基于关键局部语义对齐的小样本图像分类算法

孙 哲,郑 旺,郭朋飞
(燕山大学信息科学与工程学院,河北秦皇岛 066004)

摘 要: 小样本分类旨在从少量标记样本中学习识别新类. 目前基于局部描述符的小样本分类方法因考虑了局部特征在可见类和不可见类中的一致性取得了较好的分类性能. 然而,基于局部描述符的表示方法存在邻近表示信息冗余、部分表示与图像语义无关、可解释性差等问题. 鉴于此,本文提出一种基于关键局部语义对齐的小样本图像分类算法(Key Local Semantic Alignment Network,KLSANet),该方法通过对齐局部语义来实现图像到类的度量以完成分类. 为了减轻图像语义无关局部对分类的影响,本文进一步设计了关键局部筛选模块并通过设置阈值筛选出关键局部块. KLSANet在三个广泛使用的基准数据集上均表现出较好的分类性能,尤其在1-shot和5-shot设置上比最优的对比方法平均提高了3.95%和2.56%. 本文的代码公布在:https://github.com/ZitZhengWang/KLSANet.

关键词: 小样本分类;局部特征;语义对齐;关键局部筛选
基金项目: 国家自然科学基金(No.62001413);河北省高等学校科学技术研究项目——青年拔尖人才项目(No.BJK2023117);燕山大学基础创新科研培育重点项目(No.2023LGZD006)
中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112(2024)10-0001-10
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20240209

Few-Shot Image Classification Algorithm Based on Key Local Semantic Alignment

SUN Zhe, ZHENG Wang, GUO Peng-fei
(*Department of Information Science and Engineering*, *Yanshan University*, *Qinhuangdao*, *Hebei* 066004, *China*)

Abstract: Few-shot classification aims to recognize new classes with a limited number of labeled samples. Currently, methods based on local descriptors achieve good performance by leveraging the consistency of local features in both visible and unseen classes. However, these methods often suffer from issues such as redundant neighboring representations, irrelevance to image semantics, and poor interpretability. To address the above problems, this paper proposes a Key Local Semantic Alignment Network(KLSANet), a few-shot image classification approach based on key local semantic alignment network, which improves few-shot image classification by aligning local semantics for image-to-class measurement. To minimize the impact of semantically irrelevant local parts, we design a key local screening module that filters out non-essential blocks using set thresholds. KLSANet demonstrates superior performance on three benchmark datasets, outperforming the best comparison methods by 3.95% and 2.56% in the 1-shot and 5-shot settings, respectively. The code is available at: https://github.com/ZitZhengWang/KLSANet.

Key words: few-shot classification; local features; semantic alignment; key local screening
Foundation Item(s): National Natural Science Foundation of China (No.62001413); Science and Technology Project of Hebei Education Department (No.BJK2023117); Key Project of Basic Innovation and Scientific Research Cultivation of Yanshan University (No.2023LGZD006)

1 引言

深度学习(Deep Learning,DL)在有大规模标记训练数据领域的性能已达到人类水平. 然而,从少量标记数据中学习仍然是个具有挑战性的任务. 缺乏训练数据通常导致模型过度拟合而无法提供良好的泛化能力. 相反,人类可以通过先前的经验和知识非常快速地从少量示例中学习. 为了弥补人工智能与人类智能的差距,小样本学习(Few-shot Learning,FSL)越来越受到

收稿日期:2024-03-04;修回日期:2024-06-17;责任编辑:赵克

研究者们的关注[1~3].

小样本学习旨在训练一个模型,使其仅通过少量标记样本即可有效地泛化到新任务上[4]. 它面临的一个巨大挑战是如何从少量样本中提取可泛化的信息来提高模型在新任务上的性能. 为解决该挑战,基于元学习的方法[5~9]通过引入跨任务的训练机制来学习可迁移的元知识以泛化到未见过的新任务上,在小样本学习领域取得了重大的进展. 基于元学习的方法受到人类有效使用先验知识的启发[9,10],获取先前学习任务的先验知识从而可以从少量数据中有效地学习新任务. 其中,先验知识的形式可能是良好的初始模型参数、优化策略、距离度量空间等. 根据先验知识的种类,现有的基于元学习的方法可以分为三类:基于优化的方法[11~15]、基于外部记忆的方法[16~20]和基于度量的方法[7,21~27]. 其中,基于度量的方法通过计算查询样本和少量支持样本间的相似度来实现小样本分类并因其简单有效而显示出巨大的潜力,代表方法如:匹配网络[7]为支持集和查询集样本学习不同的特征提取器,然后通过比较查询样本与支持样本的余弦距离来完成分类;原型网络(Prototypical Networks,PN)[21]引入类原型的概念来表示支持类,并通过比较查询样本与类原型间的欧氏距离来实现小样本图像分类.

现有的基于度量学习的方法大多数都采用图像级的特征来表示整个图像[7,21~27],并隐含地假设在可见类样本上训练得到的图像级特征依然适用于不可见类. 例如,GNN[22]将图神经网络引入小样本学习并用其学习一个基于全局特征的度量. 相比于全局图像表示,低层次的局部表示(例如:局部描述符、局部特征)更容易在不同的类中共享,并迁移到不可见的类上. 其中局部描述符(Local Descriptors,LD)是指将整个图像作为CNN的输入,并把三维特征图每个空间位置的特征视为一个局部描述符,而局部特征是使用CNN提取图像随机裁剪局部的特征. 最近,一些方法[28~37]提出使用局部描述符表示来解决小样本图像分类问题. 例如,DN4[28]将全局表示替换为深度局部描述符表示并且通过k近邻显式地利用局部描述符,证明了局部表示比全局表示更有效. ATL-Net[32]引入了任务级的注意力机制挖掘局部描述符蕴含的深层信息进一步选择重要的局部表示. 此外,AGLRs[38]和DFENet[39]同时采用图像全局特征和局部描述符计算查询样本到类的相似度. 然而,空间位置邻近的局部描述符之间通常伴随着大量冗余信息,并且很多局部描述符与图像语义缺乏关联. 加之深度网络中密集的池化操作很难从图像语义层面解释局部描述符对分类的贡献. 与此同时,对比学习的方法[40,41]取得了令人瞩目的进步,其核心思想是将图像中随机裁剪的局部区域视为单独的类并将其送入CNN进行对比学习.

由于局部描述符的局限性以及受到对比学习方法的启发,本文提出一种基于关键局部语义对齐的小样本图像分类算法(Key Local Semantic Alignment Network,KLSANet). 该方法提取图像随机裁剪产生的局部特征并通过对齐查询样本与支持类的局部语义实现了一种图像到类的度量,其核心思想是让模型学习关键的局部概念并且用支持类的局部粗略地组合出查询图像. KLSANet首先提取查询样本和支持样本随机裁剪的局部特征,然后通过k近邻在支持类的候选局部池中搜索k个最相关局部来与查询样本的局部对齐,最后度量查询图像到类的相似度完成小样本分类. 此外,为了减小随机裁剪的查询样本局部中的背景块或图像语义无关块的影响,KLSANet还设计一个关键局部筛选模块,用一种简单有效的策略过滤图像语义无关的局部块从而选出关键的查询样本局部块. 具体来说,关键局部筛选模块通过设置阈值来过滤掉与图像全局表示相似度低的查询样本候选局部. 与现有的方法不同,KLSANet用一系列随机裁剪的局部特征来表示一张图像,而不是全局表示或局部描述符. 这种局部表示方式有助于提升小样本学习的性能,原因主要有以下三点:第一,局部表示更容易在不同的类中共享,具有更强的泛化性;第二,将一系列随机裁剪的局部区域送入模型训练变相地增加了样本数量和类别多样性,有助于缓解小样本学习的过拟合问题;第三,从图像中裁剪出的局部语义更容易被人类理解,从而能够从语义层面解释其对分类的贡献. 本文在三个小样本学习基准数据集(包括CUB、Stanford Dogs和Stanford Cars)上的实验定量和定性地展示了KLSANet的有效性. 本文提出的KLSANet在三个数据集的1-shot和5-shot设置上分别比最先进的对比方法平均提高了3.95%和2.56%,比次优的方法提高了6.87%和4.12%,证明了KLSANet的有效性. 此外,本文通过可视化查询样本关键局部以及对应支持类中选出的相关局部展示了KLSANet预测结果背后的可解释性. 综上所述,本文的主要贡献可以总结为以下三点:

(1)本文提出了一种新颖的基于局部特征的小样本图像分类方法,它通过对齐查询样本和支持样本的局部语义来实现一种图像到类的度量以完成小样本图像分类任务.

(2)本文设计了一个关键局部筛选模块,该模块通过设置阈值过滤掉候选局部中与图像语义无关的局部块从而选出对分类有效的关键局部特征.

(3)本文在三个常用的小样本学习基准数据集上进行了定量和定性的实验. 实验结果表明,本文所提出算法的分类性能显著优于最先进的对比方法,并且可

视化实验展示了KLSANet的可解释性.

2　基于关键局部语义对齐小样本图像分类方法

2.1　问题描述

小样本图像分类任务通常包含一个由基类构成的训练集D_{base}和一个由新类构成的测试集D_{novel},且训练集和测试集包含的类不相交,$D_{\text{base}}\cap D_{\text{novel}}=\varnothing$. 其目标是从基类中学习来对新类进行分类. 受跨任务训练机制启发[7],小样本学习通常分为两个阶段:元训练阶段(meta-training)和元测试阶段(meta-testing). 元训练阶段,模型在一系列从D_{base}中构建的训练任务$\left\{T_{\text{train}}^1,T_{\text{train}}^2,\cdots,T_{\text{train}}^{N_T}\right\}$上跨任务元学习;对于元测试阶段,在从$D_{\text{novel}}$中构建的测试任务$T_{\text{test}}$上评估训练好的模型. 由基类构建的训练任务通常模拟由新类构建的新任务以减小训练和测试之间的差距、增强模型的泛化能力. 具体而言,训练任务的构建方式为:首先从训练集D_{base}中随机采样N个类,然后从每一类中随机采样K个样本构成支持集$S=\left\{\left(x_i,y_i\right)\right\},i\in\{1,2,\cdots,N\times K\}$,并从这$N$类的剩余样本中每类随机采样$q$个样本构成查询集$Q=\left\{\left(x_i,y_i\right)\right\},i\in\{1,2,\cdots,N\times q\}$,其中$x_i$和$y_i$分别表示图像与其对应标签,$y_i\in\{1,2,\cdots,N\}$,$S$和$Q$一起构成一个训练任务$T_{\text{train}}=\{S,Q\}$. 在$D_{\text{novel}}$上构建测试任务的过程与训练任务几乎一样,区别在于测试任务中查询样本的标签是未知的,$Q=\left\{\left(x_i\right)\right\},i\in\{1,2,\cdots,N\times q\}$. 通常,如果支持集包含$N$个不同的类且每一类有$K$个样本,那么就把目标小样本问题称为*N-way K-shot*问题.

2.2　算法概述

本文提出算法的整体流程如图1所示,它主要包含三个部分:特征提取模块(Feature Extraction Module, FEM)、关键局部筛选模块(Key Local Screening Module, KLSM)和语义相似度度量模块(Semantic Similarity Measurement Module, SSMM). 首先,特征提取模块(FEM)提取支持样本和查询样本随机裁剪的候选局部块的特征以及查询样本的全局特征. 其次,关键局部筛选模块(KLSM)从查询样本的所有候选局部特征中选出一些重要的局部块用于后续的分类. 最后,语义相似度度量模块(SSMM)通过计算支持样本候选局部特征与查询样本的重要局部特征之间的相似度矩阵对齐语义并将查询样本分类到具有最高相似度的支持类中.

为了更好地理解KLSANet,算法1总结了提出算法的推理过程.

算法1　KLSANet的推理过程

输入:训练任务集$D_T=\left\{T_{\text{train}}^1,T_{\text{train}}^2,...,T_{\text{train}}^{N_T}\right\}$,其中$T_{\text{train}}=\{S,Q\}$,$S=\left\{\left(x_i,y_i\right)\right\},i\in\{1,2,\cdots,N\times K\}$,$Q=\left\{\left(x_i,y_i\right)\right\},i\in\{1,2,\cdots,N\times q\}$.

初始化:加载预训练的特征提取模块参数θ;初始化余弦尺度参数τ为1;令m=36,α=0.2.

for T_{train} in D_T **do**　// 处理每一个训练任务
  for x^S in S **do**
    // 提取候选支持局部特征
    从支持样本x^S中随机裁剪m个候选局部块$\left\{x_S^{P_i}\right\},i\in\{1,2,\cdots,m\}$;
    提取候选支持局部特征$\left\{f_\theta\left(x_S^{P_i}\right)\right\},i\in\{1,2,\cdots,m\}$;
  end for
  for x^Q in Q **do**
    // 提取查询样本全局特征和候选查询局部特征
    提取查询样本全局特征$f_\theta\left(x^G\right)$;
    从查询样本x^Q中随机裁剪m个候选局部块$\left\{x_Q^{P_j}\right\},j\in\{1,2,\cdots,m\}$;
    提取候选查询局部特征$\left\{f_\theta\left(x_Q^{P_j}\right)\right\},j\in\{1,2,\cdots,m\}$;
    // 筛选关键查询样本局部
    $s^{P_iG}=\cos\left(f_\theta\left(x^{P_i}\right),f_\theta\left(x^G\right)\right)\quad,i\in\{1,2,\cdots,m\}$
    // 计算候选查询局部与其全局表示的相似度
    $$f_\theta^*\left(x^{P_i}\right)=\begin{cases}0 & ,s^{P_iG}\le\alpha\\ f_\theta\left(x^{P_i}\right) & ,s^{P_iG}>\alpha\end{cases}$$
    // 通过阈值过滤筛选出关键查询局部特征
    // 度量查询样本到支持类的相似度
    for c in $\{1,2,...,N\}$ **do**
      $$\boldsymbol{M}_c=\left[s^{P_iP_j}\right]_{mK\times m}=\left[\tau\cdot\cos\left(f_\theta\left(x_{S_c}^{P_i}\right),f_\theta^*\left(x_Q^{P_j}\right)\right)\right]_{mK\times m}$$
      // 计算关键查询局部与支持类c所有候选局部之间的相似度
      根据相似度选择与每个关键查询局部最相似的k个支持局部$x_{S_c}^{\hat{P}_i}\Big|_{i=1}^{k}$;
      $$p_c=\sum_{j=1}^{m}\sum_{i=1}^{k}s^{\hat{P}_iP_j}=\tau\cdot\sum_{j=1}^{m}\sum_{i=1}^{k}\cos\left(f_\theta\left(x_{S_c}^{\hat{P}_i}\right),f_\theta^*\left(x_Q^{P_j}\right)\right)$$
      // 计算查询样本到类c的相似度分数
    end for
    $$p\left(y=c\middle|x^Q\right)=\frac{\mathrm{e}^{p_c}}{\sum_{c'}\mathrm{e}^{p_{c'}}}$$
    // 推理查询样本属于支持类c的概率
  end for
end for

3　网络结构

3.1　特征提取模块

不同于现有的小样本学习方法提取全局特征或局部描述符,本文提出的方法利用特征提取模块(FEM)

提取局部块特征，这将鼓励网络学习更可能泛化到新类上的局部特征．为此，本文首先通过随机裁剪从输入图像 x 中获取 m 个候选局部块 $\{x^{P_i}\}, i\in\{1,2,\cdots,m\}$，然后将其输入特征提取模块得到对应的 m 个候选局部特征 $\{f_\theta(x^{P_i})\}, i\in\{1,2,\cdots,m\}$．特征提取模块由一个参数化为 θ 的 CNN 实现，即 f_θ: $\mathbb{R}^{C\times H\times W}\to\mathbb{R}^D$，它将用于提取支持样本和查询样本所有候选局部块的特征．

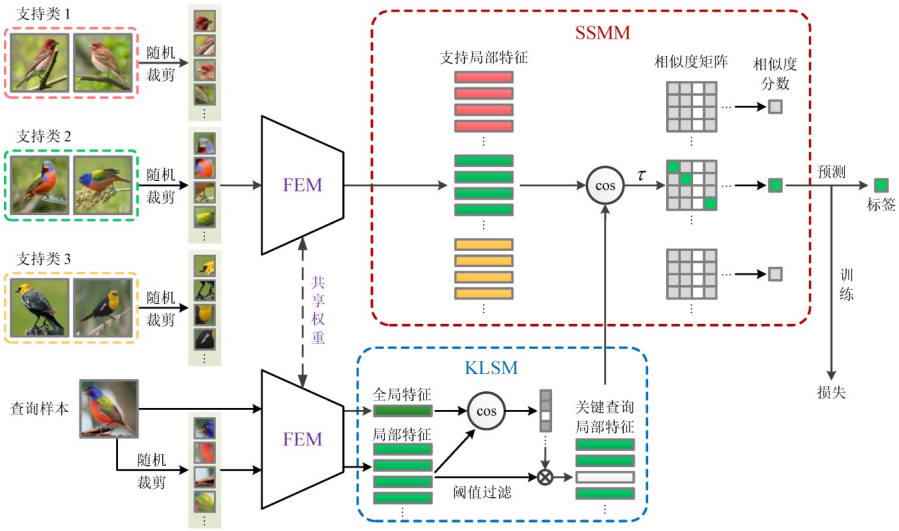


图 1　KLSANet 在 3-way 2-shot 小样本图像分类任务上的工作流程

3.2　关键局部筛选模块

由于候选局部块是通过随机裁剪生成的，因此其中不可避免地包含了一些杂乱的背景块或图像语义无关块．考虑到直接将所有候选查询局部与候选支持局部对齐并计算相似度分数将会引入噪声，从而损害分类性能．本文提出一个关键局部筛选模块(KLSM)，用一种简单有效的策略来过滤掉图像语义无关局部从而选出关键查询局部．具体而言，对于每一个查询样本 x^Q，除了用特征提取模块提取 m 个候选局部特征以外，还会提取其全局特征表示 $f_\theta(x^G)$ 用于筛选关键查询局部．筛选过程主要包含两个步骤，第一步，计算每个候选查询局部特征与全局特征表示之间的相似度，如式(1)所示：

$$s^{P_iG}=\cos\left(f_\theta(x^{P_i}),f_\theta(x^G)\right), i\in\{1,2,\cdots,m\} \tag{1}$$

其中，s^{P_iG} 表示查询样本 x^Q 的第 i 个候选局部 x^{P_i} 与其全局表示 $f_\theta(x^G)$ 之间的相似度．第二步，将相似度低于阈值 α 的候选局部特征置零以消除其对分类的负面影响，如式(2)所示：

$$f_\theta^*(x^{P_i})=\begin{cases}0, & s^{P_iG}\leqslant\alpha\\ f_\theta(x^{P_i}), & s^{P_iG}>\alpha\end{cases} \tag{2}$$

其中，α 是一个可调整的超参数，$f_\theta^*(x^{P_i})$ 表示筛选出的关键查询局部特征．这种筛选策略的基本原理是：相比于背景块或图像语义无关的块，关键局部块与图像全局表示具有更高的相似度．使用阈值筛选策略的另一个原因是关键局部块的数量会随着当前任务和局部裁剪情况而动态地变化．筛选后的查询样本关键局部将被送入语义相似度度量模块(SSMM)计算与支持样本候选局部的相似度．

3.3　语义相似度度量模块

为了更好地利用学习到的局部特征表示，本文使用语义相似度度量模块(SSMM)来对齐局部语义并计算图像到类的相似度．具体而言，给定一个支持类 c 的所有的候选局部特征 $\{f_\theta(x_{S_c}^{P_i})\}, i\in\{1,2,\cdots,mK\}$ 和筛选后的关键查询局部特征 $\{f_\theta^*(x^{P_i})\}, i\in\{1,2,\cdots,m\}$，首先计算每个关键查询局部与每个支持类候选局部之间的相似度并获得相似度矩阵 $\boldsymbol{M}_c$，如式(3)所示：

$$\begin{aligned}\boldsymbol{M}_c&=\left[s^{P_iP_j}\right]_{mK\times m}\\&=\left[\tau\cdot\cos\left(f_\theta(x_{S_c}^{P_i}),f_\theta^*(x^{P_j})\right)\right]_{mK\times m}\end{aligned} \tag{3}$$

其中，$s^{P_iP_j}$ 表示支持类的第 i 个候选局部与查询样本的第 j 个关键局部之间的相似度，K 表示支持类中的样本数，τ 表示一个可学习的尺度参数[42]．相似度矩阵 $\boldsymbol{M}_c$ 中所有元素的和在一定程度上表示查询样本与支持类之间的相似度．若直接对所有元素求和会将一些不相关的局部块间关系考虑在内从而损害相似度的准确性．因此，本文仅考虑与每个关键查询局部最相似的 k 个近邻支持局部 $x_{S_c}^{\hat{P}_i}\big|_{i=1}^{k}$ 并计算查询样本到支持类的相似度分数 p_c，即对相似度矩阵 $\boldsymbol{M}_c$ 中每列值最大的 k 个元素

求和,如式(4)所示:

$$p_c=\sum_{j=1}^{m}\sum_{i=1}^{k}s^{\hat{P}_i P_j}=\tau\cdot\sum_{j=1}^{m}\sum_{i=1}^{k}\cos\left(f_\theta\left(x_{S_c^i}^{\hat{P}_i}\right),f_\theta^*\left(x^{P_j}\right)\right)\tag{4}$$

其中,$s^{\hat{P}_i P_j}$表示第i个近邻支持局部与查询样本第j个关键局部之间的相似度. 最后,根据查询样本与各支持类的相似度分数来预测样本x^Q属于支持类c的概率,如式(5)所示:

$$p\left(y=c\left|x^Q\right.\right)=\frac{\mathrm{e}^{p_c}}{\sum_{c'}\mathrm{e}^{p_{c'}}}\tag{5}$$

4　实验结果与分析

4.1　数据集

本文在三个细粒度图像分类数据集上评估了提出算法的有效性,分别是CUB-200-2011[43]、Stanford Dogs[44]和Stanford Cars[45],其统计概况和数据划分如表1所示.

CUB-200-2011包含200种鸟类的细粒度数据集,共计11 788张彩色图像. 本文遵循Hilliard等人[46]提出的评估协议,随机采样100、50、50分别用于训练、验证和测试. Stanford Dogs是常用的细粒度图像分类任务的基准数据集,包含120个品种的狗,总计20 580张图像. 本文遵循一般的数据集划分方法[28],随机采样70、20、30分别用于训练、验证和测试. Stanford Cars也是常用的细粒度图像分类任务的基准数据集,由196类汽车,共计16 185张图像组成. 同样遵循一般的划分方法[28],随机采样130、17、49个类的图像分别用于训练、验证和测试.

表1　基准数据集统计概况及划分

数据集	总图像数	总类别数	训练/验证/测试类别划分
CUB-200-2011	11 788	200	100/50/50
Stanford Dogs	20 580	120	70/20/30
Stanford Cars	16 185	196	130/17/49

4.2　实验设置

实施细节　本文在广泛使用的5-way 1-shot和5-way 5-shot小样本分类设置上评估KLSANet的分类性能,并且使用PyTorch在GeForce RTX 4070Ti GPU上实施实验. 为公平地与其他方法比较,本文使用了两种小样本学习中常用的主干网络结构Conv-4和ResNet-12,具体实现与文献[8]一致. 在训练之前,将所有图像的尺寸调整为84×84并使用了标准的数据增强方法,包括随机裁剪、颜色抖动以及水平翻转.

训练　本文提出的KLSANet的训练过程包含两个阶段:预训练阶段和元训练阶段. 在预训练阶段,和Meta-Baseline[8]一样,本文使用标准的交叉熵损失以传统分类任务的方式在所有基类上训练一个分类器,然后移除预训练模型最后的全连接层并保留特征提取模块的参数作为第二个训练阶段的初始参数. 预训练时使用初始学习率为0.01、权重衰减为0的Adam优化器和交叉熵损失函数,batch size为128. 在元训练阶段,模型以跨任务的方式训练,每一个训练任务包含5个类,每一类包含K(1或5)个支持样本以及q=15个查询样本. 元训练前,本文进一步从每个增强图像中随机裁剪m=36个26×26的局部块来获取候选局部图像并将其作为模型的输入. 关键局部筛选模块的阈值和语义相似度度量模块的近邻数被以实验的方式设置为α=0.2和k=5. 元训练过程中使用初始学习率为0.01、权重衰减为0的Adam优化器和交叉熵损失,余弦尺度参数τ的初始值为1.

验证和测试　实验中根据验证准确率选择最佳模型,然后在具有新类的测试集上评估模型. 从测试集随机采样的新任务与训练任务一样,包含5个类,每类包含K(1或5)个支持样本以及15个查询样本. 测试结果取600个新任务的平均准确率以及95%置信区间.

4.3　与先进的方法比较

为了评估提出算法的分类性能,本节将提出的KLSANet在三个细粒度数据集上与13个最先进的小样本分类方法进行比较,结果总结在表2中.

这些对比方法包括:基于全局表示的PN[21]、GNN[22]和QPN[27],基于局部描述符的DN4[28]、DN4-DA[28]、RelationNet[29]、CovaMNet[33]、MADN4[34]、TD-SNet[35]、LMPNet[36]、BDLA[37],以及基于全局和局部描述符表示的AGLRs[38]. 对比算法中最优和次优的结果分别用粗体和下划线标记.

从表2中可以看出,本文提出的KLSANet使用ResNet-12主干网络时的准确率在三个数据集的大多数设置上显著超过所有对比方法. 具体而言,KLSANet在三个数据集的1-shot和5-shot设置上分别比最佳的对比方法平均提高了3.95%和2.56%,比次优的方法提高了6.87%和4.12%,仅在CUB数据集的1-shot和5-shot设置上甚至比最佳的对比方法提高了5.60%和4.20%. 与基于全局表示方法中最优的QPN相比,KLSANet在三个数据集的1-shot和5-shot设置上的平均提升分别达到了10.05%和4.91%,这种大幅的领先表明了基于局部特征的方法比基于全局表示的方法更适合小样本学习. 与基于局部描述符的方法相比,KLSANet比DN4、ATL-Net、BDLA等方法都有更好的表现. 例如,KLSANet在三个数据集的1-shot和5-shot设置上比基于局部描述符方法中表现良好的ATL-Net平均提升了

表2 在三个细粒度数据集(CUB、Stanford Dogs和Stanford Cars)上与最先进方法对比

方法	主干网络	CUB		Stanford Dogs		Stanford Cars	
		1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
PN	Conv-4	51.31±0.91	70.77±0.69	37.80±0.99	48.19±1.03	40.90±1.01	52.93±1.03
RelationNet	Conv-4	62.45±0.98	76.11±0.69	43.33±0.42	55.23±0.41	47.67±0.47	60.59±0.40
GNN	Conv-4	51.83±0.98	63.69±0.94	46.98±0.98	62.27±0.95	55.85±0.97	71.25±0.89
QPN	Conv-4	66.04±0.82	82.85±0.76	53.69±0.62	70.98±0.70	63.91±0.58	89.27±0.78
DN4	Conv-4	46.84±0.81	74.92±0.64	45.41±0.76	63.51±0.62	59.84±0.80	88.65±0.44
DN4-DA	Conv-4	53.15±0.84	81.90±0.60	45.73±0.76	66.33±0.66	61.51±0.85	**89.60±0.44**
CovaMNet	Conv-4	52.42±0.76	63.76±0.64	49.10±0.76	63.04±0.65	56.65±0.86	71.33±0.62
ATL-Net	Conv-4	60.91±0.91	77.05±0.67	54.49±0.92	73.20±0.69	67.95±0.84	89.16±0.48
MADN4	Conv-4	57.11±0.70	77.83±0.40	50.42±0.27	70.75±0.47	62.89±0.50	89.25±0.34
TDSNet	Conv-4	**69.34±0.89**	80.34±0.59	54.48±0.87	69.45±0.69	62.14±0.91	75.64±0.72
BDLA	Conv-4	50.59±0.97	75.36±0.72	48.53±0.87	70.07±0.70	64.41±0.84	89.04±0.45
AGLRs	Conv-4	**69.34±0.70**	**84.72±0.42**	58.85±0.69	**75.82±0.49**	**70.71±0.66**	89.42±0.33
Ours	Conv-4	66.70±0.82	83.63±0.28	52.23±0.56	70.45±0.37	54.71±0.77	78.47±0.57
PN	ResNet-12	66.09±0.92	82.50±0.58	—	—	—	—
LMPNet	ResNet-12	65.59±0.13	68.19±0.23	**61.89±0.10**	68.21±0.11	68.31±0.45	80.27±0.23
Ours	ResNet-12	**74.94±0.43**	**88.92±0.41**	**64.43±0.81**	**81.07±0.31**	**74.43±0.76**	87.84±0.45

10.15%和6.14%. KLSANet比基于局部描述符的方法好的原因可能是使用了更加丰富的、可解释性更强的局部特征表示. 值得注意的是,AGLRs方法同时采用了全局表示和局部描述符表示并且融合了全局和局部相似度,这使它在三个数据集上的整体表现大幅超过了其他对比方法. 融合全局表示和局部描述符表示的多种不同尺度的度量可能是一个更好的策略,但是本文提出的KLSANet仅使用局部特征表示就超过了AGLRs,在三个数据集的1-shot和5-shot设置上的平均提升达到了4.97%和2.62%. 这种结果有力地说明了使用局部特征表示的有效性. 需要强调的是:在小样本的背景下,更深的主干网络并不能稳定地提升模型的性能,反而会导致更严重的过拟合. 本文使用ResNet-12主干网络与其他方法对比主要是因为提出的方法KLSANet在Conv-4主干上欠拟合,无法完全展示其真实性能. 此外,与同样使用ResNet-12主干网络的LMPNet方法相比,本文提出的KLSANet在三个数据集的所有设置下都有着显著的性能提升.

4.4 消融实验

4.4.1 关键局部筛选模块的影响

本文提出的方法KLSANet利用关键局部筛选模块选出关键的查询样本局部来提高模型的分类性能,本节通过移除关键局部筛选模块或改变其阈值研究了该模块对模型准确率的影响从而验证其有效性,实验结果如表3所示. 其中w/和w/o分别表示有和没有关键局部筛选模块,而α是关键局部筛选模块中的超参数,与全局表示相似度低于α的候选查询局部将被滤除. 从表2可以看出,移除关键局部筛选模块后模型在CUB数据集5-way 5-shot设置上的准确率相比于设置合适阈值的最优结果有小幅下降,这表明关键局部筛选模块对提升模型的性能有积极作用. 此外,随着$\alpha\in\{0,0.2,0.4,0.6,0.8\}$增加,模型的准确率呈先上升后下降的趋势,当$\alpha=0.2$时准确率达到峰值,此时相比于没有关键局部筛选模块的模型提升了0.7%,这也证明了该模块的有效性. 而当α过大时,模型的准确率迅速下降. 其主要原因是过大的α使得大多数甚至全部的候选查询局部被过滤掉,从而导致模型的训练不稳定且难以收敛.

表3 CUB数据集上移除关键局部筛选模块及不同筛选阈值的影响

KLSM	α	5-way 5-shot
w/o	—	87.19±0.25
w/	0	87.59±0.25
w/	0.2	**87.89±0.43**
w/	0.4	87.05±0.25
w/	0.6	84.97±0.47
w/	0.8	57.80±0.35

4.4.2 特征表示类型的影响

为进一步验证KLSANet中局部特征度量方法的有效性,本节在CUB数据集上分别与基于局部描述符的方法DN4和在KLSANet上修改的基于全局表示的方法KLSANet*对比来研究特征表示类型的影响,实验结果

如表4所示. DN4是使用局部描述符的代表方法,且和KLSANet一样均使用了k近邻搜索来对齐相似局部,一个主要的区别是DN4依赖于局部描述符表示,而KLSANet则是基于局部特征表示. 从表4中可以观察到,提出算法KLSANet在Conv-4主干网络上的准确率大幅地超过了DN4,在CUB数据集的1-shot和5-shot设置上的领先分别达到了19.86%和8.71%. 该结果有力地证明了局部特征表示比局部描述符表示更有效. 此外,本节通过将随机裁剪局部尺寸扩大至与全局尺寸相同构建出一种基于全局表示的方法称为KLSANet*. 从表4的后两行可以看出,基于局部特征表示的KLSANet在性能上显著超过了基于全局表示的KLSANet*,在CUB数据集的1-shot和5-shot设置上的领先分别达到了6.69%和5.80%. 该结果有力地证明了局部特征表示比全局表示更有效.

表4　特征表示类型的影响

方法	主干网络	类型	CUB	
			5-way 1-shot	5-way 5-shot
DN4	Conv-4	局部描述符	46.84±0.81	74.92±0.64
KLSANet	Conv-4	局部特征	66.70±0.82	83.63±0.28
KLSANet*	ResNet-12	全局特征	68.25±0.54	83.12±0.49
KLSANet	ResNet-12	局部特征	**74.94±0.43**	**88.92±0.41**

4.4.3　k值的影响

在局部特征度量模块中,需要为查询样本的每个关键局部找到一个支持类中最相似的k个候选局部从而计算查询样本到类的相似度分数. k值的选择将会直接影响模型分类的准确率,因此如何选择合适的超参数k值是一个关键. 本节通过改变k值在三个数据集上的实验结果来研究k值对准确率的影响从而确定最佳的k值,实验结果如图2所示.

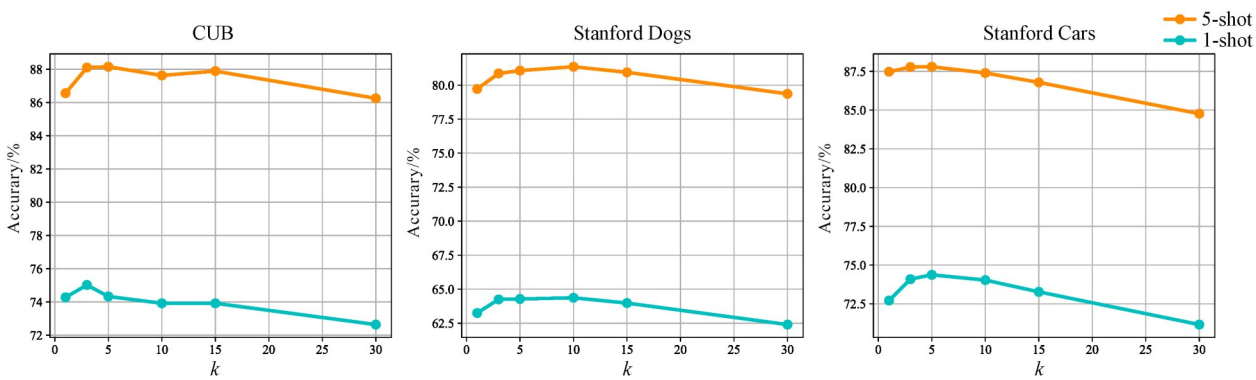


图2　三个数据集的1-shot和5-shot设置上不同k值的影响

从图2中观察到,随着$k \in \{1,3,5,10,15,30\}$的增加,模型在三个数据集的5-way 5-shot和5-way 1-shot设置上的准确率整都呈先上升后下降的趋势. 不同的是,在CUB数据集的5-shot和1-shot设置下的k值分别等于5和3时准确率达到峰值;而在Stanford Dogs数据集的5-shot和1-shot设置上大约k等于10时准确率最高;在Stanford Cars数据集的5-shot和1-shot设置上大约k等于5时准确率达峰值. 上述结果表明当选择的局部特征较少时会使模型无法区分一些类间差异小的类别,而当选择的局部过多时会使模型选中一些图像语义无关的块从而损害模型的性能.

4.5　可视化分析

给定一个查询样本的关键局部,根据与关键局部的相似度从支持类的所有候选局部中选出前k个最相似的局部来计算样本到类的相似度. 为了直观地展示提出算法的效果,本节用训练好的模型在三个数据集的测试集上实验,并且追踪了查询样本的关键局部以及对应支持类中选出的最相似的5个局部,部分结果如图3所示.

图3中展示了三个数据集上的可视化结果,对于每个数据集,从左到右各栏依次展示了查询样本(Query)、查询样本关键局部(Query parts)以及从对应支持类中选出的最相似的5个支持局部(Top5 support parts).

由图3可观察到,选出的支持局部与关键局部语义相近,这证明了提出算法正确地对齐了语义相近的局部. 例如,在CUB数据集的示例中,KLSANet可有效对齐鸟的头部、爪子、翅膀等局部的语义;在Stanford Dogs数据集中,可有效对齐狗的头部和颈部的局部语义. 此外,被选中的局部将对分类结果产生直接的贡献,同样证明了KLSANet具有较好的可解释性.

图3 关键查询局部及其对应Top-5相似度支持局部的可视化

5 结论

本文基于图像随机裁剪局部的特征研究了一种局部特征层面的度量准则，称为基于关键语义对齐的小样本分类算法(KLSANet)，其核心思想是通过k近邻搜索来对齐查询样本与支持类的局部语义从而实现一种图像到类的局部特征度量，并通过局部与图像语义的相似度筛选关键查询样本局部从而减轻图像语义无关局部的影响. 在三个基准数据集上定量和定性的实验结果均表明，本文算法显著优于许多先进的小样本学习算法并展示出良好的可解释性，均表明提出的KLSANet算法中局部特征度量机制的优越性. 此外，消融实验结果也证明了提出的关键局部筛选模块的有效性.

尽管KLSANet方法在小样本分类任务上展现出较好的性能，但仍存在一些局限性，例如需要手动调整模型超参数，且当前仅考虑了单一类型的特征表示. 因此，未来工作将聚焦于研究任务自适应的超参数学习方法以及多尺度特征表示的小样本度量方法.

参考文献

[1] WANG Y Q, YAO Q M, KWOK J T, et al. Generalizing from a few examples: A survey on few-shot learning[J]. ACM Computing Surveys, 2021, 53(3): 1-34.

[2] SONG Y S, WANG T, CAI P Y, et al. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities[J]. ACM Computing Surveys, 2023, 55(13s): 1-40.

[3] LU J, GONG P H, YE J P, et al. A survey on machine learning from few samples[J]. Pattern Recognition, 2023, 139: 109480.

[4] 葛轶洲, 刘恒, 王言, 等. 小样本困境下的深度学习图像识别综述[J]. 软件学报, 2022, 33(1): 193-210.
GE Y Z, LIU H, WANG Y, et al. Survey on deep learning image recognition in dilemma of small samples[J]. Journal of Software, 2022, 33(1): 193-210. (in Chinese)

[5] 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. New York: ACM, 2017: 1126-1135.

[6] RAVI S, LAROCHELLE H. Optimization as a model for few-shot learning[C]//International Conference on Learning Representations. Vancouver: Openreview.net, 2017: 1-11.

[7] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning[J]. Advances in Neural Information Processing Systems, 2016: 3637-3645.

[8] CHEN Y B, LIU Z, XU H J, et al. Meta-baseline: Exploring simple meta-learning for few-shot learning[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2021: 9042-9051.

[9] CAO K, BRBIC M, LESKOVEC J. Concept learners for few-shot learning[C]//International Conference on Learning Representations. Vancouver: Openreview.net, 2021: 1-17.

[10] 赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述[J]. 软件学报, 2021, 32(2): 349-369.

ZHAO K L, JIN X L, WANG Y Z. Survey on few-shot learning[J]. Journal of Software, 2021, 32(2): 349-369. (in Chinese)

[11] RUSU A A, RAO D, SYGNOWSKI J, et al. Meta-learning with latent embedding optimization[EB/OL]. (2019-03-26)[2024-03-02]. https://arxiv.org/abs/1807.05960.

[12] LI Z G, ZHOU F W, CHEN F, et al. Meta-sgd: Learning to learn quickly for few-shot learning[EB/OL]. (2017-09-28)[2024-03-02]. https://arxiv.org/abs/1707.09835.

[13] GRANT E, FINN C, LEVINE S, et al. Recasting gradient-based meta-learning as hierarchical bayes[EB/OL]. (2018-01-26)[2024-03-02]. https://arxiv.org/abs/1801.08930.

[14] OH J, YOO H, KIM C H, et al. Boil: Towards representation change for few-shot learning[EB/OL]. (2021-03-03)[2024-03-02]. https://arxiv.org/abs/2008.08882.

[15] LAI N, KAN M N, HAN C R, et al. Learning to learn adaptive classifier-predictor for few-shot learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(8): 3458-3470.

[16] SANTORO A, BARTUNOV S, BOTVINICK M, et al. Meta-learning with memory-augmented neural networks [C]//Proceedings of The 33rd International Conference on Machine Learning. New York: ACM, 2016: 1842-1850.

[17] MUNKHDALAI T, YU H. Meta networks[C]//Proceedings of the 34th International Conference on Machine Learning. New York: ACM, 2017: 2554-2563.

[18] RAMALHO T, GARNELO M. Adaptive posterior learning: Few-shot learning with a surprise-based memory module[EB/OL]. (2019-02-07)[2024-03-02]. https://arxiv.org/abs/1902.02527.

[19] PARNAMI A, LEE M. Learning from few examples: A summary of approaches to few-shot learning[EB/OL]. (2022-03-07) [2024-03-02]. https://arxiv. org/abs/2203.04291.

[20] WANG W J, DUAN L J, WANG Y X, et al. Remember the difference: Cross-domain few-shot semantic segmentation via meta-memory transfer[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 7055-7064.

[21] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning[C]. Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2017: 4080-4090.

[22] GARCIA V, BRUNA J. Few-shot learning with graph neural networks[C]//International Conference on Learning Representations. Vancouver: Openreview.net, 2018.

[23] TSENG H Y, LEE H Y, Huang J B, et al. Cross-domain few-shot classification via learned feature-wise transformation[C]//International Conference on Learning Representations. Vancouver: Openreview.net, 2020.

[24] XIE J T, LONG F, LV J M, et al. Joint distribution matters: Deep Brownian distance covariance for few-shot classification[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 7962-7971.

[25] LI W H, LIU X L, BILEN H. Cross-domain few-shot learning with task-specific adapters[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 7151-7160.

[26] 李维刚, 甘平, 谢璐, 等. 基于样本对元学习的小样本图像分类方法[J]. 电子学报, 2022, 50(2): 295-304.
LI W G, GAN P, XIE L, et al. A few-shot image classification method by pairwise-based meta learning[J]. Acta Electronica Sinica, 2022, 50(2): 295-304. (in Chinese)

[27] LI Y H, LI H X, CHEN H X, et al. Hierarchical representation based query-specific prototypical network for few-shot image classification[EB/OL]. (2021-03-21)[2024-03-02]. http://arxiv.org/abs/2103.11384.

[28] LI W B, WANG L, XU J L, et al. Revisiting local descriptor based image-to-class measure for few-shot learning [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 7253-7260.

[29] SUNG F, YANG Y X, ZHANG L, et al. Learning to compare: Relation network for few-shot learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 1199-1208.

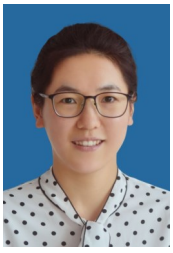
[30] CHENG J, HAO F S, LIU L, et al. Imposing semantic consistency of local descriptors for few-shot learning[J]. IEEE Transactions on Image Processing, 2022, 31: 1587-1600.

[31] HAO F S, HE F X, CHENG J, et al. Collect and select: Semantic alignment metric learning for few-shot learning [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 8459-8468.

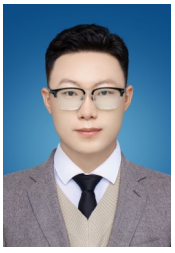
[32] DONG C Q, LI W B, HUO J, et al. Learning task-aware local representations for few-shot learning[C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. California: Morgan Kaufmann, 2020: 716-722.

[33] LI W B, XU J L, HUO J, et al. Distribution consistency based covariance metric networks for few-shot learning [C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York: ACM, 2019, 33(1): 8642-8649.
[34] LI H, YANG L, GAO F. More attentional local descriptors for few-shot learning[C]//FARKAŠ I, MASULLI P, WERMTER S. International Conference on Artificial Neural Networks. Cham: Springer, 2020: 419-430.
[35] QI Y, SUN H, LIU N Z, et al. A task-aware dual similarity network for fine-grained few-shot learning[C]//Pacific Rim International Conference on Artificial Intelligence. Cham: Springer, 2022: 606-618.
[36] HUANG H W, WU Z K, LI W B, et al. Local descriptor-based multi-prototype network for few-shot Learning[J]. Pattern Recognition, 2021, 116: 107935.
[37] ZHENG Z J, FENG X, YU H Q, et al. BDLA: Bi-directional local alignment for few-shot learning[J]. Applied Intelligence, 2023, 53(1): 769-785.
[38] ABDELAZIZ M, ZHANG Z P. Learn to aggregate global and local representations for few-shot learning[J]. Multimedia Tools and Applications, 2023, 82(21): 32991-33014.
[39] 齐妍, 孙涵. 基于判别性特征增强的小样本细粒度图像识别[J]. 计算机技术与发展, 2024, 34(1): 44-51.
QI Y, SUN H. Few-shot fine-grained image recognition based on discriminative feature enhancement[J]. Computer Technology and Development, 2024, 34(1): 44-51. (in Chinese)
[40] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//Proceedings of the 37th International Conference on Machine Learning. New York: ACM, 2020: 1597-1607.
[41] HE K M, FAN H Q, WU Y X, et al. Momentum contrast for unsupervised visual representation learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 9726-9735.
[42] ORESHKIN B, RODRÍGUEZ LÓPEZ P, LACOSTE A. TADAM: Task dependent adaptive metric for improved few-shot learning[C]. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2018: 719- 729.
[43] WAH C, BRANSON S, WELINDER P, et al. The caltech-UCSD birds-200-2011 dataset[DB/OL]. (2021) [2024-03-02]. https://authors. library. caltech. edu/records/cvm3y-5hh21.
[44] KHOSLA A, JAYADEVAPRAKASH N, YAO B P, et al. Novel dataset for fine-grained image categorization: Stanford dogs[C]//Proceedings of CVPR Workshop on Fine-Grained Visual Categorization (FGVC). Piscataway: IEEE, 2011, 2(1): 1-2.
[45] KRAUSE J, STARK M, JIA D, et al. 3D object representations for fine-grained categorization[C]//2013 IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE, 2013: 554-561.
[46] HILLIARD N, PHILLIPS L, HOWLAND S, et al. Few-shot learning with metric-agnostic conditional embeddings[EB/OL]. (2018-02-12) [2024-03-02]. https://arxiv.org/abs/1802.04376.

作者简介



孙 哲 女, 1989年12月出生于河北省邯郸市. 现为燕山大学信息科学与工程学院副教授、博士生/硕士生导师. 分别于2013年和2018年在燕山大学获得学士和博士学位. 2017年9月—2018年3月, 于澳大利亚纽卡斯尔大学博士联合培养. 主要研究方向包括多模态信号处理、小样本学习、视频理解, 在国内外发表学术论文40余篇.
E-mail: sunzhe_ysu@163.com



郑 旺 男, 1996年2月出生于新疆维吾尔自治区巴音郭楞蒙古自治州. 现为燕山大学信息科学与工程学院硕士. 主要研究方向为小样本图像分类.
E-mail: 18196480916@163.com

郭朋飞 男, 2000年12月出生于河南省南阳市. 现为燕山大学信息科学与工程学院硕士. 主要研究方向为小样本图像分类.
E-mail: gpf_xx@163.com