

面向图像分类的小样本学习算法综述

彭云聪, 秦小林, 张力戈, 顾勇翔

引用本文

彭云聪, 秦小林, 张力戈, 顾勇翔. [面向图像分类的小样本学习算法综述](#)[J]. 计算机科学, 2022, 49(5): 1-9.

PENG Yun-cong, QIN Xiao-lin, ZHANG Li-ge, GU Yong-xiang. [Survey on Few-shot Learning Algorithms for Image Classification](#)[J]. Computer Science, 2022, 49(5): 1-9.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多分支注意力增强的细粒度图像分类](#)

Fine-grained Image Classification Based on Multi-branch Attention-augmentation

计算机科学, 2022, 49(5): 105-112. <https://doi.org/10.11896/jsjcx.210100108>

[基于二分图的个性化学习任务分配](#)

Personalized Learning Task Assignment Based on Bipartite Graph

计算机科学, 2022, 49(4): 269-281. <https://doi.org/10.11896/jsjcx.210500125>

[基于图像分块与特征融合的户外图像天气识别](#)

Outdoor Image Weather Recognition Based on Image Blocks and Feature Fusion

计算机科学, 2022, 49(3): 197-203. <https://doi.org/10.11896/jsjcx.201200263>

[跨领域文本的可迁移情绪分析方法](#)

Transferable Emotion Analysis Method for Cross-domain Text

计算机科学, 2022, 49(3): 218-224. <https://doi.org/10.11896/jsjcx.210400034>

[基于混合样本自动数据增强技术的半监督学习方法](#)

Semi-supervised Learning Method Based on Automated Mixed Sample Data Augmentation Techniques

计算机科学, 2022, 49(3): 288-293. <https://doi.org/10.11896/jsjcx.210100156>

面向图像分类的小样本学习算法综述

彭云聪^{1,3} 秦小林^{1,2,3} 张力戈^{1,3} 顾勇翔^{1,3}

1 中国科学院成都计算机应用研究所 成都 610041

2 南昌理工学院 南昌 330044

3 中国科学院大学计算机科学与技术学院 北京 100049

(pengyuncong19@mails.ucas.ac.cn)

摘要 目前,以深度学习为代表的人工智能算法凭借超大规模数据集以及强大的计算资源,在图像分类、生物特征识别、医疗辅助诊断等领域取得了优秀的成果并成功落地。然而,在许多实际的应用场景中,因诸多限制,研究人员无法获取到大量样本或者获取样本的代价过高,因此研究图像分类任务在小样本情形下的学习算法成为了推动智能化进程的核心动力,同时也成为了当下的研究热点。小样本学习指在监督信息数量有限的情况下进行学习并解决问题的算法。首先,从机器学习理论的角度描述了小样本学习困难的原因;其次,根据小样本学习算法的设计动机将现有算法归为表征学习、数据扩充、学习策略三大类,并分析其优缺点;然后,总结了常用的小样本学习评价方法以及现有模型在公用数据集上的表现;最后,讨论了小样本图像分类技术的难点及未来的研究趋势,为今后的研究提供参考。

关键词: 小样本学习;图像分类;表征学习;数据扩充;迁移学习

中图法分类号 TP181

Survey on Few-shot Learning Algorithms for Image Classification

PENG Yun-cong^{1,3}, QIN Xiao-lin^{1,2,3}, ZHANG Li-ge^{1,3} and GU Yong-xiang^{1,3}

1 Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu 610041, China

2 Nanchang Institute of Technology, Nanchang 330044, China

3 School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract Presently, artificial intelligence algorithms represented by deep learning have achieved advanced results and been successfully used in fields such as image classification, biometric recognition and medical assisted diagnosis by virtue of ultra-large-scale data sets and powerful computing resources. However, due to many restrictions in the actual environment, it is impossible to obtain a large number of samples or the cost of obtaining samples is too high. Therefore, studying the learning algorithm in the case of small samples is the core driving force to promote the intelligent process, and it has also become a current research hotspot. Few-shot learning is the algorithm to learn and solve the problem under the condition of limited supervision information. Firstly, it describes the reasons why few-shot learning is difficult to generalize from the perspective of machine learning theory. Secondly, according to the design motivation of the few-shot learning algorithm, existing algorithms are classified into three categories: representation learning, data expansion and learning strategy, and their advantages and disadvantages are analyzed. Thirdly, we summarize the commonly used few-shot learning evaluation methods and the performance of existing models in public data sets. Finally, we discuss the difficulties and future research trends of small sample image classification technology to provide references for future research.

Keywords Few-shot learning, Image classification, Learning representation, Data expansion, Transfer learning

1 引言

近年来,人工智能技术在大数据时代迎来了高速发展,从

早期的学术探索迅速转变为实际应用。目前,以深度学习为代表的人工智能算法凭借超大规模数据集以及强大的计算资源,在图像分类、生物特征识别、医疗辅助诊断等领域取得了

到稿日期:2021-05-18 返修日期:2021-10-22

基金项目:国家自然科学基金(61402537);四川省科技计划资助项目(2019ZDZX0005,2019ZDZX0006,2020YFQ0056,2021YFG0034);四川省委组织部人才专项资助;全国科学院联盟合作项目(中国科学院成都分院-重庆科学技术研究院)

This work was supported by the National Natural Science Foundation of China (61402537), Sichuan Science and Technology Program (2019ZDZX0005,2019ZDZX0006,2020YFQ0056,2021YFG0034), Talents by Sichuan Provincial Party Committee Organization Department and National Academy of Science Alliance Collaborative Program(Chengdu Branch of Chinese Academy of Sciences-Chongqing Academy of Science and Technology).

通信作者:秦小林(qinxl2001@126.com)

优秀的成果并成功落地。

然而,当今现实场景中通常并不具备获得大规模可训练数据的条件,这不利于许多传统行业的智能化转型。另一方面,由于图像分类算法在实际应用中起着关键性的作用,因此面向图像分类的小样本学习的关键算法研究成为了产业智能化转型的驱动引擎之一。

深度学习以大规模数据集为前提,在图像分类、目标检测、文本分析^[1-3]等领域取得了显著的成功。然而在实际场景中,首先,由于成本、隐私、安全或道德问题,相关研究者很难或不可能获得大规模、高质量的数据及标注。例如,在医疗领域,医学影像的产生来源于病例,但少量的病例并不能够辅助机器对医疗影像进行分析。其次,在算法设计层面,研究者期望机器学习以人类的方式进行学习,即在获取少量样本的情况下,对样本进行分类和识别,并且具有快速理解新概念并将其泛化的能力。

为了能够在监督信息数量有限的情况下进行学习,针对小样本学习(Few-shot Learning)^[4-8]的研究应运而生。在小样本分类中,模型在一组具有丰富样本的类别集上进行训练,这些类称为基类,然后在类别不交叉的仅具有少量样本的另一组类别集(新类)上进行训练与测试。

目前,针对小样本学习的研究工作越来越丰富,随着深度学习的发展,涌现了很多新颖的小样本学习方法^[9-11]。例如,在模型表征阶段采用自监督学习以更好地表征出图像^[12-15];在数据扩充阶段,采用从原始域扩充或从语义空间扩充等^[11,16-17]方式来处理小样本学习任务;在学习阶段,使用迁移学习、度量学习、元学习等算法^[18-24]以更好地寻找到一个有良好泛化能力的模型。现有的小样本综述文献^[25-26]通常从算法类别的角度进行归纳总结,而本文将从理论误差分析以及算法设计动机的角度来进行综述,并覆盖近年来取得的成果。

本文首先从机器学习理论的角度描述了小样本学习难以泛化的原因;其次,依据小样本学习算法的设计动机将现有算法归为表征学习、数据扩充、学习策略三大类,并评价其优缺点;然后,总结了常用的小样本学习评价方法以及现有模型在公用数据集上的表现;最后,提出了一些有前景的研究方向,为今后的研究提供参考。

2 小样本学习核心问题

2.1 预备知识

假设一个待学习问题,它存在一个从 $\mathcal{X}$ 到 $\mathcal{Y}$ 的最优映射 h^* 。

记 $\mathcal{D}$ 是 $\mathcal{X} \times \mathcal{Y}$ 上的联合分布,存在 $\mathcal{D}$ 的独立同分布的观测采样,即训练集 D_{train} 。假设空间 $\mathcal{H}$ 指从 $\mathcal{X}$ 到 $\mathcal{Y}$ 的某一类模型(映射)的集合。

在给定假设空间 $\mathcal{H}$ 的情况下,记 $h \in \mathcal{H}$ 是从 $\mathcal{X}$ 到 $\mathcal{Y}$ 的一个映射,其期望风险为:

$$R(h; \mathcal{D}) = \mathbb{E}_{(x,y) \sim \mathcal{D}} (\ell(h(x), y)) \quad (1)$$

在 D 上的经验风险为:

$$\hat{R}_m(h; D_{\text{train}}) = \frac{1}{m} \sum_{i=1}^m \ell(h(x_i), y_i) \quad (2)$$

2.2 误差分解

由于最优映射通常并不存在于给定的假设空间中,对于

误差分析通常可做如下的误差分解^[27-29]:

$$\text{假设 } h_m = \arg \min_{h \in \mathcal{H}} \hat{R}_m(h),$$

$$h' = \arg \min_{h \in \mathcal{H}} R(h), \text{ 则有:}$$

$$\mathbb{E}[R(h_m) - R(h^*)] = \mathbb{E}[R(h') - R(h^*)] + \mathbb{E}[R(h_m) - R(h')] \quad (3)$$

因此,按照上述误差分解可以将其分类、归纳为:

(1)逼近误差,指在给定假设空间后,衡量其中映射与最优映射 h^* 的接近程度。

(2)泛化误差,指在给定数据集和假设空间时,最小化经验风险与最小化期望风险之间的差距。

如图 1 所示,本文对误差分解进行了可视化。

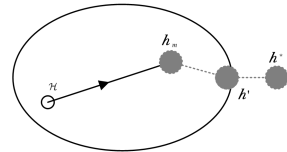


图 1 误差分解示意图

Fig. 1 Error decomposition

2.3 小样本问题的挑战

对于逼近误差而言,如果能尽可能使用表征能力更强的假设空间 $\mathcal{H}$,使得其足够复杂从而包含最优映射 h^* ,那么就可以减小逼近误差,但同时复杂的假设空间会加大泛化的难度,一定程度增大了泛化误差。

对于泛化误差而言,一般可以使用通用的泛化界进行分析,对于任意 $0 < \delta < 1$,以至少 $1 - \delta$ 的概率有:

$$R(h; \mathcal{D}) \leq \hat{R}_m(h; D) + \sqrt{\frac{\text{Capacity Measure}}{m}} \quad (4)$$

其中, Capacity Measure 代表模型的复杂度估计,通常可采用 VC 维和 Rademacher 等。通过分析可知,如果样本量太少,则很难通过经验风险逼近期望风险,这是小样本学习中最核心的问题,即经验风险最小化不一定能够泛化。

2.4 小样本问题处理动机

2.4.1 表征学习

由对逼近误差的分析可知,如果仅通过更换表征能力更强的假设空间,通常不能使总误差变小,因为这样会增大泛化误差。

小样本学习通常利用先验知识来弥补监督信息的不足。在合理的表征学习基础上,能够使用更简单的假设空间也能包含最优映射,同时由于假设空间复杂度降低,对样本量的需求减少,因此也更容易泛化。表征学习示意图如图 2 所示。

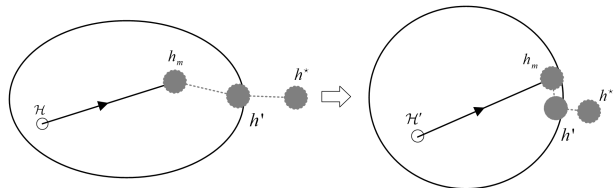


图 2 表征学习示意图

Fig. 2 Representation learning

2.4.2 数据扩充

理小样本问题最朴素的想法是通过生成样本的方式进行数据扩充,如图 3 所示。

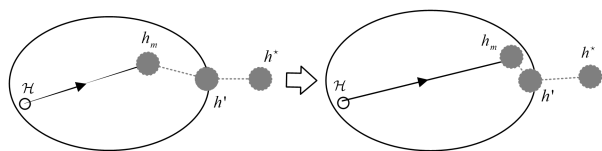


图3 数据扩充示意图

Fig.3 Data expansion

数据扩充能够缩减泛化误差上界,经验风险可以作为期望风险的良好近似值,使得最小化经验风险更加可靠。然而,如图像加噪、旋转等简单变换生成的样本并非有效样本,对于缩减泛化误差上界的帮助非常有限。

2.4.3 学习策略

如果经验风险可以作为期望风险的良好近似值,则需要考虑如何通过学习策略来优化模型至 h' 。

一方面可以使用先验知识来得到一个好的初始点,另一方面可以基于先验知识对搜索方向进行一定程度的修正。

3 小样本学习算法

小样本学习作为一个活跃的研究课题得到了广泛的研究。本节按照第2节归纳的小样本学习算法设计动机的分类,即图像表征阶段、数据扩充阶段以及学习阶段这3部分来介绍相关主题的现有研究。小样本学习算法归类示意图如图4所示。

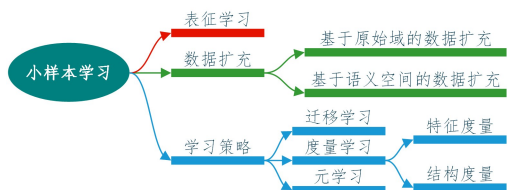


图4 小样本学习算法归类示意图

Fig.4 Categories of few-shot learning algorithms

3.1 基于表征学习的小样本算法

表征学习的动机在于将原始数据进行表征学习,从而变换到特征域进行学习。特征域具有更低的维度以及根据先验知识得到的语义信息,因此能够大幅降低学习难度。

对于表征学习,最朴素的想法是通过大量基类学习出一个特征提取器,使其能适应有限的基类、新类域差异,然后通过分类器进行识别^[30-32]。这种预训练微调的方式虽然直观、简洁,但是很难学出通用特征,在小样本学习问题上难以取得很好的效果。

近年来,随着自监督技术的不断发展,表征学习有了较大的进步,进而推动了小样本学习的发展^[33-34]。自监督学习(SSL)旨在从数据自身学习鲁棒的表示而不需要类标签,只使用数据的固有信息和结构。其核心问题是如何构造足够复杂的 pretext task^[12-15,35],从大规模的无监督数据中挖掘样本的表征信息,进而基于高级紧凑语义视觉表示来完成下游任务。

在小样本学习中,表征阶段非常重要,因此,如何构造合理的自监督学习框架进行表征学习是一个研究热点。Hjelm等^[36]提出了Deep InfoMAX (DIM),该方法与传统自监督学习构造的旋转、复原等伪任务不同,其直接基于互信息进行

学习。此方法在几个分类任务上的表现皆优于许多流行的无监督学习方法,并可与有监督学习的效果相比拟。

DIM 依据以下标准训练编码器:

(1)互信息最大化,即找到模型参数,使得原信息与表征信息的互信息最大化;

(2)统计约束,即使得编码器的输出满足某些约束,例如将表征向量约束接近于一个已知先验分布。

Bachman等^[37]在 DIM 的基础上加入了 multi-scale 机制,进一步提出了 AMDIM 模型,即不同尺度上的 feature 也可以组成正负样本对。

Chen等^[38]基于 AMDIM 进行小样本学习,并取得了良好的效果。基于 AMDIM 训练出的模型仅使用中心点学习或最近邻分类等简单算法,就可在小样本学习中取得传统小样本算法难以企及的效果。

另一方面,由于小样本学习需要鲁棒表征^[39],因此在表征学习中使用一定的正则化技术也能提升小样本学习的性能。Mangla等^[40]提出了 S2M2 方法,该方法在表征学习的过程中同时使用基于旋转的自监督技术以及基于流形混合的正则化技术。通过自监督技术学习到丰富特征流形的正则化,显著提升了小样本的学习性能。其中,流形混合指在表征后的高维语义向量中进行 Mixup^[41-42]操作,即:

$$L_{mm} = \mathbb{E}_{(x,y) \in \mathcal{D}_b} [L(Mix_{\lambda}(f_{\theta}^i(x), f_{\theta}^i(x')), Mix_{\lambda}(y, y'))] \quad (5)$$

其中, $f_{\theta}^i(\cdot)$ 代表表征函数, $Mix_{\lambda}(\cdot, \cdot)$ 代表进行 Mixup 操作。

3.2 基于数据扩充的小样本算法

数据扩充的动机在于生成非平凡、尽可能多样的样本,从而增大样本量,使得泛化误差上界减小,经验风险更加可靠。数据扩充可以分为从原始域扩充和从语义空间扩充。原始域扩充指在原始数据的基础上生成样本;语义空间扩充指在表征向量的基础上生成样本。

3.2.1 原始域的数据扩充

Wang等^[43]于2018年提出了一种通用的基于生成网络的小样本生成算法。该算法的动机在于应将生成的对学习分类器有用的样本作为目标,这个目标不同于传统的图像重建。因此,该方法通过对模型中的分类目标进行端到端的优化,联合训练生成器参数和分类器参数,通过直接学习产生对分类有用的样本,具体网络结构如图5所示。

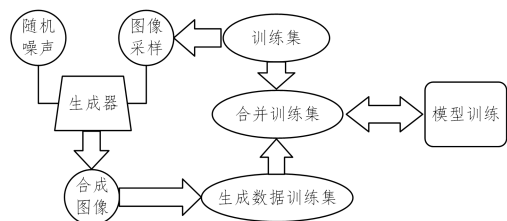


图5 生成网络结构示意图

Fig.5 Generate network structure

Azuri等^[44]于2020年提出了生成潜隐条件优化算法(GLICO),通过类内任意两个样本来表征向量的超球面插值,并将其还原为图像来生成新样本。其中超球面插值的生成方式如图6所示。

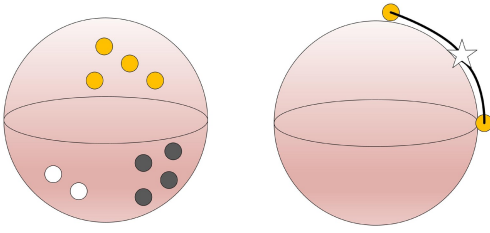


图6 GLICO 样本生成方法示意图

Fig. 6 GLICO sample generation method

该算法的本质类似于针对表征向量的 SMOTE 过采样算法,其采出的样本点全部处于已有样本点所围成的凸区域,因此其具有高置信度保证,即采出的样本点属于当前类别。

然而,由于生成的样本处于原始数据所围成的凸区域,在复杂情况下,这种生成的样本难以对分界面产生决定性的作用。

3.2.2 语义空间的数据扩充

考虑到图像重建的困难程度,因此直接从语义空间生成样本也是一种不错的选择。

Schwartz 等^[45]于 2018 年提出了 Delta-Encoder,该方法通过偏移量学习来生成样本,其生成方式如图 7 所示。这种采样方式可以轻松采样出原始数据所围成的凸区域以外的样本,这些样本对分界面将产生较大的改善作用。然而,由于其仅考虑任意两个样本组成的样本对,而忽略了类内、类间流形,因此难以具备高置信度。

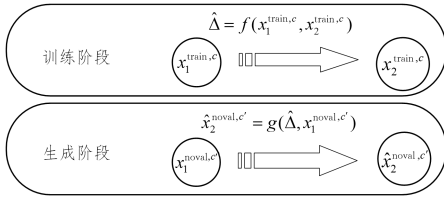


图7 Delta-Encoder 示意图

Fig. 7 Diagram of Delta-Encoder

Chen 等^[17]在 Delta-Encoder 的基础上进一步提出了一种新的生成框架,即多样性迁移网络(Diversity Transfer Network)。它学习迁移已知类别中潜在的多样性,并将其与支持集的特征相结合,从而在特征空间域中生成样本。生成方式如图 8 所示。

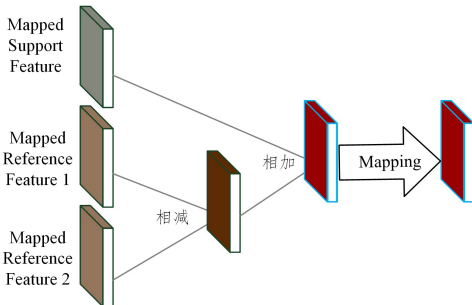


图8 多样性样本生成示意图

Fig. 8 Sample generation method for diversity transfer network

这种生成方式类似于 Delta-Encoder 的偏移量学习,区别在于多样性迁移的学习是通过最小化元分类损失来进行

学习,而非基于特征重构损失进行学习。

3.3 学习策略

3.3.1 迁移学习

利用迁移学习^[46-47]可以减小模型训练的代价,同时达到让神经网络适应小样本数据的目的。迁移学习的思想是,相似任务之间的学习存在一定共通性。其本质是基于源域数据和目标域数据之间共同的特征、关系以及模型参数共享等方式进行迁移学习,并将模型用于新类数据分类。

如果存在足够相似的任务,可以将从相关任务中学到的预训练模型的 θ_0 作为一个良好的初始化。在大规模数据的预训练情况下, θ_0 能够通过固定特征提取器、微调分类器等方式,很快地适应当前任务。如果预先训练的 θ_0 不足以完全适应任务,可以引入一个附加参数 δ ,将模型参数扩展为 $\theta = \{\theta, \delta\}$,再进行学习。

具体的迁移方式可以分为基于特征的迁移、基于模型的迁移和基于关系的迁移。

Chen 等^[32]提出了一个详细的实证研究,结果表明,在基类和新类之间存在有限域差异的情况下,减小类内差异会显著提升小样本学习算法的性能,同时,通过使用较深的主干网络或者将线性分类器替换为余弦距离分类器,可以减小类内差异。Chen 等提出的具有标准微调的基线方法 Baseline++ 相比传统小样本学习算法仍有竞争力。

Hu 等^[48]提出了基于迁移特征分布的小样本学习方式,该方法基于两个步骤:首先通过幂律变换预处理特征向量,使它们更接近高斯型分布;然后使用最优传输算法将特征的分布与所考虑的类的分布对齐。

求解类中心及样本分类问题可描述为:

$$\{\hat{l}(f_i)\}, \{\hat{c}_j\} = \arg \min_{\{l(f_i)\} \in C, \{c_j\}} \sum_i (f_i - c_{l(f_i)})^2 \quad (6)$$

其中, $l(f_i)$ 代表第 i 个样本特征向量的标签, c_j 代表第 j 类样本的类中心。考虑到通过最大后验概率估计这些中心类似于最小化 Wasserstein 距离,因此,受最优传输问题(Optimal Transport Problem)的启发,采用 Sinkhorn mapping 算法来解决该问题。此方法可以在各种特征提取器和数据集上带来精度的大幅提高,2020 年期间在多个数据集上保持了最先进的成果。

3.3.2 度量学习

度量学习是一种非常直观的解决小样本问题的方法。具体可以分为基于特征的度量学习和基于结构的度量学习。

(1) 特征度量

基于特征的度量学习一般是将图像映射到一个高维语义空间,然后基于语义空间高维向量的某种度量方式来计算样本间的差异,从而进行分类任务。

最经典的度量学习是 Koch^[31]早期提出的深度卷积孪生网络。该算法让一对样本基于相同的特征提取网络进行特征提取,并基于欧氏距离进行相似性度量,进而完成分类。虽然这种方法简单、直观,但在复杂情况下难以达到理想的效果。

Vinyals 等^[49]提出了匹配网络(Matching Networks),它在特征度量学习的基础上加入了外部记忆增强,将一个带

标签的支持集和一个未标记的样本映射到其标签上,从而无须进行微调以适应新的类别。

Sung 等^[50]于 2018 年提出了端到端的相关网络(Relation Network,RN),其核心在于学习一个深度距离度量,用于度量样本的相似性。此算法首先在表征阶段将图像表征为高维特征向量,然后在相关阶段将待查询样本与支持集样本的高维特征向量进行拼接,最后根据回归的方式求出相关值,进而完成分类任务。

(2)结构度量

结构度量指通过探索样本间的结构关系来完成分类任务,常见的结构度量算法有图神经网络及其衍生算法。

图神经网络(GNN)方法在小样本学习中的核心是通过主要特征向量进行度量学习、结构学习等,从而依据节点分类或者边分类的框架进行直推式推理。由于图神经网络模型的节点和边可以表示更多的图像信息,因此图神经网络对样本间的复杂关系有更强的表示能力,这也有助于探索更多小样本数据集中潜在的信息。

Garcia 等^[51]基于图模型消息传递机制来研究小样本问题,并且可以扩展至半监督等问题。其核心是在完成图像表征后,通过图结构的消息传递与邻接矩阵更新进行节点分类,根据表征的特征向量与自行定义或者训练所得的度量函数 φ 来构建邻接矩阵 $\mathbf{A}$ 。

$$\tilde{\mathbf{A}}_{i,j}^{(k)} = \varphi_{\theta}(x_i^{(k)}, x_j^{(k)})$$

(7)

其总体流程为:首先进行图像表征,根据表征向量得到初始的邻接矩阵后,进行图卷积操作;然后更新邻接矩阵,再进行图卷积,重复上述操作数次;最后根据每个节点的特征向量进行分类。

2019 年, Kim 等^[52]提出了一种新的 Edge-Labeling 图神经网络,将深度神经网络应用于 Edge-Labeling 图,用于小样本学习。与传统的基于图网络的节点分类不同,此算法通过迭代地更新边的标签信息来刻画类内的相似性和类间的不相似性,从而完成对样本的聚类。

3.3.3 元学习

元学习通过初始值学习、多阶段优化等方式来实现学习

的目的。目前元学习已经在多个领域取得优异的效果,例如针对图像分类的元伪标签算法已经达到了当前最好的效果^[53];针对标签噪声的元学习算法^[54]也大幅改善了带噪声模型的性能。

在小样本问题中,以任务为单位进行多次学习的算法都可以宽泛地被称为元学习算法,因为它代表着从多次任务中学习到任务级别的经验并用于小样本分类,这也是学会学习的过程。但目前许多元学习算法在实验上并没有取得特别理想的效果,有待我们更深入地研究元学习算法,并进一步构造出适合小样本学习的元学习框架。

Finn 等^[55]于 2017 年提出了与模型无关的元学习算法 MAML。此算法与任何用梯度下降训练的模型兼容,并适用于各种不同的学习问题,其主要学习的对象是模型初始化参数,并通过学习这种超参数来控制学习过程。具体而言, MAML 不在意初始化参数在多个相似任务上当前的表现,而在意由其训练出的最终模型参数的表现。Nichol 等^[56]在 MAML 的基础上进行改进,提出了 Reptile 算法。参数优化时, Reptile 仍然考虑学习一个良好的参数初始化,由于 Reptile 在每个任务学习中使用随机梯度下降而不要求使用微分,因此,相比 MAML, Reptile 所需的计算量和内存都更少。

Ravi^[57]提出了一种新的思路,即直接学习一个可以直接输出梯度更新的优化器。其核心在于借助 LSTM 序列特性,根据损失梯度和训练数据直接学习出模型参数的更新方向,即:

$$\phi_i = \theta - \alpha f(\theta, \mathcal{D}_i^r, \nabla_{\theta} \mathcal{L})$$

(8)

此外,元学习具有良好的结合性,常与其他算法相结合。Sun 等^[58]将迁移学习和元学习相结合,提出了元迁移学习,其利用迁移学习减少模型需要更新的参数量,同时可以使用更深的网络结构提升小样本学习性能。

3.4 算法总结与分析

上文依据小样本学习算法设计动机的不同,将现有算法归为表征学习、数据扩充、学习策略三大类别,下文将对这 3 种方法进行总结与分析,并且详细比较它们的优缺点,如表 1 所列。

表 1 各类算法优缺点对比

Table 1 Comparison of advantages and disadvantages of various algorithms

算法大类	算法小类	优点	缺点
表征学习	自监督学习	对图像有更好的表征,小样本情境下的学习性能有明显提升	在小样本场景中,如何构建高效的自监督伪任务仍然具有挑战性
	正则化	对数据分布的微小变化具有鲁棒性	并非所有的正则化方法都能提升性能,因此需要根据具体任务进行选择
数据扩充	原始域扩充	直观	大多数相关模型训练困难,并且更侧重于真实图像的重建,而非重建出具有分类意义的图片,因此难以对分类效果产生较大的积极影响
	语义域扩充	相比原始域扩充训练更容易,并且易生成对分类效果产生积极影响的样本	在高度抽象的语义空间中,生成样本的策略有待进一步探究
学习策略	迁移学习	性能较好,易于实现	需要存在与目标域相近的源域,否则可能导致负迁移
	度量学习	计算便捷,适合直推推理	极度依赖表征模型的性能
	元学习	可以通过任务级别的学习结合多层优化,学习更深层次的知识	复杂度较高,当前性能不够理想,有待进一步改进

4 小样本学习数据集及实验对比

4.1 小样本学习常用数据集

对于小样本分类问题,用于评估算法性能的数据集主要有 Mini-ImageNet,CUB-200,CIFAR100 等。

Mini-ImageNet 数据集由来自 ImageNet 数据集的 100 个类的子集组成,每个类包含 600 幅图像,每张图像的尺

寸为 84×84 像素。

CUB 数据集包含鸟类的 200 个子类,共计 11 788 幅图像,图像分辨率较高,同时也是细粒度的视觉分类任务中最广泛使用的数据集。

CIFAR100 数据集包含 100 个类别,每类包含 600 幅图像,每张图像的尺寸为 32×32 像素。

上述各个数据集的样本示例如图 9 所示。

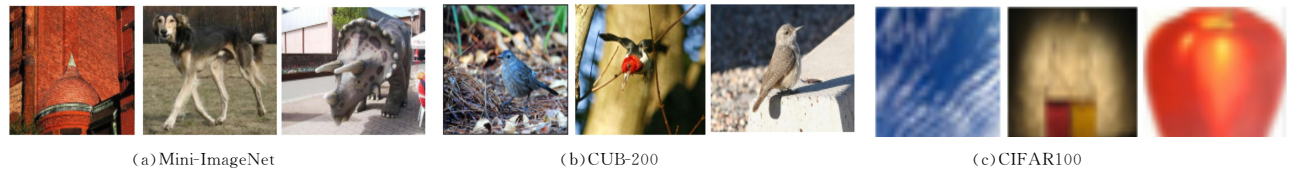


图 9 3 种公用数据集的样本示例图

Fig. 9 Sample of three common datasets

4.2 算法评估:N-way K-shot

小样本图像分类算法通常以 N-way K-shot 进行算法度量。N-way K-shot 指模型是在一组类(基类)上训练的,这些类具有丰富的样本,然后在类别不交叉的另一组类(新类)上

进行测试,即从新类中任意选取 N 类,每类抽取 K 个样本进行训练,然后进行准确率测试,重复数次抽取测试并取平均值。

4.3 实验对比分析

表 2 列出了各种算法在各数据集上的实验结果对比。

表 2 各算法在公用数据集上的实验结果对比

Table 2 Comparison of experimental results of each algorithm on common datasets

Method	Type	Mini-ImageNet		CUB		CIFAR100	
		5 way-1shot	5 way-5shot	5 way-1shot	5 way-5shot	5 way-1shot	5 way-5shot
Meta-Learner LSTM ^[57] (2017)	学习策略	43.40	60.60	40.40	49.70	—	—
Matching Nets ^[49] (2016)	学习策略	46.60	60.00	49.30	59.30	50.50	60.30
MAML ^[55] (2017)	学习策略	48.70	63.10	38.40	59.10	49.30	58.30
Meta-SGD ^[59] (2017)	学习策略	50.47	64.03	53.34	53.34	53.83	70.40
RELATION NET ^[50] (2018)	学习策略	50.44	65.32	—	—	—	—
GCN ^[51] (2018)	学习策略	50.30	66.40	—	—	—	—
Baseline++ ^[32] (2019)	学习策略	51.87	75.68	69.55	85.17	—	—
MAML++ ^[60] (2018)	学习策略	52.15	68.32	—	—	—	—
DEML ^[23] (2018)	学习策略	58.49	71.28	71.28	77.11	61.62	77.94
MTL ^[58] (2019)	学习策略	61.20	75.50	—	—	—	—
EGNN ^[52] (2019)	学习策略	—	76.37	—	—	—	—
LEO ^[22] (2018)	学习策略	61.76	77.59	68.22	78.27	—	—
Delta-Encoder ^[45] (2018)	数据扩充	59.90	69.70	69.80	82.60	66.70	79.80
DTN ^[17] (2020)	数据扩充	63.45	77.91	72.00	85.10	71.50	82.80
S2M2-R ^[40] (2020)	表征学习	64.93	83.18	80.68	90.85	—	—
LaplacianShot ^[61] (2020)	表征学习	75.57	84.72	80.96	88.68	—	—
AmdimNet ^[38] (2019)	表征学习	76.82	90.98	77.09	89.18	—	—
PT-MAP ^[48] (2020)	学习策略	82.92	88.82	91.55	93.99	—	—

从横向对比来看,在各个数据集中 5way-5shot 的分类正确率明显高于 5way-1shot 的分类正确率,这与理论分析一致,即更多的训练数据会使得经验风险更加可靠。

从纵向对比来看,随着时间的推移,不同模型之间的性能出现了明显提升,说明小样本算法近年来取得了较大的突破,例如经典的小样本学习算法 MAML 与 LaplacianShot 等算法的正确率在多个数据集上皆相差较大。

从时间先后的角度进行分析可以发现,早期对于小样本问题的探索主要集中于以元学习、迁移学习为主的学习策略研究,此阶段的代表性算法有 MAML,RELATION NET,EGNN 等。但是只考虑学习策略存在一定的局限性,因此早期算法的正确率普遍较低,在 Mini ImageNet 数据集上 5way-

1shot 的正确率均在 70% 以下。随着表征学习研究的深入,小样本学习算法的性能取得了较大提升,以正则化表征为代表的 LaplacianShot 算法与以自监督为代表的 AmdimNet 算法在多个数据集上的性能大大领先于早期仅考虑学习策略的算法。这与前文的理论分析相符,即有效的表征学习能够同时降低两种风险,因此表征学习往往能带来阶段性的提升。此外,由于数据扩充类算法能一定程度地缩小泛化误差,并具有良好的结合性,因此会进一步提升当前算法的性能,例如表 1 中 Delta-Encoder 于 2018 年获得了较好的性能。

上述分析可以归纳为:首先,基于表征学习的算法能带来阶跃性的提升;其次,学习策略算法中,迁移学习目前取得了最好的效果,但是基于图网络的度量学习以及元学习仍具有

巨大的进步空间;最后,数据扩充技术通常能与现有前沿算法相结合并进一步提升算法性能,因此数据扩充技术常被作为提升算法性能的工具之一。

结束语 小样本学习算法作为推动传统产业智能化转型的关键技术,它的工程意义以及理论意义都使其得到了广泛的关注。目前小样本学习算法已经在图像识别、光谱识别、工业制造、文本分类、雷达识别、农业病害检测、医疗检测等领域^[3,62-71]得到了广泛应用。

本文从机器学习理论中误差分解的角度描述了小样本学习存在的核心问题,并根据理论分析归纳小样本学习算法的设计动机,按照设计动机将现有算法归为表征学习、数据扩充、学习策略三大类。同时,对每类方法的核心思想以及代表性算法进行了对比分析,并结合实验结果概括了每类技术的优缺点以及未来的发展方向。

在现有研究方法的基础上进行分析与展望发现,该领域还存在以下具有挑战性的研究方向。

(1)考虑如何通过自监督、无监督等方式来进一步提升图像表征效果。在表征学习的基础上,能够使用更简单的假设空间使其包含最优映射,同时由于假设空间复杂度降低,对样本量需求减少,更容易泛化。

(2)考虑设计更加通用高效的数据扩充算法,同时考虑其高置信度与高区分度,并合理权衡其比重。

(3)考虑在小样本学习中更灵活、高效地使用元学习思想。

(4)考虑当仅存在异质化任务,即不存在类似的基类训练集,只能使用与目标任务存在一定差异的公开数据集时,如何设计高效的小样本学习算法。

参 考 文 献

- [1] REN S, HE K, GIRSHICK R B, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [C]//International Conference on Neural Information Processing Systems. 2015.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet Classification with Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.
- [3] YAN L, ZHENG Y, CAO J. Few-shot Learning for Short Text Classification[J]. Multimedia Tools Applications, 2018, 77(22): 29799-29810.
- [4] LI F F, FERGUS R, PERONA P. One-shot Learning of Object Categories[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2006, 28(4): 594-611.
- [5] LAKE B M, SALAKHUTDINOV R, TENENBAUM J B. One-shot Learning by Inverting a Compositional Causal Process [C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. 2013: 2526-2534.
- [6] TANG K D, TAPPEN M F, SUKTHANKAR R, et al. Optimizing One-shot Recognition with Micro-set Learning [C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010: 3027-3034.
- [7] MUNKHDALAI T, YUAN X, MEHRI S, et al. Rapid Adaptation with Conditionally Shifted Neurons [C]//International Conference on Machine Learning. 2018: 3664-3673.
- [8] FINK M. Object Classification From a Single Example Utilizing Class Relevance Metrics [C]//International Conference on Neural Information Processing Systems. 2004: 449-456.
- [9] XU Z, ZHU L, YANG Y. Few-shot Object Recognition from Machine-labeled Web Images [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1164-1172.
- [10] SIMON C, KONIUSZ P, NOCK R, et al. Adaptive Subspaces for Few-shot Learning [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 4136-4145.
- [11] CHEN Z, FU Y, ZHANG Y, et al. Multi-level Semantic Feature Augmentation for One-shot Learning[J]. IEEE Transactions on Image Processing, 2019, 28(9): 4594-4605.
- [12] DOERSCH C, GUPTA A, EFROS A A. Unsupervised Visual Representation Learning by Context Prediction [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1422-1430.
- [13] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context Encoders: Feature Learning by Inpainting [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2536-2544.
- [14] ZHANG R, ISOLA P, EFROS A A. Colorful Image Colorization [C]//European Conference on Computer Vision. 2016: 649-666.
- [15] GIDARIS S, SINGH P, KOMODAKIS N. Unsupervised Representation Learning by Predicting Image Rotations [C]//International Conference on Learning Representations. 2018.
- [16] RITCHIE D, WANG K, LIN Y A. Fast and Flexible Indoor Scene Synthesis via Deep Convolutional Generative Models [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.
- [17] CHEN M, FANG Y, WANG X, et al. Diversity Transfer Network for Few-shot Learning [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 10559-10566.
- [18] SANTORO A, BARTUNOV S, BOTVINICK M, et al. Meta-learning with Memory-augmented Neural Networks [C]//International Conference on Machine Learning. 2016: 1842-1850.
- [19] WANG Y X, HEBERT M. Learning from Small Sample Sets by Combining Unsupervised Meta-training with CNNs [C]//International Conference on Neural Information Processing Systems. 2016: 244-252.
- [20] JAMAL M A, QI G J. Task Agnostic Meta-learning for Few-shot Learning [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 11719-11727.
- [21] MISHRA N, ROHANINEJAD M, CHEN X, et al. A Simple Neural Attentive Meta-Learner [C]//International Conference on Learning Representations. 2018.

- [22] RUSU A A, RAO D, SYGNOWSKI J, et al. Meta-Learning with Latent Embedding Optimization[C] // International Conference on Learning Representations. 2018.
- [23] ZHOU F, WU B, LI Z. Deep Meta-learning: Learning to Learn in The Concept Space[J]. arXiv:1802.03596, 2018.
- [24] BERTINETTO L, HENRIQUES J, TORR P, et al. Meta-learning with Differentiable Closed-form Solvers[C] // International Conference on Learning Representations. 2019.
- [25] ZHAO K L, JIN X L, WANG Y Z. Survey on Few - shot Learning [J]. Journal of Software, 2021, 32(2): 349-369.
- [26] LIU Y, LEI Y B, FAN J L, et al. Survey on Image Classification Technology Based on Small Sample Learning [J]. Acta Automatica Sinica, 2021, 47(2): 297-315.
- [27] BOTTOU L, CURTIS F E, NOCEDAL J. Optimization Methods for Large-scale Machine Learning [J]. Siam Review, 2018, 60(2): 223-311.
- [28] BOTTOU L, BOUSQUET O. The Tradeoffs of Large Scale Learning[C] // International Conference on Neural Information Processing Systems. 2007: 161-168.
- [29] WANG Y, YAO Q, KWOK J T, et al. Generalizing from A Few Examples: A Survey on Few-shot Learning[J]. ACM Computing Surveys, 2020, 53(3): 1-34.
- [30] QI H, BROWN M, LOWE D G. Low-shot Learning with Imprinted Weights[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 5822-5830.
- [31] KOCH G. Siamese Neural Networks for One-shot Image Recognition[C] // International Conference on Machine Learning Deep Learning Workshop. 2015.
- [32] CHEN W Y, LIU Y C, KIRA Z, et al. A Closer Look at Few-shot Classification[C] // International Conference on Learning Representations. 2019.
- [33] GIDARIS S, BURSUC A, KOMODAKIS N, et al. Boosting Few-shot Visual Learning with Self-supervision[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 8059-8068.
- [34] SU J C, MAJI S, HARIHARAN B. When does Self-supervision Improve Few-shot Learning? [C] // European Conference on Computer Vision. 2020: 645-666.
- [35] MISRA I, MAATEN L V D. Self-supervised Learning of Pre-text-invariant Representations[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 6707-6717.
- [36] HJELM R D, FEDOROV A, LAVOIE-MARCHILDON S, et al. Learning Deep Representations by Mutual Information Estimation and Maximization[C] // International Conference on Learning Representations. 2018.
- [37] BACHMAN P, HJELM R D, BUCHWALTER W. Learning Representations by Maximizing Mutual Information Across Views[J]. arXiv:1906.00910, 2019.
- [38] CHEN D, CHEN Y, LI Y, et al. Self-supervised Learning for Few-shot Image Classification[J]. arXiv:1911.06045, 2019.
- [39] TSAI Y H, HUANG L K, SALAKHUTDINOV R. Learning Robust Visual-Semantic Embeddings[C] // Proceedings of the IEEE International Conference on Computer Vision. 2017: 3571-3580.
- [40] MANGLA P, KUMARI N, SINHA A, et al. Charting the Right Manifold: Manifold Mixup for Few-shot Learning[C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020: 2218-2227.
- [41] GUO H, MAO Y, ZHANG R. Mixup as Locally Linear Out-of-manifold Regularization[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2019: 3714-3722.
- [42] ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup: Beyond Empirical Risk Minimization[J]. arXiv:1710.09412, 2017.
- [43] WANG Y X, GIRSHICK R, HEBERT M, et al. Low-shot Learning from Imaginary Data[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7278-7286.
- [44] AZURI I, WEINSHALL A D. Generative Latent Implicit Conditional Optimization when Learning from Small Sample[C] // International Conference on Pattern Recognition. 2020.
- [45] SCHWARTZ E, KARLINSKY L, SHTOK J, et al. Delta-Encoder: An Effective Sample Synthesis Method for Few-shot Object Recognition[C] // International Conference on Neural Information Processing Systems. 2018: 2850-2860.
- [46] DHILLON G S, CHAUDHARI P, RAVICHANDRAN A, et al. A Baseline for Few-Shot Image Classification[C] // International Conference on Learning Representations. 2019.
- [47] WANG Y X, HEBERT M. Learning to Learn: Model Regression Networks for Easy Small Sample Learning[C] // European Conference on Computer Vision. 2016: 616-634.
- [48] HU Y, GRIPON V, PATEUX S. Leveraging the Feature Distribution in Transfer-based Few-shot Learning[J]. arXiv: 2006.03806, 2020.
- [49] VINYALS O, BLUNDELL C, LILICRAB T, et al. Matching Networks for One shot Learning[C] // International Conference on Neural Information Processing Systems. 2016: 3637-3645.
- [50] SUNG F, YANG Y, 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. 2018: 1199-1208.
- [51] GARCIA V, BRUNA J. Few-Shot Learning with Graph Neural Networks[C] // International Conference on Learning Representations. 2018.
- [52] KIM J, KIM T, KIM S, et al. Edge-labeling Graph Neural Network for Few-shot Learning[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 11-20.
- [53] PHAM H, DAI Z, XIE Q, et al. Meta Pseudo Labels[J]. arXiv: 2003.10580, 2020.
- [54] LI J, WONG Y, ZHAO Q, et al. Learning to Learn from Noisy Labeled Data[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 5051-5059.
- [55] FINN C, ABBEEL P, LEVINE S. Model-agnostic Meta-learning

- for Fast Adaptation of Deep Networks[C]//International Conference on Machine Learning. 2017:1126-1135.
- [56] NICHOL A, SCHULMAN J. Reptile: A Scalable Metalearning Algorithm[J]. arXiv:1803.02999, 2018.
- [57] RAVI S, LAROCHELLE H. Optimization as a Model for Few-shot Learning[C]//International Conference on Learning Representations. 2017.
- [58] SUN Q, LIU Y, CHUA T S, et al. Meta-transfer Learning for Few-shot Learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:403-412.
- [59] LI Z, ZHOU F, CHEN F, et al. Meta-SGD: Learning to Learn Quickly for Few-shot Learning[J]. arXiv:1707.09835, 2017.
- [60] ANTONIOU A, EDWARDS H, STORKEY A. How to Train your MAML[J]. arXiv:1810.09502, 2018.
- [61] ZIKO I, DOLZ J, GRANGER E, et al. Laplacian Regularized Few-shot Learning[C]//International Conference on Machine Learning. 2020:11660-11670.
- [62] SHYAM P, GUPTA S, DUKKIPATI A. Attentive Recurrent Comparators[C]//International Conference on Machine Learning. 2017:3173-3181.
- [63] SNELL J, SWERSKY K, ZEMEL R. Prototypical Networks for Few-shot Learning[C]//International Conference on Neural Information Processing Systems. 2017:4080-4090.
- [64] TRIANTAFILLOU E, ZEMEL R S, URTASUN R. Few-Shot Learning Through an Information Retrieval Lens[C]//International Conference on Neural Information Processing Systems. 2017.
- [65] LU Y Q, MIN W Q, DUAN H, et al. Few-shot Food Recognition Combining Triplet Convolutional Neural Network with Relational Network [J]. Computer Science, 2020, 47 (1):136-143.
- [66] KANG B, LIU Z, WANG X, et al. Few-shot Object Detection via Feature Reweighting[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:8420-8429.
- [67] ROSTAMI M, KOLOURI S, EATON E, et al. Deep Transfer Learning for Few-shot SAR Image classification [J]. Remote Sensing. 2019, 11(11):1374.
- [68] WANG L, BAI X, ZHOU F. Few-Shot SAR ATR Based on Conv-BiLSTM Prototypical Networks[C]//2019 6th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR). 2019:1-5.
- [69] FAN Q, ZHUO W, TANG C K, et al. Few-shot Object Detection with Attention-RPN and Multi-relation Detector[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:4013-4022.
- [70] LIU W, ZHANG C, LIN G, et al. Crnet: Cross-reference Networks for Few-shot Segmentation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:4165-4173.
- [71] LI Y, YANG J. Meta-learning Baselines and Database for Few-shot Classification in Agriculture [J]. Computers and Electronics in Agriculture, 2021, 182:106055.



PENG Yun-cong, born in 1998, post-graduate. His main research interests include few-shot learning and theory of statistical machine learning.



QIN Xiao-lin, born in 1980, Ph.D, professor, Ph.D supervisor. His main research interests include automatic reasoning and swarm intelligence.

(责任编辑:李亚辉)