Visual Computing Institute
Computer Vision
Prof. Dr. Bastian Leibe

RWTH AACHEN UNIVERSITY

Point2Vec for Self-Supervised Representation Learning on Point Clouds

Karim Abou Zeid* Jonas Schult* Alexander Hermans Bastian Leibe
RWTH Aachen University *equal contribution
{abouzeid,schult,hermans,leibe}@vision.rwth-aachen.de

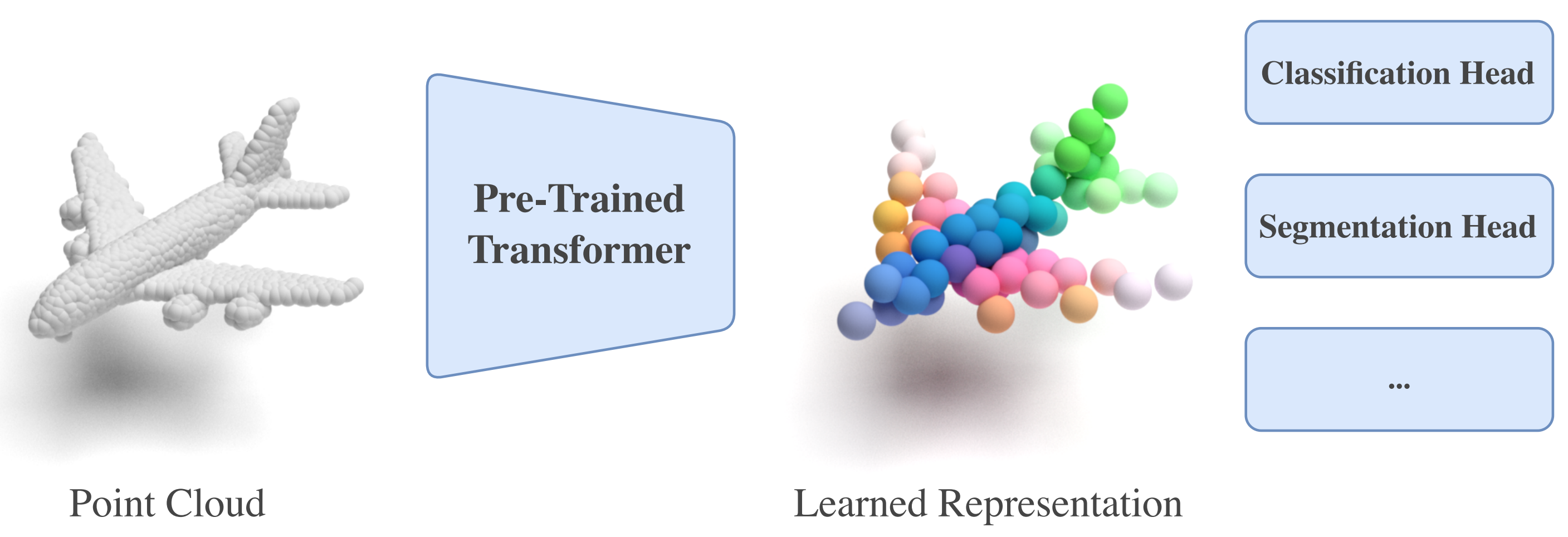
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Abstract

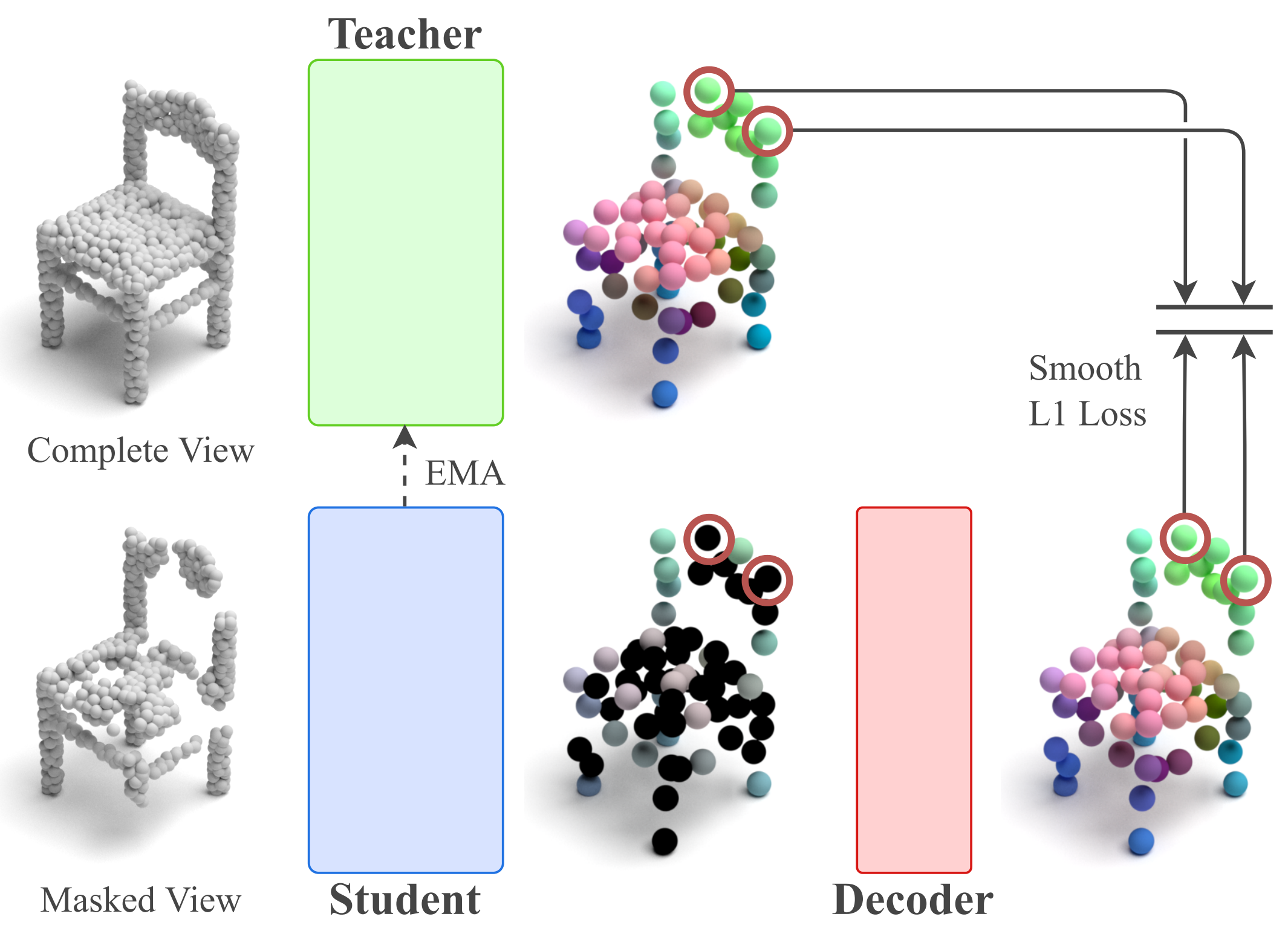
We extend data2vec to the point cloud domain and show promising results on several downstream tasks. However, our analysis reveals that disclosing positional information can expose the object's overall shape to the student, which hinders data2vec from learning strong representations. To address this 3D-specific shortcoming, we propose point2vec, which unleashes the full potential of data2vec-like pre-training on point clouds. Our experiments show that point2vec outperforms other self-supervised Transformer-based methods on shape classification on ModelNet40 and ScanObjectNN. Our results suggest that the learned representations are both transferable and strong.

Representation Learning on Point Clouds

- Self-supervised learning of representations from unlabeled point clouds.
- Learned representations can be used for downstream tasks such as classification, segmentation, etc.

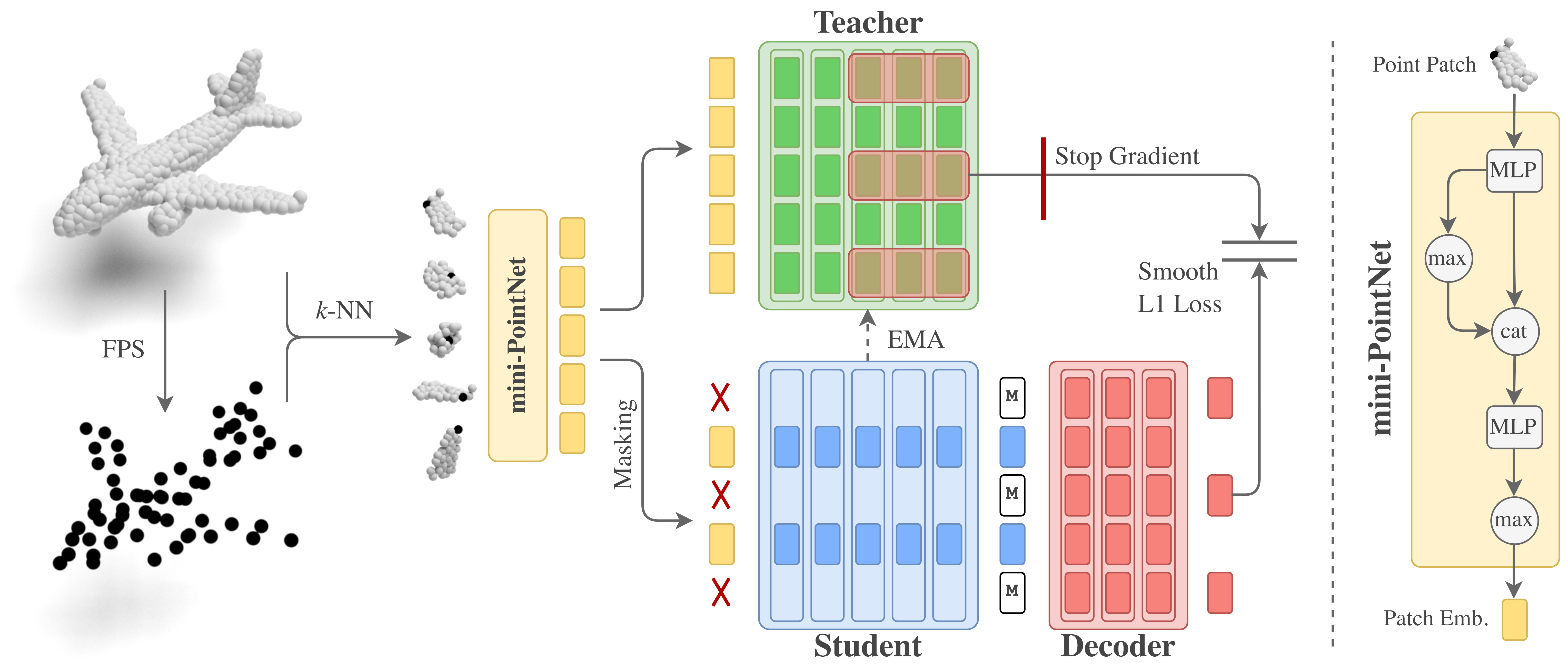


High-Level Overview of Point2Vec

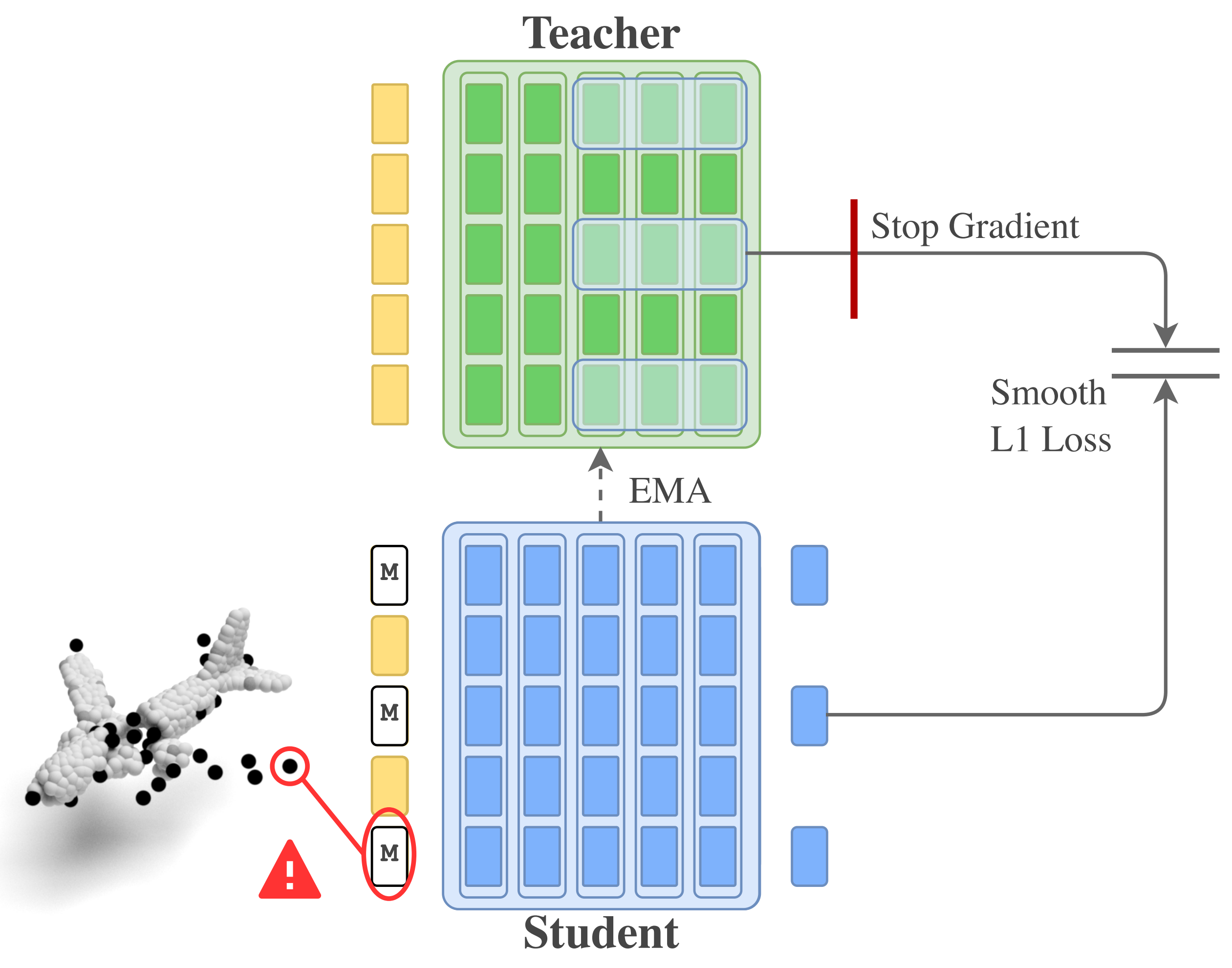


A teacher network 🟩 predicts latent representations using a complete view of the point cloud. A student network 🟦 predicts the same representations, but from a partially masked view. A shallow decoder 🟥 then reconstructs the latent representations of the masked regions ●. The student and the decoder are optimized, whereas the teacher is an exponential moving average of the student.

Our Point2Vec Pre-Training Method

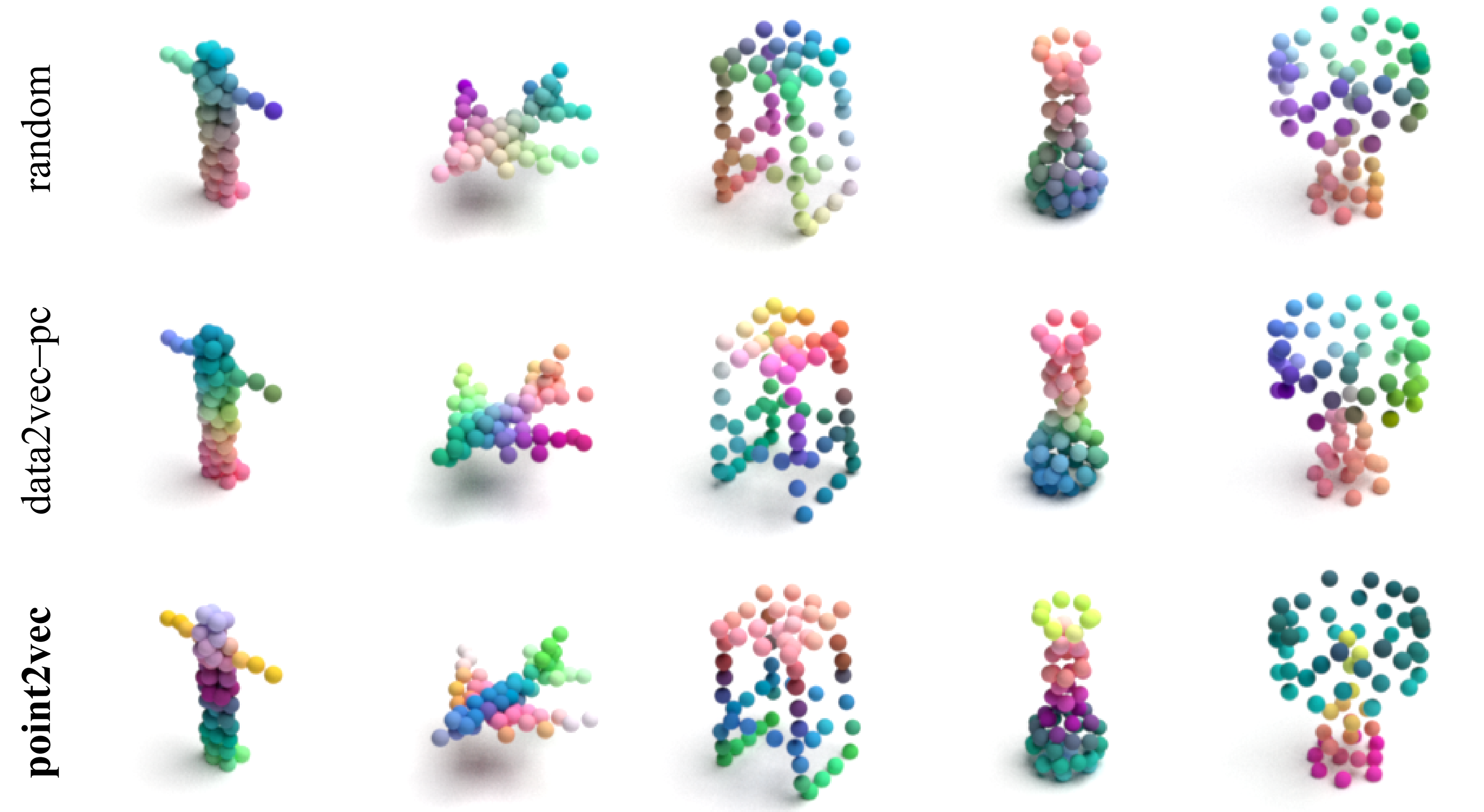


Leakage of Positional Information in Data2Vec–pc



When applying data2vec to point cloud data, the positional information of the mask embeddings reveal the overall shape of the point cloud to the student.

Qualitative Results (PCA projection)

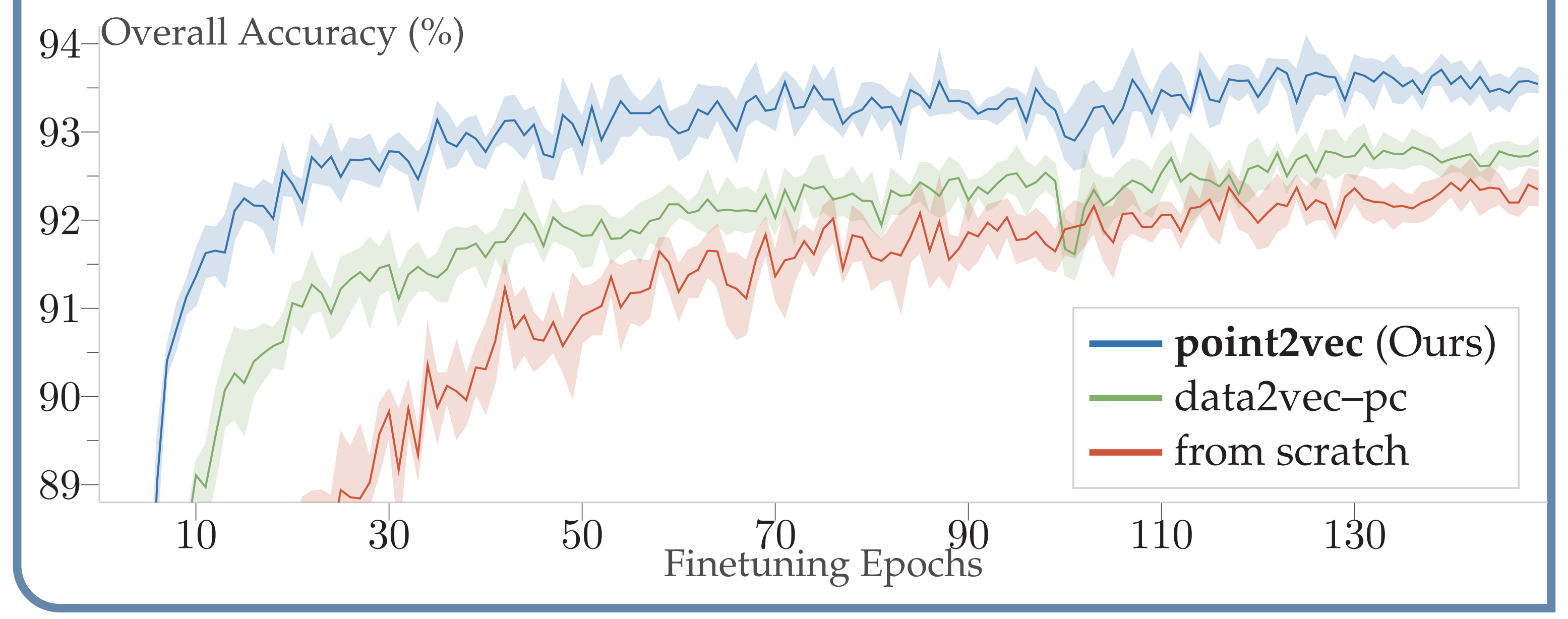


We use PCA to project the learned representations into RGB space. Both a random initialization and data2vec–pc pre-training show a fairly strong positional bias, whereas point2vec exhibits a stronger semantic grouping without being trained on downstream dense prediction tasks.

Results

	Classification (Overall Acc.)		Part Seg. (mIoU$_I$)
	ModelNet40	**ScanObjNN**	**ShapeNetPart**
Point-BERT	93.2	83.1	85.6
MaskPoint	93.8	84.6	86.0
Point-MAE	93.8	85.2	86.1
Point-M2AE	94.0	86.4	**86.5**
from scratch	93.3	84.3	85.7
data2vec–pc	93.6 +0.3	85.5 +1.2	85.9 +0.2
point2vec (Ours)	**94.8** +1.2	**87.5** +2.0	86.3 +0.4

Fine-Tuning Learning Curve (ModelNet40)



Pretext Task and Dataset Ablation

	Overall Accuracy	
Pretext Task	**ModelNet40**	**ScanObjNN**
none, *i.e.* training from scratch	93.3	84.3
classification (ShapeNet)	93.2	82.9
point2vec (ModelNet40)	93.9	84.4
point2vec (ShapeNet)	**94.8**	**87.5**

Conclusion

- Point2vec is a self-supervised representation learning approach which unleashes the full potential of data2vec-like pre-training on point clouds.
- It achieves remarkable results on various downstream tasks, surpassing other self-supervised learning approaches in shape classification as well as few-shot learning on well-established benchmarks.

More Information

vision.rwth-aachen.de/point2vec