WILEY | Hindawi

Research Article

An Effective Method for Sensing Power Safety Distance Based on Monocular Vision Depth Estimation

Leixiong Wang,[1] **Bo Wang** [iD]**,**[1] **Shulong Wang,**[1] **Fuqi Ma,**[1] **Xuzhu Dong,**[1] **Liangzhong Yao,**[1] **Hengrui Ma,**[1] **and Mohamed A. Mohamed** [iD]**²**

[1]*School of Electrical and Automation, Wuhan University, Wuhan, Hubei 430072, China*
[2]*Electrical Engineering Department, Faculty of Engineering, Minia University, Minia 61519, Egypt*

Correspondence should be addressed to Bo Wang; whwdwb@whu.edu.cn

As an important index of risk protection, the safety distance is crucial to ensure the safe and stable operation of the power system and the safety of personnel's life. Traditional monitoring methods are difficult to balance recognition accuracy and convenience. Therefore, this paper presents a power safety distance sensing method based on monocular visual images to achieve the recognition of the safety distance of external damage in complex scenes of transmission corridors, and proposed a power density depth distance model. In this model, a codec network with skip-connection to extract features and aggregate shallow and deep features for input power system images. Then, the regularization method, migration learning strategy, cosine annealing learning strategy, and data enhancement strategy are used to further optimize the model, so as to obtain a model with good accuracy and generalization in complex conditions. The effectiveness and superiority of the proposed method are verified in comparison to other external damage monitoring methods. The experimental results showed that the proposed method has high accuracy for the distance of external damage in the actual scenario. Moreover, the method has good generalizability, which can be easily deployed in video monitoring systems on different transmission corridors.

1. Introduction

Because the power system has the characteristics of high voltage, strong current, and outward discharge, the power system has strict safety distance standards, operational management measures, and other means to prevent various short-circuit, fire, explosion, and personal injury accidents caused by human body and construction appliance touching or being too close to the charged object [1]. According to the causes of different safety accidents, it can be learned that the control effect of means such as restricting personnel and apparatus from entering the charged areas by means of five preventions and other safety regulations is limited [2]. When power enterprises are under severe pressure of overhaul and maintenance, there are problems such as the poor implementation of safety production responsibilities and lax control of operating sites. For example, during the equipment reconstruction of 500 kV shipping substation of

Chongqing Electric Power Company in 2021, due to the insufficient distance between crane lifting equipment and electrified equipment, a bus trip accident was caused. It shows the power system lacked an effective safe distance sensing method [3]. It can be seen that the research on the method of measuring the power safety distance is of great importance to ensure the safe operation of power equipment and the safety of personnel [4].

Currently, safety distance sensing methods for power systems primarily include manual measurement, LiDAR [5], and video monitoring methods [6]. Electric power workers often rely on experience or use theodolite to determine whether there is insufficient safety distance in the inspection section. However, because of the subjective factors of electric power workers, interference from trees and buildings, and visual bias, it is difficult for workers to effectively and accurately determine whether the safety distance is below the standard. The LiDAR method mainly obtains the spatial

geometric structure of the inspected object through inspection by drones equipped with LiDAR, which has the advantages of high-ranging accuracy, strong directionality, and no ground clutter interference [7]. However, the cost is high and requires drones, which cannot be inspected in real-time and is not conducive to the real-time identification of safety hazards. In addition, the processing of laser point cloud data has a high degree of difficulty [8]. The intelligent video surveillance method mainly uses binocular images for depth estimation and safety distance discrimination through the similar triangle principle [9]. Due to the limitation of the early depth estimation principle, this method has a high false alarm and missed alarm rate [10]. Therefore, in order to prevent the occurrence of major accidents affecting the national production life, the power industry needs a power safety distance awareness method that can balance detection accuracy and ease of deployment [11].

Over the past few years, with the rapid development of deep learning, deep neural networks with strong adaptive functionality have been widely accepted by academics [12]. Depth estimation based on deep learning can construct models that correlate image information and depth information to obtain the depth information of the scene [13, 14]. The depth estimation technique based on deep learning gives better results [15]. Currently, they can be categorized as supervised, unsupervised, and semisupervised according to the degree of use of true depth distance [16]. As a supervised approach, the literature [17] achieved good depth estimation performance based on adaptive interval segmentation through deep residual networks for depth-valued classification. The literature [18] uses an unsupervised approach to train the network to obtain depth information of images using the geometric constraint information of neighboring frames of a monocular video stream with multiple frames, reducing the data usage limitation and obtaining promising results. As a semisupervised approach, the literature [19] introduces real depth maps as supervised information in an unsupervised framework, and achieves a blend of supervised and unsupervised by using a more powerful supervised signal for training. Depth estimation techniques based on deep learning have shown great progress in performance [20], and their application in the field of power safety distance perception has become possible.

Summarizing the previous literature on transmission corridor monitoring methods, it can be seen that these methods are difficult to combine both detection accuracy and ease of deployment and are difficult to apply on large-scale transmission corridors. However, in practical applications, transmission line monitoring usually faces problems such as large monitoring area scope, the coexistence of near and distant objects, random operation area, and complex image background [21]. Under the cost limitation, the current method has many problems such as low detection accuracy and low monitoring efficiency. In order to improve the recognition accuracy and efficiency of transmission line safety distance and enhance the generalization capability of transmission line safety distance monitoring, this paper constructs a monocular image-based power safety distance

sensing method and proposes a power density depth model based on supervised depth estimation for existing transmission corridor video monitoring systems. The main contributions of this paper are summarized as follows:

(1) A power safety distance sensing method of external damage based on deep learning is constructed to achieve the real-time recognition of the safety distance of external damage, which adapts to most transmission corridor scenarios and improves the monitoring efficiency of the safety distance of external damage

(2) A power density depth model is proposed, which is based on a supervised depth estimation approach using a coder-decoder network architecture with jump connections for image feature extraction aggregation, direct output of spatial distance information in complex scenes of transmission corridors through the network, and multiple optimization strategies to achieve high generalization and recognition accuracy

(3) The effectiveness and superiority of the proposed power safety distance sensing method for safety distance of external damage identification are verified in comparison to traditional manual measurement, LiDAR, and video monitoring methods

The rest of the paper is organized as follows: Section 2 introduces the power safety distance sensing method network structure. Section 3 presents the structure parameter optimization and data enhancement of the power density depth model. In Section 4, experiment results are presented to verify the proposed method, followed by conclusions. Section 5 is the conclusion.

2. Power Safety Distance Sensing Method Based on Power Density Depth Distance Model

This paper proposes a power security distance sensing method based on the power density depth distance sensing model according to the dense depth model [22]. The specific process of this method is shown in Figure 1. This paper adopts the method of "offline training + online application" to build the power safety distance sensing method, the model is constructed and trained offline, and then the trained model is deployed in the transmission corridor video monitoring system for online application. Under the optimization of the regularization method, migration learning strategy, cosine annealing learning strategy, and data enhancement strategy, the offline training phase is mainly to learn the mapping relationship between image pixel information and the corresponding depth distance information by power density depth distance model.

The online application is mainly to input the current moment's images into the trained power density depth distance model to quickly obtain spatial distance information in complex scenes of transmission corridors.
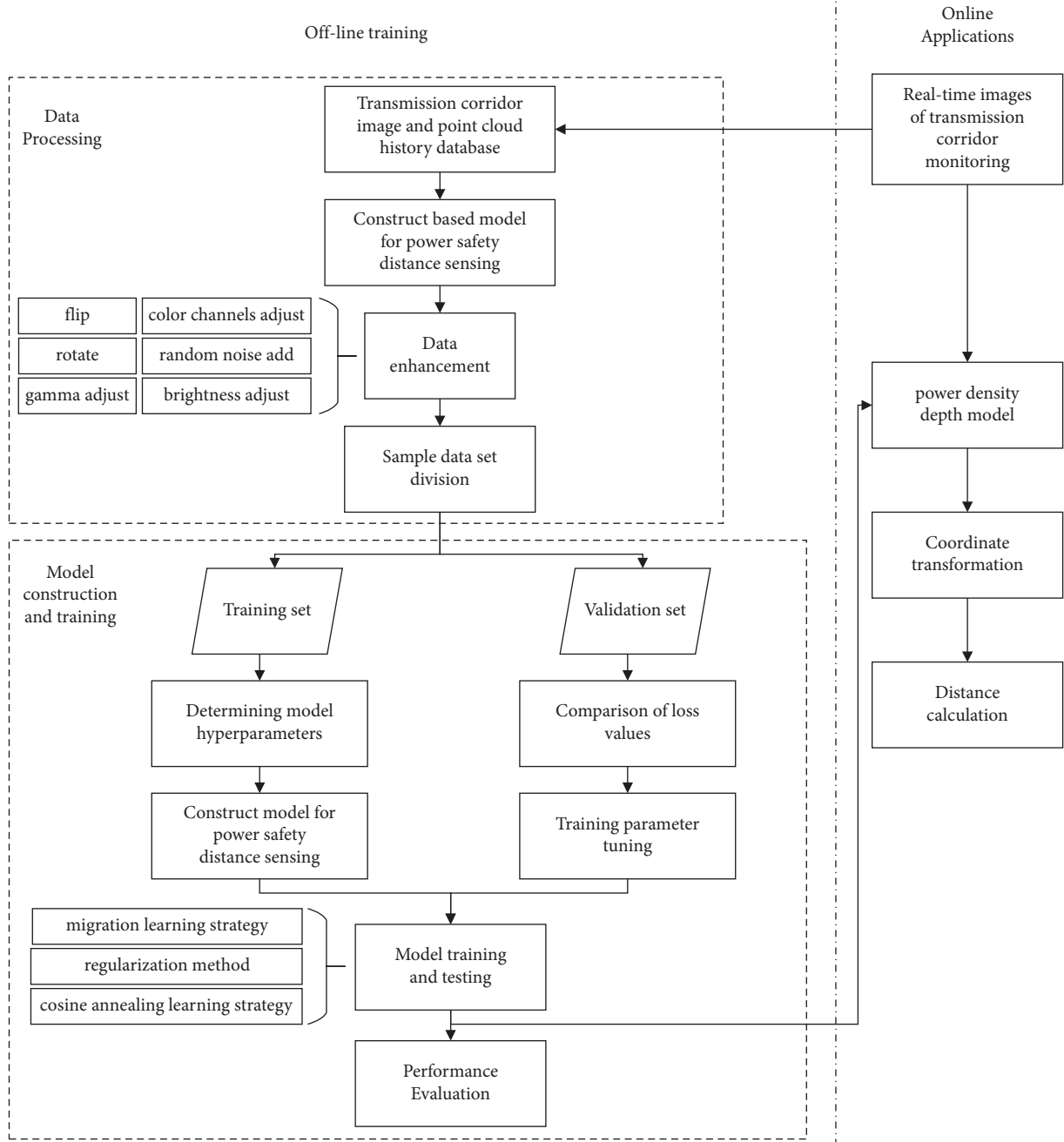
FIGURE 1: Flowchart of power safety distance sensing method.

Then, the spatial distance information based on the pixel coordinate system is transformed into the depth estimation structure based on the real coordinate system. At last, the coordinate points are manually selected to calculate the distance and judge whether the distance is lower than the safety distance standard.

2.1. Depth Feature Extraction Codec Network Based on DenseNet. The network model is shown in Figure 2; the network extracts the features of the input power system image, aggregates the shallow features and deep features, and extracts the fine structure features to ensure that the network can effectively use the context information provided by the

deep features to help the depth estimation of a single point. The shallow features including object contour and position information are effectively used to improve the overall accuracy of the depth estimation algorithm.

The core network of the method is a skip-connected coder-decoder network which is based on the convolutional neural network. In this paper, DenseNet-169 [23] is used for feature extraction of power monocular images as an encoder. The last layer of each convolutional block of the encoder is two bilinear sampling blocks and ReLU activation function with parameters for downsampling, which can obtain more spatial features while reducing the difficulty of calculation.
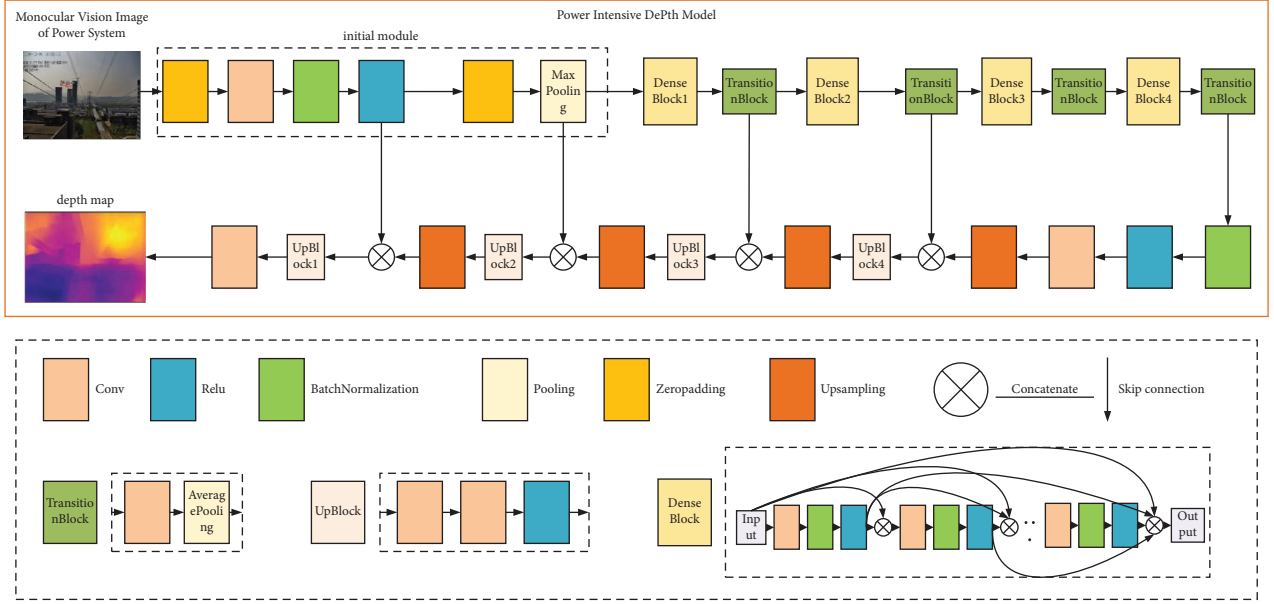
Figure 2: Power density depth distance sensing model.

Input a color monocular image. If the number of convolutional layers is the output eigenvector of the current layer, which is expressed as follows:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right), \tag{1}$$

where the ReLU function is $(\cdot)$, the output of the current layer is x_j^l, and the convolutional operation is $*$, when the single convolutional kernel of the current convolutional layer is k_{ij}^l and the table convolutional layer offset is b_j^l.

If the number of convolutional layers is m, Formula (2) is the output eigenvector of the pool layer.

$$x_j^m = f\left(\beta_j^m \text{down}\left(x_j^{m-1}\right) + b_j^m\right), \tag{2}$$

where the softmax activation function is $f(\cdot)$, the connection weight is β_j^m, the input of the current pool layer is x_j^{m-1}, the input matrix summation operation is expressed as $\text{down}(\cdot)$, and the current offset is b_j^m.

In this paper, the decoder is composed of convolutional operation and bilinear upsampling operation. The convolutional block of the corresponding encoder is jump connected to the upsampling block of the corresponding decoder. While expanding the feature map, the fine edge structure feature map is obtained to reduce the feature loss. The feature map is the depth map directly output after the convolutional operation. The resolution of the output depth map of the algorithm is 1/2 of the input image.

2.2. Model Loss Function. The main meaning of the loss function of the algorithm is to minimize the depth difference between the predicted depth image $\widehat{D}$ and the original depth image D, and the image detail distortion of the reconstructed depth image $\widehat{D}$.

The composition of the loss function of this algorithm is shown in Formula (3).

$$L_s = \frac{\lambda}{n}\left(\sum_P \left|D_P - \widehat{D}_P\right| + \sum_P \left(\left|g_x\left(D_P, \widehat{D}_P\right)\right| + \left|g_y\left(D_P, \widehat{D}_P\right)\right|\right)\right.$$
$$\left. + \sum_P \frac{1}{2}\left(1 - \text{SSIM}\left(D_P, \widehat{D}_P\right)\right)\right). \tag{3}$$

In this loss function algorithm, D is the original depth image, $\widehat{D}$ is the reconstructed depth image, p is the pixel point, n is the total number of pixels, and $\lambda = 0.1$ is the weight parameter of depth loss.

The first line on the right side of the equation is depth loss, which means that the pixel difference of the pixel corresponding to the same position p of the reconstructed depth image $\widehat{D}$ and the original depth image D is calculated.

The second line on the right of the equation is the loss of depth smoothness, which represents the minimum second gradient $L1$ criterion defined on the depth image gradient g, where g_x and g_y calculate the difference between the x and y components of the depth image gradient, respectively.

The third line on the right side of the equation is the appearance matching loss structure similarity item, SSIM [24]. SSIM is a commonly used measure in image reconstruction task and expressed as shown in formula (4).

$$\text{SSIM}\left(D, \widehat{D}\right) = \frac{\left(2\mu_D\mu_{\widehat{D}} + c_1\right)\left(2\sigma_{D\widehat{D}} + c_2\right)}{\left(\mu_D^2 + \mu_{\widehat{D}}^2 + c_1\right)\left(\sigma_D^2 + \sigma_{\widehat{D}}^2 + c_2\right)}. \tag{4}$$

In SSIM algorithm, μ_D is the average value of the original depth image D, $\mu_{\widehat{D}}$ is the average value of the reconstructed depth image $\widehat{D}$, σ_D^2 is the variance of D, $\sigma_{\widehat{D}}^2$ is the variance of $\widehat{D}$, $\sigma_{D\widehat{D}}$ is the covariance of D and $\widehat{D}$, and $c_1 = (k_1K)^2$ and

$c_2 = (k_2 K)^2$ are the constants used to maintain stability. K is the dynamic range of the pixel values, $K = 255$, $k_1 = 0.01$, and $k_2 = 0.03$.

The reciprocal of depth is used in the actual training prediction of this algorithm. D_{origin} is the original depth map and $D = M/D_{\text{origin}}$ is the target depth map; M is the maximum depth in the scene.

2.3. Coordinate System Transformation.

Camera imaging is to change the object to the photosensitive element of the camera through multiple coordinate systems, in which the coordinate systems involved are as follows: world coordinate system $(O_w - X_w Y_w Z_w)$, which describes the real position of the camera, in M; camera coordinate system $(O_c - X_c Y_c Z_c)$, the origin is the optical center and the unit is m; image coordinate system $(o - xy)$, the origin is the midpoint of the imaging plane and the unit is mm; pixel coordinate system (uv), the origin is the upper left corner of the image and the unit is pixel. As shown in Figure 3, p is a point in the world coordinate system; point p with coordinate (x, y) is the imaging point of point p in the image. (u, v) is the co-ordinate of the pixel coordinate system corresponding to the point. f is the focal length of the camera, representing the distance from o to O_c.

Through the abovementioned coordinate system conversion, a conversion Formula (4) from the pixel to the world coordinate system can be obtained:
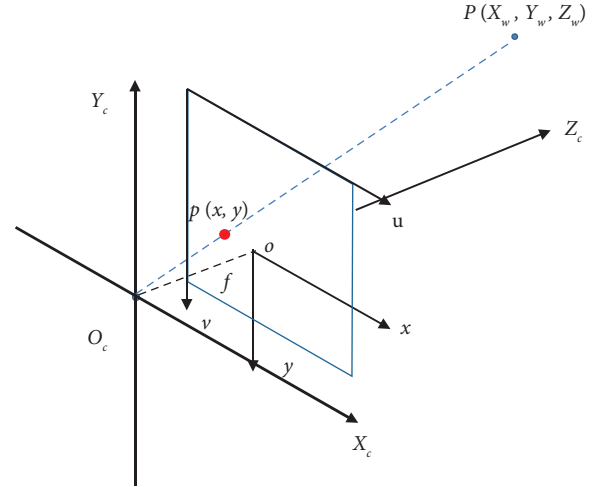

FIGURE 3: The coordinate system involved in camera imaging.

enter the charged area. The distance between coordinate points is calculated as follows:

The Euclidean distance between two points is directly obtained from the three-dimensional coordinates $(X_k, Y_k, \text{and } Z_k)$ and $(X, Y, \text{and } Z)$ of the two points in the electrically selected monocular image of the power system:

$$d = \sqrt{(X_k - X)^2 + (Y_k - Y)^2 + (Z_k - Z)^2}. \quad (6)$$

After the coordinate system conversion of two points of a manually selected power system monocular image, the distance between two points can be quickly calculated.

3. Structure Parameter Optimization and Data Enhancement of Power Density Depth Model

To meet the need for high accuracy perception of power safety distance, a regularization strategy is used to optimize the power density depth model structure. The model parameters are optimized by the migration learning strategy and cosine annealing learning strategy in order to improve generalization and detection speed. Finally, the sample data are enhanced to improve the stability and accuracy of the network. Figure 4 shows an example of actual sample data.

The monocular image of the power system usually has far and near scenes, and the texture and optical flow characteristics in the image are not obvious. Secondly, the background of the image is more complex, and the image background changes with the change of the four seasons.

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{1}{dx} & 0 & u_0 \\ 0 & \dfrac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \quad (5)$$

where the coordinate value of the coordinate point in the camera coordinate system is Z_c, which is obtained by the power density depth distance sensing model. The internal parameters of the camera are the length and width of a single pixel, represented by f_x and f_y. The central coordinate of the imaging surface is $(u_0$ and $v_0)$. The external parameters of the camera are the rotation matrix R and offset matrix T. In this paper, the internal and external parameters of the camera are obtained through Zhang Zhengyou calibration.

2.4. Safe Distance Calculation.

After the conversion of the pixel coordinate system and the real coordinate system, and the manual selection of the coordinate points, it can be determined whether the three-dimensional coordinates

3.1. Model Structure Optimization Based on Regularization.

In order to reduce the influence of model overfitting, this paper uses a regularization strategy to optimize the model structure, reduce the influence of some parameters on the model, and ensure that the model has a good training effect on the actual data set.

The specific optimization step is to add regularization terms to the loss functions of the convolutional and pooling layers to reduce the sum of values of the parameters, and to improve the accuracy and generalization of the model.

FIGURE 4: Example of actual sample data.

3.1.1. Convolutional Layer Optimization. The loss function L_{conv} can reduce the difference between output and input and simplify the feature expression. The following is the convolutional loss function Formula (6) after *L*2 regularization:

$$L_{\text{conv}}\left(W, x^{(l)}\right) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(y_l - W * x_{\text{il}}\right)^2 + \lambda R_{L2}\left(w\right)\right],$$
(7)

where the convolutional kernel parameter with quantity *n* is $W = [w_1, w_2, \cdots, w_n]^T$. The first item is the expression ability of the model. The convolutional operation is represented by $*$, and the input of sample *l* with dimension *m* is $x^l = [x_{1l}, x_{2l}, \cdots, x_{\text{ml}}]^T$. The actual depth label of the sample *l* is y_l.

The second item is the regularization term representing the complexity of the parameters. This paper uses *L*2 regularization $\lambda R_{L2}(w)$ to reduce the sum of parameter squares and prevent overfitting; λ is the regularization factor and $R_{L2}(w) = \sum_{j=1}^{n} w_j^2$.

3.1.2. Pooling Layer Optimization. The following is the loss function Formula (7) of the pooling layer after *L*2 regularization:

$$L_{\text{fc}}\left(U, x^{(k)}\right) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(y_k - Ux_{\text{ik}}\right)^2 + \lambda R_{L2}\left(u\right)\right],$$
(8)

where the first item is the expression ability of the model, and the pooling factor with the number of *t* is $U = [u_1, u_2, \cdots, u_t]^T$. The second item is the regularization term representing the complexity of the parameters, λ is the regularization factor, and $R_{L2}(u) = \sum_{j=1}^{t} u_j^2$.

When the regularization factor is too small, the reduction of model parameters is small, the model is still easy to overfit, and the model generalization is limited; when the regularization factor is too large, the number of model parameters is sharply reduced, and the whole network becomes a simple approximately linear network, the model features fitting ability is seriously reduced, and the model accuracy is decreased in detail. In order to balance the relationship between the feature-fitting ability of the model and the parametric size of the model, this paper repeatedly verifies the setting of the regularization factor size based on the sample data set to improve the accuracy and generalization of the model.

3.2. Migration Learning Shared Parameter Strategy. Migration learning, which is widely used in image recognition of power systems, will be used to train the target data using network parameters learned from source datasets [25, 26]. Migration learning can be divided into four basic methods: based on sample migration, based on feature migration, based on model migration and based on relationship migration [27]. This paper uses a model-based migration learning method by sharing the parameter information of the pretraining mode to realize the migration

from large source domain datasets to specific learning tasks in the target domain.

For real images of power systems with complex backgrounds, the convolutional neural network is difficult to extract image features effectively and accurately, and it is difficult to obtain enough samples to train the model in practice. These factors will reduce the accuracy and generalization of model distance sensing. Therefore, this paper uses the training strategies of transfer learning, deleting the top-level DenseNet-169 network pretrained on ImageNet [28] related to the original network classification task to extract features from the actual sample data as an encoder, effectively improving the accuracy and generalization of the model.

3.3. Cosine Annealing Learning Strategy.

The learning rate of a model is an important factor that affects its accuracy. The loss value of the network decreases too slowly at a lower learning rate. Networks may be trapped in local optimum or divergent when learning rates are high. During algorithm training, network parameters are set by random initialization, so in order to reduce the loss quickly, the network needs to set a larger learning rate. After several iterations, the learning rate should be reduced to avoid local optimum or divergence caused by too fast updating of network parameters. In this paper, we use the learning strategy of cosine annealing, and the formula is as follows:

$$\eta_t = \eta_{\min}^i + \frac{1}{2}\left(\eta_{\max}^i - \eta_{\min}^i\right)\left(1 + \cos\left(\frac{T_{\mathrm{cur}}}{T_i}\pi\right)\right). \tag{9}$$

In (9), i is the current index value, $\eta_{\max}^i$ and $\eta_{\min}^i$ represent the maximum and minimum learning rates, respectively, they define the range of learning rates. T_{cur} denotes the number of epochs currently being trained, and T_i represents the total number of epochs in the i-th training. The initial learning rate was set to 0.0001, the minimum learning rate to 0.00001, the maximum learning rate set to 0.001, and the training epoch to 200.

3.4. Data Enhancement.

Referring to the methods of Eigen [29], this paper expands the sample data to make the model have better generalization ability. Specific data enhancement operations include the following:

(1) Sample data are flipped horizontally, rotated 90 degrees and 180 degrees; the probability is 50%

(2) Sample data gamma values are randomly chosen from the (0.5, 1.5) range; the probability is 50%

(3) Sample data color channels are randomly multiplied by random numbers in the (0.5, 1.5) range for color adjustment; the probability is 50%

(4) Randomly add 30% noise to the sample with a 50% probability

(5) The sample data are randomly multiplied by random numbers in the (0.5, 1.5) range for brightness adjustment with a probability of 50%

As shown in Figure 5, the sample data are flipped, rotated, gamma adjusted, color channels adjusted, random noise added, and brightness adjusted.

4. Performance Test of Power Safety Distance Sensing Method Based on Power Density Depth Distance Sensing Model

In order to test the effectiveness of the power safety distance sensing method based on the power density depth model, this paper tests and compares the binocular camera [30], SFMlearner unsupervised depth estimation [31], Mono-Depth semisupervised depth estimation [32], DenseDepth supervised depth estimation, and the power security distance sensing methods based on power density depth-sensing model to verify the validity of the methods presented in this paper.

4.1. Experimentation Environment and Dataset Description.

This algorithm is based on Keras deep learning framework. The computer is configured as Windows 10 operating system, 8-core Core i7 processor, GTX2060 graphics card, 16 G memory.

As shown in Figure 6, the datasets for the experiment are obtained from the transmission corridor monitoring data from a province in China in recent years. The dataset is mainly based on the transmission channel scene with flat ground and three-dimensional buildings. There are 2025 pairs of RGB image pairs taken by binocular cameras and corresponding depth maps, ranging from 5 to 250 m, which are almost based on the point cloud data of the transmission corridor and filled in the corresponding pixel depth deficit under the guidance of the literature [33]. Based on the average depth of each pair of data, the dataset can be divided into 4 distance scenarios, respectively. The average depth and image quantity of each scenario are shown in Table 1.

4.2. Evaluating Indicator.

To evaluate and compare the performance of various depth estimation techniques, this paper adopts a common method of performance evaluation for depth estimation techniques. The method has five evaluation indices: AbsRel (absolute relative error), RMSE (root mean square error), RMSE-log (logarithmic root mean square error), SqRel (relative square error), and % correct (threshold accuracy). This accuracy measure is used as the accuracy [34] by calculating the ratio of pixels whose maximum value is less than the threshold T to the total pixels. The formulas for these indicators are as follows:

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Figure 5: Continued.

(i)

(j)

FIGURE 5: Example of actual sample data enhancement: (a) sample image, (b) rotation, (c) sample image, (d) gamma value adjustment, (e) sample image, (f) color adjustment, (g) sample image, (h) random noise, (i) sample image, and (j) brightness adjustment.

FIGURE 6: Examples of the dataset.

TABLE 1: The detail of depth images datasets.

Resolution	Average depth	Training	Testing	Total
1920 * 1080	50	213	46	259
	100	275	87	362
	150	410	157	567
	200	621	216	837
Total	150	1519	506	2025

$$\text{AbsRel} = \frac{1}{n} \sum_{p} \frac{\left|D_p - \widehat{D}_p\right|}{D_p},$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{p} \left|D_p - \widehat{D}_p\right|^2},$$

$$\text{RMSE} - \log = \sqrt{\frac{1}{n} \sum_{p} \left|\log D_p - \log \widehat{D}_p\right|^2}, \qquad (10)$$

$$\text{SqRel} = \frac{1}{n} \sum_{p} \frac{\left|D_p - \widehat{D}_p\right|^2}{D_p},$$

$$\%\text{correct} = \max\left(\frac{\widehat{D}_p}{D_p}, \frac{D_p}{\widehat{D}_p}\right) = \delta < T.$$

In the equation, D_p is the true depth value of the pixel point p in the initial depth image, $\widehat{D}_p$ stands for the estimated depth value of the pixel point p in the prediction depth image, n is the total number of pixels, and T represents the threshold value. In this paper, $T = 1.25$.

4.3. Depth Map Display.
Scene depth is the distance from the scene to the camera imaging center, which is usually visualized by depth maps. The color depth map uses color values to represent the depth of image pixels [35], as shown in Figure 7. In particular, in order to avoid the problem of too large loss function value caused by too large original depth value of transmission line scene, which affects network training, the DenseDepth depth estimation method uses the reciprocal of depth in actual training prediction, so its depth map performance result is opposite to other methods except for the binocular camera [36, 37].

4.4. Setting of Regularization Factor.
For selecting the appropriate regularization factor, this paper repeats the test verification based on the power density depth model, and the results are shown in Table 2. Table 2 shows that when $\lambda = 100$, % correct is the highest and the detection speed is relatively fast. Therefore, $\lambda = 100$ is selected for the regularization factor.
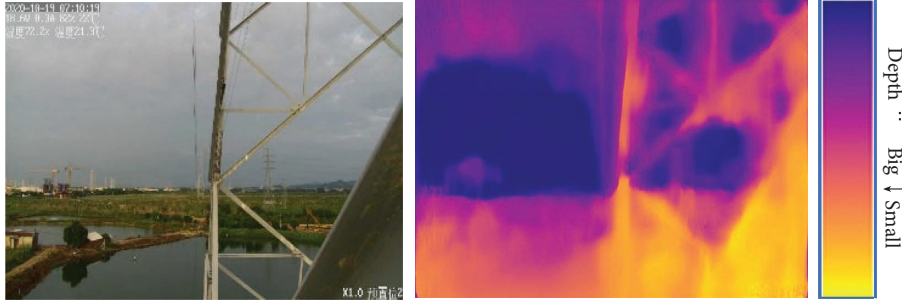
Figure 7: Color depth map representation.

Table 2: Influence of λ value on the detection precision.

λ value	Number of iterations	% correct (%)	Speed (ms per sheet)
0	5000	83.12	102
		84.46	79
50	5000	84.45	77
		84.98	63
100	5000	84.93	71
		85.81	59
200	5000	84.77	71
		84.23	66
1000	5000	83.23	61
		83.41	55

4.5. Performance Evaluation of Depth and Distance Sensing Method. In order to test the performance of the power-intensive depth model proposed in this paper, this section makes a qualitative and quantitative comparative analysis of five methods: binocular camera, unsupervised depth estimation, semisupervised depth estimation, supervised depth estimation, and the power-intensive depth model proposed in this paper. The test results are shown in Tables 2 and 3.

4.5.1. Qualitative Analysis. Table 2 shows the comparison results between the depth estimation method and other methods on the data set in this paper. Absolute relative error, root mean square error, logarithm root mean square error, relative square error, detection speed, and accurate threshold are used in Table 3. The model proposed in this paper has significantly improved in error, detection speed, and threshold accuracy. According to Table 4, the comprehensive analysis is as follows:

(1) Due to the characteristics of transmission corridor scene image texture, inconspicuous optical flow features, and large scene range, the traditional binocular camera technology and the unsupervised depth estimation SFMlearner algorithm have a threshold accuracy $\delta < 1.5$ of no more than 70%, and the processing speed is greater than 1 sec/each, making it difficult to achieve the transmission corridor scene ranging performance requirements.

(2) Semisupervised depth estimation MonoDepth achieves more accurate depth estimation by introducing the binocular right view into the model as an additional supervised signal on the basis of reducing the difficulty of data set acquisition. However, matching corresponding pixels between binocular images of transmission corridors is difficult, and the reconstruction process of this method is vulnerable in interference. Its root mean square error RMSE is 5.9764, and the percentage of pixels with large errors in the prediction results is large. Therefore, the method is less stable in the transmission corridor scenario.

(3) The supervised depth estimation DenseDepth and the model proposed in this paper directly predict the corresponding pixel depth values for the input monocular images based on real point cloud data, and their threshold accuracy reaches about 80%. By optimizing the structure of the original DenseDepth model through regularization and reasonably reducing the model parameters, the detection speed of the proposed model is improved by 31% compared with that of DenseDepth, and thanks to the optimized model training process by data augmentation strategy, migration learning strategy, and cosine annealing learning strategy; the network features are extracted and fitted well. The RMSE of the proposed model is 5.4645 and the threshold accuracy is 85.36%. In summary, the generalization and accuracy of the power density depth model are improved compared with the initial DenseDepth, which meets the requirements of the transmission corridor scenario for ranging performance.

4.5.2. Quantitative Analysis. The qualitative comparison is shown in Table 4. The power density depth model proposed in this paper has a good depth estimation effect, more local details can be obtained, and ensure that the boundary of the object is obvious. Meanwhile, it has good scene generalization and adaptability. According to Table 3, the comprehensive analysis is as follows:

(1) The model proposed in this paper can obtain more local details. As shown in the detection results of the tower crane in the third row of the image, the power line in the fourth row of the image, and power line and distant trees in the fifth row of the detection

TABLE 3: Comparison between six kinds of computer vision methods on accuracy and speed.

| Method | Lower is better | | | | | Higher is better |
	AbsRel	RMSE	RMSE-log	SqRel	Speed (per picture)	Correct (%)
Binocular camera	0.3398	7.9764	0.4283	1.8364	2.0 s	64.46
SFMlearner	0.3081	7.6983	0.4887	2.8249	1.0 s	68.42
MonoDepth	0.1898	5.9764	0.2283	0.8364	89 ms	77.88
DenseDepth	0.1550	5.8630	0.2150	0.9051	1.0 s	79.98
Power density depth model	0.1378	5.4645	0.1932	0.8363	69 ms	85.36

image in Table 4, by optimizing the model parameters through transfer learning and cosine annealing learning strategies, the feature extraction network of the proposed method can extract more deep feature information, so as to restore more scene details. It can not only estimate the depth information of smaller tower cranes, power lines, excavators, and other objects, but also better restore the scene level. However, as shown in the upper right area of the image in row 6 of Table 4, the method proposed in this paper may also cause the problem of depth estimation error.

(2) The model proposed in this paper has good depth estimation continuity. As shown in Table 4, the boundaries of the buildings in the upper left area of the detection image in row 3 and the power lines in the image in row 4 are clear and well correspond to the RGB image. This method can effectively use the shallow and deep features, reduce the loss of spatial context features and scale context features, and obtain good depth estimation performance at the object boundary.

(3) The model proposed in this paper has good scene generalization and adaptability. As shown in the image detection results in rows 1 and 2 in Table 4, the scene test image with uneven ground and no three-dimensional building is quite different from the data set in this paper. However, it is obvious that by optimizing the model structure through the regularization method, the proposed method has a good depth estimation effect. It is suitable for the depth information estimation of distant details and can better estimate the boundary depth information of objects in the scene.

In summary, the power density depth model proposed in this paper optimizes the model structure by regularization method and optimizes the model training process by migration learning and cosine annealing learning strategies, and finally the model has a more reasonable parameter training effect and quantity and has a better depth estimation effect. Meanwhile, the method can effectively utilize shallow and deep features through a skip-connected coder-decoder network, reduce the loss of spatial context features and scale context features, and estimate the depth information of distant details appropriately, which can better estimate the boundary depth

information of objects in the scene, with higher accuracy and better scene generalization and adaptability to the transmission corridor scenes.
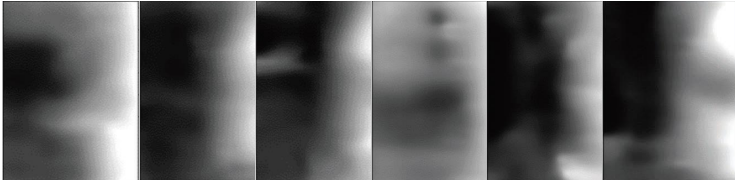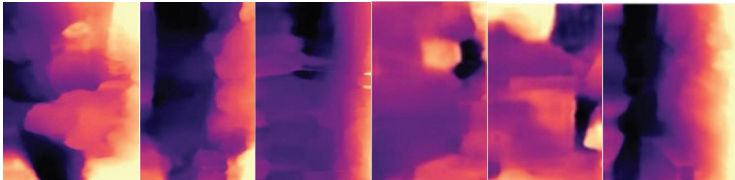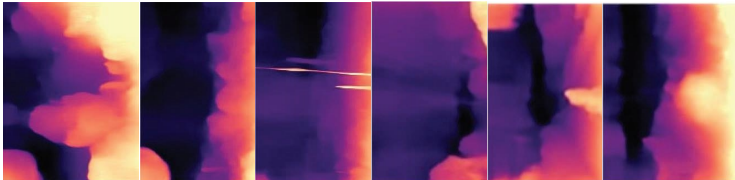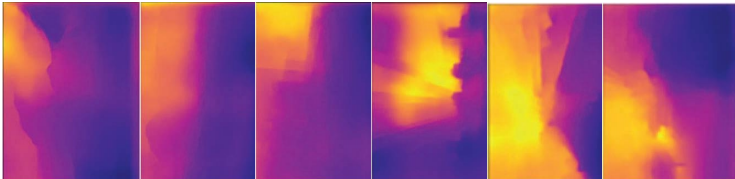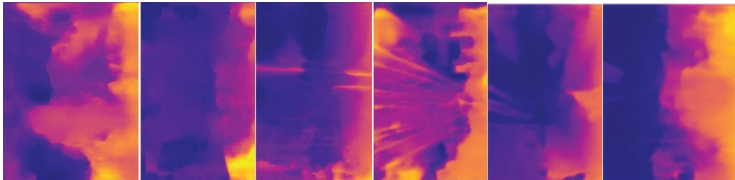
4.6. Result Analysis of Power Security Distance Sensing Method Based on Power Density Depth Model. In this paper, the point cloud data are used to calculate the two-point distance as the real distance data to realize the comparative quantitative analysis of the detection results of the power security distance sensing method based on the power density depth model. The experimental comparison results are shown in Figures 8 and 9 and Table 5.

Comprehensive analysis shows that as shown in Figure 8, when the method proposed in this paper detects the transmission channel image of flat ground and three-dimensional buildings similar to the model training data set, the relative error between the calculated safety distance and the point cloud data is the smallest, which is 11.067%. As shown in Figure 8, when the proposed method detects the transmission channel image with a large difference from the model training data set, the error is the largest, which is 24.295%. Overall, the total average relative error of the proposed method is 18.329%. While reducing the cost, the relative error difference between the proposed method and point cloud data is less than 20%. This method adopts monocular depth estimation technology based on deep learning. On the basis of reducing the cost of safe distance perception, it can perceive the distance between far and near, and improve the accuracy of monocular depth estimation through image, so as to combine detection accuracy and ease of deployment.

At the same time, there are many sources of error in the comparison of results, including the error of training data acquisition, the error caused by the calibration of internal and external parameters of the camera, the error caused by obtaining the coordinates of different objects and wire pixels in the image, and so on.

Meanwhile, this paper compares the proposed method with several commonly used transmission corridor monitoring methods, including traditional manual measurement methods, video monitoring methods, and laser point cloud diagnosis methods. The performance indicators include whether to support ranging, the relative error of ranging, false alarm rate, processing time of each image, and whether to support multiple scenes, and the comparison results are shown in Table 6.

TABLE 4: Comparison of detection results of six depth estimation methods.

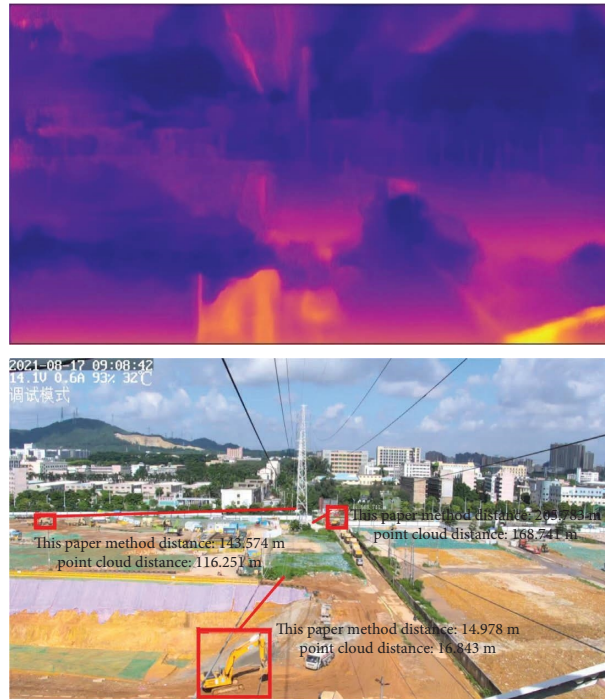| Test picture	Binocular camera	SFMlearner	MonoDepth	DenseDepth	Power density depth model

FIGURE 8: Image 1 of the safety distance sensing result of this method.

FIGURE 9: Image 2 of the safety distance sensing result of this method.

Results show that in the transmission corridor scenario, the monitoring performance of the traditional manual measurement method is fair, but the method cannot achieve real-time monitoring of the transmission corridor and cannot meet the growing demand for external breakage risk control in the transmission corridor. Although the common

monitoring methods based on video monitoring have now been used on a large scale in the transmission corridor due to their low cost, real-time monitoring, and good adaptability to multiple scenes, the method is only through image recognition technology to detect the external broken object, missing distance information; resulting in an alarm once the monitoring perspective inside the external broken object, the false alarm rate is extremely high, monitoring efficiency is low. Although the laser point cloud diagnosis method [7] has the lowest relative error in distance measurement as well as false alarm rate, the method requires a 3D point cloud model for a single transmission corridor and is based on this model for subsequent safety distance measurement of external damage. Due to the large coverage of power system transmission corridors, the method does not support multiple scenarios and real-time detection, and the application cost is extremely high, making it difficult to promote its application. In contrast, by supervised depth estimation, the proposed method can be based on the existing transmission corridor video monitoring system to achieve real-time measurement of the safety distance of external damage at a low cost. By learning the existing transmission corridor images and point cloud data, the applicability of this method can cover most transmission corridors and support multiple scenarios, while the false alarm rate of the distance measurement error of this method is relatively low, which can meet the demand of transmission corridor monitoring.

Overall, the power security distance sensing method based on the power density depth model in this paper is a general method for power security distance sensing; simply use the data corresponding to the scenarios of power security distance perception for training. In this paper, based on the

TABLE 5: Safety distance measurement result of this method.

Number	The distance measured by the method in this paper (m)	The distance measured by the point cloud data (m)	Absolute error (m)	Relative error (%)
1	10.491	12.447	1.956	15.714
2	11.732	10.333	1.399	13.539
3	143.574	116.251	11.860	23.503
4	19.952	25.607	5.655	22.084
5	30.662	40.502	9.840	24.295
6	86.859	75.721	11.138	14.709
7	14.978	16.842	1.864	11.067
8	26.714	32.027	5.313	16.589
9	29.901	24.950	4.951	19.843
10	205.783	168.741	27.042	21.951

TABLE 6: Performance comparison of different power safety distance detection methods.

Methods	Technology	Whether to measure the distance	Average relative error (%)	False alarm rate (%)	Processing speed	Whether to support multiple scenes
Traditional monitoring methods	Manual measurement	Yes	10.325	42.44	Non-real-time	Yes
Common monitoring methods	Video monitoring	No	No	67.89	Real-time	Yes
Literature [7]	LiDAR	Yes	3.8971	8.345	Non-real-time	No
The method proposed in this paper	Supervised depth estimation	Yes	11.067	15.35	Real-time	Yes

examination of power system single objective images to reduce the cost of the perception of the safe distance, the method can effectively discover the safe distance between manually selected monitoring points. At the same time, it can have high accuracy and speed of the measurement of the safe distance.

5. Conclusion

To tackle the problem of the recognition of the safety distance of external damage, a power safety distance sensing method based on monocular visual images is proposed to achieve the recognition of safety distance of external damage in a complex scene of transmission corridors. The specific conclusions are as follows:

(1) A power safety distance sensing method based on supervised depth estimation is constructed, which can be used to obtain spatial distance information at low cost by inputting monocular images and realize the safety distance of detecting external damage based on the existing monitoring system, which can effectively improve the monitoring efficiency

(2) A power density depth distance model based on the convolutional neural network is proposed and optimized by the regularization method, migration learning strategy, cosine annealing learning strategy, and data enhancement strategy, which can obtain spatial distance information in complex scenes of transmission corridors while maintaining good accuracy and generalizability.

However, when the proposed method is used for feature extraction, it is easy to ignore the features of the prospective part, resulting in feature loss. This problem will be improved in the follow-up work.

Data Availability

The data used to support the findings of this study are included within the paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] F. Q. Ma, B. Wang, X. Z. Dong, H. G. W., P. Luo, and Y. Y. Zhou, "Safety image interpretation of power industry: basic concepts and technical framework," *Proceedings of the CSEE*, vol. 21, pp. 1–17, 2021.

[2] F. Q. Ma, B. Wang, X. Z. Dong, H. G. W., P. Luo, and Y. Y. Zhou, "Power vision edge intelligence: power depth vision acceleration technology driven by edge computing," *Power System Technology*, vol. 44, no. 6, pp. 2020–2029, 2020.

[3] P. E. N. G. Xiangyang, C. H. E. N. Chi, and X. U. Xiaogang, "Transmission corridor safety distance diagnosis based on point cloud and unmanned aerial vehicle loaded airborne

laser scanning," *Power System Technology*, vol. 18, no. 11, pp. 3262–3267, 2014.

[4] H. Zhao, Y. Zhao, and Y. Jiang, "Supervision and management integrated safety management system based on electric power internet of things[J]," *Electric Safety Technology*, vol. 22, no. 9, pp. 22–25, 2020.

[5] Y. I. N. Jinhua and S. U. N. Chaoyang, *Z H E N G Yanchun*, vol. 28, no. 7, Beijing, China, Beihang University, 2007

[6] H. U. A. N. G. Junjie, L. I. U. Xiaobo, and A. O. Yu, "Research of binocular vision monitoring schemes based on transmission line passage ways," *Electrical Automation*, vol. 41, no. 1, pp. 38–40, 2019.

[7] C. Chen, X. Y. Peng, S. Song, K. Wang, J. J. Qian, and B. S. Yang, "Safety distance diagnosis of large scale transmission line corridor inspection based on LiDAR point cloud collected with UAV," *Power System Technology*, vol. 39, no. 8, pp. 39–42, 2012.

[8] Z. Xu, M. Liu, G. Yang, and N. Li, "Application of interval analysis and evidence theory to fault location," *IET Electric Power Applications*, vol. 3, no. 1, pp. 77–80, 2009.

[9] J. Jiang and X. Zhang, "Depth estimation methods based on computer vision," *Electro-Optic Technology Application*, vol. 26, no. 1, pp. 51–55, 2011.

[10] L. Gao, *Algorithm Research and Design of Transmission Line Video Anti-break System*, LanZhou University of Technology, Lanzhou, China, 2018.

[11] X. Han, *Research on Measurement Method of Distance for Crisscross Span of Overhead Transmission Line Based on Machine Vision*, North China Electric Power University, Beijing, China, 2016.

[12] P. Luo, B. Wang, and H. Ma, "Defect recognition method with low false negative rate based on combined target detection framework," *High Voltage Engineering*, vol. 47, no. 2, pp. 454–464, 2021.

[13] A. Saxena, S. H. Chung, and A. Y. Ng, "Learning depth from single monocular images," in *Proceedings of the 18th International Conference on Neural Information Processing Systems Conference (NIPS)*, pp. 1161–1168, Cambridge, UK, June 2005.

[14] F. Dellaert, S. M. Seitz, C. E. Thorpe, and S. Thrun, "Structure from motion without correspondence," in *Proceedings of the 2000 IEEE conference on computer vision and pattern recognition (CVPR)*, vol. 15, pp. 557–564, Hilton Head, SC, USA, June 2000.

[15] L. I. Yang, X. Chen, Y. Wang, and M. Liu, "Progress in deep learning based monocular image depth estimation," *Laser and Optoelectronics Progress*, vol. 56, no. 19, pp. 9–25, 2019.

[16] T. Bi, Y. Liu, D. Weng, and Y. Wang, "Survey on supervised learning based depth estimation from a single image," *Journal of Computer-Aided Design and Computer Graphics*, vol. 30, no. 8, pp. 1383–1393, 2018.

[17] S. F. Bhat, I. Alhashim, and P. Wonka, "AdaBins: depth estimation using adaptive bins," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4009–4018, Virtual, July 2021.

[18] J. Watson, O. Mac Aodha, V. Prisacariu, B. Gabriel, and M. Firman, "The temporal opportunist: self-supervised multi-frame monocular depth," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1164–1174, Virtual, May 2021.

[19] V. Guizilini, J. Li, R. Ambrus, S. Pillai, and A. Gaidon, "Robust semi-supervised monocular depth estimation with reprojected distances," in *Proceedings of the Conference on Robot Learning*, pp. 503–512, Osaka, Japan, September 2020.

[20] C. Shi, *Depth Estimation of Monocular Image Based on Deep Learning*, Beijing University of Chemical Technology, Beijing, China, 2020.

[21] Z. Liu, X. Miao, J. Chen, and H. Jiang, "Review of visible image intelligent processing for transmission line inspection," *Power System Technology*, vol. 44, no. 3, pp. 1057–1069, 2020.

[22] I. Alhashim, P. Wonka, I. Alhashim, and P. Wonka, "High quality monocular depth estimation via transfer learning," 2018, https://arxiv.org/abs/1812.11941.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and Q. Kilian, "Weinberger. Densely connected convolutional networks," in *Proceedings of the 2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, Honolulu, HI, USA, July 2016.

[24] W. Zhou, A. C. Bovik, H. R. Sheikh, and P. E. Simoncelli, "Image quality assessment: from error visibility to structural similarit," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 2004.

[25] M. A. Peng and Y. Fan, "Small sample smart substation power equipment component detection based on deep transfer learning," *Power System Technology*, vol. 44, no. 3, pp. 1148–1159, 2020.

[26] H. Yan and J. Chen, "Insulator string positioning and state recognition method based on improved YOLOV3 algorithm," *High Voltage Engineering*, vol. 46, no. 2, pp. 423–431, 2020.

[27] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[28] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "Imagenet: a large-scale hierarchical image database," in *Proceedings of the 2009 ieee conference on computer vision and pattern recognition (cvpr)*, pp. 248–255, Miami, FL, USA, June 2009.

[29] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," in *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS)*, pp. 2366–2374, Cambridge, MA, USA, December 2014.

[30] X. Li and D. Zhang, "Binocular image ranging algorithm baseol on downsampling clustering," *Journal of Liaoning University Natural Sciences Edition*, vol. 47, no. 3, pp. 277–283, 2020.

[31] T. Zhou, M. Brown, N. Snavely, G. David, and Lowe, "Unsupervised learning of depth and ego-motion from video," in *Proceedings of the The 30th IEEE Conference on Computer Vision and Pattern Recognition*, vol. 7, Honolulu, HA, USA, July 2017.

[32] P. Chen, A. H. Liu, Y.-C. Liu, and Y. C. F. Wang, "Towards scene understanding: unsupervised monocular depth estimation with semantic-aware representation," in *Proceedings of the 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 2619–2627, Long Beach, CA, USA, June 2019.

[33] Y. Zhang and F. Thomas, "Deep depth completion of a single RGB-D image," in *Proceedings of the 2018 IEEE/CVF*

conference on computer vision and pattern recognition (CVPR), pp. 175–185, Salt Lake City, UT, USA, June 2018.

[34] H. Xu, *Research on Depth Estimation Algorithms for Monocular Image*, Shandong University, Jinan, China, 2018.

[35] L. I. U. Yi-ying, *Depth Estimation from Monocular Image Based on Deep Convolutional Neural Networks*, Xidian University, Xi'an, China, 2019.

[36] F. Ma, B. Wang, J. Zhou et al., "An effective risk identification method for power fence operation based on neighborhood correlation network and vector calculation," *Energy Reports*, vol. 7, no. 2021, pp. 6995–7003, 2021.

[37] J. Liu, R. Jia, W. Li et al., "High precision detection algorithm based on improved RetinaNet for defect recognition of transmission lines," *Energy Reports*, vol. 6, pp. 2430–2440, 2020.