基于动态图卷积的图像情感分布预测

苏育挺[1,2]，王 骥[2]，赵 玮[1]，井佩光[1,2]

（1. 天津大学 电气自动化与信息工程学院，天津 300072；2. 天津大学 国际工程师学院，天津 300072）

摘 要：针对图像情感分布学习中，视觉特征与高阶情感语义之间存在语义鸿沟以及情感标签具有主观性和模糊性的问题，提出了一种情感语义动态图卷积网络模型。该模型通过情感激活模块自动定位情感语义区域，从而有效挖掘契合情感语义的内容表征；通过动态图卷积模块自适应地捕获图像情感标签之间的语义关联性；最终构建并行结构输出联合局部语义和标签相关性的情感预测分布。在3个公开情感数据集上的实验结果证明了本文算法在图像情感分布预测任务中的有效性。

关键词：信息处理技术；视觉情感计算；动态图卷积；标签分布式学习

中图分类号：TP391 **文献标志码**：A **文章编号**：1671-5497(2023)09-2601-10

DOI：10.13229/j.cnki.jdxbgxb.20211169

Dynamic graph convolutional neural network for image sentiment distribution prediction

SU Yu-ting[1,2], WANG Ji[2], ZHAO Wei[1], JING Pei-guang[1,2]

(1. *School of Electrical and Information Engineering, Tianjin University, Tianjin* 300072, *China*; 2. *Tianjin International Engineering Institute, Tianjin University, Tianjin* 300072, *China*)

Abstract: Aiming at the problem that there exists semantic gap between visual features and high-level emotional semantics and the subjectivity and ambiguity of emotional labels in image sentiment distribution learning, this paper proposes an Emotional Semantic Dynamic Graph Convolution Network (ESDGCN). In this framework, the Emotion Activation Module (EAM) is constructed to automatically locate the emotional semantic regions to effectively mine the content representation that fits the emotional semantics. In addition, the Semantic Dynamic Graph Convolution Network (SDGCN) is to adaptively capture the semantic relevance between labels. Finally, we adopt the parallel structure to jointly consider local semantic emotional information and label correlations. Experimental results on three open emotional datasets demonstrate the effectiveness of the proposed method.

Key words: information processing technology; visual sentiment computing; dynamic graph convolution; label distribution learning

收稿日期：2021-11-08.

基金项目：国家自然科学基金项目(61802277).

作者简介：苏育挺(1972-)，男，教授，博士. 研究方向：多媒体信息处理，多媒体信息安全，图像视频压缩编码. E-mail：ytsu@tju.edu.cn

通信作者：井佩光(1988-)，男，副教授，博士. 研究方向：短视频语义分析及理解，跨媒体智能计算，机器学习. E-mail：pgjing@tju.edu.cn

0 引 言

近年来,用户逐渐倾向于在社交网络中使用图像表达自己的情感,视觉情感计算受到了广泛关注。作为计算机视觉语义认知领域中的重要分支,视觉情感计算在面部情感识别[1-4]、多模态情感预测[5-7]、多媒体信息检索[8,9]、图像情感标注[10,11]等多个领域具有广阔的应用前景。

与时尚分析[12,13]、图像记忆度预测等任务类似,图像情感分析也具有一定程度的主观性和模糊性,因此,Geng[14]提出使用标签分布式学习(Label distribution learning,LDL)构建从实例到连续标签分布之间的映射关系。区别于传统单标签学习和多标签学习,标签分布式学习可以描述样本实例各个标签的重要程度。近年来,标签分布式学习在视觉情感计算领域中受到了越来越多的关注[15-20]。由Plutchilk情感轮[21]等心理研究可知,不同情感之间是彼此关联的。基于此类问题,诸多学者借助低秩思想挖掘标签之间的隐性相关性[22,23]或借助聚类方法、高斯图模型等构建显式关联结构[24-26]挖掘标签分布中的语义关联性。

虽然低秩算法在挖掘情感间的语义关联性层面取得了较为突出的进展,但由于其并非端到端学习,所以预测性能的优劣很大程度上取决于提取特征的优劣。随后,部分学者将循环神经网络(Recurrent neural network, RNN)、卷积神经网络(Convolutional neural network,CNN)等深度学习网络框架应用于视觉情感分析中[27-34],其可以有效提升视觉特征鲁棒性和稳定性,但在挖掘高阶情感语义关联性方面仍存在欠缺。近年来,图卷积网络(Graph convolutional network,GCN)在语义关系建模上取得了较大进展并成功应用于自然语言处理、计算机视觉等多个领域[35-37]。

尽管上述网络在图像分类任务中取得了显著效果,但上述图卷积网络中的图邻接矩阵均是基于标签共现性等先验信息统计得到的,故嵌入的图关联信息是全局静态的,这会导致局部与局部图像之间存在关联性不一致,即频率偏差问题。除此之外,由于图邻接矩阵为全局静态,随着网络层数的不断加深易出现过平滑现象。

为了解决上述问题,本文提出情感语义动态图卷积网络模型(Emotional semantic dynamic convolutional network,ESDGCN)用于预测图像情感分布。该模型使用动态图卷积对情感标签之间的语义关联性进行建模,其可以依据每张情感图像自适应地挖掘情感语义标签之间的特定依赖关系,从而有效捕获情感图像的局部相关性,并有效避免了静态图卷积引起的频率偏差、自适应性差和过平滑问题。为了更好地构建动态图卷积中的图节点向量,本文引入情感激活模块(Emotion activation module,EAM),利用弱监督检测定位情感激活区域并生成对应的类别语义特征从而引导图关系学习;为进一步增强动态图卷积中的语义关联性,搭建语义注意力机制(Semantic attention module,SAM)使得语义相关性较高的图节点标签被重点关注,从而强化图邻接矩阵的学习。

1 算法模型

1.1 整体网络结构

为了解决上述问题,本文使用动态图对情感标签之间的语义关系进行建模,其主要涉及两个主要模块:情感激活模块和基于语义注意力机制的动态图卷积模块(Semantic dynamic graph convolutional network,SDGCN)。其中,情感激活模块EAM使用弱监督检测的方式定位各个类别的情感区域并生成对应的情感语义特征;基于语义注意力机制的动态图卷积模块SDGCN则用于对情感语义特征之间的相关关系进行建模,并自适应地捕获特定图像情感标签之间的局部依赖性,本文网络结构如图1所示。

1.2 情感激活模块EAM

情感激活模块EAM主要用于感知和定位各个情感类别的激活区域,从而生成一系列类别向量并作为节点特征输入到后期的SDGCN模块进行情感分布预测。

对于包含K个示例的情感数据集$\{\boldsymbol{I}_j,\boldsymbol{D}_j\}$,$\boldsymbol{I}_j$为第$j$张情感图像,$\boldsymbol{D}_j\in\boldsymbol{R}^{1\times C}$为对应图像的$C$维情感分布分数。情感分布式学习通过挖掘情感图像中的特征信息预测得到C维情感分布分数。区别于one-hot或者multi-hot类型的离散概率分布,情感分布分数$\boldsymbol{D}_j$是累加和为1的情感隶属度向量,其每个值不仅可以判断各个标签能否描述该情感图像,同时还能指出各个标签对于描述情感图像的相对重要程度。

为了推断情感激活区域,本文使用弱监督检测方法生成无需定位信息的软激活图。具体而言,$\boldsymbol{X}\in\boldsymbol{R}^{H\times W\times N}$为从基础网络中提取得到的特征

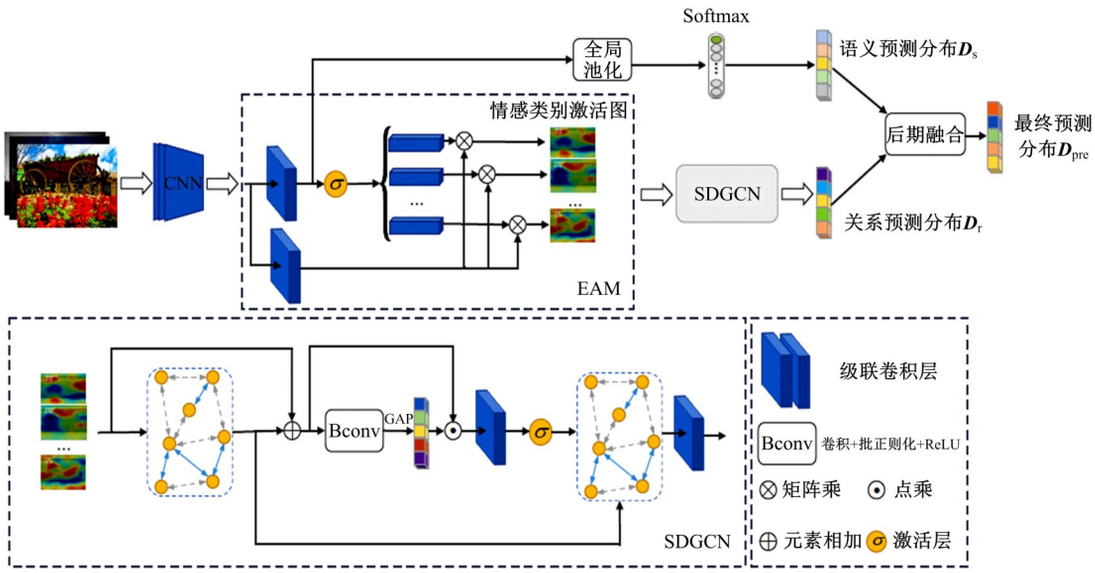

图 1　基于动态图卷积的图像情感分布预测模型

Fig. 1　Emotional dynamic graph convolution network for image sentiment distribution prediction

图输入,其中H、W、N分别对应特征图的高度、宽度和通道数。随后EAM将卷积滤波器作为响应检测器捕获特定情感类别的激活图$\boldsymbol{M}=[\boldsymbol{M}_1,\boldsymbol{M}_2,\cdots,\boldsymbol{M}_c]\in\boldsymbol{R}^{H\times W\times C}$。这些情感激活图与维度变换后的特征$\boldsymbol{X}'\in\boldsymbol{R}^{H\times W\times N'}$耦合得到情感类别响应矩阵$\boldsymbol{Z}=[\boldsymbol{Z}_1,\boldsymbol{Z}_2,\cdots,\boldsymbol{Z}_c]\in\boldsymbol{R}^{C\times N'}$,其中$\boldsymbol{Z}_c\in\boldsymbol{R}^{1\times N'}$为每个情感类别的响应向量,可表示为:

$$\boldsymbol{Z}_c=\boldsymbol{M}_c\boldsymbol{X}'=\sum_{i=1}^{H}\sum_{k=1}^{W}\boldsymbol{M}_c^{(i,k)}\boldsymbol{X}'^{(i,k)} \tag{1}$$

类激活映射(Class activation mapping,CAM)借助全连接层的权重对特征图进行加权重叠生成类激活特征图,其可以显式地表征分类决策主要来自于哪一些类特征图[38]。区分于CAM模块使用全局平均池化后的全连接层权重与特征图进行耦合得到类响应图,本文使用卷积滤波器输出代替CAM的全连接层权重,并将滤波输出经过激活函数Sigmoid正则化处理后通过上行网络的全局空间池化(Global spatial pooling,GSP)和归一化激活函数Softmax得到上行语义激活网络的情感语义预测分布$\boldsymbol{D}_s^{(j)}\in\boldsymbol{R}^{1\times C}$,表示为:

$$\boldsymbol{D}_s^{(j)}=\delta(f_{GSP}(\boldsymbol{Z}^{(j)})) \tag{2}$$

式中:$\boldsymbol{Z}^{(j)}\in\boldsymbol{R}^{C\times N'}$为第$j$张图像EAM模块输出的情感类别激活输出;$\boldsymbol{D}_s^{(j)}$为对应第$j$张情感图像的情感语义预测分布,其为一系列累加和为1的连续情感得分值。

1.3　基于语义注意力机制的动态图卷积SDGCN

利用情感激活模块EAM学习得到的情感类别响应$\boldsymbol{Z}$,本文进一步引入基于语义注意力机制的动态图卷积SDGCN挖掘不同情感语义标签之间的相关关系从而引导情感分布式学习。

不同于大多数基于静态图卷积的多标签图像识别模型[31-33],这里使用SDGCN自适应地挖掘特定图像的语义关联性。静态图卷积模型针对所有样例生成全局统一的图关联矩阵,且大多数图关联矩阵是通过标签共现频率等全局先验信息获得的,从而缺乏自适应性和对局部相关性的关注,随着网络深度的不断增大,静态图卷积网络通常会出现过平滑的现象。而动态图卷积针对每张图像都能自适应地生成特定的图关联矩阵,能很好地捕获情感图像的局部相关性。

对应一个动态图$G=(V,E)$而言,V和E分别为节点集和边集,动态图卷积网络的输入主要分为两部分:节点特征矩阵$\boldsymbol{Z}$和动态图邻接矩阵$\boldsymbol{A}_d$。在本文中,每个节点对应1种情感,$\boldsymbol{Z}_i^{(j)}\in\boldsymbol{R}^{1\times N'}$,$i=1,2,\cdots,C$为对应第$j$张图像第$i$个节点的情感类别向量,由EAM模块弱监督学习定位得到,N'为节点特征向量的维度;$\boldsymbol{A}_d^{(j)}\in\boldsymbol{R}^{C\times C}$为动态图邻接矩阵,用于挖掘第$j$张图像的相关关系。每个动态图卷积层可以描述为:

$$\boldsymbol{H}^{(l,j)}=\delta(\boldsymbol{A}_d^{(l,j)}\boldsymbol{H}^{(l-1,j)}\boldsymbol{W}_d^{(l,j)}) \tag{3}$$

式中:$\boldsymbol{H}^{(l,j)}\in\boldsymbol{R}^{C\times N_l}$,$l\in\{0,1,2\}$对应第$j$张情感图像通过第$l$层SDGCN网络的语义输出;

$N_l, l\in\{0,1,2\}$为第l层SDGCN网络中各个节点特征的维度;$\delta(\cdot)$为非线性激活函数,在这里使用LeakyReLU函数;$\boldsymbol{W}_{\mathrm{d}}^{(l,j)}\in\boldsymbol{R}^{N_{l-1}\times N_l}$为第$j$张图像在第$l$层SDGCN的状态更新矩阵,其和$\boldsymbol{A}_{\mathrm{d}}^{(0,j)}$均为随机初始化得到并在训练过程中不断更新。在动态图卷积学习的过程中,图邻接矩阵$\boldsymbol{A}_{\mathrm{d}}^{(l,j)}$将情感语义相关信息$\boldsymbol{Z}^{(l,j)}$扩散到所有节点,接收到信息的各个情感节点通过线性变换$\boldsymbol{W}_{\mathrm{d}}^{(l,j)}$不断更新节点向量的状态$\boldsymbol{H}^{(l,j)}$。

为了让语义相关性更高的情感特征被重点关注,这里引入语义注意力机制SAM强化图邻接矩阵$\boldsymbol{A}_{\mathrm{d}}$的学习。首先,利用残差思想将$\boldsymbol{Z}^{(j)}$和$\boldsymbol{H}^{(1,j)}$叠加保留更为丰富的相关信息;随后,将叠加后特征通过BCONV模块(由卷积层、批处理层和非线性激活层级联组成的卷积块)用于捕获更为稳定的相关信息,BCONV模块通过在卷积层后级联批正则化BN层和激活层从而保持训练过程中语义关联的稳定性;最后,将训练得到的稳定语义关联特征通过全局平均池化操作将全局空间信息压缩为情感通道向量$\boldsymbol{u}^{(j)}\in\boldsymbol{R}^{C\times 1}$,其具体可以表示为:

$$\boldsymbol{u}_c^{(j)}=f_{\mathrm{GAP}}(f_{\mathrm{Bconv}}(\boldsymbol{Z}_c^{(j)},\boldsymbol{H}_c^{(1,j)}))=\frac{1}{N_1}\sum_{i=1}^{N_1}\sigma(BN(f_{\mathrm{conv}}(\boldsymbol{Z}_{c,i}^{(j)},\boldsymbol{H}_{c,i}^{(1,j)}))) \tag{4}$$

式中:$\boldsymbol{u}_c^{(j)}$为第j张情感图像对应的权重向量的第c个元素值;$f_{\mathrm{GAP}}(\cdot)$和$f_{\mathrm{Bconv}}(\cdot)$分别对应全局平均池化层和BCONV模块;$\sigma(\cdot)$和$BN(\cdot)$分别对应BCONV模块中的非线性激活层和批正则化操作,这里使用ReLU作为激活层。

随后,本文将学习得到的$\boldsymbol{u}$作为权重对原始节点特征进行加权点乘,从而使得$\boldsymbol{u}$能针对不同实例的图节点特征$\boldsymbol{H}$而动态地调整,从而进一步强化了不同情感图节点之间的局部相关关系。本文进一步利用卷积层生成动态邻接矩阵$\boldsymbol{A}_{\mathrm{d}}$并归一化处理,具体如下所示:

$$\boldsymbol{A}_{\mathrm{d}}^{(j)}=\sigma(f_{\mathrm{conv}}(\boldsymbol{u}^{(j)}\odot(\boldsymbol{Z}^{(j)}+\boldsymbol{H}^{(1,j)}))) \tag{5}$$

式中:$\boldsymbol{A}_{\mathrm{d}}^{(j)}\in\boldsymbol{R}^{C\times C}$为学习得到的动态图邻接矩阵;$\odot$为哈达玛积,表示对应元素点乘;$f_{\mathrm{conv}}(\cdot):\boldsymbol{R}^{C\times N_1}\rightarrow\boldsymbol{R}^{C\times C}$为卷积层,用于维度转换;$\sigma(\cdot)$为激活函数,用作归一化处理,其选用Sigmoid函数。

学习得到的图邻接矩阵$\boldsymbol{A}_{\mathrm{d}}$会引导那些强相关的情感语义特征在下层GCN网络中被重点关注,从而得到嵌入局部情感语义关系的自适应特征$\boldsymbol{H}^{(l)}$,随后将自适应特征$\boldsymbol{H}^{(l)}$依次通过分类器得到最终的情感关系预测分布$\boldsymbol{D}_{\mathrm{r}}$,具体如下所示:

$$\boldsymbol{D}_{\mathrm{r}}^{(j)}=\delta(f_{\mathrm{cls}}(\boldsymbol{H}^{(2,j)})) \tag{6}$$

式中:$f_{\mathrm{cls}}(\cdot):\boldsymbol{R}^{C\times N_2}\rightarrow\boldsymbol{R}^{C\times 1}$为分类器,这里使用卷积层作为分类函数;$\delta(\cdot)$为归一化激活函数。

1.4 后期融合和情感分布预测

整体而言,本文借助情感激活模块EAM感知定位情感激活区域,并通过上行语义激活网络得到情感语义预测分布$\boldsymbol{D}_{\mathrm{s}}$;借助基于语义注意力机制的动态图卷积模块SDGCN自适应捕获特定图像的情感关联性,并通过下行网络得到情感关系预测分布$\boldsymbol{D}_{\mathrm{r}}$。随后为了综合考虑情感语义和情感局部关联性的影响,本模型将语义预测分布和关系预测分布进行后期融合,得到最终的情感预测分布,其具体可以表示为:

$$\boldsymbol{D}_{\mathrm{pre}}^{(j)}=\lambda\boldsymbol{D}_{\mathrm{s}}^{(j)}+(1-\lambda)\boldsymbol{D}_{\mathrm{r}}^{(j)} \tag{7}$$

式中:λ为情感关系分布和情感语义分布之间的平衡参数。

在训练过程中,本文使用MSE损失衡量情感预测分布与真实分布之间的误差情况,其具体可以表示为:

$$\mathrm{MSE}(\boldsymbol{D}_{\mathrm{pre}},\boldsymbol{D}_{\mathrm{real}})=\frac{1}{K}\sum_{j=1}^{K}\sum_{i=1}^{C}(\boldsymbol{D}_{\mathrm{real}}^{(i,j)}-\boldsymbol{D}_{\mathrm{pre}}^{(i,j)})^2 \tag{8}$$

式中:K、C分别对应情感图像个数和情感标签维度。

2 实验和结果分析

2.1 实验数据集和基本设置

为了有效评估本文提出的基于动态图神经网络的情感分布式学习模型,本文分别在3个公开情感数据集上进行实验,包括Flickr-LDL[39]、Twitter-LDL[39]和Emotion6[40]。其中Flickr-LDL和Twitter-LDL是南开大学视觉情感计算实验室从已有的数据集和网络上抽取的大规模情数据集,其分别包含10 700和10 045张真实场景图像,并采用8类经典情感标签进行标注。Emotion6则采用7种情感标签进行标注,主要包含6种基础情感(生气、恶心、愉悦、害怕、悲伤、吃惊)和中性情感,共包含1980张情感图像。

为了评估ESDGCN的性能,本文使用ResNet-101作为本次实验的Backbone网络,其中上下行网络最后一层全连接层或卷积层的输出维度设置为类别数,用于输出预测的情感分布分数。在训练过程中,将3个数据集随机切分成80%的

训练集和20%的测试集,并将EAM通道数设置为1024,将SDGCN模块的节点特征维度设置为1024。在训练过程中,为了防止过拟合的现象,本文实验采用如下数据增强方法:①对输入图像进行随机裁剪,其尺寸大小设置为448×448;②对图像进行随机水平翻转。除此之外,此次实验选择SGD作为训练优化器,并设置动量为0.9,正则化项权值衰减为10^{-5}。为了保证模型能够快速收敛,本文使用在ImageNet数据集的训练参数作为本次实验的预训练模型,并针对Backbone网络和ESDGCN模型分别设置初始学习率为0.1和0.01。为了保证模型能够完全且有效收敛,本文实验针对不同数据集设置不同的训练批次,其中针对大规模数据集Flickr-LDL和Twitter-LDL,训练批次设置为100;而针对较小规模的数据集Emotion6,设置训练批次为50。为了防止模型由于初始学习率设置不当陷入局部最优的情况,本文使用ReduceLROnPlateau学习率调整机制不断调整训练阶段的学习率,其在损失不再下降或者LDL指标趋于平稳的情况下按一定比率降低学习率,同时设置等待时间监视指标是否有所改善。

2.2　评价标准

针对视觉情感标签分布预测任务而言,如何有效衡量和评价预测分布与真实分布的相似情况是十分重要的。当前有很多指标用于衡量两个分布之间的相似度或距离,本文分别从不同指标族中抽取6种互斥的评价指标综合评价本模型的表现情况,其分别为Kullback-Leibler散度(KL)、Chebyshev距离(Cheb)、Sorensendist距离(Soren)、SquaredChord距离(SqC)、Cosine相似度(Cos)和Intersection相似度(Inter)。其中KL、Cheb、Soren和SqC用于衡量分布之间的距离,其值越小代表预测精度越高;Cos和Inter用于衡量分布之间的相似度,其值越大代表预测精度越高。

2.3　收敛性实验

为了验证本文算法ESDGCN的收敛性能,分别从损失函数和分布预测性能指标两个角度评估本文算法的收敛性和有效性。如图2所示,ESDGCN算法在Flickr-LDL数据集上训练前期损失急剧下降,并在70次训练迭代后出现新一轮下降,这主要是由于实验设置的学习率下降为原有的1/10导致的;最后在100次训练迭代后逐渐平稳,体现了较快的收敛速度和较优的稳定性能。

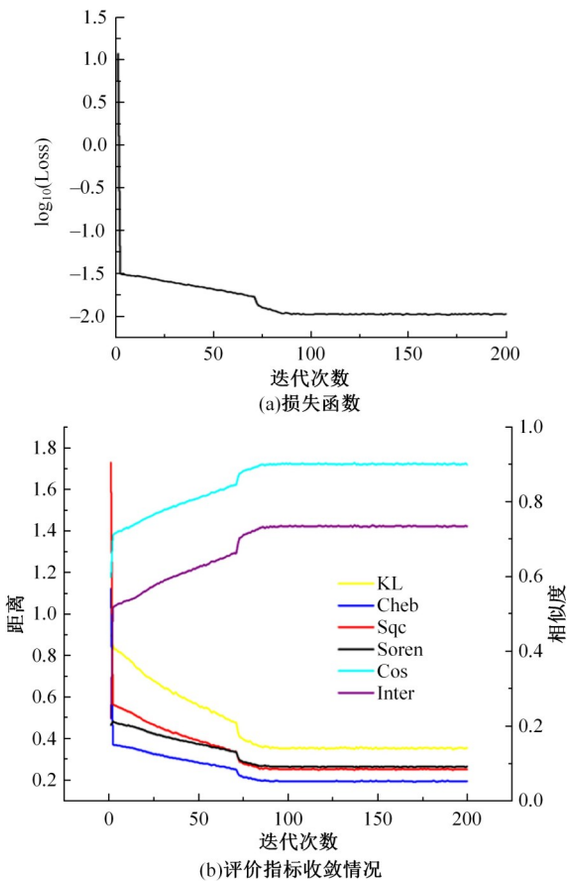

图2　ESDGCN损失函数和各个评价指标收敛情况(Flickr-LDL)

Fig. 2　Loss curve and convergence curve of evaluation indexes for Flickr-LDL dataset

类似地,图2(b)展示了ESDGCN算法不同预测指标随迭代次数的变化情况,距离和相似指标均在100次迭代后达到平稳,具有较好的收敛性。

2.4　消融实验

为了验证各个模块和分支的有效性,以Flickr-LDL情感数据集为例,本次实验在固定其他条件不变的情况下,分别在Backbone分支(B)、仅有EAM分支(B+E)、仅有动态图卷积DGCN分支(B+D)、包含相关注意力机制SAM的DGCN分支(B+D+S)、双路分支融合(B+E+D+S)的情况下进行实验,结果如表1所示。由

表1　不同模块组合下的情感标签分布式学习性能对比

Table 1　Comparison of different combination of module

方法	KL	Cos	Inter	Cheb	SqC	Soren
B	0.467	0.804	0.602	0.355	0.581	0.392
B+E	0.553	0.816	0.640	0.293	0.375	0.367
B+D	0.482	0.820	0.659	0.276	0.404	0.366
B+D+S	0.428	0.843	0.669	0.251	0.347	0.331
B+E+D+S	0.369	0.847	0.705	0.249	0.338	0.327

表1可以看出:①EAM对基础网络的预测性能有明显提升作用,表明情感区域弱监督学习可以在不增加标注负担的情况下提供更为丰富、精准的情感语义信息提升预测精度;②SDGCN分支的性能优于EAM分支,表明挖掘情感标签之间的语义关联性相比于挖掘精准鲁棒的语义特征具有更为重要的作用,SDGCN在EAM模块输出的情感语义响应向量的基础上进一步动态地挖掘特定图像的语义相关性,提升了情感预测精度;③DGCN+SAM分支的性能优于DGCN的性能,证明语义注意力机制能够进一步强化图邻接矩阵的学习;④分支融合后的性能优于各个模块单独作用的情况,证明同时挖掘语义区域和情感语义相关性对于视觉情感标签分布预测任务有显著作用。

2.5 敏感性实验

本节对本文提出的算法模型ESDGCN进行了敏感性实验。此次实验主要研究上下支路比例、图卷积网络层数和图卷积图节点特征维度对于视觉情感预测效果的影响。

表2为本文模型ESDGCN在Flickr-LDL数据集上取不同支路比例因子下的情感预测情况。从表2可以看出,当比例因子为0.5时,视觉情感标签分布预测性能最优,过大或者过小的比例因子会导致情感语义相关性缺乏或情感特征不丰富等问题,从而使得情感标签分布预测性能降低。

表2 不同上下支路比例对情感标签分布式学习性能的影响

Table 2 Comparison of different branch ratios

Ratios	KL	Cos	Inter	Cheb	SqC	Soren
0.0	0.427	0.828	0.666	0.27	0.347	0.353
0.1	0.488	0.811	0.650	0.278	0.390	0.368
0.3	0.575	0.795	0.625	0.292	0.398	0.381
0.5	0.369	0.847	0.705	0.249	0.338	0.327
0.7	0.570	0.784	0.626	0.302	0.448	0.396
0.9	0.450	0.834	0.664	0.262	0.332	0.342
1.0	0.553	0.816	0.640	0.293	0.375	0.367

表3为本文模型ESDGCN在Flickr-LDL数据集上设置不同图卷积层个数的情感预测分数情况。从表3可以看出,两层动态图卷积下的ESDGCN在Flickr-LDL数据集上表现性能最佳,单层图卷积和3层图卷积对应模型参数量与Flickr-LDL数据集实际样本量不匹配,分别存在欠拟合和过拟合的问题。

表3 不同图卷积层数对情感分布式学习性能的影响

Table 3 Comparison of different number of GCN layers

层数	KL	Cos	Inter	Cheb	SqC	Soren
1	0.520	0.819	0.651	0.273	0.343	0.349
2	0.369	0.847	0.705	0.249	0.338	0.327
3	0.461	0.823	0.671	0.274	0.340	0.345

表4为本文模型ESDGCN在Flickr-LDL数据集上采用不同图结构的情感预测分布分数情况。静态图网络分支(T)的邻接矩阵由训练集数据一次构建,在模型训练过程中保持不变。从表4中可以看出:两层静态图级联网络结构固定,表现性能最差;而两层包含相关注意力机制SAM(S)的DGCN(D)级联参数较多导致模型过拟合;两层动态图卷积DGCN级联表现性能最好。图3展示了部分图像的预测结果。

表4 不同图卷积结构对情感分布式学习性能的影响

Table 4 Comparison of different GCN structure

图结构	KL	Cos	Inter	Cheb	SqC	Soren
T+T	0.549	0.801	0.626	0.290	0.366	0.374
D+D	0.436	0.818	0.666	0.284	0.335	0.355
S+S	0.504	0.801	0.660	0.285	0.402	0.368

表5为本文模型ESDGCN在Flickr-LDL数据集上不同图节点特征维度下的情感预测分数情况。从表5可以看出:低维度的节点特征难以稳定地挖掘各个情感标签的语义表征,当图节点特征维度为1024时,对应情感预测性能表现最佳;当图节点特征维度大于1024时,预测精度随着图节点特征维度的陡增而下降,故最终选择图节点特征维度为1024。

2.6 对比实验

为了证明本文模型在视觉情感标签分布预测任务中的有效性,将ESDGCN模型与近几年的先进算法进行对比,主要包括基于低秩结构的标签分布预测算法、基于卷积神经网络的算法。对于基于低秩结构的标签分布预测算法而言,本文使用在对应情感数据集上预训练得到的backbone网络提取视觉特征,并使用主成分分析(Principal component analysis,PCA)算法进行降维获得300维提取特征作为标签分布式学习(Label distribution learning,LDL)算法输入,对比的LDL算法包括AA-BP[14]、AA-KNN[14]、CPNN[16]、EDL-LRL[24]、LDLLC[41]、LDL-SCL[23]。对于卷积神经网络模型而言,本文依次对3种经典卷积神经网络模型使用ImageNet预训练并在情感数据集上进行微调获得情感预测结果,3种对应的经典卷积神经网络为AlexNet、VGGNet和

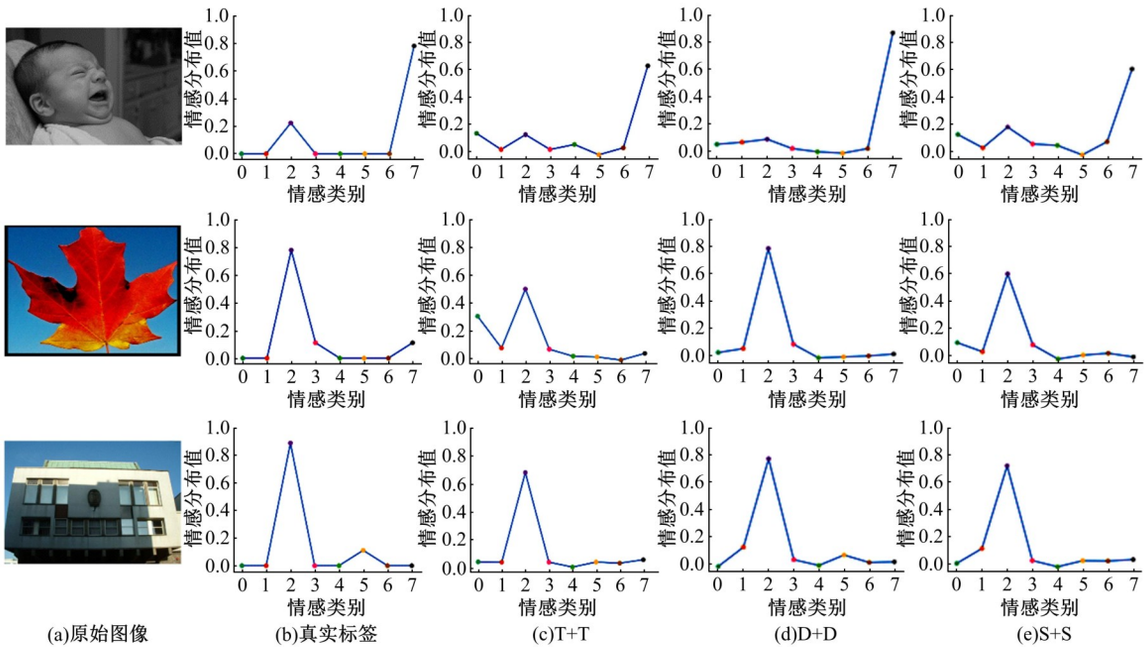

图 3　不同图卷积结构的预测结果

Fig. 3　Prediction results of different graph convolution structures

表 5　不同图节点特征维度对情感标签分布式学习性能的影响

Table 5　Comparison of different dimension of GCN layers

特征维度	KL	Cos	Inter	Cheb	SqC	Soren
256	0.474	0.840	0.674	0.255	0.311	0.326
512	0.506	0.818	0.649	0.274	0.359	0.357
1024	0.369	0.847	0.705	0.249	0.338	0.327
2048	0.511	0.827	0.659	0.265	0.337	0.341
4096	0.511	0.819	0.651	0.273	0.339	0.350

表 6　ESDGCN 算法同其他算法在 Flickr-LDL 数据集的对比

Table 6　Comparison of ESDGCN with others for Flickr-LDL

算法	KL	Cos	Inter	Cheb	SqC	Soren
AA-KNN[14]	0.737	0.777	0.599	0.308	0.447	0.401
CPNN[16]	1.001	0.695	0.538	0.353	0.555	0.462
EDL-LRL[24]	0.864	0.791	0.596	0.303	0.463	0.402
LDLLC[41]	0.785	0.768	0.570	0.329	0.503	0.430
LDL-SCL[23]	0.731	0.769	0.529	0.357	0.555	0.471
AlexNet	0.480	0.834	0.656	0.262	0.335	0.343
VGGNet	0.479	0.844	0.668	0.255	0.317	0.329
ResNet101	0.467	0.804	0.602	0.355	0.581	0.392
ACPNN*[17]	1.179	0.650	0.506	0.378	0.614	0.494
JCDL*[34]	0.528	0.837	0.676	0.266	0.292	0.348
SSDL*[42]	0.450	0.849	0.646	0.267	0.356	0.349
ESDGCN	0.369	0.847	0.705	0.249	0.338	0.327

ResNet101。除此之外,本文还对比了最新的3种基于CNN的情感分布预测模型,分别为ACPNN[17]、JCDL[34]和SSDL[42]。ESDGCN和各个对比算法在3个不同公开数据集上的具体实验结果如表6～表8所示,其中,"*"表示该行指标是笔者自行复现该算法,在相应数据集上评估得到的结果。

由表6可以看出:①本文ESDGCN模型从6个标签分布指标整体上看性能最佳;②在所有LDL算法中,CPNN和ACPNN性能表现较差,原因在于二者都未将标签关联性学习考虑在内;③在考虑标签关联性学习的算法中,EDL-LRL算法性能优于LDLLC,二者都对标签矩阵直接施加正则化约束,但是前者针对聚类后的标签组施加低秩约束,而后者针对全体标签施加低秩约束,可以看出挖掘标签间的局部相关性对提升模型精度有较大作用;④本文ESDGCN算法的性能优于其他挖掘局部相关性的LDL算法,证实针对特定图像自适应地生成图邻接矩阵的性能要优于通过聚类方式获取得到的组级别局部相关结构;⑤AlexNet、VGG-Net、ResNet101的整体性能优于基于低秩结构算法,证实深度学习在提取大规模情感数据集的特征信息上具有较优的表现,但三者性能差于本文算法,证实通过软检测的方式定位局部情感语义区域提取的高层级语义特征表现性能优于低层级视觉特征的表现;⑥JCDL和

表7 ESDGCN算法同其他算法在Twitter-LDL数据集的对比

Table 7 Comparison of ESDGCN with other for Twitter-LDL

方法	KL	Cos	Inter	Cheb	SqC	Soren
AA-KNN[14]	2.628	0.763	0.570	0.345	0.542	0.430
CPNN[16]	1.179	0.735	0.552	0.358	0.547	0.448
EDL-LRL[24]	2.837	0.525	0.376	0.504	0.855	0.623
LDLLC[41]	1.541	0.523	0.367	0.512	0.875	0.633
LDL-SCL[23]	1.034	0.515	0.430	0.577	1.447	0.664
AlexNet	0.489	0.855	0.679	0.251	0.316	0.320
VGGNet	0.501	0.869	0.676	0.249	0.306	0.334
ResNet101	0.522	0.830	0.649	0.274	0.340	0.351
ACPNN*[17]	1.502	0.642	0.481	0.413	0.678	0.519
JCDL*[34]	0.543	0.855	0.698	0.254	0.283	0.345
SSDL*[42]	0.514	0.859	0.685	0.253	0.291	0.339
ESDGCN	0.408	0.862	0.693	0.247	0.334	0.328

表8 ESDGCN算法同其他算法在Emotion6数据集的对比

Table 8 Comparison of ESDGCN with others for Emotion6

方法	KL	Cos	Inter	Cheb	SqC	Soren
AA-KNN[14]	0.708	0.602	0.538	0.353	0.356	0.462
CPNN[16]	0.564	0.685	0.569	0.331	0.295	0.431
EDL-LRL[24]	3.699	0.780	0.653	0.279	0.404	0.327
LDLLC[41]	0.424	0.796	0.664	0.247	0.210	0.336
LDL-SCL[23]	0.405	0.788	0.637	0.268	0.219	0.363
AlexNet	0.506	0.743	0.619	0.276	0.246	0.384
VGGNet	0.384	0.825	0.676	0.234	0.238	0.316
ResNet101	0.472	0.750	0.619	0.279	0.325	0.383
ACPNN*[17]	1.950	0.475	0.403	0.476	0.701	0.597
JCDL*[34]	0.438	0.805	0.668	0.251	0.260	0.325
SSDL*[42]	0.400	0.803	0.658	0.237	0.242	0.369
ESDGCN	0.286	0.835	0.725	0.228	0.260	0.307

SSDL性能整体较优但略逊于ESDGCN算法,证实引入图卷积网络进行标签关系建模对卷积神经网络情感分布预测精度有提升作用。

表7为ESDGCN算法和其他对比算法在Twitter-LDL数据集上的表现。从表7中可以看出:ESDGCN比其他方法获得了更好的预测性能,显示了其在大规模情感数据集的情感标签关系建模方面具有优势;基于卷积神经网络的深度学习算法整体优于基于低秩结构的LDL算法,原因在于神经网络在识别和提取大规模数据集的语义特征上具有较大优势。在基于深度卷积神经网络的各个对比算法中,SSDL在Twitter-LDL的表现性能较优但性能劣于ESDGCN,证实动态挖掘标签间相关结构对情感分布的预测精度有提升作用。

表8为ESDGCN算法和其他对比算法在Emotion6数据集上的表现。从表8中可以看出,本文ESDGCN方法获得了最好的预测性能。在所有LDL算法中,LDL-EDL、LDLLC和LDL-SCL的预测精度处于较高水平,这表明挖掘标签之间的关联结构对引导情感标签分布预测有较大的优势。然而三者的性能均差于本节提出的ESDGCN算法,这也证实借助动态图卷积进行标签关系建模相比于使用低秩结构挖掘标签关联性而言的性能更优。除此之外,在基于深度学习的对比算法中,VGGNet基础网络在Emotion6数据集上表现较好,但预测精度不及ESDGCN,这也证实引入图卷积网络进行标签关系建模通常能对经典卷积神经网络情感分布的预测精度有提升作用。

3 结束语

本文提出一种端到端的基于动态图卷积神经网络的图像情感分布预测算法ESDGCN,其主要由两个模块组成:情感激活模块EAM和基于语义注意力机制的动态图卷积网络模块SDGCN。EAM模块主要用于自动定位情感语义区域,从而挖掘契合情感语义的内容表征,SDGCN模块主要用于自适应地捕获特定情感图像标签之间的语义关联性并使用语义注意力机制SAM对其中强关联信息进行重点关注,最终通过两个模块的联动学习输出联合局部语义信息和标签相关性的情感预测分布。实验证实,使用EAM可以在不添加标签负担的情况下提供鲁棒性更强且丰富的语义内容表示;除此之外,SDGCN在EAM学习得到的标签内容表征基础之上进一步动态地挖掘情感标签的依赖关系并显著提升了模型的预测精度;本文模型在3个公开情感数据集上的实验结果表明,联合挖掘语义区域和情感相关性对情感分布预测任务而言具有显著优势。

参考文献:

[1] Zhou L, Fan X, Ma Y, et al. Uncertainty-aware cross-dataset facial expression recognition via regularized conditional alignment[C]//Proceedings of ACM International Conference on Multimedia, New York, USA, 2020: 2964-2972.

[2] Farzaneh A H, Qi X. Discriminant distribution-agnostic loss for facial expression recognition in the wild

[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Piscataway, USA, 2020: 406-407.

[3] 卢洋, 王世刚, 赵文婷, 等. 基于离散Shearlet类别可分性测度的人脸表情识别方法[J]. 吉林大学学报: 工学版, 2019, 49(5): 1715-1725.
Lu Yang, Wang Shi-gang, Zhao Wen-ting, et al. Facial expression recognition based on separability assessment of discrete Shearlet transform[J]. Journal of Jilin University (Engineering and Technology Edition), 2019, 49(5): 1715-1725.

[4] 方明, 陈文强. 结合残差网络及目标掩膜的人脸微表情识别[J]. 吉林大学学报: 工学版, 2021, 51(1): 303-313.
Fang Ming, Chen Wen-qiang. Face micro-expression recognition based on ResNet with object mask[J]. Journal of Jilin University (Engineering and Technology Edition), 2021, 51(1): 303-313.

[5] Huang F, Wei K, Weng J, et al. Attention-based modality-gated networks for image-text sentiment analysis[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2020, 16(3): 1-19.

[6] Ji R, Chen F, Cao L, et al. Cross-modality microblog sentiment prediction via bi-layer multimodal hypergraph learning[J]. IEEE Transactions on Multimedia, 2018, 21(4): 1062-1075.

[7] Jian M, Dong J, Gong M, et al. Learning the traditional art of Chinese calligraphy via three-dimensional reconstruction and assessment[J]. IEEE Transactions on Multimedia, 2019, 22(4): 970-979.

[8] Yang J, She D, Lai Y K, et al. Retrieving and classifying affective images via deep metric learning[C]// Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, USA, 2018: 491-498.

[9] Yao X, She D, Zhao S, et al. Attention-aware polarity sensitive embedding for affective image retrieval [C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, USA, 2019: 1140-1150.

[10] Li Z, Liu J, Zhu X, et al. Image annotation using multi-correlation probabilistic matrix factorization [C]//Proceedings of the ACM International Conference on Multimedia, New York, USA, 2010: 1187-1190.

[11] Li Z, Tang J, He X. Robust structured nonnegative matrix factorization for image representation[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 29(5): 1947-1960.

[12] Yang X, Song X, Feng F, et al. Attribute-wise explainable fashion compatibility modeling[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2021, 17(1): 1-21.

[13] Yang X, Song X, Han X, et al. Generative attribute manipulation scheme for flexible fashion search[C]// Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, USA, 2020: 941-950.

[14] Geng X. Label distribution learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(7): 1734-1748.

[15] Peng K C, Chen T, Sadovnik A, et al. A mixed bag of emotions: model, predict, and transfer emotion distributions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, USA, 2015: 860-868.

[16] Geng X, Yin C, Zhou Z H. Facial age estimation by learning from label distributions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(10): 2401-2412.

[17] Yang J, Sun M, Sun X. Label distribution learning via augmented conditional probability neural network [C]//Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, USA, 2017: 224-230.

[18] Zhou Y, Xue H, Geng X. Emotion distribution recognition from facial expressions[C]//Proceedings of the ACM International Conference on Multimedia, New York, USA, 2015: 1247-1250.

[19] Ren T, Jia X, Li W, et al. Label distribution learning with label-specific features[C]//Proceedings of the International Joint Conference on Artificial Intelligence, San Mateo, USA, 2019: 3318-3324.

[20] Zhao S, Yao H, Gao Y, et al. Continuous probability distribution prediction of image emotions via multitask shared sparse regression[J]. IEEE Transactions on Multimedia, 2016, 19(3): 632-645.

[21] Plutchik R. Emotions: a general psychoevolutionary theory[J]. Approaches to Emotion, 1984(1984): 197-219.

[22] Xu M, Zhou Z H. Incomplete label distribution learning[C]//Proceedings of the International Joint Conference on artificial intelligence, San Mateo, USA, 2017: 3175-3181.

[23] Jia X, Li Z, Zheng X, et al. Label distribution learning with label correlations on local samples[J]. IEEE Transactions on Knowledge and Data Engineering,

2019, 33(4): 1619-1631.

[24] Jia X, Zheng X, Li W, et al. Facial emotion distribution learning by exploiting low-rank label correlations locally[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Piscataway, USA, 2019: 9841-9850.

[25] Chen T, Yu F X, Chen J, et al. Object-based visual sentiment concept analysis and application[C]//Proceedings of the ACM International Conference on Multimedia, New York, USA, 2014: 367-376.

[26] Su Y T, Zhao W, Jing P G, et al. Exploiting low-rank latent gaussian graphical model estimation for visual sentiment distribution[J]. IEEE Transactions on Multimedia, 2022,25: 1243-1255.

[27] 缪裕青, 雷庆庆, 张万桢, 等. 多视觉目标融合的图像情感分析研究[J]. 计算机应用研究, 2021, 38(4): 1250-1255.
Miao Yu-qing, Lei Qing-qing, Zhang Wan-zhen, et al. Research on image sentiment analysis based on multi-visual object fusion[J] Application Research of Computers, 2021, 38(4): 1250-1255.

[28] 盛家川, 陈雅琦, 王君, 等. 深度学习结构优化的图像情感分类[J]. 红外与激光工程, 2020, 49(11): 264-273.
Sheng Jia-chuan, Chen Ya-qi, Wang Jun, et al. Image sentiment classification via deep learning structure optimization[J] Infrared and Laser Engineering, 2020, 49(11): 264-273.

[29] Chen T, Borth D, Darrell T, et al. Deepsentibank: visual sentiment concept classification with deep convolutional neural networks[J/OL]. [2021-10-25]. https://arxiv.org/pdf/1410.8586v1.pdf

[30] Zhu X, Li L, Zhang W, et al. Dependency exploitation: a unified CNN-RNN approach for visual emotion recognition[C]//Proceedings of the International Joint Conference on Artificial Intelligence, San Mateo, USA, 2017: 3595-3601.

[31] Campos V, Salvador A, Giró-i-Nieto X, et al. Diving deep into sentiment: understanding fine-tuned CNNs for visual sentiment prediction[C]//Proceedings of the 1st International Workshop on Affect & Sentiment in Multimedia, New York, USA, 2015: 57-62.

[32] Campos V, Jou B, Giro-i-Nieto X. From pixels to sentiment: fine-tuning CNNs for visual sentiment prediction[J]. Image and Vision Computing, 2017, 65(1): 15-22.

[33] You Q, Luo J, Jin H, et al. Robust image sentiment analysis using progressively trained and domain transferred deep networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, USA, 2015: 381-388.

[34] Yang J, She D, Sun M. Joint image emotion classification and distribution learning via deep convolutional neural network[C]//Proceedings of the International Joint Conference on Artificial Intelligence, San Mateo, USA, 2017: 3266-3272.

[35] 徐冰冰, 岑科廷, 黄俊杰, 等. 图卷积神经网络综述[J]. 计算机学报, 2020, 43(5): 755-780.
Xu Bing-bing, Cen Ke-yan, Huang Jun-jie, et al. A survey on graph convolutional neural network[J] Chinese Journal of Computers, 2020, 43(5): 755-780.

[36] Chen T, Xu M, Hui X, et al. Learning semantic-specific graph representation for multi-label image recognition[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, Piscataway, USA, 2019: 522-531.

[37] He T, Jin X. Image emotion distribution learning with graph convolutional networks[C]//Proceedings of the International Conference on Multimedia Retrieval, New York, USA, 2019: 382-390.

[38] Zhou B, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Los Alamitos, USA, 2016: 2921-2929.

[39] Yang J, Sun M, Sun X. Label distribution learning via augmented conditional probability neural network [C]//Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, USA, 2017: 224-230.

[40] Peng K C, Chen T, Sadovnik A, et al. A mixed bag of emotions: model, predict, and transfer emotion distributions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, USA, 2015: 860-868.

[41] Jia X, Li W, Liu J, et al. Label distribution learning by exploiting label correlations[C]//Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, USA, 2018: 3310-3317.

[42] Xiong H, Liu H, Zhong B, et al. Structured and sparse annotations for image emotion distribution learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, USA, 2019: 363-370.