Off-position detection based on convolutional neural network

Tianbing Zhang[1],Wang luo[1, a],Qiwei Peng[1],Gongyi Hong[1],Min Feng[1], Yuan Xia[1], Lei Yu[1],Xu Wang[2]and Yang Li[2]

[1]*Nari Group Corporation (State Grid Electric Power Research Institute),China*
[2]*State Grid Xinjiang Electric Power Science Research Institute, China*

Abstract. As a part of the intelligent video surveillance, off-position detection, which needs a real-time and precise algorithm, is used to detect whether the person on duty is absent from working position. This work is necessary for improving efficiency and reducing human resource consumption. Considering the excellent performance of convolutional neural network in image classification, we first propose a method for off-position detection using CNN in this paper and get good results. Furthermore, we introduce a new dataset for working position by generating crops from video frames. Then we randomly generate 224×224 crops from training images to fine-tune our deep neural network.
Keywords: off-position detection; CNN; classification.

1 Introduction

There are lots of methods for off-position detection such as object detecting and tracking. But background modeling, foreground object detecting and tracking can't satisfy real-time or achieve a high precision in many cases. A good performance convolutional neural network has achieved makes it an outstanding algorithm in classification and many other computer vision tasks. In our work, we regard off-position detection as a two-class classification problem, and use CNN to train a model to detect whether the person is absent.

The pre-trained model based on the large-scale dataset is important to train a network for a specific task [6]. Since there are millions parameters in neural network, using thousands of images to train the model will cause serious over-fitting. Transferring the pre-trained model is useful for obtaining better initial parameters, avoiding local minima and getting rich low-level features.

[a]Corresponding author: luowang@sgepri.sgcc.com.cn

Figure 1. Examples of off-position detection. The red boxes indicate there is no person.

In this paper, we focus on the classification for working position. Owing to no dataset available for working position, a new dataset is built. Before training specific network using a small dataset, the parameters of the pre-trained model are transferred. To obtain a pre-trained network, we use Image Net [5] which is the largest database including millions of labeled images in thousands of categories. Then the convolutional neural network is employed to train the model for off-position detection. In order to avoid over-fitting, we obtain 224×224 crops from each training image randomly. Meanwhile dropout is set to 0.5 to solve this problem. As far as we know, this is first work which applies convolutional neural network to off-position detection. We show some examples that the person is away from working position in Figure 1.

In section 2, we review applications of deep learning in image processing and traditional methods for image classification. In section 3,we introduce the dataset and proposed method. Section 4 describes the details of experimental setup and presents the results of off-position detection. Section 5 is the conclusion of our work.

2 Related work

Deep learning is a new branch of machine learning and has shown its good performance in a lot of areas: automatic speech recognition, natural language processing and image recognition. Deep learning has achieved very good results in the field of image processing, and has been widely used in image classification [1],object detection and semantic segmentation. Then, we review previous works on image classification and deep learning networks.

2.1 Image classification

According to the different features extracted from the image, different classes of image can be distinguished by image classification algorithm. Image classification is given some predefined labels as well as training sets to predict the label of any unknown images. There are some kinds of traditional methods for image classification .Generally, these works first extract the features of image, and then

train the classifier by coded features. Traditional feature extraction methods like SIFT [11] and HOG [4]are dependent on human experiences, so they sometimes can't get satisfied results. These extracted features are finally fed to the classifiers such as random forest [3] or SVM [7].

Sometimes it is difficult to choose a proper feature extraction method, because the effect of these methods relies heavily on feature extraction. So we need a universal method which can perform well for most cases. Convolutional neural network uses the whole image as input, and obtains the features by different convolution filters from different layers.

2.2 Deep convolutional networks

Several deep architectures are constructed such as AlexNet[10],GooLeNet [13] and VGG-Net [12].AlexNet as a typical convolutional neural network, consists of 5 convolutional layers,2 fully connected layers, and a label layer with 1000 nodes. Compared with AlexNet, VGG-Net has more layers, usually 16 to 19 layers and the size of convolutional layer is smaller than AlexNets. VGG-Net won the first and the second place on the ILSVRC 2014 localization and classification challenges respectively. A better performed and more complicated architecture is GoogLeNet, of which the benefits are experimentally proved on the ILSVRC 2014 detection and classification challenges, where it outperforms other current excellent networks. Our network is based on AlexNet, since this network is small and saves storage.

Since there are millions of parameters in the network, we need to pay attention to the over-fitting problem. To solve this problem, a large-scale database is used to pre-train a model. Then the specific model is trained by a small-scale dataset using fine-tune processing. And dropout [8] is another trick which randomly drops some of the hidden layer nodes during training period.

3 Proposed method

In this section, we first introduce the dataset for off-position detection. Then the structure of the convolutional neural network are described.

3.1 Dataset for off-position detection

Since there is no existing dataset for working position to our knowledge, a new dataset is built. We show some examples in Figure 2.Positive examples are images that the person on duty is absent. Obviously negative examples are images that the person is present. Examples are all from surveillance video frames. While the person is absent from working position, the video scene is almost still. Obtain examples directly from video frame can lead to over-fitting due to a large number of parameters and a small size of training dataset. Therefore we build our dataset by obtaining training images from video frames and resize them to 256×256 pixels. Then we generate 224×224 crops from each training image randomly during the training phase. This trick can make our convolutional network more robust for slight shake. There are 2367 images in our dataset. We use 1556 images to fine-tune the pre-trained network and get the model for our work, then use the rest 811 images to test.

Figure 2 . Examples of the dataset. The first line is the examples of the person on duty is absent and the second line is present.

3.2 Deep convolutional network for off-position detection

Our convolutional network is based on Alex convolutional network, a classic framework .This architecture uses ReLU to replace traditional activation functions such as tanh or sigmoid activation function. And it improves the recognition rate by locally normalizing the response of the same layer adjacent nodes. Our network consist of 5 convolution layers, 2 normalization layers,3 pooling layers and inner product layer ,then outputs a single label for classification. The structure of neural network is shown in Figure 3.And the configuration of our convolutional neural network is shown in Table 1.When the output probability is higher than 0.7,we consider that the person is absent from working position.

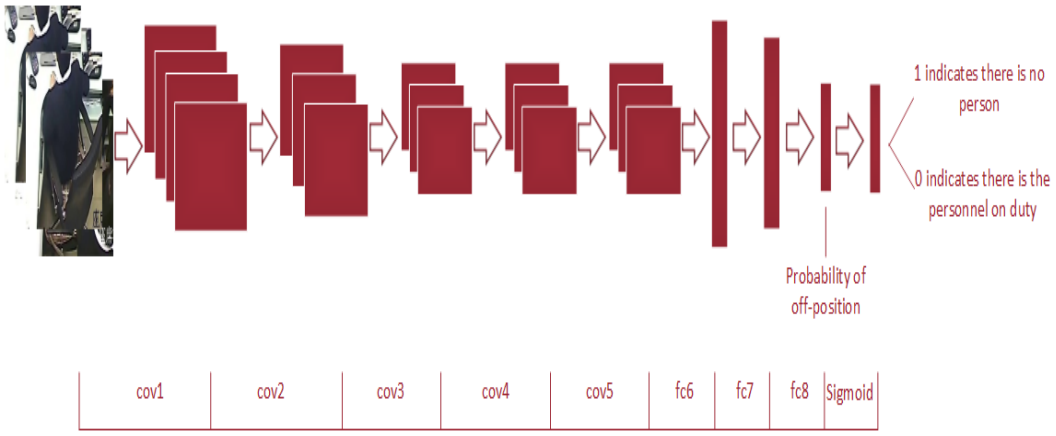

Figure 3 . The structure of neural network. Here, “conv”indicates the convolutional layer and “fc” indicates the full connected layer.

To train a network optimized for this single-label classification, we use Euclidean function as our loss function. The backward propagation and gradient descent optimization [2] are adopted to iteratively update the parameters of model. The dropout rate is set to 0.5, which means 50% parameters are randomly not chosen to be used in every iteration, to avoid over-fitting.

4 Experiment

In this section, we introduce the implementation details as well as the experimental setup. Then we analyze the proposed network and show the results of off-position detection.

Table 1. The configuration of our convolutional network.

name	Kernel size	stride	pad	Output size
input	-	-	-	224×224×3
cov1	11	4	100	104×104×96
cov2	5	1	2	52×52×256
cov3	3	1	1	26×26×384
cov4	3	1	1	26×26×384
cov5	3	1	1	26×26×256
fc6	6	1	0	8×8×4096
fc7	1	1	0	8×8×4096
fc8	1	1	0	1×1×1

4.1 Implementation details

Training images from video frames are resized to 256×256 pixels .To construct the dataset for working position, we generate 224×224 crops from training images randomly. This trick achieve a good performance in improving the robustness for small shake of video screen. We randomly split our dataset into training or testing sets:65% images for training and others for testing.

We based our network on AlexNet. The dimension of fc8 is 1000 in pre-trained convolutional neural network, since Image Net consist of 1000 classes totally. For our dataset, we use an inner product layer with 1 output to replace fc8.Parameters optimization is based on the Euclidean loss function. The back propagation and gradient descent are employed to update the parameters in100000 iterations. Our network is based on the Caffe [9] framework. The initial learning rate is set to 0.00001 and the learning rate is continually adjusted while training period. The dropout rate is set to 0.5,which means in every iteration there are 50% parameters are not updated.

4.2 Results of evaluation

During the test period, the test image is fed to the convolutional neural network. The probability output higher than the threshold indicates that the person on duty is absent from working position.

Figure 4 . There are some test results. Here, red box indicates that there is no person while blue box indicates that there is the person on duty.

Several results of the detection are shown in Figure 4.While the person is present from working position, the accuracy rate is 95.71%.And while the person is absent ,the accuracy rate is 96.26%.We can see that our method can achieve 95.93% accuracy rate totally. The high accuracy rate indicates that our network achieves a pretty good results and the network can detect whether the person is absent accurately. But there are still some false or missing detection, especially when the part of body in the detection box. We can improve our dataset for adding some images in which there are only part of bodies to positive examples.

5 Conclusion

In this paper, we build a new dataset for working position. A convolutional neural network is employed to a single-label classification. This is the first time that CNN is used for off-position detection. The back propagation as well as gradient descent optimization are applied to optimize the network parameters. The experimental results shows the great performance of our network. In the future, we are interested in improving our deep network to adapt to more complex scenes of working position.

Acknowledgment

This work was supported in part by the Natural Science Foundation of Jiangsu Province (Grants No. BK 20130107).

References

1. I. S. A. Krizhevsky and G. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.
2. C. M. Bishop and N. M. Nasrabadi. Pattern recognition and machine learning. 2006.

3. L. Breiman. Random forests. Machine learning, 2001.
4. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005.
5. J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang,E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. arXiv preprintarXiv:1310.1531, 2013.
6. D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent,and S. Bengio. Why does unsupervised pre-training help deep learning? The Journal of Machine LearningResearch,11:625–660, 2010.
7. M. A. Hearst, S. Dumais, E. Osman, J. Platt, andB. Scholkopf. Support vector machines. Intelligent Systemsand their Applications, 1998.
8. G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, andR. R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580, 2012.
9. Y. Jia. Caffe: An open source convolutional architecture for fast feature embedding. Inhttp://caffe.berkeleyvision.org/,2013.
10. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenetclassification with deep convolutional neural networks. InAdvances in neural information processing systems, pages 1097–1105, 2012.
11. D. Lowe. Object recognition from local scale-invariant features.In ICCV, 1999.
12. K. Simonyan and A. Zisserman. Very deep convolutionalnetworks for large-scale image recognition. arXiv preprintarXiv:1409.1556, 2014.
13. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed,D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich.Going deeper with convolutions. Eprint Arxiv, 2014.