International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Hyperparameter Optimization and Regularization on Fashion-MNIST Classification

Greeshma K V, Sreekumar K

Abstract—Nowadays the most exciting technology breakthrough has been the rise of the deep learning. In computer vision Convolutional Neural Networks (CNN or ConvNet) are the default deep learning model used for image classification problems. In these deep network models, feature extraction is figure out by itself and these models tend to perform well with huge amount of samples. Herein we explore the impact of various Hyper-Parameter Optimization (HPO) methods and regularization techniques with deep neural networks on Fashion-MNIST (F-MNIST) dataset which is proposed by Zalando Research. We have proposed deep ConvNet architectures with Data Augmentation and explore the impact of this by configuring the hyperparameters and regularization methods. As deep learning requires a lots of data, the insufficiency of image samples can be expand through various data augmentation methods like Cropping, Rotation, Flipping, and Shifting. The experimental results show impressive results on this new benchmarking dataset F-MNIST.

Keywords—Data Augmentation, Convolutional Neural Network (CNN), Hyperparameter Optimization, Deep Learning, Fashion-MNIST

I. INTRODUCTION

Deep learning models have made a great breakthrough for the image classification tasks in computer vision. Deep learning often refers to some hidden elements as hyperparameters as they are one of the most crucial components of any deep learning applications. Hyperparameters are the fine tuning elements that live outside the model but that can heavily influence its behavior and the performance of the model immensely dependent on the selection of right hyperparameter. Within this paper a ConvNet model has built with various regularization methods and hyperparameter optimization techniques for recognizing images of fashion objects using the Fashion-MNIST (F-MNIST) dataset. F-MNIST is a fashion products image dataset be made up of 60,000 training set and 10,000 test set samples including 10 categories of 28x28 grayscale images. Figure 1 shows all the class labels, names and some images in F-MNIST.

Revised Manuscript Received on 30 July 2019.
***** Correspondence Author
Greeshma K V*, Department of Computer Science & I.T., Amrita School of Arts & Sciences, Kochi Amrita Vishwa Vidyapeetham, India.
Sreekumar K, Department of Computer Science & I.T., Amrita School of Arts & Sciences, Kochi Amrita Vishwa Vidyapeetham, India.

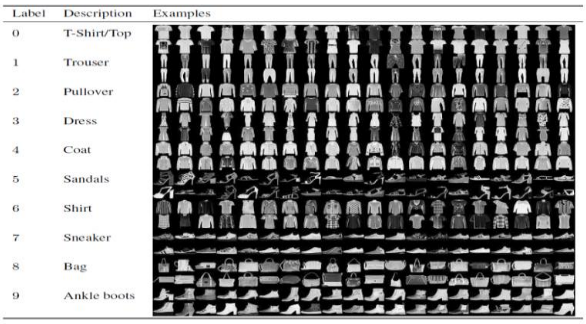


Fig. 1. **Fashion-MNIST Dataset Images**

The regularization technologies like Batch normalization (BN) [8] and Dropout [18] are commonly used to ward off overfitting as regards insufficient training images. Data augmentation is a technique of producing new identical sample images to the training data, which can be considered as one sort of regularization method [14]. For the image classification problems, data augmentations are frequently used in the pioneer research works [17]. For preventing overfitting we have used many effective techniques such as data augmentation and dropout. Most common and the easiest technique to avoid overfitting on the image data is to artificially expand the training image samples. The recently-introduced technique "dropout" [18] is regularization method for neural networks suggested by Srivastava, et al., (2014) where randomly selected neurons ignored or "dropped out" during training.

II. BACKGROUND AND RELATED WORK

In image classification different methods are used such as methods based on low-level image feature representation which consider image as a collection of low-level characteristics like texture, shape, size, color, etc. and methods based on mid-level visual feature constructions for image classification tasks. Nowadays, usage of deep neural networks and neural-networks to obtain image representation is trending. Such architectures allows us to extract features from a specified layer of trained neural network and then use extracted feature maps as a numeric image representation. There are a large number of publications related to the image processing with neural networks. Our work is related to this type of research, where CNN are used for classifying images. Image classification in the fashion domain has numerous benefits and applications and has various research works have been presented about it. One among the previous studies has reviewed deep neural networks is able to attain record breaking outputs on very challenging dataset using supervised learning (Krizhevsky et al., (2012) [11]). Their network contains 5 CNN layers and 3 fully-connected layers.

Retrieval Number: B3092078219/19©BEIESP
DOI: 10.35940/ijrte.B3092.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

They worked one among the largest ConvNets on the ImageNet dataset subsets and achieved best ever results reported on this. This neural network includes a number of novel and unusual features that increase the performance such as relu nonlinearity, overlapping pooling etc. and decrease the time for training. They have used various effective methods for reducing overfitting, which are data augmentation and dropout. Fashion-MNIST dataset has been presented by Zalando Research (Xiao et al., 2017 [19]). F-MNIST is proposed to intend for a direct drop-in substitute for the classical MNIST handwritten digits dataset which has been considered as the benchmark for machine learning techniques, as it contains the same structure, image format and size of train and test set splits. They have provided some results of classification in this paper to form a benchmark on this dataset. All algorithms presented on that were repeated five times by shuffling the training data and the mean of the accuracy on the test data were reported on it. Dufourq et al., 2017 [4] suggest EDEN (Evolutionary DEep Networks), an effective neuro-evolutionary algorithm which includes the strengths of deep networks and genetic algorithms. The search area of the neural network model is explored by them by adding supplementary features like optimization of the embedding layers in their study.

In latest research studies, ConvNets has been used for classifying images. In the work of S. Bhatnagar, D. Ghosal, and Kolekar M. H. (2017) [1], F-MNIST categorization is conducted to categorize groups of fashion article images. They have demonstrated 3 different ConvNet models and applied residual skip connections and batch normalization (BN) for ease and speed of the learning process.

F-MNIST is a kind of more challenging task than classical MNIST dataset. Original MNIST dataset—commonly used as the "Hello World" of machine learning applications in computer vision, is overused, too easy and cannot represent modern computer vision tasks. Researchers at Zalando company have developed a new image classification dataset called F-MNIST in hopes that it should be a substitute for original MNIST [13] dataset. This newly introduced dataset contains images of various products of clothing and accessories—such as t-shirts, coats, shoes, and other fashion items. Each image is a 28x28 grayscale fashion article image, related with a label from ten categories (t-shirt/top to ankle boots). F-MNIST is the most challenging dataset and gives us a lot more room for improving the model. Hence it could be a potential substitute for classical MNIST.

III. PROPOSED METHODOLOGY

Classification of images is used in various applications, ranging from facial recognition to self-driving cars. ConvNets are current state-of-the-art models for object classification. ConvNets are being used everywhere. For getting started with image classification the handwritten digits MNIST dataset is easier and mostly overused.

We propose to classify fashion products images using hyperparameters optimization methods and regularization techniques implementing with CNN. Almost in all computer vision tasks ConvNets are being used. ConvNets mainly consists of three phases. In the first phase a convolution operation occurs in between filters or kernels and input image of very small size and a feature map is produced. Each kernel in a ConvNet learns different features of the image. The convolution operation in ConvNet is simply a mathematical operation i.e. multiplication of the filter and image matrix. The convolution function between a 2D filter Q and 2D image P is,

$$C(m,n) = (P * Q)(m,n) = \sum_i \sum_j P(i,j)Q(m-i,n-j) \quad (1)$$

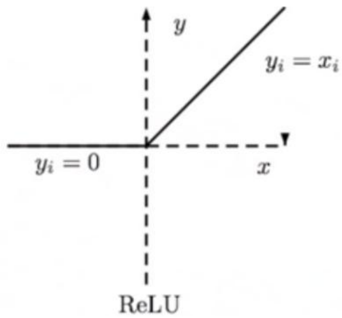
It can also be expressed as,

$$C(m,n) = (Q * P)(m,n) = \sum_i \sum_j P(m-i,n-j)Q(i,j) \quad (2)$$

For a 3x3 filter size the equation becomes,

$$C(m,n) = (Q * P)(m,n) = \sum_{i=1}^{3} \sum_{j=1}^{3} P(m-i,n-j)Q(i,j) \quad (3)$$

The second phase of ConvNet model is Activation layer. Activation function introduces non-linearity to the model. Most prominent activation functions are ReLU (Rectified Linear Unit) [15], Tanh and Sigmoid. ReLU activation function is implemented in the proposed models. Usually, ReLU function is most popularly used in almost all the ConvNets.



$$R(z) = \max(0,z)$$

$$R(z) = \begin{cases} 0 & for \quad z < 0 \\ z & for \quad z \geq 0 \end{cases}$$

ReLU is best for hidden layers. $R(z)$ is zero when z is less than zero and $R(z)$ is equal to z when z is above or equal to zero. Other alternatives are sigmoid, tanh and other activation functions depending on the task. They are a crucial part of neural networks. The third phase is pooling function which is applied to resize the dimension of the input image to avoid overfitting. ConvNets often use pooling layers to decrease the size of the representation. Suppose we have a 4 x 4 input, and you want to apply a type of pooling called max pooling. And the output of this particular implementation of max pooling will be a 2 x 2 output. To demonstrate, if

$$A = \begin{bmatrix} 1 & 3 & 2 & 1 \\ 2 & 9 & 1 & 1 \\ 1 & 3 & 2 & 3 \\ 5 & 6 & 1 & 2 \end{bmatrix}$$

then the result of this 2 x 2 max pooling operation will be

$$MaxP(A) = \begin{bmatrix} 9 & 2 \\ 6 & 3 \end{bmatrix}$$

In this work two different neural network architectures are proposed: 2 ConvNets and 4 ConvNets. The first one includes two convolutional layers. The last one is composed of 4 convolutional layers. These models with hyperparameters like epochs and batch size, optimizers and activation functions and regularization ways such as data augmentation and dropout usage allows us to achieve good results with this model architecture compared with others.

Current work the main focus is to classify fashion images in F-MNIST dataset which is a new challenging alternative to MNIST dataset. We show limitations and shortcomings of complex and simple methods. Therefore, we show necessity of hyperparameters and regularization techniques usage in this problem. Usage of a benchmarking dataset is the well-known and promising approach for classifying images. Herewith HPO and regularization approach are used and showing how we can utilize available knowledge from initial dataset to achieve significantly better performance. We show how to overcome the complexities we faced. In fully-supervised deep neural network architectures with limited training data will dramatically overfit the training data. Final classification model is built on joined methods that allow us to benefit both from using the HPO and regularization methods.

A. Image Preprocessing

The F-MNIST database contains 70000 images of dimension 28x28. These images and their corresponding labels are separated as training data and test data. To prepare the data for training, some processing have applied on the images like resizing images, normalizing the pixel values etc. After doing the necessary processing on the image information's, the label data, we have converted it into categorical formats like label '5' should be represented as a vector format of [0, 0, 0, 0, 0, 1, 0, 0, 0, 0] to build the model.

Fig. 2.Some sample Images in Fashion-MNIST Dataset with their labels

Each image has 28 x 28 resolutions. The CNN accepts image input shape in a specific format. So we have reshaped our input. All the images in our dataset are in grayscale. Normalization is applied on the input images for getting the dimensions in same scale. For that images are rescaled so that each pixel in image data lies in [0, 1] interval format instead of [0, 255]. Then we have applied the one-hot encoding technique for the labels. In this process the label which is an integer here is transformed into a vector which includes only one '1' for corresponding label position and the rest of the elements will be '0'.

B. Convolutional Neural Network (CNN or ConvNets)

Among various deep learning architectures, ConvNets stands out for its unprecedented performance on computer vision. ConvNet is an Artificial Neural Network inspired by biological visual cortex and been successfully applied to image processing tasks. A special kind of artificial neural network is ConvNet which contains at least one convolutional layer. A typical ConvNet takes an input image, pass it through a set of layers convolution, non-linear activation, pooling (downsampling) and fully connected, and retrieve an output of classification labels. This output of this CNN layer is an activation map.

The first ConvNets architecture of the model defined in this paper consists of 2 convolution layers succeeded by activation, pooling, fully connected and softmax layers respectively. Multiple filters are used at each ConvNet layer, for various types of feature extraction. In our first ConvNet layer 32 numbers of filters of the dimension (3, 3) is given and in the second layer 64 filters of (3, 3) is applied. In the second ConvNet model 4 convolution layers followed by Batch Normalization, relu, maxpooling and dropout. First two convolution layers contain 32 numbers of filters and next two with 64 filters. Each filter has the dimension of (3, 3).

C. Regularization Techniques

1) Dropout

In neural networks the regularization technique used to reduce overfitting by preventing co-adaptations on training data is dropout. While training neural network the technique dropout is used which randomly dropping out the neurons in the learning stage. After pooling layer and fully connected layer, dropout is introduced in this architecture to reduce over-fitting problem. In our model after pooling layer dropout ignored 25% of the neurons and disabled 50% of the neurons after fully connected layer.

2) Data Augmentation

When working with deep neural network models data preparation is required. Object recognition tasks are more complex and increasingly it requires data augmentation. The deep networks require very huge amount of sample data to attain the best performance. One technique to get more data for training is image augmentation; it artificially creates new training images by applying transformations on the data. To improve the performance of deep neural networks which is used for building a classifier of images using very little data, data augmentation technique can be applied. The method of artificially creating new images for training by applying transformations such as random rotations, shifts, shear, flips etc. is known as data augmentation.

D. Hyperparameter Tuning

1) Optimizers

Optimization algorithms help us to minimize or maximize the objective function. Minimizing the loss by the training process is very important and has a main role in the operation of training of the neural network model. The two optimizers used in these architectures are Adam [9] and Adadelta [20] for optimization of the loss function. Adam work well across a wide range of deep learning architectures. Adam usually outperforms the rest followed very closely by the other adaptive learning rate methods, Adagrad and Adadelta. Adam optimizer can be calculated as

$$\Delta\theta_t = -\frac{\eta}{\sqrt{\hat{v}_t} + \epsilon}\hat{m}_t \qquad (4)$$

Adadelta is another popular gradient descent technique for optimization of loss function which is also used for the model parameters in our model. Adadelta prevents learning rate decay and it is an extension of Adagrad. The Adadelta rule is represented as

$$\Delta\theta_t = -\frac{\eta}{E[g^2]_t + \epsilon} g_t \qquad (5)$$

$$\Delta\theta_{t+1} = \theta_t + \Delta\theta_t$$

2) Batch size and Number of Epochs

Mini-batch is usually preferable in the learning process of ConvNets. A range of 16 to 128 batch size is a good choice to test with. ConvNet is sensitive to batch size. In this model we have used 64 and 128 as batch size for training images. Number of epochs is the number of complete pass through the entire training set. The number of epochs has increased until the difference of training and the test error is very small. Here, we have checked with 40 and 60 epochs.

3) Activation Function

Activation function is just a thing that should be added to the output at the end of any neural network. This is used to obtain the output of the neural network like yes or no. Depending upon the function it maps the resulting values in between -1 to 1 or 0 to 1 etc. ReLU is really popular in the last few years and it is used in this models.

E. Other parameters

To increase the training speed of the model Batch Normalization was done. To keep the dimensions to be same in the layers after convolution layer the batch normalization applied and it decrease the time for training obviously. The categorical cross entropy is used as loss function because the current problem is the classification of multi-class. This helps for calculating the fault rate value between the actual and predicted.

Fig. 3.Model Loss per Epochs in CNN + HPO +Reg Method

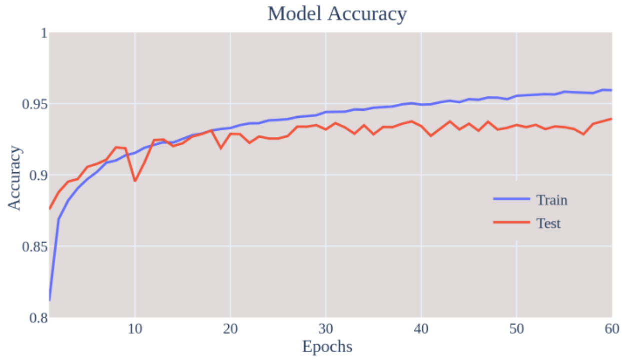


Fig. 4.Accuracy per Epochs in CNN + HPO +Reg Method

Overall comparing with base models, the CNN + HPO + Reg model has high accuracy and low loss in both test and training set as shown in Fig. 3 and Fig. 4. The 32 filters generated in first convolution layer is shown in Fig. 5.

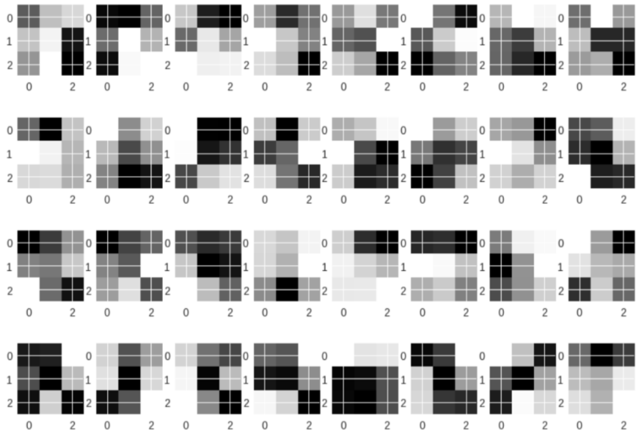


Fig. 5.32 learnable filters with 3x3 kernel size in first layer

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Fashion-MNIST Dataset

In this work, we use F-MNIST dataset, which contains of 60,000 images of training and 10,000 images of test. Each gray scale image has a dimension of 28-by-28 pixels and grouped into ten categories from T-shirt/top to Ankle boots as displayed in Fig. 1. Figure 6 shows that two sample images in F-MNIST dataset, pictures of T-Shirt and Sandal.

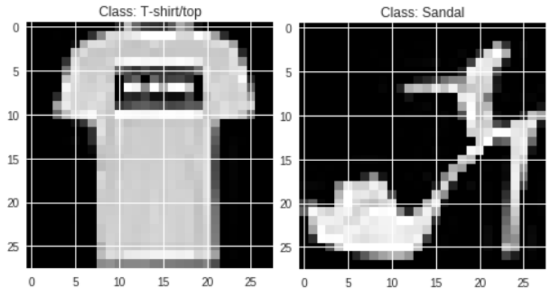


Fig. 6.Two images from two classes of Fashion-MNIST Dataset

We present some state-of-the-arts results to form a benchmark for F-MNIST.

Retrieval Number: B3092078219/19©BEIESP
DOI: 10.35940/ijrte.B3092.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

All neural network architecture results with several ConvNets models configuring hyperparameters and applying regularization are shown in below Table I. By comparing the results of best models published in the literatures [19], [4] and [1]. In literature [19], SVC (Support Vector Classifier) is applied. Table II Shows that these literature results with our best ConvNet performance.

TABLE I. ACCURACY AND LOSS OF F-MNIST WITH OUR MODELS

Model	Parameter	Accuracy		Loss	
		Train	Test	Train	Test
	Adadelta, BS-128, Epochs – 40	0.9945	**0.9352**	0.0238	**0.2469**
CNN2	Epochs – 60	0.9978	**0.9367**	0.0114	0.2876
	Adam, BS-128, Epochs – 40	0.9982	0.9319	0.0092	0.3098
CNN2	Epochs – 60	**0.9993**	0.9302	**0.0045**	0.3372
	Adadelta, BS-64, Epochs – 40	0.9890	0.9324	0.0379	**0.2617**
CNN2	Epochs – 60	0.9931	0.9317	0.0227	0.3109
	Adam, BS-64, Epochs – 40	0.9961	0.9275	0.0157	0.3147
CNN2	Epochs – 60	**0.9990**	0.9334	**0.0047**	0.3670
CNN4 + HPO + Reg	Adam, BS-64, Epochs – 60	0.9594	**0.9399**	0.1100	**0.2037**

CNN2 – 2 Convolutional Layers; CNN4 – 4 Convolutional Layers; HPO – Hyper Parameter Optimization;
Reg – Regularization; BS – Batch Size; DA – Data Augmentation

In the above table we can see that the accuracy and loss of testing and training set. We have achieved 99.90% training accuracy and 0.47% training loss when we have used Adam optimizer and 64 mini-batch size in 60epochs with 2 ConvNet layers. We got the maximum testing accuracy of 93.99% and minimum testing loss 20.37% with hyperparameter optimization and regularization techniques used with 4 ConvNet layers. In this model we have implemented Adam optimizer with batch size 64 and iterated the model till 60 epochs.

TABLE II. ACCURACY RESULTS ON F-MNIST TEST DATA WITH LITERATURES

Model	*Test Accuracy*
SVC [19]	0.8970
EDEN [4]	0.9060
CNN2 [1]	0.9117
CNN2 + BN + Skip [1]	0.9254
CNN4 + HPO + Reg	**0.9399**

SVC – Support Vector Classifier; EDEN – Evolutionary DEep Networks; CNN2 – 2 Convolutional Layers;
BN- Batch Normalization; Skip – Residual Skip Connections; CNN4 – 4 Convolutional Layers

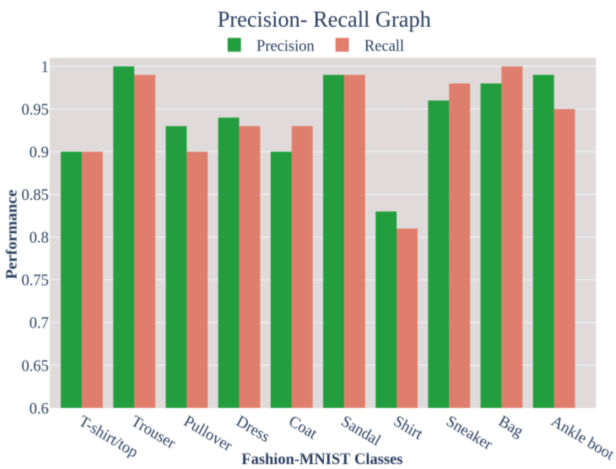


Fig. 7.Precision-Recall Graph of 10 classes

Comparing with the accuracy results on F-MNIST dataset test data results with various models in literature, the CNN4 + HPO + Reg model shows better accuracy results of 93.99% with minimum loss. Figure 8 displays some of the accuracy and loss graphs when using various hyperparameter optimization and regularization techniques. These HPO techniques are varying with their batch size, epochs, optimizers, convolution layers etc.

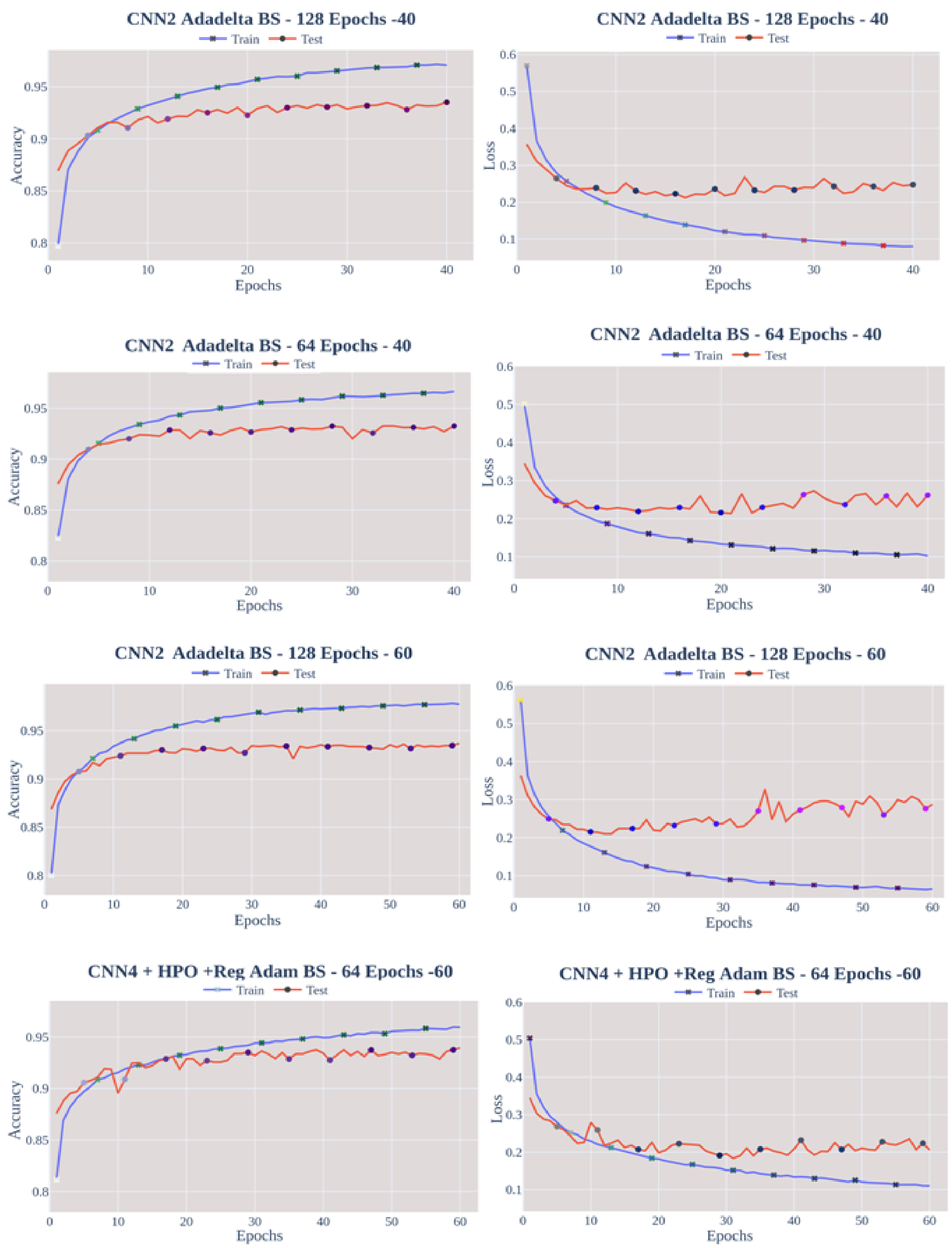


Fig. 8. Accuracy and Loss Graphs using various HPO

TABLE III. CLASSIFICATION REPORT

Class	*Recall*	*Precision*	*F1 Score*
T-shirt/top	0.90	0.90	0.90
Trouser	0.99	1.00	0.99
Pullover	0.90	0.93	0.92
Dress	0.93	0.94	0.94
Coat	0.93	0.90	0.92
Sandal	0.99	0.99	0.99
Shirt	0.81	0.83	0.82
Sneaker	0.98	0.96	0.97
Bag	1.00	0.98	0.99
Ankle boot	0.95	0.99	0.97
Overall	**0.94**	**0.94**	**0.94**

Generally used measures for tasks like image classification are Recall, Precision and F1 Score. These results for 10 different categories are shown in Table III.

When compared with other categories the scores of recall, precision and f1-score are very less for Shirt and T-Shirt/top. Shirts, Pullover, T-Shirt/top and Coats also show low scores. Among other categories of images, these 4 categories are frequently misclassified. The reason for this low metrics score is its similarity in such small images of 28-by-28 pixels. In the below Table IV we can clearly see the different categories which are the major causes of error in predicting the correct image. The main sources of error in the proposed model are from Shirt and T-Shirt/Top class.

TABLE IV. CONFUSION MATRIX : CNN4 + HPO + REG MODEL

Class	***Label***	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
T-shirt/top	**0**	**898**	0	13	7	2	1	72	0	7	0
Trouser	**1**	0	**988**	0	7	1	0	2	0	2	0
Pullover	**2**	14	0	**902**	6	36	0	41	0	1	0
Dress	**3**	15	1	10	**932**	12	0	29	0	1	0
Coat	**4**	0	0	17	20	**932**	0	27	0	4	0
Sandal	**5**	0	0	0	0	0	**995**	0	3	0	2
Shirt	**6**	75	0	29	19	54	0	**815**	0	8	0
Sneaker	**7**	0	0	0	0	0	6	0	**984**	1	9
Bag	**8**	0	0	0	1	0	0	0	0	**999**	0
Ankle boot	**9**	0	0	0	0	0	8	0	37	1	**954**

V. CONCLUSION

With hyperparameter optimization and regularization techniques used with four layer ConvNets we were capable of attain an accuracy of 93.99%. We can clearly see how by tuning various hyperparameters like optimizers, batch size, number of epochs and regularization methods such as image augmentation and dropout increase the overall performance and significantly decrease the training time. F-MNIST can be a best drop-in substitution for MNIST although it is more difficult than MNIST dataset. We can implement or serve these models with hyperparameter tuning and regularization techniques for various types of image classification tasks and this dataset should be very much challenging when doing machine learning tasks. Configuring hyperparameters needs years of experience and it is a black art, but tuning these parameters we can achieve magical results solving various computer vision tasks.

REFERENCES

1. Bhatnagar, S., Ghosal, D., & Kolekar, M. H. (2017, December). Classification of fashion article images using convolutional neural networks. In Image Information Processing (ICIIP), 2017 Fourth International Conference on (pp. 1-6). IEEE.
2. Cireşan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. arXiv preprint arXiv:1202.2745.
3. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on (pp. 248-255). Ieee.
4. Dufourq, E., & Bassett, B. A. (2017, November). EDEN: Evolutionary deep networks for efficient machine learning. In Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech), 2017 (pp. 110-115). IEEE.
5. Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). Cambridge: MIT press.
6. Greeshma K. V., & Sreekumar K. Fashion-MNIST Classification Based on HOG Feature Descriptor Using SVM, unpublished.
7. Iliukovich-Strakovskaia, A., Dral, A., & Dral, E. (2016). Using pre-trained models for fine-grained image classification in fashion field. In Proceedings of the First International Workshop on Fashion and KDD, KDD (pp. 31-40).
8. Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.
9. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
10. Kori, A., Krishnamurthi, G., & Srinivasan, B. (2018). Enhanced Image Classification With Data Augmentation Using Position Coordinates. arXiv preprint arXiv:1802.02183.
11. 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).
12. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
13. LeCun, Y. (1998). The MNIST database of handwritten digits. http://yann. lecun. com/exdb/mnist/.
14. Lemley, J., Bazrafkan, S., & Corcoran, P. (2017). Smart Augmentation Learning an Optimal Data Augmentation Strategy. IEEE Access, 5, 5858-5869.
15. Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10) (pp. 807-814).
16. Seo, Y., & Shin, K. S. (2019). Hierarchical convolutional neural networks for fashion image classification. Expert Systems with Applications, 116, 328-339.
17. Shijie, J., Ping, W., Peiyi, J., & Siping, H. (2017, October). Research on data augmentation for image classification based on convolution neural networks. In Chinese Automation Congress (CAC), 2017 (pp. 4165-4170). IEEE.
18. 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.
19. Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747.
20. Zeiler, M. D. (2012). ADADELTA: an adaptive learning rate method. arXiv preprint arXiv:1212.5701.
21. Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham.