Semantic Scene Completion from a Single Depth Image
SUPPLEMENTAL MATERIAL

Shuran Song Fisher Yu Andy Zeng Angel X. Chang Manolis Savva Thomas Funkhouser
Princeton University
http://sscnet.cs.princeton.edu

1. SUNCG Dataset Statistics

In this section, we present several statistics related to our SUNCG dataset. We start by providing the basic statistics of scene structure and physical size for 3D scenes in our dataset, and then move on to talk about higher-level statistics regarding object categories, room types, and object-room relationships.

Scene Structure Statistics Figure 1 illustrates the distribution of number of rooms and number of floors per scene in the SUNCG dataset. The 3D scenes in our dataset are range from single room studio to multi-floor houses. The average and median number of rooms per-house are 8.9 and 7 respectively. The average and median number of floors per-house are 1.3 and 1 respectively.

Physical Size Statistics All object meshes and 3D scenes in the SUNCG dataset are measured in real-world spatial dimensions (units are in meters). Figure 2 shows statistics related to physical size over three levels: rooms, floors and houses.

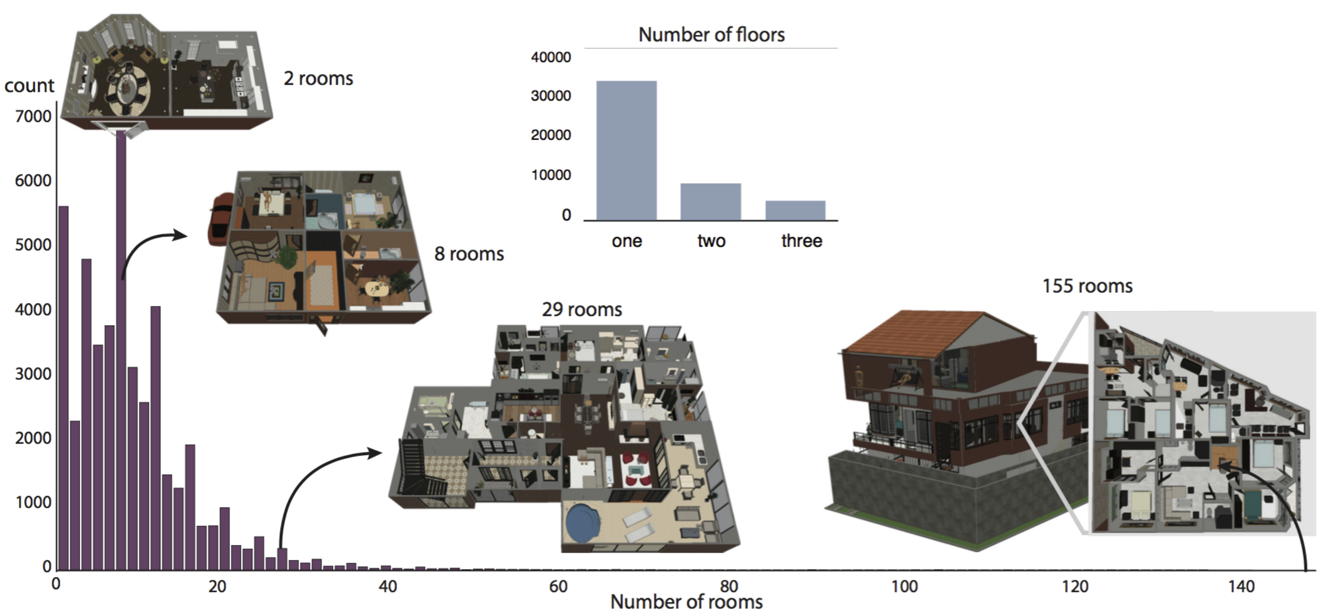

Figure 1. **Scene structure statistics.** Distribution of number of rooms and number of floors in our SUNCG dataset. Our dataset contains large variety of 3D indoor scenes such as small studios, multi-room apartments, and multi-floor houses.

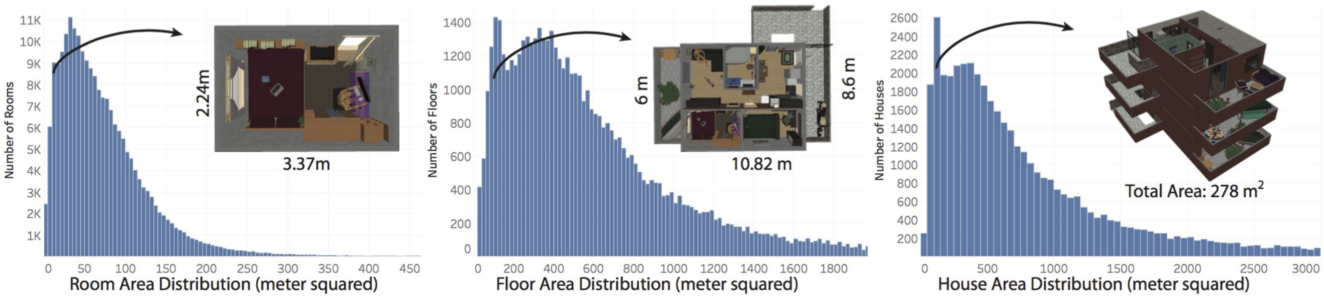

Figure 2. **Distribution of physical sizes** (in meters^2) per room, floor, and house of the SUNCG dataset.

Figure 3. Distribution of different room types in the SUNCG dataset (left), and examples of rooms per room type (right).

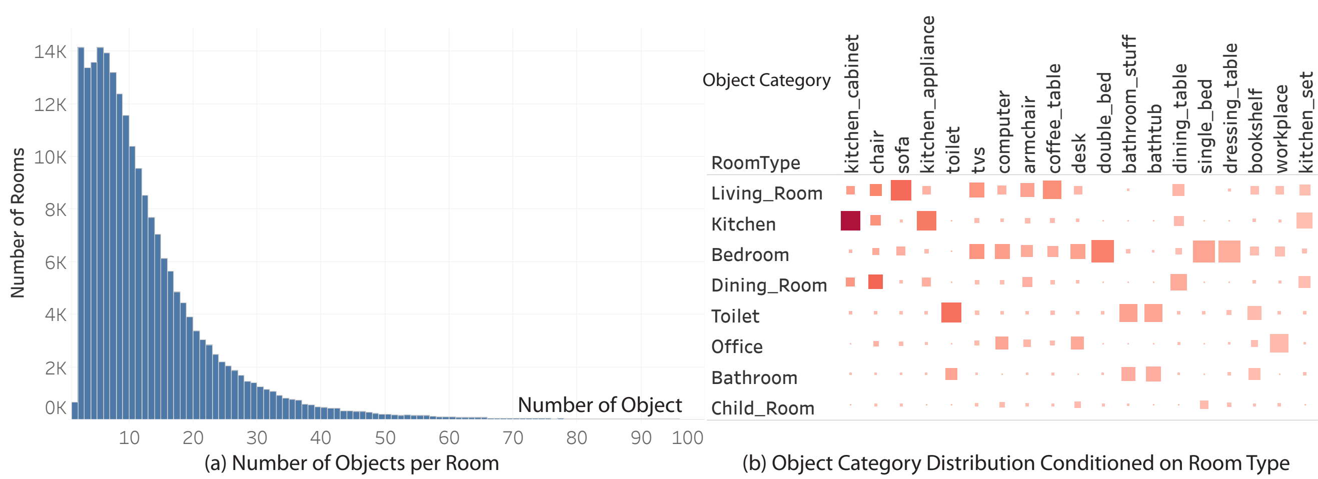

Figure 4. **Object-Room Relationship.** On the left we show the distribution of number of objects in each room. On average there are more than 14 objects in each room. On the right we show the object category distribution conditioned on different room type. Size of the square shows the frequency of given object category appears in the certain room type. The frequency is normalized for each object category. As expected, object occurrences are tightly correlated with the room type. For example, kitchen counters has a very high chance to appear in kitchen, tvs are more likely to appear in living room or bedroom, while chairs appear frequently in many room types.

Room Type Statistics Figure 3 shows the room type distribution and several example rooms per type from our dataset. In total, we have 24 room types that are labeled by the user during creation. These labels include: living room, kitchen, bedroom, child room, dining room, bathroom, toilet, hall, hallway, office, guest room, wardrobe, room, lobby, storage, boiler room, balcony, loggia, terrace, entryway, passenger elevator, freight elevator, aeration, and garage. The four most common room types in our dataset are bedroom, living room, kitchen and toilet, which agrees with the distribution in real-world living spaces.

Object Category Statistics Figure 5 shows overall object category occurrence in the SUNCG dataset. Figure 6 shows examples of object models from the object library, which contains a diverse set of common furniture and objects for common living spaces. Furthermore, during the creation of the 3D scenes, users have the flexibility to reshape, resize, and re-apply texture to objects to better fit the room style, which further improves the dataset diversity.

Object-Room Relationships With complete object and room type annotations, we can further study the object-room relationships in our dataset. Figure 4 (a) shows the distribution of number of objects per room. Figure 4 (b) shows the distribution of object categories conditioned on different room types. On average there are more than 14 objects in each room. The occurrence and arrangements of these objects in rooms provide rich contextual information that we can learn from.

2. Implementation Details

We implement our network architecture in Caffe [2]. We randomly initialize all layers by drawing weights from the Xavier algorithm [1], and initialize biases to 0. We train with a fixed learning rate of 0.01. We run SGD with a momentum of 0.99, and weight decay of 0.0005. During training, each mini-batch has one volume, and we accumulate gradients over four iterations and update the weights once afterwards. Therefore our effective mini-batch size is four. Figure 7 and 8 show more results and error visualizations.

References

[1] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Aistats*, volume 9, pages 249–256, 2010.

[2] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

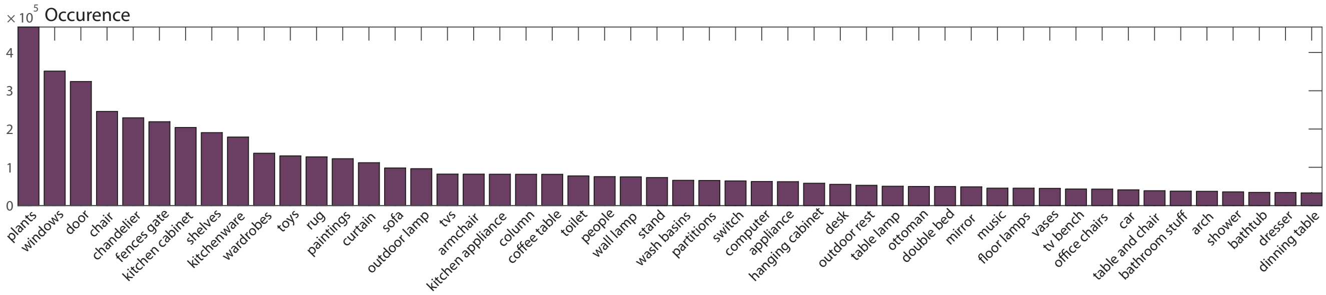

Figure 5. **Distribution of object categories in the SUNCG dataset.** We have 84 object categories in total. Here we show the top 50 object categories with highest number of occurrences in our dataset.

Figure 6. **Object meshes of selected object categories.** The total number of unique object meshes per category in the object library is listed in parentheses. During the creation of 3D scenes, users have the flexibility to reshape, resize, and reapply textures to objects to better fit the room arrangement and style, which further improves the diversity of our dataset.

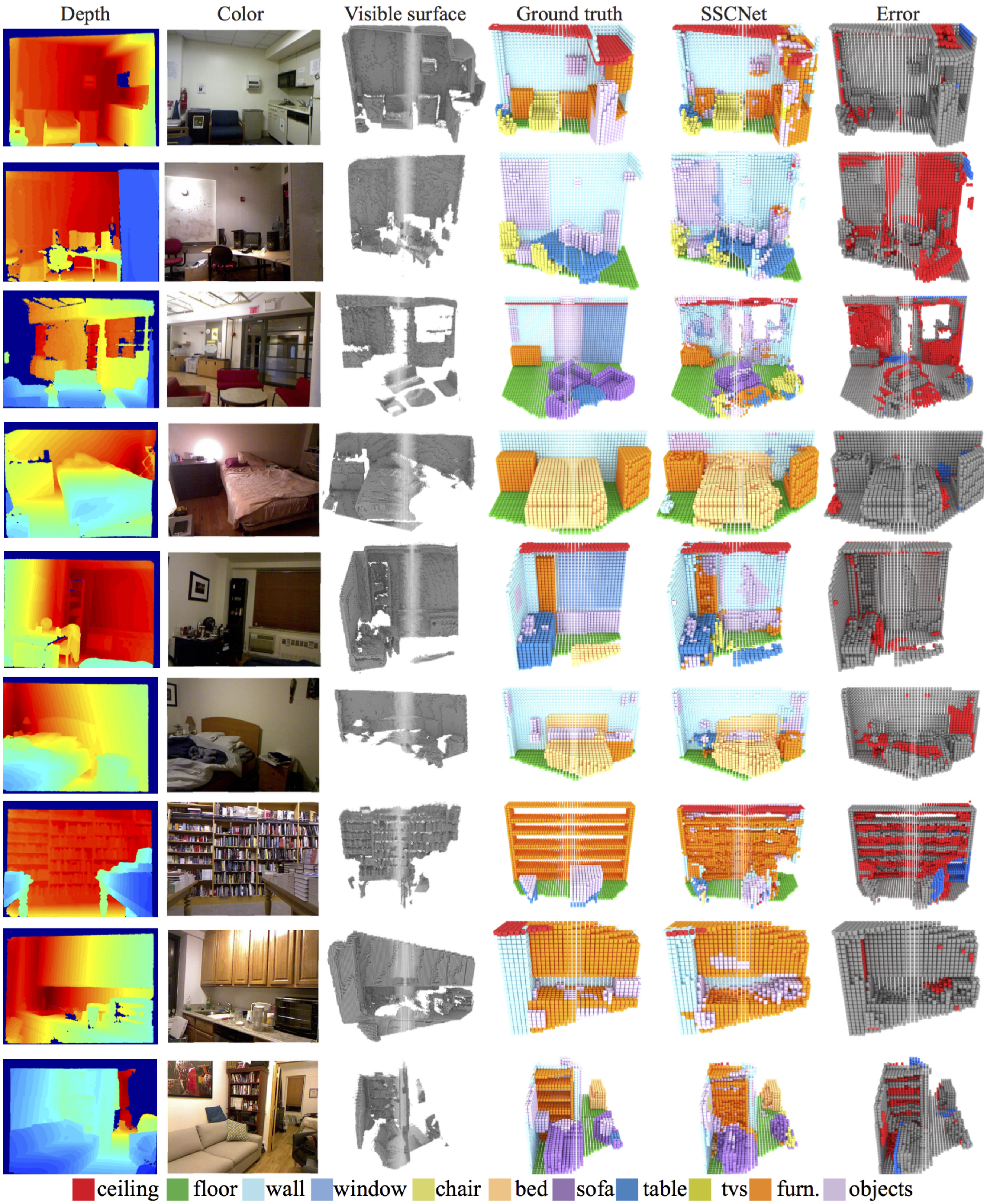

Figure 7. **Results and error visualization.** The first three columns show the input depth map, corresponding color image and visible surface. The fourth and fifth columns show the ground truth and prediction results for the semantic scene completion task. The sixth column visualizes the error of completion result in the evaluation region (not include voxels in observed free space, or outside field of view). Gray voxels indicates true positive, red voxels indicates false positive and blue indicates false negative for the binary scene completion task.

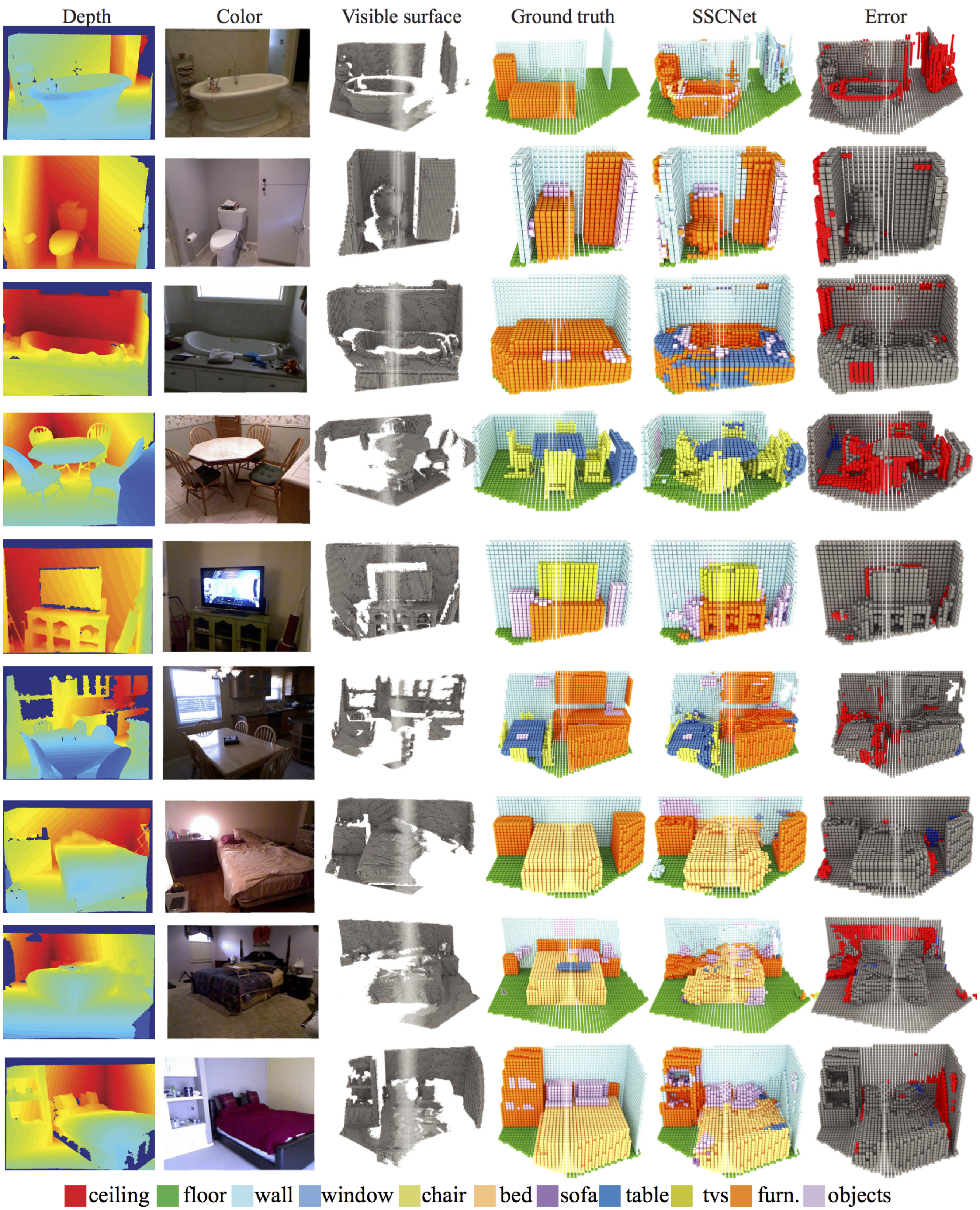

Figure 8. **Results and error visualization.** see Figure 7.