Revisiting Pose-Normalization for Fine-Grained Few-Shot Recognition

会议：CVPR2020

代码：https://github.com/Tsingularity/PoseNorm\_Fewshot.

**摘要**

小样本细粒度的分类需要一个模型来仅基于少量图像来学习不同类别（例如鸟类）之间细微的细微区别。 这就要求姿势，关节和背景具有明显的不变性。 一种解决方案是使用姿势归一化表示：首先在每个图像中定位语义部分，然后通过表征每个部分的外观来描述图像。 尽管此类表示在完全监督的分类中不受欢迎，但我们证明了它们对于小样本的细粒度分类非常有效。 通过最小化模型容量的增加，姿势归一化可将浅层和深层体系结构的精度提高10％到20个百分点，将其更好地推广到新领域，并且对多个小样本算法和网络主干有效。

**1.简介**

最小化微调的泛化能力对于学习的神经模型至关重要，不仅对未见的数据而且对未见的数据类型也至关重要。 考虑一下图1所示的任务。我们仅从几种鸟类获得了一张图像（或非常少的图像），仅从这些信息中，我们就必须学会识别它们。 众所周知，人类非常擅长此快速学习任务[19]，但是机器仍在努力：尽管在视觉识别方面取得了巨大进步，并且经过两年的专注研究，但在几个小样本基准上的性能仍然远远低于完全有监督的方法的性能。

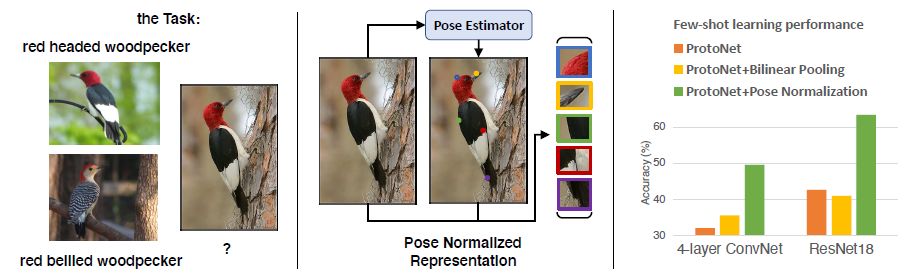


图1.左：细粒度的小样本任务。 对象共享相同的零件结构，并且类别之间的细微差别。 中：基于简单姿态估计器的姿态标准化表示可以捕获语义部分信息。 右图：在浅层和深层骨干上，姿势标准化都可以显着提高小样本的学习性能。 具有我们表示形式的浅层架构（4层ConvNet + Pose规范化）甚至比没有它的更深层的黑盒网络（ResNet18）表现更好。

实际上，这是一个问题，尤其是对于细粒度的分类问题（例如图1中的问题）。 在这种情况下，不同的类别可能有数百种，而正确标记这些类别所需的专业知识和精力会使注释变得昂贵。 在一起，这使得为细粒度分类收集大标签的训练集非常困难，有时很难做到。 因此，神经网络处理细粒度，小样本学习的能力对于实际应用至关重要。

在此任务上机器和人的绩效之间存在巨大差距的原因是什么？ 一个直观的假设是，人类使用了更加稳定的特征表示，这对于大的空间变形是不变的。 例如，在鸟类分类任务中，我们可以使用鸟类各个部分的属性来表征鸟类图像：喙的形状，鸟翼的颜色，有无冠。 这样的特征不仅对图像背景的变化是不变的，而且对照相机姿势和清晰度的变化也不变，这使我们能够有效地感知各种鸟类及其单个图像的相似性和差异性。

这种特征化是“姿势归一化”的，在重新发现卷积网络之前已被探索为细粒度分类的有希望的方向[32]。然而，研究人员发现，使用黑盒架构进行端到端训练，并且不进行姿势归一化，可以大大提高标准基准（尽管进行了一致的修改，例如双线性池[16]）。 的确，近年来，年度细粒度分类挑战[1]（https://sites.google.com/view/fgvc6/home.）的获奖者主要集中在这些黑盒架构上。 姿势归一化的直观思想已被抛弃。

在本文中，我们将重新进行姿势归一化以处理小样本，细粒度分类，并演示其在这种情况下有用。姿势归一化是通过对卷积架构进行非常简单的修改来实现的，增加了很少的新参数（与先前的方法相比，网络大小增加了两倍或更多[34，10]）。 我们的方法与小样本学习技术和骨干神经体系结构的选择正交。 我们用三种不同的小样本学习技术，两种不同大小的主干架构以及三种鸟类和飞机的细粒度分类数据集来评估我们的方法。 我们发现：

1.姿势归一化可全面提高收益，在某些情况下，准确性可提高20点以上，而新颖的类则无需注释任何部分。

2.在所有设置中，姿势归一化的性能都优于对神经体系结构的黑盒修改，例如双线性池。

3.即使只有5％的基类训练数据带有姿势注释，姿势归一化的优点也很明显。

4.姿势归一化对于浅层和深层网络体系结构均有效。 具有姿势归一化功能的浅层网络优于较深层的黑盒网络。

我们观察到的巨大性能提升，再加上体系结构本身的简单性，表明了在细粒度，小样本的分类中姿势归一化的强大功能。

**2.相关工作**

细粒度识别是计算机视觉中的经典问题，并且是反复出现的挑战[1]。 尽管我们专注于鸟类的分类[26]，但提出的想法也适用于其他细粒度的任务，例如识别飞机模型[17]，汽车[15]或对象具有相同零件集合的任何其他问题。 在细粒度识别的背景下，Farrell等人[6]提出了姿势归一化的思想：预测对象的各个部分并将每个部分的外观记录为描述符。 此后，已经探索了该想法的许多版本，包括改变零件的类型[32、10、33]，零件检测器[31]，以及将这些想法与神经网络相结合[34]。 最后一个与我们的工作最相似。 但是，所有这些方法都与完全监督的识别有关，而在这里我们只看小样本识别。

姿势归一化还启发了零件无监督的黑盒模型。 Lin等[16]引入双线性池作为这种归一化的推广，我们在工作中将其与这种思想进行了比较。 空间变压器网络[14]显式实例化无监督的姿势归一化并端到端对其进行训练。 还提出了这种直觉的其他实例[4，11，21]。 但是，这些无人监督的方法增加了相当大的复杂性和计算量，因此很难分辨仅姿势标准化的好处。 相比之下，我们专注于一种轻量，直接，语义的方法，以表明姿势归一化而不是增加网络功能可以提高性能。

小样本学习方法可以大致分为以下三类：1）转移学习基线在基类上训练标准分类网络，然后在冻结表示上为新颖的类学习新的线性分类器。 最近的工作表明这具有竞争力[3，27，18]。 2）元学习技术训练“学习者”：该功能可映射带有标签的小型训练集和测试图像以测试预测。 示例包括ProtoNet [20]，MatchingNet [25]，RelationNet [22]和MAML [7]。 这些学习者有时可能会包括学习到的数据增强[28]，某些方法使用姿势注释[14]来训练。3）权重生成技术为新类别生成分类权重[8，9]。

大多数小样本学习方法使用黑盒网络体系结构，在有足够的标记数据的情况下，黑盒网络体系结构可以很好地发挥作用，但在高度受限的小样本学习场景中可能会受到影响。Wertheimer和Hariharan [29]重新研究了Lin等人的双线性池。 [16]，并发现它运作良好。 他们还引入了一种简单，有效的本地化归一化表示形式，但仅限于粗略的对象边界框，而不是细粒度的部分。 朱等[35]引入了一个语义指导的多注意模块来帮助零样本学习，但是它是完全不受监督的。我们将实验与无监督的基线进行比较。

姿势归一化会增加常见变异模式的不变性。 增加不变性的另一种方法是使用学习的数据增强[12、28、5]。 但是，这通常需要大量的额外网络和大量计算。 相反，我们专注于轻量级方法。

还要注意，我们的基线之一[8]已经优于最近的增强方法[28]。在以下各节中，我们首先概述小样本识别。 然后，我们证明了特征的姿势归一化可以在一系列小样本学习算法中充当即插即用的网络层。

**3.小样本识别**

小样本学习的目标是构建一个学习器，该学习器仅给出少量标记的示例即可产生有效的分类器。 在经典的小样本设置中，首先为学习者提供一个大的标记集（表示集Drepre），其中包括许多来自基类Ybase的标记图像。 学习者必须使用此数据设置其参数以及任何超参数。 然后，它遇到了一组不相干的新颖类Ynovel，从中它获得了一小组参考图像Drefer。 然后，学习者必须从该集合中为新类学习分类器。

在大多数技术中，我们可以将学习者分为三个模块：特征图提取器、特征聚合器和学习算法。

特征图提取器通常被实现为具有可学习参数的深度卷积神经网络。 对于每个输入图像x，网络产生对应的特征图张量，其中C，H，W分别表示特征图的通道，高度和宽度尺寸。

特征集合器 是通过参数化的转换，将特征图转换为全局特征向量：，其中d是潜在维数。通常，是全局平均池化模块。

学习算法提取数据集S中的训练特征向量和相应标签，以及测试特征向量v，并输出标签ˆp的概率分布：。 为了我们的目的，我们考虑三种代表性的方法：

迁移学习遵循标准的网络预培训和微调程序。由具有学习的权重矩阵和softmax激活的简单线性分类器实现。 函数, 与同时进行训练，以最小化Drepre中数据的标准交叉熵损失。为了使模型适应新颖的类，特征提取器参数被冻结，并且在函数上训练新的线性分类器 Drefer中的新类。

原型网络[20]是一种代表性的元学习方法，该方法通过平均每个类别中的特征向量来为每个类别生成原型表示。 然后，是一个非参数分类器，它根据数据点的特征向量与每个类原型之间的距离来分配类概率。 每个训练情节从基本类别Ybase中抽取N个类别，并从每个类别中抽取少量支持集和图像查询集。 支持图像形成类原型，而查询集上的N-way分类会产生损失，并向参数添加相应的更新梯度。

在动态小样本学习[8]中，再次是线性（或余弦）分类器，但不是直接在Drefer上进行微调，而是由学习的权重生成器G生成分类器。训练过程包括两个 阶段。 首先是有关Drepre的标准分类培训。 在第二阶段中，特征提取器参数被冻结。为了训练生成器G，该算法从Ybase中随机选择几个“假”新颖类，并将它们视为真正新颖，使用分类器权重进行分类 由G生成并最小化这些类的模拟“测试”示例上的分类损失。

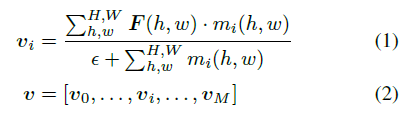
**4.姿势归一化特征向量**

两种直觉激发了我们提出的方法。 首先，对于细粒度的识别，两类之间的外观差异往往非常小。 在小样本设置中，算法很难捕获这些细微的差异，因为只有少数示例可供参考。使用姿势归一化将特征表示聚焦在每个图像的信息量最大的部分上，这将有助于学习过程。 其次，由于细粒度识别涉及相似的对象，因此它们可能共享相同的语义结构。因此，很有可能在基类上训练的姿势估计器将泛化，甚至推广到未见的新类。

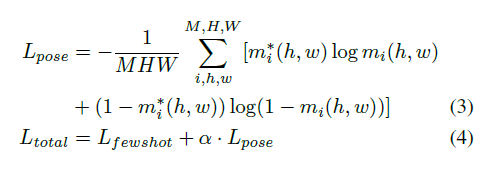
我们假设M个不同的部分。零件注释可用于Drepre中的（某些）基础类训练样本，但不适用于新颖的类。 我们将每个图像x的零件批注格式化为M×H×W位置张量，其中H×W是特征图的空间分辨率。

现在，我们介​​绍提取姿势归一化特征向量的方法。为此，网络必须首先估计姿势。 我们使用一个非常小的两层卷积网络。它在特征图张量上运行,从特征图提取器的中间层提取。在最后一层使用sigmoid激活 并为所有带注释的部分生成一个热图位置预测。我们故意使用一个小的并在中重用计算为了最大程度地减少附加参数可能对分类器最终性能的影响。 改进的性能应表明，姿势信息对于细粒度的小样本学习是有用的，而不是一个更大的网络。

给定heatmap m和特征图F，我们必须构造一个特征向量v。将m中的每个通道用作特征图的空间注意掩码，从而生成一个注意归一化的特征向量。 将这M个零件描述向量串联起来，将生成图像的最终表示向量。 形式上，将表示为特征图F中位置（h，w）的特征向量，将表示为对于第i个零件类别在位置（h，w）的热图像素值，。的计算公式为：



= 训练期间的损失是地面真值部分位置热图之间的逐像素对数损失之和和预测的热图m，以及最初的小样本分类损失：



是一个平衡的超参数。 为了促进在分类分支中的学习，用于小样本分类的特征向量最初从地面真实部分注释热图中生成而不是预测的热图m。 然后，冻结姿势估计网络的参数。在新颖类的后续适应/微调和评估/推断阶段，根据预测的热图m计算特征向量。图2概述了我们的方法。

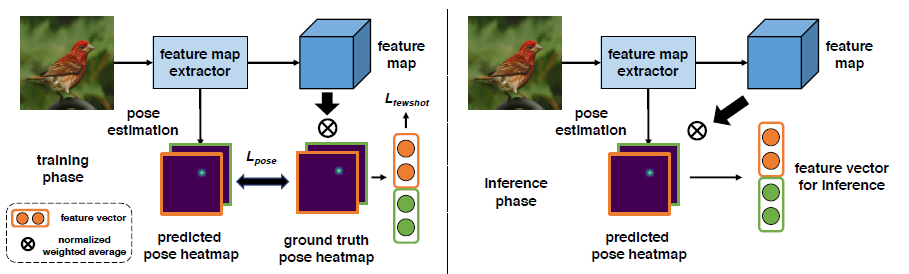


图2.用于训练和推理的姿势归一化框架。姿势估计器将网络主干的中间输出作为输入，并生成姿势热图预测。通过将每个热图应用为对特征图的关注来计算特征向量。最终表示是这些向量的级联。在该示例中，零件数M ＝ 2。

请注意，尽管我们在训练过程中假设有一组固定的一致的零件标签，但我们不需要零件在所有对象中始终出现，也不需要任何特定的对象包含所有指定的零件。 因此，我们的姿势估计器应该广义地概括：依赖于各个部分（例如汽车，家具，昆虫）的外观的对象的任何细粒度分类都适用于此方法。

**5.实验**

**5.1数据集和实施细节**

我们尝试使用CUB数据集[26]，该数据集包含来自200个类别的11,788张图像。 它还为每个图像包括15个零件批注，因此M = 15。 按照[29，3]中的评估设置，我们将数据集随机分为100个基础，50个验证和50个新类。 基本类别图像构成了表示集。 对于每个验证类和新颖类，我们随机抽取其图像的20％，以形成参考集。 其余的新颖图像形成查询集，该查询集用于评估算法。请注意，我们的模型只能在基类中访问零件批注。验证或新颖类中的任何图像都没有零件注释信息，包括它们的参考集和查询集。

NABird评估：CUB的评估集中只有50个新颖的类，这可能会使评估产生干扰。 小样本学习算法之间的精度差异在域移位的情况下也显着降低[3]。 因此，为了验证我们提出的方法的鲁棒性和泛化能力，我们还在另一个更大的鸟类数据集NABird [23]（NA）上评估了我们的CUB模型，该模型在与CUB删除重叠后包含418个类别和 35,733张图像。和以前一样，我们从每个类别中随机抽取20％的图像以形成参考集。 其余图像构成查询集。

网络主干：对于特征图提取器，以前的工作[20，29，8]采用一种标准架构：具有批处理归一化和ReLU的4层，64通道卷积网络。 在此设置下，输入图像尺寸为84×84，输出特征图为64×10×10。 更深的骨干网可以显着减少这些方法之间的性能差异[3]，因此，除了4层网络之外，我们还通过一些技术改进来训练和评估ResNet18 [13]骨干网，从而提高了所有模型的性能。我们将最后一个块的第一个卷积和下采样层的步幅从2更改为1。因此，最后一个块的输出大小保持为14×14，而不是7×7。 我们还向原始ResNet18的最后一层添加了具有批处理归一化的1×1卷积，这将通道数从512减少到了32。经过修改的ResNet18的输入大小仍为224×224，但是输出大小为32×14×14。

姿势估计模块：姿势估计网络的层由Conv-BN-ReLU-Conv组成，其中Conv表示3×3卷积。在4层ConvNet中，以第二次之后的特征图为输入卷积。中两个卷积层的输入/输出通道数分别为64/30和30 / M，其中M为零件类别数。在ResNet18中，将第三个块的特征图作为输入，相应的卷积通道大小为256/64和64 / M。可以看出，与原始骨干网相比，引入的可学习参数的数量很少。

**5.2基线方法**

对于小样本学习算法，我们分别将转移学习，原型网络和动态小样本学习分别表示为转移，原型和动态。 我们在所有学习算法和网络骨干网中，将我们提出的姿势归一化方法（PN）与以下特征聚合方法进行了比较：

平均池化是最简单的方法，通常在以前的工作中采用。 如果未另外指定功能聚合器，则所有后续模型均使用平均池。

我们还提出了一个基线，该基线与定位器一起训练该平均池特征提取器和分类器，而后者在测试时将被丢弃。 这个表示为MT的多任务模型检查姿势估计是否仅在小样本训练中充当正则化函数。

双线性池（BP）[16]是有效的模块，用于扩展潜在特征空间并提高细粒度视觉分类器的表达能力。 最近的工作[29]发现BP可以适应原型网络，在不增加参数数量的情况下提高性能。

Few-shot localization（FSL）[29]在表示和参考集中使用边界框注释。 该模型可在对对象进行分类之前学习对其进行定位，从而提高了小样本分类的准确性。 由于此模型的本地化者是通过原型学习的，因此不会引入任何其他卷积层。

边界框归一化（bbN）与基于边界框的方法进行了更直接的比较，该方法不需要新颖类的框注释。 我们使用PN模型，但是将M设置为2，然后训练定位器根据用于基础类训练数据的地面真值边界框将图像分为前景/背景区域。

无监督姿势归一化（uPN）基于无监督定位[29]，这是一种竞争性定位方法，其中，根据来自一组学习参数向量的特征距离，将特征图划分为多个软区域。 遵循相同的核心思想，我们引入M = 15个学习的，与类别无关的姿势向量，并根据到每个位置的每个向量的相对特征距离在空间上划分特征图。 我们平均合并结果的15个软区域（好像它们是15个预测的零件位置），以生成用于分类器的特征向量。 姿势矢量是学习的参数，经过端到端训练，并与分类器体系结构一起使用，不需要零件注释或单独的定位模块。

此外，我们包括模型的oracle版本：具有地面真实姿势（PN\_gt）的姿势归一化。

**5.3小样本识别结果**

我们首先在上训练所有模型，然后使用验证集为每个模型选择最佳的超参数和停止点。 然后，我们使用中有限的一组标记的新颖类图像在上对它们进行评估。 对于评估指标，我们使用全方位评估[29、12、28]，而不是通常采用的5-way任务。需要算法来同时区分所有新颖类，这是一个更具挑战性的设置。 对于参考图像的数量，我们考虑了[29]提出的标准1-shot/ 5-shot和all-shot设置，即, 应用中每个新颖类别的所有标记图像。

对于CUB，所有样本的结果显示在表1中。对于1和5样本设置，我们在图3和4中绘制了600个随机生成的测试情节的平均值。95％的置信区间均小于0.6个百分点。

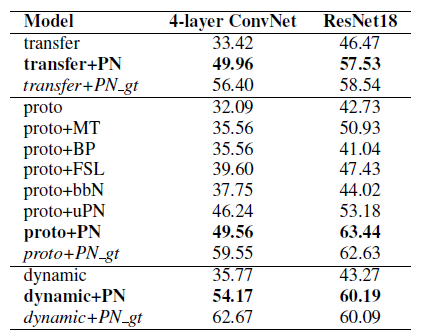


表1. CUB数据集上不同模型的小样本分类结果。 模型是通过小样本学习算法然后通过特征表示方法来组织的。姿势归一化对于浅层网络和深层网络骨干网的所有三种小样本学习算法都显着提高了性能。

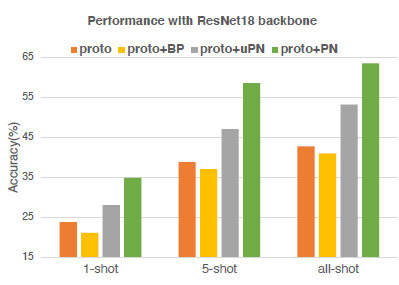


图3. CUB的准确性比较。 所有模型均使用ResNet18原型网络。 在所有设置下，姿势归一化主导其他方法。

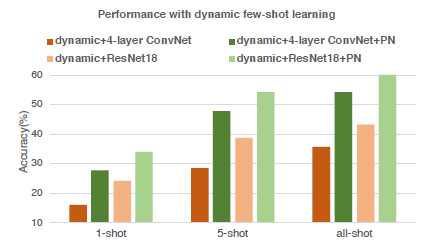


图4. CUB上的动态小样本学习模型的性能比较。 姿势归一化所带来的准确性提升是显着且一致的。

使用以上在CUB上训练的模型，然后使用和对NA进行相同的评估。 NA中的新颖类数量很多（418），并且每个类的图像数量不平衡。 因此，我们仅报告全部类全样本结果于表2中，包括所有测试样本的平均准确度和每个类别的平均准确度。 在上述每种设置中，我们平均对proto，proto + uPN和proto + PN模型进行了8次以上的试验。95％的置信区间均在0.9个百分点以内。

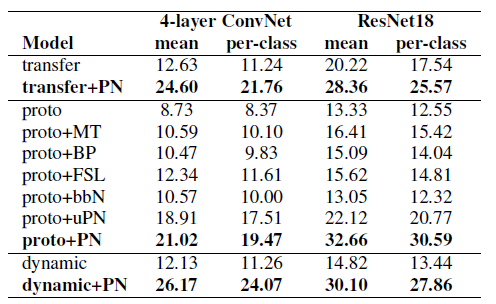


表2. CUB模型在NA上的性能。 在这个新领域中，姿势归一化所带来的性能提升仍然非常重要。 表现与CUB观察结果一致。

从这些实验结果，我们得出以下结论：

1.姿势归一化在（平均池）基准上提供了显着且一致的性能提升。 浅层和深层网络骨干，所有三种小样本学习方法以及两个评估数据集的准确性均得到提高。 在CUB的全类，全样本设置下，在所有三种学习算法中，4层ConvNet的准确度增益始终大于15点，在ResNet18上达到20点。 具有姿势归一化功能的浅层网络甚至可以胜过其深层网络。

2.在所有设置中，姿势归一化均优于其他聚合函数，包括黑盒修改（双线性池），基于边界框定位（FSL和bbN）的技术以及无监督的姿势归一化。 它也胜过多任务训练，这表明关键是标准化而不是额外的辅助损失。

3.姿势信息比粗略的对象定位更有效。 在表1中，PN和bbN贡献了相似数量的新可学习参数，但是PN中细粒度的姿势信息导致其性能优于bbN，后者仅关注于粗糙边界框。 通过将PN与PN gt进行比较，我们发现更好的姿态估计器可能会为性能带来更大的提升。

**5.4姿势注释数量的影响**

尽管零件位置通常比细粒度的专家级标签便宜（请参阅[23]仔细标注的管道），仍然可能难以收集高质量的零件注释。 因此，我们考虑模型的消融，其中有限数量的训练图像带有零件批注。 对于其余图像，不计算Lpose，并且预测姿势热图生成用于分类器训练的特征向量，而不是地面真相。

我们评估具有浅主干和深主干的CUB上的原型网络，并更改带有零件批注的图像百分比。 结果如图5所示。姿势归一化在训练期间对注释稀疏性非常强健（当可用性超过30％时，波动小于5分），并且即使只有5％的姿势注释可用，也始终优于BP。

**5.5 FGVC飞机的评估**

我们在细粒度飞机分类[17]（FGVC-Aircraft）上评估了这些结论的普遍性，该分类包含跨越100个飞机模型的10,000张图像。 按照与CUB相同的比率，我们将类别分为50个基本类别，25个验证类别和25个新颖类别。 引用/查询拆分如5.1节所述。 由于此数据集不包含任何零件批注，因此我们使用独立的数据集OID-Aircraft [24]（OID）来联合训练我们的姿势归一化模块。 OID包含6357个图像（忽略了与FGVC共享的图像），以及每个图像5个零件批注（因此M = 5）。 OID不包含分类标签。

每次训练迭代都从OID和FGVC采样图像批次。 OID图像用于计算Lpose，而FGVC图像使用预测的姿势热图来获取特征向量。 结果显示在表3中。尽管姿态估计器是在不相交的图像上训练的，但它仍然可以有效地提高飞机的识别性能。 我们得出的结论是，姿势归一化适用于细粒度的几次镜头分类任务。 将这种方法扩展到非细粒度的任务或特定于类的部分并不是一件容易的事，但可能是将来研究的重要方向。

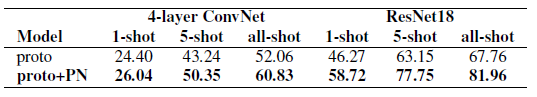


表3. FGVC-飞机数据集上所有三个评估设置下的少量结果。 结果平均超过五次试验。

**6.分析**

**6.1模型解释**

尽管具有准确性，但我们希望姿势归一化表示能够被人类解释，这与先前的黑匣子表示不同。 为了调查模型实际学习的内容，我们进行了两个实验来分析学习的姿势归一化表示。 两者都使用带有ResNet18主干网的proto + PN模型。

零件重要性：每种鸟类都可能具有一组特别可区分的零件属性。 为了验证我们的模型是否学到了这一点，我们进行了以下测试。对于每个类别，当从表示中删除相应的零件特征向量时，我们将迭代零件并计算测试精度。 准确度下降的幅度可以解释为模型学到的每个零件对于此类的重要性。 我们在图6中形象地了解了这对三种物种的重要性，并将其与现场指南中的物种描述进行了比较。 我们的网络分数在很大程度上符合专家的判断。

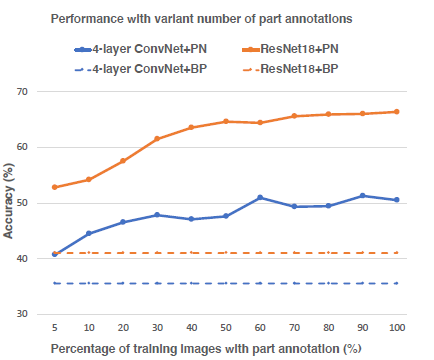


图5.零件注释稀疏时，姿势归一化的少量测试精度。 性能下降非常小。 即使在训练期间只有5％的注释可用，姿势归一化也比双线性合并更好。

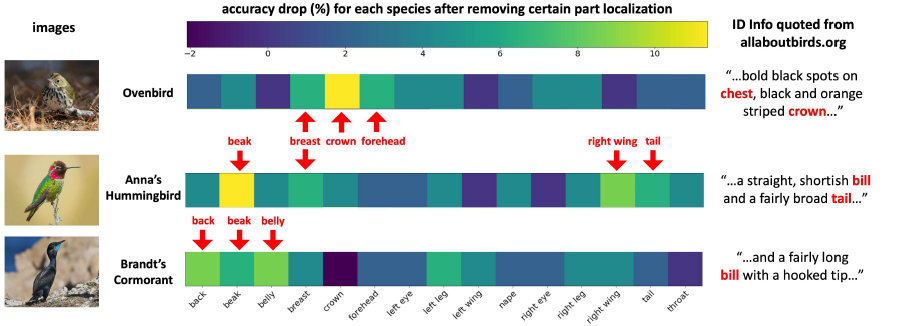


图6.去除单个零件矢量（零件重要性）后，可视化所选鸟类的准确度下降。 右边引用了鸟类专家关于如何识别这些物种的描述。 估计的零件重要性与专家的判断非常吻合。

最近的邻居：不同的鸟类可能共享相同的零件属性； 例如，加利福尼亚鸥和环嘴鸥具有相同的喙形。 因此，在姿势归一化中，这两只鸟的喙向量应接近，因为部分向量被设计为以与类无关的方式编码区域信息。 为了验证这一点，我们在参考集中找到了与查询图像/部分对中给定向量最接近的部分向量的前5个图像。图7中给出了四个随机示例。通常，我们的假设成立-描述每个查询图像中给定部分的向量确实可以推广到其他物种。

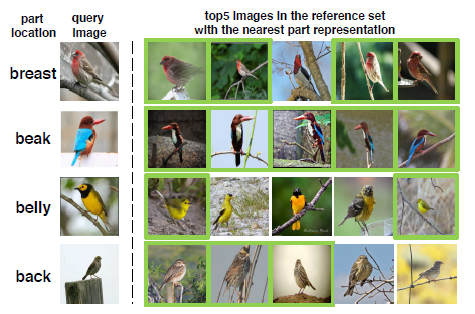


图7.对于给定的零件位置，零件向量与查询图像最接近的图像。 如果图像与查询图像属于同一类别，则图像用绿色框标记。 我们看到零件表示捕获了跨类的零件位置的语义上有意义的属性。

**6.2姿态估算**

在评估姿势估计[30，2]的先前工作之后，我们针对浅层网络骨干和深层网络骨干在不同的阈值处计算归一化PCK（使用边界框的对角线归一化）。 结果如图8所示。我们看到这两个估计都可以得出准确的结果。 尽管更深的网络主干确实可以提供更好的估计，但是这种提升也非常有限。我们认为，更复杂的姿势估计器可以在小样本识别中获得更好的结果。

**6.3无监督姿势归一化**

我们注意到，从分类的角度来看，无监督姿势归一化也表现良好（参见表1）。如图9所示，具有无监督姿势归一化的更深的主干确实会产生局部关键点，这可能有助于分类。 但是，请注意，这些关键点的语义跨图像含义不一致（图9，前两行）。 预测也是不稳定的，有时不同的渠道会提供相似的热图（图9，底行）。 这种不一致之处可以帮助解释为什么在细粒度，小样本的分类中，机器发现的零件不如手工设计的零件好。

**7.结论**

我们展示了一个简单，轻巧的姿势归一化模块可以在测试时无需任何部件注释的情况下，在细粒度的几次拍摄识别中始终带来巨大的性能提升。我们的结果适用于浅层和深层网络骨干网，多个一次性学习算法和多个域。除了显着提高准确性外，我们还证明了姿势归一化表示是人类可以高度解释的。因此，我们强烈建议将姿势标准化作为细粒度的少量学习社区重新审视的一般区域。