PAPER • OPEN ACCESS

Transmission Line Scene Classification Based on Light-VGGNet

View the article online for updates and enhancements.

Transmission Line Scene Classification Based on Light-VGGNet

Hongxing Wang[1], Zheng Huang[1*], Bin Liu[1], Xiang Huang[1], Wei Han[2] and Hongchen Li[3]

[1] JiangSu Frontier Electric Technology Co., Ltd, 58 Suyuan Avenue, Jiangning District, Nanjing, Jiangsu Province, China
[2] Shenzhen DY Power AI Co., Ltd, 1st floor, Building 17, Block B, Jiyue City, 133 Tangxing Road, Taoyuan Street, Nanshan District, Shenzhen, China
[3] College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, No.29, Jiangjun Avenue, Jiangning District, Nanjing, Jiangsu Province, China
Email: hz10@vip.qq.com

Abstract. In recent years, power departments have gradually utilized UAV to carry out regular inspections of transmission lines to ensure the safety and stability of power systems. However, a large number of images without useful information are usually captured during the UAV inspection. These images that contain no transmission line information are transmitted to the ground station with informative images, which leads to the surge of workload. To solve this problem, we propose an intelligent image filtering method based on the VGGNet, which is called Light-VGGNet. Firstly, this paper collects the aerial images captured by cameras during the UAV inspection and builds an aerial dataset of transmission line scenes. Then, the Light-VGGNet is proposed, which achieves much lower memory consumption and faster running speed than those of the VGGNet. We use the aerial image dataset to train the proposed network. Finally, the best weight is loaded and utilized to predict the label of the test dataset.

1. Introduction

With the development of human society, electric power has become one of the most critical energy sources. To achieve large-scale power transmission and meet the power demand, power departments build transmission lines in various regions. Therefore, the safe and regular running of transmission lines is essential for the stable and reliable operation of power systems [1]. However, due to the impact of natural disasters, transmission lines always suffer from many kinds of faults, such as insulator burst and tower collapse, which will severely reduce the efficiency of power transmission [2]. To address this challenge, the power departments need to use different methods to evaluate the health condition of transmission lines regularly. Due to the wide distribution of transmission lines, manual inspection suffers from poor safety and low efficiency [3]. Therefore, it has become a trend to use UAV with cameras to carry out the transmission line inspection.

The information extracted from aerial images is used to evaluate the health of transmission line components such as insulators and towers. However, UAV based inspection will collect large numbers of useless data, which does not contain transmission line information. If useless data and informative data are transmitted to the ground station at the same time, it will lead to a surge of work and seriously affect the efficiency of transmission line inspection. To solve this problem, we use an image

classification algorithm to remove useless data and retain informative data. This method can effectively improve the efficiency of transmission line inspection and enhance the reliability of transmission line operation.

With the development of deep learning and GPU, the CNN (Convolutional Neural Network) is widely used in different image processing tasks, such as image classification, object detection, and semantic segmentation. Moreover, CNNs have achieved excellent performance in these tasks above [4]. Traditional machine learning algorithms need to use image feature extraction algorithms such as SIFT (Scale-invariant Feature Transform), LBP (Local Binary Pattern), and HOG (Histogram of Oriented Gradient) to extract image features in advance. However, there is a significant gap between the image features extracted by these algorithms and the high-level image semantics, so the traditional machine learning algorithm is not suitable for processing the images with complex backgrounds [5]. CNNs use convolution layers, pooling layers, and fully connected layers to extract image feature information. As the depth of the network increases, CNN can extract the high-level semantics of images. Hence, CNN can achieve excellent performance in different image processing tasks [6].

In this paper, Light-VGGNets are proposed based on VGGNets, and we train the networks on the transmission line scene dataset. The experimental results show that Lghit-VGGNets achieve much lower memory consumption and faster running speed than those of VGGNets on the transmission line scene dataset. Hence, Light-VGGNets are suitable to be deployed in the UAV system.

2. Related Works

In 1998, LeCun first proposed the LeNet and applied it to the character recognition task [7]. However, due to poor performance, LeNet did not attract much attention. In 2012, Krichevsky et al. proposed AlexNet and won the ILSVRC 2012 competition champion [8]. Subsequently, more and more researchers have proposed CNNs with better performance, such as VGGNet [9] and GooLeNet [10]. The conclusion of VGGNet shows that the depth of CNN is a crucial component to gain better classification performance. GooLeNet was proposed by Christian Szegedy to reduce computation complexity and gain various receptive fields. This proposed network consists of Inception modules that can extract different image features with various kernel sizes. These receptive fields created operations that captured sparse correlation patterns in the new feature map stack [11]. With the increase of layers, the accuracy of the training dataset will decrease, which is caused by the vanishing gradient problem. To solve this, He et al. proposed ResNet in 2015 and won the ILSVRC 2015 competition champion [12].

Previous researchers have made valuable results in scene classification tasks. In Ref. [13], Hua et al. proposed a novel aerial image classification network, which consists of three modules: (2) a label-wise feature parcel learning module, (2) an attentional region extraction module, and (3) a label relational inference module. Then they evaluate the proposed model on two aerial image datasets. In Ref. [14], Zheng et al. proposed a method to solve the aerial scene classification task based on extracting CNN activations from the last convolutional layer of pre-trained CNN. In Ref. [15], the authors proposed a bidirectional LSTM-based sub-network for aerial image classification. Then, they validate the effectiveness of their model on the UCM dataset and DFC15 dataset.

3. Methods

VGGNet is regarded as a famous landmark during the development process of CNN. The architecture of VGGNet consists of 3×3 convolutional layers with the ReLU activation function. Following the activation function is a max-pooling layer and fully connected layers which using a softmax activation function. However, due to its structure, VGGNet suffers from large memory consumption. Hence, VGGNet is not suitable to be adapted in UAV systems. To reduce the number of network parameters, we propose Light-VGGNets and utilize them to solve the scene classification task.

3.1. OCM

A proposed structure is proposed to replace the 3×3 convolution layers in VGGNets, which is called OCM (optimized convolution module). The architecture of OCM is as shown in figure 1. The group convolution is adopted into OCM to reduce the number of network parameters. Consider a feature map x that is passed through the OCM. The input size and the output size of the convolution layer are both $n \times n \times d$. The input x is split into two parts, which are called x_1 and x_2. A 3×3 convolution layer is operated on x_1 and x_2 to gain larger receptive fields. However, the lack of information between x_1 and x_2 will lead to poor classification performance. To facilitate the information flow in the network, the connection between x_1 and x_2 is added before the convolution layer. It could be found that the output y_1 contains the information of x_1 and x_2. Hence, the OCM can extract more information than a 3×3 convolution layer. To alleviate the vanishing gradient problem, we introduce identity mapping into VGGNets. The output of the OCM y is given as formula (1).

$$\begin{cases} y_1 = F\left(x_1 + x_2\right) \\ y_2 = F\left(x_2\right) \\ y = x + C\left(y_1, y_2\right) \end{cases} \tag{1}$$

where $F(\cdot)$ denotes the convolution operation, and $C(y_1, y_2)$ refers to the concatenation of y_1 and y_2.
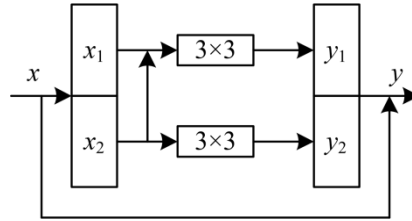
Figure 1. The architecture of the OCM.

3.2. OPM

By analyzing the network structure of VGGNets, it could be found that the three fully connected layers in VGGNets contain most of the network parameters. However, there is one fully connected layer in ResNets. Inspired by this, the fully connected layers in VGGNets are replaced by one fully connected layer, and the last maxpool layer is replaced by an average pool layer, as suggested in Ref. [12]. This new module is called OPM (optimized predict module), as shown in figure 2. Compared with the traditional predict module in VGGNets, the OPM can reduce the number of network parameters significantly.
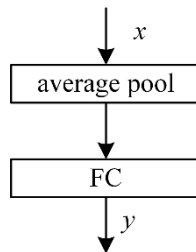
Figure 2. The architecture of the OPM.

3.3. Implementation Details

The layouts of VGGNets and Light-VGGNets are shown in table 1. In this paper, all the networks are trained using SGD (Stochastic Gradient Descent) with a mini-batch size of 10. We set the momentum,

the weight decay, and the initial learning rate to 0.9, 10-4, and 0.001. The loss function and activation function are set to binary cross-entropy and sigmoid.

Table 1. The layouts of VGGNets and Light-VGGNets.

VGG-16	Light-VGG-16	VGG-19	Light-VGG-19
input (224×224 RGB image)			
conv3-64×2	OCM-64×2	conv3-64×2	OCM-64×2
maxpool			
conv3-128×2	OCM-128×2	conv3-128×2	OCM-128×2
maxpool			
conv3-256×3	OCM-256×3	conv3-256×4	OCM-256×4
maxpool			
conv3-512×3	OCM-512×3	conv3-512×4	OCM-512×4
maxpool			
conv3-512×3	OCM-512×3	conv3-512×4	OCM-512×4
maxpool	average pool	maxpool	average pool
FC×3	FC	FC×3	FC

4. Experiments

The dataset used in this paper is composed of aerial images captured by cameras during UAV inspection. The size of each image is 224 × 224. The aerial images are divided into positive and negative samples. Positive samples contain the information of critical components of transmission lines, while negative samples not. The schematic diagram of positive and negative samples is shown in figure 3.

(a)   (b)
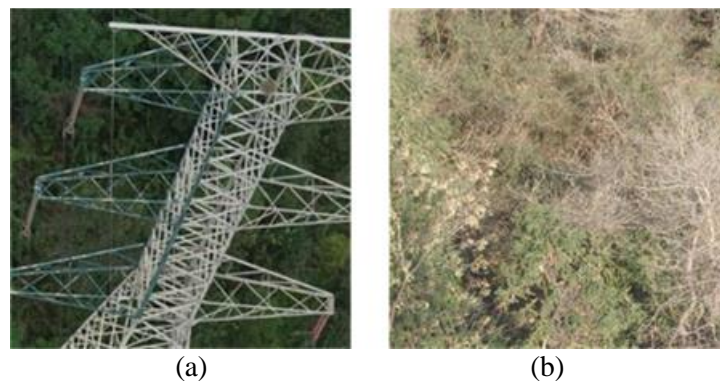
Figure 3. (a) The positive sample; (b) The negative sample.

As suggested in Ref. [13], the classification performance of CNN is evaluated by test accuracy and F1-measure. TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) are essential indexes of classification tasks. The precision rate and recall rate are given as equation (2).

$$\begin{cases} \text{precision} = \dfrac{TP}{TP+FP} \\ \text{recall} = \dfrac{FP}{FP+TN} \end{cases} \tag{2}$$

According to the above analysis, accuracy and F1-measure can be calculated by equation (3).

$$\begin{cases} \text{accuracy} = \dfrac{TP+TN}{TP+FN+FP+TN} \\[2mm] \text{F1-measure} = \dfrac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \end{cases} \qquad (3)$$

The experimental results are listed in table 2. It can be found that Light-VGGNets gain slightly lower accuracy and F1-measure than those of VGGNets. The networks will run on-line. Hence, the real-time performance of networks should be evaluated. We measure the memory consumption and average running time of each network. It could be found that Light-VGGNets achieve much lower memory consumption and faster running speed than those of VGGNets. Hence, Light-VGGNets proposed in this paper can be regarded as the compromise between classification performance and real-time performance. Figures 4 and 5 show the confusion matrixes of VGGNets and Light-VGGNets.

Figure 4. (a) The confusion matrix of VGG-16; (b) The confusion matrix of the Light-VGG-16.

Figure 5. (a) The confusion matrix of VGG-19; (b) The confusion matrix of the Light-VGG-19.

Table 2. The accuracy, F1-measure, memory consumption, and average running time of VGGNets and Light-VGGNets.

	Accuracy	F1-measure	Memory consumption	Average running time
VGG-16	98.80%	98.80%	512M	0.007s
Light-VGG-16	98.73%	98.72%	28.2M	0.005s
VGG-19	98.68%	98.67%	532M	0.008s
Light-VGG-19	98.15%	98.14%	38.3M	0.006s

5. Conclusion

In this paper, Light-VGGNets are proposed and applied to transmission line scene classification. Experimental results show that Light-VGGNets gain slightly lower classification performance than VGGNets. However, Light-VGGNets achieve much much lower memory consumption and faster

running speed than those of VGGNets. Hence, Light-VGGNets can be regarded as a compromise between classification performance and real-time performance. Light-VGGNets are more suitable to be deployed in UAV systems.

References
[1]   Jenssen R and Roverso D 2018 Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning *International Journal of Electrical Power & Energy Systems* **99** 107-120.
[2]   Deng C, Wang S, Huang Z, et al. 2014 Unmanned aerial vehicles for power line inspection: A cooperative way in platforms and communications *Journal of Communications* **9** (9) 687-692.
[3]   Zhou Z, Zhang C, Xu C, et al. 2018 Energy-efficient industrial Internet of UAVs for power line inspection in smart grid *IEEE Transactions on Industrial Informatics* **14** (6) 2705-2714.
[4]   Chan T H, Jia K, Gao S, et al. 2015 PCANet: A simple deep learning baseline for image classification? *IEEE transactions on image processing* **24** (12) 5017-5032.
[5]   Bengio Y, Courville A and Vincent P 2013 Representation learning: A review and new perspectives *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35** (8) 1798-1828.
[6]   Hinton G E, Osindero S and Teh Y W 2006 A fast learning algorithm for deep belief nets *Neural Computation* **18** (7) 1527-1554.
[7]   Lecun Y and Bottou L 1998 Gradient-based learning applied to document recognition *Proceedings of the IEEE* **86** (11) 2278-2324.
[8]   Krizhevsky A, Sutskever I and Hinton G E 2012 Imagenet classification with deep convolutional neural networks *Advances in Neural Information Processing Systems* 1097-1105.
[9]   Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale image recognition *arXiv preprint arXiv:1409.1556*.
[10]   Szegedy C, Liu W, Jia Y, et al. 2015 Going deeper with convolutions *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pp 1-9.
[11]   Alom M Z, Taha T M, Yakopcic C, et al. 2018 The history began from alexnet: A comprehensive survey on deep learning approaches *arXiv preprint arXiv:1803.01164*.
[12]   He K, Zhang X, Ren S, et al. 2016 Deep residual learning for image recognition *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* pp 770-778.
[13]   Hua Y, Mou L and Zhu X X 2020 Relation network for multilabel aerial image classification *IEEE Transactions on Geoscience and Remote Sensing*.
[14]   Zheng X, Yuan Y and Lu X 2019 A deep scene representation for aerial scene classification *IEEE Transactions on Geoscience and Remote Sensing* **57** (7) 4799-4809.
[15]   Hua Y, Mou L and Zhu X X 2019 Recurrently exploring class-wise attention in a hybrid convolutional and bidirectional LSTM network for multi-label aerial image classification *ISPRS Journal of Photogrammetry and Remote Sensing* **149** 188-199.