软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2020,31(7):1933–1942 [doi: 10.13328/j.cnki.jos.005934]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

面向细粒度草图检索的对抗训练三元组网络*

陈 健[1], 白 琮[1], 马 青[1,2], 郝鹏翼[1], 陈胜勇[3]

[1](浙江工业大学 计算机科学与技术学院,浙江 杭州 310023)

[2](浙江工业大学 理学院,浙江 杭州 310023)

[3](天津理工大学 计算机科学与工程学院,天津 300384)

通讯作者: 白琮, E-mail: congbai@zjut.edu.cn

摘 要: 将草图作为检索示例用于图像检索称为基于草图的图像检索,简称草图检索.其中,细粒度检索问题或类内检索问题是2014年被研究者提出并快速成为广受关注的研究方向.目前研究者通常用三元组网络来解决类内检索问题,且取得了不错的效果.但是三元组网络的训练非常困难,很多情况下很难收敛甚至不收敛,且存在着容易过拟合的风险.借鉴循环生成对抗训练的思想,设计了 SketchCycleGAN 帮助提高三元组网络训练过程的效率,以对抗训练的方式使其参与到三元组网络的训练过程中,通过充分挖掘数据集自身信息的方式取代了利用其他数据集进行预训练的过程,在简化训练步骤的基础上取得了更好的检索性能.通过在常用的细粒度草图检索数据集上的一系列对比实验,证明了所提方法的有效性和优越性.

关键词: 基于草图的图像检索;细粒度检索;三元组网络;对抗训练

中图法分类号: TP391

中文引用格式: 陈健,白琮,马青,郝鹏翼,陈胜勇.面向细粒度草图检索的对抗训练三元组网络.软件学报,2020,31(7):1933–1942. http://www.jos.org.cn/1000-9825/5934.htm

英文引用格式: Chen J, Bai C, Ma Q, Hao PY, Chen SY. Adversarial training triplet network for fine-grained sketch based image retrieval. Ruan Jian Xue Bao/Journal of Software, 2020,31(7):1933–1942 (in Chinese). http://www.jos.org.cn/1000-9825/5934.htm

Adversarial Training Triplet Network for Fine-grained Sketch Based Image Retrieval

CHEN Jian[1], BAI Cong[1], MA Qing[1,2], HAO Peng-Yi[1], CHEN Sheng-Yong[3]

[1](School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China)

[2](College of Science, Zhejiang University of Technology, Hangzhou 310023, China)

[3](College of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China)

Abstract: Sketch based image retrieval means that the sketch is used as the query in the retrieval. Fine-grained image retrieval or intra-categoryretrieval was proposed in 2014 and attracted more attentions quickly. Triplet network is often used to do fine-grained retrieval and get promising performance. However, training triplet network is quite difficult, it is hard to converge and easy to over-fit in some situations. Inspired by the adversarial training, this study proposes SketchCycleGAN to improve the efficiency of the triplet network training process. In this proposal, pre-training the networks with other database is replaced by mining the information inside the database with the help of adversarial training. That could simplify the training procedure with better performance. This proposal could get better performance than other state-of-the-art methods in a series of experiments executed on widely used databases for fine-grained sketchbased retrieval.

* 基金项目: 国家重点研发计划(2018YFB1305200); 浙江省自然科学基金(LY18F020032, LY18F020034); 浙江省教育厅项目(Y201839922)

Foundation item: National Key R&D Program (2018YFB1305200); Natural Science Foundation of Zhejiang Province of China (LY18F020032, LY18F020034); Zhejiang Provincial Department of Education of China (Y201839922)

本文由“多媒体内容的多维度相似性计算与搜索”专题特约编辑蒋树强研究员、刘青山教授、孙立峰教授、李波教授推荐.

收稿时间: 2019-05-02; 修改时间: 2019-07-11; 采用时间: 2019-09-17; jos 在线出版时间: 2020-01-13

CNKI 网络优先出版: 2020-01-14 11:25:43, http://kns.cnki.net/kcms/detail/11.2560.TP.20200114.1125.017.html

Key words: sketch based image retrieval; fine-grained retrieval; triplet network; adversarial training

手绘草图是一种自然、直观的表达人们意图的方式,且随着智能便携设备的发展,手绘草图可以方便地从各种移动终端上获取.因此,手绘草图提供了一种可随时随地描述特定物体外观、结构和姿态的方式.以其作为检索示例的基于草图的图像检索(sketch based image retrieval,简称 SBIR)在文字或图像示例缺失或难以描述检索需求时可以提供额外的、补充的图像搜索方法.通常来讲,手绘草图具有很高的抽象性且只粗略描述了待检物体或场景的整体形状或显著形状,但是待检图像数据库往往又是真实的图片.两者之间的不同主要体现在以下 3 个方面[1]:(1) 视觉显示上的不同:手绘草图只有整体形状或显著形状,但是图片具有详细的色彩、纹理及形状特征描述;(2) 内容上的不同:手绘草图没有背景,但是图片一般包含背景.(3) 抽象程度的不同:手绘草图对图片的抽象存在着随机干扰、简化描述和不现实的失真等.上述不同使得基于草图的图像检索成为一个非常具有挑战性的课题.自 20 世纪 90 年代以来,基于草图的图像检索就开始进入研究者的视野,最早的工作在 1992 年被提出[2],并在类间图像检索方面取得了极大的发展[3],之后,越来越多的学者进入到了这个领域,提出了很多富有建设意义的方法[4,5].随着深度学习技术的发展,人们发现其不仅能应用到图像分类的任务上[6],在手绘草图检索上也有着很大的应用价值[7,8].随着基于草图的图像检索在电子商务和公共安全上新需求的不断显现及研究者对基于草图的图像检索的重新认识,人们发现以往的研究大都忽视了草图也具备描述细粒度差异的能力.因此,基于草图的细粒度图像检索(fine-grained sketch based image retrieval,简称 FG-SBIR)引起了研究者的关注.其目的是在特定类别物体范围内检索与草图示例最相似的物体图像[9].

基于草图的细粒度图像检索的概念在 2014 年由 Li 等人[10]首先提出,他们利用可变部件模型检测物体并表征物体以完成细粒度检索.目前该研究作为一个新兴的研究方向正处方兴未艾之际,最近几年的研究方法主要是利用深度学习来建立草图和图像之间的共同特征空间,进而完成相似性度量.Sangkloy 等人[11]首先将深度学习的方法引入 FG-SBIR.Li 等人[12−14]基于深度学习进行图像理解与图像检索,并对弱监督学习与哈希在其中的应用进行了研究.Yu 等人[15]利用 3 个 Sketch-a-Net[16]组建了一个三元组网络来完成 FG-SBIR.Song 等人[17]在三元组网络之外引入了属性排序.Huang 等人[18]提出用 Recurrent Neural Network 和 Convolutional Neural Network 来共同组建一个三元组网络的方法.Zhang 等人[19]将哈希表示集成到三元组网络中.Pang 等人[20]将三元组网络中锚样本由手绘草图替换为手绘草图与合成草图的合集.由以上分析可知,三元组网络由于对细节具有较好的区分能力而受到了基于草图的细粒度图像检索领域研究者的重视.但是,三元组网络具有训练难度大的特点,其存在收敛速度慢甚至不收敛的问题,并且在训练时容易在训练集上过拟合导致最终在测试时表现变差.

本文提出了一种基于对抗训练三元组网络的细粒度草图检索方法,将循环生成对抗网络(CycleGAN)的思想引入到三元组网络的训练过程中.通过训练循环生成对抗网络实现手绘草图与图像之间在一定程度上的互相转换,利用手绘草图生成图像,并将其作为三元组一份子输入到三元组网络中进行训练,使得网络的整个训练过程更为平滑并取得更好的效果.同时,这种方式省去了寻找其他额外的数据集对三元组网络进行预训练的过程,因此整个训练过程显得更为简单、直观,将此方法应用到其他场景时也能更为方便地实现.本文提出的方法具有以下特点.

(1) 提出了一种基于手绘草图的细粒度图像检索方法,通过引入生成对抗训练,充分挖掘数据集自身的信息来摆脱用其他数据集预训练的依赖,进而提升训练效率和效果,取得了更好的检索精度.

(2) 提出 SketchCycleGAN,实现手绘草图与真实图像之间的相互转换,将输入的手绘草图锚样本转换到真实图像所在的空间作为新的锚样本,拉近锚样本与正负样本之间的距离,使三元组网络的训练过程更为平滑.

(3) 实验结果表明,所提出的方法可以通过使用所提出的 SketchCycleGAN 减少训练过程复杂性的同时明显提升 FG-SBIR 的性能,相比其他方法存在较大优势.

1 生成式对抗网络与三元组网络概述

1.1 生成式对抗网络

生成式对抗网络(generative adversarial network,简称 GAN)[21]是一种生成式建模方法,利用生成器(generator)与判别器(discriminator)进行对抗来生成数据.生成器的目的是从训练数据样本中学习潜在的真实数据分布,并通过接受噪声等方式生成新的数据样本.判别器的目的是正确判别输入样本是来源于训练数据样本还是由生成器生成的.生成器与判别器通过对抗彼此不断进行优化,以提升各自的生成能力或判别能力,整个对抗式生成网络的学习就是寻找生成器与判别器间的纳什均衡.

GAN 把对抗的概念引入到了机器学习领域中,它可以自动学习真实样本的数据分布,并且可以有效地生成能够建立自然性解释的数据样本.GAN 作为一种生成式方法,往往由诸如深度神经网络等结构组成,这使其对生成样本的维度没有限制,极大程度上扩展了能够生成的数据样本的范围,这在需要生成高维数据的问题中显得尤为重要.此外,采用深度神经网络结构增加了模型设计时的自由度,它能够很便利地整合各类不同的损失函数,以应对不同的任务要求.在此基础上,GAN 还能自动学习潜在的损失函数,即其判别器可以自动学习出较好的判别方法,这在面对不同的问题时非常有用.

在 GAN 训练过程中,以生成器与判别器之间的对抗作为训练准则,利用反向传播的方式对模型进行训练.与传统的使用马尔可夫链方法进行训练的生成模型相比,生成式对抗网络无需进行各种近似推理,不用考虑复杂的变分下界,这不仅大大降低了训练的难度,在训练的效率上也有着非常大的改善.在生成数据样本的过程中,GAN 能够直接对新的数据样本进行采样及推断而无需考虑繁杂琐碎的采样序,这对于生成新数据样本的效率有很大的提升.由于 GAN 采用了对抗的方式进行训练,而不是对真实数据样本直接进行复制或者平均,从而在生成的数据样本的多样性上也有着很大优势.

循环生成式对抗网络(CycleGAN)[22]是用于图像上的一个 GAN 结构,它可以实现图像与图像之间的相互转换,与其他使用对抗式生成网络进行图像相互转换的模型如 pix2pix[23]相比,它没有训练数据样本必须完全匹配的限制,因此可以将其更广泛地使用到各类不同的任务当中.

CycleGAN 由两个镜像对称的 GAN 组成,即一共存在两个判别器和两个生成器.两个 GAN 各自拥有一个判别器,而两个生成器是共享的.这两个生成器分别代表两个映射函数 $G:X\to Y$ 与 $F:Y\to X$,判别器 D_X 用于区分 x 和 $F(y)$,判别器 D_Y 则用于区分 y 和 $G(x)$,其总体结构表现为一个环形网络.在训练过程中,CycleGAN 使用两部分损失函数,一部分用于使生成样本与目标样本保持一致分布,另一部分为循环一致性损失(cycle consistency loss)[22],用于确保 G 与 F 不产生矛盾.

对于生成器 $G:X\to Y$ 以及判别器 D_Y,其损失函数定义如下:

$$\mathcal{L}_{GAN}\left(G,D_Y,X,Y\right)=\mathbb{E}_{y\sim p_{data}(y)}\left[\log D_Y\left(y\right)\right]+\mathbb{E}_{x\sim p_{data}(x)}\left[\log\left(1-D_Y\left(F\left(x\right)\right)\right)\right] \tag{1}$$

类似地,对于映射关系:$Y\to X$ 和其判别器 D_X,其损失函数定义如下:

$$\mathcal{L}_{GAN}\left(F,D_X,Y,X\right)=\mathbb{E}_{x\sim p_{data}(x)}\left[\log D_X\left(x\right)\right]+\mathbb{E}_{y\sim p_{data}(y)}\left[\log\left(1-D_X\left(F\left(y\right)\right)\right)\right] \tag{2}$$

循环一致性损失使用了 L1 范数的形式,其定义为

$$\mathcal{L}_{cyc}\left(G,F\right)=\mathbb{E}_{x\sim p_{data}(x)}\left[F\left(G\left(x\right)\right)-x_1\right]+\mathbb{E}_{y\sim p_{data}(y)}\left[G\left(F\left(y\right)\right)-y_1\right] \tag{3}$$

即 CycleGAN 的最终损失函数为

$$\mathcal{L}\left(G,F,D_X,D_Y\right)=\mathcal{L}_{GAN}\left(G,D_Y,X,Y\right)+\mathcal{L}_{GAN}\left(F,D_X,Y,X\right)+\lambda_c\mathcal{L}_{cyc}\left(G,F\right) \tag{4}$$

最终优化的目标函数与原始的 GAN 很相似,其定义如下:

$$G^*,F^*=\arg\min_{G,F}\max_{D_X,D_Y}\mathcal{L}\left(G,F,D_X,D_Y\right) \tag{5}$$

利用这种结构,将手绘草图转换到真实图像所在的空间,再把转换的结果作为一个全新的锚样本,与旧的锚样本所对应的正负样本组成新的三元组,用到接下来的训练过程中.

1.2 三元组网络

三元组网络(triplet network)[24]是从孪生网络(siamese network)[25]延伸出来的一种结构,其所解决的问题也与孪生网络基本一致:即完成在数据样本的类别非常多或难以确定,且每个类别的训练样本数量又非常少的情况下的视觉学习任务.它相比单一的网络结构能够取得更好的效果,因此在人脸识别等领域有着广泛的应用[26].

与孪生网络不同的是,三元组网络的每组训练数据由 3 个样本组成:一个锚样本(anchor,表示为 x),一个与锚样本属同类的正样本(positive example,表示为 x^+),一个与锚样本属异类的负样本(negative example,表示为 x^-).三元组网络对 x^+与 x^-相对 x 的欧式距离进行编码,其行为可以表示为

$$TripletNet\left(x,x^-,x^+\right)=\begin{bmatrix}\left\|Net\left(x\right)-Net\left(x^-\right)\right\|_2\\ \left\|Net\left(x\right)-Net\left(x^+\right)\right\|_2\end{bmatrix}\in\mathbb{R}_+^2 \tag{6}$$

当采用相似的 CNN 结构时,三元组网络往往相比孪生网络表现得更好,主要在于三元组网络在对比损失函数(contrastive loss)[27]的基础上,构建了一个新的损失函数,对于样本相对类内及类间的各自距离间的差值添加了一个限制(margin),并取其值为 1.其损失函数定义为

$$Loss\left(d_+,d_-\right)=\left\|\left(d_+,d_--1\right)\right\|_2^2=const\cdot d_+^2 \tag{7}$$

其中,d_+与 d_-定义为

$$d_+=\frac{e^{\left\|Net(x)-Net\left(x^+\right)\right\|_2}}{e^{\left\|Net(x)-Net\left(x^+\right)\right\|_2}+e^{\left\|Net(x)-Net\left(x^-\right)\right\|_2}} \tag{8}$$

$$d_-=\frac{e^{\left\|Net(x)-Net\left(x^-\right)\right\|_2}}{e^{\left\|Net(x)-Net\left(x^+\right)\right\|_2}+e^{\left\|Net(x)-Net\left(x^-\right)\right\|_2}} \tag{9}$$

2 细粒度手绘草图图像检索

2.1 循环生成式对抗网络设计

预训练的过程对于神经网络的训练十分重要,它不仅影响着网络的收敛速度,某种程度上也决定了网络最后的效果和精度,不同的预训练过程可能会导致最终的结果相差巨大.CycleGAN 作为一种 GAN 结构,它不仅可以实现不同数据之间的转换,还能生成一些新的数据样本以供使用,这就启发了一种辅助训练的方式,即使用这些新生成的数据样本来辅助训练过程,不但免去了寻找其他额外数据集来进行预训练的过程,也免去了复杂且难以评估的预训练流程,简化了训练流程,使训练过程更为直观,并充分挖掘了数据集自身有效信息.

基于草图的图像检索的任务是通过手绘草图检索相对应的真实图像,即需要实现手绘草图与真实图像之间的相互匹配与距离评估.因此,通过使用图像与图像间的相互转换,将手绘草图在一定程度上转换得与真实图像很相似,有助于帮助后面进行手绘图像与真实图像的相互匹配.由于最终的任务是使用手绘草图进行检索而非真实图像的边缘图(edge map),而手绘草图是抽象和标志性的,无法保证手绘草图与期望检索出的真实图像完全匹配,具体的区别如图 1 所示.

Fig.1 The difference between sketch and edge map

图 1 手绘草图与边缘图的区别

因为 CycleGAN 可以实现图像间的转换且没有训练数据样本必须完全匹配的限制,因此本文基于 CycleGAN 的思想设计了一个 SketchCycleGAN 结构来实现手绘草图与待检索图像之间的相互转换,如图 2 所示.整个结构由两个完全镜像的 GAN 构成,两个生成器分别负责 *A* 生成 *B* 和 *B* 生成 *A* 的功能,两个判别器则各自判断是否是真实的 *A* 或 *B*,再以循环的方式将其组合起来.每个生成器拥有 3 个卷积层及对应的 3 个反卷积层,在中间有 4 个残差网络块来控制所需的参数数量,每个判别器由 5 个卷积层组成.这里与图像转换的任务不同,并不需要手绘草图完全转化为真实草图.事实上,完全匹配是不可能的,因而在生成器中仅使用了 4 个残差网络块来避免训练资源对生成器倾斜过多,以防产生浪费甚至导致参数过多而产生过拟合问题.

在训练时,SketchCycleGAN 同时接受手绘草图 *A* 以及该草图对应的真实图像 *B* 作为输入,将 *A* 和 *B* 及生成的假 *B* 与假 *A* 分别输入到对应的判别器中计算损失,与此同时,假 *B* 与假 *A* 分别输入到彼此的生成器中生成重构 *A* 和重构 *B*,再与原始的 *A* 和 *B* 进行比较计算循环一致性损失,整个训练表现出环的形态.同时,原始的 *A* 作为原本的锚样本,其生成的假 *B* 即为用作进一步组成三元组的新的锚样本.

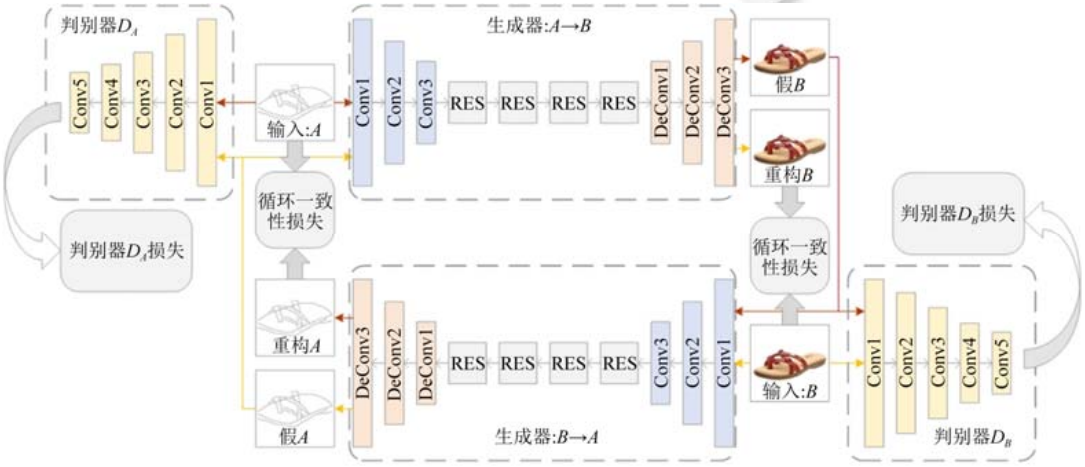


Fig.2 The architecture of SketchCycleGAN

图 2 SketchCycleGAN 网络结构

2.2 三元组网络设计

本文最终的任务是实现细粒度的手绘图像检索,由于训练数据样本数量很少类别却很多,单一的网络结构在这任务上无能为力.因此所提方法使用了 3 个相同的基于 Sketch-a-Net 的改进网络结构组成三元组网络来进行训练.由于 Sketch-a-Net 是用于手绘草图分类的网络结构,参照基准方法[15]去除最后 Sketch-a-Net 中用于分类的最后一层,在 fc7 后添加 L2 标准化层(L2 normalization layer)进一步对 fc7 的输出进行处理且作为最终特征的输出层.设定 fc7 的神经元数量为 256 个,即设定输出特征的维度为 256,并从图像中计算边缘图组成训练三元组数据样本.最终使用的三元组网络的网络结构如图 3 所示.

将三元组网络表示为 $f_\theta(\cdot)$,对于一个三元组 $t=(x,x^+,x^-)$,分别对应草图、相应的正确自然图像与错误自然图像,其损失函数定义为

$$L_\theta(t)=\max\left(0,\Delta+D\left(f_\theta(x),f_\theta\left(x^+\right)\right)-D\left(f_\theta(x),f_\theta\left(x^-\right)\right)\right) \tag{10}$$

其中,$D(x,y)$表示 x 与 y 之间的欧式距离(Euclidean distance),而Δ为锚样本相对正样本及负样本各自距离之间的差值的限制.最终的目标函数定义为

$$\min_\theta \sum_{t\in T} L_\theta(t)+\lambda_{Tri}R(\theta) \tag{11}$$

其中,T是所有的训练三元组数据样本,θ为整个三元组网络的所有参数,$R(\cdot)$是 L2 正则化项,λ_{Tri}是一个常数,用来控制正则化项的影响.

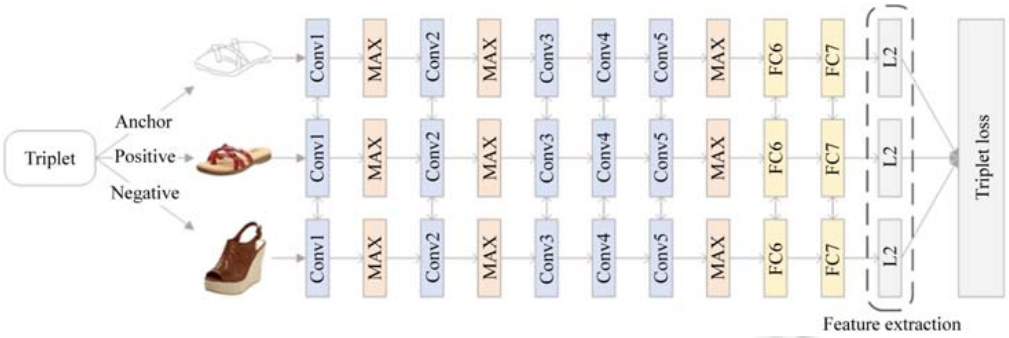


Fig.3 The architecture of triplet network

图 3 三元组网络结构

2.3 循环对抗生成训练三元组网络

三元组网络在细粒度的识别问题上有着很好的效果,但同时也存在收敛速度慢甚至不收敛的问题,因此,网络参数初始化的好坏对三元组网络最后的效果有很大的影响,尤其是当训练数据很少的时候,而本文任务正是处于这种情况.为了使网络的参数获得合适的初始化值,基准方法[15]通过另外寻找一个或多个其他比较大的数据集,例如 TU-Berlin Sketch 事先对网络进行预训练,之后再用实验的数据集对网络进行微调(fine-tune),这不仅使训练过程变得很复杂,需花费大量时间,并且对预训练数据集的选择及处理也会引申出更多的问题.

本文所提方法使用 SketchCycleGAN 生成数据来训练三元组网络,从而不需要另外的数据集对网络进行复杂的预训练.具体流程如图 4 所示,训练时锚样本与正样本作为输入进入到 SketchCycleGAN 中对其进行训练,同时,锚样本通过 SketchCycleGAN 生成新的数据样本作为新的锚样本与正样本及负样本组成三元组,对三元组网络进行初步的训练.由于新生成的锚样本会比原本的锚样本更接近正样本及负样本,这有助于三元组网络在训练前期快速地收敛提升训练的效率,最后,原始的三元组对三元组网络作进一步的训练微调,最终使网络收敛到效果最好.

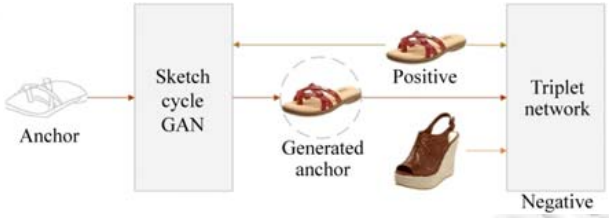


Fig.4 SketchCycleGAN for triplet network

图 4 SketchCycleGAN 训练三元组网络

2.4 算法流程

本文所提方法的整个算法流程可以分为训练与检索两个过程.

在训练过程中,首先将草图与所对应的自然图像成对地输入到 SketchCycleGAN 当中进行训练,与此同时,取所生成的自然图像作为新的锚样本,与对应的正负样本组合成三元组适当地训练三元组网络,之后直接用原始的草图作为锚样本,对三元组网络作进一步的微调训练,直到损失函数趋于稳定,即网络的参数收敛.

检索时,将检索的草图及检索的图像数据库都输入到已完成训练的三元组网络中,并从 L2 层提取出对应的特征作为检索所用特征,通过计算余弦距离(Cosine distance)对检索的结果进行排序,完成整个检索过程.

3 实验结果与分析

本文使用 TensorFlow[28]来实现图 2 所示的 SketchCycleGAN 结构及图 3 所示的三元组网络结构,并将它们

结合在一起,整个网络的参数在开始时都采用随机初始化的方式,即没有使用其他的数据集预先进行训练.训练时先使用SketchCycleGAN训练三元组网络,再单独使用三元组对三元组网络作进一步的微调.所有实验在配备有i7-8700K CPU,16G内存和GTX-1080显卡的工作站上进行.

验证实验在专用的细粒度手绘草图图像检索公开数据集:QMUL-Shoe和QMUL-Chair上进行.采用Top正确率(acc)作为检索的评价指标,其定义如下:

$$acc.@K=\frac{\sum_{q=1}^{Q}rel(q)}{Q} \tag{12}$$

其中,Q是查询数量,K是人为指定的范围,通常有K=1及K=10两种指标,本文也采用这两种指标,当检索的正确结果存在于返回结果的前K个时,$rel(q)$=1,否则$rel(q)$=0.

3.1 数据集

QMUL-Shoe[15]数据集是细粒度图像检索数据集,所有数据样本属于鞋子,由419对手绘草图及图像组成.所有图像大小皆为256×256像素.数据集中的304对数据样本作为训练数据集参与训练过程.剩下的115对数据样本作为测试数据集,仅在测试阶段使用.

QMUL-Chair[15]数据集是细粒度图像检索数据集,所有数据样本属于椅子一类,由297对手绘草图及图像组成.所有图像大小皆为256×256像素.数据集中的200对数据样本作为训练数据集参与训练过程.剩下的97对数据样本作为测试数据集,仅在测试阶段使用.

TU-Berlin Sketch[29]数据集是一个手绘草图的分类数据集,由20 000张256×256大小的手绘草图组成,并且均匀分布在250个类别当中.在实验中,该数据集仅仅作为设计对比实验时所使用的预训练数据集而不参与测试过程,以验证使用SketchCycleGAN辅助训练的优越性.

3.2 SketchCycleGAN有效性分析

为了证明在所提方法中SketchCycleGAN的有效性,我们设计了对照实验以进行比较,由于两个数据集十分相似,我们仅使用QMUL-Shoe来进行实验比较.第1个对比实验是一个单独的三元组网络,对其不进行任何的预训练,所有的训练参数值都采用随机初始化的方式.第2个对比实验采用先使用分类数据集预训练的方式,在三元组网络L2层后面添加一个分类层,然后使用分类数据集TU-Berlin Sketch进行预训练,由于TU-BerlinSketch拥有250个类别共20 000张手绘草图,对于实验数据集QMUL-Shoe而言过于巨大,为了防止预训练过多导致过拟合,仅训练50个周期,并确保分类准确率保持在0.5左右,然后去除分类层使用三元组进行训练.

对于SketchCycleGAN,设定其循环一致性损失的系数λ_C=10,对所有的生成器及判别器均使用Adam优化来进行优化,学习率皆设定为0.000 2;对于三元组网络,设定参数λ_{Tri}=0.000 1以及参数Δ=0.3,同样使用Adam优化器对其进行优化,设定其学习率为0.000 1,并将输出的特征维度设置为256维以保证与其他方法比较时的公平性,最终实验的详细结果见表1.

Table 1 Effectiveness of SketchCycleGAN on QMUL-Shoe

表1 SketchCycleGAN在QMUL-Shoe上的有效性

名称	acc.@1	acc.@10
Triplet network	0.313	0.817
TU-Berlin sketch pretraining+Triplet network	0.346	0.851
SketchCycleGAN+triplet network (Our proposal)	**0.427**	**0.913**

从实验结果可以看出,当不对三元组网络进行任何处理而直接采用随机的初始化值进行训练时,其展现的性能比较低.当使用TU-Berlin Sketch数据集先对网络进行预训练后,虽然增加了训练步骤的复杂性,但是最后检索的性能有了明显的提升.由此可以确定,在使用三元组训练前的网络初始化值的好坏对网络最终性能有着很大的影响.最后,本文所提方法在两项指标上皆取得了最高的分数,可以证明SketchCycleGAN发挥了很大的

作用,不仅省去了寻找其他数据集进行预训练的步骤,还提高了最终检索的性能.

3.3 QMUL-Shoe和QMUL-Chair结果分析

由于 QMUL-Shoe 数据集与 QMUL-Chair 数据集非常相似,因此实验中对这两个数据集的设置基本一致.实验中各种参数的具体设置与前面相同.在这两个数据集上,为了证明本文所提方法的优越性,我们将其性能与 BoW-HOG+rankSVM[30]、Dense-HOG+rankSVM[15]、ISN Deep+rankSVM[16]、3DS Deep+rankSVM[31]、TSN without data aug.[15]、TSN with data aug.[15]和 GDH@128-bit[19]进行了比较.上述各种方法大致介绍如下.

BoW-HOG+rankSVM[30]:提取 HOG 特征,用 rankSVM[32]模型进行排序检索.

Dense-HOG+rankSVM[15]:在密集网格上连接 HOG 特征,信息更为丰富,用 rankSVM 模型进行排序.

ISN Deep+rankSVM[16]:使用 Sketch-a-Net 的神经网络结构提取特征,用 rankSVM 模型进行排序.

3DS Deep+rankSVM[31]:使用 3DShapeCNN 的神经网络结构进行特征提取,用 rankSVM 模型进行排序.

TSN without data aug.[15]与 TSN with data aug.[15]:使用由 Sketch-a-Net 构成的基本三元组网络提取特征,计算欧式距离进行排序检索.前者不包含数据增强步骤,后者包含数据增强步骤.

GDH@128-bit[19]:使用生成式的域迁移方法,并加入了注意力模型,引入了哈希方法进行特征表示,比较哈希值的汉明距离(Hamming distance)进行排序.

详细结果见表 2 和表 3.

Table 2 Comparative results against baselines on QMUL-Shoe

表 2 与其他方法在 QMUL-Shoe 数据集上的比较结果

名称	acc.@1	acc.@10
BoW-HOG+rankSVM[30]	0.174	0.687
Dense-HOG+rankSVM[15]	0.244	0.652
ISN Deep+rankSVM[16]	0.200	0.626
3DS Deep+rankSVM[31]	0.052	0.217
TSN without data aug.[15]	0.373	0.861
TSN with data aug.[15]	0.391	0.878
GDH@128-bit[19]	0.357	0.843
Our proposal	**0.427**	**0.913**

通过表 2 所示结果可以发现,本文所提方法在 QMUL-Shoe 数据集上取得了很好的效果,与其他先进方法相比,在 acc.@1 和 acc.@10 两项指标上均取得了更好的性能且提升巨大.这是由于我们引入了 SketchCycleGAN 来辅助参与训练过程,使得最终三元组网络能够更好地收敛,更为精确地对原始图像进行特征表达,体现在检索时能够较为精准地评估图像彼此之间的距离与差异,使得最终的检索性能有了很大的提升.由于所提方法不需要额外的数据集参与训练,训练过程更为方便,整体上相较其他先进方法有着很大优势.

Table 3 Comparative results against baselines on QMUL-Chair

表 3 与其他方法在 QMUL-Chair 数据集上的比较结果

名称	acc.@1	acc.@10
BoW-HOG+rankSVM[30]	0.289	0.670
Dense-HOG+rankSVM[15]	0.526	0.938
ISN Deep+rankSVM[16]	0.474	0.825
3DS Deep+rankSVM[31]	0.061	0.268
TSN without data aug.[15]	0.644	0.956
TSN with data aug.[15]	0.691	0.979
GDH@128-bit[19]	0.671	**0.990**
Our proposal	**0.733**	0.984

通过表 3 所示结果可以发现,对于 QMUL-Chair 数据集,本文所提方法仍然有着不错的表现,在 acc.@1 上取得了最高的性能并且提升很大,在 acc.@10 上虽然处于第二的位置,但仅以非常小的数值差距低于 GDH@128-bit,且与其他方法相比仍然具有一定的优势.这主要是由于 QMUL-Chair 数据集相对而言比较简单,且在 acc.@10 这项指标上各类方法的数值都比较高,因此在该项数值上存在一定的瓶颈.

4 结束语

本文提出了一种基于对抗训练三元组网络的细粒度草图检索方法,提出了 SketchCycleGAN,以原始锚样本生成新的锚样本,与正负样本组成全新的三元组对三元组网络进行训练,使得三元组网络的训练过程变得更为平滑高效,最终取得了很好的检索效果.本文所提方法还在训练的步骤和复杂度上存在一定的优势.通过 SketchCycleGAN 充分挖掘了数据集自身的信息,从而做到了在不依赖外部数据集预训练或复杂预处理的情况下仍使三元组网络很好地收敛,使得整个训练过程显得更为简洁、直观.通过在训练方法上的比对实验及与最近几年所提方法的比对,证明了所提方法在细粒度草图检索任务上的有效性.

References:

[1] Li Y, Li W. A survey of sketch-based image retrieval. Machine Vision and Applications, 2018,29(7):1083–1100.

[2] Kato T, Takio K, Otsu N, *et al*. A sketch retrieval method for full color image database-query by visual example. In: Proc. of the 11th IAPR Int'l Conf. on Pattern Recognition. 1992. 530–533.

[3] Xu D, Alameda-pineda X, Song J, *et al*. Cross-paced representation learning with partial curricula for sketch-based image retrieval. IEEE Trans. on Image Processing, 2018,27(9):4410–4421.

[4] Li B, Liang S, Sun ZX. Sketch retrieval based on topological relations. Computer Science, 2005,(12):227–231 (in Chinese with English abstract).

[5] Liang S, Sun ZX. Small sample incremental biased learning algorithm for sketch retrieval. Ruan Jian Xue Bao/journal of Software, 2009,20(5):1301–1312 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/3274.htm [doi: 10.3724/SP.J.1001.2009.03274]

[6] Bai C, Huang L, Chen JN, Pan X, Chen SY. Optimization of deep convolutional neural network for large scale image classification. Ruan Jian Xue Bao/Journal of Software, 2018,29(4):1029-1038 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5404.htm [doi: 10.13328/j.cnki.jos.005404]

[7] Fan YC, Tan XH, Zhou MQ, Zheng X. A scale invariant local descriptor for sketch based 3D model retrieva. Chinese Journal of Computers, 2017,40(11):2448–2465 (in Chinese with English abstract).

[8] Yu MY, Wu H, Guo XY, Jia Q, Guo H. Sequential feature based sketch recognition. Computer Science, 2018,45(S2):198–202 (in Chinese with English abstract).

[9] Song J, Yu Q, Song YZ, *et al*. Deep spatial-semantic attention for fine-grained sketch-based image retrieval. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision (ICCV). IEEE, 2017. 5552–5561.

[10] Li Y, Hospedales T, Song YZ, *et al*. Fine-grained sketch-based image retrieval by matching deformable part models. In: Proc. of the British Machine Vision Conf. British Machine Vision Association. 2014. 115.1–115.12.

[11] Sangkloy P, Burnell N, Ham C, *et al*. The Sketchy database: Learning to retrieve badly drawn bunnies. ACM Trans. on Graphics (TOG), 2016,35(4):1–12.

[12] Li Z, Tang J. Weakly supervised deep metric learning for community-contributed image retrieval. IEEE Trans. on Multimedia, 2015,17(11):1989–1999.

[13] Li Z, Tang J, Mei T. Deep collaborative embedding for social image understanding. IEEE Trans. on Pattern analysis and Machine Intelligence, 2019,41(9):2070–2083.

[14] Tang J, Li Z. Weakly supervised multimodal hashing for scalable social image retrieval. IEEE Trans. on Circuits and Systems for Video Technology, 2017,28(10):2730–2741.

[15] Yu Q, Liu F, Song YZ, *et al*. Sketch me that shoe. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016. 799–807.

[16] Yu Q, Yang Y, Liu F, *et al*. Sketch-a-Net: A deep neural network that beats humans. Int'l Journal of Computer Vision, 2017,122(3):411–425.

[17] Song J, Song YZ, Xiang T, *et al*. Deep multi-task attribute-driven ranking for fine-grained sketch-based image retrieval. BMVC, 2016, 1–11.

[18] Huang F, Cheng Y, Jin C, *et al*. Deep multimodal embedding model for fine-grained sketch-based image retrieval. In: Proc. of the 40th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval-SIGIR 2017. New York: ACM Press, 2017. 929–932.

[19] Zhang J, Shen F, Liu L, *et al*. Generative domain-migration hashing for sketch-to-image retrieval. In: Ferrari V, Hebert M, Sminchisescu C, *et al*., eds. Proc. of the ECCV 2018. Cham: Springer Int'l Publishing, 2018. 304–321.

[20] B KP, Li D, Song J, *et al*. Deep factorised inverse-sketching. In: Proc. of the ECCV 2018. 2018. 37–54.

[21] Goodfellow I, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial nets. In: Advances in Neural Information Processing Systems. 2014. 2672–2680.

[22] Zhu JY, Park T, Isola P, *et al*. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2017. 2223–2232.

[23] Isola P, Zhu JY, Zhou T, *et al*. Image-to-image translation with conditional adversarial networks. In: Proc. of the IEEE Conf. on Computer Vision and PATTERN Recognition. 2017. 1125–1134.

[24] Hoffer E, Ailon N. Deep metric learning using triplet network. In: Proc. of the Int'l Workshop on Similarity-Based Pattern Recognition. Cham: Springer-Verlag, 2015. 84–92.

[25] Bromley J, Guyon I, LeCun Y, *et al*. Signature verification using a "siamese" time delay neural network. In: Advances in Neural Information Processing Systems. 1994. 737–744.

[26] Cheng D, Gong Y, Zhou S, *et al*. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 1335–1344.

[27] LeCun Y, Huang FJ. Loss functions for discriminative training of energy-based models. In: Proc. of the AIStats. 2005,6:34.

[28] Abadi M, Barham P, Chen J, *et al*. Tensorflow: A system for large-scale machine learning. In: Proc. of the 12th {USENIX} Symp. on Operating Systems Design and Implementation ({OSDI} 16). 2016. 265–283.

[29] Eitz M, Hays J, Alexa M. How do humans sketch objects. ACM Trans. on Graphics, 2012,31(4):44:1–44:10.

[30] Li Y, Hospedales TM, Song YZ, *et al*. Free-hand sketch recognition by multi-kernel feature learning. Computer Vision and Image Understanding, 2015,137:1–11.

[31] Wang F, Kang L, Li Y. Sketch-based 3D shape retrieval using convolutional neural networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 1875–1883.

[32] Joachims T. Optimizing search engines using clickthrough data. In: Proc. of the 8th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. ACM, 2002. 133–142.

附中文参考文献:

[4] 李彬,梁爽,孙正兴.基于空间关系的手绘草图检索.计算机科学,2005,(12):227–231.

[5] 梁爽,孙正兴.面向草图检索的小样本增量有偏学习算法.软件学报,2009,20(5):1301–1312. http://www.jos.org.cn/1000-9825/3274.htm [doi: 10.3724/SP.J.1001.2009.03274]

[6] 白琮,黄玲,陈佳楠,潘翔,陈胜勇.面向大规模图像分类的深度卷积神经网络优化.软件学报,2018,29(4):1029–1038. http://www.jos.org.cn/1000-9825/5404.htm [doi: 10.13328/j.cnki.jos.005404]

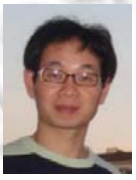
[7] 樊亚春,谭小慧,周明全,郑霞.基于局部多尺度的三维模型草图检索方法.计算机学报,2017,40(11):2448–2465.

[8] 于美玉,吴昊,郭晓燕,贾棋,郭禾.基于时序特征的草图识别方法.计算机科学,2018,45(S2):198–202.

陈健(1995－),男,学士,主要研究领域为基于内容的图像检索.

郝鹏翼(1986－),女,博士,讲师,CCF 专业会员,主要研究领域为机器学习,图像处理.

白琮(1981－),男,博士,副教授,博士生导师,CCF 专业会员,主要研究领域为计算机视觉,多媒体信息处理.



陈胜勇(1973－),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为计算机视觉.

马青(1982－),女,讲师,主要研究领域为图像检索.