographical Survey Earth Explorer website	Average localization error of less than 8 meters	Google Earth™	Improved localization accuracy in environments with few landmarks.
Arturo and Martinez [54]	CNN	'GreySeqNet', a compact CNN for autonomous drone racing using greyscale images	Gazebo simulator generated dataset	Average camera pose error of 31cm, Frequency: up to 83Hz	Not Applicable	Enabled high-speed and precise maneuvering in drone racing.
Pi *et al.* [55].	CNN	Mapping approaches (PROC and PROS) with CNN models for orthogonal view localization	Not specified	mAP: Model-P 97%, Model-O 51%	Not Applicable	Effective for real-time localization and mapping with RGB camera inputs.
Cabrera-Ponce *et al.* [56].	CNN & RNN	Streamlined CNN architecture for deducing GPS positions from single aerial images	Datasets with images taken from 25m and 20m elevations	Prediction error: 2.8 to 6.1 meters, Speed: 103fps	GeoTagged images	Rapid and accurate geolocation using single aerial images.
Cao *et al.* [59].	CNN	Enhanced DenseNet with quality-aware template matching for UAV localization in satellite images	University-1652	Localization accuracy within 5 meters	Google Maps	Showcased significant increase in localization accuracy with enhanced DenseNet.
Amer *et al.* [63].	CNN	Deep CNNs and regression models for UAV trajectory following in GPS-denied environments	Synthetic environments using Unreal Engine	Improved path following accuracy with visual cues	Not Applicable	Focused on trajectory maintenance using visual inputs, demonstrating potential for path-based UAV localization.

TABLE 1. *(Continued.)* **Comprehensive comparison of the reviewed methods for aerial visual localization.**

Luo *et al.* [64].	CNN	KeepEdge: Edge computing and visual information for parcel delivery localization	UAV and street-view images from Anhui University's 21 buildings. 35582 images total	Improved accuracy of UAV deliveries in complex environments	street-view images	Enhanced UAV localization accuracy using edge computing in GPS-denied environments for precise parcel delivery.
Zhuofan Cui *et al.* [65].	CNN	Vision Transformer with cross-view consistent attention for geo-localization	University-1652	Recall (R1): 91.5% (Euclidean), Accuracy improvement in image retrieval	Not Specified	Introduced cross-view consistent attention improving matching accuracy across UAV and satellite images.
Zahedi *et al.* [66]	FFNN & RNN	FFNN for mapping UAV data to target positions; seq2seq RNN for out-of-view target location prediction.	Data collected in a MATLAB/Simulink simulation environment	FFNN: MAE reduced for x-location (6.27m) and y-location (5.94m). RNN: Stable MAE across future time steps.	MATLAB/ Simulink simulation environment.	Demonstrates superior performance of neural networks over Kalman filter in localization accuracy and predictive stability.
Ahn *et al.* [68].	SNN	Contrastive learning for aerial and satellite image matching for UAV localization.	Simulated aerial and satellite images from Google Earth™, incorporating visual variations.	Top-3 precision for image retrieval at 67%, RMSE for location estimation at 36.4m.	Google Earth™	The framework effectively predicts UAV global coordinates by matching aerial images with satellite imagery, showcasing robustness in visually varied conditions.
Schleiss [70].	GAN	Translated aerial images into map-like representations, then matched with Open Street Map data for UAV localization without GNSS.	Aerial imagery over Bonn, Germany, and its surrounding areas, with Open Street Maps for ground truth.	Median localization error of 22.7m over selected paths with suitable landmarks. overall median error of 40m including images without matchable features.	Open Street Maps.	The method achieves high matching rates and is robust against temporal changes, capable of incorporating multiple landmark types for accurate UAV localization.
Bodi MA *et al.* [73]	DRL	Lidar-based UAV formation localization and DRL control in GPS-denied areas with collision risk evaluation and efficient learning via dynamic experience replay.	Real-world UAV flight data and simulations.	Showed improved formation accuracy and robust control in GPS-denied environments; the Lidar-based method achieved relative localization with high precision.	Not specified.	Demonstrated effective UAV formation flight control using Lidar for localization and DRL for maneuvering in GPS-denied environments, with significant improvements in accuracy and robustness over traditional methods.

TABLE 2. Validation environments.

Reference	Environment Size	Altitude of UAV	Note
Amer *et al.* [49]	Different districts with size 500m x 500m	Different altitudes, with a maximum altitude of 333m	Focuses on district-level localization in Cairo, Egypt using urban geotagged images from Google and Bing maps. Simulates drone footage with Bing maps.
Nassar *et al.* [51]	1200m and 500m Trajectory path	300m	Integrates SIFT, RANSAC, and U-Net for registration, tested with simulated and real UAV flights over urban areas.
Mughal *et al.* [52]	2 km² area was used in training and testing	Not specified	Uses HD aerial images for feature matching and localization, emphasizing accuracy within pre-stored maps.
Goforth and Lucey [53]	5.9km by 7.5km (Training)	0.2km and 0.22km	Training with a dataset from New Jersey, USA. Tested on urban (village) and rural (gravel pit) environments using the senseFly eBee drone. Focuses on GPS-denied environments.
Arturo and Martinez [54]	Gazebo simulator generated dataset	Not specified	Utilizes simulated data for drone racing, demonstrating maneuvering capabilities in a controlled environment.
Pi *et al.* [55]	426 inches in width (X-axis) and 333 inches in length (Y-axis).	Not specified	proposed mapping approaches with CNN models for orthogonal view localization, emphasizing effectiveness for real-time mapping and localization with RGB camera inputs.
Cabrera-Ponce *et al.* [56]	4000 Images from 25m and 20m elevations	25m and 20m	Streamlines CNN and RNN for deducing GPS positions from aerial images, focusing on rapid and accurate geolocation using GeoTagged images.
Cao *et al.* [59]	1.6 km x 1.6km - University-1652 dataset	300m to 400m	Enhances DenseNet for template matching between UAV and satellite images for visual localization, introducing multi-scale feature fusion and quality-aware template matching.
Amer *et al.*(2021) [63]	Unreal Engine simulated environment - several paths with a range of lengths [145-412]m	Simulated altitudes	Uses CNN and regression for UAV path following without GPS, focusing on navigation in GPS-denied environments with synthetic data for training.
Luo *et al.* [64]	35582 Images from Anhui University's campus - Size is not specified	[30-60]m	Employs Edge computing for parcel delivery localization in GPS-denied environments, enhancing UAV accuracy with visual and street-view images for precise delivery.
Zhufian *et al.* [65]	Not specified	Not specified	Details on environment size, altitude, and additional notes need to be extracted from the document.
Zahedi *et al.* [66]	MATLAB/Simulink simulation environment	Simulated altitudes	Uses FFNN and RNN for target geo-localization and location prediction, outperforming traditional Kalman filter-based methods in accuracy and consistency.

TABLE 2. *(Continued.)* **Validation environments.**

Ahn *et al.* [68]	Dataset of 1.3k images, 720*720 resolutions capturing 200m x 200m each	Not specified	Framework employs CNN-based Siamese Neural Network with contrastive learning for aerial-satellite image matching.
Schleiss [70]	Covers an area of 560m long and 680m wide with 4121 overlapping images	Approximately 300m	Uses GANs for translating aerial images into map-like representations for matching with Open Street Maps data.
Bodi *et al.* [73]	Not specified	Not specified	Details on environment size, altitude, and additional notes need to be extracted from the document.

different environmental and weather conditions, which often is kept private. The scarcity of publicly available benchmark datasets impedes the robust comparison of different methods. In the pursuit of assessing diverse methods, evaluation metrics are of paramount importance. In this context, the mean absolute error and root mean square errors are widely used. However, different studies may use different evaluation metrics. The disparate adoption of different metrics across different studies hampers the seamless comparison of performance. As such, a comprehensive and holistic assessment of methods necessitates adopting a widely acknowledged set of metrics. Such an approach would enable a holistic and fair appraisal of methods. Besides the aforementioned hurdles, a majority of existing studies lack details about critical model configurations and hyper-parameters (e.g., number of layers and number of neurons), which hinders the fair comparison of various methods' performance. The scarcity of shared source codes further amplifies the problem and makes it even harder to compare results.

- Ethical, privacy, regulatory, and safety implications: Despite their enticing potential and expected benefits for real-world applications, aerial visual localization methods bring forth a new array of ethical, privacy, and safety concerns. These concerns stem from the inherent reliance on high-resolution images acquired through onboard cameras and come into sharp focus, particularly when UAVs navigate over private properties or sensitive areas. Therefore, it becomes an imperative priority to design and implement stringent protocols to both preserve privacy rights and responsibly manage the troves of data gathered during localization processes. Another key aspect to consider is compliance with relevant national and international regulations. Aviation authorities and governing bodies outline and impose regulations and restrictions on UAVs' operations. Deep learning-based visual localization techniques must adhere to these mandates, encompassing airspace restrictions, requisite flight permits, and privacy-preserving guidelines, among others. Furthermore, as deep learning-based techniques integrate into navigation and localization, worries arise about system reliability and trustworthiness. This is especially critical in civilian areas, where any failure could lead to accidents or property damage, endangering human lives and infrastructure. To mitigate these concerns, a robust system design becomes a must, proactively preventing damage and hazards stemming from potential failures of such techniques. Having that said, using deep learning-based visual localization technology ethically and responsibly is of utmost importance. This entails ensuring that the technology is used for legitimate purposes and by legal and ethical standards. Transparency in data usage, obtaining informed consent, and maintaining accountability are critical factors that must be considered. By addressing these implications, the integration of visual localization in UAVs can be done responsibly and beneficially while preserving privacy, ensuring safety, and complying with regulations and ethical standards.

VI. FUTURE RESEARCH DIRECTIONS

Potential future directions may consider several avenues.

- Learning and training paradigms:
  - Researchers may investigate the possibility of using different learning and training methods such as online, transfer, and compact learning techniques. Online learning methods, allow models to learn and adapt to changes on-the-fly during operation. This also has been suggested by Cabrera-Ponce et al. [56]. Another interesting research path is investigating the use of transfer learning to overcome the perennial issue of training datasets scarcity. This involves investigating how models trained in one sort of environment perform when used in a different environment. This not only helps to alleviate the problem of data scarcity but also sheds light on the adaptability and generalizability of deep learning. In addition, active learning strategies become essential in scenarios where data collection is expensive. An active learning model autonomously selects the most valuable data for acquisition and learning. Given the pivotal role of data quality and quantity in the effectiveness of deep

learning methods, active learning strategies require further exploration in the context of localization.
  - Future work could involve deploying the proposed algorithm by Amer et al. [63] on a real drone. Additionally, Generative Adversarial Networks (GANs) could be used to apply style transfer between synthetic and real training images to improve system performance and reduce the size of the training dataset. Investigating the incorporation of a Siamese network into the current model for UAV geolocalization is another direction.
  - Future research could investigate using other visual bandwidths, such as thermal imagery, and creating ad-hoc reference points using multiple cooperative UAVs. Solutions to improve detection and mapping under low-light conditions could also be explored [55].
- Advancing model architectures and ensemble techniques:
  - The evolution of DNN architectures remains a focal point of research. As such, researchers may examine the impact of deeper models on localization accuracy. This would involve designing new deep neural network architectures or customizing/modifying existing ones such as PoseNet and CompactPN. Aligned with this direction, combining different types of neural networks (e.g., CNNs, RNNs, LSTMs, etc.) or using ensemble learning could improve the localization performance. Investigating the impact of compact models might be another research direction as they can offer significant speed advantages. More research could be done to investigate how to make these models even faster without losing accuracy.
  - Future work could experiment with larger datasets covering multiple cities and improve the semantic segmentation pipeline by exploiting additional UAV videos. The dataset should help in developing an end-to-end deep learning algorithm that replaces traditional computer vision methods. Efficient semantic segmentation methods such as ESPNet and ShuffleSeg could also be investigated. Additionally, the SSM component can be improved by replacing the demanding dictionary search and heuristics with a CNN inspired by existing models. Pose estimation and image-to-image registration could be experimented with for calibration and sequential frame registration components [51].
- Enhancing model interpretability and trustworthiness:
  - The trustworthiness of deep learning models can be substantially bolstered by focusing on interpretability. The development of models capable of producing more interpretable and explainable results is a noteworthy direction. This helps in understanding these models' underlying decision-making processes and enhances their acceptance and adoption in critical real-world scenarios.
- Exploring data fusion for enhanced context:
  - Since the effectiveness of deep learning methods greatly relies on the quality and quantity of training data, future research could explore data fusion of multimodal sources, such as altimeter data, IMU, or even LIDAR. This could offer additional context and potentially enhance accuracy.
- Real-World assessment of model robustness:
  - While controlled settings have been used in most contemporary studies, the practicality and resilience of these models must be extensively examined beyond the laboratory setting. Such research will shed light on the models' ability to adapt to the complexities and uncertainties seen in real-world localization applications.
  - Future work could involve experimenting with larger datasets, improving the efficiency of semantic segmentation methods, and replacing heuristic search methods with CNN-based approaches. Investigating pose estimation and image-to-image registration for calibration and sequential frame registration components is also suggested [52].
  - Future work could focus on adding more efficiency to the method to achieve faster operation frequencies than those obtained in the current work. Another direction is exploring the use of two or more cameras [54].

VII. CONCLUSION

GNSS is a cornerstone for the safe and efficient operations of UAVs. However, GNSS signals are susceptible to blockage, reflection, and spoofing, resulting in navigation loss and potentially UAV loss. To address these challenges, researchers have turned to visual navigation methods as alternative solutions to conventional navigation methods. This paper provides insights into the potential of UAV visual localization and the challenges associated with its implementation, focusing on deep Learning-based visual navigation systems. In Figure 3, a classification of visual localization methods is presented. We have systemically examined state-of-the-art deep learning-based methodologies against eight criteria: ML-model, data, reference imagery, environment, altitude, metrics, and results, as summarized in Tables 1- 2.

Aerial visual localization using deep learning techniques holds immense potential for various industries and applications. Nevertheless, it faces several challenges including evaluating and comparing different methods, assessing performance, addressing ethical, privacy, regulatory, and safety implications, and complying with relevant regulations. Existing studies often lack details on critical model configurations and hyper-parameters, making it difficult to compare results. Furthermore, ensuring the reliability and trustworthiness of these systems is crucial, particularly in the civil context. Integrating deep learning-based visual

TABLE 3. Key terminologies in UAV localization and their descriptions.

Abbreviation	Full Form	Description
UAV	Unmanned Aerial Vehicles	Aircraft systems operated without a human pilot onboard.
GNSS	Global Navigation Satellite System	A satellite system that provides geo-spatial positioning with global coverage.
GPS	Global Positioning System	A satellite-based navigation system used to determine the ground position of an object.
INS	Inertial Navigation Systems	Navigation aids that use a computer, motion sensors, and rotation sensors to continuously calculate the position, orientation, and velocity of a moving object without needing external references.
NLOS	Non-Line-Of-Sight	describe signals that travel along indirect paths to reach the receiver, often obstructed by physical objects.
AVL	Absolute Visual Localization	The process of determining an exact position in space using visual information.
RVL	Relative Visual Localization	Identifying the position relative to known locations or objects, using visual information.
SLAM	Simultaneous Localization and Mapping	A technique where a device can create a map of its environment while navigating through it.
SSD	Sum of Squared Differences	A method used in computer vision to measure the similarity between two images.
PF	Particle Filter	A sequential Monte Carlo method used for estimating the state of a system where the true state cannot be measured directly.
NCC	Normalized Cross-Correlation	A measure of similarity between two signals, commonly used in image processing.
PDF	Probability Density Function	A function that describes the likelihood of a random variable to take on a given value.
EKF	Extended Kalman Filter	An algorithm that extends the Kalman filter to nonlinear systems.
MI	Mutual Information	A measure of the mutual dependence between two variables.
RMSE	Root Mean Square Error	A standard way to measure the error of a model in predicting quantitative data.
IMU	Inertial Measurement Units	Electronic devices that measure and report a body's specific force, angular rate, and sometimes the magnetic field surrounding the body.
SIFT	Scale-Invariant Feature Transform	An algorithm in computer vision to detect and describe local features in images.
ORB	Oriented FAST and Rotated BRIEF	A fast robust feature detector and descriptor, useful in computer vision tasks.
HOG	Histogram of Oriented Gradient	A feature descriptor used in computer vision and image processing for object detection.
OF	Optical Flow	The pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer and the scene.
CNNs	Convolutional Neural Networks	A class of deep neural networks, most commonly applied to analyzing visual imagery.
RNNs	Recurrent Neural Networks	A class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence.
PV	Perspective View	An approximate representation, on a flat surface, of an image as it is seen by the eye.
OV	Orthogonal View	A method of representing three-dimensional objects in two dimensions, where the view direction is orthogonal to the projection plane.
mAP	Mean Average Precision	A metric used to evaluate the accuracy of object detectors.
FC	Fully Connected	A type of neural network layer where each neuron is connected to all neurons in the previous and next layers.
SSM	Semantic Shape Matching	A process of matching shapes based on their semantic meaning and geometric properties.
PROC	Perspective to orthogonal conversion based on Reference Objects' Coordinates	A method to transform perspective views to orthogonal views using the coordinates of reference objects.
PROS	Perspective to Orthogonal Conversion based on Reference Objects' Size	A method to transform perspective views to orthogonal views based on the size of reference objects.
SURF	Speeded Up Robust Features	A patented local feature detector and descriptor that can be used for tasks such as object recognition or 3D reconstruction.
SGD	Stochastic Gradient Descent	Optimization algorithm used to minimize the function by iteratively moving towards the minimum value of the gradient.

TABLE 3. *(Continued.)* **Key terminologies in UAV localization and their descriptions.**

FCNN	Fully Connected Neural Network	A neural network where all neurons between layers are connected, commonly used for classification tasks.
GANs	Generative Adversarial Networks	A class of machine learning frameworks where two networks, a generator and a discriminator, are trained simultaneously to generate new data with the same statistics as the training set.
AP	Average Precision	A metric that combines recall and precision for ranked retrieval results, which is useful for evaluating classification models.
FFNN	Feed-Forward Neural Network	An artificial neural network where connections between nodes do not form a cycle, typically used in straightforward classification and regression tasks.
ReLU	Rectified Linear Unit	A nonlinear function that outputs the input directly if it is positive, else, it will output zero. It's used to add nonlinearity to neural networks.
LSTM	Long Short-Term Memory	A recurrent neural network capable of learning order dependence in sequence prediction problems.
SNN	Siamese Neural Network	A neural network architecture that compares a pair of inputs by passing them through symmetric subnetworks and is commonly used for tasks that involve finding similarity or relationship between two inputs.
GAN	Generative Adversarial Network	A single network within GANs focused on generating data. The term can refer to either the generator or discriminator within the GAN framework.
OSM	Open Street Maps	A collaborative project to create a free editable map of the world, widely used for geospatial data and mapping applications.
IoU	Intersection over Union	A metric used in object detection to measure the overlap between predicted and ground truth bounding boxes.
RL	Reinforcement Learning	A type of machine learning where an agent learns to make decisions by performing actions in an environment to achieve some rewards.
DDPG	Deep Deterministic Policy Gradient	An algorithm in reinforcement learning that uses a neural network to approximate the policy and the value functions.
RANSAC	Random Sample Consensus	An iterative method to estimate parameters of a mathematical model from a set of observed data which contains outliers.

localization technology ethically and responsibly is essential, involving transparency, informed consent, and accountability.

Future directions may explore various learning and training methods, including online, transfer, and compact learning techniques. Investigating the impact of deeper models on localization accuracy is another avenue, with new deep neural network architectures or customizing existing ones. Combining different neural networks or ensemble learning could further enhance localization performance. Compact models offer speed advantages, but further research is needed to optimize their speed without compromising accuracy. Active learning strategies are essential in scenarios where data collection is expensive. Future studies could incorporate data from multiple sources, such as altimeter data, IMU, or LIDAR, to improve accuracy. Testing these models in diverse real-world conditions is crucial for understanding their robustness and practical applicability.

APPENDIX

See Table 3.

ACKNOWLEDGMENT

ChatGPT (openai.com) has been used to proofread and enhance the language and readability of the article.

REFERENCES

[1] B. Fan, Y. Li, R. Zhang, and Q. Fu, "Review on the technological development and application of UAV systems," *Chin. J. Electron.*, vol. 29, no. 2, pp. 199–207, Mar. 2020.

[2] G. N. Muchiri and S. Kimathi, "A review of applications and potential applications of UAV," in *Proc. Sustain. Res. Innov. (SRI) Conf.*, 2022, pp. 280–283.

[3] D. C. Tsouros, S. Bibi, and P. G. Sarigiannidis, "A review on UAV-based applications for precision agriculture," *Information*, vol. 10, no. 11, p. 349, Nov. 2019.

[4] S. A. H. Mohsan, M. A. Khan, F. Noor, I. Ullah, and M. H. Alsharif, "Towards the unmanned aerial vehicles (UAVs): A comprehensive review," *Drones*, vol. 6, no. 6, p. 147, Jun. 2022.

[5] P. K. Panigrahi and S. K. Bisoy, "Localization strategies for autonomous mobile robots: A review," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 34, no. 8, pp. 6019–6039, Sep. 2022.

[6] N. Gyagenda, J. V. Hatilima, H. Roth, and V. Zhmud, "A review of GNSS-independent UAV navigation techniques," *Robot. Auto. Syst.*, vol. 152, Jun. 2022, Art. no. 104069.

[7] G. Zhang and L.-T. Hsu, "Intelligent GNSS/INS integrated navigation system for a commercial UAV flight control system," *Aerosp. Sci. Technol.*, vol. 80, pp. 368–380, Sep. 2018.

[8] N. S. Hashemi, R. B. Aghdam, A. S. B. Ghiasi, and P. Fatemi, "Template matching advances and applications in image analysis," 2016, *arXiv:1610.07231*.

[9] S. Lee, J. Ryu, and W. Byun, "Survey on navigation satellite system and technologies," *Electron. Telecommun. Trends*, vol. 36, no. 4, pp. 61–71, 2021.

[10] A. Couturier and M. A. Akhloufi, "Relative visual localization (RVL) for UAV navigation," *Proc. SPIE*, vol. 10642, pp. 213–226, May 2018.

[11] A. Couturier and M. A. Akhloufi, "A review on absolute visual localization for UAV," *Robot. Auton. Syst.*, vol. 135, Jan. 2021, Art. no. 103666.
[12] N. Kubo, K. Kobayashi, and R. Furukawa, "GNSS multipath detection using continuous time-series C/N0," *Sensors*, vol. 20, no. 14, p. 4059, Jul. 2020.
[13] P. Dhomane and R. Mathew, "Counter-measures to spoofing and jamming of drone signals," Tech. Rep., 2020.
[14] Z. Wu, Y. Zhang, Y. Yang, C. Liang, and R. Liu, "Spoofing and anti-spoofing technologies of global navigation satellite system: A survey," *IEEE Access*, vol. 8, pp. 165444–165496, 2020.
[15] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, Dec. 2016.
[16] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017.
[17] Y. Lu, Z. Xue, G.-S. Xia, and L. Zhang, "A survey on vision-based UAV navigation," *Geo-Spatial Inf. Sci.*, vol. 21, no. 1, pp. 21–32, Jan. 2018.
[18] S. Leung and E. J. Shamwell, "Outdoor visual localization: A survey," Army Res. Lab., Combat Capabilities Develop. Command, Adelphi, MD, USA, Tech. Rep., 2020.
[19] R. O. Duda, D. G. Stork, and P. E. Hart, *Pattern Classification*. Hoboken, NJ, USA: Wiley, 2006.
[20] K. Thyagarajan, *Digital Image Processing With Application to Digital Cinema*. Evanston, IL, USA: Routledge, 2005.
[21] M. B. Hisham, S. N. Yaakob, R. A. A. Raof, A. B. A. Nazren, and N. M. Wafi, "Template matching using sum of squared difference and normalized cross correlation," in *Proc. IEEE Student Conf. Res. Develop. (SCOReD)*, Dec. 2015, pp. 100–104.
[22] Z. Cui, W. Qi, and Y. Liu, "A fast image template matching algorithm based on normalized cross correlation," *J. Phys., Conf. Ser.*, vol. 1693, no. 1, Dec. 2020, Art. no. 012163.
[23] G. J. Van Dalen, D. P. Magree, and E. N. Johnson, "Absolute localization using image alignment and particle filtering," in *Proc. AIAA Guid., Navigat., Control Conf.*, Jan. 2016, p. 0647.
[24] J. P. Lewis, "Fast template matching," in *Proc. Vis. Interface*, vol. 95, Quebéc City, QC, Canada, 1995, pp. 15–19.
[25] A. Yol, B. Delabarre, A. Dame, J.-É. Dartois, and E. Marchand, "Vision-based absolute localization for unmanned aerial vehicles," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2014, pp. 3429–3434.
[26] R. Szeliski, *Feature Detection and Matching*. Cham, Switzerland: Springer, 2022, pp. 333–399.
[27] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 2, Sep. 1999, pp. 1150–1157.
[28] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[29] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. 9th Eur. Conf. Comput. Vis.*, Graz, Austria. Springer, May 2006, pp. 404–417.
[30] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571.
[31] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[32] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 23–79, Jan. 2021.
[33] M. Shan, F. Wang, F. Lin, Z. Gao, Y. Z. Tang, and B. M. Chen, "Google map aided visual navigation for UAVs in GPS-denied environment," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2015, pp. 114–119.
[34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[35] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
[36] A. Carrio, C. Sampedro, A. Rodriguez-Ramos, and P. Campoy, "A review of deep learning methods and applications for unmanned aerial vehicles," *J. Sensors*, vol. 2017, no. 1, pp. 1–13, 2017.
[37] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep learning vs. traditional computer vision," in *Advances in Computer Vision*, K. Arai and S. Kapoor, Eds., Cham, Switzerland: Springer, 2020, pp. 128–144.
[38] R. Atienza, *Advanced Deep Learning With TensorFlow 2 and Keras: Apply DL, GANs, VAEs, Deep RL, Unsupervised Learning, Object Detection and Segmentation, and More*. Birmingham, U.K.: Packt, 2020.
[39] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, pp. 1–74, Mar. 2021.
[40] M. Coşkun, A. Uçar, Ö. Yildirim, and Y. Demir, "Face recognition based on convolutional neural network," in *Proc. Int. Conf. Modern Electr. Energy Syst. (MEES)*, Nov. 2017, pp. 376–379.
[41] P. Kamencay, M. Benco, T. Mizdos, and R. Radil, "A new method for face recognition using convolutional neural network," *Adv. Electr. Electron. Eng.*, vol. 15, no. 4, pp. 663–672, Nov. 2017.
[42] S. Sharma, K. Shanmugasundaram, and S. K. Ramasamy, "FAREC—CNN based efficient face recognition technique using dlib," in *Proc. Int. Conf. Adv. Commun. Control Comput. Technol. (ICACCCT)*, May 2016, pp. 192–195.
[43] Z. Ouyang, J. Niu, Y. Liu, and M. Guizani, "Deep CNN-based real-time traffic light detector for self-driving vehicles," *IEEE Trans. Mobile Comput.*, vol. 19, no. 2, pp. 300–313, Feb. 2020.
[44] T.-D. Do, M.-T. Duong, Q.-V. Dang, and M.-H. Le, "Real-time self-driving car navigation using deep neural network," in *Proc. 4th Int. Conf. Green Technol. Sustain. Develop. (GTSD)*, Nov. 2018, pp. 7–12.
[45] B. T. Nugraha, S.-F. Su, and Fahmizal, "Towards self-driving car using convolutional neural network and road lane detector," in *Proc. 2nd Int. Conf. Autom., Cognit. Sci., Opt., Micro Electro-Mech. Syst., Inf. Technol. (ICACOMIT)*, Oct. 2017, pp. 65–69.
[46] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical image analysis using convolutional neural networks: A review," *J. Med. Syst.*, vol. 42, no. 11, pp. 1–13, Nov. 2018.
[47] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.
[48] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[49] K. Amer, M. Samy, R. ElHakim, M. Shaker, and M. ElHelw, "Convolutional neural network-based deep urban signatures with application to drone localization," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 2138–2145.
[50] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[51] A. Nassar, K. Amer, R. ElHakim, and M. ElHelw, "A deep CNN-based framework for enhanced aerial imagery registration with applications to UAV geolocalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 1513–1523.
[52] M. H. Mughal, M. J. Khokhar, and M. Shahzad, "Assisting UAV localization via deep contextual image matching," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2445–2457, 2021.
[53] H. Goforth and S. Lucey, "GPS-denied UAV localization using pre-existing satellite imagery," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 2974–2980.
[54] J. A. Cocoma-Ortega and J. Martinez-Carranza, "A compact CNN approach for drone localisation in autonomous drone racing," *J. Real-Time Image Process.*, vol. 19, no. 1, pp. 73–86, Feb. 2022.
[55] Y. Pi, N. D. Nath, and A. H. Behzadan, "Deep neural networks for drone view localization and mapping in GPS-denied environments," in *Proc. 18th Int. Conf. Comput. Civil Building Eng. (ICCCBE)*, 2020, pp. 1–16, doi: 10.46421/2706-6568.37.2020.
[56] A. A. Cabrera-Ponce and J. Martinez-Carranza, "Convolutional neural networks for geo-localisation with a single aerial image," *J. Real-Time Image Process.*, vol. 19, no. 3, pp. 565–575, Jun. 2022.
[57] A. Kendall, M. Grimes, and R. Cipolla, "PoseNet: A convolutional network for real-time 6-DOF camera relocalization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2938–2946.

[58] 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.
[59] Y. Cao, K. Ren, and Q. Chen, "Template matching based on convolution neural network for UAV visual localization," *Optik*, vol. 283, Jul. 2023, Art. no. 170920.
[60] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
[61] Z. Zheng, Y. Wei, and Y. Yang, "University-1652: A multi-view multi-source benchmark for drone-based geo-localization," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 1395–1403.
[62] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012.
[63] K. Amer, M. Samy, M. Shaker, and M. ElHelw, "Deep convolutional neural network based autonomous drone navigation," *Proc. SPIE*, vol. 11605, pp. 16–24, Jan. 2021.
[64] H. Luo, T. Chen, X. Li, S. Li, C. Zhang, G. Zhao, and X. Liu, "KeepEdge: A knowledge distillation empowered edge intelligence framework for visual assisted positioning in UAV delivery," *IEEE Trans. Mobile Comput.*, vol. 22, no. 8, pp. 4729–4741, Aug. 2022.
[65] Z. Cui, P. Zhou, X. Wang, Z. Zhang, Y. Li, H. Li, and Y. Zhang, "A novel geo-localization method for UAV and satellite images using cross-view consistent attention," *Remote Sens.*, vol. 15, no. 19, p. 4667, Sep. 2023.
[66] R. Zahedi, E. Ceh-Varela, R. Selje II, H. Cao, and L. Sun, "Neural network based approaches to mobile target localization and tracking using unmanned aerial vehicles," in *Proc. AIAA Scitech Forum*, Jan. 2020, p. 0392.
[67] L. Utkin, M. Kovalev, and E. Kasimov, "An explanation method for Siamese neural networks," in *Proc. Int. Sci. Conf. Telecommun., Comput. Control*. Springer, 2019, pp. 219–230.
[68] S. Ahn, H. Kang, and J. Lee, "Aerial-satellite image matching framework for UAV absolute visual localization using contrastive learning," in *Proc. 21st Int. Conf. Control, Autom. Syst. (ICCAS)*, Oct. 2021, pp. 143–146.
[69] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.
[70] M. Schleiss, "Translating aerial images into street-map-like representations for visual self-localization of UAVs," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 575–580, Jun. 2019.
[71] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
[72] 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*.
[73] B. Ma, Z. Liu, F. Jiang, W. Zhao, Q. Dang, X. Wang, J. Zhang, and L. Wang, "Reinforcement learning based UAV formation control in GPS-denied environment," *Chin. J. Aeronaut.*, vol. 36, no. 11, pp. 281–296, Nov. 2023.

OMAR Y. AL-JARRAH received the B.Sc. degree in computer engineering from Yarmouk University, Irbid, Jordan, in 2005, the M.Sc. degree in computer engineering from The University of Sydney, Sydney, NSW, Australia, in 2008, and the Ph.D. degree in electrical and computer engineering from Khalifa University, Abu Dhabi, Abu Dhabi, in 2016. He has more than 12 years of combined academic and industrial experience. He is currently an Assistant Professor with Jordan University of Science and Technology, Irbid. His main research interests include machine learning, intrusion detection, big data analytics, autonomous and connected vehicles, unmanned aerial vehicles, and knowledge discovery in various applications.

AHMED S. SHATNAWI received the B.S. degree from the Department of Computer Engineering, Jordan University of Science and Technology, in January 2007, the M.S. degree in software engineering from George Mason University, in 2012, and the Ph.D. degree in engineering from the University of Wisconsin–Milwaukee, in 2017. He is currently an Associate Professor with the Department of Software Engineering and the Department of Network Engineering and Security, Jordan University of Science and Technology, with extensive experience in software engineering and information security. His research interests include software engineering, information security, cryptography, and human–computer interaction. He is especially interested in finding better ways to design software systems that are safe, secure, and reliable to use.

MOHAMMAD M. SHURMAN received the B.Sc. degree in electrical and computer engineering from Jordan University of Science and Technology, Irbid, Jordan, in 2000, and the M.Sc. and Ph.D. degrees in computer engineering-wireless networks from The University of Alabama in Huntsville (UAH), Huntsville, AL, USA, in 2003 and 2006, respectively. He is currently with the Department of Network Engineering and Security, Jordan University of Science and Technology. His research interests include wireless ad hoc networks, security and key management of wireless networks, wireless sensor networks, the IoT, network coding, mobile edge computing (MEC), fog computing, wireless communication, mobile networks, software-defined networks (SDN), 5G technology, and blockchains.

OMAR A. RAMADAN received the B.Sc. degree in mechanical engineering from Jordan University of Science and Technology, in 2021, where he is currently pursuing the M.Sc. degree in mechatronics engineering. His passion for robotics and automation led him to join MARS Robotics, in 2021, where he serves as a Research and Development Engineer and adeptly takes on the role of Scrum Master. His research interests include advanced control, precise localization, machine learning, and innovative applications in mechatronics, with a special interest in unmanned aerial vehicles (UAVs).

SAMI MUHAIDAT (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2006. From 2007 to 2008, he was an NSERC Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Canada. From 2008 to 2012, he was an Assistant Professor with the School of Engineering Science, Simon Fraser University, BC, Canada. He is currently a Professor with Khalifa University and an Adjunct Professor with Carleton University, Ottawa, ON, Canada. His research interests include advanced digital signal processing techniques for wireless communications, intelligent surfaces, MIMO, optical communications, massive multiple access techniques, backscatter communications, and machine learning for communications. He is an Area Editor of IEEE TRANSACTIONS ON COMMUNICATIONS and a Guest Editor of IEEE NETWORK Special Issue on "Native Artificial Intelligence in Integrated Terrestrial and Non-Terrestrial Networks in 6G." He served as a Senior Editor and an Editor for IEEE COMMUNICATIONS LETTERS, an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, and an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.