**Adaptive Subspaces for Few-Shot Learning**

会议：CVPR2020

代码：<https://github.com/chrysts/dsn_fewshot>

**摘要**

对象识别要求具有泛化能力，以避免过度拟合，尤其是在样本极少的情况下。 通常在元学习的框架下对有限样本进行概括，使学习技术具有在动态环境中快速适应的能力，并被证明是终身学习的重要方面。 在本文中，我们通过介绍由少量样本构成的动态分类器，为小样本学习提供了一个框架。 子空间方法被用作动态分类器的中心块。 我们将凭经验证明，这种建模方法可以抗干扰（例如，离群值），并且在监督和半监督的小样本分类任务上产生竞争性结果。 我们还开发了一种判别形式，可以进一步提高准确性。

**1.简介**

各种研究表明，计算机视觉，语音识别和自然语言理解等许多深度学习技术如果可用的注释数量有限，将无法生成可靠的，可以很好地泛化的模型。 除了与注释数据相关的工作外，在某些情况下，精确的注释可能会变得不适和当前任务。这种困难的一个主要例子是对象检测标签，它需要注释对象的边界框，如[1]中所述。 在其他一些情况下，标记过程可能需要专业知识（例如手语识别[2]）。

与当前的深度学习趋势相反，人类只能从几个例子中学习新的对象。 反过来，这为人类提供了终身学习能力。 受这种学习能力的启发，开发了几种方法来研究有限样本的学习[3-12]。 这类学习被称为“小样本学习”（FSL），它已经通过嵌入学习的各种思路得以解决[4，13、14]，适应技术[7、8]甚至生成模型[3、15]。

在这项工作中，我们首先将FSL公式化为两个阶段的学习范式，即**1**.学习通用特征提取器，然后**2**.学习根据有限的数据动态的生成分类器。 我们将证明，许多最新的FSL技术都非常适合这种学习范例。 此外，我们将证明将FSL视为上述范例将是有益的，并为我们提供了将FSL形式化的工具。

一旦建立了两阶段学习范式，我们将把注意力转向如何从有限的数据可靠地生成分类器。 除了有限的注释外，我们还将显示许多富挑战性的FSL问题中的要求是从高维数据中学习分类器。最终归结为学习来自高维数据的对称函数（对称函数是在给定参数时具有相同值的函数，而与参数的顺序无关。）为此，我们做出了另一贡献，并建议使用在视觉数据建模方面具有悠久历史的子空间来构造对称函数[16-19]。 这与以前的研究大不相同，在先前的研究中，对称函数是通过pooling的形式来实现的（例如，在[20]中求平均值）。

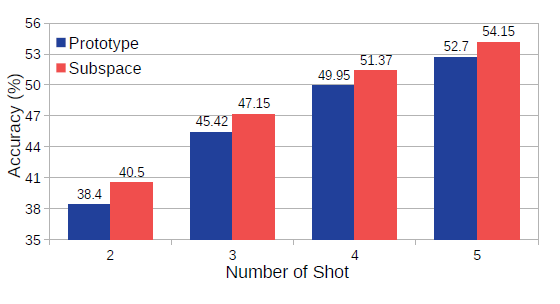


图1：使用少量（2-5）图像评估的原型和子空间分类器的准确性。 特征提取器在ImageNet上经过ResNet-34培训。 原型是同一类内的少量图像的平均池，而子空间是类特定的基向量。 原型和子空间是直接从生成的特征构建的，而没有其他可学习的参数。

作为一个激励性的例子，我们将最新的原型网络[20]与我们使用CUB数据集[21]提出的子空间方法进行比较和对比。 为此，对于通用特征提取器，我们使用了在ImageNet [22]上训练的ResNet-34。 我们考虑了具有不同样本的四个FSL问题（具体为2到5个），并在图1中报告了原型网络和子空间方法的准确性。正如稍后将详细说明的那样，在原型网络中，一种构造低样本分类器的方法是平均每个类别中的所有样本。 除自然选择外，平均还得到**1**的支持。在[23]中，表明对于适当的变换ρ和φ，集合X上的所有对称函数都可以写成。2.在[11]中，作者指出，一类中样本的平均值与softmax学习到的分类器的参数高度相关，因此，人们希望平均值也能反映FSL中一类的真实参数。 尽管如此，我们观察到我们的子空间解决方案始终且轻松地胜过原型网络。 这个令人信服的结果是我们在有监督和半监督FSL上进行的全面实验所得出的结果（例如，参见表1和表5），这表明在小样本情况下，存在更好的方法可以从有限的观测值中建立分类器，而基于子空间的观测器就是我们的建议。

贡献。 总而言之，我们在这项工作中做出了以下贡献：

i. 在生成动态分类器的框架内制定小样本学习解决方案。

ii. 我们建议使用子空间扩展现有动态分类器。 我们依赖一个公认的概念，该概念表明二阶方法可以更好地归纳分类任务。

iii.我们还引入了一种判别性表述，在训练过程中鼓励子空间之间的最大区别。 该解决方案进一步提高了性能。

iv.我们证明了我们的方法可以利用未标记的数据，因此适合半监督的小样本学习和转导设置问题。在我们的实验中评估了这种变体的鲁棒性。

**2.相关工作**

在本节中，我们回顾了有关分类任务的小样本学习和子空间方法的文献。 最初引入小样本学习来模仿人类的学习能力。 一些早期的作品使用生成模型和相似性学习来捕获对象的零件和几何构造内的变化[3、15、24]。这些作品使用手工制作的功能来进行小样本分类。文献[15]中提出的星座模型考虑了对象部分进行推理。 这些零件的几何结构有助于区分不同的对象。 此外，Torralba等。 [24]利用视觉对象的相似特征，但是模型没有利用几何结构。 另一个非深层解决方案是Lake等人的工作。文献[3]使用一组原语（strokes）对小样本分类进行建模。 上面的几类分类方法没有经过端到端的训练，给定的任务也不是偶然的。

深度学习在从图像中学习区分特征方面非常成功。Santoro等[25]和Vinyals等[4]尝试使用端到端的深度神经网络解决小样本分类。 在大多数情况下，受情节训练的网络旨在从有限的数据中推断特定任务的基础区分模型。元学习还可以用于获得快速自适应网络。 一个突出的想法是学习神经网络参数（权重）的初始值。 通过适当的初始化，可以使用有限样本的反向传播来期望网络适应不同的任务。 Sachin等[8]使用长期短期记忆（LSTM）嵌入了w.r.t. 给定任务来训练网络。 MAML [7]不使用LSTM编码梯度，但是它仍然可以执行元学习，通常具有更好的性能。作为扩展，MAML ++ [26]使用重要性方案来权衡梯度更新期间的损失。MetaNets [27]是另一个快速自适应网络，混合了所谓的快速权重和固定权重。 快速权重通过反向传播而改变，而固定权重不变。 因此，可以将这种方法视为仅应用于选定权重的一种优化。

基于度量学习的FSL是最接近我们工作的方向。 匹配网络[4]和孪生网络[13]学习按样本度量，这意味着到样本的距离用于确定查询的标签。 在原型网络中[20]，Snell等人。 将想法从样本扩展到了按类别度量。 来自特定类别的所有样本的描述符被分组并视为类别原型。 随后将原型用于推理。学习类表示和查询之间的非线性关系可以通过神经网络进行建模，如关系网络中所示[14]。 学习了基本度量标准，以保持共享相同类标签的特征向量之间的较小距离。 Qiao等指出，网络的激活与其分类器（最终层）的权重相关，并主张由激活制成的原型足以进行分类。其他作品使用特征关注模块[28，29]来调制特征以进行小样本学习[30，31]。

最近的几项工作针对的是小样本半监督学习（FS-SSL）。 Garcia等。 [32]利用图神经网络进行半监督设置，其中未标记的数据通过图神经网络（GNN）与标记的数据连接。 然后，将从GNN中提取的特征用于对查询进行分类。 Ren等人提出的另一种FS-SSL协议[33]表明，未标记图像有助于从支持集中提取样本，从而提高小样本分类的性能。 在[33]中提出的方法是基于原型网络[20]，其原型通过使用未标记图像进行了改进。

**3.问题设置**

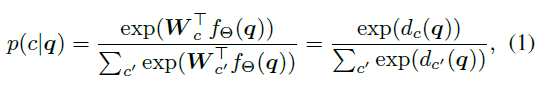
我们首先定义小样本学习中使用的术语。 对于每个迭代，都会以元学习的方式训练一些样本。 为了获得训练有素的模型，使用所谓的情节对数据进行采样。 情节由两个集合组成，即支持集S和查询集Q。此学习范例描述了在每次迭代中给定碎片数据的情况下机器如何提高其能力。具体来说，深度嵌入学习每集的标签和输入量有限。 这种学习范例是众所周知的N-way-K-shot分类（例如20-way-1-shot和5-way-5-shot）。 我们介绍（N-way, K-shot)）小样本学习的符号。每个情节或任务均由支持集和 查询集，其中xi，j表示来自类i的第j个样本，。 在半监督环境中，情节内还存在一个未标记的数据集。

一个相关的问题是半监督的小样本学习，其中将未标记的数据提供给模型。 在文献中，对于半监督的小样本学习，考虑了各种配置，例如[32-34]。 在这项工作中，我们遵循[33]中具有挑战性的协议，其中引入了所谓的干扰器。 因此，情节包括支持集S，查询集Q和未标记集R。将支持（标记）S和查询Q集配置为小样本学习。 另外，提供了未标记的集合R来辅助情节内的分类任务。 在未标记的集合中，有来自两个不同来源的样本：支持类别和干扰项类别。 顾名思义，来自干扰项类别的样本与分类任务无关，代表了支持集之外的类别。

**4.提议的方法**

**4.1初步**

我们考虑两个阶段的小样本学习问题：特征提取器和动态分类器。 令是从输入空间X到由神经网络实现的D维表示的映射，并且是特定于类的集合。 我们将小样本学习的问题表述为生成动态分类器。 为此，沿着softmax层的神经网络的最后一层实现了：



其中Wc是c类的权重，q为query查询样本。 然后，FSL的问题可以理解为一旦提供新任务就可以生成W。 为了展示此设置，我们在下面讨论成对分类器，原型和非线性分类器。

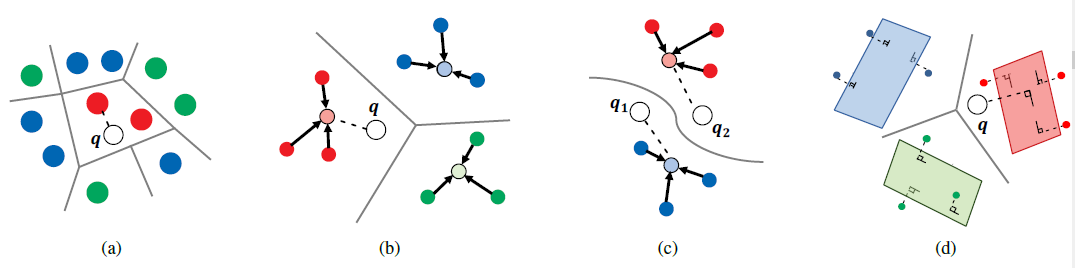


图2：用于小样本分类的各种分类器。 （a）匹配网络创建成对分类器。（b）原型网络根据同一类别中的样本创建均值分类器。（c）关系网络产生非线性分类器。（d）我们提出的方法使用子空间创建分类器。

Pair-Wise分类器。 如图2（a）所示，可以通过计算样本之间的相似性直接从样本中构建分类器。 使用这个的开创性的工作分类器是匹配网络[4]。 样本通过LSTM和关注模块嵌入。 但是，这种方法不会造成不变性。 输入图像的顺序会影响准确性。 分类器权重Wc替换为函数g（·）（例如LSTM）以编码样本。 然后，归类特定样本，并使用余弦相似度进行预测。

原型分类器。 基于文献[11]中对小样本分类的观察，最后一个完全连接层的参数与原型相关。 因此，从原型生成分类器。 通过引入一个简单的多层感知器，来自最终激活层的特征向量的平均值将用于执行小样本分类。通过直接学习特征嵌入的原型网络[20]也证实了这一观察。 以下某些作品还将原型用作动态分类器，例如[35，36]。 因此，Wc被代替。 此外，由于执行平均操作以生成分类器，因此该方法保留了对称属性（图像顺序不变）。 在图2（b）中描绘了该图示。

非线性二进制分类器。 这种方法利用了决策边界的非线性。 关系网络使用非线性二进制分类器来计算相似度，如图2（c）所示。 令，是可学习的分类器（比较器）。 我们可以重新定义等式1为，其中σ是非线性函数（例如sigmoid）。 即使此分类器不使用softmax函数，它也遵循生成分类器的原理，该分类器学习数据点对的比较。

**4.2针对Few-Shot分类的子空间**

我们建议通过子空间对点进行建模。 每个子空间Zi具有由的基底；n≤D，且。 我们的目标是学习特征提取器以生成子空间（即函数），使得生成的空间适合子空间分类器。

表示类别c的子空间的基底可以通过矩阵分解例如奇异值分解（SVD）来获得。 我们强调，获得鲁棒子空间的更多相关技术可能会改善算法。 尽管如此，我们的目标是评估用于小样本学习的子空间建模概念是否合理，因此我们在实现中选择了截断的SVD。

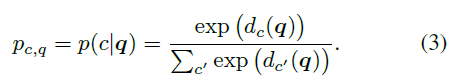
**4.3子空间分类器**

在提高分类器的功能上，高阶信息比低阶信息更有用。 子空间方法可以形成鲁棒的分类器。 下面，我们描述如何创建子空间并基于该子空间进行分类。 由编码的一组新样本可以表示为，其中。子空间上的分类方法之一是找到数据点与其在子空间上的投影之间的最近距离。 为此，从计算出特定类别的投影矩阵Pc。 现在，可以将查询投影到Pc上，并执行基于从查询到投影到Pc的最短距离（在原始空间中）的分类。 我们一般的子空间分类器定义为：



且μc可解释为点与子空间之间的偏移量。因此，Pc是矩阵Bc带有跨越线性子空间的正交基底的截断矩阵（因此，）。

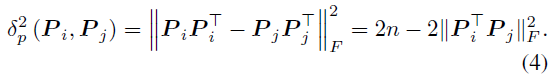
我们使用softmax函数将分配给类c的查询的概率定义为：



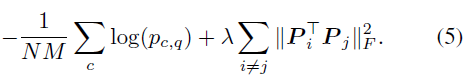
现在，我们可以最小化等式3的负对数，并更新。 为了训练整个框架，需要通过SVD进行反向传播，这在现代深度学习包（例如PyTorch [37]）中可用。 此后，我们将我们提出的方法称为深度子空间网络（DSN）。

**4.4区分性深子空间网络**

我们在这一部分中的目标是通过学习导致更多区分性子空间的表示来增强DSN。 为此，我们利用了格拉斯曼几何[38]并提出在训练过程中最大化子空间之间的距离。 这可以通过使用格拉斯曼式投影度量轻松实现，该度量具有多个有用的属性（参见[39]）。 更具体地说，给定两个子空间Pi和Pj的基底，将投影度量定义为：



通过最小化来实现最大化投影度量，从而产生以下损失：



算法1说明了训练DSN的步骤。 我们的总体流程如图3所示。

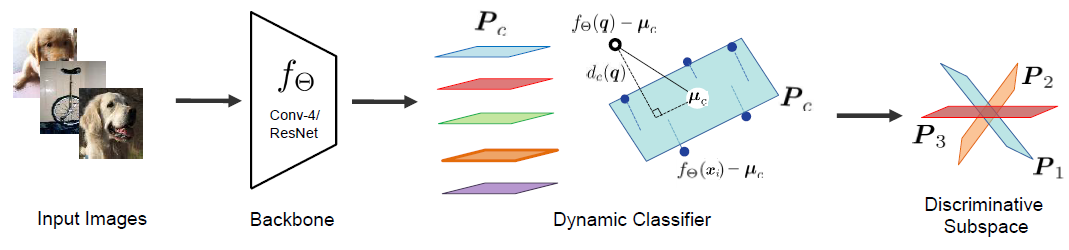
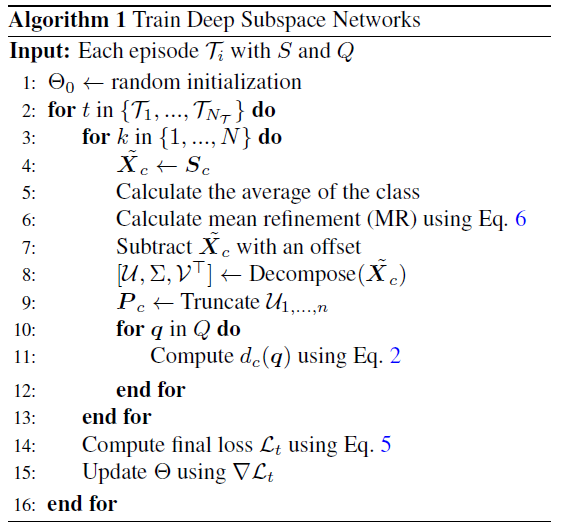
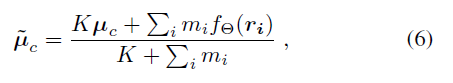


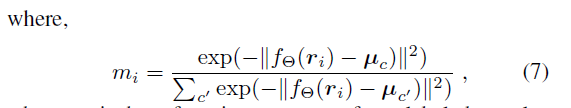
图3：我们方法的总体流程。 子空间分类器将每个类的分类器替换为单个向量。然后应用判别方法以最大化子空间之间的余量。



**4.5用于半监督FewShot学习的DSN**

接下来，我们扩展了第4.2节中开发的模型，以解决半监督的小样本学习。 这样做时，我们需要利用未标记的数据以使我们的数据更好地适应子空间。 我们通过根据以下方式细化每个类别的中心（均值细化）来实现





其中mi是未标记样本的软分配分数。为了在存在干扰因素的情况下工作，我们使用假平均值为零的伪类，如[33]。 我们凭经验观察到，对方法的这种简单修改可以改善结果，而无需改进矩阵分解步骤。此外，该技术还适用于使用查询集作为未标记数据来细化类均值的转导设置。

备注1.据我们所知，在[40，41]和我们的初步研究[42]中，子空间已用于解决FSL。 这项工作与TAPNET [40]之间的主要区别在于，我们的方法中的投影是特定于类的，而TAPNet使用的是特定于任务的投影。我们的初步工作[42]在大约8个月之前进行，Devos和Grossglauser [41]的工作具有相同的精神，可以被视为FSL的特定于类的子空间方法。

**5.实验**

下面我们对比和评估我们在四个具有挑战性的数据集上使用最新技术的方法，这些数据集是mini-ImageNet [8]，tiered-ImageNet [33]，CIFAR [43]和Open MIC [44]。 此外，我们在整个实验中使用了几个CNN主干，例如[20]中实现的4卷积层（Conv-4）和[45]中使用的ResNet-12，用于标准的小样本分类。 我们遵循一般做法，使用N-way K-shot和15个查询图像评估模型。 在进行摄动分析和半监督少发（SS-FSL）分类时，采用了Conv-4。 所有数据集都提供了深子空间网络（DSN）的报告结果。

Mini-ImageNet。 小型ImageNet [8]包含60,000张ImageNet [46]数据集的图像。 Mini-ImageNet中的图像尺寸为84×84，分别代表100个类别，其中64、16和20个类别分别用于训练，验证和测试。 每个类都有来自[8]的图像列表之后的600图像。 从先前的工作（例如[47]）清楚地表明，CNN主干会影响性能。因此，我们采用4卷积层（4-Conv）和ResNet12进行公平比较。 我们还将mini-ImageNet用于40％标记数据的半监督分类。

tiered-ImageNet。 该数据集也派生自ImageNet，但是与小型ImageNet相比，它包含的类更加广泛。 来自20个不同类别的351个类别用于培训，来自6个不同类别的97个类别用于验证，而来自8个不同类别的160个类别用于测试。 我们遵循4-Conv和ResNet-12主干的实现，并且在mini-ImageNet上的图像大小为84×84。

CIFAR-100。 我们对CIFAR-FS数据拆分进行评估。 这些数据集上的所有图像均为32×32，每类的样本数为600。CIFAR-FS数据集[49]是包含CIFAR-100 [43]的所有100个类的快照学习基准。 数据集分为64个，16个和20个，分别用于训练，验证和测试。

Open-MIC。 该数据集[44]包含来自10个博物馆展览空间的图像。 在此数据集中，有866个类别，每个类别1-20张图像。 图像会经历各种光度学和几何畸变，这些类的性质通常很细腻，因此使小样本学习问题很困难。 我们使用的协议和基线在[55]中提出，但排除了最容易分类的类别，以便可以测试1样本以上，然后重新运行SoSN [55]方法。 数据集分为四个子集：p1 =（shn + hon + clv），p2 =（clk + gls + scl），p3 =（sci + nat），p4 =（shx + rlc）。 协议[55]假定对p1→p2，p2→p3，p3→p4和p4→p1进行评估，其中x→y表示对子集x进行训练并对子集y进行测试。 在一个子集中进行训练并在另一个子集中进行测试描绘了小样本学习问题，因为每个博物馆中的对象都具有不同的背景。 请注意，我们用少于3个示例消除了类，并重新运行了实验中的所有算法。

**5.1Few-shot学习**

我们遵循常规做法，并在进行小样本学习和分类时在mini-ImageNet，tiered-ImageNet，CIFAR-FS和Open MIC上评估我们的方法。 mini-ImageNet的CNN架构与[47]中使用的相同，具有4个卷积层（Conv-4）和ResNet-12 [56]。 同时，仅ResNet-12用于CIFAR-FS和tiered-ImageNet。 我们使用ADAM [57]优化Conv-4，使用SGD优化ResNet-12。 为了公平地比较，我们进行了类似的实验设置。

按照其他方法训练Conv-4骨干而无需增加数据，并且每5K情节将学习率降低一半。 我们训练了5次1发和5发，然后在测试Conv-4时应用了相同的分类任务设置。 注意，使用Conv-4的原型网络[20]也经过了5路训练和测试。 ResNet-12的训练是通过数据增强进行的，最初将学习率设置为0.1，然后分别在时期12、30和45将其调整为0.003、0.00032和0.00014。 此外，[45]中的训练策略每批使用15样本，10张查询图像和8情节。 我们从一个验证集进行交叉验证，并为所有实验设置λ= 0.03。 评估准确性超过1000集。

通过设计，我们的方法需要多个样本来标识子空间的跨度。 因此，对于1样本的情况，我们通过翻转支持图像通过数据增强来生成额外的样本。

结果。 下面，我们提供基于Conv-4和ResNet-12的结果进行全面比较。 请注意，不同的骨干网可能会影响小样本学习的性能。 对于mini-ImageNet，表1显示了我们的方法在各种CNN主干以及5道5发和1发样本数量方面均优于最新方法。 我们的方法还可以受益于查询集的平均细化（MR）。 在具有更多参数（例如ResNet-12 [56]）的更深的CNN上，我们的方法甚至更好。

在5-way-1-shot和5-shot时，我们的性能比MetaOpt-SVM [45]高1.3％。 我们的方法也始终优于其他方法，如ImageNet和CIFARFS数据集（请参见表2和3）。

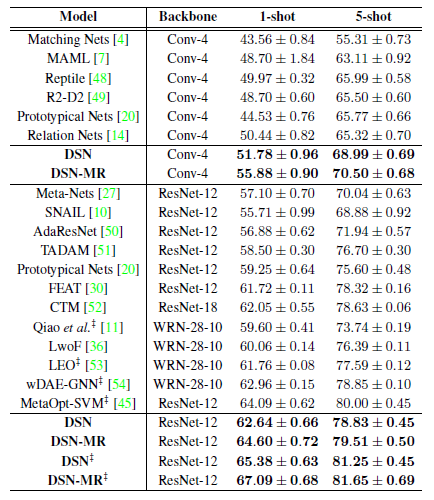


表1：与现有技术的比较。 在mini-ImageNet数据集上具有95％置信区间的5-Shot小样本分类结果，该数据集具有用于1-shot和5-shot的各种主干。 ‡的方法包括用于训练模型的训练和验证集。

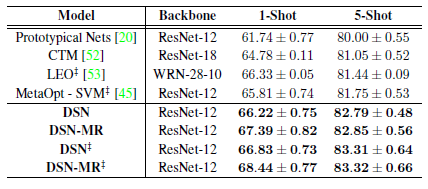


表2：在tiered-ImageNet上以95％置信区间进行的5-shot小样本分类结果。 ‡的方法包括用于训练模型的训练和验证集。

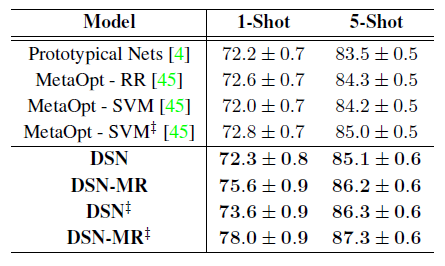


表3：使用ResNet-12在CIFAR-FS数据集上以95％置信区间进行的5-way小样本分类结果。‡的方法包括用于训练模型的训练和验证集。

在open-MIC数据集（请参见表4）上，可以观察到类似的趋势。 我们的方法优于用于少量学习的最先进的嵌入方法（即，匹配网[4]，原型网[20]和二阶相似性网络（SoSN）[55]）。 结果表明，我们的子空间表示对于Open MIC数据集造成的各种光度和几何失真具有鲁棒性，并且可以很好地建模此数据集中包含的细粒度概念。 OpenMIC包含具有不同类型对象的不同展览。 与其他方法相比，我们的模型可以推广到Open MIC上不同的对象子集，增益大约为2％。

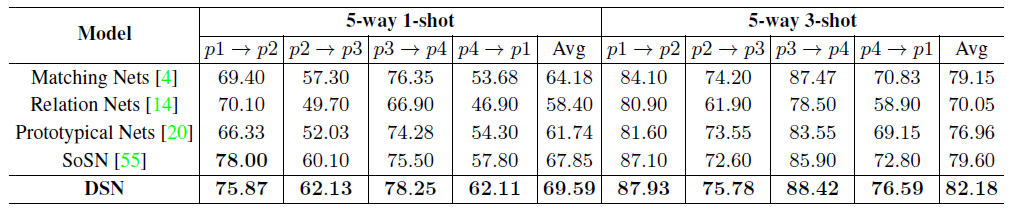


表4：在Open MIC数据集上使用Conv-4进行5-way-1-shot和3-shot的小样本分类结果。

**5.2半监督的Few-shot学习**

对于本节中的实验，我们使用具有[4]卷积层的嵌入体系结构，如[33]中所述。 我们遵循了[33]提出的实验设置。 支持和查询集的带标签部分的情节组成类似于小样本分类任务，但是，每个情节中都提供了一个额外的未贴标签的情节。 我们的模型在mini-ImageNet和tiered-ImageNet上分别具有40％和10％的标记数据的100K情节上进行了训练。我们使用了ADAM求解器[57]，然后随着权重的衰减将学习率设置为0.001，并将每10K集的学习率降低为一半。

训练是在半监督环境中进行的，也使用了未标记的环境。 未标记的集合由支持集合和干扰项类别中的样本组成。 支持培训和干扰培训的数量设置为五个，以进行培训和测试。在训练阶段，未标记集中的样本数量为50（每个班级有五个样本）。 在测试阶段，未标记的集合由每个类别的20个样本组成。 该查询集每个类别有20个样本用于测试目的。 分别将mini-ImageNet和tiered-ImageNet上的半监督小样本学习的λ设置为0.03和0.005。

结果。 评估准确性超过600集。 结果是标记和未标记集的10个随机拆分的平均值。 表5中详细介绍的半监督实验表明，我们的方法通过利用未标记的数据来提高性能。 我们的结果与半监督学习（SS-FSL）上的原型网络，其中包含soft K-means（非屏蔽）和masked K-means（屏蔽），如[33]所建议。

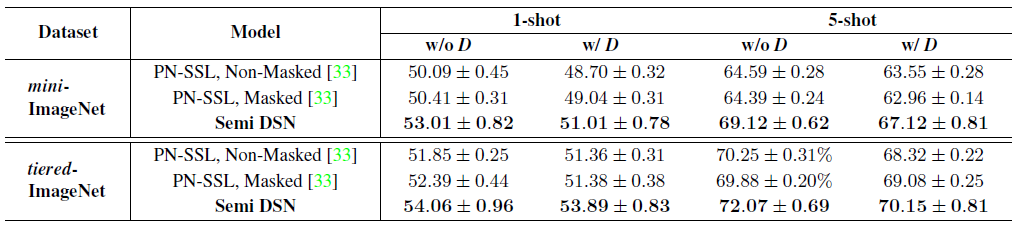


表5：在微型ImageNet和分层ImageNet上使用Conv-4分别获得40％和10％的标记数据的5次半监督小样本分类结果。 我们显示有（w / D）和没有干扰因素（w / o D）的分类结果。

**5.3消融研究**

区别项。 下面是烧蚀研究执行区分项。 等式4中的区分性项鼓励不同类的子空间之间的正交性。 此项可提高小样本分类任务的性能。 在给出Conv-4骨架的情况下，我们在表6中研究了该机制的结果。 根据结果​​，我们得出结论，网络学习了相互区分的判别子空间。 这项经验研究证明，判别项可以提高性能，并为判别提供更多判别子空间。

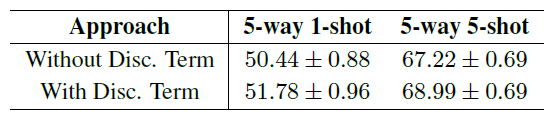


表6：使用Conv-4在mini-ImageNet上带有或不带有区分项的DSN的分类准确率很少。

子空间维数。 与其他模型（例如匹配网络，原型网络和关系网络）相比，我们的DSN附带了一个额外的超参数，即子空间的维数（即n）。 根据经验，我们建议使用n = K − 1来训练和测试我们的模型。 实际上，DSN对n表现出很大的鲁棒性，从而使我们的模型训练变得简单。 我们观察到，使用Conv-4主干网在mini-ImageNet上选择n从2到K − 1不会显着影响性能（±0.5％）。

**6.讨论**

对干扰的鲁棒性。 有人可能会争辩说，噪声是否会在小样本学习中带来问题。 但是，收集数据时某些噪声模式可能并不明显。因此，不能保证数据没有噪声。 我们在实验中观察到，标准方法的性能随着对信号的少量扰动而显着降低，如图4所示。但是，我们基于子空间的模型可以很好地处理这种噪声。

计算复杂度。 我们的DSN方法的计算复杂度为O（min（ND2K，NDK2）），其中K，N和D分别是样本数量，路数和特征维数。 与原型网络方法（即O（NDK））的复杂性相比，由于使用了SVD步骤，我们的方法要慢一些。 然而，为了解决SVD的复杂性，可以使用快速近似SVD算法[58]。

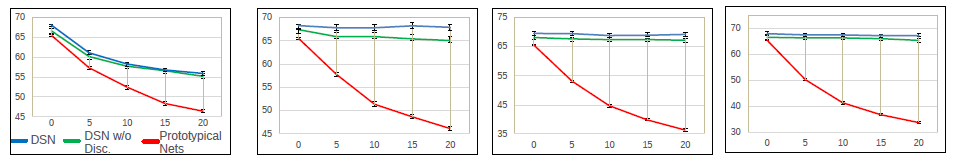


图4：使用Conv-4在mini-ImageNet上进行5-way-5-shot时存在异常值和附加噪声的实验。 显示了DSN，无判别式术语的DSN和原型网络的结果（请参见图例）。 第一列显示了在支持样本中引入异常值的影响（异常值类别与样本的支持类别不相交）。 第二，第三和第四列分别显示了引入根据高斯分布随机产生的有噪声样本的影响，该样本分别具有随机平均值和σ= {0.15，0.3，0.4}。 性能是通过w.r.t. 离群值和噪声样本（x轴）的数量不断增加。

**7.结论**

本文介绍了DSN，这是一种新颖的小样本学习方法，它通过仿射子空间采用了小样本学习模型。 从经验上讲，我们表明通过DSN学习的表示形式在广泛的有监督和半监督的小样本镜头问题中都具有表达力。 两者都接受过元学习训练，并且在训练模型时以前没有看到测试集。 事实证明，子空间模型可以代表子空间上的几个数据点，因此可以大大改善现有模型。

在DSN中，每个类别分类器均由其所有样本形成的子空间表示，这意味着每个类别均通过其训练数据点的范围进行建模。 我们证明了DSN在小样本学习中对噪声具有鲁棒性。 我们的实验表明，通过简单地鼓励子空间彼此分离，可以获得更高的分类精度。