YOLOv3

主干特征提取网络（Darknet-53）

输入是一个416\*416\*3的图片 -> 进行下采样，宽高会被不断的压缩，通道数不断的扩张

我们可以获得一堆的特征层，可以表示输入

1. Inputs: 416 416 3
2. Conv2D 32 × 3 × 3 -> 416 × 416 × 32
3. Residual Block 1 × 64 -> 208 × 208 × 64
4. Residual Block 2 × 128 - > 104 × 104 × 128
5. Residual Block 8 × 256 -> 52 × 52 × 256 会和 26×26×512特征层上采用的结果进行堆叠->5 次卷积-> 3×3的卷积和1× 1的卷积分别进行分类预测和回归预测 -> 13 × 13× 75(3×25)
6. Residual Block 8 × 512 -> 26× 26 × 512 会和 13×13×1024特征层上采样的结果进行堆叠 ->5 次卷积-> 3×3的卷积和1× 1的卷积分别进行分类预测和回归预测 -> 13 × 13× 75(3×25)
7. Residual Block 4 × 1024 -> 13 × 13 × 1024 -> 5次卷积 -> 3×3的卷积和1× 1的卷积分别进行分类预测和回归预测 -> 13 × 13× 75(3×25)

25的解释：20 + 1 + 4， VOC数据集分20类（20为属于某一个类的概率），1（是否有物体）， 4（调整先验框参数）