Journal of Information & Computational Science 12:17 (2015) 6391–6400 November 20, 2015
Available at http://www.joics.com

Convolution Neural Networks Embedding K-means

Manchang Gu [a,*], Siwen Li [a], Li Li [b]

[a] *School of Computer Science, Sichuan University, Chengdu 610065, China*

[b] *State Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University Chengdu 610064, China*

Abstract

In the computer vision domain, Convolution Neural Networks (CNN) is becoming a popular and promising method for object recognition. However the traditional CNN on intermediate layers lacks of transparence. This paper proposes an improved method called k-means CNN (k-CNN) by using a modified k-means algorithm. In this way, we boost the discriminative capacity of the first convolution layer to make the feature representation more learnable. Specifically, we observe that k-CNN can speed up the training time introduced by k-means. We also contribute to find the approximately optimal number of clusters based on computing the sum squared error. Experimental results show k-CNN achieving competitive and efficient performances on CIFAR-10, SVHN and MNIST datasets.

Keywords: Convolution Neural Networks; K-means; Recognition

1 Introduction

Several deep convolution neural networks based approaches have recently been substantially improving the state of the art in image classification, achieving stunning progress on ILSVRC2014[1]. Compare to the previous works on visual representation, the feature maps trained by CNN can preserves quantities of global spatial information while the classical technique such as SIFT [7] and HoG [8] features only could grasp the locally invariant features, i.e., intuitively simple Gabor-like oriented edge detectors are imposed on handcrafted features. Without resorting to carefully handpicked features mentioned above, CNN is aim to automatically learning rich hierarchical features in a multi-layer framework. Due to the fascinating ability of learning complex pattern, by means of extracting the sophisticated feature representations layer by layer. In general, the first convolution layer learns the directed edges or corners features with random noise pixels existing. Then the second and higher layers detect higher features like parts or even abstract concept.

It is noteworthy that the originally proposed LeNet-5 only has five layers in 1990s [2]. Nevertheless the deeper networks make a success in ILSVRC2014, as a captivating exemplar GoogleNet

*Corresponding author.
Email address: manchanggu@gmail.com (Manchang Gu).

[1] ILSVRC2014 stands for ImageNet Large-Scale Visual Recognition Challenge 2014.

1548–7741 / Copyright © 2015 Binary Information Press
DOI: 10.12733/jics20107075

composed of 22 deep layers. Inspired by the architecture of 'going deeper' thought, this indicates that accumulating more multi-layer neural networks may lead to better results. However, the convolution and full connected layers spend most time and largely parameters. To address this problem, we attempt to run a modified k-means algorithm on training dataset to initialize the first convolution layer filter. In this way, feature maps highlight learning discriminative target within the image. On one hand, k-means clustering is a method of vector quantization as a well-known and effective module in various disciplines. But the selection of k value is a sensitive problem which is correlated with the domain knowledge. On the other hand, the last fully connected layers are not only always prone to overfitting in order that lead to poor ability of generalization but also heavily over-parameterized [3].

2 Convolutional Neural Networks

Convolutional Neural Networks (CNN) which is a supervised feed-forward network has sparked renaissance during the past few years in various areas of computer vision. Although CNN primitively focuses on recognizing the handwritten digits, CNN has regained popularity on account of increasing big data like ImageNet [15] and greatly advanced GPU (Graphic Processing Unit). Therefore, CNN has spurred a flurry of breakthroughs on image classification [10], object detection [27], visual tracking [12]. CNN is different from one another in how convolution and subsample layers are realized and arranged. Then the network can be finetuned with standard backpropagation algorithm.

2.1 Convolution Layer

The convolution layer convolves the whole input with a bank of learned filters, each producing one feature map in the output image. As illustrated in Fig. 1, we use filters size with $R \times Q$ as $\omega[r,q]$ while it connects $I_x \times I_y$ input neuron map size followed by convolution layer. The denotation of St defines the stride of skipping pixel and n denotes the current layer. Thus j the output 2D map size can be formulated as:

$$\zeta_x = \frac{I_x - R}{St} + 1, \quad \zeta_y = \frac{I_y - Q}{St} + 1$$

and the operation can be defined as:

$$\zeta^n[i,j] = \mathcal{F}\left(\sum_{R=0}^{R-1}\sum_{Q=0}^{Q-1}\zeta^{n-1}[i+R, j+Q]\,\omega[r,q]\right)$$

where F represents the activation function, $\zeta[i,j]$ is the outcome of pixel at the output layer ζ.

2.2 Subsample Layer

The purpose of the subsample layer is to provide shift invariance by reducing feature map dimensions. There have been emerged different choices for subsample layer: max pooling, average pooling, stochastic pooling [23], etc. A subsample layer will combine the convolved activations

You are reading a preview. **Would you like to access the full-text?**

Access full-text

[16] Ngiam Jiquan, Zhenghao Chen, Daniel Chia, Pang W. koh, Quoc V. Le, Andrew Y. Ng, Tiled convolutional neural networks, In Advances in Neural Information Processing Systems, 2010, 1279-1287

[17] Yangqing Jia et al., Caffe: Convolutional architecture for fast feature embedding, Proceedings of the ACM International Conference on Multimedia, ACM, 2014

[18] Torralba Antonio, Robert Fergus, William T. Freeman, 80 million tiny images: A large data set for nonparametric object and scene recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(11), 2008, 1958-1970

[19] Szegedy Christian et al., Going deeper with convolutions, arXiv preprint arXiv: 1409.4842, 2014

[20] Sermanet Pierre, Soumith Chintala, Yann LeCun, Convolutional neural networks applied to house numbers digit classification, 2012 IEEE 21st International Conference on Pattern Recognition (ICPR), 2012

[21] Mairal Julien et al., Convolutional kernel networks, Advances in Neural Information Processing Systems, 2014

[22] Krizhevsky Alex, G. Hinton, Convolutional deep belief networks on CIFAR-10, Unpublished Manuscript (2010)

[23] Matthew D. Zeiler, Rob Fergus, Stochastic pooling for regularization of deep convolutional neural networks, arXiv preprint arXiv: 1301.3557, 2013

[24] Alex krizhevsky, Geoffrey Hinton, Learning Multiple Layers of Features from Tiny Images, Master's Thesis, Department of Computer Science, University of Toronto, 2009

[25] Strehl Alexander, Joydeep Ghosh, Raymond Mooney, Impact of similarity measures on web-page clustering, Workshop on Artificial Intelligence for Web Search (AAAI 2000), 2000

[26] Matthew D. Zeiler, Rob Fergus, Visualizing and understanding convolutional networks, ECCV 2014, Springer International Publishing, 2014, 818-833

[27] Yi Sun, Xiaogang Wang, Xiaoou Tang, Deep convolutional network cascade for facial point detection, 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2013

[28] Ciresan Dan, Ueli Meier, Jrgen Schmidhuber, Multi-column deep neural networks for image classification, 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012