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

## First impression

According to the abstract and information of the paper as well as my personal experience, as people using smart devices, there is a growing demand for continuous speech and handwriting recognition. Compared with the traditional pattern recognition, LeCun et al. introduced a new automatic machine learning based on multilayer neural networks by extracting needed features from pixel images rather than the previous method that using traditional hand-crafted feature collector.

The reason that I choose this paper is that many important methods of machine learning (such as GTN, SVM and K-NM) have been mentioned or applied in this article and it is helpful for my understanding of machine learning and future study.

## Introduction

This report will contain explanations and review of a chosen machine learning literature. The chosen literature is "Gradient-Based Learning Applied to Document Recognition" which is written by LenCun, Bottou, Bengio and Haffner in 1998. Its critiques will be divided into content, innovation, technical quality, application and X-factor and presentation so that make report clearer and easier to understand.

## Content

### Relevant background and Handwritten digit recognition

This literature paper first introduced the development and history of pattern recognition as well as the evolution of the traditional pattern. In the early days, with the abundance and diversity of natural data, combined with manual production, which led to a low- accuracy recognition system. With the development of the technology, people come up with the first module, called the feature extractor, convert the input data model to lower-dimensional data or short strings for easy comparison and matching. However, with the growth of demand for handwriting recognition in the market and according to the evidence that a big part of modern commercial OCR (Optical character recognition) systems use some form of multilayer neural network. The handwritten digit recognition based on gradient-learning, as a new approach, lead to more accurate introduced by LenCun, Bottou, Bengio and Haffner (1998).

LenCun et al. introduce convolutional networks, which can extract features from pixel images rather than the traditional hand-crafted feature extractor. In addition, in convolutional networks, architecture needs to use and incorporate knowledge by local patterns and relevant technology. Convolutional neural networks have fewer parameters compare with the fully connected neural networks. Therefore, the core of the convolutional neural network, named LeNet-5 is explained by the author.

### LeNet-5

In addition to input parameters which is a 32\*32 pixel image, more than the character in the database, the framework mainly consists of 7 layers including C1, S2, C3, S4, C5, F6 and RBF output, in addition, all of them contain trainable parameters. Layer C1 is the first layer including six feature maps whose size is 28\*28 and each map connect to a 5\*5 neighborhood in the input parameters so that the connection from input work; The next layer is a subsampling called S2 also with 6 feature images that have 14\*14 area; Followed by a convolutional layer called C3 with 16 feature maps which have the size of 10\*10; After that is a subsampling layer named S4 which has 16 maps with the size of 5\*5 and each map also connect with a 2\*2 neighborhood. The fifth convolutional section, as a fully-connect layer, which has 48120 parameters. Finally, F6, as the last layer connecting with the C5, is also a fully-connect section, which contain 84 units. Euclidean Radial Basis Function（RBF）units, as the last input layer, play the role of criterion based on loss function.

### Methods comparison and results

LenCun et al. argue that design an individual recognition is a benchmark for comparing recognition methods. NIST is a database to train and test system which including training set named SD-3 and testing set named SD-1. However, SD-3 has more advantages than SD-1 in recognition. According to the results, such as LeNet-5, after about 10 tests, the error rate was stable at 0.95%, and after 19 tests, it was stable at 0.35%. Therefore, in LeNet-5, over-training occur will improve the accuracy of the system.

### Classifier Comparison

This section illustrates and compares convolutional neural networks with other trainable classification methods, such as K-NN, SVM, PCA and so on. Considering to different aspects including training time, designer selection, memory requirement and recognition time. Each classification approach has its pros and cons, for example, K-NN classification has the advantage that no training time but large memory requirement and recognition time. The author found that Boosted LeNet-4 performs best and has the lowest error rate, with 0.7%, followed by LeNet-5 and SVM ploy, both 0.8%.

## Innovation

As for innovation, I would like to talk about LeNet-5 architecture.

### LeNet-5 architecture

LeNet-5 model, as the first proposed and successfully applied in the pattern recognition of many commercial Banks, Later, some of its design ideas and methods were learned and emulated by many new technologies, such as AlexNet (2012), GoogleNet (2014) and so on.

LeNet-5 model is small and easy to understand, especially when trained on the MNIST. And it is very friendly for beginners. LeNet-5 combine the convolutional layer and subsampling layer to reduce the missing of the connection between each layer. Another innovation is that using the distributed code as an output rather than sigmoid units, because sometimes classifiers are used to reject non-characters.

## Technical quality

I think the technical innovation of the paper is very high. Both there are some research information results between LeNet-5 and other methods as well as classifier comparison. And get a low error rate and high accuracy in the results.

### The quality of LeNet-5 implementation

The author put and train different versions of LeNet-5 model in the MNIST database. Iterate through the training data for each section and the result of the error rate was stable at 0.95% at 10 tests. And after 19 tests, it was stable at 0.35%. Therefore, in LeNet-5, over-training occur will improve the accuracy of the system. And the accuracy of the systems depends on the quality of the pieces and cut from the recognizer and segmented character.

### Classifier Comparison

As we mentioned, the results of different various classifier also provide strong evidence of the technical quality. According to the graph, the author found that Boosted LeNet-4 performs best and has the lowest error rate, with 0.7%, followed by LeNet-5 and SVM ploy, both 0.8%. Therefore, the LeNet-4 and LeNet-5 models have a important values in image recognition application.

## Application and X-factor

As we can see in the paper, Convolutional Neural Networks, as a useful technical support for processing two-dimensional shapes, has made progress and breakthroughs in many aspects. A checking reading based on GCN technique is widely used in industrial deployment, utilizing segmentation transformer, recognition graphics and composition transformer, which make it very cost-efficient and can produce solutions quickly.

Although, as the development of the machine learning technique and computing capacity, LeNet-5 is replaced by more and more techniques in ImageNet Large Scale Visual Recognition Challenge ([ILSVRC](https://en.wikipedia.org/wiki/ImageNet#ImageNet_Challenge)). Such as AlexNet (2012), GoogleNet (2014) and so on. However, they are not suitable for some document paper areas because of the expensive computing and equipment. In my opinion, I think LeNet-5 also has its function in document recognition, such as Identification of bank check signatures and Text digital recognition.

I also find some interesting application in other paper, such as action recognition. Pyramid pooling based convolutional neural network and combine the action recognition targets, restructure into three-dimensional convolutional neural network provides the Ability to collect unstable objects.

## Presentation

The quality of this paper is good enough for the standard of a literature article as its well-structured and very comprehensive introduction of the Convolutional neural network and relevant methods. The paper uses statistics, background-history, many graphs and labs examples to help readers to understand the information and techniques in the article, such as LeNet-5 architecture and GTNs as well as results of various classifiers methods. However, I think the presentation of the paper need to be improved. Firstly, it is very hard for me to find some applications in the paper, so that I have to search for other sources. Secondly, the paper will be more attractive if it provides more information and processions about GTN development and relevant application.

## References

[1] LEC98: LeCun, Y., Bottou, L., Bengio, Y.& Hanffner, P. 1998, 'Gradient Based Learning Applied to Document Recognition',Proceedings of the IEEE, pp.2278-2324

[2] KIPF, T. (2016). *How powerful are Graph Convolutional Networks?*. [online] Tkipf.github.io. Available at: https://tkipf.github.io/graph-convolutional-networks/ [Accessed 30 Sep. 2016].

[3] Das, S. (2017). *CNN Architectures: LeNet, AlexNet, VGG, GoogLeNet, ResNet and more ….*. [online] Medium. Available at: https://medium.com/@sidereal/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5 [Accessed 17 Nov. 2017].

[4] Bhandare, A., Bhide, M., Gokhale, P. and Chandavarkar, R. (2019). *Applications of Convolutional Neural Networks*. (IJCSIT) International Journal of Computer Science and Information Technologies, pp.Vol. 7 (5), 2016, 2206-2215.