

DOI:10.16356/j.1005-2615.2022.04.018

小样本图像分类中的类别信息融合网络

张 玉¹, 尚志华², 郭晓楠², 黄福玉³, 刘毅志⁴

(1. 郑州师范学院信息科学与技术学院, 郑州 450044; 2. 中国科学技术大学信息科学技术学院, 合肥 230026;
3. 北京中科院研究院, 北京 100049; 4. 湖南科技大学计算机科学与工程学院, 湘潭 411201)

摘要: 小样本图像分类任务要求模型仅从少量的图像样本中学到新类别的正确分类方法, 是一种特殊的分类任务。然而, 以往大多数小样本工作都单独处理来自不同类别的样本, 而没有充分利用到不同类别间的信息。本文提出了一种新的类别融合网络(Category-fusion network, CFN), 通过融合来自不同类别的样本信息, 同时挖掘类别内和类别间的信息。CFN 的重要部分是一个融合映射的学习, 即如何融合样本中的特征, 从而映射出网络参数。其中的一个重要问题是融合映射是否应该随不同的输入样本而改变。本文设计了 3 个不同的模块: 具有固定映射的类无关模块、融合映射仅依赖于期望学习的目标类别的半相关模块和完全相关的模块, 其中融合映射完全依赖于输入样本。本文的网络可以通过学习多个类别的样本之间的关系来进行类别概念的学习, 并生成融合信息的分类器。实验结果表明, 本文网络在广泛应用的 MiniImageNet 数据集上得到了 60.03% 的分类精度。

关键词: 小样本学习; 图像分类; 类内信息; 类间信息

中图分类号: TP183

文献标志码: A

文章编号: 1005-2615(2022)04-0715-08

Category Information Fusion Network in Few-Shot Image Classification

ZHANG Yu¹, SHANG Zhihua², GUO Xiaonan², HUANG Fuyun³, LIU Yizhi⁴

(1. College of Information Science and Technology, Zhengzhou Normal University, Zhengzhou 450044, China;

2. Department of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China;

3. Beijing Research Institute, University of Science and Technology of China, Beijing 100049, China;

4. Department of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China)

Abstract: Few-shot image classification is a special task where the model learns to build correct concepts of categories from only a few examples. Due to the frequent occurrence of few-shot scenarios, it has aroused extensive research. However, most previous few-shot models process examples from different categories individually without considering inter-classes information. We propose a novel category-fusion network (CFN) to exploit the intra- and inter-classes information simultaneously by fusing the information of examples from different categories. The key part of CFN is the learning of a fusion map, that is, how to fuse the features in the sample to map out the network parameters. There is an important problem that whether the fusion mapping should change with different input examples. To explore this problem, we design three different modules: (1) the class-irrelevant module with a fixed mapping; (2) the semi-relevant module where the fusion mapping only depends on the target category whose knowledge is expected to be learned; (3) the fully-relevant module where the fusion mapping totally depends on input examples. Our network can build the

基金项目: 国家自然科学基金重点项目 (U19B2023); 河南省本科高校青年骨干教师培养计划 (2021GGJS170); 湖南省教育厅科学研究重点项目 (19A172)。

收稿日期: 2021-07-03; **修订日期:** 2021-11-10

通信作者: 尚志华, 男, 博士研究生, E-mail: shangzh@mail.ustc.edu.cn。

引用格式: 张玉, 尚志华, 郭晓楠, 等. 小样本图像分类中的类别信息融合网络[J]. 南京航空航天大学学报, 2022, 54(4): 715-722. ZHANG Yu, SHANG Zhihua, GUO Xiaonan, et al. Category information fusion network in few-shot image classification[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2022, 54(4): 715-722.

concept of a certain category by learning from examples of several categories, and generates a classifier with fused information. The experiments show the effectiveness of our network in few-shot learning, which obtains 60.03% in accuracy on the widely used MiniImageNet dataset.

Key words: few-shot learning; image classification; intra-classes information; inter-classes information

深度学习被广泛应用于各种图像识别任务中并取得了显著的效果提升^[1-6],但是基础的深度学习框架通常需要大规模标注的数据集作为监督,例如 ImageNet 数据集^[7]等,才能得到能够准确分类的模型。然而,在一些领域获取海量数据集的代价很高,尤其是一些安全领域,由于一些类别的图像出现频次低,采集标注难度大,因此这些类别往往可供学习样本数量极少,这种情况下,基础的深度学习方式并不适用。同时,一个已经在特定类别训练良好的深度学习模型在应用到新类别时,通常需要大量的新类别样本,这与人类的学习方式有极大的不同。人类在积累了足够多的知识后,能够仅从少量的图片中学习到新类别的概念,并能准确地识别该类别的其他图片。为了缩小深度学习模型与人之间的这种差距,许多研究人员致力于研究如何用深度学习网络模拟上述的人类学习行为,即小样本图像分类问题。

小样本图像分类问题的目的是让模型能够通过少量的样本学习新的类别,从而能够对新类别的样本进行分类。一些先前的工作在小样本分类中取得了很大的成功^[8-15],Snell等^[10]学习一个嵌入,从而能够使用最近邻或线性分类器对来自新类别的图像进行识别。Sung等^[11]则直接使用基于度量的模型对新类别的查询图像进行分类,无需更新网络参数。Qiao等^[15]利用预先训练的卷积神经网络提取样本的

特征,并利用这些特征生成新的类别分类器。但是,当计算某个类别的参数时,上述工作均只使用该类别中的样本,而忽略其他类别中的样本。

为了在小样本分类任务中更充分地利用不同类别样本中的信息,本文的重点是在前向传播中利用类间信息。其网络框架如图1所示。在所示五分类任务中,从每个类别中选取1张样例图像输入网络,经过特征提取器和融合模块获得融合后的各类别特征,并通过分类器参数预测器生成分类器。之后提取查询图像特征并通过生成的分类器获得预测结果。本文网络中,来自不同类别的样本的信息通过一个可学习的映射融合在一起,通过这种方式,可以利用类间信息,学习到更准确的新类别概念。但是如何对不同的输入样本改变融合映射,以及如何合理地改变融合映射,都是本文需要关注的问题。因此,本文设计了3个不同的融合模块来进行融合映射,从而探讨上述问题:(1)类无关模块。使用常量映射来融合输入样本的特性。通过类无关模块,在训练过程中直接学习融合映射,不因测试样本而改变。(2)半相关模块。学习生成动态映射。当学习一个新的类别时,半相关模块分别为属于该类别和其他类别的输入样本计算两种权值,然后将它们通过权值融合在一起。(3)完全相关模块。计算不同类别样本特征的内积,然后利用这些内积生成融合映射。

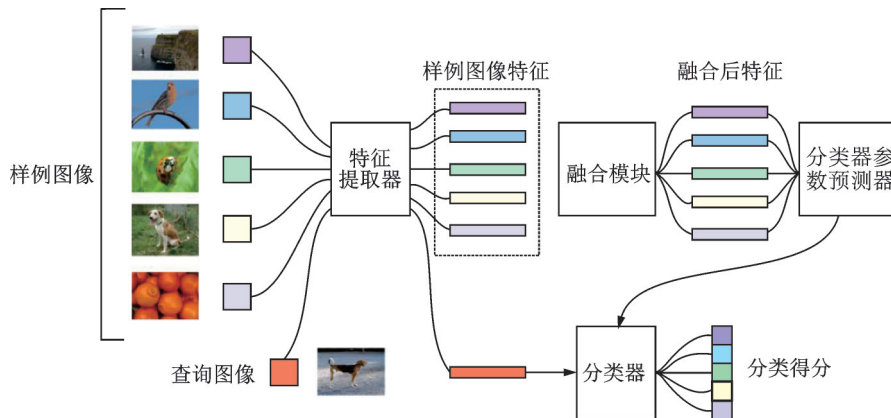


图1 本文网络框架图

Fig.1 Framework of our network

1 相关工作

近年来,许多研究者对小样本学习产生了兴趣。其中,小样本分类问题的目标是让机器学习模

型在学习了一些类别的大量数据后,对于新的类别只需要少量的样本就能快速学习,达到在这些类别上的精确分类。对于一个具体的小样本任务,设其新的类别数为 C ,每个类别的样本个数为 K ,则称

其为 C -way K -shot 任务。许多小样本学习方法^[8,10-12,15]可被归纳为3类:基于元学习的方法、基于优化器的方法和基于度量学习的方法。以上方法都能够从类似的任务中提取一些可转移的知识用于进行新任务。基于元学习的方法。元学习方法^[16]旨在训练1种元学习模型,这种模型可以从1个新任务的几个训练例子中快速学习出1个新的模型。具体来说,MAML方法^[12]学习了1个良好的初始条件,用于对小样本问题进行微调而不会严重过拟合。神经网络的初始权值可以在几个梯度下降步骤中进行微调,以适用于新的分类任务。本文网络是一种元学习网络,用于学习可迁移的融合样本信息和生成分类器的方法。元学习方法的快速泛化能力源自其训练机制,在训练过程中产生的梯度被用来作为快速权重的生成。模型包含1个元学习器和1个基准学习器,元学习器用于学习元学习任务之间的泛化信息,并使用存储机制保存这种信息,基准学习器则用于快速适应新的任务,并和元任务交互产生预测输出。

基于优化器的方法。基于优化器的方法认为普通的梯度下降方法难以在小样本任务的场景下拟合,因此通过调整优化方法来完成小样本分类的任务。文献^[17]提出了在样本较少的情况下,原有的分类任务中基于梯度的优化器算法失效的原因。Finn等^[12]提出了1种新的优化方法,能够学习模型的初始化参数,使得一步或几步迭代后在新任务上的精度最大化。本文方法可以学习任意标准模型的参数,并让该模型能快速适配。Antoniou等^[18]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^[19]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。

基于度量学习的方法。度量学习方法的目的是学习图像的特征表示,并以前馈的方式对查询和样本图像进行分类。原型网络^[10]学习了1种嵌入,这样模型就可以通过计算到样本图像的距离来分类查询图像。关系网络^[11]则学习1个深度距离度量来比较样本图像和查询图像。这里的深度距离度量比起欧几里得距离和余弦距离更加复杂,由卷积神经网络计算得到。TADAM^[8]学习了1个任务相关的度量空间,该度量空间针对不同的任务度量不同尺度的距离。TADAM用向量表示任务,并将其输入到后续网络中,计算相应的度量尺度。Cai等^[20]提出了一种生成式匹配网络,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。该方法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。同时,Cai等^[20]还提出了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Long short-term memory (LSTM)对记忆间隙进行编码,不仅提高了图像特征的表达能力,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度量,相比于余弦距离,计算复杂度更加简单。Zhou等^[21]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。

除了以上常见的3类方法之外,还有一些方法被应用于小样本图像分类任务中,基于迁移学习的小样本图像分类有3种实现方式:基于特征^[22-23]、基于相关性^[23-26]和基于共享参数^[27-28]的方式等。在一些工作中,对偶学习^[21]、贝叶斯学习^[29]和图神经网络^[30-32]等也被用于处理小样本图像分类问题。与本文最相关的方法是Qiao等^[15]中的激活中预测参数。该方法的目的是学习1个参数预测器,以便在给定特定类别的样本图像时,模型能够生成用于查询图像的分类器参数。一般来说,小样本学习是一种多分类的任务。但是文献^[15]中的参数预测器仅从A类别的样本中学习生成A类别的分类器参数,在前向传播中忽略了类间信息。相比之下,本文网络根据来自不同类别的所有样本图像预测分类器的参数,与文献^[15]相比只需要很小的额外计算量。

2 类别信息融合网络模型

在小样本学习任务中,数据集分为两部分,训练集中包含的类别为 C_{train} ,且每个类别的样本比较多,测试集包含的类别为 C_{test} ,每个类别中只包含很少的样本。其中,训练集和测试集的样本类别没有交集,即 $C_{\text{train}} \cap C_{\text{test}} = \emptyset$ 。在 C -way K -shoty 分类任务的每个训练片段中,首先从训练集中随机选择 C 个类别,并在每个类别中随机抽取 K 个有标注的样本。将类别 y 中的 K 个样本的集合表示为 S_y ($y \in \{1, 2, \dots, C\}$),即样本集。这 C 个类别中剩余的一部分样本作为查询集,即 $Q = \left\{ (x_i, y_i) \right\}_{i=1}^Q$ 。在测试时,样本集和查询集以类似的方式从测试集

中采样得到。小样本模型的目标是在每个训练或测试阶段中,从样本集合中学习分类方法并对查询集合中的图像进行分类。用 $A \in \mathbf{R}^{C \times n}$ 表示1个样本的特征矩阵,其中 n 为图像特征的维数, A 中的第 y 行是 S_y 中所有示例特征向量的平均值。 $L \in \mathbf{R}^{C \times n}$ 表示融合特征矩阵,其中第 y 行同时包含第 y 类信息和样本的类间信息。本文工作的重点是学习融合映射 $\phi: A \rightarrow L$ 。正如引言中提到,设计了3个不同的模块来分别学习 ϕ 。

本文算法步骤主要分为两部分:(1)融合特征矩阵的计算,本文设计了多种可选方法,在第2.1节融合映射模块中分别进行了详细描述;(2)基于融合特征矩阵进行分类器参数预测并对查询图像进行分类,在第2.2节类别融合网络中进行了表述。

2.1 融合映射模块

2.1.1 类别无关模块

期望类别无关能够学习到一种通用的、不变的融合映射来进行小样本的学习。同时,在小样本学习中,由于采样的随机性和训练集/测试集的标记空间的差异性,参数模型往往不能很好地从训练样本中学习。因此,设计了一个具有少量额外参数的类别无关模块(Class-irrelevant module, CIM)。CIM通过在 $\mathbf{R}^{C \times C}$ 中的矩阵 W_{CIM} 直接学习融合映射。如图2所示,CIM通过1个全连接层将 A 映射为 L ,即

$$L = W_{\text{CIM}} A \quad (1)$$

CIM中参数较少,但有2个优点:(1)不对实验仪器产生更多的运算要求;(2)可以避免过拟合。此外,CIM中也没有添加更多人为约束。因此,CIM的性能可以直观地说明利用类间信息的有效性。

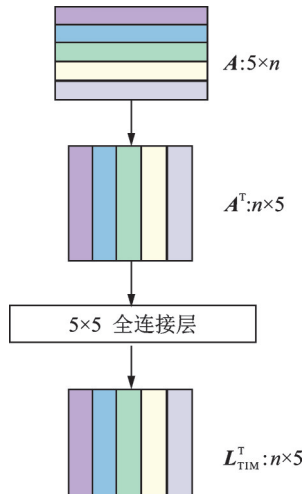


图2 类别无关模块结构图

Fig.2 Illustration of the class-irrelevant module architecture

2.1.2 半相关模块

由于CIM没有考虑融合映射和输入示例之间的关系,因此设计一个半相关模块(Semi-relevant module, SRM)将这种关系进行融合。SRM就是为了学习这种关系并动态生成融合映射而设计的。用 $W_{\text{SRM}} \in \mathbf{R}^{C \times C}$ 表示由SRM生成的融合映射矩阵,在 W_{SRM} 中,每个元素是1个融合权值,它与两个类别相关,分别是输入示例所属的源类别和输出特征所属的目标类别。如果是来自目标类别的样本,它们与目标类别的逻辑关系相同。如果不是来自目标类别的样本,它们与目标类别的逻辑关系相同,在融合中权重相近。因此,在SRM中增加了1个约束,即只考虑融合权值与目标类别之间的相关性来控制参数的数量。如果用完全连接的层来实现SRM,那么约束至少可以减少一半参数。因此,给定样本矩阵 A ,SRM生成两个权重向量 $\alpha, \beta \in \mathbf{R}^{C \times 1}$,并可计算 W_{AFM} 和 L ,即

$$W_{\text{SRM},i,j} = \begin{cases} \alpha_i + \beta_i & i=j \\ \beta_i & i \neq j \end{cases} \quad (2)$$

$$L = W_{\text{SRM}} A = \alpha_{\text{repeat}} \cdot A + \beta A_{\text{arg}} \quad (3)$$

式中: α_i, β_i 分别代表 α 和 β 中的第 i 个元素; $\alpha_{\text{repeat}} \in \mathbf{R}^{C \times n}$, 其每一列为 α ; $A_{\text{arg}} \in \mathbf{R}^{1 \times n}$ 为 A 中每行的平均值; (\cdot) 表示 hadamard 乘积。由于SRM只考虑融合映射与目标类别之间的关系,本文称之为“半相关”,结构图如图3所示。图中 α 的“ \times ”运算表示沿某一维重复 α 以匹配 A 的大小并计算它们的 hadamard 积 β 积下方的“ \times ”运算则表示矩阵乘法。

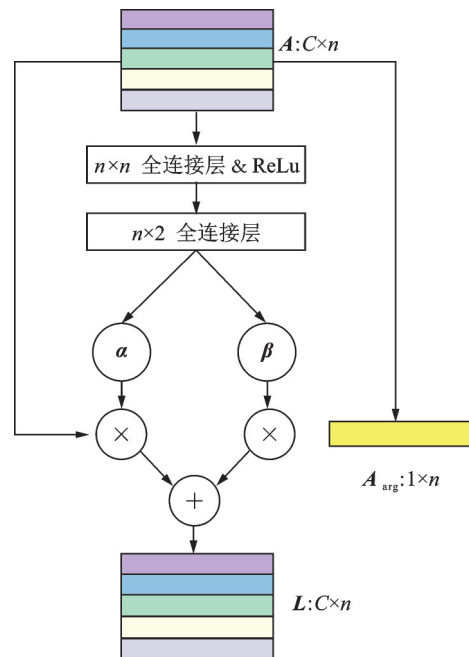


图3 半相关模块结构图

Fig.3 Illustration of the semi-relevant module architecture

2.1.3 全相关模块

SRM 生成带有强约束的融合映射,全相关模块(Fully-relevant module, FRM)则根据没有约束的输入学习融合映射。但是,由于图像特征的维数较高,如果 FRM 直接从层间完全连接图像特征生成融合映射,则 FRM 参数过多,容易导致过拟合。因此,首先计算样本的内积。内积包含了样本之间的关系,用这种方法可以显著减少参数数量,FRM 结构图如图 4 所示。FRM 中计算 L 的方法为

$$L = w(AA^T)A \quad (4)$$

式中 $w(\cdot)$ 为 FRM 生成的融合映射的函数,即图 4 中的 2 个全连接层。本文对全连接层的输入和输出进行了重新设计,以便进行后续计算。

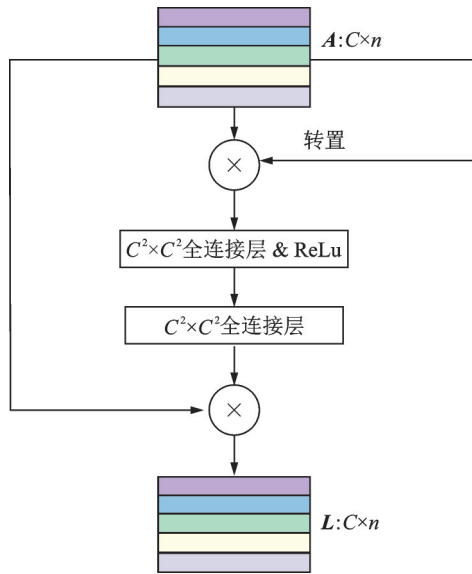


图 4 全相关模块结构图

Fig.4 Illustration of the fully-relevant module architecture

2.2 类别融合网络

2.2.1 参数预测

由于最终目标是分类,因此类别融合网络(Category-fusion network, CFN)期望生成 1 个融合特征的分类器。Qiao 等^[15] 中的研究表明,图像特征与分类器参数具有较高的相关性。同时,CFN 学习的融合特征包含了大量的图像特征信息。因此,CFN 以与 Qiao 等^[15] 类似的方式生成分类器。CFN 通过 Ψ 将 L 映射到 W_L 的线性映射学习预测分类器的参数,其中 L 是最终分类的参数,即

$$W_L = L\Psi \quad (5)$$

$$f(x_i, y) = W_{L,y}a(x_i) \quad (6)$$

式中: $W_{L,y}$ 为类 y 的分类器中的参数; $a(\cdot)$ 为利用卷积神经网络实现的特征提取器。

2.2.2 目标函数

为了训练 CFN,通过最小化分类损失在训练集上训练本文的模型,有

$$L(\phi, \Psi) = \sum_{(x_i, y_i) \in Q} \left(-\phi(A)_y \Psi a(x_i) + \log \sum_{y'=1}^C e^{\phi(A)_{y'} \Psi a(x_i)} \right) \quad (7)$$

式中 $\phi(A)_y$ 为 $\phi(A)$ 中的第 y 列。在实验中,分别使用 CIM、SRM 和 FRM 作为融合映射模块,而 ϕ 代表其中之一。

2.2.3 测试过程

在测试过程中,CFN 首先使用预先训练的特征提取器 $a(\cdot)$ 提取样本集合和查询集中图像的特征向量。然后,CFN 用融合映射 ϕ 对实例进行融合。使用融合特性, Ψ 生成 1 个分类器,该分类器可以对整个查询集合进行分类。

3 实验与分析

3.1 数据集

在 MiniImageNet 数据集上评估了本文方法,该数据集是较大规模的 ILSVRC-15 数据集中的一部分。该数据集由来自 100 个类别的 60 000 张彩色图像组成,其中每个类别中有 600 个样本图像。遵循 Qiao 等^[15] 提出的划分方式,选取 80 个类别用于训练,20 个类别用于测试。在实验中预处理图像像素的方式为:首先,将图像的大小调整为 92×92 ,然后随机裁剪它们,以 80/92 的比例进行训练,并以相同的比例进行测试。

3.2 具体步骤

训练过程中,本文通过使用两种不同的基础网络骨架得到两种特征提取器,分别是简单卷积模块组成的网络^[9] 和宽残差网络(WRN-28-10)^[33]。按照 Qiao 等^[15] 中的方法将这些网络进行了调整以适应 MiniImageNet 数据集。比起 CIFAR 数据集上的网络,本文网络中将 2 倍下采样改为 3 倍下采样,并在最后添加 1 个全局平均池化层。

本文提出了 5 种分类的 CFN 网络。对于 CIM,使用无激活层的单级全连接层,输入和输出维度均为 5,如图 2 所示。SRM 则是由 2 个完全连接的层和 2 个层之间的 ReLu 层组成的顺序网络。第 1 层的输入和输出尺寸与图像特征相同,第 2 层的输出为 α 和 β 。FRM 通过 2 个全连接层实现,其中输入和输出的大小都是 $5 \times 5 = 25$ 。在第 1 个全连接层之后还有 1 个 ReLu 层。使用 1 个 $n \times n$ 的全连接层来实现参数预测器 Ψ 。

在训练过程中,首先在训练集上对网络进行常规的多分类训练。无论在小样本学习还是常规的分类网络学习中,特征提取器的目的都是为了得到具有鉴别性的图像特征。虽然这部分不是关于小样本学习的主要研究,但在实验中证实了其对结果的显著影响。预训练之后,将融合映射 ϕ 和参数预测器 Ψ 一起训练。按照现有的小样本学习工作的常规设置,进行了5-way 1-shot和5-shot分类。在每个训练/测试集中,查询集都由来自每个类别的15张图像组成。即在1次5-way 1-shot实验中,使用 $1 \times 5 = 5$ 个样本来生成分类器,并在 $15 \times 5 = 75$ 张图像上进行分类测试。

3.3 方法对比

在每一个小样本学习步骤中,将在5个类别上验证模型的分类准确性。与Sung等^[11]中一致,最终的准确率平均在600个随机生成的测试集上进行测试。表1对比了本文方法和以往方法的准确性。其中方法栏中上半部分方法使用3.2节中提到的简单卷积网络骨架,下半部分采用的是残差网络作为网络骨架。与其他使用简单卷积网络的方法相比,CIM、SRM和FRM都有较大的提升,FRM在1样本和5样本场景上都获得了最高的精度。这一结果表明,融合实例信息对小样本学习是有效的。在以简单网络为骨架的本文网络中,FRM的性能优于CIM,说明根据输入实例改变融合映射是可行的。但是,在以WRN为骨架时,每个模块对精度的提升要小于简单卷积网络作为骨架的情况。本文认为这应该是由于过拟合的原因,因为WRN-28-10的参数数量是简单骨架网络的几百倍。由于过拟合严重,SRM和FRM

表1 MiniImageNet数据集上小样本的学习精度

Table 1 Few-shot 5-way accuracies on MiniImageNet %

方法	1-shot	5-shot
匹配网络 ^[19]	43.56±0.84	55.31±0.73
元学习LSTM ^[19]	43.44±0.77	60.60±0.71
MAML ^[16]	48.70±1.84	63.11±0.92
关系网络 ^[11]	50.44±0.82	50.44±0.82
Qiao等-简 ^[15]	54.53±0.40	67.87±0.20
本文网络-简-CIM	56.13±0.86	69.11±0.69
本文网络-简-SRM	56.36±0.85	69.34±0.72
本文网络-简-FRM	56.54±0.83	69.74±0.69
TADAM ^[8]	58.5	76.7
Qiao等-WRN ^[15]	59.60±0.41	73.74±0.19
本文网络-WRN-CIM	60.03±0.83	74.15±0.66
本文网络-WRN-SRM	59.88±0.82	73.35±0.70
本文网络-WRN-FRM	59.16±0.83	73.75±0.71

的性能受到限制。然而,CIM仍然达到了最好的性能,证明了CIM的鲁棒性。CIM具有更少的参数,它可以学习较通用的类别无关的融合映射。因此,在过拟合条件下,该算法具有较好的鲁棒性和性能。综上所述,虽然3个模块有不同的优点,但本文提出的融合样本方法的有效性是毋庸置疑的。

3.4 分析与讨论

3.4.1 CIM学习情况分析

CIM的设计思路是在不考虑样本差异的情况下学习融合映射。为了分析CIM学到的内容,将CIM的融合映射矩阵的绝对值显示在表2中。为了便于观察,每个部分的值均进行了除以最大值的归一化。观察可以发现对角线上的值支配着矩阵,这意味着对于每个目标类别,属于这个类的样本对它的影响最大。虽然非对角元素的值较小,但它们的总和也比较可观,这也说明了类间信息对最终分类是有帮助的。此外还可观察到,沿对角线,最大值为1.0,最小值为0.52,这表明CIM并没有平等地融合每个类的样本。推测这是因为每个融合的图像特征都包含了所有类别的信息,CIM学到了一个适合的策略来利用样本信息,而不是平均地融合。

表2 CIM融合映射矩阵绝对值

Table 2 Absolute values of CIM fusion mapping matrix

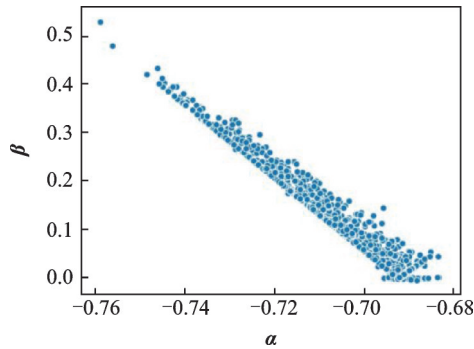
目标类别	源类别					每行中非对角元素之和
	A	B	C	D	E	
A	0.78	0.07	0.05	0.11	0.22	0.45
B	0.09	0.83	0.05	0.11	0.23	0.48
C	0.10	0.06	0.52	0.10	0.22	0.48
D	0.10	0.07	0.03	1.00	0.22	0.42
E	0.08	0.06	0.05	0.12	0.68	0.31

3.4.2 SRM学习情况分析

与CIM不同,SRM采用动态融合映射,而 α 和 β 是理解SRM融合映射特性的关键。图5显示了它们在600个训练步中的分布。结果表明, α 的分布区间较小,而 β 的分布区间较大。这说明该方法可以调整非目标类别实例对融合映射的影响,最终根据实例改变融合映射。此外, α 和 β 分布相对密集,这意味着不同输入的融合映射具有相似性。

3.4.3 分类器中参数的相关性

通过观察对应分类器中参数的相关性来分析这3个模块。由于最终目标是进行分类,所以分类器的参数直接影响结果。在每1训练/测试步中,

图5 SRM中 α 和 β 的分布Fig.5 Distribution of α and β in SRM

每个类别都有1组分类器参数。本文计算了每一集不同参数的平均相关性,并在图6中显示了600步的相关性的核密度估计。显然这3个模块都降低了分类器参数的相关性,这意味着分类器更加关注类别之间的差异,而忽略了多余的相似信息。这个结果是合理的,因为没有类间的信息,分类器不能学习类别之间的差异是什么,只衡量查询图像与样本图像的相似程度。毫无疑问,学习这些额外的知识可以帮助分类器做出正确的预测。同时,从图6中可见,模块之间也存在差异。图中“Base”表示没有融合映射的类别融合网络,FRM的相关性最低,说明FRM学习融合的方法比其他2种方法更能提高分类器的识别性能。

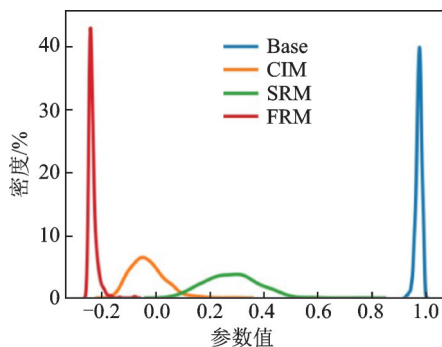


图6 分类器参数相关性的核密度估计

Fig.6 Kernel density estimation of correlation of classifier parameters

4 结 论

本文提出了一种新的类别信息融合网络,该网络可以通过融合样本信息来充分利用样本中的类间信息。本文的网络能够学习前向传播中类别之间的差异,并生成一个更有分辨力的分类器。此外,设计了3个模块,以不同的方式融合不同类别的信息,并讨论了它们各自的优缺点。3种模块间相互独立,可以根据不同任务上的选择单一模块使用,同时3种模块也可以在模型中并行计算,分别计算相应的融合特征及其分类器,最终对多个分类器进行模型融合

得到最终结果。本文网络在MiniImageNet数据集上实现了最先进的性能,在每个新类别1个样本和5个样本场景下分别得到了60.03%和74.15%的分类精确度,超越了基准网络的分类性能。实验结果证明了该方法对类间信息的有效利用。

参考文献:

- [1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Proceedings of Advances in Neural Information Processing Systems. London: MIT, 2012: 1097-1105.
- [2] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//Proceedings of International Conference on Learning Representations. San Diego, USA: ICLR, 2015: 1-5.
- [3] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [4] XIE H, MAO Z, ZHANG Y, et al. Double-bit quantization and index hashing for nearest neighbor search [J]. IEEE Transactions on Multimedia, 2019, 21(5): 1248-1260.
- [5] FANG S, XIE H, ZHA Z J, et al. Attention and language ensemble for scene text recognition with convolutional sequence modeling[C]//Proceedings of the 26th ACM International Conference on Multimedia. [S.l.]: ACM, 2018: 248-256.
- [6] XIE H, YANG D, SUN N, et al. Automated pulmonary nodule detection in CT images using deep convolutional neural networks[J]. International Journal of Computer Assisted Radiology and Surgery, 2019, 85: 109-119.
- [7] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2009: 248-255.
- [8] ORESHKIN B N, RODRIGUEZ P, LACOSTE A. TADAM: Task dependent adaptive metric for improved few-shot learning[C]//Proceedings of Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc, 2019.
- [9] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning[C]//Proceedings of Advances in Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc, 2016.
- [10] SNELL J, SWERSKY K, ZEMEL R S. Prototypical networks for fewshot learning[C]//Proceedings of

- Advances in Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc, 2017.
- [11] SUNG F, YANG Y, ZHANG L, et al. Learning to compare: Relation network for few-shot learning [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018.
- [12] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks [C]//Proceedings of International Conference on Machine Learning. Sydney, NSW, Australia: JMLR, 2017.
- [13] KOCH G, ZEMEL R, SALAKHUTDINOV R. Siamese neural networks for one-shot image recognition [C]//Proceedings of ICML Deep Learning Workshop. [S.l.]: ICML, 2015.
- [14] SANTORO A, BARTUNOV S, BOTVINICK M, et al. Meta-learning with memory-augmented neural networks [C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2016: 1842-1850.
- [15] QIAO S, LIU C, SHEN W, et al. Few-shot image recognition by predicting parameters from activations [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 7229-7238.
- [16] 谭晓阳, 张哲. 元强化学习综述 [J]. 南京航空航天大学学报, 2021, 53(5): 653-663.
TAN Xiaoyang, ZHANG Zhe. Review on meta reinforcement learning [J]. Journal of Nanjing University of Aeronautics & Astronautics, 2021, 53(5): 653-663.
- [17] RAVI S, LAROCHELLE H. Optimization as a model for few-shot learning [C]//Proceedings of 5th International Conference on Learning Representations. Toulon, France: ICLR, 2017: 24-26.
- [18] ANTONIOU A, EDWARDS H, STORKEY A. How to train your MAML [C]//Proceedings of ICLR 2019. New Orleans, USA: ICLR, 2019.
- [19] NICHOL A, SCHULMAN J. Reptile: A scalable metalearning algorithm [J]. arXiv Preprint, arXiv: 1803.02999, 2018.
- [20] CAI Q, PAN Y Y T. Memory matching networks for one-shot image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 4080-4088.
- [21] ZHOU L, CUI P, YANG S, et al. Learning to learn image classifiers with informative visual analogy [J]. arXiv Preprint, arXiv: 1710.06177, 2017.
- [22] CHOI J, KRISHNAMURTHY J, KEMHAVI A, et al. Structured set matching networks for one-shot part labeling [C]//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018. Salt Lake, USA: IEEE, 2018: 3627-3636.
- [23] HARIHARAN B, GIRSHICK R. Low-shot visual recognition by shrinking and hallucinating features [C]//Proceedings of IEEE International Conference on Computer Vision, ICCV 2017. Venice, Italy: IEEE, 2017: 3037-3046.
- [24] LIU Z, LI J, SHEN Z, et al. Learning efficient convolutional networks through network slimming [C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2736-2744.
- [25] DENTON E L, ZAREMBA W, BRUNA J, et al. Exploiting linear structure within convolutional networks for efficient evaluation [C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, MA, USA: MIT Press, 2014: 1269-1277.
- [26] JADERBERG M, VEDALDI A, ZISSERMAN A. Speeding up convolutional neural networks with low rank expansions [J]. Arxiv Preprint, arxiv: 1405-3866, 2014.
- [27] OQUAB M, BOTTOU L, LAPTEV I, et al. Learning and transferring mid-level image representations using convolutional neural networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE Computer Society, 2014: 1717-1724.
- [28] QI H, BROWN M, LOWE D G. Low-shot learning with imprinted weights [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 5822-5830.
- [29] YOON J, KIM T, DIA O, et al. Bayesian model-agnostic meta-learning [C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Toronto, ON, Canada: IEEE, 2018: 7343-7353.
- [30] GARCIA V, BRUNA J. Few-shot learning with graph neural networks [C]//Proceedings of the 6th International Conference on Learning Representation. [S.l.]: ICLR, 2018.
- [31] KIM J, KIM T, KIM S, et al. Edge-labeling graph neural network for few-shot learning [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 11-20.
- [32] LIU Y B, LEE J, PARK M, et al. Learning to propagate labels: Transductive propagation network for few-shot learning [C]//Proceedings of ICLR 2019. [S.l.]: ICLR, 2019.
- [33] ZAGORUYKO S, KOMODAKIS N. Wide residual networks [C]//Proceedings of British Machine Vision Conference. France: [s.n.], 2016: 1-15.