Deep learning and transfer learning approaches for image classification

Sajja Tulasi Krishna, Hemantha Kumar Kalluri

ABSTRACT--- ***Deep Learning is one of the machine learning areas, applied in recent areas. Various techniques have been proposed depends on varieties of learning, including un-supervised, semi-supervised, and supervised learning. Some of the experimental results proved that the deep learning systems are performed well compared to conventional machine learning systems in image processing, computer vision and pattern recognition. This paper provides a brief survey, beginning with Deep Neural Network (DNN) in Deep Learning area. The survey moves on the Convolutional Neural Network (CNN) and its architectures, such as LeNet, AlexNet, GoogleNet, VGG16, VGG19, Resnet50 etc. We have included transfer learning by using the CNN's pre-trained architectures. These architectures are tested with large ImageNet data sets. The deep learning techniques are analyzed with the help of most popular data sets, which are freely available in web. Based on this survey, conclude the performance of the system depends on the GPU system, more number of images per class, epochs, mini batch size.***

Keywords: Convolutional Neural Network (CNN); Deep Learning (DL); Machine Learning (ML); Pre-trained Network; Transfer Learning.

1. INTRODUCTION

Deep Learning is a buzzword in computer world at present. It becomes more popular in many real time applications and it is the wider part of the Machine Learning. Critically, Deep Learning takes a lot of data, which can make decisions about new data. This data is passed through Neural Networks, known as Deep Neural Networks (DNN). A lot of deep learning methods are use neural networks, because of which, the deep learning models are popular as deep neural networks. In Deep learning, Convolutional Neural Network (CNN) [1, 2] is popular type of deep neural networks. CNN eliminate the need for manual feature extraction like traditional features extraction algorithms, such as SIFT, LBP etc. The CNN directly extract the features from a set of raw image data. Related features are not pre-trained; they will learn when the networks are on the train on a group of images. This automated way of feature extraction is the most accurate learning models for computer vision tasks such as object detection, classification, recognition. Traditional Machine Learning approaches done the feature extraction manually and the classification algorithm classify the objects separately. But, In Deep Learning approaches the network itself extract the features without user interpretation also classify the objects. This is shown in Figure 1.

2. CONVOLUTIONAL NEURAL NETWORKS (CNN)

Convolutional Neural Network formed with the help of different layers to perform the image classification task. The architecture of the CNN contains the different layers as follows:

1. **Input Layer:** This input layer, accepts raw images, and forwarded to further layers for extracting features.
2. **Convolution Layer:** After input layer, next layer is convolution layer. In this layer number of filters is applied on images for finding features from images. These features are used for calculating the matches at testing phase.

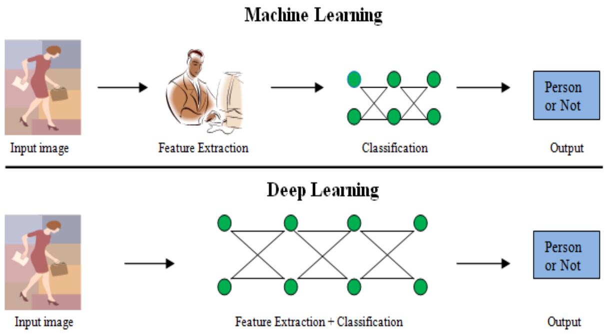

Fig. 1: Machine Learning vs. Deep Learning

3. **ReLU (Rectified Linear Unit):** After convolution layer the next layer is Rectified Linear Unit or ReLU. This layer replaces the negative number of the convolution layer with zero (0), which helps for faster and effective training.
4. **Pooling:** Extracted features are sent to pooling layer. This layer captures large images and reduces them, and reduces the parameters to preserve important information. It preserves maximum value from each window.
5. **Fully Connected Layer:** The final layer is a fully connected layer, which takes up high- level filtered images and translates them into labels with categories.
6. **Softmax Layer:** This layer present just before the output layer. This layer gives the decimal probabilities to each class. Those decimal probabilities are in between 0 and 1.

The first four stages are called feature extraction stages and last two are called classification stages are shown in Figure 2.

Revised Manuscript Received on February 11 , 2019.

Sajja Tulasi Krishna, Vignan's Foundation for Science Technology and Research deemed tobe University.

Hemantha Kumar Kalluri, Vignan's Foundation for Science Technology and Research deemed tobe University. (E-mail: hemanth_mtech2003@yahoo.com)

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

2.1 Transfer learning

In deep learning, the model trains with a large volume of data and learns model weight and bias during training. These weights are transferred to other network models for testing. The new network model can start with pre-trained weights [3].

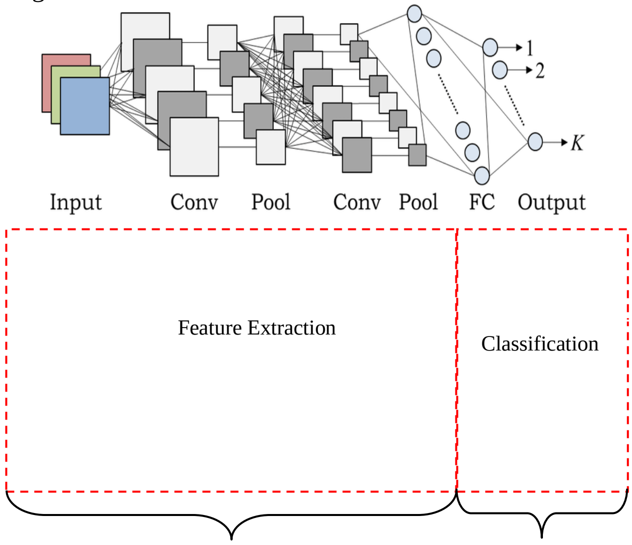

Fig. 2: CNN layers

A pre-trained model is already trained on the same domains. There is a lot of pre-trained architectures are available, the reasons for using pre-trained models are mentioned below as:

First, it requires more computational power to train the huge models on large datasets.

Second, too much time taking to train the network up to number of weeks. Train the new network with pre-trained weights can speed up the learning process.

The top-5 error rates of CNN's architecture are placed in Table 1.

Table 1: top-5 error rates of network Architectures

CNN Architectures	Top-5 error rate (%)
LeNet[4]	28.2
AlexNet[5]	16.4
VGG[6]	7.30
GoogleNet[7, 8]	6.70
ResNet[9]	3.57

The pre-trained models are explained as follows:

LeNet:

LeNet [4] is first Convolutional architecture, which consist of two convolutional layers with ReLu and average pooling layers, followed by another convolutional layer, which is used for flattening, then two fully connected layers and ultimately one softmax layer.

AlexNet:

AlexNet [5] is much deeper neural network than the LeNet. In this network Rectified Linear Unit (ReLU) is used to add non-linearity, it speeds up the network. This network has five convolutional layers, three fully connected layers followed by output layer and also contains 62.3 million parameters.

VGG:

The full form of VGG [6] is Visual Geometry Group. Normally VGG network contains VGG16 and VGG19. In this network the large size kernels are replaced with the multiple number of 3x3 filters, because of this we extract complex features at low cost.

GoogleNet:

This GoogleNet[7, 8] achieves good accuracy but it required high computational power because the order of calculations are very high. GoogleNet was replaced with average pooling after last convolutional layer instead of fully connected layers at the end; this will reduces the number of parameters.

ResNet:

So far, while increasing the network depth automatically accuracy also increases. But some problems arise along with network depth in ResNet [9]. Increased depth that required changing the weights, which arises the end of the network, the prediction becomes small at the initial layers. Another one is huge parameter space it required. To prevent these problems residual modules are came into picture. ResNet50 and ResNet152 are example networks of ResNet.

The Network architectures and their error rates are discussed in Table 1. Now, parameters used in these networks, developed years are shown in Table 2. While network training time of these many no. of parameters are used by the network.

Table 2: CNN architectures with number of parameters

CNN Architecture	Year	Developed by	No. of Parameters
LeNet[4]	1998	YannLeCun et al.	60000
AlexNet[5]	2012	Alex K et al	62.3 million
VGGNet[6]	2014	Simonyan, Zisserman	138 million
GoogleNet[7, 8]	2014	Google	4 million
ResNet[9]	2015	Kaiming He	25 million

3. RESULTS & DISCUSSIONS

Krizhevsky, A, et.al.[5] have proposed a network with five convolutional layers and three fully connected layers. The researchers got best top--1 and top--2 error rates of 37.5% and 17% respectively. The network trained nearly five to six days on GPU. The network tested with the ImageNet Fall 2011 dataset. Hana D, et.al.[10]Used data-augmentation technique, it is used to elaborate existing data sets to avoid the over fitting with convolutional neural network architectures. Transfer Learning is used on pre-trained networks such as AlexNet, VGG16, ResNet-152, achieved better classification accuracy by using data augmentation, which is efficient to particularly when the training data is less. To test the data augmentation phase, researchers used six datasets, those are flowers102, dogs, caltech101, event8, 15 scenes, 67 Indore scene. Pak, M.

et.al. [11]. Reviewed the most popular architectures, and discussed about the reduced error rates and high accuracy rates. These architectures achieve tremendous results over the machine learning algorithms.

Table 3: Existing methods, data sets, results

S.No	Authors	Methods used	Data sets used	Training (Images/%)	Testing (Images/%)	Accuracy
1	Sharma N, et. al.[1]	Transfer learning with AlexNet, GoogleNet, ResNet50	CIFAR10	50000	10000	GoogleNet-68.95%, ResNet50-52.55%, AlexNet-13%
			CIFAR100	50000	10000	
2	Krizhevsky, A, et.al.[5]	Proposed a network contains 5 convolutional and 3 fully connected layers	ImageNet Fall 2011, 15M images, 22K categories,	7.5M	7.5M	Error rates top-5 : 37.5%, top-1 : 17.0%.
3	Hana D, et al.[10]	Transfer learning, web data augmentation technique with Alex,vgg16,resnet-152	Flowers102	8189		92.5%
			dogs	20580		79.8%
			Caltech101	9146		89.3%
			event8	1579		95.1%
			15scene	4485		90.6%
			67 Indore scene	15620		72.1%
4	Hussain, M.et.al. [12]	Transfer Learning with Inceptiov3 model	CIFAR10	50000	10000	70.1%
			Caltech Face	12150		65.7%
						500 epochs-91%
						4000 epochs-96.5%
5	Loussaief, S, et. al. [14]	CNN deep learning	Caltech 101	9146		96%
6	Dutta S, et. al.[19]	Compare NN models	DR	77%	23%	LeNet-72%, VGGNet-76%
			CT	72%	28%	LeNet-71%, VGGNet-78%
7	Ma, B., et. al. [21]	An autonomous learning algorithm automatically generate Genetic DCNN architecture	MNIST,CIFAR10,CIFAR100	90%	10%	99.72% on MNIST, 89.23% on CIFAR10, 66.70% on CIFAR100
8	Devikar, P. [26]	Google's Inception-v3	Dogs			96%
9	Lee, S. J., et. al. [27]	CNN+AdaBoost	CIFAR-10	80%	20%	88.4%

Hussain, M.et.al. [12] , In their work Inceptionv3 architecture of CNN was used. It is one of the transfer learning model and most accurate model for image classification; they achieved 3.46% of top-5 error rate. Originally this Inceptionv3 model had been trained on the ImageNet dataset. For transfer learning the researcher tested on Caltech faces and CIFAR10 datasets. Experimentally researchers proved that the accuracy influenced on the following stages:

1. Number of images in a dataset of each class, if more number of images then the computational time and power required.
2. Increased the full training cycles on total training data, gradually the accuracy improved.
3. Accuracy is depends on the image category also.

Sharma N, et. al.[1], analyze the performance of the AlexNet, GoogleNet and ResNet50 with two datasets

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

CIFAR10, CIFAR100. They concluded, to get the high accuracy, there is a need to increase the number of layers in the network for training. Rawat, W, et.al. [14]In this paper the authors presented the in-depth literature survey, development of deep learning techniques and layers of the Convolutional network. Loussaief, S, et. al. [15] had been compared Bag of Features (BoF) in Machine Learning techniques with CNN based deep learning techniques. BoF method and CNN applied on Caltech101 dataset to compare the result analysis. The feature vector size of both the methods equal, but CNN out performed on BoF. They got 96% accuracy by used CNN for image classification.

Maggiori, E. et al. [16] proposed, size of the data set is influenced on accuracy of classification. To show the experimental analysis they used Tiny-ImageNet and MiniPlaces2 datasets. They are randomly split data set into different training and testing images on different runs, then the results shows that the while increasing the training size automatically accuracy increased. Ballester, P.et al. [17] experimentally proved that, CNN's architectures are not suitable for sketch image classification. They did not classify the images correctly. Researchers applied on TU-Berlin sketch dataset. Alom, M. Z, et. al. [18] gave the in depth survey about the deep learning approaches, pre-trained models. Dutta S, et. al.[19] have compare the Neural network models in performance wise for medical image processing. LeNet and VGGNet have used for comparison on Diabetic Retinopathy (DR) and computed tomography (CT) emphysema data sets. LeNet achieved 72% and 71% accuracies on DR and CT datasets, VGGNet achieved 76% and 78% accuracies on DR and CT datasets. The researchers got these results on CPU based system; they thought that accuracies may be increased on GPU based system.

Pouyanfar, S.,et. al.[20] In this paper the researchers presents in-depth review of traditional approaches and current approaches in image, text and speech processing. They are analyzed on deep learning applications also. Ma, B., et. al. [21]the genetic DCNN design network has proposed, an Autonomous learning algorithm that can create DCNN structure automatically based on the genetic algorithm. This proposed architecture got highest accuracy than AlexNet, VGG, ResNet on MNIST, CIFAR10, CIFAR100 datasets. The proposed architecture got 99.72% on MNIST dataset, 89.23% on CIFAR10 and 66.70% on CIFAR100. Recent, deep learning, the best set of the algorithm was exploited to learn Identify the properties and simultaneous information patterns for active learning [22]. Early studies shown, on the impact of suitable dataset size in transfer learning using CNN, shows that transfer learning would range from: 500 to 1000 images per each class [23]. In medical field, transfer learning with CNN can classify the abdominal ultrasound images very effectively [24]. In [25], Shaha, M., et. al. used transfer learning to compare the performance of fine tuned network VGG19 with AlexNet and VGG16 on three parameters precision, recall and F-score, for performance measuring they used GHIM10K and CalTech256 datasets.

Devikar, P. [26] this can be done separating the network in a specific layer, adding many new ones using a dataset at layers and last network. Lee, S. J.,et. al. [27] to enhance the image identification accuracy, Combine a CNN with Adaboost algorithm. After CNN trained to feature extraction model, Adaboost algorithm is used for ensemble learning. Applications of convolutional neural networks are widely works on object detection and classification [28, 30].A Novel deep learning model was proposed for image classification with an auto encoder and an estimator on MNIST and FERET data sets [29].

Literature survey is summarized in Table 3. This table contains the earlier researchers proposed methods, data sets and how much percentage of data sets is used for training and testing and results they got.

4. DATA SETS

Most of the researcher used the following data sets for their experimental analysis to measure the network performance.

MNIST:

MNIST database, collection of hand-written digits (0-9), the training folder contains 60000 images, and a testing folder contains 10000 images. This is a subset of large data set available in NIST. The size of each image in this dataset is 28x28 pixels.

Caltech101:

Caltech101 is a dataset consists of total 8677 images of 101 type image categories. Each category approximately contains 40 to 800 images. The size of each image is nearly 300x200 pixels.

CIFAR10:

CIFAR10 [18] dataset consists of 10 classes with 60000 images. Each class consists of 6000 images. Official data includes 50000 training image samples and 10000 testing image samples. The size of the each image is 32x32 color images. For example, some of class labels are automobile, bird, cat, deer, dog, horse, ship, truck.

CIFAR100:

CIFAR-100 [18] data set consists of 100 classes containing 60000 images. Each class contains 600 images. Among those images, they are divided 500 training samples and 100 testing samples per each class. These 100 classes are grouped into twenty super classes.

Flower17:

This dataset contains 17 category flower images, with 80 images for each class. These flowers are taken from common flowers in the UK. In these image classes intra class variations and inter class similarities are present.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Scope & Motivation:

From the earlier researchers, image classification accuracy may increased with more number of epochs, large data sets with more number of images per class and another one is GPU based system, compared to CPU systems GPU system will give more accurate performance. The increasing minimum batch size per each iteration also may increase the system performance.

5. CONCLUSION

In this, we have provided a survey of deep learning and its network architectures. We have gone through the various state-of-the-art deep learning architectures. We explained about the pre-trained models, by applying transfer learning. Moreover, we discussed about the some of the popular data sets with large volume of data. From this survey we concluded that the size of data may affect the accuracy and the number of epochs also. Based on this survey motivation we are going to implement face recognition system with accurate performance in minimum time.

REFERENCES

1. Sharma, N., Jain, V., & Mishra, A. (2018). “An Analysis of Convolutional Neural Networks for Image Classification”, *Procedia Computer Science,* Vol.132, pp.377-384.
2. Fukushima &Kunihiko. (1988). "Neocognitron: A hierarchical neural network capable of visual pattern recognition". *Neural networks*, Vol. 1.2 pp. 119-130.
3. Pan, SinnoJialin, &Qiang Yang. (2010)."A survey on transfer learning", *IEEE Transactions on knowledge and data engineering, IEEE,* Vol. 22.10, pp. 1345-1359.
4. LeCun, &Yann, (1998). "Gradient-based learning applied to document recognition", *Proceedings of the IEEE, IEEE*. Vol. 86.11, pp. 2278-2324.
5. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012).”Imagenet classification with deep convolutional neural networks”, *In Advances in neural information processing systems*, pp. 1097-1105.
6. Simonyan, Karen, & Andrew Zisserman (1998)," Very deep convolutional networks for large-scale image recognition", *arXiv preprint arXiv*, pp. 1409.1556.
7. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., &Rabinovich, A. (2015). “Going deeper with convolutions”, *In Proceedings of the IEEE conference on computer vision and pattern recognition, IEEE (CVPR)*, pp. 1-9.
8. Szegedy, C., Ioffe, S., Vanhoucke, V., &Alemi, A. A. (2017). “Inception-v4, inception-resnet and the impact of residual connections on learning”. *In Association for the Advancement of Artificial Intelligence (AAAI),* Vol. 4, p. 12.
9. He, K., Zhang, X., Ren, S., &Sun, J. (2016). “Deep residual learning for image recognition”. *In Proceedings of the IEEE conference on computer vision and pattern recognition, IEEE (CVPR)*, pp. 770-778.
10. Hana D., Qigang Liu , &Weiguo Fan. (2017), “A New Image Classification Method Using CNN transfer learning and Web Data Augmentation”, Expert Systems with Applications, *Elsevier*, Vol. 95, pp. 43-56.
11. Pak, M., & Kim, S. (2017).”A review of deep learning in image recognition”. *In Computer Applications and Information Processing Technology (CAIPT), 4th International Conference on IEEE*, pp. 1-3.
12. Hussain, M., Bird, J. J.,&Faria, D. R. (2018). “A Study on CNN Transfer Learning for Image Classification”. *In UK Workshop on Computational Intelligence, Springer, Cham,* pp. 191-202.
13. Rawat, W., & Wang, Z. (2017). “Deep convolutional neural networks for image classification: A comprehensive review”. *Neural computation*, Vol. 29(9), pp. 2352-2449.
14. Loussaief, S., &Abdelkrim, A. (2018). “Deep learning vs. bag of features in machine learning for image classification”, In 2018, *International Conference on Advanced Systems and Electric Technologies (IC_ASET), IEEE*, pp. 6-10.
15. Maggiori, E., Tarabalka, Y., Charpiat, G., &Alliez, P. (2017). “High-resolution image classification with convolutional networks”. *In Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International, IEEE*, pp. 5157-5160.
16. Ballester, P., &de Araújo, R. M. (2016). “On the Performance of GoogLeNet and AlexNet Applied to Sketches”, *In Association for the Advancement of Artificial Intelligence (AAAI)*, pp. 1124-1128.
17. Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Hasan, M., Van Esesn, B. C., &Asari, V. K. (2018). “The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches”. *arXiv preprint arXiv:1803.01164*.
18. CIFAR10 and CIFAR100 data sets: https://www.cs.toronto.edu/~kriz/cifar.html
19. Dutta S, Manideep BC, Rai S, &Vijayarajan V. (2017) “A comparative study of deep learning models for medical image classification.” *IOP Conference Series: Materials Science and Engineering*, Vol.263, No.4, pp.042097.
20. Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., &Iyengar, S. S. (2018). “A Survey on Deep Learning: Algorithms, Techniques, and Applications”. *ACM Computing Surveys (CSUR),* 51(5), pp.92.
21. Ma, B., & Xia, Y. (2018). “Autonomous Deep Learning: A Genetic DCNN Designer for Image Classification”. *arXiv preprint arXiv:1807.00284*.
22. Ranganathan, H., Venkateswara, H., Chakraborty, S., &Panchanathan, S. (2017). “Deep active learning for image classification”. *In Image Processing (ICIP), 2017 IEEE International Conference, IEEE*. pp. 3934-3938.
23. Soekhoe, D., van der Putten, P., &Plaat, A. (2016). “On the impact of data set size in transfer learning using deep neural networks”. *In International Symposium on Intelligent Data Analysis, Springer, Cham*, pp. 50-60.
24. Cheng, P. M., &Malhi, H. S. (2017). “Transfer learning with convolutional neural networks for classification of abdominal ultrasound images*”. Journal of digital imaging, Springer,* Vol. 30(2), pp.234-243.
25. Shaha, M., &Pawar, M. (2018). “Transfer Learning for Image Classification”. In 2018 *Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE*, pp. 656-660.
26. Devikar, P. (2016). ”Transfer Learning for Image Classification of various dog breeds”, *International Journal of Advanced Research in Computer Engineering and Technology (IJARCET)*, Vol. 5(12), pp. 2707-2715.
27. Lee, S. J., Chen, T., Yu, L., & Lai, C. H. (2018). “Image classification based on the boost convolutional neural network”. *IEEE Access*, Vol. 6, pp. 12755-12768.
28. Aarthi, R., &Harini, S. (2018). “A Survey of Deep Convolutional Neural Network Applications in Image Processing”. *International Journal of Pure and Applied Mathematics*, Vol. 118 No. 7, pp. 185-190.

29. Ososkov, G., &Goncharov, P. (2017). Shallow and deep learning for image classification. *Optical Memory and Neural Networks, Springer*, Vol. 26(4), pp. 221-248.
30. Maggiori, E., Tarabalka, Y., Charpiat, G., &Alliez, P. (2017). High-resolution image classification with convolutional networks. *In Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*, *IEEE,* pp. 5157-5160.