

吉林大学学报(工学版)
Journal of Jilin University(Engineering and Technology Edition)
ISSN 1671-5497,CN 22-1341/T

《吉林大学学报(工学版)》网络首发论文

题目: 基于滑动特征向量的小样本学习方法
作者: 曹洁, 屈雪, 李晓旭
DOI: 10.13229/j.cnki.jdxbgxb20200532
收稿日期: 2020-07-15
网络首发日期: 2020-11-13
引用格式: 曹洁, 屈雪, 李晓旭. 基于滑动特征向量的小样本学习方法. 吉林大学学报(工学版). <https://doi.org/10.13229/j.cnki.jdxbgxb20200532>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于滑动特征向量的小样本学习方法

曹洁^{1,2}, 屈雪³, 李晓旭¹

(1.兰州理工大学 计算机与通信学院, 兰州 730050; 2. 甘肃省城市轨道交通智能运营工程研究中心, 兰州 730050; 3.兰州理工大学 电气工程与信息工程学院, 兰州 730050)

摘要: 针对在小样本学习中, 几个样本的特征图不足以描述整个类特征空间, 导致误分类的问题。本文提出了滑动特征向量神经网络 (slip feature vectors neural network, 简称 SFV), SFV 集合样本局部滑动特征向量构建类特征空间, 利用图片特征向量一类特征向量的度量方式, 分类查询样本。通过融合特征块的边缘信息以及位置结构的相关性, 最大限度地利用了深层特征图信息, 扩充类特征空间。并在几个基线数据集上, 均取得了不错的效果, 尤其在细粒度数据集上, 达到了最佳精度。

关键词: 计算机应用技术; 计算机视觉; 小样本学习; 局部特征; 度量学习

中图分类号: TP183

文献标志码: A

DOI: 10.13229/j.cnki.jdxbgxb20200532

Few-shot learning based on slip feature vectors

CAO Jie^{1,2}, QU Xue³, LI Xiaoxu¹

(1. School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China;
2. Engineering Research Center of Urban Railway Transportation of Gansu Province, Lanzhou 730050, China;
3. School of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China)

Abstract: In the few-shot learning, the extremely limited number of examples per class can hardly represent the class distribution effectively, which is the main reason for misclassification. To tackle this problem, we propose Slip Feature Vectors Neural Network (SFV for short), which used the local slip feature vectors of samples per class to construct the class feature spaces, and also utilized the image-to-class measure to classify the query samples. It expands the class feature space by adding the edge information of feature blocks and correlation of their position and structures to maximize the utilization of the deep feature maps when the sample is extremely limited, which can ease over fitting problem caused by small sample data. Experimental study on benchmark datasets consistently shows its superiority over the related other framework, especially on fine-grained datasets, it achieves state-of-the-art.

Key words: computer application technology; computer vision; few-shot learning; local features; metric learning

收稿日期: 2020-07-15.

基金项目: 国家自然科学基金 (No.61906080, No.61763028) .

作者简介: 曹洁 (1966—), 女, 教授, 硕士. 研究方向: 人工智能. E-mail: caoj@lut.edu.cn

通信作者: 李晓旭(1982—), 女, 副教授, 博士. 研究方向: 计算机视觉与机器学习. E-mail: lixiaoxu@lut.edu.cn.

0 引言

如今，利用神经网络的大数据图像分类可以达到人类的分类水平。但在现实生活中，一些涉及隐私及安全的领域，很难获取大规模标记样本。并且人类具有快速学习的能力，能够利用以前的知识及少量标记样本，学习到新的概念。为了模拟人的分类行为，提出了小样本图像分类^[1](few-shot learning)。由于样本量极少，且需要分类从未见过的新类，因此，如何提高模型的泛化能力并改善过拟合问题是小样本分类需要解决的重要难题。

最近，研究者们开始用迁移学习解决小样本学习问题，通过迁移知识，达到快速分类的目的。因此，小样本学习方法大致可以分为基于元学习和基于度量学习两类。基于元学习方法^{[2][3]}是通过迁移元知识，使模型获得快速学习的能力。2019年提出的baseline^[18]，利用微调预训练模型，达到快速泛化到新类的目的。Finn等人提出的MAML^[2]通过对初始参数进行一步或多步的梯度调整，为模型找到一个好的初始化，来达到仅用少量数据就能快速适应新任务的目的。方法^[3]构建了一个基于LSTM的优化器，用于为模型提供参数更新指引。还有一些元学习方法需要借助外部记忆模块，2018年提出的MM-Net^[9]将支持集的特征存储到外部记忆模块中，以便在推理阶段中使用。同时，构建一个上下文学习器，预测出嵌入模块的参数。但是由于此类方法存在与时间线性相关的隐藏状态，因此很难训练其复杂的内存寻址结构^[9]。基于度量学习^{[4][5][6][7]}的方法则是通过迁移嵌入模块参数和度量函数（比如欧式距离等），使训练器学习支持集和查询样本之间的同类判别规则。Kochetal等人利用两个相同的网络^[10]，分别输入两个样本，判定两个样本是否为同一类。2016年提出的匹配网络联合了注意力机制和记忆模块机制，从而避免了微调^[5]。原型网络^[6]提出类原型的概念，将同类样本特征图的均值作为类原型，通过比较查询样本和各类原型之间的距离，为查询样本分配类别。关系网络^[7]由原型网络衍生而来，用关系模块代替原型网络的欧式距离，达到学习度量方式的目的。

上述方法均是将深层特征图作为一个整体，用于特征比较。而2019年李等人提出的DN4^[4](Deep Nearest Neighbor Neural Network)利用“词袋”的概念，将支持集中同类样本的局部特征描述符集合起来，构成类的局部特征描述符集，采用图片-类的度量方式，通过比较查询样本和支持集中类的局部描述符之间的余弦距离，通过KNN算法，为查询样本分配得分最高的类别。DN4性能的提升主要得益于以下几个方面：

第一，利用局部特征进行分类，缓解了过拟合

的问题。第二，集合同类样本的局部特征构成类级特征，利用同类样本视觉模块的可交换性，即：一个样本可以由同类其他样本的局部特征组装拼接而成，增加了模型的泛化能力。

但无论是前文提到的方法还是DN4方法都忽略了对深层特征图信息的再提取。深层特征图表征了样本的深层特征分布，是影响分类结果的关键因素。DN4中将局部特征看成独立的特征，丢失了深层特征之间位置及结构的相关性。而前文利用全局特征作为比较特征的方法，虽然保留了这种相关性，但是在小样本学习的场景下，几个样本的特征图不足以完整地表示类的特征分布。因此，在尽可能保留深层特征图信息的前提下，扩充样本特征空间使其逼近类特征空间，本文提出了滑动特征向量神经网络（简称SFV）。

本文贡献如下：

- (1) 利用滑动窗口提取滑动特征块，融合特征块之间的边缘信息和位置结构的相关性，充分利用深层特征图。
- (2) 利用滑动特征向量构建类特征空间，扩充类特征空间，使其更逼近类特征分布。
- (3) SFV在多个基准数据集上取得了不错的效果，并在细粒度数据集上取得最佳精度。

1 滑动特征向量神经网络

常见的小样本分类的神经网络大致分为两个模块：深层嵌入模块 Φ 和特征比较模块 φ ，通过比较支持集和查询样本经 Φ 提取的深层特征图，利用 φ 为查询样本分配得分最高的类别。本文为了提取类滑动特征向量，SFV在 Φ 和 φ 中插入了滑动特征生成模块 S 。SFV模型图如图一所示。

SFV由深层嵌入模块 Φ ，滑动特征生成模块 S 和特征比较模块 φ 构成。这三部分整合到一个框架中，采用端到端的训练方式，三个模块之间严格划分，移植性高，滑动特征生成模块 S 可与其它嵌入模块或比较模块拼接。

特征提取模块 Φ ，用于提取支持集和查询样本的深层特征表示。样本经 Φ 得到大小为 $c * h * w$ 的深层特征表示，其中， c 代表通道数， h ， w 分别代表特征图的长和宽。即对于样本 X ： $x \rightarrow \Phi(x) \in R^{c * h * w}$ 。

滑动特征生成模块 S 。 S 将大小为 $c * a * a$ 的滑动窗覆盖到 Φ 得到的深层特征图上，步长为 s 。生成 n 个大小为 $c * a * a$ 的滑动特征块，其中，

$n = (h - a) / s + 1$, 即 $\Phi(X)' = [x_1, x_2 \dots x_n] \in R^{n * c * a * a}$, 其中 x_i 是第 i 个滑动特征块。为了缓解过拟合, 降低相似标准, 对滑动特征块做全局平均池化, 即: $\Phi(X)' = [x_1, x_2 \dots x_n] \in R^{n * c * a * a} \rightarrow [y_1, y_2 \dots y_n] \in R^{n * c}$, 其中 y_i 代表第 i 个滑动特征向量。

特征比较模块 φ , 用于比较类和查询样本的滑动特征向量之间的相似性, 通过 KNN, 得到与查询样本特征向量最相似的 k 个类特征向量, 分别对这 $n * k$ 个相似性分数求和, 从而得到查询样本与该类的相似分数。分别求出 q 和 C 个类的相似分数, 为 q 分配得分最高的类别。即:

$$\varphi(S(q), S(C)) = \sum_{i=1}^n \sum_{j=1}^k d(q_i, C_i^j)$$

其中 , $S(q) = [q_1, q_2 \dots q_n] \in R^{c * n}$ 和

$S(C) = [c_1, c_2 \dots c_{Kn}] \in R^{c * Kn}$ (K 为 C -way K -shot 对应的 K 值) 分别对应查询样本和支持集中类的滑动特征向量, $d(\cdot)$ 为距离函数, SFV 中为余弦距离函数:

$$d(q_i, C_i^j) = \frac{q_i^T C_i^j}{\|q_i\| \cdot \|C_i^j\|}$$

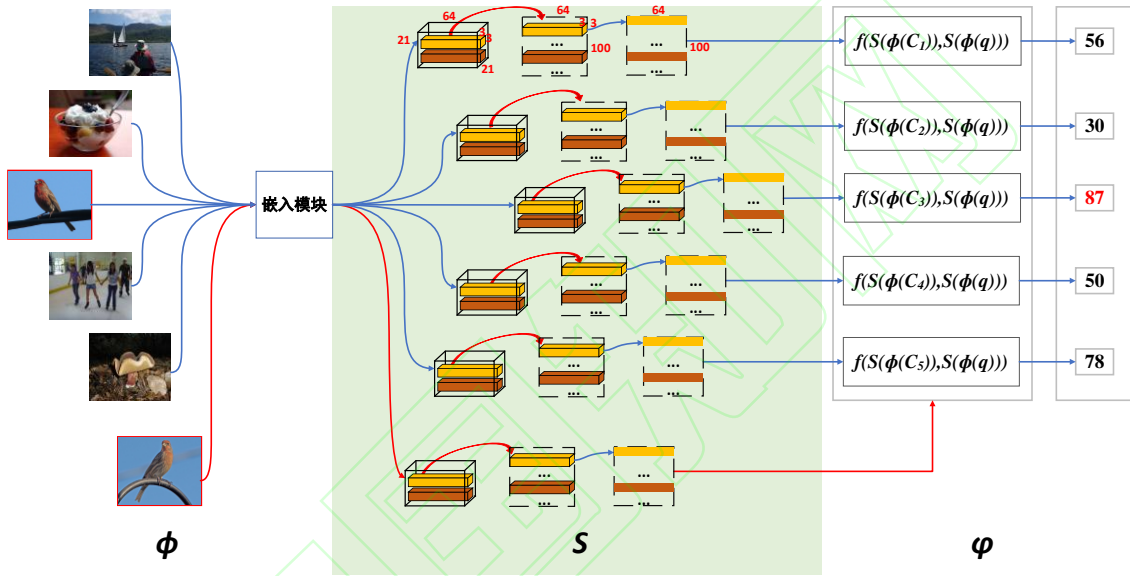


图 1 5-way 1-shot 下滑动特征向量神经网络 (SFV) 结构图

Fig.1 Illustration of proposed SFV for a few-shot learning task in 5-way k-shot setting

因此, 整个框架的流程是每次任务, 将支持集 S 和查询样本 q 输入到嵌入模块 Φ 中, 得到其对应的深层特征图 $\Phi(S)$ 和 $\Phi(q)$, 将 $\Phi(S)$ 和 $\Phi(q)$ 送入滑动特征生成模块 S 中, 得到对应的滑动特征向量, 再送入到特征比较模块 φ 中, 得到 q 和各个类的相似分数, 将相似分数送入到优化器中, 优化整个网络。

2 实验设置

2.1 数据集

本文所有实验均在以下四个数据集上实现。图片的大小设定为 84×84 。

miniImagnet: 常见的通用数据集, 包含 100 个类, 每类 600 张。按照方法^[4]的标准, 将 100 个类分为 64, 16, 20, 分别作为训练集, 验证集, 测试集。

CUB-200^[12]: 常见的细粒度数据集之一, 包含 200 类, 共 6033 张鸟类图片。将 200 个类分为 130, 20, 50, 分别作为训练集, 验证集, 测试集。

Stanford Cars^[13]: 常见的细粒度数据集之一, 包含 196 类, 共 16,185 张车类图片。同样, 将 196 类分为 130, 17, 49 类分别作为训练集, 验证集, 测试集。

Stanford Dogs^[14]: 常见的细粒度数据集之一, 包含 120 类, 共 20,580 张狗类图片。同样, 将 120 类分为 70, 20, 30, 作为训练集, 验证集, 测试集。

2.2 网络结构

SFV 嵌入模块采用了常见的 Conv-64F, 该网络结构包含四个卷积模块, 每个模块中包含一层卷积层, 一个 BN 层, 一个 LeakyReLU 层, 每一层卷积层的通道数均为 64, 卷积核大小为 3×3 。前两个卷积模块中, 加入了 2×2 的最大池化层。

2.3 实验设置

本文在 pytorch 框架下, 基于以上四个数据集进行 5-way K-shot 分类任务。训练策略和方法^[4]相同, 采用分段任务策略, 在训练阶段, 设置 30 个周期, 共 30,000 个分段, 每个分段为一个元任务, 每次任务包含 5 个类, 对 5-way 5-shot 而言, 每类包含 5 个标记样本, 10 个未标记样本, 分别作为支持集和查询集。5-way 1-shot 的支持集中每类有 1 个标记样本, 查询集中每类有 15 个未标记样本。本文使用 Adam 优化算法, 初始学习率为 0.005, 每 10 个周期学习率减半。在测试阶段, 我们以 95% 的置信区间, 在测试集中随机取样 600 个分段, 将 top-1 的平均精度作为测试结果。

表 1 各算法 5-way K-shot 在 miniImagenet 上的平均精度 (95% 的置信区间) (*指的是在相同设置下的复现结果)

Table 1 Mean accuracies of 5-way K-shot tasks on miniImageNe, with 95% confidence intervals (*Results re-implemented in same setting for a fair comparison)

模型		嵌入模块	5-way 1-shot	5-way 5-shot
基线实验	DN4	Conv-64F	51.24±0.74	71.02±0.64
	DN4*	Conv-64F	51.78±0.80	70.15±0.73
	全局特征 KNN*	Conv-64F	44.54±0.76	49.52±0.75
度量学习方法	匹配网络	Conv-64F	43.56±0.84	55.31±0.73
	原型网络	Conv-64F	48.45±0.96	66.53±0.51
	关系网络	Conv-64F	50.44±0.82	65.32±0.70
	SFV	Conv-64F	53.81±0.81	71.98±0.68
基于元学习方法	Baseline*	Conv-64F	36.34±0.58	54.88±0.67
	Meta-Learner LSTM	Conv-32	43.44±0.77	60.60±0.71
	MAML	Conv-32F	48.70±1.84	63.11±0.92
	TADAM	ResNet-12	58.50±0.30	76.70±0.30
	MM-Net	Conv-32F	53.37±0.48	66.97±0.35
	LEO	WRN-28-10	61.76±0.08	77.59±0.12

3 实验结果

3.1 通用数据集分类

常见小样本分类性能测试是在 miniImagenet 上测试精度。实验结果如表 1 所示, 其中, KNN 算法中的 $k=1$ 。表 1 中, 在 5-way 5-shot 上, DN4 和 SFV 比全局 KNN 精度分别高 20.63%、22.46%, 说明利用局部特征构建类特征空间, 扩充特征空间的有效性。此外, SFV 比匹配网络、原型网络、关系网络、DN4 在 5-way 1-shot 上分别提升了 10.25%, 5.36%, 3.37%, 2.03%, 在 5-way 5-shot 上, 也有 16.67%, 5.45%, 6.66%, 1.83% 的提升。原因在于其他的基于度量学习方法使用的是样本级特征, 不足以描述

2.4 基线方法设置

基线实验: 全局特征 KNN 和 DN4。全局特征 KNN 将全局特征图直接放入余弦比较模块中, 用于测试局部特征的有效性, DN4 用于测试滑动特征向量的有效性。

除了两个基线实验, 表 1 还将比较了几个最优的基于度量学习方法和基于元学习方法。本文的方法属于基于度量学习, 同属于此类的方法还有匹配网络^[5]、原型网络^[6]、关系网络^[7]。为了实验的可比性, 本文重新运行了这三种算法, 使得他们的嵌入模块和数据分配方式与 SFV 中的一样。对于基于元学习的方法, 表 1 包括 Meta-Learner LSTM^[3], MAML^[2], TADAM^[15], MM-Net^[9], LEO^[16], baseline^[18]。

类特征空间^[4], 而 SFV 利用滑动特征向量构成的类特征空间分类, 更逼近真实的类特征空间, 缓解了过拟合; 并且还利用了同类样本局部特征的可交换性, 提升模型的泛化能力。SFV 相对于 DN4, 提升原因在于 SFV 进一步扩展了类特征空间, 利用滑动窗口操作, 增加了局部特征块与块之间边缘信息以及位置和结构的相关性, 证明滑动特征块的有效性。

与元学习方法相比, SFV 依然可以取得不错的效果。在 5-way 5-shot 的任务中, SFV 相较于 baseline、Meta-Learner LSTM、MAML、MM-Net 分别提升了 17.1%, 11.38%, 8.87%, 5.01%。baseline、LEO 这类元学习方法性能的提升, 很大一部分来源于预训练网络或是外部记忆存储模块, 而 SFV 采

用端到端的训练方式，不需要预训练。

3.2 细粒度数据集分类

为了测试 SFV 在细粒度数据集上的表现，表 2 列举了各个框架在各个细粒度数据集上的精度。细粒度分类比通用数据集分类难度更大，因为细粒度数据集具有类内距离大，类间距离小的特征。一般来说，细粒度分类需要大量的标记样本训练模型。

由表 2 可以看出，SFV 模型在细粒度数据集上的 1-shot 表现尤其优异，在 CUB-200 上，与度量学习方法相比，SFV 的精度分别比全局 KNN、DN4、匹配网络、原型网络的精度分别高 16.28%，3.53%，12.51%，20.45%。与基于元学习方法相比，SFV 的精度分别比 MAML，baseline 高 11.26%，25.81%；在 5-way 5-shot 中，SFV 依然能保持最高的精度，分别比基于度量学习方法各框架精度高 26.58%，

0.93%，16.09%，30.39%，分别比基于元学习方法的框架精度高 12.39%，24.58%。SFV 相较于全局 KNN、匹配网络、原型网络、MAML、baseline 等利用全局深层特征图分类的方法，精度提升的原因在于利用局部特征构建了类特征空间，在细粒度数据集中，各类样本类间距离小，背景高度相似，分类结果容易受背景噪声的干扰，而 SFV 将深层特征图分解为局部特征，有利于将目标区域和背景区域区分开来，减少这种噪声干扰。而 SFV 相较于 DN4，融合了特征块的边缘信息和位置结构的相关性，扩充了细粒度特征比较空间。另一方面，SFV 扩大了感受野范围，降低了相似的标准，缓解了过拟合。而基于元学习方法的 baseline 虽然利用了预训练，但是本文并没有采取数据增强，baseline 仍然存在比较严重的过拟合，因此，SFV 的精度高于 baseline。

表 2 各算法 5-way K-shot 在各细粒度数据集上的平均精度（95%的置信区间）（*指的是在相同设置下的复现结果）

Table 2 The mean accuracies of the 5-way k-shot tasks on the fine-grained datasets, with 95% confidence intervals (*Results re-implemented in the same setting for a fair comparison)

模型		Stanford Cars		Stanford Dogs		CUB-200	
		5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
基于度量学习方法	全局 KNN	39.08±0.77	41.61±0.71	41.92±0.82	45.91±0.82	41.53±0.81	49.01±0.80
	DN4	59.84±0.80	88.65±0.44	45.41±0.76	63.51±0.62	46.84±0.81	74.92±0.64
	DN4*	60.73±0.85	88.45±0.49	48.90±0.82	69.63±0.74	54.28±0.92	74.66±0.75
	匹配网络	34.80±0.98	44.70±1.03	35.80±0.99	47.50±1.03	45.30±1.03	59.50±1.01
	原型网络	40.90±1.01	52.93±1.03	37.59±1.00	48.19±1.03	37.36±1.00	45.28±1.03
	SFV	65.18±0.85	88.39±0.48	52.19±0.86	69.86±0.72	57.81±0.93	75.59±0.74
基于元学习方法	MAML*	44.34% ±0.81	61.42% ±0.71	43.64% ±0.85	56.62% ±0.75	46.55±0.88	63.20% ±0.79
	Baseline*	28.53% ±0.52	40.41% ±0.58	29.58% ±0.49	42.24% ±0.63	32.00% ±0.58	51.01% ±0.69

3.3 讨论

3.3.1 消融实验

为了验证特征块、滑动操作、全局平均池化的有效性，表三设置了相应方法在 CUB-200 上的消融实验。实验 II 全局 KNN-全局平均池化网络与全局 KNN 网络相比，对深层特征图进行了全局平均池

化。同样，实验 IV 滑动特征块网络与 SFV 相比，没有全局平均池化，实验 III 特征块网络与 IV、V 相比，没有滑动操作。实验 I、II 和 IV、V 两两一对，用于测试全局平均池化对精度的影响，实验 I、III 用于测试利用特征块比较的有效性。实验 III、IV 用于测试滑动操作的有效性。

表 3 SFV 的消融实验在 CUB-200 上 5-way K-shot 的平均精度（95%的置信区间）

Table 3 The ablation study results of the 5-way K-shot tasks on CUB-200, with 95% confidence intervals

模型	CUB	
	5-way 1-shot	5-way 5-shot
全局 KNN 网络(I)	41.53±0.99	49.01±0.80
全局 KNN-全局平均池化网络(II)	41.04±0.99	47.35±0.92
特征块网络(III)	54.55±0.94	69.75±0.8
滑动特征块网络(IV)	56.25±0.92	74.37±0.76
滑动特征向量网络 SFV(V)	57.81±0.93	75.59±0.74

在表 3 中, 由实验 I、II 可得, 对于全局 KNN 网络, 加入全局平均池化会导致精度下降。分别在 1-shot 和 5-shot 任务上, 下降了 0.49% 和 1.66%。但对于由实验 IV、V 可知, 对于滑动特征网络, 加入全局平均池化则可以提高模型精度, 在 1-shot 和 5-shot 任务上, 分别提高了 1.56% 和 1.22%。说明加入全局平均池化会导致特征空间减小, 在类似全局 KNN 网络这类特征量不足方法的情况下, 特征空间的减小, 会导致特征量不足, 精度下降。而 SFV 就是为了解决比较特征量不足提出的, 减小特征空间有助于减小网络弹性, 加快收敛速度。并且全局平均池化可以削弱特征块的细节信息, 降低特征块相似的标准, 从而缓解过拟合。

由实验 III 和 IV 可得, 加入滑动操作分别在 1-shot 和 5-shot 上提升了 1.7% 和 4.62%, 说明了滑动操作有利于提取更多有效特征, 丰富特征空间, 提高模型精度。

3.3.2 滑动特征块大小对实验结果的影响

在 SFV 中, 存在超参数特征块大小 a 。特征块过大, 导致最终特征向量数量少, 不足以描述特征空间。特征块过小, 不仅会造成运算的负担, 还会加重背景等无关特征对结果的影响。因此, 本文在 CUB-200 上进行了滑动特征块大小为 $a \in \{1, 3, 5\}$ 的测试实验, 实验结果如表 4 所示。

表 4 不同超参数 a 值下 SFV 在 CUB-200 上 5-way K-shot 的平均精度

a 的取值	5-way 1-shot	5-way 5-shot
$a=1$	54.28 \pm 0.92	74.66 \pm 0.75
$a=3$	57.81\pm0.93	75.59\pm0.74
$a=5$	57.58 \pm 0.95	73.46 \pm 0.75

表 5 不同 shot 值下各框架在 CUB-200 数据集上关于 5-way K-shot 的平均精度

模型	5-way 1-shot	5-way 3-shot	5-way 5-shot
全局 KNN*	41.53 \pm 0.81	48.71 \pm 0.91	49.01 \pm 0.80
DN4*	54.28 \pm 0.92	68.72 \pm 0.82	74.66 \pm 0.75
SFV	57.81\pm0.93	70.90\pm0.83	75.59\pm0.74

在 5-way 1-shot 实验中, a 为 3 时实验结果最好。而在 5-way 5-shot 实验中, 特征块过大或过小, 都会降低精度。因此, a 值的选择应该由具体任务决定。在本实验中, a 的取值为 3, 考虑到支持集样本

和查询样本目标位置的不一致性, 查询样本中滑动步长为 2, 支持集样本中滑动步长为 1。

3.3.3 shot 数对实验结果的影响

为了验证 SFV 中同类样本中 shot 数对实验结果的影响, 文中测试了在不同 shot 数下, 各个框架在 CUB-200 上 5-way K-shot 任务的实验对比情况, 实验结果如表 5 所示。

在表 5 中, 各框架 5-way K-shot 任务精度均随 shot 数的增加而增加。在 5-way 1-shot 的任务中, SFV 分别比全局 KNN, DN4 的精度高 16.28%, 3.53%, 在 5-way 3-shot 的任务中, SFV 分别比全局 KNN, DN4 的精度高 22.19%, 2.18%, 而在 5-way 5-shot 中, SFV 比全局 KNN, DN4 的精度高 26.58%, 0.93%。说明由于类局部特征空间存在, SFV 和全局 KNN 相比, 更适合 shot 数多的任务, 而由于加入了特征块的边信息和位置结构的相关性, 与 DN4 相比, SFV 更适合 shot 数少的任务。

4 结束语

本文深入探究了利用滑动局部特征向量构建类特征空间的方法, 并提出了滑动特征向量神经网络 (SFV), 重点阐述了滑动特征向量的引入, 可以提高模型的泛化性, 缓解过拟合, 并在各数据集上验证了此说法。但是, 在进行特征空间比较时, 应该尽可能比较判别特征的相似性, SFV 没有侧重于提取判别特征。我们下一步工作旨在利用判别特征构建类特征空间, 缓解过拟合的同时, 提高模型精度。

参考文献

- [1] Wang Y, Yao Q, Kwok J T, et al. Generalizing from a few examples: A survey on few-shot learning[J]. ACM Computing Surveys (CSUR), 2020, 53(3): 1-34.
- [2] C. Finn, P. Abbeel, and S. Levine. Model-agnostic meta-learning for fast adaptation of deep networks[C]// International Conference on Machine Learning (ICML), Sydney, Australia, International Convention Centre, 2017, 70: 1126-1135
- [3] S. Ravi and H. Larochelle. Optimization as a model for few-shot learning[C]// International Conference on Learning Representations (ICLR), San Juan, 2016.
- [4] Li W, Wang L, Xu J, et al. Revisiting local descriptor based image-to-class measure for few-shot learning[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, 2019: 7260-7268.
- [5] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[C]// Neural Information Processing Systems, Barcelona, 2016: 3630-3638.
- [6] J. Snell J, Swersky K, Zemel R. Prototypical networks for

- few-shot learning[C]//neural information processing systems. Long Beach, CA ,2017: 4077-4087.
- [7] 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 ,2018: 1199-1208.
- [8] N. Mishra, M. Rohaninejad, X. Chen, et al. A simple neural attentive meta-learner[C]//International Conference on Learning Representations, Canada,2018.
- [9] Cai Q, Pan Y, Yao T, et al. Memory matching networks for one-shot image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4080-4088.
- [10] Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition[C]//International Conference on Machine Learning, Lille, France ,2015, 2.
- [11] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [12] P. Welinder, S. Branson, T. Mita, et al. Caltech-UCSD Birds 200[J]. Technical Report CNS-TR-2010-001, California Institute of Technology, 2010.
- [13] Krause J, Stark M, Deng J, et al. 3d object representations for fine-grained categorization[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops, Sydney, Australia,2013: 554-561.
- [14] A. Khosla, N. Jayadevaprakash, B. Yao, et al. Novel dataset for fine-grained image categorization: Stanford dogs[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs ,USA,2011,volume 2, page 1
- [15] Oreshkin, B., Lopez, et al. Tadam: Task dependent adaptive metric for improved few-shot learning[C]// Neural Information Processing Systems,Canada,2018,721-731
- [16] Rusu, A. A., Rao, et al. Meta-learning with latent embedding optimization[C]// In International Conference on Learning Representations, New Orleans ,2019.
- [17] Lee, K., Maji, et al. Meta-learning with differentiable convex optimization[C]//In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition , Long Beach, CA,2019,10657-10665
- [18] Wei-Yu Chen Yen-Cheng Liu, Zsolt Kira, et al. A closer look at few-shot classification[C]//International Conference on Learning Representations, New Orleans ,2019