A General Graph Spectral Wavelet Convolution via Chebyshev Order Decomposition

Nian Liu[1], Xiaoxin He[1], Thomas Laurent[2],

Francesco Di Giovanni[3], Michael M. Bronstein[3,4], Xavier Bresson[1]

[1]National University of Singapore, [2]Loyola Marymount University

[3]University of Oxford, [4]AITHYRA, Austria

Convolution Theorem

Kernel

$$\boldsymbol{\kappa} * \boldsymbol{X} = F^{-1}(F(\boldsymbol{\kappa}) \cdot F(\boldsymbol{X}))$$

Inverse Transform — Transform

Classical Graph Convolution

Vector-valued Kernel

$$\boldsymbol{\kappa} *_G \boldsymbol{X} = \boldsymbol{U}\left(diag(\theta_\lambda) \cdot \boldsymbol{U}^\top \boldsymbol{X}\right)$$

Fourier Basis

How to further advance …

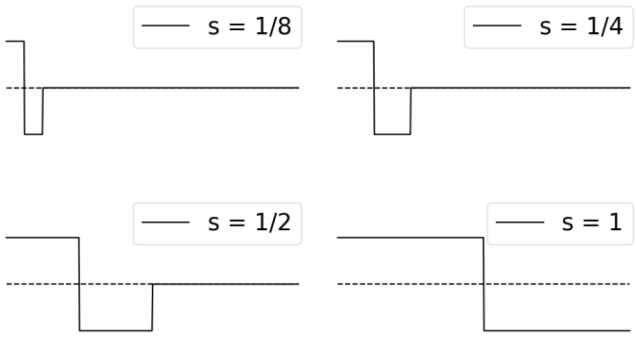
Improve the Basis $\boldsymbol{\kappa} * \boldsymbol{X} = F^{-1}(F(\boldsymbol{\kappa}) \cdot F(\boldsymbol{X}))$

- Graph Fourier Basis
  - $\boldsymbol{U}$: constant resolution & fixed pattern

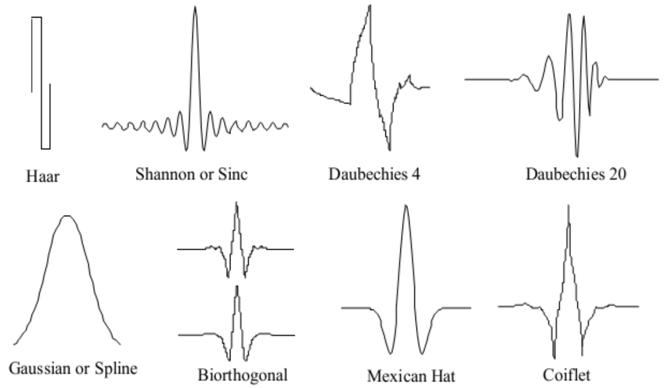
Improve the Basis

$$\boldsymbol{\kappa} * \boldsymbol{X} = F^{-1}(F(\boldsymbol{\kappa}) \cdot F(\boldsymbol{X}))$$

- Graph Fourier Basis
  - $\boldsymbol{U}$: constant resolution & fixed pattern
- Wavelet Basis
  - $\Psi_{s,a}(x) = \frac{1}{s}\Psi\left(\frac{x-a}{s}\right)$, s: scale, a: location



Multiple resolutions & Scaled receptive fields



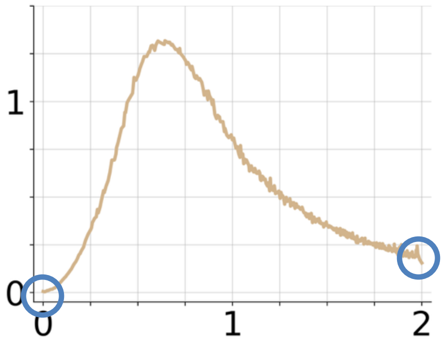
Adaptive patterns & Learnable

Improve the Basis

$$\boldsymbol{\kappa} * \boldsymbol{X} = F^{-1}(F(\boldsymbol{\kappa}) \cdot F(\boldsymbol{X}))$$

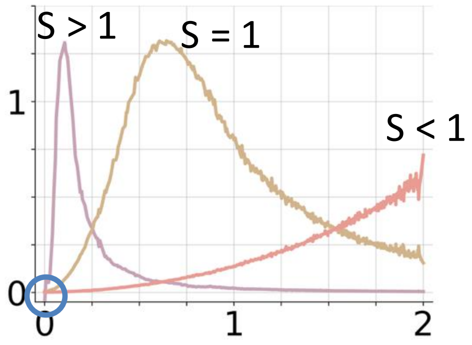
- Graph Wavelet Bases[1]

Unit wavelet $\Psi = Ug(\lambda)U^{\top}$

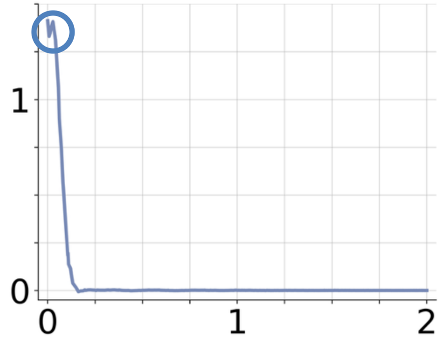


Wavelet Admissibility Criteria[2]

Multiple scales $g(s\lambda)$



Scaling function $\Phi = Uh(\lambda)U^{\top}$



Supplement direct current signals

$$C_{\Psi} = \int_{-\infty}^{\infty} \frac{|g(\lambda)|^2}{|\lambda|} d\lambda < \infty. \quad , \quad g(\lambda = 0) = 0 \text{ and } \lim_{\lambda \to \infty} g(\lambda) = 0$$

How to design graph wavelets …

[1] Hammond, D. K., Vandergheynst, P., & Gribonval, R. (2011). Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, *30*(2), 129-150.
[2] Mallat, S. (1999). A wavelet tour of signal processing.

Improve the Basis

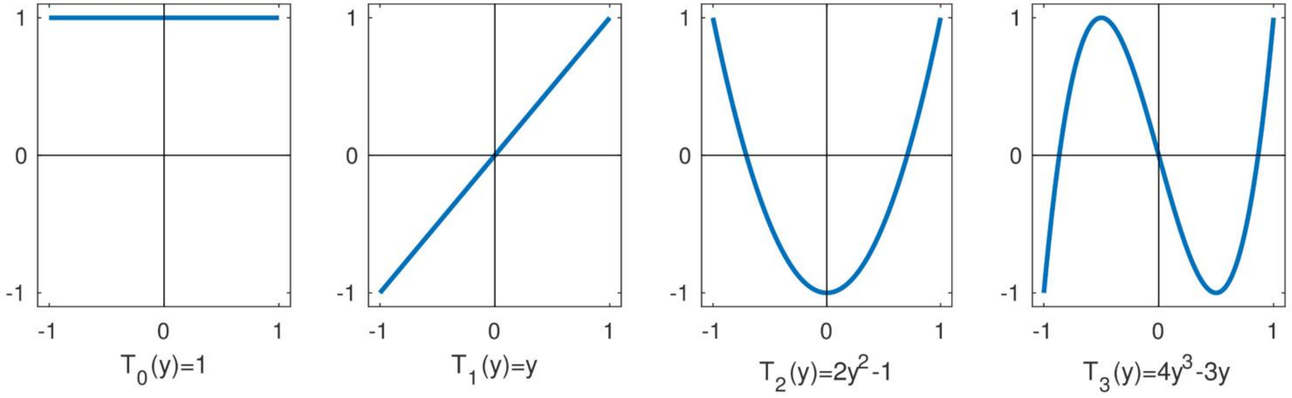
$$\boldsymbol{\kappa} * \boldsymbol{X} = F^{-1}(F(\boldsymbol{\kappa}) \cdot F(\boldsymbol{X}))$$

➢ Design of Graph Wavelet Bases – Chebyshev expansion[1]

- Approximate any function using polynomials

$$T_k(y) = \begin{cases} 1 & k = 0 \\ y & k = 1 \\ 2yT_{k-1}(y) - T_{k-2}(y) & k > 1 \end{cases} \Rightarrow f(y) = \frac{1}{2}c_0 + \sum_{k=1}^{\infty} c_k T_k(y)$$

- Waveform



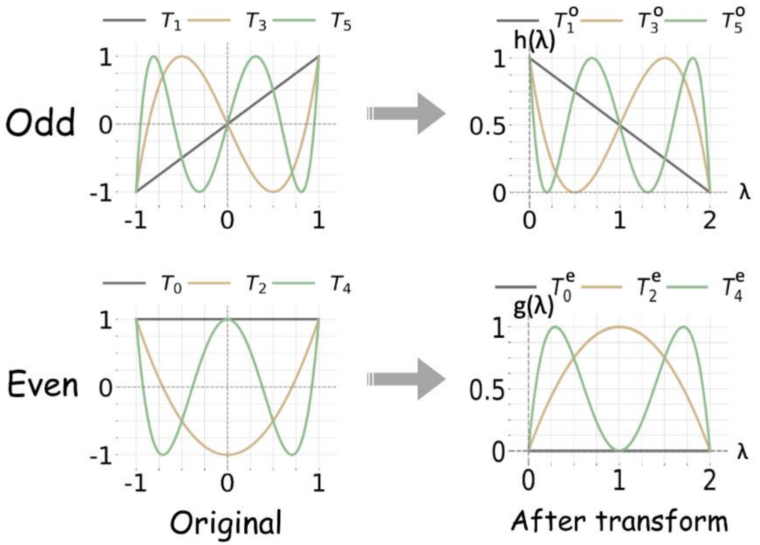
[1] Hammond, D. K., Vandergheynst, P., & Gribonval, R. (2011). Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, *30*(2), 129-150.

Improve the Basis $\boldsymbol{\kappa} * \boldsymbol{X} = F^{-1}(F(\boldsymbol{\kappa}) \cdot F(\boldsymbol{X}))$

- Chebyshev Decomposition

Separately introduce odd terms and even terms from Chebyshev polynomials into the approximation of scaling function and wavelet.

$$T_k(y) \rightarrow 1/2 \cdot (-T_k(y-1)+1)$$



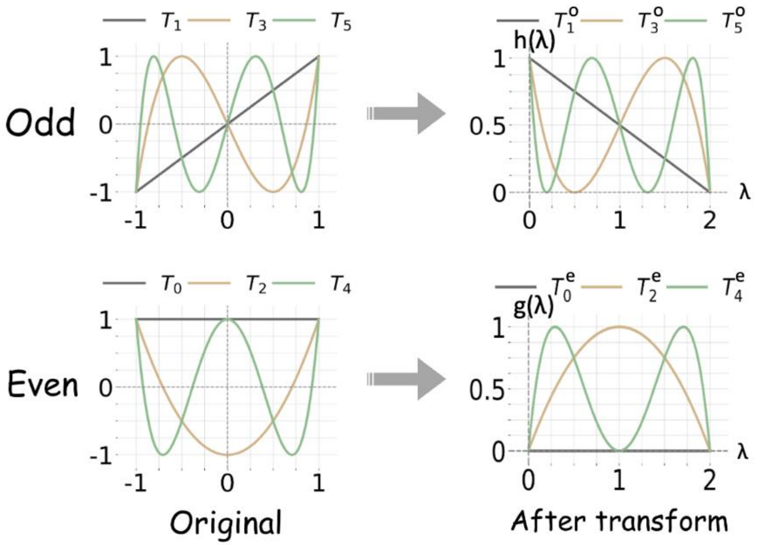
Improve the Basis

$$\boldsymbol{\kappa} * \boldsymbol{X} = F^{-1}(F(\boldsymbol{\kappa}) \cdot F(\boldsymbol{X}))$$

➢ Chebyshev Decomposition

Separately introduce odd terms and even terms from Chebyshev polynomials into the approximation of scaling function and wavelet.

$$T_k(y) \rightarrow 1/2 \cdot (-T_k(y-1)+1)$$



- Scaling Function

$$h(\boldsymbol{\Lambda}) = \sum_i^{\rho} b_i T_i^o(\boldsymbol{\Lambda})$$

- Wavelet

$$g(\boldsymbol{\Lambda}) = \sum_i^{\rho} a_i T_i^e(\boldsymbol{\Lambda})$$

- Scales

$$\tilde{\boldsymbol{s}} = \sigma(\text{Mean}(\boldsymbol{W_s}\hat{\boldsymbol{Z}} + \boldsymbol{b_s})) \cdot \overline{\boldsymbol{s}}$$

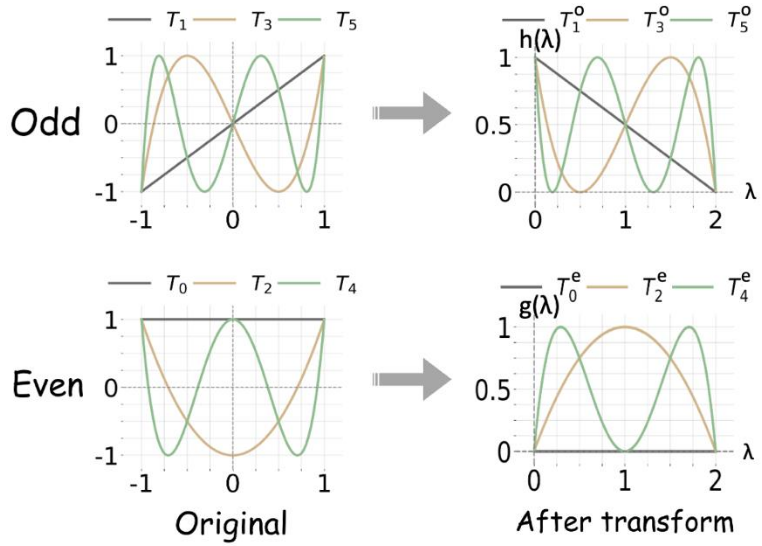
Improve the Basis

$$\boldsymbol{\kappa} * \boldsymbol{X} = F^{-1}(F(\boldsymbol{\kappa}) \cdot F(\boldsymbol{X}))$$

➢ Chebyshev Decomposition

Separately introduce odd terms and even terms from Chebyshev polynomials into the approximation of scaling function and wavelet.

$$T_k(y) \rightarrow 1/2 \cdot (-T_k(y-1)+1)$$



- Theoretically correct
- Easily available
- Arbitrarily complex
- Adaptively learnable
- Multiple ranges

…

Improve the Kernel $\kappa * X = F^{-1}(F(\kappa) \cdot F(X))$

- Vector-valued Kernel, diag(θ_λ)
  - Scale global frequency patterns, e.g., low and high frequencies
  - Not suitable for wavelet signals → localized, node-specific patterns

[1] Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. NeurIPS, 29.
[2] Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.
[3] Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020). Fourier neural operator for parametric partial differential equations. arXiv preprint arXiv:2010.08895.

Improve the Kernel $\boldsymbol{\kappa} * \boldsymbol{X} = F^{-1}(F(\boldsymbol{\kappa}) \cdot F(\boldsymbol{X}))$

- Vector-valued Kernel, diag(θ_λ)
  - Scale global frequency patterns, e.g., low and high frequencies
  - Not suitable for wavelet signals → localized, node-specific patterns
- Matrix-valued Kernel, FNO[3]

$$\boldsymbol{H}^{(l+1)} = \mathcal{F}^{-1}(R_\theta \cdot \mathcal{F}(\boldsymbol{H}^{(l)})),\ R_\theta \in \mathbb{R}^{N \times d \times d}$$

  - No constraints, More parameters
  - Over-fitting → weight sharing

$$\mathbb{M} *_{\mathcal{G}} \boldsymbol{X} = \mathcal{F}^{-1}\mathbb{M} \circ \mathcal{F}(\boldsymbol{X}) = \mathcal{F}^{-1}(\mathrm{MLP}(\mathcal{F}(\boldsymbol{X})))$$

$R_\theta: N \times d \times d \rightarrow d \times d$

[1] Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. NeurIPS, 29.
[2] Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.
[3] Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020). Fourier neural operator for parametric partial differential equations. arXiv preprint arXiv:2010.08895.

Integrate two factors

- Wavelet based Graph Convolution -- WaveGC

$$H^{(l+1)} = \sigma\left(\left[\Phi\mathbb{S}\circ\Phi H^{(l)} || \Psi_{s_1}\mathbb{W}_1 \circ \Psi_{s_1} H^{(l)} || \ldots || \Psi_{s_J}\mathbb{W}_J \circ \Psi_{s_J} H^{(l)}\right]\cdot W\right)$$

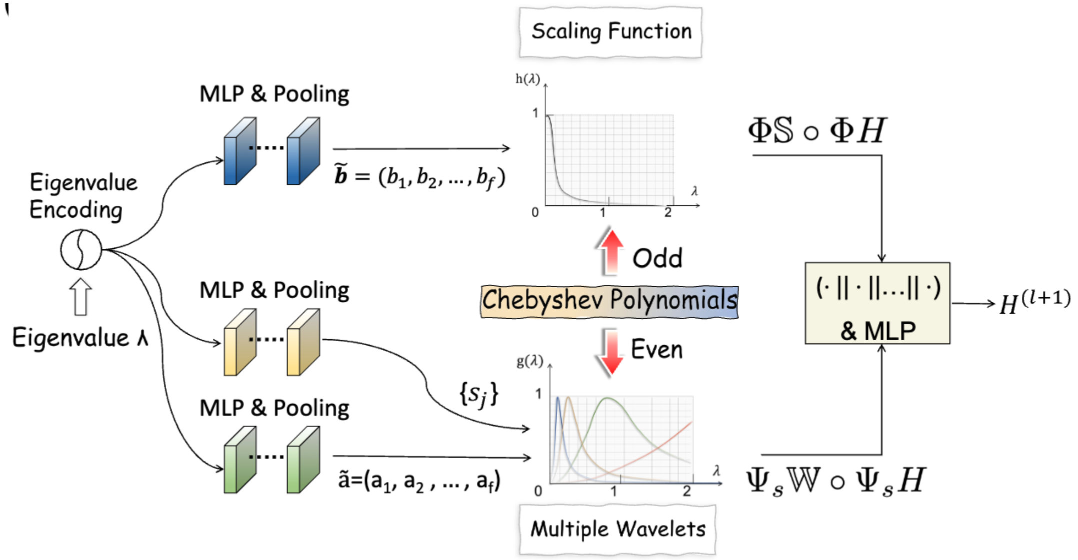
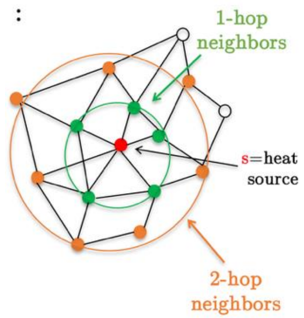


Table 1: Comparison between classical graph convolution and WaveGC.

	Classical Graph Convolution	WaveGC
Formula	$\sigma(U\text{diag}(\theta_\lambda)U^\top H \cdot W)$	$\sigma([\Phi\mathbb{S}\circ\Phi H \,\|\|\, \Psi_s\mathbb{W}\circ\Psi_s H]\cdot W)$
Kernel	$\text{diag}(\theta_\lambda)$ (Vector)	$\mathbb{S}$ / $\mathbb{W}$ (Matrix)
Bases	$U^\top$ (Fourier basis)	Φ / Ψ_s (Scaling / Wavelet basis)

Theorem – diverse receptive fields, scale s



- Compare $\sigma(\Psi_s HW)$ vs. $\sigma\left(\sum_{j=0}^{K} \tau_j A^j HW\right), \tau_j \in [0,1]$

Theorem 4.2 (Short-range and long-range receptive fields). *Given a large even number $K > 0$ and two random nodes a and b, if the depths m_Ψ and m_A are necessary for $\sigma(\Psi_s HW)$ and $\sigma(\sum_{j=0}^{K} \tau_j A^j HW)$ to induce the same amount of mixing $mix_{y_G}(b, a)$, then the lower bounds of m_Ψ and m_A, i.e. L_{m_Ψ} and L_{m_A}, approximately satisfy the following relation when scale $s \to 0$:*

$$L_{m_\Psi} \approx \frac{P}{K} L_{m_A} + \frac{2|E|}{K\sqrt{d_a d_b}} \frac{mix_{y_G}(b, a)}{\gamma} \cdot \frac{1}{(\alpha^2 s^{2K})^{m_\Psi}}. \tag{12}$$

Or, if $s \to \infty$, the relation becomes:

$$L_{m_\Psi} \approx \frac{P}{K} L_{m_A} - \frac{2|E|}{K(K+1)^{2m_A} \tau_P{}^{2m_A} \sqrt{d_a d_b}} \frac{mix_{y_G}(b, a)}{\gamma}, \tag{13}$$

where $P < K$ and $(\tau_P A^P)_{ba} = \max\{(\tau_m A^m)_{ba}\}_{m=0}^{K}$. d_a and d_b are degrees of two nodes, and $\alpha = \frac{C \cdot 2^K (K+1)}{K!}$. $\gamma = \sqrt{\frac{d_{max}}{d_{min}}}$, where d_{max}/d_{min} is the maximum / minimum degree in the graph.

Numerical Results

Table 2. Qualified results on short-range tasks compared to baselines. **Bold**: Best, Underline: Runner-up, OOM: Out-of-memory. All results are reproduced based on source codes.

Model	CS	Photo	Computer	CoraFull	ogbn-arxiv
	Accuracy ↑	Accuracy ↑	Accuracy ↑	Accuracy ↑	Accuracy ↑
GCN	92.92±0.12	92.70±0.20	89.65±0.52	61.76±0.14	71.74±0.29
GAT	93.61±0.14	93.87±0.11	90.78±0.13	64.47±0.18	71.82±0.23
APPNP	94.49±0.07	94.32±0.14	90.18±0.17	65.16±0.28	71.90±0.25
Scattering	94.77±0.33	92.10±0.61	85.68±0.71	57.65±0.84	66.23±0.19
Scattering GCN	95.18±0.30	93.07±0.42	88.83±0.44	61.14±1.13	71.18±0.76
SGWT	94.81±0.23	92.45±0.62	85.19±0.59	55.04±1.12	69.08±0.30
GWNN	90.75±0.59	94.45±0.45	90.75±0.59	64.19±0.79	71.13±0.47
UFGConvS	95.33±0.27	93.98±0.59	88.68±0.39	61.25±0.93	70.04±0.22
UFGConvR	95.46±0.33	94.34±0.34	89.29±0.46	62.43±0.80	71.97±0.12
WaveShrink-ChebNet	94.90±0.30	93.54±0.90	88.20±0.65	58.98±0.69	OOM
DEFT	95.04±0.32	94.35±0.44	91.63±0.52	68.01±0.86	72.01±0.20
WaveNet	94.91±0.29	94.09±0.63	92.06±0.33	57.65±1.05	71.37±0.14
SEA-GWNN	95.11±0.37	94.35±0.50	89.88±0.64	66.74±0.79	72.64±0.21
WaveGC (ours)	**95.89±0.34**	**95.37±0.44**	**92.26±0.18**	**69.14±0.78**	**73.01±0.18**

Numerical Results

Table 3. Qualified results on long-range tasks compared to baselines. **Bold**: Best, Underline: Runner-up, OOM: Out-of-memory, All results are reproduced based on source codes.

Model	VOC	PCQM	COCO	Pf	Ps
	F1 score ↑	MRR ↑	F1 score ↑	AP ↑	MAE ↓
GCN	12.68±0.60	32.34±0.06	08.41±0.10	59.30±0.23	34.96±0.13
GINE	12.65±0.76	31.80±0.27	13.39±0.44	54.98±0.79	35.47±0.45
GatedGCN	28.73±2.19	32.18±0.11	26.41±0.45	58.64±0.77	34.20±0.13
Scattering	16.58±0.49	33.90±0.27	16.44±0.79	56.80±0.38	26.77±0.11
Scattering GCN	30.45±0.36	33.73±0.45	30.27±0.60	62.87±0.64	26.43±0.20
SGWT	31.22±0.56	34.04±0.05	32.97±0.53	60.23±0.27	25.39±0.21
GWNN	25.60±0.56	32.72±0.08	13.39±0.44	65.47±0.48	27.34±0.04
UFGConvS	31.27±0.39	33.94±0.24	23.15±0.55	65.83±0.75	27.08±0.58
UFGConvR	31.08±0.33	34.08±0.20	26.02±0.48	65.29±0.82	27.50±0.21
WaveShrink-ChebNet	18.80±0.85	32.56±0.11	11.12±0.46	61.12±0.53	27.45±0.06
DEFT	35.98±0.20	34.25±0.06	30.14±0.49	66.95±0.63	25.06±0.13
WaveNet	28.60±0.15	33.19±0.20	23.06±0.18	64.63±0.27	25.88±0.01
SEA-GWNN	31.97±0.55	29.89±0.26	24.33±0.23	68.75±0.20	25.64±0.31
WaveGC (ours)	**41.63±0.19**	**34.50±0.02**	**35.96±0.22**	**69.73±0.43**	**24.83±0.11**

Effectiveness of Wavelet Basis

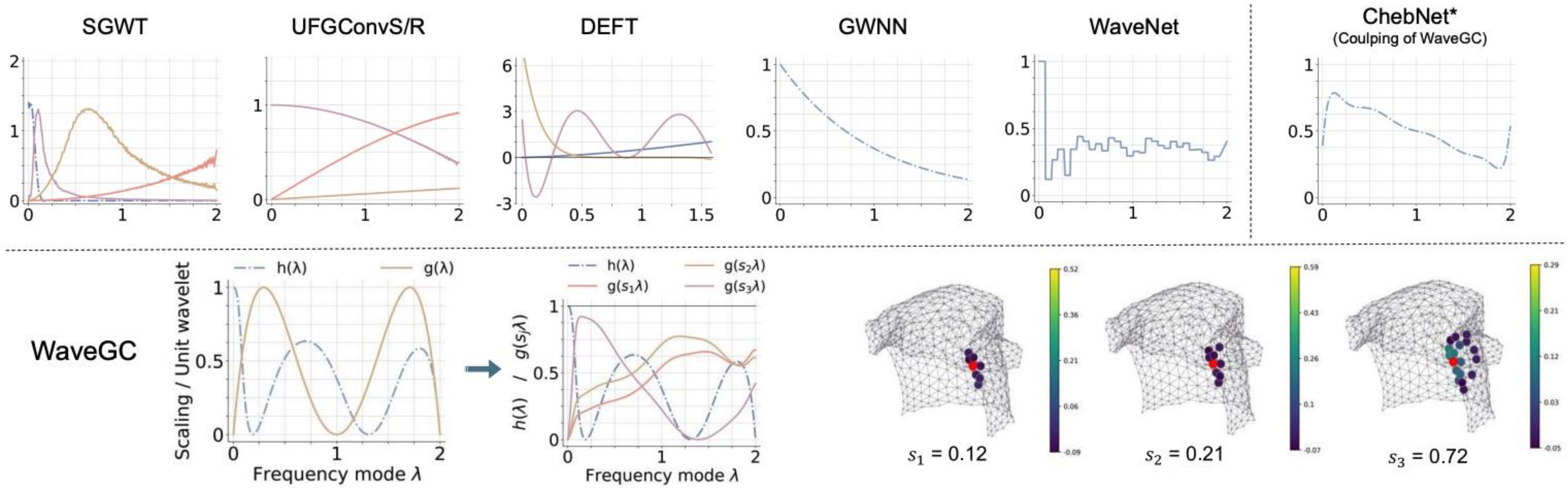


Figure 2. The spectral and spatial visualization of different bases on PascalVOC-SP.

IMPACT OF THE LEARNED SCALES

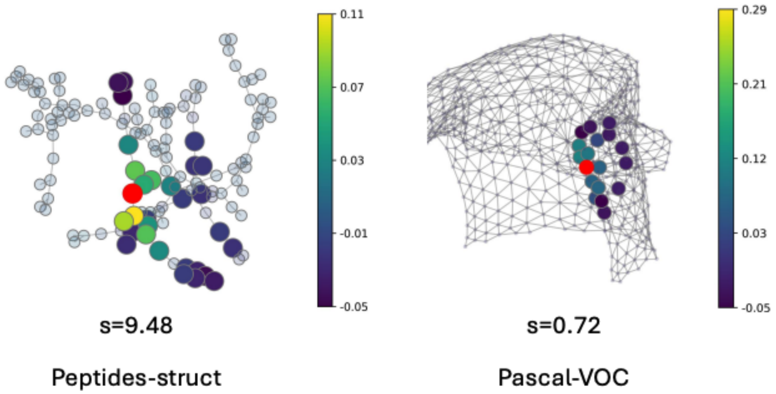


Figure 6. Visualizations of receptive fields for Peptides-struct (Ps) and Pascal-VOC (VOC) at their largest scale s.

Table 10. Comparison of average and max receptive fields of Ps and VOC.

	Peptides-struct	Pascal-VOC
Avg. Receptive Field	3.02	0.74
Max Receptive Field	9	3
Avg. Shortest Path	20.89	10.74

IMPACT OF THE LEARNED SCALES

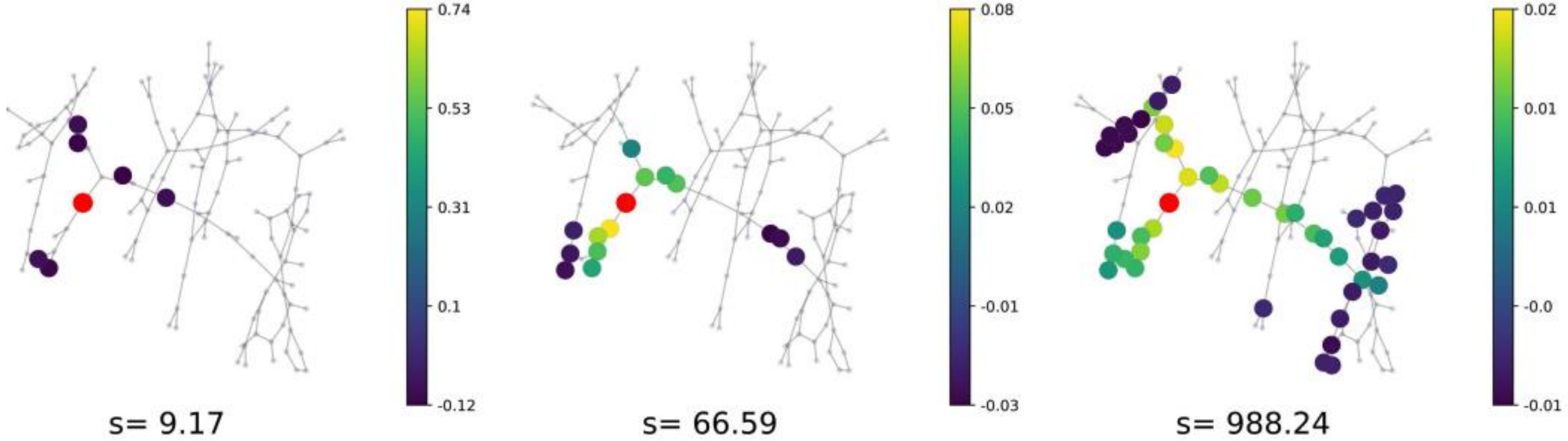


Figure 7. Visualizations of receptive fields for Peptides-func (Pf) at extreme scales.

Thanks for Listening!