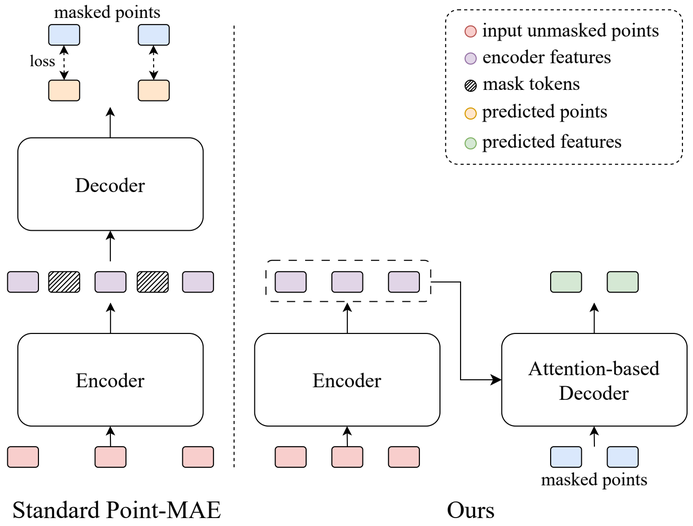
3D Feature Prediction for Masked-AutoEncoder-Based Point Cloud Pretraining

TEXAS The University of Texas at Austin

Siming Yan, Yuqi Yang, Yuxiao Guo, Hao Pan, Peng-shuai Wang, Xin Tong, Yang Liu, Qixing Huang

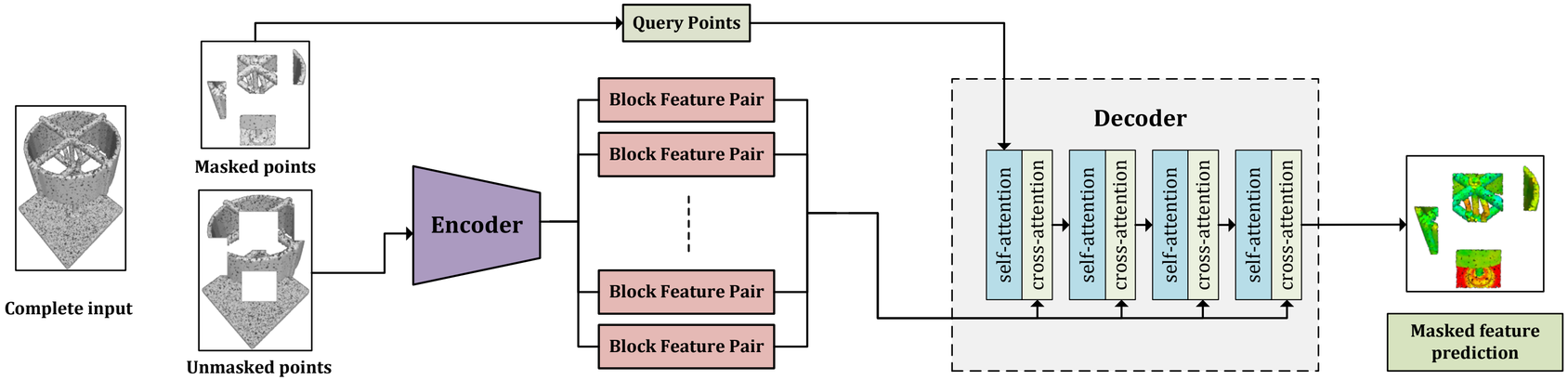
Microsoft

Introduction



This paper introduces a novel masked autoencoder based point cloud pre-trianing methd **MaskFeat3D.** Instead of recovering the positions of points, this method focuses on recovering features like surface normals and variations using a unique attention-based decoder. This decoder operates independently from the encoder. We tested the effectiveness of our approach and decoder design with various encoder types for point cloud training and showed that our pre-trained networks perform well on different point cloud analysis tasks.

Method



Feature Visualization

Point Cloud Point Normal Surface Variation

Experiment Results

Method	ScanObjectNN			ShapeNetPart	
	OBJ-BG	OBJ-ONLY	PB-T50-RS	ins. mIoU	cls. mIoU
PointViT	79.9	80.6	77.2	85.1	83.4
MinkowskiNet	84.1	86.1	80.1	85.3	83.2
PointNeXt	91.9	91.0	88.1	87.1	84.7
MaskFeat3D (PointViT)	91.7(91.6)	90.0(89.6)	87.7(87.5)	86.3(86.3)	84.9(84.8)
MaskFeat3D (MinkowskiNet)	85.1(85.0)	87.0(86.7)	80.8(80.6)	85.6(85.5)	83.5(83.5)
MaskFeat3D (PointNeXt)	**92.7**(92.6)	**92.0**(91.9)	**88.6**(88.5)	**87.4**(87.4)	**85.5**(85.5)

Method	MN40 Linear
OcCo	89.6
PointBERT	87.4
PointMAE	88.5
MaskFeat3D	**91.1**

Method	Backbone	ScanNet		SUN RGB-D	
		AP_{25}	AP_{50}	AP_{25}	AP_{50}
MaskFeat3D	VoteNet	63.3	41.0	61.0	36.5
MaskFeat3D	Point-M2AE	67.5	50.0	-	-
MaskFeat3D	CAGroup3D	**75.6**	**62.3**	**67.2**	**51.0**