




Article

# Robust Autonomous Vehicle Computer-Vision-Based Localization in Challenging Environmental Conditions

Sergei Chuprov <sup>1</sup>, Pavel Belyaev <sup>2</sup>, Ruslan Gataullin <sup>2</sup>, Leon Reznik <sup>1,\*</sup>, Evgenii Neverov <sup>2</sup>  
and Ilia Viksnin <sup>2</sup>

<sup>1</sup> Department of Computer Science and Global Cybersecurity Institute, Rochester Institute of Technology, Rochester, NY 14623, USA; sc1723@rit.edu

<sup>2</sup> Department of Computer Science, Saint Petersburg Electrotechnical University “LETI”, St. Petersburg 197022, Russia; piubeliaev@etu.ru (P.B.); rusfiner@mail.ru (R.G.); wixnin@mail.ru (I.V.)

\* Correspondence: leon.reznik@rit.edu

**Abstract:** In this paper, we present a novel autonomous vehicle (AV) localization design and its implementation, which we recommend to employ in challenging navigation conditions with a poor quality of the satellite navigation system signals and computer vision images. In the case when the GPS signal becomes unstable, other auxiliary navigation systems, such as computer-vision-based positioning, are employed for more accurate localization and mapping. However, the quality of data obtained from AV’s sensors might be deteriorated by the extreme environmental conditions too, which infinitely leads to the decrease in navigation performance. To verify our computer-vision-based localization system design, we considered the Arctic region use case, which poses additional challenges for the AV’s navigation and might employ artificial visual landmarks for improving the localization quality, which we used for the computer vision training. We further enhanced our data by applying affine transformations to increase its diversity. We selected YOLOv4 image detection architecture for our system design, as it demonstrated the highest performance in our experiments. For the computational platform, we employed a Nvidia Jetson AGX Xavier device, as it is well known and widely used in robotic and AV computer vision, as well as deep learning applications. Our empirical study showed that the proposed computer vision system that was further trained on the dataset enhanced by affine transformations became robust regarding image quality degradation caused by extreme environmental conditions. It was effectively able to detect and recognize images of artificial visual landmarks captured in the extreme Arctic region’s conditions. The developed system can be integrated into vehicle navigation facilities to improve their effectiveness and efficiency and to prevent possible navigation performance deterioration.

**Keywords:** computer vision; robust vehicle localization; data quality; machine learning



**Citation:** Chuprov, S.; Belyaev, P.; Gataullin, R.; Reznik, L.; Neverov, E.; Viksnin, I. Robust Autonomous Vehicle Computer-Vision-Based Localization in Challenging Environmental Conditions. *Appl. Sci.* **2023**, *13*, 5735. <https://doi.org/10.3390/app13095735>

Academic Editors: Nadjim Horri, William Holderbaum and Fabrizio Giulietti

Received: 30 March 2023

Revised: 25 April 2023

Accepted: 28 April 2023

Published: 6 May 2023



**Copyright:** © 2023 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

Navigation is one of the most important autonomous vehicle (AV) functions that performs positioning in the environment and decides on further mobility actions. The failure of accurate localization decreases AV effectiveness and efficiency in other operations, and this might result in safety accidents. Contemporary AVs primarily rely on global navigation satellite systems (GNSSs) to perform navigation and localization, such as GPS or GLONASS, which provide a global coverage and a high accuracy position. For example, contemporary GPS-based methods, which are the most commonly used GNSS systems, are capable of achieving a subcentimetric accuracy (e.g., RTK DGPS [1]). This level of accuracy, which could be further complemented by the measurements from other AV’s on-board sensors, is sufficient for AVs to safely and efficiently navigate their surroundings. However, in some territories, e.g., dense urban or low populated areas with frequent extreme environmental conditions such as the Arctic region, the GPS signal can become unstable or even absent,

which exacerbates navigational operations. In addition to communication obstructions and environmental conditions, other factors can lead to GPS failures, such as electronic interference that may be caused by a variety of sources, e.g., other electronic devices or power lines; signal jamming that can be intentionally activated by various parties or malicious attackers; and software and hardware malfunctions in GPS-related components that can cause failures. To mitigate these factors, multi-sensor perception may be employed for localization and mapping improvement. The data from multi-modality sensors, such as an on-board camera and an inertial measurement units (IMUs), might be fused in order to determine the AV's position in relation to the surrounding objects. However, these sensor devices might also fail or be maliciously attacked, which leads to data quality (DQ) degradation and jeopardizes the AV's security and safety [2,3]. Obviously, the localization accuracy in this case depends on the quality of data acquired from the sensors. The data from multiple sensors might be fused to enhance its accuracy and quality; however, the sensors need to be properly selected in this case [4]. This operation might employ various sensor selection techniques, such as genetic-algorithms-based ones [5].

In this paper, we consider computer-vision-based AV localization as a complimentary system to perform navigational operations. Since computer vision utilizes intelligent machine learning (ML) techniques, its operation involves initial procedures of the ML model training and validating. In most cases, such procedures commonly employ only high quality training data [6], which allows ML-based systems to achieve reliable performance on high quality testing images. However, our previous investigations demonstrated that using real life images with the degraded quality images dramatically affects the performance of ML-based computer vision classification [7,8]. To prevent ML performance degradation, multiple techniques might be leveraged, including the increase of training set diversity and the re-training of ML-based systems on samples of varying DQ [9].

The AV's navigation operation largely depends on how the sensor system is designed. One of the well-known sensors used for AV navigation is LiDAR. For example, Krishnamoorthi et al. [10] considered a navigation and localization system consisting of LiDARs and on-board cameras. The system performed path segmentation from point cloud data, which could be further used in computer vision and AV localization. AVs, as well as unmanned aerial vehicles, can rely on their own sensors to analyze the environment. The navigation and localization system allows AVs to make decisions about further actions without the operator. Arash et al. [11] discussed ways to improve the navigation system for the Internet of Drones using ML and deep learning algorithms. Their survey [12] refers to 37 articles devoted to assuring the system's reliability and fault tolerance. Data from sensors can be processed by computer vision algorithms to solve navigation and monitoring problems, as presented in [13]. The practical applications of autonomous robots may be diverse. Such robots are especially useful when the manual operator's control is associated with significant risks to their health; while our research studies the employment of AVs in Arctic region extreme conditions, other authors, for example, consider AV operation in underground mines [14] and in radioactive conditions [15].

The problem of designing a reliable navigation system is related to the task of ensuring the safety of the vehicle. Amiri et al. [12] deliberated on AV navigation security issues related to the system's reliability and fault tolerance. The task of controlling an autonomous vehicle was considered by Liu et al. [16]. The technology was focused on the evolution from vehicle state estimation and trajectory tracking control in AVs at the microscopic level to collaborative control in connected and automated vehicles at the macroscopic level. The task of developing a navigation system for electric vehicles is more specific. Various scenarios for driving electric vehicles were considered by Chen et al. [17]. The results presented in our work also contribute to improving the safety of using AVs in various applications, e.g., autonomous transport navigation in extreme weather conditions.

The issue of sensor fusion is well covered in the literature. For example, Xia et al. [18] discussed AV kinematic and dynamic synthesis for sideslip angle estimation based on a

consensus Kalman filter. Gao et al. [19] studied slip angle estimation by considering a signal measurement characteristic.

In our work, we investigated methods of improving navigation performance in extreme environmental and poor GPS signal communication conditions. We designed and developed a complimentary AV computer-vision-based localization system to make it robust in harsh environmental conditions that cause the degradation of images obtained from the on-board camera. This computer vision recognition system should integrate with the conventional GPS or even replace it when it becomes unavailable. In particular, we investigated and verified our design on an Arctic region use case, as, nowadays, many authorities seek to automate their transportation operations in this region, and AVs are deemed to become a prominent solution in this field [20]. An example of the resource excavation facilities cluster in the Arctic region is demonstrated in Figure 1. Highly limited natural visual landmarks and changing weather conditions in this region make navigation a real challenge for the AVs. In our particular use case, we considered specialized artificial visual tags allocated over the AV operation area. These tags contain GPS coordinates encoded via AprilTag [21], which is needed to be detected and properly recognized by AVs in order to perform localization in case of GPS unavailability. To implement our design, we employed Nvidia Jetson AGX Xavier as the computing device for the developed ML-based system verification. To enhance our ML-based model robustness toward DQ degradation, we generated images of the artificial visual tags affected by the various lightning and harsh weather conditions, such as blizzards and heavy rain. In addition, we employed several types of affine transformations to these images to diversify our training set. We utilized the produced set of images to re-train our ML-based image classification model. Then, we evaluated the performance of the developed ML-based model to detect and recognize the artificial visual tags from the varied distances.



**Figure 1.** An example of the fossil fuels excavation facilities cluster located in the Arctic region. As one can see, the landscape is covered with snow and ice and might be characterized as mostly flat and uniform, with very few distinctive features. The extreme weather conditions and low levels of precipitation in the region result in a lack of any vegetation forms that could serve as visual markers. The absence of large geological formations such as mountains or canyons also contributes to the lack of visual landmarks.

This paper's major contribution is our computer-vision-based localization system design that might be employed by AVs in order to improve their localization and mapping accuracy in case of GPS signal unavailability. The advantage of the proposed system is its robustness regarding GPS signal instability and image quality degradation caused by extreme environmental conditions. Our verification use case demonstrated that the

developed solution is able to effectively detect and recognize images of artificial visual landmarks in the extreme conditions of Arctic region.

## 2. Background

In our study, we considered the Arctic region as a great use case that poses both challenges of input DQ degradation and GPS signal instability. This region experiences a long winter, lasting up to 10 months, which creates severe problems for AV operation. The air temperature can drop to  $-60$  degrees Celsius, and the strong cold winds can reach up to 30 m/s. The increased frequency of snowfall and blizzards, as well as the small possibility of manual search and rescue in the case of emergency situations, exacerbate the challenges. Such climatic conditions significantly affect AV technical component operation. The excess resource consumption, fast battery discharge, and failures of electronic devices that are not adapted for low temperatures are some of the issues faced. The remote and inaccessible nature can limit the GPS signal's availability, thus making navigation challenging for AVs. The extreme weather conditions, low visibility, and harsh terrain can further compound the challenge of GPS signal loss [22], thus potentially leaving AVs stranded and unable to navigate safely.

To address the challenges of AV navigation in such arduous areas as the Arctic region, AVs may use alternative navigation and localization systems in addition to GPS, e.g., IMU that relies on inertial sensors to track changes in position and movement direction or a computer-vision-based system, which employs on-board cameras to identify and track the position based on visual landmarks. These systems can complement GPS-based localization, thereby providing redundancy and ensuring that AVs can continue to navigate safely and accurately, even in the event of GPS signal loss. Some developments have been introduced to realize the AV's computer-vision-based localization system. Marinho et al. [23] presented an approach to improve the mobile robot navigation via extracting the information from the topological map. The approach was designed mainly for outdoor areas; however, it requires a sufficient number of visual reference points to navigate the robot effectively.

Data-driven ML methods are also widely used to approach AV navigation [24,25]. Here, learning approaches relying on expert reference [26] or based on reinforcement learning [27] may be employed. These learning approaches make it possible to improve AV navigation performance when the navigation experience is increasing over the operation time. This might happen in the conditions of a dynamic environment, which is typical for operating in the real world. However, ML-based navigation approaches may poorly generalize and easily "forget" previous knowledge that is displaced by recent experience. This inevitably leads to AV overall navigation performance degradation. The analysis of existing AV computer-vision-based navigation and localization approaches demonstrates that these methods usually require a substantial number of visual landmarks in order to maintain the level of navigation performance that is acceptable by the user and application requirements. Most of the presented approaches are primarily acceptable for indoor navigation, as its environment can often be characterized by a large number of informative reference points. Outdoor navigation is much more challenging, as its environment usually provides insufficient visual information for accurate positioning and localization, especially in specific areas such as the Arctic region. In these conditions, computer-vision-based navigation and localization methods might not be effective. To approach this challenge, additional artificial visual landmarks may be introduced to the AV's operating environment for more accurate localization. The AprilTag is one of the known examples of such artificial landmarks [21]. These tags may carry various information that might be used by AVs to perform localization and mapping on the terrain; for example, they might provide the current GPS coordinates where these tags are allocated. However, artificial visual landmarks are still required to be properly recognized by the AV's computer-vision-based system in extreme environmental conditions. Petrov [28] proposed a system for navigating AVs in a winter terrain by extracting the vector of the visual navigation marker. This approach might be effectively employed when the visual conditions are more or less stable,

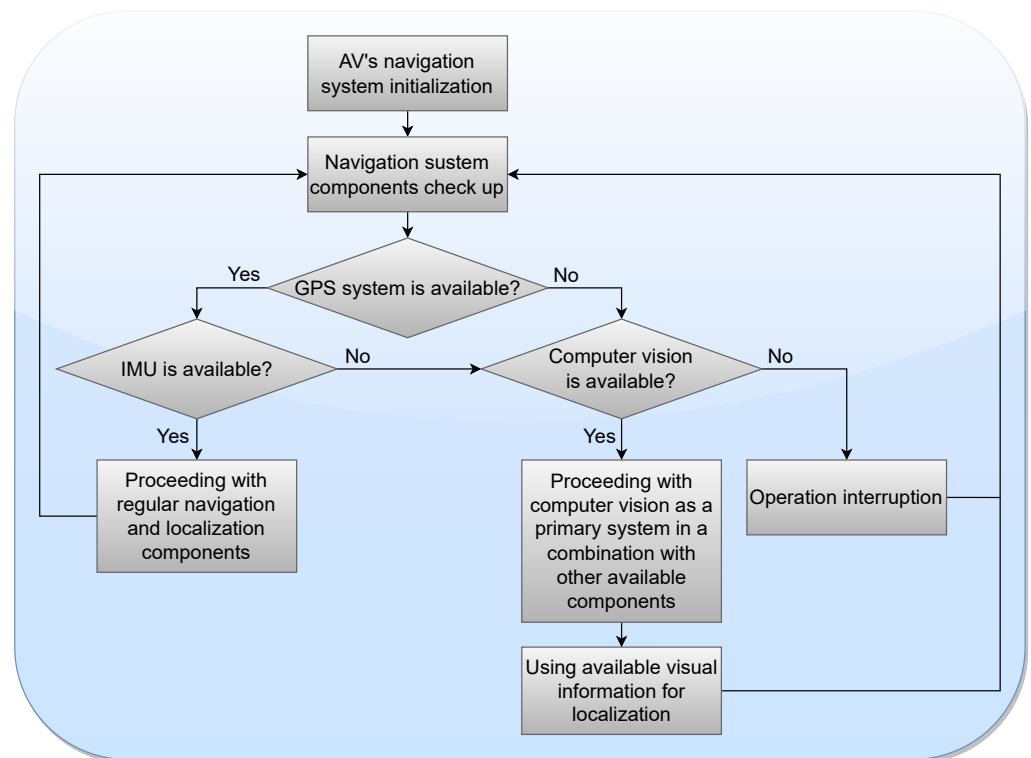
without significant environmental condition degradation. Ali and Shahhood [29] developed a system for navigation in a 3D space, including open terrain areas. The disadvantage of the solution is that it also does not consider the AV's operating conditions, which makes it unsuitable for complex environments.

In this paper, we focus on rather complex, dynamic, and weather-dependent environmental conditions that affect the quality of the obtained visual information in various ways. Specifically, we developed a computer-vision-based AV localization system and enhanced its robustness regarding the on-board camera image quality degradation caused by harsh environmental conditions. One of the significant challenges we addressed is the low informative features visibility caused by the unstable environment. The extreme weather can cause camera lenses to freeze or fog, which can further degrade the image quality and affect the computer-vision-based localization performance. Additionally, the reflective properties of snow and ice can create glare and distortions in the images captured by the system, thereby further degrading the DQ. We present a solution capable of adapting to those complex weather conditions by employing an ML model for a computer-vision-based localization system. As the visual information extractor, we employed a convolutional neural network (CNN), as it demonstrated the highest performance in our empirical investigation. To enhance the robustness of our ML-based model, we produced images affected by various complex environmental conditions and employed them to re-train the model. To further assure the ML model performance on the low quality data, we expanded the training dataset by modifying it using affine transformations. Below, we describe the architecture and major components of the developed AV computer-vision-based localization system and how it might be integrated in the overall navigation process.

### 3. Integration of the Computer-Vision-Based Localization into the AV's Navigation Stack

The generic AV navigation process can be described as follows: if GPS is available, it is employed as the primary AV navigation system to perform route planning and proceeding to the target destination. By default, computer-vision-based localization is considered as the auxiliary navigation system (ANS). Depending on the AV's composition, it might also employ other ANSs, such as IMU or radar. ANSs might be used in combination with GPS or with each other (which is a typical practice) or individually if one or more navigation components fail [30]. ANSs are routinely used to correct the GPS navigation error and for precise positioning and obstacles detection. Based on the available data from the navigation and localization system components, the AV makes decisions on further driving actions. Figure 2 schematically represents the routine process of decision making depending on the navigation and localization components' availability and their performance. For the sake of demonstration, an IMU was used here as another ANS that the AV was equipped with. Initially, the availability of the GPS signal was verified. With a successful response from the GPS, the system state switches to the IMU availability verification. The main problem with relying solely on the IMU for the navigation and localization operations is the complex environmental conditions in which the AV has to operate. When primarily relying on the IMU, the navigation error is rapidly accumulated [31], which ultimately leads to the AV's disorientation. In this case, such an error can be corrected by the GPS coordinates; however, the GPS system might be unavailable for a considerable time. In this case, localization and further navigation might be performed based on computer vision in combination with other available ANSs.

The problem of time synchronization for a IMU/GNSS is also challenging for reliable navigation, and it is addressed in the following way: the AV's sensors operate in series, not in parallel. In the sequential operation of sensors, the data is transmitted from one system to another, thus receiving additional parameters for work. The information first passes through the IMU, then through the GNSS. The use of AprilTags makes it possible to verify inaccuracies in the IMU and correct their impact on the navigation performance.



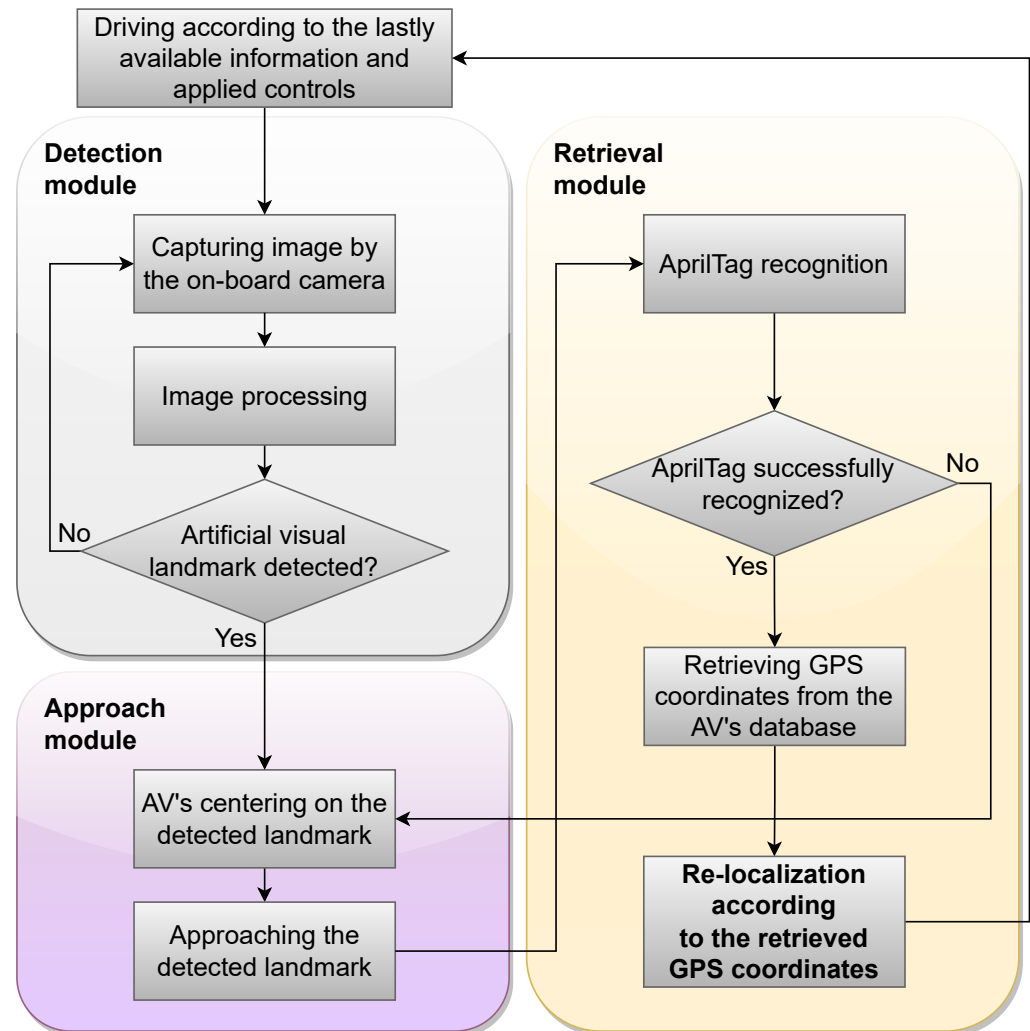
**Figure 2.** Schematic representation of how the proposed computer-vision-based localization might be integrated into the AV navigation system process—our design example. The process might be represented as an iterative algorithm that incorporates verification of the navigation and localization facilities the AV is equipped with, and it employs them if available.

The computer-vision-based localization system should be capable of detecting and retrieving the information from the artificial visual landmark in the current location. The architecture of the proposed computer-vision-based localization incorporates the following basic modules:

- A detection module;
- An approach module;
- A retrieval module.

The architecture of these modules and the connections between them are represented in Figure 3. The detection module operates based on the implemented and trained ML model, the approach module is implemented based on the Nvidia Isaac “Followme” module [32], and the retrieval module utilizes the AprilTag recognition function [21]. Each module performs unique functions within the overall AV navigation system. The detection module is responsible for detecting the artificial visual landmark on which the AprilTag with the visually encoded GPS coordinates is printed. The approach module allows for pointing and holding the detected landmark in the field of the AV’s on-board camera view while the AV is traveling closer to properly recognize the tag. The retrieval module is responsible for retrieving the information on the current GPS coordinates from the visual tag, and it performs the AV’s localization adjustment according to these coordinates. In the case of two or more landmarks appearing in the on-board camera’s field of view simultaneously, the one that is detected with a higher performance is selected for proceeding toward it. If multiple visual landmarks have a similar detection performance, then their bounding boxes are compared, and the decision is made based on the higher intersection over union (IoU) value. If the IoU value is also similar, the visual landmark is selected randomly. In our empirical study, we assumed that the visual landmarks were allocated with the distance of 20 m between each other, which makes multiple detection almost impossible. The detection module is the initial component that informs the other two modules if there

are any artificial visual landmarks detected. The other two modules operate in parallel in order to reduce the time for the information retrieval. Such a combination of actions makes it possible to minimize the time required to perform localization and mapping. Instead of approaching within a certain distance to the tag to properly recognize it, the retrieval process is constantly performed throughout the AV's approach.



**Figure 3.** Schematic representation of the proposed computer-vision-based system architecture and its operation—our design example. The proposed architecture is composed of three modules: detection, approach, and retrieval. The detection module is responsible for analyzing the visual data obtained from the on-board camera and detecting artificial landmarks. The approach module is responsible for keeping the visual landmark in the AV's on-board camera field of view by adjusting the movement while approaching. The retrieval module is responsible for recognizing the AprilTag and retrieving the information on the GPS coordinates.

The detection module utilized the YOLOv4 ML object detection model [33] that was pre-trained on the MS COCO dataset [34]. This is a reliable version of a well-known architecture that outperforms other popular ML models in terms of both performance and efficiency [33].

The approach module operation was based on the software provided by the Nvidia Isaac "Followme" software development kit [32]. This software allows for AVs to perform the recognition of the detected object and correct the movement actions while approaching it. The original programming logic for the "Followme" component implies that the AV should be close enough to the AprilTag to recognize it. When considering extreme environmental

conditions, this might be impractical, as it will deteriorate the AV's operation effectiveness and efficiency. To minimize the approach time in order to improve the overall navigation and localization system efficiency, we introduced our modifications into the "Followme" source code. Instead of initially detecting the AprilTag and using it as a reference point for the approach, we switched to the visual artificial landmark as the initial object of interest for detection. Harsh weather conditions significantly deteriorate the AprilTag's visibility and detectability, especially from a large distances. Artificial visual landmarks themselves are rather large and more visible in such conditions, which makes them good candidates to be utilized as reference points. Hence, the AV can detect these artificial visual landmarks from a longer distance and spend less resources for this detection, which makes navigation more efficient.

Once the visual landmark is detected, it is approached until the information is retrieved successfully from the AprilTag, which is printed on this landmark. The successful information retrieval means that the current location is extracted from the AprilTag. The AV stores the ID of the AprilTag, which is related to the GPS coordinates represented in the format of  $DD.DDDDD, DD.DDDDD$ , where  $DD$  stands for the longitude and latitude values. An example of the encoded data is 40.17312, 42.84121. When the AprilTag is recognized, the corresponding coordinates are retrieved from the AV's database. After retrieving the GPS coordinates, the current coordinates and path of the AV is updated, and all passed landmarks are recorded. After re-localization according to the retrieved coordinates, the system responsible for the AV's routing updates the route to the target destination by considering the current location, and the AV proceeds to the target destination according to the process displayed in Figure 3.

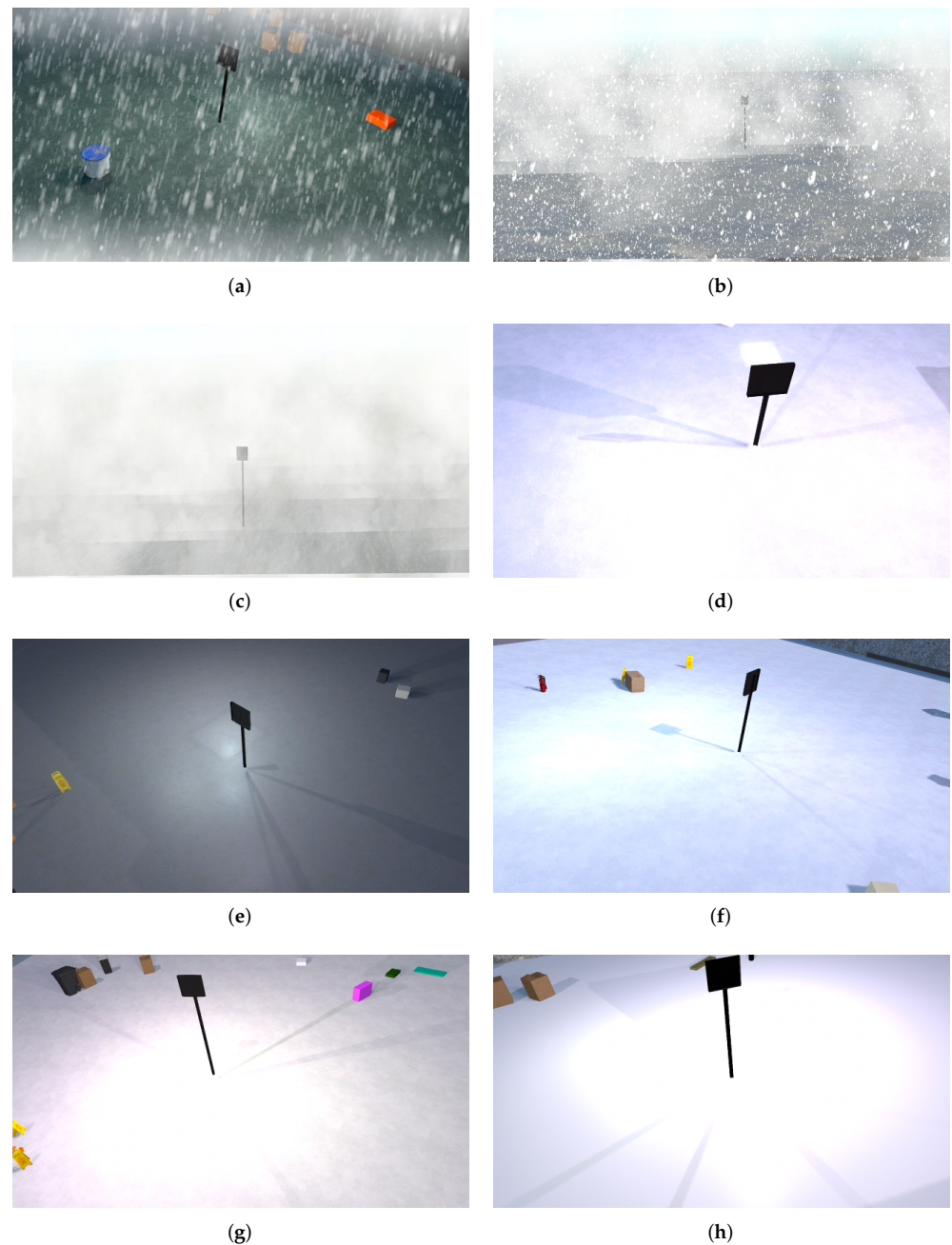
#### 4. Employed Data and Transformations

Despite the numerous publicly available image datasets of various sizes, categories, and designs for different knowledge domains, those datasets are usually composed of high quality data. For our considered specific Arctic region use case, to produce the ML model that is robust regarding possible DQ degradation, we also needed images affected by the extreme environmental conditions. We found the available public datasets not suitable for our specific use case, so we developed the visual artificial landmark model with the printed AprilTag by employing the Nvidia Isaac Sim application [32]. Alongside this application, we also utilized the Unity3D software package [35] to generate images of the visual tags in various extreme environment conditions that were captured from various distances and angles. The Nvidia Isaac Sim made it possible to configure numerous parameters, such as object location, illumination, overlapping with other objects, etc. With these settings, it was possible to produce a substantial collection of the unique data samples to train ML models. An example of the designed visual artificial landmark is demonstrated in Figure 4. Considering the features of the Arctic region terrain, we produced a substantial set of images capturing the artificial visual landmark in various complex environmental conditions, such as snow drifts, blizzards, insufficient illumination, etc. In Figure 4, we represent a few instances of the produced images. The constructed artificial visual landmark consisted of two parts: the holder and the panel for the AprilTag. The height of the holder was 1 m, the size of the panel was 40 by 40 cm, and the size of the printed AprilTag was 30 by 30 cm.

From the initial  $\approx 4000$  original good quality images, we generated around 20,000 samples of various qualities that were captured from various angles and distances. To further increase the image diversity, we also employed affine transformation techniques. Random affine transformations are geometrical vector modifications that are introduced to existing images. Such transformation methods as rotation, scale modification, shift, and reflection might be applied, which make it possible to produce a visually distinct image without discarding the essential details and structure of the image. Affine transformations are widely used in computer vision applications to expand the diversity of the samples in order to train more generalizable models. This approach is especially useful in training deep neural



networks that require a large amount of diverse data to achieve acceptable performance. In our work, we employed such affine transformations as rotation, scale modification, shift, and reflection alongside the changing environment to produce images that would be infeasible to obtain in real conditions. By employing the affine transformations, we expanded our data by 3126 additional samples. Then, we manually labeled the produced data samples to employ them for further training procedures. The produced data was split into 70% training, 15% validation, and 15% test batches.



**Figure 4.** Artificial visual landmark image samples generated in Unity3D software package. The figure demonstrates multiple image examples of varying quality under diverse environmental conditions: (a) heavy rain; (b) heavy snowfall; (c) blizzard; (d–h) captured from varied distances, angles, and in various illumination conditions.

## 5. Empirical Study

The produced labels were represented by the class ID and point coordinates of the visual landmark in the  $[x, y, x + w, y + h]$  format, where  $w$  and  $h$  are the width and height of the bounding box, respectively,  $x$  is the initial coordinate of the bounding box on the horizontal axis, and  $y$  is the initial coordinate of the bounding box on the vertical axis. As the employed ML model was initially pre-trained on the MS COCO dataset [34], we only re-trained the fully connected layer, which used the Softmax activation function. As a computing device for training, we employed the Nvidia Jetson AGX Xavier [36]. This device is known for its low power consumption of only 10W, which allows it to be employed in autonomous mobile robotic devices. The device makes it possible to utilize CUDA cores for real-time calculations. The employed device possesses the following technical characteristics:

- AI performance: 32 TOPS (trillion operations per second);
- GPU: Nvidia Volta architecture with 512 Nvidia CUDA cores and 64 tensor cores;
- CPU: 8-core 64-bit Nvidia Carmel processor with Armv8.2 architecture, 8 MB L2, and 4 MB L3;
- Deep learning accelerator: 2 NVDLA Engines;
- Computer vision accelerator: 2 PVA accelerators.

For the ML model performance evaluation metric, we employed Average Precision ( $AP$ ), which incorporates precision and recall and is usually employed for evaluating ML performance [37].  $AP$  can be characterized as the area under the precision–recall curve, which shows how the accuracy and completeness of the algorithm change at various classification probability thresholds.

The IoU is a metric that is usually employed to evaluate the object detection algorithm's performance. Typical threshold values indicate correct detection when the IoU is more than 50%. The IoU value is calculated as the relationship between the intersection areas of the predicted bounding box and the actual ground truth bounding box. Since we are evaluating accuracy for the detection and classification algorithm, we also employed the mean  $AP$  ( $mAP$ ) metric, which can be calculated by Equation (1):

$$mAP = \sum_{q=1}^Q AP(q), \quad (1)$$

where  $Q$  is the number of classes;  $q$  is the class; and  $AP(q)$  is the  $AP$  value for the  $q$ 's class.

To select an ML model architecture for our use case, we performed an ablation study where we tested how various ML model architectures performed on a testing set of the produced data samples. We tested ML architectures based on a CNN (YOLOv4); Haar cascades; and based on ML with a histogram of directional gradients (HOG) used as a descriptor for a support vector machine (SVM). We pre-trained these models on the same cohort of the produced data and then tested them over the prepared testing set. We employed ML accuracy as the evaluation metric for this ablation study. The ML performance results are presented in Table 1. According to the results, the YOLOv4 object detector that is based on CNN architecture demonstrated the highest accuracy in the studied conditions. Based on these results, we selected the YOLOv4 as an image processor for the designed AV computer-vision-based localization system. Furthermore, the choice of the YOLOv4 was justified by the fact that, at the time of the research, this architecture was recommended for achieving robust real-time computation performance using the employed Nvidia Jetson AGX Xavier computational device.

**Table 1.** Results of the ablation study for evaluating the performance of three distinct ML architectures over the data samples produced with Unity3D. The testing data includes image samples affected by the extreme weather conditions and captured from various angles.

	YOLOv4, %	Haar Cascades, %	HOG + SVM, %
Accuracy	78.75	42.39	49.17

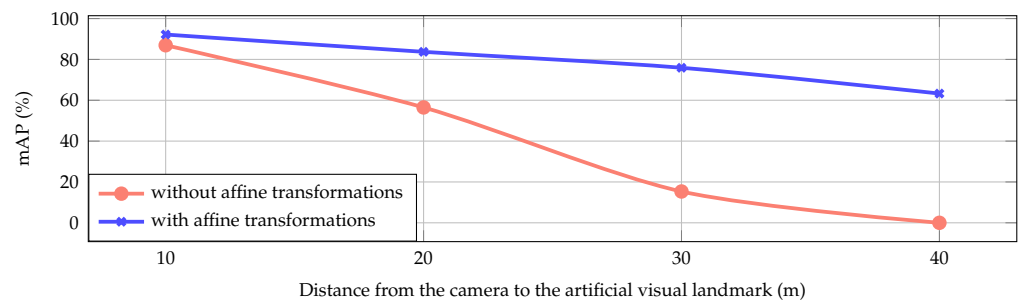
Initially, we re-trained our model for 2000 epochs on the images without affine transformations and tested it on the produced images of the artificial visual landmark captured from the varied distances (from 5 to 20 m) and affected by various environmental conditions. The first column in Table 2 represents the performance results achieved by this model. As one can see, the model poorly detected the target objects in the images and provided unacceptable performance. We then continued to further re-train the model, but, in this case, it was only trained on the images produced with affine transformations. After further training, we test the model again over the similar images employed in the former case. These testing results are represented in the second column of Table 2. The model was able to achieve substantial performance improvement with the 83.8% for the mAP and 78.73% for the average IoU.

**Table 2.** Results comparing the effect of re-training the ML model on the data modified with affine transformations. The affine transformations include such geometrical vector modifications as rotation, scale modification, shift, and reflection.

	Model without Affine Transformations, %	Model with Affine Transformations, %
mAP	25.8	83.8
Average IoU	46.31	78.73

To evaluate the performance of the produced model in various weather conditions, we formed four distinct testing data cohorts based on the distance to the artificial visual landmarks in the images. In our experiments, we employed images captured from the distance of 10, 20, 30, and 40 m. Figure 5 represents the performance demonstrated by the model on these data cohorts. For the sake of comparison, we also presented the results for the model trained on the data without affine transformations. From the results, one can see the employment of various augmentation techniques over the training data, wherein the quality was significantly deteriorated by extreme environmental conditions, which made it possible to substantially enhance the ML performance in the case of testing sample quality degradation. However, as the distance between the visual landmark and the camera increased, the tag recognition performance dropped. A distance of 30 m from the tag contributed to a 60% ML performance decline. Furthermore, when starting from a distance of 40 m away from the tag, the model completely stopped recognizing the tag.

In addition to extreme weather conditions, we also considered changing illumination as another quality degradation factor. In the Arctic region, daylight is highly limited in most times of the year, which poses an additional challenge for AV computer-vision-based localization systems. To evaluate how our model was robust regarding the changes in illumination, we formed a number of testing data cohorts. They were generally divided into two categories: daylight, which incorporates images captured in the lightning conditions of 155 LUX, and nighttime, with images captured in the extremely limited lightning conditions of 5 LUX. Each of these categories included three sub-categories of images captured from the varied distances of 0.5, 2, and 5 m. Table 3 demonstrates the results of the ML model performance over the employed images cohorts.



**Figure 5.** The tag recognition performance by the ML model trained with and without affine transformations tested on images affected by various extreme weather conditions and captured from different distances. The results are presented for the model trained only on the produced images without re-training on data modified by affine transformations (orange plot) and for the models further re-trained on the data modified by the affine transformations (blue plot).

As one can see, in the case of the 0.5 m distance, the ML model demonstrated high performance in both daylight and nighttime conditions, with an accuracy of 97% and 94%, respectively. However, increasing the distance from the camera to 2 m affected the performance significantly, and even critically, in the nighttime conditions, while, in case of the daytime, the performance was still over 80%. For the nighttime images, it dropped below 50%, which substantially exacerbated the AV's computer-vision-based localization. Further increasing the distance to 5 m completely prevented the ML model from selecting the correct operation in the nighttime. However, for the images captured in the daytime, even though the demonstrated performance was impractical, it was still above the 50% boundary.

**Table 3.** Comparison of the performance demonstrated by the ML model in recognizing AprilTags captured under different illumination conditions and from various distances. The distance varied from 0.5 to 5 m, and the illumination was 155 LUX for the daytime and 5 LUX for the nighttime.

Distance, m	Accuracy (Daylight), %	Accuracy (Nighttime), %
0.5	97	94
2	83	43
5	54	0

## 6. Limitations and Potential Future Directions

Both our empirical study and the developed solution have several limitations. First, as the use case scenario for our solution, we considered the AV's operation in a structured environment with natural or artificial visual landmarks allocated in advance over the operation area. An AV has to possess prior knowledge on the environment, which might be realized with an off-line map and relationships between the known landmarks and GPS coordinates stored in the AV's memory. Second, instead of producing and employing real data for training the ML model, we utilized the well-known Unity3D software package to generate artificial samples designed to represent affects by extreme weather conditions. This allowed us to significantly extend the size and the diversity of the dataset in order to train the model to be more robust in varying conditions. In addition, in our research, we concentrated only on improving the robustness of the computer-vision-based localization system against the low quality data affected by the harsh weather conditions. We did not consider additional influences that these conditions might have had on the AV's navigation and localization quality, such as failures of the other AV systems and elements responsible for driving control.

Our approach improved the value and contributions developed in our work. As we showed in our experiments, pre-trained ML models further trained on augmented low quality data become more robust regarding DQ degradation. Since Unity provides a reliable

simulation environment for designing and testing AI-based solutions for industrial robotic systems before their actual implementation and deployment, the artificial samples we produced can be considered as robust and close-to-real. Similar training operations might be performed with real images in the case of implementing our solution in practice. In our development, we concentrated only on enhancing the robustness of computer-vision-based localization against the DQ degradation caused by extreme weather conditions, which we effectively demonstrated in our results. Despite our value for the detrimental effects of other processes and factors that may influence the navigation, localization, and driving operations, they are out of our paper's scope.

As the potential future research directions, we can propose research on the optimization of the landmarks' design, their spatial distribution, and the sensor fusion. The optimal size and visual characteristics of the artificial visual landmarks play important roles in their early detection and optimal approaching by the AVs. Comprehensive research resulting in recommendations on the shape, appearance, and other visual characteristics for improving artificial visual landmark detection in extreme environmental conditions by the computer-vision-based systems would help to enhance the navigation and localization quality and performance. Another direction is the landmarks' optimal allocations over the AV's operation area. This is a complex problem that requires the consideration of numerous factors, such as the AV's technical characteristics, landscape features, dynamic environment conditions, operation area dimensions, starting and target destination locations, etc. One more direction is sensor fusion optimization aimed at selecting appropriate sensors and integrating the data acquired from them in order to improve the overall DQ. Progress in this direction would make it possible to effectively produce data whose quality satisfies the user and application requirements. This would allow for enhancing the quality of intelligent navigation.

## 7. Conclusions

AV navigation is a challenging problem in the conditions of GNSS signal instability or absence, which might often be the case in certain regions as diverse as dense architecture urban areas or low populated areas with extreme environmental conditions. To address this problem, data from other AV sensor systems are commonly employed to maintain the required localization and mapping accuracy. However, the quality of measurements produced by those sensors might be affected by various factors; for example, the quality of images from the on-board camera might be degraded by harsh weather conditions. This leads to the AV's navigation performance deterioration, decreases the AV's effectiveness and efficiency, and might jeopardize the safety of AVs and surrounding objects. In this paper, we proposed an enhanced design of a computer-vision-based localization system that can be integrated into AV navigation facilities to improve their localization and mapping accuracy in extreme environmental conditions where GPS signals are unstable or absent. Based on our empirical study results, the following conclusions can be formulated. A CNN-based ML model architecture (in particular, YOLOv4), demonstrated performance that was higher than the two other investigated architectures and can be employed as an image processor for AV localization applications. However, as our research showed, even the model trained on a comprehensive set of images representing affects by various extreme environmental conditions, was insufficient to achieve performance acceptable for effective and efficient AV operation. Employing data augmentation techniques such as affine transformations and further training the model on this data significantly improved the model robustness regarding testing image quality degradation. Indeed, the further trained model was able to demonstrate a performance higher than 80% for the degraded images captured from the distance of up to 20 m. In the changing illumination conditions, the model was able to achieve high performance in the distance of up to 2 m in daylight. Unfortunately, in the nighttime, when the distance increased above 0.5 m, the model was hardly able to properly recognize the tag.

**Author Contributions:** Conceptualization, L.R. and I.V.; methodology, S.C.; software, P.B.; validation, P.B.; formal analysis, R.G.; investigation, P.B.; resources, E.N.; data curation, P.B. and E.N.; writing—original draft preparation, S.C.; writing—review and editing, R.G. and L.R.; visualization, S.C.; project administration, L.R., R.G., E.N. and I.V.; funding acquisition, L.R. and I.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work of Sergei Chuprov and Leon Reznik was supported by the United States Military Academy (USMA), Grant Number W911NF-20-1-0337. The work of Pavel Belyaev, Ruslan Gataullin, Evgenii Neverov, and Ilia Viksnin was supported by the Ministry of Science and Higher Education of the Russian Federation, known as “Goszadanie” №075-01024-21-02 of 29.09.2021 (project FSEE-2021-0014).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study is available on request from the authors: I.V. and P.B. The data is not publicly available due to it is a commercial company’s intellectual property.

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

## References

1. Stateczny, A.; Specht, C.; Specht, M.; Brčić, D.; Jugović, A.; Widźgowski, S.; Wiśniewska, M.; Lewicka, O. Study on the positioning accuracy of GNSS/INS systems supported by DGPS and RTK receivers for hydrographic surveys. *Energies* **2021**, *14*, 7413. [[CrossRef](#)]
2. Chuprov, S.; Viksnin, I.; Kim, I.; Melnikov, T.; Reznik, L.; Khokhlov, I. Improving Knowledge Based Detection of Soft Attacks Against Autonomous Vehicles with Reputation, Trust and Data Quality Service Models. In Proceedings of the 2021 IEEE International Conference on Smart Data Services (SMDS), Chicago, IL, USA, 5–10 September 2021; pp. 115–120.
3. Chuprov, S.; Viksnin, I.; Kim, I.; Reznikand, L.; Khokhlov, I. Reputation and trust models with data quality metrics for improving autonomous vehicles traffic security and safety. In Proceedings of the 2020 IEEE Systems Security Symposium (SSS), Crystal City, VA, USA, 1 July 2020–1 August 2020; pp. 1–8.
4. Khokhlov, I.; Chuprov, S.; Reznik, L. Integrating Security with Accuracy Evaluation in Sensors Fusion. In Proceedings of the 2022 IEEE Sensors, Dallas, TX, USA, 30 October 2022–2 November 2022; pp. 1–4.
5. Chuprov, S.; Reznik, L.; Khokhlov, I.; Manghi, K. Multi-Modal Sensor Selection with Genetic Algorithms. In Proceedings of the 2022 IEEE Sensors, Dallas, TX, USA, 30 October 2022–2 November 2022; pp. 1–4.
6. Reznik, L. *Intelligent Security Systems: How Artificial Intelligence, Machine Learning and Data Science Work for and against Computer Security*; IEEE Press-John Wiley & Sons: Hoboken, NJ, USA, 2022.
7. Chuprov, S.; Reznik, L.; Obeid, A.; Shetty, S. How Degrading Network Conditions Influence Machine Learning End Systems Performance? In Proceedings of the IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), New York, NY, USA, 2–5 May 2022; pp. 1–6.
8. Chuprov, S.; Satam, A.N.; Reznik, L. Are ML Image Classifiers Robust to Medical Image Quality Degradation? In Proceedings of the 2022 IEEE Western New York Image and Signal Processing Workshop (WNYISPW), Rochester, NY, USA, 4 November 2022; pp. 1–4.
9. Chuprov, S.; Khokhlov, I.; Reznik, L.; Shetty, S. Influence of Transfer Learning on Machine Learning Systems Robustness to Data Quality Degradation. In Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 18–23 July 2022; pp. 1–8.
10. Rajathi, K.; Gomathi, N.; Mahdal, M.; Guras, R. Path Segmentation from Point Cloud Data for Autonomous Navigation. *Appl. Sci.* **2023**, *13*, 3977. [[CrossRef](#)]
11. Heidari, A.; Jafari Navimipour, N.; Unal, M.; Zhang, G. Machine Learning Applications in Internet-of-Drones: Systematic Review, Recent Deployments, and Open Issues. *ACM Comput. Surv.* **2023**, *55*, 1–45. [[CrossRef](#)]
12. Amiri, Z.; Heidari, A.; Navimipour, N.; Unal, M. Resilient and dependability management in distributed environments: a systematic and comprehensive literature review. *Clust. Comput.* **2022**, *26*, 1565–1600. [[CrossRef](#)]
13. Kim, I.; Matos-Carvalho, J.P.; Viksnin, I.; Simas, T.; Correia, S.D. Particle Swarm Optimization Embedded in UAV as a Method of Territory-Monitoring Efficiency Improvement. *Symmetry* **2022**, *14*, 1080. [[CrossRef](#)]
14. Kim, H.; Choi, Y. Development of Autonomous Driving Patrol Robot for Improving Underground Mine Safety. *Appl. Sci.* **2023**, *13*, 3717. [[CrossRef](#)]
15. Mascarich, F.; Kulkarni, M.; De Petris, P.; Wilson, T.; Alexis, K. Autonomous mapping and spectroscopic analysis of distributed radiation fields using aerial robots. *Auton. Robots* **2022**, *47*, 139–160. [[CrossRef](#)]
16. Liu, W.; Hua, M.; Deng, Z.; Huang, Y.; Hu, C.; Song, S.; Gao, L.; Liu, C.; Xiong, L.; Xia, X. A systematic survey of control techniques and applications: From autonomous vehicles to connected and automated vehicles. *arXiv* **2023**, arXiv:2303.05665.

17. Chen, G.; Hua, M.; Liu, W.; Wang, J.; Song, S.; Liu, C. Planning and Tracking Control of Full Drive-by-Wire Electric Vehicles in Unstructured Scenario. *arXiv* **2023**, arXiv:2301.02753.
18. Xia, X.; Hashemi, E.; Xiong, L.; Khajepour, A. Autonomous Vehicle Kinematics and Dynamics Synthesis for Sideslip Angle Estimation Based on Consensus Kalman Filter. *IEEE Trans. Control Syst. Technol.* **2022**, *31*, 179–192. [[CrossRef](#)]
19. Gao, L.; Xiong, L.; Xia, X.; Lu, Y.; Yu, Z.; Khajepour, A. Improved vehicle localization using on-board sensors and vehicle lateral velocity. *IEEE Sens. J.* **2022**, *22*, 6818–6831. [[CrossRef](#)]
20. Ryghaug, M.; Haugland, B.T.; Søråa, R.A.; Skjølvold, T.M. Testing Emergent Technologies in the Arctic: How Attention to Place Contributes to Visions of Autonomous Vehicles. *Sci. Technol. Stud.* **2022**, *35*, 4–21. [[CrossRef](#)]
21. Olson, E. AprilTag: A robust and flexible visual fiducial system. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 3400–3407.
22. Yastrebova, A.; Höyhty, M.; Boumard, S.; Lohan, E.S.; Ometov, A. Positioning in the Arctic region: State-of-the-art and future perspectives. *IEEE Access* **2021**, *9*, 53964–53978. [[CrossRef](#)]
23. Marinho, L.B.; Rebouças Filho, P.P.; Almeida, J.S.; Souza, J.W.M.; Souza Junior, A.H.; de Albuquerque, V.H.C. A novel mobile robot localization approach based on classification with rejection option using computer vision. *Comput. Electr. Eng.* **2018**, *68*, 26–43. [[CrossRef](#)]
24. Zeng, J.; Ju, R.; Qin, L.; Hu, Y.; Yin, Q.; Hu, C. Navigation in Unknown Dynamic Environments Based on Deep Reinforcement Learning. *Sensors* **2019**, *19*, 3837. [[CrossRef](#)] [[PubMed](#)]
25. Wang, Y.; He, H.; Sun, C. Learning to Navigate Through Complex Dynamic Environment With Modular Deep Reinforcement Learning. *IEEE Trans. Games* **2018**, *10*, 400–412. [[CrossRef](#)]
26. Kahn, G.; Villafior, A.; Ding, B.; Abbeel, P.; Levine, S. Self-supervised deep reinforcement learning with generalized computation graphs for robot navigation. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; pp. 5129–5136.
27. Han, S.H.; Choi, H.J.; Benz, P.; Loaiciga, J. Sensor-Based Mobile Robot Navigation via Deep Reinforcement Learning. In Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Shanghai, China, 15–17 January 2018; pp. 147–154.
28. Dmitriy, P. Navigation System of an Autonomous Mobile Robot for Work in the Far North. Master’s Thesis, Tomsk Polytechnic University, Tomsk, Russia, 2020. Available online: <http://earchive.tpu.ru/handle/11683/61872> (accessed on 29 March 2023).
29. Ali, D.A.; Shahhood, F. Modular and hierarchical approach to agent training to explore a three-dimensional environment. In Proceedings of the Student Science Spring, Moscow, Russia, 1–30 April 2021; pp. 251–252.
30. Chuprov, S.; Marinenkov, E.; Viksnin, I.; Reznik, L.; Khokhlov, I. Image processing in autonomous vehicle model positioning and movement control. In Proceedings of the 2020 IEEE 6th World Forum on Internet of Things (WF-IoT), New Orleans, LA, USA, 2–16 June 2020; pp. 1–6.
31. Liu, X.; Zhou, Q.; Chen, X.; Fan, L.; Cheng, C.T. Bias-error accumulation analysis for inertial navigation methods. *IEEE Signal Process. Lett.* **2021**, *29*, 299–303. [[CrossRef](#)]
32. Monteiro, F.F.; Vieira, A.L.B.; Teixeira, J.M.X.N.; Teichrieb, V. Simulating real robots in virtual environments using NVIDIA’s Isaac SDK. In Proceedings of the Anais Estendidos do XXI Simpósio de Realidade Virtual e Aumentada, Rio de Janeiro, Brazil, 28–31 October 2019; pp. 47–48.
33. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
34. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014.
35. Ogre Development Team. UNITY COMPUTER VISION—Free Tools and Content for Generating Synthetic Data. Available online: <https://unity.com/products/computer-vision> (accessed on 25 March 2023).
36. NVIDIA Jetson AGX Xavier. AI Platform for Autonomous Machines. Available online: <https://www.nvidia.com/en-us/autonomous-machines/jetson-agx-xavier/> (accessed on 20 April 2023).
37. Padilla, R.; Netto, S.L.; da Silva, E.A.B. A Survey on Performance Metrics for Object-Detection Algorithms. In Proceedings of the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 1–3 July 2020; pp. 237–242.

**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.