Integration of Image Decomposition Methods and CNN for Image Classification

Mahammad Ismayilov, Dalia Čalnerytė

Kaunas University of Technology,
Studentų g. 50, Kaunas, Lithuania
mahammad.ismayilov@ktu.edu, *dalia.calneryte@ktu.lt*

Abstract. This study explores integrating Haar wavelet decomposition techniques with convolutional neural networks for image classification on the MNIST dataset. The research demonstrates that without losing significant accuracy by applying the 1-level, 2-level, and 3-level decomposition techniques, the model can reduce the dimensionality and the number of parameters required by the convolutional neural network model. During the training, the 1-level Haar CNN results achieved optimal performance, demonstrating competitive accuracy and computational efficiency compared to the baseline CNN model. This approach highlights the potential of wavelet decomposition techniques to enhance CNN performance with limited computational resources.

Keywords: Convolutional Neural Network (CNN), Haar Wavelet Decomposition, Image Classification, Image Decomposition, MNIST Dataset.

1 Introduction

Image classification is one of the main types of image processing tasks, employed in the stages of diagnosing diseases [1], automatic defect identification [2,3], surveillance systems [4], and other problems. A hybrid deep neural network model was proposed to classify different eczema types [1], a convolutional neural network (CNN) model was implemented to identify flaws in the real-time food packaging control [2], and random forest model was employed in real-time home textile fabric defect inspection system [3]. Image classification problem can be defined as assigning a class for an image based on its content. Artificial intelligence (AI) methods like CNN enabled to automate processes and therefore reduce the human power needed to perform the repetitive tasks. It also enabled transferring experts' knowledge and making the decision processes easier.

One of the main problems in the image processing tasks is the collection of the dataset that is large enough to train the model. Since most of the AI

DOI: https://doi.org/10.15388/LMITT.2025.6

methods are based on the idea that models learn the features relevant for the classes from the raw data, the training procedure usually requires large number of images in the dataset to achieve practically acceptable results. Moreover, complex datasets that demonstrate high diversity in the features representing the objects from the same class lead to the need for training large models and this results in the need for large computational and memory resources. Conventional image processing technologies, such as Fourier transformation or discrete wavelet decomposition, can be applied to highlight representative features in the image and speed up the training process of the classification model. Such integration of conventional image preprocessing techniques and AI methods ensures the sustainable pipeline for image classification that balances computational resources and accuracy.

In this paper, the methodology to integrate conventional image decomposition methods with convolutional neural networks is presented. It investigates the impact of multi-level decomposition levels as a preprocessing step to reduce input dimensionality and the number of parameters. The approach is experimentally demonstrated with the MNIST (Modified National Institute of Standards and Technology) dataset, which is a subset of a larger NIST handwritten digits dataset from 0 to 9 with 60,000 examples of the training set and 10,000 examples of the test set, where each image is 28x28 pixels in size [5].

2 Related Work

Various integrations of wavelet decomposition and CNN were presented in research related to image classification. It is possible to simply use wavelet decomposition as feature extraction method and use the obtained features in machine learning model for classification. For example, Haar wavelet decomposition features were applied to monitor meat quality and classify it as fresh, frozen, and rotten based on its texture [6]. The property of the Haar decomposition to extract features of repetitive structures was employed to develop a real-time home textile fabric defect inspection machine [3]. After several modifications, such as brightness compensation and Gaussian blur filter, the results of Haar wavelet transform were used as input in Random Forest Classifier. The developed system demonstrated high accuracy and short inference time [3].

One of the most popular approaches of wavelet decomposition and CNN combination is wavelet pooling in CNN [7]. Although it demonstrated good performance in texture classification and image annotation tasks compared to the conventional CNN models with much higher number of parameters, the wavelet pooling is a computationally expensive procedure and can be difficult to implement in practical tasks [8].

Another approach is based on image feature extraction using wavelet decomposition and combining the obtained results with output of different CNN levels in various ways. The wavelet transformation was applied to decompose hyperspectral images of three tea types (black, green, and yellow) and used as input for a method based on a lightweight CNN and support vector machine [9].The CNN-enhanced multi-level Haar wavelet features fusion network was proposed to alleviate the issue that ordinary CNN learns mainly spatial characteristics and do not take into account the spectral features [10]. Spectral features extracted using Haar transforms were combined with conventional CNN output at different layers in the fire surveillance system [4]. The proposed model resulted in lower computational cost and reduced number of false alarms compared to the ordinary CNN. The wavelet function was incorporated into CNN blocks as activation function in the model developed to detect breast cancer [11]. The wavelet decomposition was used to replace the early convolution layers in the deep learning model design to classify medical images [12]. It was demonstrated that fixed feature extraction method reduced the number of parameters the neural network needs to learn and therefore speeds up the training process. In addition, the models with wavelet decomposition were more stable compared to the one that used convolutions, thus, it leads to wider application possibilities in the data-limited domains [12]. The issue related to data insufficiency to train the deep CNN model for hyperspectral polarimetric synthetic aperture radar (PolSAR) imagery classification stimulated incorporating Haar wavelets as an effective feature extraction technique for a three-branch deep CNN model in order to improve accuracy and mitigate noise [13].

3 Methodology

The proposed method, shown in Figure 1 below, first applies decomposition to the images. Then, the decomposition results are stacked into the array for multi-channel input in the CNN model, which afterward is trained and evaluated.

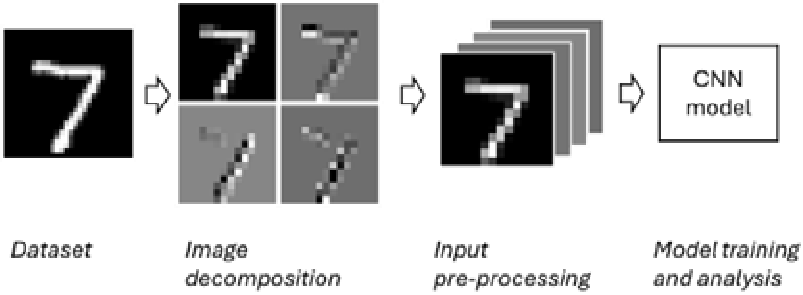

Figure 1. The proposed method

3.1 Image Decomposition

The Haar wavelet transform was initially introduced for signal processing, and it was called the 1D wavelet transform. This wavelet transform is based on decomposing every signal into approximation and detail coefficients, where the Haar wavelet function is crucial for this decomposition as it defines how the signal is split into detail and approximation components [14]. Mathematically, the function can be written as:

$$\psi(t) = \begin{cases} 1, & 0 \leq t < \frac{1}{2} \\ -1, & \frac{1}{2} \leq t < 1 \\ 0, & otherwise \end{cases} \tag{1}$$

Here *ψ(t)* is the Haar wavelet function, often called the "mother wavelet", and *t* is an independent variable, which is typically referred to as position in image processing.

The Haar wavelet transform can be extended to 2D for image decomposition, which can be computed using the 1D Haar wavelet decompositions. The multi-level decomposition technique is a continuation of the non-standard approach for two-level decomposition and more, as the standard method does not inherently define multi-level processing.

From the visualization aspect, Figure 2 illustrates the image decomposition to sub-bands, arranging approximate coefficient (A), horizontal (H), vertical (D), and diagonal (D) components, and so on. The approximation captures the low-frequency part, representing the overall structure of the image, and the horizontal component contains high-frequency information along the horizontal direction, while the vertical component contains high-frequency information along the vertical direction. Additionally, the diagonal component captures high-frequency variations in both directions.

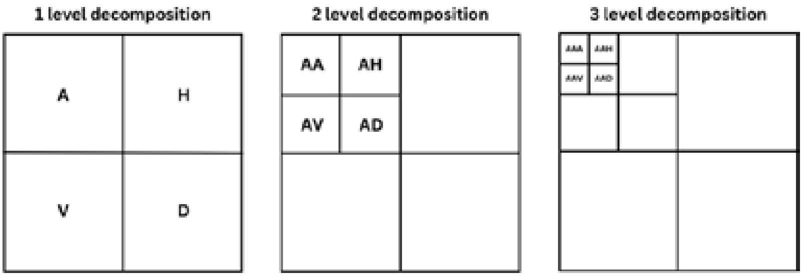

Figure 2. Visualization of the decomposition levels in Haar wavelet transform

Afterward, the 2-level decomposition applies the Haar transform again but only to the approximation sub-band from the first level, which generates a further smoothed approximation (AA) and produces horizontal, vertical, and diagonal details of the approximations (AH, AV, AD). Then, the process repeats the same on the AA sub-band from the 2-level, which creates the third-level approximation (AAA) and horizontal, vertical, and diagonal details of the second-level approximation (AAH, AAV, AAD).

4 Model Evaluation

The convolutional neural network model, integrated with different decomposition levels of the Haar wavelet transform, is defined by the following layers, as shown in Figure 3 below. The main difference between the 1-level, 2-level, and 3-level decomposition is that the CNN model has different number of channels depending on the Haar decomposition level, resulting in different feature resolutions and abstraction levels.

This CNN architecture has been chosen because the MNIST dataset is relatively simple, and even simple models can achieve good accuracy. Moreover, the simplicity of the model is ideal for clearly demonstrating the impact of the Haar wavelet transform.

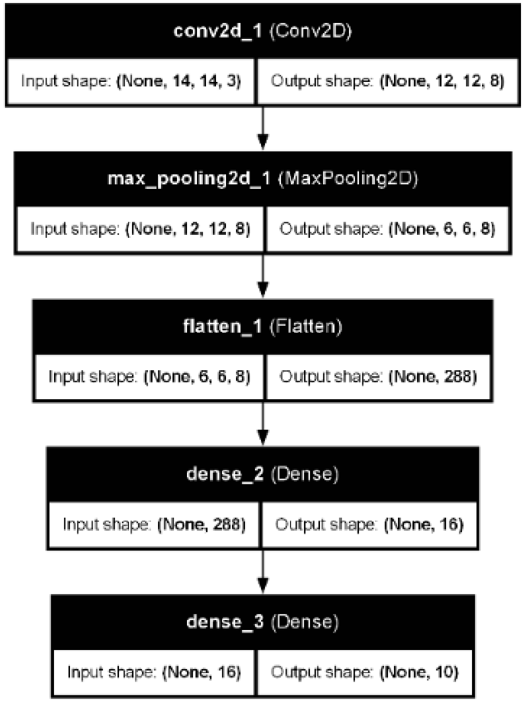

Figure 3. The architecture of the CNN 3-level decomposition model

CNN models with 3 different Haar decomposition levels were compared to a baseline CNN trained on MNIST images containing monochrome 70,000 handwritten digits with 60,000 training and 10,000 test images. Both models share a similar CNN architecture, differing primarily in the input dimensions due to the Haar preprocessing step. After training the models, benchmark results were provided to compare the number of parameters, accuracy, loss, and training time.

Accuracy measures how often the model correctly predicts the outcome:

$$Accuracy = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\hat{y}_i = y_i) \tag{2}$$

where y_i represents the true label of the i^{th} sample in the dataset, $\hat{y}_i$ are the predicted labels produced by the model and $n_{samples}$ is the total number of samples in the dataset. The function $1(\widehat{y_i} = y_i)$ returns *1* if the predicted label matches the actual label.

Categorical Cross-entropy is used as the loss function, which is suitable for multi-class classification where the labels are integer-encoded. For multiclass, the cross-entropy loss formula is shown in equation 3:

$$Loss = -\sum_{i=1}^{C} y_i * log(\widehat{y_i}) \tag{3}$$

where y_i is the true label from the one-hot encoded target vector, $\widehat{y_i}$ is the predicted probability for class *i* and *C* is the number of classes.

5 Results

The models experimented with MNIST dataset images and normalized to the [0, 1] range by dividing pixel values by 255 and trained for 5 epochs due to the dataset‘s simplicity. A Haar wavelet transform was applied at three decomposition levels, initially reducing the image resolutions to 14×14, 7×7, and 4×4 pixels, and the 1-level and 2-level decompositions were resized back to 14x14 pixels for uniform multi-channel CNN input, as shown in Figure 4 below.

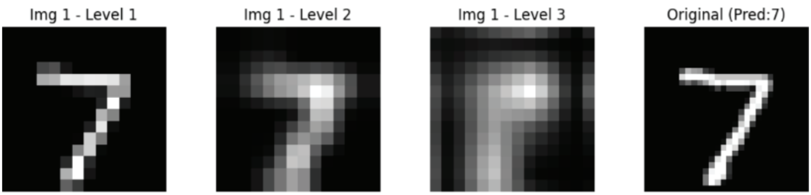

Figure 4. Decomposition level results of CNN model

The CNN models were trained using the TensorFlow library on a laptop equipped with the 12th Gen Intel Core i7-12700H processor, 16 GB of RAM, and an NVIDIA GeForce RTX 3060 GPU with 6 GB video memory.

In general, all models illustrated excellent performance on the MNIST dataset, which is shown in Table 1 below. Nevertheless, comparing the wavelet decomposition levels with the CNN Baseline model, Haar CNN 2-level and Haar CNN 1-level achieved the highest accuracy with 97.50% and 97.45% of results. Moreover, Haar CNN 1-level gained the lowest test loss of 0.0845 and required fewer parameters, indicating slightly better generalization than other models.

Table 1. Benchmark results of Baseline CNN and Haar CNN models

Model	Training Time (s)	Parameters	Test Loss	Test Accuracy
Baseline CNN	9.96	21898	0.0919	0.9725
Haar CNN 1-level	10.87	4874	0.0845	0.9745
Haar CNN 2-level	8.76	4946	0.0893	0.9750
Haar CNN 3-level	8.47	5018	0.1162	0.9651

The unified plots in Figure 5 with training and validation accuracy/loss illustrate that no major overfitting was observed. However, comparing all Haar CNN models, the Baseline CNN model exhibits slightly faster convergence and final accuracy. The loss comparison plot shows a steady decrease in loss for all models, with Haar CNN Level 1 achieving the lowest final loss among the Haar models.

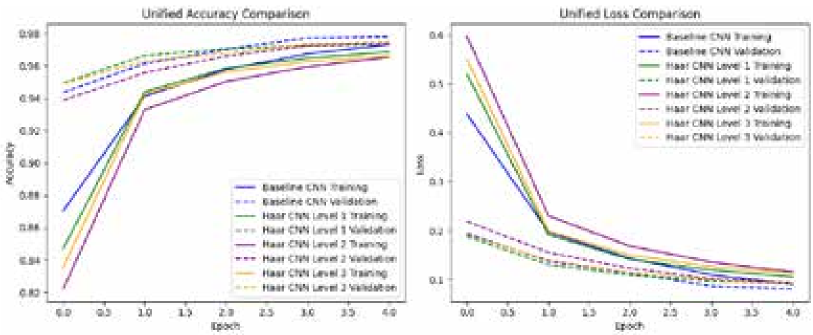

Figure 5. Accuracy and loss comparison results

The Confusion Matrix results in Figure 6 below showed that the Baseline CNN achieves minimal misclassifications for digits like „0“, „1“, „3“, „6“, „7“, and „9“; however, compared with the Haar CNNs, the 1-level decomposition model presents improved accuracy for specific digits such as „5“ and „8“, indicating that the wavelet decomposition technique captures edge and shape variations more effectively. Although Haar CNN 2 and 3 level models offer slight adjustments, they introduce misclassifications in digits like „6“ and „4“, making them less consistent.

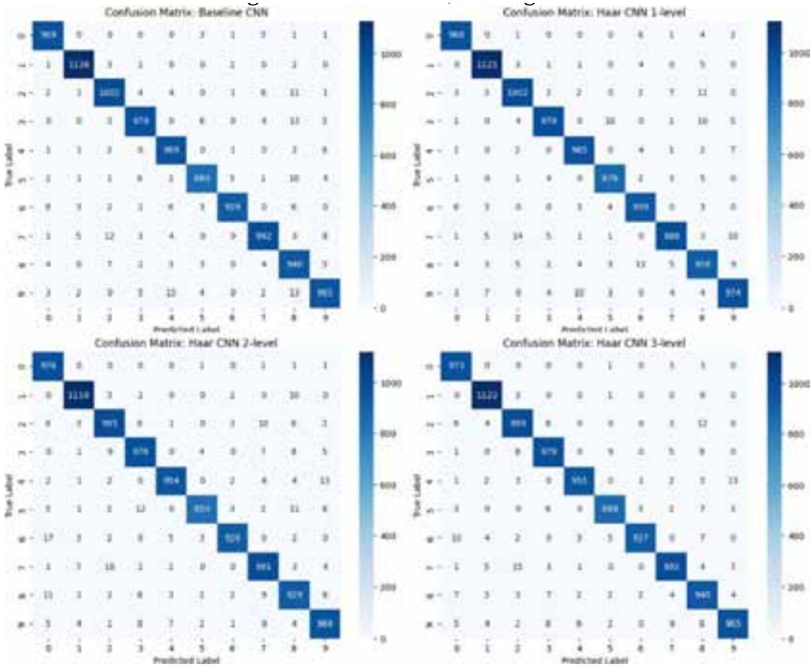

Figure 6. Confusion matrix results

6 Conclusions

To sum up, combining convolutional neural network and Haar wavelet transform has shown that the decomposition method can successfully reduce the dimensionality needed by CNN models. With the lowest test loss and competitive accuracy compared to the baseline CNN model, the Haar CNN with 1-level decomposition shows the best possible balance between accuracy, computational efficiency, and model complexity out of all the assessed models. However, this research was focused on approximation coefficients, omitting the detail coefficients, which contain valuable directional and edge-specific information. Future research will integrate detail coefficients to leverage the Haar decomposition's capabilities fully and will look at methods that methodically lower the number of factors needed at each level of decomposition to improve efficiency and model generalization, which is particularly advantageous in settings with limited resources.

References

[1] Rasheed, A., Umar, A. I., Shirazi, S. H., Khan, Z., Nawaz, S., & Shahzad, M. (2022). Automatic eczema classification in clinical images based on hybrid deep neural network. Computers in Biology and Medicine, 147, 105807. https://doi.org/10.1016/j.compbiomed.2022.105807

[2] Medus, L. D., Saban, M., Francés-Víllora, J. V., Bataller-Mompeán, M., & Rosado-Muñoz, A. (2021). Hyperspectral image classification using CNN: Application to industrial food packaging. Food Control, 125, 107962. https://doi.org/10.1016/j.foodcont.2021.107962

[3] Barman, J., Wu, H.-C., & Kuo, C.-F. J. (2022). Development of a real-time home textile fabric defect inspection machine system for the textile industry. Textile Research Journal, 92(23–24), 004051752211114. https://doi.org/10.1177/00405175221111477

[4] Huang, L., Liu, G., Wang, Y., Yuan, H., & Chen, T. (2022). Fire detection in video surveillances using convolutional neural networks and wavelet transform. Engineering Applications of Artificial Intelligence, 110, 104737. https://doi.org/10.1016/j.engappai.2022.104737

[5] Deng, L. (2012). The MNIST database of handwritten digit images for machine learning research [Dataset]. Papers with Code. https://paperswithcode.com/dataset/mnist

[6] Kiswanto, K., Hadiyanto, H., & Sediyono, E. (2024). Meat Texture Image Classification Using the Haar Wavelet Approach and a Gray-Level Co-Occurrence Matrix. Applied System Innovation, 7(3), 49. https://doi.org/10.3390/asi7030049

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

[8] Zhao, L., & Zhang, Z. (2024). A improved pooling method for convolutional neural networks. Scientific Reports, 14, 1589. https://doi.org/10.1038/s41598-024-51258-6

[9] Cui, Q., Yang, B., Liu, B., Li, Y., & Ning, J. (2022). Tea category identification using wavelet signal reconstruction of hyperspectral imagery and machine learning. Agriculture, 12(8), 1085. https://doi.org/10.3390/agriculture12081085

[10] Guo, W., Xu, G., Liu, B., & Wang, Y. (2022). Hyperspectral image classification using CNN-enhanced multi-level Haar wavelet features fusion network. IEEE Geoscience and Remote Sensing Letters, 19, 6008805. https://doi.org/10.1109/LGRS.2022.3167535

[11] Oyelade, O. N., & Ezugwu, A. E. (2022). A novel wavelet decomposition and transformation convolutional neural network with data augmentation for breast cancer detection using digital mammogram. Scientific Reports, 12, 5913. https://doi.org/10.1038/s41598-022-09983-z

[12] Masquelin, A. H., Cheney, N., Kinsey, C. M., & Bates, J. H. T. (2021). Wavelet decomposition facilitates training on small datasets for medical image classification by deep learning. Histochemistry and Cell Biology, 155(2), 309–317. https://doi.org/10.1007/s00418-020-01961-y

[13] Jamali, A., Mahdianpari, M., Mohammadimanesh, F., Bhattacharya, A., & Homayouni, S. (2022). PolSAR image classification based on deep convolutional neural networks using wavelet transformation. IEEE Geoscience and Remote Sensing Letters, 19, 4510105. https://doi.org/10.1109/LGRS.2022.3185118

[14] Talukder, K. H., & Harada, K. (2007). Haar wavelet based approach for image compression and quality assessment of compressed image. IAENG International Journal of Applied Mathematics, 36(1), 1–9. https://doi.org/10.48550/arXiv.1010.4084