Analysis of Accuracy and Efficiency in Standard CNN and Highway CNN Techniques

Moreau J.T.
Department of Ophthalmology, Bordeaux Vision Sciences Institute, Bordeaux, France

KEYWORDS: Alexnet, Highway CNN, Evolutionary Highway CNN, CNN.

ABSTRACT
In recent years, convolutional networks have shown breakthrough performance in image classification and detection. The main reason behind the performance of convnets is that they are inspired from the mammal's visual cortex. In this paper, we have investigated the performance of four models that are Alexnet, Highway Convolutional Neural Network, Convolutional Neural Network and an evolutionary approach on highway convolutional neural network on the basis of train loss, test loss, train accuracy and test accuracy. These models are tested on two datasets that are WANG dataset and Simpsons dataset. In WANG dataset, Alexnet model achieved the highest test accuracy of 0.2625 and the highest train accuracy of 0.2193. Evolutionary Highway CNN has the least train loss of 0.1599 and CNN has the least test loss of 0.1604. In Simpsons dataset, Evolutionary Highway CNN has the Highest test accuracy of 0.5780. Highway CNN has the highest train accuracy of 0.5662 and for loss domain; Evolutionary Highway CNN has the least train and test loss

INTRODUCTION
This paper discussed about Alexnet, Convolutional Neural Network, Highway Convolutional Neural Network and Evolutionary Highway Neural Network. All these networks are tested with two different data sets.

A. Alexnet & Algorithm
This neural network was coined by three people who are Alex Krizhevsky et al.[1]. This was best performed CNN in 2012 as it won ILSVRC - 2012 competition and achieved a winning top-5 test error rate of 15.3% where top-5 error means that the rate at which given an image, the model does not output the correct label with its top-5 prediction. The next 26.2% achieved by the second-best entry. The network was tested on Imagenet dataset. The Imagenet dataset consists of 15 million high resolution images and having around 22000 categories. The network's first convolutional network having 96 convolutional filters of size 11 x 11 x 3 and having strides of 4 pixels followed by pooling layer and normalization layers which is called as local response normalization. The pooling layer is called as overlapping pooling because the pooling units are spaced by $s = 2$ pixels having a size of$z = 3$ pixels. As $s < z$, overlapping pooling is obtained. This scheme reduces the top-1 and top-5 error rate by 0.4% and 0.3%. Local response normalization is used after the pooling layer because local response normalization is based on the concept of Lateral Inhibition which means the capacity of excited neuron to subdue its neighbours. The activation used in convolutional layer is ReLU which have unbounded activation and thus local response normalization layer is useful when dealing with ReLU activation function. The same process is continued another time but now the second convolutional layer has 256 convolutional filters of size 5 x 5 x 48. Now the third, fourth and fifth convolutional filters are arranged one after the another where third convolutional layer has 384 filters of size 3 x 3 x 256, fourth convolutional layer has 384 filters of size 3 x 3 x 192 and fifth layer has 256 filters of size 3 x 3 x 192 and then overlapping pooling and local response normalization layer. At last one fully connected layer has 4096 neurons with dropout function and another fully connected layer having 4096 neurons with dropout function. Dropout function is used reduce over fitting. Both the fully connected layer uses the ReLU activation function. At last outer layer having softmax activation function is used for multiple classifications. The data was first augmented before training.

B. Convolutional Neural Network& Algorithm
The name convolutional was first coined by LeCun et al.[6]. The convolutional network was inspired by the mammal's visual cortex whose model was proposed in the research by D.H.Hubel and T.N.Wiesel. So, in a Convolutional Neural Network, first the convolutional layer sees the image by scanning one complete image through receptive field. Through the receptive field, the convolutional layers extract the most important features from the image. After the important featureS are extracted, it is then written to the activation map. Now this activation map is subjected to the pooling layer for down sampling process. Down sampling means to reduce the size of the image to get the best extracted feature and each convolutional layer consists of ReLU activation function. The ReLU activation function is used for the normalization purpose. After that we flatten the outputs which came from the pooling layer into small dimension vectors through the flatten layer. Now it is processed to the fully connected layers for the classification tasks. The output layer consists of the softmax function for

classifying multiple or more than two cases.

Step 1. *The input is feed into the input layers.*
Step 2. *The convolutional layer is used having 32 number of neurons and uses ReLU activation function.*
Step 3. *This convolutional layer is followed by pooling layer having kernel size of 2 units.*
Step 4. *Step 2 and step 3 is repeated again.*
Step 5. *The flatten layer is applied to convert the output from the pooling layer into small dimension vectors.*
Step 6. *The fully connected layer is applied having 128 numbers of neurons for classification purpose.*
Step 7. *The last layer is the output layer which uses softmax activation function for multiple classifications.*

C. Highway Convolutional Neural Network& Algorithm

Highway Convolutional Neural Network are those networks in which the convolutional layers are stacked one after the another where each convolutional layer uses the concept of highway network and the size of the filters of each convolutional layer is smaller than before. This means that the filter size is maximum for the first convolutional layer. There are 16 numbers of filters in each convolutional layer and having ELU activation function. This is followed by the pooling layer and the fully connected layer for the classification task. The first fully connected layer has 128 numbers of neurons and the second fully connected layer have 256 numbers of neurons and both have the ELU activation function. The output layer consists of softmax activation function which is used for multiple classification tasks. Now as each convolutional layer uses the concept of the highway network, thus each convolutional layer employs two gated function. First is the shared gate which is based on the concept of usual convolutional layer and the other is the transformed gate which manages the amount of information that should be propagated from the convolutional layer. This is done with the help of transform matrix $W_{T.}$ The inverse of this defines the amount of information which is coming from input x and is stated as $1 - T(x, W_T)$. Hence the output of the Highway Convolutional layer is

$$y = H(x, W_C).T(x, W_T) + x.\left(1 - T(x, W_T)\right) \tag{1}$$

Where,W_C is the weight matrix combined with the usual convolutional layer and y denotes the output. The research paper which is referred for this paper uses this Highway Convolutional Neural Network on the MNIST dataset.

Step 1. *The input is feed into the input layer*
Step 2. *Two highway convolutional layers are staked one after the another having 16 number of filters and uses ELU activation function.*
Step 3. *The he pooling layer is applied for down sampling.*
Step 4. *Step 2 and step 3 are the loop of 2 units.*
Step 5. *Two fully connected layers are stacked one after the another where first fully connected layer have 128 number of neurons and the second fully connected layers have 256 number of neurons.*
Step 6. *The last is the output layer which uses the softmax activation function for multiple classifications.*

D. Evolutionary Highway Convolutional Neural Network& Algorithm

This network was proposed by the Oliver Kramer [3] where he applied evolutionary approach on the Highway Convolutional Neural Network. The evolutionary approach is very useful when it comes to the problem of black box optimization with local optima. In this, for the evolution of Highway Convolutional Neural Network a new child z' ∈ B^N is generated based on a single parent z where B^N means binary representation with bit string length. The (1+1)-EA which is employed by the network evolution checks if the fitness of the child is equal to the fitness of its parent. This process is repeated until the process gets terminated. EA employs bit flip mutation with probability σ. The evolutionary approach uses two methods that are mutation rate control and niching. The mutation rate control process is done through Rechenberg's rule where if high success rate with$\frac{g}{G} > \frac{1}{5}$, then the mutation rate σ is increased by multiplication $\sigma' = \sigma.\tau$ where $\tau < 1$. The other process is the Niching which is used to overcome local optima where local optima mean the best solution of a problem within its small neighbourhood of possible solutions. The benefit of overcoming this problem is that, if the child is worse than its parent, then it is not rejected with some probability η but is optimized by κ generation. The process within niching is that the fitness of last child replaces the fitness of the last parent only when the fitness of child is better than the parent otherwise original parent will be the base in the main optimization branch. Highway Convolutional Neural Network is evolved in such a way that four highway convolutional layers are stacked one after the another followed by the pooling layer after the 4^{th} convolutional layer in the loop of 4 units where each highway convolutional layers consists of 24 number of filters and uses ReLU activation function. After that 2 fully connected layers are used where the first fully connected layer has 128 neurons and this layer uses softsign activation function and the other fully connected layer has 32 neurons and uses ELU activation function. As the network was implemented on the MNIST dataset, thus the output layer uses softmax activation function as it is used for more than 2 classification tasks. The Highway Convolutional Neural Network generated by the evolutionary approach gave the stable convergence of 0.94 accuracy level when tested on MNIST dataset.

Step 1. *Initialize $z \epsilon B^N$ randomly.*
Step 2. *Repeat*
If niching mode is equal to true for κ generation and $f(x) > f(z_n)$ then replace z with z_n, niching mode is equal to false.
Under the if condition mutate z to z_n with bit flip.
Under if condition adapt σ with Rechenberg and replace z with z' if $f(z)$ is greater than or equal to $f(z')$.
Step 3. *Else with probability η, $z_n = z$ then replace z with z' and count generation with κ, niching mode is equal to true.*
Step4. *This whole process is repeated until the termination condition is met.*

LITERATURE REVIEW

Alex Krizhevsky et al.[1] presented a paper on ImageNet Classification with Deep Convolutional Neural Networks in which they proposed a network called Alexnet implemented on Imagenet dataset where they achieved 67.4% top-1 error rate and 40.9% top-5 error rate and ReLU activation function is used in each convolutional layer followed by overlapping pooling and local response normalization.

Rupesh Kumar Srivastava et al.[2] presented a paper on Highway Networks in which they have given solution of a problem when the training becomes more slow because of very deep networks. Convolutional Neural Networks are combined with Highway networks to form Highway Convolutional Neural Networks and when tested on CIFAR-10 dataset given maximum accuracy of 92.24% when two highway convolutional layers are used. Furthermore 50 hidden layers highway networks are trained on MNIST and CIFAR-100 dataset.

Oliver Kramer[3] presented a paper on Evolution of Convolutional Highway Networks where he proposed a model in which the Highway Convolutional neural network is modified through evolutionary approach. This method was tested on MNIST dataset where the evolutionary approach towards Highway Convolutional Neural Network has shown 0.94 accuracy level at a stable rate.

Djork-Arné Clevert et al.[4] presented a paper on FAST AND ACCURATE DEEP NETWORK LEARNING EXPONENTIAL LINEAR UNITS (ELU) also called as ELU where they applied ELU activation function in the network having more than 5 hidden layers and output preformed ReLU activation function, LReLU and PReLU activation function. ELU network reaches the 20% top-5 error rate after 160K iteration where ReLU takes 200K iteration when tested on imagenet dataset and when tested on CIFAR-10 dataset, it gives the test error of 6.55% and in CIFAR-100 dataset, it gives test error of 24.28%.

Abien Fred M. Agarap[5] presented a paper on Deep Learning using Rectified Linear Units (ReLU) where softmax activation function is compared with ReLU activation function on two architecture that is VGG-like CNN network and FFNN network. For FFNN, 97.77% accuracy generated when applied on MNIST dataset having ReLU activation function and 97.98% accuracy when having softmax activation function, 89.06% accuracy is generated when applied on Fashion-MNIST dataset having ReLU activation function and 89.35% accuracy when having softmax activation function. 90.64% accuracy is generated when applied on WDBC dataset having ReLU activation function and 90.40% accuracy when having softmax activation function. For VGG-like CNN, 91.74% accuracy is generated when tested in MNIST dataset having ReLU activation function and 95.36% accuracy when having softmax activation function. 85.84% accuracy is generated when applied on Fashion-MNIST dataset having ReLU activation function and 86.08% accuracy when having softmax activation function.

Yann LeCun et al.[6] presented a paper on Convolutional Networks for Images, Speech and Time Series where he proposed that as convolutional network is inspired by the mammal’s cortex, it removes the need for the feature extraction but normalizing the image for orientation is still required and fully invarient recognition is still beyond reach.

Xavier Glorot et al.[7] presented a paper on Understanding the difficulty of training deep feedforward neural networks where they compared Deep Neural Networks having 5 hidden layers on the basis of activation function which are Softsign, Softsign N, Tanh, Tanh N and Sigmoid where N means normalized initialization. When tested on Shapeset dataset, Softsign N has the least test error rate of 16.06%, on MNIST dataset , Softsign has the least test error rate of 1.64%, on CIFAR-10, Tanh N has the least test error rate of 52.92% and on Imagenet dataset, Softsign N has the least test error rate of 68.13%.

Nitish Srivastava et al.[8] presented a paper on Dropout: A Simple Way to Prevent Neural Networks from Overfitting where they proposed dropout function on the neural networks to reduce over fitting. Due to this the error rate becomes less when tested on datasets like MNIST, TIMIT, CIFAR-10 and CIFAR-100, Street View House Numbers, ImageNet, Reuters-RCV1 and Alternative Splicing dataset. Using dropout function with DBM when tested on MNIST dataset gives the error rate of 0.79%. Dropout with Convnet and maxpooling when tested on Street View House Numbers gives the error rate of 2.47%. Dropout with Convnet and maxpooling when tested on CIFAR-10 gives error rate of 12.61% and on CIFAR-100 gives error rate of 37.20%. Dropout with Convolutional networks when tested on Imagenet dataset gives error rate of 16%. Dropout with DBM(8 layers and 4 layers) when tested on TIMIT dataset gives error rate of 19.7%. Dropout Neural Network is compared with Bayesian Neural Network where Bayesian Neural Network achieved 623 bits code quality more than dropout neural network. Furthermore comparison is also made on basis of regularization methods on MNIST where Dropout+Max-norm have the least test classification rate of 1.05%.

Gao Huang et al.[12] presented a paper on Densely Connected Convolutional Networks where they have stacked a number of convolutional layers and tested on four datasets that are CIFAR-10, CIFAR-100, SVHN and Imagenet. On Imagenet dataset, DenseNet-161 (k = 48) gets the least top-1 and top-5 error rate of 22.33 and 6.15. On CIFAR-10 and CIFAR-100 datasets, DenseNet-BC gets the error rate of 5.19 and 19.64. On CIFAR-10+ and CIFAR-100+ , error rate achieved where 3.46 and 17.18. On SVHN dataset, DenseNet (k = 24) achieved the least error rate of 1.59.

Lei Zhang et al.[11] presented a paper on ROAD CRACK DETECTION USING DEEP CONVOLUTIONAL NEURAL NETWORK where they compared three models that are ConvNets, Boosting and SVM on the basis of Precision, Recall and F_1 score. ConvNets achieved the best Precision value of 0.8696, Recall value of 0.9251 and F_1 score of 0.8965.

Rupesh Kumar Srivastava et al.[15] presented a paper on Training Very Deep Networks where they having tested dense convolutional networks and highway networks on MNIST, CIFAR-10 and CIFAR-100 datasets. Highway networks having 151K parameters have resulted in 99.55% accuracy on MNIST dataset. Highway networks with 19 numbers of layers gets the highest accuracy of 92.46% on CIFAR-10 dataset. On CIFAR-100 dataset, highway networks has achieved the highest accuracy of 67.76% when compared to other models like Maxout, dasNet, NiN, DSN and All-CNN.

Mariusz Bojarski et al.[14] presented a paper on End to End Learning for Self-Driving Cars where they have proposed their network in which they used first three convolutional layers with a 2x2 strides and 5x5 kernels and last two convolutional layers non-strided 3x3 kernel size. On simulation test, they got the Autonomy value of 100%. 100% speed precision was achieved; position precision was 64.8% and comfort value of 89.6%.

Li Xu et al.[10] presented a paper on Deep Convolutional Neural Network for Image Deconvolution where they had inversed the kernel for deconvolutional into convolutional network. When the proposed work was compared with different models, they get the better performance of 26.23dB on disk sat. Kernel, 26.01dB on disk kernel, 27.76dB on motion sat. Kernel and 27.92dB on motion kernel type.

Javier Ruiz-del-Solar et al.[16] presented a paper on A Survey on Deep Learning Methods for Robot Vision where they compared different models like C3D (1 net) + linear SVM[21], VGG-3D + C3D[22], Ng et al. [18], Wu et al. [19], Guo et al. [17] and Lev et al. [20] on the basis of accuracy where the last models obtains the best accuracy of 94.0%. The models where tested on UCF-101 dataset.

Yann LeCun et al.[9] presented a paper Deep learning where they have given brief description about supervised learning, Backpropagation to train multilayer architecture, Convolutional Neural Network, Image Understanding with deep convolutional neural network, Distributed representations and language processing, Recurrent Neural Networks and the future of deep learning.

Table-1. Table showing detailed parameters of the paper which were taken into consideration in literature review.

S.No	Title of the paper	Author name	Journal or conference name	Methods Used	Results
1.	ImageNet Classification with Deep Convolutional Neural Networks[1]	Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton	NIPS'12 Proceedings of the 25th International Conference on	Alexnet	67.4% top-1 error rate and 40.9% top-5 error rate.

			Neural Information Processing Systems		
2.	Highway Networks[2]	Rupesh Kumar Srivastava, Klaus Greff and Jurgen Schmidhuber	ICML 2015 Deep Learning workshop	Highway networks are combined with convolutional networks	Accuracy of 92.24% was obtained when tested on CIFAR-10 dataset.
3.	Evolution of Convolutional Highway Networks[3]	Oliver Kramer	Neural and Evolutionary Computing	Evolutionary based Highway convnet was compared with highway convnet.	Evolutionary convnet achieved 0.94 accuracy at a stable rate when tested on MNIST dataset.
4.	FAST AND ACCURATE DEEP NETWORK LEARNING EXPONENTIAL LINEAR UNITS (ELU)[4]	Djork-Arné Clevert, Thomas Unterthiner & Sepp Hochreiter	ICLR	ELU activation function	Test error of 6.55% and in CIFAR-100 dataset, it gives test error of 24.28% was achieved using ELU activation function. Also 20% top-5 error rate was achieved at 160K iteration when tested Imagenet dataset.
5.	Deep Learning using Rectified Linear Units (ReLU)[5]	Abien Fred M. Agarap	Neural and Evolutionary Computing (cs.NE); Computer Vision and Pattern Recognition (cs.CV)	softmax activation function is compared with ReLU activation function on two architecture that is VGG-like CNN network and FFNN network	97.77% accuracy when used ReLU and 97.98% accuracy when used softmax for FFNN on MNIST dataset. 91.74% accuracy when used ReLU and 95.36% accuracy when used softmax for VGG like CNN on MNIST dataset. 89.06% accuracy when used ReLU and 89.35% when used softmax for FFNN on Fashion-MNIST dataset. 85.84% accuracy when used ReLU and 86.08% accuracy when used softmax for Fashion MNIST dataset. 90.64% when used ReLU and 90.40% when used softmax for FFNN on WDBC.
6.	Convolutional networks for images, speech, and time series[6]	Yann LeCun and Yoshua Bengio	The handbook of brain theory and neural networks	Convolutional Neural Network	-
7.	Understanding the difficulty of training deep feedforward neural networks[7]	Xavier Glorot, Yoshua Bengio	Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics	Used 5 hidden layers on the basis of activation function which are Softsign, Softsign N, Tanh, Tanh N and Sigmoid	16.06% test error for Softsign N on Shapeset, 1.64% test error for Softsign on MNIST, 52.92% test error for Tanh N on CIFAR-10 and 68.13% test error for Softsign N on ImageNet dataset.

8.	Dropout: A Simple Way to Prevent Neural Networks from Overfitting[8]	Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov		Dropout with convnet where convnet is followed by max-pooling layer.	Using dropout function with DBM when tested on MNIST dataset gives the error rate of 0.79%. With Convnet and maxpooling when tested on Street View House Numbers gives the error rate of 2.47%. With Convnet and maxpooling when tested on CIFAR-10 gives error rate of 12.61% and on CIFAR-100 gives error rate of 37.20%. With Convolutional networks when tested on Imagenet dataset gives error rate of 16%. With DBM(8 layers and 4 layers) when tested on TIMIT dataset gives error rate of 19.7%. Dropout + Max-norm have the least test classification rate of 1.05%.
9.	Densely Connected Convolutional Networks[12]	Gao Huang, Zhuang Liu, Laurens van der Maaten and Kilian Q. Weinberger	CVPR	Number of convolutional layers are stacked and then tested on four dataset.	Densenet-161 achieved the least error top-1 error rate of 22.33 and top-5 error rate of 6.15 on Imagenet dataset. CIFAR-10 and CIFAR-100 datasets, DenseNet-BC gets the error rate of 5.19 and 19.64. On CIFAR-10+ and CIFAR-100+, error rate achieved where 3.46 and 17.18. On SVHN dataset, DenseNet (k = 24) achieved the least error rate of 1.59.
10.	ROAD CRACK DETECTION USING DEEP CONVOLUTIONAL NEURAL NETWORK[11]	Zhang, Lei & Yang, Fan & Zhang, Yimin & Julie Zhu, Ying	ICIP	Three models ConvNets, Boosting and SVM are compared on the basis of Precision, Recall and F_1 score	Precision value of 0.8696, Recall value of 0.9251 and F_1 score of 0.8965 was achieved for road crack detection.
11.	Training Very Deep Networks[15]	Rupesh Kumar Srivastava, Klaus Greff and Jurgen Schmidhuber	Neural Information Processing Systems Conference	dense convolutional networks and highway networks are tested on MNIST, CIFAR-10 and CIFAR-100 datasets	151K parameters has resulted in 99.55% accuracy on MNIST and 19 number of layers gets the highest accuracy of 92.46% on CIFAR-10 for Highway networks. On CIFAR-100 dataset, highway networks has achieved the highest

					accuracy of 67.76%
12.	End to End Learning for Self-Driving Cars[14]	Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D. Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, Xin Zhang, Jake Zhao, Karol Zieba		first three convolutional layers witht a 2x2 strides and 5x5 kernels and last two convolutional layers non-strided 3x3 kernel size	Autonomy and speed precision value achieved was 100%, position precision was 64.8% and comfort value of 89.6% for convolutional layers having a 2x2 strides and 5x5 kernels and last two convolutional layers non-strided 3x3 kernel size.
13.	Deep Convolutional Neural Network for Image Deconvolution[10]	L. Xu, J. S. Ren, C. Liu, and J. Jia	Advances in Neural Information Processing Systems	They had inversed the kernel for deconvolutional into convolutional network.	Performance of 26.23dB on disk sat. Kernel, 26.01dB on disk kernel, 27.76dB on motion sat. kernel and 27.92dB on motion kernel for the proposed work.
14.	A Survey on Deep Learning Methods for Robot Vision[16]	Javier Ruiz-del-Solar, Patricio Loncomilla and Naiomi Soto		Four different models are compared on the basis of accuracy.	Lev et al. [20] proposed model has achieved the highest accuracy of 94.0%.
15.	Deep learning[9]	Y. LeCun, Y. Bengio, and G. Hinton		Brief description about deep learning methods is given.	-

DATASET

Two different data sets are used in this paper. First data set is WANG data set consisting of five hundred images of five different classes and the second data set is Simpsons data set consisting of 1390 images of four different classes.

A. Data Set 1

The first dataset used for the test cases is the WANG[24, 25] dataset. The dataset is sub divided into 5 classes namely class 1, class 2, class 3, class 4 and class 5. Class 1 consisting of African images, class 2 consisting of Beach images, class 3 consisting of image of Monuments, class 4 consisting of buses images and class 5 consisting of Dinosaur images.

Africa Beach Monuments Buses Dinosaurs

Figure 1: One example image from each of the five classes of the used database

This database was utilized broadly to test the different features in light of the fact that the measure of the database and the accessibility of class data takes into account execution assessment. Every one of the images is of size 256 x 256 pixels. For the execution purpose, 285 images where taken as the training samples and 80 images where taken as the testing samples.

B. Data Set 2

The second dataset used for the test cases is the Simpsons[23] dataset. The dataset is sub divided into 4 classes namely class 1, class 2, class 3 and class 4. Class 1 consisting of Abraham Grampa images, class 2 consisting of

Agnes skinner images, class 3 consisting of image of Apu and class 4 consisting of Barney Gumble images.

Abraham Grampa Agnes Skinner Apu Barney Gumble

Figure 2: One example image from each of the five classes of the used database

This database was utilized broadly to test the different features in light of the fact that the measure of the database and the accessibility of class data takes into account execution assessment. Every one of the images is of size 293 x 290 pixels. For the execution purpose, 1390 images where taken as the training samples and 500 images where taken as the testing samples.

RESULTS

This section shows the values of four parameters that are compared on the basis of four different models. For optimization of each model, Adam optimizer is used and trained up to 40 iterations.

Table-2. Table showing comparative analysis of four different types of parameter when tested on WANG dataset

Networks	Train Loss	Test Loss	Train Accuracy	Test Accuracy
Alexnet	0.3122	0.2950	0.2193	0.2625
CNN	0.1604	0.1604	0.1897	0.2125
Highway CNN	0.3305	0.3200	0.1736	0.2000
Evolutionary	0.1599	0.2146	0.1602	0.2625

Table-3. Table showing comparative analysis on four different types of parameters when tested in Simpsons dataset

Networks	Train Loss	Test Loss	Train Accuracy	Test Accuracy
Alexnet	0.2239	0.2140	0.5521	0.5720
CNN	0.1286	0.1269	0.5641	0.5440
Highway CNN	0.2169	0.2320	0.5662	0.5360
Evolutionary Highway CNN	0.1278	0.1242	0.5616	0.5780

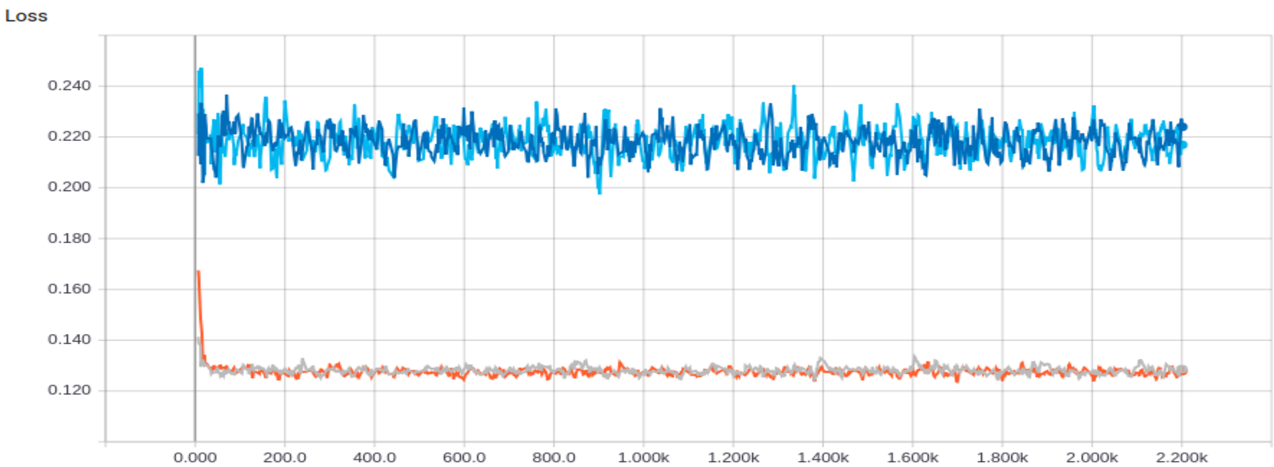

Figure 3: Figure shows the train loss on Simpsons dataset.

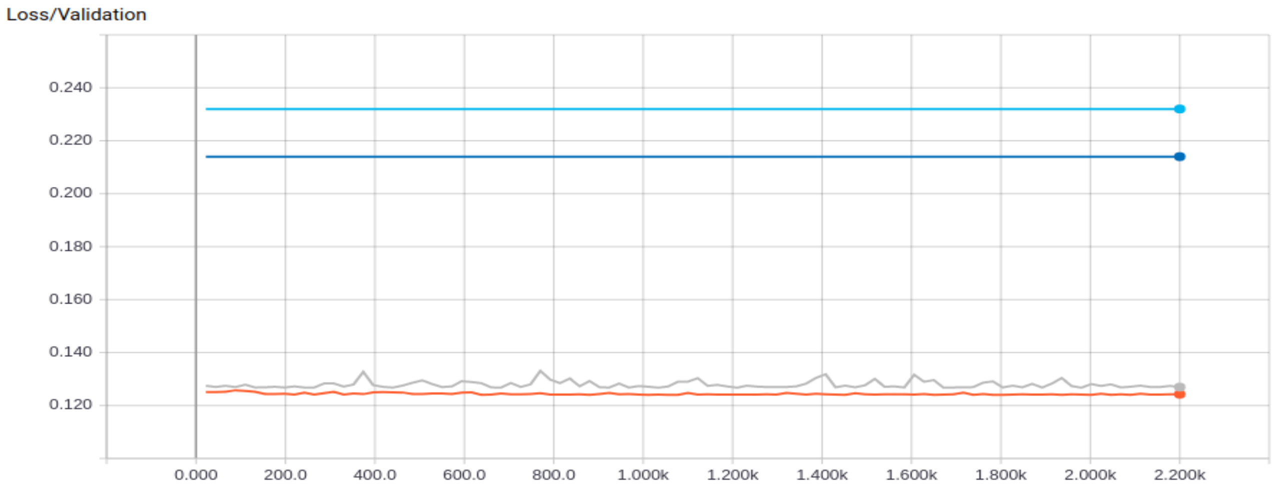

Figure 4: Figure shows the test loss on Simpsons dataset.

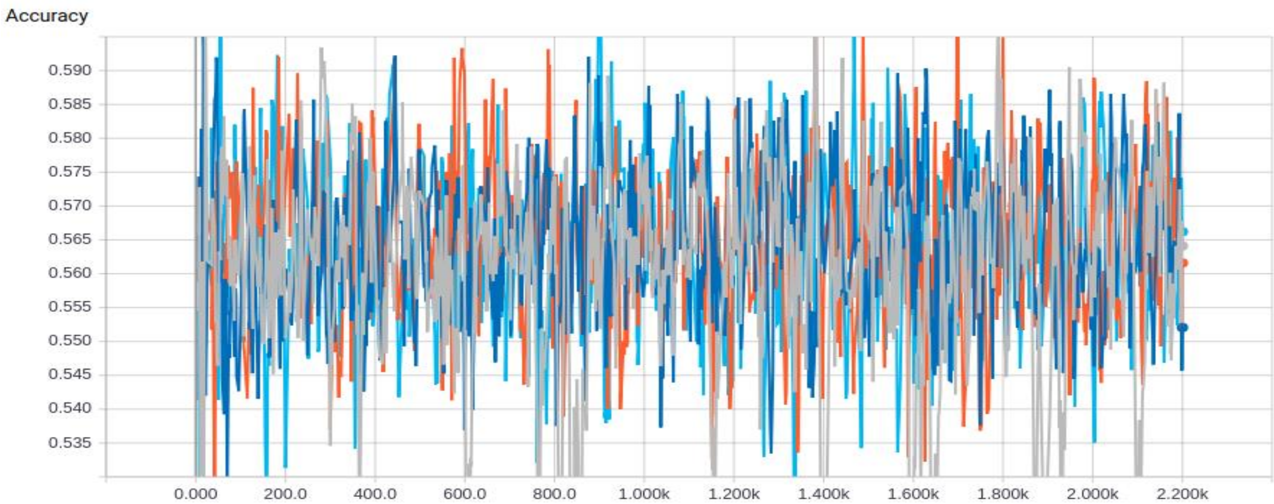

Figure 5: Figure shows the train accuracy on Simpsons dataset.

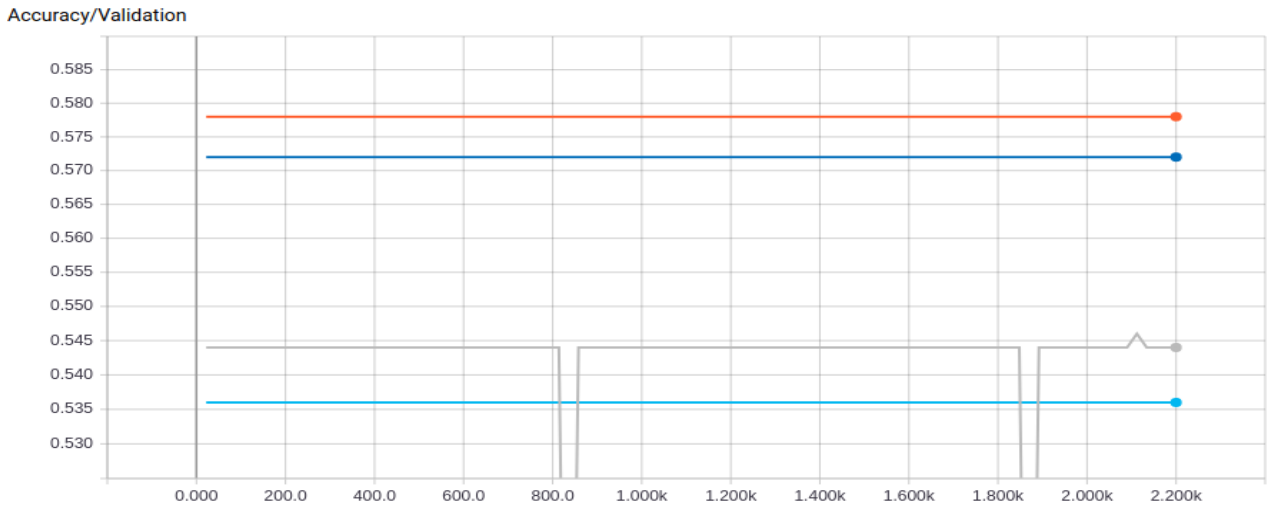

Figure 6: Figure shows the test accuracy on Simpsons dataset.

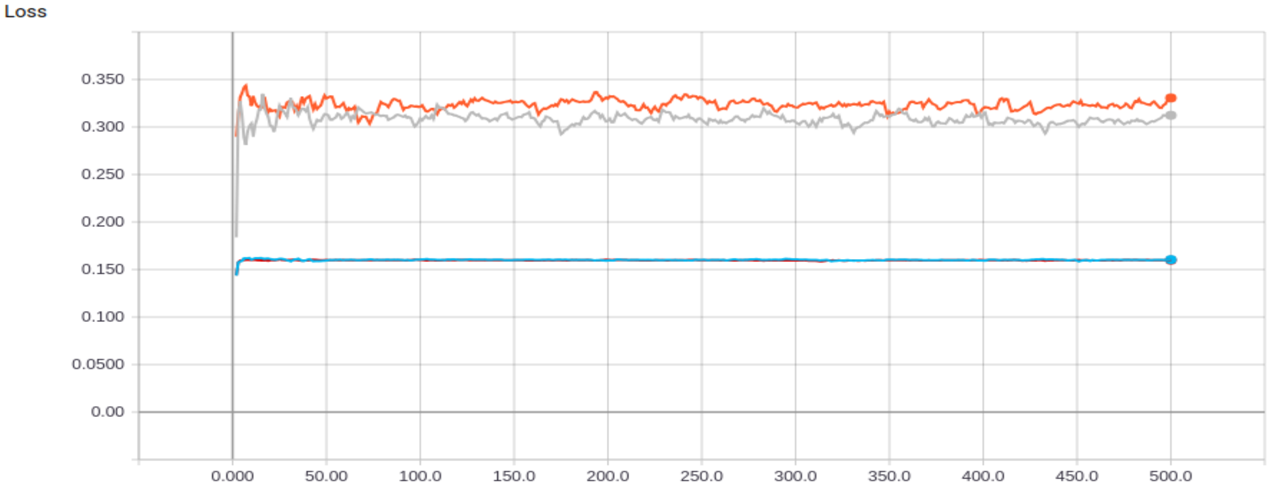

Figure 7: Figure shows the train loss on WANG dataset.

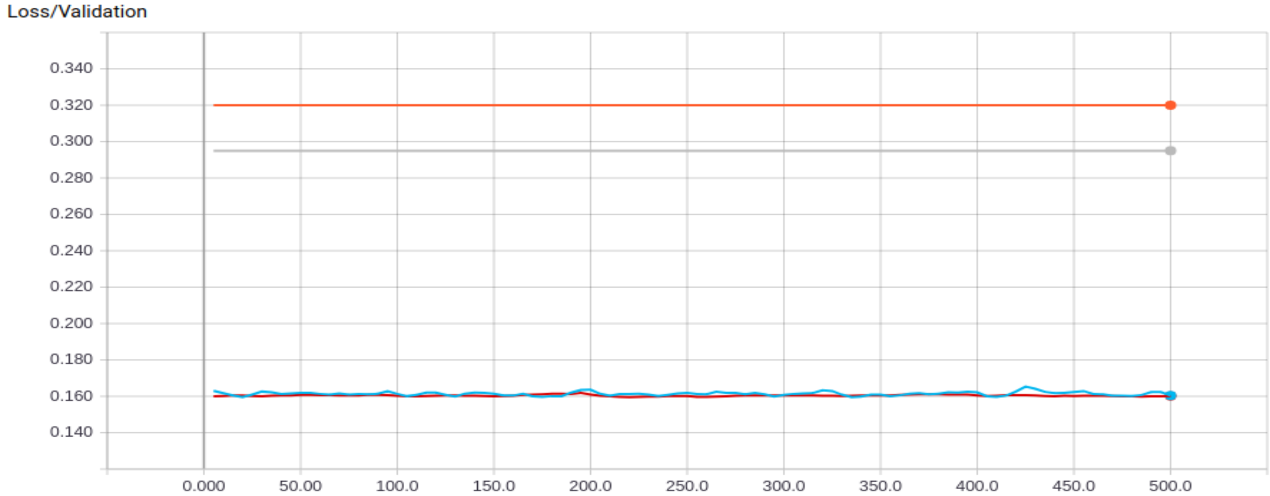

Figure 8: Figure shows the test loss on WANG dataset.

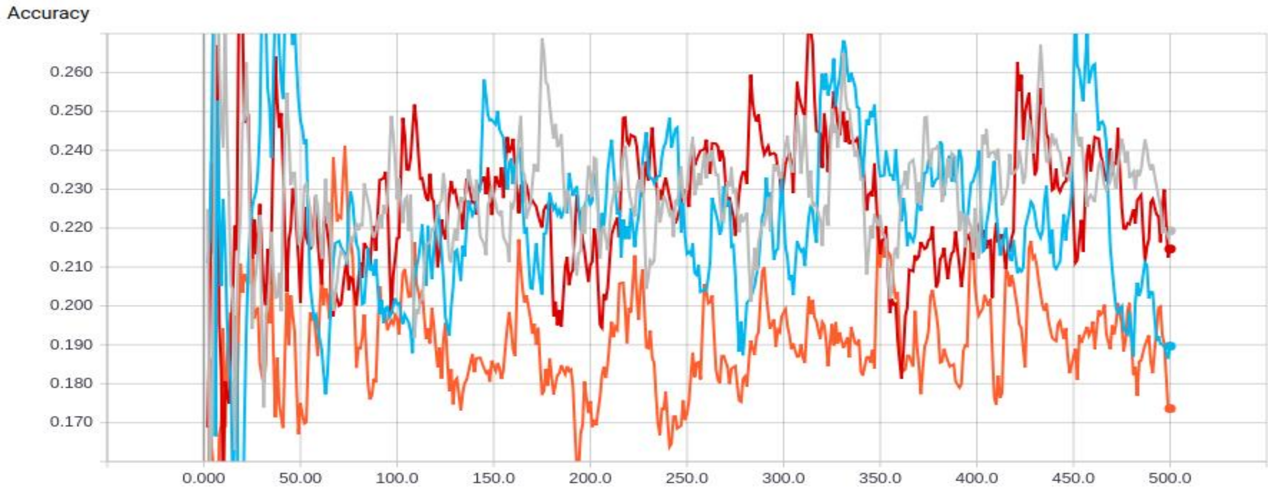

Figure 9: Figure shows the train accuracy on WANG dataset.

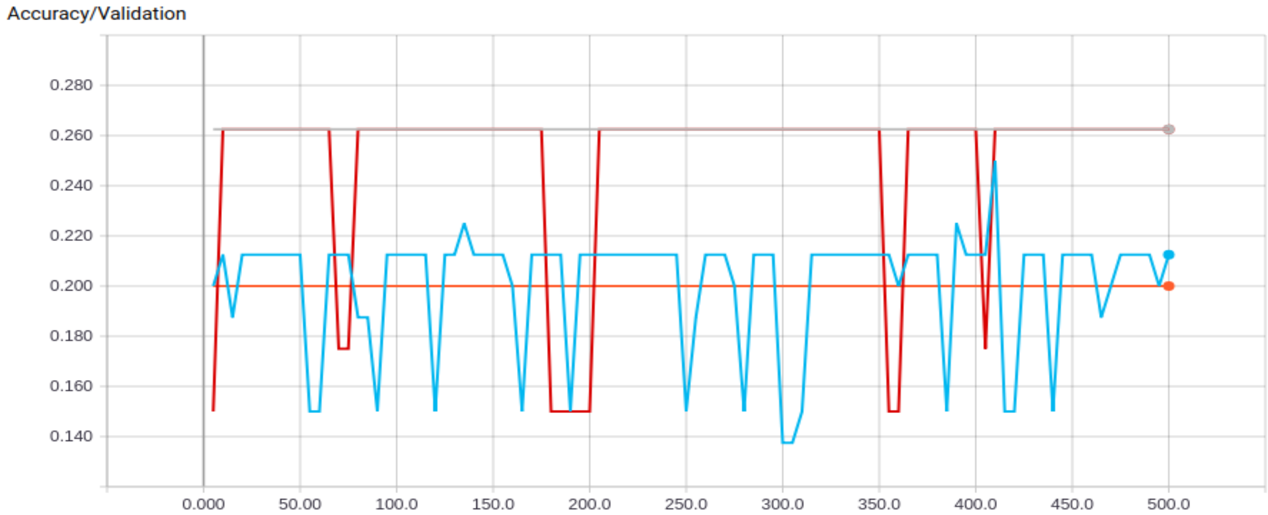

Figure 10: Figure shows the test accuracy onWANG dataset.

The graphs that are acquired while training and testing on Simpsons dataset, grey line signifies original convolutional neural network, blue line signifies Alexnet, orange line signifies evolutionary highway convolutional neural network and light blue line signifies original highway convolutional neural network. For the graphs that are acquired during training and testing of WANG dataset, grey line signifies Alexnet, orange line signifies original Highway convolutional neural network, red line signifies evolutionary highway convolutional neural network and light blue line signifies original convolutional neural network. All the graphs are acquired from the Tensorboard and for the classification process Tensorflow is used, a Deep Learning framework developed by Google.

CONCLUSION

A review of four different types of models is given that are Convolutional Neural Network, Alexnet, Highway Convolutional Neural Network and Evolutionary Highway Convolutional Neural Network. To accomplish this task, two different datasets are used which are WANG dataset and other one is Simpsons dataset. Evolutionary Highway CNN achieved the highest test accuracy of 0.5780 and least test loss of 0.1242. For WANG dataset, highest test accuracy was achieved by Alexnet and Evolutionary Highway CNN and least test loss was achieved by CNN model.

REFERENCES

1. Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012, Vol. 1, Pages 1097-1105.
2. Rupesh Kumar Srivastava, Klaus Greff and Jurgen Schmidhuber, “Highway Networks”, presented at ICML 2015 Deep Learning workshop, 2015, Vol. 2.
3. Oliver Kramer, “Evolution of Convolutional Highway Networks”, Neural and Evolutionary Computing, 2017, Vol. 1.
4. Djork-Arné Clevert, Thomas Unterthiner & Sepp Hochreiter, “FAST AND ACCURATE DEEP NETWORK LEARNING EXPONENTIAL LINEAR UNITS (ELUS)”, ICLR, 2016.
5. Abien Fred M. Agarap, “Deep Learning using Rectified Linear Units (ReLU)”, Neural and Evolutionary Computing (cs.NE); Computer Vision and Pattern Recognition (cs.CV), 2018.
6. Yann LeCun and Yoshua Bengio, “Convolutional networks for images, speech, and time series”, The handbook of brain theory and neural networks, 1998, Pages 255-258.
7. Xavier Glorot, Yoshua Bengio; Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, PMLR 9:249-256, 2010.
8. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov; 15(Jun):1929−1958, 2014
9. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015.
10. L. Xu, J. S. Ren, C. Liu, and J. Jia, “Deep Convolutional Neural Network for Image Deconvolution,” in Advances in Neural Information Processing Systems 27, 2014, pp. 1790–1798.
11. Zhang, Lei & Yang, Fan & Zhang, Yimin & Julie Zhu, Ying. (2016). Road crack detection using deep convolutional neural network. 10.1109/ICIP.2016.7533052.

12. Gao Huang, Zhuang Liu, Laurens van der Maaten and Kilian Q. Weinberger, “Densely Connected Convolutional Networks”, CVPR, 2017.
13. Mnih V., Hinton G.E. (2010) Learning to Detect Roads in High-Resolution Aerial Images. In: Daniilidis K., Maragos P., Paragios N. (eds) Computer Vision – ECCV 2010. ECCV 2010. Lecture Notes in Computer Science, vol 6316. Springer, Berlin, Heidelberg.
14. Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D. Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, Xin Zhang, Jake Zhao, Karol Zieba, “End to End Learning for Self-Driving Cars”, 2016.
15. Rupesh Kumar Srivastava, Klaus Greff and Jurgen Schmidhuber, “Training Very Deep Networks”, Neural Information Processing Systems Conference, 2015.
16. Javier Ruiz-del-Solar, Patricio Loncomilla and Naiomi Soto, “A Survey on Deep Learning Methods for Robot Vision”, 2018.
17. Huiwen Guo, Xinyu Wu, Wei Fengab. Multi-stream deep networks for human action classification with sequential tensor decomposition. Signal Processing, Volume 140, November 2017, Pages 198-206.
18. Y.H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, G. Toderici. Beyond short snippets: deep networks for video classification. IEEE Conference on Computer Vision and Pattern Recognition(CVPR) (2015), pp. 4694-4702.
19. Z. Wu, Y.G. Jiang, X. Wang, H. Ye, X. Xue. Multi-stream multi-class fusion of deep networks for video classification. ACM on Multimedia Conference (2016), pp. 791-800.
20. G. Lev, G. Sadeh, B. Klein, L. Wolf. RNN fisher vectors for action recognition and image annotation. European Conference on Computer Vision(ECCV) (2016), pp. 833-850.
21. Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, Manohar Paluri. Learning Spatiotemporal Features with 3D Convolutional Networks. ArXiv:1412.0767 (2014).
22. Farzad Husain, Babette Dellen and Carme Torra. Action Recognition based on Efficient Deep Feature Learning in the Spatio-Temporal Domain. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
23. https://github.com/alexattia/SimpsonRecognition
24. Jia Li, James Z. Wang, ``Automatic linguistic indexing of pictures by a statistical modeling approach,'' IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 9, pp. 1075-1088, 2003.
25. James Z. Wang, Jia Li, Gio Wiederhold, ``SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries,'' IEEE Trans. on Pattern Analysis and Machine Intelligence, vol 23, no.9, pp. 947-963, 2001.