

Stop-band Energy Constraint for Orthogonal Tunable Wavelet Units in Convolutional Neural Networks for Computer Vision problems

An D. Le, Hung Nguyen, Sungbal Seo, You-Suk Bae and Truong Q. Nguyen

Content

- Motivation
- Previous Work
- Proposed Method
- Experiments and Results
 - Image Classification
 - With ResNet18 on CIFAR10 and DTD
 - Extended study with ResNet34 on DTD and CIFAR10
 - As an encoder in an Anomaly Detection pipeline
- Conclusion

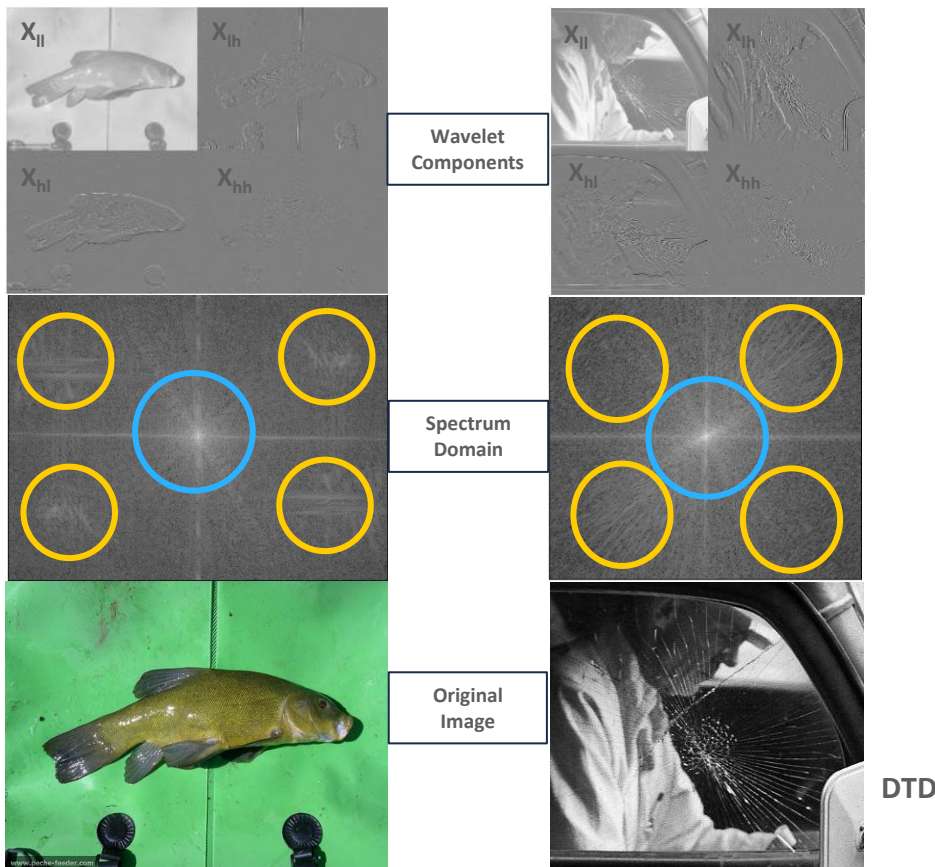
Motivation

Motivation

- High resolution images contain important information in both low-pass and high-pass components.
- The detailed features are more prominent in high-pass components.
- Both low-pass and high-pass components of an input image contain important information which should be used.
- The tunable wavelet filters improved CNN processing, particularly for images with significant high-frequency content.
- The flexibility of the tunable filters resulted in an unclear distinction between high-pass and low-pass behaviors.

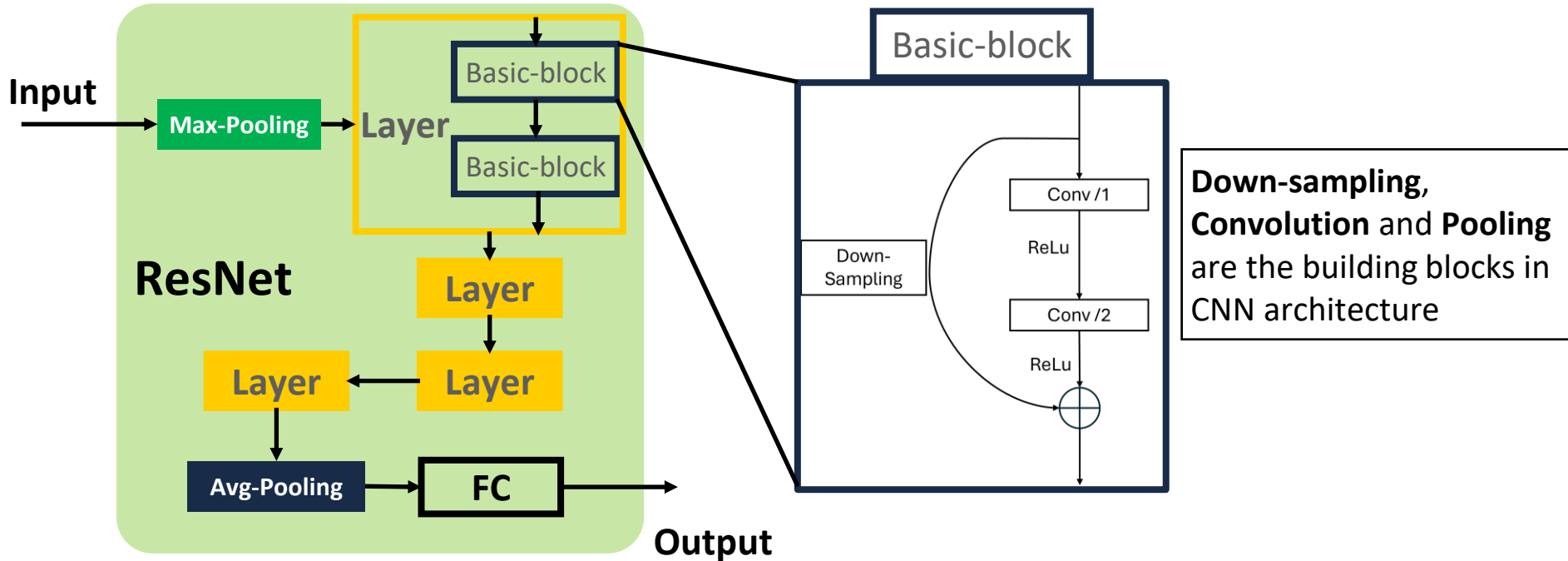
→ Stopband Energy Constraint (SBE) is proposed to enforce distinct low-pass and high-pass characteristics in tunable wavelet filters.

ImageNet1K

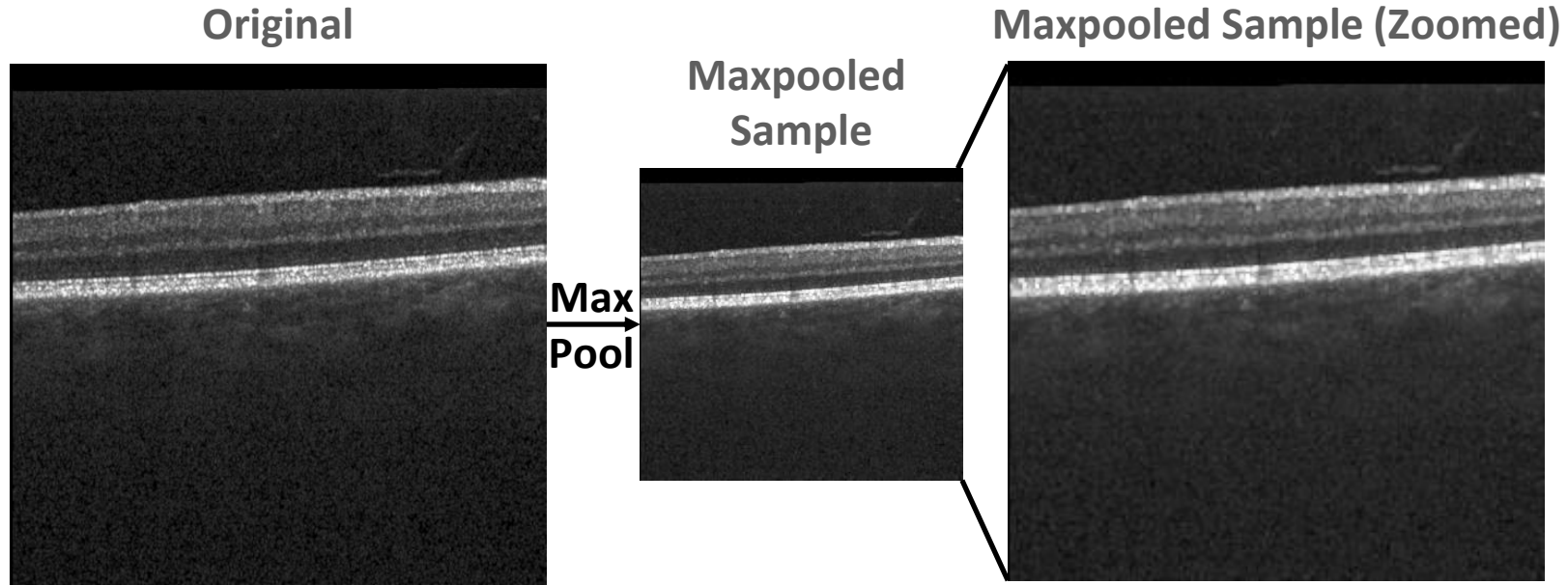


Previous Work

Convolution Neural Network: ResNet Architecture

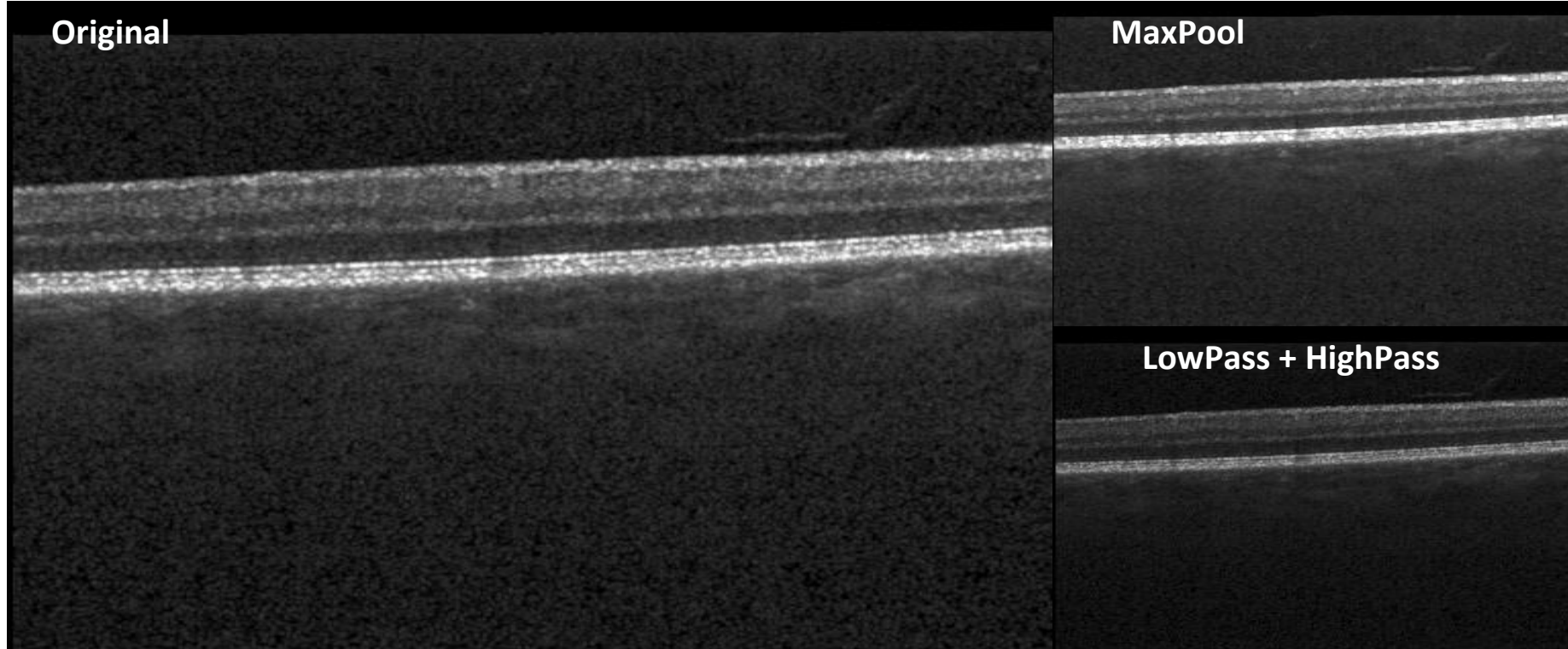


Maxpool Feature



Maxpool discards details or high-frequency information/features in the input (Lowpass Filter)

Visualization of Combining Approximated and Detailed Information



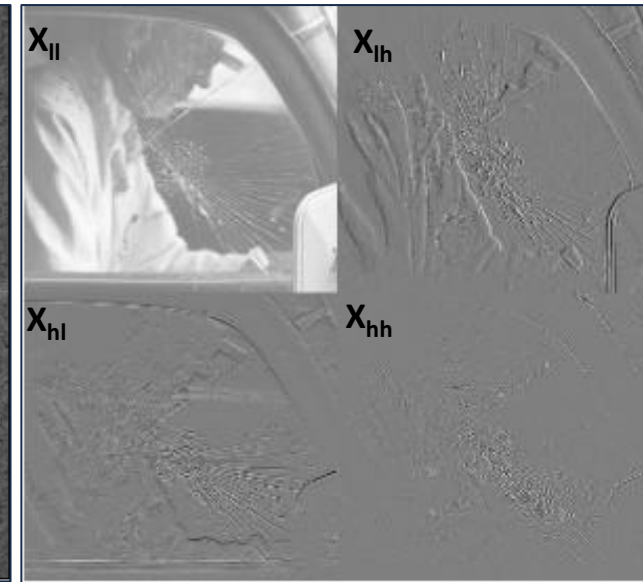
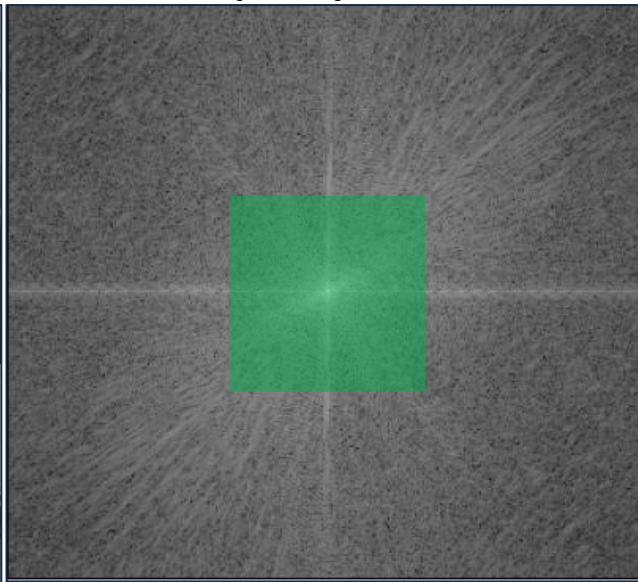
Low-pass based Pooling

- X_{ll} is the low-pass component, X_{lh} , X_{hl} , and X_{hh} are the high-pass components in Discrete Wavelet Decomposition.
- The area **inside** the **green** rectangle is the **low frequency** area, and the area **outside** the **green** rectangle is the **high frequency** area
- Other poolings such as Spectral-pooling (pooling from frequency domain) and WaveCNet (only using low-pass components) only focus on low-frequency information.

Original

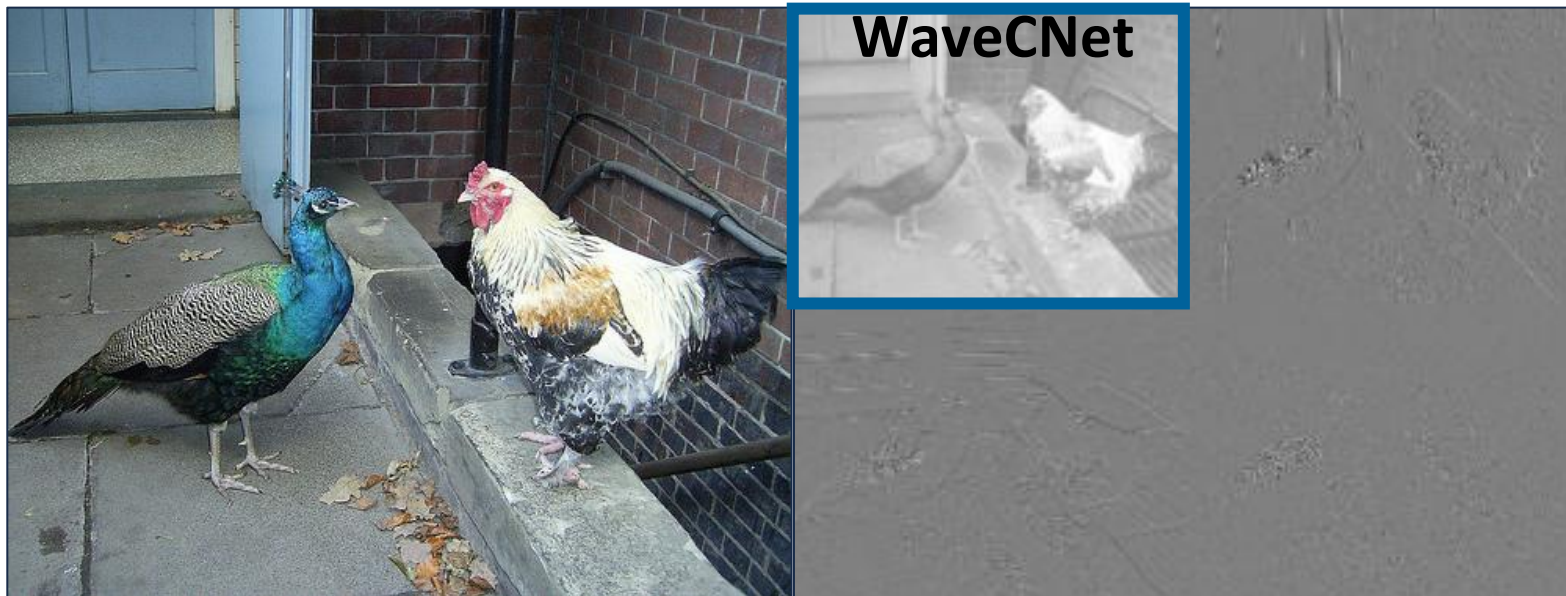
Frequency Domain

Discrete Wavelet Decomposition



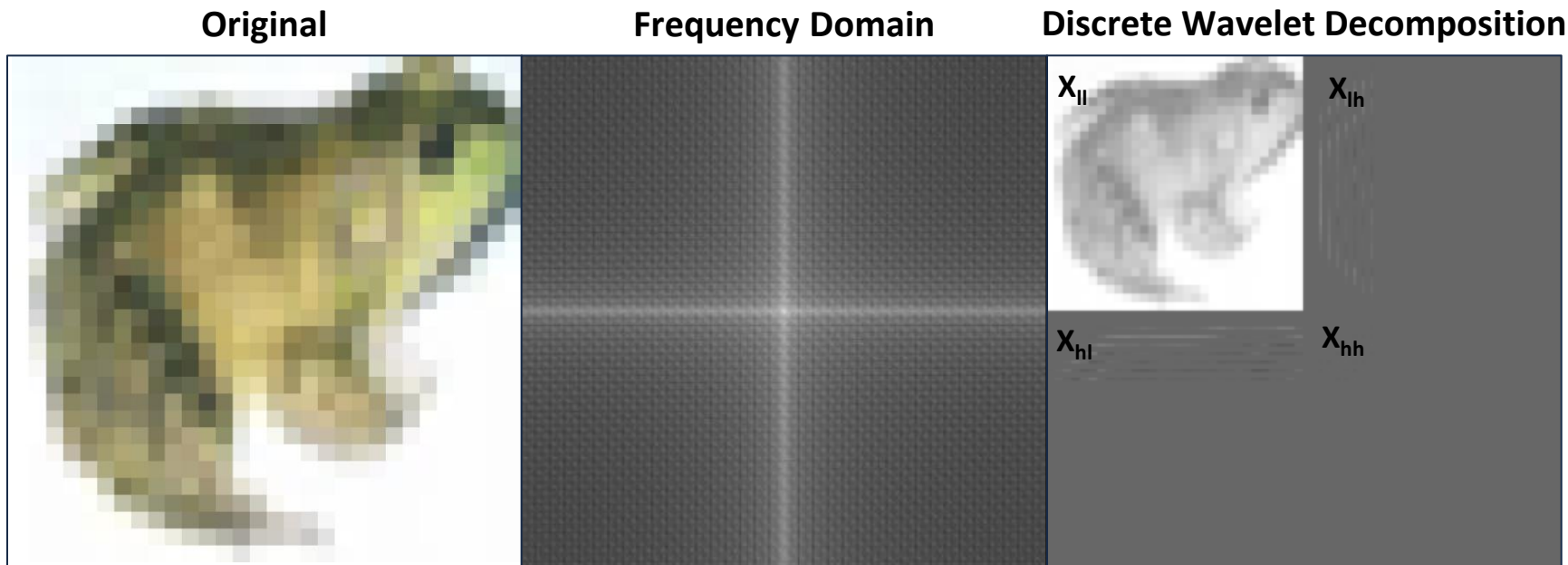
High-resolution Example and Pooling in WaveCNet

- **WaveCNet** uses **predefined wavelet functions** replace **MaxPooling**, **Down-sampling**, and **stride-2 convolution blocks** in **CNN**.
- **WaveCNet** uses only the **approximate/low-pass component** from the DWT decomposition results.



Low-resolution Data and Expectation

- Low-resolution samples do not contain much details or high-frequency features.
- Combining low and high frequency information in CNN may not show much performance improvement.



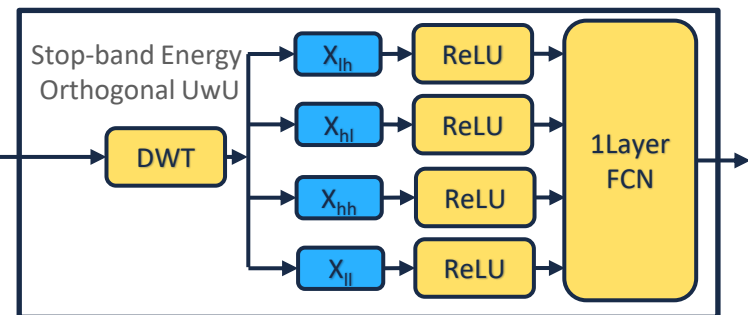
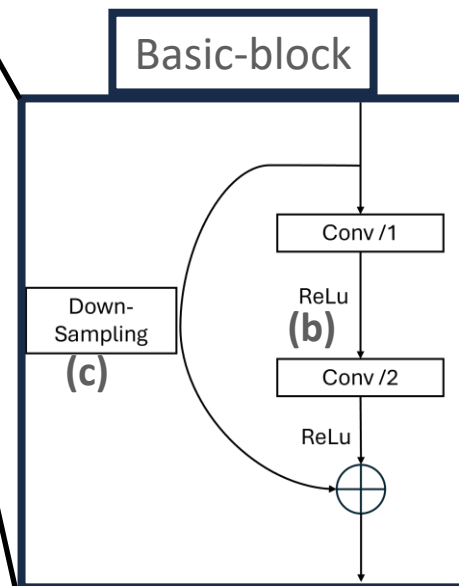
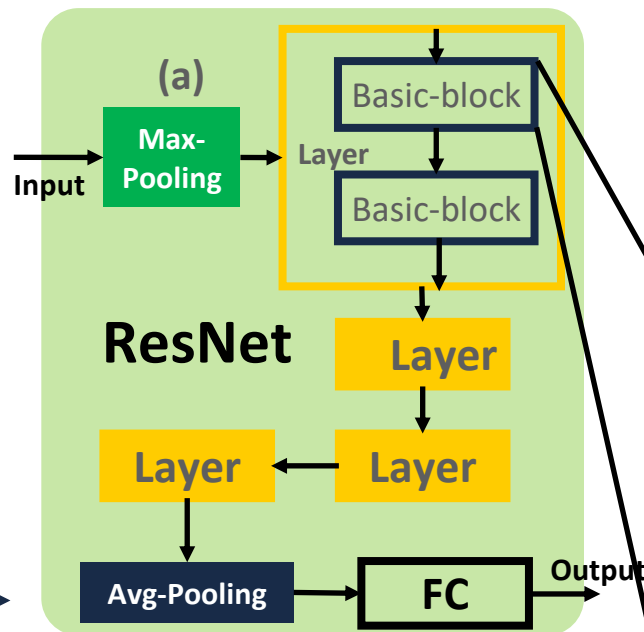
Proposed Method

Proposed Method Implementation in CNN

(a) Replacing MaxPool blocks.

(b) Replacing Stride 2 Convolution blocks.

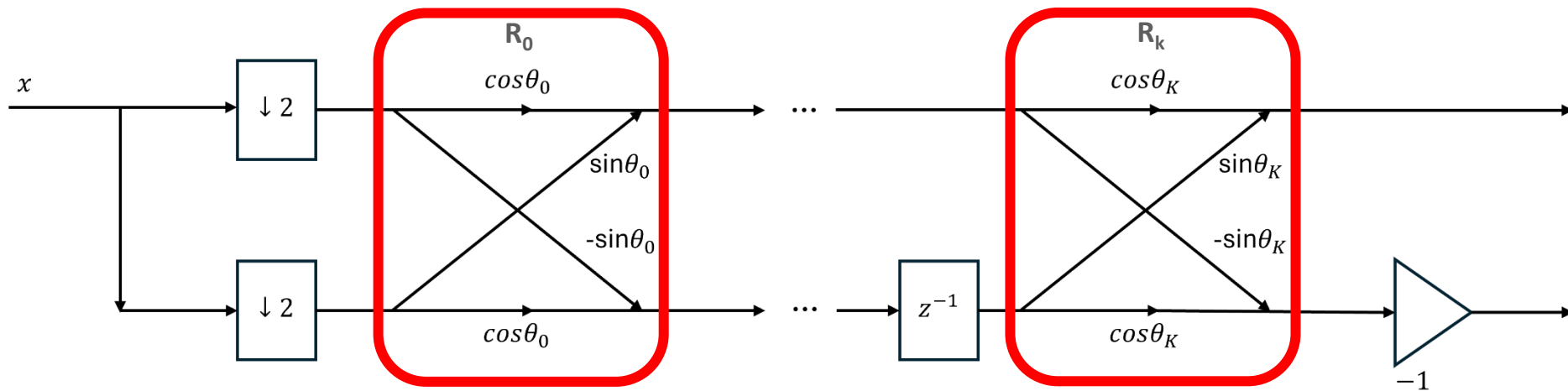
(c) Replacing down-sampling blocks.



Proposed Method

- Stop-band Energy Constraint Orthogonal Lattice Structure Universal Wavelet Unit (SBE-OrthLatt-UwU):
 - a wavelet unit with tunable coefficients.
 - uses Stop-band Energy Constraint (SBE) the enforces low-pass and high-pass characteristics.
 - integrates into Universal Wavelet Unit with Orthogonal Lattice structure (OrthLatt-UwU).
 - generates decomposed components.
 - outputs the optimal feature map from the decomposed components

Universal Wavelet Unit: Orthogonal Lattice Structure of The Filter Bank



- Lattice Structure for the Analysis Bank:**

$$\begin{aligned}
 \begin{pmatrix} \mathbf{H}_{0,K}(z) \\ \mathbf{H}_{1,K}(z) \end{pmatrix} &= \begin{pmatrix} \mathbf{H}_{0,K}(z) \\ -z^{-L} \mathbf{H}_{0,K}(-z^{-1}) \end{pmatrix} = \mathbf{H}_{P,K}(z^2) \begin{pmatrix} 1 \\ z^{-1} \end{pmatrix} \\
 &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \underbrace{\begin{bmatrix} \cos \theta_K & \sin \theta_K \\ -\sin \theta_K & \cos \theta_K \end{bmatrix}}_{\mathbf{R}} \begin{bmatrix} 1 & 0 \\ 0 & z^{-2} \end{bmatrix} \cdots \begin{bmatrix} \cos \theta_0 & \sin \theta_0 \\ -\sin \theta_0 & \cos \theta_0 \end{bmatrix} \begin{pmatrix} 1 \\ z^{-1} \end{pmatrix}
 \end{aligned}$$

The Orthogonal Lattice structure satisfies the Perfect Reconstruction Condition.

Proposed method: Stop-band Energy Constraint

• Stop-band Energy Constraint Loss:

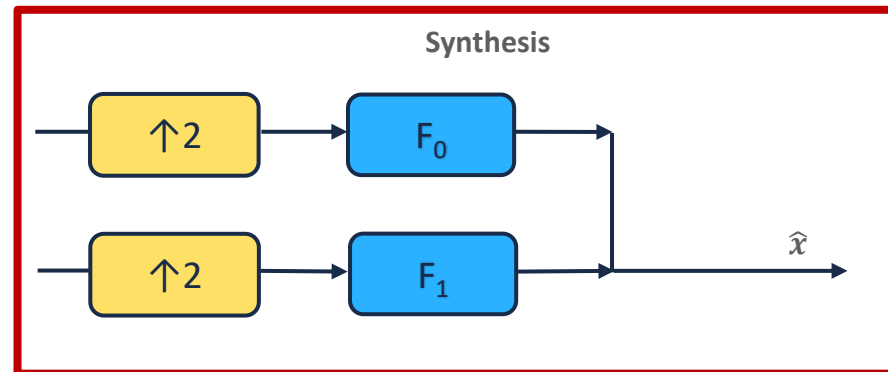
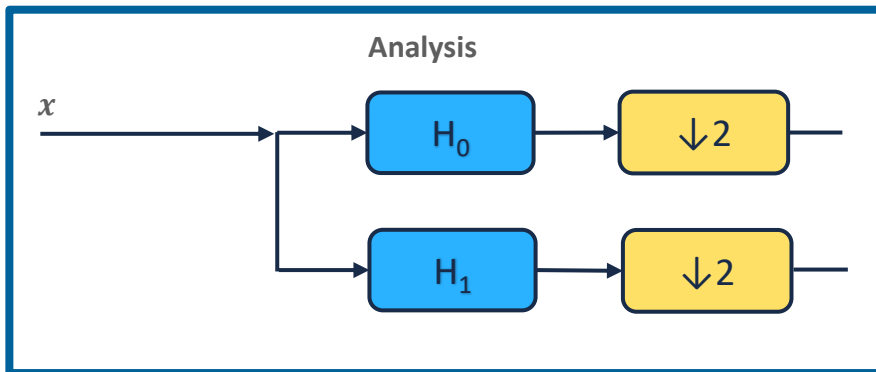
$$L_{SBE} = \frac{E_{sbn}}{E_{total}} = \frac{1}{2K} \frac{\sum_{i=w_{si}}^K |H_0(e^{jw_i})|^2}{\sum_{k=0}^{N-1} h_0[k]^2}$$

• Final Loss:

$$L_{total} = (1 - \alpha)L_{CE} + \alpha L_{SBE}$$

Anti-Aliasing Condition

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



Discrete Wavelet Decomposition

Filters: $\mathbf{L} = \begin{bmatrix} \dots & \dots & \dots \\ \dots & l_{-1} & l_0 & l_1 \\ \dots & & \dots & l_{-1} & l_0 & l_1 & \dots \\ & & & \dots & \dots \end{bmatrix}$ $\mathbf{H} = \begin{bmatrix} \dots & \dots & \dots \\ \dots & h_{-1} & h_0 & h_1 \\ \dots & & \dots & h_{-1} & h_0 & h_1 & \dots \\ & & & \dots & \dots \end{bmatrix}$

Analysis:

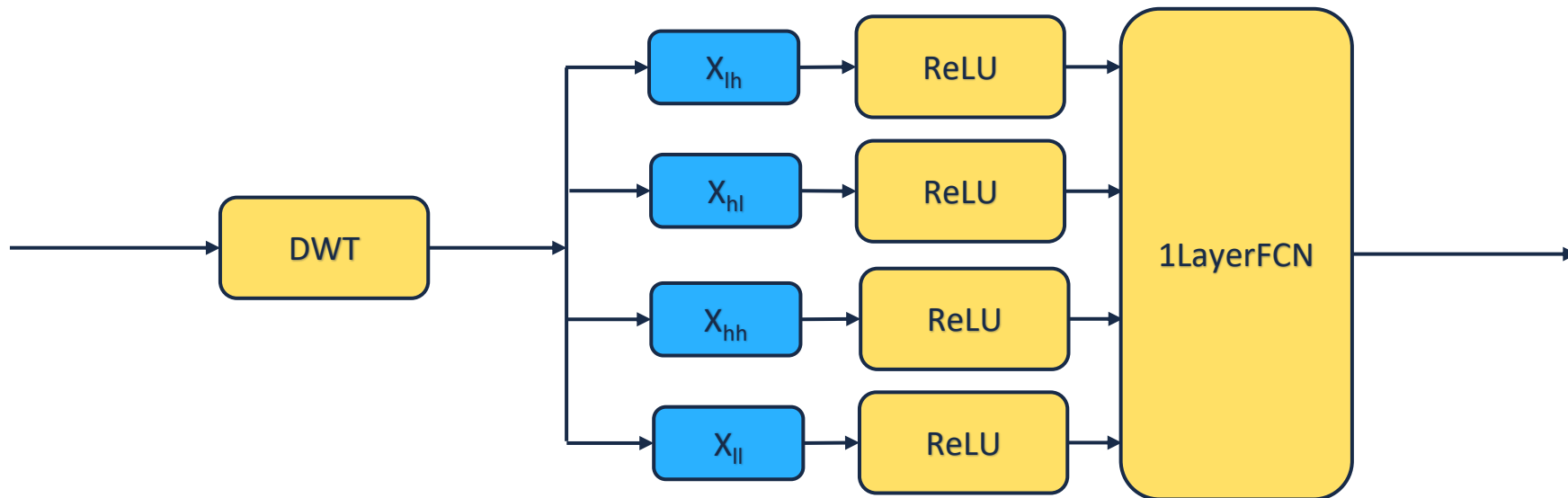
$$\begin{aligned} \mathbf{X}_{ll} &= \mathbf{L}\mathbf{X}\mathbf{L}^T, & \mathbf{X}_{lh} &= \mathbf{H}\mathbf{X}\mathbf{L}^T \\ \mathbf{X}_{hl} &= \mathbf{L}\mathbf{X}\mathbf{H}^T, & \mathbf{X}_{hh} &= \mathbf{H}\mathbf{X}\mathbf{H}^T \end{aligned}$$

Synthesis:

$$\mathbf{X} = \mathbf{L}^T \mathbf{X}_{ll} \mathbf{L} + \mathbf{L}^T \mathbf{X}_{lh} \mathbf{H} + \mathbf{H}^T \mathbf{X}_{hl} \mathbf{L} + \mathbf{H}^T \mathbf{X}_{hh} \mathbf{H}$$

Sub-band Component Fusion

Stop-band Energy Orthogonal UwU

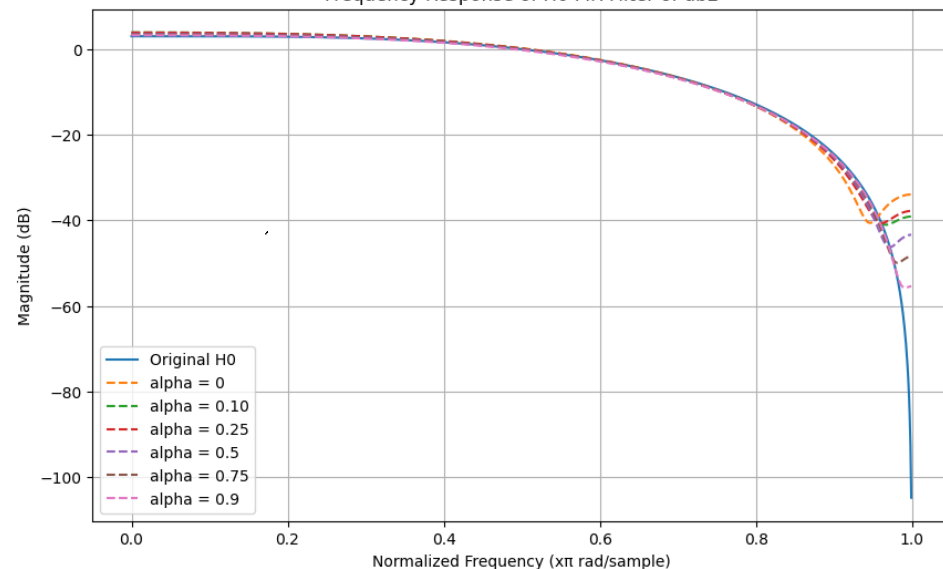


Experiments and Results

Frequency Response Analysis of SBE with different α value

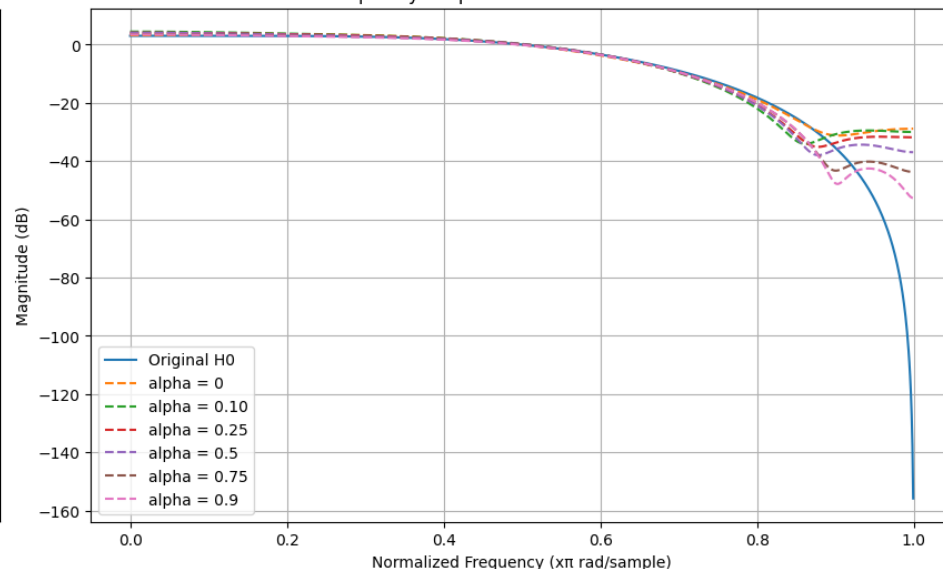
DB2

Frequency Response of H0 FIR Filter of db2



DB3

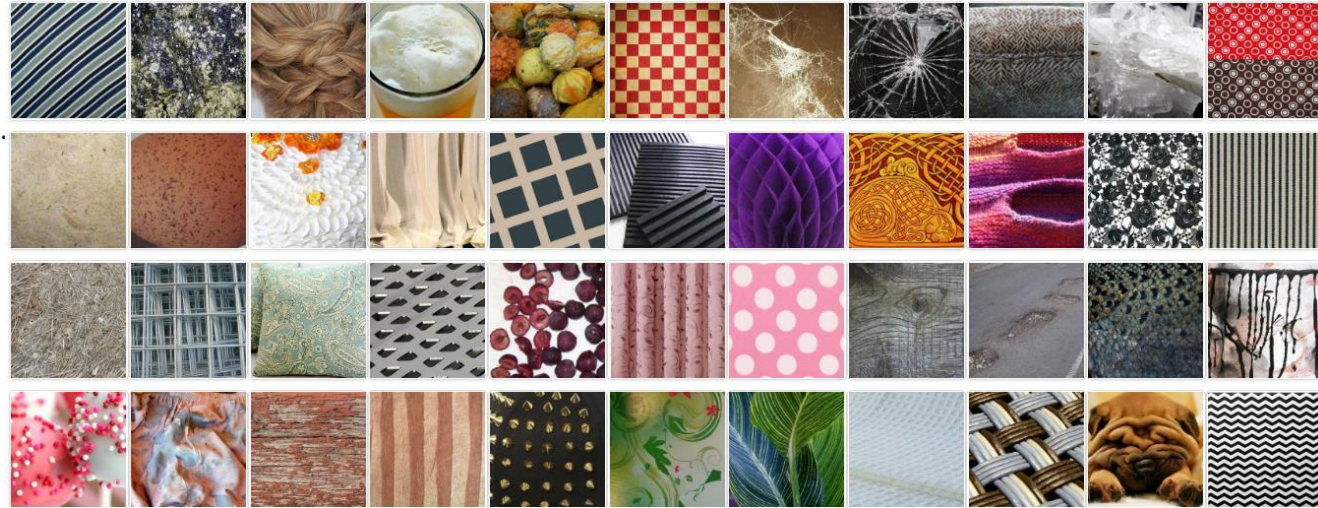
Frequency Response of H0 FIR Filter of db3



Test on Image Classification task

DTD (Describable Textures Dataset)

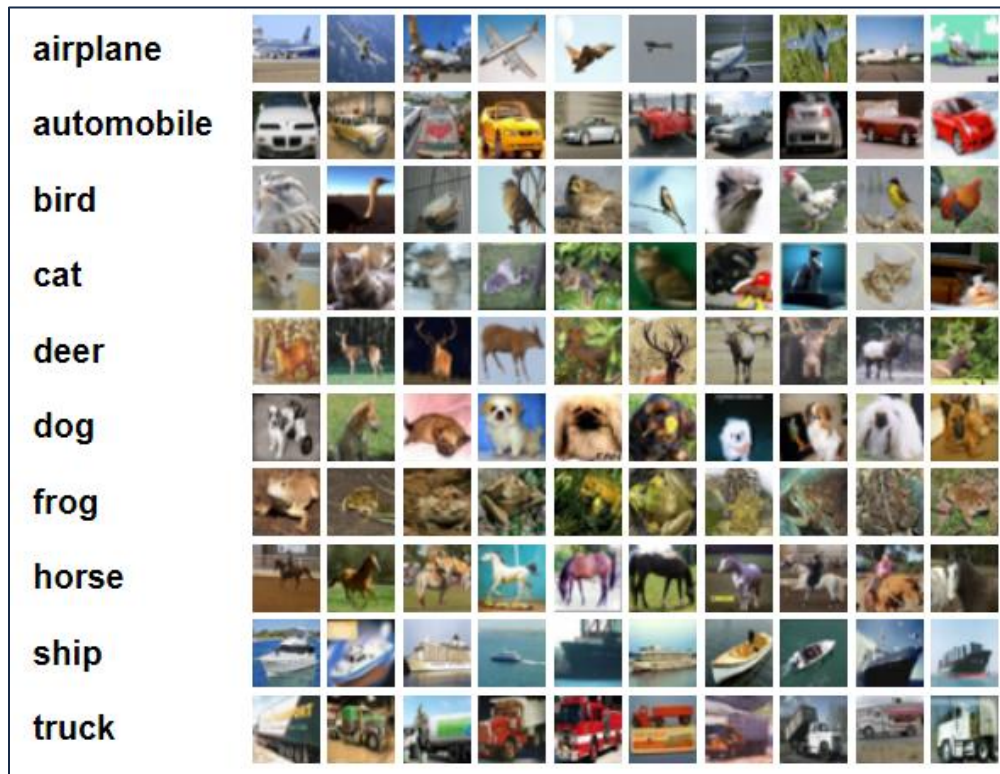
- Consisting of 5640 images for 47 categories.
- Have 120 images for each class.
- Dimensions: from 300x300 pixels to 640x640 pixels



M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed and A. Vedaldi, "Describing Textures in the Wild," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 3606-3613, doi: 10.1109/CVPR.2014.461.

CIFAR10

- Consisting of 60,000 images for 10 classes
- Training set: 50,000 images
- Test set: 10,000 images
- Dimension: 32x32 pixels color images



SBE-OrthLatt-UwU DTD and CIFAR (changing α for SBE)

	CIFAR10 ResNet18: 92.44%			DTD ResNet18: 33.99%		
	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$
2-Tap Haar	94.54%	94.83%	94.11%	41.86%	45.69%	47.55%
4-Tap DB2	94.92%	94.54%	94.07%	42.98%	43.94%	46.65%
6-Tap DB3	94.60%	94.46%	94.25%	38.40%	44.10%	46.28%
8-Tap DB4	93.85%	94.48%	94.50%	39.04%	39.10%	46.54%

DTD (vs other wavelet-based method)

Wavelet	Accuracy			
None (Baseline)	33.99			
	OrthLatt UwU [1] (without SBE)	Ours (with SBE)	PR-Relax UwU [2]	WaveCNet [3]
2-Tap Haar	40.37%	47.55%	43.67%	25.53%
4-Tap DB2	43.51%	46.65%	40.69%	23.62%
6-Tap DB3	40.59%	46.28%	35.90%	24.36%
8-Tap DB4	40.80%	46.54%	39.79%	26.91%

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

[2] 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.

[3] Q. Li, L. Shen, S. Guo and Z. Lai, "Wavelet Integrated CNNs for Noise-Robust Image Classification," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 7243–7252, doi: 10.1109/CVPR42600.2020.00727.

CIFAR10 (vs other wavelet-based method)

Wavelet	Accuracy			
None (Baseline)	92.44%			
	OrthLatt UwU [1] (without SBE)	Ours (with SBE)	PR-Relax UwU [2]	WaveCNet [3]
2-Tap Haar	94.27	94.83	94.97	94.76
4-Tap DB2	95.05	94.92	94.66	94.93
6-Tap DB3	95.03	94.60	93.37	94.56
8-Tap DB4	95.11	94.50	94.21	93.81

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

[2] 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.

[3] Q. Li, L. Shen, S. Guo and Z. Lai, "Wavelet Integrated CNNs for Noise-Robust Image Classification," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 7243–7252, doi: 10.1109/CVPR42600.2020.00727.

Compare with other architectures on CIFAR10 (ResNet18)

Models	CIFAR10 (Accuracy)
ResNet18 (baseline)	92.44%
ResNet18 (SpectralPool) [1]	92.50% (+0.06%)
Resnet18 (DiffStride) [2]	92.90% (+0.46%)
WA-CNN ResNet18 [3]	92.57%(+0.13%)
WaveCNET (DB2) ResNet18 [4]	94.93% (+2.49%)
Orthogonal-LatticeUwU (8-tap DB4) ResNet18	95.11% (+2.67%)
PR-relaxation (2-tap Haar) ResNet18	94.97%(+2.52%)
StopbandEnergyOrthogonal-UwU (alpha=0.25, 4-tap DB2) ResNet18	94.92% (+2.48%)

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

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

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

[4] Q. Li, L. Shen, S. Guo and Z. Lai, "Wavelet Integrated CNNs for Noise-Robust Image Classification," 2020 IEEE/CVF Conference on Computer

[5] 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.

ResNet34 on DTD and CIFAR10

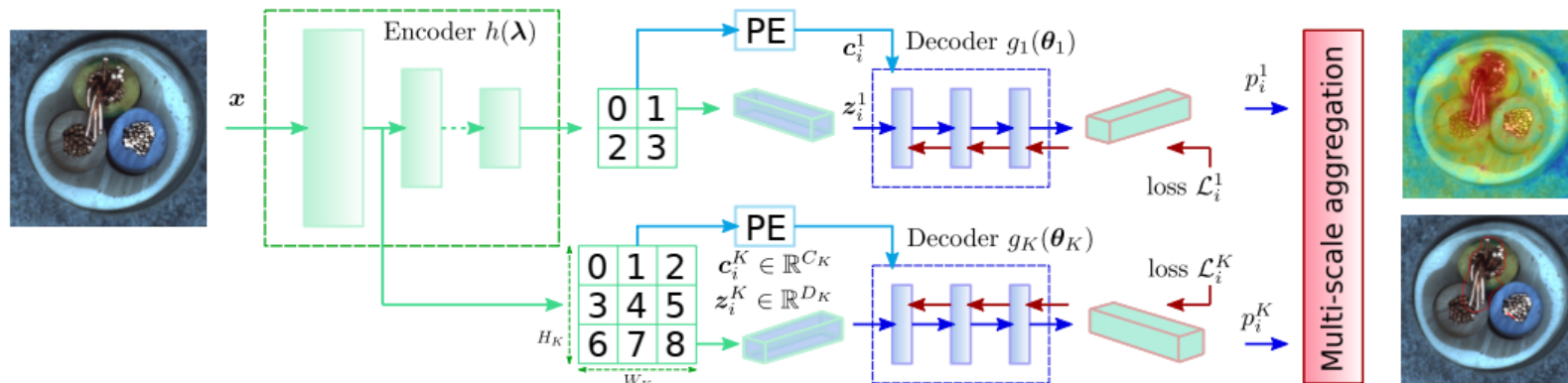
ResNet34 Extension

Wavelet	DTD				
None (Baseline)	24.47%				
	Orthogonal-LatticeUwU (without SBE)	Ours ($\alpha = 0.25$) (without SBE)	Ours ($\alpha = 0.5$) (without SBE)	Ours ($\alpha = 0.75$) (without SBE)	WaveCNet
2-tap Haar	41.49%	36.70%	41.49%	43.14%	32.39%
4-tap Haar	38.88%	36.65%	43.03%	45.00%	32.23%

Wavelet	CIFAR10		
None (Baseline)	94.33%		
	Orthogonal-LatticeUwU (without SBE)	Ours (without SBE)	WaveCNet
2-tap Haar	95.44%	94.71% ($\alpha = 0.5$)	95.07%
4-tap Haar	95.61%	94.63% ($\alpha = 0.75$)	95.12%

Anomaly Detection

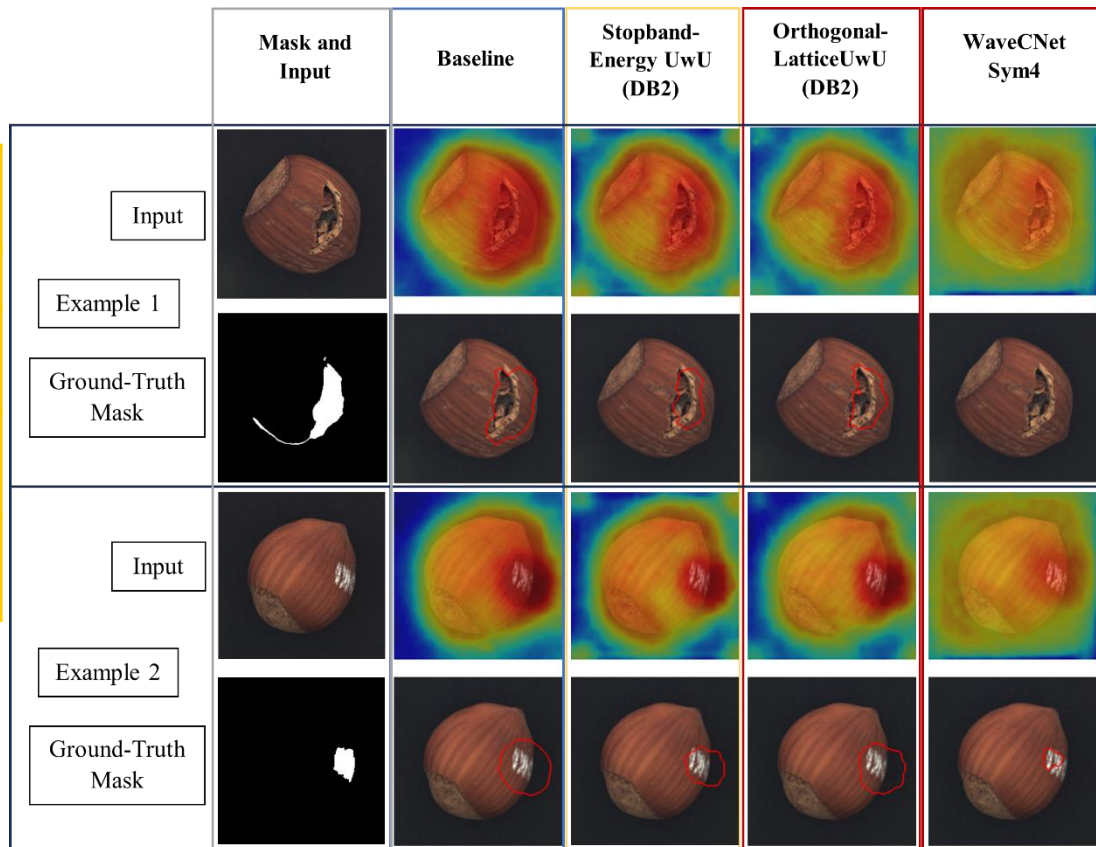
Model used for Anomaly Detection CFLOW (WACV 2022) (Baseline)



- Encoder with Image Classification backbone (ResNet18)
- Decoder with normalizing flow architecture to estimate the unknown density of z to by an estimated parameterized density.
- The decoder is a sequence of invertible transformations to estimate the likelihood of the feature vectors.
- PE (Positional Encoder) is used to provide spatial conditions to the decoder.
- The estimates probabilities from all the scales are up-sampled to the original resolution and then aggregated for anomaly score calculation.

Anomaly Detection

Models	SEG-AUROC	DET-AUROC
Baseline ResNet18 CFLOW-AD	96.45%	92.46%
OrthLatt-UwU 4-tap DB2 ResNet18 CFLOW-AD	97.15%	89.57%
WaveCNet Sym4 ResNet18 CFLOW-AD	96.20%	87.61%
SBE-OrthLatt-UwU (DB2, 0.75) ResNet18 CFLOW-AD	96.99%	94.46%



Conclusion

- The Stop-band energy (SBE) constraint is introduced to enhance frequency selectivity in tunable wavelets.
- The SBE constraint improves CNN performance in high-frequency tasks like image classification and anomaly detection.
- The SBE constraint in CNN implementation boosts performance on texture-rich datasets, maintaining competitive results on low-resolution ones.
- Though not always improving top-1 accuracy, it enhances the model's ability to capture fine details.
- The approach is computationally efficient and broadly applicable, with future extensions planned for Transformer architectures.

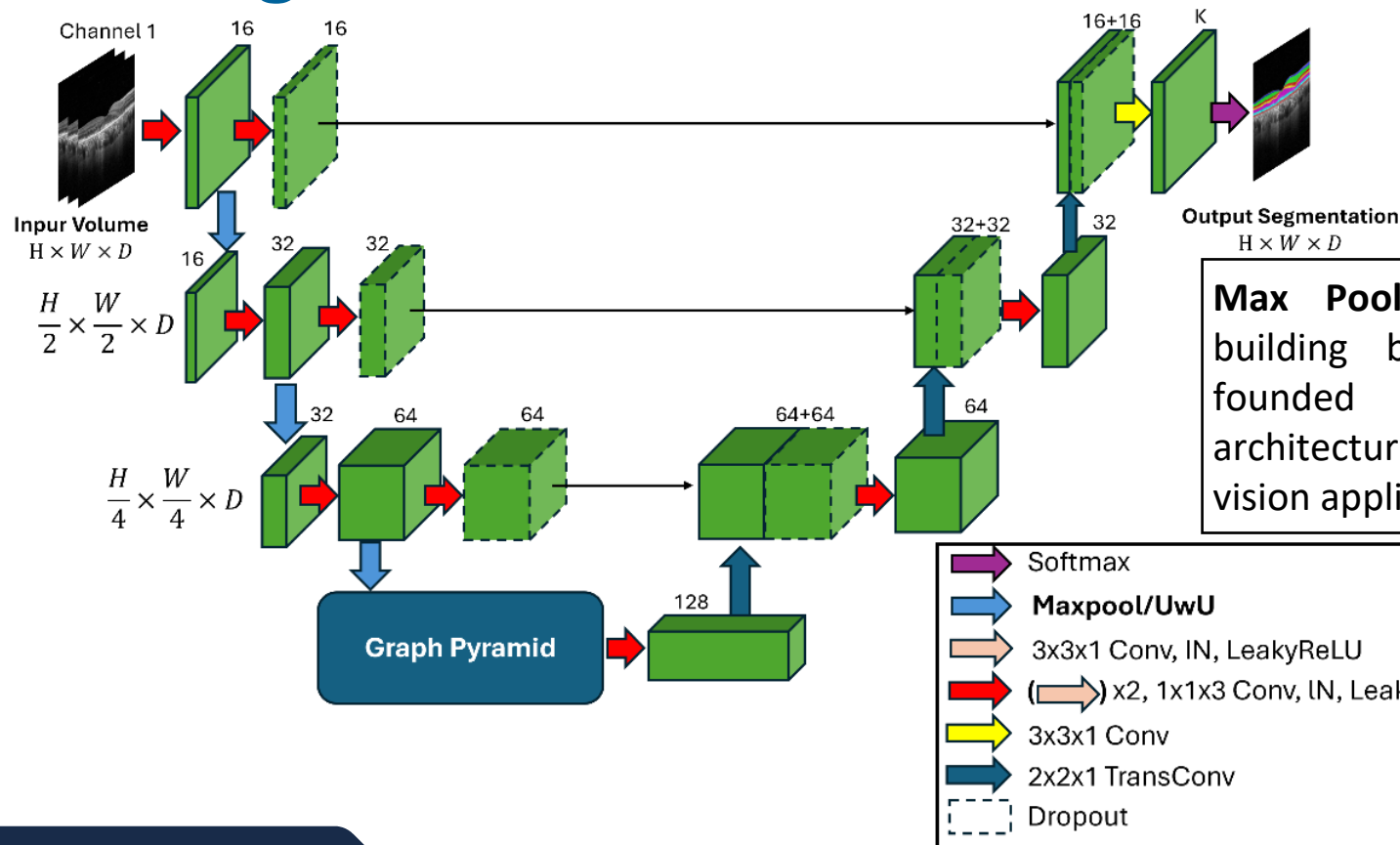
Thank you

Reference

- [1] G. Denis, I. Shun and K. Kazuki, "CFLOW-AD: Real-Time Unsupervised Anomaly Detection With Localization via Conditional Normalizing Flows", *WACV*, pp. 98-107, 2022.
- [2] A. Krizhevsky, "Learning multiple layers of features from tiny images," 2009
- [3] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed and A. Vedaldi, "Describing Textures in the Wild," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 3606-3613, doi: 10.1109/CVPR.2014.461.
- [4] A. D. Le, S. Jin, Y.-S. Bae, and T. Q. Nguyen, "A lattice-structurebased trainable orthogonal wavelet unit for image classification," *IEEE Access*, vol. 12, pp. 88715–88727, 2024.
- [5] 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.
- [6] Q. Li, L. Shen, S. Guo and Z. Lai, "Wavelet Integrated CNNs for Noise-Robust Image Classification," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 7243-7252, doi: 10.1109/CVPR42600.2020.00727.
- [7] R. Zhang, "Making convolutional networks shift-invariant again," in *ICML*, 2019.
- [8] R. Riad, O. Teboul, D. Grangier, and N. Zeghidour, "Learning strides in convolutional neural networks," *ICLR*, 2022.
- [9] X. Zhao, P. Huang, and X. Shu, "Wavelet-attention cnn for image classification," *Multimedia Syst.*, vol. 28, p. 915–924, jun 2022.
- [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.

Backup

3D Segmentation Model



Max Pooling is a common building block that can be founded in most CNN architectures and computer vision applications.

Implementation in CNN

- (a) Replacing MaxPool blocks.
- (b) Replacing Stride 2 Convolution blocks.
- (c) Replacing down-sampling blocks.

