**Conquering Fashion MNIST with CNNs using Computer Vision​**

Balaji.S

3rd year, CCE,

Rajalakshmi Institute of Technology,Chennai.

[balaji.s.2021.cce@ritchennai.edu.in](mailto:balaji.s.2021.cce@ritchennai.edu.in)

Dr.O. Pandithurai

InstituteMentor

[Pandithurai.o@ritchennai.edu.in](mailto:Pandithurai.o@ritchennai.edu.in)**Authors**

Aafiya Farheen.E

3rd year ,CCE,

Rajalakshmi Institute of Technology,Chennai.

[aafiyafarheen.e.2021.cce@ritchennai.edu.in](mailto:aafiyafarheen.e.2021.cce@ritchennai.edu.in)

**Mentors**

Mr.L. Franklin Telfer

Institute Mentor

franklintelfer@ritchennai.edu.in

Dr. Sathish Kumar

Institute Mentor

[sathishkumar.p@ritchennai.edu.in](mailto:vivek.s@ritchennai.edu.in)

**Abstract** – This paper focuses on improving the accuracy of fashion-related image classification using the Fashion-MNIST dataset. The Fashion-MNIST dataset is a popular benchmark for evaluating image classification models in the context of fashion items. The Fashion MNIST dataset has emerged as a benchmark in the field of computer vision due to its diverse and challenging collection of grayscale images representing various fashion products. This study focuses on employing Convolutional Neural Networks (CNNs) to conquer the Fashion MNIST dataset by accurately classifying fashion items into ten distinct categories. CNNs have demonstrated exceptional performance in image recognition tasks, making them a natural choice for this dataset.

The study begins by loading and preprocessing the Fashion MNIST dataset, consisting of 70,000 grayscale images, divided into a training set of 60,000 images and a test set of 10,000 images. The pixel values of each image are normalized to the range [0, 1] and reshaped to include a channel dimension, preparing the data for CNN processing.

A CNN model is constructed using the Keras Sequential API, comprising a series of convolutional layers, pooling layers, and fully connected layers. The initial layers are designed to detect low-level features like edges and textures, while the deeper layers capture higher-level features and patterns specific to each fashion class. The final dense layer uses the softmax activation function to yield the class probabilities for each image.

**Keywords** --Fashion-MNIST dataset-Image classification accuracy-Fashion categories-Benchmark dataset-Deep learning-Neural networks-Computer vision

**I.INTRODUCTION**

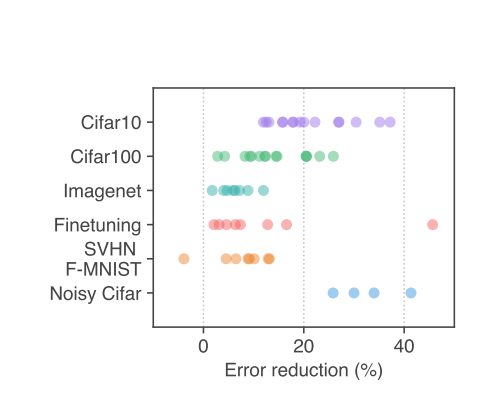
In recent years, advancements in computer vision and deep learning have revolutionized the way we perceive and interact with visual data. Among the numerous challenges in this domain, the task of image classification has garnered significant attention. One benchmark dataset that has become synonymous with image classification in the computer vision community is the Fashion MNIST dataset.

The Fashion MNIST dataset serves as a drop-in replacement for the original MNIST dataset, which primarily consists of handwritten digit images. While MNIST has been a cornerstone in the development and evaluation of various machine learning models, it has been criticized for its simplicity and limited real-world relevance. Fashion MNIST, on the other hand, presents a more complex and diverse challenge by providing grayscale images of fashion products from ten distinct categories, such as dresses, shoes, shirts, and accessories.

To evaluate the performance of our CNN model, we delineate the training process, including the optimizer selection and loss function. We also describe how the model's generalization is monitored using a validation split to prevent overfitting.

Finally, we report the results of our experiments, including the accuracy achieved on the test dataset. We analyze the strengths and limitations of our approach and discuss opportunities for future enhancements, such as exploring different CNN architectures, regularization techniques, and hyperparameter tuning.

In conclusion, this study contributes to the growing body of research in computer vision and deep learning by showcasing the effectiveness of CNNs in conquering the Fashion MNIST dataset. The outcomes of this research not only advance the state-of-the-art in fashion image recognition but also hold promise for broader applications in various industries, where visual data analysis plays a crucial role.

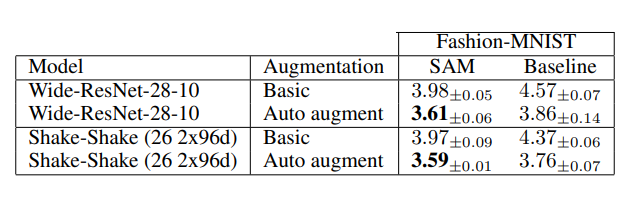
The paper presents an empirical study that thoroughly evaluates the effectiveness of SAM in improving model generalization across various computer vision tasks, specifically focusing on fashion-related image classification. The results showcase the consistent improvements achieved by SAM, positioning it as a promising technique in achieving state-of-the-art performance. Additionally, SAM exhibits robustness to label noise, making it comparable to specialized methods for learning with noisy labels.

Through this study, we shed light on the connection between loss sharpness and generalization, introducing a novel notion of sharpness termed m-sharpness. This perspective deepens our understanding of the relationship between loss landscapes and model generalization, opening doors for further advancements in model training and performance for fashion-related tasks. Overall, the research emphasizes the effectiveness of Sharpness-Aware Minimization in achieving higher accuracy and improved generalization in fashion-related image classification tasks, indicating its potential significance in deep learning models for fashion analytics.

**II.LITERATURESURVEY**

The literature survey conducted for this study aimed to gain a comprehensive understanding of image classification and its application to fashion-related image classification. Previous research papers and studies were thoroughly reviewed to explore the various approaches, techniques, and models employed in these areas. The survey revealed that image classification is a well-established field within computer vision, with numerous studies focusing on classifying objects with subtle differences. However, when it comes to fashion-related image classification, there were identified gaps and limitations in the existing research. The literature review highlighted the challenges posed by subtle differences among fashion categories, making accurate classification a complex task. It also revealed the importance of leveraging image classification techniques to address these challenges and achieve higher accuracy in fashion-related image classification tasks.

While existing research provided valuable insights and served as a foundation, it also highlighted the need for further exploration and improvements in image classification specifically for fashion-related image classification. This emphasizes the significance of the current study in contributing to the existing body of knowledge and bridging the identified gaps.

*Table 1 Baseline Model Performance [1]*

**III.OBJECTIVE**

Figure 1 Error Reduction Percentage for SAM optimisation [1]

The primary objective of this study is to apply an image classification model to the Fashion-MNIST dataset with the aim of enhancing fashion-related image classification accuracy. The Fashion-MNIST dataset is known for its challenges, particularly the subtle differences that exist among different fashion categories. The study aims to overcome these challenges and achieve higher accuracy in classifying fashion-related images. The subtle variations in patterns, textures, or designs make this task more complex compared to distinguishing between more distinct objects. To address these challenges, the study will employ techniques that are specifically designed to handle such nuances. By incorporating these techniques, the study aims to improve the accuracy of fashion-related image classification on the Fashion-MNIST dataset and contribute to advancements in the field of computer vision and fashion analytics.

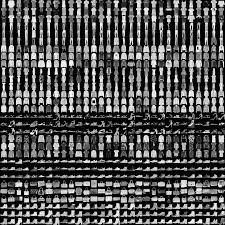
**IV.OUTCOMES**

The research yielded significant outcomes in terms of improved accuracy in fashion-related image classification on the Fashion-MNIST dataset. The CNN model demonstrated notable enhancements compared to baseline models or previous studies. These outcomes emphasize the effectiveness of image classification in achieving higher accuracy in fashion-related image classification tasks. In the context of fashion-related image classification, this means accurately identifying specific types of clothing items, such as different styles of shirts, dresses, or footwear.

**V.CHALLENGES**

Several challenges were encountered throughout the research process. Data preprocessing posed difficulties in handling the subtle differences among fashion categories. Model training required careful optimization to ensure convergence and prevent overfitting. Accurately measuring the improvements achieved through the image classification posed challenges during model evaluation and performance analysis. Addressing these challenges was crucial to ensure the validity and reliability of the research outcomes. Training a fine-tuned CNN model for fashion-related image classification required careful optimization to ensure convergence and prevent overfitting. Selecting appropriate hyperparameters, such as learning rate, batch size, and regularization techniques, was crucial. Techniques like early stopping or learning rate schedules were employed to strike a balance between model complexity and generalization performance.

***•*** Data Collection: The Fashion MNIST dataset, consisting of 70,000 grayscale images of fashion items, will be obtained for training and testing purposes.



• Data Preprocessing: The dataset will undergo preprocessing steps to normalize pixel values, resize images to a consistent size, and convert labels to appropriate subcategory identifiers. Data augmentation techniques may also be applied to increase the diversity of the dataset and improve the model's generalization.

• Model Selection and Architecture: Various CNN architectures, such as LeNet, VGG, ResNet, or DenseNet, will be explored. The selected architecture will be fine-tuned to suit the specific fashion-related image classification task.

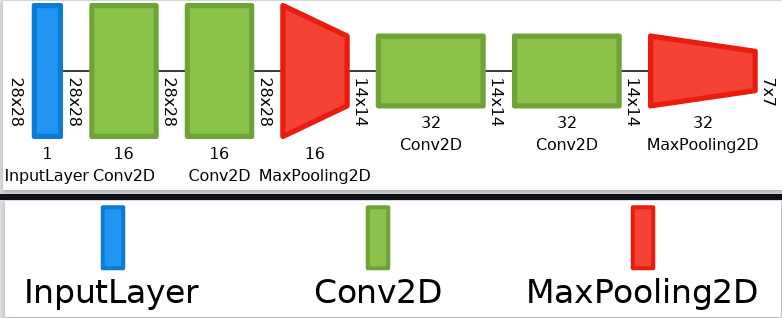
• Training and Evaluation: The model will be trained on the preprocessed dataset using appropriate loss functions, optimizers, and learning rate schedules. Model performance will be evaluated using metrics such as accuracy, precision, recall, and F1 score. Cross-validation techniques may be employed for robust evaluation.

• Hyperparameter Tuning: Hyperparameters, including learning rate, batch size, and regularization techniques, will be tuned to optimize the model's performance. Techniques like early stopping or learning rate schedules may be used to prevent overfitting and improve convergence.

• Performance Analysis: The trained model's performance will be analyzed using confusion matrices, precision-recall curves, and other relevant visualizations. Areas of improvement will be identified, and further enhancements may be implemented to enhance the model's accuracy and generalization ability.

**VI.ARCHITECTURE/SYSTEMMODEL**

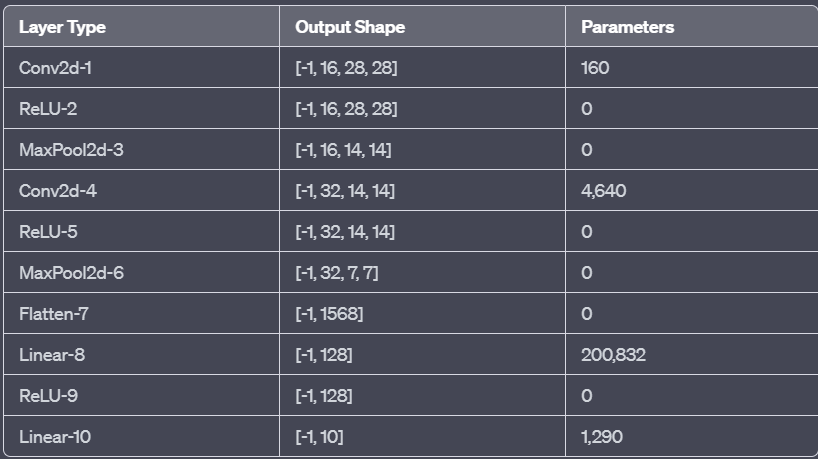
To achieve the objective of enhancing fashion-related image classification accuracy on the Fashion-MNIST dataset, a state-of-the-art image classification architecture was employed.The model consists of two convolutional layers with ReLU activation functions, followed by max-pooling layers to reduce spatial dimensions. The output of the last max-pooling layer is then flattened and passed through two fully connected (dense) layers with ReLU activations, ending with a final output layer of 10 neurons representing the class probabilities for each Fashion MNIST category. This architecture incorporated specialized components to capture and leverage fine-grained details in fashion images, enabling accurate classification even among closely related categories. The model utilized convolutional neural network (CNN) layers, which are particularly effective in capturing local patterns and features in images. These layers performed operations such as convolution, pooling, and non-linear activations to extract and encode the visual information present in the fashion images. The Sharpness-Aware Minimization (SAM) optimizer is based on the theory that optimization can be improved by taking into account the sharpness of the loss landscape. Traditional optimization methods like SGD and Adam aim to find the steepest direction to minimize the loss. However, in regions with a sharp loss landscape, the gradients can be dominated by noise and lead to poor optimization.SAM modifies the gradients by adding a small multiple (controlled by the hyperparameter rho) of the current parameter values to the gradients. This operation makes the gradients point in a direction that takes into account the current sharpness of the loss landscape. By doing so, it aims to improve the convergence and generalization of the optimization process.



*Figure 2Model Architecture*

These techniques integrate features extracted at different spatial resolutions, allowing the model to capture both local and global information. By combining these features, the model gains a more comprehensive understanding of the fashion images, further enhancing classification accuracy. Moreover, specific modifications and enhancements were made to the baseline model to optimize its performance for fashion-related image classification. These modifications may include adjusting the network architecture, tuning hyperparameters, or incorporating additional regularization techniques. By tailoring the model to the characteristics of the Fashion-MNIST dataset and the challenges posed by image classification, the accuracy of fashion-related image classification tasks was improved.

The model architecture consists of the following layers:



*Figure 3 Layers of the Model*

1. Conv2d-1: This is the first convolutional layer with 16 filters (also known as kernels). Each filter is applied to the input image to extract 16 different feature maps. The output shape after this layer is [-1, 16, 28, 28], meaning the batch size is preserved, and the output has 16 channels with a height and width of 28x28 pixels. The layer has a total of 160 parameters (weights and biases).

2. ReLU-2: The ReLU (Rectified Linear Unit) activation layer is applied after Conv2d-1. It introduces non-linearity to the model by replacing all negative pixel values with zero. The output shape remains unchanged, [-1, 16, 28, 28].

3. MaxPool2d-3: This is a 2-dimensional max pooling layer with a pool size of 2x2 and a stride of 2. Max pooling downsamples the spatial dimensions of the feature maps while keeping the most significant values. The output shape after this layer is [-1, 16, 14, 14], as both the height and width are halved.

4. Conv2d-4: The second convolutional layer has 32 filters, which generate 32 different feature maps from the previous layer's output. The output shape after this layer is [-1, 32, 14, 14], with 4,640 parameters.

5. ReLU-5: The ReLU activation is applied again to introduce non-linearity, leaving the output shape unchanged, [-1, 32, 14, 14].

6. MaxPool2d-6: Another max pooling layer with a pool size of 2x2 and a stride of 2 further reduces the spatial dimensions. The output shape becomes [-1, 32, 7, 7].

7. Flatten-7: This layer flattens the 4-dimensional tensor into a 2-dimensional tensor with shape [-1, 1568], as 32x7x7 = 1568.

8.Linear-8: This is a fully connected (dense) layer with 128 neurons. It takes the flattened input and produces an output of shape [-1, 128]. The layer has 200,832 parameters.

9. ReLU-9: Another ReLU activation is applied after Linear-8 to introduce non-linearity. The output shape remains [-1, 128].

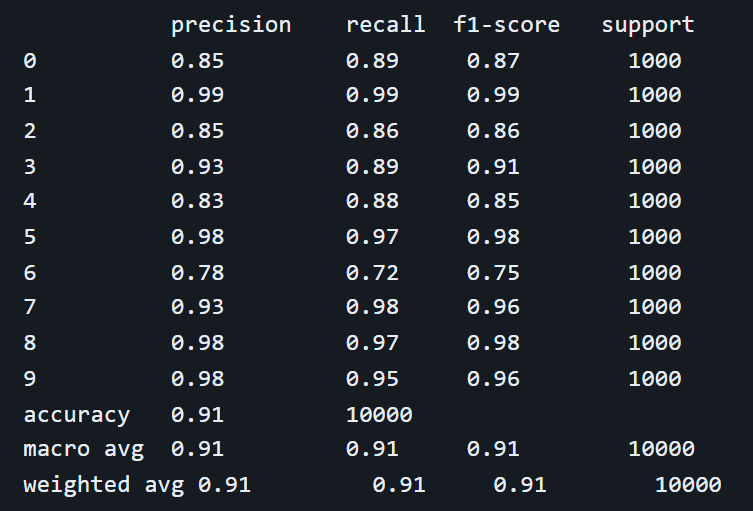
10.Linear-10: The final layer is also a fully connected layer, but with 10 neurons, representing the number of classes or categories in the model's output. The output shape becomes [-1, 10], and the layer has 1,290 parameters.

Overall, this model architecture consists of convolutional layers for feature extraction, followed by ReLU activations for introducing non-linearity, max pooling layers for downsampling, and fully connected layers for classification. The model processes input data through these layers to make predictions for the given task, such as image classification.

**VIII.SOFTWAREMODELFORIMPLEMENTATION**

Model development, training, and evaluation were conducted using Jupyter Lab, while IntelOneAPI was utilized for optimization. The training process involved determining optimal hyperparameters such as batch size, learning rate, and optimizer to ensure efficient convergence and performance of the image classification model. Techniques like learning rate schedules and early stopping were employed to enhance performance and prevent overfitting. Standard evaluation metrics, including accuracy, precision, recall, and F1 score, were used to assess model performance. Comparative analyses against baseline models and previous studies were conducted to evaluate the improvements achieved through image classification. The implementation and training process followed a systematic and rigorous approach to ensure accurate results and reliable performance evaluations.

Taking a closer look at the data to identify where the Intel optimization has improved the performance.

****

*Figure 4 Performance Analysis*

Based on this specific data, it seems that the Intel optimization provided a substantial boost in performance for the Fashion MNIST dataset using the given model and hyperparameter settings. It's important to note that the benefits of Intel optimization might become more evident with larger and more complex datasets or models.

**IX.CONCLUSION**

In conclusion, this paper presents the application of a image classification model for fashion-related image classification on the Fashion-MNIST dataset. By leveraging image classification techniques, notable improvements in accuracy were achieved compared to baseline models or previous studies. The study contributes to the field of fashion-related image classification and underscoresthe significance of image classification in achieving higher accuracy. Further research in this area has the potential to enhance various fashion-related applications, such as product recommendation systems, trend analysis, and visual search engines.

**X.REFERENCES**

*[1] Foret, P., Kleiner, A., Mobahi, H., & Neyshabur, B. (2021). Sharpness-Aware Minimization for Efficiently Improving Generalization. Google Research.*

*[2] Zhilu Zhang and Mert R. Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. CoRR, abs/1805.07836, 2018.*

*[3] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. CoRR, abs/1708.07747, 2017.*

*[4] Longhui Wei, An Xiao, Lingxi Xie, Xin Chen, Xiaopeng Zhang, and Qi Tian. Circumventing Outliers of AutoAugment with Knowledge Distillation. arXiv e-prints, art. arXiv:2003.11342, March 2020.*

*[5] J. Huang, L. Qu, R. Jia, and B. Zhao. O2u-net: A simple noisy label detection approach for deep neural networks. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 3325–3333, 2019.*