SSLayout360: Semi-Supervised Indoor Layout Estimation from 360° Panorama

Supplementary Material

Phi Vu Tran
Flyreel AI Research
vuptran@flyreel.co

A. SSLayout360 Architecture Details

Figure A1 shows an expanded version of the SSLayout360 architecture depicted in Figure 2 of the main paper.

B. Implementation Details

B.1. Dataset Details

Table B1 provides summary statistics of the datasets used in the empirical evaluation of this paper. We utilize the standard training, validation, and test splits provided by Zou *et al*. [12, 13] for PanoContext, Stanford-2D3D, and MatterportLayout. For PanoContext evaluation, we follow the authors' protocol of combining the training split of PanoContext with the entire Stanford-2D3D dataset for a total of 963 labeled training examples. Similarly, we augment the Stanford-2D3D training split with the entire PanoContext dataset, resulting in a combined total of 916 labeled training instances, for the evaluation of the Stanford-2D3D dataset. In our semi-supervised experiments, we follow the standard practice of combining the training and validation dataset splits, discarding all label information, as the source of unlabeled data [4, 7, 8]. Thus, we use 1,009 unlabeled instances for semi-supervised experiments on PanoContext, 949 on Stanford-2D3D, and 1,837 on MatterportLayout.

Structured3D is a large photo-realistic *synthetic* dataset comprising 21,835 panoramas of rooms in 3,500 diverse indoor scenes with ground truth cuboid and non-cuboid layout annotations. We set aside the first 3,000 scenes (corresponding to 18,362 images) for training, and equally split the last 500 scenes into 250 validation (1,776 images) and 250 test (1,697 images) sets. We pre-train HorizonNet on 18,362 synthetic images and perform transfer learning on MatterportLayout via supervised fine-tuning. In our semi-supervised experiments, we leverage HorizonNet pre-trained on Structured3D as a strong layout predictor initialized with synthetic data to further evaluate SSLayout360 on MatterportLayout.

Dataset	# Training	# Validation	# Testing	# Unlabeled
PanoContext [10]	413	46	53	1,009
Stanford-2D3D [1]	404	33	113	949
MatterportLayout [13]	1,650	187	458	1,837
Structured3D [11]	18,362	1,776	1,697	–

Table B1. Summary statistics of the datasets used in the empirical evaluation. See text for more details.

Cuboid Layout	Training Epochs				
Dataset	20 labels	50 labels	100 labels	200 labels	All labels
PanoContext	200 (50)	200 (100)	200 (100)	200 (100)	300 (300)
Stanford-2D3D	200 (50)	200 (100)	200 (100)	200 (100)	300 (300)
Non-Cuboid Layout	**Training Epochs**				
Dataset	50 labels	100 labels	200 labels	400 labels	1,650 labels
Structured3D FT	100 (50)	100 (50)	100 (50)	100 (50)	100 (50)
MatterportLayout	200 (50)	200 (100)	200 (100)	300 (100)	300 (300)

Table B2. The number of training epochs as a function of labeled examples for each dataset under consideration across both supervised and semi-supervised settings. The number in parenthesis indicates the semi-supervised counterpart. `All labels` refers to 963 labels for PanoContext and 916 labels for Stanford-2D3D. `Structured3D FT` refers to supervised and semi-supervised fine-tuning experiments using the Structured3D synthetic dataset.

B.2. Training Epochs

Table B2 summarizes the number of training epochs we run in our supervised and semi-supervised experiments, expressed as a function of labeled examples for each dataset under consideration. In the supervised setting, we define an epoch as one pass over available labeled examples in $\mathcal{D}_L$. In the semi-supervised setting, an epoch is defined as one pass over all unlabeled examples in $\mathcal{D}_U$. In short, we train our models between 50 and 300 epochs, depending on the setting and how many unlabeled instances are used in combination with labeled examples.

For the `Structured3D FT` row, we first train HorizonNet on 18,362 synthetic images for 50 epochs, and then transfer the learned weights to evaluate on MatterportLayout

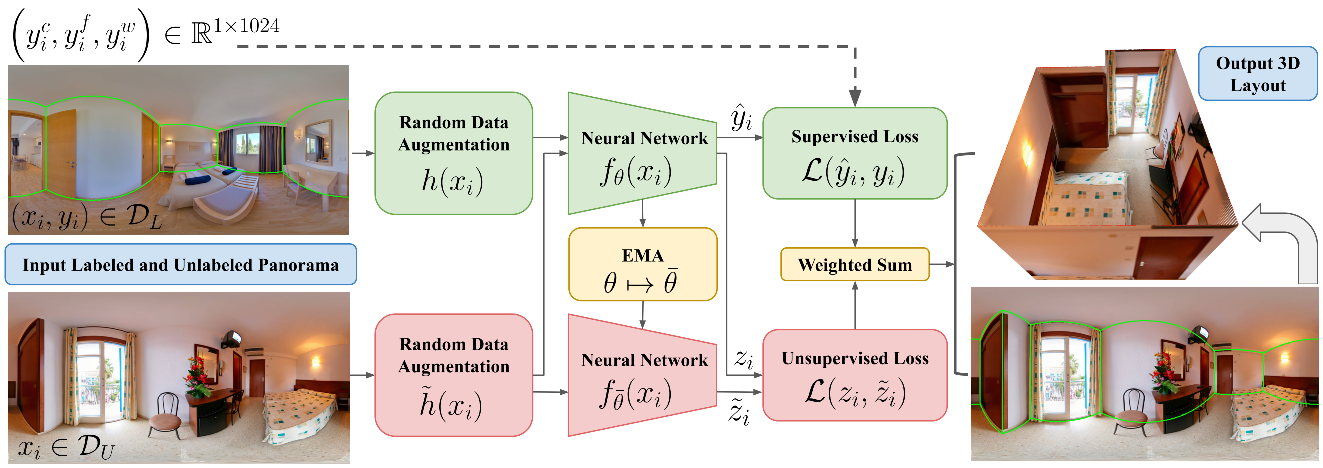

Figure A1. An illustration of the SSLayout360 architecture for semi-supervised indoor layout estimation from a 360° panoramic scene.

via supervised fine-tuning with an additional 100 epochs when using 50, 100, 200, 400, or 1,650 labels. In the semi-supervised setting, we leverage pre-trained HorizonNet as a strong predictor and train SSLayout360 to convergence with only 50 epochs when using 50, 100, 200, 400, or 1,650 labels. The reduced number of training epochs is an expected benefit when performing transfer learning.

Lastly, (not shown in Table B2) in MatterportLayout experiments with 4,000 unlabeled images, we train SSLayout360 for 100 epochs when using 100, 200, 400 labels, and 300 epochs using 1,650 labels. For MatterportLayout experiments with 10,454 unlabeled images, we train SSLayout360 for 100 epochs when using 100, 200, 400, or 1,650 labels.

B.3. Hyper-Parameters

Table B3 summarizes the shared hyper-parameters as a function of dataset for both supervised and semi-supervised settings. Not shown in Table B3 are three hyper-parameters specific to our SSLayout360 algorithm, which are kept constant in all semi-supervised experiments: the unsupervised (or consistency) loss weight $\lambda = 1$, the ramp-up period T@30%, and EMA decay coefficient $\alpha = 0.999$. We anneal the learning rate hyper-parameter after each training step t according to the polynomial schedule: $\text{lr} \times (1 - {}^{t}/{}_{t_{\text{max}}})^{0.5}$. In general, we fix the hyper-parameters constant and do not attempt to tune them on a per-dataset or per-experiment basis, which can limit the real-world applicability of our method.

B.4. Model and Data Perturbation

SSLayout360 relies on two sources of noise perturbation for its success: random model dropout [5] and input data augmentation. In conjunction with Algorithm 1, we employ the following procedure to obtain consistent student-teacher predictions for semi-supervised indoor layout estimation:

1. Separate the input data source into labeled and unlabeled branches. The unlabeled branch consists of all available training and validation examples, but without ground truth label information.
2. Apply random data augmentation consisting of panoramic stretching [6], horizontal rotation, left-right flipping, and gamma correction to the labeled branch as input to the *student* model.
3. Apply the same set of data augmentation to the unlabeled branch, but without gamma correction and with a different random seed, as another set of input to the *student* model.
4. Perturb the resulting output of Step 3 with gamma correction as (noisy) input to the *teacher* model.
5. Enforce the consistency constraint on the student-teacher model outputs from Steps 3-4 per Algorithm 1 and Equation (4) described in the main paper.

With this procedure, we introduce noise perturbation to the teacher's unlabeled input via gamma correction with $\gamma \in [0.5, 2]$. We also rely on random dropout with probability 0.5 at each forward pass for additional model perturbation to help regularize the teacher's unsupervised targets.

Dataset	Batch Size	Learning Rate	Adam β_1, β_2 [2]
PanoContext	8 (8)	0.0003	0.9, 0.999
Stanford-2D3D	8 (8)	0.0003	0.9, 0.999
Structured3D	8 (–)	0.0003	0.9, 0.999
MatterportLayout	4 (4)	0.0001	0.9, 0.999
Extra Unlabeled	4 (12)	0.0001	0.9, 0.999

Table B3. Shared training hyper-parameters as a function of dataset for both supervised and semi-supervised settings. The `Batch Size` column indicates the blend of labeled (unlabeled) images for semi-supervised experiments or just labeled images for supervised experiments. In the `Extra Unlabeled` experiments, we use a mini-batch with a mixture of 4 labeled and 12 unlabeled examples to accelerate model training.

PanoContext	3D IoU (%) ↑				
Method	20 labels 20 images	50 labels 50 images	100 labels 100 images	200 labels 200 images	963 labels 963 images
HorizonNet	61.48 ± 2.07	63.84 ± 2.87	65.43 ± 1.30	75.76 ± 0.62	83.55 ± 0.31
SSLayout360	62.21 ± 1.65	67.15 ± 1.25	69.14 ± 1.23	77.55 ± 0.89	82.56 ± 1.05
	Corner Error (%) ↓				
Method	20 labels 20 images	50 labels 50 images	100 labels 100 images	200 labels 200 images	963 labels 963 images
HorizonNet	3.51 ± 0.79	2.78 ± 0.90	3.17 ± 0.27	1.07 ± 0.15	0.70 ± 0.02
SSLayout360	3.09 ± 0.55	2.37 ± 0.56	2.02 ± 0.40	1.06 ± 0.27	0.75 ± 0.09
	Pixel Error (%) ↓				
Method	20 labels 20 images	50 labels 50 images	100 labels 100 images	200 labels 200 images	963 labels 963 images
HorizonNet	5.68 ± 0.52	5.03 ± 0.45	4.75 ± 0.06	3.17 ± 0.17	1.97 ± 0.03
SSLayout360	5.52 ± 0.48	4.35 ± 0.36	4.49 ± 0.28	3.03 ± 0.53	2.03 ± 0.19

Stanford-2D3D	3D IoU (%) ↑				
Method	20 labels 20 images	50 labels 50 images	100 labels 100 images	200 labels 200 images	916 labels 916 images
HorizonNet	62.20 ± 3.98	68.27 ± 1.45	69.94 ± 3.64	74.95 ± 3.69	82.79 ± 0.90
SSLayout360	69.19 ± 2.34	70.66 ± 1.62	74.21 ± 1.33	77.75 ± 1.30	84.54 ± 0.59
	Corner Error (%) ↓				
Method	20 labels 20 images	50 labels 50 images	100 labels 100 images	200 labels 200 images	916 labels 916 images
HorizonNet	2.70 ± 0.60	1.64 ± 0.12	1.66 ± 0.20	1.50 ± 0.18	0.64 ± 0.02
SSLayout360	2.19 ± 0.61	1.79 ± 0.42	1.43 ± 0.16	1.13 ± 0.15	0.63 ± 0.02
	Pixel Error (%) ↓				
Method	20 labels 20 images	50 labels 50 images	100 labels 100 images	200 labels 200 images	916 labels 916 images
HorizonNet	5.03 ± 0.51	3.95 ± 0.22	3.77 ± 0.39	3.69 ± 0.26	2.13 ± 0.05
SSLayout360	4.24 ± 0.41	4.07 ± 0.15	3.54 ± 0.11	3.27 ± 0.35	1.99 ± 0.04

Table C1. Quantitative cuboid layout results, without the use of unlabeled data, evaluated on the **PanoContext (left)** and **Stanford-2D3D (right)** test sets averaged over four independent runs with different random seeds. Number format: mean value ± standard deviation.

Mixed Corners	3D IoU (%) ↑				
Method	50 labels 50 images	100 labels 100 images	200 labels 200 images	400 labels 400 images	1,650 labels 1,650 images
HorizonNet	63.44 ± 0.56	68.79 ± 0.49	72.25 ± 0.50	74.46 ± 0.35	79.12 ± 0.37
SSLayout360	63.21 ± 0.73	68.71 ± 0.22	73.01 ± 0.24	75.38 ± 0.72	80.06 ± 0.26
	2D IoU (%) ↑				
Method	50 labels 50 images	100 labels 100 images	200 labels 200 images	400 labels 400 images	1,650 labels 1,650 images
HorizonNet	67.17 ± 0.65	72.06 ± 0.49	75.16 ± 0.53	77.15 ± 0.36	81.54 ± 0.31
SSLayout360	66.76 ± 0.74	71.83 ± 0.24	75.78 ± 0.26	77.99 ± 0.70	82.35 ± 0.29
	δ_1 ↑				
Method	50 labels 50 images	100 labels 100 images	200 labels 200 images	400 labels 400 images	1,650 labels 1,650 images
HorizonNet	0.76 ± 0.01	0.84 ± 0.01	0.89 ± 0.01	0.91 ± 0.01	0.94 ± 0.01
SSLayout360	0.77 ± 0.01	0.85 ± 0.01	0.89 ± 0.01	0.91 ± 0.01	0.95 ± 0.01
	RMSE ↓				
Method	50 labels 50 images	100 labels 100 images	200 labels 200 images	400 labels 400 images	1,650 labels 1,650 images
HorizonNet	0.41 ± 0.01	0.34 ± 0.01	0.30 ± 0.01	0.28 ± 0.01	0.23 ± 0.01
SSLayout360	0.41 ± 0.01	0.34 ± 0.01	0.29 ± 0.01	0.27 ± 0.02	0.23 ± 0.01

Table C2. Quantitative non-cuboid layout results, without the use of unlabeled data, evaluated on the MatterportLayout test set with mixed corners averaged over four independent runs with different random seeds. Number format: mean value ± standard deviation.

C. SSLayout360 without Unlabeled Data

Tables C1 – C2 compare the fully supervised HorizonNet with SSLayout360 without the use of unlabeled data. This setting corresponds to fully supervised learning with consistency regularization. In this setting, SSLayout360 only uses the available labeled training examples as the source of "unlabeled data" to generate unsupervised proxy targets for enforcing the consistency constraint. We observe that for PanoContext and Stanford-2D3D in Table C1, SSLayout360 without unlabeled data produces slightly better results than the supervised HorizonNet counterpart across most settings and metrics. Similarly, for MatterportLayout in Table C2, SSLayout360 gives competitive or slightly better results than the supervised HorizonNet counterpart. Our findings from these experiments corroborate previous SSL literature that regularization can slightly improve supervised learning without unlabeled data [3, 8, 9]. In scenarios with additional unlabeled data, SSLayout360 provides a significant boost in accuracy performance over the supervised baselines.

D. Additional MatterportLayout Results

We report supervised and semi-supervised MatterportLayout results for rooms having 4, 6, 8, and 10 or more corners in Tables D1 – D2. These results show that the effective use of unlabeled data, in combination with labeled data, improves room layout estimation with increasing scene complexity across most settings and metrics under consideration.

E. Qualitative 3D Layout Reconstruction

Figures E1 – E3 present qualitative 3D layout reconstruction results on select MatterportLayout test instances with increasing layout complexity, ranging between 6 and 12 corners, using the post-processing algorithm proposed by Sun *et al*. of HorizonNet [6]. These results are obtained from our SSLayout360 model trained on 100 labeled and 4,000 unlabeled images. Using only 6% of the available 1,650 labels, our SSLayout360 model is able to produce convincing 3D layout reconstruction of complex indoor scenes.

4 Corners	3D IoU (%) ↑				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	63.21 ± 0.59	69.09 ± 0.53	73.62 ± 0.57	76.10 ± 0.52	81.41 ± 0.58
SSLayout360	**68.07 ± 0.23**	**73.91 ± 0.44**	**77.31 ± 0.40**	**79.37 ± 0.41**	**82.81 ± 0.72**
	2D IoU (%) ↑				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	67.22 ± 0.68	72.63 ± 0.52	76.78 ± 0.61	78.99 ± 0.48	84.02 ± 0.48
SSLayout360	**71.99 ± 0.28**	**77.31 ± 0.42**	**80.28 ± 0.40**	**82.18 ± 0.40**	**85.22 ± 0.80**
	δ_1 ↑				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	0.75 ± 0.01	0.83 ± 0.01	0.88 ± 0.01	0.91 ± 0.01	**0.95 ± 0.01**
SSLayout360	**0.80 ± 0.01**	**0.88 ± 0.01**	**0.91 ± 0.01**	**0.93 ± 0.01**	**0.96 ± 0.01**
	RMSE ↓				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	0.40 ± 0.01	0.31 ± 0.01	0.27 ± 0.02	0.25 ± 0.02	**0.18 ± 0.01**
SSLayout360	**0.33 ± 0.01**	**0.26 ± 0.01**	**0.23 ± 0.01**	**0.20 ± 0.01**	**0.17 ± 0.01**

6 Corners	3D IoU (%) ↑				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	67.67 ± 0.60	73.31 ± 0.82	77.23 ± 0.58	78.04 ± 0.34	81.93 ± 0.55
SSLayout360	**70.48 ± 0.35**	**76.38 ± 0.58**	**79.55 ± 0.52**	**80.32 ± 0.24**	**82.82 ± 0.68**
	2D IoU (%) ↑				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	71.92 ± 0.63	76.68 ± 0.90	80.05 ± 0.57	80.62 ± 0.30	84.40 ± 0.60
SSLayout360	**74.57 ± 0.37**	**79.67 ± 0.60**	**82.12 ± 0.60**	**82.86 ± 0.28**	**85.05 ± 0.81**
	δ_1 ↑				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	0.76 ± 0.01	0.87 ± 0.01	0.93 ± 0.01	0.94 ± 0.01	**0.95 ± 0.02**
SSLayout360	**0.79 ± 0.01**	**0.90 ± 0.01**	**0.95 ± 0.01**	**0.95 ± 0.01**	**0.96 ± 0.01**
	RMSE ↓				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	0.35 ± 0.01	0.28 ± 0.02	0.24 ± 0.01	0.23 ± 0.01	**0.21 ± 0.03**
SSLayout360	**0.31 ± 0.01**	**0.24 ± 0.01**	**0.21 ± 0.01**	**0.21 ± 0.01**	**0.20 ± 0.02**

Table D1. Quantitative layout results for **4 corners (left)** and **6 corners (right)** evaluated on the MatterportLayout test set averaged over four independent runs with different random seeds. Number format: mean value ± standard deviation.

8 Corners	3D IoU (%) ↑				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	63.14 ± 1.20	67.97 ± 0.39	68.30 ± 0.56	70.30 ± 0.91	**74.54 ± 0.59**
SSLayout360	**65.97 ± 0.13**	**68.39 ± 0.63**	**69.96 ± 0.42**	**72.05 ± 0.82**	74.51 ± 0.54
	2D IoU (%) ↑				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	66.29 ± 1.31	**70.87 ± 0.34**	70.87 ± 0.67	72.60 ± 0.92	**76.58 ± 0.58**
SSLayout360	**68.94 ± 0.16**	70.86 ± 0.62	**72.38 ± 0.42**	**74.34 ± 0.80**	76.31 ± 0.63
	δ_1 ↑				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	0.79 ± 0.01	0.87 ± 0.01	0.89 ± 0.01	0.90 ± 0.01	**0.94 ± 0.01**
SSLayout360	**0.83 ± 0.01**	**0.90 ± 0.01**	**0.90 ± 0.01**	**0.91 ± 0.01**	**0.94 ± 0.01**
	RMSE ↓				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	0.41 ± 0.02	**0.34 ± 0.01**	**0.33 ± 0.01**	**0.33 ± 0.02**	**0.29 ± 0.01**
SSLayout360	**0.36 ± 0.01**	**0.34 ± 0.01**	**0.32 ± 0.02**	**0.31 ± 0.01**	**0.29 ± 0.01**

10+ Corners	3D IoU (%) ↑				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	57.82 ± 1.11	60.52 ± 0.80	61.49 ± 1.23	**64.91 ± 1.99**	67.90 ± 0.69
SSLayout360	**60.53 ± 1.71**	**62.33 ± 1.14**	**64.15 ± 0.95**	**66.52 ± 1.13**	**70.24 ± 0.98**
	2D IoU (%) ↑				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	59.85 ± 1.22	62.56 ± 0.79	63.59 ± 1.15	**67.19 ± 2.13**	69.75 ± 0.61
SSLayout360	**62.49 ± 1.75**	**64.28 ± 1.17**	**66.13 ± 0.89**	**68.42 ± 1.12**	**71.98 ± 1.02**
	δ_1 ↑				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	0.80 ± 0.01	0.84 ± 0.01	0.85 ± 0.01	0.86 ± 0.01	0.88 ± 0.01
SSLayout360	**0.82 ± 0.01**	**0.87 ± 0.01**	**0.87 ± 0.01**	**0.89 ± 0.01**	**0.90 ± 0.01**
	RMSE ↓				
Method	50 labels 1,837 images	100 labels 1,837 images	200 labels 1,837 images	400 labels 1,837 images	1,650 labels 1,837 images
HorizonNet	0.56 ± 0.02	0.53 ± 0.02	**0.50 ± 0.03**	**0.47 ± 0.03**	**0.44 ± 0.03**
SSLayout360	**0.52 ± 0.03**	**0.49 ± 0.01**	**0.49 ± 0.02**	**0.46 ± 0.03**	**0.41 ± 0.02**

Table D2. Quantitative layout results for **8 corners (left)** and **10+ corners (right)** evaluated on the MatterportLayout test set averaged over four independent runs with different random seeds. Number format: mean value ± standard deviation.

Figure E1. Qualitative layout estimation of exemplar MatterportLayout test instances. **Left:** A comparison of ground truth layout (green lines) with our SSLayout360 model (magenta lines) trained on 100 labeled and 4,000 unlabeled images under equirectangular view. **Right:** 3D layout reconstruction. Best viewed electronically.

Figure E2. Qualitative layout estimation of exemplar MatterportLayout test instances. **Left:** A comparison of ground truth layout (green lines) with our SSLayout360 model (magenta lines) trained on 100 labeled and 4,000 unlabeled images under equirectangular view. **Right:** 3D layout reconstruction. The transparent regions denote walls hidden from the camera field-of-view. Best viewed electronically.

Figure E3. Qualitative layout estimation of exemplar MatterportLayout test instances. **Left:** A comparison of ground truth layout (green lines) with our SSLayout360 model (magenta lines) trained on 100 labeled and 4,000 unlabeled images under equirectangular view. **Right:** 3D layout reconstruction. The transparent regions denote walls hidden from the camera field-of-view. Best viewed electronically.

References

[1] Iro Armeni, Sasha Sax, Amir R. Zamir, and Silvio Savarese. Joint 2D-3D-Semantic Data for Indoor Scene Understanding. arXiv:1702.01105, 2017. 1

[2] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*, 2015. 2

[3] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning. *Transactions on Pattern Analysis and Machine Intelligence*, 41:1979–1993, 2017. 3

[4] Avital Oliver, Augustus Odena, Colin Raffel, Ekin D. Cubuk, and Ian J. Goodfellow. Realistic Evaluation of Semi-Supervised Learning Algorithms. In *Advances in Neural Information Processing Systems*, 2018. 1

[5] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014. 2

[6] Cheng Sun, Chi-Wei Hsiao, Min Sun, and Hwann-Tzong Chen. HorizonNet: Learning Room Layout With 1D Representation and Pano Stretch Data Augmentation. In *Computer Vision and Pattern Recognition*, pages 1047–1056, 2019. 2, 3

[7] Antti Tarvainen and Harri Valpola. Mean Teachers are Better Role Models: Weight-Averaged Consistency Targets Improve Semi-Supervised Deep Learning Results. In *Advances in Neural Information Processing Systems*, 2017. 1

[8] Phi Vu Tran. Exploring Self-Supervised Regularization for Supervised and Semi-Supervised Learning. In *NeurIPS Workshop on Learning with Rich Experience: Integration of Learning Paradigms*, 2019. 1, 3

[9] Xiaohua Zhai, Avital Oliver, Alexander Kolesnikov, and Lucas Beyer. S4L: Self-Supervised Semi-Supervised Learning. In *International Conference on Computer Vision*, 2019. 3

[10] Yinda Zhang, Shuran Song, Ping Tan, and Jianxiong Xiao. PanoContext: A Whole-Room 3D Context Model for Panoramic Scene Understanding. In *European Conference on Computer Vision*, 2014. 1

[11] Jia Zheng, Junfei Zhang, Jing Li, Rui Tang, Shenghua Gao, and Zihan Zhou. Structured3D: A Large Photo-realistic Dataset for Structured 3D Modeling. In *European Conference on Computer Vision*, 2020. 1

[12] Chuhang Zou, Alex Colburn, Qi Shan, and Derek Hoiem. LayoutNet: Reconstructing the 3D Room Layout from a Single RGB Image. In *Computer Vision and Pattern Recognition*, pages 2051–2059, 2018. 1

[13] Chuhang Zou, Jheng-Wei Su, Chi-Han Peng, Alex Colburn, Qi Shan, Peter Wonka, Hung-Kuo Chu, and Derek Hoiem. 3D Manhattan Room Layout Reconstruction from a Single 360 Image. arXiv:1910.04099, 2019. 1