Shallow Neural Network Model for Hand-drawn Symbol Recognition in Multi-Writer Scenario

Sounak Dey, Anjan Dutta, Josep Lladós, Alicia Fornés and Umapada Pal
Computer Vision Center, Computer Science Department,Universitat Autònoma de Barcelona, Barcelona, Spain
Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India
Email: {sdey, adutta, josep, afornes}@cvc.uab.es, umapada@isical.ac.in

***Abstract*—One of the main challenges in hand drawn symbol recognition is the variability among symbols because of the different writer styles. In this paper, we present and discuss some results recognizing hand-drawn symbols with a shallow neural network. A neural network model inspired from the LeNet architecture has been used to achieve state-of-the-art results with very less training data, which is very unlikely to the data hungry deep neural network. From the results, it has become evident that the neural network architectures can efficiently describe and recognize hand drawn symbols from different writers and can model the inter author aberration.**

I. Introduction

Symbol recognition is one of the central topics of Graphics Recognition [3]. A lot of effort has been made in the last decade to develop good symbol and shape recognition methods inspired in either structural or statistic pattern recognition approaches. The presence of hand drawn symbols increases the difficulty of classification: there is a high variability in writing style, with different sizes, shapes and intensities, increasing the number of touching and broken symbols. In addition, working with old documents even increases the difficulties in these stages because of paper degradation and the frequent lack of a standard notation.

Symbol recognition in graphical document images can be seen as a particular case of shape recognition. The proposed methods for recognizing symbols can be roughly divided: (1) structural, (2) statistical. Structural methods usually use graph based representation to achieve the goal [3]. The statistical methods formulate the symbol recognition problem as a particular case of shape recognition. Their major focus is the expressiveness and compactness in the shape description. Zhang and Lu [10] reviews the main techniques used in this field, mainly classified in contour-based descriptors (i.e. polygonal approximations, chain code, shape signature, and curvature scale space) and region-based descriptors (i.e. Zernike moments, ART, and Legendre moments [7]). A good statistical shape descriptor should guarantee inter-class separability and intra-class similarity, even when describing noisy and distorted shapes. It has been proved that some descriptors, robust with some affine transformations and occlusions in printed symbols, are not efficient enough for handwritten symbols.

The success of convolutional neural network (CNN) [5] in different fields of computer vision, *i.e.*, natural language processing [2], handwriting recognition [8], semantic image segmentation [6] etc has increased interests on extending these frameworks to other domains. This fact motivates us to investigate with the neural network models for writer dependent symbol recognition problem. In this paper, we design a shallow neural network which is inspired by the previously proposed LeNet model [5], and present results and discussions with respect to the symbol recognition problem in multi writer scenario. For experiments, we have used the NicIcon dataset [9] which contains multi writer hand drawn symbols.

The rest of the paper is organized as follows: in Section II, we describe the shallow network model. Experimental results obtained by the shallow network model are discussed in Section III. In Section IV, we made the conclusions and some ideas for future research work have been discussed.

II. Shallow network for graphics recognition

We get motivated by the LeNet model [5], which is known to work well on digit classification tasks. A slightly different version from the original LeNet implementation is used by replacing the sigmoid activations with Rectified Linear Unit (ReLU) activations for the neurons (1). We also replace the softmax layer with log softmax, such that probability of being one class will not be too small.

$$f(x) = max(0, x) \tag{1}$$

The design of our network model contains the essence of CNNs that are still used in larger models. In general, it consists of a convolutional layer followed by a pooling layer, another convolution layer followed by a pooling layer, and then two fully connected layers similar to the conventional multilayer perceptrons. A 2-dimensional input image is fed into two sets of convolutional and pooling layers. This output is then fed to a fully connected layer and a softmax classifier. Compared to multilayer, fully connected perceptrons, our modified LeNet model can recognize images better. This is due to the following three properties of the convolution:

- The 3D nature of the neurons: a convolutional layer is organized by width, height and depth. Neurons in each layer are connected to only a small region in the previous layer. This region is called the receptive field.
- Local connectivity: A CNN utilizes the local space correlation by connecting local neurons. This design guarantees that the learned filter has a strong response to local input features. Stacking many such layers generates

a non-linear filter that is more global. This enables the network to first obtain good representation for small parts of input and then combine them to represent a larger region.

- Weight sharing: In a CNN, computation is iterated on shared parameters (weights and bias) to form a feature map. This means that all the neurons in the same depth of the output respond to the same feature. This allows the network to detect a feature regardless of its position in the input. In other words, it is shift invariant.

For the classification problem at hand we use the negative log likelihood loss (2) because we had an unbalanced training set for very less images. This can also be thought of as the multi-class cross-entropy loss.

$$C = -\frac{1}{n}\sum_{n} \ln(a_y^L) \tag{2}$$

In the above equation a is the neuron's output and y is the corresponding desired output. The detailed specifications of the architecture is provided in Table I and Table II

TABLE I
OVERVIEW OF OUR SHALLOW NETWORK ARCHITECTURE

Layer	Size	Parameters
Convolution	$10 \times 5 \times 15$	stride $= 1$
Pooling	$10 \times 2 \times 2$	stride $= None$
Convolution	$20 \times 5 \times 5$	stride $= 1$
Pooling + Dropout	$20 \times 2 \times 2$	stride $= None, p = 0.5$
Fully Connected + Dropout	4096	$p = 0.7$
Fully Connected	14	

TABLE II
TRAINING HYPER-PARAMETERS

Parameter	Value
Initial Learning Rate (LR)	0.01
Learning Rate Decay	0
Weight Decay	0
Momentum	0.5
Batch Size	100

III. EXPERIMENTAL RESULTS AND DISCUSSION

We have tested our method for shape recognition tasks, and for this purpose, we have used the NicIcon datasets [9]. This dataset is composed of $26,163$ handwritten symbols of 14 classes from 34 different writers and it is commonly used for online symbol recognition, but offline data is also available. The dataset is already divided into three subsets (training, validation and test) for both writer dependent and independent settings. Depending on the setting, $9,300$, $6,200$ and $10,700$ symbols are contained in the training, validation and test sets, respectively. We have extracted individually every symbol from the scanned forms, and then binarized and scale-normalized in an image of 100×100 pixels. In Table III, we provide the results obtained by our network model, which shows that our model outperforms the many of the existing approaches in the writer independent(WI) scenario.

TABLE III
ACCURACY RATE (%) COMPARISON OF DIFFERENT METHODS ON THE NICICON DATASET [9].

Method	BSM [4]	nrBSM [1]	NRAM+nrBSM [1]	ShallowNet
NicIcon WI	90.02	91.09	95.18	**97.01**

IV. CONCLUSIONS

In this paper we have used a shallow neural network model for graphical symbol recognition in multi-writer scenario. The main advantage of this network is the lesser number of parameters, which is quite simple to learn. We also show unlike many convolutional neural network (CNN) we can train a shallow CNN with unbalanced small dataset. Additionally the results also demonstrate the capacity of the network to capture the structure of the shape and deal with large deformations. The proposed deep learning model outperforms many state-of-the-art methods that usually work with hand crafted descriptors. In the future, it would be interesting to use the proposed network model hand drawn architectural symbol classification task.

ACKNOWLEDGMENT

This work has been partially supported by the European Union's research and innovation program under the Marie Skłodowska-Curie grant agreement No. 665919, the Spanish project TIN2015-70924-C2-2-R, the PIF grant supported by Univeristat Autónoma de Barcelona (PIF contract 10/2016-09/2019). The Titan X Pascal and the Titan Xp used for this research was donated by the NVIDIA Corporation.

REFERENCES

[1] J. Almazán, A. Fornés, and E. Valveny, "A non-rigid appearance model for shape description and recognition," *PR*, vol. 45, no. 9, pp. 3105–3113, 2012.

[2] A. Conneau, H. Schwenk, L. Barrault, and Y. LeCun, "Very deep convolutional networks for natural language processing," *CoRR*, vol. abs/1606.01781, 2016.

[3] P. Dosch and E. Valveny, "Report on the second symbol recognition contest," in *GREC*, 2006, pp. 381–397.

[4] S. Escalera, A. Fornés, O. Pujol, P. Radeva, G. Sánchez, and J. Lladós, "Blurred shape model for binary and grey-level symbol recognition," *PRL*, vol. 30, no. 15, pp. 1424–1433, 2009.

[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015, pp. 3431–3440.

[7] P. Salembier and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*. John Wiley & Sons, Inc., 2002.

[8] S. Sudholt and G. A. Fink, "Phocnet: A deep convolutional neural network for word spotting in handwritten documents," in *ICFHR*, 2016, pp. 277–282.

[9] D. Willems, R. Niels, M. van Gerven, and L. Vuurpijl, "Iconic and multi-stroke gesture recognition," *PR*, vol. 42, no. 12, pp. 3303 – 3312, 2009.

[10] D. Zhang and G. Lu, "Review of shape representation and description techniques," *PR*, vol. 37, no. 1, pp. 1–19, 2004.