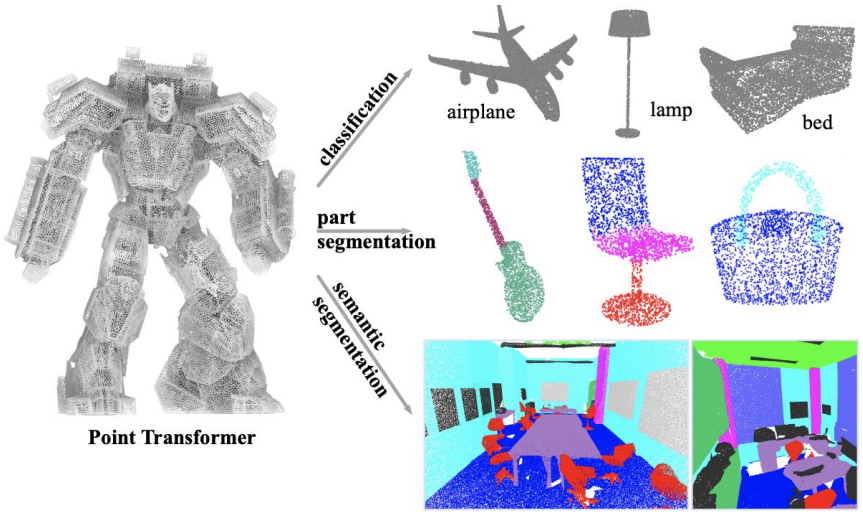
Point Transformer



Authors: Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip Torr, Vladlen Koltun

Presented by: Connor Vines, Daniel Zhang, Han Lin

Arguments

Arguments

1. Can be used as a backbone for a variety of tasks (paper shows segmentation and classification) not just object detection like the other paper.
2. Performance: Results & Inference Speed, and applications
3. Can be more easily built on (Point transformer v2 and v3 and Stratified Transformer for 3D Point Cloud Segmentation) & versatility

Argument #1 - Technical Novelty

A general Transformer architecture for 3D point clouds that can be applied to multiple tasks including semantic segmentation, object classification, and object parts segmentation.

Argument #2 - Performance

Table 1: Semantic Segmentation on S3DIS

Method	OA	mAcc	mIoU
HPEIN	87.2	68.3	61.9
MinkowskiNet	-	71.7	65.4
KPConv	-	72.8	67.1
PointTransformer	**90.8**	**76.5**	**70.4**

Table 2: Shape Classification on ModelNet40

Method	mAcc	OA
Point2Sequence	90.4	92.6
KPConv	-	92.9
InterpCNN	-	93.0
PointTransformer	**90.6**	**93.7**

Table 3: Object Part Seg. on ShapeNetPart

Method	cat. mIoU	ins. mIoU
PointCNN	84.6	86.1
InterpCNN	84.0	86.3
KPConv	**85.1**	86.4
PointTransformer	83.7	**86.6**

Method	ScanNetV2		SUN RGB-D	
	AP_{25}	AP_{50}	AP_{25}	AP_{50}
BoxNet*	49.0	21.1	52.4	25.1
VoteNet*	60.4	35.5	58.3	33.4
3DETR-m	65.0	47.0	59.0	32.7
H3DNet	**67.2**	**48.1**	**60.1**	**39.0**
HGNet	-	-	**61.6**	**34.4**
BRNet	**66.1**	**50.9**	**61.1**	**43.7**
GroupFree3D	**69.1**	**52.8**	**63.0**	**45.2**
FCAF3D	**71.5**	**57.3**	**64.2**	**48.9**

Argument #3 - Research Impact

Point transformer [PDF] thecvf.com

H Zhao, L Jiang, J Jia, PHS Torr… - Proceedings of the …, 2021 - openaccess.thecvf.com

… We show that **Point Transformers** are remarkably effective in 3D deep learning tasks, both at … In particular, **Point Transformers** set the new state of the art on large-scale semantic …

☆ Save 99 Cite Cited by 1410 Related articles All 8 versions »

An end-to-end transformer model for 3d object detection [PDF] thecvf.com

I Misra, R Girdhar, A Joulin - Proceedings of the IEEE/CVF …, 2021 - openaccess.thecvf.com

… Furthermore, we show **3DETR** is applicable to 3D tasks … Unlike DETR, **3DETR** does not employ a ConvNet backbone, … be easily incorporated in **3DETR** to further improve its performance…

☆ Save 99 Cite Cited by 369 Related articles All 6 versions »

Point Transformer V3: Simpler, Faster, Stronger [PDF] arxiv.org

X Wu, L Jiang, PS Wang, Z Liu, X Liu, Y Qiao… - arXiv preprint arXiv …, 2023 - arxiv.org

… Therefore, we present **Point Transformer V3** (… **point** clouds organized with specific patterns. This principle enables significant scaling, expanding the receptive field from 16 to 1024 **points** …

☆ Save 99 Cite Cited by 2 Related articles All 2 versions »

An End-to-End Transformer Model for 3D Object Detection

Ananya, Andrew, Kaan, Sizhe

Arguments

Argument 1: More Efficient in Learning from Sparse Data

- 3DETR is designed to efficiently learn from sparse 3D point cloud data. Its architecture, draws from the Transformer model, is optimized to handle the inherent sparsity and irregularity of point clouds.
- Because 3DETR doesn’t need dense sampling/complex network layers, it can directly process sparse 3D point cloud data, making it more efficient
- This contrasts with Point Transformer, which may rely more on dense sampling or intricate network layers to achieve its performance
-

Argument 2: Less model Complexity and Overhead

- Point Transformer's specialized self-attention mechanisms come with increased model complexity and computational overhead, compared to 3DETR which avoids complex 3D-specific operations
- Minimal modifications to the vanilla Transformer block, utilizing standard Transformer components with non-parametric queries keep 3DETR's architecture straightforward
- This computational efficiency makes the model easier to deploy in the real world, compared to models requiring extensive 3D domain customization

Argument 3: More Versatile

- 3DETR's design philosophy emphasizes minimal modifications to the Transformer architecture, aiming for a model that not only excels in 3D object detection but also possesses inherent generalizability across different datasets and potentially across different 3D understanding tasks.
- This contrasts with Point Transformer, whose task-specific optimizations, while effective, may limit its ability to perform equally well across a broader range of datasets or when applied to tasks slightly different from those it was originally designed for.