A Joint Faster RCNN and Stereovision Algorithm for Vegetation Encroachment Detection in Power Line Corridors

Shuaiang Rong
Department of Electrical and Computer Engineering
University of Illinois at Chicago
Chicago, USA
srong4@uic.edu

Lina He
Department of Electrical and Computer Engineering
University of Illinois at Chicago
Chicago, USA
lhe@uic.edu

***Abstract*— With the rapidly increasing construction of power lines in complex terrains, vegetation encroachment of overhead power lines can cause severe consequences, e.g., wildfires and cascading failures. Therefore, it is significant to detect vegetation encroachment of overhead power lines to ensure secure and reliable operation of power systems. This paper proposes an automatic joint algorithm to fast and accurately detect vegetation encroachment of overhead power lines in real-time. This algorithm integrates a deep learning method, i.e., the region convolution neural network (Faster RCNN) into a stereovision algorithm to process the 2D image data captured by the vision sensors located on a power tower. The generated bounding boxes (bboxes) only can represent vegetation regions in 2D images that contain the features of vegetation. The stereovision algorithm is applied in this paper to process the bboxes in 2D images to generate the corresponding 3D information for vegetation location, including the height of the vegetation. The proposed algorithm has been tested on the realistic vegetation encroachment image data from an utility. The results show that the proposed algorithm can fast and accurately detect vegetation encroachment of overhead power lines.**

Keywords—deep learning, vegetation detection, Faster RCNN, stereovision algorithm

I. INTRODUCTION

The reliable power supply greatly depends on secure and stable operation of power lines. In current power systems, overhead power lines are widely used in transmission and distribution systems as key components of power systems. With the rapidly increasing consumption of electric power, the operation environment of overhead power lines has become more complex especially in the undeveloped natural areas. The vegetation encroachment of overhead power lines poses severe threats to the bare conductors of overhead power lines due to the growth of vegetation and wind conditions. The accidents caused by the conflicts of external forces between the vegetation and overhead power lines have become the prominent issues, which can lead to serious consequences such as wildfires and cascading failures.

In recent years, wildfires caused by short circuit faults of overhead power lines in vegetarian areas commonly happen. Among them, the Camp Fire in 2018 is the most destructive wildfire in California history and caused at least 85 casualties and destroyed 14,000 residences [1]. According to the investigation, the electric utility, i.e. the Pacific Gas and Electric (PG&E), was accused of being responsible for this catastrophe since wind-driven blazes was sparked by the PG&E-owned power lines. As a consequence of the $30 billion liability, the PG&E announced bankruptcy in the early of 2019 [2]. To prevent this kind of catastrophes from happening again, a fast and accurate inspection method of vegetation encroaching power lines is strongly required.

The traditional inspections are mainly conducted by manual methods, i.e. patrols on foot or helicopters. However, these methods demand much labor force and time. They usually have a high rate of false and missing judgments on the inspection of vegetation encroachment due to the limited human eyesight. Therefore, those traditional inspections cannot meet the requirements of secure and reliable operation of power lines with the rapidly increasing overhead power lines nowadays. To reduce the labor and time of inspection, vision sensors, such as, surveillance cameras, have been mounted on power towers to obtain image or video data for inspection. However, the received image and video data of overhead power line corridors from vision sensors are still processed manually in present.

To improve the efficiency and accuracy of vegetation encroachment detection in power line corridors, several automatic methods have been discussed in the existing literatures. The airborne light detection and ranging (LiDAR) method has a good performance of detection of vegetation encroachment [3]. The work of [4] applied unmanned aerial vehicles (UAVs) for the LiDAR inspection to improve efficiency. However, the high costs of the light sensors and UAVs lead to great economic burdens. Alternatively, the image recognition received a great attention in the vegetation detection in power line corridors with advanced image processing technologies [5-6]. Some researchers studied on the application of image recognition on vegetation detection in overhead power line corridors using image data from drones and satellites [7-8]. The applied methods in those papers mainly focus on the recognition of the tree crowns in aerial images, which cannot achieve an accurate and efficient monitoring due to the long distance from satellites and the patrol routines [9]. In order to achieve the real-time monitoring of power lines, visions sensors are mounted on power towers to capture image and video data. However, it is very challenging to recognize vegetation in the image data received from vision sensors due to the complex backgrounds and wide variety of morphology of trees. To the best of the authors' knowledge, this challenge has not been well solved yet. In recent years, deep learning has greatly improved the performance of image recognition and provides opportunities to overcome this challenge [10-11].

To fill this gap, this paper proposes a joint algorithm that integrates a deep learning algorithm into a stereovision algorithm for online vegetation encroachment detection. This proposed algorithm is applied to detect vegetation in the 2D images captured by the vision sensors mounted on power towers. The deep learning algorithm used in this paper is the Faster RCNN, which is developed based on the regional convolution neural network (RCNN) and the Fast RCNN [12-14]. Instead of using the selective search (SS) to propose region proposals, the RPN is proposed in the Faster RCNN to improve the detection speed greatly [15]. This paper applies

978-1-7281-5508-1/20/$31.00 ©2020 IEEE

the Faster RCNN algorithm to process the 2D images to generate the bboxes, which represent the regions of vegetation in 2D images that contain the features of vegetation. In order to obtain the corresponding 3D information for the vegetation location, including the height of vegetation, this paper applied a computer vision algorithm, i.e. the stereovision algorithm to process the bboxes [16].

This paper is organized as follows. Section II outlines the proposed algorithm and introduces the experimental prototype. Section III introduces the principles and process of the Faster RCNN. Section IV shows the testing results of vegetation detection and discusses the effectiveness of the proposed algorithm. Section V is the conclusion.

II. Methodology and Experimental Prototype

A. Outline of the proposed algorithm

The proposed joint algorithm mainly contains two parts, i.e. the stereovision algorithm and the Faster RCNN algorithm. The structure of the proposed algorithm is illustrated in Fig. 1.

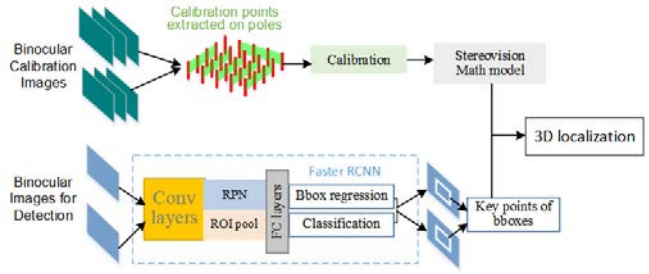

Fig. 1. Structure of proposed algorithm

Firstly, in order to generate the 3D information of the target object, the stereovision model is built. The stereovision model can project information in 2D binocular images to corresponding 3D infromation. This model is constructed through the calibration of the binocular cameras based on Zhang's calibration method [16]. This method uses a checkerboard to calibrate the cameras. Usually, the scale of the monitoring scene in power line corridors is very large. The distance between two power towers can be up to several hundred meters. As a result, a checkerboard is not practical for calibration in this large-scale 3D space. Therefore, this paper applies a calibration pole to form 2D planes in the large area.

Then the binocular images are sent to the Faster RCNN to detect the regions of vegetation in the 2D images. The Faster RCNN mainly consists of two steps. In the first step, the RPN is used to generate bboxes representing region proposals of the target object. Afterwards, the Fast RCNN is used to classify and modify bboxes. At the end, the two sets of key points of bboxes of the highest score from 2D images are then sent to the stereovision model to generate the corresponding 3D information.

B. Experimental prototype

In order to develop the stereovision model, the images of calibration objects need to be generated and processed using the calibration algorithm. This experiment is carried out to generate the calibration image data to develop the stereovision model. The experimental prototype is set up in an experimental laboratory. Binocular vision sensors are two surveillance cameras mounted on a DJ 300 power tower. The received images from those cameras are transmitted to the base station for analysis using a local area network (LAN). The bandwidth of the LAN is 100Mbps (megabits per second). In practice, the rate is usually 6~10MB/s. Since a picture captured from the camera is about 280KB with the size of 1920×1080 pixels, the possible maximum packet transmission time of a picture from the vision sensors to the base station is 0.046s. This time is nearly the same as the frame rate of the video from the surveillance cameras. Therefore, the LAN can support real-time detection of vegetation in power line corridors.

In this paper, the image data are obtained on the operation platform in the base station by intercepting photos and recording videos. The cameras are set approximately 13 meters high from the ground， and two cameras adopt convergent stereopsis to ensure a shared view of overlooking the power line corridor. The experimental setup is shown in Fig. 2.

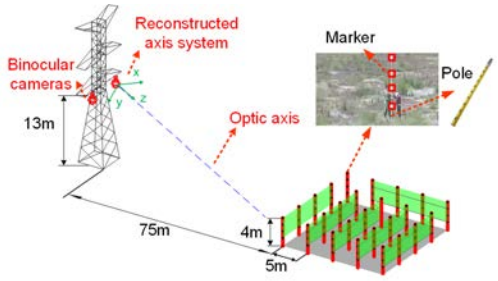

Fig. 2. Experimental setup

The calibration pole is stretched to 4 meters and placed on the ground vertically with a white marker on the point of each meter, which is extracted as calibration points. The pole is then placed transversely and longitudinally every 5 meters to form a matrix comprising 6 calibration planes shown as green in Fig. 2. With the extracted calibration points, the stereovision model consists of internal and external parameters of the binocular cameras can be calculated. It is noted that the reconstructed coordinate system is based on the left camera coordinate system in this stereovision model.

III. Vegetation Detection Using Faster RCNN

According to the proposed algorithm, the 2D information of the vegetation need to be generated to obtain the corresponding 3D information of the vegetation. In this paper, the Faster RCNN is applied to recognize the vegetation based on the features of vegetation and generate the bboxes that represent the vegetation regions in 2D images.

A. Faster RCNN

The Faster RCNN mainly consists of the RPN and the Fast RCNN. They both share all the data of the convolution layers. The structure of the Faster RCNN is shown in Fig. 3.

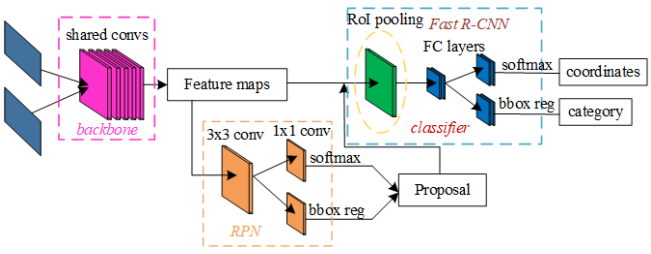

Fig. 3. Structure of Faster RCNN

The RPN predicts the regions and identify the probability of the region containing a target object. Then region proposals of the target object are generated. They are sent to the Fast RCNN block for object recognition and detection.

In the process of the Faster RCNN, binocular images are firstly sent to the shared conv layers to exact feature maps. The shared conv layers contain convolution layers, rectified linear unit (ReLU) layers, and pooling layers. The feature maps are shared with the subsequent RPN and fully connected (FC) layers [17]. In the RPN, a softmax layer is used to determine whether the region proposals contain the object, and then the bbox regression layer is used to modify the regions to obtain accurate proposals. The region of interest (RoI) pooling layer collects the input feature maps and proposals, extracts the proposed feature maps, and sends them to the subsequent FC layers in the Fast RCNN. Then the category and final exact position of the bboxes are obtained from the softmax and bbox regression layers, respectively.

The defined CNN in this paper is pre-trained with the CIFAR-10 database including 50,000 images with 10 categories [18]. This database is selected due to the small size of the images, which is 32×32 pixels. This can help to decrease the training time on a limited graphic processing unit (GPU). This paper develops a CNN, which consists of three parts. An image input layer is used in the first part of CNN to restrict the type and size of the input images. Then 3 convolutional layers, each has 32 5×5 filters and is followed by a ReLU layer and a pooling layer, are used in the middle part. The final layers consist of FC layers with 64 neurons and a softmax loss layer.

B. RPN

In the RPN, a sliding window as a 3×3 convolution kernel is used to slide on the feature map, and the center corresponds to a position on the map. In order to detect the object in different sizes and shapes, each sliding window are set to use 3 different sizes and aspect ratios of anchor boxes in each position, which is shown on the right of the Fig. 4.

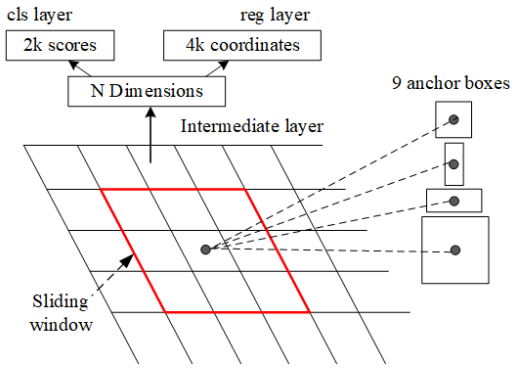

Fig. 4. Anchors in RPN

Then 9 anchors are generated in each position to predict the bboxes in different sizes and shapes of the input images. As a result, 9×H×W×N region proposals are generated, where the symbols H and W represent the height and width of the feature map respectively and the symbol N is the dimension of the feature map. Then the data of the region proposals are transformed to a N dimension feature vector to be fed into the two FC networks, i.e. the classification (cls) layer and the regression (reg) layer to determine the category and regress the bboxes of the anchors. The reg layer has 4×9 outputs that have 4 coordinate parameters representing the location of the bboxes of the 9 proposals in the original graph. The cls layer has 2×9 outputs, which determines whether the original map area contains the object. The principle of this process are introduced as follows.

The training of RPN is a process to minimize the loss function of RPN,which is represented in (1),

$$L(\{p_i\},\{u_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

In which the symbol i represents the index of the anchor, the symbol p_i represents the probability of containing the object, and the symbol p_i^* represents the label of the anchor. In each position, labels are assigned to the anchors. The positive label represented by 1 is assigned to the anchor when the intersection-over-union (IoU) overlap of the anchor with a ground truth box (GT) is the highest or greater than 0.7. The negative label is assigned when the IoU of the anchor is less than 0.3. The symbol t_i represents coordinates of the predicted bbox represented by (t_x, t_y, t_w, t_h), the symbol t_i^* represents coordinates of the GT corresponding to the positive anchors, and the term $L_{cls}(p_i, p_i^*)$ represents the log loss over object and non-object categories, which is expressed by (2),

$$L_{cls}(t_i, t_i^*) = -\log[p_i^* p_i + (1-p_i^*)(1-p_i)] \quad (2)$$

and the term $L_{reg}(t_i, t_i^*)$ represents the regression loss function of reg layer, which is expressed by (3).

$$L_{reg} = R(t_i - t_i^*) \quad (3)$$

In which the symbol R represents the $smooth_{L_1}$ function defined by (4).

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & otherwise \end{cases} \quad (4)$$

The L_1 norm loss function is chosen since it's less sensitive to the outliers of data. The $smooth_{L_1}$ function adjusts L_1 norm loss to be smooth on zero point to get only one derivative value for better converge results. The symbols $\{p_i\}$ and $\{u_i\}$ represent the outputs of the cls and reg layers, respectively. The symbols N_{cls} and N_{reg} are the normalization value of the cls and reg layers, respectively. The symbol λ represents an average weight, which is with a default of 10 and is aimed to scale the outputs of the cls and reg layers to the same numerical level.

C. Alternating training

During the training of the Faster RCNN, the RPN and Fast RCNN are trained alternatively to share convolution layers. The first step is to train the RPN with the pre-trained network and the network model is updated. The second step is to train the Fast RCNN with the same pre-trained model, and the network is updated again, but the convolution layers are not shared yet. The third step is to initialize PRN training using the network model from the step two. Meanwhile, the shared convolution layers are fixed and unique layers of the RPN are fine-tuned. The fourth step is to fine-tune the unique layers of the Fast RCNN with the shared convolution layers that has been fixed.

IV. CASE STUDY

A case study is carried out to verify the effectiveness and feasibility of the proposed algorithm. The image data received from the vision sensors mounted on a power tower are postprocessed with the proposed algorithm to generate the stereovision model and the trained Faster RCNN model. Then a test is carried out to verify the effectiveness of the models. The 2D images that contain features of vegetation are processed by the Faster RCNN algorithm to generate bboxes, which represent the vegetation regions. Then corresponding 3D information are generated with the stereovision model for location and height measurement. The applied images for testing are realistic vegetation encroachment image data, which are from the power line monitoring system of an utility.

A. Stereovision calibration

With the extracted calibration points, the calibration is carried out to generate stereovision model. The visualization of external parameters of the stereovision model that represent the positions of the cameras and planes is shown in Fig. 5.

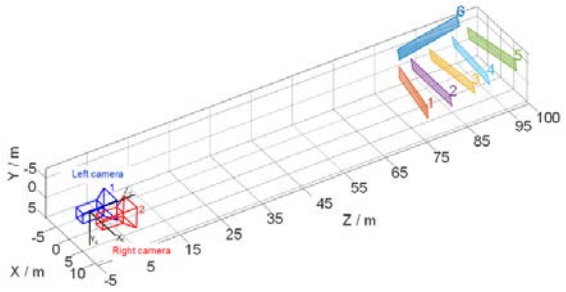

Fig. 5. External parameters of stereovision

Since the positions of the objects in the reconstructed coordinate system in Fig. 5 match with the real positions in Fig. 2, an effective stereovision model is accomplished.

B. Vegetation detection

The CNN is first trained with CIFAR database. Then for the training of vegetation detection, the online image samples of trees in different sizes are collected and processed manually into a training data set. A few training samples of trees are presented in Fig. 6.

Fig. 6. Samples of tree data

With the training data fed into the pre-trained network, a Faster RCNN detector is generated to be used for vegetation detection. The images applied for vegetation encroachment are captured from a realistic monitoring system.

Fig. 7. Rectification of images with epipolar lines

As shown in Fig. 7, images are firstly rectified with the epipolar lines to eliminate vertical parallax, which can simplify the reconstruction calculation of stereovision [16]. After the images are fed to the Faster RCNN detector, several region boxes are generated, and the corresponding bboxes with the highest score are selected for reconstruction by the stereovision, which can be seen from Fig. 8.

Fig. 8. Region detection of vegetation using Faster RCNN

In Fig. 8, the corresponding regions of the vegetation are successfully detected. 6 key points in the corner and middle of the boundary lines of the region boxes are selected as feature points of the vegetation in 2D images. Then the key points are processed with the stereovision model to generate the corresponding 3D information of the vegetation. The image coordinates of key points are calculated with the coordinates (x, y, w, h) of the bboxes, which represent the coordinates (x, y) of the upper left corner points, the width coordinate (w), and the height coordinate (h) of the bboxes. The results of 2D image coordinates and 3D coordinates of the key points are shown in TABLE I.

TABLE I. MEASUREMENT RESULTS OF KEY POINTS

Points	Image coordinates (pixel)				3D coordinates (m)		
	Left camera		*Right camera*				
	x	y	x	y	x	y	z
1	166	189	11	195	-7.4131	-1.3642	28.2804
2	306	189	122	195	-3.2124	-1.2034	24.9227
3	446	189	234	195	0.1023	-1.0751	22.2454
4	166	478	11	477	-7.7001	6.1394	29.1216
5	306	767	122	758	-3.3890	12.1033	25.7474
6	446	478	233	477	0.0843	4.7456	22.4961

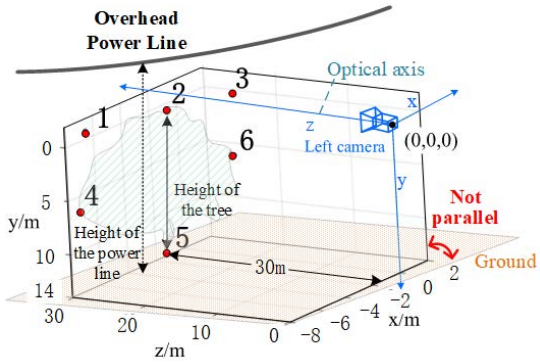

Fig. 9. 3D coordinates of key points in reconstruction coordinate system

It can be seen that the 6 key points are successfully reconstructed as shown in Fig. 9. The reconstruction

coordinate system is based on the left camera coordinate system. In the 3D coordinates shown in Fig. 9, the coordinates of the camera are set as (0, 0, 0). The Fig. 9 also shows that point 5 at the bottom of the tree is 30 meters away from the camera on the tower, which matches with reality. Considering the optical axis (i.e., the z axis) is not parallel with ground, so the reconstructed coordinate system is also not based on the ground, which indicates the X-Z plane is not parallel with the ground and the y axis is not vertical to the ground. Therefore, the tree is "floating" in this reconstructed system instead of vertically standing on the ground. As a result, the height of the tree needs to be calculated by calculating the distance between the top point 2 and bottom point 5. In (5), the symbol Dab is used to represent the distance between two points a and b, whose coordinates are represented by (xa, ya, za) and (xb, yb, zb), the distance between 2 points can be obtained below,

$$Dab = \sqrt[2]{(xa - xb)^2 + (ya - yb)^2 + (za - zb)^2} \quad (5)$$

After calculation, the distance between point 2 and point 5 is 13.3334 m. Comparing with the realistic measurement of height of the tree that is 12.9 m, the result only has an error of 3.36 %. Since the shape and the background of the power lines are both simple, the features of the power lines can be easily detected from the 2D images. Through reconstruction, the height of the power lines can be easily obtained. With the proposed algorithm, the height of the tree is calculated. Comparing the heights of the vegetation and power lines, the vegetation encroachment of the power lines can be effectively detected.

V. Conclusion

The vegetation encroachment poses a severe threat to the safe operation of overhead power lines. To solve this problem, this paper proposed a joint algorithm that integrates the Faster RCNN algorithm with the stereovision algorithm for vegetation encroachment detection in power line corridors. Based on the binocular images from vision sensors mounted on the power tower, the bboxes in 2D images that contain features of vegetation are obtained through the Faster RCNN algorithm. Then the corresponding 3D information for vegetation location are generated by the stereovision model based on the bboxes in 2D images. The results showed that the proposed joint Faster RCNN and stereovision algorithm can fast and accurately detect vegetation encroachment in power line corridors.

References

[1] R. V. Alejandra, "California's Camp fire was the costliest global disaster last year, insurance report shows," *Los Angeles Times*, Los Angeles, Jan. 11, 2019.

[2] J. Efstathiou, and R. Varghese, "A PG&E bankruptcy may be what California needs for a utility fix," *Bloomberg News*, New York, Jan. 14, 2019.

[3] S. Ashidate, S. Murashima and N. Fujii, "Development of a helicopter-mounted eye-safe laser radar system for distance measurement between power transmission lines and nearby trees," *IEEE Trans. Power Del.*, vol. 17, no. 2, pp. 644–648, Apr. 2002.

[4] S. J. Mills, M. P. Castro; Z. Li, et al., "Evaluation of aerial remote sensing techniques for vegetation management in power-line corridors," *IEEE Trans. Geosci. Remote Sens*., vol. 48, no. 9, pp. 3379–3390, Sept. 2010.

[5] J. Ahmad, A. S. Malik, L. Xia, et al., "Vegetation encroachment monitoring for transmission lines right-of-ways: A survey," *Electric Power Syst. Research*, vol. 95, pp. 339–352, Feb. 2013.

[6] F. Mirallès, N. Pouliot and S. Montambault, "State-of-the-art review of computer vision for the management of power transmission lines," In *Proc. Int. Conf. Appl. Robotics Power Industry*, Foz do Iguassu, Brazil, 2014, pp. 1–6.

[7] C. Sun, R. Jones, H. Talbot, et al., "Measuring the distance of vegetation from powerlines using stereo vision," *ISPRS Journal Photogrammetry Remote Sens.*, vol.60, no. 4, pp. 269–283, June 2006.

[8] Y. Kobayashi, G. G. Karady, G. T. Heydt and R. G. Olsen, "The Utilization of Satellite Images to Identify Trees Endangering Transmission Lines," *IEEE Trans. Power Del.*, vol. 24, no. 3, pp. 1703–1709, July 2009.

[9] Z. Li, R. Hayward, J. Zhang, and Y. Liu, "Individual tree crown delineation techniques for vegetation management in transmission line corridor," In *Proc. Digital Image Comp. Tech. Appl.*, Canberra, Australia, 2008, pp. 148–154.

[10] K. He, X. Zhang, S. Ren, and J.Sun, "Deep residual learning for image recognition," In *Proc. The IEEE Conf. Comp. Vision Pattern Recognit. (CVPR)*, Las Vegas, United States, 2016, pp. 770–778.

[11] V. N. Nguyen, R. Jenssen, and D. Roverso, "Automatic autonomous vision-based transmission line inspection: A review of current status and the potential role of deep learning," *Int. Journal Elec. Power & Energy Syst.*, vol. 99, pp. 107–120, 2018.

[12] S. Ren, K. He, R. Girshick, et al., "Faster R-CNN: Towards real-time object detection with region proposal networks," *Neural Inf. Processing Syst. Conf.*, 2015, pp. 91–99.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," In *Proc. IEEE Conf. Comp. Vision Pattern Recognit.*, Columbus, OH, 2014, pp. 580–587.

[14] R. Girshick, "Fast R-CNN," In *Proc. IEEE Int. Conf. Comp. Vision*, 2015, pp. 1440–1448.

[15] J. R. Uijlings, K. E. A. Van De Sande, T. Gevers, et al. "Selective search for object recognition," *Int. Journal Comp. Vision*, vol.104, no.2, pp.154–171, Sep. 2013.

[16] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol.22, no. 11, pp. 1330–1334, Dec. 2000.

[17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," In *Proc. IEEE Conf. Comp. Vision Pattern Recognit.*, Boston, United states, Jun. 2015, pp. 3431–3440.

[18] A. Krizhevsky. "Learning multiple layers of features from tiny images," M.S. thesis, Dept. of Comp. Science, Univ. Toronto, Canada, 2009.

Shuaiang Rong received the B.S degree in Automation in 2015 and M.S. degree of electric engineering in 2018 both from Shanghai University of Electric Power. She is currently pursuing Ph.D. degree in electrical engineering at the University of Illinois at Chicago. Her research interest includes machine learning, image processing, renewable energy integration and power systems.

Lina He received the Ph.D. degree in electrical engineering at University College Dublin, Ireland in 2014, and the B.S. and M.S. degrees from Huazhong University of Science and Technology, China, in 2007 and 2009, respectively. She is currently an Assistant Professor in department of electrical and computer engineering, University of Illinois at Chicago, Chicago, USA. She was a Project Manager and Senior Consultant at Siemens headquarter in Germany and Siemens US from 2014 to 2017. Her research interests include renewable energy integration, power systems and coordination with power electronics, wide-area protection and cybersecurity, machine learning, and image processing,.