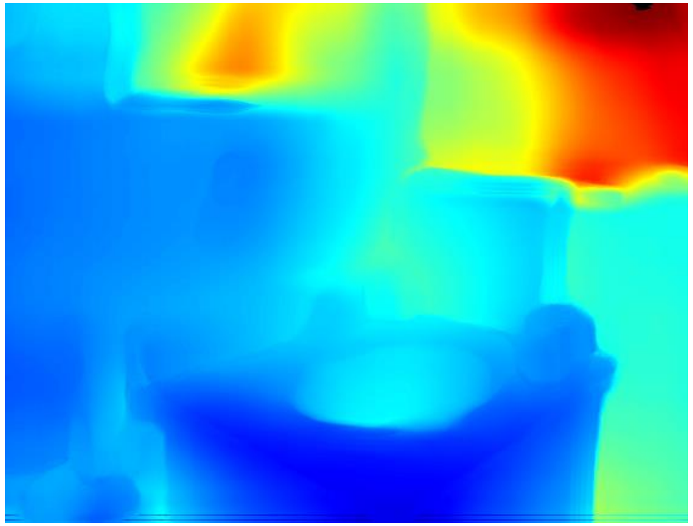
GeoNet: Geometric Neural Network for Joint Depth and Surface Normal Estimation

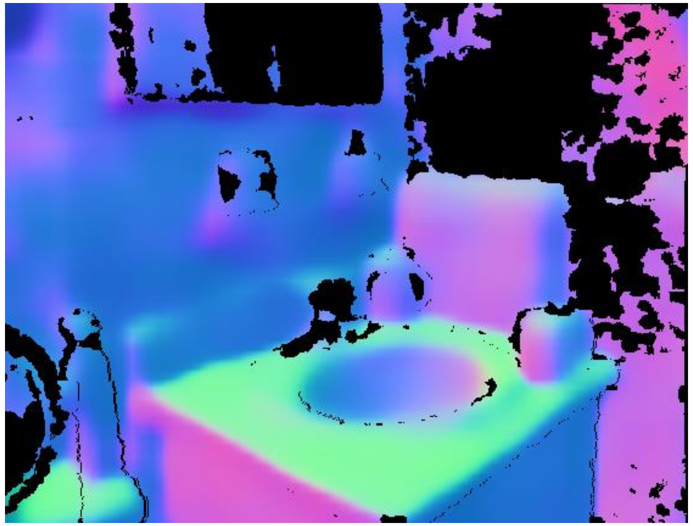
Joint work with Renjie Liao, Zhengzhe Liu, Raquel Urtasun, Jiaya Jia

Problem

Image



Depth



Normal

Prior work

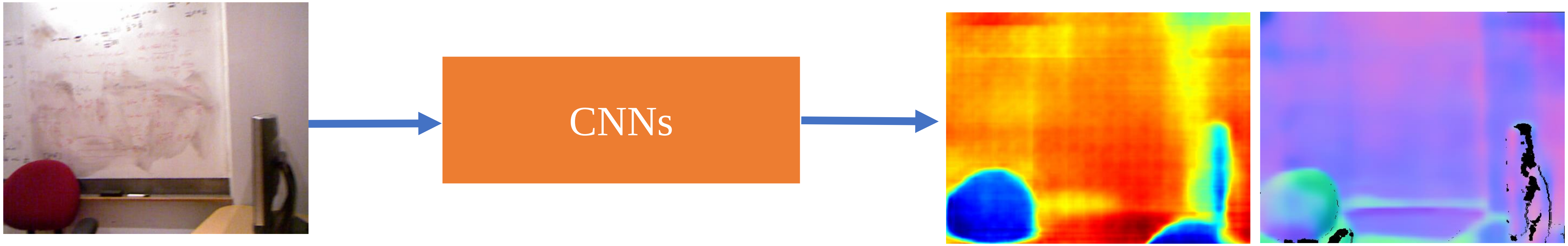
- CNN based depth/normal prediction from single image

  Discrete-continuous depth estimation from a single image [Liu et al. 2014].
  Towards unified depth and semantic prediction from a single image [Wang et al. 2015].
  Designing deep networks for surface normal estimation [Wang et al. 2015].
  Predicting depth, surface normals and semantic labels… [Eigen et al. 2015].
  Monocular depth estimation using neural regression forest [Roy et al. 2016].
  Marr revisited: 2d-3d alignment via surface normal prediction [Bansal et al. 2016].
  Depth from a single image by harmonizing overcomplete local network predictions [Chakrabarti et al. 2016].
  Coupled depth learning [Baig et al. 2016].
  Deeper depth prediction with fully convolutional residual networks [Laina et al. 2016].
  Surge: Surface regularized geometry estimation from a single image [Wang et al. 2016]
  Multi-scale continuous crfs as sequential deep networks for monocular depth estimation [Xu et al. 2017].
  DORN: Deep Ordinal Regression Network for Monocular Depth Estimation [Fu et al. 2018].
  Joint task-recursive learning for semantic segmentation and depth estimation [Zhang et al. 2018].

Prior work

2D point wise regression problem.

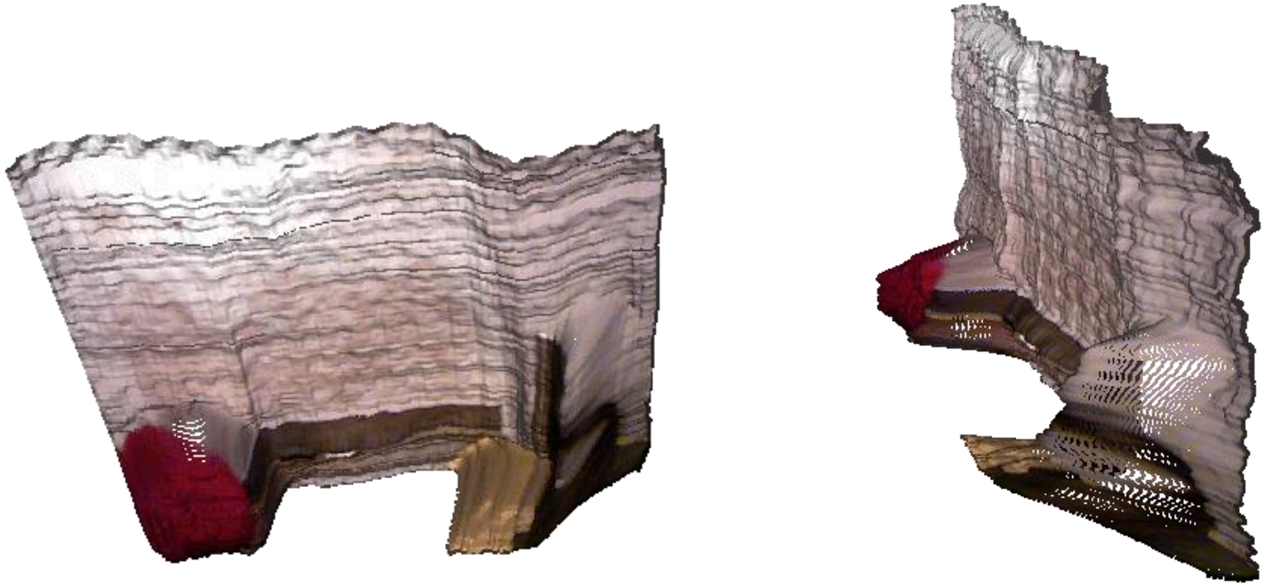
CNN are good at improving metric result.



CNN based pipeline

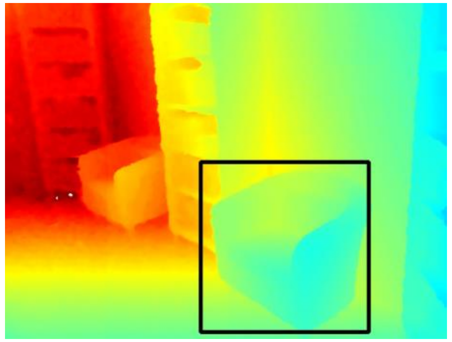
Prior work

Input image



Predicted result visulized in 3D
[Laina et al. 2017]

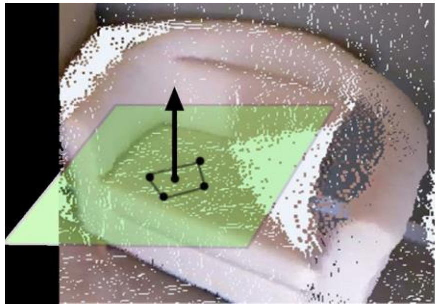
Motivation



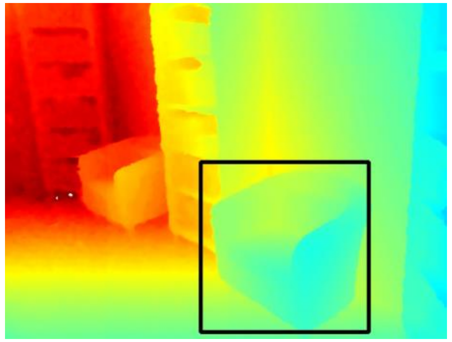
Depth

3D Point Cloud

Motivation

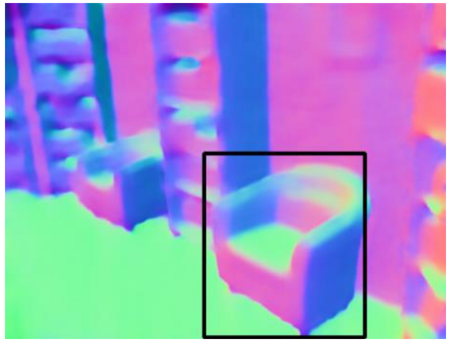


Depth to Normal



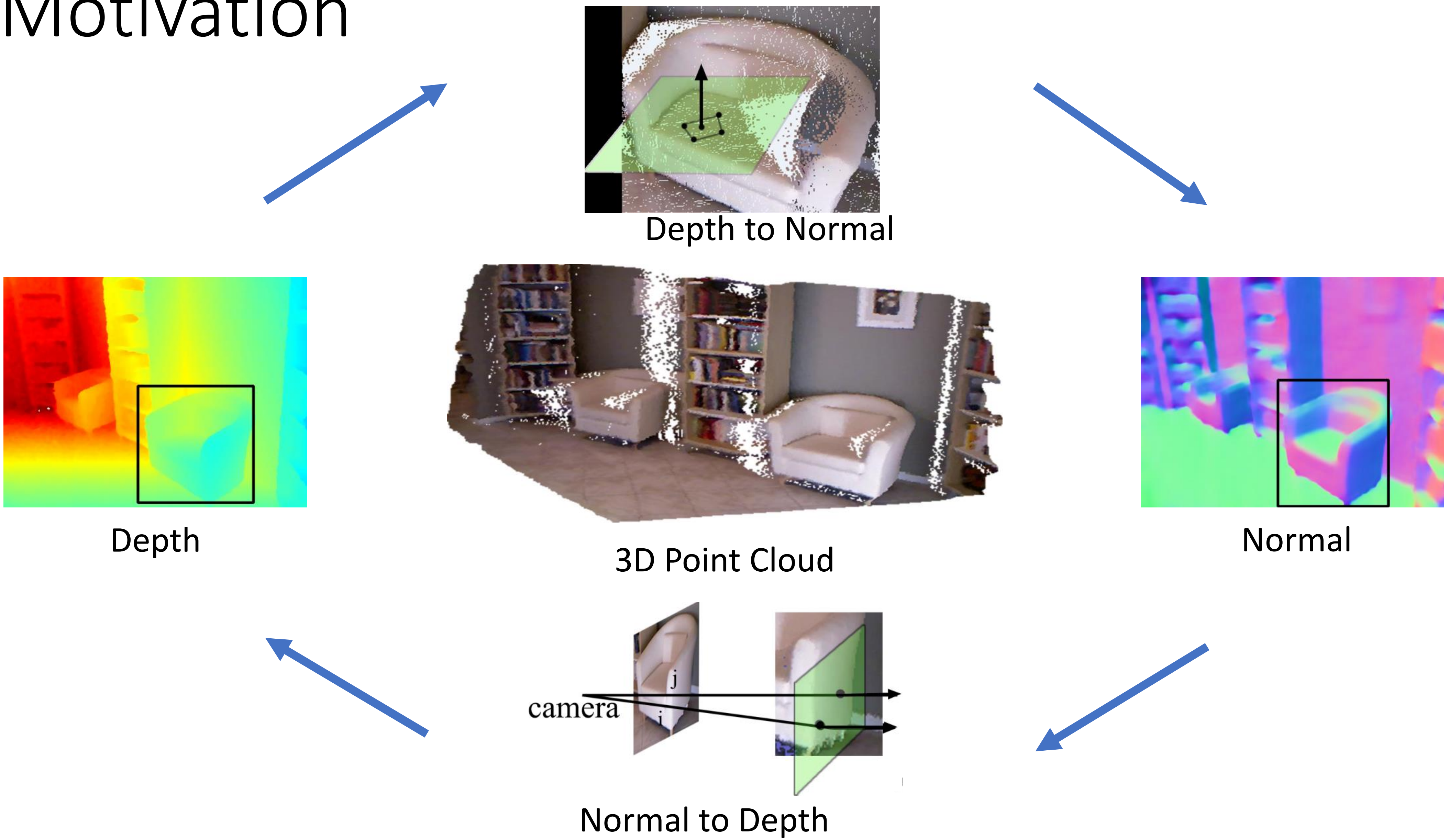
Depth

3D Point Cloud



Normal

Motivation


Depth to Normal
Depth
3D Point Cloud
Normal
camera
j
i
Normal to Depth

Challenge

Depth and Normal are geometrically related.

Incorporate geometrical relationship into deep neural network is non-trivial.

Challenge

- Verification experiments

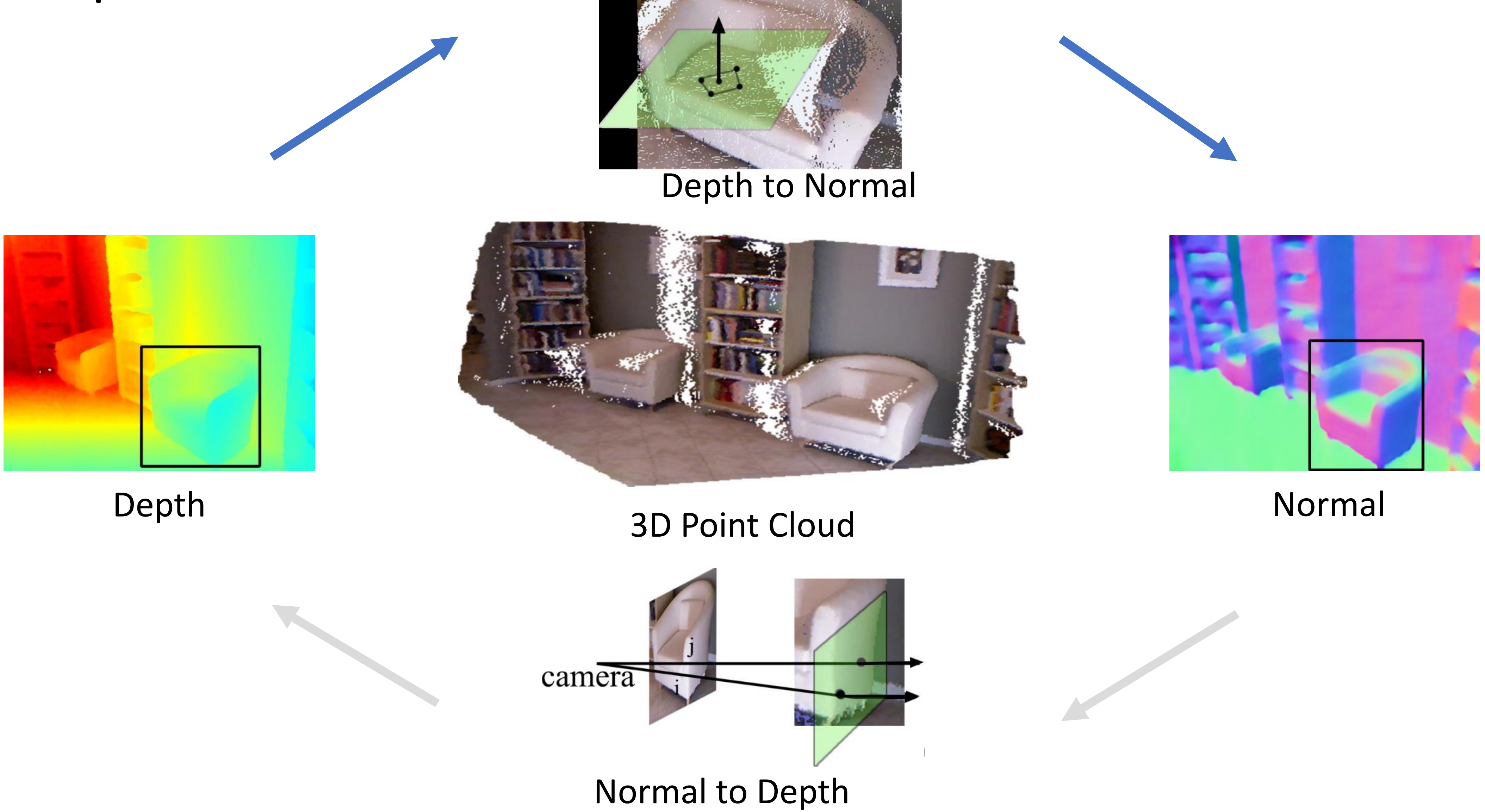
Challenge

- Verification experiments

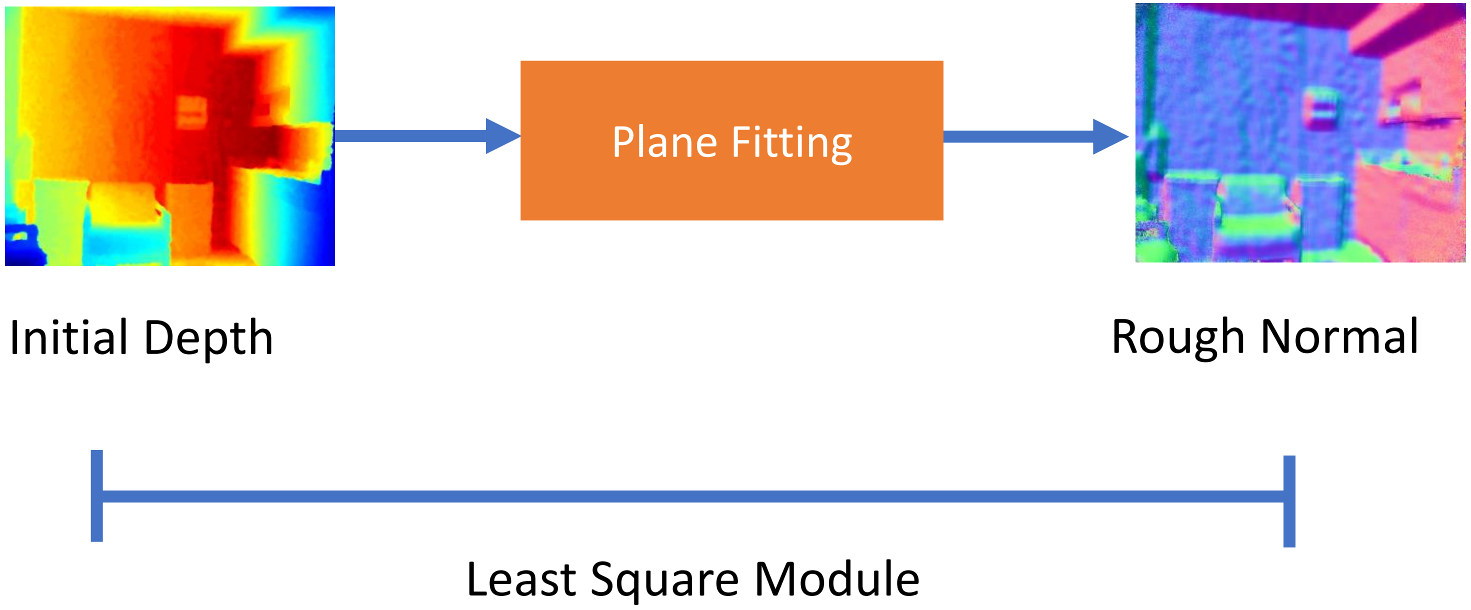
- Results:

  Best result (median error 37.6°, mean error 39.5°, rmse 44°)

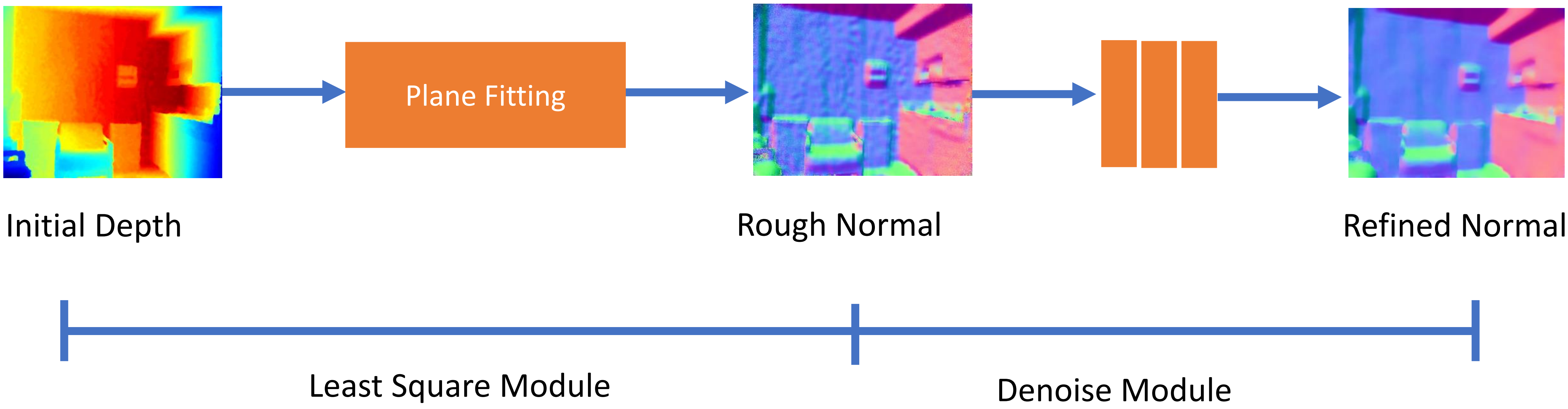
Depth to Normal



Depth to Normal Modeling



Depth to Normal Modeling

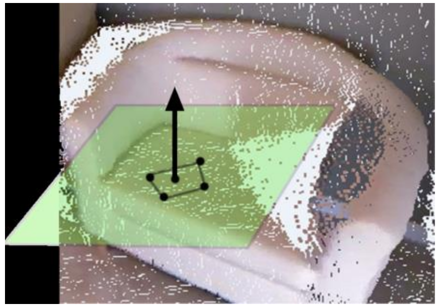


CNN and Geometric Condition

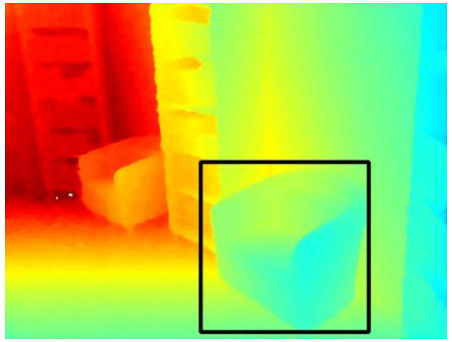
	Metric	CNN	Least Square	Full-model
Error	mean	39.5	11.5	8.2
	median	37.6	6.4	3.0
	RMSE	44.0	18.8	15.5
Accuracy	$\delta < 11.25^\circ$	6.1	70.0	80.0
	$\delta < 22.50^\circ$	21.4	86.7	90.3
	$\delta < 30.00^\circ$	35.5	91.3	93.5

CNN cannot converge

Normal to Depth

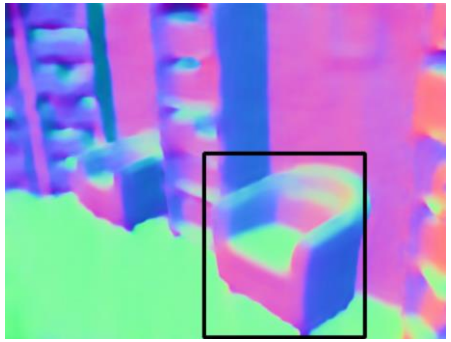


Depth to Normal

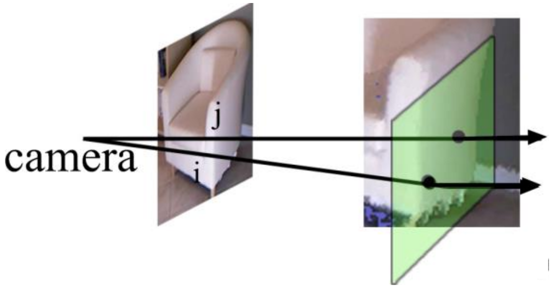


Depth

3D Point Cloud

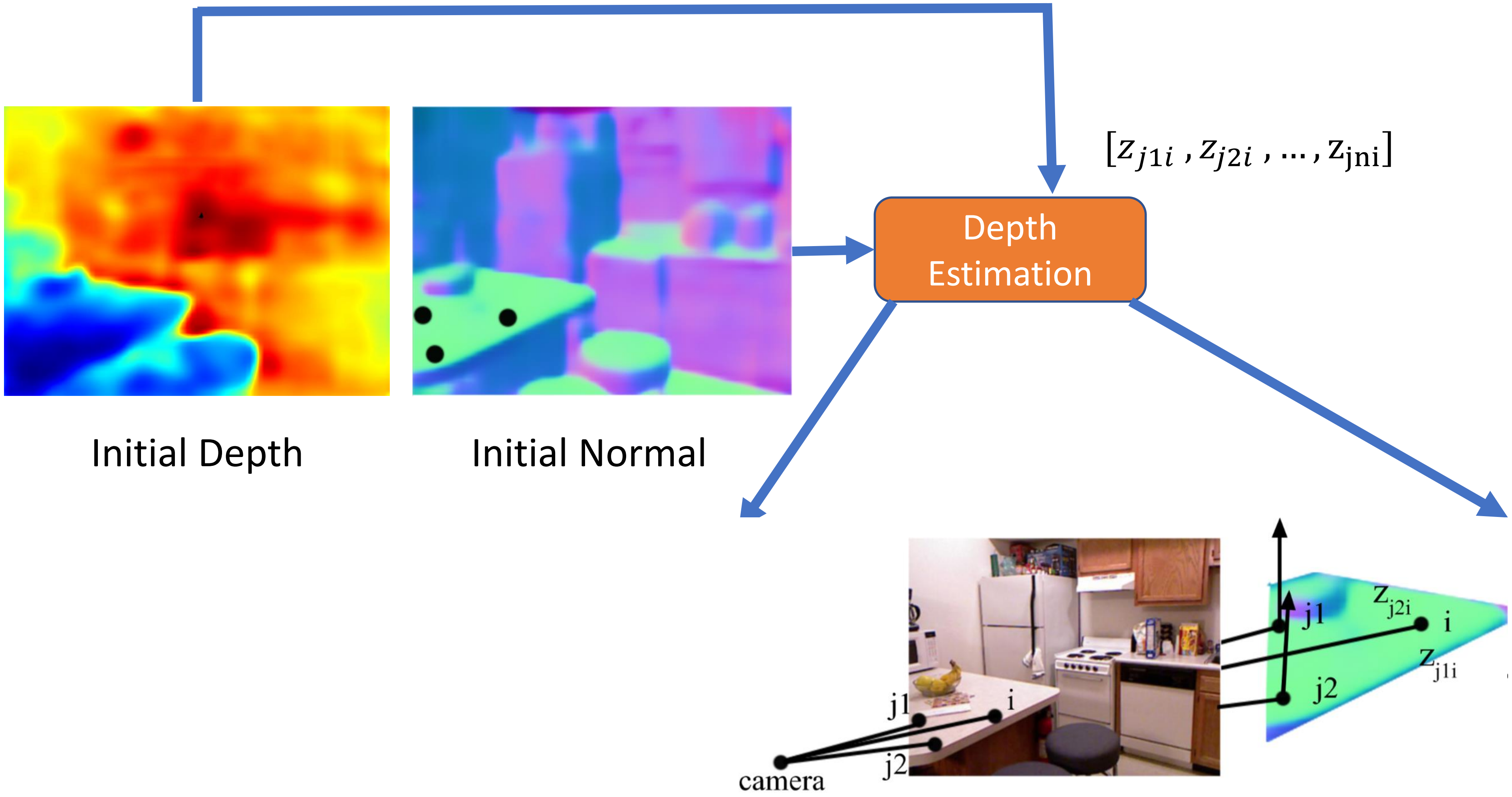


Normal

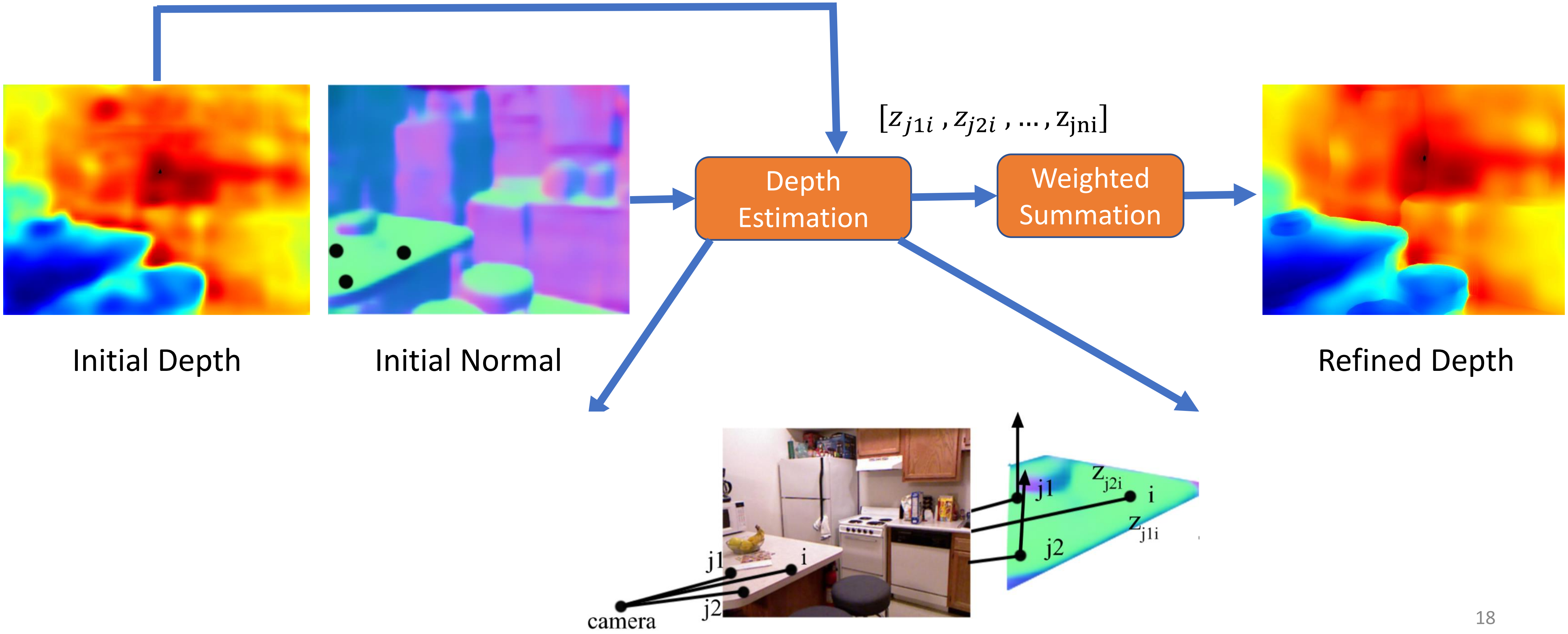


Normal to Depth

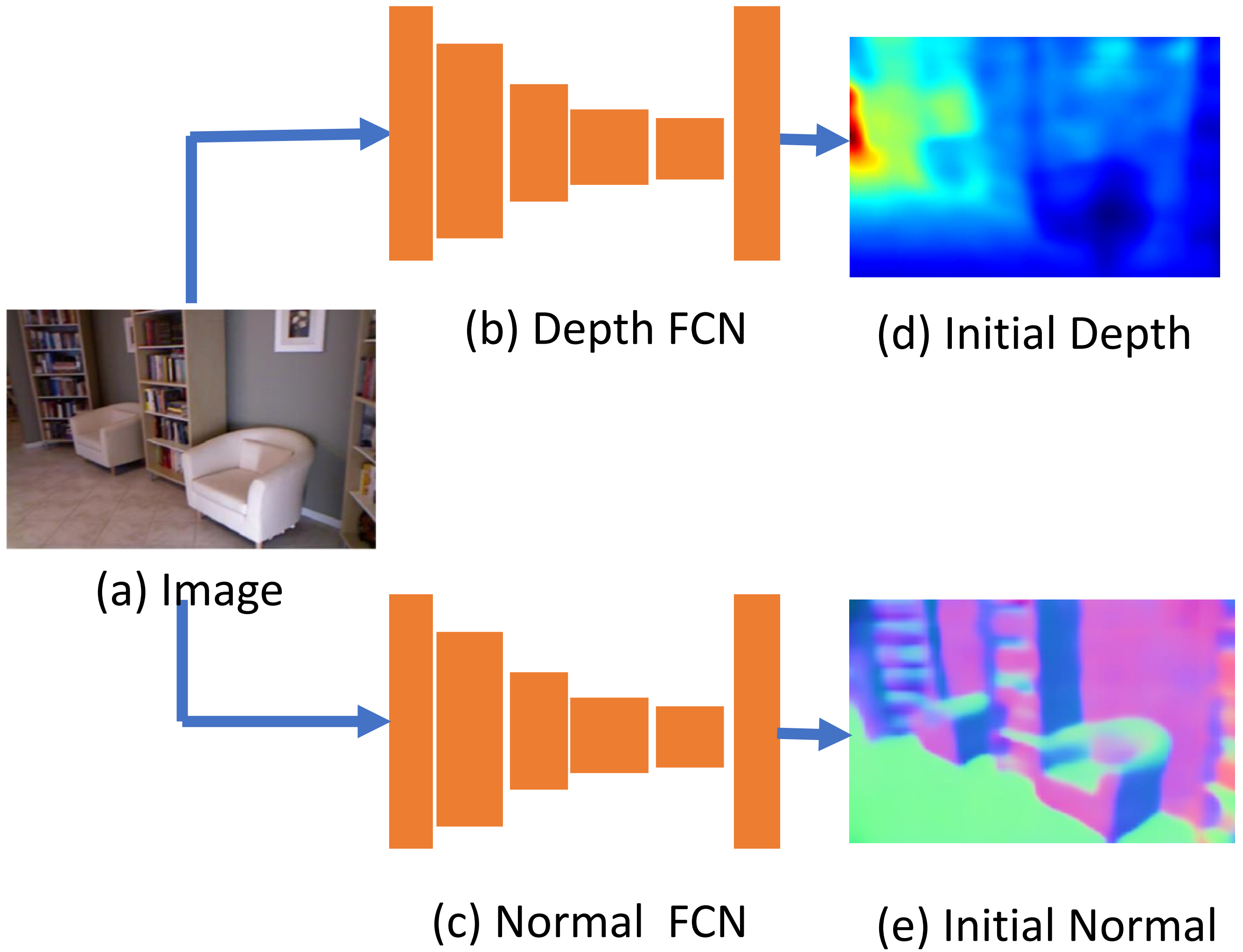
Normal to Depth Modeling



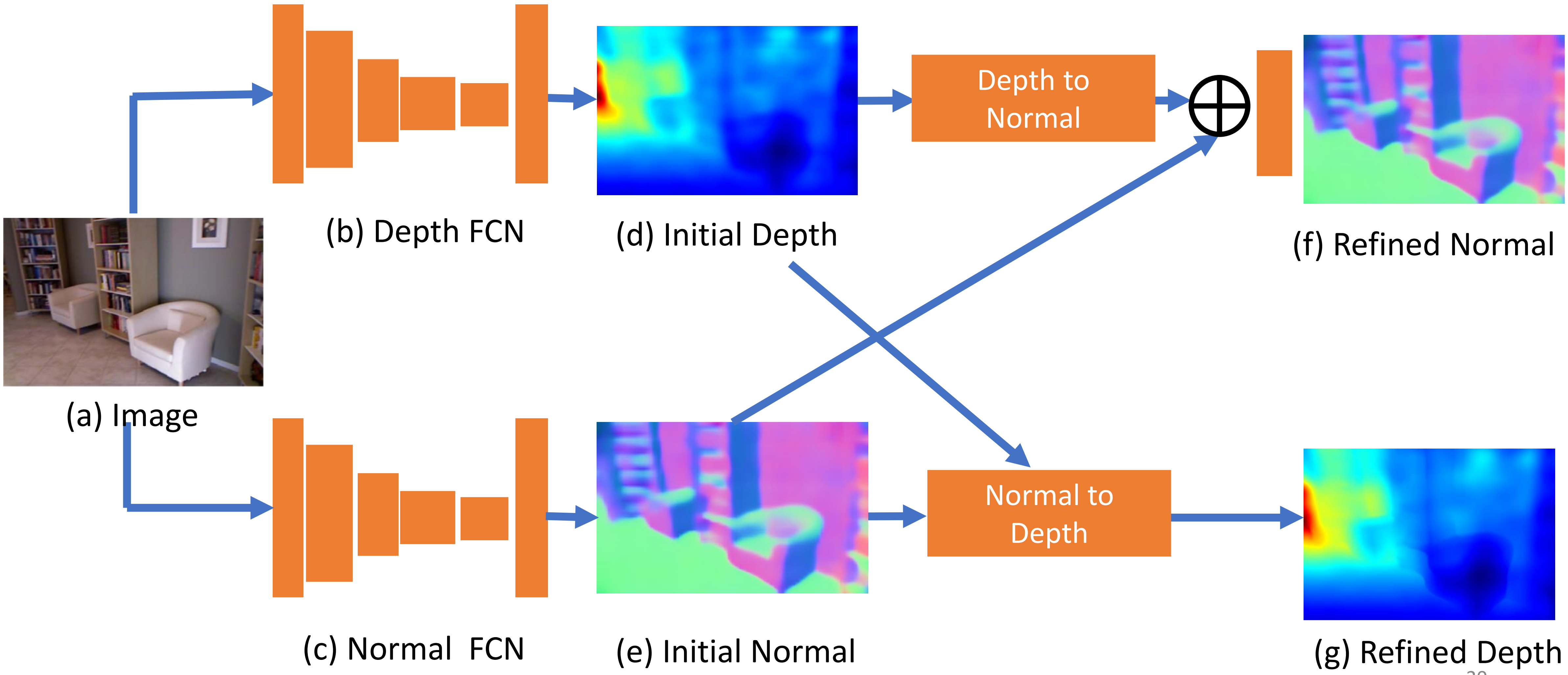
Normal to Depth Modeling



Overall Model



Overall Model



Results

Evaluation: Normal

	Metric	Baseline	Our
Error	mean	19.4	19.0
	median	12.5	11.8
	rmse	27.0	26.9
Accuracy	$\delta < 11.25^{\circ}$	46.0	48.4
	$\delta < 22.50^{\circ}$	70.3	71.5
	$\delta < 30.00^{\circ}$	78.9	79.5

Predict dormal i: n_i

Ground truth normal i : n_i^*

Angle δ_i: $\delta_i = \arccos(n_i^T n_i^*)$

Mean error: $\frac{1}{N}\sum_i \delta_i$

Median error: Median value in set $\{\delta_i, \delta_2, \dots, \delta_N\}$

rmse: $\sqrt{\frac{1}{N}(\sum_i \delta_i^2)}$

Accuracy $\delta < \theta$: $\frac{\sum_i sign\,(\delta_i - \theta)}{N}$

Evaluation: Depth

	Metric	Baseline	Our
Error	rmse	0.626	0.626
	log10	0.068	0.065
	rel	0.155	0.149
Accuracy	$\delta < 1.25$	76.8	78.6
	$\delta < 1.25^2$	95.1	95.6
	$\delta < 1.25^3$	98.9	99.0

Predict depth i: d_i

Ground truth depth i : d_i^*

rMSE: $\sqrt{\frac{1}{N}(\sum_i |d_i - d_i^*|^2)}$

Log10: $\frac{1}{N}\sum_i |log(d_i) - log(d_i)|$

Relative: $\frac{1}{N}\sum_i \frac{|d_i - d_i^*|}{d_i^*}$

Accuracy with threshold t: percentage of pixels subject to $\max\left(\frac{d_i}{d_i^*}, \frac{d_i^*}{d_i}\right) < t$.

Consistency Evaluation

	Metric	Baseline	Our	Improve
Error	mean	42.2	34.9	17.3%
	median	39.8	31.4	21.1%
	rmse	48.9	41.4	15.3%
Accuracy	$\delta < 11.25^{\circ}$	9.8	15.3	56.1%
	$\delta < 22.50^{\circ}$	25.2	35.0	38.9%
	$\delta < 30.00^{\circ}$	35.9	47.7	32.9%

Transform depth to norm and compare the transformed result with the predictions from the corresponding model.

2D Illustration

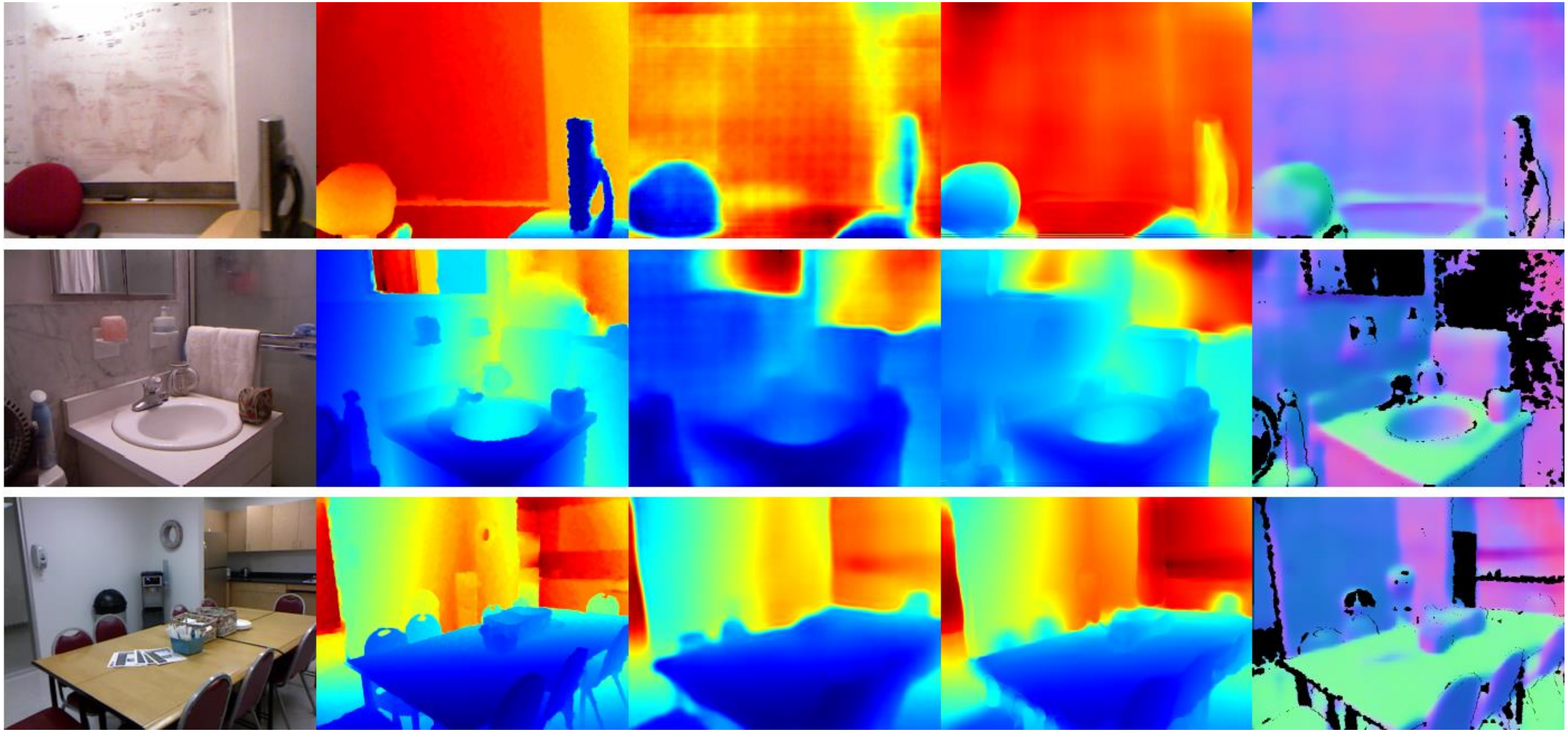
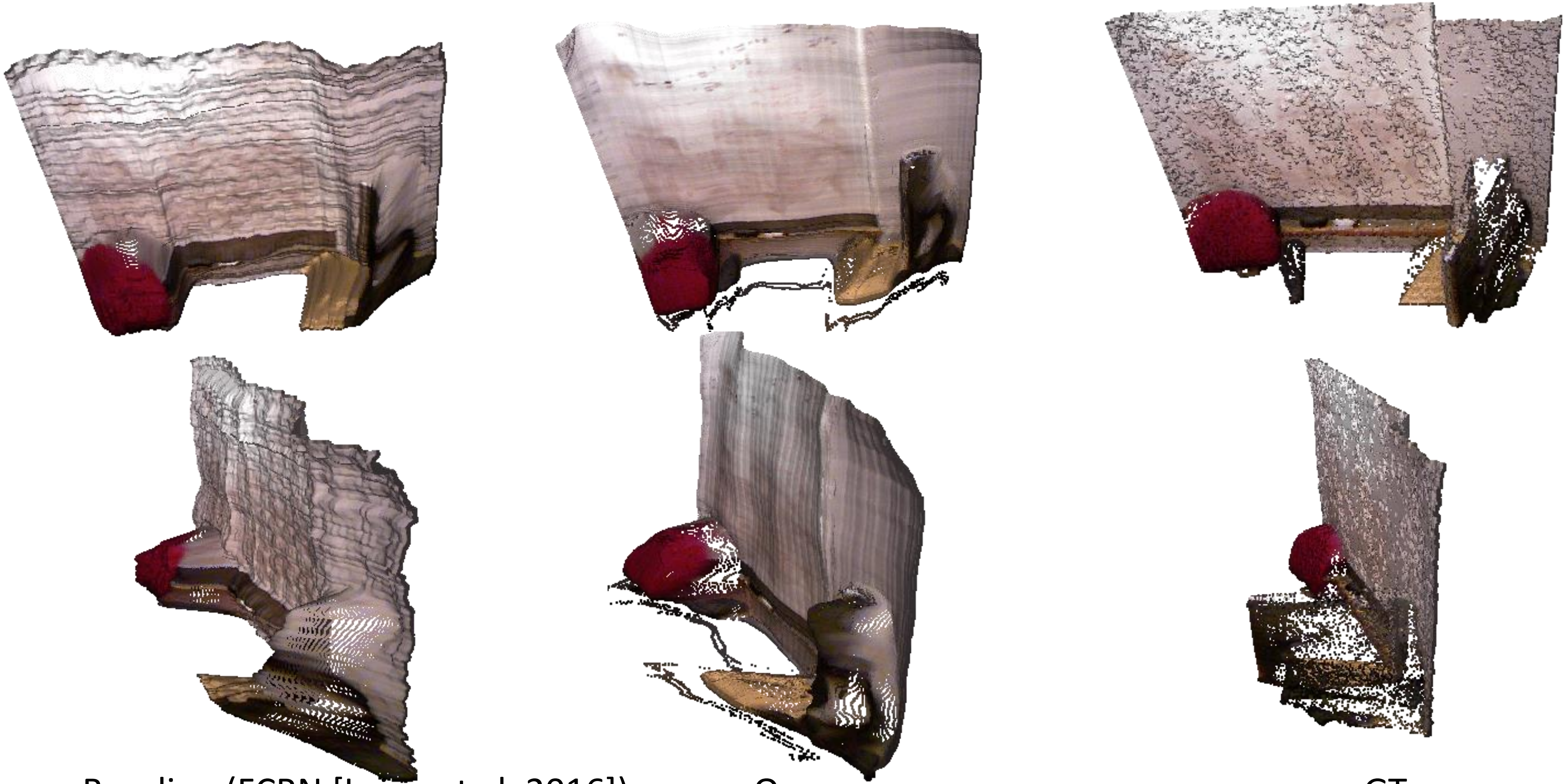


Image GT-Depth Baseline (FCRN) Our-depth Our-Normal

3D Illustration



Baseline (FCRN [Laina et al. 2016]) Our GT