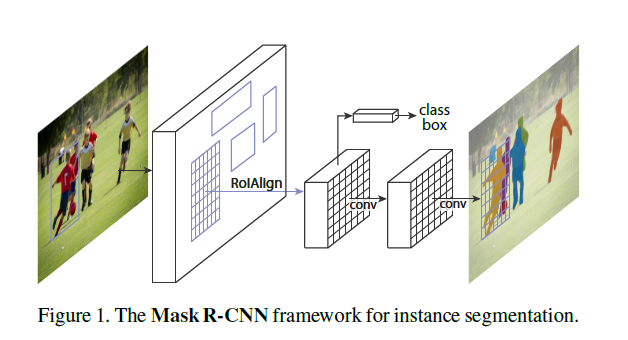
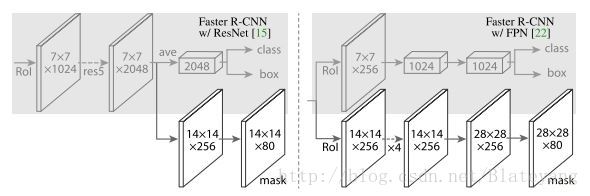
**利用Mask-RCNN进行人体姿态估计**

**Mask R-CNN简介**



Mask R-CNN是何凯明等人在Faster R-CNN基础上提出的一个优秀的目标实例分割模型。该模型能够有效地检测图像中的目标并为每个实例生成高质量的分割掩码。如下图所示，该模型通过在Faster R-CNN已存在的bbox识别分支旁并行地添加一个用于预测目标掩码的分支。掩码分支是一个应用到每个RoI上的小型FCN（全卷积网络），能够预测RoI中每个像素所属的类别，从而实现准确的实例分割。



Mask R-CNN的技术要点主要有三点：

1，在原有的Faster R-CNN基础上采用FCN结构添加了一个并行的mask分支用于生成目标的掩模。

2，由于实例分割对目标掩模的位置精度要求高，为了解决RoIPooling存在的misalignment问题（坐标映射和划分RoI bins时存在量化误差，导致实际的RoI不能严格对齐真实的RoI），作者引入了RoIAlign操作，使用双线性差值（在两个方向分别进行一次线性插值）计算出在每个RoI bin的四个规则采样位置的输入特征的精确浮点值，然后使用max或average整合结果。

3，mask分支的损失函数使用的是基于sigmoid的平均二值交叉熵损失函数，而不是FCN中常用的基于softmax的多类交叉熵损失，这样允许网络为每一类输出对应的mask，避免类间竞争，使得mask与类别预测解耦分离。

**Mask-RCNN用于人体姿态估计**

Mask R-CNN采用的是 “自上而下”的人体姿态估计思路，先检测人的边界候选框，再用single的方法进行估计。

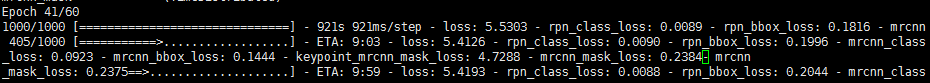
对于每一个生成的候选区域，当检测到该区域包含人这一种类时，会对人体身上每一个关键点的位置进行one-hot编码。M个掩码对应于人体的M个关键点类型之一

One-hot独热编码为一位有效编码，利用K位状态寄存器来对N个状态进行编码，每个寄存器都是相互独立的,在任何条件下只有一位one-hot编码可以作为分类变量的二进制向量表示。分配关键点时，人体每个部位的关键点对应于一个one-hot掩码，训练的目标最终是得到一个m\*m的二进制mask，当中只有一个像素被标记为关键点，其余像素均为背景。对于每一个关键点的位置，进行最小化平均交叉熵损失检测，K个关键点是被独立处理的。

训练过程：

项目工程放在服务器：/ data/ai/zgm/Keypoints-of-humanpose-with-Mask-R-CNN，用了2和3号两个GPU,直接在终端中输入：CUDA\_VERSION\_DEVICE=2,3 python train\_human\_pose.py进行训练

目前分三个阶段进行训练，第一阶段epoch=40,lr=0.002；第二阶段epoch=60,lr=0.0002；第三阶段epoch=100,lr=0.00002;目前第一阶段，第二阶段正在跑；

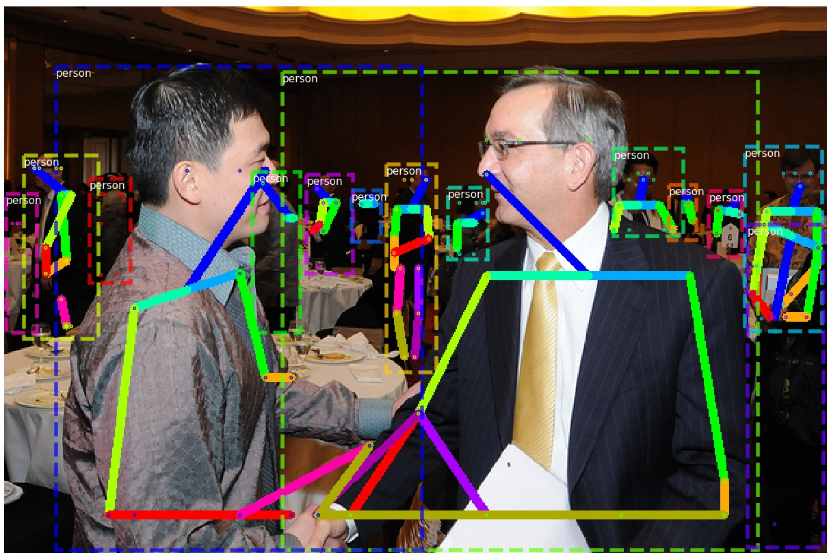


测试过程：

利用预训练好的模型：mask\_rcnn\_coco\_humanpose.h5进行测试

直接在终端中输入：CUDA\_VERSION\_DEVICE=2,3 python demo\_human\_pose.py

测试结果如下：



效果一般，下面会将三个阶段全部训练完后再次测试。