Supplementary Material for HodgeFormer : Transformers for Learnable Operators on Triangular Meshes through Data-Driven Hodge Matrices

Akis Nousias
K3Y Labs
anousias@k3y.bg

Stavros Nousias
Technical University of Munich
stavros.nousias@tum.de

A. End-to-end Architecture

A.1. Multi-head Hodge Attention on vertices, edges and faces

This section presents a detailed analysis of the HodgeFormer attention mechanism, demonstrating how the multi-head Hodge Attention component operates across different mesh elements. By combining the Hodge star operator formulation (Eq. 11) with the attention-based operators (Eq. 5), the Hodge Laplacians L_v, L_e, and L_f for vertices (0-forms), edges (1-forms), and faces (2-forms) are reformulated as the attention-based expressions given in equations S1, S2, and S3, respectively.

$$
\begin{aligned}
L_v &:= \star_0^{-1}(x_v) \cdot d_0^T \cdot \star_1(x_e) \cdot d_0 \\
&:= \sigma\left(\frac{Q_v K_v^T}{\sqrt{d_h}}\right) \cdot d_0^T \cdot \sigma\left(\frac{Q_e K_e^T}{\sqrt{d_h}}\right) \qquad \text{(S1)} \\
L_e &:= d_0 \cdot \star_0^{-1}(x_v) \cdot d_0^T \cdot \star_1(x_e) + \\
&\quad \star_1^{-1}(x_e) \cdot d_1^T \cdot \star_2(x_f) \cdot d_1 \\
&:= d_0 \cdot \sigma\left(\frac{Q_v K_v^T}{\sqrt{d_h}}\right) \cdot d_0^T \cdot \sigma\left(\frac{Q_{e1} K_{e1}^T}{\sqrt{d_h}}\right) + \\
&\quad \sigma\left(\frac{Q_{e2} K_{e2}^T}{\sqrt{d_h}}\right) \cdot d_1^T \cdot \sigma\left(\frac{Q_f K_f^T}{\sqrt{d_h}}\right) \cdot d_1 \qquad \text{(S2)} \\
L_f &:= d_1 \cdot \star_1^{-1}(x_e) \cdot d_1^T \cdot \star_2(x_f) \\
&:= d_1 \cdot \sigma\left(\frac{Q_{e2} K_{e2}^T}{\sqrt{d_h}}\right) \cdot d_1^T \cdot \sigma\left(\frac{Q_f K_f^T}{\sqrt{d_h}}\right) \qquad \text{(S3)}
\end{aligned}
$$

The updated features for each mesh element are computed by applying the respective Hodge Laplacian to the corresponding value vectors:

$$
\begin{aligned}
x_v &= L_v \cdot V_v \quad \text{(vertex features)} \qquad \text{(S4)} \\
x_e &= L_e \cdot V_e \quad \text{(edge features)} \qquad \text{(S5)} \\
x_f &= L_f \cdot V_f \quad \text{(face features)} \qquad \text{(S6)}
\end{aligned}
$$

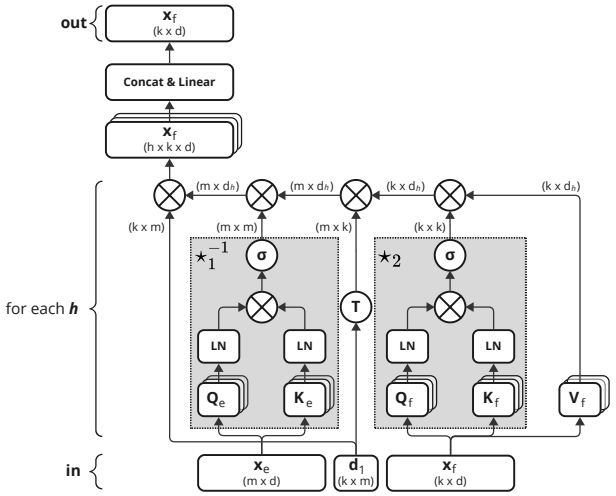


Figure S1. Multi-head Hodge Attention applied to latent face features x_f. The multi-head attention mechanism learns data-driven Hodge Star matrices $\star_1^{-1}$ and $\star_2$.

This formulation ensures that the attention mechanism respects the topological structure of the mesh while enabling information flow between different dimensional elements (vertices, edges, faces). The formation of these updated features is illustrated in Figures S2, S3, and S1, which visualize the computational flow described in equations (S4), (S5), and (S6), respectively.

A.2. Computational complexity and memory requirements

Consider a HodgeFormer layer operating on vertex features x_v by applying the operator $L_v := \star_0^{-1}(x_v) \cdot d_0^T \cdot \star_1(x_e) \cdot d_0$. Also, let x_v and x_e have dimensions (n_v, d) and (n_e, d) respectively, and assume that for practical applications $n_e \approx 3n_v$ via Euler's formula on meshes. The layer performs:

(a) Application of corresponding maps W_Q, W_K and W_V on x_v and x_e with complexity $O(n_v \cdot d^2)$ and $O(n_e \cdot d^2)$.

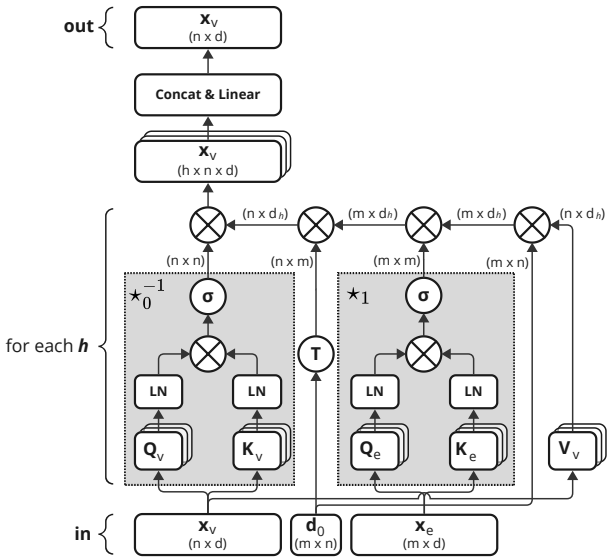


Figure S2. Multi-head Hodge Attention applied to latent vertex features x_v. The multi-head attention mechanism learns data-driven Hodge Star matrices $\star_0^{-1}$ and $\star_1$.

(b) Sparse-dense matrix multiplication of V_v with d_0 and d_0^T of dimension (n_e, n_v) with complexity $O(n_e d)$ where n_e is the number of nonzero entries of d_0.

(c) Computation of Hodge matrices $\star_0^{-1}(x_v)$ and $\star_1(x_e)$, based on the product QK^T between elements of dimensions (n, d) and $(\sqrt{n}, d)$ with complexity $O(n_v^{1.5} \cdot d)$ and $O(n_e^{1.5} \cdot d)$ respectively. The operation is performed via gather operations and the full matrix is not materialized.

(d) Multiplication of Hodge matrices with feature vector V_v results in complexity $O(n_v^{1.5} \cdot d)$ and $O(n_e^{1.5} \cdot d)$.

The resulting complexity is $O(nd^2) + O(n^{1.5}d)$ with d fixed and usually much smaller than n. In comparison, the eigendecomposition of a sparse (n, n) matrix has complexity $O(kn^2)$ for calculating the first k eigenvectors. This holds for Laplacian positional embeddings as well as for spectral features such as Heat Kernel Signature (HKS). Tab. S1 presents experiments evaluating the computational and memory requirements.

Mesh Size (n_v)	2^8	2^{10}	2^{12}	2^{14}
Compute Time (ms)				
HF Encoder (Train)	5.78	8.98	29.72	197.9
HF Encoder (Infer)	2.50	3.42	10.96	66.13
HF Layer (Infer)	1.08	1.91	9.42	55.25
Peak Memory Usage (GBs)				
HF Encoder (Train)	0.12	0.40	2.54	19.23
HF Encoder (Infer)	0.09	0.31	2.08	15.89
HF Layer (Infer)	0.09	0.31	2.08	15.89

Table S1. Metrics for different mesh sizes with respect to the number of vertices, for a 1-layer HodgeFormer (HF) end-to-end architecture (training and inference) as well as a standalone HodgeFormer layer (inference). The model operates on vertices with a latent embedding dimension $d = 256$ and hidden MLP dimension of $d_h = 512$. Measured on an Nvidia RTX 4090 GPU.

B. Results

B.1. Robustness Analysis

To further evaluate the robustness of the HodgeFormer architecture, we perform an analysis, where we evaluate on meshes with added noise, different topology, removed triangles as well as on incomplete meshes. Specifically, we produce variants of the *Human* dataset's test set as follows:

- **Gaussian Noise:** We add Gaussian Noise (GN) ϵ, of several levels $\lambda \in \{0.005, 0.01, 0.02\}$ calculated w.r.t. the diagonal of the axis-aligned bounding box of each model, i.e. $\epsilon = \lambda \cdot |BB_{diag}|$.
- **QEM Remeshing:** We use the models of the original *Human* dataset from [4] and remesh them using the Quadratic Error Metric (QEM) to different target face resolution, i.e. $[1000, 2000]$.
- **Face Removal:** We assign to each face a probability p to be randomly removed, and produce dataset variants with different probabilities in the range $[0.01, 0.20]$.
- **Patch removal:** We assign to each face a probability $p = 0.005$ to be randomly selected and remove a large patch of k neighbors where $k \sim \mathcal{U}(8, 15)$.

For all methods, with the exception of QEM Remeshing, we use the test set of the *Human* dataset. Ground truth for evaluation is defined via nearest-neighbor on the original meshes. We evaluate the pre-trained networks reported in Tab. 5, on the mutated datasets and report the results in Tab. S2. The HodgeFormer architecture retains well its performance under noise addition, remeshing, removed faces and added holes.

Incomplete Meshes: Additionally, we select two mesh models from the test set of the Human Body Dataset, one with accurate and one with inaccurate segmentation results, and gradually remove body parts from them, similar to [1]. Then, we produce segmentations using a pre-trained HodgeFormer network as showcased in Fig. S5. The segmentation results remain reasonable and consistent with the initial predictions of the network on these mesh models, even on the one with inaccurate segmentation.

B.2. Training on One Sample per Class

We repeat the experiment performed by [5], where a model is trained on the 30-class *SHREC11* [3] dataset on splits of only one sample per class (split-1). The model follows the same

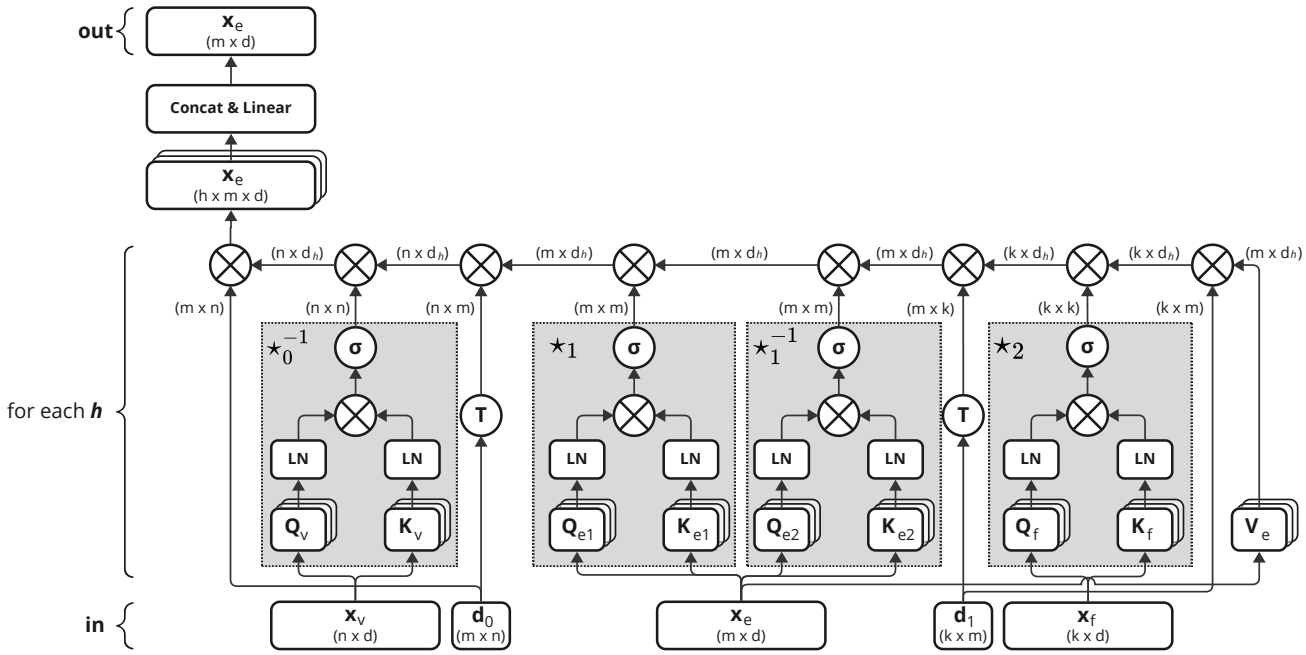


Figure S3. Multi-head Hodge Attention applied to latent edge features x_e. The multi-head attention mechanism learns data-driven Hodge Star matrices $\star_0^{-1}$ and $\star_1$.

HUMAN Test Set Variant	Accuracy	Acc.Drop
Original	90.3%	n/a
Gaussian Noise ($\lambda = 0.005$)	88.3%	2.0%
Gaussian Noise ($\lambda = 0.010$)	87.3%	3.0%
Gaussian Noise ($\lambda = 0.020$)	81.1%	9.2%
QEM Remesh ($1000F$)	87.2%	3.1%
QEM Remesh ($2000F$)	86.2%	4.1%
Face Removal ($p = 0.04$)	87.8%	2.5%
Face Removal ($p = 0.10$)	85.9%	4.4%
Face Removal ($p = 0.20$)	81.6%	8.7%
Patch removal	86.8%	3.5%

Table S2. Robustness evaluation of the HodgeFormer architecture. We report the performance of pre-trained networks on mutated variations of the *Human* dataset test set. The HodgeFormer architecture retains well its performance under noise addition, remeshing, removed faces and added holes.

Method	SHREC11 (split-10)	SHREC11 (split-1)
SubDivNet [2]	99.5%	36.5%
MWFormer [5]	100.0%	41.7%
HodgeFormer (ours)	98.7%	**45.9%**

Table S3. Classification performance of different methods on the 30-class *SHREC11*[3] dataset trained on splits of only 1 sample per class (split-1). Results for (split-10) are also reported for comparison.

architecture as Sec.5 with a learning rate of 0.0025 to compensate for the small dataset size. The results are presented in Tab. S3 along with (split-10) results for comparison.

B.3. Variations Study on Mixing Strategies

This section examines the effect of mixing HodgeFormer and Transformer layers on segmentation performance. Table S4 summarizes the performance on the *COSEG* Vases dataset under varying layer compositions. We denote N_H as the number of HodgeFormer layers and $R_{H:T} = (N_H : N_T)$ the ratio of the number of HodgeFormer layers to Vanilla Transformer layers. As shown in Table S4, carefully mixing Transformer with HodgeFormer layers has an effect on the final model performance. In general, from the conducted experiments, different mixing strategies would fit different datasets. For the *COSEG* Vases the best result (93.04%) was obtained by adding two Transformer layers on top of the HodgeFormer layers. Additional layouts of HodgeFormer and Transformer layers are showcased in Sec. B.6.

B.4. Variation Study on Label Smoothing

This section examines the effect of label smoothing in the cross-entropy loss on the model's performance for classification and segmentation tasks. Table S5 reports classification results on *SHREC11* and segmentation results on *COSEG* Vases under different label smoothing values.

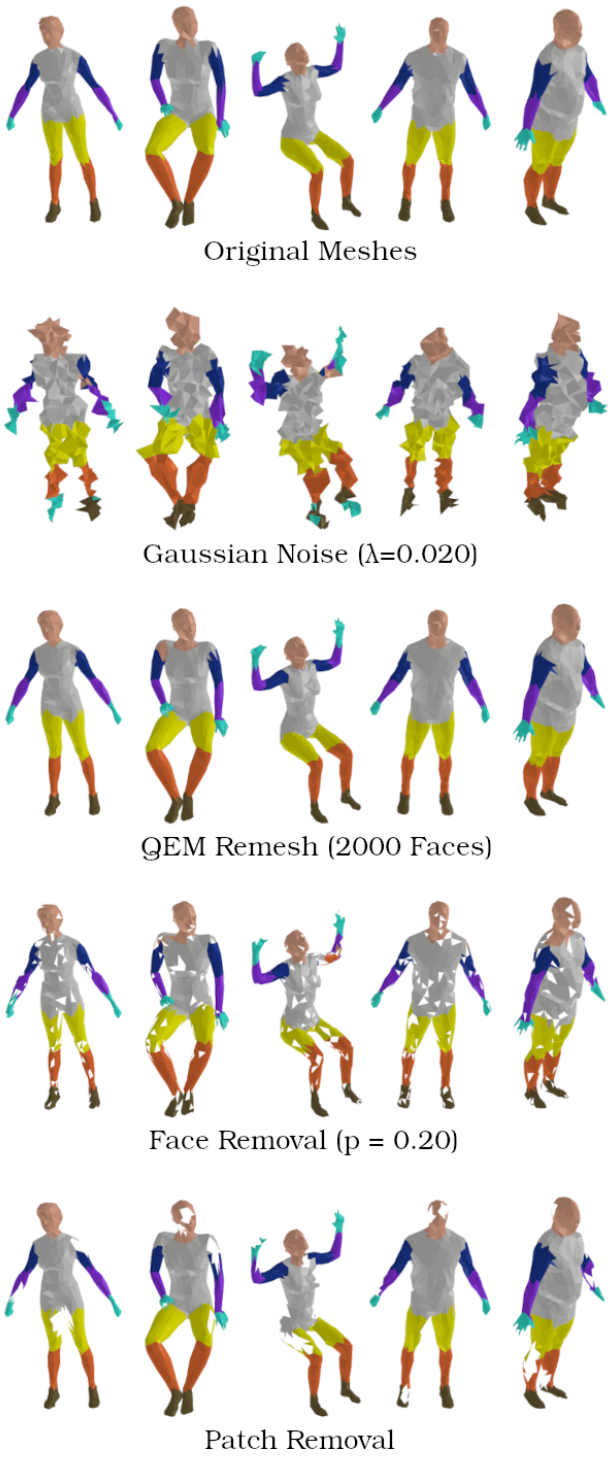


Figure S4. Mesh segmentation results on selected models from the mutated Human Body test set. The segmentation results remain reasonable and consistent with the initial predictions of the network.

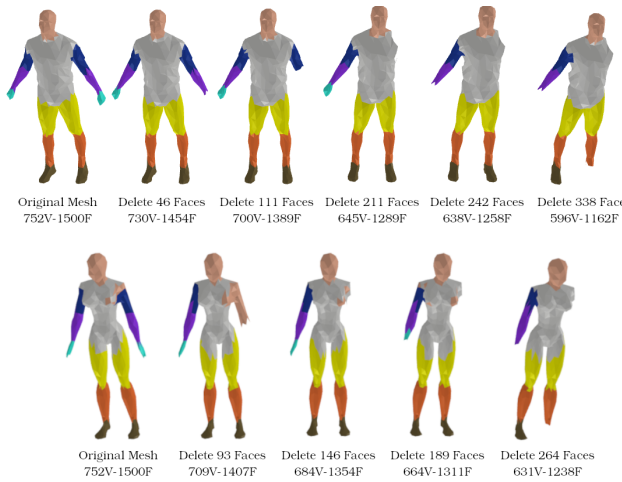


Figure S5. Mesh segmentation results on selected models from the mutated Human Body test set. The segmentation results remain reasonable and consistent with the initial predictions of the network.

Layers	**Vases (COSEG)**
$N_H = 4$ & $R_{(N_H,N_T)} = 4:0$	92.02%
$N_H = 4$ & $R_{(N_H,N_T)} = 1:1$	92.85%
$N_H = 4$ & $R_{(N_H,N_T)} = 2:1$	92.59%
$N_H = 4$ & $R_{(N_H,N_T)} = 4:2$	93.04%

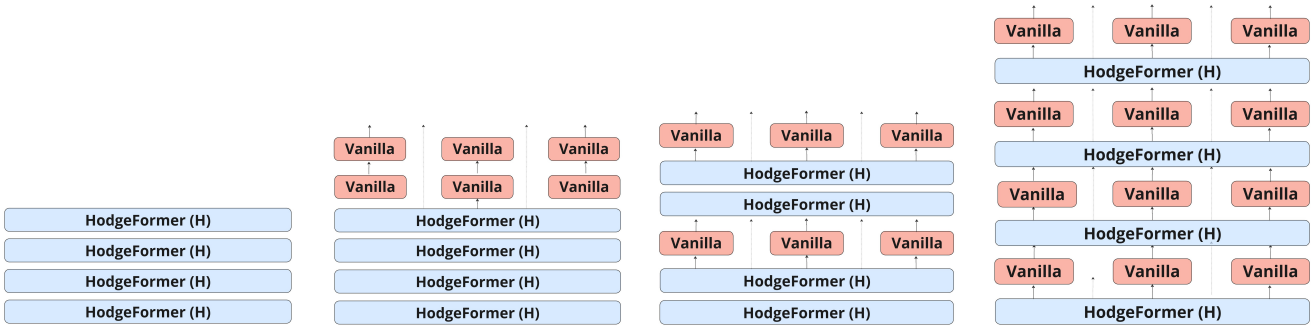
Table S4. Variation experiments comparing different strategies of mixing HodgeFormer with Transformer layers and their effect on the performance, evaluated on the *COSEG* Vases dataset.

Label Smoothing	**SHREC11** (split-10)	**Vases** (COSEG)
0.00	97.50%	94.03%
0.05	98.33%	94.08%
0.10	98.67%	94.12%
0.20	98.70%	94.30%
0.40	98.33%	93.25%

Table S5. Variation study on the effect of label smoothing on the model's performance for classification and segmentation tasks.

B.5. Eigenfunctions of the corresponding learned operator

We compute the eigenfunctions of the corresponding learned operators and visualize them on the mesh geometry, providing insights into how the HodgeFormer layer captures and processes geometric features at different scales and orientations. Figure S7 at the top, presents the learned attention maps for two sample meshes of *COSEG* Aliens dataset measured in layer 6 of the architecture. Figure S7 at the bottom presents the first five eigenfunctions of a *COSEG* Aliens dataset sample.



(a) (b) (c) (d)

Figure S6. Visualization of mixing strategies of HodgeFormer and vanilla Transformer layers used in Sec. B.3. N_H corresponds to the number of HodgeFormer transformer layers and $R_{(N_H,N_T)}$ corresponds to the ratio of HodgeFormers to vanilla transformers: a) $N_H = 4$ & $R_{(N_H,N_T)} = 4:0$, b) $N_H = 4$ & $R_{(N_H,N_T)} = 4:2$, c) $N_H = 4$ & $R_{(N_H,N_T)} = 2:1$, d) $N_H = 4$ & $R_{(N_H,N_T)} = 1:1$

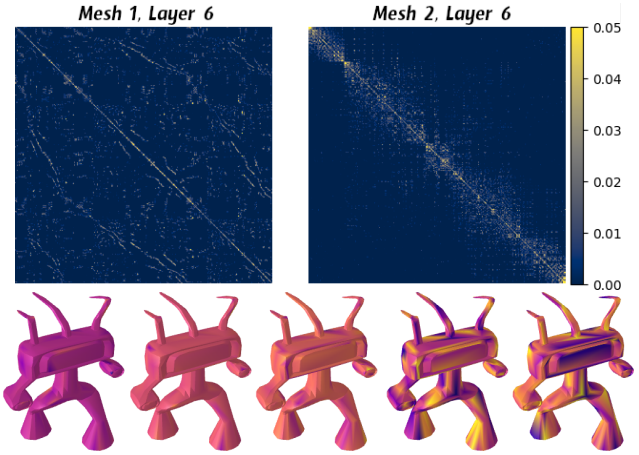
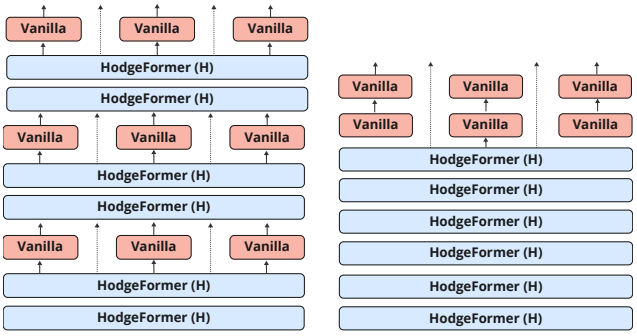


Figure S7. Qualitative illustration. Above: Learned attention maps from two samples of the COSEG Aliens dataset. Below: First few eigenfunctions of the corresponding learned operator, $L_6, head_1$ for the 1st sample.



(a) (b)

Figure S8. HodgeFormer architecture variations. For *Human* dataset, our highest performing model was a 6-layered HodgeFormer model mixed with 2 Transformer layers in a $2:1$ ratio, as presented in (a), whereas for *Vases*, *Chairs* and *Aliens* datasets, our highest performing model was a 6-layered HodgeFormer model followed by two Transformer layers, as shown in (b).

B.6. Hodgeformer & Transformer Layer Layouts

This section includes visualizations of HodgeFormer architecture variations and interleaving strategies, along with notation clarifications. N_H corresponds to the number of HodgeFormer transformer layers and $R_{(N_H,N_T)}$ corresponds to the ratio of HodgeFormers to vanilla transformers.

References

[1] Qiujie Dong, Zixiong Wang, Junjie Gao, Shuangmin Chen, Zhenyu Shu, and Shiqing Xin. Laplacian2mesh: Laplacian-based mesh understanding. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):492–502, 2023. 2

[2] Shi-Min Hu, Zheng-Ning Liu, Meng-Hao Guo, Jun-Xiong Cai, Jiahui Huang, Tai-Jiang Mu, and Ralph R Martin. Subdivision-based mesh convolution networks. *ACM Transactions on Graphics (TOG)*, 41(3):1–16, 2022. 3

[3] Z. Lian, A. Godil, B. Bustos, M. Daoudi, J. Hermans, S. Kawamura, Y. Kurita, G. Lavoué, H. V. Nguyen, R. Ohbuchi, Y. Ohkita, Y. Ohishi, F. Porikli, M. Reuter, I. Sipiran, D. Smeets, P. Suetens, H. Tabia, and D. Vandermeulen. Shrec'11 track: Shape retrieval on non-rigid 3D watertight meshes. In *Eurographics Workshop on 3D Object Retrieval, EG 3DOR*, pages 79–88, 2011. 2, 3

[4] Haggai Maron, Meirav Galun, Noam Aigerman, Meirav Trope, Nadav Dym, Ersin Yumer, Vladimir G Kim, and Yaron Lipman. Convolutional neural networks on surfaces via seamless toric covers. *ACM Transactions on Graphics (TOG)*, 36(4):1–10, 2017. 2

[5] Hao-Yang Peng, Meng-Hao Guo, Zheng-Ning Liu, Yong-Liang Yang, and Tai-Jiang Mu. MWFormer: Mesh Understanding with Window-based Transformer. *Computers & Graphics*, 115: 382–391, 2023. Publisher: Elsevier BV. 2, 3