

一种基于人脸信息补完的单图像三维人脸重建方法*

郭家富, 严利民**

(上海大学 微电子研究与开发中心, 上海 200444)

摘 要: 针对三维人脸重建中缺失部分人脸信息的侧面人脸重建效果关注不足, 提出了一种基于人脸信息补完的单图像三维人脸重建方法。首先, 使用正面人脸图像生成三维人脸模型, 然后通过这个模型获得侧面人脸图像, 使用这一组相对应的图片来训练网络, 使得最终通过单张侧面人脸图像重建出具有完整人脸信息的三维人脸模型。实验结果证明, 这种方法稳定可靠, 生成的三维人脸模型具有高精度、纹理完整等优点。

关键词: 三维人脸重建; 人脸补完; 侧面人脸

中图分类号: TP391.41 **文献标志码:** A **文章编号:** 1005-488X(2022)01-0015-07

A Single-image Three-dimensional Face Reconstruction Method Based on Face Information Completion

GUO Jiafu, YAN Limin

(Microelectronics Research and Development Center, Shanghai University, Shanghai 200444, CHN)

Abstract: The main direction of 3D face reconstruction lies in the accurate reconstruction of the surface shape, and insufficient attention has been paid to the effect of the side face reconstruction with missing part of the face information. To solve this problem, a single-image 3D face reconstruction method was proposed based on face information completion. Firstly, the frontal face image was used to generate a three-dimensional face model, which was used to obtain a side face image, then this set of corresponding pictures was used to train the network, so that a single side face image was reconstructed finally with three-dimensional face model with complete face information. Experimental results prove that this proposed method is stable and reliable, and the generated 3D face model has the advantages of high precision and complete texture.

Key words: 3D face reconstruction; face completion; profile face

引 言

随着3D电影、动画的发展与普及, 人们对于画

质的需求正在从二维迈向三维。对于人像来说, 面部是其中最为重要的一部分^[1], 具有传递人的身份、情感、意图等关键信息的功能。获取三维人脸的方

收稿日期: 2021-09-06

* 基金项目: 国家自然科学基金(61774101)

作者简介: 郭家富(1996—), 男, 硕士研究生, 研究领域为集成电路设计与人工智能;

严利民(1971—), 男, 博士, 副教授, 研究领域为集成电路设计及系统集成、新型显示技术和计算机视觉。

(E-mail: yanlm@shu.edu.cn)

** 通讯作者

法主要有两种,分别是硬件采集和从人脸图像进行重建。硬件采集的优点是精度高,但是这种方法存在局限,需要通过硬件设备来采集。相对地,基于单图像的三维人脸重建工作目前也有相当多研究^[2-6],但是这些研究普遍针对正面的人脸图像建立,对于缺失信息的侧面人脸图像则表现不佳。

最常见的基于单图像的三维重建方法是由Vetter等人提出的三维可形变模型(3D Morphable Face Models, 3DMM)^[2],3DMM是从一组二维人脸图像的三维人脸扫描数据集中学习而来。该方法使用主成分分析法来生成线性3D人脸模型,并由两组PCA基函数表示,分别为表面形状(Shape)和纹理(Texture)。由于其高优化复杂性,早期基于3DMM的方法会出现局部最优解和不良初始化,并且会非常耗时。近年来基于3DMM改进的方法比较经典的有3DDFA^[3],它使用CNN来回归3DMM参数,从而获得精度更高的表面形状。但是由于缺少用来训练的带有标注的三维人脸数据,导致重建结果的泛化性能不够。2DASL^[4]提出了一项自监督的三维人脸重建技术,该技术在二维和三维中互生成,达成闭环以获得更好的重建效果。但是这些方法主要依赖于线性模型,导致生成的模型受限于模型空间,而且对于输入图像中不可视部分并没有进行过多的处理,在侧面人脸图像作为输入时表现不佳。

除了基于3DMM的方法,还有一些研究使用其他方法来表示三维人脸^[5-6]。Aaron提出一种端

到端的方法进行重建^[5],该方法使用体素来表示三维人脸,可以跳过模型直接生成三维人脸。但是这种方法会导致语义信息的丢失,并且结构过于复杂导致速度过慢。冯瑶^[6]提出了PRnet,使用UV位置图来表示三维人脸,该方法使用了一个轻量级的编码器-解码器网络来实现三维人脸的重建。这种方法速度快一些,但是也没有处理侧面人脸图像中的不可视部分,导致对侧面人脸图像重建时出现失真。

针对非正面人脸图像作为输入时三维重建出现失真的现象,提出一种自监督训练的方法,将建立的三维人脸表面缺失的纹理补完,从而获得具有高精度、完整的、通用的三维人脸。

1 三维人脸表面重建及表面纹理补完方法

算法分为两步,流程如图1所示。首先,通过UV位置图^[6]生成三维人脸表面形状。其次,将不同的人脸输入图像大致分为两种情况,小角度人脸和大角度人脸。小角度人脸指正面或接近正面的人脸图像,对于这种情况,可以直接从输入图像上提取到足够的信息,用以填充纹理。而对于大角度的人脸,从输入图像上提取的信息就不足以获得完整的人脸表面纹理,文章的主要研究目标就是将人脸表面纹理缺失部分补完。最后,将完整的纹理与三维人脸表面形状合成得到三维人脸模型。

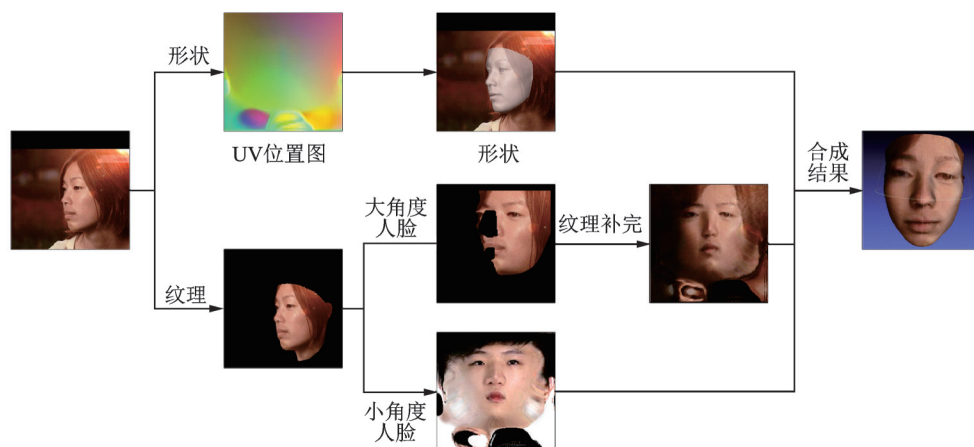


图1 方法概述流程

Fig.1 Process of proposed method

1.1 三维人脸重建

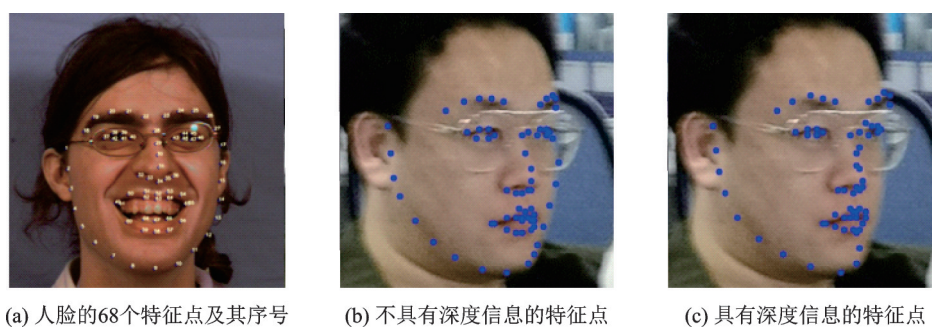
三维人脸重建分为两个部分,分别是表面形状和纹理,其中表面形状指人脸的几何形状,纹理指

在表面形状上填充的颜色信息。对于表面形状的重建,首先要获取人脸的68个特征点^[7]信息,用于人脸区域的选定和表面形状重建的约束。文中选用

UV位置图来表示三维人脸,这种方法可以进行数据降维,使用二维图像来表示三维人脸信息。最后,使用一个端到端的网络来进行三维重建,获取三维人脸表面形状。

为选定人脸区域,文中使用了文献[8]中的3D-FAN来获得人脸面部的68个特征点,相较于之前的面部特征点获取方法,该方法可以直接预测

人脸图像中的3D信息,从而预测出具有深度信息的特征点。使用具有深度信息的人脸特征点有助于三维人脸表面形状参数的重建,对于侧面人脸图像可以获得更加精准的结果。如图2所示,图2(a)所示为面部的68个特征点,图2(b)是只有二维信息的特征点,图2(c)是具有深度信息的特征点。



(a) 人脸的68个特征点及其序号

(b) 不具有深度信息的特征点

(c) 具有深度信息的特征点

图2 人脸68个特征点示意图

Fig.2 Schematic of 68 feature points of the face

UV贴图是用于轻松包装纹理的3D模型表面的平面表示,创建UV贴图的过程称为UV展开。文中使用的UV位置图就是将三维的空间信息存储到二维的图像中:将空间中的点坐标 (x, y, z) 转换成颜色值

RGB,并将其与三维人脸模型表面纹理展开的UV贴图相对应。如图3所示,图3(a)为原始人脸图像,图3(b)为三维人脸模型表面纹理展开的UV贴图,图3(c)为存储到二维的三维空间信息的UV位置图。



(a) 人脸图像

(b) UV纹理图

(c) UV位置图

图3 UV贴图示意图

Fig.3 Schematic of 3 UV maps

使用UV位置图表示三维人脸,也就是使用二维图像来表示三维信息,为了使用输入图像来生成UV位置图,文中采用了一个编码器-解码器结构的网络来预测UV位置图。编码器是残差网络,由一个卷积层和十个残差块构成。解码器使用反卷积网络,由十四层转置卷积层构成,损失函数如下:

$$Loss = \sum \| pos(u, v) - p\tilde{os}(u, v) \| \cdot W(u, v) \quad (1)$$

其中 $pos(u, v)$ 是预测的UV位置图, $p\tilde{os}(u, v)$ 是已有的真实UV位置图, $W(u, v)$ 代表面部区域所占权重,文中将其设为(68个特征点):(五官):(其他面部区域)=16:9:4。人脸特征点是构成人脸形状的

主体骨架,所以设置权重最高;其次是在人脸上比较突出的五官,有眼睛、鼻子和嘴巴;最后才是面部的其他区域。文献[6]中将这个比例设置为16:4:3,但是部分结果中器官重建出现阴影,所以文中将之设置为16:9:4。

在纹理(Texture)方面,如图4所示,根据输入图4(a)以及预测的三维人脸表面形状图4(b),可以获得视点上的颜色信息图4(c),再通过UV位置图将其展开图4(d),这个展开图是根据原图展开得来,所以并不完整,只包含了输入图像所具有的信息,但是对于具有较完整面部信息的正面图片来说,这

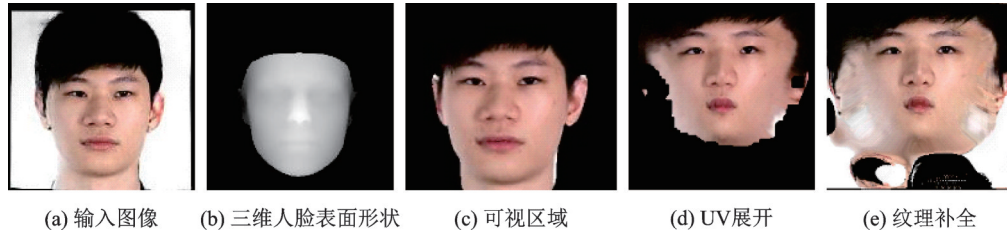


图 4 纹理生成示意图

Fig.4 Schematic of texture generation

些信息已经足够用来构建完整的三维人脸模型,再将剩余位置补全如图 4(e)所示。

1.2 三维人脸表面信息补完

为了填补侧面人脸图像作为输入时,生成的三维人脸模型中不可见顶点上的纹理,文中提出一种自监督的方法,通过正面人脸输入所建立的完整三

维人脸模型,构建出一个训练对用以训练。主要思路如图 5 所示,将三维人脸乘上一个旋转矩阵,再将其渲染回二维平面获得侧面人脸图像,通过这种方法获得一个训练对:同一人脸的正面图像与侧面图像。通过旋转所获得的图像,即为侧面人脸图像,需要进行纹理补完,而相对应的输入图像则可以作为监督来对网络进行训练。

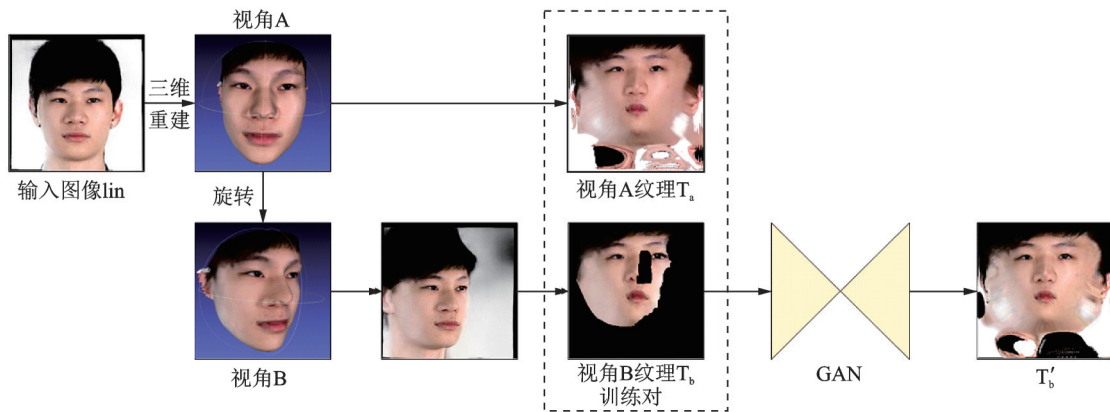


图 5 纹理补完方法流程图

Fig.5 Flow chart of texture completion method

网络架构方面,选择使用生成对抗网络(Generative Adversarial nets, GAN)^[9]来生成完整的纹理,生成对抗网络结构由两部分组成,生成器 G 和鉴别器 D,二者在对抗中达到平衡,从而生成更为逼真的图像。生成器由编码器,转换器和解码器三个部分组成。鉴别器使用 Pix2PixHD^[10]中的多尺度鉴别器,这种鉴别器已经被证明^[11]具有更好的训练效果。多尺度鉴别器与普通鉴别器的区别在于其分别在原图和原图的 1/2, 1/4 三个不同的尺度上进行鉴别并对结果取平均,这里的不同尺度是指在编码器的不同层提取的图像特征上进行鉴别。

损失函数分为三部分,第一部分是生成器和鉴别器的对抗损失:

$$L_{GAN}(G, D_k) = E_{I_{in}}[\log D_k(I_{in}, T_b')] + E_{T_b'}[\log(1 - D_k(I_{in}, G(I_{in})))] \quad (2)$$

对抗损失是为了完成生成对抗网络的任务,鉴别器

D 要足够接近 1 来区分输出图像与真实图像,而生成器要“欺骗”鉴别器,使得鉴别器 D 的结果接近 0,即:

$$\min_G \max_{D_1, D_2, D_3} \sum_{k=1,2,3} L_{GAN}(G, D_k) \quad (3)$$

第二个损失函数是特征匹配损失,由于生成器必须生成多个尺度的中间数据,这种损失函数可以使训练更加稳定。令鉴别器 D_k 第 i 层的特征提取为 $D_k^{(i)}$,则特征匹配损失为:

$$L_{FM}(G, D_k) = E_{(I_{in}, T_b')} \sum_{i=1}^m \frac{1}{n_i} [\|D_k^{(i)}(I_{in}, T_b') - D_k^{(i)}(I_{in}, G(I_{in}))\|_1] \quad (4)$$

其中 m 代表总层数, n_i 代表第 i 层的总元素数。

最后,再加上一个感知损失来规范生成结果的身份,将生成结果 T_b' 和输入图像 I_{in} 分别送入 VGG16^[12]提取特征,令 $F^{(i)}$ 为鉴别器第 i 层提取的特

征,则感知损失为:

$$L_C = \sum_{i=1}^m \frac{1}{n_i} [\|F^{(i)}(T'_b) - F^{(i)}(G(I_{in}))\|_1] \quad (5)$$

其中 m 代表总层数, n_i 代表第 i 层的总元素数。总的损失函数为:

$$\min_G ((\max_{D_1, D_2, D_3} \sum_{k=1,2,3} L_{GAN}(G, D_k)) + \lambda_1 (\sum_{k=1,2,3} L_{FM}(G, D_k)) + \lambda_2 (L_C)) \quad (6)$$

这样训练出的生成对抗网络,可以使用侧面人脸图像作为输入,通过上节中建立的三维人脸表面模型,生成完整的三维人脸表面纹理。这种方法在训练数据缺失的情况下,仍然可以生成较好的结果。

2 实验

使用 300W-LP^[13] 人脸数据作为三维人脸表面形状的训练集,使用 MS-Celeb-1M 经过 ArcFace 清洗的数据集 MS1MV2^[14] 作为纹理补全的训练集,使用 AFLW2000-3D^[15] 和 AFLW-LFPA^[16] 作为测试集。使用 3DDFA, 2DASL, PRNet 作为三维人脸重建的对比数据,同时增补部分人脸正面化算法^[17-19] 作为纹理补完的对比数据。使用均方差作为表面

形状的比较标准:

$$NME = \frac{1}{N} \sum_{k=1}^N \frac{\|x_k - y_k\|_2}{d} \quad (7)$$

其中 N 是三维人脸上的总顶点数, d 为输入图像中人脸的两眼内宽, x_k 代表预测的三维人脸的第 k 个顶点, y_k 是真实三维人脸数据中与 x_k 对应的顶点。

表 1 是在 AFLW2000-3D 和 AFLW-LFPA 中各个方法的均方差结果,可以看出在正面,即 $0 \sim 30^\circ$ 的情况下,三维人脸表面形状重建结果与目前表现较好的两种方法持平;在具有较大角度的人脸图像重建中,文中的方法要略好一些,这是因为使用了预测的三维人脸特征点作为三维重建的辅助手段。

图 6 是正面人脸重建效果对比图,图 7 是侧面人脸重建效果对比图,可以看出,没有人脸补完的算法对于人脸图像中不可视区域的处理是比较粗

表 1 各种方法重建结果均方差结果

Tab.1 NME results in different methods

方法	AFLW2000-3D				AFLW-LFPA
	$0 \sim 30^\circ$	$30^\circ \sim 60^\circ$	$60^\circ \sim 90^\circ$	平均	平均
3DDFA	3.78	4.54	7.93	5.42	4.72
2DASL	2.75	3.44	4.41	3.53	2.85
PRNet	2.75	3.51	4.61	3.62	2.93
Ours	2.76	3.41	4.37	3.51	2.83



图 6 正面三维人脸重建效果对比图

Fig.6 Comparison of the results with different methods for frontal 3D face reconstruction

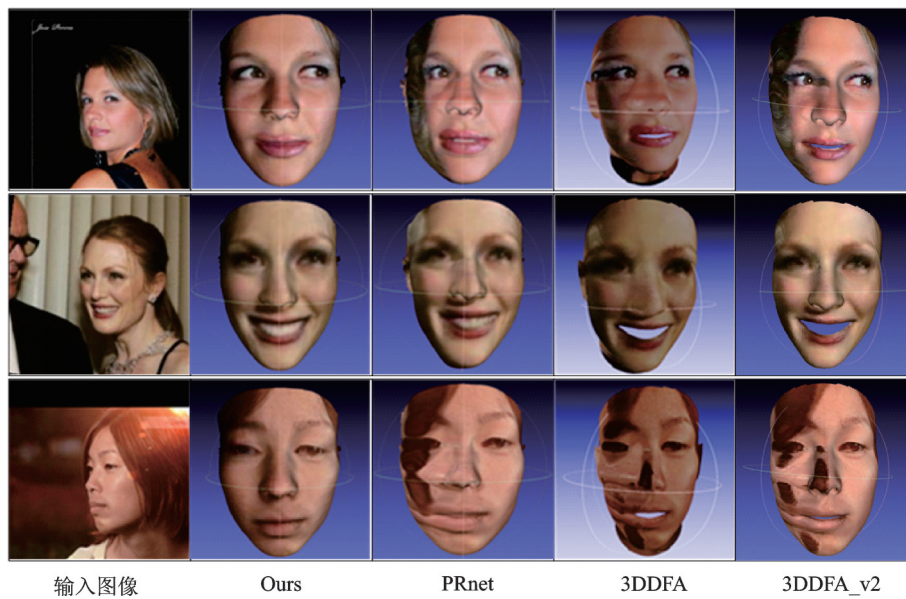


图7 侧面三维人脸重建效果对比图

Fig.7 Comparison of the results with different methods for side 3D face reconstruction

糙的,文中的方法很好地补全了这一区域的缺失。
图8是与一些人脸正面化的结果对比图,可以看出

就算是与专门做人脸补完的算法比较,文中的方法也是具有优势的。

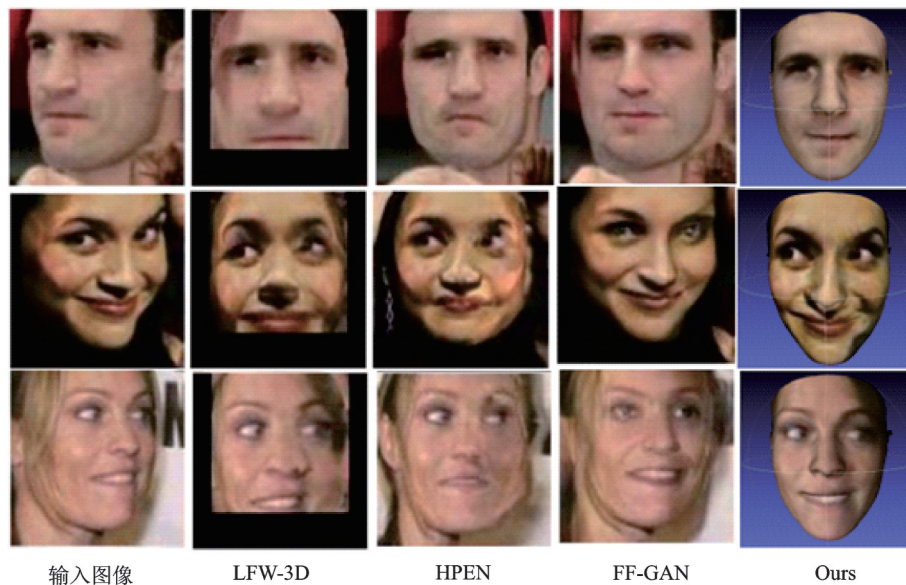


图8 与人脸正面化的结果对比图

Fig.8 Comparison with the results of face frontalization

3 结 论

为了提升三维人脸重建的效果,在三维人脸表面形状方面,使用了具有深度信息的特征点作为输入进行重建,对侧面人脸的重建具有较好的效果。对于侧面人脸的重建任务,提出了一种方法,通过自监督学习训练,使用生成对抗网络来生成三维人脸表面缺失的纹理,从而解决了侧面人脸重建信息

缺失的问题。经验证,文中提出的方法可以取得较好的补完效果。

参 考 文 献

- [1] 姚立平, 潘中良. 基于改进的 HOG 和 LBP 算法的人脸识别方法研究[J]. 光电子技术, 2020, 40(2): 114-118.
- [2] Blanz V, Vetter T. A morphable model for the synthesis of 3D faces[C]. Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, New York, United

- States, 1999: 187-194.
- [3] Zhu X, Liu X, Lei Z, et al. Face alignment in full pose range: A 3d total solution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 41(1): 78-92.
- [4] Tu X, Zhao J, Xie M, et al. 3D Face Reconstruction from a single image assisted by 2D face images in the wild[J]. IEEE Transactions on Multimedia, 2020, 99:1.
- [5] Jackson A S, Bulat A, Argyriou V, et al. Large pose 3D face reconstruction from a single image via direct volumetric CNN regression[C]. Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017: 1031-1039.
- [6] Feng Y, Wu F, Shao X, et al. Joint 3d face reconstruction and dense alignment with position map regression network[C]. Proceedings of the European Conference on Computer Vision (ECCV), Löwenbräu Keller, 2018: 534-551.
- [7] Pandzic I S, Forchheimer R. MPEG-4 facial animation [M]. Chennai India: John Wiley & Sons, Limited, 2002.
- [8] Bulat A, Tzimiropoulos G. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks)[C]. Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017: 1021-1030.
- [9] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Advances in Neural Information Processing Systems, 2014, 3:2672-2680.
- [10] Wang T C, Liu M Y, Zhu J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018: 8798-8807.
- [11] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017: 1125-1134.
- [12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]. Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 2015: 1375-1388.
- [13] Zhu X, Lei Z, Liu X, et al. Face alignment across large poses: A 3d solution[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016: 146-155.
- [14] Deng J, Guo J, Xue N, et al. Arcface: Additive angular margin loss for deep face recognition [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019: 4690-4699.
- [15] Jourabloo A, Liu X. Large-pose face alignment via CNN-based dense 3D model fitting [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016: 4188-4196.
- [16] Guo J, Zhu X, Yang Y, et al. Towards fast, accurate and stable 3d dense face alignment [C]. Computer Vision - ECCV 2020: 16th European Conference, Glasgow, UK, August 23 - 28, 2020, Proceedings, Part XIX 16. Springer International Publishing, 2020: 152-168.
- [17] Hassner T, Harel S, Paz E, et al. Effective face frontalization in unconstrained images [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015: 4295-4304.
- [18] Zhu X, Lei Z, Yan J, et al. High-fidelity pose and expression normalization for face recognition in the wild [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015: 787-796.
- [19] Yin X, Yu X, Sohn K, et al. Towards large-pose face frontalization in the wild [C]. Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017: 3990-3999.

(上接第14页)

- [32] Kana J B, Vignaud G, Gibaud A, et al. Thermally driven sign switch of static dielectric constant of VO₂ thin film[J]. Optical Materials, 2016, 54:165-169.
- [33] Ghasemi M, Choudhury P K, Baqir M A, et al. Metamaterial absorber comprising chromium-gold nanorods-based columnar thin films[J]. Journal of Nanophotonics, 2017, 11(4):043505.
- [34] Artmann K. Berechnung der Seitenversetzung des totalreflektierten strahles [J]. Annalen der Physik, 1948, 437 (1-2) : 87-102.
- [35] Sreekanth K V, Ouyang Q, Sreejith S, et al. Phase change material based low loss visible frequency hyperbolic metamaterials for ultrasensitive label free biosensing [J]. Advanced Optical Materials, 2019, 7(12):1900081.
- [36] Wei C W, Cen M J, Chui H C, et al. Numerical study of biosensor based on α -MoO₃/Au hyperbolic metamaterial at visible frequencies[J]. Journal of Physics D Applied Physics, 2021, 54 (3):034001.
- [37] Skinner J J, Lim W K, Bedard S, et al. Protein dynamics viewed by hydrogen exchange [J]. Protein Science, 2012, 21 (7):996-1005.
- [38] Wu Q, Li N B, Wang Y, et al. A 2D transition metal carbide MXene-based SPR biosensor for ultrasensitive carcinoembryonic antigen detection [J]. Biosensors & Bioelectronics, 2019, 144:111697.
- [39] Grauslund L R, Calvaresi V, Pansegrau W, et al. Epitope and paratope mapping by HDX-MS combined with SPR elucidates the difference in bactericidal activity of two anti-nad monoclonal antibodies [J]. Journal of the American Society for Mass Spectrometry, 2021, 21(7):1575-1582.