International Information and Engineering Technology Association
Advancing the World of Information and Engineering

WaveTexNeT: Ensemble Based Wavelet-Xception Deep Neural Network Architecture for Color Texture Classification

Philomina Simon[1*], Uma Vijayasundaram[2]

[1] Department of Computer Science, University of Kerala Thiruvananthapuram, Kerala 695581, India
[2] Department of Computer Science, Pondicherry University Kalapet, Puducherry 605014, India

Corresponding Author Email: philomina.simon@gmail.com

ABSTRACT

Recognizing real visual textures in the nature have been a challenging task since they are complex and stochastic. In spite of several decades of research, classifying the real world color textures are still challenging because of the intricate nature of the textures and the lack of substantial improvement of accuracy in benchmark datasets. Deep Learning techniques have found to be effective in identifying and classifying the texture patterns to a larger extent, but it could not capture spectral information and achieve excellent results for natural images. In this paper, we propose a deep convolutional neural network architecture, WaveTexNeT that combines Wavelet convolutional neural networks (WaveletCNN) and Xception model with luminance information for classifying real-world natural textures. Spectral and spatial features are extracted from WaveletCNN and Xception model. The highlight of the work is the utilization of spectral and spatial information along with luminance for texture classification. A color space image data augmentation technique is proposed that use luminance images from YIQ model for color texture classification. This work also throws light into the significance of luminance information for texture classification. Experimental analysis of the work reports that WaveTexNeT captures better feature representations and outperforms the accuracy obtained using the state-of-the-art methods. WaveTexNeT obtained an accuracy of 90.34% and 95.01% for the describable and material perception texture datasets DTD and FMD respectively.

1. INTRODUCTION

The visual real world is comprised of rich diverse textures present in the nature from histopathological images from cells, tissue abnormalities, defect classification, natural visual scenes. Texture [1, 2] depicts the image surface characteristics such as roughness, irregularity, coarseness, smoothness and it can be used for texture classification. Classifying real life textures have wide applications in machine vision and visually textured object recognition. Even though color is an important cue in interpreting images, we cannot identify a natural image without analyzing the texture. Textures serve as a vital and robust cue in classifying and recognizing regions or objects. The prominent applications include classifying forest species, rocks, wood species, fabrics, satellite images, natural images, food grains, textural defect and tumors. Different color spaces have distinct color representation patterns [3]. The significance of the color spaces other than RGB for better color representations in texture is not often investigated. Better human perception can be modelled by considering the luminance obtained from the different color models. In spite of several decades of research, classifying the real-world color textures are still challenging task because of the complex nature of the textures and the less improvement in accuracy in benchmark datasets [4]. It is challenging to distinguish the texture classes that we encounter in the daily life.

Representing real life textures can be challenging and tedious. Due to the advancements in technology, deep learning and machine learning approaches continue to grab better attention in texture classification research. Machine learning and Deep learning based methods capture the discriminative features in an image. Deep learning can address the challenges of texture classification to a great extent. But still there is a need for better texture representations. This research gap is addressed in the proposed approach. Wavelet based approaches can be effectively used for color texture classification. But this area is still largely unexplored and needs investigation [5]. Van de Wouwer et al. [6] discussed that color along with texture capture more discriminative information and the choice of the color space is very important. So it is better to propose a model that takes advantage of texture features captured in spatial and spectral domain with a color space information for better classification.

WaveTexNeT, the proposed work is developed based on deep learning and wavelet based approaches. This method has good generalization ability and better feature representations that improves the performance of the texture classification. Wavelet based representations [7] generated with Convolutional Neural Network (ConvNet/ CNN) improves the classification accuracy. Wavelets exhibit multiresolution property. So, wavelet texture representations extract features through multilevel wavelet decomposition from high and low frequency components. The innovation of the work is two fold. First objective is to make use of luminance information in deep architectures for classifying color textures. Second objective is to develop a deep architecture that incorporates spectral and

the spatial feature representation along with the luminance information for classifying real world textures.

The contribution of the work is summarized as follows:

- A deep neural network WaveTexNeT is proposed which extracts the spectral and the spatial feature representation along with the luminance information for classifying real world textures.
- A new color space data augmentation technique is proposed which augments luminance images obtained from YIQ color model along with RGB images to perform texture classification.
- Wavelet CNN and Xception are combined in this work to capture the spectral frequency features and spatial information using depth wise separable convolution and skip connections. This model has lesser number of trainable parameters.
- Performance analysis reveals that WaveTexNeT produces excellent results than the state -of- the- art methods in color texture classification in benchmark datasets DTD and FMD.

Section 2 discusses prior literature related to texture representations and wavelet based deep learning methods.Section 3 provides the details regarding the methodology adopted in this research work. Section 4 demonstrates the experimental work and analysis conducted. Section 5 sums up the highlights of the work.

2. LITERATURE SURVEY

With the progress of the hardware infrastructure and computational facilities, various deep learning and machine learning approaches have been discussed in the research field of color texture classification. We have explored CNN based and wavelet based deep learning methods in color texture classification.

2.1 Texture representations using deep learning

Computer vision tasks have widely used different ConvNet deep algorithms for obtaining better evaluation results. Tivive et al. [8] is one of the early works in which the CNN-based texture classification is discussed which focuses on classifying texture patterns in an image. This CNN network preserves the spatial arrangement of the input image and uses the sigmoid function for classification. Dixit et al. [9] discussed an optimized CNN for texture classification. Deep feature-based optimization is performed to optimize CNN features using the nature-inspired Whale Optimization Algorithm for texture classification. Liu et al. [10] proposed genetic algorithm-based CNN, GANet, to handle scale variations in classifying textures. Genetic algorithm adaptively changes the filters applied in CNN hidden layers, for learning more prominent texture patterns. Zhang et al. [11] designed Deep Texture Encoding Network, where the encoding layer transfers the CNN features from pre-trained models thereby capturing domain-specific texture features. Deep neural network architectures, PCANet, RandNet, LDANet [12] are proposed for the classification. It simplifies the processing by utilizing Cascaded PCA, hashing, and block histogram construction. This CNN does not require regularization parameters and optimizers. Dai et al. [13] presented a Bilinear CNN model, FASON that captures the first order and second-order information within the features from the deep network. Tan and Le [14] presented a work EfficientNet architecture by scaling in the network depth,

width and resolution which obtained good results. Sajjadi et al. [15] performed single image super resolution through exploiting automated synthesis of texture content by EnhanceNet which focused on creating realistic textures. Gowda et al. [16] discussed a ColorNet model which investigates the significance of color spaces and deep learning model DenseNet for image classification. Roy et al. [17] presented a fusion based deep learning method TexFusionNet for texture classification where AlexNet and VGG features are fused.

2.2 Wavelet based representations for texture classification

Wavelets are of interest to researchers. A wavelet is a mathematical representation useful for solving problems in signal and image processing domain. 2D-wavelets are used for image processing applications. The concept of wavelet has been successfully implemented for better texture analysis. Deep Wavelet representations are gaining prominence since it can exhibit spectral characteristics of an image. Wavelet based texture representations are found to be efficient for classification tasks for the three decades. Wavelet analyses a texture image into different frequency components at different resolution scales (i.e multiresolution). Multiresolution techniques [5] intend to transform texture images into a representation in which both spatial and frequency information are present. Wavelet decomposition provides a complete texture image representation and performs decomposition based on scale and orientation. During the last decade, multiresolution techniques are increasingly being applied to image processing problems. Global features are quickly and efficiently extracted from images through these techniques. Human visual system is keen in perceiving multiresolution analysis for extracting detail from natural scenes. Wavelets provide sparse representations for processing smooth images.

Arivazhagan et al. [18] studied the significance of using Discrete Wavelet Transform (DWT) for texture analysis. In the work [18], first level decomposition texture image capture LH1, HL1, HH1 denote the detail coefficients and subband LL1 represents approximate coefficients. In the next level of decomposition, LL1 is further divided and decomposed. The features generated from this frequency information is efficient in classifying a textured image. Wavelet pyramid decomposition captures the high and the low frequency components within a texture image. Wavelet based texture representations [19] found to be efficient for classification task in medical field, defect detection, remote sensing from last three decades. Fujieda et al. [20] explored a texture based deep neural networks, Wavelet CNN in which frequency domain-based approach is incorporated into the CNN. Wavelet CNN [21] considered spectral information and performs multiresolution analysis through wavelet transform. The spectral analysis of CNN also captures the scale-invariant features of the texture patterns. Bruna and Mallat [22] developed wavelet scattering network (ScatNet) where convolution filters are represented as gabor wavelets that captures spectral features. These gabor wavelets preserve high frequency content in images. Tao et al. [23] developed WMACapsNet which utilized both spectral and spatial information through multi scale wavelet features and self-attention blocks. Liu et al. [24] presented a CNN network, CWTACapsNet based on multi scale wavelet feature decomposition, quantization and tensor blocks. The tensor blocks exploits the dependencies across the cross channels,which prove the efficiency of CWTACapsNet.

Multilevel convolution network is successfully used for tasks including image restoration [25] and classification of ECG heart beats [26]. Khalil and Adib [26] developed end-to end wavelet architecture for heart disease prediction based on ECG signal features and wavelet features. Usually in image classification tasks, Convolution Neural Network (CNN) are prone to noise interferences while processing. WaveCNet [27] utilizes DWT to decompose the texture feature maps into low and high frequency bands. This WaveCNet work is robust to noise and also improves the classification accuracy. Zhao et al. [28] focused on the significance of wavelet attention modules for image classification. This method extracted better feature representations from attention level based high frequency components and preserved the structures from low frequency components. For addressing the research gap and getting motivated from the above discussed related works, wavelet based deep architecture, WaveTexNeT is proposed in this work.

3. METHODOLOGY-WAVETEXNET

The proposed method WaveTexNeT is inspired from the fact that the ConvNets cannot capture the spectral information which is relevant in processing the from texture images. Wavelet CNN [20, 21] has proven to be a promising method in capturing the spectral features that constitutes the approximation features including the low frequency especially texture and detail features including the high frequency spectral features such as edges, boundaries.

3.1 Spatial-spectral information

Texture images can be processed both in spatial and spectral domain. Spatial information provides the detailed information captured from textures based on the manipulation of pixel intensity values in neighbourhood. Spectral information depicts the frequency related information present in texture images. The low frequency regions depict the smooth areas, while the high frequency regions represent edges, contours and boundaries. Rich set of spatial-spectral features can be captured from texture images. Texture classifications research based on spatial-spectral information needs further attention and investigation [29]. Spatial- Spectral methods have the capability of reducing noise and lessen the influence of noisy pixels. Spatial-spectral methods improve the performance of a texture classification system. Deep CNN architectures provide the spatial information by the successive convolutions with the kernel. CNN captures image local features and merges the extracted the local features to obtain higher-order image features through aggregation and pooling techniques, and classification task can be successfully performed using the higher-order features. Spectral approaches transform the images into frequency domain using a set of spatial filters. The spectral information can be obtained from texture images at different scales and orientations that constitute the features. This approach has been well studied in texture classification and results are promising [18]. Feature extraction in the frequency domain has an advantage. A spatial filter can be easily made selective by enhancing certain frequencies while suppressing the others. This explicit selection of certain frequencies is difficult to control in CNNs. Rather than relying CNNs to learn performing spectral analysis, it would be good to combine the spatial and spectral representation based on a

multiresolution analysis using wavelet transform. So, in this work, spatial and spectral features are captured using Xception model and Wavelet CNN.

3.2 WaveTexNeT architecture

WaveTexNeT investigates different resolutions by four level decompositions through Wavelet CNN and the depth wise separable convolutions by Xception model. In WaveTexNeT, new architecture is proposed by exploiting the spectral multiresolution features from wavelet CNN and high level deep spatial features from the Xception model. The wavelet CNN model is concatenated with the Xception model for better texture feature representations by incorporating the spatial and spectral features. Wavelet CNN make use of Discrete Wavelet Transform (DWT) for capturing multiresolution and spectral information by applying convolution operation and down sampling. The working of WaveTexNeT is given as follows. The input RGB texture images are augmented with the luminance images from YIQ model. These images are fed to wavelet CNN for multiresolution analysis using the 4-level wavelet decomposition. These wavelet coefficients are generated. The input images are also fed to Xception and images undergo depth wise convolution. In Xception, we are training the network with learned weights and the features are generated. Luminance information also assist in getting the prominence in this network and hence textures can be represented better. The features captured from the Wavelet CNN and Xception are concatenated. Then, global average pooling is applied and after that flattening is done to obtain single 1D vector. Dense layer, batch normalization and ReLU activation and dropout is carried out two times. Finally, fully connected layer performs texture classification. Together with Wavelet CNN, Xception and the luminance information obtained a better feature representation and hence better classification for texture images are achieved. WaveTexNeT architecture is provided in Figure 1.
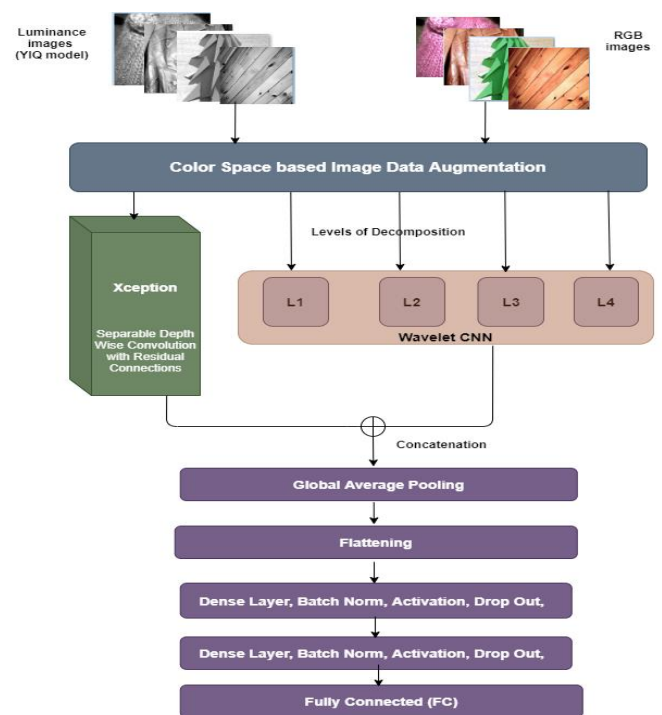

Figure 1. WaveTexNeT architecture

3.3 Color space image augmentation with luminance

Image data augmentation [30, 31] is a technique for providing different variations of image data through geometrical transformations or other methods. hence these techniques address the issues on training using less or limited data. In the work, we propose a new color space augmentation technique where the initial dataset is augmented with luminance images. This is inspired from the fact that texture images can better be represented with other color spaces along with RGB. Here in the work, color information is preserved from RGB images and the texture features from luminance images. Since the dataset DTD and FMD contains limited number of images, we augment the dataset with the luminance images from the YIQ model [32]. Luminance channel captures significant texture features present in the image. Luminance images are captured from the Y channel of the YIQ model. The other data augmentation techniques used are shearing, zooming and horizontal flip on the training data. We generated the luminance images from the Y channel. In the work, I and Q, the chromatic components are not used as RGB preserves the color information. To analyse the efficiency of YIQ model, we compared the approach with luminance images (Y channel) from YCbCrcolor model [33]. Here also, we used only Y images to generate the luminance images.

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.595 & -0.274 & -0.321 \\ 0.211 & -0.522 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 65.48 & 128.55 & 24.97 \\ -37.78 & -74.16 & 111.93 \\ 111.96 & -93.75 & -18.21 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \tag{2}$$

The conversion of RGB to YIQ and RGB to YCbCrcolor space is illustrated in Eqns. (1) and (2).

3.4 Significance of luminance

Texture and color are unique cues for representing and capturing visual scenes of the nature for texture classification. Human vision is more sensitive to luminance details than chrominance in an image [34]. Luminance corresponds to the light energy emitted from the source based on the spectral sensitivity of eye. Luminance channel is significant and it would be good to separate luminance from other color models for extracting the structural details and better feature representation for texture classification problems [35]. The significance of the luminance information needs to be better studied for texture analysis applications. The literature shows that the color spaces can improve accuracy in texture classification systems [36]. The effect of luminance in color models with deep features needs further investigation. It is significant to investigate the influence of luminance information for color texture classification. Since deep networks capture chrominance information than luminance, it is worth exploring the significance of luminance information in WaveTexNeT architecture.

3.5 Wavelet convolution neural network

In Wavelet Convolution Neural Network termed as Wavelet CNN [20], the input images are represented as four multiresolution levels of decomposition which extract better texture features. Wavelet CNN extract features in multiresolution analysis in frequency domain. Wavelet CNN uses the 3x3 convolutional kernels, stride 2 with padding 1x1 for capturing the spectral features. Stride and padding is applied to the input image to lower the feature dimensions. Multiresolution analysis is performed in input images for the decomposition. An illustration of single level wavelet decomposition is depicted in Figure 2. 2D wavelet decompositions are represented in approximation (low) (LL1) and detail (high) wavelet coefficients (LH1, HL1 and HH1) and concatenated for each R, G and B channel. The global average pooling is applied for convolution modules in the final stage and then fully connected layer is applied.
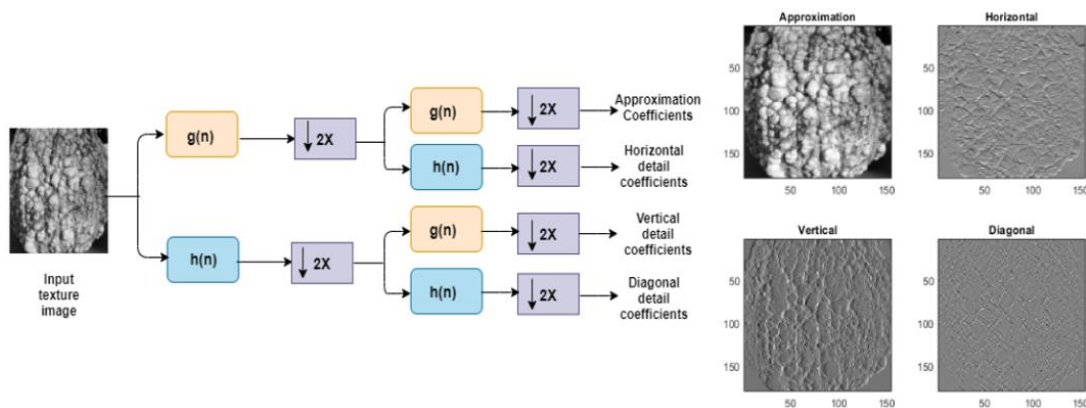

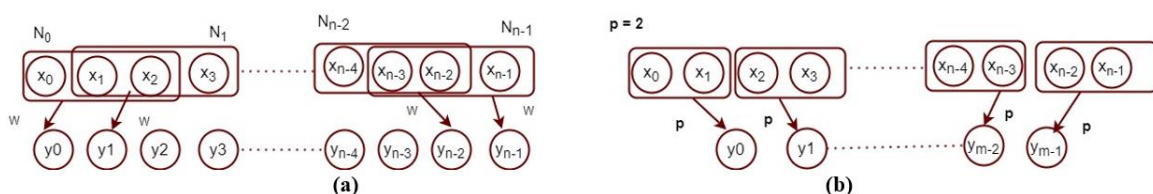**Figure 2.** 2D Wavelet decomposition

Figure 3. Representation of conv operation and pooling in Wavelet CNN

In WaveTexNeT, pooling and the convolution operation is considered as down sampling and filtering thereby establishing a relation between convolutional neural networks and multiresolution decomposition. The frequency domain offers an advantage for feature extraction. By increasing certain frequencies while suppressing others, a spatial filter may be readily made selective. In CNNs, this explicit selection of specific frequencies is difficult to regulate. Wavelet CNN incorporate spectral techniques into CNNs, through multiresolution analysis. This model is trained with images of size 224x224. In wavelet CNN, multiresolution analysis can be considered as repeatedly applying convolution and pooling layers on low-frequency parts with a specific pair of convolution filters.

Convolution layers are expressed as weighted sum of the neighbouring pixel intensities and Pooling layer is expressed as computing the mean and carry out down sampling. The representation of convolution and pooling [20] is given in Figure 3.

The Convolution operation [20] can be expressed as:

$$y = x * w \tag{3}$$

where,

$$w = (w_0, w_1 \ldots \ldots \ldots, w_{o-1}) \in R^o$$

$$x = (x_0, x_1 \ldots \ldots \ldots \ldots, x_{n-1}) \in R^n$$

$$y = (y_0, y_1 \ldots \ldots \ldots \ldots, y_{n-1}) \in R^n$$

x is an input vector in n-dim space, w is the arbitrary weight values or kernel filter and y is the output vector. Pooling is applied to reduce the dimensions of the feature maps just after the convolution processing. Pooling is expressed as

$$y = (x * \boldsymbol{p}) \downarrow p \tag{4}$$

where, p is a parameter that represents the pooling support, $p = \frac{1}{p}, \ldots \ldots \frac{1}{p} \in R^n$ represents averaging filter. The value $p = 2$ indicates that output dimensions are reduced to one half of inputs by computing mean pair wise. $\downarrow$ represents the downsampling operation. Estimation of averaging filter indicates convolution operation via p followed by downsampling operation and using stride value p. From Eq. (3) and (4), we can express y as

$$y = (x * \boldsymbol{h}) \downarrow p \tag{5}$$

where, $\boldsymbol{h}$ is the pair of convolution filters with $p = 2$.

$$x_{low} = (x * \boldsymbol{h}_{low}) \downarrow 2 \tag{6}$$

$$x_{high} = (x * \boldsymbol{h}_{high}) \downarrow 2 \tag{7}$$

The Eq. (6) and (7) relates convolution operation to multi resolution decomposition that considers convolution operation with kernels utilizing low frequency and high frequency information. We can again reframe the equation as:

$$x_{low,s+1} = (x_{low,s} * \boldsymbol{h}_{low,s}) \downarrow 2 \tag{8}$$

$$x_{high,s+1} = (x_{high,s} * \boldsymbol{h}_{high,s}) \downarrow 2 \tag{9}$$

The kernels $h_{high,s}$ and $h_{low,s}$ are considered as wavelet function and scaling function for multi resolution wavelet decomposition respectively. The parameter s in the Eq. (8) and (9) represents the level of decomposition. In the work, we use four levels of wavelet decomposition. The trigger activation function we used is Rectified Linear Unit (ReLU).

3.6 Xception architecture

Xception [37] is a deep architecture expanded as Extreme Xception that is motivated from the Inceptionv3, which entirely relies upon the concept of depth wise separable convolution and the skip connections. There are three modules in Xception namely entry module, middle module and exit module and this CNN model uses the depth wise separable convolutions and the residual connections [38]. A depth-wise separable convolution layer separates each channel of the input and filter distinctly, convolves them by each channel, and later splits one element of 3 channels to be convoluted until all elements have been convoluted. The depth-wise separable convolution layer reduces the number of parameters compared with the conventional convolution layer. The algorithm also has some residual structure that skips over the block of the depth-wise separable convolution layers. Instead, it has the blocks of the conventional convolution layer and the batch normalization layer. Xception Architecture is displayed in Figure 4.
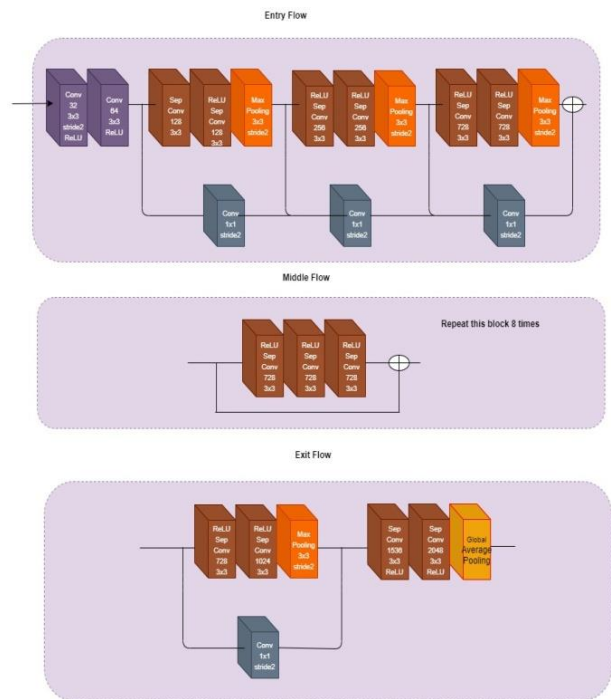

Figure 4. Model representation of Xception

4. EXPERIMENTAL ANALYSIS AND RESULTS

The proposed method is tested with benchmark datasets DTD and FMD. The next section covers the details of the dataset used.

4.1 Datasets

In color texture natural scene classification, the datasets Describable Texture Dataset (DTD) and Flickr Material

Dataset (FMD) are popularly used datasets to illustrate the various real life and complex textures that we encounter in the day today life. These are intricate datasets and assessed to be the challenging benchmark data for the texture classification.

Describable Textures dataset (DTD) is a colored-texture database [39] that contains the real-world textures in nature such as cobwebbed, braided, dotted, blotchy and frilly. DTD contains images with attribute-based texture representations, In DTD, texture images are extracted from the web rather than being captured or generated in a controlled setting. In color texture classification, DTD depicts the various natural describable textures annotated with adjectives that describe the properties. DTD comprises of texture images that have describable attributes and it captures unique patterns within a texture. The DTD dataset is considered to be challenging because of the complex nature of the real life textures and the lesser interclass variance. DTD has 120 images each for 47 texture categories adding up to 5640 images. The glimpse of DTD dataset is given in Figure 5.

Figure 5. View of DTD dataset

The Flickr Material Database (FMD) [40] is another challenging texture dataset that captures various categories of the material appearances. FMD was created with the objective of acquiring wide range of material surfaces [41]. Within the fabric class, there are four sets of fabric surfaces that exhibit different material properties, colors and sizes. FMD provide a view of realistic surface textures that account to the material recognition that is encountered in the everyday life. FMD dataset has 10 classes, and each class contains 100 images. FMD is developed to comprehend the human perception of material classes and also to design Artificial Intelligence (AI) based systems for material recognition. It is still challenging to develop computer vision systems that match human performance on FMD dataset. A glimpse of the FMD can be seen in Figure 6.

Figure 6. View of FMD dataset

4.2 Experimental setup

The experiments are conducted in system configuration Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz, 64 bit Operating system, x64 based processor, 8GB RAM, NVIDIA GeForce MX150. All the experiments are performed in Python Colab Pro environment. Python Keras and TensorFlow are used for implementation. The experiments are implemented on challenging benchmark DTD and FMD datasets. Since the number of training images are less for the datasets, they are augmented with the luminance (Y channel) images from YIQ model for each class along with the RGB images for learning the deep features. In the experiments on DTD, each class have 240 images with 47 classes each add up to a total of 11280 images used for training and testing. In FMD, each class have 200 images with 10 classes adding to a total of 2000 images used for training and testing. These datasets are split randomly and 70% of data are used for training and 30% of data are used for testing.

4.3 Learning parameters

Total number of parameters for the WaveTexNeT model is 34,055,319 where number of trainable parameters is 33,987,863 and non-trainable parameters amount to 12,928. The number of parameters of the proposed model is less than VGG19, ResNet101, ResNet152, InceptionResNet, NasNetLarge. The batch size is selected as 8 for both the datasets. The default parameters of WaveletCNN and Xception architecture are used in this implementation. The loss function used for multiclass texture classification is categorical cross entropy which captures the error rate as loss function in deep learning models. The optimizer used in the work is Stochastic Gradient Descent (SGD). An optimizer is used in deep architecture to train a deep model and minimise the residual. Learning rate used is 0.01. SGD with momentum helps to compute the gradient vectors faster in proper orientations and results in faster convergence. Drop out is fixed as 0.5 to prevent the overfitting. Accuracy curve is generated by computing the accuracy over 50 and 100 epochs on FMD and DTD respectively with learning rate 0.01. The parameters used in WaveTexNeT are summarized in Table 1.

Table 1. Parameters used for WaveTexNeT

Proposed Model		DTD dataset	FMDdataset
	Accuracy	90.34%	95.01%
	No of Epochs	100	50
WaveTexNeT(Luminance from YIQ model)	Batch Size	8	8
	Learning rate	0.01	0.01
	Optimizer used	SGD	SGD

4.4 Result analysis

The detailed analysis of the experiments conducted is illustrated in this section. Wavelet based method, WaveTexNeT is tested on DTD and FMD to prove the efficiency in classifying the natural textures. In this section, experimental results are analyzed.

4.4.1 DTD dataset

Table 2. Precision, recall & F1 score on DTD with YIQ luminance

Class	Precision	Recall	F1 score
Banded	0.89	0.86	0.87
Blotchy	0.80	0.82	0.81
Braided	0.92	0.97	0.95
Bubbly	0.94	0.89	0.91
Bumpy	0.88	0.83	0.86
Chequered	1.00	0.93	0.96
Cowebbed	0.88	0.93	0.91
Cracked	0.89	0.90	0.90
Cross hatched	0.89	0.81	0.85
Crystalline	0.95	0.96	0.95
Dotted	0.93	0.97	0.95
fibrous	0.91	0.89	0.90
fleckled	0.82	0.93	0.87
Freckled	1.00	0.94	0.97
Frilly	0.96	0.93	0.94
Gauzy	0.84	0.89	0.86
Grid	0.82	0.86	0.84
Grooved	0.77	0.96	0.85
Honey combed	0.84	0.96	0.90
Interlaced	0.97	0.93	0.95
Knitted	0.99	0.96	0.97
Lacelike	0.92	0.96	0.94
Lined	0.95	0.83	0.89
Marbled	0.86	0.83	0.85
Matted	0.88	0.92	0.90
Meshed	0.82	0.85	0.84
Paisley	0.91	0.97	0.94
Perforated	0.82	0.94	0.88
Pitted	0.88	0.82	0.85
Pleated	0.97	0.92	0.94
Polka dotted	1.00	0.94	0.97
Porous	0.94	0.83	0.88
Potholed	0.97	0.96	0.97
Scaly	0.93	0.86	0.89
Smeared	0.91	0.81	0.85
Spiralled	0.85	0.88	0.86
sprinkled	0.93	0.90	0.92
stained	0.94	0.82	0.87
Stratfied	0.93	0.90	0.92
Stripped	0.92	0.97	0.95
Studded	0.90	0.88	0.89
Swirly	0.91	0.89	0.90
Veined	0.96	0.94	0.95
Waffled	0.89	0.97	0.93
Woven	0.90	0.86	0.88
Wrinkled	0.88	0.88	0.88
zigzagged	0.95	1.00	0.97
Over all Accuracy 90.34 %			

The results obtained on DTD dataset are discussed. Figure 7(a) and figure 7(b) illustrates the accuracy value and the loss value obtained on DTD using WaveTexNeT. As observed, the learning curve is smooth and it reached convergence within 100 epochs. Accuracy curve shows that an accuracy of 90.34% is obtained. Precision- Recall- F1 score values obtained on

DTD using WaveTexNeT are shown in Table 2. F1 score for the proposed approach on DTD resulted in a high F1 Score of 0.97 for 4 classes viz. Freckled, Knitted, Potholed and Zigzagged. The proposed approach is compared with other existing approaches in terms of accuracy and the results are shown in Table 3. Cimpoi et al. [42] used combination of fisher vector CNN (FV-CNN) and fully connected (FC-CNN) to get an accuracy of 69.80% which is based on VGG model. Simon and Uma [43] developed a texture classification method for classifying deep features with support vector machine (SVM) classifier and obtained 66.49% accuracy. Tao et al. [23] used a wavelet multi attention capsule network (WMACapsNet) which explores spatial and spectral features. The graphical analysis of accuracy for the DTD dataset is provided in Figure 8. Cimpoi et al. [39] also proposed another method based on Improved Fisher Vector (IFV) and Deep Convolutional network Activation Features (De-CAF) which obtained an accuracy of 66.7%. Dai et al. [13] presented a Bilinear CNN model, FASON, captured an accuracy of 72.9% on DTD dataset. Mao et al. [44] discussed an end-to-end Deep residual pooling network based on obtained an accuracy of 76.62%. Xue et al. [45] used Deep Pooling Network obtained 73.2% accuracy. Fujieda et al. [20] discussed a wavelet convolution neural network for classifying the textures that reported an accuracy of 59.8%. Liu et al. [24] used CWTACapsNet, with wavelet and compressed tensor self attention model to get 81.52%. Simon and Uma [46] utilizes deep features and luminance information along with the machine learning classifier and produced an accuracy of 73.63%. WaveTexNeT obtained a good accuracy of 89.01% for luminance from YCbCr, 90.34% for luminance from YIQ color space. It is clear that the model is performing well. From this result analysis, it is proved that the proposed WaveTexNeT along with YCbCr and YIQ color model is performing better when compared to conventional state-of-the-art methods.
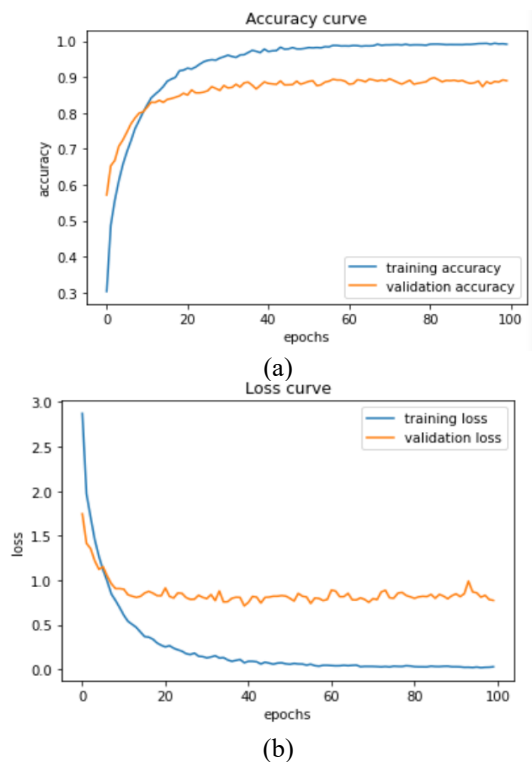

(a)

(b)

Figure 7. (a) Accuracy curve and (b) Loss curve of WaveTexNeT for DTD dataset

Table 3. Accuracy analysis on DTD -WaveTexNeT model

Authors	Method	Accuracy (%)
Cimpoi et al. [42]	FC-CNN + FV-CNN	69.80
Simon and Uma [43]	Deep Features + SVM	66.49
Cimpoi et al. [39]	IFV+DeCAF	66.70
Dai et al. [13]	FASON(conv4+ conv5)	72.90
Fujieda et al. [20]	Wavelet CNN	59.80
Xue et al. [45]	Deep Encoding Pooling Network	73.20
Simon and Uma [46]	DeepLumina	73.63
Liu et al. [24]	CWTACapsNet	81.52
Tao et al. [23]	WMACapsNet	79.52
Mao et al. [44]	Deep Residual Pooling Network	76.62
WaveTexNeT (Proposed Method)	**Wavelet CNN+ Xception + Luminance(YCbCr)**	**89.01**
WaveTexNeT (Proposed Method)	**Wavelet CNN+ Xception + Luminance (YIQ)**	**90.34**

Figure 8. Graphical representation of accuracy analysis on DTD

4.4.2 FMD dataset

The Figure 9(a) and Figure 9(b) illustrate the accuracy value and the loss value respectively for WaveTexNeT on FMD dataset. This curve is generated by computing the accuracy over 50 epochs. As observed, the learning curve is smooth and it reached convergence within 50 epochs. This curve obtained an accuracy of 95.01%. Precision, Recall and F1 score on FMD is shown in Table 4. F1 score for the proposed approach on FMD resulted in a high F1 score of 0.97 for 2 classes viz. Fabric and Foliage. The comparative analysis with existing approaches is performed and the results are shown in Table 5. The graphical representation of accuracy analysis on FMD is given in Figure 12. The proposed work, WaveTexNeT is analysed with the luminance from two color spaces, YIQ and YCbCr. In order to assess the efficiency, extensive experiments have been carried out with the color spaces YIQ and YCbCr. The model FC-CNN + FV-CNN proposed by Song et al. [47] obtained an accuracy of 83.20% on FMD. The model [43] developed based on deep features and SVM produced an accuracy of 84.50% accuracy. Cimpoi et al. [39] developed a model based on Improved Fisher Vector (IFV) and Deep Convolutional network Activation Features (De-CAF) secured an accuracy of 65.50%. Bell at al. [48] used the Scale invariant feature transform (SIFT) and IFV for getting 69.60% accuracy. Deep Lumina [46] obtained an accuracy of 90.15%. Mao et al. [44] presented a method Deep residual pooling network obtained an accuracy of 85.72%. Jbene et al. [49] Performed an analysis with Xception and SIFT-FV to obtain the performance of 86.10% on FMD. Experiments conducted demonstrate that the results in YIQ color model are excellent and promising. WaveTexNeT demonstrated an exceptional accuracy of 94.18% for YCbCr and 95.01% for
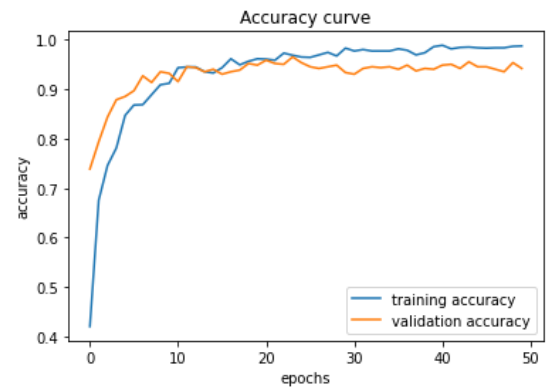
YIQ luminance on DTD. So it is inferred, that the proposed WaveTexNeT model is providing better accuracy.

Table 4. Precision, recall and F1 score on FMD with YIQ Luminance
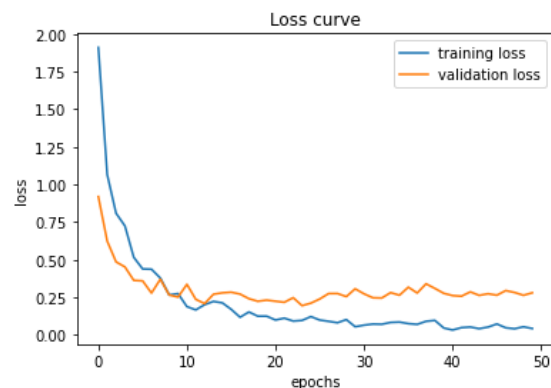
Class	Precision	Recall	F1 score
Fabric	0.98	0.95	0.97
Foliage	0.95	0.98	0.97
Glass	0.97	0.93	0.95
Leather	0.94	0.97	0.95
Metal	0.95	0.93	0.94
Paper	0.95	0.95	0.95
Plastic	0.95	0.93	0.94
Stone	0.92	0.95	0.93
Water	0.98	0.93	0.96
Wood	0.92	0.97	0.94
Overall Accuracy 95.01%			

(a)

(b)

Figure 9. (a) Accuracy curve and (b) Loss curve of WaveTexNeT for FMD dataset

Table 5. Accuracy analysis on FMD -WaveTexNeT model

Authors	Method	Accuracy (%)
Song et al. [47]	FC-CNN + FV-CNN	83.20
Simon and Uma [43]	Deep Features + SVM	84.50
Cimpoi et al. [39]	IFV+DeCAF	65.50
Bell et al. [48]	SIFT-IFV+fc7	69.60
JBene et al. [49]	Xception + SIFT-FV	86.10
Simon and Uma [46]	DeepLumina	90.15
Mao et al. [44]	Deep Residual Pooling Network	85.72
WaveTexNeT (Proposed Method)	Wavelet CNN+ Xception + Luminance (YCbCr)	**94.18**
WaveTexNeT (Proposed Method)	Wavelet CNN+ Xception + Luminance (YIQ)	**95.01**

The confusion matrix for the WaveTexNeTanalysed for the color spaces YIQ and YCbCr on FMD datset is displayed in Figure 10 and Figure 11 respectively.
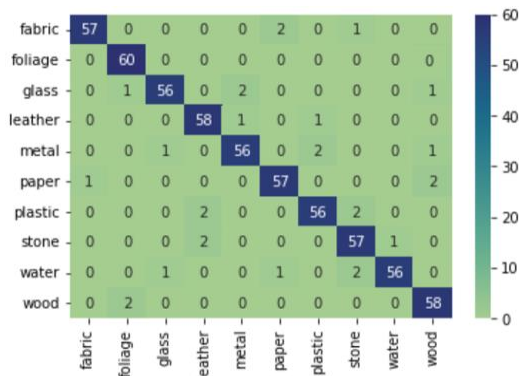

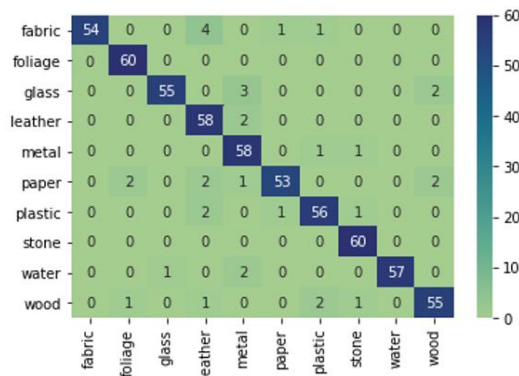**Figure 10.** WaveTexNeT - Confusion Matrix on FMD in YIQ color space

Figure 11. WaveTexNeT - Confusion Matrix on FMD in YCbCrcolor space
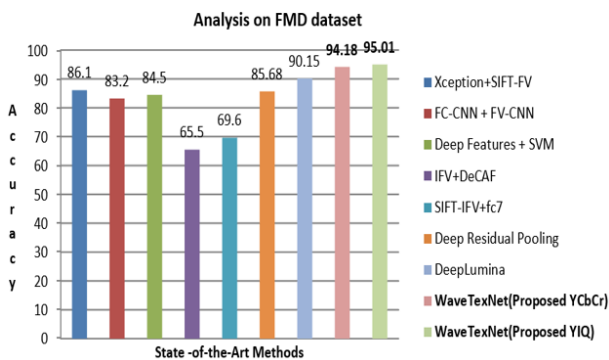

Figure 12. Graphical representation of accuracy analysis on FMD

5. CONCLUSION

In this work, a deep learning method is proposed, WaveTexNeT that combines the wavelet and deep learning approach from Wavelet CNN and Xception model respectively for visual natural texture classification. This deep network combines the spatial and the spectral features of the describable natural textures with luminance information. A new color space image data augmentation technique is introduced that augments luminance images for the color texture classification. Experiments are conducted to assess the efficiency of the WaveTexNeT quantitatively and estimated confusion matrix. The texture class-wise precision, recall and F1-score on DTD and FMD are also evaluated by conducting the experiments. Extensive experiments have been carried out with YIQ and YCbCr color models. Experiments demonstrate that luminance from YIQ proved to be a better color model for texture classification. The proposed method obtained excellent and promising accuracy of 90.34% on DTD and 95.01% on FMD challenging texture datasets than other state -of -the -art techniques.

REFERENCES

[1] Petrou, M.M., Kamata, S.I. (2021). Image processing: dealing with texture. John Wiley & Sons. https://doi.org/10.1002/047003534X

[2] Mirmehdi, M. (2008). Handbook of texture analysis. Imperial College Press. https://doi.org/10.1142/p547

[3] Velastegui, R., Yang, L., Han, D. (2021). The importance of color spaces for image classification using artificial neural networks: A review. In International Conference on Computational Science and Its Applications, pp. 70-83. https://doi.org/10.1007/978-3-030-86960-1_6

[4] Liu, L., Chen, J., Fieguth, P., Zhao, G., Chellappa, R., Pietikäinen, M. (2019). From BoW to CNN: Two decades of texture representation for texture classification. International Journal of Computer Vision, 127(1): 74-109. https://doi.org/10.1007/s11263-018-1125-z

[5] Livens, S., Scheunders, P., Van de Wouwer, G., Van Dyck, D. (1997). Wavelets for texture analysis, an overview. Sixth International Conference on Image Processing and Its Applications, 2: 581-585. https://doi.org/10.1049/cp:19970958

[6] Van de Wouwer, G., Scheunders, P., Livens, S., and Van Dyck, D. (1999). Wavelet correlation signatures for color texture characterization. Pattern Recognition, 32(3): 443-451. https://doi.org/10.1016/S0031-3203(98)00035-1

[7] Liu, J.W., Zuo, F.L., Guo, Y.X. (2021). Research on

improved wavelet convolutional wavelet neural networks. Applied Intelligence, 51: 4106-4126. https://doi.org/10.1007/s10489-020-02015-5

[8] Tivive, F.H.C., Bouzerdoum, A. (2006). Texture classification using convolutional neural networks. In TENCON 2006-2006 IEEE Region 10 Conference, pp. 1-4. https://doi.org/10.1109/TENCON.2006.343944

[9] Dixit, U., Mishra, A., Shukla, A., Tiwari, R. (2019). Texture classification using convolutional neural network optimized with whale optimization algorithm. SN Applied Sciences, 1(6): 1-11. https://doi.org/10.1007/s42452-019-0678-y

[10] Liu, L., Chen, J., Zhao, G., Fieguth, P., Chen, X., Pietikäinen, M. (2019). Texture classification in extreme scale variations using GANet. IEEE Transactions on Image Processing, 28(8): 3910-3922. https://doi.org/10.1109/TIP.2019.2903300

[11] Zhang, H., Xue, J., Dana, K. (2017). Deep ten: Texture encoding network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 708-717. https://doi.org/10.1109/CVPR.2017.309

[12] Chan, T.H., Jia, K., Gao, S., Lu, J., Zeng, Z., Ma, Y. (2015). PCANet: A simple deep learning baseline for image classification? IEEE transactions on image processing, 24(12): 5017-5032. https://doi.org/10.1109/TIP.2015.2475625

[13] Dai, X., Yue-Hei Ng, J., Davis, L.S. (2017). Fason: First and second order information fusion network for texture recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7352-7360. https://doi.org/10.1109/CVPR.2017.646

[14] Tan, M., Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning, pp. 6105-6114. https://doi.org/10.48550/arXiv.1905.11946

[15] Sajjadi, M.S., Scholkopf, B., Hirsch, M. (2017). Enhancenet: Single image super-resolution through automated texture synthesis. In Proceedings of the IEEE International Conference on Computer Vision, pp. 4501-4510. https://doi.org/10.1109/ICCV.2017.481

[16] Gowda, S.N., Yuan, C. (2018). ColorNet: Investigating the importance of color spaces for image classification. In Asian Conference on Computer Vision, pp. 581-596. https://doi.org/10.1007/978-3-030-20870-7_36

[17] Roy, S.K., Dubey, S.R., Chanda, B., Chaudhuri, B.B., Ghosh, D. K. (2020). Texfusionnet: an ensemble of deep Cnn feature for texture classification. In Proceedings of 3rd International Conference on Computer Vision and Image Processing, pp. 271-283. https://doi.org/10.1007/978-981-32-9291-8_22

[18] Arivazhagan, S., Ganesan, L. (2003). Texture classification using wavelet transform. Pattern Recognition Letters, 24(9-10): 1513-1521. https://doi.org/10.1016/S0167-8655(02)00390-2

[19] Scheunders, P., Livens, S., Van de Wouwer, G., Vautrot, P., Van Dyck, D. (1998). Wavelet-based texture analysis. International Journal on Computer Science and Information Management, 1(2): 22-34. https://doi.org/10.1.1.40.1818/

[20] Fujieda, S., Takayama, K., Hachisuka, T. (2018). Wavelet convolutional neural networks. arXiv preprint arXiv:1805.08620. https://doi.org/10.48550/arXiv.1805.08620

[21] Fujieda, S., Takayama, K., Hachisuka, T. (2017).

Wavelet convolutional neural networks for texture classification. arXiv preprint arXiv:1707.07394. https://doi.org/10.48550/arXiv.1707.07394

[22] Bruna, J., Mallat, S. (2013). Invariant scattering convolution networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8): 1872-1886. https://doi.org/10.1109/TPAMI.2012.230

[23] Tao, Z., Wei, T., Li, J. (2021). Wavelet multi-level attention capsule network for texture classification. IEEE Signal Processing Letters, 28: 1215-1219.https://doi.org/10.1109/LSP.2021.3088052

[24] Liu, X., Shan, C., Zhang, Q., Cheng, J., Xu, P. (2021). Compressed wavelet tensor attention capsule network. Security and Communication Networks, 9949204. https://doi.org/10.1155/2021/9949204

[25] Liu, P., Zhang, H., Lian, W., Zuo, W. (2019). Multi-level wavelet convolutional neural networks. IEEE Access, 7: 74973-74985. https://doi.org/10.1109/ACCESS.2019.2921451

[26] Khalil, M., Adib, A. (2020). An end-to-end multi-level wavelet convolutional neural networks for heart diseases diagnosis. Neurocomputing, 417: 187-201. https://doi.org/10.1016/j.neucom.2020.07.056

[27] Li, Q., Shen, L., Guo, S., Lai, Z. (2021). WaveCNet: wavelet integrated CNNs to suppress aliasing effect for noise-robust image classification. IEEE Transactions on Image Processing, 30: 7074-7089. https://doi.org/10.1109/TIP.2021.3101395

[28] Zhao, X., Huang, P., Shu, X. (2022). Wavelet-attention CNN for image classification. Multimedia Systems, 28(3): 915-924. https://doi.org/10.1007/s00530-022-00889-8

[29] Xu, J.L., Gowen, A.A. (2020). Spatial-spectral analysis method using texture features combined with PCA for information extraction in hyperspectral images. Journal of Chemometrics, 34(2): e3132. https://doi.org/10.1002/cem.3132

[30] Shorten, C., Khoshgoftaar, T.M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1): 1-48. https://doi.org/10.1186/s40537-019-0197-0

[31] Xiao, Y., Decencière, E., Velasco-Forero, S., Burdin, H., Bornschlögl, T., Bernerd, F., Warrick, E., Baldeweck, T. (2019). A new color augmentation method for deep learning segmentation of histological images. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 886-890. https://doi.org/10.1109/ISBI.2019.8759591

[32] Robinson, G.S. (1977) Color edge detection. Optical Engineering, 16(5): 479-484. https://doi.org/10.1117/12.7972120

[33] CCIR Recommendation. 601-2. (1990). Encoding parameters of digital television for studios. International Telecommunications Union, Geneva. https://cd.sc.vilga.org/DOCS/mice/MPEG/ccir601.html.

[34] Al-Mualla, M., Canagarajah, C.N., Bull, D.R. (2002). Video coding for mobile communications: efficiency, complexity and resilience. Elsevier.

[35] Prabhakar, K.R., Srikar, V.S., Babu, R.V. (2017). Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. IEEE International Conference on Computer Vision (ICCV), pp. 4724-4732. https://doi.org/10.1109/ICCV.2017.505

[36] Tkalcic, M., Tasic, J.F. (2003). Colour spaces: perceptual,

historical and applicational background. The IEEE Region 8 EURCON 2003. Computer as a Tool, 1: 304-308. https://doi.org/10.1109/EURCON.2003.1248032

[37] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251-1258. https://doi.org/10.1109/CVPR.2017.195

[38] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90

[39] Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., Vedaldi, A. (2014). Describing textures in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3606-3613. https://doi.org/10.1109/CVPR.2014.461

[40] Sharan, L., Rosenholtz, R., Adelson, E. (2009). Material perception: What can you see in a brief glance? Journal of Vision, 9(8): 784-784. https://doi.org/10.1167/9.8.784

[41] Sharan, L., Rosenholtz, R., Adelson, E.H. (2014). Accuracy and speed of material categorization in real-world images. Journal of Vision, 14(9): 12-12. https://doi.org/10.1167/14.9.12

[42] Cimpoi, M., Maji, S., Vedaldi, A. (2015). Deep filter banks for texture recognition and segmentation. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 3828-3836. https://doi.org/10.1109/CVPR.2015.7299007

[43] Simon, P., Uma, V. (2020). Deep learning based feature extraction for texture classification. Procedia Computer Science, 171: 1680-1687. https://doi.org/10.1016/j.procs.2020.04.180

[44] Mao, S., Rajan, D., Chia, L.T. (2021). Deep residual pooling network for texture recognition. Pattern Recognition, 112: 107817-107817. https://doi.org/10.1016/j.patcog.2021.107817

[45] Xue, J., Zhang, H., Dana, K. (2018). Deep texture manifold for ground terrain recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 558-567. https://doi.org/10.1109/CVPR.2018.00065

[46] Simon, A.P., Uma, B.V. (2022). DeepLumina: A method based on deep features and luminance information for color texture classification. Computational Intelligence and Neuroscience. https://doi.org/10.1155/2022/9510987

[47] Song, Y., Li, Q., Feng, D., Zou, J.J., Cai, W. (2016). Texture image classification with discriminative neural networks. Computational Visual Media, 2(4): 367-377. https://doi.org/10.1007/s41095-016-0060-6

[48] Bell, S., Upchurch, P., Snavely, N., Bala, K. (2015). Material recognition in the wild with the materials in context database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3479-3487. https://doi.org/10.1109/CVPR.2015.7298970

[49] Jbene, M., El Maliani, A.D., El Hassouni, M. (2019). Fusion of convolutional neural network and statistical features for texture classification. In 2019 International Conference on Wireless Networks and Mobile Communications (WINCOM), pp. 1-4. https://doi.org/10.1109/WINCOM47513.2019.8942469