

## 一种用于细粒度人脸识别的眼镜去除方法

毛亮<sup>①</sup> 薛月菊<sup>\*②</sup> 魏颖慧<sup>③</sup> 朱婷婷<sup>①</sup>

<sup>①</sup>(深圳职业技术学院粤港澳大湾区人工智能应用技术研究院 深圳 518055)

<sup>②</sup>(华南农业大学电子工程学院 广州 510642)

<sup>③</sup>(西安电子科技大学数学与统计学院 西安 710071)

**摘要:**为解决眼镜遮挡会降低人脸识别性能的难点,借鉴深度卷积神经网络在超分辨率方面的成功应用,该文提出一种用于细粒度人脸识别的眼镜自动去除方法ERCNN。用卷积层、池化层、MFM特征选取模块和反卷积层设计ERCNN网络模型,自动学习戴眼镜和未戴眼镜人脸图像对之间的映射关系,实现端到端的眼镜去除。然后,收集大量监控场景下的人脸图像,以及互联网上公开的人脸图像作为训练集;同时构建SLLFW数据集,作为眼镜去除和人脸识别的测试集。最后,通过与传统的眼镜去除方法进行对比试验,该文算法的各项评价指标优于传统方法,能有效的去除真实人脸图像中眼镜;同时在SLLFW人脸数据集上形成的全框眼镜、半框眼镜和无框眼镜人脸数据集上对多种人脸识别算法进行对比试验。试验表明,在FAR为1%的情况下,利用该文方法对F-SLLFW, H-SLLFW和R-SLLFW数据集的人脸图像进行眼镜去除后, SphereFace算法的TAR分别达到90.05%, 91.14%和92.33%,比未去除眼镜的识别率分别提高了3.92%, 3.08%和1.26%;同样,在FAR为0.1%的情况下,比SphereFace算法的TAR分别提高了10.06%, 4.29%和2.13%,说明该文方法有助于提升细粒度人脸识别的识别精度。

**关键词:** 人脸识别; 深度卷积网络; 特征选择; 眼镜去除

中图分类号: TN911.73; TP391.4

文献标识码: A

文章编号: 1009-5896(2021)05-1448-09

DOI: 10.11999/JEIT200176

## An Eyeglasses Removal Method for Fine-grained Face Recognition

MAO Liang<sup>①</sup> XUE Yueju<sup>②</sup> WEI Yinghui<sup>③</sup> ZHU Tingting<sup>①</sup>

<sup>①</sup>(Institute of Applied Artificial Intelligence of the Guangdong-Hong Kong-Macao Greater Bay Area, Shenzhen Polytechnic, Shenzhen 518055, China)

<sup>②</sup>(College of Electronic Engineering, South China Agricultural University, Guangzhou 510642, China)

<sup>③</sup>(School of Mathematics and Statistics, Xidian University, Xi'an 710071, China)

**Abstract:** In order to solve the problem that eyeglasses reduce often the performance of face recognition, based on the successful application of deep convolution neural network in super-resolution, This paper proposes an automatic eyeglasses removal method ERCNN (Eyeglasses Removal CNN) for fine-grained face recognition. Specifically, the ERCNN network which is designed based on the convolution layer, pool layer, MFM (Max Feature Map) feature selection module and deconvolution layer, are automatically learned the mapping relationship between facial images with eyeglasses and their counterparts without eyeglasses to realize end-to-end eyeglasses removal. Then, massive facial images are captured through surveillance equipment and collected from the Internet as the training set. And, SLLFW data set is established, which is used as the test set of eyeglasses removal and face recognition. The experiment show that the proposed method can better effectively remove the eyeglasses from the real facial image than the traditional eyeglasses removal methods, and the evaluation index of the method is better than other methods. In addition, several face recognition methods are tested separately on the facial images formed by SLLFW data set. Experiments show that when the FAR

收稿日期: 2020-03-17; 改回日期: 2020-07-20; 网络出版: 2020-07-27

\*通信作者: 薛月菊 xueyueju@163.com

基金项目: 国家科技支撑计划(2015BAD06B03-3)

Foundation Item: The National Science and Technology Support Program (2015BAD06B03-3)

(False Accept Rate) is 1%, the TAR (True Accept Rate) of the Sphreface method reaches 90.05%, 91.14% and 92.33%, which is 3.92%, 3.08% and 1.26% higher than the Sphreface method is not used to remove the eyeglasses from the F-SLLFW, H-SLLF and R-SLLFW, respectively. Similarly, when the FAR 0.1%, the TAR of Sphreface method is increased by 10.06%, 3.08% and 1.26% respectively. Therefore, the proposed method can better improve the recognition accuracy of fine-grained face recognition.

**Key words:** Face recognition; Deep Convolution Neural Network (DCNN); Feature selection; Eyeglasses removal

## 1 引言

人脸识别一直是模式识别和计算机视觉领域最活跃的研究方向之一。在过去的二十年中,人脸识别技术尽管已经取得了长足的进展,但面向现实的人脸识别系统仍然面临着诸多挑战。这些挑战主要来自于人脸图像中可能存在的多种多样的难以预测的变化,如表情、姿势、光照、分辨率和遮挡等。这些变化可能导致训练图像和待识别图像之间发生严重的数据偏移,从而导致人脸识别率降低<sup>[1]</sup>。其中眼镜是常见的遮挡物体,一般可分为全框、半框和无框3类,如图1所示,从左到右表示戴全框、戴半框和戴无框眼镜的人脸图像,这里指的眼镜不包括太阳镜、运动眼镜等。实际应用中,由于相似人脸这类细粒度特性和眼镜遮挡的影响,人脸识别的准确率会出现显著的下降。因此,本文以如何去除眼镜作为主要研究对象。

近年来,研究人员提出了一些从人脸图像中提取特征和去除眼镜的方法,如2004年,Wu等人<sup>[2]</sup>提出一种自动去除眼镜的方法,利用戴眼镜的人脸图像与不戴眼镜的对应部分之间的统计映射去除眼镜,但对于一些无框眼镜,很难去除;2005年,Du等人<sup>[3]</sup>提出了一个从正面人脸图像中提取并摘除眼镜的方法。首先利用自适应二值化检测正面人脸图像中眼镜遮挡的部分,然后使用基于PCA重构的递归误差补偿方法,虽然对人工合成戴眼镜人脸图像可进行眼镜去除,但对光照变化较敏感;2005年,Park等人<sup>[4]</sup>提出了一种从人脸正面图像中去除眼镜的方法。首先检测被眼镜遮挡的区域,然后通过PCA重构的递归误差补偿方法,生成一幅自然的无眼镜人脸图像,有效地解决了眼镜遮挡问题,提高了人脸识别系统的性能;2007年,Heo等人<sup>[5]</sup>采用多视觉主动外观模型进行拟合的方法,将非正面图像转换为正面人脸,并选择一组正面人脸作为基



图1 戴不同类型眼镜的人脸图像

向量,对被遮挡的人脸图像进行重建,较好地去除了人脸图像中的眼镜遮挡;2010年,Jia等人<sup>[6]</sup>提出了一种基于相位一致性和渐进修补的眼镜检测与去除算法。该方法利用相位一致性和对比度不敏感的特点,较好地提高了人脸识别率;2010年,Wang等人<sup>[7]</sup>提出了一种基于主动外观模型的眼镜去除方法。首先利用主动外观模型粗略地检测出遮挡区域,然后采用椭圆模型来拟合眼睛的位置并提取眼睛区域,最后通过迭代误差补偿生成合成图像,有效地解决了厚边眼镜遮挡人脸后人脸识别率低的问题;2013年,Cheng等人<sup>[8]</sup>提出了一种去除眼镜的疲劳检测方法。首先检测人脸区域,通过形态学操作去除眼镜,然后检测眼睛区域,并采用模板匹配方法进行跟踪,确定眼睛的开闭状态,最后根据眼睛的连续状态确定疲劳状态,准确率达100%;2014年,李根等人<sup>[9]</sup>提出了基于思维进化的机器学习与局部特征结合的方法,首先根据遮挡与无遮挡人脸图像的局部特征进行对比,记录对比相似度作为局部区域的得分,然后对所有局部区域进行趋同过程和异化过程的演化,得到无遮挡区域及遮挡物体区域,实现遮挡人脸条件下的人脸识别,取得较好的识别效果,但验证的数据集较小,且不具备相似人脸这类细粒度特性;2014年,Guo等人<sup>[10]</sup>提出了一种新的眼镜自动去除方法,利用SVM进行眼镜检测,然后采用PCA重构的递归误差补偿方法处理人脸图像,最后通过定量测量,较好地提升了人脸识别的性能;2015年,吴从中等人<sup>[11]</sup>提出基于稀疏表达分类的去除遮挡的方法。首先求出有遮挡人脸图像在无遮挡人脸图像训练集上的稀疏系数,然后根据稀疏系数进行恢复重建,得到去遮挡的人脸图像,但该方法非常依赖眼镜的先验知识;2015年,Fernández等<sup>[12]</sup>提出一种基于局部二进制模式和对准的自动眼镜检测算法,取得了较好的效果,但该方法对于不同类型的眼镜,包括运动眼镜、太阳眼镜、安全眼镜、阅读眼镜等检测的鲁棒性较差;2017年,Sandhan等人<sup>[13]</sup>通过对眼镜镜片反射特性的分析,得到约束条件,并将反射层的梯度调整为高度系数,在几次迭代中会消除反射的影响,最后在正面人脸图像上具有较好的去除反射

效果,提高了虹膜的检测率;2018年,曹志义等人<sup>[14]</sup>提出了一种改进的生成对抗网络模型的人脸还原算法,在已有网络的基础上增加一个还原层,使得测试图像可以通过改进的对抗网络生成对应的高置信度图像,极大地缩短了网络的收敛时间;2019年,毛亮等人<sup>[15]</sup>提出了一种用于细粒度人脸识别的眼镜去除方法的中国发明专利。综上所述,在近几年的发展历程中,传统的眼镜去除方法由简单的PCA不断向更复杂的算法演变,但眼镜去除效果始终无法达到较为理想的状态,而以上方法都是在较少的未戴眼镜和戴眼镜人脸数据上进行试验,没有大规模的戴眼镜和未戴眼镜人脸数据,无法很好地分析眼镜遮挡对人脸识别算法的影响,尤其是细粒度人脸识别。

随着深度学习的快速发展,深度卷积神经网络不论是在解决高层次视觉问题,还是解决低层次视觉问题上都取得重大的突破。2014年,Dong等人<sup>[16]</sup>提出一种基于深度卷积神经网络的单幅图像的超分辨率重构方法,只需做很少的预处理和后处理,可得到低分辨率和高分辨率图像之间映射关系,实现端到端的处理。但是,如果利用该方法直接进行戴眼镜区域的重构,无法得到较理想的效率。

基于上述算法的启发,本文提出一种端到端的用于细粒度人脸识别的眼镜去除方法,简称ERCNN(Eyeglasses Removal CNN)。首先,在文献<sup>[17,18]</sup>方法的基础上,对本文所有数据集进行人脸检测和人脸对齐,形成 $128 \times 128$ 和 $128 \times 96$ 的归一化人脸图像,然后对处理后的人脸,划分3个局部区域,如图2所示,分别标记为Part1, Part2和Part3,其中Part1和Part3为人脸的未处理区域,作为重构人脸图像的局部区域;Part2为包含戴眼镜或未戴眼镜的眼睛区域,作为眼镜去除算法处理输入输出图像。同时利用卷积层、池化层、MFM特征选取模块和反卷积层设计ERCNN网络,将Part2区域输入到网络中进行自动学习戴眼镜和未戴眼镜人脸图像对之间的映射关系,实现端到端的眼镜去除。最后,在自建的SLLFW测试集上进行眼睛去除验证,并利用现有的人脸识别方法在戴有全框、半框和无框3种不同类型眼镜的人脸图像上验证算法对人脸识别率的影响程度。

## 2 试验数据

### 2.1 试验数据采集及预处理

试验数据来源于监控场景、互联网和证件照等人脸图像。其中监控场景的人脸图像采集于2016年11月至2017年12月间,设备的安装高度距离地面 $200 \sim 250$  cm处,图像分辨率为 $1920 \times 1080$ 像素,采集时间分为上午,下午和晚上3个时间段。为保证人脸的多样性,收集到的人脸数据包括未戴眼镜、不同姿态、不同性别和年龄等,总共35975张。对于所收集到的人脸图像,采用人工合成的方法,将全框、半框和无框3种眼镜,合成到原始的人脸图像上,得到61105张戴眼镜的人脸图像。如图2(a)所示,第1列为原始人脸图像,第2列为人工合成戴眼镜人脸图像。

为了得到更好的眼镜去除效果,本文利用文献<sup>[17,18]</sup>中人脸检测和人脸对齐方法,将收集到的数据进行预处理,利用提取到的人脸眼镜关键点对图像进行水平旋转,从而克服视场角的姿态变化影响<sup>[19]</sup>,进而归一化到 $128 \times 128$ 像素和 $112 \times 96$ 像素两种不同尺度的人脸图像,如图2(a)所示,第3列为原始人脸图像对齐后的结果,第4列为戴眼镜人脸图像对齐后的结果。

为了验证眼镜去除算法的有效性,本文在细粒度人脸识别数据FGLFW<sup>[20]</sup>的基础上,通过人工筛选剔除姿态较大的人脸图像,形成SLLFW数据集。SLLFW数据集由8901张人脸图像构成,其中6091张人脸图像来自互联网,余下的2810张人脸图像中,有2232张未戴眼镜的人脸图像,578张真实戴眼镜的人脸图像。为了验证本文方法的性能,采用人工合成的方法,将全框、半框和无框3种眼镜,合成到SLLFW数据集中未戴眼镜的人脸图像上。如图2(b)所示,第1列为原始人脸数据,第2列为戴全框眼镜人脸图像,第3列为戴半框眼镜人脸图像,第4列为戴无框眼镜人脸图像。

### 2.2 训练集和测试集准备

本文收集的原始人脸图像共35975张,经过人工预处理后,形成与之对应的戴不同类型眼镜的人脸图像,总共61105张,作为训练集。在训练前需进行垂直镜像翻转以及顺时针 $30^\circ$ 和 $45^\circ$ 旋转的数据扩增,形成扩增后的训练样本集,共计183315张人



(a) 人脸对齐后的图像

(b) 未戴眼镜和戴不同眼镜的人脸图像

图2 人脸数据集

脸图像。SLLFW数据集中未戴眼镜人脸图像和与之对应的人工合成戴眼镜人脸图像作为眼镜去除方法的验证集，其余578个真实戴眼镜的人脸图像作为眼镜去除的测试集，进一步验证本文方法的性能。同时，SLLFW数据集还可作为人脸识别测试集，验证眼镜去除后对人脸识别性能的影响。

### 3 眼镜去除方法

#### 3.1 网络结构

借鉴SRCNN<sup>[16]</sup>的思路，本文设计了一种新的眼镜去除深度卷积神经网络(Eyeglasses Removal CNN, ERCNN)，通过端对端的学习，自动构建戴眼镜的人脸图像和不戴眼镜的人脸图像之间的映射关系，实现戴眼镜人脸图像的自动去除，网络结构如图3所示。首先，以戴眼镜的人脸图像作为输入，大小为128×128，将输入图像切分为20×128，64×128和64×128 3个图像区域，分别为Part1，Part2和Part3，其中Part2包含完整的戴眼镜的眼睛区域；然后，将Part2区域作为卷积层的输入，用于网络中MFM (Max Feature Map)单元进行特征选择和最大元素操作，再利用反卷积、平均池化和逐个元素加权求和的操作，对Part2区域实现重构，进而得到去除眼镜后的眼睛区域Part2\_new；最后，将输出的Part2\_new区域与原始的Part1区域和Part3区域进行合并，得到完整的去除眼镜的人脸图像。

具体地，ERCNN网络中使用卷积核大小为9×9，padding为4，步长为1的Conv1来提取Part2的高维特征，输出256维通道的特征图，然后通过最大池化层Pool1，MFM1(由切片层Slice1和取最大的元素操作层Eltwise1实现)和激活函数ReLU1处理后，输出大小为32×64，128维通道的特征图。紧接着，继续使用和Conv1相同参数的Conv2堆叠卷积，再经过MFM2(由切片层Slice2和取最大的元素操作层Eltwise2实现)和激活函数ReLU2处理后，将输入的高维特征图非线性映射到另一个高维特征图，输出大小为32×64，128维通道的特征图。然后，利用卷

积核大小为4×4，padding为1，步长为2的反卷积结构DConv3将高维特征图上采样2倍到原图大小的64×128像素，输出128维通道的特征图；通过平均池化层Pool2对该特征图进行处理，并将其特征值乘以权重0.2，与自身特征值进行元素求和操作(Eltwise3)，输出大小为64×128像素，128维通道的特征图；对输出的特征图，使用卷积核大小为1×1，步长为1的Conv4将高维特征图降维重构得到去除眼镜的眼睛区域Part2\_new。最后将去除眼镜后的眼睛区域Part2\_new与Part1和Part3合并，输出去除眼镜人脸图像。

相比更深层的网络VGG<sup>[21]</sup>、GoogLeNet<sup>[22]</sup>和ResNet<sup>[23]</sup>等，本文设计的ERCNN在网络层数和参数方面都较少，模型占用内存较小，运行效率较高，具有很好的实时性。

#### 3.2 MFM单元

ERCNN网络使用的MFM单元结构，采用Wu等人<sup>[18]</sup>提出Maxout Activation作为网络的卷积层激活函数，输出MFM作为特征选择单元，对卷积层神经元添加了竞争关系，抑制低激活的神经元，使网络适用于不同的数据，拥有更好的泛化能力。同样，ERCNN网络在卷积层后使用MFM层，可以起到增强模型性能的作用。

假定输入MFM的卷积层为 $x^n \in R^{H \times W}$ ，其中 $n = \{1, 2, \dots, 2N\}$ 为特征图通道，输入通道数为 $2N$ ， $W$ 和 $H$ 分别为特征图的宽度和高度。MFM操作将输入的两个特征图取对应元素最大值输出一个特征图，计算公式为

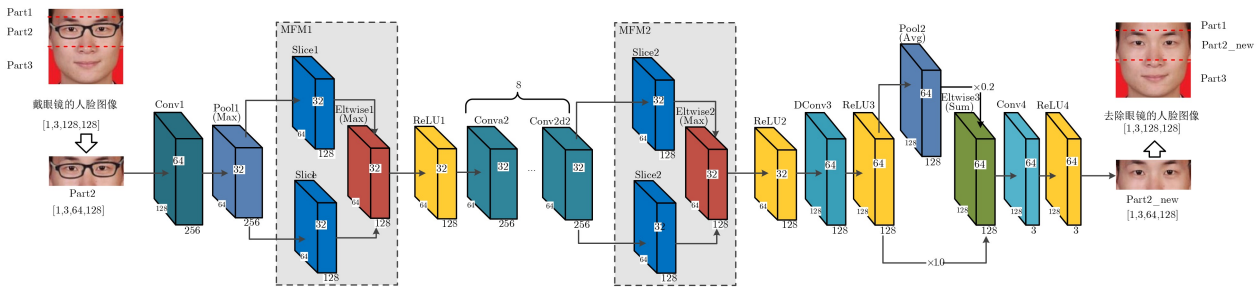
$$\hat{x}_{ij}^k = \max(x_{ij}^k, x_{ij}^{k+N}) \quad (1)$$

式中， $1 \leq k \leq N$ ， $1 \leq i \leq H$ ， $1 \leq j \leq W$ ，通过MFM操作后输出的 $\hat{x}^n \in R^{H \times W \times N}$ 。

#### 3.3 损失函数

ERCNN网络模型采用Euclid Loss<sup>[18]</sup>监督信号来构建损失函数

$$L = \frac{1}{2m} \sum_{i=1}^m \|x_i - y_i\|^2 \quad (2)$$



Conv2a, ..., Conv2d, Conv4为卷积层；ReLU3为激活函数

图3 ERCNN的网络结构

式中,  $L$ 为Euclid损失函数,  $x_i \in \mathbf{R}^d$ 表示第*i*类的第*i*个感兴趣区域的特征,  $m$ 为训练过程中小批量数据的数量,  $y_i \in \mathbf{R}^d$ 表示第*y*类的真实特征。

## 4 试验与分析

### 4.1 试验平台

试验训练及测试采用32 GB内存、2块Nvidia Geforce GTX 1080型号的GPU, Intel I7-7700型号CPU的硬件平台和Ubuntu16.04操作系统。在Caffe深度学习框架<sup>[24]</sup>上, 采用Python作为编程语言实现本文算法。

### 4.2 评价标准

本文对于眼镜去除处理后的人脸图像质量评价, 采用均方误差MSE(Mean Squared Error)、峰值信噪比PSNR(Peak Signal-to-Noise Ratio)、结构相似度SSIM(Structural SIMilarity)、信息保真度准则IFC(Information Fidelity Criterion)4个指标评价。其中, MSE是比较原图与处理后图像的可视误差; PSNR近似于人对重建质量的感知; SSIM用于比较参考信号和失真信号的结构。作为结构相似性理论的实现, SSIM从图像组成的角度将结构信息定义为独立于亮度、对比度的, 反映场景中物体结构的属性, 并将失真建模为亮度、对比度和结构3个不同因素的组合。IFC则是将图像质量评价问题作为信息保真问题来处理。

具体计算公式为:

(1) MSE表示为 $m \times n$ 的无噪声单色图像 $I$ 与它的近似图像 $K$ 的均方差值。

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i, j) - K(i, j)\|^2 \quad (3)$$

其中,  $I(i, j)$ 表示图像 $I$ 中第*i*行、第*j*列的像素值;  $K(i, j)$ 表示图像 $K$ 中第*i*行、第*j*列的像素值;  $m$ 和 $n$ 分别表示图像 $I$ 的高和宽。当MSE值越小时, 说明两幅图像越接近, 即近似图像 $K$ 的质量越好; 当两幅图像完全相同时, MSE的取值为0。

(2) PSNR通常通过MSE计算。

$$\text{PSNR} = 10 \lg \left( \frac{\text{MAX}_I^2}{\text{MSE}} \right) = 20 \lg \left( \frac{\text{MAX}_I}{\sqrt{\text{MSE}}} \right) \quad (4)$$

其中,  $\text{MAX}_I$ 是无噪声图像 $I$ 的最大像素可能值, 当用8位二进制来表示每个像素的颜色时, 这个值是255。更一般的情况, 当每个像素用 $B$ 位二进制来表示时,  $\text{MAX}_I = 2^B - 1$ 。

(3) SSIM表示为参考信号和失真信号的结构信息的度量。

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (5)$$

其中,  $\mu_x$ 和 $\sigma_x^2$ 分别是 $x$ 的均值和方差,  $\mu_y$ 和 $\sigma_y^2$ 分别是 $y$ 的均值和方差,  $\sigma_{xy}$ 是 $x$ 和 $y$ 的协方差,  $C_1 = (K_1L)^2, C_2 = (K_2L)^2$ 是用来维持稳定的参数,  $L$ 是像素值的动态范围,  $K_1 = 0.01, K_2 = 0.03$ 。

(4) IFC表示为源图像和失真图像之间共享的统计信息, 并不是失真程度的度量, 而是保真度的度量。

$$\text{IFC} = \sum_{k \in \text{subbands}} I(C^{N_k, k}; D^{N_k, k} | s^{N_k, k}) \quad (6)$$

其中,  $C^{N_k, k}$ 表示来自第*k*个子带的随机场 $C^k$ 的 $N_k$ 系数,  $D^{N_k, k}$ 和 $s^{N_k, k}$ 类似。

### 4.3 模型训练

本文使用随机梯度下降法, 对ERCNN以端端的联合方式进行训练, 其中设置mini-batch大小为64, 冲量为0.9, 权值的衰减系数为 $5 \times 10^{-4}$ , 最大迭代次数为 $5 \times 10^6$ 次。在 $1 \times 10^6$ 次迭代后, 每隔 $1 \times 10^6$ 次保存一个模型, 在迭代 $3.18 \times 10^6$ 时停止, 最终获取精度最高的模型。

## 5 结果与分析

### 5.1 去除眼镜的结果

利用ERCNN方法, 在SLLFW数据集上采用不同人脸对齐方式分别对不同类型的眼镜, 如全框、半框和无框3种眼镜进行去除, 如图4所示。其中第1列为原始人脸图像; 第2列第1行为戴全框眼镜的人脸图像, 第2行为眼镜去除后的效果; 第3列第1行为戴半框眼镜的人脸图像, 第2行为眼镜去除后的效果; 第4列第1行为戴无框眼镜的人脸图像, 第2行为眼镜去除后的效果。由图可知, 全框眼镜和半框眼镜的去除效果较好, 相比原始人脸图像的差异较小; 无框眼镜的去除效果一般, 去除的人脸图像上仍然有比较明显的眼镜残留痕迹, 但对眼部区域的遮挡影响较小。

相比传统的PCA方法处理的结果, 本文方法ERCNN在各项评价指标上具有一定的优势, 如表1所示, 其中采用ERCNN和PCA方法<sup>[2]</sup>利用 $128 \times 128$ 的对齐方式在全框、半框和无框3种类型的眼镜去除后与不戴眼镜的原始人脸图像, 按照指标PSNR, SSIM, MSE和IFC得到的平均值为35.3, 0.98, 19.01和7.2。特别在MSE指标上, ERCNN相比PCA提高了10.4倍; 在SSIM, IFC和PSNR指标上, ERCNN方法比PCA分别高出0.12, 0.1和10.1。因此, 从各项指标分析, 利用ERCNN处理的人脸图像有较好的效果。如果利用 $112 \times 96$ 的对齐方式, 在全框、半框和无框3种类型的眼镜去除后与不戴眼镜的原始人脸图像的各项指标均值为36.0, 0.99, 16.37和7.5, 相比 $128 \times 128$ 的人脸对齐方式,



图 4 人脸眼镜去除的结果

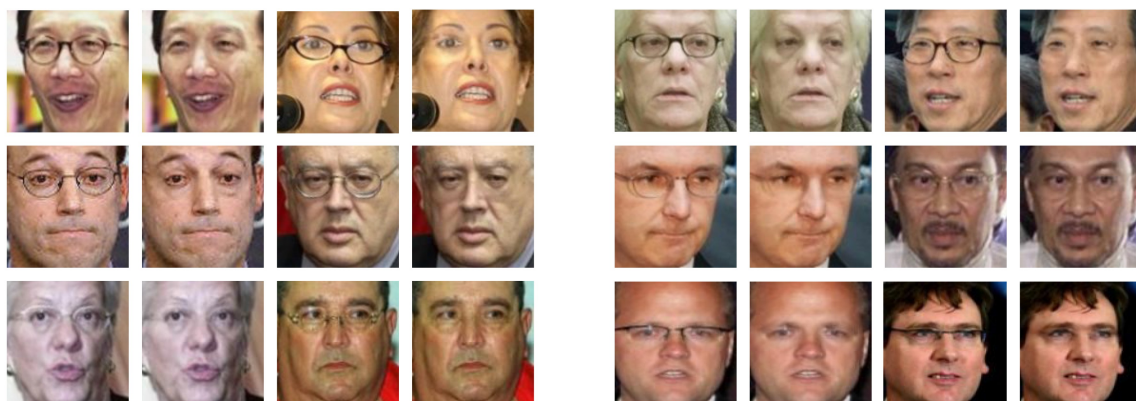
表 1 不同方法以及不同类型的眼镜去除性能对比

人脸对齐	眼镜类型	方法	评价指标				
			PSNR (dB)	SSIM	MSE	IFC	
128×128	全框	ERCNN	34.7	0.98	22.03	7.0	
		PCA	24.2	0.85	247.22	6.1	
	半框	ERCNN	35.7	0.99	17.5	7.3	
		PCA	25.5	0.87	183.27	6.2	
	无框	ERCNN	35.7	0.99	17.5	7.5	
		PCA	26.0	0.88	163.34	6.3	
	平均值	ERCNN	<b>35.3</b>	<b>0.98</b>	<b>19.01</b>	<b>7.2</b>	
		PCA	<b>25.2</b>	<b>0.86</b>	<b>197.94</b>	<b>6.2</b>	
	112×96	全框	ERCNN	35.6	0.99	17.91	7.3
			PCA	24.3	0.84	241.59	4.5
半框		ERCNN	36.3	0.99	15.24	7.6	
		PCA	25.0	0.85	205.63	4.5	
无框		ERCNN	36.1	0.99	15.96	7.7	
		PCA	24.9	0.85	210.42	4.6	
平均值		ERCNN	<b>36.0</b>	<b>0.99</b>	<b>16.37</b>	<b>7.5</b>	
		PCA	<b>24.7</b>	<b>0.84</b>	<b>219.21</b>	<b>4.5</b>	

每项指标都有一定程度的提升, 说明112×96的对齐方式更适合ERCNN的眼镜去除方法。

为了更好的说明本文方法的性能, 从SLLFW数据集中选取真实戴眼镜人脸图像作为ERCNN的测试集, 如图5所示。图5(a)和5(b)中, 第1列和第3列为真实戴不同类型眼镜的人脸图像, 第2列和第4列为经过本文方法进行眼镜去除后的人脸图像。

利用主观评价的方法, 图5(a)和图5(b)中全框、半框和无框眼镜的去除效果都较好, 不仅保持了戴眼镜区域原有的信息, 还没有明显的残留眼镜信息。因此, 利用本文方法对以上3种不同类型的眼镜得到较好的效果。但对于墨镜和带有颜色的眼镜的人脸图像, 利用本文方法进行处理后, 获得的效果较差, 如图6所示, 第1列为戴墨镜的人脸图像, 无法利用本文方法进行眼镜去除; 第2列是戴全框眼镜的人脸, 由于边框较细, 眼镜去除的效果不太明显; 第3列为戴无框眼镜的人脸图像, 由于眼镜带有颜色, 通过ERCNN方法处理后, 仍然无明显变化。



(a) 戴全框和无框眼镜人脸的眼镜去除结果

(b) 戴全框和半框眼镜人脸的眼镜去除结果

图 5 真实戴眼镜人脸的眼镜去除的结果

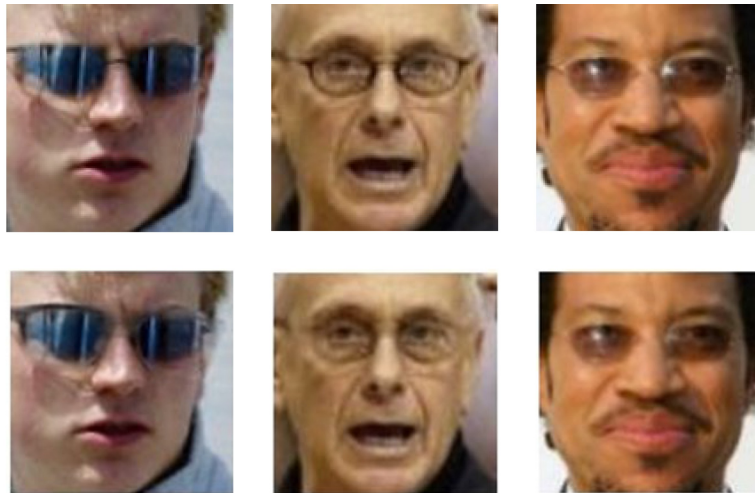


图6 真实戴眼镜人脸的眼镜去除效果不佳的结果

## 5.2 去除眼镜后对细粒度人脸识别结果的影响

为了验证本文方法对细粒度人脸识别的影响,选择4种公开的人脸识别算法,标记为①,②,③,④<sup>[17,25-27]</sup>,对LFW人脸数据集、文献[10]的细粒度人脸数据集和本文的戴全框眼镜的人脸数据集、戴半框眼镜的人脸数据集和戴无框眼镜的人脸数据集,分别标记F-SLLFW, H-SLLFW和R-SLLFW进行验证。本文中人脸识别的评价指标采用TAR(True Accept Rate)和FAR(False Accept Rate)<sup>[26]</sup>,其中TAR表示正确接受的比例,即相同人的分数大于设定阈值的数量除以相同人比较的次数;FAR表示错误接受的比例,即不同人的分数大于设定阈值的数量除以不同人比较的次数,如表2所示。

如表2所示,在FAR为1%的情况下,①,②,③和④算法在LFW数据集上的TAR,比在SLLFW数据集上分别下降了5.86%,19.7%,6.54%和3.29%;同样,在FAR为0.1%的情况下,①,②,③和④算法的TAR分别下降了11.17%,29.04%,10.47%和5.66%,说明在细粒度人脸数据集上,在FAR的越小时,不同的人脸识别算法都未达到较高的准确率。在实际应用中,人脸识别中不仅存在相似人脸,还会存在眼镜遮挡情况。因此,利用本文方法对3种不同类型的眼镜进行去除后,再利用公开的人脸识别算法进行验证。表2中,当FAR为1%时,在F-SLLFW数据集上算法④的TAR为86.13%,比其他算法分别提高了5.26%,30.36%和2.93%;当FAR为0.1%时,算法④的TAR为71.13%,比其他算法分别提高了3.46%,30.36%和2.93%,说明算法④在未去除眼镜时,人脸识别准确率优于其他算法。当利用本文方法对F-SLLFW数据集进行眼镜去除后,在FAR为1%的情况

下,算法④的TAR为90.05%,比未去除眼镜时提高了3.92%;在FAR为0.1%的情况下,算法④的TAR为81.79%,比未去除眼镜时提高了10.06%。同样,利用本文方法对H-SLLFW和R-SLLFW数据集进行眼镜去除后,在FAR为1%的情况下,算法④的TAR为分别为91.14%和92.33%,比未去除眼镜时提高了3.08%和1.26%;在FAR为0.1%的情况下,算法④的TAR提高了4.29%和2.13%。说明对人脸进行眼镜去除后,在不同的FAR条件下,都有助于提升人脸识别的准确率。

## 5.3 实时性测试结果

在处理速度上,本文方法在Intel Core i7-7700

表2 不同人脸识别的识别性能对比(%)

数据集	FAR	TAR			
		①	②	③	④
LFW	1	99.03	98.13	99.07	<b>99.40</b>
	0.1	95.50	91.27	95.90	<b>97.04</b>
SLLFW	1	93.17	78.43	92.53	<b>96.11</b>
	0.1	84.33	62.23	85.43	<b>91.38</b>
F-SLLFW	1	80.87	55.77	83.20	<b>86.13</b>
	0.1	68.27	39.87	70.70	<b>71.73</b>
本文方法对F-SLLFW进行处理后	1	86.47	64.47	88.63	<b>90.05</b>
	0.1	77.07	48.10	77.87	<b>81.79</b>
H-SLLFW	1	83.10	60.70	86.27	<b>88.06</b>
	0.1	72.80	43.53	75.10	<b>78.36</b>
本文方法对H-SLLFW进行处理后	1	87.60	67.07	89.93	<b>91.14</b>
	0.1	80.70	45.53	78.83	<b>82.65</b>
R-SLLFW	1	87.07	66.10	89.37	<b>91.70</b>
	0.1	79.53	49.67	76.73	<b>82.76</b>
本文方法对R-SLLFW进行处理后	1	88.53	67.50	89.13	<b>92.33</b>
	0.1	81.07	47.80	77.43	<b>84.89</b>

CPU 3.60 GHz、内存16GB、显卡为GTX1080的设备上，对128×128像素人脸图像的处理时间为3 ms，模型占用内存大小为11.9 MB，说明使用该方法在不会增加太多计算资源和内存资源的情况，满足实际应用的需求。

## 6 结论

本文设计了一种用于细粒度人脸识别的眼镜去除网络ERCNN，适用于对全框、半框和无框3种眼镜的去除。具体地：(1)利用大量的监控场景、互联网和证件照的人脸图像构建数据集，保证了人脸图像的多样性，作为研究对象，设计了一种ERCNN网络，利用常用的网络结构，融合MFM单元和反卷积层，很好地实现了眼镜去除。在眼镜去除精度方面，本文方法在PSNR, SSIM, MSE和IFC上的平均值为35.3, 0.98, 19.01和7.2，相比PCA方法分别提高了10.4倍, 0.12, 0.1和10.1，在各项指标上都具有一定优势。在识别速度上，本文方法的处理速度仅为3 ms，也具有较好的实时性。(2)在细粒度人脸识别测试集上进行对比试验，当FAR为1%时，本文方法对F-SLLFW, H-SLLFW和R-SLLFW数据集的人脸图像进行眼镜去除后，SphereFace算法的TAR分别达到90.05%, 91.14%和92.33%，比未去除眼镜的识别率分别提高了3.92%, 3.08%和1.26%；同样，当FAR为0.1%时，比SphereFace算法的TAR分别提高了10.06%, 4.29%和2.13%，说明本文方法有助于提升细粒度人脸识别的识别精度。

## 参考文献

- [1] 李小新, 梁荣华. 有遮挡人脸识别综述: 从子空间回归到深度学习[J]. 计算机学报, 2018, 41(1): 177-207. doi: [10.11897/SP.J.1016.2018.00177](https://doi.org/10.11897/SP.J.1016.2018.00177).  
LI Xiaoxin and LIANG Ronghua. A review for face recognition with occlusion: From subspace regression to deep learning[J]. *Chinese Journal of Computers*, 2018, 41(1): 177-207. doi: [10.11897/SP.J.1016.2018.00177](https://doi.org/10.11897/SP.J.1016.2018.00177).
- [2] WU Chenyu, LIU Ce, SHUM H Y, *et al.* Automatic eyeglasses removal from face images[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(3): 322-336. doi: [10.1109/TPAMI.2004.1262319](https://doi.org/10.1109/TPAMI.2004.1262319).
- [3] DU Cheng and SU Guangda. Eyeglasses removal from facial images[J]. *Pattern Recognition Letters*, 2005, 26(14): 2215-2220. doi: [10.1016/j.patrec.2005.04.002](https://doi.org/10.1016/j.patrec.2005.04.002).
- [4] PARK J S, OH Y H, AHN S C, *et al.* Glasses removal from facial image using recursive error compensation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(5): 805-811. doi: [10.1109/TPAMI.2005.103](https://doi.org/10.1109/TPAMI.2005.103).
- [5] HEO J and SAVVIDES M. Face pose correction with eyeglasses and occlusions removal[C]. 2007 IEEE Conference Biometrics Symposium, Baltimore, USA, 2007: 1-6. doi: [10.1109/BCC.2007.4430551](https://doi.org/10.1109/BCC.2007.4430551).
- [6] JIA Xiaodong and GUO Jiangling. Eyeglasses removal from facial image based on phase congruency[C]. The 3rd IEEE International Congress on Image and Signal Processing, Yantai, China, 2010: 1859-1862, doi: [10.1109/CISP.2010.5647366](https://doi.org/10.1109/CISP.2010.5647366).
- [7] WANG Yuankai, JANG J H, TSAI L W, *et al.* Improvement of face recognition by eyeglass removal[C]. The 6th IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Darmstadt, Germany, 2010: 228-231. doi: [10.1109/IHMSP.2010.64](https://doi.org/10.1109/IHMSP.2010.64).
- [8] CHENG Wenchang, LIAO H C, PAN M H, *et al.* A fatigue detection system with eyeglasses removal[C]. The 15th IEEE International Conference on Advanced Communications Technology, PyeongChang, South Korea, 2013: 331-335.
- [9] 李根, 李文辉. 基于思维进化的机器学习的遮挡人脸识别[J]. 吉林大学学报: 工学版, 2014, 44(5): 1410-1416.  
LI Gen and LI Wenhui. Face occlusion recognition based on MEBML[J]. *Journal of Jilin University: Engineering and Technology Edition*, 2014, 44(5): 1410-1416.
- [10] GUO Pei and SU Fei. Enhanced PCA reconstruction method for eyeglass frame auto-removal[C]. The 4th IEEE International Conference on Network Infrastructure and Digital Content, Beijing, China, 2014: 359-363. doi: [10.1109/ICNIDC.2014.7000325](https://doi.org/10.1109/ICNIDC.2014.7000325).
- [11] 吴从中, 刘渠芬, 詹曙. 基于稀疏表达的人脸遮挡物去除[J]. 合肥工业大学学报: 自然科学版, 2015, 38(5): 615-617, 680.  
WU Congzhong, LIU Qufen, and ZHAN Shu. A face occlusion removal method based on sparse representation[J]. *Journal of Hefei University of Technology*, 2015, 38(5): 615-617, 680.
- [12] FERNÁNDEZ A, GARCÍA R, USAMENTIAGA R, *et al.* Glasses detection on real images based on robust alignment[J]. *Machine Vision and Applications*, 2015, 26(4): 519-531. doi: [10.1007/s00138-015-0674-1](https://doi.org/10.1007/s00138-015-0674-1).
- [13] SANDHAN T and CHOI J Y. Anti-glare: Tightly constrained optimization for eyeglass reflection removal[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 1675-1684. doi: [10.1109/CVPR.2017.182](https://doi.org/10.1109/CVPR.2017.182).
- [14] 曹志义, 牛少彰, 张继威. 基于半监督学习生成对抗网络的人脸还原算法研究[J]. 电子与信息学报, 2018, 40(2): 323-330. doi: [10.11999/JEIT170357](https://doi.org/10.11999/JEIT170357).  
CAO Zhiyi, NIU Shaozhang, ZHANG Jiwei. Research on face reduction algorithm based on generative adversarial nets with semi-supervised learning[J]. *Journal of Electronics*

- & *Information Technology*, 2018, 40(2): 323–330. doi: [10.11999/JEIT170357](https://doi.org/10.11999/JEIT170357).
- [15] 毛亮, 魏颖慧, 刘三阳, 等. 一种用于细粒度人脸识别的眼镜去除方法[P]. 中国专利, 110569763A, 2019.  
MAO Liang, WEI Yinghui, LIU Sanyang, *et al.* Glasses removing method for fine-grained face recognition[P]. China Patent, 110569763A, 2019.
- [16] DONG Chao, CHEN C L, HE Kaiming, *et al.* Learning a deep convolutional network for image super-resolution[C]. The 13th European Conference on Computer Vision, Zurich, Switzerland, 2014: 184–199. doi: [10.1007/978-3-319-10593-2\\_13](https://doi.org/10.1007/978-3-319-10593-2_13).
- [17] WEN Yandong, ZHANG Kaipeng, LI Zhifeng, *et al.* A discriminative feature learning approach for deep face recognition[C]. The 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016, 9911: 499–515. doi: [10.1007/978-3-319-46478-7\\_31](https://doi.org/10.1007/978-3-319-46478-7_31).
- [18] WU Xiang, HE Ran, and SUN Zhenan. A lightened CNN for deep face representation[EB/OL]. <https://arxiv.org/abs/1511.02683v1>, 2015.
- [19] 谭小慧, 李昭伟, 樊亚春. 基于多尺度细节增强的面部表情识别方法[J]. 电子与信息学报, 2019, 41(11): 2752–2759. doi: [10.11999/EIT181088](https://doi.org/10.11999/EIT181088).  
TAN Xiaohui, LI Zhaowei, and FAN Yachun. Facial expression recognition method based on multi-scale detail enhancement[J]. *Journal of Electronics & Information Technology*, 2019, 41(11): 2752–2759. doi: [10.11999/EIT181088](https://doi.org/10.11999/EIT181088).
- [20] DENG Weihong, HU Jiani, ZHANG Nanhai, *et al.* Fine-grained face verification: FGLFW database, baselines, and human-DCMN partnership[J]. *Pattern Recognition*, 2017, 66: 63–73. doi: [10.1016/j.patcog.2016.11.023](https://doi.org/10.1016/j.patcog.2016.11.023).
- [21] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. <https://arxiv.org/abs/1409.1556v6>, 2015.
- [22] SZEGEDY C, LIU Wei, JIA Yangqing, *et al.* Going deeper with convolutions[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 1–9. doi: [10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594).
- [23] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [24] JIA Yangqing, SHELHAMER E, DONAHUE J, *et al.* Caffe: Convolutional architecture for fast feature embedding[EB/OL]. <https://arxiv.org/abs/1408.5093>, 2014.
- [25] WANG Feng, CHENG Jian, LIU Weiyang, *et al.* Additive margin softmax for face verification[J]. *IEEE Signal Processing Letters*, 2018, 25(7): 926–930. doi: [10.1109/LSP.2018.2822810](https://doi.org/10.1109/LSP.2018.2822810).
- [26] DENG Jiankang, GUO Jia, XUE Niannan, *et al.* ArcFace: Additive angular margin loss for deep face recognition[EB/OL]. <https://arxiv.org/abs/1801.07698>, 2018.
- [27] LIU Weiyang, WEN Yandong, YU Zhiding, *et al.* SphereFace: Deep hypersphere embedding for face recognition[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 6738–6746. doi: [10.1109/CVPR.2017.713](https://doi.org/10.1109/CVPR.2017.713).
- 毛亮: 男, 1983年生, 副研究员, 研究方向计算机视觉与深度学习.
- 薛月菊: 女, 1969年生, 教授, 主要研究方向机器视觉与图像处理计算机视觉与深度学习.

责任编辑: 余蓉