计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.221100177

图像情感分析的层次图卷积网络模型

谈钱辉　温佳璇　唐继辉　孙玉宝
南京信息工程大学数字取证教育部工程研究中心　南京 210044
江苏省大数据分析技术重点实验室　南京 210044
(20201222018@nuist.edu.cn)

摘　要　图像情感分析任务旨在运用机器学习模型自动预测观测者对图像的情感反应。当前基于深度网络的情感分析方法广受关注,主要通过卷积神经网络自动学习图像的深度特征。然而,图像情感是图像全局上下文特征的综合反映,由于卷积核感受野的尺寸限制,无法有效捕捉远距离情感特征间的依赖关系,同时网络中不同层次的情感特征间未能得到有效的融合利用,影响了图像情感分析的准确性。为解决上述问题,文中提出了层次图卷积网络模型,分别在空间和通道维度上构建空间上下文图卷积(SCGCN)模块和动态融合图卷积(DFGCN)模块,有效学习不同层次情感特征内部的全局上下文关联与不同层级特征间的关系依赖,能够有效提升情感分类的准确度。网络结构由4个层级预测分支和1个融合预测分支组成,层级预测分支利用SCGCN学习单层次特征的情感上下文表达,融合预测分支利用DFGCN自适应聚合不同语义层次的上下文情感特征,实现融合推理与分类。在4个情感数据集上进行实验,结果表明,所提方法在情感极性分类和细粒度情感分类上的效果均优于现有的图像情感分类模型。

关键词:图像情感分析;图卷积;全局上下文关联;层次特征关联;融合分类
中图法分类号　TP37

Hierarchical Graph Convolutional Network for Image Sentiment Analysis

TAN Qianhui,WEN Jiaxuan,TANG Jihui and SUN Yubao
Digital Forensics Engineering Research Center of the Ministry of Education,Nanjing University of Information Science and Technology,Nanjing 210044,China
Jiangsu Big Data Analysis Technology Laboratory,Nanjing 210044,China

Abstract　The image sentiment analysis task aims to use machine learning models to automatically predict the observer's emotional response to images. At present,the sentiment analysis method based on the deep network has attracted wide attention,mainly through the automatic learning of the deep features of the image through the convolutional neural network. However,image emotion is a comprehensive reflection of the global contextual features of the image. Due to the limitation of the receptive field size of the convolution kernel,it is impossible to effectively capture the dependencies between long-distance emotional features. At the same time,the emotional features of different levels in the network cannot be effectively fused and utilized. It affects the accuracy of image sentiment analysis. In order to solve the above problems,this paper proposes a hierarchical graph convolutional network model,and constructs spatial context graph convolution(SCGCN) and dynamic fusion graph convolution(DFGCN). The spatial and channel dimensions are mapped respectively to learn the global context association within different levels of emotional features and the relationship dependence between different levels of features,which could improve the sentiment classification accuracy. The network is composed of four hierarchical prediction branches and one fusion prediction branch. The hierarchical prediction branch uses SCGCN to learn the emotion context expression of single-level features,and the fusion prediction branch uses DFGCN to self-adaptively aggregate the context emotion features of different semantic levels to realize fusion reasoning and classification. Experiment results on four emotion datasets show that the proposed method outperforms existing image emotion classification models in both emotion polarity classification and fine-grained emotion classification.

Keywords　Image sentiment analysis,Graph convolution,Global context association,Hierarchical feature association,Fusion classification

随着互联网的蓬勃发展,越来越多的用户在微博、微信、　　知乎等社交平台上发布图像以分享观点和情绪。心理学研究

到稿日期:2022-11-21　返修日期:2023-04-28
基金项目:国家重点研发计划(2022YFC2405600);国家自然科学基金(62276139,U2001211)
This work was supported by the National Key Research and Development Program of China(2022YFC2405600) and National Natural Science Foundation of China(62276139,U2001211).
通信作者:孙玉宝(sunyb@nuist.edu.cn)

表明，相比文字信息，直接的视觉激励更容易激发观察者的情绪，识别图像中蕴含的情感信息有利于判断用户的情感状态，具有多种潜在应用，如意见挖掘[1]、商业智能、娱乐辅助[2]、个性化情绪预测[3]等，拥有广泛的使用场景和巨大的商业价值，图像情感分析任务也因此受到了广泛关注。

早期的情感分析研究工作主要关注手工情感特征的设计。Lang 等[4]使用底层情感特征（如颜色、纹理、构图等）的组合来预测图像情感；Zhao 等[5]则根据图像的艺术风格设计出更具表现性的情感特征，以提升预测准确度；Borth 等[6]使用由 1 200 个语义概念分类器组成的 SentiBank 检测视觉内容中的情感。以上传统方法虽然取得了一定的图像情感分类效果，但手工设计的特征情感表达能力有限，而且往往需要复杂的特征提取过程。

随着卷积神经网络（CNN）[7]的发展，学者们开始研究基于 CNN 的图像情感分析模型[8-10]，利用卷积运算自动提取情感特征进行分类。现有基于深度学习的情感分析模型进一步利用注意力机制[11]挖掘图像中的显著情感特征，进而改善网络的情感分类性能，但在情感特征学习上仍存在两方面的不足。首先，图像情感是全局上下文特征的综合反映。如图 1 所示，湖泊、白云、小屋、绿树等属性都能表达一定的情感信息，但是，由于卷积核感受野固有的尺寸限制，卷积运算无法有效捕捉图像遥远区域间的特征依赖关系，不同区域情感属性间的上下文关联没有被充分捕捉。其次，除了高级视觉语义信息，场景图中所蕴含的纹理、色彩、饱和度等底层特征也可能诱发特定类别的情绪反应，现有的情感分类网络未能充分融合利用不同层次的情感特征。这两方面的不足影响了图像情感分析的准确性。

为了解决上述两方面的挑战，本文提出了图像情感分类的层次图卷积网络模型，由 4 个层级预测分支和 1 个融合预测分支组成。层级预测分支中设计了空间上下文图卷积（SCGCN）模块，通过在空间节点上构图，将图像不同区域之间的情感内容相似度作为节点的边权重，SCGCN 能够赋予网络捕获图像整体上下文表达的能力，弥补现有基于 CNN 的模型在建模远距离特征依赖上的不足。融合预测分支设计有动态融合图卷积模块（DFGCN），将通道特征转化为节点表示。为获得通道节点之间的情感联系，进一步提出了一个邻阶矩阵激活函数，得到多层次特征通道邻阶矩阵，通过在通道连通图上进行推理，DFGCN 能够动态聚合与目标预测情感类别相关的通道节点特征，实现不同层次情感特征的融合推理。在 FI，EmotionROI，Abstract 和 Artphoto 数据集上的大量实验都证明，本文层次图卷积分类网络的分类准确度优于现有基于 CNN 的图像情感分类模型。本文的主要贡献如下：

1）提出了一种层次分类架构，利用层级预测分支分别学习 4 个语义层次上的情感特征，并在融合预测分支中实现跨层次特征交互。相比基于 CNN 的单层次情感分类模型，本文的网络架构能显著提升分类性能。

2）提出了空间上下文图卷积（SCGCN），在每个层级聚合特征坐标空间中和目标情感相关的空间节点，并用分类器预测当前层级的情感类别，让模型专注于学习特定层次上的全局上下文情感特征。

3）提出了动态融合图卷积（DFGCN），在通道维度构图并进行推理，动态融合各层级学习到的情感上下文特征，弥补现有模型在底层和中级情感特征利用上的不足。

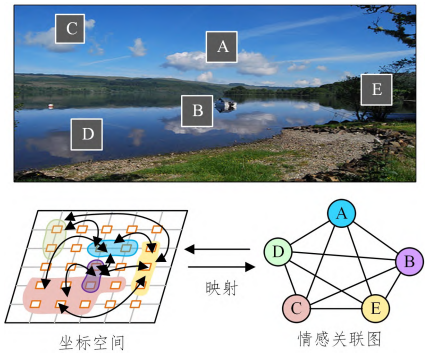


图 1 图像情感特性分析

Fig. 1 Analysis of image sentiment characteristics

1 相关工作

1.1 心理学情绪模型

心理学中有两种判断情感归属的模型，即情感维度空间（DES）模型和情感类别状态（CES）模型[2]，DES 模型用二维或三维笛卡尔空间中的连续坐标点来表示情感，而 CES 模型结合心理学的相关理论，将情感划分为不同的类别。就情感检索任务或者情感分类任务而言，以类别形式划分情感更便于使用且易于理解，因此本文的工作围绕 CES 模型展开。最基础的 CES 模型是一个粗粒度模型，即将情感划分成积极、消极两个大类。在此基础上，心理学的学者们又设计了一些相对细粒度的情感模型，其中最经典的是 Mikel 的 8 类别（愉悦、生气、敬畏、满足、厌恶、兴奋、恐惧、伤心）模型[12]。本文基于 Mikel 的 8 类别情绪轮盘模型，在粗、细两个粒度上进行情感图像的情感分类。

1.2 基于 CNN 的图像情感分类

最初，Chen 等[8]提出了 DeepSentiBank，利用形容词-名词对（ANP）和 CNN 构建了一个情感分类框架，证明基于卷积神经网络的分类模型的效果远优于支持向量机（SVM）。Yang 等提出了一个多任务学习框架[13]，联合优化情感分类和情感分布学习，提升了情感分类性能。Rao 等提出了 MldrNet[14]，将基于网络提取到的多层次情感特征按照通道维度拼接，用于情感分类，虽然利用了不同语义水平下的情感特征，但是层次特征融合方式较为简单，不能突出多层次特征在情感表达上的差异性。

最近的情感分类研究将重心放在局部情感表达上，其方法大致可以分为两类：基于硬性注意力机制和基于软性注意力机制的方法。基于硬性注意力机制的方法[15-17]通过目标检测网络发掘图像中物体的边界框，获得局部视觉特征，以此增强网络的情感识别性能，但忽略了复杂场景下不同语义特征之间的关联。为解决此问题，Zhang 等[18]提出了 OSSCM，利用语义情感关联模块计算 bounding box 提取到的图像局部区域相关性；Hu 等[19]提出了 GOSR，利用目标检测网络获得高级对象语义特征并用图结构表示复杂场景图中的对象

语义联系，最后通过图卷积优化对象信息，这与本文方法较为相关。但是上述方法需要预先使用 Faster-RCNN 等目标检测网络来提取图像中的具体前景对象，使得此类基于目标检测的情感分类网络只适用于描述真实世界的场景图，对于 Abstract 数据集中的抽象艺术图片，这类方法不能框出具体对象。此外，基于硬性注意力的情感分类方法大都利用基于网络的最后一层的高级情感特征来获取 bounding box，没有利用到色彩、纹理等与情感相关的底层特征。相比硬性注意力机制，软性注意力机制能够利用空间和通道注意力更精细地突出局部情感区域。基于通道注意力机制，Yang 等[11]提出了 WSCNet，用于计算不同通道特征的响应权重，并对原通道特征加权相加后生成情感注意力图，用热力图直观地突出情感表达强烈的区域，但是受制于 CNN 的感受野限制，WSCNet 更倾向于提取局部情感特征，不能捕捉远距离情感要素依赖。

1.3 图表示学习

近年来，在图表示学习方面的研究成果表明，图结构是关系推理最有效的方法之一。在早期的研究中，学者们就根据图网络提出条件随机场(CRF)[20]和随机游走网络[21]来进行有效的图像分割。近年来，已有大量的研究将卷积推广到基于图的数据中[22-25]。在这些方法中，图卷积网络(GCN)作为对切比雪夫展开一阶近似的简化模型，缓解了卷积算子对局部特征的过度关注。Wang 等[26]提出通过 GCN 捕获目标检测器所捕获的图像区域之间的关系，但是所提算法需要预先使用目标检测网络定位高级语义对象的边界框。Chen 等[27]提出了一个通用的可训练图推理模块，用于图像中不相交和远距离区域关系之间的推理，避免了目标检测过程导致的计算负担。利用 GCN 的推理能力，本文构建了图像情感分析的层次图卷积网络模型，分别在空间维度与特征通道维度进行构图，有效捕获不同层次特征的全局上下文和层级特征间的关联性进行情感分类。

2 本文算法

本文提出了一个端到端的层次化图卷积分类网络来预测图像情感。网络核心模块包括基干网络、空间上下文图卷积(SCGCN)和动态融合图卷积(DFGCN)。通过层次化的组合，构建的网络包括 4 个层次特征预测分支和 1 个融合预测分支组成。层次特征预测分支利用 SCGCN 专注于学习当前层级的全局情感上下文特征，在单一层级特征上进行情感分类；融合预测分支利用 DFGCN 动态融合层级预测分支学习到的层级特征，并将其用于最终的情感分类。为了更好地指导网络学习，本文设计了多层次联合损失函数，同时优化单层次特征学习和融合特征学习。下文将详细介绍网络的总体结构、SCGCN、DFGCN 以及损失函数设计。

2.1 层级图卷积分类网络框架

如图 2 所示，本文选用 ResNet101 作为主干网络，提取 4 个层级的情感特征$\{C_2,C_3,C_4,C_5\}$，其中 C_2 保留了更多的图像空间信息，其纹理、色彩等底层特征表达较强，但情感语义水平较低，C_5 虽然丢失了一定空间信息，但情感语义表达水平较高。本文遵循 FPN[28]的设计，自顶向下地融合相邻层级的层次特征，利用较高层级特征增强底层特征的情感语义表达。此外，本文在每个层级上使用 SCGCN 增强层次特征间层的上下文交互，得到优化后的层次特征$\{G_2,G_3,G_4,G_5\}$。计算过程如下：

$$\begin{cases} G_l=\mathrm{SCGCN}(\mathrm{GAP}(C_l)+G_{l+1}), & l=2,3,4 \\ G_l=\mathrm{SCGCN}(\mathrm{GAP}(C_l)), & l=5 \end{cases} \tag{1}$$

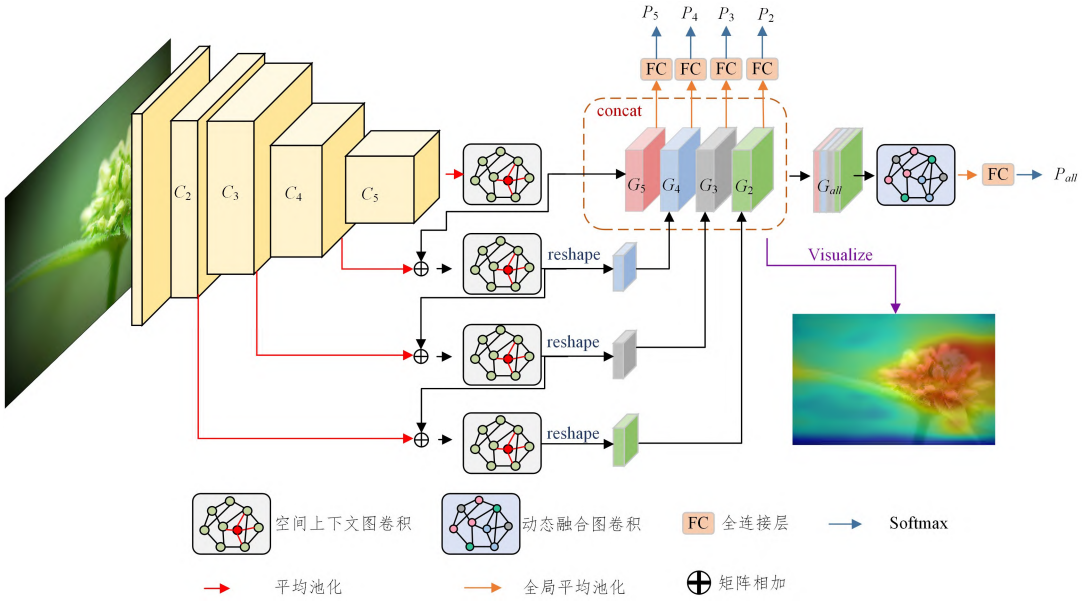


图 2 图像情感分类的层次图卷积网络模型框图

Fig. 2 Diagram of hierarchical graph convolutional network for image sentiment classification

模型情感识别能力的关键在于提取具有辨别力的情感特征，相比单个分类器指导下的层次特征提取，在每个层级设置分类器有利于模型专注于学习本层级的情感表示，这对细粒度情感分类尤为重要，因此本文在每个层级都设置情感预测分支，得到的情感预测概率为：

$$P_l=\mathrm{softmax}(\mathrm{FC}(\mathrm{GAP}(G_l))),l=2,3,4,5 \tag{2}$$

其中，GAP 将情感特征池化为 256×1 的情感向量，FC 代表全连接层。

得到层次特征的空间上下文表示$\{G_2,G_3,G_4,G_5\}$后，全局预测分支将其按通道维度拼接，再利用 DFGCN 动态聚合多层次通道信息，得到全局情感特征 G_{all}，并进行最终的图像情感分类，计算式如下：

$$G_{all}=[G_2,G_3,G_4,G_5] \tag{3}$$

$$P_{all}=\text{softmax}(\text{FC}(\text{GAP}(G_{all}))) \tag{4}$$

其中，$[\cdot]$表示将通道维度的特征拼接操作，P_{all}为全局情感预测概率。

2.2 空间上下文图卷积

现有基于 CNN 的图像情感分析方法不能充分捕获远距离情感区域的关联，情感表征能力有限，而图卷积网络能够通过节点间的关系推理提取全局情感特征。因此，本文提出空间上下文图卷积(SCGCN)，在空间维度上构图，计算空间节点的情感相似度，实现图像的全局情感感知。

由于卷积操作能保留原始图像的空间信息，层级特征 $C_l\in R^{H\times W\times C}$中的每一个像素就对应原始图像的一块区域，如图 3 所示。为了捕获图像区域在情感上的空间上下文关联，本文在层级预测分支中引入 M 层 SCGCN，首先将层次特征 $C_l\in R^{H\times W\times C}$视作空间网格上的节点表示 $X=\{x_1,x_1,\cdots,x_L\}$，$L=W\times H$ 代表节点数量。设第 m 个 SCGCN 的输入情感特征为 $H_S^{(m)}\in R^{L\times C}$，第一个 SCGCN 的输入为 $H_S^{(0)}=X$，利用 M 层 SCGCN 动态聚合空间节点得到第 l 层的空间上下文情感特征 $G_l=H_S^{(M+1)}$。SCGCN 的构建如图 3 所示，首先在坐标空间上构建了一个无向全连通图 $Graph_S=(H_S,E_S,A_S)$，其中 H_S 代表图 $Graph_S$ 中的节点，E_S 为连接节点的边，$\boldsymbol{A}_S$ 为空间邻阶矩阵，其中的元素描述了边的权重，如式(5)、式(6)所示。本文用节点向量之间的余弦距离衡量空间节点 x_i,x_j 所对应图像区域的情感内容相似度 $sim(x_i,x_j)$，并通过 softmax 函数归一化情感内容相似度系数，进而得到边的权重的具体数值。

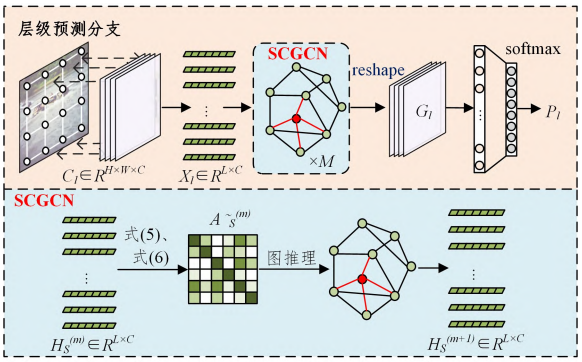


图 3 SCGCN 和层级预测分支

Fig. 3 Single level sentiment prediction branch with SCGCN

$$sim(x_i,x_j)=\frac{\langle\phi(x_i),\phi(x_j)\rangle}{\|\phi(x_i)\|\|\phi(x_j)\|} \tag{5}$$

$$A_s^{i,j}=\frac{\exp(sim(x_i,x_j))}{\sum_{k=1}^{L}\exp(sim(x_i,x_k))} \tag{6}$$

其中，$\langle\cdot\rangle$代表向量的内积运算，$\|\cdot\|$代表l_2范数，$\phi(x_i)=\boldsymbol{w}\cdot x_i$ 和 $\varphi(x_j)=\boldsymbol{w}'\cdot x_j$ 为两个线性变换操作，用于提升模型的泛化性能[26,29]，$\boldsymbol{w},\boldsymbol{w}'\in R^{C\times C}$是可学习网络参数矩阵，$\exp(\cdot)$为指数运算。

在完成情感坐标空间构图后，本文通过图卷积在图上进行推理，以实现邻居节点间的信息交互，其表达式如下：

$$\widetilde{\boldsymbol{A}}_s^{(m)}=\boldsymbol{A}_s^{(m)}+\boldsymbol{I} \tag{7}$$

$$H_s^{(m+1)}=\sigma(\widetilde{\boldsymbol{A}}_s^{(m)}H_s^{(m)}\boldsymbol{\theta}_s^{(m)}) \tag{8}$$

其中，$H_S^{(m)}$ 是第 m 层图卷积的激活值，$\boldsymbol{\theta}_c^{(m)}\in R^{C\times C}$为第 m 层图卷积的线性变换矩阵，σ 为 ReLU 激活函数。为防止节点在图卷积运算中与其他节点做加权求和时忽略节点自身的信息，本文在 $\boldsymbol{A}_s^{(m)}$的基础上加上单位矩阵 $\boldsymbol{I}$ 得到最终的邻阶矩阵 $\widetilde{\boldsymbol{A}}_s^{(m)}$。单位矩阵 $\boldsymbol{I}$ 也起到了跳跃连接的作用，缓解了训练过程中梯度消失的问题，使得节点间的信息传递变得更加稳定。GCN 中冗余的卷积层会导致过平滑现象的出现，使得模型无法聚合属于目标情感类别的特征节点，因此本文将 SCGCN 的数量 M 设置为 2。

2.3 动态融合图卷积

通过层级预测分支，网络能够利用 SCGCN 在各个层级上学习空间上下文特征$\{G_2,G_3,G_4,G_5\}$，并在每个层级进行情感预测。区别于语义分割等传统视觉任务，底层和高级情感特征在图像情感的表达上都发挥着重要作用，因此融合多层次情感特征能够实现更准确的情感分类效果。MldrNet[14]将多层次特征在通道维度上拼接以实现特征融合，利用融合特征实现了更准确的情感分类。但是，在深度网络中，图像的通道特征可以看作对特定类别情感的响应，不同层次特征的通道响应具有一定的关联性，将特征拼接的简单融合方式不能有效捕捉层次通道上的响应关系。因此，本文提出了 DFGCN，将多层次特征的一个通道视为图中的节点，并设计了一个邻阶矩阵激活函数，将通道节点之间的关系映射为数值在 0～1 之间的权重，以此来描述节点之间的情感关联程度，最后使用图卷积实现通道特征的跨层次特征融合。

如图 4 所示，$H_c^{(m)}\in R^{H\times W\times 4C}$为第 m 个 DFGCN 的输入，$H_c^{(0)}=G_{all}$则为将$\{G_2,G_3,G_4,G_5\}$拼接融合得到的初始全局情感特征。通过将 $H_c^{(m)}\in R^{H\times W\times 4C}$展开成 $4C$ 个维度为 $L=W\times H$ 的向量，可以得到通道特征的节点表示。为了捕获不同层次通道特征之间的情感关联，需要构建描述通道节点关系的通道邻阶矩阵。首先，使用全局平均池化将多层次情感特征 $H_c^{(m)}$ 转变为维度为 $4C$ 的向量 $\boldsymbol{V}=\{v_1,v_2,\cdots,v_{4C}\}$，向量中的每一个元素可以看作对应特定类别的情感相应强度，接着将向量 $\boldsymbol{V}=\{v_1,v_2,\cdots,v_{4C}\}$送入邻接矩阵激活函数 $F(\cdot)$，得到通道节点之间的关联矩阵 $\boldsymbol{A}_c^{(m)}$，矩阵元素的计算式如式(9)、式(10)所示。

$$F(\boldsymbol{V})_{i,j}=\frac{2}{1+e^{|v_i-v_j|}} \tag{9}$$

$$A_{c_{i,j}}^{(m)}=F(\boldsymbol{V})_{i,j}\times w_{i,j} \tag{10}$$

其中，v_i,v_j 分别代表向量 $\boldsymbol{V}$ 中的第 i,j 个元素，它们分别对应特征 $H_c^{(m)}$ 中第 i,j 个通道的情感响应强度。邻阶矩阵激活函数 $F(\cdot)$能够将 v_i,v_j 在情感响应上的关系映射为 0～1 之间的值 $F(V)_{i,j}$，通过将 $F(V)_{i,j}$乘以可学习参数 $w_{i,j}$可以增强网络的泛化性能，得到节点关联矩阵 $\boldsymbol{A}_c^{(m)}$。

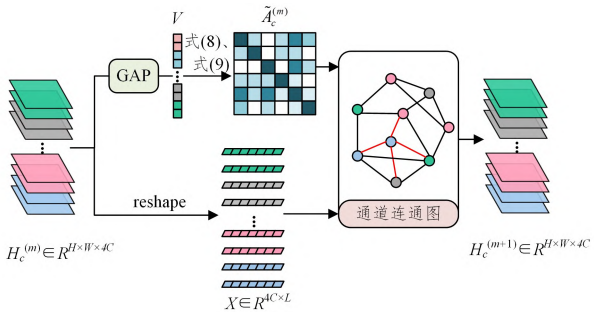


图 4 动态融合图卷积 DFGCN

Fig. 4 Dynamic channel graph convolution network

类似于 2.2 节中的 SCGCN，本文在 $\boldsymbol{A}_c^{(m)}$ 的基础上加上单位矩阵 $\mathbf{I}$ 得到最终的通道邻阶矩阵 $\widetilde{\boldsymbol{A}}_c^{(m)}$，避免节点特征在推理时过分关注邻居节点而忽略节点本身的情感信息。

$$\widetilde{\boldsymbol{A}}_c^{(m)}=A+\mathbf{I} \tag{11}$$

在完成通道维度的构图之后，使用 DFGCN 来动态聚合不同层次的通道特征，类似于 3.3 节的图卷积操作，DFGCN 图卷积计算可以用下列公式描述：

$$X_c^{(m)}=\text{Project}(H_c^{(m)}) \tag{12}$$

$$X_c^{(m+1)}=\sigma(\widetilde{\boldsymbol{A}}_c^{(m)}X_c^{(m)}\boldsymbol{\theta}_c^{(m)}) \tag{13}$$

$$H_c^{(m+1)}=\text{BackProject}(X_c^{(m+1)}) \tag{14}$$

式(12)中的 Project(·)将二维图像特征投影到节点空间，得到第 m 层的节点特征 $X_c^{(m)}$。式(13)通过图推理得到全局关系聚合后的节点特征 $X_c^{(m+1)}$，其中 $\boldsymbol{\theta}_c^{(m)}\in R^{4C\times 4C}$ 为可学习参数矩阵，可以提升网络的泛化能力，σ 代表 ReLU 激活函数。通过式(14)的反投影操作，可以将节点特征重新映射成二维图像特征。经过 N 层 DFGCN 就能得到最终的动态融合特征 $G_{\text{all}}=H_c^{(N)}$。为了避免过度平滑现象的出现，本文将 DFGCN 的数量 N 设置为 2。

2.4 多层次情感联合学习

2.1 节中，本文利用 SCGCN 增强层次情感特征的空间上下文表达，并且在每个层级上都进行情感分类学习并输出情感分类概率 P_l，$l\in\{2,3,4,5\}$，单层次的情感分类预测能帮助 SCGCN 将图像区域间的关系和情感属性相关联，在每个层级学习更具辨别性的层次特征，但是由于单层分支的网络学习容易过拟合，模型在不同层次的情感预测结果会发生冲突。为了缓解层级预测冲突，本文引入融合预测分支，利用 DFGCN 抑制和目标情感表达关联较小的层次通道特征，使用多层次融合之后的情感特征进行分类能够得到最终的情感预测概率 P_{all}。最后，本文基于交叉熵损失函数，联合优化层次情感特征学习和全局情感预测，损失函数如式(15)所示。

$$L=-\frac{1}{4N}\sum_{l=1}^{4}\sum_{m=1}^{N}\sum_{c=1}^{C}Y_m\times P_{m,c}^{l}-\frac{1}{N}\sum_{m=1}^{N}\sum_{c=1}^{C}Y_m\times P_{m,c}^{\text{all}} \tag{15}$$

其中，$l\in\{2,3,4,5\}$ 代表 4 个层级，Y_m 为第 m 个样本的真实情感标签，$P_{m,c}^{l}$ 代表网络在第 l 层级上将第 m 个样本预测为第 c 个情感类别的概率。

3 实验结果与分析

3.1 实验设置

1)数据集.

本文在 4 个广泛使用的基准数据集上进行实验，分别是 FI-8 数据集[30]、Abstract[31]、ArtPhoto[31] 和 EmotionROI[32] 数据集。

(1)FI-8 数据集。该数据集中的图像采集自社交媒体 Fliker 和 Instagram，总共有 23 308 张图片。根据 Mikels 等[12] 的研究，数据集中的每一张图片都被标记成 8 种情绪类别。由于其中一些图片在互联网上已经不存在，因此本文抓取其中的 21 826 张图片进行实验，按照 80%，5% 和 15% 的比例将数据集划分成训练集、验证集和测试集。另外，将 FI-8 数据集中的图像按照积极和消极进行归类，构成 FI-2 数据集，以测试本文网络在粗粒度情感分类上的有效性，数据集的划分比例和 FI-8 相同。

(2)Abstract，ArtPhoto 和 EmotionROI 数据集。这 3 个数据集是小规模情感数据集，包含 395，806 和 1 980 张情感图像，其中 Abstract 包含 395 张抽象艺术图像，ArtPhoto 包含 806 张与美学相关的艺术图像，EmotionROI 包含 1 980 张来自真实世界的场景图。本文利用网络在这 3 个小规模数据集上的预测结果证明了层次图卷积分类网络对于不同类型图像情感识别的泛化性和准确性。遵循文献[21-22]中的实验设置，将数据集划分成训练集(80%)和测试集(20%)。

2)实验环境

本文算法以及对比算法的实验运行环境如下：处理器为 Intel(R) Core(TM) i7-10700K CPU @ 3.80 GHz；独立显卡为 NVIDIA RTX 3090(24 GB)；内存(RAM)为 32.0 GB；系统类型为 Ubuntu 18.04；开发语言为 python3；深度学习框架为 Pytorch。

3)实验相关参数设置

本文中，batchsize 设定为 64，使用带动量(Momentum)的 SGD 优化器进行训练，动量设置为 0.9，权重衰减(weight_decay)设置为 0.01。对于 FI 数据集，将预训练模型 ResNet101 和本文新增网络模块的参数分别设置为 0.001 和 0.01，每隔 10 个 epoch 学习率减小为原来的 1/10，总共训练 35 个 epoch，再使用测试集测试训练好的模型。对于 Abstract，ArtPhoto 和 EmotionROI 这 3 个小型数据集，先使用 FI 数据集预训练网络，再将得到的网络在小规模数据集上进行微调，将优化器学习率设置为 0.000 1，其他设置与 FI 数据集保持相同。为了降低过拟合风险，除了设置权重衰减，本文还对训练数据使用了数据增强，包括随机裁剪并缩放为 448×448 的大小、随机水平翻转以及数据归一化处理。

4)评价指标

本文使用分类准确率(Accuracy)来评估模型的性能，准确率指测试样本中正确预测的样本数量与样本总数量的比值，以百分数的形式表示，如式(16)所示。

$$\text{分类准确率}=\frac{\text{正确预测的样本数量}}{\text{样本总数量}}\times 100\% \tag{16}$$

3.2 对比实验

本节将本文算法的情感分类准确率与 10 个先进的模型进行了对比，分别是 Sentibank、DeepSentibank、PCNN[33]、Zhu 等[9] 提出的模型、Resnet101[34]、MldrNet、Yang 等提出的模型、OSSCM、Wscnet 和 GOSR。其中除手工设计特征算法 Sentibank 外，其余算法均是基于深度视觉特征的。具体实验结果如表 1 所列，表中粗体表示最优结果。

表 1 不同算法的情感分类准确率对比

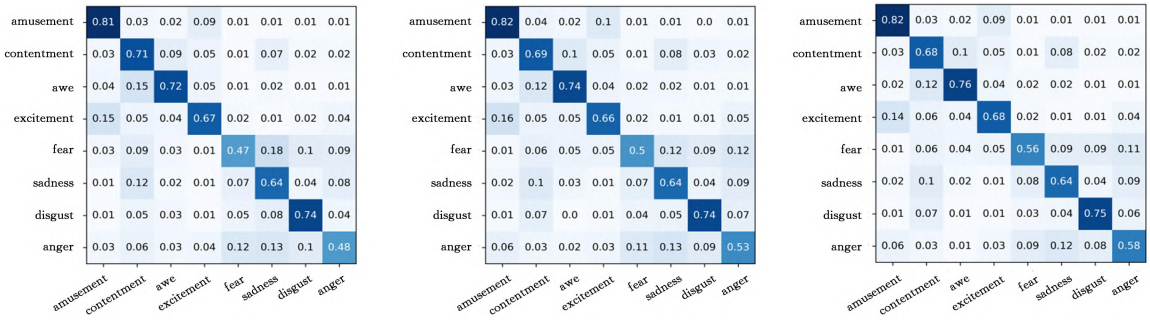
Table 1 Comparison of sentiment classification accuracy of different algorithms

（单位：%）

Method	FI-2	FI-8	Artphoto	EmotionROI	Abatract
Sentibank	56.47	—	67.85	66.18	64.95
DeepSentibank	64.39	—	70.11	70.11	68.73
PCNN	75.35	56.16	70.96	73.58	70.84
Zhu et al.	84.26	—	75.50	80.52	73.88
ResNet50	85.43	64.74	70.93	79.28	73.07
Resnet101	85.92	66.16	71.08	79.67	73.36
MldrNet	—	67.24	—	—	—
Yang et al	86.35	—	74.80	81.26	76.03
OSSCM	—	69.32	—	—	—
Wscnet	88.25	70.07	80.25	—	—
GOSR	85.40	70.80	—	83.0	—
ours	**90.67**	**71.54**	**81.43**	**85.37**	**83.65**

从表 1 可以看出，本文算法在所有数据集上都取得了最好的分类精度，尤其是在 FI-2，FI-8 和 EmotionROI 上，比现有效果最好的算法 GOSR 的性能提高了 5.27%，0.74%和 2.37%，这得益于本文层次分类网络的全局感受野和多层次特征融合，本文将在 3.3 节进行大量消融实验，验证本文网络各模块的有效性。

为了更详细地展示本文算法在每个情感类别上的细粒度分类效果，图 5 给出了本文模型、WSCNet 和 GOSR 在 FI-8 数据集上的情感细粒度分类混淆矩阵。可以看到，消极情感类别（fear，sadness，disgust，anger）更容易与其他类别相混淆，这是因为 FI 数据集中的积极情感样本数目大于消极情感类别，导致分类器更倾向于将消极情感预测为积极。对比混淆矩阵的可视化结果可以看出，与 GOSR 和 WSCNet 相比，本文网络在消极情感类别，尤其是 fear 和 anger 两个类别上也能获得更好的情感识别性能。



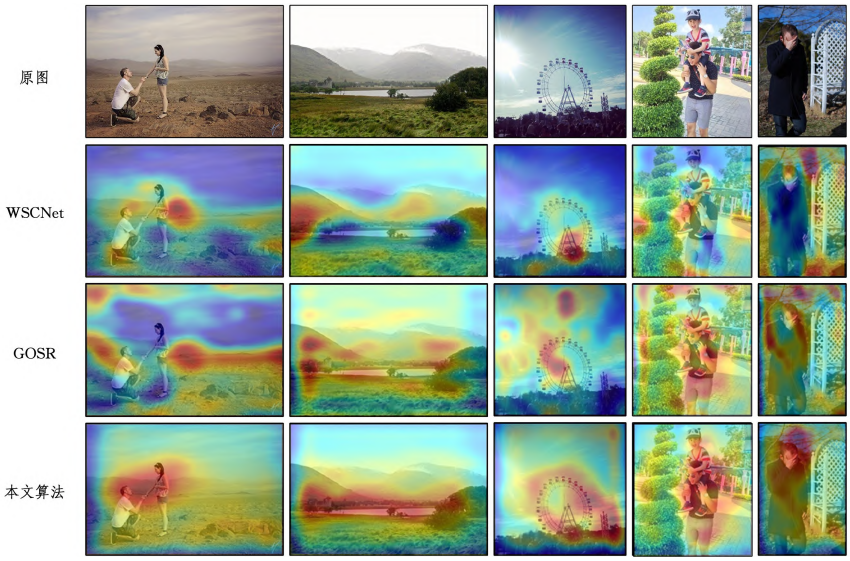
（a）WSCNet （b）GOSR （c）本文算法

图 5 本文算法和最先进算法在 FI-8 数据集上的分类混淆矩阵

Fig. 5 Classification confusion matrix of our method and the state-of-the-art algorithm on FI-8 dataset

利用类激活映射（CAM）[25]能够可视化网络感知到的情感区域，直观地展现网络的情感特征提取能力，因此利用 CAM 生成的热力图直观展示了本文算法在情感特征提取上的优势。

从图 6 可以看出，相比 WSCNet 和 GOSR，本文算法更能充分捕获情感元素的远距离关系依赖。



注：第 1 行为原图，第 2—4 行分别为 Wscnet、GOSR 和本文算法算法的情感专注区域。

图 6 不同算法的情感感知区域对比

Fig. 6 Comparison of sentiment perception regions of different algorithms

例如，在图 6 中的第 3 列中，WSCNet 和 GOSR 只能感知摩天轮的部分区域，但本文网络能够感知到摩天轮的整体轮廓，能更充分地捕获摩天轮这一情感要素。第 4 列中，GOSR 由于利用趋势性损失函数优化 CNN 分支的情感分类，相比 WSCNet 能更准确地提取情感特征，但也同时关注到较多的情感不相关区域，这在一定程度上会影响最终的分类结果。本文算法得益于图卷积带来的全局感受野，能够更准确地定位图像中影响情感反映的重要

特征，同时可避免情感不相关区域带来的噪声干扰。

3.3 消融实验

为了验证网络各模块的有效性，本文在 ResNet101 的基础上设置两大类消融实验。第一类实验对比 4 种分类架构，验证本文层级分类网络的有效性。该组实验中网络不加入 SCGCN 和 DFGCN，仅仅利用卷积算子提取图像特征。第二类消融实验将不包含 SCGCN 和 DFGCN 的层次分类网络作为 baseline，通过后续增加 SCGCN 和 DFGCN 来验证本文提出的两种图卷积在空间和通道上下文推理上的作用。

首先，如图 7 所示，设计了 4 类情感分类架构。图 7(a)单纯使用 ResNet101 的最后一层输出情感预测概率；图 7(b)在 ResNet101 的 4 个层级上分别进行情感分类；图 7(c)先将 ResNet101 的 4 个层级的情感特征池化成统一尺度，再按通道拼接进行层次融合，利用融合得到的全局情感特征识别图像情感；图 7(d)给出了本文的层次分类架构，先在单个层级上学习情感表示，再将 4 个尺度情感特征融合，并联合优化层次分类和全局情感分类。为了简便框图，图 7 中省略了全连接层 FC。

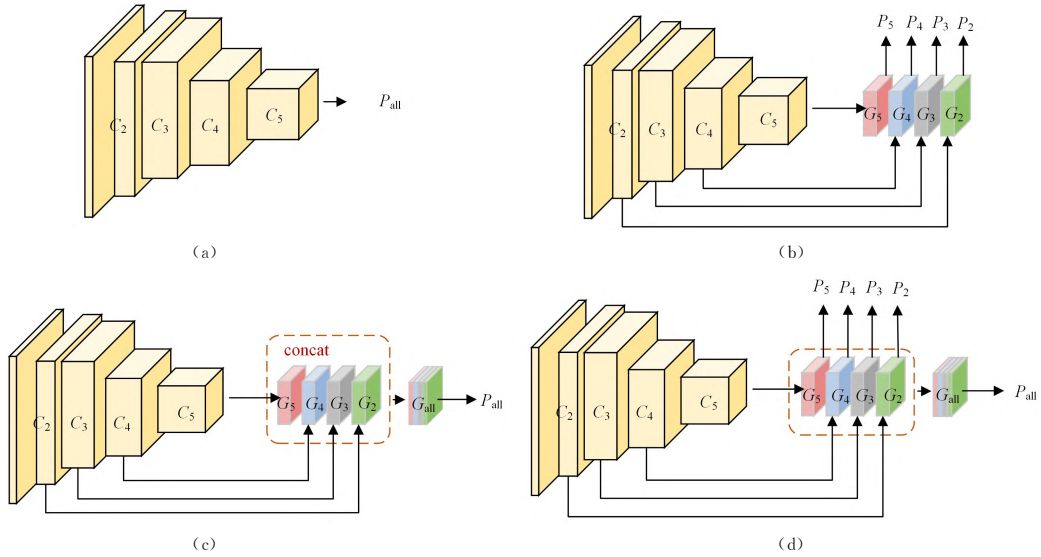


图 7 4 种情感分类架构

Fig. 7 Four sentiment classification architectures

消融实验的结果如表 2 所列，其中 Acc_2，Acc_3，Acc_4，Acc_5，Acc_{all} 分别为网络基于层次特征 C_2，C_3，C_4，C_5 和多层次融合特征 C_{all} 计算出的情感分类准确度。从图 7(b)可以看出，不同层次特征对于情感表达的贡献不同，其情感分类准确度有一定差异。对比图 7(a)可以发现，相比仅仅用 ResNet101 提取的高级情感语义进行情感分类，同时优化 4 个层次的情感特征学习能获得更好的情感识别效果。图 7(c)的结果表明，将 4 个层次的特征进行拼接融合，使得模型能够同时利用图像高级语义和底层视觉特征进行情感分类，提升情感分类的性能。为了验证本文多层次情感联合学习框架的优势，本文在图 7(b)的基础上设置实验 7(d)，联合单层次分类学习和融合情感学习，可以发现单层次分类准确度和融合分类准确度都能获得很好的提升，证明本文的层次分类框架能够使模型在不同层级上学习到更准确的情感表征。

表 2 4 种分类架构的准确度对比

Table 2 Accuracy comparison of four classification architectures

(单位：%)

方法	Acc_2	Acc_3	Acc_4	Acc_5	Acc_{all}
图 7(a)所示模型	—	—	—	—	66.16
图 7(b)所示模型	67.24	67.31	67.57	67.88	—
图 7(c)所示模型	—	—	—	—	68.13
图 7(d)所示模型	69.03	69.16	69.22	69.43	69.84

图 7(d)所示的模型为 baseline，本文通过进一步的消融实验来说明 SCGCN、DFGCN 以及特征自顶向下融合的作用，实验结果如表 3 所列。

表 3 消融实验情感分类准确度的对比

Table 3 Accuracy comparison of sentiment classification in ablation experiments

(单位：%)

消融算法	Baseline	SCGCN	DFGCN	自顶向下融合	FI-8
B	√	—	—	—	69.84
B+D	√	—	√	—	70.67
B+S	√	√	—	—	70.39
B+F	√	—	—	√	70.01
B+S+D	√	√	√	—	71.36
B+F+D+S	√	√	√	√	71.54

表 3 中第 1 列中的 B，D，S，F 分别代表 baseline，SCGCN，DCGCN 和 FPN 中自顶向下的特征融合，“+”代表在网络中嵌入对应的组件。可以看出，在 baseline 的基础上自顶向下地融合相邻层级情感特征能够在一定程度上提升情感分类性能，因为高层级情感特征在情感语义的表达上更为强烈，通过逐级自顶向下融合不仅可以增强底层特征的情感语义表达，还能在一定程度上减缓层级预测冲突，提升多层次融合后的特征质量。在 baseline 的每一层级加入 SCGCN 能够将情感分类准确率提升 0.65%，因为 SCGCN 能够构建图像区域间的情感关联图，利用图卷积在全局角度聚合单一层级的情感特征，弥补 CNN 带来的感受野缺陷。另外，通过在融合预测分支中引入 DFGCN，本文网络能够将 baseline 的分类准确率提升 0.83%，这主要得益于 DFGCN 赋予的通道特征跨层次交互能力。通过在 baselline 的基础上加上 SCGCN、DFGCN

和自顶向下融合，本文的层次图卷积分类网络相比 baseline 能获得显著的分类性能提升，证实了本文网络在图像情感分类上的有效性。

如图 8 所示，本文还提供消融实验的可视化效果图，直观展示本文核心组件 DFGCN 和 SCGCN 在情感特征提取上的效果。虽然表 3 所列的实验结果表明，在 baseline 的基础上增加自顶向下的特征融合能在一定程度上提升分类准确度，但从“B+F”组的情感可视化图可以看出，网络在关注老鹰、孩童、浪花等情感要素的同时，给予了部分情感不相关区域同等的关注，如在图 8(b)中，除了老鹰这一前景对象，还提取到了大量关于天空的背景信息。从“B+F+D”组可以看出，通过加入 DFGCN 能够缓解这一现象，因为 DFGCN 能够利用通道连通图跨层次构建情感响应的关联，在特征融合时动态调整不同层次特征的权重，帮助网络选择与目标情感相关的层次特征。比如，图 8(a)、图 8(b)中，墙壁、天空等背景信息的情感表达就得到了有效抑制。最后，为了说明 SCGCN 的作用，本文在层次分类网络的每一层引入 SCGCN，对比图 8 中的第 3，4 列图片可以发现，加入 SCGCN 后网络能更准确且充分地提取情感特征，例如，图 8(b)中，第 4 列可视化出的老鹰整体轮廓相比第 3 列更为清晰，图 8(d)组中，“B+F+D”虽然能感知到图像左上部分的摩托车，但是对右下角的骑行者的感知不够充分，这得益于 SCGCN 的空间上下文推理能力，第 4 列图像中左上和右下的骑行者虽然在图像中空间距离较为遥远，但都能被模型所捕获。

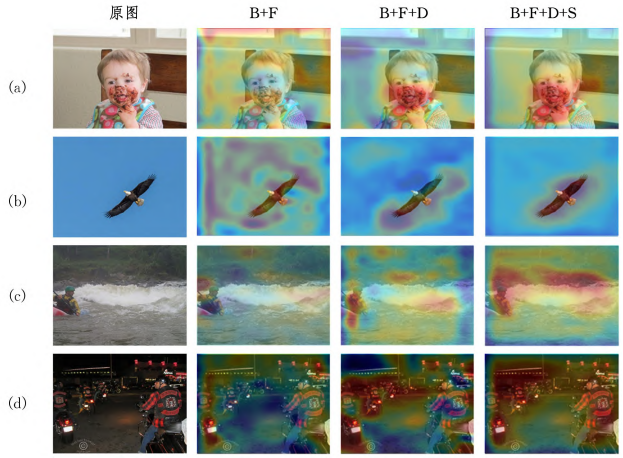


图 8 在 baseline 的基础上引入各模块的情感注意力图

Fig. 8 Sentiment attention map after each module is introduced based on baseline

类似于通道注意力机制，本文的 DFGCN 模块能够增强与目标情感类别相关的层次通道特征。但是目前的通道注意力机制，如 SENet[35]，CBAM[36]，ECANet[37] 等，静态地计算不同通道特征的重要性权重，通道特征间的关系没有被充分利用。本文的 DFGCN 能够利用图卷积动态建模层次通道特征间的情感关联，通过将本文网络中的 DFGCN 替换成 SENet，CBAM 和 ECANet 中的通道注意力模块，可以验证 DFGCN 在图像情感分析任务上的有效性。实验结果如表 4 所列。

表 4 DFGCN 与现有通道注意力模块的情感分类准确度的对比

Table 4 Accuracy comparison of sentiment classification between DFGCN and existing channel attention modules

（单位：%）

Method	FI-2	FI-8
SE	90.12	70.74
CBAM	90.23	70.96
ECA	90.16	70.31
DFGCN	**90.67**	**71.54**

在 FI-2 和 FI-8 上的粗、细两个粒度上的情感分类结果表明，相比现有的通道注意力模块，本文的动态通道图卷积能获得更高的情感分类准确度。

结束语 本文提出了一种基于 SCGCN 和 DFGCN 的层次图卷积分析网络，用于图像情感分类任务，模型在每一个层级使用 SCGCN 捕获图像局部区域之间的情感关联，弥补了基于 CNN 的图像情感分析模型在全局上下文建模方面的缺陷。通过层次分类架构，模型能够聚焦于学习每一层级上的情感空间嵌入，提高层次特征的情感表征能力。最后使用基于 DFGCN 的特征融合分支，模型能够动态构建不同层次通道特征的上下文联系，从而聚合和目标情感类别相关的通道特征，实现跨层次特征的高效融合。本文利用层级融合之后的全局情感特征进行情感分类，在 4 个公开数据集上的实验表明，本文方法在情感极性分类和细粒度情感分类上都能获得比现有方法更优秀的分类性能。

参 考 文 献

[1] YOU Q, LUO J, JIN H, et al. Cross-modality consistent regression for joint visual-textual sentiment analysis of social multimedia[C]// Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. 2016: 13-22.

[2] ZHAO S, YAO X, YANG J, et al. Affective image content analysis: Two decades review and new perspectives[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(10): 6729-6751.

[3] ZHAO S, YAO H, GAO Y, et al. Predicting personalized emotion perceptions of social images[C]// Proceedings of the 24th ACM International Conference on Multimedia. 2016: 1385-1394.

[4] LANG P J, BRADLEY M M, CUTHBERT B N. International affective picture system(IAPS): affective ratings of pictures and instruction manual. (Rep. No. A-8)[R]. 2008.

[5] ZHAO S, GAO Y, JIANG X H, et al. Exploring Principles-of-Art Features For Image Emotion Recognition[J]. ACM, 2014: 47-56.

[6] BORTH D, CHEN T, JI R, et al. Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content[C]// Proceedings of the 21st ACM International Conference on Multimedia. 2013: 459-460.

[7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[8] CHEN T, BORTH D, DARRELL T, et al. Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks[J]. arXiv: 1410.8586, 2014.

[9] ZHU X, LI L, ZHANG W, et al. Dependency Exploitation: A Unified CNN-RNN Approach for Visual Emotion Recognition

[C]// IJCAI. 2017:3595-3601.
[10] PANDA R, ZHANG J, LI H, et al. Contemplating visual emotions: Understanding and overcoming dataset bias[C]// Proceedings of the European Conference on Computer Vision(ECCV). 2018:579-595.
[11] SHE D, YANG J, CHENG M M, et al. Wscnet: Weakly supervised coupled networks for visual sentiment classification and detection[J]. IEEE Transactions on Multimedia, 2019, 22(5): 1358-1371.
[12] MIKELS J A, FREDRICKSON B L, LARKIN G R, et al. Emotional category data on images from the International Affective Picture System[J]. Behavior Research Methods, 2005, 37(4): 626-630.
[13] YANG J, SHE D, SUN M. Joint Image Emotion Classification and Distribution Learning via Deep Convolutional Neural Network[C]// IJCAI. 2017:3266-3272.
[14] RAO T, XU M, LIU H, et al. Multi-scale blocks based image emotion classification using multiple instance learning[C]// 2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016:634-638.
[15] SUN M, YANG J, WANG K, et al. Discovering affective regions in deep convolutional neural networks for visual sentiment prediction[C]// 2016 IEEE International Conference on Multimedia and Expo(ICME). IEEE, 2016:1-6.
[16] YANG J, SHE D, SUN M, et al. Visual sentiment prediction based on automatic discovery of affective regions[J]. IEEE Transactions on Multimedia, 2018, 20(9): 2513-2525.
[17] YOU Q, JIN H, LUO J. Visual sentiment analysis by attending on local image regions[C]// Thirty-First AAAI Conference on Artificial Intelligence. 2017.
[18] ZHANG J, CHEN M, SUN H, et al. Object semantics sentiment correlation analysis enhanced image sentiment classification[J]. Knowledge-Based Systems, 2020, 191: 105245.
[19] HU Y, WEN G, CHAPMAN A, et al. Graph-based visual-semantic entanglement network for zero-shot image recognition [J]. IEEE Transactions on Multimedia, 2021, 24: 2473-2487.
[20] CHANDRA S, USUNIER N, KOKKINOS I. Dense and low-rank gaussian crfs using deep embeddings[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 5103-5112.
[21] BERTASIUS G, TORRESANI L, YU S X, et al. Convolutional random walk networks for semantic image segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:858-866.
[22] YAN J D, JIA C Y. Text Classification Method Based on Double-Graph Neural Network Information Fusion [J]. Computer Science, 2020, 49(8): 230-236.
[23] ZHOU F Q, CHENG W Q. Sequence recommendation based on globally enhanced graph neural network [J]. Computer Science, 2022, 49(9): 55-63.
[24] ZHOU H Y, ZHANG D Q. Multi-center data-oriented hypergraph convolutional neural network and its application [J]. Computer Science, 2022, 49(3): 129-133.
[25] LI Z M, ZHANG Y P, LIU Y J, et al. Point Cloud Representation Learning Based on Deformable Graph Convolution [J]. Computer Science, 2022, 49(8): 273-278.
[26] WANG X, GUPTA A. Videos as space-time region graphs [C]// Proceedings of the European conference on computer vision(ECCV). 2018:399-417.
[27] CHEN Y, ROHRBACH M, YAN Z, et al. Graph-based global reasoning networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 433-442.
[28] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2117-2125.
[29] LIU D, PURI R, KAMATH N, et al. Composition-aware image aesthetics assessment[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020: 3569-3578.
[30] YOU Q, LUO J, JIN H, et al. Building a large scale dataset for image emotion recognition: The fine print and the benchmark [C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2016.
[31] MACHAJDIK J, HANBURY A. Affective image classification using features inspired by psychology and art theory[C]// Proceedings of the 18th ACM International Conference on Multimedia. 2010:83-92.
[32] PENG K C, SADOVNIK A, GALLAGHER A, et al. Where do emotions come from? predicting the emotion stimuli map[C]// 2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016:614-618.
[33] YOU Q, LUO J, JIN H, et al. Robust image sentiment analysis using progressively trained and domain transferred deep networks[C]// Twenty-ninth AAAI Conference on Artificial Intelligence. 2015.
[34] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
[35] JIE H, LI S, GANG S. Squeeze-and-Excitation Networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2018.
[36] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module[C]// Proceedings of the European Conference on Computer Vision(ECCV). 2018:3-19.
[37] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2020.

TAN Qianhui, born in 1998, postgraduate. His main research interests include deep learning and image sentiment analysis.

SUN Yubao, born in 1983, Ph.D, professor, is a member of China Computer Federation. His main research interests include deep learning theory and applications and image sentiment analysis.

(责任编辑:喻藜)