FPT Software AICenter | McGill University | UAB

Range-aware Positional Encoding via High-order Pretraining: Theory and Practice

Viet Anh Nguyen ♠ Khang Nhat Ngo ♣ Truong Son Hy♢,∗

♠FPT AI Center, Vietnam ♣McGill University, Canada ♢The University of Alabama at Birmingham, United States

∗: Correspondence to thy@uab.edu

Introduction

Based on Wavelet Positional Encoding of Ngo et al. (2023), we propose HOPE-WavePE (High-Order Permutation Equivariant Wavelet Positional Encoding) a novel pre-training strategy for positional encoding that is equivariant under the permutation group and is sensitive to the length and diameter of graphs downstream tasks. Since our approach relies solely on the graph structure, it is domain-agnostic and adaptable to datasets from various domains, therefore paving the wave for developing general graph structure encoders and graph foundation models. We theoretically demonstrate that such equivariant pretraining schema can approximate the training target for abitrarily small tolerance. We also evaluate HOPE-WavePE on graph-level prediction tasks of different areas and show its superiority compared to other methods

1 Method Overview

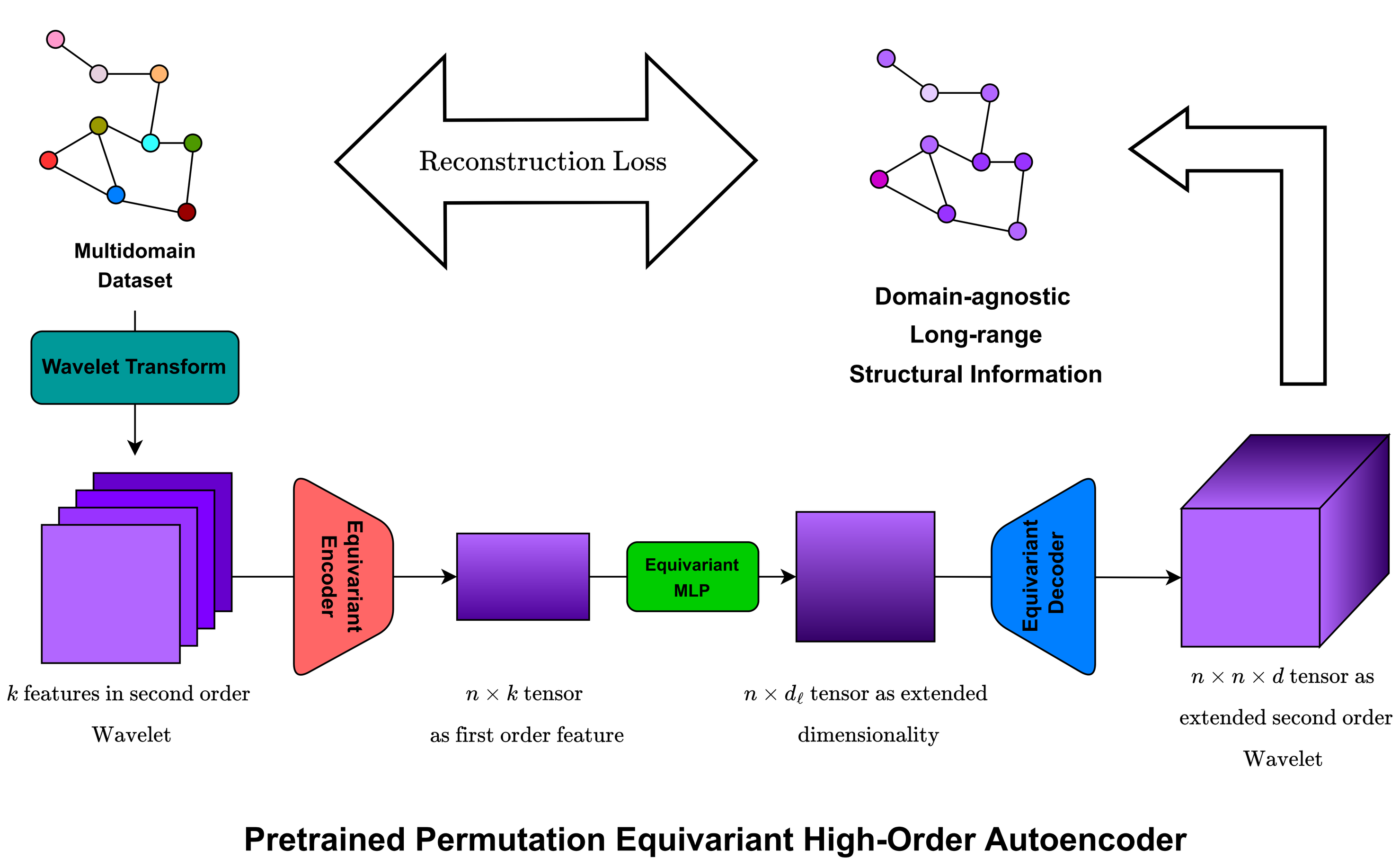

Figure 1: Our proposed high-order autoencoder can learn domain-agnostic information.

We inherited the Wavelet PE [1] to capture multiresolutional feature on graphs. Afterwards, we propose a novel encoding-decoding scheme that collects the most important features on Wavelet. The output is reflected with the masked adjacencies via the loss function

$$\mathcal{L} = \sum_{i=1}^{r} \text{BinCrossEntropy}\left(\mathbf{M}_i \odot \mathbf{A}_{s_i}, \mathbf{M}_i \odot \widehat{\mathbf{A}}_{s_i}\right).$$

where $\mathbf{A}_{s_i}$ and $\widehat{\mathbf{A}}_{s_i}$ are ground truth and predicted adjacency matrices respectively.

Our method is domain-agnostic as there is only the adjacency matrix is needed as input, ignoring domain features on downstream graph tasks. Moreover, our novel masking technique not only greatly reduces bias to certain hop lengths during training which enable long-range learning, but also how this simple mathematical scheme can filter out out-of-diameter hop lengths for shorter graphs. This latter part enables our method **range-awareness**. On top of all, HOPE-WavePE is also designed to follow the **permutation equivariance** constraint, which is inherently crucial on graphs.

2 Visualizations

Figure 2: Patterns learnt by HOPE-WavePE,nodes with resembling colors indicate highly-correlated node features.

3 Theoretical Results

Theorem 3.1. *For any $\epsilon > 0$ and real coefficients $\theta_1, \theta_2, \ldots, \theta_d$, there exists a HOPE-WavePE parametrization $\varphi : \mathbb{R}^{n\times n\times d} \to \mathbb{R}^{n\times n}$ such that*

$$\left\| \varphi(\mathbf{Z}) - \sum_{j=1}^{r} \theta_j \mathbf{A}_{s_j} \right\| < \epsilon.$$

4 Experiments

Method	MUTAG	PROTEINS	NCI1	NCI109	IMDB-B	IMDB-M
RWK	79.2 ± 2.1	59.6 ± 0.1	>3 days	-	-	-
GK ($k = 3$)	81.4 ± 1.7	71.4 ± 0.3	62.5 ± 0.3	62.4 ± 0.3	-	-
PK	76.0 ± 2.7	73.7 ± 0.7	82.5 ± 0.5	-	-	-
WL kernel	90.4 ± 5.7	75.0 ± 3.1	**86.0** ± 1.8	-	73.8 ± 3.9	50.9 ± 3.8
DCNN	-	61.3 ± 1.6	56.6 ± 1.0	-	49.1 ± 1.4	33.5 ± 1.4
DGCNN	85.8 ± 1.8	75.5 ± 0.9	74.4 ± 0.5	-	70.0 ± 0.9	47.8 ± 0.9
IGN	83.9 ± 13.0	76.6 ± 5.5	74.3 ± 2.7	72.8 ± 1.5	72.0 ± 5.5	48.7 ± 3.4
GIN	89.4 ± 5.6	76.2 ± 2.8	82.7 ± 1.7	-	75.1 ± 5.1	52.3 ± 2.8
PPGNs	90.6 ± 8.7	**77.2** ± 4.7	83.2 ± 1.1	**82.2** ± 1.4	73.0 ± 5.8	50.5 ± 3.6
Natural GN	89.4 ± 1.6	71.7 ± 1.0	82.4 ± 1.3	-	73.5 ± 2.0	51.3 ± 1.5
GSN	**92.2** ± 7.5	76.6 ± 5.0	83.5 ± 2.0	-	**77.8** ± 3.3	**54.3** ± 3.3
SIN	-	76.4 ± 3.3	82.7 ± 2.1	-	**75.6** ± 3.2	**52.4** ± 2.9
CIN	**92.7** ± 6.1	**77.0** ± 4.3	**83.6** ± 1.4	**84.0** ± 1.6	**75.6** ± 3.7	**52.7** ± 3.1
GIN + HOPE-WavePE (ours)	**93.6** ± 5.8	**79.5** ± 4.81	**84.5** ± 2.0	**84.1** ± 1.9	**76.0** ± 3.7	**52.7** ± 2.9

Table 1: Experiments on TU datasets.

5 Conclusions

We have introduced HOPE-WavePE, a novel high-order permutation-equivariant pretraining method specifically designed for graph-structured data. Our approach leverages the inherent connectivity of graphs, eliminating reliance on domain-specific features. This enables HOPE-WavePE to generalize effectively across diverse graph types and domains. The superiority of HOPE-WavePE is demonstrably proven through both theoretical and empirical analysis. Finally, we have discussed the potential of HOPE-WavePE as a foundation for a general graph structural encoder. A promising future direction will be to focus on optimizing the scalability of this approach.

References

[1] Nhat Khang Ngo, Truong Son Hy, and Risi Kondor. Multiresolution graph transformers and wavelet positional encoding for learning long-range and hierarchical structures. *The Journal of Chemical Physics*, 159(3):034109, 07 2023.