PointClustering: Unsupervised Point Cloud Pre-training using Transformation Invariance in Clustering

Fuchen Long, Ting Yao, Zhaofan Qiu, Lusong Li and Tao Mei

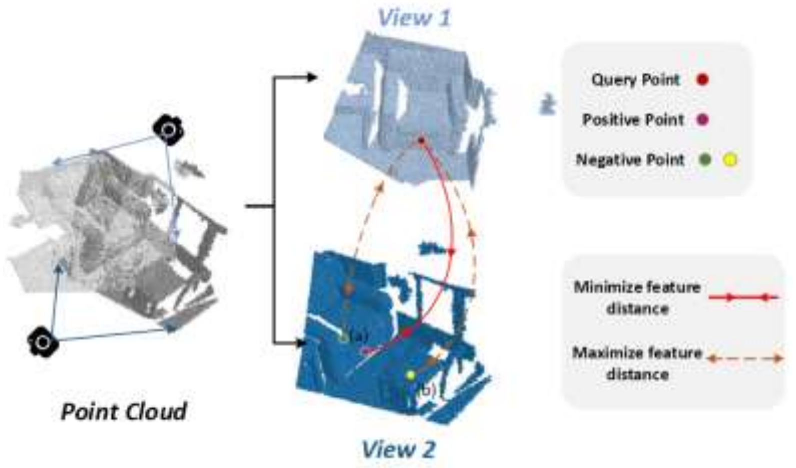
HiDream.ai Inc

Poster ID: 116

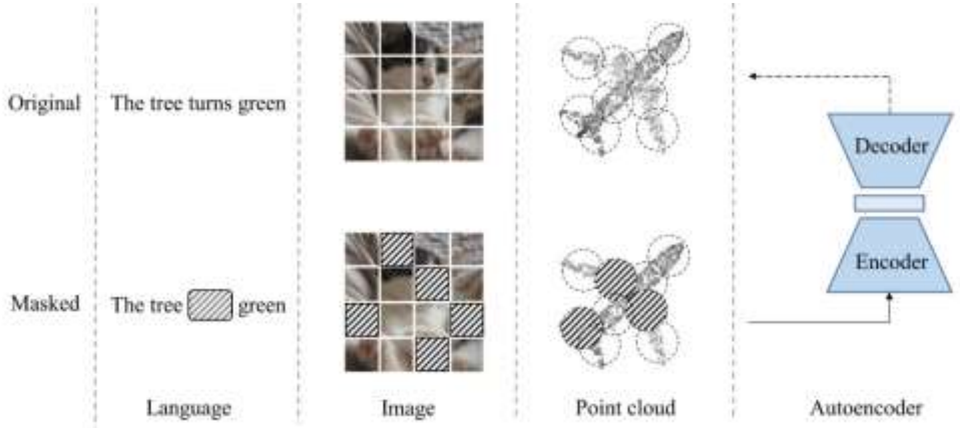
Paper Tag: THU-PM-116

New Point Cloud Pre-training Paradigm: PointClustering

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

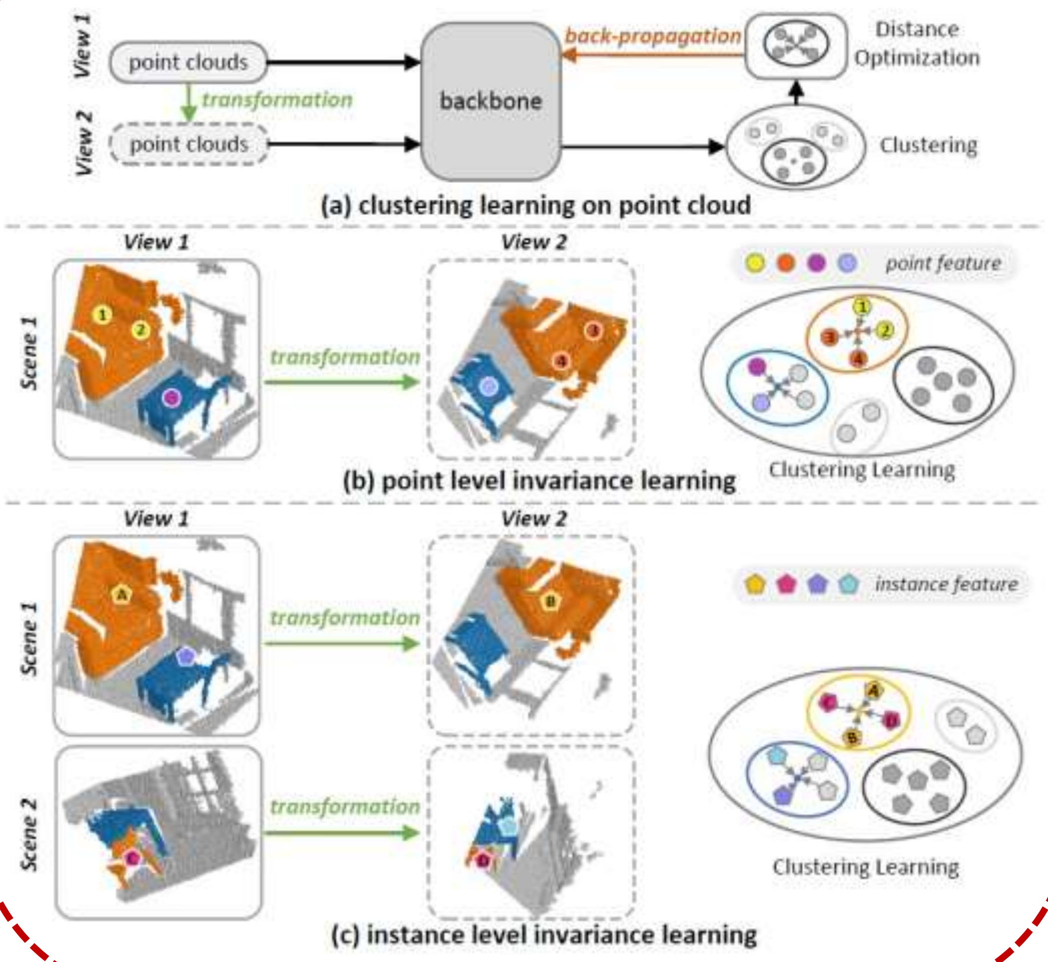


Contrastive Learning



Reconstruction

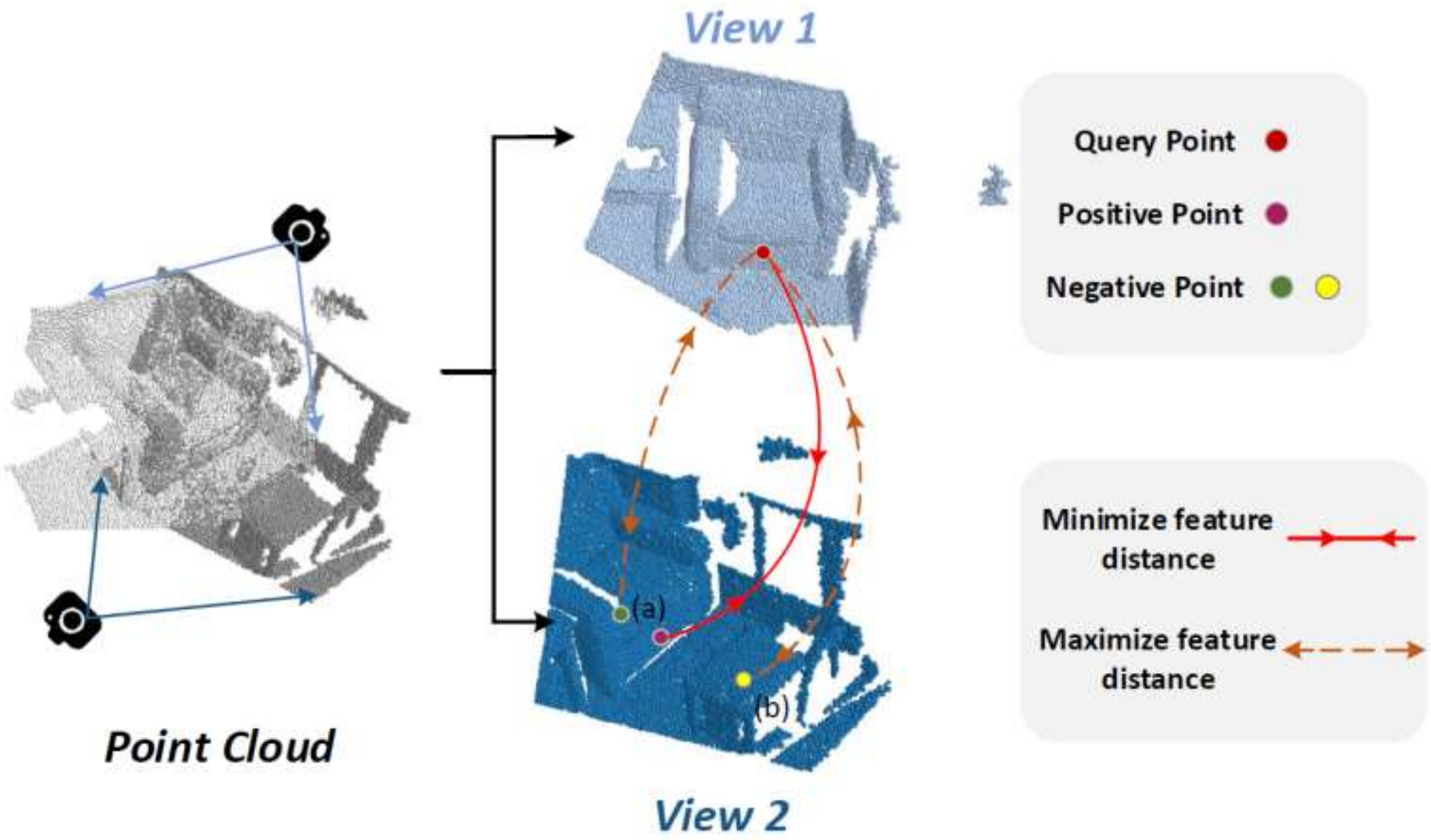
Previous Learning Paradigms



Our Proposal

Unsupervised Point Cloud Pre-training

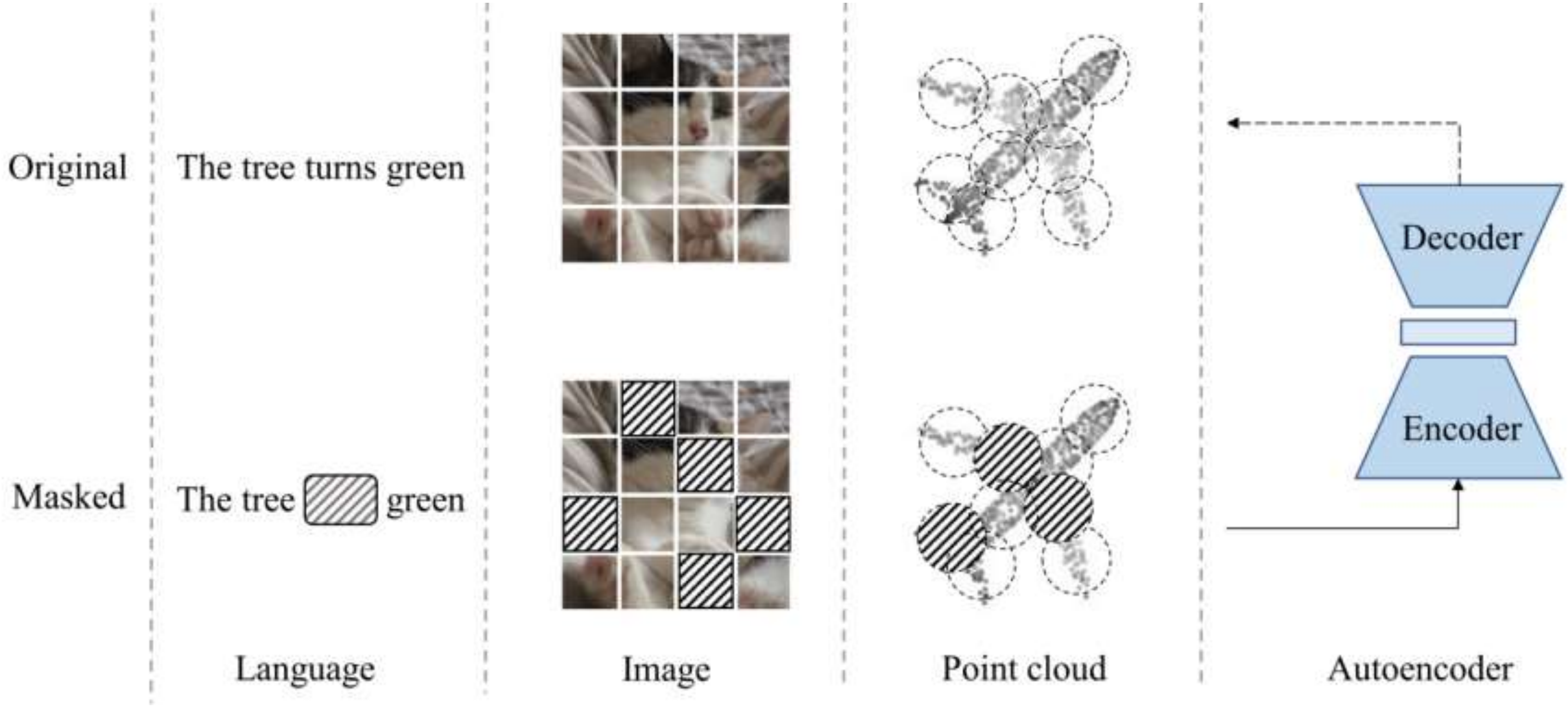
- Contrastive Learning



Xie *et al.* PointContrast: Unsupervised Pre-training for 3D Point Cloud Understanding. In ECCV, 2020.

Unsupervised Point Cloud Pre-training

- Reconstruction



Pang *et al*. Masked Autoencoders for Point Cloud Self-supervised Learning. In ECCV, 2022.

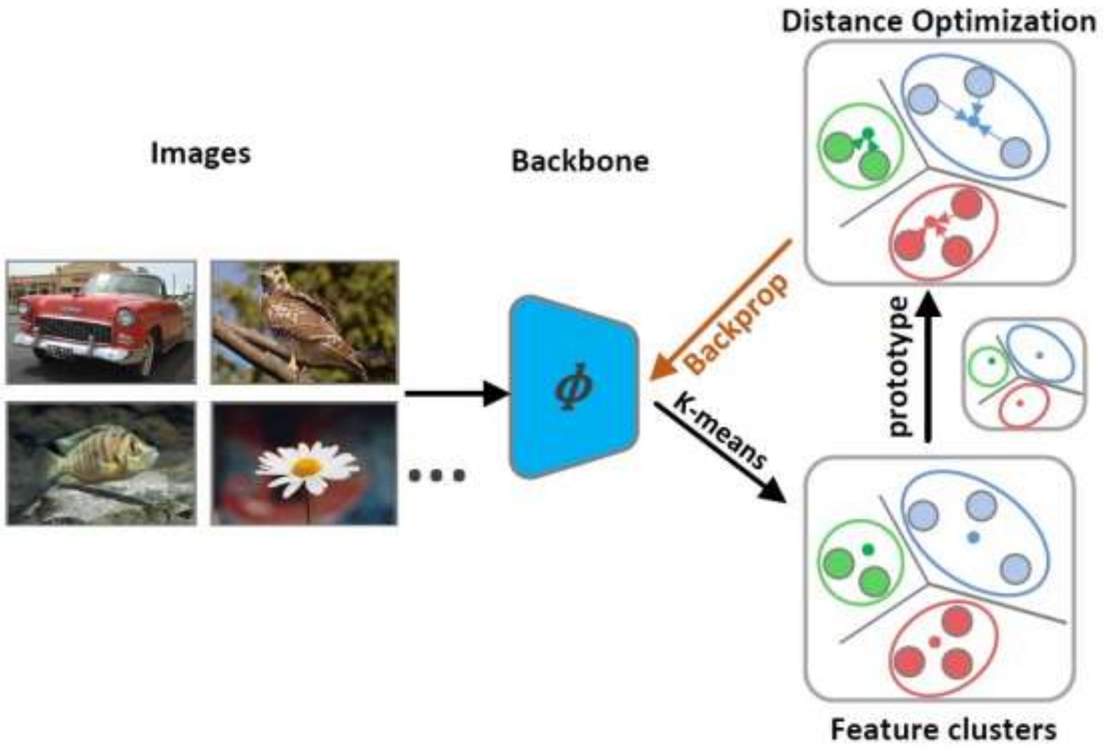
Limitation & Solution

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

- Sample-specific Unsupervised Learning
  - **Semantics of instances** are not fully explored

- New Paradigm of Point Clustering
  - Clustering estimates data distribution **holistically**
  - **Class-level** semantic information mining

Deep Clustering for Image



InfoNCE-based Clustering Loss:

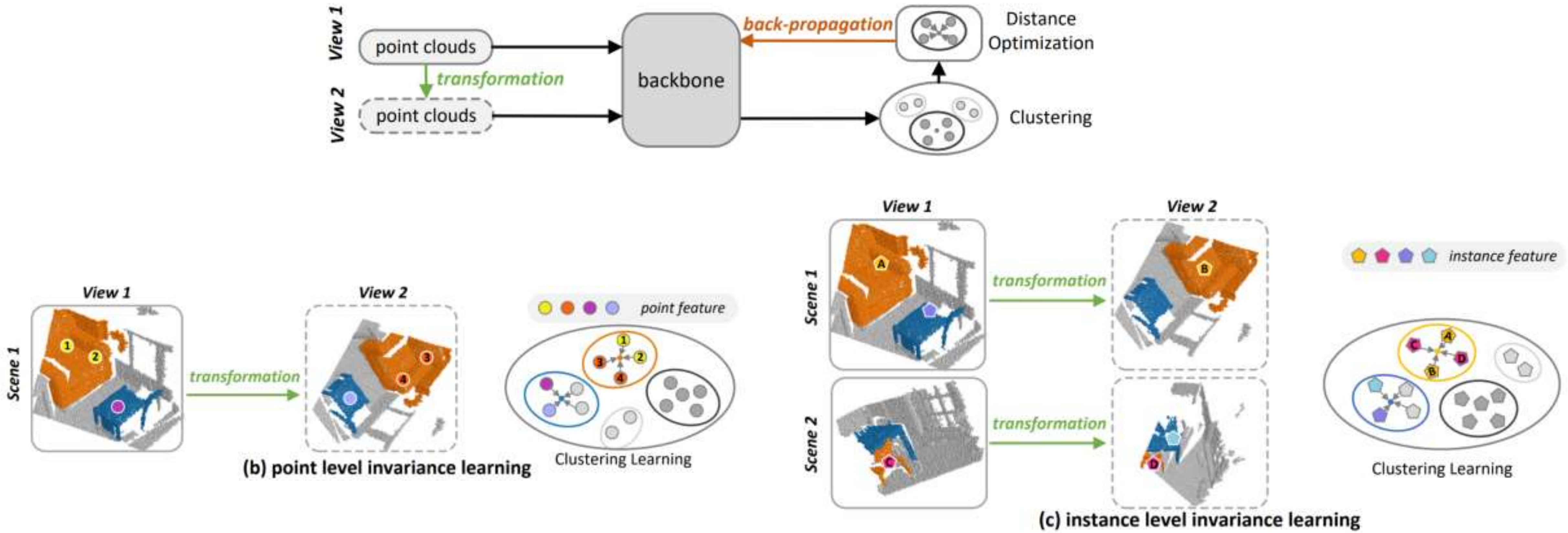
$$L_c(f_i, \mathbf{u}, \mathbf{y}) = -\log \frac{\exp\left(f_i \cdot u_{y_i}/\tau\right)}{\sum_{j=0}^{K-1} \exp\left(f_i \cdot u_j/\tau\right)},$$

- Iterative Learning in each epoch
  - Feature extraction and K-means clustering
  - Similarity optimization between feature and clustering prototypes
- Semantic Exploration
  - Class-level information exploration
- Directly apply to Point Clouds
  - Ignore inherent geometry of 3D point data

Caron *et al.* Deep Clustering for Unsupervised Learning of Visual Features. In ECCV, 2018.

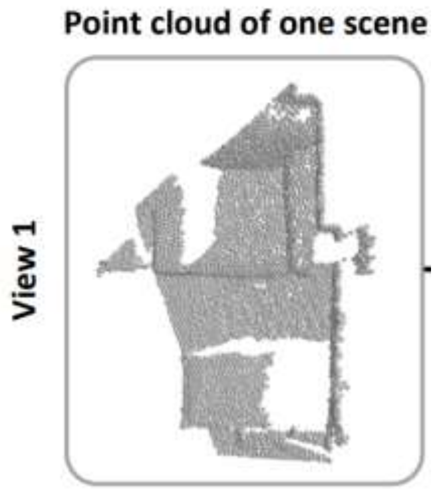
Clustering Learning on Point Clouds

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

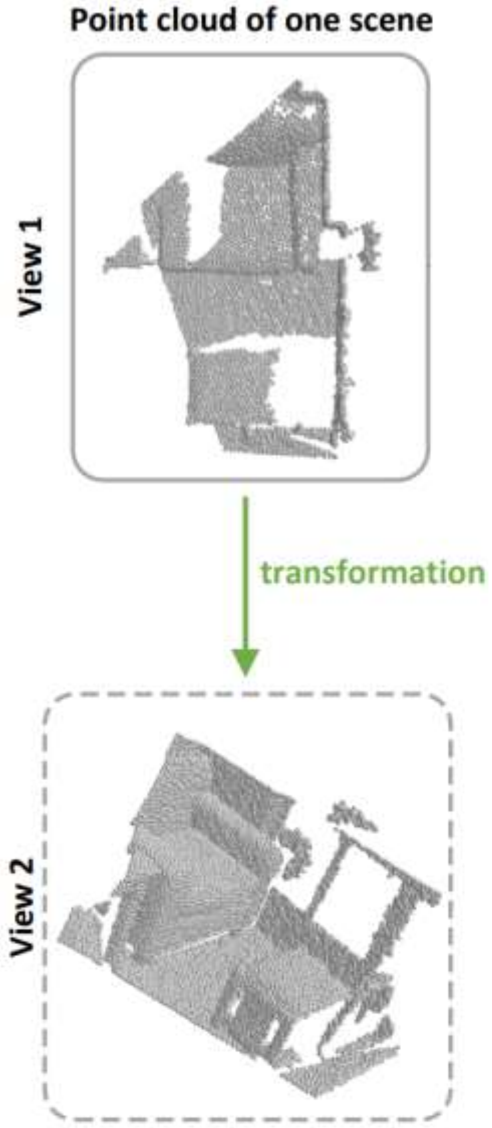


- Feature invariance learning as the inductive bias in clustering
- Exploitation of geometric and semantics for transformation invariance

Point-level Invariance Learning



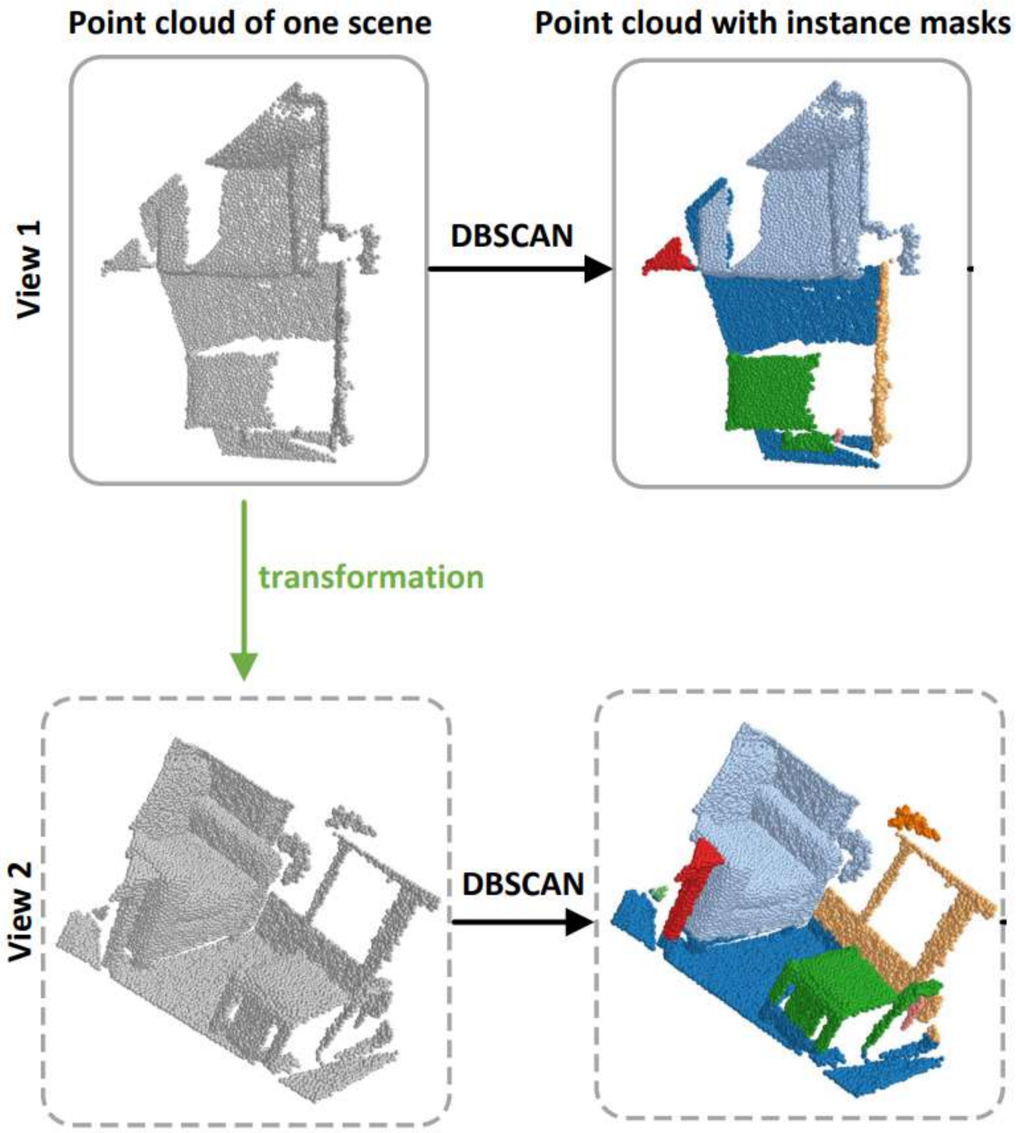
Point-level Invariance Learning



Apply data transformation (e.g., rotation) on each scene to generate two views

Point-level Invariance Learning

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA



Apply DBSCAN to cluster points into instances

Point prototype set of each view

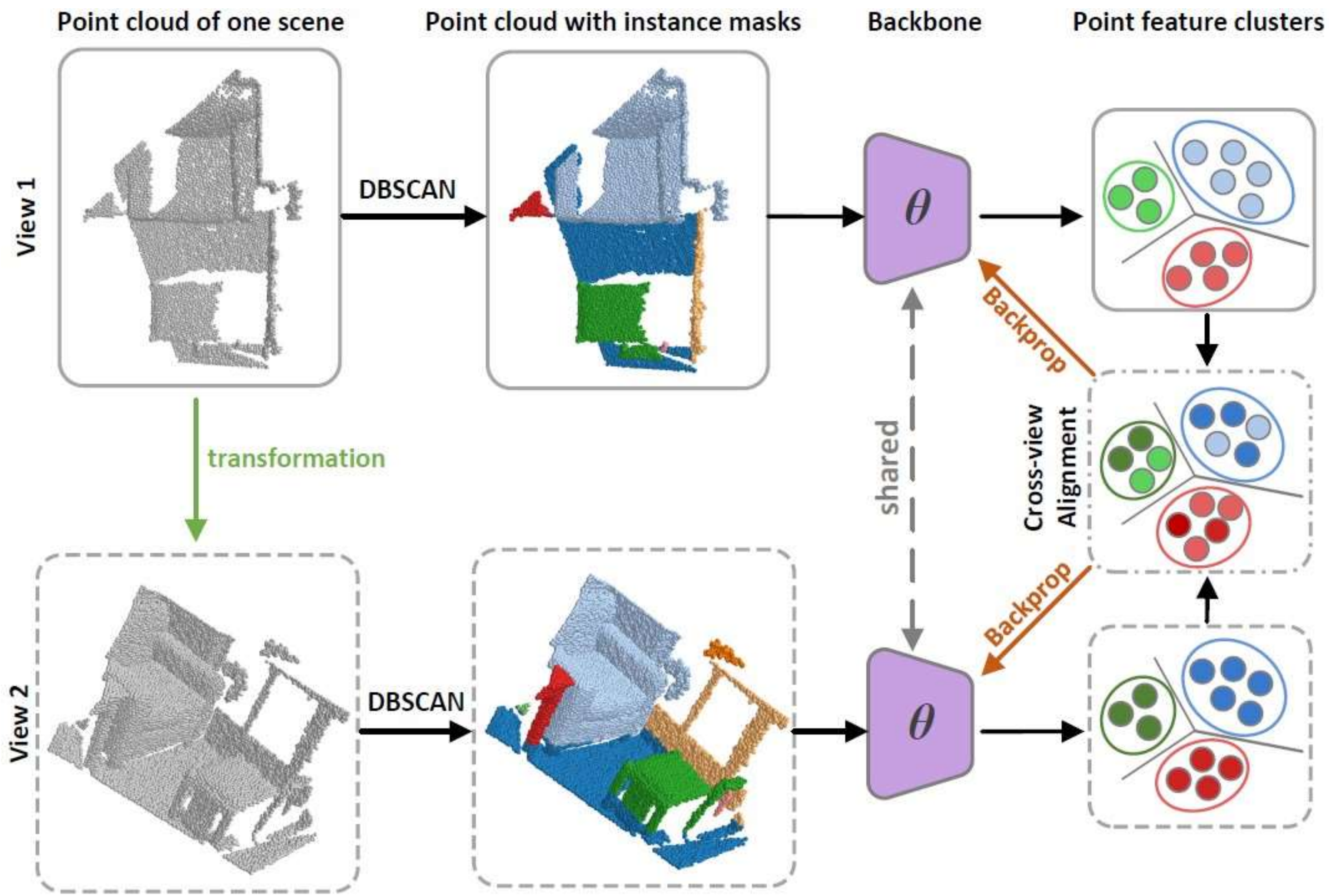
$$\mathbf{u}^{P_1} \text{ and } \mathbf{u}^{P_2}$$

Assigned point label set of each view

$$\mathbf{y}^{P_1} \text{ and } \mathbf{y}^{P_2}$$

Ester et al. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In SIGKDD, 1996.

Point-level Invariance Learning



Point feature extraction of two views

$$f_i^{P_1} \text{ and } f_i^{P_2}$$

Inner-view clustering loss

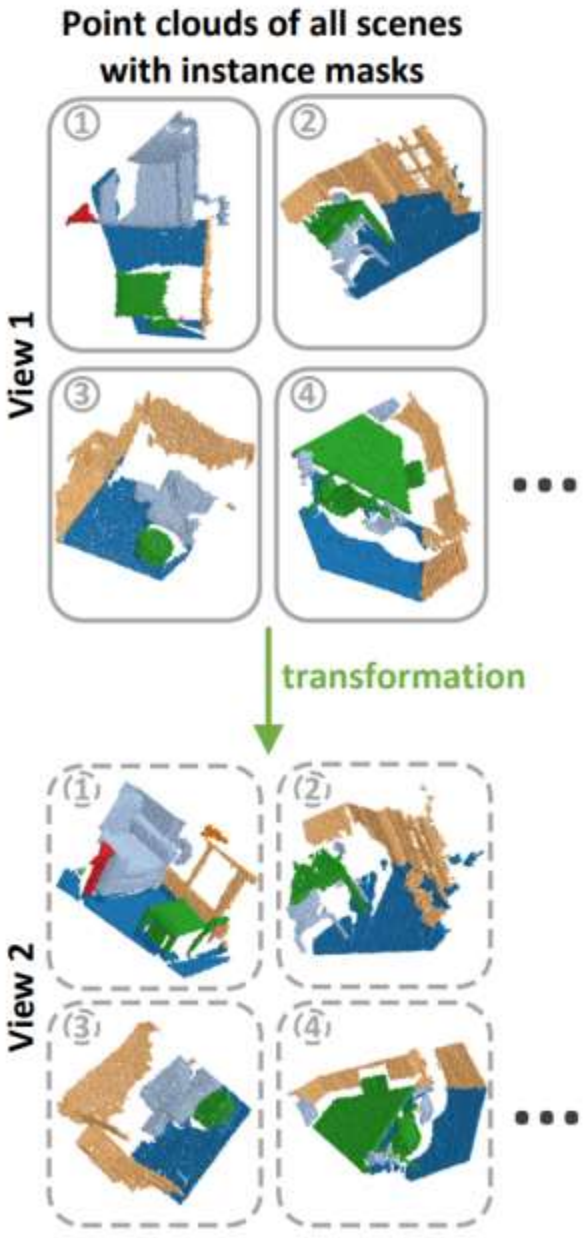
$$L_{ine}^{P} = L_c(f_i^{P_1}, \mathbf{u}^{P_1}, \mathbf{y}^{P_1}) + L_c(f_i^{P_2}, \mathbf{u}^{P_2}, \mathbf{y}^{P_2}),$$

Cross-view clustering loss

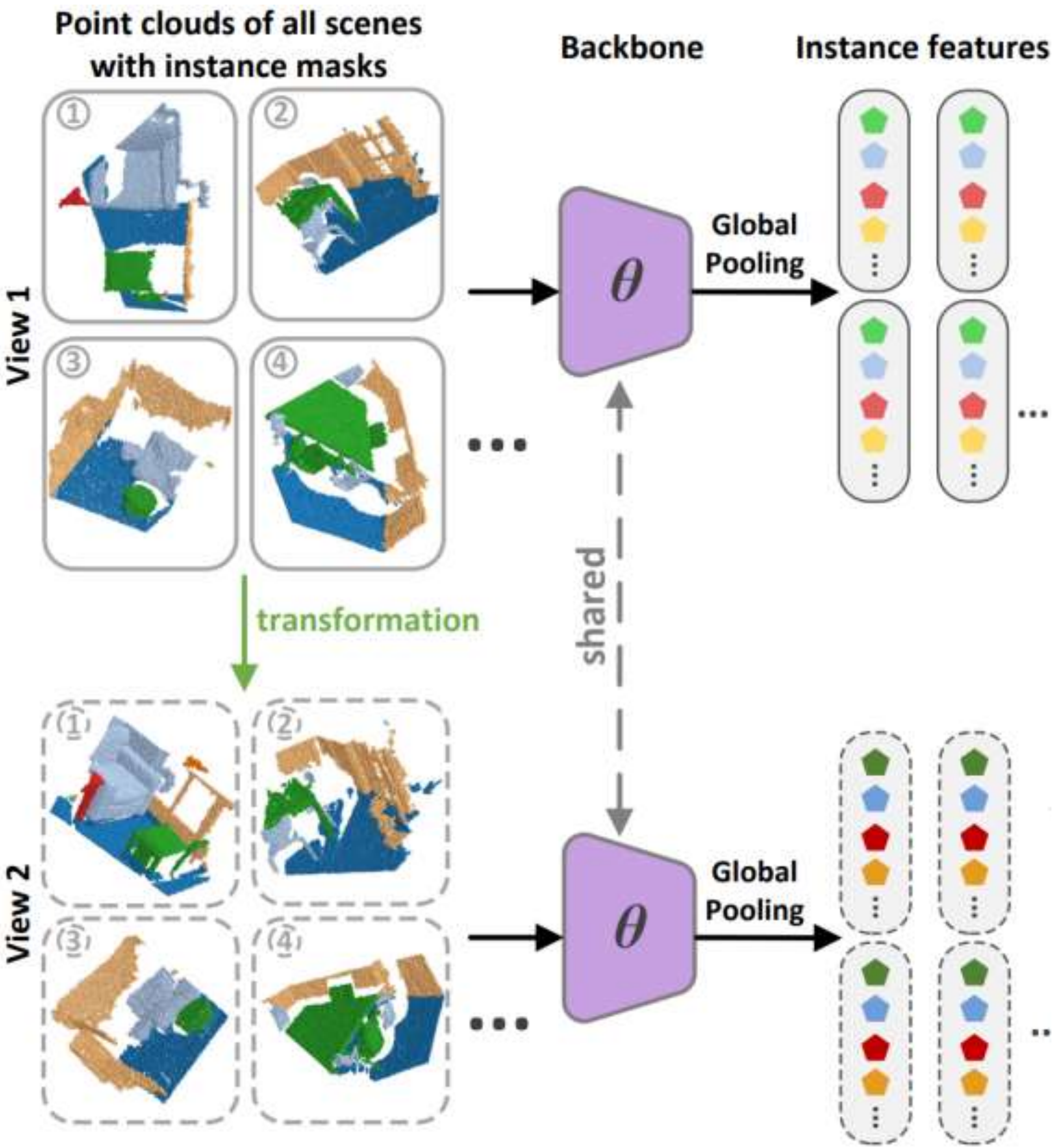
$$L_{cro}^{P} = L_c(f_i^{P_1}, \mathbf{u}^{P_2}, \mathbf{y}^{P_2}) + L_c(f_i^{P_2}, \mathbf{u}^{P_1}, \mathbf{y}^{P_1}).$$

Point-level loss $L^P = L_{ine}^P + L_{cro}^P.$

Instance-level Invariance Learning



Instance-level Invariance Learning



Instance-level invariance learning on **all instances over the entire dataset**

Instance feature is obtained through **globally pooling point features**

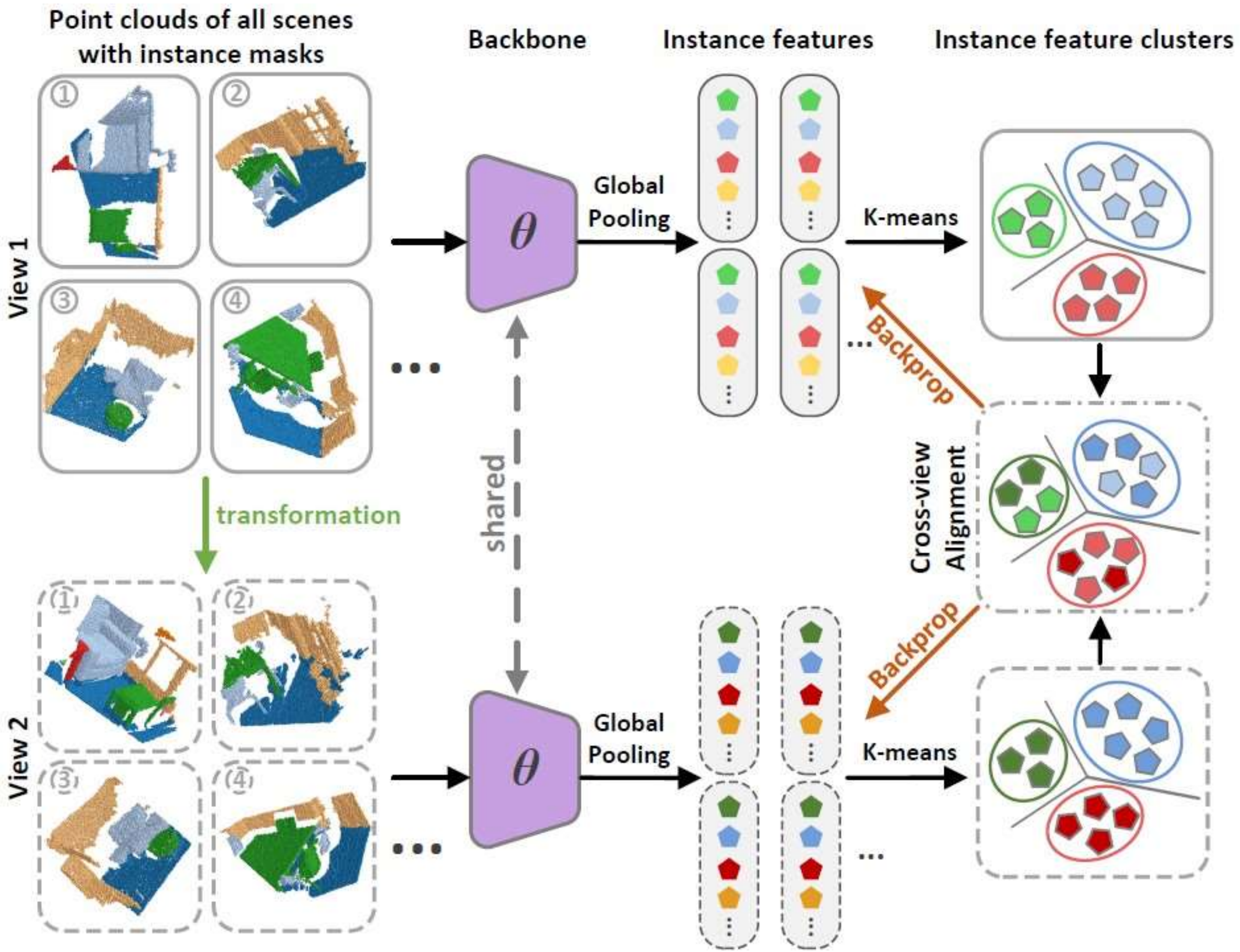
Instance prototype set of each view

$$\mathbf{u}^{I_1} \text{ and } \mathbf{u}^{I_2}$$

Assigned instance label set of each view

$$\mathbf{y}^{I_1} \text{ and } \mathbf{y}^{I_2}$$

Instance-level Invariance Learning



Instance feature of two views

$$f_i^{I_1} \text{ and } f_i^{I_2}$$

Inner-view clustering loss

$$L^I_{ine} = L_c(f_i^{I_1}, \mathbf{u}^{I_1}, \mathbf{y}^{I_1}) + L_c(f_i^{I_2}, \mathbf{u}^{I_2}, \mathbf{y}^{I_2}).$$

Cross-view clustering loss

$$L^I_{cro} = L_c(f_i^{I_1}, \mathbf{u}^{I_2}, \mathbf{y}^{I_2}) + L_c(f_i^{I_2}, \mathbf{u}^{I_1}, \mathbf{y}^{I_1}).$$

Instance-level loss $L^I = L^I_{ine} + L^I_{cro}.$

Overall loss $L_{ov} = L^P + L^I.$

Experiments

- Datasets
  - Pre-training Dataset
    - ScanNet: 2.5M RGB-D scaning frames
    - Extract 190K 3D scans from 1,200 depth video sequences
    - Sample 8,192 points of each scans for pre-training
  - Datasets of Downstream Tasks

Dataset	Statistic	Task	Gain
ModelNet40 [62]	9.8K train, 2.5K val	Object Cls.	**+3.0% Acc**
ScanObjectNN [56]	11.4K train, 2.9K val	Object Cls.	**+10.1% Acc**
ShapeNetPart [69]	14.0K train, 2.9K val	Part Seg.	**+1.6% mIoU**
PartNet [40]	17.1K train, 2.5K val	Part Seg.	**+4.3% mIoU**
S3DIS [3]	199 train, 67 val	Semantic Seg.	**+6.7% mIoU**
ScanNetV2 [10]	1.2K train, 312 val	Semantic Seg.	**+5.7% mIoU**

Performance Comparisons

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

- Comparisons with the state-of-the-art methods

Performances on classification

Approach	Backbone	ModelNet40	ScanObjectNN
Scratch	PointNet++	90.7	77.9
DepthContrast [74]	PointNet++	91.3	-
GLR [50]	PointNet++	93.0	-
ReSp [52]	DGCNN	92.4	-
OcCo [60]	DGCNN	93.0	-
PointClustering	PointNet++	**94.1** (+3.4)	**84.5** (+6.6)
Scratch	SR-UNet	90.1	76.2
PointContrast [65]	SR-UNet	91.2	-
PointClustering	SR-UNet	**93.6** (+3.5)	**83.7** (+7.5)
Scratch	PointViT	91.5	77.2
Point-BERT [70]	PointViT	93.2	83.1
MaskPoint [34]	PointViT	93.8	84.3
Point-MAE [41]	PointViT	93.8	85.2
MaskSurf [73]	PointViT	93.4	85.8
PointClustering	PointViT	**94.5** (+3.0)	**87.3** (+10.1)

Performances on part segmentation

Approach	Backbone	ShapeNetPart	PartNet
Scratch	PointNet++	84.9	42.5
OcCo [60]	DGCNN	85.0	-
ReSp [52]	DGCNN	85.3	-
PointClustering	PointNet++	**85.9** (+1.0)	**47.0** (+4.5)
Scratch	SR-UNet	84.7	38.9
PointContrast [65]	SR-UNet	85.1	41.5
PointClustering	SR-UNet	**86.0** (+1.3)	**42.1** (+3.2)
Scratch	PointViT	85.1	45.8
Point-BERT [70]	PointViT	85.6	-
MaskPoint [34]	PointViT	86.0	-
MaskSurf [73]	PointViT	86.1	-
Point-MAE [41]	PointViT	86.1	-
PointClustering	PointViT	**86.7** (+1.6)	**50.1** (+4.3)

Performances on semantic segmentation

Approach	Backbone	S3DIS	ScanNetV2
Scratch	PointNet++	55.3	57.9
OcCo [60]	DGCNN	58.0	-
PointClustering	PointNet++	**61.2** (+5.9)	**62.6** (+4.7)
Scratch	SR-UNet	68.2	70.3
DepthContrast [74]	SR-UNet	71.5	71.2
CSC [23]	SR-UNet	72.2	73.8
PointContrast [65]	SR-UNet	70.9	74.1
PointClustering	SR-UNet	**73.2** (+5.0)	**75.5** (+5.2)
Scratch	PointViT	58.9	60.1
Point-MAE [41]	PointViT	60.0	-
MaskSurf [73]	PointViT	61.6	-
PointClustering	PointViT	**65.6** (+6.7)	**65.8** (+5.7)

PointClustering achieves better performances with three kinds of point backbone, i.e., PointNet++, SR-UNet and PointViT , on all benchmarks

Performance Comparisons

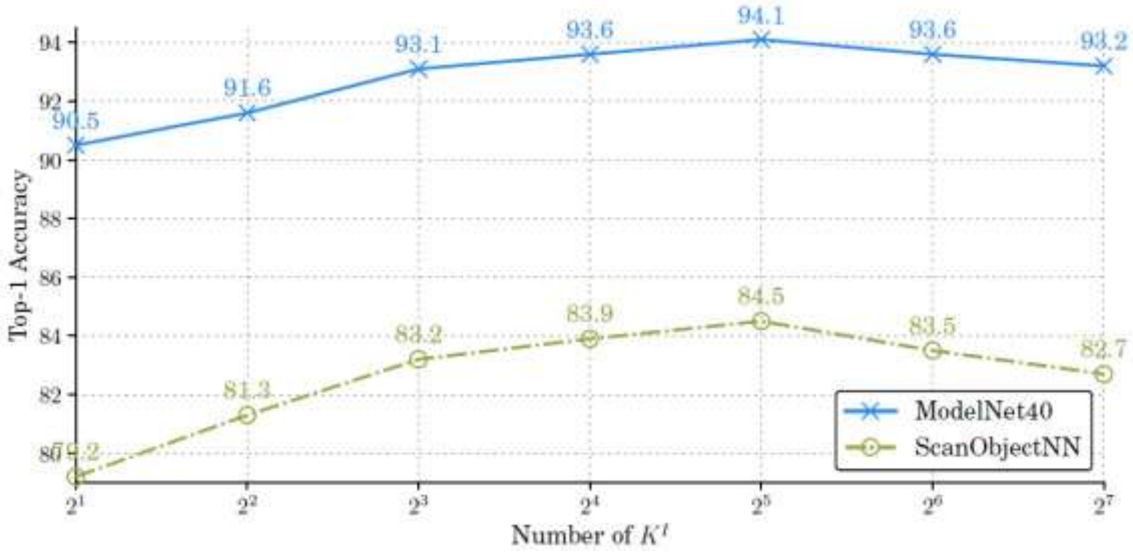
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

- Ablation Studies

Ablation studies on different invariance learning

Model: point-level inv.	Model: instance-level inv.	ModelNet40	ScanObjectNN
Scratch		90.7	77.9
SceneClustering		91.0	78.1
PointClustering$^-$		91.5	80.1
✓		93.0	82.6
	✓	93.4	83.1
✓	✓	**94.1**	**84.5**

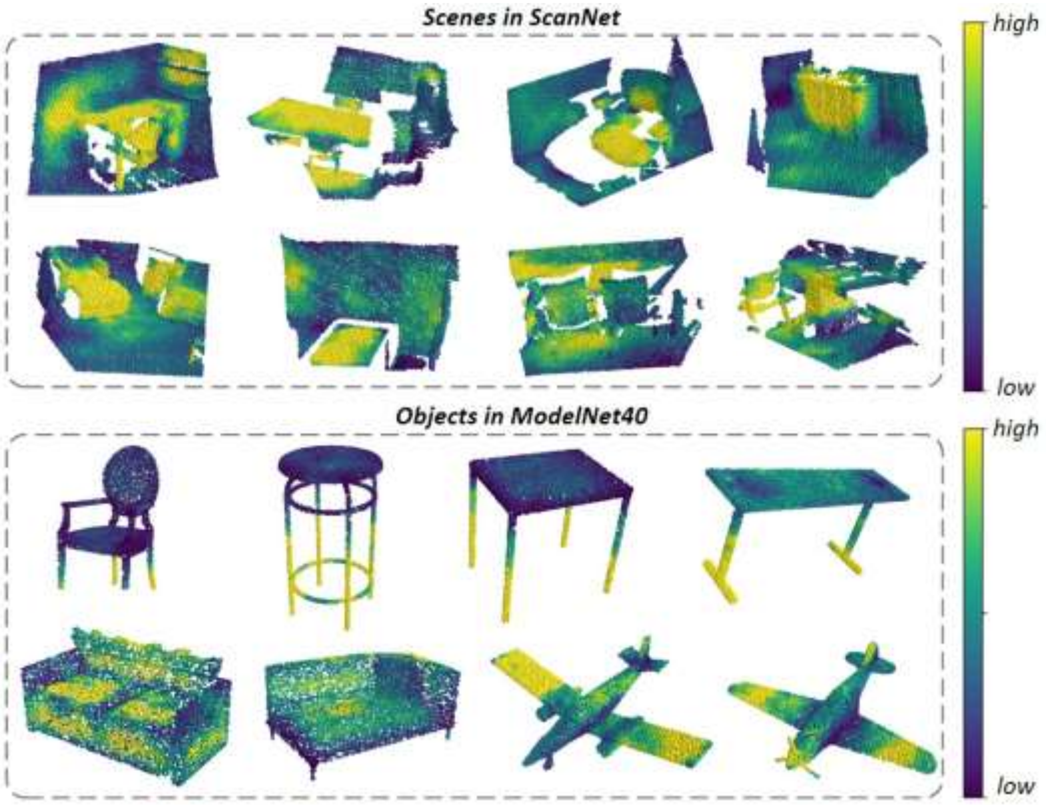
Ablation studies on instance clustering number K



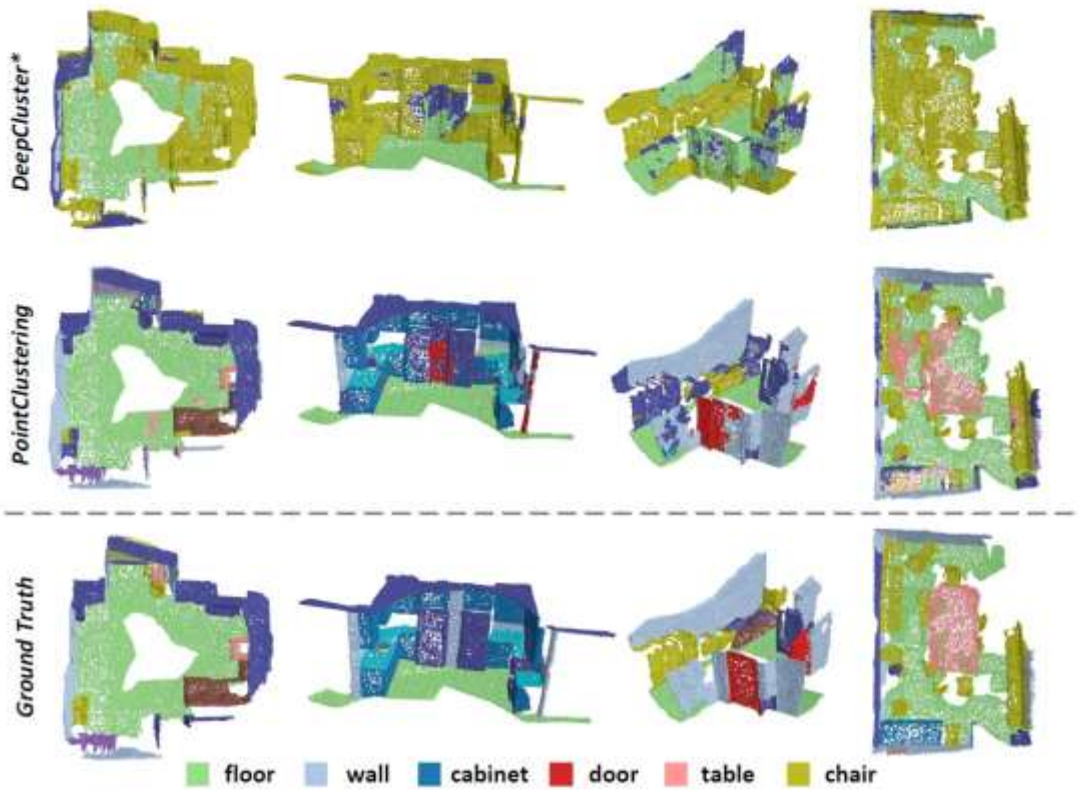
Visualization

- Point-level and Instance-level

Point feature visualization on ModelNet40



Unsupervised semantic segmentation on ScaneNetV2



JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Thanks!

longfc.ustc@gmail.com

Source Code: https://github.com/FuchenUSTC/PointClustering