Geometric constraints

Created	@March 20, 2022 1:10 PM
Property	

link:

https://openaccess.thecvf.com/content_ICCV_2019/papers/Yin_Enforcing_Geometric_Constraints_of_Virtual_Normal_for_Depth_Prediction_ICCV_2019_paper.pdf

Backgrounds

- high-order 3D geometric constraints is important
  - design a loss term
    - enforce geometric constraints(virtual normal) by randomly sampled 3 points in 3D space
    - accuracy improved
  - byproduct
    - can recover good structures of point cloud & surface normal, directly from depth
    - no training process for new-sub models
- TEST
  - NYU Depth-V2
  - KITTI

Conclusion

> SOTA(2019)

- VNL, a long-range geometric constraints
  - not only extract from neighbor, but the whole
  - robustness
  - accuracy
  - simple

Intro

Monocular Depth Prediction

- object ← → single monocular camera

Problem

- Endeavous are made by using **local** geometric constraints
    - In common cameras, they can be fluctuated
    - Extracted from neighbor, so not the whole picture of the scene

Work in the paper

> virtual normal

- reconstruct 3D **point cloud** from **estimated** depth map
    - RGB pixel (2D) → 3D
- from point cloud generated, randomly select 3 points with large distance → *virtual plane*, defined by *virtual normal(VN)*
- How much does VN differ from real one can be used to loss metrics

Related Work

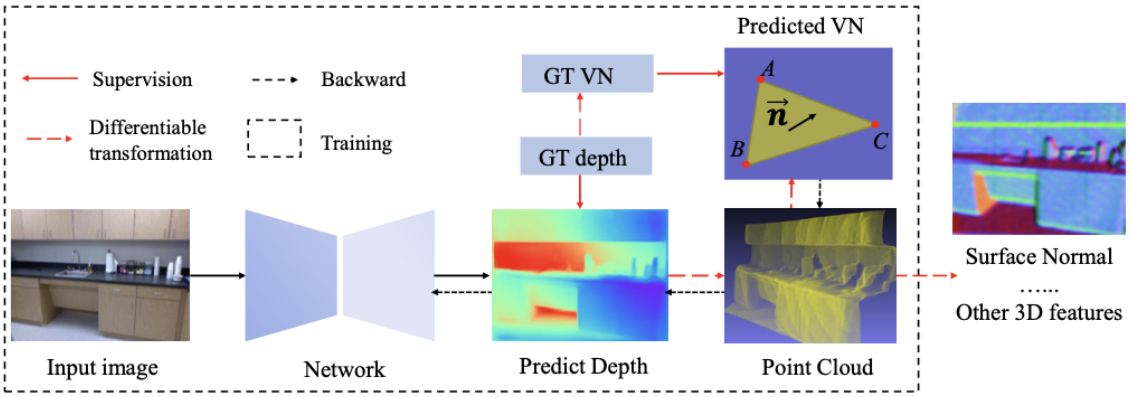
Monocular depth prediction

- most early work focus on Pixel-Wise (or Point-Wise), using very deep neural network
    - active work
    - passive work
- CRF(continuous conditional random field) focus extract information from neighbor, which is Pair-Wise
- High-order method

Surface Normal

Work

main architecture



main pipeline of the structure

- Pixel-Wise supervision like others, producing D_{pred}
- With VN, do Geometric Constraints supervision from D_{pred}, producing reconstructed point cloud P_{pred}
- Under GT VN, GT depth, producing D_{gt} and P_{gt}

High-order Geometric Constraints

Surface Normal

- calculate method: (local method)

  https://www.researchgate.net/publication/221070884_Comparison_of_surface_normal_estimation_methodsfor_range_sensing_applications

  - accuracy is influenced by sampling methods, which is less robust
  - common camera is less accurate

Virtual Normal

- 2D pixel $p_i(u_i, v_i)$ → 3D point $P_i(x_i, y_i, z_i)$
- projection formula:

$$z_i = d_i, x_i = \frac{d_i(u_i - u_0)}{f_x}, y_i = \frac{d_i(v_i - v_0)}{f_y}$$

  d_i is the depth, f_x, f_y is focal depth along x, y axis

- Sample method: (randomly select N groups, each has 3 points)
  - the 3 are non-linear, which can be represented to a plane
    - α, β is hyperparameter ($\alpha = 120^\circ, \beta = 30^\circ$)

$$\{\alpha \geq \angle(\overrightarrow{P_A P_B}, \overrightarrow{P_A P_C} \geq \beta, \alpha \geq \angle(\overrightarrow{P_B P_C}, \overrightarrow{P_B P_A} \geq \beta,)\}$$

  - long-range restriction
    - $\theta = 0.6m$ is hyperparameter

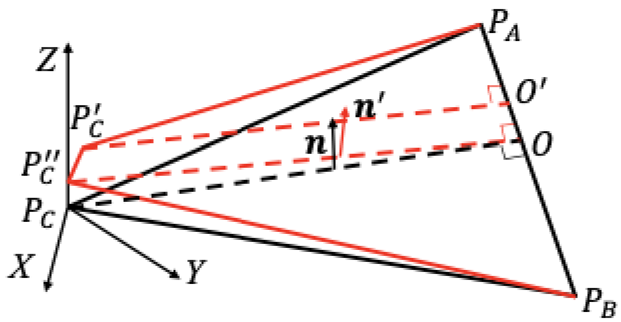
$$\{||\overrightarrow{P_k P_m}|| > \theta, k, m \in group\}$$

  - normal vector

$$\mathbf{n_i} = \frac{\overrightarrow{P_{Ai} P_{Ci}} \times \overrightarrow{P_{Ai} P_{Bi}}}{||\overrightarrow{P_{Ai} P_{Ci}} \times \overrightarrow{P_{Ai} P_{Bi}}||}$$

Robustness to Depth Noise

- The difference between $\mathbf{n'}$ produced by noise and original $\mathbf{n}$ is small



geometric demonstration, math demonstration is valid too.

- local method only focuses on low-order characteristics, however, VN notices high-order ones

Virtual Normal Loss (VNL)

- naive method

$$\mathscr{L}_{VN} = \frac{1}{N}(\Sigma_{i=0}^{N}||\mathbf{n_i}^{pred} - \mathbf{n_i}^{gt}||)$$

Pixel-wise Depth Supervision

- combine with VNL and weighted cross-entropy loss (WCEL)

$$\mathscr{L} = \mathscr{L}_{WCE} + \lambda\mathscr{L}_{VN}$$

  - $\lambda = 5$ is a trade-off parameter

Experiments

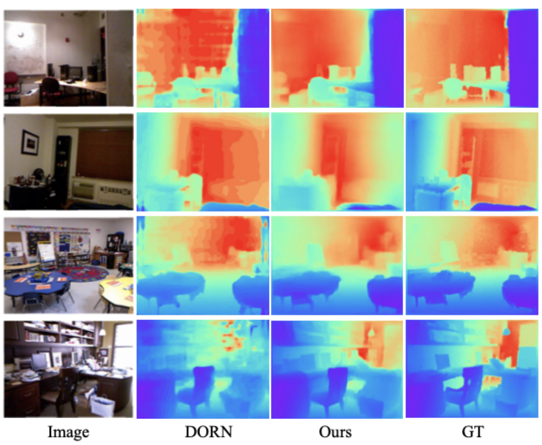
Dataset

- NYU-V2
  - indoor dataset
- KITTI
  - outdoor dataset

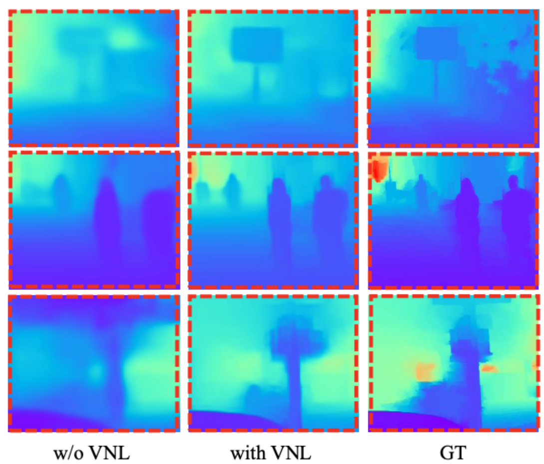
Training Model

- pre-trained ResNet-101 on ImageNet as backbone
  - base $\alpha = 0.0001$, with a decaying
  - batch size=8
  - 10 epochs (NYU-large, KITTI)
  - 40 epochs (NYU-Small)
  - more in paper...

Results



Indoor result compared with DORN(SOTA)



Outdoor result compared with with/out VNL

Table 1. Results on NYUD-V2. Our method outperforms other state-of-the-art methods over all evaluation metrics.

Method	rel	log10	rms	δ_1	δ_2	δ_3
	Lower is better			Higher is better		
Saxena *et al*. [35]	0.349	-	1.214	0.447	0.745	0.897
Karsch *et al*. [20]	0.349	0.131	1.21	-	-	-
Liu *et al*. [29]	0.335	0.127	1.06	-	-	-
Ladicky *et al*. [23]	-	-	-	0.542	0.829	0.941
Li *et al*. [25]	0.232	0.094	0.821	0.621	0.886	0.968
Roy *et al*. [32]	0.187	0.078	0.744	-	-	-
Liu *et al*. [28]	0.213	0.087	0.759	0.650	0.906	0.974
Wang *et al*. [38]	0.220	0.094	0.745	0.605	0.890	0.970
Eigen *et al*. [7]	0.158	-	0.641	0.769	0.950	0.988
Chakrabarti [3]	0.149	-	0.620	0.806	0.958	0.987
Li *et al*. [26]	0.143	0.063	0.635	0.788	0.958	0.991
Laina *et al*. [24]	0.127	0.055	0.573	0.811	0.953	0.988
DORN [12]	0.115	0.051	0.509	0.828	0.965	0.992
Ours	**0.108**	**0.048**	**0.416**	**0.875**	**0.976**	**0.994**

Result on NYU-V2

Table 2. Results on KITTI. Our method outperforms other methods over all evaluation metrics except rms.

Method	δ_1	δ_2	δ_3	rel	rms	rms (log)
	Higher is better			Lower is better		
Make3D [35]	0.601	0.820	0.926	0.280	8.734	0.361
Eigen *et al*. [8]	0.692	0.899	0.967	0.190	7.156	0.270
Liu *et al*. [28]	0.647	0.882	0.961	0.114	4.935	0.206
Semi. [22]	0.862	0.960	0.986	0.113	4.621	0.189
Guo *et al*. [14]	0.902	0.969	0.986	0.090	3.258	0.168
DORN [12]	0.932	0.984	0.994	0.072	**2.727**	0.120
Ours	**0.938**	**0.990**	**0.998**	**0.072**	3.258	**0.117**

Results on KITTI

Ablation Studies

Effectiveness

> shown above

Samples

> more is better

Recover from Estimated Depth

Point Cloud

Surface Normal