付军秀, 袁野, 徐康, 等. 基于轻量化 PFLD 的人形机器人面部和头部的镜像行为方法研究[J]. 智能计算机与应用,2026,16(7):49-56. DOI:10.20169/j.issn.2095-2163.24121101

基于轻量化 PFLD 的人形机器人面部和头部的镜像行为方法研究

付军秀[1,2], 袁 野[1,2], 徐 康[1,2], 傅柯婷[1,2], 任钦泽[1,2], 刘 娜[1,2]
(1 上海理工大学 机器智能研究院, 上海 200093; 2 上海理工大学 健康科学与工程学院, 上海 200093)

摘 要: 在自然的人机互动中,人形机器人能够精确地模仿人类的面部表情和头部动作是非常关键的。为了实现这种镜像行为,本文提出了一种基于轻量化 PFLD(A Practical Facial Landmark Detector)网络的实时人形机器人头部和面部镜像行为方法研究。首先,设计了一款由 14 个舵机控制的仿真机器人头,使其能够展现丰富的面部表情和头部姿态。其次,对 PFLD 模型进行了优化和轻量化,同时采用 FSA-Net 来检测头部姿态。通过建立面部关键点和头部姿态与舵机之间的映射关系,实现了人类头部动作到机器人头部动作的实时镜像。实验结果显示,优化后的模型在关键点检测的准确性提升 15.4%、模型参数减少 15.5%,面部表情相似度差距在 20%范围内,头部运动的均方误差在 3.5 以内。该研究为人形机器人头部动作的精确模仿提供了有效方法,显著提升了人机交互的自然性、实时性及准确度。

关键词: 镜像行为; 人形机器人; 轻量级网络; 面部表情; 头部动作
中图分类号: TP242 **文献标志码**: A **文章编号**: 2095-2163(2026)07-0049-08

Research on the mirror-behavior method for humanoid robot facial and head movements based on lightweight PFLD

FU Junxiu[1,2], YUAN Ye[1,2], XU Kang[1,2], FU Keting[1,2], REN Qinze[1,2], LIU Na[1,2]
(1 Institute of Machine Intelligence, University of Shanghai for Science and Technology, Shanghai 200093, China;
2 School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: In natural human-machine interaction, the ability of humanoid robots to accurately mimic human facial expressions and head movements is critical. This study proposes a real-time method for humanoid robot head and face mirroring behavior based on a lightweight PFLD network. A simulated robot head controlled by 14 servos is designed to exhibit diverse facial expressions and head postures. The PFLD model is optimized and made lightweight, and FSA-Net is employed to detect head posture. A mapping relationship is established between facial key points, head posture, and servo movements, enabling real-time mirroring of human head movements. Experimental results show a 15.4% improvement in key point detection accuracy and a 15.5% reduction in model parameters. The facial expression gap is within 0.2, and the mean square error of head movements is within 3.5. This method provides an effective approach for precise imitation of humanoid robot head movements, enhancing the naturalness, real-time performance, and accuracy of human-computer interaction.

Key words: mirror behavior; humanoid robots; lightweight network; facial expression; head movement

0 引 言

随着人机交互技术的快速发展,仿真人头在模仿人类精细表情方面的能力变得尤为重要。当摄像头捕捉到人类表情并使机器人实时模仿时,这不仅显著提升了沟通的有效性,还在互动中增强了信任感[1]。因此,使机器具备准确理解和模拟人类表情的能力,对于提升人机交互的质量至关重要。在交互过程中,为实现情感体验的增强和情绪状态的精准表达,机器人需要具备实时生成逼真的头部动作与精细面部表情的能力[2]。这一技术广泛应用于人性化机器人、虚拟主持人等场景[3],并在娱乐、教育、医疗等领域展现出巨大的应用潜力。同时,还将深刻影响未来社会的人机互动模式及关系演变[4]。

作者简介: 付军秀(2001—),男,硕士研究生。主要研究方向:计算机视觉。
通信作者: 刘 娜(1985—),女,博士,硕士生导师。主要研究方向:机器视觉,深度学习。Email:15538361047@163.com。

收稿日期: 2024-12-11

综上所述,在伺服约束和硬件限制下,开发能够实时生成镜像表情和头部运动的技术,对推动更加智能化与人性化的产品设计具有重要意义。

近年来,镜像行为技术取得了显著进展,尤其是在数字人领域。通过数字人技术,能够在虚拟环境中模拟人类的面部特征、表情、手势和姿势变化[5-6]。这种技术依赖于对虚拟人物模型的精确控制,如在数字人中,用户可以调整每个顶点和面片的状态,从而实现细致的动作和表情变化。然而,数字人和机器人之间仍存在较大差距。与虚拟模型不同,仿人机器人需要依靠伺服舵机来驱动头部运动,但由于机器人内部空间的限制,无法为每个细节配置相应的舵机[7-8]。因此,如何在硬件限制下实现高度逼真的面部表情和头部动作的镜像,成为人形机器人头部设计中的一个关键问题。

除了机械结构的硬件限制外,提高图像中面部关键点提取的精度至关重要,因为这将直接影响机器人头部镜像行为的真实性。传统的人脸特征点提取方法,如基于点分布模型的方法,计算效率有限且对噪声敏感,鲁棒性差[9-10]。线性回归方法使用的特征点较少,参数难以调整。相比之下,深度学习技术在人脸特征点提取方面具有更强的学习能力和适应性。本文提出了采用修改后的 PFLD 模型[11]和 FSA-Net 技术结合的方案,以增强面部特征提取与头部姿态检测的准确性,并为实现更为真实和流畅的机器人镜像行为奠定基础[12]。

此外,在面部表情和头部动作的镜像行为中,实时高效的交互更符合人们的期望。因此,除了考虑到准确性、还要考虑到实时性,镜像行为是否流畅直接影响到了人机交互的体验。然而,因为处理器的内存和计算能力的限制,过大的深度学习模型会导致内存不足,从而产生高延迟的问题[13]。

针对上述问题,本文提出了基于轻量级网络的机器人面部和头部的镜像行为。首先,研发了一个仿真人头,皮肤使用硅胶材质,以确保柔软和灵活性。同时,精密机械结构和伺服控制技术保证了模仿行为的准确性。其次,使用轻量级模块替换普通的卷积模块对深度学习模型进行了一定的轻量化,并且使用 Kolmogorov-Arnold Networks(KAN)替代全连接层用于减少参数量,但考虑到轻量化会导致精确度降低,所以本文采用了特征金字塔结构,对特征进行更好的提取,通过此方法能够实时准确检测 98 个面部特征点的坐标[14]。接下来,使用 FSA-Net 检测头部姿态[15]。最后,对检测到的人脸特征点和头部姿态进行变换和归一化,建立与仿人机器人头部最优舵机位移的映射关系。

1 相关工作

一个自然有效的机器人表情模仿行为对人类的面部关键点识别的准确性和实时性有着比较高的要求[16]。在以往的研究中,Xia 等学者[17]提出了一种基于稀疏局部补丁的 Transformer 面部特征点检测。Zhang 等学者[18]提出了一种混合训练中提取人脸特征深度的方法。但是以上在嵌入式设备上运行时,因为模型参数量较大,运算次数较多,从而影响仿真头部的镜像的实时性。因此,本文中的面部特征点检测是基于轻量级的深度学习模型,在保证精度有一定的提升的情况下,对网络模型的计算量和参数量进行压缩,使得模型能够在相同的嵌入式设备上有较好的实时性。本文采用了权重剪枝的方法直接将模型中某些权重的数值设为零,从而减少参数数量[19]。由于模型中全连接层的参数量比较大,所以本文使用了 KAN 网络用于替换全连接层,并在网络设计中加入轻量化模块,进一步减小模型尺寸。深度学习网络保证了人脸特征点识别的准确性和鲁棒性,经过减少网络的参数可以加快面部特征点的检测速度,从而保证了从人到人形机器人镜像行为的高精度和实时性。

本文使用了深度学习模型来提取 98 个面部特征点。然后,根据人体面部运动的规律,找到运动部分相对应的关键点,使用部分相对应的关键点的曲率,同时计算并归一化这些特征点的欧式距离。最后,将计算的结果映射到机器人头的舵机位移上,从而达到面部表情模仿效果。这种基于面部特征点之间欧氏距离的伺服驱动方法类似于人体肌肉控制,可以更好地模仿面部细节的变化[20]。同时,使用深度学习来对头部的欧拉角进行推断,从而完成头部姿态的模仿。

2 仿真人形头部的镜像行为实现

本文开发了一个仿真人形机器人头作为镜像面部表情和头部运动行为的平台。图 1 为本文方法的流程框架,首先使用相机捕捉图片,然后通过 2 个深度学习模型分别计算出 98 个面部特征点的坐标和头部的俯仰角。计算某些面部特征点的欧式距离和头部俯仰角进行归一化,最后将归一化的结果映射到 14 个舵机的位移上,从而实现仿真人头的镜像行为。

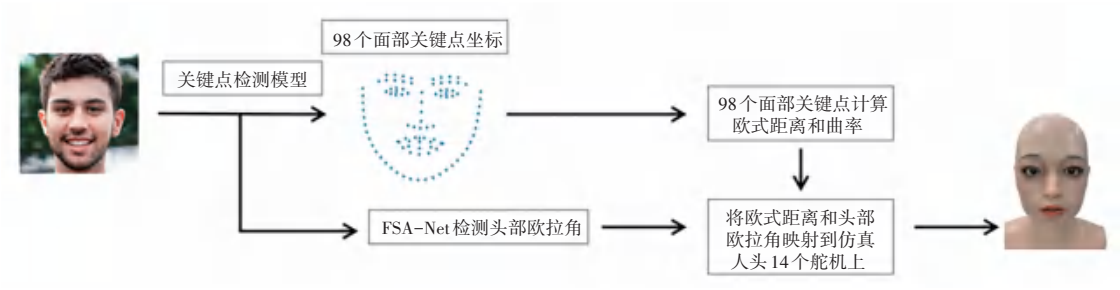


图 1　仿真人头的镜像行为流程

Fig. 1　Flow chart of simulated robot head mirror behavior

2.1　仿真人形头部的设计和控制

为了追求自然的人机交互，开发了一个仿人机器人平台，真实地模仿人类的面部表情和头部运动。该平台集成了带有微处理器的软皮机器人头部、先进的伺服控制系统和精密的机械结构。还可以准确地复制人类面部运动和头部的运动，从而展示各种各样的面部表情和头部动作。图 2 是硬件流程图。面部数据通过摄像头捕获并传输到电脑设备。使用轻量化的深度学习模型，检测 98 个面部关键点坐标和头部俯仰角以确定最佳伺服位置，然后通过串口传输给 STM32F103C8T6 微处理器。随后，微处理器驱动相应的伺服系统，使机器人头部能够动态地反映人类的面部表情和头部运动。



图 2　硬件流程图

Fig. 2　Flow chart of the hardware

2.1.1　仿真机器人头部设计

仿真机器人头部平台可分为头部框架和内部模块。头部框架是根据真实人脸进行 3d 打印的，外部与仿真硅胶皮肤紧密贴合，赋予人形外观和触感，头部内部的空心空间用于放置机械控制结构。机械控制结构由绳结构和半球结构组成。采用一对半球结构控制眼睑的开合，上下眼睑相互偏离的角度为 50°~80°，同时眼睛是高清摄像头、可以获取仿真机器人头所看到的画面。眉毛的上下运动、脸颊的凹陷和凸起都是通过舵机牵拉不同的钢丝绳引起硅胶变形实现的。眉毛上下运动的最大范围为 5 mm，脸颊凹陷和凸起的最大范围也为 5 mm。使用 3 个舵机的控制实现头部点、摇、摆的运动，驱动头部实现头部运动跟踪。

2.1.2　机器人头部平台的控制

部署在机器人头部的 STM32F103C8T6 微处理器通过串口与电脑通信。微处理器从电脑获得归一化的最优舵机旋转角度。由于伺服的旋转角度由高阶脉宽的持续时间决定，微处理器将归一化位移数据的值对应于高阶脉宽的持续时间来驱动伺服，从而实现归一化位移数据到伺服位移的映射关系。表 1 给出了仿真机器人头的舵机的功能。机器人头部装有 14 个舵机，用于控制嘴巴、眼球、嘴角、眼睑、眉毛以及颈部的旋转角度。这些舵机允许独立或联合控制左右眼睑、眉毛和嘴角。尽管仿人机器人的自由度不及人类，但却能够展示面部表情和头部各方向的旋转，为机器人镜像行为生成提供了稳定可靠的可视化平台。

表 1　头部舵机编号及相应控制

Table 1　Head steering gear number and corresponding control

舵机	作用	舵机	作用
S_1	左闭眼 / 睁眼	S_8	右闭眼 / 睁眼
S_2	左微笑	S_9	右微笑
S_3	左眉毛	S_{10}	右眉毛
S_4	点头	S_{11}	摇头
S_5	摆头	S_{12}	张 / 闭嘴
S_6	左眼球上下	S_{13}	右眼球上下
S_7	左眼球左右	S_{14}	右眼球左右

2.2　轻量化深度学习模型

2.2.1　轻量化人脸关键点检测模型

(1) 模型总体设计。在面部表情模拟中，人脸关键点的准确度直接影响了表情模拟的相似性。同时，模型的处理效率也直接关系到表情模拟的流畅性。所以本文使用了一个轻量级的面部关键点检测模型，模型用于检测面部的 98 个关键点的坐标，通过计算将坐标映射到舵机上从而实现面部表情模仿。图 3 为轻量化的人脸关键点检测模型的总体架构。在网络模型中，使用了深度可分离卷积、KAN 网络、权重剪枝方法对模型进行了轻量化，加快了模型的训练速度和推理速度，在精度稳定不变的情况下减小了模型的尺寸。

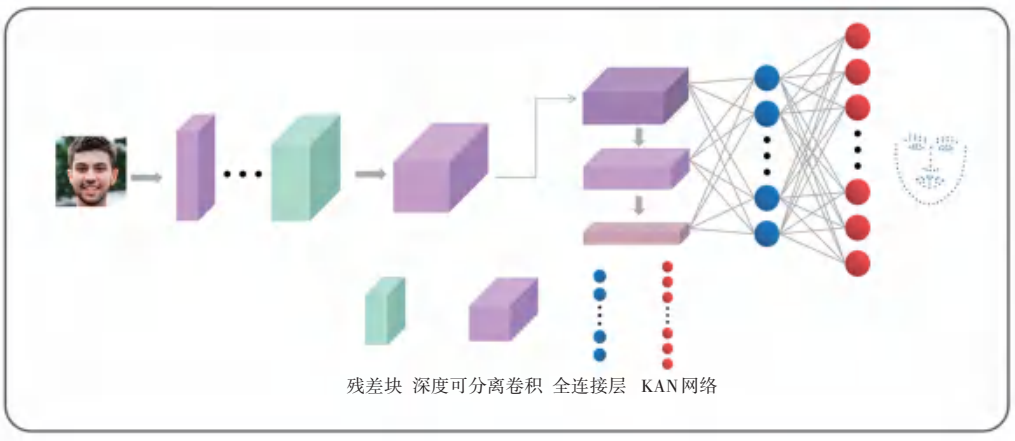


图 3 轻量化 PFLD 模型结构

Fig. 3 Lightweight PFLD model structure

轻量级人脸特征点识别网络的结构包含 13 个残差块、2 个标准卷积层和 1 个 KAN 网络层,其中残差块中使用的是深度可分离卷积。该轻量级网络的计算复杂度,可以用下式来描述:

$$C = C_{\text{conv}_1} + C_{\text{conv}_2} + C_{\text{KAN}} + \sum_{n=1}^{13} C_{\text{conf}}^{n} \quad (1)$$

其中,C_{conv_1} 和 C_{conv_2} 表示 2 个标准卷积的计算复杂度;C_{KAN} 表示 2 个完全连接层的计算复杂度;C_{conf}^{n} 表示第 n 个残差块的计算复杂度。

(2)权值剪枝方法。权值剪枝旨在降低训练的计算成本,提升模型的运行速度。本方法通过对网络各层连接权值进行全局评估,根据评估结果对网络各层赋予权重,再通过迭代剪枝和调整过程优化网络结构。权重剪枝流程如图 4 所示。

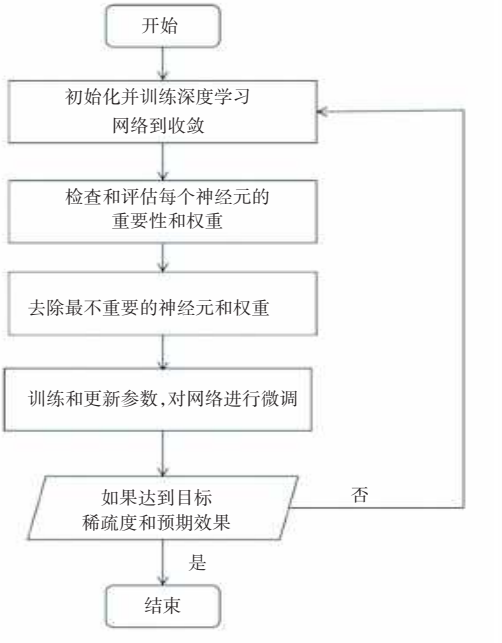


图 4 权重剪枝流程图

Fig. 4 Flowchart of weight reduction

在初始阶段,网络通过定期训练来学习加权连接。模型训练至收敛状态后,开始评估连接权值的重要性,为剪枝操作做好准备。接下来,对权值低于设定阈值的连接进行剪枝,将这些连接从网络中移除,从而将原本密集的网络转化为稀疏网络。

剪枝操作完成后,为了避免对网络精度产生负面影响,需要对剩余的稀疏连接进行再训练。同时,设定稀疏性目标,以保证网络在剪枝后的稀疏度与准确性。每次剪枝后,系统会重新计算当前的稀疏度目标,并检查是否达到预期的稀疏度。

在剪枝计划结束时,如果达到预期的稀疏度目标,就会得到一个更小的网络。每轮剪枝后,保留连接权值以便在发生错误剪枝时能够快速初始化网络。经过多次迭代和调整后,最终可以确定在保持所需精度的情况下连接数最少的网络结构,从而得到一个高效的轻量级深度神经网络,适用于快速检测面部特征。经过权重剪枝模型大小从 10.1 MB 变成了 6.5 MB,并且精度保持不变。

2.2.2 FSA-Net 模型的集成与应用

在本研究中,面临的关键挑战是实现对仿真人头模型头部姿态的精确估计,这要求模型不仅要有高准确度,还要具备实时性。本文选择了 FSA-Net 模型作为解决方案的核心组件。

FSA-Net 是一种先进的深度学习模型,专为头部姿态估计任务设计。该模型由研究团队基于最新的深度学习理论和技术构建,旨在通过精确地识别头部的 Roll、Pitch 和 Yaw 角来改善人机交互和虚拟现实应用。

2.3 从面部关键点和头部 RPY 角到舵机的映射关

为了更准确、快速地提取人脸特征点,采用轻量级网络对 98 个人脸特征点的二维坐标进行跟踪。

这些点构成了面部关键点集合 P：

$$P = \{p \mid p_i = (x_i, y_i), 0 \leqslant i \leqslant 98\} \tag{2}$$

其中，(x_i, y_i) 表示第 i 个人脸特征点的坐标。此外，仿人机器人通过伺服位移实现镜像行为，伺服位移矢量 S 可以表示为：

$$S = (S_1, S_2, \cdots, S_m), \quad m = 10 \tag{3}$$

轻量化网络实时跟踪的当前关键点集合 P。因此，S 和 P 之间的关系可以表示为：

$$S = F(P) \tag{4}$$

头部舵机分布如图 5 所示。一个舵机可以被多个面部特征点控制，机器人头部平台的每个部分都由一个独立的伺服控制。

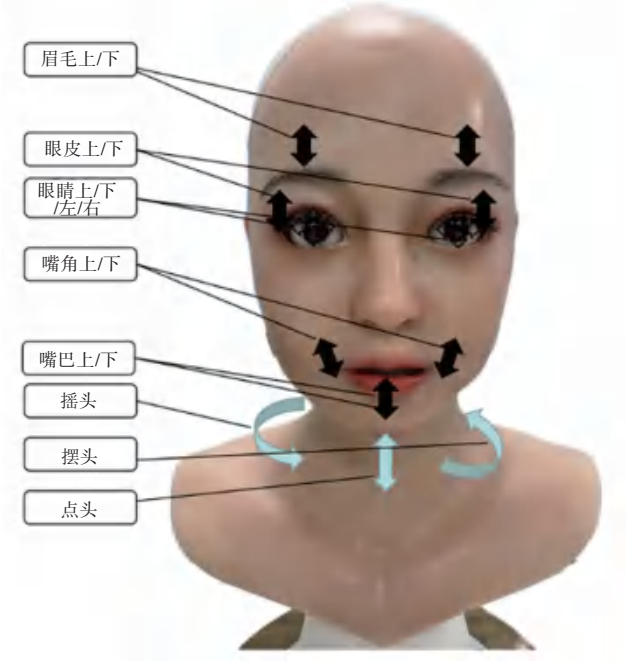


图 5　头部舵机分布图

Fig. 5　Head steering gear distribution

对于每个舵机的运动控制，计算其相对应的关键点的曲率。曲率(Curvature)通常用来描述曲线或曲面在某一点的弯曲程度。例如，当嘴巴由紧闭到张开的过程中，曲率逐渐增大。曲率的计算公式如下：

$$K = \frac{|x'(t)y''(t) - y'(t)x''(t)|}{|x'(t)^2 + y'(t)^2|^{\frac{3}{2}}} \tag{5}$$

用嘴巴举例，由于嘴部的关键点是离散的，所以计算曲率首先确定坐标点的顺序，确保坐标点是按照在曲线上的顺序排列的。然后，使用多项式进行拟合，本文采用的是二次多项式来拟合这些点，其二次多项式为：

$$y = ax^2 + bx + c \tag{6}$$

其中，a、b 和 c 是多项式的系数，本文通过关键点坐标来确定这些系数。要找到最佳拟合的多项式，本文通过最小化误差的平方和来求解系数。通过建立和解决正规方程(Normal Equations)来实现。对于二次多项式，本文构建方程组如下：

$$\begin{cases} \sum x_i^4 a + \sum x_i^3 b + \sum x_i^2 c = \sum x_i^2 y_i \\ \sum x_i^3 a + \sum x_i^2 b + \sum x_i c = \sum x_i y_i \\ \sum x_i^2 a + \sum x_i b + nc = \sum y_i \end{cases} \tag{7}$$

将其转化为矩阵表示并且求解，转化为：

$$\boldsymbol{Ax} = \boldsymbol{B} \tag{8}$$

其中：

$$\boldsymbol{A} = \begin{bmatrix} \sum x_i^4 & \sum x_i^3 & \sum x_i^2 \\ \sum x_i^3 & \sum x_i^2 & \sum x_i \\ \sum x_i^2 & \sum x_i & n \end{bmatrix} \tag{9}$$

$$\boldsymbol{x} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} \sum x_i^2 y_i \\ \sum x_i y_i \\ \sum y_i \end{bmatrix} \tag{10}$$

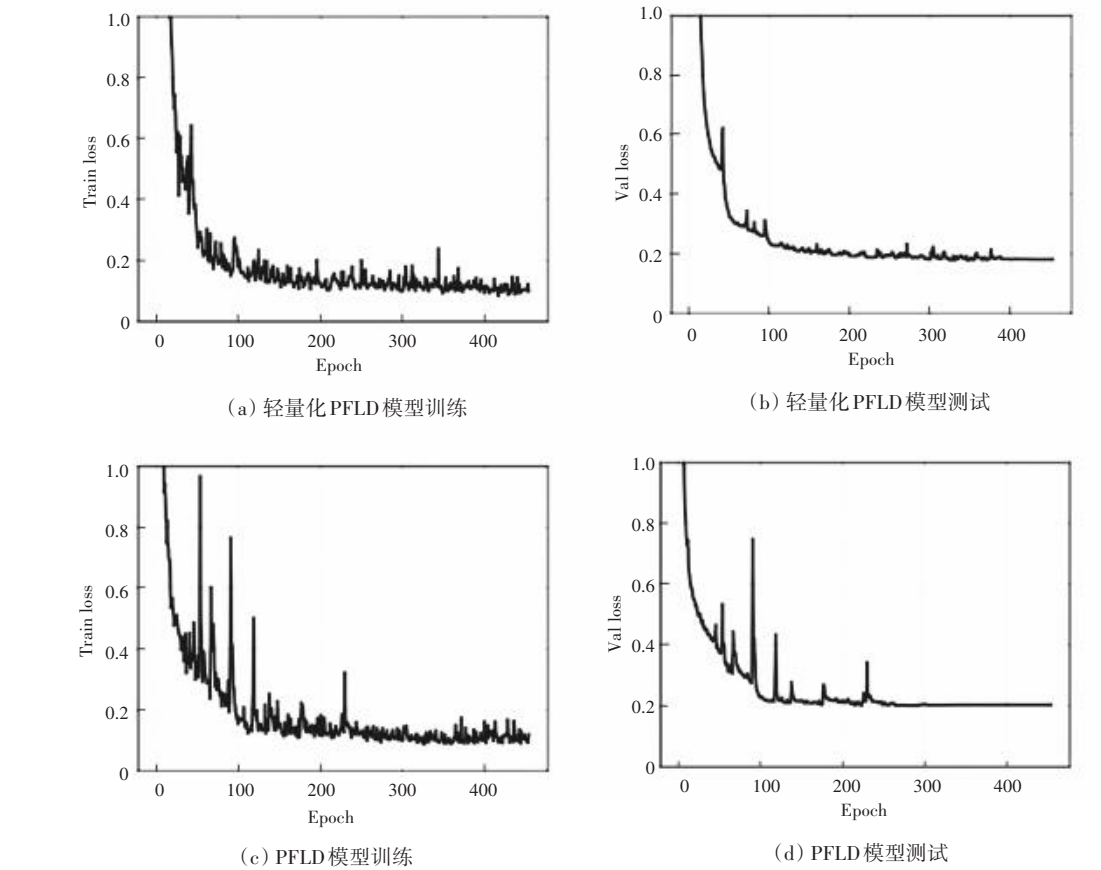
使用逆矩阵将 $\boldsymbol{x}$ 计算出来，从而求出这个二次多项式，取这些点的中间点计算曲率，从而判断嘴巴张开的程度。同理，眼睛、眉毛、嘴角的运动都可以用此来实现。

3　实验结果

为了衡量仿人机器人在模仿人类行为时的相似度和响应速度，本文设计了 3 个阶段的实验。首先，将本文的方法与原始的 PFLD 模型在 WFLW 数据集[21]上进行了对比，WFLW 是由商汤科技提供的人脸关键点检测数据集，包含了 10 000 张人脸图像(7 500 训练集/2 500 测试集)，标注了 98 个关键点，已成为评估人脸关键点检测模型鲁棒性的基准数据集之一。通过这些对比实验，验证了本文优化后的轻量级深度学习网络在准确性和实时处理方面的表现。其次，本文采用了改进后的轻量级模型来驱动仿人机器人的镜像行为，并对其效果进行了评估。本文利用表情识别技术来确定情感状态，并比较人类表演者与机器人模拟出的表情的概率分布，以此作为衡量表情模仿行为相似性的一个指标。最后，本文利用陀螺仪来记录机器人和人类参与者的头部姿态角度，这些数据将用于分析两者在头部运动模仿行为上的相似性。

3.1 轻量级深度学习的评价

为了评估改进的 PFLD 轻量级面部特征点识别算法的低延迟和高精度,本文在 WFLW 数据集上进行训练和测试。此外,还将未修改的 PFLD 模型与本文提出的方法进行了比较。在实验中,WFLW 数据集被输入到 2 个网络中,并在相同的条件下进行训练。在训练过程中记录和可视化了 2 个网络的训练精度和训练损失。对比实验表明,深度学习网络经过轻量化处理后仍能收敛,其轻量化的 PFLD 模型在测试集的 loss 低于原模型在测试集上的 loss,其中修改过后的模型 loss 为 0.186,原模型的 loss 为 0.22,如图 6 所示。使用累积误差分布(Cumulative Errors Distribution curve, CED)曲线进行进一步验证,其中 CED 曲线的横坐标表示归一化误差的阈值,纵坐标表示在这个误差阈值以下,预测关键点与真实关键点匹配成功的样本比例,结果如图 7 所示。这一改进不仅提高了模型的实用性,还为在资源受限的设备上部署面部特征点识别提供了可能。总体而言,本文的轻量级方法在 WFLW 数据集上展现出了优异的准确率和良好的预测性能。



(a) 轻量化 PFLD 模型训练　(b) 轻量化 PFLD 模型测试

(c) PFLD 模型训练　(d) PFLD 模型测试

图 6　在 WFLW 数据集上 2 种模型的训练/测试曲线

Fig. 6　Training/test curves of the two models on the WFLW dataset

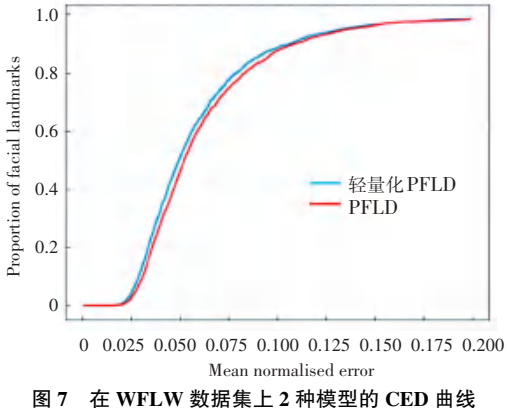


图 7　在 WFLW 数据集上 2 种模型的 CED 曲线

Fig. 7　CED curves of the two models on the WFLW dataset

为了进一步评估本文的方法在轻量级处理效果和实时性方面的优势,记录了在 WFLW 数据集上训练 2 个网络时,每个历元的平均训练时间、总训练时间、模型大小。WFLW 数据集经过数据处理,其中训练集有 75 000 张图片,测试集有 2 500 张图片。轻量化后的 PFLD 模型在每个训练周期(epoch)中耗时 1.154 min,模型大小为 6.5 MB。而原始 PFLD 模型则在每个训练周期中耗时 1.56 min,且其大小为 7.7 MB。

时间提升了 26%,参数量缩减了 15.5%。总地来说,本文提出的方法在人脸特征点在推理相同大

小的数据时具有较低的延迟。

3.2　面部表情模仿的评价

在仿人机器人上对 98 个人脸特征点进行舵机伺服位移映射，实现仿人机器人面部对人类表演者面部的镜像行为。为了对人类表演者和人形机器人之间的面部表情的镜像行为进行定性评估，建立了一个实用的验证系统来评估本文的方法。在实验中，在人脸前方约 0.2 m 处安装一个摄像头采集人脸图像，采集到的图像输入修改过的 PFLD 模型中，驱动仿真机器人头再现人脸的面部表情。

在实验中，招募 2 名人类受试者参加一项初步研究。要求 2 名被试做出不同的面部表情，并记录下这些面部表情和仿真机器人对应的镜像行为，评价其相似度，可以用来直接反映真实的人到仿人机器人的面部表情映射。此外，为了分析具体的相似度，将人类表演者和人形机器人的面部表情输入相同的面部表情识别算法，并计算相应的情绪类别和预测概率，以评估面部表情反映行为的相似度。识别结果如图 8 所示。图 8 中，纵坐标表示情绪识别的预测概率，不同颜色表示不同的情绪类别。

结果表明，机器人可以产生与人类受试者相同的表情类型，而模仿的相似性（即预测概率的差异）可以保持在 0.2 以内。进一步验证了该方法的有效性。

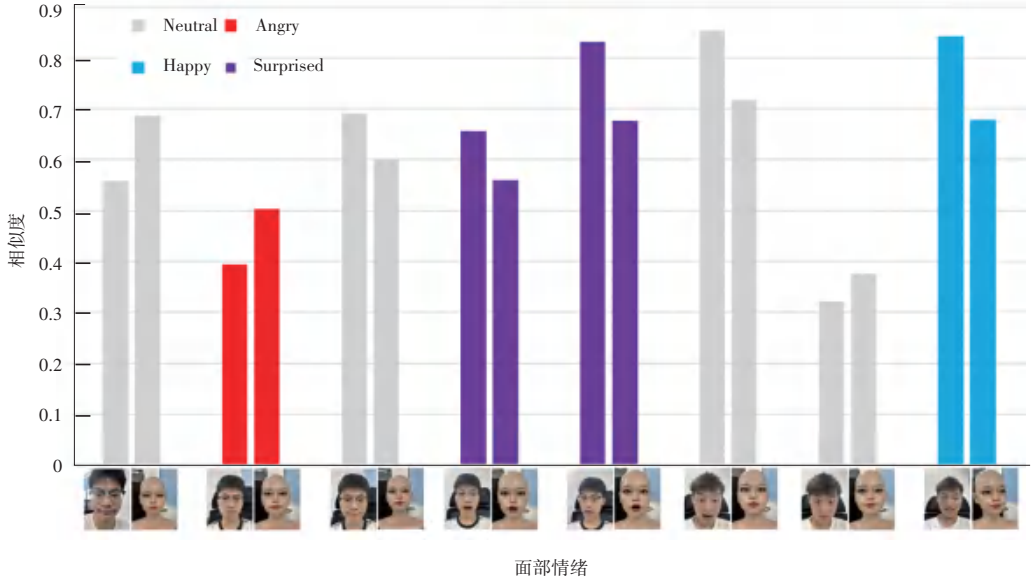


图 8　面部情绪相似度

Fig. 8　Facial emotional similarity

3.3　头部镜像运动行为评价

为了评估机器人头部运动镜像行为的相似性，使用九轴数字倾角仪分别获取实验者人头和仿真人头的 RPY 角数据，并将此数据进行比较，即可显示出头部镜像运动行为的相似性。

实验中，在机器人头部平台的顶部安装上九轴数字倾角仪，用于捕捉机器人头部运动。图 9 展示了实验者头部旋转和本文的方法对头部旋转的 3 个欧拉角的比较。结果表明，通过九轴数字倾角仪采集实验者的头部运动数据的曲线和机器人头部数据的曲线具有相同的趋势，两者曲线有一定的拟合性。

为了更清楚地了解误差的大小，本文使用了均方根误差（Root Mean Squared Error，RMSE）分别对俯仰角、横滚角和航向角进行计算。求出俯仰角的均方根误差为 3.192 7，横滚角的均方根误差为 3.447 3，航向角的均方根误差为 3.085 1。

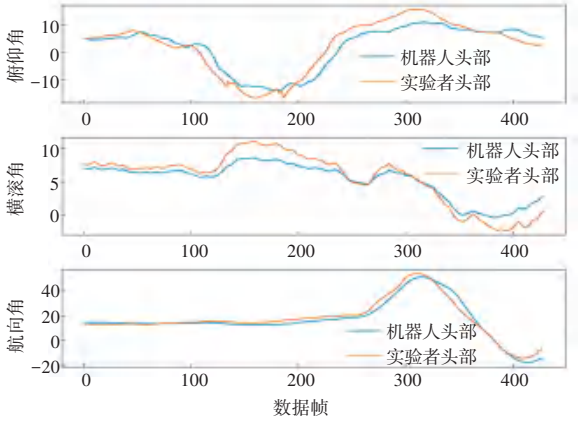


图 9　头部旋转的 3 个欧拉角对比曲线

Fig. 9　Comparison curves of three Euler angles of head rotation

4 结束语

在自然人机交互中,实现对人类面部表情和头部动作的实时且高度仿真的镜像反应是至关重要的。但是,关于仿人机器人复制人类面部表情和头部动作的研究目前还相对匮乏。现有的一些研究主要集中在将人类的这些行为转移到虚拟形象上,而非机器人。此外,大多数现有研究忽略了在嵌入式设备上运行的深度学习算法的实时性和准确性。

为此,本研究提出了一种轻量级的深度学习网络仿人机器人面部表情模仿方法。首先,开发了一个 14 自由度的人形机器人头部平台,实现了面部表情和头部运动的镜像行为。其次,对 PFLD 模型进行轻量化,用于识别 98 个人脸关键特征点。然后将面部特征点的运动转换为表情变化,并建立了映射关系,以优化机器人头部动作的伺服控制。实验结果表明,所提的轻量级网络模型能够在保证高精度的前提下实现人脸特征点的实时识别。此外,为了进一步评估镜像行为的模仿精度,使用面部表情识别算法验证面部表情镜像行为的准确性,使用九轴数字倾角仪验证头部运动镜像行为的准确性。

本文的研究也存在一些局限性。例如,本文的机器人头部平台还比较不发达,有很多改进的可能性,特别是在其机械结构、数据传输和电力系统消耗方面。未来,希望将脑电图、语音等能够反映情绪变化的信息结合起来,加入仿人机器人中,实现情绪的多模态表达。

参考文献

[1] Toichoa E A, Mohammed W M, Martinez L J L. Emotion-driven analysis and control of human-robot interactions in collaborative applications [J]. Sensors, 2021, 21(14): 4626.

[2] Rawal N, Stock-Homburg R M. Facial emotion expressions in human-robot interaction: A survey [J]. International Journal of Social Robotics, 2022, 14(7): 1583-1604.

[3] Yan H, Ang M H, Poo A N. A survey on perception methods for human-robot interaction in social robots [J]. International Journal of Social Robotics, 2014, 6: 85-119.

[4] 李新潮, 袁野, 刘娜. SignBot:一种基于轻量深度神经网络的人形手语机器人[J]. 小型微型计算机系统, 2023, 44(12): 2193-2200.

[5] Osman A. Realistic digital human characters: challenges, models and training algorithms [D]. Tübingen: Universität Tübingen, 2024.

[6] Yoon L, Yang D, Kim J, et al. Placement retargeting of virtual avatars to dissimilar indoor environments [J]. IEEE Transactions on Visualization and Computer Graphics, 2020, 28(3): 1619-1633.

[7] Berns K, Hirth J. Control of facial expressions of the humanoid robot head ROMAN [C]// Proceedings of 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, NJ: IEEE, 2006: 123-130.

[8] Jaeckel P, Campbell N, Melhuish C. Towards realistic facial behaviour in humanoids: mapping from video footage to a robot head [C]// Proceedings of 2007 IEEE 10th International Conference on Rehabilitation Robotics. Piscataway, NJ: IEEE, 2007: 278-283.

[9] Zhang F C, Liang X M, Sun Y Q, et al. POFMakeup: A style transfer method for Peking Opera makeup [J]. Computers and Electrical Engineering, 2022, 104: 108425.

[10] Milborrow S, Nicolls F. Locating facial features with an extended active shape model [C]//Proceedings of the 10th European Conference on Computer Vision. Cham: Springer, 2008: 504-513.

[11] Guo X, Li S, Yu J, et al. PFLD: A practical facial landmark detector[J]. Neurocomputing, 2019, 367: 11-20.

[12] Yang T Y, Chen Y T, Lin Y Y, et al. FSA-Net: Learning fine-grained structure aggregation for head pose estimation from a single image [C]// Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2019: 8142-8151.

[13] 于长东, 毕晓君, 韩阳, 等. 基于轻量化深度学习模型的粒子图像测速研究[J]. 光学学报, 2020, 40(7): 0720001.

[14] Hassibi B, Stork D G. Second order derivatives for network pruning: Optimal brain surgeon[C]//Proceedings of Advances in Neural Information Processing Systems 5 (NIPS 1992). Denver, CO, USA: Morgan Kaufmann, 1992: 164-171.

[15] Liu Z, Wang Y, Vaidya S, et al. KAN: Kolmogorov-Arnold Networks[J]. Neural Networks, 2024, 172: 106123.

[16] Tao Y, Zhang M, Su Y, et al. Exploring college English language learners' social knowledge construction and socio-emotional interactions during computer-supported collaborative writing activities [J]. Journal of Computer Assisted Learning, 2022, 31(5): 613-622.

[17] Xia J, Qu W, Huang W, et al. Sparse local patch transformer for robust face alignment and landmarks inherent relation learning [C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2022: 12345-12355.

[18] Zhang S, Pan X, Cui Y, et al. Learning affective video features for facial expression recognition via hybrid deep learning [J]. IEEE Access, 2019, 7: 32297-32304.

[19] 李艳丽. 卷积神经网络模型压缩方法在图像分类中的应用[D]. 长春: 吉林大学, 2021.

[20] Wang L, Zhang Y, Feng J, et al. On the Euclidean distance of images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(8): 1334-1339.

[21] Wu W, Qian C, Yang S, et al. Look at boundary: A boundary-aware face alignment algorithm [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 2129-2138.