ELSD: Efficient Line Segment Detector and Descriptor

Haotian Zhang[1] Yicheng Luo[1] Fangbo Qin[2] Yijia He Xiao Liu[1*]
[1]Megvii Technology [2]Institute of Automation, CAS
{zhanghaotian,luoyicheng}@megvii.com qinfangbo2013@ia.ac.cn
heyijia2016@gmail.com liuxiao@megvii.com

1. Appendix

1.1. Implementation details

Our ELSD uses the backbone of U-shape network that adopts ResNet34[4] as the encoder and optionally Hourglass Network[10] as the backbone. We conduct standard data augmentation for the training set, including horizontal/vertical flip and random rotate. Input images are resized to 512×512. Our model is trained using ADAM[6] optimizer with a total of 170 epochs on four NVIDIA RTX 2080Ti GPUs and an Inter Xeon Gold 6130 2.10 GHz CPU. The initial learning rate, weight decay, and batch size are set to $1e-3$, $1e-5$, and 16 respectively. The learning rate is divided by 10 at the 100th and 150th epoch. It is recommended to train line detector firstly and then jointly train with descriptors, since the line descriptor branch is easier to learn compared to the line detector branch. See the source code in the supplementary file for more details.

1.2. Qualitative Results on Line Segment Detector

We show more visualization results on the Wireframe dataset[5] and YorkUrban dataset[2] in Figure 1 and Figure 2. The configurations for visualization of different methods are as follows:

- The *a-contrario* validation of LSD[3] is set to $-log\ \epsilon = 0.01 \times 1.75^8$.
- The thresholds in line verification of L-CNN[14], HAWP[12] and HT-HAWP[9] are set to 0.98, 0.95 and 0.99 respectively, where the PR curve of sAP^{10} achieves maximum F-score on Wireframe dataset.
- The threshold of root-point detection in TP-LSD is set as 0.43, where the PR curve of sAP^{10} also achieves maximum F-score.
- For Our ELSD, the threshold of mid-point's score after Non-Centerness Suppression is set to 0.22 for the same purpose.

*Corresponding author. This paper is supported by the National Key R&D Plan of the Ministry of Science and Technology (Project No. 2020AAA0104400).

Methods	Dimension	Precision(%)	Recall(%)	F-Score(%)
LBD	78	69.3	63.8	66.4
LLD	64	57.5	43.6	49.6
DLD	8	52.2	42.6	47.0
WLD	16	67.0	57.2	61.7
Ours	256	72.6	**77.1**	74.7
	64	**73.5**	76.2	**74.8**
	36	72.2	75.3	73.7
	16	68.4	69.7	69.1
	8	60.3	60.1	60.2

Table 1: Line descriptor evaluation by line matching.

1.3. Qualitative Results on Line Descriptor

To perform the quantitative and qualitative evaluation for line matching using different descriptors, we select about 1000 image pairs from ScanNet[1] dataset that includes large viewpoint change, rotation change, and scale change. We further visualize the line matching results of LBD[13], LLD[11] and our 64-dimensional descriptor. We use the OpenCV implementation of 72-dimensional LBD descriptors and the official model of LLD descriptors. Note that we find the nearest neighbors to match lines across descriptors and perform cross-checking. The results are shown in Figure 3.

1.4. Comprehensive comparison on Line Description

We add the comparison with the official DLD[8] and WLD[7] models, under the same evaluation setting in Section 4.4. We also train ELSD with setting the same descriptor dimension as DLD and WLD respectively. The whole results are shown in Table 1. Our descriptors outperform DLD and WLD, mainly due to ELSD's tightly-coupled learning of detector and descriptor.

1.5. The ablation study for the use of static and dynamic lines.

We train ELSD using only static lines, and the result is reported in Table 2. Using both static and dynamic lines in-

No.	Static	Dynamic	Dimension	Precision(%)	Recall(%)	F-Score(%)
1	✓	✓	64	**73.6**	**76.2**	**74.8**
2	✓		64	71.1	71.6	71.3

Table 2: Ablation study on dynamic lines in descriptor learning.

deed boosts the performance, compared to only using static lines.

References

[1] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 2017. 1

[2] Patrick Denis, James H. Elder, and Francisco J. Estrada. Efficient edge-based methods for estimating manhattan frames in urban imagery. In David Forsyth, Philip Torr, and Andrew Zisserman, editors, *Computer Vision – ECCV 2008*, pages 197–210, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. 1

[3] Rafael Gioi, Jeremie Jakubowicz, Jean-Michel Morel, and Gregory Randall. Lsd: A fast line segment detector with a false detection control. *IEEE transactions on pattern analysis and machine intelligence*, 32:722–32, 04 2010. 1

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 1

[5] 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*, June 2018. 1

[6] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. 1

[7] M. Lange, Claudio Raisch, and A. Schilling. Wld: A wavelet and learning based line descriptor for line feature matching. In *VMV*, 2020. 1

[8] M. Lange, F. Schweinfurth, and A. Schilling. Dld: A deep learning based line descriptor for line feature matching. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5910–5915, 2019. 1

[9] Yancong Lin, Silvia L. Pintea, and Jan C. van Gemert. Deep hough-transform line priors. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, 2020. 1

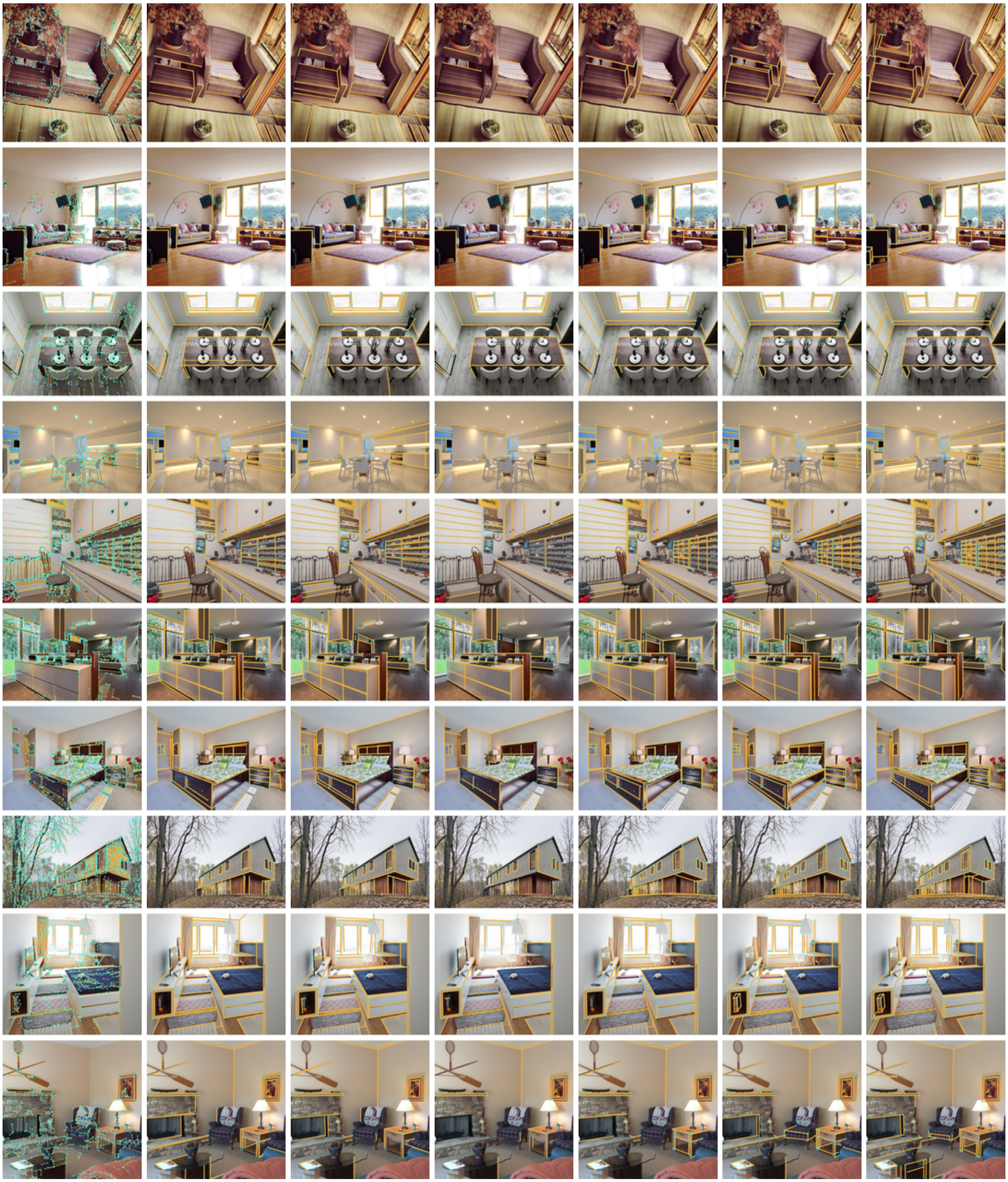
[10] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hourglass networks for human pose estimation. In *ECCV (8)*, pages 483–499, 2016. 1

[11] A. Vakhitov and V. Lempitsky. Learnable line segment descriptor for visual slam. *IEEE Access*, 7:39923–39934, 2019. 1

[12] Nan Xue, Tianfu Wu, Song Bai, Fudong Wang, Gui-Song Xia, Liangpei Zhang, and Philip H.S. Torr. Holistically-attracted wireframe parsing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1

[13] Lilian Zhang and Reinhard Koch. An efficient and robust line segment matching approach based on lbd descriptor and pairwise geometric consistency. *Journal of Visual Communication and Image Representation*, 24(7):794–805, 2013. 1

[14] Yichao Zhou, Haozhi Qi, and Yi Ma. End-to-end wireframe parsing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 1



(a) LSD (b) L-CNN (c) HAWP (d) HT-HAWP (e) TP-LSD (f) ELSD (Ours) (g) GT

Figure 1: Visualization of line detection methods on Wireframe dataset.

(a) LSD (b) L-CNN (c) HAWP (d) HT-HAWP (e) TP-LSD (f) ELSD (Ours) (g) GT

Figure 2: Visualization of line detection methods on YorkUrban dataset.

(a) LBD (b) LLD (c) ELSD (Ours) (d) GT

Figure 3: Line matching results using descriptors. Note that the blue lines are predicted by our ELSD, and the green / red lines represent the true / false positive matches.