计算机科学
COMPUTER SCIENCE

一种鲁棒的双教师自监督蒸馏哈希学习方法

苗壮，王亚鹏，李阳，王家宝，张睿，赵昕昕

引用本文

苗壮，王亚鹏，李阳，王家宝，张睿，赵昕昕. 一种鲁棒的双教师自监督蒸馏哈希学习方法[J]. 计算机科学, 2022, 49(10): 159-168.

MIAO Zhuang, WANG Ya-peng, LI Yang, WANG Jia-bao, ZHANG Rui, ZHAO Xin-xin. Robust Hash Learning Method Based on Dual-teacher Self-supervised Distillation[J]. Computer Science, 2022, 49(10): 159-168.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210800050

一种鲁棒的双教师自监督蒸馏哈希学习方法

苗 壮 王亚鹏 李 阳 王家宝 张 睿 赵昕昕
陆军工程大学指挥控制工程学院 南京 210007
(emiao_beyond@163.com)

摘 要 为了提高无监督哈希学习的性能,实现鲁棒的哈希图像检索,提出了一种鲁棒的双教师自监督蒸馏哈希学习方法。该方法包括自监督双教师学习和鲁棒哈希学习两个阶段:第一阶段设计了一种改进的聚类算法,有效提高了硬伪标签的标注精度,而后通过微调教师网络得到了图像的初始软伪标签;第二阶段提出了一种结合混合去噪和双教师共识去噪策略的软伪标签去噪方法,有效去除了初始软伪标签中的噪声,而后利用蒸馏学习将双教师网络中的信息通过去噪软伪标签传递给学生网络,进而获得无标签图像的鲁棒哈希码。在CIFAR-10,FLICKR25K和EuroSAT上进行了实验,实验结果表明,与TBH方法相比,在CIFAR-10上所提方法的MAP平均提高了18.6%;与DistillHash方法相比,在FLICKR25K上所提方法的MAP平均提高了2.4%;与ETE-GAN方法相比,在EuroSAT上所提方法的MAP平均提高了18.5%。

关键词:哈希学习;自监督学习;知识蒸馏;图像检索;噪声标签
中图法分类号 TP391

Robust Hash Learning Method Based on Dual-teacher Self-supervised Distillation

MIAO Zhuang,WANG Ya-peng,LI Yang,WANG Jia-bao,ZHANG Rui and ZHAO Xin-xin
Command and Control Engineering College,Army Engineering University of PLA,Nanjing 210007,China

Abstract In order to improve the performance of unsupervised hash learning and achieve robust hashing image retrieval,this paper proposes a novel robust hash learning method based on dual-teacher self-supervised distillation. Specifically,the proposed method contains two stages:a self-supervised dual-teacher learning stage and a robust hash learning stage. In the first stage,a modified cluster algorithm is designed to effectively improve the accuracy of hard pseudo labels. Then,we fine-tune the teacher networks by hard pseudo labels to get the initial soft pseudo labels. In the second stage,we filter the initial soft pseudo labels by our soft pseudo label denoising method,which combines a hybrid denoising strategy and a dual-teacher denoising strategy. Then,we train the student network with the denoised soft pseudo labels by knowledge distillation,so that robust hash codes for label-free images are obtained. Extensive experiments on CIFAR-10,FLICKR25K and EuroSAT datasets show that the proposed robust hash learning method outperforms the state-of-the-art methods. In detail,the MAP of our method is 18.6% higher than that of the TBH method on CIFAR-10,2.4% higher than that of the DistillHash method on FLICKR25K,and 18.5% higher than that of the ETE-GAN method on EuroSAT.

Keywords Hash learning,Self-supervised learning,Knowledge distillation,Image retrieval,Noisy labels

1 引言

随着深度学习技术的发展,有监督哈希学习取得了巨大突破,特别是最近提出的深度有监督哈希学习方法,在图像检索领域取得了令人瞩目的成就[1-3]。虽然这些方法不断刷新着特定数据集的检索精度,但是它们需要大量具有可靠标签的数据,这对许多现实世界的应用提出了很高的要求。相反,深度无监督哈希学习提供了一种不需要任何人工标签的高效学习框架,可以在没有人工标签的情况下学习到无标签图像的二值化哈希码,并用于高效的图像检索,因此深度无监督哈希学习成为了新的研究热点。

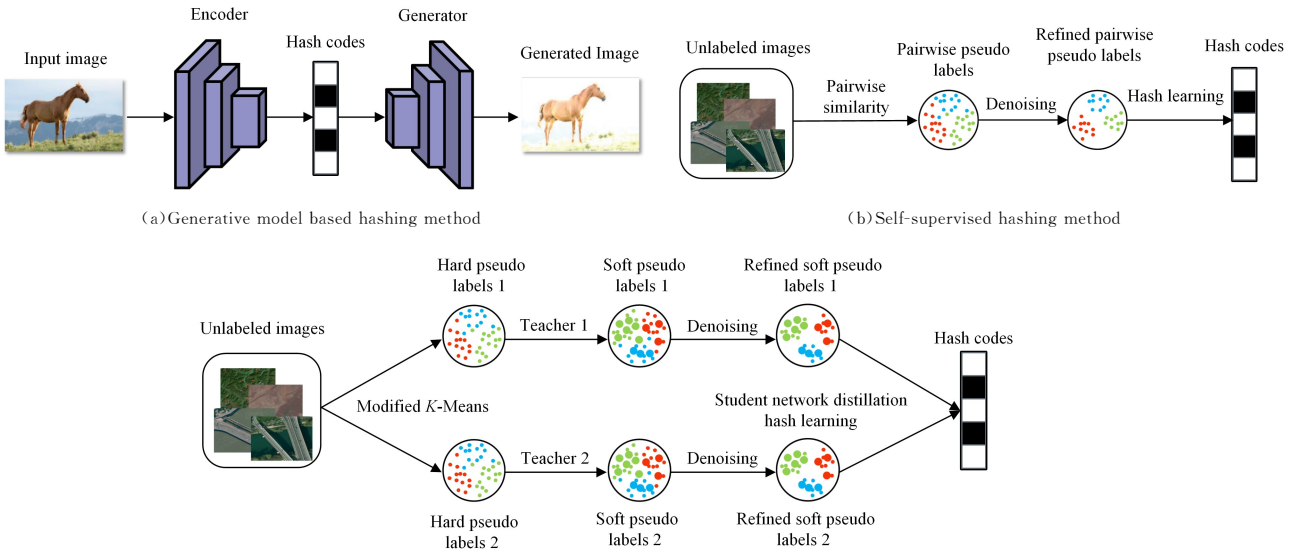
现有的深度无监督哈希学习方法可大致分为生成哈希方法和自监督哈希方法两类。生成哈希方法(见图1(a))主要基于自编码器[4]、对抗生成网络[5]等生成模型实现哈希学习,如TBH(Twin-Bottleneck Hashing)方法[6]、ETE-GAN(End-To-End Generative Adversarial Networks)方法[7]和BGAN

到稿日期:2021-08-05 返修日期:2021-12-08
基金项目:国家自然科学基金(61806220);国家重点研发计划(2017YFC0821905)
This work was supported by the National Natural Science Foundation of China(61806220)and National Key Research and Development Program of China(2017YFC0821905).
通信作者:李阳(solarleeon@outlook.com)

(Binary Generative Adversarial Networks) 方法[8]；自监督哈希方法(见图 1(b))主要基于图像对相似性生成伪标签，而后利用伪标签监督哈希网络学习，如 DistillHash 方法[9]和 SSDH(Semantic Structure-based unsupervised Deep Hashing)方法[10]。在以上两类方法中，自监督哈希方法具有更强的特征提取能力，且易于训练，因此成为了目前深度无监督哈希学习的主流方向。然而，自监督哈希方法中的伪标签存在大量噪声，噪声标签的存在将会导致模型性能退化，因此噪声标签成为了影响自监督哈希学习性能的关键因素。

为了减少伪标签中的噪声标签，DistillHash 方法[9]通过最优贝叶斯分类器筛选更加可信的相似或不相似图像对，提高了图像对伪标签的精度；SSDH 方法[10]则对伪标签进行了细化，除传统的“相似”和“不相似”两类伪标签外，还加入了“不确定”伪标签。但是，由于这些方法自身伪标签标注算法性能的限制，且未在去噪后做进一步鲁棒学习处理，因此噪声标签的问题仍未得到较好解决，这导致深度无监督哈希学习的性能与深度有监督哈希学习仍有很大差距。

本文从噪声标签的角度出发，对自监督哈希方法进行了重新思考，我们认为：如何有效减少噪声标签，并进一步控制深度网络对噪声标签的过拟合，是影响自监督哈希学习性能的核心问题。因此，本文在标签去噪的基础上，通过增加自监督双教师学习机制进一步提高了伪标签标注精度，通过增加鲁棒哈希学习机制增强了哈希网络的稳定性，进而使得无监督哈希检索精度得到有效提高(见图 1(c))。



(a)Generative model based hashing method

(b)Self-supervised hashing method

(c)Our roubst dual-teacher self-supervised distillation hasing method

图 1 不同深度无监督哈希方法的对比

Fig. 1 Comparison between different deep unsupervised hashing methods

基于上述理念，本文提出了一种鲁棒的双教师自监督蒸馏哈希学习方法(见图 2)，该方法包括自监督双教师学习和鲁棒哈希学习两个阶段。第一阶段的主要目的是将无标签数据转换为可靠的有标签数据，第二阶段的主要目的是将所得到的有噪声标签的数据转换为鲁棒的二进制哈希码。由于这两个阶段的主要任务不同，本文根据任务特点设计了不同的学习方法。具体来说，在第一阶段，本文通过改进 K-Means 聚类算法有效提高了伪标签的标注精度，并通过微调教师网络得到了图像的初始软伪标签。在第二阶段，本文通过混合去噪和双教师共识去噪策略有效去除了初始软伪标签中的噪声，而后通过蒸馏哈希学习，从有噪声的软伪标签中提取得到更加鲁棒的哈希码，并用于高效的图像检索。

综上所述，本文的主要贡献可以概括为：

(1)提出了一种两阶段的自监督哈希学习框架。该框架通过自监督双教师学习和鲁棒哈希学习两个阶段，提高了伪标签标注精度，增强了哈希网络的稳定性。

(2)提出了一种鲁棒的蒸馏哈希学习方法。该方法从噪声标签的角度出发，通过软伪标签去噪有效降低了软伪标签中的噪声；同时该方法利用蒸馏学习将双教师网络中的信息通过去噪软伪标签传递给学生网络，进而获得无标签图像的鲁棒哈希码。

(3)本文在 3 个常用数据集 CIFAR-10，FLICKR25K 和 EuroSAT 上进行了实验，实验结果表明，本文方法在所有实验条件下均优于其他方法。与 TBH 方法[6]相比，在 CIFAR-10 上本文方法的 MAP 平均提高了 18.6%；与 DistillHash 方法[9]相比，在 FLICKR25K 上本文方法的 MAP 平均提高了 2.4%；与 ETE-GAN 方法[7]相比，在 EuroSAT 上本文方法的 MAP 平均提高了 18.5%，充分验证了本文方法的有效性。

2 相关工作

2.1 深度无监督哈希学习方法

近年来，研究者们提出了多种深度无监督哈希学习方法[8-13]。与深度有监督哈希学习方法相比，深度无监督哈希学习方法可以在没有数据标签的情况下学习哈希表示。深度无监督哈希学习方法中的生成哈希方法通常使用生成模型

来实现哈希编码。例如，TBH 方法[6]通过两个 Bottleneck 结构将自适应图和生成自编码器结合，有效挖掘了图像间的相似性信息；ETE-GAN 方法[7]通过在 GAN 网络训练中加入相似性损失，提高了 GAN 网络的特征判别性；HashGAN 方法[11]通过共享编码器和判别器参数，有效避免了 GAN 网络的过拟合，从而提高了无监督哈希学习的性能。深度无监督哈希学习方法中的自监督哈希方法[9-10,14]则主要基于伪标签生成等代理任务实现哈希学习。例如，DistillHash 方法[9]通过筛选高可信度的相似或不相似图像对构建相应的伪标签，而后利用选出的图像对训练带哈希层的深度网络；SSDH 方法[10]通过设置阈值将图像对伪标签细化为 3 类，在提高图像对伪标签准确性的同时充分利用了训练数据。虽然这些方法在伪标签标注时进行了去噪工作，能够在一定程度上减小标签噪声的影响，但由于伪标签标注算法性能的限制，且未在去噪后做进一步的鲁棒学习处理，导致这些方法学习到的哈希码仍然不够鲁棒。

为此，本文设计了一种改进的聚类算法，有效提高了伪标签标注精度，而后利用混合去噪、双教师共识去噪和双教师蒸馏等多种手段实现了噪声鲁棒的哈希学习，进一步提高了哈希学习的性能。

2.2　自监督学习

自监督学习的核心思想是通过代理任务从无标签数据中学习知识[15]，其中最有代表性的两类是基于对比[16-18]的自监督学习方法和基于伪标签[19-21]的自监督学习方法。在基于对比的自监督学习方法中，SimCLR 方法[16]将图像的变换作为正例，将其他不同图像作为该图像的负例，通过拉近正例距离、拉远负例距离进行学习，但需要设置非常大的训练批次，因此需要高性能硬件的支持；MoCo 方法[17]提出了动量更新训练方式，一定程度上降低了基于对比的自监督学习方法对硬件的要求。在基于伪标签的自监督学习方法中，Deep Cluster[19]方法利用 K-Means 聚类簇作为伪标签来迭代地学习视觉表示，成功将聚类和深度学习进行了结合，但由于 K-Means 算法在聚类时面临空聚类和聚类不平衡的问题[19]，导致该方法的性能相对较低；Asano 等则提出了一种新的伪标签标注方法 SeLa(Self Labeling)[20]，该方法将求解伪标签分配问题作为最优运输问题的一个实例，实现了快速求解，但由于无法使用预训练网络，该方法需要大量训练数据从头开始训练深度网络。

受到基于伪标签自监督学习方法的启发，本文设计了一种基于聚类的伪标签标注方法。首先，本文通过使用等量约束，有效提高了 K-Means 算法的聚类效果，避免了空聚类和聚类不平衡问题[19]；其次，与 Deep Cluster 方法和 SeLa 方法不同的是，本文保留了教师网络的预测概率分布，并将其作为软伪标签来指导学生网络进行哈希学习，而不是直接采用“one-hot”的硬伪标签，从而充分保留了教师网络的知识。此外，本文使用了两个不同的教师网络进行伪标签标注，这使得本文方法对噪声标签具有更强的鲁棒性。

2.3　噪声鲁棒学习

在众多计算机视觉任务中，噪声鲁棒学习是一个重要而又具有挑战性的研究课题[22]。噪声标签的存在对深度学习有较大影响，往往会导致深度模型的性能退化。在深度有监督学习中，已经提出了诸如 Bootstrap[23]和重加权[24]等技术来缓解该问题，这些方法可以充分挖掘训练数据中的有效信息，同时调整每个样本的损失。但是，对损失的调整无法保证始终是正确的，这将导致错误调整的累积，影响损失调整的效果，尤其是当噪声标签比例较大时。为此，本文提出了一种混合去噪方法，从软伪标签置信度和聚类距离两个维度对训练样本进行过滤，有效降低了伪标签中的噪声。

此外，受近期蒸馏学习工作[25-26]的启发，本文在得到去噪软伪标签后，进一步使用知识蒸馏来提高无监督哈希学习的性能。知识蒸馏[27]通常用于模型压缩，即通过蒸馏将教师网络的知识传递给轻量级的学生网络，而后使用学生网络执行下游任务。本文则将知识蒸馏用于提高模型鲁棒性，增强深度网络对噪声标签的耐受能力。与传统的蒸馏方法[27]不同，本文方法中的教师网络和学生网络都是在无监督的情况下进行学习的。同时，为了避免单个教师在预测时的“偏见”，本文采用双教师共同指导学生网络进行哈希学习，有效提高了学生网络的鲁棒性。

3　本文方法

本文方法的主要思路是：通过聚类和双教师蒸馏学习，从两个无监督教师网络中学习鲁棒的哈希表示。对于一个包含 N 张无标签图像的数据集 $X=\{x_i\}_{i=1}^{N}$，深度无监督哈希学习旨在学习非线性哈希函数 Hash(·)，使得原始图像 x_i 可以通过 $Hash(\cdot)$ 被编码成紧凑的 h 比特哈希码 $b_i=\mathrm{Hash}(x_i)$，$b_i\in\{+1,-1\}^h$。由于在无监督场景下没有可供使用的图像标签，若要实现 Hash(·)函数学习，就需要充分挖掘图像内容的语义相关信息[9,14]。为了进一步去除自监督学习过程中噪声标签带来的负面影响，本文在自监督学习过程中将约束聚类和鲁棒学习相结合，提出了一种鲁棒的双教师自监督蒸馏哈希学习方法，实现了带噪声伪标签数据的鲁棒哈希学习。

图 2 给出了本文的总体框架，主要包含自监督双教师学习和鲁棒哈希学习两个阶段。在第一阶段中，首先使用两个教师网络分别对无标签图像进行特征提取；然后利用改进的聚类算法对图像深度特征进行聚类，根据聚类簇划分给图像标注上硬伪标签，如将第一个聚类簇划分中特征对应的图像硬伪标签标注为“[1,0,…,0]”；而后利用得到的硬伪标签分别微调两个教师网络，提高教师网络对新数据的适应性，在完成微调后，即可利用两个教师网络分别对图像进行预测，保留图像的预测概率分布并将其作为初始软伪标签。在第二阶段中，首先利用混合去噪和双教师共识去噪对初始软伪标签进行过滤，得到最终去噪软伪标签；而后利用最终去噪软伪标签对学生网络进行蒸馏哈希学习。在完成学生网络哈希学习后，即可通过学生网络哈希层的输出提取图像哈希码，用于鲁棒的哈希图像检索。

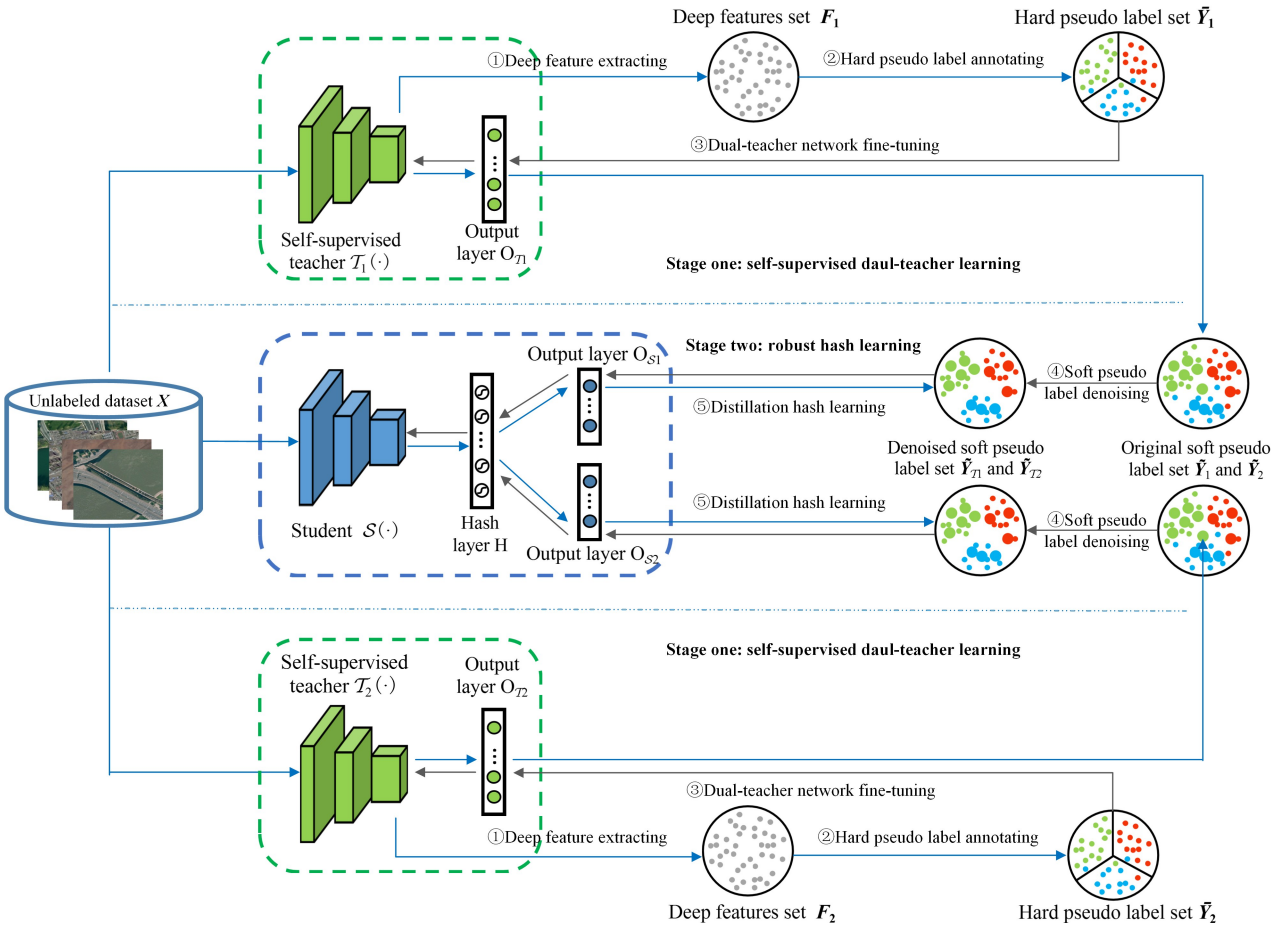


图 2 本文两阶段自监督哈希学习框架

Fig. 2 Our two-stage self-supervised hash learning framework

3.1 自监督双教师学习

自监督双教师学习阶段主要包括深度特征提取、硬伪标签标注和双教师网络微调 3 个步骤。本阶段的目的是：生成无标签图像的初始软伪标签，为鲁棒哈希学习提供监督信息。

3.1.1 深度特征提取

选取两个不同的预训练深度网络 $\mathcal{T}_1(\cdot)$ 和 $\mathcal{T}_2(\cdot)$ 作为自监督双教师网络，其骨干网参数分别为 $\theta_{\mathcal{T}1}$ 和 $\theta_{\mathcal{T}2}$。而后利用 $\mathcal{T}_1(\cdot)$ 和 $\mathcal{T}_2(\cdot)$ 分别对无标签数据集 $\boldsymbol{X}=\{\boldsymbol{x}_i\}_{i=1}^N$ 中的图像进行特征提取，得到深度特征集 $\boldsymbol{F}_1$ 和 $\boldsymbol{F}_2$。

$$\boldsymbol{F}_1=\{\boldsymbol{f}_{i1}\mid \boldsymbol{f}_{i1}=\mathcal{T}_1(\theta_1,x_i)\}_{i=1}^N \tag{1}$$

$$\boldsymbol{F}_2=\{\boldsymbol{f}_{i2}\mid \boldsymbol{f}_{i2}=\mathcal{T}_2(\theta_2,\boldsymbol{x}_i)\}_{i=1}^N \tag{2}$$

其中，$\boldsymbol{f}_{i1}$ 和 $\boldsymbol{f}_{i2}$ 分别为 $\mathcal{T}_1(\cdot)$ 和 $\mathcal{T}_2(\cdot)$ 对图像 x_i 提取到的深度特征。

3.1.2 硬伪标签标注

本文基于聚类对无标签图像的硬伪标签进行标注。为了提高硬伪标签标注的精度，本文设计了一种等量约束聚类算法 K*，该算法在 K-Means 算法的基础上进行了改进。K* 在聚类过程中将数据划分为大小相等的聚类簇，有效避免了 K-Means 算法在聚类过程中的空聚类和聚类不平衡问题[19]。算法 K* 的具体过程如下。

(1)初始化。利用式(1)和式(2)得到深度特征集 $\boldsymbol{F}_1$ 和 $\boldsymbol{F}_2$ 作为聚类样本。设置聚类簇数为 k，最大迭代次数为 M，聚类簇最大容量为 $D=N/k$，当前迭代次数为 $m=0$。

本文算法在深度特征集 $\boldsymbol{F}_1$ 和 $\boldsymbol{F}_2$ 取值范围内分别随机选取 k 个坐标点作为聚类中心集 $\boldsymbol{U}_t$。

$$\boldsymbol{U}_t=\{\boldsymbol{u}_{jt}\}_{j=1}^k,t=1,2 \tag{3}$$

其中，t 表示深度特征集编号，j 表示聚类簇编号，$\boldsymbol{u}_{jt}$ 表示深度特征集 $\boldsymbol{F}_t$ 的第 j 个聚类中心，$\boldsymbol{U}_t$ 表示深度特征集 $\boldsymbol{F}_t$ 对应的聚类中心集。然后初始化聚类簇划分 $\boldsymbol{C}_t$：

$$\boldsymbol{C}_t=\{C_{jt}\mid C_{jt}=\phi\}_{j=1}^k,t=1,2 \tag{4}$$

其中，C_{jt} 表示深度特征集 $\boldsymbol{F}_t$ 对应的第 j 个聚类簇。在初始化过程中，每个聚类簇均置为空集。在后续迭代过程中，聚类簇中的样本会根据样本分配结果动态调整。

(2)样本分配。分别计算 $\boldsymbol{f}_{i1}$ 和 $\boldsymbol{f}_{i2}$ 到各聚类中心 $\boldsymbol{u}_{j1}$ 和 $\boldsymbol{u}_{j2}$ 的距离 d_{ij1} 和 d_{ij2}。在未达到最大容量 D 的聚类簇中，分别选取聚类簇 C_{j1} 和 C_{j2}，使距离 d_{ij1} 和 d_{ij2} 最小，并将 $\boldsymbol{f}_{i1}$ 和 $\boldsymbol{f}_{i2}$ 分别加入聚类簇 C_{j1} 和 C_{j2}。

(3)聚类中心集更新。首先更新上轮聚类中心集 $\boldsymbol{U}_t'=\boldsymbol{U}_t$，然后更新本轮聚类中心集 $\boldsymbol{U}_t$，即分别计算各聚类中心 $\boldsymbol{u}_{jt}$。

$$\boldsymbol{u}_{jt}=\frac{1}{|C_{jt}|}\sum_{\boldsymbol{f}_{it}\in C_{jt}}\boldsymbol{f}_{it},t=1,2 \tag{5}$$

(4)终止条件判断。更新当前迭代次数 $m=m+1$，然后分别对比本轮聚类中心 $\boldsymbol{U}_t$ 和上轮聚类中心 $\boldsymbol{U}_t'$，若聚类中心无变化或 m 达到最大迭代次数 M，则输出聚类簇划分 $\boldsymbol{C}_t$，否则按式(4)将 $\boldsymbol{C}_t$ 各聚类簇置为空集，并返回步骤(2)。

在完成聚类后，按照样本 $\boldsymbol{f}_{i1}$ 和 $\boldsymbol{f}_{i2}$ 所在的聚类簇 C_{j1} 和

C_{j2} 给该样本分配两个“one-hot”硬伪标签 $\bar{\boldsymbol{y}}_{i1}\in\boldsymbol{R}^k$ 和 $\bar{\boldsymbol{y}}_{i2}\in\boldsymbol{R}^k$。由此可以得到数据集 $\boldsymbol{X}$ 对应的两个硬伪标签集 $\bar{\boldsymbol{Y}}_1=\{\bar{\boldsymbol{y}}_{i1}\}_{i=1}^N$ 和 $\bar{\boldsymbol{Y}}_2=\{\bar{\boldsymbol{y}}_{i2}\}_{i=1}^N$。

3.1.3　双教师网络微调

双教师网络微调的目的在于提高两个教师网络对新数据的适应能力,进而利用微调后的两个教师网络对无标签图像进行预测,得到无标签图像的初始软伪标签,并将其用于指导鲁棒哈希学习。

在得到无标签数据集 $\boldsymbol{X}$ 对应的硬伪标签集 $\bar{\boldsymbol{Y}}_1$ 和 $\bar{\boldsymbol{Y}}_2$ 后,本文以 $\bar{\boldsymbol{Y}}_1$ 和 $\bar{\boldsymbol{Y}}_2$ 为监督信息,并通过 CE(Cross Entropy) 损失分别对两个教师网络 $\mathcal{T}_1(\cdot)$ 和 $\mathcal{T}_2(\cdot)$ 进行微调。假设 $\mathcal{T}_1(\cdot)$ 和 $\mathcal{T}_2(\cdot)$ 对图像 $\boldsymbol{x}_i$ 的预测概率分布分别为 $\tilde{\boldsymbol{y}}_{i1}\in\boldsymbol{R}^k$ 和 $\tilde{\boldsymbol{y}}_{i2}\in\boldsymbol{R}^k$,则可利用硬伪标签构建 CE 损失,用于微调两个教师网络 $\mathcal{T}_1(\cdot)$ 和 $\mathcal{T}_2(\cdot)$ 的参数 $\theta_{\mathcal{T}1}$ 和 $\theta_{\mathcal{T}2}$。

$$L_{CE1}=-\sum_{i=1}^{N}\sum_{j=1}^{k}\bar{\boldsymbol{y}}_{i1}^j\log(\tilde{\boldsymbol{y}}_{i1}^j)\tag{6}$$

$$L_{CE2}=-\sum_{i=1}^{N}\sum_{j=1}^{k}\bar{\boldsymbol{y}}_{i2}^j\log(\tilde{\boldsymbol{y}}_{i2}^j)\tag{7}$$

微调两个教师网络时,采用随机梯度下降法对教师网络参数进行更新。在完成教师网络微调后,即可分别使用两个教师网络 $\mathcal{T}_1(\cdot)$ 和 $\mathcal{T}_2(\cdot)$ 对无标签图像 $\boldsymbol{x}_i$ 进行预测,得到该图像的初始软伪标签 $\tilde{\boldsymbol{y}}_{i1}\in\boldsymbol{R}^k$ 和 $\tilde{\boldsymbol{y}}_{i2}\in\boldsymbol{R}^k$,进而得到数据集 $\boldsymbol{X}$ 对应的两个初始软伪标签集 $\tilde{\boldsymbol{Y}}_1=\{\tilde{\boldsymbol{y}}_{i1}\}_{i=1}^N$ 和 $\tilde{\boldsymbol{Y}}_2=\{\tilde{\boldsymbol{y}}_{i2}\}_{i=1}^N$。

3.2　鲁棒哈希学习

鲁棒哈希学习包括软伪标签去噪和学生网络蒸馏哈希学习两个步骤。本阶段的主要目的是:首先通过软伪标签去噪去除伪标签中的部分噪声,提高伪标签精度;然后通过蒸馏哈希学习进一步提高学生网络对噪声的鲁棒性,实现鲁棒的哈希检索。

3.2.1　软伪标签去噪

本文利用双教师网络的预测概率分布作为图像的初始软伪标签,软伪标签比硬伪标签包含更多的语义信息,但仍无法避免噪声标签带来的负面影响。为此,本文设计了一种简单有效的软伪标签去噪方法,该方法包含混合去噪和双教师共识去噪两种策略。

混合去噪策略充分融合了置信度去噪和距离去噪的优势,能够有效去除初始软伪标签集 $\tilde{\boldsymbol{Y}}_1$ 和 $\tilde{\boldsymbol{Y}}_2$ 中的噪声,得到混合去噪软伪标签集 $\tilde{\boldsymbol{Y}}_{h1}$ 和 $\tilde{\boldsymbol{Y}}_{h2}$。在完成混合去噪后,双教师共识去噪策略能进一步保留 $\tilde{\boldsymbol{Y}}_{h1}$ 和 $\tilde{\boldsymbol{Y}}_{h2}$ 对应图像公共部分的软伪标签,得到最终的去噪软伪标签集 $\tilde{\boldsymbol{Y}}_1$ 和 $\tilde{\boldsymbol{Y}}_2$。

(1)混合去噪策略

1)置信度去噪。在使用硬伪标签集 $\bar{\boldsymbol{Y}}_1$ 和 $\bar{\boldsymbol{Y}}_2$ 微调教师网络时,正确的硬伪标签往往更容易被教师网络拟合,其对应的软伪标签预测概率较高,而错误硬伪标签对应的软伪标签预测概率则相对较低[28]。因此,我们可以通过设置阈值 e 对初始软伪标签进行筛选,只保留置信度高于 e 的软伪标签,由此得到置信度去噪软伪标签集 $\tilde{\boldsymbol{Y}}_{c1}$ 和 $\tilde{\boldsymbol{Y}}_{c2}$($\tilde{\boldsymbol{Y}}_{c1}\subseteq\tilde{\boldsymbol{Y}}_1$,$\tilde{\boldsymbol{Y}}_{c2}\subseteq\tilde{\boldsymbol{Y}}_2$)。

2)距离去噪。本文的硬伪标签集 $\bar{\boldsymbol{Y}}_1$ 和 $\bar{\boldsymbol{Y}}_2$ 是基于聚类算法得到的。聚类时在高维深度特征空间中,各类别交界处的样本往往更容易混淆。因此,本文根据深度特征 $\boldsymbol{f}_{ti}$ 到各自聚类中心 $\boldsymbol{u}_{jt}$ 的距离,在初始软伪标签集中按比例 r 保留距离聚类中心 $\boldsymbol{u}_{jt}$ 最近样本对应的软伪标签,得到距离去噪软伪标签集 $\tilde{\boldsymbol{Y}}_{d1}$ 和 $\tilde{\boldsymbol{Y}}_{d2}$($\tilde{\boldsymbol{Y}}_{d1}\subseteq\tilde{\boldsymbol{Y}}_1$,$\tilde{\boldsymbol{Y}}_{d2}\subseteq\tilde{\boldsymbol{Y}}_2$)。

在完成以上两个去噪步骤后,分别保留 $\tilde{\boldsymbol{Y}}_{c1}$ 和 $\tilde{\boldsymbol{Y}}_{d1}$、$\tilde{\boldsymbol{Y}}_{c2}$ 和 $\tilde{\boldsymbol{Y}}_{d2}$ 的公共部分,即可得到混合去噪软伪标签集 $\tilde{\boldsymbol{Y}}_{h1}$ 和 $\tilde{\boldsymbol{Y}}_{h2}$($\tilde{\boldsymbol{Y}}_{h1}=\tilde{\boldsymbol{Y}}_{c1}\cap\tilde{\boldsymbol{Y}}_{d1}$,$\tilde{\boldsymbol{Y}}_{h2}=\tilde{\boldsymbol{Y}}_{c2}\cap\tilde{\boldsymbol{Y}}_{d2}$)。

(2)双教师共识去噪策略

值得注意的是,在得到混合去噪软伪标签集 $\tilde{\boldsymbol{Y}}_{h1}$ 和 $\tilde{\boldsymbol{Y}}_{h2}$ 后,由于 $\tilde{\boldsymbol{Y}}_{h1}$ 和 $\tilde{\boldsymbol{Y}}_{h2}$ 为使用两个不同教师网络得到的混合去噪软伪标签,因此 $\tilde{\boldsymbol{Y}}_{h1}$ 和 $\tilde{\boldsymbol{Y}}_{h2}$ 对应的图像集 $\boldsymbol{X}_1$ 和 $\boldsymbol{X}_2$ 可能不同。基于 $\tilde{\boldsymbol{Y}}_{h1}$ 对应的图像集 $\boldsymbol{X}_1\subseteq\boldsymbol{X}$ 和 $\tilde{\boldsymbol{Y}}_{h2}$ 对应的图像集 $\boldsymbol{X}_2\subseteq\boldsymbol{X}$,求取 $\boldsymbol{X}_1$ 和 $\boldsymbol{X}_2$ 的公共部分 $\boldsymbol{X}^*=\boldsymbol{X}_1\cap\boldsymbol{X}_2$,即可得到 $\boldsymbol{X}^*$ 对应的最终去噪软伪标签集 $\tilde{\boldsymbol{Y}}_{\mathcal{T}1}$ 和 $\tilde{\boldsymbol{Y}}_{\mathcal{T}2}$($\tilde{\boldsymbol{Y}}_{\mathcal{T}1}\subseteq\tilde{\boldsymbol{Y}}_{h1}$,$\tilde{\boldsymbol{Y}}_{\mathcal{T}2}\subseteq\tilde{\boldsymbol{Y}}_{h2}$)。最终去噪软伪标签集 $\tilde{\boldsymbol{Y}}_{\mathcal{T}1}$ 和 $\tilde{\boldsymbol{Y}}_{\mathcal{T}2}$ 分别为教师网络 $\mathcal{T}_1(\cdot)$ 和 $\mathcal{T}_2(\cdot)$ 对 $\boldsymbol{X}^*$ 标注的两组软伪标签集,而 $\boldsymbol{X}^*$ 则为最终用于训练学生网络的图像集合。

3.2.2　学生网络蒸馏哈希学习

在学生网络蒸馏哈希学习中,本文使用两个教师网络 $\mathcal{T}_1(\cdot)$ 和 $\mathcal{T}_2(\cdot)$ 共同指导学生网络 $S(\cdot)$ 进行学习,即利用最终去噪软伪标签集 $\tilde{Y}_{\mathcal{T}1}$ 和 $\tilde{Y}_{\mathcal{T}2}$ 共同监督学生网络 $S(\cdot)$ 进行学习。本文使用双教师共同指导学生网络哈希学习的原因是:单个教师网络对数据可能存在“偏见”,会导致在学生网络蒸馏哈希学习过程中无法得到最优的学习效果。相反,当使用两个不同教师网络共同指导学生网络哈希学习时,本文的双教师学习机制可以纠正单个教师网络的误差,有效减轻单个教师噪声标签的负面影响。

在学生网络蒸馏哈希学习时,本文采用 KL(Kullback-Leibler) 损失指导学生网络 $S(\cdot)$ 学习。假设 $S(\cdot)$ 对图像 $\boldsymbol{x}_i$ 的预测概率分布为 $\tilde{y}_{iSt}\in\boldsymbol{R}^k$,则学生网络蒸馏哈希学习的损失可表示为:

$$L_{KL}=\sum_{t=1,2}\sum_{i=1}^{N'}\tilde{\boldsymbol{y}}_{it}\log\frac{\tilde{\boldsymbol{y}}_{it}}{\tilde{\boldsymbol{y}}_{iSt}}\tag{8}$$

其中,N' 为 $\boldsymbol{X}^*$ 的数据量。通过最小化 L_{KL} 即可求解得到学生网络 $S(\cdot)$ 的参数 θ_S。

学生网络 $S(\cdot)$ 包含由 tanh 函数激活的哈希层 H。在求解得到 $S(\cdot)$ 参数后,通过对哈希层 H 输出的哈希特征 $\boldsymbol{v}$ 进行二值化即可得到图像对应的哈希码。假设 $S(\cdot)$ 的哈希层 H 对图像 $\boldsymbol{x}_i$ 提取的哈希特征为 $\boldsymbol{v}_i\in\boldsymbol{R}^h$,$h$ 为哈希码长度,则图像 $\boldsymbol{x}_i$ 的哈希码 $\boldsymbol{b}_i$ 可表示为:

$$\boldsymbol{b}_i=\mathrm{sign}(\boldsymbol{v}_i)\tag{9}$$

4　实验

4.1　数据集

本文共选取了 3 个公开数据集,包含 2 个自然图像数据集 CIFAR-10 和 FLICKR25K,以及 1 个遥感场景数据集 EuroSAT。各数据集的具体参数如下:1) CIFAR-10,该

数据集是一个常用的单标签图像数据集，共包含 10 个类别，每类6000 张，共 60000 张图像；2）FLICKR25K，该数据集是一个常用的多标签不平衡图像数据集，共包含 24 个类别，各类别图像数量从几百到几千不等，共 25000 张图像；3）EuroSAT，该数据集是一个常用的单标签遥感场景图像数据集，共包含 10 个类别，每类 2000～3000 张，共 27000 张图像。

4.2 实验设置

本文全部实验均基于 PyTorch1.4 深度学习框架，使用了 2 块 Geforce RTX 2080 Ti 显卡进行测试。等量约束聚类算法 K^* 最大迭代次数 M 为 10。在自监督双教师学习阶段，双教师网络采用 ResNet-101 和 ResNet-152 作为骨干网络，采用随机梯度下降优化器微调教师网络参数，学习率设置为 0.001，训练批次大小设置为 32，训练次数设置为 10 轮。在鲁棒哈希学习阶段，采用随机梯度下降优化器更新学生网络参数，学习率设置为 0.01，训练批次大小设置为 32，训练次数设置为 50，置信度去噪阈值 e 为 0.8，距离去噪保留比例 r 为 0.85。为统一标准，本文实验采用 DistillHash 方法[9]中所使用的深度网络 VGG16 作为学生网络的骨干网。本文选取两种常用指标 MAP（Mean Average Precision）和 P-R（Precision-Recall）作为实验评价标准。

在数据集划分上，对于 CIFAR-10 和 FLICKR25K 数据集，本文参考 DistillHash 方法的设置，CIFAR-10 的训练集为从每类图像中随机抽取 500 张，测试集为从每类图像中随机抽取 1000 张，数据库为除测试集外的 50000 张图像；FLICKR25K 的训练集为从 25000 张图像中随机抽取 5000 张，测试集为从其余图像中随机抽取 2000 张，数据库为除测试集外的 23000 张图像；对于 EuroSAT 数据集，本文参考 ETE-GAN 方法[7]的设置，训练集为从每类图像中随机抽取 500 张，测试集为从每类图像中随机抽取 100 张，数据库为除测试集外的 26000 张图像。

4.3 实验结果与分析

4.3.1 检索结果对比

在 CIFAR-10 和 FLICKR25K 上，本文与 13 种方法进行了比较，其中 SH[29]，LSH[30]，PCAH[31]，SpH[32]，DSH[33]，ITQ[34]为传统无监督哈希学习方法，而 SGH[13]，DeepBit[12]，BGAN[8]，SSDH[10]，SSDH＋PSO[35]，TBH[6]，DistillHash[9]为深度无监督哈希学习方法，如表 1 所列。从表 1 中可以看出，本文方法的检索精度在两个数据集上所有长度哈希码下均达到了最优性能，这是因为本文提出的双教师自监督蒸馏哈希方法能够有效降低伪标签中的噪声，显著提高哈希检索性能。在 CIFAR-10 上，本文方法的 MAP 比 TBH 方法[6]平均提高了 18.6%；在 FLICKR25K 上，本文方法的 MAP 比 DistillHash[9]方法平均提高了 2.4%。从以上分析可以看出，本文方法在各种条件下均有较好表现，特别是在单标签图像检索数据集 CIFAR-10 上的提升效果明显。

表 1 在 CIFAR-10 和 FLICKR25K 上图像检索精度的对比

Table 1 Comparison of image retrieval accuracy on CIFAR-10 and FLICKR25K

Methods	CIFAR-10(MAP@50000)				FLICKR25K(MAP@23000)			
	16 bit	32 bit	64 bit	128 bit	16 bit	32 bit	64 bit	128 bit
SH[29]	0.161	0.158	0.151	0.154	0.592	0.592	0.602	0.621
LSH[30]	0.132	0.158	0.167	0.179	0.583	0.589	0.593	0.601
PCAH[31]	0.143	0.159	0.173	0.184	0.609	0.611	0.603	0.607
SpH[32]	0.144	0.167	0.178	0.184	0.611	0.603	0.634	0.625
DSH[33]	0.162	0.188	0.192	0.206	0.607	0.612	0.612	0.615
ITQ[34]	0.194	0.209	0.215	0.219	0.619	0.632	0.635	0.648
SGH[13]	0.180	0.183	0.189	0.190	0.616	0.628	0.625	0.621
DeepBit[12]	0.220	0.241	0.252	0.253	0.593	0.593	0.620	0.635
BGAN[8]	0.525	0.531	0.562	—	—	—	—	—
SSDH[10]	0.257	0.256	0.259	0.260	0.662	0.673	0.673	0.677
SSDH+PSO[35]	0.286	0.286	0.287	—	0.692	0.692	0.699	—
TBH[6]	0.532	0.573	0.578	—	—	—	—	—
DistillHash[9]	0.284	0.285	0.287	0.290	0.696	0.706	0.708	0.700
Ours	**0.733**	**0.747**	**0.760**	**0.765**	**0.712**	**0.724**	**0.731**	**0.739**

注：每列加粗字体为最优结果，加下划线字体为次优结果

在 EuroSAT 数据集上，本文与 7 种方法进行了比较，其中 SKLSH[36]，SP[37]，SpH[32]，ITQ[34]为传统无监督哈希学习方法，而 GAN_MS_ESM[7]，CNN_MS[7]，ETE-GAN[7]为深度无监督哈希学习方法，如表 2 所列。从表 2 中可以看出，本文方法在所有长度哈希码下均达到了最优性能。本文方法的 MAP 比 ETE-GAN[7]方法平均提高了 18.5%，这充分说明本文方法在遥感场景图像检索领域也有较好表现，进一步验证了本文方法的泛化能力。

此外，本文在 FLICKR25K 数据集上绘制了 10 种方法的 P-R 曲线，以便对各种方法进行更全面的比较，结果如图 3 和图 4 所示。从图 3 和图 4 中可以看出，本文方法的 P-R 曲线明显优于其他方法的 P-R 曲线性能，这也充分说明了本文方法可以达到比其他方法更高的精度。

表 2 在 EuroSAT 上图像检索精度对比

Table 2 Comparison of image retrieval accuracy on EuroSAT

Methods	EuroSAT(MAP@100)		
	32 bit	64 bit	128 bit
SKLSH[36]	0.138	0.144	0.140
SpH[32]	0.445	0.477	0.512
ITQ[34]	0.455	0.502	0.505
GAN_MS_ESM[7]	0.647	0.656	0.679
CNN_MS[7]	0.275	0.284	0.272
ETE-GAN[7]	0.615	0.658	0.683
Ours	**0.820**	**0.839**	**0.853**

注：每列加粗字体为最优结果，加下划线字体为次优结果

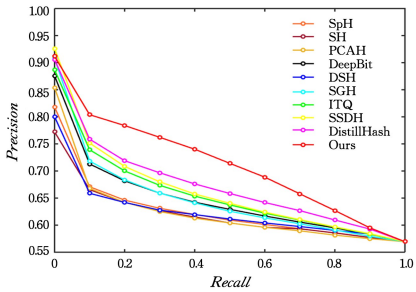


图 3　在 FLICKR25K 上 16 位哈希码 P-R 曲线

Fig. 3　P-R curves with 16 bits hash codes on FLICKR25K

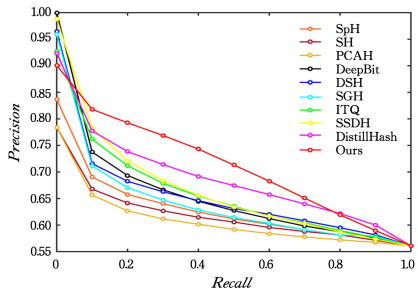


图 4　在 FLICKR25K 上 32 位哈希码 P-R 曲线

Fig. 4　P-R curves with 32 bits hash codes on FLICKR25K

4.3.2　消融实验

为了验证本文方法各组成部分的作用，本文在 CIFAR-10 上设计了 6 组消融实验，对硬伪标签生成方法、混合去噪策略及双教师去噪蒸馏进行了消融分析，实验结果如表 3 所列。表 3 中，前 3 组实验主要是为了对不同硬伪标签生成方法的性能进行对比，因此均以硬伪标签为监督信息，学生网络直接利用硬伪标签构建 CE 损失以进行自监督学习；后 3 组实验主要是为了对混合去噪策略及双教师去噪蒸馏的性能进行对比，均以软伪标签为监督信息，学生网络均利用不同类型的软伪标签构建 KL 损失以进行自监督蒸馏学习。

对比表 3 中的实验 1—实验 3 可知，相比 Deep Cluster[19] 和 SeLa[20] 硬伪标签生成方法，本文设计的等量约束聚类硬伪标签生成方法能够进一步提高硬伪标签的标注精度，使得哈希检索 MAP 相比 Deep Cluster 硬伪标签生成方法平均提高了 51.4%，相比 SeLa 硬伪标签生成方法平均提高了 30.2%。

对比表 3 中的实验 3 和实验 4 可知，当均采用本文等量约束硬伪标签生成方法时，利用软伪标签 KL 损失进行自监督蒸馏学习比利用硬伪标签 CE 损失进行自监督分类学习得到的哈希检索性能更好，哈希检索 MAP 提高了 1.7%。这说明利用软伪标签进行自监督蒸馏学习能够有效减轻伪标签中的噪声对哈希学习的负面影响，提高学生网络的噪声鲁棒性，进而提升哈希检索的性能。

对比表 3 中的实验 4 和实验 5 可知，当实验 5 引入混合去噪策略后，相比实验 4 能进一步去除软伪标签中的噪声，使得哈希检索 MAP 提高 1.5%。这说明本文的混合去噪策略有效降低了软伪标签中的噪声，提高了软伪标签的可用性。进一步地对比表 3 中的实验 5 和实验 6 可知，当实验 6 进一步引入双教师共识去噪策略和双教师蒸馏学习后，软伪标签中的噪声可以得到进一步过滤，同时双教师蒸馏哈希方法充分利用了两个教师的“共识”，进一步提升了哈希检索性能，使得哈希检索 MAP 提高 4.4%。

表 3　在 CIFAR-10 上的图像检索消融实验

Table 3　Ablation experiments of image retrieval on CIFAR-10

Number	Settings	CIFAR-10(MAP@50000)			
		16 bit	32 bit	64 bit	128 bit
1	ResNet-101+Deep Cluster[19] hard pseudo labels	0.133	0.161	0.171	0.181
2	ResNet-101+SeLa[20] hard pseudo labels	0.341	0.366	0.396	0.391
3	ResNet-101+our hard pseudo labels	0.667	0.676	0.684	0.674
4	ResNet-101+our soft pseudo labels	0.681	0.689	0.693	0.706
5	ResNet-101+our soft pseudo labels+hybrid denoising	0.695	0.706	0.711	0.716
6	(ResNet-101+ResNet-152)+our soft pseudo labels+hybrid denoising	**0.733**	**0.747**	**0.760**	**0.765**

注：每列加粗字体为最优结果，加下划线字体为次优结果

综合以上分析，本文方法通过设计等量约束聚类硬伪标签生成方法、混合去噪策略以及双教师去噪蒸馏 3 种机制，较好地解决了自监督深度哈希学习中噪声标签过多及噪声标签过拟合问题。等量约束聚类硬伪标签生成方法通过在传统 *K*-Means 算法中加入等量约束，有效提高了硬伪标签的标注精度；混合去噪策略在置信度和距离两个维度上对噪声标签进行过滤，去除了软伪标签中的部分噪声；双教师去噪蒸馏则通过双教师共识及软伪标签蒸馏学习有效提高了学生网络的鲁棒性。

4.3.3　硬伪标签可视化

硬伪标签精度反映了伪标签中的噪声比例，对软伪标签生成和蒸馏哈希学习均有较大影响。为了进一步说明本文硬伪标签生成方法的有效性，我们利用 ResNet-101 作为教师网络，使用本文设计的等量约束聚类硬伪标签生成方法，针对 CIFAR-10 和 EuroSAT 的训练集数据进行了硬伪标签混淆矩阵可视化，结果如图 5 和图 6 所示。

从图 5 和图 6 中可以看出，本文方法得到的硬伪标签均保持了较高精度。在 CIFAR-10 数据集上，硬伪标签总精度超过了 79%，而在 EuroSAT 数据集上也超过了 73%。在两个数据集内部，由于类内差异、类间差异各不相同，各类别精度也不尽相同，如 CIFAR-10 数据集中第 2 类和第 7 类精度超过了 90%，而第 3 类却不足 70%；EuroSAT 数据集中第 1 类、第 7 类和第 9 类精度也超过了 90%，而第 4 类、第 5 类和第 10 类则出现了较大比例的混淆，导致硬伪标签精度较低。因此，如何增强此类易混淆类别间的可区分度，是下一步提升硬伪标签精度的关键。

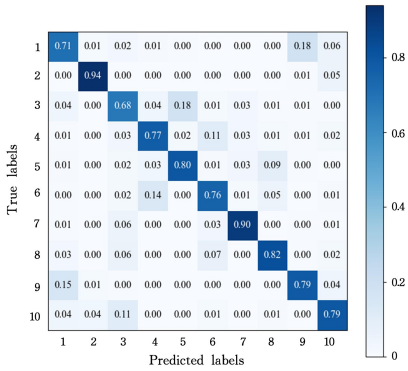


图 5 在 CIFAR-10 上的硬伪标签混淆矩阵

Fig. 5 Hard pseudo label confusion matrix on CIFAR-10

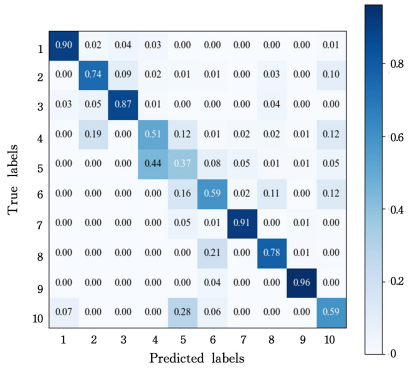


图 6 在 EuroSAT 上的硬伪标签混淆矩阵

Fig. 6 Hard pseudo label confusion matrix on EuroSAT

4.3.4 混合去噪超参数设置

混合去噪过程包含置信度去噪和距离去噪两个步骤。置信度去噪中的阈值 e 和距离去噪中的保留比例 r 对去噪性能有一定影响。为进一步确定超参数 e 及 r 的取值，本文分别以 ResNet-101 和 ResNet-152 为教师的骨干网，在 CIFAR-10 数据集上 64 位哈希码下进行了测试，实验结果如图 7 和图 8 所示。

图 7 给出了当置信度去噪阈值 e 取不同值时，CIFAR-10 上不同教师网络对应的检索精度。其中，横坐标 e 表示置信度去噪阈值，e 的取值越大，保留的软伪标签置信度就越高，此时被过滤的噪声数据也越多；而当阈值 e 取值为 0 时，则表示不进行置信度去噪，此时保留所有软伪标签。图 8 给出了当距离去噪保留比例 r 取不同值时，CIFAR-10 上不同教师网络对应的检索精度。图 8 中，横坐标 r 表示距离去噪的保留比例，当 r 取值为 1 时，表示保留所有软伪标签。

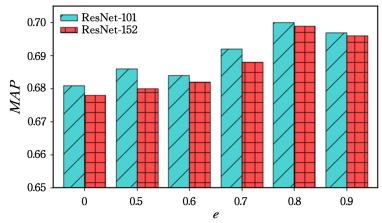


图 7 在 CIFAR-10 上不同置信度去噪阈值 e 的检索精度

Fig. 7 Retrieval accuracy with different e on CIFAR-10

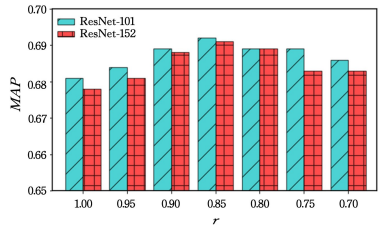


图 8 在 CIFAR-10 上不同距离去噪保留比例 r 的检索精度

Fig. 8 Retrieval accuracy with different r on CIFAR-10

从图 7 可以看出，当阈值 e 取 0 时，此时未进行任何去噪操作，MAP 最低。当阈值 e 逐渐增大时，MAP 开始提高，说明软伪标签中噪声得到了有效去除；当阈值 e 取 0.8 时，两个教师网络对应的 MAP 均达到最大值；当阈值 e 继续增大时，MAP 开始下降，这可能是因为阈值 e 过大会过滤掉过多软伪标签，导致训练数据不足，进而使得性能下降。

从图 8 可以看出，当保留比例 r 取 1 时，此时未进行任何去噪操作，MAP 最低；当保留比例 r 逐渐减小时，MAP 开始提高，说明通过丢弃远离聚类中心的数据可以去除噪声标签；当保留比例 r 取 0.85 时，两个教师网络对应的 MAP 均达到最大值。

综合以上分析，置信度去噪阈值 e 取 0.8，距离去噪保留比例 r 取 0.85 时，本文提出的混合去噪方法的性能达到最优。

4.3.5 哈希图像检索结果

为了对本文哈希检索方法的效果进行可视化，我们在 EuroSAT 数据集上利用 32，64 和 128 位 3 种不同长度的哈希码进行了图像检索，如图 9 所示。

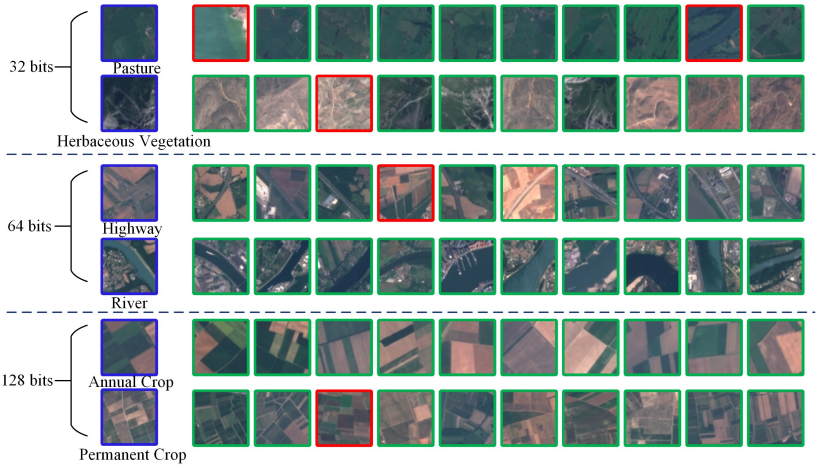


图 9 在 EuroSAT 上不同哈希编码的图像检索结果(电子版为彩图)

Fig. 9 Image retrieval results of different hash bits on EuroSAT

图 9 给出了 3 种条件下部分图像按相似度由大到小排名前 10 的检索结果，其中最左侧一列蓝色框图像为查询图像，右侧为对应图像的相似度由大到小排名前 10 的检索结果，绿色框表示正确检索的结果，红色框表示错误检索的结果。从图 9 中可以看出，该数据集存在类内差异大、类间差异小的现象，例如在 Herbaceous Vegetation 类别中，同时存在绿色外观图像和黄色外观图像，而 Annual Crop 类别和 Permanent Crop 类别间的纹理则较为相似。即使在这种困难情况下，本文方法仍可以取得较好的检索结果。这进一步说明，本文双教师自监督蒸馏哈希学习方法可有效提升哈希学习的鲁棒性。

结束语　本文提出了一种鲁棒的双教师自监督蒸馏哈希学习方法，该方法不需要人工标签就可以实现高效的哈希图像检索。在该方法的第一阶段中，首先通过等量约束聚类算法对无标签图像深度特征进行聚类，得到图像硬伪标签；其次利用硬伪标签对教师网络进行微调，使得教师网络进一步适应数据，提升教师网络的特征提取能力；最后利用教师网络对图像进行预测，保留图像的预测概率分布并将其作为初始软伪标签。在该方法的第二阶段中，本文通过混合去噪策略，有效降低了初始软伪标签中的噪声；而后利用双教师的“共识”，保留两组混合去噪软伪标签的公共数据，得到最终的去噪软伪标签并将其作为学生网络蒸馏哈希学习的监督信息，进一步提升了检索性能。本文在 CIFAR-10，FLICKR25K 和 EuroSAT 这 3 个公开数据集上进行了测试，本文方法的平均检索精度比当前主流方法分别提高了 18.6%，2.4%和 18.5%。实验结果表明，本文方法对各类数据均有良好的鲁棒性。

参 考 文 献

[1] WANG J，KUMAR S，CHANG S F. Semi-supervised hashing for large-scale search[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2012，34(12)：2393-2406.

[2] SHARIF U，MEHMOOD Z，MAHMOOD T，et al. Scene analysis and search using local features and support vector machine for effective content-based image retrieval[J]. Artificial Intelligence Review，2019，52(2)：901-925.

[3] PENG Y，ZHANG J，YE Z. Deep reinforcement learning for image hashing[J]. IEEE Transactions on Multimedia，2019，22(8)：2061-2073.

[4] ZHANG B，QIAN J. Autoencoder-based unsupervised clustering and hashing[J]. Applied Intelligence，2021，51(1)：493-505.

[5] GOODFELLOW I J，POUGET-ABADIE J，MIRZA M，et al. Generative adversarial networks[C]// Proceedings of Advances in Neural Information Processing Systems. 2014：2672-2680.

[6] SHEN Y，QIN J，CHEN J，et al. Auto-encoding twin-bottleneck hashing[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE，2020：2815-2824.

[7] CHEN X，LU C. An end-to-end adversarial hashing method for unsupervised multispectral remote sensing image retrieval[C]// Proceedings of the IEEE International Conference on Image Processing. IEEE，2020：1536-1540.

[8] SONG J，HE T，GAO L，et al. Binary generative adversarial networks for image retrieval[C]// Proceedings of the AAAI Conference on Artificial Intelligence. AAAI，2018：394-401.

[9] YANG E，LIU T，DENG C，et al. Distillhash：unsupervised deep hashing by distilling data pairs[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE，2019：2946-2955.

[10] YANG E，DENG C，LIU T，et al. Semantic structure-based unsupervised deep hashing[C]// Proceedings of the International Joint Conference on Artificial Intelligence. 2018：1064-1070.

[11] DIZAJI G，ZHENG F，SADOUGHI N，et al. Unsupervised deep generative adversarial hashing network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE，2018：3664-3673.

[12] LIN K，LU J，CHEN C S，et al. Learning compact binary descriptors with unsupervised deep neural networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE，2016：1183-1192.

[13] DAI B，GUO R，KUMAR S，et al. Stochastic generative hashing[C]// Proceedings of the International Conference on Machine Learning. PMLR，2017：913-922.

[14] LI Y，WANG Y，MIAO Z，et al. Contrastive self-supervised hashing with dual pseudo agreement[J]. IEEE Access，2020，8：165034-165043.

[15] HUANG J，DONG Q，GONG S，et al. Unsupervised deep learning by neighbourhood discovery[C]// Proceedings of the International Conference on Machine Learning. PMLR，2019，97：2849-2858.

[16] CHEN T，KORNBLITH S，NOROUZI M，et al. A simple framework for contrastive learning of visual representations[C]// Proceedings of the International Conference on Machine Learning. PMLR，2020，119：1597-1607.

[17] HE K，FAN H，WU Y，et al. Momentum contrast for unsupervised visual representation learning[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE，2020：9726-9735.

[18] BACHMAN P，HJELM R，BUCHWALTER W. Learning representations by maximizing mutual information across views[C]// Proceedings of Advances in Neural Information Processing Systems. 2019：15509-15519.

[19] 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. Berlin：Springer，2018：139-156.

[20] ASANO Y，RUPPRECHT C，VEDALDI A. Self-labelling via simultaneous clustering and representation learning[C]// Proceedings of the International Conference on Learning Representations，OpenReview，2020.

[21] CARON M，MISRA I，MAIRAL J，et al. Unsupervised learning of visual features by contrasting cluster assignments[C]// Proceedings of Advances in Neural Information Processing Systems. 2020.

[22] HAN B，YAO Q，LIU T，et al. A survey of label-noise represen-

tation learning: past, present and future[EB/OL]. (2021-02-20)[2021-04-27]. https://arxiv.org/abs/2011.04406.

[23] REED S, LEE H, ANGUELOV D, et al. Training deep neural networks on noisy labels with bootstrapping[C]// Proceedings of the International Conference on Learning Representations. 2015.

[24] SHU J, XIE Q, YI L, et al. Meta-weight-net: Learning an explicit mapping for sample weighting[C]// Proceedings of Advances in Neural Information Processing Systems. 2019: 1919-1930.

[25] LI Y, YANG J, SONG Y, et al. Learning from noisy labels with distillation[C]// Proceedings of the International Conference on Computer Vision. IEEE, 2017: 1928-1936.

[26] HUANG Z, ZOU Y, BHAGAVATULA V, et al. Comprehensive attention self-distillation for weakly-supervised object detection[C]// Proceedings of Advances in Neural Information Processing Systems. 2020.

[27] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[J]. Computer Science, 2015, 14(7): 38-39.

[28] NORTHCUTT C, JIANG L, CHUANG I. Confident Learning: Estimating Uncertainty in Dataset Labels[J]. Journal of Artificial Intelligence Research, 2021, 70: 1373-1411.

[29] WEISS Y, TORRALBA A, FERGUS R. Spectral hashing[C]// Proceedings of Advances in Neural Information Processing Systems, Curran Associates, Inc.. 2008: 1753-1760.

[30] ANDONI A, INDYK P. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions[J]. Communications of the ACM, 2008, 51(1): 117-122.

[31] WANG X, ZHANG L, JING F, et al. AnnoSearch: Image auto-annotation by search[C]// Proceedings of the 2006 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2006: 1483-1490.

[32] HEO J, LEE Y, HE J, et al. Spherical hashing[C]// Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 2957-2964.

[33] JIN Z, LI C, LIN Y, et al. Density sensitive hashing[J]. IEEE Transactions on Cybernetics, 2014, 44(8): 1362-1371.

[34] GONG Y, LAZEBNIK S, GORDO A, et al. Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(12): 2916-2929.

[35] WEI Y, TIAN D, SHI J, et al. Optimizing non-differentiable metrics for hashing[J]. IEEE Access, 2021, 9: 14351-14357.

[36] RAGINSKY M, LAZEBNIK S. Locality-sensitive binary codes from shift-invariant kernels[C]// Proceedings of Advances in Neural Information Processing Systems, Curran Associates, Inc.. 2009: 1509-1517.

[37] XIA Y, HE K, KOHLI P, et al. Sparse projections for high-dimensional binary codes[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015: 3332-3339.

MIAO Zhuang, born in 1976, associate professor, is a member of China Computer Federation. His main research interests include artificial intelligence, pattern recognition and computer vision.

LI Yang, born in 1984, associate professor, is a member of China Computer Federation. His main research interests include computer vision, deep learning and image processing.

(责任编辑:喻藜)