

基于小样本学习的图像分类技术综述

刘颖^{1,2} 雷研博¹ 范九伦¹ 王富平^{1,2} 公衍超^{1,2} 田奇³

摘要 图像分类的应用场景非常广泛,很多场景下难以收集到足够多的数据来训练模型,利用小样本学习进行图像分类可解决训练数据量小的问题. 本文对近年来的小样本图像分类算法进行了详细综述,根据不同的建模方式,将现有算法分为卷积神经网络模型和图神经网络模型两大类,其中基于卷积神经网络模型的算法包括四种学习范式:迁移学习、元学习、对偶学习和贝叶斯学习;基于图神经网络模型的算法原本适用于非欧几里得结构数据,但有部分学者将其应用于解决小样本下欧几里得数据的图像分类任务,有关的研究成果目前相对较少. 此外,本文汇总了现有文献中出现的数据集并通过实验结果对现有算法的性能进行了比较. 最后,讨论了小样本图像分类技术的难点及未来研究趋势.

关键词 迁移学习, 元学习, 对偶学习, 贝叶斯学习, 图神经网络

引用格式 刘颖, 雷研博, 范九伦, 王富平, 公衍超, 田奇. 基于小样本学习的图像分类技术综述. 自动化学报, 2021, 47(2): 297–315

DOI 10.16383/j.aas.c190720

Survey on Image Classification Technology Based on Small Sample Learning

LIU Ying^{1,2} LEI Yan-Bo¹ FAN Jiu-Lun¹ WANG Fu-Ping^{1,2} GONG Yan-Chao^{1,2} TIAN Qi³

Abstract Image classification is widely used in different fields. However in many scenarios, it is difficult to collect sufficient data to train the model for classification. Small sample learning provides a solution to this problem. This paper provides a comprehensive survey on recent small sample learning techniques for image classification. Existing algorithms are divided into two categories based on different modeling methods: Convolution based neural network model and graph based neural network model. The algorithms based on convolution neural network model include four learning paradigms: transfer learning, meta learning, dual learning and Bayesian learning. The algorithm based on graph neural network model was originally suitable for non Euclidean data, but some scholars applied it to solve the image classification task of Euclidean data under small samples, which is relatively scarce. In addition, this paper summarizes the existing data sets in literature and compares the performance of existing algorithms through experiments. Finally, the challenges and research trends in small sample image classification are discussed.

Key words Transfer learning, meta learning, dual learning, Bayesian learning, graph neural network

Citation Liu Ying, Lei Yan-Bo, Fan Jiu-Lun, Wang Fu-Ping, Gong Yan-Chao, Tian Qi. Survey on image classification technology based on small sample learning. *Acta Automatica Sinica*, 2021, 47(2): 297–315

图像分类是一个经典的研究课题,典型的图像分类算法涉及两个问题,一是如何对图像特征进行

更好的表示,二是如何学习好的分类参数. 随着卷积神经网络 (Convolutional neural networks, CNN) 的设计越来越深层化,图像特征的代表能力越来越强,同时也能对图像进行自动分类. 在 CNN 提出之前,人类通过人工设计的图像描述符对图像特征进行提取,效果卓有成效,例如尺度不变特征变换 (Scale-invariant feature transform, SIFT)^[1]、方向梯度直方图 (Histogram of oriented gradient, HOG)^[2], 还有词袋模型 (Bag-of-words, BoW)^[3-6] 等,但是人工设计特征通常需要花费很大精力,并且不具有普适性. 基于 CNN 的深度学习在 2012 年的 ILS-VRC 挑战赛上取得了巨大成功,同时由于大数据、计算硬件的发展以及反向传播算法^[7]的提出,深度学习在图像分类领域^[8]表现出优异性能.

深度学习的优势有赖于大数据,在大数据的驱动下,机器能够进行有效学习,然而数据量不足会

收稿日期 2019-10-17 录用日期 2020-04-16

Manuscript received October 17, 2019; accepted April 16, 2020
公安部科技强警 (2016GABJC51), 国家自然科学基金 (61671377, 61802305), 陕西省国际科技合作计划 (2018KW-003), 西安邮电大学研究生创新基金项目 (CXJJLY2019087) 资助

Supported by Science and Technology Project Fund under Ministry of Public Security of China (2016GABJC51), National Natural Science Foundation of China (61671377, 61802305), The Project of International Scientific and Technological Cooperation and Exchange in Shaanxi Province of China (2018KW-003), Innovation fund of Xi'an University of Posts and Telecommunications (CXJJLY2019087)

本文责任编辑 赖剑煌

Recommended by Associate Editor LAI Jian-Huang

1. 西安邮电大学图像与信息处理研究所 西安 710021 2. 电子信息现场勘验应用技术公安部重点实验室 西安 710021 3. 华为云深圳 518000

1. Institute of Image and Information Processing, Xi'an University of Posts and Telecommunications, Xi'an 710021 2. Key Laboratory of the Ministry of Public Security, Xi'an 710021 3. Huawei Cloud, Shenzhen 518000

导致模型出现过拟合等问题, 虽然简单的数据增强 (Data augmentation, DA) 和正则化技术可以缓解该问题, 但是该问题并没有得到完全解决^[9], 故针对小数据集的小样本学习成为了解决这类问题的关键技术. 2003 年 Li 等首次提出了 One-shot learning 问题并利用贝叶斯框架^[10-11]对视觉对象进行学习, 与之相比, 深度学习技术在小样本图像分类任务上具有更大的潜力. 小样本学习的产生有两个因素, 一是训练数据量少. 比如在医疗领域, 医学影像的产生来源于病例, 通常少量的病例并不能够辅助机器对医疗影像进行分析. 二是让机器学会以人类的方式进行学习, 即类人学习. 人类能够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^[12], 小样本学习的目的之一就是让机器以人类的学习方式完成任务. 小样本学习目前在目标识别^[13]、图像分割^[14]、图像分类与检索^[15]等多种图像处理任务中都有应用, 文献 [16-19] 从样本数量、带标签样本的数量、先验知识的作用等不同角度对小样本学习技术进行了分类讨论和分析.

在 2010 至 2015 年间, 大量的文献利用语义迁移来解决训练样本不足的问题. 例如, Mensink 等^[20-21]借鉴了聚类和度量学习的方法对 ImageNet 数据集进行分类, 并且探索了 KNN (K-nearest neighbor) 和 NCM (Nearest class mean) 分类器, 通过对每个类的语义描述, 学习一个度量, 使其在训练和测试类别间共享, 达到迁移效果; 文献 [22] 将语义知识迁移扩展到直推式学习, 利用已知类别推测未知类别的表示, 计算未知类别的样本相似性, 在构建数据空间分布关系时, 将数据投影到低维的语义空间, 再进一步寻找数据的空间分布, 并在 AwA (Animals with attributes)^[23]、ImageNet^[24] 以及 MPII composites activities^[25] 数据集上取得很好的分类效果; 文献 [26] 提出直推式多视图嵌入框架来解决领域漂移问题, 利用异构多视图标签传播来解决原型稀疏性问题, 有效利用了不同语义表示提供的互补信息, 并在 AwA、CUB (Caltech-UCSD-Birds)^[27]、USAA (Unstructured social activity attribute)^[28] 数据集上取得了很好效果; Fu 等^[28]为解决带有稀疏和不完整标签的社交媒体数据的属性学习问题, 利用零样本学习思想提出了一种学习半潜在在属性空间的模型, 它能够表达用户自定义和潜在的属性信息, 在 USAA 数据集上取得很好的效果. 这些文章针对的多是零样本学习问题, 本文主要介绍小样本下的图像分类算法, 因此综述算法更集中于 Few-shot learning.

近年来现有文献中基于小样本学习的图像分类算法都是采用深度学习. 将深度学习中的技术用于小样本学习中, 比如使用数据增强技术来增加样本的数量, 通过注意力机制和记忆力机制来对图像特征进行提取, 设计提取特征网络和分类器之间的映射关系, 与此同时, 迁移学习、元学习、对偶学习、贝叶斯学习以及图神经网络方法也被用于小样本图像分类的任务. 本文的小样本学习算法与其他几篇小样本文献相比较有两点区别, 一是本文阐述的小样本学习算法是针对图像分类任务, 文献 [16-19] 中的小样本学习算法不仅应用于图像分类, 还有识别、分割等图像任务以及小样本学习在语音、视频中的算法应用, 本文集中地对小样本图像分类算法进行了分类并归纳总结; 二是本文探索了不同的网络建模方式, 将小样本图像分类算法分为卷积神经网络模型和图神经网络模型两大类, 卷积神经网络模型主要基于 CNN 对图像数据建模, 图神经网络模型进一步将 CNN 应用在图神经网络上, 通过图结构中的节点和边来对图像数据建模.

本文的结构如下, 第 1 节介绍了小样本图像分类的流程, 小样本图像分类数据集和实验评价指标; 第 2 节将现有小样本图像分类算法按照数据结构类型分为卷积神经网络模型和图神经网络模型两大类并进行详细介绍; 第 3 节通过实验数据对比分析了各种算法的性能; 第 4 节总结了小样本图像分类面临的技术挑战并讨论了未来研究趋势; 第 5 节总结全文.

1 小样本图像分类介绍

1.1 基于小样本学习的图像分类算法

小样本学习是指训练类别样本较少的情况下, 进行相关的学习任务, 一般地, 我们也希望机器通过学习大量的基类 (Base class) 后, 仅仅需要少量样本就能快速学习到新类 (New class). 通常情况下, 小样本学习能够利用类别中的少量样本, 即一个或者几个样本进行学习. 类别下的训练样本只有一个的情况下, 小样本学习被称为 One-shot learning^[10], 类别下的训练样本有多个的情况下, 称为 Few-shot learning^[29-30], Few-shot learning 包含有 One-shot learning 的情况.

小样本图像分类流程如图 1 所示, 包括准备数据集、构建用于图像特征提取的网络和设计分类器三个步骤. 以下对建立小样本图像分类流程的三个步骤进行具体介绍.

1.1.1 数据集处理

本文将处理小样本图像数据集的方式分为两

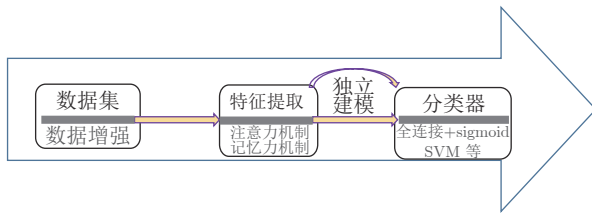
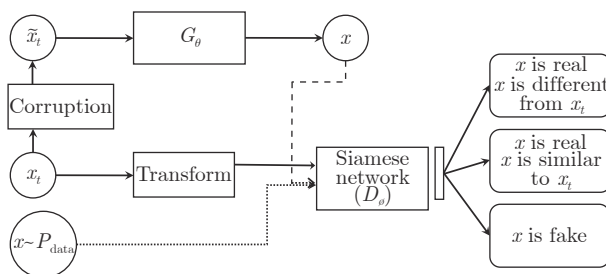


图 1 小样本图像分类流程

Fig.1 The procedure of small sample image classification

种. 一是进行数据增强, 将数据集进行量级扩增和模式扩增, 量级扩增是指针对数据量级的扩大, 模式扩增是指在量级扩增的同时, 让不同的样本包含更多的语义特性. 训练模型时, 如果每类包含大量的样本, 且这些样本包含大量的模式, 则在缓解过拟合问题的同时, 模型会具有更好的鲁棒性和泛化能力. 除了图像样本数目的扩增, 图像特征的增强也是一种数据增强的方式^[31]; 二是不对小样本数据集进行处理, 在只有少量样本的情况下, 让模型适应数据, 针对数据的特点进行建模^[32]. 对于小样本数据集, 设计用于提取表示能力强的特征的网络架构往往非常重要.

小样本图像数据集不足以让模型捕捉到足够的数据模式, 而应用于深度学习的卷积神经网络又需要大量带标签的训练样本, 因此最直接的解决办法就是进行数据增强^[33]. 数据增强包括三种方式: 一是通过旋转、加噪、裁剪、压缩等操作^[34], 这种数据增强的方式在处理样本不足的问题中都有应用; 二是利用生成模型生成新的样本或者生成新的特征. 例如, Jia 等提出使用贝叶斯方法生成新的样本^[35], 生成对抗网络可为训练样本提供更多的生成图像, 增加训练样本的多样性. 2017 年 Mehrotra 等提出了生成网络 + 孪生网络的网络结构^[36], 生成对抗网络由生成器和判别器组成, 如图 2 所示, 利用生成器进行数据增强, 孪生网络用于判别器. 孪生网络第一次被 Bromley 等在 1993 年提出并用来解决一

图 2 生成对抗网络 + 孪生网络^[36]Fig.2 Generative adversarial networks + siamese networks^[36]

种匹配问题的签名验证^[37]; 三是通过函数变换, 将图像中需要关注的物体本身或者其模式通过某种变换, 达到增加样本数量或者特征数量的目的. 例如, Dixit 等利用含有属性标注的语料库来生成新的样本^[38]. 如果收集到的数据包含少量带标签的样本和大量未带标签的样本, 可以利用带标签的样本去标注未带标签的样本, 产生的伪样本也可以作为增加的训练样本^[39].

1.1.2 特征提取

这个过程主要是为适应数据分布建立特征提取模型, 该模型能够提取图像的有效特征, 图像特征的有效性可解释为, 对于一个模型, 提取到的图像特征可以对图像进行有效表示, 达到更好的分类效果. 为了提高图像特征的有效性, 注意力机制^[31]、记忆力机制^[34]等技术被应用于小样本图像分类算法中.

1) 注意力机制

注意力机制是人类视觉所特有的大脑信号处理机制, 人类视觉通过快速扫描全局图像, 获得重点关注目标区域, 然后抑制其他无用信息, 极大地提高了视觉信息处理的效率与准确性. 注意力模型在图像领域被广泛使用, 它借鉴了人类的注意力机制, 在对图像的处理过程中, 始终关注感兴趣的部分区域, 因此在建立提取图像特征模型的过程中, 注意力机制扮演着将图像信息进一步提取为有效信息的角色, 学习不同局部的重要性.

注意力机制在数学形式上可以理解为加权求和, 通常情况下使用 Softmax 形式, 并通过引入新的参数来弥补模型的拟合能力. 文献 [40] 在对图像特征提取的过程中使用了单一注意力机制, 文献 [41] 认为单一的注意力机制对图像信息的提取不够充分, 提出利用多注意力机制将类别标签信息与视觉信息联系起来, 减小视觉信息与语义信息之间的鸿沟.

2) 记忆力机制

循环神经网络 (Recurrent neural network, RNN) 解决了短期记忆的问题, 其变体长短期记忆网络 (Long short-term memory, LSTM) 解决了短期记忆和长期依赖的问题. 在图像领域, 基于 LSTM 的记忆力机制近年来发挥着重要作用, 它增强了模型的拟合能力和图像特征的代表能力. 小样本图像分类算法对于记忆力机制的应用可分为两类, 一是直接利用 LSTM 对图像进行编码, 提高图像特征的代表能力^[9]; 二是利用读写控制器对记忆信息进行写入和读出^[42].

1.1.3 分类器

分类器的设计取决于图像特征的有效性和分类器与图像特征之间的适应性. 分类器与图像特征之

间的适应性可解释为,在假定图像特征具有有效性的前提下,分类器能够最大程度地区分不同类别的图像特征.通常小样本图像分类中所使用的分类器,大多数是在卷积神经网络的最后一层构建带有 Softmax 的全连接层,或者对提取的图像特征应用 K 近邻 (K-nearest neighbor, KNN) 算法,还有对分类器的权重进行重新生成,使模型同时适用于基类数据集和新类数据集,目前现有的元学习方法一般不研究将基类和新类一起进行分类的问题.

分类器本质上是对特征进行相似性度量并对不同类别进行最大程度的区分.传统机器学习中最简单的度量方式 KNN,可对每个样本间的距离进行度量,然后进行相似性距离排序,例如,可以利用 1-nearest neighbor,即 1-NN 来完成 One-shot learning 任务,但实验表明分类效果并不好^[43].除此之外,还可以通过支持向量机 (Support vector machine, SVM) 进行分类度量^[44-46].近邻成分分析 (Neighborhood component analysis, NCA)^[47]及其非线性方法^[48]、基于集合的弱度量方法^[49]等都是样本特征度量工作的范畴.在小样本图像分类任务中,也有利用余弦距离和欧氏距离以及点乘方式对特征距离进行度量.

一般地,当模型学习到新的类别后,会忘记之前学习过的类别,与之前所做工作不同的是,Gridaris 等^[50]提出了基于注意力机制的分类器权重生成器,通过重新设计分类器来适应分类器权重和图像特征之间的匹配程度,使模型同时适用分类基类和新类样本.类似于上述工作,Chen 等^[51]将线性分类器替换为基于距离的分类器,以比较两种分类器在不同数据集上的优劣.

1.2 数据集及分类指标

1.2.1 小样本公用数据集介绍

近年来,现有文献中的小样本公用数据集主要包括: Omniglot^[52]、CIFAR-100^[53]、Mini-ImageNet^[9]、Tiered-ImageNet^[54]和 CUB-200^[27].从数据量级的大小来看,数据集 Tiered-ImageNet,不仅数据量级较大、类别多,而且每类包含的样本也多;量级较小的数据集,例如 Mini-ImageNet 数据集、CIFAR-100 数据集、CUB-200 数据集,这类数据集类别较少、类内样本数相对较多;Omniglot 数据集类别较多,但是相对类内样本少.从数据类型的复杂度来看, Omniglot 数据集属于字符类型的图像,包含的类型和模式较为简单,对其进行实验往往分类精度较高;其他数据集都属于自然图像,包含的图像模式较为复杂,对其进行实验分类精度往往相对较低.

上述数据集的相关信息如表 1 和图 3 所示.小样本图像分类使用的公用数据集图像类别均达到或超过 100 类,总体数据量均超过 10 000, Tiered-ImageNet 数据集达到 77 多万, Omniglot 数据集和 CUB-200 数据集的平均类内样本数未达到 100, Tiered-ImageNet 数据集的平均类内样本数超过 1000.从本文第 3 节的实验分析将看到,类别越多,类内样本越多,越有利于进行小样本图像分类,这表明数据量级的大小对小样本图像分类结果具有一定的影响.

表 1 小样本公用数据集的数量信息

Table 1 Quantitative information of small sample public data sets

数据集	数据数量	类别数量	平均类内样本
Omniglot ^[52]	32 460	1 623	20
CIFAR-100 ^[53]	60 000	100	600
Mini-ImageNet ^[9]	60 000	100	600
Tiered-ImageNet ^[54]	778 848	608	1 281
CUB-200 ^[27]	11 788	200	58

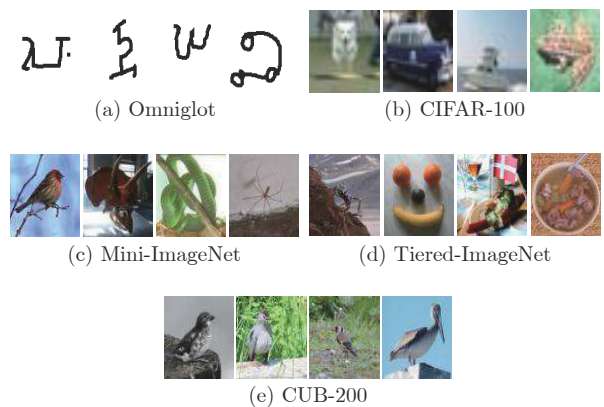


图 3 小样本公用数据集样本示例

Fig. 3 Sample examples of small sample public data sets

1.2.2 评价指标

小样本图像分类算法的实验评价指标通常称为 N -way K -shot^[9].也有使用 top-1 和 top-5 来评价图像分类精度^[55]. N -way K -shot: 选取 N 类图像样本,每类图像选取 K 个样本或样本对,一般地, $N \in \{5, 10, 15, 20\}$, $K \in \{1, 5\}$.模型训练阶段,构建好训练模型并在选取的 $N \times K$ 个样本或样本对上进行训练;在验证阶段和测试阶段,选取 N 类样本中的 K 个样本或者样本对,执行 N -way K -shot 分类任务.根据预测结果来确定预测类别,预测类别与实际类别相符的准确率即为评价指标. Top-1: 指预测排名第一的类别与实际结果相符的准确率. Top-5:

指预测排名前五的类别包含实际结果的准确率。

2 小样本图像分类算法

针对不同类型数据的建模方式, 本文将小样本图像分类算法分为卷积神经网络模型和图神经网络模型. 根据学习范式, 卷积神经网络模型可分为迁移学习、元学习、对偶学习和贝叶斯学习. 基于迁移学习的小样本图像分类有三种实现方式, 基于特征、基于相关性和基于共享参数; 基于元学习的小样本图像分类有三种实现方式, 基于度量、基于优化和基于模型; 基于对偶学习的小样本图像分类有两种实现方式, 一是利用自动编码器, 二是利用生成对抗网络. 本节将对以上小样本图像分类算法进行详细介绍, 并在 Omniglot 数据集、Mini-ImageNet 数据集、CIFAR-100 数据集和 CUB-200 数据集上进行实验分析.

2.1 迁移学习

利用迁移学习^[56-57]可以减小模型训练的代价, 同时达到让卷积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是有相同规律可寻的, 并且学习第 n 个任务比第 1 个任务要更为简单^[58]. 迁移学习关注的是目标任务, 给定一个源域 D_s 和一个学习任务 T_s , 一个目标域 D_t 和一个学习任务 T_t , 迁移学习的目的是使用在 D_s 和 T_s 上的知识帮助提高在目标域 D_t 上的预测函数 $f_t(x)$ 的学习, 以更好地执行学习任务 T_t , 其中 $D_s \neq D_t$ 或 $T_s \neq T_t$. 如果迁移学习中的源数据和目标数据不同但是具有相关性^[59], 则需要进一步处理. 比如, 使用多源域的决策知识预测目标域的样本标签^[60].

如图 4 所示, 从小样本图像分类的流程来看, 迁移学习是在图像特征提取阶段实现的. 具体的迁移方式可以分为基于特征的迁移、基于共享参数的迁移和基于关系的迁移. 如果将基类作为源域数据, 将新类作为目标域数据, 以基类数据到新类数据的知识迁移为例, 基于特征的迁移是找出基类数据和新类数据之间共同的特征, 通过特征变换的方式将基类数据的知识进行迁移, 用于新类数据分类. 该方法存在的难点在于, 一是寻找基类数据和新类数据的共同特征, 二是采用何种方式对特征进行迁移; 基于关系的迁移是建立基类数据和新类数据之间相关知识的映射, 通过这种关系映射来进行学习. 该方法的难点在于, 一是如何确定映射关系, 二是如何建立映射关系; 基于共享参数的迁移需要找到基于基类数据模型和基于新类数据模型之间的共享参数或者相同的先验分布, 利用这些参数或者先验分布进行知识迁移. 该方法的难点在于如何寻找共享

参数和确定先验分布. 在寻找源域数据和目标域数据的共同特征、知识映射关系以及模型的共享参数和先验分布时, 重要的是搭建能够有效提取图像特征的网络结构以及适用的知识迁移方式. 需要建立可持续学习的模型时, 小样本迁移学习不仅需要保证模型对目标域数据有效, 而且还要确保模型在源域数据和目标域数据都有不错的分类效果.

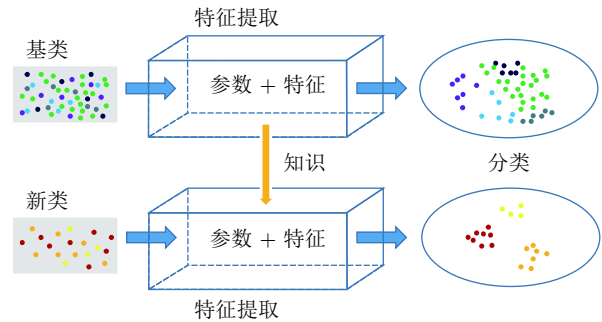


图 4 迁移学习

Fig.4 Transfer learning

2.1.1 基于特征的迁移学习

Hariharan 等^[61]利用基类样本间的模式对新类样本进行相同变换, 达到增加训练样本数量的目的. 具体来讲, 从类 A 中抽取两个样本, 这两个样本间存在某种变换模式, 再从类 B 中取出一个样本, 对这个样本实施和类 A 中两个样本间同样的变换模式以生成新的样本. 该方法使用 ImageNet1k 数据集, 将其分为基类数据集和新类数据集, 基类中含有大量训练样本, 新类含有少量训练样本. 训练模型分为两个阶段, 一是表征学习阶段, 对数据增强后的基类数据进行特征提取, 并构建分类器; 二是小样本学习阶段, 利用基类数据和新类数据共同训练模型, 以获取基类数据和新类数据的共同特征, 并将表征学习阶段提取的特征用于对基类和新类进行分类. 为了使分类器同时适应基类数据和新类数据, 如式 (1) 所示, 提出一个新的损失函数, 用来减小模型

$$Loss = \min_{W, \phi} L_D(\phi, W) + \lambda L_D^{SGM}(\phi, W) \quad (1)$$

在基类和新类上学习能力的差异, 其中, $L_D(\phi, W)$ 表示在基类数据上的损失, 平方梯度 (Squared gradient magnitude, SGM) 损失 $L_D^{SGM}(\phi, W)$ 表示基类与新类之间的差异所造成的损失, 参数 λ 通过交叉验证确定, 以保证构建一个在基类和新类都适用的分类器. Choi 等^[62]针对素描图像和自然图像的小样本学习问题提出了一种结构性集合匹配网络 (Structured set matching networks, SSMN), 利用的是样本间的相关性. 该模型在自建的多标签的素

描图像集合、自然图像集合、素描和自然图像混合的集合间的三个数据集中都有不错的效果. 它利用图像的多标签信息进行域内或跨域迁移, 通过 CNN 和双向 LSTMs 对基类样本和新类样本的局部特征进行提取并映射进同一空间中, 计算局部和全局相似度实现图像分类.

2.1.2 基于关系的迁移学习

将知识压缩进一个单一的模型已经被 Bucilua-ana 等证明是可行的^[63], 进一步地, 2014 年 Hinton 等首次提出了知识蒸馏的概念^[64], 通过引入相对复杂的教师网络, 来诱导精简、低复杂度的学生网络的训练, 将知识从教师网络中迁移到压缩的学生网络中^[65], 实现知识迁移. 学生网络可以通过对教师网络进行修剪^[66-68] 或者压缩^[69-72] 得到, 也可以重新设计一个新的网络架构. 知识蒸馏的目的就是在减少网络架构的同时把网络的知识保留下来, 为了达到这一目的, 提出了一个新的温度参数 Tem , 将输出的概率 (硬目标) 进行软化, 如式 (2) 所示,

$$q = \frac{\exp\left(\frac{z_i}{Tem}\right)}{\sum_i \exp\left(\frac{x_i}{Tem}\right)} \quad (2)$$

其中 z_i 是 Softmax 层的前一层输出, q 是软化后的概率输出 (软目标). 教师网络的预测输出除以温度参数 Tem 之后, 做 Softmax 变换, 可以获得软化的概率分布 (软目标), 数值介于 0 ~ 1 之间, 取值分布较为缓和, 即对于样本的所属类别分别给出一个或大或小的概率, 而不是确定的 0 或 1. Tem 数值越大, 分布越缓和; 而 Tem 数值减小, 容易放大错误分类的概率, 引入不必要的噪声. 针对较困难的分类或检测任务, Tem 通常取 1, 确保教师网络中正确预测的贡献. 硬目标则是样本的真实标注, 可以用 One-hot 矢量表示. 总体的损失设计为软目标与硬目标所对应的交叉熵的加权平均, 其中软目标交叉熵的加权系数越大, 表明迁移诱导越依赖教师网络的贡献, 这对训练初期阶段是很有必要的, 有助于让学生网络更轻松地鉴别简单样本, 但训练后期需要适当减小软目标的比重, 让真实标注帮助鉴别困难样本. 另外, 教师网络的推理性能通常要优于学生网络, 而模型容量则无具体限制, 因此带有标签的训练数据越多, 教师网络推理精度越高, 越有利于学生网络的学习. Kimura 等^[39] 在 2018 年提出了一种不同于之前的知识蒸馏方法, 仅仅需要使用少量的训练样本就可以实现知识迁移. 利用少量可得带标签的训练样本训练一个高斯过程的模型作为教师网络, 以克服过拟合问题, 接着如同知识蒸

馏一样, 将模型中的知识迁移到学生网络模型中, 同时使用诱导点^[73]作为增加的训练样本, 对学生网络进行训练. 2019 年有研究人员提出在学生网络中添加 1×1 卷积层^[74], 实现知识蒸馏^[75]. 该算法使用预训练模型 VGG 或者 ResNet 等作为教师网络, 在学生网络中添加 1×1 卷积层, 进一步减少参数的数量, 保持特征映射尺度不变的同时增加网络的非线性, 使用最小二乘回归将其与教师网络进行对齐, 即对学生网络进行诱导学习, 经过网络训练后得到最终的学生网络.

2.1.3 基于共享参数的迁移学习

Oquab 等^[76] 采用微调策略. 该算法中, 对图像进行多块分解实现数据增强, 加强了模型以局部视角识别图像的能力. Oquab 等认为卷积神经网络提取的中层特征能够对图像进行很好的表示, 利用在 ImageNet 数据集上预训练的模型^[8], 对图像中层特征进行提取, 并重新构建分类层, 构建新的网络对数据集分类.

Qi 等^[77] 提出将迁移学习和增量学习进行结合, 通过对分类器的权重进行处理来实现增量零训练. 该算法利用卷积神经网络作为特征提取器以共享参数, 对新样本进行特征提取后, 产生一个分类权重向量, 将其扩展进预训练的分类器权重中, 以应对新样本的分类任务.

除了将迁移学习与增量学习进行结合, 也可对特征提取器与分类器间的映射关系进行独立建模. Qiao 等^[55] 在 2018 年提出直接从激活函数层预测分类参数的算法 (Predicting parameters from activations, PPA), 一般地, 最后一层激活函数层与分类层间有相应的权重连接, 在激活函数和分类层之间建立一个分类参数预测器, 可以更好地对分类器的参数进行调整, 匹配图像特征与分类器.

2.2 元学习

元学习又叫做学会学习, 是机器学习领域一个重要的研究方向, 它解决的是学会如何学习的问题. 传统的机器学习研究模式是: 获取特定任务的数据集, 每次再利用这些数据集从头开始训练模型. 然而, 人类可以通过获取以往的经验, 对同类型的任务或有共性的任务进行快速学习, 这是因为人类懂得如何学习. 如图 5 所示, 如果把特征提取视为机器在数据集上学习的过程, 那么元学习器就是要评估这个学习过程, 也就是让机器学习学习的过程, 即通过学习获得学习经验, 利用这些经验再去对最终的目标任务进行评估. 一种常见的元学习方式是将学习算法编码进卷积神经网络中, 包括基于距离

度量的元学习和基于模型的元学习. 基于距离度量的元学习将图像映射到一个度量空间并使用某种度量方式计算不同图像样本的差异, 度量方式包括固定距离度量^[40] (欧氏距离、余弦距离或点乘) 和非固定距离度量^[62] (例如使用 Sigmoid 计算距离得分); 基于模型的元学习通过构建元模型来获得经验知识^[78], 再利用这些经验去评估最终的分类任务. 另一种元学习方式是基于优化的元学习, 基于优化的元学习目的是使网络具有一个好的初始化^[79].

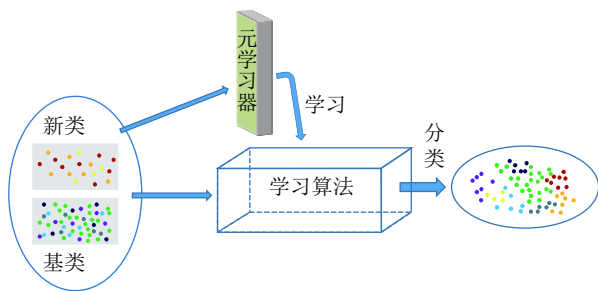


图 5 元学习
Fig. 5 Meta learning

2.2.1 基于度量的元学习

采用固定距离度量方式. 2015 年 Koch 等^[43] 针对字符识别提出了深度卷积孪生网络, 利用全局仿射变换增加训练数据集. 该算法训练一个孪生网络对样本进行相似性判决, 即让样本对通过完全相同的网络结构, 利用欧氏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测试样本进行分类. 虽然该方法提出不依赖先验知识, 易于简化模型, 但是缺少先验知识的辅助信息, 在较为复杂的数据集上进行小样本图像分类任务时难以达到好的效果. Vinyals 等使用余弦距离度量, 设计的匹配网络 (Matching networks, MN)^[9] 经由基于深度特征的度量学习和外部存储器增强的神经网络启发, 可以从小数据集中快速地学习新的概念, 同时避免微调, 且对细粒度图像分类 (Fine-grained image classification) 任务有很好的适应性. 匹配网络采用 “episode” 的形式, 即从原始数据中采样出带有标签的任务集合, 然后从任务集合中获取支持集和目标集构成元任务, 通过对支持集的训练, 来最小化目标集上的误差, 很多元学习的文章对数据集都采取 “episode” 的形式. 通过使用注意力机制和用于上下文嵌入的双向长短期记忆网络 (Long-short term memory, LSTM)^[80], 对图像特征进行充分提取. 计算测试样本标签的数学表达式如式 (3),

$$\hat{y} = \sum_{i=1}^k a(\hat{x}, x_i) y_i \quad (3)$$

其中 x_i, y_i , 是来自支持集 $S = (x_i, y_i)_{i=1}^k$ 的样本及其对应标签, \hat{x} 代表测试样本. 可以看出测试样本标签的输出是支持集中标签的线性组合, a 是一个注意力机制, 通过在余弦距离上使用 Softmax 来实现, 可认为是对于标签的加权系数, 用来衡量支持集中训练样本和测试样本的相关程度. Bartunov 等^[40] 在 2018 年提出了一种生成式匹配网络 (Generative matching networks, GMN), 认为新样本的生成服从某一条件概率分布, 使用该分布生成新样本来进行数据增强, 增加了样本的多样性, 因此不要求训练数据本身具有丰富的多样性, 少量的数据即可进行小样本图像分类任务. 与之前的匹配网络不同, 该方法不是针对样本进行直接匹配, 而是将样本映射到语义嵌入空间, 在嵌入空间中利用条件似然函数对样本的语义特征向量进行匹配, 减小了特征空间和语义空间的鸿沟. Cai 等^[42] 提出了端到端的记忆匹配网络 (Memory matching networks, MMN), 是一种利用内部存储来进行记忆编码的元学习方法, 它将提取到的图像特征用记忆写入控制器压缩进记忆间隙, 然后利用上下文学习器, 即双向的 LSTM 对记忆间隙进行编码, 不仅提高了图像特征的代表能力, 而且能够探索类别之间的关系, 其输出为未标注样本的嵌入向量, 记忆读入控制器通过读入支持集的嵌入向量, 将两者点乘作为距离相似度量, 相比于余弦距离, 计算复杂度更加简单. Snell 等提出的原型网络 (Prototypical networks, PN)^[29] 需要计算类别原型. 通过学习一个度量空间, 在这个度量空间内, 分类器可以根据样本到类别原型间的距离, 来对样本进行分类. 每个类别原型, 可以通过对每个类别中所有样本在度量空间的向量求平均得到, 使用欧氏距离来判断样本所属的类别.

Choi 等采用非固定距离度量^[62], 针对素描图像和自然图像的小样本图像分类问题提出了一种结构性集合匹配网络 (Structured set matching networks, SSMN), 从执行任务的角度来看该方法属于元学习. 该模型利用 RNN 对图像间的所有标签对应的局部信息进行局部相似度计算, 并将局部特征和全局特征进行结合, 利用多标签数据增强图像的解释性, 但同时也增加了标注数据的工作量. 人类在辨别事物的时候, 习惯对不同的事物进行比较, 根据这个简单的思想, Sung 等^[30] 在 2018 年提出的端到端的相关网络 (Relation network, RN), 学习一个深度距离以度量元学习任务的不同样本. 相关

网络由两个模块组成,嵌入模块和相关模块,嵌入模块对不同的样本进行特征提取,相关模块将不同样本的特征进行拼接进而得出不同样本间的相关性度量分数. Zhou 等^[81]提出了基于嵌入回归的视觉类比网络,学习低维的嵌入空间,再从嵌入空间中学习到分类参数的线性映射函数,对新类分类时,将新类样本与学习到基类的嵌入特征进行相似度度量.

2.2.2 基于模型的元学习

一般地,元学习分为两个阶段:执行在不同任务上的元级模型的慢速学习和执行在单个任务中的基准模型的快速学习^[82-83],元级模型的目的是学习不同任务中的通用知识,然后将其传递给基准模型,来帮助在单个任务上的学习. Munkhdalai 等^[78]在 2017 年使用卷积神经网络构造了一种元网络 (Meta networks, meta-Nets),用于跨任务间的学习,遵循之前的工作,将元学习问题分为两个阶段. 它提出了一种更快的学习方法是利用一个神经网络去预测另一个神经网络的参数,生成的参数称为快权值,用来加快网络的学习速度,普通的基于随机梯度下降 (Stochastic gradient descent, SGD) 等优化的参数被称为慢权值. 该模型的训练由元级模型和基准模型共同来执行,包括元信息的获取,快权重的产生和慢权重的更新. 基准模型使用损失梯度信息作为元信息,在不同元任务中获取元信息并存储在外部设备. 元级模型对存储在外部存储设备的元信息进行获取,并通过预测产生快权值来加速网络学习. Zhou 等^[32]在 2018 年提出的深度元学习模型 (Deep meta learning, DML),能够在概念空间中进行学习,而不是在传统上的视觉空间. 该模型由三个模块组成:概念生成器、元学习器和概念判决器. 为了让元学习有一个好的特征表示,概念生成器使用深度残差网络,以捕捉高级语义概念,然后利用概念判决器进行信息反馈和优化,同时将提取到的特征作为概念用于元学习器进一步地学习. 该方法通过在语义概念上的学习减少了视觉空间和语义空间的鸿沟,使用深度残差网络来构建更加复杂的模型,以适应复杂的数据模式. Santoro 等^[34]提出利用 RNN 架构加外部存储记忆的方式. 外部存储的使用使模型精度获得了提高,但是同时也降低了模型的效率并占用了大量的存储空间. Sun 等^[84]提出的元-迁移学习 (Meta-transfer learning, MTL) 利用迁移知识有效减少模型更新的参数,同时构建更深层次的网络增加模型的复杂度.

元学习模型通用方法. 针对小样本数据量较少的特点, Wang 等^[85]在 2018 年提出了对于任何元学习模型都适用的数据增强方式,针对生成模型捕

捉的样本模式不足这一问题,提出使用生成模型来产生新的样本,这些样本是由真实的样本和噪声向量通过 3 层带有 Relu 非线性激活函数的多层感知机 (Multi-layer perception, MLP) 产生的,利用生成的样本和原有的样本共同对模型进行训练. 该数据增强方式能够结合任何元学习算法进行使用,由于采用生成网络产生新样本,因此实验结果的好坏取决于生成新样本的质量. Wang 等^[86]提出了一个模型回归网络 (Model regression networks),利用的是分类器之间的相关性. 该方法认为在小样本数据中学习到的分类器和在大样本数据中学习到的分类器之间存在一种变换,可以通过深度回归网络进行学习,同时该变换作为一种先验知识可以帮助在小样本数据上的分类任务.

2019 年有研究学者将增量学习与元学习进行结合,提出的注意力吸引网络^[87] (Attention attractor networks, AAN) 模型不仅在新类上表现良好,而且不会遗忘在基类上学习到的知识. 如图 6,训练阶段 A,在基类上进行预训练模型,学习分类参数 W_a ,阶段 B 结合注意力机制并利用每次学习一个新任务的分类参数 W_b ,阶段 C 将 W_a 和 W_b 作为基类和新类的分类参数 W_b^* 用来对元任务进行测试. 对于给定的新任务都会学习一个参数 W_b ,代表该任务在执行分类时的贡献,使得分类器更加灵活适用,而且对单个新样本的分类也更加容易.

2.2.3 基于优化的元学习

针对小样本数据集的微调策略,采用的是将模型在大数据集上进行预训练,然后在小数据集上进行简单微调. 然而经过预训练的模型并不能保证对于微调有一个很好的初始化参数. 基于优化的元学习能够保证网络学习到一个好的初始化,使模型对新任务更易于微调.

Finn 等^[79]在 2017 年提出了一种与模型无关 (Model-agnostic meta-learning, MAML) 的元学习算法. 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. 简单地讲,如果在模型中加入新的任务,每个不同的任务会产生不同的损失,利用模型在该任务上的损失进行参数优化,使其快速适用于新的分类任务. 然而 MAML 对神经网络结构非常敏感,导致训练过程不稳定, Antoniou 等^[88]提出对 MAML 进行优化,进一步提高了系统的泛化性能,加快了网络的收敛速度,减少了计算开销. Nichol 等^[89]提出的基于优化的元学习模型 Reptile,也是通过学习网络参数的初始化,与 MAML 不同的是, Reptile 在参数

优化时不要求使用微分. Ravi 等^[90]提出的基于梯度的优化算法, 使用基于 LSTM 的元学习模型去学习一个网络的初始化, 它能够捕捉到单个任务的短期知识和所有任务的长期知识, 以便更好地提取特征用于解释图像.

2.3 对偶学习

为了降低机器对大量标注样本的依赖, 以及在强化学习中减少机器与环境交互的次数, 对偶学习作为一种新的学习范式应运而生. 现实生活中, 很多有实用价值的人工智能任务往往是成对出现的, 例如, 在图像领域, 图像识别和图像生成都有重要

的应用, 属于对偶任务. 如果根据对偶任务来训练模型, 利用任务到任务的反馈信息, 就能克服模型对数据的依赖问题^[91].

神经网络中的自编码器 (Auto encoder, AE) 就是对偶学习的一个特例, 包括两个部分: 编码和解码. 变体算法有降噪自编码器^[92]、堆叠卷积自编码器^[93]和变分自编码器^[94]等, 其应用主要为数据去噪、数据降维以及数据生成. 在进行小样本图像分类的过程中, 由于数据量小, 会进行数据增强操作, 但是数据增强有时候会产生噪声数据, 对于样本数据的多样性表示不足, 可能不会对决策边界产生影响, 需要引入额外的语义知识. 如图 7 所示, Chen

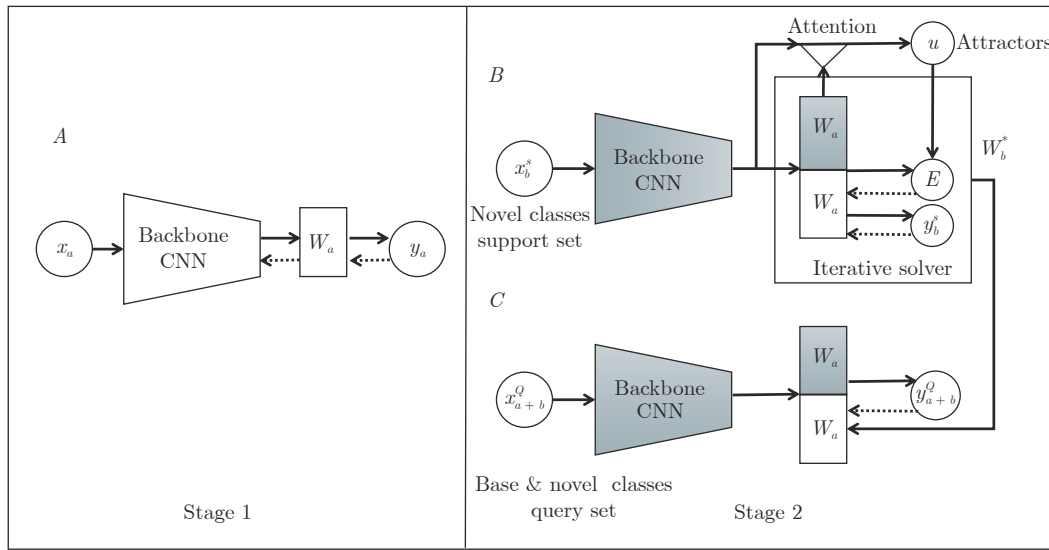


图 6 注意力吸引网络结构^[87]

Fig.6 Attention attractor networks structure^[87]

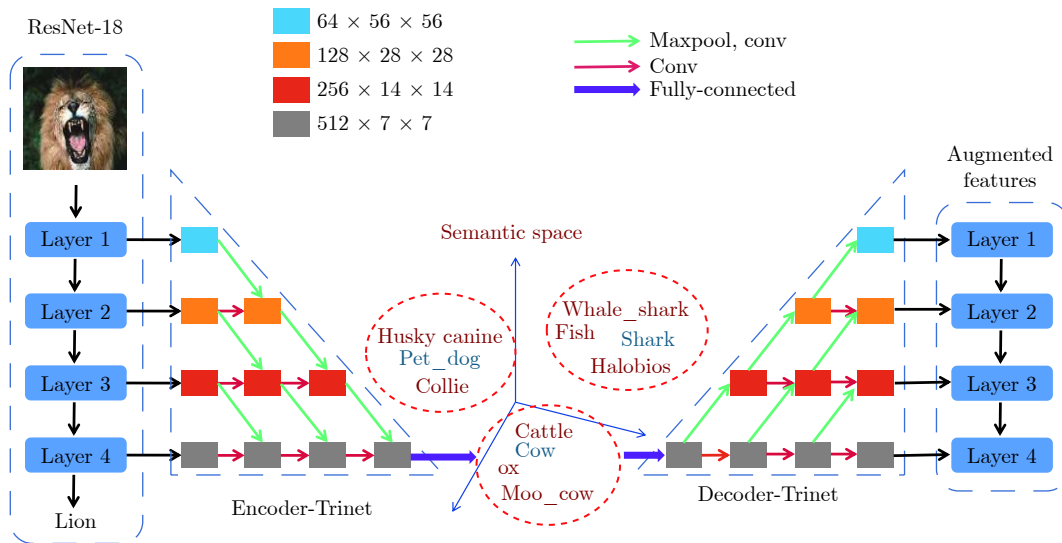


图 7 编码—解码机制^[31]

Fig.7 Coding-decoding mechanism^[31]

等^[31]在2018年提出语义特征增加的算法 (Semantic feature augmentation, SFA) ResNet-18 + 对偶 TriNet 网络, 利用编码 - 解码机制在图像特征和语义空间进行变换实现特征增加, 增加的特征能够丰富图像的语义多样性, 引入额外的语义信息. 使用 ResNet-18 网络对图像特征进行提取, 编码器 Encoder-Trinet 将提取的特征映射到语义空间, 在语义空间中, 对语义特征添加噪声进行高斯扰动, 假设语义空间中的特征值的微小变化允许在保持语义信息的同时形成潜在的类内变换, 然后采用解码器 Decoder-Trinet 将语义特征映射回多层的 ResNet-18 特征空间, 即实现特征增加. 该算法是一个端到端的网络框架, 能够在多层的图像特征空间和语义空间中学习映射关系, 适用于多种网络结构.

2.4 贝叶斯学习

深度学习基于大数据通过多层网络实现对抽象概念的理解, 显然, 数据量越多其效果越好, 假如没有那么多的大数据该如何进行抽象概念的理解. 对人类来说, 即便没有知识的积累, 没有相应的专业知识, 我们也能够照猫画虎, 这有点类似贝叶斯学习的方式^[95]. 贝叶斯学习是利用参数的先验分布, 由小样本信息得到的后验分布, 直接求出总体分布. 贝叶斯学习理论使用概率去表示所有形式的不确定性, 通过概率规则来实现学习和推理过程. 更具体的来说, 贝叶斯学习并不去求解最优的参数值 θ , 而是假设参数 θ 本身符合某个分布, 即先验概率 $P(\theta)$, 随后利用训练样本得到条件概率分布 $P(X|\theta)$, 根据贝叶斯公式我们便能求得样本的总体分布, 如式 (4),

$$P(\theta|X) = \frac{P(\theta)P(X|\theta)}{P(X)} \quad (4)$$

其中 $P(X)$ 为样本 X 服从的分布. 将贝叶斯学习与深度学习结合为贝叶斯深度学习, 此时网络的权重 W_i 和偏置 b 由确定的值变成某种分布. Lake 等在2011年提出层次贝叶斯程序学习 (Hierarchical Bayesian program learning, HBPL), 对观测像素进行结构解释, 解决了字符识别的问题, 但是其参数空间太大^[12], 因此, 又提出了一种基于组合和因果关系的层次贝叶斯模型^[96], 解决了对大数据集的依赖问题. 在2015年的贝叶斯框架中提出了一种新的计算模型, 用来模拟人类的学习能力^[52], 同时加入了元学习, 可以从现有的字符中抽象出其部件, 再根据不同部件的因果关系创造新的字符, 从而形成丰富的概念. 2018年 Kim 等^[97]将贝叶斯方法应用在与模型无关的元学习算法中, 将基于梯度的元学习与非参数变量的推理结合起来, 使用贝叶斯先验

防止元学习模型过拟合, 但也提升了模型的复杂度.

2.5 图神经网络模型

现实生活中的大量问题都可以被抽象成图模型^[98], 图 $G = (V, E)$ 作为一种数据结构, 由节点 V 和边 E 的集合组成, 能够表达复杂的数据关系. 传统的机器学习方法很难处理图神经网络信息, 充分挖掘图中蕴含的知识是一项非常具有挑战的任务. 在深度学习时代, 将图与深度学习进行融合成为了一项重要的工作. 本节所述的图神经网络 (Graph neural network, GNN) 模型是将 CNN 用于图神经网络上, 并对欧几里得小样本图像数据进行分类.

图神经网络在2005年首次被 Gori 等^[99]和 Scarselli 等^[100]提出, 用传统的方法处理图结构数据是将其转换为一组平面向量, 然而以这种方式处理数据, 重要的拓扑信息可能丢失, GNN 扩展了递归神经网络, 使有向图、无向图、循环图等得以被处理, 作为一种可训练的网络其中固定节点可被分别调整. Bruna 等^[101]和 Henaff 等^[102]提出学习图拉普拉斯的光滑谱乘子, 是将 CNN 泛化到非欧氏空间的一种尝试, 但是其计算代价非常高. Defferrard 等^[103]和 Kipf 等^[104]通过学习图拉普拉斯的多项式解决了计算代价的问题. Li 等^[105]和 Sukhbaatar 等^[106]进一步放宽模型的限制, 对网络层内的权重解耦, 同时提出门控机制进行非线性更新.

在小样本学习中, 为了将图神经网络应用于规则的欧几里得图像数据, Satorras 等^[107]在2018年提出了一个端对端的图卷积网络结构 (Graph convolutional network, GCN), 用于捕捉任务中具有不变性的特征. 图神经网络模型由输入图片的集合构成, 图神经网络中的节点与集合中的图片信息相关, 边是一个可训练的、用来度量相邻两个节点相似性的参数. 如图8所示, 将5个样本通过式 (5) 构建出图邻接矩阵, 接着通过图卷积得到节点的嵌入向量, 然后用式 (5) 依次更新图, 用图卷积更新节点嵌入, 这样就构成了一个深度的图卷积神经网络,

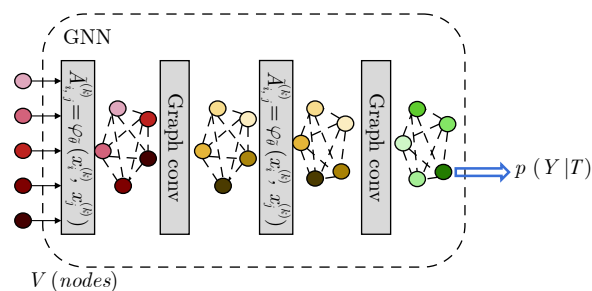


图8 图卷积神经网络^[107]

Fig.8 Graph convolution neural network^[107]

最后使用交叉熵损失函数计算图卷积神经网络的输出概率 p . $\varphi_{\tilde{\theta}}(\cdot)$ 的具体形式为式 (6), 使用 MLP 来计算图邻接矩阵 $\tilde{A}_{i,j}^{(k)}$. 其中, $x_i^{(k)}$ 和 $x_j^{(k)}$ 为输入样本. T 表示输入的学习任务.

$$\tilde{A}_{i,j}^{(k)} = \varphi_{\tilde{\theta}}(x_i^{(k)}, x_j^{(k)}) \quad (5)$$

$$\varphi_{\tilde{\theta}}(x_i^{(k)}, x_j^{(k)}) = MLP_{\tilde{\theta}}(abs(x_i^{(k)} - x_j^{(k)})) \quad (6)$$

文献 [107] 利用图节点的标签信息, 隐式地对类内的相似性和类间的不相似性进行建模, 与之相反, Kim 等 [108] 在 2019 年提出的基于边标签的图卷积神经网络 (Edge-labeling graph neural network, EGNN), 将数据集分为多个元任务, 包括支持集和查询集, 通过直接探索类内的相似性和类间的不相似性来迭代地更新边标签信息, 通过预测边标签对样本进行显式聚类. Liu 等 [109] 提出了一种转导式的传播网络 (Transductive propagation network,

TPN), 该算法利用元学习框架和流型假设, 通过对特征嵌入参数和图神经网络构建的参数进行联合学习, 将标签从标注样本传递到未标注样本, 提高了模型的泛化能力.

3 实验对比分析

上文描述的现有基于小样本学习的图像分类算法被归纳为卷积神经网络模型和图神经网络模型两大类, 具体如图 9 所示.

3.1 各种算法在公用数据集上的实验测试结果分析

1) 三种基于元学习的小样本图像分类算法各有优势, 此外训练模型时学习的类别越多, 类内样本越少, 分类效果越不好.

Omniglot 数据集是字符图像, 背景单一, 内容简单. 从表 2 中可以看出, 基于元学习的小样本算

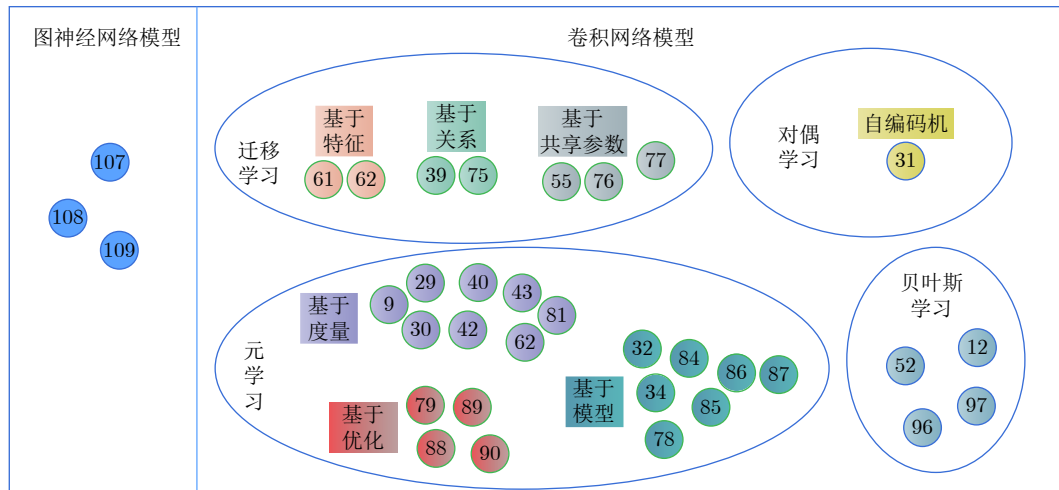


图 9 小样本图像分类算法概况

Fig. 9 Overview of small sample image classification algorithms

表 2 基于元学习的 Omniglot 实验结果

Table 2 Experimental results of Omniglot based on meta learning

		Omniglot			
		5way-1shot	5way-5shot	20way-1shot	20way-5shot
基于度量的元学习	MN ^[9]	98.12	99.63	94.40	98.78
	文献 [40]	90.80	96.70	77.00	91.00
	MMN ^[42]	99.28	99.77	97.16	98.93
	PN ^[29]	98.80	99.18	92.11	97.57
	RN ^[30]	99.48	99.60	97.67	98.97
基于模型的元学习	Meta-Nets ^[78]	98.00	99.60	96.90	98.50
基于优化的元学习	MAML ^[79]	98.79	99.48	93.43	95.33
	文献 [88]	—	—	97.65	99.33
	Reptile ^[89]	97.50	99.87	93.75	97.68

法在 Omniglot 数据集上的 N -way K -shot 分类结果非常好. 然而, 学习类别越多, 样本越少, 分类效果越不好, 因此 20way-1shot 的实验结果相对其他 N -way K -shot 分类结果较低.

小样本图像分类算法中, 基于度量的元学习算法在 Mini-ImageNet 数据集上学习到好的度量空间可提高分类效果. 如表 3 所示, 基于度量的元学习算法中, MMN 使用了记忆力机制, 加强了图像特征的表示能力, 可以学习到一个好的度量空间.

表 3 基于元学习的 Mini-ImageNet 实验结果

Table 3 Experimental results of Mini-ImageNet based on meta learning

		Mini-ImageNet	
		5way-1shot	5way-5shot
基于度量的元学习	MN ^[9]	44.38	57.78
	PN ^[29]	44.43	66.04
	RN ^[30]	50.13	64.33
	MMN ^[42]	53.37	66.97
基于模型的元学习	DML ^[32]	58.49	71.28
	AAN ^[87]	54.89	62.37
	MTL ^[84]	61.20	75.50
基于优化的元学习	MAML ^[79]	43.09	60.63
	Reptile ^[89]	48.21	66.00
	文献 [90]	43.44	60.00
	文献 [88]	52.15	68.32

小样本图像分类算法中, 基于模型的元学习算法通过学习丰富的图像语义特征帮助在 Mini-ImageNet 数据集上分类. 其中, DML 利用深度残差网络作为概念生成器, 可以构建表达能力更大的网络结构, 产生更好的语义特征.

小样本图像分类算法中, 基于优化的元学习算法具有快速学习的能力. 其与基于模型的元学习算法相比分类结果较差, 通过学习网络参数的初始化, 模型微调于新类时会更加适应, 该类算法能够快速对新样本进行训练, 其分类效果依赖于优化策略以及对新样本的适应.

2) 图卷积网络模型中, 对类内样本的相关性和

类间样本的不相关性建模, 在 Omniglot 和 Mini-ImageNet 数据集上能够产生更好的分类效果.

如表 4 所示, GCN、TPN 以及 EGNN 在 Omniglot 数据集上都取得了很好的分类精度, 在更为复杂的图像数据集 Mini-ImageNet 上, EGNN 的分类效果好于 GCN 和 TPN.

GCN 模型利用样本间的相关性建立模型, 但忽略了样本间存在的差异性. TPN 对模型的建立是利用了样本间的相关性和不相关性. EGNN 则利用卷积网络对样本间的相关性和不相关性进行学习, 进一步复杂化了模型, 同时增强了模型的非线性化, 提高了模型的表达能力.

3) 当小样本图像分类算法提取到丰富的高层语义特征或者在特征提取和分类器之间设计好的映射函数时, 能产生较好的分类效果.

表 5 所示, 分别从各类中挑选出的性能最好的算法进行比较, 在 Mini-ImageNet 数据集上各算法的 5way-1shot 分类精度接近于 60.0%, 5way-5shot 的分类精度均高于 70.0%, 其中 SFA 和 EGNN 达到了 76.0%, 这四种算法分别是迁移学习算法 PPA、元学习算法 DML、对偶学习算法 SFA、基于图卷积神经网络的算法 EGNN, 其中 PPA 算法通过激活函数来预测分类器中的分类参数, 相当于在高层语义特征和分类器之间做一个映射, 使分类器对于不同语义特征的选择更加精确; DML 算法利用深度残差网络提取到图像的高级语义特征; SFA 算法通过编码—解码机制, 对编码器映射到语义空间中的实例特征扰动, 再利用解码机产生丰富的图像特征. EGNN 算法对类内样本关系和类间样本关系进行建模, 能够对图像信息进行强有力的表示. 可以看出, 通过对图像的高层语义特征的利用, 提高了小样本图像分类的精度.

3.2 小样本学习算法在轮胎花纹数据集上的实验结果

为进一步分析现有小样本图像分类算法的表现, 本节实验在西安邮电大学图像与信息处理研究所依托与公安部门合作的平台所自建的轮胎花纹图像数据集^[110]上进行.

表 4 基于图卷积网络的 Mini-ImageNet、Omniglot 实验结果

Table 4 Experimental results of Mini-ImageNet and Omniglot based on graph convolutional network

	Omniglot				Mini-ImageNet	
	5way-1shot	5way-5shot	20way-1shot	20way-5shot	5way-1shot	5way-5shot
GCN ^[107]	99.26	99.72	97.66	99.10	53.03	64.78
TPN ^[109]	99.26	99.44	96.48	98.59	54.44	67.05
EGNN ^[108]	99.75	99.77	98.62	99.62	62.34	75.77

表 5 迁移学习、元学习、对偶学习和图神经网络模型实验结果

Table 5 Experimental results of transfer learning, meta learning, dual learning and graph neural network model

	Mini-ImageNet	
	5way-1shot	5way-5shot
迁移学习 PPA ^[55]	59.60	73.74
元学习 DML ^[32]	58.49	71.28
对偶学习 SFA ^[31]	57.95	76.64
图神经网络模型 EGNN ^[108]	62.34	75.77

轮胎花纹分类的研究源于交通事故及公安案件处理中轮胎花纹匹配的实际需求. 该数据集是目前公开用于学术研究的最大的轮胎花纹数据集, 包含轮胎表面花纹数据和轮胎压痕花纹数据各 80 类, 每类 30 张不同亮度不同尺度和不同旋转角度的图片, 如图 10 所示. 实验测试分别在表面花纹图像数据、压痕花纹图像数据、及两种图像混合数据上进行 (因为实际需求往往需要进行表面花纹和压痕花纹的比对). 实验中 46 类用于训练, 10 类用于验证, 13 类用于测试, 轮胎混合花纹数据集包含同样的类别, 不同的是每类 160 张图像.



图 10 轮胎花纹数据集样本示例

Fig. 10 Sample examples of tire patterns data sets

为研究基于元学习的小样本学习算法、通过编码—解码进行语义特征增强的小样本学习算法和基于图神经网络的小样本学习算法对轮胎花纹图像分类的效果, 分别对以下 5 个算法进行了实验: 基于优化的小样本元学习算法^[79], 基于模型的小样本元学习算法^[78], 基于度量的小样本元学习算法^[30], 基于编码—解码结构的小样本对偶学习算法^[31], 基于图神经网络的小样本学习算法^[107]. 表 6 为实验测试结果, 通过五组实验对比可以看出:

- 1) 通过编码—解码结构进行的语义特征增强能够提高分类精度.
- 2) 5 组实验的分类精度在混合数据集上均相对

表 6 在轮胎花纹数据集上的测试结果对比

Table 6 Test results comparison of various algorithms on tire patterns data set

算法	轮胎数据集	分类精度	
		5way-1shot	5way-5shot
文献 [79]	表面	67.09	85.55
	压痕	77.66	87.32
	混合	46.03	64.00
文献 [78]	表面	53.46	78.42
	压痕	66.13	80.45
	混合	42.80	63.53
文献 [107]	表面	77.46	89.52
	压痕	77.76	92.00
	混合	58.04	79.98
文献 [31]	表面	72.71	91.03
	压痕	76.42	91.76
	混合	51.84	81.02
文献 [30]	表面	63.97	81.60
	压痕	73.71	84.54
	混合	48.21	65.20

较低, 这是因为同一类轮胎混合花纹图像中包含了表面花纹和压痕花纹两种既相关又有差异的数据, 造成类间相似度降低, 从而分类任务难度增加.

相比其他算法, 基于图神经网络的小样本学习算法在轮胎花纹表面数据集和压痕数据集上的分类精度差异最小, 而且在混合花纹数据集上的分类精度最高. 这说明基于图神经网络的小样本学习算法适用于轮胎花纹数据集的分类研究. 下一步工作中, 我们将对比更多算法, 并进行更进一步的研究.

3.3 不同模型的小样本图像分类算法的讨论

针对第 3.1 节和第 3.2 节的实验分析结果, 本节进一步分析了各类算法之间的特点, 并分别对卷积神经网络模型和图神经网络模型进行讨论.

1) 对卷积神经网络模型的讨论.

如表 7 所示, 卷积神经网络模型中的迁移学习、元学习、对偶学习都可以使用数据增强的方式来解决小样本图像分类问题, 当增强的样本具有较大的数量和丰富的语义内容时, 小样本数据集的分类结果会有所提升.

对偶学习目前在小样本图像分类中的主要应用是数据增强, 不同于之前的图像变换方法, 对偶学习中可以利用自编码器在图像的视觉特征空间和语义特征空间之间相互变换, 它可以和现有的图像特征提取模型进行结合, 利用自编码器寻找好的数据增强方式.

表 7 小样本图像分类算法的对比
Table 7 Comparison of small sample image classification algorithms

算法	数据增强	训练策略	分类度量方式	数据集
基于特征的迁移学习	文献 [61] SSMN ^[62] 函数变换增加训练样本	CNN 表示学习阶段 + 小样本学习阶段	全连接层 + Softmax	ImageNet
		CNN + LSTM 局部特征度量 + 全局特征度量	嵌入向量 + 点乘	DiPART、PPM、Cross-DiPART-PPM
基于关系的迁移学习	文献 [39] 文献 [75] 采用伪样本数据	CNN 知识蒸馏	全连接层 + Softmax	MNIST
		CNN 1×1卷积核知识蒸馏	全连接层 + Softmax	CIFAR-10、CIFAR-100
基于共享参数的迁移学习	文献 [76] 文献 [77] PPA ^[59] 裁剪	CNN 预训练模型 + 微调	全连接层 + Softmax	PASCAL VOC 2007、PASCAL VOC 2012
		CNN 分类权重嵌入	全连接层 + Softmax	CUB-200
		CNN 在激活函数和 Softmax 之间建模, 预测类别的分类参数	全连接层 + Softmax	Mini-ImageNet
基于度量的元学习	文献 [43] MN ^[9] 仿射变换	CNN 孪生网络 + 距离度量	嵌入向量 + 欧氏距离	Omniglot
		CNN + LSTM 注意力模块 + 样本间匹配	嵌入向量 + 余弦距离	Omniglot、Mini-ImageNet
	MMN ^[42] PN ^[29] RN ^[30] 旋转	CNN + bi-LSTM 记忆读写控制模块 + 样本间匹配	嵌入向量 + 点乘	Omniglot、Mini-ImageNet
	CNN 聚类 + 样本间原型度量	嵌入向量 + 欧氏距离	Omniglot、Mini-ImageNet	
	CNN 不同样本在特征空间比较	全连接层 + Softmax	Omniglot、Mini-ImageNet	
基于优化的元学习	文献 [81] MAML ^[79] Reptile ^[80] 旋转	CNN 利用嵌入特征回归分类参数 + 不同样本映射到同一嵌入空间进行相似性度量	全连接层 + Softmax	Omniglot、Mini-ImageNet
		利用基于梯度的学习来更新每个元任务的参数		Omniglot、Mini-ImageNet
		将梯度下降计算的参数与初始化参数的差用于参数梯度更新		Omniglot、Mini-ImageNet
基于模型的元学习	文献 [90] Meta-Nets ^[78] DML ^[82] 旋转	利用 LSTM 模型学习优化算法		Mini-ImageNet
		CNN + LSTM记忆模块 + Meta learner + Base learner	全连接层 + Softmax	Omniglot、Mini-ImageNet
		CNN 概念生成器 + 概念判决器 + Meta learner	全连接层 + Softmax	CUB-200、CIFAR-100、Mini-ImageNet
基于模型的元学习	文献 [34] 文献 [85] 文献 [86] 函数变换增加训练样本	CNN + LSTM对样本和标签进行绑定编码使用外部记忆存储模块	全连接层 + Softmax	Omniglot
		CNN 利用数据增强提升元学习	全连接层 + Softmax	ImageNet
		CNN 减小大数据集和小数据集分类器间的差异	SVM	CUB-200
自动编码器	文献 [87] AAN ^[87] 对样本生成相应的分类参数	CNN 注意力模块 + 增量学习 + 元学习	全连接层 + Softmax	Mini-ImageNet、Tiered-ImageNet
		CNN 通过扰动语义空间特征实现样本特征增加	全连接层 + Softmax	CUB-200、CIFAR-100、Mini-ImageNet
图卷积神经网络	SFA ^[31] 使用编码 - 解码机制进行特征增加	GCN 利用图节点标签信息, 隐式地对类内和类间样本关系进行建模	全连接层 + Softmax	Omniglot、Mini-ImageNet
		GCN 通过预测边标签, 显式地对类内和类间样本进行建模	全连接层 + Softmax	Mini-ImageNet、Tiered-ImageNet
		GCN 流型假设 + 标签传播	全连接层 + Softmax	Mini-ImageNet、Tiered-ImageNet

相对于迁移学习、元学习和对偶学习, 贝叶斯学习目前在小样本学习中的应用较少, 可以更好地应用于训练数据量较少的情况, 但需要指定参数的先验分布, 而且对于样本的独立性要求较高, 但是现实生活中的样本和类别都具有一定的相关性, 因此建模方式存在偏差, 可将其与其他小样本图像分类方法相结合。

目前小样本图像分类中应用最多的是迁移学习和元学习, 两种方法都可以借助预训练模型来进一步学习, 或者借助迁移学习思想和元学习策略对小样本数据进行训练, 迁移学习更多侧重于得到表示

性更强的迁移特征, 元学习在度量方式、模型设计以及初始化策略上都有考量. 同时, 对传统机器学习分类器的使用也使得模型的解耦性增强, 更好地进行网络架构的设计, 其中基于欧氏距离、余弦距离和点乘方式度量在基于度量的元学习中使用较多.

2) 对图神经网络模型的讨论.

本文所述的图神经网络模型利用 CNN 对欧几里得图像数据进行特征提取, 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, 因此图神经网络对于样本间的复杂关系有更强的表示能力, 也有助于探索更多潜在于小样本数据集中的信息.

EGNN 相比较于 GCN 和 TPN 其模型的复杂度更高, 体现在对图中相邻节点关系的表示上, EGNN 不仅利用了类内样本间的相关性, 而且也对类间样本的不相关性进行建模, 再通过迭代不断地学习类内相关性和类间不相关性. 从表 4 中可以发现, 其在 Mini-ImageNet 数据集上的 5way-5shot 分类精度达到了 75.77%. 图神经网络对数据的建模方式与卷积神经网络模型不同, 它能够将在图像间的联系以图连接的形式呈现出来, 图中的边可以显式表达这种联系, 图模型构建以及图的更新方式目前还有待更多研究.

4 技术挑战与未来研究趋势

目前, 小样本图像分类算法在模式较为简单的字符型数据集 Omniglot 上已取得很好的分类结果, 但是对于相对复杂的数据集, 虽然分类结果不断提升, 但是仍然不理想. 利用数据增强、正则化、对特征提取过程建模等方式, 可以有效地缓解小样本带来的过拟合问题, 也能够增强图像特征的代表能力, 但仍然需要在克服过拟合问题和增强图像的代表能力之间进行权衡. 除此之外, 小样本图像分类仍然面临一些挑战, 本节将对此进行介绍, 同时从技术角度对小样本图像分类未来的研究趋势进行展望.

4.1 小样本图像分类面临的挑战

1) 权衡过拟合问题和图像特征表示能力

小样本图像分类模型往往需要克服过拟合问题, 同时又要从少量的样本中学习能够表示图像的有效特征. 迁移学习中对小样本数据集进行特征提取^[61], 元学习中从元任务中获取元信息^[29] 等都需要对图像特征进行提取, 为了缓解过拟合问题, 通常使用的网络结构较为简单, 不足以对图像中蕴含的信息进行有效表达, 而 Resnet 网络^[87] 和其他残差网络^[32] 能够加深网络的层数, 记忆模块能够对历史信息进行存取和使用^[34, 62, 78, 90], 从而增强了图像特征的代表能力. 因此, 如何权衡过拟合问题和图像特征表示能力是小样本图像分类需要面临的挑战.

2) 不同应用领域的小样本图像分类

从上述的实验分析中可以看出, 多数小样本图像分类算法, 在模式简单、背景单一的字符型数据集 Omniglot 上具有非常好的分类效果^[30, 79, 89], 在模式较为复杂的其他类型的数据集, 同一个小样本图像分类算法在不同的小样本数据集上的分类结果具有一定的差异^[31-32]. 针对不同应用领域图像数据内容的不同特点, 如何设计合适的小样本图像分类算

法, 或者具有一定普适性适应不同数据集的算法, 这也是小样本图像分类目前的难点.

4.2 小样本图像分类未来的研究方向

1) 应用注意力机制

小样本学习的训练样本量较少, 提取到的信息相对有限, 可以利用注意力机制在有限的训练样本下, 提取到对图像具有表示性更强的特征, 并且使得该特征能够显著影响分类效果. 小样本学习从本质上讲是想让机器学会人类的学习方式以及泛化能力, 人类能够在图像识别任务中很好地利用注意力机制, 此外, 注意力机制能够提高神经网络的可解释性^[111], 软注意力机制和硬注意力机制^[112]、自注意力机制^[113]、互注意力机制^[114] 等注意力模型, 其直观性、通用性以及可解释性能够对小样本图像分类任务提供重要帮助.

2) 将 CNN 中图像特征的标量表示替换为向量表示

CNN 利用卷积核能够检测出相应的图像特征, 但如果样本不够丰富, 一些重要信息就会检测不到, 比如位置等信息, 因此, CNN 需要更多的样本来增强它的性能, 提高图像特征的代表性. 胶囊网络通过向量来对图像特征进行表示, 向量中可以包含任意个值, 每个值代表当前需要识别的物体的一个特征, 而传统的卷积操作是通过线性加权求和的结果, 得到的是标量. 胶囊网络利用动态路由算法进行信息传递, 它需要较少的训练数据, 而且能够保留图像特征的位置和姿态信息, 对旋转、平移以及其他仿射变换也有很强的鲁棒性^[115].

5 结束语

本文针对当前基于小样本学习的图像分类算法进行了归类总结, 依据对不同数据类型的建模方式, 将小样本图像分类算法分为卷积神经网络模型和图神经网络模型两大类, 其中, 卷积神经网络模型又分为迁移学习、元学习、贝叶斯学习和对偶学习四种学习范式, 并针对数据集处理、特征提取和分类器设计三个环节, 对两类算法进行了详细介绍. 迁移学习更多侧重于得到表示性更强的迁移特征; 元学习在度量方式、模型设计以及初始化策略上都有考量; 贝叶斯方法目前难以单独应用于小样本图像分类; 对偶学习应用于小样本图像分类的是编码—解码结构, 可进行数据增强; 图神经网络可侧重于对图像间关系进行建模. 最后针对目前小样本图像分类算法的不足, 分析了小样本图像分类面临的挑战, 同时从技术角度探索了小样本图像分类的未来研究趋势.

References

- 1 Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, **60**(2): 91–110
- 2 Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, USA: IEEE, 2005. 886–893
- 3 Csurka G, Dance C R, Fan L X, Willamowski J, Bray C. Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision, Proceedings of the European Conference on Computer Vision. Prague, Heidelberg, Berlin: Springer 2004. 1–22
- 4 Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). New York, USA: IEEE, 2006. 2169–2178
- 5 Perronnin F, Sánchez J, Mensink T. Improving the fisher kernel for large-scale image classification. In: Proceedings of the 11th European Conference on Computer Vision. Heraklion, Greece: Springer, 2010. 143–156
- 6 Sivic J, Zisserman A. Video google: A text retrieval approach to object matching in videos. In: Proceedings of the 9th IEEE International Conference on Computer Vision. Nice, France: IEEE, 2003. 1470–1477
- 7 Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors. *Nature*, 1986, **323**(6088): 533–536
- 8 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, USA: ACM, 2012. 1106–1114
- 9 Vinyals O, Blundell C, Lillicrap T, Kavukcuoglu K. Matching networks for one shot learning. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: ACM, 2016. 3637–3645
- 10 Li F F, Fergus R, Perona P. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(4): 594–611
- 11 Li F F, Fergus R, Perona P. A bayesian approach to unsupervised one-shot learning of object categories. In: Proceedings of the 9th IEEE International Conference on Computer Vision. Nice, France: IEEE, 2003. 1134–1141
- 12 Lake B M, Salakhutdinov R, Gross J, Tenenbaum J B. One shot learning of simple visual concepts. *Cognitive Science Society*, 2011, **33**: 2568–2573
- 13 Hariharan B, Girshick R. Low-shot visual object recognition. arXiv: 1606.02819, 2016.
- 14 Shaban A, Bansal S, Liu Z, Essa I, Boots B. One-shot learning for semantic segmentation. In: Proceedings of the 2017 British Machine Vision Conference. London, UK: BMVC, 2017.
- 15 Gui L K, Wang Y X, Hebert M. Few-shot hash learning for image retrieval. In: Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW). Venice, Italy: IEEE, 2017. 1228–1237
- 16 Qi G J, Luo J B. Small data challenges in big data era: A survey of recent progress on unsupervised and semi-supervised methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, DOI: [10.1109/TPAMI.2020.3031898](https://doi.org/10.1109/TPAMI.2020.3031898)
- 17 Shu J, Xu Z B, Meng D Y. Small sample learning in big data era. arXiv: 1808.04572, 2018.
- 18 Wang Y Q, Yao Q M, Kwok J, Ni L M. Generalizing from a few examples: A survey on few-shot learning. arXiv: 1904.05046, 2019.
- 19 Wang Y Q, Yao Q M. Few-shot learning: A survey. arXiv: 1904.05046, 2019.
- 20 Mensink T, Verbeek J, Perronnin F, Csurka G. Metric learning for large scale image classification: Generalizing to new classes at near-zero cost. In: Proceedings of the 12th European Conference on Computer Vision. Florence, Italy: Springer, 2012. 488–501
- 21 Mensink T, Verbeek J, Perronnin F, Csurka G. Distance-based image classification: Generalizing to new classes at near-zero cost. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(11): 2624–2637
- 22 Rohrbach M, Ebert S, Schiele B. Transfer learning in a transductive setting. In: Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, USA: ACM, 2013. 46–54
- 23 Lampert C H, Nickisch H, Harmeling S. Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(3): 453–465
- 24 Deng J, Dong W, Socher R, Li L J, Li K, Li F F. Imagenet: A large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA: IEEE, 2009.
- 25 Rohrbach M, Stark M, Schiele B. Evaluating knowledge transfer and zero-shot learning in a large-scale setting. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2011. 1641–1648
- 26 Fu Y W, Hospedales T M, Xiang T, Gong S G. Transductive multi-view zero-shot learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(11): 2332–2345
- 27 Wah C, Branson S, Welinder P, Perona P, Belongie S. The caltech-ucsd birds-200-2011 dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011.
- 28 Fu Y W, Hospedales T M, Xiang T, Gong S G. Learning multimodal latent attributes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(2): 303–316
- 29 Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: ACM, 2017. 4077–4087
- 30 Sung F, Yang Y X, Zhang L, Xiang T, Torr P H S, Hospedales T M. Learning to compare: Relation network for few-shot learning. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 1199–1208
- 31 Chen Z T, Fu Y W, Zhang Y D, Jiang Y G, Xue X Y, Sigal L. Semantic feature augmentation in few-shot learning. In: Proceedings of the European Conference on Computer Vision. Prague, Heidelberg, Berlin: Springer, 2018.
- 32 Zhou F W, Wu B, Li Z G. Deep meta-learning: Learning to learn in the concept space. arXiv: 1802.03596, 2018.
- 33 Gan Ying-Jun, Zhai Yi-Kui, Huang Yu, Zeng Jun-Ying, Jiang Kai-Yong. Research of facial beauty prediction based on deep convolutional features using double activation layer. *Acta Electronica Sinica*, 2019, **47**(3): 636–642
(甘英俊, 翟懿奎, 黄聿, 增军英, 姜开永. 基于双激活层深度卷积特征的人脸美丽预测研究. *电子学报*, 2019, **47**(3): 636–642)
- 34 Santoro A, Bartunov S, Botvinick M, Wierstra D, Lillicrap T. Meta-learning with memory-augmented neural networks. In: Proceedings of the 33rd International Conference on Machine Learning. New York, USA: ACM, 2016. 1842–1850
- 35 Jia Y Q, Darrell T. Latent task adaptation with large-scale hierarchies. In: Proceedings of the 2013 IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013. 2080–2087
- 36 Mehrotra A, Dukkipati A. Generative adversarial residual pairwise networks for one shot learning. arXiv: 1703.08033, 2017.

- 37 Bromley J, Guyon I, LeCun Y, Säckinger E, Shah R. Signature verification using a “Siamese” time delay neural network. In: Proceedings of the 6th International Conference on Neural Information Processing Systems. Denver, USA: ACM, 1993. 737–744
- 38 Dixit M, Kwitt R, Niethammer M, Vasconcelos N. AGA: Attribute-guided augmentation. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 7455–7463
- 39 Kimura A, Ghahramani Z, Takeuchi K, Iwata T, Ueda N. Few-shot learning of neural networks from scratch by pseudo example optimization. In: Proceedings of the 2018 British Machine Vision Conference. Newcastle, UK: British Machine Vision Association, 2018.
- 40 Bartunov S, Vetrov D. Few-shot generative modelling with generative matching networks. In: Proceedings of the 21st International Conference on Artificial Intelligence and Statistics. Playa Blanca, Spain: JMLR, 2018. 670–678
- 41 Wang P, Liu L Q, Shen C H, Huang Z, Van Den Hengel A, Shen H T. Multi-attention network for one shot learning. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 2721–2729
- 42 Cai Q, Pan Y W, Yao T, Yan C G, Mei T. Memory matching networks for one-shot image recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 4080–4088
- 43 Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. In: Proceedings of the 32nd International Conference on Machine Learning. Lille, France: International Machine Learning Society, 2015.
- 44 Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 815–823
- 45 Taigman Y, Yang M, Ranzato M A, Wolf L. Web-scale training for face identification. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 2746–2754
- 46 Ma Yue-Feng, Liang Xun, Zhou Xiao-Ping. A fast sparse algorithm for least squares support vector machine based on global representative points. *Acta Automatica Sinica*, 2017, **43**(1): 132–141
(马跃峰, 梁循, 周小平. 一种基于全局代表点的快速最小二乘支持向量机稀疏化算法. *自动化学报*, 2017, **43**(1): 132–141)
- 47 Goldberger J, Roweis S, Hinton G, Salakhutdinov R. Neighbourhood components analysis. In: Proceedings of the 17th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2004. 513–520
- 48 Salakhutdinov R, Hinton G. Learning a nonlinear embedding by preserving class neighbourhood structure. In: Proceedings of the 11th International Conference on Artificial Intelligence and Statistics. Naha, Japan: JMLR, 2007. 412–419
- 49 Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification. *The Journal of Machine Learning Research*, 2009, **10**: 207–244
- 50 Gidaris S, Komodakis N. Dynamic few-shot visual learning without forgetting. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 4367–4375
- 51 Chen W Y, Liu Y C, Kira Z, Wang Y C F, Huang J B. A closer look at few-shot classification. In: Proceedings of the 7th International Conference on Learning Representations. New Orleans, Louisiana, United States: arXiv.org, 2019.
- 52 Lake B M, Salakhutdinov R, Tenenbaum J B. Human-level concept learning through probabilistic program induction. *Science*, 2015, **350**(6266): 1332–1338
- 53 Krizhevsky A. Learning multiple layers of features from tiny images. Technical Report, University of Toronto, Canada, 2009.
- 54 Ren M Y, Triantafillou E, Ravi S, Snell J, Swersky K, Tenenbaum J B, et al. Meta-learning for semi-supervised few-shot classification. In: Proceedings of the 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings. Vancouver, Canada: arXiv.org, 2018.
- 55 Qiao S Y, Liu C X, Shen W, Yuille A L. Few-shot image recognition by predicting parameters from activations. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 7229–7238
- 56 Aytar Y, Zisserman A. Tabula rasa: Model transfer for object category detection. In: Proceedings of the 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011. 2252–2259
- 57 Tommasi T, Orabona F, Caputo B. Safety in numbers: Learning categories from few examples with multi model knowledge transfer. In: Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 3081–3088
- 58 Thrun S. Is learning the n-th thing any easier than learning the first? In: Proceedings of the 8th International Conference on Neural Information Processing Systems. Denver, USA: MIT Press, 1996. 640–646
- 59 Pan S J, Yang Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010, **22**(10): 1345–1359
- 60 Ji Ding-Cheng, Jiang Yi-Zhang, Wang Shi-Tong. Multi-source transfer learning method by balancing both the domains and instances. *Acta Electronica Sinica*, 2019, **47**(3): 692–699
(季鼎承, 蒋亦樟, 王士同. 基于域与样例平衡的多源迁移学习方法. *电子学报*, 2019, **47**(3): 692–699)
- 61 Hariharan B, Girshick R. Low-shot visual recognition by shrinking and hallucinating features. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 3018–3027
- 62 Choi J, Krishnamurthy J, Kembhavi A, Farhadi A. Structured set matching networks for one-shot part labeling. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake, USA: IEEE, 2018. 3627–3636
- 63 Bucilua C, Caruana R, Niculescu-Mizil A. Model compression. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM, 2006. 535–541
- 64 Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. *Computer ence*, 2015, **14**(7):38–39
- 65 Ba L J, Caruana R. Do deep nets really need to be deep? In: Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2014. 2654–2662
- 66 Han S, Mao H Z, Dally W J. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. In: Proceedings of the 2016 International Conference on Learning Representations. Barcelona, Spain: MIT Press, 2016. 2654–2662
- 67 Li H, Kadav A, Durdanovic I, Samet H, Graf H P. Pruning filters for efficient ConvNets. In: Proceedings of the 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings. Toulon, France: arXiv.org, 2017.
- 68 Liu Z, Li J G, Shen Z Q, Huang G, Yan S M, Zhang C S. Learning efficient convolutional networks through network slimming. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 2736–2744
- 69 Denton E, Zaremba W, Bruna J, LeCun Y, Fergus R. Exploiting linear structure within convolutional networks for efficient

- evaluation. In: Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2014. 1269–1277
- 70 Jaderberg M, Vedaldi A, Zisserman A. Speeding up convolutional neural networks with low rank expansions. In: Proceedings of the 2014 Conference on British Machine Vision Conference. Nottingham, UK: British Machine Vision Association, 2014.
- 71 Zhang X Y, Zou J H, He K M, Sun J. Accelerating very deep convolutional networks for classification and detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(10): 1943–1955
- 72 Kim Y D, Park E, Yoo S, Choi T, Yang L, Shin D. Compression of deep convolutional neural networks for fast and low power mobile applications. In: Proceedings of the 4th International Conference on Learning Representations. San Juan, Puerto Rico: IEEE, 2016.
- 73 Snellson E, Ghahramani Z. Sparse Gaussian processes using pseudo-inputs. In: Proceedings of the 18th International Conference on Neural Information Processing Systems. Vancouver, Canada: MIT Press, 2005. 1257–1264
- 74 Lin M, Chen Q, Yan S C. Network in network. In: Proceedings of the 2nd International Conference on Learning Representations. Banff, AB, Canada: arXiv.org, 2014.
- 75 Li T H, Li J G, Liu Z, Zhang C S. Knowledge distillation from few samples. arXiv: 1812.01839, 2018.
- 76 Oquab M, Bottou L, Laptev I, Sivic J. Learning and transferring mid-level image representations using convolutional neural networks. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014. 1717–1724
- 77 Qi H, Brown M, Lowe D G. Low-shot learning with imprinted weights. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 5822–5830
- 78 Munkhdalai T, Yu H. Meta networks. In: Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: ACM, 2017. 2554–2563
- 79 Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: ACM, 2017. 1126–1135
- 80 Hochreiter S, Schmidhuber J. Long Short-term memory. *Neural Computation*, 1997, **9**(8): 1735–1780
- 81 Zhou L J, Cui P, Yang S Q, Zhu W W, Tian Q. Learning to learn image classifiers with informative visual analogy. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019.
- 82 Mitchell T M, Thrun S. Explanation-based neural network learning for robot control. In: Advances in Neural Information Processing Systems 5. Denver, USA: ACM, 1992. 287–294
- 83 Vilalta R, Drissi Y. A perspective view and survey of meta-learning. *Artificial Intelligence Review*, 2002, **18**(2): 77–95
- 84 Sun Q R, Liu Y Y, Chua T S, Schiele B. Meta-transfer learning for few-shot learning. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 403–412
- 85 Wang Y X, Girshick R, Hebert M, Hariharan B. Low-shot learning from imaginary data. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 7278–7286
- 86 Wang Y X, Hebert M. Learning to learn: Model regression networks for easy small sample learning. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016. 616–634
- 87 Ren M Y, Liao R J, Fetaya E, Zemel R S. Incremental few-Shot learning with attention attractor networks. In: Advances in Neural Information Processing Systems 32. Vancouver, Canada: MIT Press, 2019.
- 88 Antoniou A, Edwards H, Storkey A. Antoniou A, Edwards H, Storkey A. How to train your MAML. In: Proceedings of the 7th International Conference on Learning Representations. New Orleans, USA: arXiv.org, 2019.
- 89 Nichol A, Schulman J. Reptile: A scalable metalearning algorithm. arXiv: 1803.02999, 2018.
- 90 Ravi S, Larochelle H. Optimization as a model for few-shot learning. In: Proceedings of the 5th International Conference on Learning Representations. Toulon, France: IEEE, 2017.
- 91 Qin Tao. Dual learning: A new paradigm of machine learning [Online], available: https://mp.weixin.qq.com/s/gwzy-Qu9YzxVyLT-L_mOicw, December 7, 2016.
- 92 Vincent P, Larochelle H, Bengio Y, Manzagol P A. Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine Learning. Helsinki, Finland: ACM, 2008. 1096–1103
- 93 Masci J, Meier U, Cireřan D, Schmidhuber J. Stacked convolutional auto-encoders for hierarchical feature extraction. In: Proceedings of the 21st International Conference on Artificial Neural Networks. Espoo, Finland: Springer, 2011. 52–59
- 94 Kingma D P, Welling M. Auto-encoding variational bayes. In: Proceedings of the 2nd International Conference on Learning Representations. Banff, Canada: arXiv.org, 2014.
- 95 Chen Xiao-Liang, Human like concept learning of small sample and deep reinforcement learning of big data [Online], available: <http://blog.sciencenet.cn/blog-1375795-964076.html>, May 22, 2016.
- 96 Lake B M, Salakhutdinov R, Tenenbaum J B. One-shot learning by inverting a compositional causal process. In: Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, Spain: ACM, 2013. 2526–2534
- 97 Kim T, Yoon J, Dia O, Kim S, Bengio Y, Ahn S. Bayesian model-agnostic meta-learning. In: Advances in Neural Information Processing Systems 31. Montréal, Canada: MIT Press, 2018.
- 98 Mao Guo-Yong, Zhang Ning. Building connected undirected simple graph from degree of vertex. *Computer Engineering and Applications*, 2011, **47**(29): 40–41, 106 (毛国勇, 张宁. 从节点度数生成无向简单连通图. 计算机工程与应用, 2011, **47**(29): 40–41, 106)
- 99 Gori M, Monfardini G, Scarselli F. A new model for learning in graph domains. In: Proceedings of the 2005 IEEE International Joint Conference on Neural Networks. Montreal, Canada: IEEE, 2005. 729–734
- 100 Scarselli F, Gori M, Tsoi A C, Hagenbuchner M, Monfardini G. The graph neural network model. *IEEE Transactions on Neural Networks*, 2009, **20**(1): 61–80
- 101 Bruna J, Zaremba W, Szlam A, LeCun Y. Spectral networks and locally connected networks on graphs. In: Proceedings of the 2nd International Conference on Learning Representations. Banff, Canada: arXiv.org, 2014.
- 102 Henaff M, Bruna J, LeCun Y. Deep convolutional networks on graph-structured data. arXiv: 1506.05163, 2015.
- 103 Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: ACM, 2016. 3837–3845
- 104 Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. In: Proceedings of the 5th International Conference on Learning Representations. Toulon, France: arXiv.org, 2017.
- 105 Li Y J, Tarlow D, Brockschmidt M, Zemel R S. Gated graph sequence neural networks. In: Proceedings of the 4th International

al Conference on Learning Representations. San Juan, Puerto Rico: IEEE, 2016.

- 106 Sukhbaatar S, Szlam A, Fergus R. Learning multiagent communication with backpropagation. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: ACM, 2016. 2244–2252
- 107 Satorras V G, Bruna J. Few-shot learning with graph neural networks. In: Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada: IEEE, 2018.
- 108 Kim J, Kim T, Kim S, Yoo C D. Edge-labeling graph neural network for few-shot learning. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 114–20
- 109 Liu Y B, Lee J, Park M, Kim S, Yang E, Hwang S J, Yang Y. Learning to propagate labels: Transductive propagation network for few-shot learning. In: Proceedings of the 7th International Conference on Learning Representations. New Orleans, LA, USA: IEEE, 2019.
- 110 Key Laboratory of Electronic Information Processing for Applications in Crime Scene Investigation, Ministry of Public Security. The pattern image database of image processing team of Xi'an University of Posts and Telecommunications [Online], available : <http://www.xuptciip.com.cn/show.html?database-lthhhw>, January 1, 2019.
- 111 Li J W, Monroe W, Jurafsky D. Understanding neural networks through representation erasure. In: Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: ACM, 2017.
- 112 Xu K, Ba J L, Kiros R, Cho K, Courville A, Salakhutdinov R, et al. Show, attend and tell: Neural image caption generation with visual attention. In: Proceedings of the 32nd International Conference on International Conference on Machine Learning. Lille, France: ACM, 2015. 2048–2057
- 113 Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing systems. Long Beach, USA: ACM, 2017. 5998–6008
- 114 Hou R B, Chang H, Ma B P, Shan S G, Chen X L. Cross attention network for few-shot classification. In: Advances in Neural Information Processing Systems 32. Vancouver, Canada: MIT Press, 2019. 4005–4016
- 115 Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules. In: Advances in Neural Information Processing Systems 30. Long Beach, USA: MIT Press, 2017. 3856–3866



刘颖 西安邮电大学通信与信息工程学院教授. 主要研究方向为图像视频检索. 本文通信作者.

E-mail: liuying_ciip@163.com

(LIU Ying Professor at the School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications. Her research interest covers image and video retrieval. Corresponding author of this paper.)



雷研博 西安邮电大学通信与信息工程学院硕士研究生. 主要研究方向为图像检索.

E-mail: 15829051768@163.com

(LEI Yan-Bo Master student at the School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications. His main research interest is image retrieval.)



范九伦 西安邮电大学校长. 主要研究方向为模式识别与图像处理.

E-mail: jiulunf@xupt.edu.cn

(FAN Jiu-Lun President of Xi'an University of Posts and Telecommunications. His main research interest covers pattern recognition and image processing.)



王富平 西安邮电大学通信与信息工程学院讲师. 主要研究方向为图像特征提取.

E-mail: wangfuping@xupt.edu.cn

(WANG Fu-Ping Lecturer at the School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications. His main research interest is image feature extraction.)



公衍超 西安邮电大学通信与信息工程学院讲师. 主要研究方向为视频编码, 码率控制.

E-mail: gongyanchao@xupt.edu.cn

(GONG Yan-Chao Lecturer at the School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications. His research interest covers video coding and rate control.)



田奇 华为云首席科学家. 研究方向为多媒体检索, 计算机视觉, 机器学习和模式识别.

E-mail: qitian@cs.utsa.edu

(TIAN Qi Chief scientist of Huawei Cloud. His research interest covers multimedia retrieval, computer vision, machine learning and pattern recognition.)