

融合注意力机制的深度神经网络的人脸表情皮沟识别方法

王长清¹

WANG Changqing

摘要

在复杂光照条件下,由于光照的不均匀性和方向性,皮沟图像的小邻域系和皮沟纹理特征基团会发生形变,影响人脸皮肤的细微纹理特征的提取。为此,提出一种融合注意力机制的深度神经网络方法进行人脸表情皮沟识别。利用神经网络从预处理后的图像中精确捕捉到与表情变化紧密相关的局部特征,由此引入注意力机制,分别关注不同通道和不同位置的特征信息,增强对皮沟纹理的捕捉能力,使网络能够更加聚焦于对表情识别具有关键作用的特征信息,即皮沟的细微纹理信息。并将其输入到全连接层中,通过 SoftMax 函数转化为表情分类的概率分布,实现人脸表情皮沟的准确识别。实验结果表明,该识别方法不仅保持了高分类精度,还显著提升了识别速度,为实际应用场景中的快速、准确人脸表情识别提供了强有力的支持。

关键词

融合注意力机制;深度神经网络;人脸表情皮沟;皮沟识别

doi: 10.3969/j.issn.1672-9528.2025.04.006

0 引言

人脸表情识别作为计算机视觉的一个重要研究领域,近年来在多领域得到了广泛的应用。然而,在实际场景中,由于光照、角度、个体差异等多种因素的干扰,人脸表情识别仍面临诸多挑战。传统的计算机视觉技术往往难以准确捕捉和识别复杂多变的人脸表情特征。为解决这一问题,研究人员提出了多种识别方法。其中,改进型的 MobileNet 的轻量级人脸表情识别方法主要针对传统深度学习模型在人脸表情识别上计算资源消耗大、模型复杂度高的问题而设计的。该方法通过引入 MobileNet V3 的轻量化设计,减少了网络层数和调整通道数,从而显著降低了模型的参数量^[1]。在不同的光照条件下,由于没有针对性地聚焦关键特征,该模型难以区分哪些是真正与表情相关的皮沟特征变化,哪些是光照引起的变化,导致在不同光照场景下的识别结果不稳定。基于残差网络的人脸表情识别方法利用深度残差网络的结构优势,通过引入连续小卷积以及人脸关键点的几何特征融合实现表情识别^[2]。该方法在融合特征时,未对皮沟特征做特殊处理,部分关键信息可能会在融合过程中被稀释,或者淹没于大量其他特征里,这会影响模型捕捉表情细微差别的能力,进而使识别精度受限。

针对现有方法的局限性,本文创新性地提出了一种结合注意力机制的深度神经网络方法,专门用于人脸表情皮沟识别。注意力机制模仿了人脑中的注意分配过程,使得系统能够专注于信息的核心部分,同时忽略无关细节。通过深度融合皮沟特征与注意力机制,旨在进一步提升人脸表情识别的精确度,实现对人脸表情皮沟特征的精确识别和有效分类。

1 人脸表情皮沟图像预处理

在人脸表情皮沟识别过程中,皮沟(皮肤上的微小纹路)图像是重要的信息来源,其中蕴含着丰富的表情特征^[3]。然而,原始图像由于数据源或采集装置的差异,存在分辨率与像素尺寸不一致的问题,并且容易受到噪声、光照不均等因素的干扰,影响识别结果^[4]。为解决这些问题,本文提出了一种对人脸表情图像进行预处理的方法。预处理过程旨在有效减轻光照和阴影的影响,削弱噪声,从而提升图像的整体质量和稳定性。通过这些预处理措施,确保图像中皮沟的清晰度和辨识度,为后续皮沟特征的精确提取和表情识别奠定坚实的基础。

假设原图像为 $I(x, y)$, 目标图像为 $I'(x', y')$, 其中 (x, y) 和 (x', y') 分别表示原图像和目标图像的坐标。双线性插值可以通过公式实现:

$$I'(x', y') = (1-a)(1-b)I(x_1, y_1) + a(1-b)I(x_2, y_1) + (1-a)bI(x_1, y_2) + abI(x_2, y_2) \quad (1)$$

1. 广东茂名幼儿师范专科学校 广东茂名 525000

[基金项目] 广东省教育厅 2023 年度普通高校特色创新项目“基于深度学习的人脸表情识别教学评价反馈系统研究”(2023KTSCX383)

式中： $a = \frac{x' - x_1}{x_2 - x_1}$ ， $b = \frac{y' - y_1}{y_2 - y_1}$ ，且 (x_1, y_1) 和 (x_2, y_2) 是原图像

中与目标图像坐标 (x', y') 相邻的 4 个像素点。

利用方向梯度直方图 (HOG) 算法框架下的 Gamma 校正步骤调整图像对比度，以执行输入图像的颜色空间标准化。Gamma 校正旨在通过特定的数学变换调整图像的对比度，有效减轻由光照不均匀或阴影造成的视觉影响，并削弱噪声的干扰，从而增强图像的整体质量和稳定性^[5-6]。Gamma 校正的数学表达式为：

$$I_{out}(x', y') = c \cdot I(x', y')^\gamma \quad (2)$$

式中： c 是一个常数用于调整亮度； γ 是 Gamma 值，用于调整对比度。

在完成上述预处理步骤后，可得到人脸浅层皮沟图像，如图 1 所示。

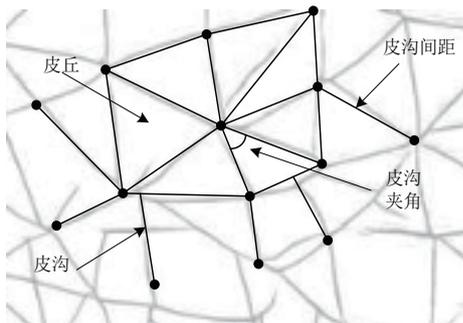


图 1 浅层皮沟图像示例

在图 1 所展示的图像中，通过控制皮沟的间距与夹角来确保图像保持高度的清晰度和可辨识度。皮沟间距被调控在 0.3~0.5 ms 的区间内，而皮沟夹角则大致维持在 25.2° 左右。这些预处理措施为后续皮沟特征提取和表情识别奠定了坚实的基础。

2 利用深度神经网络提取皮沟特征

人脸表情具有多样性和复杂性，这些差异在皮沟特征上得以体现，这些特征对于精确识别表情至关重要^[7]。为此，本研究提出运用深度神经网络（卷积神经网络 CNN）进行皮沟特征提取。深度神经网络凭借其多层卷积和池化操作，能够从预处理后的图像中精确捕捉到与表情变化紧密相关的局部特征，如边缘和纹理等，并在后续的全连接层中将这此特征进行整合，从而全面理解人脸表情^[8]。此外，深度神经网络对于不同光照条件、角度变化和表情差异具有较强的鲁棒性。

本文所提出的深度卷积神经网络 (CNN) 采用了七层结构设计，包括两个卷积层、两个池化层、两个全连接层及一个分类层。基于卷积神经网络的特征提取模型如图 2 清晰展示。

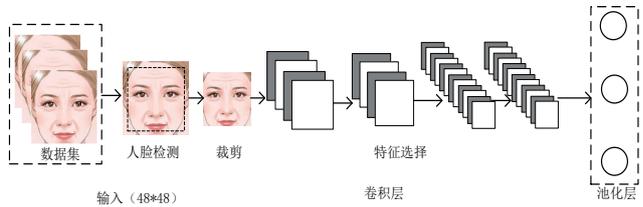


图 2 基于卷积神经网络的特征提取模型

在 CNN 网络中，卷积层用于提取预处理后图像中的局部特征，如边缘、纹理等，这些特征在皮沟特征的提取中尤为重要。通过一组可学习的滤波器对输入数据进行卷积运算，并经过激活函数的转换，卷积层能够生成一系列新的特征映射图^[9]。卷积操作具体表示为：

$$Y = \sigma(W \cdot I_{out}(x', y') + b) \quad (3)$$

式中： W 是滤波器； b 是偏置项； σ 是激活函数； Y 是卷积操作后生成的新特征映射图。

池化层对卷积层输出的特征映射图进行下采样，以降低特征图的维度并减少计算量，同时尽可能地保留关键信息。池化窗口的大小为 $k \times k$ ，步长为 s ，则最大池化操作可以表示为：

$$Y_{pool} = \max_{pooling}(Y, k, s) \quad (4)$$

式中： Y_{pool} 是池化后的特征映射图； $\max_{pooling}$ 函数在每个 $k \times k$ 的窗口内选择最大值，并按照步长 s 移动窗口。

通过卷积层和池化层的交替使用，CNN 网络能够逐步提取出人脸图像中的皮沟特征以及其他关键信息。这些特征在后续的全连接层中进一步整合，形成对人脸表情的全面理解，并最终用于表情识别等任务。

3 人脸表情皮沟识别

在人脸表情皮沟识别过程中，利用 CNN 网络提取出的人脸图像中的皮沟特征以及其他关键信息各自包含了表情识别的不同方面。皮沟特征反映了皮肤纹理的细微变化，对于捕捉微妙的表情变化至关重要；而其他关键信息，如边缘、形状等，则提供了人脸的整体结构和形态信息^[10]。将这些特征进行融合，可以综合利用各种信息，提高表情识别的准确性和鲁棒性。在特征融合过程中，不同特征对于表情识别的贡献度是不同的。一些特征可能更加关键，而另一些特征则可能相对次要。为了优化特征融合的效果，本研究引入了注意力机制。注意力模块可以分别关注不同通道和不同位置的特征信息，从而增强对皮沟纹理的捕捉能力，以提高表情皮沟识别的准确性和鲁棒性。

设注意力机制层的权重为 W_{att} ，偏差为 b_{att} ，则注意力权重 A 可以表示为：

$$A = \sigma(W_{att} \cdot Y_{pool} + b_{att}) \quad (5)$$

在特征融合阶段，多尺度卷积核的集合为 $\{K_1, K_2, \dots, K_n\}$ ，其中 K_i 表示大小为 $i \times i$ 的卷积核。

对于每个卷积核 K_i ，多尺度特征融合后的注意力参数 A_i 可以表示为：

$$A_i = \sigma(W_{att} \cdot \text{Conv}(K_i, Y_{pool}) + b_{att}) \quad (6)$$

式中： $\text{Conv}(K_i, Y_{pool})$ 是使用卷积核 K_i 对特征图进行卷积操作经过这一系列的处理，获取人脸表情特征的注意力参数^[11]。

为了融合这些多尺度注意力参数，采用加权平均方法，则最终融合的注意力权重可以表示为：

$$A'_i = \sum_{i=1}^n w_i \cdot A_i Y_{pool} \quad (7)$$

式中： w_i 是每个尺度特征的权重，可以通过学习得到。

在得到融合的注意力权重后，将其与原始特征图进行逐元素相乘，可得到融合了注意力机制的特征图，形成更加全面和准确的表情特征表示。这一过程表示为：

$$Y'_{pool} = A'_i \otimes Y_{pool} \quad (8)$$

式中： \otimes 表示逐元素相乘。

综上，经过注意力机制融合处理的图像，包含丰富的表情皮沟特征^[12]。将这些特征输入到全连接层进行进一步处理。随后，这些特征通过 SoftMax 函数被转化为表情分类的概率分布^[13]。设分类器的权重为 W_{cls} ，偏差为 b_{cls} ，则输入图像 Y'_{pool} 属于第 c 类表情的概率 $P(c|Y'_{pool})$ 可以表示为：

$$P(c|Y'_{pool}) = \text{soft max}(W_{cls} \cdot Y'_{pool} + b_{cls})c \quad (9)$$

在上述分类阶段，经过充分训练的模型会对输入图像的特征表示进行细致的评估，以确定其最有可能属于上述七种不同表情类别中的哪一种。根据最大概率原则，模型会选择具有最高概率的表情标签作为最终的识别结果。

4 实验分析

4.1 实验准备

本实验在某大学人工智能与机器学习实验室展开，该实验室配备了高性能计算机、专业图像处理软件和先进的人脸表情识别系统。服务器型号为 Dell PowerEdge R740，配备了 Intel Xeon Scalable Gold 6248 CPU (2.5 GHz, 20 核心)，128 GB DDR4 ECC REG 内存，以及四块 NVIDIA Tesla V100 SXM2 32 GB GPU，操作系统为 Ubuntu 18.04 LTS。这台服务器将作为深度学习模型的训练和推理平台。实验过程中，CNN 层数为 5 层，卷积核大小为 3×3 ，步长为 1，激活函数为 ReLU。在全连接层部分，共有 2 层，神经元个数分别为 128 和 7 (对应 7 种表情类别)。

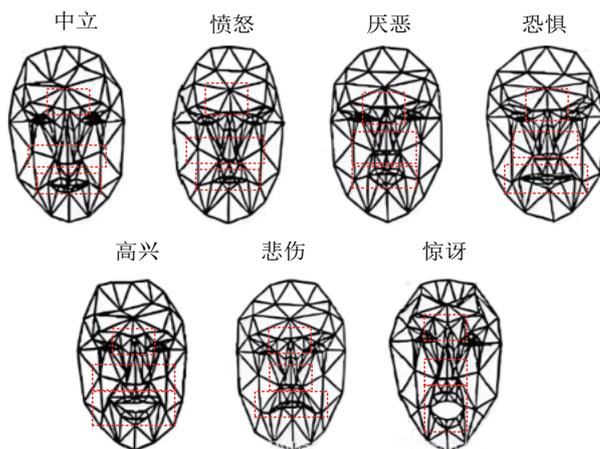
在实验过程中，选择 CK+ 数据集作为研究对象。该数据集共包含 5 000 张图片，将其中 2 900 张图片用于训练集、

训练模型；剩余的 2 100 张图片作为测试集，用于评估模型的性能。CK+ 数据集内容丰富，涵盖了愤怒、厌恶、恐惧、高兴、悲伤、惊讶以及中立等多种表情类别，每种表情类别均包含 300 张图片。

愤怒、厌恶、恐惧、高兴、悲伤、惊讶和中立的表情皮沟识别结果如图 3 所示。



(a) 表情样例



(b) 识别结果

图 3 表情识别结果

从图 3 可以看出，所提方法可以有效地识别出人脸面部的皮沟特征，并基于所识别出的特征，辨别出人脸的表情。

4.2 实验结果与分析

在人脸表情皮沟识别的研究中，对比本文提出的识别方法与现有方法——改进型的 MobileNet 的轻量级人脸表情识别方法(方法 1)和基于残差网络的人脸表情识别方法(方法 2)在分类精度以及识别速度上的表现。

4.2.1 ROC 曲线

通过本次实验，对比了本文提出的人脸表情皮沟识别方法与另外 2 种方法在 ROC 曲线上的表现。对比结果如图 4 所示。

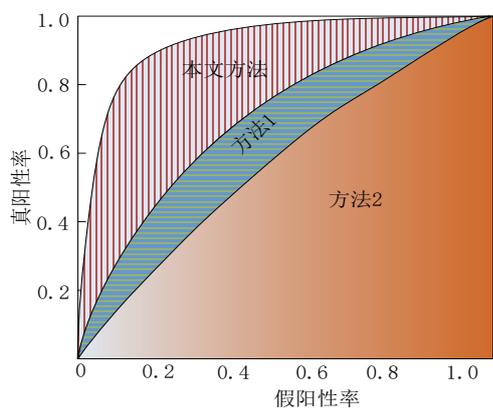


图4 三种方法的 ROC 曲线

如图4所示,本文提出的识别方法在ROC曲线上的表现显著优于方法1和方法2。在实现高TPR的同时,保持了较低的FPR,表明本文方法能够在准确识别出真正的人脸表情皮沟特征的同时,有效减少误判的可能性。这一表现反映了本文方法具有更高的分类精度和更好的泛化能力。这主要在于通过引入注意力机制,本文方法能够有效地突出关键特征并抑制无关信息,使得模型能够更专注于那些对于分类最为关键的人脸表情皮沟特征。这一步骤不仅提高了模型的分类精度,还增强了模型的泛化能力,使其在面对不同人脸和不同表情时能够保持稳定的性能。

4.2.2 识别速度

本次实验对比了本文方法、方法1和方法2在7种不同表情类别(愤怒、厌恶、恐惧、高兴、悲伤、惊讶和中立)上的平均识别时间。对比结果如图5所示。

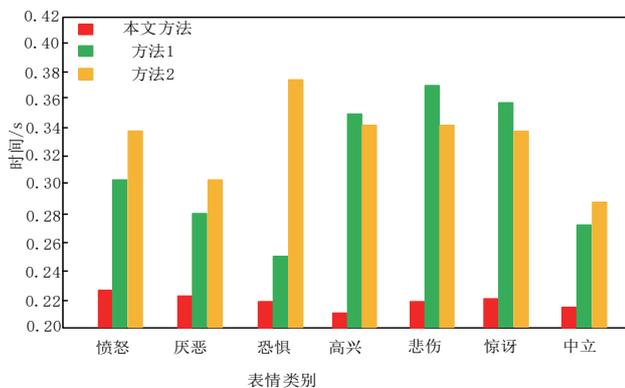


图5 三种方法的识别速度

分析图5可知,本文方法在识别速度上具有显著优势。对于所有测试的表情类别,本文方法的平均识别时间均显著低于方法1和方法2。特别是在高兴表情的识别上,本文方法的速度提升尤为突出。结果表明,本文方法通过引入注意力机制,有效突出了关键特征并抑制了无关信息,从而显著减少了模型所需处理的信息量,进而提高了识别速度。

5 结语

本文所提的融合注意力机制的神经网络方法,在人脸表情皮沟识别领域取得了显著的效果。注意力机制的引入使模型能精准聚焦人脸表情的关键皮沟特征,极大提升了识别精度。实际测试表明,该方法在不同光照、角度和表情变化场景下表现优异,克服了传统方法难以解决的干扰问题。

参考文献:

- [1] 吴振荣,邱卫根,张立臣.改进型的 MobileNet 的轻量级人脸表情识别方法[J].信息技术,2024(8):44-50.
- [2] 赵艳芹,姜昊.基于残差网络的人脸表情识别方法[J].黑龙江科技大学学报,2024,34(3):469-473.
- [3] 焦阳阳,黄润才,万文桐,等.基于图像融合与深度学习的人脸表情识别[J].传感器与微系统,2024,43(3):148-151.
- [4] 李星燃,张立言,姚树婧.结合特征融合和注意力机制的微小表情识别方法[J].计算机科学,2022,49(2):4-11.
- [5] 卢晋.基于自建表情数据集和深度神经网络的人脸表情识别方法[J].深圳职业技术学院学报,2022,21(5):3-8.
- [6] 林德贵,邱富杭,余清清.融合 Gamma 校正与多任务卷积神经网络的人脸识别[J].信息技术与信息化,2023(11):179-182.
- [7] 刘经纬,许少伟.基于深度学习的图像中人脸检测与表情识别[J].信息系统工程,2024(7):70-73.
- [8] 岳也,温瑞萍,王川龙.带有特征信息卷积神经网络的人脸识别算法[J].工程数学学报,2024,41(3):410-420.
- [9] 蒋文豪.多尺度注意力机制下的人脸表情识别算法设计[J].现代计算机,2024,30(4):29-33.
- [10] 卢莉丹,夏海英,谭玉枚,等.注意力引导局部特征联合学习的人脸表情识别[J].中国图象图形学报,2024,29(8):2377-2387.
- [11] 雒翠萍,聂志刚.基于深度学习的人脸表情识别研究与应用[J].甘肃科技,2022,38(1):23-25.
- [12] 王威,许玉燕,王新,等.融合注意力机制和轻量级卷积神经网络的胸部 CT 影像分类方法研究[J].中国生物医学工程学报,2024,43(4):429-437.
- [13] 李校林,张鹏飞,董玉晖,等.基于改进型 Softmax Loss 损失函数的面部表情识别[J].计算机仿真,2023,40(5):234-256.

【作者简介】

王长清(1981—),男,河南信阳人,硕士,副教授,研究方向:人工智能及大数据应用。

(收稿日期:2024-11-28)