Title

A Novel Learnable Orthogonal Wavelet Unit Neural Network with Perfection Reconstruction Constraint Relaxation for Image Classification

Authors

Le, An D

Jin, Shiwei

Bae, You Suk

et al.

Publication Date

2023-12-07

DOI

10.1109/vcip59821.2023.10402772

A Novel Learnable Orthogonal Wavelet Unit Neural Network with Perfection Reconstruction Constraint Relaxation for Image Classification

An D Le[1], Shiwei Jin[1], You Suk Bae[2], Truong Nguyen[1]
[1]Jacobs School of Engineering, University of California San Diego, California, USA
[2]Department of Computer Engineering, Technology University of Korea, Gyeonggi-do, Korea
d0le@ucsd.edu, sjin@eng.ucsd.edu, ysbae@tukorea.ac.kr, tqn001@eng.ucsd.edu

***Abstract*—CNNs utilize lowpass frequency features, evidenced by max pooling operations that preserve dominant features. Addressing this, we used a one-layer FCN using both coarse and detailed components via DWT. We then developed UwU, a learnable wavelet-based unit that integrates the PR constraint relaxation, allowing feature map component fine-tuning. Distinctively, UwU's coefficients are trainable, unlike prior studies. This is the first work that utilizes PR constraint relaxation to enhance CNNs. Our innovative techniques serve to enhance stride-convolution, pooling, and downsampling units in CNNs. We tested these improved units using the ResNet family architectures against traditional frequency and wavelet-based units. Performance metrics from CIFAR10, ImageNet1K, and the DTD show promising results. Particularly, while CIFAR10 results are on par with other methods, there's a marked performance improvement on ImageNet1K and DTD. Further, integrating UwU into the ResNet18-based encoder of the CFLOW-AD system yields competitive anomaly detection results, particularly for hazelnut images in the MVTecAD dataset.**

***Index Terms*—Computer vision, Image classification, Wavelet transforms, Anomaly detection**

I. Introduction

Max pooling, prevalent in VGG [1], DenseNet [2], Mobilenets [3], [4], and ResNets [5], prioritizes locally dominant features, often omitting finer details and causing aliasing artifacts [6]. While frequency-based methods [7], [8] focus on low-frequency features, wavelet techniques like WaveCNet [9] mainly use lowpass components. However, Wavelet-Attention CNNs [9] and CWNN [10] incorporate both coarse and detailed components. Notably, high-frequency details are crucial for high-resolution images.

As shown in Fig. 1, the CIFAR10 [11] sample has most information concentrated in the low-frequency region or the approximation component. In contrast, all other samples from ImageNet1K [12], MVTecAD [13], [14], and DTD [15] have information in both low (approximation) and high (detail) frequency regions. Especially in the "cracked" DTD sample shown in the fourth column of Fig. 1, the decomposed $\mathbf{X}_{ll}$ component, showing lowpass or approximation information, contains only few features of the "cracked" texture, which is dominantly shown in detail or highpass components such as $\mathbf{X}_{hl}$, $\mathbf{X}_{lh}$, and $\mathbf{X}_{hh}$. For enhanced CNN classification, we employ all DWT multi-resolution components and integrate a one-layer FCN to find optimal feature maps. We introduce UwU, a novel learnable orthogonal wavelet unit, notable for its PR constraint relaxation and trainable coefficients—unlike conventional static ones. These innovations have been embedded into ResNet18, enhancing downsampling, pooling, and stride-convolution processes in CNNs. The networks were validated with the baseline ResNet18 along with its wavelet-based variants like WaveCNet [9], Wavelet-Attention CNN [16], CWNN [10] and spectrum-based variants such as SpectralPooling [7] and DiffStride [8] on CIFAR10 [11], ImageNet1K [12], and DTD [15] datasets. We further applied the proposed units in the encoder of the CFLOW-AD [17] pipeline to the anomaly detection task for the hazelnut category in the MVTecAD [13], [14] dataset. In summary:

- We propose a novel pooling method consisting of a Discrete Wavelet Transform (DWT) and a one-layer FCN

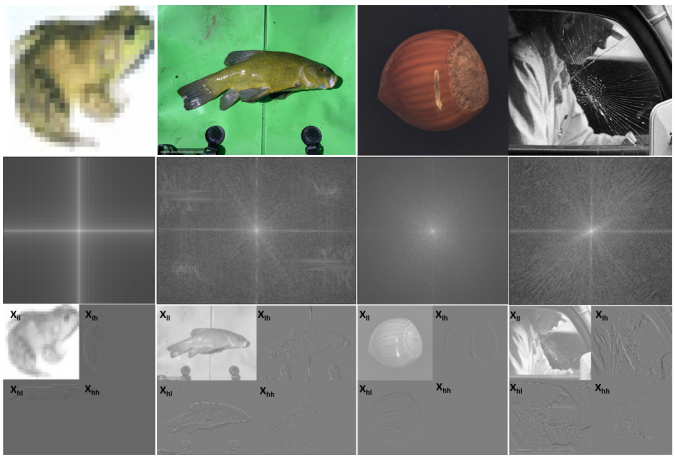

Fig. 1. From left to right, wavelet (Haar) and frequency representations of the samples from CIFAR10 (first column), ImageNet1K (second column), MVTecAD (third column), and DTD (fourth column). The original images (top row) are shown with its frequency representation (middle row) and wavelet representation (bottom row). $\mathbf{X}_{ll}$, $\mathbf{X}_{lh}$, $\mathbf{X}_{hl}$, and $\mathbf{X}_{hh}$ show the coarse approximation and details wavelet representations.

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Innovative Human Resource Development for Local Intellectualization support program (ITP-2023-2020-0-01741) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation).

for processing all subband components of DWT.

- Inspired by the perfect reconstruction constraint for the filter bank, which consists of distortion and aliasing, we introduce a new distortion loss function to design a new set of filter in DWT, while maintaining aliasing cancellation.
- We apply the proposed methods on a wide range of image classification datasets: CIFAR10, ImageNet1K, and DTD and achieve excellent performance.
- The proposed units are also implemented in the CFLOW-AD anomaly detection model and tested on the MVTecAD dataset.

II. Related Works

Max pooling in CNNs down-samples feature maps, retaining maximal values to preserve prominent features [18], [19]. However, without filtering, it can lead to aliasing artifacts, violating the sampling theorem [6]. This results in frequency overlaps in signal processing and Moiré patterns in imaging. Furthermore, max pooling can compromise object structures in deep networks [9].

Spectral pooling [7] uses the DFT to pool in the frequency domain, emphasizing lower frequencies and omitting higher ones. However, its non-differentiability with respect to strides [8] requires predefined hyper-parameters for each down-sampling layer. While DiffStride [8] optimizes stride numbers and cropping sizes via back-propagation, detail information in the feature map still gets discarded.

DWT-based methods use wavelet bases in CNNs. Through DWT or FWT [20], [21], these methods enable CNNs to operate in the wavelet domain, facilitating PR and avoiding aliasing. Such techniques are evident in recent image classification works [9], [10], [16]. In [9], [22], only approximation components of wavelets are used. In [22], a second-level wavelet decomposition is used, but only its second-level sub-bands are used for reconstruction. In [9], the first-level approximation is employed for feature maps. In [16] attention maps are derived from vertical and horizontal details, and in [10], dual-tree complex wavelet transformation (DT-CWT) is used, and its decomposed components are averaged as the down-sampling output. Distinctively, our UwU offers trainable coefficients and integrates PR constraint relaxation, diverging from prior fixed-coefficient approaches.

III. Proposed Method

We introduce a novel learnable universal wavelet unit (UwU) with an orthogonal filter bank. Paired with a one-layer FCN, it optimizes feature maps and can serve as down-sampling and pooling units. Additionally, replacing a stride-convolution layer with a non-stride version, followed by UwU, retains detail components in the convolution output.

A. Universal Wavelet Unit with PR Constraints

DWT analysis and synthesis depend on filter banks meeting the PR constraint. This ensures that signals decomposed by analysis filters are perfectly reconstructed by synthesis

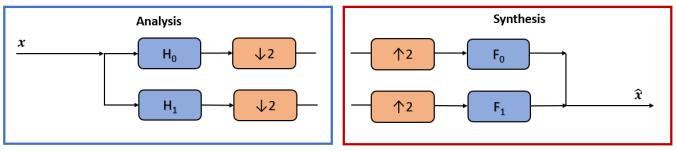

Fig. 2. Two-channel filter bank architecture.

ones. The PR consists of both Aliasing Cancellation and No Distortion/Half-Band conditions [20], [21]. In recent works [9], [16], predefined orthogonal wavelets such Haar, Daubechies, and Symlet [20], [21], [23] are used. In this work, we propose a wavelet unit with trainable coefficients which would still abide to Aliasing Cancellation and Half-band conditions. The filter bank structure is shown in Fig. 2.

In Fig. 2, the analysis and synthesis parts of the filter bank are shown in the blue and red rectangle boxes, respectively. $\mathbf{H}_0$ and $\mathbf{H}_1$ are, correspondingly, low-pass and high-pass filters for the analysis part of the filter bank; whereas $\mathbf{F}_0$ and $\mathbf{F}_1$ are, respectively, low-pass and high-pass filters for the synthesis part of the filter bank. To satisfy the Alias Cancellation condition of an orthogonal filter bank, with $\mathbf{h}_0 = [h(0), h(1), ..., h(N-1)]$ as coefficients of $\mathbf{H}_0$ with N taps, the coefficients of the other filters in the filter bank can be found with alternating flip, order flip and alternating signs relations, which can be expressed as follows [21]:

$$\begin{cases} \text{Order Flip:} & \mathbf{f}_0(n) = \mathbf{h}_0(N-1-n) \\ \text{Alternating Flip:} & \mathbf{h}_1(n) = (-1)^n \mathbf{h}_0(N-1-n) \\ \text{Alternating Sign:} & \mathbf{f}_1(n) = -(-1)^n \mathbf{h}_0(n). \end{cases} \quad (1)$$

In (1), $\mathbf{f}_0$, $\mathbf{h}_1$, and $\mathbf{f}_1$ are filter coefficients of $\mathbf{F}_0$, $\mathbf{H}_1$, and $\mathbf{F}_1$, respectively. From the relations presented in (1), the filter bank satisfies the anti-aliasing condition of the PR constraint. Moreover, with Aliasing Cancellation condition, only filter coefficients of $\mathbf{H}_0$ need to be found, which reduces the number of parameters needed for the analysis part used in a classification model. In addition, to fullfill the PR constraint, the filter coefficients need to satisfy the Half-band condition as follows [21]:

$$\mathbf{P}(z) + \mathbf{P}(-z) = 2, \quad (2)$$

where $\mathbf{P}(z) = \mathbf{H}_0(z)\mathbf{H}_0(z^{-1})$. From (2), the condition for the filter coefficients can be expressed as follows:

$$\begin{cases} \sum_{n=0}^{N-1} h(n)^2 = 1 \\ \sum_{0<n,n+2l<N-1}^{N-1} h(n)h(n+2l) = 0, \quad \text{for } 0 < l \leq N/2. \end{cases} \quad (3)$$

From (3), the loss function for the PR constraint based on Half-band condition can be mathematically expressed as follows:

$$\mathbf{L_{PR}} = |1 - \sum_{n=0}^{N-1} h(n)^2|^2 + \sum_{l=1}^{N/2} \Big(\sum_{0<n,n+2l<N-1}^{N-1} h(n)h(n+2l) \Big)^2 \quad (4)$$

From (1) and (4), PR constraint is implemented to train the filter bank analysis. The relaxation of the PR constraint can

be done by multiplying $\mathbf{L_{PR}}$ with a factor α. In our image classification study using Cross-Entropy $\mathbf{L_{CE}}$, an α of 0.01 gave the best results for all image experiments. A higher α strengthens the No Distortion constraint, but relaxing it allows more coefficient fine-tuning. The total loss function $\mathbf{L}$ is expressed as follows:

$$\mathbf{L} = \mathbf{L_{CE}} + \alpha \mathbf{L_{PR}} \tag{5}$$

B. 2D Implementation

From the filter coefficients satisfying the PR constraint stated in the previous section, the high-pass and low-pass filter matrices $\mathbf{H}$ and $\mathbf{L}$ are computed to find the approximation $\mathbf{X}_{ll}$, and other detail components $\mathbf{X}_{lh}$, $\mathbf{X}_{hl}$, and $\mathbf{X}_{hh}$. $\mathbf{L}$ can be computed as follows:

$$\mathbf{L} = \mathbf{D}\widehat{\mathbf{H}}, \tag{6}$$

where $\mathbf{D}$ is the downsampling matrix and $\widehat{\mathbf{H}}$ is a Toeplitz matrix with filter coefficients of $\mathbf{H}_0(z)$. $\mathbf{H}$ has a similar form as $\mathbf{L}$ with filter coefficients of $\mathbf{H}_0(z^{-1})$. Using $\mathbf{H}$ and $\mathbf{L}$, $\mathbf{X}_{ll}$, $\mathbf{X}_{lh}$, $\mathbf{X}_{hl}$, and $\mathbf{X}_{hh}$ are computed as follows:

$$\begin{aligned} \mathbf{X}_{ll} &= \mathbf{L}\mathbf{X}\mathbf{L}^T, \quad \mathbf{X}_{lh} = \mathbf{H}\mathbf{X}\mathbf{L}^T, \\ \mathbf{X}_{hl} &= \mathbf{L}\mathbf{X}\mathbf{H}^T, \quad \mathbf{X}_{hh} = \mathbf{H}\mathbf{X}\mathbf{H}^T, \end{aligned} \tag{7}$$

C. Implementation in CNN architectures

We implemented the proposed units in ResNet family architectures. For downsampling and pooling layer, the proposed UwU is followed by a onelayer-FCN. In addition, the 2 stride-convolution is replaced with a non-stride convolution block followed by the proposed UwU. The implementation is shown in Fig. 3.

IV. Experiments

We applied our proposed method to the ResNet family architectures, training and testing on the CIFAR10 [11], ImageNet1K [12], and DTD [15] datasets. For wavelets with 2, 4, 6, and 8 coefficients, we trained our units on ResNet18 using the proposed UwU. the trainable parameters in UwU were initialized using Haar coefficients for 2 taps, DB2 and Symlet2 for 4 taps, DB3 and Symlet3 for 6 taps, and DB4 and Symlet4 for 8 taps. In addition to the original wavelets, networks were also trained with WaveCNet. The best results for CIFAR10 and ImageNet1K were compared to the performance of baseline ResNet18, WaveCNet ResNet18, CWNN-ResNet18 in [9], SpectralPool-ResNet18 and DiffStride-ResNet18 in [8], and WaveletAttention CNN ResNet18 (WA-CNN-ResNet18) in [16]. We expanded our research to include ResNet34 and ResNet50 on DTD and ImageNet1K. Furthermore, the proposed units were incorporated into the CFLOW-AD [17] encoder for anomaly detection in the hazelnut category of MVTecAD [13], [14].

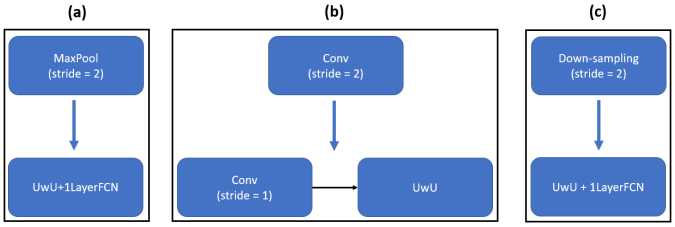

Fig. 3. Implementation of the proposed unit in CNN architecture, replacing max-pool (a), stride-convolution (b), and downsampling (c) functions.

TABLE I
ACCURACY OF UWU ON RESNET18 FOR CIFAR10.

Wavelet	**Accuracy**(%)		
None (Baseline)	92.44		
	UwU ResNet18	one-layer FCN-Wavelet ResNet18	WaveCNet ResNet18
2-Tap Haar	94.97	**95.01**	94.76
4-Tap DB2	94.66	**95.21**	94.93
4-Tap Symlet2	95.13	**95.17**	94.79
6-Tap DB3	93.37	**94.56**	**94.56**
6-Tap Symlet3	94.42	**94.57**	94.35
8-Tap DB4	**94.21**	94.07	93.81
8-Tap Symlet4	94.79	94.79	**94.84**

TABLE II
ACCURACY OF UWU AND OTHER APPROACHES WITH RESNET18 ARCHITECTURE ON CIFAR10.

Models	**Accuracy**(%)
baseline-ResNet18 [5]	92.44
SpectralPool-ResNet18 [7]	92.50 (+0.06)
DiffStride-ResNet18 [8]	92.90 (+0.46)
WA-CNN-ResNet18 [16]	92.57 (+0.13)
WaveCNet-ResNet18 Sym4 [9]	94.84 (+2.40)
one-layer FCN-Wavelet DB2 ResNet18(ours)	**95.21 (+2.77)**
UwU-4Tap-initSymlet2 ResNet18(1LayerFCN)(ours)	95.13 (+2.69)

A. Image Classification with ResNet18

We applied ResNet18 with the proposed units (Fig. 3) on CIFAR10, ImageNet1K, and DTD which is a dataset consisting of images full of textures and details. Beside the baseline and WaveCNet-ResNet18 models, the proposed units are compared given reported performances of Wavelet-Attention CNN ResNet18 [16] (WA-CNN-ResNet18), SpectralPool-ResNet18 and DiffStride-ResNet18 from [8], and Convolutional Wavelet Neural Network ResNet18 (CWNN-ResNet18) from [9] on CIFAR10 and ImageNet1K datasets.

1) On CIFAR10: In general, the proposed methods show a comparable improvement to the baseline ResNet18 compared to WaveCNet or Original Wavelet (with one-layer FCN). The performance is shown in Table I. The best results from Table I are compared with the reported performances of Wavelet-Attention CNN ResNet18 [16] (WA-CNN-ResNet18), SpectralPool-ResNet18 and DiffStride-ResNet18 from [8]. The best performances on CIFAR10 are shown in Table II. The best UwU on CIFAR10 has a higher accuracy than the baseline and a competitive performance compared with the original DB2.

2) On ImageNet1K: In general, the UwU shows the best improvement for the 2-Tap case. The results are shown in Table III. The best results from Table III are then compared with the reported performances of SpectralPool-ResNet18 and DiffStride-ResNet18 from [8], the best performance from WaveCNet-ResNet18-cohen2.2 [9], CWNN-ResNet18 [9], shown in Table IV. We observe that one-layer FCN-Wavelet Sym2 ResNet18 and UwU-2Tap-initHaar ResNet18

TABLE III
ACCURACY OF UWU ON RESNET18 FOR IMAGENET1K.

Wavelet	Accuracy(%)		
None (Baseline)	69.76		
	UwU ResNet18	one-layer FCN-Wavelet ResNet18	WaveCNet ResNet18
2-Tap Haar	**71.96**	71.78	71.47
4-Tap DB2	71.78	**71.80**	71.48
4-Tap Symlet2	71.87	**71.90**	—
6-Tap DB3	70.51	**70.57**	70.35
6-Tap Symlet3	**71.38**	70.82	—
8-Tap DB4	70.51	**70.57**	70.35
8-Tap Symlet4	**71.76**	71.45	—

TABLE IV
ACCURACY OF UWU AND OTHER APPROACHES WITH RESNET18 ARCHITECTURE ON IMAGENET1K.

Models	Accuracy(%)
baseline-ResNet18 [5]	69.76
SpectralPool-ResNet18 [7]	69.93 (+0.17)
DiffStride-ResNet18 [8]	69.72 (-0.04)
WaveCNet-ResNet18-cohen2.2 [9]	71.62 (+1.86)
CWNN-ResNet18 [10]	70.06 (+0.3)
one-layer FCN-Wavelet Symlet2 ResNet18(ours)	71.90 (+2.14)
UwU-2Tap-initHaar ResNet18(1layerFCN)(ours)	**71.96 (+2.2)**

TABLE V
ACCURACY OF UWU ON RESNET18 FOR DTD

Wavelet	Accuracy(%)		
None (Baseline)	27.45		
	UwU ResNet18	one-layer FCN-Wavelet ResNet18	WaveCNet ResNet18
2-Tap Haar	**35.43**	34.47	25.32
4-Tap DB2	**35.21**	28.78	23.62
4-Tap Symlet2	**33.19**	30.27	29.47
6-Tap DB3	33.72	**34.10**	24.36
6-Tap Symlet3	**32.71**	27.61	26.76
8-Tap DB4	**31.70**	30.43	26.91
8-Tap Symlet4	**32.39**	29.04	31.22

(one-layer FCN) achieve the best results, with UwU-2Tap-initHaar ResNet18 (one-layer FCN) having a slightly better performance. On a GTX 1080 TI setup, the average inference times for the baseline, WaveCNet ResNet18 2-Tap Haar, and UwU ResNet18 2-Tap Haar are 0.335, 0.334, and 0.333 seconds respectively, indicating similar computing times.

3) On DTD: From Table V, the proposed UwU shows a clear improvement to the baseline ResNet18 compared to WaveCNet and original wavelet with a one-layer FCN, except for the case of DB3. In this experiment, the advantage of the proposed UwU is clearly shown when the networks are applied on DTD which contains many texture and detail information.

B. ResNet34 and ResNet50 on DTD and ImageNet1K

For DTD, we tested the proposed unit on ResNet34 and ResNet50, comparing with WaveCNet for 2 and 4 taps, shown in Table VI. On ImageNet1K, ResNet34 with our unit using Symlet2 for initialization outperformed WaveCNet with Cohen4.4 and baseline models, with the accuracy of 74.65%, 74.61% and 73.30%, respectively. UwU consistently tops performance on DTD and ImageNet1K, with a more significant accuracy gain on DTD.

C. As the Encoder of CFLOW-AD on MVTecAD (Hazelnut)

The ResNet18 with the proposed unit is used as the encoder in the CFLOW-AD [17] pipeline for the anomaly detection task on hazelnut images of MVTecAD [13], [14] and shows a comparable performance to the baseline ResNet18 along with the WaveCNet ResNet18 encoders. The models were evaluated with localization and detection AUROCs, shown in Table VII. Defect detection result examples are visualized in Fig. 4.

TABLE VI
ACCURACY OF UWU ON RESNET34 AND RESNET50 FOR DTD

Wavelet	ResNet34		ResNet50	
None (Baseline)	16.91%		13.30%	
	Ours	WaveCNet	Ours	WaveCNet
2-Tap Haar	**25.37**%	24.26%	**16.81**%	13.35%
4-Tap DB2	**27.34**%	23.78%	**21.12**%	19.73%
4-Tap Sym2	**27.23**%	21.70%	**18.94**%	13.72%

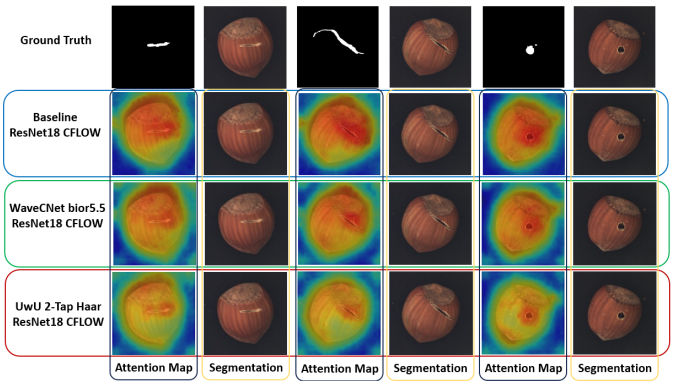

Fig. 4. Anomaly detection for hazelnut objects in the MVTecAD. The first row shows pairs of the ground-truth mask (left) and the corresponding input (right). From second to fourth rows, pairs of the attention map (left) and the segmentation (right) overlaid on the corresponding input. The attention map using the proposed approach (last row) shows excellent localization property in the anomaly area.

TABLE VII
LOCALIZATION AND DETECTION AUROCS OF CFLOW-AD PIPELINE WITH THE UWU, WAVECNET, AND BASELINE RESNET18 ENCODERS FOR HAZELNUT CATEGORY IN MVTECAD

Models	Localization AUROC	Detection AUROC
Baseline ResNet18 CFLOW-AD	96.51%	94.21%
UwU 2-Tap (iniHaar) ResNet18 CFLOW-AD	96.83%	93.32%
WaveCNet ResNet18 CFLOW-AD (bior5.5)	97.15%	85.43%

V. CONCLUSION

We present UwU, a learnable orthogonal wavelet unit paired with a one-layer FCN for optimal CNN feature maps. A defining feature of UwU is its learnable coefficients, setting it distinctly apart from earlier works that relied on static, untrainable coefficients. Its design meets the Aliasing Cancellation condition, halving trainable parameters, and with our new loss function, adheres to the PR constraint flexibly. We have integrated these techniques into the ResNet family architectures, yielding competitive results on CIFAR10, ImageNet1K, and DTD. Applied to the ResNet18 encoder in the CFLOW-AD pipeline, the proposed methods show promise in hazelnut anomaly detection. In future work, we will apply the technique to object detection and image segmentation.

REFERENCES

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.

[2] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2017.

[3] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *CoRR*, vol. abs/1704.04861, 2017.

[4] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[6] R. Zhang, "Making convolutional networks shift-invariant again," in *ICML*, 2019.

[7] O. Rippel, J. Snoek, and R. P. Adams, "Spectral representations for convolutional neural networks," in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'15, (Cambridge, MA, USA), p. 2449–2457, MIT Press, 2015.

[8] R. Riad, O. Teboul, D. Grangier, and N. Zeghidour, "Learning strides in convolutional neural networks," *ICLR*, 2022.

[9] Q. Li, L. Shen, S. Guo, and Z. Lai, "Wavelet integrated cnns for noise-robust image classification," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7243–7252, 2020.

[10] Y. Duan, F. Liu, L. Jiao, P. Zhao, and L. Zhang, "Sar image segmentation based on convolutional-wavelet neural network and markov random field," *Pattern Recognition*, vol. 64, pp. 255–267, 2017.

[11] A. Krizhevsky, "Learning multiple layers of features from tiny images," 2009.

[12] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

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

[14] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger, and C. Steger, "The mvtec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection," *International Journal of Computer Vision*, vol. 129, no. 4, p. 1038–1059, 2021.

[15] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, , and A. Vedaldi, "Describing textures in the wild," in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[16] X. Zhao, P. Huang, and X. Shu, "Wavelet-attention cnn for image classification," *Multimedia Syst.*, vol. 28, p. 915–924, jun 2022.

[17] D. Gudovskiy, S. Ishizaka, and K. Kozuka, "CFLOW-AD: Real-time unsupervised anomaly detection with localization via conditional normalizing flows," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 98–107, January 2022.

[18] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, (Madison, WI, USA), p. 111–118, Omnipress, 2010.

[19] J. Weng, N. Ahuja, and T. Huang, "Cresceptron: a self-organizing neural network which grows adaptively," in *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*, vol. 1, pp. 576–581 vol.1, 1992.

[20] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[21] G. Strang and T. Q. Nguyen, "Wavelets and filter banks," 1996.

[22] T. Williams and R. Li, "Wavelet pooling for convolutional neural networks," in *International Conference on Learning Representations*, 2018.

[23] I. Daubechies, *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, 1992.