FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

Using Deep Convolutional Neural Networks for Object Recognition and Object Tracking

---

Bachelor-Thesis

submitted by

Thomas Engels

born in

Aachen

submitted at

14th September 2017

Supervisor:
Prof. Dr. Alexander Ferrein,
Prof. Dipl.-Inf. Ingrid Scholl

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfaßt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, dass alle Stellen der Arbeit, die wörtlich oder sinngemäß aus anderen Quellen übernommen wurden, als solche kenntlich gemacht sind und dass die Arbeit in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegt wurde.

Aachen, den 14. September 2017

Vorname Name

At first I wanted to open my thesis with an useful definition for *artificial intelligence*, but since the term is misleading and debatable I couldn't really define what *artificial intelligence* is. Then again I realized how we utilize the term *intelligence* for this matter:

The moment we fully understand and can describe an seemingly intelligent behavior by an artificial entity with an algorithm, we wouldn't call it intelligent any longer.

Contents

1 Introduction **1**
- 1.1 General Introduction to this Thesis' Topic . . . 2
- 1.2 Short Overview about the Motivation and Challenge of the Thesis . . . 3
- 1.3 Short Explanation of the Approach . . . 4
- 1.4 Overview about Each Chapter of the Thesis . . . 4

2 Related Works **7**
- 2.1 Which Related Works Exists in Known Literature? . . . 7
  - 2.1.1 Segmentation with the Help of Ground Truth Information . . . 8
  - 2.1.2 Other Related Works with Different Methods or Objectives . . . 9
  - 2.1.3 Further Related Works: . . . 12
- 2.2 Wherein Differs the Approach of the Thesis from State of the Art? . . . 13

3 Mathematical/Technical Background **17**
- 3.1 Used Models . . . 17
  - 3.1.1 Tensor . . . 17
  - 3.1.2 Artificial Neural Network . . . 18
  - 3.1.3 Convolution . . . 27
  - 3.1.4 Deconvolution . . . 35
- 3.2 Unwanted Effects in Neural Networks . . . 37
  - 3.2.1 Overfitting . . . 37
  - 3.2.2 Vanishing Gradient Problem . . . 40
  - 3.2.3 Exploding Gradient Problem . . . 41
- 3.3 Which Mathematical Methods are Known in Literature? . . . 43
  - 3.3.1 The Neuron . . . 43
  - 3.3.2 Forwardpropagation . . . 44
  - 3.3.3 Backpropagation . . . 46
  - 3.3.4 SGD . . . 50
  - 3.3.5 Convolution . . . 51

4 Approach **59**
  - 4.0.6 Prerequisites . . . 59
- 4.1 Description of the Own Approach . . . 60
  - 4.1.1 Object Detection . . . 60
  - 4.1.2 Object Localization . . . 64
- 4.2 Object Tracking . . . 69

5 Experimental Results **75**
5.1 Which Criteria for Evaluation of the Approach Exists? 75
5.2 Which Experiments were carried out? 75
5.3 Which Data were Gained? Are those Data and Experiments Significant? 77
5.4 Statistical Evaluation 80
5.5 Detail of Implementation 83

6 Further Analysis **89**
6.1 Future Works 89
6.2 Summary and Conclusion 90
Literature 93

List of Figures

1.1 [SLJ+14, Page 3, Fig. 1.] . . . 1

2.1 [Zha16, Page 6, Fig. 3.] . . . 10
2.2 Network flow . . . 10
2.3 Network flow with PG . . . 11
2.4 R-CNN caption boxes . . . 11
2.5 [chr17] . . . 12
2.6 [Page 3, Fig.1. YPW+07] . . . 12

3.1 Perceptron-Network . . . 19
3.2 illustration logistic regression . . . 19
3.3 Feed Forward Network with one Hidden Layer . . . 20
3.4 curved separation . . . 20
3.5 topological distortion of the domain space . . . 21
3.6 cutout of t-SNE visualizations of word embeddings. from [TRB10] . . . 22
3.8 recognized features captured at different layers, from [ZF13] . . . 24
3.10 transfer functions . . . 25
3.11 df/dx of transfer functions . . . 25
3.12 simple CNN, from [LKF10] . . . 27
3.13 one convolutional unit over a certain path of input pixels . . . 28
3.14 convolutional layer, from [Ola14a] . . . 28
3.16 processed dataset . . . 29
3.17 local optimum . . . 31
3.18 Example sentences generated by multimodal RNN [KF15, Page 7, Fig. 6.] 32
3.19 unlikely events for a Neural Network. from [GDDM13, Page 26, Fig. 6.] 32
3.21 model AlexNet [KSH12, Page 2, Fig. 2.] . . . 33
3.22 inception modules [SLJ+14, Page 5, Fig. 2.] . . . 35
3.24 overfitting . . . 38
3.25 vanishing gradient problem . . . 40
3.23 inference pattern (checker board artifact) . . . 42
3.26 a single neuron . . . 43
3.27 SGD illustration . . . 51
3.28 (3.28a source *CIFAR-10 dataset*) . . . 53
3.29 convolution on multiple batches . . . 55

4.1 residual network of the thesis approach . . . 60
4.2 Residual learning: a building block, from [HZRS15] . . . 61
4.3 ResNet, from [HZRS15] . . . 62
4.5 ResNet training curves (18 and 34 layers), from [HZRS15, Page 5, Fig. 4.] 63
4.6 image decompression pattern through deconvolution . . . 66
4.7 convolutional network with appended deconvolutional network, . . . 67

4.9 object localization with option 1 67
4.11 object localization with option 2 68
4.13 object localization with option 3 69
4.14 processing of the image “pyramid” 70
4.15 inference signal filter 71
4.16 image fragmentation 72

5.1 object localization with option 2 78
5.3 Cat or no-Cat Net 79
5.5 Cat or no-Cat Net zoomed 80
5.7 original cat image zoomed for comparison 80

Abbreviations

AI Artificial intelligence

AWS Amazon Web Services 59

CIFAR Canadian Institute for Advanced Research 59

CMOS Sensor active-pixel sensor using CMOS technology 90

CMOS Complementary Metal-Oxide-Semiconductor viii

CNN Convolutional Neural Network 27

CPU Central Processing Unit

DSSD Deconvolutional Single Shot Detector 12

FLOPS Floating Point Operations Per Second 62

GPU Graphics Processing Unit

HSV Hue Saturation Value (image color space)

ILSVRC ImageNet Large Scale Visual Recognition Challenge 33

JcF Jacobian of a Composition 48

Q Learning Quality Learning 13

R-CNN Region Based CNN 11

RAM Random-Access Memory

ReLU Rectified Linear Unit 26

ResNet Residual Network 35

RGB Red Green Blue (image color space)

RNN Recurrent Neural Network 21

ROI Region of Interest 79

SGD Stochastic Gradient Descent 22

SIFT Scale-Invariant Feature Transform 90

SSD Single Shot Detector 12

VGG Net a Neural Network by the VGG team 34

VResNet Vanilla Residual Network (a basic version of a deep ResNet) 1

ZF Net Zeiler Fergus Net (personal name) 8