



## Article

# Dual-View Stereovision-Guided Automatic Inspection System for Overhead Transmission Line Corridor

Yaqin Zhou <sup>1,2</sup>, Chang Xu <sup>1</sup>, Yunfeng Dai <sup>3</sup>, Xingming Feng <sup>3</sup>, Yunpeng Ma <sup>1,2</sup> and Qingwu Li <sup>1,2,\*</sup><sup>1</sup> College of Internet of Things Engineering, Hohai University, Changzhou 213022, China<sup>2</sup> Jiangsu Key Laboratory of Power Transmission and Distribution Equipment Technology, Hohai University, Changzhou 213022, China<sup>3</sup> State Grid Yancheng Power Supply Company, Yancheng 224000, China

\* Correspondence: 20031644@hhu.edu.cn

**Abstract:** Overhead transmission line corridor detection is important to ensure the safety of power facilities. Frequent and uncertain changes in the transmission line corridor environment requires an efficient and autonomous UAV inspection system, whereas the existing UAV-based inspection systems has some shortcomings in control model and ground clearance measurement. For one thing, the existing manual control model has the risk of striking power lines because it is difficult for manipulators to judge the distance between the UAV fuselage and power lines accurately. For another, the ground clearance methods based on UAV usually depend on LiDAR (Light Detection and Ranging) or single-view visual repeat scanning, with which it is difficult to balance efficiency and accuracy. Aiming at addressing the challenging issues above, a novel UAV inspection system is developed, which can sense 3D information of transmission line corridor by the cooperation of the dual-view stereovision module and an advanced embedded NVIDIA platform. In addition, a series of advanced algorithms are embedded in the system to realize autonomous control of UAVs and ground clearance measurement. Firstly, an edge-assisted power line detection method is proposed to locate the power line accurately. Then, 3D reconstruction of the power line is achieved based on binocular vision, and the target flight points are generated in the world coordinate system one-by-one to guide the UAVs movement along power lines autonomously. In order to correctly detect whether the ground clearances are in the range of safety, we propose an aerial image classification based on a light-weight semantic segmentation network to provide auxiliary information categories of ground objects. Then, the 3D points of ground objects are reconstructed according to the matching points set obtained by an efficient feature matching method, and concatenated with 3D points of power lines. Finally, the ground clearance can be measured and detected according to the generated 3D points of the transmission line corridor. Tests on both corresponding datasets and practical 220-kV transmission line corridors are conducted. The experimental results of different modules reveal that our proposed system can be applied in practical inspection environments and has good performance.

**Keywords:** automatic transmission line corridor inspection system; dual-view stereovision; power line detection and tracking; 3D reconstruction; aerial image classification; ground clearance measurement and detection



**Citation:** Zhou, Y.; Xu, C.; Dai, Y.; Feng, X.; Ma, Y.; Li, Q. Dual-View Stereovision-Guided Automatic Inspection System for Overhead Transmission Line Corridor. *Remote Sens.* **2022**, *14*, 4095. <https://doi.org/10.3390/rs14164095>

Academic Editors: Levente Klein, Andre Dias and Mohammad Awrangjeb

Received: 8 June 2022

Accepted: 18 August 2022

Published: 21 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



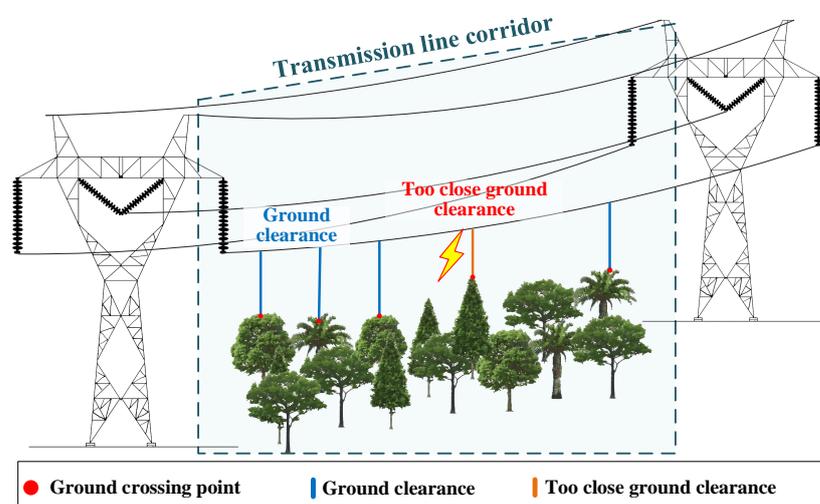
**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Overhead transmission lines are an important part of power delivery infrastructure [1–3]. The area of land and space occupied by the path of an overhead transmission line is defined as the line corridor. Within the line corridor, too close a distance between transmission lines and objects on the ground may cause arcs or short circuits, which seriously threaten the safety of power equipment and people [4,5]. A large number of electrical accidents are caused by growing plants, changing topography, large engineering machinery, and so on. In order to avoid these electrical accidents, transmission lines in the corridor area must be kept at a safe distance from objects on the ground. Therefore,

regular inspection of the overhead transmission line corridor is critical to a power system's safe operation.

As can be seen in Figure 1, ground clearance refers to the vertical distance between transmission line conductors and the corresponding objects on the ground, which is a key indicator to judge the security of the transmission line corridor. In addition, the safety of ground clearance should be judged by the voltage of the transmission line and the category of the ground object. Taking the 220 kV transmission line as an example, according to the relevant regulations in China, the minimum safe vertical clearance between power lines and buildings is 6 m, while that between power lines and trees is 3.5 m. Therefore, effective transmission line corridor detection requires three steps. Firstly, accurate spatial location of power lines and ground objects should be achieved. Secondly, the type of ground objects under the power lines should be determined accurately. Finally, the ground clearances between power lines and corresponding ground objects are measured to diagnose the safety of the transmission line corridor. It can be seen that inspection of the transmission line corridor is a very important but challenging job.



**Figure 1.** Schematic diagram of ground clearance of transmission corridors.

In the early stages, transmission line corridor inspection mainly relies on the fact that the staff directly carry out ground clearance measurement by some professional measuring tools such as altimeters or theodolites. However, manual inspection is time-consuming, inefficient, and dangerous because it requires the staff to stand near the underside of the transmission line and perform inspection tasks of power line locations, crossing object selection, and ground clearance measurement by visual observation [6]. With the continuous expansion of the grid scale, it is an inevitable trend to replace manual inspection with intelligent means [7]. In this paper, 'Intelligent means' refers to the use of advanced artificial intelligence technologies such as unmanned aerial technology, image processing, machine learning, deep learning, 3D reconstruction, and so on to carry out transmission line corridor inspection. By using intelligent means, some tasks can be performed by computers. For example, through 'intelligent means', staff can observe transmission lines from a distance with the cameras carried by the drones. Besides, power lines can be distinguished from the images by some algorithms rather than human eye observation. Intelligent means can, therefore, significantly improve the efficiency and quality of transmission line inspections.

At present, unmanned aerial vehicles (UAVs) have become one of the main tools for inspection of transmission lines due to their efficiency and flexibility [8,9]. The UAVs are equipped with aerial cameras or LiDARs and controlled by professional staff to collect information of the transmission line corridor for inspection. Though the UAV-based inspection method has significantly improved the efficiency of UAV transmission line inspection, it still has some shortcomings in both control mode and ground measurement in practical application.

**In terms of the control model.** UAV-based inspection technology is highly dependent on manual control. The common UAV-based inspection method needs manual control, which may threaten the safety of the equipment due to operational mistakes. It is quite a challenging task to control the UAV and successfully complete the transmission line detection. UAV operators should be fully familiar with transmission line environments to determine the correct flight path [10,11]. Besides, they also should have good enough operational skill to maintain a safe distance between the UAV and power lines; the long distance between the operator and transmission lines makes it difficult to gauge accurately by human vision. The misjudgment of distance may lead to the UAV colliding with power lines, threatening the safety of electrical equipment. As a result, the control model of UAV-based inspection needs to be further improved to replace manual operation. Introducing a vision-based autonomous inspection system is a potential approach to realize UAV automatic control [3]. Therefore, many studies of power line detection have been proposed for the auxiliary positioning of power lines [12,13]. For example, Zhang et al. [14] detected power lines by gray-level operator and prior knowledge from power line images with fixed orientation. Abdelfattah and Srikanth et al. [15,16] attempted to detect power lines and transmission towers by adopting the You Only Look At CoefficientTs (YOLACT) method. In spite of power line detection being extensively studied, they usually work well in preprocessed datasets, while they may perform poorly in real-world situations. Besides, few studies of power line 3D reconstruction and tracking have been conducted, where the latter is the important operation for automatic control. There is still no effective vision-based solution to achieve UAV automatic inspection for the transmission line corridor.

**In terms of the ground clearance measurement.** At present, UAV-based methods usually perform ground clearance measurement of transmission line corridors by using LiDAR. LiDAR has high accuracy but requires special offline software to process the large amount of original 3D point cloud data [17]. Inefficient offline processing and expensive cost seriously restrict its application and popularization. In addition, some researchers attempted to recover 3D information of the transmission line corridor by the UAV equipped with a monocular camera [14,18]. However, this is also time-consuming and difficult to implement because it requires a large number of repeatedly taken aerial images of the transmission line corridor. In conclusion, frequent and uncertain changes in the transmission line corridor environment requires a kind of real-time and low-cost inspection method. The existing means of ground clearance measurement for the UAV-based method do not satisfy the above conditions.

As the spatial distribution of different transmission lines and ground objects is very different, a single-binocular camera is difficult to meet the needs of multiple tasks in transmission line detection. Vision systems with multiple cameras have wide application in video-surveillance and autonomous navigation, which has attracted the attention of researchers. Therefore, adopting multiple stereo cameras with different views is the potential solution. Caron G. et al. [19] presented an intrinsic and extrinsic calibration approach for a hybrid stereo rig involving multiple central camera types. Strauß T. et al. [20] used binary patterns that surround each checkerboard to solve the association problem of checkerboard corners over time and between different cameras. Then, they adopted a sparse nonlinear least squares solver to estimate the optimal parameter set of multiple cameras with nonoverlapping fields of view. Figueiredo, R. et al. [21] adopted a reconfigurable multistereo camera system for Next-Best-View (NBV) planning and demonstrated advantages of multiple stereovision camera designs for autonomous drone navigation. In our previous work, we constructed a perpendicular double-baseline trinocular vision module by three cameras, and conducted theoretical exploration and verification of the power line 3D reconstruction and ground clearance measurement based on stereovision [22]. To further address the practical issues in transmission line corridor inspection analyzed above, a dual-view stereovision-guided automatic inspection system for the transmission line corridor is developed, inspired by these previous works above. Two stereo cameras of the proposed system have different views and subtasks. One set of stereo cameras is used to track power lines in the air and realize autonomous drone navigation. Another set of

stereo cameras is used to observe the ground objects under the power lines and assist in the ground clearance measurement. Besides, corresponding novel algorithms are proposed and applied to perform power line detection and tracking, automatic flight strategy formulation, and ground clearance measurement. The main contributions of this paper are presented as follows.

- **A novel automatic inspection system for the transmission line corridor is developed.** In this system, a dual-view stereovision module and an embedded NVIDIA platform are mounted on the UAV to perceive the surrounding environment. The dual-view stereovision module consists of two binocular cameras. One binocular camera is used to acquire the images of power lines and provide information for automatic flight. Another binocular camera is used to identify and locate the ground objects under power lines. The embedded NVIDIA platform is used to process information acquired by the dual-view stereovision module and achieve power line detection and reconstruction, aerial image classification, and ground clearance measurement. Different images from two views can be acquired and processed synchronously by the collaboration of the embedded NVIDIA platform and the dual-view stereovision module. Then, automatic inspection and real-time ground clearance measurement can be achieved simultaneously.
- **A real-time automatic flight strategy formulation method based on power line tracking is proposed.** We firstly propose an edge-assisted dual-refinement power line detection network, which can detect more high-level semantic cues and extract power lines from binocular images more accurately. Then, the detection results are utilized to locate power lines and calculate the distance between the UAV and power lines by binocular-vision-based 3D reconstruction processing. According to the calculated distance, the following flight points will be formulated one by one and used to control the automatic movement of the UAV along the power lines.
- **We present a ground clearance measurement strategy with ground object identification to accurately detect the safety of transmission line corridors.** In the proposed strategy, an aerial image classification method based on light-weight semantic segmentation network is proposed, which can classify ground objects effectively. After that, 3D points of ground objects are reconstructed by the advanced feature point matching method and concatenated with 3D points of corresponding power lines. Based on 3D points of the transmission line corridor, a plumb line between power lines and corresponding ground objects could be determined; then, the ground clearance could be calculated and detected without difficulty.

The rest of this paper is organized as follows. In Section 2, the architecture and information processing pipeline of the proposed dual-view stereovision-guided automatic transmission line corridor inspection system is elaborated in detail. Section 3 presents algorithms embedded the system, including power line detection, power line 3D reconstruction, automatic flight strategy formulation, ground object classification, and ground clearance measurement and detection. Section 4 demonstrates the proposed inspection system and experimental results. Section 5 briefly concludes this paper.

## 2. Automatic Transmission Line Corridor Inspection System Overview

### 2.1. Hardware Platform and System Architecture

As can be seen in Figure 2, the proposed automatic inspection system employs a refitted DJI Matrice 300 RTK (M300) <https://www.dji.com/cn/matrice-300/specs/>, (accessed on 17 August 2022) quadrotor as the platform. The detailed specifications of the proposed inspection platform are shown in Table 1. With the DJI guidance visual positioning system, the flight state of the UAV could be accurately estimated. The definition of visual positioning system can be seen in [23]. The DJI guidance visual positioning system can estimate positioning information by fusing the information of IMU, magnetic compass, cameras, and other sensors in the DJI UAV. The official positioning accuracy of the UAV is supplied in Table 1, including horizontal positioning accuracy—1 cm + 1 ppm, and vertical positioning

accuracy—1.5 cm + 1 ppm. It can be seen that the DJI M300 RTK UAV has high positioning accuracy. In this work, we focus on estimating the relative position relationship between the UAV and the power lines. Therefore, the UAV positioning information we used in the system is generated by the guidance visual positioning system of DJI UAV directly.



Figure 2. Overall framework of the proposed UAV-based inspection platform.

Table 1. Details of the proposed inspection platform.

Hardware Device	Hardware Model	Specification
UAV	DJI M300 RTK	Dimensions: 810(L) × 670(W) × 430(H) mm Max Takeoff Weight: 9 kg Max Wind Resistance: 15 m/s Max Flight Time: 55 min Max Speed: 17 m/s Hovering Accuracy: ±0.1 m Horizontal Positioning Accuracy: 1 cm + 1 ppm Vertical Positioning Accuracy: 1.5 cm + 1 ppm
Embedded Platform	Nvidia Jetson AGX Xavier developer kit	GPU: 512 Cores and 64 Tensor-Cores AI Performance: 32 Tops Memory: 32 GB 256-bit LPDDR4x 136.5 GB/s Power: 10 W   15 W   30 W Mechanical: 100 mm × 87 mm (L × W)
Dual-View Stereo Vision Module	LenaCV <sup>®</sup> CAM-AR0135-3T16 Lena Computer Vision Company	Upper binocular camera Resolution: 1280 × 960 pixels Frame Rate: 30 fps Focal Length: 25 mm Baseline Distance: 104.49 mm
		Lower binocular camera Resolution: 1280 × 960 pixels Frame Rate: 30 fps Focal Length: 16 mm Baseline Distance: 101.45 mm

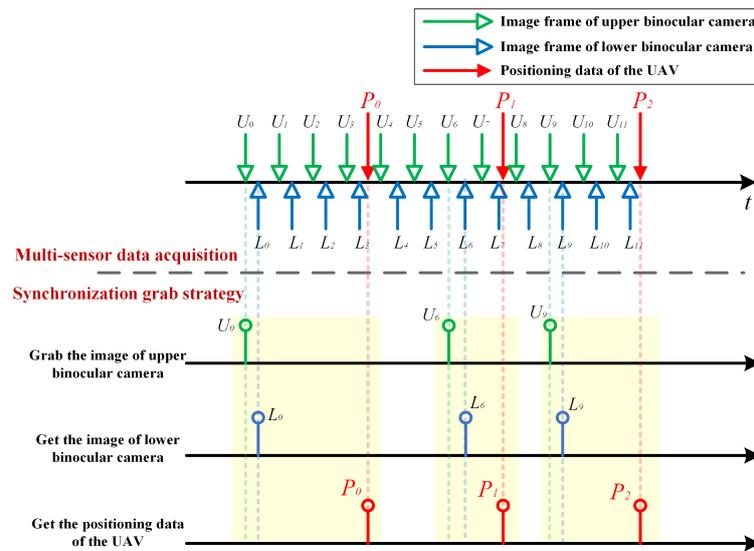
The self-designed, dual-view stereovision module is composed of two binocular cameras. In detail, one binocular camera is mounted on the top of the UAV at an upward angle

to locate and track the power lines, which is named the upper binocular camera in this paper. Moreover, in order to improve the accuracy of power line detection, the upper binocular camera is set with dip angles in both the upward and horizontal directions. Specifically, the upper binocular camera has an elevation angle of 5 degrees. Besides, the angle between the baseline of the upper binocular camera and the horizontal line is 10 degrees. Another binocular camera is hung under the UAV with a downward inclination of about 60 degrees for aerial image classification and ground objects discrimination; this is called the lower binocular camera. The NVIDIA Jetson AGX Xavier is a kind of AI computer for autonomous machines, delivering the performance of a GPU workstation in an embedded module under 30 W (<https://developer.nvidia.com/embedded/jetson-agx-xavier-developer-kit>, accessed on 17 August 2022)). It is adopted as an embedded platform to process upper and lower binocular images to achieve 3D environment perception and power line tracking in real-time. Meanwhile, the Universal Asynchronous Receiver/Transmitter (UART) connects the embedded NVIDIA platform and UAV, and enables communication between them. By developing the DJI onboard SDK (OSDK) (<https://developer.dji.com/onboard-sdk/>, accessed on 17 August 2022). and DJI payload SDK (PSDK) (<https://developer.dji.com/payload-sdk/>, accessed on 17 August 2022), the binocular images and position of the UAV could be processed in the embedded NVIDIA platform and calculate the next target flight point in real-time; then, the automatic inspection can be achieved.

## 2.2. Pipeline of Dual-View Stereovision-Guided Automatic Inspection Strategy

The long-distance transmission line corridor inspection can be divided into several short inspection tasks between adjacent transmission towers. The detailed processing steps within each short inspection task are shown in Figure 2. During the automatic inspection, the flight height of the UAV is similar to the power lines and the head direction of the UAV is controlled to keep perpendicular to the power lines. The upper binocular images are processed to detect and reconstruct the 3D power lines, and the lower binocular images are adopted for the reconstruction of the ground scene. The body-origin world coordinate of the UAV adopts the IMU as the coordinate origin. Therefore, the IMU coordinates can be used as the positioning coordinates of the UAV. Through the joint calibration results of the binocular cameras and internal IMU, the 3D points of power lines and ground scene are concatenated in the body-origin world coordinate system. Thus, in this way, the 3D information around transmission line corridors are acquired for automatic navigation and ground clearance measurement. In general, the IMU is just used as the reference of UAV positioning coordinates. The positioning information for navigation is estimated by fusing the information of the IMU, magnetic compass, and other sensors in the UAV. The positioning accuracy of the DJI UAV is shown in Table 1.

In order to synchronize the upper and downward cameras with the IMU, we performed two steps. The first step is to synchronize the local time of the embedded NVIDIA platform to the system times of the DJI UAV by utilizing the OSDK and PSDK. The UAV will send positioning data packets at a frequency of 1 Hz. The binocular image acquisition frequencies of the upper and downward cameras are both 5 Hz. It is obvious that the data acquisition frequencies of the UAV positioning module and binocular cameras are different. Therefore, in the second step, we designed a multisensor synchronous capture strategy to achieve the data synchronization. As shown in the Figure 3, the images obtained by the upper binocular camera are used as trigger information to grab other data within the same time interval. When we capture an image from the upper binocular camera, we will capture the first image from the lower binocular image and the first positioning data from the UAV to form a set of synchronous data. In the figure, the area with a light-yellow background represents a set of synchronized data. For example,  $\{U_0, L_0, P_0\}$  is a set of synchronized data, so are  $\{U_6, L_6, P_1\}$  and  $\{U_9, L_9, P_2\}$ .



**Figure 3.** Schematic diagram of multisensor data synchronous capture strategy.

In order to achieve automatic flight, we need enough aerial binocular images to satisfy the training requirements of power line detection and the aerial image classification network. Thus, before performing automatic flight, we controlled the UAV with the help of the pilot to capture sufficient aerial binocular images in advance along the 10 km-long transmission line in Jiangsu Province, China. Details of the datasets will be described in the experimental part. Then, we chose a part of the route to test our system on site. The location and photos of the typical flight route are shown in Figure 2.

### 3. Methods

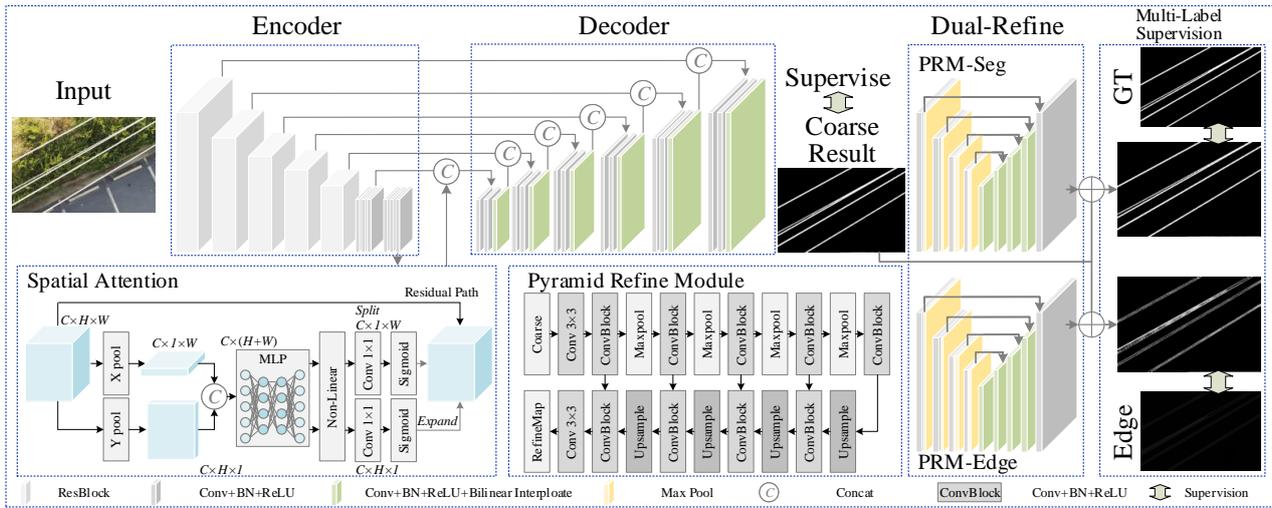
#### 3.1. Real-Time Automatic Flight Strategy Formulation Based on Power Line Tracking

Since the pixels of power lines only occupy a small part of the whole image, the feature points on power lines are difficult to detect and match. Thus, we propose an edge-assisted dual-refinement power line detection network to locate power lines. Then, the 3D power lines are reconstructed based on the epipolar constraint of binocular vision. Finally, according to the structure of the 3D power lines, the target flight point of the UAV is generated in the world coordinate system and the motion planning is achieved with the continuous target points.

##### 3.1.1. Edge-Assisted Power Line Detection

Since power lines may possess different numbers, angles, and thicknesses in aerial images, it is difficult to extract enough semantic cues from aerial images to achieve accurate power line detection. By integrating the idea of multitask learning, we propose an edge-assisted dual-refinement power line detection network. As shown in Figure 4, the proposed network consists of two important modules. The U-Net-like encoder–decoder module is adopted to predict the coarse detection results. Two layers are added behind the ResNet 18 backbone to detect more high-level semantic cues. Then, high-level features are further enhanced with the proposed spatial attention module. The framework of decoder is nearly symmetrical to the encoder except for the shortcut with lower layers.

By fusing high-level global contexts and low-level details, the coarse detection result is achieved and supervised. Furthermore, the coarse result is reconstructed to generate the edge map and refined result. By the generation and supervision of the edge information, the coarse result is further refined and the high-precision detection of the power lines is realized.



**Figure 4.** Overall framework of the proposed power line detection network.

### (1) Spatial attention module

Although the modified encoder module is deeper than the original ResNet 18, the overall receptive field is still restricted within the fully convolutional network. Thus, the spatial attention module is proposed to handle this problem. The details of the proposed spatial attention module are shown in Figure 4. The input feature  $X \in \mathbb{R}^{C \times H \times W}$  is firstly aggregated along the second and third dimension to encode positional information. Two attention maps  $\tilde{X}^h \in \mathbb{R}^{C \times H \times 1}$  and  $\tilde{X}^w \in \mathbb{R}^{C \times 1 \times W}$  are obtained to model the long-range dependence with different spatial direction. This kind of long-range information could enhance the location capability of the proposed network. The encoded output  $\tilde{X}^w$  at width  $w$  and the output  $\tilde{X}^h$  at height  $h$  could be formulated as

$$\tilde{X}^w = \frac{1}{H} \sum_{i=1}^H X(i, w), \tilde{X}^h = \frac{1}{W} \sum_{i=1}^W X(h, i). \quad (1)$$

These two attention maps are further concatenated to fuse the different coordinate attention together as

$$\tilde{X}^f = \text{Conv}_{1 \times 1} \left( \text{Cat}[\tilde{X}^h, \text{trans}(\tilde{X}^w)] \right), \quad (2)$$

in which  $\text{Conv}_{1 \times 1}$  denotes the convolutional layers with size  $1 \times 1$ ,  $\text{Cat}[* , *]$  denotes concatenation operator, and  $\text{trans}(\cdot)$  denotes the transpose operation.

Then, the fused attention map  $\tilde{X}^f$  is further squeezed to  $\tilde{X}^f \in \mathbb{R}^{C \times (H+W)}$ . Multi-layer Perceptron (MLP) is adopted to enhance the attention vectors via the relationship with others. The enhanced attention vectors  $S^f \in \mathbb{R}^{C \times (H+W)}$  could be calculated as  $S^f = \delta(\text{MLP}(\tilde{X}^f))$ , where  $\delta$  denotes the nonlinear activation function. We then split the  $S^f$  along the second dimension into  $S_w \in \mathbb{R}^{C \times W}$  and  $S_h \in \mathbb{R}^{C \times H}$ . Another two convolution layers with size  $1 \times 1$  are adopted to handle the unsqueezed  $S_w \in \mathbb{R}^{C \times W}$  and  $S_h \in \mathbb{R}^{C \times H}$  as

$$S'_h = \delta(\text{Conv}_{1 \times 1}(S_h)), S'_w = \delta(\text{Conv}_{1 \times 1}(S_w)). \quad (3)$$

Finally,  $S'_h$  and  $S'_w$  are expanded to the same size with  $X$  and used as weight maps to generate the output in a residual way as  $X_{out} = X \otimes S'_h \otimes S'_w + X$ , in which  $\otimes$  denotes the pixelwise multiplication.

### (2) Dual-refinement module

To segment power lines from aerial images more accurately, the dual-refinement module is proposed to optimize power line detection results. Specifically, the pyramid refinement module (PRM) is proposed for different tasks in two paths, respectively. One-

path PRM of the dual-refinement module is adopted to extract the edges of power lines from the coarse results. Another PRM is adopted to refine both the boundary and region of the coarse results. The structure of PRM likes U-Net [24], which is composed of an encoder–decoder network with multiple residual paths. Different from the structure of traditional U-Net, the layers of PRM are shallow to maintain efficiency. The PRM of the edge detection is removed in the inference stage. The outputs of PRMs could be considered as residual maps, which are added with the coarse result to obtain the refined edge map  $P_e$  and refined detection results  $P_d$ , respectively. Finally, the coarse detection results  $P_c$  are supervised with the binary cross-entropy (BCE) loss function as

$$\begin{cases} \mathcal{L}_{total} = \mathcal{L}_{bce}(GT_{pl}, P_c) + \mathcal{L}_{bce}(GT_{pl}, P_d) + \beta \times \mathcal{L}_{bce}(GT_{edge}, P_e) \\ \mathcal{L}_{bce}(GT, P) = \sum_{i=1}^N \{GT(i) \times \log(P(i)) + (1 - GT(i) \times \log(P(i)))\} \end{cases} \quad (4)$$

in which  $N$  denotes the total number of the pixels;  $\beta$  denotes the weight of edge supervision and is set as 0.5 by default.

### 3.1.2. Binocular-Vision-Based Power Line 3D Reconstruction

The texture feature of power lines is too similar to distinguish, which makes it hard for stereo matching algorithms to define matching cost and search for correct matching point pairs. Thus, it is difficult to reconstruct the power lines. To solve the above problem, we firstly detect the power lines in the rectified binocular images. The power line detection results are refined and thinned to single-pixel width. For a certain point on the power line, its  $Y$  coordinate values in the binocular images are the same. Based on this epipolar constraint, all point pairs with the same  $Y$  values on power lines are matched. For the matching points of the first power line in stereo images,  $(x_i^l, y_i^l)$  and  $(x_i^r, y_i^r)$ , the disparity of the matching points could be calculated as  $d = |x_i^l - x_i^r|$ . Then, the first power line could be converted into the 3D coordinate as  $P = \{p_1, p_2, \dots, p_i, \dots, p_N\}$ ,  $p_i = (x_i, y_i, z_i)$ , where  $p_i$  denotes the 3D power line points and  $P$  denotes the 3D point sets of power lines.  $N$  represents the total number of 3D power line points.

### 3.1.3. Power Line Tracking Guided Automatic Flight Strategy Formulation

In this way, power lines of current stereo images are reconstructed in the camera coordinate system  $o_c - x_c y_c z_c$ . To control the DJI M300 UAV, the target flight path point in the world-coordinate system  $o_w - x_w y_w z_w$  needs to be calculated. The origin of world coordinate system is generally a fixed point in the real world, just as the takeoff point. It is difficult to calculate the offset between the UAV and this point.

Thus, as shown in Figure 5, the body-origin world coordinate system  $o'_w - x'_w y'_w z'_w$  is built, which adopts the IMU as the coordinate origin. The transformation relationships between  $o'_w - x'_w y'_w z'_w$  with  $o_c - x_c y_c z_c$  have only rotation and translation operations. Considering the mounting angle of the binocular camera, the rotation operations could be represented as

$$q = \begin{bmatrix} q_w \\ q_x \\ q_y \\ q_w \end{bmatrix} = \begin{bmatrix} \cos(\theta/2) \cos(\varepsilon/2) \cos(\alpha/2) + \sin(\theta/2) \sin(\varepsilon/2) \sin(\alpha/2) \\ \cos(\theta/2) \cos(\varepsilon/2) \sin(\alpha/2) - \sin(\theta/2) \sin(\varepsilon/2) \cos(\alpha/2) \\ \sin(\theta/2) \cos(\varepsilon/2) \cos(\alpha/2) + \cos(\theta/2) \sin(\varepsilon/2) \sin(\alpha/2) \\ \cos(\theta/2) \sin(\varepsilon/2) \cos(\alpha/2) - \sin(\theta/2) \cos(\varepsilon/2) \sin(\alpha/2) \end{bmatrix}, \quad (5)$$

in which  $\alpha$ ,  $\theta$ , and  $\varepsilon$  are the rotation angles along  $x_c$ -axis,  $y_c$ -axis, and  $z_c$ -axis, respectively. In this way, all power line points  $(X_c, Y_c, Z_c)$  in the camera coordinate system could be converted with the quaternion as

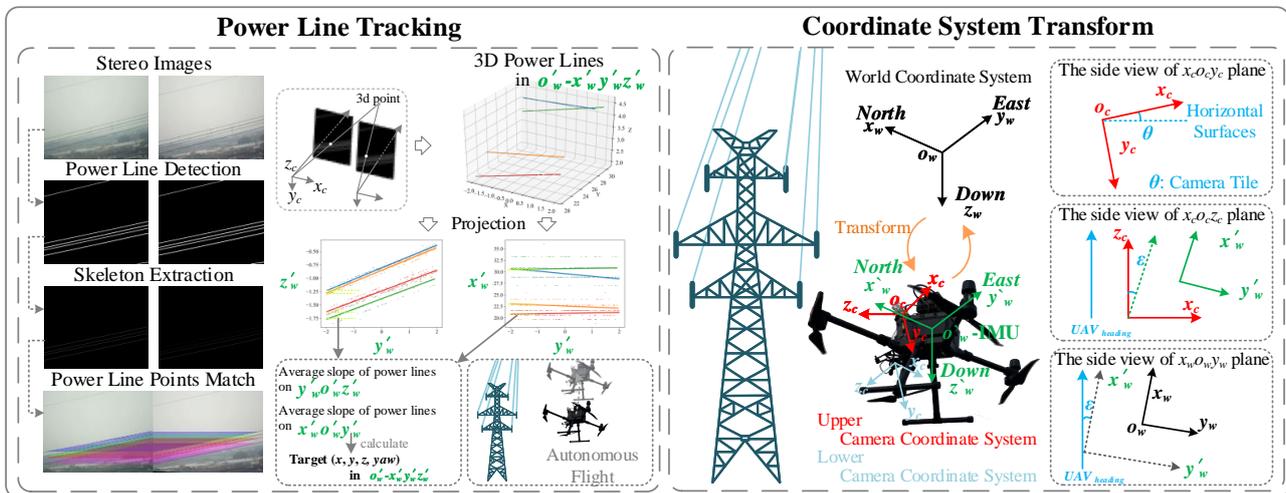
$$\begin{bmatrix} 0 \\ X'_w \\ Y'_w \\ Z'_w \end{bmatrix} = q \begin{bmatrix} 0 \\ X_c \\ Y_c \\ Z_c \end{bmatrix} q^{-1} + \begin{bmatrix} 0 \\ t_x \\ t_y \\ t_z \end{bmatrix}, \tag{6}$$

where  $(X'_w, Y'_w, Z'_w)$  is the coordinate of power line points in  $o'_w - x'_w y'_w z'_w$ ;  $t_x, t_y$ , and  $t_z$  are the distances between the origins and two coordinate systems; and  $\theta$  is the declination angle between the UAV heading and North, which could be obtained by the magnetic compass on the DJI UAV. Other parameters in the rotation and translation operators are calculated by the open-source camera/IMU calibration toolbox Kalibr [25], which could jointly calibrate the camera and internal IMU with millimeter precision. After all power line points are converted into  $o'_w - x'_w y'_w z'_w$ , the target flight point with the yaw value  $(x, y, z, yaw)$  in  $o'_w - x'_w y'_w z'_w$  is calculated for motion planning. Power line points are projected into the horizontal  $x'_w o'_w y'_w$  and vertical  $y'_w o'_w z'_w$  planes. Then, the average slopes of the projected power lines on these two planes are calculated as  $s_1$  and  $s_2$ , respectively. For a predefined flight distance  $L$  of each movement, the  $y$  value of target point could be calculated as

$$y = \sqrt{L / ((s_1)^2 + (s_2)^2)}. \tag{7}$$

Thus, the coordinate value of the target point  $P_f$  could be calculated as

$$P_f = (y \times s_1, y \times s_2, \arctan(s_1)), \tag{8}$$



**Figure 5.** Details of the proposed power line tracking algorithm and the transformation of different coordinate systems.

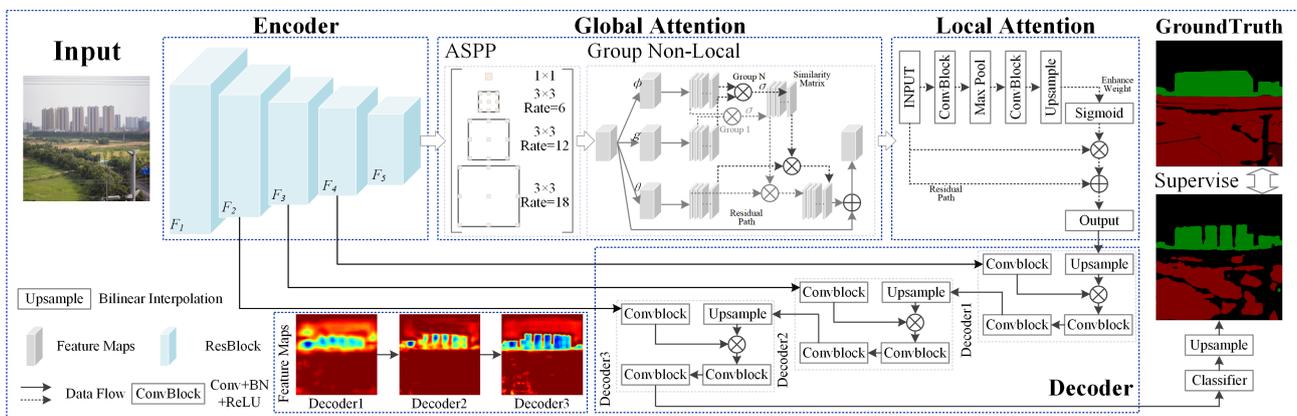
The target point is generated in  $o'_w - x'_w y'_w z'_w$ , which represents the offset in the world coordinate system. With the connection of the UAV and embedded NVIDIA platform, the UAV flies to this point immediately when receiving the command. When the UAV reaches the target point of current stereo images, the binocular images are reselected as the next key frame when the UAV hovering, and the next target point is recalculated to complete the automatic flight and inspection.

### 3.2. Ground Clearance Measurement and Detection of Transmission Corridors

#### 3.2.1. Ground Object Classification Based on Light-Weight Semantic Segmentation Network

In the actual environment of transmission line corridors, buildings and plants on the ground may pose a threat to the safe operation of the power lines. Different types of ground objects require different safe distances from power lines. In order to judge the safety of ground clearance correctly, an aerial image classification method based on light-weight semantic segmentation network is proposed to classify ground objects.

The overall framework of this method is shown in Figure 6, the global and local contexts are extracted to assist semantic segmentation, which could improve the performance of aerial image classification. The ResNet 18 is adopted as the encoder module within the light-weight semantic segmentation network. Multilevel features are extracted within the encoding process, which are denoted as  $F_i, i \in \{1, 2, 3, 4, 5\}$ .  $F_5$  is enhanced by the proposed global–local attention module. In the decoding stage, low-level CNN features  $F_2, F_3,$  and  $F_4$  are gradually fused with high-level features to recover more spatial information and refine the contour of instances.



**Figure 6.** Overall framework of the proposed aerial image classification method based on light-weight semantic segmentation network.

Global–local attention module is proposed to enhance the high-level features of the aerial images [26]. Specifically, for the multiscale features  $F \in \mathbb{R}^{H \times W \times C}$  extracted with the atrous spatial pyramid pooling (ASPP) module, the group nonlocal (G-NL) module is proposed to enlarge the receptive field. The features  $F$  are first embedded into the key, value, and query spaces as

$$F_{\theta} = F \times W_{\theta}, F_{\phi} = F \times W_{\phi}, F_g = F \times W_g, \quad (9)$$

in which ‘ $\times$ ’ denotes the matrix multiplication operator.  $W_{\theta}, W_{\phi},$  and  $W_g \in \mathbb{R}^{C \times C'}$  denote the embedding vectors, which could decrease the feature dimensions to reduce computation.

Suppose that  $\delta$  denotes the SoftMax activation operator and  $Y \in \mathbb{R}^{H \times W \times C'}$  denotes the output features; traditional nonlocal block generally calculates the similarity matrix and achieves the final output as  $Y = \delta(F_{\theta} \times F_{\phi}) \times F_g$ . Thus, by calculating the relationship between each pixel and the global information, the features are enhanced via the long-range dependence. However, due to the large number of feature points, the ability of nonlocal module for long-distance information is restricted. Thus, the concept of grouping is introduced into the nonlocal module to improve the ability to capture global clues. The  $F_{\theta}, F_{\phi},$  and  $F_g$  are first grouped along the channel dimension into  $F_{\theta}^k, F_{\phi}^k,$  and  $F_g^k \in \mathbb{R}^{H \times W \times t}$  with  $t = C' / \text{num\_group}$ . The embedding spaces with the same group number are processed as the original nonlocal module.

Assume that  $Y^i \in \mathbb{R}^{H \times W \times t}$  is the output of the  $i$ -th group. Then,  $Y$  is obtained by integrating the attentional outputs of all groups together.

$$\begin{cases} Y^i = \delta(F_\theta^i \times F_\phi^i) \times F_g^i, i \in \{1, 2, \dots, t\} \\ Y = \text{Cat}[Y^1, Y^2, \dots, Y^i, \dots, Y^t], i \in \{1, 2, \dots, t\} \end{cases} \quad (10)$$

The output of G-NL is calculated as  $F_{global} = F + YW_c$ , in which  $W_c$  is adopted to recover the channel information. However, when too much attention is paid to global information, some small areas may be ignored. Local Distribution is desired to be learned in each channel. The local context information of  $F_{global}$  is adopted to calculate the weight map of each channel. The weight map  $M \in \mathbb{R}^{H \times W \times C}$  is multiplied with the global feature to generate the final output as  $F_{out} = F_{global} + (M \otimes F_{global})$ .

$$\begin{cases} M = \delta(\text{Upsample}(M')) \\ M' = \text{Convblock}(\text{max\_pool}(\text{Convblock}(F_{global}))) \end{cases} \quad (11)$$

After the high-level features are extracted and enhanced, the decoder module is proposed to gradually recover spatial information with the shortcut of low-level features. As shown in Figure 6, the decoding process is divided into three stages. For stage 1 of the decoder module, it could be formulated as

$$F_1^D = \text{Convblock}^{\times 2}(\text{Upsample}(F_{out}) \otimes \text{Convblock}(F_4)). \quad (12)$$

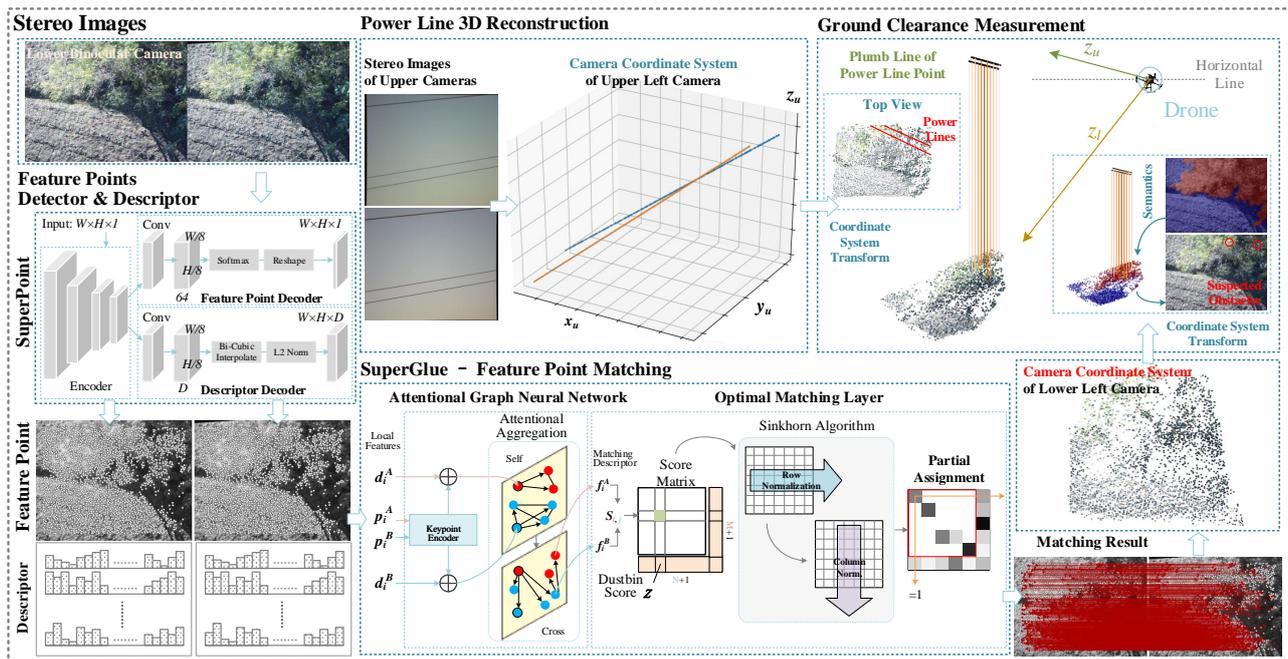
The other two stages of decoder module are the same as stage 1 except for different inputs and the change in channel numbers. Finally, the feature  $F_3^D$  is obtained with a channel size of 32 and the same spatial size as a quarter of the input image.  $F_3^D$  is further sent to the classifier module to generate the features with a channel of the class number. Given the prediction map and the corresponding ground truth, the cross-entropy loss is adopted to supervise the training process.

### 3.2.2. Ground Clearance Measurement and Detection

As shown in Figure 7, in order to calculate the ground clearance between power lines and the ground objects, we reconstruct the local 3D model of the transmission corridor according to upper and lower binocular images. Then, the ground points corresponding to the power line points are calculated with the plumb lines in the real world. Finally, ground clearance suspected obstacles are calculated combined with the aerial image classification results. The details of processes mentioned above are described in the following.

#### (1) 3D point reconstruction of ground objects based on feature point matching

In order to obtain the 3D information of the ground, we need to calculate the displacement between corresponding pixels. However, as the objects of binocular images are acquired at long distances and the scenes are complex, it is difficult for dense stereo matching methods to achieve satisfactory performance. Traditional sparse feature matching methods such as SIFT and ORB could only extract few corresponding points, which may lose some information within key positions. Thus, we adopt the advanced feature point detector SuperPoint [27] and feature matching method SuperGlue [28] to achieve accurate feature point matching. The framework of these two deep-learning methods are shown in Figure 7 and the detailed steps are described in the following.



**Figure 7.** The process of the proposed clearance distance measurement and suspect obstacles detection methods.

As shown in Figure 7, an encoder–decoder network is adopted in SuperPoint to extract high-level features for two decoder modules. In the interest point decoder, the features are first compressed into 64 channels with the 64-kernel convolutional layers. To minimize the checkerboard artifacts introduced by deconvolutional layers or upsampling layers, the subpixel convolution layers are applied on the SoftMax results to generate the one-channel response map. Meanwhile, inspired by the Universal Correspondence Network, the semidense grid of descriptor with size  $H/8 \times W/8$  is generated by the proposed descriptor decoder. Then, the semidense grid is upsampled with the bicubic interpolation and L2-normalizes operators to generate the descriptors with dimension  $D$ . The left and right images are both sent into the same SuperPoint. The key point  $P_i^A$  with descriptors  $d_i^A$  of the left image and the key point  $P_j^B$  with descriptors  $d_j^B$  of the right image are obtained by the SuperPoint algorithm in an end-to-end manner, in which  $i = \{1, 2, \dots, M\}$  and  $j = \{1, 2, \dots, N\}$ .

Traditional matching methods generally adopt Nearest Neighbor (NN) to search corresponding points and filter incorrect matches. These methods may ignore the assignment structure of feature points and discard visual information, which may result in unsatisfied and inaccurate matching performance. The SuperGlue adopts a deep neural network to solve the optimization problem and calculate the final target to obtain the soft partial assignment matrix  $P \in [0, 1]^{M \times N}$ . As shown in Figure 7, the attentional graph neural network is designed to enhance the local feature descriptor with long-range and global information. The Multilayer Perceptron (MLP) is adopted to couple the visual appearance  $d_j^*$  and the position of feature points  $p_j^*$ , thereby increasing the dimension of low-dimensional features. Then, a single multiplex graph is constructed, in which the nodes are the key points of both images. Through the message passing with interimage edges and cross-image edges, the receptive field is enlarged and the specificity of features is boosted with global information to achieve the feature aggregation. Then, the optimal matching layer is constructed to learn the score matrix  $S \in \mathbb{R}^{M \times N}$  and the Sinkhorn algorithm is adopted to maximize the total score  $\sum_{i,j} S_{i,j} P_{i,j}$  to obtain the soft partial assignment matrix.

In this way, the accuracy matching point pairs  $[(x_l^i, y_l^i), (x_r^i, y_r^i)]_{\beta}^{i=1}$  are obtained by the SuperPoint–SuperGlue algorithm, in which  $\beta$  denotes the total number of matched points. The 3D information of these points in the lower camera coordinate could be calculated as

$$Z^i = \frac{b_{low} * f_{low}}{x_l^i - x_r^i}, X^i = \frac{(x_l^i - x_l^0)}{f_{low}} Z^i, Y^i = \frac{(y_l^i - y_l^0)}{f_{low}} Z^i, \quad (13)$$

in which  $(X^i, Y^i, Z^i)$  denotes the 3D coordinates of the matched points and  $(x_l^0, y_l^0)$  denotes the primary point;  $b_{low}$  and  $f_{low}$  denote the baseline distance and focal length, respectively.

## (2) Ground clearance measurement and detection based on local 3D points of the corridor

As described in Section 3.1.3 and Figure 5, the translation and rotation between the lower camera coordinates with the body-origin world coordinate system are calibrated by Kalibr toolbox. In this way, 3D points calculated from the lower binocular camera are transformed into the body-origin world coordinate system. Thus, the local 3D model of the transmission corridor can be generated by concatenating the 3D points of the upper and lower binocular cameras in the same coordinate system.

Since the power line points and ground points are both converted into the world coordinate system, the vertical line in the real-world is parallel to the  $z$ -axis of the world coordinate system. The power line and ground points are all projected to the  $x'_w o'_w y'_w$  plane. Thus, through the bird's-eye view, the corresponding points of power lines on the ground could be located. However, since the matched ground points are sparse, there may be no corresponding ground points for some power line points. Thus, the 24-pixel neighborhood of the corresponding points is averaged as the height of the central point if the central point is unmatched. Finally, combined with the classification results of aerial image and the safe clearance distance of different ground objects, the security of the power line for the current frame is recorded.

Finally, the system can detect the safety of ground clearance easily according to rules. Combined with the classification results of aerial image and clearances of different ground objects, the security of the power line for the current frame is recorded and the suspected obstacles are detected.

## 4. Results

In this section, extensive experiments are conducted to evaluate the proposed power line detection, 3D reconstruction, and tracking methods. The effectiveness and robustness of the proposed ground clearance measurement are indicated in the following.

### 4.1. Experiments of Power Line Detection

#### 4.1.1. Datasets and Implementation Details

The proposed power line detection network is built with the commonly used deep-learning library PyTorch [29]. For the training process, the batch size and initial learning rate are set as 4 and 0.01, respectively. The learning rate becomes one tenth of the preset value every ten epochs. The Adam optimizer is adopted for training process. The proposed network is trained in an end-to-end manner on the NVIDIA RTX 3090 GPU. For the verification of the proposed power line detection network, the publicly available dataset and the self-build dataset are adopted for the fair experiment.

## (1) Transmission Towers and Power Lines Aerial-image Dataset (TTPLA)

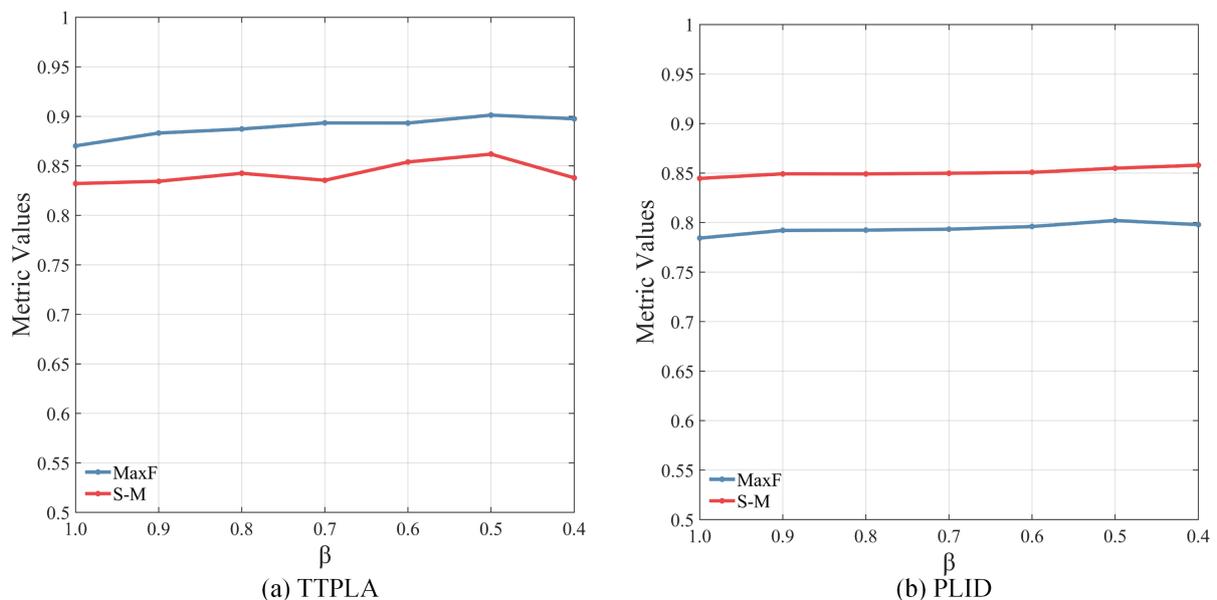
Abdelfattah et al. [15] proposed the Transmission Towers and Power Lines Aerial-image (TTPLA) dataset, which labeled the transmission towers and power lines for the instance segmentation. The UAV does not follow a fixed flight pattern for power line aerial images acquisition. Thus, there are great differences in the shape and distribution of power lines. There is a total of 1242 images in this dataset, in which 1000 images are selected as the training dataset and the others are used for testing.

## (2) Power Line Image Dataset (PLID)

The proposed inspection platform is used to acquire aerial images along the power lines. The distance and the shooting angle between power lines and the UAV change within a certain range. Thus, the orientations, pixel widths, and radius curvatures of power lines in all images are similar. We adopted 1000 images to build the dataset named Power Line Image Dataset (PLID), in which 600 images are selected as the training dataset and the others are used for testing.

We compare the proposed power line detection method with four edge detection methods including Canny, LSD [30], HED [31], and RCF [32]. Saliency detection could detect and segment salient objects into single-channel grayscale images, which are consistent with power line detection results. Therefore, three state-of-the-art saliency detection methods including PiCANet [33], EGNNet [34], and PFANet [35] are also selected for comparison. Note that the annotations of all datasets are converted into the binary form, and all these deep-learning-based methods are produced by training on the four datasets, respectively. The power line detection task could be considered as a kind of binary segmentation task such as salient object detection and edge detection. We adopt three commonly used metrics to quantitatively evaluate the proposed method including Max F-measure, MAE, and S-measure, which could evaluate the detection results at the pixel and structure levels.

As can be seen in Figure 8, we set  $\beta = \{0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$  and test the metric values on TTPLA dataset and PLID dataset. It can be seen that the network has better performance when we set  $\beta = 0.5$ .



**Figure 8.** Analysis of the power line detection method performance effected by different values of  $\beta$ .

## 4.1.2. Visual Comparison

As shown in Figure 9, the directions of power lines are quite different, which increases the difficulty of power line detection. The traditional methods based on edge and line detection all have unsatisfactory performance. Although HED and RCFNet could detect

the power lines with high accuracy, the pixel widths of the power lines are wider than the ground truth. Inaccurate power line widths may generate more errors in the thinning and refinement process. The detection results on the proposed PLAD are shown in Figure 10. Although the shooting distance and shooting angle between power lines and the UAV are fixed, the detection results of many methods are noisy. Meanwhile, power lines are broken in the detection results of many algorithms, which may result in inaccurate power line counting. Due to the introduction of edge features, power lines are segmented with fine boundary and accurate pixel widths in the proposed method. The proposed power line detection network has better performance on these two datasets.

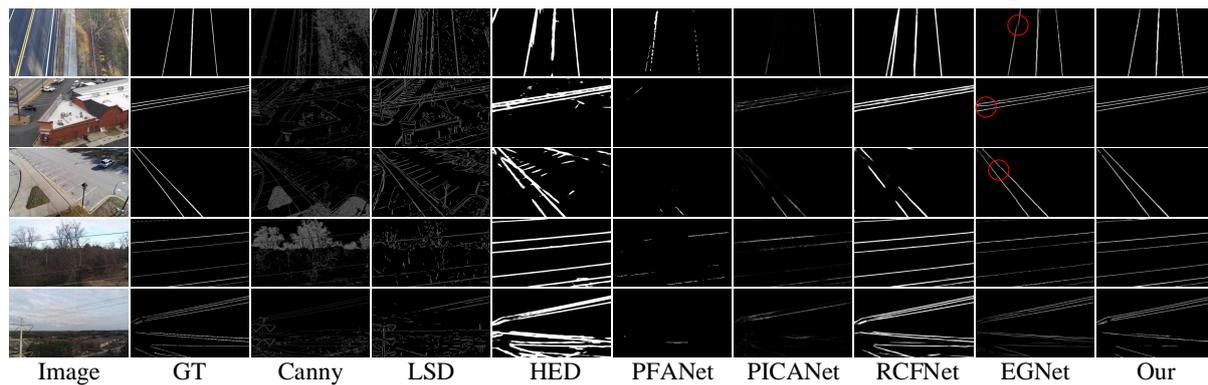


Figure 9. Comparison of power line detection results on the TTPLA dataset.

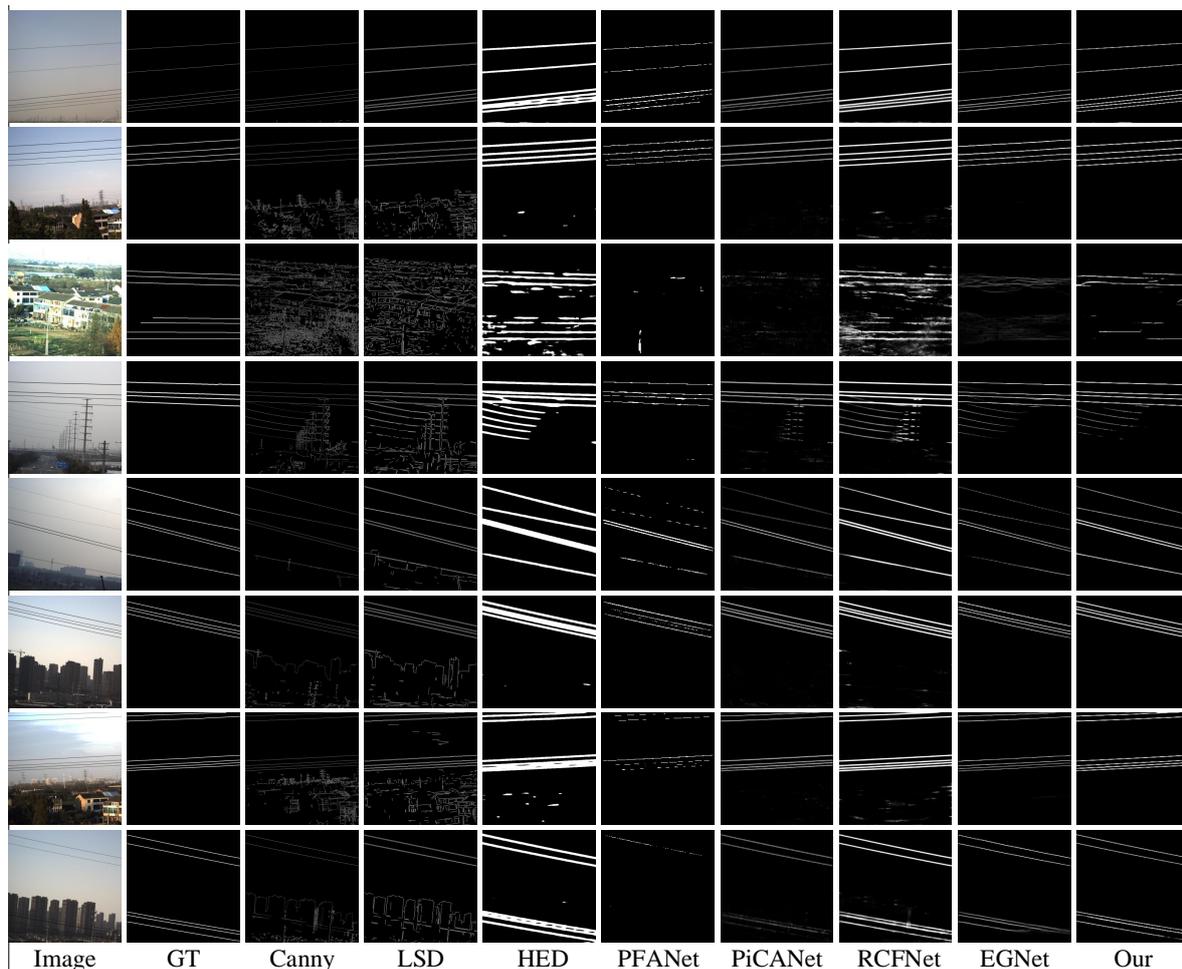


Figure 10. Comparison of power line detection results on the PLAD dataset.

#### 4.1.3. Quantitative Evaluation

The quantitative performance comparisons between our method and seven methods on the adopted datasets are shown in Table 2. The proposed method achieves better performance under different datasets and metrics. Since the training set of TTPLA has more images, the power line detection network could be trained better and achieve higher numerical results. Though the max F-measure of RCF is higher than that of other methods, the detection results are noisy when the background is complex. The max F-measure and S-measure are improved by about 1.5% and 1.8% compared with the second-best methods RCFNet and PiCANet on these two datasets. The quantitative experimental results show that the proposed method could realize the real-time detection of power lines with high accuracy, which could be adopted for 3D reconstruction.

**Table 2.** Comparison of power line detection results with other methods on two datasets.

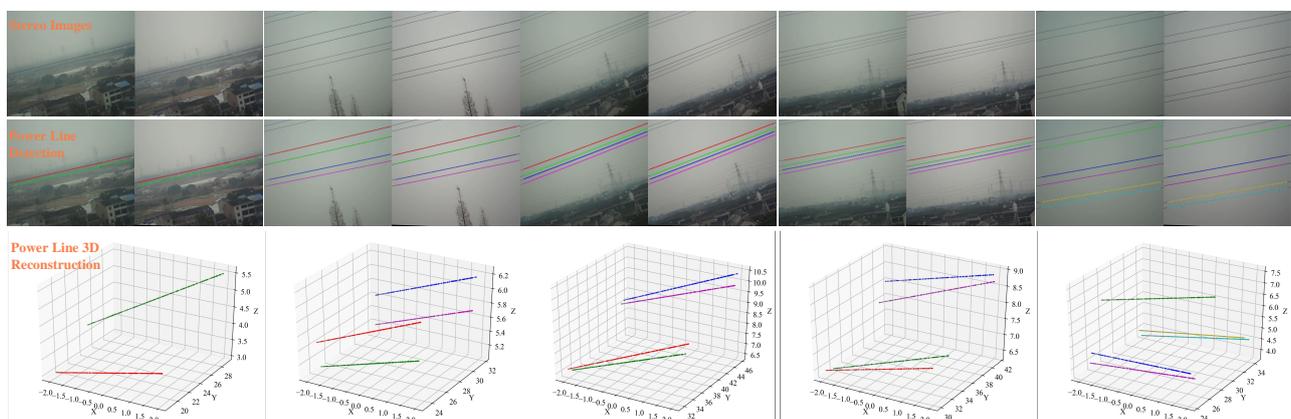
Dataset	Metries	Canny	LSD	HED	PFANet	PiCANet	RCFNet	EGNet	Ours
TTPLA	MaxF	0.1537	0.2506	0.6034	0.5109	0.7242	0.8730	0.7972	0.9012
	MAE	0.1944	0.0715	0.0526	0.0390	0.0405	0.0227	0.0314	0.0209
	S-M	0.4660	0.5235	0.6997	0.5690	0.6659	0.8460	0.7322	0.8618
PLID	MaxF	0.5331	0.4300	0.4819	0.6967	0.7992	0.8002	0.7320	0.8021
	MAE	0.0415	0.0422	0.0620	0.0266	0.0271	0.0377	0.0270	0.0204
	S-M	0.5716	0.6030	0.6236	0.6959	0.8338	0.7719	0.7738	0.8549

#### 4.2. Experiments of Power Line Tracking

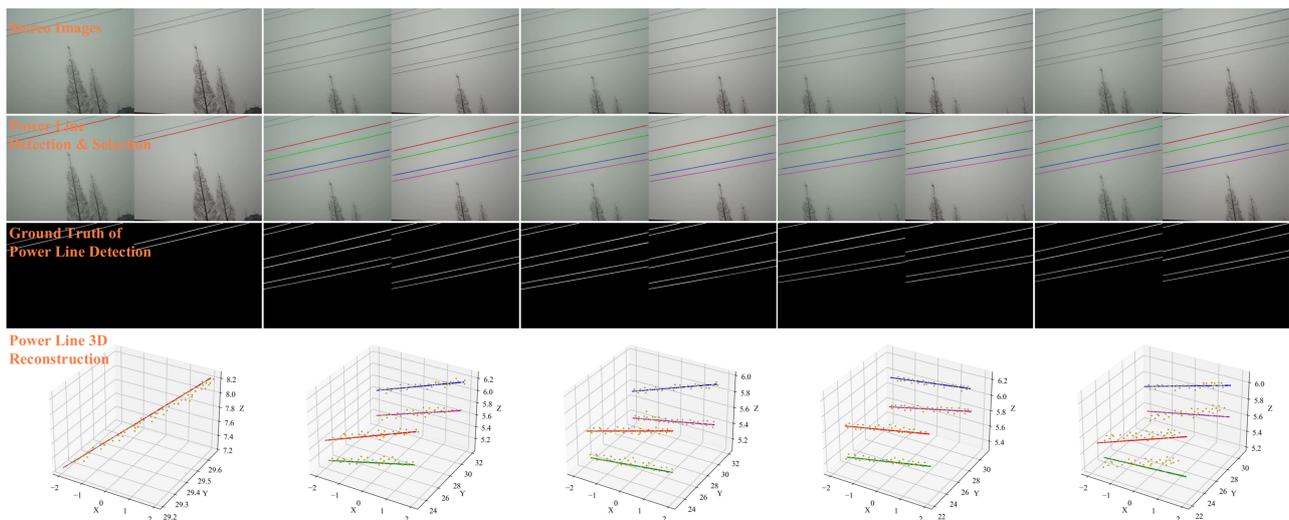
##### 4.2.1. Performance of Power Line 3D Reconstruction

To evaluate the performance of the proposed power line tracking and autonomous inspection algorithms, we conduct several flight experiments in the actual environment. Taking the selected 220-kV transmission line as an example, the transmission tower contains a total of 14 power lines within the low, medium, and high phases of transmission lines and the uppermost layer of ground wires. Power lines are distributed on both sides of the power tower. The width of the transmission channel is 8 m. When inspecting and shooting from the horizontal side, power lines on both sides are acquired in the same image. It is difficult for human eyes to judge the distance between power lines and the UAV.

Figures 11 and 12 shows the stereo images acquired by the upper binocular camera in the real inspection process. The short power lines are not reconstructed according to the preset parameters. Meanwhile, the power line points are all constructed and fitted with the proposed methods. The 3D reconstruction results of power lines are presented in the body-origin world coordinate system. The y-axis of the body-origin world coordinate system denotes the distances between power lines and the UAV.



**Figure 11.** Power lines 3D reconstruction results of different scenes.



**Figure 12.** Power lines 3D reconstruction results of several continuous frames. (Discrete points indicate the corresponding LiDAR points).

Figure 11 shows power line detection and 3D reconstruction results in different scenes. In other binocular images, power lines are distributed on both sides of transmission towers. It can be seen that the distances between two sides of power lines are about 8 m, which also coincides with the reality.

Figure 12 shows some power line detection and 3D reconstruction results of several continuous frames. It can be seen that the power lines can be detected accurately in the binocular images. The power lines are almost parallel to each other in each frame, which coincides with the reality. We map the sparse point clouds of the Livox Avia LiDAR to the body-origin world coordinate system as a reference for accuracy evaluation. The point cloud reconstructed by our algorithm is relatively consistent with the point cloud distribution of Livox Avia LiDAR. It illustrates that our 3D reconstruction results of power lines are relatively accurate even in the binocular images with obvious interference. Therefore, our localization method is encouraging from the perspective of scenario adaptability.

In order to decrease the magnetic-field interference and maintain the safety of the inspection system, we kept the distance between the UAV and power lines above 20 m. As shown in Figure 11, power lines in the first column of binocular image are on the same side of the transmission tower. It can be seen from the y-axes that the distance between the UAV and the nearest power line is about 24 m. Besides, the distances of other results in Figures 11 and 12 are all maintained above 20 m, which shows that our method can effectively control the distance between the UAV and the power lines.

#### 4.2.2. Experiments of Autonomous Flight

We conduct several experiments in the real-world around the transmission lines to evaluate the performance of the proposed power line tracking and autonomous flight strategies. The experimental position has a strong RTK signal, which could improve the accuracy of recorded trajectories. The calculation points of key frames are drawn in blue circles in Figure 13. The distances between each calculation key point are about 4 m. The trajectory of the UAV is relatively smooth, which could reflect the stability and robustness of the power line tracking.

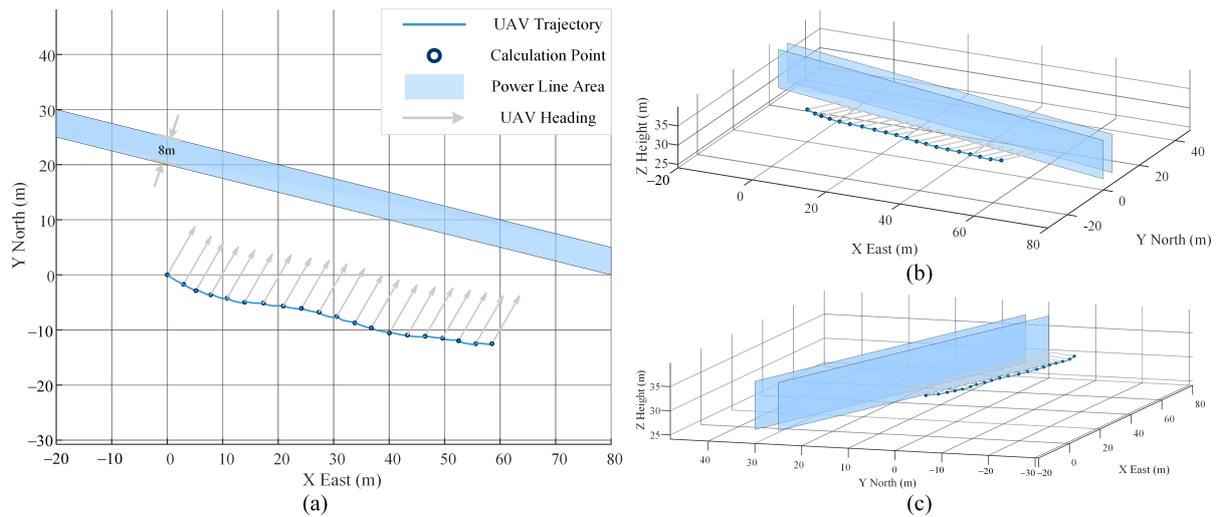


Figure 13. The detailed trajectory of a flight. (a) Altitude view; (b,c) are different planar views.

In order to further show the effectiveness and robustness of the proposed autonomous flight strategy, we use the UAV to track different power lines. As shown in Figure 14, the developed inspection platform could fly along the power lines with different heights. Although the pendulous traits of power lines mean the shapes of power lines are not straight, the trajectories of the inspection platform all have a similar shape with the curve of power lines.

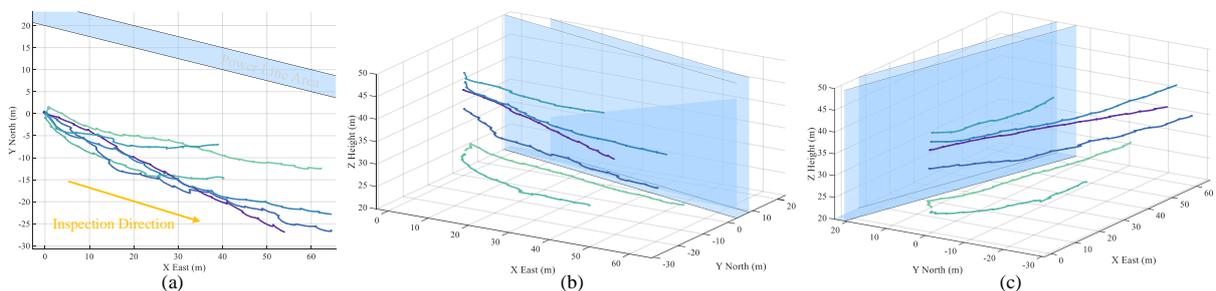


Figure 14. Trajectories of several flight tests. (a) Altitude view; (b,c) are different planar views.

#### 4.3. Experiments of Ground Clearance Measurement and Detection

In this section, we carry out experiments on the ground clearance measurement and detection function of the system. Since the voltage level of the tested transmission line is 220 kV, the detection rule set of ground clearance in the system is shown in Table 3. In the following content, we will test the performance of the system in terms of the ground object classification function and ground clearance detection function, respectively.

Table 3. Min safety ground clearances between different ground objects of 220 kV transmission lines.

Ground Objects	Buildings	Trees	Roads	Rivers
The min safety ground clearance	6.0 m	3.5 m	8.0 m	7.0 m

##### 4.3.1. Performance of the Proposed Aerial Image Classification Network

###### (1) Datasets and implementation details

In order to acquire enough scene styles, we collect aerial images by controlling the UAVs flight for about 10 km. A total 200 images are selected for the training and testing of the proposed semantic segmentation. Five classes are annotated including buildings, trees, roads, rivers, and others. The original resolution of the images is  $2448 \times 2048$ . As can be

seen in Figure 15, we divide 200 images into five parts according to the typical background objects in them. We randomly sample 3/4 of the five groups of data for training. Before the system test on site, about 1/3 of the remaining data are used for inference and verifying the effectiveness of the classification algorithm. The input images resolution of the network should be resized to  $512 \times 512$ . The details of the aerial image classification dataset are listed in Table 4.

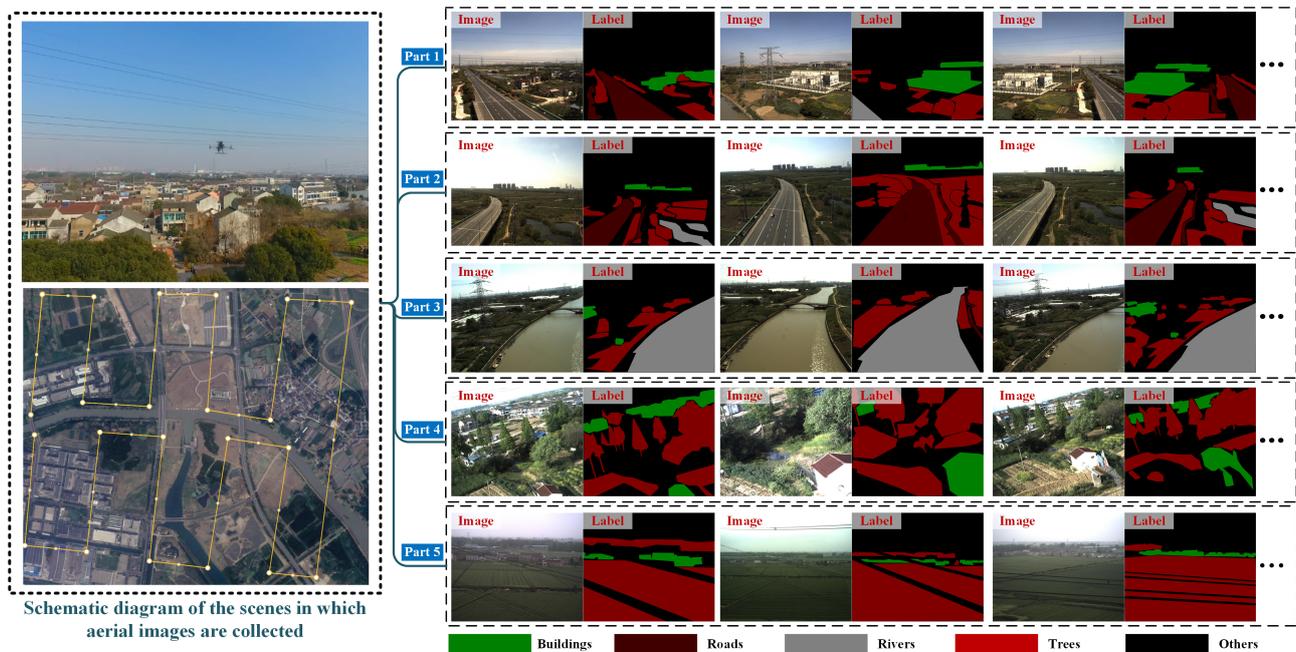


Figure 15. Samples of the aerial image classification dataset.

Table 4. Details of the aerial image classification dataset.

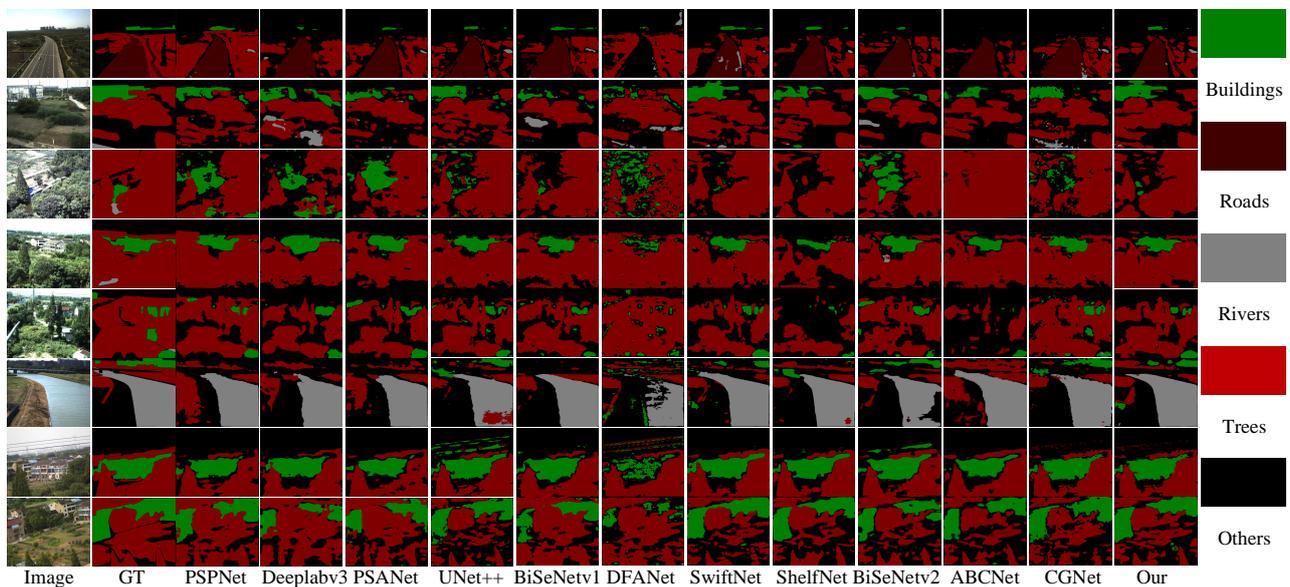
Part	Typical Background Objects	Total Number	Original Image Resolution	Number for Training	Number for Inference	Input Image Resolution
1	Buildings, Roads	35	2448 × 2048	27	8	512 × 512
2	Roads, Trees	42		32	10	
3	Rivers, Trees	32		24	8	
4	Buildings, Trees	44		33	11	
5	Trees	47		35	12	

The numbers of pixels in all classes are extremely unbalanced, which increases the difficulty of semantic segmentation for aerial images. A total 11 state-of-the-art semantic segmentation networks are adopted for a fair comparison. The proposed semantic segmentation network is also built with the PyTorch framework. The batch size and initial learning rate are set as 4 and 0.01, respectively. The learning rate of the global–local attention and decoder module are set as ten times of backbone network. Other training settings and parameter settings are consistent with the training of power line detection.

## (2) Visual comparison

To qualitatively validate the effectiveness of the proposed semantic segmentation network, we firstly visualize the segmentation results generated by our method and 11 comparative methods in Figure 16. As shown in Figure 16, introduced by the class-aware edge, the proposed method could segment the image with fine boundary. Compared with the segmentation results of other methods, the proposed method could segment the ground objects more completely. The boundaries of big objects such as roads or rivers in the segmentation results are continuous and accurate. Meanwhile, due to the huge scale

differences of categories in the proposed dataset, comparison methods generally have poor classification and segmentation performance for small areas.



**Figure 16.** Aerial image classification results of the proposed and the state-of-the-art networks.

### (3) Quantitative evaluation

The evaluation results of the proposed semantic segmentation network and the comparison with other state-of-the-art methods on the proposed dataset are shown in Table 5. The proposed method obtains 73.90% in terms of mean IoU and achieves the best performance among the 11 state-of-the-art methods. The mIoU of the proposed method is increased by about 3% compared with the second-best method. Due to the adoption of the shallow backbone network, the model size of the proposed method is about 12.9 M parameters. In terms of inference speed, the proposed network is tested on the embedded NVIDIA platform and the runtime is only 0.18 s, which could satisfy the requirement of real-time calculation.

**Table 5.** Comparison of segmentation results with state-of-the-art methods on the proposed dataset.

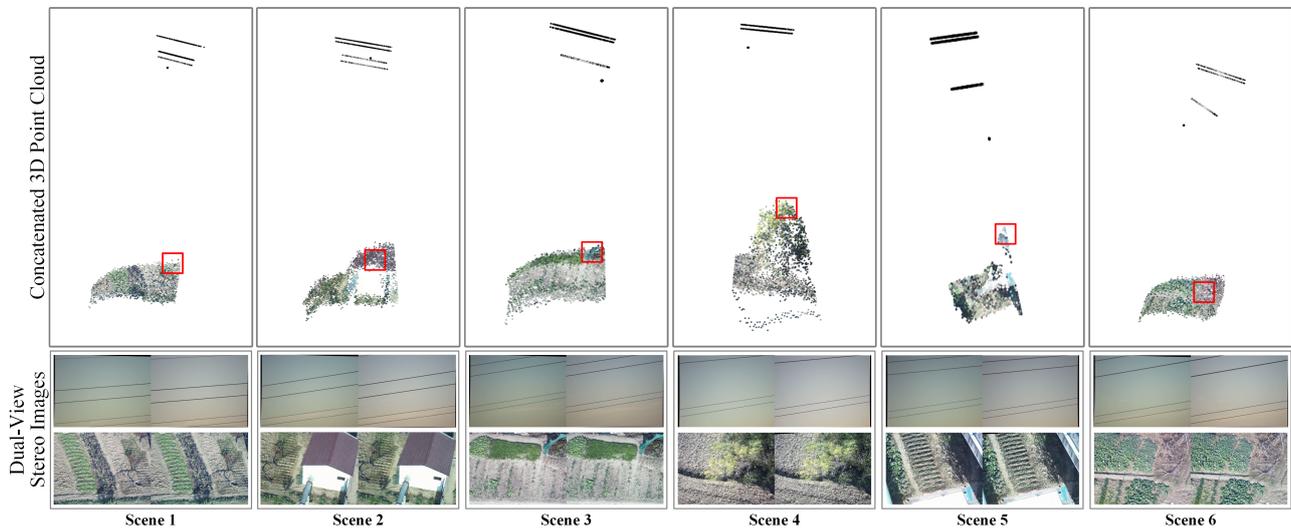
Model	Year and Ref.	Class IoU (%)					mIoU (%)	Param (M)
		Trees	Buildings	Rivers	Roads	Others		
PSPNet	2017 [36]	68.11	61.66	81.16	62.47	75.74	68.35	46.7
DeepLab	2017 [37]	60.10	61.56	78.16	56.00	73.29	63.96	39.0
PSANet	2018 [38]	65.72	60.20	79.83	57.87	75.07	65.91	48.3
UNet++	2018 [39]	66.86	69.55	79.84	61.02	76.97	69.32	9.2
BiSeNetv1	2018 [40]	64.34	61.17	78.43	56.05	73.62	65.00	13.3
DFANet	2019 [41]	62.72	24.07	37.55	0.01	62.94	31.44	2.2
SwiftNet	2019 [42]	66.32	72.62	84.43	60.18	76.34	70.89	11.8
ShelfNet	2019 [43]	62.98	66.72	80.78	59.13	74.19	67.40	14.5
BiSeNetv2	2020 [44]	65.41	60.01	70.82	57.11	73.47	63.33	3.3
ABCNet	2021 [45]	61.73	61.02	71.81	47.63	71.98	60.55	13.4
CGNet	2021 [46]	65.85	61.82	79.63	55.07	74.47	65.59	0.5
Ours	--	69.71	74.40	86.59	64.91	78.50	73.90	12.9

#### 4.3.2. Performance of the Clearance Distance Measure Method

##### (1) Visualization of generated local 3D points of the transmission line corridor

Through the concatenation of the 3D point clouds of the upper and lower binocular cameras, the spatial distribution relationship between power lines and the ground can be

displayed directly and visually. As shown in Figure 17, because of the long focal length of the cameras, the field-of-view of the cameras are relatively narrow. With the adopted SuperPoint and SuperGlue feature matching methods, many points of the ground are matched and the ground point cloud are reconstructed with high accuracy.



**Figure 17.** 3D reconstruction results of the inspection scenes.

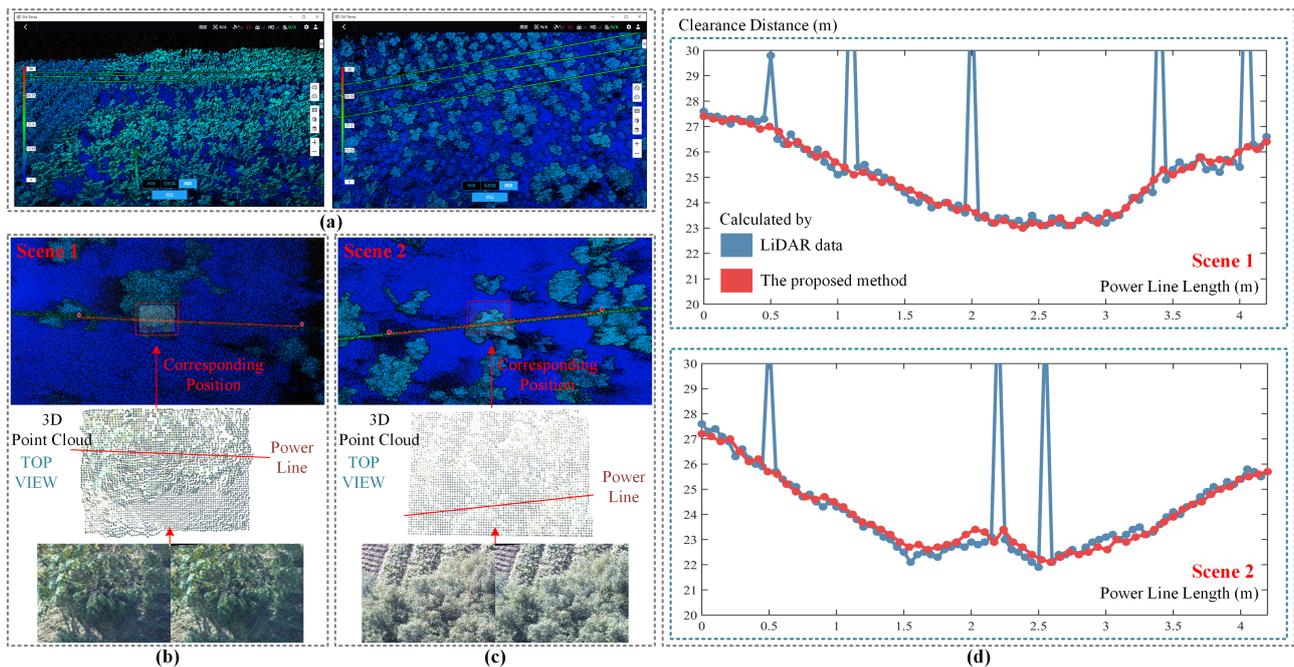
Combined with the joint calibration results of lower camera, upper camera, and IMU inside the UAV, the 3D power line points and ground points are converted into the body-origin world coordinate system. The black point in Figure 17 is the position of the UAV. It can be seen from Figure 17 that the 3D information near the transmission line corridor is obtained. Due to the fixed shooting distances between the UAV and power lines, the power lines in each image are generally about 5 m-long and the clearances are about 15 m. In addition, Table 6 shows the time of different information processing stages required to reconstruct the six scenes in Figure 17. It can be seen that the total processing time of each set of data is about 1.6 s.

**Table 6.** The time of different information processing stages required to reconstruct the six scenes in Figure 17.

Time Consuming	Scene 1	Scene 2	Scene 3	Scene 4	Scene 5	Scene 6
Power line detection time	0.212s	0.208 s	0.214 s	0.205 s	0.201 s	0.205 s
Power line 3D reconstruction time	0.525 s	0.568 s	0.498 s	0.475 s	0.535 s	0.528 s
Ground object classification time	0.196 s	0.213 s	0.199 s	0.203 s	0.204 s	0.205 s
Ground object 3D reconstruction time	0.314 s	0.303 s	0.315 s	0.318 s	0.314 s	0.317 s
Ground clearance measurement time	0.256 s	0.244 s	0.252 s	0.238 s	0.248 s	0.251 s
The next target point recalculation time	0.105 s	0.107 s	0.111 s	0.105 s	0.108 s	0.105 s
Total time consuming	1.608 s	1.643 s	1.589 s	1.544 s	1.610 s	1.611 s

## (2) Quantitative evaluation of ground clearance measurement results

To quantitatively evaluate the accurateness of the proposed ground clearance measurement algorithm, we use the UAV with Livox Avia LiDAR to scan around the transmission lines, and reconstruct the 3D point cloud in conjunction with professional offline software named 'Dajiang Zhitu(Power Version)'. Besides, the Livox Avia LiDAR has high ranging accuracy, and its maximum detection range and random error of distance measurement are about 450 m and 2 cm, respectively. Figure 18a exhibits local 3D point clouds of transmission lines acquired by Livox Avia LiDAR and its professional software.



**Figure 18.** Experiments of ground clearance measurement. (a) Local 3D point clouds of transmission line acquired by Livox Avia LiDAR and its professional software. (b,c) Two different scenes for evaluation. (d) The clearance distances of power line points calculated by the LiDAR data and the proposed method.

Figure 18b,c show two typical scenes of transmission line, which are reconstructed by the Livox Avia LiDAR and our system, respectively. Specifically, the red box marked in the LiDAR point cloud shown in the upper part overlaps with the reconstruction result of our dual-view stereovision-guided system shown in the lower part. Subsequently, the clearance distances between the power lines and objects below in the two overlapping scenes measured by two means are shown in Figure 18d.

As the laser of LiDAR has certain penetration, some LiDAR rays can cross the gaps of the trees, hit the ground, and then reflect back to the receiving module of the LiDAR. Therefore, these data points with much longer LiDAR distances shown in Figure 18d are normal, which show the distance between the power lines and the ground under the trees. Our image-based clearance distance method is unable to penetrate the trees for measurement. Therefore, our system can only measure the clearance distance between the power lines and tree crowns. Since the points calculated by the proposed methods are relatively sparse, the number of calculated distance points are less than that of LiDAR. Except the penetrating points of the LiDAR, our clearance distance measurement points are close to other corresponding points of LiDAR. This shows that the proposed system in this paper has good performance close to LiDAR, which benefits from the advanced proposed algorithms and effective cooperation between various modules.

Meanwhile, it can be seen in Table 7 that the clearances calculated by the proposed method are similar to the LiDAR data. The mean absolute error between the proposed method and the LiDAR data is about 0.168 m. Thus, the accuracy of the proposed ground clearance measurement method is evaluated, which could meet the requirements of actual inspection applications. Meanwhile, integrated with the proposed inspection platform, the inspection of transmission line is safer and more efficient. The red boxes in Figure 17 indicate the positions with the minimum ground clearance distances for all scenes. The measured distances are further compared with the minimum clearance distance of each class. As shown in Table 8, the safe level of each scene could be judged with high accuracy.

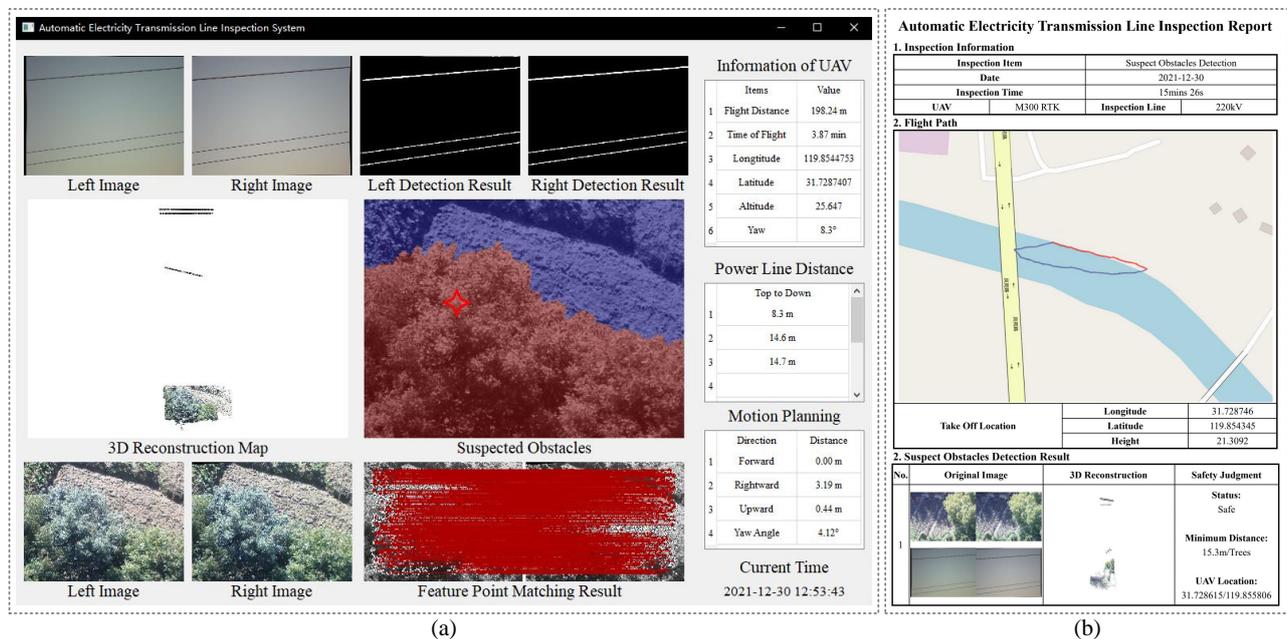
**Table 7.** The mean and standard deviation of the errors within the average of calculated ground clearance distances.

Types	Average Clearances		Error	
	LiDAR	Proposed Method	Mean Absolute (m)	Standard Deviation (m <sup>2</sup> )
Scene 1	24.843	24.871	0.168	0.0176
Scene 2	23.910	23.940	0.208	0.0199

**Table 8.** Results of ground clearance detection in the tested 220-kV transmission line corridor.

Scene	Category of Ground Object	Measured Clearance	Safety Standards	Judgment Results
1	Trees	14.9 m	3.5 m	Safe
2	Buildings	11.4 m	6.0 m	Safe
3	Trees	16.7 m	3.5 m	Safe
4	Trees	6.9 m	3.5 m	Safe
5	Buildings	10.3 m	6.0 m	Safe
6	Trees	20.2 m	3.5 m	Safe

Finally, the interface of the inspection platform is shown in Figure 19. During the inspection process, the state of the UAV, the acquired images, and processing results are aggregated for the image transmission. The interface could be monitored on the observation screen by the pilot in real-time. Thus, the pilot could obtain the real distance between the UAV and power lines to improve the inspection safety. The inspection report could be generated when the inspection is finished, which could further improve the automation of the system.

**Figure 19.** Interface and the generated patrol report of the inspection platform. (a) The visual interface of the automatic electricity transmission line inspection system; (b) The report of the automatic electricity transmission line inspection system.

## 5. Discussion

In this paper, a dual-view stereovision-guided automatic inspection system for the transmission line corridor is developed. The system consists of a DJI UAV, an advanced embedded NVIDIA platform, and a dual-view stereovision module, which can synchronously sense 3D information of power lines and ground objects on transmission line corridors.

Based on the hardware platform, we further develop a dual-view stereovision-guided automatic inspection strategy. Using this strategy can control the UAV to move along the power lines automatically and measure the ground clearance in real-time. Specifically, we firstly locate power lines in upper binocular images accurately by the proposed edge-assisted power line detection method. Experiments show that the proposed power line detection method performs well even in images with complex backgrounds compared with other state-of-the-art methods. Then, 3D points of the detected power lines are reconstructed based on epipolar geometry constraint. After that, the next guidance flight point is generated to control the automatic movement of the UAV. Besides, the lower binocular images are processed by the proposed aerial image classification method to classify categories of ground objects. Meanwhile, 3D points of ground objects are reconstructed and concatenated with 3D points of power lines to generate the local 3D point cloud of the transmission line corridor. Finally, based on plumb line determination and object classification results, ground clearance is measured and detected correctly. The inspection tests in real-world environments have demonstrated that the proposed automatic flight and inspection strategy can satisfy the real-time and accuracy requirements for practical inspection applications.

The magnetic field generated by the high-voltage cable will affect the flight state of the UAV. During the previous development of our system, we have taken into account the effect of the transmission line magnetic field on the UAV. We find that the DJI UAV will be unable to fly normally when it detects excessive magnetic interference. Therefore, we will conduct the magnetic compass calibration of the UAV in an environment without strong magnetic field interference. Besides, in order to decrease the magnetic field interference and keep the safety of the inspection system, we kept the distance between the UAV and power lines above 20 m. At the present stage, we control the interference of the magnetic field of the transmission line by keeping the proper distance between the drone and the transmission lines. As can be seen in Figure 13, the gray arrows are the heading directions (declination angles) generated by the magnetic compass from the UAV DJI. The angles between gray arrows and power lines are almost maintained near 90 degrees, which indicates that the declination angles are not greatly affected by the magnetic field. The anti-electromagnetic interference performance of the UAV system is essential for transmission line inspection. When the magnetic field becomes stronger, increasing the distance to the power line will not be feasible. Therefore, we will further analyze, study, and optimize the anti-electromagnetic interference performance of the UAV system in future work.

Transmission line inspection is an extremely complex and challenging task. There are still many interesting directions to study. For example, more inspection functions could be embedded into the proposed system to make the inspection more convenient. Thus, we will research the defect detection methods of power line such as broken strand, icing, etc. and integrate all these algorithms into a unified framework. Besides, we will further study the automatic flight strategy of power towers based on the proposed UAV platform and attempt to enable the UAV system to automatically accomplish electric fitting detection tasks of transmission line such as insulators, vibration dampers, and so on.

## 6. Conclusions

In this paper, a dual-view stereovision-guided automatic inspection system for the transmission line corridor is developed. Beneficial from the proposed power line detection, ground object classification, and stereovision-based measurement methods, our system could perceive and reconstruct power lines and ground areas, thus realizing autonomous flight and ground clearance measurements. The inspection experiments in real-world environments have demonstrated that the proposed system could satisfy the real-time and accuracy requirements for practical inspection applications.

Our further work will focus on three aspects to extend the functions of the UAV system. Firstly, we will further analyze, study, and optimize the anti-electromagnetic interference performance of the UAV system. Secondly, we will research the defect detection methods of power line such as broken strand, icing, and so on; this is also important for transmission line corridor inspection. Afterwards, we will further study the automatic flight strategy

of power towers based on the proposed UAV platform, and attempt to enable the UAV system to automatically accomplish electric fitting detection tasks of transmission lines such as insulators, vibration dampers, and so on.

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

**Funding:** This research was funded by the National Natural Science Foundation of China, grant number 62001156, and the Key Research and Development Plan of Jiangsu Province, grant number BE2019036 and BE2020092.

**Data Availability Statement:** The data are not publicly available due to the confidentiality of the research projects.

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

## References

1. Tao, X.; Zhang, D.; Wang, Z.; Liu, X.; Zhang, H.; Xu, D. Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks. *IEEE Trans. Syst. Man Cybern. Syst.* **2020**, *50*, 1486–1498. [[CrossRef](#)]
2. Ola, S.R.; Saraswat, A.; Goyal, S.K.; Jhajharia, S.; Rathore, B.; Mahela, O.P. Wigner Distribution Function and Alienation Coefficient-based Transmission Line Protection Scheme. *IET Gener. Transm. Distrib.* **2020**, *14*, 1842–1853. [[CrossRef](#)]
3. Choi, H.; Yun, J.P.; Kim, B.J.; Jang, H.; Kim, S.W. Attention-based Multimodal Image Feature Fusion Module for Transmission Line Detection. In *IEEE Transactions on Industrial Informatics*; Springer: Berlin/Heidelberg, Germany, 2022; p.1. [[CrossRef](#)]
4. Kandanaarachchi, S.; Anantharama, N.; Muñoz, M.A. Early Detection of Vegetation Ignition Due to Powerline Faults. *IEEE Trans. Power Deliv.* **2021**, *36*, 1324–1334. [[CrossRef](#)]
5. Shi, S.; Zhu, B.; Mirsaiedi, S.; Dong, X. Fault Classification for Transmission Lines Based on Group Sparse Representation. *IEEE Trans. Smart Grid* **2019**, *10*, 4673–4682. [[CrossRef](#)]
6. Jiang, S.; Jiang, W.; Huang, W.; Yang, L. UAV-Based Oblique Photogrammetry for Outdoor Data Acquisition and Offsite Visual Inspection of Transmission Line. *Remote Sens.* **2017**, *9*, 278. [[CrossRef](#)]
7. Nguyen, V.N.; Jenssen, R.; Roverso, D. Automatic Autonomous Vision-Based Power Line Inspection: A Review of Current Status and The potential Role of Deep Learning. *Int. J. Electr. Power Energy Syst.* **2018**, *99*, 107–120. [[CrossRef](#)]
8. Yang, L.; Fan, J.; Liu, Y.; Li, E.; Peng, J.; Liang, Z. A Review on State-of-the-Art Power Line Inspection Techniques. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 9350–9365. [[CrossRef](#)]
9. Zhai, Y.; Yang, X.; Wang, Q.; Zhao, Z.; Zhao, W. Hybrid Knowledge R-CNN for Transmission Line Multifitting Detection. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [[CrossRef](#)]
10. Zhao, Z.; Qi, H.; Qi, Y.; Zhang, K.; Zhai, Y.; Zhao, W. Detection Method Based on Automatic Visual Shape Clustering for Pin-Missing Defect in Transmission Lines. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 6080–6091. [[CrossRef](#)]
11. Guan, H.; Sun, X.; Su, Y.; Hu, T.; Wang, H.; Wang, H.; Peng, C.; Guo, Q. UAV-Lidar Aids Automatic Intelligent Power Line Inspection. *Int. J. Electr. Power Energy Syst.* **2021**, *130*, 106987. [[CrossRef](#)]
12. Song, B.; Li, X. Power Line Detection From Optical Images. *Neurocomputing* **2014**, *129*, 350–361. [[CrossRef](#)]
13. Zhao, W.; Dong, Q.; Zuo, Z. A Method Combining Line Detection and Semantic Segmentation for Power Line Extraction from Unmanned Aerial Vehicle Images. *Remote Sens.* **2022**, *14*, 1367. [[CrossRef](#)]
14. Zhang, Y.; Yuan, X.; Li, W.; Chen, S. Automatic Power Line Inspection Using UAV Images. *Remote Sens.* **2017**, *9*, 824. [[CrossRef](#)]
15. Abdelfattah, R.; Wang, X.; Wang, S. TTPLA: An Aerial-Image Dataset for Detection and Segmentation of Transmission Towers and Power Lines. In Proceedings of the Asian Conference on Computer Vision (ACCV), Virtual, 30 November–4 December 2020.
16. Vemula, S.; Frye, M. Real-Time Powerline Detection System for an Unmanned Aircraft System. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC 2020), Toronto, ON, Canada, 11–14 October 2020; pp. 4493–4497. [[CrossRef](#)]
17. Li, Y.; Ibanez-Guzman, J. Lidar for Autonomous Driving: The Principles, Challenges, and Trends for Automotive Lidar and Perception Systems. *IEEE Signal Process. Mag.* **2020**, *37*, 50–61. [[CrossRef](#)]
18. Zhang, Y.; Yuan, X.; Fang, Y.; Chen, S. UAV Low Altitude Photogrammetry for Power Line Inspection. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 14. [[CrossRef](#)]
19. Caron, G.; Eynard, D. Multiple camera types simultaneous stereo calibration. In Proceedings of the IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 2933–2938. [[CrossRef](#)]

20. Strauß, T.; Ziegler, J.; Beck, J. Calibrating multiple cameras with non-overlapping views using coded checkerboard targets. In Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Qingdao, China, 8–11 October 2014; pp. 2623–2628. [[CrossRef](#)]
21. Figueiredo, R.; Hansen, J.; Fevre, J.; Brandão, M.; Kayacan, E. On the Advantages of Multiple Stereo Vision Camera Designs for Autonomous Drone Navigation. *arXiv* **2021**, arXiv:2105.12691.
22. Zhou, Y.; Li, Q.; Wu, Y.; Ma, Y.; Wang, C. Trinocular Vision and Spatial Prior Based Method for Ground Clearance Measurement of Transmission Lines. *Appl. Opt.* **2021**, *60*, 2422–2433. [[CrossRef](#)] [[PubMed](#)]
23. Chang, C.W.; Lo, L.Y.; Cheung, H.C.; Feng, Y.; Yang, A.S.; Wen, C.Y.; Zhou, W. Proactive Guidance for Accurate UAV Landing on a Dynamic Platform: A Visual–Inertial Approach. *Sensors* **2022**, *22*, 404. [[CrossRef](#)] [[PubMed](#)]
24. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI, Munich, Germany, 5–9 October 2015; pp. 234–241.
25. Rehder, J.; Nikolic, J.; Schneider, T.; Hinzmann, T.; Siegwart, R. Extending Kalibr: Calibrating the Extrinsic of Multiple IMUs and of Individual Axes. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 4304–4311. [[CrossRef](#)]
26. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-Local Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
27. DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperPoint: Self-Supervised Interest Point Detection and Description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, UT, USA, 18–23 June 2018.
28. Sarlin, P.E.; DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperGlue: Learning Feature Matching With Graph Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
29. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2019; Volume 32.
30. Grompone von Gioi, R.; Jakubowicz, J.; Morel, J.M.; Randall, G. LSD: A Line Segment Detector. *Image Process. Line* **2012**, *2*, 35–55. [[CrossRef](#)]
31. Xie, S.; Tu, Z. Holistically-Nested Edge Detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
32. Liu, Y.; Cheng, M.M.; Hu, X.; Wang, K.; Bai, X. Richer Convolutional Features for Edge Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–16 July 2017.
33. Liu, N.; Han, J.; Yang, M.H. PiCANet: Learning Pixel-Wise Contextual Attention for Saliency Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
34. Zhao, J.X.; Liu, J.J.; Fan, D.P.; Cao, Y.; Yang, J.; Cheng, M.M. EGNet: Edge Guidance Network for Salient Object Detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019.
35. Zhao, T.; Wu, X. Pyramid Feature Attention Network for Saliency Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
36. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–16 July 2017.
37. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)]
38. Zhao, H.; Zhang, Y.; Liu, S.; Shi, J.; Loy, C.C.; Lin, D.; Jia, J. PSANet: Point-wise Spatial Attention Network for Scene Parsing. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
39. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Granada, Spain, 20 September 2018; pp. 3–11.
40. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
41. Li, H.; Xiong, P.; Fan, H.; Sun, J. DFANet: Deep Feature Aggregation for Real-Time Semantic Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
42. Wang, H.; Jiang, X.; Ren, H.; Hu, Y.; Bai, S. SwiftNet: Real-Time Video Object Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 1296–1305.
43. Zhuang, J.; Yang, J.; Gu, L.; Dvornek, N. ShelfNet for Fast Semantic Segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Seoul, Korea, 27 October–2 November 2019.
44. Liu, P.; Zhang, H.; Yang, G.; Wang, Q. An Improved Image Segmentation Method of BiSeNetV2 Network. In *Proceedings of the 4th International Conference on Control and Computer Vision*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 13–17.

45. Liu, Y.; Chen, H.; Shen, C.; He, T.; Jin, L.; Wang, L. ABCNet: Real-Time Scene Text Spotting with Adaptive Bezier-Curve Network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 13–19 June 2020.
46. Wu, T.; Tang, S.; Zhang, R.; Cao, J.; Zhang, Y. CGNet: A Light-Weight Context Guided Network for Semantic Segmentation. *IEEE Trans. Image Process.* **2021**, *30*, 1169–1179. [[CrossRef](#)] [[PubMed](#)]