

基于 DCT 的模糊面部超分辨率性别识别模型研究

阎卫东¹, 韩璐², 张野¹, 刘国奇²

(1. 沈阳建筑大学土木工程学院, 辽宁 沈阳 110168;
2. 沈阳建筑大学计算机科学与工程学院, 辽宁 沈阳 110168)

摘要:为解决智能监控领域中存在的基于人脸图像的性别识别准确率不高的问题,本研究构建了 DCT-FSR-CNN 模型。该模型以相邻两帧面部图像作为超分辨率的双输入,通过离散余弦变换和局部超分辨率技术对图像进行融合处理以提高图像整体分辨率;对 Alexnet 网络进行优化改进,通过融合网络中间层进行特征值优化以提高分类器的准确率。同时,将该模型与其他 4 种性别识别模型进行实验对比,在现实的模糊面部图像数据集上的实验结果准确率和运行速度均有明显提升,证明了该模型的应用性和有效性。

关键词:图像融合;超分辨率;卷积神经网络;性别识别;计算机视觉

中图分类号:TP183

文献标志码:A

性别作为人类基本属性之一,在个体识别和社会交往中起着极其重要的作用。性别认知是人类普遍具有的能力,如果计算机具有能够识别性别的能力,不仅可以在许多项目中发挥非常大的作用,还可以提高其他相关应用程序的性能。

通过面部外观准确地识别出性别是人脸识别中的热点研究方向。李同宇等^[1]提出一种基于特征融合的人脸图像性别识别方法。Levi G 等^[2]通过实验表明,与其他方法相比,通过使用深度卷积神经网络可以显著提高年龄估计和性别识别的性能。Lin G S 等^[3]提出了一种基于多区域特征提取和信息融合的性别分类方案,以实现无约束图像的性别分类。Duan M 等^[4]提出了一种年龄和性别分类的混合深度学习模型,通过集

成用于从输入图像中提取特征的卷积神经网络和极限学习机进行识别与分类。Dagher I 等^[5]通过使用聚类和增量学习来提高支持向量机的性别分类精度。Lin C J 等^[6]通过基于双输入卷积神经网络的特征融合算法和参数优化方法来进行人脸性别识别。陈文兵等^[7]通过优化卷积层提取的特征和使用误差最小化极限学习机作为分类器以实现面部图像性别识别。Ranjan R 等^[8]将单独的卷积神经网络与深度卷积神经网络的中间层相融合,使用多任务学习算法对融合的特征进行操作,利用任务之间的协同作用提高整体性能。余婵娟^[9]等针对人脸性别出现的特征冗余问题提出一种基于稀疏表示的特征选择方法。

虽然基于人脸的性别识别已是一个较为成熟的技术,但现如今仍然有大量的老旧设

备,从设备中提取出的图像模糊,导致性别识别的结果准确率低,故提高模糊图像的性别识别准确率成为城市安全智能监控领域中非常重要的一个课题。为解决上述问题,研究提出一种 DCT-FSR-CNN 混合结构模型,其模块流程如图 1 所示。该模型通过使用基于离散余弦变换(Discrete Cosine Transform, DCT)的图片融合技术和超分辨率技术对图像进行数据增强,通过提取监控视频中的相邻帧图像,在图像块融合后,进行超分辨率处理,最后通过优化后分类器进行性别识别任务。在现实数据集上的实验结果表明,该模型既提高了基于人脸图像的性别识别精度,也有效地提升了计算速度。

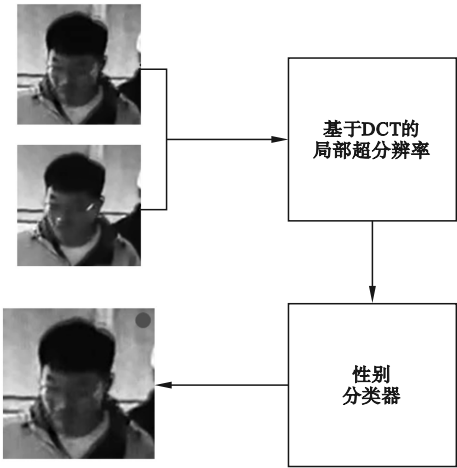


图1 DCT-FSR-CNN 模型流程

一、基于 DCT 的局部超分辨率

图像融合是对两张及以上图像中的表征信息进行提取和选择,然后将这些表征信息组合成一张完整的图像,保留其必要或有效

$$F(u,v) = c(u)c(v) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j) \cos\left[\frac{(2i+1)\pi u}{2N}\right] \cos\left[\frac{(2j+1)\pi v}{2N}\right] \quad (1)$$
$$c(u) = \begin{cases} \sqrt{\frac{1}{N}}, & u=0 \\ \sqrt{\frac{2}{N}}, & u \neq 0 \end{cases} \quad (2)$$

式中: u 和 v 为频域的坐标; $c(u)$ 为变换系数; $F(u,v)$ 为变化后的频域系数; i 和 j 为原始图像中像素点的坐标; $f(i,j)$ 为原始图像中像素点 (x,y) 的灰度像素值; N 为图像块

的信息,同理可得 $c(v)$ 。DCT-FSR-CNN 模型把图像融合与超分辨率技术结合起来,将双输入的低分辨率图像划分成规格为 8×8 的正方形图像块,并对两幅图像中相同位置块进行基于离散余弦变化计算的对比选择,选择出高质量图像块。对高质量图像块使用基于块的超分辨率方法,相比于进行图像全局超分辨率处理,基于块的超分辨率方法可以提升模型运行效率。基于 DCT 的局部超分辨率算法步骤如下:

1. 基于 DCT 的最佳质量块选择

记相邻两帧源图像分别为 P_A 和 P_B ,对图像的尺寸规整化设定为 $P \times Q$ 。将输入图像均分割成 k 个 8×8 的图像块,图像 P_A 的第 n 块记为 B_n^A ,图像 P_B 的第 n 块记为 B_n^B ,其中 $n=1,2,\dots,k$ 。图像块 B_n 的 DCT 输出即可表示为集合 $D=D_0,D_1,\dots,D_{k-1}$,计算公式如式(1)所示。该模块流程如图 2 所示,具体算法步骤如下。

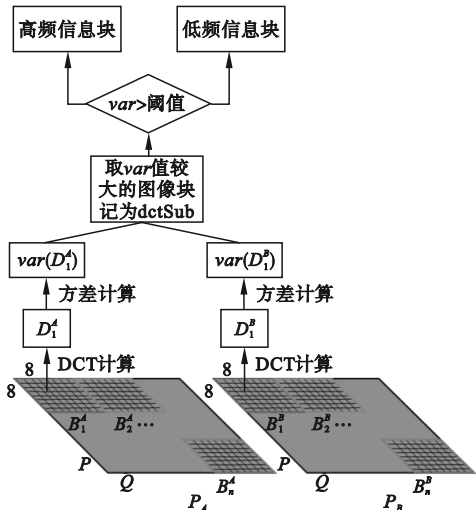


图2 DCT 高质量块选择流程

- 尺寸。
- (1) 对 $D_n^R (n=1,2,\dots,k; R=A,B)$ 通过归一化、计算均值进行方差求值得到 $var(D_n^R)$ 。
 - (2) 将得到的块 $var(D_n^A)$ 和 $var(D_n^B)$ 进行值大小对比选择值更大,记为 $dctSub$, $dctSub$ 即为通过离散余弦变换计算和比较得到的高质量块。
 - (3) 由于对图像块像素值进行过 DCT

计算处理,所以需要对图像块进行反余弦变换,即二维离散反余弦变换计算 (Inverse

$$f(i,j) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} c(u)c(v)F(u,v)\cos\left[\frac{(2i+1)\pi u}{2N}\right]\cos\left[\frac{(2j+1)\pi v}{2N}\right] \quad (3)$$

式中: $F(u,v)$ 为 DCT 变换后的系数; $c(u)$ 和 $c(v)$ 同式(2)。

(4)最终得到的 $\text{idct}(\text{dctSub})$ 为 8×8 的图像块就是选择出的高质量块。为了避免超分辨率处理包含信息少的图像块导致整体的处理时间增加,将图像块分为高频信息块和低频信息块,直接对高频信息图像块进行操作。通过计算每个图像块的方差值来分析图像块的内容,如果图像块的方差值大于预定

Discrete Cosine Transform, IDCT), 得到 $\text{idct}(\text{dctSub})$ 。IDCT 公式如式(3)所示。

义值,则对该图像块进行超分辨率处理,如果该图像块的方差值小于预定义值,则将图像块进行上采样处理。

2. 局部超分辨率技术

超分辨率技术的核心是学习低分辨率图像 X 和高分辨率图像 Y 之间端到端的映射函数关系,并通过图像 X 重构出图像 Y。

采用 FSRCNN 模型对高频信息图像块进行超分辨率操作,模型流程如图 3 所示。

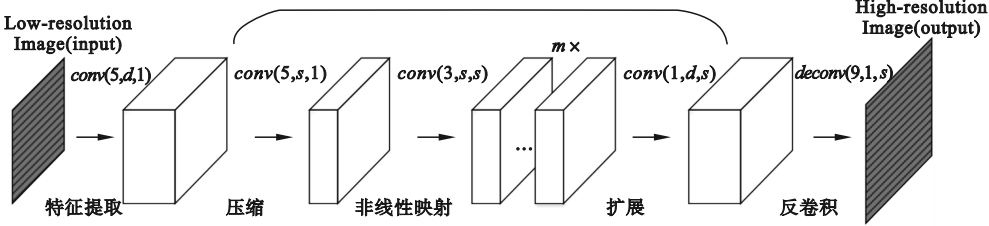


图 3 FSRCNN 流程

(1)特征提取。该层对选择的高频信息图像块进行特征提取。使用一组参数表示为 $\text{conv} = (f_1, n_1, c_1)$ 的滤波器进行卷积操作。当 $f_1 = 5$ 的条件下图像可以得到较好的效果和较小的信息损失, c_1 预设为 1, 将 n_1 设为 LR 图像的特征图维度 d , 故可以表示为 $\text{conv} = (5, d, 1)$ 。

(2)压缩。该层对特征图进行压缩。通过增加一个收缩层以降低 LR 图像维度 d , 实现降低计算成本。将滤波器大小设置为 $f_2 = 1$, 通过更少的滤波器数量 $n_2 = s, (s \ll d)$, 将特征维度从 d 降到 s 。故可以表示为 $\text{conv} = (1, s, d)$ 。

(3)非线性映射。该层是超分辨率技术中核心的部分,其最重要的两个参数即为特征图的通道数 d 和卷积层层数 m 。该层采用中等大小的滤波器 $f_3 = 3$, 使用 m 个 3×3 层来代替原单个宽层。故第三层可以表示为 $\text{conv} = (3, s, s)$ 。

(4)扩展。该层是收缩层的逆过程。使用低维特征生直接成 HR 图像会导致重建图像质量低,因此需要对特征图进行一个维度

扩展的操作。非线性映射层之后设置一个扩展层实现维度扩展。为与收缩层保持一致,卷积核的数量上选取与压缩层数量的个数,卷积核采用尺寸为 1×1 的滤波器。该层可以表示为 $\text{conv} = (1, d, s)$ 。

(5)反卷积层。该层进行卷积的逆运算,输出重建的 HR 高频信息图像块。在超分辨率技术的最后增加一个反卷积层,用于对特征执行图像尺寸放大和聚合的操作。若滤波器以 $\text{stride} = m$ 进行图像卷积,则得到 $1/m$ 倍的输出;反之,若将 stride 设置为 $1/m$, 则得到 m 倍输出。

3. 输出图像

通过图像融合将图像块融合进其在图像中对应的位置,并使用一致性检测得到最终高质量图像。基于 DCT 的局部超分辨率流程如图 4 所示。

二、卷积神经网络性别分类器

卷积神经网络被广泛应用于计算机智能识别领域中^[10],随着卷积神经网络层数的增加所适用的任务更加复杂,为了优化特征提

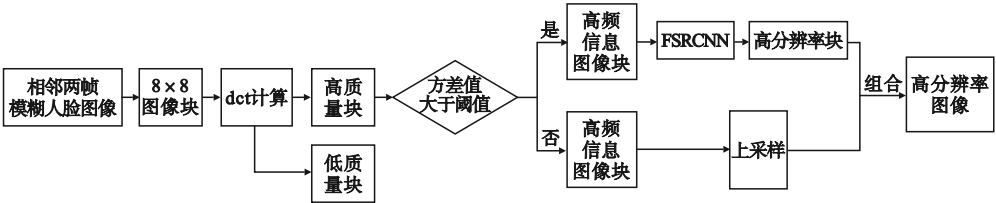


图 4 基于 DCT 的局部超分辨率流程

取,通过将卷积神经网络中间层特征图进行融合,构建一个单一的卷积神经网络模型用于性别识别任务。对 AlexNet^[11] 进行特征层融合改进作为分类器进行目标任务,由于卷积神经网络的特征在网络中是分层分布的,深浅不同的层数输出的特征值用于执行不同类别的任务。较低层特征所表达的信息较细致,较高层特征适合于更复杂的任务^[12]。因此,为了提高性别识别任务的性能,应通过融合部分网络中间层的特征代替融合所有中间层^[13]。

以尺寸为 224×224 的图片作为该网络模型的输入。模型由 5 个卷积层和 3 个全连接层构成,在每个卷积层后都会连接一个 ReLU 激活函数层对神经元进行非线性激活,池化层也均采用 3×3 的滤波器以步长为 2 进行最大池化操作。

第一个卷积层使用 96 个滤波器对图像以步长为 4 进行卷积,随后接入 LRN 层;在与下一个卷积层进行连接前,先连接一个池化层对特征图进行缩小,输出特征图大尺寸为 $27 \times 27 \times 96$ 。第二个卷积层需要对像素进行全局扫描,因此将步长参数设置为 1,卷积核采用尺寸为 5×5 的滤波器对特征图进行卷积,同样在接入下一个卷积层前通过 LRN 层和最大池化层对特征图进行处理。第三个卷积层使用 384 个滤波器对输入特征进行卷积,输出特征值为 $13 \times 13 \times 384$ 。第四五个卷积层是相连接的,均使用尺寸为 3×3 的滤波器以步长为 1 对特征图进行卷积,在第五层时卷积核个数由 384 减少到 256。在经过第五层卷积层以后连接一个最大池化层,输出 $6 \times 6 \times 256$ 的特征值。至此该模型中卷积模块结束,虽然在前几个卷积层中有很大的计算量,但是由于卷积神经网络中卷积层共享权重的特性,所以参数量很

小,都在 1M 左右,而全连接层部分的参数量占比全局最高。在全连接层前增加了两个带有 dropout 的网络层,在 9 层和 10 层分均通过 4 096 个神经元对特征图进行连接,并随机失活丢掉一些节点,最后使用 ReLU 函数进行非线性激活。

使用 AlexNet 作为基础模型进行中间层特征融合,挑选其网络的 pool1、conv3 和 pool3 层进行融合。由于不同层特征图具有不同尺寸,分别为 $27 \times 27 \times 96$, $13 \times 13 \times 384$ 和 $6 \times 6 \times 256$,所以不能通过直接将特征连接起来的方式进行融合,需要对特征图进行维度统一。利用 conva 和 convb 子卷积层对 pool1、conv3 和 pool3 进行维度统一,将特征图尺寸统一为 $6 \times 6 \times 256$ 。该模型中卷积神经网络分类器结构如图 5 所示。

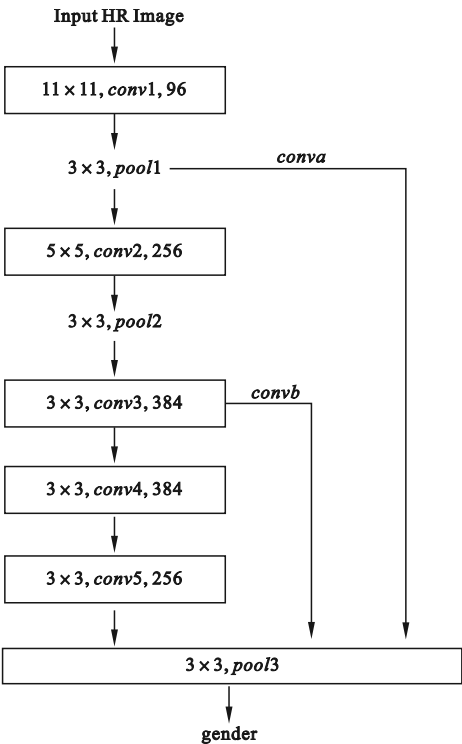


图 5 分类器结构

在训练过程中使用 R—CNN Face 网络的权重初始化该网络参数,用 *softmax* 层对特征值进行分类,性别识别所用到的损失函数如公式(4)所示。

$$loss_G = -(1 - g) \times \log(1 - p_g) - g \times \log(p_g)$$

(4)

式中: g 为性别,值为 1 代表女性,值为 0 代表男性; G 为损失函数标识; p_g 为计算结果的二维概率向量。

三、实验结果与分析

1. 训练与测试数据集

通过构建并使用现实图片数据集进行实验以评估验证该模型。首先通过在低分辨率监控视频上无差别截取视频片段,由于监控视频相邻帧存在大量重叠和冗余,而视频三帧之间目标物体位移较大,故选取相邻两帧进行图片融合效果最佳,每段视频随机抽取相邻两帧为一组,图片大小为 196×196 ,格式为 *JPG*,该实验数据集中含有 700 张现实图片,其中 400 张男性,300 张女性。将数据集划分为 10 个子集,每个子集中男女比例均为 4:3,随机抽取 9 个数据集作为训练集,剩余 1 个为测试集。提出的 DCT - FSR - CNN 模型依赖于 8GB 显存 AMD R5 处理器。数据集部分展示如图 6 所示。



图 6 现实人脸数据集

2. 基于 DCT 的超分辨率实验

通过 *opencv* 提取视频相邻帧,将相邻帧通过 *opencv* 中的 *face alignment* 进行面部对齐,对输入数据进行检验,查看图像尺寸是否一致,若不一致则统一图片格式为 $196 \times$

196,同时检验图像输入个数是否为 2,输入图像是否为灰度图,若不是则对输入图像进行灰度处理。将预处理后的两张面部图像作为基于 DCT 的局部超分辨率模块输入。

对 10 个子集进行实验对比,将各子集平均峰值信噪比作为实验对比参数,对比结果如图 7 所示,实验结果表明经过处理的图像峰值信噪比 (PSNR) 得到明显提升。

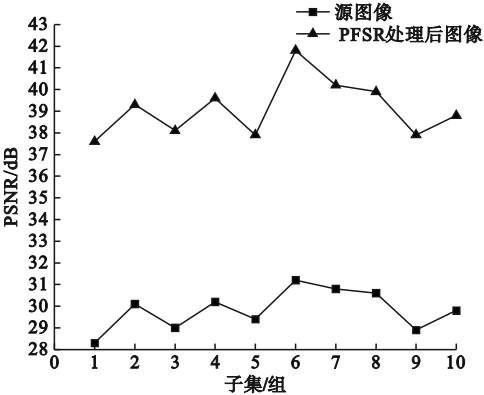


图 7 超分辨率前后 PSNR 值对比

3. 模型实验

为了进行模型评估,将本文所提出的 DCT-FSR-CNN 模型与 4 种模型进行对比实验,模型分别为 RCNN-Gender、CNN-ELM、CNN-SE-ELM 和 LNetS-ANet。对比过程中使用相同数据集进行训练与测试,具体对比结果如表 1 所示。

在该模型的训练和测试过程中,性别识别结果为男性的男性样本记为 TP (True Positive);性别识别结果为女性的女性样本记为 TN (True Negative);性别识别结果为男性的女性样本记为 FP (False Positive);性别识别结果为女性的男性样本记为 FN (False Negative)。具体计算公式如下。

准确率公式:

$$acc = \frac{TP + TN}{TP + TN + FP + FN}$$

(5)

召回率公式:

$$recall = \frac{TP}{TP + FN}$$

(6)

整体评价指标 F_1 - Score 公式:

$$F_1 - Score = 2 \times \frac{pre \times recall}{pre + recall}$$

(7)

式中: acc 为准确率; $recall$ 为召回率; pre 为查准率。

表 1 模型对比结果

模型	acc	recall	F_1 —Score
RCNN-Gender	0.76	0.8	0.78
CNN-ELM	0.81	0.83	0.82
CNN-SE-ELM	0.83	0.82	0.82
LNets-ANet	0.85	0.86	0.85
DCT-FSR-CNN	0.89	0.88	0.88

实验证明:DCT-FSR-CNN 性别识别模型在现实数据集测试中取得的效果表现最优,准确率为 0.89。实验图像中蓝色圆点表示识别结果为男性,粉色圆点识别结果女性,实验结果如图 8 所示。



(a)DCT-FSR-CNN模型实验结果(男)



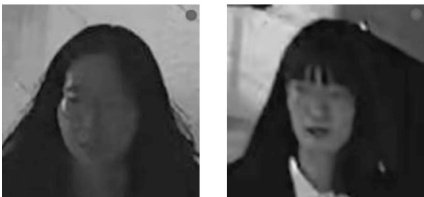
(b)DCT-FSR-CNN模型实验结果(女)



(c)其他模型分类错误(男)



(d)本研究模型分类正确(男)



(e)其他模型分类错误(女)



(f)本文模型分类正确(女)

图 8 模型实验结果

四、结 语

针对图像分辨率低的问题,提出了 DCT-FSR-CNN 性别识别模型。将相邻两帧图像作为该模型的双输入,通过离散余弦变换计算筛选出高质量块,并将图像块分为高频信息图像块和低频信息图像块。将高频信息图像块通过超分辨率技术提高分辨率,与低频信息块进行整合,实现图像的局部超分辨率。得到的整合后的图像作为分类器的输入进行性别识别任务。通过融合网络中间层进行特征值优化以提高分类器准确率。在现实世界模糊人脸数据集上进行实验,证明了该模型具有较强的应用性,实验结果证明了该模型的识别有效性。实验结果表明该模型实验结果精确度、召回率、 F_1 —Score 在无约束的模糊现实数据集上有明显提升。

参考文献:

[1] 李同宇,李卫军,覃鸿. 基于特征融合的人脸图像性别识别[J]. 智能系统学报,2013,8(6):505-511.

[2] LEVI G, HASSNCER T. Age and gender classification using convolutional neural networks [C] // Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) workshops, Boston, MA: IEEE,2015:34-42.

[3] LIN G S, CHANG M, CHANG Y, et al. A gender classification scheme based on multi-region feature extraction and information fusion for unconstrained images [J]. Multimedia tools and applications,2016(75):9775-9795.

[4] DUAN M, LI K, YANG C, et al. A hybrid deep learning CNN-ELM for age and gender classification [J]. Neurocomputing, 2018, 275:

448 – 461.

[5] DAGHER I, AZAR F. Improving the SVM gender classification accuracy using clustering and incremental learning [J]. Expert systems, 2019,36(3) :e12372.

[6] LIN C J, LIN C H, JENG S Y. Using feature fusion and parameter optimization of dual-input convolutional neural network for face gender recognition[J]. Applied science, 2020, 10 (9) : 3166.

[7] 陈文兵,李育霖,陈允杰. 一种基于 CNN-SE-ELM 的年龄和性别识别模型[J]. 计算机工程与科学, 2021, 43 (5) : 872 – 882

[8] RANJAN R, PATEL V M, CHELLAPPA R. Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition [J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 41 : 121 – 135.

[9] 余婵娟,方昱春. 稀疏表示的特征选择方法在人脸性别和种族识别中的应用[J]. 计算机应用与软件, 2016, 33 (1) : 138 – 141.

[10] 孟庆成, 万达, 吴浩杰, 等. 基于卷积神经网络的混凝土裂缝图像识别方法[J]. 沈阳建筑大学学报 (自然科学版), 2021, 23 (5) : 832 – 840.

[11] KRIZHEVSKY A, SSUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60 (6) : 84 – 90.

[12] CHENG J L, CHENG H L, SHIOU Y J. Using feature fusion and parameter optimization of dual-input convolutional neural network for face gender recognition [J]. Applied science, 2020, 10 (9) : 3166.

[13] GUO S L, MIN K C, YU J C, et al. A gender classification scheme based on multi-region feature extraction and information fusion for unconstrained images [J]. Multimedia tools and applications, 2016, 75 (8) : 9775 – 9795.

Research on Fuzzy Facial Super-Resolution Gender Recognition Model Based on DCT

YAN Weidong¹, HAN Lu², ZHANG Ye¹, LIU Guoqi²

(1. School of Civil Engineering, Shenyang Jianzhu University, Shenyang 110168, China; 2. School of Computer Science and Engineering, Shenyang Jianzhu University, Shenyang 110168, China)

Abstract: To solve the problem of low accuracy of gender recognition based on face images in the field of intelligent monitoring, the DCT-FSR-CNN model is proposed. The model uses two adjacent facial images as the double input of super-resolution and fuses the images with discrete cosine transform (DCT) and local super-resolution technology to improve the overall resolution of the images. Optimize and improve the Alexnet network to enhance the accuracy of the classifiers and the eigenvalue optimization is carried out through the fusion network middle layer. At the same time, the model is experimentally compared with four other gender recognition models, and the accuracy and operation speed of the experimental results on the real and fuzzy facial image data set are significantly improved, which proves the applicability and effectiveness of the model.

Key words: image fusion; super-resolution; convolutional neural networks; gender recognition; computer vision

(责任编辑:王丽娜 英文审校:贾凡华)