IMPROVED OPTICAL CHARACTER RECOGNITION WITH DEEP LEARNING

TAN CHIANG WEI

UNIVERSITI TEKNOLOGI MALAYSIA
IMPROVED OPTICAL CHARACTER RECOGNITION WITH DEEP LEARNING

TAN CHIANG WEI

A project report submitted in fulfilment of the requirements for the award of the degree of
Master of Engineering

Faculty of Electrical Engineering
Universiti Teknologi Malaysia

JANUARY 2018
I would like to dedicate my project report to my beloved family
ACKNOWLEDGEMENT

Throughout the reporting process, I got countless help from lecturers, friends and a lot of people for their support. Without their contribution, this report will never be completed.

My first man to be grateful is my project supervisor, Dr. Usman Ullah Sheikh. He gave me a lot of support, guidance and criticism to ensure that my project can be completed in time. Without him, it will be difficult to complete the report. I am also would like to take this opportunity to thank Universiti Teknologi Malaysia (UTM) for all the lecture and useful knowledge throughout the 2 year course of Master Degree in this university.

I would also want to thank Intel Microelectronic Sdn. Bhd. for sponsoring my school fees in order to study in UTM. Among my colleagues, I would like to thank my manager Teh Eng Keong and James Lim, as well as my mentors, Lim Mui Liang and Lim Wen Chin, for covering my work while I’m doing my research and write this report. Finally, I want to sincerely thank my family, especially my parents, for giving me unlimited support and encouragement.
ABSTRACT

Optical Character Recognition (OCR) plays an important role in the retrieval of information from pixel-based images to searchable and machine-editable text formats. For instance, OCR is typically used in many computer vision applications such as in automatic signboard recognition, language translation as well as in the process of digitizing scanned documents. However, compared to old documents or poorly printed documents, printed characters are typically broken and blurred, which makes the character recognition in potentially far more complicated. Although there are several OCR applications which utilizes techniques such as feature extraction and template matching for recognition, these methods are still not accurate enough for recognition. In this work, deep learning network (transfer learning with Inception V3 model) is used to train and perform OCR. Deep learning network is implemented and trained using Tensorflow Python API that supports Python 3.5+ (GPU version) which is available under the Apache 2.0 open source license. The Inception V3 network is trained with 53,342 character images consisting of noises which are collected from receipts and newspapers. From the experiment results, the system achieved significantly better recognition accuracy on poor quality of text character level and resulted in an overall 21.5% reduction in error rate as compared to existing OCRs. Besides, there is another experiment conducted to further analyze the root causes of text recognition failure and a solution to overcome the problem is also proposed. Analysis and discussion were also made on how the different layer’s properties of neural network affects the OCR’s performance and training time. The proposed deep learning based OCR has shown better accuracy than conventional methods of OCR and has the potential to overcome recognition issue on poor quality of text character.
ABSTRAK

# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DECLARATION</td>
<td>ii</td>
</tr>
<tr>
<td></td>
<td>DEDICATION</td>
<td>iii</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGEMENT</td>
<td>iv</td>
</tr>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td></td>
<td>ABSTRAK</td>
<td>vi</td>
</tr>
<tr>
<td></td>
<td>TABLE OF CONTENTS</td>
<td>vii</td>
</tr>
<tr>
<td></td>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td></td>
<td>LIST OF ABBREVIATIONS</td>
<td>xiii</td>
</tr>
<tr>
<td></td>
<td>LIST OF APPENDICES</td>
<td>xiv</td>
</tr>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Project Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Challenges</td>
<td>2</td>
</tr>
<tr>
<td>1.3</td>
<td>Objectives and Scope</td>
<td>3</td>
</tr>
<tr>
<td>1.4</td>
<td>Project Report Organization</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>LITERATURE REVIEW</td>
<td>5</td>
</tr>
<tr>
<td>2.1</td>
<td>Background of Deep Learning</td>
<td>5</td>
</tr>
<tr>
<td>2.2</td>
<td>Convolutional Neural Network’s Structure</td>
<td>7</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Convolutional Layer</td>
<td>10</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Pooling/Subsampling Layer</td>
<td>11</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Non-linear Functions</td>
<td>12</td>
</tr>
<tr>
<td>2.2.3.1</td>
<td>Continuous Trigger</td>
<td>13</td>
</tr>
<tr>
<td>2.2.3.2</td>
<td>Rectified Linear Unit (ReLu)</td>
<td>14</td>
</tr>
</tbody>
</table>
2.2.4 Fully connected layers 14

2.3 Transfer Learning 15

2.3.1 Categorization of Transfer Learning Techniques 16

2.4 Related Work of OCR 19

2.4.1 OCR with Clustering Method 19

2.4.2 OCR with Support Vector Machine Classifier 20

2.4.3 OCR with Deep Learning 21

2.4.4 OCR with Transfer Learning 22

2.4.5 The Impact of Training Data 24

2.5 Chapter Summary 24

3 METHODOLOGY 25

3.1 Design Model 25

3.2 Image Processing and Segmentation 27

3.3 Labeling 29

3.4 Training Sets Collection 30

3.5 Testing Sets Distribution 34

3.6 Helper Function for Showing Result 35

3.6.1 Plot Example Error 36

3.6.2 Plot Confusion Matrix 36

3.7 Chapter Summary 37

4 EXPERIMENT RESULTS 38

4.1 First Network Model Selection 38

4.2 Second Network’s Implementation 39

4.2.1 Number of Fully Connected Layer Test 40

4.2.2 Training Iteration Test 42

4.2.3 Batch Size Test 43

4.2.4 Activation Function Test 44

4.3 Benchmark with Existing OCR 45

4.4 Text Character’s Noise Pattern Analysis 47
5 CONCLUSION AND FUTURE WORK

5.1 Research Outcomes

5.2 Future Works

REFERENCES

Appendix A
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Comparison between traditional machine learning and transfer learning [1]</td>
<td>17</td>
</tr>
<tr>
<td>2.2</td>
<td>Different types of transfer learning [1]</td>
<td>18</td>
</tr>
<tr>
<td>2.3</td>
<td>Related works on OCR using clustering method [2, 3, 4]</td>
<td>20</td>
</tr>
<tr>
<td>2.4</td>
<td>Related works about OCR using neural network method [5, 6]</td>
<td>21</td>
</tr>
<tr>
<td>2.5</td>
<td>Related works about OCR using deep neural network method [7, 8, 9]</td>
<td>22</td>
</tr>
<tr>
<td>2.6</td>
<td>Related works about OCR using transfer learning method [10, 11]</td>
<td>23</td>
</tr>
<tr>
<td>3.1</td>
<td>Total training samples for 94 characters</td>
<td>30</td>
</tr>
<tr>
<td>3.2</td>
<td>Sample of confusion matrix table</td>
<td>37</td>
</tr>
<tr>
<td>4.1</td>
<td>First Network Selection Table</td>
<td>38</td>
</tr>
<tr>
<td>4.2</td>
<td>Iteration Test Result</td>
<td>42</td>
</tr>
<tr>
<td>4.3</td>
<td>Drop out test table</td>
<td>44</td>
</tr>
<tr>
<td>4.4</td>
<td>Table Benchmark of a9t9 OCR and proposed OCR</td>
<td>45</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

FIGURE NO. | TITLE | PAGE
---|---|---
2.1 | Deep neural network diagram [12] | 5
2.3 | Training of neural networks [14] | 7
2.4 | Illustration of a human’s neuron and machine neuron’s model [14] | 8
2.5 | Typical block diagram of a CNN [14] | 9
2.6 | Pictorial representation of convolution process [14] | 11
2.7 | LeNet-5 network diagram and the learned features [15] | 11
2.8 | Pictorial representation of average and max pooling [14] | 12
2.9 | non-linear function examples [14] | 13
2.10 | Pictorial representation of ReLU [14] | 14
2.11 | Processing of a fully connected layer [14] | 15
2.12 | Transfer learning comparison diagram [1] | 16
2.13 | An overview of different types of transfer [1] | 18
2.14 | Related works about the impact of training data [16] | 24
3.1 | Proposed model of OCR system | 25
3.2 | Transfer-learning based on Inception V3 Model Diagram | 26
3.3 | Example of image processing | 28
3.4 | Example of segmentation results | 28
3.5 | Dataset arrangement by labeling | 29
3.6 | Testing samples distribution diagram | 34
3.7 | Screenshot of Tensorflow cmd displaying result | 35
3.8 | Plot example errors | 36
4.1 | Second network block diagram | 40
4.2 | One layer test | 40
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3</td>
<td>Two layer test</td>
<td>41</td>
</tr>
<tr>
<td>4.4</td>
<td>Three layer test</td>
<td>41</td>
</tr>
<tr>
<td>4.5</td>
<td>Accuracy vs Iteration graph</td>
<td>42</td>
</tr>
<tr>
<td>4.6</td>
<td>Batch size vs accuracy and time usage graph</td>
<td>43</td>
</tr>
<tr>
<td>4.7</td>
<td>Activation function vs accuracy graph</td>
<td>45</td>
</tr>
<tr>
<td>4.8</td>
<td>Test Sample 1</td>
<td>46</td>
</tr>
<tr>
<td>4.9</td>
<td>Test Sample 2</td>
<td>46</td>
</tr>
<tr>
<td>4.10</td>
<td>Test Sample 3</td>
<td>46</td>
</tr>
<tr>
<td>4.11</td>
<td>Test Sample 4</td>
<td>46</td>
</tr>
<tr>
<td>4.12</td>
<td>Noise patterns of hand-crafted data</td>
<td>47</td>
</tr>
<tr>
<td>4.13</td>
<td>Text character’s noise pattern analysis graph</td>
<td>47</td>
</tr>
<tr>
<td>4.14</td>
<td>Transfer value of images</td>
<td>48</td>
</tr>
</tbody>
</table>
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDR</td>
<td>Correct Detection Rates</td>
</tr>
<tr>
<td>CNN</td>
<td>Convolutional Neural Network</td>
</tr>
<tr>
<td>HOG</td>
<td>Histogram of Oriented Gradients</td>
</tr>
<tr>
<td>MSER</td>
<td>Maximally Stable Extremal Regions</td>
</tr>
<tr>
<td>OCR</td>
<td>Optical Character Recognition</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
</tbody>
</table>
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Segmentation Scripts</td>
<td>57</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Project Background

The ability to perform human functions such as reading machines is an ancient dream. However, over the last few years, reading by a machine is no longer a dream and has grown to become a truth. Text character recognition commonly deals with the recognition of optically processed characters which is also called as optical character recognition (OCR). The basic idea of OCR is to convert any hand written or printed text into data files that are able to be edited and read by machine. With OCR, any article or book can be scanned directly and the editable text format can then be easily converted from a computer. The OCR system has two major advantages which are the ability to increase productivity by reducing staff involvement and storing text efficiently. More generally, the areas where this system can be applied are postal departments, banks, publication industry, government agencies, education, finance, health care [7].

The universal OCR system consists of three main steps which are image acquisition and preprocessing, feature extraction and classification [7]. Image preprocessing phase cleans up and enhances the image by noise removal, correction, binarization, dilation, color adjustment and text segmentation etc. Feature extraction is a technique for extracting and capturing certain pieces of information from data. In the classification phase, the portion of the divided text in the document image will be mapped to the equivalent textual representation.

Nowadays, there are several existing OCR solutions which are commonly used in machine learning research and pattern recognition. Unfortunately, there is
still a challenging problem for recognizing broken or faded English characters. The performance of OCR directly depends on the quality of input image or document, thus making the character recognition in scene images is potentially far more complicated. In addition, English characters with poor quality are typically obtained from old printed documents that are usually caused by damaged print cartridges. Unfortunately, these training samples are yet to be found in the existing solution. In order to recognize poor quality English characters, an improved OCR with sufficient training data distribution is needed.

In traditional machine learning research, many people think that the feature vectors of test and training data are provided from the same source. However, this may not be truth in some of the OCR research cases. In the concept of transfer learning, training samples can be used to pre-train a network in the source domain, and these well-trained learning characteristics can be delivered and benefit from the training process in the target domain of the second network. In recent years, traditional methods in the field of OCR research have been almost substituted by deep learning methods such as Convolutional Neural Networks (CNN). An idea is proposed by Oquab et al. that is using the CNN to learn image representations on a large annotation dataset can adequately transfer this information to other visual recognition tasks with a limited amount of training data [17]. Yejun Tang et al. proposed another idea is to add an adaptation layer in CNN using transfer learning, which achieves performance improvement in historical Chinese character recognition tasks [10]. Inspired by these works, the proposed method in this project is going to apply a deep neural network with transfer learning method for broken English character recognition problems.

1.2 Challenges

From previous section, the problem statement is simply explained. The existing conventional OCR with machine learning is trained based on hand-written text and good quality printed text. There is still a challenge for poor quality (broken, blurred and incomplete) English text character recognition. In addition, due to insufficient labeled training samples of poor quality English character, neural network used for
OCR will suffer from imbalanced training data distribution issue. However, the data-labeling process requires new train data which is very costly to train up a new network to recognize poor quality text characters. The process will also consume a huge amount of training time as well.

Furthermore, there is also another challenge where the performance of deep neural network will potentially be affected by the new training data distributions. For example, a neural network is pre-trained to recognize good quality of "O" character, and then if the network trained again with different "broken" pattern of poor quality of "O" character, the weights adjusted in the network will actually negatively be affected by the new training data. Philippe Henniges et al. explained that if training with over-represented class distributions, this will cause the performance of neural network to degrade [16]. From the challenges stated above, the classification and training data distribution is the most crucial stage and a challenge in this project. The aim of this work is to improve an OCR method with deep learning network that will apply transfer learning concept and achieve the high accuracy performance while keeping the training time short.

1.3 Objectives and Scope

The first objective is to collect training materials with a set of blur, incomplete English text characters in images. Next is to develop an OCR method by using deep learning neural network approach. Moreover, investigate the method that will achieve high accuracy while reducing the training time. Lastly to benchmark the performance of the proposed OCR with existing OCR methods.

Scope of this project is mainly focused on the classification part such as network structure adjustment and training data distribution. However, some existing solutions such as OpenCV will be used and applied for the image processing part (segmentation and filtering). While going through preprocessing, each image will segment out the text character instead of word, so as similar to the input for classification.
Besides that, dataset will be prepared in png or jpg format. The text language is English only, where the font size and type of the text dataset are typically depend on the data-set’s resources. In addition, the dataset will be collected from old newspapers and receipts. Tensorflow will be used as the framework in this project on Windows with NVIDIA GPU using Python and C++ programming language.

1.4 Project Report Organization

Chapter 2 presents the literature review on Convolutional Neural Network, Transfer Learning and related works. In Chapter 3, the design model, methodology, training data collection, testing data distribution plan, and fine-tuning methods are explained. In Chapter 4, experiments results are presented with a summary of the network behavior and benchmark of this work. In Chapter 5, this work is concluded with recommendations for future works.
REFERENCES


39. Al-Mubaid, H. and Umair, S. A. A new text categorization technique using
distributional clustering and learning logic. *IEEE Transactions on Knowledge and Data Engineering*, 2006. 18(9): 1156–1165.


52. SPACE, O. Free OCR API and Online OCR. URL https://ocr.space/.