增.对于在手绘草图数据集上训练的对比深度网络,本文使用的手绘草图训练库是 TU-Berlin[1].这个手绘草图数据集是目前最大的和最通用的数据集,包含了 20 000 张手绘草图.但是,数据量不足以驱动深度卷积神经网络,容易对数据产生过拟合现象.所以,采用与文献[20]相同的数据扩增方式,即通过镜像、(水平、垂直)平移、或者旋转,将数据扩增到 32×32×11×2 倍.

在本文的实验中,训练的实验数据集有 ImageNet、TU-Berlin 手绘数据集和 Flickr15k 数据集.实验中,本文将实验数据集大体按照 7:3 的比例进行分割.所有模型在训练集完成训练并收敛之后,再在测试集上进行测试,从而获得每个模型的检索精度.

3.3 评价标准

本文使用的评价标准是 mAP、平均准确率(mean average precision).本文中对于 SBIR 中的 mAP 定义如下.

$$mAP=\frac{1}{N}\sum_{q\in N_q}\sum_{r\in R}\frac{RT_i/IT_i}{N_R} \tag{4}$$

其中,q 表示在检索序列里面的检索手绘草图,N_q 是检索序列中的手绘草图的总数,RT_i 是指检索出来的正样本的索引号,IT_i 表示的是在检索序列中正样本的排序索引,N_R 表示正样本的总数.

3.4 实验结果

本文方法与现有方法的实验对比结果见表 1.为了充分验证本文网络框架的有效性,对于现有的深度学习网络,设置了 3 种类型的训练方式来使现有的 CNN 网络的学习能力达到最大化.

(1) 第 1 类是单独在自然彩色图片 ImageNet 上训练的深度卷积神经网络特征,其中,使用 LeNet 的数据集 MINIST 是因为手写数字数据集和我们的手绘草图在视觉上有一定的相似性.

(2) 第 2 类是单独在手绘草图数据库 TU-Berlin 上训练的深度卷积神经网络,用所学习的特征进行检索.

(3) 第 3 类是使用训练好的深度卷积网络,在 Flickr15k 数据集上微调,然后使用再次收敛的网络所生成的特征进行检索.

Table 1 Retrieval performance of different methods

表 1 不同方法的检索表现

名称	描述	训练数据集	检索数据集	mAP
VGG-16	深度网络(自然图片数据集上训练)	ImageNet	Flickr15k	0.179 3
VGG-19		ImageNet		0.226 1
VGG		ImageNet		0.288 2
Alexnet		ImageNet		0.293 5
Sketch-A-Net(single)	深度网络(手绘草图数据集上训练)	TU-Berlin	Flickr15k	0.153 8
AlexNet-Sketch		TU-berlin		0261 6
VGG-Sketch		TU-berlin		0.265 8
Siamese CNN[18]	深度网络(数据集 Flickr15k 上训练)	Flickr15k	Flickr15k	0.195 4
Sketch-A-Net(single)		Flickr15k		0.237 4
Sketch-Photo Mode[1,17]		Flickr15k		0.361 7
AlexNet		Flickr15k		0.381 1
VGG		Flickr15k		0.429 3
Quardruplet Network[19]		Flickr15k		0.433 0
VGG-16		Flickr15k		0.458 2
VGG-19		Flickr15k		0.501 7
多层卷积神经网络	本文框网络框架	Flickr15k	Flickr15k	0.557 4

表 1 中,Sketch-A-Net(single)网络结构是参考引用文献[1]中的网络参数实现出来的单一尺度的深度网络,其中,本文使用单一图片尺度为 256×256.

从表 1 可以得到的结论如下:多层融合卷积神经网络的跨域建模的视觉表达和多层深度融合卷积神经网络的特征表达模型对于手绘草图检索(SBIR)领域中手绘草图和自然图片具有较强的拟合能力;本文模型所生成的特征向量的检索准确率远超基于自然图片训练的模型,如主流的 Alexnet 网络和 VGG 框架等,同时,对于基于手绘草图数据集训练的深度卷积神经网络的性能也有较大的提升.另外,在第 3 类实验中,本文通过与同领域

基于手绘草图跨域检索方法进行比较,如 Sketch-Photo Model[17]、Siamese CNN[18]与 Quardruplet Network[19]等,实验结果表明,本文的多层卷积神经网络比同类方法的检索准确率大致提升了 12%~20%.从表 1 的实验结果可以看出本文方法相对于现有方法在网络性能上的提升.

为了揭示手绘草图和自然图片不同层次之间的内在联系,对不同视觉层次的检索结果进行分析,实验的对比结果见表 2.

Table 2 Retrieval results on different layers

表 2 不同层次的检索效果

层次	层次 1	层次 2	层次 3
mAP	0.410 1	0.458 2	0.529 8

在表 2 中,实验中使用的检索特征是 VGG-19 网络 FC7 的特征向量,在拥有最少的空间、语义信息的第 1 层,获得了相对较低的检索精度;而在具有最细致语义信息和空间信息的第 3 层,获得了 52.9%的检索精度,仅仅比本文的多层融合卷积神经网络少 3%的精度.同时,这也是本文的特征融合策略有效性的证明.

为了探索不同的特征融合策略对于手绘草图检索(SBIR)的影响,我们对不同的融合策略进行了实验对比,实验结果见表 3.

Table 3 Impacts of different feature fusion strategies on retrieval results

表 3 特征融合策略对于检索结果的影响

特征融合策略	均值融合	串联融合	权重融合
mAP	0.360 4	0.557 4	0.438 1

从表 3 可知,不同的特征融合策略对于提升或者拉低本文最终的融合特征向量的辨别力产生了非常大的影响,尤其是一个较差的融合策略能够将本文的多层特征的检索精度从 55.74%降低到 36.04%.所以,一个合适的特征融合策略对于多层融合卷积神经网络能力的提升具有非常重要的意义.

最后,多层融合卷积神经网络对 Flickr15k 数据集的检索结果如图 6 所示.

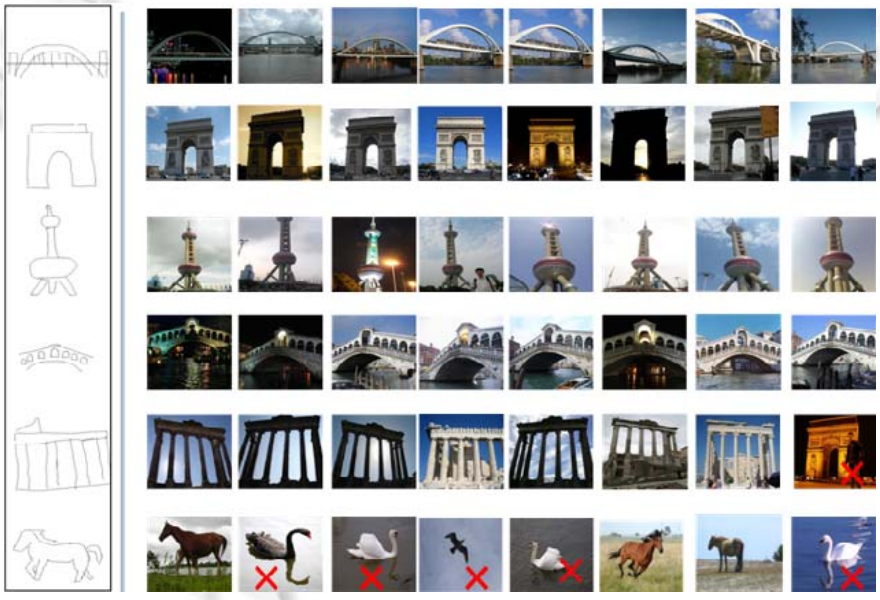

Fig.6 Retrieval performance of multi-layer fusion CNN on Flickr15k

图 6 多层融合卷积神经网络在 Flickr15k 上的检索表现

如图6所示,“×”表示错误的检索结果.其中最后一行展示的是多层融合卷积神经网络最差的检索表现,主要是因为 Flickr15k 数据集中这类图片的训练数据远远不足的缘故.

4 总结与未来工作

在本文中,“多层”的概念一直是模型中至关重要的主题,而在实验中发现,本文的多层视觉表达方法对于手绘草图或者自然图片具有数据扩增的潜力.与此同时,本文的理论中仍存在着一个猜想:对于手绘草图和自然图片的边缘图,是否它们的抽象层次越高,它们之间的相似度就越大?在实验中,本文的3层视觉表达的确显示了这样的趋势.

关于设置多少层才能最佳地利用好手绘草图和彩色自然图片中的层次信息的问题,本文的实验仅仅证明了多层表达能够进一步提高检索表现,所以仍然需要进行更多的实验来解答这个问题.正如本文的实验部分所示,抽象的程度越高,在自然图片的边缘图中的背景噪音就越少.所以我们猜想:在某一个合适的抽象层次,能够获得一个最好的视觉表达和特征表示.

本文的特征采用的特征融合策略都比较简单,与此同时,特征融合阶段也是提高多层卷积神经网络的最终特征的辨别力的关键环节.所以,一种更好的特征融合方法将会为手绘草图检索(SBIR)的检索精度带来巨大的提升.

References:

[1] 1Eitz M, Hays J, Alexa M. How do humans sketch objects? ACM Trans. on Graph, 2012,31(4):44:1−44:10. [doi: 10.1145/2185520.2185540]

[2] Fu H, Zhou S, Liu L, *et al*. Animated construction of line drawings. ACM Trans. on Graphics, 2011,30(6):1−10. [doi: 10.1145/2070781.2024167]

[3] Yu Q, Yang Y, Song YZ, *et al*. Sketch-a-Net that beats humans. arXiv preprint arXiv:1501.07873, 2015. [doi: 10.5244/C.29.7]

[4] 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):Article No.119. [doi: 10.1145/2897824.2925954]

[5] Lim JJ, Zitnick CL, Dollár P. Sketch tokens: A learned mid-level representation for contour and object detection. In: Proc. of the 2013 IEEE Conf. on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2013. 3158−3165. [doi: 10.1109/CVPR.2013.406]

[6] Arbelaez P, Maire M, Fowlkes C, *et al*. Contour detection and hierarchical image segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2011,33(5):898−916. [doi: 10.1109/TPAMI.2010.161]

[7] 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. Washington: IEEE Computer Society, 2016. 799−807. [doi: 10.1109/CVPR.2016.93]

[8] Su H, Maji S, Kalogerakis E, *et al*. Multi-View convolutional neural networks for 3D shape recognition. In: Proc. of the 2015 IEEE Int'l Conf. on Computer Vision. Washington: IEEE Computer Society, 2015. 945−953. [doi: 10.1109/ICCV.2015.114]

[9] Lowe DG. Distinctive image features from scale-invariant keypoints. Int'l Journal of Computer Vision, 2004,60(2):91−110. [doi: 10.1023/B:VISI.0000029664.99615.94]

[10] Mori G, Belongie S, Malik J. Efficient shape matching using shape contexts. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005,27(11):1832−1837. [doi: 10.1109/TPAMI.2005.220]

[11] 11Cao Y, Wang C, Zhang L, *et al*. Edgel index for large-scale sketch-based image search. In: Proc. of the 2011 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Washington: IEEE Computer Society, 2011. 761−768. [doi: 10.1109/CVPR.2011.5995460]

[12] 12Hu R, Collomosse J. A performance evaluation of gradient field hog descriptor for sketch based image retrieval. Computer Vision and Image Understanding, 2013,117(7):790−806. [doi: 10.1016/j.cviu.2013.02.005]

[13] 13−Xiao C, Wang C, Zhang L, *et al*. Sketch-Based image retrieval via shape words. In: Proc. of the 5th ACM Int'l Conf. on Multimedia Retrieval. New York: ACM Press, 2015. 571−574. [doi: 10.1145/2671188.2749360]

[14] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Proc. of the 2012 Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2012. 1097−1105. [doi: 10.1145/3065386]

[15] Le Cun Y, Bottou L, Bengio Y, *et al*. Gradient-Based learning applied to document recognition. Proc. of the IEEE, 1998,86(11): 2278−2324. [doi: 10.1109/5.726791]

[16] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[17] Bui T, Ribeiro L, Ponti M, *et al*. Generalisation and sharing in triplet convnets for sketch based visual search. arXiv preprint arXiv: 1611.05301, 2016.

[18] Qi Y, Song YZ, Zhang H, *et al*. Sketch-Based image retrieval via Siamese convolutional neural network. In: Proc. of the 2016 IEEE Int'l Conf. on Image Processing (ICIP). Los Alamitos: IEEE Computer Society Press, 2016. 2460−2464. [doi: 10.1109/ICIP.2016.7532801]

[19] Seddati O, Dupont S, Mahmoudi S. Quadruplet networks for sketch-based image retrieval. In: Proc. of the 2017 ACM Int'l Conf. on Multimedia Retrieval. New York: ACM Press, 2017. 184−191. [doi: 10.1145/3078971.3078985]

[20] Yu Q, Yang Y, Song YZ, *et al*. Sketch-a-Net that beats humans. arXiv preprint arXiv:1501.07873, 2015.

[21] Feng GH, Sun ZX, Viard-Gaudin C. Stroke fragmentation using geometry features and hidden Markov model. Ruan Jian Xue Bao/Journal of Software, 2009,20(1):1−10 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/3383.htm [doi: 10.3724/SP.J.1001.2009.03383]

[22] Liu YJ, Pang YP, Lu ZQ, *et al*. Sketch based image retrieval based on chamfer distance transform and bag of mid maps descriptor. Journal of Computer-Aided Design & Computer Graphics, 2016,28(12):2168−2174 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-9775.2016.12.017]

[23] Deng J, Dong W, Socher R, *et al*. Imagenet: A large-scale hierarchical image database. In: Proc. of the 2009 IEEE Conf. on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2009. 248−255. [doi: 10.1109/CVPR.2009.5206848]

附中文参考文献:

[21] 冯桂焕,孙正兴,Viard-Gaudin C.使用几何特征与隐 Markov 模型的手绘笔画图元分解.软件学报,2009,20(1):1−10. http://www.jos.org.cn/1000-9825/3383.htm [doi: 10.3724/SP.J.1001.2009.03383]

[22] 刘玉杰,庞芸萍,路子奇,等.结合距离变换和隐层图词包的手绘图像检索方法.计算机辅助设计与图形学学报,2016,28(12): 2168−2174. [doi: 10.3969/j.issn.1003-9775.2016.12.017]

于邓(1992−),男,山东潍坊人,硕士,主要研究领域为计算机图形学,模式识别,机器学习,手绘检索与识别.

刘玉杰(1971−),男,博士,副教授,CCF 专业会员,主要研究领域为计算机图形图像处理,多媒体数据分析,多媒体数据库,多媒体数据压缩.

邢敏敏(1992−),女,硕士,主要研究领域为图像处理,行人检测.

李宗民(1965−),男,教授,博士生导师,CCF 高级会员,主要研究领域为计算机图形学,图像处理,模式识别.

李华(1956−),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图形图像处理.