

# 基于深度学习的人脸颜值评估系统<sup>①</sup>

王冰冰, 麦成源, 庄杰颖, 潘家辉, 梁 艳

(华南师范大学 软件学院, 佛山 528225)

通信作者: 梁 艳, E-mail: [liangyan@m.scnu.edu.cn](mailto:liangyan@m.scnu.edu.cn)



**摘 要:** 针对人脸颜值评估系统正确率和实时性低的问题, 提出了一种基于深度学习的人脸颜值评估系统. 该系统利用基于 HOG 特征的方法进行人脸检测, 采用 FaceNet 预训练模型提取人脸特征值, 提出基于 Softmax 分类层和 ReLU 回归层的双层决策模型, 并结合人脸局部特征量化值进行人脸颜值评估. 在 SCUT-FBP5500 数据集上进行实验, 结果表明该系统正确率为 78.58%, 单张图片的平均评估时间为 2.98 s, 能满足实际应用的需求.

**关键词:** 人脸颜值评估; FaceNet; 双层决策模型; 局部特征量化

引用格式: 王冰冰, 麦成源, 庄杰颖, 潘家辉, 梁艳. 基于深度学习的人脸颜值评估系统. 计算机系统应用, 2022, 31(1): 99-104. <http://www.c-s-a.org.cn/1003-3254/8291.html>

## Facial Attractiveness Evaluation System Based on Deep Learning

WANG Bing-Bing, MAI Cheng-Yuan, ZHUANG Jie-Ying, PAN Jia-Hui, LIANG Yan

(School of Software, South China Normal University, Foshan 528225, China)

**Abstract:** To address the problem of low accuracy and poor real-time performance of the facial attractiveness evaluation system, we propose a new facial attractiveness evaluation system based on deep learning. In this system, the HOG feature-based method and the FaceNet pre-training model are used for face detection and facial feature extraction respectively. Furthermore, a two-layer decision model based on the Softmax classification layer and ReLU regression layer is proposed, which is combined with the quantized values of local facial features to evaluate the facial attractiveness. Experimental results on the SCUT-FBP5500 dataset show that the accuracy of the system is 78.58%, and the average evaluation time of a single image is 2.98 seconds, which can meet the needs of practical applications.

**Key words:** facial attractiveness evaluation; FaceNet; two-layer decision model; local feature quantification

## 1 引言

### 1.1 研究背景

“颜值”是用来衡量外貌美丑状态的一个流行词汇, 体现了人们对外貌的重视. 在追求高颜值的今天, 如何快速、客观评价一个人的颜值是值得研究的课题. 随着计算机视觉的发展, 人脸识别技术和深度学习算法逐渐普及, 众多研究者尝试采用人工智能的方法对人脸颜值进行自动评估. 人脸颜值评估系统可为人脸颜值提供客观的评价, 具有一定的理论研究价值. 此外, 人脸颜值评估系统可应用于美妆产品、图片美化、社

交网络等场景, 具有极大的市场应用价值.

### 1.2 研究现状

近年来, 随着互联网技术的不断发展, 为了满足人们对美貌的追求, 许多人脸颜值评估系统应运而生. 国内推出了多款操作简便的颜值打分器 APP, 仅需上传图片即可预测出用户的颜值分数, 但评估准确率较低, 对于美丑的判定与大众审美存在较大的偏差. PrettyScale 是国外推出的一款专业在线颜值点评系统, 主要根据黄金比例的美学原理对用户的颜值进行评估, 但该系统操作较为复杂, 需要用户手动调整面部、眼睛、鼻

<sup>①</sup> 基金项目: 国家自然科学基金面上项目 (62076103); 广州市重点领域研发计划 (202007030005)

收稿时间: 2021-03-29; 修改时间: 2021-04-29; 采用时间: 2021-05-14; csa 在线出版时间: 2021-12-17

子、嘴巴等各部分的位置。

早期的人脸颜值评估研究主要采用传统的机器学习方法,比如,管业鹏等人<sup>[1]</sup>根据黄金比例的标准和人脸面部的1/3比例,提出了一种基于图像的非监督人脸颜值评估方法。李立琛<sup>[2]</sup>采用梯度方向直方图、几何特征、颜色直方图、尺度不变特征转换和Gist空间包络特征及其融合特征对人脸图像进行颜值评估,平均准确率达63%。毛慧芸等人<sup>[3]</sup>提出17维特征提取方法,并结合C4.5决策树学习方法,对510幅中国女性人脸图像进行颜值评估,分为4个评价等级,准确率达到71.6%。华南理工大学的谢多睿<sup>[4]</sup>设计了一款颜值达芬奇系统,主要采用基于几何特征与基于表现特征两种传统的机器学习方法进行颜值评估,同时结合了角度矫正和群众打分等功能,使得系统的测评结果更加合理。

传统的机器学习方法需要手工提取特征,泛化性能较弱,难以应对复杂的情况。因此,近年来较多学者采用基于深度学习的解决方案。相较于传统机器学习方法,深度学习通过对各层网络逐层的学习,更容易发现颜值评估的规律,具有广阔的应用空间。2015年,Xie等人<sup>[5]</sup>公开了一个用于人脸颜值评估的数据集,并设计3个不同网络层次的卷积神经网络(convolutional neural networks, CNN)模型进行测试,实验结果表明,采用CNN模型的分类效果远高于传统机器学习方法。甘俊英等人<sup>[6]</sup>则在此基础上进行了改进,通过减少网络层数和减小图像的尺寸以提高模型的实时性,并通过微调数据的方法进一步提高预测的准确性。Zhai等人<sup>[7]</sup>提出了一种继承不同尺度特征的颜值评估方法BeautyNet,在LSFBD数据集<sup>[8]</sup>上进行实验,准确率达67.48%。与传统机器学习方法相比,基于深度学习的方法在评估准确率上有所提高,但训练更为复杂的深度学习模型,需要大量的数据。

为了实现人脸颜值的智能评估,本文研发了一款基于深度学习的颜值评估系统,利用图像的方向梯度直方图(histogram of oriented gradient, HOG)特征<sup>[9]</sup>进行人脸检测,采用FaceNet预训练模型进行人脸特征值的提取,提出基于Softmax分类层和ReLU回归层的双层决策模型,并综合考虑人脸局部特征进行人脸颜值评估,不仅提高了颜值评估的准确率,还具有较高的实时性,满足实际应用的需求。

## 2 基于深度学习的人脸颜值评估系统

### 2.1 系统描述

本文设计了一款基于深度学习的人脸颜值评估系统,框架图如图1所示,结合双层决策模型和综合局部特征量化值,实现人脸颜值的实时评估,一定程度上避免了人工干预等主观性较强的评价。系统主要分为四个模块:人脸检测模块、人脸特征提取模块、人脸颜值评估模块、局部特征值量化模块。整体设计的思路如下:用户通过拍照或者本地相册输入需要评估的人脸图片,利用基于HOG特征的人脸检测器进行人脸检测,然后利用FaceNet预训练模型提取人脸特征向量,并将其作为双层决策模型的输入,采用双层决策模型进行训练,最终将人脸局部特征进行量化,并结合量化后的特征值得到颜值评估得分结果。

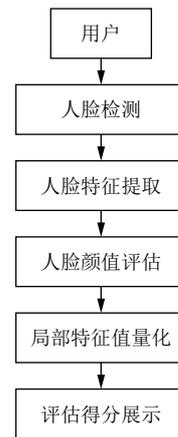


图1 人脸颜值评估系统框架图

### 2.2 人脸检测处理流程

用户通过拍照或选择照片的方式将图片上传至本系统,系统将会调用Dlib中的基于HOG特征<sup>[10]</sup>的分类器对获取到的照片进行人脸检测。人脸检测的流程如下:首先读取包含人脸矩形框阈值、滑动窗口大小和扫描滤波器参数的文件;其次选择尺度参数,利用滑动窗口对图像的不同尺度进行活动扫描,并计算出每个窗口的HOG特征;再根据文件中给出的滤波器参数计算每个窗口的响应值,对图像进行初步的判断,若检测的结果为人脸,则对其进行标记。经过一轮活动扫描后,图像的每个尺度会出现多次标记的情况。针对较高重叠率的人脸,我们在图像扫描完成后,利用非极大值抑制(non maximum suppression, NMS)方法<sup>[11]</sup>,根据重叠率的阈值和计算得到的滤波响应值来消除重叠多余

且响应值低的人脸,从而得到最终的人脸检测区域.人脸检测的流程图如图2所示.

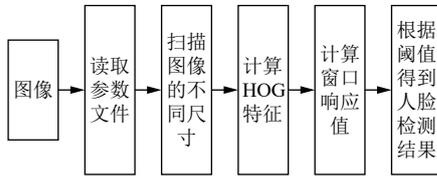


图2 人脸检测流程图

### 2.3 基于 FaceNet 预训练模型的人脸特征提取方法

FaceNet 预训练模型能够有效提取图片中具有复杂特征或难以直接量化的特征,并且提取的全局特征向量不受图片几何随机变换的影响,给接下来的双层决策模型带来较好的效果. FaceNet 框架图如图3所示.其中, FaceNet 主体模型采用一个由3个带有残差连接的 Inception 模块和1个 Inception v4 模块组成极深度网络 Inception ResNet-v2. Batch 为输入的人脸图像样本的批尺寸; L2 为特征归一化; Embeddings 则是经过极深度网络和特征归一化后生成的特征向量; Triplet Loss 代表的是三元损失函数,即通过学习使得类别内部的样本距离小于不同类别样本的距离.

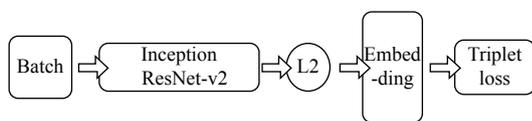


图3 FaceNet 预训练模型流程图

三元损失函数的定义如下:

$$T = \sum_i^N \left[ \|F(x_i^a) - F(x_i^m)\|_2^2 - \|F(x_i^a) - F(x_i^n)\|_2^2 + \omega \right]_+ \quad (1)$$

其中,  $x_i^a$ 、 $x_i^m$ 、 $x_i^n$  分别为第  $i$  个样本的正面人脸图(锚点)、与样本同类的人脸图(正样本)以及与样本不同类的人脸图(负样本),  $F(x_i^a)$ 、 $F(x_i^m)$ 、 $F(x_i^n)$  为3幅人脸图像对应的特征,  $N$  为样本个数,  $\omega$  是为平衡收敛性与精确度设置的超参数.在训练过程中,我们参考文献[12]将  $\omega$  设置为 0.2.

### 2.4 人脸颜值评估模块的原理与实现

人脸颜值评估模块的目的是对人脸图像进行颜值测评.若直接利用分类模型进行等级的分类,虽然可以得到较高的准确率,但评估的结果相对变得模糊,难以合理地区分每个人的颜值评估结果.另一方面,若直接

根据原始数据集中的连续性分数作为标签,训练一个回归型的模型,则会因为数据分布不均,导致均方根误差 (root mean square error, RMSE)、均方误差 (mean square error, MSE) 等相关指标较大,评估的结果准确性不高.

因此,本文将 FaceNet 模型中提取的人脸特征作为人脸颜值评估模块的输入,并将数据集分成“较低”“一般”“较高”3种等级,分别训练3种回归模型:较低颜值回归模型、一般颜值回归模型和较高颜值回归模型.此外,采用 ReLU 激活函数增加回归模型的非线性表达能力,通过梯度下降法对参数进行优化.

在人脸颜值评估模块的具体实现中,我们结合分类模型和回归模型,采用基于 Softmax 分类层和 ReLU 回归层的双层决策模型评估人脸颜值.首先,将 FaceNet 预训练模型提取的人脸特征向量作为双层决策模型的输入,通过双层决策模型中的 Softmax 分类层预测该人脸特征所处的颜值等级,然后通过对对应等级的回归模型(较低颜值回归模型、一般颜值回归模型和较高颜值回归模型),得到连续的人脸颜值评估数值,最后将该数值转换成百分制,即可获得百分制形式表示的颜值评估结果.具体流程如图4所示.

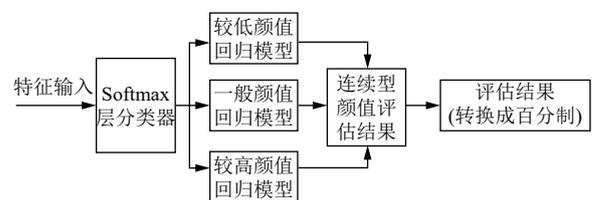


图4 双层决策模型流程图

### 2.5 局部特征值量化

考虑到单一评价会影响评估的客观性,本文根据文献[13]提出的东方人脸颜值测评标准,引入了5个主要局部特征值,分别为:下颞角度、眉心至鼻下/鼻下至下颌、人脸长宽比例、微笑角度、唇厚度比(下/上).传统审美观将人脸圆润对称视为美的标准,脸型不应过分尖瘦,同时,人脸长宽比应符合黄金比例,因此,我们通过分析数据集“较高”等级的人脸局部特征,并结合传统审美观,最终确定了5个局部特征的标准参考值:下颞角度为  $120^\circ$ ,眉心至鼻下与鼻下至下颌的比例为 1:1,人脸长宽比例为 1:0.618,微笑角度为小于  $180^\circ$ ,下唇与上唇厚度比值大于 1.具体每个局部特征的标准参考值如表1所示.

表1 标准局部特征值表

标准参考值	局部特征描述
120°	下颚角度
1/1	眉心至鼻下/鼻下至下颌
1/0.618	人脸长宽比例
<180°	微笑角度
>1	唇厚度比(下/上)

根据人脸图像各个局部特征值与对应的标准参考值之间的差异性,我们设置了相应的加分规则,如表2所示.规则设置的原则是:当局部特征值与标准参考值越接近,说明该特征越接近传统的大众审美,因此可以获得加分,反之,不得分.比如,下颚角度越接近120°,加分越多,当与标准参考值的差值超过30°,即脸型太宽或太尖,则不加分.

表2 局部量化特征加分规则表

局部特征	规则	加分值
下颚角度(°)	$ X-120  \leq 20$	2
	$20 <  X-120  \leq 30$	1
	$ X-120  > 30$	0
眉心至鼻下/鼻下至下颌	$ X-1  \leq 0.2$	2
	$0.2 <  X-1  \leq 0.3$	1
	$ X-1  > 0.3$	0
人脸长宽比例	$1.418 \leq X \leq 1.818$	2
	$1.218 \leq X < 1.418$ & $1.818 < X \leq 2.018$	1
	其他	0
微笑角度(°)	$180 - X \geq 10$	1
	$180 - X \leq 10$	0
唇厚度比(下/上)	$X-1 > 0$	1
	$X \leq 0$	0

### 3 实验和测试

#### 3.1 系统架构和实验环境

我们基于 Windows 系统和 Ubuntu 虚拟机搭建了系统的开发环境,主要通过 PyCharm 开发工具、Android Studio 开发工具、MySQL 数据库组成,如图5所示.系统的功能主要有 人脸检测、基于 FaceNet 模型的特征提取、颜值评估 3 大模块构成.在业务层中,采用基于 Softmax 分类层和 ReLU 回归层的双层决策模型、MySQL 数据库,同时使用第三方资源 Dlib 和 OpenCV 开源计算机视觉库.

#### 3.2 实验过程

##### 3.2.1 数据集收集及均衡化

###### (1) 数据集的来源与收集

本文采用 SCUT-FBP5500 数据集<sup>[5]</sup>,其中包含 5 500

张正面人脸图像,年龄分布在 15~60 岁.人脸图像包括了 2 000 名亚洲女性、2 000 名亚洲男性、750 名加索男性和 750 名高加索女性.本系统主要针对亚洲人脸颜值进行评估,因此我们选取了数据集中 4 000 张亚洲人脸图像用于模型训练和算法测试.每张图像分别由 60 个 18~27 岁的评分员进行评分,共 5 个等级,并将评分的均值作为数据集每张图片的标签. SCUT-FBP5500 数据集的部分原始数据如图6所示.

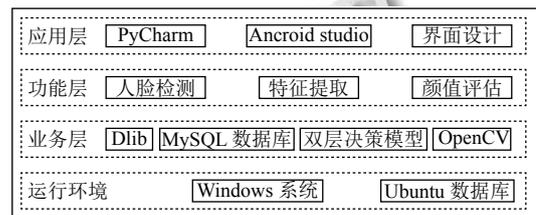


图5 系统架构图



图6 数据集的部分原始数据

###### (2) 数据集的扩充与处理

我们发现 SCUT-FBP5500 数据集在评分两端(接近 1 分和 5 分)的数据较少,而在中间分数段的数据居多.虽然该数据集符合正常的人脸颜值分布,但数据集的分布不均将会影响模型的训练效果,因此需要进一步收集数据并对数据集进行扩充与处理.

本文参考了英国 IM 杂志亚洲最时尚面孔名单,扩充数据集,使高分数据增多,并去除掉质量较差的图片,增强后的数据集共 4 794 张,然后将其按 7:3 的比例划分为训练集和测试集进行训练.增强的部分原始数据如图7所示.

由于在实际生活中,存在于评分两端分数段的数据较少,为使数据集分布均匀,我们调整分数段,合并

相关分类,将5个等级划分为3个等级:设定1-2分数区间为类别“较低”,3分数区间为类别“一般”,4-5分数区间为类别“较高”。



图7 数据集增强后的部分原始数据

### 3.2.2 模型训练过程和主要参数设置

本研究对基准的分类层模型进行结构上的改进,同时利用数据增强提高模型评估的性能.训练过程主要分为4步.

(1) 利用基于HOG特征的人脸检测器检测人脸区域,截取数据集图像中人脸的感兴趣区域(region of interest, ROI),从而避免了模型参数的冗余.

(2) 将训练集输入到FaceNet预训练模型中后,初步提取人脸的图像特征,得到人脸特征集.

(3) 在双层决策模型训练中,首先利用人脸特征集对3分类的Softmax分类器进行训练,然后将人脸特征集的5个等级划分为3个等级,分别训练对应等级的回归模型.

(4) 最后利用测试集计算双层决策模型的准确率.在双层决策模型测试中,将测试集输入到FaceNet预训练模型以提取人脸特征值,然后通过Softmax分类层预测出等级,并选择对应的回归模型进行颜值的评估.本文将双层决策模型最后的输出值进行四舍五入处理,综合原始的标签得到最终的准确率.Softmax分类器训练过程准确率如图8所示,当总迭代次数达到300次时,训练集准确率趋于收敛.

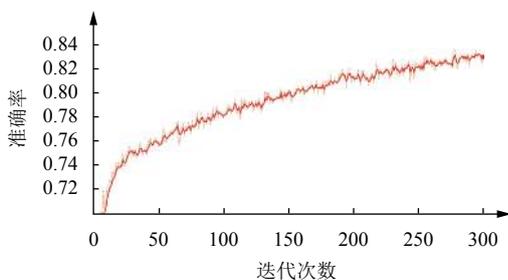


图8 双层决策模型中的分类层训练集准确率

主要参数设置如下:

- (1) 学习率为0.001;
- (2) 总迭代次数为300次;
- (3) 一次训练所选取的样本数(batch size)为32.

### 3.2.3 实验结果与分析

#### (1) 数据集处理结果与分析

为了确定数据增强的方法,解决数据集分布不均的问题,我们对以下3种数据增强方法进行对比:

① 变换增强处理:对评分两端的图片进行随机图像变换与增强;

② DCGAN模型<sup>[14]</sup>处理:采用DCGAN模型对评分两端的图片训练,生成类似图片.

③ 人工处理:人工扩充数据集,使高分数据增多,并去掉质量较差的图片.

我们基于不同的数据集处理方法,运用本文所提出的FaceNet+双层决策模型+局部特征量化值进行训练,结果如表3所示.

表3 基于不同数据集处理方法的模型训练结果

数据集处理方法	训练样本数	测试样本数	总训练时间(小时)	正确率(%)
原始数据	2 800	1 200	5.17	74.06
变换增强处理	3 850	1 650	5.91	72.95
DCGAN模型处理	3 850	1 650	6.20	74.72
人工处理	3 355	1 439	5.86	78.58

从表3可以看到,采用DCGAN模型处理和人工处理的方法对原始数据进行增强,均可以提高评估的准确率.但是,采用DCGAN模型训练生成图片的时间成本代价较高,总训练时间为6.20小时,而且生成图片的质量不高.与另外两种数据增强方法相比,采用人工处理的方法不仅总训练时间最短,而且准确率最高,达78.58%,因此本文最终确定采用人工处理方法进行数据增强.

#### (2) 颜值评估方法比较与分析

为了验证本文方法的有效性,我们把本文方法与FaceNet+Softmax、FaceNet+双层决策模型方法的性能进行比较,实验结果如表4所示.采用正确率为性能评价指标,其计算公式如下:

$$ACC = \frac{N_{true}}{N_{predict}} \quad (2)$$

其中,ACC为正确率, $N_{true}$ 为正确分类的图片数量, $N_{predict}$ 为总预测的人脸图片数量.

表4 不同颜值评估方法结果对比 (%)

检测方法	正确率
FaceNet+Softmax	73.04
FaceNet+双层决策模型	75.45
本文方法	78.58

与另外两种颜值评估方法相比,本文方法经数据扩充后,利用FaceNet+双层决策模型+局部特征量化值进行人脸颜值评估,正确率比采用FaceNet+Softmax和FaceNet+双层决策模型的方法分别提高了5.54%和3.07%,证明了双层决策模型与局部特征量化值的结合能有效地提高分类性能,最终提高系统的整体性能。

### (3) 系统响应时间分析

我们对系统的响应时间进行了测试,输入20张人脸图片进行颜值评估,共花费59.6 s,平均每张图片的评估速度为2.98 s,处理单张图片的最长时间为3.23 s,最短时间为2.16 s,其变化幅度也是在用户可接受的范围之内,因此,本文提出的方法能够满足用户实际应用需求。

## 4 总结

本文将人工智能技术应用于美妆领域,研发了一款基于深度学习的人脸颜值评估系统,可实时对用户的妆容进行评估,提高用户的美妆技能和质量。该系统利用FaceNet预训练模型进行人脸特征提取,结合双层决策模型和人脸局部特征量化值进行人脸颜值评估。在经过数据增强后的SCUT-FBP5500数据集进行实验,本系统获得78.58%的准确率,人脸图片的平均评估时间为2.98 s,能满足实际应用的需求。但影响人脸颜值的因素诸多,如肤色、发型、情感状态等。如何进一步提取综合性更强的特征,实现更贴近大众审美标准的人脸颜值评估系统,是我们未来要继续研究的问题。

### 参考文献

- 管业鹏,夏慧明,时勇杰.非监督人脸美貌度评价.电子器件,2012,35(2):244-248.[doi:10.3969/j.issn.1005-9490.2012.02.029]
- 李立琛.基于深度自学习的人脸美丽预测研究[硕士学位论文].江门:五邑大学,2014.
- 毛慧芸,金连文,杜明辉.基于几何特征及C4.5的人脸美丽分类方法.模式识别与人工智能,2010,23(6):809-814.[doi:10.3969/j.issn.1003-6059.2010.06.010]
- 谢多睿.人脸美丽的特征分析及其云系统设计[硕士学位论文].广州:华南理工大学,2017.
- Xie DR, Liang LY, Jin LW, et al. SCUT-FBP: A benchmark dataset for facial beauty perception. Proceedings of 2015 IEEE International Conference on Systems, Man, and Cybernetics. Hong Kong: IEEE, 2015. 1821-1826. [doi: 10.1109/SMC.2015.319]
- 甘俊英,黄聿,翟懿奎,等.基于轻量级卷积神经网络的人脸美丽预测.五邑大学学报(自然科学版),2017,31(3):42-48.
- Zhai YK, Cao H, Deng WB, et al. BeautyNet: Joint multiscale CNN and transfer learning method for unconstrained facial beauty prediction. Computational Intelligence and Neuroscience, 2019, 2019: 1910624. [doi: 10.1155/2019/1910624]
- Zhai YK, Huang Y, Xu Y, et al. Benchmark of a large scale database for facial beauty prediction. Proceedings of 2016 International Conference on Intelligent Information Processing. Wuhan: ACM, 2016. 21. [doi: 10.1145/3028842.3028863]
- Dalal N, Triggs B. Histograms of oriented gradients for human detection. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego: IEEE, 2005. 886-893. [doi: 10.1109/CVPR.2005.177]
- 刘梦佳,赵建国.基于SVM的疲劳驾驶人脸图像特征检测算法.半导体光电,2020,41(2):278-282.
- Neubeck A, van Gool L. Efficient non-maximum suppression. Proceedings of the 18th International Conference on Pattern Recognition. Hong Kong: IEEE, 2006. 850-855. [doi: 10.1109/ICPR.2006.479]
- Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 815-823. [doi: 10.1109/CVPR.2015.7298682]
- 韩晓旭.东方女性人脸颜值量化算法研究[硕士学位论文].西安:西北大学,2017.
- Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434, 2015.