BIORTHOGONAL TUNABLE WAVELET UNIT WITH LIFTING SCHEME IN CONVOLUTIONAL NEURAL NETWORK

An Le[1], Hung Nguyen[1], Sungbal Seo[2], You-Suk Bae[2], Truong Nguyen[1]
[1]Electrical and Computer Engineering Department, University of California San Diego, La Jolla, CA 92093, USA
{d0le,hun004,tqn001}@ucsd.edu
[2]Department of Computer Engineering, Tech University of Korea, Siheung 15073, Korea
{sungbal,ysbae}@tukorea.ac.kr

***Abstract*—This work introduces a novel biorthogonal tunable wavelet unit constructed using a lifting scheme that relaxes both the orthogonality and equal filter length constraints, providing greater flexibility in filter design. The proposed unit enhances convolution, pooling, and downsampling operations, leading to improved image classification and anomaly detection in convolutional neural networks (CNN). When integrated into an 18-layer residual neural network (ResNet-18), the approach improved classification accuracy on CIFAR-10 by 2.12% and on the Describable Textures Dataset (DTD) by 9.73%, demonstrating its effectiveness in capturing fine-grained details. Similar improvements were observed in ResNet-34. For anomaly detection in the hazelnut category of the MVTec Anomaly Detection dataset, the proposed method achieved competitive and well-balanced performance in both segmentation and detection tasks, outperforming existing approaches in terms of accuracy and robustness.**

***Index Terms*—Anomaly detection, Computer vision, Discrete wavelet transforms, Feature extraction, Image processing, Image recognition, Machine learning, Supervised learning, Wavelet coefficients, Wavelet transform.**

I. Introduction

Max pooling, a key component in CNN architectures such as ResNets [1], emphasizes dominant features but discards fine details, leading to aliasing artifacts [2]. While frequency-based methods [3], [4] focus on low-frequency components, wavelet-based models like WaveCNet [5] predominantly use low-pass filters. However, models such as Wavelet-Attention CNNs [6] incorporates both coarse and fine-grained details, which is crucial for high-resolution image processing.

As depicted in Fig. 1, the CIFAR-10 dataset [7] predominantly consists of low-frequency information, whereas MVTecAD [8], [9] and DTD [10] exhibit features distributed across both low- and high-frequency domains. In the "cracked" DTD sample shown in the second column of Fig. 1, the low-pass component $\mathbf{X}_{ll}$ retains only minimal texture details, whereas the high-pass components $\mathbf{X}_{hl}$, $\mathbf{X}_{lh}$, and $\mathbf{X}_{hh}$ effectively capture its distinctive features. This underscores the importance of maintaining both high- and low-frequency information within CNN architectures. Previous studies [11], [12] used wavelet decomposition and perfect reconstruction to retain full image information, improving performance. While tunable wavelet filters [11], [12] enhanced CNNs, especially for high-frequency images, they rely on orthogonal wavelets

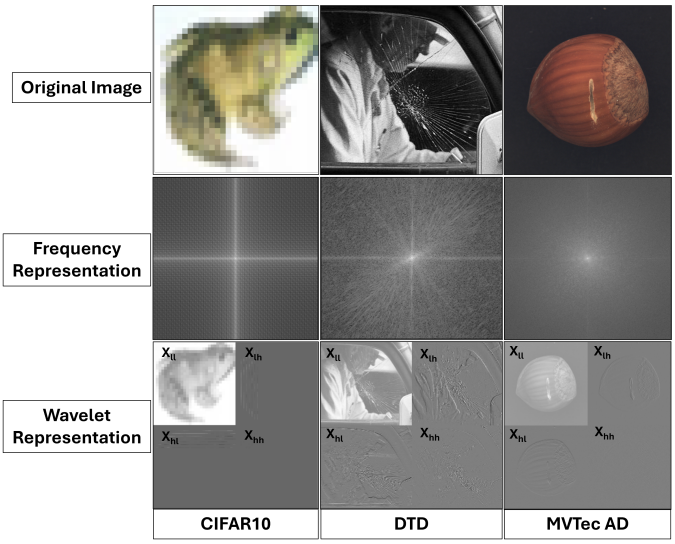

Fig. 1. From left to right, wavelet (Haar) and frequency representations of the samples from CIFAR10 (first column), DTD (second column), and MVTecAD (third column). The original images (top row) are shown with its frequency representation (middle row) and wavelet representation (bottom row). $\mathbf{X}_{ll}$,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

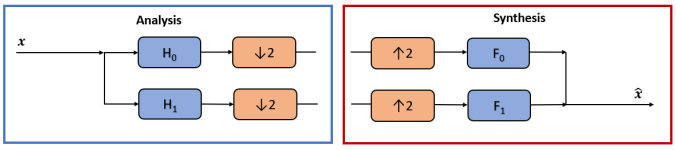

Fig. 2. Two-channel filter bank architecture.

This work was supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT)(IITP-2024-2020-0-01741).

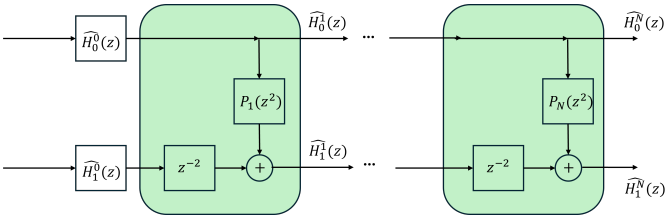

Fig. 3. The analysis of a biorthogonal filter bank constructed with lifting scheme. $P_k(z)$ is the lifting step function for k in the range from 1 to N.

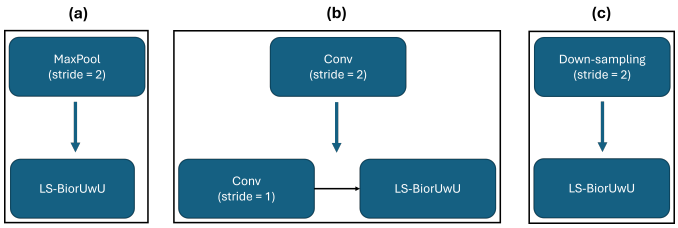

Fig. 5. Implementation of the proposed unit in CNN architecture, replacing max-pool (a), stride-convolution (b), and downsampling (c) functions.

and require equal filter lengths. To overcome these limitations, we propose a biorthogonal tunable wavelet unit based on a lifting scheme (LS-BiorUwU), which relaxes these constraints, allowing greater design flexibility. Integrated into ResNet architectures, the proposed unit improves classification on CIFAR-10 and DTD and serves as a feature extractor in the CFLOW-AD anomaly detection pipeline [13], tested on the hazelnut category of MVTecAD [8], [9]. Our approach enhances CNN performance in both image classification and anomaly detection. In summary:

- We propose LS-BiorUwU, a novel biorthogonal tunable wavelet unit based on the lifting scheme that relaxes the orthogonality and equal filter length constraints imposed by existing orthogonal wavelet units.
- We integrate the proposed unit into ResNet architectures trained on CIFAR-10 and DTD, achieving strong classification performance, especially on DTD.
- We incorporate the unit into the CFLOW-AD anomaly detection model and evaluate it on the MVTecAD dataset.

II. Related Works

Max pooling in CNNs downsamples feature maps by selecting maximum values, preserving key features [14], [15]. However, without filtering, it introduces aliasing artifacts, leading to frequency overlap and Moiré patterns [2], and can distort object structures in deeper networks [5]. Wavelet-based methods apply discrete wavelet transforms (DWT, FWT) to process features in the wavelet domain [16], [17], improving image classification [5], [6], [18]. Existing approaches rely on predefined wavelet functions, primarily using approximation components [5], with limited reconstruction from higher-order decompositions. Wavelet-Attention CNN [6] incorporates attention maps from detail components. Trainable wavelet filters, such as those in [11], relaxed perfect reconstruction constraints, while the units in [12] enforced an orthogonal lattice structure. However, these methods rely on orthogonal wavelets and require equal filter lengths.

To address these limitations, we propose a biorthogonal tunable wavelet unit based on a lifting scheme, which removes these constraints and enables greater flexibility in wavelet design.

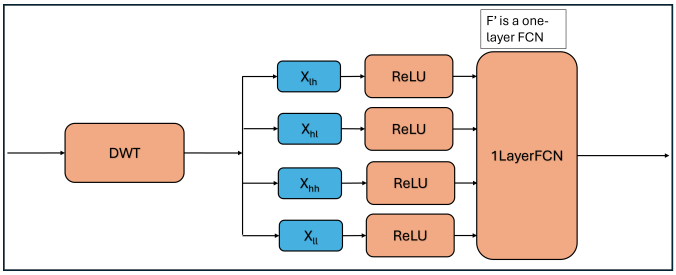

Fig. 4. Diagram of low-pass and high-pass component implementation. The signal goes from left to right. The results from DWT go to ReLU functions to become the inputs of a one-layer FCN. Because an FCN can take inputs of arbitrary sizes, the one-layer FCN can read the decomposed components and finetune the trainable coefficients to optimally combine the decomposed components. The fine-tuned one-layer FCN combines the inputs to find the optimal feature map.

III. Proposed Method

A. Lifting Scheme for Tunable Biorthogonal Wavelet Filters

Using a lifting scheme, the tunable biorthogonal wavelet unit (LS-BiorUwU) relaxes the constraint of orthogonality to biorthogonality and allows unequal filter lengths in the filter bank of the wavelet unit.

In the filter bank structure demonstrated in Fig. 2, the analysis, shown in the blue rectangle box, and synthesis, shown in the red rectangle box, parts of the filter bank have the function of decomposing and reconstructing signals, respectively. H_0 and H_1 are, correspondingly, low-pass and high-pass filters for the analysis part of the filter bank; whereas F_0 and F_1 are, respectively, low-pass and high-pass filters for the synthesis part of the filter bank. With L taps, H_0 and H_1 have $\mathbf{h}_0 = [h_0(0), h_0(1), ..., h_0(L-1)]$ and $\mathbf{h}_1 = [h_1(0), h_1(1), ..., h_1(L-1)]$ as their coefficients, respectively. In orthogonal wavelets, $\mathbf{h}_0$ and $\mathbf{h}_1$ are required to have the same length and are related. This requirement can be relaxed if the filter bank is constructed with a lifting scheme from an orthogonal wavelet. Hence, H_0 and H_1 can be represented as a matrix multiplication as follows:

$$\begin{bmatrix} H_0(z) \\ H_1(z) \end{bmatrix} = \begin{bmatrix} \widehat{H_0^N}(z) \\ \widehat{H_1^N}(z) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ P_N(z^2) & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & z^{-2} \end{bmatrix} \cdots \begin{bmatrix} 1 & 0 \\ P_1(z^2) & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & z^{-2} \end{bmatrix} \begin{bmatrix} \widehat{H_0^0}(z) \\ \widehat{H_1^0}(z) \end{bmatrix}, \quad (1)$$

in which $\widehat{H_0^N}(z)$ and $\widehat{H_1^N}(z)$ are the final filter pairs constructed after N lifting steps from the $\widehat{H_0^0}(z)$ and $\widehat{H_1^0}(z)$ pair of

TABLE I
THE RECURSIVE IMPLEMENTATION OF THE LIFTING SCHEME FOR 1 AND 2 LIFTING STEPS IS TESTED BY COMPARING ITS $\mathbf{H}_0$ AND $\mathbf{H}_1$ COEFFICIENTS WITH THOSE OF BIOR1.3 AND BIOR1.5.

1 Lifting Step	**Recursive Lifting Scheme**	$\mathbf{h}_0$	0.7071, 0.7071
		$\mathbf{h}_1$	-0.0880, -0.0880, 0.7071, -0.7071, 0.0880, 0.0880
	Bior1.3 Ground-Truth	$\mathbf{h}_0$	0.7071, 0.7071
		$\mathbf{h}_1$	-0.0880, -0.0880, 0.7071, -0.7071, 0.0880, 0.0880
2 Lifting Steps	**Recursive Lifting Scheme**	$\mathbf{h}_0$	0.7071, 0.7071
		$\mathbf{h}_1$	0.0166, 0.0166, -0.1215, -0.1215, 0.7071, -0.7071, 0.1215, 0.1215, -0.0166, -0.0166
	Bior1.5 Ground-Truth	$\mathbf{h}_0$	0.7071, 0.7071
		$\mathbf{h}_1$	0.0166, 0.01666, -0.12156, -0.12156, 0.70716, -0.70716, 0.12156, 0.12156, -0.01666, -0.0166

an orthogonal wavelet. In addition, for k in the range from 1 to N, $P_k(z)$ is the lifting step function, which can be represented as follows:

$$P_k(z) = -a_k + a_k z^{-2k} \text{ for } k \text{ in } [1, N], \tag{2}$$

where a_k is the tunable parameter in the biorthogonal wavelet unit. The lifting scheme can be visualized in Fig. 3. In addition, the proposed tunable biorthgonal wavelet with lifting scheme can be implemented as the following recursive algorithm:

$$\begin{bmatrix} \widehat{H_0^k}(z) \\ \widehat{H_1^k}(z) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ P_k(z^2) & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & z^{-2} \end{bmatrix} \begin{bmatrix} \widehat{H_0^{k-1}}(z) \\ \widehat{H_1^{k-1}}(z) \end{bmatrix}$$
$$= \begin{bmatrix} \widehat{H_0^{k-1}}(z) \\ -a_k \widehat{H_0^{k-1}}(z) + z^{-2} \widehat{H_1^{k-1}}(z) + a_k z^{-4k} \widehat{H_0^{k-1}}(z) \end{bmatrix}, \tag{3}$$

for k in the range from 1 to N. In this work, Haar or Bior1.1 is used for $\widehat{H_0^0}(z)$ and $\widehat{H_1^0}(z)$ initialization.

B. 2D Implementation

From the coefficients $\mathbf{h}_0$ and $\mathbf{h}_1$, the high-pass and low-pass filter matrices $\mathbf{H}$ and $\mathbf{L}$ are derived to compute the approximation component $\mathbf{X}_{ll}$ and the detail components $\mathbf{X}_{lh}$, $\mathbf{X}_{hl}$, and $\mathbf{X}_{hh}$ of the input $\mathbf{X}$. The matrix $\mathbf{L}$ is computed as follows:

$$\mathbf{L} = \mathbf{D}\widehat{\mathbf{H}}, \tag{4}$$

where $\mathbf{D}$ denotes the downsampling matrix, and $\widehat{\mathbf{H}}$ is a Toeplitz matrix formed using the filter coefficients of $\mathbf{H}_0(z)$. The matrix $\mathbf{H}$ follows the same structure as $\mathbf{L}$ but is derived from the filter coefficients of $\mathbf{H}_0(z^{-1})$. Using $\mathbf{H}$ and $\mathbf{L}$, the components $\mathbf{X}_{ll}$, $\mathbf{X}_{lh}$, $\mathbf{X}_{hl}$, and $\mathbf{X}_{hh}$ are computed as follows:

$$\begin{aligned} \mathbf{X}_{ll} &= \mathbf{L}\mathbf{X}\mathbf{L}^T, \quad \mathbf{X}_{lh} = \mathbf{H}\mathbf{X}\mathbf{L}^T, \\ \mathbf{X}_{hl} &= \mathbf{L}\mathbf{X}\mathbf{H}^T, \quad \mathbf{X}_{hh} = \mathbf{H}\mathbf{X}\mathbf{H}^T. \end{aligned} \tag{5}$$

C. Implementation in CNN architectures

The proposed units were incorporated into ResNet family architectures. In downsampling and pooling layers, the UwU is followed by a one-layer fully connected network (FCN), as depicted in Fig. 4. Furthermore, the stride-2 convolution is replaced with a non-stride convolution block, followed by the proposed LS-BiorUwU, as shown in Fig. 5.

IV. EXPERIMENTS AND RESULTS

This section integrates the proposed LS-BiorUwU unit into CNN architectures. First, the lifting scheme implementation for initialization is examined, followed by an analysis of the frequency response of the tuned coefficients. Next, the unit is applied to ResNet18 (tested on CIFAR-10 and DTD with 1, 2, and 3 lifting steps) and ResNet34 (with 1, 2, and 3 lifting steps). Finally, its effectiveness is evaluated within the CFLOW-AD pipeline for anomaly detection on the Hazelnut class in the MVTecAD dataset.

A. Lifting Scheme for Tunable Biorthogonal Wavelet Filters Coefficient Analysis

This section examines the recursive implementation of the lifting scheme for LS-BiorUwU, starting from the Haar/Bior1.1 wavelet with 1 and 2 lifting steps. Higher lifting steps are not considered due to diminishing returns in classification improvement. The coefficients used are $a_1 = \frac{880}{7071}$ for 1 step and $a_1 = \frac{405}{2357}$, $a_2 = \frac{-166}{7071}$ for 2 steps. Table I shows that the resulting $\mathbf{h}_0$ and $\mathbf{h}_1$ closely match Bior1.3 and Bior1.5, confirming the proposed scheme's ability to construct biorthogonal wavelets with correct lifting steps and initial orthogonal wavelet filters. The LS-BiorUwU unit is then integrated into a ResNet18 pooling layer trained on CIFAR-10, and its 2-step version (LS-BiorUwU-2Step) is analyzed. As shown in Fig. 6, the tuned H_1 filter retains high-pass characteristics.

B. Image Classification: CIFAR10 and DTD

This section evaluates the performance of LS-BiorUwU units with 1, 2, and 3 lifting steps in ResNet architectures on CIFAR-10 and DTD. CIFAR-10 [7] consists of 60,000 low-resolution (32×32) color images across 10 classes, with 50,000 for training and 10,000 for testing. In contrast, DTD [10] (Describable Textures Dataset) includes 5,640 high-resolution images across 47 categories, emphasizing rich textures and high-frequency details. The LS-BiorUwU units were primarily tested on ResNet18 and compared to the tunable orthogonal lattice wavelet unit (OrthLatt-UwU) with 2-tap filters initialized with Haar/Bior1.1 [12]. For LS-BiorUwU-1Step and LS-BiorUwU-2Step, Bior1.3 and Bior1.5 filter coefficients were used to determine a_1 and a_2 for initialization. For LS-BiorUwU-3Step, the a_1 and a_2 values from LS-BiorUwU-2Step were used, while a_3 was set to a value near zero for initialization.

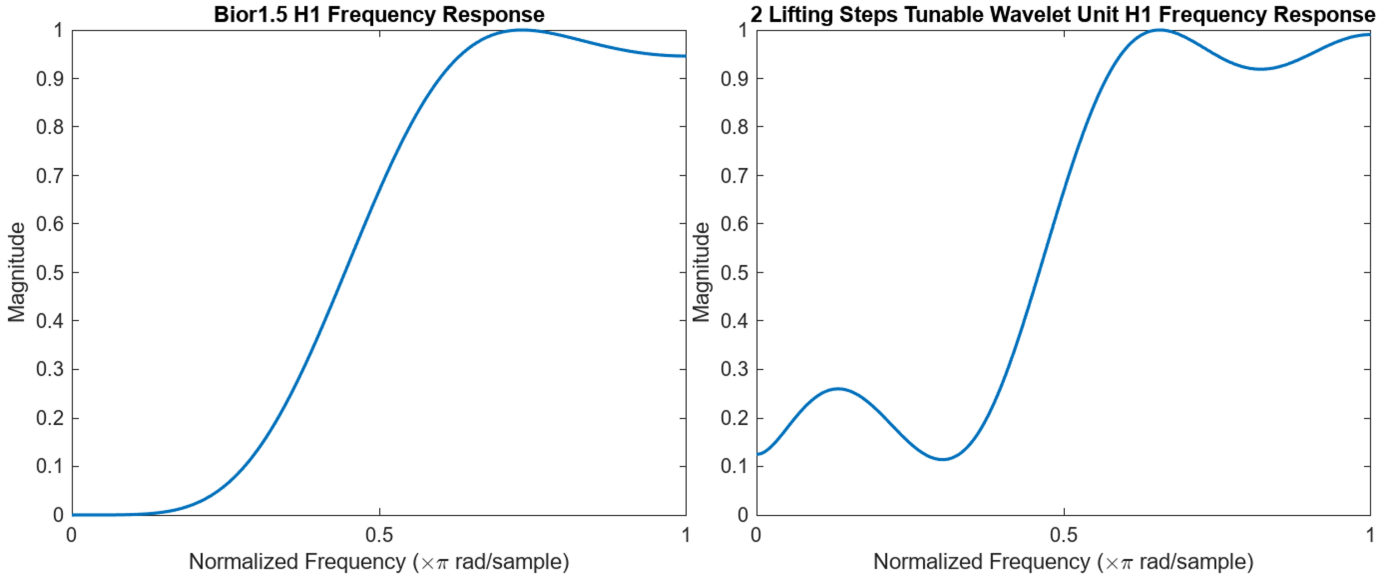

Fig. 6. Frequency Response Analysis of the Bior1.5 H_1 (left) and the tuned coefficients of H_1 of LS-BiorUwU after 2 lifting steps (right).

TABLE II
ACCURACY OF LS-BIORUWU WITH 1, 2 AND 3 LIFTING STEPS, ALONG WITH ORTHLATT-UWU-2TAP EVALUATED ON RESNET18 FOR DTD AND CIFAR10.

DTD			
Baseline: 33.99%			
OrthLatt-UwU-2Tap	LS-BiorUwU-1Step	LS-BiorUwU-2Step	LS-BiorUwU-3Step
40.37%	42.71%	43.67%	43.72%
CIFAR10			
Baseline: 92.44%			
OrthLatt-UwU-2Tap	LS-BiorUwU-1Step	LS-BiorUwU-2Step	LS-BiorUwU-3Step
94.97%	94.08%	94.56%	94.39%

TABLE III
ACCURACY OF LS-BIORUWU WITH 1, 2 AND 3 LIFTING STEPS, ALONG WITH ORTHLATT-UWU-2TAP EVALUATED ON RESNET34 FOR DTD AND CIFAR10.

DTD			
Baseline: 24.47%			
OrthLatt-UwU-2Tap	LS-BiorUwU-1Step	LS-BiorUwU-2Step	LS-BiorUwU-3Step
41.49%	41.76%	41.76%	42.45%
CIFAR10			
Baseline: 94.33%			
OrthLatt-UwU-2Tap	LS-BiorUwU-1Step	LS-BiorUwU-2Step	LS-BiorUwU-3Step
95.44%	94.45%	94.56%	94.36%

1) ResNet18: LS-BiorUwU with 1, 2, and 3 lifting steps was implemented in ResNet18 and tested on CIFAR-10 and DTD, representing datasets with low-resolution, low-pass features and high-resolution, high-pass details, respectively. As shown in Table II, LS-BiorUwU-ResNet18 outperformed the baseline ResNet18 across all lifting step initializations. While LS-BiorUwU did not surpass OrthLatt-UwU-2Tap on CIFAR-10, it showed a clear advantage on DTD. Nevertheless, it shows that the proposed LS-BiorUwU can still perform well on low-resolution data.

For DTD, which contains high-resolution images with detailed features, increasing the lifting steps and the order of the high-pass filter led to better results. Since OrthLatt-UwU-2Tap (Bior1.1/Haar) and LS-BiorUwU (1, 2, and 3 lifting steps) share the same 2-Tap low-pass filter length, adding more lifting steps improved DTD accuracy by increasing the order of the high-pass filter. However, LS-BiorUwU-3Step provided only limited improvement over LS-BiorUwU-2Step, likely due to suboptimal initialization in LS-BiorUwU-3Step. These results suggest that LS-BiorUwU maintains competitive performance on low-resolution datasets, while demonstrating superior accuracy on high-resolution images that contain rich details and high-frequency features—benefiting from the use of higher-order high-pass filters.

2) Extended Study with ResNet34: In this section, LS-BiorUwU with 1, 2, and 3 lifting steps was implemented in ResNet34, tested on CIFAR-10 and DTD, and compared against the baseline ResNet34 and OrthLatt-UwU-2Tap. As shown in Table III, while LS-BiorUwU achieves comparable performance on CIFAR-10, it consistently outperforms both baseline ResNet34 and OrthLatt-UwU-2Tap ResNet34 at every lifting step on DTD. Additionally, increasing the number of lifting steps leads to better performance, aligning with the trends observed in the ResNet18 experiments. This suggests that the performance gains of LS-BiorUwU extend to deeper neural network architectures as well.

TABLE IV
SEGMENTATION AND DETECTION AUROCS OF CFLOW-AD PIPELINE WITH THE BASELINE RESNET18, ORTHLATT-UWU-2TAP, AND LS-BIORUWU-2STEP ENCODERS FOR HAZELNUT CATEGORY IN MVTECAD.

Models	**SegAUROC**	**DetAUROC**
Baseline [12]	96.45%	92.46%
OrthLatt-UwU-2Tap [12]	97.20%	89.21%
LS-BiorUwU-2Step	97.21%	92.11%

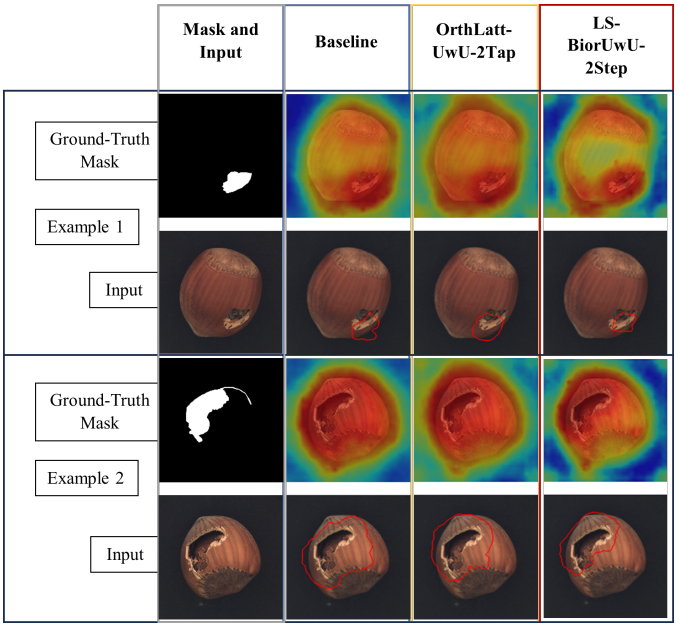

Fig. 7. Anomaly detection on hazelnut objects from the MVTec AD dataset with two examples. From left to right: the first column presents the mask and input image, while the second to fourth columns display heatmaps and defect segmentation results from the baseline, OrthLatt-UwU-2Tap, and LS-BiorUwU-2Step, respectively.

C. Anomaly Detection: MVTecAD

In this experiment, LS-BiorUwU-2Step ResNet18, trained on the DTD dataset, was used as an encoder in the CFLOW-AD pipeline [13] for anomaly detection on hazelnut images from the MVTecAD dataset [8], [9]. The LS-BiorUwU-2Step ResNet18 encoder was compared against baseline ResNet18 and OrthLatt-UwU-2Tap, as shown in Table IV. While OrthLatt-UwU-2Tap achieved a segmentation AUROC of 97.20%, LS-BiorUwU-2Step ResNet18 slightly outperformed it with 97.21%, while also achieving a significantly higher detection AUROC of 92.11%. Fig. 7 presents heatmaps from detection models, demonstrating that LS-BiorUwU-2Step ResNet18 provides better segmentation and localization performance compared to the baseline and OrthLatt-UwU-2Tap encoders.

V. Conclusion

This study introduces a biorthogonal tunable wavelet unit based on a lifting scheme, which removes constraints on orthogonality and filter length, enabling greater design flexibility and enhancing image classification and anomaly detection in CNNs. Results show that increasing lifting steps and the order of the high-pass filter improves performance on high-frequency feature images while maintaining competitive results on low-resolution datasets. Additionally, the proposed approach balances segmentation and detection, leading to improved anomaly detection performance. However, the study has some limitations, as the current approach applies lifting steps in a single direction, from the low-pass filter to the high-pass filter, to increase the high-pass filter order. Future work can explore a dual lifting scheme to adjust both low-pass and high-pass filter orders, offering greater design flexibility. Additionally, while the current tunable biorthogonal wavelet units demonstrate good performance, their stopband attenuation can be further improved. This can be addressed by incorporating a stopband-attenuation constraint for better frequency selectivity in future research.

References

[1] 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.

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

[3] 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.

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

[5] 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.

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

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

[8] 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.

[9] 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.

[10] 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.

[11] A. D. Le, S. Jin, Y. S. Bae, and T. Nguyen, "A novel learnable orthogonal wavelet unit neural network with perfection reconstruction constraint relaxation for image classification," in *2023 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pp. 1–5, 2023.

[12] A. D. Le, S. Jin, Y.-S. Bae, and T. Q. Nguyen, "A lattice-structure-based trainable orthogonal wavelet unit for image classification," *IEEE Access*, vol. 12, pp. 88715–88727, 2024.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] A. D. Le, S. H. Yassin, W. R. Freeman, A. Heinke, D.-U. G. Bartsch, S. Borooah, S. Jin, T. Nguyen, and C. An, "Wavelet deep learning network for objective retinal functional estimation from multimodal retinal imaging," in *Ophthalmic Medical Image Analysis* (A. Bhavna, H. Chen, H. Fang, H. Fu, and C. S. Lee, eds.), (Cham), pp. 32–41, Springer Nature Switzerland, 2025.