**Boosting Few-Shot Visual Learning with Self-Supervision**

会议：ICCV2019

代码：<https://github.com/valeoai/BF3S>

摘要

小样本学习和自我监督的学习解决了同一问题的不同方面：如何训练几乎没有标签数据或没有标签数据的模型。 小样本学习的目标是优化方法和模型，这些方法和模型可以有效地学习以识别低数据状态下的模式。 自我监督学习则将重点放在未标记的数据上，并寻找监督信号以馈入高容量深层神经网络。 在这项工作中，我们利用了这两个领域的互补性，并提出了一种通过自我监督来改善小样本学习的方法。 我们将自我监督作为小样本学习管道中的辅助任务，使特征提取器可以在只使用少数带注释的样本的情况下，学习更丰富，更易传递的视觉表示。 通过自我监督，我们的方法可以自然扩展到在小样本中使用来自其他数据集的各种未标记数据。 我们报告了一系列架构，数据集和自我监督技术的持续改进。

**1.简介**

基于深度学习的方法在各种图像理解任务上取得了令人印象深刻的结果，例如图像分类[22、29、52、54]，对象检测[47]或语义分割[4]。 但是，要成功学习这些任务，卷积神经网络（convnet）需要访问大量手动标记的训练数据。如果不是这样，经过训练的卷积网络可能会遭受测试数据泛化性能不佳的困扰。 相比之下，人类完全有能力从几个示例中学习新的视觉概念，毫不费力地将其推广到新数据。小样本学习[8，9，26，30，55]旨在赋予人造感知系统类似的能力，尤其是在现代深度学习的帮助下。

因此，尽管视觉训练很少（例如，每个类只有1或5个示例），但小样本视觉学习的目标是设计能够有效学习识别一组课程的识别模型。 为了避免由于数据稀缺而导致的过度拟合，小样本学习算法依赖于转移学习技术，并具有两个学习阶段。 在第一阶段，通常使用一组不同的类（称为基类）来训练模型，该类与大量带注释的训练示例相关联。 这个阶段的目标是让小样本模型获得可转移的视觉分析能力，通常以学习的表示形式获得，在第二阶段中动员起来。 在下一步中，该模型确实学会了识别新颖的类，而这是在第一个学习阶段看不到的，每个类仅使用几个训练示例。

小样本学习涉及自我监督的表示学习[6、7、14、31、40、61]。 后者是一种无监督学习的形式，可在仅使用图像中存在的视觉信息定义的无注释借口任务上训练模型。 这项自我监督任务的目的是使模型学习图像表示形式，这将在转移到其他图像理解任务时很有用。 例如，在Doersch[6]等人的开创性工作中，通过在预测图像补丁相对位置的自我监督任务上进行训练，网络设法学习成功地转移到对象识别，对象检测和语义分割的视觉任务的图像表示。 因此，就像在小样本学习中一样，自我监督方法也有两个学习阶段，第一个学习具有借口自我监督任务的图像表示，第二个使这些表示适应感兴趣的实际任务。 同样，两种学习方法都试图限制深度学习方法对大量标记数据可用性的依赖。

受小样本学习和自我监督学习之间联系的启发，我们建议将两种方法结合起来以提高小样本学习模型的转移学习能力。 更具体地说，我们将自我监督的损失添加到训练损失中，这种损失是小样本模型在其第一个学习阶段就最小化的（见图1）。 因此，我们在此阶段人为地增加了小样本模型可以解决的训练任务。 我们认为，这种任务增强会迫使模型学习更多样化的图像特征集，从而反过来提高了模型的适应能力，而无需训练数据。 此外，由于自我监督不需要数据注释，因此可以在第一个学习阶段添加额外的未标记数据。通过这种方式扩展训练数据大小和种类，人们可能希望学习更强大的图像功能，并通过小样本学习获得进一步的性能提升。 在极端情况下，在第一个学习阶段仅使用未标记的数据，从而完全取消对基类的使用，也很有吸引力。 我们将证明，由于自我监督的任务，这些半监督和非监督体制的确可以用于小样本物体识别。

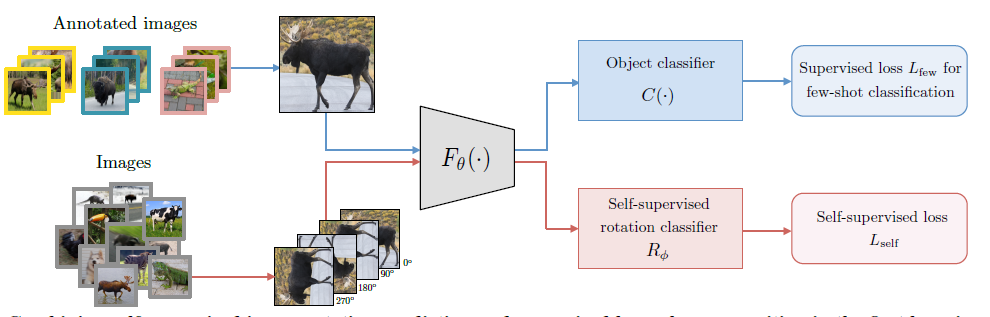


图1：在小样本系统的第一个学习阶段中，将自我监督的图像旋转预测与监督的基类识别相结合。 我们在多任务设置中使用带注释的数据（顶部分支）和未带注释的数据（底部分支）训练特征提取器F（·）。 我们使用带注释的数据来训练具有小样本镜头分类损失Lfew的对象分类器C（·）。 对于自我监督任务，我们从带注释的集合（以及可选地从另一组非带注释的图像）中采样图像。 在这里，我们为每个输入图像生成四个旋转，用F（·）处理它们，并用自监督损失Lself训练旋转分类器R。相对补丁程序位置自我监督的管道与此类似。

总而言之，我们的工作主要有以下几个方面：

（1）我们建议将自我监督融入到小样本学习算法的训练目标中。 目的是通过较少的培训数据来提高后者适应新颖课程的能力。

（2）我们通过对Mini-Imagenet，CIFAR-FS，tiered-Mini-Imagenet和ImageNetFS小样本数据集进行详尽的定量实验来研究增加的自我监督损失的影响。 在所有这些方面，自我监督可以改善小样本学习的成绩，从而获得最新的成果。

（3）最后，我们将拟议的小样本识别框架扩展到半监督和无监督的设置，在前者中获得了进一步的性能提升，并向后者展示了我们的框架提供了一种协议，用于评估在小样本对象上的自我监督表示识别非常简单快捷。

**2.相关工作**

小样本学习。 有大量的小样本学习方法，其中包括：基于梯度下降的方法[1，10，39，45]，它们学习如何通过模型快速使模型适应给定的小样本识别任务。 梯度下降迭代次数少； 基于度量学习的方法，用于学习小样本任务的查询（即测试图像）与支持图像（即训练图像）之间的距离度量[26、33、53、55、57、59]； 通过访问存储用于该任务的训练示例的内存模块来学习将测试示例映射到类标签的方法[11、25、35、38、50]； 在给定的训练数据很少的情况下，学习如何为每个新类别生成分类器[12、16、43、44、13]或多层神经网络[3、18、19、58]的权重的方法; 从减少的数据量中“半透明”类的其他示例的方法[21，57]。

在我们的工作中，我们考虑了度量学习类别中的两种方法，即原型网络[53]和余弦分类器[12，43]，因为它们既简单又灵活。然而，所提出的辅助自我监督与其他几种小样本类法兼容。

自我监督学习。 它是无监督学习的最新范例，它定义了无注释的借口任务，以为特征学习提供代理监督信号。预测图像的颜色[31、61]，图像块的相对位置[6、40]，已应用于图像[14]的随机旋转或图像[42]的缺失中心部分是用于自我监督特征学习的许多方法[15、32、36、56、62]中的某些方法。 直觉是，通过解决此类任务，训练后的模型提取了可用于其他下游任务的语义特征。 在我们的案例中，我们考虑了一个多任务设置，其中我们通过受监督的最终任务和辅助的自监督借口任务的联合监督来训练主干卷积网络。与大多数多任务设置旨在同时在所有任务上取得良好效果[27]不同，我们的目标是通过利用代理任务的监督来提高主要任务的性能，如[37]所示。 我们希望在小样本情况下，从可用数据中挤出可概括的功能非常重要，将自我监督用作辅助任务将带来改善。另外，与我们的工作有关，Chen等[5]最近在生成的对抗网络[17]中添加了旋转预测自我监督功能，从而显着提高了合成图像的质量。

**3.方法**

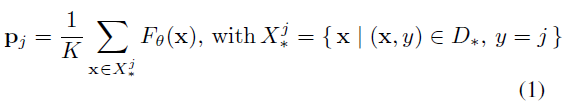
如前所述，小样本学习算法具有两个学习阶段和两个相应的类集。在这里，我们将在第一个学习阶段使用的基类训练集定义为，其中x∈I是大小为Nb的标签集Yb中带有标签y的图像。另外，我们将第二学习阶段使用的Nn个新颖类的训练集定义为Dn = {（x，y）}⊂I×Yn，其中每个类有K个样本（基准中K = 1或5）。有人谈论N-way K-shot学习。 重要的是，标签集Yn和Yb不相交。在本节的其余部分中，我们首先在第3.1节中介绍我们考虑的两种标准的一次性学习方法，并在第3.2节中介绍建议的方法以通过自我监督提高其性能。

在本节的其余部分中，我们首先在第3.1节中介绍我们考虑的两种标准的一次性学习方法，并在第3.2节中介绍建议的方法以通过自我监督提高其性能。

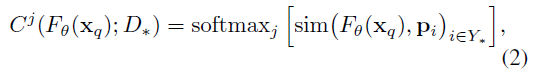
**3.1. Explored** **few-shot learning methods**

所有few-shot算法的主要组成部分是特征提取器Fθ（·），这是一个带参数的卷积网络。 给定图像x，特征提取器将输出d维特征Fθ（x）。 在这项工作中，我们考虑了以下两种代表性的少拍算法：原型网络（PN）[53]和余弦分类器（CC）[12、43]。 它们非常相似，主要区别在于第一个学习阶段：只有CC会与特征提取器一起学习实际的基础分类器，而PN仅依赖于类级别的平均值。

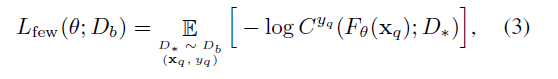
原型网络（PN）[53]。 在此方法的第一阶段，特征提取器Fθ（·）是在采样的小样本分类子问题上学习的，这些子问题类似于目标小样本分类问题。 在此学习阶段的每个训练阶段中，都会对N \*个基础类的子集Y ∗⊂Yb进行采样（它们称为“支持类”），并且针对每个样本，从Db中随机选取K个训练示例。 这产生训练集D \*。 给定当前特征提取器Fθ，每个类j∈Y ∗（其“原型”）的平均特征计算如下：



并用于构建简单的基于相似度的分类器。 然后，给定来自支持类但与D ∗中的样本不同的新图像xq，分类器为每个类j输出归一化分类得分

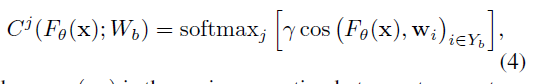


其中sim（·，·）是相似度函数，可以是余弦相似度或负平方欧几里得距离。 因此，实际上，图像xq将被分类为其最接近的原型。 请注意，分类器以D ∗为条件，以计算类原型。 第一学习阶段最终等于迭代地最小化后续损失w.r.t.  ：



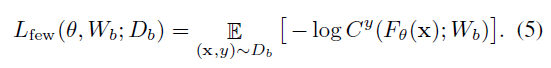
其中（xq，yq）是D ∗中定义的支持类的训练样本，但不同于D ∗中的图像。 在第二学习阶段，特征提取器Fθ被冻结，新颖类的分类器被简单地定义为，原型的定义如（1）中D \* = Dn。

余弦分类器（CC）[12，43]。 在CC小样本学习中，第一阶段在对基础类进行分类的（标准）监督任务下，将特征提取器Fθ与基于余弦相似度的分类器一起训练。 表示Wb = [w1，...，wNb] d维分类权向量的矩阵，输入图像x的归一化分数读取



其中cos（·，·）是两个向量之间的余弦运算，标量是softmax运算符的逆温度参数。具体来说，控制softmax运算符生成的概率分布的峰值

第一个学习阶段旨在最小化w.r.t.  和Wb负对数似然损失为：



使用基于余弦相似度的分类器而不是基于标准点积的分类器的原因之一是，前者学习的特征提取器减少了类内差异，因此可以更好地归纳到新颖的类上。 类似于PN，权重向量wj可以解释为基类的学习原型，将输入图像特征与基类进行比较以进行分类。

与PN一样，第二阶段归结为通过对Dn中关联的K个样本进行简单平均来为每个新类别计算一个代表特征wj，并以相同的方式定义最终分类器C（.; [w1···wNn]） 如（4）中所示。

**3.2 通过自我监督促进****小样本学习**

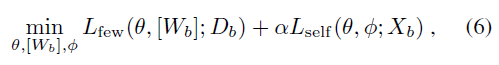
在学习的第一阶段，小样本学习会遇到重大挑战。 在第二阶段，如何使特征提取器学习可用于很少训练数据的新颖类的图像特征？ 考虑到这一目标，我们建议利用自我监督型特征学习的最新进展来进一步改进当前的小样本学习方法。

通过解决可以平凡监督的非平凡的借口任务，例如从图像强度中恢复图像的颜色，鼓励网络学习丰富的通用图像特征，这些特征可以转移到其他下游任务，例如图像分类。 在第一个学习阶段，我们建议通过包括这种自我监督任务以及识别基类的主要任务来扩展特征提取器Fθ（。）的训练。

我们考虑了两种将自我监督结合到少量学习算法中的方法：（1）通过使用基于自我监督任务的辅助损失函数，以及（2）在训练过程中以半监督方式利用未标记的数据。 我们在下面描述这两种技术。

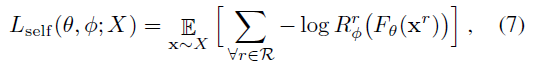
**3.2.1基于自我监督的辅助损失**

通过在第一个学习阶段添加辅助的自我监督损失，我们将自我监督合并到小样本学习算法中。 更正式地说，让是应用于集合Xb = {x | （x，y）∈Db}中被剥夺了类别标签的训练示例。 损失是特征提取器的参数和仅专用于自我监督任务的网络的参数的函数。 小样本学习的第一个训练阶段现为



其中Lfew代表PN小样本loss（3）或CC小样本loss（5），在后一种情况下带有附加参数Wb（因此使用括号表示法）。 正超参数控制自我监督项的重要性。 图1提供了该方法的说明。

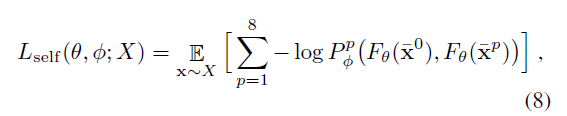
图像旋转。 在此任务中，卷积必须在R = {0°，90°，180°，270°}的四个可能的2D旋转中识别出一个应用于图像的旋转（请参见图1）。 具体来说，给定图像x，我们首先创建其四个旋转副本{x r | r∈R}，其中x r是旋转了r度的图像x。 基于从这种旋转图像中提取的特征Fθ（x r），新网络Rφ尝试预测旋转类别r。 因此，此任务的自我监督损失定义为：



其中X是非旋转图像的原始训练集，Rr（·）是旋转r的预测归一化分数。 直观地讲，为了很好地完成此任务，模型应减少对像ImageNet数据集所特有的朝上的图像的偏见，并学习更多的功能以区分低数据类型中的类。

相对补丁程序位置。 在这里，我们根据图像创建随机的补丁对，然后在八个可能的位置中预测第二个补丁的位置。 到第一个，例如“在左侧和上方”或“在右侧和下方”。具体来说，给定图像x，我们首先将其划分为3×3网格上的9个区域，并在每个区域内采样一个补丁。让我们将x 0表示为中心图像块，将x 1···x 8表示为按字典顺序排列的八个邻居。 我们计算每个补丁的表示，然后通过级联生成补丁特征对。 我们训练一个完全连接的网络，以预测每对对的位置。

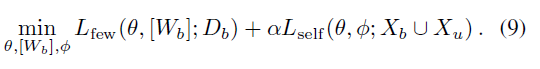
此任务的自我监督丢失定义为：



其中X是一组图像，是相对位置p的预测归一化分数。 凭直觉，即使在存在遮挡和背景杂波的情况下，执行此任务的良好模型也应在某种程度上识别对象和零件。请注意，为了防止模型学习诸如色差[6]之类的低级图像统计信息，请使用积极的色彩增强功能对色块进行预处理（即，以0.66的概率转换为灰度并将每个色块的像素分别归一化为 均值和单位标准偏差为零）。

**3.2.2半监督小样本学习**

训练损失（6）中的自我监督项Lself不依赖于类别标签。 我们可以轻松地将其扩展为从其他未标记的数据中学习。 的确，如果除了Db之外还有一组未标记图像的Xu，我们可以通过将第一个学习阶段重新定义为以下内容，使自我监督的任务从中受益：



通过训练特征提取器Fθ来最大程度地减少这些未标记图像上的自我监督损失，我们打开了它的视觉范围，希望这将进一步提高其容纳稀缺数据的新颖类的能力。这可以看作是对小样本方法的半监督训练方法。这种半监督训练方法的一个有趣的方面是，它不需要多余的未标记数据来自与标记数据集Db中相同的（基本）类。 因此，它具有更大的灵活性。 未标记数据的来源比标准的半监督方法更重要。

**4.实验结果**

在本节中，我们将第4.2节和第4.3节中的自我监督评估为辅助损失函数，作为在半监督训练中利用未标记数据的一种方式。 最后，在第4.4节中，我们使用几次快照对象识别任务来评估自我监督方法。 我们在附录A和附录B中作为补充材料提供了更多的实验结果以及实施细节。

数据集。 我们对四个小样本图像数据集进行了实验，这些图像是MiniImageNet [55]，分层MiniImageNet [46]，CIFAR-FS [2]和ImageNet-FS [20]。 MiniImageNet包含100个从ImageNet数据集[48]中随机选择的类（即64个基本类，16个验证类和20个新颖的测试类）； 每个类别有600张尺寸为84×84像素的图像。 tiered-MiniImageNet包含从ImageNet中随机选择的608个类（即351个基本类，97个验证类和160个新颖的测试类）； 总共总共有779张，165张图像，尺寸为84×84。最后，ImageNet-FS是基于原始ImageNet的一些基准测试，具有389个基类，300个新颖的验证类和311个新颖的测试类。 CIFAR-FS是通过将100个CIFAR-100类划分为64个基本类，16个验证类和20个新颖的测试类而创建的小样本数据集。 该数据集中的图像尺寸为32×32像素。

评估指标。小样本的分类算法会在其第二学习阶段（当将学习的分类器应用于来自新颖类的测试图像时）基于分类准确性进行评估。 更具体地说，从一组可用的新颖类中采样了大量的N-way K-shot任务。 通过从可用的测试（验证）类中随机选择Nn个新颖类，然后在所选类中随机选择每类K个训练图像和M个测试图像（确保训练图像和测试图像不重叠）来创建每个任务。 分类准确率是在Nn×M张测试图像上测得的，并在所有采样的小样本任务中取平均值。 除非另有说明，否则对于所有实验，我们使用M = 15，Nn = 5，K = 1或K = 5（分别为1次和5次设置）。ImageNet-FS是一个更具挑战性的基准：它同时使用所有Nn = 311个新类，根据新类和基类的分类准确性，对ImageNet验证集上的方法进行评估。

**4.1实现细节**

网络架构。 我们使用四种不同的特征提取器架构Fθ：Conv-4-64 [55]，Conv-4-512，WRN28-10 [60]和ResNet10 [22]。 特定于旋转预测任务的网络Rφ（·）作为输入获得Fθ的输出特征图，并实现为卷积网络。 给定两个补丁，特定于相对补丁位置任务的网络Pφ（·，·）获取以Fθ作为输入提取的特征向量的级联，并将其转发到两个完全连接的层。 有关更多详细信息，请参见附录B.1。

在第一个学习阶段进行培训优化程序。我们使用小批量随机梯度下降（SGD）优化训练损失。 对于标记的数据，我们同时应用识别Lfew和自我监督Lself损失。 对于半监督训练，在每个步骤中，我们对包含标记数据（使用损失和未标记数据）组成的微型批次进行采样，在这种情况下，我们仅应用Lself。有关更多详细信息，请参见附录B.2和B.3。

执行相对补丁定位任务。 由于补丁定位任务中补丁的色彩激增，而且补丁的大小比原始图像小9倍，因此特征提取器从其中“看到”的数据图片分布与原始图像的分布非常不同。为了克服这个问题，我们对从补丁中提取的特征施加了额外的辅助分类损失。 具体来说，在CC的第一个学习阶段，我们将图像的9个面片的特征Fθ（¯xp）合并（例如，通过级联或平均），然后将余弦分类器（4）应用于结果特征（该分类器 不能与应用于原始图像特征的分类器共享其权重向量）。 注意，该基于补丁的辅助分类损失具有与原始分类损失Lfew相同的权重。 同样，在第二学习阶段，我们不使用基于补丁的分类器。

**4.2 自监督作为辅助损失函数**

旋转预测作为辅助损耗。 我们首先研究了将旋转预测作为自我监督对余弦分类器（CC）和原型网络（PN）的小样本学习算法的影响。 我们使用MiniImageNet和CIFAR-FS数据集进行了这项研究，并在表1和表2中分别报告了结果。 对于CC情况，我们使用无自我监督但经过训练可识别图像的所有4个旋转版本的CC模型作为强基准。使用该基线的原因是，在第一个学习阶段，模型“看到”与旋转预测自我监督模型相同类型的数据，即旋转图像。 请注意，尽管第一个学习阶段的轮换人数有所增加，但在第二个阶段，该模型仅用作小说类的训练示例图像的正立版本。 但是，当适应新类时，使用旋转增强可以改善基线模型的分类性能。 因此，为了进行公平的比较，我们还将旋转增强应用于具有旋转预测自我监督的CC模型。 对于PN情况，我们不使用旋转增强，因为在我们的实验中这会导致性能下降。

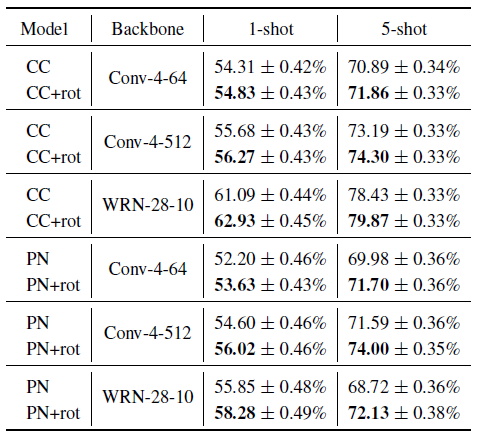


表1：在MiniImageNet上作为辅助损耗的旋转预测。具有95％置信区间（使用2000集）的新类的平均5种分类准确性。

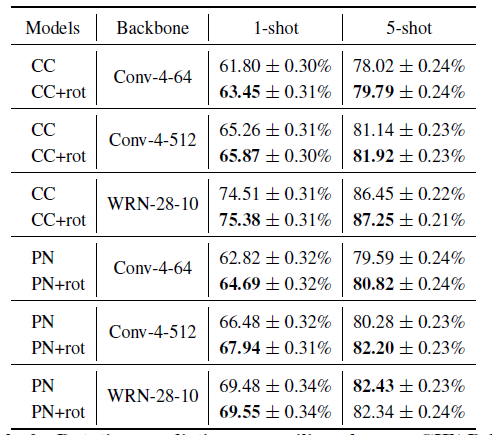


表2：作为CIFAR-FS辅助损耗的旋转预测。具有95％置信区间（使用5000集）的新颖类的平均5种分类准确性。

表1和表2中的结果表明（1）的确，增加旋转预测自我监督可以改善镜头分类性能，并且（2）对于WRN-28-10等高容量架构，性能提高更为显着。 在表3中，我们使用ResNet10主干网和CC提供了具有挑战性的ImageNet-FS [20]基准测试的结果。 我们的方法在识别新类，以及同时识别新颖和基本类（全部）上有挺高。

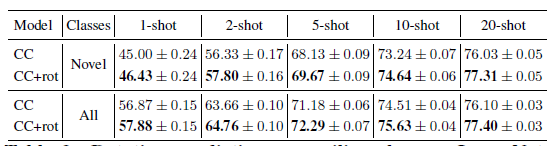


表3：旋转预测作为ImageNetFS上的辅助损耗[20]。 ResNet10骨干网的前5位准确性。 CC的基本类的准确性为92.91％，CC + rot的基本类为93.52％。对于这些结果，我们没有使用旋转增强。

相对斑块位置预测作为辅助损耗。 如第3.2.1节所述，我们考虑了第二项自我监督任务，即相对补丁位置预测。 为简单起见，我们将其评估仅限于针对Conv-4-64和Conv-4-512骨干网的CC算法，在我们的实验中，这些算法被证明比PN性能更好，并且更易于训练。 另外，我们仅考虑MiniImageNet数据集，而不考虑CIFAR-FS，因为后者包含大小为32×32的缩略图，从中提取补丁没有意义：它们的大小必须小于8×8像素，即 对于评估的架构而言太小了。 我们在表4中报告了MiniImageNet的结果。作为强有力的基线，我们使用了无自我监督的CC模型，但有第4.1节中所述的基于补丁的辅助分类损失。

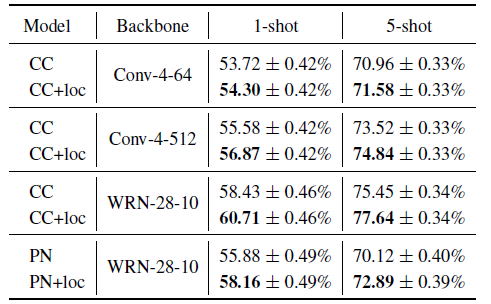


表4：MiniImageNet上作为辅助损耗的相对补丁位置。具有95％置信区间（使用2000集）的小说类的平均5种分类准确性。

根据表4的结果，我们观察到：（1）相对补丁位置也设法提高了一次性拍摄分类的性能，并且，如在轮播预测的情况下，这种改进对于高容量网络体系结构更为重要。 （2）此外，与旋转预测情况相比，相对面片位置通常可提供较小的性能改进。

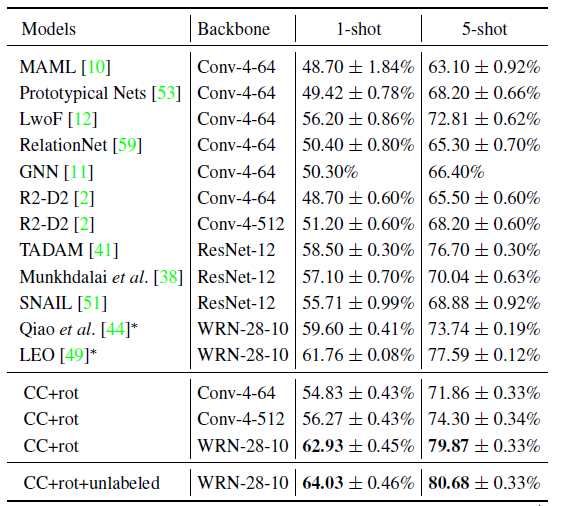


表5：与MiniImageNet上先前工作的比较。 ：还使用验证类进行培训。 有关CC + rot +未标记模型的说明，请参见第4.3节。

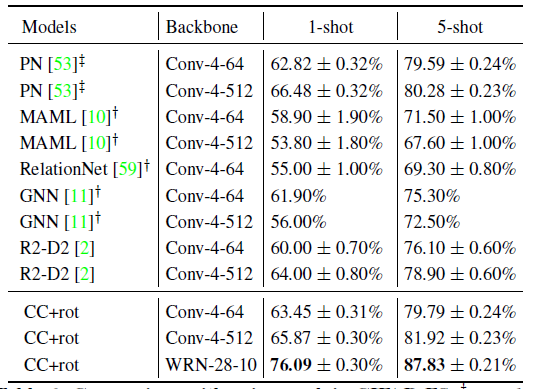


表6：与CIFAR-FS中先前工作的比较。 †：[2]的结果。 ‡：我们的执行情况。

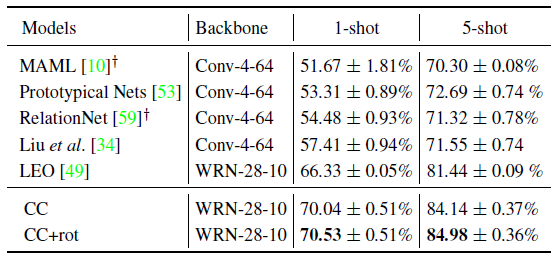
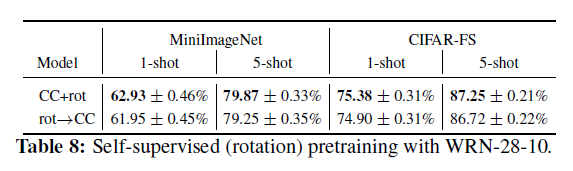


表7：旋转预测作为tiered-MiniImageNet上的辅助损耗。 新类的平均5方向准确性为95％（使用2000集）。 †：[34]的结果。

与先前工作的比较。 在表5、6和7中，我们分别将我们的方法与MiniImageNet，CIFAR-FS和分层MiniImageNet数据集上的先前的小样本方法进行了比较。 对于我们的方法，我们使用CC和旋转预测自我监督，前者取得较好的效果。在所有情况下，我们都可以以远远超过现有方法的方式获得最先进的结果。 例如，在MiniImageNet的1拍和5拍设置中，我们的性能优于以前的领先方法LEO [49] 分别约为1.2和2.3个百分点。



与自我监督的预训练进行比较。 在小样本学习中利用自我监督的另一种方法是对小样本模型的特征提取器进行自我超负荷预训练。 在表8中，我们使用高容量WRN-28-10主干为CC和旋转预测任务（条目“ rot→CC”）提供了此类结果。 我们观察到，与所提倡的方法相比，它提供的改进较小。

附录A中提供了更详细的结果。

**4.3。 半监督的小样本学习**

接下来，我们评估提出的半监督训练方法。 在这些实验中，我们使用带有旋转预测自我监督的CC模型。 我们执行两种类型的半监督实验：（1）使用来自相同基类的未标记数据进行训练，以及（2）使用不是来自基类的未标记数据进行训练。

训练来自相同基类的未标记数据。从MiniImageNet的基类中，我们仅将训练图像的一小部分（例如，每个类别的图像的5％）用作带注释的训练数据，而将其余图像（例如95％）用作未标记的数据 在半监督训练中。 我们在表9的前两个部分中提供了结果。将所建议的半监督训练方法与没有自我监督的CC模型和有自我监督但没有求助于未标记数据的CC模型进行了比较。 结果表明，确实，我们的方法通过利用未标记的图像设法提高了小样本分类性能。与任等人相比。 [46]也提出了一种半监督方法，我们使用Conv-4-64和20％注释的方法比使用Conv-4-64和40％注释的方法（即，我们的51.21％和68.89％ MiniImageNet的准确度与1次和5次设置的50.41％和64.39％相比）。

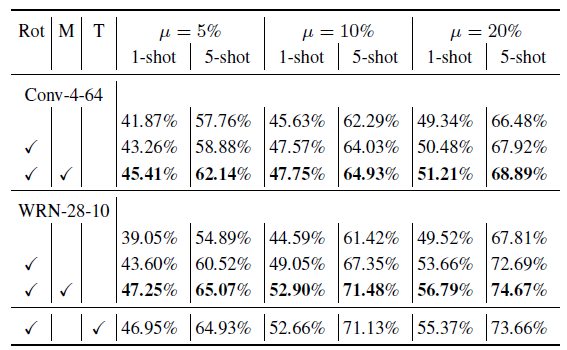


表9：在MiniImageNet上以旋转预测作为自我监督的半监督训练。μ是MiniImageNet的基类训练图像在训练期间用作注释数据的百分比。Rot表示添加自我监督，M表示使用（其余）MiniImageNet训练数据集作为未标记的数据，T表示使用分层的MiniImageNet训练数据集作为未标记的数据。

使用非标签数据（不是来自基类）进行训练。这是一个更现实的设置，因为很难将未标记的图像限制为与基类相同。 对于本实验，我们使用分层的MiniImageNet基类的训练图像减去与MiniImageNet的基类，验证或测试类通用的类作为未标记的数据。 总共使用了分层MiniImageNet中的408、726张未标记图像。报告结果位于表9的最后一行。确实，即使在这种困难的情况下，我们的半监督方法仍然能够利用未标记的数据并提高分类性能。 此外，我们进行了额外的实验，其中我们使用100％的MiniImageNet训练图像和来自分层MiniImageNet的未标记数据来训练基于WRN-28-10的模型。 该模型在MiniImageNet上的1次和5次设置分别达到了64.03％和80.68％的准确度（请参见表5的条目“ CC + rot + unlabeled”），相对于已经非常强大的CC + rot模型（请参见 表5）。

**4.4 Few-shot对象识别可评估自我监督的表示形式**

鉴于我们的框架允许将任何类型的自我监督学习方法与所采用的小样本学习算法轻松组合，我们还建议将其用作比较/评估不同自我监督方法有效性的替代方法。 为此，我们框架的唯一所需更改是在第一个学习阶段中唯一使用自学习损失（即，现在不使用任何已标记的数据）（有关实现的详细信息，请参阅附录B.2）。 然后，可以将第二学习阶段获得的小样本模型的性能用于评估所考虑的自我监督方法。

比较竞争的自我监督技术并不是一件容易的事，因为必须通过设置另一个利用人为学习的表示的人为设计任务来完成[6，28]。 取而代之的是，鉴于小样本和自我监督学习的目标非常相似，我们认为所提出的比较方法对于评估不同的自我监督特征可能更有意义。 此外，（1）它获得了更强大的统计数据，这些统计数据是通过对具有数千种不同类别的课程和训练/测试样本的情节进行评估得出的；（2）速度非常快（只需几分钟即可评估），（3） 与其他协议相比，它要简单得多（无需使体系结构适应新任务或微调参数，因此无需超参数调整）。 这些优点可能使该协议在自我监督研究中实际上有用。

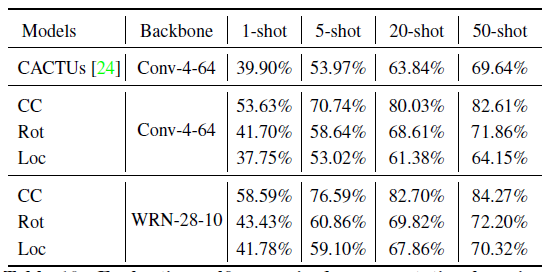


表10：在MiniImageNet上的几次快照识别中评估自我监督的表示学习方法。 Rot表示旋转预测任务，Loc表示相对面片定位任务，CC表示余弦分类器的监督方法。

为了说明我们的观点，我们在表10中提供了MiniImageNet数据集上此类评估的定量结果，用于旋转预测和相对面片位置预测的自监督方法。 对于自我监督的训练，我们使用MiniImageNet的基类的训练图像，而针对小样本的分类步骤，我们使用MiniImageNet的测试类。 我们观察到，与CC的监督方法相比，探索的自我监督方法实现了相对竞争的分类性能，并获得了与其他更复杂，无监督的系统同等或更好的结果。 在以后的工作中，我们将在此评估环境中对自学表示法进行更详细，更彻底的比较。

**5.结论**

受小样本和自我监督学习之间紧密联系的启发，我们建议在小样本识别模型的训练过程中增加基于自我监督的辅助损失。 目标是仅使用少量训练数据来提高后者识别新颖课程的能力。 我们在MiniImagenet，CIFAR-FS，tiered-MiniImagenet和ImageNet-FS小样本数据集上进行的详细实验表明，确实增加了自我监督可以显着改善少拍分类性能，从而使所采用的少拍模型得以实现 最新的结果。 此外，自我监督损失的无注释性质使我们能够以半监督方式利用各种未标记数据，从而进一步提高了分类性能。 最后，我们证明了所提出的框架还可以用于评估基于小样本物体识别的自我监督或非监督方法。