第二篇论文疑问点

1. 随着深度学习的发展，人们发现卷积神经网络具有极好的学习特征的能力。卷积神经网络的原理以及简单的理解流程。

答：CNNs主要是做三件事情

1. 读取图片：图片是由一个又一个的像素点构成，而每一个像素点，都是由一个0~255的数字组成，所以，在计算机“看”来，一张图片，就是一个又一个的数字。
2. 特征提取：利用卷积核即过滤器。过滤器的作用是：寻找图片的特征。过滤器会在图片上从头到尾“滑过”一遍。每滑到一个地方，就将该地方的图像特征提取出来。具体提取过程大概是：每一个过滤器中的数值，都是算法自己学习来的，不需要我们费心去设置。每当过滤器在输入特征图框中，缓慢滑过时，我们用过滤器中的每一个值，与输入特征图框中的对应值相乘、再相加。所得到的结果的数值就是我们所提取的“特征图片”。
3. 图片分类：将提取后的特征图，进行坐标划分（x，y），x为一个像素格子中的数字，y我们给每张图片打上的标签：这张图是“猫”、那张图是“狗”…… 统统这些标签，就是y。有很多组（x,y）时，计算机会自己去学习寻找x与y之间的关系，这个“关系”，就是θ。将特征图所有的X展开，并依靠“全连接神经网络”，找到x-y之间的关系θ。输出结果的模型得到的最终结果，表示“图片为某种类别的概率”。注意：这里的概率之和，永远为1。最终根据最高概率的数值判断出图片的真实内容。
4. 特征图来检测不同大小的目标的流程以及原理

答：同问题一。

1. 经过深层次卷积后的特征图拥有丰富的语义信息，解释一下小这句话的原理。特征图如何经过深层次卷积

答：卷积核的通道数和输入特征图的通道数一致

一卷积核对应一个输出特征图

卷积核的大小由自己决定

浅层卷积后底层特征图语义信息比较少，位置信息比较丰富。

经过多次卷积，即深层卷积特征图语义信息越丰富，位置信息越少。

1. 特征提取器的工作原理

同问题1中的回答，特征提取原理。

1. 目标框回归的运行机制、原理、目的以及作用。
2. 答：目标框的坐标预测计算公式：做一个中心化的处理。提高对位置信息判断的准确度。

做回归为了让神经网络去学习。在最小的特征图上，由于其感受野最大，故应该用其来检测大目标，所以大尺度的特征图应该应用小尺寸的先验框，小尺寸的特征图应该应用大尺度的特征图来进行预测框的回归。目标框回归的目的就是要寻找某种映射关系，使得候选目标框(Region Proposal)的映射无限接近于真实目标框(Ground Truth)。对真实目标框的预测，通过预测相对位置的方法预测出目标框相对于坐上角的相对坐标。

1. 先验框和预测框的逻辑关系，以及分别的运行机制和原理，什么是先验框？什么是预测框是用来干嘛的？

答：先验框与预测框的关系，虚线框表示先验框，实线框表示预测框。将原始图片根据特征图尺寸划分成个网格单元，每个网格单元会预测3个预测框，每个预测框包含4个坐标信息和1个置信度信息。当真实框中某个目标中心坐标落在某个网格中时，就由该网格预测这个目标。

1. 特征图大小与先验框的关系。二者之间的作用机制。以及在模型训练过程中起到什么样的作用？

答:将原始图片根据特征图尺寸划分成个网格单元，每个网格单元会预测3个预测框，每个预测框包含4个坐标信息和1个置信度信息。即特征图被划分成多个网格单元，每个网格单元会预测三个预测框。当真实框中某个目标中心坐标落在某个网格中时，就由该网格预测这个目标。预测框通过先验框平移缩放得到，说明特征图的每个单元网格与先验框是一对多的关系。作用就是提高对小目标检测的准确度。

1. 论文对网络模型的增强的四个方面进行改进的一个总体的逻辑流程。

答：要写增强的点：

（1）数据增强：把拼接后的图片输入到神经网路，采用8张图片，增加随机性，一个图片中小目标更多，提高神经网络的敏感性，增加小目标检测的准确性

（2）新尺度的过程：目标框回归公式的修改。