An Empirical Study for the Deep Learning Models

Monika Sethi¹, Sachin Ahuja², Vinay Kukreja³
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India.
E-mail: sethi.monika4@gmail.com

Abstract. Deep Learning (DL) models have tested to be very powerful in solving many hard problems. Especially, those are related to computer vision, text, speech, and classification. However, the blueprint of such models requires large space and elaboration that needs to be examined. Convolutional Neural Network (CNN) is the most popular neural network that can extract the features automatically as compared to conventional machine learning algorithms (CMLA). Our aim in this paper is to lessen the human attempt required to layout architectures by the use of a gadget architecture development process that allows the exploration of huge design space by automating sure version construction, alternative generation, and assessment. The main operations in CNN are Convolution, Pooling, Flattening, Full Connection between the input and output layer. The dataset taken as CIFAR 10 having 60,000 color images of 10 different classes is considered for the study where, various classes represent the images of cars, trucks, frogs, horses, trucks, cats, cars, airplanes, ships, and deer. It is expected that the performance of the CNN model can be further improved by using the deeper network architecture, or by an increasing number of epochs or data augmentation. In this paper, an attempt has been made to explain simple and deeper CNN models on the CIFAR 10 dataset and the comparison has been carried out to check the accuracy achieved from both the models.

Keywords: Deep Learning, Supervised Learning, Convolutional Neural Network, CIFAR 10

1. Introduction
Artificial Intelligence (AI) is vastly growing in both academics and industries in the past few years. CMLA was restricted to process the input raw data in its original form. It was a very challenging task to construct a representation (feature) extractor. It involved years of efforts and too many experts in that field, to convert the raw data in its suitable form (feature vector) from where the learnable machine could learn the different patterns to classify the input [1].

Deep learning (DL) is a subset of machine learning (ML), which is further a subfield of AI, shown in Figure 1. DL is composed of an enormous multilayer network as shown in Figure 2. Each layer consists of artificial neurons (basic unit of artificial neural networks) that are used to extract features automatically without the intervention of human beings. These features are the representation of the raw data (labeled or unlabelled data) that is required for applications like image classification, speech recognition, and many more [1] [2]. With these abilities, DL has achieved remarkable success in diverse areas over ML and AI technologies [3].
DL is categorized into four types named supervised, unsupervised, semi-supervised, and reinforcement learning [4] as shown in Figure 3.

This technique uses labeled data for input to the network [5]. In this technique, the network architecture has several inputs (independent variables) and their corresponding output (dependent variable). For example, in expression $y=mx+c$; $y$ is the dependent variable whose value is dependent upon input variable $x$. $c$ is a biased value that is fixed for all the inputs. The intelligent agent predicts the output $\hat{y}$ and matches it with the actual one $y$. The agent calculates the loss function (error) that is the difference between $y$ and $\hat{y}$ and then alters the network learnable parameters to reduce the cost error function. Post full training, the agent is capable to get the correct output corresponding to input with high accuracy. CNN, Deep Neural Networks (DNN), Recurrent Neural Networks (RNN) including Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) is the different techniques under supervised deep learning.

Unsupervised learning does not use any labeled data rather has only independent variables that are
used as predictors [6]. This algorithm also does not have any dependent variables thus the machine
does not have any data that can be used as a reference while predicting new data. It allows the
machine to predict output without any guidance. So, it can be used to perform more complex
processing as compared to the supervised machine learning Algorithm.

Semi supervised learning technique is a combination of both supervised and unsupervised learning
algorithms. During the learning process, it uses both labeled as well as unlabeled data as input [5], [6].
Deep Reinforcement Learning (DRL) started with Google DeepMind in the year 2013 [7]. The
purpose of DeepMind was to introduce an artificial agent capable to attain the almost identical level of
generalization and performance. Similar to human beings, the agents are capable to learn at their own
to have successful policies or actions used for long-term rewards. This way of learning using trial and
error from hit and miss is termed as reinforcement learning [8]. Like humans, the agents also enhance
their knowledge from the raw data (as input) without hand-crafted feature extraction. As a result, DRL
is the blend of both reinforcement learning and deep learning.

2. Deep Learning CNN for Image Classification:
Convolution Neural Networks (CNNs) is a class of deep learning, has a significant role in analyzing
and interpreting images [9]. In 1970, the basic architecture was already built and used to work with
imagery data in many useful applications. After the breakthroughs in neural networks like the
realization of dropout [10] [11][12], rectified linear units (Relu) [13], and highly computational
graphical processing units (GPUs), CNN's became feasible for complex images classification.
Nowadays, larger CNNs are used to classify a large data set consisting of complex images. In this
study, CNN is used for the problem of classifying images on CIFAR 10 [14].

2.1. Background of CNN (ConvNet)
Similar to neural networks, CNN is consisting of several neurons with learnable parameters (weights)
and biases. Every neuron in the network is subjected to many inputs and then calculates the weighted
sum over them. Post that passes the sum through an activation function (ReLu- generally used in
middle hidden layers) generating an output at the output layer (where the sigmoid function is used for
the classification). The architecture of CNN consists of a stack of layers lying between the input and
output layer as depicted in Figure 4. The hidden layers between the input and output layer of the
ConvNet consist of Convolutional layers, ReLu layer, Pooling (Max pooling), and fully connected
layer. The classification is done at the output layer. Each layer has a set of neurons organized into
three dimensions named height, width, and depth. For example, in CIFAR 10 dataset, the dimension of
each image is 32*32*3 (width*height*depth). Where, width*height is the resolution of the image and
depth (red, green, and blue for color image) is the no of channels. In each layer, the neurons are
coupled to the minor region of the previous layer, rather than to all neurons as in the case of fully
connected.
The basic building block of CNN is similar to the structure of neurons in the human brain and it was inspired by the neurobiological model of the Visual Cortex [15].

**Figure 4.** Key Operations in CNN

2.2. **CNN Model Implementation on CIFAR 10**

Dataset CIFAR 10 is available with 60,000 color images with 32*32 resolutions of 10 different classes. This class represents the images of cars, trucks, frogs, horses, trucks, cats, cars, airplanes, ships, and deer.

These 60000 images are divided into two different datasets as a training set and test dataset. In the training set, 50,000 images are further divided into 5 different batches. In each training batch, 10,000 pictures are randomly selected from the training set. Each picture in training and test dataset is properly labeled with the class name that is for the information from which category the picture belongs and in each batch, there are 10,000 images. The test dataset has 10,000 images with exactly 1000 images from each class.

But on the other hand, in the training dataset although there are 5000 images from each class there is randomness in the number of images from one test batch to another. All the classes are completely mutually exclusive. There is no overlap between the images of two classes.

In this study, two CNN models (Simple and Deeper) on CIFAR 10 [13] are presented. Both models use CNN architecture except that the numbers of hidden layers are different. Only one hidden layer is used in model 1 and model 2 has 4 hidden layers. Each hidden layer performs mainly four key operations that are convolution, max pooling, and flattening, full connection. The basic framework used for both models is shown in Figure 5.

**Figure 5.** CNN Model Framework
The dataset is divided into two-part training set data (83.3%) and test set data (16.67%). The network is trained through CNN using a training dataset and then the prediction is made on the test dataset. The performance is evaluated based on the accuracy achieved on the test data. The architecture used for the CNN model 1 is presented in Figure 6. The number of hidden layers (depends upon experiment) can be increased to enhance the accuracy for the classification prediction at the output layer. The depth of the CNN model is based on the number of hidden layers used; the more hidden layer makes the network more and deeper [16]. The simple and deeper CNN model’s architectures are shown in Figures 6 and 7 respectively.

![Figure 6. CNN Model 1 Architecture](image)
3. Results & Discussion:
When CNN models 1 and 2 implemented for 100 epochs on CIFAR 10, the following results are shown in Figure 8 and Figure 9 is obtained. In CNN model 1, training accuracy after 100 epochs is 55.46% (Figure 8) and in CNN model 2 it is 79.32% (Figure 9), which depicts that the accuracy increases with the number of convolution layers and moreover the total number of parameters learned using CNN model 1 is only 923,914 whereas using CNN model 2 are 1,250,858.
Figure 8. CNN Model 1 Output after each epoch

Figure 9. CNN Model 2 Output after each epoch

The training and validation losses are decreasing after each epoch in both of the models as shown in Figure 10 and 11 respectively.

Figure 10. CNN Model 1 Training loss vs. Validation loss

Figure 11. CNN Model 2 Training loss vs. Validation loss

On the other hand, training and validation accuracy are increasing in model 1 as well as in model 2. But the rate of increase of accuracy is higher in model 2 and it has higher number of convolution layers are more (i.e. model 2). The trained model is evaluated on the test dataset and retrieves an accuracy of 52.67% and 78.10% for CNN model 1 and model 2 respectively shown in Figure 12 and 13.
4. Conclusion:
In deep learning, classification can be categorized as supervised or unsupervised learning, whereas in supervised technique the network learns from the input data/images fed into it and applies that learning to classify the new observation. CNN is a particular group of an artificial neural network under deep learning that operates on the concept of neurons in human beings. In the paper, the researchers fitted the simple and deeper CNN model on CIFAR 10 dataset. It is identified that the CNN model’s performance is further improved by executing the model over many epochs, image data augmentation, and using the deeper network architecture. The researchers have successfully explained simple and deeper CNN models on CIFAR 10 and carried out the comparison to check the accuracy achieved from both the models. Therefore, a conclusion can be drawn that deeper CNN performs better in terms of the classification prediction to the Simple CNN model.

References
[1] Lecun Y, Bengio Y, and Hinton G, 2015 Deep learning Nature, vol. 521, no. 7553, pp. 436–444.
[2] Schmidhuber J, 2015 Deep Learning in neural networks: An overview Neural Networks, vol. 61, pp. 85–117.
[3] Arel I, Rose D, 2010 Deep machine learning-a new frontier in artificial intelligence research [research frontier]. IEEE computational intelligence magazine, 5(4), 13-18.
[4] Alom M. Z. Taha, T M, Yakopcic, C, Westberg, S, Sidike, P, Nasrin, M S., and Asari, V. K. 2018 The history began from alexnet: A comprehensive survey on deep learning approaches. arXiv preprint arXiv:1803.01164.
[5] Zhu, X. and Goldberg, A. B. 2009 Introduction to semi-supervised learning. Synthesis lectures on artificial intelligence and machine learning, 3(1), 1-130.
[6] Chapelle, O., Scholkopf, B. and Zien, A., 2009. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. IEEE Transactions on Neural Networks, 20(3), pp.542-542.
[7] Wu, I. C., Lee, C. S., Tian, Y., and Müller, M. 2018. Guest Editorial Special Issue on Deep Reinforcement Learning and Games. IEEE Transactions on Games, 10(4), 333-335.
[8] Kaelbling, L. P., Litman, M. L., & Moore, A. W. 1996 Reinforcement learning: A survey. Journal of artificial intelligence research, 4, 237-285.
[9] Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2017. Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.
[10] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014 Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.
[11] Rice, L., Wong, E., & Kolter, Z. 2020 Overfitting in adversarially robust deep learning. International Conference on Machine Learning, pp. 8093-8104.
[12] H. Wu and X. Gu 2015 Max-pooling dropout for regularization of convolutional neural networks Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9489, pp. 46–54.
[13] A. F. Agarap Deep learning using Rectified Linear Units (ReLU) arXiv preprint arXiv:1803.08375.
[14] CIFAR 10 Dataset https://www.cs.toronto.edu/~kriz/cifar.html, accessed on 30/11/2020 at 11:30 AM.
[15] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998 Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
[16] Ba, J., & Caruana, R. 2014 Do deep nets really need to be deep? Advances in neural information processing systems, 27, 2654-2662.