A Basic Study on Human Detection for Train Operation Video

Jie Zhang Hiroshi Watanabe

Graduate School of Global Information and Telecommunication Studies, Waseda University

Abstract: To ensure the safety of train service, utilization of information technology has been paid attention. Accordingly, we are studying on developing an intelligent video analysis system which is able to detect abnormality automatically from the video of train operation and warm the staff simultaneously. This paper introduces a fundamental investigation into human detection technology in the field of video of train operation by means of deep learning, which is known as a new type of neural network.

1. Introduction

With the development of image analysis technology, various of human detection methods using the combination of learning methods and image features have been proposed.

In the typical usage, HOG (Histograms of Oriented Gradients) and SVM (Support Vector Machines) or Boosting are used to object recognition. However, they only capture low-level edge information and it seems to be difficult to design features that effectively capture mid-level cues or high-level representation. However, recent developments in machine learning, known as "Deep Learning", have shown how hierarchies of features can be learned directly from data. The approach dispense with designing of feature and is able to avoid overtraining.

In this paper, we introduce one type of deep learning which is called Convolutional Neural Network, and show it's architecture. We also demonstrate its effectiveness and show the results of experiment in the end of this paper.

2. Related Work

We selected deep learning to realize effective human detection. For comparison purpose, we picked up two types of features in the conventional approaches. One has been mentioned in the part of introduction called Histograms of Oriented Gradients and the other is known as Local Binary Patterns.

On the other hand, we chose Boosting as learning method because it has always outperformed Support Vector Machines in the conventional experiments.

We randomly picked 3500 images and 1000 images from datasets for training and testing, respectively. As a result, deep learning provides better recognition rate of 96.5% for overall data.

3. Proposed Method

Since we confirmed the better recognition rate provided by deep learning, we picked up this convolutional neural network for passenger detection from train operation video. The architecture of our network is summarized in Figure 1, which contains of 7 learned layers - 3 convolutional, 2 sub-

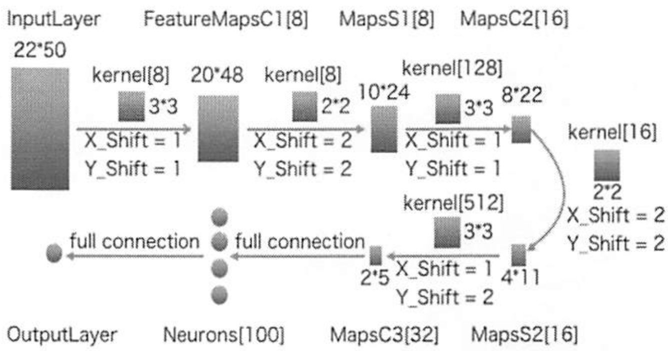

Figure 1: Illustration of the architecture of our CNN.

sampling and 2 fully connected. Below, we describe some of the novel or unusual features of our network architecture.

3.1 Activation Function

Selection of an optimal activation function is an important part of the neural network. Generally speaking, the activation function should be symmetric, and the neural network should be trained to a value that is lower than the limit of the function.

In our neural network, we select hyperbolic tangent as the activation function because it is completely symmetric. This function is a good choice also because of the hyperbolic tangent is easy to obtain its derivative. Since y = tanh(x), the derivative will be 1 - y^2. This result means that we can calculate the derivative given only the output of the activation function. This merit will help us to reduce back propagation time which is an iterative process that starts with the last layer and moves backwards through the layers until reach to the first layer.

3.2 Learned Linear Combination Subsampling

Subsampling have been widely used in the convolutional neural networks to reduce the dimensionality of a high-dimensional output of the convolutional layer. In the domain of 2D filter outputs, subsampling can also be thought of as increasing the position invariance of the filters.

In practice, we used 3 kinds of subsampling methods. The learned linear combination subsampling reached to the optimality. Figure 2 shows their calculation procedures.

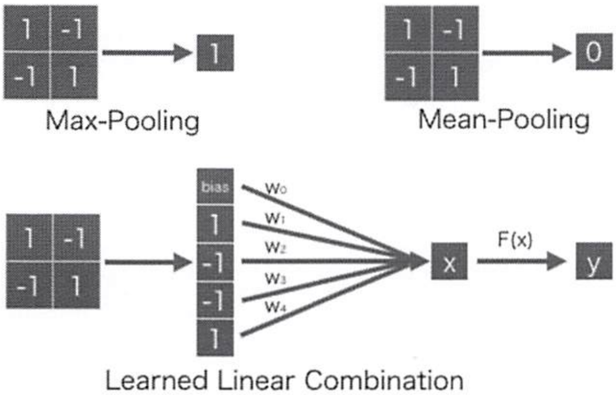

Figure 2: Calculation of Max-Pooling, Mean-Pooling and Learned Linear Combination.

4. System Overview

The system of human detection using proposed method can be divided into acquisition unit to detect region, detection unit and error feedback unit. The detection region acquisition unit detects and perceives regions of human appears, such as platform, railroad crossings and so on. The detection unit recognizes and locates human from detection region. The error feedback unit records false-positive detections, and save them to negative file.

5. Experimental Result

We prepared 7090 images including 1350 positive images and 5740 negative images manually and automatically cut from several train operation videos via YouTube© for training. For testing, we downloaded a train operation video for our simulation. The results of human detection are shown in Figure 3 as a typical case.

At the first detection to the frame, we got five positive results including one false-positive detection. Through the error feedback unit, we added this false-positive detection into the training set and retrained the network. The retesting results are shown in Figure 4, which modified the false-positive detection.

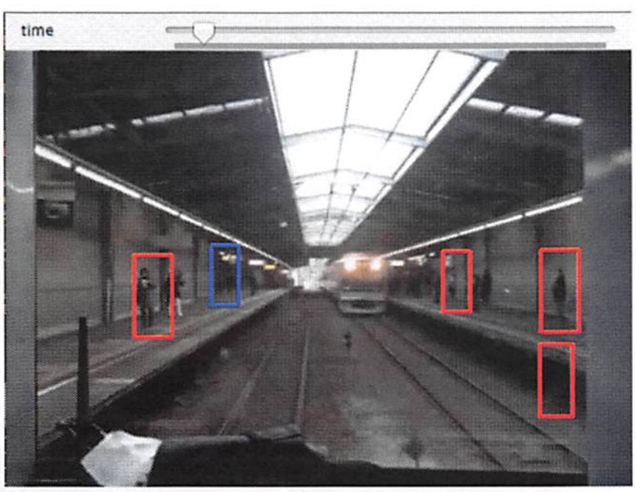

Figure 3: Human detection without error feedback unit.

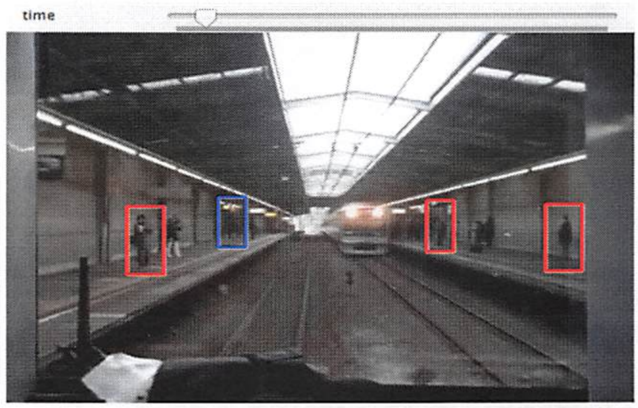

Figure 4: Human detection with error feedback unit.

6. Conclusion

As a basic study on human detection for train operation video, the effectiveness of the deep convolutional neural network has been demonstrated. From the experimental results, although the resolution of video is slightly low, we still acquired an acceptable error rate which is averagely 0.25/frame. Further development is expected in the future works.

Acknowledgment

The research results have been achieved by "Research and Development on Fundamental and Utilization Technologies for Social Big Data", the Commissioned Research of National Institute of Information and Communications Technology (NICT), JAPAN.

References

1. A. Krizhevsky.: "Convolutional Deep Belief Networks on CIFAR-10," http://www.cs.utoronto.ca/~kriz/conv-cifar10-aug2010.pdf.
2. D. Ciresan, U. Meier, and J. Schmidhuber.: "Multi-column Deep Neural Networks for Image Classification," http://people.idsia.ch/~juergen/cvpr2012.pdf.
3. Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, L.D. Jackel, et al.: "Handwritten Digit Recognition with a Back-Propagation Network," http://yann.lecun.com/exdb/publis/pdf/lecun-90c.pdf.
4. Y. LeCun, K. Kavukcuoglu, and C. Farabet.: "Convolutional Networks and Applications in Vision," In Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on, pages 253-256, May 30 2010 - June 2 2010.

Graduate School of Global Information and Telecommunication Studies
Waseda University
〒169–0072 Shillman Hall-401, 3-14-9 Okubo, Shinjuku, Tokyo
Phone: +81–3–5286-2509
E-mail: ketu.tyowu@ruri.waseda.jp