

# MULTI-MODULAR ARCHITECTURE BASED ON CONVOLUTIONAL NEURAL NETWORKS FOR ONLINE HANDWRITTEN CHARACTER RECOGNITION

*Emilie POISSON\*, Christian VIARD GAUDIN\*, Pierre-Michel LALLICAN\*\**

\* Image Video Communication, IRCCyN UMR CNRS 6597  
EpuN – Rue Christian Pauc – BP 50609 – 44306 NANTES Cedex 3 – France  
{Emilie.Poisson ; Christian.Viard-Gaudin} @polytech.univ-nantes.fr

\*\* VISION OBJECTS – 9, rue Pavillon – 44980 Ste Luce sur Loire – France  
pmlallican@visionobjects.com

## ABSTRACT

In this paper, several convolutional neural network architectures are investigated for online isolated handwritten character recognition (Latin alphabet). Two main architectures have been developed and optimised. The first one, a TDNN, processes online features extracted from the character. The second one, a SDNN, relies on the off-line bitmaps reconstructed from the trajectory of the pen. Moreover, an hybrid architecture called SDTDNN has been derived, it allows the combination of on-line and off-line recognisers. Such a combination seems to be very promising to enhance the character recognition rate. This type of shared weights neural networks introduces the notion of receptive field, local extraction and it allows to restrain the number of free parameters in opposition to classic techniques such as multi-layer perceptron. Results on UNIPEN and IRONOFF databases for online recognition are reported, while the MNIST database has been used for the off-line classifier.

## 1. INTRODUCTION

Handwriting recognition is classically separated in two distinct domains : online and offline recognition. These two domains are differentiated by the nature of the input signal. For offline recognition, a static representation resulting from the digitalisation of a document is available. Many different applications currently exist, such as, check, form, mail or technical document processing. Whereas, online recognition systems are based on dynamic information acquired during the production of the handwriting. They require specific equipment, allowing the capture of the trajectory of the writing tool. Mobile communication systems (Personal Digital Assistant,

electronic pad, smart-phone) more and more integrate this type of interface and it is still important to improve the recognition performances for these applications while respecting strong constraints on the number of parameters to be stored and on the processing speed.

The first objective of this work is to optimize a neural network architecture less conventional than a Multi-Layer Perceptron (MLP) and to allow for a very great robustness with respect to deformations and disturbances. Accordingly, we opted for the study, the development and the test of a Convolutional Neural Network (CNN). Indeed, as stressed by the recent article [6], it presents remarkable properties to handle directly 2D patterns avoiding the subtle stage of the extraction of relevant features.

A second objective, within the framework of an online recognition system is to study the complementarities of the static and dynamic representation of a character [1]. Indeed, two different pen trajectories can correspond to the same graphic pattern and the same character class. In this case, the static representation will be more robust. Conversely, a given character can have distinct templates which are produced by very close movements, giving an advantage to the dynamic representation. We can expect in these conditions that an approach combining the two types of information will allow the improvement of the performances of recognition. Various experiments related to this combination are carried out in this work.

A modular architecture has been define. It allows for many possible configurations: basic MLP, CNN processing (either the static data or the on-line data) with or without a coupling stage at the output level or on the hidden layers.

## 2. CONVOLUTIONAL NEURAL NETWORKS

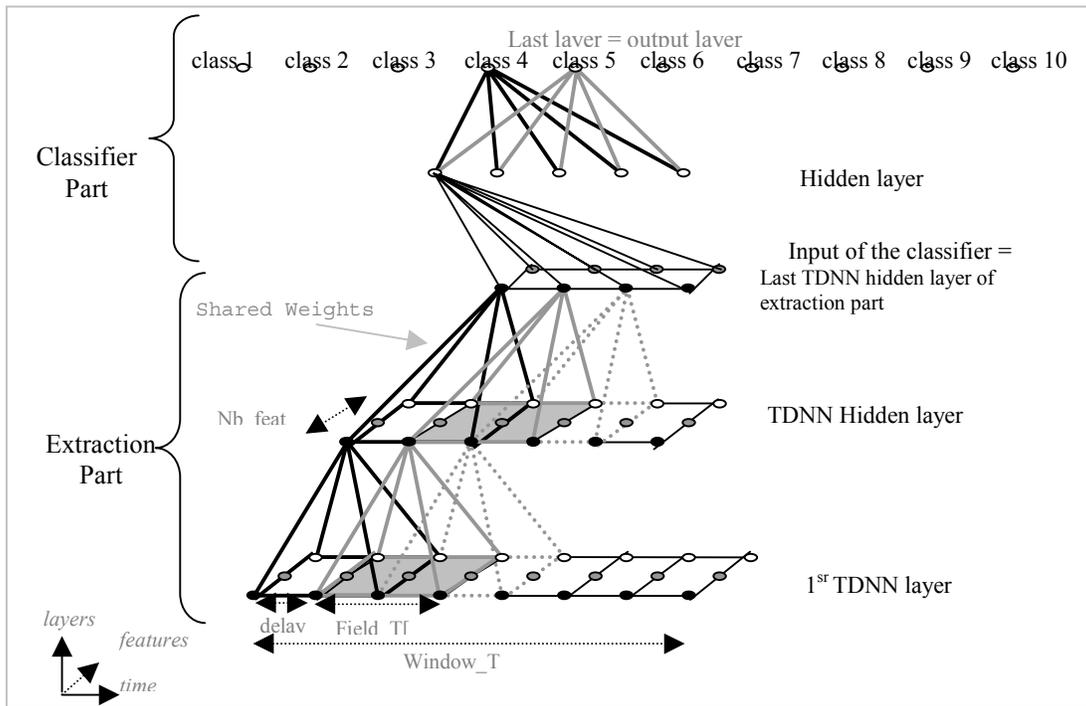


Figure 1 : TDNN architecture

The first important experiments on neural networks for handwriting recognition have been proposed in the late eighties [7]. The architecture of these networks was basically Multi-Layer Perceptron with back-propagation learning. More recently, Convolutional Neural Networks [4] have been derived from MLP, they incorporate important notions such as weight sharing and convolution receptive fields. In that sense, they are capable of a local, shift-invariant feature extraction process.

A perceptron has a fully-connected architecture, one of its main deficiencies is that the topology of the input is ignored : the input variables can be presented in any order without affecting the result of the training. For a CNN, a hidden neuron is connected to a subset of neurons from the preceding layer. It is the local receptive field for this neuron. Thus, each neuron can be seen as a specific local feature detector unit. Furthermore, the weight sharing constraint reduces the number of parameters in the system, facilitating thus the generalization process. This type of network has been applied successfully for digit recognition [4]. Two types of CNN are presented in the following sections. First, a TDNN which is used to process the on-line data, then, a SDNN is introduced to handle the off-line data.

### 2.1. A TDNN architecture

The TDNN, Time Delay Neural Network is a neural network with temporal shift which was first introduced for speech recognition [8]. It has since been transposed for

sequential data (see Penacée [4], LeNet5 [8]). It is thus particularly suited to process online handwriting signals. We have carefully defined the topology of the network : size of the receptive fields, number of layers, constraints on the weight sharing and also the learning algorithms (1st/2nd order) [9]. The selected architecture of the TDNN consists of two principal parts (see figure 1). The first, corresponding to the lower layers, implements the successive convolutions which enable it to gradually transform a sequence of feature vectors into another sequence of higher order feature vectors. The second part corresponds to a traditional MLP, it receives as input all the outputs of the extraction part.

We used online data (X, Y) from the Unipen [3] and IRONOFF [11] databases. They were resampled in order to avoid the influence of the pen speed and to obtain a fixed number of points per sample (50 points). Then, a preprocessing module extracts normalized features from each point: position (2), direction (2), curvature (1), pen status (1), for a total of 7 characteristics per point (see figure 2). Concerning learning, the network is trained with a traditional technique based on a stochastic gradient. It gives, according to our tests, results as good as a second order learning method.

Table 1 presents the comparative performances obtained with the best configurations for a MLP and a TDNN.

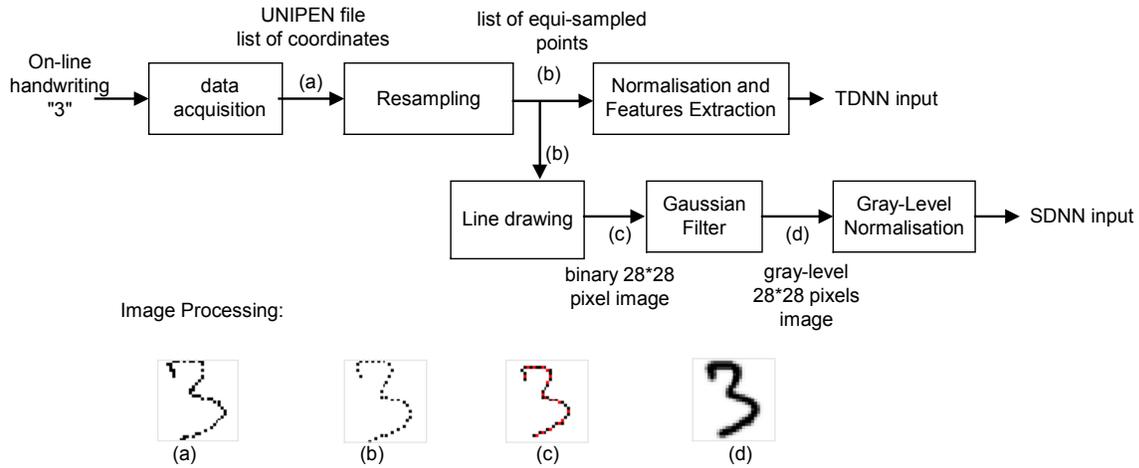


Figure 2: Preprocessing

	Learning set	Test set	TDNN %	MLP %
<b>UNIPEN database</b>				
10 Digits	10 423	5 212	<b>97,9</b>	<b>97,5</b>
26 Lowercase	34 844	17 423	<b>92,8</b>	<b>92,0</b>
26 Uppercase	17 736	8 869	<b>93,5</b>	<b>92,8</b>
<b>IRONOFF database</b>				
10 Digits	3 059	1 510	<b>98,4</b>	<b>98,2</b>
26 Lowercase	7 952	3 916	<b>90,7</b>	<b>90,2</b>
26 Uppercase	7 953	3 926	<b>94,2</b>	<b>93,6</b>

Table 1: TDNN and MLP Recognition Performances.

We can emphasize the significantly higher performances obtained by the TDNN on the three subsets: digits, lowercase and uppercase characters. Indeed, this allows it to decrease the error rate up to 16% on the digit set. In addition, the TDNN architecture requires less storage capacity due to its constraint on the weight sharing. For example, the number of coefficients reduces from 36,110 for a MLP (100 neurons on the hidden layer) to 17,930 for the TDNN-digit, (receptive field: 20, delay: 5, local features: 20, 100 hidden units for the classifier). This is a factor two reduction rate. Consequently, TDNN architecture presents real advantages for embedded applications. Moreover, it is established that with equal performances (same bias), the simpler a system is, the better its capacities of generalization (lower variance) are [2], known as the famous principle of the Occam's razor "Pluralitas non est ponenda sine necessitate".

## 2.2. The SDNN architecture

With the TDNN, the temporal nature of the data is exploited by the recognition system. It often allows the system to raise ambiguities and to identify more easily

some characters. On the other hand, some variations in the stroke ordering are disturbing. It concerns, for instance, the temporal position of diacritic marks or some post-retracing. In this case, the pictorial representation is more stable, and can be learned by a SDNN. Like the TDNN, the SDNN is a Convolutional Neural Network, it is a generalization of the TDNN to a 2D topology.

The meta parameters to be fixed for this network relate to the size of the receptive fields, the space shifts, the number of local features and the number of hidden layers for the extractor part and the classifier part. They were experimentally determined and the best compromise was obtained with two hidden layers, a convolutional window of 6\*6, a shift of 2, 20 local feature units and a linear classifier. These experiments were conducted on the MNIST offline isolated digit database [6]. The inputs of the network correspond to a 28\*28 image whose gray levels are normalized between [- 1,1].

Neural Networks	Number of Free Parameters	Learning recognition rate	Test recognition rate
MLP on pixels	159 010	99,4%	98,2%
MLP on features [10]	36 610	99,2%	98,6%
SDNN LeNet5 [6] (pixels)	60 000	-	99,05%
<b>Proposed SDNN (pixels)</b>	<b>18 370</b>	<b>99,9%</b>	<b>98,8%</b>

Table 2: Performances on MNIST database (Learning set: 60 000 digits, Test set : 10 000 digits)

The results (table 2) are on the same line as for the TDNN : on the one hand, the performances are a little bit higher than those with a MLP, while on the other hand, a

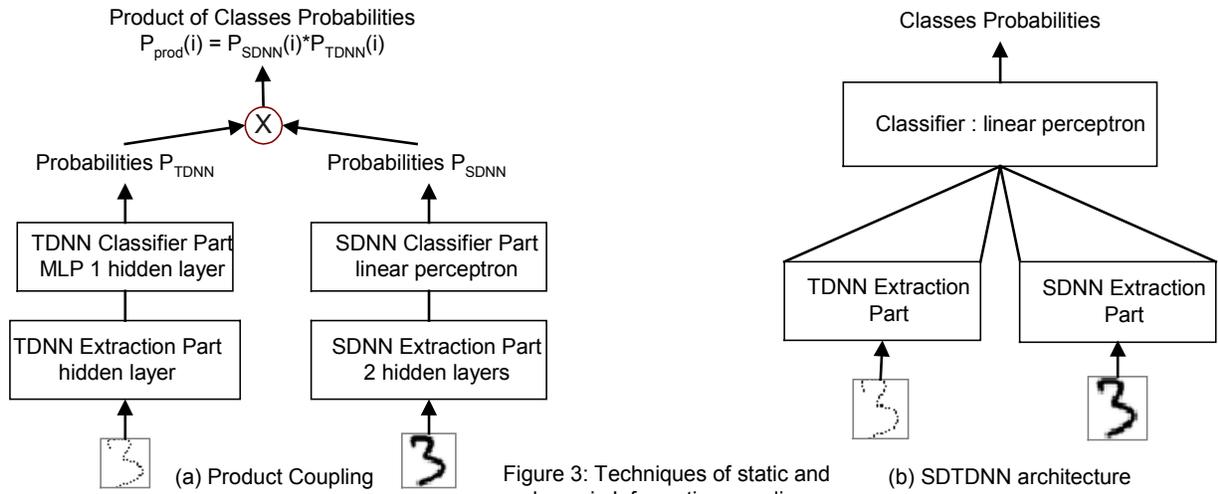


Figure 3: Techniques of static and dynamic Information coupling

significant reduction in the number of weights has been achieved, which is a major goal for portable applications with low storage capacities.

We want, in fact, to use this offline recognizer for data available originally as sequences of points (Unipen and IRONOFF databases). It is thus necessary to synthesize images from the pen trajectories. This transformation is obviously much easier than the reverse transformation [5], figure 2 illustrates the various stages of this pretreatment. We can consequently test the same databases as those used to validate the TDNN.

line information with respect to the static one where all ordering information has been lost. Secondly, the behaviors of the two recognizers are not fully correlated. For instance, one third of the failures of the TDNN are correctly recognized by the SDNN. As expected, these two recognizers complement each other.

	Learning set	Test set	SDNN %	MLP %
<b>UNIPEN Database</b>				
10 Digits	10 423	5 212	<b>95,4</b>	<b>94,4</b>
26 lowercase	34 844	17 423	<b>86,6</b>	<b>85,4</b>
26 uppercase	17 736	8 869	<b>89,5</b>	<b>87,5</b>
<b>IRONOFF Database</b>				
10 Digits	3 059	1 510	<b>94,3</b>	<b>91,8</b>
26 lowercase	7 952	3 916	<b>80,5</b>	<b>77,8</b>
26 uppercase	7 953	3 926	<b>89,9</b>	<b>87,1</b>

Table 3: SDNN and MLP Recognition Rate on Unipen and IRONOFF databases transformed in offline images.

### 3. TDNN AND SDNN CROSSED PERFORMANCES

TDNN and SDNN offer each very interesting recognition performances. It is interesting to study their respective behavior to estimate the potential profits that we can expect from a coupling of the two systems. Table 4 displays the cross-distribution of success (OK) and failure (KO) with respect to the two recognizers.

From this table, we can notice several interesting points. First, the recognizer exploiting the on-line data, the TDNN, outperforms (+2.4 %) the recognizer processing the off-line image. This confirms the superiority of the on-

KO	110 (2,1 %)	<b>38</b> (0,7 %)	72 (1,4 %)
OK	5 102 (97,9 %)	4 937 (94,8 %)	<b>165</b> (3,1 %)
TDNN / SDNN		4 975 (95,5 %)	237 (4,5 %)
Total : 5212 ex		OK	KO

Table 4: SDNN and TDNN cross evaluation on Unipen-digit database

### 4. COOPERATION OF ONLINE AND OFFLINE INFORMATION

Two coupling techniques have been tested. One at the output level, the other at the hidden layer level, see Fig 3.

#### 4.1. Combination at the output level

In this configuration, called “product coupling”, the final outputs are the product of the outputs of the TDNN and of the SDNN being separately trained. Consequently, it gives the geometrical mean of the posterior probabilities Prob(C|O) of the classes, obtained with the Softmax transfer function on the output units of each network.

Table 5 shows the interest of the product coupling technique. It allows the error rate to be reduced nearly 15% on the test Unipen digit database when compare to the best of the two recognisers, the recognition rate being increased from 97.9 % to 98.2 %. Among the examples

which were correctly classified (OK) by only one of the two classifiers (165+38 → 3.8 %), most of them are correctly classified by the product coupling (147 + 29 → 3.2 %). It remains only 0,5 % (18+9 → 0.5 %) which do not take advantage of the product coupling. Furthermore, some examples (72 → 1.4 %) which were not correctly classified by both recognisers are now correctly classified (3 → 0.1 %).

	OK TDNN OK SDNN	OK TDNN KO SDNN	KO TDNN OK SDNN	KO TDNN KO SDNN	Total
OK Product	4937 (94,8%)	147 (2,8%)	29 (0,5%)	3 (0,1%)	5 116 <b>(98,2%)</b>
KO Product	0	18 (0,3%)	9 (0,2%)	69 (1,3%)	96 (1,8%)
Total	4937 (94,8)	165 (3,1%)	38 (0,7%)	72 (1,4%)	5212

Table 5: Effect of the product coupling on UNIPEN digits.

#### 4.2. The SDTDNN

With the previous architecture, the combination module does not take advantage of the training which was done separately for each classifier. In order to integrate the training of the combination function, we built a multi modular architecture, called SDTDNN, for Space Displacement and Temporal Delay Neural Network. This structure (see fig 3.b) has a unique output layer which is fully connected to the concatenation of the hidden layers of both classifiers.

	TDNN 20/5/20 MLP 100		Product Coupling		SDTDNN	
Unipen	#Para meters	% Reco	#Para meters	% Reco	#Para meters	% Reco
Digit	17 930	97,9	36 300	98,2	13 392	97,9

Table 6: Compared SDTDNN performances

Up to now this architecture (SDTDNN = TDNN 10/2/20+SDNN 6/2/6 – 6/2/20+MLP lin). has reached the same level of performance as the TDNN alone but with fewer parameters. We believe that there is still room for improvement with this architecture. In fact, the trade-off is in favour of “product coupling” for the best recognition rate and in favour of the SDTDNN for minimizing the number of parameters of the system.

#### 5. CONCLUSION

We have presented here a new multi-modular architecture based on Convolutional Neural Networks

intended to be integrated on mobile systems of low capacities. We have demonstrated the superiority of online data with respect to offline data and that using both of them allows either an increase in the recognition performances or a decrease the classifier complexity in terms of memory requirement. These results show that this architecture offers a good compromise performance /complexity within the framework of concerned applications. We think that it is still possible to improve this compromise and to consider the extent of its use to an online cursive words recognition system.

#### 6. REFERENCES

- [1] F. Alimoglu, E. Alpaydin, “Combining Multiple Representations and Classifiers for Pen-based Handwritten Digit Recognition”, *ICDAR’97*, pp. 637-660, Ulm-Allemagne, 1997.
- [2] C.M. Bishop, “Neural Networks for Pattern Recognition”, *Oxford University Press*. ISBN0-19-853849-9, p 116-161, 1995.
- [3] I. Guyon, L. Schomaker, S. Janet, M. Liberman , and R. Plamondon, “First UNIPEN benchmark of on-line handwriting recognizers organized by NIST”. *Technical Report* BL0113590-940630-18TM, AT&T Bell Laboratories, 1994.
- [4] I. Guyon, J. Bromley, N. Matic, M. Schenkel, H. Weissman, “Penacee: A Neural Net System for Recognizing On-line Handwriting”, In E. Domany, J. L. van Hemmen, and K. Schulten, editors, *Models of Neural Networks*, volume 3, pp. 255-279, Springer, 1995.
- [5] P-M. Lallican, C. Viard-Gaudin, S. Knerr, « From Off-line to On-line Handwriting Recognition », *IWFHR’2000*, Amsterdam, Netherlands, pp. 303-312, September 11-13, 2000.
- [6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Intelligent Signal Processing*, pp. 306-351, 2001.
- [7] Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, “Handwritten digit recognition with a backpropagation neural network”. In D. Touretzky editor, *Advances in Neural Information Processing Systems 2*, pp. 396-304, 1990.
- [8] Y. LeCun and Y. Bengio, "Convolutional Networks for Images, Speech, and Time-Series," in *The Handbook of Brain Theory and Neural Networks*, (M. A. Arbib, ed.), 1995.
- [9] E. Poisson, C. Viard-Gaudin, « Réseaux de neurones à convolution : reconnaissance de l'écriture manuscrite non contrainte », *Valgo 2001* (ISSN 1625-9661), N° 01-02, 2001.
- [10] Y.H. Tay, “Off-line Handwriting Recognition using artificial Neural Network and Hidden Markov Model”- PhD University of Nantes and University Teknologi Malaysia, 2002.
- [11] C. Viard-Gaudin, P.M. Lallican, S. Knerr, P. Binter, "The IRESTE ON-OFF (IRONOFF) Handwritten Image Database", *ICDAR’99*, pp. 455-458, Bangalore, 1999.