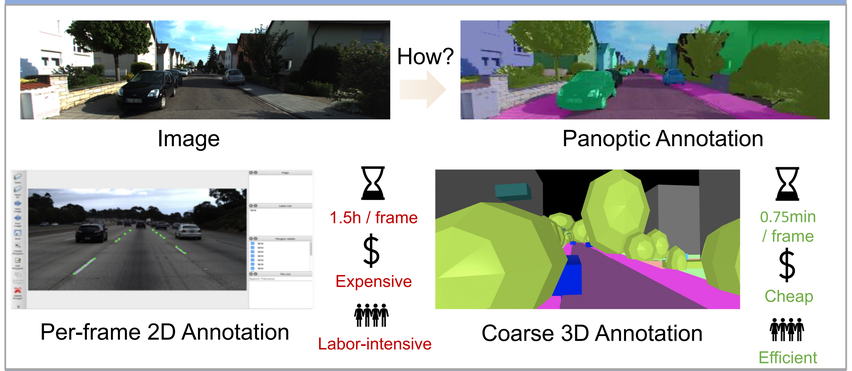
Panoptic NeRF : 3D-to-2D Label Transfer for Panoptic Urban Scene Segmentation

Xiao Fu[1*], Shangzhan Zhang[1*], Tianrun Chen[1], Yichong Lu[1], Lanyun Zhu[2], Xiaowei Zhou[1], Andreas Geiger[3] , Yiyi Liao[1]

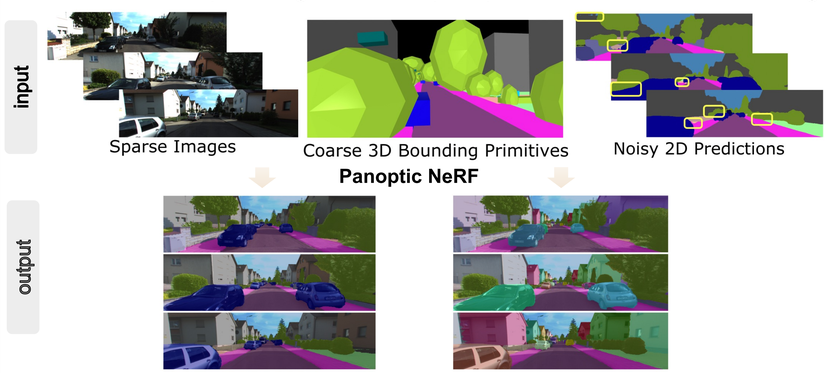
[1]Zhejiang University, [2]Singapore University of Technology and Design, [3] University of Tübingen and MPI-IS

Motivation



Introduction

Idea : Infer in 3D space to generate 2D panoptic labels via volume rendering



Multi-view and Spatio-temporally Consistent Semantic and Panoptic Segmentations

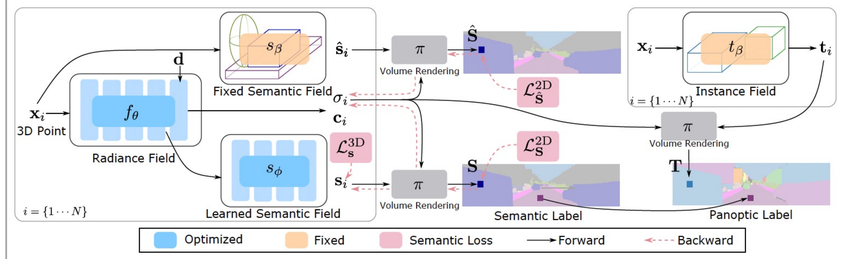
Contributions

- End-to-end 3D-to-2D label transfer in terms of volume rendering
- State-of-the-art performance compared to existing label transfer methods
- Enable rendering semantic/instance labels at novel viewpoints

Challenges

- **Poor Geometry** - With sparse inputs views in driving scenario, vanilla NeRF suffers from poor geometry reconstruction that leads to wrong panoptic maps
- **Label Ambiguity** - At overlapping regions of the 3D bounding primitives, label ambiguity yields inaccurate 2D labels

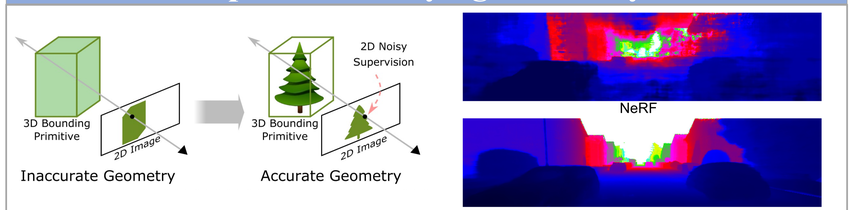
Method Overview



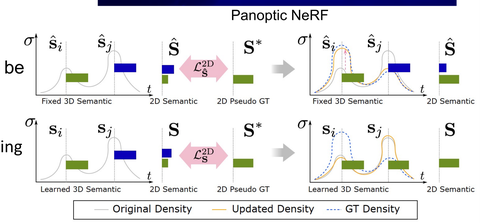
Fixed Semantic Field $s_\beta : \mathbf{x} \in \mathbb{R}^3 \mapsto \hat{\mathbf{s}} \in \mathbb{R}^{M_s}$ Learned Semantic Field $s_\phi : \mathbf{x} \in \mathbb{R}^3 \mapsto \mathbf{s} \in \mathbb{R}^{M_s}$

Radiance Field $f_\theta : (\mathbf{x} \in \mathbb{R}^3, \mathbf{d} \in \mathbb{S}^2) \mapsto (\sigma \in \mathbb{R}^+, \mathbf{c} \in \mathbb{R}^3)$ Instance Field $t_\beta : \mathbf{x} \in \mathbb{R}^3 \mapsto \mathbf{t} \in \mathbb{R}^{M_t}$

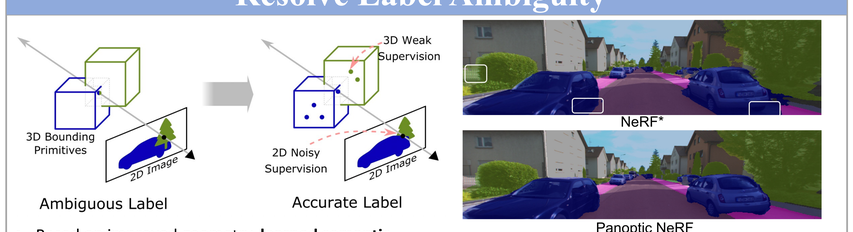
Improve Underlying Geometry



- By applying $\mathcal{L}_{\hat{\mathbf{S}}}^{2D}$ to $\hat{\mathbf{S}}$ rendered via **fixed semantic field**, underlying geometry can be improved
- In contrast, $\mathcal{L}_{\mathbf{S}}^{2D}$ applied to $\mathbf{S}$ rendered via **learned semantic field** may "cheat" by updating the semantics



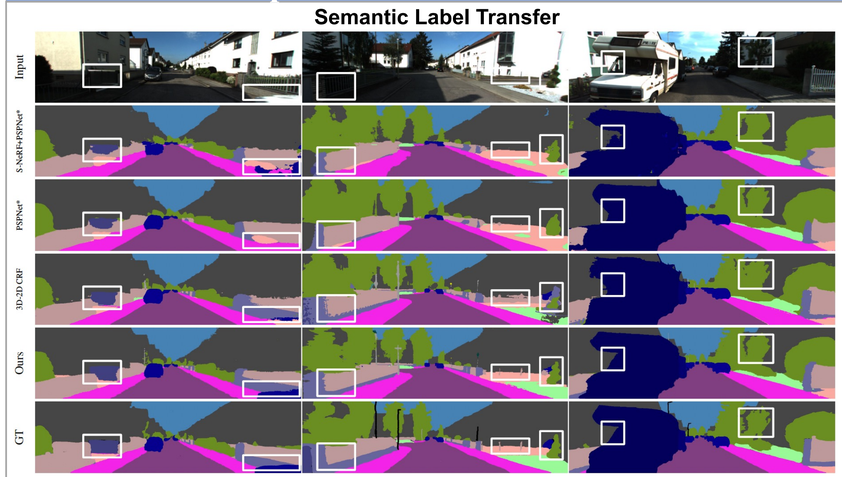
Resolve Label Ambiguity



- Based on improved geometry, **learned semantic field** further resolves label ambiguity at overlapping regions of 3D bounding primitives via 2D and 3D weak supervisions
- 2D pseudo-label usually performs well on frequently occurring classes, 3D supervision further suppresses noise of 2D predictions

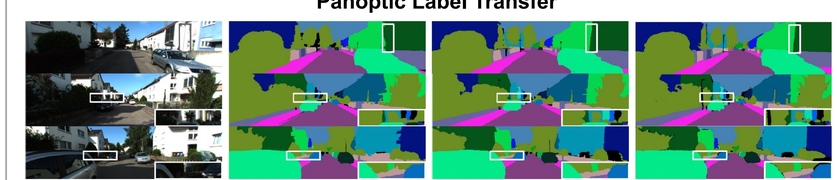
Experiments on KITTI-360

Semantic Label Transfer



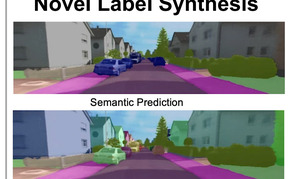
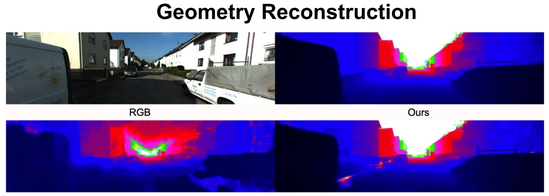
Method	Road	Park	Sdwlk	Terr	Bldg	Vegt	Car	Trler	Crvn	Gate	Wall	Fence	Box	Sky	mIoU	Acc	MC
FC CRF + Manual GT	90.3	49.9	67.7	62.5	88.3	79.2	85.6	48.9	78.1	23.4	35.3	46.5	42.0	92.7	63.6	89.1	85.41
S-NeRF + Manual GT	87.0	35.8	64.7	58.2	83.4	76.3	70.3	93.5	76.5	41.4	44.0	52.6	29.0	92.0	64.6	86.8	88.98
S-NeRF + PSPNet*	94.6	52.9	77.7	65.0	88.0	80.8	87.9	58.3	86.0	36.0	44.1	56.8	42.2	90.9	68.6	90.5	93.42
PSPNet*	95.5	49.7	77.5	66.7	88.9	82.4	91.6	46.5	83.1	24.2	43.3	51.3	51.1	89.3	67.2	90.7	91.79
3D Primitives + GC	81.7	31.0	45.6	22.5	59.6	56.7	63.0	61.7	37.3	61.6	28.8	50.6	39.5	50.3	49.3	73.4	86.56
3D Mesh + GC	91.7	53.1	67.2	31.4	81.3	72.1	85.2	93.5	86.0	65.2	40.7	59.7	54.4	65.6	67.7	86.0	94.99
3D Point + GC	93.5	59.0	76.1	37.2	82.0	74.1	87.5	94.7	85.7	66.7	59.4	65.9	58.6	68.0	72.0	87.9	**96.51**
3D-2D CRF	95.2	64.2	83.8	67.9	90.3	84.2	92.2	93.4	90.8	68.2	**64.5**	70.0	55.8	**92.8**	79.5	92.8	94.98
Ours	**95.6**	**68.4**	**84.1**	**69.5**	**91.0**	**84.4**	**93.0**	**94.8**	**93.7**	**71.6**	63.1	**74.4**	**59.7**	91.7	**81.1**	**93.2**	95.02

Panoptic Label Transfer



Method		PQ	SQ	RQ	PQ†	Method		PQ	SQ	RQ	PQ†
3D-2D CRF	All	62.2	79.1	76.9	64.9	Ours	All	**64.4**	**79.3**	**79.6**	**66.9**
	Things	60.7	79.5	**75.2**	60.7		Things	**61.9**	**80.7**	**75.2**	**61.9**
	Stuff	63.0	**78.9**	77.9	67.3		Stuff	**65.8**	78.6	**82.0**	**69.8**

Novel Label Synthesis



Geometry Reconstruction