

EfficientNet

Rethinking Model Scaling for Convolutional Neural Networks

代码: https://gitcode.net/mirrors/lukemelas/efficientnet-pytorch?utm_source=csdn_github_accelerator

论文出发点或背景

卷积神经网络通常都是先在固定资源预算下开发设计，然后如果资源有多余的话再将模型结构放大以便获得更好的精度。在本篇论文中，我们系统地研究了模型缩放并且仔细验证了**网络深度、宽度和分辨率之间的平衡可以导致更好的性能表现**。基于这样的观察，我们提出了一种新的缩放方法——使用一个简单高效的复合系数来完成对深度/宽度/分辨率所有维度的统一缩放。我们在MobileNets和ResNet上展示了这种缩放方法的高效性。

公认的更大的输入图像尺寸带来更多运算量的同时能够帮助提升准确率

论文创新思路

我们的方法使用一组固定的缩放系数统一缩放网络深度、宽度和分辨率。举个例子，如果想使用 2^N 倍的计算资源，我们可以简单的对网络深度扩大 α^N 倍、宽度扩大 β^N 、图像尺寸扩大 γ^N 倍，这里的 α, β, γ 都是由原来的小模型上做微小的网格搜索决定的常量系数

论文方法介绍

直观地说，复合缩放方法是有意義的，因为如果输入的图像更大，那么网络就需要更多的层来增加接受域，并需要更多的通道来在更大的图像上捕获更细粒度的模式。

更深的卷积神经可以捕捉到更丰富、更复杂的特征，并在新任务上很好地推广。然而，由于梯度消失的问题，更深层次的网络也更难训练

缩放网络宽度通常用于小尺寸模型，更宽的网络往往能够捕获更细粒度的特征，并且更容易被训练。然而，非常宽但较浅的网络往往难以捕获更高层次的特征。

更高的分辨率确实可以提高精度，但对于非常高的分辨率，精度增益会降低

扩大网络宽度、深度或分辨率的任何维度都可以提高精度，但对于更大的模型，精度增益会降低。

对于更高分辨率的图像，我们应该增加网络深度，这样更大的接受域可以帮助捕获类似的特征，在更大的图像中包含更多的像素。相应地，当分辨率更高时，为了在高分辨率图像中使用更多的像素捕获更细粒度的模式，我们也应该增加网络宽度

为了追求更好的精度和效率，在连续网络缩放过程中，平衡网络宽度、深度和分辨率的所有维度都是至关重要的。

Φ 是一个用户指定的系数，它控制有多少资源可用于模型缩放，而 α 、 β 、 γ 则指定如何将这些额外的资源分别分配给网络宽度、深度和分辨率。

实际效果

模型缩放的高效性严重地依赖于baseline网络

个人理解

EfficientNet的base结构是利用结构搜索搜出来的，然后使用compound scaling规则放缩，得到一系列表现优异的网络。对于深度、宽度、通道数，如果只单纯增加三个维度中的一个网络准确率提升会变慢，容易达到饱和。如果同时增加这三个维度，准确率的提升会变快。当分辨率更高时，为了在高分辨率图像中捕获更细粒度的模式，我们也应该增加网络宽度，同时也应该加深深度，来增大感受野，因而三者其实是一种彼此相关的模式。