

# 基于样本旋转的生成困难样本的深度度量学习方法

张鸽<sup>1,2</sup>, 闫京<sup>1,2</sup>, 魏巍<sup>1,2\*</sup>, 梁吉业<sup>1,2</sup>

(1. 山西大学 计算机与信息技术学院, 山西 太原 030006;

2. 山西大学 计算智能与中文信息处理教育部重点实验室, 山西 太原 030006)

**摘要:** 现有深度度量学习方法通过构造困难样本生成方法指导模型高效训练, 基于代数计算的困难样本生成方法具有简单、高效的优点。然而这类方法缺乏对数据整体分布的考虑, 导致生成的样本随机性较强、模型收敛缓慢。针对该问题, 将三元组中的正样本以其所属的类中心为轴, 旋转至锚点与该类中心连线的反向延长线上, 提出一种基于样本旋转的困难样本生成方法, 给出了一种新的损失函数, 构建了一种基于样本旋转的生成困难样本的深度度量学习模型(RHS-DML), 有效提升了模型的训练效率。在 Cars196, CUB200-2011 以及 Stanford Online Products 数据集上进行了图像检索的实验, 与代数计算方法中对称生成样本方法进行了比较, 结果表明, 本文提出算法的检索性能相较于对称样本生成方法, 在三个数据集上分别高出 2.4%, 0.7%, 1.4%。

**关键词:** 深度度量学习; 困难样本生成; 多类 N 元组损失; 代数计算

**中图分类号:** TP181 **文献标志码:** A **文章编号:** 0253-2395(2024)05-0973-09

## Sample Rotation-based Hard Sample-Generating Methods for Deep Metric Learning

ZHANG Ge<sup>1,2</sup>, YAN Jing<sup>1,2</sup>, WEI Wei<sup>1,2\*</sup>, LIANG Jiye<sup>1,2</sup>

(1. School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China;

2. Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan 030006, China)

**Abstract:** Existing deep metric learning methods guide efficient training of the model by constructing hard sample generation methods. The hard sample generation methods based on algebraic computation have the advantages of simplicity and efficiency. However, such methods lack consideration of the overall data distribution, resulting in strong randomness of the generated samples and slow convergence of the model. To address this problem, we propose a hard sample generation method based on sample rotation by rotating positive samples in a triad to the reverse extension of the line connecting the anchor point and the class center on the axis of the class to which they belong, and give a new loss function to construct a deep metric learning model (RHS-DML) for generating hard samples based on sample rotation, effectively improving the training efficiency of the model. Experiments on image retrieval were conducted on the Cars196, CUB200-2011, and Stanford Online Products datasets, and compared with the symmetric sample generation method in algebraic computing. The results showed that the retrieval performance of the algorithm proposed was 2.4%, 0.7%, and 1.4% higher than the symmetric sample generation cost method on the three datasets, respectively.

**Key words:** deep metric learning; hard sample generation; multi-class n-pair loss; algebraic calculations

收稿日期: 2023-03-06; 接受日期: 2023-06-05

基金项目: 国家自然科学基金(61976184)

作者简介: 张鸽(1998-), 女, 山西运城人, 硕士, 研究方向为机器学习。E-mail: 1628621898@qq.com

\* 通信作者: 魏巍(WEI Wei), E-mail: weiwei@sxu.edu.cn

引文格式: 张鸽, 闫京, 魏巍, 等. 基于样本旋转的生成困难样本的深度度量学习方法[J]. 山西大学学报(自然科学版), 2024, 47(5): 973-981. DOI: 10.13451/j.sxu.ns.2023106

### 0 引言

机器学习算法中往往使用欧氏距离度量样本间的相似性,然而欧氏距离度量平等地对待样本属性中的所有成分,忽略了特征重要性的差异。为解决该问题,Xing等<sup>[1]</sup>提出距离度量学习(Distance Metric Learning),通过将同类样本之间的距离拉近,异类样本之间的距离推远,学习更好的距离度量,提升模型性能。然而,随着信息技术的发展,数据变得越来越复杂,传统度量学习算法已经无法应对复杂的数据。随着深度学习的发展,深度度量学习应运而生。凭借深度神经网络强大的非线性拟合能力,深度度量学习在图像检索<sup>[2]</sup>、人脸识别<sup>[3-4]</sup>、行人重识别<sup>[5-7]</sup>等领域得到了广泛的应用。

现有深度度量学习方法普遍利用神经网络将数据映射到一个嵌入空间,在该嵌入空间中,同类样本之间的距离更近,异类样本之间的距离更远。以孪生网络为例,如图1所示,Hu等<sup>[8]</sup>在训练集中使用随机采样方法随机选择样本构造Batch,Batch中的样本两两组合形成样本对,样本对中的每个样本通过孪生网络得到嵌入表示,依据得到的嵌入表示,利用损失函数计算样本对之间的相似性,从而拉近正样本对之间的距离,推远负样本对之间的距离。

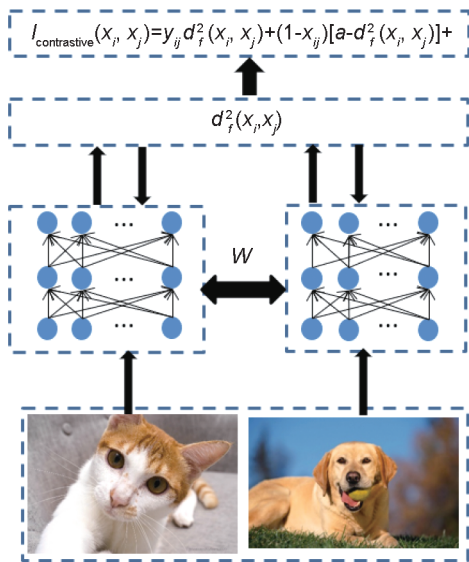


图1 深度度量学习的流程图

Fig. 1 The flowchart of deep metric learning

然而现有深度度量学习的训练过程存在以下三个问题<sup>[9]</sup>:(1)若训练集中有 $n$ 个样本,将会产生 $n^2$ 个样本对,若训练样本很多,将会导致模型收敛缓慢;(2)如果Batch中多数为损失较小的容易样本对,少数为损失较大的困难样本对,所得到的模型的判别能力弱;(3)若训练集中仅包含容易样本对,将会导致损失函数无梯度,则不利于模型收敛;因而困难样本挖掘方法的研究具有一定的理论价值。

困难样本挖掘方法有一些主流的工作,Ge<sup>[10]</sup>引入层次树的思想,首先将训练集中的每个类作为叶子节点,并将相似度高的类合并为父节点,最终基于树结构在相似度高的类中随机选择样本构造Batch。Xu等<sup>[11]</sup>构建一种非对称度量学习框架,通过设计两路数据流交错对齐结构,在相同采样规模下建立了更丰富的数据关系,从而提升了模型的泛化能力。由于现有困难样本挖掘方法的计算成本较高,Suh等<sup>[12]</sup>依据类到样本之间的距离在训练集中挖掘一些难负类,并从所选的类中选择样本构造Batch,以提高模型的训练效率。然而,现有困难样本挖掘方法往往只关注数据集中部分损失较大的困难样本对,而忽视数据集中损失较小的容易样本对,导致模型无法准确刻画嵌入空间的全局结构。

为解决上述问题,研究人员利用生成对抗网络<sup>[13-14]</sup>、自编码器<sup>[15-16]</sup>以及代数计算<sup>[17-18]</sup>生成困难样本。为解决负样本中困难样本数量较少的问题,Duan等<sup>[13]</sup>提出深度对抗度量学习(Deep Adversarial Metric Learning, DAML)框架,基于对抗训练方式利用容易负样本生成困难负样本,该方法有效地提升了算法性能。在基于三元组的深度度量学习方法中,现有困难样本挖掘方法大多挖掘对于当前网络困难样本的三元组,而非在全局优化网络方面真正重要的三元组,为解决该问题,Zhao等<sup>[14]</sup>提出用于困难三元组生成的对抗性网络以提升模型性能。由于现有困难样本挖掘方法仅利用训练集的子集,不足以全面地表征嵌入空间的全局几何,Zheng等<sup>[15]</sup>提出难度感知的深度度量学习框架(Hardness-aware Deep Metric Learning, HD-ML),该框架对嵌入进行线性插值,自适应地

调节生成样本的困难程度,并为再循环训练生成相应标签,从而充分利用所有样本的隐藏信息训练模型。

上述方法将被忽视的损失较小的容易样本合成为困难样本,从而得到容易样本携带的隐藏信息,利用携带的隐藏信息训练模型。这些方法有效地提升了模型性能,但需要一个额外的子网络生成困难样本,使得模型尺寸变大、超参数变多,最终导致模型训练缓慢。因而引入了使用代数计算生成困难样本的方法<sup>[17-18]</sup>, Ko等<sup>[18]</sup>受到 Query Expansion 和 Dataset Augmentation 的启发,提出了嵌入扩张的方法,该方法通过样本组合生成包含增强信息的样本,并进行困难样本挖掘,以使用信息量最大的嵌入表示训练模型。在对称生成样本<sup>[17]</sup>方法中,以坐标原点与其中每一个正样本的连线为对称轴,锚点与正样本按照该对称轴分别生成对称样本(如图2所示),该方法不需要额外的子网络,且没有超参数。然而在生成样本时,缺乏对数据整体分布的考虑,导致生成样本随机性较强、模型收敛缓慢。

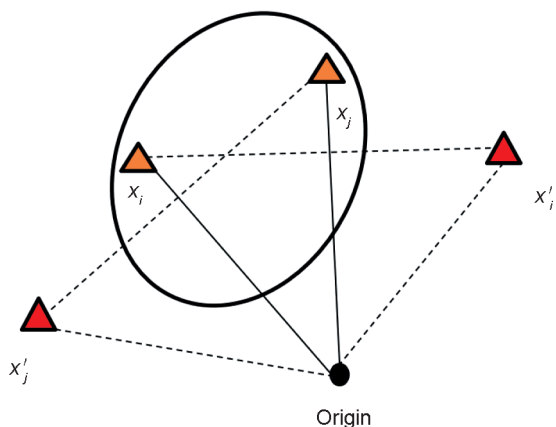


图2 对称生成困难样本示意图

Fig. 2 Schematic diagram of symmetric generation of hard samples

为解决上述问题,本文提出一种基于样本旋转的困难样本生成方法。结合该方法,给出了一种新的损失函数,构建了一种基于样本旋转的生成困难样本的深度度量学习模型(Deep Metric Learning model for generating Hard samples based on Sample Rotation, RHS-DML)。本文的主要贡献如下:

(1)通过将三元组中的正样本以其所属的

类中心为轴旋转至锚点与该类中心连线的反向延长线上,提出了一种新的困难样本生成方法,结合类中心约束,保证了生成的困难样本与同类样本之间的一致性。

(2)提出了基于样本旋转的生成困难样本的深度度量学习模型(RHS-DML),有效地提升了模型的训练效率。

(3)通过在 Cars196、CUB200-2011 以及 Stanford Online Products 三个数据集上进行图像检索的实验,证明了相较于对称样本方法,本文算法性能在三个数据集上分别高出 2.4%, 0.7%, 1.4%。

本文其余部分的内容如下,第1节介绍了相关工作的背景,第2节讲述了基于样本旋转的生成困难样本的深度度量学习模型,第3节给出了所提方法在图像检索任务的实验结果,最后,在第4节中描述了结论。

## 1 相关工作

### 1.1 深度度量学习

深度度量学习算法凭借高效的性能受到了人们的广泛关注。Kaya等<sup>[19]</sup>将深度度量学习模型划分为三个部分,分别是采样策略、损失函数以及网络结构。其中,损失函数与网络结构通常联合使用,Chopra等<sup>[20]</sup>提出对比损失的概念,对比损失首先通过孪生网络将数据映射到嵌入空间,在该嵌入空间中,使用欧氏距离拉近同类样本之间的距离,推远异类样本之间的距离。由于对比损失没有考虑同类样本与异类样本之间的相对关系,Schroff等<sup>[3]</sup>提出三元组损失,该损失通过选择一个锚点,使得包含该样本的负样本对之间的距离能够比包含该样本的正样本对之间的距离大于一个间隔,但该方法仅考虑一个负类,未考虑其他负类。因而,Sohn<sup>[21]</sup>提出多类N元组损失,在多类N元组损失中,选取N个类,每类选择两个样本,在其中一个类中的样本作为锚点及其正样本时,其他N-1类的正样本作为该类中锚点的负样本,因此多类N元组损失是将锚点与N-1个负样本进行配对,从而实现同类样本对的相似度高于所有异类样本对的相似度。Song等<sup>[22]</sup>提出提升结构化损失,他们针对三元组损失未曾充分利用Batch中所有样本对的问

题,提出提升结构化损失的概念,该损失对于Batch中每个正样本对中的每个样本,分别计算距离它们最近的负样本,选择其中距离最近的负样本计算其损失。而Wen等<sup>[4]</sup>提出中心损失的概念,这种损失通过将样本特征、样本到类中心的距离以及Softmax损失一同作为损失函数。随后,Mishchuk等<sup>[23]</sup>提出了成对聚类损失的概念,该损失是所有正样本到聚类中心的距离加上一个间隔之和小于异类样本到此聚类中心的距离。同年,代理损失被提出<sup>[24]</sup>,它解决了三元组损失采样困难的问题,将原始样本用代理点来近似,因而,约束同类样本对和异类样本对的距离就转化为约束锚点与同类样本对应代理点和锚点与异类样本对应代理点的距离。

### 1.2 困难样本挖掘

Ge<sup>[10]</sup>引入层次树的思想,首先将训练集中的每个类作为叶子节点,并将相似度高的类合并为父节点,最终基于树结构在相似度高的类中随机选择样本构造Batch。由于现有深度度量学习对所有的数据学习单个度量,导致模型过拟合及泛化能力弱,Sanakoyeu等<sup>[25]</sup>受到分治法的启发,将数据聚类,在每个类中学习一个度量构造Batch。现有的困难样本挖掘方法需要复杂的计算成本,导致现有深度度量学习方法需要在训练速度与预测精度之间进行权衡,Suh等<sup>[12]</sup>依据类到样本之间的距离从训练集中选择一些难负类,而后在所选的类中搜索样本构造Batch。Gajić等<sup>[26]</sup>对谱哈希方法进行改进,使用改进之后的谱哈希方法对数据划分,之后提出负样本集(Bag of Negatives, BON)策略构造Batch。Xu等<sup>[11]</sup>构建一种非对称度量学习框架,通过设计两路数据流交错对齐结构,在同等采样规模下建立了更丰富的数据关系,从而提升了模型的泛化能力。Wang等<sup>[27]</sup>认为无论采样方案有多复杂,困难样本挖掘方法在本质上都受到Batch大小的限制。因此提出跨批处理内存(Cross Batch Memory, XBM)结构,跨Batch构建样本对。

### 1.3 困难样本生成

然而,现有困难样本挖掘方法通常只关注数据集中部分损失较大的困难样本对,而忽视数据集中损失较小的容易样本对,导致模型无法准确刻画嵌入空间的全局结构。因而引入了

困难样本生成方法,现有的困难样本生成方法分为三类,使用生成对抗网络生成困难样本,使用自编码器生成困难样本以及使用代数计算生成困难样本。

为解决负样本中困难样本数量较少的问题,Duan等<sup>[13]</sup>提出深度对抗度量学习(DAML)框架,利用对抗训练的方式将容易的负样本生成为困难负样本,该方法有效提升了算法性能。在基于三元组的深度度量学习方法中,现有困难样本挖掘方法大多挖掘对于当前网络困难的三元组,而不是在全局优化网络方面真正重要的三元组的问题。针对该问题,Zhao等<sup>[14]</sup>提出用于困难三元组生成的对抗性网络以提升算法性能。Wang等<sup>[28]</sup>认为改进度量形式的方法无法在测试集上生成可靠的度量,提出一个自适应间隔深度对抗性度量学习框架(AM-DAML),利用大量常见的容易负样本生成潜在的困难负样本,并应用它们来促进稳健的度量学习,为防止模型在训练过程中出现过拟合或欠拟合的问题,提出一种自适应间隔损失,在负样本(包括对抗性负样本和原始负样本)和正样本之间保留了一个灵活的间隔。Zhu等<sup>[29]</sup>提出一种稳健样本生成方法来构造信息三元组。所提出的困难样本生成方法是一个两阶段合成框架,通过有效的正、负样本生成器分别在两个阶段产生困难样本。第一阶段用分片线性操作拉近锚点与正样本对,并通过巧妙设计条件生成对抗网络提高生成困难样本的质量,以降低模型塌缩的风险。第二阶段利用一个自适应反向度量约束来生成最终的困难样本。

Lin等<sup>[16]</sup>为解决由于不区分类内差异而使得模型过拟合的问题,提出深度变分度量学习(Deep Variational Metric Learning, DVML)框架,明确对类内方差进行建模,并拆分出类内不变性,通过学习到的类内方差分布,同时生成困难样本,以提高模型的鲁棒性。由于现有困难样本挖掘方法仅利用训练集子集,不足以全面地表征嵌入空间的全局几何,Zheng等<sup>[15]</sup>提出难度感知的深度度量学习框架(HDML),该框架对嵌入进行线性差值,自适应调节生成样本的困难程度,并为再循环训练生成相应的标签,从而充分利用所有样本的隐藏信息训练模型。

Ko等<sup>[18]</sup>为解决使用生成对抗网络或自编码器导致模型尺寸变大、超参数变多、训练速度变慢等问题,提出嵌入扩张的方法,该方法通过样本组合生成包含增强信息的合成样本,并进行困难样本挖掘,以使用信息量最大的嵌入表示训练模型。实验证明,该方法效果明显好于使用其他困难样本生成方法。同年,Gu等<sup>[17]</sup>提出对称生成困难样本的方法来解决上述生成对抗网络或自编码器所存在的问题,该方法将样本对中每个样本与坐标轴之间的连线作为对称轴对称生成样本,之后进行困难样本挖掘,其实验结果可以看出,对称生成困难样本方法的性能较优。然而,在生成样本时缺乏对数据整体分布的考虑,导致生成的样本随机性较强、模型收敛缓慢。

## 2 基本原理

本节首先介绍了符号定义以及本文使用的深度度量学习损失函数,之后提出了一种新的困难样本生成方法,即基于样本旋转的困难样本生成方法。其次,在提出了基于样本旋转的生成困难样本的深度度量学习模型(RHS-DML)后,讲述了模型训练过程。

假设 $I$ 是数据空间, $\mathcal{X}$ 是 $d$ 维嵌入空间。定义一个映射 $f:I \rightarrow \mathcal{X}$ ,该映射为从数据空间 $I$ 到嵌入空间 $\mathcal{X}$ 的映射,其参数由深度神经网络决定。对数据集 $I$ 映射到 $d$ 维嵌入空间 $\mathcal{X}$ 的特征样本 $X = \{x_1, x_2, \dots, x_n\}$ , $n$ 为样本个数。假设样本 $x_i$ 的类标签记为 $l_i$ , $l_i \in \{1, 2, \dots, m\}$ ,其中 $m$ 为类的个数。类中心的集合为 $C = \{c_1, c_2, \dots, c_m\}$ 。

本文所使用的损失函数为多类 $N$ 元组损失(Multi-class  $N$ -pair),具体描述如下:

$$L_{N\text{-pair}} = \frac{1}{|P|} \sum_{(x_i, x_j) \in P} \log \left( 1 + \sum_{x_k: l_k \neq l_i} \exp(S_{x_i, x_k} - S_{x_i, x_j}) \right), \quad (1)$$

其中 $P$ 表示正样本对集合, $|\cdot|$ 表示集合内包含的样本对的数量, $S_{x_i, x_j} = f_{x_i}^T f_{x_j}$ 表示样本 $x_i$ 与 $x_j$ 在嵌入空间中的相似度。

### 2.1 基于样本旋转的困难样本生成方法

对于任意给定样本 $x_i$ ,将其设置为锚点。选择与 $x_i$ 来自同一类的样本 $x_j$ 作为正样本构建正样本对 $(x_i, x_j)$ 。样本 $x_i$ 与 $x_j$ 的类标签相同,

即 $l_i = l_j$ ,记样本 $x_i$ 所属类的类中心为 $c_i$ ,则样本 $x_j$ 所属类的类中心也为 $c_i$ 。为充分考虑样本所在类的整体分布情况,以样本所属类的类中心为约束,将正样本 $x_j$ 以类中心 $c_i$ 为轴进行旋转;此外,为生成在类中心约束下最困难的正样本,将正样本 $x_j$ 旋转至锚点 $x_i$ 与类中心 $c_i$ 连线反向延长线上,生成困难样本 $x_j'$ 。此时初始正样本 $x_j$ 与生成样本 $x_j'$ 到类中心 $c_i$ 的距离相等,同时,样本 $x_j'$ 到锚点 $x_i$ 的距离最远,以此保证该样本为最困难的样本。如图3所示,椭圆形表示类的整体分布,不同形状表示为不同类的样本,下图中三角形与正方形即为两个类的样本,图中红色形状为生成的困难样本,在椭圆中心的橙色三角形表示样本 $x_i$ 的类中心 $c_i$ ,红色三角形表示生成的困难样本 $x_j'$ 。

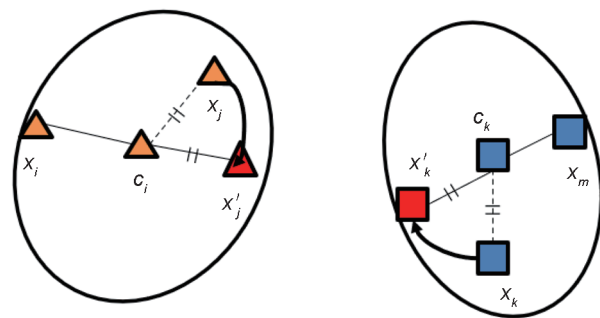


图3 基于样本旋转的困难样本生成方法示意图

Fig. 3 Schematic diagram of hard sample generation method based on sample rotation

利用向量的相关知识,通过下列方法可得到生成的困难正样本:

样本在通过神经网络得到嵌入向量后,可知锚点 $x_i$ ,原始正样本 $x_j$ 以及该类的类中心 $c_i$ 。假设锚点 $x_i$ 到生成的困难正样本 $x_j'$ 之间的向量为 $r_{x_i, x_j'}$ ,其单位向量为 $u_{x_i, x_j'}$ ,锚点 $x_i$ 到该类的类中心 $c_i$ 之间的向量为 $r_{x_i, c_i}$ ,类中心 $c_i$ 到原始正样本 $x_j$ 之间的向量为 $r_{c_i, x_j}$ ,类中心 $c_i$ 到生成的困难正样本 $x_j'$ 之间的向量为 $r_{c_i, x_j'}$ 。

$$u_{x_i, x_j'} = \frac{c_i - x_i}{\|c_i - x_i\|}, \quad (2)$$

$$r_{x_i, c_i} = c_i - x_i, \quad (3)$$

$$\|r_{c_i, x_j}\| = \|r_{c_i, x_j'}\| = \|x_j - c_i\|, \quad (4)$$

$$r_{x_i, x_j'} = r_{x_i, c_i} + r_{c_i, x_j'} = u_{x_i, x_j'} * (\|r_{x_i, c_i}\| + \|r_{c_i, x_j'}\|), \quad (5)$$

$$x_j' = x_i + r_{x_i, x_j'}. \quad (6)$$

通过上述公式的计算,即可得到生成的困

难正样本  $x_j'$ 。

## 2.2 基于样本旋转的生成困难样本的深度度量学习模型(RHS-DML)

为了更好地利用生成的困难正样本,在生成困难正样本之后,进行困难样本挖掘,选择损失值大的样本对参与训练。本文的损失函数在原始的多类  $N$  元组损失 (Multi-class  $N$ -pair) 上进行了改进,提出了新的损失函数如下:

$$L_{N\text{-pair}}^{\text{RHS}} = \frac{1}{|P|} \sum_{(x_i, x_j) \in P} \log \left( 1 + \sum_{x_k: l_i \neq l_k} \exp \left( \max_{(x_p, x_n) \in \hat{N}_{l_i \neq l_k}} S_{x_p, x_n} - S_{x_i, x_j'} \right) \right), \quad (7)$$

其中  $P$  表示正样本对集合,  $\hat{N}_{l_i \neq l_k}$  表示不同类的负样本对集合,  $x_p$  与  $x_n$  分别表示为在生成正样本之后,两个类中的锚点与正样本进行两两组合,形成负样本对,在进行困难样本挖掘之后找到的最难的负样本对,  $|\cdot|$  表示集合内含有的样本对的数量,  $S_{x_p, x_n} = f_{x_p}^T f_{x_n}$  表示样本  $x_p$  与  $x_n$  在嵌入空间中的相似度。

本文提出的基于样本旋转的生成困难样本的深度度量学习模型图如下所示:

本文首先将训练集及类中心通过神经网络,得到样本集  $X$  和类中心集  $C$ 。之后随机选择  $N$  对正样本对(图4中,不同形状表示不同类),并基于类中心旋转生成困难样本的方法生成困难正样本(如图4中,所有红色形状均为生成的样本),而后进行困难样本挖掘,选择损失值大的样本对,使用提出的损失函数计算损失。

## 3 实验分析

本节首先介绍实验所用的数据集和实验设置,并将提出的方法与目前代表性的方法进行对比分析,然后进行了消融实验分析和收敛速度的分析。

### 3.1 实验数据集

本文采用的数据集为深度度量学习的标准数据集,分别为 Cars196<sup>[30]</sup>, CUB200-2011<sup>[31]</sup> 以及 Stanford Online Products<sup>[22]</sup> 数据集。

Cars196 数据集共有 16 185 张图片,分为 196 个类,实验时使用前 98 个类的 8 054 张图片用于训练,使用剩下 98 个类的 8 131 张图片用于测试。

CUB200-2011 数据集共有 11 788 张图片,分

为 200 个类,实验时使用前 100 个类的 5 864 张图片用于训练,使用剩下 100 个类的 5 924 张图片用于测试。

Stanford Online Products 数据集共有 120 053 张图片,分为 22 634 个类,实验时使用前 11 318 个类的 59 551 张图片用于训练,使用剩下 11 316 个类的 60 502 张图片用于测试。

### 3.2 实验设置

在实验中,利用 Python 编程语言和 TensorFlow 深度学习框架实现,在 Nvidia Geforce RTX 3090 24 G 服务器上运行。

使用 ImageNet 预训练的 GoogLeNet 和 Xavier 方法来随机初始化一个全连接层。此外,模型采用 Adam 优化器进行优化,其学习率设置为 0.000 1。所有图像均归一化为  $256 \times 256$ ,水平翻转并随机裁剪为  $227 \times 227$ ,所有特征向量的嵌入大小设置为 512 维。

### 3.3 实验结果

实验所采用的任务是深度度量学习中经典任务之一,图像检索任务。所采用的评价指标为召回率,即 R@K。在 Cars196 和 CUB200-2011 数据集上,所设置的  $K$  分别为 1、2、4、8,而在 Stanford Online Products 数据集上,所设置的  $K$  分别为 1、10、100。

本文将提出的基于样本旋转的生成困难样本的深度度量学习方法(RHS-DML)与代表性的方法进行了比较,分别是 Sohn<sup>[21]</sup> 提出的多类  $N$  元组损失 (Multi-class  $N$ -pair), Duan 等<sup>[13]</sup> 提出的 DAML, Gu 等<sup>[17]</sup> 提出的 Symmetrical Synthesis (Symm+ $N$ -pair), Ko 等<sup>[18]</sup> 提出的 Embedding Expansion (EE+ $N$ -pair), Song 等<sup>[32]</sup> 提出的 Facility Location (FacilityLoc) 以及 Chen 等<sup>[33]</sup> 提出的深度连通图度量学习 (Deep Consistent Graph Metric Learning, CGML)。其中 DAML 属于深度度量学习使用生成对抗网络生成困难样本, Symm+ $N$ -pair 和 EE+ $N$ -pair 属于深度度量学习使用代数计算生成困难样本的方法。表1—表3分别是在 Cars196, CUB200-2011 以及 Stanford Online Products 三个深度度量学习标准数据集上所提出的模型(RHS-DML)与对比的方法进行比较的结果(表中将召回率最高的数据加粗表示)。由表1—表3可知,在三个数据集

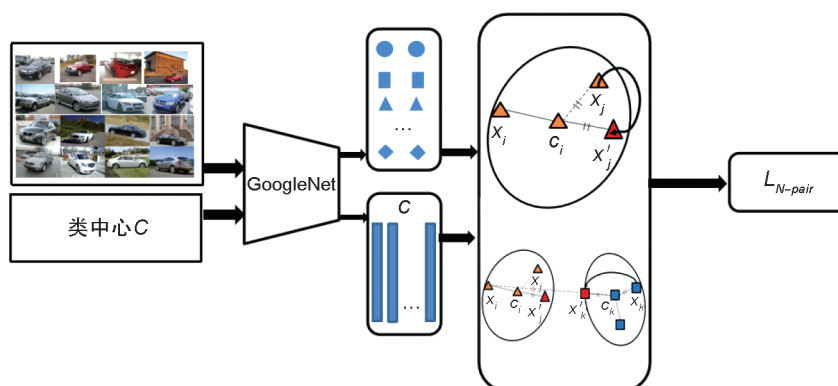


图4 基于样本旋转的生成困难样本的深度度量学习框架

Fig. 4 The deep metric learning framework based on rotation to generate hard samples

上提出的模型(RHS-DML)都具有较为明显的优势。

表1 在Cars196数据集上进行图像检索任务比较的实验结果(%)

Table 1 Experimental results (%) for image retrieval task comparison on Cars196 dataset

模型	R@1	R@2	R@4	R@8
Multi-class $N$ -pair(2016)	68.9	78.9	85.8	90.9
DAML(2018)	75.1	83.8	89.7	93.5
Symm+ $N$ -pair(2020)	76.5	84.3	90.4	94.1
EE+ $N$ -pair(2020)	72.5	81.1	87.6	92.5
FacilityLoc(2017)	58.1	70.6	80.3	87.8
CGML(2021)	75.8	84.4	90.5	94.4
RHS-DML	78.9	86.3	91.9	97.5

表2 在CUB200-2011数据集上进行图像检索任务比较的实验结果(%)

Table 2 Experimental results (%) for image retrieval task comparison on CUB200-2011 dataset

模型	R@1	R@2	R@4	R@8
Multi-class $N$ -pair(2016)	51.9	64.3	74.9	83.2
DAML(2018)	52.7	65.4	75.5	84.3
Symm+ $N$ -pair(2020)	55.9	67.6	78.3	86.2
EE+ $N$ -pair(2020)	55.2	67.4	77.7	86.4
FacilityLoc(2017)	48.2	61.4	71.8	81.9
CGML(2021)	52.1	64.2	75.4	84.5
RHS-DML	56.6	68.3	79.8	87.6

### 3.4 消融实验

表4为在Cars196数据集上进行的消融实验结果,提出的方法使用的损失函数为多类 $N$ 元组损失(Multi-class  $N$ -pair),在得到类中心后,锚点按照类中心将原始正样本旋转到锚点的反向延长线上,因此在进行消融实验时,所采用的对比

方法是多类 $N$ 元组损失(Multi-class  $N$ -pair)以及按照坐标原点,锚点将原始正样本旋转到锚点反向延长线上的方法,即RHS-DML-O。

表3 在Stanford Online Products数据集上进行图像检索任务比较的实验结果(%)

Table 3 Experimental results (%) for image retrieval task comparison on Stanford Online Products dataset

模型	R@1	R@10	R@100
Multi-class $N$ -pair(2016)	66.4	82.9	92.1
DAML(2018)	68.4	83.5	92.3
Symm+ $N$ -pair(2020)	73.2	86.7	94.8
EE+ $N$ -pair(2020)	73.5	87.5	94.4
FacilityLoc(2017)	67.0	83.7	93.2
CGML(2021)	68.4	84.3	93.2
RHS-DML	74.6	88.6	95.7

表4 在Cars196数据集上RHS-DML与RHS-DML-O与 $N$ -pair损失进行图像检索任务比较的实验结果(%)

Table 4 Experimental results (%) comparing RHS-DML with RHS-DML-O with  $N$ -pair loss for image retrieval on the Cars196 dataset

	R@1	R@2	R@4	R@8
Multi-class $N$ -pair(2016)	68.9	78.9	85.8	90.9
RHS-DML-O	76.2	84.2	90.3	92.6
RHS-DML	78.9	86.3	91.9	97.5

由表4可知,按照坐标原点旋转的方法(Deep Metric Learning model for generating Hard Samples based on Origin Rotation, RHS-DML-O)相较于未旋转的Multi-class  $N$ -pair方法效果更好,说明样本旋转在困难样本生成方法中起到了积极作用,而提出的基于样本旋转的困难样本生成方法(RHS-DML)比按照坐标原点旋转的方法(RHS-DML-O)效果好,说明按

照类中心旋转对于模型训练起到了积极作用。

图5展示了提出的方法与对称生成方法在Cars196数据集上的损失,从图中可以看出提出的方法在1200步左右就已经变得很平缓,而对称生成方法在3400步开始才开始收敛,实验结果说明,提出的方法具有更快的收敛速度。

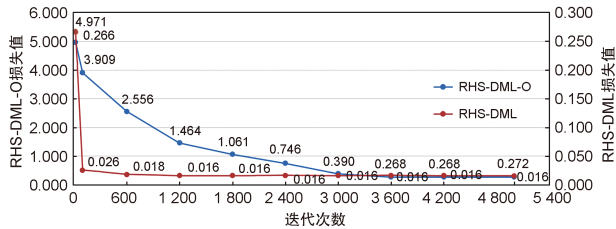


图5 在Cars196数据集上RHS-DML与对称生成方法的损失进行的比较

Fig. 5 Comparison of RHS-DML on the Cars196 dataset with the losses of the symmetric synthesis method

## 4 结论

本文通过借鉴代数计算的思想,提出了基于样本旋转的困难样本生成方法,设计了一种新的损失函数,构建了基于样本旋转的生成困难样本的深度度量学习模型(RHS-DML),通过在三个数据集上进行图像检索任务的实验,验证了提出的方法性能优于代表性的深度度量学习的方法,缓解了基于代数计算的困难样本生成的模型收敛速度慢的问题。

### 参考文献:

- [1] XING E P, NG A Y, JORDAN M I, *et al.* Distance Metric Learning, with Application to Clustering with Side-information[C]//Proceedings of the 15th International Conference on Neural Information Processing Systems. New York: ACM, 2002: 521-528. DOI: 10.5555/2968618.2968683.
- [2] GORDO A, ALMAZÁN J, REVAUD J, *et al.* Deep Image Retrieval: Learning Global Representations for Image Search[M]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 241-257. DOI: 10.1007/978-3-319-46466-4\_15.
- [3] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a Unified Embedding for Face Recognition and Clustering[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2015: 815-823. DOI: 10.1109/cvpr.2015.7298682.
- [4] WEN Y D, ZHANG K P, LI Z F, *et al.* A Discriminative Feature Learning Approach for Deep Face Recognition [M]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 499-515. DOI: 10.1007/978-3-319-46478-7\_31.
- [5] ZHAO L M, LI X, ZHUANG Y T, *et al.* Deeply-learned Part-aligned Representations for Person re-identification [C]//2017 IEEE International Conference on Computer Vision (ICCV). New York: IEEE, 2017: 3219-3228. DOI: 10.1109/iccv.2017.349.
- [6] ZHAO H Y, TIAN M Q, SUN S Y, *et al.* Spindle Net: Person re-identification with Human Body Region Guided Feature Decomposition and Fusion[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2017: 1077-1085. DOI: 10.1109/CVPR.2017.103.
- [7] 叶钰, 王正, 梁超, 等. 多源数据行人重识别研究综述[J]. 自动化学报, 2020, 46(9): 1869-1884. DOI: 10.16383/j.aas.c190278.
- [8] YE Y, WANG Z, LIANG C, *et al.* A Survey on Multi-source Person re-identification[J]. *Acta Autom Sin*, 2020, 46(9): 1869-1884. DOI: 10.16383/j.aas.c190278.
- [9] HU J L, LU J W, TAN Y P. Discriminative Deep Metric Learning for Face Verification in the Wild[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2014: 1875-1882. DOI: 10.1109/CVPR.2014.242.
- [10] 刘冰, 李瑞麟, 封举富. 深度度量学习综述[J]. 智能系统学报, 2019, 14(6): 1064-1072. DOI: 10.11992/tis.201906045.
- [11] LIU B, LI R L, FENG J F. A Brief Introduction to Deep Metric Learning[J]. *CAAI Trans Intell Syst*, 2019, 14(6): 1064-1072. DOI: 10.11992/tis.201906045.
- [12] GE W F. Deep Metric Learning With Hierarchical Triplet Loss[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer, 2018: 269-285. DOI: 10.1007/978-3-030-01231-1\_17.
- [13] XU X Y, YANG Y H, DENG C, *et al.* Deep Asymmetric Metric Learning via Rich Relationship Mining[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2019: 4076-4085. DOI: 10.1109/CVPR.2019.00420.
- [14] SUH Y, HAN B, KIM W, *et al.* Stochastic Class-based Hard Example Mining for Deep Metric Learning[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2019: 7251-7259. DOI: 10.1109/CVPR.2019.00742.
- [15] DUAN Y Q, ZHENG W Z, LIN X D, *et al.* Deep Adversarial Metric Learning[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 2780-2789. DOI: 10.1109/CVPR.2018.00294.

- [14] ZHAO YIRU, JIN ZHONGMING, QI GUOJUN, *et al.* An Adversarial Approach to Hard Triplet Generation [C]//Proceedings of the European Conference on Computer Vision. Cham: Springer, 2018: 508–524. DOI: 10.1007/978-3-030-01240-3\_31.
- [15] ZHENG W Z, LU J W, ZHOU J. Hardness-aware Deep Metric Learning[C]//IEEE Transactions on Pattern Analysis and Machine Intelligence. New York: IEEE, 2020: 72–81. DOI: 10.1109/TPAMI.2020.2980231.
- [16] LIN XUDONG, DUAN YUEQI, DONG QIYUAN, *et al.* Deep Variational Metric Learning[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer, 2018: 714–729. DOI: 10.1007/978-3-030-01267-0\_42.
- [17] GU G, KO B. Symmetrical Synthesis for Deep Metric Learning[J]. *Proc AAAI Conf Artif Intell*, 2020, **34**(7): 10853–10860. DOI: 10.1609/aaai.v34i07.6716.
- [18] KO B, GU G. Embedding Expansion: Augmentation in Embedding Space for Deep Metric Learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2020: 7253–7262. DOI: 10.1109/CVPR42600.2020.00728.
- [19] KAYA M, BILGE H Ş. Deep Metric Learning: a Survey[J]. *Symmetry*, 2019, **11**(9): 1066. DOI: 10.3390/sym11091066.
- [20] CHOPRA S, HADSELL R, LECUN Y. Learning a Similarity Metric Discriminatively, with Application to Face Verification[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). New York: IEEE, 2005: 539–546. DOI: 10.1109/CVPR.2005.202.
- [21] SOHN K. Improved Deep Metric Learning with Multi-class N-pair Loss Objective[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. New York: ACM, 2016: 1857–1865. DOI: 10.5555/3157096.3157304.
- [22] SONG H O, XIANG Y, JEGELKA S, *et al.* Deep Metric Learning via Lifted Structured Feature Embedding [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 4004–4012. DOI: 10.1109/CVPR.2016.434.
- [23] MISHCHUK A, MISHKIN D, RADENOVIĆ F, *et al.* Working Hard to Know Your Neighbor's Margins: Local Descriptor Learning Loss[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. California, USA: Curran Associates Inc, 2017: 4829–4840.
- [24] MOVSHOVITZ-ATTIAS Y, TOSHEV A, LEUNG T K, *et al.* No Fuss Distance Metric Learning Using Proxies [C]//2017 IEEE International Conference on Computer Vision (ICCV). New York: IEEE, 2017: 360–368. DOI: 10.1109/ICCV.2017.47.
- [25] SANAKOYEU A, TSCHERNEZKI V, BÜCHLER U, *et al.* Divide and Conquer the Embedding Space for Metric Learning[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2019: 471–480. DOI: 10.1109/CVPR.2019.00056.
- [26] GAJIĆ B, AMATO A, GATTA C. Fast Hard Negative Mining for Deep Metric Learning[J]. *Pattern Recognit*, 2021, **112**: 107795. DOI: 10.1016/j.patcog.2020.107795.
- [27] WANG X, ZHANG H Z, HUANG W L, *et al.* Cross-batch Memory for Embedding Learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2020: 6387–6396. DOI: 10.1109/CVPR42600.2020.00642.
- [28] WANG Z Y, WANG Y G, DONG B, *et al.* Adaptive Margin Based Deep Adversarial Metric Learning[C]//2020 IEEE 6th Intl Conference on Big Data Security on Cloud (Big-DataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS). New York: IEEE, 2020: 100–108. DOI: 10.1109/BigDataSecurity-HPSC-IDS49724.2020.00028.
- [29] ZHU C, HU Z, DONG H H, *et al.* Construct Informative Triplet with Two-stage Hard-sample Generation[J]. *Neurocomputing*, 2022, **498**(C): 59–74. DOI: 10.1016/j.neucom.2022.05.032.
- [30] KRAUSE J, STARK M, JIA D, *et al.* 3D Object Representations for Fine-grained Categorization[C]//2013 IEEE International Conference on Computer Vision Workshops. New York: IEEE, 2014: 554–561. DOI: 10.1109/ICCVW.2013.77.
- [31] WAH CATHERINE, BRANSON STEVE, WELINDER PETER, *et al.* The Caltech-UCSD Birds-200-2011 Dataset [R]. CNS-TR-2010-001, Pasadena: California Institute of Technology, 2010.
- [32] SONG H O, JEGELKA S, RATHOD V, *et al.* Deep Metric Learning via Facility Location[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2017: 2206–2214. DOI: 10.1109/CVPR.2017.237.
- [33] CHEN B H, LI P Y, YAN Z Y, *et al.* Deep Metric Learning with Graph Consistency[J]. *Proc AAAI Conf Artif Intell*, 2021, **35**(2): 982–990. DOI: 10.1609/aaai.v35i2.16182.