An Ensemble of Deep Learning Architectures for Automatic Feature Extraction

Fatma Shaheen and Brijesh Verma
Centre for Intelligent Systems
Central Queensland University, Brisbane, Australia
{f.shaheen, b.verma}@cqu.edu.au

Abstract— **This paper presents a novel ensemble of deep learning architectures for automatic feature extraction. Many ensemble techniques have been recently proposed and successfully applied to real world applications. The existing ensemble techniques can achieve high accuracy however the accuracy depends on features they use and features are extracted by a separate model for feature extraction. As deep learning architectures such as Convolutional Neural Networks (CNNs) can automatically extract features, it is a good idea to explore their feature extraction ability in an ensemble. Therefore the purpose of this research is to propose an ensemble of CNNs and find out the answer of whether or not an ensemble of CNNs can perform better than the traditional ensemble techniques which use a separate feature extraction. To find an answer of the research question, an ensemble of CNNs, an ensemble of MLPs, a CNN and an MLP are implemented and evaluated on the same benchmark datasets. A large number of experiments were conducted and the results showed that the proposed ensemble of CNNs can automatically extract features and achieve better accuracy but takes a higher number of epochs than other ensembles on some real-world image datasets.**

Index Terms— **Ensemble of CNNs, Deep Learning, Feature Extraction, Convolutional Neural Network;**

I. INTRODUCTION

An ensemble of deep learning architectures is a process of combining a number of deep learners with a fusion technique. An ensemble of neural networks is not a new idea as many ensembles using various machine learning techniques have been recently proposed and evaluated [1-4]. Deep learning architectures can automatically extract and classify image features which make such architectures very efficient and attractive for real-world image parsing applications. Deep learning is not a new concept, however many deep learning architectures have been recently developed and evaluated because of significant improvement in fast computing infrastructure. Many deep learning architectures (e.g. CNNs) contain feature extraction and classification processes together. In traditional techniques, normally features are extracted by a feature extraction technique using various algorithms and then features are identified by a classification technique so the classification [1-15] is a very important task in many real world applications in particular face recognition [6], handwriting recognition [7], medical diagnosis [8,9], customer identification for online banking [10], forecasting in environmental science [11] and many more.

Deep learning based classifiers can learn features and achieve better accuracy than many existing classifiers. A large amount of research on deep learning based CNNs has already been conducted and published. A new version of a deep recurrent of visual attention model has been introduced [16] that uses a deep recurrent neural network trained with reinforcement learning to find the most relevant areas of the input image. This model was first applied to the MNIST dataset and then a real-world Multi-Digit Street View House Number (SVHN) dataset. It was found that multi-digit house number recognition using this model was more successful compared to the performance of the current state-of-the-art convolutional neural networks.

A different form of a recurrent convolutional architecture based model suitable for large-scale visual learning was proposed in [17] that is end-to-end trainable. This model was applied and evaluated on a benchmark of video recognition dataset. The dataset has included over 12,000 videos categorized into 101 human action classes.

A deep neural network with a clustering algorithm was proposed in [18] for reducing the number of correlated parameters and improving the text categorization accuracy. A new input patch extraction method for feature extraction was employed to reduce the redundancy between filters at neighboring locations. Accuracy obtained on an image recognition STL-10 dataset was 74.1% with a test error rate of 0.5% on MNIST dataset. A deep learning technique for robotic hand grasp detection was proposed in [19]. A two-step cascaded system with two deep networks was used, where the top detections from the first network are re-evaluated by the second network.

Deep learning has been combined with an ensemble of neural networks, one such approach is proposed in [20]. The approach was applied to black box image classification problem with 130 thousand of unlabelled samples. Although deep architectures have recently been applied to many application tasks, it is important to understand the ensemble of such deep architectures and compare it with traditional techniques. The complexity of deep architectures makes it difficult to use it for some large scale image processing tasks.

In the past few years, several papers have shown that ensemble techniques can deliver outstanding performance in learning and reducing the test error. An ensemble model with 5 convnets [12] achieved very good performance on the ImageNet 2012 classification benchmark. It achieved a top 1 error rate of 38.1%, compared to the top -1 error rate of 40.7% given by the single model. In [13], it was shown that by using an ensemble of 6 convnets, the top 1 error was reduced from 40.5% to 36.0%.

Many other traditional ensembles have also been in existence for a long time. Breiman introduced [4] the concept of bagging more than 20 years ago which helped us gaining an understanding of how ensemble of classification and regression trees work when they were trained by taking random samples from the whole dataset.

In this paper, we propose an ensemble of CNNs and conducted experiments for ensemble of CNNs and MLPs to

answer the following research questions (i) What is the performance of an ensemble of CNNs with an automatic feature extraction? (ii) How does an ensemble of CNNs perform on complex dataset in comparison with a traditional ensemble of MLPs, a single MLP and a single CNN?

This paper consists of 5 sections. The rest of the paper is organized as follows. Section II describes the proposed ensemble of deep learning architectures. Section III presents the experiments and results. A discussion of results is presented in Section IV. Finally Section V presents the conclusion.

II. Proposed Ensemble Methodology

The proposed ensemble methodology using deep learning architectures is shown in Fig 1. The traditional ensemble of image based MLPs is used for comparison purposes and it is shown in Fig 2. The details of both ensembles are described below in the following subsections.

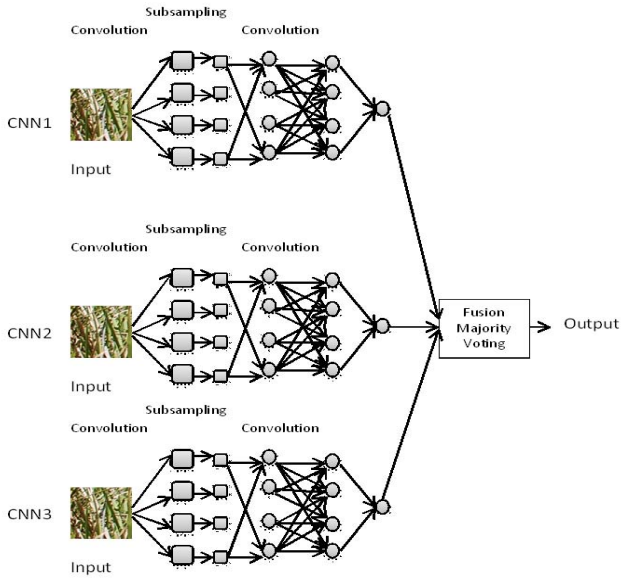

Fig. 1: Automatic Feature Extraction based Ensemble of CNNs

In the proposed ensemble of deep learning architectures, a CNN has been used as a single classifier. In this research, the ensemble architecture used three CNNs without investigating appropriate number of CNNs in ensemble because the purpose of this research is not to find appropriate ensemble parameters. Each CNN in ensemble contains standard layers such as convolutional layer, max pooling layer and fully connected layer. In the convolutional layer, a set of filters is used and every filter can have a variable size. The window size 28x28 and filter size 5x5 were used in each CNN. The max-pooling layer operates independently at each depth slice of the input and resizes it spatially using the max operator. Each CNN is separately trained and then decision is combined using majority voting.

In the proposed ensemble of MLPs, the full image is the input to an ensemble of three Multi-Layer Perceptrons (MLPs). In each MLP, a backpropagation training algorithm is used for the training. The number of hidden neurons and the training epochs are varied to obtain the best accuracy. An overview of this method is presented in Fig 2.

The steps for research methodology are listed below.

Step 1: Image dataset.
Step 2: Train and test ensemble of CNNs.
Step 3: Train and test ensemble of MLPs.
Step 4: Train and test a single CNN.
Step 5: Train and test a single MLP.
Step 6: Repeat Steps 1-5 for different datasets.

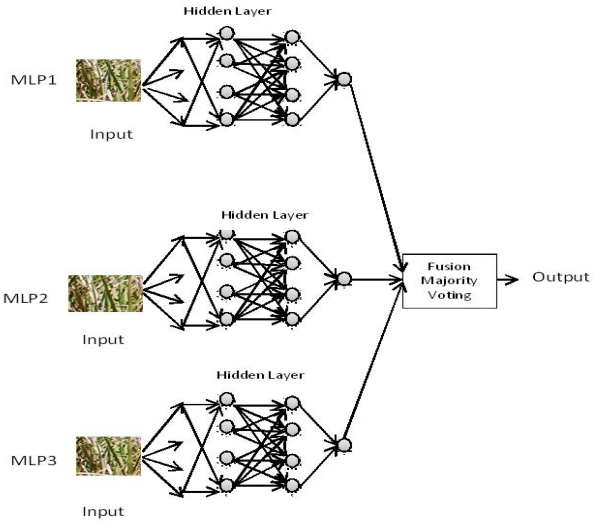

Fig. 2: Image-based Ensemble of MLPs

III. Experiments and Results

The experiments for this research have been conducted on a number of real-world datasets. The first dataset used in this research is MNIST (Mixed National Institute of Standards and Technology). It consists of handwritten digits. The dataset has a training set of 60,000 examples, and a test set of 10,000 examples. MNIST dataset [21] is a good benchmark for evaluating various learning techniques as it has been used by many researchers. The second dataset used in this research is cow heat dataset [22]. This dataset is collected from cow paddock and used to detect heat in cows. The dataset is divided into two categories (a) changed color due to the heat and (b) unchanged. The third dataset is roadside vegetation dataset [23] which is used to identify fire risk based on roadside vegetation. The dataset contains 600 images of 7 different classes (i.e. grass-brown, grass-green, road, sky, soil, tree-leaf, and tree-stem). The dataset has been divided into training and test sets. The training set consists of 75% data and test set consists of 25% data.

The results of experiments on three datasets mentioned above are shown below in Tables I to X. The results using the proposed ensemble of CNNs, traditional ensemble of MLPs, image based MLPs and CNNs are presented and compared.

Table I shows the results obtained from ensemble of CNNs on MNIST dataset. Table II shows the results from image-based ensemble of MLPs with same parameter settings. The results obtained by the proposed ensemble shows 99.33% and traditional ensemble of image-based MLPs shows 95.2% accuracy.

Table I: Accuracy [%] using Ensemble of CNNs on MNIST

#Epochs	Accuracy on Training Set	Accuracy on Test Set
150 1000 1010	99.21	99.33

Table II: Accuracy [%] using Ensemble of MLPs on MNIST

#Epochs	#Hidden Neurons	Accuracy on Training Set	Accuracy on Test Set
50 53 55	12 12 12	78.43	95.21

Table III and Table IV show the results of ensembles on roadside-vegetation dataset. Again ensemble of CNNs shows higher accuracy than the traditional ensemble of MLPs.

Table III: Accuracy [%] using Ensemble of CNNs on Roadside-vegetation dataset

Epochs	Accuracy on Training Set	Accuracy on Test Set
1000 1050 1100	95.45	88.76

Table IV: Accuracy [%] using Ensemble of MLPs on Roadside-vegetation dataset

Epochs	Hidden Units	Accuracy on Training Set	Accuracy on Test Set
50 50 50	16 16 16	80.40	64.12
50 50 45	16 16 20	79.15	73.53

Table V and Table VI show the results obtained by ensemble of CNNs and ensemble of image-based MLPs on cow dataset. It can be seen from the tables that on cow dataset image-based ensemble of MLPs confirms similar test accuracy in comparison with the results obtained by ensemble of CNNs.

Table V: Accuracy [%] using Ensemble of CNNs on Cow heat dataset

#Epochs	Accuracy on Training Set	Accuracy on Test Set
1000 1050 1100	100	100

Table VI: Accuracy [%] using Ensemble of MLPs on Cow heat dataset

#Epochs	#Hidden Neurons	Accuracy on Training Set	Accuracy on Test Set
100 100 101	6 16 6	100	100
100 100 100	6 16 12	100	93.3

Tables VII–X show the results with single classifiers including CNNs and MLPs on all three datasets. The results show that CNN can achieve similar accuracies as ensembles on cow dataset. However proposed ensemble shows much higher accuracy on other two datasets.

Table VII: Accuracy [%] using CNN architecture on all three dataset

#Epochs	50	100	1000
MNIST dataset			
Training Accuracy	92.00	94.30	98.19
Test Accuracy	98.71	98.92	98.99
Vegetation dataset			
Training Accuracy	52.86	72.29	93.20
Test Accuracy	42.1	66.17	72.71
Cow heat dataset			
Training Accuracy	62.86	94.29	100.00
Test Accuracy	80.00	86.67	100.00

Table VIII: Accuracy [%] using image-based MLP on Roadside vegetation data. The only highest accuracies obtained for hidden neurons are listed below.

#Hidden Neurons	#Epochs	Accuracy on Training Set	Accuracy on Test Set
6	100	72.11	57.65
12	100	94.47	68.24
16	100	95.48	70.00
24	50	86.43	68.82
120	100	98.74	70.59

Table IX: Accuracy [%] using image-based MLP on Cow heat dataset. The only highest accuracies obtained for hidden neurons are listed below.

#Hidden Neurons	#Epochs	Accuracy on Training Set	Accuracy on Test Set
6	100	100.00	100.00
12	100	100.00	92.31
16	100	100.00	100.00
24	50	100.00	84.62
120	50	100.00	84.62

Table X: Accuracy [%] using image-based MLP on MNIST dataset. The only highest accuracies obtained for hidden neurons are listed below.

#Hidden Neurons	#Epochs	Accuracy on Training Set	Accuracy on Test Set
6	1000	100.00	86.70
12	50	100.00	93.30
16	50	100.00	86.70
24	100	100.00	86.70
120	50	100.00	80.00

IV. Discussion

The experiments on a number of different datasets were conducted to answer the research questions introduced in introduction section of this paper. The first dataset was a standard MNIST digit classification dataset that has been used by many researchers around the globe for evaluating deep learning algorithms and architectures. The experiments were further conducted on slightly different and challenging real-world local datasets which are roadside vegetation dataset and cow dataset. Although ensemble of CNNs has performed well in all experiments and produced highest results in terms of accuracy (99.33% on MNIST dataset, 100% on cow dataset and 88.76% on vegetation dataset), it is worth to note that ensemble of CNNs took longer time in terms of epochs for each dataset to achieve the highest accuracy. Traditional image based MLPs and ensemble of MLPs have performed as good as ensemble of CNNs on only cow dataset. The ensemble of CNNs has performed well for all datasets, therefore it is appropriate to state that ensemble of CNNs is the best performer and able to automatically extract features and classify them. The results suggest that the proposed ensemble of CNNs is most suitable technique for not only MNIST dataset but for other real world image datasets. The comparative analysis is shown below in Fig 3.

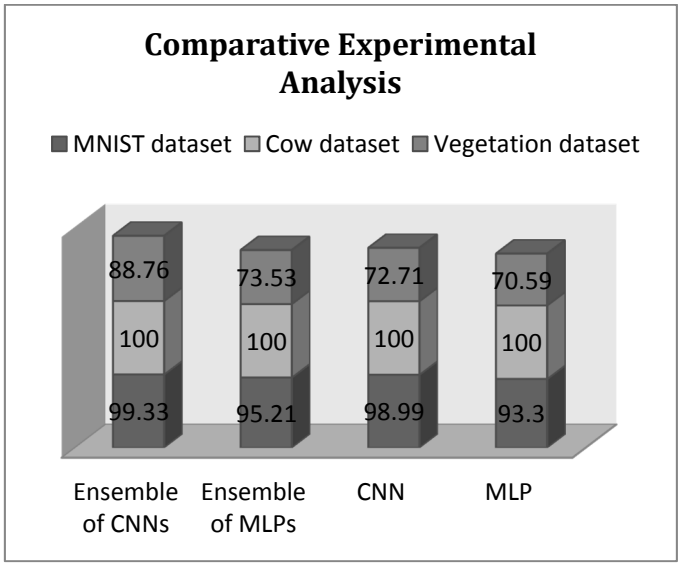

Fig 3: Comparative experimental analysis

V. Conclusion

This paper presented a novel ensemble of CNNs and evaluated the impact of its automatic feature extraction and classification abilities on a number of real-world datasets. A detailed analysis of the classification accuracy of ensemble of CNNs, ensemble of MLPs, CNNs and MLPs was conducted. The ensemble of CNNs was firstly evaluated on MNIST, Cow and Vegetation datasets and then the ensemble of image based MLPs, single CNNs and single image based MLPs were evaluated on the same benchmark dataset so that a comparison of performance could be conducted. The full image was used as an input to ensembles of CNNs, ensemble of MLPs, CNNs and MLPs. Similar experimental conditions were used for the training and testing of each model.

The systematic experiments suggest that ensemble of CNNs with an automatic feature extraction based image classification performs the best but it takes longer time to learn. It has been found that, for some real-world datasets a simple ensemble of traditional MLPs can have equivalent performance with a small number of epochs in comparison to ensemble of CNNs.

The proposed ensemble of CNNs has outperformed all other ensembles and single classifiers including CNNs. It has obtained 99.33% accuracy on MNIST dataset, 100% accuracy on cow dataset and 88.76% accuracy on vegetation dataset which are the highest among published accuracy on these datasets. This research will be further extended by considering all ensemble parameters and more benchmark datasets.

VI. References

1. Z. Lu, X. Wu and J.C. Bongard, “Active learning through adaptive heterogeneous ensembles“, IEEE Transactions on Knowledge and Data Engineering, 2015. 27(2): pp. 368-381.
2. V. Cheplygina, D.M. Tax and M. Loog, “Dissimilarity-based ensembles for multiple instance learning”, IEEE Transactions on Neural Networks and Learning Systems, 2016. **27**(6): pp. 1379-1391.
3. B. Verma and A. Rahman, “Cluster-oriented ensemble classifier: Impact of multicluster characterization on ensemble classifier learning”, IEEE Transactions on Knowledge and Data Engineering, 2012. 24(4): pp. 605-618.
4. L. Breiman, “Bagging predictors”, Machine Learning, 1996. 24(2): pp. 123−140,.
5. J. Schmidhuber, "Deep learning in neural networks: an overview", Neural Networks, 2015. 61: pp. 85-117.
6. R.S. Ahmad, K.H. Mohamad, S.S. Liew and R. Bakhteri, "Convolutional neural network for face recognition with pose and illumination variation", International Journal of Engineering and Technology (IJET), 2014. 6(1): pp. 44-57.
7. H. Lee and B. Verma, “Binary segmentation algorithm for English cursive handwriting recognition”, Pattern Recognition, 2012, 45 (4): pp. 1306-1317.
8. B. Sahiner, H.P. Chan, N. Petrick, D. Wei, M.A. Helvie, D.D. Adler and M.M. Goodsitt, "Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images", IEEE Transactions on Medical Imaging, 1996, 15(5): pp. 598 610.
9. B. Verma and S. Hassan, “Hybrid ensemble approach for classification”, Applied Intelligence, 2011, 34 (2):pp.258-278.
10. J.L. Marzo i Lázaro, "Enhanced convolution approach for CAC in ATM networks, an analytical study and implementation", 1997: Universitat de Girona.
11. B. Klein , L. Wolf and Y. Afek, "A dynamic convolutional layer for short range weather prediction", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4840-4848.
12. A. Krizhevsky, I. Sutskever and G.E. Hinton, "ImageNet classification with deep convolutional neural networks", Advances in Neural Information Processing Systems, 2012, pp. 1097-1105.
13. M.D. Zeiler and R. Fergus,"Visualizing and understanding convolutional network", European Conference on Computer Vision, 2014, pp. 818-833.
14. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1-9.
15. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, 2014.
16. J. Ba, V. Mnih and K. Kavukcuoglu, "Multiple object recognition with visual attention", arXiv preprint arXiv:1412.7755, 2014.
17. J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S.Venugopalan, K. Saenko and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2625-2634.
18. A. Dundar, J. Jin and E. Culurciello, "Convolutional clustering for unsupervised learning", arXiv preprint arXiv:1511.06241, 2015.

19. I. Lenz, H. Lee and A. Saxena,"Deep learning for detecting robotic grasps", The International Journal of Robotics Research, 2015, 34(4-5): pp. 705-724.
20. L. Romaszko, "A deep learning approach with an ensemble-based neural network classifier for black box ICML 2013 contest", Workshop on Challenges in Representation Learning, ICML, 2013, pp. 1-3.
21. Y. LeCun, C. Cortes and C. J.C. Burges,"The MNIST database of handwritten digits", 2016. [Online], Available: http://yann.lecun.com/exdb/mnist/, [Accessed: 08 September 2016]
22. F. Shaheen, M. Asaf and B. Verma “Impact of automatic feature extraction in deep learning architecture”, Proceedings of the International Conference on Digital Image Computing Techniques and Applications, 2016, pp. 1-6.
23. L. Zhang, B. Verma and D. Stockwell, "Class-semantic color-texture textons for vegetation classification", Proceedings of the International Conference on Neural Information Processing, Springer, 2015, pp. 354-362.