GroundNet: Monocular Ground Plane Normal Estimation with Geometric Consistency

Yunze Man, Xinshuo Weng, Xi Li, Kris Kitani

Zhejiang University, Carnegie Mellon University

{yzman,xilizju}@zju.edu.cn, {xinshuow, kkitani}@cs.cmu.edu

Background

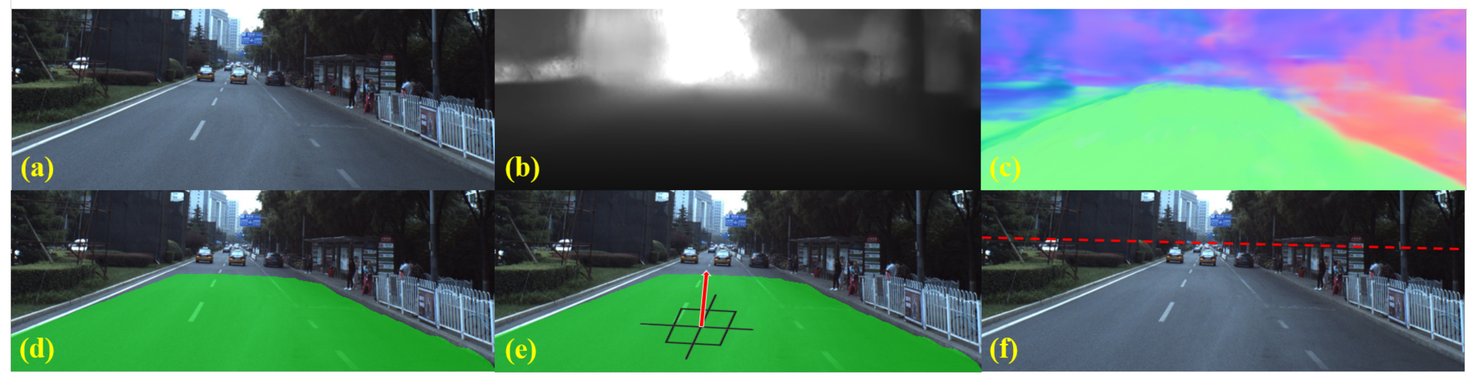
- **Problem Definition:** Given a single RGB image I, to estimate the 3D orientation of the ground plane, which is usually represented by an normal vector $\mathbf{n} = [n_x, n_y, n_z]$.
- **Geometric Model:** Pinhole Camera Model

$$P_i = \begin{bmatrix} \lambda u_i \\ \lambda y_i \\ \lambda \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = K_c Q_i$$

where K_c is intrinsic matrix, f_x and f_y are the focal length along the x and y directions respectively.

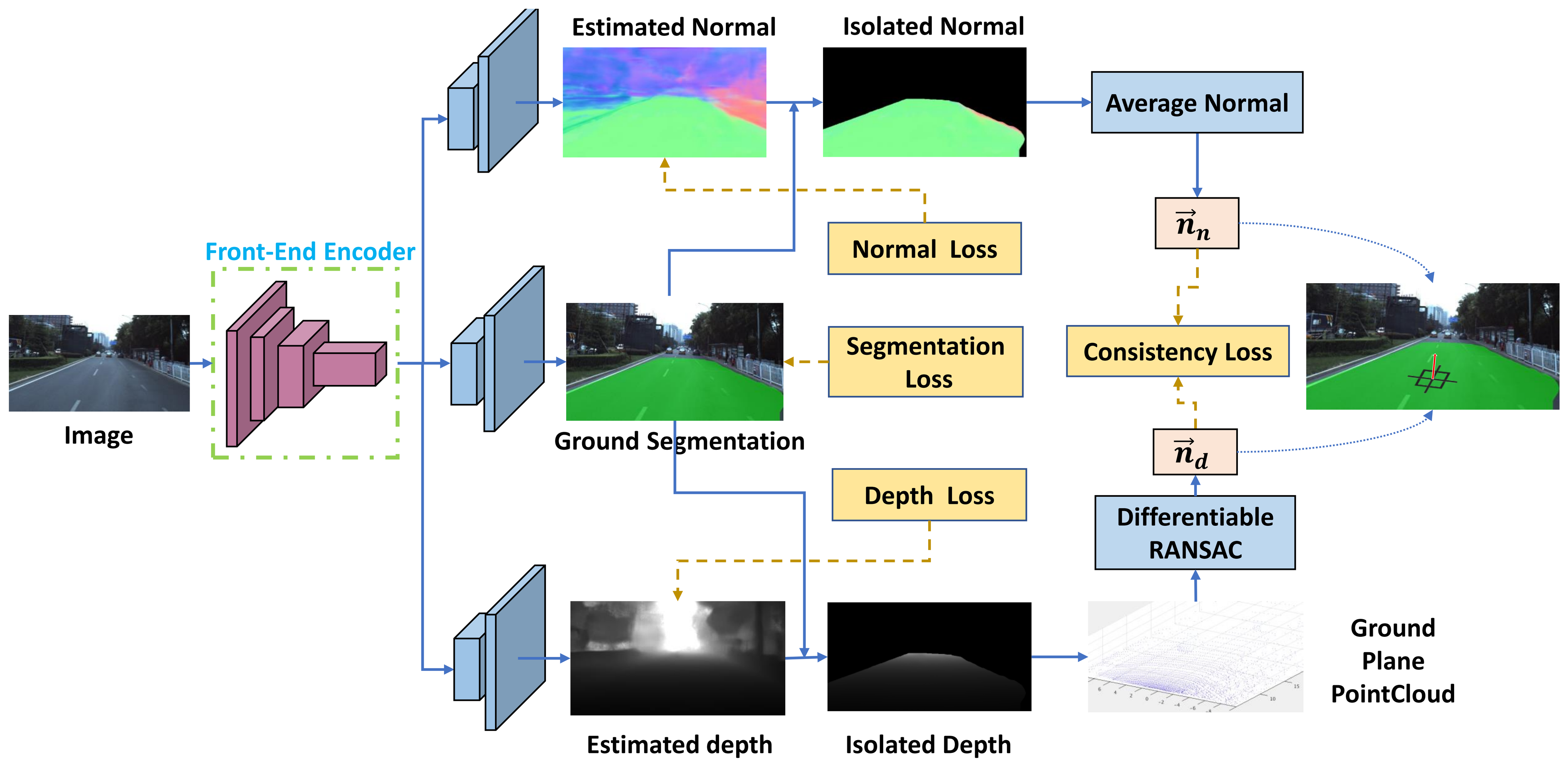
- **Datasets:** Real-world autonomous driving dataset
  - **KITTI:** A popular autonomous driving dataset
  - **ApolloScape:** New big dataset for scene parsing.

Motivation



- Estimating the 3D orientation of the ground plane is an important pre-processing step for ground robots, wearable camera systems and autonomous driving.
- Previous work either highly rely on the reliability of ***low-level feature*** extraction, or directly estimate the ***horizon line*** without considering the underlying 3D geometric constraints.
- ***Geometric correlation*** between depth and normal provides essential supervision, allowing better learning of 3D information through multi-modality interaction and refinement
- Therefore, we propose GroundNet for end-to-end ground plane normal estimation from monocular image, which leverage deep feature extraction with additional 3D geometric supervision.

GroundNet Framework



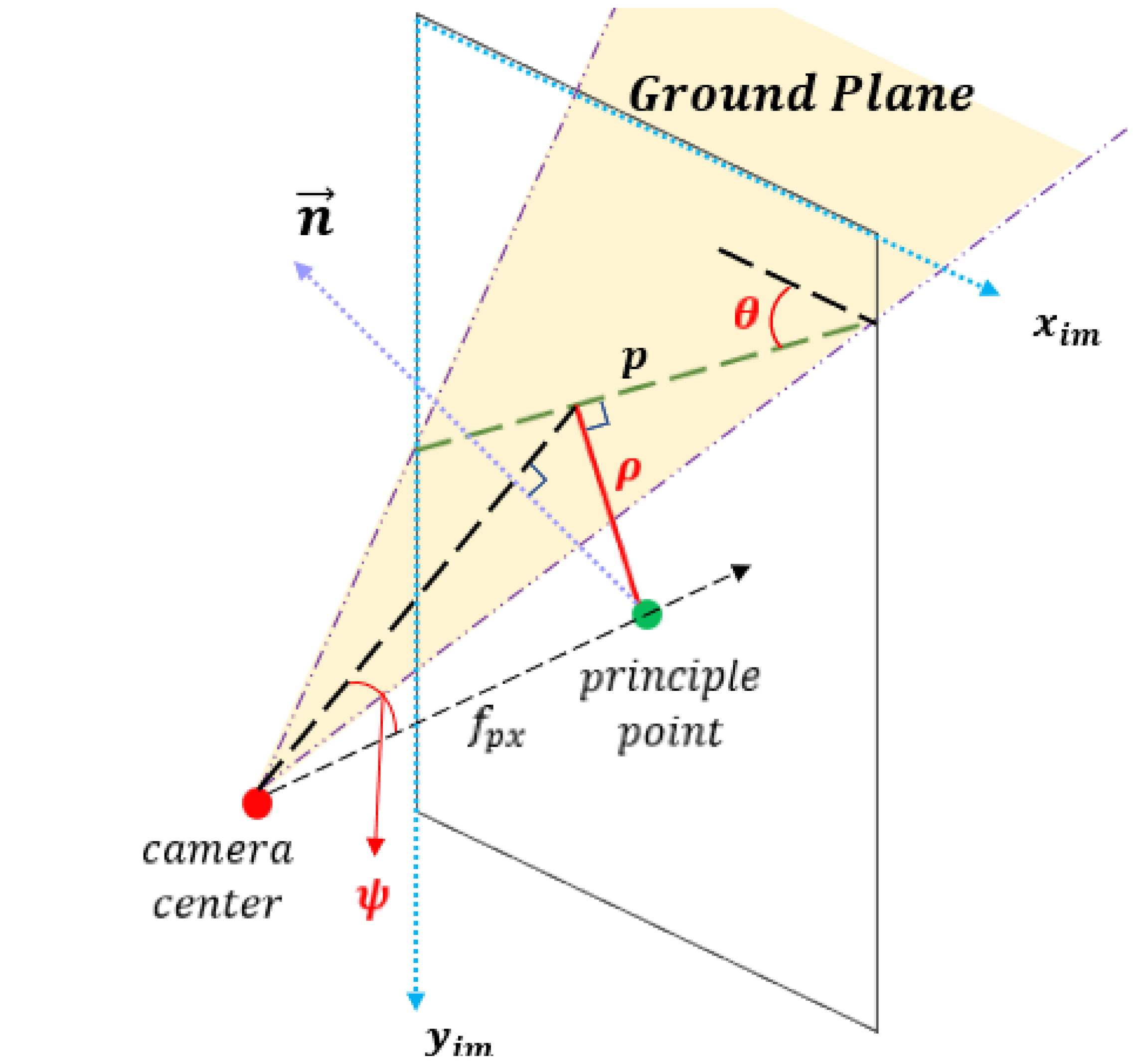
- ***Front-End***: A fully convolutional encoder first outputs three streams: (1) surface normal stream (2) depth stream (3) ground seg stream.
- ***Depth Stream***: A pixel-wise depth estimation, followed by back-projection to 3D. Then a plane fitting module will give a grounf normal estimation. We try Least Square (LS) and Differentiable-RANSAC (D-RANSAC) module.
- ***Normal Stream***: A dense normal estimation followed by ground average
- ***Ground Seg Stream***: An auxiliary Task to get rid of various outdoor noise
- ***Geometric Consistency***: Based on local linear orthogonality assumption, formalize as a quadratic minimization.

Ablation Study

	KITTI			ApolloScape		
	0.05	0.10	0.15	0.05	0.10	0.15
normal	6.73	7.99	9.35	5.78	7.47	8.24
depth (LS)	3.01	4.47	6.52	2.97	4.39	5.66
depth (DSAC)	2.92	4.29	6.41	2.74	4.25	5.38
joint (LS)	2.74	4.06	5.93	2.68	4.02	5.23
joint (DSAC)	**2.65**	**3.84**	**5.41**	**2.49**	**3.72**	**4.87**

- ***Joint Training*** validates the effectiveness geometric consistency
- ***DSAC*** module outperforms the vanilla least square model

Evaluation Metrics



- Horizon line is ***equivalent*** to ground normal.

Horizon Line Estimation

- Red: GT. Green: GroundNet. Gray: VP.
- Cyan: Perceptual. Yellow: DeepHorizon.

Normal Estimation

- Red: GT. Blue: GroundNet. Yellow: SkipNet.

P4D-13, ACM Multimedia (MM), 2019