一．分类与定位

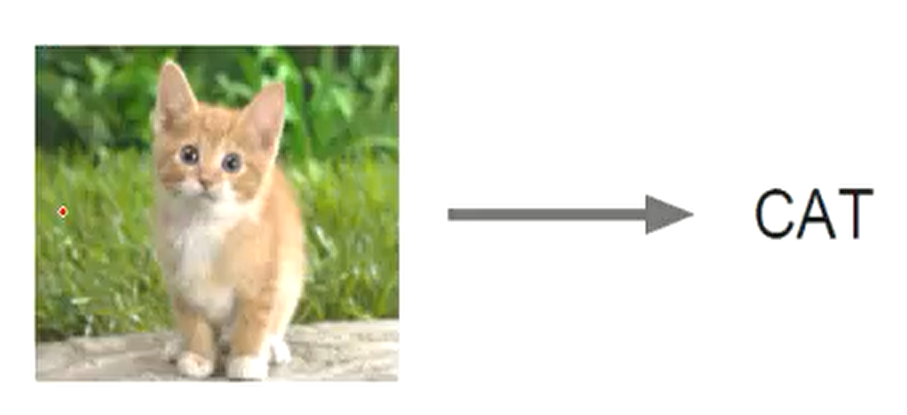
（一）定义

我们不光要对物体进行分类，还要对物体在图片中的位置进行定位。

（二）分类任务

输入图片，输出分类的标签，评估的标准是分类的准确性。

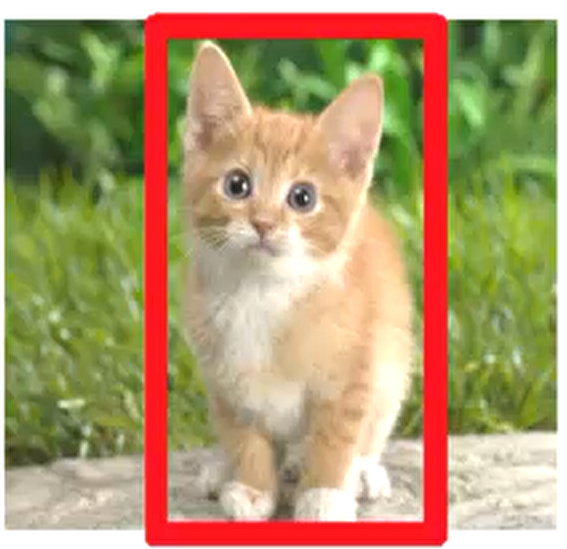
如下图所示：输入一张图片，输出标签为CAT。



（三）定位任务

输入图片，输出一个图片中的一个方框，该方框用向量（x,y,w,h）表示。X是左上角x轴坐标，y是左上角y轴坐标，w是方框的宽度，h是方框的高度。

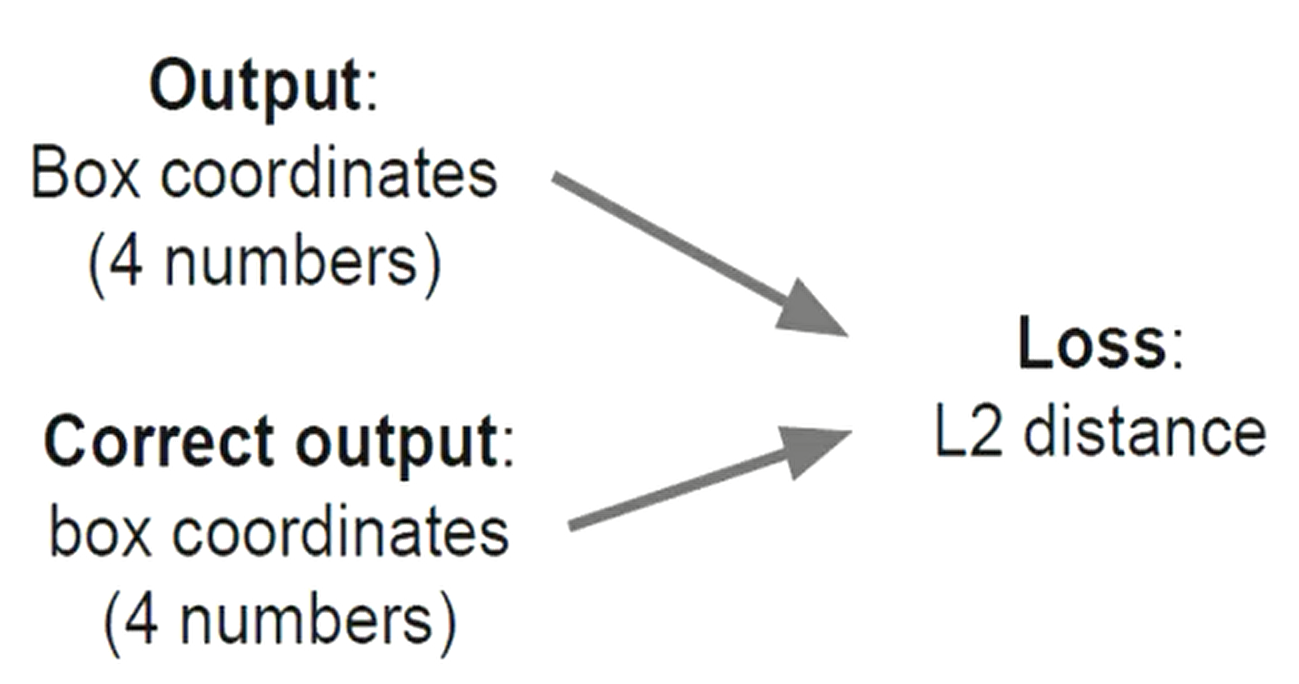
评估的标准是：正确方框位置和输出的方框位置的重合程度。重合程度越高，那么定位越准确。



二．定位过程

在机器学习中，我们学习了回归的概念，通过计算输出值和正确值的差距损失值，反向传播修改参数，直到损失值越来越小，输出值与正确值差距减小，满足我们的需求。

定位的过程也是类似的，通过卷积神经网络计算出方框的位置，计算和正确方框位置的损失值，方向传播修改参数，使损失值越来越小，输出的方框和正确方框越来越接近。



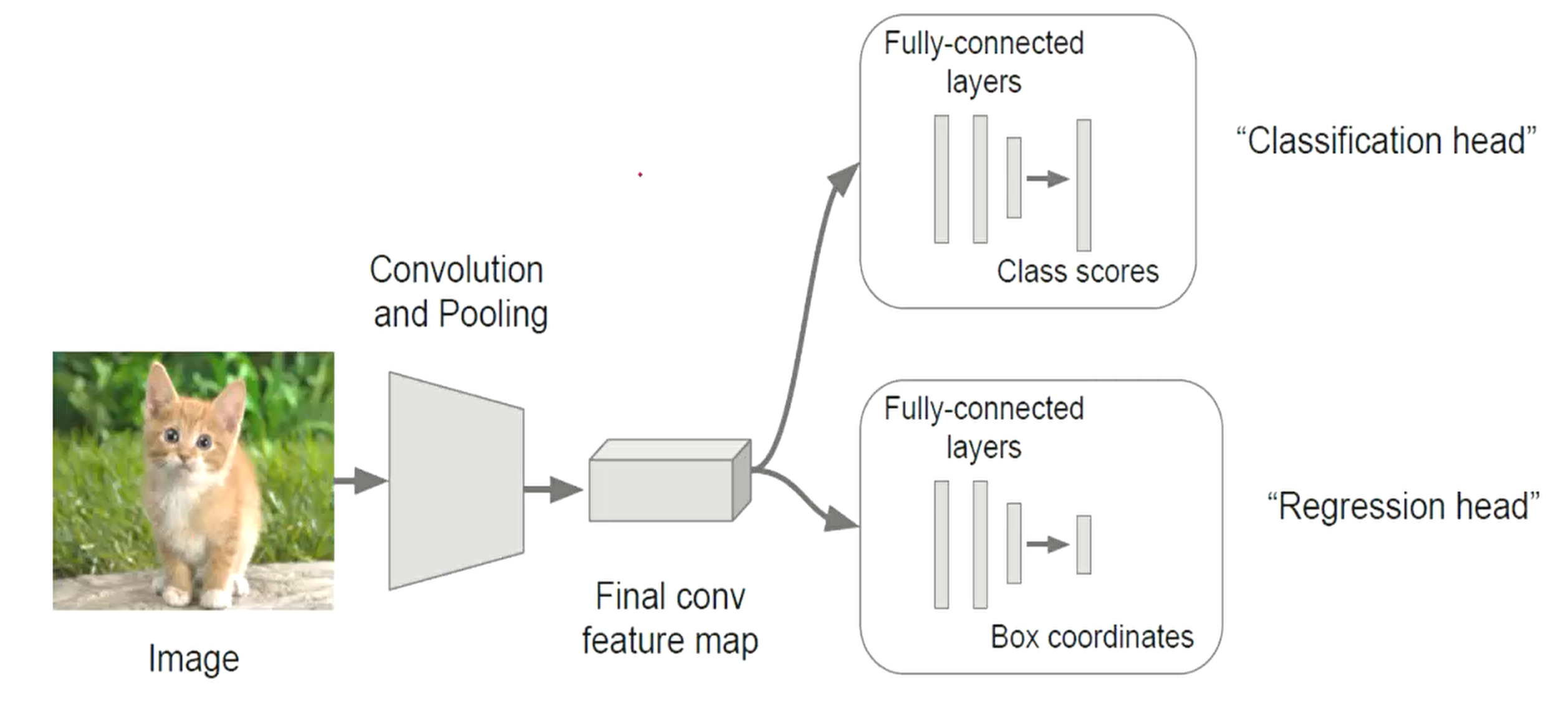
三．分类和定位过程

（一）总体过程

输入图片，通过卷积神经网络提取特征，在得出的最后一层的卷积特征图后，设置两种全连接层任务，一种是分类任务，另一种是回归任务。

分类全连接神经网络：正常的全连接分类神经网络，将图片的类别标出来。

回归全连接神经网络：通过回归得出最优的方框位置。



（二）具体步骤

1. 训练一个分类模型或者微调一个分类模型（AlexNet，VGGNet）

所谓微调一个分类模型指的是将已经很成熟的模型调一下基本参数，已适用于本次分类任务。这样，避免了自己训练一个模型的成本。

2. 给网络添加一个新的全连接层用于回归方框的位置参数。

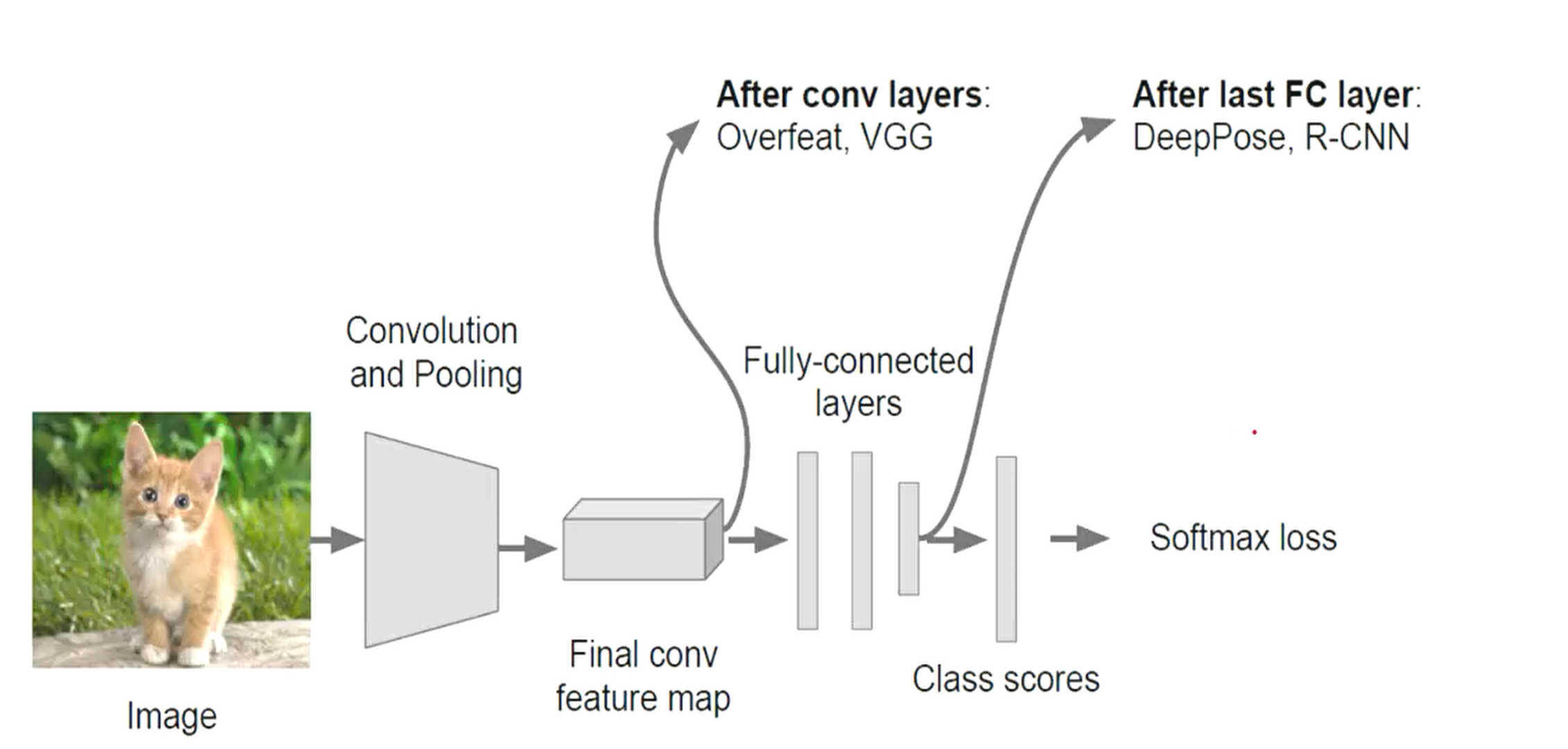
3.用随机梯度下降（SGD）和欧氏距离损失（L2 LOSS）来训练回归层。

4. 同时测试分类和回归。

（三）回归层应该添加在哪里呢？

1. 添加在最后一层卷积之后。如Overfeat，VGG

2.添加在最后的全连接层之后。如DeepPose, R-CNN

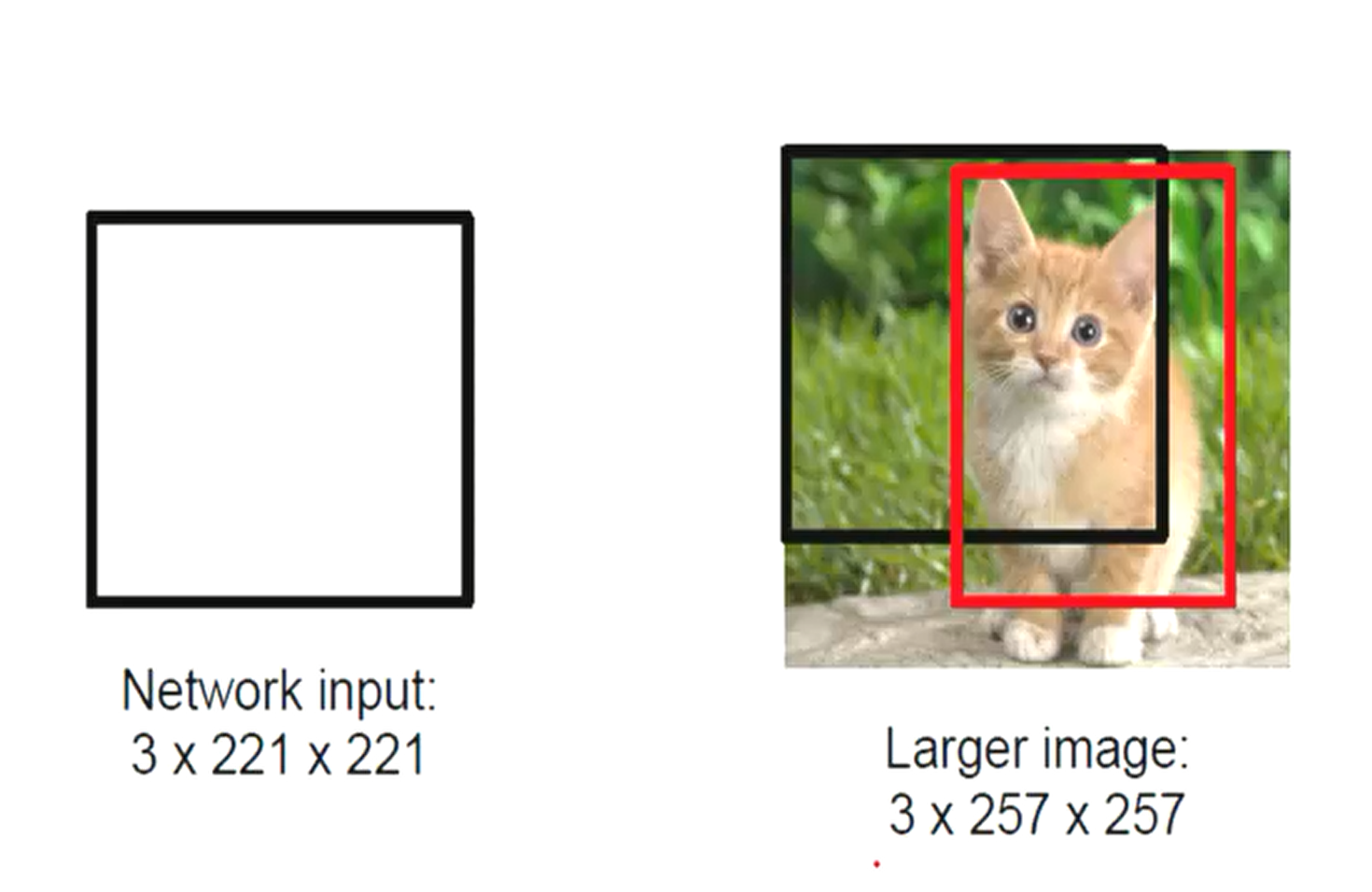


四．滑动窗口

滑动窗口就是为了在输入图片中找出猫，通过在输入图片中滑动设置好的窗口，得出每个窗口中分类为猫的概率值。找出概率值最高的窗口区域，作为定位的区域。

那么现在问题出现了，滑动窗口的尺寸不可能总是和图片中猫的尺寸相适应，可能有的图片中猫的尺寸非常小，那么滑动窗口的尺寸相对于该图片中的猫的尺寸非常大，那么就很难检测出猫，计算出概率值。因此，不同的图片需要不同的预处理 ，通过调整输入图片的尺寸，猫的尺寸也得以调整，以适应滑动窗口的尺寸，便于物体的检测。

如下所示：



左图为滑动窗口，右图为输入图像，在图像中滑动四次，得出的最大概率值区域用于定位回

归。

