计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190800145

基于注意力模型的手绘图像检索方法

李宗民[1] 李思远[1] 刘玉杰[1] 李 华[2]
1 中国石油大学(华东)计算机与通信工程学院 山东 青岛 266580
2 中国科学院计算技术研究所 北京 100190
(lizongmin@upc.edu.cn)

摘 要 针对手绘图像检索领域中手绘图像的特征稀疏、手绘本身易于形变等问题,文中提出了一种基于注意力模型的特征提取方法,通过精确提取手绘图像中的语义特征来获得高效准确的检索结果。首先使用卷积神经网络作为提取语义特征的基础框架;然后在有监督训练的过程中引入了注意力模型机制,通过在卷积神经网络的最后一层卷积层后引入注意力结构块的方法来定位出有效的语义特征,其中注意力结构块由空间注意力结构和通道注意力结构联合组成;最后通过融合不同层次的语义特征形成最终的特征描述子,达到高精度的检索,在基准数据库 Flickr15k 上的实验结果表明所提方法是可行有效的。此外,在手绘图像分类任务中,提出的注意力机制大幅提高了分类精度。

关键词:手绘检索;注意力模型;卷积神经网络;手绘分类
中图法分类号 TP391.41

Sketch-based Image Retrieval Based on Attention Model

LI Zong-min[1], LI Si-yuan[1], LIU Yu-jie[1] and LI Hua[2]
1 College of Computer & Communication Engineering, China University of Petroleum, Qingdao, Shandong 266580, China
2 Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

Abstract To solve the problems of the sparse features and the geometric distortion of hand-drawn images in the research field of SBIR (sketch based image retrieval), a new feature extraction method based on attention model is proposed in this paper. The retrieval results can be obtained efficiently and accurately by accurately extracting the semantic features of hand-drawn images. Firstly, convolutional neural network is used as the basic framework for extracting semantic features, and then the supervised training process is carried out. Attention model mechanism is introduced to locate effective semantic features by adding attention block after the last convolution layer of the convolution neural network, and the attention block is composed of spatial attention structure and channel attention structure. Finally, the final feature descriptor is formed by the fusion of semantic features in different layers, to realize high retrieval accuracy. The experimental results on benchmark Flickr15k dataset proves the feasibility and effectiveness of the proposed method. In addition, the proposed attention model can greatly improve the classification accuracy in the task of sketch classification.

Keywords Sketch-based image retrieval, Attention model, Convolutional neural network, Sketch classification

1 引言

手绘是少数的能够直接表示视觉内容的视觉描述符之一,在计算机视觉领域具有重要的研究意义。近年来,随着触屏设备的普及和大规模自由手绘草图数据集[1-2]的出现,手绘草图在人机交互系统(HCI)中得到了广泛的应用。而手绘图像检索(Sketch Based Image Retrieval,SBIR)所拥有的灵活、方便、易操作等优势正不断地吸引着用户。为了实现高效准确的手绘图像检索效果,本文的目标集中在提取有区分性且高效稳定的手绘图像特征描述子,本文主要研究如何利用注意力模型高效地提取手绘图像中有效的语义特征。

现有的工作中,主流的手绘图像检索方法大多通过边缘检测的方法来提取自然图像的边缘,以降低自然图像和手绘草图之间的语义差异[3-5],而且多选用 SIFT[6],HOG[7] 等手工特征来表示提取边缘,但是这样会产生大量背景噪声。在全局语义特征提取方面,随着近年来卷积神经网络在计算机视觉研究中取得的巨大成功,基于卷积神经网络框架所提取的深度特征表现出了优于手工特征的巨大优势,因此手绘检索任务引入了基于卷积神经网络的深度特征[8-11],然而目前的 CNNs 多是为彩色自然图像设计,不能高效准确地描述手

到稿日期:2019-08-28 返修日期:2019-12-16 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61379106,61379082,61227802);山东省自然科学基金(ZR2013FM036,ZR2015FM011)
This work was supported by the National Natural Science Foundation of China (61379106,61379082,61227802) and Natural Science Foundation of Shandong Province (ZR2013FM036,ZR2015FM011).
通信作者:李思远(875031416@qq.com)

绘图像，在数据库 Flickr15K 上的实验表明，直接使用卷积神经网络提取特征，其检索精度并未有较大提升[10]，这说明在手绘图像检索领域需要更适合手绘图像特性的特征提取方法。

手绘草图由干净的白色背景以及黑色线条构成，具有高度抽象性、稀疏性等特性，不同的手绘者在描述同一物体时，其手绘作品也往往具有较大的差异性。因此，本文在卷积神经网络的基础上引入了注意力机制，通过对注意力模块进行训练，使模型能提取出有效描述手绘草图轮廓信息的语义特征。并且我们在实验中通过可视化训练的注意力模块发现，注意力模块所选取的特征位置点多集中在笔画曲率变化较大的位置，这符合特征选取的预期，从而使得模型提取的特征不仅能精确描述手绘草图，而且对于手绘草图的形变具有较好的鲁棒性。最后本文将学习得到的特征命名为 DASF(Deep Attention Sketch Features)，并在手绘图像检索基准数据集上做了充分的对比实验，验证了其在手绘检索任务中的可行性。此外，在实验过程中发现，通过不同的训练机制，本文提出的注意力模型可以在手绘图像检索任务和手绘图像分类任务中都取得优异的表现，进一步验证了本文方法的可行性和有效性。

2 相关工作

2.1 手绘图像检索

现有的大多数手绘图像检索技术延续了传统的基于内容的图像检索的技术路线。下面分别对传统技术方案以及基于深度学习的方案进行介绍。

传统的手绘图像检索方法被作为形状匹配问题下的分支，因此许多传统的二值形状描述符被应用到手绘图像检索任务中(如 shape band[12]，shape context[13]，chamfer matching[14]，shape-index feature[15])，这些特征描述符在一些简单的手绘草图数据集中具有良好的性能，但是当手绘草图的形状复杂、变化多样时就失去了准确描述图像的能力。而传统的手工特征，如 SIFT[6]，HOG[7] 等，因为对图像的梯度变化把握较准确，因此也被引入手绘检索任务中，但是这些手工特征本身是针对彩色自然图像设计的，对于手绘草图这种由简单的黑白线条构成的图像则缺乏描述性，因此在手绘检索任务和分类任务中的表现并不好。文献[16]提出了一种新的非对称特征映射概念，通过多内核评估来提升手绘图像检索的精度，但是对于手绘草图的形变则缺乏鲁棒性。

近年来，随着深度学习的发展，卷积神经网络在计算机视觉任务中表现出了优良的性能，因此在手绘图像检索这一任务中也引入了使用卷积神经网络提取特征的方法，Wang 等[9]通过对手绘草图数据集做数据增强，通过切割、旋转、提取笔画等操作来扩充草图数据集以训练卷积神经网络，然后利用训练好的卷积神经网络来共同提取手绘草图和彩色自然图像的深度特征，并计算其相似度。文献[8]提出使用 Siamese Network 框架来共同训练手绘草图和彩色图像的边缘图，通过训练把手绘草图和彩色自然图像的边缘图映射到相同的特征空间中。文献[10]首次将“有序”这一概念引入手绘草图的检索任务中，通过卷积神经网络分层次提取手绘草图的特征来提升检索的精度。为了降低手绘草图和彩色自然图像的语义差距，文献[17]首先利用 GAN 把手绘草图生成为彩色自然图像，然后提取特征进行检索。文献[18]利用分段的训练机制融合了分类训练、Siamese 网络训练以及三分支网络训练来共同提取有效的特征描述子，大幅度提升了检索的精度。虽然深度学习的引入使得手绘图像检索任务的结果大幅提升，但是上述方法都没有解决手绘草图的稀疏性和易于形变等问题，因此本文提出了一种基于注意力机制的卷积神经网络来进一步扩展深度学习在手绘图像研究中的应用。

2.2 注意力机制

注意力机制的本质在于使计算机模仿人类观察物品的方式，已被广泛应用到计算机视觉和自然语言处理领域，包括图像语义生成[19]、图像分类[20]等。在不同的注意力模型中，软注意力模型是最常用的，其具有可导性，通过端到端训练可以提供一个分配不同权重的模板，从而能够模仿人类观察物体的方式。与软注意力模型不同，硬注意力模型是不可微的，其训练过程往往通过强化学习来完成。从注意力机制的关注域来划分，可以把注意力机制分为空间域和通道域。文献[21]在基于内容的图像检索任务中引入了空间注意力机制，在卷积层中加入空间注意力模块，通过训练来选取图像中的关键点，从而提升检索精度。在细粒度图像识别任务中，文献[22]通过引入空间注意力模块来提取细粒度图像特征。Hu 等[23]提出在卷积神经网络中不同通道之间也应该有不同的比重，因此他们提出了 SE-Net 模型，通过在训练过程中对特征图进行压缩训练来得到不同通道的权重，并在图像分类任务中提升了分类精度。文献[24]通过多层感知机将空间注意力机制和通道注意力机制连接起来，进一步提升了注意力模型在卷积神经网络中的应用，但是该注意力模型在手绘图像识别任务上却没有良好的表现，因为在处理手绘草图这种稀疏特征的对象时，该模型并不能准确地定位特征关键点。文献[25]提出通过注意力模块来学习手绘草图的细粒度属性，以完成细粒度手绘图像检索任务，但是该方法并不能学习到手绘草图的轮廓关键点信息，因此其特征描述子仍然缺乏对手绘草图的完整描述性。因此，针对以上问题，本文提出了一种新的注意力模型，在手绘图像检索任务和手绘图像分类任务中都取得了良好的表现。

3 本文方法

本文提出在卷积神经网络中引入注意力模型作为基础网络框架，其整体流程如图 1 所示。

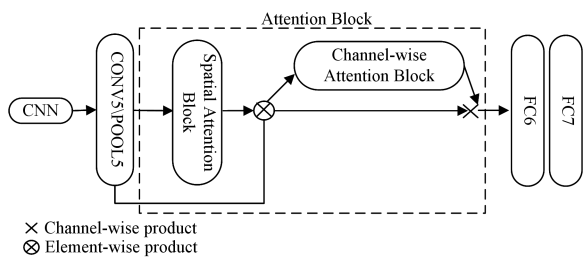

图 1 引入注意力模型的基础模型框架

Fig. 1 Basic model framework with attention module

在手绘图像检索任务中，以此模型作为手绘草图和彩色自然图像边缘图的共同特征提取器，通过把手绘草图和彩色自然图像边缘图映射到相同特征空间中来计算其相似度，再

对其进行排序从而得到检索结果。为了验证注意力模型在手绘研究中的有效性,本文除了在手绘图像检索任务中进行实验验证外,还通过不同训练机制使用该网络框架在手绘图像分类任务中进行了实验对比。

3.1 注意力模型框架

在本文方法中,注意力模块起到特征选择的作用,对于卷积神经网络的任意一层卷积层来说,注意力模块将卷积层的输出作为一个输入,然后通过训练生成一个空间域的选择器和一个通道域的选择器,最后利用该选择器对原卷积层输入做选择滤波,并把输出作为新的输入传输到卷积神经网络的下一层。已有的基于卷积神经网络的方法多选用 VGG16[26] 作为基础网络框架,为了验证注意力模型的有效性,本文方法中的卷积神经网络同样选取 VGG16,并且本文选择把注意力模块引入到 VGG16 框架的最后一层卷积层之后。注意力模块的详细流程如图 2 所示。

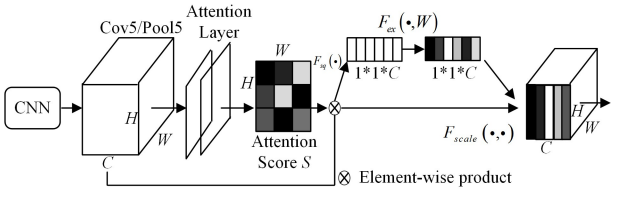

图 2 注意力模块流程图

Fig. 2 Flow chart of attention module

首先定义卷积神经网络的最后一层卷积层输出的特征图为 $X\in R^{H\times W\times C}$,其中 H 和 W 代表特征图的大小,C 代表特征图的通道数。对于特征图中任意一点 (i,j) 的特征向量 $\boldsymbol{f}_{i,j}$,通过式(1)计算其对应的得分:

$$s_{i,j}=F_{spp\_att}(\boldsymbol{f}_{i,j};W_{sa}) \tag{1}$$

$$\alpha_{i,j}=\mathrm{Re}LU(s_{i,j}) \tag{2}$$

其中,$F_{spp\_att}(\cdot)$ 代表空间域注意力模块的映射函数,本文方法中的空间域注意力模块选取两层卷积核大小为 1 的卷积层作为映射函数进行训练,W_{sa} 代表空间域注意力模块学习到的参数权重。针对手绘草图的稀疏性,为了使空间域注意力模块能滤去无用的特征向量,本文选用 RELU 激活函数得到空间域的掩膜 $\alpha_{i,j}$。另外,为了使得空间注意力模块能够学习到特征图中的轮廓关键点信息,本文方法把空间注意力模块放在特征图之后进行学习,而通道注意力模块则是放在空间注意力模块之后进行加权聚合。这与针对彩色自然图像设计的空间域注意力模型[21,24]是不同的。然后,本文方法把输入的特征图和其对应的得分掩膜图像做点积运算,得到新的特征图 $f_{i,j}^{sp\_att}$,如式(3)所示:

$$f_{i,j}^{sp\_att}=\alpha_{i,j}\cdot\boldsymbol{f}_{i,j} \tag{3}$$

在空间域对特征图进行注意力掩膜滤波运算之后,考虑到在手绘草图中没有纹理、颜色等特征信息,在通道域中并不是每一个卷积核的运算都是有效的,因此本文的注意力模型在空间域注意力模块之后又引入了通道域的注意力模块,对特征图的不同通道做特征选择。定义空间域滤波后的特征图为 $X^s\in R^{H\times W\times C}$,并且在空间域上对每个通道的特征图进行压缩得到特征向量 $\boldsymbol{z}$,其第 c 个通道所对应的特征值的计算公式如下:

$$z_c=F_{sq}(u_c)=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}u_c(i,j) \tag{4}$$

与空间域注意力模块一样,通道域注意力模块同样选用 RELU 激活函数。

$$s=F_{ex}(\boldsymbol{z},W)=\delta(g(\boldsymbol{z},W)) \tag{5}$$

$$\delta(g(\boldsymbol{z},w))=\delta(W_2\delta(W_1\boldsymbol{z})) \tag{6}$$

$$s=\delta(W_2\delta(W_1\boldsymbol{z})) \tag{7}$$

其中,δ 代表 RELU 激活函数,而 W_1 和 W_2 与 SE-Net 一样选择两层全连接层,本文选取的是 RELU 激活函数,能够更好地滤去无用的通道特征。最后,通道域注意力模块任一通道的输出为:

$$\tilde{X}_c=F_{scale}(x_c^s,s_c)=s_c\cdot x_c^s \tag{8}$$

由此可以得到注意力模块的最终输出为 $\tilde{X}=[\tilde{X}_1,\tilde{X}_2,\cdots,\tilde{X}_c]$,并将其作为特征图输入卷积神经网络的下一层全连接层中。

3.2 不同任务中的不同训练机制

为了验证本文所提出的注意力模型的有效性,本文在手绘图像检索任务和手绘图像分类任务中都做了实验验证,并针对不同任务选取了不同的训练机制。

对于手绘检索任务,本文以 3.1 节提出的注意力模型为基础网络框架,利用 Siamese-Net[27] 的训练机制来训练注意力模型。在训练过程中,把手绘草图和对应的彩色自然图像的边缘图作为图像对输入卷积神经网络中,通过训练把两个不同域的图像映射到相同特征空间来提取特征,以完成手绘到彩色自然图的检索任务。而对于手绘草图的分类任务,同样使用本文提出的注意力模型作为基础网络框架,但是选取通用的交叉熵损失函数来训练注意力模型。

3.3 相似性度量

本文给定一幅手绘草图和若干张彩色自然图像的边缘图,将其输入上述引入注意力模型的卷积神经网络框架中,并提取最后一层全连接层的输出作为图像的特征描述符 $\boldsymbol{f}^{FC7}$,假定 $\boldsymbol{f}_s^{FC7}$ 表示手绘草图的特征向量,$\boldsymbol{f}_e^{FC7}$ 表示彩色自然图像边缘图的特征向量,则手绘草图和彩色图像的相似度使用 $\boldsymbol{f}_s^{FC7}$ 和 $\boldsymbol{f}_e^{FC7}$ 的欧氏距离来计算。

为了进一步验证注意力模型的有效性,本文除了提取全连接层的输出作为图像描述符外,还提取了注意力模块的输出 $\tilde{X}$,并对其做全局平均池化得到长度为 512 的特征向量 $\boldsymbol{f}^{att}$,然后采用前特征融合的方式,把 $\boldsymbol{f}^{att}$ 和 $\boldsymbol{f}^{FC7}$ 进行了拼接,最后得到长度为 4608 的特征向量作为图像的特征表示,并利用欧氏距离来衡量相似度。

4 实验

4.1 基准数据库

为了验证注意力模型在手绘研究中的可行性,本文在手绘图像检索任务和手绘分类任务中都进行了实验验证,因此本文选取了各自任务下的基准数据集做实验对比。

(1) Flickr15k[2]。在手绘图像检索任务中,本文选取 Flickr15k 针对手绘图像的检索准确率进行对比实验。Flickr15k 是手绘图像检索任务中的基准数据集之一,包含 60 个类别的自然图像,共 14660 幅,以及非专业手绘创作者绘制的 33 个类别的手绘草图(共 329 幅)。

(2) TU-Berlin[1]。在手绘图像分类任务中,本文则选取

TU-Berlin 数据集做实验对比，它是目前手绘分类任务中最大最常用的基准数据集，共包含 20 000 幅手绘草图，分别属于 250 个类别。随机抽取 2/3 的手绘草图作为训练集，将余下图像作为测试集。

4.2 实验设置

为了充分验证本文方法的优越性，设置不同的实验流程来进行对比。

4.2.1 训练集划分

在手绘检索任务中，随机抽取 3/4 的手绘草图作为训练集，将余下图像作为测试集，而在手绘分类任务中则随机抽取 2/3 的手绘草图作为训练集，将余下图像作为测试集。训练迭代的次数设定为 20。

4.2.2 数据增强

数据增强 DA(Data Augmentation)是一种有效的抑制过拟合以及提升模型鲁棒性的方法。因此，本文针对有无数据增强进行对比实验。

本文按照文献[18]中的数据增强方式分别对手绘检索任务和手绘分类任务的数据做了相同的数据增强。首先将手绘草图和自然图像都随机裁剪为 224×224 的分辨率，然后对训练集做了随机翻转增强，再利用文献[28]提出的分组增强的方式，依据手绘草图的笔画重要性将其分为 4 组进行分组扩展，不同的组则分别保留不同重要性的笔画，以此可以产生新的手绘草图，最后在 4 个组中随机抽取其中的手绘草图增添为训练集。

通过数据增强的方式，本文方法在手绘检索任务以及手绘图像分类任务中都取得了较大的提升。

4.2.3 查询拓展

查询拓展 QE(Query Expansion)是一种改进检索效果的标准方法，利用排名靠前的结果来得到新的检索结果[29]。其流程为首先由图像描述符得到原始查询结果，再根据原始查询结果中排名靠前的图像特征计算特征向量的平均值，并将其作为新的查询特征，以此来求出新的查询结果。本文方法中的图像特征描述子为 4608 维特征向量，并且选取前 3 个查询结果计算其平均值，因此新的查询特征包含了前 3 个查询结果中的特征信息，能够更精准地描述目标。通过查询拓展，本文方法在手绘检索任务中取得了较大的精度提升。

4.3 比较方法

在手绘图像检索任务中，本文分别选取了基于传统手工特征的方法和基于深度学习的方法进行实验对比。在基于传统手工特征的方法中，本文挑选了 ShapeContext[13]，GF-HOG[2]，BOW[6]，AFM[16]等方法，BOW 词包模型选取 SIFT 特征描述子作为特征。在基于深度学习的方法中，为了验证注意力模型能够提升深度模型在手绘检索任务中的表现，本文选取了 CBAM[24]，Sketch-A-Net[28]，Siamese-Net[27]，VGG16[26]，Multi-Regression[18]，Attribute-Attention[25]作为对比。为了排除基础网络结构对检索精度的影响，本文选取了 VGG16 和 Sketch-A-Net 网络进行对比实验，其中 Sketch-A-Net 是目前手绘图像分类任务中最优秀的网络框架，本文分别以 Sketch-A-Net 和 VGG16 为基础网络，同样使用 Siamese 训练机制，然后提取其最后一层全连接层作为图像特征向量计算检索精度。本文在手绘检索任务中的评价标准是平均准确率 MAP(Mean Average Precision)。

在手绘分类任务中，本文同样选取了基于手工特征的传统分类方法和基于深度学习的方法进行实验对比。传统手绘分类方法包括 HOG-SVM[1]，multi-kernel SVM (MKL-SVM)[30]，FisherVector SpatialPolling (FV-SP)[31]，DE-HOG[32]。基于深度学习的方法则选取了目前在手绘分类任务中精度较高的 Sketch-A-Net[28]，DSF[9]，FBin[33]，DeepEmbedding[34]，Sketch-R2CNN[35]。同时为了排除本文方法中使用的基础卷积神经网络 VGG16 对分类精度的影响，本文使用单独的 VGG16 做分类训练进行实验对比。

4.4 实验对比分析

根据评价标准，本文分别在手绘图像检索任务和手绘图像分类任务中进行对比实验。

表 1 列出了在手绘图像检索任务中不同方法在 Flickr15K 数据集上检索的平均准确率。从表 1 可以看出，本文所提出的基于注意力模型的框架在手绘图像检索任务中是有效的，在引入了数据增强和查询拓展之后其提取的最终特征描述子的平均准确率超过了所有对比方法的平均准确率。与基于手工特征的传统算法相比，本文方法的精度提升了约 40%，与现有的深度学习方法相比，本文方法的精度提升了 10%～20%。并且通过与 CBAM[24]的实验结果进行对比，验证了本文针对手绘草图设计的注意力模型相比基于彩色图像设计的注意力模型更具有优越性。通过与 Attribute-Attention[25]进行对比可知，本文所提出的针对手绘草图轮廓关键点的注意力模型具有更高的检索精度，这说明只针对细粒度属性进行注意力机制的学习在一般手绘检索任务中具有一定的局限性。另外，可以看到数据增强和查询拓展对于检索的平均精度具有较大的提升，能够得到最好的检索效果。

表 1 Flickr15K 上不同方法的检索精度

Table 1 Retrieval accuracy of different methods on flickr15K

(单位：%)

方法	训练数据集	平均准确率
BOW(sift)	—	5.83
ShapeContext	—	8.14
GF-HOG	—	12.79
HOG	—	13.09
Sketch-A-Net	Flickr15k	34.25
Siamese-Net	Flickr15k	19.54
VGG16	Flickr15K	32.84
AFM	Flickr15K	57.90
CBAM	Flickr15K	40.23
Multi-Regression	Flickr15K	53.26
Attribute-Attention	Flickr15K	42.36
DASF(ours)	Flickr15K	44.54
DASF+DA(ours)	Flickr15K	53.67
DASF+DA+QE(ours)	Flickr15K	58.60

为了对比验证数据增强和查询拓展对本文方法在手绘检索任务中的影响，本文绘制了不同方法在 Flickr15K 数据集上检索结果的 PR 曲线图，如图 3 所示。由图 3 可知，数据增强和查询拓展都对检索的结果有着较好的提升，并且数据增强对于检索精度的提高优于查询拓展。另外，通过对比本文方法和不加注意力模型的 VGG 以及 Sketch-A-Net 方法可以看到，本文提出的 3 种方法的 PR 曲线都较为平滑，这也验证了本文所提出的注意力模型能够使最终学习到的特征描述子对手绘草图有着较高的鲁棒性，进一步验证了本文方法的优越性。

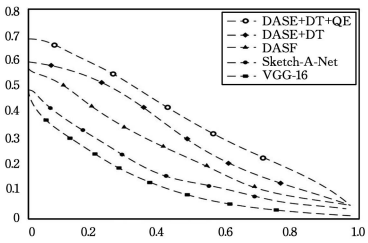

图 3 PR 曲线图

Fig. 3 PR curves

表 2 列出了在手绘图像分类任务中不同方法在 TU-Berlin 数据集中的分类精度。

表 2 TU-Berlin 上不同方法的分类精度

Table 2 Classification accuracy of different methods on TU-Berlin

（单位：%）

方法	训练数据集	分类准确率
HOG-SVM	TU-Berlin	56.0
MKL-SVM	TU-Berlin	65.8
FV-SP	TU-Berlin	68.9
FBin	TU-Berlin	73.8
DSF	TU-Berlin	77.3
Sketch-A-Net	TU-Berlin	74.9
DeepEmbedding	TU-Berlin	82.2
DE-HOG	TU-Berlin	80.6
Sketch-R2CNN	TU-Berlin	83.2
VGG16	TU-Berlin	70.2
DASF(ours)	TU-Berlin	78.2
DASF+DA(ours)	TU-Berlin	83.6

从表 2 可以看出，本文提出的注意力模型在手绘图像分类任务中同样是有效的，基于注意力的分类模型在 TU-Berlin 数据集上的分类精度比传统算法提升了 10%～20%，并且在经过数据增强之后其分类精度在基于深度学习的方法中也达到了最高，相比于目前最优的手绘分类网络 Sketch-A-Net 提升了 8%，相比于 DSF 算法有 6%的提升，而且相比于 DeepEmbedding 也提升了 1%，与 VGG16 网络模型的比较结果可以验证本文提出的注意力模型的有效性。

4.5 可视化实验分析

本文对注意力模型在空间域中的权重做了可视化分析。图 4 为训练之后空间域注意力模块为手绘草图所分配的权重的热力图。从图中可以看出，注意力模块为手绘草图的笔画连接处以及笔画的拐点处分配了较大的权重，因此，在手绘草图发生形变时，本文所提出的注意力模型可以准确地把握物体的整体形状，而且对手绘草图的笔画长度以及线宽的变化具有较好的鲁棒性。

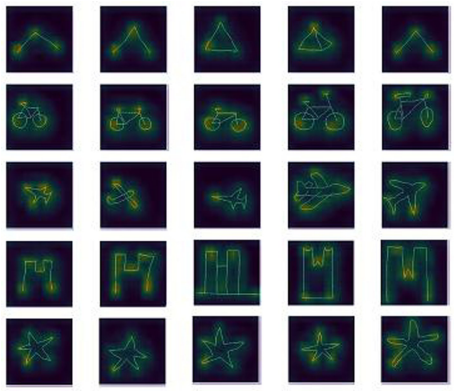

图 4 空间域注意力模块的可视化结果

Fig. 4 Visualization results of spatial attention module

图 5 给出了采用本文提出的注意力模型对 Flickr15K 数据集的手绘图像进行检索的结果。图 5 中，第一行是手绘图像，下方是检索出的 Top-5 自然图像场景，其中红框标注了错误的检索结果。可以看到，这些错误的检索结果和对应的手绘草图也有着相似的形状。这说明本文算法虽然对手绘草图的整体轮廓能够进行较为准确的描述，但是对于手绘草图的细节信息仍具有一定的局限性。下一步工作将尝试融合细粒度信息来进一步提升检索效果。

图 5 Flickr15K 上的部分检索结果（电子版为彩色）

Fig. 5 Partial retrieval results on Flickr15k

结束语 与已有的基于手绘图像的检索技术相比，本文提出了一种针对手绘图像语义特征的混合域注意力模型，为手绘草图的语义特征设计了一种新的筛选机制，通过对注意力模型的训练，注意力模块能够筛选出手绘草图中重要位置的语义特征并为其分配较高的权重。实验结果表明，本文提出的注意力模型对手绘图像的研究具有较强的实用性。

参 考 文 献

[1] EITZ M, HAYS J, AlEXA M. How do humans sketch objects? [J]. Acm Transactions on Graphics, 2012, 31(4): 44.

[2] HU R, COLLOMOSSE J. A performance evaluation of gradient field hog descriptor for sketch based image retrieval[J]. Computer Vision and Image Understanding, 2013, 117(7): 790-806.

[3] EITZ M, HILDEBLAND K, BOUBEKEUR T, et al. A descriptor for large scale image retrieval based on sketched feature lines [C]// Proceedings of Eurographics Symposium on Sketch-based Interfaces & Modeling. ACM, 2009: 29-36.

[4] HU R, BARNARD M, COLLOMOSSE J P. Gradient field descriptor for sketch based retrieval and localization[C]// Proceedings of IEEE International Conference on Image Processing. IEEE, 2010: 1025-1028.

[5] EITZ M, HILDEBRAND K, BOUBEKEUR T, et al. Sketch-Based Image Retrieval: Benchmark and Bag-of-Features Descriptors[J]. IEEE Transactions on Visualization and Computer Graphics, 2011, 17(11): 1624-1636.

[6] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[7] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2005: 886-893.

[8] YU Q, SONG Y Z, ZHANG H, et al. Sketch-based image re-

trieval via Siamese convolutional neural network[C]// Proceedings of IEEE International Conference on Image Processing. IEEE Computer Society Press, 2016.
[9] WANG X, DUAN X, BAI X. Deep Sketch Feature for Cross-domain Image Retrieval[J]. Neurocomputing 2016, 207: 387-397.
[10] LIU Y J, YU D, PANG Y P. Sketch Based Image Retrival Based on Multi-layer Semantic Feature and Deep Convoluntional Neural Network[J]. Journal of Computer-Aided Design and Computer Graphics, 2018, 30(4): 651-657.
[11] LIU Y J, PANG Y P, LU Z Q, et al. Sketch Based Image Retrieval Based on Chamfer Distance Transform and Bag of Mid Maps Descriptor [J]. Journal of Computer-Aided Design & Computer Graphics, 2016, 28(12): 2168-2174.
[12] BAI X, LI Q, LATECKI L J, et al. Shape band: A deformable object detection approach[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009.
[13] MORI G, BELONGIE S, MALIK J. Efficient shape matching using shape contexts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(11): 1832-1837.
[14] THAYANANTHAN A, STENGER B, TORR P H S, et al. Shape context and chamfer matching in cluttered scenes[J]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, 1: 127-133.
[15] XIA G S, DELON J, GOUSSEAU Y. Shape-based Invariant Texture Indexing[J]. International Journal of Computer Vision, 2010, 88(3): 382-403.
[16] TOLIAS G, CHUM O. Efficient Contour Match Kernel[J]. Image & Vision Computing, 2018, 76: 14-26.
[17] LIU Y J, DOU C H, ZHAO Q L. Sketch Based Image Retrival with Conditional Generative Adversarial Network[J]. Journal of Computer-Aided Design and Computer Graphics, 2017, 29(12): 2336-2342.
[18] BUI T, RIBEIRO L, PONTI M, et al. Sketching out the details: Sketch-based image retrieval using convolutional neural networks with multi-stage regression[J]. Computers & Graphics, 2018, 71: 77-87.
[19] LU J, XIONG C, PARIKH D, et al. Knowing when to look: Adaptive attention via a visual sentinel for image captioning[EB/OL]. [2016-02-06]. https://arxiv.org/abs/1612.01887.
[20] MNIH V, HEESS N, GRAVES A, et al. Recurrent Models of Visual Attention[J]. Advances in neural information processing systems, 2014, 2: 2204-2212.
[21] NOH H, ARAUJO A, SIM J, et al. Large-Scale Image Retrieval with Attentive Deep Local Features[C]// Proceedings of IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, 2017.
[22] XIAO N T, XU N Y, YANGN K, et al. The application of two-level attention models in deep convolutional neural network for fine-grained image classification[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2015: 2.
[23] HU J, LI S, ALBANIE S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 99: 1-1.
[24] WOO S, PARK J, LEE J Y, et al. Convolutional block attention module[C]// Proceedings of the EuropeanConference on Computer Vision (ECCV). 2018: 3-19.
[25] SONG J, YU Q, SONG Y Z, et al. Deep Spatial-Semantic Attention for Fine-Grained Sketch-Based Image Retrieval[C]// IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, 2017.
[26] SIMONYAN K, ZISSERMN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2017-06-15]. https://arxiv.org/abs/1409.1556.
[27] CHOPRA S, HADSELL R, LECCUN Y. Learning a similarity metric discriminatively, with application to face verification [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2005: 539-546.
[28] YU Q, YANG Y, SONG Y Z, et al. Sketch-a-net that beats humans[EB/OL]. [2017-06-15]. https://arxiv.org/abs/1501.07873.
[29] JOLY A, BUISSON O. Logo retrieval with a contrario visual query expansion[C]// International Conference on Multimedia. 2009.
[30] LI Y, HOSPEDALES T M, SONG Y Z, et al. Free-hand sketch recognition by multi-kernel feature learning[J]. Computer Vision and Image Understanding, 2015, 137: 1-11.
[31] SCHNEIDER, ROSALIA G, TUYTELAARS T. Sketch classification and classification-driven analysis using Fisher vectors[J]. ACM Transactions on Graphics, 2014, 33(6): 1-9.
[32] ZHONG Y, ZHANG H G, GUO J S, et al. Directional Element HOG for Sketch Recognition[C]// International Conference on Network Infrastructure and Digital Content (IC-NIDC). 2018.
[33] PRABHU A, BATCHU V, GAJAWADA R, et al. Hybrid Binary Networks: Optimizing for Accuracy, Efficiency and Memory [C]// IEEE Winter Conference on Applications of Computer Vision (WACV). 2018, 10: 821-829.
[34] MISHRA, SINGH A K. Deep Embedding using Bayesian Risk Minimization with Application to Sketch Recognition[EB/OL]. [2018-12-6]. https://arxiv.org/abs/1812.02466.
[35] LI L, ZOU C, ZHENG Y, et al. Sketch-R2CNN: An Attentive Network for Vector Sketch Recognition[EB/OL]. [2018-11-20]. https://arxiv.org/abs/1811.08170.

LI Zong-min, born in 1965, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include computer graphics, image processing, and scientific computing visualization.

LI Si-yuan, born in 1996, postgraduate. His main research interests include computer vision, image processing, image retrieval, and sketch image recognition.