ECCV'20 ONLINE

23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON

COMPUTER VISION

WWW.ECCV2020.EU

Learning 3D Part Assembly from A Single Image

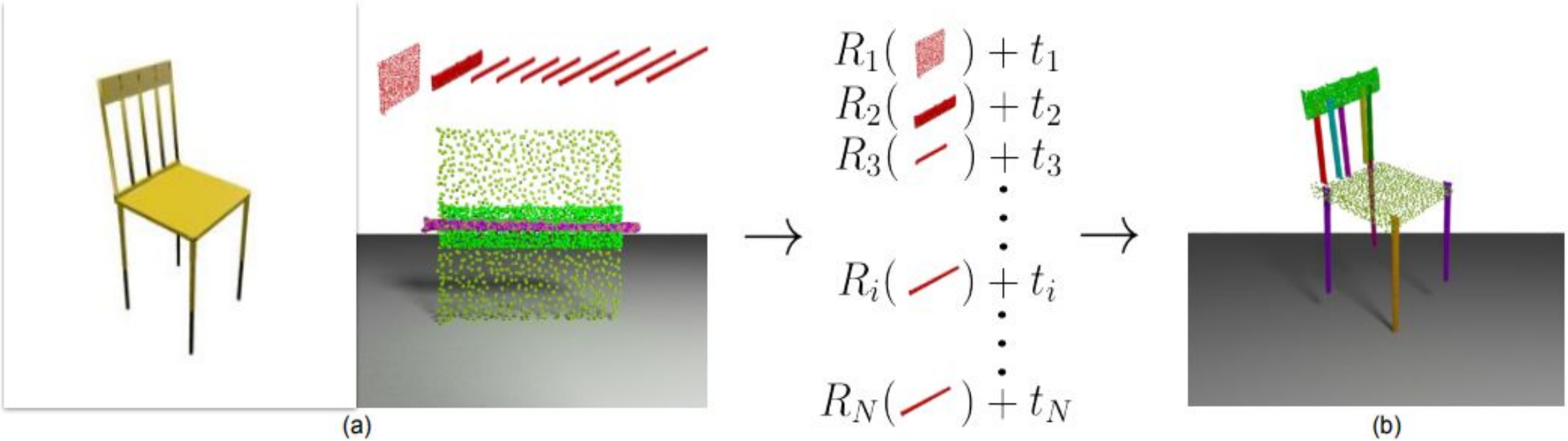
Yichen Li*[1], Kaichun Mo*[1], Lin Shao[1], Minhyuk Sung[2], Leonidas Guibas[1]

Stanford University [1]

Adobe [2]

Problem Definition

Image-Guided Shape Structure Decoding



Input Image | Input Part Geometry | Structure Decoding | Assembled Shape

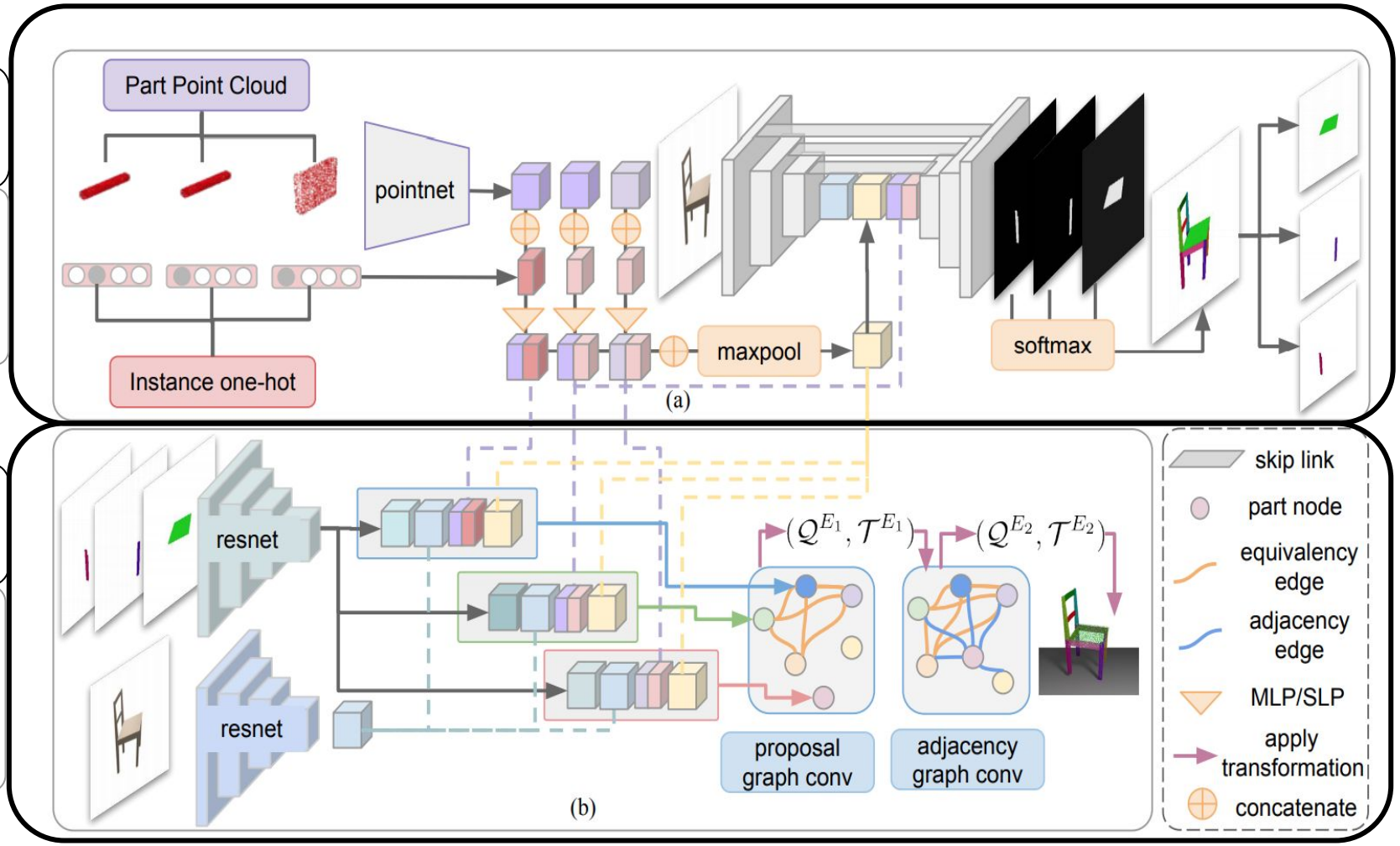
Approach

segmentation

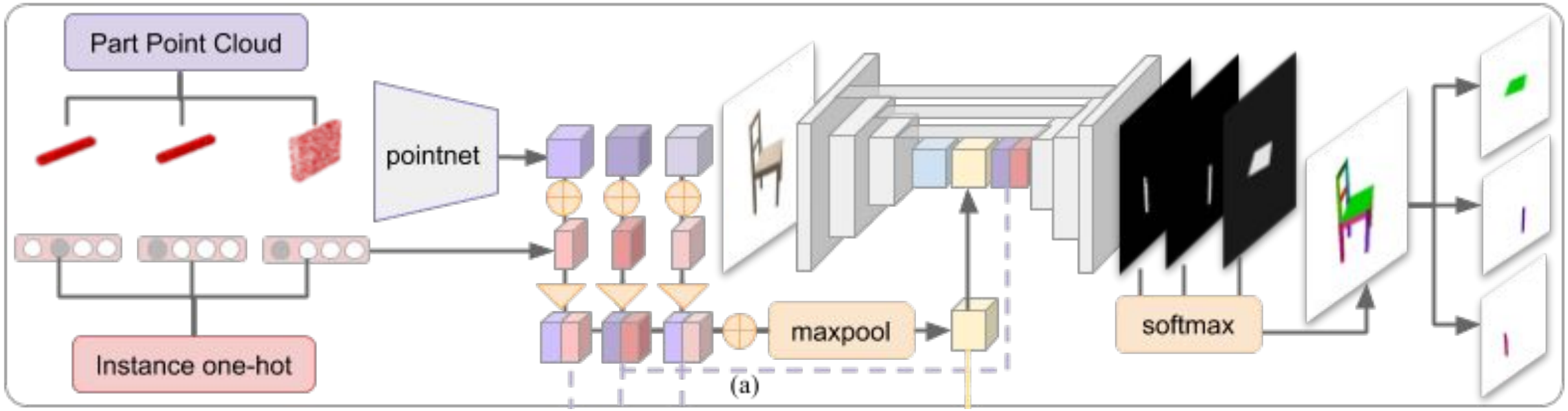
Geometry Conditioned Image Segmentation

pose prediction

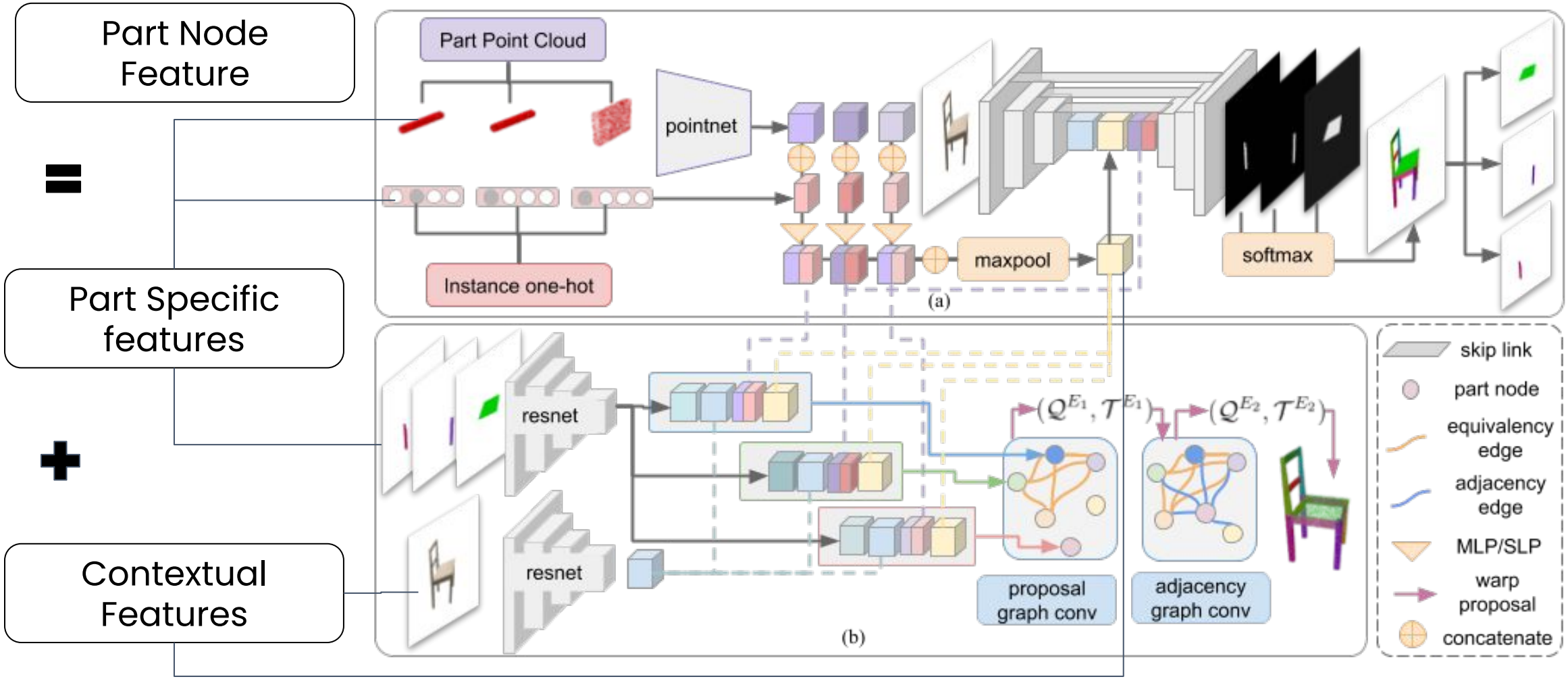
Graph Decoding of Part Geometry Conditioned Shape Structure



Instance Segmentation



Graph-Conv Pose Induction



Loss Functions

$$\mathcal{L}_{mask_i} = -\frac{\sum_{u,v\in[m,m]} \hat{M}_i^{(u,v)} \cdot M_{\mathcal{M}(i)}^{(u,v)}}{\sum_{u,v\in[m,m]} \left(\hat{M}_{\mathcal{M}(i)}^{(u,v)} + M_i^{(u,v)} - \hat{M}_{\mathcal{M}(i)}^{(u,v)} \cdot M_i^{(u,v)}\right)}.$$

$$\mathcal{L}_{pose} = \sum_{i=1}^{N} (\lambda_1 \times \mathcal{L}_T + \lambda_2 \times \mathcal{L}_C + \lambda_3 \times \mathcal{L}_E) + \lambda_4 \times \mathcal{L}_W$$

L2 Translation loss

Shape Holistic chamfer distance loss

L2 Rotation loss

Chamfer distance Rotation loss

Evaluation Metric

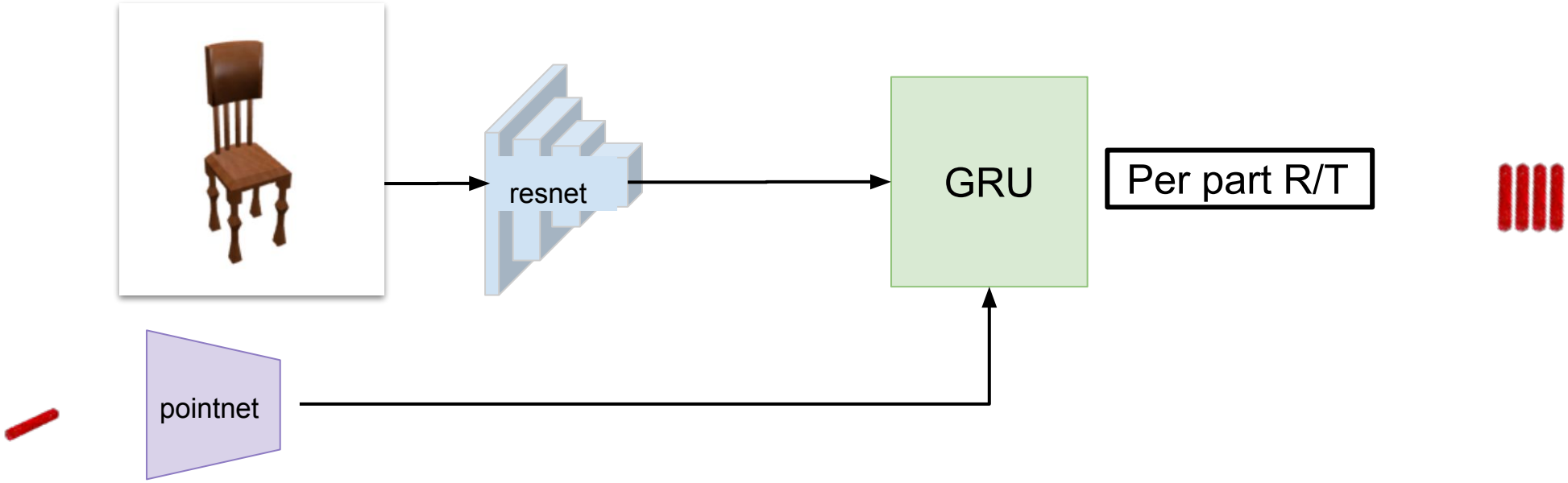
Part Accuracy

$$PA = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\left(\left\|(\hat{R}_{\mathcal{M}(i)}(p_i) + \hat{t}_i) - (R_i(p_i) + t_i)\right\|_{chamfer} < \tau\right)$$

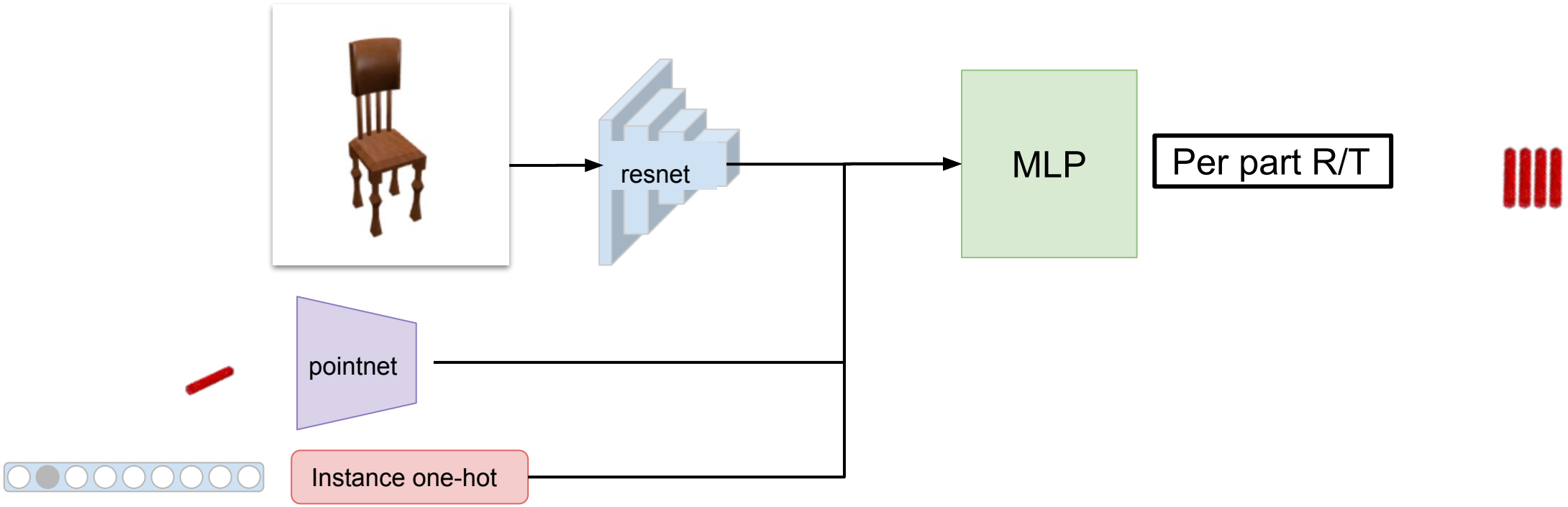
Shape Chamfer Distance

$$SC(S, \hat{S}) = \frac{1}{N \cdot d_{pc}} \sum_{x \in \hat{S}} \min_{y \in S} \|x - y\|_2 + \frac{1}{N \cdot d_{pc}} \sum_{y \in S} \min_{x \in \hat{S}} \|x - y\|_2$$

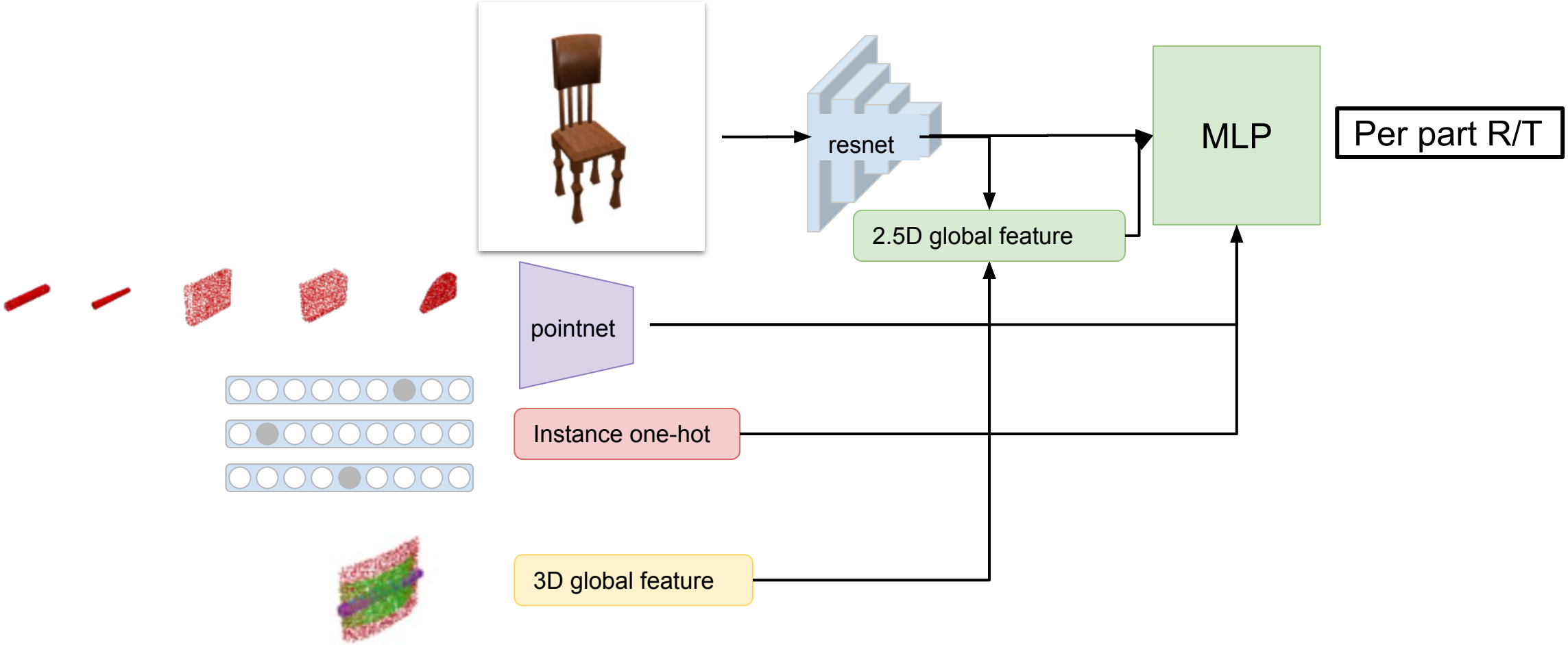
Baseline 1 - GRU



Baseline 2 - InsOneHot



Baseline 3 - GlobalFeat

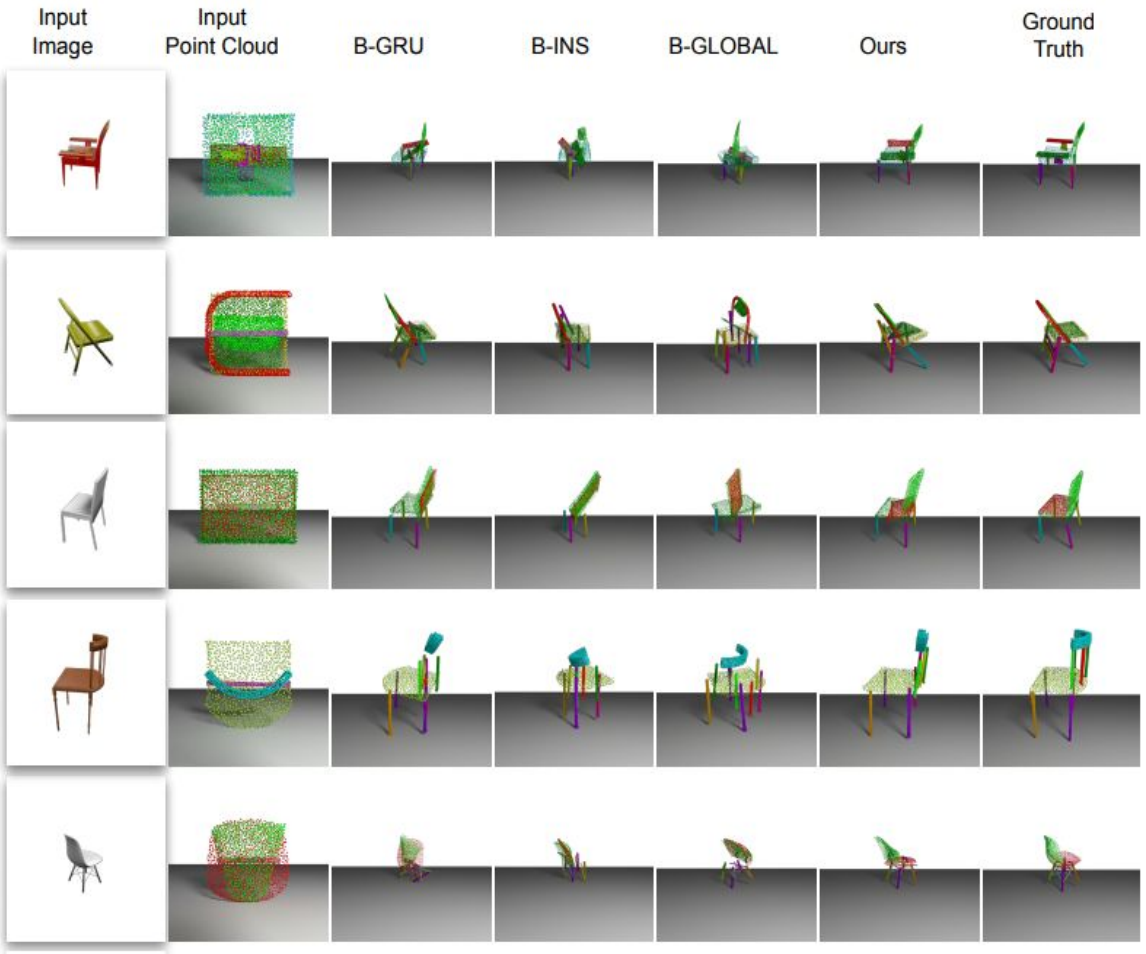
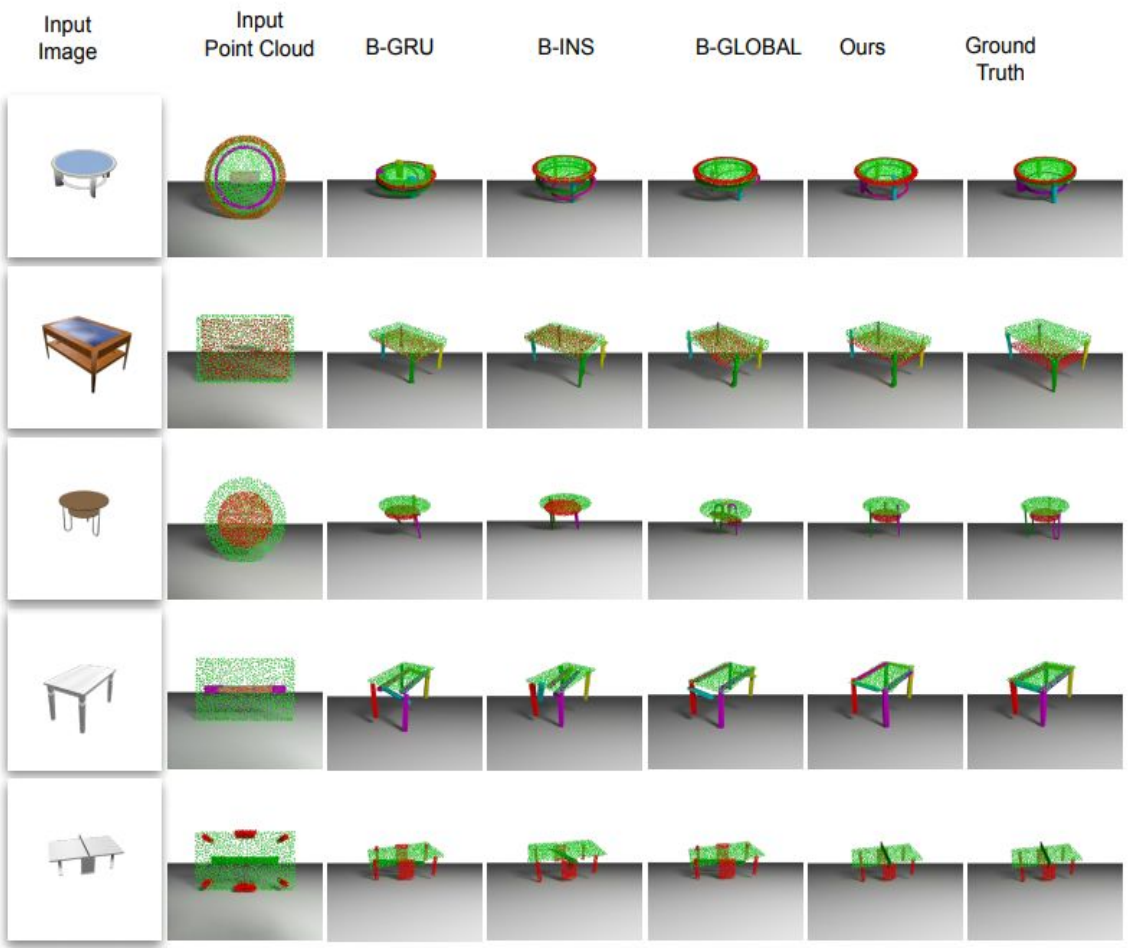


Quantitative Results

Table 1: Part Accuracy and Assembly Chamfer Distance(CD)

Modality	Method	Part Accuracy ↑			Assembly CD ↓		
		Chair	Table	Cabinet	Chair	Table	Cabinet
Level-3	B-GRU	0.310	0.574	0.334	0.107	0.057	0.062
	B-InsOnehot	0.173	0.507	0.295	0.130	0.064	0.065
	B-Global	0.170	0.530	0.339	0.125	0.061	0.065
	Ours	**0.454**	**0.716**	**0.402**	**0.067**	**0.037**	**0.050**
Mixed	B-GRU	0.326	0.567	0.283	0.101	0.070	0.066
	B-InsOnehot	0.286	0.572	0.320	0.108	0.067	0.061
	B-Global	0.337	0.619	0.290	0.093	0.062	0.0677
	Ours	**0.491**	**0.778**	**0.483**	**0.065**	**0.037**	**0.043**

Table 2: Visible and Invisible Part Accuracies

Modality	Method	Part Accuracy (Visible) ↑			Part Accuracy (Invisible) ↑		
		Chair	Table	Cabinet	Chair	Table	Cabinet
Level-3	B-GRU	0.3182	0.598	0.353	0.206	0.481	0.304
	B-InsOnehot	0.178	0.572	0.291	0.104	0.369	0.289
	B-Global	0.174	0.563	0.354	0.120	0.427	0.269
	Ours	**0.471**	**0.753**	**0.455**	**0.270**	**0.557**	**0.358**
Mixed	B-GRU	0.335	0.593	0.302	0.180	0.267	0.258
	B-InsOnehot	0.295	0.592	0.346	0.133	0.275	0.279
	B-Global	0.334	0.638	0.320	0.184	0.349	0.227
	Ours	**0.505**	**0.803**	**0.537**	**0.262**	**0.515**	**0.360**

Qualitative Results