**YOLOv1**

**前向过程**

**输入输出**

图示, 示意图

描述已自动生成输入图片尺寸：448×448×3

输出：算法将每张图像分成S×S个grid cell，每个grid cell预测B个预测框bounding box各自的x,y,w,h,confidence五个参数，一张图片一共预测个值，C个类别的

在YOLOv1模型中S=7,B=2,C=20(数据集中共有20个类) ，即预测个值

**各预测值含义**

：类条件概率，当前grid cell已经包含物体的条件下的各类别的概率

x,y：bbox的中心点相对于grid cell左上角的坐标

w,h：bbox的相对于整幅图像的宽度和高度

confidence：置信度，包括该bbox是否包含物体和该bbox预测的有多准两重信息

各个预测值的意义应当从其所要拟合的值来看

**网络结构**

图示, 工程绘图

描述已自动生成

7×7×64-s-2中的-s-2表示stride(步幅)为2，不加-s-2则默认步幅为1

每个Conv层都默认same填充，即输出大小等于输入大小/步幅

Same填充：

设输入图像尺寸M，卷积核大小K，步长S，在输入图像周围填充P圈0，则应满足

每个Conv层后都添加leaky-relu激活

**预测推断**

图形用户界面

描述已自动生成

每个grid cell预测一组，表示当前grid cell包含物体情况下各类别的概率，由B个bbox共享，其最大值对应的类别在图中用不同颜色表示

框线粗细表示每个bbox的confidence大小

对于每个bbox，计算，表示其对应的各个类别置信度，用于NMS

：如果预测框负责检测物体则取1，否则为0

：预测框与真实框的交并比

***Problem***

较大的物体和处在多个grid cell的边界上的物体可能被多个bbox预测(尽管损失函数也在避免这种情况发生，但其无法降至0)

***Solution***

置信度过滤：将confidence<0.5的预测框的置为0

非极大值抑制(Non-Maximal Suppression)

遍历全部C个类别中每个类别：

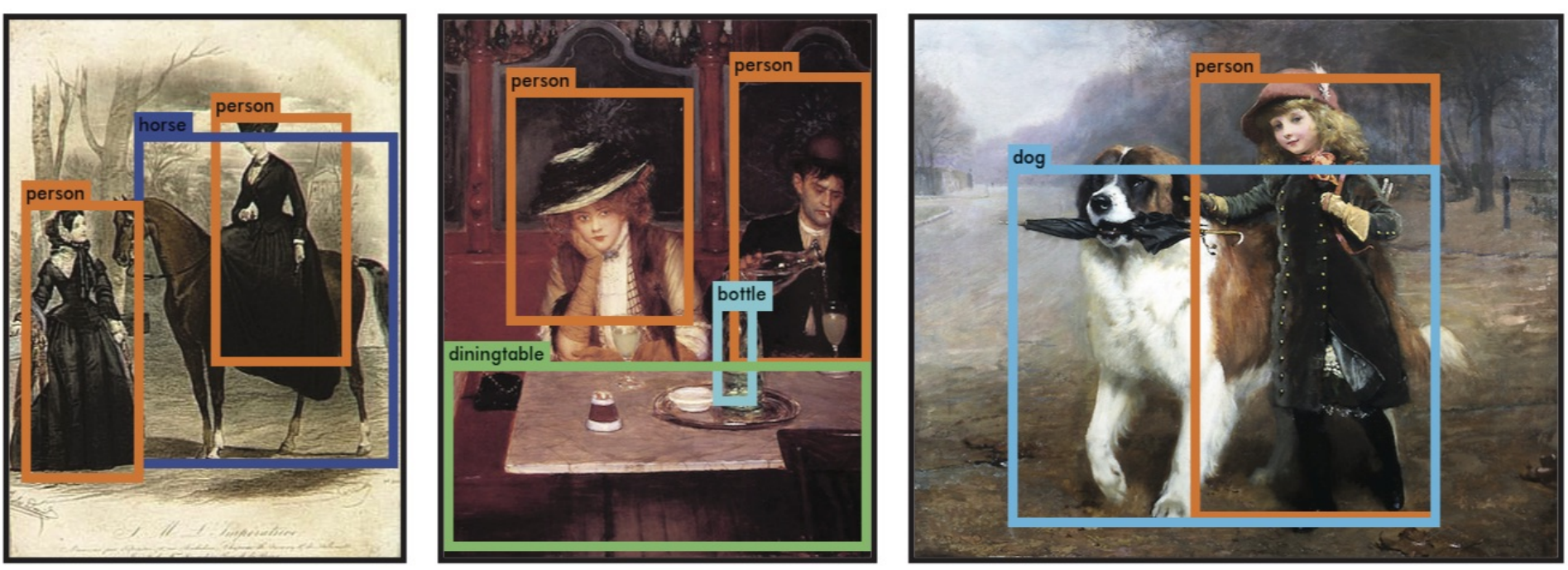
按从高到低将 bbox排序

将最大的bbox与其后所有非零对应的IoU比，如果IoU> 0.5，则置该

按从高到低将bbox重新排序

取下一个非零的bbox重复上述操作，直到某bbox后都为零或其后已无bbox

**可视化**



遍历全部C个类别中的每个类别：

若存在不为0，则绘出其对应的bbox

**损失函数**

**损失函数表达式**

符号说明：

：第i个grid cell的第j个bbox负责检测object则为1，否则为0. 当时，与ground truth的IoU最大的bbox负责检测object；当时，B个bbox都不负责检测物体，即

：第i个grid cell的第j个bbox不负责检测object则为1，否则为0

：第i个grid cell是否包含object，也即是否有ground truth的中心点落在此grid cell中则为1，否则为0

：ground truth中心点相对于grid cell(总共S\*S个)左上角位置(的偏移量，

注：x轴正方向向右，y轴正方向向下

：ground truth相对于整幅图像的宽高，

：每个bbox的confidence，直接由回归得到

：为confidence的标签值

，在条件下，为bbox与ground truth的IoU；在条件下，

：每个grid cell预测的C个类的类别概率

：在条件下，object标注类别的为1， 其余为0

**损失函数各部分的功能**

：负责检测物体的bbox的中心点定位误差

：负责检测物体的bbox的宽高误差

：负责检测物体的bbox的置信度误差

：不负责检测物体的bbox的置信度误差

：负责检测物体的grid cell的分类误差

**损失函数设计过程中存在的问题及解决方案**

***Problem1***

8维(2×4)的定位误差和20维的分类误差一视同仁不合理

***Solution1***

更重视8维的坐标预测，赋予其对应的损失较大的权值

***Problem2***

在一幅图像中，很多grid cell中没有object，因此很多bbox也就不负责检测object，如果Loss的1~5各项之间系数相同，将会导致第4项的值相对于其他项过大，压倒了其他项的梯度，这可能导致网络不稳定甚至发散

***Solution2***

对不负责检测object的bbox的置信损失，赋予小的权值，对负责检测object的bbox的置信损失和类别损失赋予正常权值1

***Problem3***

若用绝对误差，一般小框的绝对误差和较小，大框的绝对误差和较大，将之一视同仁，对大框不公平(对其惩罚过多)

***Solution3***

加根号缩小大框和小框误差的差距，可以降低这种不公平