End-to-End Wireframe Parsing

Yichao Zhou Haozhi Qi Yi Ma

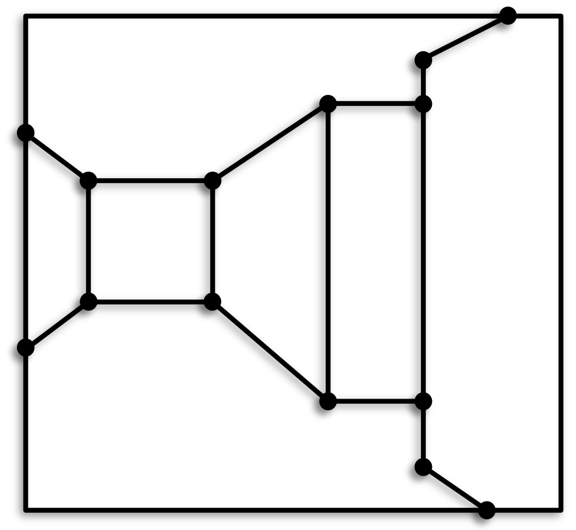
University of California, Berkeley

ICCV 2019 Seoul, Korea

Wireframe Representation

Our representation of wireframes is based on the notation from graph theory:

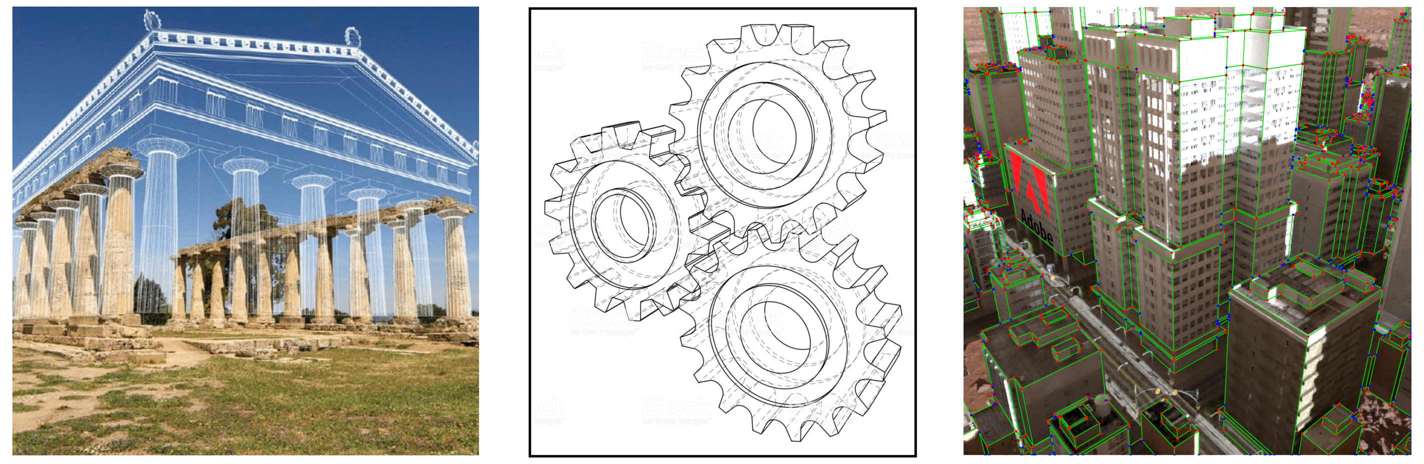
- Let $W = (V, E)$ be a wireframe;
- V is the set of junction indices;
- $E \subseteq V \times V$ is the set of lines;
- For each $\forall i \in V$: $\boldsymbol{p}_i$ represents its coordinate in image space.



Why Wireframe?

- Editable CAD representation;
- Clean geometry from priors of man-made environments;
- Compact, easy for content sharing and transmission;
- High-level structuralized features, as opposed to local features such as SIFT or line segments;
- Now possible with recent advances in deep learning.

Applications



(a) Augmented Reality (b) CAD Reconstruction (c) 3D Editing

Related Work

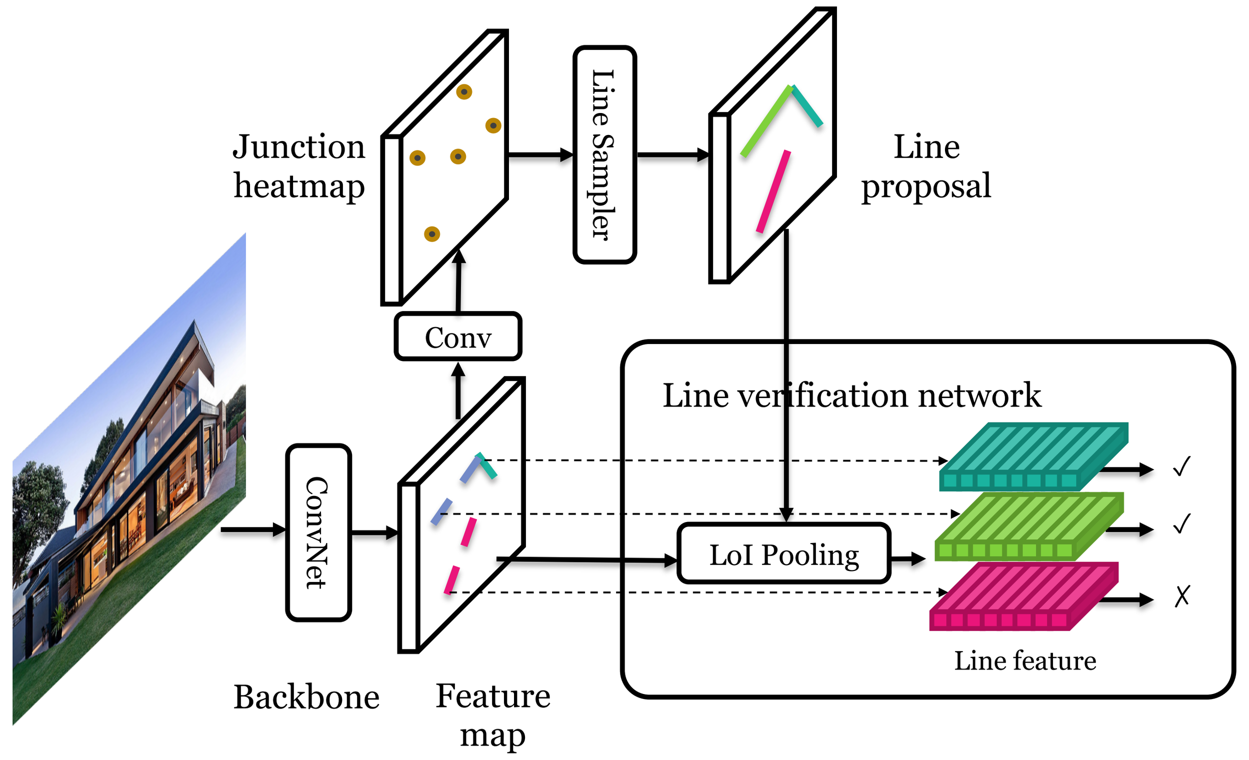
Previous Methods [1,2,3]

- Two-stage algorithms
- First, use neural networks to predict pixelwise heat maps
- Next, apply heuristic algorithms to turn the pixel-wise heat maps into a vectorized format

Our Method (L-CNN)

- End-to-end trainable
- Directly outputs vectorized wireframe, including junctions and lines
- Easy to implement in modern neural network frameworks

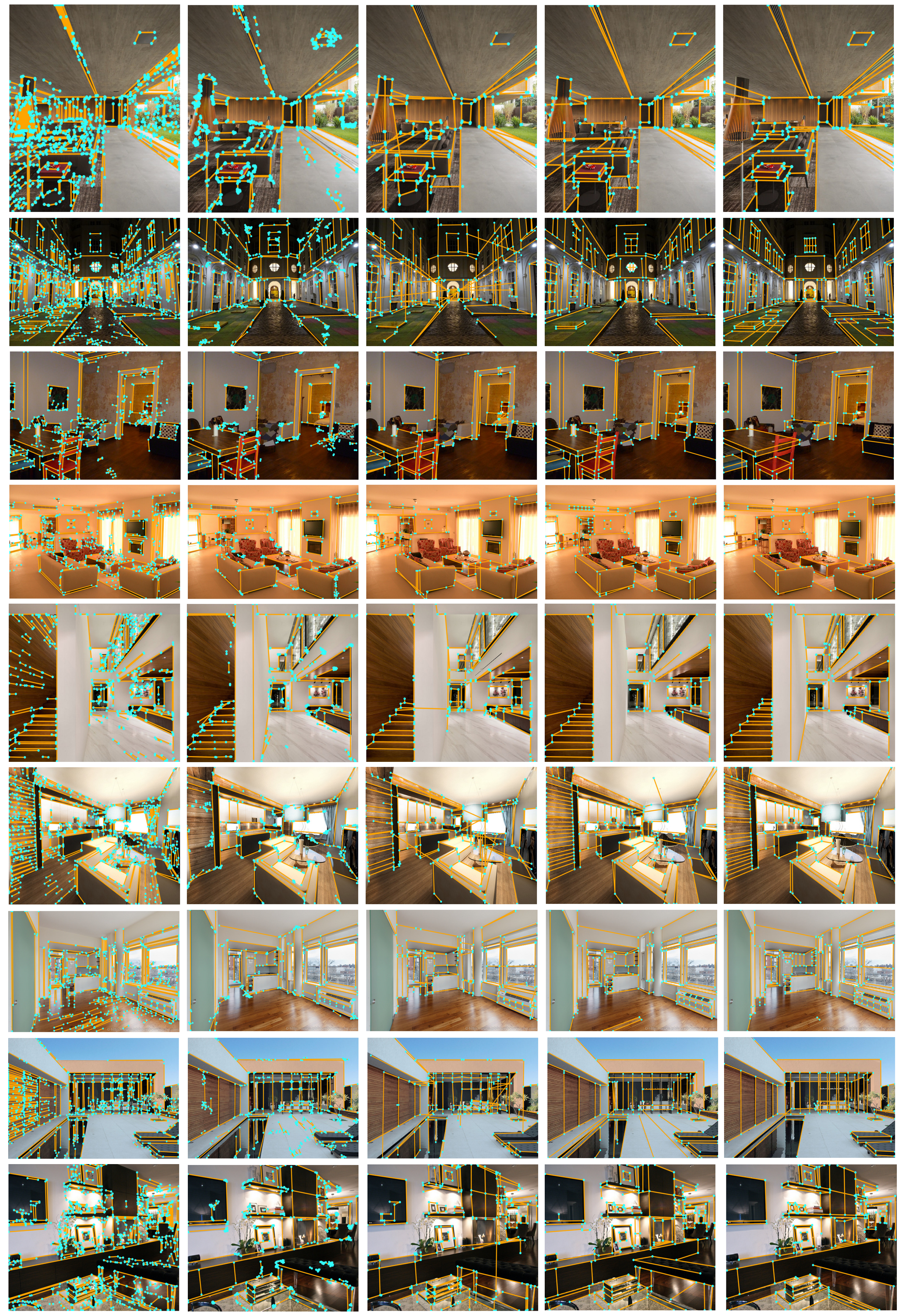
Methods



Datasets and Settings

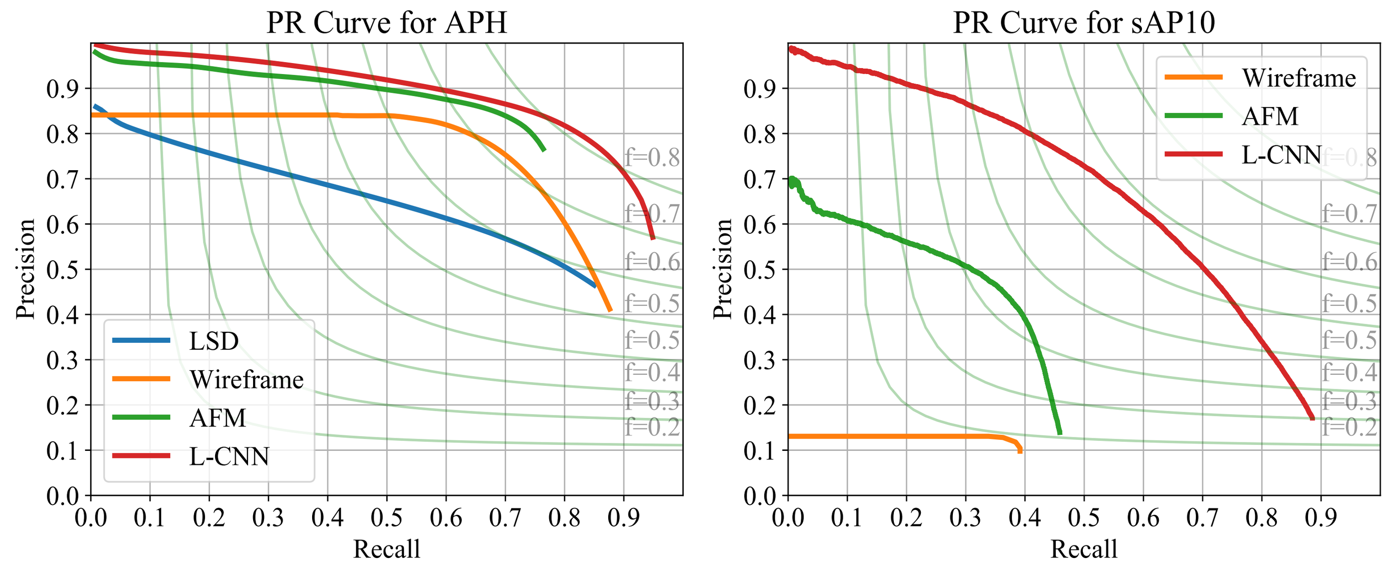
- Experiments on ShanghaiTech dataset [1];
  - Training set: **5,000** images;
  - Testing set: **462** images;
- Trains and tests on a single NVIDIA GTX 1080Ti.

Qualitative Measures



(a) LSD [2] (b) AFM [3] (c) Wireframe [1] **(d) L-CNN** (e) Ground Truth

Quantitative Measures



	sAP[10]	mAP[J]	AP[H]	F[H]
LSD [2]	/	/	52.0	61.0
Wireframe [1]	5.1	40.9	67.8	72.6
AFM [3]	24.4	23.3	69.5	77.2
L-CNN	**62.9**	**59.3**	**83.0**	**81.2**

Acknowledgement

This work is partially supported by Sony US Research Center, FHL Vive Center for Enhanced Reality, Berkeley BAIR, and Bytedance Research Lab. We thank Kenji Tashiro of Sony for helpful discussions. We also thank Cecilia Zhang of Berkeley for her comments on the draft of this paper.

Reference

1. Kun Huang, Yifan Wang, Zihan Zhou, Tianjiao Ding, Shenghua Gao, and Yi Ma. Learning to parse wireframes in images of man-made environments. In CVPR, 2018.
2. Rafael Grompone Von Gioi, Jeremie Jakubowicz, Jean-Michel Morel, and Gregory Randall. LSD: A fast line segment detector with a false detection control. PAMI, 2010.
3. Nan Xue, Song Bai, Fudong Wang, Gui-Song Xia, Tianfu Wu, and Liangpei Zhang. Learning attraction field representation for robust line segment detection. In CVPR, 2019.

Berkeley UNIVERSITY OF CALIFORNIA