**See Better Before Looking Closer: Weakly Supervised Data Augmentation Network for Fine-Grained Visual Classification**

博客：<https://blog.csdn.net/David_Yang_7/article/details/100712821>

代码：<https://github.com/GuYuc/WS-DAN.PyTorch>

**摘要**

通常采用数据扩充来增加训练数据量，防止过度拟合并改善深度模型的性能。 但是，实际上，诸如随机图像裁剪之类的随机数据增强效率很低，并且可能会引入许多不受控制的背景噪声。 在本文中，我们提出了弱监督数据增强网络（WS-DAN），以探索数据增强的潜力。 具体来说，对于每张训练图像，我们首先通过弱监督学习来生成注意力图，以表示对象的区分部分。接下来，我们增强由这些注意图引导的图像，包括注意裁剪和注意下降。 提出的WS-DAN将分类准确性提高了两倍。 在第一阶段，由于会提取出更具区别性的部分特征，因此可以更好地查看图像。 在第二阶段，注意力区域提供了对象的准确位置，这确保我们的模型能够更近地观察对象并进一步提高性能。 在常见的细粒度视觉分类数据集中进行的综合实验表明，我们的WS-DAN超越了最新技术，证明了其有效性。

**1.简介**

数据扩充是一种常用的策略，可以通过引入更多的数据差异来增加训练数据的数量，从而改善深度模型的泛化性。它已被证明在大多数计算机视觉任务中都是有效的，例如对象分类，检测和分割。 深度模型有各种数据增强功能，包括图像裁剪，旋转和颜色失真。 以前的作品通常选择随机数据扩充来对其训练数据进行预处理。例如，随机图像裁剪可以生成具有不同平移和比例的图像，从而提高深度模型的鲁棒性。 但是，裁剪区域是随机采样，并且其中很大一部分包含许多背景噪声，这可能会降低训练效率，影响提取特征的质量并抵消其好处。 为了提高数据增强的效率，模型应注意目标对象的空间信息。

精细视觉分类（FGVC）旨在将下属类别归类为基本类别，例如鸟类种类，汽车模型和飞机类型。 FGVC具有挑战性，原因有以下三个主要原因：（1）类内差异很大。 属于同一类别的物体通常呈现出明显不同的姿势和视点。 （2）组间差异小。 属于不同类别的对象可能会非常相似，除了一些细微的差异，例如 鸟头的颜色样式通常可以确定其类别； （3）培训数据有限。 通常标记细粒度类别需要专业知识和大量注释时间。 由于这些原因，仅通过最新的粗粒度卷积神经网络（CNN），例如VGG [23]，ResNet [7]和Inception [26]，很难获得准确的分类结果。

正如最近的工作[5、38、30]所指出的那样，FGVC的关键步骤是在多个对象的零件中提取更具区分性的局部特征。 但是，很难定义对象的各个部分，并且各个对象之间的差异也很大。 此外，标记这些物体的零件需要额外的人工费用。 在这项工作中，我们利用弱监督学习仅通过图像级注释来定位可区分对象的部分。 我们没有提出关注区域边界框[1，5]，而是通过卷积生成的注意力图来表示对象的部分或视觉模式。 我们还提出了双线性注意力集中和注意力正则化损失来弱化注意力集中过程。 与其他零件本地化模型相比[38，30]，我们的模型可以更轻松地定位大量对象零件（超过10个），从而获得更好的性能。

在获得对象零件的位置后，我们提出了以注意力为导向的数据增强方法，以有效地增强训练数据，并解决上述类别间高方差和类别间低方差的问题。对于不同的细粒度类别，除很少差异外，对象通常非常相似。 注意裁切可以通过裁切和调整部分区域的大小来区分出它们，以区分更多的局部特征。 对于相同的细粒度类别，如果模型仅关注少数几个对象的零件，则由于姿势和视点差异而将这些零件封闭时，很可能会预测出错误的类别。 因此，从不同对象的零件中提取局部特征至关重要。 我们的注意力下降会从图像中随机删除对象的一个​​区域，以鼓励网络从其他对象的区域中提取出具有区别性的特征。 因此，通过以注意力为导向的数据扩充，我们的模型可以在多个对象的各个部分中提取出更具区别性的特征，这意味着可以更好地看到对象。

以注意力为导向的数据扩充的另一个好处是，我们可以准确地定位对象，这使我们的模型可以更近地看待对象并优化预测。对于每个测试图像，将对对象类别进行粗略的预测。 该模型首先根据原始图像预测对象的区域和类别的粗略概率。随后，扩大对象的区域并预测精细阶段的可能性。

总之，这项工作的主要贡献是：

1.我们提出了“弱监督学习”方法，以生成注意力图来表示可区分对象部分的空间分布，并提取连续的局部特征以解决细粒度的视觉分类问题。

2.基于注意力图，我们提出了以注意力为导向的数据增强，以提高数据增强的效率，包括注意裁剪和注意力下降。 注意裁剪随机修剪并调整注意部分之一的大小，以增强局部特征的表示。 注意降落会从图像中随机删除关注区域之一，以鼓励模型从多个区分部分提取特征。

3.我们利用注意力图来准确地定位整个对象并将其放大以进一步提高分类精度。

在常见的细粒度视觉分类数据集中进行的实验表明，我们的数据扩充可以显着提高细粒度分类和对象定位的准确性，这超过了最新的方法和基准。

本文的其余部分安排如下。 我们首先在第2节中回顾相关的工作，包括数据扩充和细粒度的视觉分类，然后在第3节中描述拟议的弱监督数据增强网络（WS-DAN）。在第4节中，进行了全面的实验以证明有效性 WS-DAN。 最后，在第5节中得出结论。