Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis

Patrice Y. Simard, Dave Steinkraus, John C. Platt

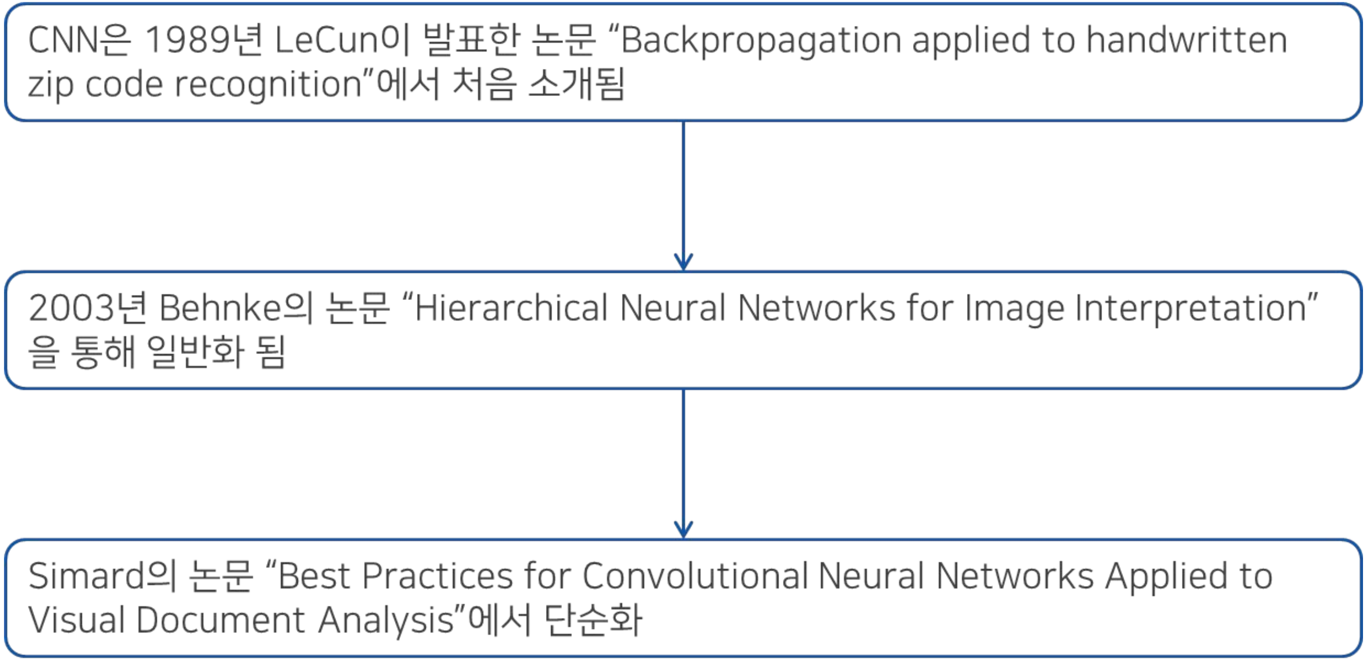
2020 . 03 . 31

산업경영공학과

이종호

Background Knowledge

Convolutional Neural Network (CNN)의 역사



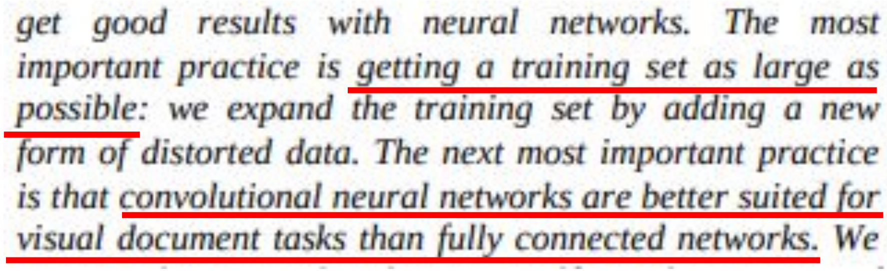
CNN은 1989년 LeCun이 발표한 논문 “Backpropagation applied to handwritten zip code recognition”에서 처음 소개됨

2003년 Behnke의 논문 “Hierarchical Neural Networks for Image Interpretation”을 통해 일반화 됨

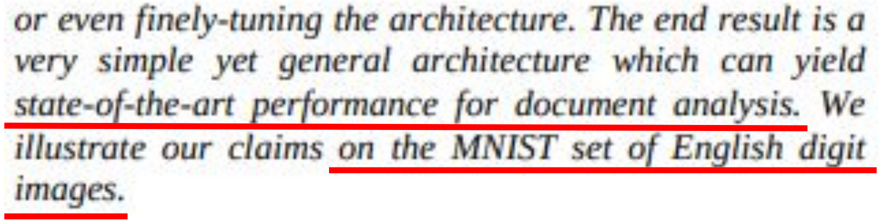
Simard의 논문 “Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis”에서 단순화

Abstract

1. Getting a training set as large as possible
2. Using CNN better suited for visual document tasks than fully connected networks



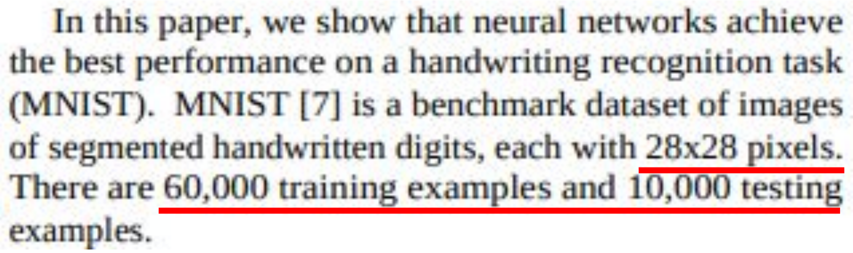
get good results with neural networks. The most important practice is getting a training set as large as possible: we expand the training set by adding a new form of distorted data. The next most important practice is that convolutional neural networks are better suited for visual document tasks than fully connected networks. We



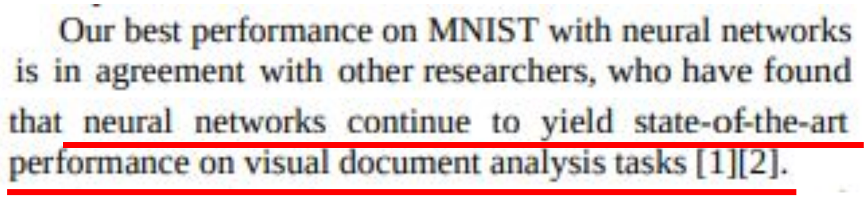
or even finely-tuning the architecture. The end result is a very simple yet general architecture which can yield state-of-the-art performance for document analysis. We illustrate our claims on the MNIST set of English digit images.

DataSet

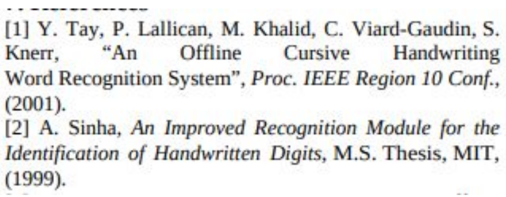
MNIST DATASET



In this paper, we show that neural networks achieve the best performance on a handwriting recognition task (MNIST). MNIST [7] is a benchmark dataset of images of segmented handwritten digits, each with 28x28 pixels. There are 60,000 training examples and 10,000 testing examples.

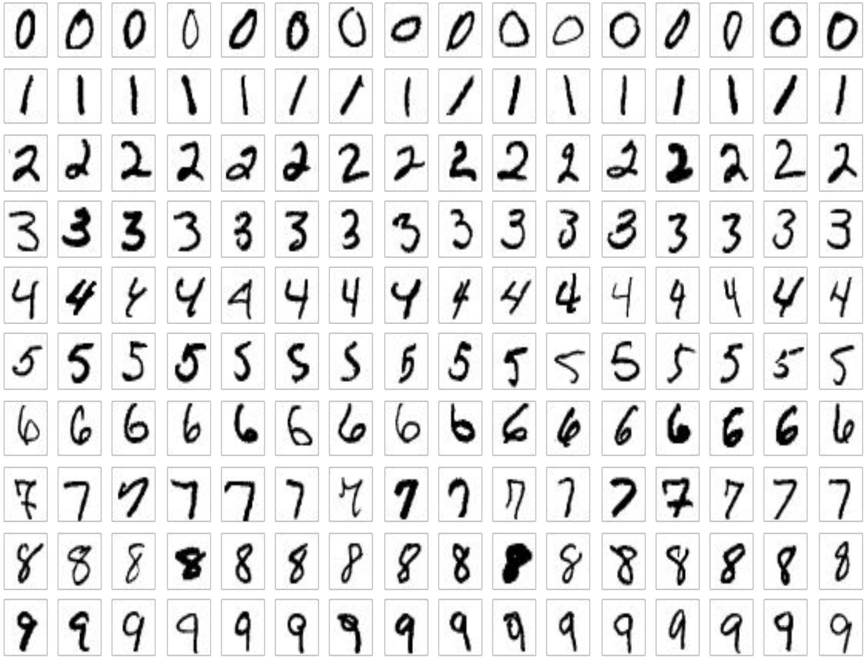


Our best performance on MNIST with neural networks is in agreement with other researchers, who have found that neural networks continue to yield state-of-the-art performance on visual document analysis tasks [1][2].

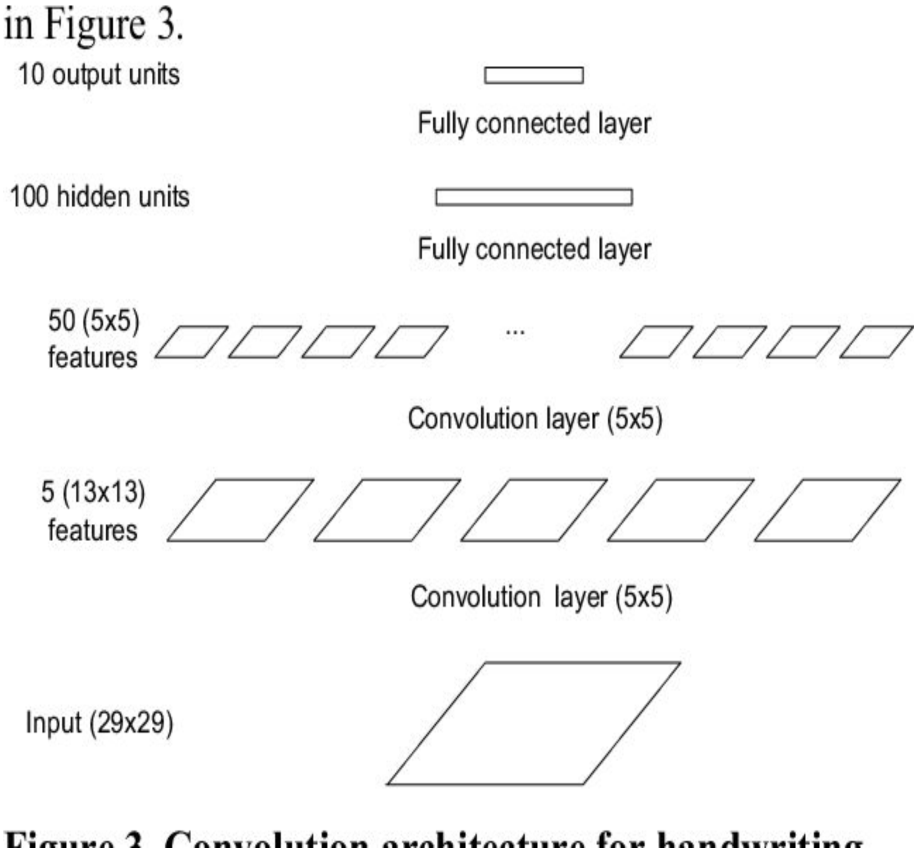


[1] Y. Tay, P. Lallican, M. Khalid, C. Viard-Gaudin, S. Knerr, "An Offline Cursive Handwriting Word Recognition System", *Proc. IEEE Region 10 Conf.*, (2001).

[2] A. Sinha, *An Improved Recognition Module for the Identification of Handwritten Digits*, M.S. Thesis, MIT, (1999).



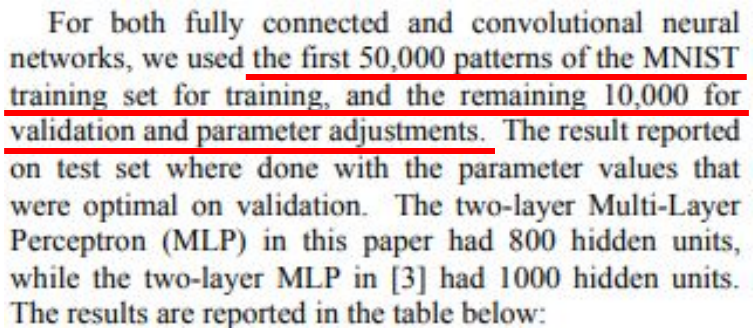
Overall architecture for MNIST



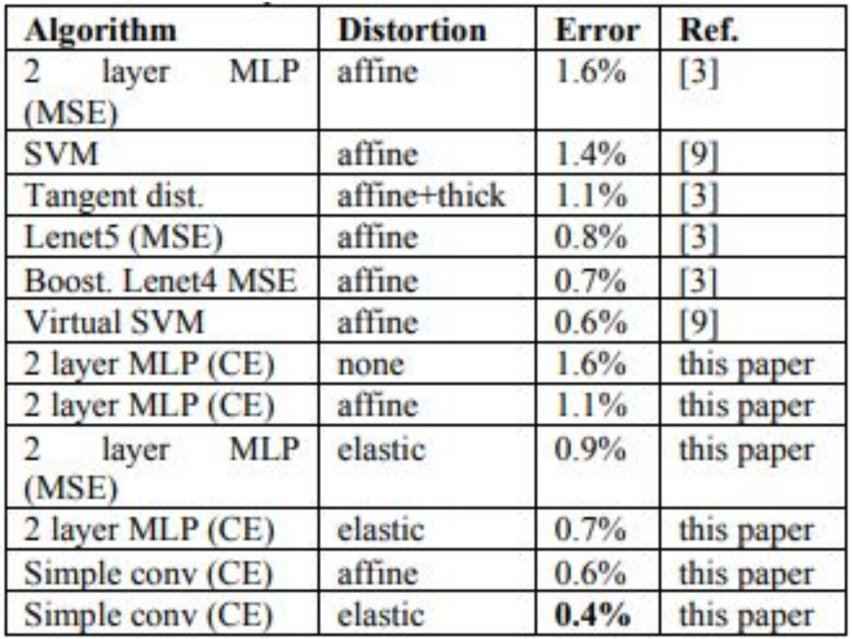
in Figure 3.

Figure 3. Convolution architecture for handwriting

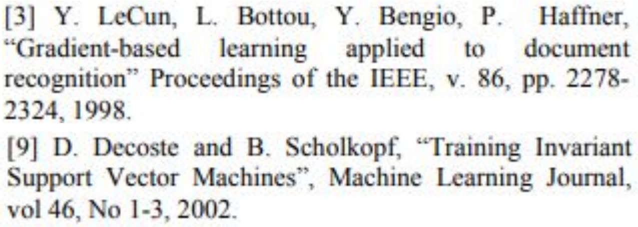
Result



For both fully connected and convolutional neural networks, we used the first 50,000 patterns of the MNIST training set for training, and the remaining 10,000 for validation and parameter adjustments. The result reported on test set where done with the parameter values that were optimal on validation. The two-layer Multi-Layer Perceptron (MLP) in this paper had 800 hidden units, while the two-layer MLP in [3] had 1000 hidden units. The results are reported in the table below:



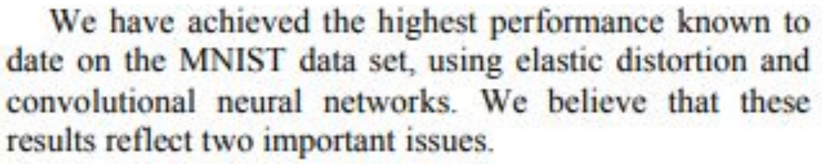
Algorithm	Distortion	Error	Ref.
2 layer MLP (MSE)	affine	1.6%	[3]
SVM	affine	1.4%	[9]
Tangent dist.	affine+thick	1.1%	[3]
Lenet5 (MSE)	affine	0.8%	[3]
Boost. Lenet4 MSE	affine	0.7%	[3]
Virtual SVM	affine	0.6%	[9]
2 layer MLP (CE)	none	1.6%	this paper
2 layer MLP (CE)	affine	1.1%	this paper
2 layer MLP (MSE)	elastic	0.9%	this paper
2 layer MLP (CE)	elastic	0.7%	this paper
Simple conv (CE)	affine	0.6%	this paper
Simple conv (CE)	elastic	**0.4%**	this paper



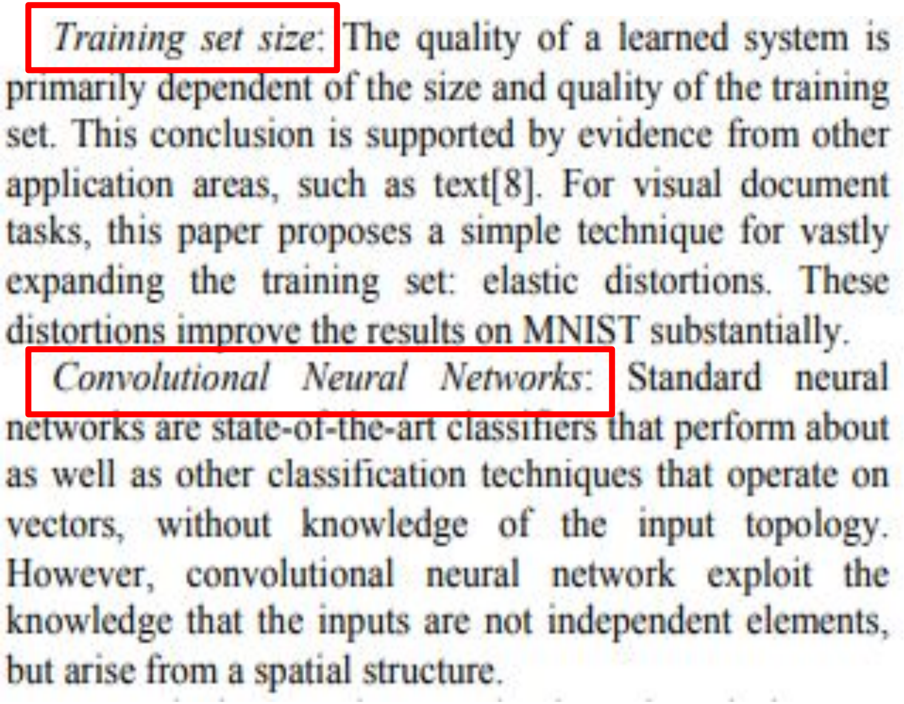
[3] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition" Proceedings of the IEEE, v. 86, pp. 2278-2324, 1998.

[9] D. Decoste and B. Scholkopf, "Training Invariant Support Vector Machines", Machine Learning Journal, vol 46, No 1-3, 2002.

Conclusions



We have achieved the highest performance known to date on the MNIST data set, using elastic distortion and convolutional neural networks. We believe that these results reflect two important issues.



Training set size: The quality of a learned system is primarily dependent of the size and quality of the training set. This conclusion is supported by evidence from other application areas, such as text[8]. For visual document tasks, this paper proposes a simple technique for vastly expanding the training set: elastic distortions. These distortions improve the results on MNIST substantially.

Convolutional Neural Networks: Standard neural networks are state-of-the-art classifiers that perform about as well as other classification techniques that operate on vectors, without knowledge of the input topology. However, convolutional neural network exploit the knowledge that the inputs are not independent elements, but arise from a spatial structure.

참고자료

https://excelsior-cjh.tistory.com/79

https://excelsior-cjh.tistory.com/180

https://excelsior-cjh.tistory.com/152?category=940399