Mid-Semester Examination Assignment

Course: EE 678 - Wavelets

Date: 17 September 2024

Group Members:

Mrudul Jambhulkar	21d070044
Bhavik Yadav	21d070090
Akhilesh Chauhan	21d070010

Contents

Chapter 1

Group Information and References

(a) Group Information

Title: A CNN-RNN Hybrid Model with 2D Wavelet Transform Layer for Image Classification
Group Number: 14
Group Members: Bhavik Yadav - 21d070090 , Mrudul Jambhulkar - 21d070044 , Akhilesh Chauhan - 21d070010

(a) References

The reference consulted for this report is :

1. Z. Dong, R. Zhang and X. Shao, ”A CNN-RNN Hybrid Model with 2D Wavelet Transform Layer for Image Classification,” 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), Portland, OR, USA, 2019, pp. 1050-1056, doi: 10.1109/ICTAI.2019.00147.

Chapter 2

Literature Survey

2.1 INTRODUCTION

The advancements in Machine Learning (ML) has led in Convolutional Neural Network (CNN) models being in great demand for ML tasks like object detection, image classification, natural language processing, etc. Currently, the models operate primarily in the spatial domain, i.e. given an image, features or results would be extracted based only on the pixel values. Thus, the arrangement of pixels is used to determine important features. This has mainly helped in identifying basic features like edges, slopes, shapes from the initial layers of a CNN model and interpolating it to identify further complex features.

However, there is still progress to be made in the spectral analysis of an image. Spectral analysis focuses mainly on how the frequency of pixel intensities change. This is needed in various cases like enhancing an image by removing the noise frequencies or identifying unclear edges which couldn't be done accurately in spatial domain. This has particularly important applications in the medical and space industries.

Current models employ the use of 2D Wavelet Transform between two layers leading to development of wavelet CNNs. An example of one can be found in [1] where the learned features and the 2D wavelet transform are combined to build a new architecture. Therefore the use of Discrete Wavelet Transform (DWT) in the case of learning new features is fairly exciting for us.

2.2 THE 2D WAVELET TRANSFORM LAYER

In this work, 2D Wavelet Transform is modeled as a filter. The two-directional RNN is also used to compress the input to half the dimension to capture certain features that a normal CNN might miss in large images. CNNs sometimes struggle in looking at long-range dependencies. For example, in a panoramic image where related pixels are a bit apart, or when there are similar objects far apart in an image. This is where RNN helps by capturing these relationships.

After filtering this compressed input with the 2D Wavelet Transform, it is convolved with the learned matrix to generate the output. The relation is given by the equation in the paper, as follows:

$$O^{(i)} = \sum_{c=0}^{C-1} \left(I'^{(i)} * \phi_j + \left| I'^{(i)} * \psi_{j,\theta} \right| \right) * A$$

We can find the weight matrix by applying the Gaussian and Locally Smoothed Mean (LSM) filters. The algorithm discussed in [1] provides an iterative method using an LSM filter for refining the weights. The Gaussian filter is needed to remove the noise and ensure the updated weight matrix transitions smoothly.

First, a matrix is initialised using Gaussian filtering and then the LSM filter is applied for each input and output channels. Then the algorithm repeatedly updates the learnable matrix A by multiplying with the filter output. The Gaussian filter simply applies a weighted average function at a point by using the Gaussian function values as the weights. This helps in smoothening or blurring the output image to reduce the noise. The filter function is given by the Gaussian function:

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2} \right)$$

where (x, y) are the coordinates of a point, and σ is the standard deviation.

The LSM filter helps in locally stabilising the weight values leading to better weight updates. The LSM filter works by simply averaging over the pixels values over a small window.

For computing the memory cost, we can consider the number of learning parameters. Considering C_i input parameters, C_{i+1} output parameters, and a $S \times S$ convolutional kernel, the memory cost will be:

$$S^2 C_i C_{i+1}$$

Every kernel has C_i weights combined with the distinct weights of the output layer. In [1], the scale factor J and the number of orientations K are also considered, adding additional JK parameters.

For the computational cost, a traditional CNN has a cost of $S^2 C_{i+1}$ per input. The 2D Wavelet Transform employs a different type of cost, namely Dual-Tree Complex Wavelet Transform (DTCWT), which involves 4 DWTs per input. Each tree in DTCWT produces one part of the complex wavelet coefficient (either real or imaginary) using wavelet transforms. While it is more computationally extensive than DWT, it provides better accuracy and shift invariance. Thus, the total DTCWT cost is as follows:

$$8S(1 - 2^{-2J})$$

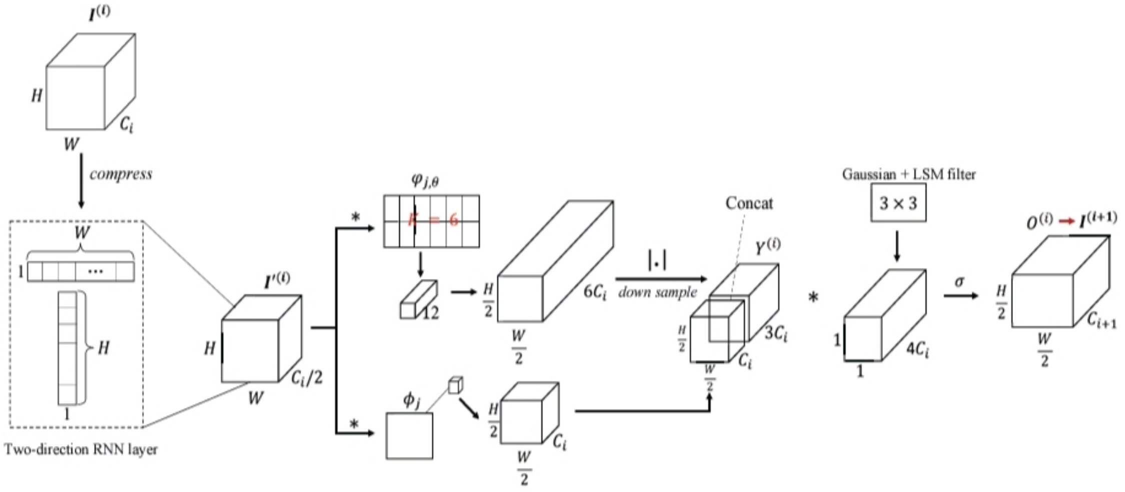

Fig. 1. Block Diagram of proposed 2D Wavelet Transform (2D-WT) Layer.

Figure 2.1: Working of the 2D-DWT Layer [1]

This adds up to the total combining cost, where the coefficients are combined using dyadic wavelet combination. This reduces the spatial size to a quarter, thus the net cost is traditionally lower than for a CNN when the number of output channels is greater than 4

2.3 THE CNN-RNN HYBRID MODEL

In the design of 2D-WT layers, we adopt a VGG-like network as the basis architecture. As indicated in Fig. given below, our hybrid CNN-RNN model utilizes one single-stream deep network wherein 2D-WT layers are added through replacing the last two convolutional layers in the primary network stream, which are depicted in dashed boxes. Output layers: The output layers comprised the 2D-WT layers with responses at different hierarchical levels; the sum of which appeared in a common output layer. The two Fully Connected (FC) layers computed the final classification probabilities.

Refer to Table given below for details about our CNN-RNN hybrid model, built upon the VGG-11 base architecture. Pooling and batch-normalization layers are introduced between each pair of convolutional layers. Input image size = 3×H×W.

The final prediction in the baseline VGG-like networks may lose significant feature details due to the downsampling effect of the pooling layer. We introduce skip connections that link the final prediction layer with earlier layers that have preserved finer feature details. Architecture of the skip connections is depicted in Fig. given below. A 2D-WT layer with batch normalization will form the skip layer connecting the features from the low-scale and high-scale convolutional layers.

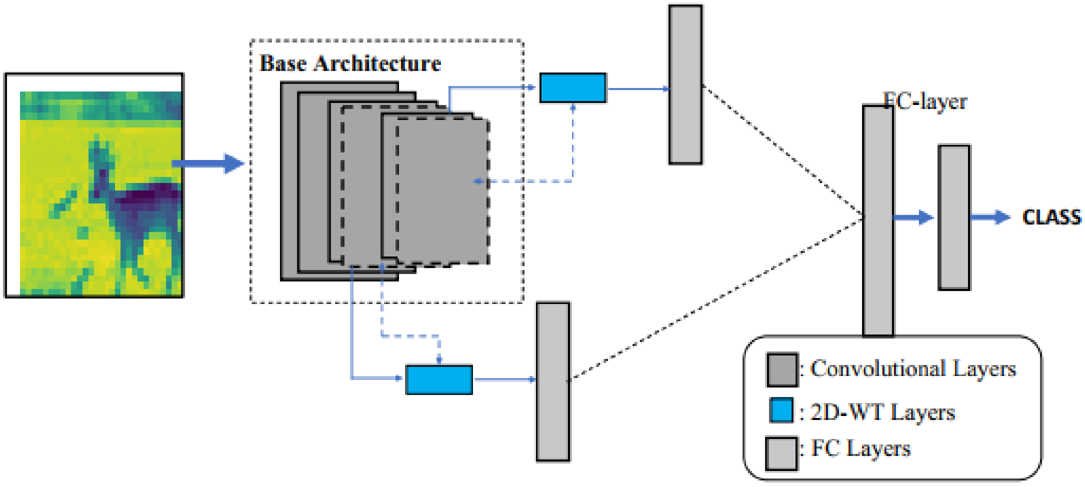

Figure 2.2: CNN-RNN hybrid network with 2D-WT skip layers.

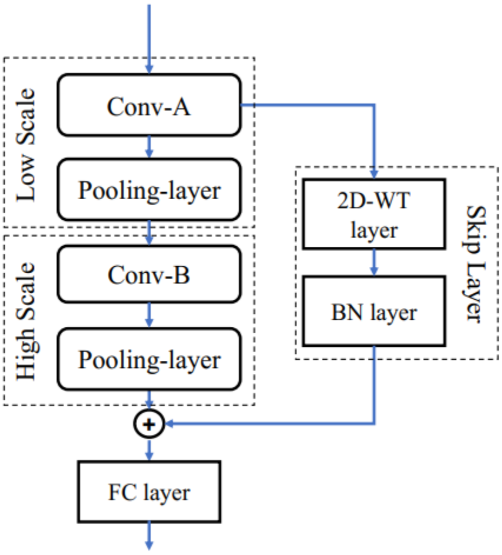

Figure 2.3: The architecture of skips in CNN-RNN hybrid network.

Table 2.1: CNN-RNN Hybrid Architecture Based on VGG-11.

Layer's Name	**Output Size**
conv-A	$64 \times \frac{H}{2} \times \frac{W}{2}$
conv-B	$128 \times \frac{H}{4} \times \frac{W}{4}$
conv-C	$256 \times \frac{H}{8} \times \frac{W}{8}$
conv-D	$256 \times \frac{H}{16} \times \frac{W}{16}$
Conv-E	$512 \times \frac{H}{16} \times \frac{W}{16}$
2D-WT	$512 \times \frac{H}{32} \times \frac{W}{32}$
conv-F	$512 \times \frac{H}{32} \times \frac{W}{32}$
2D-WT	$512 \times \frac{H}{32} \times \frac{W}{32}$

2.4 EXPERIMENTS

The paper validates the 2D-WT layer-based CNN-RNN hybrid model by performing several experiments. The authors have demonstrated the performance of the model on three well-known public datasets: CIFAR-10, CIFAR-100, and Tiny ImageNet. CIFAR-10 and CIFAR-100 both have 32x32 pixel images that fall into 10 and 100 categories, respectively, while Tiny ImageNet contains super small 64x64 pixel images across 200 classes. Experiments are conducted in PyTorch, and a single NVIDIA GTX 1070 is used for each experiment.

This bottom-up architecture can be understood as a VGG-like network (TABLE 2); six convolutional layers are used for CIFAR, and eight convolutional layers are applied for Tiny ImageNet. Hyperparameters that remain consistent across experiments include an initial learning rate of 0.01, weight decay set to 10^{-5}, and a batch size of 128. There is a learning rate decay during training: for the CIFAR model, a learning rate of 0.01 is used for the first 80 epochs, after which it decays to 0.001 for another 40 epochs. For the Tiny ImageNet model, the learning rate starts at 0.01 and decays to 0.001 after 30 epochs, with an additional 15 epochs of training.

For ablating analysis, an ablation study(Table 3) on the contribution of the 2D-WT layer is performed by comparing 2D-WT to convolutional layers called invariant layers. The experiments lead the better performance of network variants involving the 2D-WT layer on the performances of the network variants with invariant layers on the three mentioned datasets. The results point out the fact that the inclusion of the 2D-WT layer widely enhances the classification accuracy and robustness, irrespective of which convolutional layer was substituted. This depicts the superior extraction feature ability of the 2D-WT layer, capturing both spatial and spectral information rather than the conventional layers.

In addition to the comparisons at the layer level, the paper discusses the benefits of skip-layer connections between lower and higher convolutional layers that protect the details of features because of down-sampling. This architecture of skip layers is specially designed to further enhance classification results.

As shown in Table 4, the network-level experiments further reveal that the proposed hybrid CNN-RNN model with enhanced 2D-WT outperforms the baseline VGG architectures in terms of classification accuracy. For example, on CIFAR-10 and CIFAR-100, the

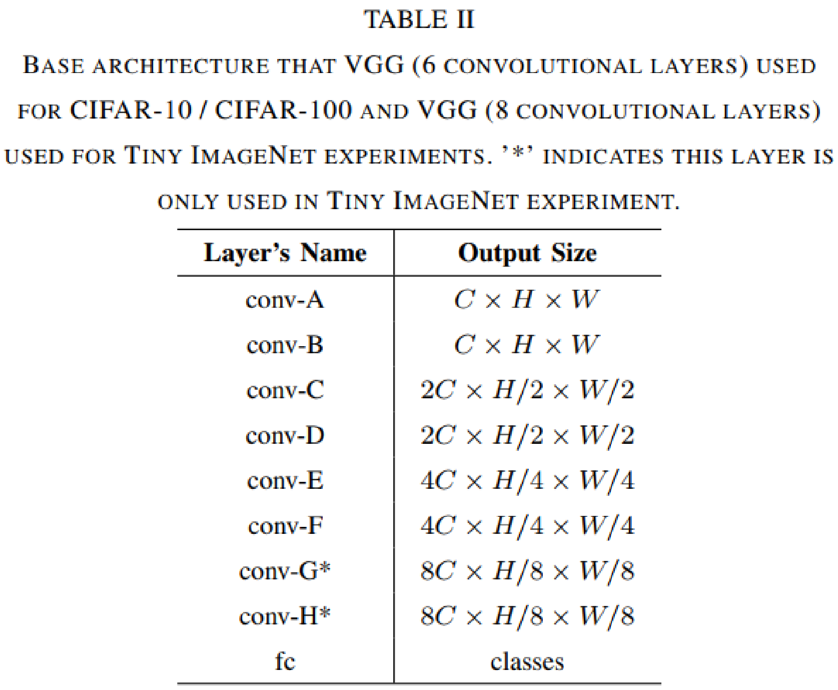

TABLE II

BASE ARCHITECTURE THAT VGG (6 CONVOLUTIONAL LAYERS) USED FOR CIFAR-10 / CIFAR-100 AND VGG (8 CONVOLUTIONAL LAYERS) USED FOR TINY IMAGENET EXPERIMENTS. '*' INDICATES THIS LAYER IS ONLY USED IN TINY IMAGENET EXPERIMENT.

Layer's Name	**Output Size**
conv-A	$C \times H \times W$
conv-B	$C \times H \times W$
conv-C	$2C \times H/2 \times W/2$
conv-D	$2C \times H/2 \times W/2$
conv-E	$4C \times H/4 \times W/4$
conv-F	$4C \times H/4 \times W/4$
conv-G*	$8C \times H/8 \times W/8$
conv-H*	$8C \times H/8 \times W/8$
fc	classes

Figure 2.4: TABLE 2(of the paper)

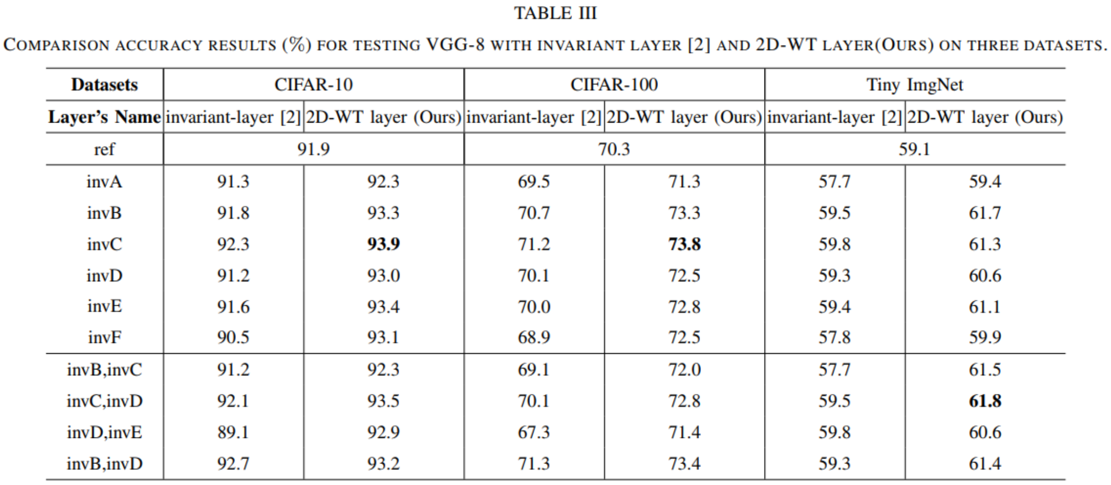

TABLE III

COMPARISON ACCURACY RESULTS (%) FOR TESTING VGG-8 WITH INVARIANT LAYER [2] AND 2D-WT LAYER(OURS) ON THREE DATASETS.

Datasets	CIFAR-10		CIFAR-100		Tiny ImgNet	
Layer's Name	invariant-layer [2]	2D-WT layer (Ours)	invariant-layer [2]	2D-WT layer (Ours)	invariant-layer [2]	2D-WT layer (Ours)
ref	91.9		70.3		59.1	
invA	91.3	92.3	69.5	71.3	57.7	59.4
invB	91.8	93.3	70.7	73.3	59.5	61.7
invC	92.3	**93.9**	71.2	**73.8**	59.8	61.3
invD	91.2	93.0	70.1	72.5	59.3	60.6
invE	91.6	93.4	70.0	72.8	59.4	61.1
invF	90.5	93.1	68.9	72.5	57.8	59.9
invB,invC	91.2	92.3	69.1	72.0	57.7	61.5
invC,invD	92.1	93.5	70.1	72.8	59.5	**61.8**
invD,invE	89.1	92.9	67.3	71.4	59.8	60.6
invB,invD	92.7	93.2	71.3	73.4	59.3	61.4

Figure 2.5: TABLE 3 (of the paper)

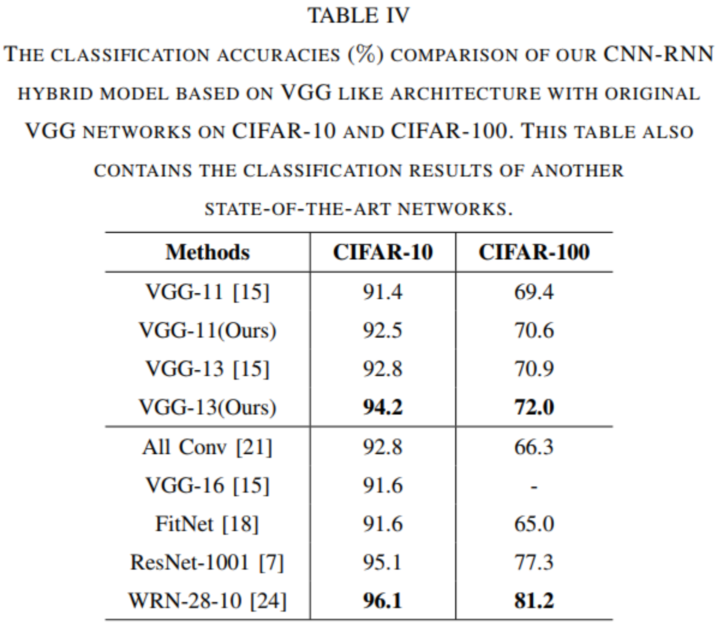

TABLE IV

The classification accuracies (%) comparison of our CNN-RNN hybrid model based on VGG like architecture with original VGG networks on CIFAR-10 and CIFAR-100. This table also contains the classification results of another state-of-the-art networks.

Methods	**CIFAR-10**	**CIFAR-100**
VGG-11 [15]	91.4	69.4
VGG-11(Ours)	92.5	70.6
VGG-13 [15]	92.8	70.9
VGG-13(Ours)	**94.2**	**72.0**
All Conv [21]	92.8	66.3
VGG-16 [15]	91.6	-
FitNet [18]	91.6	65.0
ResNet-1001 [7]	95.1	77.3
WRN-28-10 [24]	**96.1**	**81.2**

Figure 2.6: TABLE 4 (of the paper)

VGG-11 model with 2D-WT layers is able to achieve a higher accuracy than its traditional counterpart. The performance is comparable to other state-of-the-art architectures, such as ResNet-100 and WRN-28-10, but at fewer parameters and convolutional layers. This demonstrates the efficiency of the proposed model, which balances high accuracy with computationally efficient performance.

Fig. 2.6 Graphs for the accuracy curves of classification models during the epochs of training of the model. As the performance of the model is considered to be the best ever attained to date, the effectiveness of the model can be observed from the graph. The x-axis on the graph represents the epoch of training, and y-axis denotes the accuracy of the classification model. It is evidently observed that the hybrid CNN-RNN model with 2D-WT layers is remarkably more accurate than the architectures, VGG-like. Hence, from this criterion also, the contribution of 2D-WT layers gets reinforced.

Besides, the practicality of the model can be seen through the Fig. 2.7 with visual test results from the CIFAR-10 dataset. The classification results of images are right as a probability of being categorized in categories is shown with a bar chart on the right side. This visual result proves that the hybrid network consisting of 2D-WT layers with CNN-RNN is highly applicable.

Table 5. Comparison between number of parameters and convolutional layers with proposed architecture and other state-of-the-art models. Hybrid architecture has a nice trade-off balance for both high performance and efficiency computationally. The inclusion of skip layers does augment the no of parameters, but the model remains unaffected in terms of accuracy; these are areas that should be explored in further work.

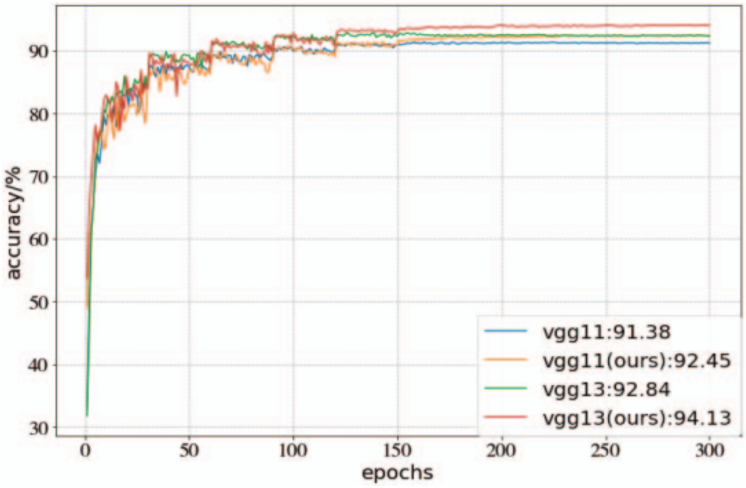

Figure 2.7: Visual classification results of test sets in CIFAR-10.

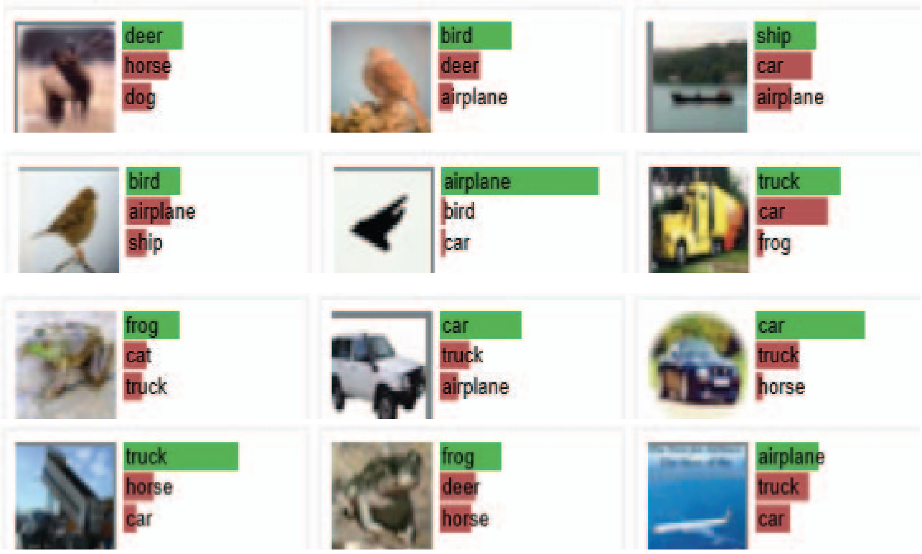

Figure 2.8: Visual classification results of test sets in CIFAR-10.

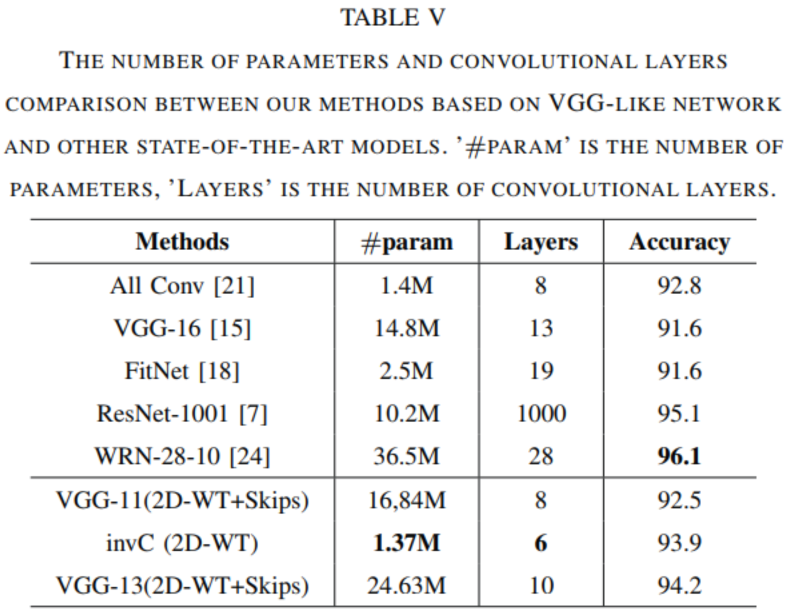

TABLE V

The number of parameters and convolutional layers comparison between our methods based on VGG-like network and other state-of-the-art models. ’#param’ is the number of parameters, ’Layers’ is the number of convolutional layers.

Methods	#param	Layers	Accuracy
All Conv [21]	1.4M	8	92.8
VGG-16 [15]	14.8M	13	91.6
FitNet [18]	2.5M	19	91.6
ResNet-1001 [7]	10.2M	1000	95.1
WRN-28-10 [24]	36.5M	28	**96.1**
VGG-11(2D-WT+Skips)	16,84M	8	92.5
invC (2D-WT)	**1.37M**	**6**	93.9
VGG-13(2D-WT+Skips)	24.63M	10	94.2

Figure 2.9: TABLE 5(of the paper)

2.5 CONCLUSION

In short, the application of wavelet transforms mainly the 2D-Wavelet Transform (2D-WT) layers within deep learning models has been a promising approach towards better feature extraction for image classification. Related literature indicates that 2D-WT produces strong frequency decomposition based on better accuracy as well as higher robustness in the classification process. In addition, the hybrids in CNN-RNN architectural form a strong potential system that can leverage the spatial feature extraction capabilities and sequential recognition patterns of such a network. Hybrid architecture-based models have been presented with such potential utility, reviewed in this paper for possible performance more so than traditional networks, especially on complex datasets. All of these models, and especially the CNN-RNN hybrids with 2D-WT layers, represent promising directions for future work in such image processing tasks as object recognition and segmentation.

Chapter 3

Wavelets and Machine Learning Structures

(b) Wavelets and Convolutional Neural Networks/ Machine Learning Structures/ Deep Learning Structures

Explain the connection(s) between wavelets/ filter banks and convolutional neural networks/ machine learning structures/ deep learning structures that you have explored. Bring out specifically, how wavelets have come into the realm of convolutional neural networks/ machine learning structures/ deep learning structures in the reference(s) that you have studied, in some depth.

Solution : A Convolutional Neural Network (CNN) is a type of deep learning algorithm that offers superior performance with image, speech, or audio signal inputs. CNNs are widely used in image processing for tasks such as image classification, detection, and segmentation. In this paper, a 2D wavelet transform (WT) layer is employed in a CNN-RNN hybrid model to improve the accuracy of image classification. A VGG-like architecture with multiple layers is used, where the last two layers are replaced by 2D-WT layers.

The 2D-WT layer computes the wavelet transform operation, given by the following equation:

$$WT(I(c,u(x,y))) = \{I(c,u(x,y)) * \phi_J(u(x,y)), |I(c,u(x,y)) * \psi_{j,\theta}(u(x,y))|\}$$

where $\phi_J(u(x,y))$ is the low-pass filter, and $\psi_{j,\theta}$ is the mother wavelet filter. The layer combines the invariant terms and low-pass terms and convolves them with the weight matrix A to produce an output that serves as input to the next layer. Convolving with the mother wavelet $\psi_{j,\theta}$ (the high-pass filter) extracts high-frequency information, while convolving with the low-pass filter $\phi_J(u(x,y))$ captures low-frequency information. The output $O(i)$ is computed as:

$$O(i) = \sum_{c=0}^{C-1} (I'(i) * \phi_j + |I'(i) * \psi_{j,\theta}|) * A$$

where $I'(i)$ is the compressed input $I(i)$, generated by a two-direction RNN layer. The operation involves convolving $I'(i)$ with both the low-pass filter ϕ_j and the high-pass filter $\psi_{j,\theta}$, followed by combining these results and convolving with the weight matrix A to produce the output.

The 2D-WT layer, combined with batch normalization, acts as a skip layer that connects the outputs of final convolution layers with earlier ones to recover information lost during downsampling .

Chapter 4

Economy and Explainability in Machine Learning/ Deep Learning

(c) Economy and Explainability/ Interpretability

What does economy mean in the context of machine learning/ deep learning/ neural networks? What does explainability/ interpretability mean in the context of machine learning/ deep learning/ neural networks?

Economy refers to making the CNN efficient in terms of the usage of resources such as reducing the number of units, minimizing the number of layers, or lowering memory costs (i.e., the number of parameters) and computational costs. In the context of CNNs, economy is achieved by creating a network that is less spread out and more compact, while still maintaining high performance. Efficient design can lead to faster training and inference times, reduced power consumption, and overall better scalability, especially in large-scale models or real-time applications.

Explainability refers to how wavelets assist in providing explanations for what the CNN is learning, as opposed to allowing the network to independently learn its own features. Wavelets help in uncovering the relationships between input data attributes and model outputs, thus enabling us to explain the nature and behavior of the machine learning or deep learning model.

Interpretability involves understanding how wavelets contribute to interpreting the model's weights and features to determine the final output. Wavelets provide transparency by making it easier to observe the inner workings of the machine learning or deep learning model. This transparency allows us to understand exactly why and how the model is generating its predictions, which is achieved by analyzing the transformations and features the model learns during training.

Chapter 5

Benefits of Wavelets in Machine Learning Systems

(d) Benefits of Wavelets/ Filter Banks in Machine Learning

Explain the benefits that have accrued, from this union of wavelets/ filter banks and convolutional neural networks/ machne learning structures/ deep learning structures. In particular, emphasize and identify the benefits in terms of economy and/or explainability/ interpretability.

- **Economy:**
  - **Number of layers :** The CNN-RNN-2D WT hybrid model(with VGG-11 and VGG-13 as the base network) was compared with networks such as All Conv , VGG-16 , FitNet , ResNet-100 and WRN-28-10 . The hybrid network takes fewer layers than other models and provides reasonable accuracy .
  - **Memory Cost :** For a traditional convolutional layer with C_i input channels and C_{i+1} output channels and kernel size S the number of parameters is give by:

    $$\#params = S^2 C_i C_{i+1}$$

    .

    In this hybrid model , number of input channels are first compressed from C_i to $Ci/2$. Consider scale J=1 and number of orientations K=6 , so in this case the number of parameters will be given by :

    $$\#params = (JK+1)\frac{C_i}{2}C_{i+1} = \frac{7}{2}C_iC_{i+1}$$

    .

    In VGG-16 or VGG-19 models kernel size is 3X3 . Thus using 2D-WT layers instead of the traditional convolutional layers clearly reduced the number of parameters .

The hybrid model was compared with All Conv , VGG-16 , FitNet , ResNet-100 and WRN-28-10 and was known to take fewer parameters than them .

  - **Computational Cost :** A traditional convolutional layer with kernel size S has S^2C_{i+1} multiplies per input pixel. The hybrid model utilises dual-tree complex wavelet transform (DTCWT) for each input channel . A regular discrete wavelet transform has $2S(1-2^{-2J})$ multiplies for scale J, (DTCWT) has 4 discrete wavelet transforms (DWTs) for an input , so total cost = $8S(1-2^{-2J})$. Usually , $S = 6$ and $J = 1$, thus cost = 36 . After including the cost of combining process ,

$$cost = C_{i+1} + 36$$

For most of the CNNs $C_{i+1} = 10$ so the computational cost for the 2D-WT layers is much less than traditional convolutional layer .

- **Explainability/ interpretability :** The WT layer helps in learning different harmonics from the feature map. The compressed feature map, obtained after the two LSTM layers, is passed to the 2D-WT layer. The WT layer decomposes this compressed map to capture both high-frequency and low-frequency information.

  This decomposition is performed by convolving the feature map with the mother wavelet $\psi_{j,\theta}(u(x,y)) = 2^{-j}\psi(2^{-j}R_{-\theta}u(x,y))$ and the low-pass scaling function $\phi_J(u(x,y)) = 2^{-J}\phi(2^{-J}u(x,y))$. The transformation can be represented as:

$$I(c,u(x,y)) = \{I(c,u(x,y)) * \phi_J(u(x,y)), |I(c,u(x,y)) * \psi_{j,\theta}(u(x,y))|\}$$

- **Avoiding overfitting :** Standard deep learning architectures (without wavelets) used in image classification primarily focus on minimizing the loss function to improve the accuracy of feature extraction. However, they do not emphasize understanding the filtering layers or analyzing the relationship between input data and model output. The use of wavelets helps improve the explainability and interpretability of the model, which in turn helps prevent overfitting.

- **Skip connection :** In this hybrid model, the 2D-WT layer acts as a skip layer to improve the accuracy of image classification. In original VGG-like architectures with many layers, some feature information may get lost due to downsampling by pooling layers. The 2D-WT layer, combined with batch normalization, serves as a skip layer that connects the outputs of low-scale and high-scale convolution layers to recover information lost during downsampling. The 2D-WT layer provides an alternative path for gradients to flow. During backpropagation, the long chain of multiplication with values less than 1 may cause the gradient to become very small as we reach the earlier layers of the deep architecture. This issue, known as the vanishing gradient problem, can hinder stable training and convergence. The 2D-WT layer helps in recovering feature information and mitigates the vanishing gradient problem, ensuring stable training and convergence.

- **Improved accuracy :** CNN-RNN Hybrid framework with 2D-WT layers as the skip layers with VGG-like architectures as the base networks are known to display better classification accuracy when compared to other models for same parameters and same number of layers .

Chapter 6

Relevance to Semester Course Project

(e) Application to Course Project

List and explain, how the ideas that you have explained in Questions (b) and (d) above play a useful role in the specific vertical/ theme that your group has chosen for its semester course project, if you have been able to identify the same.

Solution : The vertical chosen by our group is - Application of wavelets/ time-frequency/ multiresolution/ filter banks in Magnetic Resonance Imaging (MRI) . This hybrid model can be utilized for tumor detection and classification in MRI scan images. For example, the VGG-16 framework is commonly used for brain tumor detection using MRI scans. Implementing our hybrid model(with VGG-16 as base architecture) will provide better accuracy and robustness. It will also offer low memory and computation cost . This model can also be applied for image segmentation tasks and can be used for locating a tumor in an image and labeling each pixel as tumor or background. Thus , CNN-RNN hybrid model with 2D WT layer has many applications in the field of MRI for tasks like image classification , object detection and image segmentation .

Chapter 7

Original Ideas for Economy/ Explainability Improvements

(f) Additional Ideas on Wavelets and Machine Learning Integration

If you have been able to think beyond what the references have described and have come out with some of your own relevant ideas to bring economy/ explainability/ usefulness through the union of wavelets/ filter banks and machine learning/ deep learning/ neural networks, please explain the same, clearly and unambiguously.

Solution : The model's interpretability can be further enhanced by employing a Multilayer Wavelet Attention (MWA) mechanism instead of a simple 2D Wavelet Transform (WT) in our hybrid CNN RNN model. A Frequency Attention Mechanism (FAM) can be incorporated alongside the 2D wavelet transform to improve the network's learning ability and filter out noise. FAM introduces a Global Average Pooling (GAP) layer to compress the global information of the mixed feature map. The global pooling for the i-th feature signal, z_i, is computed as: $z_i = \mathrm{GAP}(LG_i) = \frac{1}{L'} \sum_j LG_i(j)$

Here, LG_i represents the i-th feature signal of the mixed feature map LG. After this, encoding and decoding operations are performed to capture the importance of these signals. This process involves two convolutional layers: the first layer uses the ReLU activation function, while the second layer uses the sigmoid activation function. Finally, matrix multiplication is performed to propagate the weight information into the CNN, enabling the network to focus on the most important features. This method allows the network to neglect noise and learn only from the relevant features and will thus provide better explainability or interpretability . (*Reference : Huan Wang, Zhiliang Liu, Dandan Peng, Ming J. Zuo, Interpretable convolutional neural network with multilayer wavelet for Noise-Robust Machinery fault diagnosis, Mechanical Systems and Signal Processing, Volume 195, 2023, 110314, ISSN 0888-3270, https://doi.org/10.1016/j.ymssp.2023.110314.*)