An Effective Normalization Algorithm at Deep Learning for Passenger Detection in Railway Operation Videos

Jie Zhang*, Hiroshi Watanabe*

Graduate School of Global Information and Telecommunication Studies, Waseda University

Recent academic developments in the field of machine learning, known as "Deep Learning", have shown how hierarchies of features can be learned. In previous studies, we proposed a deep convolutional neural network to realize effective passenger detection which achieved an error rate of 5%. In this work, an effective normalization algorithm is proposed in order to reduce the error rate of detection. In our experiments, it is proved that the algorithm we proposed provide a better results with lower error rate. Moreover, in contrast to the previous studies, the detection rate was improved through the use of proposed algorithm. Combining our proposed normalization algorithm with our detection network, the passenger recognition results reached optimality.

1. Introduction

Recently, "Deep Learning" has been focused on in the area of artificial intelligence and machine learning. Deep learning is one of the convolutional neural networks to emulate human neural circuit. The advantage of this neural network is a capability to obtain an optimum feature automatically by its self-learning scheme. Specifically, the convolutional neural network is suitable for object recognition in images [1].

A convolutional neural network consists of multiple layers which are concatenated with some optional layers. As optional layers, "Normalization layer", "Rectification layer", and "Max-pooling layer" can be used. However, the effect of these layers including parameter settings is unclear. In this paper, the effect of these layers for the human detection application is investigated.

2. Rectification Layer

Rectification is a function to take R(x) = max (0, x) for input x. In recent studies, the rectification layer is used in the learning of the multilayered neural network widely [2]. In our experiments, we place this rectification layer after the convolutional layer, which is activated by the function of hyperbolic.

3. Max-pooling Layer

Max-pooling is an operation that extracts the maximum value of neurons in a local region [3]. Because it's stepwise translation invariance, the layer of max-pooling has a significant effect on the accuracy of image recognition. In our experiments, recognition rate will be investigated when the max-pooling layer is placed after the convolutional layer and rectification layer.

4. Normalization Layer

Normalization in a multilayer convolutional neural network is a process of generalization by modifying the data of feature maps according to a criterion. In our experiments, the normalization layer will be added only after the input layer. And the normalization process can be carried out according to the following formula [4].

$$b_{x,y}^{i} = a_{x,y}^{i} \div \left(k + \alpha \times \sum_{j=\max(0,\, i-n\div 2)}^{\min(N-1,\, i+n\div 2)} \left(a_{x,y}^{j}\right)^2 \right)^{\beta}$$

5. Proposed Method

The architecture of our proposed convolutional neural network is summarized in Fig.1, which is contains of 10 learned layers - 1 normalization, 5 convolutional, 2 max-pooling and 2 fully connected.

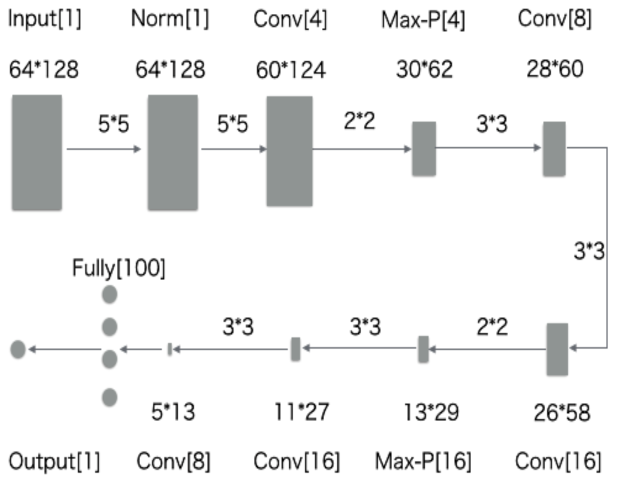

Fig.1 Illustration of the architecture of our CNN

〒169-0072
Shillman Hall-401, 3-14-9 Okubo, Shinjuku, Tokyo, Japan
e-mail: ketu.tyowu@fms.co.jp

6. Experiment

In order to clarify the effects of "Normalization layer", "Rectification layer", and "Max-pooling layer" in the convolutional neural network for passenger recognition, five kinds of convolutional neural networks will be constructed. The combinations of these layers are shown below.

1) Convolutional
2) Convolutional + Rectification
3) Convolutional + Max-pooling
4) Convolutional + Rectification + Max-pooling
5) Normalization + Convolutional + Max-pooling

6.1 Experimental Outline

In these experiments, we try to recognize the passenger images from evaluation sample which is have no intersection with learning sample. And "INRIA Person Dataset" will be used. The learning sample contains 3500 images and evaluation sample contains 1000 images are automatically and randomly pick from the dataset.

6.2 Experimental Parameter

The parameters corresponding to experiment 1~5 are shown in Table 1~5.

Table 1 Parameters of experiment 1

	Width*Height*Maps	Filter	Stride
Input	64*128*1	--	--
Conv	60*124*4	5*5	1*1
Conv	29*61*8	4*4	2*2
Conv	27*59*16	3*3	1*1
Conv	13*29*16	3*3	2*2
Conv	11*27*16	3*3	1*1
Conv	5*13*8	3*3	2*2
Fully	1*1*100	--	--
Output	1*1*1	--	--

Table 2 Parameters of experiment 2

	Width*Height*Maps	Filter	Stride
Input	64*128*1	--	--
Conv	60*124*4	5*5	1*1
Rect	60*124*4	--	--
Conv	29*61*8	4*4	2*2
Rect	29*61*8	--	--
Conv	27*59*16	3*3	1*1
Rect	27*59*16	--	--
Conv	13*29*16	3*3	2*2
Rect	13*29*16	--	--
Conv	11*27*16	3*3	1*1
Rect	11*27*16	--	--
Conv	5*13*8	3*3	2*2
Fully	1*1*100	--	--
Output	1*1*1	--	--

Table 3 Parameters of experiment 3

	Width*Height*Maps	Filter	Stride
Input	64*128*1	--	--
Conv	60*124*4	5*5	1*1
Max-P	30*62*4	--	2*2
Conv	28*60*8	3*3	1*1
Conv	26*58*16	3*3	1*1
Max-P	13*29*16	--	2*2
Conv	11*27*16	3*3	1*1
Conv	5*13*8	3*3	1*1
Fully	1*1*100	--	--
Output	1*1*1	--	--

Table 4 Parameters of experiment 4

	Width*Height*Maps	Filter	Stride
Input	64*128*1	--	--
Conv	60*124*4	5*5	1*1
Rect	60*124*4	--	--
Max-P	30*62*4	--	2*2
Conv	28*60*8	3*3	1*1
Rect	28*60*8	--	--
Conv	26*58*16	3*3	1*1
Rect	26*58*16	--	--
Max-P	13*29*16	--	2*2
Conv	11*27*16	3*3	1*1
Rect	11*27*16	--	--
Conv	5*13*8	3*3	1*1
Fully	1*1*100	--	--
Output	1*1*1	--	--

Table 5 Parameters of experiment 5

	Width*Height*Maps	Filter	Stride
Input	64*128*1	--	--
Norm	64*128*1	--	1*1
Conv	60*124*4	5*5	1*1
Max-P	30*62*4	--	2*2
Conv	28*60*8	3*3	1*1

Conv	26*58*16	3*3	1*1
Max-P	13*29*16	--	2*2
Conv	11*27*16	3*3	1*1
Conv	5*13*8	3*3	1*1
Fully	1*1*100	--	--
Output	1*1*1	--	--

7. Experimental Result

Experimental Results are shown in Table 6. It shows that the passenger recognition results reached optimality by using proposed method.

Table 6 Experimental Results

Experiment	True Positive	True Negative	Average
1	95.2%	97.2%	96.6%
2	94.9%	97.6%	96.8%
3	96.2%	98.6%	97.9%
4	96.9%	98.4%	98.0%
5	96.9%	99.0%	98.4%

8. Conclusion

From the experimental results, the effectiveness of normalization has been demonstrated. It is proved that the algorithm we proposed provide a better results with lower error rate. Moreover, in contrast to the previous studies, the detection rate was improved through the use of proposed algorithm. Combining our proposed normalization algorithm with our detection network, the passenger recognition results reached optimality.

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) (2010)
2) Y. Sun, X. Wang, and X. Tang. Deep Convolutional Network Cascade for Facial Point Detection. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2013)
3) J. Zhang, D. Ishii, and H. Watanabe: "Effect of Rectification and Max-pooling at Deep Learning for Object Recognition," IPSJ National Conference, 2P-08, March (2015) [in Japanese]
4) J. Zhang, D. Ishii, and H. Watanabe: "Effect of Normalization at Deep Convolutional Neural Networks for Human Recognition" IEICE National General Conference, D12-24, March (2015) [in Japanese]
5) J. Zhang, H. Watanabe: "A basic study on human detection for train operation video," PCSJ-IMPS 2014, No.I-5-01, pp.188-189, Nov. (2014)