JOURNAL OF IMAGE AND GRAPHICS

主办：中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

# 中国图象图形学报

2020
12
VOL.25

ISSN1006-8961
CN11-3758/TB

数字浮雕建模 *P2647*

# 中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第25卷第12期（总第296期）

2020年12月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

**主管单位** 中国科学院
**主办单位** 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

**主　　编** 吴一戎
**编辑出版** 《中国图象图形学报》编辑出版委员会
**通信地址** 北京市海淀区北四环西路19号
**邮　　编** 100190
**电子信箱** jig@radi.ac.cn
**电　　话** 010-58887035
**网　　址** www.cjig.cn

**广告发布登记号** 京朝工商广登字20170218号
**总 发 行** 北京报刊发行局
**订　　购** 全国各地邮局
**海外发行** 中国国际图书贸易集团有限公司
（邮政信箱：北京399信箱　邮编：100048）
**印刷装订** 北京科信印刷有限公司

## Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

**Superintended by** Chinese Academy of Sciences
**Sponsored by** Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

**Editor-in-Chief** Wu Yirong
**Editor, Publisher** Editorial and Publishing Board of Journal of Image and Graphics
**Address** No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
**Zip code** 100190
**E-mail** jig@radi.ac.cn
**Telephone** 010-58887035
**Website** www.cjig.cn

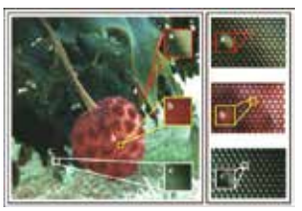
**Distributed by** Beijing Bureau for Distribution of Newspapers and Journals
**Domestic** All Local Post Offices in China
**Overseas** China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
**Printed by** Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

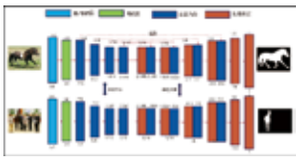
**国外发行代号** M1406
**国内邮发代号** 82-831
**国内定价** 60.00元

2020年12月 第25卷 第12期（总第296期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

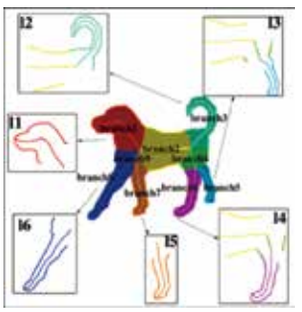
# 目次



光场显著性检测研究综述(第2465页)



融合注意力机制与知识蒸馏的孪生网络压缩(第2563页)



面向可解释性的物体拓扑结构骨架表征方法(第2587页)


## 综述

光场显著性检测研究综述
刘亚美，张骏，张旭东，孙锐，高隽 …… 2465

## 图像处理和编码

自适应卷积的残差修正单幅图像去雨
王美华，何海君，李超 …… 2484

局部特定空间关系统计特征的RVIN噪声检测器
于海雯，易昕炜，徐少平，林珍玉，刘蕊蕊 …… 2494

全局与局部属性一致的图像修复模型
孙劲光，杨忠伟，黄胜 …… 2505

## 图像分析和识别

关键语义区域链提取的视频人体行为识别
马淼，李贻斌，武宪青，高金凤，潘海鹏 …… 2517

针对形变与遮挡问题的行人再识别
史维东，张云洲，刘双伟，朱尚栋，暴吉宁 …… 2530

多特征融合的行为识别模型
谭等泰，李世超，常文文，李登楼 …… 2541

结构特征下的可撤销人脸识别
孙浩浩，邵珠宏，尚媛园，陈滨，赵晓旭 …… 2553

融合注意力机制与知识蒸馏的孪生网络压缩
耿增民，余梦巧，刘峡壁，吕超 …… 2563

联合聚焦度和传播机制的光场图像显著性检测
李爽，邓慧萍，朱磊，张龙 …… 2578

## 图像理解和计算机视觉

面向可解释性的物体拓扑结构骨架表征方法
危辉，余莉萍 …… 2587

自监督深度残差函数映射网络的3维模型对应关系计算
杨军，马中昌 …… 2603

融合自编码器和one-class SVM的异常事件检测
胡海洋，张力，李忠金 …… 2614

抗高光的光场深度估计方法
王程，张骏，高隽 …… 2630

## 计算机图形学

不同类型高度图控制浅浮雕模型生成方法
尚菁，余文杰，王美丽 …… 2647

## 遥感图像处理

残差密集空间金字塔网络的城市遥感图像分割
韩彬彬，张月婷，潘宗序，台宪青，李芳芳 …… 2656

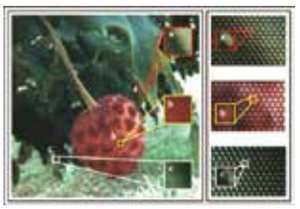
结合判别相关分析与特征融合的遥感图像检索
葛芸，马琳，储珺 …… 2665

改进卷积网络的高分遥感图像城镇建成区提取
侯博文，闫冬梅，郝伟，黄青青，苏秀琴，李青雯 …… 2677

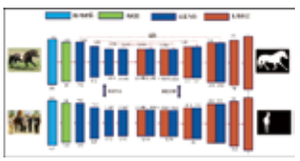
总目次 …… I

Volume 25, Number 12
Published December 16, 2020

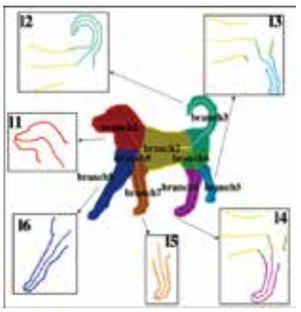
# CONTENTS
## JOURNAL OF IMAGE AND GRAPHICS



Review of saliency detection on light fields(P2465)



Combining attention mechanism and knowledge distillation for Siamese network compression (P2563)



Skeleton characterization of object topology toward explainability (P2587)


### Review

Review of saliency detection on light fields
Liu Yamei, Zhang Jun, Zhang Xudong, Sun Rui, Gao Jun ····· 2465

### Image Processing and Coding

Single image rain removal based on selective kernel convolution using a residual refine factor
Wang Meihua, He Haijun, Li Chao ····· 2484

Random-valued impulse noise detector using local spatial structure statistics
Yu Haiwen, Yi Xinwei, Xu Shaoping, Lin Zhenyu, Liu Ruirui ····· 2494

Image inpainting model with consistent global and local attributes
Sun Jinguang, Yang Zhongwei, Huang Sheng ····· 2505

### Image Analysis and Recognition

Human action recognition in videos utilizing key semantic region extraction and concatenation
Ma Miao, Li Yibin, Wu Xianqing, Gao Jinfeng, Pan Haipeng ····· 2517

Person re-identification based on deformation and occlusion mechanisms
Shi Weidong, Zhang Yunzhou, Liu Shuangwei, Zhu Shangdong, Bao Jining ····· 2530

Multi-feature fusion behavior recognition model
Tan Dengtai, Li Shichao, Chang Wenwen, Li Denglou ····· 2541

Cancelable face recognition with fusion of structural features
Sun Haohao, Shao Zhuhong, Shang Yuanyuan, Chen Bin, Zhao Xiaoxu ····· 2553

Combining attention mechanism and knowledge distillation for Siamese network compression
Geng Zengmin, Yu Mengqiao, Liu Xiabi, Lyu Chao ····· 2563

Saliency detection on a light field via the focusness and propagation mechanism
Li Shuang, Deng Huiping, Zhu Lei, Zhang Long ····· 2578

### Image Understanding and Computer Vision

Skeleton characterization of object topology toward explainability
Wei Hui, Yu Liping ····· 2587

Correspondence Calculation of 3D Models by a Self-Supervised Deep Residual Functional Maps Network
Yang Jun, Ma Zhongchang ····· 2603

Anomaly detection with autoencoder and one-class SVM
Hu Haiyang, Zhang Li, Li Zhongjin ····· 2614

Anti-specular light-field depth estimation algorithm
Wang Cheng, Zhang Jun, Gao Jun ····· 2630

### Computer Graphics

Bas-relief generation controlled by different height maps
Shang Jing, Yu Wenjie, Wang Meili ····· 2647

### Remote Sensing Image Processing

Residual dense spatial pyramid network for urbanremote sensing image segmentation
Han Binbin, Zhang Yueting, Pan Zongxu, Tai Xianqing, Li Fangfang ····· 2656

Remote sensing image retrieval combining discriminant correlation analysis and feature fusion
Ge Yun, Ma Lin, Chu Jun ····· 2665

Urban built-up area extraction using high-resolution remote sensing images with an improved convolutional neural network
Hou Bowen, Yan Dongmei, Hao Wei, Huang Qingqing, Su Xiuqin, Li Qingwen ····· 2677

**中图法分类号**:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2020)12-2505-12

**论文引用格式**:Sun J G, Yang Z W and Huang S. 2020. Image inpainting model with consistent global and local attributes. Journal of Image and Graphics,25(12):2505-2516(孙劲光,杨忠伟,黄胜. 2020. 全局与局部属性一致的图像修复模型. 中国图象图形学报,25(12):2505-2516)[DOI:10.11834/jig.190681]

# 全局与局部属性一致的图像修复模型

孙劲光,杨忠伟,黄胜

辽宁工程技术大学电子与信息工程学院,葫芦岛　125105

**摘　要**:**目的**　图像修复是计算机视觉领域的研究热点之一。基于深度学习的图像修复方法取得了一定成绩,但在处理全局与局部属性联系密切的图像时难以获得理想效果,尤其在修复较大面积图像缺损时,结果的语义合理性、结构连贯性和细节准确性均有待提高。针对上述问题,提出一种基于全卷积网络,结合生成式对抗网络思想的图像修复模型。**方法**　基于全卷积神经网络,结合跳跃连接、扩张卷积等方法,提出一种新颖的图像修复网络作为生成器修复缺损图像;引入结构相似性(structural similarity,SSIM)作为图像修复的重构损失,从人眼视觉系统的角度监督指导模型学习,提高图像修复效果;使用改进后的全局和局部上下文判别网络作为双路判别器,对修复结果进行真伪判别,同时,结合对抗式损失,提出一种联合损失用于监督模型的训练,使修复区域内容真实自然且与整幅图像具有属性一致性。**结果**　为验证本文图像修复模型的有效性,在 CelebA-HQ 数据集上,以主观感受和客观指标为依据,与目前主流的图像修复算法进行图像修复效果对比。结果表明,本文方法在修复结果的语义合理性、结构连贯性以及细节准确性等方面均取得了进步,峰值信噪比(peak signal-to-noise ratio,PSNR)和结构相似性的均值分别达到 31.30 dB 和 90.58%。**结论**　本文提出的图像修复模型对图像高级语义有更好的理解,对上下文信息和细节信息把握更精准,能取得更符合人眼视觉感受的图像修复结果。

**关键词**:图像修复;全卷积神经网络;扩张卷积;跳跃连接;对抗式损失

# Image inpainting model with consistent global and local attributes

Sun Jinguang, Yang Zhongwei, Huang Sheng

*School of Electronic and Information Engineering, Liaoning Technical University, Huludao　125105, China*

**Abstract**: **Objective** Image inpainting is a hot research topic in computer vision. In recent years, this task has been considered a conditional pattern generation problem in deep learning that has received much attention from researchers. Compared with traditional algorithms, deep-learning-based image inpainting methods can be used in more extensive scenarios with better inpainting effects. Nevertheless, these methods have limitations. For instance, their image inpainting results need to be improved in terms of semantic rationality, structural coherence, and detail accuracy when processing the close association among global and local attributed images, especially when dealing with images involving a large defect area. This paper proposes a novel image inpainting model based on the fully convolutional neural network and the idea of generative adversarial network to solve the above problems. This model optimizes the network structure, loss constraints, and training strategies to obtain improved image inpainting effects. **Method** First, this paper proposes a novel image inpainting network as a generator to repair defective images by using effective methods in the field of image processing. A network

**收稿日期**:2019-12-30;**修回日期**:2020-03-17;**预印本日期**:2020-03-26

**基金项目**:国家自然科学基金项目(61702241,61602226)

**Supported by**:National Natural Science Foundation of China (61702241, 61602226)

framework based on a fully convolutional neural network is then built in the form of an encoder-decoder. For instance, we replace part of convolutional layers in the network decoding stage with dilated convolution. We also apply dilated convolution superposition with multiple dilation rates to obtain a larger input image area compared with ordinary convolution in small-size feature graphs and then effectively increase the receptive field of the convolution kernel without increasing the calculation amount to develop a better understanding of images. We also set long-skip connections in the corresponding stage of encoding-decoding. This connection strengthens the structural information by transmitting low-level features to the decoding stage. The setting enhances the correlation among deep features and reduces the difficulties in network training. Second, we introduce structural similarity (SSIM) as the reconstruction loss of image inpainting. This image quality evaluation index is built from the perspective of the human visual perception system and differs from the common mean square error (MSE) loss per pixel. This index comprehensively evaluates via an experiment the similarity between two images in their brightness, contrast, and structure. Structural similarity, as the reconstruction loss of an image, can effectively improve the visual effects of image inpainting results. We use the improved global and local context discriminator as a two-way discriminator to determine the authenticity of the inpainting results. The global context discriminator guarantees the consistency of attributes between the image inpainting area and the entire image, whereas the local context discriminator improves the detailed performance of the image inpainting area. Combined with adversarial loss, this paper proposes a joint loss to improve the performance of the model and reduce the difficulties in its training. By drawing lessons from the training mode of generative adversarial networks, we presents a novel method to alternately train image inpainting network and image discriminative network, which obtains an ideal result. In practical applications, we only use image inpainting network to repair defective images. **Result** To verify the effectiveness of the proposed image inpainting model, we compare the image inpainting effect of this model with that of mainstream image inpainting algorithms on the CelebA-HQ dataset by using subjective perception and objective indicators. To achieve the best inpainting effect in controlled experiments, we use official versions of codes and examples. The image inpainting result is taken from loading pre-training files or online demos. We place the specific defect mask onto 50 randomly selected images as test cases and then apply different image inpainting algorithms to repair and collect statistics for the comparison. The CelebA-HQ dataset is a cropped and super-resolution reconstructed version of the CelebA dataset, which contains 30 000 high-resolution face images. The human face represents a special image that not only contains specific features but also an infinite amount of details. Therefore, face images can fully test the expressiveness of the image inpainting method. Considering the algorithm consistent attribute of the global and local images in the controlled experiment, experiment results show that the image inpainting model demonstrates some improvements in its semantic rationality, structural coherence, and detail performance compared with other algorithms. Subjectively, this model has a natural edge transition and a very detailed image inpainting area. Objectively, this model has a peak signal-to-noise ratio (PSNR), and SSIM of 31.30 dB and 90.58% on average, respective, both of which exceed those of mainstream deep learning-based image inpainting algorithms. To verify its generality, we test the image inpainting model on the Places2 dataset. **Conclusion** This paper proposes a novel image inpainting model that shows improvements in terms of network structure, cost, training strategy, and image inpainting results. This model also provides a better understanding of the high-level semantics of images. Given its highly accurate context and details, the proposed model obtains better image inpainting results from human visual perception. We will continue to improve the effect of image inpainting and explore the conditional image inpainting task in the future. Our plan is to improve and optimize this model in terms of network structure and loss constraint to reduce losses in an image during the feature extraction process under a controllable network training setup. We shall also try to make the defect mask do more work with channel domain attention mechanism to further improve the quality of image inpainting. We also plan to analyze the relationship between image boundary structure and feature reconstruction. We aim to improve the convergence speed of network training and the quality of image inpainting by using an accurate and effective loss function. Furthermore, we would use human-computer interaction or presupposed condition to affect the results of image inpainting, which explores more practical values of the model.

**Key words**: image inpainting; fully convolutional neural network; dilated convolution; skip connection; adversarial loss

## 0 引　言

图像修复(image inpainting)是一种利用缺损图像中已知部分的信息预测缺损区域的内容,允许使用替代内容去填充目标区域的技术。其最终目的是保证修复后的图像整体结构连贯统一,修复区域边缘处过渡自然,修复内容细节丰富合理,最好能够使观察者无法分辨图像是否经过修复。

图像修复的概念自 Bertalmio 等人(2000)首次提出至今已有 20 年的时间,随着互联网产业的高速发展,图像修复技术的应用范围愈发广泛,而且顺应社会的需求,提出了很多优秀的图像修复算法并不断推陈出新,主要包括传统算法和基于深度学习的方法两大类。

传统算法根据修复策略分为基于结构扩散的方法(Tsai 等,2001;Shen 和 Chan,2002)和基于纹理合成的方法(Criminisi 等,2004;Shen 等,2009)以及在此基础上提出的基于样本的图像修复方法(Hays 和 Efros,2007;强振平 等,2019a)。基于结构扩散的方法仿照工匠在修复艺术品时由轮廓到细节的过程(强振平 等,2019b),设计算法时优先考虑边界的连贯性和局部区域的一致性,通过偏微分方程等算法将待修复图像中已知的信息向缺损区域扩散,达到修复图像的目的,该类方法对细小的刮痕、较小的非纹理区域的修复效果较好,但面对较大区域图像缺损、图像中大物体移除等任务时难以取得理想的效果。基于纹理合成的方法以及在此基础上提出的基于样本的图像修复方法在纹理细节修复上取得了很好的效果,但在图像高级语义理解和图像的全局结构把握等方面还有所欠缺,在相关信息有限的情况下,经常出现图像修复区域与图像全局语义不一致等问题,不能取得令人满意的图像修复效果。

深度学习方法逐渐用于图像处理任务,其中图像修复任务被定义为条件图像生成问题进行研究。Pathak 等人(2016)提出的上下文编码网络(context encoder,CE)是最早使用卷积神经网络(convolutional neural network,CNN)进行图像修复的方法之一,它对编码—解码网络进行训练,并结合对抗性网络的思想设计损失函数,相较于传统算法取得了不错的图像修复效果,但该方法受限于网络结构,只适合固定大小的低分辨率图像修复任务且修复痕迹明显。Iizuka 等人(2017)以全卷积神经网络结构为框架,使用 Yu 和 Koltun(2016)提出的扩张卷积(dilated convolutions)和双判别网络对上下文编码器进行改进,使修复网络能够修复任意不规则形状且缺损面积较大的图像,但图像修复结果需要后处理(Yang 等,2017)才能达到理想的修复效果,增大了图像修复代价的同时破坏了网络的完整性。Yu 等人(2018)在此基础上提出一种粗修—精修的网络结构并引入注意力机制(attention mechanism),在一定程度上提高了修复效果,但仍存在面对较大面积缺损区域时修复效果不理想的问题。Wang 等人(2018)提出了生成式多列卷积神经网络结构,使用不同尺寸的卷积核来充分提取特征,并设计了一种新颖的置信驱动的图像重建损失,使用了很多技巧去提高图像修复质量,达到了十分优秀的视觉效果,但在处理大型数据集,尤其是数据集中对象或场景类别较多时修复效果不理想,会出现网络参数难以拟合、修复结果的结构和纹理模糊等问题。Portenier 等人(2018)通过扩展输入图像的通道数,将轮廓约束条件、颜色约束条件等人工干预信息传入网络模型,达到了预设条件可以干预图像修复结果的目的。

针对现有方法的不足,本文借鉴了 U-Net(Ronneberger 等,2015)和生成式对抗网络(Goodfellow 等,2014)的设计思路,综合运用图像处理领域成熟有效的方法,从网络结构、损失约束和训练策略等方面对上下文编码网络进行改进,提出了一种全局与局部属性一致的图像修复模型。实验结果表明,本文方法在主观感受和客观指标方面均取得了较大的进步,图像修复结果在全局与局部属性上具有一致性,并且在色彩重构、局部细节等方面的表现突出,更加符合人眼视觉感受。

## 1 本文方法

本文提出一种全局与局部属性具有一致性的图像修复模型。模型使用新颖的图像修复网络对输入的附加有缺损掩模的图像进行修复,使用全局和局部上下文判别网络作为附加网络。在训练过程中,通过对抗式损失,约束图像修复网络的权重学习过程,使图像修复网络能够真实地修复图像。通过对模型采用分阶段的训练方法,在加速网络参数拟合

的同时,进一步提高图像修复质量。本文提出的图像修复模型架构如图 1 所示。

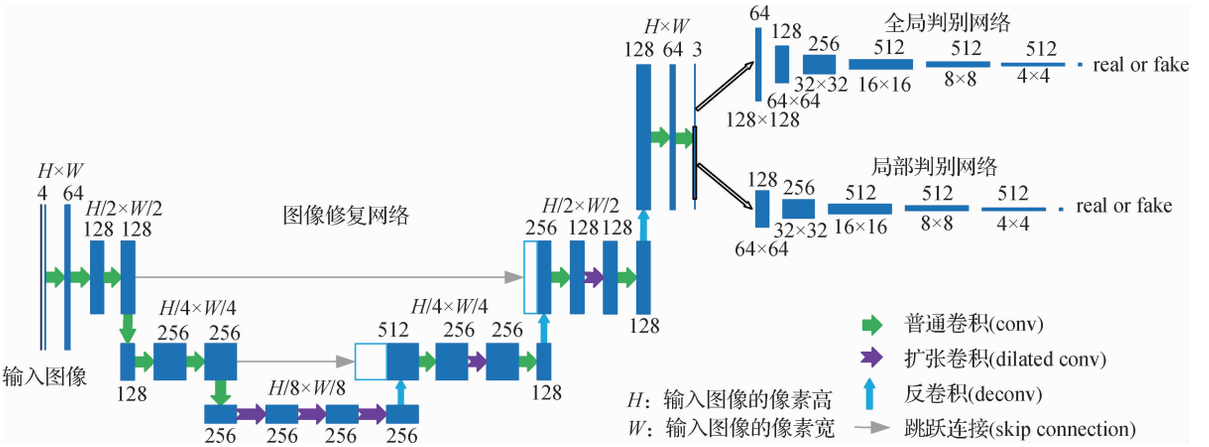



图 1　图像修复模型框架

Fig. 1　Construction of image inpainting model

### 1.1　图像修复网络

图像修复网络(见表 1)以全卷积神经网络为基础,按照编码器—解码器的形式构建。图像修复网络的输入是由能够指示缺损区域的二进制单通道掩模(1 表示待修复的像素)和 3 通道 RGB 原始图像共同组成的 4 通道复合图像,网络的输出为 3 通道 RGB 图像修复结果。为增强网络对全局语义的理解以及对局部细节的把握,本文在网络结构中引入了在图像处理领域已被证明有效的扩张卷积和跳跃连接等方法,通过网络结构上的设计优化提高图像修复效果。

#### 1.1.1　基于扩张卷积的感受野增强

为了保证图像修复质量且保持图像修复网络的输入输出尺寸统一,本文采用步长(stride)为 2 的跨步卷积(strided convolution)替代了可能会丢失关键信息的池化层,并控制对输入图像进行下采样(sub-sampled)的次数,但这会给后续操作带来感受野的问题。定义网络第 $i$ 层卷积核的感受野大小为

$$RF_i = RF_{i-1} + \left((K_i - 1) \times \prod_{n=1}^{i-1} S_n\right) \tag{1}$$

式中,$RF_{i-1}$ 为上一层卷积核的感受野,$K_i$ 为本层卷积核尺寸,$S_n$ 为第 $n$ 层卷积步长。

小尺寸的卷积核感受野不足,网络模型无法得到全局视野,无法有效理解图像高级语义,可能会造成修复区域与全局语义不一致的情况;大尺寸的卷积核又会带来计算量上的指数级增长,给模型训练带来困难。扩张卷积(又称空洞卷积)通过在卷积的卷积核中注入空洞,用 0 填充这些空洞,此时卷积核实际有效计算点数量不变,达到了在保持计算

表 1　图像修复网络结构

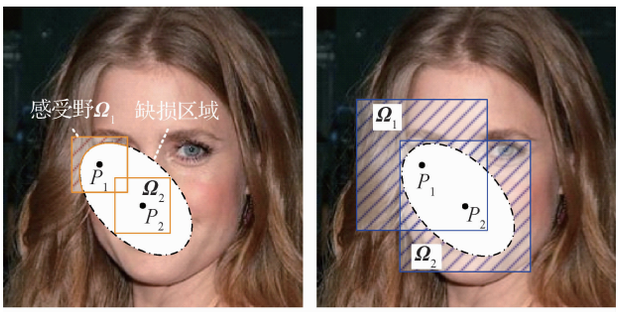
Table 1　Architecture of the image completion network

| 操作类型 | 卷积核 | 扩张率 | 步长 | 输出尺寸 |
|---|---|---|---|---|
| 卷积 | 5×5 | 1 | 1×1 | 64 |
| 卷积 | 3×3 | 1 | 2×2 | 128 |
| 卷积 | 3×3 | 1 | 1×1 | 128 |
| 卷积 | 3×3 | 1 | 2×2 | 128 |
| 卷积 | 3×3 | 1 | 1×1 | 256 |
| 卷积 | 3×3 | 1 | 1×1 | 256 |
| 卷积 | 3×3 | 1 | 2×2 | 256 |
| 扩张卷积 | 3×3 | 2 | 1×1 | 256 |
| 扩张卷积 | 3×3 | 4 | 1×1 | 256 |
| 扩张卷积 | 3×3 | 8 | 1×1 | 256 |
| 反卷积 | 4×4 | 1 | 1/2×1/2 | 256 |
| 卷积 | 3×3 | 1 | 1×1 | 256 |
| 扩张卷积 | 3×3 | 8 | 1×1 | 256 |
| 卷积 | 3×3 | 1 | 1×1 | 128 |
| 反卷积 | 4×4 | 1 | 1/2×1/2 | 128 |
| 卷积 | 3×3 | 1 | 1×1 | 128 |
| 扩张卷积 | 3×3 | 16 | 1×1 | 128 |
| 卷积 | 3×3 | 1 | 1×1 | 128 |
| 反卷积 | 4×4 | 1 | 1/2×1/2 | 128 |
| 卷积 | 3×3 | 1 | 1×1 | 64 |
| 卷积 | 3×3 | 1 | 1×1 | 3 |

量不变的前提下增大卷积核尺寸，进而增大感受野的目的。扩张卷积的等效卷积核尺寸为

$$K = k + (k - 1)(d - 1) \tag{2}$$

式中，$k$ 为标准卷积核尺寸，$d$ 为扩张卷积的扩张率（dilation rate）。普通卷积与扩张卷积的感受野大小对比如图 2 所示。




(a) 普通卷积感受野　　(b) 扩张卷积感受野

图 2　扩张卷积感受野示意图

Fig. 2　The diagram of dilated convolution receptive field

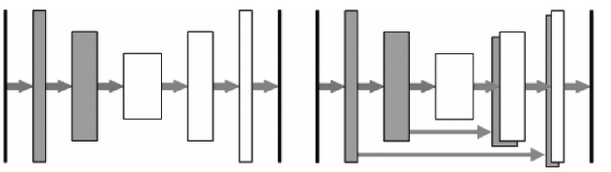
((a) ordinary convolution; (b) dilated convolution)

在较小尺寸的特征图上使用扩张卷积，可以有效捕获比普通卷积更大的输入图像面积。通过设定卷积步长为 1，依靠填充参数（padding）的变化，每层扩张卷积输出的特征图（feature map）的大小保持不变，因此该层输入图像的全图特征信息都得以保留，再通过多层扩张率倍增（$2^1$，$2^2$，$2^3$，…）的扩张卷积叠加，缓解了扩张卷积本身存在的表征点遗漏的问题，降低了图像特征信息丢失，为之后的修复纹理生成打下良好的基础，符合图像修复任务的需要。

1.1.2　基于跳跃连接的特征增强

深度学习通常通过加深神经网络的方式，将低层特征逐层映射以便于提取出更加重要的特征，但这样做会导致在信息逐层传递的过程中丢失部分可能重要的特征，导致深层特征之间的相关性降低，直接影响到模型的表达能力，同时也会造成梯度消失、网络模型训练难度增大等问题。

为解决上述问题，He 等人（2016）提出了残差网络（residual network，ResNet），通过在部分层之间增加短的捷径，即跳跃连接（skip-connection），构建一个残差块（residual block）来保证重要的特征信息能够传递到网络的深层。残差网络的跳跃连接只连接上下邻层，因此被称为短连接。Ronneberger 等人（2015）提出长连接架构网络 U-Net，相较于传统的自编码器（auto-encoder），U-Net 的网络结构左右完全对称，在编码和解码的对应阶段使用长跳跃连接，通过叠加（concatenate）的方式将低层次的特征传递给解码网络，使解码网络可以融合不同规模的特征，提高图像修复结果的细节表现，本文方法采用的就是长跳跃连接。自编码器与 U-Net 的结构对比如图 3 所示。



(a) 自编码器　　(b) U-Net

图 3　自编码器和 U-Net 结构对比

Fig. 3　Comparison of auto-encoder and U-Net

((a) auto-encoder; (b) U-Net)

如上所述，本文提出的模型在图像修复网络中没有使用池化层。对于池化层缺失可能造成的结构性信息缺失，本文通过长跳跃连接传递低层特征到解码阶段进行结构性信息强化。经后期实验结果证明，长跳跃连接的引入能够在有效降低网络模型训练难度的同时提高图像修复质量。

在设计图像修复网络时，为实现压缩图像并提取特征的同时尽可能减少信息损失，在编码阶段控制压缩次数，仅将图像尺寸压缩为原图的 1/8。在此基础上，配合多层不同扩张率的扩张卷积、长跳跃连接补充特征图信息，确保了图像修复模型在解码阶段有足够的特征信息用于在缺损区域生成清晰合理的修复纹理，达到理想的图像修复效果。

### 1.2　图像判别网络

图像判别网络是一个由全局上下文判别网络和局部上下文判别网络共同组成的双路并行判别网络，具有判别图像是真实的还是经过修复的能力。网络依靠卷积神经网络将图像逐层压缩成小的特征向量，再通过全连接层将网络的输出融合在一起，得到与图像真实的概率相对应的连续值，最后使用 sigmoid 函数，使得该值在[0，1]范围内，表示图像是真实图像而不是经过修复的概率。

全局上下文判别网络的输入为缩放至 256 × 256 像素的整幅图像，输出为输入图像是真实图像的概率。全局上下文判别网络的作用是监督图像修复网络真实地修复缺损图像，确保修复区域与全图具有上下文属性一致性。全局上下文判别网络的结

构详见表2。

表2　全局上下文判别网络结构

Table 2　Architecture of the global discriminator

| 操作类型 | 卷积核 | 步长 | 输出尺寸 |
|---|---|---|---|
| 卷积 | 5×5 | 2×2 | 64 |
| 卷积 | 5×5 | 2×2 | 128 |
| 卷积 | 5×5 | 2×2 | 256 |
| 卷积 | 5×5 | 2×2 | 512 |
| 卷积 | 5×5 | 2×2 | 512 |
| 卷积 | 5×5 | 2×2 | 512 |
| 全连接 | – | – | 1 024 |

注：“–”表示无相关参数。

局部上下文判别网络在结构上与全局上下文判别网络基本一致，因输入图像尺寸不同，在设计网络结构时移除了第1层卷积，该判别网络的输入是包含缺损区域的128 × 128像素范围的图像，当图像为真实图像时，则随机选取全图1/4大小的图像块作为输入。局部上下文判别网络的作用是增强修复区域的细节表现，降低生成纹理的模糊程度。局部上下文判别网络的结构详见表3。

表3　局部上下文判别网络结构

Table 3　Architecture of the local discriminator

| 操作类型 | 卷积核 | 步长 | 输出尺寸 |
|---|---|---|---|
| 卷积 | 5×5 | 2×2 | 128 |
| 卷积 | 5×5 | 2×2 | 256 |
| 卷积 | 5×5 | 2×2 | 512 |
| 卷积 | 5×5 | 2×2 | 512 |
| 卷积 | 5×5 | 2×2 | 512 |
| 全连接 | – | – | 1 024 |

注：“–”表示无相关参数。

全局和局部上下文判别网络各司其职，依靠对抗式损失，帮助图像修复网络，提高图像修复质量。与Iizuka等人(2017)将两个子判别网络输出的连续值拼成一个1×2 048维向量后再做判别的做法不同，本文模型的两个子判别网络各自输出一个结果，根据子判别网络的权重综合计算得到最终的判别结果，这样做既有助于判别网络的拟合，又能够提高判别网络的精度，间接提高了图像修复质量。需要说明的是，本文子判别网络的权重是经反复实验得出的经验值，在实际应用中可以根据数据集的特点适当调节权重大小以达到理想的训练效果。

### 1.3　损失函数

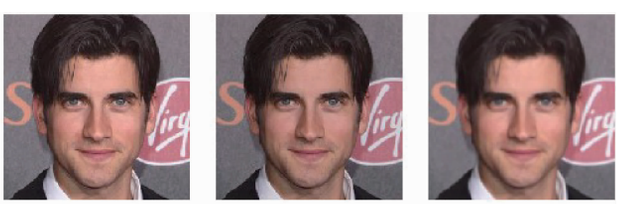
本文为使图像修复网络模型能够真实地修复缺损的图像，取得令人满意的图像修复效果，联合使用了多个损失函数来减小修复结果与原始图像的差距，提高图像修复效果。

#### 1.3.1　重构损失

在图像修复的相关工作中，一般采用均方误差(mean square error，MSE)或平均绝对值误差(mean absolute error，MAE)作为重构损失函数，其中较为常见的是均方误差。对于两幅尺寸均为 $m \times n$ 的图像，均方误差的定义为

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[X(i,j) - Y(i,j)]^2 \tag{3}$$

根据式(3)可知，传统基于均方误差的损失只是简单地以对位元素相乘(element-wise)的方式计算修复图像与原始图像的像素差的平方，然后全图求平均得到最终结果，此种计算方式不足以表达人的视觉系统对图像的直观感受。图4是图像经不同处理的视觉效果对比，图4(b)(c)分别是将灰度值调整为原始图像0.9倍的图像和经过高斯模糊处理后的图像。从人眼视觉的角度明显感觉到调整灰度值的图像比经过高斯模糊处理的图像清晰，更接近于原始图像。分别计算两幅经过调整的图像与原始图像的MSE距离，高斯模糊的图像与原始图像的距离(20.258 1)远小于降低灰度值的图像与原始图像的距离(105.887 6)，即高斯模糊的图像质量比降低灰度值的图像高，这与人眼视觉感知结果不符，因此使用均方误差损失作为图像重构损失来指导图像修复所得到的修复结果可能在结果数据上表现出色，但在人眼视觉感受方面效果不佳。



(a) 原图　(b) 降低灰度值　(c) 高斯模糊

图4　图像经不同处理的视觉效果对比

Fig. 4　Comparison of different image processing((a) original image; (b) intensity reduction; (c) Gaussian blur)

针对这个问题，本文采用 Wang 等人(2004)提出的结构相似性(structural similarity，SSIM)作为图像重构损失来指导图像修复，从人眼视觉感受的角度谋求更好的修复效果。SSIM 分为亮度、对比度和结构相似度，计算分别为

$$l(\boldsymbol{x},\boldsymbol{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{4}$$

$$c(\boldsymbol{x},\boldsymbol{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{5}$$

$$s(\boldsymbol{x},\boldsymbol{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{6}$$

式中，$\mu_x$，$\mu_y$ 为图像 $\boldsymbol{x}$，$\boldsymbol{y}$ 的均值；$\sigma_x^2$，$\sigma_y^2$ 为图像 $\boldsymbol{x}$，$\boldsymbol{y}$ 的方差；$\sigma_{x,y}$ 为图像 $\boldsymbol{x}$，$\boldsymbol{y}$ 的协方差；$C_1 = (k_1L)^2$，$C_2 = (k_2L)^2$ 为两个常数，避免除零，$L$ 为像素值的范围；$k_1 = 0.01$，$k_2 = 0.03$ 为默认值。

定义两图相似度的计算式为

$$SSIM(\boldsymbol{x},\boldsymbol{y}) = l(\boldsymbol{x},\boldsymbol{y}) \cdot c(\boldsymbol{x},\boldsymbol{y}) \cdot s(\boldsymbol{x},\boldsymbol{y}) \tag{7}$$

令 $C_3 = C_2/2$，则对比度与结构相似度公式可进行化简，得到最终的 SSIM 计算式为

$$SSIM(\boldsymbol{x},\boldsymbol{y}) = \frac{(2\mu_x\mu_y + C_2)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{8}$$

同样以图 4 为例，用结构相似性去衡量两幅经过调整的图像与原始图像的相似度，高斯模糊的图像与原始图像的相似度(0.926 0)小于降低灰度值的图像与原始图像的相似度(0.989 9)，这与人眼视觉感知得到的结果相符。

用函数 $C(\boldsymbol{x},\boldsymbol{M}_c)$ 表示图像修复网络，其中 $\boldsymbol{x}$ 为输入图像，$\boldsymbol{M}_c$ 是与输入图像同尺寸的缺损掩模，二进制单通道掩模 $\boldsymbol{M}_c$ 中缺损区域的像素被赋值为 1，其他区域的像素均赋值为 0。定义图像重构损失为

$$S(\boldsymbol{x},\boldsymbol{M}_c) = 1 - SSIM(\boldsymbol{x},C(\boldsymbol{x},\boldsymbol{M}_c)) \tag{9}$$

#### 1.3.2　对抗式损失

本文中，图像判别网络的损失作为对抗式损失在网络模型训练中起到重要作用，将神经网络的标准优化问题转化为极小—极大优化问题。同理，图像判别网络可以用函数 $D(\boldsymbol{x},\boldsymbol{M}_d)$ 表示。在网络模型训练的第 3 阶段中，图像修复网络和图像判别网络串联训练、联合优化，对于整个图像修复模型，优化函数定义为

$$\min_C \max_D E\left[\log D(\boldsymbol{x},\boldsymbol{M}_d) + \log\left(1 - D(C(\boldsymbol{x},\boldsymbol{M}_c),\boldsymbol{M}_c)\right)\right] \tag{10}$$

式中，$\boldsymbol{M}_d$ 表示随机掩模，$E$ 表示期望，期望值为一个训练批次中图像 $\boldsymbol{x}$ 的像素平均值.

结合图像重构损失和对抗式损失，得到最终的联合优化函数为

$$\min_C \max_D E[\lambda_1 S(\boldsymbol{x},\boldsymbol{M}_c) + \lambda_2 \log D(\boldsymbol{x},\boldsymbol{M}_d) + \lambda_2 \log\left(1 - D(C(\boldsymbol{x},\boldsymbol{M}_c),\boldsymbol{M}_c)\right)] \tag{11}$$

式中，$\lambda_1$ 和 $\lambda_2$ 分别为图像重构损失和对抗式损失的权重，经多次实验总结得到权重参考值为 $\lambda_1 = 0.9$，$\lambda_2 = 0.01$。联合优化函数适用于训练的第 3 阶段，即图像修复网络与图像判别网络联合训练，此阶段对图像修复模型进行微调，提高图像修复效果。

### 1.4　模型训练

本文为加速网络模型拟合并提高图像修复质量，参考 Salimans 等人(2016)的工作，采用分阶段训练的方式，交替训练图像修复网络与图像判别网络，具体训练步骤如下：

输入：附加缺损掩模的图像 $\boldsymbol{X}$。

输出：修复后的图像 $\boldsymbol{Y}$。

1)依据批大小(batch-size)从训练集中随机抓取图像并做缩放、随机翻转等预处理。

2)第 1 阶段训练：

(1)用随机大小的缺损掩模 $\boldsymbol{M}_c$ 遮盖训练图像，缺损区域用数据集平均像素值填充；

(2)将处理后的训练图像输入图像修复网络修复图像，此时图像判别网络不参与训练；

(3)根据输入和输出图像，用式(9)计算图像重构损失并更新图像修复网络参数。

3)第 2 阶段训练：

(1)用随机大小的缺损掩模 $\boldsymbol{M}_d$ 遮盖训练图像，缺损区域用数据集平均像素值填充；

(2)使用参数固定的图像修复网络修复图像，采用二分类交叉熵损失(binary cross entropy loss，BCE_loss)计算图像判别损失并更新图像判别网络参数。

4)第 3 阶段训练：

(1)用随机大小的缺损掩模 $\boldsymbol{M}_c$ 遮盖训练图像，缺损区域用数据集平均像素值填充；

(2)将图像修复网络与图像判别网络联合训练，用式(11)计算图像修复损失，根据联合损失对整体网络模型进行微调。

本文实验在 Ubuntu18.04.2 系统下基于深度学

习框架 Pytorch 1.1.0 进行训练和测试。硬件环境为 Intel® Core™ i7-8700K 处理器(3.70 GHz)、32 GB 内存、NVIDIA GTX 1080Ti 显卡。图像修复模型进行一次完整的训练需要 21 h 左右,整个训练周期约为 2 周。

## 2 实验结果与分析

为验证本文方法在图像修复任务中的有效性,将本文方法与图像修复领域中 3 项优秀工作在主观效果和客观指标方面进行横向对比,通过人眼主观感受和客观数据综合评估本文方法。

本文方法主要在 CelebA-HQ 数据集上进行训练和测试。CelebA-HQ 数据集是在人脸数据集 CelebA 数据集的基础上对图像进行筛选后经裁剪、超分辨重建得到的数据集,包含 30 000 幅高分辨率人脸图像。之所以选择 CelebA-HQ 数据集作为本文实验数据集,一方面是因为人脸是最特殊的图像,既包含特定的五官结构,又在细节方面有无限变化的可能,能充分考验网络模型的修复能力;另一方面是因为图像修复领域多数相关工作均使用过该数据集,有助于对比实验真实有效地进行。为防止网络模型在训练过程中出现过拟合等问题,本文在实际使用过程中实施了随机反转、打乱数据等操作。为证明本文方法的通用性,本文提出的图像修复模型在 Places2 数据集上也进行了训练和测试。

为确保对比实验中各算法均达到最好的修复效果,本文涉及的对比实验所用代码、示例均为该算法作者发布的官方版本。图像修复结果均通过加载网络模型及官方预训练文件或使用线上演示(live-demo)得到。同时,为公平地对 4 种图像修复算法进行对比评价,本文展示的图像修复效果均为网络原始输出,未经过任何后处理。

### 2.1 主观效果对比

图 5 是 4 种基于深度学习的图像修复算法对 3 组 6 幅图像进行图像修复所得到的结果。6 幅图像分为 3 组,分别表示单一五官缺损、五官局部缺损和面部大面积缺损 3 种图像缺损情况。图 5(b)为附加了缺损掩模的图像,即神经网络的输入。图 5(c)—(f)为对比算法和本文方法的图像修复效果。为验证算法的鲁棒性,每组图像中男女图像各一幅,且肤色、五官等外貌差异较为明显。

由实验结果可见,Iizuka 等人(2017)提出的改进上下文编码器的方法受限于网络结构,在高分辨率数据集上表现不佳,虽正确修补了缺损的五官,但修复区域出现较明显的像素噪声,修复痕迹明显,尤其在面部大面积缺损的情况下,修复结果五官轮廓模糊,细节表现不佳。Yu 等人(2018)提出的结合注意力机制的渐进修复网络在小区域图像缺损的修复结果上表现良好,细节较为丰富,缺损边缘处过渡自然,没有明显的修复痕迹,但在面部大面积缺损的情况下修复效果不佳,缺损区域的修复效果没有统一,修复痕迹较为明显。Wang 等人(2018)提出的生成式多列卷积网络模型是 3 组对比算法中表现最好的,在各种缺损情况下均能准确地修复缺损区域,五官轮廓清晰,细节信息处理到位,但缺损边缘处过度不自然,修复区域与其他区域存在一定程度的色差,需要经过后处理才能达到最佳效果,这增加了图像修复成本。相比而言,本文方法的修复效果表现良好,在各种情况下均能准确地修复图像。得益于图像修复网络结构中扩张卷积的使用,图像修复网络能捕捉到缺损区域周围更大范围的区域,因此本文方法在五官局部缺损的情况下,能根据图像已知的信息去修复缺损区域,相比于对比算法,图像修复结果更接近原始图像。长跳跃连接和对抗式损失的作用更多体现在增强图像的细节表现和全局与局部属性一致性方面。由图 5 可见,本文方法得到的图像修复结果在缺损边缘处过渡自然,五官细节较为丰富,修复区域与整幅图像具有属性一致性。从图像修复结果的主观视觉感受对比可以发现,本文方法的修复结果更为真实自然,这要归功于本文采用更符合人眼视觉系统的结构相似性(SSIM)而非均方误差(MSE)作为图像重构的损失函数。本文方法在 Places2 数据集下的测试结果如图 6 所示。

### 2.2 客观指标对比

为了定量衡量图像修复效果,以平均 $L_1$ 误差(mean $L_1$ loss)、平均 $L_2$ 误差(mean $L_2$ loss)、峰值信噪比(peak signal-to-noise ratio, PSNR)、结构相似性指数(SSIM)这 4 项在图像修复任务中常用的图像质量评价指标对不同算法的图像修复结果进行评价分析。其中,结构相似性已在 1.3 节介绍,下面对峰值信噪比进行说明,其计算式为

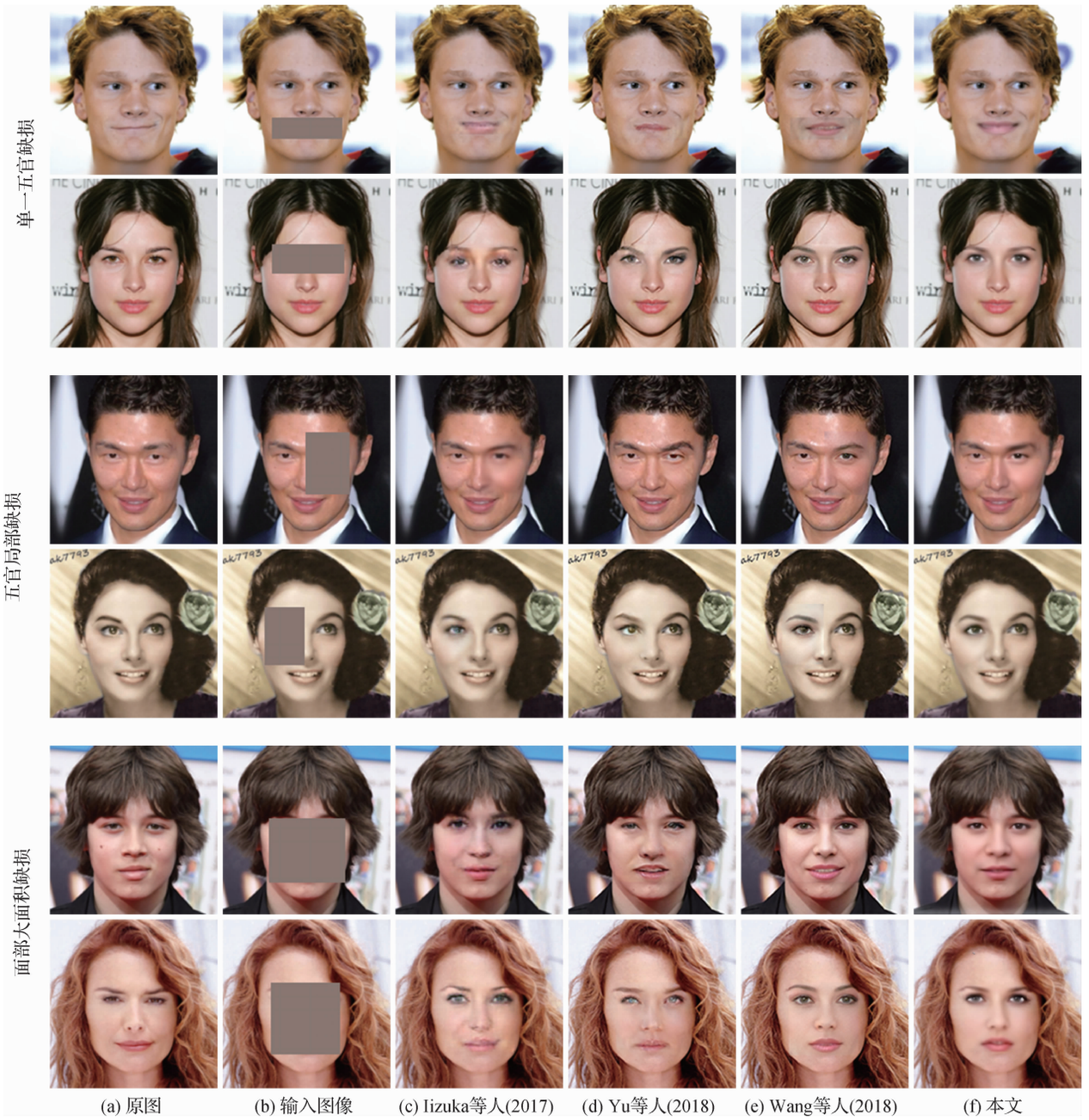



图 5　图像修复主观效果对比

Fig. 5　Comparison of subjective effects of image inpainting((a) original images; (b) input images; (c) Iizuka et al. (2017); (d) Yu et al. (2018); (e) Wang et al. (2018); (f) ours)

$$PSNR = 10 \cdot \lg\left(\frac{MAX_I^2}{MSE}\right) \tag{12}$$

式中，$MAX_I$ 为图像中可能的最大像素值。计算结果的单位是 dB，数值越大，表示修复结果失真越小。数值高于 40 dB，说明图像质量极好，即非常接近原始图像；在 30 ～ 40 dB 通常表示图像质量较好，可以察觉图像失真，但可以接受；在 20 ～ 30 dB 说明图像质量较差，失真程度不可接受。

在 4 种图像质量评价指标中，平均 $L_1$、$L_2$ 误差专注于计算两幅图像对应像素间的差异，与峰值信噪比(PSNR)类似，由于未考虑人眼的视觉特性，因此相较于结构相似性(SSIM)，使用平均 $L_1$、$L_2$ 误差和峰值信噪比指标得出的结果可能与人类主观视觉感受不完全一致。

关于测试用例图像，本文通过对每幅测试用例图像进行 4 种不同面积掩模遮盖的方式模拟不同程


图 6　本文方法在 Places2 数据集上的实验结果

Fig. 6　Experimental result on Places2 dataset

度的图像缺损，掩模颜色为数据集像素平均值。测试用例图像经附加缺损掩模处理后的效果如图 7 所示。

(a) $I_1$　(b) $I_2$　(c) $I_3$　(d) $I_4$

图 7　测试用例

Fig. 7　Test cases ((a) $I_1$; (b) $I_2$; (c) $I_3$; (d) $I_4$)

在 CelebA-HQ 测试集中，以 1∶1 的男女比例随机选取 50 幅图像进行上述操作，对经过附加缺损掩模处理的测试用例图像，采用不同的算法修复，将修复后的图像与原始图像根据上述 4 种评价指标分别计算图像间的相似度，最后将 50 幅图像得到的结果求均值得到最终结果。实验结果详见表 4。

表 4 列出了在不同面积缺损掩模遮盖的情况下，4 种图像修复算法对缺损图像进行修复后得到的修复图像与原始图像的相似性。根据实验结果分析可知，在 4 种算法从网络结构设计、损失约束等方面均对图像全局和局部属性一致性有所考虑的情况下，本文方法在 4 种评价指标上均取得了较为理想的成绩。在平均 $L_1$、$L_2$ 误差指标上，Iizuka 等人(2017)提出的改进上下文编码网络的方法在较大区域缺损情况下的指标数据似乎更加出色。造成这种结果的原因是，该项工作在模型训练过程中，对图像修复结果进行损失约束时使用的是均方误差(MSE) + 对抗式损失(GAN_loss)的联合损失，因此在更关注像素间差异的指标上的成绩可能更好。在峰值信噪比和结构相似性指标上，本文提出的方法取得了较为明显的效果，虽然 Wang 等人(2018)提出的生成式多列卷积网络模型修复较大面积缺损图像($I_4$)的结果在峰值信噪比这一评价指标上得分较高，但本文方法在更符合人眼视觉感受的结构相似性指标上表现更加出色。综合对比分析各项图像修复质量评价指标的实验数据，可以认为本文方法达到了较为理想的图像修复效果。

综上所述，无论是从主观视觉感受还是客观数据对比，本文算法的图像修复效果明显好于目前较为优秀的基于深度学习的图像修复算法，既能合理地修复缺损区域，做到修复区域边缘过渡自然，内部细节丰富真实，又能保证修复区域与图像整体具有属性的一致性，经过修复的图像整体结构连贯统一，视觉效果符合人眼视觉感受，达到图像修复任务的最终目的。

表 4　本文方法与其他图像修复算法的修复效果对比

Table 4　Comparison of inpainting effect between our method and other image inpainting methods

| 输入 | 算法 | 图像修复质量评价指标 | | | |
|---|---|---|---|---|---|
| | | 平均 $L_1$ 误差/% | 平均 $L_2$ 误差/% | 峰值信噪比/dB | 结构相似性/% |
| 图像 $I_1$ | Iizuka 等人(2017) | 13.871 8 | 1.041 4 | 30.219 0 | 89.162 8 |
| | Yu 等人(2018) | 14.736 9 | 1.769 8 | 32.197 9 | 92.374 4 |
| | Wang 等人(2018) | 21.779 3 | 1.378 3 | 34.819 1 | 93.176 7 |
| | 本文 | **13.847 0** | **0.957 6** | **35.894 5** | **93.984 9** |
| 图像 $I_2$ | Iizuka 等人(2017) | 15.225 2 | 2.158 3 | 28.741 2 | 86.341 7 |
| | Yu 等人(2018) | 18.508 6 | 3.225 1 | 29.893 0 | 90.234 7 |
| | Wang 等人(2018) | 23.632 3 | 2.721 9 | 31.075 8 | 91.983 5 |
| | 本文 | **15.132 2** | **2.054 9** | **31.669 4** | **92.725 2** |
| 图像 $I_3$ | Iizuka 等人(2017) | **17.551 9** | 4.234 6 | 25.756 2 | 82.577 3 |
| | Yu 等人(2018) | 22.585 4 | 5.644 8 | 27.026 8 | 84.986 0 |
| | Wang 等人(2018) | 25.062 9 | 4.444 3 | 29.988 0 | 88.191 6 |
| | 本文 | 18.237 9 | **4.076 7** | **30.007 5** | **89.046 9** |
| 图像 $I_4$ | Iizuka 等人(2017) | **19.702 9** | **5.318 1** | 24.144 3 | 81.033 4 |
| | Yu 等人(2018) | 21.541 3 | 7.042 0 | 25.579 8 | 83.571 8 |
| | Wang 等人(2018) | 27.629 0 | 6.396 9 | **27.926 4** | 85.410 3 |
| | 本文 | 20.196 5 | 6.060 1 | 27.613 1 | **86.570 1** |

注：加粗数据为每组最优结果。

## 3　结　论

本文提出了一种新颖的基于全卷积网络，结合生成式对抗网络思想的图像修复模型，用于更好地解决图像修复问题。该模型从图像全局语义理解、全局与局部属性一致的角度出发，针对现有工作存在的问题，借鉴图像处理领域中成熟有效的方法，提出一种新的图像修复模型，通过多方面的改进完善，使模型能够真实有效地修复不同程度缺损的图像，图像修复结果从主观视觉效果到客观数据指标都达到较为优秀的水平。需要说明的是，尽管深度学习具有强大的学习能力和表示能力，但图像修复任务涉及到的场景过于多样化，因此本文提出的是一个通用模型，训练难度与时间代价均在可接受的范围内。在处理特定问题时，对该问题所对应的样本进行针对性训练，可以获得较为理想的图像修复结果。当面对更专业的应用场景时，可以根据实际需要对网络模型进行调整并进行迁移学习(transfer learning)、微调(fine-tune)等操作，达到更好的图像修复效果。

本文提出的是一种端到端的自动图像修复模型，无法人为干预图像修复结果。未来将在本文的基础上，在继续提高图像修复效果的同时，对有条件的图像修复任务作进一步的探索，做到预设条件、人机交互可以直接或间接影响图像修复结果，从而进一步提高模型的实用价值。

## 参考文献(References)

Bertalmio M, Sapiro G, Caselles V and Ballester C. 2000. Image inpainting//Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. New York, USA: ACM Press: 417-424 [DOI: 10.1145/344779.344972]

Criminisi A, Perez P and Toyama K. 2004. Region filling and object

removal by exemplar-based image inpainting. IEEE Transactions on Image Processing, 13(9): 1200-1212 [DOI: 10.1109/TIP.2004.833105]

Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2014. Generative adversarial nets//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: ACM: 2672-2680

Hays J and Efros A A. 2007. Scene completion using millions of photographs. ACM Transactions on Graphics, 26(3): #4 [DOI: 10.1145/1275808.1276382]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Iizuka S, Simo-Serra E and Ishikawa H. 2017. Globally and locally consistent image completion. ACM Transactions on Graphics, 36(4): #107 [DOI: 10.1145/3072959.3073659]

Pathak D, Krähenbühl P, Donahue J, Darrell T and Efros A A. 2016. Context encoders: feature learning by inpainting//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2536-2544 [DOI: 10.1109/CVPR.2016.278]

Portenier T, Hu Q Y, Szabó A, Bigdeli S A, Favaro P and Zwicker M. 2018. Faceshop: Deep sketch-based face image editing. ACM Transactions on Graphics, 37(4): #99 [DOI: 10.1145/3197517.3201393]

Qiang Z P, He L B, Chen X and Xu D. 2019a. Image inpainting using image structural component and patch matching. Journal of Computer-Aided Design and Computer Graphics, 31(5): 821-830 (强振平, 何丽波, 陈旭, 徐丹. 2019a. 利用图像结构成分的优先块匹配图像修复方法. 计算机辅助设计与图形学学报, 31(5): 821-830) [DOI: 10.3724/SP.J.1089.2019.17368]

Qiang Z P, He L B, Chen X and Xu D. 2019b. Survey on deep learning image inpainting methods. Journal of Image and Graphics, 24(3): 447-463 (强振平, 何丽波, 陈旭, 徐丹. 2019b. 深度学习图像修复方法综述. 中国图象图形学报, 24(3): 447-463) [DOI: 10.11834/jig.180408]

Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]

Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A and Chen X. 2016. Improved techniques for training GANs//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: ACM: 2234-2242

Shen B, Hu W, Zhang Y M and Zhang Y J. 2009. Image inpainting via sparse representation//Proceedings of 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing. Taipei, China: IEEE: 697-700 [DOI: 10.1109/ICASSP.2009.4959679]

Shen J H and Chan T F. 2002. Mathematical models for local nontexture inpaintings. SIAM Journal on Applied Mathematics, 62(3): 1019-1043 [DOI: 10.1137/S0036139900368844]

Tsai A, Yezzi A and Willsky A S. 2001. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. IEEE Transactions on Image Processing, 10(8): 1169-1186 [DOI: 10.1109/83.935033]

Wang Y, Tao X, Qi X J, Shen X Y and Jia J Y. 2018. Image inpainting via generative multi-column convolutional neural networks//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: ACM: 329-338

Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4): 600-612 [DOI: 10.1109/TIP.2003.819861]

Yang C, Lu X, Lin Z, Shechtman E, Wang O and Li H. 2017. High-resolution image inpainting using multi-scale neural patch synthesis//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 4076-4084 [DOI: 10.1109/CVPR.2017.434]

Yu F, Koltun V. 2016. Multi-Scale Context Aggregation by Dilated Convolutions[EB/OL]. https://arxiv.org/pdf/1511.07122.pdf

Yu J H, Lin Z, Yang J M, Shen X H, Lu X and Huang T S. 2018. Generative image inpainting with contextual attention//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 5505-5514 [DOI: 10.1109/CVPR.2018.00577]

## 作者简介

孙劲光,1962 年生,女,教授,博士生导师,主要研究方向为计算机图形学、图像处理、知识工程。
E-mail: sunjinguang@lntu.edu.com

杨忠伟,通信作者,男,硕士研究生,主要研究方向为图像处理、计算机视觉。
E-mail:yzw_comvision@163.com

黄胜,男,硕士研究生,主要研究方向为图像处理、目标检测。
E-mail: 18342356464@163.com