

# 深度图像聚类的交替归一化与类别均匀先验优化方法

朱翌明, 马征\*

(西南交通大学信息科学与技术学院, 四川成都 611756)

**摘要:**深度图像聚类采用深度学习方法对无标签图像数据的类簇结构进行分析,然而,缺乏类别标签提供确定信息,无监督深度图像聚类可能输出不确定聚类预测导致噪声信息,不利于性能提升和应用发展。因此,提出一种基于交替归一化和类别均匀先验的聚类预测优化方法,校正低置信度预测,提升了深度图像聚类性能。同时,该方法与模型结构和训练过程耦合度较低,可实现深度图像聚类的跨模型优化。通过在多个数据集上的实验结果表明,该方法对多种深度图像聚类模型具有较好的聚类预测优化效果。

**关键词:**无监督学习;深度聚类;聚类预测优化;跨模型优化

中图分类号:TP183 文献标志码:A 文章编号:1001-2486(2025)04-151-07



论文  
拓展

## Optimization method for deep image clustering based on alternating normalization and category-wise uniform prior

ZHU Yiming, MA Zheng\*

(School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China)

**Abstract:** Deep image clustering was employed to analyze the cluster structure of unlabeled image data through deep learning techniques. However, due to the absence of class labels that provide definitive information, uncertain clustering predictions may be yielded by unsupervised deep image clustering, introducing noise information that was found detrimental to performance enhancement and application development. Therefore, a clustering prediction optimization method based on alternating normalization and category-wise uniform prior was proposed to correct low confidence predictions and improve deep image clustering performance. At the same time, the method had a low degree of coupling with the model structure and training process, enabling cross-model optimization for deep image clustering frameworks. Experimental results on multiple datasets reveal that the effective clustering prediction optimization is achieved for various deep image clustering models through the approach.

**Keywords:** unsupervised learning; deep clustering; clustering prediction optimization; cross-model optimization

深度图像聚类利用深度学习学习方法学习有利于聚类分析的图像特征,可以实现在无标签情况下识别图像集合的类簇并分析类别结构,同时为其他模型以无监督方式提供数据的类别信息,促进了许多相关的无监督图像分析处理任务的发展。

深度学习模型的成功得益于从成对的数据-标签对中进行数据表征学习的有监督训练方式,例如图像分类、目标检测等模型。然而,无监督学习场景下,缺少现成的类别标签。因此一些方法<sup>[1-4]</sup>通过表征学习和聚类学习联合训练的方式训练模型。例如 DeepCluster<sup>[1]</sup>通过对卷积神经网络提取的图像特征采用  $k$ -means 聚类得到类别信

息,从而训练图像特征提取模型。然而,采用这种联合训练的方式会导致模型易受初始化的影响,导致模型依赖于低级特征,降低模型泛化能力<sup>[5]</sup>。因此两阶段训练模型先进行表征学习,然后再进行聚类学习。最近邻语义聚类<sup>[5]</sup>(semantic clustering by adopting nearest neighbors, SCAN)模型利用深度表征学习进行数据特征提取,然后采用高置信度的聚类预测作为伪标签微调模型。许多相关无监督任务通过结合深度聚类以无监督方式获取数据的类别信息,从而辅助模型训练。交换视图聚类<sup>[6]</sup>(swapping assignments between multiple views, SwAV)和 Masked siamese<sup>[7]</sup>基于聚

收稿日期:2024-12-15

基金项目:四川省科技计划资助项目(2023YFG0100)

第一作者:朱翌明(1995—),男,陕西安康人,博士研究生,E-mail:zym@my.swjtu.edu.cn

\*通信作者:马征(1980—),男,四川成都人,教授,博士,博士生导师,E-mail:zma@swjtu.edu.cn

引用格式:朱翌明,马征.深度图像聚类的交替归一化与类别均匀先验优化方法[J].国防科技大学学报,2025,47(4):151-157.

Citation: ZHU Y M, MA Z. Optimization method for deep image clustering based on alternating normalization and category-wise uniform prior[J]. Journal of National University of Defense Technology, 2025, 47(4): 151-157.

类预测实现无监督图像表征学习;一些方法<sup>[8-11]</sup>基于聚类预测实现无监督语义分割和目标检测等问题;广义类别发现 (generalized categories discovery, GCD)<sup>[12]</sup>和 SimGCD (simple GCD)<sup>[13]</sup>提出了广义类别发现问题,同时进行标记数据的分类和无标记数据的聚类学习。

由于缺少真实标签提供确定的类别信号,样本可能由于不确定预测被分配错误的类别,从而对深度图像聚类及其相关任务模型的训练过程普遍产生不利影响。针对这个问题,在无监督目标重识别任务中,Ge 等<sup>[14]</sup>提出“同步平均教学”优化聚类预测;Zheng 等<sup>[15]</sup>采用层次聚类方法在线动态提高聚类预测质量。在基于聚类的表征学习任务中,SCAN<sup>[5]</sup>和 DeepCluster<sup>[1]</sup>模型都采用丢弃低置信度预测的方式提高伪标签质量。文献[16]结合了许多现有缓解标签噪声的鲁棒学习方法,包括噪声标签识别<sup>[17]</sup>,半监督模型重训练<sup>[18]</sup>,多模型联合训练<sup>[19]</sup>。尽管以上研究都提出了对深度图像聚类进行优化的方法,但都是根据特定任务提出的针对性方法,因此与具体任务过程和模型结构的耦合度较高,导致这些优化方法应用范围较小,难以推广到更大范围的数量逐渐增加的深度图像聚类相关模型。

因此,为了能够实现跨模型地优化深度图像聚类及相关模型的训练过程,本文提出基于交替归一化和类别均匀先验的深度图像聚类跨模型优化方法。首先以不确定度为依据,将一个训练批次 (minibatch) 样本的聚类预测分为低置信度和高置信度预测。然后设定高置信度预测为参考集,低置信度预测为待优化集,采用交替归一化 (alternative normalization, AN) 方法<sup>[20]</sup>逐个处理并更新低置信度预测,使得该批次聚类预测的类别分布逼近类别均匀先验分布,从而校正该训练批次的低置信度 (不确定性) 预测。进一步地,基于计算的聚类预测的不确定度得到损失函数的样本加权系数,强化低置信度聚类预测的优化效果。在深度嵌入聚类<sup>[2]</sup> (deep embedding cluster, DEC)、SCAN、SwAV 和 SimGCD 四种不同的深度聚类相关模型上验证了方法的跨模型优化效果。实验表明,该方法在 DEC 和 SCAN 模型上分别在 mnist、STL-10 和 cifar-10 三个数据集上最高提升了 2.1%、9.5% 和 0.3% 的准确度。对基于在线聚类的对比表征学习模型 SwAV 应用该方法,使得基于该表征模块在 cifar-10 数据集上的图像分类性能提高了 2.5% 的分类准确率。对广义类别发现模型 SimGCD 运用该方法优化无标签数据的

聚类学习过程,同时提高了新旧类别数据的分类性能。通过采用主成分分析 (principal components analysis, PCA) 和  $t$  分布随机邻域嵌入 ( $t$ -distributed stochastic neighbor embedding,  $t$ -SNE)<sup>[21]</sup>两种方法对 SwAV 提取的视觉特征进行降维和可视化,发现该方法优化了位于决策边界附近的低置信度预测,提高了特征的判别性。

## 1 方法

为了实现跨模型优化聚类预测,提出通过优化模型输出的聚类预测概率分布实现跨模型优化。首先将聚类预测的优化过程建模为最优传输问题;然后利用该问题的求解过程对模型输出的低置信度聚类预测进行校正;最后对具有较高不确定度的训练样本赋予较高的损失函数权值,强化第一步的优化效果。

### 1.1 基于最优传输理论建模聚类优化问题

假设模型对训练样本的聚类预测为  $\mathbf{P}_{yx} = p(y|x) \subseteq \mathbf{R}_+^{N \times M}$ ,  $\mathbf{P}_{yx}$  是关于类别  $y$  和样本  $x$  的联合概率矩阵,即  $\mathbf{P}_{yx}$  表示模型将训练样本  $X$  分配到类别  $Y$  的分配过程, $N$  为样本数, $M$  为类别数; $\mathbf{R}_+^{N \times M}$  为  $N \times M$  的正实数矩阵。进一步假设存在更优的类别分配  $\mathbf{Q}_{yx} = q(y|x)$ ,此时基于  $\mathbf{Q}_{yx}$  训练深度聚类模型的损失函数为:

$$J = -\frac{1}{N} \sum_{i=1}^N \sum_{y=1}^M q(y|x_i) \ln p(y|x_i) \quad (1)$$

文献[22]证明对于无监督聚类为了避免  $\mathbf{Q}_{yx}$  退化为平凡解 (导致所有样本分配为同一类别),需要限制  $\mathbf{Q}_{yx}$  满足均等化分约束,因此可将此约束优化问题建模为最优传输问题,此时  $\mathbf{Q}_{yx}$  满足传输多面体结构<sup>[22]</sup>:

$$U(\mathbf{r}, \mathbf{c}) := \{ \mathbf{Q} \subseteq \mathbf{R}_+^{M \times N} \mid \mathbf{Q}\mathbf{e} = \mathbf{r}, \mathbf{Q}^T \mathbf{e} = \mathbf{c} \} \quad (2)$$

其中,  $\mathbf{e}$  表示维度匹配的全 1 向量,  $\mathbf{r}$  和  $\mathbf{c}$  分别代表  $\mathbf{Q}_{yx}$  的行向量之和与列向量之和,  $U(\mathbf{r}, \mathbf{c})$  为  $\mathbf{Q}_{yx}$  的解空间。采用均匀分布作为  $\mathbf{Q}_{yx}$  解空间的约束,即  $\mathbf{r} = \frac{1}{M} \cdot \mathbf{e}$ , 此时聚类模型的预测倾向于将  $N$  个样本平均分配到  $M$  个类别上,同时缓解模型训练时对样本数量占比高的类别的过拟合问题。

此时,可将式(3)重写为最优传输问题形式,并且通过增加熵正则化项提高在大量样本时的优化效率<sup>[23]</sup>,即:

$$\min_{\mathbf{Q} \in U(\mathbf{r}, \mathbf{c})} \langle \mathbf{Q}, -\ln \mathbf{P} \rangle + \frac{1}{\lambda} KL(\mathbf{Q} \parallel \mathbf{rc}^T) \quad (3)$$

其中,  $\langle \cdot \rangle$  表示矩阵内积,  $KL$  表示 KL 散度 (Kullback-Leibler divergence),  $\mathbf{rc}^T$  表示  $M \times N$  的

矩阵,  $\lambda$  表示调节正则项的权重。

该形式的求解一般采用交替归一化方法(也称 sinkhorn 算法)<sup>[23]</sup>,即:

$$\forall y: \alpha_y \leftarrow [\mathbf{P}^A \boldsymbol{\beta}]_y^{-1} \quad (4)$$

$$\forall x: \beta_x \leftarrow [\boldsymbol{\alpha}^T \mathbf{P}^A]_x^{-1} \quad (5)$$

交替迭代上述两个公式,即可使  $\mathbf{Q}_{yx}$  满足类别均匀约束条件。其中,  $\boldsymbol{\alpha}$  和  $\boldsymbol{\beta}$  为缩放系数向量,  $\alpha_y$  和  $\beta_x$  为每次迭代的临时值,  $\mathbf{P}^A$  为  $\mathbf{P}$  的  $\lambda$  次幂。每一次迭代涉及矩阵和向量的乘法运算,其计算复杂度为  $O(NM)$ ,由于  $M$  为固定的类别数量,因此计算复杂度随样本数量  $N$  线性增长。此时可实现基于  $\mathbf{P}_{yx}$  得到优化后的  $\mathbf{Q}_{yx}$ <sup>[23]</sup>,即:

$$\mathbf{Q}_{xy} = \text{diag}(\boldsymbol{\alpha}) \mathbf{P}^A \text{diag}(\boldsymbol{\beta}) \quad (6)$$

## 1.2 基于交替归一化和类别均匀先验校正低置信度预测

通过 1.1 小节将聚类优化问题建模为最优传输问题,可对聚类模型的预测进行优化。为了降低方法对模型结构和训练过程的影响,提高对不同聚类模型的跨模型适用性,进一步提出对低置信度预测的针对性校正。

### 1.2.1 区分高低置信度预测

给定一批训练样本 ( $batch\_size = n$ ),首先采用  $top-k$  信息熵作为聚类预测的不确定度指标,即先对预测的概率值进行降序排序,然后选取前  $k$  个概率值计算信息熵。假设聚类预测概率分布  $p$  的  $top-k$  概率值为  $\bar{p}_1, \bar{p}_2, \dots, \bar{p}_k$ ,则  $top-k$  信息熵为:

$$H_{top-k}(p) = - \sum_{i=1}^k \bar{p}_i \ln \bar{p}_i \quad (7)$$

计算出聚类预测的不确定度后,设定一个阈值  $\tau$ ,不确定度大于等于  $\tau$  的即为低置信度预测,反之则为高置信度预测。

### 1.2.2 校正低置信度预测

假设一批样本的聚类预测经过 1.2.1 节方法区分为高置信度预测集  $\mathbf{A} \subset \mathbf{R}^{a \times M}$ ,低置信度预测集  $\mathbf{B} \subset \mathbf{R}^{(n-a) \times M}$ ,即  $\mathbf{P} = [\mathbf{A} \ \mathbf{B}]^T$ 。

取一个低置信度预测  $\mathbf{b} \in \mathbf{B}$ ,与高置信度预测集  $\mathbf{A}$  按行拼接作为待优化的预测矩阵  $\hat{\mathbf{P}} \subset \mathbf{R}^{(a+1) \times M}$ ,即  $\hat{\mathbf{P}} = [\mathbf{A} \ \mathbf{b}]^T$ 。采用 3.1 小节描述的交替归一化方法,对  $\hat{\mathbf{P}}$  进行处理,得到优化后的  $\tilde{\mathbf{Q}} = [\mathbf{A} \ \hat{\mathbf{b}}]^T$ ,从而得到校正后的  $\hat{\mathbf{b}}$ 。

### 1.2.3 低置信度预测更新

通过对低置信度预测集  $\mathbf{B} \subset \mathbf{R}^{(n-a) \times M}$  中的每一个待优化预测  $\mathbf{b}$  按 1.2.2 节方法进行校正,可优化该批样本中所有低置信度预测从而得到  $\hat{\mathbf{b}}$ ,

替换  $\mathbf{P}$  中对应位置的  $\mathbf{b}$ ,而高置信度预测则无须调整,即可得到该批样本优化后的聚类预测。算法 1 给出了低置信度聚类预测优化的伪代码。

### 算法 1 低置信度聚类预测的优化

Alg. 1 Optimization of low confidence clustering predictions

输入:聚类预测集合  $\mathbf{P}$ ,  $batch\_size = n$ ,置信度阈值  $\tau$

输出:优化后的聚类预测  $\mathbf{Q}$

```

1. for each  $p$  in  $\mathbf{P}$ 
2.   由式(7)计算  $top-k$  信息熵  $H_{top-k}(p)$ 
3.   if  $H_{top-k}(p) > \tau$ 
4.      $\mathbf{B}$ .append( $p$ ) //低置信度集合  $\mathbf{B}$  添加元素  $\mathbf{P}[i]$ 
5.      $N$ .append( $p$ .index) //集合  $N$  记录  $p$  在  $\mathbf{P}$  中的位置
6.   else
7.      $\mathbf{A}$ .append( $p$ ) //高置信度集合  $\mathbf{A}$  添加元素  $\mathbf{P}[i]$ 
8.   end if
9. end for each
10. for each  $b \in \mathbf{B}$ 
11.    $\mathbf{P}'$ .append( $\mathbf{A}$ ,  $b$ ) //将  $\mathbf{A}$  和  $b$  拼接为待优化集合  $\mathbf{P}'$ 
12.   由式(4)和式(5)对  $\mathbf{P}'$  进行交替归一化
13.    $\mathbf{B}'$ .append( $b$ )
14. end for each
15. for each ( $b$ ,  $i$ ) in ( $\mathbf{B}'$ ,  $N$ )
16.    $\mathbf{P}[i] = b$  //优化后的  $b$  覆盖原位置的元素
17. end for each
18.  $\mathbf{Q} = \mathbf{P}$ 

```

## 1.3 基于不确定度加权样本损失

对 1.2.1 小节计算的样本聚类预测的不确定度采用 softmax 函数进行归一化,得到样本损失值的加权系数,使得不确定度高的低置信度预测具有较高的权值。

给定一批样本  $\hat{\mathbf{X}} = \{x_1, x_2, \dots, x_n\} \subseteq \mathbf{X}$ ,其聚类预测  $\hat{\mathbf{p}} = \{p_1, p_2, \dots, p_n\}$ ,通过式(9)计算其不确定度:  $H_{top-k}(\hat{\mathbf{p}}) = [h_1, h_2, \dots, h_n]$ 。采用 softmax 函数得到归一化权重系数为:

$$\mathbf{a} = F_{\text{softmax}}(H_{top-k}(\hat{\mathbf{p}})) \quad (8)$$

与样本  $\hat{\mathbf{X}}$  的损失函数  $L$  的值向量进行点积实现对样本损失的加权:

$$L = - \frac{1}{N} \sum_{i=1}^N \sum_{y=1}^M q(y|x_i) \ln p(y|x_i) \quad (9)$$

其中,  $L$  表示模型的损失函数,具体形式视模型而定,分类问题中一般为交叉熵损失。

## 2 实验

以 DEC 和 SCAN 作为深度聚类基线模型,展示方法对深度聚类模型的聚类性能优化的有效性。

然后对无监督表征学习方法 SwAV 和广义类别发现模型 SimGCD 应用所提方法优化聚类预测以提高相应性能;通过可视化 SwAV 模型学到的图像表征以及对比下游的图像分类性能,展示了方法的跨模型优化能力;最后以 DEC 和 SCAN 为基线模型进行消融对比试验,证明方法两部分的必要性。

## 2.1 实验配置

### 2.1.1 数据集设置

为了与基线模型进行合适的对比,分别按照文献[2,5,6,13]实验的配置进行数据集设置。采用了 mnist<sup>[24]</sup>、cifar-10/100<sup>[25]</sup> 和 STL-10<sup>[26]</sup> 图像数据集进行实验。mnist 数据集即手写数字图像数据集,包含 60 000 幅训练集和 10 000 幅测试集的 28 像素×28 像素的黑白图像,有 0~9 共 10 个类别。cifar-10 包含 60 000 幅 10 个类别的彩色图像,图像大小为 32 像素×32 像素,cifar-100 则包含 100 个类别,每个类别包括 500 幅训练图像和 100 幅测试图像;STL-10 是与 cifar-10 类似的彩色图像数据集,但是具有更高的 96 像素×96 像素的分辨率,共有 100 000 幅无标记图像和 13 000 幅标记图像,采用 13 000 幅标记图像以方便性能评估;为了公平比较,参考文献[2],同样采用 ResNet-50 提取的特征作为 DEC 模型的输入。

### 2.1.2 实验环境配置与超参数设置

所有实验基于 PyTorch 深度学习框架,采用 NVIDIA RTX 3090 显卡进行计算加速。基线模型源码和运行环境基于原文献公布的官方代码库,基线模型的超参数配置如学习率、训练轮次 epochs、批次大小、最优化方法和网络结构等参考了文献[2,5,6,13]。

## 2.2 聚类性能优化实验

### 2.2.1 性能指标

1) 准确度:模型正确分类的样本个数占样本总数的比例。

2) 标准化互信息(normalized mutual information, NMI):计算模型聚类结果与标准分类之间的相似度,取值为 0~1 之间的实值,越高代表聚类结果越好。

### 2.2.2 基线模型

深度聚类模型的发展过程中出现了两种主要类型,按模型的训练阶段可以分为联合训练模型和两阶段训练模型,本文按此分类分别选取了具有代表性的基线模型。其中 DEC 采取了基于自编码器和聚类联合训练的模型设计,SCAN 则采用了先训练语义表征学习模块后进行聚类训练和

微调的两阶段模型。

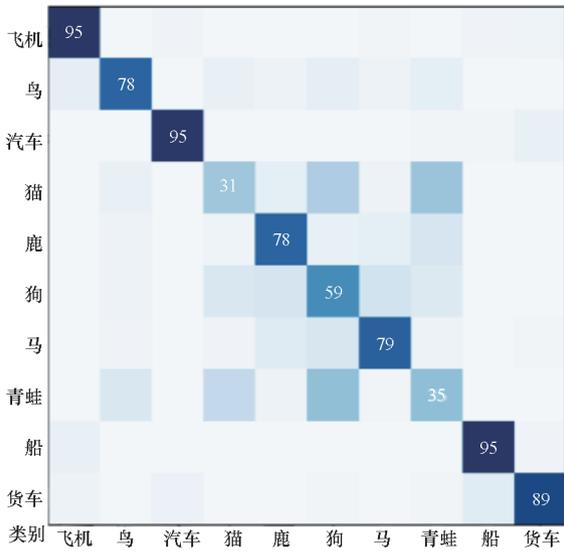
### 2.2.3 实验结果

表 1 对比基线模型使用方法前后聚类性能的差异。优化聚类预测后,DEC 和 SCAN 模型在 mnist 和 STL-10 两个数据集上获得了准确度和 NMI 性能较大提升。在 cifar-10 上 SCAN 的聚类性能提升效果相对较低,可能因为 SCAN 在 cifar-10 数据集的拟合效果较好,因此低置信度预测较少,从而影响方法优化效果。而在 cifar 数据集上,DEC 和 SCAN 模型本身的聚类性能有待提高,因此存在更多可优化的低置信度预测,相对优化效果更明显,侧面说明方法优化低置信度预测的合理性。图 1 给出了优化前后的聚类预测混合矩阵。表 1 的 SwAV 和 simGCD 方法相关行数据给出了方法对基于深度聚类的表征学习模型 SwAV 和广义类别发现模型 SimGCD 的优化效果,表明该方法具有较强的跨模型适用能力。表 1 中 RUC + SCAN 方法一行展示了与同类方法鲁棒无监督聚类(robust unsupervised cluster, RUC)<sup>[19]</sup> 优化效果对比。所提方法虽然对基线模型的优化效果略低于该方法,但是所提方法具有更好的跨模型适用性。表 2 给出了不同模型的超参数设置,结合表 1 的聚类优化结果可以看出,本文方法对于采用不同训练轮次(epochs)和多种训练批次的

表 1 聚类和分类性能优化结果

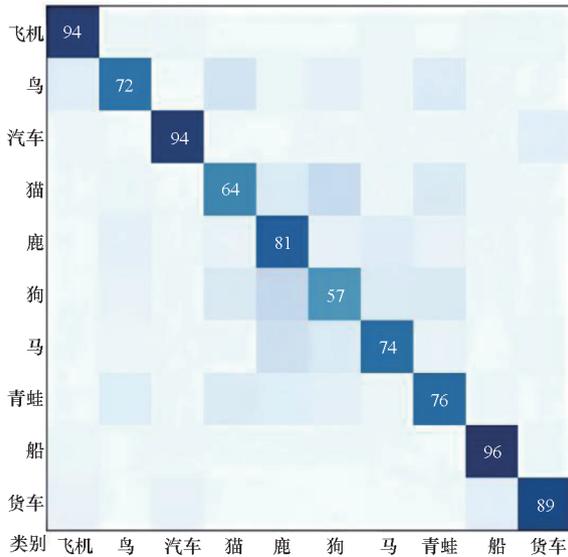
Tab. 1 Optimization results of clustering and classification performance

方法	数据集	准确度/%	标准化互信息
DEC	mnist	84.30	81.60
	STL-10	80.60	75.40
Ours + DEC	mnist	86.40	82.70
	STL-10	90.10	81.70
SCAN	cifar-10	88.04	79.21
	cifar-100	50.70	48.60
Ours + SCAN	STL-10	81.58	70.45
	cifar-10	88.30	79.32
	cifar-100	50.60	47.90
SwAV	STL-10	82.83	71.43
	cifar-10	68.30	—
Ours + SwAV	cifar-10	70.80	—
SimGCD <sup>[16]</sup>	cifar-100	79.13	—
Ours + SimGCD	cifar-100	80.72	—
RUC <sup>[19]</sup> + SCAN	cifar-10	89.50	—
Ours + SCAN	cifar-10	88.30	79.32



(a) 优化前的混合矩阵图

(a) Confusion matrix before optimization



(b) 优化后的混合矩阵图

(b) Confusion matrix after optimization

图 1 SCAN 模型优化前后聚类结果混合矩阵图

Fig. 1 Confusion matrix plot of clustering results before and after optimization of the SCAN model

表 2 超参数设置

Tab. 2 Hyperparameter setting

方法	学习率	训练轮次	训练批次	最优化方法
DEC	0.001	200	256	SGD
SCAN	0.001	100	128	Adam
SwAV	0.001	100	32	Adam

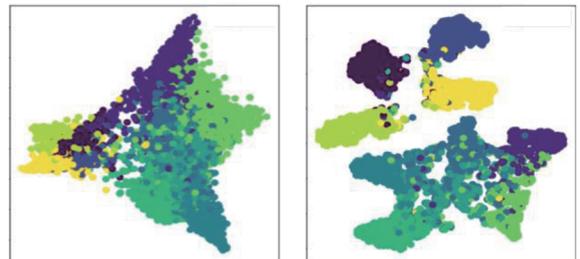
模型都具有较好的聚类优化效果,体现出较强的跨模型优化能力。

### 2.3 优化对比表征学习 SwAV 及其下游分类任务

针对一种基于在线聚类的对比学习模型 SwAV,验证方法在聚类辅助相关模型时的优化能力。采用所提方法对聚类预测进行优化,使聚类预测更准确反映正负样本的类别信息。图 2 给出了优化前后,SwAV 模型提取的图像表征可视化结果,可以看出优化后模型提取的不同类别图像表征之间具有更大的间隔,具有更好的判别性。在此基础上,表 1 SwAV 方法相关的两行展示采用优化方法后,基于 SwAV 预训练的模型进行下游的分类任务微调,分类性能提升了 2.5%,说明了该方法对基于在线聚类的无监督深度图像处理模型具有较好的优化能力。

### 2.4 定性分析

图 2(a) ~ (b)展示了 SwAV 预训练的表征学习模型提取的 cifar-10 图像特征,图 2(c) ~ (d)为采用所提方法优化后模型提取的图像特征,采用了 PCA 和 t-SNE 对模型提取的高维图像特征进行降维和可视化。可以看出,采用所提方法后不同类簇特征之间的距离更大,说明采用该方法



● 飞机 ● 鸟 ● 汽车 ● 猫 ● 鹿 ● 狗 ● 马 ● 青蛙 ● 船 ● 货车

(a) 优化前 PCA 可视化

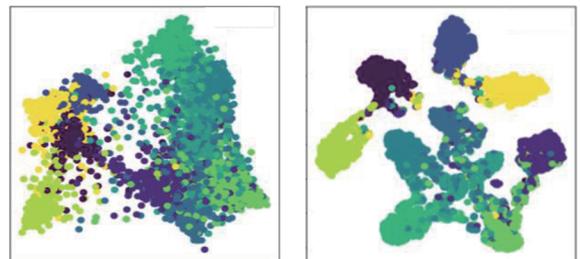
(a) PCA visualization

before optimization

(b) 优化前 t-SNE 可视化

(b) t-SNE visualization

before optimization



● 飞机 ● 鸟 ● 汽车 ● 猫 ● 鹿 ● 狗 ● 马 ● 青蛙 ● 船 ● 货车

(c) 优化后 PCA 可视化

(c) PCA visualization

after optimization

(d) 优化后 t-SNE 可视化

(d) t-SNE visualization

after optimization

图 2 优化前后 SwAV 提取的 cifar-10 图像特征可视化

Fig. 2 Visualization of the cifar-10 image features extracted by the SwAV before and after optimization

优化不确定预测后,优化了位于决策边界附近的样本预测,使得表征模型学得的图像特征更具有判别性,从而提升后续任务的性能。

### 2.5 消融对比实验

基于 DEC 和 SCAN 模型验证所提方法两个部分的必要性。表 3 为在 STL-10 数据集上分别测试采用不同方法配置下的性能,“√”表示采用了对应方法部分。第一行数据为未采用所提方法的基线性能;第二行和第一行对比,采用所提方法优化聚类预测后,两个模型在聚类准确度分别提升了 7.9% 和 1.25%;第三行和第一行对比,在没有进行低置信度预测校正的情况下,增加低置信度样本的损失权重会降低模型性能 0.7%;第四行数据则说明采用所提方法校正并加权低置信度预测可以进一步提高聚类性能,取得最高的性能提升。

表 3 消融实验结果

Tab. 3 Results of the ablation study

校正	重加权	DEC		SCAN	
		准确度/%	标准化互信息	准确度/%	标准化互信息
		80.60	75.40	81.58	70.45
√		88.50	79.90	82.83	71.43
	√	79.30	74.40	—	—
√	√	90.10	81.70	83.36	72.19

### 3 结论

针对深度图像聚类的聚类预测优化问题,提出基于交替归一化和类别均匀先验的优化方法,旨在通过校正低置信度预测实现跨模型优化深度聚类预测。提出的方法对模型结构和训练过程的耦合度较低,因此可以高效地优化不同类型深度聚类相关模型。未来可以考虑利用相关先验信息,例如类簇数量或尺寸等约束信息,进一步提高聚类预测性能,同时进行更深入的理论分析。

### 参考文献 (References)

[1] CARON M, BOJANOWSKI P, JOULIN A, et al. Deep clustering for unsupervised learning of visual features [C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018.

[2] XIE J Y, GIRSHICK R, FARHADI A. Unsupervised deep embedding for clustering analysis [C]// Proceedings of the 33rd International Conference on Machine Learning, 2016:

478 - 487.

[3] HAN S, PARK S, PARK S, et al. Mitigating embedding and class assignment mismatch in unsupervised image classification [C]// Proceedings of the 16th European Conference on Computer Vision (ECCV), 2020: 768 - 784.

[4] CARON M, TOUVRON H, MISRA I, et al. Emerging properties in self-supervised vision transformers [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 9630 - 9640.

[5] VAN GANSBEKE W, VANDENHENDE S, GEORGOULIS S, et al. SCAN: learning to classify images without labels [C]// Proceedings of the 16th European Conference on Computer Vision (ECCV), 2020: 268 - 285.

[6] CARON M, MISRA I, MAIRAL J, et al. Unsupervised learning of visual features by contrasting cluster assignments [C]// Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020: 9912 - 9924.

[7] ASSRAN M, CARON M, MISRA I, et al. Masked siamese networks for label-efficient learning [C]// Proceedings of the European Conference on Computer Vision (ECCV), 2022: 456 - 473.

[8] WEN X, ZHAO B C, ZHENG A L, et al. Self-supervised visual representation learning with semantic grouping [C]// Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS 2022), 2022.

[9] CHO J H, MALL U, BALA K, et al. PiCIE: unsupervised semantic segmentation using invariance and equivariance in clustering [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 16789 - 16799.

[10] WU H, ZHAO S J, HUANG X, et al. Commonsense prototype for outdoor unsupervised 3D object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 14968 - 14977.

[11] VS V, GUPTA V, OZA P, et al. MeGA-CDA: memory guided attention for category-aware unsupervised domain adaptive object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 4514 - 4524.

[12] VAZE S, HANT K, VEDALDI A, et al. Generalized category discovery [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 7482 - 7491.

[13] WEN X, ZHAO B C, QI X J. Parametric classification for generalized category discovery: a baseline study [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023: 16544 - 16554.

[14] GE Y X, CHEN D P, LI H S. Mutual mean-teaching: pseudo label refinery for unsupervised domain adaptation on person re-identification [C]// Proceedings of the International Conference on Learning Representations (ICLR), 2020.

[15] ZHENG Y, TANG S X, TENG G L, et al. Online pseudo label generation by hierarchical cluster dynamics for adaptive person re-identification [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 8351 - 8361.

[16] PARK S, HAN S, KIM S, et al. Improving unsupervised image clustering with robust learning [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

- Recognition (CVPR), 2021: 12273 – 12282.
- [17] LI J N, SOCHER R, HOI S C H. DivideMix: learning with noisy labels as semi-supervised learning[C]//Proceedings of the International Conference on Learning Representations (ICLR), 2020.
- [18] BERTHELOT D, CARLINI N, GOODFELLOW I, et al. MixMatch: a holistic approach to semi-supervised learning[C]//Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), 2019.
- [19] HAN B, YAO Q M, YU X R, et al. Co-teaching: robust training of deep neural networks with extremely noisy labels[C]//Proceedings of the Neural Information Processing Systems, 2018.
- [20] SINKHORN R, KNOPP P. Concerning nonnegative matrices and doubly stochastic matrices [J]. Pacific Journal of Mathematics, 1967, 21(2): 343 – 348.
- [21] VAN DER MAATEN L, HINTON G. Visualizing data using  $t$ -SNE[J]. Journal of Machine Learning Research, 2008, 9(86): 2579 – 2605.
- [22] ASANO Y M, RUPPRECHT C, VEDALDI A. Self-labelling via simultaneous clustering and representation learning[C]//Proceedings of the International Conference on Learning Representations (ICLR), 2020.
- [23] CUTURI M. Sinkhorn distances: lightspeed computation of optimal transport [C]//Proceedings of the Conference on Neural Information Processing Systems, 2013.
- [24] LECUN Y, CORTES C. The MNIST database of handwritten digits[EB/OL]. [2024-03-01]. [https://www.lri.fr/~marc/Master2/MNIST\\_doc.pdf](https://www.lri.fr/~marc/Master2/MNIST_doc.pdf).
- [25] KRIZHEVSKY A. Learning multiple layers of features from tiny images [M]. Handbook of Systemic Autoimmune Diseases, 2009.
- [26] COATES A, LEE H, NG A Y. An analysis of single-layer networks in unsupervised feature learning[C]//Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, 2011: 215 – 223.