

自监督增广的监督分类学习增强^{*}

耿传兴^{1,2}, 谭正豪^{1,2}, 陈松灿^{1,2}

¹(南京航空航天大学 计算机科学与技术学院, 江苏 南京 211106)

²(模式分析与机器智能工信部重点实验室 (南京航空航天大学), 江苏 南京 211106)

通信作者: 陈松灿, E-mail: s.chen@nuaa.edu.cn



摘要: 借助预置任务创建的免费监督信号/标记, 自监督学习 (SSL) 能学得无标记数据的有效表示, 并已在多种下游任务中获得了验证. 现有预置任务通常先对原视图数据作显式的线性或非线性变换, 由此形成了多个增广视图数据, 然后通过预测上述视图或变换的对应标记或最大化视图间的一致性达成学习表示. 发现这种自监督增广 (即数据自身与自监督标记的增广) 不仅有益无监督预置任务而且也有益监督分类任务的学习, 而当前鲜有工作对此关注, 它们要么将预置任务作为下游分类任务的学习辅助, 采用多任务学习建模; 要么采用多标记学习, 联合建模下游任务标记与自监督标记. 然而, 下游任务与预置任务间往往存在固有差异 (语义, 任务难度等), 由此不可避免地造成二者学习间的竞争, 给下游任务的学习带来风险. 为挑战该问题, 提出一种简单但有效的自监督多视图学习框架 (SSL-MV), 通过在增广数据视图上执行与下游任务相同的学习来避免自监督标记对下游标记学习的干扰. 更有意思的是, 借助多视图学习, 设计的框架自然拥有了集成推断能力, 因而显著提升了下游分类任务的学习性能. 最后, 基于基准数据集的广泛实验验证了 SSL-MV 的有效性.

关键词: 机器学习; 监督分类学习; 自监督学习; 多视图学习; 数据增广

中图法分类号: TP181

中文引用格式: 耿传兴, 谭正豪, 陈松灿. 自监督增广的监督分类学习增强. 软件学报, 2023, 34(4): 1870–1878. <http://www.jos.org.cn/1000-9825/6433.htm>

英文引用格式: Geng CX, Tan ZH, Chen SC. Self-supervisedly Augmented Supervised Classification Enhancement. Ruan Jian Xue Bao/Journal of Software, 2023, 34(4): 1870–1878 (in Chinese). <http://www.jos.org.cn/1000-9825/6433.htm>

Self-supervisedly Augmented Supervised Classification Enhancement

GENG Chuan-Xing^{1,2}, TAN Zheng-Hao^{1,2}, CHEN Song-Can^{1,2}

¹(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

²(MIT Key Laboratory of Pattern Analysis and Machine Intelligence (Nanjing University of Aeronautics and Astronautics), Nanjing 211106, China)

Abstract: With the free supervised signals/labels created by pretext tasks, self-supervised learning (SSL) can learn effective representation from unlabeled data, which has been verified in various downstream tasks. Existing pretext tasks usually first perform explicit linear or nonlinear transformations on the original view data, thus forming multiple augmented view data, then learn the representation by predicting the corresponding transformations or maximizing the consistency among the above views. It is found that such self-supervised augmentations (i.e., the augmentations of the data itself and self-supervised labels) benefit the learning of not only the unsupervised pretext tasks but also the supervised classification task. Nevertheless, few work focus on this at present, while existing works either take the pretexts as the auxiliary of downstream classification task and adopt the multi-task learning or jointly model the downstream task labels and self-supervised labels in a multi-label learning way. Actually, there are inherent differences between downstream and pretext tasks

* 基金项目: 国家自然科学基金 (62076124, 62106102); 中央高校基本科研业务费 (NJ2020023)

收稿时间: 2021-06-07; 修改时间: 2021-07-17; 采用时间: 2021-08-16; jos 在线出版时间: 2022-09-30

CNKI 网络首发时间: 2022-11-15

(e.g., semantic, task difficulty, etc.), which inevitably result in the competitions between them and bring risks to the learning of downstream tasks. To challenge this issue, this study proposes a simple yet effective SSL multi-view learning framework (SSL-MV), which avoids the learning interference of self-supervised labels on downstream labels through performing the same learning as downstream tasks on the augmented data views. More interestingly, with the multi-view learning, the proposed framework naturally owns the integration inference ability, which significantly improves the performance of downstream supervised classification tasks. Extensive experiments on benchmark datasets demonstrate the effectiveness of SSL-MV.

Key words: machine learning; supervised classification learning; self-supervised learning (SSL); multi-view learning; data augmentation

基于强大的表示学习能力,深度卷积神经网络(CNNs)^[1]近年来已在计算机视觉和自然语言处理等领域取得突破性进展。当前,训练一个能力强的CNNs模型通常需要大量标记数据,然而在各种实际任务中,收集这样的数据既耗时代价又高。相比之下,无标记数据既廉价又易得。因此,如何有效利用它们一直是机器学习研究的热点之一^[2]。最近,一种新的无监督表示学习范式——自监督学习(SSL)^[3]广受关注,其利用预置任务提供的免费监督信号来从无标记数据中学得有效表示,并已在图像分类、图像分割、变化点检测等多种下游任务上获得有效验证^[4]。

从多视图视角考察^[5-7],现有预置任务通常先对原视图数据作显式的线性或非线性的变换,由此形成了原数据多个所谓的增广视图,然后通过预测上述视图或变换的对应标记或最大化视图间的一致性达成学习表示,旨在借助这些变换或视图自身提供的免费监督信号来引导表示的学习。基于此,许多预置任务已被提出,它们大致可分为基于变换的预置任务(如旋转^[8]、拼图^[9]等)和基于对比的预置任务^[10-12]。另外,也有一些基于聚类的预置任务^[13,14],它们首先基于当前模型进行聚类,然后使用聚类索引分配标记,通过迭代地执行此过程来提升表示的质量。特别地,由于易于实现及计算的高效,基于变换的预置任务已广泛用于多种学习场景,如域适应^[15]、鲁棒学习^[16]、分布外检测^[17]等。本文也将此种类型的预置任务作为主要研究对象。下面以预测旋转^[8]为例简单介绍基于变换的预置任务,该任务先对原视图数据执行旋转变换 $\{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$ 以生成多组变换数据,每组数据可看作原视图数据的一个增广视图,然后预测所有这些数据的变换,即自监督标记,我们称这里的增广为自监督增广(数据自身与自监督标记的增广)。

请注意,相比于自监督增广,传统数据增广技术^[18]仅通过各种变换显式地生成更多的训练样本来扩充训练集,并未有效利用变换自身所附带的监督信号(自监督标记)。在监督分类场景中,这些增广的样本通常会被分配与其原样本相同的标记。然而,如果这些增广数据相对原数据产生较大的分布差异,再强制这种标记的不变性可能存在损害性能的风险,如数字图像分类 $\{6 \text{ vs } 9\}$,作旋转增广。我们发现,若将自监督标记引入下游任务则可有效降低前述风险。换句话说,自监督增广不仅有益无监督预置任务而且也有益监督分类任务^[19]的学习,而当前鲜有工作对此关注,现有工作或将预置任务作为下游分类任务的学习辅助,采用多任务学习建模;或采用多标记学习,联合建模下游任务标记与自监督标记^[20]。然而,下游任务与预置任务间往往存在固有差异(语义、任务难度等),由此不可避免地造成二者学习间的竞争,给下游任务的学习带来风险。

为挑战该问题,本文进一步探究了自监督增广的监督分类学习增强,并提出一种简单但有效的自监督多视图学习框架(SSL-MV),其隐式地将自监督标记纳入下游任务模型的构建,在利用自监督增广的同时有效缓解预置任务与下游任务间的竞争。具体地,SSL-MV通过在增广数据视图上执行与下游任务相同的学习来避免自监督标记对下游标记学习的干扰。更有意思的是,借助多视图学习,我们的框架自然拥有了集成推断能力,显著提升了下游分类任务的学习性能。图1展示了所提学习框架与现有框架的区别,具体地,本文贡献总结如下。

(1) 提出一种简单但有效的自监督多视图学习框架(SSL-MV),其隐式地将自监督标记纳入下游分类任务模型的构建,在利用自监督增广的同时有效缓解了预置任务与下游任务间的竞争。

(2) 特别地,借助多视图学习,SSL-MV自然拥有了集成推断能力,显著提升了下游分类任务的学习性能。

(3) 基准数据集上的广泛实验验证了所提学习框架的有效性。

本文第1节介绍现有利用自监督增广增强监督分类性能的方案。第2节具体描述本文提出的自监督多视图学习框架SSL-MV。第3节在基准数据集上进行了实验,验证了本文方法优于现有方法。最后总结全文,并对未来工作进行展望。

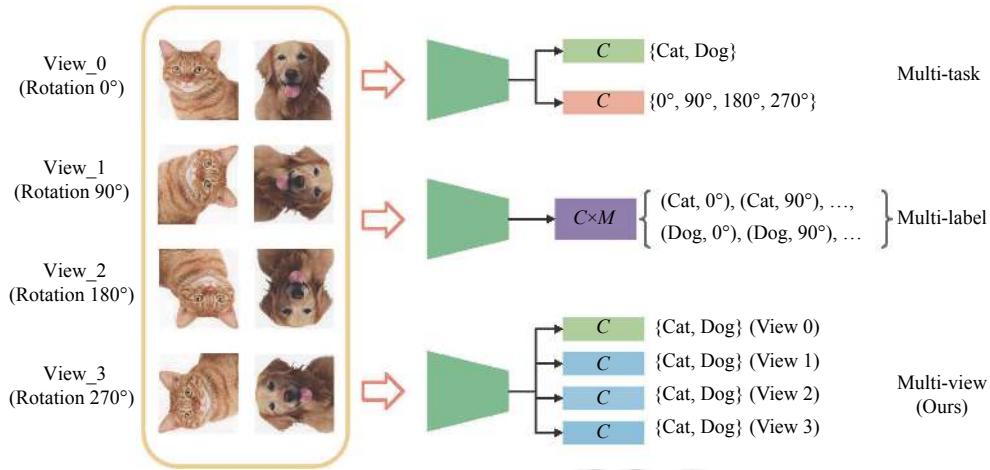


图 1 所提学习框架与现有学习框架的差异

1 相关工作

令 $D = \{(x_i, y_i)\}_i^N \subseteq R^{N \times d}$ 表示训练集, 其中 $y_i \in \{1, 2, \dots, C\}$ 是对应的标记, C 是类标记的个数. 令 $z = f(x; \theta)$ 表示 CNNs 特征提取器, 其中 θ 是网络参数. 令 L_{CE} 表示交叉熵损失函数, $\sigma(\cdot; u)$ 表示 Softmax 分类器, 即 $\sigma_i(z; u) = \exp(u_i^T z) / \sum_k \exp(u_k^T z)$. 令 \tilde{x}_j 表示作用变换 T_j 后的增广样本, 其中 $j = 0, 1, \dots, M$ (自监督标记), T_0 表示恒等变换.

将 SSL 用于辅助监督分类学习的一种简单且直接的方案就是将 SSL 看作一种数据增广策略, 像传统增广技术一样显式扩充训练集, 其损失函数为:

$$L_{DA}(x, y; \theta, u) = \frac{1}{M} \sum_{j=1}^M L_{CE}(\sigma(\tilde{z}_j; u), y) \tag{1}$$

其中, $\{T_j\}_{j=1}^M$ 是预定义变换, $\tilde{x}_j = T_j(x)$ 是作用变换 T_j 后的增广视图样本, $\tilde{z}_j = f(\tilde{x}_j; \theta)$ 是其通过 CNNs 后得到的嵌入表示, $\sigma(\cdot; u)$ 是监督分类器. 可见, 优化上述损失将迫使分类器 $\sigma(\cdot; u)$ 对变换 $\{T_j\}_{j=1}^M$ 保持不变. 然而, 通过某些变换产生的增广 (视图) 样本可能与原训练样本的统计特征存在显著差异, 即数据分布产生非常大的漂移, 因此一味强制分类器的这种不变性可能没有意义, 甚至损害分类器性能, 例如数字图像分类 {6 vs. 9}, 作旋转增广. 我们发现, 若将自监督标记也同时引入下游任务的建模则该风险能有效降低. 事实上, 已有少量工作开始相关探究.

1.1 自监督增广多任务学习框架

如前所述, 在基于变换的预置任务中, 模型通常学习预测何种变换 (自监督标记) 作用于给定的样本. 因此, 将自监督标记引入下游分类任务建模的一种直接方式就是采用多任务的学习框架, 我们这里称其为自监督多任务学习框架^[20], 其在下游分类任务与预置任务间共享特征空间的同时, 对这两个任务的损失进行优化.

$$L_{MT}(x, y; \theta, u, v) = \frac{1}{M} \sum_{j=1}^M L_{CE}(\sigma(\tilde{z}_j; u), y) + L_{CE}(\sigma(\tilde{z}_j; v), j) \tag{2}$$

其中, $\sigma(\cdot; u)$ 和 $\sigma(\cdot; v)$ 分别表示下游监督分类任务与预置任务的分类器. 请注意, 损失函数的第 2 项可看作第 1 项分类损失的一个额外约束, 因此其仍会隐式地迫使分类器习得对变换保持不变的能力, 依旧存在前述损害分类性能的风险^[20].

1.2 自监督增广的多标记学习框架

为了进一步消除这种不变性对下游任务分类器的影响, 有学者^[20]开始尝试采用多标记学习的方式联合建模

下游任务的类标记与自监督标记, 我们这里称其为自监督多标记学习框架. 具体地, 他们构建了一个联合的 Softmax 分类器 $\rho(\cdot; w)$ 来表示联合概率 $P(i, j|\bar{x}) = \rho(w_{ij}^T \bar{z}) / \sum_{k,l} \exp(w_{kl}^T \bar{z})$, 整个目标函数如下:

$$L_{SLA}(x, y; \theta, w) = \frac{1}{M} \sum_{j=1}^M L_{CE}(\rho(\bar{z}_j; w), (y, j)) \tag{3}$$

其中, $L_{CE}(\rho(\bar{z}; w), (i, j)) = -\log \rho_{ij}(\bar{z}; w)$. 在推断时, 由于事先知道何种变换被应用, 因此该建模方式无需考虑所有 $N \times M$ 种标记来预测原始标记, 仅需使用条件概率预测 (SLA-SI) 即可.

$$P(i|\bar{x}_j, j) = \exp(w_{ij}^T \bar{z}_j) / \sum_k \exp(w_{kj}^T \bar{z}_j),$$

其中, $\bar{z}_j = f(\bar{x}_j)$. 此外, 对于所有变换 $\{T_j\}_{j=1}^M$, 该学习框架也允许一个集成推断 (SLA-SG).

$$P_{\text{aggregated}}(i|x) = \frac{\exp(s_i)}{\sum_{k=1}^C \exp(s_k)}, \text{ 其中 } s_i = \frac{1}{M} \sum_{j=1}^M w_{ij}^T \bar{z}_j.$$

2 自监督增广多视图学习框架

从优化的视角看, 尽管 L_{MT} 与 L_{SLA} 都显式建模了下游分类任务的类标记与自监督标记, 且相比增加额外约束的 L_{MT} , L_{SLA} 更易优化, 但由于预置任务与下游分类任务间固有的差异 (语义、任务难度等), 同时优化这两个任务 (即预测下游分类标记与预测自监督标记) 不可避免地会存在竞争, 由此给下游任务的学习带来风险. 为挑战该问题, 本文另辟蹊径, 尝试隐式地将自监督标记纳入下游分类模型的构建, 旨在有效利用自监督标记的同时避免对其优化, 以缓解因语义、任务难度差异过大带来的负面竞争. 我们具体通过在增广数据视图上执行与下游任务相同的学习来避免自监督标记对下游标记学习的干扰, 并提出了一个简单但有效的自监督多视图学习框架 (SSL-MV), 如图 2 所示, 其由一个共享的 CNNs 特征提取器 $f_{\text{shared}}(x; \theta)$ 和来自不同增广数据视图独立的分类器 $\phi_j(\cdot; \vartheta_j)$ 构成, 我们定义损失函数为公式 (4).

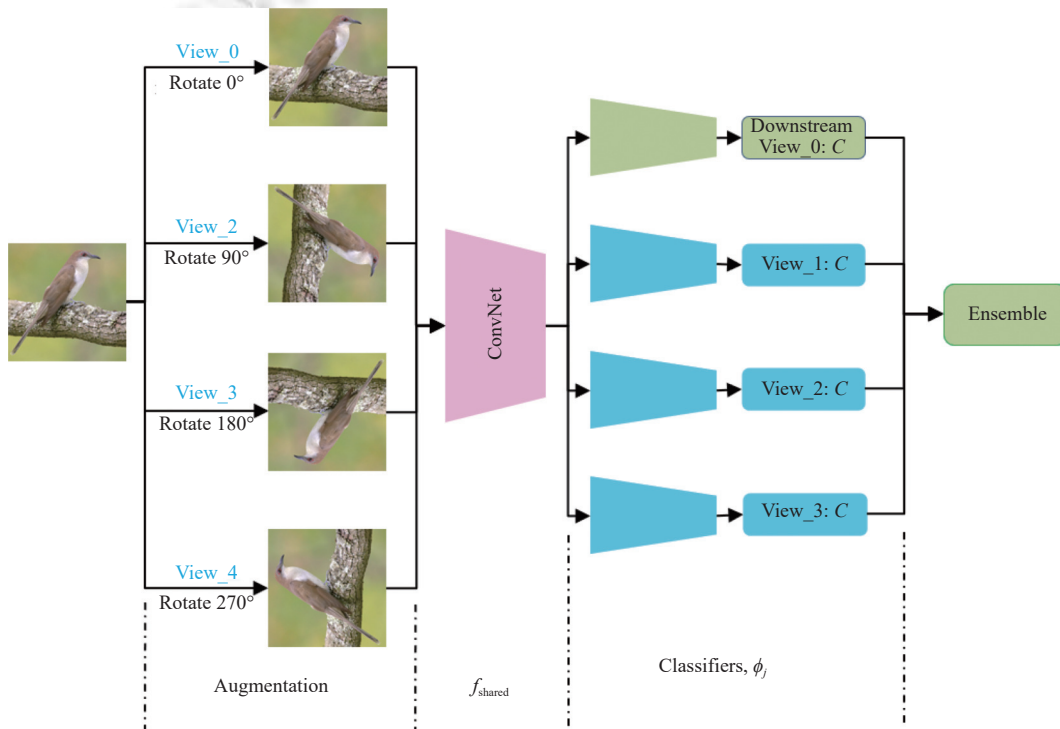


图 2 SSL-MV 学习框架概述 (以旋转为例)

$$L_{MV}(x, y; \theta', \theta) = \frac{1}{N} \sum_{j=0}^M \sum_{i=1}^N L_{CE}(\varphi_j(f_{\text{shared}}(\tilde{x}_j; \theta'); \theta_j), y) \quad (4)$$

直觉上, 这样处理使得下游任务与增广数据视图上的任务在语义或任务难度层面基本处在同层级, 由此会促发一种良性竞争学习, 因而显著提升下游任务的性能. 在 SSL-MV 训练完成后, 我们考虑了两种推断, 一种是单独推断, 我们称之为 SSL-MV(s) (实验结果见第 3.3 节); 此外, 借助多视图学习框架, SSL-MV 本质具备集成推断的能力, 我们称之为 SSL-MV(e). 有意思的是, 简单集成来自不同增广视图的预测结果, SSL-MV(e) 在多个基准数据集上的性能大幅度超越现有方法, 获得新 SOTA 性能 (实验结果见第 3.4 节).

3 实验与结果

3.1 实验设置

3.1.1 数据集与模型

参照文献 [20,21], 在 4 个常用的监督分类基准数据集上评估了我们的学习框架, 它们是 CIFAR10/100^[22], TinyImageNet^[23] 和一个细粒度数据集 MIT67^[24]. 另外, 表 1 描述了这些数据集的具体细节. 对于所用 CNNs 模型, 为保证实验的公平性, 同样参照文献 [20,21]: ResNet32 用于处理 CIFAR10/100, ResNet18 用于处理 TinyImageNet 和 MIT67.

表 1 实验数据集

数据集	类别数	训练集	测试集
CIFAR10	10	50000	10000
CIFAR100	100	50000	10000
TinyImageNet	200	100000	10000
MIT67	67	5360	1340

3.1.2 训练细节

对于所有基准数据集, 使用随机梯度下降 SGD 的优化方法, 学习率参数设置为 0.1, 动量参数设置为 0.9. 对于 CIFAR10/100, 设置权重衰减参数为 0.0005, 批次大小设置为 128, 训练 200 轮. 对于 TinyImageNet 权重衰减参数设置为 0.0001, 批次大小设置为 64 (受限于 GPU 算力), 训练 100 轮. 对于 MIT67, 权重衰减参数设置为 0.0005, 批次大小设置为 32, 训练 180 轮. 请注意, 为保证实验公平, 这里设置的批次大小及训练轮数均参照文献 [20]. 此外, 对所有数据集, 在训练轮数的 50% 和 75% 时, 分别将学习率参数衰减为原来的 10%.

3.1.3 预置任务的选择

本文主要关注的是监督分类任务, 因此将整幅图像作为输入非常重要. 参考文献 [20], 这里重点关注了两种预置任务, 即旋转和色彩置换. 对于旋转, 采用了 4 种常用的变换, 即 $\{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$; 对于色彩置换, 采用了 6 种变换, 即 $\{RGB, RBG, GRB, GBR, BRG, BGR\}$. 图 3 可视化了这些变换作用在原图像上的示例.

3.2 共享 CNNs 模块 (block) 实验

参照文献 [20], 本文采用两种 ResNet 网络架构^[25]构建 SSL-MV 学习框架, 其中 ResNet32 (用于 CIFAR10/100) 有 3 个 CNNs 模块 (block), ResNet18 (用于 TinyImageNet 与 MIT67) 有 4 个 CNNs 模块. SSL-MV 在所有视图上的分类任务均共享同一个 CNNs 特征提取器 f_{shared} , 因此一个自然的问题就是如何选取共享的 CNNs 模块. 为处理该问题, 本节探究了共享不同数目 CNNs 模块对性能的影响. 以 CIFAR100 为例, 图 4 展示了采用不同变换后的实验结果. 可以看到, 无论旋转还是色彩置换, 共享 3 个 CNNs 模块均因不同视图任务间的竞争而造成性能的显著降低. 对于旋转, 共享 2 个 CNNs 模块要略优于共享一个, 而对于色彩置换, 得到的结果刚好与前者相反, 这从侧面表明, 不同变换对原数据分布扰动是不一样的. 另外, 已有研究表明, CNNs 的浅层通常能捕获较通用的底层特征^[26], 如边缘、角等, 而其深层则可捕获与任务相关的高层 (语义) 特征. 因此, 考虑到这些变换对数据分布扰动带来的不确定性及 CNNs 网络特性, 为获得更通用的学习框架, 本文 SSL-MV 的所有实验均仅共享一个 CNNs 模块.

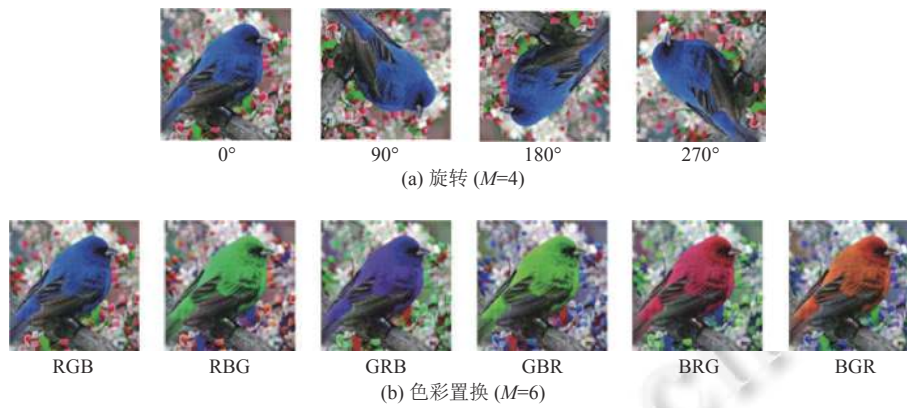


图3 旋转及色彩置换示例可视化

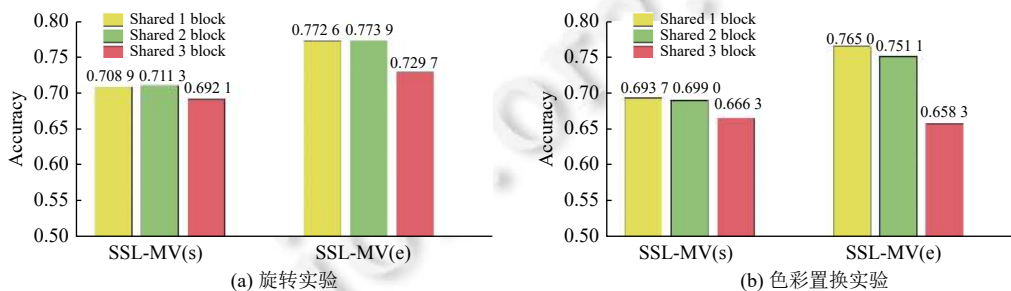


图4 在 CIFAR100 上的共享 CNNs 模块实验

3.3 单独推断实验

本节为展示 SSL-MV 单独推断的优势, 我们与现有方法进行了对比, 其中 Baseline 表示传统监督分类方法, DA 表示仅利用了自监督数据增广的方法, MT 表示自监督多任务学习方法, SLA+SI^[20]表示自监督多标记学习方法 (单独推断). 表 2 以旋转变换为例, 报告了 SSL-MV(s) 与上述这些方法的对比结果. 可以发现以下结论.

表 2 单独推断实验结果 (以旋转为例) (%)

数据集	Baseline	DA	MT	SLA+SI	SSL-MV(s)
CIFAR10	92.39	90.44	90.79	92.50	93.47
CIFAR100	68.27	65.73	66.10	68.68	70.89
TinyImageNet	63.11	60.21	58.04	63.99	64.37
MIT67	54.75	—	—	—	58.28

(1) 旋转变换确实一定程度上使原数据分布发生较大的分布漂移, 如若对增广数据仍强制对变换 (自监督) 标记的不变性, 如 DA 或 MT, 则会显著损害下游分类任务的性能.

(2) 从优化层面, 相比 MT 需要一个额外的变换不变性约束, SLA+SI 更易优化因而更具优势, 故性能也得到显著提高.

(3) 尽管 SLA+SI 获得了较好的性能, 但下游分类任务与预置任务间固有的差异, 如语义、任务难度等, 二者间的学习 (预测下游任务的类标记与预测自监督标记) 竞争仍存在损害下游分类任务性能的风险. 而我们的 SSL-MV 通过在增广数据视图上执行与下游分类任务相同的任务, 隐式地将自监督标记纳入下游分类模型的构建, 在有效的利用自监督标记的同时, 大大降低了学习任务间的竞争. 更有意思的是, 因 SSL-MV 所有视图任务在语义、任务难度等方面基本处于同一层级, 不同视图任务在学习的同时获得一种有益竞争, 可见我们单独推断的 SSL-MV(s) 在所有基准数据集上均获得了最优性能.

3.4 集成推断实验

本节评估了 SSL-MV 的集成推断能力 (SSL-MV(e)). 请注意, 尽管 MT 也将自监督标记纳入下游分类模型的构建, 但由于预测目标的独立性, 本身缺乏集成推断的能力. 因此, 本节重点与 SLA 的集成推断 (SLA-AG)^[20]进行了对比. 表 3 报告了实验结果, 可以发现以下结论.

表 3 集成推断实验结果 (%)

数据集	变换	Baseline	SSL-MV(s)	SLA+AG	SSL-MV(e)
CIFAR10	旋转	92.39	93.47	94.50	95.18
CIFAR100		68.27	70.89	74.14	77.26
TinyImageNet		63.11	64.37	66.95	69.51
MIT67		54.75	58.28	64.85	61.49
CIFAR10	色彩置换	92.39	93.52	92.51	95.01
CIFAR100		68.27	69.37	69.14	76.32
TinyImageNet		63.11	64.25	64.15	70.51
MIT67		54.75	57.24	59.99	60.30

(1) 集成推断相对于单独推断具有显著的优势, 例如在所有数据集及其相应变换下, SSL-MV(e) 领先 SSL-MV(s) 平均约 4.27 个百分点.

(2) 对于 CIFAR10, 在旋转变换下, 我们的 SSL-MV(e) 略优于 SLA-AG (95.18% vs. 94.50%), 但在色彩置换下, SSL-MV(e) 显著领先 SLA-AG (95.01% vs. 92.51%), 这表明相比 SLA-AG, SSL-MV(e) 更能有效利用自监督增广所带来的数据多样性, 这也启发我们进一步增加变换的丰富性 (即提高增广样本的多样性) 可能会进一步提高集成推断的性能. 为验证该论断, 我们在第 3.5 节专门进行了组合变换实验.

(3) 尽管在 MIT67 旋转变换的设置下, SSL-MV(e) 分类性能低于 SLA-AG, 但在其他数据集及其相应的变换下, SSL-MV(e) 均取得了性能上的显著优势, 如 CIFAR100 与 TinyImageNet, 在旋转变换下 (4 个视图), SSL-MV(e) 领先 SLA-AG 近 3 个百分点, 而在色彩置换下 (6 个视图), SSL-MV(e) 更是领先后者约 6 个百分点.

3.5 组合变换实验

本节通过变换的组合增加增广数据视图的多样性, 来验证第 3.3 节的论断. 为此, 我们额外引入一种常用的图像增强变换, 即图像锐化 (Sharpness(γ)), 其中参数 γ 控制着图像的锐化程度, 当 $\gamma=0$ 时返回一张模糊的图像, $\gamma=1$ 时返回原图像. 具体地, 我们采用了 $\gamma=0, 0.5, 1, 1.5$ 这 4 种锐化变换. 表 4 报告了相应的实验结果, 可以看到, 变换的不同, 生成数据视图数目的不同均会影响性能, 但我们的 SSL-MV(s) 均显著优于 Baseline, 这再次表明本文的学习框架有效缓解了预置任务与下游分类任务的竞争. 更重要的是, 随着视图数目的增多 (即视图数据多样性的增强), 我们集成推断 SSL-MV(e) 的性能得到持续性的提高, 充分验证了我们的论断.

表 4 组合变换实验结果

旋转	锐化	视图数	SSL-MV (%)	
			SSL-MV(s)	SSL-MV(e)
0°	$\gamma=1$	1	68.27 (Baseline)	
0°, 180°	$\gamma=1$	2	70.87	75.08
0°, 90°, 180°, 270°	$\gamma=1$	4	70.89	77.26
0°	$\gamma=0, 0.5, 1, 1.5$	4	69.36	75.87
0°, 180°	$\gamma=0, 0.5, 1, 1.5$	8	69.39	77.69
0°, 90°, 180°, 270°	$\gamma=0, 0.5, 1, 1.5$	16	69.17	77.72

3.6 与独立集成的对比

由第 3.3 节可以看到, 本文的集成推断 SSL-MV(e) 相比其方法 (如 SLA+AG) 在性能方面确实有显著的提升, 但不可否认的是, 相比这些方法, SSL-MV(e) 拥有更多的可训练参数. 因此, 一个自然的问题是: SSL-MV(e) 性能的

提升是否仅因为模型参数量的提升?为进一步验证 SSL-MV(e) 的有效性更多源自算法本身设计,本节专门进行了 SSL-MV(e) 与独立集成 (IE) 的对比实验。

这里仍以旋转为例,独立集成 (IE) 是指为每一个旋转视图数据使用一个单独的网络来训练,最终在测试阶段集成/融合所有预训练网络的输出结果。此外,为进一步验证我们方法的优势,被赋予独立集成的 SLA+AG,即 IE+SLA+AG^[20],也被作为基线方法。请注意,粗略地估计,IE 与 IE+SLA+AG 的(可训练)参数量均高于 SSL-MV(e)。表 5 报告了实验结果,可以看到,SSL-MV(e) 仍获得了较为显著的性能优势,分类准确率先 IE 平均约 1.71 个百分点,领先 IE+SLA+AG 平均约 0.48 个百分点。这充分表明 SSL-MV(e) 有效性更多源自算法设计本身,而非仅依靠模型可训练参数的增加。

表 5 与独立集成的对比 (以旋转为例) (%)

数据集	Baseline	IE	IE+SLA+AG	SSL-MV(e)
CIFAR10	92.39	94.36	95.10	95.18
CIFAR100	68.27	74.82	76.40	77.26
TinyImageNet	63.11	68.18	69.01	69.51

4 总结与展望

本文探究自监督增广的监督分类增强学习,分析了现有方法的不足,并从多视图视角出发,提出一个简单但有效的自监督多视图学习框架 (SSL-MV),该框架隐式地将自监督标记嵌入下游分类模型的构建,在有效利用自监督标记的同时,有效避免了其对下游分类标记学习的干扰。在基准数据集上的广泛实验验证了所提学习框架的有效性。下一步工作是探究本文所提模型的泛化能力及拓展其至更具挑战的开放、动态场景,如开集识别^[27]等。此外,对于自监督学习,由于预置任务对下游任务通常是不可知或后见之明的,因此如何设置有效的预置任务及如何更有效地利用预置任务仍是一个开问题,也需进一步探究。

References:

- [1] 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]
- [2] Figueiredo MAT, Jain AK. Unsupervised learning of finite mixture models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002, 24(3): 381–396. [doi: 10.1109/34.990138]
- [3] Jing LL, Tian YL. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2021, 43(11): 4037–4058. [doi: 10.1109/TPAMI.2020.2992393]
- [4] Ericsson L, Gouk H, Hospedales TM. How well do self-supervised models transfer? In: *Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021. 5410–5419. [doi: 10.1109/CVPR46437.2021.00537]
- [5] Wang Z, Chen SC, Gao DQ. A novel multi-view learning developed from single-view patterns. *Pattern Recognition*, 2011, 44(10–11): 2395–2413. [doi: 10.1016/j.patcog.2011.04.002]
- [6] Wang Z, Xu J, Chen SC, Gao DQ. Regularized multi-view learning machine based on response surface technique. *Neurocomputing*, 2012, 97: 201–213. [doi: 10.1016/j.neucom.2012.05.027]
- [7] Zhao BY, Zhang CQ, Chen L, Liu XW, Li ZC, Hu QH. Generative model for partial multi-view clustering. *Acta Automatica Sinica*, 2021, 47(8): 1867–1875 (in Chinese with English abstract). [doi: 10.16383/j.aas.c200121]
- [8] Gidaris S, Singh P, Komodakis N. Unsupervised representation learning by predicting image rotations. In: *Proc. of the 2018 Int'l Conf. on Learning Representations (ICLR)*. Vancouver: OpenReview.net, 2018.
- [9] Noroozi M, Favaro P. Unsupervised learning of visual representations by solving jigsaw puzzles. In: *Proc. of the 14th European Conf. on Computer Vision*. Amsterdam: Springer, 2016. 69–84. [doi: 10.1007/978-3-319-46466-4_5]
- [10] Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: *Proc. of the 37th Int'l Conf. on Machine Learning*. Vienna: PMLR, 2020. 1597–1607.
- [11] He KM, Fan HQ, Wu YX, Xie SN, Girshick R. Momentum contrast for unsupervised visual representation learning. In: *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020. 9726–9735. [doi: 10.1109/CVPR42600.2020.00975]

- [12] Caron M, Misra I, Mairal J, Goyal P, Bojanowski P, Joulin A. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in Neural Information Processing Systems*. 2020.
- [13] Caron M, Bojanowski P, Joulin A, Douze M. Deep clustering for unsupervised learning of visual features. In: *Proc. of the 15th European Conf. on Computer Vision*. Munich: Springer, 2018. 139–156. [doi: [10.1007/978-3-030-01264-9_9](https://doi.org/10.1007/978-3-030-01264-9_9)]
- [14] Asano YM, Rupprecht C, Vedaldi A. Self-labelling via simultaneous clustering and representation learning. In: *Proc. of the 2019 Int'l Conf. on Learning Representations (ICLR)*. Addis Ababa, Ethiopia: OpenReview.net, 2019.
- [15] Sun Y, Tzeng E, Darrell T, Efros AA. Unsupervised domain adaptation through self-supervision. arXiv:1909.11825, 2019.
- [16] Hendrycks D, Mazeika M, Kadavath S, Song D. Using self-supervised learning can improve model robustness and uncertainty. *Advances in Neural Information Processing Systems*. Vancouver, 2019.
- [17] Tack J, Mo S, Jeong J, Shin, J. CSI: Novelty detection via contrastive learning on distributionally shifted instances. *Advances in Neural Information Processing Systems*. 2020.
- [18] Zhong Z, Zheng L, Kang GL, Li SZ, Yang Y. Random erasing data augmentation. *Proc. of the AAAI Conf. on Artificial Intelligence*, 2020, 34(7): 13001–13008. [doi: [10.1609/aaai.v34i07.7000](https://doi.org/10.1609/aaai.v34i07.7000)]
- [19] Huang KQ, Ren WQ, Tan TN. A review on image object classification and detection. *Chinese Journal of Computers*, 2014, 37(6): 1225–1240 (in Chinese with English abstract). [doi: [10.3724/SP.J.1016.2014.01225](https://doi.org/10.3724/SP.J.1016.2014.01225)]
- [20] Lee H, Hwang SJ, Shin J. Self-supervised label augmentation via input transformations. In: *Proc. of the 37th Int'l Conf. on Machine Learning*. PMLR, 2020. 5714–5724.
- [21] Tan ZH. On the generalization ability of deep learning strategies [MS. Thesis]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2021 (in Chinese with English abstract).
- [22] Krizhevsky A. Learning multiple layers of features from tiny images. Technical Report, Toronto: University of Toronto, 2009.
- [23] Le Y, Yang X. Tiny imagenet visual recognition challenge. CS 231N7, 2015. http://cs231n.stanford.edu/reports/2015/pdfs/yle_project.pdf
- [24] Quattoni A, Torralba A. Recognizing indoor scenes. In: *Proc. of the 2009 IEEE Conf. on Computer Vision and Pattern Recognition*. Miami: IEEE, 2009. 413–420. [doi: [10.1109/CVPR.2009.5206537](https://doi.org/10.1109/CVPR.2009.5206537)]
- [25] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In: *Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 770–778. [doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90)]
- [26] Bengio Y. Deep learning of representations: Looking forward. In: *Proc. of the 1st Int'l Conf. on Statistical Language and Speech Proc.* Tarragona: Springer, 2013. 1–37. [doi: [10.1007/978-3-642-39593-2_1](https://doi.org/10.1007/978-3-642-39593-2_1)]
- [27] Geng CX, Huang SJ, Chen SC. Recent advances in open set recognition: A survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2021, 43(10): 3614–3631. [doi: [10.1109/TPAMI.2020.2981604](https://doi.org/10.1109/TPAMI.2020.2981604)]

附中文参考文献:

- [7] 赵博宇, 张长青, 陈蕾, 刘新旺, 李泽超, 胡清华. 生成式不完整多视图数据聚类. *自动化学报*, 2021, 47(8): 1867–1875. [doi: [10.16383/j.aas.c200121](https://doi.org/10.16383/j.aas.c200121)]
- [19] 黄凯奇, 任伟强, 谭铁牛. 图像物体分类与检测算法综述. *计算机学报*, 2014, 37(6): 1225–1240. [doi: [10.3724/SP.J.1016.2014.01225](https://doi.org/10.3724/SP.J.1016.2014.01225)]
- [21] 谭正豪. 关于深度学习中的某些策略泛化性能的研究 [硕士学位论文]. 南京: 南京航空航天大学, 2021.



耿传兴(1990—), 男, 博士, CCF 专业会员, 主要研究领域为机器学习, 模式识别.



陈松灿(1962—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为机器学习, 模式识别, 神经计算.



谭正豪(1997—), 男, 硕士, 主要研究领域为机器学习, 模式识别.