滕伟楠. 基于表情识别的文娱节目效果评价方法研究[J]. 智能计算机与应用,2024,14(8):143-150. DOI:10.20169/j.issn.2095-2163.240824

基于表情识别的文娱节目效果评价方法研究

滕伟楠
(西安石油大学 计算机学院, 西安 710065)

摘 要: 本研究旨在探讨一种基于表情识别的文娱节目效果评价方法。首先,针对 Emotion-Domestic 数据集中丰富的表情图像数据,通过构建和训练改进后的 ResNet34 神经网络进行人脸表情分类;其次,为了使算法具备处理多人脸表情的能力,扩充了多人脸表情图像数据到数据集中,通过引入 OpenCV 分类器在自动标定和裁剪出尽可能多的人脸图像,输出每个人脸的表情分类,并将每个观众的表情分类结果映射到原多人脸照片上,以直观地展示观众的情感反应;最后,引入了统计方法,以百分比来量化图片中每种表情的分布情况,为节目效果评估提供客观指标。本文方法应用于多个情节爆发点时,可以形成时序性的节目整体效果评价,从而为文娱节目的效果评价和改进提供了一种新的思路,具有潜在应用前景。

关键词: 表情识别; 文娱节目; 效果评价; ResNet34; OpenCV
中图分类号: TP399 **文献标志码**: A **文章编号**: 2095-2163(2024)08-0143-08

Research on the effect evaluation method of entertainment programs based on expression recognition

TENG Weinan
(School of Computer Science, Xi′an Shiyou University, Xi′an 710065, China)

Abstract: The purpose of this study is to explore an evaluation method based on expression recognition for the effect of entertainment programs. Firstly, according to the abundant expression image data in the Emotion-Domestic dataset, the improved ResNet34 neural network was constructed and trained to classify facial expressions. Secondly, in order to make the algorithm have the ability to process multi-face expressions, the multi-face expression image data was expanded into the dataset, and the OpenCV classifier was introduced to automatically calibrate and crop as many face images as possible, output the expression classification of each face, and map the expression classification results of each audience to the original multi-face photos, so as to intuitively display the audience′s emotional response. Finally, a statistical method was introduced to quantify the distribution of each expression in the picture as a percentage, which provided an objective index for the evaluation of the program effect.the method can be extended to multiple points of the complete program for multi-point effect analysis,provides a new idea for the conception and production,and has potential application prospects.

Key words: expression recognition; entertainment programs; effect evaluation; ResNet34; OpenCV

0 引 言

在当今数字化时代,娱乐产业正经历着电子信息技术推动的前所未有的创新和变革,其中的文娱节目如小品、相声等作为中国文化瑰宝的代表,一直以来以其幽默诙谐的表演风格吸引着广大观众。如何通过受众的反馈来量化地评价小品、相声的娱乐效果,并帮助创作者不断地改进、优化和升级文娱作品,是一个值得深入研究的课题。近年来,随着计算机视觉和深度学习技术的迅速发展,将这一目标转化为现实变得越来越有希望。

表情识别技术作为计算机视觉领域的热点研究方向,致力于从人脸表情中准确地识别情感类别。通过将表情识别技术应用于小品相声演出的效果评价中,可以实时捕捉观众的情感反馈,为创作者和表演者提供有益的反馈指导。通过这种技术的引入,可以不断地深化演员与观众之间的情感互通和交互,强化文娱节目的娱乐效果,进一步提升观众的视听体验。

在探索表情识别技术的应用过程中,深度卷积

作者简介: 滕伟楠(1998-),男,硕士研究生,主要研究方向:计算机视觉,深度学习。Email:tengweinan1998@163.com

收稿日期: 2024-02-29

神经网络的发展成为关键驱动力。从最早的 AlexNet[1] 成为 2012 年 ImageNet[2] 大规模视觉识别挑战赛冠军,其验证了深度卷积神经网络的高效性,做出了非常多的贡献:提出了一种卷积层加全连接层的卷积神经网络结构;首次使用 ReLU 函数[3] 作为神经网络的激活函数;首次使用 Dropout 正则化[4] 来控制过拟合;首次加入动量的小批量梯度下降算法加速了训练过程的收敛;使用数据增强[5] 策略极大地抑制了训练过程的过拟合;利用了 GPU 的并行计算能力,加速了网络的训练与诊断。

ZFNet[6] 与 AlexNet 网络结构基本一致,但其将 AlexNet 的第一个卷积层的卷积核大小改为 7×7,并将第一、二个卷积层的卷积步长都设置为 2、且增加了第三、四个卷积层的卷积核个数。VGG[7] 模型使用了尺寸更小的 3 * 3 卷积核串联来获得更大的感受野,并且放弃了使用 11×11 和 5×5 这样的大尺寸卷积核,去掉了更深,非线性更强,网络参数更少,并且去掉了 AlexNet 中的局部相应归一化(LRN)层。GoogleNet[8] 的出现,提出了一种 Inception 结构,能保留输入信号中的更多特征信息、并且去掉了 AlexNet 的前两个全连接层,采用了平均池化,这一设计使得 GoogleNet 只有 500 万参数,比 AlexNet 少了 12 倍。这些经典网络模型为图像分类任务的性能带来了巨大的提升。

然而,随着网络模型的不断加深,出现了梯度消失[9] 和梯度爆炸[10] 等问题,导致难以训练更深的网络。ResNet[11](Residual Network)作为一种新型的深层卷积神经网络,通过引入残差块的概念,成功地解决了这些问题,允许构建数十甚至上百层的网络。ResNet 的创新在于,其引入了跳跃连接,使得信息可以直接跨越多个层次,从而使网络更易于训练和优化。这种设计方式使得 ResNet 可以构建更深、更复杂的网络结构,为图像分类任务带来了突破性的性能提升。

ResNet 有 ResNet18、ResNet34、ResNet50、ResNet101、ResNet152 几种结构方式。其中 ResNet18、ResNet34 是相对浅层的网络,其余 3 种是更深层的网络。本文将基于 ResNet34 进行表情识别方面的研究。

1 相关理论

1.1 残差网络

残差网络是一种深度卷积神经网络架构,其突出特点是引入了残差模块[11](Residual Blocks),使得网络可以轻松地构建数十甚至上百层的深度,有效地解决了深层网络训练中的梯度消失和梯度爆炸问题,从而实现更好的性能和收敛速度,是本文进行表情识别反馈的核心网络。

如图 1 所示,残差模块的主要思想是恒等映射的思想。假设卷积层学习的变换为 $F(X)$,残差结构的输出是 $H(X)$,则有 $H(X) = F(X) + X$。若图 1 从下至上分别为第 20 层、第 21 层、第 22 层,这样设计的好处是,即使第 21 层没有起到任何作用,那么 21 层的输出和 20 层的输出至少是一样的,不至于降低性能,这便是恒等映射的思想,保留了网络输出等于输入的可能性,可以让正向的信息流一直被传递下去。对于反向传播来说,即使 Conv 层的梯度为 0,也不会出现梯度消失,可以有两条路线进行反向传播,在 Conv 层梯度为 0 时,其也可以回到 X。即残差结构,能够避免普通的卷积层堆叠存在的信息丢失问题,保证前向信息流的流畅,同样,残差结构也能够应对梯度反传过程中的梯度消失问题,保证反向梯度流的通畅。

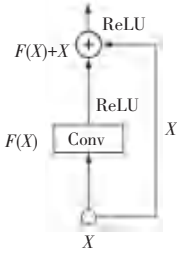


图 1 残差模块恒等映射结果

Fig. 1 Residual module identity mapping results

在 ResNet 的不同版本中,残差结构也存在一定的差异。代表性的有两种残差结构,在 ResNet-18 和 ResNet34 中的残差模块仅包含两个小型的 3×3 卷积层(如图 2),而在 ResNet-50 和 ResNet-101 中的残差模块引入了“瓶颈结构”的残差模块,每个残差模块是由一个 1×1 卷积层、一个 3×3 卷积层和一个 1×1 卷积层组成(如图 3)。如图 4 所示,可以将残差结构看作一种集成模型。

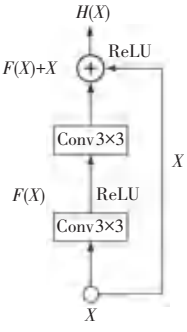


图 2 简单残差结构

Fig. 2 Simple residual structure

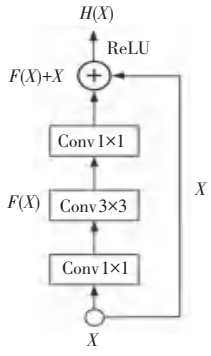


图 3　“瓶颈”结构残差模型

Fig. 3　Residual model of "bottleneck" structure

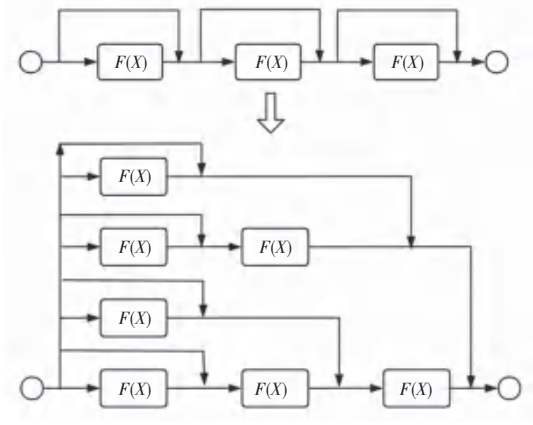


图 4　“瓶颈”残差结构的展开

Fig. 4　Development of the "bottleneck" residual structure

1.2　批归一化

批归一化[12]是一种用于神经网络的正则化技术，其有助于加速网络的训练，并且可以减轻梯度消失和梯度爆炸问题。批归一化通过规范化每一层的输入，使得输入分布更加稳定，从而加速训练收敛过程。如图 5 所示，举例对于 Conv1 的输入 image1 已经是满足某一分布的特征矩阵，但是对于 Conv2 的输入 Feature map 就不一定满足某一分布规律了，而 Batch Normalization 的目的就是使 Feature map 满足均值为 0，方差为 1 的分布规律。

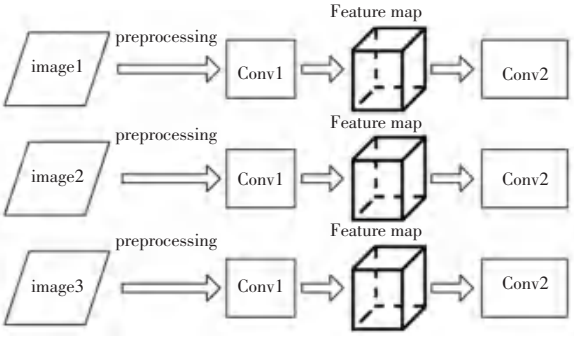


图 5　模型内的传输

Fig. 5　Transfer within the model

批归一化的公式如下：

$$\mu = \frac{1}{m}\sum x_i \tag{1}$$

$$\delta^2 = \frac{1}{m}\sum (x_i - \mu)^2 \tag{2}$$

$$\hat{x} = \frac{x - \mu}{\sqrt{(\delta^2 + \varepsilon)}} \tag{3}$$

$$y_i = \gamma^{\hat{x}_i} + \beta \tag{4}$$

以上是对于输入特征 x，在每个小批次上 BN 的计算公式，其中 μ 是均值，δ^2 是方差，$\hat{x}$ 为标准化流程，将其标准化为均值为 0，方差为 1 的分布，y_i 是尺度和偏移的变换，ε 是一个很小的常数，避免分母为 0，γ 和 β 是可以学习的参数。

在训练过程中，每个小批次的均值和方差会在计算中使用。然而，在推理（测试）阶段，通常使用移动平均来估计整体数据集的均值和方差，以便保持模型在不同批次上的一致性。

批归一化通过标准化输入数据并引入可学习的参数方式，使得神经网络的训练更加稳定和高效。能加速训练收敛，提升模型性能，有助于控制过拟合。

1.3　迁移学习

迁移学习（Transfer Learning）[13]是深度学习领域中的一种重要技术，其通过将从一个任务（源任务）中学到的知识应用到另一个相关任务（目标任务）上，从而加速目标任务的训练并提高性能。其核心思想是利用源任务的知识来初始化目标任务的模型，或者在目标任务上进行微调，以适应不同的数据分布，具体内容如下：

（1）预训练[14]模型：在迁移学习中，通常会使用一个在大规模数据集上进行预训练的模型。例如在 ImageNet 数据集上进行的图像分类任务。这个预训练模型在源任务上学到了有用的特征表示，包含了丰富的视觉信息或其他领域的知识。

（2）冻结层[15]特征提取器：在迁移学习中，通常会将预训练模型的底层（通常是卷积层）称为特征提取器，而顶层则是与源任务相关的任务特定层。通常情况下，特征提取器的权重会在目标任务上保持冻结，以保留源任务学到的特征表示。

（3）微调顶层：在目标任务上，只需微调模型的顶层，这包括最后的全连接层或分类器层。通过微调，模型可以适应目标任务的数据分布，从而提高性能。

（4）领域适应：在某些情况下，源任务和目标任务

可能具有不同的数据分布,这被称为领域间的偏移。为了解决领域适应问题,可以采用领域自适应方法等技术,来减少源任务和目标任务之间的分布差异。

(5)多任务迁移学习:一个模型需要同时处理多个相关任务的情况下,可以通过多任务迁移学习的方法来共享模型的特征提取器,并为每个任务定制特定的顶层来实现多任务学习。

迁移学习有助于避免在目标任务上从零开始训练深度神经网络,大大减少了训练时间的数据需求,本文的方法研究中,便应用到了迁移学习预训练模型。

1.4 多人脸检出方法

对于表情分类,ResNet34 模型只能对单人脸图片的表情进行分类得到结果,但在文娱节目观看中,往往会有多人进行观看,若只能输出单人脸分类结果,则难以得出符合现实场景的评价结论,因此需要一种对多人脸进行检出的方法。

Harr(Haar)级联检测器是一种对象检测算法,其使用 Haar-like 特征进行特征提取,然后通过级联的方式进行快速而准确的目标检测,可以使用预训练的 Haar 级联检测器(Haar Cascades)[16] 来解决此问题。该算法最初是由 Viola 和 Jones 在 2001 年提出的,后被广泛应用于人脸检测等领域,其可以有效解决多人脸框定问题。

开源的 OpenCV 中,实现并优化了 Haar 级联检测器。通过 Haar 级联检测器可以对预处理后的多人脸照片进行人脸目标框定,并将框定的人脸框进行裁剪,分别输入 ResNet34 模型进行预测得到分类结果,再将每个人脸的表情分类结果绘制在框定人脸后的图片中,作为结果进行输出。将 OpenCV 与 ResNet34 结合,OpenCV 完成多人脸框定,ResNet34 完成表情分类,可以更好地实现在多人场景下对文娱节目效果的反馈和评价。

2 基于表情识别的效果评价模型构建

2.1 体系架构及应用思路

本文提出一种如何将表情识别应用到文娱节目效果评价中的应用思路。

首先可通过以下 3 种方式获取有效观众照片:

(1)在需要测试的情节爆发点处拍摄观众照片输入本网络进行预测分类。

(2)在需要测试的情节爆发点处截取观众视频图片输入本网络进行预测分类。

(3)在需要测试的情节爆发点处截取摄像头帧图像输入本网络进行预测分类。

在通过以上方式成功获取观众照片后,事先确认好测试的效果表情以及阈值。例如,本次测试的情节爆发点为笑点情节,则标准即为 happy 表情和 surprise 表情之和大于某阈值,通常设为 50%,即评定该笑点情节效果优秀。

在做好以上准备工作后,便可将待测试照片输入模型,输出一张多人脸表情分类结果,结果包括尽可能多的人脸的位置以及表情分类,并且输出每种表情所占百分比结果。

将百分比结果和事先设定的阈值比较,即可得出该情节爆发点的效果好坏,实现对文娱节目的效果评价。体系架构与应用思路如下:

(1)应用方事先准备:首先应用方通过拍摄情节爆发点处的观众图片、截取情节爆发点处的视频图片、获取情节爆发点处的摄像头帧图片等方式得到待测试图片,并且事先规定待测试表情和阈值。

(2)本文研究方法:首先进行图片预处理,用 OpenCV 框定人脸位置并裁切,其次将人脸裁剪结果输入模型得到分类,最后输出结果及百分比。

(3)单点结果分析:与阈值对比实现单点效果评价。

(4)节目总体分析:将多点结果生成折线图与理想折线图比较,实现对节目总体的分析。

2.2 网络架构设计

本文基于 ResNet34 构建文娱节目效果评价的表情识别模型,需将待测试的情节爆发点处的观众图像进行一系列预处理(如尺寸调整、灰度化、归一化等),使图像适合作为 ResNet34 模型的输入。使用预训练好的网络模型,迁移学习到本任务当中,因为所识别的表情有 7 类,所以将预训练的模型的最终输出层节点数改为 7,并代替原始模型的输出层(全连接层)为新的输出层。最终通过分类器得到分类结果并使用验证集对训练后的模型进行评估,网络结构见表 1。

2.3 损失函数

在基于表情识别的文娱节目效果评价方法研究中,损失函数的设计不仅涉及到模型的训练,更直接地影响着模型对于表情识别任务的优化。通过合理设计损失函数,能够在训练过程中引导模型更好地理解和学习不同表情类别之间的差异,从而使其在实际应用中能够准确地进行分类。考虑到 Emotion-Domestic 数据集中类别数量不均衡的情况,本文采用加权交叉熵损失函数,以更好地适应不同表情类别的训练需求,进一步增强模型在面对少数类别样本时的表现。

表 1　ResNet-34 网络结构

Table.1　Network structure of ResNet-34

Layer name	网络层	参数数量	输出尺寸
Conv1	7×7,64	9 408	112×112×64
Conv2_x	3×3 max pool		56×56×64
Conv3_x	$\begin{bmatrix}3\times3 & 64\\ 3\times3 & 64\end{bmatrix}\times3$	36 928	56×56×64
Conv4_x	$\begin{bmatrix}3\times3 & 128\\ 3\times3 & 128\end{bmatrix}\times4$	73 984	28×28×128
Conv5_x	$\begin{bmatrix}3\times3 & 256\\ 3\times3 & 256\end{bmatrix}\times6$	295 936	14×14×256
Conv6_x	$\begin{bmatrix}3\times3 & 512\\ 3\times3 & 512\end{bmatrix}\times3$	1 180 160	7×7×512
全局平均池化层	Average pool		1×1×512
全连接层			1×1×7

加权交叉熵损失函数在传统的交叉熵损失[17]的基础上,引入了类别权重的概念,以平衡不同类别样本的重要性。针对 Emotion-Domestic 数据集中数量不均衡的类别,本文对于每个类别设置一个相应的权重,这个权重可以基于类别的数量比例、经验设定或其他合理策略。

具体地,对于一个样本,加权交叉熵损失[18]的计算公式如下:

$$L(y,p)=-\sum_{i=1}^{N} w_i \times y_i \times \log(p_i) \qquad (5)$$

其中,N 是类别的数量;y_i 是样本的真实类别标签的第 i 个元素,0 或 1,表示是否属于该类别;p_i 是模型的预测概率分布的第 i 个元素;w_i 是为类别 i 设置的权重。

3　实验结果及分析

3.1　数据集

本研究中,选择 Emotion-Domestic 人脸表情数据集作为主要数据来源。该数据集包含来自 7 个不同情感类别的表情图像,分别为愤怒、厌恶、恐惧、快乐、悲伤、惊讶和中性。这种多样的情感类别使得数据集适用于模型的训练和评估,也与论文目的相一致。Emotion-Domestic 数据集中的图像总数为 54 601张图片,研究中将数据集中的 80%用于训练,10%用于验证,10%用于测试;训练集共有 44 599 张照片,验证集共有 5 000 张照片,测试集共有 5 002 张照片。具体类别分布见表 2。

数据集中的图像来自于真实场景,具有真实的情感表达,因此对于情感分析和表情识别的研究具有一定的代表性。数据集部分实例如图 6 所示。

表 2　Emotion-Domestic 数据集分布

Table 2　Emotion-Domestic data set distribution

类别	样本数量	训练集	验证集	测试集
愤怒	3 771	3 072	349	350
厌恶	3 598	2 970	313	315
恐惧	1 755	1 448	153	154
快乐	26 453	21 614	2 419	2 420
悲伤	3 164	2 588	289	287
惊讶	3 839	3 113	364	362
中性	12 021	9 794	1 113	1 114

图 6　数据集示例

Fig. 6　Data set example

3.2　软硬件环境设置及评估指标

本次实验所采用的操作系统为 64 位 Window10,处理器 Intel(R) Core(TM) i5-8300H,机带 RAM 为 16.0GB;软件环境 Python3.7,深度学习框架 pytorch2.0.1(GPU),编译器为 pycharm2022。

为了进一步提升本模型性能,采用了迁移学习和数据增强的方法,以此来减少模型过拟合,对模型进行 100 个 epoch 的训练,batch_size 设置为 64,并使用加权交叉熵损失,Adam 优化器[19],学习率设置为 0.01 以及学习率衰减[20]。

本文将使用准确率(*Accuracy*)以及损失(*Loss*)评估指标来衡量基于 ResNet34 模型的 7 类表情识别任务的性能。这些指标可以帮助更全面地了解模型在不同情感类别上的分类能力和整体性能。

准确率是评估模型在整个数据集上正确分类的样本比例。其是最常用的评估指标之一,可以通过以下公式计算:

$$Accuracy = \frac{\text{正确分类的样本数}}{\text{总样本数}} \times 100 \quad (6)$$

3.3 实验步骤

论文中重点研究了通过训练 ResNet 网络模型,来对人脸表情进行识别分类,以解决在文娱节目中的情节爆发点处观察观众的情感反馈的目的,实现步骤设计如下:

(1)数据准备。下载 Emotion-Domestic 数据集并将其划分为训练集、验证集、测试集。使用 Pytorch 的 torchvision.transforms 模块对图像进行预处理(包括缩放、归一化和数据增强等)。

(2)模型构建。构建 ResNet34 模型,通过迁移学习使用预训练的权重进行初始化;调整 ResNet-50 的输出层节点为 7,以匹配 7 个情感类别。

(3)使用训练集对模型进行训练,在训练过程中,利用批量梯度下降算法进行参数优化,记录训练损失和准确率,方便后面分析。

(4)使用验证集评估模型的性能,计算验证集上的准确率,用于了解模型的泛化能力和性能。

(5)超参数调整。根据验证集的表现,调整模型的超参数(如学习率、批量大小),以进一步提升模型的性能。

(6)模型测试。使用测试集对最终的模型进行测试,计算在测试集上的准确率,以全面评估模型在实际应用中的性能。

3.4 实验结果分析

为了直观展示训练过程,通过绘制曲线来直观展示训练集和测试集在网络模型上训练的损失(*Loss*)和准确度(*Accuracy*),实验结果如图 7、图 8 所示。

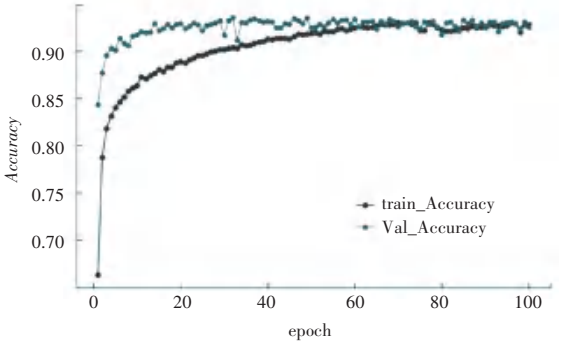


图 7 训练集验证集准确度曲线

Fig. 7 Training set verifies set Accuracy curve

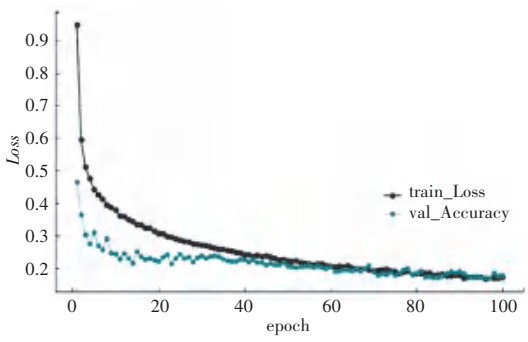


图 8 训练集验证集损失曲线

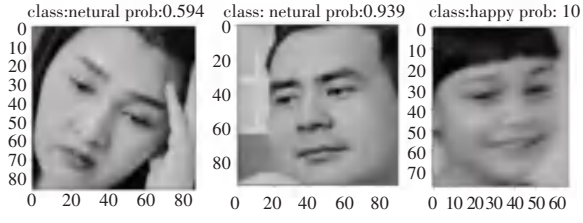
Fig. 8 Training set verification set Loss curve

由图可知,ResNet34 网络在训练 100 个 epoch 之后训练集和验证集的损失达到最低,此时的准确率也达到了 92%左右。

3.5 文娱节目情节爆发点的表情结果评价分析

在测试中,以多人观看节目照片为例进行效果评价。将该照片输入到训练好的模型,模型会尽可能多的对人脸进行检测,输出包括每个检测人脸的预测结果及分数、并输出带有尽可能多人脸的目标框和预测结果的整张照片(如图 9、图 10 所示),并且在图 9 中可以看到,较模糊的表情也可以进行比较准确的分类。

(a) 原图



(b) 预测结果

图 9 示例 1

Fig. 9 Example 1

模型还会输出每种表情所占的百分比,用户可以自行设定阈值类别和数值,如对于煽情类情节表情类别可规定为 sad 及某数值等。本例中,假设在笑点场景下,规定 happy 和 surprise 百分比之和需大于 50%,即可评定该情节爆发点效果优秀,经模型

反馈，在示例 1 图片中，happy 表情所占比例为 33.3%<50%，即评定该笑点情节效果较差；而在示例 2 图片中，happy 表情所占比例为 100%>50%，即评定该笑点情节效果优秀（如图 11）。

（a）原图

（b）预测结果

图 10　示例 2

Fig. 10　Example 2

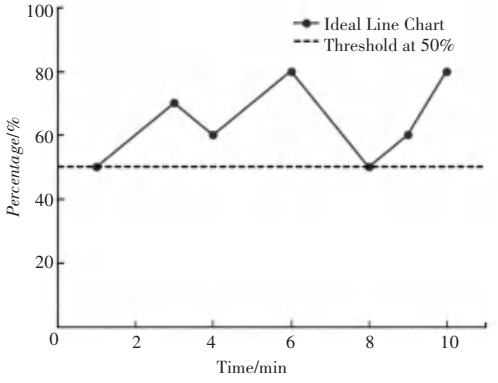


图 11　理想节目效果折线图

Fig. 11　Line chart of ideal program effect

在示例 1 中，中性表情百分比为 66.7%，快乐表情占比 33.33%；示例 2 中，快乐百分比占比为 100%。

3.6　节目效果评价方法研究

上述实验着重研究了对于文娱节目的情节爆发点处的表情分类与效果评价，而对于如何将单点研究转变为多点研究，最后发展为对于文娱节目总体的效果评价，本文提供一种应用思路。针对待效果评价的文娱节目，可事先设置多个待检测的情节爆发点，并且事先假设该情节爆发点的百分比值，作为每个情节爆发点的理想效果。通过以上数据，便可形成横轴为时间轴，纵轴为效果百分比，通过在横轴上标定情节爆发点，以及每个情节爆发点的假设的百分比数值，便可生成理想效果下的节目效果折线图，如图 11 所示。用户事先设置了 7 个可能的情节爆发点，并假定其效果百分比，并且设置阈值 50% 来衡量该情节的效果好坏。应用上述思想到本文研究方法中，截取每个情节爆发点的观众图片，并分别将观众图片输入模型，输出带有尽可能多的人脸表情分类及百分比数值，便可生成在实际模型分类结果下的节目效果折线，输出在理想效果下的节目效果折线图上，如图 12 所示。

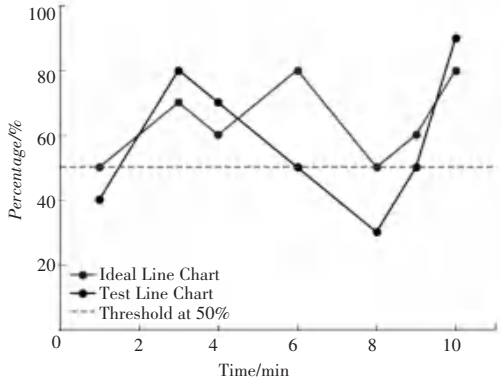


图 12　理想与测试折线图对比

Fig. 12　Ideal and test line chart comparison

得到上述折线图后，本文提供两种思路，以折线图为媒介对节目总体效果进行评价：

1）面积比较法

通过对比理想折现图和实际折线图的面积大小，衡量通过模型分类得到的多点效果是否达到预期。

2）多点比较法

通过统计比较在每个情节爆发点下，理想百分比和实际百分比的大小，得出有哪些情节爆发点达到了预期的效果，又有多少情节爆发点没有达到理想的效果，来对具体的情节爆发点进行针对性的调整，如图 12 所示。在预先设定的 7 个情节爆发点中，有 4 个情节爆发点没有达到预期效果，并且有 2 个情节爆发点没有达到衡量效果好坏的 50% 阈值，便可针对性的对没有达到预期的情节爆发点进行针对性的调整。

通过以上方法，便能以本文研究的单点效果评价延伸到对文娱节目整体的效果评价，提高了本文方法的通用性、实用性，能更好的对文娱节目的效果进行反馈。

4 结束语

在本研究中存在的问题主要有：

(1)将人脸框定分类器在侧面脸和遮挡脸部的检测方面表现不够灵敏。实验中发现在遇到侧脸和带遮挡脸部检测时,目标框的框定存在局限,这一问题部分源于训练数据的不足,因为训练数据主要集中在正面脸部的图像上。此外,模型架构和特征提取方法也可能需要进一步改进,以提高在复杂情况下的性能。

(2)本文使用的 Emotion-Domestic 数据集虽然在表情识别领域具有代表性,但其情感类别和观众情感可能与文娱节目演出的实际情境存在差异,并且数据集仅有 7 种通用表情分类。在实际情况中,每种表情都有多种程度,例如开心表情具有微笑、大笑、狂笑等多种细分种类,该数据则无法更好的展现表情的程度类别。在后续的研究中会收集更多与文娱节目相关的情感数据,以更好地适应实际应用场景,对表情进行更细致的划分。以提高在反馈系统中的应用价值。

本文通过对 ResNet34 模型在观众表情识别效果评价实验中的详细研究,得出了模型在表情识别任务上的准确率、情感分布、观众反应等方面的实际效果。实验结果表明,ResNet34 模型在观众表情识别方面具有良好的性能,能够准确捕捉观众在情节爆发点处的情感反馈。本研究提出了一种在文娱节目演出中应用观众表情识别技术进行效果评价的思路,为文娱节目领域引入了一种新颖的反馈机制。在后续研究中,可以将该单点研究技术进行延伸,如生成情节折线图或者加入对文娱节目中观众的掌声或欢呼的检测等等,更好的对文娱节目效果进行评价反馈。

参考文献

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

[2] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.

[3] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]//Proceedings of the 27th International Conference on Machine Learning (ICML-10). IEEE, 2010: 807-814.

[4] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

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

[6] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//Proceedings of Computer Vision-ECCV 2014: 13th European Conference.IEEE,2014: 818-833.

[7] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[8] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015: 1-9.

[9] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[C]//Proceedings of the 13th International Conference on Artificial Intelligence and Statistics. IEEE, 2010: 249-256.

[10] PASCANU R, MIKOLOV T, BENGIO Y. On the difficulty of training recurrent neural networks [C]//Proceedings of International Conference on Machine Learning. IEEE, 2013: 1310-1318.

[11] 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.IEEE, 2016: 770-778.

[12] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C]// Proceedings of International Conference on Machine Learning. IEEE, 2015: 448-456.

[13] PAN S J, YANG Q. A survey on transfer learning [J]. Transactions on Knowledge and Data Engineering, 2009, 22(10): 1345-1359.

[14] ERHAN D, COURVILLE A, BENGIO Y, et al. Why does unsupervised pre-training help deep learning? [C]//Proceedings of the 13th International Conference on Artificial Intelligence and Statistics. IEEE, 2010: 201-208.

[15] ZAGORUYKO S, KOMODAKIS N. Wide residual networks[J]. arXiv preprint arXiv:1605.07146, 2016.

[16] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2001: I-I.

[17] BISHOP C M, NASRABADI N M. Pattern Recognition and Machine Learning[M]. New York: Springer, 2006.

[18] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS [J]. Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.

[19] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

[20] BOTTOU L, CURTIS F E, NOCEDAL J. Optimization methods for large-scale machine learning[J]. SIAM Review, 2018, 60(2): 223-311.