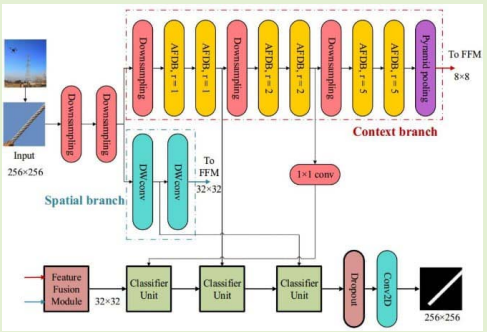
Efficient Parallel Branch Network With Multi-Scale Feature Fusion for Real-Time Overhead Power Line Segmentation

Zishu Gao, Guodong Yang, *Member, IEEE*, En Li, *Member, IEEE*, Zize Liang, and Rui Guo

***Abstract*—Image-based segmentation of overhead power lines is critical for power line inspection. Real-time segmentation helps the inspection robot avoid obstacles or land on the wire during the inspection task. It is challenging for several studies to achieve real-time overhead power line segmentation with high accuracy. In addition, cluttered background brings great difficulties to overhead power lines segmentation. To address these issues, an efficient parallel branch network for real-time overhead power line segmentation is proposed. Our framework combines a context branch that generates useful global information with a spatial branch that preserves high-resolution segmentation details. The asymmetric factorized depth-wise bottleneck (AFDB) module is designed in the context branch to achieve more efficient short-range feature extraction and provide a large receptive field. Furthermore, the subnetwork-level skip connections in the classifier are proposed to fuse long-range features and lead to high accuracy. Experiments demonstrate that our framework achieves more than 90% segmentation accuracy.**



***Index Terms*—Real-time segmentation, lightweight network, dilated depth-wise convolution, power line inspection.**

I. Introduction

OVERHEAD power line inspection is a critical task in power grid systems [1]. Typically, the inspection is performed by power line inspection workers, but manual inspection can be dangerous and time-consuming. The smart grid has developed a lot with the advancement of robots and UAVs [2], and robot inspections gradually replaced manual inspections. The useful power inspection sensors and methods play a role in promoting power line segmentation. They not only make power line segmentation more and more accurate, but also enables real-time segmentation. Real-time segmentation helps the inspection robot avoid obstacles or land on the wire during the inspection task.

An overhead power line can be seen as a line when viewed from a distance [3]. With the development of inspection robots and sensors, many traditional detection approaches have been formulated. Some studies have used millimeter-wave radar [4], [5] to detect overhead transmission lines. However, it is difficult for mobile robots to carry radar equipment because of weight and power concerns. Therefore, many image-based methods are gradually developed.

Overhead power line segmentation is a key problem of power line detection. The wire usually runs through the entire image, and it is usually assumed to appear in the image as a long line. Thus, many researchers prefer a line-based approach [6]. In [7], adaptive binarization is used to pre-segment the region of wire and the Hough transform is adopted to extract the line features. The ratio detection operator and fast Hough transform help to extract the pixels representing wires and detect straight lines [8]. In [9], a histogram of oriented segments method is leveraged to both establish the main angle orientation of power lines and discard segments that are not power lines in the images. However, it is often difficult for these methods to separate the wire from the image when there are buildings or roads in the background. Because

Manuscript received January 28, 2021; accepted February 24, 2021. Date of publication February 26, 2021; date of current version April 16, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1307400 and in part by the National Natural Science Foundation under Grant U1713224 and Grant 61973300. The associate editor coordinating the review of this article and approving it for publication was Prof. Yudong Zhang. *(Corresponding author: Guodong Yang.)*

Zishu Gao, Guodong Yang, En Li, and Zize Liang are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: gaozishu2016@ia.ac.cn; guodong.yang@ia.ac.cn; en.li@ia.ac.cn; zize.liang@ia.ac.cn).

Rui Guo is with State Grid Shandong Electric Power Company, Jinan 250001, China (e-mail: guoruihit@qq.com).

Digital Object Identifier 10.1109/JSEN.2021.3062660

the difference between the wire and the background in the infrared band is significant, infrared images can be used to help segment the wires [10].

However, as the power line inspection robot attempts to land on the wire, the wire in the image changes to a quadrilateral region with spiral strips. Deep learning-based methods can independently leverage the line and stripe features to segment power lines [11]–[14]. In [15], the multi-scale and structured features are fully exploited to detect the power line. Spatial information is crucial to these segmentation methods [16]. Fully convolutional networks (FCN) [17], which are the foundation of the segmentation framework, utilize VGG as a backbone [18] to extract spatial features. DeepLab v2 [19] and DeepLab v3 [20] employ atrous spatial pyramid pooling to replace the pooling layer and encode the spatial details. A pyramid pooling module is adopted in pyramid scene parsing network (PSPNet) [21], which uses multi-scale average pooling to extract the global context. Multi-Path refinement networks (RefineNet) [22] designs a residual convolution unit with multi-resolution fusion to obtain the spatial features of the images. However, these methods are limited by large computational costs.

In recent studies, lightweight segmentation methods have been gradually developed in an attempt to balance computational cost and segmentation accuracy. ENet [23] is designed based on SegNet [24], and it prunes redundant channels to achieve real-time segmentation; however, it loses spatial information and yields inferior segmentation accuracy in experiments. Efficient residual factorized convNet (ERFNet) [25] uses a non-bottleneck-1D module with a factorized filter, which accelerates its execution time. MobileNet [26] adopts depth-wise convolution and pointwise convolution to reduce the parameters of the model. MobileNet v2 [27] employs an inverted residual structure with linear bottlenecks to reduce computational cost and increase accuracy. MobileNet v3 [28] performs a blockwise search using platform-aware NAS to further improve performance. The ShuffleNet [29] unit is composed of pointwise group convolution and channel shuffle operations, which significantly reduces the computational cost while maintaining accuracy. ShuffleNet v2 [30] employs the channel split operator to improve the performance of ShuffleNet.

The context guided network (CGNet) [31] uses the context guided block to capture information of both local features and surrounding context, and further improves the segmentation ability. ESPNet [32] introduces pointwise convolutions and the spatial pyramid of dilated convolutions, which is based on the reduce-split-transform-merge strategy, to save memory footprint. ESPNet v2 [33] replaces point-wise convolution and dilated convolutions with group point-wise convolution and depth-wise dilated separable convolutions to reduce the parameters of the convolution kernel. The deep feature aggregation network (DFANet) [34] initially uses two versions of Xception and aggregates spatial information through a sub-network and sub-stage cascade, respectively. However, a shallow network weakens the ability to identify features.

In recent studies, multi-branch methods have been gradually developed [35], [36]. The fast semantic segmentation network (Fast-SCNN) [37] proposes the learning to downsample module to extract features for both branches. However, this method has a low capacity for segmentation experiments. In [38], both visible light and infrared images are fed into a modified Fast-SCNN framework to increase the detection performance. The two branches in ContextNet include a deep path at small resolution and a shallow path at full resolution [39]. The bilateral segmentation network (BiseNet) [40] uses the Attention Refinement Module and Feature Fusion Module to decode features. However, the learning capacity of multi-branch networks is reduced owing to the simple connection of features.

To achieve high-quality segmentation with low computational complexity, we propose a novel parallel branch overhead line segmentation network inspired of ContextNet [39], which can also deal with the challenge caused by cluttered background. There are two shortcomings for the ContextNet. The first one is that the method uses two separate branches to extract features separately, which increases the computational cost. The second one is that ContextNet adopts a simple feature fusion module with addition operation and did not make full use of the extracted features. In this paper, the first shortcoming is solved by adopting two standard convolutional layers to extract low-level features for the parallel branch simultaneously. The second one is solved by proposing AFDB module and subnetwork-level skip connections in the classifier, which gather both short-range and long-range information to produce better segmentation results.

There are three main contributions in our paper:

- The asymmetric factorized depth-wise bottleneck (AFDB) module generates short-range context information while keeping small network size.
- The subnetwork-level skip connections in the classifier gather long-range features from intermediate layers and lead to high segmentation accuracy.
- The efficient parallel branch framework achieves more than 90% accuracy and it outperforms other state-of-the-art approaches.

The remainder of this paper is organized as follows: Section II introduces the proposed segmentation network, including context branch, spatial branch, feature fusion model and classifier. Section III provides the experiments and segmentation performance analysis. The conclusion of this paper is drawn in Section IV.

II. Proposed Segmentation Framework

In this section, we expect to make efficient use of short-range and long-range features to achieve comparable performance. Therefore, we implement an efficient parallel branch network for real-time overhead power line segmentation. Fig. 1 shows the overview of the network architectures. It includes the context branch, spatial branch, feature fusion model, and classifier. We call this line segmentation network LSNet for convenience.

Before the context branch and spatial branch, we employ two standard convolutional layers to implement downsampling. One Batch normalization and ReLU follow each convolutional layer.

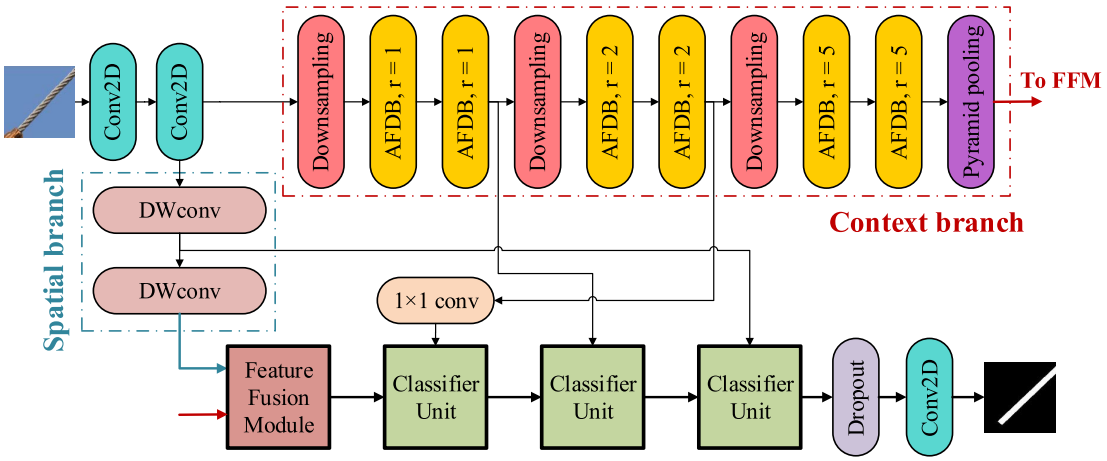


Fig. 1. An overview of the proposed LSNet. The context branch leverages AFDB module to achieve short-range feature extraction. The spatial branch helps to capture abundant spatial information. The feature fusion module solves the information independence between two branches. In addition, a classifier with subnetwork-level skip connections is designed to recover the long-range features and boost the segmentation performance. The r represents dilated rate.

A. Context Branch and Spatial Branch

Fig. 2 shows the different residual blocks. They are adopted to many lightweight detection and segmentation network to solve the efficiency limitation. However, Fig. 2(a) and Fig. 2(b) use only pointwise convolutional layer, which bring a lot of computation burden [29]. To balance the efficiency and segmentation performance, we propose the AFDB module in the context branch, which uses the strength of the dilated depth-wise separable convolution (DDWconv) and factorized convolution. DDWconv enables the framework to obtain features from many originally neighboring pixels and enlarges the receptive field. The factorized convolution leverages $1 \times n$ and $n \times 1$ convolution kernel instead of $n \times n$ convolution kernel, which reduces the parameters of the model while remaining the same receptive kernel.

As shown in Fig. 2(c), the AFDB module is designed in ResNet bottleneck style. First we split the input channel into two half branches and merge the features from two branches, which is a kind of feature reuse. Then we leverage asymmetric factorized depth-wise convolution, one branch is ($n \times 1$ and $1 \times n$) and the other is ($1 \times n$ and $n \times 1$) with dilation. By adopting this modification, we enlarge the width of the module to make efficient use of layers. The kernel size of the AFDB module is 3. Each factorized convolutional layer is followed by ReLU. The two pair of factorized depth-wise convolutional layers leverage affluent features with low computational cost. Next, the outputs of the two paths are concatenated for the same channel with the input. Identify mapping is used to combine the inputs and the outputs. Finally, the channel shuffle operation is performed to gather the spatial information. Note that we omit the nonlinearity between depth-wise and pointwise convolutions.

The whole context branch is composed of asymmetric factorized depth-wise bottleneck (AFDB) module, downsampling, and pyramid pooling module (PPM). Downsampling is single 3×3 convolution with a stride of 2 and helps to reduce computation while obtaining deeper network to gather context information. The pyramid pooling module is inspired by [21] and obtains different region-based features.

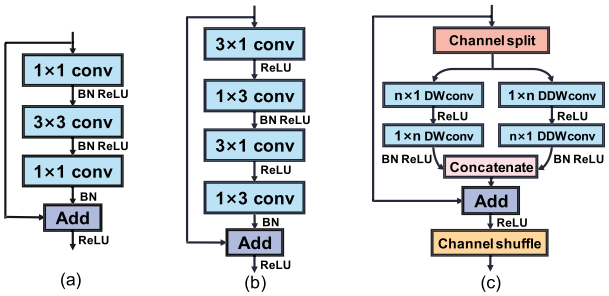


Fig. 2. Comparison of different residual modules. From left to right are (a) ResNet bottleneck, (b) non-bottleneck-1D, (c) our AFDB module, where DW represents depth-wise separable convolution, DDW represents dilated depthwise separable convolution.

The spatial branch aims to capture high-resolution features and refine the information with the context. It consists of two DWconv layers, where the first layer has a stride of 1, and the other has a stride of 2. The output of the channels of each layer is 64 and 128.

B. Feature Fusion Model

Some networks use shallow feature fusion module and classifier to refine the features [37] [39]. Although they ensured efficiency, the accuracy needs to be improved. To attain a balance between segmentation performance and speed, we propose a simpler feature fusion module and a more complicated classifier with subnetwork-level skip connections. The feature fusion module solves the information independence between two branches. As shown in Fig. 3, given the differing resolution of the features, we use the upsampling layer to enlarge the resolution of the context outputs. Then, we concatenate the two outputs of the context branch and the spatial detail branch. The convolutional layer and batch normalization are after the concatenation.

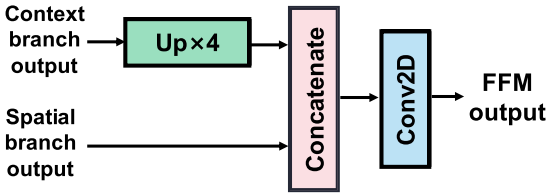


Fig. 3. Feature fusion module. It assists with integration of the multi-scale features from different resolutions.

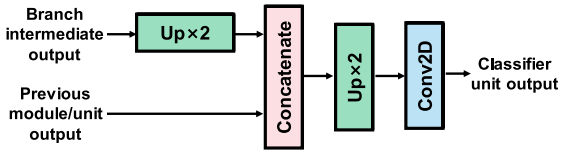


Fig. 4. The classifier unit used in LSNet. The subnetwork-level skip connections improves the ability to recover long-range features.

C. Classifier With Subnetwork-Level Skip Connections

We found that fusing intermediate layers boots the segmentation accuracy. Therefore, we propose a classifier unit with subnetwork-level skip connections. This classifier fuses the long-range context and spatial information, which not only refines image detail but also recovers precise edges in the image. The classifier unit is shown in Fig. 4. It consists of upsampling layers, concatenation and convolutional layer. The inputs of the classifier unit includes not only previous unit output but also intermediate layers in parallel branches. The upsamlping layer makes the features of the two inputs have the same resolution, so that the features can be concatenated together. The convolutional layer is used to bridge the semantic gaps. We attempted to adopt the addition operation in place of concatenation for reduced parameters. However, the addition operation performed worse than concatenation in our experiments.

Three classifier units make up the decoder of the model. The classifiers fuse information of different scales step by step, which can combine contextual information more precisely than simply stacking multi-scale features [19]. In addition, it does not cause too much computational cost since the features have small resolution.

Dropout layer and convolutional layer follow the classifier units and the details of the whole module is shown in Fig. 1. In addition, our network uses cross-entropy as the loss function.

III. Experiments

In this section, we carry out several experiments to evaluate our segmentation model. First, we describe the dataset construction, implementation details and segmentation criteria. Next, we discuss the segmentation criteria. Then we compared the segmentation performance with other state-of-the-art method. In addition, we analyze the training process of our method and demonstrate the segmentation performance of our proposed method. Finally, we provide the ablation evaluation.

A. Experiment Implementation

The overhead power line datasets were acquired by daily inspection including 1000 original images. Each image contains 1 to 3 wires and has been resized to $256 \times 256 \times 3$. We apply a 3-fold cross-validation to avoid the contingency of our conclusion. The 1000 real images are divided into three folds with the respective number (334,333,333). And each fold of the data is used for testing the model in turn and the rest two folds are used for training. In addition, the training set is augmented by random rotation, mirroring, and synthesizing data with complex backgrounds to better simulate various imaging environments. In detail, the first fold is augmented by rotating with the angles of 90° and flipping left and right. The second fold is augmented by rotating with the angles of 180° and flipping up and down. The third fold is augmented by rotating with the angles of 270° and synthesizing data with complex backgrounds. The training set samples are made into VOC format.

We implement our framework in Pytorch and our experiments are performed on an NVIDIA Tesla V100 graphics card. The batch size of 8 was applied and the total training epochs were set to 100 epochs to analyse the loss and pixel accuracy during the training process. We adopted the stochastic gradient descent optimizer (SGD) to update the weight parameters using a learning rate of 0.001 and a learning rate decay of 1×10^{-3}. In addition, Nesterov momentum was adopted and its momentum value was set to be 0.9.

During the test processing, the segmented grayscale probabilistic mask of the object was binarized. The binarization threshold of 0.4 was applied in our experiments.

TABLE I
Confusion Matrix of Segmented Mask Prediction, Where Positive Denotes Target and Negative Denotes Background

Ground truth	Prediction	Category
Positive	Positive	True Positive (TP)
Negative	Positive	False Positive (FP)
Positive	Negative	False Negative (FN)
Negative	Negative	True Negative (TN)

B. Evaluation Indicators

To evaluate the quality and efficiency of the segmentation, several evaluation indicators are adopted. For the binary semantic segmentation, TP, TN, FP, FN can be adopted to evaluate the prediction of every pixel in the segmentation result. Table I explains the define of the four categories.

We apply overall accuracy, precision, recall, F1 and Intersection over Union (IoU) as the similarity measure and frames per second (FPS) as the runtime measure. Through three-fold cross-validation, the metric of the segmented performance of each fold is calculated separately, and then the average value is used as the final result. The metrics are defined as follows:

$$Overall\,Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

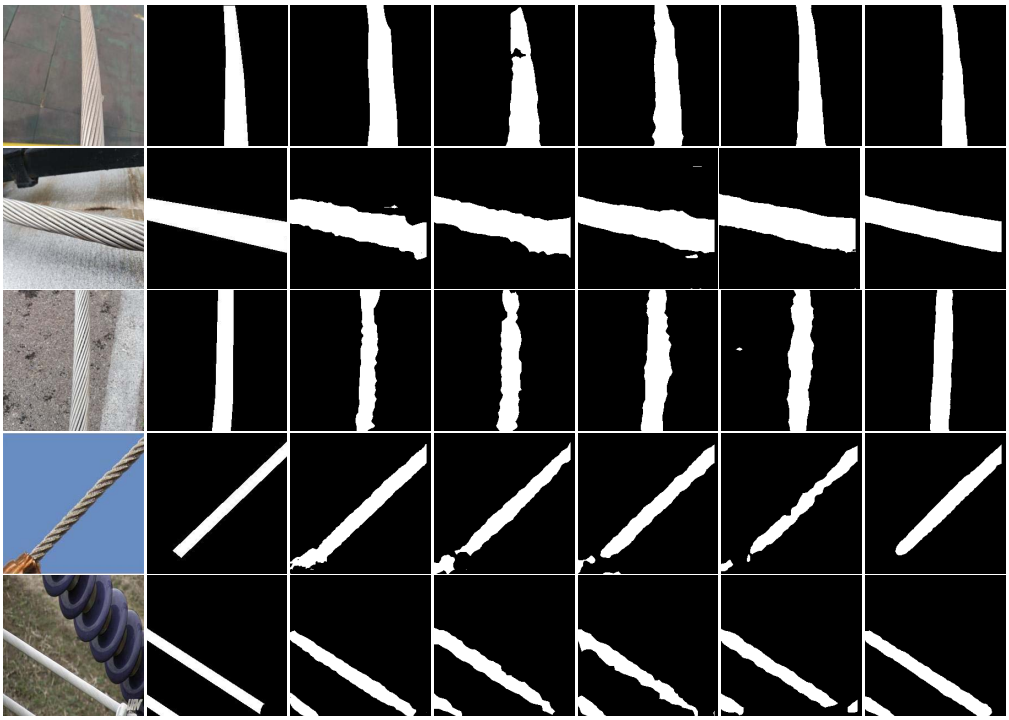


Fig. 5. Qualitative comparison of the different models.

TABLE II
COMPARATIVE PERFORMANCE OF DIFFERENT MODELS

Model	Overall Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	mIoU (%)
LEDNet	94.59	94.91	94.94	94.92	93.95
Fast-SCNN	95.68	95.58	95.54	95.56	95.15
ICNet	96.50	98.60	97.40	97.99	96.08
ContextNet	97.56	98.68	98.51	98.59	97.23
LSNet	97.61	98.62	98.64	98.63	97.31

$$F1 = \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + FP + FN} \tag{4}$$

$$IoU = \frac{TP}{TP + FN + FP} \tag{5}$$

In addition, Receiver operating characteristic (ROC) curve is adopted to evaluate the power line segmentation performance. It calculates a series of sensitivity and specificity at various threshold settings. The larger the area under the curve (AUC), the higher the accuracy of segmentation.

C. Comparison of Different Networks

To compared the proposed network with other state-of-the-art methods, we evaluate performance using five metrics. Table II shows the evaluation metrics in comparison to those of state-of-the-art methods including ICNet [35], Fast-SCNN [37], ContextNet [39], and LEDNet [41]. It can be seen that our proposed method outperforms others in comprehensive metrics overall accuracy, recall, F1 score and IoU. ContextNet has higher precision but lower recall than our proposed method.

Fig. 5 presents the qualitative segmentation results. It can be seen that the proposed LSNet resembles the ground truth closely. These results demonstrate that LSNet makes significant improvements owing to the efficient context branch and classifier. The ContextNet gets applicable segmentation performance for some images. However, some segmented mask from ContextNet contains outliers from the complex background, because it does not make full use of the extracted features with simple feature fusion module. The boundary of the output segmented from ICNet is distorted due to simple convolutional layers for feature extraction. The masks segmented from Fast-SCNN also have blurry boundaries, because they did not have enough layers to extract feature modules in order to reduce network parameters. The outputs from LEDNet miss some details due to the inaccurate segmentation on the edge of the power line.

The parameter and runtime analysis are shown in Table III. It can be seen that all the methods achieve real-time segmentation. In addition, we can see that LEDNet is slower than the proposed LSNet. LSNet is slightly slower compared

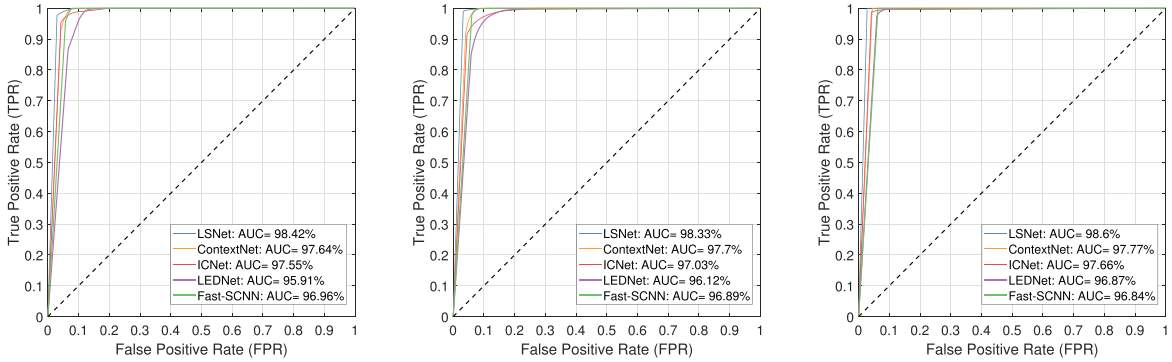


Fig. 6. The performance of line segmentation in terms of ROC curves and AUC values on the three test folds. From left to right are 1st test fold, 2nd test fold, and 3rd test fold.

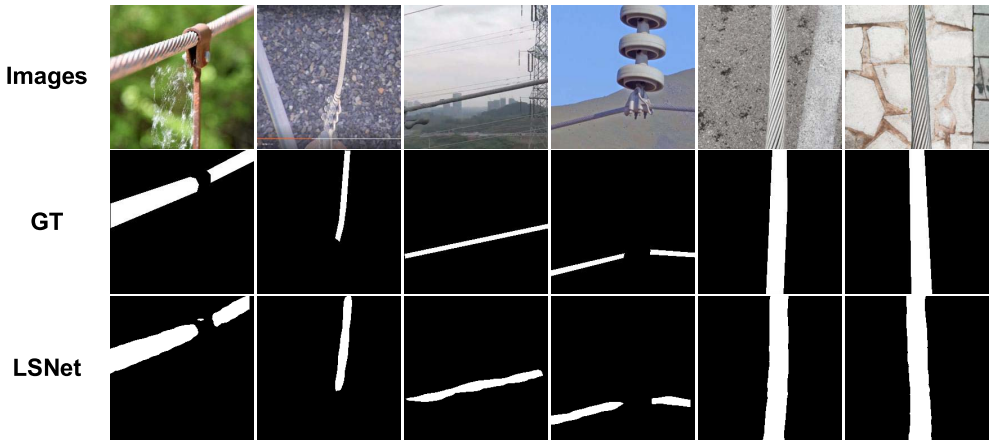


Fig. 7. The segmentation performance of some typical samples.

TABLE III
ACCURACY, PARAMETERS AND RUNTIME COMPARISON OF DIFFERENT METHODS

Model	mIoU(%)	Parameters (M)	Speeding Time(ms)
LEDNet	93.95	0.94	0.935
Fast-SCNN	95.15	1.11	0.964
ContextNet	97.23	0.85	0.843
ICNet	96.08	6.68	1.324
LSNet	97.31	5.36	1.191

TABLE IV
MIOU COMPARISON OF EACH COMPONENT. CB: CONTEXT BRANCH; SB: SPATIAL BRANCH; PPM: PYRAMID POOLING MODULE

Model	mIoU
CB	0.874
CB + PPM	0.932
CB + SB + PPM	0.973

to other methods, but it gets higher segmentation performance. ContextNet, ICNet, and Fast-SCNN are committed to minimizing network parameters even though there is a certain loss of accuracy. Our target is to improve the power line segmentation for power line inspection while satisfying the real-time segmentation. Therefore, our method improves segmentation performance with minor loss in runtime.

D. Performance of Data Augmentation and Proposed Method

To verify the reliability of the datatset and the 3-fold cross-validation, we plot the ROC curves of the segmentation results over all three test folds. As shown in Fig. 6, all the methods show effective segmentation performance in different test folds. And our proposed model presents larger AUC value on all the test folds. It also demonstrates that our dataset has a certain degree of reliability. The three-fold cross-validation can effectively prevent the chance of experimental results.

The segmentation performance of power lines taken at different distances is shown in Fig. 7. The power line in the distance appear as line, and that in the vicinity can be seen as an Quadrangular. We can see that our proposed method can segment these two types of power lines well.

E. Ablation Evaluation

In this subsection, we validate the contribution of the modified component of the proposed framework. Table IV shows that, with the help of the pyramid pooling module (PPM), the mIoU of the model increases by 6.6%. In addition, segmentation performance improves from 93.2% to 97.3% owing to the spatial branch. This demonstrates that the spatial branch obtains abundant spatial information and boosts the segmentation performance.

TABLE V
DETAILED PERFORMANCE COMPARISON OF DIFFERENT FEATURE FUSION MODULES AND CLASSIFIERS

Model	mIoU
Addition	0.919
FFM + Conv2d	0.964
FFM + Classifier	0.973

To fully integrate the features of the two branches, we introduced an efficient feature fusion model and a classifier with subnetwork-level skip connections. Table V shows the different choices of feature fusion modules and classifiers. Addition means that two branches are simply added. The results demonstrate that feature fusion model is capable of integrating more multi-scale information and improving the segmentation performance. Many networks adopt a simple classifier using one convolutional layer (Conv2d) to decode the segmentation map. We compare the classifier using multi-scale skip connections with the simple Conv2d. It can be seen that although the computational complexity increases, the classifier with subnetwork-level skip connection yields more accurate segmentation results.

IV. CONCLUSION

An efficient parallel branch framework for real-time overhead power line segmentation is proposed in this paper. The context branch leverages asymmetric factorized depth-wise bottleneck module to achieve efficient short-range feature extraction with a large receptive field. The spatial branch helps to capture abundant spatial information. In addition, a classifier with subnetwork-level skip connections can be used to effectively recover the long-range features and boost the segmentation performance. The experiment results show that our proposed model performs better than state-of-the-art networks.

In the future, we will further improve the model architecture for high-resolution images. In addition, we will conduct overhead power line defect detection.

REFERENCES

[1] D. B. Rawat and C. Bajracharya, "Detection of false data injection attacks in smart grid communication systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1652–1656, Oct. 2015.

[2] L. Yang, J. Fan, Y. Liu, E. Li, J. Peng, and Z. Liang, "A review on state-of-the-art power line inspection techniques," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9350–9365, Dec. 2020.

[3] T. Mao *et al.*, "Development of power transmission line defects diagnosis system for UAV inspection based on binocular depth imaging technology," in *Proc. 2nd Int. Conf. Electr. Mater. Power Equip. (ICEMPE)*, Apr. 2019, pp. 478–481.

[4] S. Futatsumori, C. Amielh, N. Miyazaki, K. Kobayashi, and N. Katsura, "Helicopter flight evaluations of high-voltage power lines detection based on 76 GHz circular polarized millimeter-wave radar system," in *Proc. 15th Eur. Radar Conf. (EuRAD)*, Sep. 2018, pp. 218–221.

[5] S. Futatsumori, C. Amielh, K. Morioka, A. Kohmura, N. Miyazaki, and N. Yonemoto, "Investigation of circular polarization for 76 GHz helicopter collision avoidance radar to improve detection performance of high-voltage power lines," in *Proc. Eur. Radar Conf. (EURAD)*, Oct. 2017, pp. 295–298.

[6] L. Wang and H. Wang, "A survey on insulator inspection robots for power transmission lines," in *Proc. 4th Int. Conf. Appl. Robot. for Power Ind. (CARPI)*, Oct. 2016, pp. 1–6.

[7] D. Zhao, G. Yang, E. Li, and Z. Liang, "Design and its visual servoing control of an inspection robot for power transmission lines," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2013, pp. 546–551.

[8] C. Shuai, H. Wang, G. Zhang, Z. Kou, and W. Zhang, "Power lines extraction and distance measurement from binocular aerial images for power lines inspection using UAV," in *Proc. 9th Int. Conf. Intell. Hum.-Mach. Syst. Cybern. (IHMSC)*, Aug. 2017, pp. 69–74.

[9] A. Cerón, I. Mondragón, and F. Prieto, "Onboard visual-based navigation system for power line following with UAV," *Int. J. Adv. Robot. Syst.*, vol. 15, no. 2, 2018, Art. no. 1729881418763452.

[10] Y. Zhang, J. Wang, H. Miao, Z. Wang, G. Zhang, and Y. Li, "Identification technology of transmission line defects based on UAV infrared image," in *Proc. 12th Int. Conf. Intell. Comput. Technol. Autom. (ICICTA)*, Oct. 2019, pp. 199–206.

[11] H. Yang, C. Huang, L. Wang, and X. Luo, "An improved encoder-decoder network for ore image segmentation," *IEEE Sensors J.*, early access, Aug. 14, 2020, doi: 0.1109/JSEN.2020.3016458.

[12] Y. Hu, "Image segmentation based on velocity feature vector for moving target extraction," *IEEE Sensors J.*, vol. 20, no. 20, pp. 11983–11991, Oct. 2020.

[13] C.-M. Lin, C.-Y. Tsai, Y.-C. Lai, S.-A. Li, and C.-C. Wong, "Visual object recognition and pose estimation based on a deep semantic segmentation network," *IEEE Sensors J.*, vol. 18, no. 22, pp. 9370–9381, Nov. 2018.

[14] Z. Qiu, F. Yan, Y. Zhuang, and H. Leung, "Outdoor semantic segmentation for UGVs based on CNN and fully connected CRFs," *IEEE Sensors J.*, vol. 19, no. 11, pp. 4290–4298, Jun. 2019.

[15] H. Zhang, W. Yang, H. Yu, F. Xu, and H. Zhang, "Combined convolutional and structured features for power line detection in UAV images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 1306–1309.

[16] Y. Huang, Z. Tang, D. Chen, K. Su, and C. Chen, "Batching soft IoU for training semantic segmentation networks," *IEEE Signal Process. Lett.*, vol. 27, pp. 66–70, 2020.

[17] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Sep. 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[19] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[20] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: https://arxiv.org/abs/1706.05587

[21] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2016, pp. 6230–6239.

[22] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1925–1934.

[23] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A deep neural network architecture for real-time semantic segmentation," 2016, *arXiv:1606.02147*. [Online]. Available: https://arxiv.org/abs/1606.02147

[24] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Jan. 2017.

[25] E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo, "ERFNet: Efficient residual factorized ConvNet for real-time semantic segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 263–272, Jan. 2018.

[26] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," Apr. 2017, *arXiv:1704.04861*. [Online]. Available: https://arxiv.org/abs/1704.04861

[27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[28] A. Howard *et al.*, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Nov. 2019, pp. 1314–1324.

[29] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 6848–6856.

[30] N. Ma, X. Zhang, H. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient cnn architecture design," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 122–138.
[31] T. Wu, S. Tang, R. Zhang, J. Cao, and Y. Zhang, "CGNET: A light-weight context guided network for semantic segmentation," *IEEE Trans. Image Process.*, vol. 30, pp. 1169–1179, Dec. 2020.
[32] S. Mehta, M. Rastegari, A. Caspi, L. G. Shapiro, and H. Hajishirzi, "ESPNet: Efficient spatial pyramid of dilated convolutions for semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 561–580.
[33] S. Mehta, M. Rastegari, L. Shapiro, and H. Hajishirzi, "ESPNetv2: A light-weight, power efficient, and general purpose convolutional neural network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9190–9200.
[34] H. Li, P. Xiong, H. Fan, and J. Sun, "DFANet: Deep feature aggregation for real-time semantic segmentation," *arXiv: Comput. Vis. Pattern Recognit.*, 2019.
[35] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "ICNet for real-time semantic segmentation on high-resolution images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 9514–9523.
[36] D. Mazzini, "Guided upsampling network for real-time semantic segmentation," Jul. 2018, *arXiv:1807.07466*. [Online]. Available: https://arxiv.org/abs/1807.07466
[37] R. P. K. Poudel, S. Liwicki, and R. Cipolla, "Fast-SCNN: Fast semantic segmentation network," Feb. 2019, *arXiv:1902.04502*. [Online]. Available: https://arxiv.org/abs/1902.04502
[38] H. Choi, G. Koo, B. J. Kim, and S. Woo Kim, "Real-time power line detection network using visible light and infrared images," in *Proc. Int. Conf. Image Vis. Comput. New Zealand (IVCNZ)*, Dec. 2019, pp. 1–6.
[39] R. P. K. Poudel, U. Bonde, S. Liwicki, and C. Zach, "ContextNet: Exploring context and detail for semantic segmentation in real-time," May 2018, *arXiv:1805.04554*. [Online]. Available: https://arxiv.org/abs/1805.04554
[40] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "BiSeNet: Bilateral segmentation network for real-time semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 325–341.
[41] Y. Wang *et al.*, "LEDNet: A lightweight encoder-decoder network for real-time semantic segmentation," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2019, pp. 1860–1864.

Zishu Gao received the B.S. degree in automation from Hunan University, Changsha, China, in 2016. She is currently pursuing the joint Ph.D. degree in control theory and control engineering with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, and the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China. Her current research interests include computer vision and robotics.

Guodong Yang (Member, IEEE) received the B.S. degree in mining engineering from the Huainan Mining Institute, Anhui, China, in 1991, the M.S. degree in safety technology and engineering from the Shandong Mining Institute, Shandong, in 1994, and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences (IACAS), Beijing, China, in 2002. From 1994 to 1999, he was an Assistant Professor with the Mechanical and Electrical Engineering Department, Shandong Agricultural University, Shandong. He is currently a Professor with the State Key Laboratory of Management and Control for Complex Systems, IACAS. His research interests include robotics, computer vision, and manufacturing systems.

En Li (Member, IEEE) received the B.E. degree in automation from the Southwest University of Science and Technology, Mianyang, China, in 2002, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2007. His current research interests include optical measurement, automatic control, and optoelectronic systems.

Zize Liang received the B.E. degree in automation and the M.E. degree from Tsinghua University, Beijing, China, in 1986 and 1989, respectively. His current research interests include optical measurement sensors, 3-D reconstruction, and power electronic technology.

Rui Guo received the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 2007. He is currently the Chief Expert of the State Grid Shandong Electric Power Research Institute, Jinan, China. His current research interests include electric robots and power electronic technology.