Supplementary Material:
3D Scene Reconstruction with Multi-layer Depth and Epipolar Transformers

1. System Overview

We provide an overview of our 3D reconstruction system and additional qualitative examples in our **supplementary video** (see project website).

2. Training Data Generation

As we describe in Section 5.1 of our paper, we generate the target multi-layer depth maps by performing multi-hit ray tracing on the ground-truth 3D mesh models. If an object instance is completely occluded (i.e. not visible at all from the first-layer depth map), it is ignored in the subsequent layers. The Physically-based Rendering [8] dataset ignores objects in "person" and "plant" categories, so those categories are also ignored when we generate our depth maps. The complete list of room envelope categories (according to NYUv2 mapping) are as follows: wall, floor, ceiling, door, floor_mat, window, curtain, blinds, picture, mirror, fireplace, roof, and whiteboard. In our experiments, all room envelope categories are merged into a single "background" category.

In Figure 1 we provide a visualization of our multi-layer depth representation. In Figure 3, we provide a visualization of the evaluation metrics in our paper.

3. Representing the Back Surfaces of Objects

Without the back surfaces, ground truth depth layers ($\bar{D}_{1,3,5}$) cover only 82% of the scene geometry inside the viewing frustum (vs. 92% including frontal surfaces — refer to Table 1 in our paper for full comparison). Figure 2(a) visualizes $\bar{D}_{1,3,5}$, without the back surfaces. This representation, *layered depth image* (LDI) [4], was originally developed in the computer graphics community [4] as an algorithm for rendering textured depth images using parallax transformation. Works based on prediction of LDI or its variants [7, 9] therefore do not represent the invisible back surfaces of objects. Prediction of back surfaces enables volumetric inference in our epipolar transformation.

4. Multi-layer Depth Prediction

See Figure 8 for network parameters of our multi-layer depth prediction model. All batch normalization layers have momentum 0.005, and all activation layers are Leaky ReLUs layers with $\alpha = 0.01$. We use In-place Activated BatchNorm [3] for all of our batch normalization layers. We trained the network for 40 epochs. The meta parameters (learning rates, momentum, batch size, epochs, etc) are the same for all the networks in our system.

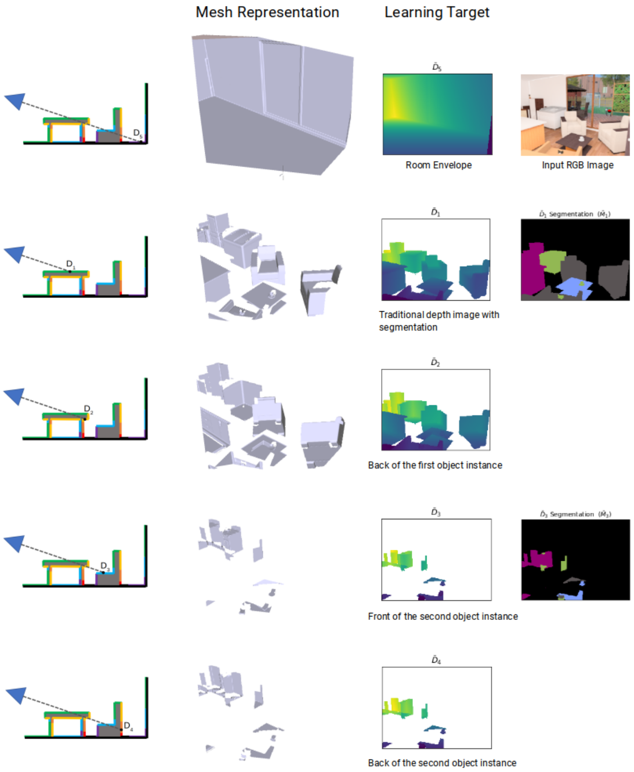

Mesh Representation
Learning Target
D_5
$\bar{D}_5$
Room Envelope
Input RGB Image
D_1
$\bar{D}_1$
$\bar{D}_1$ Segmentation ($\bar{M}_1$)
Traditional depth image with segmentation
D_2
$\bar{D}_2$
Back of the first object instance
D_3
$\bar{D}_3$
$\bar{D}_3$ Segmentation ($\bar{M}_3$)
Front of the second object instance
D_4
$\bar{D}_4$
Back of the second object instance

Figure 1: Layer-wise illustration of our multi-layer depth representation in 3D. Table 1 in our paper reports an empirical analysis which shows that the five-layer representation ($\bar{D}_{1,2,3,4,5}$) covers 92% of the scene geometry inside the viewing frustum.

5. Multi-layer Semantic Segmentation

See Figure 9 for network parameters of multi-layer semantic segmentation. We construct a binary mask for all

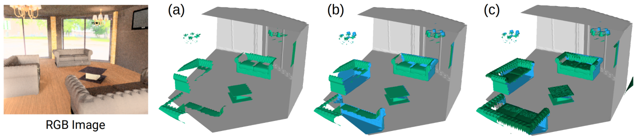


Figure 2: Illustration of ground-truth depth layers. (a, b): 2.5D depth representation cannot accurately encode the geometry of surfaces which are nearly tangent to the viewing direction. (b): We model both the front and back surfaces of objects as seen from the input camera. (c): The tangent surfaces are sampled more densely in the additional virtual view (dark green). Table 3 in our paper shows the effect of augmenting the frontal predictions with the virtual view predictions.

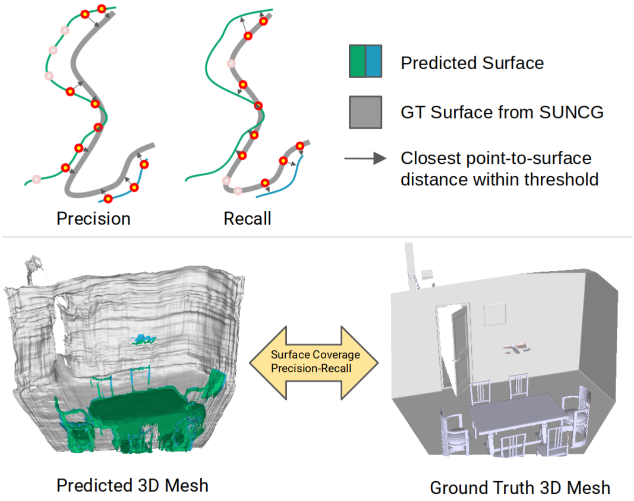


Figure 3: Illustration of our 3D precision-recall metrics. *Top*: We perform a bidirectional surface coverage evaluation on the reconstructed triangle meshes. *Bottom*: The ground truth mesh consists of all 3D surfaces within the viewing frustum and in front of the room envelope. We take the union of the predicted meshes from different views in world coordinates. This allows us to perform a layer-wise evaluation (e.g. Figure 8 in our paper).

foreground objects, and define segmentation mask M_l as all non-background pixels at layer l. As mentioned in section 3.1, D_1 and D_2 the same segmentation due to symmetry, so we only segment layers 1 and 3. The purpose of the foreground object labels is to be used as a supervisory signal for feature extraction F_{seg}, which is used as input to our Epipolar Feature Transformer Networks.

6. Virtual-view Prediction

See Figure 11 and 12 for network parameters of our virtual-view height map prediction and segmentation models. The height map prediction network is trained to minimize foreground pixel losses. At test time, the background mask predicted by the segmentation network is used to zero out the floor pixels. The floor height is assumed to be zero in world coordinates. An alternate approach is minimizing both foreground and background losses and thus allowing the height map predictor to implicitly segment the floor by predicting zeros. We experimented with both architectures and found the explicit segmentation approach to perform better.

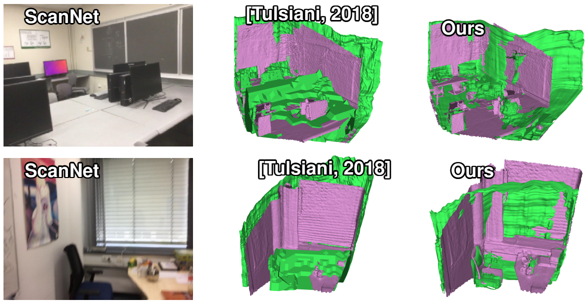


Figure 4: Evaluation of 3D reconstruction on the ScanNet [1] dataset, where green regions are detected objects and pink regions are ground truth.

7. Voxelization of Multi-layer Depth Prediction

Given a 10m^3 voxel grid of resolution 400 (equivalently, 2.5cm^3) with a bounding box ranging from (-5,-5,-10) to (5,5,0) in camera space, we project the center of each voxel into the predicted depth maps. If the depth value for the projected voxel falls in the first object interval (D_1, D_2) or the occluded object interval (D_3, D_4), the voxel is marked occupied. We evaluate our voxelization against the SUNCG ground truth object meshes inside the viewing frustum, voxelized using the Binvox software which implements z-buffer based carving and parity voting methods. We also voxelize the predicted Factored3D [6] objects (same meshes evaluated in Figure 8 of our paper) using Binvox under the same setting as the ground truth. We randomly select 1800 examples from the test set and compute the intersection-over-union of all objects in the scene. In addition to Figure 4 of our paper, Figure 7 shows a visualization of our voxels, colored according to the predicted semantic labeling.

8. Predictions on NYU and SUNCG

Figures 5 and 6 show additional 3D scene reconstruction results. We provide more visualizations of our network outputs and error maps on the SUNCG dataset in the last few pages of the supplementary material.

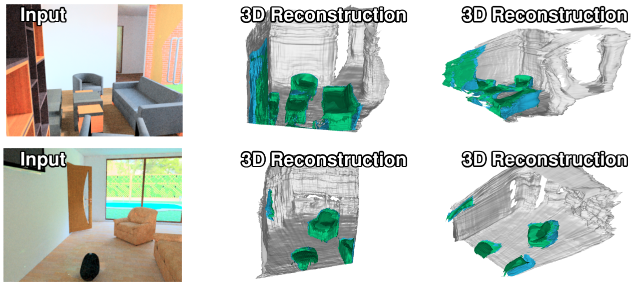


Figure 5: Evaluation of single image scene reconstruction on SUNCG [5].

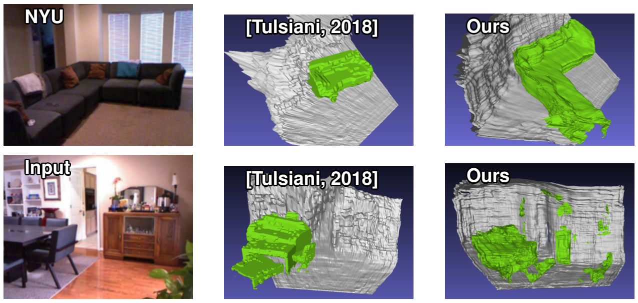


Figure 6: Evaluation of single image scene reconstruction on NYUv2 [2].

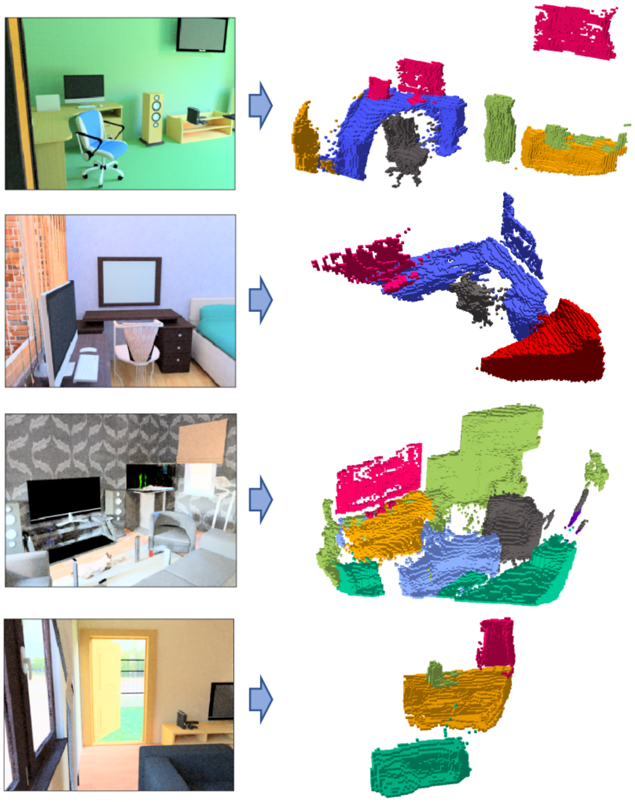


Figure 7: Volumetric evaluation of our predicted multi-layer depth maps on SUNCG.

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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2

[2] Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2012. 3

[3] Samuel Rota Bulò, Lorenzo Porzi, and Peter Kontschieder. In-place activated batchnorm for memory-optimized training of dnns. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2018. 1

[4] Jonathan Shade, Steven Gortler, Li-wei He, and Richard Szeliski. Layered depth images. 1998. 1

[5] Shuran Song, Fisher Yu, Andy Zeng, Angel X Chang, Manolis Savva, and Thomas Funkhouser. Semantic scene completion from a single depth image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 3

[6] Shubham Tulsiani, Saurabh Gupta, David Fouhey, Alexei A Efros, and Jitendra Malik. Factoring shape, pose, and layout from the 2d image of a 3d scene. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2

[7] Shubham Tulsiani, Richard Tucker, and Noah Snavely. Layer-structured 3d scene inference via view synthesis. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. 1

[8] Yinda Zhang, Shuran Song, Ersin Yumer, Manolis Savva, Joon-Young Lee, Hailin Jin, and Thomas Funkhouser. Physically-based rendering for indoor scene understanding using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 1

[9] Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, and Noah Snavely. Stereo magnification: Learning view synthesis using multiplane images. In *ACM Transactions on Graphics (SIGGRAPH)*, 2018. 1

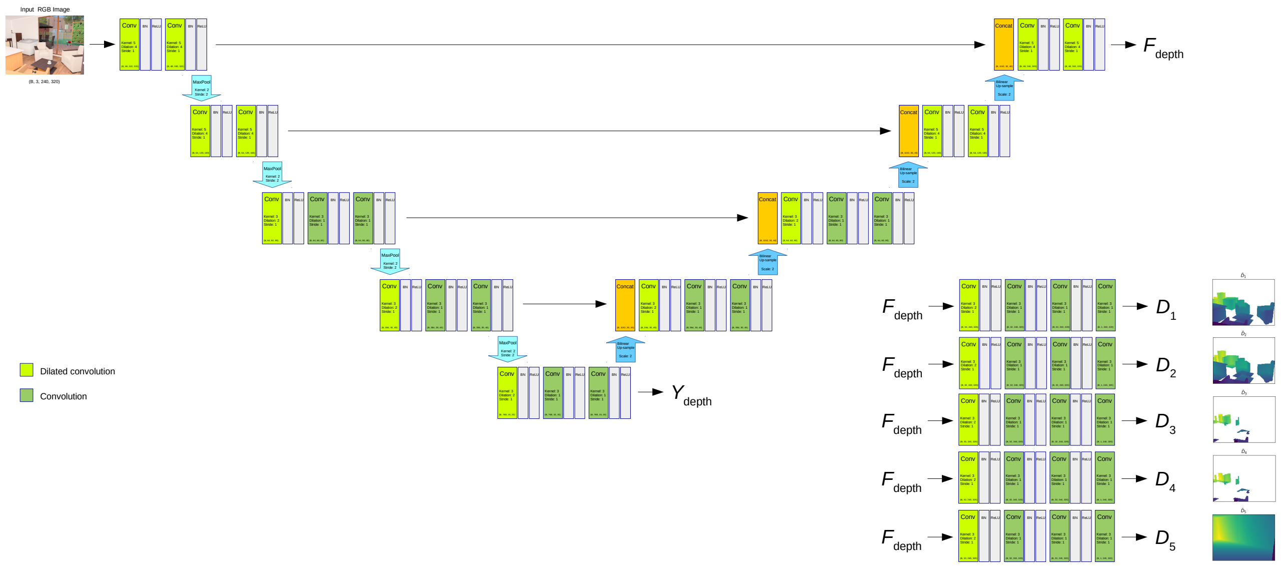


Figure 8: Network architecture for multi-layer depth prediction. The horizontal arrows in the network represent skip connections. This figure, along with following figures, is best viewed in color and on screen.

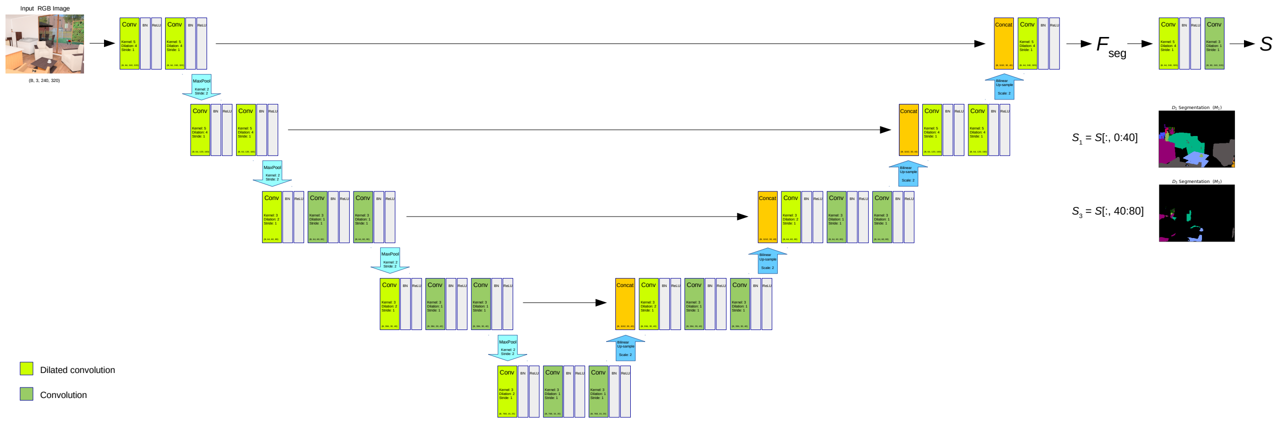


Figure 9: Network architecture for multi-layer semantic segmentation network. (Best viewed in color and on screen)

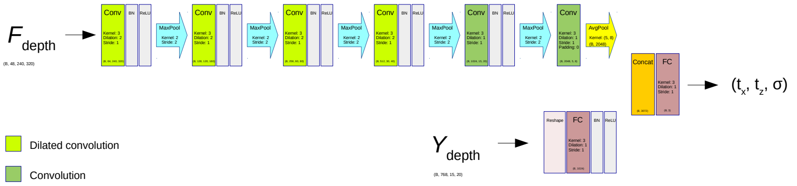


Figure 10: Network architecture for virtual camera pose proposal network. (Best viewed in color and on screen)

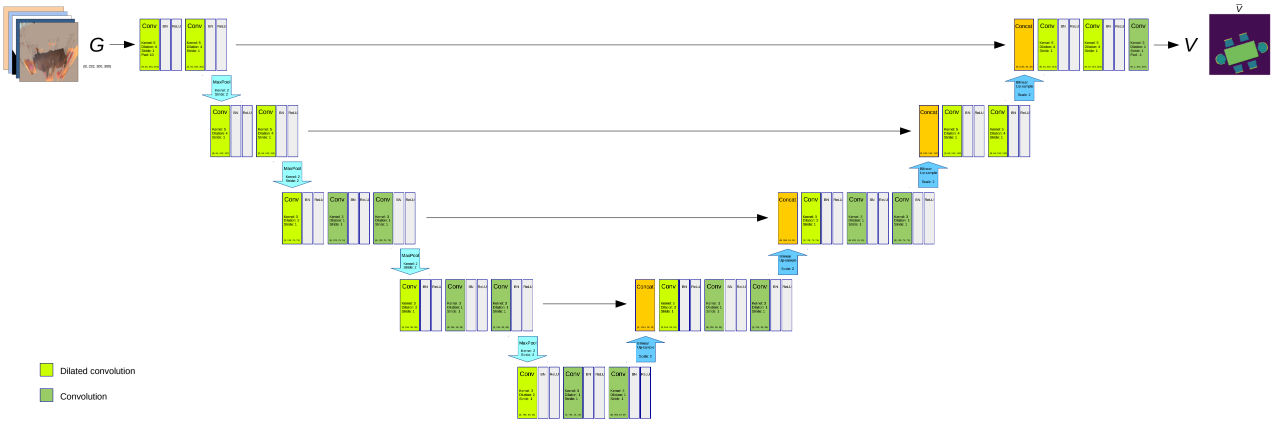


Figure 11: Network architecture for virtual view surface prediction network. (Best viewed in color and on screen)

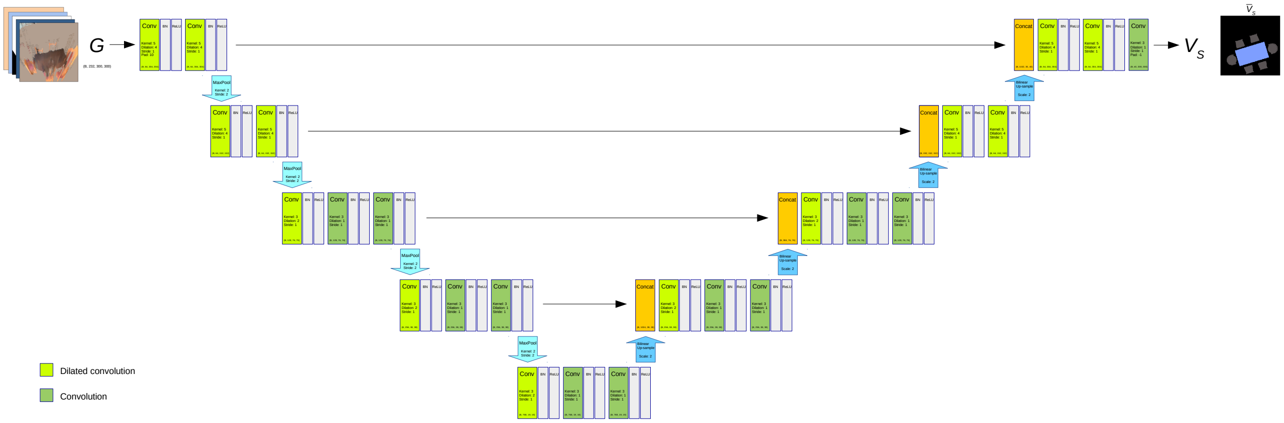
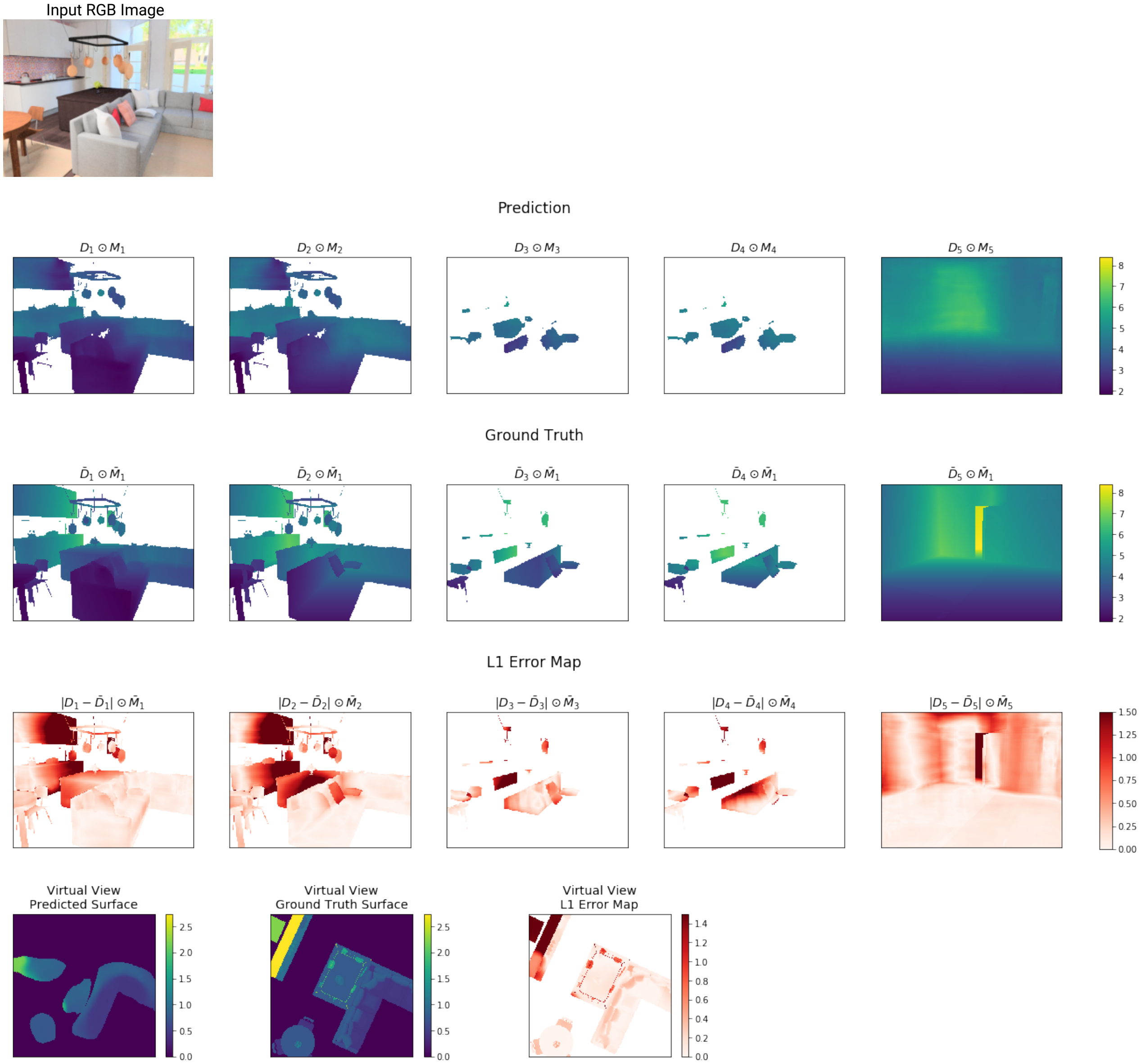
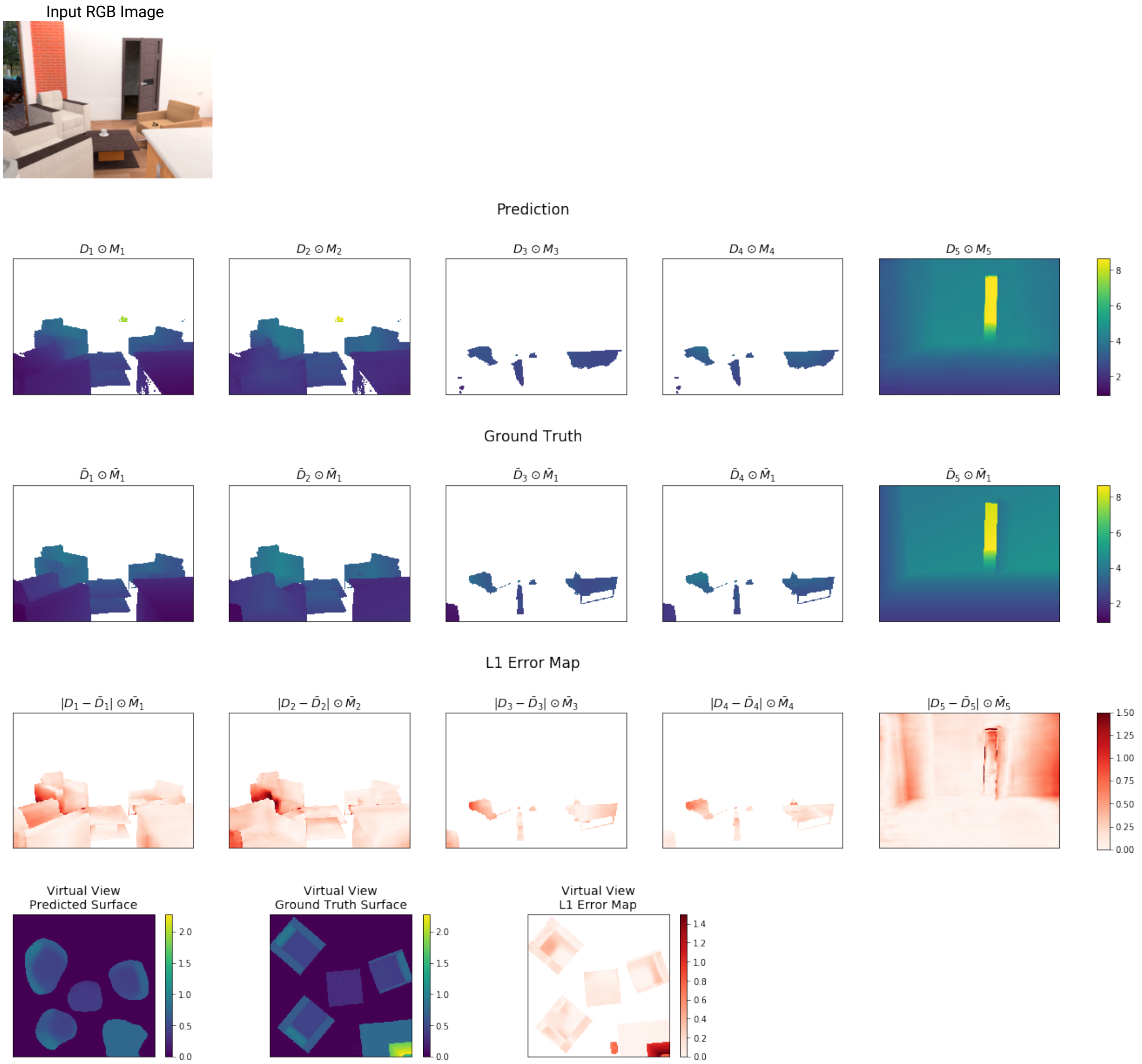
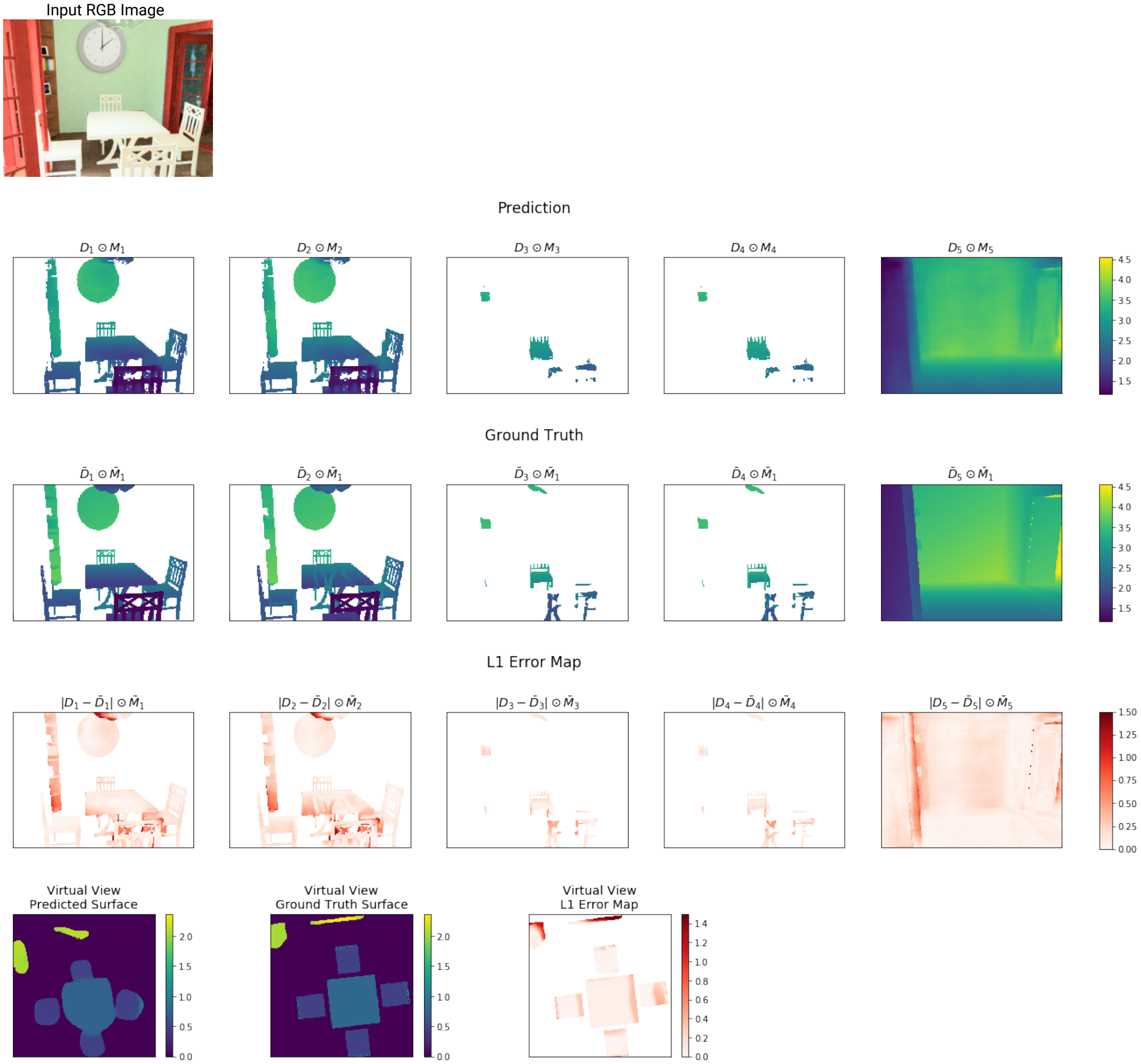


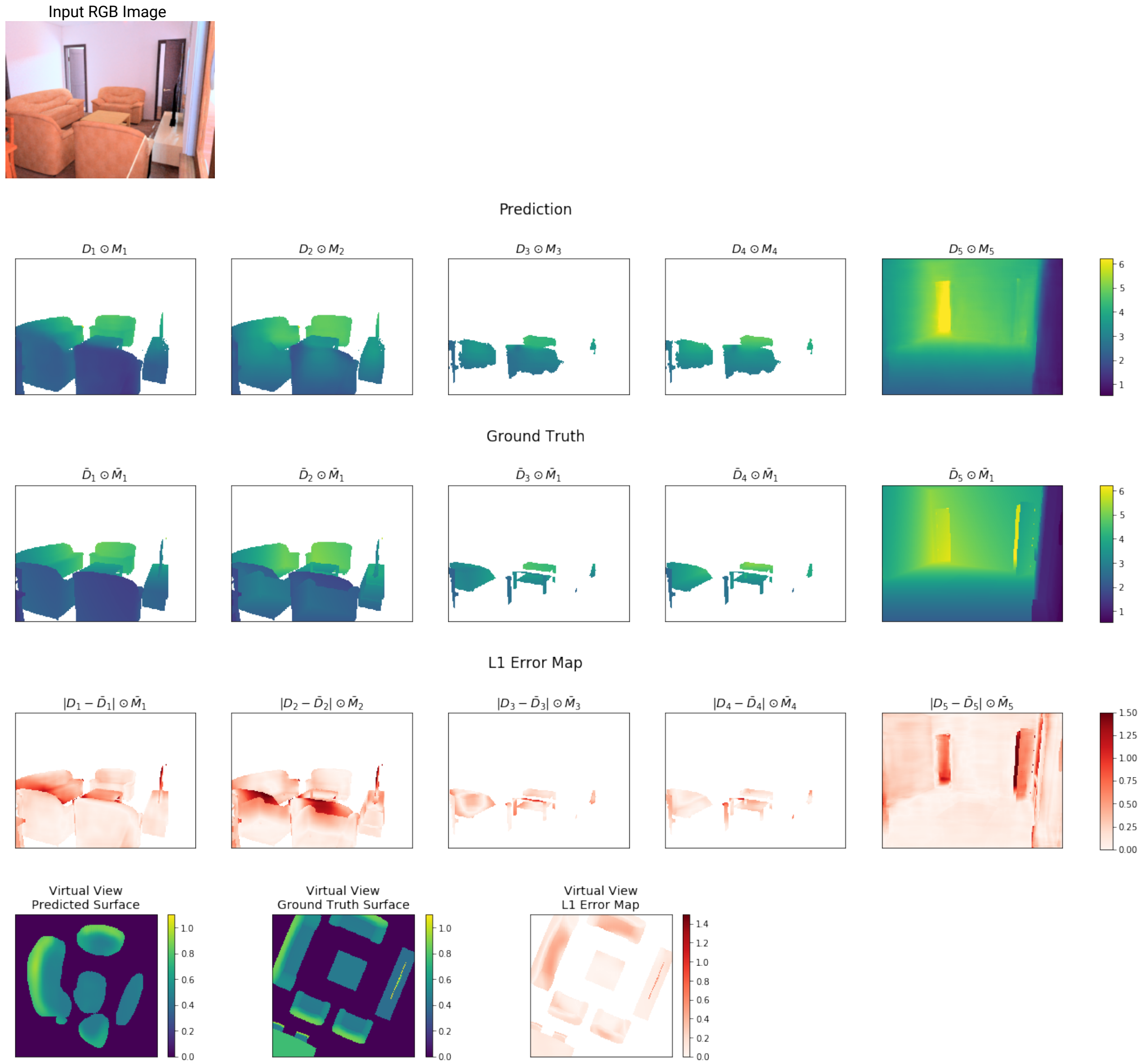
Figure 12: Network architecture for virtual view segmentation network. (Best viewed in color and on screen)

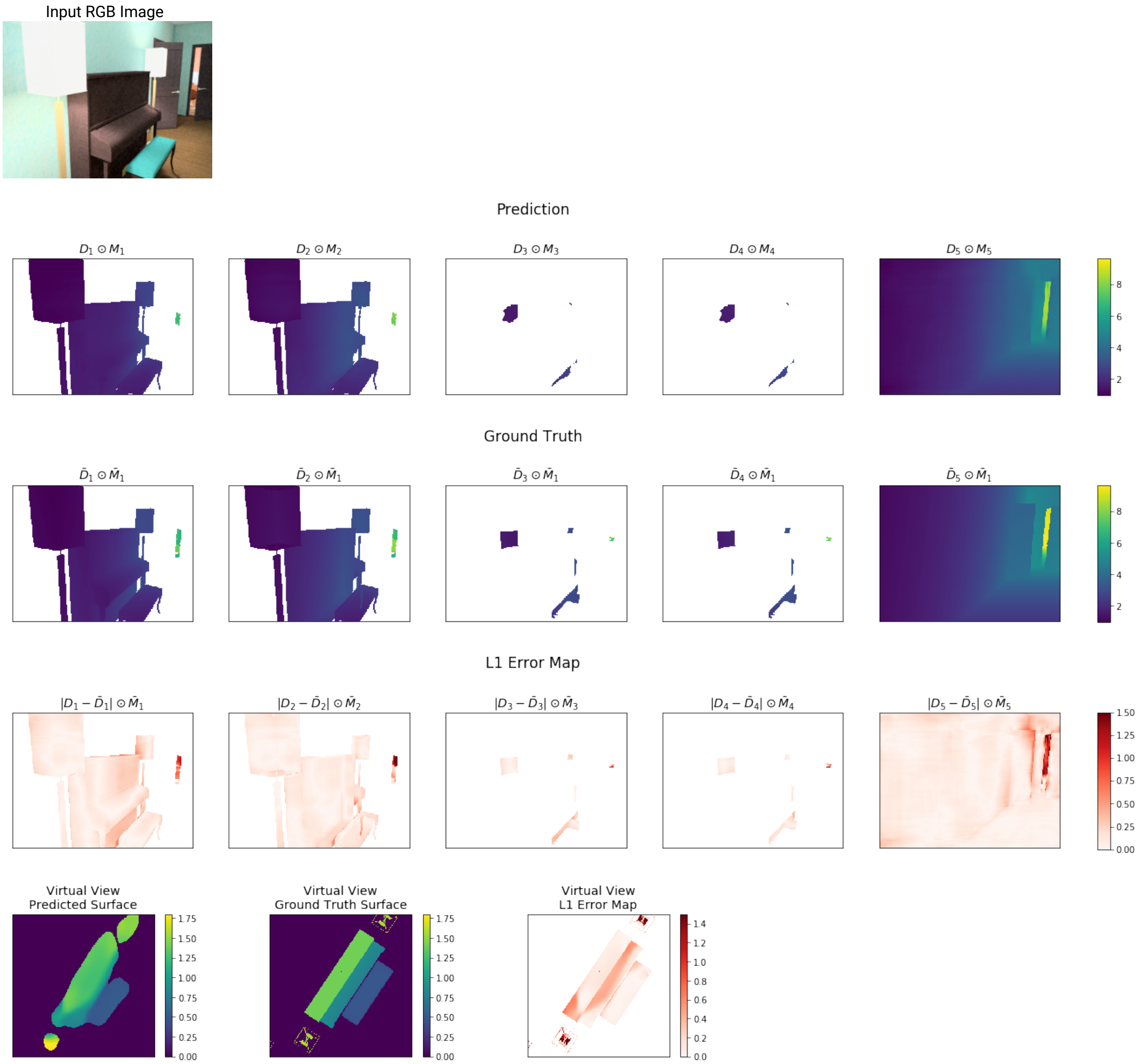

Input RGB Image
Prediction
$D_1 \odot M_1$
$D_2 \odot M_2$
$D_3 \odot M_3$
$D_4 \odot M_4$
$D_5 \odot M_5$
8
7
6
5
4
3
2
Ground Truth
$\bar{D}_1 \odot \bar{M}_1$
$\bar{D}_2 \odot \bar{M}_1$
$\bar{D}_3 \odot \bar{M}_1$
$\bar{D}_4 \odot \bar{M}_1$
$\bar{D}_5 \odot \bar{M}_1$
8
7
6
5
4
3
2
L1 Error Map
$|D_1 - \bar{D}_1| \odot \bar{M}_1$
$|D_2 - \bar{D}_2| \odot \bar{M}_2$
$|D_3 - \bar{D}_3| \odot \bar{M}_3$
$|D_4 - \bar{D}_4| \odot \bar{M}_4$
$|D_5 - \bar{D}_5| \odot \bar{M}_5$
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Virtual View
Predicted Surface
2.5
2.0
1.5
1.0
0.5
0.0
Virtual View
Ground Truth Surface
2.5
2.0
1.5
1.0
0.5
0.0
Virtual View
L1 Error Map
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0

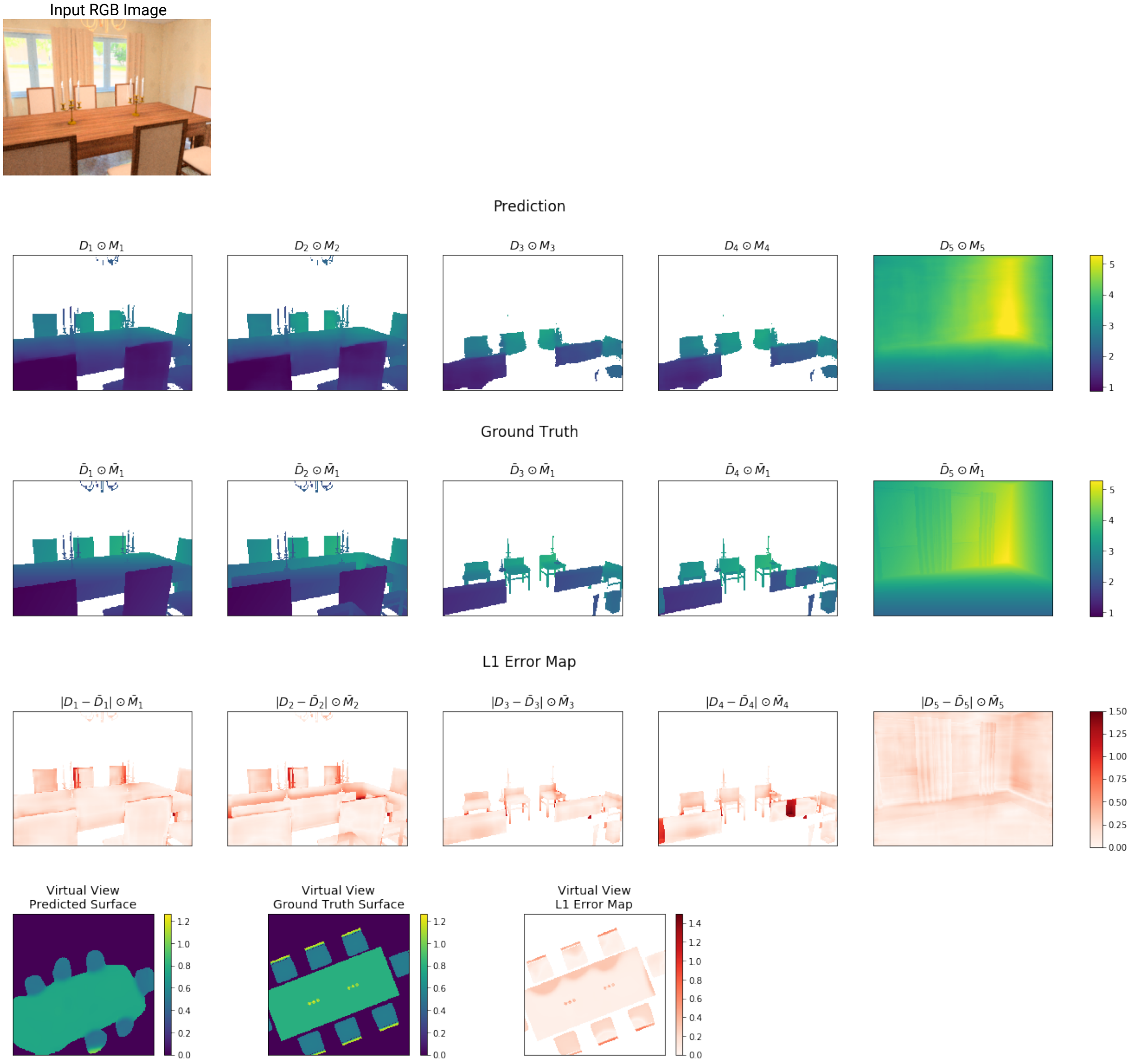


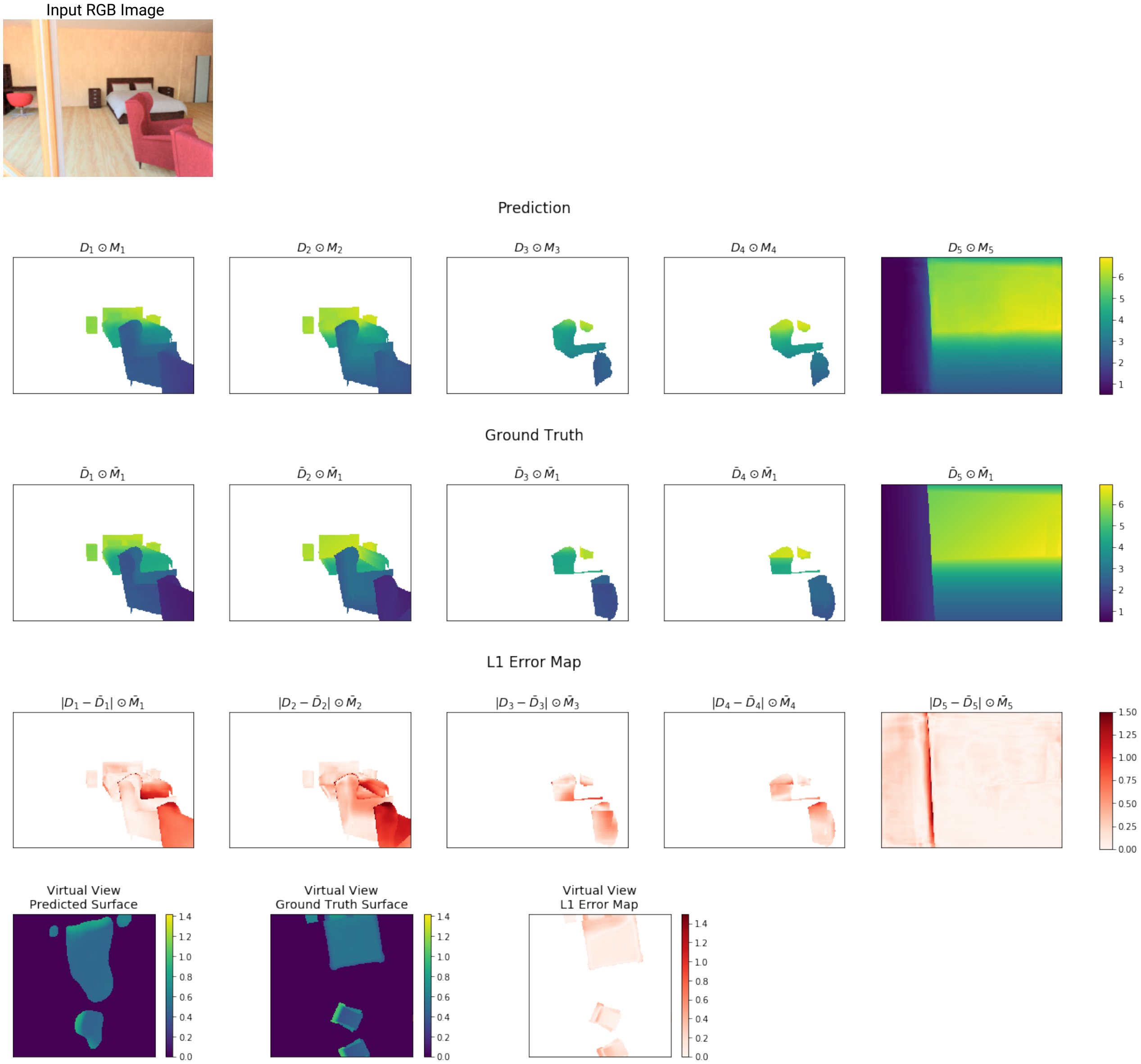
Input RGB Image
Prediction
$D_1 \odot M_1$
$D_2 \odot M_2$
$D_3 \odot M_3$
$D_4 \odot M_4$
$D_5 \odot M_5$
2
4
6
8
Ground Truth
$\bar{D}_1 \odot \bar{M}_1$
$\bar{D}_2 \odot \bar{M}_1$
$\bar{D}_3 \odot \bar{M}_1$
$\bar{D}_4 \odot \bar{M}_1$
$\bar{D}_5 \odot \bar{M}_1$
2
4
6
8
L1 Error Map
$|D_1 - \bar{D}_1| \odot \bar{M}_1$
$|D_2 - \bar{D}_2| \odot \bar{M}_2$
$|D_3 - \bar{D}_3| \odot \bar{M}_3$
$|D_4 - \bar{D}_4| \odot \bar{M}_4$
$|D_5 - \bar{D}_5| \odot \bar{M}_5$
0.00
0.25
0.50
0.75
1.00
1.25
1.50
Virtual View
Predicted Surface
0.0
0.5
1.0
1.5
2.0
Virtual View
Ground Truth Surface
0.0
0.5
1.0
1.5
2.0
Virtual View
L1 Error Map
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

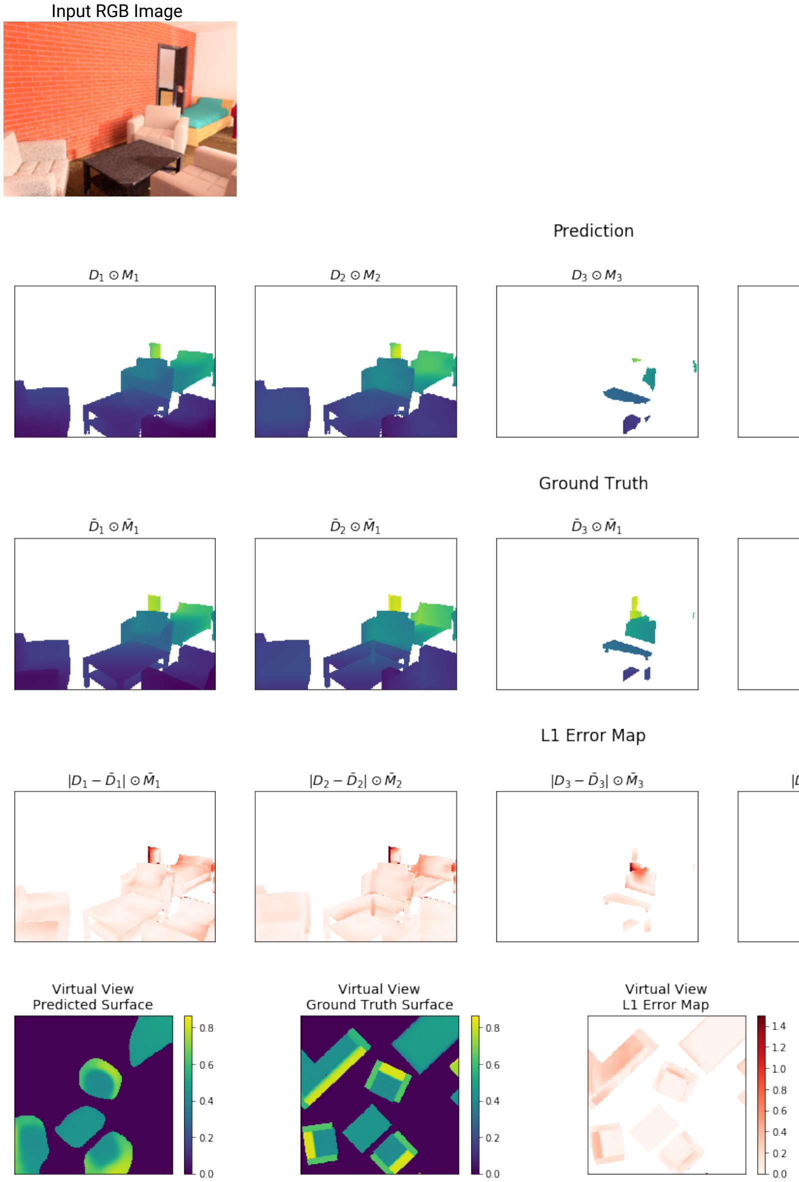

Input RGB Image
Prediction
$D_1 \odot M_1$
$D_2 \odot M_2$
$D_3 \odot M_3$
$D_4 \odot M_4$
$D_5 \odot M_5$
4.5
4.0
3.5
3.0
2.5
2.0
1.5
Ground Truth
$\bar{D}_1 \odot \bar{M}_1$
$\bar{D}_2 \odot \bar{M}_1$
$\bar{D}_3 \odot \bar{M}_1$
$\bar{D}_4 \odot \bar{M}_1$
$\bar{D}_5 \odot \bar{M}_1$
4.5
4.0
3.5
3.0
2.5
2.0
1.5
L1 Error Map
$|D_1 - \bar{D}_1| \odot \bar{M}_1$
$|D_2 - \bar{D}_2| \odot \bar{M}_2$
$|D_3 - \bar{D}_3| \odot \bar{M}_3$
$|D_4 - \bar{D}_4| \odot \bar{M}_4$
$|D_5 - \bar{D}_5| \odot \bar{M}_5$
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Virtual View
Predicted Surface
2.0
1.5
1.0
0.5
0.0
Virtual View
Ground Truth Surface
2.0
1.5
1.0
0.5
0.0
Virtual View
L1 Error Map
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0

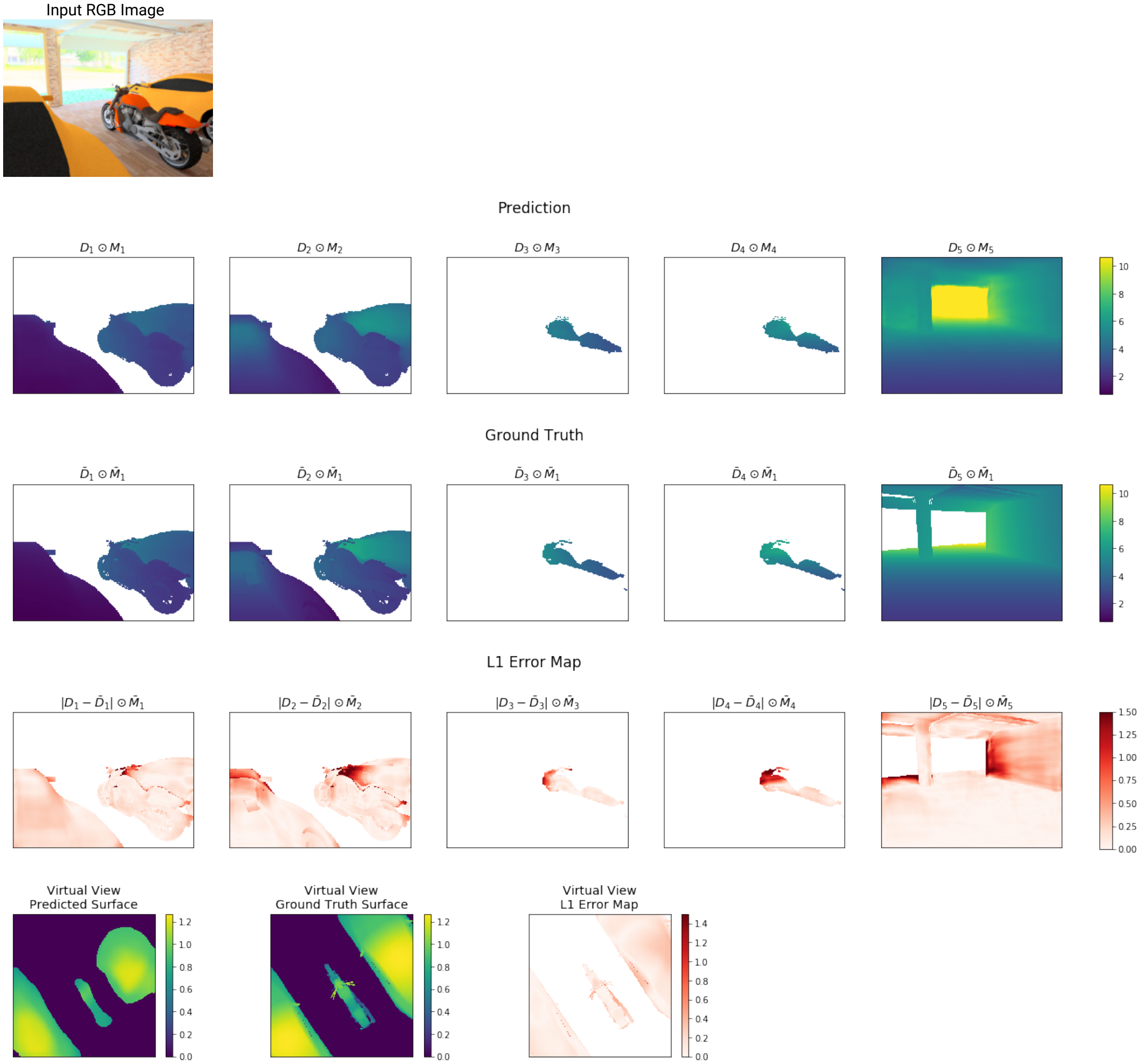

Input RGB Image
Prediction
$D_1 \odot M_1$
$D_2 \odot M_2$
$D_3 \odot M_3$
$D_4 \odot M_4$
$D_5 \odot M_5$
1 2 3 4 5 6
Ground Truth
$\bar{D}_1 \odot \bar{M}_1$
$\bar{D}_2 \odot \bar{M}_1$
$\bar{D}_3 \odot \bar{M}_1$
$\bar{D}_4 \odot \bar{M}_1$
$\bar{D}_5 \odot \bar{M}_1$
1 2 3 4 5 6
L1 Error Map
$|D_1 - \bar{D}_1| \odot \bar{M}_1$
$|D_2 - \bar{D}_2| \odot \bar{M}_2$
$|D_3 - \bar{D}_3| \odot \bar{M}_3$
$|D_4 - \bar{D}_4| \odot \bar{M}_4$
$|D_5 - \bar{D}_5| \odot \bar{M}_5$
0.00 0.25 0.50 0.75 1.00 1.25 1.50
Virtual View
Predicted Surface
0.0 0.2 0.4 0.6 0.8 1.0
Virtual View
Ground Truth Surface
0.0 0.2 0.4 0.6 0.8 1.0
Virtual View
L1 Error Map
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4

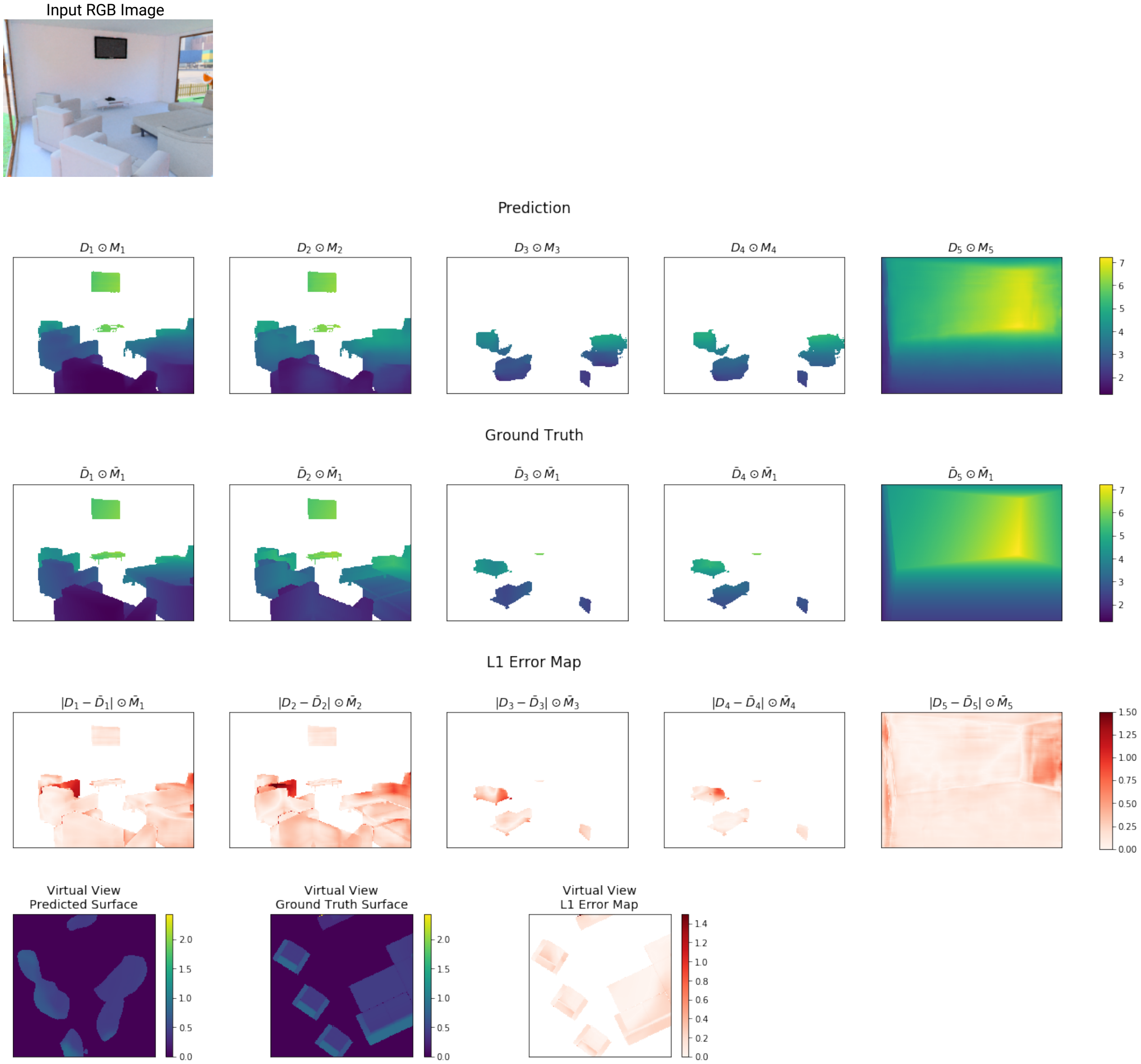

Input RGB Image
Prediction
$D_1 \odot M_1$
$D_2 \odot M_2$
$D_3 \odot M_3$
$D_4 \odot M_4$
$D_5 \odot M_5$
8
6
4
2
Ground Truth
$\bar{D}_1 \odot \bar{M}_1$
$\bar{D}_2 \odot \bar{M}_1$
$\bar{D}_3 \odot \bar{M}_1$
$\bar{D}_4 \odot \bar{M}_1$
$\bar{D}_5 \odot \bar{M}_1$
8
6
4
2
L1 Error Map
$|D_1 - \bar{D}_1| \odot \bar{M}_1$
$|D_2 - \bar{D}_2| \odot \bar{M}_2$
$|D_3 - \bar{D}_3| \odot \bar{M}_3$
$|D_4 - \bar{D}_4| \odot \bar{M}_4$
$|D_5 - \bar{D}_5| \odot \bar{M}_5$
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Virtual View
Predicted Surface
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Virtual View
Ground Truth Surface
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Virtual View
L1 Error Map
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0

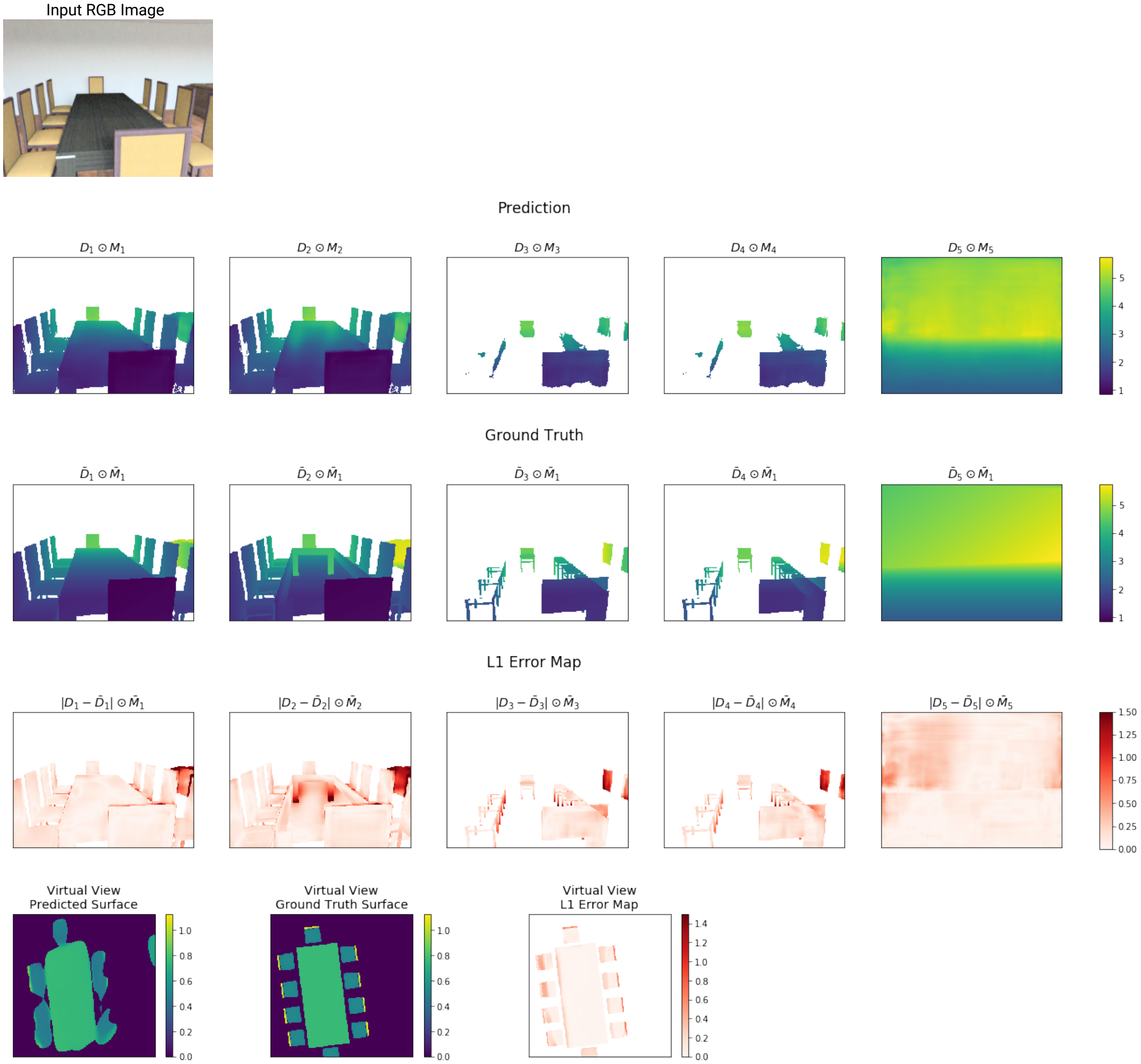

Input RGB Image
Prediction
$D_1 \odot M_1$
$D_2 \odot M_2$
$D_3 \odot M_3$
$D_4 \odot M_4$
$D_5 \odot M_5$
5
4
3
2
1
Ground Truth
$\bar{D}_1 \odot \bar{M}_1$
$\bar{D}_2 \odot \bar{M}_1$
$\bar{D}_3 \odot \bar{M}_1$
$\bar{D}_4 \odot \bar{M}_1$
$\bar{D}_5 \odot \bar{M}_1$
5
4
3
2
1
L1 Error Map
$|D_1 - \bar{D}_1| \odot \bar{M}_1$
$|D_2 - \bar{D}_2| \odot \bar{M}_2$
$|D_3 - \bar{D}_3| \odot \bar{M}_3$
$|D_4 - \bar{D}_4| \odot \bar{M}_4$
$|D_5 - \bar{D}_5| \odot \bar{M}_5$
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Virtual View
Predicted Surface
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Virtual View
Ground Truth Surface
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Virtual View
L1 Error Map
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0

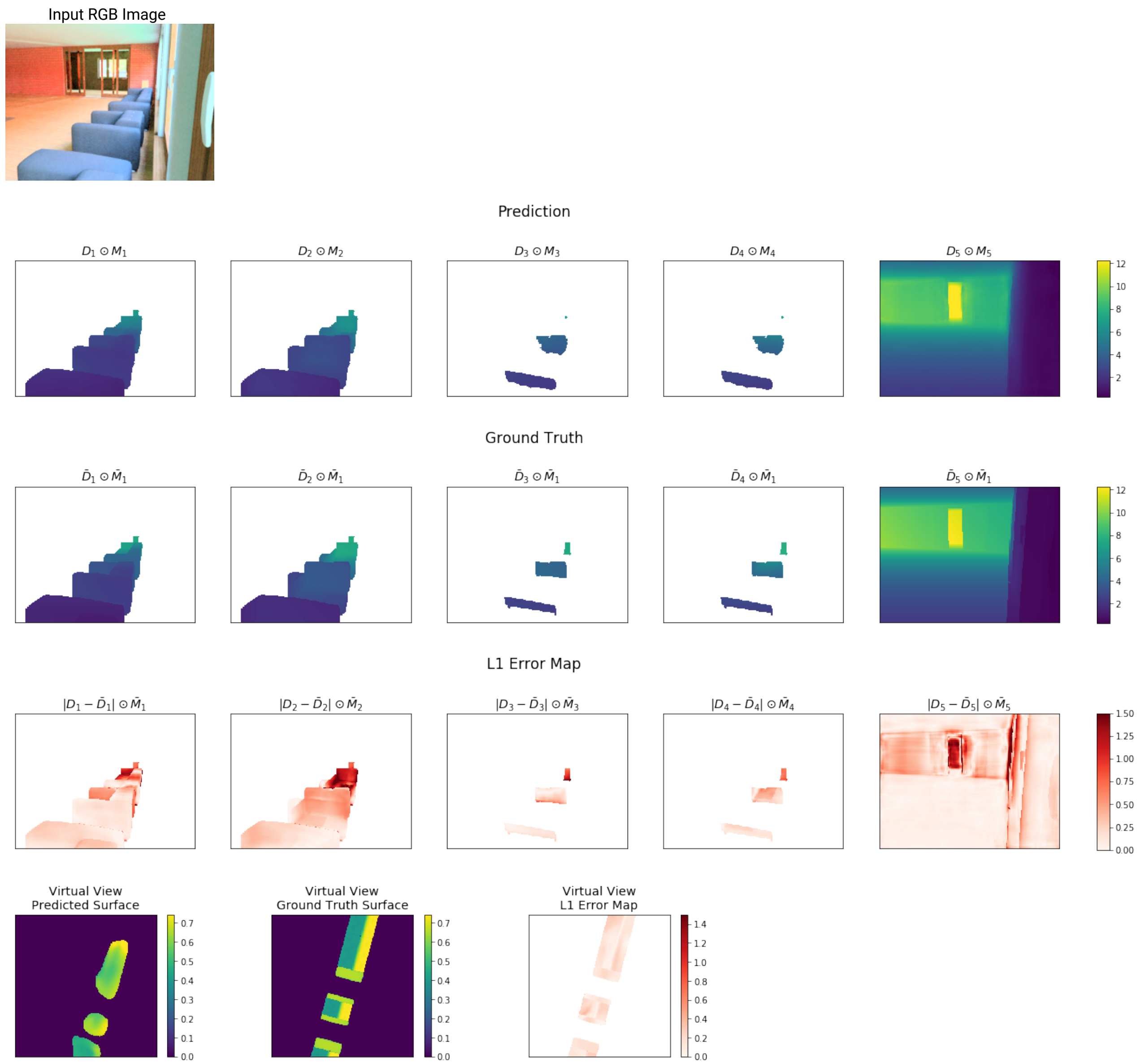

Input RGB Image
Prediction
$D_1 \odot M_1$
$D_2 \odot M_2$
$D_3 \odot M_3$
$D_4 \odot M_4$
$D_5 \odot M_5$
1
2
3
4
5
6
Ground Truth
$\bar{D}_1 \odot \bar{M}_1$
$\bar{D}_2 \odot \bar{M}_1$
$\bar{D}_3 \odot \bar{M}_1$
$\bar{D}_4 \odot \bar{M}_1$
$\bar{D}_5 \odot \bar{M}_1$
1
2
3
4
5
6
L1 Error Map
$|D_1 - \bar{D}_1| \odot \bar{M}_1$
$|D_2 - \bar{D}_2| \odot \bar{M}_2$
$|D_3 - \bar{D}_3| \odot \bar{M}_3$
$|D_4 - \bar{D}_4| \odot \bar{M}_4$
$|D_5 - \bar{D}_5| \odot \bar{M}_5$
0.00
0.25
0.50
0.75
1.00
1.25
1.50
Virtual View
Predicted Surface
Virtual View
Ground Truth Surface
Virtual View
L1 Error Map
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4


Input RGB Image
Prediction
$D_1 \odot M_1$
$D_2 \odot M_2$
$D_3 \odot M_3$
$D_4 \odot M_4$
$D_5 \odot M_5$
7
6
5
4
3
2
Ground Truth
$\bar{D}_1 \odot \bar{M}_1$
$\bar{D}_2 \odot \bar{M}_1$
$\bar{D}_3 \odot \bar{M}_1$
$\bar{D}_4 \odot \bar{M}_1$
$\bar{D}_5 \odot \bar{M}_1$
7
6
5
4
3
2
L1 Error Map
$|D_1 - \bar{D}_1| \odot \bar{M}_1$
$|D_2 - \bar{D}_2| \odot \bar{M}_2$
$|D_3 - \bar{D}_3| \odot \bar{M}_3$
$|D_4 - \bar{D}_4| \odot \bar{M}_4$
$|D_5 - \bar{D}_5| \odot \bar{M}_5$
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Virtual View
Predicted Surface
0.8
0.6
0.4
0.2
0.0
Virtual View
Ground Truth Surface
0.8
0.6
0.4
0.2
0.0
Virtual View
L1 Error Map
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0


Input RGB Image
Prediction
$D_1 \odot M_1$
$D_2 \odot M_2$
$D_3 \odot M_3$
$D_4 \odot M_4$
$D_5 \odot M_5$
10
8
6
4
2
Ground Truth
$\bar{D}_1 \odot \bar{M}_1$
$\bar{D}_2 \odot \bar{M}_1$
$\bar{D}_3 \odot \bar{M}_1$
$\bar{D}_4 \odot \bar{M}_1$
$\bar{D}_5 \odot \bar{M}_1$
10
8
6
4
2
L1 Error Map
$|D_1 - \bar{D}_1| \odot \bar{M}_1$
$|D_2 - \bar{D}_2| \odot \bar{M}_2$
$|D_3 - \bar{D}_3| \odot \bar{M}_3$
$|D_4 - \bar{D}_4| \odot \bar{M}_4$
$|D_5 - \bar{D}_5| \odot \bar{M}_5$
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Virtual View
Predicted Surface
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Virtual View
Ground Truth Surface
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Virtual View
L1 Error Map
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0


Input RGB Image
Prediction
$D_1 \odot M_1$
$D_2 \odot M_2$
$D_3 \odot M_3$
$D_4 \odot M_4$
$D_5 \odot M_5$
7
6
5
4
3
2
Ground Truth
$\bar{D}_1 \odot \bar{M}_1$
$\bar{D}_2 \odot \bar{M}_1$
$\bar{D}_3 \odot \bar{M}_1$
$\bar{D}_4 \odot \bar{M}_1$
$\bar{D}_5 \odot \bar{M}_1$
7
6
5
4
3
2
L1 Error Map
$|D_1 - \bar{D}_1| \odot \bar{M}_1$
$|D_2 - \bar{D}_2| \odot \bar{M}_2$
$|D_3 - \bar{D}_3| \odot \bar{M}_3$
$|D_4 - \bar{D}_4| \odot \bar{M}_4$
$|D_5 - \bar{D}_5| \odot \bar{M}_5$
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Virtual View
Predicted Surface
2.0
1.5
1.0
0.5
0.0
Virtual View
Ground Truth Surface
2.0
1.5
1.0
0.5
0.0
Virtual View
L1 Error Map
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0


Input RGB Image
Prediction
$D_1 \odot M_1$
$D_2 \odot M_2$
$D_3 \odot M_3$
$D_4 \odot M_4$
$D_5 \odot M_5$
1
2
3
4
5
Ground Truth
$\bar{D}_1 \odot \bar{M}_1$
$\bar{D}_2 \odot \bar{M}_1$
$\bar{D}_3 \odot \bar{M}_1$
$\bar{D}_4 \odot \bar{M}_1$
$\bar{D}_5 \odot \bar{M}_1$
1
2
3
4
5
L1 Error Map
$|D_1 - \bar{D}_1| \odot \bar{M}_1$
$|D_2 - \bar{D}_2| \odot \bar{M}_2$
$|D_3 - \bar{D}_3| \odot \bar{M}_3$
$|D_4 - \bar{D}_4| \odot \bar{M}_4$
$|D_5 - \bar{D}_5| \odot \bar{M}_5$
0.00
0.25
0.50
0.75
1.00
1.25
1.50
Virtual View
Predicted Surface
0.0
0.2
0.4
0.6
0.8
1.0
Virtual View
Ground Truth Surface
0.0
0.2
0.4
0.6
0.8
1.0
Virtual View
L1 Error Map
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4


Input RGB Image
Prediction
$D_1 \odot M_1$
$D_2 \odot M_2$
$D_3 \odot M_3$
$D_4 \odot M_4$
$D_5 \odot M_5$
12
10
8
6
4
2
Ground Truth
$\bar{D}_1 \odot \bar{M}_1$
$\bar{D}_2 \odot \bar{M}_1$
$\bar{D}_3 \odot \bar{M}_1$
$\bar{D}_4 \odot \bar{M}_1$
$\bar{D}_5 \odot \bar{M}_1$
12
10
8
6
4
2
L1 Error Map
$|D_1 - \bar{D}_1| \odot \bar{M}_1$
$|D_2 - \bar{D}_2| \odot \bar{M}_2$
$|D_3 - \bar{D}_3| \odot \bar{M}_3$
$|D_4 - \bar{D}_4| \odot \bar{M}_4$
$|D_5 - \bar{D}_5| \odot \bar{M}_5$
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Virtual View
Predicted Surface
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Virtual View
Ground Truth Surface
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Virtual View
L1 Error Map
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0