

基于空间注意力和类协方差度量的小样本学习^①

李国强^② 王天雷 龚 宁 王俊妍

(燕山大学电气工程学院 秦皇岛 071000)

摘 要 近年来,小样本学习逐渐被学术界广泛研究,其旨在使模型在给定样本较少的情
况下完成一系列任务。目前基于度量学习的元学习算法被广泛应用于小样本学习中,本
文利用度量学习的思想,对基准元度量学习算法原型网络进行改进,提出了注意力类协方
差原型网络。首先,为增加模型的泛化能力,提出 IBN-Resnet12 作为特征提取器;接着,
在特征提取模块后加入了空间注意力模块,有效地增强了局部特征;最后提出类协方差度
量作为最终的度量分类器,完成了对图像特征维度间相关性的建模。本文在小样本学习
经典数据集上进行实验,证明了模型的有效性;同时还进行了大量消融实验,证明了模型
改进中各个部分的有效性。

关键词 小样本学习;元度量学习;泛化能力;空间注意力;类协方差度量

0 引 言

随着深度学习的发展,图像识别技术得到了飞速的进步,在大数据的驱动下,图像分类、目标检测、语义分割等领域都取得了显著的成果。当前的深度学习算法大多数基于卷积神经网络(convolutional neural networks, CNN),随着硬件设备的迭代更新,在训练量足够的前提下,CNN 具有提取特征速度快、提取特征准确等一系列优点。依靠这些优点,CNN 逐渐取代了传统图像识别中利用人工标注获取特征的方法,广泛应用于图像识别领域^[1]。然而当数据量不足时,现有的基于 CNN 的深度学习算法逐渐暴露出泛化能力差、过拟合严重等缺点。小样本学习(few-shot learning)旨在当训练数据量较少的情况下,利用相关算法完成计算机视觉领域的各个任务,近些年来逐渐成为计算机视觉领域的研究重点之一^[2-3]。

目前处理小样本学习问题的方法大致可分为以下几种:基于元学习的小样本学习策略、基于迁移学

习的小样本学习策略和基于图神经网络的小样本学习策略等^[4]。其中基于元学习的小样本学习应用最为广泛,其目的是训练神经网络学会利用以往的经验,使模型学会自己学习,其显著特点为测试集(test set)的类别为训练集(train set)从未见过的新类。而基于元学习的小样本学习算法又可以分为基于优化的元学习算法、基于模型的元学习算法和基于度量学习的元学习算法。

基于度量的元学习算法思想简单易懂,同时在多个数据集上都取得了较好的效果,进而成为处理小样本学习的最主要算法之一。根据度量方式不同,基于度量的元学习分为固定距离度量以及非固定距离度量^[5]。本文借鉴了基于度量的元学习算法思想,在经典的小样本学习模型的基础上,提出了基于注意力类协方差的原型网络(attention-based class covariance prototypical network, ACCPN)。首先,为增加模型的泛化能力,特征嵌入模块在 IBN-Resnet^[6]的基础上,提出了 IBN-Resnet12;同时为加强局部特征,在特征提取器后引入了空间注意力模

^① 河北省自然科学基金(F2020203003)资助项目。

^② 男,1984年生,博士,副教授;研究方向:神经网络;联系人,E-mail:zhihuiyuang@163.com。
(收稿日期:2021-06-29)

块(spatial attention model, SAM)。对于分类器部分, 本文将类协方差距离作为小样本学习度量分类器中的距离度量方式。实验证明, 该方法能够在不增加额外参数的前提下, 提升小样本图像分类的准确率。

1 相关工作

1.1 问题描述

小样本学习是指在给定训练样本较少的前提下, 完成相关的包括图像分类、语义分割、目标检测等任务, 其目标是希望模型经过训练后, 仅通过少量的样本就能快速学习新类。其实现过程可分为元训练阶段 D_{base} 和元测试阶段 D_{novel} 。区别于普通的图像识别, 小样本学习以“任务”为单位, 每一“任务”又包含支持集 D_s 和查询集 D_q 。其评价指标通常为 C -way N -shot, 即在经过元训练阶段 D_{base} 多个“任务”训练过后, 在元测试阶段 D_{novel} 的支持集 D_s 中每一个“任务”有 C 个类别, 每个类别有 N 个样本^[7]。小样本学习的最终目的就是在 D_{novel} 中只通过 D_s 的少量样本就能识别出 D_q 中的样本。

1.2 小样本学习方法介绍

当前处理小样本学习多利用元学习的思想, 而基于元学习的小样本学习算法又可以分为基于优化的元学习算法、基于模型的元学习算法和基于度量学习的元学习算法等。

基于优化的元学习算法的目标为寻找一个好的初始化参数, 在测试阶段只需微调就能使网络在面对从未见多的新样本时能够很快寻找到最优初始值。文献[8]在2017年提出了一种与模型无关的元学习算法(model-agnostic meta-learning, MAML)。该算法提出的模型无关性元学习算法, 使用少量的梯度迭代步骤就可以学习到适用于新任务的参数, 能够匹配任何使用梯度下降法训练的模型。

基于模型的元学习算法旨在学习不同模型中的通用知识, 并用通用知识预测基准模型参数, 并能加快训练速度。文献[9]在2017年使用卷积神经网络构造了一种元网络(meta networks, Meta-Nets), 其思想是学习一种跨任务的元级学习, 利用外部记忆模块存储每个类别对应的权重信息和表征信息,

利用该信息进行预测, 结果表明该方法能在保证准确率的同时对新任务实现快速参数化。

基于度量的元学习算法借鉴了度量学习的思想, 简单易懂, 且拥有较高的准确率。2015年Koch等人^[10]针对小样本学习提出了深度卷积孪生网络, 该算法训练一个孪生网络对样本进行相似性度量, 即让样本对通过完全相同的网络结构, 利用欧式距离对从样本中学习到的特征进行相似性度量, 根据学习到的特征映射对验证集进行分类。而后Snell等人^[11]提出原型网络(prototypical networks, PN), 对于给定数据集, 计算每一类的类平均值作为类原型, 测试时, 利用测试集提取的特征与各类的类原型之间的距离进行分类。Sung等人^[12]在2018年提出了端到端的相关网络(relation network, RN), 相关网络由嵌入模块和关系模块组成, 嵌入模块用于提取图像特征, 关系模块用于计算相关性得分, 利用相关性得分进行分类。王年等人^[13]在关系网络的基础上, 融合inception块和感受野块, 提出了一种基于改进关系网络的小样本学习方法。文献[14]在欧氏距离的基础上借鉴了线性规划中的运输问题, 提出了一种新的用于小样本学习度量分类器中的距离度量方式, 计算查询集和支持集图像的各个图块之间的最佳匹配代价来表示二者之间的相似程度。

此外, 近些年还出现了区别于元学习的小样本学习的方法, 文献[15]提出利用自监督的方法增强小样本数据, 文献[16]引用外部扩展集辅助训练。

2 模型介绍

当前基于元度量学习的小样本学习算法摒弃了传统深度学习的全连接层作为最终的分类器, 而选择参数量较小且不易过拟合的距离度量方法作为小样本学习分类器。距离度量方式大致可分为基于固定距离的度量和基于非固定距离的度量。其中基于固定距离的度量多借鉴原型网络(PN)的思想, 首先利用特征提取模块提取支持集 D_s 图像特征; 而后计算每一类的类原型作为该类的类代表, 具体地, 对于第 n 类的类原型 C_n 可以表示为

$$C_n = \frac{1}{K} \sum_{i=1}^k f_{\theta}(x_{ni}) \quad (1)$$

式中 K 表示第 n 类有 K 个样本, f_θ 表示特征提取器, x_{ni} 表示第 n 个类别的第 i 个样本。而后对于查询集 D_q 的样本 x_i , 提取其特征 $f_\theta(x_i)$, 再与每一类原型计算距离, 再根据距离利用 softmax 函数完成最终的分类:

$$P_{ni} = \frac{\exp(-d(f_\theta(x_i), C_n))}{\sum_{j=1}^C \exp(-d(f_\theta(x_i), C_j))} \quad (2)$$

式中 $d(m, n)$ 表示 m 和 n 之间的距离, P_{ni} 表示 x_i 属于类别 n 的概率, C 表示每个任务有 C 类。目前大多小样本学习算法的特征提取模块多为 4 层卷积神经网络或者 Resnet 的变体 Resnet-12, 这种简单的网络结构虽然能在少量样本的情况下较好地提取特征, 但并未解决小样本学习中常见的泛化能力差等问题, 且由于每一类的样本数量过少, 在计算各类的原型时得到的并不是该类真正的类原型。同时, 基于度量的小样本学习大多采用简单的欧式平方距离以及余弦距离作为最终的距离度量分类器, 而这两种距离度量方式忽略了样本特征维度间的相关性, 在计算类间距离时默认了不同类别的协方差具有一致性。而对于小样本学习任务, 同一类特征维度具有高度相关性, 对特征维度相关性的建模能很

大程度上提升小样本学习的准确率。

为解决上述问题, 本文从特征提取器和分类器两方面对原型网络(PN)进行了改进, 提出了注意力类协方差原型网络(ACCPN)。

对于特征提取器, 本文提出了 IBN-Resnet12 作为特征提取网络, 并在 IBN-Resnet12 后添加了空间注意力模块(SAM)^[17], 更高效提取特征的同时增加了模型的泛化能力, 在样本较少的情况下, 更加突出重要的特征; 对于分类器, 本文提出将类协方差距离作为度量分类器的距离度量方式, 有效解决了 PN 的度量分类器中欧式距离存在的问题, 其总结构图如图 1 所示。对于一个小样本学习任务 Γ , 首先将训练集分为支持集(support set)和查询集(query set), 将 support set 和 query set 图像输入到特征提取器中得到特征 U_s 以及 U_q , 利用式(1)计算 support set 中每一类的类原型 C_k ; 接着将 U_q 与 C_k 进行类协方差距离度量, 并利用 softmax 函数得到最终的分类结果。

本文在小样本数据集 Miniimagenet^[18] 和 Tieredimagenet^[19] 上进行了实验, 证明了模型整体的有效性; 同时本文还做了大量消融实验, 证明了各个部分的有效性。

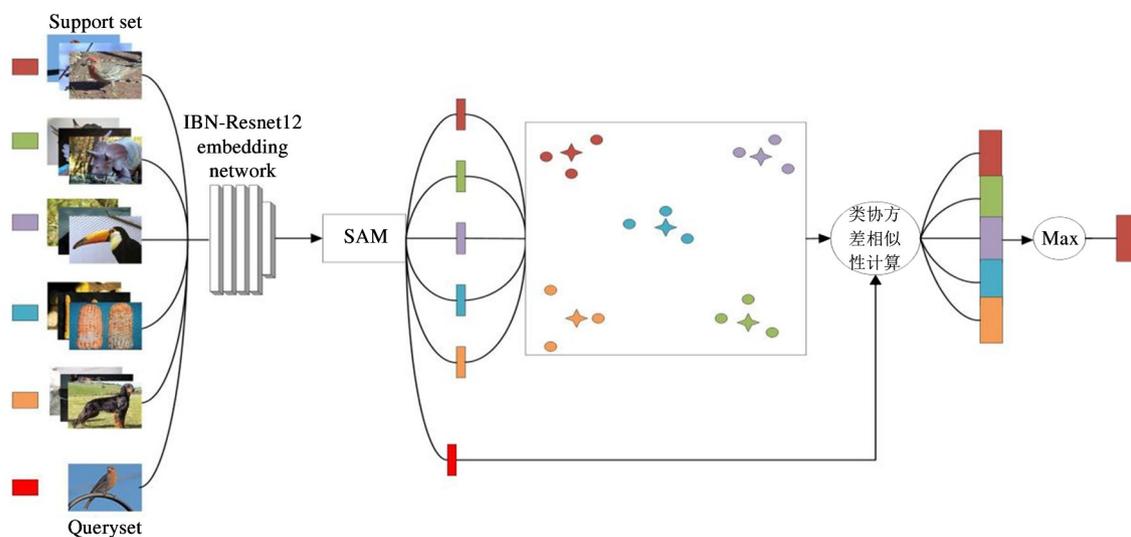


图 1 ACCPN 总体结构图

2.1 特征提取模块

在小样本学习中, 由于 support set 数量极少, 训练得到的模型很容易出现过拟合严重、泛化能力差、

局部特征不明显等问题。针对该问题, 本文改进了传统小样本学习的特征提取模块, 提出了 IBN-Resnet12 作为特征提取网络, 增强了模型的泛化能力,

同时在特征提取网络后加入了空间注意力模块 (SAM),有效增强了局部特征。IBN-Resnet12 以及 SAM 共同构成了本文的特征提取模块。

IBN-Resnet12 结构图如图 2 所示,其设计原理是将批归一化(batch normalization, BN)^[20]和实例归一化(instance normalization, IN)^[21]结合,IN 和 BN 在本质上都是归一化的方法,IN 针对单一的图像,而 BN 针对一个批次的图像。Pan 等人^[6]提出 IN 对图像外观变化具有不变性,BN 可以保存内容相关信息,故 IN 常被用于图像风格迁移等底层视觉任务中,而 BN 则常被用于目标识别、分类等高层任务中。如果将 IN 和 BN 结合起来,则同时提高了模型的泛化能力和学习能力。本文在浅层同时使用 IN 和 BN,而在网络深层只使用 BN,提出了 IBN-Resnet12。实验证明,该网络对小样本图像分类效果有一定的提升。

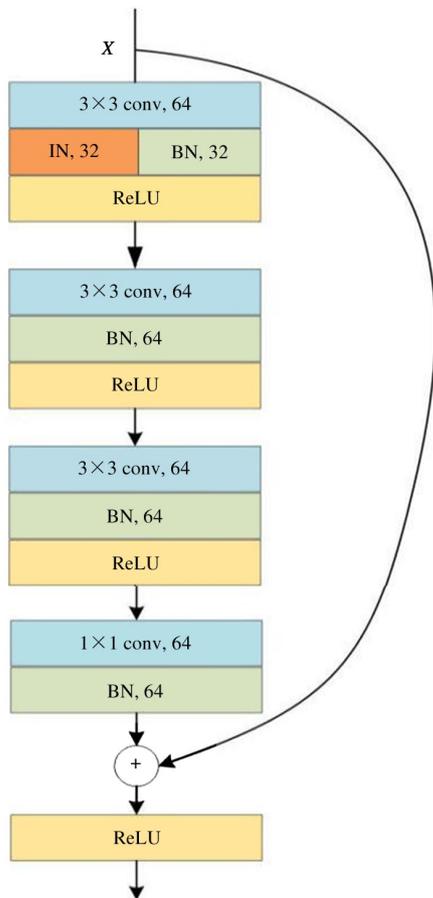


图 2 IBN-Resnet12

当前大多数基于度量的小样本学习在经过特征提取网络后直接将特征输入到相似性度量模块。而

由于各类的样本较少,利用该方法计算得到的类原型并非该类真正的类原型。本文利用空间注意力模块(SAM)增强局部特征,对更重要的局部特征赋予更大的权重,使经过特征提取模块并计算得到的类原型更加接近真实的类原型。空间注意力模块如图 3 所示。

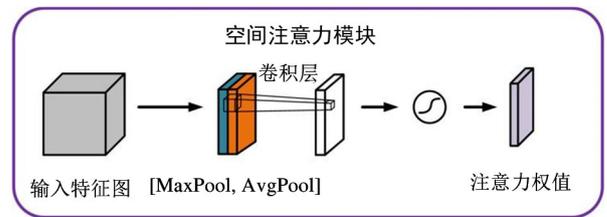


图 3 空间注意力模块

空间注意力是对通道进行压缩,对于输入到 SAM 的特征,首先做一个基于通道的 Max pooling 和 Average pooling,然后在通道维度上进行合并,再经过一个卷积降维为一个通道,最后利用 Sigmoid 函数生成权重 $\sigma_{i,j}(i \in (1,h), j \in (1,w))$ 。其过程可由式(3)表示。

$$M_s(F) = \sigma(\text{conv}^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) = \sigma(\text{conv}^{7 \times 7}([F_{\text{avg}}^s; F_{\text{max}}^s])) \quad (3)$$

其中 conv 表示一层卷积神经网络,7 × 7 表示表示卷积核大小。

给定一张尺寸为 $H \times W$ 的图像 A,经过 2.1 小节中的 IBN-Resnet12 特征提取模块,得到了大小为 $h \times w \times d$ 的特征向量矩阵 U ,其中 h, w 表示长、宽, d 表示通道数,同时还可以将 U 看作 $h \times w$ 个局部特征的和,每个局部特征包含 d 个单位,即每一张图像经过特征提取拥有 $h \times w$ 个 d 维的局部特征^[22]。本文在 IBN-Resnet12 特征提取后利用 SAM 强化局部特征,经过 IBN-Resnet12 模块得到的特征为 $U = [u^{1,1}, u^{1,2}, \dots, u^{i,j}, \dots, u^{h,w}]$,其中 $u^{i,j} \in R^{1 \times 1 \times c}$ 代表空间位置 $(i, j), i \in \{1, 2, \dots, h\}, j \in \{1, 2, \dots, w\}$,而后经过由式(4)表示的 SAM 过程,得到图像特征中各个局部特征的权值,最后将原特征与权值相乘,得到由 SAM 强化后的图像特征 U_{SAM} 。

$$U_{\text{SAM}} = [\sigma_{1,1} u^{1,1}, \dots, \sigma_{i,j} u^{i,j}, \dots, \sigma_{h,w} u^{h,w}] \quad (4)$$

相比于只利用 IBN-Resnet12 提取的特征 U , U_{SAM} 能有效地突出更重要的局部特征,而在计算类

原型时,由于每一类的样本较少,利用 U_{SAM} 计算得到的类原型更接近该类的真实类原型。经实验证明,SAM 模块的添加,有效地提升了小样本图像分类的准确率。

2.2 类协方差距离度量

原型网络(PN)采用固定距离度量方式,其侧重点在于特征提取部分。在度量分类器的选择上,通过比较欧式距离和余弦距离,最终选择了欧氏距离作为度量分类器的距离度量方式。然而欧氏距离在度量特征之间的距离时,有以下缺点:(1) 欧氏距离并未考虑特征维度之间的关联性;(2) 欧氏距离默认了所有特征服从同一分布,而在小样本学习中,对分布进行建模能有效地提升模型性能。

本文提出类协方差距离度量,定义类协方差距离为基于类别的协方差距离的加权和。协方差距离是有效计算两个样本相似度的方法;对于一个均值为 $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_p)^T$, 协方差矩阵为 \boldsymbol{Q} 的多变量 $\boldsymbol{x} = (x_1, x_2, x_3, \dots, x_p)^T$, 协方差距离如式(5)所示。

$$D_K(\boldsymbol{x}) = \sqrt{(\boldsymbol{x} - \boldsymbol{\mu})^T (\boldsymbol{Q})^{-1} (\boldsymbol{x} - \boldsymbol{\mu})} \quad (5)$$

协方差距离解决了上述欧氏距离在计算小样本学习中支持集类原型和查询集特征相似度时的缺点。具体的,协方差距离在度量时通过协方差矩阵考虑到了特征维度间的关联性,且成功解决了欧氏距离对于类内样本相对于原型的分布不敏感这一缺点^[23]。协方差距离和欧氏距离的比较如图4所示。

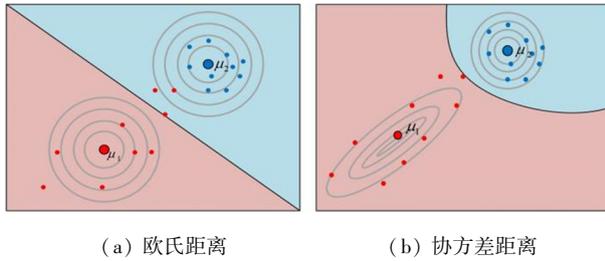


图4 两种距离比较

对于一个小样本学习任务 Γ , 输入图像经过特征提取模块,得到支持集 D_s 的特征 $f_\theta(x_s)$ 和查询集 D_q 的特征 $f_\theta(x_q)$, 然后利用式(1)计算支持集各类的类原型 C_K , K 表示第 K 类,则分类概率计算公式为

$$p(y_i^* | f_\theta^f(x_i^*); S^\Gamma) = \text{soft max}(-d_k(f_\theta^f(x_i^*); C_k)) \quad (6)$$

式中, d_k 表示协方差距离,其协方差矩阵为 \boldsymbol{Q}_k^f , 表示特定于任务 Γ 和类别 K 的协方差矩阵。故如何计算该协方差矩阵为建模的关键,由于小样本学习中支持集的样本数量远小于特征空间维度,因此为保证可逆,使用一种正则化方法确定 \boldsymbol{Q}_k^f 。

$$\boldsymbol{Q}_k^f = \lambda_k^f \boldsymbol{M}_k^f + (1 - \lambda_k^f) \boldsymbol{M}^f + \beta \boldsymbol{I} \quad (7)$$

式中, \boldsymbol{I} 为单位矩阵; β 为矩阵缩放系数,根据特征的大小决定,本实验取 $\beta = 0.5$; \boldsymbol{Q}_k^f 由两个协方差矩阵加权组合而来,其中 \boldsymbol{M}_k^f 是类内协方差矩阵, \boldsymbol{M}^f 是类间协方差矩阵;通过对所有 $x_i \in S_k^f$ 的特征值 $f_\theta^f(x_i)$ 进行估计得到 \boldsymbol{M}_k^f 。

$$\boldsymbol{M}_k^f = \frac{1}{|S_k^f| - 1} \sum_{(x_i, y_i) \in S_k^f} (f_\theta^f(x_i) - c_k)(f_\theta^f(x_i) - c_k)^T \quad (8)$$

其中, S_k^f 表示支持集中属于类别 K 的样本,在单样本时,将 \boldsymbol{M}_k^f 定义为零矩阵。对于 \boldsymbol{M}^f , 其计算方式与 \boldsymbol{M}_k^f 相同,不同的是 \boldsymbol{M}^f 针对于 S^f 中所有的样本,定义为任务 Γ 所有类的协方差矩阵。比例系数 λ_k^f 由式(9)计算。

$$\lambda_k^f = |S_k^f| / (|S_k^f| + 1) \quad (9)$$

对于 C -way N -shot 的小样本学习任务 Γ , 当 $N = 1$ 时, $\boldsymbol{Q}_k^f = 0.5 \boldsymbol{M}^f + \beta \boldsymbol{I}$, 此时 \boldsymbol{Q}_k^f 只依赖于 \boldsymbol{M}^f 和 β ; 当 N 大于 1 时, \boldsymbol{Q}_k^f 逐渐由 \boldsymbol{M}_k^f 、 \boldsymbol{M}^f 、 β 三者共同决定,且 N 越大,类内协方差矩阵 \boldsymbol{M}_k^f 对 \boldsymbol{Q}_k^f 的影响越大。

支持集图像 D_s 经过 IBN-Resnet12 的特征提取以及 SAM 增强局部特征后,得到经过特征增强的图像特征 U_{SSAM} ; 对于第 i 类,利用式(1)计算该类的类原型 C_i 。查询集图像 D_q 同样经过特征提取模块以及 SAM 模块得到该图像的特征 U_{qSAM} ; 随后将支持集得到的各类的类原型 C_i 与查询集得到的特征 U_{qSAM} 送入到相似性度量模块,利用相似性得分,将查询集图像 D_q 进行分类。其流程如算法 1 所示。

由算法 1 可以得到 ACCPN 训练的全部流程;首先划分数据。将数据集分为训练集、验证集和测试集,取训练集 D_{base} 并将其分为 T 个任务,每个任务由支持集 D_s 和查询集 D_q 组成。接着将划分好的任务输入到特征提取模块中,计算支持集的类原型 C 。而后计算查询集 D_q 中各个样本与不同类别 C 的协

方差距离 d 。最后利用 d 确定该样本的类别。利用交叉熵损失函数计算损失,经过 R 次迭代,得到训练好的模型 F^R 。ACCPN 是基于 PN 网络的改进,由于其加入了注意力机制,故在反向传播时需要迭代的参数除了 CNN 还有 SAM,使得 ACCPN 在时间复杂度上略高于 PN,但 SAM 只包含少量卷积层,故该模型以牺牲时间复杂度来换取准确率策略具有可行性。

算法 1

输入:

- 训练集 D_{base}
- 任务 T 的数量 T
- 每一任务支持集类别数量 C
- 每一类别样本数量 N
- 查询集数量 S
- 迭代轮数 R

输出:

第 R 轮的模型 F^R
 $F = \text{CNN} + \text{SAM}$
 初始化模型 F^1 参数 ϕ, θ 以及学习率 α
 $r = 1$
 Repeat
 For i, batch in 训练集 D_{base} :
 随机选择 $C \times N$ 个样本构成支持集 D_s
 随机选择 $C \times S$ 个样本构成查询集 D_q
 根据式(1)计算 D_s 类原型 C
 根据式(8)计算 M_k^r, M^r
 根据式(9)计算比例系数 λ_k^r
 根据式(7)计算协方差矩阵 Q_k^r
 根据式(5)计算 D_q 与 C 的类协方差距离 D_k
 根据式(6)计算分类概率 P
 End for
 利用梯度下降,更新模型参数 ϕ, θ, α
 $r + 1$
 Until r 等于 R
 Return CNN + SAM 模型 F^R

3 实验

为验证上述算法的有效性,本文在小样本学习经典数据集 Miniimagenet 和 Tieredimagenet 上进行大量实验。

3.1 数据集介绍

Miniimagenet 是从大型计算机视觉数据集 Imagenet^[24]上节选的部分数据。2016年 DeepMind 团队将 Miniimagenet 数据集用于小样本图像分类,自此该数据集成为了小样本学习研究的基准数据集。Miniimagenet 包含 100 个类,每一类包含 600 个样本,每张图像大小为 84×84 像素。本文将该数据集按 6:2:2 的比例分为训练集、验证集和测试集。

Tieredimagenet 也是从 Imagenet 数据集中选取的部分数据,由文献[17]于 2018 年提出,并应用于小样本学习领域。其一共包含 34 个大类,每个大类又包含 10~30 个小类,共计 608 个类别,779 165 张图像;将 608 个类别划分为 351 个训练集,97 个验证集,160 个测试集,具体划分情况如表 1 所示。

表 1 Tieredimagenet 数据集划分情况

	训练集	验证集	测试集	全部数据
大类/个	20	6	8	34
小类/个	351	97	160	608
图像数量/张	448 695	124 261	206 209	779 165

3.2 实验设置

本实验的测试平台是在 Nvidia TITAN Xp (GPU 12 GB),编程环境为 Python 3.6,使用的深度学习框架为 Pytorch;优化器选择 Adam 优化器,采用固定学习率为 0.001。对于小样本学习任务,本文选择 5-way 5-shot 和 5-way 1-shot 2 个评价指标,即测试阶段的支持集随机选取 5 个类别,每种类别分别选取 5 个以及 1 个样本;最终准确率由测试阶段所有任务的平均准确率计算得到。在训练阶段,支持集的类别数 (support-way) 和样本数 (support-shot) 会直接影响模型训练的好坏和最终的准确率;对于 Miniimagenet 和 Tieredimagenet 2 个数据集,本文设置的参数基本一致。5-way 5-shot 具体实验设置如表 2 所示,5-way 1-shot 实验设置如表 3 所示。

表中 query-shot 为查询集每一类样本数量, max-epoch 表示迭代的总轮数, save-epoch 表示保存模型的轮数, task-num 表示任务的数量。

表2 5-way 5-shot 实验设置

训练阶段		测试阶段	
support-way	15	support-way	5
support-shot	5	support-shot	5
query-shot	20	query-shot	20
max-epoch	200	task-num	1000
save-epoch	20		
task-num	100		

表3 5-way 1-shot 实验设置

训练阶段		测试阶段	
support-way	20	support-way	5
support-shot	1	support-shot	1
query-shot	20	query-shot	20
max-epoch	200	task-num	1000
save-epoch	20		
task-num	100		

3.3 实验结果分析

本文在原型网络基础上提出了基于注意力和类协方差的原型网络 (ACCPN), 并在小样本图像分类经典数据集 Miniimagenet 和 Tieredimagenet 上进行大量实验来证明整体算法的有效性, 同时进行相关消融实验以证明各个部分的有效性。

3.3.1 Miniimagenet 实验对比

为验证整体模型的有效性, 本文将 ACCPN 与经典小样本学习算法进行对比, 其实验结果如表 4 所示。由表可知, 本文对比了 Matching Networks^[25]、PN(4-conv)^[11]、PN(Resnet12)^[11]、MAML^[8]、Relation Networks^[12]、AdaResnet^[26] 和 SNAIL^[27] 等小样本学习算法, 本文提出的方法 ACCPN 在 5-way 5-shot 和 5-way 1-shot 任务上都取得了显著的效果, 其中 ACCPN 在 5-way 5-shot 任务上取得了所有对比算法中最好的效果。对于 5-way 5-shot 任务, ACCPN 相比于基准算法 PN(4-Conv) 以及 PN(Resnet12) 的性能提升分别为 6.73% 和 3.06%; 对于 5-way 1-shot 任务, ACCPN 相比于基准算法 PN(4-Conv) 以及 PN(Resnet12) 的性能提升为 6.64% 和 3.09%。为证明各个改进部分的有效性, 本文进行了大量的消融实验, 其结果如表 5 所示。基准模型为 PN(Resnet12)。

由表 4 可知, 相比于其他的小样本学习算法, 本文提出的方法 ACCPN 在小样本图像分类中取得了显著的效果。而由表 5 可知, 本文提出的 IBN-Resnet12, SAM 以及类协方差度量 (class covariance, CC) 对于小样本图像分类都具有一定的效果。对于 5-way 5-shot 任务, IBN-Resnet12 结构相比于 Resnet12 结构的精度提升为 0.70% 左右; 加入了 SAM 的 PN 相比于基准 PN 的精度提升约为 0.72%; 本文提出的类协方差度量 (CC) 对于最终精度的提升效果最明显, 相对于基准 PN, 该方法精度提升为 2.6% 左右。对于 5-way 1-shot 任务, IBN-Resnet12 结构相比于 Resnet12 结构的精度提升为 0.71%; 加入了 SAM 的 PN 相比于基准 PN 的精度提升约为 0.50%; 类协方差度量 (CC) 相对于基准 PN 提升了 1.58% 的精确度。

表4 Miniimagenet 数据性能比较/%

方法	5-way 5-shot	5-way 1-shot
Matching Networks ^[25]	55.31	43.56
PN(4-conv) ^[11]	66.40	48.42
PN(Resnet12) ^[11]	70.07	52.48
MAML ^[8]	63.11	47.70
AdaResnet ^[26]	71.94	56.88
SNAIL ^[27]	68.88	55.17
Relation Networks ^[12]	70.20	52.19
ACCPN	73.13	55.57

表5 Miniimagenet 消融实验结果/%

基准方法为 PN	5-way 5-shot	5-way 1-shot
Resnet12	70.07	52.48
Resnet12 + SAM	70.78	52.94
Resnet12 + CC	72.62	54.06
IBN-Resnet12	70.75	53.19
IBN-Resnet12 + CC	72.80	54.02
IBN-Resnet12 + SAM	71.01	53.96
IBN-Resnet12 + SAM + CC	73.13	55.57

3.3.2 Tieredimagenet 实验对比

本节按照上述 Miniimagenet 的实验对比方法, 在 Tieredimagenet 上同样进行了大量的实验以及消融实验, 验证了 ACCPN 模型的有效性, 结果如表 6

所示。本文在 Tieredimagenet 数据集上与 PN(4-conv)^[11]、PN(Resnet12)^[11]、Relation Networks^[12]以及其他小样本学习算法进行了比较。由表 6 得知, ACCPN 在 Tieredimagenet 数据集上取得了较好的效果,对于 5-way 5-shot 任务, ACCPN 相对于 PN(4-conv), 准确率提升了 7.76%, 相对于 PN(Resnet12), 准确率提升了 1.67%。对于 5-way 1-shot 任务, ACCPN 相对于 PN(4-conv), 准确率提升了 6.09%, 相对于 PN(Resnet12), 准确率提升了 1.55%。为证明各个改进部分的有效性, 本文进行了大量的消融实验, 其结果如表 7 所示。

表 6 Tieredimagenet 数据性能比较/%

方法	5-way 5-shot	5-way 1-shot
PN(4-conv)	71.69	52.31
PN(Resnet12)	77.78	56.85
Relation Networks	78.32	54.48
WDSI ^[15]	78.90	56.76
ROT ^[28]	79.87	59.62
ACCPN	79.45	58.40

表 7 Tieredimagenet 消融实验结果/%

基准方法为 PN	5-way 5-shot	5-way 1-shot
Resnet12	77.78	56.85
Resnet12 + SAM	78.46	57.24
Resnet12 + CC	79.02	58.01
IBN-Resnet12	78.81	57.49
IBN-Resnet12 + CC	79.21	58.11
IBN-Resnet12 + SAM	79.02	57.92
IBN-Resnet12 + SAM + CC	79.45	58.40

由表 7 的消融实验结果得知, 本文提出的 IBN-Resnet12、SAM 模块以及类协方差度量(CC)对于小样本图像分类都具有一定的效果。对于 5-way 5-shot 任务, IBN-Resnet12 结构相比于 Resnet12 结构的精度提升了 1.03% 左右, 加入了 SAM 的 PN 相比于基准 PN 的精度提升约为 0.68%, 本文提出的类协方差度量(CC)相对于基准 PN 提升精度为 1.24% 左右。对于 5-way 1-shot 任务, IBN-Resnet12 结构相比于 Resnet12 结构的精度提升了 0.64%, 加入了 SAM 的 PN 相比于基准 PN 的精度提升约为

0.39%, 类协方差度量(CC)相对于基准 PN 提升了 1.16% 的精确度。

3.3.3 模型复杂度分析

由上述实验可知 ACCPN 相对于基准的 PN 以及其他小样本学习算法有了较大的性能提升, 接下来进行该模型的复杂度分析。ACCPN 由特征提取器和分类器构成, 并非端到端的结构, 故本文分别对特征提取器和分类器进行复杂度分析。首先对于特征提取部分, 相对于 PN, 本文的 ACCPN 在 CNN 后加入了 SAM, 故在参数计算时也只是加入了 SAM 的参数量。具体参数量及时间复杂度对比如表 8 所示。

表 8 ACCPN 和 PN 参数量及时间复杂度对比

	总参数量/个	浮点数/个
ACCPN	7 996 898	2 474 583 954
PN	7 996 800	2 474 581 504

由表 8 可得, ACCPN 相对于 PN 的总参数量只多了 98, 而时间复杂度可用浮点数来衡量, 浮点数意指浮点运算量, 由表 8 可得, ACCPN 的浮点数相比于 PN 只多了 2405。

对于分类器部分, 由于 ACCPN 需要计算类协方差距离, 故比 PN 的欧氏距离计算所需的时间要长。但由于类协方差距离是固定距离计算, 不引入额外的参数, 故总参数量是不变的。

综上所述, ACCPN 在只增加少量参数及时间复杂度的前提下, 准确率有了较大的提升。

4 结论

本文提出了一种基于自注意力和类协方差度量的原型网络(ACCPN), 相对于基准原型网络, 本文从特征提取模块和分类器部分都做了改进。首先利用 IN 的图像外观不变性, 提出 IBN-Resnet12 作为特征提取结构, 使模型在保持学习能力的同时, 增加模型泛化能力; 其次在特征提取模块后嵌入空间注意力模块, 增强局部特征; 最后提出类协方差距离度量, 突出了图像特征间的关联性。本文在小样本学

习经典数据集 Miniimagenet 和 Tieredimagenet 上做了大量实验,证明了模型的有效性。

小样本学习旨在训练模型在给定样本较少的情
况下完成一系列任务,而当前小样本学习方法多基
于度量学习的思想,而忽略了对样本数据分布的建
模,故小样本学习未来的发展应重点放在样本本身,
依据不同样本的数据分布特点建立泛化性更好的小
样本学习模型。同时,当前小样本学习任务多为图
像分类,未来的应用场景应拓展到目标检测、语义分
割和文本识别等多方面。

参考文献

- [1] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back propagating errors[J]. *Nature*, 1986,323(6088):533-536
- [2] LI F F, FERGUS R, PERONA P. One-shot learning of object categories [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(4): 594-611
- [3] LI F F, FERGUS R, PERONA P. A Bayesian approach to unsupervised one-shot learning of object categories[C] //Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France, 2003:1134-1141
- [4] 刘颖,雷研博,范九伦,等. 基于小样本学习的图像分类技术综述[J]. *自动化学报*, 2021,47(2):297-315
- [5] 魏胜楠,张景异,陈亮,等. 自适应局部关系网络的小样本学习方法[J]. *沈阳理工大学学报*, 2021,40(4): 35-41
- [6] PAN X, PING L, SHI J, et al. Two at once: enhancing learning and generalization capacities via IBN-Net [EB/OL]. <http://arxiv.org/abs/1807.09444>; arXiv, (2018-07-25), [2021-06-29]
- [7] QIAO S Y, LIU C X, SHEN W, et al. Few-shot image recognition by predicting parameters from activations[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake, USA, 2018: 7229-7238
- [8] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks [C] //Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 2017: 1126-1135
- [9] MUNKHDALAI T, YU H. Meta networks[C]//Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 2017: 2554-2563
- [10] KOCH G, ZEMEL R, SALAKHUTDINOV R. Siamese neural network for one-shot image recognition [C] //International Conference on Machine Learning, Lille, France, 2015: 1-14
- [11] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few shot learning[C] //Advances in Neural Information Processing Systems, Long Beach, USA, 2017: 4077-4087
- [12] SUNG F, YANG Y X, ZHANG Li, et al. Learning to compare: relation network for few-shot learning [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake, USA, 2018:1199-1208
- [13] 王年,孟树林,吴洛天,等. 基于改进关系网络的小样本学习[J]. *安徽大学学报(自然科学版)*, 2020,44(4):38-44
- [14] ZHANG C, CAI Y, LIN G, et al. DeepEMD: few-shot image classification with differentiable earth mover's distance and structured classifiers [C] //2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, 2020:12200-12210
- [15] SU J C, MAJI S, HARIHARAN B. When does self-supervision improve few-shot learning? [C] //European Conference on Computer Vision, Glasgow, UK, 2020: 645-666
- [16] ZHANG M L, ZHANG J H, LU Z W, et al. IEPT: instance-level and episode-level pretext tasks for few-shot learning[C]//International Conference on Learning Representations, Vienna, Austria, 2021: 1-16
- [17] ZHANG M, WANG S, YANG D, et al. Spatial attention model based target detection for aerial robotic systems [J]. *International Journal of Intelligent Robotics and Applications*, 2019, 3(5): 4721-4734
- [18] VINIYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning[C] //Advances in Neural Information Processing Systems, Barcelona, Spain, 2016:3630-3638
- [19] REN M, TRIANTAFILLOU E, RAIN S, et al. Meta-learning for semi-supervised few-shot classification [EB/OL]. <http://arxiv.org/abs/1803.00676>; arXiv, (2018-03-02), [2021-06-29]

- [20] SIMON M, RODNER E, DENZLER J. ImageNet pre-trained models with batch normalization [J]. *Machine Learning*, 2016,18(15):457-468
- [21] ULYANOV D, VEDALDI A, LEMPITSKY V. Instance normalization: the missing ingredient for fast stylization [EB/OL]. <http://arxiv.org/abs/1607.08022>; arXiv, (2016-07-27), [2021-06-29]
- [22] 李涛涛. 基于局部描述子的小样本学习方法研究 [D]. 南京: 南京大学计算机科学与技术系, 2019
- [23] 李航. 统计学习方法 [M]. 北京: 清华大学出版社, 2019:206-265
- [24] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C] // *Advances in Neural Information Processing Systems*, Lake Tahoe, USA, 2012: 1-9
- [25] CA I Q, PAN Y, YAO T, et al. Memory matching networks for one-shot image recognition [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 4080-4088
- [26] MISHRA N, ROHANINEJAD M, CHEN X, et al. A simple neural attentive meta-learner [EB/OL]. <http://arxiv.org/abs/1707.03141>; arXiv, (2017-07-11), [2021-06-29]
- [27] MUNKHDALAI T, YUAN X, MEHRI S, et al. Rapid adaptation with conditionally shifted neurons [EB/OL]. <http://arxiv.org/abs/1712.09926>; arXiv, (2017-12-28), [2021-06-29]
- [28] GIDARIS A, BURSUC N, KOMODAKIS P, et al. Cord, boosting few-shot visual learning with self-supervision [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 2019: 8058-8067

Few shot learning based on spatial attention and class covariance metrics

LI Guoqiang, WANG Tianlei, GONG Ning, WANG Junyan
(School of Electrical Engineering, Yanshan University, Qinhuangdao 071000)

Abstract

In recent years, few-shot learning has gradually been widely studied in academia, which aims to make models accomplish a series of tasks with a small number of given samples. At present, metric-based meta learning algorithms are widely used in few-shot learning. The idea of metric learning is used to improve the prototype network of benchmark metric learning algorithm and the attention class covariance prototypical network (ACCPN) is proposed. First, IBN-Resnet12 is proposed as the feature extraction part to increase the generalization ability of the model; then, the spatial attention module is added after the feature extraction module to effectively enhance the local features; finally, the class covariance metric is used as the final metric classifier to complete the modeling of correlation among image feature dimensions. The experiments are conducted on the few-shot learning classical datasets to prove the effectiveness of the model; also, a large number of ablation experiments are conducted in this paper to prove the effectiveness of each part of the model improvement.

Key words: few-shot learning, metric-based meta learning, generalization capability, spatial attention, class covariance metric