

基于多尺度卷积神经网络的人脸润饰检测算法^①



张 萌, 王晓峰, 胡姣姣, 张德鹏

(西安理工大学 理学院, 西安 710054)

通讯作者: 张 萌, E-mail: mengzhang0727@163.com

摘 要: 针对目前人脸润饰检测算法特征提取复杂、识别率低的问题, 提出了一种基于多尺度卷积神经网络的人脸润饰检测算法. 不同于传统的卷积神经网络, 本文的网络增加了图像预处理, 利用基于方向梯度直方图 (Histograms of Oriented Gradient, HOG) 特征的人脸提取算法从原始图像中提取出人脸部分; 在第一个池化层后连接局部归一化 (Local Response Normalization, LRN) 层, 加速模型的收敛; 提出了多尺度卷积层, 将大小为 1×1 , 3×3 和 5×5 的卷积核进行级联, 提高模型分类效果. 实验结果表明, 本文算法的检测精度在人脸润饰数据集 LFW 和 ND-IIITD 分别达到 99.5% 和 92.9%, 相比于主流网络结构和最新人脸润饰检测算法, 检测精度有显著提高.

关键词: 深度学习; 卷积神经网络; 人脸润饰检测; 局部归一化; 多尺度卷积

引用格式: 张萌, 王晓峰, 胡姣姣, 张德鹏. 基于多尺度卷积神经网络的人脸润饰检测算法. 计算机系统应用, 2019, 28(10): 130-137. <http://www.c-s-a.org.cn/1003-3254/7072.html>

Facial Image Retouching Detection Algorithm Based on Multi Scale-Convolutional Neural Network

ZHANG Meng, WANG Xiao-Feng, HU Jiao-Jiao, ZHANG De-Peng

(Faculty of Sciences, Xi'an University of Technology, Xi'an 710054, China)

Abstract: In order to solve the problem that the existing facial retouching detection algorithm has complex feature extraction and low recognition rate, we proposed a facial retouching detection algorithm based on Multi-Scale-Convolutional Neural Network (MS-CNN). Different from the traditional CNN, MS-CNN adds image preprocessing, which uses Histograms of Oriented Gradient (HOG) feature-based facial extraction algorithm to extract the facial part from the original image. It connects the Local Response Normalization (LRN) layer after the first pooling layer to accelerate the convergence of the model. A multi-scale convolution layer is proposed, which cascades convolution kernel of 1×1 , 3×3 , and 5×5 to improve the classification accuracy. The experimental results show that the detection accuracy of the proposed algorithm is 99.5% in LFW data set and 92.9% in ND-IIITD data set, respectively. Compared with the mainstream network structure and existing facial retouching detection algorithms, the detection accuracy of the proposed algorithm is significantly improved.

Key words: deep learning; convolutional neural network; facial retouching detection; local response normalization; multi-scale convolution

① 基金项目: 国家自然科学基金 (61772416); 陕西省教育厅重点实验室项目 (17JS098); 陕西省技术创新引导项目 (2018XNCG-G-02)

Foundation item: National Natural Science Foundation of China (61772416); Key Laboratory Program of Education Bureau, Shaanxi Province (17JS098); Technology Innovation Guiding Project of Shaanxi Province (2018XNCG-G-02)

收稿时间: 2019-03-01; 修改时间: 2019-03-29; 采用时间: 2019-04-08; csa 在线出版时间: 2019-10-15

随着 Photoshop 等图像润饰工具的广泛使用,越来越多的图像编辑软件被用于图像润饰.如图 1,每对图像的左边为原始图像,右边为对应的润饰图像,从图 1 可以看出图像润饰可改变图像的原始内容并且篡改痕迹肉眼难以辨认.部分恶意篡改图像如被大量地用于科学研究和法庭证据等,无疑将会对政治和社会稳定产生严重的影响,因此数字图像取证技术的研究显得尤为重要.



图 1 原始图像及其润饰图像的例子

1 相关工作

图像润饰通常改变了图像的颜色、纹理、亮度,常见操作包括对图像进行局部模糊^[1]等,目前的润饰检测算法主要有传统算法和深度学习两类,下文将分别对两类算法进行介绍.

1.1 传统人脸润饰检测算法

传统的人脸润饰检测算法主要依赖于人工设计判别特征,最后建立特征分类模型实现润饰图像分类.2011年,Ananya 等人^[2]首次将润饰图像作为篡改图像来检测,提出用期望最大化算法 (Expectation-Maximization, EM) 实现图像分类,但该算法只能将篡改图像种类的范围缩小,无法彻底区分润饰图像与复制粘贴图像.2011年,李晓飞等人^[3]提出了基于双拉普拉斯变换的润饰图像盲检测技术,该方法使用 KD-Tree 和分层聚类法来匹配图像块,应用 7-tap 拉普拉斯滤波器消除误匹配,最终实现润饰图像的局部润饰检测.该算法对于未压缩的和高分辨率的压缩图像有效,但是对于不同色调和饱和度的图像有误检,对低质量的压缩图像误检较多.2011年,Farid 等人^[4]引入一个定量的润饰图像度量标准,通过建模估计图像的几何和光度变化,用 8 个统计量将图像按润饰程度分为 5 级,其中几何修正量化为 4 个统计量,即待测对象面部和身体的运动幅度的平均值和标准差,光度修正也量化为 4 个统计量,包括局部平滑/锐化滤波器在空间边界的均值和标准差,和结构相似性指标度量 (Structural Similarity

Index Metric, SSIM) 的均值和标准差.该度量标准能定量的判断图像的润饰程度,缺点是需要原始图像作参考.2013年,Shah 等人^[5]使用人类感知来检测润饰图像,首先采用 PCA 技术对待测图像进行人脸识别,然后将待测图像与其润饰图像做对比,最后通过人类的感知确定待测图像是否润饰图像.

1.2 基于深度学习的人脸润饰检测算法

2006年,加拿大多伦多大学教授 Hinton 在顶尖学术刊物《科学》上发表了一篇文章^[6],开启了深度学习在学术界和工业界的浪潮.近几年随着深度学习在图像分类^[7-11]、目标检测^[12]、图像语义分割^[13]等领域取得的一系列突破性的研究成果,研究者们开始将深度学习应用于图像取证领域.2015年,Kang 等人^[14]提出了一种基于卷积神经网络的中值滤波检测方法,通过添加滤波器层来改进传统的 CNN 模型,添加的滤波层可以抑制由于图像边缘和纹理的存在引起的干扰,使得中值滤波留下的迹线可以成功暴露,这是深度学习应用在取证领域的第一个研究成果,在此之后,卷积神经网络陆续应用在其他取证领域,例如图像重拍摄检测^[15]、相机源识别^[16]、图像隐写分析^[17]、图像拼接及复制粘贴篡改检测^[18]等.随着深度学习在图像取证领域的广泛应用,研究者们也开始将深度学习用于人脸润饰检测.在 2016年,Bharati 等人^[19]设计了第一个 ND-IIITD 人脸润饰数据库,首次提出用监督的深度玻尔兹曼机 (Supervised Deep Boltzmann Machine, SDBM) 来提取特征,支持向量机来对原始图像和润饰图像分类,检测精度比传统人脸润饰检测算法高 38.3%.2017年,Bharati 等人^[20]又提出了一种新的半监督自动编码器来提取特征,并用支持向量机分类,但这两个算法的特征提取和图像分类都是分步进行的,不是端到端的润饰图像检测算法.

1.3 本文工作

由上文可知,传统人脸润饰检测算法手动设计特征复杂且识别率低,现有深度学习算法无法实现端到端的润饰图像检测,为了解决以上问题,本文提出了多尺度卷积神经网络模型 (Multi Scale-Convolutional Neural Network, MS-CNN),不同于传统的 CNN,本文提出的 MS-CNN 有 3 个优点:增加基于方向梯度直方图 (Histograms of Oriented Gradient, HOG) 特征的人脸提取预处理,从待测图像中提取人脸部分作为网络的唯一输入;在首次池化操作后加入局部归一化层 (Local Response Normali-

zation, LRN), 加速网络收敛; 由于传统 CNN 中使用的单层单卷积核难以捕获人脸润饰图像与原始图像间微小的类间差别, 因此提出了多尺度卷积层, 将不同尺寸的 1×1 , 3×3 和 5×5 的卷积核进行了级联, 以提高检测精度.

本文算法在人脸润饰数据集 LFW 和 ND-IIITD 进行了测试, 实验结果表明提出的算法与 4 种网络结构和最新人脸润饰检测算法相比, 检测精度均有大幅提升.

2 基于 HOG 特征的人脸提取预处理

2005 年 Dalal 提出了基于 HOG 特征^[21]的行人检测算法, 近几年该算法被越来越多的应用于人脸检测和人脸识别领域. 考虑到人脸润饰检测的目标区域为图像中的人脸部分, 因此在预处理层利用基于 HOG 特征的人脸提取算法从待测图像中提取出人脸部分, 算法步骤如下:

Step 1. 对图像 $I(x, y)$ 进行平方根的伽马压缩得到 $I(x, y)^{gamma}$, 实验中压缩参数 $gamma$ 取 0.5;

Step 2. 结合象限信息求像素点 (x, y) 处的梯度幅值 $G(x, y)$ 和梯度方向 $\theta(x, y)$:

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (1)$$

$$\theta(x, y) = \begin{cases} \arctan \frac{G_y(x, y)}{G_x(x, y)} + \pi, & \arctan \frac{G_y(x, y)}{G_x(x, y)} < 0 \\ \arctan \frac{G_y(x, y)}{G_x(x, y)}, & \text{其他} \end{cases} \quad (2)$$

其中, $G_x(x, y)$ 与 $G_y(x, y)$ 分别为像素点 (x, y) 水平和垂直方向的梯度分量, 求出梯度方向后将梯度方向 $\theta(x, y) \in [0^\circ, 180^\circ]$ 划分为 9 个区间.

Step 3. 将图像划分为 8×8 不重叠的单元格, 检测窗口在图像上滑动的步长和块在检测窗口滑动的步长

均为 8 个像素, 检测窗口大小设置为 128×64 , 块大小为 16×16 . 每个单元格中像素点 (x, y) 的梯度方向 $\theta(x, y)$ 按区间段投票, 计算区间段 i 及其相邻区间段的梯度投票 $v_i (i = 0, \dots, 8)$.

$$\begin{aligned} v_i &= (1 - \alpha) * G(x, y) \\ v_{neighb} &= \alpha * G(x, y) \end{aligned} \quad (3)$$

其中 α 为权重, 为了简化计算复杂度, 实验中取权重 $\alpha = 0.5$ ^[21]. 这样每个单元格生成一个 9 维的特征 $V_{unit}^i = (v_0^i, \dots, v_8^i)$, 每个块由四个单元格组成, 因此每个块生成一个 36 维的特征向量 $V_{block} = (V_{unit}^1, \dots, V_{unit}^4)$, 再对特征向量 V_{block} 归一化得到 V , 图像内所有块的特征向量 V 组合得到最终特征向量 V_{HOG} , 最后将 V_{HOG} 送入支持向量机来提取图像中的人脸部分.

3 多尺度卷积神经网络结构设计

多尺度卷积神经网络 (MS-CNN) 模型包括 3 部分: 数据输入、特征提取和分类, 图 2 为 MS-CNN 的结构图. 数据输入层分批次将图像送进特征提取部分, 每个批次可处理 40 张图像; 特征提取部分共 6 个卷积层, 每个卷积层都有一个 xavier 核, 使输入输出的数据服从式 (4) 的均匀分布. 同时为了防止网络过拟合, MS-CNN 中添加了 3 个池化层, 常见的池化操作包括最大池化、平均池化和重叠池化, 本文选用最大池化来对图像做下采样; 分类部分包括全链接层、Dropout 层和 Softmax 层. 下文将详细描述 MS-CNN 网络结构的设计.

$$W \sim U \left[-\frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}}, \frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}} \right] \quad (4)$$

其中, n_j 和 n_{j+1} 分别表示输入、输出该卷积层的样本个数.

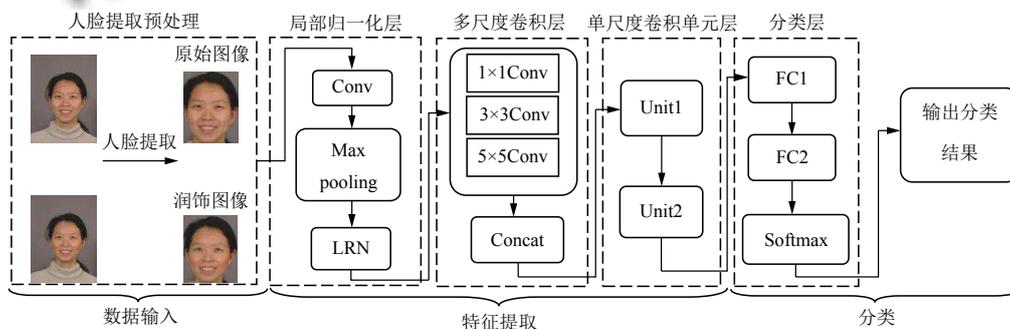


图 2 提出的多尺度卷积神经网络 (MS-CNN) 模型

3.1 MS-CNN 网络超参数

MS-CNN 首先将输入的彩色图像的大小统一调整为 128×128 , 数据输入层的 $batch_size$ 取 40, 即 MS-CNN 每个批次处理 40 张图像, 直到处理完图像库中的所有图像, 特征提取和分类部分的超参数见表 1.

表 1 MS-CNN 的超参数

层名	层类型	卷积核 尺寸	填充 步长	输出特征维数
C1	Convolution	11×11	0 2	$96 \times 59 \times 59$
C2	Max pooling	3×3	0 2	$96 \times 29 \times 29$
C3	LRN	-	-	$96 \times 29 \times 29$
MS	Multi-scale	1×1	0 1	$64 \times 29 \times 29$
	Convolution	3×3	1 1	$64 \times 29 \times 29$
Unit1	Convolution	5×5	2 1	$64 \times 29 \times 29$
	Convolution	5×5	1 1	$128 \times 27 \times 27$
	Convolution	5×5	1 1	$128 \times 25 \times 25$
	Max Pooling	2×2	0 2	$128 \times 13 \times 13$
Unit2	Convolution	5×5	1 1	$256 \times 11 \times 11$
	Max Pooling	2×2	- 2	$256 \times 5 \times 5$
FC1	Fully connected	-	-	256
FC2	Fully connected	-	-	2
S	Softmax	-	-	2

3.2 特征提取

特征提取部分是 MS-CNN 的核心, 包括 LRN 层、多尺度卷积层和单尺度卷积单元三部分, 下文详细介绍了每个部分.

3.2.1 局部归一化 (LRN) 层

一张 128×128 的图像首先进入数据输入层, 从数据输入层输出后进入第一个卷积层, 该层设有 96 个 11×11 的大尺寸卷积核 k_{ij}^1 , 连接输入的第 i ($i = 1, \dots, 40$)个图像与本层的第 j ($j = 1, \dots, 96$)个卷积核, 卷积生成 96 个大小为 59×59 的特征图 x_j^1 ($j = 1, \dots, 96$), 在卷积层后紧跟一个最大池化层, 对特征图 x_j^1 做下采样. 第一个卷积与池化操作的公式如下:

$$\begin{cases} x_j^1 = \text{ReLU} \left(\sum_{i=1}^{batch_size} x_i * k_{ij}^1 + b_j^1 \right) \\ x_j^{pool} = \beta_j^{pool} \text{down}(x_j^1) + b_j^{pool} \end{cases} \quad (5)$$

其中, b_j^1 表示第一层卷积的偏置, $\text{ReLU}(\cdot)$ 为激活函数, $batch_size$ 表示 MS-CNN 每个批次处理图像的个数, 本文中取 $batch_size = 40$; x_j^{pool} 为池化层输出的特征图, β_j^{pool} 和 b_j^{pool} 分别表示池化层的权重和偏置, $\text{down}(\cdot)$ 为最大池化操作.

LRN 仿造生物学上活跃的神经元对相邻神经元的

侧抑制现象^[8], 对局部神经元的活动创建竞争机制, 使得其中响应比较大的值变得相对更大, 并抑制其他反馈较小的神经元, 因此在池化层后添加 LRN 层有助于网络的快速收敛. 该层对输入的 96 个特征图 x_j^{pool} 分别按像素做归一化处理, 输出归一化后的特征图 x_j^{LRN} ($j = 1, \dots, 96$):

$$x_j^{LRN} = \frac{x_j^{pool}}{\left(k + \alpha \sum_{q=\max(0, j-n/2)}^{\min(N-1, j+n/2)} (x_q^{pool})^2 \right)^\beta} \quad (6)$$

其中, x_q^{pool} 表示池化层输出的第 q ($q = 1, \dots, N$)个特征图, $N = 96$ 为特征图总数, n, k, α 和 β 为参数, 在本文中分别取 $n = 5, k = 2, \alpha = 10^{-4}, \beta = 0.75$.

3.2.2 多尺度卷积层

计算机视觉领域的图像分类问题类间差别较大, 多为不同类别的分类问题, 例如猫狗分类、汽车卡车分类等. 而图像取证的类间差别微小, 传统的卷积神经网络难以区分, 例如文献^[14]中为了放大中值滤波取证的类间差别, 增加了中值滤波残差层作为预处理, 同样文献^[15]在预处理层利用拉普拉斯滤波器来放大重拍摄图像与原始图像间的类间差别. 而润饰图像取证属于图像取证中的一类, 润饰图像与原始图像间的差异也很小, 多为图像局部的亮度、纹理、饱和度等, 因此传统单尺度的卷积层难以捕获这种微小的类间差别, 因此本文提出多尺度卷积层, 该层级联了 3 种大小分别为 $1 \times 1, 3 \times 3$ 和 5×5 的 xavier 卷积核各 64 个, 分别对 LRN 层输出的特征图 x_j^{LRN} 做卷积, 多尺度卷积操作表示为:

$$\begin{cases} f_i^{1 \times 1} = \text{ReLU} \left(\sum_{i=1}^{n_{LRN}} x_j^{LRN} * k_{ij}^{1 \times 1} + b_i^{1 \times 1} \right) \\ f_i^{3 \times 3} = \text{ReLU} \left(\sum_{i=1}^{n_{LRN}} x_j^{LRN} * k_{ij}^{3 \times 3} + b_i^{3 \times 3} \right) \\ f_i^{5 \times 5} = \text{ReLU} \left(\sum_{i=1}^{n_{LRN}} x_j^{LRN} * k_{ij}^{5 \times 5} + b_i^{5 \times 5} \right) \end{cases} \quad (7)$$

其中, $k_{ij}^{1 \times 1}$ 为 1×1 卷积核, 连接 LRN 层的第 j 个输出特征图 x_j^{LRN} 与多尺度卷积层的第 i 个输出特征图 $f_i^{1 \times 1}$, $b_i^{1 \times 1}$ 为偏置, n_{LRN} 为 LRN 层输出特征图的个数, $n_{LRN} = 96$, $\text{ReLU}(\cdot)$ 为激活函数, $f_i^{3 \times 3}$ 与 $f_i^{5 \times 5}$ 同理.

为了使特征图 $f_i^{1 \times 1}$ 、 $f_i^{3 \times 3}$ 与 $f_i^{5 \times 5}$ 大小相同, 针对 3 种不同尺寸的卷积核分别设定填充参数 pad : $pad_{1 \times 1} = 0$, $pad_{3 \times 3} = 1$, $pad_{5 \times 5} = 2$. 最后由 $\text{Concat}(\cdot)$ 将特征图 $f_i^{1 \times 1}$ 、

$f_i^{3 \times 3}$ 与 $f_i^{5 \times 5}$ 连接,得到最终由多尺度卷积层输出的特征图 $f_i^{Multi_Scale}$,该操作表示为:

$$f_i^{Multi_Scale} = Concat(f_i^{1 \times 1}, f_i^{3 \times 3}, f_i^{5 \times 5}) \quad (8)$$

3.2.3 单尺度卷积单元

多个多尺度卷积层堆叠会产生大量的参数,降低网络训练效率,因此本文没有采取对多尺度卷积层进行简单的堆叠,而是在它之后增加了两个单尺度卷积单元 $unit1$ 和 $unit2$.每个单尺度卷积单元都有两个卷积层,一个最大池化层,每个卷积层后都有一个 $ReLU(\cdot)$ 激活函数来加速网络收敛. $unit1$ 和 $unit2$ 的结构类似但卷积核的个数不同,卷积核个数分别为128和256, $unit1$ 和 $unit2$ 的具体结构和详细参数见3.1节.

3.3 分类

分类层分3部分:2个全连接层、Dropout层和Softmax层.全连接层将特征提取部分得到的所有二维特征图转化为全连接的一维特征向量;Dropout层^[9]以概率 $p = 0.5$ 随机的将部分隐藏神经元的输出清零,避免网络过拟合;最后由Softmax分类器将多个神经元的输出映射到(0, 1)区间内,实现润饰图像与原始图像的分类.

4 实验

4.1 实验环境

本文实验采用的深度学习平台为Caffe^[22],电脑硬件环境的内存为8GB,CPU为Intel(R)Core(TM)i7,GPU为GTX 1050 Ti,操作系统为64位Ubuntu 16.04.

4.2 实验数据

带标签的人脸(Labeled Faces in the Wild, LFW)数据集:LFW是一个用于研究无约束的人脸识别的数据库,包含了13 233张人脸图像,图像大小均为 250×250 .我们从LFW中随机选取1000张人脸图像,利用Meitu软件中的润饰功能,即:人像美容中的智能美容,对1000张人脸进行润饰,这2000张图像组成LFW润饰数据集,该数据集的部分示例图见图3,每对示例图的左边为原始图像,右边为对应的润饰图像.

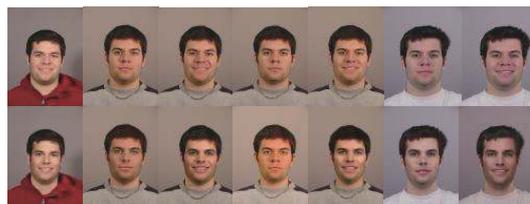


图3 LFW润饰数据集的示例

ND-IIITD人脸润饰数据集:ND-IIITD是Bharati^[19]提出的第一个人脸润饰数据集,图像大小均为 1200×1600 ,示例见图4,图4(a)(b)中的第一行为原始图像,第二行为对应的润饰图像.该数据库对325个对象进行照片采集,每个对象采集8张照片,因此该数据集共有2600张原始图像,并利用PortraitPro Studio Max软件对每个对象中的7张照片进行人脸润饰操作,该润饰操作对男性女性分别预设了7种不同的润饰操作组合,这些润饰操作都不同程度的改变了面部特征,包括:皮肤的纹理,眼睛、鼻子或嘴唇的形状,眼睛的颜色等等.



(a) 女性7种不同的润饰操作示例图



(b) 男性7种不同的润饰操作示例图

图4 ND-IIITD人脸润饰数据集的示例

实验中人脸润饰数据集LFW与ND-IIITD的训练集及测试集数量见表2.

表2 训练样本和测试样本数量

数据集	原始	润饰	总共	训练集	测试集
LFW	1000	1000	2000	1600	400
ND-IIITD	2600	2275	4875	3900	975

4.3 实验参数设置

到目前为止,深度学习平台Caffe总共提供了6种优化算法,本实验中选取的优化算法为随机梯度下降算法(Stochastic Gradient Descent, SGD).

在模型训练中,基础学习率(base_lr)为0.001,在训练过程中对基础学习率进行调整,学习率调整策略(lr-policy)为“inv”,学习率调整策略“inv”的系数gamma和power分别为0.0001和0.75.上一次梯度更新的权重(momentum)为0.9,基础网络权值衰减项(weight_decay)为0.0005,用来防止网络过拟合.每训练1000次在屏幕上打印一次训练详情,方便观察检测精度和

损失值随迭代次数的变化结果,最大迭代次数(max-iter)为100 000次,为了节省训练时间,运行模式(slover_mode)选GPU.

4.4 性能评估

4.4.1 方法有效性验证

为了验证本文提出的MS-CNN网络结构的有效性,分别在人脸润饰数据集LFW和ND-IIITD上进行测试,实验结果见表3和图5.

表3 MS-CNN在不同迭代次数下的检测精度和损失值

迭代次数	LFW 人脸润饰图像库		ND-IIITD 人脸润饰图像库	
	检测精度	损失值	检测精度	损失值
1000	0.7289	0.5861	0.5901	0.4847
2000	0.8500	0.3805	0.7386	0.3697
3000	0.8859	0.1453	0.7893	0.3013
4000	0.9234	0.1219	0.8166	0.0913
5000	0.9375	0.0856	0.8353	0.0639
6000	0.9422	0.0592	0.8562	0.0282
7000	0.9516	0.0513	0.8643	0.0272
8000	0.9570	0.0413	0.8697	0.0081
9000	0.9594	0.0261	0.8712	0.0042
10 000	0.9648	0.0231	0.8750	0.0035
20 000	0.9719	0.0144	0.8938	0.0017
30 000	0.9727	0.0130	0.9063	0.0012
40 000	0.9773	0.0073	0.9138	1.8138×10^{-4}
50 000	0.9789	0.0030	0.9150	1.7344×10^{-4}
60 000	0.9844	0.0025	0.9163	1.4122×10^{-4}
70 000	0.9867	0.0024	0.9163	1.0943×10^{-4}
80 000	0.9891	0.0011	0.9175	2.0887×10^{-5}
90 000	0.9898	0.0010	0.9188	2.0586×10^{-5}
100 000	0.9930	5.1511×10^{-4}	0.9200	1.3091×10^{-5}

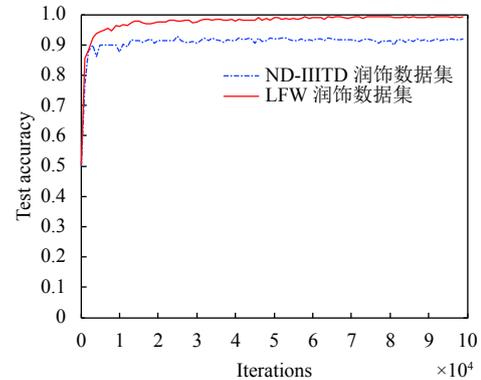
从表3和图5可以看出,检测精度随着迭代次数的增加先快速增长然后趋于平稳,且LFW数据集的检测精度略优于ND-IIITD数据集,原因是ND-IIITD数据集的润饰操作更精细,图像分辨率更高,因此润饰图像与原始图像的像素间差异较小,使得ND-IIITD数据集比LFW数据集较难检测,故检测精度略低.

损失值代表实际值与模型估计值之间的误差,损失值越小说明模型拟合的越好.图5(b)为本文构建的模型在不同数据集上的训练损失值变化情况,随着迭代次数的增加,两个数据集上的损失值均逐渐递减并趋于稳定,说明本文提出的模型拟合的较好.

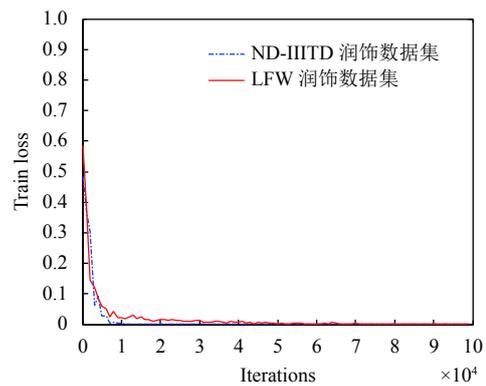
4.4.2 性能比较

近几年用于图像分类的卷积神经网络层出不穷,其中LeNet^[7]与AlexNet^[8]是比较经典的卷积神经网络结构,Mlpconv网络是Yan等人^[10]提出的一种增强的

非线性网络结构,ResNet^[11]是一种全新的网络结构,即深度残差网络.为了检验文本提出网络模型的性能,我们在相同的数据集上,对迭代次数为10万次的网络模型进行测试,每个模型测试10次取平均值,实验结果见表4.



(a) 检测精度变化曲线图



(b) 训练损失值变化曲线图

图5 MS-CNN在人脸润饰数据集LFW和ND-IIITD上的检测性能

表4 不同网络结构的检测精度

网络结构	数据集	检测精度(%)	模型大小(MB)
LeNet ^[7]	LFW	89.4	84.2
	ND-IIITD	83.0	
AlexNet ^[8]	LFW	94.2	114.3
	ND-IIITD	53.8	
Mlpconv ^[10]	LFW	90.6	52.6
	ND-IIITD	85.8	
ResNet-50 ^[11]	LFW	97.6	188.6
	ND-IIITD	90.4	
MS-CNN	LFW	99.5	23.5
	ND-IIITD	92.9	

从表4可以得出,在LFW人脸润饰数据集上,本文算法的检测精度高于LeNet网络10.1%,高于AlexNet网络5.3%,高于Mlpconv网络8.9%,高于ResNet-50网络1.9%;在ND-IIITD人脸润饰数据集上,本文算法的

检测精度高于 LeNet 网络 9.9%, 高于 AlexNet 网络 39.1%, 高于 Mlpconv 网络 7.1%, 高于 ResNet-50 网络 2.5%。

模型大小反应了网络结构在训练过程中参数量的大小, 模型越大神经网络需要学习的参数越多, 层之间传输的数据流越大, 更消耗计算资源。ResNet-50 的网络层数最深参数量最大, 因此模型最大, 而本文算法在保证检测精度较高的情况下参数量较小, 节约了计算资源。

综上, 本文提出的算法与以上 4 种网络结构相比检测精度较高、模型较小, 因此性能较好。

除此之外, 本文还与人脸润饰检测领域的最新算法文献[4]和文献[19]进行对比, 文献[4]属于传统手动设计特征的算法, 作者将图像的几何、光度变化和感知失真作为判别特征, 利用支持向量实现分类; 文献[19]属于深度学习类的算法, 作者利用有监督的深度玻尔兹曼机来提取图像特征, 支持向量实现分类。本文算法、文献[4]和文献[19]在 ND-IIITD 人脸润饰数据集上的实验结果见表 5。

表 5 检测精度对比

算法	检测精度 (%)
手动设计特征+SVM ^[4]	48.8
深度玻尔兹曼机+SVM ^[19]	87.1
提出的 MS-CNN	92.9

由表 5 可知本文算法的检测精度分别比文献[4]和文献[19]高 44.1% 和 5.8%, 并且文献[4]和文献[19]都将特征提取和图像分类分步进行, 而本文提出的 MS-CNN 实现了端到端的人脸润饰图像检测。

5 结论

由于人脸润饰检测的类间差别远小于计算机视觉, 传统 CNN 的检测精度不高, 针对该问题, 本文提出了一种多尺度卷积神经网络模型 MS-CNN 用于人脸润饰检测, 该算法不仅避免了手动设计提取特征的复杂, 而且实现了端到端人脸润饰检测。在图像预处理阶段, 本文算法利用基于 HOG 特征的人脸提取算法提取待测图像的人脸部分; 通过添加局部归一化层来加速模型收敛; 提出多尺度卷积层使模型能学习到更丰富的纹理特征。实验结果表明, 与现有的算法相比, MS-CNN 具有较高的检测精度。

今后的工作主要围绕增强 MS-CNN 对人脸润饰检测的通用性进行, 进一步优化网络结构并提高检测精度。

参考文献

- 1 Qazi T, Hayat K, Khan SU, *et al.* Survey on blind image forgery detection. IET Image Processing, 2013, 7(7): 660–670. [doi: 10.1049/iet-ipr.2012.0388]
- 2 Ananya U, Mudenagudi U. Detection of doctored images using bispectral analysis. Proceedings of 2011 IEEE International Conference on Image Information Processing. Shimla, India. 2011. 1–6.
- 3 李晓飞, 申铉京, 陈海鹏. 基于 Bi-Laplacian 润饰图像的盲鉴别算法. 计算机应用, 2011, 31(1): 239–242. [doi: 10.3969/j.issn.1000-386X.2011.01.072]
- 4 Kee E, Farid H. A perceptual metric for photo retouching. Proceedings of the National Academy of Sciences of the United States of America, 2011, 108(50): 19907–19912. [doi: 10.1073/pnas.1110747108]
- 5 Shah H, Shinde P, Kukreja J. Retouching detection and steganalysis. IJEIR, 2013, 2(6): 487–490.
- 6 Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science, 2006, 313(5786): 504–507. [doi: 10.1126/science.1127647]
- 7 Lecun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278–2324. [doi: 10.1109/5.726791]
- 8 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA. 2012. 1097–1105.
- 9 Hinton GE, Srivastava N, Krizhevsky A, *et al.* Improving neural networks by preventing co-adaptation of feature detectors. arXiv: 1207.0580.
- 10 Lin M, Chen Q, Yan SC. Network in network. International Conference on Learning Representations. Banff, Canada. 2014.
- 11 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2015. 770–778.
- 12 Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2013. 580–587.
- 13 Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640–651. [doi: 10.1109/TPAMI.2016.2572683]

- 14 Chen JS, Kang XG, Liu Y, *et al.* Median filtering forensics based on convolutional neural networks. *IEEE Signal Processing Letters*, 2015, 22(11): 1849–1853. [doi: [10.1109/LSP.2015.2438008](https://doi.org/10.1109/LSP.2015.2438008)]
- 15 Yang PP, Ni RR, Zhao Y. Recapture image forensics based on laplacian convolutional neural networks. *Proceedings of the 15th International Workshop on Digital Forensics and Watermarking*. Beijing, China. 2016. 119–128.
- 16 Bondi L, Baroffio L, Güera D, *et al.* First steps toward camera model identification with convolutional neural networks. *IEEE Signal Processing Letters*, 2017, 24(3): 259–263. [doi: [10.1109/LSP.2016.2641006](https://doi.org/10.1109/LSP.2016.2641006)]
- 17 Qian YL, Dong J, Wang W, *et al.* Deep learning for steganalysis via convolutional neural networks. *Proceedings of SPIE 9409, Media Watermarking, Security, and Forensics*. San Francisco, CA, USA. 2015. 94090J.
- 18 Rao Y, Ni JQ. A deep learning approach to detection of splicing and copy-move forgeries in images. *Proceedings of 2016 IEEE International Workshop on Information Forensics and Security (WIFS)*. Abu Dhabi, United Arab Emirates. 2017. 1–6.
- 19 Bharati A, Singh R, Vatsa M, *et al.* Detecting facial retouching using supervised deep learning. *IEEE Transactions on Information Forensics and Security*, 2016, 11(9): 1903–1913. [doi: [10.1109/TIFS.2016.2561898](https://doi.org/10.1109/TIFS.2016.2561898)]
- 20 Bharati A, Vatsa M, Singh R, *et al.* Demography-based facial retouching detection using subclass supervised sparse autoencoder. *Proceedings of 2017 IEEE International Joint Conference on Biometrics (IJCB)*. Denver, CO, USA. 2017. 474–482.
- 21 Dalal N, Triggs B. Histograms of oriented gradients for human detection. *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego, CA, USA. 2005. 886–893.
- 22 Jia YQ, Shelhamer E, Donahue J, *et al.* Caffe: Convolutional architecture for fast feature embedding. *Proceedings of the 22nd ACM international conference on Multimedia*. Orlando, FL, USA. 2014. 675–678.