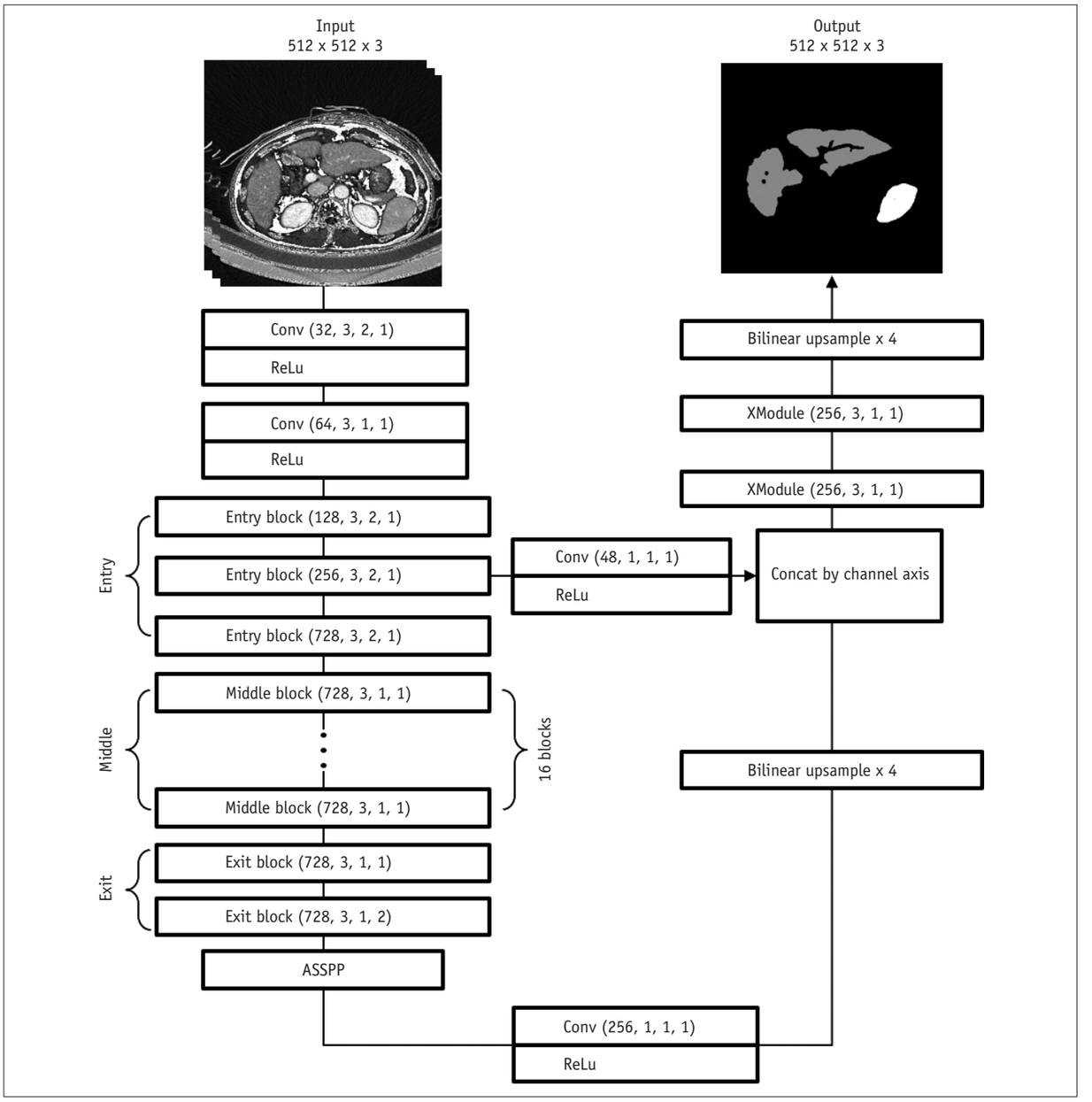
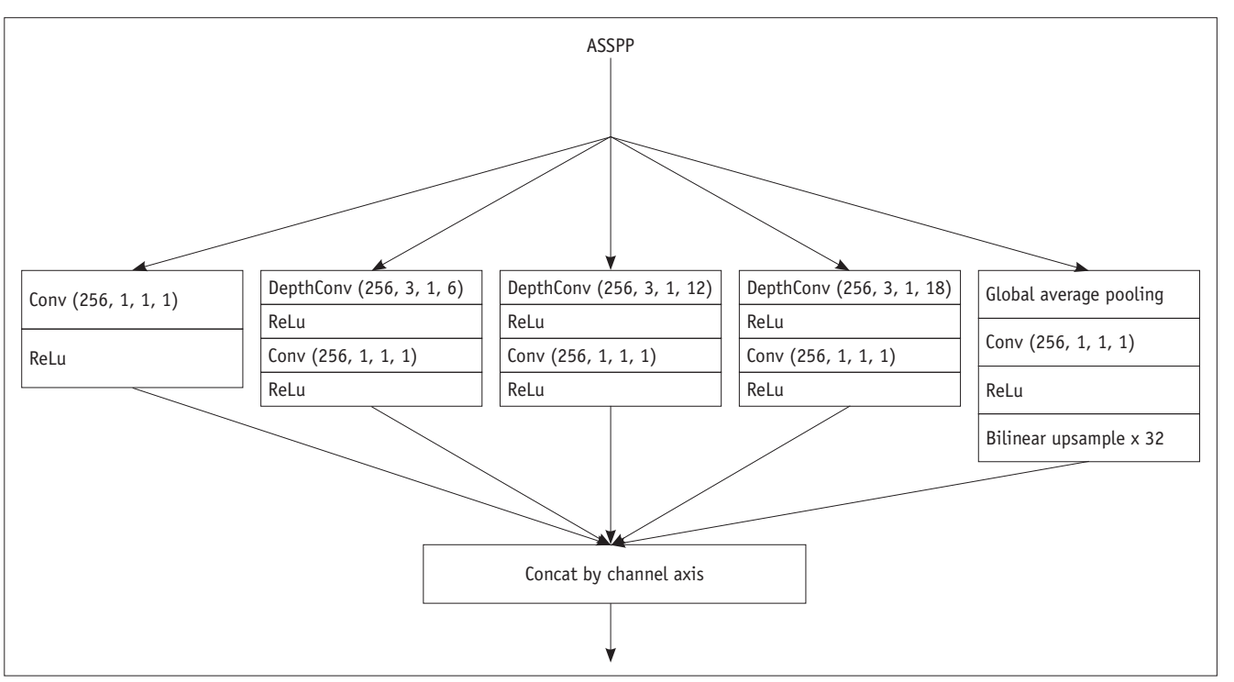


A



B

Supplementary Fig. 2. Diagrams showing detailed architecture of deep learning algorithm. Numbers in parentheses of diagrams indicate hyperparameters (filter, kernel, stride, and rate).

A. Overall model architecture. Model receives three consecutive CT images using 2.5-dimensional input set-up. Encoder consists of Xception model and ASSPP unit. Input images are processed by two blocks of Conv and ReLU, followed by three entry blocks with 128, 256, and 728 filters that downsample feature maps by factor of two; sixteen middle blocks; two exit blocks; and ASSPP block. Decoder upsamples output of ASSPP block after 1 x 1 Conv. There is skip connection between encoder and decoder, where output of second entry block of encoder undergoes 1 x 1 Conv and Concat to upsampled features by channel axis. Following two XModule and bilinear upsampling layers, model returns logit map with same size of input CT images (512 x 512 x 3 channels) where probabilities of liver, spleen, and background for center section of input CT images are assigned to channel axis. Note that all Conv layers are followed by batch normalization layer. **B.** Details of ASSPP unit. ASSPP unit consists of 1 x 1 Conv layer, Xmodule containing DepthConv layers with different Conv rates of 6, 12, and 18, and global average pooling layer. These layers are then concatenated by channel axis. ASSPP = Atrous Separable Spatial Pyramid Pooling, Concat = concatenation, Conv = convolution, DepthConv = depth convolution, ReLU = rectified linear unit, XModule = Xception modules