A Appendix

In this section, we provide more details on experimental settings, several additional experiments, and the visual interpretation of the results of the ablation experiments: attention types.

A.1 Experimental Settings

Table 6: **Experimental settings.** S denotes the segmentation task, C denotes the classification task.

Configuration	value
device	2 NVIDIA GeForce RTX 3090 GPUs
optimizer	AdamW
learning rate	$5e$-4
weight decay	$5e$-2
learning rate schedule	cosine
warmingup epochs	10
augmentation	ScaleAndTranslate
batch size	32(C), 16(S)
number of points	$1,024$(C), $2,048$(C, S), $8,192$(C)
number of patches	64(C), 128(C, S), 512(C)
patch size	32
training epochs	300
loss function	Cross-Entropy(C), Negative Log-Likelihood(S)

A.2 Additional Experiments

In this subsection, we evaluate the performance of rotation robustness of the proposed methods, and some additional experimental results.

Table 7: **Additional experiments on rotation robustness.** We test the rotation robustness of Point-GT on classification tasks for ModelNet40. The models are trained on the training set without rotation and evaluated with two rotation settings. Therein, Z denotes the test set that is randomly rotated along the Z-axis and $SO(3)$ denotes the test set with random rotation followed by $SO(3)$.

Methods	Z	$SO(3)$
Transformer[45]	27.1	11.6
Point-GT-G (Ours)	**29.1**	11.5
Point-GT-DM (Ours)	28.9	**12.1**

Table 8: **Additional shape classification results on SHREC 15 [22].**We report the type of input feature and the accuracy (%). SHREC 15 is a non-rigid shape dataset with 1,200 shapes from 50 categories. Shapes in SHREC15 are 2D surfaces embedded in 3D space. We follow a ratio of seven to three to split the training and validation sets. During the evaluation, we randomly sampled 2,048 points from each shape as input. All the results of the proposed approach illustrated are evaluated without voting.

Methods	Input Feature	Accuracy
PointNet++[28]	XYZ	60.18
GeoNet[15]	XYZ	94.67
Transformer[45]	XYZ	82.00
Point-GT-G (Ours)	XYZ	85.43
Point-GT-DM (Ours)	XYZ	83.43
Point-MAE[26]	XYZ	90.29
Point-GT-MAE (Ours)	XYZ	**97.14**

A.2.1 Comparison Between Graph-based GSA and Data Manifold-based GSA

Since there are some differences in results between the graph-based GSA and the data manifold-based GSA, we give explanations from two aspects: methodology and visual interpretation experiments.

Methodology First, for the data with complex structures and missing data points, e.g., ScanObjectNN dataset, Table 2 shows that graph-based GSA performs worse than manifold-based GSA. This phenomenon may be probably because manifold-based GSA captures the better geodesic features for the data with noise.

Secondly, for the noise-free data, e.g., ShapeNetPart dataset, the computing of geodesic is easier to calculate. Graph-based GSA and manifold-based GSA have similar results, illustrated in Table 4. The difference in the results possibly results from the incidental bias of the geodesic distances computing.

Thirdly, for datasets with small sample sizes, such as few-shot ModelNet40 and SHREC 15, the graph-based GSA seems to show a more robust performance, e.g., Table 3 and Table 8. This probably denotes that the data manifold-based method is less efficient for learning the geodesic features of small datasets.

Finally, compared with manifold-based GSA, the shortest path algorithm of graph-based GSA is non-differentiable. In the paper, graph-based GSA first utilizes the k-NN to compute the local neighborhood, and then Floyd's shortest path algorithm is used to estimate the geodesic distance of all pairs of the points. This process is separated and non-differentiable in the whole model and has complex computing. Although the powerful performance of PyTorch and Autograd makes the network work successfully, some imprecise gradient propagation may lead to worse results. As for the data manifold-based GSA, the projection manifold, Oblique Manifold, is differentiable, which can avoid such a dilemma. Overall, since the differentiability often plays an important role in the global optimization, this possibly leads to the difference in results.

Visual interpretation In Section A.3 and Fig. 4, we have illustrated the attention heat maps of the several attention types mentioned in the ablation experiments.

A.3 Visual Interpretation

In this subsection, we show a visual interpretation of the differences in performance of the attention types mentioned in ablation experiments in Fig. 4. In the caption, the "Dot." denotes the standard dot-product self-attention, and the "Dot.-less" denotes the "Dot." without linear transformation. The results of the quantitative ablation experiments are presented in Section 4.4 in the paper.

The Fig. 4 indicates the following results:

- The Dot-less self-attention ("Dot.-less") gives a better focus area than standard dot-product self-attention ("Dot."). This confirms our assumption that a general linear transformation may lead to reducing the ability of geometric structure feature extraction of the point cloud, i.e., a general linear transformation reduces the performance of the self-attention mechanism.
- The proposed GSA methods give more precise focus areas than the "Dot." and the "Dot.-less" methods. This confirms our motivation that similar Euclidean distances can not reveal the geodesic distances and geometric semantics. In short, the dot product multiplication in Euclidean space is insufficient to capture accurate internal non-Euclidean structures of point cloud objects.
- The graph-based GSA and the data manifold-based GSA show similar attention patterns, which confirms the validity of our proposed methods.

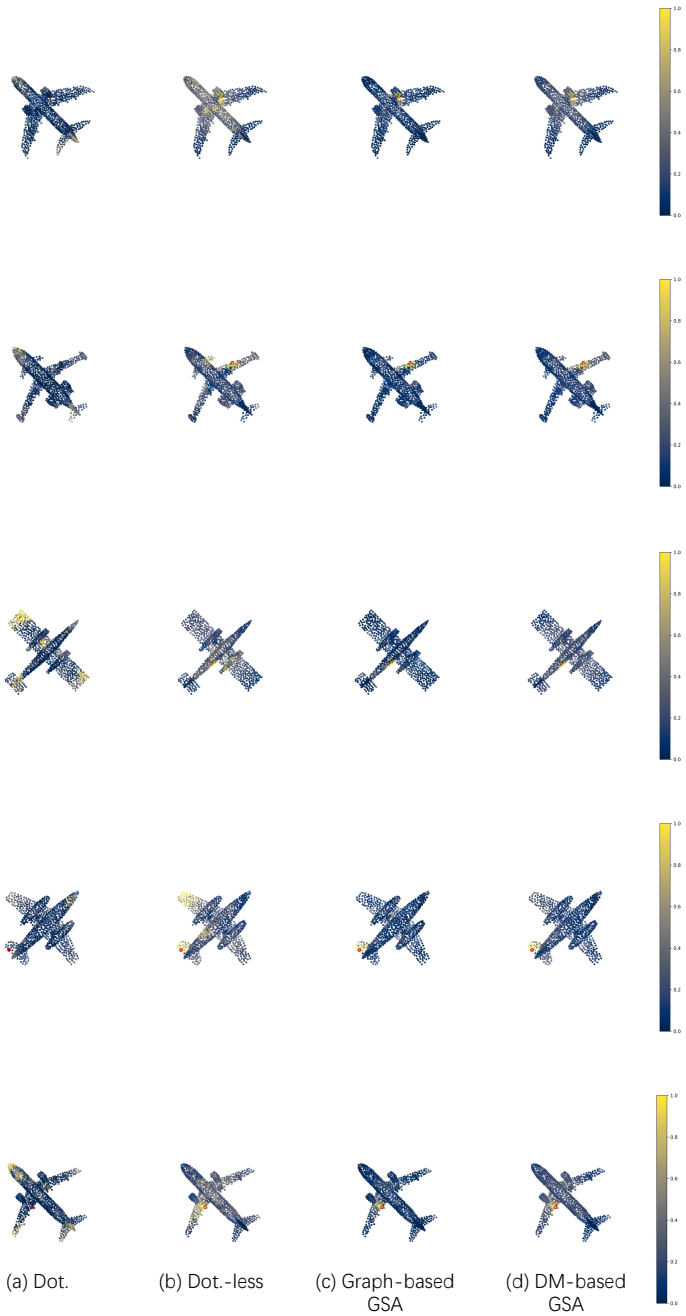

Figure 4: **Attention Heat Maps.** From left to right, attention heat maps of (a) Dot., (b) Dot.-less, (c) Graph-based GSA and (d) DM (Data Manifold)-based GSA are illustrated. The red point is the query point, and a higher attention score means a higher attention weight.