Color Image Classification on Edge by Fusing Visual Feature Extractors and Weightless Neural Networks

Marina S. R. P. Magalhães[1], Leandro S. de Araújo[1]

[1] Institute of Computing – Universidade Federal Fluminense – RJ, Brazil

piragibemarina@gmail.com, leandro@ic.uff.br

Abstract. *Weightless Neural Networks (WNNs) are a promising alternative for the image classification problem in an edge computing environment due to their low computational cost. However, the direct use of this model requires converting the image to grayscale to facilitate the conversion of an image into its binary representation. This work investigates the performance of WNNs for the color image classification problem without adopting the grayscale conversion strategy. Several image preprocessing and representation strategies are evaluated, including grayscale conversion, feature extraction techniques widely adopted in the field of computer vision such as VLAD and Fisher Vector, and Convolutional Neural Networks (CNNs). Several experimental scenarios were conducted with the CIFAR-10 dataset, evaluating the models with 7 feature extractors (5 CNN architectures) in terms of accuracy, precision, recall, F1-score and execution time. The results show that hybrid models with CNNs outperform other feature extractors, achieving 89.57% accuracy with the ConvNeXt model, surpassing the accuracy of pure WNN (44.01%) and resulting in lower computational costs compared to pure CNNs.*

1. Introduction

Image classification is a fundamental task in the field of Computer Vision, playing a critical role in numerous real-world applications, including facial recognition, moving object detection, and autonomous driving systems [Wu et al. 2025]. Feature extraction is a crucial stage in image processing and classification pipelines to represent images in a more compact and informative manner, enabling them to be effectively used as inputs for classification models. Convolutional Neural Networks (CNNs) revolutionized the field of image classification, providing a automatically learning of the hierarchical feature representations directly from raw image data through convolution and pooling operations [Lecun et al. 1998]. However, these networks are computationally intensive and require substantial energy consumption due to gradient-based optimization and backpropagation processes. Consequently, deploying CNNs on resource-constrained platforms, such as Internet of Things (IoT) devices and Edge Computing environments, remains a significant challenge [Xie et al. 2019].

Weightless Neural Networks (WNNs) represent an alternative machine learning paradigm based on lookup tables for learning patterns from binarized data. Models such as WiSARD offer unidirectional training and low computational complexity, making them attractive for hardware-limited applications. Nevertheless, their performance strongly depends on the quality of input preprocessing, since binarization restricts data representation to binary values [Santiago et al. 2020]. Recent WNN-based models, including DWN [Bacellar et al. 2024], ULEEN [Susskind et al. 2023a], and BTHOWeN [Susskind et al. 2023b], employ grayscale image conversion as a preprocessing step. While this approach simplifies the input representation, it inevitably discards valuable color information that may be relevant for classification.

This work investigates the application of Weightless Neural Networks to the classification of color images from the CIFAR-10 dataset and compares their performance with conventional CNN-based models. To address the limitations of binary representations, different preprocessing and feature extraction strategies are explored, including grayscale conversion, VLAD [Jégou et al. 2010] and Fisher Vector [Sánchez et al. 2013]. Furthermore, a hybrid architecture is proposed, combining the convolutional and pooling layers of CNNs with a WiSARD-based classifier replacing the traditional fully connected layer. The main contributions of this work are summarized as follows: (i) a WiSARD-based architecture for color image classification that preserves color information without grayscale conversion; (ii) two feature extraction methods based on image descriptors: Fisher Vector and VLAD; (iii) a hybrid CNN–WiSARD training framework that employs a fine-tuned WiSARD classifier on feature representations extracted from a pre-trained CNN; (iv) a comprehensive assessment of the proposed approaches with respect to classification performance (Accuracy, Precision, Recall, and F1-Score) and execution time.

2. Weightless Neural Network - WiSARD Model

WiSARD (Wilkie, Stonham and Aleksander's Recognition Device) is a Weightless Neural Network model originally proposed in the 1980s [Aleksander et al. 1984]. Unlike conventional neural networks that rely on weighted connections, WiSARD stores input patterns directly in Random Access Memories (RAMs), enabling recognition through comparisons between previously memorized patterns and the input presented to the network.

During training, the input data is transformed into a binary representation and partitioned into tuples of size N. Each tuple defines a memory address that contributes to the learning process. This procedure, known as the addressing map, determines how input bits are assigned to individual RAM nodes [Santiago et al. 2020]. For each class in the dataset, a discriminator is created, consisting of a collection of RAM nodes responsible for storing and recognizing patterns associated with a specific class. Each input tuple is connected to one, and only one, RAM within the discriminator. During the training phase, only the discriminator corresponding to the ground-truth class is updated, following a learning mechanism similar to that adopted by other lookup table (LUT)-based models.

Finally, the inference phase is performed to evaluate the model's ability to recognize previously unseen patterns. Test samples undergo the same preprocessing pipeline used during training, including binarization, and are subsequently presented to all discriminators in the network. Each discriminator verifies whether the tuples extracted from the input have been previously stored in its corresponding RAM nodes. Whenever a match is found, the discriminator receives a score contribution, meaning that active memory locations increment its final score. After all tuples have been evaluated, each discriminator produces a total score, and the input is assigned to the class associated with the highest score [Santiago et al. 2020].

3. Methodology

The methodology adopted in this study was designed to analyze the performance of different approaches for image classification on the CIFAR-10 dataset, including CNNs, WiSARD, and hybrid architectures. To achieve this objective, the experimental procedure was divided into four main stages: (i) selection and organization of the CIFAR-10 dataset for the classification task; (ii) application of preprocessing and feature extraction techniques, including specific transformations and image preparation according to the target approach (CNN-based, WiSARD-based, or hybrid); (iii) implementation of the classification models, comprising standalone CNNs, WiSARD models employing different encoding techniques, and a hybrid framework integrating CNN convolutional layers with a WiSARD classifier through a fine-tuning strategy; and (iv) evaluation of the obtained results using the selected performance metrics.

CIFAR-10 dataset. CIFAR-10 dataset is one of the most widely used benchmarks for image recognition and machine learning tasks [Krizhevsky 2012]. It consists of 60,000 color images with a resolution of 32 × 32 pixels, distributed across 10 object categories: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. The dataset is divided into 50,000 training images and 10,000 test images, with the latter being used to evaluate the performance and generalization capability of the proposed models.

Feature Extractors. For the WiSARD-based approaches, different image preprocessing strategies were investigated to evaluate the impact of distinct feature extraction methods. In the first scenario, the images were directly converted into tensors and provided as input to the WiSARD model. A second approach evaluated the effectiveness of grayscale conversion, where the images were transformed into grayscale, flattened into one-dimensional vectors, and subsequently supplied to the network. For the remaining feature extraction methods, namely VLAD and Fisher Vector, only basic preprocessing operations were applied, consisting of tensor conversion and data flattening, while preserving the original RGB color channels.

- **Fisher Vector**, introduced in [Sánchez et al. 2013], is an image representation technique that encodes local descriptors, such as SIFT features, using a probabilistic model. A *Gaussian Mixture Model* (GMM) is first trained to capture the distribution of the descriptors, and each image is then represented by its deviations from the learned distribution. The resulting feature vector provides a compact and discriminative summary of the image's visual characteristics.
- **VLAD (Vector of Locally Aggregated Descriptors)**, proposed in [Jégou et al. 2010], is an image representation technique that aggregates local descriptors using *K-means* clustering. For each cluster, the residuals between the descriptors and their assigned centroid are accumulated and concatenated into a single vector. After normalization, the resulting representation provides a compact and discriminative summary of the image, making it suitable for image retrieval and classification tasks.

For the CNN-based approaches, all models were initialized with weights pretrained on the ImageNet dataset. Therefore, CIFAR-10 images were resized to 224 × 224 pixels to match the input requirements of the selected architectures, except for InceptionV3, which requires images of size 299 × 299 pixels. Additionally, pixel values were normalized and converted into tensors following the conventions of the PyTorch framework.

Classification Models. The experimental setup was organized into three main evaluation tracks. The first track focused on state-of-the-art CNNs. Five architectures were selected for this purpose: ResNet50, ResNet18, ConvNeXt, InceptionV3, and VGG16. All models were initialized with weights pre-trained on the ImageNet dataset. They were adapted to the target task through a fine-tuning procedure, in which the original classification head (with 1000 output classes) was replaced by a new fully connected layer containing 10 neurons, corresponding to the number of classes in the CIFAR-10 dataset. The second track evaluated the performance of the WiSARD model under different input representations. Initially, simple preprocessing strategies were considered, such as direct flattening of the input images and grayscale conversion. In addition to these baseline configurations, the behavior of the model was also analyzed when combined with the three previously introduced feature extraction techniques: VLAD and Fisher Vector. Finally, a hybrid approach was investigated, integrating the convolutional feature extraction layers of pre-trained CNNs with WiSARD acting as the final classification module.

Evaluation Metrics. The performance of the proposed models was evaluated using standard classification metrics, namely Accuracy, Precision, Recall, and F1-score. In addition, computational performance was assessed by measuring the execution time required for feature extraction, model training, and inference.

4. Experiments and Results

Experimental setup All experiments were conducted on a workstation equipped with an Intel Core i7-10700 processor, 16 GB of RAM, and an NVIDIA GeForce RTX 3060 GPU with 12 GB of dedicated memory. Each experimental configuration was executed ten times, and the reported results correspond to the average values obtained across all runs. All CNN models were trained for 15 epochs. The batch size was set to 128 forResNet18 and ResNet50, and 64 for the remaining architectures. A learning rate of (1×10^{-3}) and a weight decay coefficient of (8×10^{-4}) were adopted for all CNN models. In addition, the Cross-Entropy Loss function was used throughout the experiments. For the WiSARD models, a fixed 8-bit distributed thermometer encoding was employed in all experiments.

CNN results. Table 1 reports accuracy, train and test time for CNN models using GPU. Among the evaluated models, ConvNeXt achieved the best performance, reaching an accuracy of 94.31%. ResNet18 was the fastest architecture, requiring 2896.17 seconds for training. In contrast, VGG16 exhibited the longest training time, requiring 9294.16 seconds. A similar trend was observed for ResNet50 and ConvNeXt, where larger models generally required longer training times. An exception was InceptionV3, which, despite being smaller than ResNet50, required nearly twice the training time.

WiSARD results. Table 2 reports accuracy, train and test time for WiSARD models using only CPU. The results obtained using non-CNN feature extractors indicate that the *Standard* configuration (binarization of all three color channels) achieved the best performance, reaching an accuracy of 44.01%. This result highlights the limitations of handcrafted or fixed-feature approaches in capturing and generalizing complex visual patterns when compared to deep learning-based methods. It was also observed that removing color

Table 1. Results of performance of CNN models.

Model	Accuracy	Precision	Recall	F1	Train (s)	Test (s)
ResNet18	$92.80\% \pm 0.41\%$	92.93%	92.80%	92.80%	2896.17	18.52
ResNet50	$92.97\% \pm 0,88\%$	93.18%	92.97%	92.97%	3885.79	23.28
ConvNeXt	$94.31\% \pm 1.23\%$	94.52%	94.31%	94.29%	7999.56	30.67
InceptionV3	$93.72\% \pm 0.26\%$	93.82%	93.72%	93.73%	7483.29	41.51
VGG16	$91.26\% \pm 0.51\%$	91.49%	91.26%	91.26%	9294.16	40.53

Table 2. Results of performance of WiSARD with different feature extractors.

Model	Accuracy	Precision	Recall	F1	Train (s)	Test (s)
Standard	$44.01\% \pm 0.19\%$	44.14%	44.01%	43.23%	45.97	40.27
Gray Scale	$38.50\% \pm 0.24\%$	38.74%	38.50%	38.00%	13.86	12.51
VLAD	$22.80\% \pm 0.22\%$	25.05%	22.80%	21.59%	152.71	52.53
Fisher Vector	$22.73\% \pm 0.20\%$	22.04%	22.73%	21.90%	45.97	120.59
ResNet18	$79.35\% \pm 0.29\%$	79.53%	79.35%	79.25%	69.49	18.10
ResNet50	$83.75\% \pm 0.09\%$	83.90%	83.75%	83.73%	144.71	47.26
ConvNeXt	$89.57\% \pm 0.18\%$	89.69%	89.57%	89.56%	149.97	37.95
InceptionV3	$78.74\% \pm 0.10\%$	79.10%	78.74%	78.77%	198.32	62.36
VGG16	$53.34\% \pm 0.20\%$	72.68%	53.34%	53.47%	629.17	320.32

information by converting the images to grayscale negatively affected classification performance, reducing the accuracy to 38.50%. This finding suggests that color constitutes an important discriminative feature for the classification task and contributes significantly to the model's ability to distinguish between classes. When comparing the best result obtained with the standard feature extractors with CNN feature extractors, the classification accuracy increased from 44.01% to 89.57%. Although this performance does not surpass that of the standalone CNN models, which achieved a maximum accuracy of 94.31%, the results demonstrate that WiSARD can effectively benefit from more sophisticated feature extraction techniques and that its performance is highly dependent on the quality of the input representation. Another notable advantage of the hybrid approach was the substantial reduction in execution time (only CPU). Similar to the feature extraction experiments, the combination of CNN-based feature extraction with WiSARD yielded favorable results in terms of computational efficiency, where the shortest execution time was approximately 1.5 minutes, considerably lower than those observed for the standalone CNN models, without using a GPU.

5. Conclusion

This work investigates WNN for the image classification task on the CIFAR-10 dataset, and several input preprocessing techniques were explored, including basic transformations such as grayscale conversion, classical feature extraction methods such as VLAD and Fisher Vector. In addition, state-of-the-art CNNs were evaluated both as standalone classifiers and as feature extractors combined with WiSARD through a fine-tuning strategy. Although the hybrid CNN–WiSARD approach did not match the performance of the standalone CNNs, it substantially outperformed the isolated WiSARD models. Once again, ConvNeXt produced the best results among the hybrid architectures, achieving an accuracy of 89.57%. From a computational efficiency perspective, all WiSARD-based approaches exhibited lower execution times, energy consumption, and carbon footprint when compared to standalone CNNs. As future work, we will investigate the application of explainability techniques, such as Grad-CAM, to CNN-based models in order to gain deeper insights into the visual patterns learned during training and their influence on classification decisions.

Acknowledgements

This research was funded by CNPq (grant 307770/2025-7).

References

Aleksander, I., Thomas, W., and Bowden, P. (1984). Wisard·a radical step forward in image recognition. *Sensor Review*, 4(3):120–124.

Bacellar, A. T. L., Susskind, Z., Breternitz Jr., M., John, E., John, L. K., Lima, P. M. V., and França, F. M. G. (2024). Differentiable weightless neural networks. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org.

Jégou, H., Douze, M., Schmid, C., and Pérez, P. (2010). Aggregating local descriptors into a compact image representation. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3304–3311.

Krizhevsky, A. (2012). Learning multiple layers of features from tiny images. *University of Toronto*.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

Santiago, L., Verona, L., Rangel, F., Firmino, F., Menasché, D. S., Caarls, W., Breternitz Jr, M., Kundu, S., Lima, P. M., and França, F. M. (2020). Weightless neural networks as memory segmented bloom filters. *Neurocomputing*, 416:292–304.

Susskind, Z., Arora, A., Miranda, I. D. S., Bacellar, A. T. L., Villon, L. A. Q., Katopodis, R. F., de Araújo, L. S., Dutra, D. L. C., Lima, P. M. V., França, F. M. G., Breternitz Jr., M., and John, L. K. (2023a). Uleen: A novel architecture for ultra-low-energy edge neural networks. *ACM Trans. Archit. Code Optim.*, 20(4).

Susskind, Z., Arora, A., Miranda, I. D. S., Villon, L. A. Q., Katopodis, R. F., de Araújo, L. S., Dutra, D. L. C., Lima, P. M. V., França, F. M. G., Breternitz, M., and John, L. K. (2023b). Weightless neural networks for efficient edge inference. In *Proceedings of the International Conference on Parallel Architectures and Compilation Techniques*, PACT '22, page 279–290, New York, NY, USA. Association for Computing Machinery.

Sánchez, J., Mensink, T., and Verbeek, J. (2013). Image classification with the fisher vector: Theory and practice. *International Journal of Computer Vision*, 105.

Wu, P., He, X., Dai, W., Zhou, J., Shang, Y., Fan, Y., and Hu, T. (2025). A review on research and application of ai-based image analysis in the field of computer vision. *IEEE Access*, 13:76684–76702.

Xie, R., Jia, X., Wang, L., and Wu, K. (2019). Energy efficiency enhancement for cnn-based deep mobile sensing. *IEEE Wireless Communications*, 26(3):161–167.