# 第五页：

包括darknet53 fpn yolo head 当有一张图片输入到yolov3中时，我们首先会使用darknet53对图片进行特征提取 1.resize到416\*416大小2.再经过卷积层和残差块进行特征提取3.最终经过darknet53会得到三个有效特征层 这三个特征层分别是 输入的图片压缩3、4、5次的结果将三个特征层传入到FPN （特征金字塔）作用是做加强特征融合五次卷积->卷积和上采样->堆叠 重复多次 构建结束将三个五次卷积后的结果传入到yolo head中 得到预测结果255怎么来的？由于每一个网格点都具有三个先验框，所以255 = 3 \* 85其中的85可以拆分为4+1+80，其中的4代表先验框的调整参数(也就是我们需要四个参数对先验框进行调整来得到预测结果)，1代表先验框内是否包含物体，80代表的是这个先验框的种类(coco数据集)，由于coco分了80类，当然如果是自己的数据集(比如只区分猫和狗 那么可以把80改成2)所以这里是80。如果YoloV3只检测两类物体，那么这个85就变为了4+1+2 = 7。

# 第六页：

batch=64：训练中每次使用的图像数量，也称为批处理大小。这里设置为64，表示在每次模型更新时，使用64张图像来计算梯度和更新模型权重。subdivisions=16：批处理的子分割数。这个参数将批处理分割成更小的部分，每个子部分有16个图像。这有助于节省内存并提高训练效率。width=608、height=608、channels=3：这些参数定义了模型期望接收的图像尺寸和通道数。通常情况下，这表示输入图像的宽度和高度为608像素，通道数为3（RGB图像）。momentum=0.9：优化算法中的动量参数。它表示在更新模型参数时，考虑历史梯度的比重。较高的动量可以加速收敛。decay=0.0005：权重衰减（weight decay）的速率。这个参数控制着模型权重在训练过程中衰减的速度，有助于防止过拟合。angle=0、saturation=1.5、exposure=1.5、hue=0.1：这些参数通常是数据增强的一部分。数据增强通过对训练图像进行旋转、调整饱和度、曝光和色调等方式，增加训练数据的多样性，有助于提升模型的鲁棒性和泛化能力。

# 第八页：

# 遍历模型的输出for output in outputs: # 遍历每个输出的检测结果 for detection in output: # 提取置信度得分 scores = detection[5:] # 获取具有最高得分的类别索引 class\_id = np.argmax(scores) # 获取对应类别的置信度得分 confidence = scores[class\_id] # 检查置信度是否大于0.5 if confidence > 0.5: # 计算边界框的位置信息 center\_x = int(detection[0] \* width) center\_y = int(detection[1] \* height) w = int(detection[2] \* width) h = int(detection[3] \* height) x = int(center\_x - w / 2) y = int(center\_y - h / 2) # 将边界框位置、置信度和类别信息存储到相应的列表中 boxes.append([x, y, w, h]) confidences.append(float(confidence)) class\_ids.append(class\_id)