IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

Improved Fabric Defect Detection Using a Vision Transformer and ResNet Hybrid Model

Vishwath Ramachandran[1], Rohit Madhavan S[1], Philip Anand[1], Arjun Vishanth[1], Pradeep K V[2]
Centre for Advanced Data Science, Vellore Institute of Technology, Chennai, India[1]
School of Computer Science Engineering, Vellore Institute of Technology, Chennai, India[2]

Abstract: *Fabric defect detection plays a vital role in ensuring the production of high-quality textiles. Manual inspection methods are time-consuming, subjective, and error-prone, necessitating the development of automated systems. This research paper proposes a novel approach to fabric defect detection by combining the strengths of Vision Transformer (ViT) and ResNet-50 architectures into a hybrid model. A notch filter coupled with a fast Fourier transform is also used to improve the quality of the training dataset. The hybrid model aims to capture both local and global information within fabric images, thereby enhancing defect detection accuracy. Extensive experiments on a publicly available fabric defect dataset demonstrate that the proposed hybrid model outperforms the individual ViT and ResNet-50 models and other state-of-the-art fabric defect detection approaches. The results showcase a superior accuracy of 98.5% for our proposed architecture, which is significantly higher than the 93.4% and 96.5% achieved by ResNet and ViT, respectively.*

Keywords: Fabric defect, Machine Learning, ResNet-50, Vision Transformer, Hybrid Model.

I. INTRODUCTION

The textile industry is dedicated to producing high-quality fabrics that meet customer expectations and maintain a competitive edge in the market. Fabric defects, including holes, stains, and irregularities, can significantly affect textiles' aesthetic appeal and functional performance, resulting in customer dissatisfaction and financial losses for manufacturers. Cloths and consumer goods with fabric tears or any other defects can significantly affect brand reputation. This can in turn result in monetary losses. Multiple apparel companies have gone out of business in the past due to the reduction in the quality of the fabrics they make.

Currently, the detection of defects in fabrics is done by manual inspection. However, this method suffers from multiple drawbacks. The manual inspection of fabrics is time-consuming and expensive. The process is also prone to errors due to subjectivity or carelessness. As more resources are invested into the manual detection of defects in fabrics, the price of the end-product increases resulting in consumers paying higher for clothes and other fabric-related goods. Foreign materials caught in the fabric greatly reduce the quality of the material and could harm the end user if undiscovered. Sometimes certain defects such as loose threads are too small to be identified by the naked eye. Initially, the aesthetic of the fabric will not be affected, but it could grow and reduce the structural integrity of the fabric over time. To overcome these limitations, computer vision techniques and machine learning algorithms have emerged as promising solutions for automating fabric defect detection. Artificial intelligence is often optimized for identifying patterns. Since the images of fabric defects can be classified as irregularities in the pattern of fabrics, AI is well suited for the task of identifying them. These machine-learning tools are often more accurate and faster than manual labour when trained with good-quality datasets. They are also less expensive and unlike manual labour, machine learning models will improve over time, providing better accuracy as the models are updated over the years. Using machines to detect defects also means that they do not need breaks, decreasing production time. There are several deep learning techniques that are being implemented to perform the task of detecting defects such as Convolutional Neural Networks (CNNs) [1–5] and Image Transformers [6–9].

The architecture proposed should be a significant improvement to the way in which defects in most types of fabrics are identified in the industry. The classification explored in this paper is binary classification, although it is perfectly

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

possible to perform multi-class classification with the same architecture on larger datasets. The authors go into a more detailed overview of the architecture in the upcoming sections.

II. SELECTING A DEEP-LEARNING MODEL

There are several instances of papers that have used image transformation techniques for identifying defects in fabrics. Chi-ho Chan and Grantham K. H. Pang made use of image processing techniques such as image contrasting and Fourier transforms for preprocessing fabric images [6]. This transform converted the image from the spatial domain to the frequency domain and was used for feature extraction. The high-frequency components obtained were useful for detecting defects by extracting features such as variance, skewness, and kurtosis. K.L. Mak et Al [7] also made a defect detection scheme where the important texture features are extracted and used to construct morphological filters for textile fabric defect detection. The performance of these transforms does not meet the current standards for detecting fabric defects by themselves. However, they can be used for preprocessing and feature extraction before being passed on to a machine-learning model [6]. Junpu Wang et Al [8] introduced a Defect transformer that incorporated a CNN and a transformer into an efficient hybrid transformer architecture for defect detection. They emphasized the advantages of a model that uses CNNs to focus on local features and Transformers to focus on global features.

There are numerous cases of Convolutional Neural Networks being used to identify defects in fabrics [8-20,38-51]. JF Jing et Al [10] used a pre-trained deep CNN for transfer learning where the method outperformed many state-of-the-art models when compared against quality and robustness. W Ouyang et Al [2] developed a deeplearning algorithm by combining the techniques of image pre-processing, fabric motif determination, candidate defect map generation, and convolutional neural networks (CNNs). The accuracy tested with this model exceeded 98%. The computational efficiency of CNNs as well as how they excel at capturing local patterns and spatial dependencies make them excellent candidates for detecting fabric defects. B Fang et Al [1] constructed fabric datasets with the help of convolutional neural networks (CNNs) integrated with attention mechanisms and frequency domain filtering to improve the accuracy of the CNN model. The compatibility of the CNNs and frequency domain filtering here is encouraging and could be implemented with similar models. However, there are several advancements in the area of CNN models that have resulted in interesting candidates for advanced CNN models.

S Chakraborthy et Al [11] proposed a novel methodology that demonstrated the application of convolutional neural network (CNN) integrated with other advanced CNN models such as a visual geometric group (VGG), DenseNet, ImageNet, Inception, and Xception deep learning networks to compare model performance. The results concluded that VGG provided better results compared to a simple CNN model. H Xie et Al [12] proposed a robust fabric defect detection method and compared the VGG-16, YOLOV2, and ResNet models to improve the defect localization accuracy. Although the VGG family provides better accuracy against YOLOV2 and other state-of-the-art models, ResNet outperforms VGG due to its innovative residual connections, enabling the successful training of deeper neural networks without suffering from the degradation problem. ResNet was introduced by Kaiming He et Al [13] as a more advanced version of ImageNet. Diving deeper into ResNet, there are several papers that utilize it for defect detection [14–18]. A novel model based on the ResNet architecture called DefectNet was used by Li et Al [14] for defect detection and achieved an accuracy of 91.57%. The architecture of ResNets enhances the information flow, improves gradient propagation, and ultimately leads to higher accuracy in various computer vision tasks, including fabric defect detection.

In order to improve upon the ResNet architecture, we intend to incorporate features from another type of deep learning model known as transformers. Traditionally, convolutional neural networks (CNNs) have been the dominant architecture for image recognition tasks. However, transformers have shown remarkable success in language processing and image classification tasks. Yao et al. [19] proposed a modified Vision Transformer abbreviated as ViT which possessed the ability to compress token vectors with utmost efficacy and achieved superior accuracy over several advanced Transformer models such as SOTA and DeiT. The Vision Transformer (ViT) excels in fabric defect detection due to its ability to effectively capture long-range dependencies and model complex visual patterns. The model's self-attention mechanisms allow it to learn global contextual information, enabling accurate detection and classification of fabric defects, even in the presence of intricate and diverse patterns. Yao et al. also put forth another model named MaxViT which achieved an accuracy of 88.7% by creating a hybrid model with ViT and ImagNet-21K [20]. By

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

combining the strengths of both models, it leverages the local feature extraction capabilities of advanced CNN models and the global self-attention mechanism of the Vision Transformer. This integration allows for better modelling of long-range dependencies and enhances the network's ability to capture both local and global contextual information, leading to improved performance in complex visual tasks such as fabric defect detection. By replacing the ImageNet with the better-performing ResNet [11–13], the authors of this paper propose that a ResNet and ViT hybrid model would incorporate and emphasize the best features of both models.

III. MOTIVATION AND RESEARCH OBJECTIVES

The binary classification of fabric defects is far too time-consuming and error-prone to be done manually at an industrial scale. So, our motivation is the need for an automated state-of-the-art model capable of identifying these fabric defects with high accuracy. Such a model is too complex for classic computer vision-based programs to perform accurately, so several deep-learning models were proposed. A hybrid model consisting of two of the highest-performing models was used here. The research objectives are:

- To apply a deep learning neural network model to identify defects in fabrics.
- To show that the hybrid model performs better than the individual ResNet and ViT models.
- To produce results with high accuracy when given small datasets of a few hundred images.

IV. METHOD

4.1 System Configuration

The data was trained and validated utilizing an Intel Core i9 9820X (10 Cores, 20 Threads, up to 4.10Ghz) CPU and an NVIDIA GeForce RTX 2080 GPU. The code was written in a Python environment and utilized the Tensorflow machine learning framework library. The data were randomly divided into training, testing, and validation using stratified random sampling to ensure the data was homogeneous. The model has an adaptive learning rate with a batch size of 4. The number of training images is sufficient as over a thousand labeled images are available for testing, which is adequate for training a ViT model.

4.2 Dataset Generation

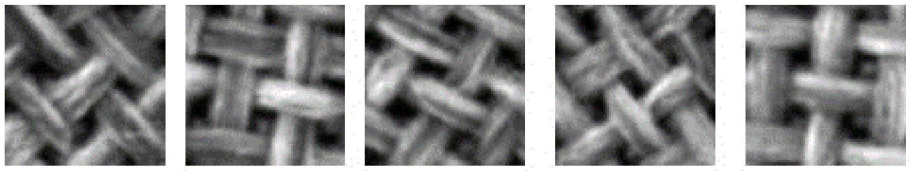

Fig. 1. Images from the dataset representing good fabrics

Fig. 2. Images from the dataset representing damaged fabrics

The dataset is critical when constructing a deep-learning model for fabric defect detection. In a perfect environment, the dataset would be indicative of the kinds of flaws that are most likely to be found in real-life situations. It ought to have a wide variety of fabrics, made of various substances, textures, hues, and designs. A wide variety of fault types, such as stains, tears, holes, snags, and other flaws, should also be included in the dataset. In order for machine learning algorithms to be able to distinguish between various kinds of errors, the dataset must be appropriately labeled. All the pertinent defect categories that are included in the dataset should be covered by the labels, which should be consistent and precise. Another crucial factor to consider is the dataset's size. The machine learning model will be more reliable and accurate as more data becomes accessible. Overfitting, which occurs when the model becomes overly focused on

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

the training data and performs poorly on fresh data, can also be avoided with a large dataset. The ViT model requires a large dataset to train with in order to develop a powerful model. The dataset should also be balanced to ensure the model can reliably detect all types of flaws and prevent it from becoming biased towards any particular defect type. After exploring for a dataset to satisfy all these needs, we arrived at the Textile defect detection dataset as can be seen in Figures 1 and 2. The project is based on the public dataset by MVTec [21]. The main goal of this dataset is to explore self-supervised learning on texture images in order to solve anomaly detection problems and learn a robust representation of texture in lieu of traditional image processing features. The dataset is in .h5 format, organized by groups of defects, and a dataset of patches of 64x64 pixels where patches are sampled at randomized positions, and patches are extracted with different angles. About 2,720 photos of textiles with various kinds and levels of flaws are included in the dataset. The photographs are in the HDF5 format and come in a variety of dimensions, with the majority of them being 1280 x 720 pixels. They have been resized to 64x64 pixels by the model before training. The type of fabric defect found in each photograph has been named, and these labels are provided in a separate CSV file.

The dataset has 6 different classes: ('good', 'color', 'cut', 'hole', 'thread', and 'metal contamination') along with 8 different rotations (0, 20, 40, 60, 80, 100, 120, 140). To balance the dataset, there is an equal number of 'good' and 'damaged' images. There is also an equal number of each class of 'damaged' images. Given an image size, a train and test dataset are available with randomly generated patches. Source images from the train and test datasets are non-overlapping. The labeling of each defect category with a matching numerical code simplifies the processing of the data for machine-learning applications. The dataset is also balanced with each class having the same number of images. This would be useful in ensuring that the model does not have a bias to any particular class of image during classification. All these features make it an ideal dataset for building a model to detect fabric defects. This dataset can also be manipulated by the H5ToStorage object for a binary classification task.

4.3 Preprocessing

Before passing the image through the deep-learning model, preprocessing is performed to reduce noise and improve the quality of the data. The preprocessing is implemented by Fourier transforms as they are often used in various fields such as signal processing, audio, and image analysis, and data compression for this purpose [6–8]. They are transformations that can break down complex signals into their constituent sine and cosine waves of differing frequencies. Essentially, a Fourier transform represents a continuous-time signal, shown as a function of time f(t). The Fourier transform of f(t) is a function F(ω) of frequency where i is the imaginary unit, ω is the angular frequency, and the integral is taken over all time and is given by $F(\omega) = \int f(t)e^{-i\omega t}dx$. This formula breaks a signal down into sine and cosine waves of varying frequencies. The function F(ω) is a representation of the frequency domain which contains information on the strength and phase of the frequency components. Filters can be applied at this stage by altering these features to remove noise. Notch filters are one such filter and can remove noise from an image by enhancing the contrast of an image. It does this by darkening areas of high frequency and lightening areas of low frequency. The advantages of using notch filters include their ability to remove specific frequency components from an image without affecting the rest of the image. Once the filtering process has been completed, the inverse Fourier transform can then transform the signal back from the frequency domain to the time domain. The inverse Fourier transform, Where the integral is taken over all frequencies, is given by: $F(\omega) = (1/2\pi) \int f(t)e^{i\omega t}$ dx. This common form of Fourier transform is known as the discrete Fourier transform (DFT), and it is computed using the fast Fourier transform (FFT). The FFT is an algorithm that can compute DFT with N samples in O(N log N) time, making it practical for real-time processing.

4.4 ResNet-ViT Hybrid Model Architecture

The hybrid model consists of both the Vision Transformer model and the ResNet model as feature extractors for image classification. The models are trained individually and their resulting feature sets are merged to form the final model. A Residual Network (ResNet) is a deep learning model consisting of several residual blocks designed to address the vanishing gradient problem that occurs during training. A residual block is the core part of the architecture and consists of convolutional layers, batch normalization layers, and ReLU activation functions. The input is first passed through a sequence of convolutional layers, followed by batch normalization and ReLU. The output of the final layer is added to the original input called the skip connection. Using the skip connection, the network learns to identify differences

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

between the output and the input of the block. The network learns to optimize the residual function rather than the full mapping, so the gradients can propagate back to the earlier layers. This avoids the vanishing gradient problem, which occurs during the training of deep neural networks. It has a bottleneck architecture, which lowers the number of parameters and computations. It consists of a sequence of 1x1, 3x3, and 1x1 convolutional layers which reduce the dimensionality of feature maps and the computational complexity of the network.

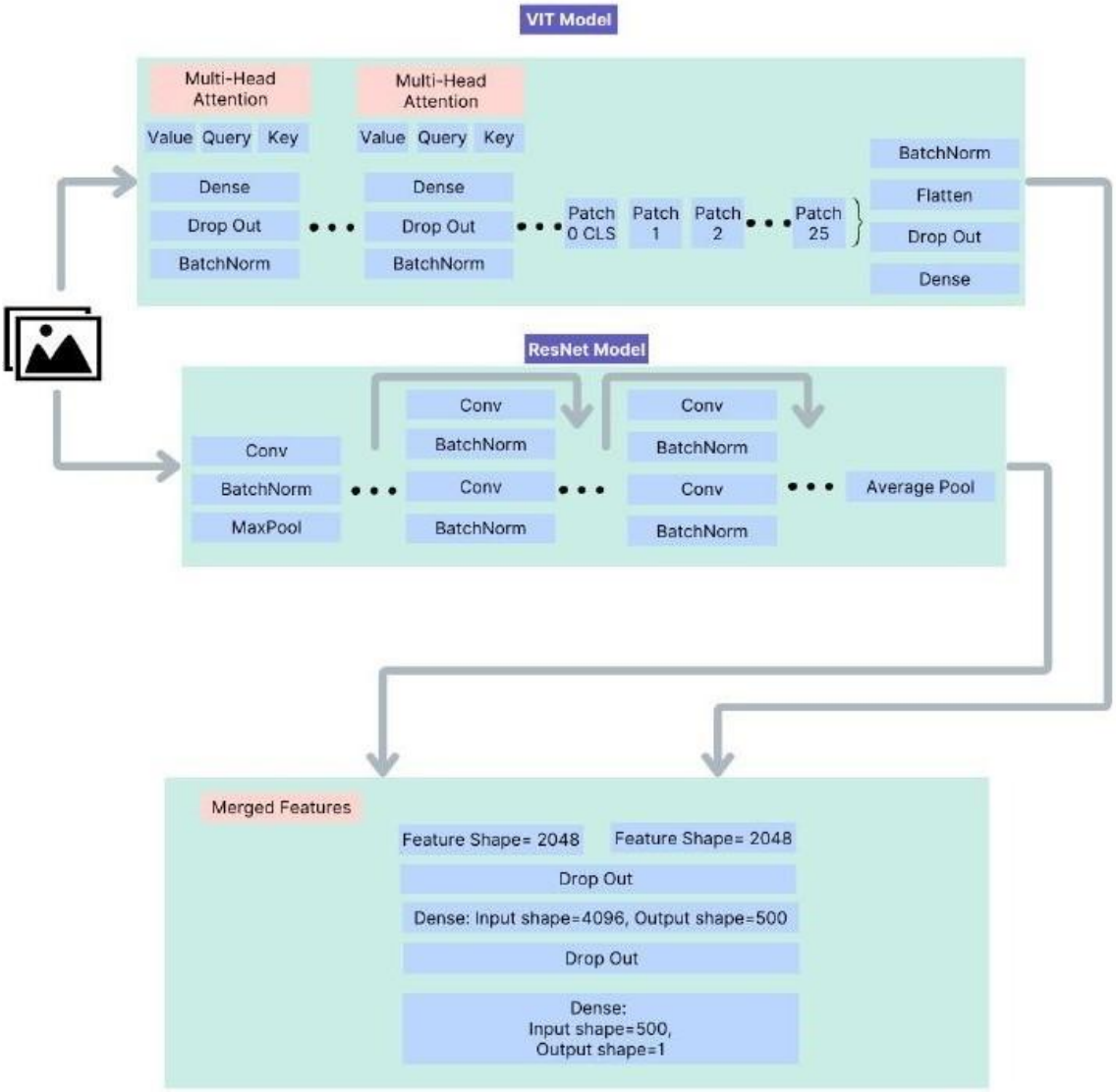

Fig. 3. Architecture of ResNet-ViT hybrid model

The ResNet model used here is known as ResNet50. The ResNet50 architecture begins with one convolutional layer along with 64 filters and a kernel size of 7x7, which is then passed through batch normalization and ReLU, followed by max-pooling with a stride of 2x2. The output is then passed through four stages containing several residual blocks. The first stage has three residual blocks with 64 filters, the second stage has four residual blocks with 128 filters, the third stage has six residual blocks with 256 filters, and the fourth stage has three residual blocks with 512 filters. The number of filters increases as the spatial resolution decreases. After the residual blocks, the output goes through an average pooling layer that computes the average of the feature maps. After which, it is passed through a fully connected layer with 1000 neurons with a softmax activation function.

The Vision Transformer(ViT) is a deep learning model that applies the transformer architecture, which was originally developed for natural language processing tasks and is now used for computer vision. The main idea behind ViT is to

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

break an image down into a sequence of patches and transform it using the transformer architecture. The input image is divided into several non-overlapping fixed-size patches, which are then arranged into a sequence of 1D vectors. These patch embeddings, along with position embeddings that represent the spatial information of the patches, are passed through a transformer encoder. The transformer encoder is made up of several layers that comprise self-attention mechanisms that allow the model to capture relationships between patches and feed-forward neural networks that help in modelling complex interactions within the image. After passing through the transformer encoder, the output sequence is processed by a classification head, typically a fully connected layer, to predict the class labels or perform other downstream tasks.

As can be seen in Figure 3, the training images are passed to both the ResNet and the ViT models. In the Resnet, the final dense layer is removed from the model to obtain the features from the final flattened layer. In the ViT model, the last hidden states from the last attention layer are taken except for the classification token. Those states are then flattened and passed through a dense layer to reduce the shape to resize the output to the same size as the ResNet features. These feature sets are merged and then passed through a few dense layers to obtain the output.

V. RESULTS AND DISCUSSIONS

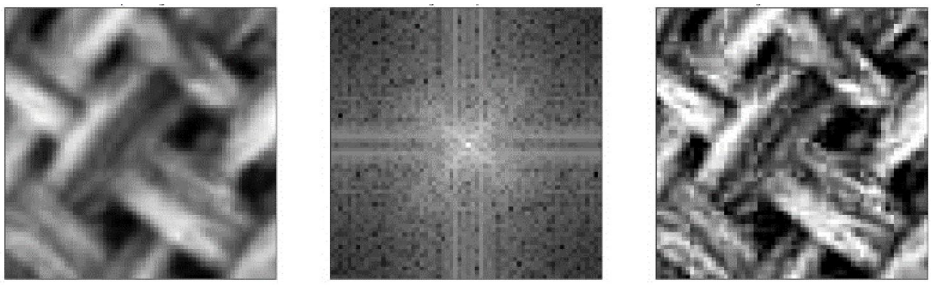

Fig. 4. a) Original image b) Magnitude spectrum of original image c) Image after applying Fourier transform

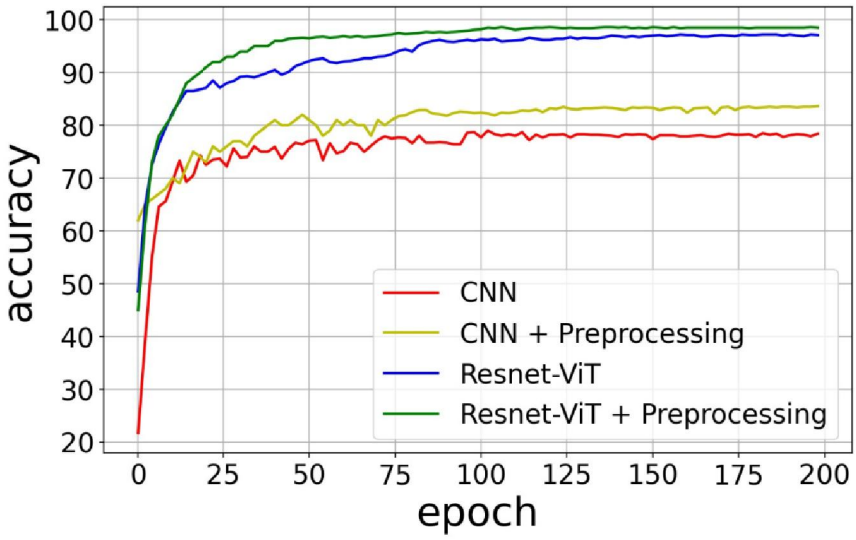

Fig. 5. Validation accuracy of CNN and ResNet-ViT hybrid models

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

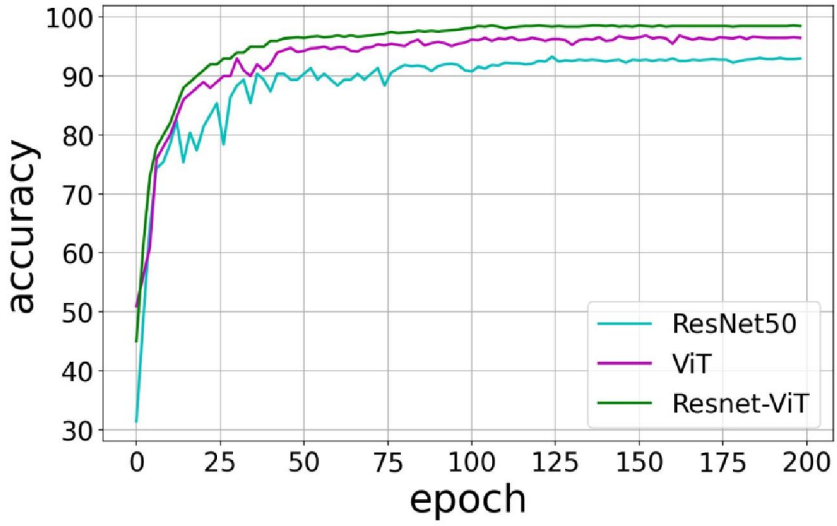

Fig. 6. Validation accuracy of models trained with preprocessing

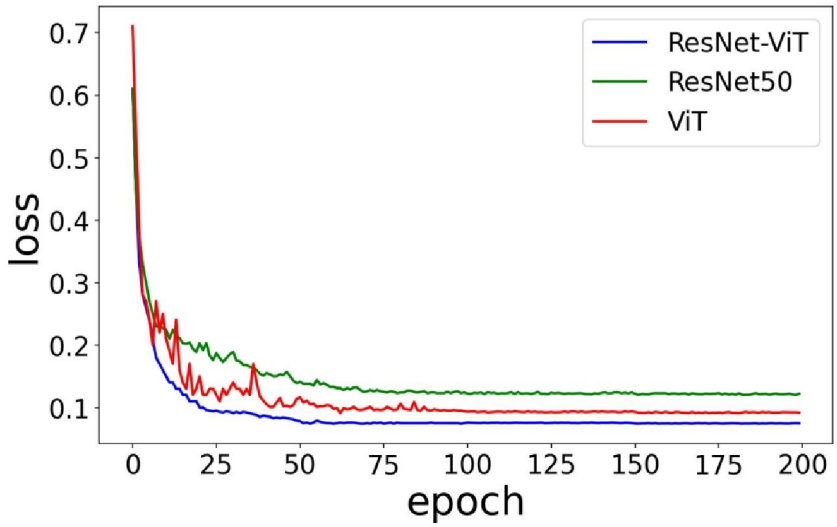

Fig. 7. Validation loss of models trained with preprocessing

TABLE I: Accuracy of models after preprocessing

Model	Training accuracy	Validation Accuracy
CNN	86.7%	83.1%
ResNet50	94.6%	93.4%
ViT	97.2%	96.5%
ResNet-ViT	99.3%	98.5%

The dataset was initially trained with a baseline CNN model. It achieved a poor training and validation accuracy of 82.3% and 78.0% respectively. In order to improve the accuracy, the images were passed through a Fourier transform which produces a magnitude spectrum of the image as shown in Figure 4. The notch filter is applied to the magnitude spectrum and is then passed through the inverse Fourier transform to produce a sharper, noise-free image. These enhanced images are then trained with the CNN model for 200 epochs, which demonstrated an improved 94.6% training accuracy and 93.4% validation accuracy as shown in Table 1 and Figure 5. The model has a loss of 0.21 which is less than the loss of 0.23 from the CNN model trained without preprocessed images. Thus, applying image transforms such as Fourier transforms and performing further fine-tuning of parameters can further improve the accuracy of the CNN models. However, CNN models do not have the efficiency of modern deep learning models. They

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

compute every node in each hidden layer. Modern models like Resnets on the other hand, implement skip-connections that allow the model to skip unnecessary computations that will not lead to any improvement in the model performance. The preprocessed images were trained with the ResNet50 model for 200 epochs and achieved a training accuracy of 94.6% and a validation accuracy of 93.4%, as can be seen in Figure 6. The model also has a loss of approximately 0.14 as shown in Figure 7. This is a significant improvement over the baseline CNN model. The size of the Resnet model is also greater than the CNN models, which also plays a factor in its better performance. One of the main flaws of the Resnet model is that it is prone to overfitting. If the model was nearing the target values and becoming more robust, the loss would decrease and the accuracy would remain the same. When run for 300 epochs, the training accuracy remained the same but the validation accuracy dropped to 89% and the loss increased to 0.31. This is because, once the model identifies the underlying data patterns, it will make incorrect generalizations which result in incorrect predictions. These incorrect predictions lead to significant changes in loss values while the training accuracy remains the same. The preprocessing using Fourier Transforms and notch filters improved the accuracy of the model without increasing the number of epochs and overfitting it.The Visual Transformer was trained with the preprocessed images and achieved a training accuracy of 97.2% and a validation accuracy of 96.5%. The ViT performs better than the ResNet50 due to its ability to retain more spatial information than a ResNet. The skip connections have more impact in ViT than in ResNets as they have a stronger effect on performance and representation similarity. The ViT model has a larger lower layer, which allows it to integrate more global information and identify quantitatively different features. The model did not face overfitting unlike the ResNet50 and had a similar loss of 0.10 when run for 300 epochs, though the accuracy did not significantly improve. However, the ResNet was able to identify thread defects slightly better than the ViT. This could be due to the ViT sacrificing its ability to identify a single class in order to build a more robust, global model. When trained with a dataset of only 200 images, the ViT model performed significantly worse than the ResNet50 and only achieved 86% accuracy while ResNet50 achieved 92%. This is because the ViT model requires a large dataset to construct a robust global model.

The hybrid ResNet-ViT model achieved a training accuracy of 99.3% and a validation accuracy of 98.5% when trained for 200 epochs. When given images without preprocessing, it achieved a lower training accuracy of 98.7% and a validation accuracy of 97.8%. The improvement of preprocessing here is significantly lower here than with the CNN model. This could be due to the hybrid model's ability to identify features that the CNN model could not identify without the removal of noise. Therefore, preprocessing would not improve this model to the same extent as the CNN model. When provided with a smaller dataset of 200 images, the hybrid ResNet-ViT model performed similarly to the ResNet model and achieved a 92% validation accuracy. Thus, the model was able to retain the features of the ResNet50 when the Visual Transformer was not able to perform with the lack of data. Therefore, this model could be used even when insufficient data is provided. However, this model would require significantly more processing than a ResNet. When the hybrid model is run for 300 epochs, the loss only slightly increased to 0.12 from the 0.08 achieved by the model run for 200 epochs. Thus, the overfitting of the ResNet model did not affect the loss of the hybrid model as much as it would have done when run individually.

One noticeable flaw in our model without image transforms is that it struggles to predict defective images with stray threads. This is most likely due to the lack of clearly defined features for this defect. The incorporation of the ResNet noticeably improves the accuracy of detecting this defect. It should be possible to further improve the accuracy of the model by implementing other models such as the Faster RCNN. Although this model does not perform as well as a ResNet, it might provide unique advantages as its structure is far less similar to a ViT. This could make it ideal for the detection of stray threads. Despite these drawbacks, the current model can have a great impact on automating the clothing industry. The replacement of manual labor will lower the cost of production, be less time-consuming, and will have a better performance. Thus, the number of defective products released to the markets will be reduced.

VI. CONCLUSION

In conclusion, In the textile industry, the proposed Fabric Defect detection system using the Resnet-ViT hybrid architectural model is an ideal approach for automating the defect detection process. The preprocessing utilizing Fourier Transforms and notch filters have improved the image quality to an extent visible in the final model's results. The final architecture shows optimistic results in the detection of defects in fabric images with high accuracy and reliability. The

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

model can remarkably decrease the time and money spent and in turn, increase the efficiency and consistency of the fabric defect detection process. The final Resnet-ViT hybrid model with the Fourier transform preprocessing achieves a validation accuracy of 98.5%, which is suitable for the fabric industry. It also performs well when given smaller datasets and the increase in loss when the ResNet is overfitted is significantly reduced. In future work, we can investigate various manufacturing processes and their unique defects. We can also investigate newer, cutting-edge models such as the SAM, Tensormask, and DeepLabV3 and incorporate them into our hybrid models. We can also apply other image transforms, such as the Hough transform, to our models to enhance features. Overall, this study highlights the potential of Deep learning mechanisms and hybrid models in improving the quality and productivity of the textile industry.

The authors report there are no competing interests to declare. No funding was received. The authors contributed equally to this work. The datasets analyzed during the current study are taken from MVTec AD — A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection [21]. The DOI of the source is https://doi.org/10.1109/CVPR.2019.00982.

REFERENCES

[1]. B. Fang, X. Long, F. Sun, H. Liu, S. Zhang, and C. Fang, Tactile-Based Fabric Defect Detection Using Convolutional Neural Network With Attention Mechanism. IEEE Trans. Instrum. Meas. , vol. 71, pp. 1-9, (2022)

[2]. W. Ouyang, J.H. B. Xu, Yuan, X.: Fabric defect detection using activation layer embedded convolutional neural network. IEEE Access 7, 70130–70140 (2019)

[3]. Jha, S.B., Babiceanu, R.F.: Deep cnn-based visual defect detection: Survey of current literature. Computers in Industry 148 (2023)

[4]. Xinghui Dong, C.J.T., Cootes, T.F.: Small defect detection using convolutional neural network features and random forests. Computer Vision – ECCV 2018 Workshops 11132 (2018)

[5]. K. Su, Q.Z., Lien, P.C.: Product surface defect detection based on cnn ensemble with rejection. IEEE Intl Conf on Dependable, Autonomic and Secure Computing, 326–331 (2019)

[6]. Chan, C.H., Pang, G.: Fabric defect detection by fourier analysis. 1999 IEEE Industry Applications Conference 3, 1743–1750 (1999)

[7]. L. Mak, P.P., Yiu, K.F.C.: Fabric defect detection using morphological filters. Image Vis. Comput. 27(10), 1585–1592 (2009)

[8]. Junpu Wang, F.Y.J.W.Z.W. Guili Xu: Transformer-based encoder-decoder model for surface defect detection. Association for Computing Machinery 211, 125–130 (2022)

[9]. Xinglong Feng, X.G., Luo, L.: An improved vision transformer-based method for classifying surface defects in hot-rolled strip steel. J. Phys.: Conf. Ser. 2082 (2021)

[10]. J.F. Jing, H.M., Zhang, H.-H.: Automatic fabric defect detection using a deep convolutional neural network. Color. Technol. 135(3), 213–223 (2019)

[11]. S. Chakraborty, M.M., Parrillo-Chapman, L.: Automatic defect detection for fabric printing using a deep convolutional neural network. Int. J. Fash. Des. Technol. Educ. 15(2), 142–157 (2022)

[12]. Xie, H., Wu, Z.: A robust fabric defect detection method based on improved refinedet. Sensors 20(15) (2020)

[13]. Kaiming He, S.R. Xiangyu Zhang, Sun, J.: Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition 20(15), 770–778 (2016)

[14]. Li, F.L., Xi, Q.: Defectnet: Toward fast and effective defect detection. IEEE Trans. Instrum. Meas. 70, 1–9 (2021)

[15]. J. Li, Y.Q.R.Z. M. Jiang, Ling, S.H.: Intelligent depression detection with asynchronous federated optimization. Complex Intell. Syst. 9(1), 115–131 (2023)

[16]. J. Li, Y.Q.R.Z. M. Jiang, Ling, S.H.: Intelligent depression detection with asynchronous federated optimization. Complex Intell. Syst. 9(1), 115–131 (2023)

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

[17]. Singh, S., Desai, K.: Automated surface defect detection framework using machine vision and convolutional neural networks. Journal of Intelligent Manufacturing 34

[18]. Zhang Y., S.P. Wa S., Y., W.: Pear defect detection method based on resnet and dcgan. Information. 12(10) (2021)

[19]. P. K. Uraon, A.V., Badholia, A.: Steel sheet defect detection using feature pyramid network and resnet. 2022 International Conference on Edge Computing and Applications, 1543–1550 (2022)

[20]. al., Z.T.: Maxvit: Multi-axis vision transformer. Computer Vision – ECCV 2022, 459–479 (2022)

[21]. P. Bergmann, D.S. M. Fauser, Steger, C.: Mvtec ad — a comprehensive real-world dataset for unsupervised anomaly detection. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 9584–9592 (2019)