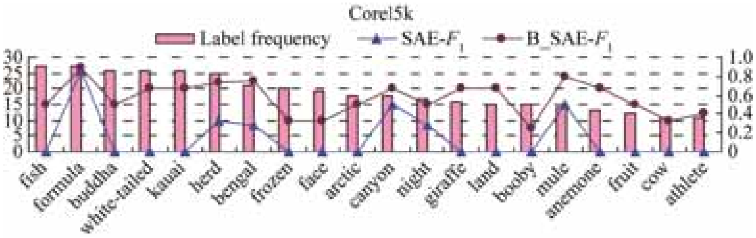


Fig.6 F_1 value of the low frequency labels of SAE and B-SAE on Corel5k

图 6 在 Corel5k 上测试 SAE 和 B-SAE 预测低频标签的 F_1 值

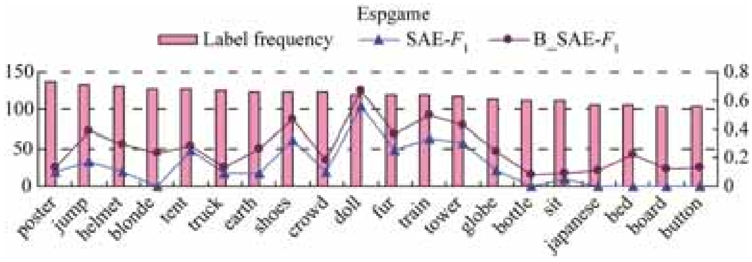


Fig.7 F_1 value of the low frequency labels of SAE and B-SAE on Espgame

图 7 在 Espgame 上测试 SAE 和 B-SAE 预测低频标签的 F_1 值

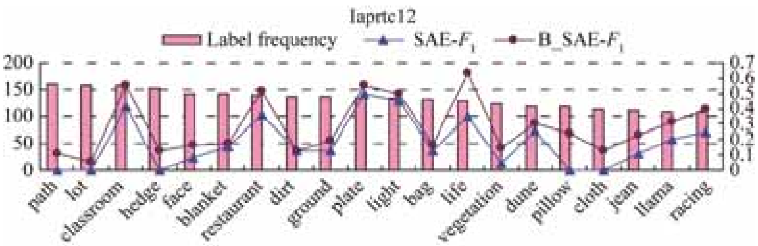


Fig.8 F_1 value of the low frequency labels of SAE and B-SAE on Iaprtc12

图 8 在 Iaprtc12 上测试 SAE 和 B-SAE 预测低频标签的 F_1 值

5.2.3 RB-SAE 算法的有效性

RB-SAE 算法是在 B-SAE 模型基础上,针对单个 B-SAE 标注结果不稳定、容易随参数变动的问题,而提出的改善算法,从而提升 B-SAE 的鲁棒性.为了说明单个 B-SAE 模型存在不稳定现象并验证该模型的效果,我们分 4 组进行测试, *Ran*(·) 产生的随机向量的分量取值服从 $N(0,1)$ 分布和 $U[0,1]$ 分布.每种加噪方式又设置两个 χ 参数,1.2 和 1.5.组内又分 3 个子 B-SAE 模型,依次设置隐层神经元个数为 450、500 和 550.总共 12 个子模型,参数设定见表 5.

Table 5 Parameter settings for testing the RB-SAE algorithm

表 5 测试 RB-SAE 算法的参数设定

子模型	常系数	随机函数	隐层神经元个数
B-SAE(1/1)	χ=1.2	Gaussian	450
B-SAE(1/2)	χ=1.2	Gaussian	500
B-SAE(1/3)	χ=1.2	Gaussian	550
B-SAE(2/1)	χ=1.5	Gaussian	450
B-SAE(2/2)	χ=1.5	Gaussian	500
B-SAE(2/3)	χ=1.5	Gaussian	550
B-SAE(3/1)	χ=1.2	Uniform	450
B-SAE(3/2)	χ=1.2	Uniform	500
B-SAE(3/3)	χ=1.2	Uniform	550
B-SAE(4/1)	χ=1.5	Uniform	450
B-SAE(4/2)	χ=1.5	Uniform	500
B-SAE(4/3)	χ=1.5	Uniform	550

使用上述 12 个 B-SAE 子模型的参数设定,分别在 3 个数据集上的执行过程如图 9~图 11 所示.

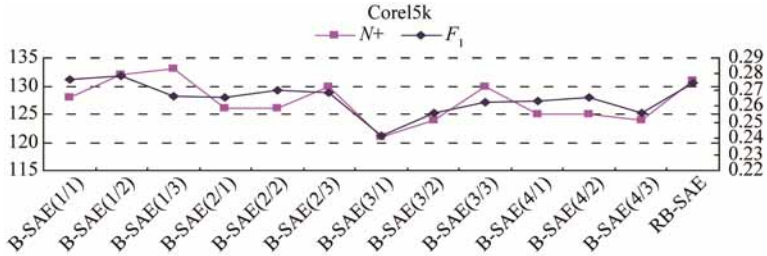


Fig.9 Results of the RB-SAE algorithm on Corel5k

图 9 在 Corel5k 上测试 RB-SAE 算法

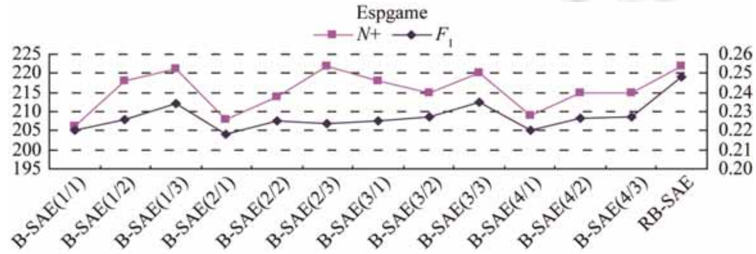


Fig.10 Results of the RB-SAE algorithm on Espgame

图 10 在 Espgame 上测试 RB-SAE 算法

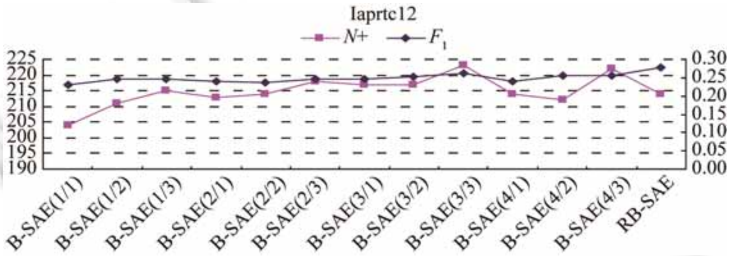


Fig.11 Results of the RB-SAE algorithm on Iaprtc12

图 11 在 Iaprtc12 上测试 RB-SAE 算法

在图 9~图 11 中,粉色折线图对应左纵轴,代表子模型至少准确预测 1 次的关键词个数 N^+,紫色折线图对应右纵轴,代表子模型的 F_1 值,每个图的最后一列为执行 RB-SAE 算法后得到的结果.从图中我们可以看出,参数设定不同,单个 B-SAE 模型得到的数据结果不稳定,要么 N^+高一点,F_1 值低一点(如图 10 中的 B-SAE(2/3),N^+较高但 F_1 值较低);要么 F_1 高一点,N^+低一点(如图 11 中的 B-SAE(1/1),F_1 值较高但 N^+较低),通过执行 RB-SAE 算法后,N^+和 F_1 值都可以取到相对较为稳定的结果.

为了方便对比 SAE 系列模型的整体效果,我们把 3 个模型的结果记录在表 6 中.SAE 为传统栈式自动编码器.B-SAE 为本文提出的平衡栈式自动编码器,这里数据对应于表 3 中加粗的部分.RB-SAE 为本文提出的鲁棒平衡栈式自动编码器算法,所给出的数据对应于图 9~图 11 的执行结果,即每个图的最后一列.

Table 6 Comparison of three auto-encoder models

表 6 3 个自动编码器模型的测试对比

模型	Corel5k				Espgame				Iaprtc12			
	P	R	F_1	$N+$	P	R	F_1	$N+$	P	R	F_1	$N+$
SAE	0.20	0.23	0.21	110	0.23	0.18	0.20	201	0.30	0.20	0.24	208
B-SAE	0.25	0.31	0.28	128	0.25	0.21	0.23	221	0.31	0.22	0.26	223
RB-SAE	0.24	0.32	0.27	131	0.27	0.23	0.25	222	0.35	0.23	0.28	214

5.2.4 ADA 的实验效果

ADA 是针对数据集不平衡问题,以模型外,即标注过程,作为出发点提出来的一种根据测试图像的高低频属性来选择标注过程的标注策略.为了说明该标注框架的合理性,我们对仅用鲁棒平衡栈式自动编码器算法(RB-SAE)和仅用语义传播算法(SP)的结果进行对比,发现大多数高频标签用 RB-SAE 得到的 F_1 值比用 SP 算法得到的 F_1 值要高,而中低频标签用 SP 算法得到的 F_1 值比用 RB-SAE 得到的 F_1 值要高,我们分别在 3 个数据集上取 20 个关键词进行统计,其中 10 个为高频词,另外 10 个为低频词,具体情况如图 12~图 14 所示.

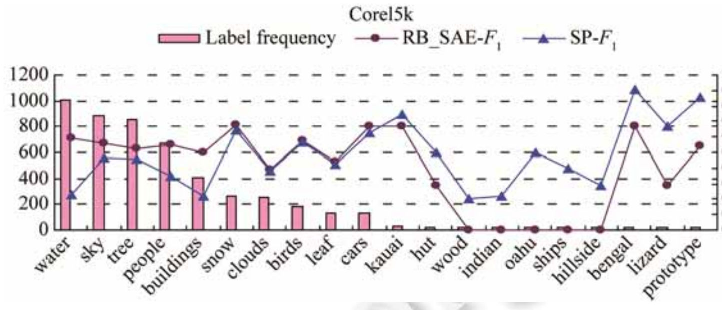


Fig.12 F_1 value of the high and low frequency labels on Corel5k computed by RB-SAE and SP

图 12 在 Corel5k 上测试 RB-SAE 和 SP 预测高低词频标签 F_1 值

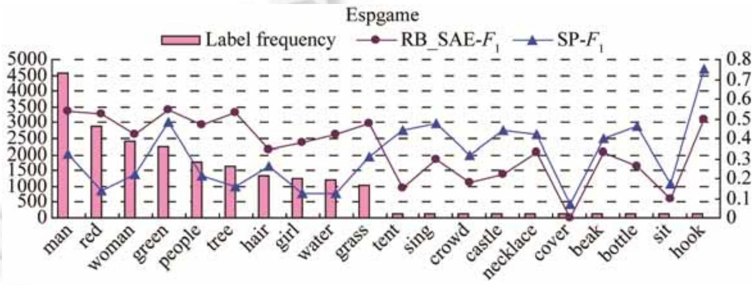


Fig.13 F_1 value of the high and low frequency labels on Espgame computed by RB-SAE and SP

图 13 在 Espgame 上测试 RB-SAE 和 SP 预测高低词频标签 F_1 值

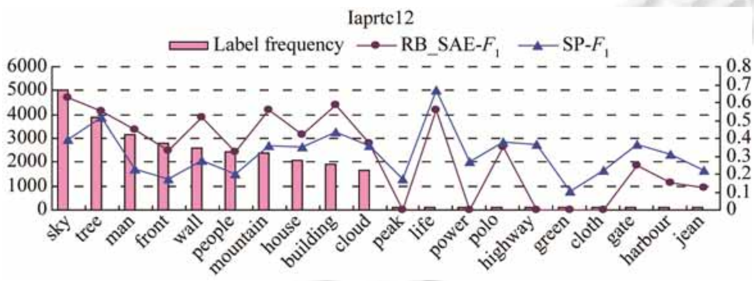


Fig.14 F_1 value of the high and low frequency labels on Iaprtc12 computed by RB-SAE and SP

图 14 在 Iaprtc12 上测试 RB-SAE 和 SP 预测高低词频标签 F_1 值

我们把提出的模型和一些经典模型以及近几年效果较好的模型在 3 个数据集上进行比较,从表 7 可以看出,我们的模型的效果大幅超过了一些经典模型,且比得上目前较好的模型.由于属性判别标注模型(ADA)综合了鲁棒自动编码器算法(RB-SAE)可以较好地预测高频属性图像和语义传播算法(SP)可以较好地预测中低频属性图像的特点(如图 12~图 14 所示),使得整体标注效果得到较大提升,在数据集 Corel5k 和 Iaprtc12 上得到的 F_1

值是最好的.又由于局部均衡数据集平衡了高频标签和低频标签的出现频次,增大了低频标签被预测的可能性,因此至少准确预测 1 次的关键词个数 N^{+}得到很大提升,在 3 个数据集上分别达到 172、251 和 280 个,大幅超过其他模型.在 3 个数据集上,我们的模型的准确率 P 和召回率 R 也均达到当前较好水平,说明我们提出的 ADA 模型是有效的.

Table 7 Comparison between the ADA model and other models

表 7 本文 ADA 模型与其他模型的对比

模型	Corel5k				Espgame				Iaprtc12			
	P	R	F_1	N+	P	R	F_1	N+	P	R	F_1	N+
MBRM[5]	0.24	0.25	0.24	122	0.18	0.19	0.18	209	0.24	0.23	0.23	223
GS[27]	0.30	0.33	0.31	146	–	–	–	–	0.32	0.29	0.30	252
JEC[6]	0.27	0.32	0.29	139	0.24	0.19	0.21	222	0.29	0.19	0.23	211
TagProp-ML[7]	0.31	0.37	0.34	146	**0.49**	0.20	0.28	213	**0.48**	0.25	0.33	227
LM3L[28]	**0.33**	0.37	0.35	146	0.40	0.26	**0.32**	239	0.44	0.28	0.34	242
NW-RNN[29]	0.29	0.32	0.30	149	–	–	–	–	0.28	0.30	0.29	259
RNN[29]	0.31	0.34	0.32	149	–	–	–	–	0.33	0.31	0.32	255
χ^2Kernel[30]	0.31	0.39	0.35	153	0.38	0.21	0.27	214	0.42	0.24	0.31	239
ANNOR-G[31]	0.22	0.29	0.25	129	0.36	**0.29**	0.32	231	0.38	0.31	0.34	242
FFSS[32]	0.27	0.33	0.30	141	0.21	0.23	0.22	221	0.29	0.29	0.29	251
MLRank[33]	0.32	0.37	0.34	151	–	–	–	–	0.38	**0.32**	0.35	259
ADA	0.32	**0.40**	**0.36**	**172**	0.35	0.21	0.26	**251**	0.42	0.30	**0.35**	**280**

5.2.5 结果展示

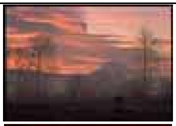
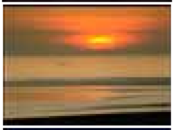
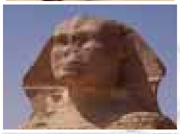
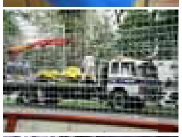
下文所给出的表 8 中的 9 幅图是本文提出的 ADA 模型的标注实例.在图像自动标注结果一列中,我们用黑色加粗的字体表示自动标注结果与人工标注结果相同,而使用斜体字体的关键词表示可以描述图像的内容,但是并没有被人工标注出来.在这里,我们并没有选择完全被标注正确的那些图像,而是选择了部分能够比较好地反映本文模型特点的一些图像.从图中可以看出,本文一些图的标注结果虽然与原始图像上的人工标注结果有区别,但却是对原始图像标注结果的有益补充,能够更加准确地描述图像的语义信息.例如,第 1 幅图像的人工标注并未将 clouds 这一关键词给标注上,而从图像的场景来看,clouds 显然要作为一个重要的关键词来描述该幅图像的场景.在第 2 幅图像中,从人的视觉角度分析,显然用 sea 这个关键词相比原始图像中的 water 更有说服力,并且原始图像中也疏漏了 beach 等从图像中可以直接得到的关键词.此外,在对抽象概念 old,area 等的描述上,原始图像中的信息并不能对其进行准确的描述,或者说,单从人的视觉角度出发,无法从图像上得到这些信息.因此,这也从另一个角度说明了人工标注存在的一些问题,可能存在漏标注,并且不同人对同一幅图像的认识也存在一定的主观差异.同一幅图像,不同的人可能给出不同的标注结果.

6 总 结

本文将栈式自动编码器(SAE)应用于图像标注任务,改善了传统的基于浅层机器学习模型标注效率低下、模型泛化能力弱等问题,并针对图像标注数据不平衡问题提出了两种解决方案:(1) 对于模型本身,我们提出了平衡栈式自动编码器(B-SAE)模型,该模型可以针对特定样本进行增强训练,解决了传统模型难以有效训练不平衡数据的问题,有效提升了低频标签的 F_1 值,并在 B-SAE 的基础上提出鲁棒平衡栈式自动编码器算法(RB-SAE),有效地解决了 B-SAE 模型本身不稳定的问题,并得到稳定的标注结果;(2) 对于标注过程,我们以待测图像作为出发点,首先为每一幅待测图像构造局部均衡数据集,并在此数据集上实现了一种有效提升中低频标签标注效果的语义传播算法(SP);然后通过判定待测图像的高低频属性来分步标注未知图像(ADA).一方面,用 RB-SAE 算法标注高频图像;另一方面,用 SP 算法标注中低频图像,标注过程取长补短,提升了整个模型的标注效果,3 个数据集上的实验验证了我们方法的有效性.此外,如何让标注模型更好地解决弱对象、背景信息、水印等噪声干扰问题是下一步的主要研究内容.

Table 8 Annotation instances of the ADA model

表 8 本文 ADA 模型的标注实例

数据集	图像	人工标注结果	ADA 模型标注结果
Corel5k		mountain, sky, sun, water	*clouds*, *peaks*, *sunset*, elephant, silhouette
		sun, water, clouds, birds	**sun**, **water**, *sea*, *beach*, **birds**
		sky, water, people, sand	*beach*, **sand**, *shadows*, coyote, maui
Espgame		blue, desert, game, man, people, yellow	**game**, **man**, **people**, soldier, **yellow**
		old, sky, stone	*brown*, *mouth*, nose, **sky**, smile
		band, group, man, red	*black*, **group**, **man**, *people*, **red**
Iaprtc12		boy, door, front, jumper	**boy**, couch, **door**, **front**, *pullover*
		area, car, fence, racetrack, racing, tree	**car**, *man*, **racing**, spectator, **tree**
		grandstand, lawn, people, round, stadium, team, uniform	**lawn**, **round**, **stadium**, **team**, **uniform**

References:

[1] Mori Y, Takahashi H. Image-to-Word transformation based on dividing and vector quantizing images with words. In: Proc. of the 1st Int'l Workshop on Multimedia Intelligent Storage and Retrieval Management. 1999. 405–409. http://xueshu.baidu.com/s?wd=Image-to-Word+transformation+based+on+dividing+and+vector+quantizing+images+with+words&rsv_bp=0&tn=SE_baiduxueshu_c1gjeupa&rsv_spt=3&ie=utf-8&f=8&rsv_sug2=1&sc_f_para=sc_tasktype%3D%7BfirstSimpleSearch%7D&rsv_n=2

[2] Duygulu P, Barnard K, Freitas JFGD, Forsyth DA. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. European Conf. on Computer Vision, 2002,2353:97–112. [doi: 10.1007/3-540-47979-1_7]

[3] Wang M, Zhou XD, Zhang JQ, Xu HT, Shi BL. Image auto-annotation via an extended generative language model. Ruan Jian Xue Bao/Journal of Software, 2008,19(9):2449–2460 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/19/2449.htm [doi: 10.3724/SP.J.1001.2008.02449]

[4] Lavrenko V, Manmatha R, Jeon J. A model for learning the semantics of pictures. In: Advances in Neural Information Processing Systems. 2004. 553–560. http://xueshu.baidu.com/s?wd=A+model+for+learning+the+semantics+of+pictures&tn=SE_baiduxueshu_c1gjeupa&cl=3&ie=utf-8&bs=Image-to-Word+transformation+based+on+dividing+and+vector+quantizing+images+with+words&f=8&rsv_bp=1&rsv_sug2=1&sc_f_para=sc_tasktype%3D%7BfirstSimpleSearch%7D&rsv_spt=3&rsv_n=2

[5] Feng SL, Manmatha R, Lawenko V. Multiple Bernoulli relevance models for image and video annotation. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. 2004,2:1002–1009. [doi: 10.1109/CVPR.2004.1315274]

[6] Makadia A, Pavlovic V, Kumar S. A new baseline for image annotation. In: Proc. of the European Conf. on Computer Vision. 2008. [doi: 10.1007/978-3-540-88690-7_24]

[7] Guillaumin M, Mensink T, Verbeek J, Schmid C. Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation. IEEE Int'l Conf. on Computer Vision, 2009,30(2):309–316. [doi: 10.1109/ICCV.2009.5459266]

[8] Lu J, Ma SP. Automatic image annotation based on concept indexing. Journal of Computer Research and Development, 2007,44(3):452–459 (in Chinese with English abstract). [doi: 10.1360/crad20070313]

[9] Qiu ZY, Fang Q, Sang JT, Xu CS. Regional context-aware image annotation. Chinese Journal of Computers, 2014,37(6):1390–1397 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2014.01390]

[10] Wang H, Huang H, Ding C. Image annotation using bi-relational graph of images and semantic labels. IEEE Conf. on Computer Vision and Pattern Recognition, 2011,42(7):793–800. [doi: 10.1109/CVPR.2011.5995379]

[11] Gao L, Song J, Nie F, Sebe N, Shen H. Optimal graph learning with partial tags and multiple features for image and video annotation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. [doi: 10.1109/CVPR.2015.7299066]

[12] Tian F, Shen XK. Image semantic annotation method for weakly labeled dataset. Ruan Jian Xue Bao/Journal of Software, 2013, 24(10):2405–2418 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4424.htm [doi: 10.3724/SP.J.1001.2013.04424]

[13] Amiri SH, Jamzad M. Efficient multi-modal fusion on supergraph for scalable image annotation. Pattern Recognition, 2015, 48(7):2241–2253. [doi: 10.1016/j.patcog.2015.01.015]

[14] Kalayeh MM, Idrees H, Shah M. NMF-KNN: Image annotation using weighted multi-view non-negative matrix factorization. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2014. 184–191. [doi: 10.1109/CVPR.2014.31]

[15] Tariq A, Foroosh H. Feature-Independent context estimation for automatic image annotation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. [doi: 10.1109/CVPR.2015.7298806]

[16] Erhan D, Bengio Y, Courville A, Vincent P. Why does unsupervised pretraining help deep learning. The Journal of Machine Learning Research, 2010,11(3):625–660.

[17] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science, 2006, 504–507. [doi: 10.1126/science.1127647]

[18] Hinton GE, Osindero S, Yw T. A fast learning algorithm for deep belief nets. Neural Computation, 2006,18(7):1527–1554. [doi: 10.1162/neco.2006.18.7.1527]

[19] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. 2012. http://xueshu.baidu.com/s?wd=ImageNet+classification+with+deep+convolutional+neural+networks&tn=SE_baiduxueshu_c1gjeupa&cl=3&ie=utf-8&bs=A+model+for+learning+the+semantics+of+pictures&f=8&rsv_bp=1&rsv_sug2=1&sc_f_para=sc_tasktype%3D%7BfirstSimpleSearch%7D&rsv_spt=3&rsv_n=2

[20] Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. The Journal of Machine Learning Research, 2010,11(12):3371–3408.

[21] Zhou B, Garcia AL, Xiao J, Torralba A, Oliva A. Learning deep features for scene recognition using places database. In: Advances in Neural Information Processing Systems. 2014. http://xueshu.baidu.com/s?wd=Learning+deep+features+for+scene+recognition+using+places+database&tn=SE_baiduxueshu_c1gjeupa&cl=3&ie=utf-8&bs=ImageNet+classification+with+deep+convolutional+neural+networks&f=8&rsv_bp=1&rsv_sug2=1&sc_f_para=sc_tasktype%3D%7BfirstSimpleSearch%7D&rsv_spt=3&rsv_n=2

[22] Pinheiro P, Collobert R. Recurrent convolutional neural networks for scene labeling. In: Proc. of the 31st Int'l Conf. on Machine Learning. 2014. 82–90. http://xueshu.baidu.com/s?wd=Recurrent+convolutional+neural+networks+for+scene+labeling&tn=SE_baiduxueshu_c1gjeupa&cl=3&ie=utf-8&bs=Learning+deep+features+for+scene+recognition+using+places+database&f=8&rsv_bp=1&rsv_sug2=1&sc_f_para=sc_tasktype%3D%7BfirstSimpleSearch%7D&rsv_spt=3&rsv_n=2

[23] Dan CC, Giusti A, Gambardella LM, Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In: Advances in Neural Information Processing Systems. 2012. 2852–2860. http://xueshu.baidu.com/s?wd=+Deep+neural+networks+segment+neuronal+membranes+in+electron+microscopy+images&tn=SE_baiduxueshu_c1gjeupa&cl=3&ie=utf-8&bs=Recurrent+convolutional+neural+networks+for+scene+labeling&f=8&rsv_bp=1&rsv_sug2=1&sc_f_para=sc_tasktype%3D%7BfirstSimpleSearch%7D&rsv_spt=3&rsv_n=2

[24] Luo P. Hierarchical face parsing via deep learning. IEEE Conf. on Computer Vision and Pattern Recognition, 2012,157(10): 2480–2487. [doi: 10.1109/CVPR.2012.6247963]

[25] Ryan K, Csaba S. Deep representations and codes for image auto-annotation. In: Advances in Neural Information Processing Systems. 2012. http://xueshu.baidu.com/s?wd=Deep+representations+and+codes+for+image+auto-annotation&tn=SE_baiduxueshu_c1gjeupa&cl=3&ie=utf-8&bs=Deep+neural+networks+segment+neuronal+membranes+in+electron+microscopy+images&f=8&rsv_bp=1&rsv_sug2=1&sc_f_para=sc_tasktype%3D%7BfirstSimpleSearch%7D&rsv_spt=3&rsv_n=2

[26] Wu JJ, Yu YN, Huang C, Yu K. Deep multiple instance learning for image classification and auto-annotation. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2015. [doi: 10.1109/CVPR.2015.7298968]

[27] Zhang S, Huang J, Huang Y, Yu Y, Li H. Automatic image annotation using group sparsity. In: Proc.of the IEEE Conf. on Computer Vision and Pattern Recognition. 2010. 3312–3319. [doi: 10.1109/CVPR.2010.5540036]

[28] Hariharan B, Zelnik-Manor L, Varma M, Vishwanathan SVN. Large scale max-margin multi-label classification with priors. In: Proc. of the 27th Int'l Conf. on Machine Learning. 2010. 423–430. http://xueshu.baidu.com/s?wd=Large+scale+max-margin+multi-label+classification+with+priors&tn=SE_baiduxueshu_c1gjeupa&cl=3&ie=utf-8&bs=Deep+representations+and+codes+for+image+auto-annotation&f=8&rsv_bp=1&rsv_sug2=1&sc_f_para=sc_tasktype%3D%7BfirstSimpleSearch%7D&rsv_spt=3&rsv_n=2

[29] Cui C, Ma J, Lian T, Wang X, Ren Z. Ranking-Oriented nearest-neighbor based method for automatic image annotation. In: Proc. of the 36th Int'l ACM SIGIR Conf. on Research and Development In Information Retrieval. 2013. 957–960. http://xueshu.baidu.com/s?wd=Ranking-Oriented+nearest-neighbor+based+method+for+automatic+image+annotation&tn=SE_baiduxueshu_c1gjeupa&cl=3&ie=utf-8&bs=Large+scale+max-margin+multi-label+classification+with+priors&f=8&rsv_bp=1&rsv_sug2=1&sc_f_para=sc_tasktype%3D%7BfirstSimpleSearch%7D&rsv_spt=3&rsv_n=2

[30] Wang Y, Dawood H, Yin Q, Guo P. A comparative study of different feature mapping methods for image annotation. In: Proc. of the Int'l Conf. on Advanced Computational Intelligence. 2015. 340–344. [doi: 10.1109/ICACI.2015.7184726]

[31] Kuric E, Bielikova M. ANNOR: Efficient image annotation based on combining local and global features. Computers and Graphics, 2015,47:1–15. [doi: 10.1016/j.cag.2014.09.035]

[32] Zhang X, Liu C. Image annotation based on feature fusion and semantic similarity. Neurocomputing, 2015,149:1658–1671. [doi: 10.1016/j.neucom.2014.08.027]

[33] Li Z, Liu J, Xu C, Lu H. MLRank: Multi-Correlation learning to rank for image annotation. Pattern Recognition, 2013,46(10): 2700–2710. [doi: 10.1016/j.patcog.2013.03.016]

附中文参考文献:

[3] 王梅,周向东,张军旗,许红涛,施伯乐.基于扩展生成语言模型的图像自动标注方法.软件学报,2008,19(9):2449–2460. http://www.jos.org.cn/1000-9825/19/2449.htm [doi: 10.3724/SP.J.1001.2008.02449]

[8] 路晶,马少平.基于概念索引的图像自动标注.计算机研究与发展,2007,44(3):452–459. [doi: 10.1360/crad20070313]

[9] 邱泽宇,方全,桑基韬,徐常胜.基于区域上下文感知的图像标注.计算机学报,2014,37(6):1390–1397. [doi: 10.3724/SP.J.1016.2014.01390]

[12] 田枫,沈旭昆.一种适合弱标签数据集的图像语义标注方法.软件学报,2013,24(10):2405–2418. http://www.jos.org.cn/1000-9825/4424.htm [doi: 10.3724/SP.J.1001.2013.04424]

周铭柯(1990－),男,福建三明人,硕士,主要研究领域为模式识别,计算机视觉,深度学习,机器学习.

杜明智(1988－),男,硕士,主要研究领域为图像处理,人工智能.

柯逍(1983－),男,博士,讲师,主要研究领域为模式识别,计算机视觉,图像处理,机器学习.