18pt

16pt

14pt

The above path represents a single image classification, and the bottom path represents semantic segmentation. Both tasks have the same front end, which consists of convolution and downsampling. After the intermediate learning sample representation is obtained, the classification task gets the sample representation of logits directly through the multi-layer perceptron, and the semantic segmentation task gets the pixel-level logits through the upsampling step.