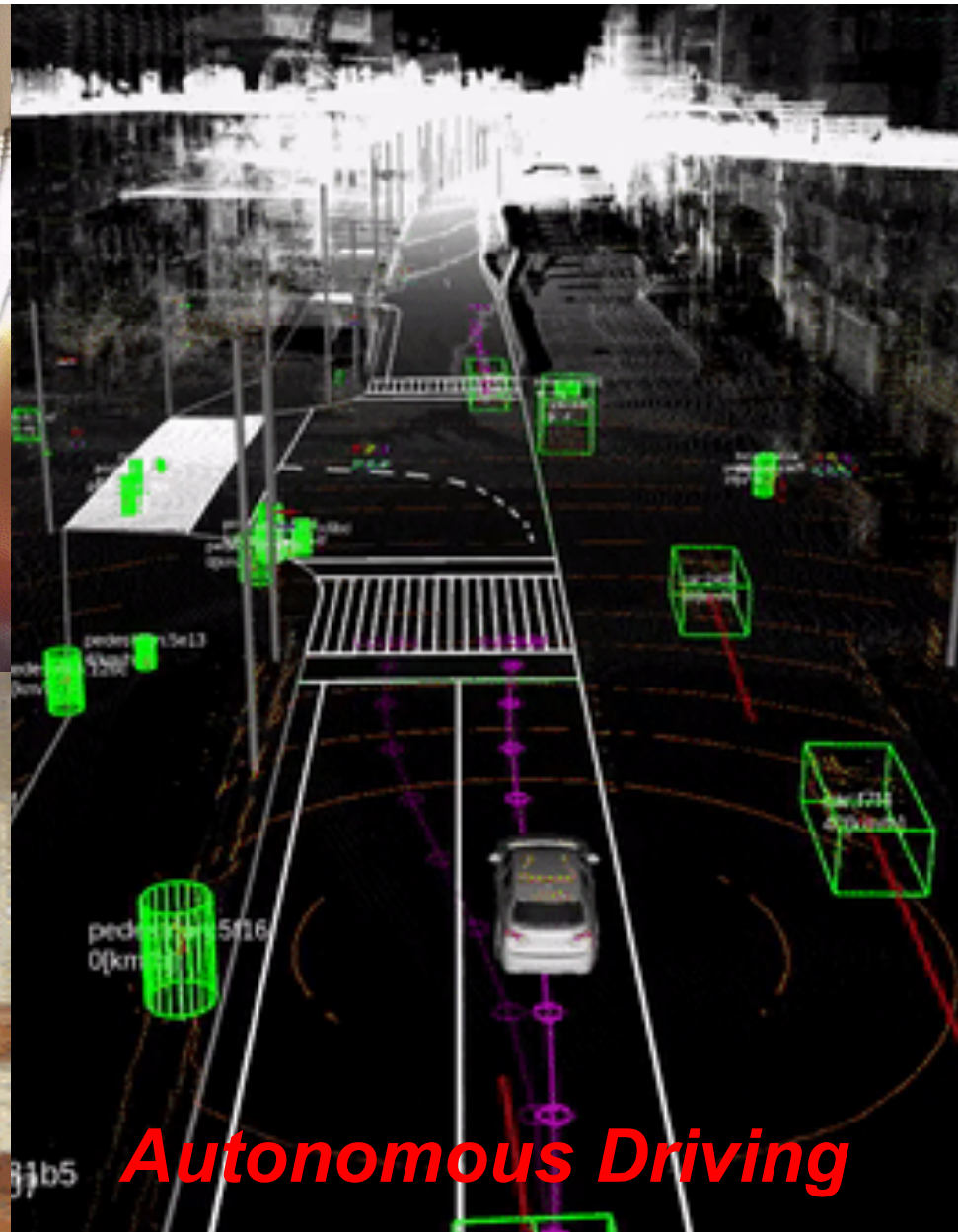
Learning 3D Representations without Full Supervision

Dr. Bo Yang

Assistant Professor
Department of Computing
The Hong Kong Polytechnic University

Group page:
https://vlar-group.github.io/

Future of AI: Intelligent systems that perceive the 3D world



Agenda

01 Background

02 Fully-supervised Methods

03 Weakly-supervised Methods

04 Unsupervised Methods

05 Future Work / Q&A

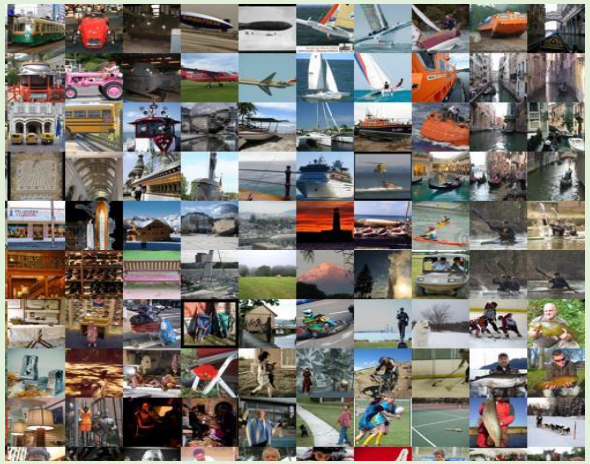
Background

01 Background

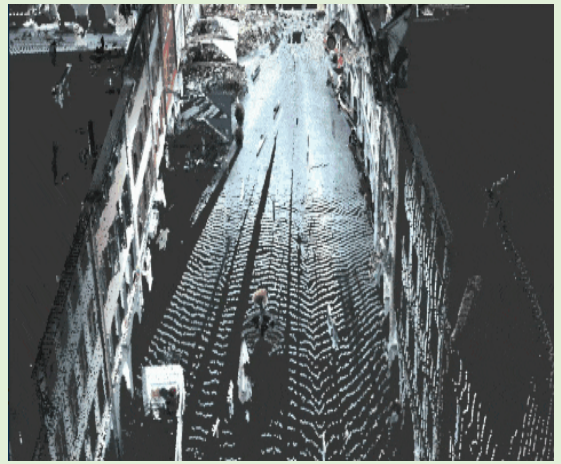
Sensors



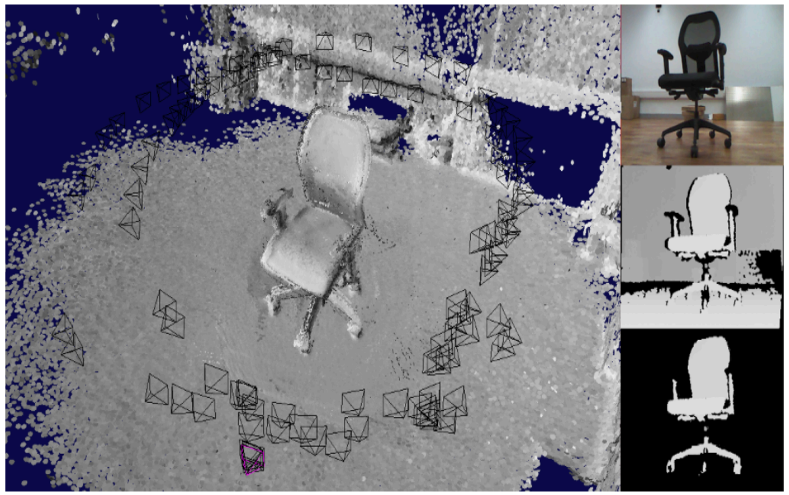
Data

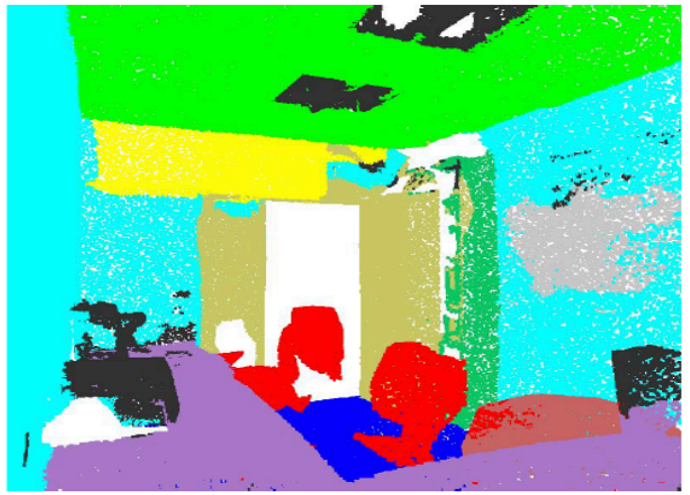

Images

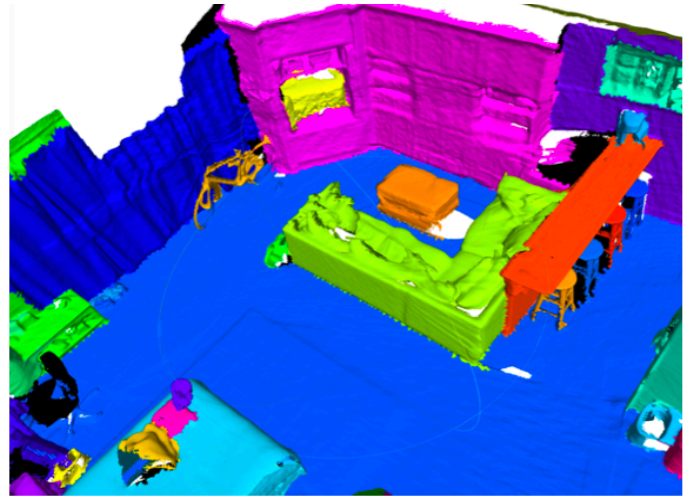

Video



Point Cloud (sequence)

01 Background







Tasks

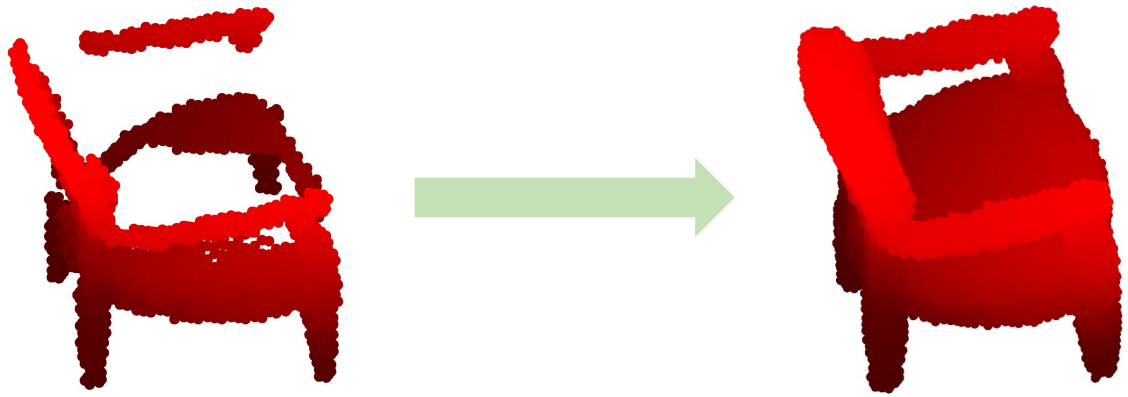
3D Modelling and Understanding

Fully-supervised Methods

02

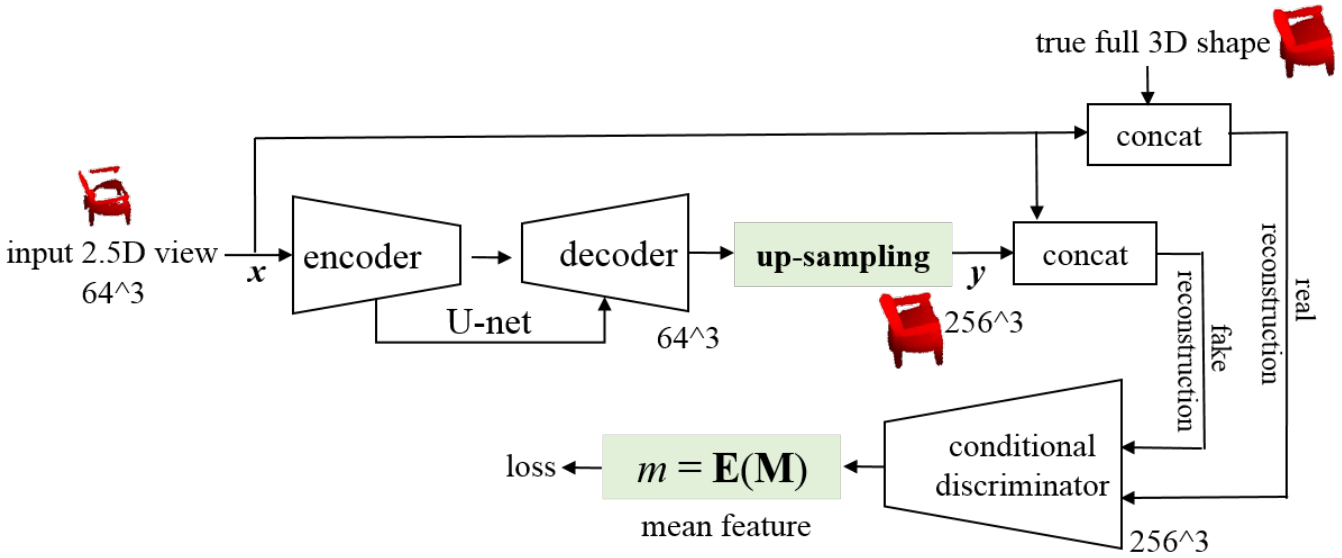
Fully-supervised Methods

Single-view Reconstruction



Single Depth View

3D Full Shape

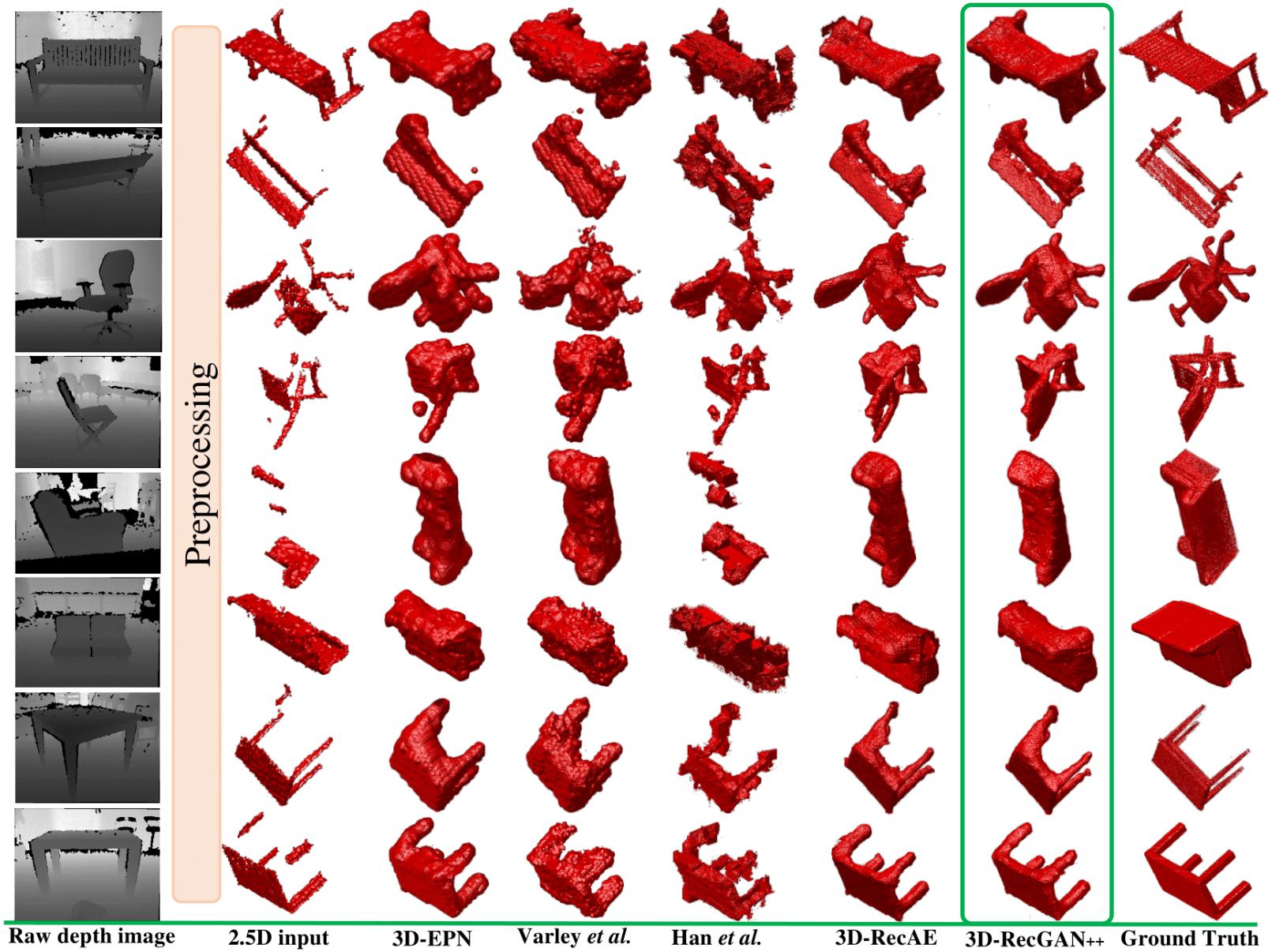


B Yang, S Rosa, A Markham, N Trigoni, H Wen, "Dense 3D Object Reconstruction from a Single Depth View", TPAMI, 2018

02

Fully-supervised Methods

Single-view Reconstruction

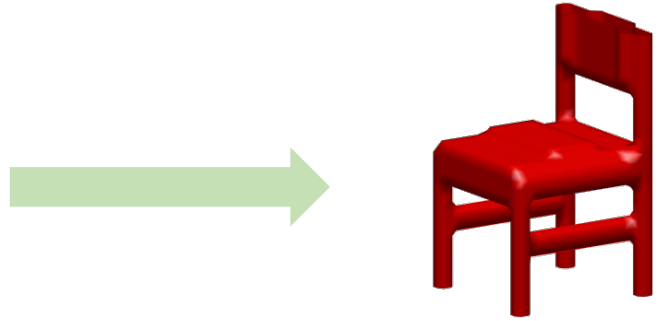


02

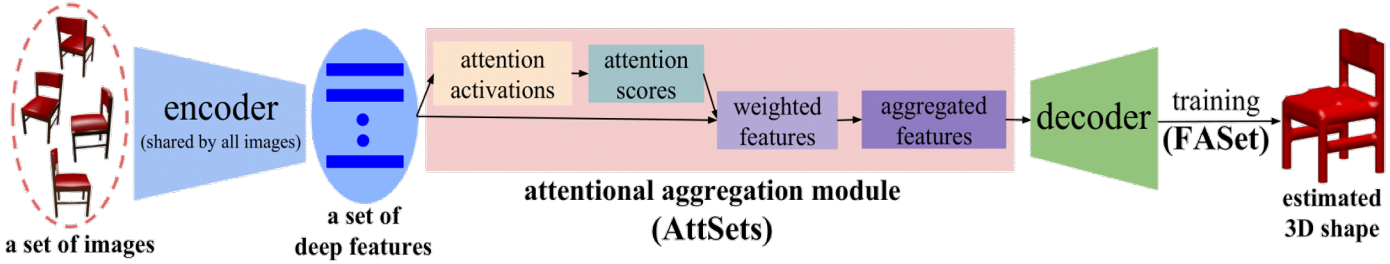
Fully-supervised Methods

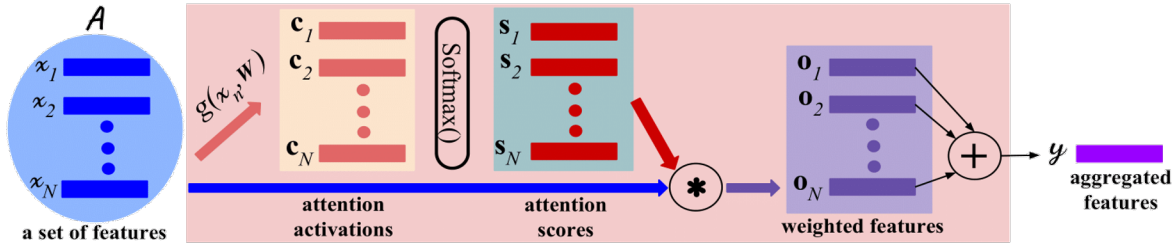
Multi-view Reconstruction

Multiple Views



3D Full Shape



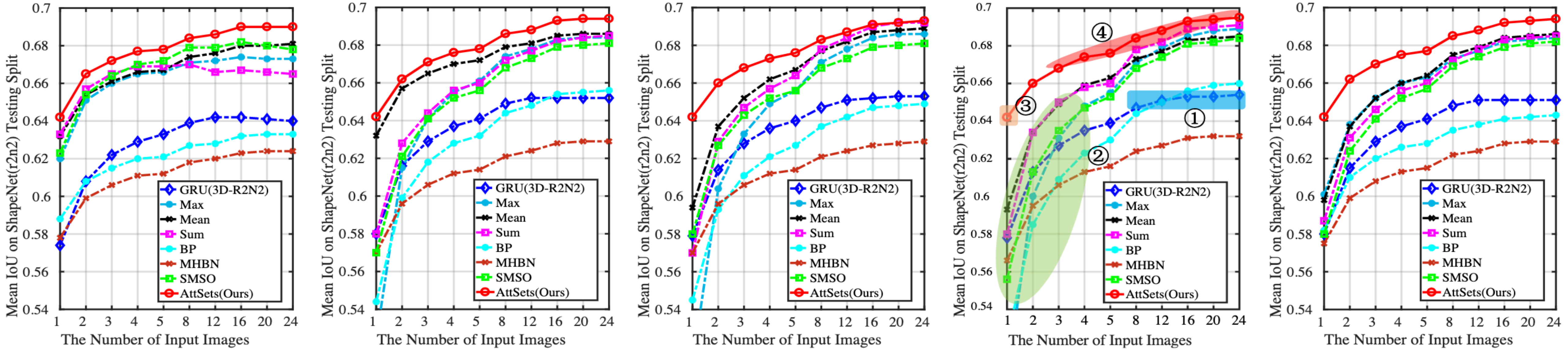


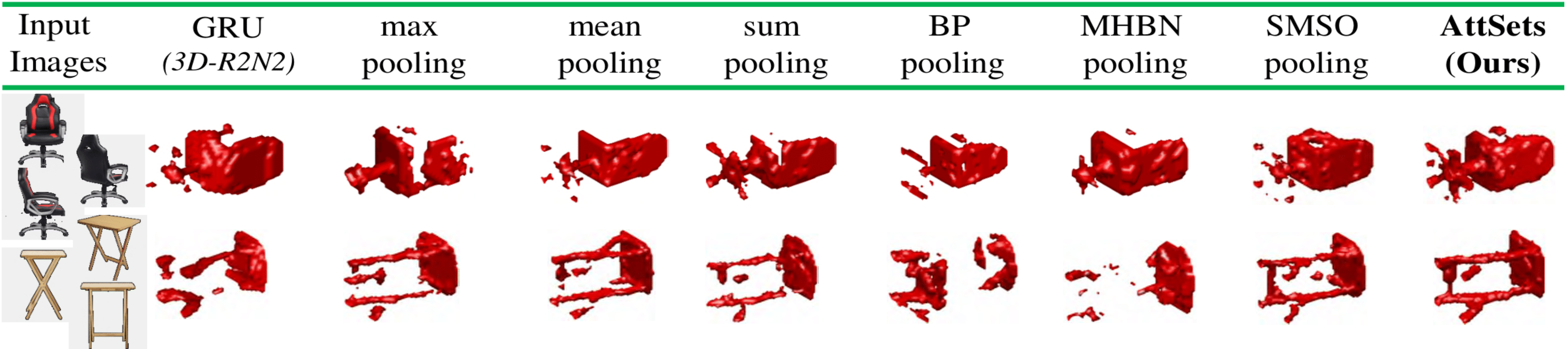
B Yang, S Wang, A Markham, N Trigoni, "Robust Attentional Aggregation of Deep Feature Sets for Multi-view 3D Reconstruction", IJCV, 2019

02

Fully-supervised Methods

Multi-view Reconstruction



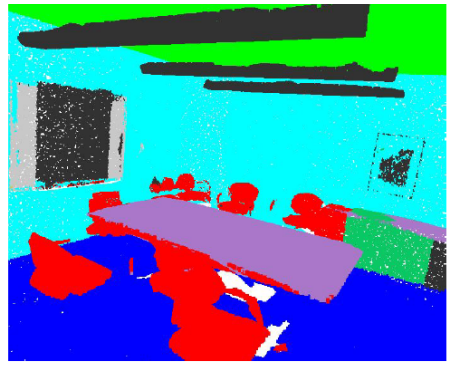


02

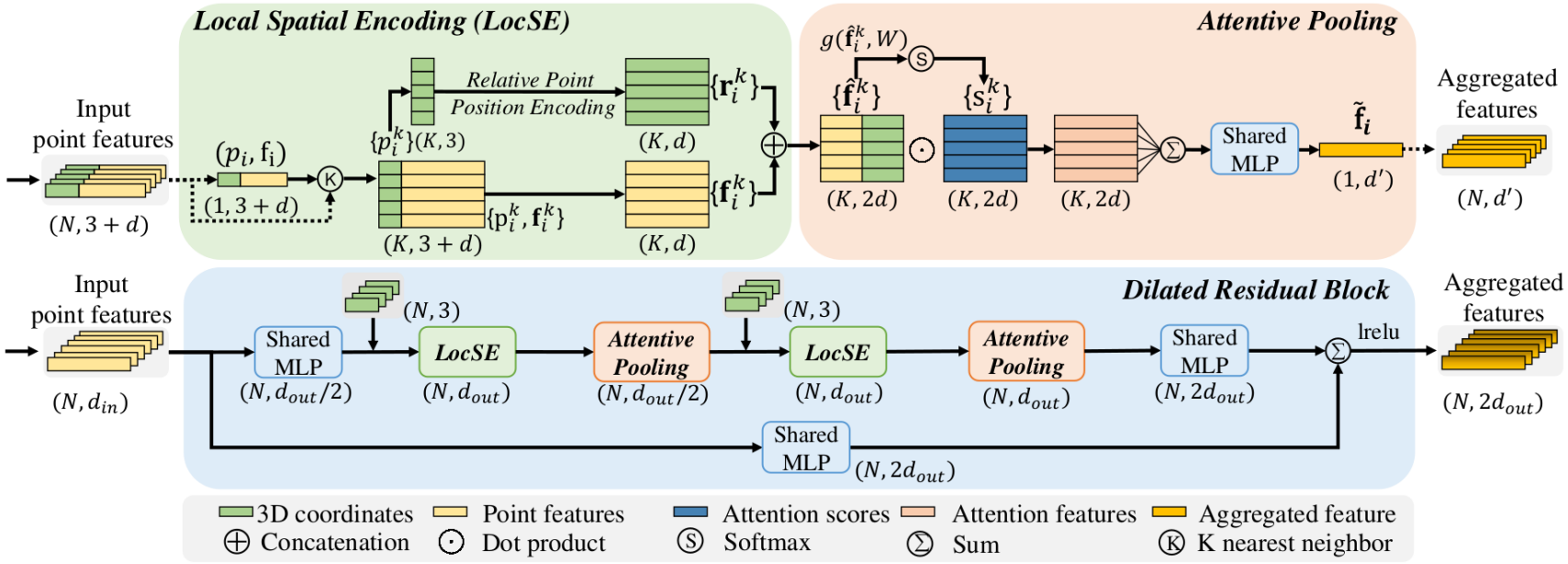
Fully-supervised Methods

Semantic Segmentation of Point Clouds

Point Cloud



Semantic Classes

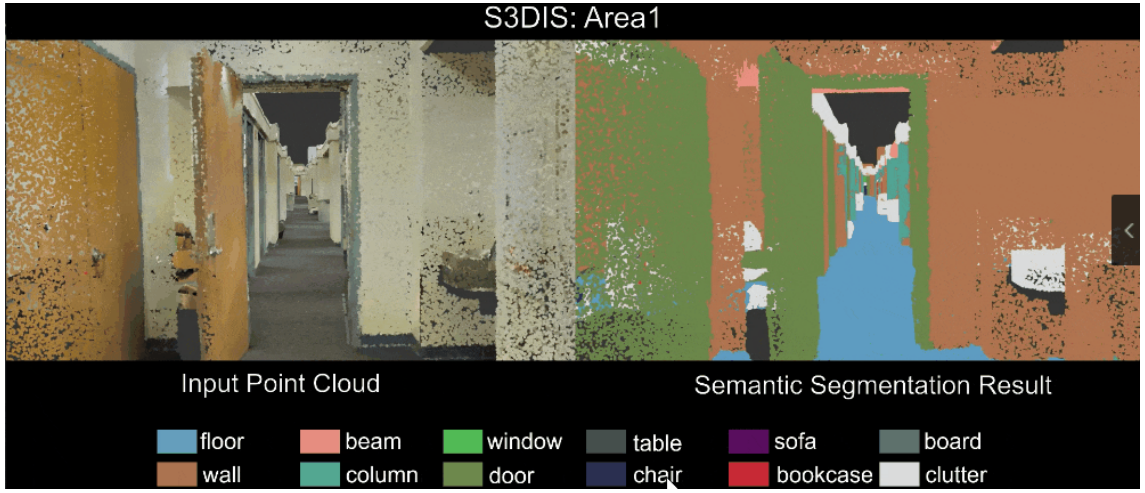


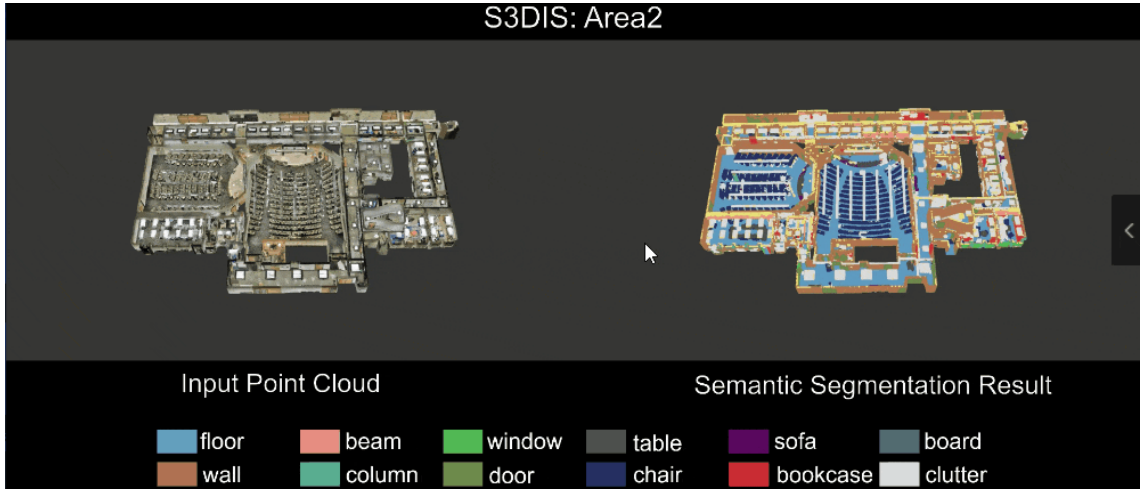
Q Hu, B Yang, L Xie, S Rosa, Y Guo, Z Wang, N Trigoni, A Markham, "RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds", CVPR, 2020*

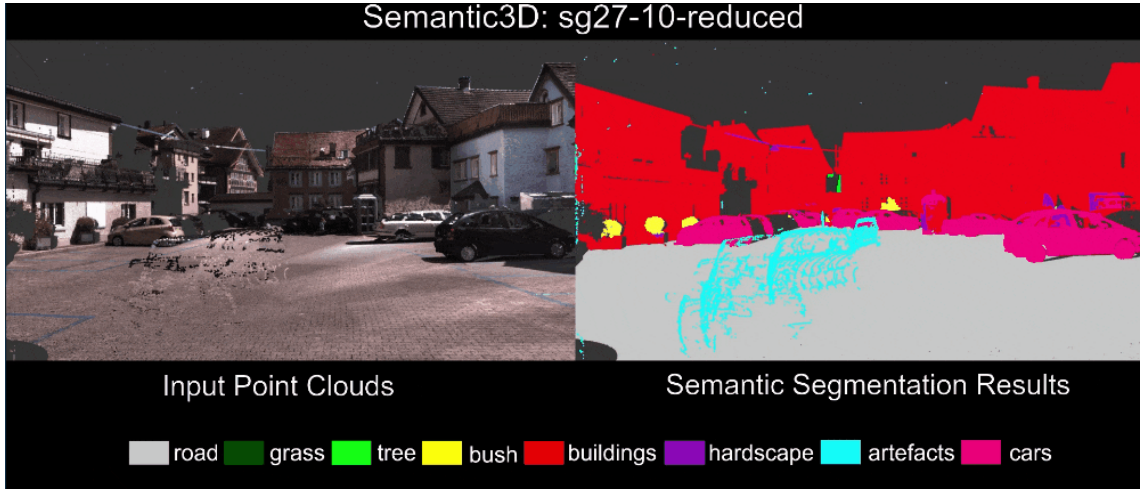
02

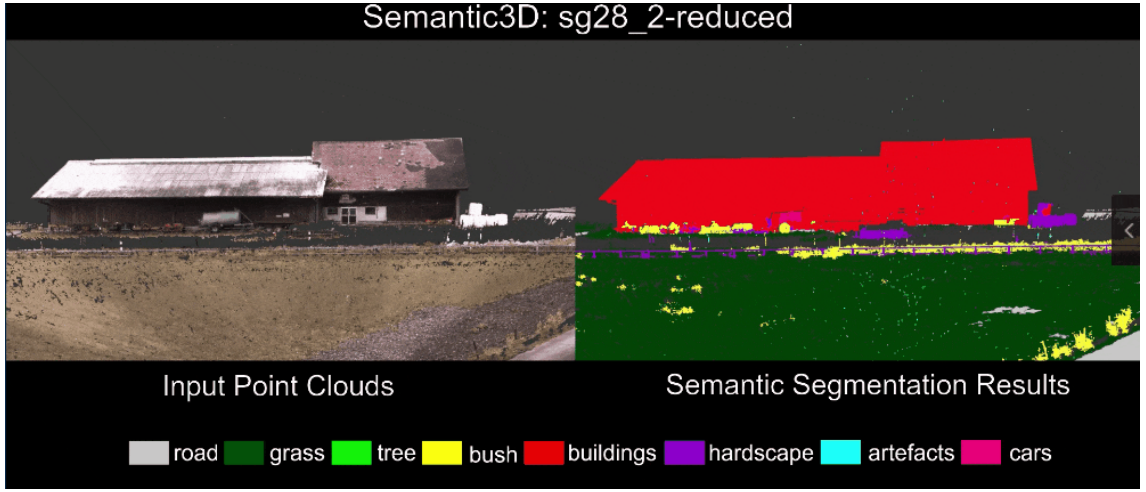
Fully-supervised Methods

Semantic Segmentation of Point Clouds





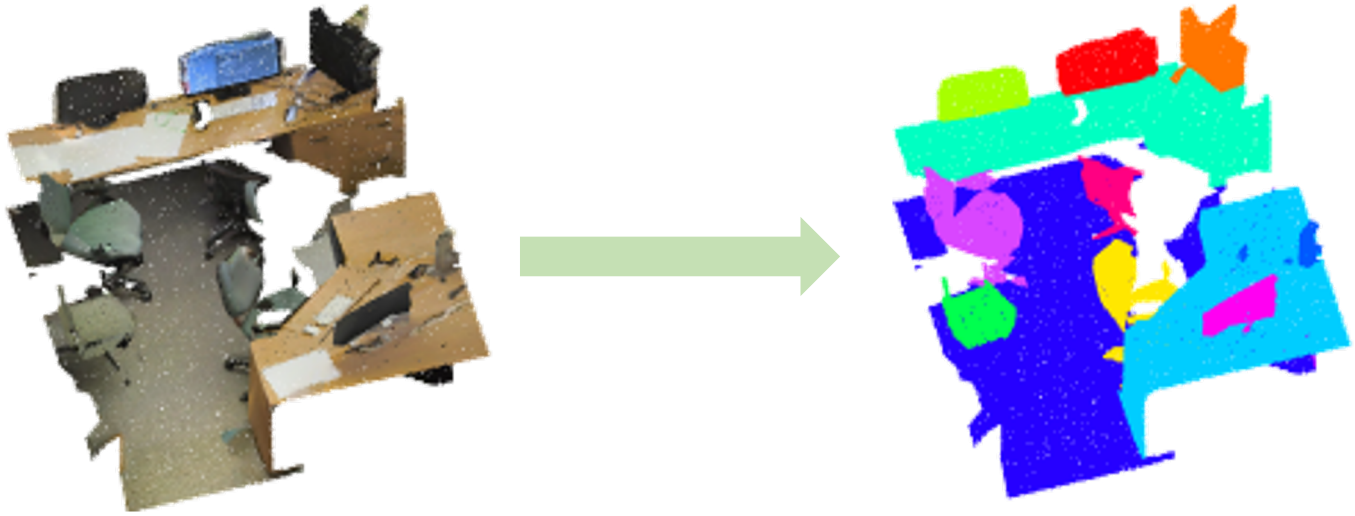




02

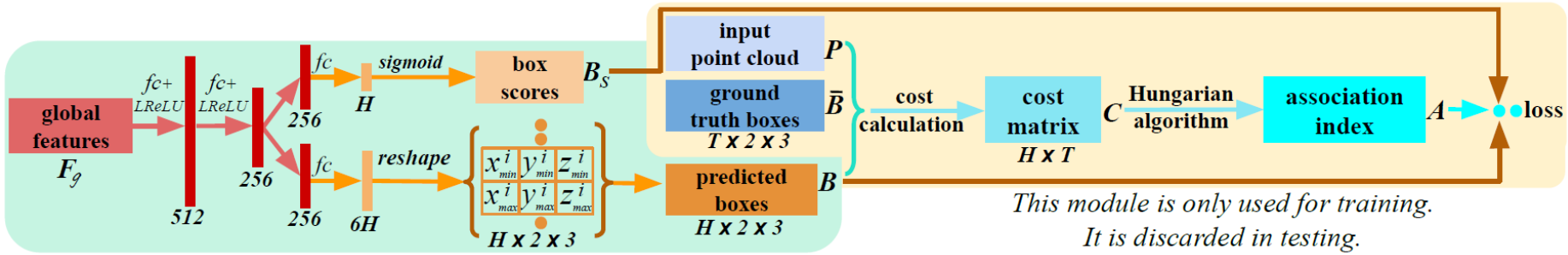
Fully-supervised Methods

Instance Segmentation of Point Clouds



Point Cloud

Instance Masks



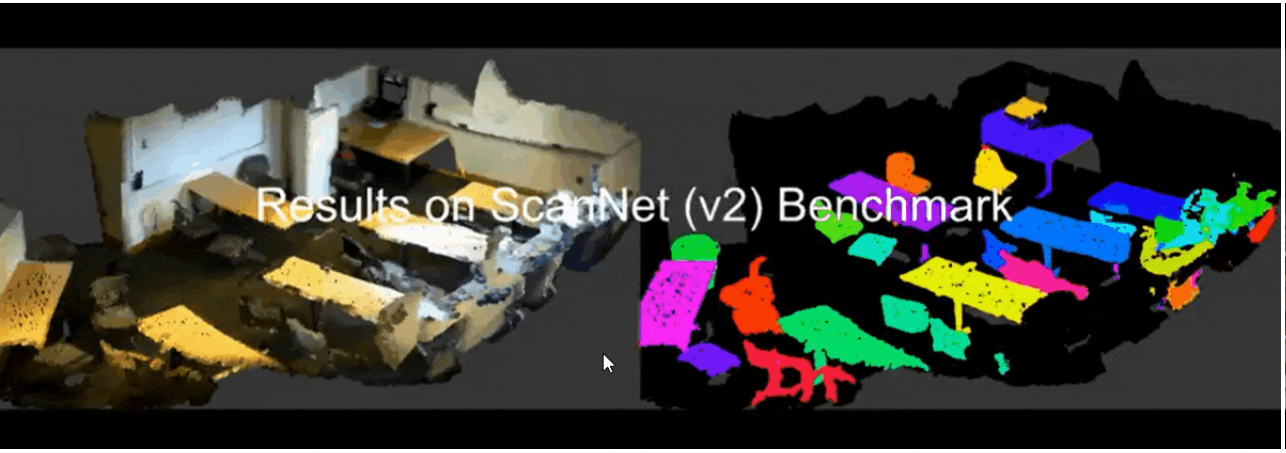
$$\boldsymbol{A} = \arg\min_{\boldsymbol{A}} \sum_{i=1}^{H}\sum_{j=1}^{T} \boldsymbol{C}_{i,j}\boldsymbol{A}_{i,j} \quad \text{subject to} \sum_{i=1}^{H} \boldsymbol{A}_{i,j} = 1, \sum_{j=1}^{T} \boldsymbol{A}_{i,j} \leq 1, j \in \{1..T\}, i \in \{1..H\}$$

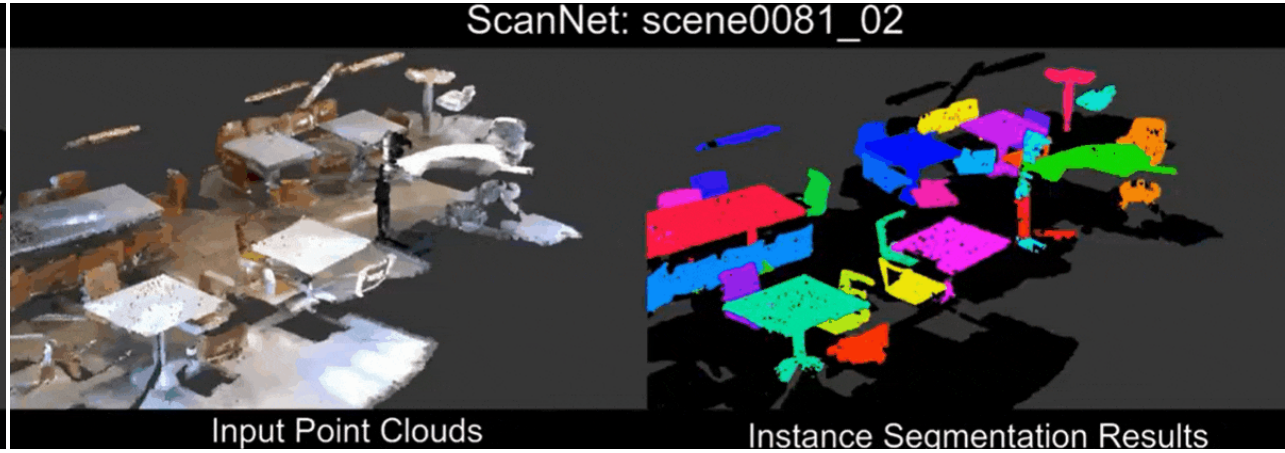
B Yang, J Wang, R Clark, Q Hu, S Wang, A Markham, N Trigoni, *"Learning Object Bounding Boxes for 3D Instance Segmentation on Point Clouds"*, *NeurIPS, 2019*

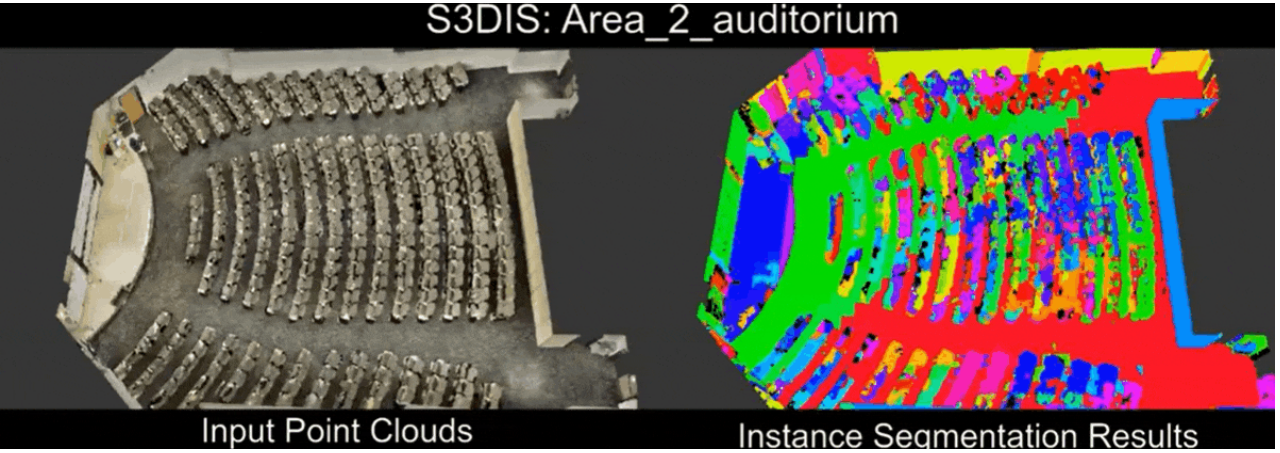
02

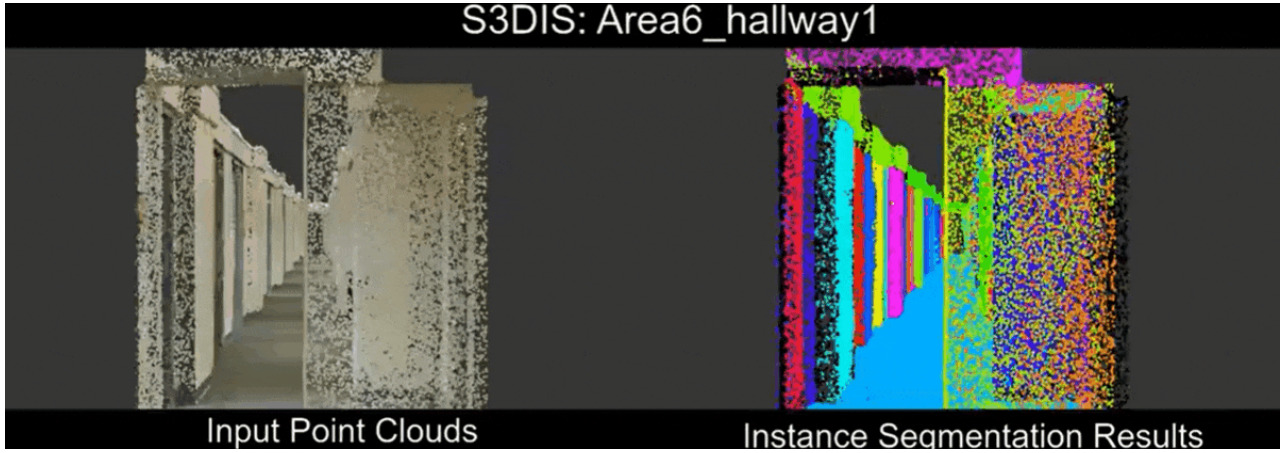
Fully-supervised Methods

Instance Segmentation of Point Clouds









Key Challenges

- Hard/Infeasible to collect 3D labels for supervision;
- Hard to generalize well to novel 3D scenarios;

Learning geometries and semantics from fewer supervision signals?

Weakly-supervised Methods

03 Weakly-supervised Methods

Semantic Segmentation of Point Clouds

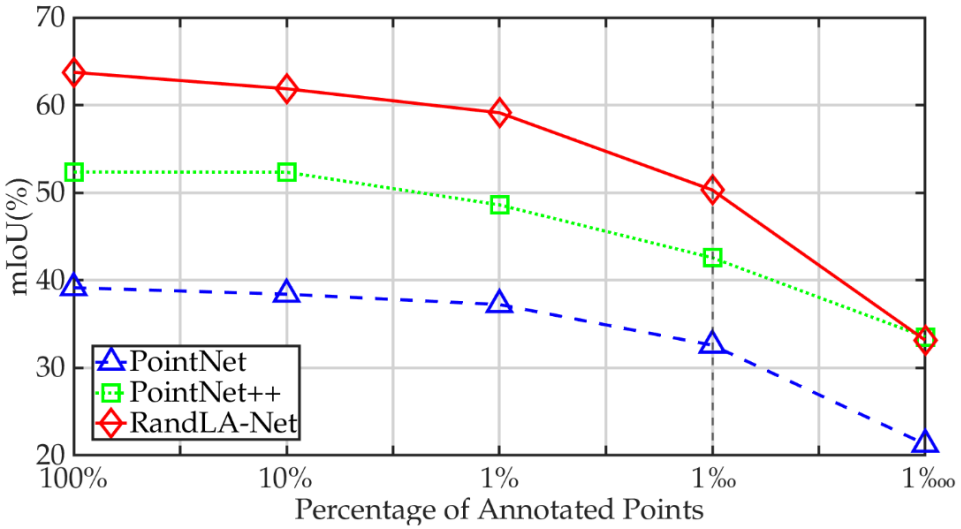
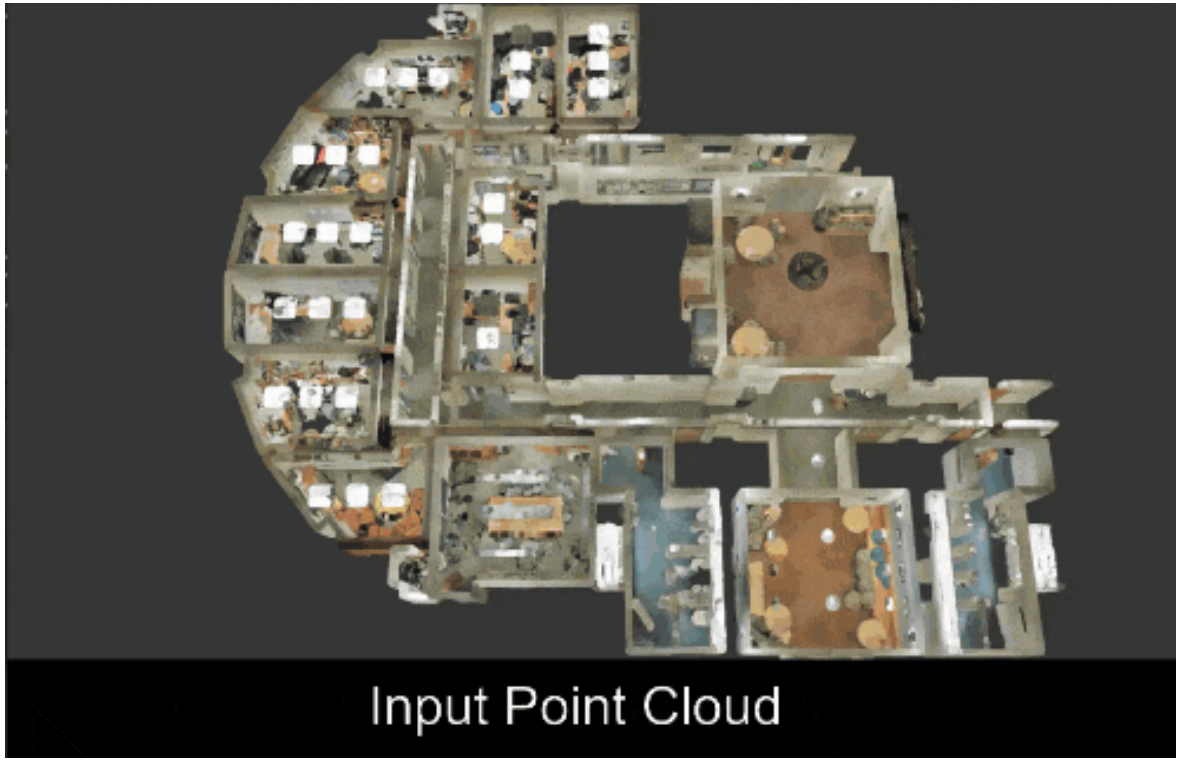


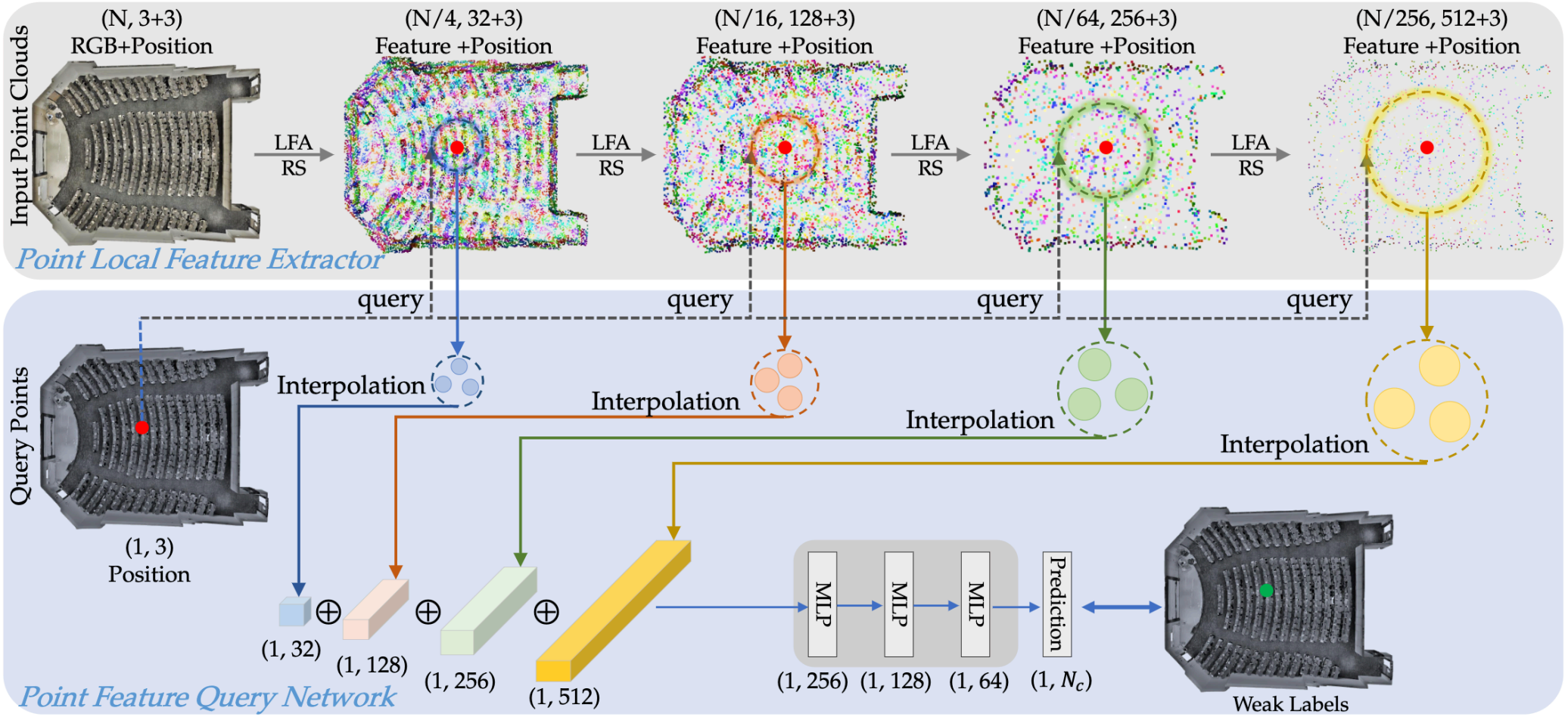
Figure 2: Benchmark results of three baselines in the *Area-5* of the S3DIS [2] dataset. Different amount of points are randomly annotated for weak supervision. (The horizontal axis uses a logarithmic scale).



03

Weakly-supervised Methods

Semantic Segmentation of Point Clouds

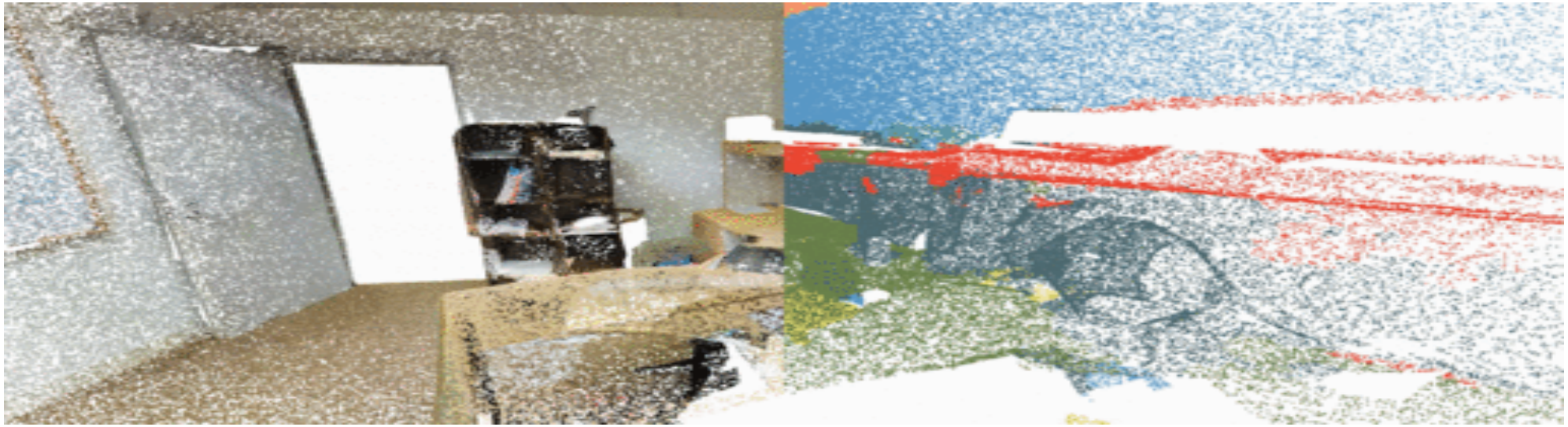


Q Hu, B Yang*, G Fang, Y Guo, A Leonardis, N Trigoni, A Markham, *"SQN: Weakly-Supervised Semantic Segmentation of Large-Scale 3D Point Clouds with 1000x Fewer Labels"*, arXiv, 2019

03

Weakly-supervised Methods

Semantic Segmentation of Point Clouds

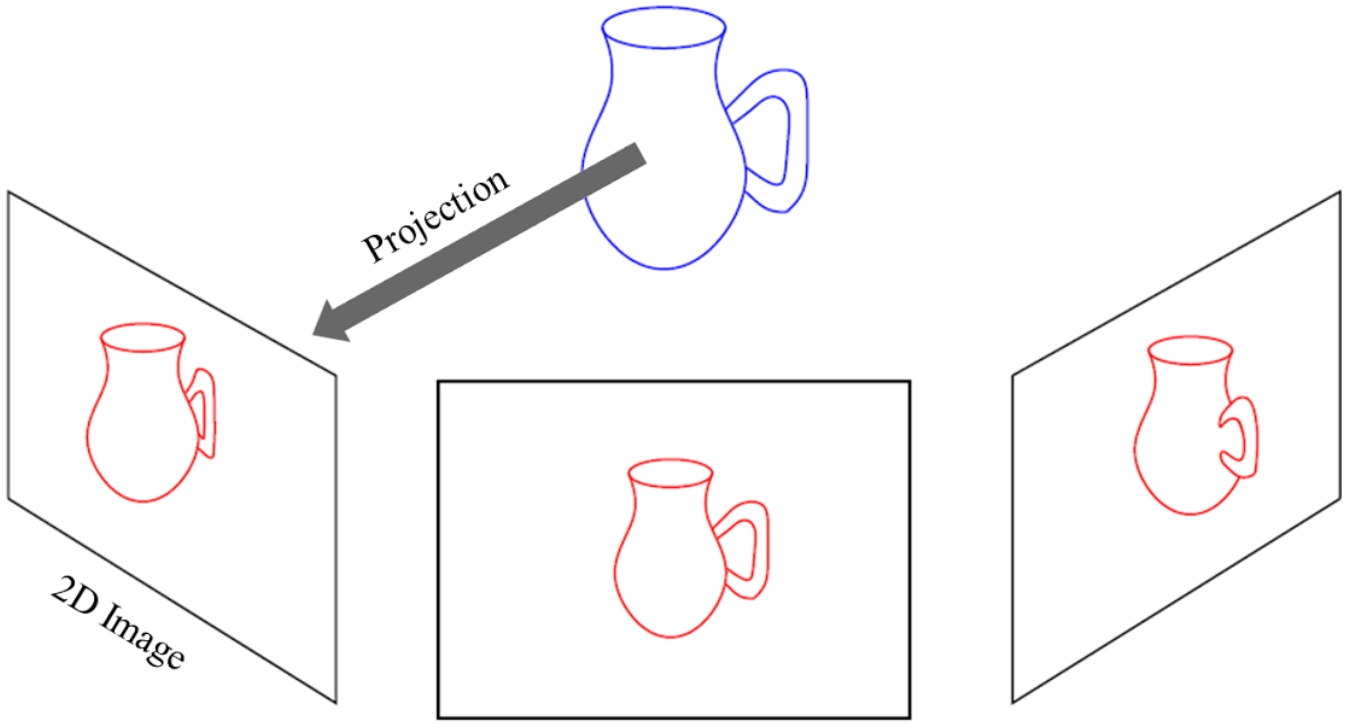


Unsupervised Methods

04

Unsupervised Methods

Multi-view Reconstruction



Multi-view Geometry

04

Unsupervised Methods

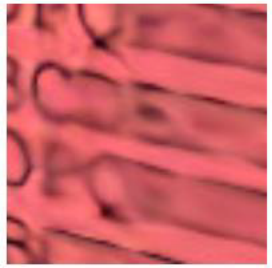
Multi-view Reconstruction

1. SRNs

- Requires optimization on every new scene;
- Global conditioning yields blurriness;

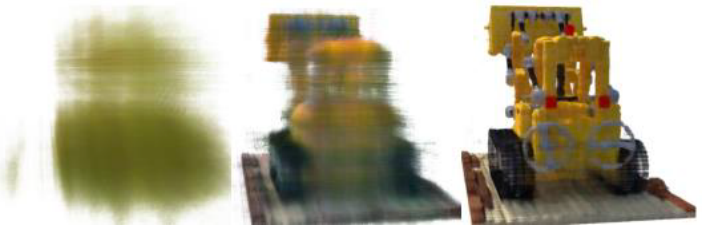
GT

SRNs Rendering



2. NeRF

- Requires retraining on every new scene;
- Relies only on pixelwise supervision;

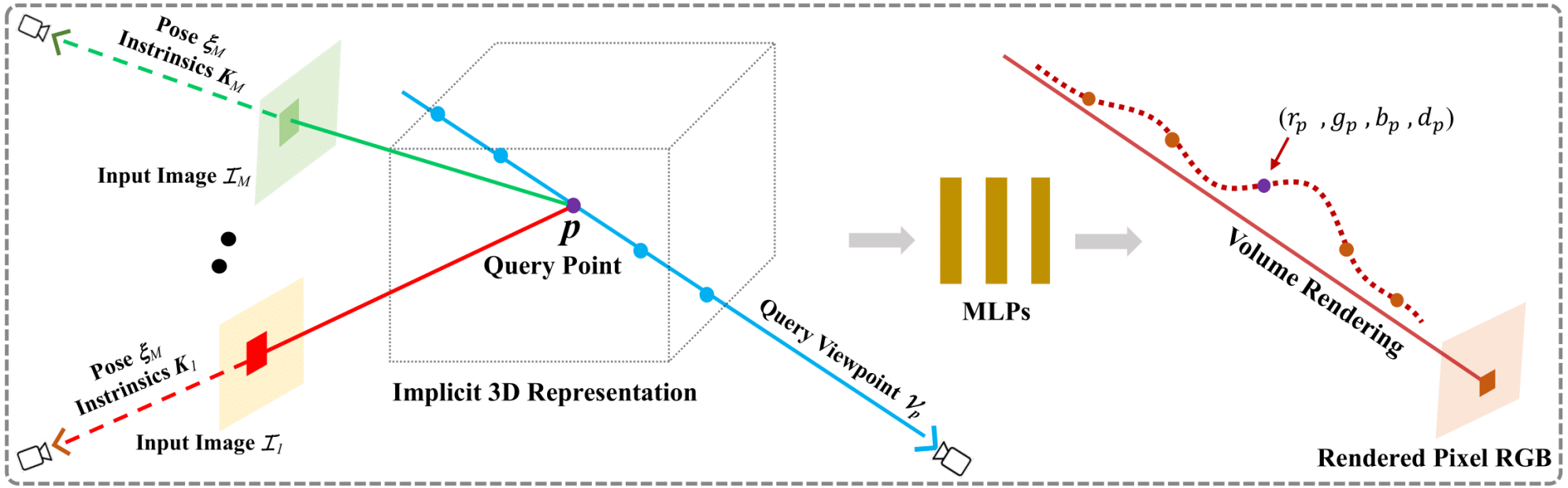


NeRF Rendering Early in Training

04

Unsupervised Methods

Multi-view Reconstruction



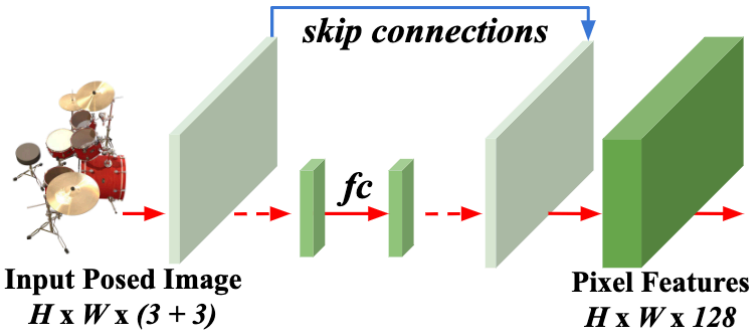
$$(r_p, g_p, b_p, d_p) = f_{\mathbf{W}}\left(\left\{(\mathcal{I}_1, \mathcal{V}_1) \cdots (\mathcal{I}_k, \mathcal{V}_k) \cdots (\mathcal{I}_K, \mathcal{V}_K)\right\}, \{x_p, y_p, z_p\}, \{x_p^v, y_p^v, z_p^v\}\right)$$

A Trevithick, B Yang, *"GRF: Learning a General Radiance Field for 3D Representation and Rendering", ICCV, 2021*

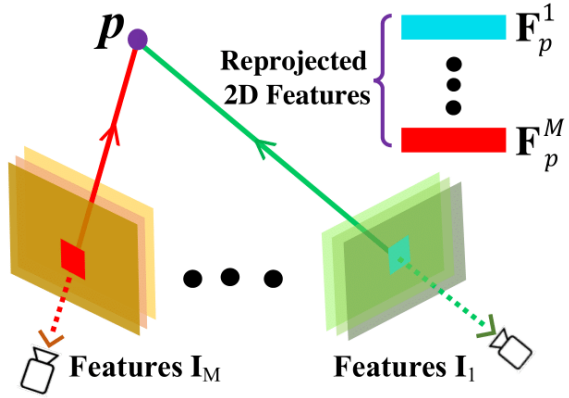
04

Unsupervised Methods

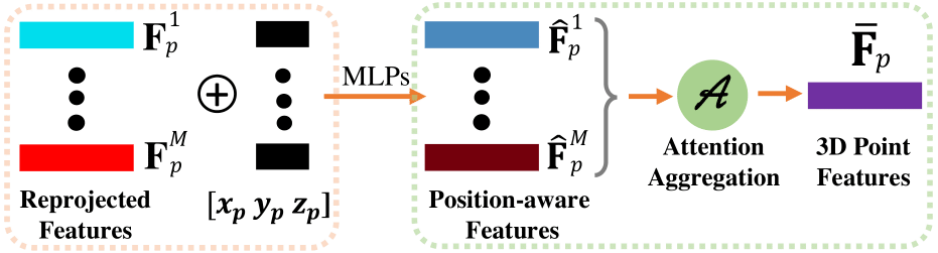
Multi-view Reconstruction



1

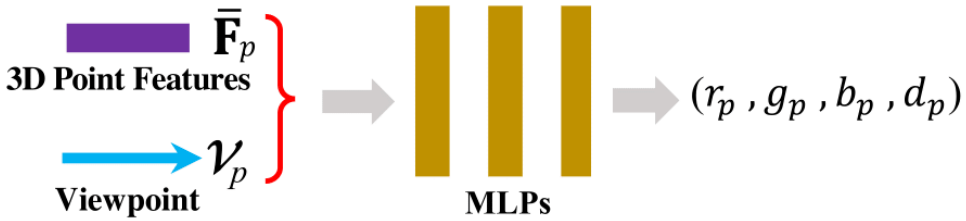


2



3

4

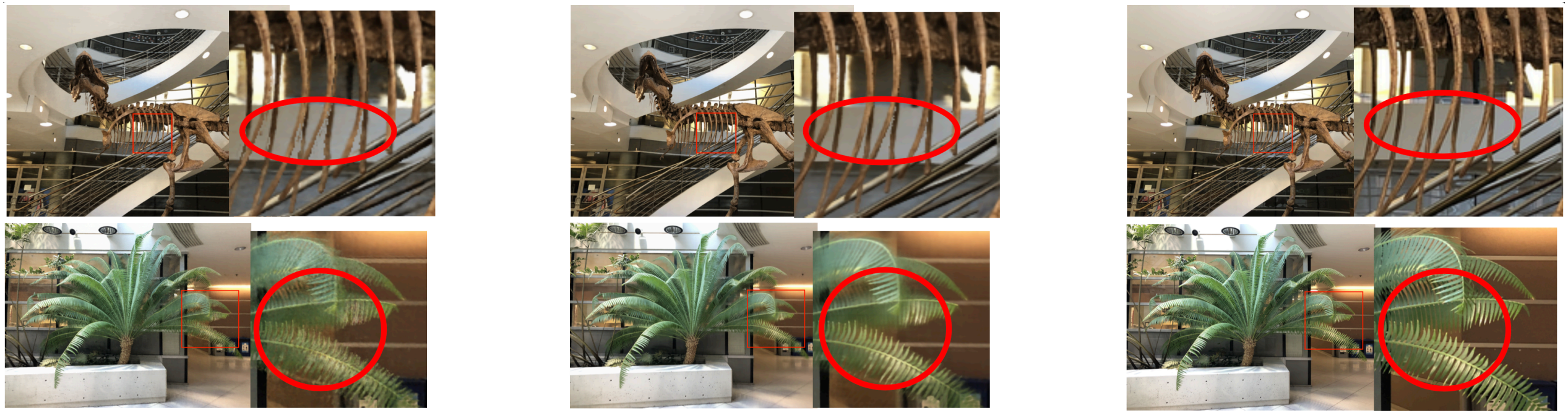


$$rgb = \int_0^{+\infty} T(t)\boldsymbol{c}(t)\sigma(t)dt, \quad T(t) = exp(-\int_0^t \sigma(s)ds)$$

04

Unsupervised Methods

Multi-view Reconstruction



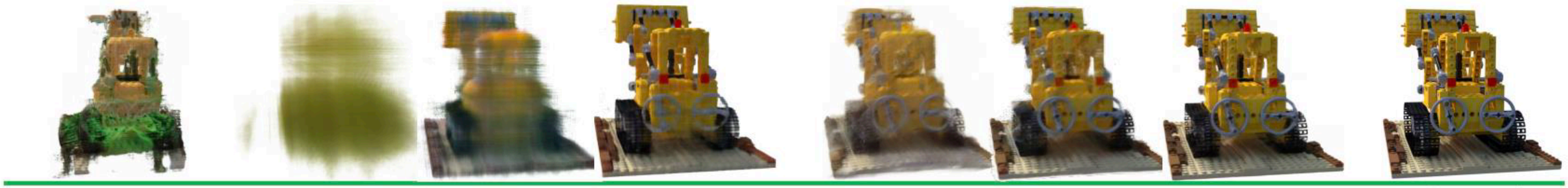
NeRF | GRF (Ours) | Ground Truth

Finer Details

04

Unsupervised Methods

Multi-view Reconstruction



GRF (Group 1) (Generalization) | NeRF (Group 2) (Trained 100, 1k, 10k iters) | **GRF (Group 2)** (Finetuned 100, 1k, 10k iters) | Ground Truth

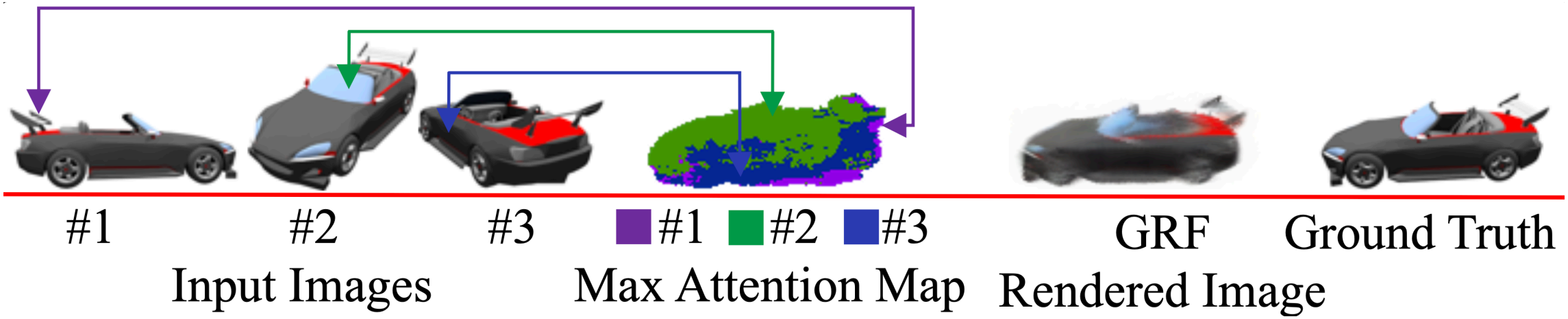
Figure 8: Qualitative results of our GRF for novel scene generalization.

Better Generalization

04

Unsupervised Methods

Multi-view Reconstruction

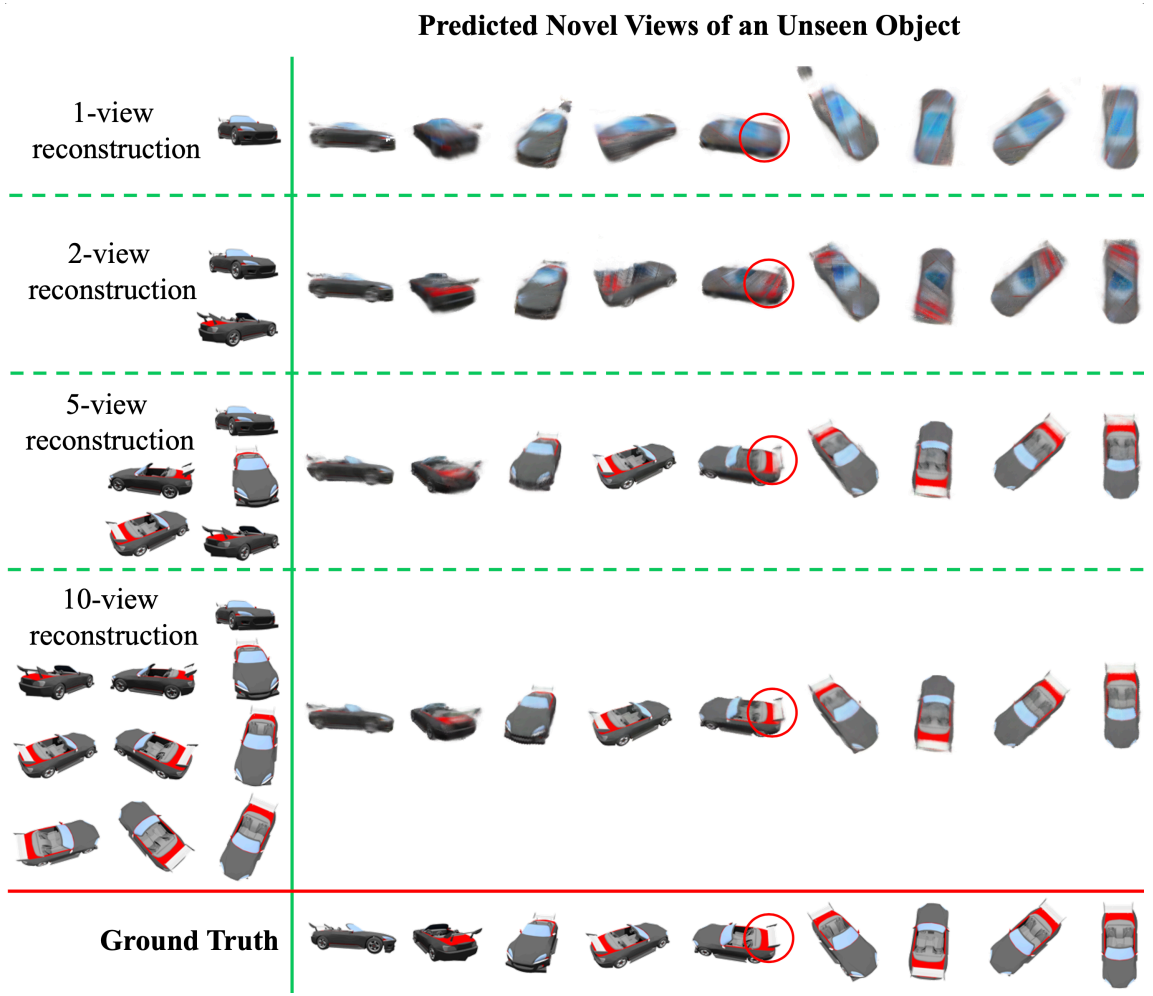


Can Handle Visual Occlusion

04

Unsupervised Methods

Multi-view Reconstruction



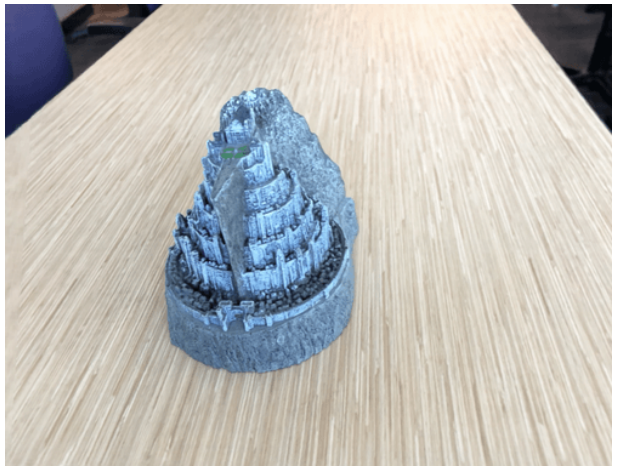
Robust to an Arbitrary Number of Input Views

04

Unsupervised Methods

Multi-view Reconstruction








Summary

- Dense human annotations may not be necessary.
- Geometric constraints and prior knowledge is the key.

Future Work

- Generalization to unseen objects, classes, and scenes.
- Unsupervised learning from geometric constraints.

05

Q/A