

# A Cascade of Deep Learning Fuzzy Rule-based Image Classifier and SVM

Plamen Angelov, *Fellow, IEEE* and Xiaowei Gu, *Student Member, IEEE*

Data Science Group,  
School of Computing and Communications,  
Lancaster University, Lancaster, UK  
E-mail: {p.angelov, x.gu3}@lancaster.ac.uk

**Abstract**—In this paper, a fast, transparent, self-evolving, deep learning fuzzy rule-based (DLFRB) image classifier is proposed. This new classifier is a cascade of the recently introduced DLFRB classifier and a SVM based auxiliary. The DLFRB classifier serves as the main engine and can identify a number of human interpretable fuzzy rules through a very short, transparent, highly parallelizable training process. The SVM based auxiliary plays the role as a conflict resolver when the DLFRB classifier produces two highly confident labels for a single image. Only the fundamental image transformation techniques (rotation, scaling and segmentation) and feature descriptors (GIST and HOG) are used for pre-processing and feature extraction, but the proposed approach significantly outperforms the state-of-art methods in terms of both time and precision. Numerical experiments based on a handwriting digits recognition problem are used to demonstrate the highly accurate and repeatable performance of the proposed approach after a very shorting training process.

**Keywords**—*deep learning; cascade; fuzzy rule-based classifier; SVM; handwriting digits recognition.*

## I. INTRODUCTION

In the recent years, human societies have witnessed the amazing successes of deep convolutional neural networks (DCNNs) at image classification problems [1]–[7]. Composed by a large number of linear and nonlinear transformation techniques, the DCNNs are able to extract high-level information from the images and perform highly accurate classification results. Being regarded as the best solution for computer vision problems, the deep convolutional neural networks have attracted lots of attentions as well as publicity [8].

Nonetheless, the DCNN architectures still have a number of unsolved questions and deficiencies. Decisions on the types of convolution kernels used in DCNNs are always *ad hoc* without clear evidence about the effectiveness of those kernels [3]–[5]. The extracted features by the DCNNs are not human interpretable, and the training process is also opaque. In addition, some publications have shown that the DCNNs are easily fooled by many unrecognizable images with near-certainty as members of a recognizable class [9]. The deficiencies of the DCNN include the requirement of huge

amounts of time and resource consumptions, and the unparallelizable training process which is usually off-line [1]–[6].

In this paper, we propose a new deep learning approach for image classification. This approach consists of the recently introduced Multi-layer Multi-model Images Classifier Ensemble (MICE) [10] and a SVM based conflict resolution classifier. The MICE is a deep learning fuzzy rule-based (DLFRB) classifier, it acts as the main engine in the proposed approach and conducts the majority of classification tasks. The DLFRB classifier is able to extract a number of highly interpretable AnYa type 0-order fuzzy rules [11] through a very short, transparent, highly parallelizable training process. The SVM based classifier serves as the auxiliary in the proposed approach. It will assist the DLFRB classifier in making decisions when there is a conflict, which means that the DLFRB classifier produces two highly confident scores for one image.

The proposed approach is highly efficient and human interpretable, it is free from the *ad hoc* decisions and user- and problem-specific parameters. It is able to recognize the handwriting digits with the currently best accuracy (achieved without elastic distortion) [6]. Numerical experiments clearly demonstrate that the proposed approach outperforms the state-of-art approaches.

## II. ARCHITECTURE OF THE PROPOSED APPROACH

The architecture of the proposed approach is depicted in Fig. 1. As we can see from the figure, the recently introduced DLFRB classifier, named MICE, is used as the main engine of the proposed approach and a SVM based conflict resolution classifier is used to support the main engine when there is a conflict in the degrees of confidence.

The proposed approach only employs the fundamental rotation, scaling and segmentation techniques for image pre-processing. The elastic distortion technique [7], which can significantly improve the generalization ability and recognition accuracy of the DCNNs [2], [4] are not used in this paper because the elastic distortion will bring randomness to the algorithm and inhibit the reproducibility [7].

The popular GIST [12] and HOG [13] descriptors are used in the proposed approach for feature extraction. Modifications

are introduced in this paper to further improve their effectiveness.

As being introduced in [10], the DLFRB classifier is able to perform highly accurate classification on the handwriting digits in the majority cases by following the “winner takes all” principle. However, it fails to react properly in the rare cases (less than 2% in the MNIST handwriting digits recognition problem [14]) in which there are two highly confident labels generated for a single image at the same time. In those rare cases, the “winner takes all” principle the DLFRB classifier follows is not an effective strategy and can easily lead to mistakes.

Therefore, in the proposed approach, a conflict resolution classifier is added as the auxiliary for the main classifier in making decisions when there are two highly confident labels produced for the same image. This conflict resolution classifier is built upon the SVM classifier [15] with polynomial kernel and it effectively improves the overall performance of this approach.

The learning process of the SVM is independent from the MICE, and thus, can be trained in parallel and will not influence the evolving nature of the proposed approach. The details of the proposed approach (MICE and the SVM based auxiliary) will be described in the following sections.

### III. MULTI-LAYER MULTI-MODEL IMAGES CLASSIFIER ENSEMBLE

As it was stated in section II, the main engine of the proposed approach is building upon the recently introduced DLFRB classifier [10]. On the basis of the original design, we add a new segmentation layer to the DLFRB classifier to improve the learning efficiency of the approach. In this section, we will briefly describe the DLFRB classifier of the proposed approach.

#### A. Training Stage

The architecture of the DLFRB classifier in the training stage is depicted in Fig. 2. From the figure we can see, the DLFRB classifier consists of the following components [10]:

1. Normalization layer, which applies linear normalization to fit the original pixel value range of  $[0, 255]$  into the range of  $[0, 1]$ .

2. Scaling layer, which resizes the training images from their original size of  $28 \times 28$  into 7 ( $S=7$ ) different sizes: 1)  $28 \times 22$ , 2)  $28 \times 24$ , 3)  $28 \times 26$ , 4)  $28 \times 28$ , 5)  $28 \times 30$ , 6)  $28 \times 32$  and 7)  $28 \times 34$ .

3. Rotation layer, which rotates the images at certain angle starting from -15 degrees going through 0 degree up to 15 degrees with an interval of 3 degrees. Therefore, 11 ( $R=11$ ) new images can be obtained from 1 image after rotation.

The scaling and rotation layers create 77

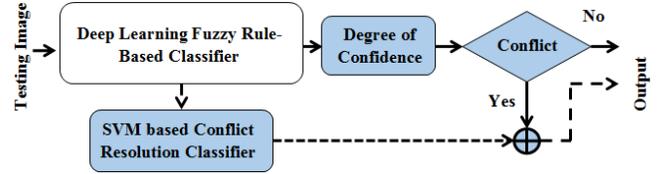


Fig.1. The diagram of the proposed approach

( $SR = 77$ ) new training sets from the original one with respect to different scaling sizes and rotation degrees.

4. Segmentation layer, which is newly added as an extension of the recently introduced DLFRB classifier (highlighted in blue). This layer is for extracting the central area ( $22 \times 22$ ) from the training images. It discards the borders which mostly consists of the white pixels with little or no information.

5. Feature descriptors. In the DLFRB classifier, we use the two commonly used descriptors: GIST [12] and HOG [13] for global feature extraction. We also introduce the following modifications to further improve the effectiveness of the extracted features for the learning process [10]:

$$\begin{cases} \mathbf{g} = \frac{G(\text{Image})}{\|G(\text{Image})\|} \\ \mathbf{h} = \frac{\kappa(1-H(\text{Image}))}{\|\kappa(1-H(\text{Image}))\|} \end{cases} \quad (1)$$

where  $\mathbf{g}$  is the  $1 \times 512$  dimensional modified GIST feature of the image used in this approach;  $\mathbf{h}$  is the  $1 \times 576$  dimensional modified HOG feature;  $G(\cdot)$  and  $H(\cdot)$  denote the GIST and HOG feature descriptors same as in [12], [13], respectively;  $\|\cdot\|$  denotes the Euclidean norm;  $\kappa(\cdot)$  is a nonlinear mapping function [16]:

$$\kappa(x) = \text{sgn}(x) \left[ \exp \left[ (1 + \text{sgn}(x)x)^2 \right] - \exp(1) \right] \quad (2)$$

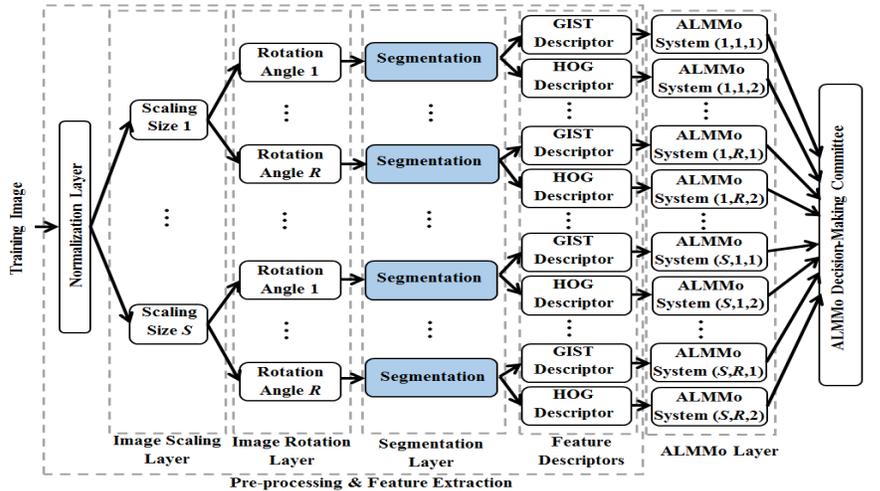


Fig.2. Architecture of the DLFRB classifier (training stage)

here  $\text{sgn}(\cdot)$  is the well-known sign function.

The GIST feature descriptor used in the proposed approach follows the default setting as described in [12]; the  $4 \times 4$  size patch is used for the HOG feature descriptor.

6. ALMMo layer. The recently introduced Autonomous Multiple Learning Model (ALMMo) system [17], [18] is employed as the learning engine to extract the local maxima from the training samples and based on these, generates AnYa type 0-order fuzzy rules [11]. The details of the learning process of the ALMMo system are described in [10], [18]. In this paper, we will focus on the identified fuzzy rules through the learning process.

After the training process, each ALMMo system will generate 10 (1 per class/digit) AnYa type fuzzy rules in the following form [11]:

$$\begin{aligned}
 & \text{IF} \left( \text{Image} \sim \text{Prototype}_{1}^{i,j} \right) \\
 & \text{OR} \left( \text{Image} \sim \text{Prototype}_{2}^{i,j} \right) \\
 & \quad \vdots \\
 & \text{OR} \left( \text{Image} \sim \text{Prototype}_{n^{i,j}}^{i,j} \right) \\
 & \text{THEN} \left( \text{Image presents digit "j"} \right)
 \end{aligned} \tag{3}$$

where  $j = 0, 1, 2, \dots, 9$ , which corresponds to digits "0" to "9";

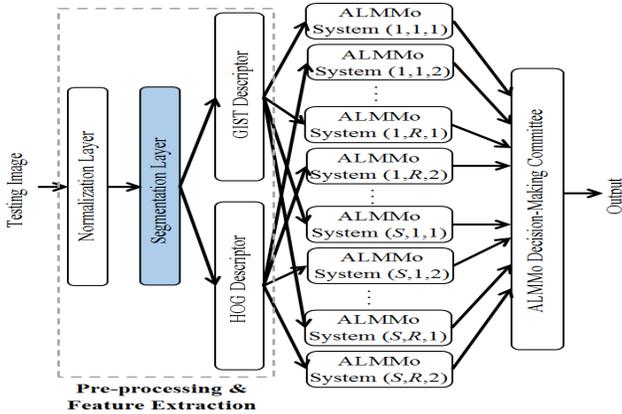


Fig.3. Architecture of the DLFRB classifier (classification stage)

TABLE I. ILLUSTRATIVE EXAMPLES OF ANYA TYPE FUZZY RULES

#	Fuzzy Rules
0	<i>IF</i> (Image ~ 0) <i>OR</i> (Image ~ 0) <i>THEN</i> (Image is "0")
1	<i>IF</i> (Image ~ 1) <i>OR</i> (Image ~ 1) <i>THEN</i> (Image is "1")
2	<i>IF</i> (Image ~ 2) <i>OR</i> (Image ~ 2) <i>THEN</i> (Image is "2")
3	<i>IF</i> (Image ~ 3) <i>OR</i> (Image ~ 3) <i>THEN</i> (Image is "3")
4	<i>IF</i> (Image ~ 4) <i>OR</i> (Image ~ 4) <i>THEN</i> (Image is "4")
5	<i>IF</i> (Image ~ 5) <i>OR</i> (Image ~ 5) <i>THEN</i> (Image is "5")
6	<i>IF</i> (Image ~ 6) <i>OR</i> (Image ~ 6) <i>THEN</i> (Image is "6")
7	<i>IF</i> (Image ~ 7) <i>OR</i> (Image ~ 7) <i>THEN</i> (Image is "7")
8	<i>IF</i> (Image ~ 8) <i>OR</i> (Image ~ 8) <i>THEN</i> (Image is "8")
9	<i>IF</i> (Image ~ 9) <i>OR</i> (Image ~ 9) <i>THEN</i> (Image is "9")

$n^{i,j}$  is the number of prototypes in the  $j^{\text{th}}$  fuzzy rule of the  $i^{\text{th}}$  ALMMo system. The 10 fuzzy rules are trained independently and there is no interaction between each other, which means they can be trained in parallel as well. Moreover, the identified fuzzy rules enable the strong transparency of the learning process, which is the major difference compared with the state-of-art DCNNs.

In the proposed approach, there are 154 ( $2SR=154$ ) ALMMo systems identified from the expanded training sets, 77 of them are trained with the GIST features and the others are trained with the HOG features. Therefore, in total, 1540 AnYa type fuzzy rules are identified. The identified fuzzy rules will play a pivotal role in the classification stage [10].

Illustrative examples of AnYa fuzzy rules identified by a single ALMMo system are visualized in Table I [10], [11]. They clearly demonstrate the advantage of the DLFRB classifier-its transparency.

7. Decision-making committee, which is of critical important in generating the final output. The operating mechanism of the decision-making committee will be described in more detail in the next subsection.

### B. Classification Stage

The architecture of the proposed DLFRB classifier in the classification stage is depicted in Fig. 3. As we can see from the figure, the first 3 layers (normalization layer, segmentation layer and feature descriptor layer) are used for pre-processing and feature extraction, the last 2 layers (ALMMo layer and decision-making committee) are for classification.

After the feature extraction, the modified GIST and HOG features are sent to the trained ALMMo system. Each ALMMo has 10 fuzzy rules, and each rule will give its output

as the score of confidence based on the “winner takes all” principle [10]:

$$\lambda_{i,j}(\mathbf{x}) = \arg \max_{k=1,2,\dots,i^{i,j}} [\mu_k^{i,j}(\mathbf{x})] \quad (4)$$

where  $\mathbf{x} = \mathbf{g}$  or  $\mathbf{h}$  is the global feature of the testing image;

$$\mu_k^{i,j}(\mathbf{x}) = \exp\left[-\|\mathbf{x} - \mathbf{p}_k^{i,j}\|^2\right]; \quad \mathbf{p}_k^{i,j}$$

is the corresponding global feature of the  $prototype^{i,j}$ ;  $j = 0, 1, 2, \dots, 9$ ;  $i = 1, 2, 3, \dots, 2SR$ .

For a better understanding, Table II is used to illustrate the process of generating the score of confidence of each fuzzy rule as tabulated in Table I based on the GIST features of the images, where  $\mu_1 \sim \mu_6$  corresponds to the scores generated based on 6 prototypes in each fuzzy rule;  $\lambda$  represents the score of confidence of each rule.

Then, every ALMMo system passes its scores of confidence corresponding to the 10 digits to the decision-making committee and the committee integrates the outputs into 10 overall scores of confidence [10]:

$$C_j^M(Image) = \frac{1}{2} \left( \frac{1}{SR} \sum_{i=1}^{SR} \lambda_{i,j}(\mathbf{g}) + \max_{i=1,2,\dots,SR} (\lambda_{i,j}(\mathbf{g})) \right) + \frac{1}{2} \left( \frac{1}{SR} \sum_{i=1}^{SR} \lambda_{i,j}(\mathbf{h}) + \max_{i=1,2,\dots,SR} (\lambda_{i,j}(\mathbf{h})) \right) \quad (5)$$

In general cases, the overall decision is made through the “winner takes all” principle by the committee as [10]:

$$Label = \arg \max_{j=0,1,\dots,9} (C_j^M(Image)) \quad (6)$$

However, in some rare cases, the highest and the second highest overall scores of confidence given by the decision-making committee are very close, which means there is a conflict. In these cases, the committee will not apply the “winner takes all” principle to decide the label, instead, it will involve the SVM based conflict resolution classifier for assistance.

In this paper, if the following condition (equation (7)) is met, the committee will involve the SVM based conflict resolution classifier for help. The way that SVM based conflict resolution classifier participates in the decision-making process will be described in the next section.

$$IF \left( C_1^{*M}(Image) < C_2^{*M}(Image) + \frac{\sigma_\lambda}{4} \right) \quad (7) \\ THEN (External Support is needed)$$

TABLE II. SCORES OF CONFIDENCE GIVEN BY THE FUZZY RULES OF EACH CLASS

Unlabelled Image	Fuzzy Rules #	$\mu_1$	$\mu_2$	$\mu_3$	$\mu_4$	$\mu_5$	$\mu_6$	$\lambda$	Label
	0	0.4752	<b>0.5335</b>	0.5128	0.4582	0.4694	0.4224	0.5335	4
	1	0.3980	<b>0.5013</b>	0.4410	0.3880	0.3891	0.4067	0.5013	
	2	0.4932	<b>0.5834</b>	0.5490	0.4925	0.5601	0.5188	0.5834	
	3	<b>0.5800</b>	0.4898	0.5032	0.4912	0.4868	0.4770	0.5800	
	4	0.7484	<b>0.8638</b>	0.7616	0.5661	0.6490	0.6406	<b>0.8638</b>	
	5	0.4240	0.5078	0.4653	<b>0.5990</b>	0.4898	0.4867	0.5990	
	6	0.4972	0.4985	<b>0.5508</b>	0.4828	0.4757	0.4622	0.5508	
	7	0.4906	0.5826	0.5996	0.3858	<b>0.6048</b>	0.4159	0.6048	
	8	<b>0.4961</b>	0.4730	0.4650	0.4681	0.4628	0.4697	0.4961	
	9	0.6221	0.5764	0.5972	0.6253	<b>0.6463</b>	0.5135	0.6463	
	0	0.4217	0.4221	0.3944	<b>0.4327</b>	0.4047	0.3472	0.4327	7
	1	0.4597	<b>0.4811</b>	0.4607	0.4683	0.4431	0.4337	0.4811	
	2	0.5529	0.5102	0.5334	<b>0.5545</b>	0.4783	0.5287	0.5545	
	3	<b>0.6079</b>	0.5153	0.5560	0.5554	0.5328	0.5299	0.6079	
	4	0.5648	0.6371	<b>0.7444</b>	0.5473	0.6041	0.5607	0.7444	
	5	0.4642	0.4539	0.4533	0.5840	<b>0.5863</b>	0.5264	0.5863	
	6	0.4808	0.4746	<b>0.5279</b>	0.5086	0.4732	0.4928	0.5279	
	7	0.5860	0.6406	<b>0.9205</b>	0.4618	0.7098	0.5296	<b>0.9205</b>	
	8	0.5484	0.5287	0.5018	<b>0.5668</b>	0.5292	0.4708	0.5668	
	9	<b>0.6849</b>	0.6359	0.6384	0.6375	0.6507	0.5520	0.6849	

where  $C_j^{*M}(Image)$  ( $j = 0, 1, \dots, 9$ ) is the ranked overall scores of confidence in the descending order;  $\sigma_\lambda$  is the standard deviation of  $C_j^M(Image)$  ( $j = 0, 1, \dots, 9$ ).

#### IV. THE SVM BASED CONFLICT RESOLUTION CLASSIFIER

In the proposed approach, a SVM based conflict resolution classifier is added to assist the DLFRB classifier when it produces two highly confident labels on one image. In this section, we will describe the SVM based conflict resolution classifier in detail.

##### A. Training Stage

The structure of the SVM based conflict resolution classifier is depicted in Fig. 4. The classifier consists of the following components:

1. Normalization layer;
2. Segmentation layer;
3. Feature descriptors;
4. Feature integration layer and
5. SVM classifier with polynomial kernel.

The first 3 layers are the same as the ones introduced in section III.

The feature integration layer is for integrating the GIST and HOG features together for training purposes. The feature integration is done by combining the HOG and GIST features

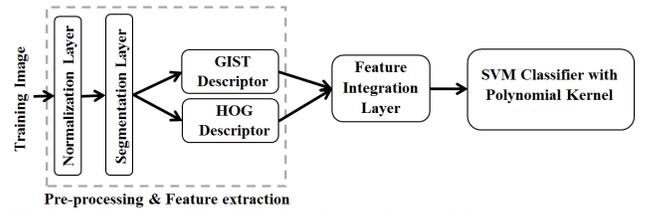


Fig.4. Architecture of the SVM based extension (training stage)

of the same image into a  $1 \times 1088$  dimensional vector  $[h, g]$ .

The SVM classifier was firstly introduced by Vapnik in 1995 [15]. The SVM classifier is currently one of the most popular classification approaches and is able to produce the state-of-art results in many problems. In the proposed approach, we use the SVM with 5-order polynomial kernel as the learning engine. We train the SVM classifier with the features extracted from the original training set only because:

- i) The training speed of the SVM classifier deteriorates significantly for large-scale datasets;
- ii) The training process of the SVM classifier cannot be parallelized;
- iii) The SVM classifier does not support online learning.

The SVM based conflict resolution classifier is fully independent from the MICE network and, thus, can be trained in parallel based on the global features extracted from the training images.

### B. Classification Stage

During the classification stage, the SVM based conflict resolution classifier will not be functioning unless being requested by the DLFRB classifier. Once the condition for external help is triggered, the DLFRB classifier will deliver the GIST and HOG features of the testing image as well as the two highly confident scores of confidence to the SVM based extension.

Then, the SVM layer will conduct a two-class classification based on the two potential classes the testing image belonging to and generate two scores of confidence, denoted by  $C_1^{*S}(Image)$  and  $C_2^{*S}(Image)$ .

The label of the testing image is decided by the following equation:

$$Label = \arg \max_{j=1,2} (C_j^{*M}(Image) + C_j^{*S}(Image)) \quad (8)$$

## V. EXPERIMENTAL DEMONSTRATION AND DISCUSSION

In this section, we will demonstrate the experimental results obtained with the MNIST dataset [14] and compare the proposed approach with the state-of-art approaches reporting the current best results (with and without elastic distortion). All the experiments are conducted on MATLAB R2015a platform using a PC with dual core i7 processor with clock

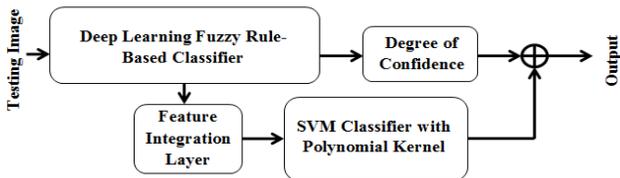


Fig.5. Architecture of the SVM based extension (classification stage)

TABLE III. TIME (IN SEC.) CONSUMPTION FOR THE LEARNING PROCESS OF EACH PARTS

Fuzzy Rule #	1	2	3	4	5	6	7	8	9	10	SVM	
Digital	"0"	"1"	"2"	"3"	"4"	"5"	"6"	"7"	"8"	"9"		
Feature	GIST	52.26	43.89	57.72	62.26	50.24	48.13	51.58	50.62	57.79	52.62	94.25
	HOG	74.52	59.09	75.38	92.86	72.48	66.94	71.61	73.59	80.40	71.80	

frequency 3.6GHz each, 16GB RAM and Window 10 operation systems.

As it has been described above, the proposed approach consists of two components: i) the DLFRB classifier as the main engine and ii) a SVM based conflict resolution classifier.

For the DLFRB classifier, the original training set is expanded into 77 different training sets using scaling and rotation. Based on the GIST and HOG features extracted from those training sets, 154 ALMMo systems are trained in parallel and 1540 AnYa type fuzzy rules are identified in total (the simplified illustration of those fuzzy rules are visualized in Table I). The SVM based external extension is trained using the features extracted from the original training set.

The DLFRB classifier is able to decide the labels of 9850 testing images with full confidence, 9825 out of which are classified correctly (accuracy is 99.75%). The incorrectly classified handwriting digits are presented in Fig. 6. 4 of the errors are "2"; 2 of them are "3", 3 of them are "4", 3 of them are "5", 5 of them are "6", 1 of them is "7", 4 of them are "8", 3 of them are "9".

With the assistance of the SVM based extension, there are 130 images correctly classified from the 150 images with dual candidate labels. The 20 errors are depicted in Fig.7. Among the 20 errors, there are 3 images that are not recognizable, which means that the two candidate labels fail to include the true label. The 3 images are marked by the red rectangles in Fig. 7. Therefore, the overall accuracy of the proposed approach is 99.55%.

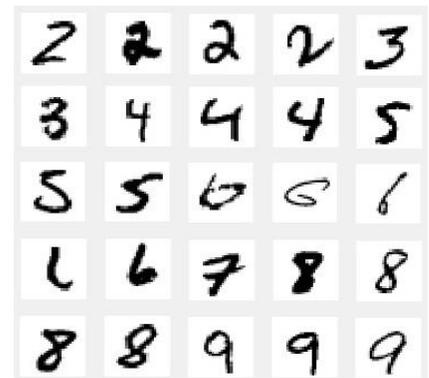


Fig.6. 25 errors made by the DLFRB classifier

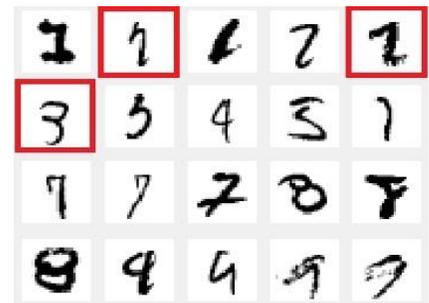


Fig.7. 20 errors made after the involvement of the SVM based conflict resolution classifier.

The training time of each AnYa type fuzzy rule of the ALMMo system based on the GIST and HOG features of the original training set are tabulated in Table III. However, note that these are indicative times

TABLE IV. COMPARISON BETWEEN THE PROPOSED APPROACH AND DIFFERENT DNN APPROACHES

Approaches	Accuracy	Training Time	PC Parameters	GPU Used	Elastic Distortion	Reproducibility	Parallelization
The Proposed Approach	99.55%	Less than 2 minute for each part of the network	Core i7-4790 (3.60GHz), 16 GB DDR3	None	NO	YES	YES
DLFRB Classifier	99.44%						NO
SVM based Conflict Resolution Classifier	99.35%						NO
Large Convolutional Neural Networks [6]	99.47%				NO	YES	NO
Committee of 35 Convolutional Neural Networks [4]	99.77%	Almost 14 hours for each one of the 35 DNNs.	Core i7-920 (2.66GHz), 12 GB DDR3	2 × GTX 480 & 2 × GTX 580	YES	NO	NO

and the exact training time is difficult to be provided because the amount of training time required by the ALMMo-0 systems based HOG and GIST features for the same training set are different. In addition, the amount of training time required by the ALMMo-0 systems based on the same type of features of different training sets also varies, though slightly. The training time consumed by the SVM based extension is also tabulated in Table III.

From Table III we can see that, the maximum training time for a fuzzy rule of AnYa type only takes less than 2 minutes. The training process of the SVM classifier also only takes less than 2 minutes. As the proposed approach is highly parallelizable, with enough computing resources, the whole training can be finished within 2 minutes for the 60000 training images. Furthermore, the core of proposed architecture (the DLFRB classifier) can be recursively, non-iteratively updated in an online, evolving scenario whereby the images are provided one by one.

The comparison between the proposed approach and the state-of-art approaches reporting the current best results (with and without elastic distortion) are tabulated in Table IV.

## VI. CONCLUSION

In this paper, a novel, fast, deep learning ensemble classifier is proposed and applied to the well-known benchmark problem of handwriting digits recognition. The proposed approach is a cascade of the recently introduced DLFRB classifier and a SVM based conflict resolution classifier. This approach only involves the most fundamental image transformation techniques and the widely used feature descriptors. Its learning process is highly parallelizable and very fast. A number of highly interpretable AnYa type fuzzy rules are identified during the training process and play a dominant role in the classification. Numerical experiments demonstrate the excellent performance of the proposed approach outperforming the state-of-art deep learning approaches by providing the highest classification accuracy without elastic distortion.

## REFERENCES

- [1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nat. Methods*, vol. 13, no. 1, pp. 35–35, 2015.
- [2] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Convolutional neural network committees for handwritten character classification," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, 2011*, vol. 10, pp. 1135–1139.
- [3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2012.
- [4] D. Cireşan, U. Meier, and J. Schmidhuber, "Multi-column Deep Neural Networks for Image Classification," in *Cvpr*, 2012, pp. 3642–3649.
- [5] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *International Conference on Learning Representations*, 2015, pp. 1–14.
- [6] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?," in *IEEE International Conference on Computer Vision*, 2009, pp. 2146–2153.
- [7] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," *Doc. Anal. Recognition, 2003. Proceedings. Seventh Int. Conf.*, pp. 958–963, 2003.
- [8] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
- [9] A. Nguyen, J. Yosinski, and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 427–436.
- [10] P. P. Angelov and X. Gu, "MICE: Multi-layer multi-model images classifier ensemble," in *IEEE International Conference on Cybernetics*, 2017.
- [11] P. Angelov and R. Yager, "A new type of simplified fuzzy rule-based system," *Int. J. Gen. Syst.*, vol. 41, no. 2, pp. 163–185, 2011.
- [12] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.
- [13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *Proc. - 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, CVPR 2005*, vol. I, pp. 886–893, 2005.
- [14] "MNIST Dataset," <http://yann.lecun.com/exdb/mnist/>.
- [15] V. N. Vapnik., *The Nature of Statistical Learning Theory*. New York: Springer, 1995.
- [16] P. P. Angelov, X. Gu, and J. C. Principe, "Fast feedforward deep learning network with automatic feature extraction," in *submitted to International Joint Conference on Neural Network*, 2016.
- [17] P. P. Angelov, X. Gu, and J. C. Principe, "Autonomous Learning Multi-model Systems from Data Streams," *IEEE Trans. Fuzzy Syst.*
- [18] P. P. Angelov and X. Gu, "Autonomous learning multi-model classifier of 0-order (ALMMo-0)," in *IEEE Conference on Evolving and Adaptive Intelligent Systems*, 2017.