Hochschule
München
University of
Applied Sciences

Bachelor Thesis
in Information Systems and Management

Label Extraction from Image via Deep Learning

Johannes Reichle
Matriculation no. 04797218

Supervisor	Prof. Dr. Rainer Schmidt
Date of Submission	XX.XX.2022

Declaration

I hereby certify that I have written Bachelor Thesis on my own and that I have not used any sources or aids other than those indicated.

Munich, the XX.XX.2022

. . . .
Johannes Reichle

Abstract

Here abstract for Bachelor Thesis.

Contents

Chapter 1

Introduction

1.1 Motivation

Nowadays it is hard to find a business process that doesn't use software for improvement. Various technologies come to be valued because of this. A recent trend is to use Deep Learning for types of problems that range from self driving cars to medical diagnosis [BRSS19]. Deep Learning is a powerful technology based on Artificial Neural Networks where data is processed in multiple layers to extract features and solve a given problem [SM19]. One area where this is especially helpful is the field of Computer Vision. Deep Learning for Computer Vision has only caught on in the recent years as the big computational cost has been met by the improvement in computer hardware [PRN+17]. Computer vision deals with extracting information from photos. This includes tasks like recognizing faces or reading text [Pri12]. Applying Deep Learning to extract equipment labels from photos fits right into this crease of applying technology for making daily problems more efficient. Combining the two fields to create value and learning about the underlying theoretical foundations and inner workings is the motivating factor of this work.

1.2 Problem description

Motivated by the wide success of Deep Learning concerning Computer Vision, the objective of this work is to implement and train a Deep Learning model that can extract equipment names from photos taken of name plates.

When determining whether automisation is an improvement four aspects have to be examined. These are time, costs, quality and flexibility. The aspects build a quadrangle that is based on the optimizing trade-off between the factors [DLRMR13].

Without software supporting the task of reading the name of the picture and typing it into the system, can take long seconds, whereas a trained Deep Learning model could complete the task in a mere instant. Therefor automisation via Deep Learning should improve the efficiency of the process when compared to manually reading and typing the information off the image.

Training costs for a Deep Learning model are very high due to the computing intensive backpropagation algorithm that tunes the network to the data. But the usage cost is low. For manual labor the opposite is the case as training a person to type in a label is done quickly and labor costs are high in comparison to the expenses for running the model.

Both Deep Learning models and human labor are not 100% accurate. It is human to make mistakes and because Deep Learning is trained only trained on a specific set of data it makes sense that not all predictions can be correct as there can always be outliers in the data. The question is whether the model can be as accurate or even better than its human counterpart. This is especially interesting when it is applied in the real world where it might have to do good in subpar situations. An example is bad image quality.

Flexibility is concerned with how well a process can adjust to changing requirements. A set of new equipment names that have to be included can pose a problem to a Deep Learning model because it is not trained for the new data. A human on the other hand should not have any problems in this regard.

The main concern for the solution's efficacy is whether it is accurate enough. Therefor this work focuses on this aspect in particular.

1.3 Methodology

The goal of this work is to implement and train a Deep Learning model to read in labels from photos. The emerging artifact can be used to solve the problem detailed in 1.2. The expository instantiation is helpful to gain more understanding the artifact as it is common in design science. In particular this is justificatory knowledge on the design on the Deep Learning model and Machine Learning way of approaching problems. This is important in order to apply it and to optimize existing research to the specific problem.

The methodology is based on action research [JP21]. It constists of a cycle of five phases: Diagnosis, Planning, Intervention, Evaluation, Reflection. The first cycle will entail an exploratory data analysis which corresponds to the Diagnosis part. Here it is important to recognize main characteristics of the images and to find outliers and other potential problems [Cox17]. The research is then extended to existing practical solutions for similar practical problems as well as proposed architectures from academic research. Theoretical knowledge about the models as well as practical information about results for similar problems contribute to the discussion about which approach is the most promissing. Combining architectures is also a viable possibility to solve the given problem. This concludes the Planning phase and will lead to a model exaptation that evolves to be the artifact at the center of this thesis. The next step is implementing and training the chosen approach which. Evaluation for of the current model follows. Storing and analyzing results of training and cross validation as well as visualizing the training progress is an important part of this. In the Reflection stage it is decided whether a new cycle should be carried out.

From the second cycle on the first three phases change as there already is a model that is to be improved. This time the Diagnosis phase entails asking questions about the existing model: What worked? Why did it work/not work? What needs to change? Changes are planned and implemented accordingly. The Evaluation and Reflection phases are not changing in the second cycle thus closing the loop. The incremental adjustments to the model are made in order to improve the accuracy. This includes possibly adjusting the architecture, hyperparameter tuning and preprocessing approaches like image compression.

1.4 Expected results and outlook

The research into the theoretical foundation of Deep Learning and into possible approaches leads to a strong understanding of the underlying technology. This is helpful to produce a comparison of approaches that is based on theoretical as well as practical knowledge. The goal is to find out which approach work best for the chosen practical problem and why that is the case. Implementation and training of the most promissing one is yielding the artifact this work revolves around. The process of optimization not only improves the solution to the problem (see 1.2) but is also used to learn more about the implemented approach.

Integrating the artifact into the business process is not an issue that is discussed in this work. Nor will model feedback and over time iteration be part.

The intendet structure of the thesis with dependencies between chapters can be found in figure 1.1.

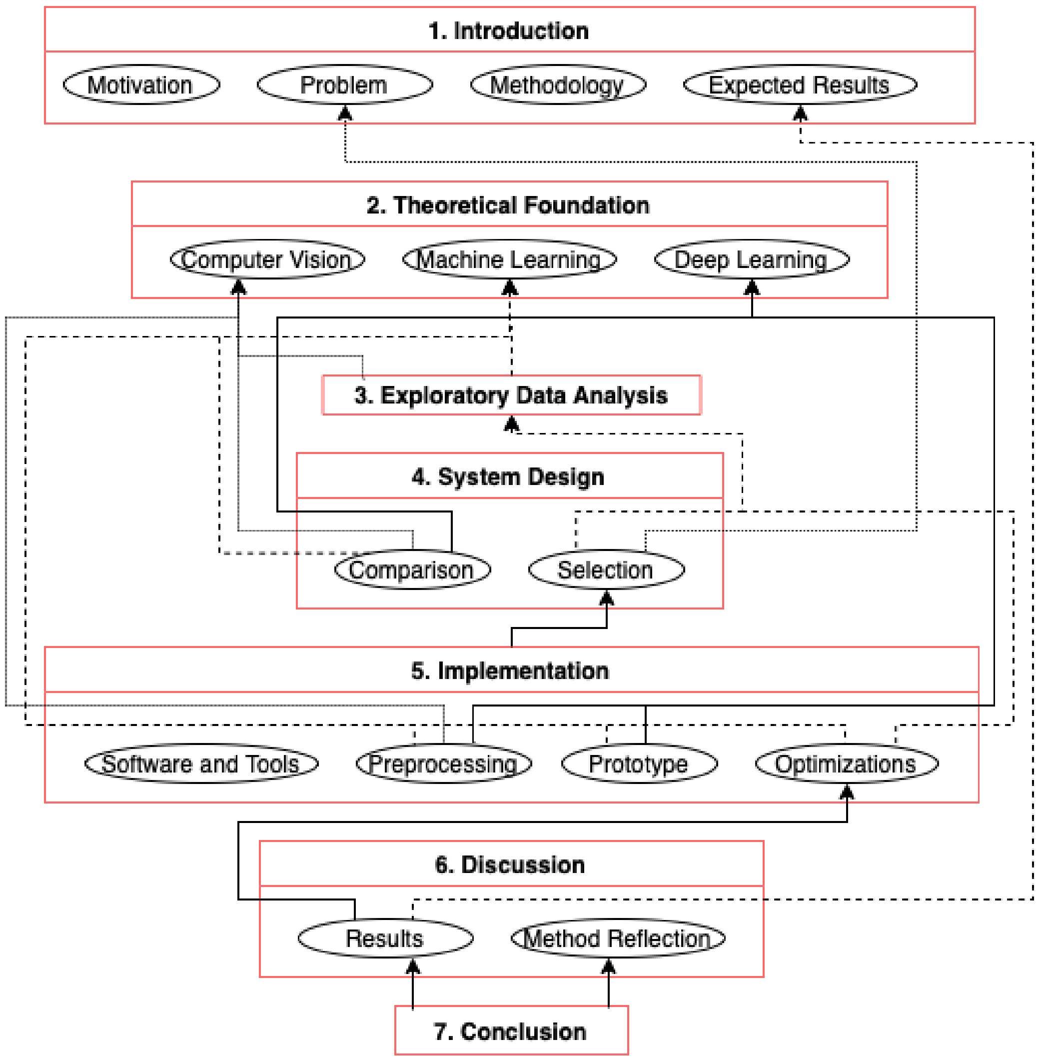

Figure 1.1: Chapters with subchapters and dependencies

Chapter 2

Theoretical Foundation

2.1 Opical Character Recognition

Deep Learning based OCR [ZJG+20] What is OCR: process of converting images of typed, handwritten or printed text into machine-encoded one includes two sub frameworks: text detection and text recognition

2.1.1 Text detection

Detect position coordinates containing text in input image Text detection more challenging

Two object detection methods

- R-CNN
  views detection problem as classification problem
  CNN to extract deep features of proposals by selective search
  Use SVM to classify with features
- YOLO extract feature maps on entire image
  directly regress bounding boxes on feature maps

YOLO generally slower but more accurate

Deep Learning in Character Recognition Considering Pattern Invariance Constraints [OOK15]

2.1.2 Character recognition

Recognize text based on position coordinates

2.2 Machine Learning

1. Supervised — Unsupervised
2. Loss Function
3. Optimization techniques: Stochastic-Batch Gradient Descent, GD Momentum, Adam
4. Errors metrics
5. Bias-Variance tradeoff (including Regularization)

2.3 Deep Learning

1. ANN / MLP
2. CNN
3. RNN

Chapter 3

Exploratory Data Analysis

Chapter 4

System Design

Search for specific information

4.1 Approach comparison

4.1.1 Approach Research

GitHub implementation

Two models that can be used in conjunction
detection [Beo21b]
uses RetinaNet structure [LGG+18] applies techniques from textboxes++ [LSB18]
character recognition [Beo21a]
needs cropped text area as input
uses CRNN [SBY15] → end-to-end learning, LSTM fir arbitrary length of input and output, no need to apply detection and cropping to each single character

Tesseract

Open Source OCR engine [Smi07]

- uses Deep Learning (found c++ code for layers in repo)
- Processing in step-by-step pipeline, some unusual stages
  1. Line and Word finding
  1.1. Line finding
  1.2. Baseline Fitting
  1.3. Fixed Pitch Detection and Chopping
  1.4. Proportional Word Finding
  2. Word Recognition

2.1 Chopping Joined Characters
2.2 Accociating Broken Characters
3. Static Character Classifier
3.1 Features
3.2 Classification
3.3 Training Data
4. Linguistic Analysis
5. Adaptive Classifier

EAST

An Efficient and Accurate Scene Text Detector

4.1.2 Comparison

4.2 Approach selection

Chapter 5

Implementation

5.1 Software and Tools

5.2 Preprocessing

5.3 Prototype

5.4 Optimizations

Chapter 6

Discussion

6.1 Results

6.2 Method reflection

Chapter 7

Conclusion

Appendix A

References

List of Figures

List of Tables

Bibliography

[Beo21a] Beom. *CRNN (CNN+RNN)*, September 2021. original-date: 2018-01-14T07:52:25Z.

[Beo21b] Beom. *Text Detector for OCR*, August 2021. original-date: 2019-03-12T05:11:06Z.

[BRSS19] Valentina Emilia Balas, Sanjiban Sekhar Roy, Dharmendra Sharma, and Pijush Samui, editors. *Handbook of Deep Learning Applications*, volume 136 of *Smart Innovation, Systems and Technologies*. Springer International Publishing, Cham, 2019.

[Cox17] Victoria Cox. *Translating Statistics to Make Decisions: A Guide for the Non-Statistician*. Apress, Berkeley, CA, 2017.

[DLRMR13] Marlon Dumas, Marcello La Rosa, Jan Mendling, and Hajo A. Reijers. *Fundamentals of Business Process Management*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[JP21] Paul Johannesson and Erik Perjons. *An Introduction to Design Science*. Springer International Publishing, Cham, 2021.

[LGG+18] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. *Focal Loss for Dense Object Detection*. arXiv:1708.02002 [cs], February 2018. arXiv: 1708.02002.

[LSB18] Minghui Liao, Baoguang Shi, and Xiang Bai. *TextBoxes++: A Single-Shot Oriented Scene Text Detector*. IEEE Transactions on Image Processing, 27(8):3676–3690, August 2018. arXiv: 1801.02765.

[OOK15] Oyebade K. Oyedotun, Ebenezer O. Olaniyi, and Adnan Khashman. *Deep Learning in Character Recognition Considering Pattern Invariance Constraints*. International Journal of Intelligent Systems and Applications, 7(7):1–10, June 2015.

[Pri12] Simon J. D. Prince. *Computer Vision: Models, Learning, and Inference*. Cambridge University Press, June 2012. Google-Books-ID: PmrICLzHutgC.

[PRN+17] Moacir Antonelli Ponti, Leonardo Sampaio Ferraz Ribeiro, Tiago Santana Nazare, Tu Bui, and John Collomosse. *Everything You Wanted to Know about Deep Learning for Computer Vision but Were Afraid to Ask*. In 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), pages 17–41, October 2017. ISSN: 2474-0705.

[SBY15] Baoguang Shi, Xiang Bai, and Cong Yao. *An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition*. arXiv:1507.05717 [cs], July 2015. arXiv: 1507.05717.

[SM19] Ajay Shrestha and Ausif Mahmood. *Review of Deep Learning Algorithms and Architectures*. IEEE Access, 7:53040–53065, 2019. Conference Name: IEEE Access.

[Smi07] R. Smith. *An Overview of the Tesseract OCR Engine*. In Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), volume 2, pages 629–633, September 2007. ISSN: 2379-2140.

[ZJG+20] Zhenyao Zhao, Min Jiang, Shihui Guo, Zhenzhong Wang, Fei Chao, and Kay Chen Tan. *Improving Deep Learning based Optical Character Recognition via Neural Architecture Search*. In 2020 IEEE Congress on Evolutionary Computation (CEC), pages 1–7, July 2020.

Appendix B

Code

code here