얼굴 검출을 위한 캐스케이드 CNN 정확도에 관한 연구

우위네마 조세린*, 이해연*
*금오공과대학교 소프트웨어공학과
e-mail: zenejoselyne@kumoh.ac.kr

A Study on Cascaded CNN Accuracy for Face Detection

Uwinema Joseline*, Hae-Yeoun-Lee**,
*Dept of Software Engineering, Kumoh National Institute of Technology

Abstract

Convolutional Neural Network is arguably the most popular deep learning architecture that is one of the most attractive area of research since it has various applications including face detection and recognition. The cascaded CNN operates at multiple resolution and rejects the background regions in the fast low resolution stages. By considering that advantage, we carry out the study on accuracy of cascaded CNN for face detection applications. The key point for our study is to analysing and improving the accuracy of cascaded CNN by applying simulations of algorithm where by we used Google's Tensorflow GPU as deep learning framework.

1. Introduction

Neural networks concepts is a generic name for a large class of machine learning algorithms, including; hopefield networks, fully connected neural networks, convolutional neural networks, recurrent neural networks, long short term neural networks and so on, which is mostly trained with an algorithm known as back-propagation as mentioned in [1]. In the late eighties the denomination in neural networks and machine learning in general was fully connected neural networks and they have a large number of parameters or weights. By contrast, the convolutional neural networks which could be considered essentially a not fully connected neural nets which means that each neuron is connected to only a few neurons in the previous layer and neurons share weights. Thus, such kind of networks have proven successfully especially in the field of computer vision as well as natural language processing. the success of convolutional neural networks was the main reason why neural networks which recently called deep learning has become a hot topic for researches in the past 6 years.

Particularly, since their introduction by LeCun et al in the early 1990′s, Convolutional Network or convnets have demonstrated excellent performance at tasks such as face detection and hand-written digit classification [2, 3]. The convnets are made up of neurons that have learnable weights and biases. Convnet architectures make the explicit assumption that the inputs are images, which allows us to encode certain properties into the architecture. These play the role of making forward the function more efficient to implement and widely reduce the amount of parameters in the network [4] which is one of the key point of determining the accuracy of cascaded CNN for face detection. Following the emerging trend of exploring deep learning for face detection, in this paper, we present the method used to determine the accuracy of cascaded CNN for face detection by extending the previous study done on CNN cascade for face detection algorithm [5].

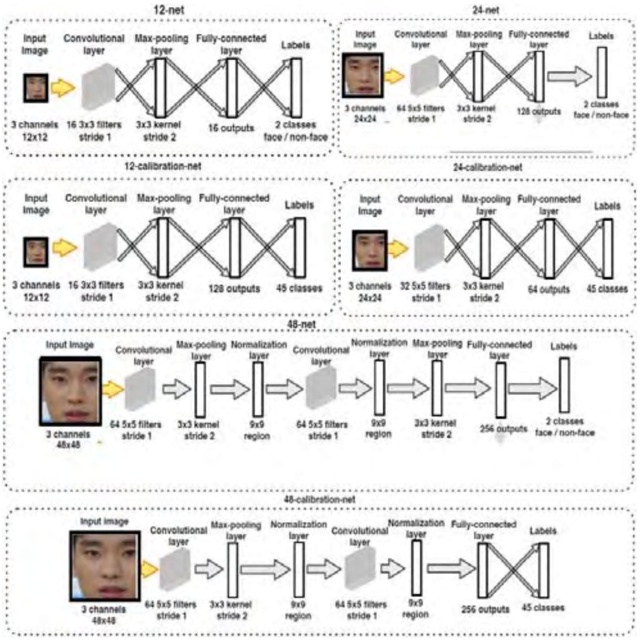
2. Related works

many researches prove that CNN is go-to model on image related fields whereby in terms of accuracy they blow ideas out of the water. The main advantage of CNN compared to its antecedent researches is that it automatically detects the important features without any human interaction The detection approach can be considered as a special type of object detection task in computer vision. Researchers thus have attempted to tackle face detection by exploring some successful deep learning techniques for generic object detection tasks [9]. One of very important and highly successful framework for generic object detection is the region-based CNN (RCNN) method [9, 10], which is a kind of CNN extension for solving the object detection tasks. A variety of recent advances for face detection often follow this line of research by extending the RCNN and its improved variants.

3. Proposed algorithm

In this paper, we present our contribution of the idea in [11] where we focus on determining the accuracy of the cascaded CNN while detecting the face. Compared to the previous handcrafted features, CNN can automatically learn features to capture complex visual variations by leveraging a large amount of training data and its testing phase can be easily parallelized on GPU core for acceleration [11, 12]. The GPU uses NIVIDIA GeForce 1080Ti, and the image size is set to 100.

For training process, joint training architecture known as FaceCraft is used [12, 13] to train the network at once. During training, the network takes an image of size 48 × 48 as input, and outputs one joint loss of three branches. The three branches are called x12, x24, x48 respectively, corresponding to the input size of each network.



(Fig. 1) CNN Used for Face Detection

From Fig. 1 shown above, the cascaded CNN method for face detection has more advantages in efficiency than traditional cascade, where by different stages in the cascade can be jointly trained to achieve better performance.

Table 1 is a description of some CNN Parameters used to increase accuracy in 12-net. 24-net and 48-net

(Table 1) CNN parameters for each net

Parameters	12-net	24-net	48-net
Learning rate	0.001	0.001	0.001
Threshold	0.1	0.003	0.003
input channels	3	5	5
batch size	3	5	5

3.1 Cascaded CNN concept and structure

Practically, Cascaded CNN can have varied settings for accuracy computation trade off [13]. Mainly, the cascaded CNN for face detection contains three stages. In each stage, one detection network and one calibration network is applied [13]. There are six CNNs in total that including 3 CNNs for face vs non-face binary classification and 3 CNNs for bounding box calibration [14]. In this case training process is quiet complicated but training samples are carefully prepared for all stages and optimize each and every network These CNNs analyzed based on AlexNet to apply ReLU non-linearity function after the pooling layer and fully-connected layer [17], and also drop-out before classification or regression layer.

Non Maximum Suppression(NMS) was applied to reject the highly overlapped boxes.

3.2 Dataset

In training by using cascaded CNN algorithm, our dataset called 3R. 3R contains about 26000 images that have faces and 27000 images that have no faces. 3R is collected from online social network, the image on which is a reflection of the real world images in everyday life. To add negative samples, we also use images in PASCAL VOC2012 [4] that do not contain persons as background image. In total, the dataset contain 47211 images with 82987 faces and about 32000 background images.

4. Experimental results

In this study, we carry out experimental results on face detection dataset to determine the accuracy of cascaded CNN based algorithm as shown in Table 2.

(Table 2) Cascade CNN-based algorithm accuracy

Stages	epoch	iteration	High Accuracy calculated(%)
12-net	from 0-9	3000/10000 6000/10000 9000/10000	75%
24-net	from 0-9	3000/10000 6000/10000 9000/10000	85%
48-net	from 0-9	3000/10000 6000/10000 9000/10000	87%

For joint loss, each branch has face vs non-face detection classificational loss and a bounding-box regressional loss. By adding the two with loss weights we get the joint loss function as illustrated below;

$$L_{\text{joint}} = \lambda_1 L_{x12} + \lambda_2 L_{x24} + \lambda_3 L_{x48},$$

Where Lx12, Lx24 and Lx48 are different losses of three branches and λ1, λ2, and λ3 are loss weights of the three blanches.

As the results, the accuracy is calculated in the training image process where from 12-net to 48-net, the results of accuracy increases from 75% of 12-net to 87% of 48-net. the main factor that increases the accuracy is learning frequency and dropout regularization. Based on the results from Table 2 the accuracy of cascaded CNN based methods for face detection have always been accused of its runtime efficiency. Recent CNN algorithms are getting faster on high-end GPUs. However, in most practical applications, especially mobile applications, they are not fast enough.

5. Conclusion

In this paper, we have presented the study on cascaded CNN accuracy for face detection. By applying joint training method, the accuracy of cascaded CNN is evaluated and the results can be used for future researches of face detection applications

Acknowledgement

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2017R1D1 A1B03030432).

References

[1] Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, Gang Hua "A Convolutinal Neural Network Cascade for Face Detection"

[2] Fan Yang, Wongun Choi, Yuanqing Lin “Exploit All the Layers: Fast and Accurate CNN Object Detector with Scale Dependent Pooling and Cascaded Rejection Classifiers”

[3] Chenchen Zhu, Yutong Zheng, Khoa Luu, Marios Savvides “ CMS RCNN: COntextual Multi-Scale Region Based CNN for Unconstrained Face Detection”

[4] X. Shen, Z. Lin, J. Brandt, and Y. Wu. Detecting and aligning faces by image retrieval. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2013.

[5] C. Zhang and Z. Zhang. A survey of recent advances in face detection. Technical Report MSR-TR-2010-66, 2010.

[6] L. Bourdev and J. Brandt. Robust object detection via soft cascade. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 2, pages 236 - 243. IEEE, 2005. 2

[7] S. S. Farfade, M. Saberian, and L.-J. Li. Multi-view face detection using deep convolutional neural networks. arXiv preprint arXiv:1502.02766, 2015.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, 2012.

[9] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang. Probabilistic elastic part model for unsupervised face detector adaptation. In Proc. IEEE International Conference on Computer Vision, 2013.

[10] Hongwei Qin, Junjie Yan3, Xiu Li1, Xiaolin Hu3,Joint Training of Cascaded CNN for Face Detection, CVF conference 2015

[11] M.J.Er, W.Chen, S.Wu, “High speed face recognition based on discrete cosine transform and RBF neural network”, IEEE Trans. On Neural Network, Vol. 16, No. 3, PP. 679,691, 2005.

[12] D. Chen, S. Ren, Y.Wei, X. Cao, and J. Sun, “Joint cascade face detection and alignment,” in ECCV, 2014. 1, 2, 9, 10

[13] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 2879 - 2886. IEEE, 2012.

[14] S. Yang, P. Luo, C. C. Loy, and X. Tang. From facial parts responses to face detection: A deep learning

approach. arXiv preprint arXiv:1509.06451, 2015.

[15] X. Zhu and D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild,” in CVPR, 2012.

[16] H. A. Rowley, S. Baluja, and T. Kanade, “Neural network-based face etection,” TPAMI, 1998.

[17] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge, 2014.