Journal of Computer Applications
计算机应用, 2020, 40(3): 842 - 846
ISSN 1001-9081
CODEN JYIIDU
2020-03-10
http://www.joca.cn

文章编号:1001-9081(2020)03-0842-05
DOI:10.11772/j.issn.1001-9081.2019071283

基于生成对抗网络的人像修复

袁琳君[1,2],蒋　旻[1,2*],罗敦浪[1,2],江佳俊[1,2],郭　嘉[1,2]
(1. 武汉科技大学 计算机科学与技术学院,武汉 430065;　2. 智能信息处理与实时工业系统湖北省重点实验室(武汉科技大学),武汉 430065)
(*通信作者电子邮箱345467866@qq.com)

摘　要:人像修复广泛用于基于图像渲染和计算摄影的照片编辑。针对衣着的不同、高矮胖瘦的区别以及姿态的高自由度等因素给人像修复带来的困难,提出了一种基于生成对抗网络(GAN)的高效人像修复方法。算法分为两阶段:第一阶段基于编码器-解码器网络粗略修复图像,然后估计其中人体姿态信息;第二阶段基于姿态信息和GAN来精确修复人像。利用人像姿态信息来连接人像姿态关键点,形成姿态框架并执行膨胀操作,得到人像姿态掩码,以此构造人像姿态损失函数进行网络训练。实验结果表明,与Contextual Attention修复方法相比,所提方法的修复结果在结构相似度(SSIM)上提升了1%。该方法将人像姿态信息加入到修复过程中,有效地约束了待修复区域人像数据的解空间范围,加强了网络对人像姿态信息的关注程度。

关键词:人像姿态信息;生成对抗网络;图像修复;姿态掩码;人像姿态损失
中图分类号:TP391.41　　**文献标志码:**A

Portrait inpainting based on generative adversarial networks

YUAN Linjun[1,2], JIANG Min[1,2*], LUO Dunlang[1,2], JIANG Jiajun[1,2], GUO Jia[1,2]
(1. *College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan Hubei* 430065, *China*;
2. *Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System (Wuhan University of Science and Technology), Wuhan Hubei* 430065, *China*)

Abstract: Portrait inpainting was widely used in the photo editing based on image rendering and computational photography. A lot of factors including the variety in clothing, different body types such as tall, short, fat and thin size, the high freedom degree of human body pose, bring difficulties to portrait inpainting. Therefore, an efficient portrait inpainting method based on Generating Adversarial Network (GAN) was proposed. The algorithm consists two stages. During the first stage, the image was roughly inpainted based on an encoder-decoder network, and then the body pose information in the image was estimated. During the second stage, the portrait was accurately inpainted based on the pose information and GAN. Besides, the key points of the portrait pose were connected by using portrait pose information to form the pose framework and perform the dilation operation, and the portrait pose mask was obtained. Thereby, a portrait pose loss function was constructed for network training. The experimental results show that: compared with the Contextual Attention inpainting method, the proposed method has the SSIM (Structural SIMilarity index) increased by one percentage point. The method, by adding the portrait pose information into the portrait inpainting process, effectively constrains the solution space range of portrait data in the zone to be inpainted, and strengthens the network's attention to the portrait pose information.

Key words: portrait pose information; Generative Adversarial Network (GAN); image inpainting; pose mask; portrait pose loss

0　引言

填充图像的缺失像素,通常称为图像修复或补全。图像修复是计算机视觉中的一项重要任务。随着各摄像技术的发展,人脸和人像的修复[1-4]已经成为图像修复领域的重点。目前大部分研究集中在面向人脸的图像修复[3-4]和编辑[5],并已经取得了较好的效果。但对于人像的修复,研究并不像人脸修复那样火热。

人像修复的数据大部分来源于普通监视视频或个人日常拍照。在拍摄条件的限制下,人像图片的数据的质量往往会受到高光、阴影或遮挡等因素的影响。这些影响一方面造成了对受损人像修复的大量需求,另一方面也提高了人像修复技术的实现难度。另外,人脸目标在形状结构和颜色分布上具有较多共性,但人像目标由于衣着的不同、高矮胖瘦的区别以及人像姿态的高自由度等因素的影响具有更复杂多变的外观。所以,与人脸图像修复相比,人像修复面临着更大的挑战。

针对监控视频场景或个人相机拍摄的图像修复需求,本

收稿日期:2019-07-23;**修回日期:**2019-11-11;**录用日期:**2019-11-13。　**基金项目:**国家自然科学基金资助项目(41571396)。
作者简介:袁琳君(1994—),女,湖北襄阳人,硕士研究生,主要研究方向:计算机视觉、深度学习;　蒋旻(1975—),女,湖南隆回人,教授,博士,主要研究方向:计算机视觉、机器人自动导航;　罗敦浪(1994—),男,湖北武汉人,硕士研究生,主要研究方向:计算机视觉、深度学习;　江佳俊(1996—),男,湖北天门人,硕士研究生,主要研究方向:计算机视觉、深度学习;　郭嘉(1996—),女,河南安阳人,硕士研究生,主要研究方向:计算机视觉、深度学习。

文设计出一种新的对人像图片中缺失区域进行修复的方法。算法的目标:给定一张以人像为主体、并且部分区域信息完全丢失的图像,能够鲁棒地对缺失区域进行图像修复。本文提出的修复方法创新在于:本文将人像姿态估计出姿态信息引入到修复过程中,由于人像姿态的引入大大约束了待修复区域图像数据的解空间范围,所以算法仍能保持一定的鲁棒性。另外,本文利用人像姿态信息,连接人像姿态关键点,形成姿态框架并膨胀框架,得到可以遮盖图片中的人像信息的姿态掩码,根据掩码加入人像损失函数。实验证明,与其他修复方法[6]相比,本文的方法具有更好的修复性能。

1　相关工作

近年来,随着移动视频和视频数据的大规模增长,图像修复问题也受到广泛关注,并且出现了大量相关研究。 大多数图像修复研究可分为两类:基于样本的和基于深度神经网络的方法。

对于基于样本的方法,Efros 等[7]提出通过在已知区域中搜索最佳匹配块,复制最佳匹配块来填充缺失部分的方法。Criminisi 等[8]提出 bestfirst 算法用来搜索最相似的匹配块,并使用找到的匹配块来合成缺失的内容。Simakov 等[9]提出了一种基于双向相似性的全局优化方法,优化方案提高了修复的质量和分辨率。Hays 等[10]是第一个使用大型参考数据集进行孔填充的数据驱动方法,它通过在数据库中找到相似图像区域来修补图像中的漏洞。Whyte 等[11]通过几何和光度配准扩展了这种方法。Barnes 等[12]提出了一种基于补丁的数据结构,用于从图像数据库中进行有效的补丁查询。总的来说,一般基于样本的修复方法在合成纹理上性能优越,但不太适合保留边缘和结构。

在基于深度神经网络的方法方面,也出现了很多代表方法[10,13-15]。相当一部分研究采用编码器-解码器模型来进行图像修复。Pathak 等[13]提出了一种用于图像修复的编码器-解码器模型。它是把具有 64×64 中心缺失区域的待修复图像输入编码器-解码器模型,通过网络训练学习图像特征,以修复 64×64 的中心缺失区域。此外,由于 Ronneberger 等[14]提出的 U-Net 网络对传统卷积网络中的编码器-解码器模型进行了改进,加入连接,去掉了池化操作,以更好地保留图像细节,因此 U-Net 网络被广泛用于图像修复。Yan 等[15]用基于 U-Net 架构的网络进行图像修复,并且网络引入了一个特殊的移位连接层,即 Shift-Net,取得了不错的修复效果。

除了基于编码器-解码器网络模型的修复方法外,近年来,由于生成对抗网络(Generative Adversarial Network, GAN)在各种条件图像生成问题(如超分辨率、图像修复、图像翻译和图像编辑等)中产生了令人信服的结果,因此大量图像修复方法采用了基于 GAN 的架构来解决修复/补全问题。Iizuka 等[16]通过添加全局和局部鉴别器来改进 GAN 模型,保持图像在局部和全局一致,使修复的图片更加真实。此外 Yu 等[6]提出了一种全新的内容感知层来从距离遥远的区域提取近似待修复区域的特征,把内容感知层提取的特征与卷积网络相结合来修复缺失区域。Liu 等[17]提出部分卷积来修复不规则图像。然而,他们无法处理人像图像,因为没有考虑人像的高级语义结构。

在人像姿态估计方面,Chu 等[18]对该人像姿态的关键邻域增加视觉关注,来提高姿态估计的准确性。 Lifshitz 等[19]使用概率关键点投票方案进行图像定位,以获得每个身体部位的位置信息。Insafutdinov 等[20]提出一种基于卷积网络的端到端关联身体关节和特定人的 ArtTrack(Articulated Multi-person Tracking)模型,通过简化和稀疏的人像部分关系图和利用最新的方法,以更快地推理,将大量的计算工作转移到前馈卷积架构上,该架构即使在背景杂乱的情况,也能够检测并关联同一个人的身体关节,以共同推理在图像和时间范围内的身体关节的分配。这些方法大幅提高了人像姿态估计的准确性。

在人像修复方面,Wu 等[1]利用人类解析网络从缺失图片中估计人像解析图,之后在人像解析图的指导下合成输入图片的缺失区域,但由于信息缺失估计出的人像解析并不准确。与 Wu 等[1]的估计人像解析图来修复缺失图像相比,本文的方法结合姿态估计与 Yu 等[6]提出的两阶段式修复网络来进行图像修复。第一阶段粗略修复图像,根据粗略修复图像估计人像姿态信息,在第二阶段引入人像姿态信息,基于生成对抗网络来精确修复人像。另外,本文利用人像姿态信息,连接人像姿态关键点,形成姿态框架并膨胀框架,得到可以遮盖图片中的人像信息的姿态掩码,以此在损失中加入一种姿态信息损失,来约束修复图像的姿态信息,这既解决了因图像缺失信息导致的姿态修复不准确的问题,又解决了图像细节纹理修复问题。

2　基于 GAN 的人像修复网络

本文对文献[6]所提出的网络进行了扩展,网络结构如图 1 所示。方法分为两个部分:粗略修复网络和精确修复网络。

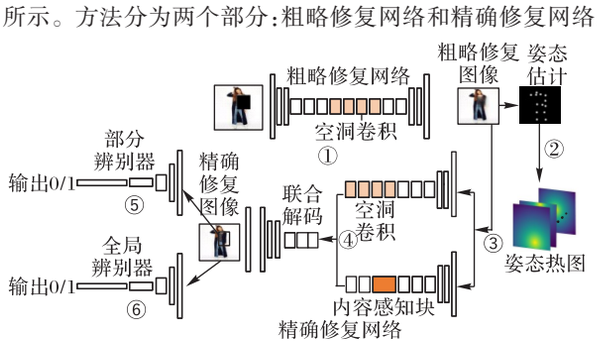


图 1　本文方法网络结构

Fig. 1　Network structure of the proposed method

2.1　粗略修复网络

第一阶段是粗略修复网络,如图 1 的步骤①,它是编码器-解码器结构。解码器的网络结构为 6 层的卷积层和 4 层的扩张卷积层,扩张卷积层是卷积层的变体,本文设置的膨胀系数 η=2,4,8,16。这样设计的扩张卷积不会产生大量的损失信息,增大了感受野。解码器的网络结构为 5 层的卷积层和 2 层的反卷积层,粗略修复图像的输入含有随机的 100×100 缺失区域的待修复图像 $\boldsymbol{x}$,最后得到 256×256×3 的粗略修复图像 $\boldsymbol{k}$。第一阶段网络的作用是对待修复图像做粗略修复预测,修复人像大致结构。

本文训练粗略修复网络使用 L1 损失函数,如式(1)所示:

$$L_{\mathrm{cro}} = E_{(k,y')}\left\|\boldsymbol{k} - \boldsymbol{y}'\right\|_1 \tag{1}$$

其中:$\boldsymbol{k}$ 表示粗略修复网络生成的粗略修复图像,$\boldsymbol{y}'$ 表示真实图像。

粗略修复网络的修复效果虽然不精确,但粗略修复网络

可以修复图像的大致结构，来提高后续姿态估计（图 1 步骤②）的准确性。

2.2 姿态估计

为了提高修复质量，在图 1 的步骤②中，本文使用 Insafutdinov 等[20]提出的一种基于卷积网络的端到端关联身体关节和特定人的姿态估计的 ArtTrack 模型，通过粗略修复图像估计出 14 个姿态位置点（包括头部、颈部、肩、肘、腕、髋、膝和脚踝等关节点）组成，再以每个姿态点为中心生成高斯分布，形成姿态热图 $\boldsymbol{p}$[21]（本文中姿态热量图维度为 256×256×14）。

2.3 精确修复网络

网络的第二阶段是以 GAN 为基础的精确修复网络。它的输入是一个三元组 $\{\boldsymbol{k},\boldsymbol{p},\boldsymbol{M}\}$，其中 $\boldsymbol{k}$ 代表粗略修复图像，$\boldsymbol{p}$ 代表姿态热图，$\boldsymbol{M}$ 代表缺失区域掩码，如图 1 的步骤③。由于姿态热图加入，在训练过程中可以加强网络对人像姿态关注及训练，使修复的图像姿态不会出现扭曲和变形。第二阶段的网络结构分为生成器和辨别器。其中生成器为编码器-解码器结构，编码器由并行的两部分组成：在顶部编码器中，其网络结构与第一阶段的编码器结构类似；底部编码器加入内容感知块，它用卷积的方法，来从已知的图像内容中匹配相似的斑块，通过在全通道上做 softmax 来找出最像待修补区域的斑块，然后使用这个区域的信息做反卷积从而来重建该修补区域。解码器采用联合解码把顶部编码器和底部编码器的编码结果结合在一起，得到 64×64×512 的图片信息输入到联合解码器中，如图 1 的步骤④。联合编码器有 2 个反卷积层，步长为 1/2，图像信息经过计算，输出为 256×256×3 的精确修复图像 $\boldsymbol{x}'$。取缺失区域的修复结果与输入的粗略修复图像相结合，如 $\boldsymbol{y}=\boldsymbol{x}'\odot\boldsymbol{M}+\boldsymbol{k}\odot(1-\boldsymbol{M})$，得到最终修复结果 $\boldsymbol{y}$。联合解码器的输出就是精确修复网络的最终输出，也是本文方法的最终修复结果。

对于本文网络的辨别器，如图 1 的步骤⑤和步骤⑥，辨别器不断地区分发生器结果的真实性（1 或 0），它可以迫使生成器产生更逼真的图像。本文采用文献[16]提出的全局辨别器和缺失区域局部辨别器相结合的方法构造辨别器模块。全局辨别器输入为修复图像 $\boldsymbol{y}$，缺失区域局部辨别器输入为缺失区域的修复图像 $\boldsymbol{y}\odot\boldsymbol{M}$。全局辨别器用于识别图像的全局一致性，而缺失区域局部辨别器的目的是识别图像局部一致性。通过两个辨别器，最终修复结果不仅实现整体一致性，还可以优化细节。

精确修复网络的损失函数包括对抗性损失 L_{GAN}，重建损失 L_{rec} 和姿态信息损失函数 L_{pose}。其中对抗性损失计算方法如式（2）所示：

$$L_{\text{GAN}}=E_{(\boldsymbol{y}',\boldsymbol{M})}[\lg(D(\boldsymbol{y}',\boldsymbol{y}'\odot\boldsymbol{M}))]+E_{(\boldsymbol{x},\boldsymbol{p},\boldsymbol{M})}\lg[1-D(G(\boldsymbol{x},\boldsymbol{p}),G(\boldsymbol{x},\boldsymbol{p})\odot\boldsymbol{M})] \tag{2}$$

其中：G 表示 GAN 中的生成器，D 表示 GAN 中的辨别器，$\boldsymbol{y}'$ 表示真实图像，$\boldsymbol{M}$ 表示为缺失区域的掩码，$\boldsymbol{p}$ 表示姿态热图。

重建损失 L_{rec} 计算修复图像和真实图像之间的 L1 距离。方法如式（3）所示：

$$L_{\text{rec}}=E_{(\boldsymbol{y},\boldsymbol{y}',\boldsymbol{M})}\left\|(\boldsymbol{y}-\boldsymbol{y}')\odot\boldsymbol{M}\right\|_1 \tag{3}$$

其中：$\boldsymbol{y}$ 表示修复网络生成的修复图像，$\boldsymbol{y}=G(\boldsymbol{x},\boldsymbol{p})$，$\boldsymbol{y}'$ 是真实图像，$\boldsymbol{M}$ 表示缺失区域掩码。图像中的缺失区域因为缺失信息，应加重对缺失区域的关注。

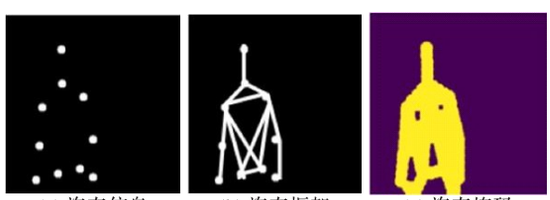
另外，本文提出了姿态信息损失函数 $Loss_{\text{pose}}$，它是针对图像中的人像信息，对人像信息进行约束的一种损失。由第 2.2 节可知，待修复的图片中的人像信息由 14 个位置信息点（包括头部、颈部、肩、肘、腕、髋、膝和脚踝等关节点）组成。本文利用文献[22]中的人像姿态信息 $\boldsymbol{p}$ 生成姿态掩码 $\boldsymbol{M}_p$ 的方法。首先连通 14 个姿态点（肩、肘和腕连接形成手臂；头部、颈部与肩连接和髋与肩部连接形成上半身；髋、膝和脚踝连接形成腿部）形成姿态框架，然后以填充圆的方法来扩充框架，直到逐渐完全覆盖住图像中的人像姿态，生成姿态掩码。生成掩码过程如图 2 所示。损失公式如式（4）所示：

$$L_{\text{pose}}=E_{(\boldsymbol{y},\boldsymbol{y}',\boldsymbol{M}_p)}\left\|(\boldsymbol{y}-\boldsymbol{y}')\odot\boldsymbol{M}_p\right\|_1 \tag{4}$$

修复网络的总损失为：

$$L_{\text{acc}}=L_{\text{rec}}+\alpha L_{\text{GAN}}+\alpha L_{\text{pose}} \tag{5}$$

α 是重建损失的权重，本文实验设置为 0.1。网络的具体训练步骤将在第 3 章详细讲述。



(a) 姿态信息 (b) 姿态框架 (c) 姿态掩码

图 2 掩码生成过程

Fig. 2 Process of mask generation

3 实验

本文实验使用的数据集 Deepfashion Dataset[23]。在这个数据集中含有大量不同人像的不同姿态图片，本文从中抽取 2 000 多张不同人像、不同姿态的代表性图片来进行网络模型的训练。

网络训练分为两个阶段。在第一阶段生成粗糙修复图像的实验中，学习率设为 1E-4，训练数据设置为 100 000 步，使用 Adam 优化器，输入数据是待修复人像图片，输出为粗略修复图像。根据第一阶段得到的粗略修复网络来估计人像姿态信息，人像姿态估计和姿态掩码生成的准确性对后续训练修复网络有比较大的影响，为了证明姿态估计和姿态掩码生成的准确性，本文做了人像姿态估计和姿态掩码生成的实验，效果如图 3 所示。可以看出通过粗略修复图像估计出的人像姿态信息大致准确，生成的姿态掩码虽然人像边界有少量分割错误（如图 3 最右上角子图中人的手部和右下角子图中人的腿部），但是从整体上来说，前景轮廓分割效果仍然具有较高的准确性，对后续修复实验有很大帮助。在精确修复网络中，本文设置训练 200 000 步，实验使用的学习率为 1E-4，使用 Adam 优化器。网络通过训练得到精确修复图像。本文修复结果如图 4 所示。可以看出，本文的人像修复方法具有很好的修复效果，能准确修复图像中的人像姿态。

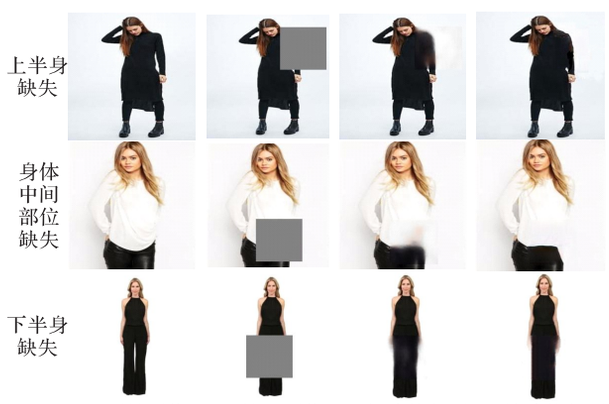
为了验证所提方法的有效性，本文提出的方法和 Contextual Attention[10]的图像修复方法进行了比较。对于人像图片，修复效果比较如图 5 所示，其中（a）是原始图像，（b）

是输入网络的缺失图片,(c)是contextual attention方法的修复效果,(d)是本文方法的修复效果。此外,还计算了它们的SSIM(Structural SIMilarity index)和PSNR(Peak Signal-to-Noise Ratio)值,如表1。从图5可以,本文修复方法有效地约束了姿态修复过程中的扭曲与变形。通过表1实验数据可知,本文提出的方法对于人像数据集的修复性能有一定提高。

(a) 姿态估计效果　(b) 掩码生成效果

图3　姿态估计和姿态掩码效果

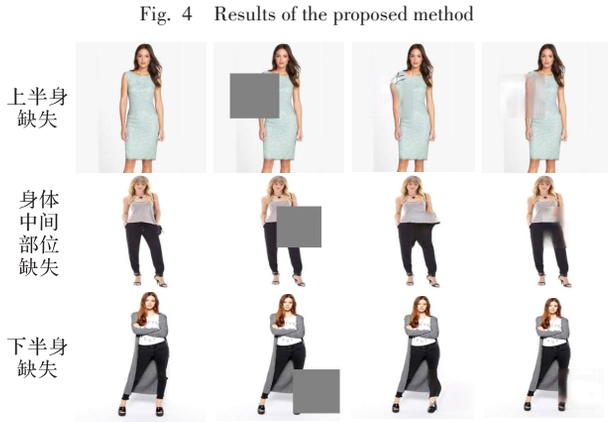
Fig. 3　Results of pose estimation and pose mask



(a) 原始图像　(b) 缺失图像　(c) 粗略修复图　(d) 精确修复图

图4　本文方法修复效果

Fig. 4　Results of the proposed method



(a) 原始图像　(b) 缺失图像　(c) Contextual Attention　(d) 本文方法

图5　本文方法与其他方法对比效果

Fig. 5　Comparison of results of the proposed method with other methods

表1　两种方法的PSNR和SSIM值的比较

Tab. 1　PSNR and SSIM results comparison of two methods

修复方法	SSIM	PSNR/dB
Contextual Attention	0.955 1	37.11
本文方法	0.961 3	38.72

为了定量测量本文方法的修复性能,在测试数据集的修复结果上进行了量化评估,并列在表2与表3中。本文主要考虑两种因素对修复效果的影响:人像缺失的部位和人像缺失部位像素占人像总像素的比例。在人像中结构最复杂的是面部结构,当缺失部位为面部时,修复效果会相对下降,当缺失部分为四肢或身体,模型会有不错的修复效果,由于上半身的结构要比下半身复杂(往往上衣比裤子或者裙子结构复杂),当缺失部分是下半身时修复效果最好;对于缺失像素占人像像素的比例这个因素,比例越大时,修复效果也会相对较差。

表2　人像不同缺失部位对结果影响的量化评估

Tab. 2　Quantitative evaluation of impact of different missing parts of portrait on results

人像缺失部位	SSIM	PSNR/dB
头部	0.936 2	37.63
上半身(不包括头部)	0.949 8	38.67
下半身	0.961 2	39.65

表3　人像缺失程度对结果影响的量化评估

Tab. 3　Quantitative evaluation of impact of degree of portrait missing on results

人像缺失部位像素占人像总像素	SSIM	PSNR/dB
<10%	0.973 1	39.83
10%~30%	0.954 9	38.57
>30%	0.915 2	37.55

4　结语

本文针对人像修复,提出将姿态信息引入到修复过程中的方法,并通过人像姿态信息生成姿态掩模,在损失函数中加入姿态信息损失函数,大大约束了修复图像的姿态信息,使其修复成合理的人像姿态的图像,减小姿态失真概率,这与实际的修复效果一致。但是,在修复脸部与衣着细节方面,修复效果还有待提升,希望在将来的工作中改进该问题,提高算法鲁棒性。

参考文献 (References)

[1] WU X, LI R-L, ZHANG F-L, et al. Deep portrait image completion and extrapolation [EB/OL]. [2019-05-12]. https://arxiv.org/pdf/1808.07757.pdf.

[2] KANAZAWA A, BLACK M J, JACOBS D W, et al. End-to-end recovery of human shape and pose[C]// Proceedings of the 31th IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7122-7131.

[3] 李波. 基于PDE的图像去噪、修补及分解研究[D]. 大连:大连理工大学,2008:10-31. (LI B. Study on PDE-based image denoising, inpainting and decomposition [D]. Dalian: Dalian University of Technology, 2008:10-31.)

[4] 王立,张勇. 弱纹理人脸图像局部破损点修复方法[J]. 计算机仿真,2018,35(11):417-420. (WANG L, ZHANG Y. Weak texture face image local damage point repair method [J]. Computer Simulation, 2018, 35(11):417-420.

[5] SHU Z, YUMER E, HADAP S, et al. Neural face editing with intrinsic image disentangling [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 5444-5453.

[6] YU J, LIN Z, YANG J, et al. Generative image inpainting with contextual attention [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 5505-5514.

[7] EFROS A A, LEUNG T K. Texture synthesis by non-parametric sampling [C]// Proceedings of the 7th IEEE International Conference on Computer Vision. Piscataway: IEEE, 1999: 1033-1038.

[8] CRIMINISI A, PEREZ P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting [J]. IEEE Transactions on Image Processing, 2004, 13(9): 1200-1212.

[9] SIMAKOV D, CASPI Y, SHECHTMAN E, et al. Summarizing visual data using bidirectional similarity [C]// Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2008: 1-8.

[10] HAYS J, EFROS A A. Scene completion using millions of photographs [J]. ACM Transactions on Graphics, 2007, 26(3): Article No. 4.

[11] WHYTE O, SIVIC J, ZISSERMAN A. Get out of my picture! Internet-based inpainting [C]// Proceedings of the 2009 British Machine Vision Conference. Durham: BMVA, 2009: No. 138.

[12] BARNES C, ZHANG F, LOU L, et al. PatchTable: efficient patch queries for large datasets and applications [J]. ACM Transactions on Graphics, 2015, 34(4): Article No. 97.

[13] PATHAK D, KRÄHENBÜHL P, DONAHUE J, et al. Context encoders: feature learning by inpainting [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 2536-2544.

[14] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]// Proceedings of the 2015 International Conference on Medical Image Computing and Computer-Assisted Intervention, LNCS 9351. Cham: Springer, 2015: 234-241.

[15] YAN Z, LI X, LI M, et al. Shift-Net: image inpainting via deep feature rearrangement [C]// Proceedings of the 2018 European Conference on Computer Vision, LNCS 11218. Cham: Springer. 2018:3-19.

[16] IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion [J]. ACM Transactions on Graphics, 2017, 36(4): Article No. 107.

[17] LIU G, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions [C]// Proceedings of the 2018 European Conference on Computer Vision, LNCS 11215. Cham: Springer, 2018: 85-100.

[18] CHU X, YANG W, OUYANG W, et al. Multi-context attention for human pose estimation [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 5669-5678.

[19] LIFSHITZ I, FETAYA E, ULLMAN S. Human pose estimation using deep consensus voting [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9906. Cham: Springer, 2016: 246-260.

[20] INSAFUTDINOV E, ANDRILUKA M, PISHCHULIN L, et al. ArtTrack: articulated multi-person tracking in the wild [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 1293-1301.

[21] BALAKRISHNAN G, ZHAO A, DALCA A V, et al. Synthesizing images of humans in unseen poses [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8340-8348.

[22] MA L, JIA X, SUN Q, et al. Pose guided person image generation [EB/OL]. [2019-05-12]. http://papers. nips. cc/paper/6644-pose-guided-person-image-generation. pdf.

[23] LIU Z, LUO P, QIU S, et al. DeepFashion: powering robust clothes recognition and retrieval with rich annotations [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1096-1104.

This work is partially supported by the National Natural Science Foundation of China (41571396).

YUAN Linjun, born in 1994, M. S. candidate. Her research interests include computer vision, deep learning.

JIANG Min, born in 1975, Ph. D., professor. Her research interests include computer vision, robot automatic navigation.

LUO Dunlang, born in 1994, M. S. candidate. His research interests include computer vision, deep learning.

JIANG Jiajun, born in 1996, M. S. candidate. His research interests include computer vision, deep learning.

GUO Jia, born in 1996, M. S. candidate. Her research interests include computer vision, deep learning.