Object tracking with deep neural networks

Santeri Salmijärvi

School of Electrical Engineering

Bachelor's thesis
Espoo Work in progress! Compiled: 14:17:59 2017/06/08

Thesis supervisor:

D.Sc. (Tech) Pekka Forsman

Thesis advisor:

M.Sc. (Tech) Mikko Vihlman

Aalto University
School of Electrical
Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
BACHELOR'S THESIS

Author: Santeri Salmijärvi

Title: Object tracking with deep neural networks

Date: Work in progress! Compiled: 14:17:59 2017/06/08 Language: English
Number of pages: 5+6

Degree programme: Bachelor's Program in Electrical Engineering

Supervisor: D.Sc. (Tech) Pekka Forsman

Advisor: M.Sc. (Tech) Mikko Vihlman

abstract in english

Keywords: keywords in english

KANDIDAATINTYÖN
TIIVISTELMÄ

Tekijä: Santeri Salmijärvi

Työn nimi: Kohteenseuranta syvillä neuroverkoilla

Päivämäärä: Work in progress! Compiled: 14:17:59 2017/06/08 Kieli: Englanti
Sivumäärä: 5+6

Koulutusohjelma: Sähkötekniikan kandidaattiohjelma

Vastuuopettaja: TkT Pekka Forsman

Työn ohjaaja: DI Mikko Vihlman

lyhyt tiivistelmä suomeksi

Avainsanat: avainsanat suomeksi

Contents

Abbreviations

DNN Deep neural network

MLP Multilayer perceptron

NN Neural network

ReLU Rectified linear unit

1 Introduction

Object tracking is a large and actively researched sub-area of computer vision. The main task for a tracker is to find and follow the desired subject in a sequence of images. The technology used for object tracking is closely related to other image analysis tasks and currently the majority of trackers are implemented as a neural network.

The field of image classification took a leap forward in 2012, when Krizhevsky et. al. presented record performance in the ImageNet-classification challenge using a convolutional network. Previous work had dismissed the network type as unfit for the task. [1] Since then, research has shifted to using convolutional networks as they have several clear advantages over other network types when used on picture analysis.

Comment by author:
go into benefits and/or give a source for the claim? maybe too specific for the introduction?

With the adoption of convolutional networks, much of the research revolves around deep neural networks. They consist of visible input and output layers with several so-called hidden layers in between them. The training of deep neural networks requires a large amount of training data and their development has been made easier by an increase in the size of applicaple datasets.

This thesis will present the architectures and principles currently used in deep neural networks tailored to object tracking tasks. The practices behind training and evaluating such networks are also introduced.

2 Background

2.1 Deep neural networks

A Deep neural network (DNN) is most commonly defined as a Neural network (NN), that has a **visible** input and output layer with several **hidden layers** between them. The distinction between visible and hidden layers is important because training only evaluates the output layer's performance. During training, a **learning algorithm** optimizes the individual hidden layers to best approximate the desired output of the whole network.

The input layer takes in the data to be processed, which typically means an array of color values in the case of object tracking. These values are then processed by the hidden layers and finally the output layer produces the target's position in the frame. These models usually come in the form of a **Feedforward neural network** or **Multilayer perceptron (MLP)**. The name comes from the fact that information flows from the input, through computations, to the output with no **feedback** connections.

Comment by author:
picture from eg. deeplearningbook page 174?

Each layer consist of several **units** with a weight and activation function. The weights of a layer are commonly represented by a matrix by which the input-vector is multiplied. Units in a layer also have the same activation function. Simply put, a unit's activation function is fed by a sum of it's weighted inputs and the result is output to the next layer alongside the other units' outputs. A commonly used unit type is the **Rectified linear unit (ReLU)**, which is defined by the activation function $g(z) = \max\{0, z\}$. It provides a nonlinear transformation while being comparable to linear models in terms of generalizing well and being easy to optimize.

Comment by author:
explain biases also

Before training, the weights of a MLP are initialized to small random values and biases to zero or small positive values. Then an algorithm called **stochastic gradient descent** is commonly applied alongside a training dataset. The basic procedure is to calculate the error of the netwok's output values compared to the desired ones using a **loss function**. It's gradient can then be calculated for example by **back-porpagation**, which feeds the errors back through the network to assign a contribution value to each unit. These values are then used to calculate the gradient of the loss function relative to the weights. Each weight is adjusted slightly to the opposite sign to minimize the loss function.

[2] Comment by author:
how to cite for the whole page?

2.2 Convolutional networks

Comment by author: what is a convolutional network, differences to a generic network
Comment by author: how do the layers function
Comment by author: benefits

3 Conclusions

conkluusiokappale

References

[1] Krizhevsky, A., I. Sutskever, and G. E. Hinton: *Imagenet classification with deep convolutional neural networks*. In *Advances in Neural Information Processing Systems*, volume 2, pages 1097–1105, 2012. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[2] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville: *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

Appendices

A Finnish summary - Suomenkielinen tiivistelmä

A Kohteenseuranta syvillä neuroverkoilla

Pitkä tiivistelmä suomeksi (3 sivua)