1. Mask R-CNN adopts the policy of detecting the object and then segmenting it. Do you see any issues in this strategy?

2. Mask R-CNN uses the features from ROI Pooling to perform the segmentation. Is there other information that you think might be useful for this task that Mask R-CNN does not use?

3. Can you think of an instance segmentation strategy that starts from semantic segmentation instead?

4. The segmentation mask is predicted using a convolution over the ROI-Pooled features. What might be the advantages and disadvantages of using a non-convolutional, standard linear layer (also called fully connected layer) instead?