1. Consider removing all downsampling and upsampling operations from the U-Net so that all layers are $3 \times 3$ convolutions and all feature maps are roughly the same height and width. What impact will this change have on the computational cost? On the expressivity of the model?

2. The U-Net uses the same number of feature channels in the contracting part (left part of Figure 1) and the expanding part. One could imagine using much fewer channels in the expanding part instead. How do these choices compare in terms of computational cost? Risk of overfitting? Model capacity?

3. The U-Net connects earlier intermediate outputs to newer intermediate outputs. In this it is similar to residual networks. How is the interconnection pattern different from residual networks?

4. Do you think a U-net like architecture might be useful for classification?

5. The U-Net model is trained from scratch. Can you think of a way of pre-training the U-Net model on ImageNet classification?