Supplementary Material

Abstract—**Recently, a number of deep-learning methods have been proposed for a closely related task: the *wireframe parsing*, which is very likely to be confused with the line segment detection task studied in our work. In this supplementary material, the difference between these two tasks is clarified through a comparison between a typical method of wireframe parsing and our proposed CGLSD.**

VI. Comparing with a wireframe parsing method

The *wireframe parsing* aims to encode large-scale geometric shapes of objects by extracting *salient* line segments and junctions from the input image [25]. Those developed deep-learning models have also been often termed 'line segment detectors' by their authors, causing a confusion with the line segment detectors (including ours) investigated in our work very easily. However, the wireframe parsing methods are not supposed to be compared with our proposed CGLSD (or, any existing line segment detectors) through a usual manner, i.e., on the same datasets and with a set of commonly-used evaluation metrics.

For clarification, a typical deep-learning method of wireframe parsing, the HAWP [26], is selected to compare with our proposed CGLSD, as depicted in Fig. 8. First, an image taken from the YorkUrban-LineSegment dataset [17] is tested, referring to Fig. 8(a). It is seen that the HAWP [26] detects much fewer line segments and thus yields a much lower F-score, when compared with that of our CGLSD. Similar observation can be made for almost every test image taken from the YorkUrban-LineSegment dataset [17]. It could be thus concluded that our CGLSD can outperform the HAWP [26] significantly. However, this is not the full story. In Fig. 8(b), an image taken from the Wireframe dataset [25] is tested. Note that the Wireframe [25] is the dataset where the HAWP (as well as all other existing wireframe parsing models) is trained. In this round of comparison, although the HAWP still detects much fewer line segments, it achieves a much higher F-score.

In conclusion, the existing deep-learning models, although also termed 'line segment detectors' frequently, are trained to detect the wireframe of a scene *only*. On the other hand, our proposed CGLSD (as well as all the existing line segment detectors reviewed in our manuscript) is designed to detect *every* line segment from the input image. For a given scene, these two kinds of methods have quite different understanding and thus are not supposed to be compared with each other by using the same evaluation metric such as the F-sore. In fact, the measurement of F-score depends heavily on how the ground-truth line segments are annotated on the input image, and the annotation results are remarkably different for these two considered tasks.

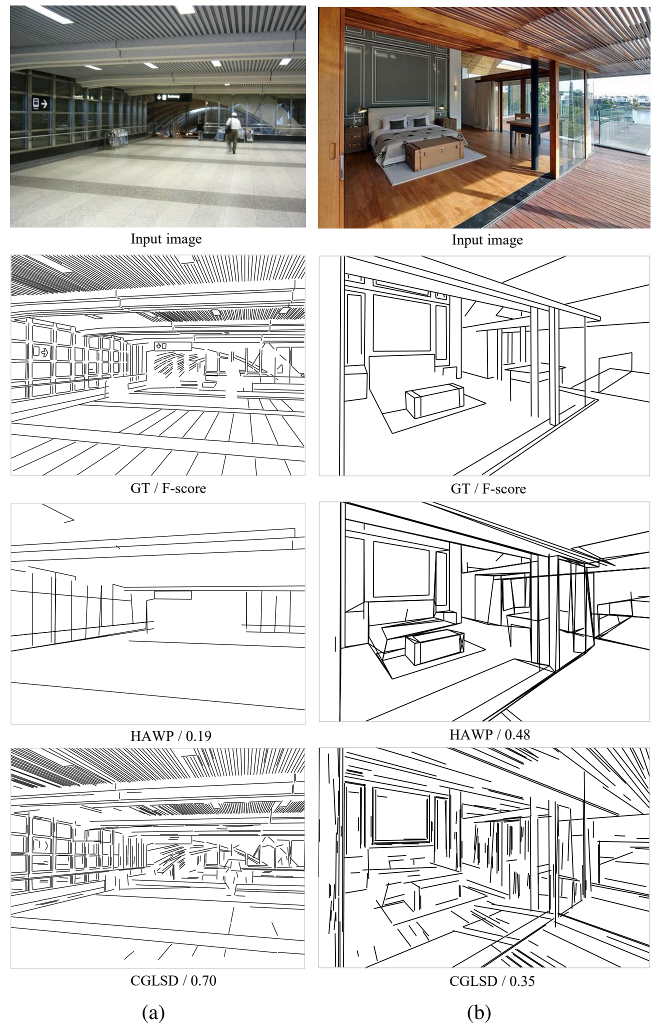

Fig. 8. A performance comparison between the HAWP [26] and our CGLSD by using (a) the image #53 taken from the YorkUrban-LineSegment dataset [17] and (b) the image #4 taken from the test set of the Wireframe dataset [25], respectively.

References

[25] K. Huang, Y. Wang, Z. Zhou, T. Ding, S. Gao, Y. Ma, Learning to parse wireframes in images of man-made environments, in: Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 626–635.

[26] N. Xue, T. Wu, S. Bai, F. Wang, G.-S. Xia, L. Zhang, P. H. Torr, Holistically attracted wireframe parsing, in: Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 2788–2797.