drones

MDPI

*Review*

# A Review on Deep Learning for UAV Absolute Visual Localization

**Andy Couturier and Moulay A. Akhloufi ***

Computer Science Department, Science Faculty, Université de Moncton, 18 Antonine-Maillet Ave, Moncton, NB E1A 3E9, Canada; andy.couturier@umoncton.ca
* Correspondence: moulay.akhloufi@umoncton.ca

**Abstract:** In the past few years, the use of Unmanned Aerial Vehicles (UAVs) has expanded and now reached mainstream levels for applications such as infrastructure inspection, agriculture, transport, security, entertainment, real estate, environmental conservation, search and rescue, and even insurance. This surge in adoption can be attributed to the UAV ecosystem's maturation, which has not only made these devices more accessible and cost effective but has also significantly enhanced their operational capabilities in terms of flight duration and embedded computing power. In conjunction with these developments, the research on Absolute Visual Localization (AVL) has seen a resurgence driven by the introduction of deep learning to the field. These new approaches have significantly improved localization solutions in comparison to the previous generation of approaches based on traditional computer vision feature extractors. This paper conducts an extensive review of the literature on deep learning-based methods for UAV AVL, covering significant advancements since 2019. It retraces key developments that have led to the rise in learning-based approaches and provides an in-depth analysis of related localization sources such as Inertial Measurement Units (IMUs) and Global Navigation Satellite Systems (GNSSs), highlighting their limitations and advantages for more effective integration with AVL. The paper concludes with an analysis of current challenges and proposes future research directions to guide further work in the field.

**Keywords:** unmanned aerial vehicles; deep learning; localization; GNSS-denied navigation

check for updates

**Citation:** Couturier, A.; Akhloufi, M.A. A Review on Deep Learning for UAV Absolute Visual Localization. *Drones* **2024**, *8*, 622. https://doi.org/10.3390/drones8110622

Academic Editor: Anastasios Dimou

Received: 16 July 2024
Revised: 26 September 2024
Accepted: 7 October 2024
Published: 29 October 2024

**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

The demand for Unmanned Aerial Vehicles (UAVs) continues to grow at an unprecedented rate. According to Drone Industry Insights, the combined commercial and recreational drone markets are projected to expand from USD 33.7 billion in 2023 to USD 54.6 billion by 2030 [1,2]. In parallel, Teal Group forecasts that worldwide procurement funding for military drones will grow from USD 14 billion in 2024 to USD 23.1 billion by 2033 [3,4]. This translates to a cumulative expenditure of USD 186.8 billion over the next decade, with an additional USD 71.8 billion anticipated for related military research and development. The growth is largely driven by advancements in UAV technology, including edge computing capabilities that facilitate real-time decision making [5,6], and significant improvements in flight endurance [7]. These factors have made UAVs indispensable tools across various industries, expanding their range of applications considerably. The notable sectors experiencing significant increases in UAV utilization include the military [8], border security [9], transport [10], search and rescue [11], wildlife surveying [12], environemental conservation [13], disaster management [14], healthcare delivery [15], commercial delivery [16], advertising [17], firefighting [18], precision agriculture [19], infrastructure inspection [20], urban planning [21], construction [22], surveying and mapping [23], media and entertainment [24], real estate [25], insurance [26], hospitality and tourism [27], and other recreational applications [28].

An indispensable requirement for the usability of UAVs in these applications is the capability for autonomous or semiautonomous operations. Without such a capability,

the practicality and effectiveness of UAVs in these contexts would be significantly compromised. In most applications, human intervention is limited to providing high-level directives to the UAV's navigation system. For example, in a survey and mapping mission, only a few high-level directives are given to the UAV, which then performs the mission entirely autonomously. Other examples include border security, where UAVs can operate 24/7 to continuously monitor for unauthorized crossings and provide real-time alerts to border patrol units [29], and search and rescue missions, where UAVs are programmed to systematically explore large, often difficult terrains to locate missing individuals [30].

To achieve autonomous operations in UAVs, precise and reliable localization is essential. Localization is an estimation process that relies on a combination of sensors, including Global Navigation Satellite Systems (GNSSs), Inertial Measurement Units (IMUs), cameras, and sometime other sensors, to gather data about the surroundings of the UAV and its movements. The majority of recent research works on the subject focus on estimating the localization of UAVs in terms of $(x, y)$ or $(lon., lat.)$, as they are the most crucial variables for autonomous operation. Its necessity stems from the fact that knowing precisely where the UAV is located is crucial for determining where it needs to go next to achieve its objectives. Even for fundamental actions such as obstacle avoidance or maintaining a stable hover, accurate localization is required. Localization allows the UAV to understand its environment, make informed decisions, and achieve high-level directives with minimal human intervention. Therefore, localization is a critical area of study for enabling meaningful autonomous flight capabilities across diverse applications.

The goal of this paper is to provide a comprehensive review of the recent literature on UAV localization, with a specific focus on deep learning techniques for UAV Absolute Visual Localization (AVL). The context of this paper is set during a time when deep learning has surpassed traditional computer vision approaches to become the state of the art in AVL, as it has in various other fields that were previously dominated by computer vision methodologies. Leveraging our extensive experience in the field, which includes continuous research since 2017 and witnessing the evolution of three generations of UAV localization approaches, we aim to equip researchers with the insights and tools required to develop robust localization techniques for GNSS-degraded or GNSS-denied environments.

A particular focus of this paper is on the development of integrated GNSS/IMU/AVL localization solutions, which remain relatively underexplored in the literature. Accordingly, this paper offers detailed background information in Section 2. This section covers the progression of UAV localization, starting from vision-less methods, moving through traditional computer vision approaches, and culminating in the rise in deep learning in the field as detailed in Section 2.1. Understanding these key developments allows researchers to build on prior work and avoid redundant efforts.

Additionally, we provide an in-depth analysis of the current state of IMU and GNSS technologies within localization frameworks. This analysis is presented in Sections 2.2 and 2.3, respectively. Given the ubiquitous presence of these sensors in nearly all UAVs, it is nearly inconceivable to design a UAV without incorporating them. Therefore, researchers working on new vision-based localization methods must inevitably engage with these sensors. Understanding their strengths and limitations is crucial for the development of integrated approaches that leverage the full potential of these technologies.

This paper provides a comprehensive exploration of recent advancements in AVL. In Section 3, we present a detailed overview of the state-of-the-art developments in AVL, including an extensive discussion of the available datasets crucial for training and evaluating these systems (Section 3.1). We categorize and analyze the three principal types of deep learning-based AVL approaches: fine-grained matching methods (Section 3.2), which focus on precise local feature correspondence; large-scale matching techniques (Section 3.3), which address scalability challenges in extensive environments; and hybrid matching strategies (Section 3.4), which combine elements of both to enhance robustness, accuracy, and real-time capabilities. In Section 4, we critically assess the persistent challenges and limitations within the field, highlighting areas where current methodologies fall short.

Finally, Section 5 proposes potential directions for future research, suggesting innovative solutions and methodologies that could address these challenges and advance the field.

Previous related works include our earlier review published in 2021 [31] and a similar review by Ali et al. in 2022 [32]. Both reviews primarily covered AVL literature from 2015 to 2019, including only a few initial works using deep learning for AVL, and thus, they lack coverage of the modern deep learning approaches that have since emerged. In contrast, our review focuses on the literature from 2019 onward and specifically on modern deep learning approaches for AVL. In 2022, Nasser et al. proposed a review on GNSS-independent UAV navigation techniques, featuring a high-level overview of perception, localization, and motion planning. However, their localization section mainly reviewed traditional computer vision methods and briefly mentioned only one work using learned features. Our work takes a more detailed approach by diving deeper into the concept of localization and providing extensive reviews of deep learning contributions in the literature. Also in 2022, Rezwan and Choi presented a comprehensive review of artificial intelligence approaches for UAV navigation [33]. However, their paper concentrated on navigation and path planning and did not cover localization. In 2023, Arafat et al. offered a review on vision-based navigation techniques for UAVs [34], addressing aspects such as mapping, obstacle avoidance, and path planning. However, the work has minimal coverage of localization and no coverage of deep learning-based localization techniques. Recently, Wang et al. proposed a scoping review on UAV-based Simultaneous Localization And Mapping (SLAM) in outdoor environments [35]. As a scoping review, it covers the literature at a higher level compared to our systematic review and only covers a few deep learning approaches, as they are less prevalent in the field of SLAM. Other related works on localization were published before 2020 and thus predate the modern deep learning techniques for localization [36–39].

Following these observations, the key differences and contributions of this paper are as follows:

- The first comprehensive review specifically dedicated to deep learning techniques for UAV AVL;
- A review of the main research developments retracing the evolution from the initial UAV localization methods without vision to modern deep learning approaches;
- A detailed examination of the current state of IMU and GNSS technologies in the optic of their integration with AVL solutions;
- An in-depth analysis and comparison of the key contributions in the AVL literature since 2019;
- A thorough inventory of publicly available datasets and source codes for AVL;
- The most comprehensive review of datasets ever contributed in the field of AVL;
- A critical discussion of the reviewed works, highlighting their limitations and the persisting challenges within the field;
- A list of potential future research directions for advancing UAV AVL technology.

## 2. Background

Recent advancements in localization and sensor technologies for Unmanned Aerial Vehicles (UAVs) have significantly influenced the design and development of new localization solutions. In terms of localization, improvements in semiconductor technology have progressively enhanced the performance and energy efficiency of edge computing devices [40,41]. These advancements have enabled researchers to increase the computational power of their localization solutions without compromising flight endurance. Consequently, research in localization has evolved from only integrating Inertial Measurement Unit (IMU)/Global Navigation Satellite System (GNSS) data to incorporating traditional computer vision techniques, and ultimately, to employing deep learning methods. This progression is elaborated in Section 2.1.

Regarding sensor technology, IMUs and GNSS remain unavoidable components in the design of UAVs [42–45]. In recent years, IMUs have experienced a remarkable generational

leap in performance, profoundly affecting the design of localization solutions. These advancements, along with the remaining challenges for IMU technology for localization, are discussed in Section 2.2.

The limitations of GNSS technology have traditionally driven the development of vision-based localization technologies, such as Absolute Visual Localization (AVL). However, recent developments, such as the deployment of advanced satellite constellations like Galileo in 2016 [46], and the introduction of message authentication technologies like Galileo Open Service Navigation Message Authentication (OSNMA) [47], are beginning to address these limitations after more than 30 years of slow progress. The drawbacks of GNSS technology and recent advancements that will inevitably impact the future design of localization solutions are detailed in Section 2.3.

### 2.1. Key Developments in UAV Localization

Early work in the field of UAV localization performed data fusion of GNSS and IMU data using sequential Bayesian filtering techniques such as Extended Kalman Filters (EKFs) [48–50]. The theoretical basis for this approach is elegant, as GNSS and IMUs are complementary in nature. GNSS provides absolute positioning, offering good long-term accuracy but poor short-term precision due to its limited update rate of 10 to 20 readings per second [51]. In contrast, IMUs offer high-frequency relative positioning with up to 32,000 readings per second [52], but suffer from an error accumulation that grows cubically in time [53]. The fusion of GNSS and IMU data aims to leverage the strengths of both systems: the IMU provides accurate short-term positioning, while the periodic GNSS updates correct the IMU drift and upgrade the solution from relative (x, y, z) to absolute coordinates (longitude, latitude, altitude). However, practical implementations of this approach are challenging or impractical for many applications due to the inherent limitations of GNSS and IMUs. The specific shortcomings of these two technologies are comprehensively discussed in Sections 2.2 and 2.3. To address these issues, modern UAVs now integrate vision-based systems alongside GNSSs and IMUs to improve localization accuracy and robustness [54,55].

Early research on integrating vision for UAV localization primarily concentrated on Visual Odometry (VO) and Simultaneous Localization And Mapping (SLAM) [39,56–65]. These methods are collectively referred to as Relative Visual Localization (RVL) [31]. Initially successful with Unmanned Ground Vehicles (UGVs), these methodologies were later adapted for UAVs [66–71]. However, they encounter significant challenges related to error accumulation, although to a lesser extent than with IMU-based localization. SLAM algorithms, despite being capable of leveraging multiple previous observations through techniques like keyframes and bundle adjustment, still rely on prior estimations for both landmark positions and UAV poses [72]. This dependence can degrade the system's performance, particularly in the absence of loop closure opportunities in large outdoor environments where UAVs are predominantly deployed [31]. Similarly, VO relies on the recursive use of past estimations to determine the current position, resulting in drift over time [73]. Although integrating inertial data with visual information, as in Visual-Inertial Odometry (VIO), can enhance accuracy [74], it does not entirely eliminate the root cause of cumulative errors [31]. This drift issue is particularly problematic in applications requiring long-duration and high-endurance operations, such as in surveying as well as defense and security scenarios. In these contexts, UAVs often operate for extended periods and cover vast areas, making bounded error a critical requirement. Furthermore, these approaches fall short when the GNSS is compromised, as it is the only source of periodic absolute positioning.

Recent research has started to explore the use of AVL, an approach specifically designed for UAVs [31]. AVL determines the position of a UAV by comparing its current camera view with previously collected and orthorectified data typically obtained from satellite imagery or aerial images from prior flights. The primary benefit of AVL is its immunity to drift over time [75]. This is achieved by continuously matching the current view of the UAV against a stored visual memory, thus providing fresh localization estimates

with each reading. Unlike systems where new estimations rely on previous observations or estimates, AVL does not and is considered an absolute source of positioning. Therfore, AVL shares more characteristics with GNSS, in contrast to VO which is the vision-based equivalent of IMU localization. However, AVL does not rely on external systems to operate, making it superior in that regard, as it is not vulnerable to interference. Like any localization approaches, AVL have its challenges which include handling large amounts of data and accurately matching images that may differ in terms of source, time of capture, lighting conditions, and resolution.

Early research efforts in AVL have made extensive use of traditional computer vision techniques as detailed in our previous review paper on the subject published in 2021 [31]. Such techniques relied on template matching [76–79] and feature point matching [80–88]. Template matching is considered a dense matching approach, where the current view of the UAV is used as a template to be searched within a reference map using a patch comparison operator. Patch comparison techniques such as normalized cross-correlation [76], mutual information [77], phase correlation [78], and normalized information distance [79] were used to compare UAV images to global maps to resolve its position. In contrast, feature point matching is a sparse matching approach where salient points are detected, often using corner detectors [89,90], with the objective of being identifiable across different, possibly independent, image captures. Descriptors are then extracted for each feature point using intensity values in their surrounding area in the image. These descriptors can then be leveraged to match feature points between two images, thus creating a basis for similarity comparison. Feature points methods used for UAV localization include SURF [80,84], abBRIEF [87], and ORB [80–82]. Most of these feature point-based localization approaches were used in combination with statistical filtering techniques such as particle filters [81,82] to create complete localization solutions.

At the time of our previous review paper's publication, deep learning was just beginning to be explored for UAV localization. Early attempts included using feature maps from specific layers of pretrained (not retrained) Convolutional Neural Networks (CNNs) as the basis for image similarity [91], and employing semantic segmentation via U-Net to create building and road masks for image alignment and matching [92,93]. Efforts also included using multi-task CNNs to simultaneously perform road segmentation, infer longitude and latitude coordinates, and generate a segmentation map indicating the UAV position on global maps [94]. Other approaches combined VGG16 [95] with a differentiable implementation of the Inverse Compositional Lucas-Kanade (ICLK) algorithm to achieve localization [96], while some researchers leveraged conditional Generative Adversarial Networks (cGANs) to transform UAV images into ones more similar to satellite imagery to facilitate template matching [97]. These initial methods were highly varied and lacked consensus on the most effective way to adapt deep learning for UAV localization. They faced several drawbacks including inferior localization accuracy compared to traditional computer vision techniques and significant computational complexity which hindered real-time processing. Additionally, CNNs trained for specific geographical regions exhibited limited applicability in diverse real-world localization scenarios. These deep learning approaches encountered many challenges mainly related to the difficulty of developing end-to-end architectures for localization and complications related to the use of dynamic input data with CNNs, which are typically trained offline. Efforts were further impeded by the need to learn large volumes of data to recognize all potential locations across global maps and the absence of large public datasets necessary to train effective models.

Recently, the initial drawbacks and challenges associated with deep learning for UAV localization have, in great part, been resolved. The current state-of-the-art approach using metric learning was proposed by three independent research groups over a short period of time [98–100]. Valaitis et al. published their paper in June 2019 [98], followed by Ahn et al. in January 2021 [99], and Couturier and Akhloufi in April 2021 [100]. Other contributions following this new approach were only published later towards the end of 2021 [101,102]. In the first three works, the triplet loss was used to train CNN models

for image similarity matching [103,104]. This approach was inspired by state-of-the-art approaches in face recognition at the time [105]. Couturier and Akhloufi were the first to propose a comprehensive UAV localization approach using metric learning and the first to test such a solution on real UAV imagery. A detailed review of these approaches as well as the current literature on AVL is presented in Section 3.

### *2.2. Inertial Measurement Unit (IMU)*

IMUs play a pivotal role for UAV localization by providing several types of sensors, including gyroscopes, accelerometers, and magnetometers in UAVs. These core sensors are often supplemented by barometers and ultrasound sensors to enhance height measurements [54]. Modern IMUs, such as the TDK InvenSense ICM-40609-D, feature accelerometers and gyroscopes capable of sampling rates as high as 32 kHz, allowing for extremely high temporal resolution [52]. For this reason, IMUs have been the primary source of high-frequency pose data in UAVs. This has made IMUs indispensable for six Degrees of Freedom (DoF) pose estimation and particularly in estimating orientation components such as yaw, pitch, and roll [39,106–109]. Furthermore, IMUs have the ability to function independently of external data sources (unlike GNSS), making them the most useful sensors to integrate with vision-based localization solutions that aim to perform localization in GNSS-degraded or GNSS-denied environments [80–83,100]. Understanding the strengths and limitations of IMUs is crucial for researchers developing advanced localization solutions, as it ensures that this potent source of precise localization data is fully utilized.

When extracting localization data from IMUs, some sensor data require more computational processing than others to yield meaningful information. For instance, accelerometer data have to be numerically integrated twice to calculate displacement and estimate the position over time [110]. Similarly, gyroscope data must be integrated once to determine the UAV rotation to estimate its orientation over time [111]. These computations are susceptible to various types of errors, such as bias, bias drift, scale factor error, scale factor drift, misalignment, and random sensor noise, all of which can affect the accuracy of the measurements [112–115]. Such errors occurring on each reading lead to an accumulation of errors over time, generally referred to as drift [116,117]. The effect of this drift is amplified by the necessary computations to obtain the desired value, leading to an error with a cubic growth rate in relation to time [53]. This remains the main challenge faced when using IMU data for localization.

To illustrate the magnitude of this phenomenon, we can examine the case of the accelerometer, which is the IMU sensor most susceptible to drift. When using an accelerometer, at least two readings are required to estimate displacement along a single axis when using the midpoint rule for integration [118]. Consequently, to calculate the total displacement along both the x and y axes, a minimum of four readings are necessary. Equation (1) demonstrates how the velocity at the $i$th iteration is computed using the midpoint rule with two readings, $a_i$ and $a_{i-1}$, over the time interval $\Delta t$. Equation (2) illustrates the subsequent integration of velocity using the midpoint rule to obtain displacement $d_i$ at the $i$th iteration. The general formula for displacement along an axis, assuming an initial displacement of 0 m and an initial velocity of 0 m/s, is presented in a summation form in Equation (3). Finally, Equation (4) provides the expression for the total displacement in both the x and y axes using the Euclidean distance and Equation (3):

$$v_i = v_{i-1} + \frac{a_i + a_{i-1}}{2} \cdot \Delta t \tag{1}$$

$$d_i = d_{i-1} + \frac{v_i + v_{i-1}}{2} \cdot \Delta t \tag{2}$$

$$d_i = \sum_{k=0}^{i} \sum_{j=0}^{k} \frac{a_j + a_{j+1}}{2} (\Delta t)^2 \tag{3}$$

$$d_{\text{total},i} = \sqrt{\left(\sum_{k=0}^{i}\sum_{j=0}^{k}\frac{a_{x,j}+a_{x,j+1}}{2}(\Delta t)^2\right)^2 + \left(\sum_{k=0}^{i}\sum_{j=0}^{k}\frac{a_{y,j}+a_{y,j+1}}{2}(\Delta t)^2\right)^2} \quad (4)$$

In our last review of the literature on AVL, written in 2020, we presented an accelerometer simulation under a worst-case scenario, where a systematic error of 50 mg (0.49 m/s$^2$) occurred at each reading [31]. Since then, some quadcopter manufacturers have increased transparency by publishing error rates after applying correction algorithms to their readings. For instance, the TDK InvenSense MPU-6000, a previous-generation IMU, is used in the Parrot Anafi USA quadcopter. This sensor has an initial calibration tolerance in the x and y axes of ±50 mg (0.49 m/s$^2$) [119], though Parrot reports an improved accuracy of ±15 mg (0.15 m/s$^2$) after compensation as detailed in their white paper on the Parrot Anafi USA quadcopter [55]. The newer Parrot ANAFI Ai is equipped with the TDK InvenSense ICM-40609-D, the latest generation of IMU for the drone market [120,121]. Note that this chip is also used in many other drones on the market such as the DJI Mini 3 [122] as demonstrated by iFixit [123]. This sensor has an initial calibration tolerance in the x and y axes of ±40 mg (0.39 m/s$^2$) [52]. Parrot reports an impressive accuracy of ±0.5 mg (0.005 m/s$^2$) after thermal and dynamic calibration [54]. This represents a significant improvement from the previous generation of sensors.

Using this information and Equation (4), we simulated the total displacement error (x, y) for the TDK InvenSense MPU-6000 and ICM-40609-D. The reading rate was set at 1 kHz for the MPU-6000 and 2 kHz for the ICM-40609-D, matching the reading rates reported by Parrot [54,55]. The results are shown in Figure 1 with two different scalings of the axes over the same data. For each sensor, the worst-case error reported by the sensor manufacturer ("MPU-6000 Worst" and "ICM-40609-D Worst") as well as the error reported by Parrot ("MPU-6000 Parrot" and "ICM-40609-D Parrot") were simulated. In these simulations, we systematically applied the reported error values to each reading.

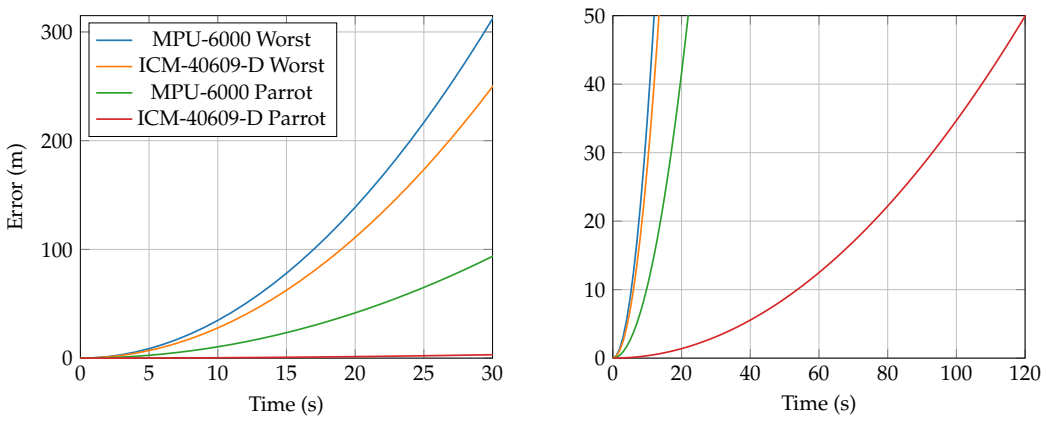


**Figure 1.** Simulation of IMUs displacement error over time.

It is important to note that the results presented reflect only the accumulation of errors in displacement estimation. These results do not directly correspond to the total localization error, which also involves additional factors such as heading errors arising from gyroscope and magnetometer measurements, as well as the specifics of the data fusion algorithms employed, which are beyond the scope of this analysis. Nevertheless, our findings provide a reliable indication of IMUs' performance in localization tasks. Since errors in heading measurements would further increase the overall localization error, the displacement errors reported here can be considered a lower bound on the expected localization error.

The first thing that stands out in the results is the significant improvement in Parrot's displacement error with the latest generation of IMU sensors and correction algorithms. The quality of these correction algorithms appears to have a substantial impact on the accuracy of the estimations. Notably, with the latest generation, the accumulated error

reaches 50 m after 120 s, whereas the previous generation reached the same error level in just 22 s. However, despite these advancements, error accumulation over time remains substantial in all cases, rapidly increasing to a level that makes an IMU alone unusable for localization. This is the main takeaway from this experiment.

The current state of IMU technology is encouraging, demonstrating remarkable generational leaps in recent years. However, the drift error associated with these solutions has not yet reached a level where IMU-only localization is feasible. Thus, other bounded estimation approaches, such as GNSS and AVL, are still necessary for robust localization solutions. Nonetheless, the significant improvements in IMU technology will impact the design of localization systems.

Primarily, these improvements mean that the frequency of updates from absolute localization sources can be reduced. In the context of integrating IMUs with AVL solutions, the possibility of higher-latency operation would provide more time for AVL algorithms to process each estimation. This extra time could be leveraged to run more sophisticated and accurate AVL solutions, such as larger and deeper deep learning models, resulting in better localization estimates than those possible with previous generations of IMUs.

Alternatively, these improvements could allow UAV designers to reduce power consumption by running fewer AVL estimations or to use lower-performance computing modules to lower manufacturing costs. Both are valid design decisions for integrating robust localization capabilities in UAVs.

### *2.3. Global Navigation Satellite System (GNSS)*

Integrating GNSS technology into UAV localization solutions requires the precise measurement of the time of flight for signals transmitted from satellites to GNSS receivers to ensure accurate positioning. However, this reliance on external signals introduces vulnerabilities to various types of signal interference, which can degrade performance or render such systems inoperable. Exploiting these vulnerabilities, it has been shown that UAVs using tightly-coupled GNSS/IMU localization can have their positioning spoofed in a covert manner, even defeating advanced integrity protection mechanisms such as Least-Squares Residual Receiver Autonomous Integrity Monitoring (LSR-RAIM) [124]. These vulnerabilities remain a significant challenge in integrating GNSS into localization solutions and serve as a primary motivation for the development of AVL solutions.

Despite these challenges, GNSS technology is ubiquitously used in UAVs, and most real-world applications of AVL solutions are expected to be deployed alongside GNSS as an additional source of absolute positioning. This dual approach is designed not to replace GNSS but to detect and counteract GNSS attacks [125–129]. For this reason, designing effective AVL solutions requires a thorough understanding of the error types and limitations of GNSS technology, as these systems will most likely need to interact together in real-world scenarios.

In many operating environments, naturally occurring interference can significantly impact GNSS performance. This type of interference has been studied in diverse contexts such as natural canyons [130], urban canyons [131], and forests [132,133]. The different types of naturally occurring interference are illustrated in Figure 2.

Signal obstruction is the most apparent issue that occurs when UAVs navigate through environments with obstacles such as buildings, mountains, or dense foliage. These obstacles block the direct Line Of Sight (LOS) between the satellite and the GNSS receiver, resulting in the absence of GNSS signals, leading to a loss of critical positioning data.

Additionally, multi-path interference is a common problem when UAVs operate near reflective surfaces such as large buildings or bodies of water. In this scenario, GNSS signals may bounce off these surfaces before reaching the UAV, causing the receiver to process both the reflected and the direct signal. This leads to discrepancies in the calculated time of flight and results in positional inaccuracies.

Furthermore, Non Line Of Sight (NLOS) situations arise when GNSS signals reach the UAV receiver indirectly due to obstructions blocking a LOS path. This can cause positioning

errors that are potentially unbounded but always positive. In such cases, errors can reach the scale of several kilometers when signals are reflected by tall, distant buildings [134].

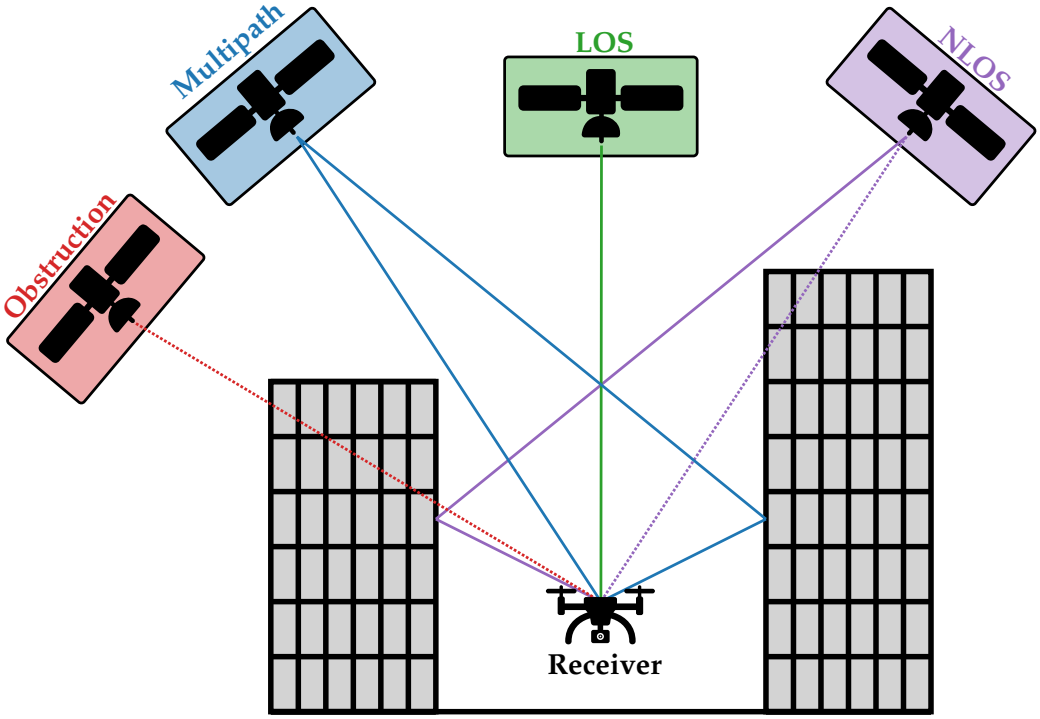


**Figure 2.** Types of GNSS signal reception.

GNSS technology is also susceptible to artificial interference generated by malicious actors which can negatively impact GNSS performance and UAV localization. Jamming is the most common and consists in the intentional broadcasting of signals at the same frequency as GNSS signals to block their reception. This interference can create either narrowband or wideband disruptions, making it difficult or impossible for GNSS receivers on UAVs to decode satellite signals accurately [135]. Such interference can lead to significant operational issues and cause a UAV relying on GNSS to lose its navigation capabilities, therefore grounding it [136]. Since February 2022, the real-life use of jamming by Russia has been extensive due to the full-scale invasion of Ukraine [137]. While the aim is presumably to disrupt Ukraine's UAVs, missiles, and other military equipment navigation, thousands of commercial airline flights in Europe and the Baltic region have also been impacted, causing significant interruptions to civil aviation operations [137,138]. The severity of the situation has led the European Union Aviation Safety Agency (EASA) to release multiple advisories on navigation interference [139]. The situation even led to the development of multiple live online maps leveraging Automatic Dependent Surveillance-Broadcast (ADS-B) signals to track Global Positioning System (GPS) jamming in near real-time [140–142]. Multiple actors across the industry, including the EASA, have observed a rise in the sophistication and danger of these attacks [139,143]. Additionally, some have called on the United States to expedite the modernization of the GPS constellation to help alleviate such issues [144]. Compared to other satellite navigation systems like China's nearly twice as large BeiDou and Europe's more modern Galileo, GPS is lagging in modernization and security improvements [144].

Another form of signal interception is meaconing, which is similar but requires simpler equipment than what is needed for a spoofing attack [145]. This attack involves capturing the original GNSS signal, delaying it, and then retransmitting it [145]. As a result, the UAV receiver misinterprets the timing and positional data, leading to incorrect navigation information [146,147]. Meaconing attacks have seen renewed attention in recent years due

to the rise in Navigation Message Authentication (NMA) deployments, as unlike spoofing, meaconing is not impacted by NMA [148].

Spoofing is the most advanced attack category used against GNSS navigation. An illustration of this concept is provided in Figure 3. The aim of this technique is to make a UAV believe it is in a different location than it truly is. It involves transmitting simulated (fake) GNSS signals to deceive the GNSS receivers, leading them to calculate incorrect positions that can be chosen by the attacker. GNSS spoofing can have very significant real-life implications. The most well-known high-profile case occurred on December 5, 2011, when an American Lockheed Martin RQ-170 Sentinel UAV was captured by Iranian forces. It was announced by the Iranian government that the UAV had been brought down using GPS spoofing [149]. This claim was initially denied by the United States government, but President Obama later acknowledged that the downed aircraft was indeed a United States (US) drone [150]. Multiple reports confirm that the RQ-170 was reverse engineered by Iran, resulting in the production of drones such as the Shahed 171 Simorgh and Shahed Saegheh [151–154]. In the past few years, spoofing attacks have become increasingly complex and diversified, now leveraging techniques such as self-consistent spoofing to synthesize false signals, signal estimation, and replay spoofing to jam or predict and rebroadcast secure codes, and advanced-form spoofing like zeroing to cancel real signals and avoid detection [155]. These advanced methods are believed to have been used in numerous real-life incidents, notably during the Russian invasion of Ukraine [156,157].

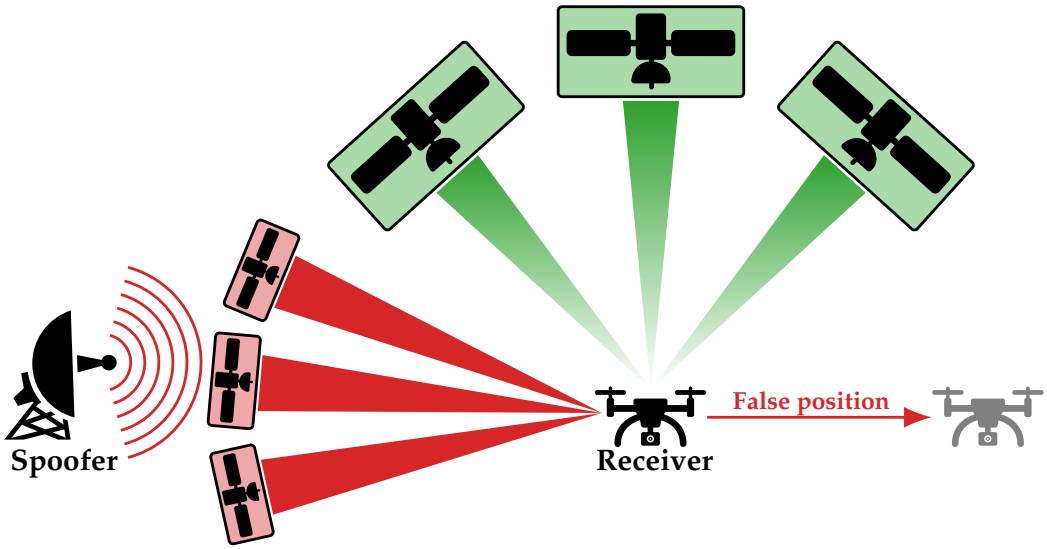


**Figure 3.** GNSS spoofing.

At the time of writing our previous review on AVL in 2020, modern anti-spoofing technologies for GNSS were still far from deployment [31]. However, in recent years, there have been significant advancements in this area [158–161]. It is crucial to acknowledge these advancements, as they will impact the way AVL solutions are used in UAVs in the future.

Spreading Code Authentication (SCA) and NMA are the two main GNSS techniques aimed at enhancing the security of satellite signals by ensuring their authenticity and integrity [160–162].

SCA involves securing spreading codes, or PseudoRandom Noise (PRN) codes, which are unique binary sequences transmitted by each GNSS satellite. These codes enable receivers to identify signals from different satellites and calculate the time of flight of each signal [163]. The prevalent method for implementing SCA involves encrypting the spreading codes, where satellites encrypt these codes and receivers subsequently decrypt them using specific cryptographic keys [164]. Provided that the decryption keys are securely maintained and the encryption method's details remain confidential, this method can effectively authenticate the spreading codes. Additionally, approaches in development

like CHIp-MEssage Robust Authentication (CHIMERA) are exploring the insertion of unpredictable watermarks into the GNSS spreading codes [165]. These watermarks are meant to be seamlessly integrated into the spreading codes, thereby minimally affecting their function and allowing legacy receivers to continue using the signals without requiring authentication [166].

NMA enhances the security of navigation messages, which are data streams transmitted by GNSS satellites containing critical information such as the satellites' position (ephemeris data), clock corrections, and system health status [163]. This technique involves appending cryptographic signatures generated using private keys to the navigation messages. Receivers can use corresponding public keys to validate these signatures, thereby verifying the authenticity and integrity of the content [167].

As researchers specializing in AVL solutions, our team has explored the integration of AVL systems in UAV platforms, particularly in environments where GNSS signals are compromised. A critical question in such integrations is the reliability of GNSS data and whether utilizing multiple constellations simultaneously could mitigate inherent GNSS limitations. In our research, we encountered challenges in sourcing concise and up-to-date information on the current capabilities and reliability of various GNSS constellations. This lack of detailed information hindered the speed and effectiveness of our feasibility analyses on several projects. While we found resources detailing the advanced capabilities of state-of-the-art multi-constellation GNSS sensors [168], the literature reviews we encountered were predominantly centered on academic research and did not sufficiently cover the latest real-world deployments [159]. Given these gaps, determining the optimal depth of integration between AVL and GNSS systems for effective real-world applications was difficult. To better support researchers and practitioners in the field, we compiled Table 1, which provides a comprehensive overview of the latest developments in SCA and NMA technologies across existing GNSS constellations. This table aims to provide a foundational reference to inform decisions regarding the integration of GNSS and the selection of GNSS constellations in localization solutions tailored to UAV platforms operating under challenging GNSS conditions.

To provide further insight into the table, we detail the latest advancements and developmental timelines of SCA and NMA technologies across different GNSS constellations. Currently, all known deployed SCA schemes are encryption-based and reserved for sensitive military and governmental applications (closed), including GPS P(Y) Code [169] and M-Code [170]; Galileo PRS [171]; GLONASS HP/VT Code [172]; BDS Authorized Service [173]; QZSS PRS [174]; and NavIC RS [175]. Galileo CAS [176] and ACAS [176,177] are encryption-based SCA schemes that are in development and are going to be available for commercial users in the future. In recent years, there has been a marked increase in advancements related to NMA schemes, significantly outpacing those in SCA technologies. In 2016, the European Union was the first to announce an NMA service plan for Galileo in order to provide OSNMA for all of its users [178–180]. While a declaration of service planned in 2023 seems to have been delayed multiple times [181–183], the service has been in a public testing phase since 2021 and is already accessible [47]. India started hinting at the development of an NMA system for its regional Navigation with Indian Constellation (NavIC) satellite system in 2019 [184] and confirmed that testing has been in progress on satellites since 2023 [185]. Japan announced the development of NMA for its regional Quasi-Zenith Satellite System (QZSS) in January 2022 [186]. The system was in a testing phase from July 2023 [187] until the start of the service was announced in April 2024 [188], and it is now fully accessible. A recent presentation at the UN Workshop on the Applications of GNSS in April 2024 by Sviatoslav Zhilenko from Roscosmos hinted at the possibility for Russia to introduce L1 authentication after 2030 for its GLObal NAvigation Satellite System (GLONASS) [189]. The type of authentication (NMA or SCA) as well as the name for the proposed technology is presently unknown. In 2017, Anderson et al. proposed CHIMERA for the United States' GPS [165]. The approach has been in development with testing on the L1C signal by the Air Force Research Laboratory since 2019 [190]. The new scheme is

anticipated to be broadcasted by the Navigation Technology Satellite 3 (NTS-3) after its launch in late 2024 [191]. While many studies exist for the potential integration of NMA in the BeiDou Navigation Satellite (BDS) system, no such development has yet been publicly announced [158,192,193].

**Table 1.** Summary of GNSS constellations with authentication schemes and usage information.

| Constellation | Authority | Coverage | Name | Type | Status/Access |
|---|---|---|---|---|---|
| GPS | United States | Global | P(Y) Code [169] | SCA | Closed |
| | | | M-Code [170] | SCA | Closed |
| | | | CHIMERA [190] | Hybrid SCA/NMA | Development |
| Galileo | European Union | Global | PRS [171] | SCA | Closed |
| | | | CAS [176] | SCA | Development |
| | | | ACAS [176,177] | SCA | Development |
| | | | OSNMA [178–180] | NMA | Open |
| GLONASS | Russia | Global | HP/VT Code [172] | SCA | Closed |
| | | | Unknown [189] | SCA or NMA | Development |
| BDS | China | Global | Authorized Service [173] | SCA | Closed |
| QZSS | Japan | Regional | PRS [174] | SCA | Closed |
| | | | QZNMA [188] | NMA | Open |
| NavIC | India | Regional | RS [175] | SCA | Closed |
| | | | NMA for NavIC SPS [185] | NMA | Development |

While GNSS reliability is steadily improving, the need for alternative absolute positioning sources like AVL is unlikely to disappear entirely. The current circumstances underscore their critical importance more than ever. Notably, during the Russia–Ukraine war, Ukraine deployed the first AVL solutions on a battlefield to counter Russia's jamming and spoofing of GNSS signals [194]. Given the slow pace of modernization of existing GNSS constellations, this reliance on alternative solutions is expected to persist for the foreseeable future. However, as GNSS signal integrity improves, it may influence how localization solutions are designed, allowing for increased trust in these signals. For this reason, we recommend that researchers in the field continue to closely monitor GNSS developments.

## 3. Current Elements and Methods in Absolute Visual Localization (AVL)

Absolute Visual Localization (AVL) is a distinctive localization method tailored for Unmanned Aerial Vehicles (UAVs). This technique leverages georeferenced global maps, which can either be satellite based or derived from prior UAV flights often through photogrammetry, to determine the position of the UAV. AVL operates by matching the current view from a typically downward-facing camera on the UAV to the reference map to determine the precise localization of the UAV. This process involves capturing images from the camera of the UAV and comparing these images with pre-stored reference images from the global map in real-time. Advanced algorithms identify key features and landmarks in both sets of images, enabling the system to correlate the current view with the appropriate map section accurately. By leveraging the georeference metadata embedded in the reference images, the system can then derive accurate localization estimates. Unlike Relative Visual Localization (RVL), which accumulates errors over time and is dependent on prior observations or estimations, AVL does not rely on past data. It can determine the localization from a single observation, making AVL inherently drift-free. AVL is therefore considered an absolute source of localization, similarly to Global Navigation Satellite Systems (GNSSs). The process is visually demonstrated in Figure 4, where the current UAV image $I_1$ is matched

with the corresponding image $I_2$ extracted from the reference map $R$ using a Convolutional Neural Network (CNN), $f(\cdot)$.

AVL is a highly effective localization method but faces significant challenges, primarily due to the task of matching imagery that can vary widely. Even if the view of the UAV is matched against its prior flights, discrepancies in lighting, weather conditions, seasons, viewing angles, altitude, and motion blur are likely to arise. Although these differences are less significant compared to images from entirely different sources, they still pose substantial challenges for accurate localization.

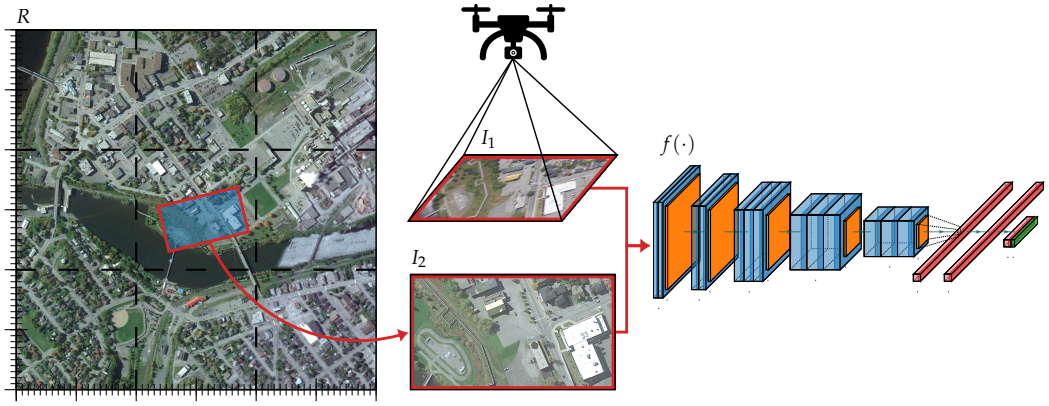


**Figure 4.** State-of-the-art Absolute Visual Localization (AVL) framework.

The complexity increases in real-world scenarios where UAV views are often compared with imagery from disparate sources such as satellites, reconnaissance aircraft, or other UAVs specialized in mapping, reconnaissance, and photogrammetry. Matching images from different airborne sensors introduces another layer of difficulty, even when using the same Complementary Metal-Oxide Semiconductor (CMOS) sensor technology. Variations in sensor size, resolution, dynamic range, noise, and noise profiles can all negatively impact the matching accuracy.

The challenge becomes even greater when comparing UAV views with satellite imagery. Unlike CMOS sensor images that only require debayering and denoising, satellite images undergo extensive processing. A key component of this process is orthorectification, which corrects distortions in imagery caused by the sensor's angle, the Earth's topography, and the curvature of the Earth [195,196]. By using accurate elevation models, like Digital Elevation Models (DEMs), orthorectification accounts for the variation in terrain to reposition image pixels correctly, ensuring that each pixel represents its true geographic location on the Earth's surface [197]. The process involves multiple steps, including adjusting for sensor tilt, terrain-induced displacement, and optical distortions from the sensor system. This process ensures a uniform scale across the image, accurately representing distances and areas and creating geospatially accurate nadir images.

Furthermore, satellite imagery processing often includes panchromatic sharpening, a data fusion method where high-resolution panchromatic (greyscale) images are combined with lower-resolution multi-spectral images to produce higher-resolution color images [198,199]. Panchromatic sharpening is frequently used on imagery from satellites such as Landsat 7/8, SPOT 6/7, and those operated by DigitalGlobe to improve image resolution [200–202]. While this technique enhances image resolution, it can introduce spectral distortion that may further impact image matching.

To address the challenges of matching imagery with significant variations, modern AVL approaches rely on CNN and Vision Transformer (ViT) architectures. These models are particularly effective in analyzing and matching image features because they can learn hierarchical representations directly from raw image data. CNNs, for instance, are adept at detecting and recognizing features such as edges, textures, and patterns at different levels of abstraction, making them robust against variations in lighting, angles, and noise [203]. Their

ability to automatically learn and optimize feature extraction surpasses that of traditional computer vision methods, which rely on manually designed features and often result in less discriminative power in complex environments [204–207].

ViT architectures further enhance this capability by incorporating attention mechanisms, which allow the model to focus on specific parts of an image that are most relevant to the task at hand [208]. This selective attention is particularly useful in scenarios with complex imagery, such as those encountered in UAV and satellite data matching. ViTs have demonstrated superior performance in tasks requiring long-range dependencies and global context, which are essential for accurately matching images with varying scales, resolutions, and perspectives [209–212].

Training CNN and ViT architectures on large, diverse datasets is essential for developing effective and accurate models [95,213,214]. Exposure to diverse samples strengthens these models' generalization ability and enhances their robustness to variations in image features and contexts. Therefore, Section 3.1 describes the publicly available datasets that are crucial for building robust AVL systems.

Recognizing the inherent complexity of the AVL problem, recent research has explored several avenues. After thorough analysis of the current literature, it was decided to categorize the current approaches in relation to their image matching techniques. This categorization stems from the fact that in learning-based AVL, image matching approaches are the central focus. The subsequent sections categorize recent developments into three key categories: fine-grained matching, which focuses on keypoints and pixel correspondences (discussed in Section 3.2); large-scale matching, which utilizes global image descriptors similar to those used in Content-Based Image Retrieval (CBIR) (covered in Section 3.3); and hybrid approaches, which integrate both techniques in a single pipeline (presented in Section 3.4). Figure 5 illustrates this categorization alongside the key concepts adjacent or related to these categories.

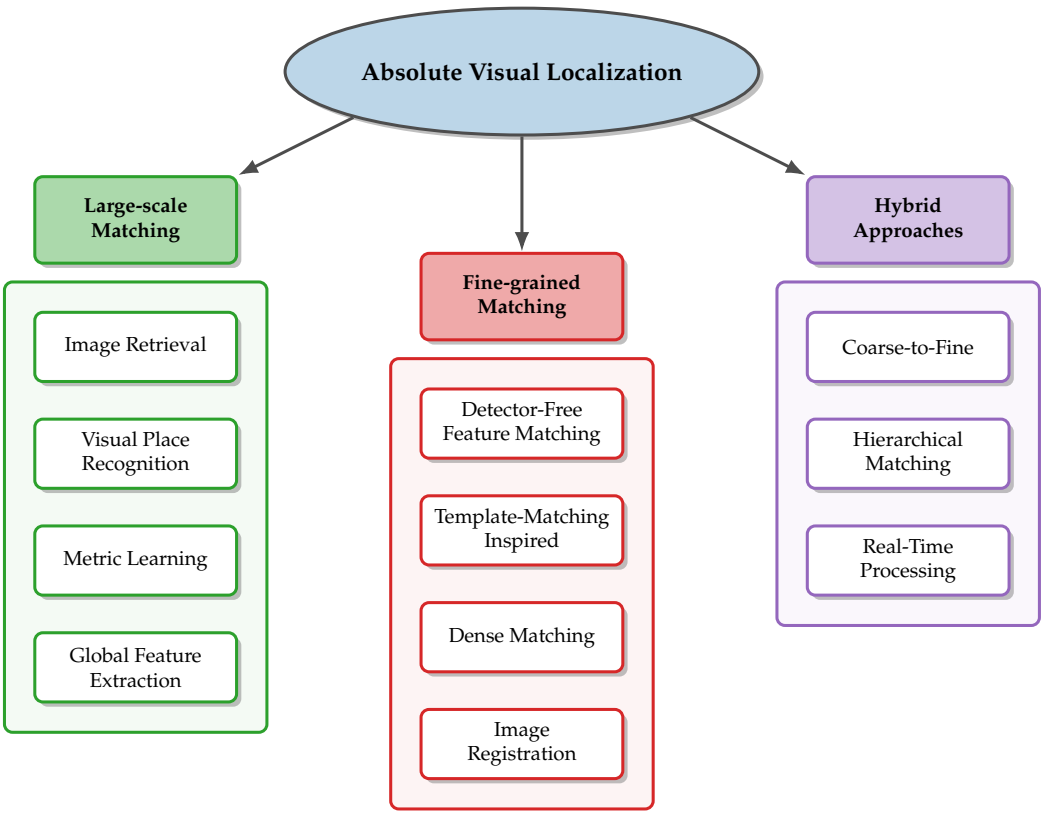


**Figure 5.** Absolute Visual Localization (AVL) methods and associated concepts.

### *3.1. Datasets*

In recent years, the field of learning-based AVL has seen a significant influx of new public datasets, greatly facilitating research and development. Currently, there are at least 12 available datasets, including 10 specifically for AVL and 2 for Cross-View GeoLocalization (CV-GL), all of which are listed in Table 2. CV-GL presents an adjacent concept to AVL, focusing on identifying the geographical location of a ground-captured query image by matching it with an aerial or satellite image of the same area. Notable datasets like University1652 [215] and SUES-200 [216] address this challenge by incorporating UAV imagery, inadvertently allowing for generating matching pairs of UAV and satellite images for AVL purposes. Consequently, these datasets are included in our list of datasets.

Among the AVL datasets, some stand out due to their unique attributes. For instance, specific datasets incorporate thermal imagery [217,218], enhancing their utility for specialized applications. Additionally, one dataset even features imagery from a Martian lander [219]. The dataset sizes vary significantly, ranging from a modest 67 MB to 121 GB. The Boson-nighttime extended dataset is particularly noteworthy for its size, offering 121 GB of real RGB satellite and infrared UAV images, which are invaluable for research in nighttime operations and low-visibility conditions.

**Table 2.** Publicly available datasets.

| Name | Release Date | Type | Size | Paper | Website | Access | Availability |
|---|---|---|---|---|---|---|---|
| AgEagle | 2018–2022 | Real RGB UAV | 75 GB | - | [220] | Open | Complete |
| University-1652 (CV-GL) | 27 February 2020 | Real RGB Sat.<br>Sim. RGB UAV | 8.6 GB | [215] | [221] | On Request | Complete |
| ATMD | 5 December 2020 | Real RGB Sat<br>Real RGB UAV | 67 MB | [222] | [223] | Open | Complete |
| SUES-200 (CV-GL) | 25 April 2022 | Real RGB Sat<br>Real RGB UAV | 5.3 GB | [216] | [224] | Open | Complete |
| ALTO | 23 May 2022 | Real RGB Sat/UAV | 29 GB | [225] | [226] | Open | Complete |
| MSDI | 9 August 2022 | Real RGB Sat/UAV | 2.5 GB | - | [227] | Open | Complete |
| WildNav | 30 September 2022 | Real RGB UAV | 1.7 GB | [228] | [229] | Open | Complete |
| Boson-nighttime | 24 February 2023 | Real RGB Sat<br>Real IR UAV | 69 GB | [217] | [230] | Open | Complete |
| DenseUAV | 11 August 2023 | Real RGB Sat/UAV | 16 GB | [231] | [232] | On Request | Complete |
| Boson-nighttime ext. | 12 March 2024 | Real RGB Sat<br>Real IR UAV | 121 GB | [218] | [233] | Open | Complete |
| JointLoc | 3 April 2024 | Sim. RGB Sat/UAV<br>Real RGB Sat/UAV | 2.43 GB | [219] | [234] | Open | Partial |
| UAV-VisLoc | 31 May 2024 | Real RGB Sat/UAV | 16.4 GB | [235] | [236] | Open | Complete |

Overall, these datasets encompass a variety of modalities, including visible spectrum and thermal infrared imagery, and they are sourced from satellites, UAVs, and even planetary landers. They include both real and simulated images, providing a broad spectrum of data types for various research applications. The following section offers detailed descriptions of these datasets, aiming to provide deeper insights into their unique attributes and potential applications.

#### 3.1.1. AgEagle

The AgEagle dataset is a comprehensive collection of UAV data captured by third-party users of AgEagle's drones, cameras, and sensors [220]. Initially created for potential customers to evaluate AgEagle's products, the dataset's high quality has also attracted interest from researchers. He et al. [237] were recently the first to use this dataset for AVL

research. The datasets were captured from 2018 to 2022 and encompass a wide variety of locations and applications, including hotel modeling, palm tree detection, soybean and corn field monitoring, mixed-use agricultural fields, industrial estate mapping, gravel quarry surveys, university campus documentation, and dam inspection. Data was collected using different drones (eBee X, WingtraOne, Asesor/9, eBee Ag) and sensors (S.O.D.A., RedEdge-P, Duet M, Duet T, Aeria X) with varying resolutions (2 cm to 11 cm) and altitudes (60 m to 285 m). The data types include RGB imagery, point clouds, raw data, orthomosaics, multi-spectral images, Digital Surface Models (DSMs), and thermal images, providing a rich resource for various applications. A summary of AgEagle's datasets is provided in Table 3.

**Table 3.** AgEagle datasets.

| Dataset | Year | Drone | Sensor | Resolution | Altitude | Area | Data types | Ref. |
|---|---|---|---|---|---|---|---|---|
| Hotel modelling | 2022 | eBee X [238] | S.O.D.A. [239] | 4.5 cm | 170 m | 1.19 $km^2$ | RGB<br>Point cloud | [240] |
| Palm tree detection | 2022 | WingtraOne [241] | RedEdge-P [242] | 2.3 cm | 60 m | - | Raw<br>Orthomosaic<br>DSM | [243] |
| Soybean field | 2022 | Asesor/9 [244] | RedEdge-P [242] | 6.5 cm | 200 m | - | Raw | [245] |
| Corn field | 2021 | eBee Ag [246] | Duet M [247] | 2.75 cm | 120 m | 2.74 $km^2$ | RGB<br>Multispectral<br>Orthomosaic | [248] |
| Mixed-use agricultural fields | 2021 | eBee Ag [246] | Duet M [247] | RGB: 2.8 cm<br>Multispectral: 11 cm | 120 m | 1.03 $km^2$ | RGB<br>Multispectral | [249] |
| Industrial estate | 2019 | eBee X [238] | Duet T [250] | RGB: 2 cm<br>Thermal: 11 cm | 84 m | 0.075 $km^2$ | RGB<br>Orthomosaic<br>Thermal<br>DSM | [251] |
| Gravel quarry | 2019 | eBee X [238] | Aeria X [252] | 3.6 cm | 122 m | 0.156 $km^2$ | RGB<br>Point cloud | [253] |
| University campus | 2019 | eBee X [238] | Aeria X [252] | 6 cm | 285 m | 2.1 $km^2$ | RGB<br>Point cloud | [254] |
| Dam inspection | 2018 | eBee X [238] | S.O.D.A. [239] | 3.4 cm | 100 m | 0.22 $km^2$ | RGB<br>Point cloud | [255] |

#### 3.1.2. University1652

The University1652 dataset is a comprehensive large-scale dataset specifically designed to advance research in CV-GL image matching tasks [215]. This dataset is notable for being the first to encompass drone-view images in such research. Imagery includes satellite images, simulated drone shots generated using Google Earth's (GE) 3D models [256], and various ground-view images sourced from Google Street View [257] and Google Image searches [258]. The dataset focus on university buildings to help reduce bias and to align better with practical drone use cases. University1652 consists of 50,218 training images representing 701 buildings across 33 universities. It also includes a substantial query set comprising 37,855 drone images, 701 satellite images, and 2579 ground images, each corresponding to the same 701 buildings from 39 universities. The gallery set is extensive as well, featuring 51,355 drone images and 951 satellite images from 951 buildings across 39 universities, and 2921 ground images covering 793 buildings across the same 39 universities. Crucially, there is no overlap between the universities in the training set and those in the test set, ensuring robust validation. The dataset offers a comprehensive multi-view aspect, providing ground-view images from various angles and synthetic drone-view images taken from different distances and orientations along a spiral flight path. On average, each building is represented by 58.38 images, with drone-view images generated by cropping every 15 frames from drone video recordings. Including the additional images from Google Image searches, there are 16.64 ground-view images per building for training.

#### 3.1.3. Aerial Template Matching Dataset (ATMD)

The Aerial Template Matching Dataset (ATMD) is a collection of 2052 high-resolution, partially overlapping aerial images [222]. These images were captured by a DJI Phantom 4 Pro [259] drone over three distinct regions in Pakistan with a total area of approximately 2 $km^2$. The regions include a 0.522 $km^2$ area from NUST Islamabad, a 0.64 $km^2$ area in DHA Rawalpindi, and a 0.664 $km^2$ in Gujar Khan District. The dataset features a diverse range of terrains, from sparsely populated residential areas with water bodies and vegetation to densely populated urban settings. The images in ATMD were taken at various times of the day, ensuring a broad spectrum of lighting and environmental conditions. Post-capture, these images were stitched together to create orthomosaics, where each pixel is geotagged. Each image in the dataset was annotated with up to 16 point-to-point correspondences between the individual satellite image and the orthomosaic. All annotations have been labeled and verified by human annotators. In addition to the primary dataset, ATMD includes orthomosaics for the same three areas but generated using Bing satellite data from OpenLayers in QGIS. However, point correspondences for these particular orthomosaics are not provided.

#### 3.1.4. SUES-200

The SUES-200 dataset is a CV-GL image matching dataset that prioritizes diversity across multiple sources, scenes, and panoramic views [216]. It encompasses imagery collected from 200 distinct locations around Shanghai University of Engineering Science (SUES). The dataset features a mixture of multi-source imagery, including satellite views obtained from AutoNavi Map and Bing Maps, paired with real drone-view images captured at varying altitudes of 150 m, 200 m, 250 m, and 300 m. At each location, the dataset includes one satellite image and 50 drone images. The drone images are taken along a curved flight path to ensure multi-angle coverage. The authors comprehensive data collection approach captures a wide array of scene types, extending beyond campus buildings to include parks, schools, lakes, and public buildings, to allow models to better generalize to real-world environments. SUES-200 is divided into training and testing sets for model development and evaluation. For the training dataset, there are images from 120 locations for each view. The drone view includes 6000 images at each height, totaling 24,000 images, while the satellite view comprises 120 images in total. In the testing dataset, the images are organized into queries and galleries. The drone query includes 16,000 images from 80 locations, with 4000 images at each height. Meanwhile, the satellite query has 80 images from 80 locations. The drone gallery is larger, consisting of 40,000 images from 200 locations with 10,000 images at each height. Lastly, the satellite gallery contains 200 images from 200 locations. The dataset emphasizes high-resolution images to avoid information loss, with drone images at a resolution of 1080 $\times$ 1080 pixels and satellite images at 512 $\times$ 512 pixels.

#### 3.1.5. Aerial-View Large-Scale Terrain-Oriented (ALTO)

The Aerial-view Large-scale Terrain-Oriented (ALTO) dataset aims to facilitate the benchmarking of Visual Place Recognition (VPR) and localization methods for UAVs [225]. It comprises two distinct helicopter trajectories both flown over Ohio and Pennsylvania in August 2017. The longer trajectory is 260 km long and stretches from Northeast Ohio Regional Airport to Zanesville VOR-DME, moving north to south. The second trajectory spans 150 km from Cambridge Municipal Airport to Rostraver Airport, near Pittsburgh, moving from west to east. The subset created from this second trajectory is known as the ICRA2022 dataset and was used in the ICRA 2022 General Place Recognition: Visual Terrain Relative Navigation challenge [260]. This subset is also refered to as the Visual Terrain Relative Navigation (VTRN) dataset by some authors. The dataset contains various terrain types including urban, suburban, forested, rural, and other natural areas. It also includes variation in viewpoints (orientation and translation) as well as varying environmental conditions like changes in illumination and time of the day. The primary imaging sensor used

for the collection is an IDS Imaging UI-5250RE-C camera [261] capturing images at a resolution of 1600 × 1200 and at a rate of 20 Hz. This camera is fitted with a Kowa LM3NC1M 3.5 mm lens [262]. The GPS-INS system consists of a NovAtel SPAN OEMV [263] paired with a LITEF LCI-1 Inertial Measurement Unit (IMU) [264] offering a horizontal position accuracy of 1.5 m Root Mean Square (RMS) and capturing data at 200 Hz. Additionally, a Loke LMC-J-0310-1 laser altimeter [265] that can measure distances up to 700 m with a resolution of 1 mm records altitude data at a rate of 20 Hz. Furthermore, the ALTO dataset includes corresponding georeferenced orthoimagery sourced from the National Agricultural Imagery Program (NAIP) of the U.S. Department of Agriculture Farm Service Agency that was collected in 2012 [266]. The images are provided in two distinct sets, and the ground truth includes the pairing of each query image (UAV image) to a satellite image with a distance difference in meters, as the images do not match perfectly.

#### 3.1.6. Manchester Surface Drone Imagery (MSDI)

The Manchester Surface Drone Imagery (MSDI) dataset is a specialized geoimagery registration dataset designed for advancing visual localization in UAVs [227]. It comprises a diverse collection of drone images including 446 downward-facing, 89 forward-facing at a 45-degree angle, and 64 forward-facing at a 0-degree angle. Each image is supplemented by a parameter matrix of the drone's camera, a transformation matrix, and a checkerboard image for camera calibration. The MSDI dataset includes images from ten locations in Manchester (total 599 images), including the Aquatics Center (80), ASDA (76), Business School (37), Energy Center (71), Holy Name Church (58), Hulme Park (47), Hulme Park at 0-degree (64) and 45-degree angles (89), Metropolitan University (29), and the Museum (48). The imagery was captured by a Parrot Anafi drone [267] flying at an altitude of 100 m and a velocity of 5 m/s, ensuring detailed aerial views. The authors of the dataset also provide image registration with corresponding satellite images from Google Map [257] and Bing Map [268] and includes a Ground Control Point (GCP) selector for simplified GCP selection and file creation in separate software [269].

#### 3.1.7. WildNav

The WildNav dataset is a collection of aerial imagery developed by the Turku Intelligent Embedded and Robotic Systems (TIERS) Lab at the University of Turku, Finland, and designed to advance research in GNSS-free, vision-based localization for UAVs operating in natural, non-urban environments [228]. The dataset comprises 202 high-resolution RGB images, each with dimensions of 4056 × 3040 pixels. These images were captured during two separate flights, resulting in 124 and 78 images respectively. The imagery was acquired using a DJI Matrice 300 UAV [270] equipped with a DJI Zenmuse H20T camera [271], which features a 12-megapixel RGB sensor, an 82.9-degree field of view, and a focal length of 4.5 mm. Images were collected at an altitude of approximately 120 m above ground level, ensuring a ground sampling distance suitable for extracting detailed features necessary for localization tasks. Each UAV image includes embedded GNSS metadata that provides ground truth coordinates. The images were taken during sunny weather conditions in the month of May and the flight area encompassed a sparsely populated region in Finland characterized by diverse natural terrains such as forests, fields, sea shores, roads, and scattered buildings. This area was specifically chosen to challenge localization algorithms by requiring them to rely on natural landmarks rather than artificial structures, which are sparse in this environment. It should be noted that corresponding satellite imagery is unfortunately not provided.

#### 3.1.8. Boson-Nighttime and Boson-Nighttime Extended

The Boson-nighttime dataset is a real-world collection of thermal aerial imagery designed to support research in thermal geolocalization and cross-modal image matching [217]. The dataset was acquired using FLIR's Boson thermal camera [272] with an 8.7 mm focal length lens, a resolution of 640 × 512 pixels, and a 50-degree horizontal field of view. The

images were taken in a nadir (downward-facing) orientation at an approximate spatial resolution of 1 m per pixel. Data acquisition involved six flights conducted between 9:00 PM and 4:00 AM to capture nighttime thermal imagery, collectively labeled as Boson-nighttime. The total ground area covered was approximately 33 $km^2$, predominantly consisting of desert terrain with small portions of farms, roads, and buildings. To generate a comprehensive thermal map, a Structure from Motion (SfM) algorithm was applied to reconstruct the thermal imagery from multiple viewpoints. Orthorectification was then performed by aligning the photometric satellite maps with the thermal maps at the same spatial resolution. Bing satellite maps were cropped to correspond with the thermal map areas and used as reference satellite imagery. The thermal map was divided into 512 × 512 pixel image crops with a stride of 35 pixels. Each thermal image crop was paired with the corresponding satellite image crop, resulting in paired datasets suitable for training and evaluation. The dataset was split into training, validation, and testing sets based on the coverage of the flights: data from three flights were used for training and validation, and the remaining three for testing. Specifically, the Boson-nighttime dataset consists of 10,256 training pairs, 13,011 validation pairs, and 26,568 test pairs of satellite and thermal image crops.

Building upon the original dataset, the Boson-nighttime extended dataset further enhances the collection by augmenting the number of satellite images without corresponding thermal images from 79,950 to 163,344 images, expanding the coverage area to approximately 215.78 $km^2$ [218]. This expansion focused on desert and farm areas adjacent to the original sampling region, incorporating a broader spectrum of geographical patterns. To ensure a robust evaluation of generalization performance, the test region from the original dataset was excluded from the generated data. Thermal images were cropped to 512 × 512 pixels. Corresponding satellite images were cropped to 512 × 512, 1024 × 1024, and 1536 × 1536 pixels, corresponding to size ratios between thermal and satellite images of 100%, 25%, and 11%, respectively.

#### 3.1.9. DenseUAV

The DenseUAV dataset is a low-altitude urban scene collection specifically developed for research in UAV self-positioning and geolocalization tasks [231]. It was collected by following a dense sampling strategy using the downward-facing camera of a UAV in real-world urban environments. This method resulted in overlapping areas between adjacent frames requiring models to capture fine-grained features and recognize spatial relationships while dealing with interference from neighboring frames thereby increasing the complexity of retrieval tasks. Comprising UAV-view and satellite-view images, the dataset was collected from fourteen universities in Zhejiang China. UAV images were captured at altitudes of 80 m, 90 m, and 100 m relative to the ground to account for variations in image scale due to flight height. Data collection used DJI Pilot's waypoints [273] with an error margin within 1 m to maintain consistent latitude and longitude coordinates. To incorporate environmental variability, images were taken under different weather conditions (sunny and cloudy days) and at random times between 6:00 AM and 6:00 PM. A fixed sampling distance of 20 m ensured uniformly spaced data and enhanced the model's ability to learn subtle differences between frames. For the satellite-view images, Google Maps imagery at zoom level 20 was used [257]. To enrich geographic diversity and help models adapt to spatial changes over time, satellite images from two different years (2020 and 2022) are included. Additionally three different scales of satellite images are provided to improve model robustness against variations in image scale. The DenseUAV dataset is structured into a training set, a test set, and a gallery set. The training set consists of 6768 UAV-view images from 2256 sampling points across ten universities and 13,536 corresponding satellite-view images. As for the test set, it includes a query set of 2331 UAV-view images from 777 sampling points across four universities and 4662 satellite-view images. Lastly, the gallery set comprises 27,297 images from 3033 sampling points, covering all fourteen universities involved in the dataset.

#### 3.1.10. JointLoc

The JointLoc dataset encompasses a collection of both simulated and real aerial images of Martian terrains, intended to support research in UAV visual localization and navigation on Mars [219]. The simulated data were created using Unreal Engine 4 [274] and Airsim [275] to model three characteristic Martian landscapes: impact craters, gravel plains, and mountainous regions. A virtual UAV captured overhead image sequences over these terrains, resulting in 925 images for craters, 760 for gravel areas, and 773 for mountainous regions. Each set features an average flight path length of 149.63 m, totaling approximately 2458 simulated images across six different trajectories. The real portion of the dataset is composed of images taken by National Aeronautics and Space Administration's (NASA) Ingenuity Mars Helicopter during the Mars 2020 mission. For each of the three terrain types (impact craters, gravel plains, and mountains), 35 representative images were selected, amounting to 105 real images. Manual calibration was performed to determine the homography matrices between these images and local maps of Mars, facilitating accurate spatial alignment for research purposes. It is important to note that only the test portion of this dataset has been made publicly available.

#### 3.1.11. UAV-VisLoc

The UAV-VisLoc dataset is a comprehensive dataset designed to benchmark visual localization methods for UAVs in real-world scenarios [235]. It comprises 6742 high-resolution drone images and 11 corresponding satellite maps, covering diverse terrains across various regions in China. The images were collected using both multi-rotor and fixed-wing UAVs at altitudes ranging from 400 m to 2000 m and at various heading angles. The inclusion of fixed-wing UAV imagery addresses the need for high-altitude data, as these UAVs are widely used in surveying, mapping, agriculture, and reconnaissance due to their endurance and capability to fly at higher altitudes. The dataset spans two seasons, summer and autumn, introducing variability in environmental conditions. The dataset also encompasses a variety of geomorphological features, including villages, towns, farms, cities, rivers, hills, and forests. This diversity is achieved by collecting data from multiple locations within China, such as Jiangsu Province, the Taihu Lake Coast, the Yangtze River Basin, Donghuayuan Town, Yunnan Province, and Gansu Province. The drone images offer a Ground Sampling Distance (GSD) ranging from approximately 0.1 to 0.2 m per pixel. Each drone image is annotated with comprehensive metadata, including the latitude and longitude of the center point, shooting height, shooting date, and drone heading angle. The satellite maps are sourced from GE [256] at a spatial resolution of 0.3 m per pixel. For each UAV flight, a satellite map covering the entire area captured by the drone is downloaded and integrated to the dataset.

### *3.2. Fine-Grained Matching*

Fine-grained matching is the process of aligning features or points between images with high precision for localization purposes. This approach uses high-resolution feature maps, often extracted at multiple scales, to capture intricate details essential for accurate matching. Multi-scale extraction helps in capturing details at various levels of granularity, essential for matching both large structures and fine details. Given its computational complexity, fine-grained matching is typically performed using a coarse-to-fine, two-step process to make the computational load manageable.

These methods are similar to traditional computer vision techniques, such as feature point matching, but they differ significantly by being detector-free. Instead of relying on keypoint detectors, fine-grained matching methods establish dense, pixel-wise matches directly through the use of deep learning models. These models generate dense feature maps where each pixel is represented by a matching vector. A refinement process is subsequently employed to accurately establish point matches using these feature maps.

Some researchers refer to these methods as template-matching inspired, due to their reliance on dense feature maps and the effort to match each pixel in one image with its

counterpart in another. This is analogous to the process in template matching, where an attempt is made to align a template with the entire image. Consequently, fine-grained matching aligns closely with the concept of dense matching, which seeks to establish correspondences across the entire image.

Once the dense feature maps are refined and precise point matches are established, the final step involves computing the homography matrix. This matrix utilizes the refined point correspondences to achieve precise image registration, culminating in an accurate alignment between images.

The remaining of the section will review a few key representative works that exemplify fine-grained matching techniques within the context of AVL. These examples underscore the advancements and practical applications of these methods, demonstrating their critical role in enhancing localization accuracy and robustness.

Sui et al. developed a UAV visual geolocalization system using a detector-free, two-stage dense image matching method [276]. This system has three modules: image preprocessing, image matching, and localization. The preprocessing module corrects the UAV images to match the satellite imagery viewpoints using a transformation matrix calculated from the localization module during the last localization estimation. The image matching module uses the corrected UAV image and satellite imagery to generate matching point pairs. It uses a ResNet-18 model [277] for feature extraction, to which a batch of two grayscale-normalized images is passed (the UAV and satellite images). Feature maps are extracted at two scales, from blocks 2 and 4 of the ResNet-18 model. In the two-stage matching process, coarse matching uses lower-resolution maps (from block 4) to increase efficiency. The authors use two implementations, a Transformer module for accuracy and a Euclidean distance approach for speed. The Transformer module [278–280] is only used in the pretraining phase of the ResNet-18, and the Euclidean distance approach is used in the fine-tuning phase and in the final localization solution. The fine matching module of the algorithm refines the results using higher-resolution feature maps (from block 2). It applies a $5 \times 5$ window (cost matrix) around the coarse matching coordinates on these maps, with the softmax function transforming this matrix into a probability distribution (confidence matrix). A Differentiable Spatial to Numerical Transform (DSNT) layer then refines the coordinates to sub-pixel precision [281]. The training is performed with the MegaDepth dataset [282]. The dataset is augmented to generate matching coordinates in pairs of images using the available camera parameters and depth values. Training occurs in two phases. Initially, only the coarse matching module is pretrained using the Transformer module and a focal loss [283]. Subsequently, fine-tuning is performed using the Euclidean approach in the coarse matching module and adding the fine matching module with an L2-norm loss. The localization module calculates a coordinates transformation matrix using RANdom SAmple Consensus (RANSAC) [284], based on the matching coordinates obtained from the image matching module. This matrix is used to convert the UAV image pixel coordinates into geographical coordinates by leveraging the satellite imagery's geographical data. The geographical location of the UAV is estimated as the coordinates of the center pixel in the UAV image, assuming a downward-pointing camera. The image matching module's performance is evaluated using the HPatches dataset [285] which includes 116 sequences of 6 images each with known homography. This dataset splits into two subsets: one with 59 sequences with significant viewpoint differences and another with 57 sequences with major changes in illumination, both natural and artificial. HPatches is not specific to UAV or satellite imagery and includes content like buildings, objects, creatures, and landscapes. The precision of the homography matrix estimated from the image matching module surpassed that of other feature point methods like LoFTR [279], SuperGlue [278], Patch2Pix [286], and SIFT [287]. Additionally, the authors tested their localization approach using the ICRA2022 dataset [260]. The results indicated an average localization error of 9.80 m across the dataset 150 km trajectory captured at 20 fps. They also developed a custom dataset using a DJI Phantom 4 RTK drone [288]. This dataset contains orthoimages and satellite navigation data of a 6 $km^2$ suburban area near Wuhan. It was captured at

300 m altitude with a resolution of 1920 × 1080 pixels and includes a corresponding satellite base image from GE [256]. The precision of localization was determined by comparing the algorithmic positioning with the GNSS positioning of the UAV, showing an average localization error of 2.24 m.

Mughal et al. developed a UAV localization solution inspired by template matching [222]. The approach is based on trainable neighborhood consensus [289–291] and leverages local patterns in dense feature correspondence neighborhoods by introducing semilocal constraints. This approach adapts a neighborhood consensus network pretrained on PASCAL-VOC and InLoc datasets. It is modified by adding softmax constraints on its correlation matrix which enables scale invariance and enhances point-to-point correspondences by using the strongest feature point matches between UAV and satellite images. The first step of the method is to extract convolutional features from both UAV and satellite orthomosaic images using a ResNet-101 model [277]. These features are then used to compute exhaustive pairwise cosine similarities, creating a correlation tensor. A 4D CNN [289] processes this tensor to learn to recognize correlation patterns and produce spatially consistent feature point matches. The approach incorporates soft mutual nearest neighbor filtering [289] before and after the CNN to reduce the weight of matching scores which are not mutual nearest neighbors. These refined correlation maps are then used to identify the best match points by extracting pairwise feature correspondences, compute a probability distribution through the softmax function, and determine the 2D indices of points with the highest probability. The top-ranked 2D points from this process are input into another fully connected network to predict precise point-to-point image correspondences. Using these correspondences, the RANSAC algorithm [284] is used to calculate a homography matrix between UAV and satellite images. The four corners of the UAV image are projected over the satellite image, the center point is extracted, the corresponding longitude and latitude coordinates are calculated from the geographic information of the orthomosaic satellite image, and used as the final localization solution. This solution assumes nadir orientation in UAV imagery. To train this model, which has approximately 27.733 million parameters, the authors have created the ATMD dataset [222]. The ResNet-101 model is initialized using pretrained ImageNet weights, while the fully connected layers are initialized with He-normal random weights [292]. The training uses an L2 regression loss and is optimized via Stochastic Gradient Descent (SGD) [293] with momentum and an early-stopping paradigm. The model's matching accuracy is verified using the following rule: a match is correct if it covers at least 90% of the area identified by labeled data. This results in a 92.35% matching rate, outperforming other deep learning and traditional feature point matching methods. Positioning accuracy is assessed by comparing the predicted region's center pixel area deviation in the orthomosaics with UAV Global Positioning System (GPS) coordinates (longitude, latitude). This shows an average positioning error of 3.594 $m^2$ and a maximum error of 31.281 $m^2$. The intersensorial registration performance of this approach is tested against the subset of Bing-sourced orthoimages present in the ATMD dataset. The tests indicate a 91.04% matching rate, an average positioning error of 3.708 $m^2$, and a maximum positioning error of 33.164 $m^2$.

### 3.3. Large-Scale Matching

Large-scale matching is a technique that uses global image descriptors to match UAV and satellite imagery. This method relies on image retrieval principles to find a UAV viewpoint image within a collection of satellite images. The approach is similar to VPR, where a previously visited location is recognized despite changes in lighting or perspective. However, matching UAV imagery with satellite imagery is more challenging due to their different characteristics and acquisition sensors.

To address this, modern methods use CNNs or ViTs to create embeddings or global descriptors for the images. Most contributions are based on metric learning, which trains a CNN or ViT to create an embedding space where similar images are closer together and

dissimilar images are farther apart. The similarity between the resulting embedding vectors can then be measured using the Euclidean distance or the cosine similarity function.

Two commonly used loss functions in this context are triplet loss [104] and triplet margin loss [294]. These loss functions work on triplets of samples: the anchor, the positive, and the negative. The goal is to minimize the distance between the anchor and the positive sample while maximizing the distance between the anchor and the negative sample by a given margin.

Some approaches attempt to regress a probability of match directly using a Siamese network, where both the satellite image and the UAV image are processed through a network in parallel. This method can provide a more straightforward way to compute matching probabilities but requires significantly more computation than metric learning approaches, where the satellite images are processed through a model only once each. This consideration is particularly important when performing computations on resource-limited embedded platforms such as UAVs. Another approach involves using a Variational AutoEncoder (VAE) [295] to encode UAV and satellite images into a shared latent space. The resulting vectors can then be used for matching. However, metric learning is more effective than VAEs for this task, as it directly optimizes for distinguishing between images, leading to a more discriminative feature space. VAEs, in contrast, focus on reconstructing input data, which results in less discriminative latent spaces, making them less suited for tasks requiring precise matching.

Large-scale matching using metric learning is the most mature and popular approach in AVL. It stands as the state-of-the-art technique, delivering high accuracy and efficiency for cross-domain image retrieval. This section will review the foundational works that established this approach, as well as highlighting key contributions in large-scale matching.

Couturier and Akhloufi developed a novel approach for UAV localization by integrating CNNs with particle filters [100]. This contribution stands out as the first in the literature to combine particle filters with CNNs for localization and is one of the three pioneering works in large-scale matching for AVL. Their approach diverges from traditional feature-matching techniques by embedding CNN-extracted features into a particle filter framework, achieving a comprehensive two degrees of freedom (x, y) (2 Degrees of Freedom (DoF)) localization solution. To train the CNN, a purpose-built dataset is constructed, comprising 50,000 images extracted from public satellite imagery using the Bing Maps API [296]. This dataset includes aerial images from 6250 random geographic locations across the United States, chosen for its diverse terrain and comprehensive coverage. An important aspect is the inclusion of oblique 45-degree bird's-eye view imagery, which closely mirrors the properties of UAV imagery due to its capture from low-flying aircraft using conventional cameras. Each location includes top-view imagery at four different zoom levels to simulate varying UAV altitudes and account for scale changes. Additionally, oblique images in four orientations (north, south, east, and west) are incorporated to introduce perspective diversity. This accounts for variations in chromaticity, illumination, altitude, perspective, orientation, and terrain. The CNN architecture used is ResNet-50 [277], and the training approach is inspired by FaceNet, a seminal work in face recognition using deep learning [104]. The network is trained using the triplet loss [104] function to optimize feature embedding, with triplets randomly selected from the dataset. Once trained, the CNN is integrated into a particle filter, which narrows the search space within the feature database and identifies potential matching locations in the reference imagery. The particle filter also leverages data from the UAV IMU to enhance localization accuracy by fusing visual information with sensor data. Experimental validation is conducted using real-world data from an 8 km flight performed with a Parrot Bebop 2 quadcopter [297]. The validation process involves two main procedures: data collection during a flight from point A to point B, and localization estimation simulating a GNSS outage while backtracking from point B to point A. The front-facing monocular camera of the UAV is oriented at an 83-degree angle from the horizon and captures images at a rate of 1 Hz. The flight traverses urban, rural, and forested areas at an altitude of 150 m and a velocity of 10 m/s. Ground truth positions

are obtained from the onboard data logger of the UAV, and the evaluation metric used is the average Euclidean distance between the estimated positions and the ground truth (GNSS). The model achieves a location identification accuracy of 82.44% on the testing set. When integrated into the particle filter, the localization solution demonstrates a localization error of approximately 32.9 m under typical conditions with the quadcopter.

Valaitis et al. introduced a novel CNN architecture for learning image similarity measures for UAV localization [98]. Their approach is recognized as one of the three foundational works in large-scale matching AVL. It leverages a pretrained VGG16 [95] network's lower layers as feature extractors, capitalizing on the network's strong discriminative properties in image classification. Inspired by advancements in face verification and recognition, they formulated the learning process as a ranking similarity learning problem using a modified triplet loss function known as the lossless triplet loss [298]. Unlike other approaches that employ the Euclidean distance to measure similarity in the embedding space, Valaitis et al. developed a custom similarity function, which aims to enhance the measurement of similarity between images by capturing more nuanced differences. This custom similarity function is shown in Equation (5), where $f_{1j}$ and $f_{2j}$ are the feature embeddings of images $im_1$ and $im_2$, respectively, $N$ is the number of dimensions, and $\beta$ and $\epsilon$ are small constants to stabilize the computation. To train and evaluate their model, the authors used a custom dataset from their previous research, consisting of 113,474 images captured from various virtual UAV flights within a robotics simulator. The images, with a resolution of 640 $\times$ 480 pixels, were rectified to eliminate camera distortion. For this study, a specialized subset was created by performing virtual flights over two satellite maps taken several years apart, resulting in temporal variations in the imagery. The source of the imagery is not mentioned in the original paper. This subset included 1188 image triplets for training, 105 for validation, and 64 for testing. Each triplet comprised an anchor image, a positive image (the same location at a different time), and a negative image (a random region from the map used for the anchor image). The images were resized to 224 $\times$ 224 pixels to match the input dimensions of the VGG16 network. Given the limited size of the training dataset, the authors chose to freeze the weights of the VGG16 layers to prevent overfitting, as fine-tuning these layers reduced the overall accuracy of similarity estimates. They implemented and compared four configurations of their model by including different numbers of VGG16 layers. Hyperparameter tuning was performed using grid search with fourfold cross-validation to optimize the model's performance. Their proposed ModelNN, which integrates the VGG16 base with additional custom layers, demonstrated the ability to learn image similarity with an image match accuracy exceeding 99% on their test dataset. However, the study acknowledged limitations due to the small training dataset and notably did not provide a complete localization solution or conduct localization validation:

$$\mathrm{sim}(im_1, im_2) = -\ln\left(-\frac{\sum_{j=1}^{N}(f_{1j} - f_{2j})^2}{\beta} + 1 + \epsilon\right), \tag{5}$$

Ahn et al. introduced a satellite image-based UAV visual localization method that uses a Siamese neural network architecture [99]. This approach aims to learn a similarity metric capable of recognizing the same location from two raw images: a UAV image and a corresponding satellite image patch. Inspired by FaceNet, the authors implemented a Siamese Neural Network (SNN) [299] with a CNN backbone to address variations in perspective, shadow angles, and the presence of vehicles [104]. Training was conducted using the triplet loss function [104], where positive and negative samples were carefully selected. Positive samples were generated by creating shifted images from the original satellite patch, with shifts limited to $\pm$25% of the total pixels to simulate slight viewpoint variations. Negative samples were taken from different patches within the same database image, excluding the original patch and its augmented versions. To simulate UAV imagery, the authors used two different sources of satellite imagery. One source was designated to represent UAV images, while the other served as the actual satellite imagery. They

constructed a custom dataset comprising 88 corresponding images from three different cities in New York and Florida. Each image covered an area of 0.16 $km^2$. The UAV images were at a resolution of 0.125 m per pixel (3200 × 3200 pixels) and the satellite images were at 0.5 m per pixel (800 × 800 pixels). These images were then divided into patches. For training and validation, approximately 30,000 and 500 image pairs, respectively, were generated with the augmented images created by rotating and adding random values to the color channels. The method was evaluated using the top-k precision score. It achieved a top-1 precision of 65% and a top-3 precision of 74%. However, the study had certain limitations: it lacked validation on real UAV data and did not provide a complete localization solution, focusing solely on place recognition. Additionally, specific details such as the CNN architecture used and the sources of the satellite imagery were not mentioned in the paper.

Liu et al. proposed a Block Joint learning Network (BJN) for UAV-satellite image matching [300]. BJN leverages multi-source joint feature learning and exploits the coupling relationships between the data sources by simultaneously extracting features from both images. Building on the work in [301], BJN incorporates the early-stage fusion of features, distinguishing it from other distance metric learning methods like [302]. BJN features two separate heads for processing UAV and satellite images, each receiving a stack of five channels comprising RGB channels, a grayscale image, and an edge image generated using the Sobel operator [303]. These heads are followed by respective transfer modules, whose outputs are then combined and passed into a fusion feature extractor. A Channel Attention (CA) module [304] is employed to optimize the weight distribution among the extracted features. The network ends with a classification stage consisting of fully connected layers that output two probabilities: the likelihood of the images being from the same scene and the likelihood of them being from different scenes. The authors synthesized a dataset using GE [256] to simulate UAV imagery by adjusting the view angle and flight height. This dataset includes two subsets: LA-500 [305], comprising 500 aerial images, and LA-10,000, which contains 10,000 aerial images. These images feature urban and suburban scenes such as bare ground, buildings, streets, and vehicles, are 500 × 500 pixels in resolution and taken from a simulated flight height of 250–300 m. The LA-10,000 dataset was used for training, while LA-500 was reserved for model evaluation. The proposed method achieved a top-1 accuracy of 63.5%, outperforming other approaches from the literature like MatchNet [306], SOSNet [307], and SOLAR [308]. Moreover, the authors validated their approach using a real UAV dataset captured with a DJI UAV (specific model not mentioned) over the campus of NorthWestern Polytechnical University (NWPU). This dataset, collected from an altitude of approximately 200 m, includes over 150 distinct urban scenes featuring grasslands, lakes, buildings, and sports courts. To further evaluate their approach, they utilized a semiquantitative method they named dense sampling by patch cropping. This process involves extracting small sections (patches) from larger images for detailed analysis. In satellite images, these patches are sampled every 20 pixels, corresponding to approximately 0.494 m per pixel, meaning each patch's center is about 10 m away from its neighboring patch. The results demonstrated that the top-2 search results had a maximum distance error of 10 m from the ground truth position and exhibited significantly higher similarity probabilities than other images.

Xiao et al. proposed an unprecedented approach for UAV visual geolocalization in low-illumination environments using thermal imagery [217]. The challenge in this approach comes from the fact that thermal sensors detect infrared radiation instead of visible light, creating notable visual inconsistency compared to those obtained using visible spectrum cameras. To address the limited availability of paired UAV thermal and satellite imagery, the authors developed domain adaptation methods to enable the leveraging of RGB satellite imagery. The approach is divided into two components, the Thermal Generative Module (TGM) and the Satellite-thermal Geo-localization Module (SGM). In the TGM, the authors generate synthetic thermal images from RGB satellite images. A pix2pix generative architecture [309] is used to generate the thermal images, creating a dataset of

paired satellite and synthetic thermal images. The model uses an adversarial structure with a generator and a discriminator that are trained as in [309] by using a ponderated sum of a Generative Adversarial Network (GAN) loss and an L1 loss. The approach for the SGM is inspired by deep VG benchmark [310] and its image retrieval process. The authors improve the approach by adding a compression layer to reduce the dimensionality of the feature maps, an approach to train with paired satellite–thermal images and a generated domain transfer dataset simultaneously, and a Domain-Adversarial Neural Network (DANN) loss [311] for unsupervised domain adaptation. The proposed approach starts by mining triplets of a query image, a positive match, and a negative match using partial database mining [310]. In partial database mining, positive matches are selected in the geographic region close to the query image and the negative matches from outside that region. During the mining, the query image is either a real thermal UAV image or a synthetic one generated from the TGM module, while the positive and negative matches are always real RGB satellite images. The final positive and negative samples of the triplets are selected as the ones with the lowest $L_2$ distance to the query image in the feature embedding space. After mining, the triplets images are passed one by one in a feature extractor, then the resulting feature map is passed to the compression layer, and then to an aggregation module using NetVLAD [312] to obtain the final global image descriptor. The resulting model is trained using a ponderated sum of the triplet margin loss [294] and the DANN loss. The DANN loss is the combination of a gradient reversal layer, a domain classifier, and a cross-entropy loss [311]. The goal of this approach is to classify the query image as the thermal class and the positive and negative matches as the satellite class and, through backpropagation and the cross-entropy loss, to force the distribution of features from the thermal and satellite images to become more similar. The authors created the Boson-nighttime dataset, which they used to train their models and validate their approach [217]. For the evaluation of thermal geolocalization performance, the authors used several metrics: Recall@N (R@N) measures the percentage of query images from which the top-N retrieved database images are within 50 m; Recall@N with tne prior location threshold $d$ ($R_d@N$) is $R@N$ with the search region limited to a radius of $d$ m from the queries; and the $L_2$ distance error with prior location threshold $d$ ($L_2^d$) measures the $L_2$ distance from the queries to the estimated position from the top-N retrieved database images within a radius of $d$ m. The results showed $R@1 = 92.1$, $R@5 = 96.9$, $R_{512}@1 = 96.5$, $R_{512}@5 = 99.1$, and $L_2^{512} = 14.7$ m. It is important to note that this paper is the most thorough that we reviewed. The authors are incredibly detailed in their work, offering all the necessary information to reproduce the experiments, including the dataset, exact architecture details, precise training parameters, and more. Additionally, their evaluation method is one of the most comprehensive we have encountered, with assessments conducted on real data and the use of multiple high-quality metrics. They have provided an exemplary model for future publications in the field.

Bianchi and Barfoot proposed an approach using a VAE to match real-time UAV images with precollected GE imagery [256] along a desired flight path [313]. The method involves training a VAE to compress these GE images into low-dimensional vector representations while retaining essential features for localization. During operation, the same VAE compresses live UAV images, which are then compared to the nearby precollected GE image vectors using an inner-product kernel. This comparison yields a distribution of weights over the corresponding GE image poses, enabling the generation of a single localization estimate along with an associated covariance to represent uncertainty. In this work, UAV images are simulated around the intended flight path using GE imagery to create the training dataset. Images are rendered every 0.5 m along the path, with additional lateral offsets up to 5 m on either side, resulting in 42 images per meter. The VAE architecture is based on the VAE framework proposed by Kingma and Welling [295] and implemented as per [314]. The encoder processes 320 $\times$ 160 grayscale GE images through six layers, each performing a 2D convolution with a stride of two followed by batch normalization, progressively doubling the number of channels. The output is reduced to a bottleneck vector of size 1000, which is determined to be the smallest size maintaining the desired accuracy.

The decoder reconstructs the images from this bottleneck vector through upsampling and convolutional layers, halving the number of channels at each step until the original image size is restored. The network is trained using a combination of photometric loss [315] and L2 loss between the corresponding layers, referred to as the skip loss. The method was evaluated using a dataset collected at the University of Toronto Institute for Aerospace Studies (UTIAS) as described in [316]. The dataset comprises six traversals of a 1132-m path over areas with roads, buildings, grass, and trees, captured under different lighting conditions (sunrise, morning, noon, afternoon, evening, and sunset). A DJI Matrice 600 Pro multirotor UAV [317] equipped with a 3-axis DJI Ronin-MX gimbal [318] was used for data collection, with a StereoLabs ZED camera [319] providing stereo images at 10 FPS. Ground truth positions were obtained using a Real-Time Kinematic (RTK) GPS system and an IMU. Experimental results indicated an average registration success rate of 97.52% across all six flights. For successful registrations, the method achieved an average longitude Root Mean Squared Error (RMSE) of 1.36 m, latitude RMSE of 0.97 m, and heading RMSE of 0.357 degrees. Including all registrations, the overall average RMSEs were 1.522 m for longitude, 0.977 m for latitude, and 0.36 degrees for heading.

### *3.4. Hybrid Matching*

Hybrid matching integrates the strengths of large-scale and fine-grained matching to achieve accurate localization in AVL applications. While standalone methods are more common, hybrid approaches offer significant advantages in terms of robustness, accuracy, and real-time performance.

The primary advantage of this two-stage approach is its balance between computational efficiency and localization accuracy. It employs a coarse-to-fine strategy, which starts with large-scale retrieval using global descriptors to narrow down the search space within extensive satellite imagery datasets. This broad search is then followed by fine-grained matching to ensure detailed, pixel-level correspondences.

By combining global image descriptors for broad searches and dense feature maps for precise alignments, hybrid methods effectively manage both the vast scale of satellite databases and the detailed requirements of UAV localization. The ability to achieve both real-time processing and fine-grained accuracy simultaneously is a pivotal benefit of hybrid matching's hierarchical method. This strategy has the potential to become state of the art in the field of AVL and to offer unprecedented levels of performance. However, the hybrid matching approach requires further contributions and maturity to realize its full potential.

Li et al. developed the Global-Local Visual Localization (GLVL) network, a novel approach for UAV localization [320]. This network adopts a two-stage visual localization strategy, integrating a large-scale retrieval module with a fine-grained matching module, enabling real-time and precise localization. The large-scale retrieval module identifies regions in satellite images that are similar to the current UAV image in order to narrow the search space for UAV view matching. Training data are used to extract triplets comprising an anchor image, a positive match, and a negative match. These images are processed through a shared encoder module, built from the layers of ResNet50 [277] up to the conv2 layer. This encoder module is shared and serves both large-scale retrieval and fine-grained matching modules to enhance efficiency. For the large-scale retrieval module, the output from the shared encoder is fed into the remainder of the ResNet50 model. A lightweight pooling layer, Generalized-Mean (GeM) [321], is employed to reduce the memory consumption of the resultant feature maps and to derive a compact image descriptor. The network is trained using a triplet loss based on the Euclidean distance [103,104]. The VTRN dataset, a subset of the ALTO dataset [225], is used to train the large-scale retrieval module. The fine-grained matching module resolves the precise UAV localization through keypoint detection and matching. Image pairs from the large-scale retrieval module are processed through the shared encoder to produce base features. These features are passed to a SuperPoint architecture for feature point extraction and description [322]. The k-Nearest Neighbors (kNN) algorithm is used to match the feature points between UAV and satellite

images [323,324]. The SuperPoint architecture is trained in a self-supervised manner using the RSSDIVCS dataset [325]. Similarly to the original SuperPoint training approach, images from the dataset are transformed using straightforward geometric transformations to create synthetic pairs of images with matching points to train the model. Both modules undergo joint training, with iterative optimization of each module at every epoch using their respective loss functions. The final localization solution is determined through a homography transformation between UAV and satellite images based on the extracted feature points. The model's evaluation involves six different scenes from two datasets. A village scene and its corresponding satellite image from [96] are used, along with five scenes from the MSDI dataset [227] and their orthorectified images from GE [256]. These scenes include both texture-rich and texture-sparse areas. The authors' approach is benchmarked against LocUAV [96], a state-of-the-art method that uses frame-by-frame predictions and performs localization through a sliding window mechanism. The results indicate that the authors' approach achieves an average localization error of 2.39 m for the village scene compared to 7.06 m for LocUAV. Additionally, they demonstrates resistance to drift over time on the village scene while LocUAV is shown to have drift. The authors have made their code and their modified datasets version publicly available [326].

## 4. Discussions and Challenges

This is our second review on the subject of Absolute Visual Localization (AVL) for Unmanned Aerial Vehicles (UAVs). We previously issued recommendations and research directions based on the literature from 2015 to 2019 in [31]. Revisiting these earlier directions allows us to discern which recommendations remain relevant, which have been resolved, and where further work is required. In this section, we revisit those prior directions and introduce new perspectives by addressing the emerging challenges and advancements in the field.

### *4.1. Algorithmic Approaches*

One of our past recommendations was to pursue more research on deep learning for AVL. The mere existence of this present paper demonstrates that deep learning research in AVL has been extensively pursued by the research community. This will evidently remain a valid research direction, but it has now become an implicit one.

Another recommendation was the development of specialized feature point methods for aerial imagery which initially appeared necessary for advancing the field. At the time, feature point approaches were the uncontested state of the art, offering superior accuracy and practical usability in real localization scenarios. It was uncertain whether a breakthrough in deep learning would surpass these approaches. However, concurrent with the publication of our earlier review, our research group and other researchers published significant work on metric learning similar to what was performed in the field of facial recognition [98–100]. This triggered a paradigm shift that allowed for learning instead of hand-crafting feature extraction, thus greatly improving the reliabilitly of AVL. Furthermore, it allowed image matching at the global descriptor level, greatly simplifying AVL implementations. While some contributions still make use of feature point approaches, such as [327], they have become the exceptions in the past few years.

We also recommended the exploration of more approaches to adapt Visual Odometry (VO) to AVL. Research in VO for Relative Visual Localization (RVL) remains active, as such approaches are implemented in commercial drone products [328,329]. However, the initial push for combining VO with satellite maps for error correction has been rendered unnecessary, as deep learning approaches have become predominant.

### *4.2. Hardware and Sensors*

One of the previous suggested research directions was the deployment of AVL approaches on embedded Graphics Processing Units (GPUs). Unfortunately, we were unable to find publications that demonstrate real-life embedded implementations of AVL algo-

rithms. As UAVs necessitate onboard embedded deployment for AVL, the lack of attention to this aspect is concerning. This research direction remains critically important, particularly as embedded GPUs, General-Purpose Graphics Processing Units (GPGPUs), and AI-capable microcontrollers continue to diversify and improve in power efficiency and performance [330–333]. If this research direction continues to be neglected, we anticipate that the adoption of AVL technology will be delayed as a result.

We also recommended developing more hyperspectral AVL approaches. We acknowledge that using spectra beyond the visible spectrum is challenging due to the significantly higher cost of sensors. Despite these challenges, the publication of two papers by Xiao et al. [217,218] has been highly valuable. These papers offer remarkable contributions on thermal imagery for AVL in low-illumination environments. Their innovative approach includes the use of real thermal imagery, Generative Adversarial Networks (GANs) and style transfer techniques to generate paired satellite and synthetic thermal images to supplement limited real-world data. Furthermore, they made both their code and datasets publicly available, thereby maximizing the impact of their research on the field [230,233]. While Xiao et al. have set a high standard, much remains to be performed in this direction, which makes it an open avenue for future research.

### *4.3. Source Code Availability*

Another significant issue highlighted in our previous paper is the scarcity of freely available implementation codes. We are pleased to report substantial improvements in this area. Since then, we have identified at least nine publications on AVL and two on Cross-View GeoLocalization (CV-GL) that provide publicly available source code. Additionally, some publications have alternative third-party reimplementations of their approaches, indicating a thriving open-source ecosystem. In total, we have identified 14 code repositories. This improvement in code availability not only facilitates reproducibility but also accelerates progress by allowing researchers to build upon existing work (e.g., [334]). This further promotes innovation and collaboration within the AVL community. Table 4 presents these research works along with their associated source code.

**Table 4.** Research works with associated publicly available source code.

| Work | Type | Paper | Ori. Code | Alt. Code |
|---|---|---|---|---|
| Long-range UAV Thermal Geo-localization with Satellite Imagery | AVL | [217] | [230] | - |
| STHN: Deep Homography Estimation for UAV Thermal Geo-localization with Satellite Imagery | AVL | [218] | [233] | - |
| A Scene Graph Encoding and Matching Network for UAV Visual Localization | AVL | [335] | [336] | - |
| UAV Localization Using Autoencoded Satellite Images | AVL | [313] | - | [337,338] |
| Vision-Based UAV Self-Positioning in Low-Altitude Urban Environments | AVL | [231] | [232] | - |
| Vision-Based GNSS-Free Localization for UAVs in the Wild | AVL | [228] | [229] | [334] |
| Leveraging Map Retrieval and Alignment for Robust UAV Visual Geo-Localization | AVL | [237] | [339] | - |
| JointLoc: A Real-time Visual Localization Framework for Planetary UAVs Based on Joint Relative and Absolute Pose Estimation | AVL | [219] | [234] | - |
| Jointly Optimized Global-Local Visual Localization of UAVs | AVL | [320] | [326] | - |
| SDPL: Shifting-Dense Partition Learning for UAV-view Geo-localization | CV-GL | [340] | [341] | [342] |
| UAV's Status Is Worth Considering: A Fusion Representations Matching Method for Geo-Localization | CV-GL | [343] | [344] | - |

### *4.4. Datasets Availability*

Previously, a major impediment to research progress was the scarcity of public data. To address this, we recommended the publication of more datasets. We are pleased to see that this issue has now been largely resolved. Between 2018 and 2024, at least 12 datasets have been made available, including 10 for AVL and 2 for CV-GL that are applicable to AVL.

However, this improvement has introduced new challenges. While most of these datasets are openly accessible, some, such as DenseUAV [345] and University1652 [346],

require request by email for access. The access-on-request process, although allowing original authors to track the usage of their datasets, can discourage broader research use due to the waiting periods and administrative procedures. Furthermore, there is a risk that valuable datasets may become underutilized or even obsolete if something prevents the original author from fulfilling access requests, such as unforeseen events or administrative issues.

Establishing open access infrastructures and robust data repositories is essential for ensuring long-term data preservation and enhancing the reproducibility of scientific research [346–348]. By fostering an open-access culture, the scientific community can prevent valuable datasets from becoming lost due to administrative issues or unforeseen events affecting dataset creators [345]. Safeguarding datasets for future generations of researchers is particularly important. Therefore, we advocate for access-on-request practices to remain an exception rather than the norm. Monitoring the impact of these practices over time will be crucial to understanding their implications on scientific progress. We recommend that future datasets avoid such restrictive processes to ensure that the collaborative and cumulative nature of scientific research is not impeded.

### *4.5. Evaluation and Result Reporting*

Similar to our previous observations, the central issue in the AVL field remains the lack of consensus or standard methods for reporting results and for validation procedures. Unfortunately, this situation has not improved since our last review. Other fields leveraging applied deep learning exhibit far superior practices in result reporting, allowing for direct comparisons between the reported findings [349–351]. This remains a significant impediment to productive research in AVL.

The selection of metrics for reporting results varies considerably among researchers, and it is evident that the existing metrics do not meet the satisfaction of many in the field. Notably, Sui et al. [276] and Mughal et al. [222] proposed what should be comparable fine-grained methodologies for AVL. However, Sui et al. assessed their method using the average localization error relative to Global Navigation Satellite System (GNSS) positioning, whereas Mughal et al. employed the average intersensorial registration error between images. These two metrics differ greatly, rendering direct comparison of the approaches impossible. Similarly to Mughal et al., Xiao et al. in [217] found it necessary to introduce new metrics. Their evaluation metrics for thermal geolocalization included Recall@N, which quantifies the percentage of query images that successfully retrieve the correct database images within 50 m; Recall@N with prior location threshold $d$ ($R_d@N$) that is a similar measure, but confines the search to a predetermined radius $d$ based on prior location estimates; and the $L_2$ distance error with a prior location threshold $d$ ($L_2^d$) that assesses the error distance in meters between the actual location of the query and the estimated location, using the top-ranked retrieved images within the radius $d$. Despite Xiao et al. providing the most comprehensive metrics encountered, which are invaluable for understanding the performance of the different variations of their approach, their limited adoption impedes our and other researchers' ability to compare their work with existing literature.

Another issue is the disparity in evaluation methods and the thoroughness of the evaluations. For instance, Valaitis et al. [98], Ahn et al. [99], and Couturier and Akhloufi [100] published similar metric learning-based approaches using triplet loss and Convolutional Neural Networks (CNNs) for large-scale image matching. However, Valaitis et al. and Ahn et al. reported only image matching results, while Couturier and Akhloufi also provided localization results. This inconsistency represents a significant difference in the thoroughness of evaluation of the proposed approaches, making it challenging to compare the localization performance across these works. Despite Valaitis et al. and Ahn et al. focusing on localization in their methodologies, the absence of extensive localization analyses makes it impossible to determine whether their approaches would be effective in a real localization scenario. Furthermore, we note that such comprehensive evaluations of localization

capabilities are often missing from contributions in the field leading to a diminished impact for these contributions.

The lack of consensus is further exacerbated by the recent proliferation of publicly available datasets and source codes, leading to researchers' lack of awareness of their existence. While the availability of these resources is beneficial, the fact that they have been published relatively recently has resulted in few studies using the same datasets. Most researchers employ different combinations of datasets for training and testing. Additionally, there is a wide variation in the UAV data used for complete localization solution validation, which is different from the training and testing data used to train the model. Some works are validated with simulation data based on satellite imagery, while others are validated on real UAV data. Using satellite data to simulate UAV data introduces significant bias in the results, which represents a substantial issue. This adds another layer of choice for researchers, further exacerbating the lack of consensus and the difficulty of comparing works in the field. We anticipate that our new review, which is the first to inventory the existing datasets for AVL, will help raise awareness of these resources, potentially reducing disparities in result reporting and facilitating more in-depth performance evaluations on real-world data.

Our recommendation to researchers in the field of AVL is to remain vigilant in regard to result reporting by comparing their work to related studies, providing a range of different metrics when in doubt, and leveraging publicly available datasets and source codes to standardize evaluations. Additionally, we advise avoiding simulated data when possible, and exercising caution when reporting results derived from such data to ensure accuracy and reliability.

## 5. Future Directions

Following this thorough review of the current literature on deep learning for Unmanned Aerial Vehicle (UAV) Absolute Visual Localization (AVL) and an analysis of the current state of the field and its remaining challenges, we propose the following research directions to guide future research:

- *Architecture and Loss Function Benchmarking for Large-Scale Matching in AVL.* While the current state of the art in deep learning for AVL is large-scale matching approaches, there remains significant potential for refinement. Future research should explore various metric learning loss functions and architectural types, particularly recent implementations from other fields, to enhance performance.
- *Fine-Grained Matching in AVL.* Fine-grained matching is an emerging area in AVL. Further research is needed to identify effective architectural design strategies that balance computational complexity and accuracy. This can help in developing more precise and efficient localization systems.
- *Hybrid Matching in AVL.* Few studies have combined large-scale and fine-grained matching within a single localization solution. Research in this direction holds the potential to develop state-of-the-art approaches that are both more efficient and more accurate than current solutions.
- *Embedded AVL Implementations.* The development and validation of embedded and real-time localization solutions on actual hardware are currently lacking. Advancing this area would significantly contribute to the field and facilitate the adoption of AVL technologies in the industry.
- *Hyperspectral AVL.* Research on hyperspectral AVL solutions is sparse. Exploring different spectra, such as Short-Wave InfraRed (SWIR), could provide effective localization solutions in challenging conditions like cloud cover and fog, thus offering valuable contributions to the field.
- *Development of Standardized Performance Evaluation Libraries.* Other fields, have developed effective and standardized performance evaluation libraries, like pycocotools [352] in object detection or py-motmetrics [353] in object tracking. Developing

similar libraries for AVL, along with automatic UAV dataset downloaders, could greatly enhance study reproducibility and advance the field.

- *Integration of AVL with Inertial Measurement Unit (IMU) Technology.* IMU data is currently an untapped source of high-quality, independent, and high-frequency localization data with the potential to greatly improve the localization performance of AVL approaches. Future research could derive ways to effectively combine IMU data with AVL methodologies, which could lead to the realization of significant gains in both accuracy and reliability.

## 6. Conclusions

This paper presents the first comprehensive review of the literature on deep learning approaches for Unmanned Aerial Vehicles (UAVs) Absolute Visual Localization (AVL). Unlike existing works, this review narrows its focus specifically to localization and dives more deeply into key contributions for the field, offering a more thorough examination of significant research contributions. As a result, three main categories of approaches are identified in the literature: fine-grained matching, large-scale matching, and hybrid approaches. Additionally, this paper reviews major research developments that have paved the way for the application of deep learning in this field, offering valuable insights for readers. A meticulous and unprecedented analysis of the current state of Global Navigation Satellite System (GNSS) and Inertial Measurement Unit (IMU), along with their associated error sources, is conducted within the context of their integration with AVL. This analysis provides a strong motivation for the continuation of the research aiming at supplementing GNSS+IMU localization solutions through the unique source of absolute positioning that is AVL for UAVs. In addition, the most comprehensive dataset review for AVL to date in the field is presented in the paper, offering valuable information for researchers looking to thoroughly evaluate their AVL methods on real-world data. The research directions proposed in our previous literature review have been revisited in the light of the recent advancements in the field. Critical discussions highlight the significant improvements, such as the increased availability of public datasets and open-source code, which have positively impacted the field. However, notable gaps remain, including the lack of standardized methods for result reporting and for approach validation, as well as insufficient interest in embedded AVL implementations. As a result, seven future research directions are proposed, ranging from the ongoing development of deep learning techniques for AVL to the creation of standardized performance evaluation libraries. Furthermore, recent years have witnessed renewed interest in the field, largely due to global conflicts, such as the Russia-Ukraine war, where UAVs are being deployed on an industrial scale. The increase in GNSS jamming and spoofing incidents has further underscored the necessity of robust AVL solutions for autonomous UAVs. Their deployment, particularly in defense and security applications, holds significant potential to influence conflict outcomes, highlighting the criticality and timeliness of research in this area.

**Author Contributions:** Conceptualization, A.C. and M.A.A.; methodology, A.C. and M.A.A.; software, A.C.; validation, A.C.; formal analysis, A.C.; investigation, A.C.; data curation, A.C.; writing—original draft preparation, A.C.; writing—review and editing, A.C. and M.A.A.; visualization, A.C.; supervision, M.A.A.; project administration, M.A.A.; funding acquisition, M.A.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was enabled in part by support provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2024-05287.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References


1. News, C.U. Commercial Drones in 2023: Trends and Growth Prospects. 2023. Available online: https://www.commercialuavnews.com/international/commercial-drones-in-2023-trends-and-growth-prospects (accessed on 12 July 2024).
2. Insights, D.I. Drone Market Report 2023–2030. 2023. Available online: https://droneii.com/product/drone-market-report (accessed on 12 July 2024).
3. Taylor, D. UAS Spending to Top $258 Billion over Next Decade: Report. 2024. Available online: https://militaryembedded.com/unmanned/payloads/uas-spending-to-top-258-billion-over-next-decade-report (accessed on 12 July 2024).
4. APDR. Worldwide Military UAS Spending to Hit $23 Billion by 2033. 2024. Available online: https://asiapacificdefencereporter.com/teal-group-worldwide-military-uas-spending-to-hit-us23-billion-by-2033/ (accessed on 12 July 2024).
5. Obiuto, N.C.; Festus-Ikhuoria, I.C.; Olajiga, O.K.; Adebayo, R.A. Reviewing the Role of AI in Drone Technology and Applications. *Comput. Sci. Res. J.* **2024**, *5*, 741–756. [CrossRef]
6. Bala, M.; Eiszler, T.; Chen, X.; Harkes, J.; Blakley, J.; Pillai, P.; Satyanarayanan, M. Democratizing Drone Autonomy via Edge Computing. In Proceedings of the 2023 IEEE/ACM Symposium on Edge Computing (SEC), Wilmington, DE, USA, 6–9 December 2023; pp. 40–52. [CrossRef]
7. Urakubo, T.; Budiyono, A. Advancements in Tilted-Rotor Unmanned Aerial Vehicles: A Comprehensive Review. *J. Instrum. Autom. Syst.* **2023**, *10*, 93–104. [CrossRef]
8. Wang, H.; Cheng, H.; Hao, H. The Use of Unmanned Aerial Vehicle in Military Operations. In *Proceedings of the Man-Machine-Environment System Engineering Conference*; Long, S., Dhillon, B.S., Eds.; Springer: Singapore, 2020; pp. 939–945. [CrossRef]
9. Koslowski, R. Drones and border control: An examination of state and non-state actor use of UAVs along borders. In *Research Handbook on International Migration and Digital Technology*; McAuliffe, M., Ed.; Edward Elgar Publishing: Cheltenham, UK, 2021; pp. 152–165. [CrossRef]
10. Gupta, A.; Afrin, T.; Scully, E.; Yodo, N. Advances of UAVs toward Future Transportation: The State-of-the-Art, Challenges, and Opportunities. *Future Transp.* **2021**, *1*, 326–350. [CrossRef]
11. Martinez-Alpiste, I.; Golcarenarenji, G.; Wang, Q.; Alcaraz-Calero, J.M. Search and rescue operation using UAVs: A case study. *Expert Syst. Appl.* **2021**, *178*, 114937. [CrossRef]
12. Corcoran, E.; Winsen, M.; Sudholz, A.; Hamilton, G. Automated detection of wildlife using drones: Synthesis, opportunities and constraints. *Methods Ecol. Evol.* **2021**, *12*, 1103–1114. [CrossRef]
13. Mohan, M.; Richardson, G.; Gopan, G.; Aghai, M.M.; Bajaj, S.; Galgamuwa, G.A.P.; Vastaranta, M.; Arachchige, P.S.P.; Amorós, L.; Corte, A.P.D.; et al. UAV-Supported Forest Regeneration: Current Trends, Challenges and Implications. *Remote Sens.* **2021**, *13*, 2596. [CrossRef]
14. Nikhil, N.; Shreyas, S.M.; Vyshnavi, G.; Yadav, S. Unmanned Aerial Vehicles (UAV) in Disaster Management Applications. In Proceedings of the 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 20–22 August 2020; pp. 140–148. [CrossRef]
15. Aggarwal, S.; Kumar, N.; Alhussein, M.; Muhammad, G. Blockchain-Based UAV Path Planning for Healthcare 4.0: Current Challenges and the Way Ahead. *IEEE Netw.* **2021**, *35*, 20–29. [CrossRef]
16. Li, F.; Kunze, O. A Comparative Review of Air Drones (UAVs) and Delivery Bots (SUGVs) for Automated Last Mile Home Delivery. *Logistics* **2023**, *7*, 21. [CrossRef]
17. Ullah, F.; Al-Turjman, F.; Qayyum, S.; Inam, H.; Imran, M. Advertising through UAVs: Optimized path system for delivering smart real-estate advertisement materials. *Int. J. Intell. Syst.* **2021**, *36*, 3429–3463. [CrossRef]
18. Jeyavel, J.; Prasad, A.A.; Shelke, K.M.; Sargade, P.D.; Thoke, U.V. Survey on Fire Fighting techniques using Unmanned Aerial Vehicles. In Proceedings of the 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 4–5 March 2021; pp. 239–241. [CrossRef]
19. Velusamy, P.; Rajendran, S.; Mahendran, R.K.; Naseer, S.; Shafiq, M.; Choi, J.G. Unmanned Aerial Vehicles (UAV) in Precision Agriculture: Applications and Challenges. *Energies* **2022**, *15*, 217. [CrossRef]
20. Lekidis, A.; Anastasiadis, A.G.; Vokas, G.A. Electricity infrastructure inspection using AI and edge platform-based UAVs. *Energy Rep.* **2022**, *8*, 1394–1411. [CrossRef]
21. Skondras, A.; Karachaliou, E.; Tavantzis, I.; Tokas, N.; Valari, E.; Skalidi, I.; Bouvet, G.A.; Stylianidis, E. UAV Mapping and 3D Modeling as a Tool for Promotion and Management of the Urban Space. *Drones* **2022**, *6*, 115. [CrossRef]
22. Rachmawati, T.S.N.; Kim, S. Unmanned Aerial Vehicles (UAV) Integration with Digital Technologies Toward Construction 4.0: A Systematic Literature Review. *Sustainability* **2022**, *14*, 5708. [CrossRef]
23. Jiang, S.; Jiang, W.; Wang, L. Unmanned Aerial Vehicle-Based Photogrammetric 3D Mapping: A survey of techniques, applications, and challenges. *IEEE Geosci. Remote Sens. Mag.* **2022**, *10*, 135–171. [CrossRef]
24. Pitas, I.; Mademlis, I. Autonomous UAV Cinematography. In Proceedings of the 30th ACM International Conference on Multimedia (MM '22), New York, NY, USA, 10–14 October 2022; pp. 7374–7376. [CrossRef]
25. Feng, W.; Dandan, L.; Qiang, C. Discussion on the Application of UAV Oblique Photography in the Registration of Rural Housing and Real Estate Integration. In Proceedings of the 2020 IEEE 5th International Conference on Intelligent Transportation Engineering (ICITE), Beijing, China, 11–13 September 2020; pp. 290–295. [CrossRef]
26. Heiets, I.; Kuo, Y.W.; La, J.; Yeun, R.C.; Verhagen, W. Future Trends in UAV Applications in the Australian Market. *Aerospace* **2023**, *10*, 555. [CrossRef]

27. Sevilla-Sevilla, C.; Mendieta-Aragón, A.; Ruiz-Gómez, L.M. Drones in hospitality and tourism: A literature review and research agenda. *Tourism Review* **2024**, *79*, 378–391. [CrossRef]
28. Tham, A. Negotiating leisure etiquette in the context of drones. *Leisure/Loisir* **2020**, *44*, 105–126. [CrossRef]
29. Tyshchuk, V.V. A Review of Legal Regulation Regarding the Use of Unmanned Aerial Vehicles for Border Security and the Impact of Global Technologies. *Int. Comp. Jurisprud.* **2024**, *10*, 61–81. [CrossRef]
30. Lyu, M.; Zhao, Y.; Huang, C.; Huang, H. Unmanned Aerial Vehicles for Search and Rescue: A Survey. *Remote Sens.* **2023**, *15*, 3266. [CrossRef]
31. Couturier, A.; Akhloufi, M.A. A review on absolute visual localization for UAV. *Robot. Auton. Syst.* **2021**, *135*, 103666. [CrossRef]
32. Ali, B.; Sadekov, R.N.; Tsodokova, V.V. A Review of Navigation Algorithms for Unmanned Aerial Vehicles Based on Computer Vision Systems. *Gyroscopy Navig.* **2022**, *13*, 241–252. [CrossRef]
33. Rezwan, S.; Choi, W. Artificial Intelligence Approaches for UAV Navigation: Recent Advances and Future Challenges. *IEEE Access* **2022**, *10*, 26320–26339. [CrossRef]
34. Arafat, M.Y.; Alam, M.M.; Moh, S. Vision-Based Navigation Techniques for Unmanned Aerial Vehicles: Review and Challenges. *Drones* **2023**, *7*, 89. [CrossRef]
35. Wang, K.; Kooistra, L.; Pan, R.; Wang, W.; Valente, J. UAV-based simultaneous localization and mapping in outdoor environments: A systematic scoping review. *J. Field Robot.* **2024**, *41*, 1617–1642. [CrossRef]
36. Lu, Y.; Xue, Z.; Xia, G.S.; Zhang, L. A survey on vision-based UAV navigation. *Geo-Spat. Inf. Sci.* **2018**, *21*, 21–32. [CrossRef]
37. Al-Kaff, A.; Martín, D.; García, F.; de la Escalera, A.; María Armingol, J. Survey of computer vision algorithms and applications for unmanned aerial vehicles. *Expert Syst. Appl.* **2018**, *92*, 447–463. [CrossRef]
38. Belmonte, L.M.; Morales, R.; Fernández-Caballero, A. Computer Vision in Autonomous Unmanned Aerial Vehicles—A Systematic Mapping Study. *Appl. Sci.* **2019**, *9*, 3196. [CrossRef]
39. Balamurugan, G.; Valarmathi, J.; Naidu, V. Survey on UAV navigation in GPS denied environments. In Proceedings of the 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi, India, 3–5 October 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 198–204. [CrossRef]
40. Palinje, M.; Sinha, S.K. Performance Evaluation Of Emerging Semiconductor Devices For Low Power Applications: A Review. In Proceedings of the 2022 10th International Conference on Emerging Trends in Engineering and Technology-Signal and Information Processing (ICETET-SIP-22), Nagpur, India, 29–30 April 2022; pp. 1–6. [CrossRef]
41. Radamson, H.; Miao, Y.; Zhou, Z.; Wu, Z.; Kong, Z.; Gao, J.; Yang, H.; Ren, Y.; Zhang, Y.; Shi, J.; et al. CMOS Scaling for the 5 nm Node and Beyond: Device, Process and Technology. *Nanomaterials* **2024**, *14*, 837. [CrossRef]
42. Bottino, V.; de Alteriis, G.; Schiano Lo Moriello, R.; Accardo, D. System Architecture Design of a UAV for Automated Cinematography in GNSS-Challenging Scenarios. In Proceedings of the 2022 IEEE 9th International Workshop on Metrology for AeroSpace (MetroAeroSpace), Pisa, Italy, 27–29 July 2022; pp. 627–632. [CrossRef]
43. Karthik, S.; Basha, N.; Ajith, V.S. Design of Hexacopter Tethered Drone. *Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET)* **2023**, *11*, 48629. [CrossRef]
44. Singh, Y.; Sangam, R.; Karmarkar, A. Assembly and Programming of Hexacopter Drone for Surveillance and Medical Kit Delivery Purposes. *Int. Res. J. Adv. Eng. Health Sci.* **2024**, *6*, 1794–1803. [CrossRef]
45. Mallick, T.C.; Bhuyan, M.A.I.; Munna, M.S. Design & implementation of an UAV (Drone) with flight data record. In Proceedings of the 2016 International Conference on Innovations in Science, Engineering and Technology (ICISET), Dhaka, Bangladesh, 28–29 October 2016; pp. 1–6. [CrossRef]
46. European Space Agency. Galileo Initial Services. 2016. Available online: https://www.esa.int/ESA_Multimedia/Transmissions/2016/12/Galileo_initial_services (accessed on 4 August 2024).
47. EUROPEAN COMMISSION, Directorate-General Defence Industry and Space, EU Satellite Navigation Programmes. GALILEO SERVICE NOTICE #09. 2021. Issue: 1.0, Service Notice to Galileo Users (SNGU) Number: 2021005. 2021. Available online: https://www.gsc-europa.eu/sites/default/files/sites/all/files/Galileo-service-notice-09-v1.0.pdf (accessed on 11 July 2024).
48. Wang, X.; Guo, J.; Cui, N. Adaptive extended Kalman filtering applied to low-cost MEMS IMU/GPS integration for UAV. In Proceedings of the 2009 International Conference on Mechatronics and Automation, Changchun, China, 9–12 August 2009; pp. 2214–2218. [CrossRef]
49. Abdelkrim, N.; Aouf, N.; Tsourdos, A.; White, B. Robust nonlinear filtering for INS/GPS UAV localization. In Proceedings of the 2008 16th Mediterranean Conference on Control and Automation, Ajaccio, France, 25–27 June 2008; pp. 695–702. [CrossRef]
50. Sasiadek, J.; Wang, Q.; Johnson, R.; Sun, L.; Zalewski, J. UAV Navigation Based on Parallel Extended Kalman Filter. In Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit. American Institute of Aeronautics and Astronautics, Guidance, Navigation, and Control and Co-Located Conferences, Dever, CO, USA, 14–17 August 2000. [CrossRef]
51. U-Blox. *UBX-M8030 Product Summary: Versatile U-Blox M8 GNSS Chips*; Technical Report; U-Blox: Phoenix, AZ, USA, 2019.
52. InvenSense, Inc. *ICM-40609-D: A TDK SmartMotion™ Solution Designed for Drone Market*; Document Number: DS-000330, Revision: 1.2; InvenSense, Inc.: San Jose, CA, USA, 2022. Available online: https://invensense.tdk.com/wp-content/uploads/2022/07/DS-000330-ICM-40609-D-v1.2.pdf (accessed on 12 July 2024).
53. Foxlin, E. Pedestrian tracking with shoe-mounted inertial sensors. *IEEE Comput. Graph. Appl.* **2005**, *25*, 38–46. [CrossRef] [PubMed]

54. Drones, P. ANAFI Ai: The 4G Robotic UAV. 2022. Available online: https://www.parrot.com/assets/s3fs-public/2022-01/whitepaperanafiai.pdf (accessed on 12 July 2024).
55. Drones, P. ANAFI USA WHITE PAPER V1.6. 2022. Available online: https://www.parrot.com/assets/s3fs-public/2022-01/white-paper-anafi-usa-v1.5.3_en.pdf (accessed on 12 July 2024).
56. Guizilini, V.; Ramos, F. Visual odometry learning for unmanned aerial vehicles. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 6213–6220. [CrossRef]
57. Depaola, R.; Chimento, C.; Anderson, M.L.; Brink, K.; Willis, A. Uav navigation with computer vision–flight testing a novel visual odometry technique. In Proceedings of the 2018 AIAA Guidance, Navigation, and Control Conference, Kissimmee, FL, USA, 8–12 January 2018. [CrossRef]
58. Caballero, F.; Merino, L.; Ferruz, J.; Ollero, A. Vision-Based Odometry and SLAM for Medium and High Altitude Flying UAVs. *J. Intell. Robot. Syst.* **2009**, *54*, 137–161. [CrossRef]
59. Cvišić, I.; Ćesić, J.; Marković, I.; Petrović, I. SOFT-SLAM: Computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles. *J. Field Robot.* **2018**, *35*, 578–595. [CrossRef]
60. More, V.; Kumar, H.; Kaingade, S.; Gaidhani, P.; Gupta, N. Visual odometry using optic flow for Unmanned Aerial Vehicles. In Proceedings of the 2015 International Conference on Cognitive Computing and Information Processing(CCIP), Noida, India, 3–4 March 2015; pp. 1–6. [CrossRef]
61. Surber, J.; Teixeira, L.; Chli, M. Robust visual-inertial localization with weak GPS priors for repetitive UAV flights. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 6300–6306. [CrossRef]
62. Goppert, J.; Yantek, S.; Hwang, I. Invariant Kalman filter application to optical flow based visual odometry for UAVs. In Proceedings of the 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), Milan, Italy, 4–7 July 2017; pp. 99–104. [CrossRef]
63. Artieda, J.; Sebastian, J.M.; Campoy, P.; Correa, J.F.; Mondragón, I.F.; Martínez, C.; Olivares, M. Visual 3-D SLAM from UAVs. *J. Intell. Robot. Syst.* **2009**, *55*, 299–321. [CrossRef]
64. Li, J.; Bi, Y.; Lan, M.; Qin, H.; Shan, M.; Lin, F.; Chen, B.M. Real-time simultaneous localization and mapping for uav: A survey. In *Proceedings of the International Micro Air Vehicle Competition and Conference*; National University of Singapore: Singapore, 2016; Volume 2016, p. 237.
65. Strydom, R.; Thurrowgood, S.; Srinivasan, M.V. Visual odometry: Autonomous uav navigation using optic flow and stereo. In Proceedings of the Australasian Conference on Robotics and Automation, ACRA. Australian Robotics and Automation Association, Melbourne, Australia, 2–4 December 2014; Volume 2.
66. Nister, D.; Naroditsky, O.; Bergen, J. Visual odometry. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, Washington, DC, USA, 27 June–2 July 2004; Volume 1, p. I. [CrossRef]
67. Forster, C.; Pizzoli, M.; Scaramuzza, D. SVO: Fast semi-direct monocular visual odometry. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 15–22. [CrossRef]
68. Leutenegger, S.; Lynen, S.; Bosse, M.; Siegwart, R.; Furgale, P. Keyframe-based visual–inertial odometry using nonlinear optimization. *Int. J. Robot. Res.* **2015**, *34*, 314–334. [CrossRef]
69. Mur-Artal, R.; Montiel, J.M.M.; Tardós, J.D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Trans. Robot.* **2015**, *31*, 1147–1163. [CrossRef]
70. Davison, A.J.; Reid, I.D.; Molton, N.D.; Stasse, O. MonoSLAM: Real-Time Single Camera SLAM. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 1052–1067. [CrossRef]
71. Engel, J.J.; Schöps, T.; Cremers, D. LSD-SLAM: Large-Scale Direct Monocular SLAM. In Proceedings of the Computer Vision–ECCV 2014, Zurich, Switzerland, 6–12 September 2014; pp. 834–849. [CrossRef]
72. SONUGÜR, G. A Review of quadrotor UAV: Control and SLAM methodologies ranging from conventional to innovative approaches. *Robot. Auton. Syst.* **2023**, *161*, 104342. [CrossRef]
73. Bettemir, O. Error analysis of visual odometry for a small size unmanned aerial vehicle. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2024**, *48*, 83–90. [CrossRef]
74. Bloesch, M.; Omari, S.; Hutter, M.; Siegwart, R. Robust visual inertial odometry using a direct EKF-based approach. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 298–304. [CrossRef]
75. Tong, P.; Yang, X.; Yang, Y.; Liu, W.; Wu, P. Multi-UAV Collaborative Absolute Vision Positioning and Navigation: A Survey and Discussion. *Drones* **2023**, *7*, 261. [CrossRef]
76. Van Dalen, G.J.; Magree, D.P.; Johnson, E.N. Absolute localization using image alignment and particle filtering. In Proceedings of the AIAA Guidance, Navigation, and Control Conference, San Diego, CA, USA, 4–8 January 2016; p. 0647. [CrossRef]
77. Yol, A.; Delabarre, B.; Dame, A.; Dartois, J.E.; Marchand, E. Vision-based absolute localization for unmanned aerial vehicles. In Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 14–18 September 2014; pp. 3429–3434. [CrossRef]
78. Wan, X.; Liu, J.; Yan, H.; Morgan, G.L.K. Illumination-invariant image matching for autonomous UAV localisation based on optical sensing. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 198–213. [CrossRef]

79. Patel, B. Visual Localization for UAVs in Outdoor GPS-Denied Environments. Master's Thesis, University of Toronto, Toronto, ON, Canada, 2019. Available online: https://www.proquest.com/openview/ff54ad672f25af4497a2176ba06f47a1/1 (accessed on 12 July 2024).
80. Couturier, A.; Akhloufi, M. Conditional Probabilistic Relative Visual Localization for Unmanned Aerial Vehicles. In Proceedings of the 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), London, ON, Canada, 30 August–2 September 2020; pp. 1–4. [CrossRef]
81. Couturier, A.; Akhloufi, M.A. UAV navigation in GPS-denied environment using particle filtered RVL. In Proceedings of the Situation Awareness in Degraded Environments 2019, Baltimore, MD, USA, 16–17 April 2019; Sanders-Reed, J.N., Arthur, J.J., III, Eds.; International Society for Optics and Photonics; SPIE: Bellingham, WA, USA, 2019; Volume 11019, pp. 188–198. [CrossRef]
82. Couturier, A.; Akhloufi, M.A. Relative visual localization (RVL) for UAV navigation. In Proceedings of the Degraded Environments: Sensing, Processing, and Display 2018, Orlando, FL, USA, 17–18 April 2018; Sanders-Reed, J.N., Arthur, J.J., III, Eds.; International Society for Optics and Photonics; SPIE: Bellingham, WA, USA, 2018; Volume 10642, pp. 213–226. [CrossRef]
83. Couturier, A. Auto-Localisation de Véhicules Aériens Autonomes avec caméras Monoculaires en Environnements Hostiles. ProQuest Dissertations, Universite de Moncton, Moncton, NB, Canada, 2019. Available online: https://www.proquest.com/dissertations-theses/auto-localisation-de-vehicules-aeriens-autonomes/docview/2584328452/se-2 (accessed on 12 July 2024).
84. Seema, B.S.; Hemanth, K.; Naidu, V.P.S. Geo-Registration of Aerial Images using RANSAC Algorithm. In Proceedings of the NCTAESD-2014, Vemana Institute of Technology, Bangalore, India, 25 April 2014; pp. 1–5.
85. Shan, M.; Wang, F.; Lin, F.; Gao, Z.; Tang, Y.Z.; Chen, B.M. Google map aided visual navigation for UAVs in GPS-denied environment. In Proceedings of the 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), Zhuhai, China, 6–9 December 2015; pp. 114–119. [CrossRef]
86. Chiu, H.P.; Das, A.; Miller, P.; Samarasekera, S.; Kumar, R. Precise vision-aided aerial navigation. In Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 14–18 September 2014; pp. 688–695. [CrossRef]
87. Mantelli, M.; Pittol, D.; Neuland, R.; Ribacki, A.; Maffei, R.; Jorge, V.; Prestes, E.; Kolberg, M. A novel measurement model based on abBRIEF for global localization of a UAV over satellite images. *Robot. Auton. Syst.* **2019**, *112*, 304–319. [CrossRef]
88. Shan, M.; Charan, A. Google map referenced UAV navigation via simultaneous feature detection and description. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–3 October 2015.
89. Rosten, E.; Drummond, T. Machine Learning for High-Speed Corner Detection. In Proceedings of the Computer Vision–ECCV 2006, Graz, Austria, 7–13 May 2006; Leonardis, A., Bischof, H., Pinz, A., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2006; pp. 430–443. [CrossRef]
90. Rosten, E.; Porter, R.; Drummond, T. Faster and Better: A Machine Learning Approach to Corner Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 105–119. [CrossRef] [PubMed]
91. Amer, K.; Samy, M.; ElHakim, R.; Shaker, M.; ElHelw, M. Convolutional Neural Network-Based Deep Urban Signatures with Application to Drone Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 2138–2145. [CrossRef]
92. Nassar, A.; Amer, K.; ElHakim, R.; ElHelw, M. A deep cnn-based framework for enhanced aerial imagery registration with applications to uav geolocalization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1513–1523. [CrossRef]
93. Nassar, A.; ElHelw, M. Aerial Imagery Registration Using Deep Learning for UAV Geolocalization. In *Deep Learning in Computer Vision: Principles and Applications*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2020; Chapter 7; pp. 183–210.
94. Marcu, A.; Costea, D.; Slusanschi, E.; Leordeanu, M. A Multi-Stage Multi-Task Neural Network for Aerial Scene Interpretation and Geolocalization. *arXiv* **2018**, arXiv:1804.01322. [CrossRef]
95. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556. [CrossRef]
96. Goforth, H.; Lucey, S. GPS-Denied UAV Localization using Pre-existing Satellite Imagery. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 2974–2980. [CrossRef]
97. Schleiss, M. Translating Aerial Images Into Street-Map Representations for Visual Self-Localization of Uavs. *ISPRS-Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *4213*, 575–580. [CrossRef]
98. Valaitis, V.; Marcinkevicius, V.; Jurevicius, R. Learning Aerial Image Similarity Using Triplet Networks. In *Numerical Computations: Theory and Algorithms*; Sergeyev, Y.D., Kvasov, D.E., Eds.; Springer: Cham, Switzerland, 2020; pp. 195–207. [CrossRef]
99. Ahn, S.H.; Kang, H.S.; Lee, J.M. Satellite Image-based UAV Localization Using Siamese Neural Network. In Proceedings of the International Conference on Artificial Life and Robotics, Virtual, 20–21 January 2021; ALife Robotics Corporation Ltd.: Oita, Japan, 2021; Volume 26, pp. 513–516. [CrossRef]
100. Couturier, A.; Akhloufi, M.A. Convolutional neural networks and particle filter for UAV localization. In Proceedings of the Unmanned Systems Technology XXIII, Online, 12–16 April 2021; Nguyen, H.G., Muench, P.L., Skibba, B.K., Eds.; International Society for Optics and Photonics; SPIE: Bellingham, WA, USA, 2021; Volume 11758, pp. 108–120. [CrossRef]
101. Dai, M.; Hu, J.; Zhuang, J.; Zheng, E. A Transformer-Based Feature Segmentation and Region Alignment Method for UAV-View Geo-Localization. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 4376–4389. [CrossRef]

102. Ahn, S.; Kang, H.; Lee, J. Aerial-Satellite Image Matching Framework for UAV Absolute Visual Localization using Contrastive Learning. In Proceedings of the 2021 21st International Conference on Control, Automation and Systems (ICCAS), Jeju, Republic of Korea, 12–15 October 2021; pp. 143–146. [CrossRef]
103. Chechik, G.; Sharma, V.; Shalit, U.; Bengio, S. Large Scale Online Learning of Image Similarity Through Ranking. *J. Mach. Learn. Res.* **2010**, *11*, 1109–1135.
104. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, 7–12 June; pp. 815–823. [CrossRef]
105. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, present, and future of face recognition: A review. *Electronics* **2020**, *9*, 1188. [CrossRef]
106. Eling, C.; Klingbeil, L.; Kuhlmann, H. Real-Time Single-Frequency GPS/MEMS-IMU Attitude Determination of Lightweight UAVs. *Sensors* **2015**, *15*, 26212–26235. [CrossRef]
107. Bodó, Z.; Lantos, B. State estimation for UAVs using sensor fusion. In Proceedings of the 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, Serbia, 14–16 September 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 000111–000116. [CrossRef]
108. Aparna, G.; Kamal, C.; Motta, R.N. IMU Based Attitude Estimation Using Adaptive Complimentary Filter. In Proceedings of the 2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS), Mumbai, India, 25–27 June 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–5. [CrossRef]
109. Öz, M.N.; Budak, S.; Kurnaz, E.; Durdu, A. Orientation Determination in IMU Sensor with Complementary Filter. *Turk. J. Forecast.* **2022**, *6*, 34–39. [CrossRef]
110. Han, S. Measuring displacement signal with an accelerometer. *J. Mech. Sci. Technol.* **2010**, *24*, 1329–1335. [CrossRef]
111. Prikhodko, I.P.; Gregory, J.A.; Bugrov, D.I.; Judy, M.W. Overcoming limitations of Rate Integrating Gyroscopes by virtual rotation. In Proceedings of the 2016 IEEE International Symposium on Inertial Sensors and Systems, Laguna Beach, CA, USA, 22–25 February 2016; pp. 5–8. [CrossRef]
112. Zhang, Q.; Niu, X.; Shi, C. Impact Assessment of Various IMU Error Sources on the Relative Accuracy of the GNSS/INS Systems. *IEEE Sens. J.* **2020**, *20*, 5026–5038. [CrossRef]
113. IV, W.S.F.; Wall, J.H.; Bevly, D.M. Characterization of Various IMU Error Sources and the Effect on Navigation Performance. In Proceedings of the 18th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2005), Long Beach, CA, USA, 13–16 September 2005; pp. 967–978.
114. Unsal, D.; Demirbas, K. Estimation of deterministic and stochastic IMU error parameters. In Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium, Myrtle Beach, SC, USA, 23–26 April 2012; pp. 862–868. [CrossRef]
115. Ribeiro, J.G.T.; de Castro, J.T.P.; Meggiolaro, M.A. An algorithm to minimize errors in displacement measurements via double integration of noisy acceleration signals. *J. Braz. Soc. Mech. Sci. Eng.* **2021**, *43*, 385. [CrossRef]
116. Valtonen Örnhag, M.; Persson, P.; Wadenbäck, M.; Åström, K.; Heyden, A. Trust Your IMU: Consequences of Ignoring the IMU Drift. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Orleans, LA, USA, 18–24 June 2022; pp. 4467–4476. [CrossRef]
117. Hardt-Stremayr, A.; Schörghuber, M.; Weiss, S.; Humenberger, M. Metrics for Real-Time Mono-VSLAM Evaluation Including IMU Induced Drift with Application to UAV Flight. In Proceedings of the Computer Vision–ECCV 2018 Workshops, Munich, Germany, 8–14 September 2019; Leal-Taixé, L., Roth, S., Eds.; Springer: Cham, Switzerland, 2019; pp. 73–87. [CrossRef]
118. Weideman, J.A.C. Numerical Integration of Periodic Functions: A Few Examples. *Am. Math. Mon.* **2002**, *109*, 21–36. [CrossRef]
119. InvenSense, I. MPU-6000 and MPU-6050 Product Specification. 2013. Available online: https://invensense.tdk.com/wp-content/uploads/2015/02/MPU-6000-Datasheet1.pdf (accessed on 12 July 2024).
120. InvenSense, I. ICM-40609-D: Designed for Drones. 2022. Available online: https://invensense.tdk.com/products/motion-tracking/6-axis/icm-40609-d/, (accessed on 12 July 2024).
121. InvenSense, I. 6-Axis MEMS Motion Sensors. 2022. Available online: https://invensense.tdk.com/products/motion-tracking/6-axis/ (accessed on 12 July 2024).
122. DJI. DJI Mini 3. 2024. Available online: https://www.dji.com/ca/mini-3 (accessed on 20 October 2024).
123. CChin; Mokhtari, S.; Rodriguez, K. DJI Mini 3 Chip ID. 2023. Available online: https://www.ifixit.com/Guide/DJI+Mini+3+Chip+ID/160247 (accessed on 12 July 2024).
124. Gao, Y.; Li, G. A GNSS Instrumentation Covert Directional Spoofing Algorithm for UAV Equipped with Tightly-Coupled GNSS/IMU. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 8501413. [CrossRef]
125. Varshosaz, M.; Afary, A.; Mojaradi, B.; Saadatseresht, M.; Ghanbari Parmehr, E. Spoofing Detection of Civilian UAVs Using Visual Odometry. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 6. [CrossRef]
126. Gu, N.; Xing, F.; You, Z. Visual/Inertial/GNSS Integrated Navigation System under GNSS Spoofing Attack. *Remote Sens.* **2022**, *14*, 5975. [CrossRef]
127. Gu, N.; Xing, F.; You, Z. GNSS Spoofing Detection Based on Coupled Visual/Inertial/GNSS Navigation System. *Sensors* **2021**, *21*, 6769. [CrossRef]

128. Xue, N.; Niu, L.; Hong, X.; Li, Z.; Hoffaeller, L.; Pöpper, C. DeepSIM: GPS Spoofing Detection on UAVs Using Satellite Imagery Matching. In Proceedings of the 36th Annual Computer Security Applications Conference (ACSAC '20), New York, NY, USA, 7–11 December 2020; pp. 304–319. [CrossRef]
129. Qiao, Y.; Zhang, Y.; Du, X. A Vision-Based GPS-Spoofing Detection Method for Small UAVs. In Proceedings of the 2017 13th International Conference on Computational Intelligence and Security (CIS), Hong Kong, China, 15–18 December 2017; pp. 312–316. [CrossRef]
130. Dymkova, S.S.; Dymkov, A.D. Experimental Studies of GNSS Errors in Rough and Wooded Mountainous Terrain. In Proceedings of the 2021 International Conference on Engineering Management of Communication and Technology (EMCTECH), Vienna, Austria, 20–22 October 2021; pp. 1–6. [CrossRef]
131. Zhao, X.; Tang, P.; Song, Q.; Jiang, T.; Wang, Y.; Tian, L.; Li, W.; Zhang, J. Experimental Analysis of Multipath Effects on GNSS Positioning in Urban Canyon. In Proceedings of the 2021 IEEE 4th International Conference on Electronic Information and Communication Technology (ICEICT), Xi'an, China, 18–20 August 2021; pp. 557–562. [CrossRef]
132. Feng, T.; Chen, S.; Feng, Z.; Shen, C.; Tian, Y. Effects of Canopy and Multi-Epoch Observations on Single-Point Positioning Errors of a GNSS in Coniferous and Broadleaved Forests. *Remote Sens.* **2021**, *13*, 2325. [CrossRef]
133. Lee, T.; Bettinger, P.; Merry, K.; Cieszewski, C. The effects of nearby trees on the positional accuracy of GNSS receivers in a forest environment. *PLoS ONE* **2023**, *18*, 1–20. [CrossRef] [PubMed]
134. McGraw, G.A.; Groves, P.D.; Ashman, B.W. Robust Positioning in the Presence of Multipath and NLOS GNSS Signals. In *Position, Navigation, and Timing Technologies in the 21st Century*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2020; Chapter 22; pp. 551–589. [CrossRef]
135. Morong, T.; Puričer, P.; Kovář, P. Study of the GNSS Jamming in Real Environment. *Int. J. Electron. Telecommun.* **2019**, *65*, 65–70. [CrossRef]
136. Moskaliuk, T.; Malatest, B. Russia Versus Ukraine and the Role of Software-Defined Radios. 2023. Available online: https://www.afcea.org/signal-media/cyber-edge/russia-versus-ukraine-and-role-software-defined-radios, (accessed on 6 August 2024).
137. Shevchenko, V. Russia Blamed for GPS Interference Affecting Flights in Europe. 2024. Available online: https://www.bbc.com/news/articles/cne900k4wvjo (accessed on 12 July 2024).
138. Slisco, A. Russia Suspected of Jamming GPS Signals for Thousands of Airline Flights. 2024. Available online: https://www.newsweek.com/russia-suspected-jamming-gps-signals-thousands-airline-flights-1893508 (accessed on 12 July 2024).
139. Khalil, J. EASA Updates Advisory on Navigation Interference. 2024. Available online: https://www.gpsworld.com/easa-updates-advisory-on-navigation-interference/ (accessed on 12 July 2024).
140. Live GPS Spoofing Tracker Map. Live GPS Spoofing Tracker Map. 2024. Available online: https://spoofing.skai-data-services.com/ (accessed on 12 July 2024).
141. GPSJAM GPS/GNSS Interference Map. GPSJAM GPS/GNSS Interference Map. 2024. Available online: https://gpsjam.org/ (accessed on 12 July 2024).
142. Flightradar24. GPS Jamming Map. 2024. Available online: https://www.flightradar24.com/ (accessed on 12 July 2024).
143. Thurber, M. GNSS Jamming and Spoofing Events Present a Growing Danger to Aviation. 2024. Available online: https://www.ainonline.com/aviation-news/air-transport/2024-03-04/gnss-jamming-and-spoofing-events-present-growing-danger (accessed on 12 July 2024).
144. Gorman, S. Global Navigation Jamming Will Only Get Worse. The U.S. Needs to Move Fast. 2024. Available online: https://spacenews.com/global-navigation-jamming-will-only-get-worse-us-needs-move-fast/ (accessed on 12 July 2024).
145. Security, M.G. 2019 Jamming and Spoofing of GNSS. 2019. Available online: https://www.maritimeglobalsecurity.org/media/1043/2019-jamming-spoofing-of-gnss.pdf (accessed on 12 July 2024).
146. Coulon, M.; Chabory, A.; Garcia-Pena, A.; Vezinet, J.; Macabiau, C.; Estival, P.; Ladoux, P.; Roturier, B. Characterization of Meaconing and its Impact on GNSS Receivers. In Proceedings of the 33rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2020), Virtual, 21–25 September 2020; pp. 3713–3737. [CrossRef]
147. Steiner, J.; Pleninger, S.; Hospodka, J. Assessing the Vulnerability of Aviation Systems to GNSS Meaconing Attacks. In Proceedings of the 2024 New Trends in Civil Aviation (NTCA), Prague, Czech Republic, 25–26 April 2024; pp. 213–218. [CrossRef]
148. Lenhart, M.; Spanghero, M.; Papadimitratos, P. Distributed and Mobile Message Level Relaying/Replaying of GNSS Signals. In Proceedings of the 2022 International Technical Meeting of The Institute of Navigation, Long Beach, CA, USA, 25–27 January 2022; pp. 56–67. [CrossRef]
149. Peterson, S.; Faramarzi, P. Exclusive: Iran Hijacked US Drone, Says Iranian Engineer. 2011. Available online: https://www.csmonitor.com/World/Middle-East/2011/1215/Exclusive-Iran-hijacked-US-drone-says-Iranian-engineer (accessed on 12 July 2024).
150. Staff, C. Iran Shows Off Downed US Drone. 2011. Available online: https://edition.cnn.com/2011/12/12/world/meast/iran-us-drone/index.html (accessed on 12 July 2024).
151. News, D. Iran Carries Successful Test Flight of Reverse Engineered RQ-170. 2014. Available online: https://web.archive.org/web/20141110130745/http://defenceradar.com/2014/11/10/iran-carries-successful-test-flight-of-reverse-engineered-rq-170/ (accessed on 12 July 2024).

152. Cenciotti, D. Iran Unveils New UCAV Modeled on Captured U.S. RQ-170 Stealth Drone. 2016. Available online: https://theaviationist.com/2016/10/02/iran-unveils-new-ucav-modeled-on-captured-u-s-rq-170-stealth-drone/ (accessed on 12 July 2024).
153. Agencies in Tehran. Iran Builds Attack Drone Similar to Captured US Model, Local Media Say. 2016. Available online: https://www.theguardian.com/world/2016/oct/01/iran-builds-thunderbolt-drone (accessed on 12 July 2024).
154. Gross, J.A.; Israel Staff, T. Iranian UAV That Entered Israeli Airspace Seems to Be American Stealth Knock-Off. 2018. Available online: https://www.timesofisrael.com/iranian-uav-that-entered-israeli-airspace-seems-to-be-american-stealth-knock-off/ (accessed on 12 July 2024).
155. Meng, L.; Yang, L.; Yang, W.; Zhang, L. A Survey of GNSS Spoofing and Anti-Spoofing Technology. *Remote Sens.* **2022**, *14*, 4826. [CrossRef]
156. Editor. Circle Spoofing Comes to Aviation-First the Baltic, Now the Mediterranean. 2024. Available online: https://rntfnd.org/2024/01/07/circle-spoofing-comes-to-aviation-first-the-baltic-now-the-mediterranean/ (accessed on 12 July 2024).
157. Pullen, S. GNSS Spoofing and Jamming in Eastern Europe. 2024. Available online: https://insidegnss.com/gnss-spoofing-and-jamming-in-eastern-europe/, (accessed on 12 July 2024).
158. Chen, X.; Luo, R.; Liu, T.; Yuan, H.; Wu, H. Satellite Navigation Signal Authentication in GNSS: A Survey on Technology Evolution, Status, and Perspective for BDS. *Remote Sens.* **2023**, *15*, 1462. [CrossRef]
159. Yuan, M.; Tang, X.; Ou, G. Authenticating GNSS Civilian Signals: A Survey. *Satell. Navig.* **2023**, *4*, 6. [CrossRef]
160. Margaria, D.; Motella, B.; Anghileri, M.; Floch, J.J.; Fernandez-Hernandez, I.; Paonni, M. Signal Structure-Based Authentication for Civil GNSSs: Recent Solutions and Perspectives. *IEEE Signal Process. Mag.* **2017**, *34*, 27–37. [CrossRef]
161. Caparra, G.; Wullems, C.; Ceccato, S.; Sturaro, S.; Laurenti, N.; Pozzobon, O.; Ioannides, R.T.; Crisci, M. *Navigation Message Authentication Schemes*; Inside GNSS: Eugene, OR, USA, 2016.
162. Qie, R.; Zhang, Z.; Xu, Y. UAV-Satellite Cross-View Image Matching Based on Siamese Network. In *Computational and Experimental Simulations in Engineering*; Li, S., Ed.; Springer: Cham, Switzerland, 2024; pp. 1139–1151. [CrossRef]
163. Blewitt, G. Basics of the GPS technique: Observation equations. *Geod. Appl. GPS* **1997**, *1*, 46.
164. Caparra, G.; Curran, J.T. On the achievable equivalent security of GNSS ranging code encryption. In Proceedings of the 2018 IEEE/ION Position, Location and Navigation Symposium (PLANS), Monterey, CA, USA, 23–26 April 2018; pp. 956–966. [CrossRef]
165. Anderson, J.M.; Carroll, K.L.; DeVilbiss, N.P.; Gillis, J.T.; Hinks, J.C.; O'Hanlon, B.W.; Rushanan, J.J.; Scott, L.; Yazdi, R.A. Chips-Message Robust Authentication (Chimera) for GPS Civilian Signals. In Proceedings of the 30th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2017), Portland, OR, USA, 25–29 September 2017; pp. 2388–2416. [CrossRef]
166. Anderson, J.; Lo, S.; Walter, T. Authentication Security of Combinatorial Watermarking for GNSS Signal Authentication. *NAVIGATION J. Inst. Navig.* **2024**, *71*, navi.655. [CrossRef]
167. Petovello, M. What Is Navigation Message Authentication? 2018. Available online: https://insidegnss.com/auto/janfeb18-SOLUTIONS.pdf (accessed on 12 July 2024).
168. Egea-Roca, D.; Arizabaleta-Diez, M.; Pany, T.; Antreich, F.; López-Salcedo, J.A.; Paonni, M.; Seco-Granados, G. GNSS User Technology: State-of-the-Art and Future Trends. *IEEE Access* **2022**, *10*, 39939–39968. [CrossRef]
169. Anthony, R.T. NAVSTAR GPS Space Segment/Navigation User Segment Interface. 2022. Available online: https://www.gps.gov/technical/icwg/IS-GPS-200N.pdf (accessed on 11 July 2024).
170. Barker, B.C.; Betz, J.W.; Clark, J.E.; Correia, J.T.; Gillis, J.T.; Lazar, S.; Rehborn, K.A.; Straton, J.R. Overview of the GPS M Code Signal. In Proceedings of the 2000 National Technical Meeting of The Institute of Navigation, The Institute of Navigation, Anaheim, CA, USA, 26–28 January 2000; pp. 542–549.
171. European GNSS Service Centre (GSC). Public Regulated Service. Available online: https://www.gsc-europa.eu/galileo/services/public-regulated-service (accessed on 11 July 2024).
172. Global Navigation Satellite System: Glonass Interface Control Document, Navigational Radiosignal in Bands L1, L2 (Edition 5.1). 2008. Available online: http://gauss.gge.unb.ca/GLONASS.ICD.pdf (accessed on 11 July 2024).
173. de Selding, P.B. Europe, China Remain at Odds over Navigation Systems. 2009. Available online: https://spacenews.com/europe-china-remain-odds-over-navigation-systems/ (accessed on 11 July 2024).
174. Cabinet Office, G.O.J. Public Regulated Service. Available online: https://qzss.go.jp/en/overview/services/sv10_public.html (accessed on 11 July 2024).
175. Indian Space Research Organisation (ISRO). Satellite Navigation Services. 2023. Available online: https://www.isro.gov.in/SatelliteNavigationServices.html (accessed on 11 July 2024).
176. Terris-Gallego, R.; Fernandez-Hernandez, I.; López-Salcedo, J.A.; Seco-Granados, G. Guidelines for Galileo Assisted Commercial Authentication Service Implementation. In Proceedings of the 2022 International Conference on Localization and GNSS (ICL-GNSS), Tampere, Finland, 7–9 June 2022; pp. 1–7. [CrossRef]
177. Boyero, J.P. Galileo Programme Status. In Proceedings of the Seventeenth Meeting of the International Committee on Global Navigation Satellite Systems, Madrid, Spain, 16–20 October 2023. Available online: https://www.unoosa.org/documents/pdf/icg/2023/PF-28/pf28.01.pdf (accessed on 11 July 2024).

178. Walker, P.; Rijmen, V.; Fernández-Hernández, I.; Bogaardt, L.; Seco-Granados, G.; Simón, J.; Calle, D.; Pozzobon, O. Galileo Open Service Authentication: A Complete Service Design and Provision Analysis. In Proceedings of the 28th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2015), Tampa, FL, USA, 14–18 September 2015; The Institute of Navigation: Manassas, VA, USA, 2015; pp. 3383–3396.
179. Fernández-Hernández, I.; Rijmen, V.; Seco-Granados, G.; Simon, J.; Rodríguez, I.; Calle, J.D. A Navigation Message Authentication Proposal for the Galileo Open Service. *NAVIGATION* **2016**, *63*, 85–102. [CrossRef]
180. Hernández, I.F.; Wyttenbach, A.F. European Satellite-Based Navigation (GNSS) for Smart Tachographs: Galileo Open Service Authentication. Presented at Tachograph Forum, 28 November 2016. 2016. Available online: https://www.ecologie.gouv.fr/sites/default/files/documents/5%20GSA%20presentation_0.pdf (accessed on 11 July 2024).
181. Simon, J.; Rodriguez, T.; Scorzolini, A.; da Silva, P.; Sbardellati, F.; Fernandez-Hernandez, I.; Damy, S.; Ibanez, D. The Galileo Open Service Navigation Message Authentication (OSNMA): The Pioneer Data Authentication Service. In Proceedings of the 37th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2024), Baltimore, MD, USA, 16–20 September 2024; Institute of Navigation (ION): Manassas, VA, USA, 2024; pp. 973–991. [CrossRef]
182. Fernandez-Hernandez, I.; Blonski, D.; de Blas, J. Galileo HAS and OSNMA. In Proceedings of the Seventeenth Meeting of the International Committee on Global Navigation Satellite Systems (ICG), Madrid, Spain, 15–20 October 2023. Available online: https://www.unoosa.org/documents/pdf/icg/2023/ICG-17/icg17.02.14.pdf (accessed on 12 July 2024).
183. Hernandez, I.F. Galileo Authentication and High Accuracy. Presentation at IGS Workshop 6th Stop: Galileo Constellation Spotlight. 2023. Available online: https://files.igs.org/pub/resource/pubs/workshop/2023/TourdelIGS6_02_FernandezHernandez.pdf (accessed on 11 July 2024).
184. Digdarsini, D.; Mehta, D. Navigation Message Authentication for NavIC System. In Proceedings of the Fourteenth Meeting of the International Committee on Global Navigation Satellite Systems, Bengaluru, India, 8–13 December 2019. Available online: https://www.unoosa.org/documents/pdf/icg/2019/icg14/WGS/icg14_wgs_08.pdf (accessed on 11 July 2024).
185. Patidar, P. Update on Navigation Message Authentication (NMA) for NavIC SPS. In Proceedings of the Seventeenth Meeting of the International Committee on Global Navigation Satellite Systems, Madrid, Spain, 15–20 October 2023. Available online: https://www.unoosa.org/documents/pdf/icg/2023/ICG-17/icg17_wgs_04_03.pdf (accessed on 11 July 2024).
186. Cabinet Office, G.o.J. Draft Specification of User Interface for Signal Authentication Service. 2023. Available online: https://qzss.go.jp/info/information/qzss_230124.html (accessed on 11 July 2024).
187. Cabinet Office, G.O.J. The Test Transmission of Signal Authentication Data. 2024. Available online: https://qzss.go.jp/en/overview/notices/qznma_230707.html, (accessed on 11 July 2024).
188. Cabinet Office, Government of Japan. Start of Signal Authentication Service. 2024. Available online: https://qzss.go.jp/en/overview/notices/qznma_240401.html (accessed on 11 July 2024).
189. Zhilenko, S. GLONASS Status. In Proceedings of the United Nations/Philippines Workshop on the Applications of Global Navigation Satellite Systems, Manila, Philippines, 22–26 April 2024. Available online: https://www.unoosa.org/documents/pdf/psa/activities/2024/GNSS2024/GNSS2024_01_03.pdf (accessed on 11 July 2024).
190. Divis, D.A. New Chimera Signal Enhancement Could Spoof-Proof GPS Receivers. 2019. Available online: https://insidegnss.com/new-chimera-signal-enhancement-could-spoof-proof-gps-receivers/ (accessed on 11 July 2024).
191. L3Harris. L3Harris Prepares Next-Generation NTS-3 Satellite for Launch in Late 2024. 2024. Available online: https://www.australiandefence.com.au/news/sponsored/l3harris-prepares-next-generation-nts-3-satellite-for-launch-in-late-2024 (accessed on 11 July 2024).
192. Zhijun, W.; Yuan, Z.; Yiming, Y.; Peng, W.; Meng, Y. BDSec: Security authentication protocol for BeiDou-II civil navigation message. *China Commun.* **2024**, *21*, 206–218. [CrossRef]
193. Zhijun, W.; Cheng, L.; Yun, Z.; Rusen, L.; Meng, Y. Anti-spoofing: Integrated information authentication of BeiDou-II civil navigation message. *China Commun.* **2024**, *21*, 242–261. [CrossRef]
194. Baker, S. Ukraine's Special Forces Have Developed New Tech That Allows Drones to Fly Without GPS, so Russia Can't Jam Them: Report. 2024. Available online: https://www.businessinsider.com/ukraine-special-forces-drone-tech-no-gps-russia-cant-jam-2024-5 (accessed on 12 July 2024).
195. Mallikarjuna Rao, G.; Mallikarjuna Rao, C.; Reddy, B.R.K.; Lalitha Parameswari, D.V.; Azeez, M. Exploring the Novice Approach to Orthorectification of Satellite Imagery. In *ICT Analysis and Applications*; Fong, S., Dey, N., Joshi, A., Eds.; Springer: Singapore, 2020; pp. 143–154. [CrossRef]
196. Hariyanto, T.; Kurniawan, A.; Pribadi, C.B.; Amin, R.A. Optimization of Ground Control Point (GCP) and Independent Control Point (ICP) on Orthorectification of High Resolution Satellite Imagery. In Proceedings of the E3S Web of Conferences, Semarang, Indonesia, 7–8 August 2019; EDP Sciences: Les Ulis, France, 2019; Volume 94, p. 02008. [CrossRef]
197. Pradhan, B.; Ahmed, A.A.; Chakraborty, S.; Alamri, A.; Lee, C.W. Orthorectification of WorldView-3 Satellite Image Using Airborne Laser Scanning Data. *J. Sens.* **2021**, *2021*, 5273549. [CrossRef]
198. Sulaiman, A.; Elashmawi, W.H.; Eltaweel, G. IHS-based pan-sharpening technique for visual quality improvement using KPCA and enhanced SML in the NSCT domain. *Int. J. Remote Sens.* **2021**, *42*, 537–566. [CrossRef]
199. Mahmoudi, F.; Karami, A. Quantitative Assessment of Transformation Based Satellite Image Pan-sharpening Algorithms. *J. Electr. Comput. Eng. Innov.* **2020**, *8*, 161–168. [CrossRef]

200. Lemeshewsky, G.P. Landsat multispectral sharpening using a sensor system model and panchromatic image. In *Proceedings of the Visual Information Processing XII*; Geradts, Z.J., Rudin, L.I., ur Rahman, Z., Schowengerdt, R.A., Reichenbach, S.E., Eds.; International Society for Optics and Photonics; SPIE: Bellingham, WA, USA, 2003; Volume 5108, pp. 295–306. [CrossRef]
201. Cheng, P. SPOT-6 and SPOT-7 Satellites. *GeoInformatics* **2015**, *18*, 24.
202. Cheng, P. Mapping Large Areas. *GeoInformatics* **2014**, *17*, 18.
203. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
204. Szeliski, R. *Computer Vision: Algorithms and Applications*; Springer: London, UK, 2010.
205. Özaydin, U.; Georgiou, T.; Lew, M. A Comparison of CNN and Classic Features for Image Retrieval. In Proceedings of the 2019 International Workshop on Content-Based Multimedia Indexing (CBMI), Dublin, Ireland, 4–6 September 2019; pp. 1–4. [CrossRef]
206. Xu, Z.; Yu, J.; Yu, C.; Shen, H.; Wang, Y.; Yang, H. CNN-based Feature-point Extraction for Real-time Visual SLAM on Embedded FPGA. In Proceedings of the 2020 IEEE 28th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM), Fayetteville, AR, USA, 3–6 May 2020; pp. 33–37. [CrossRef]
207. Shin, Y.; Balasingham, I. Comparison of hand-craft feature based SVM and CNN based deep learning framework for automatic polyp classification. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Republic of Korea, 11–15 July 2017; pp. 3277–3280. [CrossRef]
208. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30, pp. 5998–6008. [CrossRef]
209. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929. [CrossRef]
210. Ding, Y.; Yuan, W.; Zhu, Q.; Zhang, H.; Liu, X.; Wang, Y.; Liu, X. TransMVSNet: Global Context-aware Multi-view Stereo Network with Transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2022; pp. 8575–8584. [CrossRef]
211. Chen, H.; Luo, Z.; Zhou, L.; Tian, Y.; Zhen, M.; Fang, T.; McKinnon, D.; Tsin, Y.; Quan, L. ASpanFormer: Detector-Free Image Matching with Adaptive Span Transformer. In Proceedings of the Computer Vision–ECCV 2022, Tel Aviv, Israel, 23–27 October 2022; Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T., Eds.; Springer: Cham, Switzerland, 2022; pp. 20–36. [CrossRef]
212. Qin, H.; Zhou, D.; Xu, T.; Bian, Z.; Li, J. Factorization Vision Transformer: Modeling Long-Range Dependency with Local Window Cost. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, 1–14. [CrossRef]
213. Zhu, H.; Chen, B.; Yang, C. Understanding Why ViT Trains Badly on Small Datasets: An Intuitive Perspective. *arXiv* **2023**, arXiv:2302.03751. [CrossRef]
214. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS 2012), Lake Tahoe, NV, USA, 3–6 December 2017; pp. 84–90. [CrossRef]
215. Zheng, Z.; Wei, Y.; Yang, Y. University-1652: A Multi-view Multi-source Benchmark for Drone-based Geo-localization. In Proceedings of the 28th ACM International Conference on Multimedia (MM '20), New York, NY, USA, 12–16 October 2020; pp. 1395–1403. [CrossRef]
216. Zhu, R.; Yin, L.; Yang, M.; Wu, F.; Yang, Y.; Hu, W. SUES-200: A Multi-Height Multi-Scene Cross-View Image Benchmark Across Drone and Satellite. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 4825–4839. [CrossRef]
217. Xiao, J.; Tortei, D.; Roura, E.; Loianno, G. Long-Range UAV Thermal Geo-Localization with Satellite Imagery. In Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Detroit, MI, USA, 1–5 October 2023; pp. 5820–5827. [CrossRef]
218. Xiao, J.; Zhang, N.; Tortei, D.; Loianno, G. STHN: Deep Homography Estimation for UAV Thermal Geo-localization with Satellite Imagery. *arXiv* **2024**, arXiv:2405.20470. [CrossRef]
219. Luo, X.; Wan, X.; Gao, Y.; Tian, Y.; Zhang, W.; Shu, L. JointLoc: A Real-time Visual Localization Framework for Planetary UAVs Based on Joint Relative and Absolute Pose Estimation. *arXiv* **2024**, arXiv:2405.07429 [CrossRef]
220. AgEagle Aerial Systems Inc. AgEagle Resources. 2024. Available online: https://ageagle.com/resources/?filter_by=data-set (accessed on 12 July 2024).
221. layumi. University1652-Baseline. 2021. Available online: https://github.com/layumi/University1652-Baseline (accessed on 15 July 2024).
222. Mughal, M.H.; Khokhar, M.J.; Shahzad, M. Assisting UAV Localization Via Deep Contextual Image Matching. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 2445–2457. [CrossRef]
223. m-hamza mughal. Aerial Template Matching Dataset. 2020. Available online: https://github.com/m-hamza-mughal/aerial-template-matching-dataset (accessed on 15 July 2024).
224. Zhu, R. SUES-200: A Multi-Height Multi-Scene Cross-View Image Matching Benchmark Across UAV and Satellite. 2024. Available online: https://github.com/Reza-Zhu/SUES-200-Benchmark (accessed on 15 July 2024).

225. Cisneros, I.; Yin, P.; Zhang, J.; Choset, H.; Scherer, S. ALTO: A Large-Scale Dataset for UAV Visual Place Recognition and Localization. *arXiv* **2022**, arXiv:2207.12317. [CrossRef]
226. MetaSLAM. ALTO (Aerial-View Large-Scale Terrain-Oriented) Dataset. 2022. Available online: https://github.com/MetaSLAM/ALTO (accessed on 15 July 2024).
227. Zhan, M.; Morley, T.P. MSDI: A Geo-Tagged Drone Imagery for Absolute Visual Localization. 2022. Available online: https://zenodo.org/records/6977602 (accessed on 12 July 2024).
228. Gurgu, M.M.; Queralta, J.P.n.; Westerlund, T. Vision-Based GNSS-Free Localization for UAVs in the Wild. In Proceedings of the 2022 7th International Conference on Mechanical Engineering and Robotics Research (ICMERR), Krakow, Poland, 9–11 December 2022; pp. 7–12. [CrossRef]
229. TIERS. WildNav: GNSS-Free Localization in the Wild. 2023. Available online: https://github.com/TIERS/wildnav (accessed on 15 July 2024).
230. Arplaboratory. Satellite-Thermal-Geo-Localization. 2024. Available online: https://github.com/arplaboratory/satellite-thermal-geo-localization (accessed on 15 July 2024).
231. Dai, M.; Zheng, E.; Feng, Z.; Qi, L.; Zhuang, J.; Yang, W. Vision-Based UAV Self-Positioning in Low-Altitude Urban Environments. *IEEE Trans. Image Process.* **2024**, *33*, 493–508. [CrossRef]
232. Hashemi, P. Vision-Based UAV Self-Positioning in Low-Altitude Urban Environments. 2024. Available online: https://github.com/peymanHashemi/UAV_localization (accessed on 15 July 2024).
233. Arplaboratory. STHN: Deep Homography Estimation for UAV Thermal Geo-Localization with Satellite Imagery. 2024. Available online: https://github.com/arplaboratory/STHN (accessed on 15 July 2024).
234. Luo, X. JointLoc. 2024. Available online: https://github.com/LuoXubo/JointLoc (accessed on 15 July 2024).
235. Xu, W.; Yao, Y.; Cao, J.; Wei, Z.; Liu, C.; Wang, J.; Peng, M. UAV-VisLoc: A Large-scale Dataset for UAV Visual Localization. *arXiv* **2024**, arXiv:2405.11936. [CrossRef]
236. IntelliSensing. UAV-VisLoc: A Large-Scale Dataset for UAV Visual Localization. 2024. Available online: https://github.com/IntelliSensing/UAV-VisLoc (accessed on 15 July 2024).
237. He, M.; Liu, J.; Gu, P.; Meng, Z. Leveraging Map Retrieval and Alignment for Robust UAV Visual Geo-Localization. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–13. [CrossRef]
238. AgEagle. eBee X. 2024. Available online: https://ageagle.com/drones/ebee-x/ (accessed on 20 October 2024).
239. AgEagle. S.O.D.A. 2024. Available online: https://ageagle.com/drone-sensors/soda/ (accessed on 20 October 2024).
240. AgEagle Aerial Systems Inc. Hotel Modelling Dataset. 2022. Available online: https://ageagle.com/data-set/hotel-modelling/. (accessed on 20 September 2024).
241. Wingtra. WingtraOne. 2024. Available online: https://wingtra.com/mapping-drone-wingtraone/ (accessed on 20 October 2024).
242. AgEagle. RedEdge-P. 2024. Available online: https://ageagle.com/drone-sensors/rededge-p-high-res-multispectral-camera/ (accessed on 20 October 2024).
243. AgEagle Aerial Systems Inc. Palm Tree Detection Using RedEdge-P Dataset. 2022. Available online: https://ageagle.com/data-set/palm-tree-detection-using-rededge-p/. (accessed on 20 September 2024).
244. FotoAerea. Asesor/9. 2024. Available online: https://fotoaerea.com.ar/drone-para-mapeo-asesor-9/ (accessed on 20 October 2024).
245. AgEagle Aerial Systems Inc. Growing Variations Soybean Field Dataset. 2022. Available online: https://ageagle.com/data-set/growing-variations-soybean-field/ (accessed on 20 September 2024).
246. AgEagle. eBee Ag. 2024. Available online: https://ageagle.com/wp-content/uploads/2022/06/AgEagle-eBee-AG-EN-2022-2.pdf (accessed on 20 October 2024).
247. AgEagle. Duet M. 2024. Available online: https://ageagle.com/drone-sensors/duet-m/ (accessed on 20 October 2024).
248. AgEagle Aerial Systems Inc. Corn Field Mapped with eBee Ag in Brazil Dataset. 2021. Available online: https://ageagle.com/data-set/corn-field-mapped-with-ebee-ag-in-brazil/ (accessed on 20 September 2024).
249. AgEagle Aerial Systems Inc. Mixed Use Agricultural Fields eBee Ag Dataset. 2021. Available online: https://ageagle.com/data-set/mixed-use-agricultural-fields-ebee-ag/ (accessed on 20 September 2024).
250. AgEagle. Duet T. 2024. Available online: https://ageagle.com/drone-sensors/duet-t/ (accessed on 20 October 2024).
251. AgEagle Aerial Systems Inc. Industrial Estate Dataset. 2019. Available online: https://ageagle.com/data-set/industrial-estate/ (accessed on 20 September 2024).
252. AgEagle. Aeria X. 2024. Available online: https://ageagle.com/drone-sensors/aeria-x/ (accessed on 20 October 2024).
253. AgEagle Aerial Systems Inc. Gravel Quarry Dataset. 2018. Available online: https://ageagle.com/data-set/gravel-quarry/ (accessed on 20 September 2024).
254. AgEagle Aerial Systems Inc. University Campus Dataset. Available online: https://ageagle.com/data-set/university-campus/ (accessed on 20 September 2024).
255. AgEagle Aerial Systems Inc. Dam Inspection Dataset. Available online: https://ageagle.com/data-set/dam-inspection/ (accessed on 20 September 2024).
256. Google LLC. Google Earth. 2023. Available online: https://www.google.com/earth/ (accessed on 11 July 2024).
257. Google LLC. Google Maps. 2023. Available online: https://www.google.com/maps (accessed on 11 July 2024).

258. Google Images. Google Image Search. Available online: https://images.google.com (accessed on 20 September 2024).
259. DJI. DJI Phantom 4 Pro V2.0. 2024. Available online: https://www.dji.com/ca/phantom-4-pro-v2 (accessed on 20 October 2024).
260. AirLab-Carnegie Mellon University. [ICRA2022] General Place Recognition: Visual Terrain Relative Navigation. 2022. Available online: https://www.aicrowd.com/challenges/icra2022-general-place-recognition-visual-terrain-relative-navigation (accessed on 11 July 2024).
261. IDS Imaging Development Systems Inc. UI-5250RE-C-HQ PoE Rev.2.2 (AB10302). 2023. Available online: https://en.ids-imaging.com/IDS/datasheet_pdf.php?sku=AB10302 (accessed on 11 July 2024).
262. Kowa American Corp. Datasheet for Kowa LM3NC1M 3.5mm Lens. 2023. Available online: https://lenses.kowa-usa.com/wide-angle-megapixel-ncm-series/408-lm3nc1m.html (accessed on 11 July 2024).
263. NovAtel Inc. SPAN on OEMV-3 Receiver. 2023. Available online: https://novatel.com/support/previous-generation-products-drop-down/previous-generation-products/span-on-oemv-3-receiver (accessed on 11 July 2024).
264. Northrop Grumman LITEF GmbH. LITEF LCI-1 Inertial Measurement Unit. 2023. Available online: https://www.litef.de/en/industrial-solutions/products/mems-imu (accessed on 11 July 2024).
265. Kempf GmbH & Co. KG Laser Sensor LMC-J-0310-X. 2024. Available online: https://loke.de/index.php/en/laser-sensors/distanzmessung-ir/laser-sensor-lmc-j-0310-en (accessed on 20 October 2024).
266. Aerial Photography Field Office, U.S. Department of Agriculture, Farm Service Agency. National Agricultural Imagery Program. 2019. Available online: http://earthexplorer.usgs.gov (accessed on 11 July 2024).
267. Parrot Drone SAS. Parrot ANAFI-Compact and Resistant Drone with a 4K HDR Camera. 2023. Available online: https://www.parrot.com/us/drones/anafi (accessed on 11 July 2024).
268. Microsoft Corporation. Bing Maps. 2023. Available online: https://www.bing.com/maps (accessed on 11 July 2024).
269. Zhan, M.; Morley, T.P. MSDI_TOOL: MSDI Utility Tools. 2022. [CrossRef]
270. DJI. DJI Matrice 300. 2024. Available online: https://www.dji.com/ca/support/product/matrice-300 (accessed on 20 October 2024).
271. DJI. DJI Zenmuse H20 Series. 2024. Available online: https://enterprise.dji.com/zenmuse-h20-series (accessed on 20 October 2024).
272. FLIR, T. Boson-Uncooled Longwave Infrared (LWIR) OEM Thermal Camera Module. 2024. Available online: https://www.flir.ca/products/boson/ (accessed on 18 September 2024).
273. DJI. DJI Pilot App. 2024. Available online: https://www.dji.com/ca/downloads/djiapp/dji-pilot (accessed on 20 October 2024).
274. Epic Games. Unreal Engine 4. 2024. Available online: https://www.unrealengine.com/ (accessed on 19 September 2024).
275. Shah, S.; Dey, D.; Lovett, C.; Kapoor, A. AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles. In *Field and Service Robotics*; Hutter, M., Siegwart, R., Eds.; Springer: Cham, Switzerland, 2018; pp. 621–635. [CrossRef]
276. Sui, H.; Li, J.; Lei, J.; Liu, C.; Gou, G. A Fast and Robust Heterologous Image Matching Method for Visual Geo-Localization of Low-Altitude UAVs. *Remote Sens.* **2022**, *14*, 5879. [CrossRef]
277. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
278. Sarlin, P.E.; DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperGlue: Learning Feature Matching with Graph Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, 14–19 June 2020; pp. 4937–4946. [CrossRef]
279. Sun, J.; Shen, Z.; Wang, Y.; Bao, H.; Zhou, X. LoFTR: Detector-Free Local Feature Matching with Transformers. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 8918–8927. [CrossRef]
280. Wang, Q.; Zhang, J.; Yang, K.; Peng, K.; Stiefelhagen, R. MatchFormer: Interleaving Attention in Transformers for Feature Matching. In *Proceedings of the Computer Vision–ACCV 2022*; Wang, L., Gall, J., Chin, T.J., Sato, I., Chellappa, R., Eds.; Springer: Cham, Switzerland, 2023; pp. 256–273. [CrossRef]
281. Nibali, A.; He, Z.; Morgan, S.; Prendergast, L. Numerical Coordinate Regression with Convolutional Neural Networks. *arXiv* **2018**, arXiv:1801.07372. [CrossRef]
282. Li, Z.; Snavely, N. MegaDepth: Learning Single-View Depth Prediction from Internet Photos. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, 18–22 June 2018; pp. 2041–2050. [CrossRef]
283. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [CrossRef] [PubMed]
284. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]
285. Balntas, V.; Lenc, K.; Vedaldi, A.; Tuytelaars, T.; Matas, J.; Mikolajczyk, K. H-Patches: A Benchmark and Evaluation of Handcrafted and Learned Local Descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2825–2841. [CrossRef] [PubMed]
286. Zhou, Q.; Sattler, T.; Leal-Taixe, L. Patch2Pix: Epipolar-Guided Pixel-Level Correspondences. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, 19–25 June 2021; pp. 4667–4676. [CrossRef]

287. Lowe, D. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 2, pp. 1150–1157. [CrossRef]
288. DJI. Phantom 4 RTK. 2023. Available online: https://enterprise.dji.com/phantom-4-rtk (accessed on 11 July 2024).
289. Rocco, I.; Cimpoi, M.; Arandjelović, R.; Torii, A.; Pajdla, T.; Sivic, J. Neighbourhood Consensus Networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R., Eds.; Curran Associates, Inc.: New York, NY, USA, 2018; Volume 31.
290. Bian, J.; Lin, W.Y.; Matsushita, Y.; Yeung, S.K.; Nguyen, T.D.; Cheng, M.M. GMS: Grid-Based Motion Statistics for Fast, Ultra-Robust Feature Correspondence. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2828–2837. [CrossRef]
291. Schaffalitzky, F.; Zisserman, A. Automated Scene Matching in Movies. In Proceedings of the Image and Video Retrieval, London, UK, 18–19 July 2002; Lew, M.S., Sebe, N., Eakins, J.P., Eds.; Springer: Berlin/Heidelberg, Germany, 2002; pp. 186–197. [CrossRef]
292. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV'15), Santiago, Chile, 7–13 December 2015; pp. 1026–1034. [CrossRef]
293. Bottou, L. Online Algorithms and Stochastic Approximations. In *Online Learning and Neural Networks*; Saad, D., Ed.; Cambridge University Press: Cambridge, UK, 1998.
294. Balntas, V.; Riba, E.; Ponsa, D.; Mikolajczyk, K. Learning local feature descriptors with triplets and shallow convolutional neural networks. In Proceedings of the British Machine Vision Conference (BMVC), York, UK, 19–22 September 2016; Richard, C., Wilson, E.R.H., Smith, W.A.P., Eds.; BMVA Press: Durham, UK, 2016; pp. 119.1–119.11. [CrossRef]
295. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv* **2022**, arXiv:1312.6114. [CrossRef]
296. Corporation, M. Bing Maps API Documentation. 2024. Available online: https://learn.microsoft.com/en-us/bingmaps/ (accessed on 12 July 2024).
297. SAS, P.D. Parrot Bebop 2 User Guide. 2021. Available online: https://www.parrot.com/assets/s3fs-public/2021-09/bebop-2_user-guide_uk_2.pdf (accessed on 24 September 2024).
298. Coffee & Data. Lossless Triplet Loss. Available online: https://coffeeanddata.ca/lossless-triplet-loss (accessed on 24 September 2024).
299. Koch, G.; Zemel, R.; Salakhutdinov, R. Siamese neural networks for one-shot image recognition. In Proceedings of the ICML Deep Learning Workshop, Lille, France, 6–11 July 2015; pp. 1–30.
300. Liu, G.; Liu, C.; Yuan, Y. Locate Where You Are by Block Joint Learning Network. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
301. Snoek, C.G.M.; Worring, M.; Smeulders, A.W.M. Early versus late fusion in semantic video analysis. In Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA'05), New York, NY, USA, 6–11 November 2005; pp. 399–402. [CrossRef]
302. Wang, R.; Chen, J.W.; Wang, Y.; Jiao, L.; Wang, M. SAR Image Change Detection via Spatial Metric Learning With an Improved Mahalanobis Distance. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 77–81. [CrossRef]
303. Sobel, I.; Feldman, G. *A 3x3 Isotropic Gradient Operator for Image Processing*; Technical Report; Stanford Artificial Intelligence Laboratory (SAIL): Stanford, CA, USA, 1968.
304. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the Computer Vision–ECCV 2018, Munich, Germany, 8–14 September 2018; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer: Cham, Switzerland, 2018; pp. 3–19. [CrossRef]
305. OPTIMAL-HSIG. LA500 Dataset. 2021. Available online: https://github.com/OPTIMAL-HSIG/LA500 (accessed on 11 July 2024).
306. Han, X.; Leung, T.; Jia, Y.; Sukthankar, R.; Berg, A.C. MatchNet: Unifying feature and metric learning for patch-based matching. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3279–3286. [CrossRef]
307. Tian, Y.; Yu, X.; Fan, B.; Wu, F.; Heijnen, H.; Balntas, V. SOSNet: Second Order Similarity Regularization for Local Descriptor Learning. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, 15–20 June 2019; pp. 11008–11017. [CrossRef]
308. Ng, T.; Balntas, V.; Tian, Y.; Mikolajczyk, K. SOLAR: Second-Order Loss and Attention for Image Retrieval. In Proceedings of the Computer Vision–ECCV 2020, Glasgow, UK, 23–28 August 2020; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., Eds.; Springer: Cham, Switzerland, 2020; pp. 253–270. [CrossRef]
309. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, 21–26 July 2017; pp. 5967–5976. [CrossRef]
310. Berton, G.; Mereu, R.; Trivigno, G.; Masone, C.; Csurka, G.; Sattler, T.; Caputo, B. Deep Visual Geo-localization Benchmark. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 21–26 July 2022; pp. 5386–5397. [CrossRef]

311. Ganin, Y.; Lempitsky, V. Unsupervised Domain Adaptation by Backpropagation. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015; Bach, F., Blei, D., Eds.; ACM: New York, NY, USA, 2015; Volume 37; Proceedings of Machine Learning Research; pp. 1180–1189.
312. Arandjelović, R.; Gronat, P.; Torii, A.; Pajdla, T.; Sivic, J. NetVLAD: CNN Architecture for Weakly Supervised Place Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1437–1451. [CrossRef]
313. Bianchi, M.; Barfoot, T.D. UAV Localization Using Autoencoded Satellite Images. *IEEE Robot. Autom. Lett.* **2021**, *6*, 1761–1768. [CrossRef]
314. Felix, B.H.P. Variational-Autoencoder-PyTorch. 2024. Available online: https://github.com/bhpfelix/Variational-Autoencoder-PyTorch (accessed on 26 September 2024).
315. Shen, T.; Luo, Z.; Zhou, L.; Deng, H.; Zhang, R.; Fang, T.; Quan, L. Beyond Photometric Loss for Self-Supervised Ego-Motion Estimation. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 6359–6365. [CrossRef]
316. Patel, B.; Barfoot, T.D.; Schoellig, A.P. Visual Localization with Google Earth Images for Robust Global Pose Estimation of UAVs. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 21 May–31 August 2020; pp. 6491–6497. [CrossRef]
317. DJI. DJI Matrice 600 Pro. 2024. Available online: https://www.dji.com/ca/support/product/matrice600-pro (accessed on 20 October 2024).
318. DJI. DJI Ronin-MX. 2024. Available online: https://www.dji.com/ca/ronin-mx (accessed on 20 October 2024).
319. Stereolabs. ZED 2. 2024. Available online: https://www.stereolabs.com/en-ca/products/zed-2 (accessed on 20 October 2024).
320. Li, H.; Wang, J.; Wei, Z.; Xu, W. Jointly Optimized Global-Local Visual Localization of UAVs. *arXiv* **2023**, arXiv:2310.08082. [CrossRef]
321. Radenović, F.; Tolias, G.; Chum, O. Fine-Tuning CNN Image Retrieval with No Human Annotation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 1655–1668. [CrossRef]
322. DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperPoint: Self-Supervised Interest Point Detection and Description. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018. [CrossRef]
323. Fix, E.; Hodges, J.L. Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. *Int. Stat. Rev. Rev. Int. Stat.* **1989**, *57*, 238–247. [CrossRef]
324. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [CrossRef]
325. Li, Y.; Zhu, Z.; Yu, J.G.; Zhang, Y. Learning Deep Cross-Modal Embedding Networks for Zero-Shot Remote Sensing Image Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10590–10603. [CrossRef]
326. Li, H. Jointly Optimized Global-Local Visual Localization of UAVs. 2023. Available online: https://github.com/lihaoling/GLVL (accessed on 15 July 2024).
327. Van Quan, N.; Luyen, D.D.; Anh, P.H.; Thu, P.T.H.; Thanh, N.C.; Thai, V.D. Speeding Up SIFT for Faster Feature-Based UAV Visual Localization. In *Advances in Information and Communication Technology*; Nghia, P.T., Thai, V.D., Thuy, N.T., Son, L.H., Huynh, V.N., Eds.; Springer: Cham, Switzerland, 2024; pp. 83–90. [CrossRef]
328. Rostum, H.M.; Vásárhelyi, J. A review of using visual odometery methods in autonomous UAV Navigation in GPS-Denied Environment. *Acta Univ. Sapientiae Electr. Mech. Eng.* **2023**, *15*, 14–32. [CrossRef]
329. Neyestani, A.; Picariello, F.; Basiri, A.; Daponte, P.; Vito, L.D. Survey and Research Challenges in Monocular Visual Odometry. In Proceedings of the 2023 IEEE International Workshop on Metrology for Living Environment (MetroLivEnv), Milano, Italy, 29–31 May 2023; pp. 107–112. [CrossRef]
330. NVIDIA. NVIDIA Jetson Modules. 2024. Available online: https://developer.nvidia.com/embedded/jetson-modules (accessed on 15 July 2024).
331. 96Boards. HiKey970. 2024. Available online: https://www.96boards.org/product/hikey970/ (accessed on 15 July 2024).
332. BeagleBoard. BeagleBone AI. 2024. Available online: https://www.beagleboard.org/boards/beaglebone-ai (accessed on 15 July 2024).
333. Radxa. Rock Pi N10. 2023. Available online: https://wiki.radxa.com/RockpiN10 (accessed on 15 July 2024).
334. TerboucheHacene. Visual Localization for UAVs Using Satellite Imagery. 2024. Available online: https://github.com/TerboucheHacene/visual_localization (accessed on 15 July 2024).
335. Duan, R.; Chen, L.; Li, Z.; Chen, Z.; Wu, B. A Scene Graph Encoding and Matching Network for UAV Visual Localization. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 9890–9902. [CrossRef]
336. rduan036. Scene Graph Matching Demo. 2023. Available online: https://github.com/rduan036/scene-graph-matching-demo (accessed on 15 July 2024).
337. spagnoloG. UAV Localization. 2023. Available online: https://github.com/spagnoloG/uav-localization (accessed on 15 July 2024).
338. vvabi sabi. AutoEncoder for UAV Localization. 2023. Available online: https://github.com/vvabi-sabi/AE_for_UAV_localization (accessed on 15 July 2024).
339. hmf21. Leveraging Map Retrieval and Alignment for Robust UAV Visual Geo-Localization. 2024. Available online: https://github.com/hmf21/UAVLocalization (accessed on 15 July 2024).

340. Chen, Q.; Wang, T.; Yang, Z.; Li, H.; Lu, R.; Sun, Y.; Zheng, B.; Yan, C. SDPL: Shifting-Dense Partition Learning for UAV-view Geo-localization. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, 1. [CrossRef]
341. solardust8. SDPL: Shifting-Dense Partition Learning for UAV-View Geo-Localization. 2024. Available online: https://github.com/solardust8/SDPL_model (accessed on 15 July 2024).
342. C-water. [TCSVT 2024] SDPL: Shifting-Dense Partition Learning for UAV-View Geo-Localization. 2024. Available online: https://github.com/C-water/SDPL_release (accessed on 15 July 2024).
343. Zhu, R.; Yang, M.; Yin, L.; Wu, F.; Yang, Y. UAV's Status Is Worth Considering: A Fusion Representations Matching Method for Geo-Localization. *Sensors* **2023**, *23*, 720. [CrossRef]
344. Zhu, R. UAV's Status is Worth Considering: A Fusion Representations Matching Method for Geo-Localization. 2024. Available online: https://github.com/Reza-Zhu/MBF (accessed on 15 July 2024).
345. Milham, M.; Craddock, C.; Fleischmann, M.; Son, J.J.; Clucas, J.; Xu, H.Y.; Koo, B.; Krishnakumar, A.; Biswal, B.B.; Castellanos, F.X.; et al. Assessment of the impact of shared data on the scientific literature. *bioRxiv* **2017**. [CrossRef]
346. Borgman, C.L. Big Science, Little Science, and Open Science: Sustainability, Stewardship, and Knowledge Infrastructures. In *Proceedings of the Conference on the National Open Science Plan for France: From Strategy to Action 2018*; University of California Press: Berkeley, CA, USA, 2018.
347. Campbell, J. Access to Scientific Data in the 21st Century: Rationale and Illustrative Usage Rights Review. *Data Sci. J.* **2015**, *13*, 203–230. [CrossRef]
348. Levenstein, M.C.; Lyle, J. Data: Sharing Is Caring. *Adv. Methods Pract. Psychol. Sci.* **2018**, *1*, 95–103. [CrossRef]
349. Ait Nasser, A.; Akhloufi, M.A. A Review of Recent Advances in Deep Learning Models for Chest Disease Detection Using Radiography. *Diagnostics* **2023**, *13*, 159. [CrossRef] [PubMed]
350. Gayap, H.T.; Akhloufi, M.A. Deep Machine Learning for Medical Diagnosis, Application to Lung Cancer Detection: A Review. *BioMedInformatics* **2024**, *4*, 236–284. [CrossRef]
351. Guergueb, T.; Akhloufi, M.A. A Review of Deep Learning Techniques for Glaucoma Detection. *SN Comput. Sci.* **2023**, *4*, 274. [CrossRef]
352. Wu, Y. pycocotools. 2024. Available online: https://github.com/ppwwyyxx/cocoapi (accessed on 15 July 2024).
353. Heindl, C.; Toka; Valmadre, J. py-motmetrics: Benchmark Multiple Object Trackers (MOT) in Python. 2022. Available online: https://github.com/cheind/py-motmetrics (accessed on 20 October 2024).


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.