Remote sensing image-based wildfire recognition using capsnets for transmission lines

Xianping Zhao¹, Zhenyu Gao², and Gang Wen¹
¹Electric Power Research Institute, Yunnan Power Grid Company Ltd., 650217, Kunming, China
²Yunnan Power Grid Company Ltd., 650000, Kunming, China

Abstract. Transmission lines are the lifeblood of power grid operation and the lifeline of people's livelihood. Recently, large-scale overhead transmission lines frequently trip due to wildfire, leading to serious threat to security and stability of the power grid and the public’s normal use of electricity. With the application of satellite remote sensing image in the electric power sector, it becomes a reality that using remote sensing image to detect wildfire. Aiming at the characteristics of wildfires that are easily misdetected and missed, this paper proposes a remote sensing image-based wildfire recognition model based on capsule network (capsnet). The proposed architecture is evaluated on remote sensing image datasets from the Himawari-8 meteorological satellite. The proposed method has been shown to be effective in detecting wildfire in experiments, which is useful for the improvement of resist ability of wildfire for transmission lines.

1. Introduction
With the rapid increase of people's demand for electricity, electric power construction as the foundation of national development has also developed rapidly. Due to the gradual expansion of the scale of the power grid, numerous overhead transmission lines will pass through mountainous areas and forest areas in a large range. In these areas, tripping accidents of transmission lines caused by wildfires often occur, which brings serious threat to stable operation of power grid[1-3]. Remote sensing image can be used to monitor wildfire in real time, which is helpful to carry out the maintenance of transmission lines in time and is of great significance to improve the reliability of power grid.

It is a challenging problem to analyze the correct wildfire information from remote sensing image. There are many methods of remote sensing image recognition. Convolutional neural network(CNN) has been applied in remote sensing image classification due to its powerful feature learning and reasoning ability, and achieved excellent results[4-6]. Although the CNN-based method greatly improves the classification accuracy, CNN's pooling operation discards information such as object locations, which cannot fully consider the spatial relationship, so that the CNN cannot have a better precision in the recognition of wildfire.

Recently, Sara et al.[7] introduced a new network architecture called CapsNet, which is a more effective image recognition algorithm that the test accuracy for MNIST is 99.75%. CapsNet has been successfully applied in finger vein recognition, cancer screening, hyperspectral image classification and other fields in recent years[8-10]. The idea of CapsNet is to encode the relative relationships between local and entire objects. Using a group of neurons as a capsule instead of a single neuron in a traditional neural network solves the problem of not capturing the hierarchy of the image effectively.

Therefore, we believe
that this network structure can detect accurate wildfire information from remote sensing images, and propose a novel method of wildfire monitoring based on CapsNet.

However, we found that the original CapsNet could not achieve better accuracy when applied to wildfire recognition and training from scratch. Therefore, we propose an modified CapsNet model and an effective initialization mechanism to improve network performance. First, the proposed network structure named ex-CapsNet is two more convolutional layers than the original CapsNet. In addition, we extract the convolutional parameters of the pretrained CNN model and use these parameters to initialize convolutional layers and PrimaryCaps layer of the ex-CapsNet. We prove that the proposed method is more suitable for wildfire detection than the original CapsNet in the experimental section.

2. Method
This section starts with an introduction to CapsNet, followed by a detailed description to the proposed method.

2.1. CapsNet
As illustrated in Figure 1, the original CapsNet is a network architecture with three layers, convolutional layer, PrimaryCaps layer and DigitCaps layer. Where the first layer is a conventional conv_layer, whose function is to perform a local feature detection on the image and convert the image pixel into feature map as the input of PrimaryCaps layer. PrimaryCaps layer is a stack of ordinary convolutional layers that output lower-level capsules. lower-level capsules are converted into Higher-level capsules through the routing process at the DigitCaps layer.

CapsNet uses capsule instead of CNN's neuron, where neuron outputs a scalar and capsule outputs a vector. In CapsNet, a capsule is defined as a vector consisting of a group of neurons which can represent the characteristics of pattern, such as position, pose and size. The norm of vector represents the probability of the existence of a class. Higher-level capsule receive predictions from all lower-level capsules and Figure 2 shows the process by which lower-level capsules are transformed into a higher-level capsule by dynamic routing.
Figure 2. Routing Process between lower-level capsules and higher-level capsule

Since the norm of each capsule represents the existence, the Squashing function in Figure 2 is used to make the capsule compress the norm of the vector while maintaining the vector direction, ensuring that the norm value is between 0 and 1. The Squashing function expression as

\[ \nu_j = \frac{\|s_j\|^2 \cdot s_j}{1 + \|s_j\|^2 \|s_j\|} \]  

Here, calculation process of \( s_j \) as follows

\[ u_{j|i} = W_{ij} \cdot v_i, \quad s_j = \sum_i C_{ij} \cdot u_{j|i} \]  

Where \( v_i \) as the output of lower-level capsule \( i \) which detect the pattern and \( W_{ij} \) is the affine transformation matrixes that represent the spatial relationship between the capsule \( i \) and capsule \( j \). Routing coefficient \( C_{ij} \) is decided by routing process that reflects the coupling between these two capsules, which takes the following form

\[ C_{ij} = \exp\left(\frac{b_{ij}}{\sum_{i} \exp(b_{iN})}\right) \]  

At the beginning of routing process, \( b_{ij} \) is set to zero in order to ensure that each capsule \( i \) and capsule \( j \) are coupled with equal probability. Formula (3),(2) and (1) are then executed coherently and \( b_{ij} \) is updated through iteration during routing process.

\[ b_{ij} = b_{ij} + u_{j|i} \cdot v_j \]  

Update \( b_{ij} \) with the inner product of \( u_{j|i} \) and \( v_j \). When they are similar, the inner product and \( b_{ij} \) increase, and the coupling between the capsule \( i \) and the capsule \( j \) increases. Where iteration number of routing process bigger is not better. In the experiment of section 3.2, we found that the highest accuracy is achieved when the routing number is set to 2. The loss function of CapsNet is similar to SVM namely marginal loss. Each heigher-level capsule \( N \) of the DigitCaps layer is related to the loss function \( L_N \), which can be expressed as

\[ L_N = T_N \max(0, m^+ - \|v_N\|^2) + \lambda (1 - T_N) \max(0, \|v_N\| - m^-)^2 \]  

Where \( L_N \) is the margin loss of the \( N \)-th class and \( m^+ \), \( m^- \) and the proportionality coefficient \( \lambda \) are hyper-parameters. In this paper, both \( m^+ \) and \( m^- \) are set to 0.9, and \( \lambda = 0.5 \). In addition, if the \( N \)-th class exists, then \( T_N = 1 \), otherwise \( T_N = 0 \).

2.2. ex-CapsNet

We train the original CapsNet on remote sensing image datasets. But it’s not doing well in three areas: 1) the loss falls into the local minimum easily; 2) the recognition precision is lower than some classic CNN models, such as VGG-16 and Inception-v3; 3) there are too many capsules in the PrimaryCaps layer, which makes the iterative process of dynamic routing very time-consuming. Based on these three problems, we proposed the ex-CapsNet network architecture. Firstly, the remote sensing image was
input into the CNN model, and the initial feature map was extracted from the convolutional layer. Then input the initial feature map into CapsNet to get the final wildfire recognition results. The ex-CapsNet network architecture is shown in Table 1.

| name          | parameter size/stride | output shape |
|---------------|-----------------------|--------------|
| input         | -                     | (35,35,2)    |
| Conv1         | 9×9/1                 | (27,27,512)  |
| Conv2         | 9×9/1                 | (19,19,256)  |
| Conv3         | 7×7/1                 | (13,13,128)  |
| PrimaryCaps(Conv4) | 9×9×8/2               | (288,8)      |
| DigitCaps     | 288×8+8×K×16          | (K,16)       |

On the network architecture, ex-CapsNet adds only two common convolutional layers over the original CapsNet. However, this simple approach can effectively reduce the number of capsules and accelerate the iterative process of dynamic routing.

Moreover, the main change from the original CapsNet is the initialization method of the convolutional layer. We pretrain a CNN model with four convolutional layers, where the convolutional parameters are set in the same way as the convolutional layer in ex-CapsNet. In practice, PrimaryCaps layer is realized by a set of convolution operations, that is, PrimaryCaps layer consists of 8 convolution layers with a kernel of 9×9. Therefore, the convolution parameters of pre-trained CNN model can be used in convolutional layers in ex-CapsNet, and this initialization strategy can effectively avoid the loss to the local minimum. In addition, parameters in the DigitCaps layer are randomly initialized. During the iteration, update each parameter with the Adam optimizer, where the hyper-parameters ε=10^{-8}, β1=0.9, β2=0.999 and the learning rate lr is set to 0.001. At the end of ex-CapsNet, the norm of capsule N in DigitCaps is calculated by the L2 norm function, and the class corresponding to the maximum value represents the recognition result of wildfire.

3. Experiments

3.1. Dataset
In order to evaluate the performance of ex-CapsNet, the remote sensing data of Himawari-8 (H-8) meteorological satellite is used as the dataset in this paper. H-8 is an advanced geostationary meteorological satellite designed and manufactured by Japan aerospace exploration agency. H-8 has 16 observed wavelengths, which means it can take 16 different images. Where the infrared image taken at the 7-th wavelength is mainly used for natural disaster observation, while the infrared image taken at the 14-th wavelength is used for cloud (fog) observation.

We extract the main areas of these two original remote sensing images and resized the images. Therefore, two sets of image datasets with the size of 35 × 35 pixels are obtained. Combining these two images, the tensor of 35× 35×2 is used as the input of the network. Where the image of the 7-th wavelength is used to extract the wildfire information, and the image of the 14-th wavelength is used to reduce the interference of clouds (fog).

3.2. Implementation Details
In the routing process, a significant parameter that determines whether CapsNet can obtain the optimal coupling coefficient in the routing process is the number of routing iterations. Therefore, it is essential to choose a reasonable routing number and apply to subsequent experiments. In the experiment, we only changed the routing number and kept the other parameters unchanged. The experimental results are shown in Figure 3. It can be seen that the prediction accuracy first increased and then decreased with the increase of the routing number, and accuracy reached its maximum when the routing number is 2. A
small iteration number result in insufficient training, while a big iteration number may cause overfitting. In addition, the training time increases with the number of iterations. According to the experiment, the selected route frequency is 2 and apply this value to the subsequent experiments.

![Figure 3. The influence of routing number on the wildfire recognition accuracy](image1.png)

Capsule consist of a group of neurons and is the core of CapsNet. Capsules in PrimaryCaps layer are lower-level capsules that can detect small pattern in the image. While the higher-level capsules with higher dimensions in DigitCaps layer detect more complex pattern. When the dimension of the capsule is small, the detection ability of the capsule is feeble, which leads to the decline of the recognition accuracy of wildfire. In contrast, higher-dimensional capsule may contain redundant information or even noise. Considering the importance of capsule dimensions, we set a set of values ((4,8), (6,12), (8,16), (10,20)) to assess the impact of capsule dimensions. The experimental results are shown in Figure 4. Prediction accuracy reached its maximum value when PrimaryCaps layer capsule dimension is set as 8 and FinalCaps layer capsule dimension is set as 16.

![Figure 4. The influence of the dimension of the capsule on the wildfire recognition accuracy](image2.png)

3.3. Experimental Results

3.3.1. Compare to the original CapsNet

![Prediction Accuracy vs. Training Time](image3.png)

(a)original CapsNe
This section we compare the proposed network architecture with the original CapsNet, where the routing number and capsule dimension of both are using the optimal configuration in section 3.2. For fair comparison, we used the same training datasets and test datasets to train and test CapsNet and ex-CapsNet respectively. The relationship between prediction accuracy and training time of the two network architecture is shown in (a) and (b) respectively. It can be seen that ex-CapsNet is significantly better than the original CapsNet in terms of convergence speed and accuracy, which indicates that ex-CapsNet with the new architecture and initialization strategy has better performance.

3.3.2. Compare with vgg-16 and Inception-v3

VGG-16[11] is the most representative sequential CNN architecture with simple structure and strong applicability. It is characterized by the superposition of layers (convolutional layer or fully connected layer) to 16 layers, which is a deep network structure. This model ranked second in the ILSVR image classification challenge in 2014.

GoogLeNet[12] was the first winner of the competition in 2014. Compared with VGG-16, its network structure has depth not only vertically but also horizontally. This is called "Inception" structure and this paper uses the Inception-v3[13] model compared to ex-CapsNet.

The prediction results are presented in Table 2. The ex-CapsNet has a lower error rate than these two popular CNN models. The practicability of the model for wildfire recognition is further proved.

|               | precision | recall | error rate |
|---------------|-----------|--------|------------|
| VGG-16        | 91.51     | 89.92  | 9.33       |
| Inception-v3  | 92.16     | 91.42  | 8.15       |
| ex-CapsNet    | 93.04     | 90.11  | 7.69       |

4. Conclusions

Based on H-8 satellite remote sensing images, this paper proposes a CapsNet-based wildfire recognition model for transmission lines. The proposed network architecture accelerates the routing iteration process by increasing the number of convolutional layers. Secondly, more importantly, it initializes the convolutional layer of ex-CapsNet by using the convolutional parameters of the pretrained CNN model, which improves the convergence speed and prediction accuracy of the network. Compared with the original CapsNet model and two popular CNN models, ex-CapsNet model can more accurately detect wildfire. This wildfire recognition model enhances the ability to prevent and control emergencies and provides a strong support and guarantee for the security and stable operation of transmission lines.

References

[1] Wu Tian. Hu Yi, Ruan Jiang-jun, Liu Kai, Liu Ting, Chen Cheng. Air Gap Breakdown Mechanism of Model AC Transmission Line Under Forest Fires[J]. High Voltage Engineering,37(5):1115-1122 (2011)
[2] Hu Yi, Liu Kai, Wu Tian, Liu Yan, Su Ziming. Analysis of Influential Factors on Operation Safety of Transmission Line and Countermeasures[J]. High Voltage Engineering, 40(11): 3491-3499 (2014)

[3] Mphale K. Heron M. Measurement of electrical conductivity tbr a biomass fire[J]. International Journal of Molecular Sciences, 9(8): 1416-1423 (2008)

[4] Nogueira, K. Penatti, O A B. dos Santos, J.A. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognit., 61, 539–556 (2017)

[5] Hu, F. Xia G S, Hu J, Zhang L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sens., 7, 14680–14707 (2015)

[6] Zou Q., Ni L., Zhang T., Wang Q. Deep learning based feature selection for remote sensing scene classification. IEEE Geosci. Remote. Sens. Lett., 12, 2321–2325 (2015)

[7] S Sabour, N. Frosst, G. E. Hinton, "Dynamic routing between capsules", Proc. Conf. Adv. Neural Inf. Process. Syst. (NIPS), pp. 3856-3866, Dec, (2017)

[8] Yu Chengbo. Xiong Dien. Research on finger vein recognition based on capsule network[J]. Application of Electronic Technique, 44(10): 15-18 (2018)

[9] Aryan Mobiny and Hien Van Nguyen Fast CapsNet for Lung Cancer Screening. MICCAI (2) : 741-749 (2018)

[10] Deng, F. Pu, S.; Chen, X.; Shi, Y.; Yuan, T.; Pu, S. Hyperspectral Image Classification with Capsule Network Using Limited Training Samples., 18, 3153 (Sensors 2018)

[11] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs] (September 2014).

[12] Christian Szegedy et al. Going Deeper With Convolutions. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)

[13] Szegedy, C. Vanhoucke, V.; Ioffe, S.; Shlens, J. Rethinking the inception architecture for computer vision. arXiv; arXiv:1512.00567 (2015)