

## APPENDIX: PARAMETER SETTING OF ADAPTIVE O-CNN

The detailed Adaptive O-CNN encoder and decoder networks for an octree with max-depth 7 is shown in Figure 1. In the figure,  $\text{data}[d]$  represents the input feature at the  $d^{\text{th}}$  octree level.  $\text{conv}(c)$  represents the convolution operation with kernel size 3 and output channel number  $c$ .  $\text{deconv}(c)$  represents the deconvolution operation with kernel size 3, stride 2 and output channel number  $c$ . The kernel size and stride of the pooling operation are both 2.  $\text{FC}(c)$  represents the fully connected layer with output channel number  $c$ .  $\text{prediction}(c_1, c_2)$  is the *prediction module* introduced in Section 4.2, which includes two  $\text{FC}(c)$  operations. Here  $c_1$  is the number of output channels of the first  $\text{FC}(c)$  operation, and is fixed to 8. And  $c_2$  is the number of output channels of the second  $\text{FC}(c)$  operation, with which the plane parameters and the octant statuses are predicted. Since we make the octree adaptive from the  $4^{\text{th}}$  level, the value of  $c_2$  at the second and the third level in Figure 1 is set to 2, predicting whether to split an octant or not. From the  $4^{\text{th}}$  octree level the value of  $c_2$  is set to 7: 3 channels of the output are used to predict the octant fitting status: *empty*, *surface-well-approximated*, and *surface-poorly-approximated*; the other 4 channels are used to regress the plane parameters ( $\mathbf{n}, d^*$ ). The input latent code dimension of the decoder is set to 128.

We use the SGD solver to optimize the neural network, and the batch size is set to 32. In the shape classification experiment, the initial learning rate is 0.1, and it is decreased by a factor of 10 after every 10 epochs, and stops after 40 epochs. In the 3D autoencoding experiment, the initial learning rate is 0.1, and it is decreased by a factor of 10 after 100k, 200k, 250k iterations respectively, and stops after 350k iterations. In the shape prediction from a single image task, the initial learning rate is 0.1, and it is decreased by a factor of 10 after 150k, 300k, 350k iterations respectively, and stops after 400k iterations.

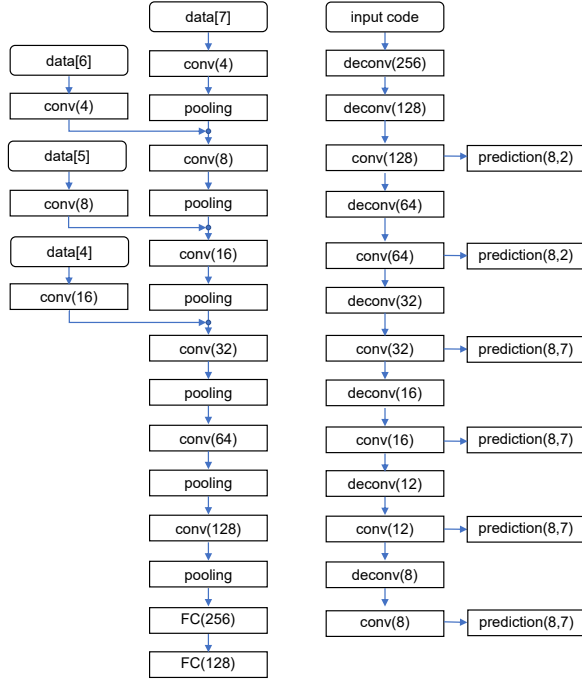


Fig. 1. Adaptive O-CNN encoder (left) and Adaptive O-CNN decoder (right) for an octree with a max-depth 7.