Part A

1.We only check for completeness for this question. If not using MyConv: -0.5

2.No results: -0.5; Not results but with code: -0.3; Low accuracy: -0.1

3. The key points we check for this question are:
- The parameter calculation for batch norm. Batch norm 2d for conv layer only requires parameters equal to the # of channels. -0.25 if this is incorrect.
- The other calculations should be straightforward. If there are obvious mistakes, -0.25, and you should have received an annotation.

4. The pre-processing step will have no effect on the output, if a > 0, due to the batch norm layer in the first block. If a <= 0, the output will be affected, unless we use min-pool instead of max-pool for a < 0. If your answer assumes a > 0, you still get full mark.

Part B

1. We only check for completeness for this question.

2. We only check for completeness for this question.

3. Skip connections help due to the following reasons (you get full mark if you give any two of those)
- It better preserves the information of the original image, and prevents information loss due to the pooling layers.
- Skip connections provide better gradient flow (prevent vanishing and exploding gradients)
- There are more parameters due to the feature concatenation, so that the model is more expressive.

4. We only check for completeness for this question

Part C

1.We only check for completeness for this question

2.We only check for completeness for this question

3.We only check for completeness for this question

4.Each of the question accounts for 0.25. The memory consumption of fine-tuning is proportional to the number of layers "being fine-tuned". (fine-tune last layer: fine-tune entire network = 1: n)

 The computational resources of fine-tuning is proportional to the number of layers in the network. (fine-tune last layer: fine-tune entire network = n: n)

5.Each of the question accounts for 0.25. The network we used in this part is a fully convolutional network. The number of parameters in the convolutional layer is not dependent on the input size, so the number of parameters will not change. On the other hand, the memory consumption is proportional to the input size. Therefore, the memory consumption quadruples as the height and width double.