drones

MDPI

Technical Note

VisionICE: Air–Ground Integrated Intelligent Cognition Visual Enhancement System Based on a UAV

Qingge Li, Xiaogang Yang *, Ruitao Lu, Jiwei Fan, Siyu Wang and Zhen Qin

Department of Missile Engineering, Rocket Force University of Engineering, Xi'an 710000, China; lqg19950105@163.com (Q.L.); lrt19880220@163.com (R.L.); fjw19900619@163.com (J.F.); wsy960328@163.com (S.W.); qinzhenpro@163.com (Z.Q.)

* Correspondence: doctoryxg@163.com

Abstract: Post-disaster search and rescue is critical to disaster response and recovery efforts and is often conducted in hazardous and challenging environments. However, the existing post-disaster search and rescue operations have problems such as low efficiency, limited search range, difficulty in identifying the nature of the target, and wrong target location. Therefore, this study develops an air–ground integrated intelligent cognition visual enhancement system based on a UAV (VisionICE). The technique combines a portable AR display device, a camera-equipped helmet, and a quadcopter UAV for efficient patrols over a wide area. First, the system utilizes wireless image sensors on the UAV and helmet to capture images from the air and ground views. Using the YOLOv7 algorithm, the cloud server calculates and analyzes these visual data to accurately identify and detect targets. Lastly, the AR display device obtains real-time intelligent cognitive results. The system allows personnel to simultaneously acquire air and ground dual views and achieve brilliant cognitive results and immersive visual experiences in real time. The findings indicate that the system demonstrates significant recognition accuracy and mobility. In contrast to conventional post-disaster search and rescue operations, the system can autonomously identify and track targets of interest, addressing the difficulty of a person needing help to conduct field inspections in particular environments. At the same time, the system can issue potential threat or anomaly alerts to searchers, significantly enhancing their situational awareness capabilities.

Keywords: unmanned aerial vehicle; YOLOv7; intelligent cognition; augmented reality

Citation: Li, Q.; Yang, X.; Lu, R.; Fan, J.; Wang, S.; Qin, Z. VisionICE: Air–Ground Integrated Intelligent Cognition Visual Enhancement System Based on a UAV. *Drones* **2023**, *7*, 268. https://doi.org/10.3390/drones7040268

Academic Editor: Diego González-Aguilera

Received: 29 March 2023
Revised: 11 April 2023
Accepted: 12 April 2023
Published: 13 April 2023

1. Introduction

Post-disaster search and rescue (SAR) is a complex and challenging endeavor. Search and rescue operations involve locating and extracting individuals trapped or injured in the aftermath of a disaster and are often conducted in hazardous and complex environments. The traditional post-disaster search and rescue operations have the problems of low search efficiency, limited search range, difficulty in identifying the nature of the target, and inaccurate positioning of target coordinates. Recent advances in technology, such as unmanned aerial vehicles (UAVs) and artificial intelligence (AI), have the potential to significantly enhance the effectiveness and efficiency of post-disaster search and rescue operations. Post-disaster searches shifted progressively to multi-dimensional, autonomous, intelligent methods due to the rapid development of intelligent technology.

Drone-integrated systems have emerged as a promising tool for search and rescue operations in recent years [1]. UAVs are irreplaceable in unique and complex environments due to their wide search range, good concealment performance, and high mobility [2,3]. By being equipped with sensors such as cameras, drones can detect and locate individuals or objects in areas that may be difficult or dangerous for human rescuers to access. In recent years, computer vision technology has made breakthroughs with the support of big data processing and high-performance cloud computing. The integration of computer

vision technology and UAV technology has effectively addressed UAV surveillance more significantly [4,5] and is a powerful tool for achieving situational awareness, target indication [6], and ground target tracking [7,8]. Therefore, combining post-disaster search and rescue and intelligent UAVs gives the searchers both ground and air perspectives. It solves the problem that field surveillance cannot be carried out under particular circumstances, making the search range larger and having higher mobility.

Augmented reality (AR) technology can enhance the capabilities of drones, allowing them to perform complex tasks and provide real-time situational awareness to operators. AR technology can effectively reflect the natural world's content and overlay virtual information into the real world. AR technology involves overlaying digital information, such as images, video, and text, onto the physical environment, creating an augmented view of reality [9–12]. When combined with drones, AR technology can provide operators with a real-time view of the drone's surroundings, as well as additional information and data, such as flight paths, obstacle detection, and telemetry. This enhances the operator's situational awareness ability, enables a more intuitive experience [13], makes it easier to control and navigate drones [14], and enables drones to perform more complex tasks.

We designed the air–ground integrated intelligent cognition visual enhancement system (VisionICE), combining AR and UAV technology according to the actual demand. The system relies on the wireless camera on the helmet and UAV to survey and shoot the target from ground and air perspectives. The cloud server recognizes and detects the returned video in real time. Finally, the AR display device receives the results of intelligent cognition and precise positioning. In real time, searchers can obtain the target recognition results and the visual experience beyond reality from air and ground perspectives. In contrast to conventional post-disaster search and rescue operations, the proposed system boasts several advantages, including precise target recognition, extensive search range, high mobility, and a straightforward process. It effectively surmounts the limitations commonly associated with traditional search methods, such as reduced efficiency, restricted field of view, and suboptimal environmental adaptability. The primary contributions of this paper can be summarized as follows.

(1) Development of an air–ground integrated intelligent cognition visual enhancement system called VisionICE. This system utilizes wireless image sensors on a drone and camera-equipped helmet to simultaneously obtain air–ground perspective images, achieving efficient patrols on a large scale in particular environments to address the issues of low efficiency and limited search range in post-disaster search and rescue operations.
(2) Based on the YOLOv7 algorithm, object detection has been achieved in scenes such as highways, villages, farmland, mountains, and forests. In practical applications, YOLOv7 can accurately identify the target class, effectively locate the target position, and achieve a detection accuracy of up to 97% for interested targets. The YOLOv7 model has a detection speed of 40 FPS, which can meet the requirements of real-time target detection and provide reliable target recognition results for searchers.
(3) Utilizing portable AR intelligent glasses, real-time display of object detection results on the cloud server and onboard computer provides searchers with an immersive visual experience. This improves the situational awareness of search personnel by issuing a potential threat or anomaly alerts. Compared to traditional post-disaster search and rescue operations, VisionICE exhibits significantly strong interactivity, experiential capabilities, and versatility.

The organization of this paper is outlined as follows: Section 2 presents a review of the relevant literature. Section 3 delineates the research methodology employed in this study. Section 4 furnishes the experimental results and subsequent analysis of the proposed algorithm. Finally, Section 5 offers the concluding remarks.

2. Related Work

2.1. Drone Search and Rescue System

UAVs are becoming increasingly popular in search and rescue missions due to their ability to quickly and efficiently cover large areas and provide real-time situational awareness. By being equipped with various sensors such as cameras, thermal imaging equipment, and LiDAR, drones can detect and locate individuals or objects in areas that are difficult or dangerous for human rescuers to access.

In search and rescue missions, drones are typically used to search for missing persons or survivors in disaster areas and identify dangerous areas or obstacles that may pose a risk to rescue personnel. At the same time, drones can provide real-time situational awareness to help decision-making and coordinate rescue work [15,16]. In addition, drones can transport medical supplies, equipment, and personnel to remote or inaccessible areas. Martinez-Alpiste et al. [1] used drones and smartphones equipped with convolutional neural networks to achieve human detection. Yang et al. [2] utilized unmanned aerial vehicles and unmanned surface vehicles to collaborate for maritime search and rescue, and used reinforcement learning (RL) to achieve path planning. Gotovac et al. [5] utilized drones to pre-acquire aerial images, and then used convolutional neural networks to improve the efficiency and reliability of search and rescue. The problem with this method is that it cannot be detected in real time.

Compared to traditional search and rescue methods, the use of drones in search and rescue missions has several advantages. Firstly, drones can quickly and efficiently cover large areas, providing a broader perspective than traditional search methods [17]. Secondly, drones can operate in hazardous environments such as fires, floods, and earthquakes, reducing the risk of death and injury for search and rescue personnel. Thirdly, drones can provide real-time data and images for ground search and rescue personnel, enabling more effective decision-making and coordination. Therefore, to address the difficulty of on-site search and rescue personnel in specific environments, this article designs and implements a comprehensive search and rescue system (VisionICE) based on unmanned aerial vehicles. Compared with existing methods, the VisionICE system improves the detection accuracy and efficiency of targets in search and rescue operations, while also possessing real-time performance.

2.2. Target Detection Algorithm

Intelligent recognition and target detection are both concepts related to computer vision, artificial intelligence (AI), and machine learning. They are interrelated but serve different purposes in the processing and analysis of visual data. Intelligent recognition refers to the process of identifying and categorizing objects or patterns within an image or video by leveraging AI and machine learning algorithms. Intelligent recognition tasks can include object recognition, face recognition, character recognition, and more. Target detection is a specific application of intelligent recognition that focuses on identifying and locating specific objects within an image or video. Target detection may involve finding objects of interest, such as people, vehicles, or animals, among a complex background. It can be said that target detection is a subcategory of intelligent recognition, as it involves the identification and localization of specific objects within a given visual scene.

The target detection algorithms based on deep learning use deep neural networks to extract shallow and high-level features of images automatically. Figure 1 illustrates the development of target detection algorithms. Compared to the traditional algorithms with manually designed components, the new feature extraction method substantially improves the accuracy and speed of target detection. Moreover, for application scenarios with high environmental complexity, large data volume, and varied target scales, the performance advantages of deep learning algorithms are more prominent. Anchor-based object detection algorithms primarily encompass two categories: two-stage algorithms [18–22] and one-stage algorithms [23–30]. As a prototypical two-stage algorithm, Faster R-CNN [19] employs Region Proposal Networks (RPN) to supplant the window selection method gov-

erned by manually designed rules, thereby achieving a more efficient acquisition of feature region candidate bounding boxes. Subsequently, the algorithm conducts classification and positional information regression on these candidate frames. While the two-stage object detection algorithm exhibits high accuracy, its processing speed is relatively slow, posing a challenge for meeting real-time performance requirements in practical application contexts.

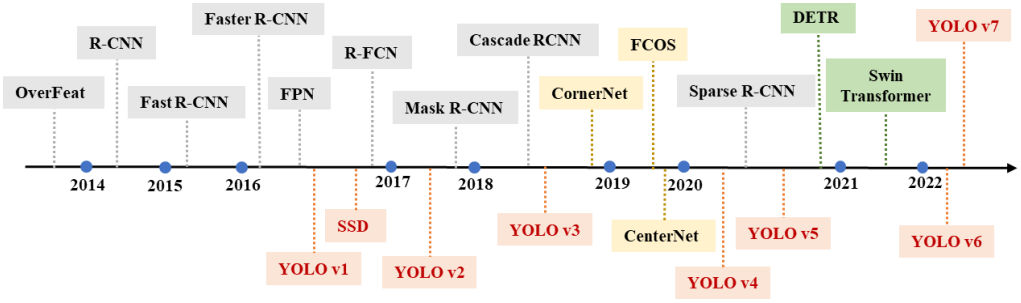

Figure 1. Target detection algorithm development history. The black, red, yellow, and green are two-stage, one-stage, anchor-free, and transformer-based target detection algorithms, respectively.

One-stage target detection algorithms, represented by the YOLO series [23,25–31], have witnessed the evolution of target detection in the era of deep learning. In contrast to object detection architectures based on candidate regions, the primary motivation behind employing a one-stage object detection algorithm with a regression-based architecture is to achieve enhanced speed and efficiency. This object detection model circumvents the extraction of candidate bounding boxes and predicts the target directly via an end-to-end methodology, converting object detection into regression prediction within a single network, thereby achieving a fundamental improvement in detection speed. YOLO's early efficiency improvement came at the expense of accuracy, due to its exclusive reliance on predicting the target bounding box on the terminal layer of the feature map. To ameliorate the localization accuracy of the regression architecture, SSD [24] uses multiple convolutional layers of different sizes for bounding box (bbox) prediction, significantly improving the localization accuracy of multi-scale targets. In addition, YOLOv2 [25] and YOLOv3 [26] also borrow the idea of Faster R-CNN and introduce an anchor box to improve the target localization accuracy.

With the improvement of the algorithm, YOLOv5 [31] has made more progress in performance with more balanced optimization of accuracy and speed. YOLOv7 [28] is the most advanced new target detector in the YOLO series. The E-ELAN module architecture designed in YOLOv7 [28] enables the framework to learn better. The E-ELAN module uses expand, shuffle, and merge cardinality to achieve the ability to continuously enhance the learning capability of the network without destroying the original gradient path. In addition, YOLOv7 [28] uses composite model scaling to balance running speed and detection accuracy, making it suitable for various computing devices. To detect and identify the targets in the air–ground view more accurately and meet the real-time requirements, we choose the current state-of-the-art YOLOv7 [28] model to achieve intelligent target detection and recognition.

2.3. Drone Augmented Reality Technology

In response to the swift progression of information technology, virtual reality and augmented reality have increasingly garnered attention within their respective fields. Virtual reality is a wholly established virtual environment that allows humans to enter a new world out of the existing environment. Augmented reality is developed from virtual reality and aims to enhance human capabilities, provide various auxiliary information for humans, become an important hub to communicate between individual humans and the information world, and connect the physical world with the information world more closely. According to the proportion of virtual and reality in a system, the system can be divided into four categories: real reality, augmented reality, enhanced virtualization, and

virtual reality, as shown in Figure 2. In turn, augmented reality and enhanced virtualization can be collectively called hybrid reality.

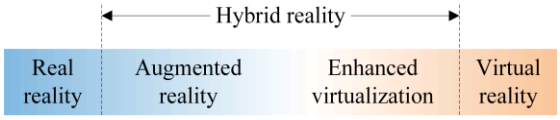

Figure 2. The relationship between augmented reality, virtual reality, and hybrid reality.

Augmented reality represents a technological approach that merges the physical world with virtual information through real-time computational processing and multi-sensor fusion [31–37]. This technology superimposes virtual information on factual data and provides an interface for human–digital world interaction [9,38]. With the development of new information technology, head-mounted augmented reality devices (e.g., AR glasses) have become the most mainstream augmented reality devices in manufacturing [39,40]. Hietanen et al. [41] designed a shared spatial model for human–machine cooperation based on depth perception to show the workspace of a fixed-frame robotic arm to the operator in real-time through 3D head-mounted AR glasses. Jost et al. [42] informed users of the location of their peripheral out-of-sight robots through AR glasses.

In recent years, researchers and practitioners have explored the use of AR technology to enhance the capabilities of drones, enabling them to perform complex tasks and providing real-time situational awareness to operators. The combination of AR technology and drones brings several advantages. Firstly, it enables operators to have a more intuitive and immersive experience, making it easier to control and navigate drones. Secondly, it enhances the situational awareness of operators, enabling them to make wiser decisions and respond quickly to constantly changing conditions. Thirdly, it enables drones to perform more complex tasks such as aerial inspections, surveying, and measurement. Kikuchi et al. [11] combined AR technology and drones, and a city digital twin method with an aerial perspective has been developed to avoid occlusion issues. Huuskonen et al. [12] determined the location of soil samples using aerial images captured by drones and guided users to the sampling point using AR intelligent glasses. Liu et al. [13] utilized AR devices to interact with autonomous drones and explore the environment. Erat et al. [14] used AR technology to obtain an external center view to help drones have stronger spatial understanding results in hazardous areas.

Drone AR technology has the advantage of virtual–real integration, which can help operators to improve the field environment perception, reduce the human brain workload and information processing stress, and provide operators with experiences beyond the real-world perception. Therefore, this paper obtains target detection results and rich visual enhancement experience in real time based on wireless AR smart glasses. Combining this with drone technology provides enhanced situational awareness for operators and enables drones to perform more complex tasks.

3. Our Approach

In developing the VisionICE system, the environment and usage habits are fully considered, and the system has strong reliability, operability, and integrity. Figure 3 illustrates the workflow of the system. VisionICE system is based on a UAV and a camera-equipped helmet, and it surveys the inspection area from ground and air perspectives. For the first perspective, the operator defines the inspection region from within the ground control station (GCS). The ground control station subsequently formulates a flight path based on the demarcated area. It transmits it to the flight platform, executing the patrol following the predetermined route. For the second, the operator wears a helmet with a camera and films a ground-view video to assist the human eye in achieving a ground search. The cloud server detects and identifies the dual-view image information in real time. Finally, it projects the target intelligent cognitive results onto AR glasses to create a hybrid

visually enhanced intelligent cognitive monitoring screen, which lays the foundation for subsequent decisions such as target tracking.

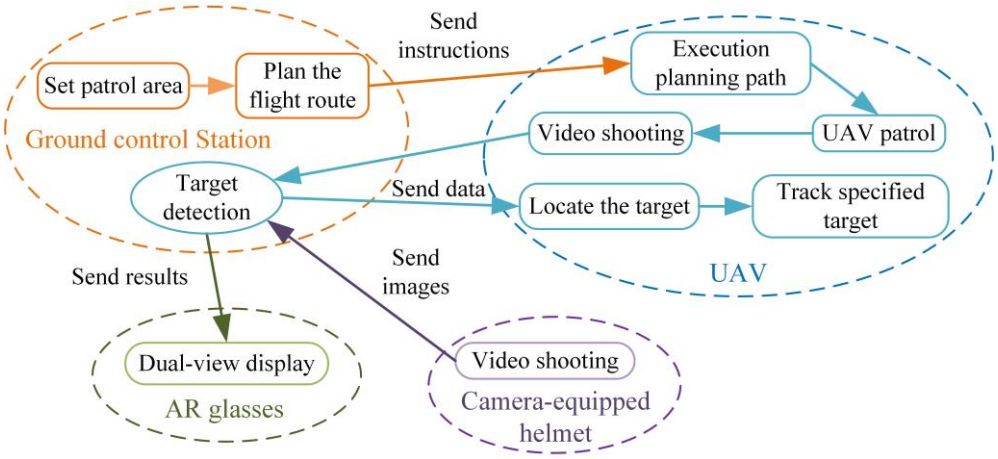

Figure 3. The VisionICE system workflow chart.

3.1. Hardware Framework

The component models of the VisionICE system are shown in Table 1. The hardware architecture of the VisionICE system primarily comprises four components: a UAV, camera-equipped helmet, portable AR display device, and cloud server, as shown in Figure 4. Among them, the helmet and UAV have image acquisition equipment, which can survey and film the target from ground and air viewpoints. The cloud server calculates and processes the returned video in real time. The portable AR display device selects BT-300 AR smart glasses to display intelligent cognitive results with hybrid visual augmentation.

Table 1. The component list of system hardware.

Systems	Component List	Specification
S500 Quadrotor UAV	Flight Controller	Pixhawk 2.4.8
	Electronic Speed Control	XXD-40A
	Motor	QM3507-680KV
	Remote Control	AT9S
	Digital Transmission Module	3DR V5 Radio
	Image Transmission Module	R2TECK-DVL1
	GPS Module	GPS M8N
	Sonar Obstacle Avoidance Module	RCWL-1605
	Power Supply System	4S Lithium Cell
	Onboard Computer	Jetson Xavier NX
	PTZ Camera	FIREFLY 8s
Helmet	Camera	IP Camera
	AR Glasses	Epson MOVERIO BT-300

3.1.1. UAV System Components

The hardware components of the UAV encompass a remote control, sonar module, GPS module, and Pan-tilt-Zoom (PTZ) camera, as illustrated in Figure 5. The UAV employs a Pixhawk flight controller, and the ground control station oversees and manages the UAV's flight to achieve target detection and identification. The UAV completes communication between the flight control system and the ground control station through the digital transmission module. The UAV uses the GPS module and the Pixhawk flight control system to achieve positioning and navigation. The UAV uses the electric PTZ to install an image acquisition module. The PTZ extends the search range of the image sensor while maintaining a stable picture. The image sensor has a maximum frame rate of 120 FPS and supports up to 1600 W pixels. The images are processed by the onboard computer.

The onboard computer is equipped with the kernel correlation filter (KCF) target tracking algorithm to realize the single target tracking function of the UAV.

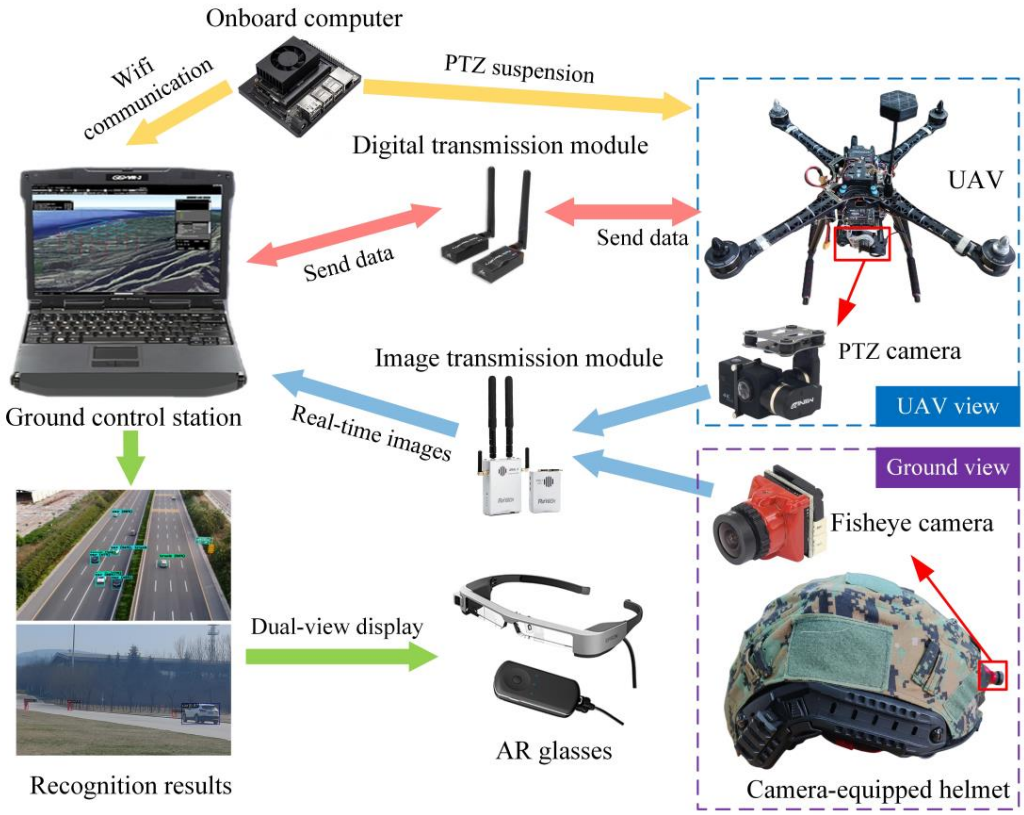

Figure 4. Hardware architecture of the VisionICE system.

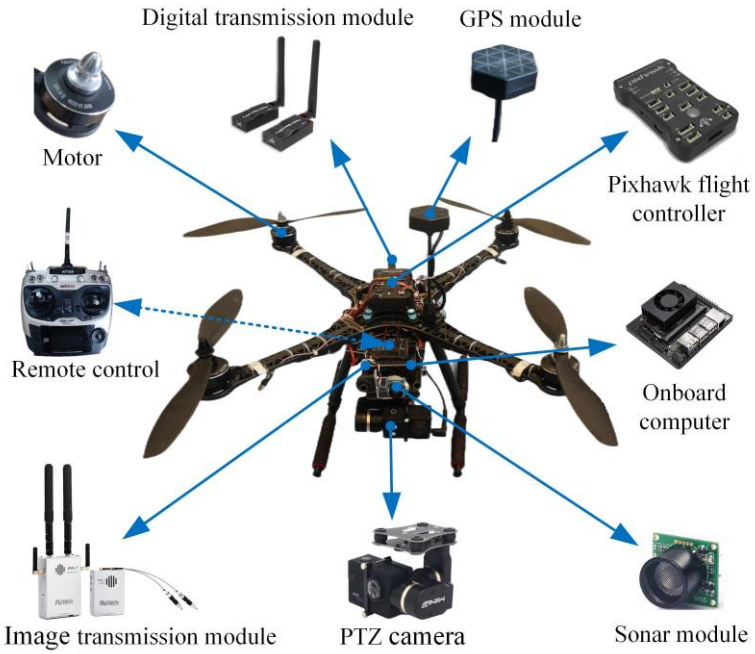

Figure 5. Hardware structure of the UAV system.

3.1.2. Camera-Equipped Helmet

The camera-equipped helmet has a lightweight image acquisition module because of its portability and range stability. The image acquisition module uses a webcam with a volume of 25 × 25 × 30 mm^3 and a weight of 30 g. Its maximum frame rate is 25 FPS, the

wide-angle level is 95 degrees, and the diagonal is 120 degrees. It supports wireless CMOS image transmission and can achieve real-time communication of images at a distance. Figure 6 depicts the camouflaged helmet equipped with the image acquisition module.

Figure 6. The camera-equipped helmet.

3.1.3. AR Smart Glasses

Epson MOVERIO BT-300 AR smart glasses are used as the remote display device for intelligent cognitive results to meet the actual use requirements of outdoor environments, as shown in Figure 7. The size of BT-300 AR glasses is only 178 × 191 × 25 mm^3, and the weight is 69 g, which reaches the level of lightweight wearable devices. BT-300 AR glasses are wirelessly connected and can interact with cloud computing devices for remote display. With no external power supply, the standby time of BT-300 AR glasses reaches 6 h. Compared to head-mounted AR display devices, the BT-300 AR glasses are small, light in weight, and have a long standby time, significantly reducing user interference. As flight glasses, the most apparent advantage of Epson MOVERIO BT-300 AR glasses is to avoid frequent switching between UAV, remote control, and ground view. BT-300 AR glasses significantly improve takeoff, landing, and low-altitude flight safety.

Figure 7. The BT-300 AR smart glasses.

3.2. Software Framework

The software components of the VisionICE system primarily consist of a UAV navigation control module, a target recognition module, and a multi-process information communication module. The UAV navigation control module is responsible for managing the attitude and position of the UAV, in addition to planning its patrol area. The target recognition module primarily focuses on target detection and identification. The multi-process information communication module chiefly ensures the coordination of information sharing among various system components, preventing process congestion from leading to software failures. Figure 8 depicts the overview of the system software architecture.

3.2.1. UAV Navigation Control Module

The operation of UAVs can be managed via remote control devices or ground control stations (GCS). The GCS serves as the primary interface between the UAV and its operators. It is responsible for monitoring, controlling, and managing the UAV's flight operations, providing real-time communication, telemetry data, and mission planning capabilities. The GCS features a user-friendly interface that allows operators to input commands, monitor the UAV's status, and visualize its position on a map. Upon the operator's demarcation of the patrol region, the GCS autonomously devises a flight trajectory. This flight plan is

subsequently conveyed to the UAV through a digital transmission apparatus, facilitating autonomous navigation along the prearranged path. When the UAV detects a search target, the GCS' detection and identification software displays the target's category and location and issues a "target found" alert.

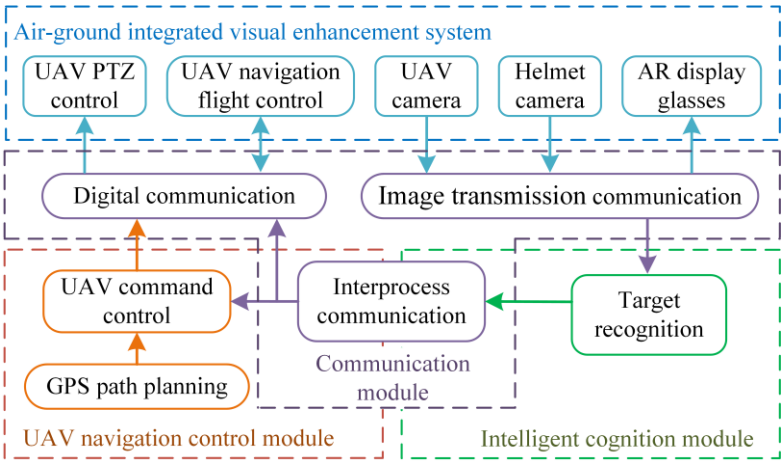

Figure 8. Overview of the system software architecture.

3.2.2. Target Recognition Module

The VisionICE system utilizes the state-of-the-art YOLOv7 object detection model. YOLOv7 has enhanced target detection accuracy through improved network modules and optimization methods. YOLOv7 increases the training cost but not the inference cost, and has a faster detection speed. Figure 9 illustrates the structure of the YOLOv7 network. First, the input image is resized to 640 × 640. Next, it is input into the Backbone network. Then the Head layer generates three feature maps with different sizes. Finally, the RepConv outputs the outcomes of the prediction.

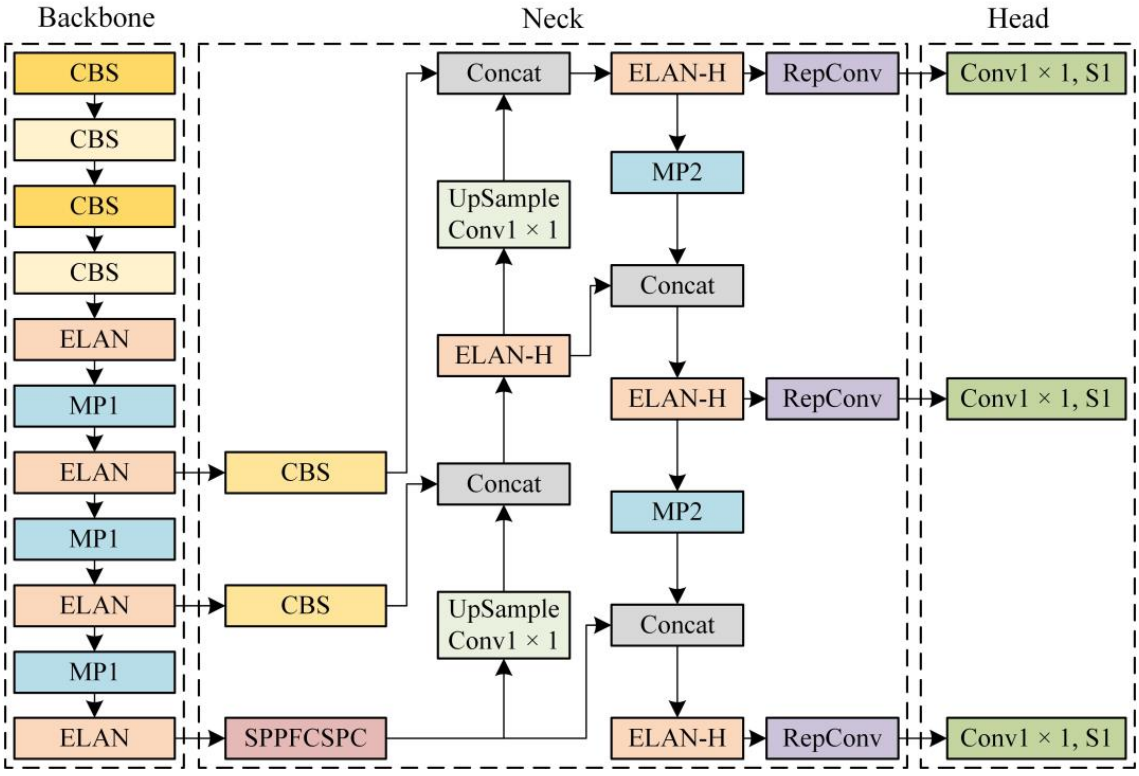

Figure 9. The YOLOv7 network structure.

The Backbone of YOLOv7 has a total of 50 layers. After four convolutional layers, followed by the ELAN module, and finally, three MPs and ELAN, the feature outputs of three scales are obtained. Among them, ELAN comprises several CBSs, including Conv, BN, and SiLU. The ELAN module's input and output feature size are kept constant. The number of channels changes in the beginning, with two CBSs, and the latter several input channels are kept the same as the output channels. The primary components of the MP layer are Maxpool and CBS, and the distinction between MP1 and MP2 is the number of channels.

The Neck of YOLOv7 is similar to YOLOv4 and YOLOv5, which is a PAFPN structure. First, it downsampled the output feature map of Backbone 32 times to obtain the feature map C5. Then after SPPCSPC, the number of channels is changed from 1024 to 512. The output of the Backbone is first fused according to top-down and C4 and C3 to obtain P3, P4, and P5. Secondly, it integrates the bottom-up with P4 and P5. In contrast to YOLOV5, YOLOV7 substitutes the downsampling layer for the MP2 layer and the CSP module for the ELAN-H module. For P3, P4, and P5 outputs from PAFPN, the number of channels is adjusted through RepConv. Finally, 1×1 convolution is used to predict objectness, class, and bbox. RepConv has the summation output of the three branches during training, and the parameters of the branches are re-parameterized to the main branch during inference.

YOLOv7 initially partitions the input image into uniformly sized $S \times S$ grid cells. It detects and classifies the detected objects if their centers fall into the grids. Each grid predicts B bounding boxes and confidence levels. Pre-set thresholds filter multiple bounding boxes to remove those with lower confidence levels. Finally, NMS is used to obtain the final detection and classification results.

YOLOv7 processes target region regression and region classification in parallel, where the target region regression contains two parameters: target center coordinates (x_i, y_i) and target size (w_i, h_i), both of which are relative quantities based on image size, ranging between $[0, 1]$. Set B to the number of prediction targets in a single grid and C to the target category. The total output of the output layer is an $S \times S \times (5B + C)$ order tensor where $5B$ contains the position, size, and probability of the target, and C represents the probability of each category in the grid. The loss function expression of the network is as follows.

$$\begin{aligned} L = {} & \lambda_{co} \sum_{i=1}^{s^2} \sum_{j=1}^{B} 1_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\ & + \lambda_{co} \sum_{i=1}^{s^2} \sum_{j=1}^{B} 1_{ij}^{obj} \left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right] \\ & + \sum_{i=1}^{s^2} \sum_{j=1}^{B} 1_{ij}^{obj} (C_i - \hat{C}_i)^2 + \lambda_{noobj} \sum_{i=1}^{s^2} \sum_{j=1}^{B} 1_{ij}^{noobj} (C_i - \hat{C}_i)^2 \\ & + \sum_{i=1}^{s^2} 1_i^{obj} \sum_{c \in \text{classes}}^{B} (p_i(c) - p_i(\hat{c}))^2 \end{aligned} \tag{1}$$

where λ_{co} denotes the coordinate weight, and $(\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i)$ represents the predicted target position and size. 1_{ij}^{obj} denotes the indicator function for the j prediction of the grid i as the target, the corresponding 1_{ij}^{noobj} denotes its indicator function for negative examples, and λ_{noobj} is the penalty factor for negative samples in the prediction results. The network weight is obtained by minimizing the loss function used to train the network. Then predict each category's probability and target location of each grid. Combining the two results, the target location and category with the highest likelihood in each grid are finally output.

Take the vehicle and pedestrian targets as examples to train the YOLOv7 model. The training process is based on the PyTorch deep learning framework. The server used in the training process has two Intel Xeon Gold 6230 (2.1 GHz/20C/27.5ML3) CPUs and two NVIDIA RTX 8000 GPUs. To better obtain the characteristics of the data and improve the performance and generalization ability of the model, the cosine annealing algorithm is used for model training. The model training process sets Epoch to 300, batch size to 8, learning

rate to 0.001, and attenuation coefficient to 0.0005. The loss value curve of the model is shown in Figure 10.

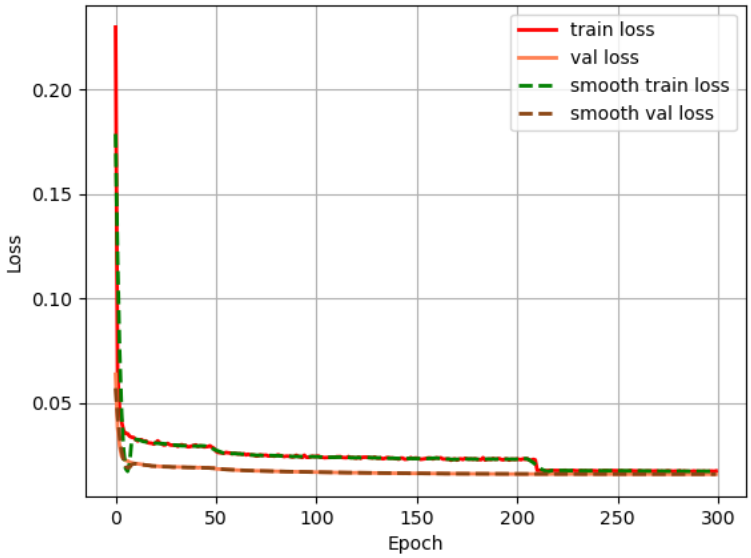

Figure 10. The loss curve of YOLOv7.

The experiment selected the mean average precision (mAP) and frames per second (FPS) as indicators to evaluate the performance of the YOLOv7 algorithm. The experimental results show that when IoU is set to 0.50, the trained mAP of the YOLOv7 model can reach 96.3%, with high target detection accuracy. When IoU increases from 0.5 to 0.95 in steps of 0.05, the mAP of the YOLOv7 model is 28.9%. The YOLOv7 model has a detection speed of 40 FPS, which can meet the requirements of real-time object detection.

3.2.3. Multi-Process Information Communication Module

Due to the complexity and unknown nature of the external environment of the Vision-ICE system, the data transmission and communication link of the whole system adopts a wireless connection, as shown in Figure 11. In the experimental and test state, the cloud computing module first acquires real-time data from the image acquisition module. Then the analyzed and processed data have to be projected onto the display device. There is a large amount of data exchange in these two links. Therefore, the quality of the data wireless communication module determines the realization and performance of the whole system function. From the system's real-time economy and portability perspective, a MERCURY wireless router is used to build the local area network (LAN). The multi-process information communication module plays a vital role in enhancing the overall robustness, efficiency, and scalability of a software system by effectively managing and coordinating inter-process communication and synchronization.

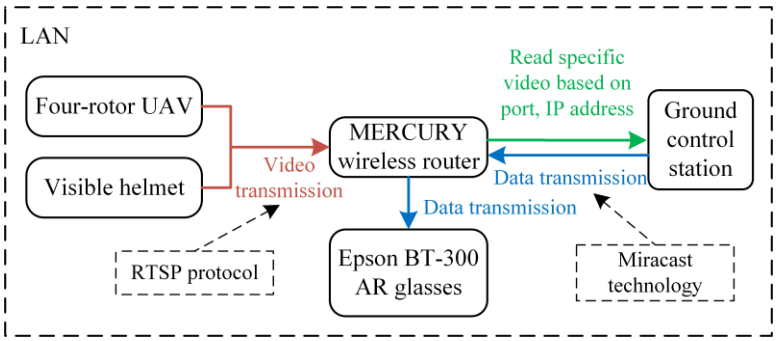

Figure 11. System wireless communication design solution.

4. Experiments and Analysis

4.1. Drone Search Flight Test

In order to thoroughly assess the feasibility and applicability of the VisionICE system, we conducted a UAV target detection assessment and a single-target tracking flight examination under authentic circumstances in this paper. Firstly, we chose a road scenario to preliminarily verify drone target detection and single-target tracking flight. The ability to detect vehicles, pedestrians, and other obstacles in real-time in highway scenes can effectively verify the real-time effectiveness of unmanned aerial vehicle systems. Due to the high speed of vehicles, different lighting conditions, and the presence of occlusion and complex backgrounds, highway scenes are particularly challenging for object detection algorithms. Object detection algorithms must be able to accurately recognize and track objects in these challenging environments to ensure the reliability of the VisionICE system. In addition, object detection by drones in highway scenes can collect real-time information about unexpected traffic accidents, which is of great significance for the search and rescue of traffic accidents.

To verify the effectiveness and real-time performance of drone target detection, we experimented on the highway to verify the effectiveness of UAV target detection and tracking flight. Figures 12 and 13 depict the detection and recognition results of the UAV flight at the height of 30 m in various highway scenarios.

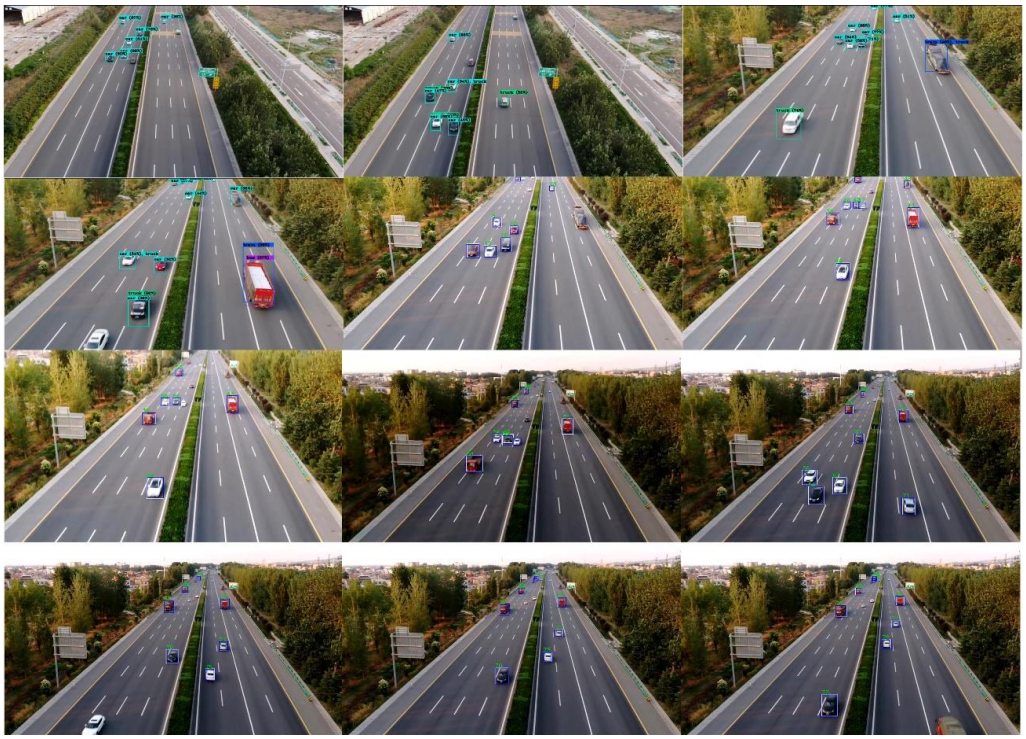

Figure 12. Vehicles detection results for scenario 1.

Figure 12 shows that the algorithm implemented in the present research exhibits superior real-time performance during practical scene assessments, as well as a noteworthy detection accuracy in small target detection. Figure 13 shows that the algorithm used in this paper has high accuracy in the actual scene test and correctly classifies and locates the vehicles and pedestrians.

Secondly, we selected the village farmland scene for single-target tracking flight testing of unmanned aerial vehicles. Due to the vast farmland scene in villages and the lack of obstacles, unmanned aerial vehicle search and rescue methods are very applicable. Drones can be used to locate and track individual targets, such as lost hikers or trapped farmers, and then guide search and rescue teams to that location using real-time images and GPS

coordinates. This can greatly reduce the time and resources required of search and rescue personnel in remote areas, and improve the safety and effectiveness of search and rescue teams. The single-target tracking flight test results in chronological order are shown in Figure 14.

Figure 13. Target detection results for scenario 2.

Figure 14. Single-target tracking flight results.

The algorithm used in this paper can successfully locate the critical target with high accuracy in the single-target tracking flight test. At the same time, the UAV can track the crucial target in real-time, and the flight process is very smooth, and it does not show significant oscillation.

In addition, drones can quickly and efficiently cover large areas of mountains and forests in search and rescue missions, especially in mountainous terrain where ground-based search effects may be difficult or dangerous. Drones can provide detailed images

of terrain and surrounding areas to detect and locate specific targets in forests, such as animals or humans. This can help rescue personnel locate and rescue people lost or injured in the forest. The search and rescue results of the VisionICE system are shown in Figure 15.

Figure 15. Search and rescue results in mountainous and forest scenarios.

From the recognition results, it can be seen that the YOLOv7 object detection algorithm can accurately locate and recognize small targets, with a maximum detection accuracy of 97%. Moreover, this algorithm is robust to changes in target behavior, enabling accurate detection of targets in different directions and postures. In addition, it can also accurately detect targets under changing lighting and shrub occlusion conditions, with high accuracy and robustness.

4.2. VisionICE System Function Display

The VisionICE system aims to use drones and camera-equipped helmets as the main data collection platform, and AR glasses as the intelligent cognitive result visualization platform. Camera-equipped helmets can protect the safety of search and rescue personnel and also capture ground video for transmission to cloud servers. The detection and recognition results of video data can assist the human eye in the ground target recognition, avoiding false positives caused by artificial subjective speculation. Drones can detect and track targets of interest within the patrol area from the air. Once the system detects a target, the ground control station will issue an alarm, allowing the drone to approach the target using a remote control or independently track the target using its target tracking algorithm. AR glasses can display real-time object detection results on cloud servers and onboard computers, and provide an augmented reality visual experience for search and rescue personnel. In addition, AR glasses prevent operators from frequently lowering their heads and raising their heads to control drones, facilitating the operation process and reducing the possibility of mistakes during operation.

The system workflow is illustrated in Figure 16. Initially, the VisionICE system delineates the patrol region via the ground control station, transmitting pertinent flight commands to the UAV through digital communication channels to govern its motion. Subsequently, the image acquisition apparatus onboard the UAV captures real-time visual data of the patrolled area. At the same time, the helmet captures the ground's visual image

information in real time. Third, the visual image signals obtained from the air and ground viewpoints are transmitted through the graphical transmission module. Fourth, the ground control station receives the video signal and then detects and identifies the target in the field of view in real time. Fifth, AR smart glasses project intelligent cognition results in real time. The searcher can obtain the visual enhancement effect of hybrid viewpoints.

Figure 16. The VisionICE system workflow. (**a**) UAV takes off and patrols according to the specified route. (**b**) UAV detects vehicle. (**c**) UAV detects pedestrians.

The VisionICE system provides a new solution for post-disaster search and rescue tasks by integrating drones, camera-equipped helmets, intelligent cognitive algorithms, and AR technology. This method can search designated areas in real-time and from multiple perspectives, providing valuable insights for search and rescue missions and other applications. In addition, the system's use of AR smart glasses enhances the searcher's situational awareness by overlaying intelligent cognitive results, further improving the efficiency and effectiveness of the search process. The workflow and functionality of this system demonstrate its potential to revolutionize object detection and tracking in various fields.

5. Conclusions

We design an air–ground integrated intelligent cognition visual enhancement system (VisionICE) based on UAVs, camera-equipped helmets, and AR glasses in this paper. The combination of helmets and drones enables operators to have both ground and air perspectives, and the use of AR glasses improves the operator's situational awareness ability. By using the YOLOv7 algorithm, the accuracy of object detection can reach 97% in scenarios such as highways, villages, farmland, and forests, achieving real-time object detection of 40 FPS. The VisionICE system improves the scope and efficiency of search and rescue, solves the problem of personnel being unable to search in special environments, and has the advantages of diverse fields of view, accurate recognition, rich visual experience,

wide application scenarios, high intelligence, and convenient operation. However, the use of the VisionICE system in search and rescue operations also faces some challenges. The challenges in terms of drones include regulatory issues such as obtaining necessary permits and complying with airspace restrictions, as well as technical challenges such as ensuring the reliability and durability of drones and their components. In addition, accurate and reliable sensor data are also needed, as well as the development of user-friendly AR interfaces and software to effectively integrate with drone hardware and control systems. Future applications of the system include battlefield surveillance, firefighting, post-disaster search and rescue, criminal investigations, anti-terrorism and peacekeeping, and many others.

Author Contributions: Conceptualization, X.Y., R.L. and Q.L.; Methodology, R.L. and Q.L.; Software, S.W. and Z.Q.; Investigation, Z.Q. and J.F.; Resources, X.Y. and R.L.; Writing—original draft preparation, J.F. and Q.L.; Writing—review and editing, R.L. and J.F.; Visualization, S.W. and Z.Q.; Supervision, X.Y. and R.L.; Project administration, X.Y. and R.L.; Funding acquisition, X.Y. and R.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 62276274, and the Aeronautical Science Fund under Grant 201851U8012.

Data Availability Statement: The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

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

References

1. 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]
2. Yang, T.; Jiang, Z.; Sun, R.; Cheng, N.; Feng, H. Maritime search and rescue based on group mobile computing for unmanned aerial vehicles and unmanned surface vehicles. *IEEE Trans. Ind. Inform.* **2020**, *16*, 7700–7708. [CrossRef]
3. Wang, Y.; Liu, W.; Liu, J.; Sun, C. Cooperative USV–UAV marine search and rescue with visual navigation and reinforcement learning-based control. *ISA Trans.* **2023**. [CrossRef]
4. Mcgee, J.; Mathew, S.J.; Gonzalez, F. Unmanned aerial vehicle and artificial intelligence for thermal target detection in search and rescue applications. In Proceedings of the 2020 International Conference on Unmanned Aircraft Systems (ICUAS), Athina, Greece, 1–4 September 2020; pp. 883–891.
5. Gotovac, S.; Zelenika, D.; Marušić, Ž.; Božić-Štulić, D. Visual-based person detection for search-and-rescue with uas: Humans vs. machine learning algorithm. *Remote Sens.* **2020**, *12*, 3295. [CrossRef]
6. Lu, R.; Yang, X.; Jing, X.; Chen, L.; Fan, J.; Li, W.; Li, D. Infrared small target detection based on local hypergraph dissimilarity measure. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 1–5. [CrossRef]
7. Lu, R.; Yang, X.; Li, W.; Fan, J.; Li, D.; Jing, X. Robust Infrared Small Target Detection via Multidirectional Derivative-Based Weighted Contrast Measure. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 1–5. [CrossRef]
8. Wang, S.; Jiang, F.; Zhang, B.; Ma, R.; Hao, Q. Development of UAV-based target tracking and recognition systems. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 3409–3422. [CrossRef]
9. Wang, X.V.; Wang, L. Augmented reality enabled human–robot collaboration. In *Advanced Human-Robot Collaboration in Manufacturing*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 395–411.
10. Unal, M.; Bostanci, E.; Sertalp, E.; Guzel, M.S.; Kanwal, N. Geo-location based augmented reality application for cultural heritage using drones. In Proceedings of the 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey, 19–21 October 2018; pp. 1–4.
11. Kikuchi, N.; Fukuda, T.; Yabuki, N. Future landscape visualization using a city digital twin: Integration of augmented reality and drones with implementation of 3D model-based occlusion handling. *J. Comput. Des. Eng.* **2022**, *9*, 837–856. [CrossRef]
12. Huuskonen, J.; Oksanen, T. Soil sampling with drones and augmented reality in precision agriculture. *Comput. Electron. Agric.* **2018**, *154*, 25–35. [CrossRef]
13. Liu, C.; Shen, S. An augmented reality interaction interface for autonomous drone. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; pp. 11419–11424.
14. Erat, O.; Isop, W.A.; Kalkofen, D.; Schmalstieg, D. Drone-augmented human vision: Exocentric control for drones exploring hidden areas. *IEEE Trans. Vis. Comput. Graph.* **2018**, *24*, 1437–1446. [CrossRef]
15. Valsan, A.; Parvathy, B.; Gh, V.D.; Unnikrishnan, R.S.; Reddy, P.K.; Vivek, A. Unmanned aerial vehicle for search and rescue mission. In Proceedings of the 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Tirunelveli, India, 15–17 June 2020; pp. 684–687.

16. Kulkarni, S.; Chaphekar, V.; Chowdhury, M.M.U.; Erden, F.; Guvenc, I. Uav aided search and rescue operation using reinforcement learning. In Proceedings of the 2020 SoutheastCon, Raleigh, NC, USA, 11–15 March 2020; pp. 1–8.
17. Silvagni, M.; Tonoli, A.; Zenerino, E.; Chiaberge, M. Multipurpose UAV for search and rescue operations in mountain avalanche events. *Geomat. Nat. Hazards Risk* **2017**, *8*, 18–33. [CrossRef]
18. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 1440–1448.
19. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*; Curran Associates: Montreal, Canada, 2015; pp. 1–9.
20. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
21. Cai, Z.; Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6154–6162.
22. Sun, P.; Zhang, R.; Jiang, Y.; Kong, T.; Xu, C.; Zhan, W.; Luo, P. Sparse r-cnn: End-to-end object detection with learnable proposals. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14454–14463.
23. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
24. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
25. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
26. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
27. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Wei, X. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv* **2022**, arXiv:2209.02976.
28. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.
29. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
30. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
31. Liu, W.; Quijano, K.; Crawford, M.M. YOLOv5-Tassel: Detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8085–8094. [CrossRef]
32. 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]
33. 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]
34. Xia, X.; Xiong, L.; Huang, Y.; Lu, Y.; Gao, L.; Xu, N.; Yu, Z. Estimation on IMU yaw misalignment by fusing information of automotive onboard sensors. *Mech. Syst. Signal Process.* **2022**, *162*, 107993. [CrossRef]
35. Liu, W.; Xia, X.; Xiong, L.; Lu, Y.; Gao, L.; Yu, Z. Automated vehicle sideslip angle estimation considering signal measurement characteristic. *IEEE Sens. J.* **2021**, *21*, 21675–21687. [CrossRef]
36. Xiong, L.; Xia, X.; Lu, Y.; Liu, W.; Gao, L.; Song, S.; Yu, Z. IMU-based automated vehicle body sideslip angle and attitude estimation aided by GNSS using parallel adaptive Kalman filters. *IEEE Trans. Veh. Technol.* **2020**, *69*, 10668–10680. [CrossRef]
37. Xia, X.; Meng, Z.; Han, X.; Li, H.; Tsukiji, T.; Xu, R.; Ma, J. Automated Driving Systems Data Acquisition and Processing Platform. *arXiv* **2022**, arXiv:2211.13425.
38. Oztemel, E.; Gursev, S. Literature review of Industry 4.0 and related technologies. *J. Intell. Manuf.* **2020**, *31*, 127–182. [CrossRef]
39. Baroroh, D.K.; Chu, C.-H.; Wang, L. Systematic literature review on augmented reality in smart manufacturing: Collaboration between human and computational intelligence. *J. Manuf. Syst.* **2021**, *61*, 696–711. [CrossRef]
40. Siew, C.Y.; Ong, S.-K.; Nee, A.Y. A practical augmented reality-assisted maintenance system framework for adaptive user support. *Robot. Comput.-Integr. Manuf.* **2019**, *59*, 115–129. [CrossRef]
41. Hietanen, A.; Pieters, R.; Lanz, M.; Latokartano, J.; Kämäräinen, J.K. AR-based interaction for human-robot collaborative manufacturing. *Robot. Comput.-Integr. Manuf.* **2020**, *63*, 101891. [CrossRef]
42. Jost, J.; Kirks, T.; Gupta, P.; Lünsch, D.; Stenzel, J. Safe human-robot-interaction in highly flexible warehouses using augmented reality and heterogenous fleet management system. In Proceedings of the 2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR), Shenyang, China, 24–27 August 2018; pp. 256–260. [CrossRef]