**DPGN: Distribution Propagation Graph Network for Few-shot Learning**

旷视研究院

会议：CVPR2020

代码：<https://github.com/megvii-research/DPGN>

**摘要**

大多数基于图网络的元学习方法都为样本的实例级关系建模。 我们进一步扩展了这一思想，以1-vs-N的方式对一个示例与所有其他示例的分布级别关系进行显式建模。我们提出了一种用于小样本学习的名为分布传播图网络（DPGN）的新颖方法。 它在每个小样本学习任务中都传达了分布级关系和实例级关系。 为了结合所有示例的分布级关系和实例级关系，我们构建了一个完整的对偶图网络，该网络由一个点图和一个分布图组成，每个节点代表一个示例。 配备了双图体系结构的DPGN在数代更新中将标签信息从标记的示例传播到未标记的示例。 在以小样本学习基准为基础的广泛实验中，DPGN在有监督的设置下要比最新结果的准确率提高许多，分别为5％~12％和在半监督下的7％~13％。

**1.简介**

深度学习的成功源于大量的标记数据[19，38]，而人类在看到很少的例子之后可以很好地概括。 这两个事实之间的矛盾引起了对小样本学习的研究的极大关注[7，20]。 小样本学习任务旨在，在给定一些标记数据（支持集）的情况下预测未标记数据（查询集）。

微调[4]是当今实践中从一个小的训练数据集中获取预测模型的实用方法。 但是，它存在过拟合问题[11]。 元学习[8]方法引入了情节（episode）的概念，以明确解决小样本问题。情节是一轮模型训练，其中在每个情节中，只有很少的示例（例如1或5个样本）是从训练数据中的每个类中随机抽样的。 元学习方法采用训练器（也称为元学习器），该训练器获取小样本训练数据并输出分类器。 这个过程称为episodic training [41]。 在元学习的框架下，提出了各种各样的假设来构建一个有效的元学习器。

近期研究的一个上升趋势是使用Graph Networks [2]处理训练数据，这是一个强大的模型，可以概括许多数据结构（列表，树），同时引入组合先验数据。 提出了小样本的GNN [10]来构建一个完整的图网络，其中每个节点特征与相应的类标签串联在一起，然后通过图网络的关注机制更新节点特征以传播标签信息。 为了进一步利用基于图的网络中的集群内相似度和集群间不相似度，EGNN [18]演示了在情境训练框架下的边缘标记图神经网络。 值得注意的是，以前的GNN小样本学习研究主要是关注节点标记或边缘标记等成对关系，而忽略了大量的实质分布关系。 此外，其他元学习方法声称可以通过情节培训来利用全局关系的好处，但这种方式是隐式的。

如图1所示，首先，我们提取支持和查询样本的实例特征。 然后，我们通过计算所有支持样本的实例级相似度来获得每个样本的分布特征。 为了利用每个示例的实例级表示和分布级表示并分别处理不同级别的表示，我们提出了双图体系结构：点图（PG）和分布图（DG）。具体而言，PG通过在每个示例上收集1-vs-n关系来生成DG，而DG通过传递每对示例之间的分布关系来精炼PG。这种循环转换充分融合了实例级和分布级之间的关系，并且这种“聚集-比较”过程的多个世代（回合）总结了我们的方法。 此外，很容易将DPGN扩展到半监督的小样本学习任务，其中支持集包含每个类别的标记和未标记样本。DPGN以相似性分布的形式在标记的样本和未标记的样本之间建立桥梁连接，这可以在半监督的小样本分类中更好地传播标记信息。

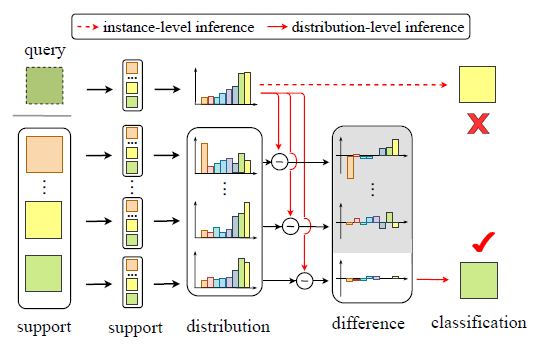


图1：我们建议的DPGN在每个样本与支持样本之间采用对比比较，以产生分布表示。 然后，在对查询样本进行分类时，它将分布级比较与实例级比较结合在一起。

我们的主要贡献概述如下：

据我们所知，DPGN是第一个将分布传播明确地并入图网络中以进行小样本学习的。 进一步的消融研究证明了分布关系的有效性。

我们设计了结合实例级和分布级关系的双重完整图网络。此框架中的循环更新策略有助于利用分布信息来增强实例功能。

对四个流行的基准数据集进行了广泛的实验，以进行小样本学习。 通过与所有最新方法进行比较，DPGN的小样本分类准确率平均显着提高了5％~12％。 在半监督任务中，我们的算法的性能比现有的基于图的小样本学习方法高7％~13％。

**2.相关工作**

**2.1图神经网络**

图神经网络最初是为处理图结构数据而设计的[34，41]。 图神经网络主要通过递归聚合和转换相邻节点来精炼节点表示。 提出了最近的方法[10、25、18]，以在小样本学习任务领域中利用GNN。TPN [25]将转导设置引入基于图的小样本学习中，该学习执行拉普拉斯矩阵将标签从支持集传播到图中的查询集。 它还通过成对的节点特征相似度传播标签的过程来考虑支持样本和查询样本之间的相似性。EGNN [18]使用样本之间的相似性/不同性，为复杂的交互动态更新节点和边缘特征。

**2.2度量学习**

小样本学习方法的另一类重点是使用度量学习方法来优化输入数据的特征嵌入。 Matching Networks [41]通过计算支持和查询集之间的嵌入距离来生成加权最近邻分类器。 原型网络[36]首先在嵌入空间中构建每个类的原型表示。 作为原型网络的扩展，IMF [1]通过自适应构造了无限的混合原型。 RelationNet [40]采用距离度量网络来学习支持和查询样本中的逐点关系。

**2.3分布学习**

分布学习理论最早是在[17]中引入的，目的是找到一种有效的算法来确定样本的分布。 提出了不同的方法[16、5、6]来有效地估计目标分布。 DLDL [9]是为分类和回归任务中的每个实例分配离散分布而不是单标签的研究之一。 CPNN [44]将要素和标签作为输入，并在其框架中仅具有一个隐藏层的情况下产生标签分布。 LDLF [35]设计了一种基于决策树算法的分布学习方法。

**2.4元学习**

一些方法采用元学习框架，该框架跨批次的任务学习元级知识。 MAML [8]是基于梯度的方法，将元学习器设计为优化器，可以在给出新示例的几个优化步骤内学习更新模型参数（例如，深度网络的所有层）。 Reptile [28]通过合并L2损失简化了元损失的计算，该损失使元模型参数朝着针对具体实例的适应模型进行更新。SNAIL [27]学习参数化的预测变量以估计模型中的参数。 MetaOptNet [21]提倡使用线性分类器代替最近邻方法，后者可以优化为凸学习问题。 LEO [33]利用编码器-解码器架构来挖掘潜在的总体表示和预测极端低数据状态下的高维参数。

**3.方法**

在本节中，我们首先提供了小样本学习任务的背景，然后详细介绍了所提出的算法。

**3.1问题定义**

小样本学习任务的目的是训练一个模型，该模型在仅给出少量样本的情况下可以很好地执行。

每个小样本任务都有一个支持集S和一个查询集Q。在给定训练数据的情况下，支持集SDtrain包含N个类别，每个类别有K个样本（即N-way-K-shot设置）， 可以表示为。 查询集QDtrain具有个样本，可以表示为。具体来说，在训练阶段，同时为支持集S和查询集Q提供数据标签。给定测试数据Dtest，我们的目标是训练一个分类器，该分类器可以将查询样本从Q Dtest准确地映射到相应的标签，而几乎不需要支持来自S Dtest的样本。 支持集和查询集的标签是互斥的。

**3.2分布传播图网络**

在本节中，我们将详细解释为小样本学习而提出的DPGN。 如图2所示，DPGN由l个世代组成，每个世代都由一个点图和一个分布图组成。 首先，通过卷积骨干提取所有样本的特征嵌入，这些嵌入被用于计算实例相似度。 其次，传递实例关系以构造分布图。 节点特征通过按照中的位置顺序聚合来初始化，并且边缘特征代表节点特征之间的分布相似性。 最后，将获得的传递给，以构造节点的更具判别性的表示，然后逐代重复上述过程。 DPGN的世代更新的简短介绍可以表示为，其中l表示第l代。

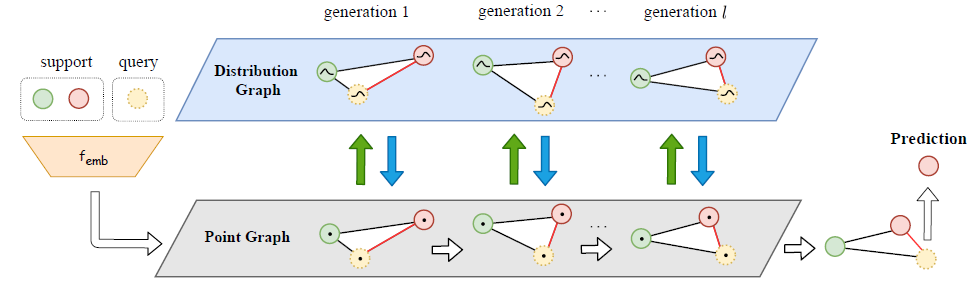


图2：DPGN的总体框架。 在此图中，我们以2-way-1shot任务为例。 从特征提取器获得的支持和查询嵌入被传递到对偶完整图（点图和分布图），以便在生成之后进行转导传播。 绿色箭头表示边缘到节点的转换（P2D，请参见第3.2.1节），该实例聚合实例相似性以构造分布表示，蓝色箭头表示另一个边缘到节点的转换（D2P，请参见第3.2.2节），将分布相似性与实例特征进行汇总。 DPGN在生成l的末尾对查询样本进行预测。

为了进一步解释，我们将和公式如下： ，，表示训练情节中的示例总数。首先由特征提取器的输出初始化。 对于每个样本xi：



其中，m表示特征嵌入的维度。

**3.2.1点到分布聚合**

**点相似度** 点图中的每个边代表实例（点）相似度，第一代边初始化如下：



其中是将实例相似度转换为一定比例的编码网络。包含两个Conv-BN-ReLU [13，15]块，其参数集为和sigmoid层。

对于世代l> 0的情况，给定可以将更新如下：



为了将边缘信息用于图形的整体视图，对进行归一化操作。

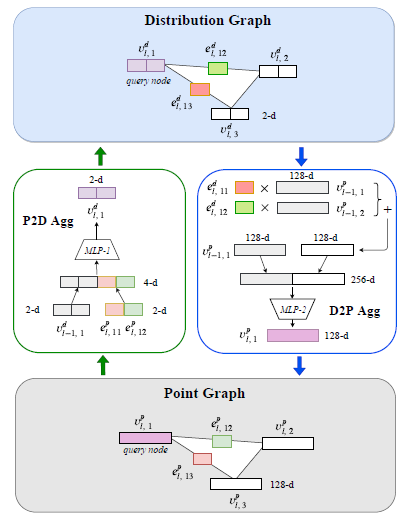
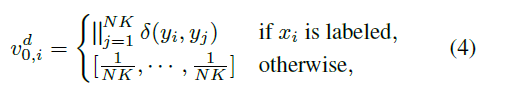


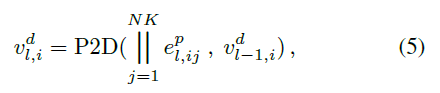
图3：有关DPGN中P2D聚合和D2P聚合的详细信息。 以2-way-1-shot任务为例。 MLP-1是P2D聚合中提到的F​​C-ReLU块，而MLP-2是D2P聚合中提到的Conv-BN-ReLU块。 绿色箭头表示P2D聚合，而蓝色箭头表示D2P聚合。 两种聚合过程都集成了其上一代的节点或边缘功能。

**P2D聚合** 在生成或更新点图中的边缘特征之后，接下来要构造分布图。 如图3所示，的目的是整合来自点图的实例关系并处理分布级别的关系。 中的每个分布特征是一个NK维特征向量，其中第j个条目中的值表示样本xi与样本xj之间的关系，而NK表示任务中支持样本的总数。 对于首次初始化：



其中，是串联运算符。是Kronecker delta函数，当yi = yj时输出1，否则输出0（yi和yj是标签）。

对于世代l> 0，可以如下更新分布节点



其中是分布图的聚合网络。 P2D在两个功能之间应用串联操作。 然后，P2D执行转换：在级联特征上，该级联特征由完全连接的层和ReLU [13]组成，参数集为。

**3.2.2分布到点聚合**

分布相似度分布图中的每个边代表不同样本的分布特征之间的相似度。 对于世代l = 0，分布相似度初始化如下：

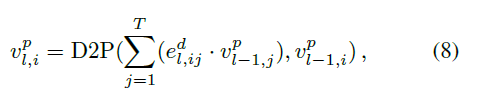


其中。编码网络使用两个Conv-BN-ReLU块转换分布相似性，最后将参数设置为和sigmoid层。 对于l> 0的世代，中的更新规则公式如下：



另外，我们对进行归一化处理。

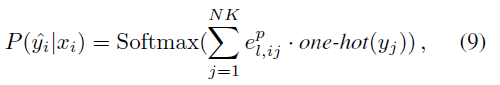
D2P聚合如图3所示，中的编码分布信息在每一代结束时流回到点图。 然后，中的节点特征通过将中的所有节点特征与边缘特征进行聚合来捕获分布关系，如下所示：



其中和是中带有参数集的点图的聚合网络。D2P将由计算出的特征与上一代节点特征进行级联，并使用两个Conv-BN-ReLU块更新级联的特征 。 在此过程之后，节点功能部件可以将分布级信息集成到实例级功能部件中，并为下一代计算实例相似性做准备。

**3.3对象**

每个节点的类别预测可以通过将DPGN的最后一代l中的对应边馈入softmax函数来计算：



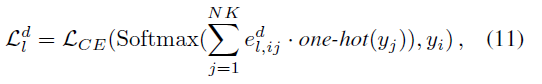
其中是给定样本xi的类别上的概率分布，而yj是支持集中第j个样本的标签。代表最后一代点图中的边缘特征。

点损失 注意，我们在点图中对每个样本进行分类预测。 因此，第l代的点损失定义如下：



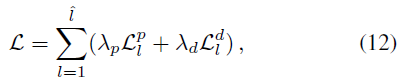
其中是交叉熵损失函数，T表示每个任务中的样本数。和yi分别是样本xi和真实标签的模型概率预测。

分布损失 为了简化培训过程并学习有区别的分布特征，我们将分布损失纳入其中，该分布损失在促进更快更好地收敛方面起着重要作用。 我们定义第l代的分配损失如下：



其中代表第l代分布图中的边缘特征。

总目标函数是上述所有损失的加权总和：



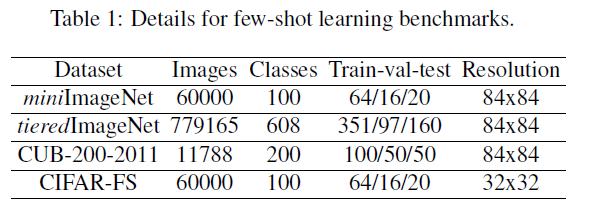
其中表示DPGN的总代，每个损失的权重设置为平衡其重要性。 在我们的大多数实验中，分别设置为1.0和0.1。

**4.实验**

**4.1数据集和设置**

**4.1.1数据集**

我们根据四个标准的一次性学习基准评估DPGN：mini-ImageNet [41]，tiered-ImageNet [31]，CUB-200-2011 [42]和CIFAR-FS [3]。 Mini-ImageNet和tiered-ImageNet是ImageNet的子集[32]。CUB-200-2011最初设计用于细粒度分类，而CIFAR-FS是CIFAR-100的子集，用于小幅分类。 如表1所示，我们按照先前的工作[41、31、4、3]的标准列出了图像编号，类别编号，图像分辨率和训练/行驶/测试分割的详细信息。



**4.1.2实验设置**

**网络架构** 我们使用四个流行的网络进行公平比较，分别是EGNN [18]，MetaOptNet [21]，CloserLook [4]和LEO [33]中使用的ConvNet，ResNet12，ResNet18和WRN。ConvNet主要由四个Conv-BN-ReLU块组成。 最后两个块还包含一个dropout层[37]。 ResNet12和ResNet18与[14]中描述的相同。 它们主要有四个块，其中分别包括一个用于ResNet12的剩余块和两个用于ResNet18的剩余块。WRN最早是在[46]中提出的。 它主要具有三个残差块，网络深度设置为[33]中的28。 所有骨干网的最后特征都由全局平均池处理，然后是具有批处理归一化功能的全连接层[15]，以获得128维实例嵌入。

**训练模式** 我们在训练之前执行数据增强，例如水平翻转，随机裁剪和颜色抖动（亮度，对比度和饱和度），这些在[11，43]中提到。我们在每次迭代中随机抽取28个元任务情节进行元训练。Adam优化器用于所有实验，初始学习率为。 我们每15000次迭代将学习率降低0.1，并将权重衰减设置为。

**评估协议** 我们以标准的小样本学习数据集的5-way-1-shot / 5-shot设置评估DPGN，mini-ImageNet，tiered-ImageNet，CUB-200-2011和CIFAR-FS。 我们遵循先前方法的评估过程[18、33、43]。 我们随机抽样10,000个任务，然后报告平均准确度（以％为单位）以及95％的置信区间。

**4.2实验结果**

主要结果 我们将DPGN的性能与几种最先进的模型（包括图形和非图形方法）进行了比较。 为了进行公平的比较，我们在mini-ImageNet，tiered-ImageNet，CIFAR-FS和CUB-200-2011数据集上采用了DPGN，并与相同主干中的其他方法进行了比较。 如表2、3和4所示，与基于图形的方法相比，所提出的DPGN优于其他现有方法并具有最新的性能。

表2：mini-ImageNet上的小样本分类准确度。表示由公共代码实现。[10，25，18]和DPGN在转导中进行了测试。

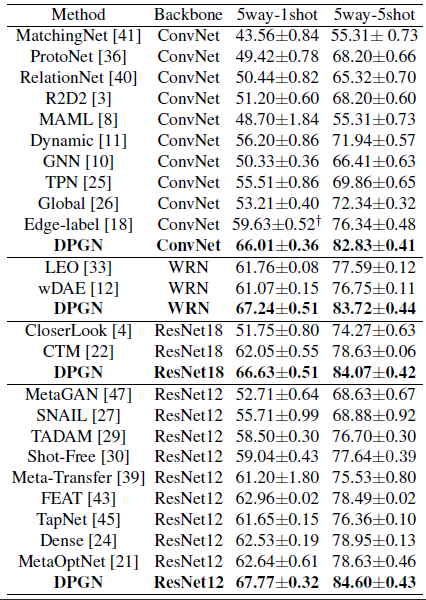


表3：tiered-ImageNet上小样本分类的准确性。表示它是通过公共代码实现的。 \*表示从[21]中报告。 [25，18]和DPGN在转导中进行了测试。

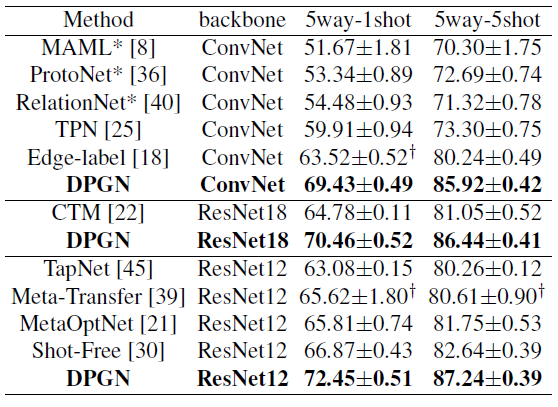
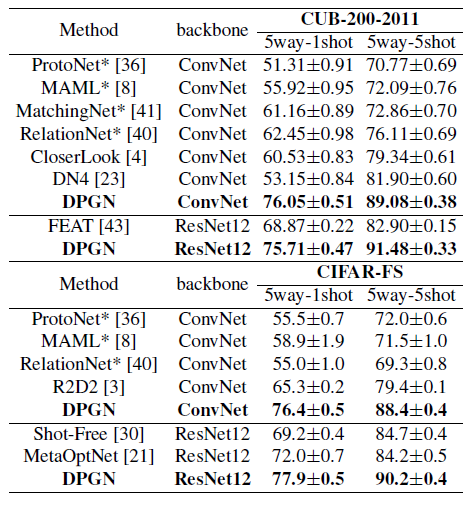


表4：CUB-2002011和CIFAR-FS的小样本分类准确度。 \*表示从[21]或[4]报告。DPGN经过转导测试。



**半监督小样本学习** 我们在半监督小样本学习中采用DPGN。 按照[25，18]，我们使用相同的条件将mini-ImageNet数据集拆分为标记和未标记部分的比例不同。 对于20％标记的半监督方案，对于每个类别中的标记和未标记数据，我们以0.2 / 0.8的比例拆分支持样本。 在半监督的小样本学习中，DPGN使用未标记的支持样本来显式构造所有其他样本的相似性分布，并且这些分布可作为查询和标记的支持之间的连接，可以将标签信息从带标签的样本传播到查询中

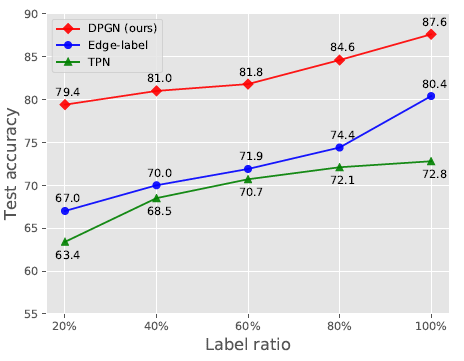
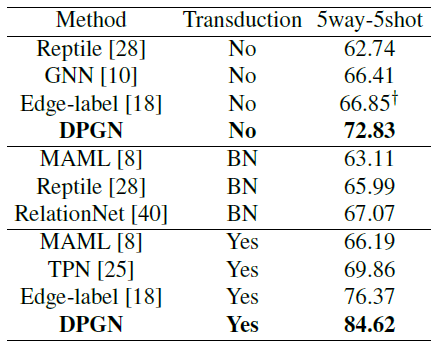


图4：在mini-ImageNet上以5-way-10-shot进行半监督的小样本学习准确性。DPGN始终远远超过TPN和EGNN

在图4中，DPGN显示了优于现有的半监督小样本方法，结果表明，当标记率降低时，利用标记数据与未标记数据之间关系的有效性。值得注意的是，在mini-ImageNet上，小样本平均分类精度方面，DPGN分别比TPN [25]和EGNN [18]高出11％~16％和7％~13％。

表5：mini-ImageNet上的转导/非转导性实验。“ BN”表示使用批归一化在测试示例之间共享信息。表示它是由作者发布的公共代码实现的。



转导传播 为了在我们的框架中验证转导设置的有效性，我们在5-way-5-shot设置下对mini-ImageNet数据集进行了转导和非转导实验。 表5显示DPGN的精度在转换设置中（与非转换设置相比）大大提高。 与仅考虑实例级功能的TPN和EGNN相比，DPGN利用查询样本之间的分布相似性，并采用双图体系结构以充分的方式传播标签信息。

high-way分类 此外，在mini-ImageNet数据集上评估了DPGN在high-way小样本场景中的性能，其结果如图5所示。观察到的结果表明，DPGN不仅超越了强大的基于图的方法[25，18]，而且还大大超越了最新的非图方法。 随着小样本任务中类的增加，它可以拓宽分布利用的视野，并使DPGN可以收集更多丰富的分布级信息以进行查询。

**4.3消融研究**

**分布图的影响** 分布图通过传播分布信息而成为DPGN的重要组成部分，因此有必要对的有效性进行定量研究。 我们通过限制分布相似性来设计实验，该相似性流到以在推理过程中的每一代中执行聚合。 具体而言，我们通过保留不同数量的特征维来掩盖边缘特征，并将其余维的值设置为零，因为零没有贡献。 图6显示了mini-ImageNet上5-way-1-shot实验的结果。显然，中的测试准确度与特征维数具有正相关关系，并且准确度增量（蓝色区域）随着特征维数的增加而降低。保持维度从0到5，DPGN的绝对值性能几乎提高了10％，结果表明分布图对我们的框架有很大的影响。

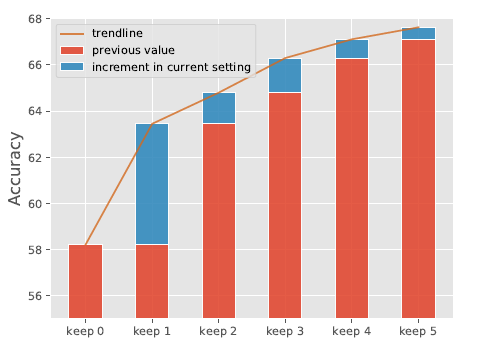


图6：通过在mini-ImageNet上以5-way-1-shot保持n个维度，的有效性。

**总代数** DPGN具有包括点图和分布图的循环体系结构，每个图分别具有节点更新和边缘更新模块。总代数是DPGN的重要组成部分，因此我们在mini-ImageNet，tiered-ImageNet，CUB-200-2011和CIFAR-FS上进行实验，以获得DPGN中不同代数的测试准确性趋势。 在图7中，随着世代数从0变为1，测试精度显着提高。 当世代数从1变为10时，测试精度会稍有提高，并且曲线在最后几代中会发生波动。 考虑由于更多的代需要更多的迭代来收敛，我们选择第六代作为测试精度和收敛时间之间的折衷。另外，为了可视化循环更新的过程，我们选择一个测试场景，其中五个查询图像的地面真实等级为[1、2、3、4、5]，并可视化实例级相似性，用于预测五个查询 样本如图8所示。热图显示DPGN经过几代后细化了实例级相似性矩阵，并对最后一代中的五个查询样本做出了正确的预测。 值得注意的是，DPGN不仅有助于更准确地预测，而且还通过使实例特征更具区分性来扩大不同类别的样本之间的相似距离，从而清理了预测热图。

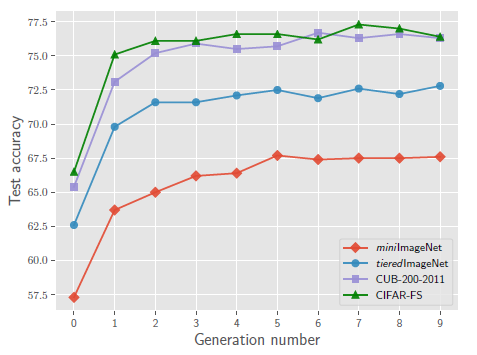


图7：mini-ImageNet，tiered-ImageNet，CUB-200-2011和CIFAR-FS上DPGN中的世代数。

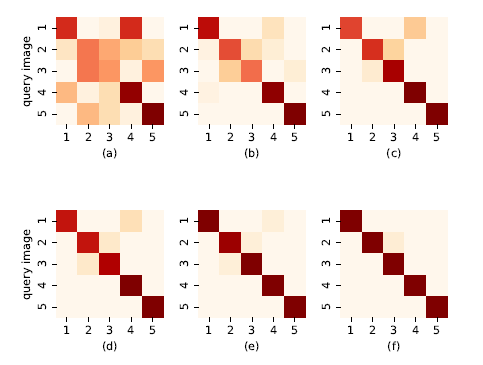


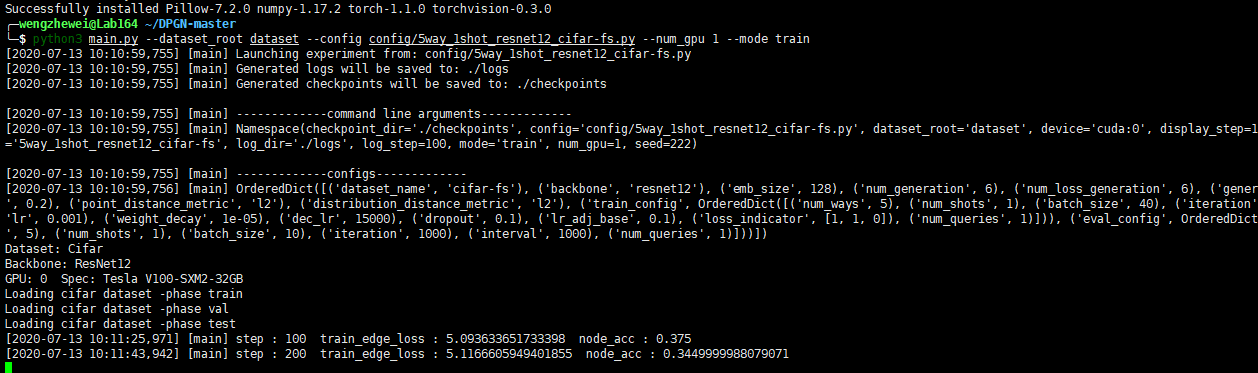
图8：每代DPGN中边缘预测的可视化。（a）到（f）表示第1到6代。深色表示更高的分数，而浅色表示较低的置信度。 左轴代表5个查询图像的索引，底轴代表5个支持类。

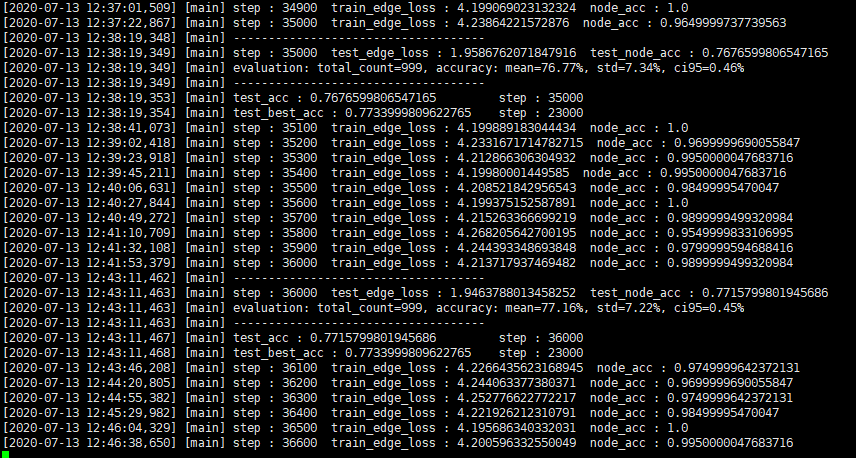
**5.结论**

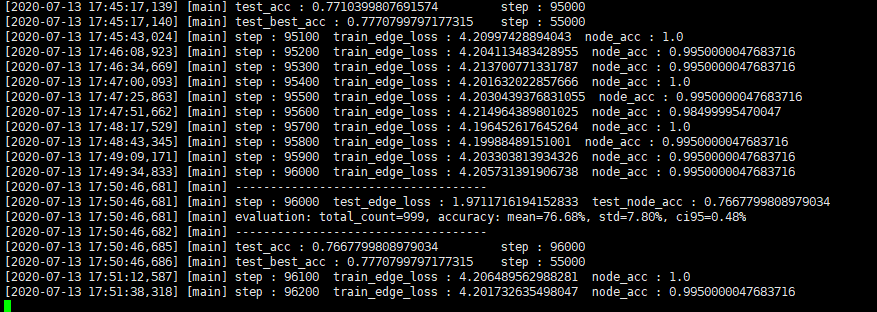
在本文中，我们介绍了用于小样本学习的分布传播图网络，这是一个完整的双重图网络，它以装备标签传播和转导的显式方式结合了实例级和分布级关系。 点和分布损失用于通过情景训练联合更新DPGN的参数。 大量的实验表明，在小样本学习基准上，我们的方法在有监督的任务中的性能比最新的最新算法高出5％~12％，在半监督的任务中7％~13％。 在以后的工作中，我们的目标是通过对与任务级关系链接的更复杂的信息进行编码，专注于高阶消息的传播。

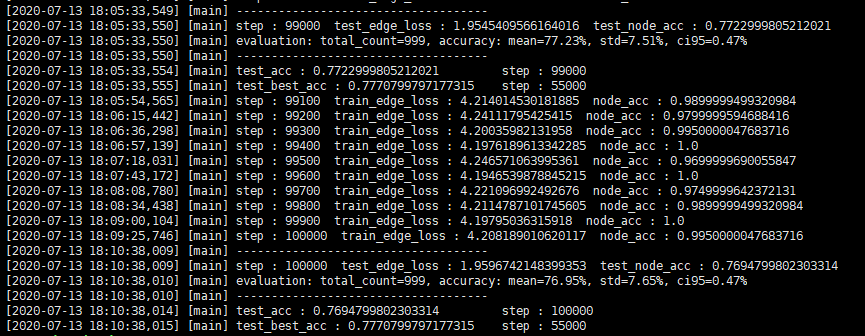
**7.论文复现**

数据集CIFAR-FS,5-way-1-shot模式，共100000个iteration，耗时8个小时。









数据集CIFAR-FS ，5-way-5-shot模式，100000个iteration

