南京邮电大学

毕业设计(论文)外文资料翻译

|  |  |
| --- | --- |
| 学 院 | 计算机学院、软件学院、网络空间安全学院 |
| 专　　业 | 数据科学与大数据技术 |
| 学生姓名 | 韦百安 |
| 班级学号 | B21111213 |
| 外文出处 | IEEE Transactions on Pattern Analysis and Machine Intelligence (2024) |

附件：1.外文资料翻译译文；2.外文原文

|  |
| --- |
| 指导教师评价：  1．翻译内容与课题的结合度： ☑优  良  中  差  2．翻译内容的准确、流畅：  ☑优  良  中  差  3．专业词汇翻译的准确性：  优  ☑良  中  差  4．翻译字符数是否符合规定要求：☑符合  不符合  　　　　　　　　　　　　　　　　　　　　　　　指导教师签名：  2025 年4月14日 |

**面向任务的通道注意力机制的细粒度小样本分类**

苏西恩·李(Sußeen Lee)、文俊·文(WonJun Moon)、玄锡·成(Hyun Seok Seong)和河在弼(Jae - Pil Heo)，IEEE会员

摘要 — 细粒度图像分类的难点主要在于类别之间的整体外观相似。因此，识别鸟类的眼睛和喙等区分细节是该任务的关键。然而，当训练数据有限时，这尤其具有挑战性。为了解决这一问题，我们提出了一种针对细粒度少样本分类的任务差异最大化（TDM）方法，该方法采用面向任务的通道注意力机制，并引入了两个新颖模块：支持注意力模块（SAM）和查询注意力模块（QAM）。SAM突出编码类别特征的通道，而QAM则为对象相关的通道分配更高的权重。查询的通道。基于这些子模块，TDM通过专注于编码通道来生成任务自适应特征同时具有分类细节和查询的，用于准确的分类敏感相似性度量支持和查询实例。虽然TDM通过任务自适应校准通道重要性来影响高速特征图，我们进一步引入了在特征提取器中间层运行的实例注意力模块（IAM），以实例方式突出显示通过扩展QAM，对象相关通道。TDM和IAM的优点及其互补优势在细粒度少样本分类任务中得到了实验验证。此外，IAM也被证明在粗粒度和跨域少样本分类中是有效的。

关键词 — 少样本分类、细粒度分类、特征对齐、注意力模块

# 1 引言

深度学习在各种视觉任务中取得了巨大进展，甚至在许多下游任务中取得了超越人类的显著性能 [6]，[12]。然而，这种性能是在有大量标注图像的情况下实现的，这需要巨大的标注成本。换句话说，如果标注图像的数量不足，性能会显著下降 [3]，[10]，[52]。因此，标注图像短缺和标注成本高的有限条件推动了少样本分类的发展 [10]，[48]，[52]，少样本分类旨在训练一个对新类别具有高度适应性的模型。为实现这一目标，少样本分类的训练主要基于情景学习策略，其中每个情景包含从数据集中采样的几个类别。此外，每个类别的图像分别划分为支持集和查询集，用于训练和评估。

基于度量的学习是少样本分类的主流方法 [20]，[48]，[49]，[52]。这些方法使用预定义或在线训练的度量学习深度表示，并基于支持集和查询集在该度量下的距离对查询进行推理。然而，由于特征提取器仅使用基础类别进行训练，通过学习到的提取器计算的新类别特征图很难形成紧密的聚类 [44]，[63]。为缓解这一问题，最近的方法利用先验知识 [25]，[63] 或提出任务动态特征对齐策略 [8]，[14]，[18]，[47]，[57]，[60]，[62]。在这两种策略中，任务动态特征对齐方法受到关注，并且可以进一步分为两个主要流派:空间对齐和通道对齐。空间对齐方法 [8]，[14]，[18]，[57]，[57]，[60] 旨在解决不同实例特征图上关键特征之间的空间不匹配问题。另一方面，由于新类别的语义特征图没有针对每个情景进行优化，通道对齐方法试图通过考虑情景的组成，使这些特征图适应目标分类任务。

尽管上述对齐方法在粗粒度少样本分类任务上取得了巨大改进，但它们对细粒度数据集的提升并不显著。这主要是因为它们只专注于利用通道或空间信息，而这些信息对于该情景可能不具有区分性。实际上，在细粒度分类中，定位具有区分性的细节很重要，因为类别很可能具有相似的整体外观 [7]，[11]，[33]，[67]。因此，对于细粒度少样本分类，还应捕捉每个类别与其他类别仅有细微差异的独特线索。

在此背景下，我们引入了一个新颖的模块，即任务差异最大化模块(Task Discrepancy Maximization，TDM)，该模块通过为每个类别对通道进行加权来定位判别区域。TDM基于按类别划分的通道权重向量，突出表示判别区域的通道，并抑制其他通道的贡献。具体而言，TDM由两个组件组成:支持注意力模块(Support Attention Module，SAM)和查询注意力模块(Query Attention Module，QAM)。给定一个支持集，SAM会为每个类别生成一个支持权重向量，该向量在判别通道上呈现高激活值。另一方面，QAM以查询集作为输入，为每个实例输出一个查询权重向量，该查询权重向量会突出与对象相关的通道。为了计算这些权重向量，需要考虑每个特征图与相应的按通道平均池化特征之间的关系。由于按通道平均池化的特征图具有对象的空间信息[27]、[58]，当通道与空间平均特征图相似时，它们很可能表示显著区域。通过结合从我们的子模块计算得到的两个权重向量，最终定义了一个特定任务的权重向量。因此，该特定任务的权重向量用于生成任务自适应特征图，以取代原始特征图。

虽然TDM是为细粒度少样本分类任务量身定制的模块，但由于TDM是为处理高级特征图而设计的，其性能在很大程度上取决于特征提取器生成的给定特征图的质量。因此，我们进一步引入实例注意力模块(Instance Attention Module，IAM)作为QAM的扩展版本，以便在特征提取阶段也能实现我们的主要思想。与QAM不同，IAM在特征提取器的中间层运行，并为每个实例计算一个通道权重向量，以像现有的注意力方法[15]、[41]、[58]一样提高特征表示的质量。由于IAM促使特征提取器关注每个实例的信息丰富的通道，因此得到的特征图包含更多与对象相关的信息和更少的背景信息。如前所述，IAM旨在在特征提取阶段补充TDM，但有趣的是，它也有助于提高一般少样本分类任务的性能。

我们的主要贡献总结如下:

（1）我们提出了一种新颖的特征对齐方法，即TDM，用于基于识别类别判别性和查询相关性通道来定义按类别划分的通道重要性，该方法专为细粒度少样本分类任务量身定制。

（2）我们进一步扩展QAM以引入IAM，将TDM的概念应用于特征提取器，这不仅在细粒度任务中补充了TDM，而且在包括粗粒度和跨领域少样本分类等更一般的场景中也有益处。

（3）我们通过在标准基准测试中取得新的最先进性能，实验验证了所提出的TDM和IAM对现有少样本分类模型的高度适用性及其优势。

# 2 相关工作

# 2.1 少样本分类

少样本图像分类研究主要有两条主线，即基于优化的方法和基于度量的方法。在早期，模型无关元学习(Model - Agnostic Meta - Learning，MAML)引入了基于优化的方法的概念，它学习适应新任务的良好初始条件。然后，元长短期记忆网络(Meta - LSTM)[42]使用基于长短期记忆网络(LSTM)的元学习器来确定通用初始点并进行有效微调。此外，元优化网络(MetaOpt - Net)[23]通过利用凸基学习器提供了一个端到端学习的微分过程。尽管这些基于优化的方法显示出了有希望的结果，但它们需要对新类别进行在线更新。

另一方面，基于度量的方法旨在通过采用预定义的[20]、[48]、[52]或在线训练的度量[49]来学习深度表示。其概念在匹配网络(MatchNet)[52]中被引入，该网络通过余弦相似度来推断查询集的类别。原型网络(ProtoNet)[48]进一步采用每个类别的平均特征作为原型，并利用它们来计算查询与每个类别之间的距离。关系网络(RelationNet)[49]没有使用预定义的度量，而是采用了一种可学习的距离度量。

如前所述，基于度量的方法通常试图减小属于同一类别的实例之间的距离。TDM是一个适用于这些基于度量的方法的模块，可提高它们的性能。具体而言，TDM能够基于自适应通道权重来测量距离，它可以动态地识别和强调判别通道，而现有技术对所有通道一视同仁。

# 2.2 特征对齐方法

在基于度量的分类中，特征对齐方法旨在进行有利于分类的距离计算。这些特征对齐方法可分为空间对齐方法和通道对齐方法。空间对齐方法[8]、[14]、[18]、[57]、[60]、[64]旨在对齐支持集和查询集的特征，以匹配对象区域。例如，CAN[14]为每对类别和查询特征图计算一个相关图，以突出对象可能存在的公共区域。CTX[8]通过注意力机制[2]测量查询实例与支持集之间的粗略空间对应关系，然后根据该对应关系为每个类别生成一个查询对齐的原型。FRN[57]通过利用岭回归的闭式解，根据查询实例的特征图重建支持集的特征图。

另一方面，通道对齐方法[18]、[47]、[60]、[62]会改变特征图，以便更好地区分新类别。具体而言，FEAT利用Transformer[31]、[51]处理支持集的特征图，以增加类别之间的距离。DMF[60]通过考虑支持集和查询对生成的动态元滤波器来对齐查询实例的每个特征图。此外，RENet[18]利用自相关和互相关对支持集和查询对的特征图进行变换，分别捕捉每个图像的结构模式并编码语义相关内容。

虽然TDM基本上是一种通道对齐方法，但与现有方法通常考虑每个类别支持集与查询实例之间的成对关系不同，TDM利用整个任务来调整特征图。

# 2.3 注意力模块

在各种下游任务中，注意力模块能显著提升性能[24]、[38]、[45]、[46]、[54]。现有的注意力方法可分为空间注意力[9]、[34]、[41]和通道注意力[15]、[58]。具体而言，为解决卷积神经网络(CNN)扩展性差的问题，SA[41]提出了一种空间注意力模块，通过基于相似度用特征图的每个网格与其他区域及其自身表示来捕捉图像中的长距离依赖关系。基于SA的成功，ViT[9]提出了仅由全连接层和多头注意力组成的网络架构，并展示了强大的空间注意力性能。

另一方面，SENet[15]提出了一种通道注意力模块，该模块通过一个简单的全连接块从空间平均特征中生成突出更具信息性通道的通道权重。CBAM[58]是另一个值得注意的通道注意力模块，它不仅使用了平均特征，还使用了最大池化特征。

在这两种注意力方法中，TDM和IAM属于通道注意力方法。然而，与现有的仅考虑单个实例信息来产生注意力结果的注意力方法不同，TDM利用整个任务为每个类别和查询实例计算一个通道权重向量。这是为了通过关注每个类别独特特征中与查询相关的特征来估计查询实例的类别。此外，IAM促使每个实例的特征图包含与对象相关的特征，以最小化TDM中背景的影响。因此，我们声称我们的模块专门用于少样本分类。

# 3 预备知识

# 3.1 问题表述

作为少样本分类问题的标准表述，我们有两个数据集:用于训练模型的元训练集 和用于评估已学习模型的元测试集 。 和 分别表示基础类别和新类别，它们不重叠(即 )。通常，少样本分类的训练和测试由情节(episode)组成。每个情节由随机采样的 个类别组成，每个类别由 个带标签的图像和 个无标签的图像组成，即 - 路 - 样本情节。带标签的图像称为支持集 ，无标签的图像称为查询集 ，而这两个集合是不相交的(即 )。支持集和查询集分别用于学习和验证。通常，通过利用支持集和查询集实例的特征图来预测查询实例的类别。如果我们将 定义为支持集中第 类的第 个实例，将 定义为查询实例，它们对应的特征图 和 表示如下:

其中 是以 为参数的特征提取器。每个特征图的形状为 ，其中 、 分别表示通道数、高度和宽度。

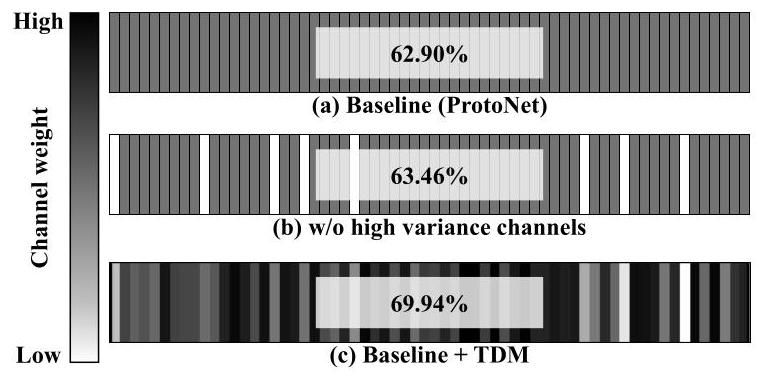


图1:CUB数据集中通道权重的影响。

子图的每一列代表通道权重，框中的数字是5 - 路1 - 样本场景的平均分类准确率，该准确率是通过从新类别中随机采10,000个情节进行评估得到的。(a) 基线方法平等对待特征图的通道，而不考虑类内通道方差(如公式(4)中定义的 )。在这种情况下，类内方差高的通道很可能会干扰类别的精确估计 —— 直观地说，同一类别的实例在某个通道上具有相似的特征会导致该通道的通道方差较低。(b) 因此，在评估阶段，我们只需为每个情节去除 值高的通道就可以得到改进(我们使用支持集和查询集计算 ，并去除 值最高的前12.5%的通道)。然而，在细粒度数据集的一个情节中，即使类别的特征图在所有通道上的 值都很低，这些特征图也可能没有针对该情节进行优化。这是因为类别共享相似的特征，例如CUB数据集中的羽毛和翅膀。(c) 因此，在细粒度数据集中，我们应该关注每个通道是否反映了不同的特征。TDM根据情节中每个类别的通道判别能力生成每个类别的通道权重。

# 3.2 动机

在基于度量的少样本学习[48]、[52]中，分类通常基于距离进行。假设这些距离是为通过特征图 的通道空间平均计算得到的 维向量 定义的，如下所示:

其中 是实例特征图 的第 个通道中位于 (h, w) 处的空间值。然后，我们按如下方式计算属于第 类的支持集和查询集的第 维的平均值:

其中 是属于第 类的实例数量， 表示第 类中第 个实例的 。基于这些，我们按如下方式定义第 类中第 个通道的逐通道方差 :

由于在向量间的距离计算中会比较相同维度的值，方差较小的维度对距离的贡献较小。因此，基于度量的少样本分类方法试图在训练阶段降低 。然而，即使训练阶段使用的类别形成了较低的 ，也不能保证验证阶段的新类别也是如此 [44]、[62]、[63]。

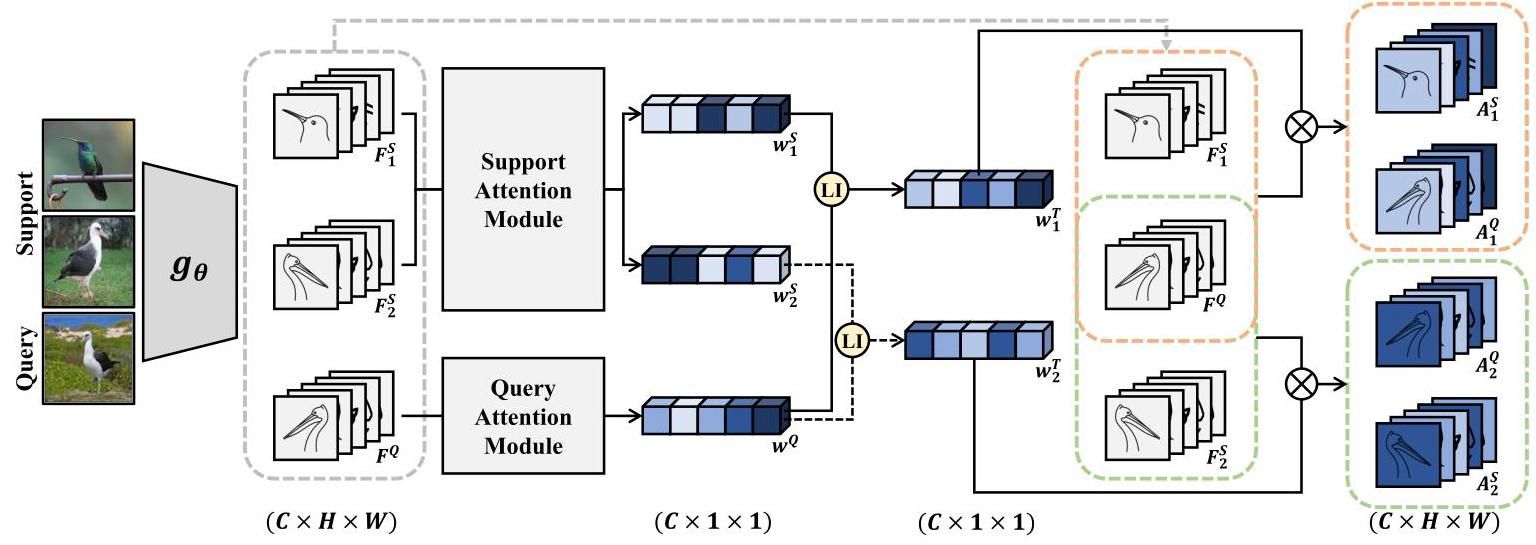


图 2:任务差异最大化(Task Discrepancy Maximization，TDM)概述。

TDM 由两个子模块组成，每个子模块都采用特征图 来生成通道权重向量 。支持注意力模块(Support Attention Module，SAM)接收支持实例的特征图作为输入，并为每个类别估计具有判别性的通道。然后，对于每个第 类，它会生成一个支持权重向量 ，该向量在这些通道中具有较高的值。另一方面，查询注意力模块(Query Attention Module，QAM)获取查询实例的特征图，并发现查询中与对象相关的通道。然后，来自 QAM 的查询权重向量 会强调那些值较高的通道。这两个子模块的权重向量通过线性插值组合，为每个第 类定义一个任务权重向量 。最后，通过将任务权重向量与原始特征图相乘，得到专注于判别区域的任务自适应特征图 。

因此，如图 1 (a) 所示，同等利用所有通道并不是一个合适的解决方案。这是因为如前所述，由于特征提取器是使用基类进行训练的，新类别中存在 较高的通道。因此，如图 1 (b) 所示，通过消除新类别中 较高的通道，仅利用 较低的通道是有效的，就像由 较低的通道组成的基类一样。

然而，在细粒度数据集中，类别属于同一个超类并共享共同特征，那么，即使通道的 较低，它们可能也不包含与其他类别的明显区别信息。因此，如图 1 (c) 所示，我们应该考虑每个通道的信息是否能与其他类别区分开来。为了实现这一点，我们将在第 4 节中介绍我们的两个新颖的通道注意力模块。此外，如第 5 节所述，我们将扩展其中一个模块以捕获实例描述性信息。

# 4 任务差异最大化:细粒度数据集的对齐方法

TDM 的整体架构如图 2 所示。给定一个由支持实例和查询实例组成的情节，首先由特征提取器计算特征图。由于特征提取器是为了找到区分基类的判别特征而训练的 [44]、[62]、[63]，因此特征图对于每个情节来说并不是最优的。为了为每个情节生成优化的特征图，TDM 通过计算表示特定任务的逐通道判别能力的特定任务权重向量来转换特征图。因此，TDM 的目标是将原始特征图细化为专注于判别细节的任务自适应特征图。在本节中，我们描述 TDM 的组件及其用途。首先，我们在 4.1 节中基于估计的显著区域定义两个逐通道代表性分数。然后，利用这些分数，我们在 4.2 节中介绍 TDM 的两个子模块:SAM 和 QAM

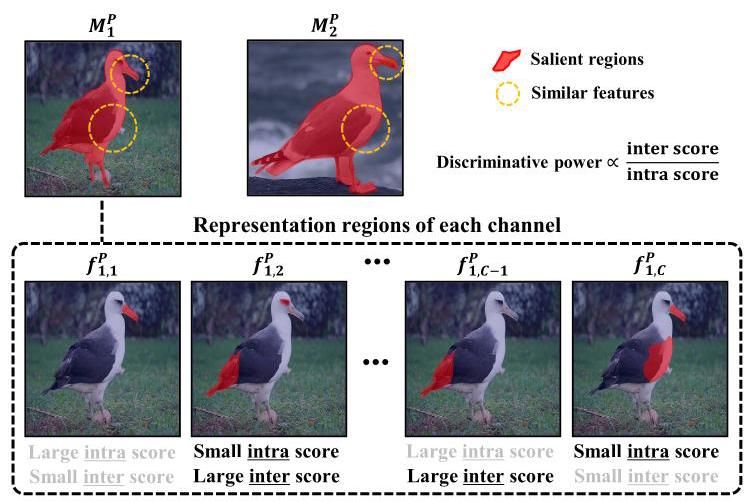


图3:通道代表性得分与判别能力之间的关系。

假设类别共享两个相似特征(胸部羽毛和喙)和两个不同特征(尾巴和眼睛)。如果一个通道对物体的各种特征进行编码，那么编码特征所覆盖的区域与显著区域高度匹配(即第2个和第C个通道)。在这种情况下，相应通道的 值较小(即 和 较小)。同样，虽然 与第 个通道对第 个物体类别的区域覆盖范围有关，但 表示第 个通道对第 个类别的判别能力。例如，(第 个通道的) 值较小，因为它只编码了另一个鸟类类别共享的特征(即胸部羽毛和喙是 和 类共享的相似特征)。相反， 值较大，因为第二个通道编码了高度可区分的特征(即尾巴和眼睛)。因此，具有判别性的通道应该是 值小且 值大。

# 4.1 通道代表性得分

给定支持集的特征图 ，对于第 类和第 个通道的每一对，我们定义两个通道代表性得分；类内得分 和类间得分《LaTeX类文件期刊》，第14卷，第8期，2015年8月 。由于每个类别可能有多个实例，我们使用一个原型[48]作为每个类别的代表。第 类的原型 计算如下:

其中 和 分别是每个类别的实例数量和第 类第 个实例的特征图。然后，对于每个原型，我们定义一个平均空间特征来表示显著物体区域。当第 类原型 的第 个通道表示为 时，相应的平均空间特征 计算如下:

在此基础上，我们进一步将第 类第 个通道的类内通道代表性得分 定义如下:

它表示第 个通道上的高激活区域与平均空间特征所表示的类显著区域的匹配程度。另一方面，第 类第 个通道的类间通道代表性得分 计算如下:

由于当 与 不同时得分较大，它表示该通道包含每个类别的判别信息的程度。直观地说，如图3所示，当一个通道的 值小且 值大时，它的区分度更高。

# 4.2 支持注意力模块(SAM，Support Attention Module)

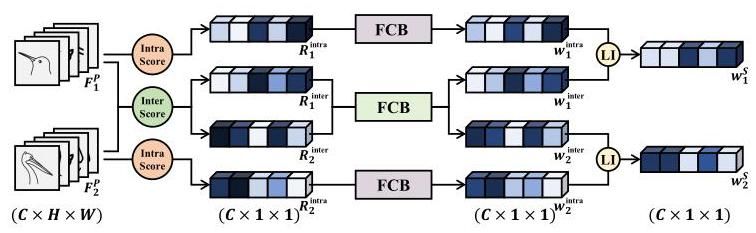


图4:支持注意力模块的示意图。

支持注意力模块(SAM)接收类别原型作为输入，并根据公式(7)和公式(8)为每个类别计算两个通道维度的代表性得分。然后，为了通过考虑这些得分的分布来反映每个通道的重要性，我们将两个得分向量 和 转换为第 类的两个权重向量 和 ，如下所示:

其中 和 表示用于生成两个权重向量的全连接块。它们的架构在表 1 中描述

第 类的支持权重向量 通过对相应的两个权重向量 和 进行线性插值得到，如下所示:

其中 是支持权重向量的平衡超参数。第 类的向量突出了第 类的独特通道，同时抑制了该情节中所有类别包含公共信息的通道。因此，当第 类的支持权重向量 与特征图相乘时，第 类的实例会聚集在一起，而其他实例则会与第 类分离。

# 4.3 查询注意力模块(QAM)

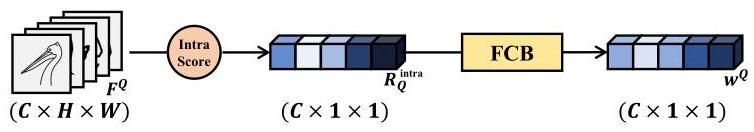


图 5:查询注意力模块的示意图

尽管第 类的支持权重向量 是为了强调第 类的独特通道而设计的，但属于第 类别的查询实例可能不具备这些特征。具体来说，如果查询实例没有与突出通道相对应的特征，支持集强调的判别性通道就会变得无用，甚至会干扰基于距离的查询类别预测。这个问题促使我们提出查询注意力模块(QAM)，以同时关注具有类别判别性且被查询实例所具备的通道。由于与支持集不同，我们没有查询实例的任何标签信息，QAM 仅利用查询实例内通道之间的关系。具体来说，QAM 计算查询实例内第 个通道的通道维度代表性得分 ，如下所示:

其中 表示特征图 的第 个通道， 是由 的通道维度平均值定义的平均空间特征。然后，查询权重向量 通过将内部得分向量 传递给全连接块 得到，如下所示:

其中 的架构在表 1 中描述。查询权重向量强调查询实例中与对象相关的通道，同时抑制其他通道。因此，查询权重向量引导模型关注与查询对象相关的信息。

# 4.4 任务差异最大化(TDM)

由于分别由 SAM 和 QAM 生成的两个权重向量 和 在目的上是互补的，我们利用它们来定义一个任务权重向量。具体来说，第 类的任务权重向量 通过对相应的支持和查询权重向量 和 进行线性插值定义，如下所示:

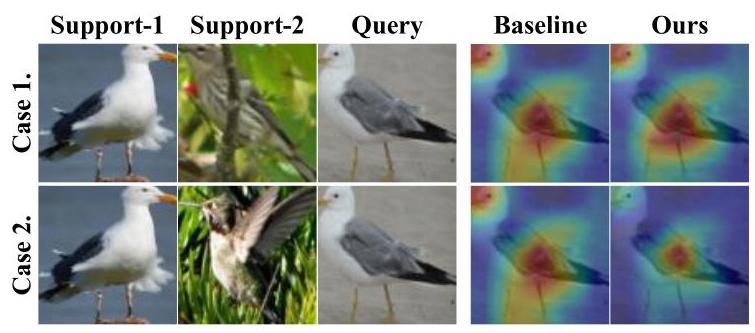


图 6:2 分类 1 样本任务的二维聚合特征激活图。(情况 1)如果不同物种的喙和翅膀不相似，TDM 认为喙和翅膀都是具有判别性的。(情况 2)然而，当鸟类的喙相似时，TDM 只将翅膀视为具有判别性的部分。

基于上述任务权重向量，将所有支持实例和查询实例的特征图转换为任务自适应特征图。具体而言，第 类支持实例的特征图 通过其对应的任务权重向量 转换为任务自适应特征图 ，如下所示:

其中， 和 分别是 的第 维的标量值和 的第 通道的标量值。另一方面，由于查询的标签不可用，因此无法指定应将哪个任务权重向量与查询的特征图相乘。因此，我们将所有任务权重向量 应用于查询的特征图 ，以生成关于所有类别的任务自适应特征图 ，如下所示:

其中， 表示类别索引， 是 的第 通道。当我们测试第 类的查询时，将使用查询对应的自适应特征图 。

例如，当将任务动态模块(TDM)应用于原型网络(ProtoNet)[48] 时，查询实例属于第 类的概率通过以下准则计算:

其中， 是距离度量， 是通过第 类支持实例的自适应特征图的平均值计算得到的原型。

# 4.5 讨论

对于一般数据集，众所周知，包含对象各种信息的特征图是有益的 [16]、[26]、[29]。另一方面，在细粒度数据集中，由于类别具有共同的整体外观，仅关注有区分性的区域是有利的 [7]、[11]、[33]、[67]。此外，在细粒度少样本分类中，与每个类别有区分性的区域几乎恒定的一般细粒度分类不同，每个类别的不同部分可能会根据情节的内容而变化。因此，动态发现情节中每个类别的不同部分是细粒度少样本分类的关键点。如图 6 所示，基线模型通过平等对待所有特征来估计查询的类别，而不考虑每个情节的构成。相比之下，任务动态模块(TDM)通过关注考虑情节后发现的有区分性的部分来预测查询的类别。这就是为什么任务动态模块(TDM)是专门为细粒度少样本分类设计的模块。

表 1:式 (9) 中的全连接块 和 以及式 (18) 中的 的架构。批量大小 B 和输入大小 C 在支持自适应模块(SAM)、查询自适应模块(QAM)和中间自适应模块(IAM)中有所不同。对于支持自适应模块(SAM)，B 是构成一个情节的类别数量， 是特征图 的通道数。另一方面，对于查询自适应模块(QAM)， 与支持自适应模块(SAM)相同，而 B 是查询的数量。在中间自适应模块(IAM)中，B 和 C 分别是中间特征图 的图像数量和通道数。

|  |  |  |
| --- | --- | --- |
| 全连接块(Fully Connected Block) | | |
| 层(Layer) | 输入形状(Input Shape) | 输出形状(Output Shape) |
| 全连接层(Fully Connected Layer) | 批次 通道数(B C) | B |
| 批量归一化(Batch Normalization) | B |  |
| 修正线性单元(ReLU) | B | B |
| 全连接层(Fully Connected Layer) | B | 批次 通道数(B C) |
| 1 + 双曲正切函数(1 + Tanh) |  | 批次 通道数(B C) |

# 5 实例注意力模块:广义查询注意力模块

由于TDM(任务驱动模块，Task-Driven Module)是为高层特征图(如最后一个卷积层)开发的，其性能可能依赖于特征提取器生成的给定特征的质量。在本节中，我们通过扩展QAM(查询注意力模块，Query Attention Module)进一步引入实例注意力模块(IAM，Instance Attention Module)，以便即使对于中间特征表示也能体现我们的设计动机。具体而言，IAM旨在为每个实例突出与对象相关的通道，而不考虑支持集或查询集。

IAM的整体架构如图7所示。IAM分别在特征提取器的中间层对每个实例进行操作，旨在强调编码与对象相关特征的通道。具体来说，IAM接收一个中间特征图 作为输入，其中 、 分别表示特征图的通道数、高度和宽度。然后，在特征图 中为第 个通道定义通道代表性得分，如下所示:

其中 和 分别是 的第 个通道和通过 的通道平均计算得到的平均空间特征。基于得分向量 ，IAM以与公式(12)中描述的QAM类似的方式推断通道权重向量 ，如下所示:

其中 是表1中描述的全连接块。随后，基于通道权重向量 ，中间特征图 被转换为注意力特征图 ，如下所示:

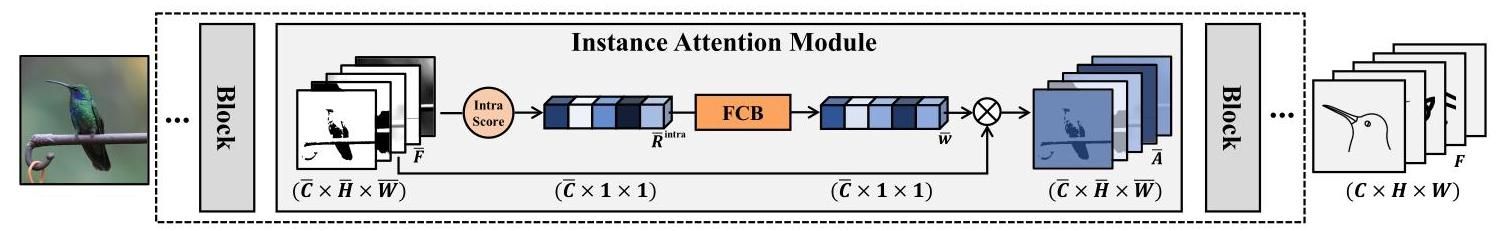


图7:实例注意力模块(IAM)的示意图。

虚线框表示图2中的特征提取器 。对于每个实例，IAM接收一个中间特征图 ，并根据每个通道 与显著对象区域之间的相似度计算通道代表性得分 。然后，它生成一个通道权重向量 ，该向量在实例的与对象相关的通道中具有较高的值。通过将中间特征图 的每个通道乘以其对应的权重 ，得到一个注意力特征图 。最后，应用了注意力 的特征图 被传递到下一层。

其中 是 的第 维的标量， 表示 的第 个通道。最后，转换后的特征图 被输入到特征提取器的下一层。

在IAM中，通道权重向量的计算和应用仅在每个实例内进行；在情景训练中不考虑支持集和查询集。因此，IAM逐实例地应用于所有支持实例和查询实例，以提高特征图的质量。此外，IAM中涉及的操作仅在一个实例内进行，额外的计算和内存开销很小，因此允许在骨干网络的中间块中使用它。

# 6 实验

在本节中，我们在细粒度分类基准上评估所提出的TDM，并进一步在细粒度和粗粒度基准上验证IAM的泛化能力。在本节的所有表格中，我们使用 表示基线的复现版本。

# 6.1 实现细节

**基线方法**。为了验证TDM(任务驱动度量，Task-Driven Metric)和IAM(实例注意力模块，Instance Attention Module)在细粒度分类问题中的有效性和适应性，我们将其应用于各种现有方法，包括原型网络(ProtoNet)[48]、深度语义网络(DSN，Deep Semantic Network)[47]、上下文网络(CTX，Context Network)[8]和特征关系网络(FRN，Feature Relation Network)[57]。另一方面，对于粗粒度分类问题，我们将IAM附加到ProtoNet、FRN和深度双对比网络(DeepBDC，Deep Bi-contrastive Network)[59]中。为了进行公平比较，我们按照FRN和DeepBDC中描述的相同超参数重新实现了每个基线模型。并且，无论是否应用TDM或IAM，都采用相同的训练和评估方案。虽然TDM通常利用公式(5)中定义的原型[48]来计算内部和外部得分，但在与CTX结合时，它会使用CTX [8]中提出的查询对齐原型。

**架构**。我们采用了近期少样本分类文献[5]、[13]、[21]、[65]、[66]中常用的模型架构；我们使用了Conv - 4和ResNet - 12。虽然这两个骨干网络都接受大小为 的图像，但根据骨干网络的不同，特征图的大小也不同。具体来说，ResNet - 12生成的特征图维度为 ，而Conv - 4生成的特征图形状为 。对于我们提出的TDM和IAM，我们额外使用了全连接层块，块的大小与特征图的通道数成正比，如表1所述。我们将IAM附加到第一个和第二个块上。公式(10)、公式(13)中的 固定为0.5。

表2:数据集的划分。其中 是总类别数， 分别是训练、验证和测试类别的数量。这些子集的类别是不相交的。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| CUB - 200 - 2011数据集 | 200 | 100 | 50 | 50 |
| 飞机 | 100 | 50 | 25 | 25 |
| 元iNaturalist数据集(meta - iNat) | 1135 | 908 | - | 227 |
| 分层元iNaturalist数据集(tiered meta - iNat) | 1135 | 781 | - | 354 |
| 斯坦福汽车数据集(Stanford - Cars) | 196 | 130 | 17 | 49 |
| 斯坦福犬类数据集(Stanford - Dogs) | 120 | 70 | 20 | 30 |
| 牛津宠物数据集(Oxford - Pets) | 37 | 20 | 7 | 10 |
| 迷你ImageNet数据集(mini - ImageNet) | 100 | 64 | 20 | 16 |
| 分层ImageNet数据集(tiered - ImageNet) | 608 | 351 | 97 | 160 |

**训练细节**。遵循基线方法[3]、[55]、[57]、[62]、[64]，我们使用标准的数据增强技术，包括随机裁剪、水平翻转和颜色抖动。为防止过拟合，我们在TDM(时域调制，Time Domain Modulation)和IAM(交互式注意力模型，Interactive Attention Model)的每个输出中添加-0.2到0.2之间的随机噪声。我们还将模块的每个输出调节到 范围内。为了进行公平比较，无论是否使用TDM或IAM，超参数和训练细节均遵循我们的基线设置。

**评估细节**。对于5分类 样本的实验，我们使用10,000个随机采样的情节进行评估，每个情节包含每个类别16个查询样本。我们报告具有95%置信区间的平均分类准确率。由于DSN(深度支持网络，Deep Support Network)和CTX(上下文模型，Context Model)在5样本情节训练的模型下表现更好，如同FRN(特征归一化网络，Feature Normalization Network)一样，因此它们的单样本性能通过5样本情节训练的模型来衡量。

# 6.2 数据集

我们使用七个基准数据集进行细粒度少样本分类:CUB - 200 - 2011、飞机数据集(Aircraft)、元自然历史图像数据集(meta - iNat)、分层元自然历史图像数据集(tiered meta - iNat)、斯坦福汽车数据集(Stanford - Cars)、斯坦福犬类数据集(Stanford - Dogs)和牛津宠物数据集(Oxford - Pets)。在粗粒度场景评估中，使用小ImageNet(mini - ImageNet)和分层ImageNet(tiered - ImageNet)数据集。每个数据集的划分信息见表2。

CUB - 200 - 2011 [53]包含200种鸟类的11,788张照片。该数据集有两种使用形式:原始形式[3]或经过人工标注边界框预处理的形式[62]、[64]。在我们的工作中，如文献[3]、[57]一样，对两种形式都进行了实验。

表3:以边界框裁剪图像作为输入在CUB数据集上的性能。“\*”表示在RENet中复现的结果。我们实现的模型的置信区间均低于0.23。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型 | 卷积-4 | | 残差网络-12(ResNet-12) | |
| 单样本(1-shot) | 五样本(5-shot) | 单样本(1-shot) | 五样本(5-shot) |
| 匹配网络(MatchNet) [52], [62], [64] | 67.73 | 79.00 | 71.87 | 85.08 |
|  | 68.87 | 82.90 | 73.27 | 85.77 |
| 深度地球移动距离(DeepEMD) [64] | - | - | 75.65 | 88.69 |
| 循环嵌入网络(RENet) [18] | - | - | 79.49 | 91.11 |
| 原型网络(ProtoNet) [48] | 62.90 | 84.13 | 78.99 | 90.74 |
| + 时域调制(TDM) | 69.94 | 86.96 | 79.58 | 91.28 |
| + 图像注意力模块(IAM) | 68.18 | 85.96 | 79.65 | 91.20 |
| + 时域调制(TDM) + 图像注意力模块(IAM) | 72.96 | 88.02 | 80.93 | 91.80 |
| [47] | 72.09 | 85.03 | 80.51 | 90.23 |
| + 时域调制(TDM) | 73.38 | 86.07 | 81.33 | 90.65 |
| + 图像注意力模块(IAM) | 75.10 | 86.66 | 82.03 | 90.67 |
| + 时域调制(TDM) + 图像注意力模块(IAM) | 74.75 | 86.89 | 82.85 | 91.47 |
| [8] | 72.14 | 87.23 | 80.67 | 91.55 |
| + 时域调制(TDM) | 74.68 | 88.36 | 83.28 | 92.74 |
| + 图像注意力模块(IAM) | 75.65 | 89.07 | 82.87 | 92.49 |
| + 时域调制(TDM) + 图像注意力模块(IAM) | 77.17 | 89.90 | 83.76 | 92.85 |
| [57] | 73.24 | 88.33 | 83.16 | 92.42 |
| + 时域调制(TDM) | 74.39 | 88.89 | 83.36 | 92.80 |
| + 图像注意力模块(IAM) | 76.29 | 89.66 | 83.63 | 92.59 |
| + 时域调制(TDM) + 图像注意力模块(IAM) | 75.49 | 89.72 | 84.17 | 93.30 |

**飞机(Aircraft)**[35]数据集包含100个飞机类别、共10000张图像。该数据集的主要挑战来自航空公司标志。具体而言，尽管飞机型号不同，但它们的航空公司标志可能相同，这使得识别任务更加困难。我们在划分训练/测试数据以及使用边界框进行图像预处理时，遵循了我们的基线模型FRN的方法。

**元iNat(meta - iNat)**[50]、[56]是一个长尾数据集。它包含1135种动物物种，每个类别的图像数量不均匀，范围在50到1000之间。对于训练和测试数据的划分，我们采用了[56]中介绍的方法，该文献最初提出了这个用于少样本分类的基准。然而，与[56]采用227路评估方案不同，我们遵循基线模型FRN，采用标准的5路评估方案。

**分层元iNat(tiered meta - iNat)**[56]与元iNat具有相同的图像。然而，不同之处在于训练和测试数据的组织方式；与元iNat不同，分层版本按超类别进行划分。因此，训练类和测试类之间存在更大的领域差距。

**斯坦福汽车(Stanford Cars)**[22]数据集包含196个汽车类别的16185张图像。我们采用了与[30]相同的数据划分协议，该文献首次将此数据集用于少样本分类任务。

**斯坦福犬类(Stanford Dogs)**[19]数据集包含来自全球120个犬种的20580张图像。与斯坦福汽车数据集类似，[30]也将其引入用于细粒度少样本分类。因此，我们遵循[30]的方法来划分这个数据集。

**牛津宠物(Oxford Pets)**[40]是另一个细粒度图像数据集，有37个宠物类别，每个类别约有200张图像。据我们所知，在本文的上一会议版本[24]之前，该数据集从未用于少样本分类任务。因此，我们像[24]中那样随机划分类别来定义训练/测试集。

**迷你ImageNet(mini - ImageNet)**[52]是少样本分类的代表性基准之一。它是ImageNet的一个子集，包含100个类别，每个类别有600张不同的图像。我们的数据集划分方法取自[52]。与上述用于验证TDM和IAM在细粒度分类中有效性的数据集不同，我们使用迷你ImageNet来评估IAM的泛化能力。

表4:以原始图像作为输入在CUB数据集上的性能。

|  |  |  |  |
| --- | --- | --- | --- |
| 模型 | 骨干网络 | 单样本(1-shot) | 五样本(5-shot) |
| 基线模型 [3] | 残差网络-18(ResNet-18) |  |  |
| 改进基线模型 [3] | 残差网络-18(ResNet-18) |  |  |
| 匹配网络 [3], [52] | 残差网络-18(ResNet-18) | 73.42±0.89 |  |
| 模型无关元学习 [3], [10] | 残差网络-18(ResNet-18) | 68.42±1.07 |  |
| 关系网络 [3], [49] | 残差网络-18(ResNet-18) |  |  |
| 半监督元学习 | 残差网络-18(ResNet-18) |  |  |
| 负余弦 [32] | 残差网络-18(ResNet-18) | 72.66±0.85 |  |
| 阿夫拉西亚比等人 [1] | 残差网络-18(ResNet-18) | 74.22±1.09 |  |
| 原型网络 [48] | 残差网络-12(ResNet-12) |  |  |
| + 时域调制(TDM) | 残差网络-12(ResNet-12) | 79.11±0.22 |  |
| + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) |  | 90.72±0.12 |
| +时域调制(TDM) + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) |  | 90.86±0.12 |
| [47] | 残差网络-12(ResNet-12) |  | 89.92±0.12 |
| + 时域调制(TDM) | 残差网络-12(ResNet-12) |  |  |
| + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) |  |  |
| +时域调制(TDM) + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) |  | 90.54±0.12 |
|  | 残差网络-12(ResNet-12) |  | 91.54±0.11 |
| + 时域调制(TDM) | 残差网络-12(ResNet-12) |  |  |
| + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) |  |  |
| +时域调制(TDM) + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) |  |  |
| [57] | 残差网络-12(ResNet-12) |  |  |
| + 时域调制(TDM) | 残差网络-12(ResNet-12) |  | 93.37±0.10 |
| + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) |  | 93.21±0.10 |
| +时域调制(TDM) + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) |  |  |

表5:飞机性能。我们实现的模型的置信区间均低于0.25。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型 | 卷积四层网络(Conv-4) | | 残差网络-12(ResNet-12) | |
| 单样本(1-shot) | 五样本(5-shot) | 单样本(1-shot) | 五样本(5-shot) |
| 原型网络(ProtoNet) [48] | 47.37 | 68.96 | 67.28 | 83.21 |
| + 时域调制(TDM) | 50.55 | 71.12 | 69.12 | 84.77 |
| + 图像注意力模块(IAM) | 49.67 | 68.57 | 69.10 | 84.04 |
| +时域调制(TDM) + 图像注意力模块(IAM) | 52.88 | 72.81 | 69.80 | 85.41 |
| [47] | 52.22 | 68.75 | 70.23 | 83.05 |
| + 时域调制(TDM) | 53.77 | 69.56 | 71.57 | 83.65 |
| + 图像注意力模块(IAM) | 54.62 | 68.87 | 72.01 | 83.36 |
| +时域调制(TDM) + 图像注意力模块(IAM) | 54.64 | 70.34 | 73.83 | 85.11 |
|  | 51.58 | 68.12 | 65.53 | 79.31 |
| + 时域调制(TDM) | 55.15 | 70.45 | 69.42 | 83.25 |
| + 图像注意力模块(IAM) | 54.70 | 70.61 | 70.93 | 82.38 |
| +时域调制(TDM) + 图像注意力模块(IAM) | 57.04 | 72.46 | 71.40 | 84.12 |
|  | 53.12 | 70.84 | 69.58 | 82.98 |
| + 时域调制(TDM) | 54.21 | 71.37 | 70.89 | 84.54 |
| + 图像注意力模块(IAM) | 54.98 | 72.12 | 71.23 | 83.66 |
| +时域调制(TDM) + 图像注意力模块(IAM) | 56.08 | 72.62 | 72.36 | 85.05 |

**分层ImageNet(tiered-ImageNet)** [43] 也是ImageNet的一个子集，但它是小样本分类领域最大的数据集。它包含601个类别，远多于mini-ImageNet的类别数量。此外，与mini-ImageNet不同，该基准数据集按超类别划分训练、评估和测试类别，因此，与分层元iNat(tiered meta-iNat)一样存在较大的领域差距。我们利用这个基准数据集来验证交互式注意力模块(IAM)，因为它是一个粗粒度分类数据集。

# 6.3 细粒度小样本分类

CUB - 200 - 2011数据集结果。表3和表4报告了我们的基线模型结果，以及将我们提出的模块——任务依赖模块(TDM)和交互式注意力模块(IAM)与之结合后的性能表现。在表3的裁剪图像实验中，TDM和IAM在所有情况下都持续提升了基线模型的性能，并且在二者同时应用时取得了最先进的分数。尽管在表4的一些设置中，IAM的效果不太明显，但TDM和IAM结合使用时，仍然表现出了优越的性能。

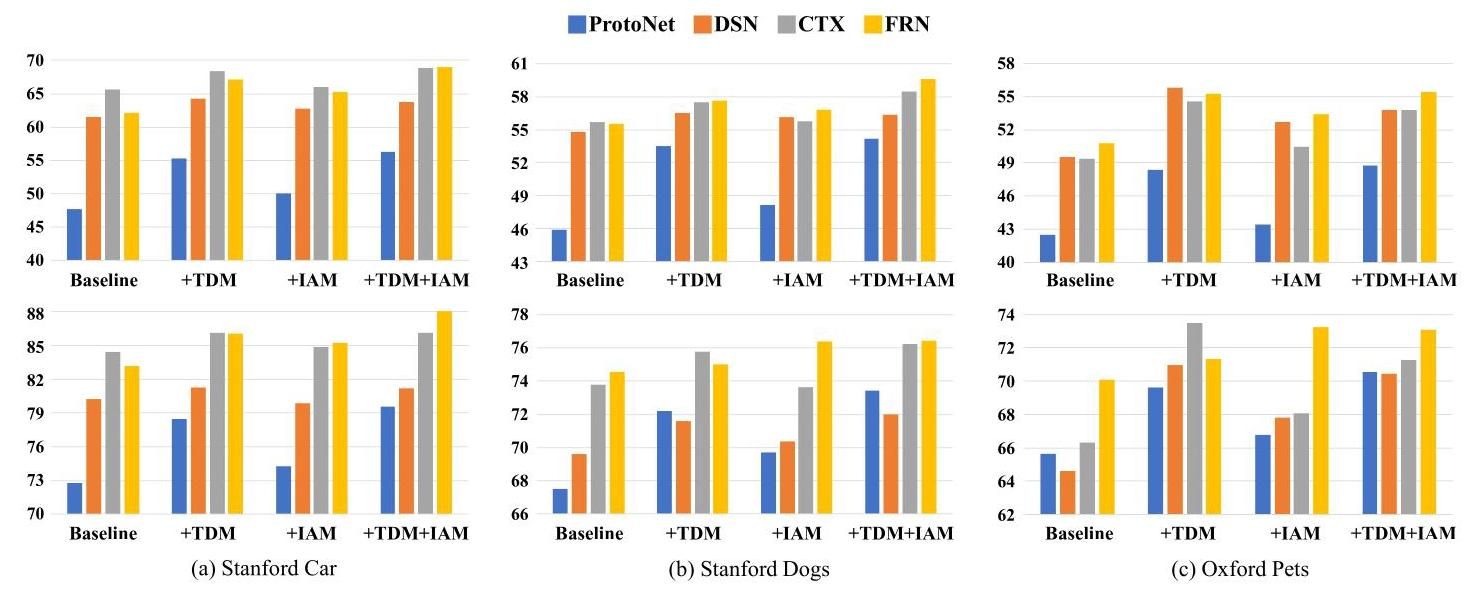


图8:(a)斯坦福汽车数据集(Stanford Car)、(b)斯坦福犬类数据集(Stanford Dogs)和(c)牛津宠物数据集(Oxford Pets)的实验结果。

每个数据集的上下两个图分别表示单样本(1-shot)和五样本(5-shot)的性能表现。每个图第一列的条形图报告了基线模型的准确率。第二列和第三列的条形图分别表示TDM和IAM与基线模型结合时的准确率。最后一列的条形图表示同时使用TDM和IAM时的性能表现。

**飞机数据集(Aircraft)结果**。如表5所示，除一种情况外，TDM和IAM无论模型大小和支持集大小如何，都能提升所有基线模型的性能，表现出一致的趋势。虽然在使用Conv - 4骨干网络的原型网络(ProtoNet)的五样本场景中，使用IAM时性能略有下降，但与仅使用TDM相比，同时采用IAM和TDM时能取得更出色的性能。

**元iNat(meta-iNat)和分层元iNat(tiered meta-iNat)数据集结果**。这些数据集适合评估模型的泛化能力，因为众所周知，在这些数据集上训练的模型由于缺乏验证集 [17]、[37]、[61] 而容易过拟合。此外，分层元iNat使得任务更加困难，因为训练集和测试集的超类别不重叠。如表6所示，TDM和IAM对过拟合和大领域差距具有鲁棒性，因为它们在大多数配置下都能提升结果。在分层元iNat的五样本场景中，TDM与特征归一化网络(FRN)结合时性能略有下降，我们认为这主要是由于FRN中可学习参数 的原因。一般来说，当存在领域差距时，较大的 表现出较好的性能。然而，我们发现TDM会使 相对较小，因为TDM有助于关注有区分性的通道。因此，除上述情况外，当同时使用TDM和IAM时，我们取得了最先进的性能。

**斯坦福汽车数据集(Stanford Cars)、斯坦福犬类数据集(Stanford Dogs)和牛津宠物数据集(Oxford Pets)结果**。虽然这些数据集包含细粒度类别，但在相关文献中并未用于评估我们的基线模型。为了进一步验证TDM和IAM的有效性，我们额外使用Conv - 4在这些数据集上进行了实验。如图8所示，除了少数仅使用IAM的情况外，TDM和IAM总体上提升了性能。具体而言，与基线模型相比，TDM在单样本和五样本场景下分别将准确率提高了4.44和3.27个百分点。另一方面，IAM分别带来了1.69和1.46个百分点的性能提升。此外，由于我们的模块相互兼容，当同时采用它们时，我们观察到性能有显著提升；在单样本和五样本情况下，准确率分别提升了4.79和3.81个百分点。

表6:使用Conv - 4骨干网络在元iNat和分层元iNat数据集上的性能表现。我们实现的模型的置信区间均低于0.23。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型 | 元iNat(meta-iNat) | | 分层元iNat(tiered meta-iNat) | |
| 单样本(1-shot) | 五样本(5-shot) | 单样本(1-shot) | 五样本(5-shot) |
| 原型网络(ProtoNet) [48] | 55.37 | 76.30 | 34.41 | 57.60 |
| + 任务动态模块(TDM) | 61.82 | 79.95 | 38.30 | 61.18 |
| + 图像注意力模块(IAM) | 59.12 | 79.83 | 37.94 | 63.47 |
| + 任务动态模块(TDM) + 图像注意力模块(IAM) | 65.10 | 81.93 | 41.87 | 64.32 |
|  | 60.06 | 76.15 | 40.83 | 58.34 |
| + 任务动态模块(TDM) | 61.87 | 78.07 | 41.00 | 58.66 |
| + 图像注意力模块(IAM) | 63.41 | 77.76 | 44.05 | 61.45 |
| + 任务动态模块(TDM) + 图像注意力模块(IAM) | 62.99 | 78.84 | 43.39 | 61.69 |
|  | 60.80 | 78.57 | 42.24 | 60.54 |
| + 任务动态模块(TDM) | 63.26 | 80.75 | 43.90 | 62.29 |
| +图像注意力模块(IAM) | 63.80 | 80.97 | 45.87 | 64.92 |
| +任务动态模块(TDM) + 图像注意力模块(IAM) | 64.96 | 81.89 | 47.40 | 66.12 |
| [57] | 61.98 | 80.04 | 43.95 | 63.45 |
| + 任务动态模块(TDM) | 63.97 | 81.60 | 44.05 | 62.91 |
| +1AM | 65.11 | 82.43 | 47.33 | 67.48 |
| +任务动态模块(TDM) + 图像注意力模块(IAM) | 65.95 | 83.30 | 46.45 | 66.55 |

在对七个基准数据集进行的大量实验中，我们验证了TDM(特征动态调制模块)和IAM(重要性感知模块)在细粒度分类任务中的优势。总结实验结果，无论使用何种数据集和基线方法，TDM都显示出了其优越性。对于IAM而言，尽管它有助于突出特征图中与目标相关的通道，但有时这对细粒度分类并无益处，因为细粒度数据集中的目标在不同类别之间可能存在过多的共同特征。然而，当IAM和TDM一起使用时，这个问题就不复存在了，因为TDM会在IAM识别出的与目标相关的特征中突出具有类别区分性的特征。另一方面，IAM有助于TDM发现更多具有区分性的特征，因为IAM为TDM提供了更多聚焦于目标的特征图。因此，我们认为TDM和IAM具有互补优势。

表7:IAM在mini - ImageNet和tiered - ImageNet数据集上的性能。FRN - EMD表示在DeepEMD中实现的FRN(特征归一化层)。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 模型 | 骨干网络 | 迷你图像网(mini-ImageNet) | | 分层图像网(tiered-ImageNet) | |
| 单样本(1-shot) | 五样本(5-shot) | 单样本(1-shot) | 五样本(5-shot) |
| 匹配网络(MatchNet) [52], [62], [64] | 残差网络-12(ResNet-12) |  |  |  |  |
| 基线++(Baseline++) [3] | 残差网络-12(ResNet-12) | 60.56±0.45 | 77.40±0.34 | - | - |
| CTM [28] | 残差网络-18(ResNet-18) |  |  |  |  |
| TADAM [39] | 残差网络-12(ResNet-12) |  |  | - | - |
| S2M2 [36] | 残差网络-12(ResNet-12) |  |  | - | - |
| 负余弦(Neg-Cosine) [32] | 残差网络-12(ResNet-12) |  |  | - | - |
| 阿夫拉西亚比等人(Afrasiyabi et al.) [1] | 残差网络-12(ResNet-12) |  |  |  |  |
| FEAT [62] | 残差网络-12(ResNet-12) |  |  |  |  |
| 深度地球移动距离(DeepEMD) [64] | 残差网络-12(ResNet-12) |  |  | 71.16±0.87 |  |
| 原型网络(ProtoNet) [48], [59] | 残差网络-12(ResNet-12) |  |  | - | - |
| + 交互式注意力模块(IAM) | 残差网络-12(ResNet-12) |  | 79.44±0.15 | - | - |
|  | 残差网络-12(ResNet-12) |  |  |  |  |
| + 交互式注意力模块(IAM) | 残差网络-12(ResNet-12) |  |  |  |  |
| 特征归一化-地球移动距离(FRN-EMD) [57], [64] | 残差网络-12(ResNet-12) | - | - |  |  |
| + 交互式注意力模块(IAM) | 残差网络-12(ResNet-12) | - | - |  |  |
| 元深度双相关(Meta DeepBDC)† [59] | 残差网络-12(ResNet-12) |  |  | - | - |
| + 交互式注意力模块(IAM) | 残差网络-12(ResNet-12) |  |  | - | - |
| 单任务学习深度双相关(STL DeepBDC)† [59] | 残差网络-12(ResNet-12) |  |  | - | - |
| + 交互式注意力模块(IAM) | 残差网络-12(ResNet-12) |  |  | - | - |

表8:模型在mini-ImageNet数据集上训练并在CUB数据集上测试的跨领域少样本分类性能。

|  |  |  |  |
| --- | --- | --- | --- |
| 模型 | 骨干网络 | 单样本(1-shot) | 五样本(5-shot) |
| 基线方法 [3] | 残差网络-18(ResNet-18) | - | 51.34±0.72 |
| 改进基线方法 [3] | 残差网络-18(ResNet-18) | - |  |
| 模型无关元学习(MAML) [3], [10] | 残差网络-18(ResNet-18) | - |  |
| 阿夫拉西亚比等人 [1] | 残差网络-18(ResNet-18) |  | 70.37±1.02 |
| 原型网络(ProtoNet) | 残差网络-12(ResNet-12) |  |  |
| + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) |  |  |
|  | 残差网络-12(ResNet-12) |  |  |
| + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) |  | 75.76±0.18 |
| 元深度双相关网络(Meta DeepBDC)† [59] | 残差网络-12(ResNet-12) |  |  |
| + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) |  | 77.71±0.15 |
| 单任务学习深度双相关网络(STL DeepBDC)† [59] | 残差网络-12(ResNet-12) |  |  |
| + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) |  |  |

表9:模型在mini-ImageNet数据集上训练并在Aircraft数据集上测试的跨领域小样本分类性能。与表8中的结果不同，此表中的STL DeepBDC(基于自训练学习的深度双对比学习)未执行蒸馏阶段，因为跳过这些阶段表现出更好的性能。

|  |  |  |  |
| --- | --- | --- | --- |
| 模型 | 骨干网络 | 单样本(1-shot) | 五样本(5-shot) |
| 原型网络(ProtoNet) [48],[59] | 残差网络-12(ResNet-12) |  |  |
| + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) |  |  |
| [57] | 残差网络-12(ResNet-12) |  |  |
| + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) | 39.77±0.17 | 63.61±0.18 |
| 元深度双相关分类器(Meta DeepBDC)† [59] | 残差网络-12(ResNet-12) |  | 59.52±0.19 |
| + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) | 37.46±0.17 |  |
| 单任务学习深度双相关分类器(STL DeepBDC)† [59] | 残差网络-12(ResNet-12) |  |  |
| + 图像注意力模块(IAM) | 残差网络-12(ResNet-12) |  |  |

# 6.4 粗粒度少样本分类

**mini-ImageNet和tiered-ImageNet数据集的实验结果**。正如第4.5节所讨论的，TDM(特征动态调制模块，Temporal Dynamic Modulation)可能不是粗粒度少样本分类任务的合适模块，因为它限制了对物体各种特征的利用。另一方面，由于IAM(实例注意力模块，Instance Attention Module)鼓励特征提取器为每个实例生成各种与物体相关的特征，我们认为IAM也有利于粗粒度少样本分类。为了验证IAM在粗粒度基准测试中的有效性，我们在mini-ImageNet和tiered-ImageNet数据集上进行了实验，结果如表7所示。可以看出，无论采用何种训练方案，IAM都能提升所有基线模型的性能。除了有效性之外，我们还强调了IAM的高适用性，因为这些结果是在没有进行任何广泛的超参数搜索或优化过程的情况下获得的。

# 6.5 跨领域少样本分类

**mini-ImageNet CUB-200-2011数据集的实验结果**。为了评估少样本分类算法的跨领域泛化能力，我们按照文献[3]、[57]的协议，在每个模型的训练集和测试集不同的情况下对其进行验证。由于细粒度类别的图像通常是由各个领域的专业人员收集的，我们认为这种设置与降低标注成本密切相关。具体来说，我们使用mini-ImageNet数据集训练每个模型，并使用CUB(原始形式)数据集对其进行验证，这与文献[3]、[57]中的做法相同。如表8所示，IAM持续提升了所有基线模型的性能，并且在没有任何适应过程的情况下达到了当前最优水平。

**mini-ImageNet 飞机数据集的实验结果**。尽管mini-ImageNet和CUB数据集之间存在较大的领域差距，但mini-ImageNet的训练类别中仍然包含两种不同类别的鸟类。因此，使用mini-ImageNet训练的每个模型可能已经学会了区分鸟类物种。另一方面，由于mini-ImageNet的训练集中没有飞机图像，对飞机类型进行分类是评估模型跨领域泛化能力的更合适设置。具体来说，我们在飞机数据集的测试集上评估使用mini-ImageNet训练的每个模型。如表9所示，我们的IAM即使在训练阶段从未见过的类别中也显示出了有效性。

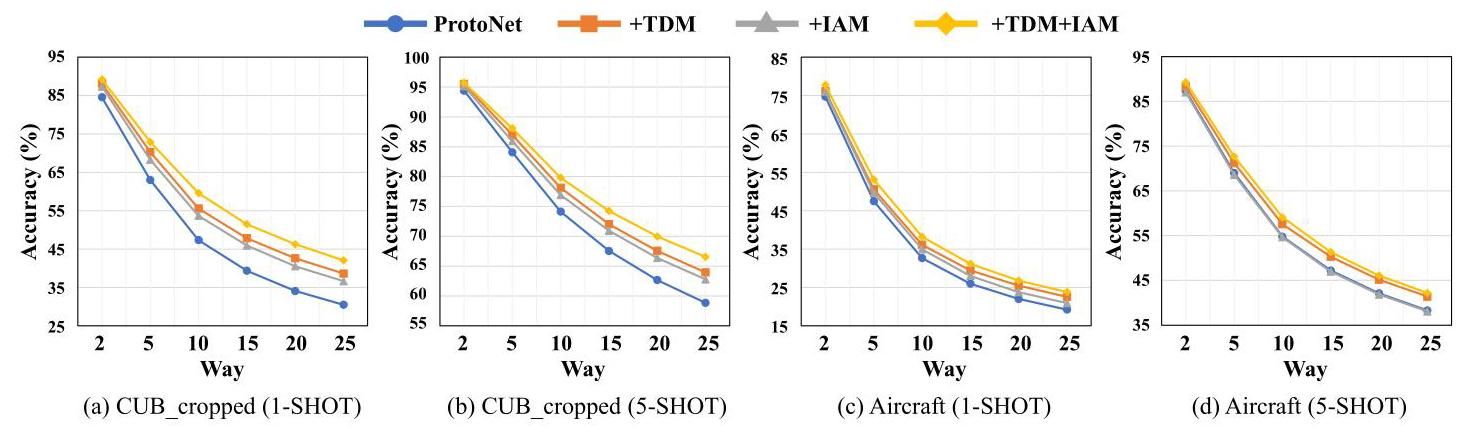


图9:不同N值下的N - 路1 - 样本和5 - 样本分类实验结果。

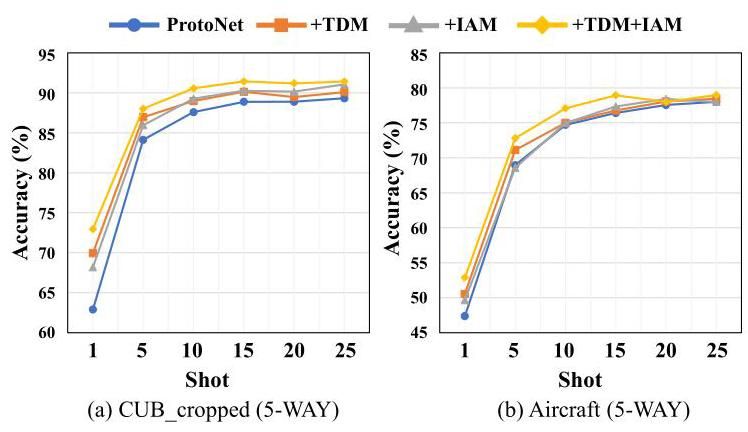


图10:不同K值下的5 - 路K - 样本分类结果。

# 7 消融实验

在本节中，我们进行消融实验。大多数消融实验是基于使用Conv - 4骨干网络的ProtoNet [48]，并使用CUB\_cropped和飞机数据集进行的。

# 7.1 改变 和 进行 - 路 - 样本分类

在第6节中，我们在各种场景下广泛验证了我们提出的模块的优点。然而，根据现有工作[48]、[52]、[57]、[59]、[64]的协议，这些实验中每个情节的类别数量固定为5。在现实场景中，类别数量可能会根据情况而变化。因此，为了验证TDM和IAM在这种场景中的有效性，我们首先评估在每个情节中包含不同数量类别 时我们的模块的性能。如图9所示，除了飞机基准测试中5 - 样本的一种情况外，与基线相比，TDM和IAM都能持续提升性能。此外，相对性能提升与类别数量成正比。这清楚地表明，我们的模块在更困难的设置(即更多类别)中更有效。

另一方面，根据现有方法[3]、[10]、[18]、[62]，每个类别的标注图像数量 在 范围内。与关于类别数量的实验类似，我们进行了标注图像数量不同的实验，结果如图10所示。如报告所示，就相对性能提升而言，我们的模块的优势在少样本情况下尤为突出。这验证了我们的模块更适合少样本场景，这也是本研究的主要任务，同时也显示了它们在多样本情况下的有效性。

表10:对SAM(语义注意力模块，Semantic Attention Module)、QAM(查询注意力模块，Query Attention Module)和IAM的消融实验。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 顺序注意力匹配器(Sequential Attention Matcher，SAM) | 正交幅度调制(Quadrature Amplitude Modulation，QAM) |  | 加州理工大学鸟类数据集裁剪版(CUB\_cropped) | | 飞机 | |
| 单样本(1-shot) | 五样本(5-shot) | 单样本(1-shot) | 五样本(5-shot) |
| - | - | - | 62.90 | 84.13 | 47.37 | 68.96 |
| ✓ | - | - | 68.53 | 85.95 | 49.45 | 69.33 |
| - | ✓ | - | 65.11 | 84.82 | 48.96 | 70.85 |
| - | - | ✓ | 68.18 | 85.96 | 49.67 | 68.57 |
| ✓ | ✓ | - | 69.94 | 86.96 | 50.55 | 71.12 |
| ✓ | - | ✓ | 71.97 | 88.00 | 51.98 | 70.67 |
| - | ✓ | ✓ | 68.87 | 86.68 | 51.69 | 69.60 |
| ✓ | ✓ | ✓ | 72.96 | 88.02 | 52.88 | 72.81 |

# 7.2 关于SAM、QAM和IAM的消融研究

由于我们的方法由三个子模块组成，即SAM、QAM和IAM，我们对这些子模块的各种组合进行了实验，以评估每个组件的贡献并确认它们的互补优势。如表10所示，除了IAM的一种情况外，每个子模块都持续提高了各个数据集的分类准确率。SAM带来的显著提升证实了，为每个类别识别并聚焦于有区分性的通道对于细粒度少样本分类至关重要。此外，尽管与SAM相比，QAM带来的提升略小，但QAM在所有测试配置中也被证明是有效的。这证实了对查询实例所具备的支持集特征赋予更多重要性的好处。另一方面，IAM的效果时有变化，在飞机数据集的5样本场景中，它可能会降低性能。这是因为当类别共享许多特征时，IAM诱导的与对象相关的特征图可能会阻碍准确的预测。然而，从第六行可以看出，SAM能够通过抑制公共特征并发现有区分性的特征来解决IAM的局限性。最重要的是，当同时使用这三个组件时，能取得最佳性能。这些结果验证了每个子模块的优点及其互补优势。

# 7.3 与余弦距离的兼容性

在本文中，与我们的基线方法[8]、[47]、[48]、[57]一样，我们在计算实例间的相似度时主要采用欧几里得距离来评估我们的方法。同时，余弦距离是其他技术[3]、[4]、[52]中采用的另一种常用度量。因此，我们还验证了我们的方法与余弦距离的兼容性，结果如表11所示。采用IAM或TDM会带来显著性能提升这一一致趋势证实了我们的方法与余弦距离度量的兼容性。

表11：与余弦距离的兼容性

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 方法 | 加州理工大学鸟类数据集裁剪版(CUB\_cropped) | | 飞机 | |
| 单样本(1-shot) | 五样本(5-shot) | 单样本(1-shot) | 五样本(5-shot) |
| 原型网络(ProtoNet) | 68.69 | 82.89 | 48.36 | 63.45 |
| + 时域调制(TDM) | 69.90 | 84.95 | 51.51 | 68.35 |
| + 图像注意力模块(IAM) | 69.72 | 84.28 | 50.69 | 66.13 |
| +时域调制(TDM) + 图像注意力模块(IAM) | 71.17 | 85.15 | 52.62 | 69.62 |

# 7.4 与现有注意力方法的比较

为了进一步验证我们的方法在细粒度少样本分类任务中相对于现有注意力模块的优势，我们将我们的方法与SENet [15]、CBAM [58]和自注意力(Self-attention) [41]进行了比较。如表12所示，TDM + IAM大幅优于现有的注意力方法。值得注意的是，我们的方法与现有模块的主要区别在于，我们在注意力模块中基于通道的代表性分数明确衡量并利用通道重要性，而现有方法依赖于不太适合少样本和细粒度场景的可学习参数。因此，这些结果证实了我们的模块在细粒度少样本分类任务中相对于现有注意力模块具有明显优势。

# 8 结论

在本文中，我们首先介绍了专为细粒度少样本图像分类量身定制的通道注意力模块，即任务差异最大化(Task Discrepancy Maximization，TDM)，它包含两个子模块:支持注意力模块(Support Attention Module，SAM)和查询注意力模块(Query Attention Module，QAM)。SAM的核心原则是强调编码类别判别信息的特征图通道，而QAM的一个作用是为查询图像集中与对象相关的通道。这些通道注意力模块能够生成更关注判别细节的任务自适应特征图，以区分细粒度类别。为了进一步提高细粒度和粗粒度少样本分类的表示能力，我们扩展了QAM以提出实例注意力模块(Instance Attention Module，IAM)。具体来说，与针对查询实例的高级特征图工作的QAM不同，IAM在中间层操作，为每个实例(无论是支持图像还是查询图像)突出与对象相关的通道。我们在几个细粒度和粗粒度图像分类基准上对所提出的模块进行了广泛评估，以验证它们相对于先前少样本分类方法在有效性和适用性方面的独特优点。

表12:与现有注意力方法的比较。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 加州理工学院鸟类数据集裁剪版(CUB\_cropped) | | 飞机(Aircraft) | |
| 单样本(1-shot) | 五样本(5-shot) | 单样本(1-shot) | 五样本(5-shot) |
| 原型网络(ProtoNet) [48] | 62.90 | 84.13 | 47.37 | 68.96 |
| + 挤压与激励网络(SENet) [15] | 69.62 | 85.90 | 48.58 | 67.84 |
|  | 69.21 | 85.37 | 48.10 | 70.03 |
|  | 69.23 | 87.49 | 50.07 | 70.41 |
| + 时间动态模块(TDM) + 图像注意力模块(IAM) | 72.96 | 88.02 | 52.88 | 72.81 |

# 参考文献

[1] 阿尔曼·阿夫拉西亚比(Arman Afrasiyabi)、让 - 弗朗索瓦·拉隆德(Jean - François Lalonde)和克里斯蒂安·加涅(Christian Gagné)。少样本图像分类的关联对齐。见《欧洲计算机视觉会议》，第18 - 35页。施普林格出版社，2020年。

[2] 德米特里·巴达诺乌(Dzmitry Bahdanau)、郑焕卓(Kyunghyun Cho)和约书亚·本吉奥(Yoshua Bengio)。通过联合学习对齐和翻译进行神经机器翻译。见《国际学习表征会议》，2015年。

[3] 陈伟宇(Wei - Yu Chen)、刘彦成(Yen - Cheng Liu)、佐尔特·基拉(Zsolt Kira)、王宇强(Yu - Chiang Frank Wang)和黄佳宾(Jia - Bin Huang)。深入研究少样本分类。见《国际学习表征会议》，2019年。

[4] 陈寅波(Yinbo Chen)、刘壮(Zhuang Liu)、徐慧娟(Huijuan Xu)、特雷弗·达雷尔(Trevor Darrell)和王小龙(Xiaolong Wang)。元基线:探索用于少样本学习的简单元学习方法。见《电气与电子工程师协会/计算机视觉基金会国际计算机视觉会议论文集》，第9062 - 9071页，2021年。

[5] 陈正宇(Zhengyu Chen)、葛吉协(Jixie Ge)、詹鹤神(Heshen Zhan)、黄思腾(Siteng Huang)和王东林(Donglin Wang)。少样本学习的帕累托自监督训练。见《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第13663 - 13672页，2021年。

[6] 邓嘉(Jia Deng)、董威(Wei Dong)、理查德·索舍尔(Richard Socher)、李佳(Li - Jia Li)、李开复(Kai Li)和李菲菲(Li Fei - Fei)。ImageNet:一个大规模分层图像数据库。见《2009年电气与电子工程师协会计算机视觉与模式识别会议》，第248 - 255页。电气与电子工程师协会，2009年。

[7] 丁瑶(Yao Ding)、周彦钊(Yanzhao Zhou)、朱毅(Yi Zhu)、叶启祥(Qixiang Ye)和焦建斌(Jianbin Jiao)。用于细粒度图像识别的选择性稀疏采样。见《电气与电子工程师协会/计算机视觉基金会国际计算机视觉会议论文集》，第6599 - 6608页，2019年。

[8] 卡尔·多尔施(Carl Doersch)、安库什·古普塔(Ankush Gupta)和安德鲁·齐斯曼(Andrew Zisserman)。交叉变换器:具有空间感知的少样本迁移。见《神经信息处理系统进展33:2020年神经信息处理系统年度会议，NeurIPS 2020，2020年12月6 - 12日，线上会议》，2020年。

[9] 阿列克谢·多索维茨基(Alexey Dosovitskiy)、卢卡斯·拜尔(Lucas Beyer)、亚历山大·科列斯尼科夫(Alexander Kolesnikov)、德克·魏森伯恩(Dirk Weissenborn)、翟晓华(Xiaohua Zhai)、托马斯·翁特希纳(Thomas Unterthiner)、莫斯塔法·德赫加尼(Mostafa Dehghani)、马蒂亚斯·明德尔(Matthias Minderer)、格奥尔格·海戈尔德(Georg Heigold)、西尔万·热利(Sylvain Gelly)等。一张图像值16×16个词:大规模图像识别的变换器。见《国际学习表征会议》，2021年。

[10] 切尔西·芬恩(Chelsea Finn)、彼得·阿贝贝尔(Pieter Abbeel)和谢尔盖·列维(Sergey Levine)。用于深度网络快速适应的与模型无关的元学习。见《国际机器学习会议》，第1126 - 1135页。模式识别与机器学习会议录，2017年。

[11] 葛伟峰(Weifeng Ge)、林祥儒(Xiangru Lin)和余一舟(Yizhou Yu)。自下而上的用于细粒度图像分类的弱监督互补部件模型。见《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第3034 - 3043页，2019年。

[12] 何恺明(Kaiming He)、张祥雨(Xiangyu Zhang)、任少卿(Shaoqing Ren)和孙剑(Jian Sun)。用于图像识别的深度残差学习。见《电气与电子工程师协会计算机视觉与模式识别会议论文集》，第770 - 778页，2016年。

[13] 洪杰(Jie Hong)、方鹏飞(Pengfei Fang)、李伟豪(Weihao Li)、张彤(Tong Zhang)、克里斯蒂安·西蒙(Christian Simon)、梅赫塔什·哈兰迪(Mehrtash Harandi)和拉尔斯·彼得森(Lars Petersson)。用于少样本学习及其他任务的强化注意力机制。见《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第913 - 923页，2021年。

[14] 侯瑞冰、常洪、马冰鹏、单世广和陈熙霖。用于少样本分类的交叉注意力网络。《神经信息处理系统进展》，第32卷，2019年。

[15] 胡杰、沈立和孙刚。挤压与激励网络。《电气与电子工程师协会计算机视觉与模式识别会议论文集》，第7132 - 7141页，2018年。

[16] 黄泽毅、王浩涵、埃里克·P·邢和黄东。自我挑战提升跨领域泛化能力。《计算机视觉 - ECCV 2020:第16届欧洲会议，英国格拉斯哥，2020年8月23 - 28日，会议录，第二部分16》，第124 - 140页。施普林格出版社，2020年。

[17] 加雷斯·詹姆斯、丹妮拉·维滕、特雷弗·哈斯蒂和罗伯特·蒂贝希拉尼。《统计学习导论》，第112卷。施普林格出版社，2013年。

[18] 姜多贤、权熙承、闵珠洪和赵民秀。用于少样本分类的关系嵌入。《电气与电子工程师协会/计算机视觉基金会国际计算机视觉会议论文集》，第8822 - 8833页，2021年。

[19] 阿迪蒂亚·科斯拉、尼坦南达·贾亚德瓦普拉卡什、姚邦鹏和李菲菲。用于细粒度图像分类的新型数据集:斯坦福犬数据集。《计算机视觉与模式识别会议细粒度视觉分类研讨会论文集》，第2卷。Citeseer，2011年。

[20] 瓦伦丁·赫鲁尔科夫、莱拉·米尔瓦哈博娃、叶夫根尼娅·乌斯蒂诺娃、伊万·奥谢列德茨和维克多·伦皮茨基。双曲图像嵌入。《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第6418 - 6428页，2020年。

[21] 金正民、金泰燮、金成雄和柳昌德。用于少样本学习的边标记图神经网络。《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第11 - 20页，2019年。

[22] 乔纳森·克劳斯、迈克尔·斯塔克、邓嘉和李菲菲。用于细粒度分类的3D对象表示。《电气与电子工程师协会国际计算机视觉研讨会论文集》，第554 - 561页，2013年。

[23] 李权俊、苏布兰苏·马吉、阿维纳什·拉维钱德兰和斯特凡诺·索阿托。基于可微凸优化的元学习。《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第10657 - 10665页，2019年。

[24] 李素恩、文元俊和许在弼。用于细粒度少样本分类的任务差异最大化。《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第5331 - 5340页，2022年。

[25] 李傲雪、黄伟然、兰旭、冯嘉时、李震国和王伟。通过自适应边界损失提升少样本学习。《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第12576 - 12584页，2020年。

[26] 李博浩、杨博宇、刘畅、刘峰、季荣荣和叶启祥。超越最大边界:少样本目标检测的类别边界均衡。《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第7363 - 7372页，2021年。

[27] 李聪聪、杜大伟、张立波、温龙胤、罗铁坚、吴艳军和朱鹏飞。用于无监督领域自适应的空间注意力金字塔网络。《欧洲计算机视觉会议》，第481 - 497页。施普林格出版社，2020年。

[28] 李宏阳、大卫·艾根、塞缪尔·道奇、马修·齐勒和王晓刚。通过类别遍历寻找少样本学习中与任务相关的特征。《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第1 - 10页，2019年。

[29] 李俊杰(Junjie Li)、王子磊(Zilei Wang)和胡小明(Xiaoming Hu)。通过擦除 - 修复学习完整特征以进行少样本分类。收录于《AAAI人工智能会议论文集》，第35卷，第8401 - 8409页，2021年。

[30] 李文斌(Wenbin Li)、王磊(Lei Wang)、徐静琳(Jinglin Xu)、霍静(Jing Huo)、高杨(Yang Gao)和罗捷波(Jiebo Luo)。重新审视基于局部描述符的图像到类别的度量用于少样本学习。收录于《IEEE/CVF计算机视觉与模式识别会议论文集》，第7260 - 7268页，2019年。

[31] 林周翰(Zhouhan Lin)、冯敏伟(Minwei Feng)、西塞罗·诺盖拉·多斯·桑托斯(Cicero Nogueira dos Santos)、于默(Mo Yu)、向兵(Bing Xiang)、周博文(Bowen Zhou)和约书亚·本吉奥(Yoshua Bengio)。一种结构化自注意力句子嵌入方法。收录于《国际学习表征会议》，2017年。

[32] 刘斌(Bin Liu)、曹越(Yue Cao)、林宇通(Yutong Lin)、李琦(Qi Li)、张政(Zheng Zhang)、龙明盛(Mingsheng Long)和胡瀚(Han Hu)。负边界很重要:理解少样本分类中的边界。收录于《欧洲计算机视觉会议》，第438 - 455页。施普林格出版社，2020年。

[33] 刘传斌(Chuanbin Liu)、谢洪涛(Hongtao Xie)、查正军(Zheng - Jun Zha)、马凌峰(Lingfeng Ma)、余凌云(Lingyun Yu)和张永东(Yongdong Zhang)。过滤与蒸馏:增强细粒度视觉分类的区域注意力。收录于《AAAI人工智能会议论文集》，第34卷，第11555 - 11562页，2020年。

[34] 刘泽(Ze Liu)、林宇通(Yutong Lin)、曹越(Yue Cao)、胡瀚(Han Hu)、魏一轩(Yixuan Wei)、张政(Zheng Zhang)、斯蒂芬·林(Stephen Lin)和郭百宁(Baining Guo)。Swin Transformer:使用移位窗口的分层视觉Transformer。收录于《IEEE/CVF国际计算机视觉会议论文集》，第10012 - 10022页，2021年。

[35] S. 马吉(S. Maji)、J. 坎纳拉(J. Kannala)、E. 拉赫图(E. Rahtu)、M. 布拉施科(M. Blaschko)和A. 韦尔达利(A. Vedaldi)。飞机的细粒度视觉分类。技术报告，2013年。

[36] 普尼特·曼格拉(Puneet Mangla)、努普尔·库马里(Nupur Kumari)、阿比舍克·辛哈(Abhishek Sinha)、梅扬克·辛格(Mayank Singh)巴拉吉·克里希纳穆尔蒂(Balaji Krishnamurthy)和维尼特·N·巴拉苏布拉马尼亚姆(Vineeth N Balasubramanian)。绘制正确的流形:用于少样本学习的流形混合。收录于《IEEE/CVF冬季计算机视觉应用会议论文集》，第2218 - 2227页，2020年。

[37] 格雷瓜尔·蒙塔冯(Grégoire Montavon)、热纳维耶芙·奥尔(Geneviève Orr)和克劳斯 - 罗伯特·米勒(Klaus - Robert Müller)。《神经网络:技巧与实践》，第7700卷。施普林格出版社，2012年。

[38] 文俊·穆恩(WonJun Moon)、玄锡成(Hyun Seok Seong)和河在弼(Jae - Pil Heo)。用于视频长尾识别的基于注意力聚合的少数类邻域扩展。预印本arXiv:2211.13471，2022年。

[39] 鲍里斯·奥列什金(Boris Oreshkin)、保罗·罗德里格斯·洛佩斯(Pau Rodríguez López)和亚历山大·拉科斯特(Alexandre Lacoste)。Tadam:用于改进少样本学习的任务相关自适应度量。《神经信息处理系统进展》，第31卷，2018年。

[40] 奥姆卡尔·M·帕克希(Omkar M Parkhi)、安德里亚·韦尔达利(Andrea Vedaldi)、安德鲁·齐斯曼(Andrew Zisserman)和CV·贾瓦哈尔(CV Jawahar)。猫和狗。收录于《2012年IEEE计算机视觉与模式识别会议》，第3498 - 3505页。IEEE，2012年。

[41] 普拉吉特·拉马钱德兰(Prajit Ramachandran)、尼基·帕尔马(Niki Parmar)、阿什什·瓦斯瓦尼(Ashish Vaswani)、欧文·贝洛(Irwan Bello)、安塞尔姆·列夫斯卡亚(Anselm Levskaya)和乔恩·什伦斯(Jon Shlens)。视觉模型中的独立自注意力机制。《神经信息处理系统进展》，第32卷，2019年。

[42] 萨钦·拉维(Sachin Ravi)和雨果·拉罗谢尔(Hugo Larochelle)。将优化作为少样本学习的模型。收录于《国际学习表征会议》，2017年。

[43] 任梦叶(Mengye Ren)、埃莱妮·特里安塔菲卢(Eleni Triantafillou)、萨钦·拉维(Sachin Ravi)、杰克·斯内尔(Jake Snell)、凯文·斯韦尔斯基(Kevin Swersky)、约书亚·B·特南鲍姆(Joshua B. Tenenbaum)、雨果·拉罗谢尔(Hugo Larochelle)和理查德·S·泽梅尔(Richard S. Zemel)。用于半监督少样本分类的元学习。收录于《第六届国际学习表征会议(ICLR)论文集》，2018年。

[44] 赖恩·罗迪(Ryne Roady)、泰勒·L·海斯(Tyler L Hayes)、罗纳德·凯姆克(Ronald Kemker)、阿耶莎·冈萨雷斯(Ayesha Gonzales)和克里斯托弗·卡南(Christopher Kanan)。开放集分类方法在大规模数据集上是否有效？《公共科学图书馆·综合》(Plos one)，15(9):e0238302，2020年。

[45] 成贤锡(Hyun Seok Seong)、文元俊(WonJun Moon)、李素彬(SuBeen Lee)和许在弼(Jae - Pil Heo)。利用隐藏正样本进行无监督语义分割。收录于《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》(Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition)，第19540 - 19549页，2023年。

[46] 沈相宪(Sang - Heon Shim)、玄相吉(Sangeek Hyun)、裴大贤(DaeHyun Bae)和许在弼(Jae - Pil Heo)。用于场景图像生成的局部注意力金字塔。收录于《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》(Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition)，第7774 - 7782页，2022年。

[47] 克里斯蒂安·西蒙(Christian Simon)、彼得·科尼乌什(Piotr Koniusz)、理查德·诺克(Richard Nock)和梅赫塔什·哈兰迪(Mehrtash Ha - rand)。用于少样本学习的自适应子空间。收录于《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》(Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition)，第4136 - 4145页，2020年。

[48] 杰克·斯内尔(Jake Snell)、凯文·斯沃尔斯基(Kevin Swersky)和理查德·泽梅尔(Richard Zemel)。用于少样本学习的原型网络。《神经信息处理系统进展》(Advances in neural information processing systems)，30，2017年。

[49] 宋富德(Flood Sung)、杨永新(Yongxin Yang)、张立(Li Zhang)、向涛(Tao Xiang)、菲利普·H·S·托尔(Philip HS Torr)和蒂莫西·M·霍斯佩代尔斯(Timothy M Hospedales)。学会比较:用于少样本学习的关系网络。收录于《电气与电子工程师协会计算机视觉与模式识别会议论文集》(Proceedings of the IEEE conference on computer vision and pattern recognition)，第1199 - 1208页，2018年。

[50] 格兰特·范·霍恩(Grant Van Horn)、奥辛·麦·奥达(Oisin Mac Aodha)、宋洋(Yang Song)、崔音(Yin Cui)、孙晨(Chen Sun)、亚历克斯·谢泼德(Alex Shepard)、哈特维希·亚当(Hartwig Adam)、彼得罗·佩罗纳(Pietro Perona)和塞尔日·贝洛尼(Serge Belongie)。iNaturalist物种分类与检测数据集。收录于《电气与电子工程师协会计算机视觉与模式识别会议论文集》(Proceedings of the IEEE conference on computer vision and pattern recognition)，第8769 - 8778页，2018年。

[51] 阿什ish·瓦斯瓦尼(Ashish Vaswani)、诺姆·沙泽尔(Noam Shazeer)、尼基·帕尔马尔(Niki Parmar)、雅各布·乌斯库雷特(Jakob Uszkoreit)、利昂·琼斯(Llion Jones)、艾丹·N·戈麦斯(Aidan N Gomez)、卢卡斯·凯泽(Łukasz Kaiser)和伊利亚·波洛苏金(Illia Polosukhin)。注意力就是你所需要的一切。收录于《神经信息处理系统进展》(Advances in neural information processing systems)，第5998 - 6008页，2017年。

[52] 奥里奥尔·维尼亚尔斯(Oriol Vinyals)、查尔斯·布伦德尔(Charles Blundell)、蒂莫西·利利克拉普(Timothy Lillicrap)、达恩·维尔斯特拉(Daan Wierstra)等。用于单样本学习的匹配网络。《神经信息处理系统进展》(Advances in neural information processing systems)，29:3630 - 3638，2016年。

[53] C. 瓦(C. Wah)、S. 布兰森(S. Branson)、P. 韦林德(P. Welinder)、P. 佩罗纳(P. Perona)和S. 贝洛尼(S. Belongie)。加州理工学院 - 加州大学圣地亚哥分校鸟类200 - 2011数据集。技术报告CNS - TR - 2011 - 001，加州理工学院，2011年。

[54] 王若彤、沈艳清、左伟亮、周三明和郑南宁。Transvpr:基于Transformer的多级注意力聚合场所识别方法。收录于《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》(Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition)，第13648 - 13657页，2022年。

[55] 王岩、赵伟伦(Wei - Lun Chao)、基利安·Q·温伯格(Kilian Q Weinberger)和劳伦斯·范德马滕(Laurens van der Maaten)。简单射击:重新审视少样本学习中的最近邻分类。预印本arXiv:1911.04623，2019年。

[56] 戴维斯·韦特海默(Davis Wertheimer)和巴拉特·哈里哈兰(Bharath Hariharan)。现实场景下带定位的少样本学习。收录于《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》(Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition)，第6558 - 6567页，2019年。

[57] 戴维斯·韦特海默(Davis Wertheimer)、唐鲁明(Luming Tang)和巴拉特·哈里哈兰(Bharath Hariharan)。基于特征图重建网络的少样本分类。收录于《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》(Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern识别》(Recognition)，第8012 - 8021页，2021年。

[58] 吴相勋(Sanghyun Woo)、朴钟灿(Jongchan Park)、李准英(Joon - Young Lee)和权仁秀(In So Kweon)。CBAM:卷积块注意力模块。收录于《欧洲计算机视觉会议论文集》(Proceedings of the European conference on computer vision (ECCV))，第3 - 19页，2018年。

[59] 谢江涛、龙飞、吕佳明、王启龙和李培华。联合分布很重要:用于小样本分类的深度布朗距离协方差。见《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第7972 - 7981页，2022年。

[60] 徐成明、傅彦伟、刘晨、王承杰、李吉林、黄飞跃、张莉和薛向阳。通过元滤波器学习动态对齐用于小样本学习。见《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第5182 - 5191页，2021年。

[61] 徐芸和罗伊斯顿·古德阿克。关于划分训练集和验证集:用于估计监督学习泛化性能的交叉验证、自助法和系统抽样的比较研究。《分析与测试杂志》，2(3):249 - 262，2018年。

[62] 叶寒佳、胡鹤翔、詹德传和沙飞。通过使用集到集函数的嵌入自适应进行小样本学习。见《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第8808 - 8817页，2020年。

[63] 张宝全、李旭涛、叶云明、黄志超和张丽莎。利用原始知识进行原型补全用于小样本学习。见《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第3754 - 3762页，2021年。

[64] 张弛、蔡宇军、林国升和沈春华。深度EMD:使用可微的地球移动距离和结构化分类器进行小样本图像分类。见《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第12203 - 12213页，2020年。

[65] 张红光、彼得·科尼乌斯、简松磊、李洪东和菲利普·H·S·托尔。重新思考类别关系:绝对 - 相对监督和无监督小样本学习。见《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第9432 - 9441页，2021年。

[66] 赵佳宝、杨一帆、林鑫、杨静和何亮。在小样本学习中更广泛地寻找以获得更好的自适应表示。见《美国人工智能协会会议论文集》，第35卷，第10981 - 10989页，2021年。

[67] 郑鹤亮、付建龙、查正军和罗洁波。在细节中寻找关键:学习三线性注意力采样网络用于细粒度图像识别。见《电气与电子工程师协会/计算机视觉基金会计算机视觉与模式识别会议论文集》，第5012 - 5021页，2019年。



李承彬(SuBeen Lee)于2020年在韩国全北国立大学(Jeonbuk National University，JBNU)获得计算机科学学士学位，于2022年在韩国成均馆大学(Sungkyunkwan University，SKKU)获得人工智能硕士学位，目前正在该校攻读人工智能博士学位。他的研究兴趣包括小样本分类、注意力机制和深度学习。



文元俊(WonJun Moon)是韩国成均馆大学(Sungkyunkwan University，SKKU)人工智能系的博士生。他分别于2021年和2022年在成均馆大学(SKKU)获得计算机科学学士和硕士学位。目前，他的研究领域包括计算机视觉和深度学习。



成贤锡(Hyun Seok Seong)于2019年在韩国成均馆大学(Sungkyunkwan University，SKKU)获得电子与电气工程学士学位，目前正在该校攻读人工智能硕博连读学位。他的研究兴趣包括用于图像分类的度量学习、机器学习和深度学习。



许在弼(Jae - Pil Heo，电气与电子工程师协会会员)分别于2008年、2010年和2015年在韩国科学技术院(Korea Advanced Institute of Science and Technology，KAIST)获得计算机科学学士、硕士和博士学位。他目前是韩国成均馆大学(Sungkyunkwan University，SKKU)的副教授。在加入成均馆大学之前，他是电子通信研究院(Electronics and Telecommunications Research Institute，ETRI)的研究员。他的研究兴趣包括计算机视觉和机器学习。