

基于多尺度三元组损失的层级图像检索算法

杨孟渭, 张索非, 吴晓富, 周全

(南京邮电大学, 江苏 南京 210003)

摘要: 现有大多数度量学习方法的目的是扩大类间样本的距离, 同时减小类内样本的距离, 在多层语义场景中, 此类度量学习方法是太不适用的。例如, 金丝猴属于猴子, 与猫不相似, 但从哺乳动物角度看, 金丝猴和猫又属于同一类。因此, 多层语义场景下存在的主要问题表现为: 在较小的语义尺度上的负对可以在较大语义尺度上的正对, 因此拉开较小语义尺度上的负对距离会损害较大语义尺度上的类结构, 反之亦然。为此, 该文提出了一种新的多尺度三元组损失 (Multi-Scale Triplet Loss, MSTL) 来解决层级图像检索问题。为了使得每个类别的嵌入空间更加紧凑, 只选取每个语义尺度标签都不相同的类别作为负样本, 同时提出了更适用于细粒度语义尺度的一种新的代理损失。最后, 通过在已有模型上赋予 MSTL 适合的权重, 以进一步提升模型的鲁棒性与检索性能。在三个 DyML 数据集上进行的大量实验表明, 该方法优于现有流行方法。

关键词: 度量学习; 多语义尺度; 层级图像检索; 对比学习; 深度学习

中图分类号: TP31

文献标识码: A

文章编号: 1673-629X(2025)04-0080-06

doi: 10. 20165/j. cnki. ISSN1673-629X. 2024. 0380

Hierarchical Image Retrieval with Multi-scale Triplet Loss

YANG Meng-wei, ZHANG Suo-fei, WU Xiao-fu, ZHOU Quan

(Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Most existing metric learning methods aim to expand the distance of inter-class samples while reducing the distance of intra-class samples, and such metric learning methods are less applicable at multiple semantic scales. For example, golden monkeys belong to monkeys, which are not similar to cats, but from the perspective of mammals, golden monkeys and cats belong to the same class. Therefore, the main problem existing in multiple semantic scales manifests itself in the fact that a negative pair at a finer semantic scale can be a positive pair at a coarser semantic scale, and thus distancing negative pairs at finer semantic scales compromises the class structure at coarser semantic scales, and vice versa. For this reason, we propose a new Multi-Scale Triplet Loss (MSTL) to solve the hierarchical image retrieval problem. In order to make the embedding space of each class more compact, only the classes with different labels at each semantic scale are selected as negative samples, and a new proxy loss that is more applicable to fine semantic scales is proposed. Finally, the robustness and retrieval performance of the model is further improved by assigning suitable weights to MSTL on the existing model. Extensive experiments on three DyML datasets show that the proposed method outperforms existing popular methods.

Key words: metric learning; multiple semantic scales; hierarchical image retrieval; contrastive learning; deep learning

0 引言

深度度量学习 (DML) 旨在学习一种由神经网络表示的函数, 用于将输入图像映射到特征嵌入空间, 在特征嵌入空间中, 语义相似的图像相互靠近, 语义不同的图像相互远离。传统的深度度量学习的目标是尽量缩小同类样本的嵌入特征距离, 同时增大不同类样本的嵌入特征距离。在多语义层次场景下, 直接使用传统单语义尺度度量学习会导致检测冲突。如图

1 所示, 一对图像在细粒度语义尺度可能被视为不同, 但在粗粒度语义尺度则可以被视为相似。图 1 中不同的形状表示不同的类别, (a) 表示在较低的语义尺度下, 不同的类别之间距离变大; (b) 表示在较高的语义尺度下, 假设这三个类别属于同一大类, 那么这些样本在这一尺度下距离变小; (c) 表示在较高的语义尺度下, 假设这三个类别属于同一大类, 最糟糕的情况是这三个类别在较小语义尺度上都不相同, 就会导致较小

收稿日期: 2024-08-07

修回日期: 2024-12-11

基金项目: 国家自然科学基金面上项目 (61876093)

作者简介: 杨孟渭 (1999-), 女, 硕士, 研究方向为计算机视觉、图像处理; 通信作者: 张索非 (1982-), 男, 博士, 讲师, 研究方向为计算机视觉。

语义尺度的检索结果也较低。Sun 等人^[1]提出了动态范围的概念,以克服上述冲突,并赋予单一深度度量在多个语义尺度上的灵活性。

为了更直观地展示多个语义尺度之间的关系,以图中较小尺度的“金丝猴”为例,它既属于“猴子”,也属于“哺乳动物”。在这种情况下,“哺乳动物”是最大的语义尺度。虽然金丝猴和白面猴在外形上并不相似,但在“猴子”等更大的尺度上,它们是相似的。所以,同一对象的相似性度量可能会因为语义尺度不同而不同。

图像检索是深度度量学习的典型应用场景之一,通常采用两种度量,即距离度量学习^[2]和排序度量学习^[3-4]。距离度量主要通过定义适当的距离度量函数来测量样本之间的相似性,如欧氏距离、余弦相似度等。设计有效度量学习的损失函数是关键挑战之一。Balntas 等人^[5]提出了困难负样本挖掘方法,利用元组

中最难挖掘的负样本来计算损失,从而自适应地关注难以挖掘的样本。此外,近年来还出现了一些变体,包括 Semi-hard triplet loss^[6]、N-pair loss^[7]和 Multi-Simi loss^[8]。Sun 等人^[1]提出了跨尺度学习(Cross-Scale Learning, CSL),这是一种优化的三元组损失方法,并将其扩展到分层设置中。不过,这种方法在嵌入空间中对正样本起拉近作用,未考虑负样本的影响。具体来说,如果细粒度上的负对是粗粒度上的正对,那么在细粒度上拉近这对样本,就会损害粗粒度上的类结构。Smooth-AP^[3]是一种基于平均精度(AP)的损失,通常用于单一语义尺度的图像检索。它表明,三元组损失是为了优化距离度量学习而不是排序度量学习,而且在使用排序度量学习进行评估时,三元组损失是次优的。Ramzi 等人^[9]提出了一种基于新的 H-AP 指标的多语义尺度检索模型,该指标侧重于对整体数据中的图像进行排序。

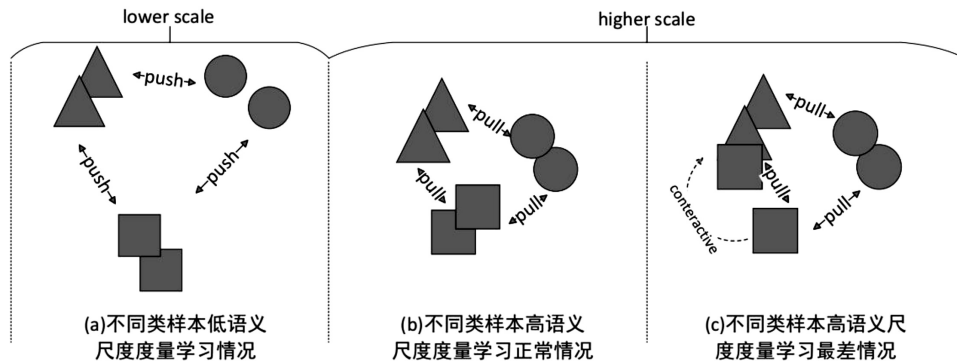


图1 依次进行度量学习的情况

在层级图像检索领域,Zheng 等人^[10]提出了 CLCD,旨在通过学习的多概念嵌入提取近似正交的图像。Zhang 等人^[11]提出了一种相似性聚合方法(MSA),该方法采用预训练的微调方案来高效地学习每个语义层的相似性关系。由于动态度量学习需要大量样本,这两种方法往往比 HAPPIER^[9]需要更多的训练时间。

为了解决上述问题,该文提出了多尺度三元组损失(Multi-Scale Triplet Loss, MSTL),通过优化三元组损失有效改善了各语义尺度的检索结果。首先,将传统的用于单一尺度的三元组损失扩展到多语义尺度。同时,在每个语义尺度上,选取相似图像为正样本,选取每个语义尺度类别都不相似的为负样本。然后,为了使同类样本与不同类样本更容易区分,在训练过程中引入了细粒度代理。最后,采用距离度量学习和排序度量学习的加权和衡量图像之间的相似性,为模型提供客观的指导,并更加注重捕捉图像在整个数据集的位置关系,从而反映数据的优先级。在三个动态度量学习数据集上进行了大量实验:实验结果表明,在相同训练历时下,该文提出的 MSTL 与 HAPPIER^[9]结

合取得了比 HAPPIER^[9]更好的性能。

1 相关工作

1.1 深度度量学习

深度度量学习(DML)是一种利用深度学习模型学习样本间相似度量度的方法。DML的关键在于设计适当的损失函数,引导深度神经网络学习样本之间的相似性关系。常用的损失函数有基于对的损失函数和基于代理的损失函数。

文献[6-7,11-13]中提出了多种基于对的损失。这些方法的重点是优化两幅图像之间的相似度指标,旨在提高正相似度,同时降低负相似度,以便更准确地捕捉图像之间的微妙关系。对于一对语义内容相似但视角或光照条件不同的图像,基于对的方法有助于区分这些细微差别,提高检索准确率。后来,一些实验^[3,6,12]也证明了挖掘有意义的数据对十分重要。

此外,还有许多研究侧重于基于代理的损失^[14-16]。这些方法通过为每个类别分配一个可学习的代理来优化样本之间的相似性,从而避免了复杂的挖掘过程。对于人脸识别任务来说,可以为每个人脸

类别学习一个代理向量,代表该类别的特征空间,捕捉每个类别中的细粒度结构,从而提高模型的泛化能力。代理学习还可应用于动态度量学习,捕捉不同类别之间的层次结构信息,从而提高相似度估计的准确性。

1.2 动态度量学习

如图 2 所示,传统的度量学习任务通常侧重于单



图 2 相似性度量说明

目前,有两种流行的动态度量学习方法。一种是距离度量学习,旨在将基于代理的三元组损失扩展到分层设置。另一种是采用分层 AP 训练方案的排名度量学习^[9],其方法是利用分层标签信息优化排名列表中两张检索图像的相对顺序。

一些研究考虑从特征嵌入的角度优化多语义尺度模型。Zheng 等人^[10]提出的 CLCD,通过学习每幅图像对应于给定语义级别的多个概念嵌入,从正图像中提取知识来学习图像概念嵌入。Zhang 等人^[11]通过划分嵌入层级,明确分离了每个语义层级的优化,以缓解跨层级冲突,减少层级间干扰,实现更好的整体优化。

1.3 三元组损失及其变体

Weinberg 和 Saul^[17]提出了大边际近邻损失作为三元组损失的原型。FaceNet^[6]正式定义了广泛使用的基于边界的损失及其变体。Balntas 等人^[5]提出通过锚点交换对三元组进行困难负样本挖掘^[5],以确保使用三元组中最难的负对来计算损失。Kihyuk Sohn^[7]为深度度量学习提出了一种可扩展的新目标,即 N-pair loss,它通过在每次更新时联合推远多个负样本,显著改善了三元组损失。Wu 等人^[12]提供了一种简单的基于边距的损失(Margin-based loss),能够在嵌入空间中均衡采样,它放宽了传统对比损失的不必要限制,并具有三元组损失的灵活性。

2 层级图像检索方法

考虑 L 个语义尺度下的层级检索问题,每个语义尺度的标签被表示为一个标签集 $\text{Label} = \{l_1, l_2, \dots, l_L\}$ 。给定一个训练集 $X = \{x_1, x_2, \dots, x_N\}$ 。 x_i^j 代表语义尺度为 j 时的 x_i 。

2.1 三元组损失回顾

设 $f(x_i)$ 是数据点 $x_i \in R^N$ 的嵌入特征,其中 $f: R^N \rightarrow R^D$ 是参数为 θ 的可微分深度网络,通常 $f(x_i)$ 的模

一语义尺度,即图像被视为相似(来自同一类别)或不相似(来自不同类别)。然而,当使用来自多个语义尺度的图像进行测试时,这种单一语义级别的学习方法可能表现不佳。

被归一化成单位长度。设 $D_{i,j}$ 为图片 x_i 和 x_j 在嵌入空间中的距离,即 $D_{i,j} = \|f_\theta(x_i) - f_\theta(x_j)\|^2$ 。FaceNet^[18]提出的三元组损失为:

$$\text{Loss}^{\text{triplet}}(\theta) = \sum_{a,p,n}^{y_a=y_p \neq y_n} \max(0, D_{a,p} - D_{a,n} + m) \quad (1)$$

当给定一个锚样本 x_a 时,公式 1 试图缩小同一类的锚样本 x_a 和正样本 x_p 在嵌入空间中的距离,同时使得不同类的锚样本 x_a 和负样本 x_n 在嵌入空间中的距离至少近 m 。但是随着训练数据集规模的快速增长,三元组数量会呈指数增长,因此训练所有可能的三元组损失对是不切实际的。

2.2 基于 Margin 的损失

困难负样本挖掘^[5]总是在高方差区域提供样本,导致噪声梯度,最终无法有效地将两个样本分开,从而导致模型坍塌。随机取样很容易集中在低方差区域,从而使损失非常小。Semi-hard negative mining^[5]通常能提供更快的收敛速度。对于层级图像检索,距离加权采样^[12]提供了广泛的示例,因此能在控制方差的同时稳定地生成信息丰富的示例。通过计算样本间的距离得到成对距离分布遵循的特定关系。

$$q(d) \propto q^{n-2} \left[1 - \frac{1}{4}d^2\right]^{\frac{n-3}{2}} \quad (2)$$

给定一个锚点 a ,距离加权采样法对负对 (a, n^*) 进行采样,采样率为:

$$\text{Pr}(n^* = n | a) \propto \min(\lambda, q^{-1}(D_{a,n})) \quad (3)$$

2.2.1 单语义尺度

Wu 等人^[12]提出了一种新的目标函数,基于边距的损失,其中新的距离加权采样可提高多个损失函数的性能。Margin-based loss 为:

$$\text{Loss}^{\text{margin}} = \sum_{i=1}^N ((\alpha + D_{i,p} - \beta)_+ + (\alpha + D_{i,n} - \beta)_+) \quad (4)$$

一般将锚样本与正样本之间的距离称为正样本对

距离;锚样本与负样本之间的距离为负样本对距离。公式中的参数 β 定义了正样本对距离和负样本对距离之间的界限;如果正样本对距离 $D_{i,p}$ 大于 β , 则损失增加;如果负样本对距离 $D_{i,n}$ 小于 β , 则损失增加。 α 控制样本的分离间隔。

2.2.2 多语义尺度

在每个语义尺度使用 Margin-based loss 进行优化,得到初始的多语义尺度三元组损失函数。

$$\text{Loss}^{\text{multi-margin}} = \sum_{j=1}^L \sum_{i=1}^N ((\alpha + D_{i,p}^j - \beta)_+ + (\alpha + D_{i,n}^j - \beta)_+) \quad (5)$$

然而,与 CSL^[2] 一样,这种方法也存在较高语义尺度影响较低语义尺度检索结果的问题。

2.3 多尺度三元组损失 (MSTL)

为缓解上述矛盾,提出 MSTL 方法,用于层级图像检索,如图 3 所示。该方法允许在高语义尺度上进行优化,而不会降低低语义尺度上的检索结果。图中,不同的形状代表最高语义尺度 L 下的不同类别。在任何语义尺度下,样本都只能远离每个语义尺度标签都不相同的样本。在为每个语义尺度选择三元组样本对时,只有最高语义尺度 L 不同的样本才会被选为负样本。所提方法在每个语义尺度上都能带来增益。

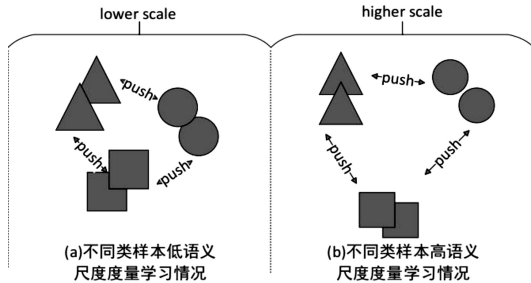


图 3 同时进行度量学习的情况

该文同时引入了一种高效的批处理方法,能从批处理的图像中依次选择有效图像作为负样本。设 $\{(x_1^l, +x_1^l), \dots, (x_N^l, +x_N^l)\}$ 是来自 N/m 个不同类的 N 个成对例子, l_j 表示语义尺度下的标签, $x_i^{l_j}$ 表示语义尺度 j 下的 x_i , $+x_i^{l_j}$ 表示 $x_i^{l_j}$ 的正样本, $x_i^{l_j}$ 表示最高语义尺度 L 下的 x_i 。如果 $\forall i(k) \leq m \times b, k(i) > m \times b$, 那么 $x_i^{l_j} \neq x_k^{l_j}$ 。设 $(x_i^{l_j}, +x_i^{l_j})$ 对应于负样本集 $\text{Negative} = \{+x_{cm+1}^{l_j}, +x_{cm+2}^{l_j}, \dots, +x_N^{l_j}\}$, 其中 $b \in [1, \frac{N}{m-1}]$, $c \in [0, \frac{N}{m}]$ 且 b 和 c 都为整数,每次从负样本集中选择一个作为负样本参与三元组损失。图 4 展示了这一批量生成过程。相应 MSTL 可表述为:

$$\text{Loss}^{\text{MSTL}} = \sum_{j=1}^L \sum_{i=1}^N ((\alpha + D_{i,p}^j - \beta)_+ + (\alpha + D_{i,n}^j - \beta)_+) \quad (6)$$

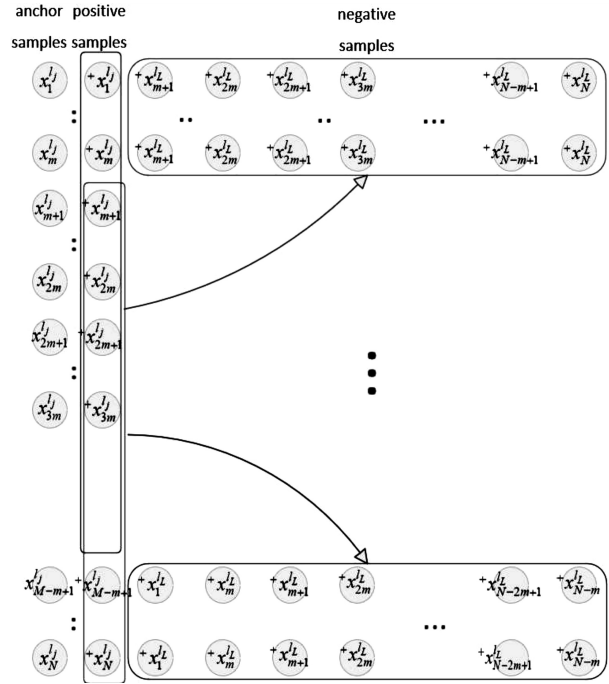


图 4 样本采样原理

2.4 模型

图 5 显示了所提方法的网络架构和训练损失。在 HAPPIER 中使用 MSTL 既能衡量图像之间的相似度,为模型提供客观指导,又能更注重捕捉图像在整个数据集集中的位置关系,进而反映出数据集的优先级。最终的方法取得了更好的检索结果。最终的损失表示为:

$$\text{Loss}^{\text{total}} = (1 - \lambda) \cdot \text{Loss}^{\text{MSTL}} + \lambda \cdot \text{Loss}^{\text{HAPPIER}} \quad (7)$$

图 5 的右上方以细粒度级别为每个类别分配了一个可学习的代理,而优化图像代理相似性则避免了复杂的挖掘过程。添加代理可使模型收敛得更快,并弥补了 MSTL 无法将相似样本拉近的缺陷。具体而言,将具有相同细粒度语义尺度的实例嵌入紧密分组。

$$\text{Loss}^{\text{proxy}} = -\log\left(\frac{\exp(\frac{v_y^T p_y}{\sigma})}{\sum_{p_z \in Z} \exp(\frac{v_y^T p_z}{\sigma})}\right) \quad (8)$$

其中, p_y 表示与嵌入 v_y 的细粒度类别相关联的标准化代理。 Z 表示代理集, σ 表示温度缩放参数。

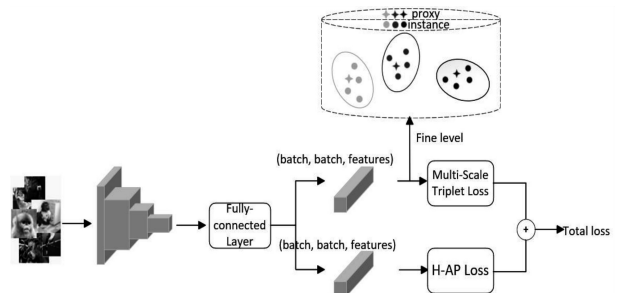


图 5 模型流程

3 实验

在本节中,将通过三个数据集对提出的 MSTL 方法进行评估。通过实验证明,将距离度量与排名度量相结合比单一度量学习能取得更好的效果。

3.1 数据集

按照文献[9]采用的三个 DyML 数据集进行实验评估,即 DyML-Vehicle、DyML-Animal 和 DyML-Product^[1]。每个数据集都有三个语义层次: $L = 3$, 包括 coarse、middle 和 fine。

这些层次构成一棵层级树,每幅图像有三个标签与之对应。根据文献[9],基础模型是在 DyML-Vehicle&Animal 上随机初始化的 ResNet34,并在 DyML-Product 的 ImageNet 上进行预训练。实验的硬件环境为 8 个 NVIDIA Tesla K80 GPU。除非另有说明,包括损失函数在内的其他实验设置与文献[1]相同。按照文献[2]的方法,采用了三个指标来评估检索性能,即平均精度(mAP)、Recall@K(R@K)和平均集合交集(ASI)。

3.2 结果

如表 1 所示,先将 MSTL 与 Margin-based loss^[14]、CSL^[1]和 HAPPIER^[9]进行比较,其中 Margin-based loss、CSL 和 MSTL 代表距离度量学习,HAPPIER 代表排序度量学习。表 1 显示了所提方法和最近四种图像检索方法在 DyML-Vehicle、DyML-Animal 和 DyML-Product 上的总体结果。与之前仅在细粒度上进行训

练的方法不同,所有这些方法都用于层级检索。可以看到,在 DyML-Animal 上,除了 mAP 比 HAPPIER 低 2.17 个百分点外,其他五个指标都明显优于 CSL,且 R@1、R@10、R@20 和 ASI 分别比 HAPPIER 高 3.7 百分点、2.97 百分点、2.7 百分点和 0.4 百分点;在 DyML-Product 上,除了 mAP 和 ASI 比 HAPPIER 低 1.64 百分点和 0.94 百分点外,R@1、R@10 和 R@20 分别比 HAPPIER 高 2.48 百分点、1.08 百分点和 1.39 百分点;在 DyML-Vehicle 上,所有 R@K 均优于 CSL,且 R@20 比 HAPPIER 高 0.3 百分点。总的来说,MSTL 在 DyML-Animal 和 DyML-Product 上获得了更好的结果,这说明距离度量学习在多语义尺度领域的表现与排序度量学习类似,R@K 也获得了显著的改进。表 1 还比较了加权距离度量学习与排序度量学习的结果,其中距离度量学习为 MSTL,排序度量学习为 HAPPIER。在 DyML-Animal 中,最新方法获得了最优的 mAP、ASI 和次优的 R@K;在 DyML-Product 中,最新方法获得了最优的 mAP、ASI 和次优的 R@K;在 DyML-Vehicle 中,虽然 MSTL 没有达到接近 HAPPIER 的结果,但最新方法达到了最好的结果。这表明,将距离度量学习与排序度量学习相结合,既能衡量图像之间的相似性,为模型提供客观指导,又能更侧重于捕捉整个数据集中图像的位置关系,从而反映数据集的优先级。

表 1 实验结果比较

数据集	Method	mAP	R@1	R@10	R@20	ASI
DyML-Animal	Margin-based	39.11	65.2	87.34	91.78	47.65
	CSL	40.56	61.0	85.59	90.43	46.43
	HAPPIER	43.9	68.7	89.27	92.82	50.68
	MSTL	41.73	72.4	92.24	95.52	51.08
	MSTL+HAPPIER	44.92	70.18	91.63	94.26	51.83
DyML-Product	Margin-based	36.9	58.63	87.46	91.7	46.42
	CSL	31.1	53.14	82.7	89.3	40.86
	HAPPIER	38.83	64.12	89.49	93.47	48.33
	MSTL	37.19	66.6	90.57	94.86	47.39
	MSTL+HAPPIER	38.9	65.55	89.17	93.87	48.6
DyML-Vehicle	Margin-based	26.78	78.9	89.98	92.81	39.47
	CSL	30.32	87.82	95.15	96.79	43.91
	HAPPIER	37.16	89.54	96.34	97.08	51.09
	MSTL	28.07	87.71	96.01	97.38	41.17
	MSTL+HAPPIER	37.97	89.97	96.55	97.52	51.22

3.3 分析

在表 2 中评估了 λ 对 mAP、ASI 和 R@K 的影响。在实际应用中,并不是所有的 λ 都能带来更好的检索结果,因此在 DyML-Animal 数据集上为 MSTL 和 HAPPIER 分配了不同的权重,以获得最佳 λ ,这对其

他两个数据集仍然有效。

从表 2 中可以看出,当 $\lambda = 0.4$ 时,mAP 和 ASI 最大;当 $\lambda = 0$ 时,R@K 最大;当 $\lambda = 1$ 时,R@K 最小。当 λ 增大时,R@K 呈现出曲折下降的趋势,虽然在较小范围内 R@K 略有提升,但整体上随着 λ 的增大,R

@K 逐渐减小,因此当 $\lambda = 0.4$ 时, $R@K$ 处于中间位置值。

表2 λ 的选择实验

λ	mAP	R@1	R@10	R@20	ASI
0	41.73	72.4	92.24	95.52	51.08
0.1	42.43	71.11	91.6	94.9	51.21
0.2	43.0	70.7	91.98	94.69	51.8
0.3	43.87	70.65	91.24	94.77	51.73
0.4	44.92	70.18	91.63	94.26	51.83
0.5	44.9	69.62	90.9	93.66	51.42
0.6	44.9	68.8	89.59	93.21	51.16
0.7	43.55	69.96	91.45	94.07	50.2
0.8	44.52	69.55	91.63	94.12	50.12
0.9	44.0	69.23	90.1	93.67	50.17
1	43.9	68.7	89.27	92.82	50.68

最后,选择 $\lambda = 0.4$,此时得到综合最优的结果。

4 结束语

该文提出了一种用于分层图像检索的多尺度三元组损失(MSTL)。通过在每个语义尺度上拉近相同标签样本的距离,同时在最高语义尺度上拉开不同标注样本的距离,避免语义尺度之间可能存在的冲突。在三个 DyML 数据集下对该方法进行了评估,结果表明 MSTL 可以使距离度量学习的结果接近排序度量学习的结果。该文还提出对距离度量学习和排序度量学习进行了加权实验,结果表明最新方法取得了最佳效果。

参考文献:

- [1] SUN Y, ZHU Y, ZHANG Y, et al. Dynamic metric learning: towards a scalable metric space to accommodate multiple semantic scales [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Nashville: IEEE, 2021: 5393-5402.
- [2] YANG L. Distance metric learning: a comprehensive survey [D]. Michigan: Michigan State University, 2006.
- [3] BROWN A, XIE W, KALOGEITON V, et al. Smooth-AP: smoothing the path towards large-scale image retrieval [C]//Proceedings of the European conference on computer vision. Glasgow: Springer, 2020: 677-694.
- [4] OCHOA X, DUVAL E. Relevance ranking metrics for learning objects [J]. IEEE Transactions on Learning Technologies, 2008, 1(1): 34-48.
- [5] BALNTAS V, RIBA E, PONSA D, et al. Learning local feature descriptors with triplets and shallow convolutional neural networks [C]//Proceedings of the British machine vision conference. London: Springer, 2016.
- [6] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2015: 815-823.
- [7] SOHN K. Improved deep metric learning with multi-class n-pair loss objective [J]. Advances in Neural Information Processing Systems, 2016, 29: 1-9.
- [8] WANG X, HAN X, HUANG W, et al. Multi-similarity loss with general pair weighting for deep metric learning [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long Beach: IEEE, 2019: 5022-5030.
- [9] RAMZI E, AUDEBERT N, THOME N, et al. Hierarchical average precision training for pertinent image retrieval [C]//Proceedings of the European conference on computer vision. Glasgow: Springer, 2022: 250-266.
- [10] ZHENG W, HUANG Y, ZHANG B, et al. Dynamic metric learning with cross-level concept distillation [C]//Proceedings of the European conference on computer vision. Glasgow: Springer, 2022: 197-213.
- [11] ZHANG D, LI Y, ZHANG Z. Multi-scale similarity aggregation for dynamic metric learning [C]//Proceedings of the 31st ACM international conference on multimedia. Ottawa: ACM, 2023: 125-134.
- [12] WU C Y, MANMATHA R, SMOLA A J, et al. Sampling matters in deep embedding learning [C]//Proceedings of the IEEE international conference on computer vision. Venice: IEEE, 2017: 2840-2848.
- [13] HADSELL R, CHOPRA S, LECUN Y. Dimensionality reduction by learning an invariant mapping [C]//Proceedings of the 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR '06). New York: IEEE, 2006: 1735-1742.
- [14] DENG J, GUO J, XUE N, et al. ArcFace: additive angular margin loss for deep face recognition [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long Beach: IEEE, 2019: 4690-4699.
- [15] 常禾雨, 司念文, 屈丹, 等. 元学习框架下基于圆损失度量的行人重识别模型 [J]. 信息工程大学学报, 2022, 23(1): 93-102.
- [16] 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.
- [17] WEINBERGER K Q, SAUL L K. Distance metric learning for large margin nearest neighbor classification [J]. Journal of Machine Learning Research, 2009, 10(2): 1-20.
- [18] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person reidentification [J]. arXiv: 1703.07737, 2017.