

基于多任务深度神经网络的图像情感区域检测

徐晶晶¹, 张俊升¹, 李腾腾¹, 祝汉城¹, 谢涛¹, 余伟²⁺

(1. 中国矿业大学 信息与控制工程学院, 江苏 徐州 221116;

2. 成都理工大学工程技术学院 电子信息与计算机工程系, 四川 乐山 614000)

摘要: 为得到准确的图像情感区域, 设计一种基于多任务深度神经网络的图像情感区域检测方法。提出的网络包括两个任务: 图像情感区域检测任务和图像情感分类任务, 两个任务的共享层为 VGG-16 网络结构, 图像的情感区域检测任务和情感分类任务通过对共享层参数的训练可以提高单个任务的泛化能力, 提升网络的性能。实验结果表明, 该方法可以准确检测图像的情感区域, 检测性能优于单任务情感区域检测方法和现有的显著性检测和物体检测方法。

关键词: 神经网络; 情感区域; 多任务; 情感分类; 深度学习

中图分类号: TP183 **文献标识码:** A **文章编号:** 1000-7024 (2020) 09-2616-07

doi: 10.16208/j.issn1000-7024.2020.09.032

Image emotion region detection based on multi-task deep neural network

XU Jing-jing¹, ZHANG Jun-sheng¹, LI Teng-teng¹, ZHU Han-cheng¹, XIE Tao¹, YU Wei²⁺

(1. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China;

2. Department of Electronic Information and Computer Engineering, The Engineering & Technical College of Chengdu University of Technology, Leshan 614000, China)

Abstract: To get image emotional region accurately, an image emotional region detection method based on multi-task deep neural network was designed. The proposed network consisted of two tasks, namely image emotion region detection task and image emotion classification task. The shared layer of the two tasks was the VGG-16 network structure. The emotional region detection task and the emotion classification task of the image improved the generalization ability of the single task by sharing network parameters for training, thereby improving the performance of the network. Experimental results show that the proposed method can detect the emotional region of the image effectively. The detection performance of the proposed method is better than that of the single-task emotion prediction method and the existing saliency detection and object detection methods.

Key words: neural network; emotional regions; multi-tasking; emotion classification; deep learning

0 引言

情感计算的概念是由 Picard 提出的, 他在《Affective Computing》一书中给出了情感计算的定义, 即情感计算是指与情感相关、来源于情感或能够对情感施加影响的计算^[1]。情感计算具有很多潜在的广泛应用, 比如情感图像检索^[2]、美学质量评价^[3]、数据挖掘^[4]和评论助手^[5]等方面。

早期针对图像的情感计算主要是利用颜色^[6]、纹理^[7,8]

和形状^[9,10]等低层次特征进行图像情感分析, 但是这些低层次特征描述图像所蕴含的高层次的情感特征的能力十分有限。随着卷积神经网络 (convolutional neural network, CNN) 在提取图像高层次特征能力的凸显, 越来越多的基于卷积神经网络的图像情感计算方法被提出。

目前, 基于卷积神经网络的情感计算方法对图像情感区域的研究甚少, 大多从图像的情感分类任务进行研究的, 且研究任务单一, 没有考虑到图像的情感区域和情感分类之间的相关性。图像激发情感的区域和激发的情感类

收稿日期: 2019-07-10; 修订日期: 2019-09-06

基金项目: 江苏省研究生科研与实践创新计划基金项目 (KYCX19_242)

作者简介: 徐晶晶 (1993—), 女, 安徽宿州人, 硕士研究生, 研究方向为情感分析; 张俊升 (1995—), 男, 安徽宿州人, 硕士研究生, 研究方向为图像质量评价; 李腾腾 (1992—), 女, 山东济宁人, 硕士研究生, 研究方向为三维重建; 祝汉城 (1989—), 男, 江苏徐州人, 博士研究生, 研究方向为深度学习和情感分析; 谢涛 (1995—), 男, 湖南娄底人, 硕士研究生, 研究方向为语义分割; +通讯作者: 余伟 (1993—), 男, 四川乐山, 硕士研究生, 研究方向为图像质量评价。E-mail: TS17060129P3@cumt.edu.cn

别是相关的, 图像情感区域是局部的, 图像情感的类别是由观看者观看图像的局部区域而激发的, 因此, 图像的情感区域检测和图像的情感分类在情感计算中都具有重要的意义。考虑到图像的情感区域检测和图像的情感分类任务的相关性, 受到多任务学习思想的启发, 本文设计了一种基于多任务深度神经网络的模型检测图像的情感区域。

1 相关工作

近年来, 与情感相关的话题在计算机视觉领域受到越来越多的学者关注, 尤其是图像的情感分类, 但是对于图像的情感区域问题的研究较少。如 You 等^[11]为解决在大规模且有噪声的数据集上训练的问题, 采用逐步训练的策略对网络架构进行微调, 提出了用于视觉情感分析的渐进卷积神经网络 (progressive CNN, PCNN); Wu 等^[12]基于情感的形容词-名词对 (adjective-noun pairs, ANP) 的标签提出了一种增强 CNN 模型来提高情感分类性能。然而上述方法均是通过图像的全局信息来对图像传达的情感进行建模的, 图像的全局信息中包含的与情感无关的冗余信息大大增加了情感计算的工作量, 所以有学者提出从图像的局部情感注意力区域进行分析与研究。

针对图像局部区域的分析与研究。Zhu 等^[13]使用弱监督网络生成局部区域预测性格特征; You 等^[14]使用注意力模型来学习图像局部区域与视觉情感之间的对应关系, 然而, 由于视觉属性检测器的性能较差, 无法准确检测图像的情感区域; Yang 等^[15]使用弱监督的方法生成情感区域来监督情感的分类, 虽然情感分类效果有提升, 但是对于情感区域检测的准确率不高, 这是因为在模型训练过程中很少考虑到多个因素对图像情感的影响, 这样容易造成模型的过拟合, 不能达到很好的泛化能力。

尽管上述方法在情感计算方面的研究都取得了一定的效果, 但是大多数都是基于图像情感的分类研究的, 没有考虑到图像情感区域检测的重要性。通常情况下, 观看者观看图像时的注意力会集中在图像的局部区域而不是整体区域。图 1 为选自 EmotionROI 数据集^[16]中的 3 幅图像以及对应的情感注意力区域, 图像的情感区域是由 15 位用户进行手工标注合成的。由图可知, 人们对图像的注意力往往在局部区域。因此, 在视觉情感分析中, 研究图像的情感区域检测方法对图像情感计算具有重要意义。

2 本文算法

图像的情感区域与图像的显著性区域以及物体区域之间有着密切的关系, 图像的显著性区域和物体区域往往是图像中突出的物体, 虽然情感区域中也包含着图像中的显著物体, 但是情感区域更多关注的是物体与物体之间的联系和更高层次的语义信息。

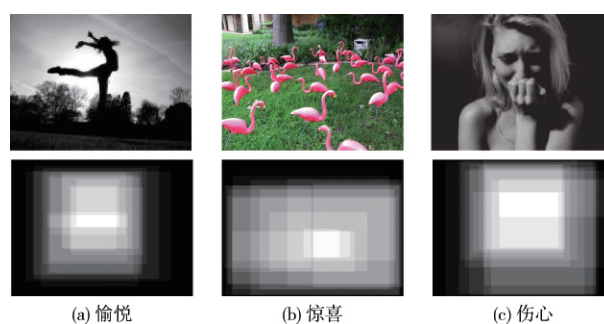


图 1 EmotionROI 情感库图像

图 2 为图像以及对应的情感区域、显著性检测^[17]区域和物体检测^[18]区域。图 2 (c) 中的显著性检测区域主要集中在狗的身体部分, 而图 2 (d) 中的物体检测区域主要集中在狗的头部, 图 2 (c) 和图 2 (d) 检测出的区域与情感区域图 2 (b) 差异较大。

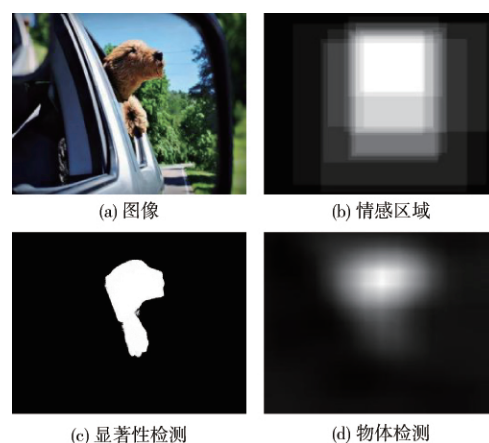


图 2 图像以及对应的情感区域、显著性检测区域和物体检测区域的对比

由图 2 可知, 图像的情感区域包含了图像中的显著物体, 但是除此之外, 情感区域还包含物体周围的一些场景信息, 因此仅通过图像的显著性检测和物体检测方法不能准确得到图像的情感区域。

为了更加准确检测图像的情感区域, 本文设计了一种基于多任务的深度神经网络来检测图像的情感区域。多任务网络结构是由情感区域检测网络和情感分类网络两部分组成, 本节将详细介绍这两个网络的结构。基于多任务深度神经网络结构如图 3 所示。

给定训练集为 N 的集合 $\{X_i, S_i, Y_i\}_{i=1}^N$, 其中 X_i 为输入的图像, S_i 为输入的带有情感区域标签的图像, $Y_i \in \{1, \dots, C\}$ 为图像 X_i 对应的情感标签。输入图像 X_i 以及对应的情感区域标签 S_i 和情感标签 Y_i , 设经过多任务网络后输出情感区域图映射为 \hat{S}_i , 情感标签为 $y_i \in \{1, \dots, C\}$ 。

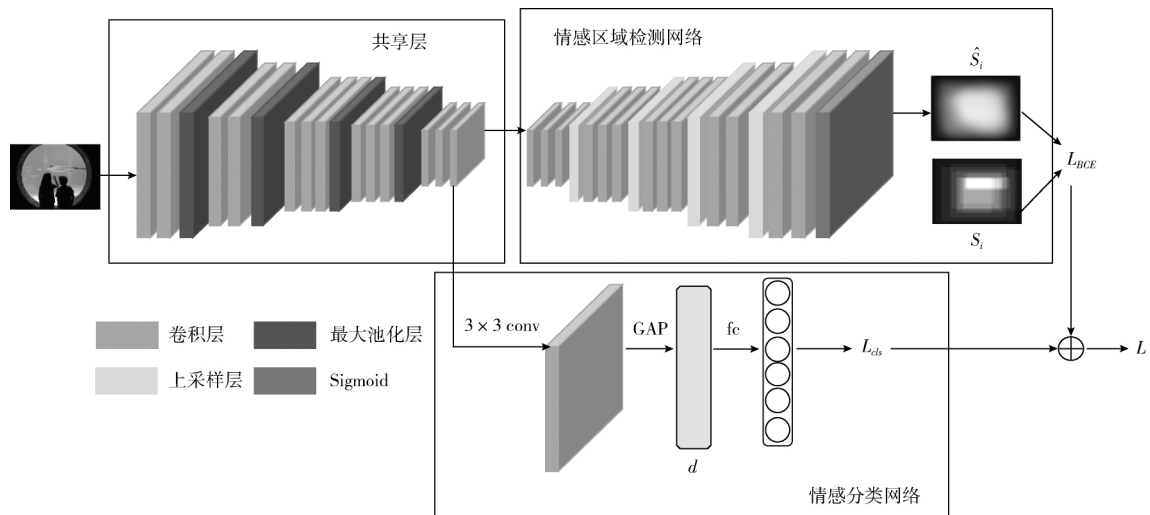


图3 基于深度学习的多任务网络结构

如图3所示,本文提出的多任务网络的共享层为VGG-16网络^[19]结构,但是省略了最后的池化层和全连接层。共享层初始化使用的权重是在ImageNet^[20]数据集上预训练得到的。

2.1 情感区域检测网络

情感区域检测网络采用编码-解码的网络结构,解码器的网络结构与编码器相同,并且解码器和编码器在结构上是对称的。编码器部分为共享层的结构,网络结构包含最大池化层减小特征图的大小;解码器部分使用上采样层,通过卷积滤波将减小的特征图恢复到与输入图像分辨率相同的大小。网络最后一层添加非线性的 1×1 的sigmoid激活层来生成情感区域映射,网络的最终输出是与输入图像大小相同的情感区域映射。

在图像情感区域检测网络训练时,损失函数定义在检测的情感区域图与其对应的情感区域标签之间,内容损失计算检测结果图与真实结果图之间的欧式距离,称为均方误差(MSE)定义为

$$L_{MSE} = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H (S_i(x, y) - \hat{S}_i(x, y))^2 \quad (1)$$

式中: W 为图像的宽度, H 为图像的高度, $S_i(x, y)$ 为情感区域标签 S_i 在 (x, y) 处的值, $\hat{S}_i(x, y)$ 为检测的情感区域图 \hat{S}_i 在 (x, y) 处的值。本文中,内容损失函数作为一个参考基准。用梯度下降法优化损失函数之前,首先对情感区域映射进行归一化处理,使其每个像素点的值都在 $[0, 1]$ 之间。为了保证检测结果均在 $[0, 1]$ 之间,本文在情感区域网络的最后一层加入了一个像素级的 Sigmoid 激活函数,其损失函数为二进制交叉熵,数学公式为

$$L_{BCE} = -\frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H (S_j \log(\hat{S}_j) + (1 - S_j) \log(1 - \hat{S}_j)) \quad (2)$$

式中: $S_j = S_i(x, y)$ 为情感区域标签 S_i 在 (x, y) 处的值, $\hat{S}_j = \hat{S}_i(x, y)$ 为检测的情感区域图 \hat{S}_i 在 (x, y) 处的值。

2.2 情感分类网络

在情感分类网络中,共享层的最后一层使用 512 个大小为 3×3 的卷积核生成 512 个通道的特征图,使用全局平均池化(global average pooling, GAP)输出维度为 512 的情感向量 d ,然后将情感向量 d 送到全连接层(full connected layer, fc),并通过 softmax 分类器输出情感类别。

在图像的情感分类过程中,输入图像 X_i 以及对应的情感标签 Y_i 需要计算输入图像的情感预测。softmax 层可以解释为情感特征向量 d 到输出的情感类别中且分配一个相对应的概率分数 $s(y_i = c | d, w_c)$,定义为

$$s(y_i = c | d, w_c) = \frac{\exp(w_c^T d)}{\sum_{c'=1}^C \exp(w_{c'}^T d)} \quad (3)$$

式中: $\{w_c\}_{c=1}^C$ 为情感向量 d 的模型权重, C 为输出的情感类别的数量, y_i 为预测的情感类别, $c \in \{1, \dots, C\}$, 为衡量模型损失,通过最小化 log 似然函数来实现分类,即分类损失函数为

$$L_{cls} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C I(y_i = c) \log s(y_i = c | d, w_c) \quad (4)$$

式中: $I(x)$ 为指示函数,如果条件 x 为真,则 $I(x) = 1$,否则为 $I(x) = 0$ 。

2.3 多任务训练

从图3可以看出,多任务网络有情感区域检测和情感分类两个输出。本文使用随机梯度下降法优化网络,同时训练情感区域检测网络和情感分类网络,提出的多任务网络模型损失函数为

$$L = L_{BCE} + L_{cls} \quad (5)$$

式中: L_{BCE} 为情感区域网络的损失函数, L_{cls} 为情感分类网络的损失函数。

3 实验结果与分析

3.1 实验设置

本文方法的性能在 EmotionROI 数据集上进行评估, EmotionROI 数据集共有 1980 幅图像, 包含生气、厌恶、恐惧、愉快、悲伤和惊喜 6 种情感。数据集中 70% 的图像作为训练集, 30% 的图像作为测试集。

本文实验运行环境为 Linux-Ubuntu16.04, python3.6, Theano 0.8.2, 开发工具为 Pycharm。输入图像的尺寸大小为 224×224 , 选择 Momentum 的随机梯度下降法优化网络, Momentum 设为 0.5, 学习率设为 0.001, 所有模型均训练 100 个 Epoch。

实验评价指标为平均绝对误差 (MAE)、查准率-查全率 (precision-recall, PR) 曲线、F-measures (F_1 , F_2 , $F_{0.5}$, $F_{\sqrt{0.3}}$) 分数、分类准确率。

平均绝对误差 (MAE) 是检测结果和真实结果之间的绝对误差的平均值定义为

$$MAE = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H |S_i(x, y) - \hat{S}_i(x, y)| \quad (6)$$

式中: W 为图像的宽度, H 为图像的高度, $S_i(x, y)$ 为真实结果 S_i 在 (x, y) 处的值, $\hat{S}_i(x, y)$ 为检测结果 \hat{S}_i 在 (x, y) 处的值。MAE 越小, 说明检测结果与真实结果越接近。

查准率定义为被正确检测区域的像素与检测区域像素的比值, 查全率定义为被正确检测区域的像素与情感区域中标记的区域像素的比值。F-measure 根据查准率和查全率定义如下

$$F_{\beta} = (1 + \beta^2) \cdot \frac{P \cdot R}{\beta^2 \cdot P + R} \quad (7)$$

式中: P 为查准率, R 为查全率, β 是可变参数, 一般常取值为 1、2、0.5、 $\sqrt{0.3}$ 。

分类准确率 (Accuracy) 为测试集中正确预测的情感类别的数目与总的测试集数目的比值。

3.2 实验结果与分析

为了验证本文提出的多任务网络的有效性, 分别在图像情感区域检测的单任务网络和图像情感分类的单任务网络上进行与多任务网络相同的实验设置的训练。单任务情感区域检测网络训练时的损失函数为式 (2) 所示, 单任务情感分类网络训练时的损失函数为式 (4) 所示。

3.2.1 情感区域检测结果

在 EmotionROI 测试集上将提出的多任务情感区域检测方法和单任务情感区域检测方法与显著性区域检测方法和物体区域检测方法进行比较分析。图 4 为不同算法检测结果的可视化图。图 (a) 为输入图像; 图 (b) 为图像的情感区域标签; 图 (c) 为多任务情感区域检测网络 (MT) 生成的情感区域图; 图 (d) 为单任务情感区域检测网络 (ST) 生成的情感区域图; 图 (e) 为显著性 GBVS^[21] 方法生成的显著图; 图 (f) 为物体检测方法 RDLM^[22] 生成的物体区域图; 图 (g) 为显著性 SDSP^[23] 方法生成的显著图; 图 (h) 为物体检测方法 OBJ^[15] 生成的物体区域图; 图 (i) 为显著性 ITTI^[24] 方法生成的显著图; 图 (j) 为显著性 LARK^[25] 方法生成的显著图; 图 (k) 为显著性 AIM^[26] 方法生成的显著图; 图 (l) 为显著性 SR^[27] 方法生成的显著图; 图 (m) 为显著性 SUN^[28] 方法生成的显著图; 图 (n) 为显著性 FT^[29] 方法生成的显著图。从图中可以看出图 (c) 和图 (d) 的情感区域与情感区域标签图 (b) 最接近, 多任务情感区域检测结果能准确表示真实的情感区域。

MAE 为情感区域检测结果与情感区域标签图 4 (b) 之间的平均绝对偏差, MAE 越小, 说明检测结果与真实结果越接近, 算法的性能越好。从图 4 可以看出本文提出的多任务情感区域网络模型的测试结果图与情感区域真实结果图的 MAE 为 0.045, 单任务情感区域检测的结果图与情感区域真实结果图的 MAE 为 0.062, 均小于 GBVS 和 RDLM 等方法的测试结果。由以上分析可知, 多任务情感区域检测网络的测试结果与真实情感区域结果最接近, 多任务情感区域检测方法的性能优于单任务情感区域检测方

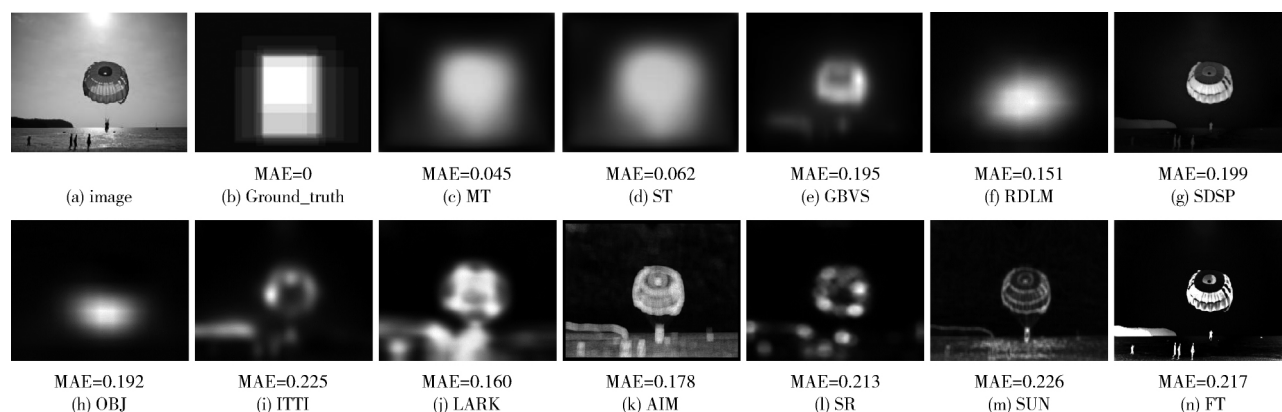


图 4 本文提出的算法在 EmotionROI 上的测试结果的情感区域

法和显著性物体检测方法的性能。

图 5 为本文提出的方法和显著性物体检测方法在 EmotionROI 数据集上测试的 PR 曲线图。从 PR 曲线中评价算法性能的好坏可以从平衡点的大小判断,平衡点是 F_1 时的取值,平衡点值越大,说明算法性能越好。其中 MT 和 ST 为多任务情感区域检测方法和单任务情感区域检测方法; AIM、FT、GBVS、ITTI、LARK、SDSP、SR 和 SUN 为显著性检测方法; OBJ 和 RDLM 为物体检测方法。

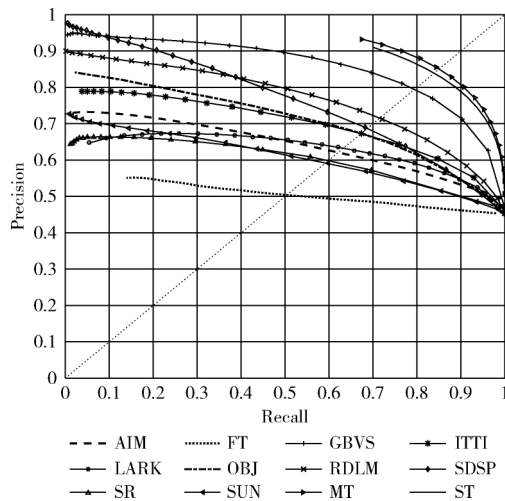


图 5 不同算法在 EmotionROI 数据库上测试的 PR 曲线图

从图 5 可以看出, MT 算法、ST 算法和 GBVS 算法的 PR 曲线值在这些方法中是最高的, MT 算法平衡点值为 0.85, ST 算法和 GBVS 算法的平衡点分别为 0.84 和 0.79。MT 算法的平衡点值最大,这表明多任务情感区域检测方法的性能要优于单任务情感区域检测和显著性物体区域检测方法。

表 1 为本文提出的多任务方法与对比方法在 EmotionROI 测试集测试的 F-measures 分数, F-measures 分数值越高表示算法性能越好。

由表 1 可知, F-measures 值最高的 3 种方法分别是 MT、ST、GBVS 方法。本文提出的基于多任务的情感区域检测方法 MT 的 F_1 、 F_2 、 $F_{0.5}$ 、 $F_{\sqrt{0.3}}$ 分数分别为 0.8525、0.9004、0.8706、0.8650; 单任务的情感区域检测方法 ST 的 F_1 、 F_2 、 $F_{0.5}$ 、 $F_{\sqrt{0.3}}$ 分数分别为 0.8422、0.8951、0.8590、0.8532; 显著性检测方法 GBVS 的 F_1 、 F_2 、 $F_{0.5}$ 、 $F_{\sqrt{0.3}}$ 分数分别为 0.8024、0.8689、0.8076、0.8028。MT 算法在对比算法中的 F_1 、 F_2 、 $F_{0.5}$ 、 $F_{\sqrt{0.3}}$ 分数值都是最高的。这表明本文提出的基于多任务的情感区域检测方法能够更加准确检测图像的情感区域,且检测性能要优于单任务情感区域检测方法和显著性检测方法。

由以上分析可知,对不同的评价指标,基于多任务的情感区域检测方法 MT 的性能都是最好的。这表明本文提

表 1 不同算法的 F-measures 的结果

	F_1	F_2	$F_{0.5}$	$F_{\sqrt{0.3}}$
GBVS	0.8024	0.8689	0.8076	0.8028
RDLM	0.7298	0.8269	0.7854	0.7211
SDSP	0.6990	0.8069	0.7023	0.6959
OBJ	0.6949	0.8099	0.6761	0.6742
ITTI	0.7047	0.8239	0.6739	0.6727
LARK	0.6822	0.8164	0.6328	0.6343
AIM	0.6693	0.8261	0.6210	0.6211
SR	0.6477	0.807	0.5992	0.5998
SUN	0.6429	0.8043	0.5917	0.5926
FT	0.6200	0.7952	0.5175	0.5232
ST	0.8422	0.8951	0.8590	0.8532
MT	0.8525	0.9004	0.8706	0.8650

出的多任务情感区域检测方法的性能优于单任务的情感区域检测方法和显著性检测方法以及物体检测方法。

3.2.2 情感分类结果

为了评估提出的多任务情感分类网络的性能,在 EmotionROI 数据集上将提出的多任务情感分类方法 (MTC) 和单任务情感 (STC) 分类方法与现有的一些方法进行比较分析。EmotionROI 数据集将情感分为生气、厌恶、恐惧、愉快、悲伤和惊喜 6 种情感。

表 2 为多任务情感分类方法和单任务情感分类方法与现有方法的分类准确率。其中 MTC 为多任务的情感分类方法, STC 为单任务的情感分类方法。从表 2 中可以看出,准确率最高的 3 种算法分别为 MTC、STC 和 DeepSentiBank, 其中 MTC 算法的准确率最高为 52.17%。

表 2 不同情感分类方法准确率结果比较

Methods	Accuracy/%
Zhao et al ^[30]	34.84
SentiBank ^[31]	35.24
DeepSentiBank ^[32]	42.53
ImageNet-AlexNet ^[33]	34.26
ImageNet-VGG16 ^[19]	37.26
ImageNet-Res101 ^[34]	40.79
STC	48.70
MTC	52.17

由以上分析可知,基于多任务的情感分类网络的性能要优于单任务的情感分类网络,这也验证了本文提出的多任务网络的有效性,基于多任务情感区域检测网络和情感分类网络的性能相比较单任务的情感区域检测网络和情感

分类网络都有所提高。

4 结束语

本文提出了一种基于多任务深度神经网络的情感区域检测方法, 多任务网络包括情感区域检测网络和情感分类网络两部分, 本文通过加入与情感区域相关的情感分类网络增加了网络共享层的泛化能力从而提高了情感区域网络的检测性能。实验结果表明, 相比较单任务的图像情感区域检测方法、显著性检测方法和物体检测方法, 本文提出的基于多任务深度神经网络的方法对图像的情感区域检测更准确, 且情感分类网络的加入不仅提高了情感区域网络的性能, 也提高了情感分类的准确率, 基于多任务的情感分类网络的性能要优于单任务的情感分类网络的性能。

参考文献:

- [1] ZHAO Sicheng, YAO Hongxun. A survey of image emotion computing [J]. Intelligent Computer and Applications, 2017, 7 (1): 1-5 (in Chinese). [赵思成, 姚鸿勋. 图像情感计算综述 [J]. 智能计算机与应用, 2017, 7 (1): 1-5.]
- [2] Pang L, Zhu S, Ngo CW. Deep multimodal learning for affective analysis and retrieval [J]. IEEE Transactions on Multimedia, 2015, 17 (11): 1-1.
- [3] Lu X, Lin Z, Jin H, et al. Rating pictorial aesthetics using deep learning [J]. IEEE Transactions on Multimedia, 2014, 17 (11): 457-466.
- [4] 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 9th ACM International Conference on Web Search and Data Mining, 2016: 13-22.
- [5] Chen Y, Chen T, Liu T, et al. Assistive image comment robot-a novel mid-level concept-based representation [J]. IEEE Transactions on Affective Computing, 2015, 6 (3): 298-311.
- [6] Li X, Song D, Zhang P, et al. Exploring EEG features in cross-subject emotion recognition [J]. Frontiers in Neuroscience, 2018, 12 (1): 162.
- [7] Hibadullah CF, Liew WC, Jo J. Colour-emotion association study on abstract art painting [C] //International Conference on Machine Learning and Cybernetics. IEEE, 2015: 488-493.
- [8] Cao D, Ji R, Lin D, et al. A cross-media public sentiment analysis system for microblog [J]. Multimedia Systems, 2016, 22 (4): 479-486.
- [9] Zhang L, Tjondronegoro D, Chandran V, et al. Towards robust automatic affective classification of images using facial expressions for practical applications [J]. Multimedia Tools and Applications, 2016, 75 (8): 4669-4695.
- [10] Xu B, Fu Y, Jiang YG, et al. Heterogeneous knowledge transfer in video emotion recognition, attribution and summarization [J]. IEEE Transactions on Affective Computing, 2018, 9 (2): 255-270.
- [11] You Q, Luo J, Jin H, et al. Robust image sentiment analysis using progressively trained and domain transferred deep networks [C] //29th AAAI Conference on Artificial Intelligence, 2015: 381-388.
- [12] Wu L, Liu S, Jian M, et al. Reducing noisy labels in weakly labeled data for visual sentiment analysis [C] //IEEE International Conference on Image Processing, 2017: 1322-1326.
- [13] Zhu H, Li L, Jiang H, et al. Inferring personality traits from attentive regions of user liked images via weakly supervised dual convolutional network [J]. Neural Processing Letters, 2019, 1 (1): 1-17.
- [14] You Q, Jin H, Luo J. Visual sentiment analysis by attending on local image regions [C] //AAAI Conference on Artificial Intelligence, 2017: 231-237.
- [15] Yang J, She D, Lai Y, et al. Weakly supervised coupled networks for visual sentiment analysis [C] //IEEE Conference on Computer Vision and Pattern Recognition, 2018: 18-22.
- [16] Peng KC, Sadovnik A, Gallagher A, et al. Where do emotions come from? Predicting the emotion stimuli map [C] //IEEE International Conference on Image Processing, 2016: 614-618.
- [17] Hou Q, Cheng MM, Hu X, et al. Deeply supervised salient object detection with short connections [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, PP (99): 1-1.
- [18] Zhou B, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization [C] //IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2921-2929.
- [19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [C] //3rd International Conference on Learning Representations, 2015: 1-14.
- [20] Liu S, Deng W. Very deep convolutional neural network based image classification using small training sample size [C] //3rd IAPR Asian Conference on Pattern Recognition. IEEE, 2016: 730-734.
- [21] Harel J, Koch C, Perona P. Graph-based visual saliency [J]. Proc of Neural Information Processing Systems, 2006, 19 (1): 545-552.
- [22] Prakash A, Moran N, Garber S, et al. Deflecting adversarial attacks with pixel deflection [C] //IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 8571-8580.
- [23] Gu Z, Li H, Zhang L. SDSP: A novel saliency detection method by combining simple priors [C] //IEEE International Conference on Image Processing, 2014: 171-175.
- [24] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20 (11): 1-1.

- 1254-1259.
- [25] Seo HJ, Milanfar P. Static and space-time visual saliency detection by self-resemblance [J]. *Journal of Vision*, 2009, 9 (12): 15-15.
- [26] Bruce NDB, Tsotsos JK. Saliency, attention, and visual search: An information theoretic approach [J]. *Journal of Vision*, 2009, 9 (3): 5-5.
- [27] Li G, Yu Y. Visual saliency based on multiscale deep features [C] // *IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 5455-5463.
- [28] Zhang L, Tong M, Marks H, et al. SUN: A Bayesian framework for saliency using natural statistics [J]. *J Vis*, 2008, 8 (7): 32. 1.
- [29] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, et al. Frequency-tuned salient region detection [C] // *IEEE Conference on Computer Vision and Pattern Recognition*, 2009: 1597-1604.
- [30] Zhao S, Gao Y, Jiang X, et al. Exploring principles-of-art features for image emotion recognition [C] // *MM 2014- Proceedings of the 2014 ACM Conference on Multimedia*, 2014: 47-56.
- [31] Borth D, Ji R, Chen T, et al. Large-scale visual sentiment ontology and detectors using adjective noun pairs [C] // *Proceedings of the 21st ACM International Conference on Multimedia*, 2013: 223-232.
- [32] Chen T, Borth D, Darrell T, et al. DeepSentiBank: Visual sentiment concept classification with deep convolutional neural networks [J]. *arXiv preprint arXiv: 1410.8586*, 2014: 1-7.
- [33] Smirnov EA, Timoshenko DM, Andrianov SN. Comparison of regularization methods for ImageNet classification with deep convolutional neural networks [J]. *AASRI Procedia*, 2014, 6 (1): 89-94.
- [34] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C] // *IEEE Conference on Computer Vision & Pattern Recognition*, 2016: 770-778.