A. Supplementary Material

A.1. Post Processing

Although the junction proposal network uses the non-maximum suppression (Section 3.3) on junctions, we find that it is still common for L-CNN to generate overlapped lines. The is often because there exist collinear junctions in the scenes. Those overlapped lines are visually unpleasant and affect quantitative performance metrics such as AP^H. Therefore, we need to remove those overlapped lines via a post-processing stage. Our strategy is inspired by the post processing technique from object detection [2]. For each pair of lines $L_i = \{(\tilde{\mathbf{p}}_i^1, \tilde{\mathbf{p}}_i^2)\}$ and $L_j = \{(\tilde{\mathbf{p}}_j^1, \tilde{\mathbf{p}}_j^2)\}$ in which L_i is ranked above L_j according the line verification network, if L_i is *close* to L_j (defined below), we

1. delete line L_j if both the projections of points $\tilde{\mathbf{p}}_j^1$ and $\tilde{\mathbf{p}}_j^2$ to L_i fall inside the segment L_i;
2. cut line L_j so that it does not overlap with L_i if only one of the projections of points $\tilde{\mathbf{p}}_j^1$ and $\tilde{\mathbf{p}}_j^2$ to L_i fall inside the segment L_i;
3. retain line L_j otherwise.

We consider line segment L_i close to L_j if and only if

$$\begin{aligned}\min(&\max(d(\tilde{\mathbf{p}}_j^1, L_i), d(\tilde{\mathbf{p}}_j^2, L_i)),\\ &\max(d(\tilde{\mathbf{p}}_i^1, L_j), d(\tilde{\mathbf{p}}_i^2, L_j))) \le \eta_S,\end{aligned}$$

where $d(\mathbf{p}, L)$ represents the distance between point $\mathbf{p}$ and line segment L. We set $\eta_S = 0.01$ in our experiments.

A.2. Junction Evaluation

Our vectorized junction mean average precision mAP^J evaluates the quality of vectorized junctions of a wireframe detection algorithm without using heat maps as in [3]. This metric is inspired by [1]. For a given ranked list of predicted junction positions, a junction is considered to be correct if the ℓ^2 distance between this junction and its nearest ground-truth is within a threshold. Using this criteria, we can draw the precision recall curve by counting the number of true positive and false position. The AP is defined to be the area under this curve. The mean AP is defined to be the average of AP over difference distance thresholds. In our implementation, we choose to average on 0.5, 1.0, and 2.0 threshold on 128×128 resolution.

One major difference between line detection and wireframe detection is that the wireframe representation contains junctions, intersections of lines. Junctions often have the physical meaning in 3D (corners or occlusiual points), and enforcing the line endpoints on junctions leads to a cleaner result. Table 3 shows the mAP^J performance of AFM [5], Wireframe [3] and our L-CNN. You can see that the junction quality of AFM is the worst because it is designed for line detection so it does not model the junction explicitly. Our L-CNN outperforms the previous wireframe parser [3] by a large margin, probably due to the introduction of non-maximum suppression, better junction proposal network, and joint training with the line verification network.

	AFM [5]	Wireframe [3]	L-CNN (ours)
mAP^J	23.3	40.9	**57.3**

Table 3: Junction accuracy of multiple wireframe and line detection algorithms on the wireframe dataset. Because the line detection algorithm AFM [5] does not output junctions, we evaluate the mAP^J with its line endpoints.

A.3. Line Features

Besides the features from the LoIPool layers, we also design and test the manual feature derived from the coordinates of lines' endpoints. For each line $L_j = (\tilde{\mathbf{p}}_j^1, \tilde{\mathbf{p}}_j^2)$, let the 6-dimension line feature vector be

$$\mathbf{F}_j = \begin{bmatrix}\tilde{\mathbf{p}}_j^{1T} & \tilde{\mathbf{p}}_j^{2T} & \left(\frac{\tilde{\mathbf{p}}_j^1 - \tilde{\mathbf{p}}_j^2}{\|\tilde{\mathbf{p}}_j^1 - \tilde{\mathbf{p}}_j^2\|_2}\right)^T\end{bmatrix},$$

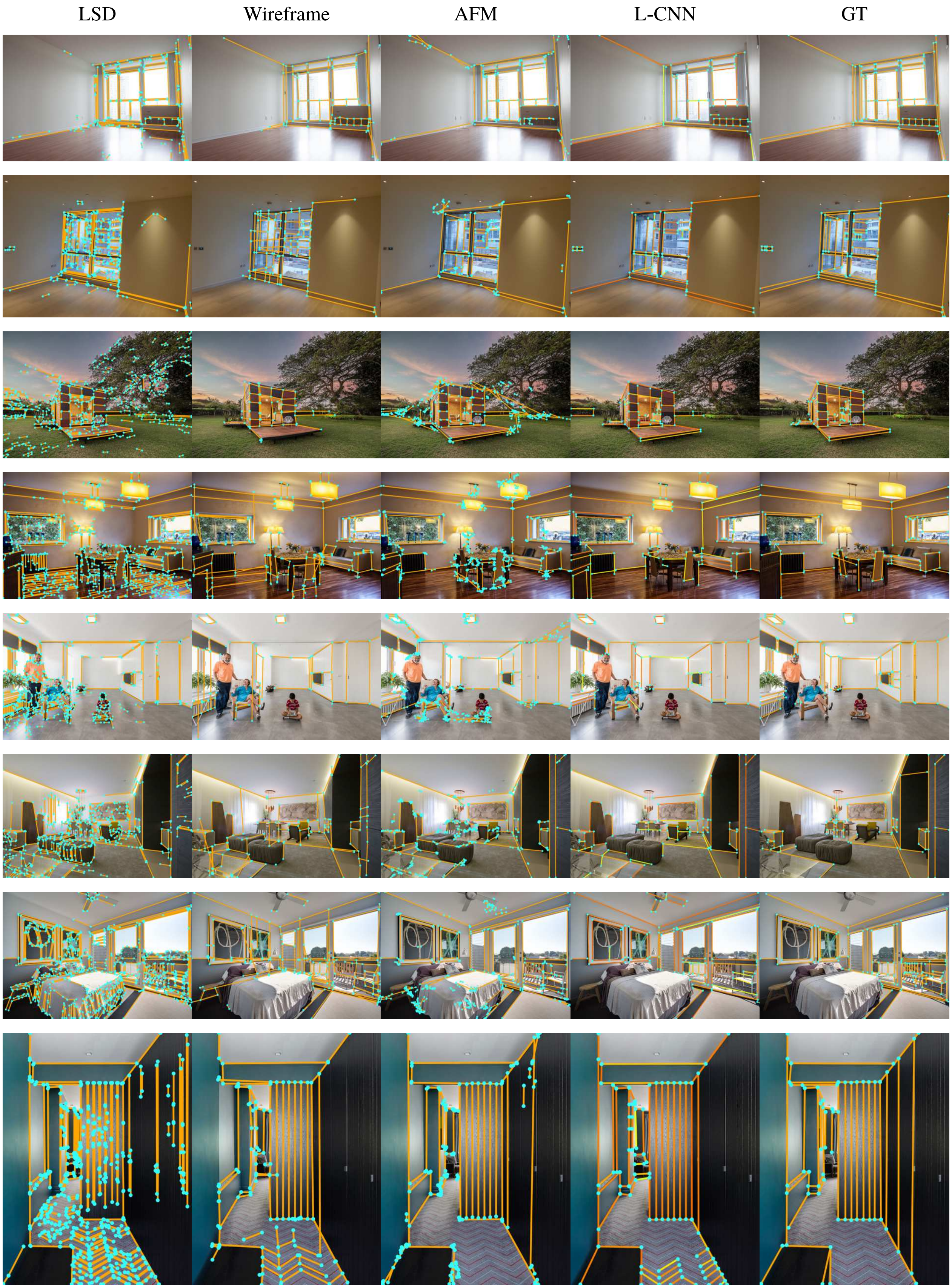
in which the first four dimensions store the coordinates of its endpoints and the last two dimensions store its normalized line directions. According to Table 4, we find that those manual features do not really improve the performance. This is probably because the features from LoIPool layers are already powerful enough for the line verification network. We also observe that the validation loss starting increasing prematurely during the training when we add the coordinate-based feature, a phenomena indicating overfitting, so we do not include this feature into our final L-CNN.

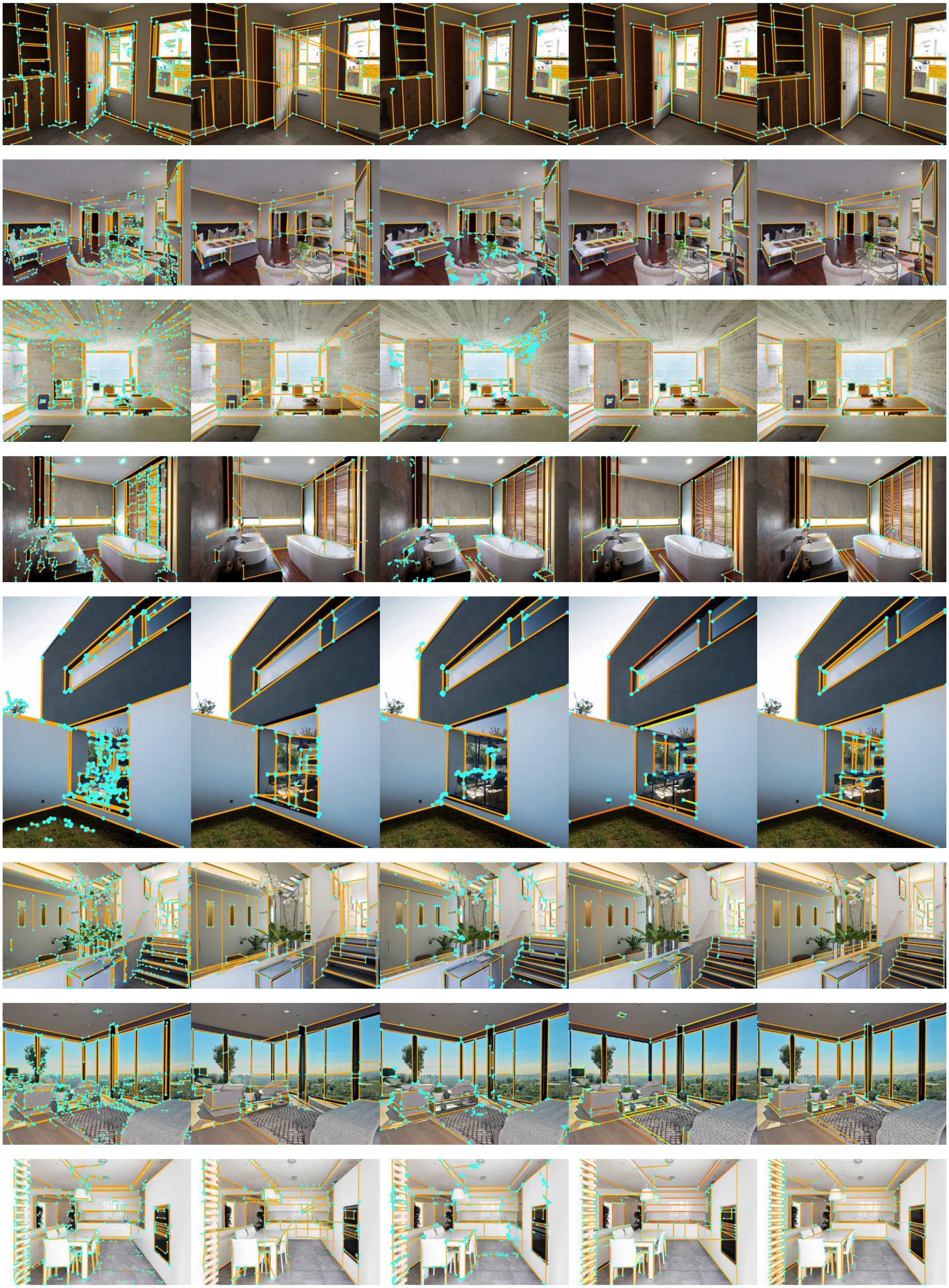
	cood	slope	sAP^5	sAP^{10}	sAP^{15}
(a)			**53.2**	**57.2**	**59.1**
(b)	✓		53.2	57.2	59.0
(c)		✓	53.1	57.1	59.0
(d)	✓	✓	52.3	56.4	58.3

Table 4: Ablation study of the coordinate-based manual line feature. The column labelled with "cood" represents whether the first four dimension of $\mathbf{F}$, the coordinates of lines' endpoints, are used as features, and the column labelled with "slope" represents whether the last two dimensions of $\mathbf{F}$, the slope of the lines, are used as features.

A.4. More Qualitative Sampled Results

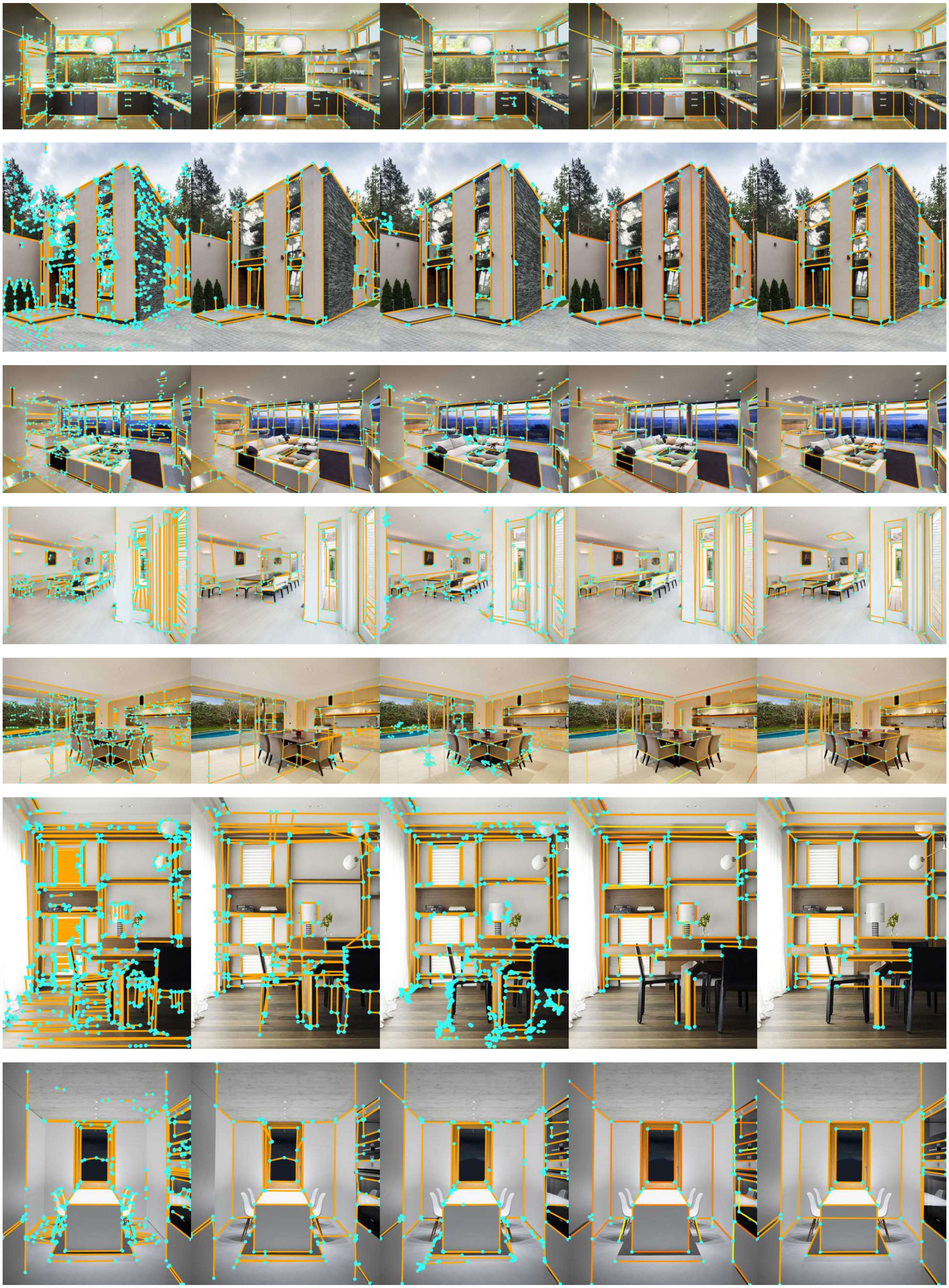
We show random sampled results from the wireframe dataset. From LSD [4], Wireframe parser [3] and AFM [5], we use the same hyper-parameters and cut-off values as in [5]. For our L-CNN, we display predicted lines with scores greater than 0.98. The colors of the lines indicate their confidences.

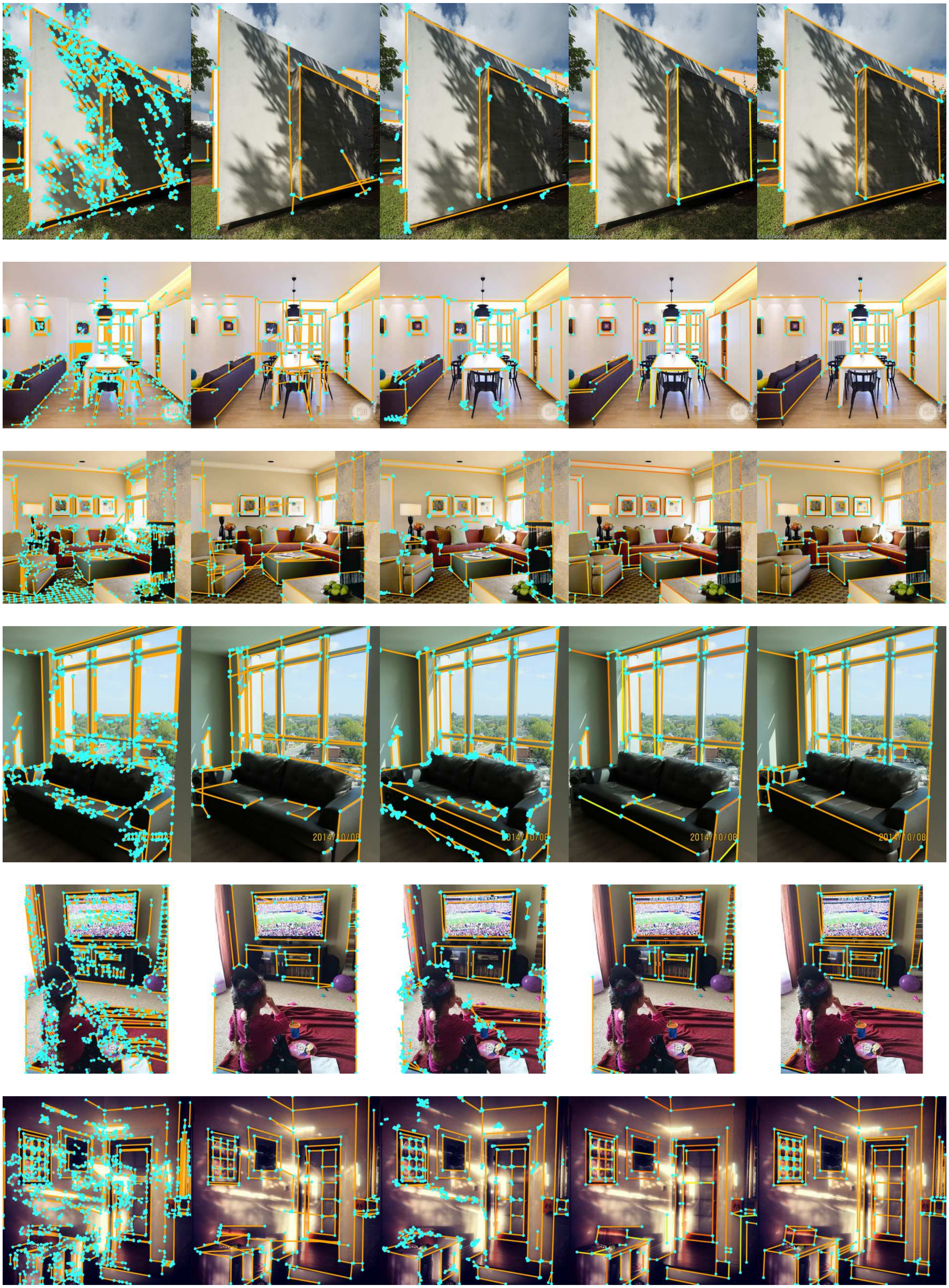

LSD
Wireframe
AFM
L-CNN
GT

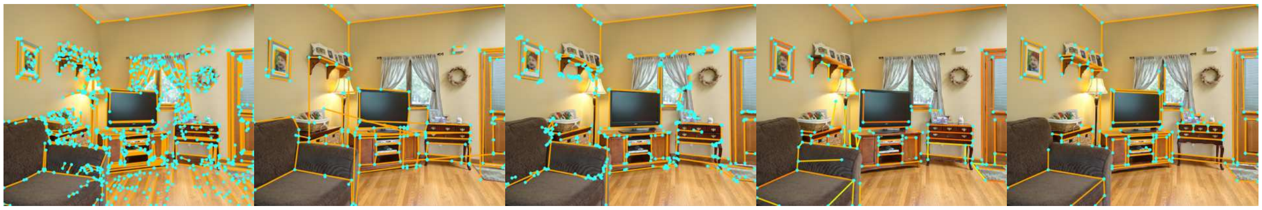



MAIN LIVING ROOM


www.rentfluff.com







References

[1] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The PASCAL visual object classes (VOC) challenge. *IJCV*, 88(2):303–338, 2010.

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014.

[3] K. Huang, Y. Wang, Z. Zhou, T. Ding, S. Gao, and Y. Ma. Learning to parse wireframes in images of man-made environments. In *CVPR*, 2018.

[4] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall. LSD: A fast line segment detector with a false detection control. *PAMI*, 32(4):722–732, 2010.

[5] N. Xue, S. Bai, F. Wang, G.-S. Xia, T. Wu, and L. Zhang. Learning attraction field representation for robust line segment detection. In *CVPR*, 2019.