Review Report on "Gradient-Based Learning Applied to Document Recognition"

Introduction

This review report is about giving an overview of the research paper- "Gradient-Based Learning Applied to Document Recognition"[1] which designed an innovating and successful multi-module architecture called Graph Transform Network for pattern recognition and also evaluate the performance of the convolutional neural network for identifying handwritten characters. The reason why this paper is selected to be analyzed is that this paper not only introduces how to achieve image identification/pattern recognition but also describe how to build a graph transformer system in detail, which is good for the future study in deep learning and image processing, which also explains how gradient-based learning algorithm like CNN is feasible for image recognition. However, even many innovations and improvements were achieved in this research, there are still some limitations to the research. In this review paper, firstly, the summarized content of the research will be described in the content section. Secondly, the innovation of the research will be explained in detailly after the content section. Thirdly, from the evaluation part of the original paper, the technical quality will be talked about. Fourthly, this report will arrive at the application of the new technology. Finally, the last section will analyze how the author organized the report.

▾ Content

In the past, most pattern recognition systems are built with the semi-automatic structure which is a combination of automatic learning techniques and hand-crafted algorithms. This literature introduces two remarkable graph transform networks(GTN) with multi-modules for segmenting, classifying, and resequencing the handwritten characters string, which estimates the manual work and increases the accuracy for the pattern recognition. On the one hand, the first GTN is combined with 4 modules the Segmentation Transformer, Character Recognition Transformer(CRT), Composition Transformer, and Beam Search Transformer, which shown as Figure 2. Specifically, the segmentation transformer will use the Heuristic Over-segmentation algorithm to divide the whole picture to a set of arcs(segments) as the input of CRT. And then the CRT which is the most important part of the research is constituted by the CNN trained with Gradient-based learning and back-propagation method. For training the CNN, such as the LeNet-5(figure 1) with 7 layers as the innovated CNN described detailly in the research, the input image is very large and consists of many pixels. For example, a fully connected network with 100 hidden units contains thousands of weights. So, CNN will use its particular function-convolution for extracting the feature data from the original image to be the feature maps, so that the gradient-based learning layer and back-propagation can handle the data. Next, the author suggests using the grammar graph and Viterbi algorithm to implement the composition transformer and the beam search transformer.
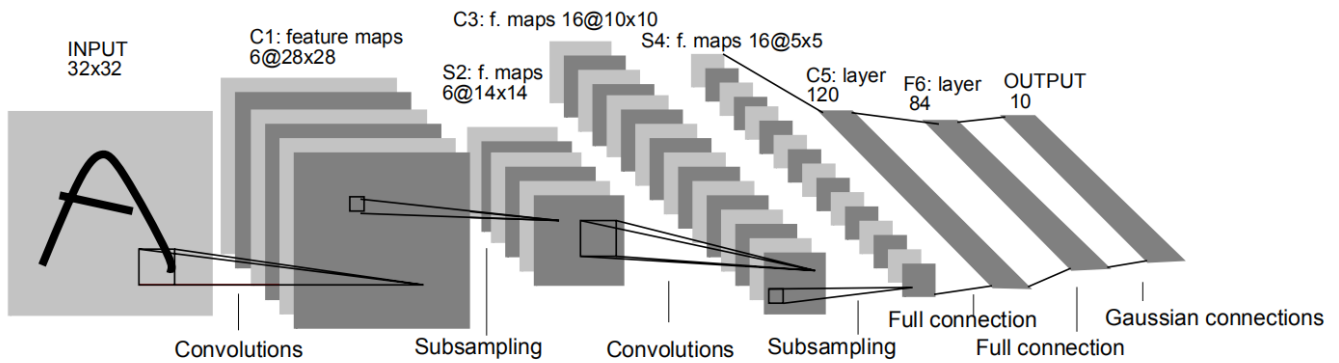
Figure 1:LeNet-5[1]

On the other hand, the second GTN(Figure 3) structure uses the Space-Displacement Neural Network which is a kind of CNN with the ability to automatically scan possible character positions of the original image to replace the segmentation transformer and Character Recognition Transformer. Besides, the composition transform is separated to be two-level composition transformer-character and word level, because the identified characters of the SDNN's output sets are uncertain, which is depended on the identified possible positions of characters. For character level, with the help of CNN, the Hidden Markov Models(HMM) which is famous for classical speech recognition can be used in the character-level composition with the reduced computational cost. For the word level, this GTN can be used for identifying continuously on-line handwritten strings, so after transforming characters, the architecture will recombine the words with a language model and output the interpretation graph.
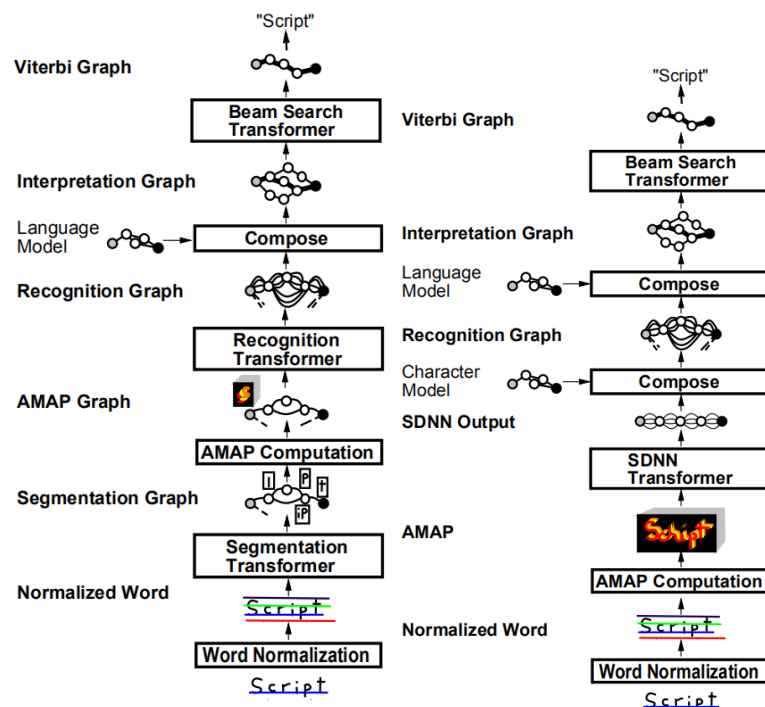

Fiture 2: first GTN with segmentation[1]　　　　　　　　　Figure 3: GTN with SDNN[1]

▼ Innovation

The most important and intelligent innovation of the research is to use the CNN to eliminate the hand-crafted part because before the research of the author, the most pattern recognition system is built in a semi-automatic architecture, which means the graph feature extraction is manual work. Therefore, It is remarkable for the importance and innovation of using the specialty of LeNet-5(Figure 1) CNN to achieve automatic feature map extracting and recognition in the research.

Also, there are two kinds of GTNs designed and experimented by the authors in the research. The authors separate the whole transformer network to be modules by the object-oriented programming(OOP) method and use the back-propagation method to compute the derivatives in the system. The innovation of the first GTN is to increase the accuracy of pattern recognition and combine the Heuristic Over-Segmentation, CNN, Viterbi algorithm and Grammar graph to fully automatical transform the graph to be words. The innovation of the second GTN is to remove the segmentation part of the first GTN by using SDNN to replace CNN and also use the HMM to achieve the character recomposition module.

What's more, the authors designed the generalized transduction for achieving the composition transform. The generalized transduction is divided as Check, Fprod, and Bprod function. For determining whether corresponding arc(s) should be created in the output graph, the Check function will compare the arcs(from recognition part) and output a boolean value. The Fprod function will be activated when the corresponding arc(s) should be built and creates new arcs and nodes between the up node and down node in the output new graph. The Bprod is called during training for propagating gradient data from the output subgraph to the data structures on the arc1 and arc2(from the recognition layer)

▾ Technical quality

The technology of the research is of high quality. Generally speaking, the authors do the literature review for past methods to pattern recognition for supporting that the pattern recognition is achieved by the hybrid-technology which combines automatic machine learning algorithm and manual feature extraction method. In the C part of section I, the authors reference several examples like the Boltzmann machines to prove that the multi-layer neural network achieved with gradient-back propagation algorithm is feasible for classifying non-linear objects, and the artificial image feature extraction can be replaced by the inner function of CNN. In my opinion, the quality of this part is good enough to explain why NN is reliable for handling such non-linear classification.

Besides, for explaining how CNN deal with the 2D graph to be the OUTPUT, in section II, the author draws the structure of the LeNet-5 CNN to show the inner cooperation between convolutions layers, subsampling layers, full connection layers, and the output layer. From figure 4, for comparing the performance of CNN with other classifying algorithms, the author experimented more than 11 algorithms like SVM, K-NN, PCA, Tangent Distance, and LeNet-1 to classify the single character image to be character, and the LeNet-5 shows the best performance(0.95% errors) without the help of artificially distorted examples(LeNet-5 0.8% errors) and boosting method(LeNet-4 0.7% errors). Also, figure 13 of section II shows the ability of LeNet-5 to identify the character from images with various degrees of noisy conditions. What's more, For the GTN with the SDNN approach, this hybrid and innovative architecture also tested with the normal and noised images and provided with a well-structured architecture diagram and algorithm coding like the generalized-composition function which is for compositing the recognized graphs to be a new graph. However, the GTNs sometimes still can not recognize some handwritten characters because the written style of different people is different. So, in my opinion, the author should deal with this situation in the report or suggest some methodology for further research for this case.
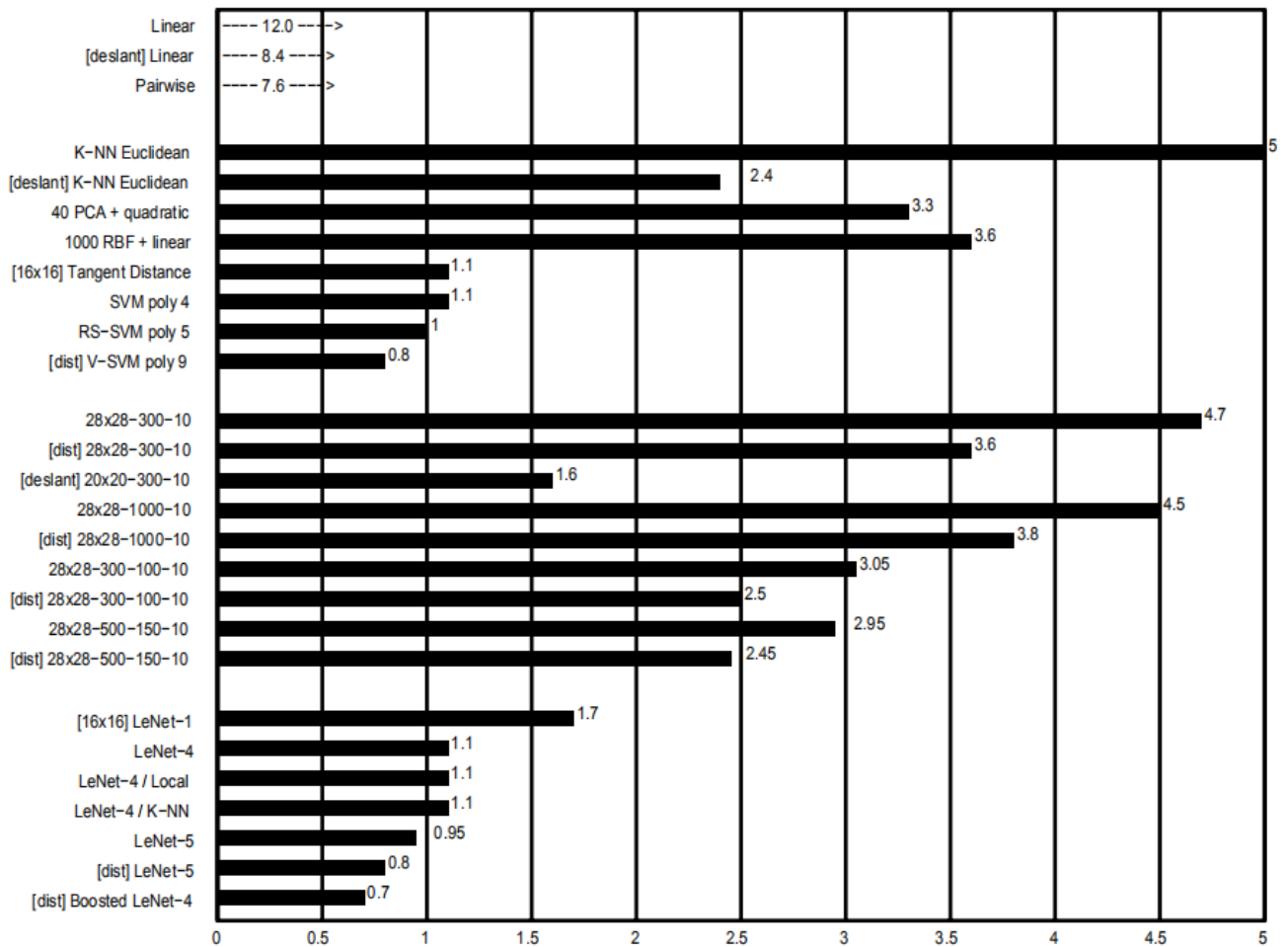
Figure 4: 11 algorithms experiments for handwritten character recognition[1]

▾ Application and X-factor

From the paper, it is known that this system is already used in the commercial area in the NCR Corporation line of the bank's check recognition systems which is used for recognizing the print or handwritten character from checks. The application has been realized and approved, because, in the United States, the system has read millions of checks per month in various banks. With the powerful document recognition ability, the system can not only be used for checks checking but also can be used for scanning the handwritten or printing bank statement as the data resource for fraud detection.

According to the Australian Payments Network Limited 2018(Figure 5)[2], Australia losses 598 thousand dollars in 2012. However, this number shows a linear increase during the past years. In 2017, 748 thousand dollar losses due to the fraud data. With the document recognition system, the bank staff can just scan the materials into the system, and don't have to manually input the data into the banking system. After scanning the data, it may be used with the fraud detection system like the financial statements fraud detection system designed by Humphery(2011). With the document recognition system, the time of manual input will be reduced, which means it can reduce the staff who need to input data by hand every day.
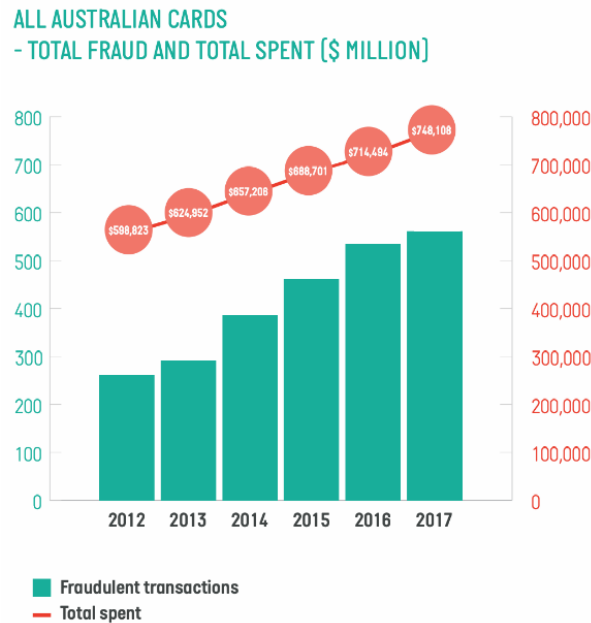
Figure 5: Australian Payments Network Limited 2018[2]

Presentation

The overall structure of the paper is well. In the introduction section, the authors introduce several fundamental knowledge about the old-fashioned semi-automatic pattern recognition system and the new technologies of machine learning, which makes the readers understand the history and limitation of pattern recognition methodology. Besides, the author uses many diagrams for presenting the clear structure of many algorithms. For example, the LeNet-5, first GTN, GTN with SDNN are all displayed with a well-structured diagram. However, the authors didn't highlight the significant parts of the formulas in the paper, In this situation, if a reader is a beginner in the machine learning area, he or she will feel very confused when they try to understand the formula. For example, the formulas of the fireworks algorithm[4] are clearly explained by Tan Y. (2010) by highlighting the keywords of the formulas. What's more, the author should summarize every argument of the formula after it is described in detailly, so that readers can compare the understanding of detail with the summary.

References

[LEC98][1]: Lecun, Y., et al. 1998. "Gradient-based learning applied to document recognition." Proceedings of the IEEE, vol. 50,no. 11 pp. 2278-2324.

[2]:Australian Payments Network Limited 2018, AustralianPaymentCardFraud-2018, Australian Payments Network Limited, viewed 1 June 2019, https://www.auspaynet.com.au/sites/default/files/2018-08/AustralianPaymentCardFraud-2018-Report.pdf. Turban, E., Sharda, R. & Delen, D. 2010, Decision Support and Business Intelligence Systems, Ninth edn, Prentice Hall Press.

[3]Humpherys, S.L., Moffitt, K.C., Burns, M.B., Burgoon, J.K. & Felix, W.F. 2011, 'Identification of fraudulent financial statements using linguistic credibility analysis', Decision Support Systems, vol. 50, no. 3, pp. 585-94.

[4]:Tan Y., Shi Y., Tan K.C. 2010, Advances in Swarm Intelligence, 1st edn, Springer, Berlin, Heidelberg.

Appendix

The orginal notebook file is available at the link:

https://github.com/geroboog/machineLearning/blob/master/A1_ReportDraft.ipynb