LEGO: Learning Edge with Geometry all at Once by Watching Videos

Zhenheng Yang[1], Peng Wang[2], Yang Wang[2], Wei Xu[3], Ram Nevatia[1]

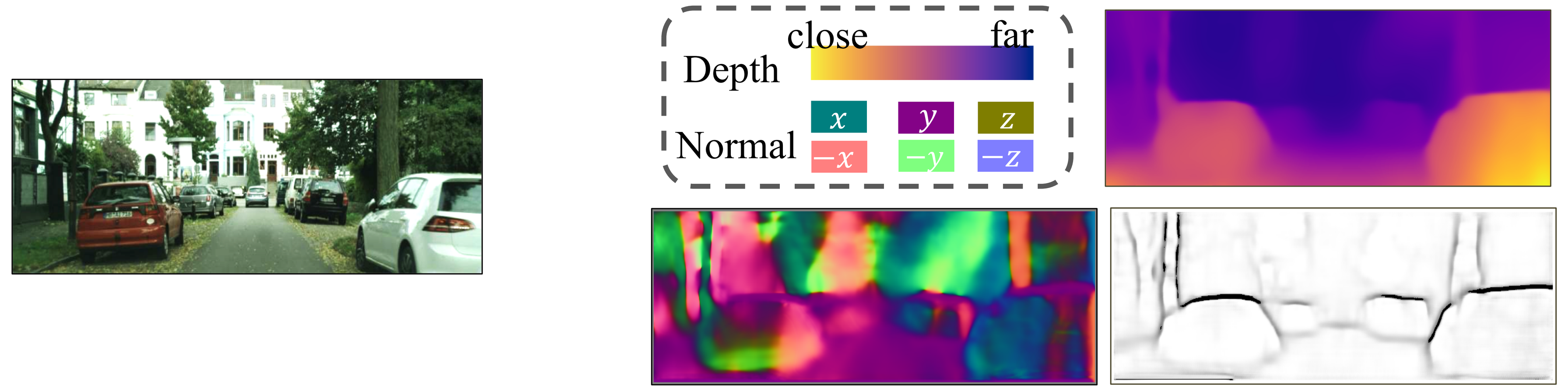
[1]University of Southern California, [2]Baidu Research,

[3]National Engineering Laboratory for Deep Learning Technology and Applications

SALT LAKE CITY • JUNE 18-22

Task

➢ Infer depth, surface normal, geometrical from single RGB image



➢ Model trained on unlabeled monocular videos

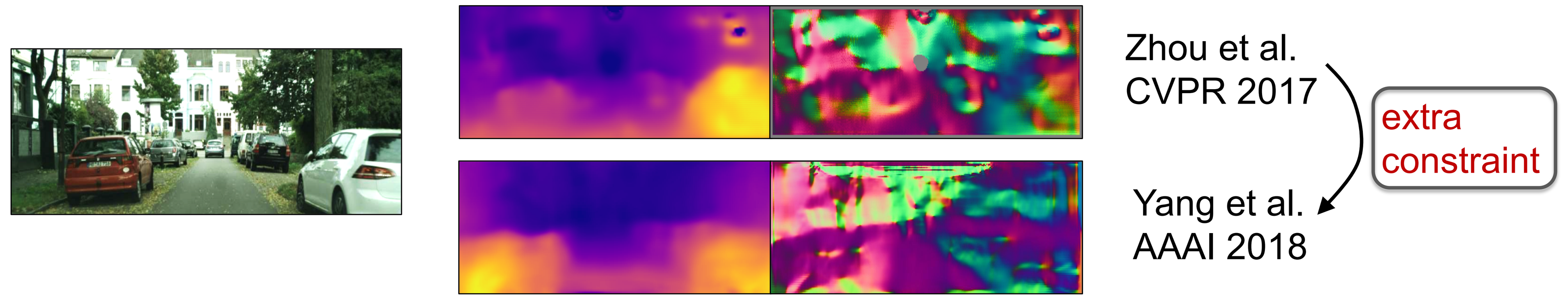
Challenge

➢ Multiple geometric cues are jointly modeled without any ground truth

Motivation

➢ Unsupervised approach is important to avoid expensive annotations

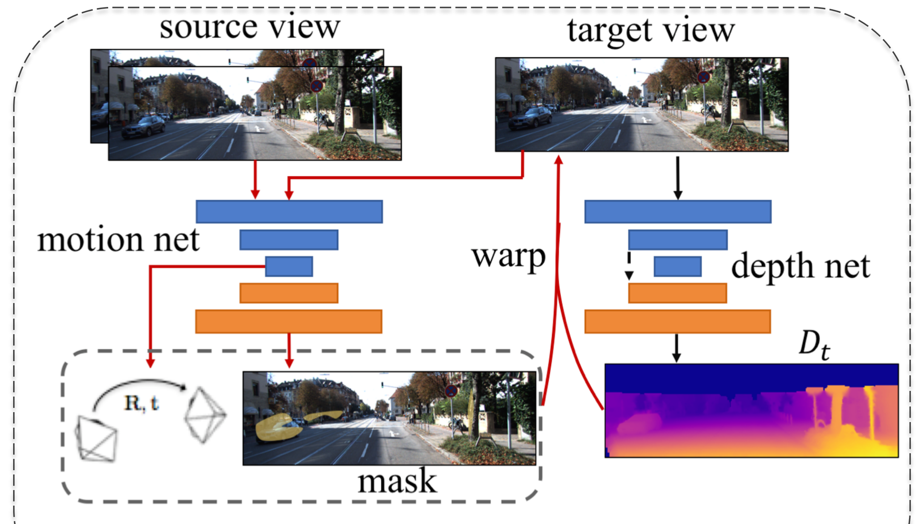
➢ Joint learning of edges with 3D geometry



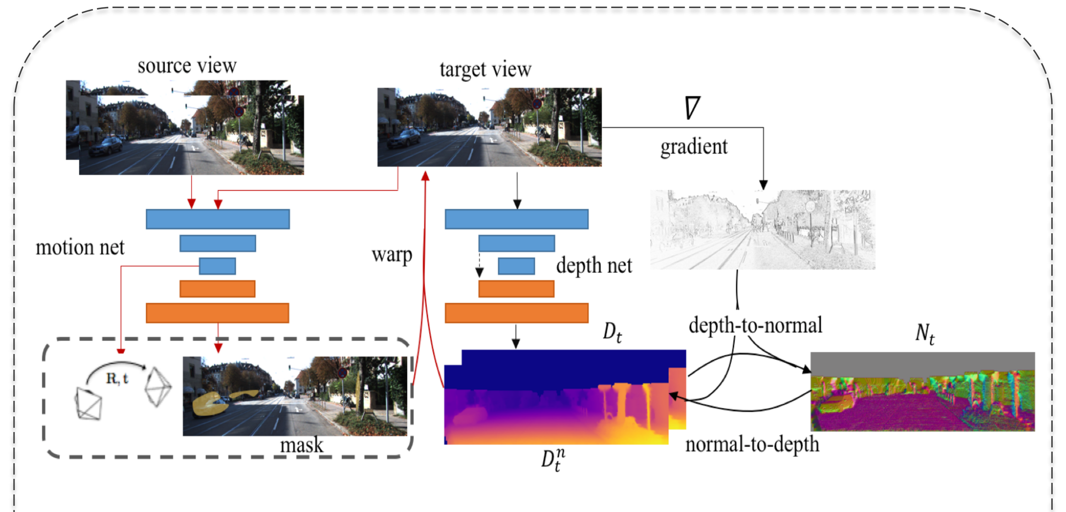
Preliminary

➢ Zhou et al. proposed to use view synthesis as supervision

➢ Yang et al. proposed to leverage extra depth-normal consistency



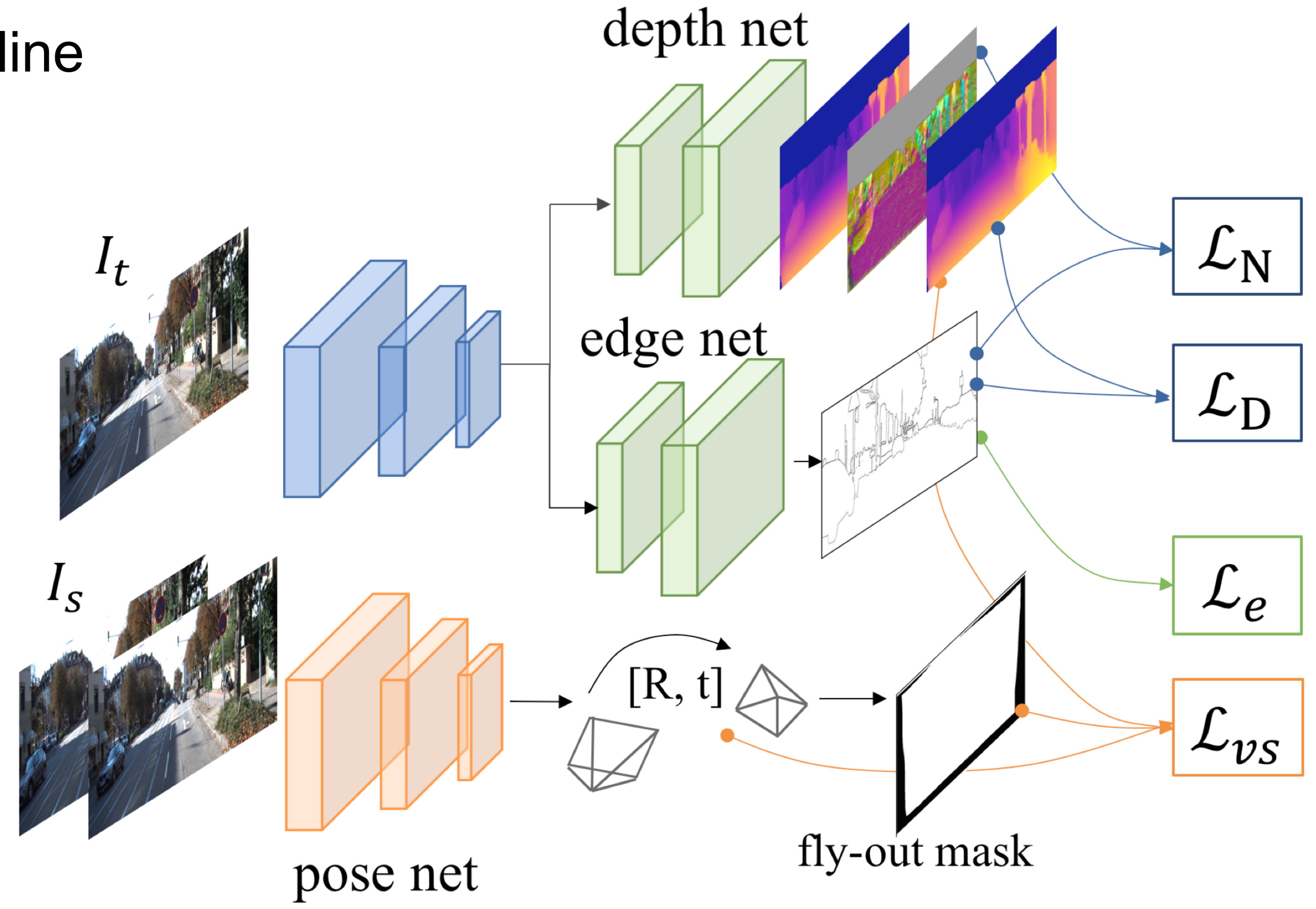
Zhou et al. CVPR 17'



Yang et al. AAAI 18'

Approach

➢ Pipeline



➢ Sub-networks: depth net, edge net, pose net

➢ Regularization terms:

$\mathcal{L}_N$

$$\mathcal{L}_N = \sum_{p_i} \sum_{x,y} \|N_t(p_i) - N_t(p_i(x,y))\|_1 \kappa(p_i, p_i(x,y))$$

Normal directions on same surface are consistent (1st order constraint)

$\mathcal{L}_D$

$$\mathcal{L}_{D_x} = \sum_{p_i} \sum_{x} \|\mathbf{g}(p_i, x)\|_1 \kappa(p_i, p_i(x)) \kappa(p_i, p_i(-x))$$

$$\mathbf{g}(p_i, x) = \frac{D_t(p_i(x)) - D_t(p_i)}{\phi(p_i(x)) - \phi(p_i)} - \frac{D_t(p_i) - D_t(p_i(-x))}{\phi(p_i) - \phi(p_i(-x))}$$

Depth changes smoothly in 3D on same surface (2nd order constraint)

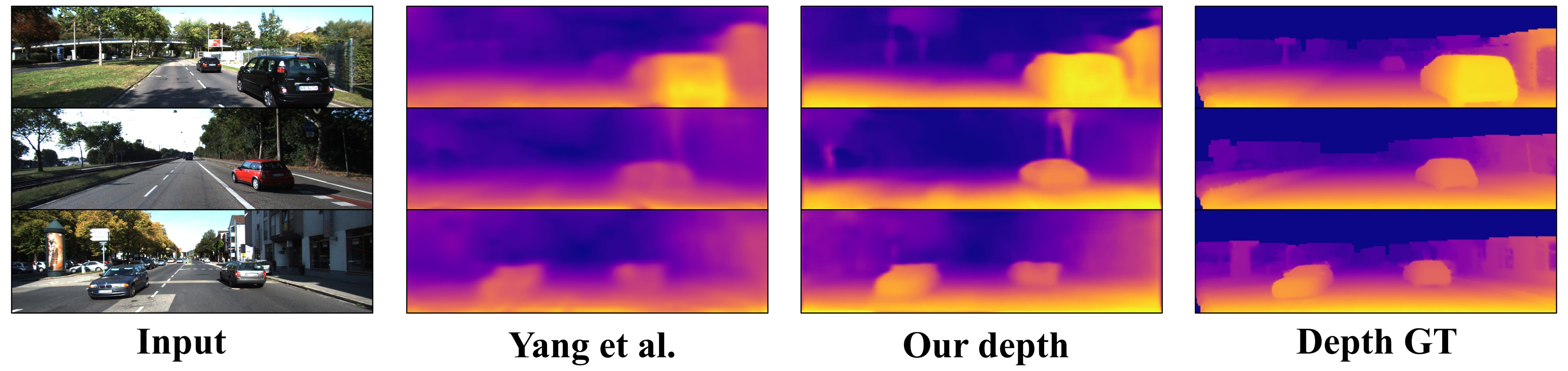
$\mathcal{L}_e$ $\quad \mathcal{L}_e(E) = \sum_{p_i} \|E_t(p_i)\|^2$ L2 regularization of edge map

$\mathcal{L}_{vs}$ $\quad \mathcal{L}_{vs}(D, \mathcal{T}) = \sum_s \sum_{p_t \in I_t} |I_t(p_t) - \hat{I}_s(p_t)|$ View synthesis

Results

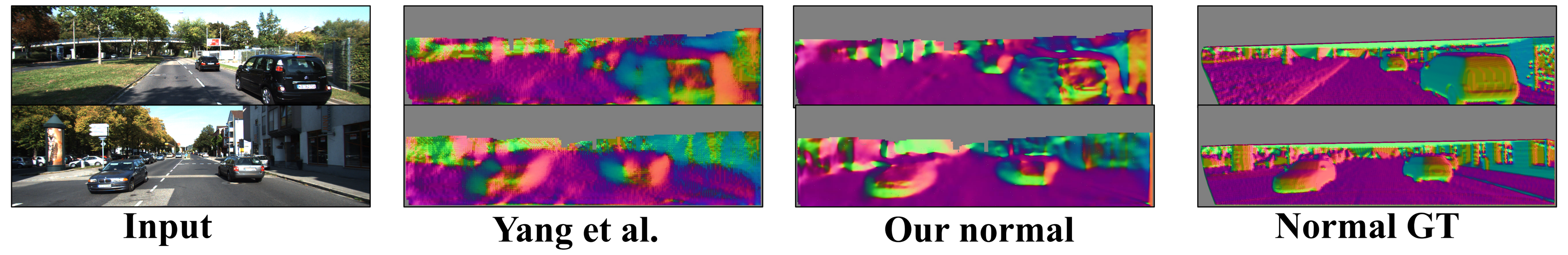
➢ Depth

Method	Test data	Supervision	Lower the better: Abs Rel	Sq Rel	RMSE	RMSE log
Godard et al.[19]	KITTI split	Pose	0.124	1.388	6.125	0.217
Zhou et al.[64]			0.216	2.255	7.422	0.299
Yang et al.[60]			0.165	1.360	6.641	0.267
LEGO (no fly-out)			0.157	1.303	6.223	0.241
LEGO			0.154	1.272	6.012	0.230
LEGO+CS (more data)			0.142	1.237	5.846	0.225

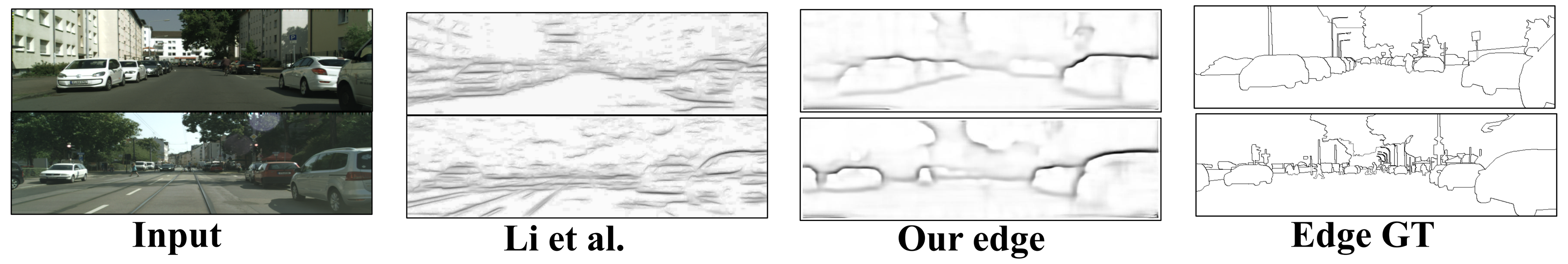


➢ Normal

Method	Mean	Median	11.25°	22.5°	30°
Godard et al.[19]	39.28	29.37	0.158	0.412	0.496
Yang et al.[60]	47.52	33.98	0.149	0.369	0.473
LEGO	**36.13**	**25.94**	**0.241**	**0.473**	**0.542**



➢ Edge



Methods	ODS	OIS	AP
Li et al.[36]	0.623	0.663	0.557
HED-D	0.457	0.503	0.467
LEGO	**0.710**	**0.731**	**0.729**

Code

Paper