Robust Convolutional Neural Networks for Image Recognition

Hayder M. Albeahdili
Dep. of Electrical and computer Engineering
University of Missouri, Columbia
Columbia, Missouri, 65211, USA

Haider A. Alwzwazy
Dep. of Electrical and computer Engineering
University of Missouri, Columbia
Columbia, Missouri, 65211, USA

Naz E. Islam
Dep. of Electrical and computer Engineering
University of Missouri, Columbia
Columbia, Missouri, 65211, USA

***Abstract*—Recently image recognition becomes vital task using several methods. One of the most interesting used methods is using Convolutional Neural Network (CNN). It is widely used for this purpose. However, since there are some tasks that have small features that are considered an essential part of a task, then classification using CNN is not efficient because most of those features diminish before reaching the final stage of classification. In this work, analyzing and exploring essential parameters that can influence model performance. Furthermore different elegant prior contemporary models are recruited to introduce new leveraging model. Finally, a new CNN architecture is proposed which achieves state-of-the-art classification results on the different challenge benchmarks. The experimented are conducted on MNIST, CIFAR-10, and CIFAR-100 datasets. Experimental results showed that the results outperform and achieve superior results comparing to the most contemporary approaches.**

Keywords*—Convolutional Neural Network; Image recognition; Multiscale input images

I. Introduction

Convolutional Neural Network (CNN) has been widely used in many real world applications, including face recognition [1, 2], image classification and recognition [3-6] and object detection [7] because it is one of the most efficient methods for extracting critical features for non-trivial tasks. CNN consists of a pipeline of alternative several different layers. Unlike neural network, CNN has three different types of layers which are considered a constituent element of CNN. Usually, Convolutional layer, subsampling layers, and fully connected layer are the main components of CNN. Also, there are some intermediate layers between those main layers that will be shown later. Then for a given task, images are passed into CNN to be processed. Passing images through several squish functions incorporated within CNN layers can lead to not leveraging some critical information used for recognition and some of the small features disappear after few layers. The reason for that is because the CNN architecture that implies like those restrictions. Specifically, both convolutional layers and max-pooling layers impose diminishing small features. To implement a robust model, small features must survive for long stages of CNN. To alleviate weaknesses inherited from former CNN models, in this work, different parameters that can influence features surviving for longer distance are explored. Deeper analysis for convolutional and max-pooling layers are presented, and then we introduce a model that has more chance for small features to survive until the final stage of CNN; specifically directly before fully connected layer.

The rest of the paper will be into five sections. In section II, prior works are presented. The most recent contemporary works are obtained. Then in section III, motivation and contribution of this work are introduced. The answer for questions, what have proposed and why it is proposed are presented in this part. Then in section IV, deploying different CNN architectures are presented. Different CNN structures are obtained in this section. Finally, experimental setup and conclusion are presented.

II. Related Work

The most dominant recent works achieved using CNN is a challenge work introduced by Alex Krizhevsky et al. [8] used CNN for challenge classification ImageNet. Various other techniques are proposed later to enhance CNN performance as demonstrated in [9, 10, 11, 12]. Recently vast works have been proposed to improve image recognition accuracy results using different methods. Thus several proposed methods are proposed for variety of applications such as image recognition [13, 14, 15, 16, 8], object detection [17, 18, 19], scene labeling [20], segmentation [21, 22], and variety of other tasks [23, 24, 25]. In addition, image recognition can be accomplished using different other approaches such as Pedro F. Felzenszwalb et al. [26] proposed a method for image recognition using Deformable Part Models (DPM). Further works are devoted using different strategies of using DPM as demonstrated in [27, 28, 29]. Varity of other methods are used for image classification such as SVM [30, 31, 32, 33], boosting [34], spatial pyramid matching [35] and different other works described in [36-39].

III. Motivation and Contribution

The state-of-the-art of image recognition specifically achieved on CIFAR-10, CIFAR-100, and MNIST is achieved using different technique as proposed in [40, 41, 42]. This work has some common procedures with prior works which can be described as follows:

- The first step is applying the pre-processing to the input images such as local contrast normalization. There are different pre-processing steps that can be applied before input images passed into deep model. In this work, pre-processing steps demonstrated by Goodfellow et al. [6] is followed.

- After pre-processing, images are fed into CNN to be trained. Multi-stage CNN is used to train and extract critical features from the input patterns used later to final scoring results. In this work different CNN architectures are used for training and extracting features. Specifically very contemporary works are recruited and incorporated for introducing new unified model which achieves the state-of-the-art image classification on the datasets used in this work. Furthermore, a robust CNN is proposed at the end of this work which accomplishes superior results comparing with recent works.
- Finally, the final outputs of CNN are evaluated for final scoring results. There are different methods to score final results of CNN either using SVM or using CNN itself by using soft-max layer build on the top of CNN. Thus soft-max layer are used in this work to evaluate and score the final recognition results.

One of the most contemporary work used CNN for image classification called Network In Network (NIN) introduced in [3] achieves superior results over several prior existing models that use deep neural network for image recognition because NIN uses different connection technique between convolutional layers than what is in conventional CNN. However, there are several factors that can influence and impact model performance leading to degrade model accuracy. Thus, in this work NIN will be recruited after diminishing its shortcomings. Weaknesses of general CNN used for image classification are various such as CNN's depth, width of the network, filter sizes, and network topology. All these are vital factors that can highly impact recognition accuracy. Consequently to diminish lethargy inherited from CNN architecture; this work endeavors to alleviate shortcomings of former networks by eliminating most limitations described earlier. Therefore in this work the most recent and very efficient methods are ensemble to be used for not trivial object recognition tasks. Variety of techniques is delved to enhance image recognition. Starting from leveraging models proposed in [3, 17] both models have several deterministically advantages over prior models as elucidating later. Both concrete models are adapted in this work for image recognition. In addition, extensive work is deliberated for exploring the impact of different parameters that can drastically influence model performance. Virtuous model is mainly instantiated to overcome drawbacks of prior deep neural network architecture used for image recognition. Finally a robust paradigm of CNN architecture is proposed at the end of this work. It achieves superior results comparing with all existing models.

Elegant CNN architectures are adapted to be used for image recognition are originally proposed for image classification [3] and object detection [17]. They are considered the robust deep neural networks models. It is worth mentioning that SPPnet proposed in [17] recruited in this work to provide multi-scale input to the image recognition model. Consequently, to best of our knowledge that image classification such as CIFAR-10, CIFAR0-100, and MNIST are trained with this like method. Providing multi-resolution input images to CNN enhances CNN accuracy drastically as it will be shown later. Furthermore, digging deeper for investigating and exploring most influential parameters is also devoted. Carefully exploring influential parameters can be best suited for mole recognition. Different model architectures are extensively analyze and investigated.. After obtaining best suited parameters, a robust model is proposed to enhance recognition performance. Proposed CNN architecture achieves best results and outperforms over most existing models. The proposed model is compared to the prior efficient works specifically compared to the prior deep neural network models. In addition, the experiments are conducted on different benchmarks for evaluation purpose. The experiments are mainly conducted on CIFAR-10, CIFAR-100, and MNIST datasets.

IV. Deploy Different CNN Architectures for Image Recognition

As illustrated earlier, this work principally is recruited two different deep neural network models named NIN and SPPnet explored in [3] and [17] respectively and implemented new unified model. Next sections start exploring in depth the influence and leveraging of incorporating both models on network architectures and how they can influence classification performance. Then the unified proposed model is an elegant model because it shortens some weaknesses inherited from former models. Thus exploring both architectures is accomplished next sections to show model's robustness on image classification.

A. Pipeline Steps of image classification

The basis CNN architecture is depicted in fig. 1. It fundamentally consists of series of stages. Part (a) presents images with multiscale to the network. Providing multi-resolution input is an essential step to gain higher accuracy. Part (b) trains the network with fed images. After choosing different scales for input images, they will feed to the CNN to extract features from different resolutions which increase the chance for small features to be enlarged using this technique. It is worth mentioning that using multi-scale input images is a method showed in [17] to increase object detection accuracy. However, we utilize it to be recruited in image recognition task. Then, finally part (c) classifies and scores input pattern. To look deeper for operations accomplished by CNN, the following steps are applied:

1) Input images are pre-processed using Goodfellow et al. [6] to be prepared for the next step.

2) After pre-processing, input images are fed to CNN. In this work a new architecture is proposed as shown in fig. 1. In addition, a robust and an efficient code are used for this purpose called Caffe [43]. It is very fast implementation which can process huge amount of data efficiently. In addition, it is very flexible to be easily adapted to different CNN architecture. The final layer of CNN has n-dimensional feature vector which is used for final classification results, where n is the number of classes for a given dataset.

3) Soft max layer is used for final scoring output. However, the length of final feature vector is anticipated to be n to match the number of classes.

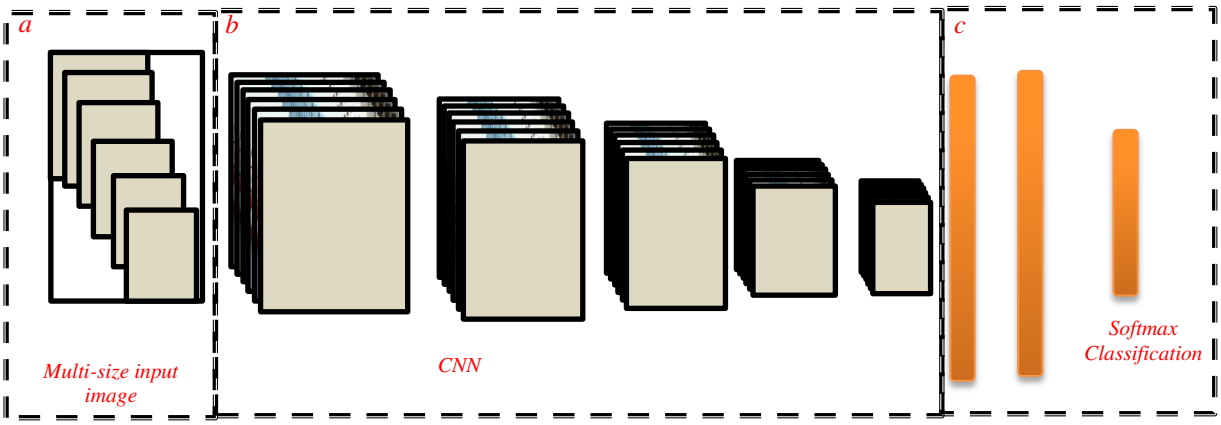

Fig. 1. simple CNN used for image classification

Fig. 1 describes CNN which has the architecture defined as 96C-96S-256C-256S-96C-96S-90F-120F-x-softmax, where C stands for Convolution layer, S is for subsampling layer, and F is for full conned layers.

It is worth mention that a dropout technique demonstrated in [12] is used in this work also to increase model performance by enhancing internal parameters and introducing more solid model. The accuracy achieved using CNN depicted in fig.1 is 0.9953, 0.83, and 0.528 on MNIST, CIFAR-10, and CIFAR-100 respectively. It is obvious that this model achieves competitive results to the most recent works. Next section provides deeper analysis and investigation for exploring and proposing more robust model.

B. *Exploring different CNN architectures*

It is obvious that the proposed network in fig. 1 achieves competitive results comparing to prior works. In addition, it accomplishes results which outperform accomplished work in [44] specifically it dominants over deep neural network approaches. Moreover, it achieves competitive results to many other approaches. The stimulating results are supportive to dig deeper and to investigate influential parameters and explore more robust model. In this part, recruited models will be used for further investigation and more effort will be put to explore more appropriate architecture for image classification. Leveraging CNN architecture is proposed in this section used for image recognition. It achieves state-of-the-art results on given benchmarks. Consequently, more parameters that can influence model performance are discussed next.

This work proposes a new topology for CNN architecture. Fig. 2 depicts the proposed model and it has drastically changes comparing with one implemented and explored in fig. 1. The proposed model inherits some leverage points from NIN. Instead of using conventional connection between convolutional layers as describe in [12, 9, 10, 11]. the robust connection proposed in NIN is incorporated in this work to increase and gain more accuracy on image classification. The size of CNN is kept the same as depicted in fig.1. The merit of this CNN architecture combines more than one elegant method such as multi-scale input images and nonlinear transformation between convolutional layers as demonstrated in [3] as shown in fig. 2.

To look deeper inside CNN and investigate the most critical parameters that can influence model performance. Fig.3 shows both convolutional and sub-sampling layers of CNN.

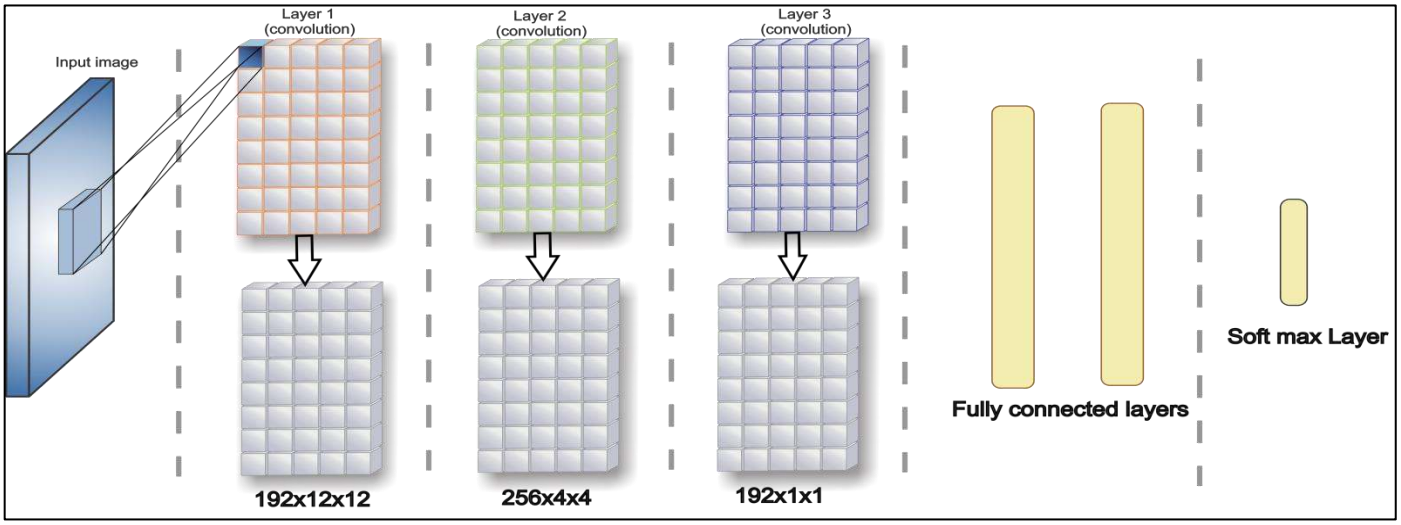

Fig. 2. CNN incorporated with two roboust models

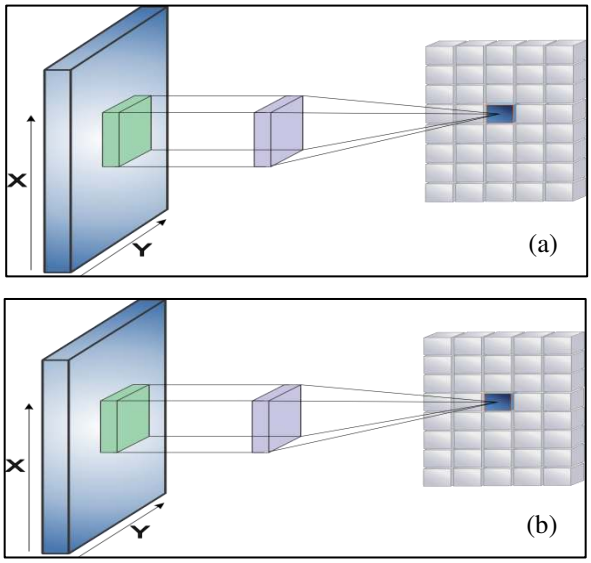

Fig. 3. CNN's layers (a) convolutional (b) max-pooling

It is clear the subsequent of alternative between these kinds of layers; it quickly diminishes the input images after few stages of CNN leading to losing vital information useful image recognition. Accordingly, CNN architectures are explored to be best suited for image classification. There are two model architectures are used in our experiments. They are shown in table I. In addition to the structure obtained in table 1, each network has more additional two fully connected layers build on the top of the final max-pooling layer. Then finally, soft-max layer is built on the top of final fully connect layer used for final scoring results. It is clear that there are two CNN architectures detailed in table1 called Network1 and Network2. It is obvious that network1 is smaller than the network2. Where, network1 consists of three convolutional layer and three max-pooling layers.

for final stage of classification. Specifically this work is dealing with small image sizes as will be obtained later. All the benchmarks used in this work have image sizes of 32x32 pixels. Consequently the small features will be not available after few stages. Therefore an elegant model of CNN architecture is proposed in this work as shown in fig. 4. It is clear that new model propose different connection than standard connection of conventional CNN. Some layers are received their connections not only directly from the layer below but also from two and three layers below. The reason for this kind of connections because small features within the input images can survive longer and will be part of the final scoring detection results. Furthermore, the first layers of CNN extract global features of input objects but as the images advance toward final fully connect layers, more accurate features are extracted.

C. Exploring Different CNN Sizes

In order to precisely analyze the influence of different CNN architectures, a new CNN architecture is proposed and carefully selected their parameter because same CNN architecture might work sufficiently for some tasks and inadequately for other tasks. Hence, in this part different deep model architectures is investigated that can fit for

TABLE I. TWO CNN ARCHITECTURES. THE ABBREVIATION CON REFERS TO CONVOLUTION. XXYXY: X REPRESENTS NUMBER OF FEATURE MAPS AND Y IS THE KERNEL SIZE. LRN AND RELU ARE ABBREVIATION FOR LOCAL RESPONSE NORMALIZATION AND RECTIFIED LINEAR UNIT RESPECTIVELY

Model name	Input size	Con1/pool1	Con2/pool2	Con3/pool3	Con4/pool4	Con5/pool5
Network1	32x32	192x5x5,str:1, ReLU	256x5x5,str:1, ReLU	192x3x3,str:1, ReLU	-	-
		2x2, LRN	2x2, LRN	2x2, LRN	–	-
Network2	32x32	192x5x5,str:1, ReLU	256x1x1,str:1, ReLU	384x1x1,str:1, ReLU	256x1x1, str:1, ReLU	192x3x3, str:1, ReLU
		3x2	3x2	–	–	Spp layer

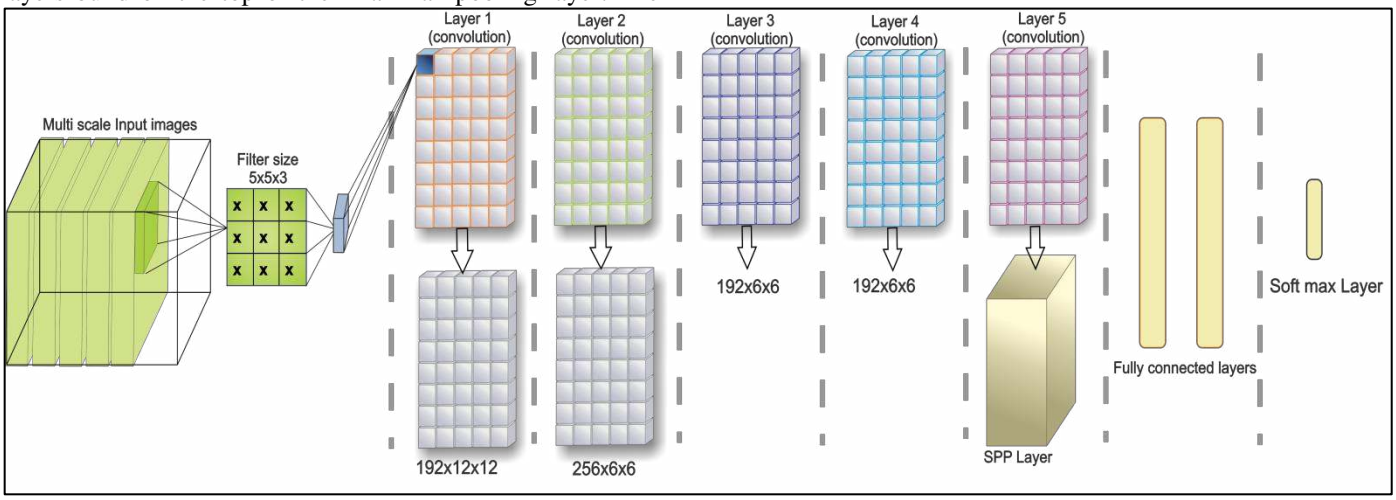

Fig. 4. CNN with five convoutional layers

V. EXPERIMENTS SETUP

In order to evaluate the proposed architecture models, extensive experiments are conducted on different challenge datasets. The most popular datasets are used for evaluation. MNIST, CIFAR-10, and CIFAR-100 are the benchmarks used in this work. To obtain the challenge accompanied with those datasets, next parts explain the related details for the datasets such as size number of image samples. It is worth mentioning that data augmentation is not used in these experiments.

A. MNIST dataset

MNIST [18] is a hand written digits 0-9. The dataset consists of 60000 samples. 50000 samples are used for training and the rest used for testing. All samples have the same size which is 28x28 pixels. The pixels are scaled to be between [0, 1] before the training. There is no preprocessing or data augmentation used in this work. The first CNN, which is named network1, structure is 192C-192S-256C-256S-192C-192C-200F-128F-10-soft-max, where C stands for Convolution layer, S is for subsampling layer, and F is for full conned layer. In this dataset, the size of mini-batches is 128 images. Test accuracy is 0.9961 % for MNIST dataset. This result is superior comparing with results [44]. A summary of the best published results on MNIST dataset is shown in Table II. Network2 which has a structured described as 192C-192S-256C-256S-384C-256C-192C-192S-400F-128F-10-soft-max achieves lower results than the prior model because MNIST might not require large network. The result achieved using network2 is 0.9958 on MNIST dataset. Comparing with other results, Table II shows the final results on MNIST. From Table II, it is obvious that both introduced morels achiever better results than what have been accomplished by Hayder et al. in [44] using their method hybrid training algorithm called Hybrid PSO-SGD which represent training algorithm using Particle Swarm Optimization and Scholastic Gradient Descent.

TABLE II. RESULTS ON MNIST DATASET

Method	Ref. #	Test Accuracy
Unsupervised Learning	[21]	0.64
What is the Best Multi-Stage	[22]	0.53
2-Layer CNN + 2-Layer NN	[23]	0.53
Stochastic Pooling	[23]	0.47
NIN + Dropout	[23]	0.47
Conv. maxout + Dropout	[24]	0.45
Hybrid PSO-SGD	[44]	0.43
Network1	**Ours**	**0.39**
Network2	**Ours**	**0.42**

B. CIFAR-10 Dataset

CIFAR-10 dataset consists of 10 classes of natural 32x32 RGB images with 50,000 samples for training and 10,000 samples for testing [19]. The same structure of network1 is used first for evaluation. The same steps are followed as in MNIST for CNN training. The performance achieved on this dataset is 86.73%.

On the other hand, the test accuracy on CIFAR-10 using network2 is 88.13% which is higher than network1 because CIFAR-10 is more challenge dataset than MNIST. Thus it requires more complicated structure. From table III, it is evident that the proposed method surpasses the other state-of-the-art works.

TABLE III. TEST SET ACCURACY RATES ON CIFAR-10 DATASET

Method	Reference #	Accuracy
Tiled CNN	[25]	73.10
Improved LCC	[26]	74.50
KDES-A	[27]	76.00
PCANet-2 (combined)	[28]	78.67
PCANet-2	[28]	77.14
K-means (Triangle, 4000 features)	[29]	79.60
Cuda-convnet2	[30]	82.00
Hybrid PSO-SGD	[44]	82.41
Network1	**[ours]**	**86.73**
Network2	**[ours]**	**88.13**

C. CIFAR-100

CIFAR-100 is one of the most challenge dataset and it has 100 classes. Images are similar to CIFAR-10 even with size. However, the main difference is that number of image samples per class are very few comparing with CIFAR-10. The total number of images is 50,000 training examples. Thus each class has 500 samples only. Testing samples has 10,000 samples. Like CIFAR-10, the pixels are scaled to be between [0, 1] before the training. Since CIFAR-100 is similar to CIFAR-10 are similar, the same setting of CNN was used for both networks.

Table IV shows the final results achieved using the proposed two models. The first network achieves 53.52% test accuracy on CIFAR-100 while network2 achieves higher accuracy which is 59.85%.

TABLE IV. TEST SET ACCURACY RATES ON CIFAR-100 DATASET

Method	Reference #	Accuracy
CONV. NET + PROBOUT	[45]	61.86%
Baseline + learned tree	[46]	63.15%
NOMP encoder	[47]	60.8%
Stochastic Pooling	[23]	57.49%
NIN	[3]	64.32%
Smooth Pooling Regions	[48]	56.29%
Beyond Spatial Pyramids	[49]	54.23%
Maxout Networks	[5]	61.43%
Network1	ours	53.52%
Network2	ours	59.85%

VI. CONCLUSION

In this work, image recognition using the deep neural network is introduced. Different model architectures are proposed by incorporating different prior elegant CNNs. Specifically both NIN and SPPnet are incorporated in a single unified model that achieves superior results comparing to former results. Then a new model is presented and outperforms prior work and accomplishes state-of-the-art results on the datasets. Also, different model architectures are introduced, and extensive parameters are discussed that can influence model performance. Deeper exploring different parameters that can be suited for CNN recognition model are presented as well. For evaluation, the experiments are conducted on challenge datasets. MNIST, CIFAR-10, and CIFAR-100 are the datasets used in this work.

FUTURE WORK

In feature work, more effort will be devoted in exploring more powerful network to handle more challenge tasks. More enhancements can be achieved by utilizing more technique to be recruited together and implemented the final model. Future works could also include more details such as reporting time consumption for each method and whether it is suitable for real-time applications or not. Also, those implemented models can be re-adapted to be used in object detection tasks.

REFERENCES

[1] S. L. Phung and A. Bouzerdoum, "A pyramidal neural network for visual pattern recognition," IEEE Transactions on Neural Networks, vol. 27, no. 1, pp. 329343, 2007

[2] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf , "DeepFace: Closing the Gap to Human-Level

[3] Min Lin, Qiang Chen, and Shuicheng Yan "Network In Network" arXiv 1312.4400v3,4 Mar 2014

[4] Dan Cires¸an, Ueli Meier and Jurgen Schmidhuber, "Multi-column Deep Neural Networks for Image Classification" CVPR 2012

[5] Julien Mairal, Piotr Koniusz, Zaid Harchaoui, and Cordelia Schmid, "Convolutional Kernel Networks" arXiv 14 Nov 2014 .

[6] Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, Yoshua Bengio "Maxout Networks" ICML 2013

[7] Dumitru Erhan, Christian Szegedy, Alexander Toshev, and Dragomir Anguelov, "Scalable Object Detection using Deep Neural Networks" CVPR, 2014.

[8] Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey. ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 25 (NIPS'2012). 2012.

[9] Marc'Aurelio Ranzato, Fu-Jie Huang, Y-Lan Boureau, Yann LeCun, "Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition" CVPR, 2007

[10] Matthew D Zeiler and Rob Fergus. Stochastic pooling for regularization of deep convolutional neural networks. arXiv preprint arXiv:1301.3557, 2013.

[11] Ian J Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. Maxout networks. arXiv preprint arXiv:1302.4389, 2013.

[12] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov" Dropout: A Simple Way to Prevent Neural Networks from Overfitting" Journal of Machine Learning Research 15 (2014) 1929-1958.

[13] Fabien Lauer, Ching Y. Suen, and G´erard Bloch "A trainable feature extractor for handwritten digit recognition" Journal Pattern Recognition 2007.

[14] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, Zhuowen Tu, " Deeply-Supervised Nets "NIPS 2014

[15] M. Fischler and R. Elschlager, "The representation and matching of pictorial structures," IEEE Transactions on Computer, vol. 22, no. 1, 1973

[16] Kevin Jarrett, Koray Kavukcuoglu, Marc'Aurelio Ranzato and Yann LeCun " What is the Best Multi-Stage Architecture for Object Recognition?" ICCV'09, IEEE, 2009.

[17] Kaiming, He and Xiangyu, Zhang and Shaoqing, Ren and Jian Sun "Spatial pyramid pooling in deep convolutional networks for visual recognition" European Conference on Computer Vision, 2014

[18] Ross Girshick, "Fast R-CNN "arXiv preprint arXiv:1504.08083, 2015

[19] X. Wang, M. Yang, S. Zhu, and Y. Lin. Regionlets for generic object detection. In ICCV, 2013. 8

[20] Karen Simonyan and Andrew Zisserman "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION" arXiv:1409.1556v5 [cs.CV] 23 Dec 2014.

[21] C. Couprie, C. Farabet, L. Najman, and Y. LeCun. Indoor semantic segmentation using depth information. Internatinal Conference on Learning Representation, 2013. 2.

[22] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. CoRR, abs/1311.2524, 2013. 4

[23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. FeiFei. Imagenet: A large-scale hierarchical image database. In CVPR, 2009. 2.

[24] L. N. Clement Farabet, Camille Couprie and Y. LeCun. Learning hierarchical features for scene labeling. PAMI, 35(8), 2013. 1, 2

[25] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng "Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations".

[26] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester and Deva Ramanan ," Object Detection with Discriminatively Trained Part Based Models"

[27] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in IEEE Conference on Computer Vision and Pattern Recognition, 2003

[28] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. OverFeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229, 2013. 1, 2.

[29] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," International Journal of Computer Vision, vol. 77, no. 1, pp. 259–289, 2008.

[30] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In CVPR, 2006. 1, 2, 5, 6, 7

[31] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. Technical report, CalTech, 2007.

[32] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. CNN features off-the-shelf: an astounding baseline for recognition. arXiv preprint arXiv:1403.6382, 2014. 1, 2.

[33] Y. Lin, T. Liu, and C. Fuh. Local ensemble kernel learning for object category recognition. In CVPR, 2007

[34] A. Opelt, M. Fussenegger, A. Pinz, and P. Auer. Weak hypotheses and boosting for generic object detection and recognition. In ECCV, 2004.

[35] K. Grauman and T. Darrell. The pyramid match kernel: Discriminative classification with sets of image features. In ICCV, 2005.

[36] P. Sermanet, S. Chintala, and Y. LeCun. Convolutional neural networks applied to house numbers digit classification. In ICPR, 2012

[37] Q. V. Le, J. Ngiam, Z. Chen, D. Chia, P. W. Koh, and A. Y. Ng, "Tiled convolutional neural networks," in NIPS, 2010.

[38] Tsung-Han Chan, Kui Jia, Shenghua Gao, Jiwen Lu, Zinan Zeng, and Yi Ma "PCANet: A Simple Deep Learning Baseline for Image Classification? "arXiv:1404.3606v2 [cs.CV] 28 Aug 2014.

[39] Adam Coates, Honglak Lee, and Andrew Y. Ng "An Analysis of Single-Layer Networks in Unsupervised Feature Learning" In AISTATS 14, 2011.

[40] Marc'Aurelio Ranzato, Fu-Jie Huang, Y-Lan Boureau, Yann LeCun, "Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition" CVPR, 2007

[41] Matthew D Zeiler and Rob Fergus. Stochastic pooling for regularization of deep convolutional neural networks. arXiv preprint arXiv:1301.3557, 2013.

[42] Ian J Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. Maxout networks. arXiv preprint arXiv:1302.4389, 2013.

[43] Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor "Caffe: Convolutional Architecture for Fast Feature Embedding" arXiv preprint arXiv:1408.5093, 2014.

[44] Hayder M. Albehadili and Naz Islam "Hybrid Algorithm For the Optimization of Training Convolutional Neural Network". Volume 6 IJACSA,No. 10, October 2015

[45] Jost T. S. and Marth R, "improving deep neural networks with probabilistic Maxout units ", ICLR 2014

[46] Nitish Srivastava and Ruslan SalaKhudinov "Discriminative transfer learning with tree-based priors, NIPS,2013

[47] Tsun-Han Lin and H. T. Kung "stable and efficient representation learning with nonnegativity constraints", ICML 2014.

[48] M. Malinowski and M. Fritz "learning smooth pooling regions for visual recognition", BMVC 2013

[49] Y. Jia, C. Huang, and T. Darrell, "Beyond spatial pyramids: Receptive field learning for pooled image features", CVPR 2012