多尺度卷积神经网络的头部姿态估计

梁令羽[1,2,3**]，张天天[1,3]，何为[1*]

[1]中国科学院上海微系统与信息技术研究所无线传感网与通信重点实验室，上海 201800；
[2]上海科技大学信息科学与技术学院，上海 200120；
[3]中国科学院大学，北京 100049

摘要 针对多尺度卷积神经网络的头部姿态估计准确率在实际应用中易受到光照、遮挡等干扰因素的影响，以及大量运算导致算法运行速度较低的问题，提出了头部姿态估计算法。使用不同尺度的卷积核对输入的头部姿态图片进行特征提取，丰富了图像特征，同时保留了图像信息，增强了算法对干扰因素的稳健性。引入 1×1 卷积对网络结构参数进行降维，降低了系统的运算量，提高了算法的时效性。实验结果表明，所提算法在 Pointing′04 和 CAS-PEAL-R1 数据库上的识别率分别为 96.5%和 98.9%，对于光照、表情、遮挡等干扰表现出较好的稳健性，具有较快的运行速度。

关键词 图像处理；头部姿态估计；卷积神经网络；多尺度卷积；1×1 卷积

中图分类号 TP391.4 **文献标识码** A **doi**：10.3788/LOP56.131003

Head Pose Estimation Based on Multi-Scale Convolutional Neural Network

Liang Lingyu[1,2,3**], Zhang Tiantian[1,3], He Wei[1*]

[1] *Key Laboratory of Wireless Sensor Network and Communication, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai* 201800, *China*;
[2] *School of Information Science and Technology, ShanghaiTech University, Shanghai* 200120, *China*;
[3] *University of Chinese Academy of Sciences, Beijing* 100049, *China*

Abstract The accuracy of head pose estimation is easy to be affected by illumination, occlusion and other disturbances in practical applications and a large number of calculations are difficult to meet timeliness of practical applications. In order to solve these problems, a method based on multi-scale convolutional neural network is proposed. The feature extraction of the input head pose image is performed by using different scale convolution kernels, which enriches the image features while preserving the image information, and enhances the robustness of the algorithm to the interference factors. At the same time, the 1×1 convolution is introduced to reduce the network structure parameters, reduce the computational complexity of the system, and improve the timeliness of the algorithm. The result of experiment shows that the recognition rates of the proposed algorithm on Pointing′04 and CAS-PEAL-R1 databases are 96.5% and 98.9%, respectively. The method shows good robustness to illumination, expression, occlusion and other disturbances, and has better operation and speed.

Key words imaging processing; head pose estimation; convolutional neural network; multi-scale convolution; 1×1 convolution

OCIS codes 100.2000; 100.5010; 110.2970

1 引 言

头部姿态估计是指计算机通过对输入图像或者视频序列进行分析、预测，确定人的头部在三维空间中的位置以及姿态参量[1]。这里的姿态参量是指人脸图像在三维空间中的旋转角度，分别为水平旋转

收稿日期：2018-12-06；**修回日期**：2019-01-09；**录用日期**：2019-01-24

基金项目：国家重点研发计划(2018YFC1505204)

* **E-mail**：wei.he@mail.sim.ac.cn；** **E-mail**：liangly@shanghaitech.edu.cn

(yaw)、垂直旋转(pitch)、平面内旋转(roll)[2]。头部姿态估计在人机交互、生物识别、虚拟现实以及疲劳检测等领域有着广泛的应用,因此对于头部姿态估计的研究具有现实意义。常用的头部姿态估计方法分为两种,基于模型的方法和基于外观的方法[3]。基于模型的方法主要是通过检测面部关键特征点来构建头部形态的几何模型,从而估计头部姿态[4-6]。基于外观的方法通常是假定获取的人脸图像和真实图像存在某种特定的关系,运用统计或者概率的方法训练关系模型来推断头部姿态[7-9]。基于模型的方法虽然能够获取到连续的头部姿态估计值,但姿态估计准确度严重依赖人脸特征点的检测精度。然而实际应用中,人脸特征点检测精度会受到光照变化、复杂背景、头部偏转以及遮挡等干扰因素的影响[10],导致检测精度明显下降,甚至无法检测到人脸特征点。此外,简单的模型对于头部姿态的描述并不准确;复杂的模型,计算困难,实际应用价值不高。基于外观的方法将对头部姿态的识别问题转化为分类问题,通过训练-学习的方法获得人脸姿态和头部姿态的对应关系,此类方法的性能依赖于人脸图像样本好坏和学习模型的设计[11]。基于外观的方法不依赖于人脸特征点的准确定位,可以预测较大偏转范围的头部姿态,因此基于外观的方法进行头部姿态估计是一个很好的选择。

头部姿态估计在实际应用中需要解决两个问题。第1个问题是解决光照、遮挡、背景等因素对识别准确率的影响。第2个问题是降低计算开销,提高系统的运行速度。由于卷积神经网络(CNN)在图像处理领域取得了丰富的成果,在字符识别、人脸识别等领域有着成熟的应用,因此采用卷积神经网络方法对于解决头部姿态在实际应用中存在的问题具有重要的研究价值。

目前,将卷积神经网络应用到头部姿态估计领域的方法不多[12]。为了解决头部姿态估计在实际应用中存在的问题,本文提出了多尺度卷积神经网络的头部姿态估计方法。该方法具有以下两方面的优点:1)所设计的卷积神经网络结构使用不同尺度的卷积核对原始人脸图像进行特征提取,在丰富图像特征表达的同时较好地保留了图像原始信息,更好地构建人脸姿态和头部姿态的对应关系模型,对于光照、遮挡等干扰因素具有较好的稳健性;2)所设计的卷积神经网络结构能够有效降低系统计算的开销,使得系统具有较高的时效性。

2 卷积神经网络

卷积神经网络是LeCun等[13]在1990年提出用来专门处理具有类似网格结构的数据神经网络,一个卷积神经网络通常由卷积层、池化层和全连接层组成。卷积层通过卷积运算来提取图像特征,卷积运算可以表示为

$$x_j^{(l)}=f\left(\sum_{i=M_j}x_i^{(l-1)}\times k_{ij}^{(l)}+b_j^{(l)}\right),\tag{1}$$

式中:l 为卷积层的层数;k 为卷积核的大小;b 为偏置项;M_j 为上一层输出的特征谱;f 为激活函数;i 为输入特征相应图的通道数;j 为输出特征相应图的通道数;$x_j^{(l)}$ 为卷积层的输出。卷积运算主要通过稀疏连接(spares interactions)、参数共享(parameter sharing)等手段来改进卷积神经网络[14]。稀疏连接是指每一层的输出单元只与上一层输入单元的一个小邻域范围内存在连接。卷积神经网络的稀疏连接性极大地降低了模型的运算量,限制了模型的拟合能力。参数共享是指在一个模型的多个函数中使用相同的参数。参数共享可以有效地减少模型需要学习的权值参数,降低了模型的存储需求。池化层常用于下采样特征响应图的局部区域,通常设置在卷积层后。常用的池化方法有均值池化和最大池化,池化操作保证了图像平移、旋转和尺度的不变性,保留主要特征的同时减少了参数数量。全连接层位于特征提取之后,有针对性地将提取出的分布式特征表示映射到样本标记空间,在整个卷积神经网络中起到“分类器”的作用[15]。

设计一个合理的卷积神经网络结构是卷积神经网络研究中的主要问题。网络结构越复杂,需要学习的参数量越多,因此需要更多的训练样本,系统计算开销较大,运行速度较慢。网络结构越简单,需要学习的参数量越少,系统计算开销较小,需要的样本也相对较少,但对特征的抽象表达能力不足,易受光照、遮挡等外在因素的干扰。

3 多尺度卷积神经网络的头部姿态估计

在实际应用过程中,相比估计连续的头部转动角度,判断头部姿态处于低头或者抬头等简单动作更有意义且更可靠。因此将头部姿态估计的识别问题转换成为对头部姿态的分类问题,通过卷积神经网络对人脸图像进行特征提取,构建多个离散角度和头部姿态的对应关系模型,利用该

模型对人脸图片进行分类，从而估计头部姿态。该方法可以有效提取对姿态敏感，对光照、遮挡、背景等影响因素稳健的特征，并且提高了系统的运算速度。头部姿态估计方法总体描述如图 1 所示。图 1 中，将输入的彩色图像进行图像预处理后，检测并裁剪出人脸区域图像，将裁剪后的人脸区域图像输入到本文设计的卷积神经网络中，得到输出的头部姿态估计结果。

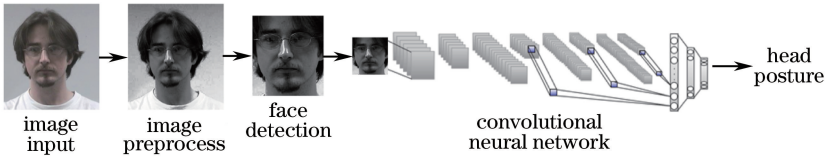

图 1 头部姿态估计流程图

Fig. 1 Head pose estimation flow chart

3.1 图片预处理及人脸检测

本文重点是设计有效的卷积神经网络模型来解决系统在实际应用中的稳健性和时效性问题，同时人脸区域的检测方法有着成熟的应用。因此只简单介绍一下图片预处理和人脸检测部分。

采用的预处理方法主要包括图像灰度化、直方图均衡化以及图像归一化 3 个步骤[16-17]。经过预处理的图像可以凸显出重要特征，减小干扰因素对图像的影响以及提高系统的运算速度。由于背景带来的噪声会严重影响头部姿态特征的选择，因此需要在提取头部姿态特征前对人脸区域进行检测。采用 Adaboost 人脸检测算法来选取人脸区域，该算法[18]基于积分图像来选取面部特征值，利用面部特征值特征构建多个弱分类器级联的强分类器，然后利用强分类器完成人脸区域检测。该算法不容易出现过拟合现象，具有较好的检测率和运行速度。对于检测到的人脸区域，将其归一化到 48×32 大小的灰度图，方便接下来训练卷积神经网络模型。

3.2 多尺度卷积神经网络结构

卷积神经网络在图像处理中的丰富成果对于解决光照、遮挡等干扰因素影响头部姿态识别准确率的问题具有参考意义。通过研究前人应用在头部姿态识别上的深度学习方法后发现，它们基本都是通过使用一种尺度卷积核提取图片底层特征信息，如文献[12]中提出基于卷积神经网络的稳健性头部姿态估计方法。然而单一卷积核在表达人脸图片信息时会遗漏较多的细节，受 InceptionNet 思想[19]的启发，本文提出的网络通过采用多个尺度卷积核对人脸图片进行特征提取，可以从原始人脸图像中得到更多的信息，丰富了图像特征。网络结构如图 2 所示，图中 conv 表示卷积层，stride 表示步长，BN 表示批归一化层，max-pool 表示最大池化，avg-pool 表示均值池化。

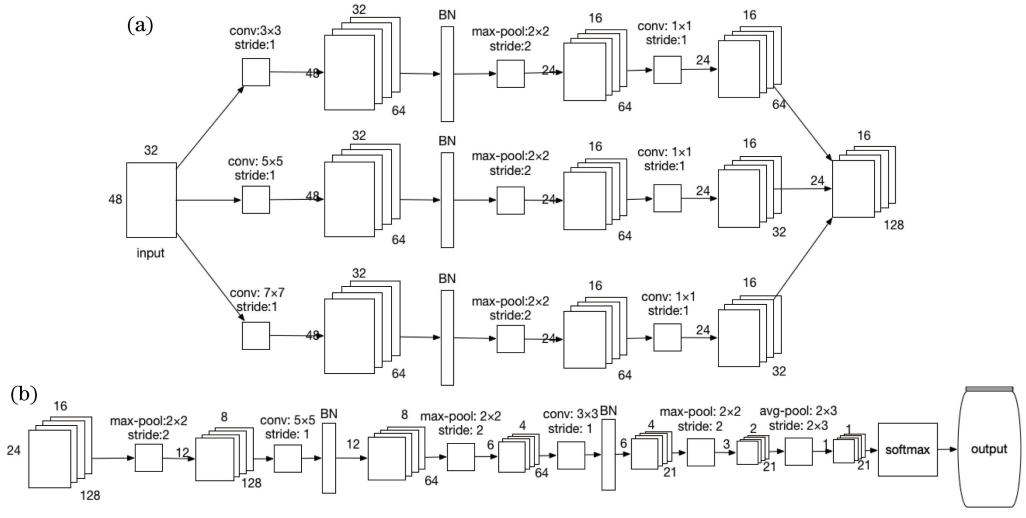

图 2 用于头部姿态估计的深度卷积神经网络结构。(a)多尺度卷积结构；(b)特征组合后处理结构

Fig. 2 Deep neural network structure for head pose estimation. (a) Multi-scale convolution structure; (b) process structure after feature combination

图 2 中采用经过提取和归一化后大小为 32×48 的单通道人脸图像作为输入，网络结构中包含 4 个卷积层，卷积层后采用 relu 激活函数增强神经网络的表达能力。相比于 sigmoid 函数、tanh 函数等激活函数，relu 函数具有克服梯度消失、加快模型的训练速度，以及更好地防止模型过拟合的特点，适用于追求系统时效性的深度卷积神经网络模型。池化层函数采用了步长为 2，大小为 2×2 的最大池化(max-pooling)层，池化后的图像尺度变为原图像大小的 1/4。

图 2(a)中采用了尺度为 3 pixel×3 pixel，5 pixel×5 pixel 以及 7 pixel×7 pixel 的卷积核对输入图片进行卷积处理。为了提高神经网络的收敛速度和拟合能力，将处理后的卷积层通过批归一化(BN)算法[20]进行处理，BN 算法步骤如下：

1) 输入：对于训练集中输入层或某个中间层输出的 batch 数据$\{x_1, x_2, \cdots, x_m\}$作为该批次的输入样本集 B，其中 x 表示输入样本，m 表示输入样本数量；

2) 输出学习参数 γ 和 β：求取该批次样本的平均值 $\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i$，该批次样本的方差 $\sigma_B^2 = \frac{1}{m}\cdot\sum_{i=1}^{m}(x_i-\mu_B)^2$，则调整后的样本 $\hat{x} = \frac{x_i-\mu_B}{\sqrt{\sigma_B^2+\varepsilon}}$，其中 ε 表示一个较小的常数值。经过 BN 算法后的输出样本集表示为$\{y_i = B_{\gamma,\beta}(x_i)\}$，其中 y 表示经过 BN 算法后的输出，$B_{\gamma,\beta}$表示 BN 层的处理，下角标 γ 和 β 表示为了避免归一化操作破坏原有样本的分布，引入可学习参数调整样本分布，由神经网络自主学习。通过 $B_{\gamma,\beta}(x_i)=\gamma\hat{x}+\beta$ 输出 γ 和 β。

BN 算法本质上是通过改变参数 γ 和 β 来优化变换数据分布的大小和位置的，从而将每层数据分布在随输入变化敏感的区域，对于激活函数而言，该区域的梯度变化通常较大。所以 BN 算法可以有效地对抗梯度消失的现象，提高卷积神经网络的收敛速度和拟合能力。

在人脸图像特征提取后，引入 1×1 卷积[21]来降低系统参数量，提高算法的运行速度。当输入通道数大于输出通道时，1×1 卷积相当于对网络结构进行降维，这种降维操作不会改变图片的大小，只会改变图像通道数，这种性质可以确保图像的完整信息得以保留。如图 2(a)所示，5×5 卷积核和 7×7 卷积核提取后特征响应图通道数均减少了一半。由于实验中发现如果仅采用单尺度卷积核，使用 3×3 大小卷积核的算法效果要好于大小为 5×5 卷积核以及 7×7 卷积核，因此使 3×3 卷积生成后特征响应图的通道数不变，即该分支的 1×1 卷积操作前输入通道数等于操作后的输出通道数。这样做有两个好处：1) 在神经网络结构中增加该 3×3 卷积核的权重，提高头部姿态的识别率；2) 配合激活函数使得图片在不损失分辨率的情况下增加非线性特性，增强了神经网络表达图像特征的能力。

在网络结构的最后，采用 2×3 的平均全局池化层来代替全连接层，能有效解决全连接层参数量大、训练速度慢且容易过拟合的问题，输出采用 Softmax 函数进行分类判别。神经网络的训练目标是最小化损失函数。采用交叉熵损失函数，损失函数的表达式为

$$l_{oss} = \sum_i y_i \ln(y_{predicted_i}) + (1-y_i)\ln(1-y_{predicted_i}), \tag{2}$$

式中：y_i 表示真实值分类；$y_{predicted_i}$表示预测值。

使用自适应性矩估计(AME)算法[22]代替传统的随机梯度下降算法迭代更新神经网络权重时，计算效率较高，内存需求较低。

4 实验结果与分析

4.1 实验准备

实验采用的数据集为 CAS-PEAL-R1 数据集和 Pointing′04 数据集，如图 3 所示。CAS-PEAL-R1 数据集是由中国科学院计算技术研究所在 2003 年完成的头部姿态数据集，其中包含 1040 人，每人有 21 幅头部姿态变化图像，共 21840 幅图像。头部姿态变化包括抬头、平视、低头 3 种垂直旋转姿态以及每种垂直旋转姿态下的 7 种水平旋转姿态。Pointing′04 数据集由 15 组图像组成，每组图像包含 2 个系列各 93 张不同姿态的同一人图像。每个系列的数据集由水平和垂直 2 个旋转角度组成，范围均为−90°到 90°，其中水平角度上每 15°保存一次头部姿态图像，垂直角度上每 30°保存一次头部姿态图像。

由于人脸区域检测相关研究和应用非常成熟，且不是本文重点。因此采用文献[18]提供的方法完成对人脸区域的检测，并归一化为 32×48 的灰度图片。部分归一化后的图像如图 4 所示。

卷积神经网络中预训练网络的参数设置如下：训练迭代次数 $N_{epochs}=1000$，每次处理样本数量 $B_{size}=128$，初始学习率为0.04，每200步缩小为原来

(a) (b)

图 3 实验用部分头部姿态库图片。(a) CAS-PEAL-R1;(b) Pointing′04

Fig. 3 Part of the experimental head posture library picture. (a) CAS-PEALR1; (b) Pointing′04

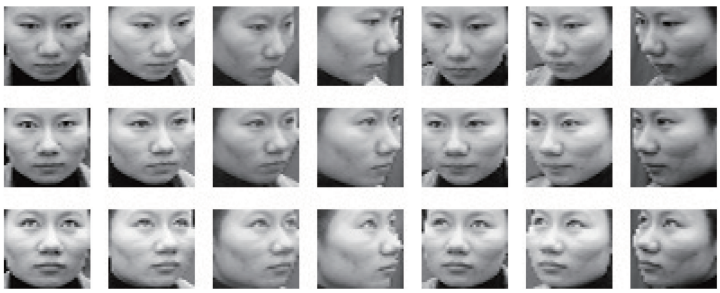

图 4 剪裁后的人脸图像

Fig. 4 Face images after cropped

的 1/2。采用的训练样本为 CAS-PEAL-R1 数据库中编号为 102 到 900 的人脸图像,测试样本采用编号为 901 到 1042 的人脸图像。

为了验证所提算法的有效性,用识别图片准确率和平均每张图片处理时间来对算法进行评估。图片识别准确率指所有测试样本的识别正确数与测试样本数之比。平均每张图片处理时间指算法处理 100 张图片后,平均每张图片所用时间。

4.2 实验结果

为了验证多尺度卷积能有效提高头部姿态的识别准确率,对比了多尺度卷积与单尺度卷积在测试集上的识别准确率,单一卷积核只采用 3×3 卷积核、5×5 卷积核以及 7×7 卷积核中的一个,设置单一卷积核卷积输出的特征响应图与多尺度卷积输出特征响应图的数量相同。训练预先设置的参数不变,实验结果如表 1 所示。多尺度卷积相比于单尺度卷积,在测试集上的头部姿态识别准确率平均提高了 6.6%,结果证明了多尺度卷积能够有效提高头部姿态的识别准确率。同时可以发现,3×3 卷积核在测试集上的识别准确率高于 5×5 卷积核以及 7×7 卷积核,说明 3×3 卷积核相比 5×5 卷积核以及 7×7 卷积核能提取更多的原始图片的细节,提高了对头部姿态识别的准确率。因此在设计多尺度卷积神经网络中,人为增加了 3×3 卷积核的比重。

表 1 多尺度卷积与单尺度卷积对比

Table 1 Multi-scale convolution vs. single-scale convolution

Convolution	Recognition accuracy on test set /%
Multi-scale convolution	98.9
3×3 single-scale convolution	94.3
5×5 single-scale convolution	93.1
7×7 single-scale convolution	93.2

表 2 给出了本文算法与近年文献提出的算法在 Pointing′04 数据集和 CAS-PEAL-R1 数据集上所取得的实验结果对比。从表 2 中可以看出:1)所提算法在 Pointing′04 数据集和 CAS-PEAL-R1 数据集上均取得了最高的识别率,说明所提算法对处理姿态估计问题的效果最优;2)所提算法在姿态识别数上相比其他算法有较大优势;此外,文献[23]中提出的算法依赖于图像特征描述符对头部姿态特征提取性能的优劣,本文提出的算法不依赖于人工选择和设计图像的特征描述符,能够从输入图像中自适应地提取细节信息,避免了人工设计特征对原始图像信息表达的不足,具有更好的识别性能。文献[24]中提出的算法在较大偏转范围的情况下,导致眼睛信息获取失败,使得头部姿态估计发生误判,本文提出的算法不依赖面部特征点定位,较大偏转姿态并不影响算法的识别性能,因此识别姿态数更多。

表 2　不同算法在 Pointing′04 和 CAS-PEAL-R1 数据库上准确率对比

Table 2　Accuracy of different algorithms on Pointing′04 and CAS-PEAL-R1

Algorithm	Pointing′04 accuracy / %	CAS-PEAL-R1 accuracy / %	Number of gestures
Algorithm of this paper	96.5	98.9	21
Cluster-glassification	94.8	96.2	12
Bayesian network[23]	92.7	93.5	6
Based onfacial feature points[24]			

实际应用中头部姿态的转动角度并不严格与数据库中相同，因此本文将算法中的 21 类姿态转换成实际应用中常见的 9 类姿态。转换前的 21 类姿态包括 PU、PM、PD，分别代表抬头、平视、低头 3 种上下旋转的姿态，以及水平方向 −45°，−30°，−15°，0°，15°，30°，45°等 7 种角度的旋转，转换后的 9 类姿态包括平视、平视左偏、平视右偏、低头、低头左偏、低头右偏、抬头、抬头左偏以及抬头右偏，转换前后关系如表 3 所示。

表 3　姿态转换关系表

Table 3　Gestures Conversion relationship table

Aftergestures conversion	Pitch attitude before gestures conversion	Yaw attitude before gestures conversion /(°)
Level	PM	0，−15，+15
Level left	PM	+30，+45
Level right	PM	−30，−45
Pitch down	PD	0，−15，+15
Pitch up	PU	0，−15，+15
Left up	PU	+30，+45
Right up	PU	−30，−45
Left down	PD	+30，+45
Right down	PD	−30，−45

为了验证算法在实际应用中的稳健性，本文测试了标准、戴口罩、戴眼镜、表情、弱光照、强光照以及复杂背景等干扰因素对头部姿态识别准确率的影响。选取了 20 名志愿者，每个志愿者采集了 7 种环境下 9 个姿态图片，共计 1260 张实验图片，部分图片如图 5 所示。

图 5　不同干扰因素下的部分头部姿态图片。(a)标准；(b)戴口罩；(c)戴眼镜；(d)表情；(e)弱光照；(f)强光照；(g)复杂背景

Fig. 5　Partial head posture pictures under different interference factors. (a) Standard; (b) with mask; (c) with glasses; (d) expression; (e) weak illumination; (f) strong illumination; (g) background

表 4 中对比了本文提出的算法和文献[23]中的算法在戴口罩、眼镜、表情、强弱光照、复杂背景和标准实验环境下头部姿态估计准确率。从表 4 的实验结果可知：1)本文提出的算法与文献[23]算法相比，在戴口罩、眼镜、表情、强弱光照、复杂背景以及标准等 7 种实验环境中的准确率分别提高了 7.6%、6.6%、3.2%、3.7%、3.9%、2.6%、2.2%，说明本文算法在各个情况下均优于文献[23]算法；2)相比于标准实验环境，本文算法在戴口罩、眼镜、表情、强弱光照、复杂背景等干扰下，识别准确率分别下降了 5.6%、2.2%、0.4%、2.4%、3.0%、0.8%，而文献[23]中的算法识别准确率下降了 9.9%、6.6%、1.4%、4.5%、4.1%、1.2%。本文算法在各个干扰环境下识别准确率的降低幅度均小于文献[23]，尤其是在口罩遮挡情况下，本文算法相比文献[23]算法的识别率下降幅度减小了 4.4%。实验结果说明本文算法在面对遮挡、表情、背景和光照等干扰因素时具有较好的稳健性。

表 4 不同干扰因素对识别率的影响

Table 4 Different interference factors effect on recognition rate

Interference factor	Accuracy of this paper /%	Accuracy of method in Ref.[23] /%
Standard	98.5	96.3
With mask	92.9	86.4
With glasses	96.3	89.7
Expression	98.1	94.9
Weak illumination	95.5	91.8
Strong illumination	96.1	92.2
Background	97.7	95.1

为了验证本文提出的算法的时效性，选择了1920 pixel×1080 pixel、1360 pixel×760 pixel 和800 pixel×600 pixel 3 种常用分辨率图片各 100 张，对比标准采用平均每张图片运行时间。采用的计算机型号为 Macbook Pro 13.3，处理器型号为 2.7 GHz Intel Core i5，内存为 8 GB、1867 MHz DDR3，未采用图形处理器(GPU)进行运算。将本文算法与现有实时头部姿态估计算法进行对比，对比结果如表 5 所示。从表 5 中可以看出本文提出的算法在 3 个分辨率上平均每张图片的识别时间最少，这说明 1×1 卷积核的引入在保证卷积神经网络对图像特征充分表达的同时，减少了网络结构中的参数量，提高了系统的运行速度，使得系统具有良好的性能，能够满足实际应用对时效性的要求。

表 5 不同分辨率下的识别时间

Table 5 Recognition time at different resolution

Resolution /(pixel×pixel)	Time of this paper /ms	Time of method in Ref. [23] /ms	Time of method in Ref. [24] /ms
1920×1080	34.3	51.3	562.1
1360× 760	32.7	48.5	451.7
800×600	30.8	45.2	349.8

5 结 论

基于多尺度卷积神经网络提出一种头部姿态估计算法，该算法能够有效估计头部姿态。针对实际应用中头部姿态易受到光照、遮挡等外部因素的影响，以及为了满足系统在实际应用中对于运行速度的要求，利用多尺度卷积核、1×1 卷积核等对传统卷积神经网络进行改进。实验结果证明，相比于现有算法，所提算法具备较好的稳健性和时效性，能够满足实际应用的要求。

参 考 文 献

[1] Alioua N, Amine A, Rogozan A, *et al*. Driver head pose estimation using efficient descriptor fusion[J]. EURASIP Journal on Image and Video Processing, 2016, 2016: 2.

[2] Ahn B, Choi D G, Park J, *et al*. Real-time head pose estimation using multi-task deep neural network[J]. Robotics and Autonomous Systems, 2018, 103: 1-12.

[3] Murphy-Chutorian E, Trivedi M M. Head pose estimation in computer vision: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(4): 607-626.

[4] Zhu X X, Ramanan D. Face detection, pose estimation, and landmark localization in the wild[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE, 2012: 2879-2886.

[5] Derkach D, Ruiz A, Sukno F M. Head pose estimation based on 3-D facial landmarks localization and regression[C]// 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), May 30-June 3, 2017, Washington D. C., USA. New York: IEEE, 2017: 820-827.

[6] Padeleris P, Zabulis X, Argyros A A. Head pose estimation on depth data based on particle swarm optimization[C]// 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, June 16-21, 2012, Providence, RI, USA. New York: IEEE, 2012: 42-49.

[7] Geng X, Xia Y. Head pose estimation based on multivariate label distribution[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 1837-1842.

[8] Huang C, Ding X Q, Fang C. Head pose estimation based on random forests for multiclass classification[C]// 2010 20th International Conference on Pattern Recognition, August 23-26, 2010, Istanbul, Turkey. New York: IEEE, 2010: 934-937.

[9] Drouard V, Horaud R, Deleforge A, *et al*. Robust head-pose estimation based on partially-latent mixture of linear regressions[J]. IEEE Transactions on Image Processing, 2017, 26(3): 1428-1440.

[10] Xia J, Pei D, Wang Q Z, *et al*. Face recognition based on local adaptive ternary derivative pattern coupled with Gabor feature [J]. Laser &

Optoelectronics Progress, 2016, 53(11): 111004.
夏军, 裴东, 王全州, 等. 融合 Gabor 特征的局部自适应三值微分模式的人脸识别[J]. 激光与光电子学进展, 2016, 53(11): 111004.

[11] Tang Y Q, Sun Z N, Tan T N. A survey on head pose estimation [J]. Pattern Recognition and Artificial Intelligence, 2014, 27(3): 213-225.
唐云祁, 孙哲南, 谭铁牛. 头部姿势估计研究综述[J]. 模式识别与人工智能, 2014, 27(3): 213-225.

[12] Sang G L, Chen H, Zhao Q J. Robust head pose estimation based on deep convolution neural networks [J]. Journal of Sichuan University (Engineering Science Edition), 2016, 48(S1): 163-169.
桑高丽, 陈虎, 赵启军. 一种基于深度卷积网络的鲁棒头部姿态估计方法[J]. 四川大学学报(工程科学版), 2016, 48(S1): 163-169.

[13] LeCun Y, Boser B, Denker J S, et al. Handwritten digit recognition with a back-propagation network [M] // Touretzky D S. Advances in neural information processing systems 2. San Francisco: Morgan Kaufmann Publishers Inc., 1990: 396-404.

[14] Ian G, Yoshua B, Aaron C. Deep learning [M]. Cambridge: MIT Press, 2016: 203-204.

[15] Ma Y J, Li X Y, Song X F. Traffic sign recognition based on improved deep convolution neural network [J]. Laser & Optoelectronics Progress, 2018, 55(12): 121009.
马永杰, 李雪燕, 宋晓凤. 基于改进深度卷积神经网络的交通标志识别[J]. 激光与光电子学进展, 2018, 55(12): 121009.

[16] Fei Y J, Shao F. Contrast adjustment based on image retrieval [J]. Laser & Optoelectronics Progress, 2018, 55(5): 051002.
费延佳, 邵枫. 基于图像检索的对比度调整[J]. 激光与光电子学进展, 2018, 55(5): 051002.

[17] Li H X, Lin Z, Shen X H, *et al*. A convolutional neural network cascade for face detection [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 5325-5334.

[18] Viola P, Jones M. Fast and robust classification using asymmetric adaboost and a detector cascade [C] // Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, December 3-8, 2001, Vancouver, British Columbia, Canada. Cambridge: MIT Press, 2002: 1311-1318

[19] Szegedy C, Liu W, Jia Y Q, *et al*. Going deeper with convolutions [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 7298594.

[20] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C] // Proceedings of the 32nd International Conference on International Conference on Machine Learning, July 6-11, 2015, Lille, France. Massachusetts: JMLR. org, 2015, 37: 448-456.

[21] Lin M, Chen Q, Yan S. Network in network [EB/OL]. (2014-03-04) [2018-11-25]. https: // arxiv. org/abs/1312.4400.

[22] Kingma D P, Ba J. Adam: a method for stochastic optimization [EB/OL]. (2017-01-30) [2018-11-25]. https: // arxiv. org/abs/1412.6980.

[23] Huang C, Ding X Q, Fang C. Head pose estimation based on random forests for multiclass classification [C] // 2010 20th International Conference on Pattern Recognition, August 23-26, 2010, Istanbul, Turkey. New York: IEEE, 2010: 934-937.

[24] Min Q S, Liu N, Chen Y T, *et al*. Head pose estimation based on facial feature point localization [J]. Computer Engineering, 2018, 44(6): 263-269.
闵秋莎, 刘能, 陈雅婷, 等. 基于面部特征点定位的头部姿态估计[J]. 计算机工程, 2018, 44(6): 263-269.