Tutorial Questions | Week 6
COSC2779 - Deep Learning

This tutorial is aimed at reviewing famous CNN architectures. Please try the questions before you join the session.

1. The GoogLeNet uses inception modules as the basic building block.

   (a) Why does inception module have multiple paths?

   > **Solution:** choosing the right kernel size for the convolution operation becomes tough. A larger kernel is preferred for information that is distributed more globally, and a smaller kernel is preferred for information that is distributed more locally.
   >
   > The Solution for this is have filters with multiple sizes operate on the same level.

   (b) The following diagram shows a variant of the basic inception model called inception V2. What is the main difference of this module compared to the basic inception block? Why is that advantageous?
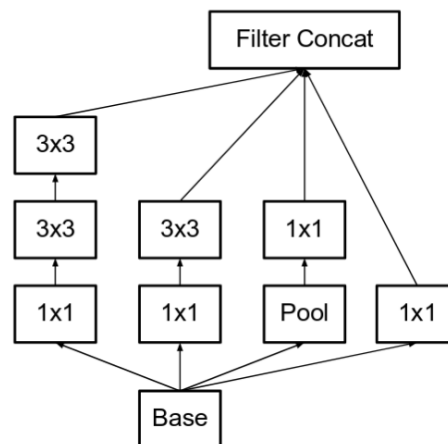
   > **Solution:** Factorize 5x5 convolution to two 3x3 convolution operations to improve computational speed. Although this may seem counterintuitive, a 5x5 convolution is 2.78 times more expensive than a 3x3 convolution. So stacking two 3x3 convolutions infact leads to a boost in performance.

   (c) What does the auxiliary classifier in the GoogLeNet do?

   > **Solution:** They provide additional training signal to the initial layers of the network, providing a solution to the vanishing gradient problem.

   (d) What is the final loss function to train GoogLeNet?

   > **Solution:** $\mathcal{L} = \text{crossentropyfinal} + \alpha_1 \text{crossentropyBranch1} + \alpha_2 \text{crossentropyBranch2}$

2. In ResNet, residual blocks usually have same input and output shapes (height and width). How is the spacial dimensions reduced in ResNet?

   > **Solution:** some residual blocks have strided convolutions in the conv layer and strided 1x1 convolutions in the skip connection.

3. Why does global average pooling based models have significantly less parameters than older CNN architectures like VGG, AlexNet?

> **Solution:** GAP reduce the number of elements in the tensor output by the convolution stage. There is only one fully connected layer after the GAP replacing the bank of FC layer usually employed at the output stages.

4. What are the main types of object detection networks. Explain the major novelties.