SharpNet: Fast and Accurate Recovery of Occluding Contours in Monocular Depth Estimation

Michaël Ramamonjisoa[1] Vincent Lepetit[1]

[1] LIGM (UMR 8049), École des Ponts, UPE

https://github.com/MichaelRamamonjisoa/SharpNet

ICCV 2019 Seoul, Korea

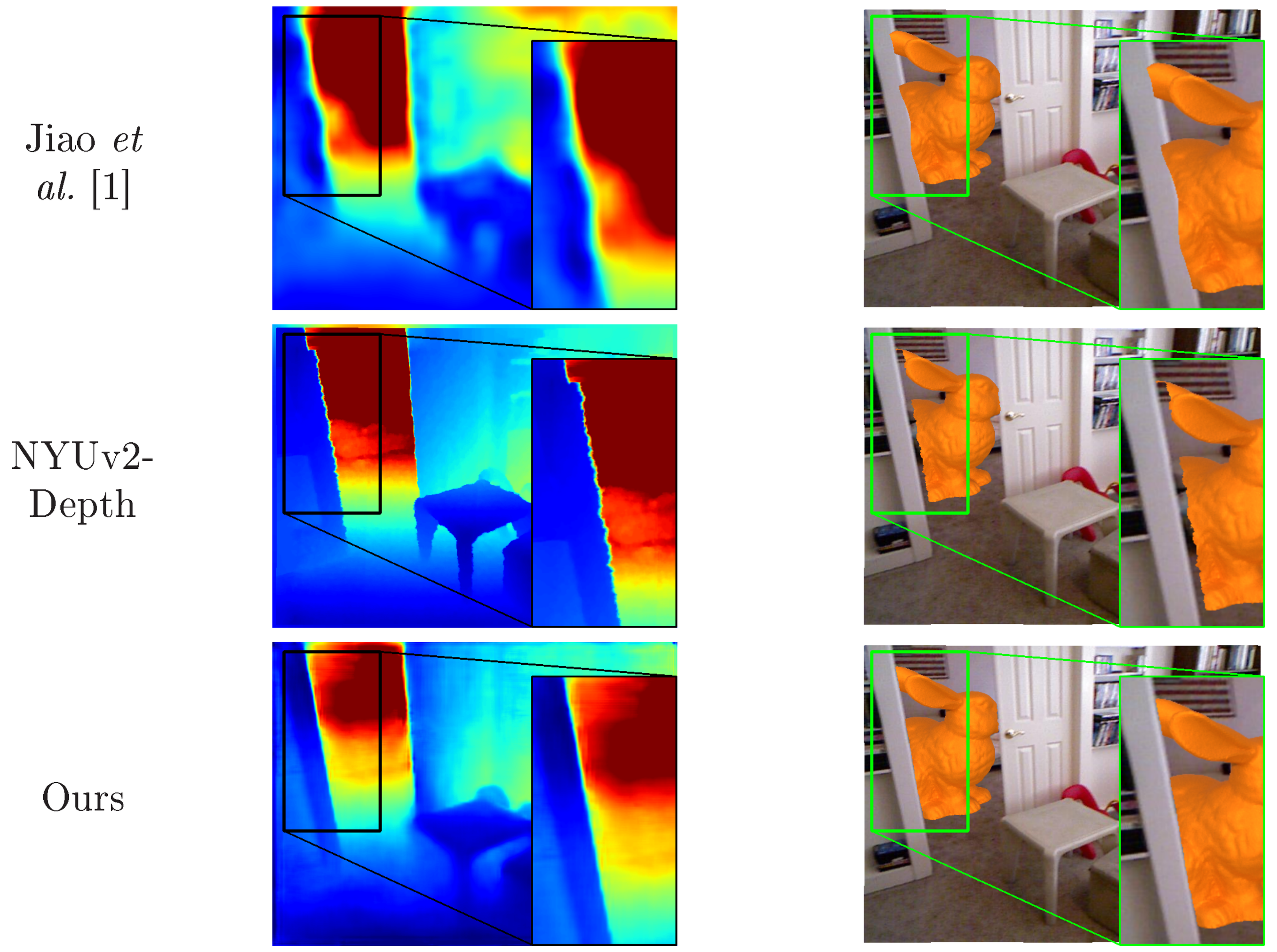
Overview

Objective

- Reconstruct surface depth from a single image with high accuracy along occluding contours

Occluding Contour (OC) $\leftrightarrow$ Depth Boundary Edge (DBE)

- Application to Augmented Reality



Contributions

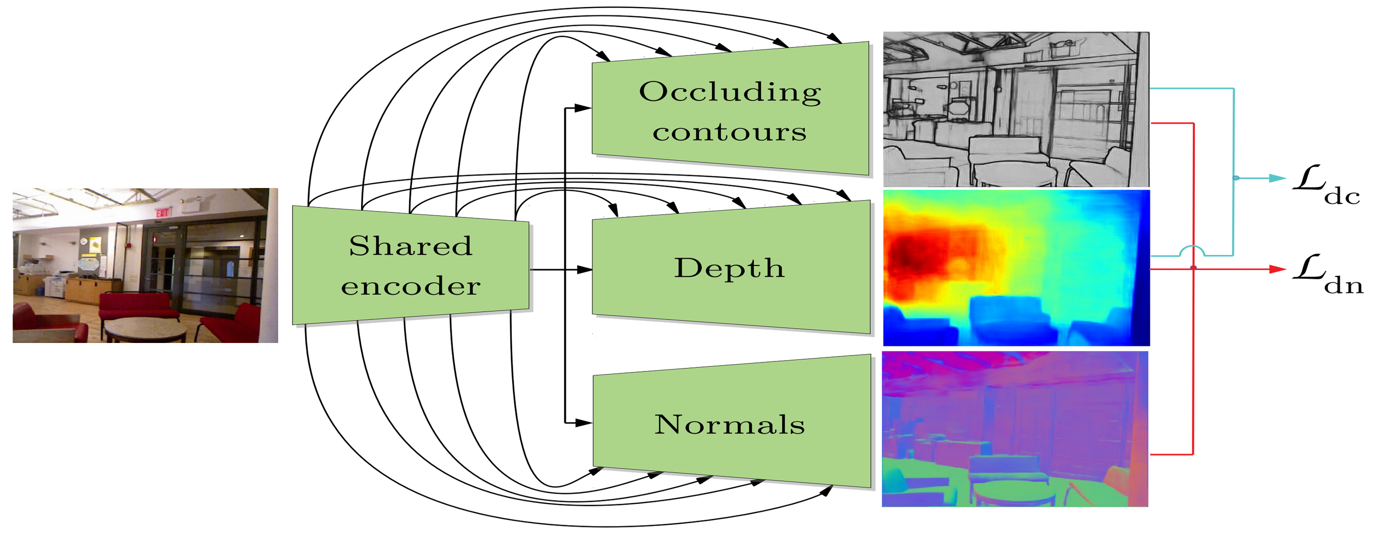
- A simple formulation of 3D local geometry constrains to improve depth estimation quality
- Our manually annotated dataset to benchmark reconstructed occluding contours accuracy of monocular depth estimation methods:

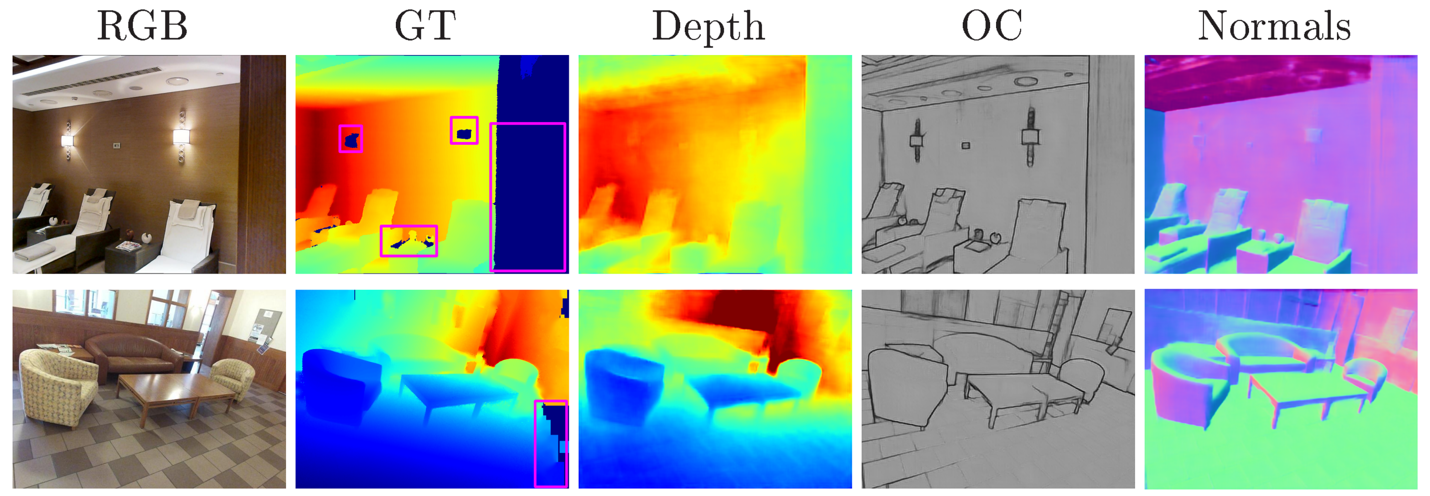
NYUv2-OC

Method

Implementation

- Multi-task learning of surface Depth $\boldsymbol{D}$, surface Normals $\boldsymbol{N}$ and Occluding Contours $\boldsymbol{C}$ (via Object Instance Contours)
- Each modality is supervised on synthetic data
- Only $\boldsymbol{D}$ is finetuned on real data while $\boldsymbol{N}$ and $\boldsymbol{C}$ serve as guidance





Loss terms

Global loss term

$$\mathcal{L} = \lambda_{\mathrm{d}}\mathcal{L}_{\mathrm{d}}(\boldsymbol{D},\widehat{\boldsymbol{D}}) + \lambda_{\mathrm{c}}\mathcal{L}_{\mathrm{c}}(\boldsymbol{C},\widehat{\boldsymbol{C}}) + \lambda_{\mathrm{n}}\mathcal{L}_{\mathrm{n}}(\boldsymbol{N},\widehat{\boldsymbol{N}}) + \mathcal{L}_{\mathrm{dc}}(\widehat{\boldsymbol{D}},\widehat{\boldsymbol{C}}) + \mathcal{L}_{\mathrm{dn}}(\widehat{\boldsymbol{D}},\widehat{\boldsymbol{N}}) \quad (1)$$

Normals $\leftrightarrow$ Depth

$$\mathcal{L}_{\mathrm{dn}}(\widehat{\boldsymbol{D}},\widehat{\boldsymbol{N}}) = \frac{1}{N}\sum_i \left(1 - \frac{<\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{n}}_i>}{\|\hat{\boldsymbol{u}}_i\|\|\hat{\boldsymbol{n}}_i\|}\right) \quad (2)$$

$$\hat{\boldsymbol{u}}_i = (\partial_x \widehat{D}_i, \partial_y \widehat{D}_i), \quad \hat{\boldsymbol{n}}_i = (\hat{N}_x^i, \hat{N}_y^i)^T \quad (3)$$

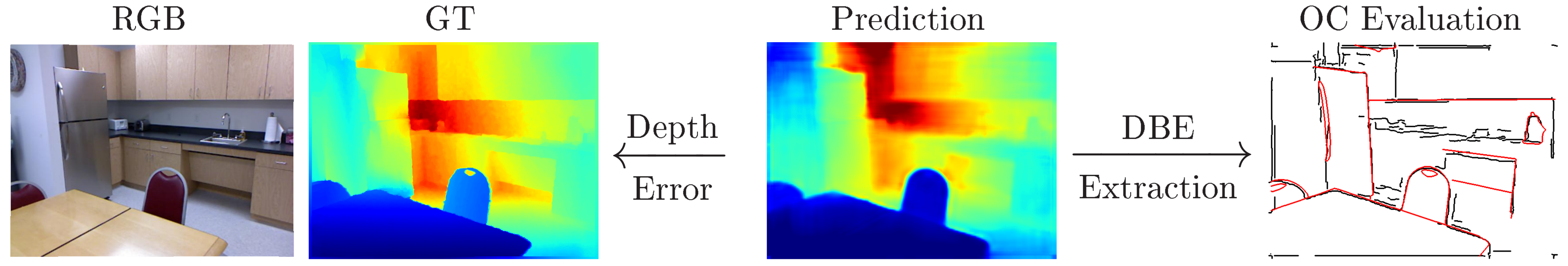
Depth $\leftrightarrow$ Occluding Contours

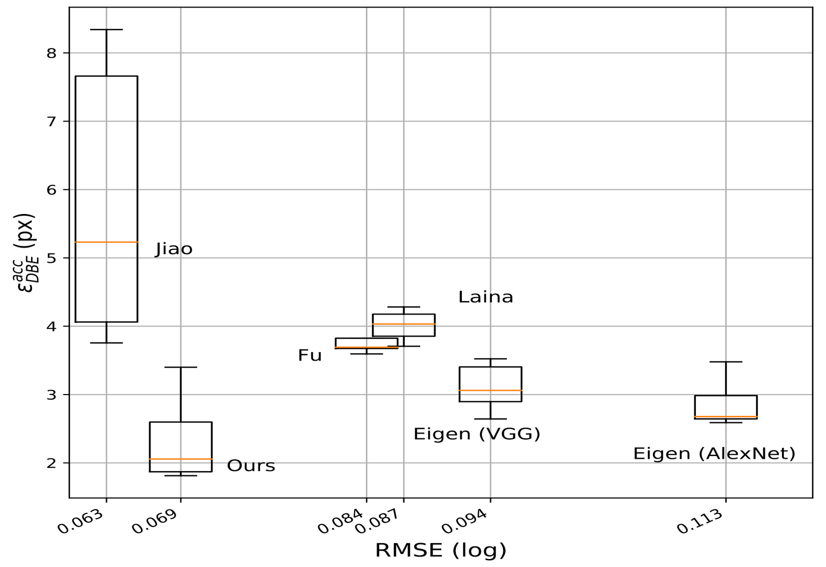
$$\mathcal{L}_{\mathrm{dc}}(\widehat{\boldsymbol{D}},\widehat{\boldsymbol{C}}) = -\frac{1}{N}\sum_i \log(\widehat{C}_i)\cdot\|\nabla(\widehat{D}_i)\|^2\|\Delta(\widehat{D}_i)\| \quad (\bigstar)$$
$$+ \frac{\mu}{N}\left(\|\widehat{\boldsymbol{C}}\|_1 - \sum_i \log(1-\widehat{C}_i)\cdot\exp(-\|\Delta(\widehat{D}_i)\|)\right). \quad (\blacklozenge) \quad (4)$$

The first term ($\bigstar$) enforces consensus between OC predictions and DBE (strong depth gradients). The second term ($\blacklozenge$) balances the first term by enforcing global sparsity of OC, especially in smooth surface depth regions.

Evaluation

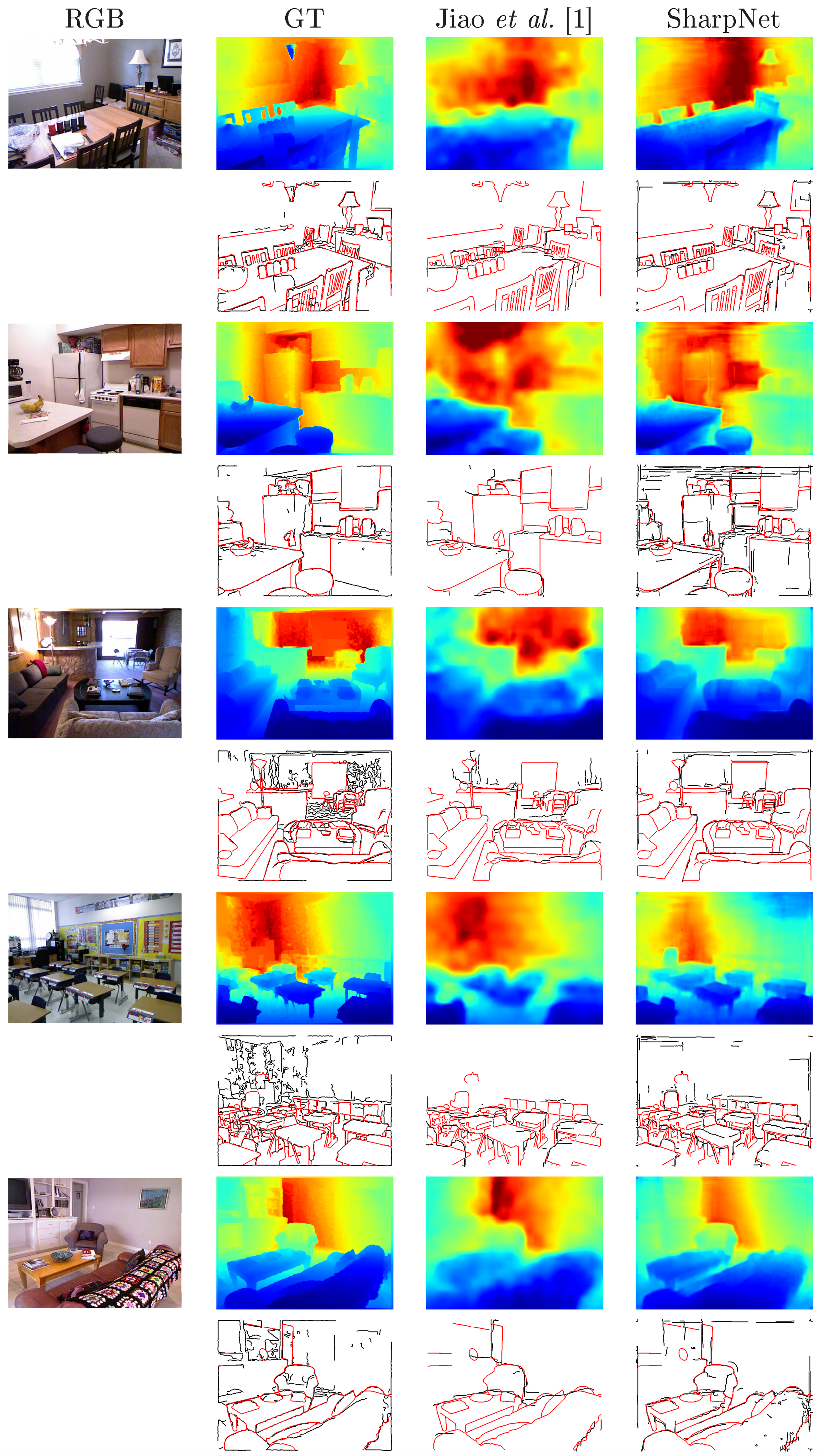
- Global depth reconstruction is evaluated on NYUv2-Depth
- OC accuracy is evaluted on NYUv2-OC using edge extraction and chamfer distance [2]





Method	Overall depth metric evaluation							Occluding contour metric evaluation			
	Accuracy ↑ ($\delta_i = 1.25^i$)			Error ↓				ϵ_{DBE}^{acc} ↓(px) $\{\sigma^-, \sigma^+\}$			
	δ_1	δ_2	δ_3	rel	$\log_{10}$	RMSE (lin)	RMSE (log)	$\{0.1, 0.2\}$	$\{0.01, 0.1\}$	$\{0.005, 0.06\}$	$\{0.03, 0.05\}$
Eigen *et al.* [3] (VGG)	0.766	0.949	0.988	0.195	0.068	0.660	0.217	2.895	3.065	<u>3.199</u>	<u>3.203</u>
Eigen *et al.* [3] (AlexNet)	0.690*	0.911*	0.977*	0.250*	0.082*	0.755*	0.259*	<u>2.840</u>	<u>3.029</u>	3.202	3.242
Laina *et al.* [4]	0.818	0.955	0.988	0.170	0.059	0.602	0.200	3.901	4.033	4.116	4.133
Fu *et al.* [5]	0.850	0.957	0.985	0.150	0.052	0.578	0.194	3.714	3.754	4.040	4.062
Jiao *et al.* [1]	**0.909**	**0.981**	**0.995**	**0.133**	**0.042**	**0.401**	**0.146**	6.389*	4.073*	4.179*	4.190*
Ours	<u>0.888</u>	<u>0.979</u>	**0.995**	<u>0.139</u>	<u>0.047</u>	<u>0.495</u>	<u>0.157</u>	**2.272**	**2.629**	**3.066**	**3.152**

Qualitative results



References

[1] J. Jiao, Y. Cao, Y. Song and R. W.H.Lau, *Look Deeper into Depth: Monocular Depth Estimation with Semantic Booster and Attention-Driven Loss*, ECCV, 2018

[2] T. Koch, L. Liebel, F. Fraundorfer and M. Körner, *Evaluation of CNN-Based Single-Image Depth Estimation Methods*, ECCV Workshops, 2018

[3] D. Eigen and R. Fergus, *Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture*, ICCV, 2015

[4] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari and N. Navab, *Deeper Depth Prediction with Fully Convolutional Residual Networks*, 3DV, 2016

[5] H. Fu, M. Gong, C. Wang, K. Batmanghelich, and D. Tao, *Deep Ordinal Regression Network for Monocular Depth Estimation*, CVPR, 2018