Learning with Limited Sample

Prithvi Raj Singh
Center for Advance Computer Studies
University of Louisiana at Lafayette
Lafayette, US
prithvi.singh1@louisiana.edu

***Abstract*—Rise of powerful Neural Networks have also created a need for significantly large dataset to train and validate the model. Many large corporation have spend hundreds of millions in data collection rather than working a way around developing algorithm to work with small data set. While some data are easy to collect, there are many real world problems that have very few data available, medical imaging, tumor detection being the one of them. In this paper, we explore how different Deep Learning tools can used on problems that has comparatively limited data set. Here we have discussed and examined the use and efficiency of different Neural Networks on image classification problem with small data set.For this paper we are focused on image classification problem with small training size. We explore the use of Conventional neural networks like CNN, and modified techniques like Transformers Compact Convolutional Transformer (CCT) and Vision Transformer (ViT) to solve an image classification problem. The CCT gives an accuracy of 98% on all of the four datasets. Transfer Learning and Data augmentation techniques used on CIFAR-10 and CIFAR-100 dataset also gives better result in comparison to regular dataset. Our objective is to find a Neural Network that can work on problem with limited data producing as less error as possible.**

***Index Terms*—Data set,Transfer Learning, Data Augmentation, training size, CNN, CCT, Transformers**

I. Introduction

Machine Learning models rely heavily on the availability of training data set, not a small amount of data set, but rather a huge data set. As we make progress in deep learning, we are also required to have huge data for the Neural Network to train on. The training, testing, and validation accuracy of all the Deep Learning model depends on the size of the data set we feed it. The AI/Machine Learning community pays huge attention to big data, but recently the concerns about unavailability of large data set for certain domains like medicine, physical science have forced the research community to work with small data. Deep Learning models are referred to has 'Data-hungry' model. Many prominent AI research organizations like Google, Facebook, IBM spend hundreds of millions of dollars in collecting large data, often unethically. This 'data-hungry' initiatives to build deep learning models is very concerning to the actual progress of AI. Data isn't readily available, and certain areas of science that can make good use of Deep Learning systems like cancer, autism detection, astronomy don't have sufficient data to feed the insatiable deep neural networks.Since many events in medicine or physical science are far more rare, collecting data or assigning a proper labels to data is very difficult.

Link to the GitHub Repo of the author.

Data collection for fields like neuro-imaging, genomics, motion tracking is difficult also because of intrinsic high cost. Inability to work with small amount of datasets has limited the scientific community in exploring different tools and model they can use [3]. Large datasets relates to the need of large computational resource to process the data and neural network. Large computational resources aren't available for many scientists, hence they can't critic or assist in the verification of 'state-of-art' machine learning algorithms subsequently slowing down the rapid advancement of AI [3]. Human beings started the investing into AI with the dream of being able to create a human like intelligence, Artificial General Intelligence (AGI). Human beings are very capable of learning new things with two or three examples. Until we can make the machine learning algorithms capable learning like human, we can't make progress towards AGI. In recent year, researchers for several field have put in lot of effort in developing neural network that can produce great results with smaller data set. Recent advancements have improved the computer vision, natural language processing, image classification problems for smaller data sets.

Many researcher, since the advent of Transformer for NLP have developed transformer that can perfrom quite well on certain datasets like Cifar-10, Cifar-100, MNIST. Although conventional neural networks perform very well on one datasets they don't necessarily perform better on other. Convolutional and Pooling layers, allow the CNN model to have great efficiency[3]. CNNs became the gold standard in computer vision since the sucess of AlexNet[3]. CNNs are know to be the best performer in the area of image classification with limited sample since they require less data and time to train. Many scientist have improvised the CNN model with data augmentation, transfer learning in order to get good efficiency.Though CNN and Transformers have shown desirable qualities for statistical inference and prediction, each comes with their limitation[3]. Many variation of transformers like ViT, CCT are introduced in last few years that have shown great results in the NLP, image processing tasks. Compact Convolution Transformer (CCT) introduced in [3],outperforms many existing variation of transformers and CNNs on certain datasets like CIFAR-10, CIFAR-100, Fashion-MNIST. Recently Recurrent Neural Network (RNN) have also been used for image classifcation problem with good sucess.

In this paper we have used CNN and transformers for image classification problem on datasets like CIFAR-10, CIFAR-100, Fashion-MNIST. The implementation of transformers on mentioned datasets shows enormous increase in the training and validation.The CCT showed top-5 accuracy of 98.2% on CIFAR-10 dataset and top 5 accuracy of 80% on CIFAR-100 dataset which is twice better than accuracy of CNN. In this paper, we have ran the existing model of transformers, CNNs with some variation to evaluate their performance figure out what might be needed to make the deep learning algorithms work better with small dataset.

Both conventional CNN and CCT performed much better on the MNIST and Fashion-MNIST datasets. The presence of smaller number of classes (10) and grayscale images (single channel) increased the training and testing efficiency on the MNIST and Fashion-MNIST dataset.CNN model can give up to 99% accuracy on Fashion-MNIST similar to CCT. Transformers model scalable while maintaining the compactness and computational efficency[Hassani]. Even though Transformers, CNNs have shown great sucess in image classification and NLP tasks, they aren't equally efficient on medical imaging and detecion task. The main theme of this paper is centered around:

- Evalauting the CNN and Transformer models on different datasets to comapre their performance on smaller datasets like CIFAR-10, MNIST.
- Breifly exploring the CCT, and ViT for image classfication problem.
- Exploring what can be done in-order to better the deep learning algorithms to work with even smaller dataset.
- Apply Data Augmentation and Transfer Learning method to check the efficiency of CNN on CIFAR-10, CIFAR-100 dataset.

II. Preliminaries

A. *Convolution Neural Networks*

Convolutional Neural Network (CNN) is the state of-the-art for image classification task. Convolutional Neural Network (CNN or ConvNet) is a especial type of multi-layer neural network. A typical CNN is composed of single or multiple blocks of convolution and sub-sampling layers, after that one or more fully connected layers and an output layer as shown in figure. A convolutional neural network is a feed-forward neural network, often with up to 20 or 30 layers. Convolutional neural networks contain many convolutional layers stacked on top of each other, each one capable of recognizing more sophisticated shapes. The architecture of a convolutional neural network is a multi-layered feed-forward neural network, made by stacking many hidden layers on top of each other in sequence. It is this sequential design that allows convolutional neural networks to learn hierarchical features. Convolution layer, Pooling layer, Fully Connected Layer, Activation function.

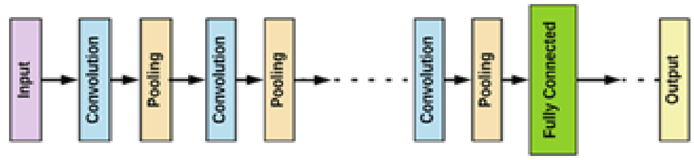


Fig 2: General representation of CNN Model.

B. *Transformer*

Transformer architecture was first introduced in 2017 by Google through 'Attention is all you need', where they implemented transformer on language translation. Several versions of the transformer have been implemented since then, particularly on NLP tasks. Transformers like Vision Transformer (ViT), Compact Convolutional Transformer (CCT) are good at image processing task. A Transformer model is made up by stacking Transformer blocks. Each Transformer block has two components: 1) a multi-head self-attention layer and 2) a token-wise feed-forward (MLP) layer. The input to these models is a sequence of vectors. These are usually embeddings of an input token sequence. The self-attention layer updates these embeddings by computing pairwise dot product attention between the input embeddings. Both layers use layer normalization and skip connections.

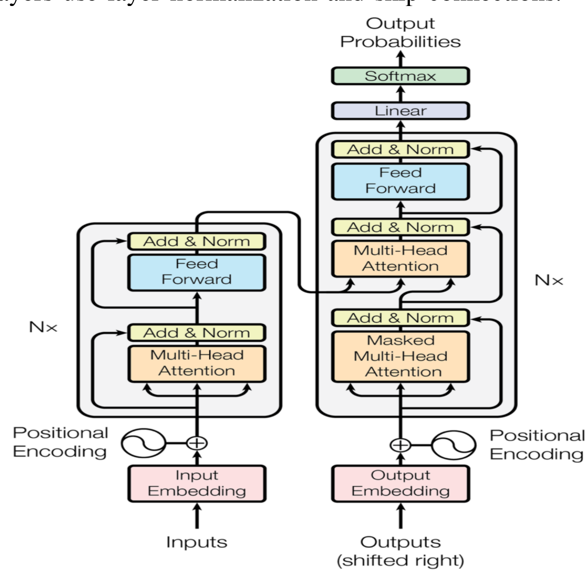


Figure 1: The Transformer - model architecture.

Vision Transformer (ViT): ViT is based on the transformer architecture originally designed for NLP task. The ViT uses image pre-processing layer which partitions the image into a sequence of non-overlapping patches followed by linear projection [Srinadh et.al]. ViT was introduced as competitor to the CNN. ViT is has less inductive bias.

Compact Convolutional Transformer (CCT): The CCT improved version of Compact Vision Transformer (CVT) which is in itself the improved version of Vision Transformer. CCT uses sequence pooling with convolutional embedding in the input layer which gives better inductive bias[3].

C. Transfer Learning

Transfer Learning involves using larger data set to train the neural network and store the knowledge gained. We reuse the stored network for different but related classification problem. We freeze the lower layers since they are trained using very large data set and replace the top classification layers. It enables the use of pre-trained models as starting point for Computer Vision and NLP tasks which require heavy computational power and larger data to obtain the optimal training and testing accuracy. Transfer Learning allows rapid progress on the performance when modeling the second neural model. Transfer Learning in Deep Learning is done through inductive transfer. Inductive transfer outweighs the model bias by allowing the model to fit different but related problems.

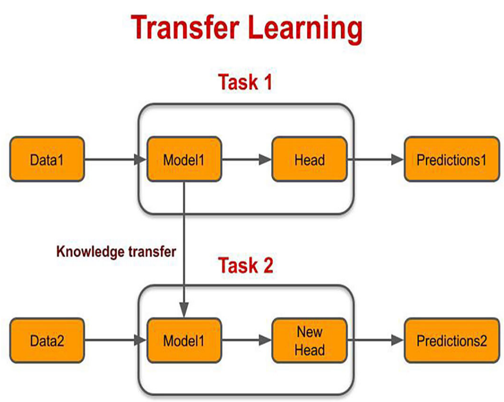


Fig 3: General representation of Transfer Learning

D. Data Augmentation

Data Augmentation is crucial in getting more training samples if we have very limited data to work with. Data Augmentation allows to increase the size of data set by doing simple changes to the existing data. We add different filters or slightly change the image by zooming in or out, rotating the image by certain angle, blurring the image, flipping, changing the brightness of the image data. Data Augmentation can help the CNN model achieve higher accuracy as shown in our experiment. Compact Convolutional Transformer (CCT) make a good use of data augmentation strategy to achieve near perfect accuracy on certain image classification problems.

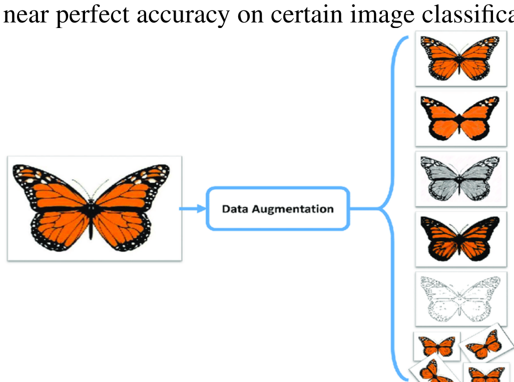


Fig 4: Representation of Data Augmentation

III. Methodology

In order to explore different models being used for learning with limited samples, we will run all the models (CNN, CCT, ViT) on different datasets and compare their efficiency. We also use the Data Augmentation and Transfer Learning techniques on few of the datasets to see whether CNN's efficiency can be increased on datasets where it performs poorly. Conventional CNN performs remarkably well when applied to MNIST and Fashion-MNIST dataset. Compact Convolutional Transformer outperforms CNN is all of the datasets we have examined. CCT can make remarkable breakthrough in the field of computer vision and NLP tasks. Proper dissection of CCT on NLP tasks is yet to be done, but we believe that CCT can perform with optimal accuracy on any NLP tasks. In this paper, we will focus on the use of CCT, CNN as the modern state of art architecture for image classification problem.

In order to optimize the CNN model perfromance on the CIFAR-10/100 datasets, we have applied data augmentation technique on both of the dataset. Transfer Learning has also shown some improvement in the use of CNN on CIFAR datasets. It is important to mention that using a dropout layer can also somewhat increase the validation accuracy of CNN by nearly 4%. Transfer Learning, Data Augmentation and Cosine loss function are three of the widely used techniques to deal with small dataset. For our experiment we have only focused on data augmentation and transfer learning. Cosine loss function can increase accuracy by almost 30% when used in place of cross-entropy loss[10].

IV. Experiments

The experiments were conducted on the free version Google Colab with GPU runtime. Google Colab offers great computational power of GPU with Disk space of 78 GB, and RAM of nearly 13 GB. GPU runtime provides great efficiency when running the Deep Neural models. The experiment as mentioned was conduceted on follwing datasets: CIFAR-10, CIFAR-100, MNIST, and Fashion-MNIST. These are some of the benchmarking datasets for Machine Learning problem. These datasets contain small amount of training samples with small resolution.

The MNIST and Fashion-MNIST dataset consist of gray scale image, a single channel, which reduces the information density. CIFAR-10 dataset consists of 60000 samples of 32x32x3 size divided into 10 diffrent classes with 6000 samples in each class. CIFAR-100 dataset consists of 60000 samples with 100 classes and 600 samples per class. The MNIST and Fashion-MNIST datasets consist of 60000 training images, and 10000 test images each of 28x28x1 size divided in 10 classes. MNIST is the dataset of hand written digits while Fashion-MNIST is the dataset of fashion and clothing items.

In all the experiments, we conducted a hyperparameter search for every different method and report the best obtained result of each method.

A. Performance of CCT on Datasets

We used the CCT on the four datasets available to compare the efficiency of CCT. The model performs remarkably well on all the four datasets, giving training accuracy in upper 90% all the time. CCT outperforms the CNN on CIFAR-100 dataset by more than double margin. CCT outperforms CNN on CIFAR-100 dataset by close to 40% margin.

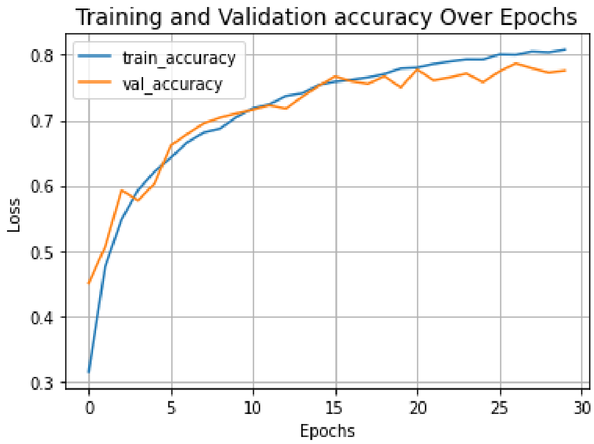


Fig 5: Train and Val Accuracy of CCT on Cifar-10.

We can see that the CCT model gives training accuracy of 80% and validation accuracy of 78%. The loss metrics of the model is quite bad. The model gave significant loss in both training and validation, about 93% and over 100% respectively.

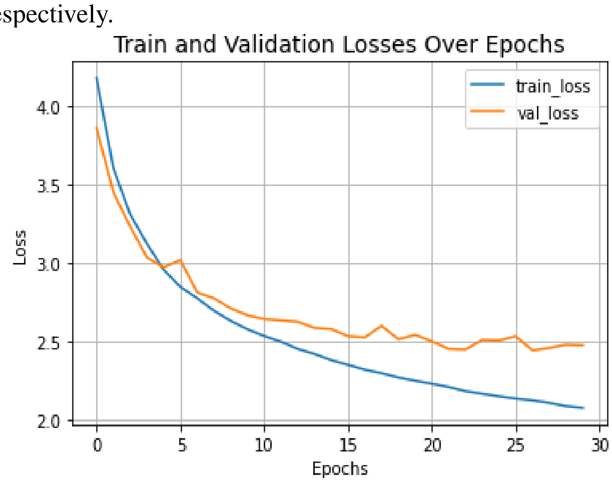


Fig 6: Representation of Losses of CCT on CIFAR-100.

From the chart above we can see that the training loss has went down steeply over the 30 epochs. The validation loss went down from initial epoch till 10 epochs and stayed closely similar for remaining epochs. The validation and training loss doesn't seem to have over-fitting problem even though the chart shows steep decline in training loss.

From above chart we can see that the training accuracy and the validation accuracy appropriately fit the plot. The CCT model achieves top training accuracy of 87%, top 5 validation accuracy of 80%. The model has a training accuracy of 58%, and validation accuracy of 50% over 30 epochs. All of the accuracy mark increase appropriately over epochs.

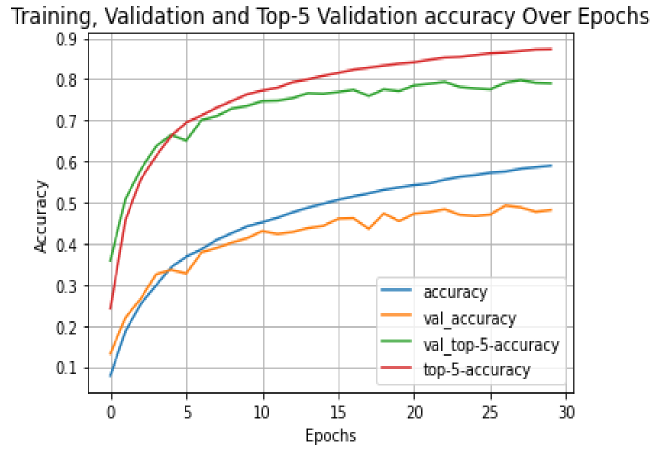


Fig 7: Performance of CCT on CIFAR-100.

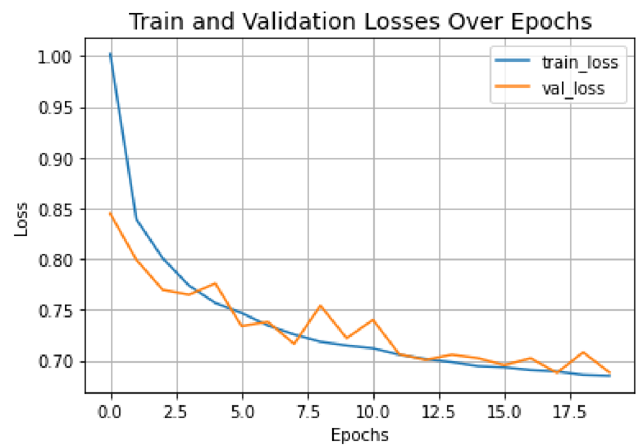


Fig 8: Loss metrics of CCT on Fashion-MNIST.

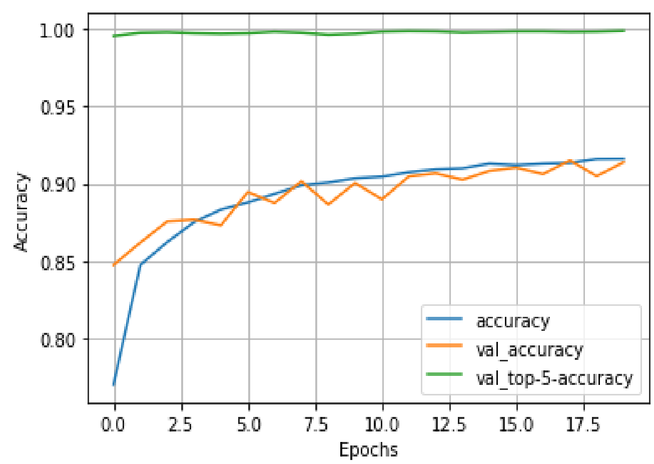


Fig 9: Representing accuracy of CCT on Fashion-MNIST.

The CCT shows great accuracy on CIFAR-10 and Fashion-MNIST dataset with test accuracy rate of 78% and 91% respectively. The CCT model gave validation accuracy of 77% with top 5 validation accuracy of near 98%, and validation accuracy of 91% with top 5 validation accuracy of near 100% when working with CIFAR-10 and Fashion-MNIST datasets respectively.

B. *Performance of CNN on Datasets*

CNN, which is the state of art image classification model, performs exceptionally well on MNIST and Fashion-MNIST dataset. CNN model fairly well on CIFAR-10, but shows worse performance on CIFAR-100 dataset.

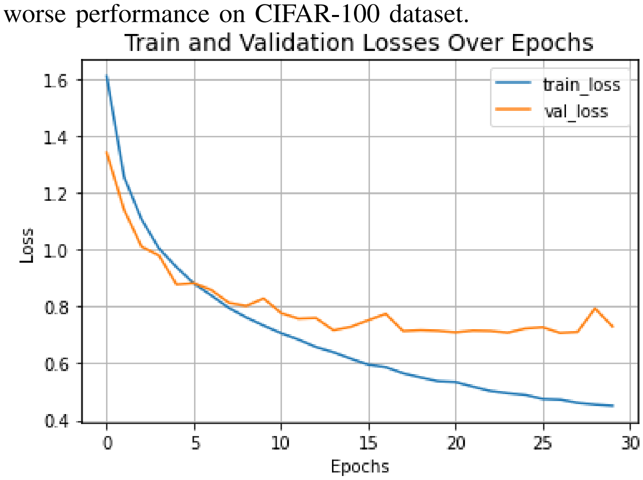


Fig 10: Losses of CNN model on CIFAR-10 dataset.

The CNN model gives around 45% overall loss on training data and top validation loss of 79% running 30 epochs. Addition of regularization technique and dropout in fully connected and dense layers can improve the performance metrics of CNN while working on either CIFAR-10 or CIFAR-100.

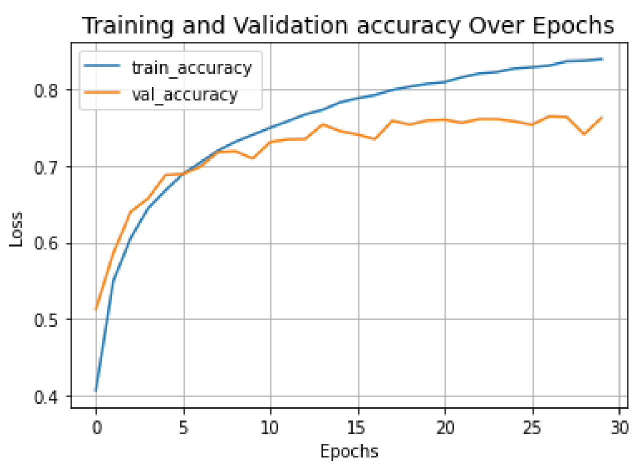
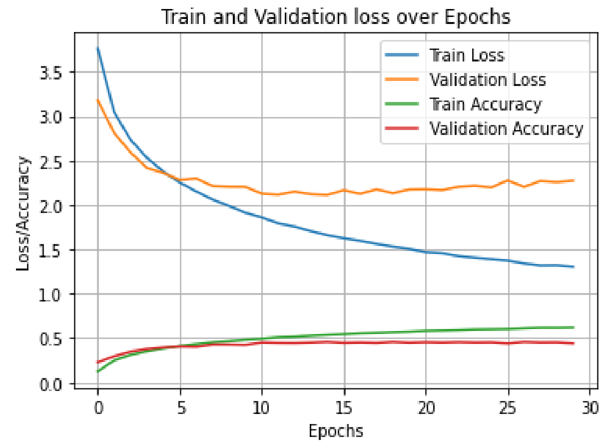


Fig 11: Accuracy of CNN model on CIFAR-10 dataset.

The CNN model have a overall test accuracy of 76% when used on CIFAR-10 dataset when the dropout layer is added in the dense layer. The model has highest training accuracy of 84% and validation accuracy of near 76%.

The CNN model shows really bad performance on CIFAR-100 dataset. The model can only achieve about 83% training accuracy, and testing accuracy of 44%. There is significant test loss, but the train loss is acceptable. The number classes in the CIFAR-100 dataset makes the conventional CNN model fail to properly train or validate. There is significant difference in the validation and training loss which indicates overfitting prob-



lem.

Fig 12: Loss and Accuracy of CNN model on CIFAR-100. We can see that the CNN on CIFAR-100 produces significant overfitting on training and validation.

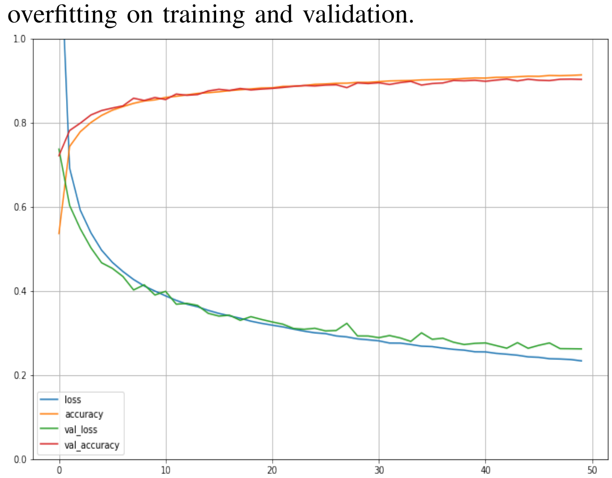


Fig 13: Accuracy and Loss of CNN on Fashion-MNIST.

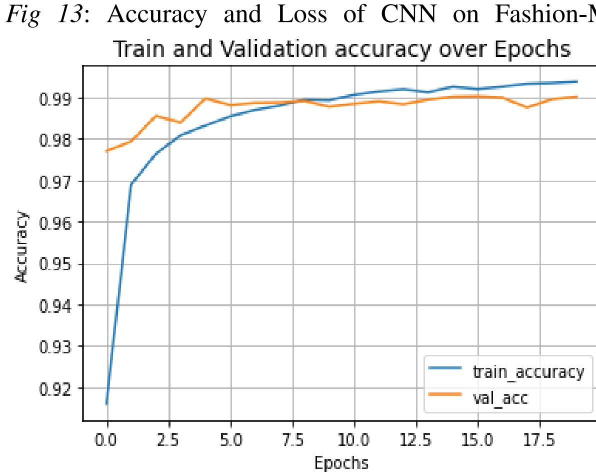


Fig 14: Training and Val Accuracy of CNN on MNIST.

The conventional CNN model works remarkably well on MNIST and Fashion-MNIST dataset.As seen in the chart the CNN model produces no overfitting in either testing or training on either dataset. The model produces near 99% testing accuracy with validation loss of only 5% on MNIST dataset. The Fashion-MNIST dataset produces testing loss of 26% but shows good testing accuracy of 90%.

C. *Classification using Vision Transformer (ViT)*

We used the Vision Transformer model to classsify images in the CIFAR-10, CIFAR-100 dataset. Although Vision transformer model doesn't give highly significant accuracy, it gives notable improvement in the training and validation accuracy.

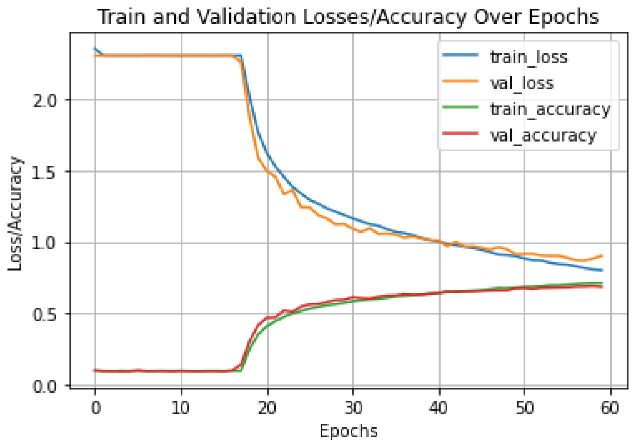


Fig 15: Train and Validation metrics on CIFAR-10.

The ViT model gives test accuracy of 68% and top 5 validation accuracy of 97%. No significant improvement can be seen in comparison to CCT or Conventional CNN, but the ViT model produces appropraite fitting. The training and validation metrics are close to each other.

When used with CIFAR-100 dataset, the ViT model,run through 100 epochs, produces test accuracy of 55% with top 5 test accuracy of 82% which is significant improvement over CNN. The model, however, produces overfitting with a significant validation loss. The model doesn't produce competitive result on CIFAR-10 dataset. ResNet50V2 implemented from scratch can give near 68% test accuracy. Though not guaranteed, we can slightly improve the performance by increasing the number of transformer layer, learning rate, weight decay.

D. *Data Augmentation and Transfer Learning*

Data Augmentation is one of the way we can increase the sample size of our data and improve the network efficiency. Tensorflow has in-built library 'ImageGenerator' that can generate images of different orientation from the provided dataset. The image generator augments the data and than replaces the original data with the augmented data. To check the usage and efficiency of data augmentation technique we used only CIFAR-10 and CIFAR-100 datasets. Transfer Learning is done with the pre-trained model like ResNet50.

The augmented data enables the CNN model to produce appropriate fitting. The validation and training metrics stay close to each other. The regular CNN model with cifar10 dataset produced slight overfitting on all metrics. The data augmentation allows the CNN model to give test accuracy of 80% with validation loss of 59%. Despite not giving any significant improvement over unaugmented data. The CNN model is better trained and validated with augmented data.

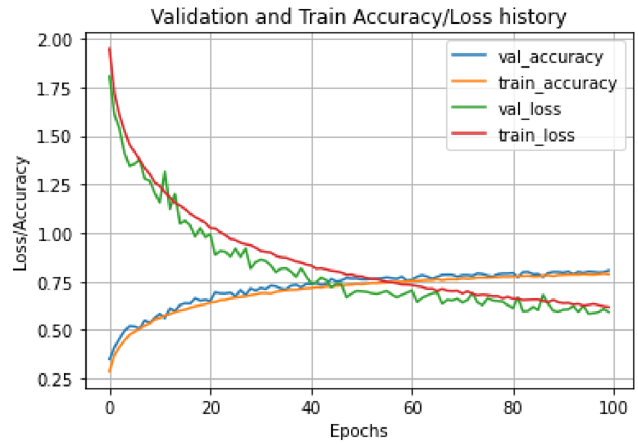


Fig 16: Performance of CNN on Augmented CIFAR-10 data.

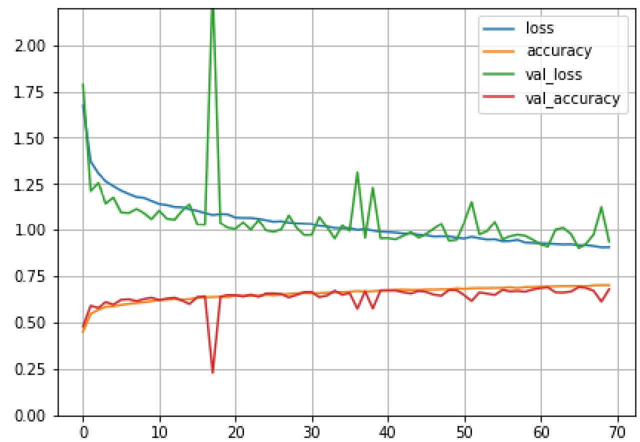


Fig 17: Transfer Learning Metrics on CIFAR-10.

Transfer Learning applied to CIFAR-10 dataset produces significant improvement than the CNN model. The model, despite only having 70% validation accuracy, gives optimal fitting of the metrics.The model produced max validation loss and accuracy at 18th epoch.

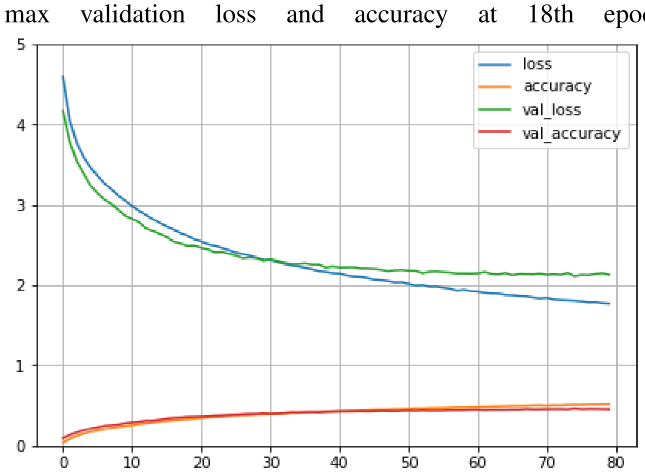


Fig 18: Transfer Learning Metrics on CIFAR-100.

Transfer Learning method don't give competitive result on CIFAR-100, but the validation accuracy and train accuracy stay close to each other. The validation accuracy is only 46% with significant validation loss of 213%.

V. Result and Analysis

Experiment was designed to find out the training and validation accuracy of CCT, CNN, ViT on four datasets. The conventional CNN give optimal accuracy when used on MNIST and Fashion-MNIST dataset however it doesn't give acceptable accuracy when used on CIFAR-10 and CIFAR-100 datasets. It is worthy to note that the presence of single channel (grayscale)images in the MNIST and Fashion-MNIST dataset helps a lot in improving the CNN's training and validation accuracy. If we are feeding in dataset with huge number of classes and small number of images per class like in CIFAR-100, the CNN model fails to give any considerable accuracy. Modified CNN like ResNet19 give optimal accuracy.

Transformer models like CCT, ViT give significantly better results in comparison to CNN. Transformer models, especially CCT, outperforms all the model and techniques we used for our image classification task. Unlike CNN, CCT model was able to increase the validation accuracy by 15% working on CIFAR-100 dataset. CCT has aslo give optimal validation accuracy on MNIST and Fashion-MNIST dataset with no significant loss like the CNN model.

Transfer Learning and Data Augmentation techniques are widely used for increasing the sample size of dataset when there is less data. There has been several implementation of these techniques on different datasets. These techniques when used properly can produce maximum accuracy with very little loss of information. We weren't able to achieve what we expected of it. There are several reasons for our failure to get the optimal metrics. We can achieve higher accuracy by increasing the number of epochs, changing the learning rate.

Computational resources is key to getting the Data Augmentation and Transfer Learning method to work. The CNN model when used with data augmentation technique needs hours to run about 80 epochs. Transfer Learning method, where we have used the pre-trained ResNet50v2 model, takes about two hours on GPU accelerated Google Colab to give result which might not be significant as in our case. Lack of faster GPU prohibited the frequent change and use of network layer for training. Access to a good computational resource is most for testing out deep neural networks.

Pre-training the model with a established network that has been trained on millions of parameter can accelerate the learning rate of any neural network. We can also improve the model's efficiency by increasing the number of epochs, increasing number of layers in transformer layer, changing the learning rate, the weight decay. Fine tuning a model using a large pre-trained high resolution dataset can also improve the overall efficiency of network. Having high number of epochs won't necessarily increase the accuracy metrics, often after reaching to certain level of validation and testing accuracy, the model stays stagnant.

It is essential to remember that every time we train and test the data, you will see different loss function cost and training accuracy. It is because the neural network chooses different weight and biases per iteration. We saw that sometime rerunning the same model increased the validation and training accuracy by 5-6%.

Metrics				
Model	CIFAR-10	CIFAR-100	Fashion-MNIST	MNIST
CNN	76.31%	44.07%	90.29%	99%
CCT	98.74%	50.05%	91.4%	-
ViT	68.76%	55%		
DA	80.6%			
TL	68.09%	45.46%		

Table Representing the validation accuracy

VI. Conclusion

For a development of human like intelligence, we most reduce the need of many samples to train the model for ordinary task. Deep Learning algorithms most learn how and what to do with as small data sample as possible. In this paper, we have shown that there are different techniques that can significantly improve the image classification problem even with small dataset, all of which come at different cost. Computational power is much needed for some of the model to work properly, but the transformer models can work perfectly with minimal hardware. Although, the research community has put huge emphasises on '*Big-Data*', it is imperative that we put equal emphasis on '*Small Data*'. Efficient network like Compact Convolutional Transformer are much needed to work on small dataset and computational power. The AI research community most continue the research for building algorithm that can perform different tasks like NLP, Image classification, object detection with the need of large sample data or computational resource. There is a pressing need for an algorithm that can perform optimally in various scientific domain like Medicine, Physical sciences. As seen from our experiment, not all model performs better on all dataset and it will be difficult to build a model that can work better on all tasks.

VII. Future Work

Nothing goes as expected, and that's true in our case too. We conducted some of the experiment with belief that it will give out the optimal results we are looking for, but we were disappointed. Because of the time constraints we couldn't subsequently make any more changes to our model to see whether it works or not. The transfer learning and data augmentation techniques is proven to increase the efficiency and effectiveness of CNN model, but our experiment didn't prove that theory, despite having no overfitting problem. We can improve the performance of these techniques by changing some parameters and running it for 200 epochs. Our GPU power resource was limited so we didn't attempt to redo the whole training and testing. With availability of proper hardware, anyone referencing our work can change certain parameters and get significant validation accuracy, about 88% with transfer learning. Our experiment proves the need for algorithm that can run fast on small GPU power. We would also like to explore the use of Cosine loss function on all four datasets to check it's efficency. Some researchers have implemented cosine loss function on the CUB, stanford cars, MIT indoor scence, and have been able to get significantly higher accuracy without having to use pre-trained model[10].

Acknowledgment

We mostly like to thank Dr. Maida for his tutorial on transformers and covering most of the used topics in this paper. We would like to express our sincere gratitude to the free version of Google Colab that allowed us to use GPU power for some experiments. We would also like thank authors of the few articles we read for the implementation of neural network. Image processing class by Dr. Chu also helped a lot with understanding transfer learning and data augmentation for image processing tasks.

References

[1] Bhojanapalli, Srinadh, Ayan Chakrabarti, Daniel Glasner, Daliang Li, Thomas Unterthiner, and Andreas Veit. "Understanding robustness of transformers for image classification." arXiv preprint arXiv:2103.14586 (2021).

[2] Vabalas, Andrius, Emma Gowen, Ellen Poliakoff, and Alexander J. Casson. "Machine learning algorithm validation with a limited sample size." PloS one 14, no. 11 (2019): e0224365.

[3] Hassani, Ali, Steven Walton, Nikhil Shah, Abulikemu Abuduweili, Jiachen Li, and Humphrey Shi. "Escaping the big data paradigm with compact transformers." arXiv preprint arXiv:2104.05704 (2021).

[4] Li Fei-Fei, R. Fergus and P. Perona, "One-shot learning of object categories," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 4, pp. 594-611, April 2006, doi: 10.1109/TPAMI.2006.79.

[5] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[6] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. Deep Learning. Cambridge (EE. UU.): MIT Press.

[7] Zhuang, Fuzhen, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. "A comprehensive survey on transfer learning." Proceedings of the IEEE 109, no. 1 (2020): 43-76.

[8] Yuan, Li, Yunpeng Chen, Tao Wang, Weihao Yu, Yujun Shi, Zihang Jiang, Francis EH Tay, Jiashi Feng, and Shuicheng Yan. "Tokens-to-token vit: Training vision transformers from scratch on imagenet." arXiv preprint arXiv:2101.11986 (2021).

[9] Barz, Bjorn, and Joachim Denzler. "Deep learning on small datasets without pre-training using cosine loss." In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1371-1380. 2020.

[10] Wilson, James, and Paul Daugherty. 2021. "Small Data Can Play A Big Role In AI". Harvard Business Review. https://hbr.org/2020/02/small-data-can-play-a-big-role-in-ai.

[11] "Machine Learning With Limited Data". 2021. Assets.Publishing.Service.Gov.Uk. https://assets.publishing.service.gov.uk

[12] "Tf.Keras.Preprocessing.Image.Imagedatagenerator". 2021. Tensorflow. https://www.tensorflow.org/.

[13] Folkman, Tyler."How To Use Deep Learning Even With Small Data". Medium. https://towardsdatascience.com/how-to-use-deep-learning-even-with-small-data-e7f34b673987.

[14] "Transfer Learning And Fine-Tuning—Tensorflow Core". 2021. Tensorflow. www.tensorflow.org/tutorials/images/transfer$_learning$.

Link to the GitHub Repo of the author. Contains codes and will be updated in future.