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

Pengecaman aksara optik (OCR) memainkan peranan penting dalam mendapatkan maklumat daripada imej ke dalam format teks yang boleh disunting oleh mesin. Contohnya, OCR biasanya digunakan dalam banyak aplikasi penglihatan komputer seperti pengenalan papan tanda automatik, terjemahan bahasa dan lain-lain. Walau bagaimanapun, pengecaman aksara telah menjadi semakin rumit ke atas dokumen lama yang terdiri daripada aksara-aksara cetak yang pecah dan kabur. Walaupun terdapat beberapa aplikasi OCR yang menggunakan teknik seperti pengekstrakan ciri dan padanan corak, tetapi kaedah ini masih tidak tepat untuk pengecaman. Dalam kerja ini, rangkaian pembelajaran mendalam (*deep learning network*) telah digunakan untuk melatih dan melaksanakan OCR dengan model Inception V3. Python Tensorflow API (versi GPU) telah digunakan untuk melatih rangkaian pembelajaran mendalam dan boleh didapati dengan lesen sumber terbuka Apache 2.0. Jaringan Inception V3 dilatih dengan 53,342 imej aksara yang dikumpulkan dari resit dan akhbar. Dari hasil percubaan, sistem mencapai ketepatan pengecaman yang lebih baik semasa menguna kualiti teks yang lebih rendah berbanding dengan OCR lain. Di samping itu, terdapat satu lagi eksperimen yang dijalankan untuk mengkaji punca-punca kegagalan pengecaman teks dan mencadangkan penyelesaian untuk mengatasi masalah ini. Analisis dan perbincangan juga dibuat tentang bagaimana sifat lapisan rangkaian neural mempengaruhi prestasi OCR dan masa latihan. OCR yang dicadangkan telah menunjukkan ketepatan yang lebih baik daripada OCR lain dan mempunyai potensi untuk mengatasi masalah mengenali kualiti teks yang tidak berkualiti.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CDR	-	Correct Detection Rates
CNN	-	Convolutional Neural Network
HOG	-	Histogram of Oriented Gradients
MSER	-	Maximally Stable Extremal Regions
OCR	-	Optical Character Recognition
SVM	-	Support Vector Machine
XML	-	Extensible Markup Language

LIST OF APPENDICES

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 colleted 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

1. Pan, S. J. and Yang, Q. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010. 22(10): 1345–1359.

2. Ho, T. K. and Nagy, G. OCR with no shape training. *Pattern Recognition, 2000. Proceedings. 15th International Conference on*. IEEE. 2000, vol. 4. 27–30.

3. Ashir, A. M. and Shehu, G. S. Adaptive clustering algorithm for optical character recognition. *Electronics, Computers and Artificial Intelligence (ECAI), 2015 7th International Conference on*. IEEE. 2015. SG–13.

4. Dundar, A., Jin, J. and Culurciello, E. Convolutional clustering for unsupervised learning. *arXiv preprint arXiv:1511.06241*, 2015.

5. Singh, D., Khan, M. A., Bansal, A. and Bansal, N. An application of SVM in character recognition with chain code. *Communication, Control and Intelligent Systems (CCIS), 2015*. IEEE. 2015. 167–171.

6. Minetto, R., Thome, N., Cord, M., Leite, N. J. and Stolfi, J. T-HOG: An effective gradient-based descriptor for single line text regions. *Pattern Recognition*, 2013. 46(3): 1078–1090. ISSN 00313203. doi:10.1016/j.patcog.2012.10.009. URL `http://linkinghub.elsevier.com/retrieve/pii/S0031320312004438`.

7. Anil, R., Manjusha, K., Kumar, S. S. and Soman, K. P. Convolutional Neural Networks for the Recognition of Malayalam Characters. In: Satapathy, S. C., Biswal, B. N., Udgata, S. K. and Mandal, J. K., eds. *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*. Cham: Springer International Publishing, vol. 328. 493–500. 2015.

8. Girshick, R., Donahue, J., Darrell, T. and Malik, J. Rich feature hierarchies

for accurate object detection and semantic segmentation tech report. *arXiv preprint arXiv:1311.2524*, 2014. 5.

9. Ko, D.-G., Song, S.-H., Kang, K.-M. and Han, S.-W. Convolutional Neural Networks for Character-level Classification. *IEIE Transactions on Smart Processing and Computing*, 2017. 6(1): 53–59.

10. Tang, Y., Peng, L., Xu, Q., Wang, Y. and Furuhata, A. CNN Based Transfer Learning for Historical Chinese Character Recognition. *Document Analysis Systems (DAS), 2016 12th IAPR Workshop on*. IEEE. 2016. 25–29.

11. Yosinski, J., Clune, J., Bengio, Y. and Lipson, H. How transferable are features in deep neural networks? *Advances in neural information processing systems*. 2014. 3320–3328.

12. Kyle, L. Applications of Deep Learning: The Good, The Bad and The Opinion, 2015. URL `https://duo.com/blog/`.

13. Andrew, L. B. Deep Learning 101 - Part 1: History and Background, 2017. URL `https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part1.html`.

14. Rein, D. Convolutional Neural Networks for Image and Video Processing, 2017. URL `https://wiki.tum.de/display/lfdv/Introduction`.

15. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X. and Wang, G. Recent advances in convolutional neural networks. *arXiv preprint arXiv:1512.07108*, 2015.

16. Henniges, P., Granger, E. and Sabourin, R. Factors of overtraining with fuzzy ARTMAP neural networks. *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*. IEEE. 2005, vol. 2. 1075–1080.

17. Oquab, M., Bottou, L., Laptev, I. and Sivic, J. Learning and transferring mid-level image representations using convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014. 1717–1724.

18. Felleman, D. J. and Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 1991. 1(1):

1–47.

19. Lee, T. S. and Mumford, D. Hierarchical Bayesian inference in the visual cortex. *JOSA A*, 2003. 20(7): 1434–1448.

20. Morris, G., Nevet, A. and Bergman, H. Anatomical funneling, sparse connectivity and redundancy reduction in the neural networks of the basal ganglia. *Journal of Physiology-Paris*, 2003. 97(4-6): 581–589.

21. Kruger, N., Janssen, P., Kalkan, S., Lappe, M., Leonardis, A., Piater, J., Rodriguez-Sanchez, A. J. and Wiskott, L. Deep Hierarchies in the Primate Visual Cortex: What Can We Learn for Computer Vision? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013. 35(8): 1847–1871.

22. Zhang, J., Tao, C. and Wang, P. A Review of Soft Computing Based on Deep Learning. *Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), 2016 International Conference on*. IEEE. 2016. 136–144.

23. G. E, H. and R, R. S. Reducing the dimensionality of data with neural networks. *Science*, 2006. 313(5786): 502–504. ISSN 0036-8075, 1095-9203. doi:10.1126/science.1129198. URL `http://www.sciencemag.org/cgi/doi/10.1126/science.1129198`.

24. Larochelle, H., Bengio, Y., Louradour, J. and Lamblin, P. Exploring strategies for training deep neural networks. *Journal of Machine Learning Research*, 2009. 10(Jan): 1–40.

25. Xavier, G. and Yoshua, B. Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research*, 2010. 9: 249–256.

26. Wang, P. and Cottrell, G. W. Basic Level Categorization Facilitates Visual Object Recognition. *arXiv preprint arXiv:1511.04103*, 2015.

27. Athiwaratkun, B. and Kang, K. Feature representation in convolutional neural networks. *arXiv preprint arXiv:1507.02313*, 2015.

28. Ciresan, D. C., Meier, U., Masci, J., Maria Gambardella, L. and Schmidhuber, J. Flexible, high performance convolutional neural networks for image classification. *IJCAI Proceedings-International Joint Conference on Artificial*

Intelligence. Barcelona, Spain. 2011, vol. 22. 1237.

29. Ciregan, D., Meier, U. and Schmidhuber, J. Multi-column deep neural networks for image classification. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE. 2012. 3642–3649.

30. Lawrence, S., Giles, C. L., Tsoi, A. C. and Back, A. D. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, 1997. 8(1): 98–113.

31. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C. and Fei-Fei, L. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 2015. 115(3): 211–252. ISSN 0920-5691, 1573-1405. doi:10.1007/s11263-015-0816-y. URL `http://link.springer.com/10.1007/s11263-015-0816-y`.

32. Maas, A. L., Hannun, A. Y. and Ng, A. Y. Rectifier nonlinearities improve neural network acoustic models. *Proc. ICML*. 2013, vol. 30.

33. Zeiler, M. D. and Fergus, R. Visualizing and understanding convolutional networks. *European conference on computer vision*. Springer. 2014. 818–833.

34. Simonyan, K. and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

35. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. R. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.

36. Springenberg, J. T., Dosovitskiy, A., Brox, T. and Riedmiller, M. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014.

37. Lin, M., Chen, Q. and Yan, S. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

38. Fung, G. P. C., Yu, J. X., Lu, H. and Yu, P. S. Text classification without negative examples revisit. *IEEE Transactions on Knowledge and Data Engineering*, 2006. 18(1): 6–20.

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.

40. Sarinnapakorn, K. and Kubat, M. Combining Subclassifiers in Text Categorization: A DST-Based Solution and a Case Study. *IEEE Transactions on Knowledge and Data Engineering*, 2007. 19(12): 1638–1651.

41. Raina, R., Battle, A., Lee, H., Packer, B. and Ng, A. Y. Self-taught learning: transfer learning from unlabeled data. *Proceedings of the 24th international conference on Machine learning*. ACM. 2007. 759–766.

42. Daume III, H. and Marcu, D. Domain adaptation for statistical classifiers. *Journal of Artificial Intelligence Research*, 2006. 26: 101–126.

43. Shimodaira, H. Improving predictive inference under covariate shift by weighting the log-likelihood function.pdf. *Journal of Statistical Planning and Inference*, 2000. 90: 227–244.

44. Zadrozny, B. Learning and evaluating classifiers under sample selection bias. *Proceedings of the twenty-first international conference on Machine learning*. ACM. 2004. 114.

45. Dai, W., Yang, Q., Xue, G.-R. and Yu, Y. Self-taught clustering. *Proceedings of the 25th international conference on Machine learning*. ACM. 2008. 200–207.

46. Wang, Z., Song, Y. and Zhang, C. Transferred dimensionality reduction. *Machine learning and knowledge discovery in databases*, 2008: 550–565.

47. T, d. The Chars74K dataset, 2012. URL `http://www.ee.surrey.ac.uk/CVSSP/demos/chars74k/`.

48. yeephy, c. tensorflow input image by tfrecord, 2017. URL `https://github.com/yeephycho/tensorflow_input_image_by_tfrecord/tree/master/src`.

49. Hvass, L. Tensorflow tutorial 06 CIFAR-10, 2016. URL `https://github.com/Hvass-Labs/TensorFlow-Tutorials/blob/master/06_CIFAR-10.ipynb`.

50. Prajit, R., Barret, Z. and Quoc, V. L. Searching for Activation Function. *arXiv*, 2017. 2.

51. Chung, H., Lee, S. J. and Park, J. G. Deep neural network using trainable activation functions. *Neural Networks (IJCNN), 2016 International Joint Conference on*. IEEE. 2016. 348–352.

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

53. Ramachandran, P., Zoph, B. and Le, Q. Searching for Activation Functions. *arXiv*, 2017. URL `https://arxiv.org/pdf/1710.05941.pdf`.