Surface Coverage Classification of UAV High Resolution Image Transmission Lines Corridor based on Total Convolution Neural Network

T Guo¹, L Xu¹, W Wang¹, Y Wei¹, Y Yang¹, J Gao² and S Jian²

¹China Southern Power Grid Guizhou Electric power CO.LTD., Guiyang, China
²School of Electrical Engineering, Wuhan University, Wuhan, China

E-mail address : whugjc@foxmail.com

Abstract. In recent years, UAV imagery has been applied more and more to the work of power line inspection. Surface coverage classification is a key step. Because of the size and amount of data of UAV high-resolution images, the accuracy requirements of traditional classification algorithms can no longer meet the practical needs. In this paper, a pixel level classification algorithm based on full convolution neural network is proposed. The total convolution neural network model reduces the error caused by other forms of deformation such as image translation, scaling, tilting and so on. In the application of UAV image, the total convolution neural network is used to classify the image, and the high dimensional information extracted from the coiling layer is used to learn the feature of the image fully, and the accuracy of the classification is improved. By comparing with the random forest algorithm, the advantages of the full convolution neural network in the classification of the surface coverage of the UAV transmission line corridor are verified, and the reference value is provided for the electric patrol line and the image classification of unmanned aerial vehicles.

1. Introduction

As an important pillar industry of the country, the electric power industry is one of the basic industries of the national economy. As an important infrastructure for wide area, long distance and large capacity transmission lines, the safety of transmission lines has a significant impact on people's daily life and national economy. Because of the long exposure to the natural environment, it is often infringed by various external factors, such as lightning strike, Storm Snow, bird damage, tree and so on, which cause the transmission lines to be easily broken down, causing a large area blackout, which seriously affects people's daily production and life and the national economy. Therefore, the power grid operation and maintenance management departments need periodic power line inspection for transmission lines to prevent and eliminate power grid safety accidents. In order to meet the needs of rapid economic development, the more and more high voltage and ultrahigh voltage power transmission lines are built, the geographical environment of the transmission corridor is becoming more and more complex. For example, through large area of water, lakes and mountains, many difficulties have been brought to the management and monitoring of the lines. The traditional aerial survey technology uses the data collected on the aircraft platform to inspect the transmission line. Because of the tedious handling procedures and poor maneuverability of the airspace control procedures, it leads to a long period and high cost. It is difficult to meet the actual engineering requirements. In recent years, with the rapid development of UAV, digital camera technology and
UAV aerial data processing software, UAV aerial survey technology has basically matured. The first step in using UAVs for power line inspection is to determine the type of surface coverage in the area where the image is located [1].

In the field of image classification, the method of supervised classification is generally used to learn the relationship between the original image and the corresponding artificial markers through the algorithm, and the relationship is used as the classification basis, and it is applied to the unmarked image so as to achieve the classification effect. The common algorithms, such as Support Vector Machine (SVM), Adaboost, random forest (Random Forest, RF) and so on, are widely used in the field of image classification. Among them, the random forest algorithm is very reliable in dealing with high-dimensional data and highly generalized data. Meanwhile, random forests are not sensitive to noise and outliers [2]. However, there are difficulties in the classification of UAV images in a relatively large scale. Because it belongs to the shallow structure model [3], for the complex original input image, its expression ability has some limitations, and can not fully learn the complex structure information. Deep Convolutional Neural Network (deep convolution neural network) is a widely used supervised learning method. It can automatically learn different levels of abstract features from multiple original images through multi-layered stacked convolution cores and pool layers. Fully Convolutional Network (FCN), as one of the deep convolution neural networks, plays a huge role in the field of image classification. It can learn high dimensional features of images, and it has the characteristics of high non deformation for other forms such as image translation, scaling, tilt and other forms. It is very suitable for the classification of large scale unmanned aerial vehicles. In this paper, the application of FCN model to the surface classification of UAV transmission line corridors is studied and verified by experiments.

2. Full Convolution Neural Network based on High Resolution UAV Images

2.1. Fully Convolution Network

In recent years, deep learning has attracted wide attention and applications in all walks of life, such as in video related fields [4], in audio related fields [5], and in language related fields [6]. In the field of image processing, convolution neural network has achieved the best results in the ImageNet ILSVRC image recognition and image classification competition [7-9]. In the field of remote sensing image classification, convolution neural network has also achieved good results [10, 11]. In particular, Long applies a deep convolutional network to the whole convolutional network to achieve the result of image classification [12]. Long instead of the full link layer in the common convolution network, the transformation from image to tag to image to image is realized. All convolution networks can achieve pixel level classification, so it has become the basic framework of many cutting-edge algorithms. Multi group convolution kernel, activation function and pooling layer are used to encode the original image and extract high-dimensional features. These high-dimensional features are decoded as input to the ascending sampling layer to achieve a pixel level classification process.

Usually, the whole convolution neural network consists of several parts:

(1) Batch Normalization (BN)

In all convolution networks, there are usually a large number of parameters that need to be calculated. When training samples are few, it is very important to prevent overfitting. Batch Normalization is the key step to prevent overfitting [13]. As shown in formula (1), each piece of data in the Batch Normalization calculation step is normalized.

\[
y = \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}} \gamma + \beta
\]  

(1)
In the form, x and y are marking input and output respectively. $\mu$ and $\sigma^2$ are the mean and variance of the input data. $\gamma$ and $\beta$ are the parameters to learn, and $\varepsilon$ is a constant for the numerical stability.

(2) Rectified Linear Units (ReLU).

The activation function used by this algorithm is the ReLU function proposed by Hinton [14]. ReLU function is the most commonly used activation function in depth learning at present. Compared with the general smooth nonlinear activation function, ReLU has the advantage of fast speed. It can save and train steps for deep networks. The specific calculation is shown in formula (2).

$$g(x) = \max(0, x)$$

(3) Convolution Layer (Conv)

Convolution layer is the core construction of the whole convolution neural network. It performs heavy computation work. The function of convolution layer is to convolution operation of input image, and the output characteristic image represents the response of each convolution kernel to the original input image [15]. Through the mobile convolution kernel, the whole image can be operated. Because each convolution kernel has the same weight, the feature can be detected no matter where the image is. The formula of the convolution layer is shown in formula (3).

$$x'_j = g(\sum_{i=0}^{M_j} x_{i-1} * k_{ij} + b_j)$$

(4) Pooling Layer

The Pooling Layer can also be called a nonlinear descending sampling layer. When a feature is found, its exact location is not as important as its relative position to other features. The emergence of pooling layer is just in line with this idea. It improves the robustness of the network and reduces the number of network parameters. There are two typical pools: mean pooling and maximum pooling. The function of mean pooling is to select the mean in the region instead of the feature. The maximum pooling is to select the maximum value in the region instead of the feature. Each algorithm has its own advantages and disadvantages. The algorithm adopts the maximum pooling.

(5) Upsampling

The Upsampling is the inverse process of the pool [16]. When we know the pool process clearly, the location of the stored data is recorded, and the corresponding data is filled to the original location in the process of up sampling.

(6) Logical regression layer

The whole convolution network is end-to-end correspondence, and input image output is also image. Output all categories through the last logical regression layer. This algorithm uses Softmax function to carry out logistic regression [17]. The maximum probability of each pixel in the output image is obtained through the Softmax layer, making it the label of the pixel, thus achieving pixel level classification.

Through the combination of the above six structures, the full convolution network used in this paper is shown in Figure 1.
2.2. UAV Data Processing

As the size of the UAV image is larger and the data information is rich, we need to preprocess the unmanned aerial images first, and we have solved the problem of the deficiency of the video card.

For larger data, we use clipping with sliding windows. The size of the sliding window used in this experiment is 128*128. In order to ensure the adequacy of the training data and prevent over fitting, we use a sliding step 32 to cut the image. In training process, training data and its corresponding label data are tailored simultaneously to generate training data sets. The FCN model is trained by training data sets.

For the data to be classified, the choice of sliding window is also 128*128. In order to ensure the efficiency of classification, we use the sliding step length of 128 to cut the image and generate the data set to be classified. Input the categorized data sets into the FCN network model, and get the prediction results of each small cutting image through the logical regression layer. The result of pixel classification can be obtained by splicing the cut data.

3. Full convolution neural network based on high resolution UAV image

3.1. High Resolution Unmanned Aerial Vehicle Sample Data

In order to verify the efficiency and accuracy of the whole convolution neural network for UAV image classification, the experimental data of Guangdong unmanned aerial vehicle (UAV) data were processed in this paper. The UAV model is KC2000 UAV, the camera is SONY A-7R2 type single counter, the UAV absolute navigation height 200-250m, the image pixel 7360*4912, and the average ground resolution of the generated image is 5cm. In the survey area, 36 images were selected for manual marking, of which 24 were used as training data and 12 as accuracy evaluation data. The ground objects are divided into six categories: towers, woodlands, grasslands, water bodies, houses and bare lands, as shown in Figure 2.
3.2. System environment

After the data set is ready, the experiment can be started. The training part of the algorithm and the data processing part of the UAV are implemented in the Ubuntu system, and the specific parameters are shown in Table 1 as follows.

| Environment parameter | values                      |
|-----------------------|-----------------------------|
| Operating system      | Ubuntu 14.04.05             |
| Computer processor    | Intel ( R ) Core(TM)i7-4770 |
| Video card            | NVIDIA Tesla K20c           |
| Computer memory       | 12.0GB                      |
| Video card memory     | 4.0GB                       |
| Integrated development environment | Python+ caffe               |

3.3. Network learning parameters

During the training network, the training data was cut into 249762 small images of 128*128. The Batch Size was set to 10 in the limited training process and a Epoch with 25000 iterations, and the training was carried out in a total of 12 Epoch. In the form of step attenuation, the learning rate is reduced, and every 4 epoch is reduced one time. The specific choice of learning parameters is shown in Table 2.

| parameter values | Parameter |
|------------------|-----------|
| Momentum         | 0.9       |
| Base learning rate | 0.001   |
| Gamma            | 0.1       |
| Weight decay     | 0.005     |
3.4. Unmanned aerial vehicle data processing

The experimental data are classified by pixel level classification based on full convolution neural network, and the comparison diagram of classification results is shown in Figure 3.

![Figure 3](image)

**Figure 3.** Comparison of Classification Results

Figure 3 (a) is the result of total convolutional network classification, and Figure 3 (b) is based on random forest classification results.

3.5. Accuracy evaluation

Data classification based on random forest and total convolution neural network is carried out on experimental data. The accuracy of the total convolution neural network for the classification of the surface coverage of the UAV transmission line corridor is verified by selecting six kinds of objects in the test data, and the precision evaluation of the electric tower, the water body, the woodland, the grassland, the bare land and the house. The accuracy, overall accuracy and Kappa coefficient of different objects are evaluated. The results are shown in Table 3 and 4. The classification results based on the full convolution neural network are more smooth than the classification results based on the random forest, and the convolution processing has played a role in reducing the noise to some extent. The accuracy of the overall precision is 5.72%.

4. Conclusion

This paper discusses the principle of the full convolution neural network and its application in the surface coverage of the low altitude unmanned aerial vehicle image transmission line corridor, and uses the unmanned aerial vehicleimage in Guangdong to carry out the classification experiment and the accuracy evaluation. It is verified that the classification results based on the full convolution neural network are due to the traditional classification method based on the random forest whether in the overall accuracy or on the Kappa coefficient. The feasibility of the algorithm is verified, and it provides reference value for the field of UAV surface coverage and other related work.

| Table 3. Result of the Random Forest |
|-------------------------------------|
| electric tower | water | woodland | grassland | bare land | buildings |
|----------------|-------|----------|-----------|-----------|-----------|
| electric tower | 90.16%| 0.44%    | 0.57%     | 0.63%     | 4.33%     | 3.87%     |
| water          | 0.33% | 87.60%   | 0.73%     | 5.63%     | 4.96%     | 0.75%     |
| woodland       | 1.18% | 0.99%    | 80.36%    | 7.29%     | 2.31%     | 1.87%     |
| grassland      | 0.55% | 2.81%    | 8.69%     | 84.23%    | 3.21%     | 0.51%     |
| bare land      | 3.57% | 5.94%    | 0.56%     | 6.57%     | 82.39%    | 0.97%     |
| buildings      | 3.21% | 0.58%    | 0.39%     | 1.80%     | 2.56%     | 91.46%    |
| Overall accuracy |       |          |           |           |           | 89.61%    |
| Kappa          |       |          |           |           |           | 86.53%    |
Table 4. Result of the Fully convolutional Neural Network

|                | electric tower | water | woodland | grassland | bare land | buildings |
|----------------|----------------|-------|----------|-----------|-----------|-----------|
| Electric tower | 94.13%         | 0.03% | 0.32%    | 0.21%     | 2.11%     | 3.20%     |
| Water          | 0.07%          | 93.32%| 0.58%    | 2.14%     | 3.87%     | 0.02%     |
| Woodland       | 0.03%          | 0.32% | 96.10%   | 2.62%     | 0.89%     | 0.04%     |
| Grassland      | 0.32%          | 1.33% | 2.26%    | 94.17%    | 1.69%     | 0.23%     |
| Bare land      | 2.32%          | 1.63% | 0.89%    | 2.55%     | 91.96%    | 0.65%     |
| Buildings      | 1.24%          | 0.32% | 0.27%    | 0.46%     | 1.37%     | 96.34%    |
| Overall accuracy|               |       |          |           |           |           |

95.33%

References

[1] Wang Y and Li J 2004 China. J. Classification Methods of Land Use/Cover Based on Remote Sensing Technology. Rsi. 1 53-9
[2] Criminisi A and Shotton J 2003. J. Decision Forests for Computer Vision and Medical Image Analysis. 273-93
[3] Joachims T 1998. R. Komplexitätsreduktion in Multivariaten Datenstrukturen. SFB475
[4] Sermonet P and Eigen D 2013. J. Integrated recognition, localization and detection using convolutional networks. Arxiv. 13126229
[5] Waibel A Hanazawa T and Hinton G 1989. J. Phoneme recognition using time-delay neural networks. IEEE Tass. 37 328-39
[6] Collobert R Weston J and Bottou L 2011. J. Natural language processing (almost) from scratch Mlr. 12 2493-537
[7] Krizhevsky A Sutskever I and Hinton G E 2012. C. Imagenet classification with deep convolutional neural networks; proceedings of the Advances in neural information processing systems.
[8] Simon K and Zisserman A 2014. J. Very deep convolutional networks for large-scale image recognition. Arxiv. 14091556
[9] Szegedy C Liu W and Jia Y 2015. C. Going deeper with convolutions; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition.
[10] Zdunek R 2015. J. Geometrical interpretation and application to spectral unmixing. Amcs. 24 233-47
[11] Feng L 2017 China. D. Extraction of construction land information from high resolution remote sensing images based on deep learning technology.
[12] Long J Shelhamer E and Darrell T 2015. C. Fully convolutional networks for semantic segmentation; proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
[13] Ioffe S and Szegedy C 2015 Japan. J. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Cs.
[14] Nair V and Ninton G E 2010. C. Rectified Linear Units Improve Restricted Boltzmann Machines; proceedings of the International Conference on Machine Learning.
[15] Hubel D H and Wiesel T N 1962. J. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Phys. 160 106
[16] Badrinarayanan V Kendall A and Cipolla R 2015. J. A deep convolutional encoder-decoder architecture for image segmentation. Arxiv. 151100561
[17] Bishop C M 1995. M. Neural networks for pattern recognition.