Appendix

Hyper-parameters of CAVnet

CAVnet adopted a U-net-like architecture (Fig.1) (Table 1) that is comprised of an encoder and a decoder. The encoder includes four down-sampling layers to reduce the image resolution and five sub-modules to extract features at different scales. Each sub-module (Table 2) fused features extracted by a convolution layer and atrous convolutional layer with different receptive fields. The features with different receptive fields and at different scales provide the network with contextual and global information. The decoder consists of four sub-modules and deconvolutional upsampling layers. The resolution is sequentially increased through the up-sampling operation until it is consistent with the resolution of the input image. The network also uses a skip connection to connect the up-sampling result with the output of the sub-module with the same resolution in the encoder as the input of the next sub-module in the decoder. Except for the last convolutional layer, batch normalization and a LeakyRelu activation function are used after each convolutional layer. A softmax activation function is used for multiple classifications in the last convolutional layer.