软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2022,33(1):193−210 [doi: 10.13328/j.cnki.jos.006342]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

小样本困境下的深度学习图像识别综述*

葛轶洲[1,2], 刘 恒[3], 王 言[3], 徐百乐[3], 周 青[1,2], 申富饶[3]

[1](通信信息控制和安全技术重点实验室, 浙江 嘉兴 314033)

[2](中国电子科技集团公司第三十六研究所, 浙江 嘉兴 314033)

[3](计算机软件新技术国家重点实验室 (南京大学), 江苏 南京 210023)

通信作者: 刘恒, E-mail: mg1937016@smail.nju.edu.cn

摘 要: 图像识别是图像研究领域的核心问题, 解决图像识别问题对人脸识别、自动驾驶、机器人等各领域研究都有重要意义. 目前广泛使用的基于深度神经网络的机器学习方法, 已经在鸟类分类、人脸识别、日常物品分类等图像识别数据集上达到了超过人类的水平, 同时越来越多的工业界应用开始考虑基于深度神经网络的方法, 以完成一系列图像识别业务. 但是深度学习方法极度依赖大规模标注数据, 这一缺陷极大地限制了深度学习方法在实际图像识别任务中的应用. 针对这一问题, 越来越多的研究者开始研究如何基于少量的图像识别标注样本来训练识别模型. 为了更好地理解基于少量标注样本的图像识别问题, 广泛地讨论了几种图像识别领域主流的少量标注学习方法, 包括基于数据增强的方法、基于迁移学习的方法以及基于元学习的方法, 通过讨论不同算法的流程以及核心思想, 可以清晰地看到现有方法在解决少量标注的图像识别问题上的优点和不足. 最后针对现有方法的局限性, 指出了小样本图像识别未来的研究方向.

关键词: 图像识别; 深度学习; 小样本学习; 数据增强; 迁移学习; 元学习

中图法分类号: TP391

中文引用格式: 葛轶洲, 刘恒, 王言, 徐百乐, 周青, 申富饶. 小样本困境下的深度学习图像识别综述. 软件学报, 2022, 33(1): 193–210. http://www.jos.org.cn/1000-9825/6342.htm

英文引用格式: Ge YZ, Liu H, Wang Y, Xu BL, Zhou Q, Shen FR. Survey on Deep Learning Image Recognition in Dilemma of Small Samples. Ruan Jian Xue Bao/Journal of Software, 2022, 33(1): 193–210 (in Chinese). http://www.jos.org.cn/1000-9825/6342.htm

Survey on Deep Learning Image Recognition in Dilemma of Small Samples

GE Yi-Zhou[1,2], LIU Heng[3], WANG Yan[3], XU Bai-Le[3], ZHOU Qing[1,2], SHEN Fu-Rao[3]

[1](Science and Technology on Communication Information Security Control Laboratory, Jiaxing 314033, China)

[2](The 36th Research Institute of China Electronics Technology Group Corporation, Jiaxing 314033, China)

[3](State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210023, China)

Abstract: Present machine learning methods have reached a higher level than human intelligence in image recognition and other tasks. However, recent machine learning methods, especially deep learning methods, rely heavily on a large number of annotation data that human cognition often does not need. This weakness greatly limits the application of deep learning method in practical problem. To solve this problem, learning from a few shot examples attracts more and more community's research interest. In order to better understand the few shot learning problem, this study extensively discusses several popular few shot learning methods, including data augmentation methods, transfer learning methods, and meta learning methods. After the processes and core ingredients of different algorithms are discussed, the advantages and disadvantages of existing methods could be clearly seen in solving few shot learning problems. At the end of this paper, the points to future research directions are highlighted in the field of few shot learning problem.

Key words: image recognition; deep learning; few shot learning; data augmentation; transfer learning; meta learning

* 基金项目: 国家自然科学基金 (61876076)

收稿时间: 2021-01-13; 修改时间: 2021-02-21; 采用时间: 2021-03-30; jos 在线出版时间: 2021-04-20

现在的机器学习方法, 尤其是基于深度神经网络的机器学习方法已经在人脸识别[1]、自动驾驶[2]、机器人[3]等图像识别相关领域取得了巨大的成就, 有的甚至已经超过人类目前的识别水平. 然而在深度学习取得巨大成就的同时, 人们发现把其应用到实际问题中却困难重重. 首先是标注数据的问题, 目前的深度学习方法需要大量的标注数据来进行训练[4], 但是实际应用中数据获取往往是困难的, 这之中既有个人隐私的问题, 比如人脸数据, 也有问题对象本身就很少的问题, 比如识别珍稀保护动物的问题, 除此之外, 数据标注工作往往需要耗费大量人力物力, 从而阻碍了深度学习技术在图像识别领域的落地. 其次是算力问题, 深度学习方法在提高算法性能的同时, 往往伴随着庞大的网络运算, 这也就使得深度学习的方法很难部署在计算资源受限的设备上, 因此在一些算力受限的应用场景, 比如自动驾驶、机器人、道路监控等问题中, 图像识别任务目前大多还是使用一些低智能化、低算力消耗的技术完成的, 这同样严重阻碍了智能化图像识别技术的发展.

与之相反, 人类的识别却是相对轻量的, 即并不需要收集大量的数据来进行学习, 更不需要长时间的思考或者计算[5]. 比如父母教新生婴儿识字, 分辨动物, 只需要简单地在家里贴上一两幅相应的字画即可, 小孩很快就会认识上面的内容. 如何在保留现在的深度学习方法强大的知识表示能力的同时, 使其可以快速从少量样本中学习到有用的知识, 这种基于小样本的图像识别问题已经逐渐引起了人们的注意.

本文将按照下面的顺序来展开讨论, 首先在第 1 节介绍小样本图像识别的问题描述, 然后会在第 2 节介绍基于数据增强的小样本学习算法, 在第 3 部分介绍基于迁移学习的算法, 在第 4 节介绍基于元学习的算法, 会在第 5 节介绍现在广泛使用的小样本图像识别问题评价指标, 并对比上面介绍的算法在该问题基准上的性能, 最后会在第 6 部分指出现有算法的不足以及未来的发展方向.

1 小样本学习简介

小样本图像识别任务需要机器学习模型在少量标注数据上进行训练和学习, 目前经常研究的问题为 N-way K-shot 形式, 即问题包括 N 种数据, 每种数据只包含 K 个标注样本[6]. 现有的小样本图像识别问题可以看做是基于深度迁移学习的图像识别问题, 这里我们把上面提到的少量标注数据称作目标数据域, 后续的识别任务都是基于目标数据所包含的类别进行的; 然后为了辅助模型的训练, 通常会引入一个和目标数据域类别互斥的辅助数据集, 和目标数据域的少量标注相反, 辅助数据集的标注样本更加丰富, 类别也更加多.

解决 N-way K-shot 形式的小样本图像识别任务, 大多数方法会从辅助数据集学习先验知识, 然后在标注有限的目标数据域上利用这些先验知识完成学习和预测任务. 在下面的章节我们会详细讨论如何基于辅助数据集来学习先验知识, 以及如何利用这些先验知识来在小样本图像识别问题上完成学习和预测.

2 基于数据增强的小样本图像识别方法

小样本图像识别任务的核心问题是标注数据不足, 所以通过算法生成人工标注数据, 来扩充原有的数据量是一种非常直观的方法[7]. 在小样本图像识别任务领域, 目前常用的数据增强方法基本上都是利用少量的标注数据来生成更多的伪数据, 比如人工合成图像, 同时需要给这些伪数据打上标签, 然后作为标注数据来辅助训练, 本质上和迁移学习的方法是异曲同工的[8]. 按照伪数据的使用方式, 可以将其划分为两种类型: 一种是使用伪数据来填补标注不足的小样本数据, 另外一种是使用伪数据来显式地锐化分类算法学习到的决策边界. 下面就这两种方法以及对应的具体算法展开讨论.

2.1 伪数据补充小样本数据

通过生成伪数据来填充标注不足的小样本数据, 是数据增强最常见的思路. 在传统的图像处理任务中, 裁剪、旋转、锐化等方法经常用来提升图像样本的多样性, 这些简单的图像增强方法可以有效地避免模型过拟合, 提升算法性能. 但是这些简单的图像增强的方法并不能有效地改善小样本学习任务的识别性能, 其中最主要的原因就是传统的数据增强方法不能很好地帮助模型度量新类 (少量标注类别) 的类内差异[9]. 比如训练数据中包含了鸟这一类别, 但是提供的数据过少, 大多都是鸟站立在枝头的图片, 那么在测试的时候, 对于飞翔在天空的鸟的图片, 模

型是很难正确分类的. 为了解决这样的问题, 一个直观的方法就是学习标注充分的同一类别数据之间的模式, 然后把这种模式应用在少量标注样本上, 产生可以较好地刻画该类类内差异的伪数据.

2.1.1 delta-encoder

文献 [10] 提出了一种基于自动编码器结构的数据增强方法, 算法框图如图 1 所示.

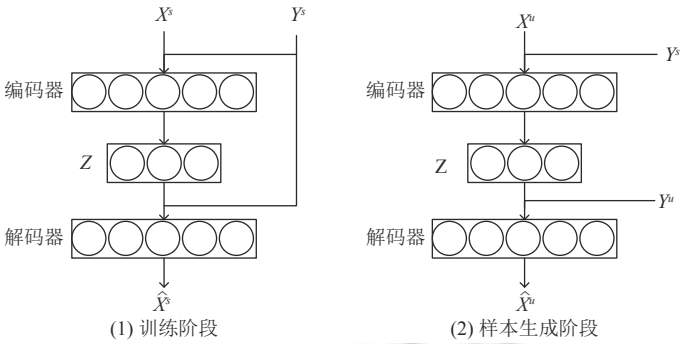


图 1 Delta-encoder 算法框图[10]

图 1 中样本 X^s 和 Y^s 来自标注充分的基类数据: X^s 是标注充分的基类数据的输入样本, Y^s 是对应的分类标签, 样本 Y^u 是只有少量标注的新类数据的分类标签,

算法的基本流程是:

(1) 在模型的训练阶段, 使用采样样本 X^s 和对应的标签 Y^s 来训练自编码网络模型, 这和之前的自编码网络模型不同之处是, 该模型解除了对中间特征 Z 的维度的限制, 使得在解码阶段, 自编码器会更多地“依赖”解码阶段的辅助数据 Y^s, 从而使得样本生成阶段, 生成的样本 $\hat{X}^s$ 即编码了同一类别的样本 X^s 和 Y^s 的类内差异, 同时也编码了新类数据 Y^u 的信息.

(2) 在模型的生成阶段, 将采样数据 X^s 和新类的分类标签 Y^u 输入训练好的自编码器, 此时自编码器利用给定的分类标签 Y^u 的信息, 让输入的原始数据 X^s 转变为具有新类数据特征的生成样本 $\hat{X}^u$, 从而完成数据增强的任务.

(3) 在模型的测试阶段, 我们经过了前面的编码器训练和伪数据生成阶段, 在测试的时候, 我们将生成的用于数据增强的数据和原始的小样本数据一块输入算法模型, 对原有的识别模型进行微调, 使其能够是识别新类别的数据.

文献 [10] 的主要贡献是提出了一种基于自编码器结构的特征差异编码器, 该模块在编码阶段学习不同类别样本之间的差异, 然后在解码阶段, 通过类内差异和源样本共同作用, 生成可以更好地刻画源样本所属类别的伪数据, 从而帮助优化了小样本学习问题.

2.1.2 dual TriNet

文献 [11] 同样采用了自编码的结构来进行数据增强, 但是和这之前提出的方法不同, 作者指出了基于实例空间进行数据增强的一些缺点, 并提出了基于语义空间的数据增强. 算法框图如图 2 所示.

下面详细介绍一下 TriNet 的算法流程, 即 TriNet 模型如何训练以及生成用于数据增强的伪数据.

(1) 首先通过 ResNet-18 网络模型提取输入图片的多级深度特征, 深度特征从浅层到深层, 分别对应了左边蓝色虚线框中的 Layer1、Layer2、Layer3 和 Layer4;

(2) 然后将上一步提到到的多级深度特征输入 TriNet 的编码器, 即图 2 中左边的虚线三角形部分, 该部分主要包含了 3 种基本操作: 绿色箭头表示串联的卷积操作和最大池化操作, 红色箭头表示卷积操作, 紫色的箭头表示全连接操作, 即矩阵变换操作. TriNet 通过上面 3 种基本操作, 将多级深度特征映射到语义空间;

(3) 通过 TriNet 编码器映射得到的语义空间是和标签的语义空间对齐的, 即狮子类 (Lion) 数据和猫类 (Cat) 数据的标签语义是相似的 (这里的标签语义可以通过预训练的自然语言处理词向量模型得到), 那么 TriNet 编码器

映射得到的狮子类语义特征在语义空间中, 是和猫类的语义特征相近的;

(4) 利用上面语义空间的性质, TriNet 模型在语义空间中寻找和输入语义特征最相似的另一个类别的语义特征, 并给其施加一定范围的高斯噪声, 输入下游的 TriNet 解码器模块;

(5) TriNet 解码器和 TriNet 编码器结构类似, 都是通过前面提到的 3 种基本操作构成的, 但是输入输出和编码器刚好相反, 即此时解码器输入的是语义空间的语义特征, 输出得到 ResNet-18 模型特征空间的深度特征. 此时解码器输出的深度特征就可以作为输入类别的伪数据, 完成数据增强任务.

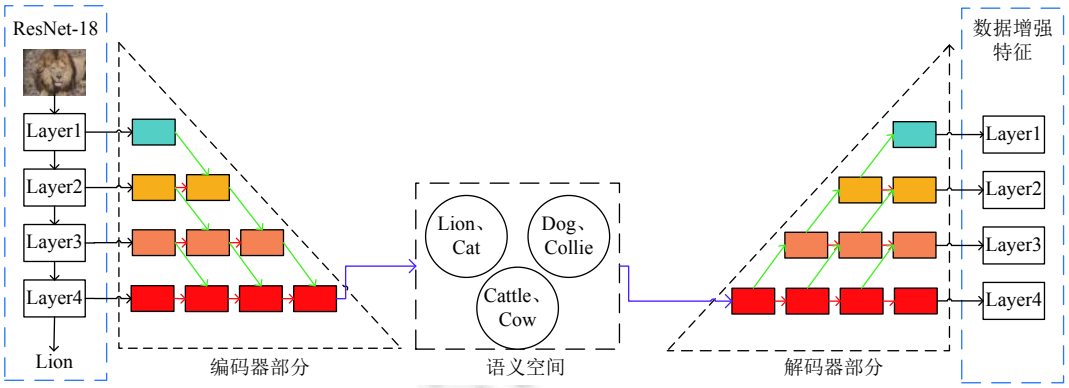


图 2 Dual TriNet 的算法框图[11]

经过充分的训练, TriNet 可以生成足够的伪数据, 将其补充进入只有少量标注的类别中, 然后我们将进行数据增强后的新任务数据输入之前预训练的识别模型, 通过微调的方法使得模型可以适应新的识别任务. 在后续的测试阶段, 我们直接将待识别的数据输入经过微调的模型, 通过识别模型的输出预测类别.

可以看到, 文献 [11] 提出的数据增强方法的输入输出不是常规的图片数据, 而是图片经过卷积神经网络结构 (这里是 4 层的 ResNet) 后得到的多层特征, 然后这里的中间层也有实际的物理含义, 即图片的语义空间, 比如基于 Word2Vec[12]得到的高维语义空间. 所谓语义空间其实是把图片数据映射到自然语言对应的特征空间中, 这是一种常用的多模态的方法.

文献 [13] 的主要贡献是提出了一种基于多级深度特征的语义空间数据增强方法. 该方法的核心是对自编码器的中间层的语义空间中的点进行调整, 从而产生新的多层特征映射. 语义空间中的调整包括两种类型:

(1) 添加高斯噪声. 通过对语义空间中的点添加一定范围内的高斯噪声, 作者认为此时解码得到的多维特征不会改变分类器的分类类别;

(2) 选取最近邻. 这个比较有实际意义, 比如图 2 所示, dual TriNet 模型将一副狮子的图片输入卷积神经网络, 得到一系列的高维特征, 这些高维特征经过编码器映射到语义空间中的一点, 即单词“Lion”对应的语义特征点. 在语义空间中, 和单词“Lion”最相邻的点, 比如是“Cat”, 那么模型就把“Cat”对应的语义特征解码, 得到一组网络特征来作为狮子的伪数据.

2.2 伪数据锐化决策边界

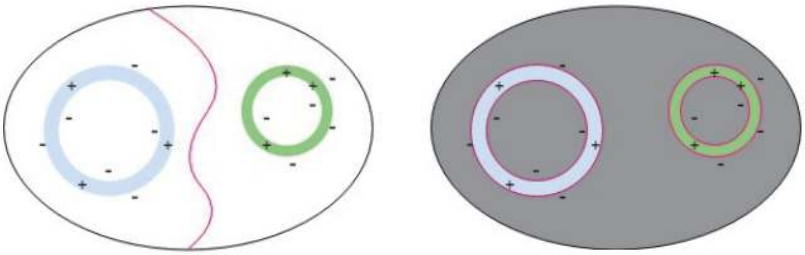
和前文讨论的伪数据增强新类标注样本的方法不同, 基于伪数据来锐化决策边界的方法对生成器的要求更低, 此时生成器不需要生成更加逼真的图片来帮助模型分类, 相反, 生成器目标更加贴近问题本身, 即利用伪数据来帮助算法锐化决策边界, 提升分类性能.

2.2.1 metaGAN

文献 [14] 提出了一种基于对抗生成网络 (GAN)[15]来生成伪数据的方法, 但是和前面提出的方法不同的是, 这篇文章并没有用伪数据来补充标注不足的类别, 而是将用 GAN 生成的伪数据作为一个新类别, 即假类 (fake class) 数据. 这种方法可以很好地和之前提出的小样本学习方法结合起来, 通过将小样本学习器和提出的数据生成

网络一起训练, 逐渐改善算法学习得到决策边界.

通过图 3 的可视化结果可以看到, 加上生成的假类数据之后, 小样本学习器有效地改善了分类算法的决策边界.



(a) 不使用 metaGAN 的决策边界 (b) 使用 metaGAN 的决策边界

图 3 metaGAN 改善决策边界对比图[14]

图 3(a) 是加入生成的假类数据前训练得到的决策边界, 其中决策边界用粉红线条表示, 蓝色和绿色区域为两个类别数据的分布区域. 可以看到, 因为标注数据的稀疏性, 算法学习到的决策边界虽然区分了当前的两个类别(颜色为蓝色和绿色的两个类别), 但是此时的决策边界并不能适应新数据的加入, 泛化性能差. 图 3(b) 是加入了生成的假类数据之后训练得到的决策边界, 其中“+”号数据代表分类类别的数据, “−”号数据代表生成的假类数据. 此时算法学习到的决策边界可以较好地刻画两个类别的分布, 得到的决策边界相较左图来说也更加锐化.

基于 metaGAN 的数据增强方法更加简单, 算法的训练和测试思路和上面两个数据增强算法类似, 但是 metaGAN 生成的伪数据不需要加入对应的类别, 而是单独作为一个噪声数据类别参与模型的微调. 因此该方法和以往的数据增强方法来说, 最大的优势在于算法简单, 对生成器的训练要求低, 因为生成的伪数据并不作为原有类别数据的补充, 而是充当锐化决策边界的角色, 所以在实际应用中往往可以取得更加好的泛化性能.

综合来说, 基于数据增强的思路来解决小样本学习问题是一种最直观的思路, 而且该类方法更加灵活, 通过设计数据增强模块生成伪数据, 将其扩充到小样本数据中, 使用混合数据直接对识别模型进行更新即可. 但是因为实际样本数目较少, 目前广泛使用的基于深度神经网络的方法在实际的数据增强中, 容易出现知识偏移以及过拟合的问题, 所以实际的应用效果会比后面介绍的几类方法差一些. 但是这种数据增强的思路对于解决实际的样本缺失问题来说更具有普遍意义, 所以将数据增强的思路融入迁移学习或者元学习的算法中, 是未来值得研究的方向.

3 基于迁移学习的小样本图像识别方法

面对标注限制的机器学习任务, 一个很自然的思路就是将模型在大数据集上进行预训练, 从中学习到一些有利于当前任务的先验知识, 从而来弥补标注数据不足的问题. 这一方法在机器学习领域, 尤其是近几年普遍使用的神经网络方法中取得了不错的效果, 下面关于为什么迁移学习[16]可以应用于小样本学习, 以及迁移学习如何应用于小样本学习进行讨论.

这里就拿在图像处理领域经常使用的卷积神经网络来举例说明. 众所周知, 卷积神经网络是通过多次卷积运算堆叠, 从图像数据中逐层提取特征, 并最终得到一个维度更低, 更利于后续全连接层的特征嵌入. 卷积神经网络为什么可以实现这么好的图像处理性能, 一直是学术界普遍关注的一个问题. 其中人们普遍认可的一个观点是, 卷积神经网络中特征的复杂性是随网络深度加深而提高的[17], 具体的信息可以通过图 4 来说明, 这里是用卷积神经网络实现的一个人脸检测的算法.

可以看到底层的神经网络学习到普遍是一些通用特征, 然后随着网络层数的加深, 特征逐渐变得特定化, 比如这里是一个人脸检测的算法, 随着网络层数的加深, 特征逐渐可以描述人的五官, 最后甚至可以表示一整张人脸信息.

对于图像处理任务而言, 用相似的网络结构模型来处理不同的任务, 网络前几层的特征一般都是相似的, 即是任务无关的, 比如卷积神经网络的卷积层参数一般是可以在不同任务之间共享的, 这也就是为什么有的网络模型

训练之初会使用一些规模较大、数据质量较高的数据集来进行预训练初始化参数的原因; 然后网络的高层特征, 以及特征提取后续的全连接层则更多地编码了任务特定信息, 不适合在任务之间共享, 需要使用目标数据来进行调整.

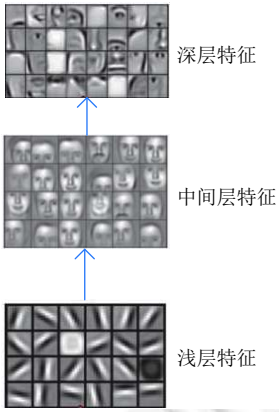


图 4 卷积神经网络逐层特征可视化[17]

前文讨论了迁移学习这一简单的思路为什么有效, 同时也指出了, 进行迁移学习最重要的任务是要完成高层特征以及特征映射下游全连接层参数的迁移. 如何进行迁移, 使得网络高层参数适合目标任务, 目前的研究主要分为了如下两种方法: 基于微调的方法, 以及基于前向传播的方法. 两种方法将会在下面的章节继续深入讨论.

3.1 基于微调的方法

深度神经网络的迁移学习方法中, 微调的方法几乎是最直观也最有效的方法. 微调的一个基本的流程如下.

(1) 在一个数据量充足的大数据集上进行预训练;

(2) 对应目标数据, 固定底层的特征权重, 只对网络高层的权重使用目标数据进行反向传播更新;

基于微调的方法步骤简单, 目前不少基于微调的方法在小样本问题上都取得了不错的效果, 下面的内容详细介绍了其中两种比较有代表性的方法.

3.1.1 Baseline++

文献 [18] 提出了一种基于微调的小样本图像识别方法, 算法的框图如图 5 所示.

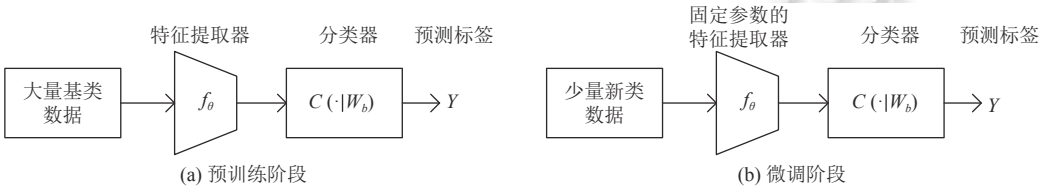


图 5 Baseline++算法流程[18]

从上述算法框图可以看到, 在预训练阶段, Baseline++方法首先从标注充分的基类数据中对网络整体参数进行训练, 这里的参数包括了特征提取器的参数 θ 和分类器的权重 W_b; 然后在微调阶段, Baseline++方法将第一步训练得到的特征提取器参数进行固定 (fixed), 然后使用标注有限的新类数据对分类器的权重 W_b 计算损失, 并更新其数值, 得到适用于小样本分类任务的一组新的分类器参数 W_n.

文献 [18] 指出了这样的观点: 以往的分类任务中, 特征提取器下游经常会使用 Softmax 分类器[19], Softmax 分类器的核心思想就是通过特征向量与分类器权重矩阵 W_b 作向量积, 通过比较向量积不同维度的数值大小来预测样本的类别. 这种方法在标注数据充分的情况下取得了较好的效果, 但是当标注数据稀疏的时候, 向量积运算反而会引入额外的向量尺度, 降低了网络预测的稳定性. 在文献 [18] 中, 作者将向量积运算改为了计算向量夹角余弦

值运算, 即去除了向量尺度对最终预测结果的影响. 通过消融实验可以证明, 在小样本学习问题上, 这种方法比原始的 Softmax 分类器分类性能更好.

3.1.2 Transductive fine-tuning

文献 [19] 提出了一种更加有效的 baseline 方法, 通过将查询集的无标签数据的香农熵[20]作为正则项加入损失函数, 使得算法可以利用无标签数据的信息, 在微调阶段学习到更加适应目标数据域的知识.

文献 [19] 的算法使用了如下的损失函数:

$$\Theta^* = \arg\min_{\Theta} \frac{1}{N_s} \sum_{(x,y)\in\mathcal{D}_s} -\log p_\Theta(y|x) + \frac{1}{N_q} \sum_{(x,y)\in\mathcal{D}_q} \mathbb{H}(p_\Theta(\cdot|x)) \tag{1}$$

其中, 符号 Θ 代表模型参数, N_s 是支持集样本数量, $\mathcal{D}_s$ 是支持集样本, p_Θ 是模型函数, N_q 是查询集样本数量, $\mathcal{D}_q$ 是查询集样本. 损失函数的前半部分就是分类任务中常见的交叉熵损失函数[21], 后半部分的正则项则相对来说比较陌生. 这里首先介绍一下香农熵的概念, 香农熵是用来度量随机变量的不确定性, 也就是这个随机变量所含有的信息量大小的. 对于一个随机变量 X, 它的香农熵定义如下:

$$\mathbb{H}(X) = -\sum_{x} P(x)\log P(x) \tag{2}$$

其中, 符号 $\mathbb{H}(X)$ 是随机变量的香农熵, $P(x)$ 代表随机变量的概率分布. 如果这个随机变量的分布相对集中, 则所含的信息量就较少 (不确定性较低), 此时香农熵就会较低; 如果这个随机变量的分布相对均匀, 极端情况下为均匀分布, 此时该随机变量的信息量就较高, 香农熵也会较高. 因此这里使用当前的网络模型对查询集上的无标签数据打上的预测得到的伪标签, 来度量模型所编码的信息量.

(1) 伪标签的各个类别预测概率分布均匀, 则说明此时的网络模型并不能对查询集中的样本给出一个较好的预测, 所以需要惩罚模型;

(2) 反之则说明模型可以较好地对新数据进行预测, 信息量或者说混乱程度更低, 此时正则项较小, 不影响模型更新.

这种方法显式地将查询集中的无监督信息编码进行了网络模型, 使得网络预测结果趋于概率集中, 通过文献 [19] 中的消融实验可以看到, 这种基于传导学习的方法取得了非常优秀的分类性能.

3.2 基于前向传播的方法

基于前向传播的方法本质上也是属于迁移学习的, 所以算法的思路和基于微调的方法类似, 唯一不同的是, 这种方法的出发点是希望减少在目标数据域上泛化的计算复杂度, 即通过一种前向传播 (forward) 的方法来显式地计算分类权重. 下面主要针对两种比较典型的算法展开讨论.

3.2.1 Dynamic Few-Shot Learning

文献 [22] 提出一种分类权重压印 (imprinting) 的方法, 算法的流程如图 6 所示.

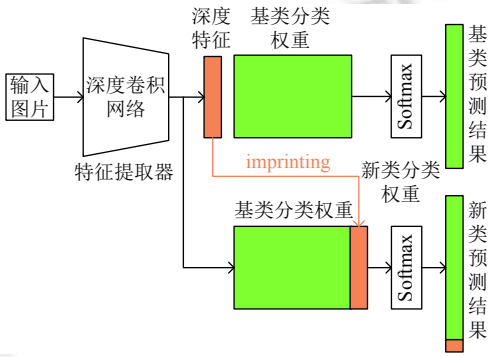


图 6 Imprinted weights 算法流程[22]

算法主要分为 3 个主要模块: 特征提取器、基类分类器以及拓展分类器. 算法采用如下的方法进行训练:

(1) 使用标注充足的基类数据集对特征提取器和基分类器进行预训练;

(2) 将标注稀疏的新类数据通过特征提取器, 得到对应的特征嵌入, 并使用每个新类的特征嵌入的平均值作为拓展分类器中该类所对应的分类权重;

(3) 使用所有的基类标注数据和新类标注数据一起对网络所有层的参数进行微调;

文献 [22] 的主要贡献是提供了一种基于新类特征映射的分类器权重初始化方法. 这种方法给出了 Softmax 分类器分类权重的一种显式计算的方法, 大大减少了反向传播带来的计算消耗, 加快了算法在新类上的泛化速度. 这种分类权重初始化方法在文献 [19] 中也进行了实验, 并且也取得了不错的实验性能.

3.2.2 Predicting Parameters from Activations

文献 [23] 与文献 [22] 的思路类似, 通过一种前向传播的思路来计算小样本数据上的分类权重, 但是文献 [23] 提出的算法引入了额外的参数映射网络, 可以学习到更加复杂的知识. 算法的参数映射网络可以如图 7 所示.

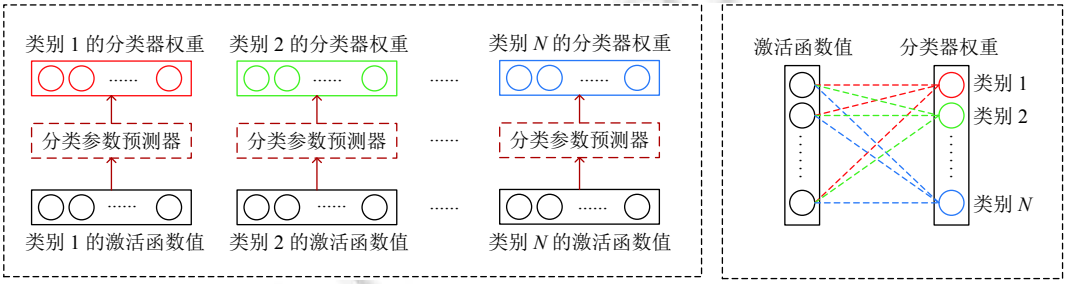


图 7 基于参数映射网络来构建分类权重[23]

图 7 的左边虚线方框描述了参数映射网络的输入输出, 以及该模块的训练过程.

(1) 该模块将各个类别的深度模型特征的平均值作为输入, 通过一个参数映射网络, 输出为映射得到的该类别的分类器权重. 本文为了方便讨论和训练, 使用了包含单层全连接层的网络来作为参数映射网络.

(2) 为了训练上述参数映射网络, 文献 [23] 从训练集中抽样出若干个小样本批数据, 通过批梯度下降的方法来训练模型.

图 7 的右边虚线方框描述了参数映射网络的推理阶段, 此时将小样本数据输入深度特征提取器, 得到每个类别对应的深度特征, 接着将其输入参数映射网络, 即可得到新类别的分类器权重, 用于该类别的预测.

总结一下和文献 [23] 提出的算法, 可以将其分为 3 个主要模块: 特征提取器、基类分类器以及权重映射网络. 算法采用如下的方法进行训练.

(1) 使用标注充足的基类数据集对特征提取器和基分类器进行预训练;

(2) 使用从基类数据采样的小样本批数据来训练参数映射网络;

(3) 使用预训练阶段学习到的特征激活到分类器权重的映射网络直接得到新类对应的 Softmax 分类器权重;

迁移学习的方法结构简单, 训练有效, 目前在几个小样本学习的数据集上都取得较好的性能. 但是迁移学习仍然存在一些改进的地方.

(1) 没有充分探索模型的可迁移性; 目前的迁移学习的方法都是在标注充分的基类数据上训练特征提取网络, 然后在标注稀疏的新类数据上微调. 但是基于深度神经网络的特征提取器的深层特征一般比底层特征编码了更多的特定信息, 采用更加有效的方法调整这些深层特征, 充分利用模型的可迁移性, 将会更加提高小样本学习算法的性能;

(2) 缺乏可解释性; 随着使用的网络模型的深度提升, 模型所编码的信息越来越缺乏可解释性, 探索深度神经网络中的可解释性, 也可以帮助后续的研究开发更加合理有效的小样本学习算法.

4 基于元学习的小样本图像识别方法

元学习[24]的目标是使得网络模型具有快速学习的能力, 快速学习是人类与生俱来的一种生存能力, 元学习方法希望模型具有像人类一样, 通过较少的示例就可以在较短的时间内学会分辨新的事物的能力. 通过元学习的问题定义可以发现, 元学习方法是处理小样本学习问题的一个重要思路. 本节将围绕 3 种用于小样本图像识别问题的元学习方法展开讨论, 这 3 种方法分别为基于优化器的小样本学习算法, 基于度量的小样本学习算法以及基于外部记忆的小样本学习算法.

4.1 基于优化学习器的小样本学习算法

基于梯度的反向传播算法[25]是深度神经网络得以进行学习的关键技术. 在监督学习的任务中, 若数据量足够多, 则模型通过随机梯度下降等优化方法可以有效的优化待学习的参数; 若在数据量较少的情况下, 仍使用上述常规的优化方法, 则会导致模型的过拟合现象, 降低模型的泛化能力. 一种直观的方法就是改进优化算法, 以应对模型在数据量较少情况下的参数学习. 本小节将介绍基于优化学习器的小样本学习方法, 首先介绍基于优化学习器的小样本学习的基本思想, 之后介绍其中具有代表性的方法.

从本文的第 3 节可以看到, 迁移学习的方法在图像处理领域中的重要作用, 其中微调方法是迁移学习在深度神经网络上的一种重要体现, 该方法直接利用在较大规模数据集上预训练得到的模型, 用于在小规模数据集上进行微调, 使得神经网络在微调之前就具有了一部分先验知识, 并将该先验知识存储于模型的参数中. 但是此种方法的出发点具有一定的启发性, 缺乏可解释, 基于优化学习器的小样本学习方法对于此问题进行了一次探索. 基于优化学习器的小样本学习是将先验知识存储于模型的参数中, 作为初始化参数, 当有一个数据规模较少的任务出现时, 模型在此初始化参数的基础上, 对参数进行进一步的优化. 基于优化学习器的小样本学习算法, 它的基本思路是将模型的初始参数, 比如神经网络中的权重初始值, 支持向量机[26]的惩罚因子等这些以往通过人为经验设定的数值, 也作为一个需要优化的参数, 设计相应的训练流程和损失函数来更新模型的初始参数. 因此需要优化的目标有两个, 一个是任务无关的先验初始化参数, 另一个是任务相关的参数. 一个好的初始化参数可以加快模型的学习速度; 弱数据较少, 还可能提高模型的性能. 基于优化学习器的小样本学习方法则学习了一个合适的初始化参数, 因此其可以帮助模型具有提升利用小规模样本进行快速学习的能力.

4.1.1 Meta-Learner LSTM

Ravi 等人[27]于 2017 年提出了 Meta-Learner LSTM 模型. 该模型的基本思想是定义一个基于 LSTM 模型[28]的元学习器, 通过训练此元学习器来优化学习器. 由于 LSTM 结构的特点, 该元学习器可以同时学习到此任务的长期记忆知识的知识和短期记忆的知识.

Meta-Learner LSTM 模型的结构如图 8 所示. 该算法首先定义了元学习器和学习器. 该模型从元数据集中采样用于训练的支持集和查询集. 在时间步 T 中, 从支持集中不断采样数据, 在每个时间步的迭代中, 基于 LSTM 的元学习器会接受来自学习器的梯度信息, 元学习器通过梯度信息不断更新初始化参数的值. 在 T 个时间步之后, 模型从查询集中进行采样, 学习器利用元学习器学习到模型参数对样本进行预测, 在查询集上得到的损失函数用于更新元学习器, 使得元学习器参数被优化. 在实际使用中, 一般是使用训练好的元学习器, 通过给出的训练数据集得到一个合适模型参数, 使用学习器根据上述初始化参数, 得到测试数据的预测结果.

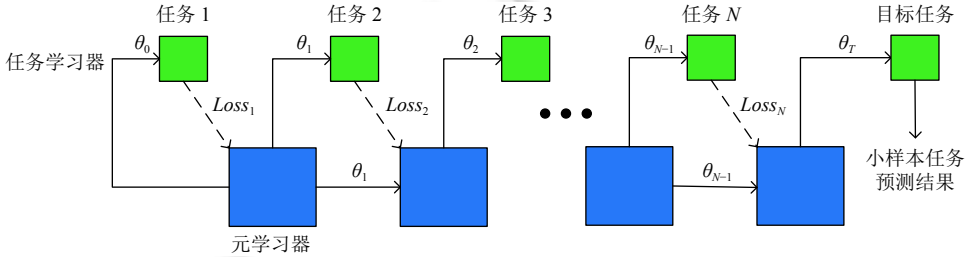


图 8 Meta-Learner LSTM 模型[27]

该算法的主要贡献是首次将序列优化问题进行了规范化, 通过使用 LSTM 这样的序列优化模型, 让模型按照顺序在不同的任务中交替训练, 使得模型具有通过少量样例, 可以快速从一个识别任务迁移到另一个识别任务中. 该算法的思路对于后续的基于优化的元学习方法具有一定的启发性, 但是由于训练数据较少, 训练 LSTM 模型所需参数规模较大, 该算法实际在小样本任务上的识别效果并没有很好.

4.1.2 MAML

Finn 等人[29]于 2017 年提出一种模型无关的元学习方法 MAML, 该方法用于处理小样本学习问题. 之所以称之为模型无关, 是因为 MAML 算法可以和其他深度学习模型相结合, 其对模型参数的更新依赖梯度优化思想, 并且不会为原模型引入其他的学习参数.

MAML 目标是使得模型可以对于样本较少的任务进行快速的适应, 其基本思想是通过优化的方法得到一个任务无关的初始参数, 从而使得模块在新的任务上可以快速收敛. MAML 将优化问题定义为双层优化问题, 内层优化针对任务相关的参数, 而外层优化针对任务无关的参数, 内层优化和外层优化都使用梯度下降方法. 这里用 T 表示采样得到的任务, 用 f 表示原始模型, 用 α 表示内层优化时的学习率, MAML 算法的目标函数定义如下:

$$\min_{\theta}\sum_{T_i\sim p(T)}L_{T_i}(f_\theta-\alpha\nabla_\theta L_{T_i}(f_\theta)) \tag{3}$$

此目标函数已经利用随机梯度下降算法完成了内层优化的更新, 即:

$$\theta_i'=\theta-\alpha\nabla_\theta L_{T_i}(f_\theta) \tag{4}$$

利用同样的优化算法对外层优化进行参数的更新:

$$\theta\leftarrow\theta-\beta\nabla_\theta\sum_{T_i\sim p(T)}L_{T_i}(f_{\theta_i'}) \tag{5}$$

其中, β 为外层优化的学习率. MAML 方法简洁但是有效, 值得注意的是除了监督学习, Finn 等人同时将 MAML 方法用于强化学习问题, 并取得了不错的结果. Nichol 等人[30]所提出的 Reptile 模型是基于一阶 MAML 模型的, Reptile 同时取消了内层优化仅更新一次的限制. 相比 MAML, 由于梯度更新由二阶转化为一阶, Reptile 的计算成本被大大节约.

Antoniou 等人[31]于 2019 年提出 MAML++模型, 针对 MAML 的不足之处进行了较全面的改进. 作者首先指出 MAML 模型的缺点: (1) 训练过程容易不稳定; (2) 由于二阶优化导致的计算成本高; (3) 由于学习率固定导致的模型灵活性低; (4) 由于缺少批量归一化[32]统计量累积导致的批量归一化效果较差. 然后对存在的问题进行优化. 对于训练不稳定问题, 作者提出多步损失优化方法, 通过改善梯度传播的方式缓解 MAML 优化过程中的不稳定性. 对于计算成本高的问题, 作者使用导数退火的方式, 通过将训练过程分段化, 在固定轮数的前一段使用一阶优化, 后一段使用二阶优化的方法, 加速训练过程. 对于模型灵活性问题, 作者通过学习内层优化学习率和方向的方式改善内层优化的灵活性, 通过针对元学习器的余弦退火方法改善外层优化的灵活性. 针对批量归一化方法效果较差的问题, 作者通过对每一步的相关信息进行统计来改善. 通过这些优化, MAML++算法对于小样本学习问题性能有明显的提升, 并且训练稳定性得到改善.

4.2 基于度量的小样本学习算法

机器学习领域中, 在样本的特征空间中使用合适的度量方法衡量不同样本之间的相似度是机器学习系统处理分类, 聚类等任务的有效手段, 因此度量学习 (metric learning)[33]是机器学习领域一个重要的研究方向. 本节主要回顾了基于度量的小样本学习方法的基本思路, 并介绍其中具有代表性的方法.

自深度学习技术发展以来, 在计算机视觉领域的图像分类任务中, 研究者一般使用卷积神经网络[34]作为图像的特征提取器, 使用全连接层作为特征的分类器, 由于大规模图像分类数据集 (如 ImageNet[35]) 被用于深度学习模型参数的训练, 因此参数可以被较好的被优化, 降低模型过拟合的风险. 在小样本学习的范式, 即 N-way K-shot 的分类任务下, 对于每个分类任务 (task), 每一类物品仅对应数量为 K 的样本 (K 通常的取值为 1, 5, 10), 因此每一个分类任务仅拥有较少的训练数据, 若直接使用常规的方法训练以全连结为基础分类器, 会加大模型过拟合的风险,

降低模型的泛化性能.

基于度量的小样本学习方法的基本思路是利用每个类别 K-shot 的数据, 学习一个特征嵌入空间, 使得在此特征嵌入空间中模型更有效的度量样本之间的相似度, 相同类别的样本相似度较高, 不同类别样本的相似度较低. 在合适的特征嵌入空间中, 利用非参数化的分类模型 (例如 K-近邻方法)N 类样本进行分类. 借助非参数化的分类模型, 基于度量的小样本学习方法相对降低了特征提取器的训练难度, 因此更适合样本较少的分类任务, 同时使得模型的结构更加灵活, 可以快速地识别新的类别.

4.2.1 Siamese Network

Koch 等人[36]于 2015 年提出基于孪生网络 (Siamese Network) 的模型. 其基本思路是利用孪生网络判断输入的两张图像的相似度, 即是否属于同一类. 在训练阶段, 对具有两组相同卷积神经网络的孪生网络进行训练, 本质为训练一个二分类器, 若样本标签为 1, 则说明两张图像属于同一类; 若标签为 0 则相反, 以此来学习两张输入图像彼此的相似度. 在测试阶段, 该模型会将待测试图像分别与任务中训练集里每个类别的图像单独输入孪生网络中, 并判断待测试图像为相似度最高的输出结果所对应的类别.

在孪生网络中, 模型利用两组卷积神经网络将两张输入图像映射到特征空间之后, 作者使用 L1 距离来度量特征空间里两个样本的相似度:

$$p=\sigma\left(\sum_j \alpha_j|h_1{}^{(j)}-h_2{}^{(j)}|\right) \tag{6}$$

其中, h_1和h_2 分别代表两个输入样本的特征向量, p 代表两个样本的相似度. 同时本文使用带有正则项的交叉熵作为损失函数, 损失函数定义如下:

$$L(x_1{}^{(i)},x_2{}^{(i)})=y(x_1{}^{(i)},x_2{}^{(i)})\log p(x_1{}^{(i)},x_2{}^{(i)})+(1-y(x_1{}^{(i)},x_2{}^{(i)}))\log(1-p(x_1{}^{(i)},x_2{}^{(i)}))+\lambda|w|^2 \tag{7}$$

其中, 符号 L 代表损失函数, p 代表网络模型, w 代表正则项. 此损失函数使得孪生网络结构训练特征嵌入空间对应的映射, 使嵌入空间中类别相同的两个样本距离相近.

孪生网络的学习思路比较清晰, 通过将数据按对组织起来, 训练一个二分类器判断输入样本是否属于同一类别. 该算法对于后续的基于度量学习的小样本学习算法起到了很好的启发意义, 但是孪生网络仍然存在不少问题, 首先是孪生网络的度量方式容易受到噪声的影响, 如果训练数据中存在一些离群点, 孪生网络测试的时候仍然会考虑这些噪声点对于待预测样本所属类别的影响. 其次是这种训练方法效率较低, 因为经过两两组合, 训练数据的规模实际上会增长若干个数量级, 一些冗余的信息也会影响模型最终的效果.

4.2.2 Match Network

Vinyals 等人[37]于 2016 年提出匹配网络 (Match Network) 模型用于小样本学习. 匹配网络通过引入注意力机制和记忆结构来对小样本数据集进行快速学习. 该网络提出两种重要的模块, 分别为上下文嵌入结构和注意力核机制. 上下文嵌入结构是指为了更好的建立支持集集中样本的联系, 在使用卷积网络对相应图像进行特征提取之后, 使用一个双向的长短期记忆 (LSTM) 网络对所有特征进行一次重编码, 利用长短期记忆网络中的各单元之间的连接建立样本之间的联系, 使编码后特征更有效的对其不同类样本进行表示. 注意力核机制是使用类似注意力结构的方法, 建立支持集中样本和待测试样本的联系. 为了建立端到端的学习过程, 本文使用的注意力核为建立在余弦距离上的 Softmax 函数, 表示如下:

$$a(\hat{x},x_i)=\frac{\mathrm{e}^{c(f(\hat{x}),g(x_i))}}{\sum_{j=1}^{k}\mathrm{e}^{c(f(\hat{x}),g(x_j))}} \tag{8}$$

此结构运用了 K-近邻分类器的思想并将其形式化为一个可导的过程, 从而使得匹配网络的参数可以被直接优化.

Match Network 较好地解决了孪生网络中存在的问题, 输入数据不需要按照成对组织, 仍然可以获得类似的基于 K-近邻的度量识别精度, 是度量学习解决小样本问题中的一个典型方法.

4.2.3 Relation Network

Sung 等人[38]于 2018 年提出关系网络 (Relation Network) 模型用于小样本学习. 在匹配网络等基础上对度量网络对结构进行了简化. 关系网络分为两部分, 分别为特征嵌入模块和关系模块. 在特征嵌入模块中, 关系网络使用卷积神经网络作为特征提取器, 分别提取支持集和待测试样本的特征. 在关系模块中, 关系网络将已经得到的待测试图像的特征分别与支持集中各类图像的特征进行融合, 本文使用的融合方式为直接拼接; 之后使用一个统一的关系预测网络预测待测试图像与各类的关系即关系分数, 判定关系分数最大的类为预测结果. 其中关系预测网络由卷积神经网络和全连接网络构成. 可以发现关系网络的结构与匹配网络相比, 更加简洁有效.

Relation Network 算法在 Match Network 模型的基础上, 将基于欧式空间的距离度量, 修改为了基于多层感知机的非线性距离度量, 该模型编码了样本之间更复杂的距离度量, 实际的识别性能会比 Match Network 更好.

4.2.4 Prototype Network

Snell 等人[39]于 2017 年提出原型网络 (Prototype Network) 结构用于小样本学习. 原型网络的主要思想是在特征空间上, 为每一个类别寻找一个特征原型, 当得到待测试图像的特征后, 分别与每个类的特征原型进行距离度量, 从而得到预测结果. 类原型定义如下:

$$c_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_\phi(x_i) \tag{9}$$

即为每个类别对应特征的平均值. 原型网络的训练过程分为两个阶段, 第 1 阶段利用训练集中的支持集得到每个类的类原型, 第 2 阶段对训练集中的查询集进行预测, 使用 Softmax 函数得到预测的分布概率, 并计算损失函数进行参数的优化.

与上述几种小样本学习方法相比, 原型网络反映了一种更简单的归纳偏差, 更有利于数据条件有限的情况, 在训练数据存在噪声的情况下, 原型网络往往会取得更好的识别性能.

4.3 基于外部记忆的小样本学习算法

前面讨论的小样本学习方法都是将从训练数据集中学习到的经验直接存储到网络模型的参数中, 并且没有使用额外的记忆存储模块; 在数据量充足的情况下, 此方式被证明是有效的, 但是当数据量较少时候, 如何更有效的存储从有效样本中所学习到的知识是一个解决小样本问题的思路. 本小节主要围绕基于外部记忆机制的小样本学习方法展开讨论, 首先介绍基于外部记忆机制的小样本学习的基本思想, 之后介绍其中具有代表性的方法.

当前计算机视觉领域的深度学习系统多以模型的参数作为存储已学习知识的媒介, 通过不断优化模型的参数使得深度学习系统具有一定的感知能力. LSTM 等循环神经网络模型虽然使用了与模型“记忆”有关的操作, 但是并没有摆脱利用模型参数作为知识存储的方式. 2014 年 Graves 等人[40]提出的神经图灵机提出一种不同的机器学习系统存储已习得知识的方式, 使用额外的记忆存储模块来实现机器学习系统的功能, 通过“写”操作将知识存入记忆存储模块, 通过“读”操作来使用知识. 该研究证明了知识的存储访问操作在机器学习任务中的有效性. 在小样本学习问题中, 数据量少是系统所需要解决的主要问题, 因为传统的机器学习系统通过规模较大的训练数据来提升模型的泛化性能, 因此如何高效地利用这些数据是小样本学习系统在设计过程中出发点. 前面所介绍的小样本学习方法主要关注的问题是如何更加有效地利用小规模数据集进行学习, 如改变机器学习系统进行优化的方式, 使得模型可以在数据量较少的基础上, 更有效的对模型参数进行优化. 相比更有效的学习方式, 基于的外部记忆机制的小样本学习方法更加关注的是如何更好的保存机器学习系统从少量样本中学习的知识, 与神经图灵机相似, 基于外部记忆的小样本学习系统的基本思路是使用额外的记忆存储结构保存知识, 从而更高效的利用数据. 通过更有效的知识存储方式帮助系统提升在小规模数据集上进行学习的能力.

4.3.1 Memory-Augmented Neural Network

Santoro 等人[41]于 2016 年提出了记忆增强神经网络模型 (memory-augmented neural network, MANN) 用于小样本学习任务. MANN 的设计收到神经图灵机的启发, 可以被认为是神经图灵机一种不同的实现方式. MANN 中基于外部记忆的模型结构如图 9 所示.

MANN 使用 LSTM 或者前馈神经网络作为模型的控制器, 并且拥有用于知识存储的外部记忆模块, 外部记忆模块如图 9 蓝色部分所示. 在控制器的每个单元中包含了读操作和写操作, 读写操作均针对外部记忆模块进行, 图 9 中蓝色有向实线描述了模型对于外部记忆模块的读写操作的过程, 红色有向实线则描述了模型反向传播的过程. 在对于输入数据进行特征表示之后, 控制器的写操作将内容写入记忆模块中, 而读操作则对于记忆模块中的内容进行检索. 对于输入数据, 控制器生成对应的键值, 并将其存入记忆模块中. 当控制器单元接受输入数据准备进行读操作时, MANN 使用余弦相似度作为检索的度量, 从而得到预测的分布.

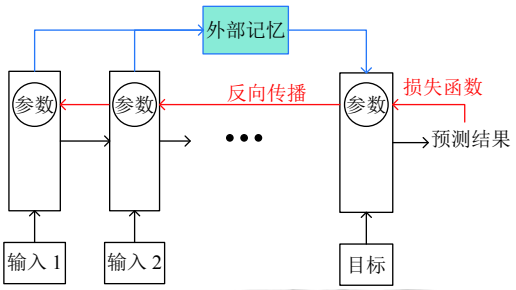


图 9 MANN 结构示意图[41]

可以发现, MANN 在对记忆模块进行访问时依据记忆的内容而不是想神经图灵机一样依据内存地址. MANN 模型结合了梯度下降方法和记忆存储方式两者的优点, 利用梯度下降方法学习一种有效的特征表示, 而利用记忆存储模块快速记忆知识.

4.3.2 MetaNet

Munkhdalai 等人[42]于 2017 年提出一种可以跨任务学习的 MetaNet 模型用于小样本学习任务. 与 MANN 不同, MetaNet 不再使用 LSTM 作为控制器, 而是将模型划分为任务不同的学习模块, 以此来区分任务学习到的相关知识和任务无关知识. MetaNet 结构的示意图如图 10 所示.

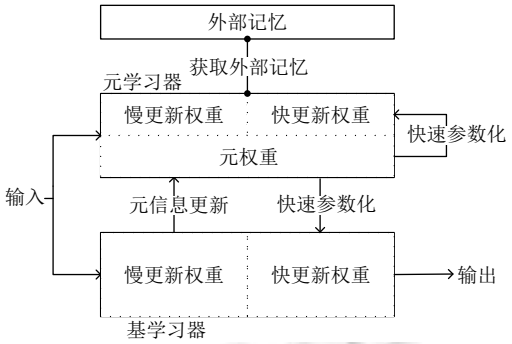


图 10 MetaNet 结构示意图[42]

MetaNet 由两个主要结构组成, 即基础学习器和元学习器, 并设置了一个外部记忆模块. 在学习的过程中, 基础学习的过程在输入任务的空间中进行, 而元学习的过程在任务无关的元空间中进行. 通过在抽象的元空间中进行学习, 元学习器可以在不同的任务之间进行持续学习, 以获得任务无关的元知识. 当系统面对一个新的任务时, 基础学习器对其进行分析, 并将分析结果以高阶形式向元学习器反馈; 元学习器通过反馈, 利用外部记忆模块, 可以快速优化自身参数和基础学习器对参数, 以适应新的任务.

MetaNet 利用将元学习器学习到的任务无关知识存储到记忆模块的方式, 可以实现对于新任务的快速学习.

基于元学习的思路来解决小样本学习问题, 是近两年该领域的研究热点, 如何划分任务通用参数和任务特定参数, 如何更加有效地训练元学习模型等课题一直具有相当的活力. 元学习算法希望学习一个可以“自主”学习的

模型, 使得模型在只有少量样本的新任务上可以快速泛化. 尽管元学习方法在小样本学习中已经取得了不错的效果, 但是该类方法仍然存在一些问题.

(1) 元学习算法优化难; 因为采用多任务交替训练的方式来更新模型, 不同任务的数据之间存在数据分布的不同, 只是简单地交替训练, 在任务数据分布差别较大的时候, 会导致最后的模型难以收敛的问题;

(2) 元学习算法缺乏相关的可解释性; 元学习算法的思路具有一定的启发性, 但是关于方法的有效性一直难以被证明, 同时元学习方法和迁移学习方法之间的区别也一直是研究者们关注的重点, 如何从理论上解释元学习的有效性, 是未来的一个重要的研究方向.

5 实验结果对比

目前小样本图像识别研究普遍使用基于 ImageNet 数据集采样得到的 mini-ImageNet[43]数据集来作为评估基准. mini-ImageNet 数据集包含了 100 个类别的数据, 其中 64 个类别作为训练集使用, 20 个类别数据作为验证集使用, 剩下的 16 个类别数据作为测试集使用. 表 1 统计了目前主流的小样本学习算法在 mini-ImageNet 数据集上的实验性能. 其中基础构架一列描述了算法使用的神经网络结构; 5-way 1-shot 的实验结果代表在包含了 5 种未知类, 每个未知类标注数据只有 1 例的情况下算法的识别准确率; 5-way 5-shot 的实验结果代表在包含了 5 种未知类, 每个未知类标注数据只有 5 例的情况下算法的识别准确率.

表 1 模型在 mini-ImageNet 数据集上的准确率结果 (%)

类型	名称	基础构架	5-way 1-shot	5-way 5-shot
基于数据增强	delta-encoder	VGG16	59.9	69.7
	dual trinet	ResNet-18	58.12±1.37	76.92±0.69
	metaGAN	Conv-4	52.71±0.64	68.63±0.67
基于迁移学习	Baseline++	Conv-4	48.24±0.75	66.43±0.63
	Transductive fine-tuning	WRN-28-10	68.11±0.69	80.36±0.50
	Imprinting*	WRN-28-10	58.47±0.66	75.56±0.52
	Activation to Parameter	WRN-28-10	59.60±0.41	73.74±0.19
基于优化的元学习	Meta-Learner LSTM	Conv-4	43.44 ± 0.77	60.60 ± 0.71
	MAML	Conv-4	48.70 ± 1.84	63.11 ± 0.92
	Reptile	Conv-4	49.97 ± 0.32	65.99 ± 0.58
	MAML++	Conv-4	52.15 ± 0.26	68.32 ± 0.44
基于度量的元学习	Match Network	Conv-4	46.6	60.0
	Prototype Network	Conv-4	49.42 ± 0.78	68.20 ± 0.66
	Relation Network	Conv-4	50.44 ± 0.82	65.32 ± 0.70
基于外部记忆的元学习	MetaNet	Conv-5	49.21 ± 0.96	—

注: 提出Imprinting算法的文献[22]并没有在Mini-ImageNet数据集上进行实验, 此处的实验结果是文献[19]中按照Imprinting算法的思想进行实验得到的结果

同时因为不同的算法使用了不同的基础网络结构, 按照算法所使用的基础网络结构对上述介绍的方法进行划分, 得到如表 2 所示的性能对比.

可以看到不同的算法使用了结构各异的基础构架来提取小样本数据的深度特征, 这些网络构架分别具有下面的特点.

(1) Conv-4 是 Yu 等人[18]在 2018 年提出的一种小样本学习网络结构. 从该模型的名字可以看出, 该模型的主要组成就是 4 层卷积层和对应的批归一化层以及池化层. 作者提出这种简单的网络模型, 主要是为了让研究者更加重视小样本学习算法中迁移性的实现, 提高小样本学习算法的可解释性.

(2) VGG-16 是由 Simonyan 等人[44]在 2014 年提出的一种卷积神经网络模型, 该模型在 2014 年的 ImageNet 图像分类和定位挑战赛中取得了优异的成绩. 它的突出特点就是简单, VGG16 的卷积层均采用相同的卷积核参

数, 池化层均采用相同的池化核参数, 整体模型是由若干卷积层和池化层堆叠组成的.

(3) ResNet-18 是 2015 年由何凯明等人[45]在现有的深度卷积神经网络基础上提出的网络模型, 该网络结构有效地改进了现有网络在多层堆叠后出现的过拟合以及训练困难等问题, 突破性地提出了残差连接卷积的算法思路, 使得网络更加专注于学习层之间的变化, 提高了网络的训练效果. ResNet-18 是该系列网络中的一种, 因为网络层数较浅, 泛化性能较好, 训练速度较快, 现在已经被学术界和工业界广泛使用;

(4) WRN-28-10 是由 Zagoruko 等人[46]于 2017 年提出的深度卷积网络模型, 该模型在 ResNet 系列网络结构的基础上进行改进, 它仅使用了 28 个卷积层的简单结构, 算法性能就可以超过 ResNet-100, 而且算法速度有了历史性的提升, 比起之前的 ResNet 几乎快了一半. 因为其速度快、精度高的原因, 现在已经在图像分类, 定位等任务中被广泛使用.

表 2　模型在 mini-Imagenet 数据集上的准确率结果 (%)

基础构架	名称	5-way 1-shot	5-way 5-shot
Conv-4	metaGAN	52.71±0.64	68.63±0.67
	Baseline++	48.24±0.75	66.43±0.63
	Meta-Learner LSTM	43.44 ± 0.77	60.60 ± 0.71
	MAML	48.70 ± 1.84	63.11 ± 0.92
	Reptile	49.97 ± 0.32	65.99 ± 0.58
	MAML++	52.15 ± 0.26	68.32 ± 0.44
	Match Network	46.6	60.0
	Prototype Network	49.42 ± 0.78	68.20 ± 0.66
	Relation Network	50.44 ± 0.82	65.32 ± 0.70
Conv-5	MetaNet	49.21 ± 0.96	—
VGG16	delta-encoder	59.9	69.7
ResNet-18	dual trinet	58.12±1.37	76.92±0.69
WRN-28-10	Transductive fine-tuning	68.11±0.69	80.36±0.50
	Imprinting	58.47±0.66	75.56±0.52
	Activation to Parameter	59.60±0.41	73.74±0.19

通过上面的实验对比可以发现:

(1) 基于迁移学习的小样本学习方法中, Transductive fine-tuning 可以取得较好的效果, 这说明通过将查询集的无标签数据的香农熵作为正则项加入损失函数, 可以帮忙模型学习到更加具有区分性的分类模型, 从而有效地提高了少量标注问题上的识别性能. 另一方面, 基于特征参数映射的 Activation to Parameter 算法, 和直接基于深度特征取平均值的 Imprinting 算法相比, 实际上提升不是很大, 这说明了预训练的深度卷积网络特征可以有效地区分不同类别, 这也为后续基于深度特征泛化的研究提供了实验支撑.

(2) 在基于元学习的小样本学习方法中, 通过对比可以发现, 基于优化的元学习方法里, MAML 方法的两个改进算法: Reptile 和 MAML++方法都取得了不错的效果, 这也证明了改善模型训练过程中的不稳定性, 对模型中的不确定性进行了刻画和考虑, 都可以有效地提升模型的性能, 这也为解决小样本学习问题提供了思路, 即考虑训练模型中存在的不确定性, 以及如何解决;

(3) 在基于度量的元学习方法中, Prototype Network 在 5-way 5-shot 的实验条件下可以得到较好的效果, Relation Network 在 5-way 1-shot 的实验条件下可以得到较好的效果, 这说明 Relation Network 提出的度量网络结构, 在样本量更少的小样本分类问题中, 更具有识别优势.

6　总结和展望

在机器学习领域之中, 不同任务机器学习任务中数据集的规模和质量是限制机器学习系统性能的重要问题. 小样本图像识别任务关注在机器学习系统在数据规模较少情况下的学习问题, 解决好小样本学习问题, 于学术界

可以帮助相关研究者更好的理解机器学习系统的内在机理, 于工业界可以有效的节约数据的标注成本, 因此近年来小样本学习领域备受研究者的关注. 在本文中, 我们主要关注图像分类任务中的小样本学习问题. 首先我们形式化的定义了图像分类任务中的小样本学习问题, 之后我们分别介绍了现有的不同种类的小样本学习模型, 包括基于数据增强的方法, 基于迁移学习的方法, 基于度量的方法, 基于优化的方法, 基于外部记忆的方法. 最后在标准数据集上比较了几类小样本图像识别模型的性能并进行分析. 我们基于对小样本学习领域总结的结果, 提出了几个发展的方向.

(1) 神经网络可解释性[47]. 尽管现阶段深度学习模型在不同领域中均取得了明显的成绩, 但是神经网络本身具有一定的黑盒性. 因此通过对于神经网络可解释性的进一步探索, 可以让研究者对于深度学习机理有更深的了解, 方便研究者根据深度学习的内在机理针对样本较少的问题做出更合理的结构上或者训练方法上的改善.

(2) 更通用的小样本学习方法. 现阶段研究者虽然开始关注更多任务中的小样本学习问题, 但是他们通常是基于设定好的任务模式进行研究, 比如小样本研究领域广泛使用的 mini-ImageNet 数据集, 每个子任务都是采用 5-way 1-shot, 或者 5-way 5-shot 这样规范的任务设定进行数据划分的, 但是实际的小样本学习系统应该是可以处理任意类别和任意标签数据的小样本识别问题的. 而且目前研究使用的小样本学习任务本质上都是从一个完整的大数据集上进行数据划分得到的, 每个子任务之间仍然存在较大的关联性. 基于更加真实的小样本任务, 以及数据组织更加宽松的数据展开研究, 是将小样本研究从理论推往实践的至关重要的一步.

(3) 增量学习问题. 目前小样本增量学习[48]已经开始被研究者所关注, 但是大部分小样本学习系统在设计的过程中并没有考虑系统的增量学习问题. 小样本识别系统在工作的初期会面对数据不足的问题, 但是随着越来越多的数据进入系统, 小样本识别系统所积攒的标注数据将会越来越多, 如何充分利用这些新进入的数据, 来改善和提高当前系统的识别系统, 对于小样本学习系统的可持续性工作至关重要. 因此将增量学习的研究和小样本学习技术结合起来, 将会有利于小样本学习技术的落地.

小样本学习领域当前仍然具有蓬勃的生机, 本文仅对于现有的图像分类任务上的小样本学习模型进行总结, 目前不同领域, 不同任务上的小样本学习问题也逐渐被研究者们所挖掘, 例如计算机视觉领域中的语义分割任务[49], 自然语言处理领域的关系抽取任务[50], 以及强化学习任务, 增量学习任务. 这些任务中的小样本学习系统在与一般系统相比较时, 性能通常存在一定的差距, 可见小样本学习领域依然有较长的一段路要走, 我们相信小样本学习领域会收到越来越多的关注.

References:

[1] Taigman Y, Yang M, Ranzato MA, Wolf L. DeepFace: Closing the gap to human-level performance in face verification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 1701–1708. [doi: 10.1109/CVPR.2014.220]

[2] El Sallab A, Abdou M, Perot E, Yogamani S. Deep reinforcement learning framework for autonomous driving. Electronic Imaging, 2017, 2017(19): 70–76. [doi: 10.2352/ISSN.2470-1173.2017.19.AVM-023]

[3] Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J. A guide to deep learning in healthcare. Nature Medicine, 2019, 25(1): 24–29. [doi: 10.1038/s41591-018-0316-z]

[4] Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge: MIT Press, 2016.

[5] Vanschoren J. Meta-learning: A survey. arXiv: 1810.03548, 2018.

[6] Fort S. Gaussian prototypical networks for few-shot learning on omniglot. arXiv: 1708.02735, 2017.

[7] Zhang HY, Cisse M, Dauphin YN, Lopez-Paz D. mixup: Beyond empirical risk minimization. In: Proc. of the 6th Int'l Conf. Paper at ICLR 2018. Vancouver, 2018.

[8] Pham H, Dai ZH, Xie QZ, Luong MT, Le QV. Meta pseudo labels. arXiv: 2003.10580, 2020.

[9] Hariharan B, Girshick R. Low-shot visual recognition by shrinking and hallucinating features. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision. Venice: IEEE, 2017. 3037–3046. [doi: 10.1109/ICCV.2017.328]

[10] Schwartz E, Karlinsky L, Shtok J, Harary S, Marder M, Kumar AD, Feris RS, Giryes R, Bronstein AM. Δ-encoder: An effective sample synthesis method for few-shot object recognition. In: Proc. of the 32nd Int'l Conf. on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2018. 2850–2860.

[11] Chen ZT, Fu YW, Zhang YD, Jiang YG, Xue XY, Sigal L. Semantic feature augmentation in few-shot learning. arXiv: 1804.05298v2,

2018.

[12] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: Proc. of the 26th Int'l Conf. on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2013. 3111–3119.

[13] Chen ZT, Fu YW, Zhang YD, Jiang YG, Xue XY, Sigal L. Multi-level semantic feature augmentation for one-shot learning. IEEE Trans. on Image Processing, 2019, 28(9): 4594–4605. [doi: 10.1109/TIP.2019.2910052]

[14] Zhang R, Che T, Ghahramani Z, Bengio Y, Song Y. MetaGAN: An adversarial approach to few-shot learning. In: Proc. of the 32nd Conf. on Neural Information Processing Systems. Montreal, 2018. 2371–2380.

[15] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: Proc. of the 27th Int'l Conf. on Neural Information Processing Systems. Cambridge: MIT Press, 2014. 2672–2680.

[16] Lu H, Zhang L, Cao ZG, Wei W, Xian K, Shen CH, Van Den Hengel A. When unsupervised domain adaptation meets tensor representations. In: Proc. of the 2017 IEEE Int'l Conference on Computer Vision. Venice: IEEE, 2017. 599–608. [doi: 10.1109/ICCV.2017.72]

[17] Lee H, Grosse R, Ranganath R, Ng AY. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proc. of the 26th Annual Int'l Conf. on Machine Learning. New York: ACM, 2009. 609–616. [doi: 10.1145/1553374.1553453]

[18] Chen WY, Liu YC, Kira Z, et al. A closer look at few-shot classification. arXiv: 1904.04232v2, 2019.

[19] Dhillon GS, Chaudhari P, Ravichandran A, Soatto S. A baseline for few-shot image classification. arXiv: 1909.02729, 2020.

[20] Grandvalet Y, Bengio Y. Semi-supervised learning by entropy minimization. In: Proc. of the 17th Int'l Conf. on Neural Information Processing Systems. Cambridge: MIT Press, 2005. 529–536.

[21] Pettersson R. Visual information. Educational Technology, 1993.

[22] Qi H, Brown M, Lowe DG. Low-shot learning with imprinted weights. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 5822–5830. [doi: 10.1109/CVPR.2018.00610]

[23] Qiao SY, Liu XC, Shen W, Yuille A. Few-shot image recognition by predicting parameters from activations. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7229–7238. [doi: 10.1109/CVPR.2018.00755]

[24] Mishra N, Rohaninejad M, Chen X, Abbeel P. A simple neural attentive meta-learner. In: Proc. of the 6th Int'l Conf. on Learning Representations. Vancouver: OpenReview.net, 2018.

[25] Nielsen MA. Neural networks and deep learning. San Francisco: Determination Press, 2015.

[26] Vapnik VN. An overview of statistical learning theory. IEEE Trans. Neural Network, 1999, 10(5): 988–999.

[27] Ravi S, Larochelle H. Optimization as a model for few-shot learning. In: Proc. of the 5th Int'l Conf. on Learning Representations. Toulon: OpenReview.net, 2017.

[28] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735–1780. [doi: 10.1162/neco.1997.9.8.1735]

[29] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proc. of the 34th Int'l Conf. on Machine Learning. Sydney: JMLR.org, 2017. 1126–1135.

[30] Nichol A, Achiam J, Schulman J. On first-order meta-learning algorithms. arXiv: 1803.02999v3, 2018.

[31] Antoniou A, Edwards H, Storkey A. How to train your MAML. arXiv: 1810.09502, 2019.

[32] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proc. of the 32nd Int'l Conf. on Int'l Conf. on Machine Learning. Lille: JMLR, 2015. 448–456.

[33] Davis JV, Kulis B, Jain P, Sra S, Dhillon IS. Information-theoretic metric learning. In: Proc. of the 24th Int'l Conf. on Machine Learning. New York: ACM, 2007. 209–216. [doi: 10.1145/1273496.1273523]

[34] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc. of the IEEE, 1998, 86(11): 2278–2324. [doi: 10.1109/5.726791]

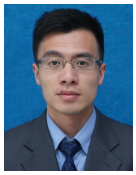
[35] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma SA, Huang ZH, Karpathy A, Khosla A, Bernstein M, Berg AC, Li FF. ImageNet large scale visual recognition challenge. Int'l Journal of Computer Vision, 2015, 115(3): 211–252. [doi: 10.1007/s11263-015-0816-y]

[36] Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. In: Proc. of the Int'l Conf. on Machine Learning Deep Learning Workshop. 2015. 2.

[37] Vinyals O, Blundell C, Lillicrap T, Kavukcuoglu K, Wierstra D. Matching networks for one shot learning. In: Proc. of the 30th Int'l Conf. on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2016. 3630–3638.

[38] Sung F, Yang YX, Zhang L, Xiang T, Torr PHS, Hospedales TM. Learning to compare: Relation network for few-shot learning. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 1199–1208. [doi: 10.1109/CVPR.

2018.00131]
[39] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. In: Proc. of the 31st Conf. on Neural Information Processing Systems. Long Beach, 2017. 4077–4087.
[40] Graves A, Wayne G, Danihelka I. Neural turing machines. arXiv: 1410.5401v2, 2014.
[41] Santoro A, Bartunov S, Botvinick M, Wierstra D, Lillicrap TP. Meta-learning with memory-augmented neural networks. In: Proc. of the 33rd Int'l Conf. on Int'l Conf. on Machine Learning. New York: JMLR, 2016. 1842–1850.
[42] Munkhdalai T, Yu H. Meta networks. In: Proc. of the 34th Int'l Conf. on Machine Learning. Sydney: JMLR.org, 2017. 2554–2563.
[43] Oreshkin BN, Rodriguez P, Lacoste A. Tadam: Task dependent adaptive metric for improved few-shot learning. In: Proc. of the 32nd Int'l Conf. on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2018. 719–729.
[44] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556v4, 2014.
[45] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778. [doi: 10.1109/CVPR.2016.90]
[46] Zagoruyko S, Komodakis N. Wide residual networks. In: Richard C, Wilson ERH, Smith WAP, eds. Proc. of the British Machine Vision Conference. New York: BMVA Press, 2016.
[47] Wu M, Hughes MC, Parbhoo S, Zazzi M, Roth V, Doshi-Velez F. Beyond sparsity: Tree regularization of deep models for interpretability. arXiv: 1711.06178, 2017.
[48] Tao XY, Hong XP, Chang XY, Dong SL, Wei X, Gong YH. Few-shot class-incremental learning. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 12183–12192. [doi: 10.1109/CVPR42600.2020.01220]
[49] Rakelly K, Shelhamer E, Darrell T, Efros A, Levine S. Conditional networks for few-shot semantic segmentation. In: Proc. of the Workshop Track-ICLR 2018. Vancouver: OpenReview.net, 2018.
[50] Gao TY, Han X, Liu ZY, Sun MS. Hybrid attention-based prototypical networks for noisy few-shot relation classification. Proc. of the AAAI Conf. on Artificial Intelligence, 2019, 33: 6407–6414.



葛轶洲(1988－), 男, 硕士, 高级工程师, 主要研究领域为信号与信息处理.

徐百乐(1989－), 男, 博士生, 主要研究方向为神经网络, 增量学习.

刘恒(1997－), 男, 硕士生, 主要研究领域为神经网络, 数据分析.

周青(1973－), 男, 研究员, 主要研究领域为通信信号处理.

王言(1997－), 男, 硕士生, 主要研究领域为数据增强, 神经网络.

申富饶(1973－), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为神经计算, 机器人智能.