用于细粒度视觉分类的弱监督数据增强网络

会议：CVPR2019,由微软亚洲研究院和中科院发表

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

背景：

随机数据增强，如随机图像裁剪之类的方式效率很低，并且可能会引入许多不受控制的背景噪声，因此在本文中提出了弱监督数据增强网络（WS-DAN），对于每张训练图像，首先通过弱监督学习来生成注意力图来表示图像中有区分性的部分，然后增强由这些注意力图引导的图像，包括注意力裁剪和注意力下降。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，这确保模型能够更近地观察对象并进一步提高性能。

简介：

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。细粒度视觉分类（FGVC）的目的是在基本级别目录下分类下级的类别，例如鸟类的种类，汽车的模型和飞机的类型。

FGVC具有挑战性，主要有以下三个原因：

1. 类内差异大。 属于同一类别的对象通常呈现出明显不同的姿势和视点
2. 类间差异小。 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. 训练数据有限。标记细粒度类别通常需要专业知识和大量的注释时间

FGVC的关键步骤是在多个对象的各个部分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。 而且，标记这些物体的零件需要额外的人工成本。本文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，通过卷积生成的注意力图来表示对象的部分或视觉模式。还提出了双线性注意力集中和注意力正则化损失来弱监督注意力集中过程。

在得到对象部件的位置后，提出了注意力指导的数据增强方法，有效地增强了训练数据，并解决了类内差异大，类间差异小的问题。对于不同的细粒度类别，除很少的差异外，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，当这些部分由于姿势和视点变化而被遮挡时，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。论文中提出注意力下降，随机从图像中删除对象区域的一个部分，以鼓励网络从其他对象区域提取出具有区别性的特征。通过以注意力为导向的数据扩充，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，从而更好地观察对象。

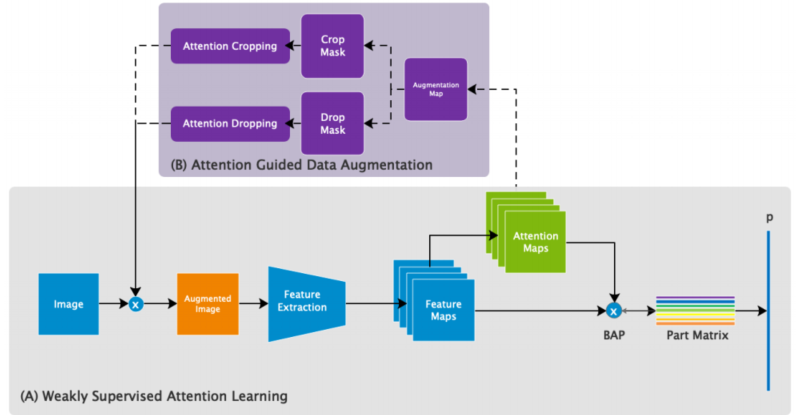
注意引导数据扩充的另一个好处是，可以精确地定位对象，这使模型可以更近地看待对象并优化预测。对于每个测试图像，将对对象类别进行粗略的预测。该模型首先根据原始图像预测对象的区域和类别的粗略概率。随后，扩大对象的区域并预测精细分类阶段的概率。

论文的主要贡献有以下三点：

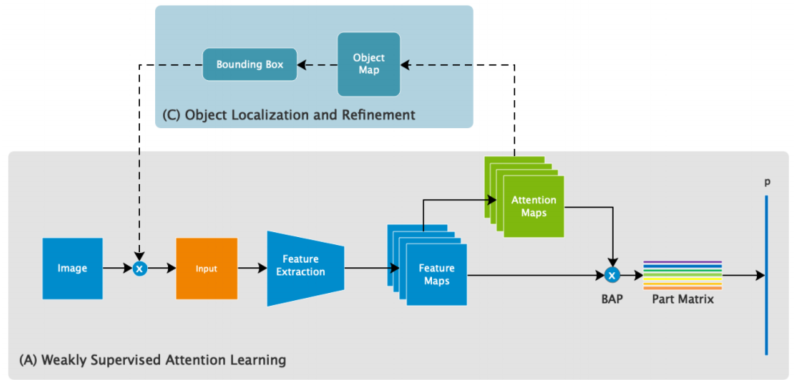
1. 提出弱监督学习，以生成注意力图来表示可区分对象部分的空间分布，并提取一系列局部特征以解决细粒度的视觉分类问题。
2. 基于注意力图，提出了以注意力为导向的数据增强，以提高数据增强的效率，包括注意裁切和注意力下降。注意裁剪随机修剪并调整注意部分之一的大小，以增强局部特征的表示。 注意降落会从图像中随机删除关注区域之一，以鼓励模型从多个区分部分提取特征。
3. 利用注意力图来精确定位整个对象并将其放大以进一步提高分类精度。

方法：

WS-DAN的网络框架如下：



1. 训练过程。（A）弱监督的注意力学习。 对于每张训练图像，将通过弱监督学习来生成注意力图，以表示对象的区别性部分（B）注意指导的数据增强。 随机选择一个注意力图来增强此图像，包括在注意力修剪和注意力下降时。 最后，原始数据和增强数据将被训练为输入数据。



1. 测试过程。首先，对象的类别概率图和注意图将通过（A）从原始图像输出，然后将根据（C）定位对象，将其放大以细化类别概率。 最后，将以上两个概率组合为最终预测。
2. 弱监督注意力学习

1.1空间表示

首先预测对象的零件区域，在训练和测试期间，该对象的位置注释（例如绑定框或关键点）不可用。在我们的方法中，我们采用弱监督学习，仅通过对象类别注释来预测对象的位置分布。

我们使用CNN提取图像I的特征，并将F∈RH×W×N表示为特征图，其中H，W和C分别表示特征层的高度，宽度和通道数。物体部分的分布由注意力图A∈RH×W×M表示，注意力图由F得出：



其中f（·）是卷积函数。Ak∈RH×W代表物体的一部分或视觉图案，例如鸟的头，汽车的轮子或飞机的机翼。M是注意图的数量，注意图将用于扩充训练数据。

我们建议通过注意力图而不是SS或RPN来建议对象的零件区域，因为前者更灵活，并且可以更轻松地在FGVC任务中进行端到端训练

1.2双线性注意力池

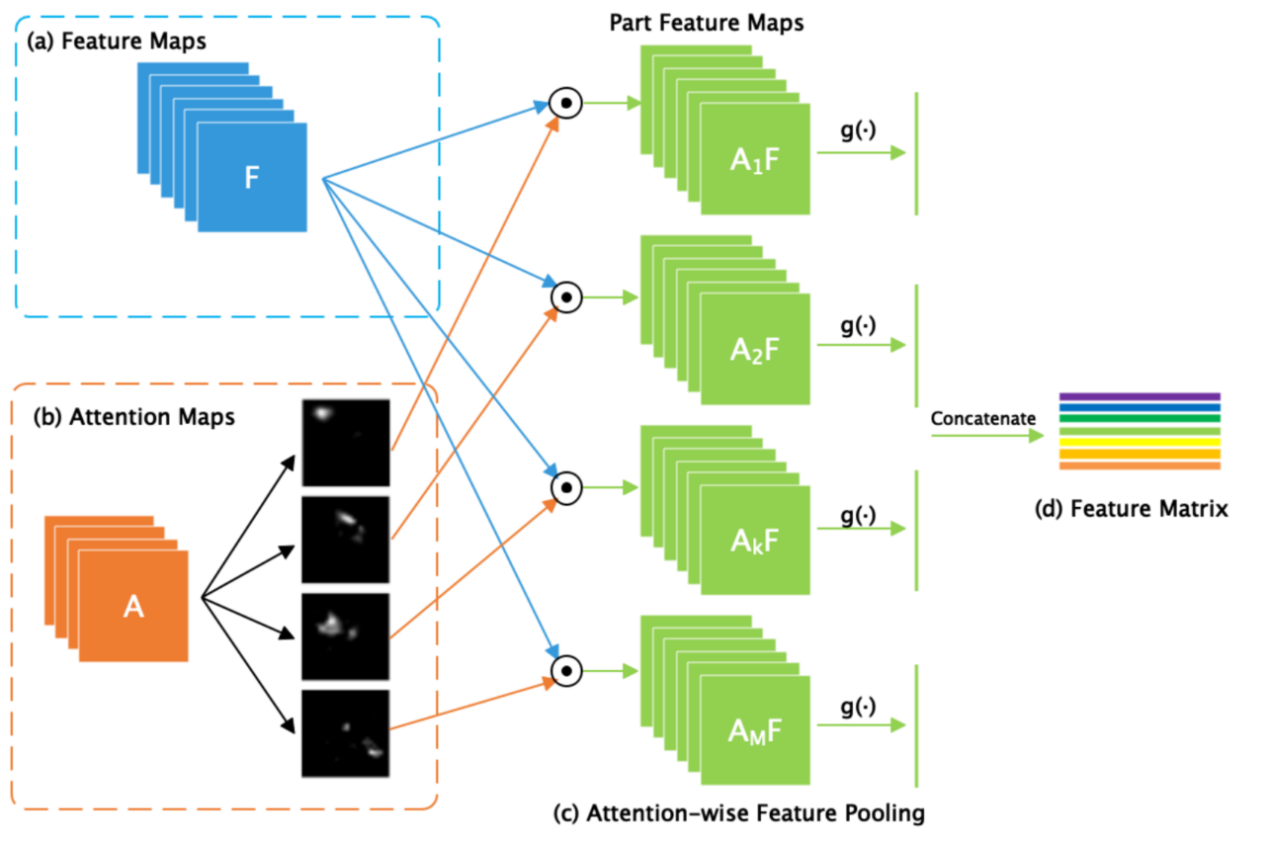


图3.双线性注意力池化的过程。网络主干网（例如Inception v3）首先分别生成特征图（a）和注意图（b），每个注意力图代表一个特定对象的部分；（c）通过将每个注意图与特征图逐元素相乘来生成零件特征图。然后，通过卷积或合并运算提取零件特征。 （d）最终特征矩阵包括所有这些零件特征

在由注意力图A表示对象的零件后，受双线性池的启发，该函数汇总了来自两流网络层的特征表示，然后提出了双线性注意力池（BAP）从这些零件中提取特征。我们将每个注意图Ak逐元素相乘特征图F，以生成M个零件特征图Fk，如下所示



其中表示两个向量逐元素相乘

然后，我们通过其他的特征提取函数g（·）进一步提取可区分的局部特征，例如全局平均池（GAP），全局最大池（GMP）或卷积，以获得第k个注意力特征fk∈R1×N，



对象的特征由部分特征矩阵P∈RM×N表示，这些矩阵由这些部分特征fk堆叠。 令Γ（A，F）表示注意力图A和特征图F之间的双线性注意力池。它可以用下面等式表示：



1.3注意力正则化

对于每个细粒度的类别，我们希望注意力图Ak可以代表相同的第k个对象的部分。受中心损失的启发，提出了针对弱监督注意力学习过程的注意力正则化损失。惩罚属于同一对象部分的特征的方差，这意味着局部特征fk将接近全局特征中心ck∈R1×N，并且注意图Ak将在第k个对象部分中被激活。损失函数如下：



ck是零件的特征中心。ck可以从零初始化，并可以通过移动平均值进行更新，



其中β控制ck的更新率，而注意力调节损失仅适用于原始图像。

1. 注意力引导的数据增强

2.1增强图

随机数据增强效率低下，尤其是在对象尺寸较小的情况下，并且会引入高百分比的背景噪声，使用注意力图，可以更有效地增强数据。 对于每个训练图像，我们随机选择其注意力图Ak来指导数据增强过程，并将其标准化为第k个增强图Ak\*∈RH×W



2.2注意力裁剪

使用增强图，我们可以放大该部分的区域并提取更详细的局部特征。 具体来说，我们首先通过将大于阈值θc∈[0，1]元素设置为1，将其他元素设置为0从中获得裁剪蒙版Ck，等式如下



然后，找到一个边界框Bk，它可以覆盖Ck的整个选定正区域，并将原始图像中的该区域作为扩展的输入数据进行放大，如图1所示。由于对象部分的比例增加，因此可以更好地看到对象 提取更多细粒度的功能。

2.3注意力丢弃

注意力正则化损失监督着每个注意力图Ak∈RH×W来表示相同的第k个对象的部分，而不同的注意力图可能集中在相似对象的部分上，为了鼓励注意力图代表区分对象的多个部分，我们建议将注意力丢弃，具体而言，如下面公式所示，通过将大于阈值θd∈[0，1]的元素设置为0，将其他元素设置为1，我们得到注意力丢弃蒙版Dk。



通过用Dk掩盖图像I将删除第k个部分区域。 放大后的图像如图1所示。由于从图像中去除了第k个对象的一部分，因此将鼓励网络提出其他区分部分，这也意味着可以更好地看到该对象：分类的鲁棒性和定位精度将得到提高

1. 对象本地化和精细化

注意导向的数据扩充还可以有助于先前的弱监督注意学习过程，这意味着可以更准确地预测对象的位置。在测试过程中，模型输出原始图像的粗略分类结果和相应的注意力图后，我们可以通过相同的网络模型预测对象的整个区域并对其进行放大以预测细粒度的结果，表示对象位置的对象图Am通过下面等式计算得到



通过粗粒度预测和细粒度预测对最终分类结果进行平均。 粗到精细预测的详细过程如算法1所示

