PEKING UNIVERSITY

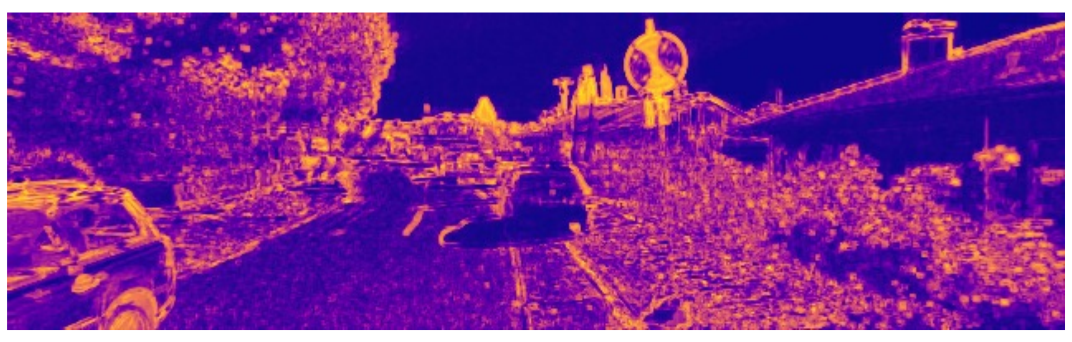
WACV WAIKOLOA HAWAII JAN 3-7 • 2023

Frequency-Aware Self-Supervised Monocular Depth Estimation

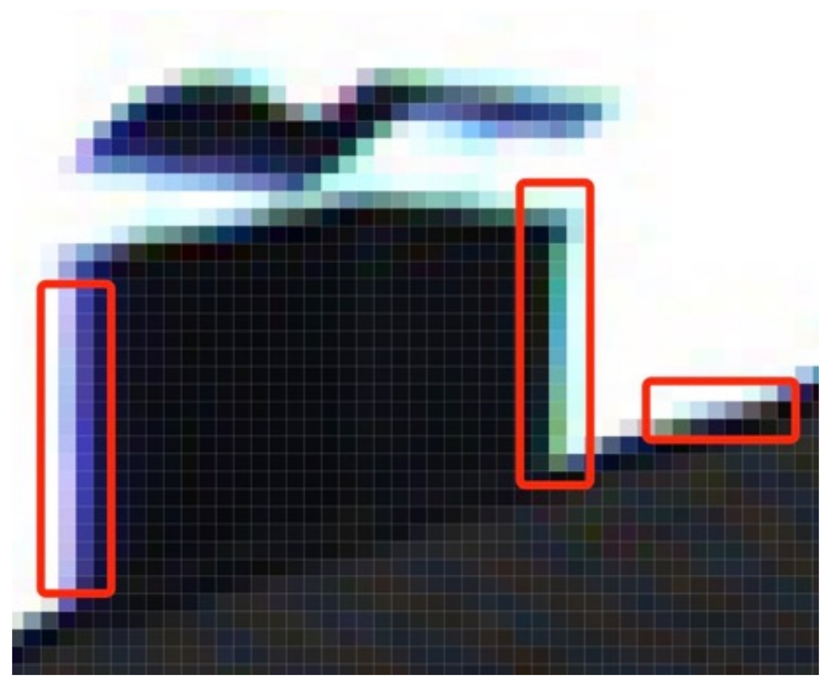
Xingyu Chen[1] Thomas H. Li[1,2,3] Ruonan Zhang[1] Ge Li ✉[1]
[1]SECE, Peking University [2]Advanced Institute of Information Technology, Peking University
[3]Information Technology R&D Innovation Center of Peking University

Does Loss at Object Boundary Makes Sense?

(a) Input RGB Image

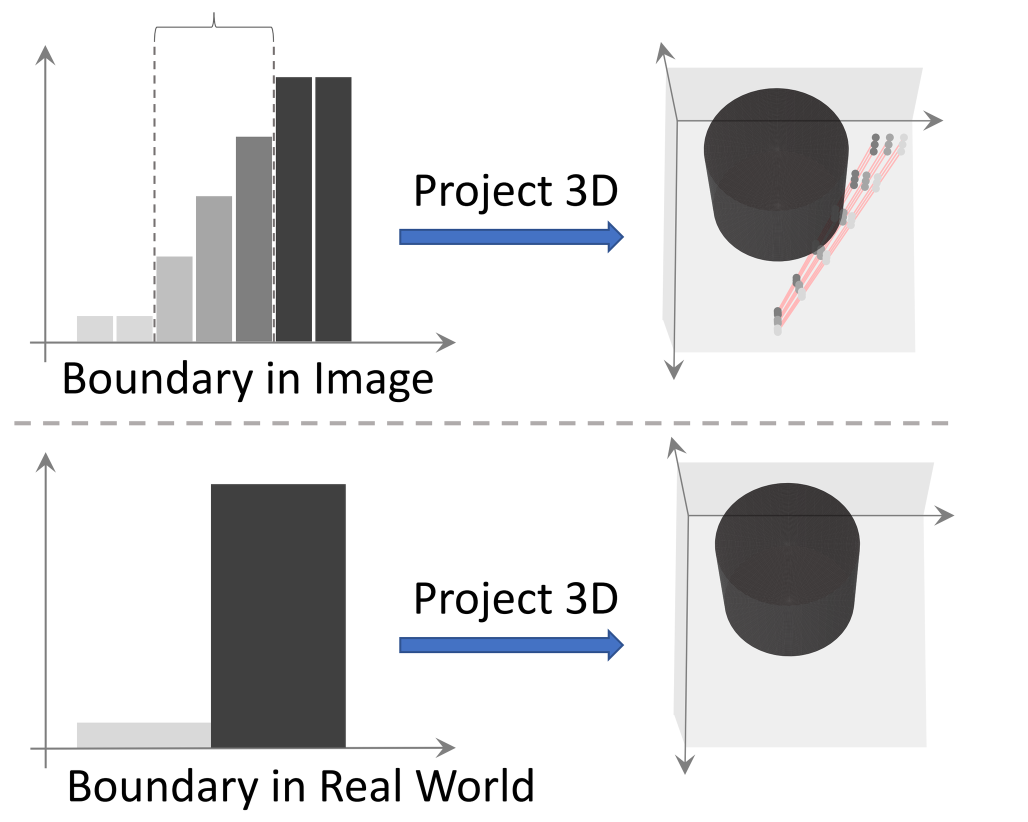


(b) Loss Map from 8/20th epoch



(c) Image Patch

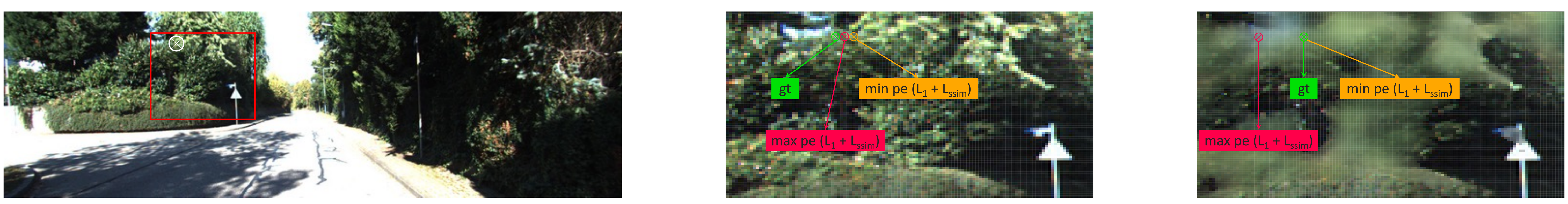
Pixels Not belonging to one deterministic object

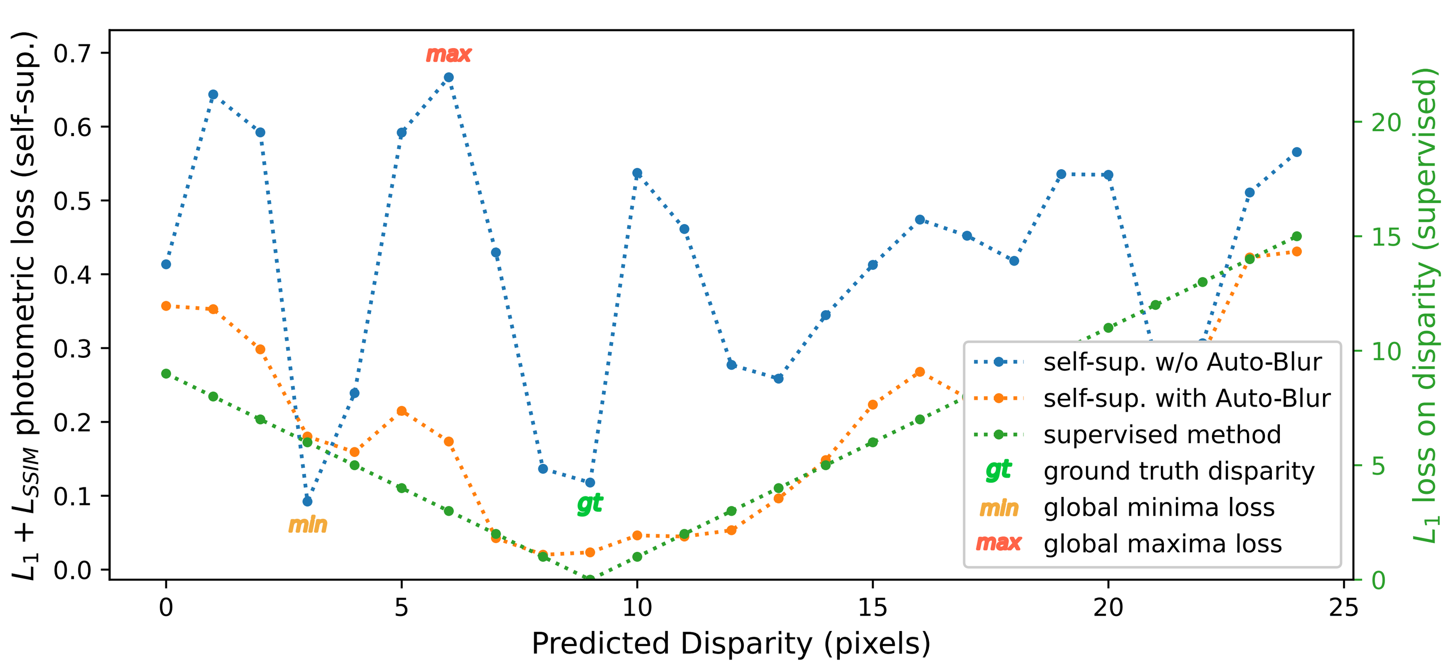


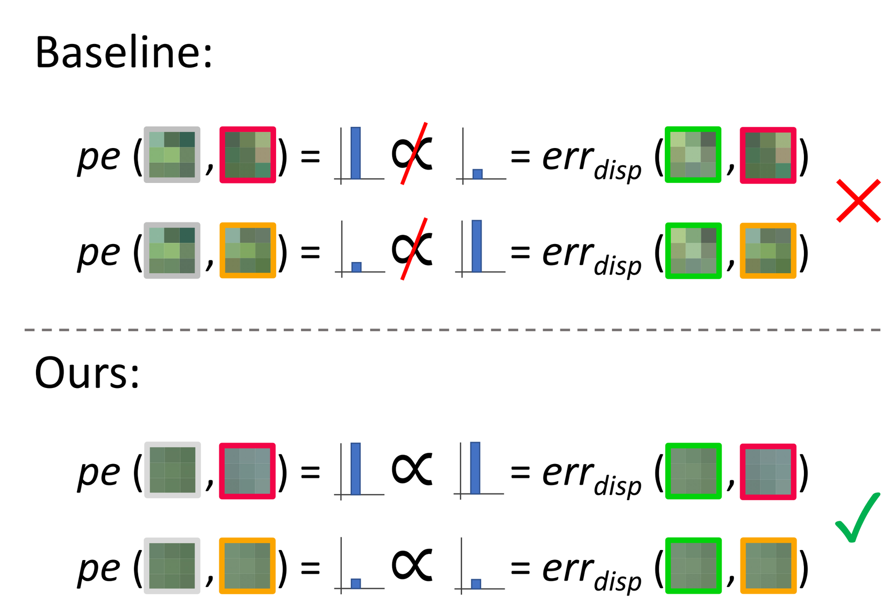
(d) Boundary Comparison

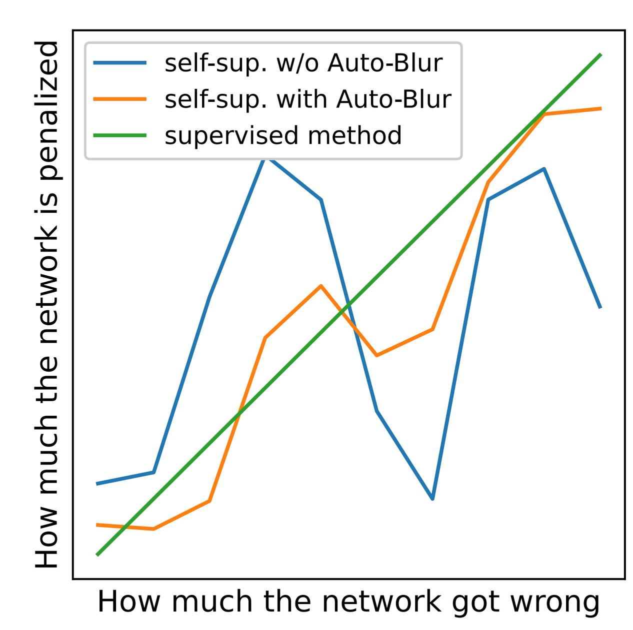
(b) On most objects, losses appear at object boundaries. **(c)** The pixels at the boundaries are gradually changed over the junction. However, these colors are ambiguous, *i.e.*, neither from the black chimney nor the white clouds. **(d)** Object boundaries in the real world are completely mutated, where one single pixel characterizes one deterministic object. However, the ambiguous pixels each contain photometric information for two objects, whereas the network predicts at most one single depth value for them. When projecting the black chimney to 3D point clouds, the ambiguous pixels detach from their main body both spatially and photometrically, regardless of the predicted depths. Hence, no pixels in the synthesized view would match them, resulting in always-large reprojection losses.

Auto-Blur: Help Photometric Loss to be ***Fair*** in High-Frequency Area



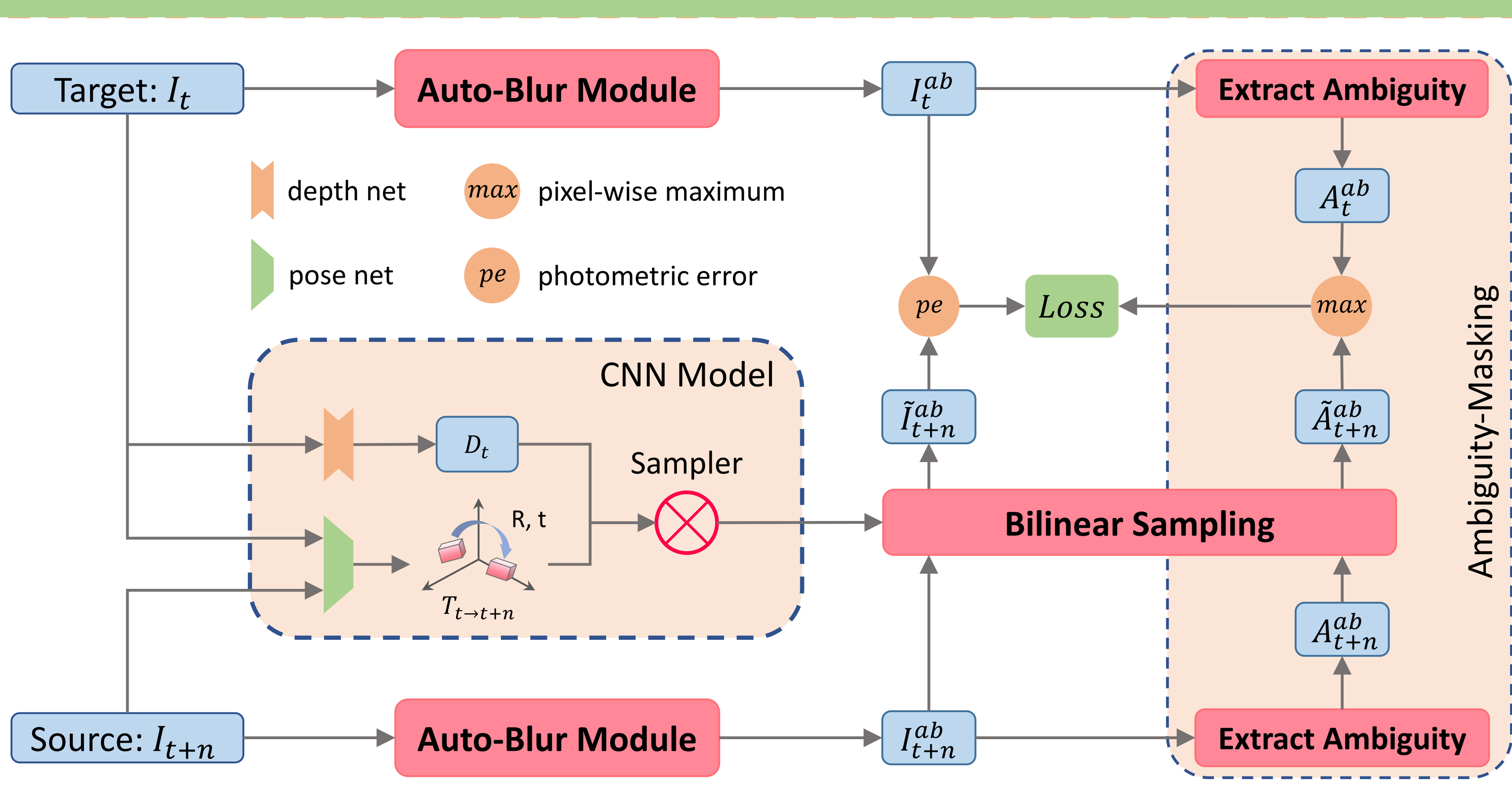






Top: A training image and its crop of the right view (stretched) with and without the proposed Auto-Blur. **Bottom**: Left is the quantitative photometric loss used in self-supervised method with/without Auto-Blur and $\mathcal{L}_1$ loss on predicted disparity used in supervised method. The middle plot (∝: proportional to) shows without Auto-Blur, disparity of *max* $\mathcal{L}_1 + \mathcal{L}_{ssim}$ photometric loss is instead more accurate than that of *min* photometric loss; the photometric loss of ground truth is even larger than some incorrect disparity, while self-supervised method augmented with our Auto-Blur does not suffer from this misjudging. Plot on the right is the qualitative analysis of the relationship between network penalty and prediction error. Supervised method exhibits the *absolutely fair* relationship. With Auto-Blur, $\mathcal{L}_1 + \mathcal{L}_{ssim}$ becomes more stable and gets closer to supervised one.

Overview and Experiments



Method	PP	Data	Extra time	AbsRel	SqRel	RMSE	RMSE log	δ_1	δ_2	δ_3
Monodepth2 no pt [12]	✗	S	-	0.130	1.144	5.485	0.232	0.831	0.932	0.968
+ Ours	✗	S	**+ 0ms**	**0.127**	**1.086**	**5.406**	**0.224**	**0.832**	**0.937**	**0.971**
Monodepth2 M [12]	✗	M	-	0.115	0.903	4.863	0.193	0.877	0.959	0.981
+ Ours	✗	M	**+ 0ms**	**0.112**	**0.834**	**4.746**	**0.189**	**0.880**	**0.961**	**0.982**
Zhou *et al.* [38]	✗	M	-	0.183	1.595	6.709	0.270	0.734	0.902	0.959
+ Ours	✗	M	**+ 0ms**	**0.142**	**1.547**	**5.433**	**0.224**	**0.840**	**0.944**	**0.974**
WaveletMonodepth [27]	✗	S	-	0.109	**0.845**	4.800	0.196	0.870	0.956	**0.980**
+ Ours	✗	S	**+ 0ms**	**0.108**	0.862	**4.786**	**0.194**	**0.875**	**0.957**	**0.980**
Monodepth2 S [12]	✗	S	-	0.109	0.873	4.960	0.209	0.864	0.948	0.975
+ Ours	✗	S	**+ 0ms**	**0.107**	**0.835**	**4.850**	**0.201**	**0.865**	**0.951**	**0.978**
FSRE-Depth [18]	✗	M	-	**0.105**	0.722	4.547	0.182	**0.886**	**0.964**	**0.984**
+ Ours	✗	M	**+ 0ms**	**0.105**	**0.711**	**4.452**	**0.181**	**0.886**	**0.964**	**0.984**
Monodepth2 MS [12]	✗	MS	-	**0.106**	0.818	4.750	0.196	0.874	0.957	0.979
+ Ours	✗	MS	**+ 0ms**	**0.106**	**0.797**	**4.672**	**0.187**	**0.887**	**0.961**	**0.982**
CADepth [35]	✗	S	-	0.107	0.849	4.885	0.204	0.869	0.951	0.976
+ Ours	✗	S	**+ 0ms**	**0.106**	**0.823**	**4.835**	**0.201**	**0.870**	**0.953**	**0.977**
Depth-Hints [32]	✗	S	-	0.109	0.845	4.800	0.196	0.870	0.956	0.980
+ Ours	✗	S	**+ 0ms**	**0.105**	**0.811**	**4.695**	**0.192**	**0.875**	**0.958**	**0.981**