Methods	Pretrained dataset	# Points	Acc.	Methods	Pretrained dataset	# Points	Acc.
LatentGAN (**?**)	SN	2k	85.7	FoldingNet (**?**)	MN	2k	84.4
FoldingNet (**?**)	SN	2k	88.4	LatentGAN (**?**)	MN	2k	87.3
PointCapsNet (**?**)	SN	2k	88.9	PointCapsNet (**?**)	MN	1k	87.5
VIPGAN (**?**)	SN	2k	90.2	Multi-task (**?**)	MN	2k	89.1
STRL (**?**)	SN	2k	90.9	MAP-VAE (**?**)	MN	2k	90.2
SSC (RSCNN) (**?**)	SN	2k	92.4	GraphTER (**?**)	MN	1k	92.0
CrossPoint (**?**)	SN	2k	91.2	GLR (RSCNN) (**?**)	MN	1k	92.2
DHGCN (DGCNN)	SN	2k	**93.2**	DHGCN (DGCNN)	MN	1k	93.0
DHGCN (AdaptConv)	SN	2k	**93.2**	**DHGCN (AdaptConv)**	MN	1k	**93.3**

Table 1: Classification results of *unsupervised* methods (including ours) on ModelNet40. 'SN/MN' denotes 'ShapeNet/ModelNet40' and '# Points' indicates the point number in pretraining.

Methods	Limited training data ratios				
	1%	2%	5%	10%	20%
FoldingNet (**?**)	56.4	66.9	75.6	81.2	83.6
MAE3D (**?**)	61.7	69.2	80.8	84.7	88.3
DHGCN	**62.7**	**72.2**	**81.3**	**86.1**	**89.1**

Table 2: Comparison results of 3D object classification with limited training data (different ratios) on ModelNet40. DGCNN is taken as the backbone.

achieves SOTA results in all 5 training data ratios, demonstrating that the features learned by our self-supervised task can be easily generalized to the point cloud classification task even with limited training data.

Classification on real-world dataset ScanObjectNN. We also conduct the classification experiment on the real-world scanning dataset ScanObjectNN (**?**), which poses great challenges for point cloud classification methods due to the involved cluttered background, noisy perturbations and occluded incomplete data. This dataset contains 15 categories, totally $2,902$ unique object instances. Here we follow the official data split strategy on three dataset variants: OBJ_ONLY, OBJ_BG and PB_T50_RS, and conduct pertaining on ShapeNet.

As shown in Table 3, DHGCN achieves the best results of 85.9% on OBJ_BG variant, exceeding all compared SOTA unsupervised methods by at least 4.2%. Our DHGCN even achieves comparable results with supervised methods, e.g., surpassing the DGCNN backbone by 3.1% on OBJ_BG. As for the PB_T50_RS variant, our method achieves an accuracy of 81.9%, slightly lower than the other two variants. We suspect that the perturbation noise in PB_T50_RS disturbs the adjacency relationship between parts (i.e., some points are incorrectly split into adjacent parts during voxelization), thus degrading the power of the learned hop distance in depicting point cloud intrinsic geometric structure.

Furthermore, despite being pretrained on the ShapeNet dataset (synthetic data), our downstream classification results on the real-world ScanObjectNN reveal that the learned geometric information is useful in mitigating the domain gap between synthetic and real-world data.

Methods	Sup.	OBJ_ONLY	OBJ_BG	PB_T50_RS
PointNet (**?**)	✓	79.2	73.3	68.2
PointNet++ (**?**)	✓	84.3	82.3	77.9
PointCNN (**?**)	✓	85.5	86.1	78.5
DGCNN (**?**)	✓	86.2	82.8	78.1
Point-BERT (**?**)	✓	88.1	87.4	83.1
Point-MAE (**?**)	✓	88.3	90.0	85.2
Jigsaw (**?**)	✗	-	59.5	-
OcCo (**?**)	✗	-	78.3	-
STRL (**?**)	✗	-	77.9	-
CrossPoint (**?**)	✗	-	81.7	-
DHGCN	✗	**85.0**	**85.9**	**81.9**

Table 3: Classification results of our method and state-of-the-art methods on ScanObjectNN. DGCNN is used as backbone. 'Sup.' denotes the method is supervised (✓) or unsupervised (✗). Results of Jigsaw, OcCo, and STRL are from CrossPoint, and "-" indicates no previous results.

Part segmentation on ShapeNet Part. We evaluate DHGCN for the shape part segmentation task on ShapeNet Part dataset (**?**), which contains $16,881$ models from 16 categories. Each model involves 2 to 6 parts, with a total number of 50 distinct part labels. Following PointNet, we sample or interpolate each model to $2,048$ points and only use point coordinates as input.

We use mean Intersection-over-Union (mIoU) as the evaluation metric, and two types of mIoU are reported in Table 4. Our self-supervised method achieves SOTA performance, which exceeds all recent unsupervised methods.

Part segmentation with limited data. We *freeze* the pretrained model and randomly sample 1% and 5% of the train set of ShapeNet Part to train several MLPs for evaluating the segmentation task with the unsupervised paradigm. Results shown in Table 5 demonstrate that our DHGCN using PAConv as backbone achieves SOTA performance, i.e., 76.9% with 1% training data and 81.9% with 5% training data for instance mIoU, which exceeds recent unsupervised methods. These results reveal that the features learned by our self-supervised hop distance reconstruction task are more expressive than other unsupervised methods in terms of part segmentation under extremely limited training data.

Please refer to supplementary for more results.

Methods	Sup.	Class mIOU	Instance mIOU
PointNet (**?**)	✓	80.4	83.7
PointNet++ (**?**)	✓	81.9	85.1
DGCNN (**?**)	✓	82.3	85.2
KPConv (**?**)	✓	85.1	86.4
PAConv (**?**)	✓	84.2	86.0
Point-BERT (**?**)	✓	84.1	85.6
LatentGAN (**?**)	✗	57.0	-
MAP-VAE (**?**)	✗	68.0	-
GrpahTER (**?**)	✗	78.1	81.9
CTNet (**?**)	✗	75.5	79.2
DHGCN	✗	**82.9**	**84.9**

Table 4: Shape part segmentation results of our method and state-of-the-art techniques on ShapeNet Part dataset. PAConv is used as backbone. 'Sup.' denotes the method is supervised learning (✓) or unsupervised learning (✗).

Methods	Limited training data ratios	
	1%	5%
SO-Net (**?**)	64.0	69.0
PointCapsNet (**?**)	67.0	70.0
Multi-task (**?**)	68.2	77.7
PointContrast (**?**)	74.0	79.9
SSC (RSCNN) (**?**)	74.1	80.1
DHGCN	**76.9**	**81.9**

Table 5: Comparison results of shape part segmentation with limited training data (different ratios) on ShapeNet Part. PAConv is taken as the backbone.

4.4 Ablation Studies

Attention mechanism. We conduct an ablation study for several model settings to verify the HGA's effectiveness, which embeds the learned hop distance matrix into edge weights. We denote the SA option as our baseline, whose switch factor λ is set to 0 in both HGA layers. HGA will degrade to general SA in this case, and the hop distance loss is disabled. Note that we train this baseline in a *supervised* manner. We compare the effect of whether the hop distance loss is calculated in each layer or only the last layer. Accuracy results of point cloud classification and hop distance prediction are reported in Table 6.

With the aid of HGA, the classification accuracy significantly exceeds that of SA baseline by 0.5%. In addition, calculating the hop distance loss in each layer leads to both higher hop distance prediction accuracy (94.6% versus 93.3%) and point cloud classification accuracy (93.3% versus 93.1%). The strong supervision of our hop distance loss leads to a more accurate learned hop distance matrix, thus producing better performance.

Gaussian kernel. The predicted hop distance will be processed by the Gaussian kernel $\mathbb{G}$ to provide more attention to neighboring parts (i.e., smaller distances yield higher weights and vice versa). The parameter σ^2 of $\mathbb{G}$ controls the decay rate between distance and edge weight. A small σ^2 causes the edge weights between remote parts to decay rapidly. For example, when $\sigma^2 = 0.2$, the weights of parts over 1-hop distance converge to 0. At this point, the receptive field degrades to 1-hop (i.e., local neighbors), resulting in a lower accuracy of 92.6%, as shown in Table 7. On the contrary, with a larger σ^2, the weights decay gently as the distance increases, enabling nodes to contribute more equally. However, this leads to a reduction in distinction due to distance, achieving only 93.0% ($\sigma^2 = 2.0$) and 92.9% ($\sigma^2 = 5.0$). The model achieves the highest accuracy of 93.3% when $\sigma^2 = 1.0$.

Attention	Sup.	Loss	Distance Acc.	Acc.
SA	✓	$\mathcal{L} = \mathcal{L}_c$	-	92.8
HGA	✗	$\mathcal{L} = \mathcal{L}_h^{(-1)}$	93.3	93.1
HGA	✗	$\mathcal{L} = \sum_l^L \mathcal{L}_h^{(l)}$	**94.6**	**93.3**

Table 6: Different attention mechanisms. Experiments are conducted on ModelNet40 with AdaptConv as the backbone. SA denotes self-attention, and HGA denotes Hop Graph Attention. Accuracy results of hop distance prediction and point cloud classification are reported.

Gaussian kernel					
σ^2	0.2	0.5	**1.0**	2.0	5.0
Acc.	92.6	93.2	**93.3**	93.0	92.9

Table 7: Ablation results on different σ^2 in the Gaussian kernel. Experiments are conducted on ModelNet40 for classification with AdaptConv as the backbone.

5 Conclusion

This paper proposes a novel self-supervised part-level hop distance reconstruction task and a novel hop distance loss to learn contextual relationships between point parts. The dynamically updated hop distances are embedded as attention weights by the proposed HGA for determining point parts' importance in feature aggregation. Our DHGCN can be easily incorporated into point-based backbones. We outperform SOTA unsupervised methods on both downstream classification and part segmentation tasks. Our model is less effective for data with large perturbations as noise leads to less accurate splitting of parts, which tends to produce misleading adjacent relationships. This will be explored in future.

Acknowledgments

This work is supported in part by the National Key Research and Development Program of China (Grant Number 2022ZD04014) and the Shaanxi Province Key Research and Development Program (Grant Number 2022QFY11-03).

provident iusto quisquam illo assumenda.Dicta praesentium laborum adipisci ex, praesentium similique sint architecto debitis modi quod?Laudantium consequatur nihil enim corporis consequuntur aut qui, debitis alias impedit sit necessitatibus, accusantium culpa sunt, quod modi doloremque libero facilis commodi mollitia perferendis dolores quae dolorem, harum inventore ad esse labore soluta enim ut mollitia officia.