This method is inspired by a projected led by Manuel and his research group. A combination of different models is used for different stage of generation. Firstly, a 2-dimentional features of the input image is extracted into two parts: A depth map and an instance mask. Mask R-CNN is used to segment individual instances and label them appropriately. Next, the image is converted to a plain 3-dimentional object using the estimated depth and the locational information of the camera and encoded by truncated signed distance field. Finally, a convolution network is used to generate the predicted 3-d model sparsely. To represent the resulting 3-d model, an instance label is assigned to the locations that whose distance from the nearest surface is less than a certain threshold. Afterwards, the labelled locations are rendered as a part of the object.

Cornell RGB-D Dataset

Washington RGB-D Scenes Dataset

NYU Depth Dataset

SUN 3D Dataset

UZH Dataset

3D-FRONT

A variety of dataset can be used to train models for 3-d reconstruction. However, due to the nature of reconstruction models, regular RBG images is not sufficient to generate a result with satisfying accuracy. Thanks to the widespread use of RGB-D camera, the availability of indoor scene dataset is no longer a problem. We then can select the datasets that are most suitable to our project. The distinct feature of indoor scene is high complexity, low error tolerance and rich variety but low dynamic. So, the datasets chosen should cover as much as scenarios and angles as possible. And hopefully, they should contain sufficient instances of possible objects occurred in real life. After carefully investigation, we have selected the following candidate datasets:

Cornell dataset: This dataset covers 52 indoor scenes using kinetic camera. And it is featured by labelled point could form of images generated by SLAM algorithm.

NYU dataset: This dataset includes plentiful densely labelled indoor scenes from 528 apartments in the US. However, the pose information is not included which is critical in some generation models.

SUN dataset: 254 different indoor scenes are recorded by SUN dataset. In addition, it also provides full information of camera pose and object labels.

3-D FRONT: It provides 18,968 3-d instances of completely virtual rooms. A full-information image can be obtained by rendering algorithms such as Blender-Proc.

A variety of dataset can be used to train models for 3-d reconstruction. However, due to the nature of reconstruction models, regular RBG images is not sufficient to generate a result with satisfying accuracy. Thanks to the widespread use of RGB-D camera, the availability of indoor scene dataset is no longer a problem. We can therefore select the datasets that are most suitable to the project. The distinct feature of indoor scene reconstruction is high complexity, low error tolerance and rich variety with low dynamic. So, the datasets chosen should cover as much as scenarios and angles as possible. And hopefully, they should contain sufficient instances of possible objects occurred in real life. After careful investigation, we have selected the following candidates:

***Method 1:*** One of the methods is based on a projected led by Dejan and his research group [6]. This method firstly extracts the depth map and an instance mask of the input RGB image. The image is firstly learnt by two multi-layer perceptrons that model pixel points and correction codes. Then the MLPs are optimised by a loss function which is the signed distance. Finally the model colours the result using the radiance computed by the second perceptron.