Image Classification Using Robust Convolutional Neural Network

Murtadha A. Majeed
Al-Imam Al-Kadhum College for Islamic science

Abstract—Convolutional Neural Network (CNN) has been widely used in variety of real world applications and it appeals many researchers because of its robustness. In this paper, CNN is used for Image Classification and it achieves high performance. Also our model is used for CIFAR-10 classification. Also it outperforms many other contemporary works. MNIST and CIFAR-10benchmarksare used for evaluation. In this work, different CNN models are proposed and they accomplished non trivial results comparing with many existing models.

Keywords— Convolutional Neural Network, Image Classification,

I. Introduction

Recently, pattern recognition becomes one of the most valuable tasks such as classification, recognition, detection, and clustering. In this work, we targeted image classification as our goal to achieve better performance. Although there are several methods used for image classification, Convolutional Neural Network (CNN) has been widely applied in many applications, including face recognition [1, 2], image classification and recognition [3-6] and object detection [7]. It is widely used in many real world applications such as safe driving systems, security surveillance, and robotic navigation. A large number ofdifficulties such as object appearance make it a challenging task. Different appearances can result from different poses and environmental factors such as illumination. Image classification becomes an even more challenging task because it takes many other considerations for image classification problem.

II. literature review

Variety of techniques are applied for image recognition including neural networks, convolutional neural networks, bag of words, random forest, Scale Invariant Feature Transform (SIFT), and Support Vector Machine (SVM). CNN is considered in this work because of its robustness and its achievement over many other challenging tasks such as object detection, segmentation, and many other tasks. To summarize prior works achieved using different methods, an algorithm proposed in [8] was used to train Boltzmann machine. Authors showed the achievement using this method. Also [9] exhibited a new method of creating nonlinear feature mapping model called DNet-kNN. In this model, authors combined both CNN and K Nearest Neighbor (KNN) to produce a robust model with dimensionality reduction. Also a different method demonstrated in [3] to increase performance for visual recognition tasks. Many other works are proposed in [10,11,12] to achieve superior results comparing with prior works.

III.steps of image classification

Image recognition can have several procedure depending on what kind of method used. However, CNN has common procedures which can described below:

- A preprocessing step is applied before begin actual training. Data augmentation and local contrast normalization are applied as demonstrated in Goodfellow et al. [13]
- Objects are fed to CNN after preprocessing step. Many contemporary works extract strong features to enhance CNN performance.
- Then, extracting features from images using trained CNN to score input candidate. Variety of techniques can be used to score input images. The

extracted features can be evaluated either using Support Vector Machine (SVM) or Soft-max built on the top of CNN layers can be used as well. In this work, SVM is used to assess the final scoring results

IV. Proposed Model Architecture

CNN generally consists of alternatives two main layers called convolution and max-pooling layer and end up with fully connected layer. All these layers are connected to each other with weights. Though, there are many different other different layers used for different purposes. The main structure of final proposed model used in our experiments is depicted in fig. 1.

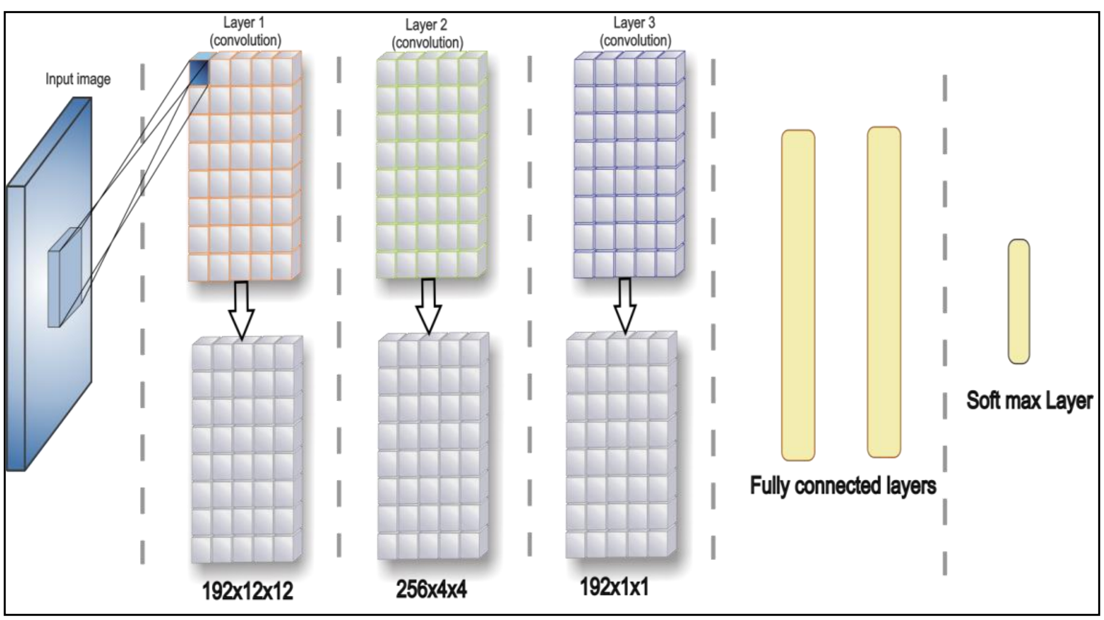

Fig. 1. Architecture of proposed CNN

As described earlier, we essentially proposed elegant CNN paradigm as depicted above. The proposed model has several advantages over many other prior models used for the same task. In this section, we explore the leverage of using many other pre-implemented techniques for

classification pipeline shown in Fig. 1 consists of three main parts. In the first part, images are prepared after pre-processing. In the next step, a large CNN is used to extract the features from the input proposals. Finally, *soft-max* is used on top of the CNN for classification. These steps match most image classification CNN approaches.

V. Benchmark Experiments

A. Overview

] and ١٤The algorithm is evaluated on two benchmark datasets: MNIST [
]. Samples for the datasets are shown in Fig 2. The CNN used in this ١٥CIFAR-10 [
work consists of alternative convolutional and max pooling layers. Fully connected layer is implemented on the top of the network. The architecture of

CNN used for each dataset is dissimilar from each other. The number of particles is 25 and they are randomly initialized with different means and variances.

(a) (b)

Fig.2. Samples of (a) MNIST (b) CIFAR-10

B. MNIST dataset

The MNIST [١٤] 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 in [0, 1] before the training. There is no preprocessing or data augmentation utilized in this work. The CNN structure is 8C-8S-24C-24S-89C-90F-10F, 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. We achieved test accuracy 99.75 for MNIST dataset. A summary of the best published results on MNIST dataset is shown in Table I.

TABLE I. RESULTS ON MNIST DATASET

Method	Ref. #	Test error
Unsupervised Learning	[١٦]	0.64
What is the Best Multi-Stage	[1٧]	0.53
2-Layer CNN + 2-Layer NN	[1٨]	0.53
Stochastic Pooling	[1٨]	0.47
NIN + Dropout	[1٨]	0.47
Conv. maxout + Dropout	[1٩]	0.45
CNN	**Ours**	**0.25**

C. CIFAR-10 Dataset

The CIFAR-10 dataset consists of 10 classes of natural 32x32 RGB images with 50,000 for training and 10,000 for testing [١٤]. The CNN used for this dataset is described as: 12C-12S-48C-48S-89C-90F-10F, which is devoted to convolutional layer with 12 feature maps, subsampling layer, and a convolutional layer with 48 feature maps, subsampling layer, and a convolutional layer with 89 feature maps, and a fully connected output layer with 90 neurons, and a fully connected output layer with 10 outputs. The subsampling layers have filters over non-overlapping region of size 2x2. We follow the same steps as in MNIST for training CNN. However, in this dataset, occurring in local optimum is faster than previous datasets so the number of times applying SGD is higher. We determine that PSO-GA needs to be united by SGD as complicated dataset used such as CIFAR-10 because the MNIST dataset is easier for classification than CIFAR-10. Nevertheless, we are still able to get the benefit of using hybrid POS-SGD. The test accuracy that we get on this dataset is 83%.

From table II, it is evident that our method surpasses the other state-of-the-art works. It is worth mentioning that we only compare with methods that used the same structure of CNN. We also did not use any other techniques that can be very useful such as dropout or drop-connect. In this work, we used exactly the same general structure proposed by Yann LeCun et al. [١٤] and we only changed the training algorithm but we keep the same configuration of CNN.

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

Method	Reference #	Accuracy
Tiled CNN	[٢٠]	73.10

Improved LCC	]٢١[	74.50
KDES-A	]٢٢[	76.00
PCANet-2 (combined)	]٢٣[	78.67
PCANet-2	]٢٣[	77.14
K-means (Triangle, 4000 features)	]٢٤[	79.60
Cuda-convnet2	] ٢٥[	82.00
Proposed CNN Model	**Ours**	**83.00**

VI.Conclusion

In this work, we have proposed and demonstrated a new training model of CNN trained on high performance GPU. In addition, we utilized high efficient deep learning toolbox called Caffe. for training Convolution Neural Network (CNN). We have established that the model is well suited for achieving nontrivial results on different datasets and surprisingly achieving state-of-the-art on these datasets. Analysis also shows that the proposed model is superior on two different benchmark datasets. Superior results are achieved on given benchmarks.

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. YanivTaigman, Ming Yang, Marc’AurelioRanzato, 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 JurgenSchmidhuber, “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, YoshuaBengio “Maxout Networks” ICML 2013
7. Dumitru Erhan, Christian Szegedy, Alexander Toshev, and DragomirAnguelov, “Scalable Object Detection using Deep Neural Networks” CVPR, 2014.
8. Ruslan Salakhutdinov and Geoffrey Hinton “Deep Boltzmann Machines”, 2009
9. Martin Renqiang Min, David A. Stanley, Zineng Yuan, Anthony Bonner, Zhaolei Zhang “Large-Margin kNN Classification Using a Deep Encoder Network” arXiv:0906.1814v1 [cs.LG] 9 Jun 2009
10. Chen-Yu Lee, SainingXie, Patrick Gallagher, Zhengyou Zhang, Zhuowen Tu, “ Deeply-Supervised Nets “NIPS 2014
11. Julien Mairal, Piotr Koniusz, Zaid Harchaoui, and Cordelia Schmid, “Convolutional Kernel Networks” arXiv 14 Nov 2014 .
12. Li Wan, Matthew Zeiler, Sixin Zhang, Yann LeCun, and Rob Fergus, “ “,International Conference on Machine Learning 2013.
13. Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, YoshuaBengio “Maxout Networks” ICML 2013
14. Yann LeCun, Leon Bottou, YoshuaBengio, and Patrick Haffiner “Gradient-Based Leaning Applied to Document Recognition” Proc of the IEEE, November 1998.
15. A. Krizhevsky. Learning multiple layers of features from tiny images. Master’s thesis, Computer Science Department, University of Toronto, 2009. 1, 6.
16. Marc’AurelioRanzato, Fu-Jie Huang, Y-Lan Boureau, Yann LeCun, “Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition” CVPR, 2007
17. Kevin Jarrett, KorayKavukcuoglu, Marc’AurelioRanzato and Yann LeCun ” What is the Best Multi-Stage Architecture for Object Recognition?” ICCV’09, IEEE, 2009.
18. Matthew D Zeiler and Rob Fergus. Stochastic pooling for regularization of deep convolutional neural networks. arXiv preprint arXiv:1301.3557, 2013.

19. Ian J Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and YoshuaBengio. Maxout networks. arXiv preprint arXiv:1302.4389, 2013.
20. Q. V. Le, J. Ngiam, Z. Chen, D. Chia, P. W. Koh, and A. Y. Ng, "Tiled convolutional neural networks," in NIPS, 2010.
21. K. Yu and T. Zhang, "Improved local coordinate coding using local tangents," in ICML, 2010.
22. L. Bo, X. Ren, and D. Fox, "Kernel descriptors for visual recognition," in NIPS, 2010
23. 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.
24. Adam Coates, Honglak Lee, and Andrew Y. Ng "An Analysis of Single-Layer Networks in Unsupervised Feature Learning" In AISTATS 14, 2011.
25. A. Krizhevsky, "cuda-convnet," http://code.google.com/p/ cuda-convnet/, July 18, 2014.