UNIVERSITY OF TARTU
Institute of Computer Science
Software Engineering Curriculum

Jaan Tohver

Optical Character Recognition for Extremely Low Quality Images

Master's Thesis (30 ECTS)

Supervisor: Gholamreza Anbarjafari, PhD

Tartu 2018

Optical Character Recognition for Extremely Low Quality Images

Abstract:

Optical character recognition (OCR) from printed and handwritten documents is a solved problem for all practical purposes. Modern OCR systems are able to achieve a 99.9% detection rate which is on par with human capabilities. Where most OCR systems fall short is when the input is of low quality, such as containing large amounts of noise or motion blur, or being of low resolution.

A real world use-case of this problem is detecting car licence plates from a security camera feed. Security cameras are often of low resolution, use high levels of compression, and have low framerate since the video footage needs to be stored for a long period of time and storage cost is paramount. In addition, cars can move unpredictably causing motion blur during the capture of a video frame.

The goal of this project is to test various methods of improving OCR accuracy with low quality input. Namely,

- Training a neural network (NN) on low quality images and testing results on low quality images.
- Training a NN on high quality images and testing on digitally enhanced low quality images.
- Training a NN on digitally enhanced low quality images and testing on digitally enhanced images.
- Using multiframe registration of frames from a video feed to improve detection quality.

The tests will be conducted using frames from a blurry, noisy, and low resolution video feed containing text. Third party solutions are used to extract areas containing text from those frames.

Some of the training and test data is gathered specifically for this task by filming, and extracting frames from an intentionally blurry video. Additional synthetic motion blur is added to some of the images. Some data is gathered from available open source databases, such as images containing vehicle licence plates.

Keywords:

OCR, Deblurring, Low Quality, Low Resolution

CERCS:

CERCS code and name: https://www.etis.ee/Portal/Classifiers/Details/d3717f7b-bec8-4cd9-8ea4-c89cd56ca46e

Contents

Unsolved issues

1 Introduction

Optical character recognition (OCR) is the conversion of text to a machine encoded form, such as converting images or printed documents to text documents. OCR is not a new concept by any means. A patent for a system which can be called the predecessor of modern OCR was already granted in 1931 [1]. Modern OCR systems are, of course, far more powerful.

1.1 Motivation

For all intents and purposes OCR is basically a solved problem for detecting text from clear and high resolution images containing printed or handwritten text. Modern systems perform on par with human capabilities in this regard, boasting a 99.9% detection rate. A problem still being researched, however, is OCR from low quality images. Low quality can mean low resolution, high levels of compression, containing artefacts generated by scanning or image capture, containing motion blur caused by the movement of the subject or the camera during image capture, or any combination thereof.

Accurate OCR from low quality images and video could solve several problems for automating processes where the system needs to detect some text. For example in an autonomous parking garage, where a camera is filming approaching cars, a system analyses the video feed and opens the gate or barrier for cars with licence plate numbers matching authorised users.

1.2 Contributions

The goal of this thesis is to test different approaches for such a system. Namely,

- Training a neural network (NN) on low quality images and testing results on low quality images.
- Training a NN on high quality images and testing on digitally enhanced low quality images.
- Training a NN on digitally enhanced low quality images and testing on digitally enhanced images.
- Using multiframe registration of frames from a video feed to improve detection quality.

Neural networks behave in very specific ways in regards to the data that was used for training them. Understanding how image enhancement and OCR networks behave under such conditions gives valuable insight into which training data to use to improve OCR accuracy for low quality images while also retaining accuracy for high resolution and high quality images.

Add accurate contribution after work is done.

1.3 Outline

The thesis is structured as follows.
Chapter 2 describes both the history and the state of the art in OCR and low quality image enhancement and discusses the advantages and drawbacks of each of the approaches.
Chapter 3 describes the research problem and the necessity of finding a solution to the problem.
Chapter 4 describes in detail the approach taken to solve the problem.
Chapter 5 describes the evaluation process of the results as well as a comparison to related work.
Chapter 5 concludes the thesis with a summary and research directions for the future

2 Background

This section gives an overview of the history of OCR and image enhancement solutions as well as research and commercial products related to the topic.

2.1 History

2.1.1 OCR

The history of OCR can be traced back to the late 1920s when the inventor Emanuel Goldberg started developing a system called "Statistical Machine" which searched microfilm archives using an optical code recognition system. He was granted a patent for his invention in 1931, which was later acquired by IBM [1].

2.1.2 Digital Image Enhancement

Digital image enhancement has been around for as long as digital images have i.e., since the 1950s [2]. In essence a very simple concept - take the pixel values of an image and, using some algorithm, modify them to produce a clearer image. It was advanced further alongside astronomical and medical research [3].

2.2 State of the Art

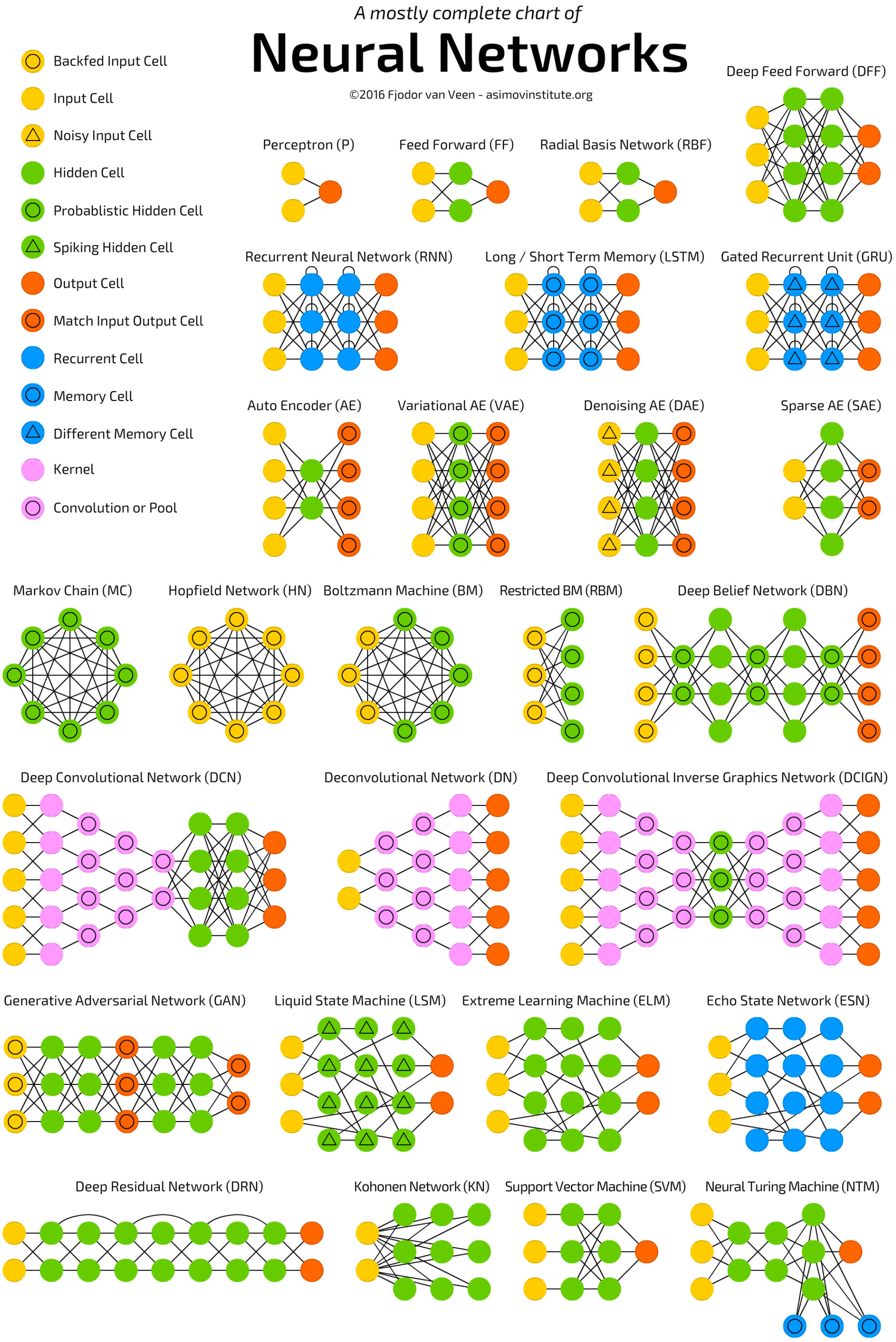
2.2.1 Neural Networks

Most modern OCR systems and image enhancement techniques are based on artificial neural networks, often called just neural networks (NNs). Which makes this a necessity is the fact that both of those problems are very dynamic and not well solvable algorithmically.

The theoretical base for NNs dates back to the end of the 1800s and is based on the structure of the human brain in which neurons interact with each other and the bonds between neurons can strengthen and weaken over time [4].

An artificial neural network consists of many simple connected processors, called artificial neurons, which produce activations based on an activation function producing real numbers in a predefined range. The simplest neural network consists of an input layer and an output layer. The neurons in the input layer produce activations based on the input. The output layer produces an output based on the activations of input layers [4].

Most NNs are not as simple as that and contain any number of hidden layers between the input and output layers. These hidden layers consist of any number of neurons which produce activations based on the output of another layer. There are various architectures used for NNs, some are combinations of other types of networks. This work mostly focuses on convolutional neural networks (CNNs) and recurrent neural networks (RNNs) [4].


A mostly complete chart of
Neural Networks
©2016 Fjodor van Veen - asimovinstitute.org
Backfed Input Cell
Input Cell
Noisy Input Cell
Hidden Cell
Probablistic Hidden Cell
Spiking Hidden Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Different Memory Cell
Kernel
Convolution or Pool
Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Deep Feed Forward (DFF)
Recurrent Neural Network (RNN)
Long / Short Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)
Markov Chain (MC)
Hopfield Network (HN)
Boltzmann Machine (BM)
Restricted BM (RBM)
Deep Belief Network (DBN)
Deep Convolutional Network (DCN)
Deconvolutional Network (DN)
Deep Convolutional Inverse Graphics Network (DCIGN)
Generative Adversarial Network (GAN)
Liquid State Machine (LSM)
Extreme Learning Machine (ELM)
Echo State Network (ESN)
Deep Residual Network (DRN)
Kohonen Network (KN)
Support Vector Machine (SVM)
Neural Turing Machine (NTM)

[5]

CNNs are primarily used for image processing. CNNs are feedforward networks (FFNs), which means that neurons do no back-propagate any information and only pass it forward to the next layer. A distinct feature of CNNs is that they consist of convolutional layers in which not all neurons are connected to all neurons in the previous and next layers [4].

RNNs are primarily used for text and speech processing. RNNs are FFNs like CNNs. However, unlike CNNs they are stateful, meaning that in addition to receiving data from the previous layer, each neuron retains data from previous passes. RNNs are also able to handle arbitrary input and output lengths unlike CNNs [4].

1979: convolution + weight replication + subsampling (Neocognitron)

2.2.2 OCR

OCR is a difficult problem since a sufficiently dynamic OCR system should be able to work on hundreds of different scripts containing tens of millions of different characters. As it stands, OCR systems are quite specific to a problem they solve, such that moving from latin script to an Urdu script is a difficult task for an OCR system [6].

Many OCR engines support over 100 languages, such as Tesseract, probably the most well known one. It started as proprietary software developed by Hewlett Packard in 1985, but was open sourced in 2005. Since then its development is continued by the community. Development has been sponsored by Google starting from 2006 [7].

Still, it is not ideal that OCR systems have to be trained to fit specific requirements. An ideal system would be able to recognise characters from whichever character set it is given as input.

CNNs and RNNs are not mutually exclusive, however, and B. Shi et al devised one of the most promising novel OCR concepts in recent history in the form of a convolutional recurrent neural network (CRNN) in their 2015 paper "An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition" [8].

Their reasoning is that popular models like CNN cannot be applied to sequence prediction, since they operate on fixed size inputs and outputs. CRNN then behaves like an RNN in that it can accept inputs and return outputs of arbitrary size, while still retaining properties of CNNs which make them invaluable in image processing [8].

2.2.3 Digital Image Enhancement

What makes digital image enhancement such a complicated research subject is the fact there is no agreed upon metric for measuring the performance of an image enhancement method. The results of a process such as image deblurring are largely subjective.

OCR is a very good context for measuring image enhancement, since the results, in this case OCR accuracy, are directly measurable before and after the enhancement process. M. D. Kim et al propose a quantitative method of measuring image deblurring accuracy in their paper "Dynamics-Based Motion Deblurring Improves the Performance of Optical Character Recognition During Fast Scanning of a Robotic Eye". In addition, they note that computation time is extremely important to modern image processing solutions [9]. Indeed, since a lot of image processing happens in smartphones and microcomputers computational requirements are crucial in comparing image enhancement methods.

In point of fact, motion deblurring is very important for OCR overall. Especially with so many photos of documents taken. In 2005, already when smartphones were not as prevalent as they are today, Xing Yu Qi at al address this concern in their paper "Motion Deblurring for Optical Character Recognition" where they handle three types of image degradation - blurring, point wise nonlinearities, and additive noise. They do this by first determining the orientation and extent of the blurred area and then recovering it [10].

Alongside the rising popularity of smartphones came the process of archiving documents by taking a photo. Scanning hardware has improved considerably since digitising documents became popular. However, in the case of photos taken by smartphones, many of the original copies of the documents have been lost since and improving the quality of existing documents is the only way to restore their readability. J. Jiao et al propose a novel CNN based two-stage technique for document deblurring [11].

A significant improvement of their approach, over previous approaches, is not requiring information about specifics of the style of the blurring for the image. They divide the image into sub-spaces and for each of those a corresponding degradative kernel space is developed. This solution not only works on different types of blurry images, but also on images that contain multiple types of blur, of which motion and focal blur are most prominent [11].

3 Background

3.1 Test Data Generation

Testing and comparing different deblurring and OCR methods requires a method to generate synthetic and consistent motion blurred images. This way the images can be divided to levels in which each level is measurably evenly blurrier than the previous.

For initial benchmarking the images are not required to be very different from each other, just consistent. Later, when retraining and testing neural networks, other images parameters, such as font size and colour also come to play. Benchmarking data is generated using a simple horizontal blur kernel as described by the following function -

Insert blur kernel function

From that function kernels of 5 different sizes are produced: 3x3, 5x5, 7x7, 9x9 and 11x11, increasing horizontal motion blur of the output in even increments.

Training and test data generation

3.2 Benchmarking

To get initial benchmark results 2 deblurring solutions and 2 OCR solutions were chosen. The deblurring solutions were "Scale-recurrent Network for Deep Image Deblurring" [**?**] and "Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring" [**?**]. The chosen OCR solutions were "Convolutional Recurrent Neural Network" [**?**] PyTorch implementation [**?**] and "Tesseract OCR Engine" [**?**].

Insert images

Insert results

3.3 Retraining OCR neural networks

A hypothesis is that training OCR networks using low quality images will improve their performance detecting text from low quality images. Both engines provide simple guides for training the network. The process was as follows:

- Get a copy of "Life and voyages of Christopher Columbus" by Washington Irving in txt format
- Generate a pdf file of the book, such that each word of the book is on a separate line while randomizing the text formatting on each page.
- Calculate the coordinates of each word.
- Generate an image file from each page of the pdf.

- Apply random intensity motion blur to each image.
- Crop each word from each image and assign labels.

4 Conclusion

Outline from the template:

What did you do?

What are the results?

Future work?

A paragraph or two about the results, including tables comparing different permutations of training data for networks.

Future work in regards to additional image and video processing methods that can be utilised for this purpose.

References

[1] G. Emanuel, “Statistical machine,” Patent 1 838 389, December, 1931. [Online]. Available: http://www.freepatentsonline.com/1838389.html

[2] (2007) Fiftieth anniversary of first digital image. [Online]. Available: https://www.sciencecodex.com/fiftieth_anniversary_of_first_digital_image

[3] A. Rosenfeld, “Picture processing by computer,” *ACM Comput. Surv.*, vol. 1, no. 3, pp. 147–176, Sep. 1969. [Online]. Available: http://doi.acm.org/10.1145/356551.356554

[4] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85 – 117, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608014002135

[5] F. V. VEEN. (2016) The neural network zoo. [Online]. Available: http://www.asimovinstitute.org/neural-network-zoo/

[6] I. Ahmad, X. Wang, R. Li, and S. Rasheed, “Offline urdu nastaleeq optical character recognition based on stacked denoising auto-encoder,” *China Communications*, vol. 14, pp. 146 – 157, 01 2017.

[7] (2018) Tesseract open source ocr engine. [Online]. Available: https://github.com/tesseract-ocr/tesseract

[8] B. Shi, X. Bai, and C. Yao, “An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 2298–2304, 2017.

[9] M. D. Kim and J. Ueda, “Dynamics-based motion deblurring improves the performance of optical character recognition during fast scanning of a robotic eye,” *IEEE/ASME Transactions on Mechatronics*, vol. PP, pp. 1–1, 01 2018.

[10] X. Y. Qi, L. Zhang, and C. L. Tan, “Motion deblurring for optical character recognition,” in *Eighth International Conference on Document Analysis and Recognition (ICDAR’05)*, Aug 2005, pp. 389–393 Vol. 1.

[11] J. Jiao, J. Sun, and N. Satoshi, “A convolutional neural network based two-stage document deblurring,” in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 01, Nov 2017, pp. 703–707.

Appendix

I. Glossary

II. Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, **Jaan Tohver**,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

   1.1 reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

   1.2 make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

   of my thesis

   Optical Character Recognition for Extremely Low Quality Images

   supervised by Gholamreza Anbarjafari, PhD

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, 16.12.2018