Universität Hamburg
Department Informatik
Knowledge Technology, WTM

Deep Learning: Neural Networks for Object Detection and Tracking Tasks

Seminar Paper

Brain Modelling

Daniel Speck
Matr.Nr. 632 13 17
2speck@informatik.uni-hamburg.de

29.05.2015

Abstract

Deep neural networks are one of the most successful learning strategies at the moment as the computing power for creating such structures rised in the past years via GPU computing. Object detection and tracking tasks can be fulfilled with these architectures.

Contents

1 Introduction

Deep learning is subcategory of machine learning and the focus of this paper will be deep neural networks in the context of deep learning.
An overview of image classification will be made [1] [2].
The visual cortex and deep learning strategies will be introduced [3].
Approaches for object detection [4] and tracking [5] via deep neural networks will be discussed.

2 Background information: Artificial neural networks

Briefly introduction for classic artificial neural networks without GPU computing.

3 Deep neural networks

Overview for classic deep neural networks. Details about different concepts and approaches.

4 Convolutional neural networks and image processing

For deep learning purposes (classic / fully-connected) multilayer perceptrons consume a sizable amount of resources for proper training when they are designed to solve complex tasks because the amount of neurons and especially weights increases rapidly with the network's size.
For example, a MLP with three layers, an input layer with 100 neurons, a hidden layer with 25 neurons and an output layer with 10 neurons for classifying images with a size of 10x10 pixels into 10 different classes would have $100 * 25 + 25 * 10 = 2,750$ weights/connections. Training this net would already result in a big time and space complexity. Moreover, as features in images capturing real world scenes are distributed in certain patterns (they cover spatially local correlation, such as shapes), there is no need to have every pixels information being processed by one neuron. Actually in most cases results would be even better, if the pixels information is pre-processed, for instance by edge detection filters but a fully connected layer of neurons is not an optimal solution for this task.
Convolutional neural networks (CNNs) are inspired by biology, instead of connecting every pixels information directly with a neuron to process its information it filters the information in the first layers [2]. This procedure is similar to the on processes happening when an biological eye receives stimuli.

The receptive field [1] has a vast amount of photoreceptor cells [2] gathering information and converging the received information on to distinctly less retinal ganglion cells [3]. This process maps several features and reduces the input dimensionality as well as distinguishes the information to separate "channels" which are then transfered to the corresponding neurons to process features such as color, motion, shapes and so on separately [3].
The idea of CNNs is based on this biological processes, the information of an input image is convolved by several filters which try to extract interesting features in the first layer and in following layers this information is pooled and subsampled [2].
Convolution itself is the applying of a function repeatedly of the output of another function and in the context of CNNs it is applying different "filters" over an image to extract the already mentioned features. A convolution layer extracts the pixel information out of an image with kernels [4].

Example for a convolution layer: In figure 1 you can see an image of an animal on the left side (original image) and a kernel processed one on the right side. The used kernel matrix was:

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \tag{1}$$

So basically each pixels information in the right, processed image is the result of applying the kernel matrix (1) on the same pixel (and the neighboring pixels) in the left image.

5 Research and field of application

5.1 Image classification

More details / information about state of the art object/image detection/classification.

5.2 Video classification

More details / information about state of the art object/video detection/classification.

[1] http://en.wikipedia.org/wiki/Receptive_field
[2] http://en.wikipedia.org/wiki/Photoreceptor_cell
[3] http://en.wikipedia.org/wiki/Retinal_ganglion_cell
[4] http://en.wikipedia.org/wiki/Kernel_(image_processing)

Figure 1: Left side: original image, right side: edge-detection kernel processed image. Original image by Michael Plotke, 28th of January, 2013. Open creative commons license. http://upload.wikimedia.org/wikipedia/commons/5/50/Vd-Orig.png and http://upload.wikimedia.org/wikipedia/commons/6/6d/Vd-Edge3.png

5.3 Object tracking

6 Conclusion

Conclusion of the paper.

References

[1] D. Ciresan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3642 – 3649, June 2012.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[3] Norbert Kruger, Peter Janssen, Sinan Kalkan, Markus Lappe, Ales Leonardis, Justus Piater, Antonio Jose Rodriguez-Sanchez, and Laurenz Wiskott. Deep hierarchies in the primate visual cortex: What can we learn for computer vision? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8):1847–1871, 2013.

[4] Christian Szegedy, Alexander Toshev, and Dumitru Erhan. Deep neural networks for object detection. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2553–2561. Curran Associates, Inc., 2013.

[5] Naiyan Wang and Dit-Yan Yeung. Learning a deep compact image representation for visual tracking. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 809–817. Curran Associates, Inc., 2013.