

文章编号 1004-924X(2021)10-2481-14

## 场内外特征融合的残缺图像精细修复

徐 涛<sup>1\*</sup>, 周纪勇<sup>2</sup>, 张国梁<sup>3</sup>, 蔡 磊<sup>1</sup>

(1. 河南科技学院 人工智能学院, 河南 新乡 453003;

2. 河南科技学院 机电学院, 河南 新乡 453003;

3. 国家电网全球能源互联网研究院有限公司, 北京 102209)

**摘要:**针对当目标特征严重缺失时现有图像修复方法未能充分利用完整区域预测缺失区域特征,造成修复结果特征不连续、细节纹理模糊等问题,本文提出一种基于场内外特征(EFIF)融合的残缺图像精细修复方法。首先,利用改进的动态记忆网络(DMN+)算法将残缺图像的场内特征及相关的场外特征融合,生成包含场内外特征的残缺图像优化图;其次,构建带有梯度惩罚约束条件的生成对抗生成网络,指导生成器对优化后的残缺图像进行粗修复,获得待修复目标的粗修复图;最后通过相关特征连贯的思想对粗修复图进一步优化,得到最终的精细修复图。在三个复杂度不同的图像数据集仿真验证,并与现有占主导地位的修复模型进行视觉效果和客观数据比对。实验结果显示:本文模型修复的结果在纹理结构上更加合理,在视觉效果和客观数据均优于其他模型,在最有挑战性的 Underwater Targe 数据集中所提算法峰值信噪比(PSNR)最高为 27.01,结构相似性指数(SSIM)最高为 0.949。

**关键词:**残缺图像;相关特征连贯;精细修复;注意机制;场内外特征

**中图分类号:** TP394.1; **文献标识码:** A **doi:** 10.37188/OPE.20212910.2481

## Fine restoration of incomplete image with external features and image features

XU Tao<sup>1\*</sup>, ZHOU Ji-yong<sup>2</sup>, ZHANG Guo-liang<sup>3</sup>, CAI Lei<sup>1</sup>

(1. School of Artificial Engineering, Henan Institute of Science and Technology, Xinxiang 453003, China;

2. School of Mechanical and Electrical Engineering, Henan Institute of Science and Technology, Xinxiang 453003, China;

3. Global Energy Interconnection Research Institute co. Ltd., Beijing 102209, China)

\* Corresponding author, E-mail: xutao@hist.edu.cn

**Abstract:** When large areas of an image are missing owing to unspecified factors, existing image restoration models usually cannot repair the image effectively, leading to repair results that suffer from discontinuity in their characteristics. This study proposes a fine restoration method for incomplete images with external and image features. First, we improved the dynamic memory networks (DMN+) in our study. DMN+ scheme combines the in-field features of an incomplete image and the related off-field features,

**收稿日期:** 2021-03-25; **修订日期:** 2021-05-24.

**基金项目:** 国家重点研发计划资助项目(No. 2019YFB1311000); 河南省科技攻关资助项目(No. 212102210161, No. 182102210302); 河南省高等学校重点科研资助项目(No. 21A520013)

generating an optimized image of the defective image containing external and image features. Next, a generative adversarial generative network with piecewise gradient penalty constraints is constructed. The network instructs the generator to perform a coarse repair on the optimized mutilated image, which results in a coarse repair image of the target to be repaired. Finally, the coarse restoration map is further optimized based on the idea of coherence of related features, and a final fine restoration image is obtained. The algorithm proposed here is verified on three image data sets with varying complexities. Moreover, the visual effects and objective data results of the proposed algorithm are compared with those of the existing dominant restoration model. The restoration results of our model are more structurally sound in terms of texture. Furthermore, our model is superior to other models in terms of both visual effects and objective data. The peak signal-to-noise ratio in the most challenging Underwater Target dataset is 27.01, with a structural similarity index of 0.949.

**Key words:** mutilated images; coherence of relevant features; fine restoration; attention mechanism; external features and image features

## 1 引言

图像修复是将因遮挡、模糊、传输干扰等各种因素造成信息缺失或损坏的图片,利用图像缺失部分邻域的信息和图像整体的结构等信息,并按照一定的信息复原技术对图像的缺失或损坏的区域进行修复<sup>[1-2]</sup>。图像修复技术具有独特的功能,已被应用到许多图像处理的场景中,例如删除图像中不需要的物体,去除目标物上的遮挡物体,修复损坏等。图像修复的核心技术是图像修复区域既要保持全局语义结构,又要保证生成逼真的纹理细节<sup>[3-5]</sup>。

传统的修复方法大多数是基于纹理合成和基于结构来实现图像信息的复原。基于纹理合成的技术通常利用低级特征的区域匹配和补丁来修复图像的缺失像素<sup>[6-9]</sup>,例如:彩色特征(RGB)值的均方差或尺度不变特征变换(SIFT)特征值<sup>[10]</sup>。此类方法对静态纹理的合成效果较好,但并不适应于图像结构复杂的场景。基于结构的图像修复方法常常依据图像信息的结构性原则,采用逐步扩散的方式修复图像,该类方法普遍适用于修复小范围缺失的图像<sup>[11-13]</sup>,当待修复目标信息大量缺失时修复效果就会明显下降。

深度学习方法的快速发展为图像修复模型开辟了一条新的路径。基于深度学习的图像修复模型是将一个深度神经网络通过在数据集中经过大量的训练,学习到图像更深层次的特征信息,从而得到更加逼真的图像修复效果<sup>[14-16]</sup>。

近年来,生成式对抗网络(GAN)作为一种无监督的深度学习模型被应用于图像修复领域<sup>[17,28]</sup>,使得基于深度学习的图像修复得到了进一步的发展,其通过编码器完成图像修复,并借助判别器对修复图像的真实性进行评价,从而有效提高了图像的修复质量。

真实环境中普遍存在遮挡、模糊、传输干扰等因素,极易造成目标特征的严重缺失,现有图像修复方法难以充分利用完整区域预测缺失区域特征,造成修复区域存在着特征不连续、细节纹理模糊等问题。基于此,本文提出一种基于场内外特征(External Features and Image Features, EFIF)融合的残缺图像精细修复方法,将从知识库中提取关于待修复目标的文本描述定义为场外特征,并将待修复目标完整区域的信息定义为场内特征,其修复结果如图1所示。首先,通过编

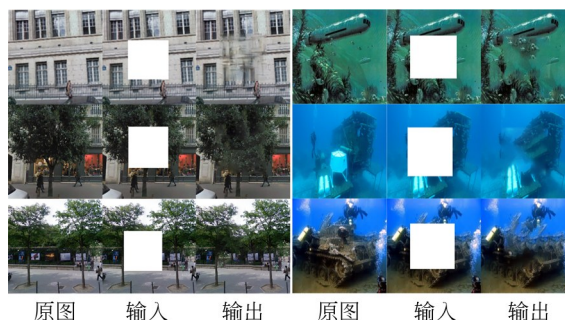


图1 本文所提出修复模型的修复结果

Fig. 1 Restoration results of the proposed restoration model in this paper

码器将场外特征编码到神经网络模型中,利用改进的动态记忆网络(DMN+)算法<sup>[18-20]</sup>对待修复目标一致的场外特征进行检索,并实现与场内特征的融合,生成包含场内外特征的残缺图像优化图,从而解决目标信息不足的问题。其次,构建带有梯度惩罚约束条件的生成式对抗网络(WGAN-GP)<sup>[21-22]</sup>,指导生成器对优化后的残缺图像进行粗修复,获得待修复目标的粗修复图。最终利用相关特征连贯的思想改进 WGAN-GP 的网络结构,得到特征连续、细节纹理清晰的精细修复图。

## 2 相关工作

### 2.1 外部知识与情景记忆

自然知识有助于提高深度学习模型的理解能力,因此,人们越来越倾向于利用场外知识库改进数据驱动模型。现有的人工智能知识库是通过人工构建的<sup>[23]</sup>,因此将知识库中的常识性知识提炼到深度神经网络是一个重要的研究领域。Wu 等人<sup>[24]</sup>将提取到常识性知识编码为向量,并将其输入到神经网络中,与视觉特征结合起来为推理提供了额外的线索。Kumar 等人<sup>[25]</sup>提出了基于情景记忆的网络模型,该模型将注意条件作为输入并经过网络进行推理从而输出结果。然而该模型是否能够直接应用于图像领域还是未知的,为了解决这一问题,Xiong 等人<sup>[20]</sup>在 DMN 模型的基础上改进了其存储和输入层,在输入融

合层采用双向的门循环单元(Gate Recurrent Unit,GRU),并提出了一种能够回答视觉问题的图像输入模型。

### 2.2 图像修复

目前的图像修复方法主要分为两类,一类是非学习的图像修复方法,一类是基于学习的图像修复方法。前者是传统的基于补丁的修复方法,后者是学习图像的深度特征,通过训练基于深度学习的修复模型最终推断缺失区域的特征信息,从而实现残缺图像的修复。非学习的方法主要是通过扩散邻近信息或复制背景最相关区域的信息来填补缺失区域<sup>[10,15]</sup>。这类方法面对缺失区域小的图像可以产生平滑和逼真的结果,然而对于缺失区域较大的图像即使采用高计算成本仍不能获得逼真结果。为了解决这一问题,Barnes 等人<sup>[26]</sup>提出了一种快速最邻计算的图像修复方法,大幅度的提高了计算速度,且能得到一个高质量的修复效果。尽管非学习方法对于表面纹理合成非常有效,但这些模型面对大面积缺失的图像具有一定的局限性。

基于学习的图像修复方法通常是利用深度学习和 GAN 策略实现残缺图像的修复。Deepak 等人<sup>[27]</sup>提出了一种基于上下特征预测的图像修复方法,该方法通过上下文编码器提取整个图像的深度特征,并对缺失部分的生成做出合理假设。同时利用像素损失函数,使生成的结果更加清晰。然而,它在生成精细纹理时的效果并不理想。为了解决这一问题,Yu 等人<sup>[29]</sup>提出了一种

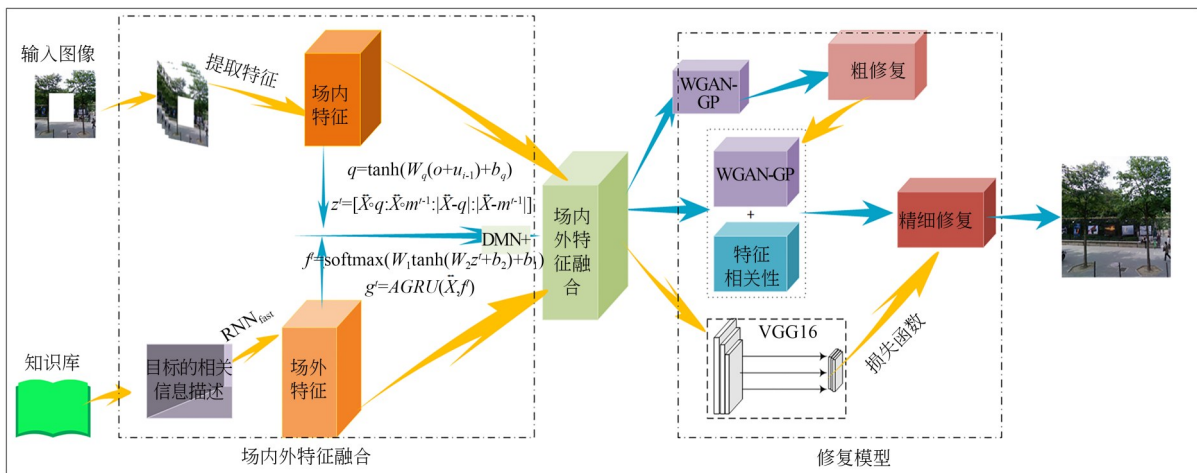


图 2 本文所提出修复模型的原理图

Fig. 2 Schematic diagram of the proposed repair model in this paper

基于上下文注意的生成式图像修复。该模型包含一个上下文注意层,其原理是将已知补丁的特征信息作为卷积处理器,通过卷积设计将生成的补丁和已知的上下文补丁进行匹配,从而得到具有精细纹理的修复效果,但是该方法未考虑残缺区域的语义相关性和特征连续性问题。为了解决这一问题,Liu 等人<sup>[22]</sup>提出了基于精细化深度学习的图像修复模型。该方法采用一种新的连贯语义层,并通过连贯语义层保留上下文语义情景结构,使其推测出来的残缺部分更加合理。

总体来看,现有修复方法仅是对场景内的特征进行深度卷积学习,在面对特征信息不足的残缺图像时,现有修复模型就会呈现出不能有效修复或修复效果精细纹理缺失的现象。因此本文结合自然知识,提出了基于场内外特征融合的残缺图像精细修复方法,其原理结构如图 2 所示。

### 3 场内外特征融合图像修复算法

#### 3.1 常识性知识的检索和嵌入

当图像目标特征严重缺失,造成数据严重不足时,现有的修复模型难以实现对其有效的修复。本文利用知识库中的常识性知识弥补原图像目标信息不足的问题。图像中目标是残缺的,由于特征信息的不足无法实现对其准确地修复。通过检测图像背景中存在的目标 $\beta$ ,根据知识库中相应的常识性知识,推测出与目标 $\beta$ 具有高度相关性的目标集合,定义为 $\{\alpha_1, \alpha_2, \dots, \alpha_k\}$ ,该目标集合中的有效信息对整体缺失区域修复工作构成直接的引导作用。将依据常识性知识获取待修复区域高相似度目标集合的过程表示为:

$$\beta \xrightarrow{\text{contact}} \{\alpha_1, \alpha_2, \dots, \alpha_k\}. \quad (1)$$

目标集合 $\{\alpha_1, \alpha_2, \dots, \alpha_k\}$ 是仅依据知识库中单一相似性对目标 $\beta$ 的初步推理描述,知识库中还包含着大量对目标集合中每个元素 $\alpha_k$ 的细节语义特征描述信息,这些细节语义信息的有效获取将极大弥补待修复目标信息不足的问题。在保证提取过程中目标特征描述与知识库中相应语义实体相匹配的基础上,本文对每个元素 $\alpha_k$ 细节语义特征描述信息的嵌入过程表示为:

$$\alpha_k \xrightarrow{\text{feature}} \{m_1, m_2, \dots, m_n\}, \quad (2)$$

其中, $\{m_1, m_2, \dots, m_n\}$ 是从知识库中检索到的对目标 $\alpha_k$ 细节常识性描述。为了对检测到的常识

性知识进行编码,本文将这些常识性描述 $\{m_1, m_2, \dots, m_n\}$ 转换成单词序列 $\{M^1, M^2, \dots, M^n\}$ ,并且通过 $m^i = w_e M^i$ 的映射关系将句子中的每一个单词映射到一个连续的向量空间中。然后通过基于循环神经网络(RNN)的编码器对这些向量进行编码:

$$h_i^t = \text{RNN}_{\text{fast}}(x_i^t, X_i^{t-1}), t \in [0, j], i \in [0, n], \quad (3)$$

其中: $x_i^t$ 表示检索到的关于物体 $\alpha_k$ 的第 $i$ 句特征描述中第 $j$ 个单词的映射向量, $X_i^{t-1}$ 表示编码器的隐藏状态。同时将双向门控循环单元(GRU)引入到编码过程中,最终编码器的隐藏状态 $h_i^t$ 表示第 $j$ 个特征描述的句子的向量重现。同时将这些文本向量定义为场外特征 $X_E$ 。

#### 3.2 特征融合

从知识库中得到的常识性知识存储到知识存储单元的内存槽中,用于推理和知识的更新。将外部知识融入到目标特征提取的过程中,用来弥补原目标物特征的不足。将 $k$ 个对象在知识库相对应特征的语义描述编码到神经网络,那么存储空间中就含有 $k \times n$ 个相关的特征向量。但是,伴随储存空间特征向量的几何式增长,将极大增加从候选知识中提取有用信息的难度。为了解决这一问题,本文通过改进的DMN+算法实现基于情景问答的场外知识检索,获取最具相关性的待修复目标特征描述,其原理如图 3 所示。

采用改进的DMN+算法对检索场外特征 $X_E$ 进行推理,以获取其特征的描述集合 $\{x_E\}$ 。利用

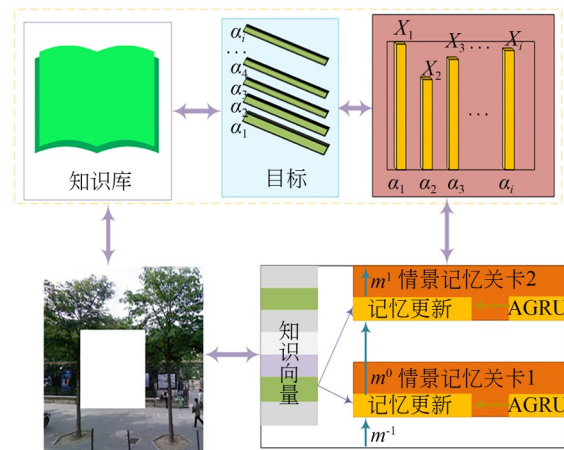


图 3 场外特征的检索与获取

Fig. 3 Retrieval and acquisition of external features

DMN+模型的注意机制生成场内外特征融合的相关性约束。为保证模型对上下文信息的理解和原始输入的记忆,对DMN+的输入层进行改进,将残缺图像  $I_{gt}$  输入到模型中,并有效的提取目标特征信息  $\bar{o}$  作为DMN+模型的第一层输入,第二层及以上层的输入为原始输入  $\bar{o}_{i+1}$  和前一层的输出  $\bar{u}_i$  的总和,如公式(4)所示:

$$\bar{o}_{i+1} = \bar{u}_i + \bar{o}. \quad (4)$$

通过对上下文的学习,使提取到的特征之间进行信息交互,以获取更多的与目标信息相关的场外特征信息,更有利于实现图像的修复。将目标特征向量输入到全连层,其处理过程的具体公式如下:

$$q = \tanh(W_q(\bar{o} + \bar{u}_{i-1}) + b_q), \quad (5)$$

$$z' = [X_E \circ q; X_E \circ m^{t-1}; |X_E - q|; |X_E - m^{t-1}|], \quad (6)$$

$$f' = \text{softmax}(W_1 \tanh(W_2 z' + b_2) + b_1), \quad (7)$$

$$g' = \text{AGRU}(X_E, f'), \quad (8)$$

其中:  $q$  表示目标  $\alpha_k$  映射后的目标向量,  $W_q, b_q$  表示映射参数。  $z'$  表示相关的场外特征  $X_E$ 、模型情景记忆  $m^{t-1}$  和被映射的目标向量  $q$  之间的交互作用,  $\circ$  表示元素相乘的运算符号,  $|\cdot|$  表示元素的绝对值,  $[\cdot]$  表示是多个元素交互运算。值得说明的是,模型情景记忆  $m^{t-1}$  和被映射的目标向量  $q$  需要先通过复制扩展达到具有相同维度,才能与

场外特征  $X_E$  进行交互运算。  $f'$  表示 softmax 层的输出,  $W_1, W_2, b_2$  和  $b_1$  表示的学习参数。  $\text{AGRU}(\cdot)$  表示 DMN+模型的注意机制,其机理是将 GRU 中的更新门用事实 K 的输出权重  $f'_k$  代替:

$$g'_n = f'_n \text{GRU}(x_E, g'_{n-1}) + (1 - f'_n) g'_{n-1}, \quad (9)$$

其中:  $g'_n$  用来表示当所有对目标  $\beta$  的特征描述都被查看后的 GRU 状态。并通过注意力机制利用当前的状态和 DMN+模型的情景记忆更新存储器的记忆状态,如公式(10)所示:

$$m' = \text{ReLU}(W_m[m^{t-1}; g'_k; q] + d_m), \quad (10)$$

其中,  $m'$  表示更新后的情景记忆状态,通过记忆有用的知识信息来弥补原目标数据的不足。利用最终情景记忆挑选与目标信息相关的场外特征信息,如下所示:

$$\bar{o}_k = \text{ReLU}(W_c[X_E; m^{t-1}] + d_c), \quad (11)$$

$$X_f = [X, X_E'], \quad (12)$$

其中:  $W_c$  和  $d_c$  表示学习参数,  $X_f$  表示目标融合后的特征。从而利用外部知识丰富了残缺图像原有的特征信息,解决了缺失目标物特征信息不足的问题。

### 3.3 修复模型构建

本文所设计的残缺图像精细修复模型由三部分网络构成,即:场内外特征融合网络,粗修复网络和精细修复网络,其总体网络结构如图4所

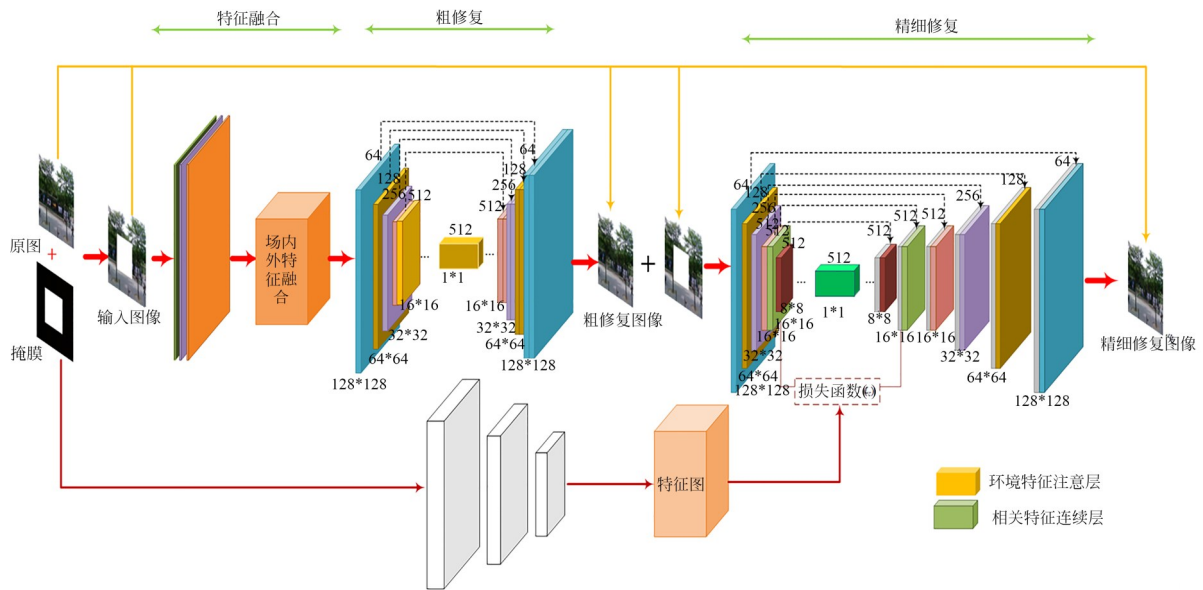


图 4 本文所提残缺图像精细修复方法的网络模型

Fig. 4 Network model of the proposed method for fine restoration of incomplete images

示。将信息残缺的图像  $I_m$  输入到场内外特征融合网络,得到一个特征信息更加丰富的输出结果,并将其输入到粗修复网络,获取粗修复图像  $I_r$ 。将待修复的图像  $I_m$  和粗修复图像  $I_r$  同时输入精细修复网络,通过提取叠加区域的有效特征信息,生成最终的精细修复图像  $I_m$ ,从而实现了对残缺图像的精细修复。下面将着重描述粗修复网络和精细修复网络的模型构建过程。

### 3.3.1 粗修复网络设计

本文所设计的粗修复网络是基于对抗神经网络策略的修复模型。它是将编码器的每一层都与解码器的对应层特征相关联。利用编码器生成待修复图像的深度特征表示,并通过解码器预测出缺失区域信息。在图像修复模型中损失要优于现有的 GAN 损失,且与重建损失函数结合时会产生一个更好的效果。基于 WGAN-GP 网络获取生成数据和原始数据之间的 Wasserstein-1 分布比较距离。其目标函数为:

$$\min_G \max_{D \in L} E_{x \sim P_r} [D(X)] - E_{\bar{x} \sim P_g} [D(\bar{X})], \quad (13)$$

其中:  $L$  表示 1-Lipschitz 函数的集合,  $P_g$  表示隐含  $\bar{x} = G(z)$  中的模型分布,  $z$  表示生成器的输入。

引入梯度惩罚约束条件,将梯度与输入掩膜相乘,其原理如下:

$$\lambda E_{\bar{x} \sim P_g} (\|\nabla_{\bar{x}} D(\bar{x}) \cdot (1 - M)\|_2 - 1)^2, \quad (14)$$

其中:  $\bar{x}$  表示从  $P_r$  和  $P_g$  采样点之间的直线采样,  $M$  表示输入掩膜的数值,缺失像素的掩膜值为 0,其它部分的掩膜值为 1,  $\lambda$  表示权重。WGAN-GP 网络中的 Wasserstein-1 距离表示为:

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} E_{(x,y) \sim \gamma} [\|x - y\|], \quad (15)$$

其中:  $\Pi(P_r, P_g)$  表示  $P_r$  和  $P_g$  采样点分布集合的  $\gamma(x, y)$  的边缘。通过对以像素为单位测量距离进行损耗评估,并利用梯度惩罚约束条件训练和优化原始 WGAN-GP 网络的生成器,使其更易学习并匹配出待修复区域的有效信息,提升粗修复网络模型的稳定性。

$$\begin{cases} n_1 = \bar{n}_1, D_{a1} = 0 \\ n_i = \frac{D_{a_i}}{D_{a_i} + D_{\max i}} \times n_{i-1} + \frac{D_{\max i}}{D_{a_i} + D_{\max i}} \times \bar{n}_i, i \in (2, k) \end{cases} \quad (18)$$

### 3.3.2 精细修复网络设计

将待修复的图像  $I_m$  和粗修复图像  $I_r$  同时输入到精细修复网络中,以促使网络更快的捕获图像中有效的特征信息,极大地提高了修复网络的修复效果。精细修复模型的网络结构与粗修复模型的相似,不同的是为了增强残缺区域的语义相关性和特征连续性,提出了一种新的相关特征连贯层,并通过特征相关性对上下语义情景结构进行保留,使其推测出来的残缺部分更加合理,如图 5 所示。

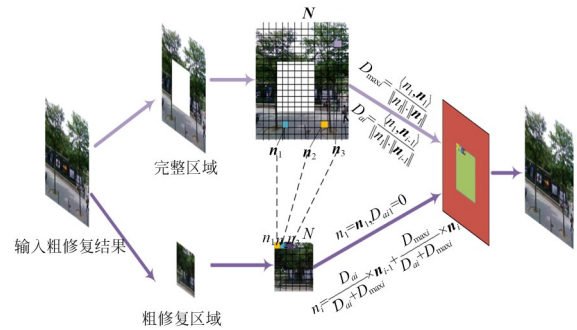


图 5 相关特征连贯原理图

Fig. 5 Coherent schematic diagram of related features

特征相关性分为搜索和生成两个阶段。对于修复区域  $N$  生成每一个补丁  $n_i (i \in (1, k), k$  表示小补丁的个数), 通过相关特征连贯层搜索与图像信息完整区域最匹配的补丁  $\bar{n}_i$ , 用于初始化  $n_i$ 。并将  $\bar{n}_i$  作为主要的特征信息, 与已经生成的补丁  $n_{i-1}$  融合, 最终反向还原出  $n_i$  作为有效的修复补丁。补丁间的相关度判断如下式所示。

$$D_{\max i} = \frac{\langle n_i, \bar{n}_i \rangle}{\|n_i\| \cdot \|\bar{n}_i\|}, \quad (16)$$

$$D_{a_i} = \frac{\langle n_i, n_{i-1} \rangle}{\|n_i\| \cdot \|n_{i-1}\|}, \quad (17)$$

其中:  $D_{a_i}$  表示这个邻近补丁之间的相似性,  $D_{\max i}$  表示最匹配的补丁  $\bar{n}_i$  和完整区域补丁  $n_i$  之间的相似性。将  $D_{a_i}$  和  $D_{\max i}$  视为生成补丁的权重, 这样每一个补丁都包含着先前补丁的信息。最终生成的补丁信息表示如下:

由于补丁的生成过程是一个迭代过程,每一个补丁在迭代的过程中均获得了更多的背景信息,且与之前生成的所有补丁( $n_1 \sim n_{i-1}$ )和  $\bar{n}_i$  都具有相关性。将  $\bar{N}$  中提取的补丁用作反卷积滤波器对  $N$  进行重建,从而获取更加逼真的修复结果。

为进一步提高图像修复的效果,本文引入特征修复识别器,通过识别器区分原始图像和已修复图像,并根据修复特征信息,计算出对抗损失函数  $D_M$  和  $D_R$ ,将  $D_M$  用于精细修复网络, $D_R$  用于粗修复网络。如式(19)所示:

$$\begin{cases} D_R = -E_{I_o}[(1 - D(I_o, I_m))^2] - E_{I_m}[D(I_m, I_o)^2] \\ D_M = -E_{I_o}[D(I_m, I_o)^2] - E_{I_m}[(1 - D(I_o, I_m))^2] \end{cases} \quad (19)$$

其中: $D$ 代表鉴别器, $E_{I_o}$ 表示所有真实取平均值的操作, $E_{I_m}$ 表示所有修复结果取平均值的操作。

### 3.4 修复模型的损失函数

为提高场内外特征融合网络检索相关场外特征的能力和优化注意机制参数,在采用改进的DMN+网络进行场内外特征融合时定义一个优化损失函数,如下式所示:

$$L_d = \sum_{f=mi}^k \max(0, \beta - S_o(x_E, m^f) + S_o(x_E, q)), \quad (20)$$

其中: $S_o(\cdot)$ 表示情景记忆和场外特征之间的匹配度, $S_R(\cdot)$ 表示场内特征和场外特征之间的匹配度。

在修复网络构建的过程中,现有的图像修复模型通常利用知觉损失提高图像修复网络的识别能力。但是,知觉损失容易误导相关语义连贯层的训练过程,本文引入一致性损失函数对现有损失函数的构建进行了改进。通过经预训练的VGG-16网络<sup>[30]</sup>提取原始图像的深度特征。并将残缺区域的原始深度特征设定为相关语义连贯层及其编码器对应层的目标,并计算出训练网络的一致性损失函数  $L_c$ ,如下式所示:

$$L_c = \sum_{z \in M} \left\| W(I_{ir})_z - \phi_m(I_o)_z \right\|_2^2 + \left\| W_d(I_{ir})_z - \phi_m(I_o)_z \right\|_2^2, \quad (21)$$

其中: $\phi_m$ 表示经预训练的VGG-16网络参数。 $W(\cdot)$ 表示编码器中相关特征连贯层的深度特征, $W_d(\cdot)$ 是解码器中相关特征连贯层对应层的

深度特征。

将 Wasserstein-1 距离作为判断条件,构建出修复网络的损失函数如下式所示:

$$L_r = \left\| I_r - I_o \right\|_1 + \left\| I_m - I_o \right\|_1. \quad (22)$$

将式(19)~式(22)所构建的特征融合优化损失函数、训练网络一致性损失函数、修复网络损失函数和粗修复网络对抗损失函数进行加权平均,获取所提残缺图像精细修复模型的总体损失函数  $L_g$ ,如下式所示:

$$L_g = \alpha_d L_d + \alpha_c L_c + \alpha_r L_r + \alpha_R D_R, \quad (23)$$

其中: $\alpha_d, \alpha_c, \alpha_r, \alpha_R$ 分别为优化损失、一致性损失、修复损失、对抗损失的权衡参数。

## 4 实验结果及分析

为验证所提残缺图像精细修复模型的有效性,分别基于 Places2<sup>[31]</sup>, Real-world Underwater Image Enhancement<sup>[32]</sup>和 Underwater Target 数据集与现有的方法进行了定性和定量的对比实验,具体实验结果及分析如下。

### 4.1 实验设置

本文设计的实验基于 Places2, Real-world Underwater Image Enhancement (RUIE) 和 Underwater Target 三个数据集。在实验中不使用任何数据标签。为了模拟图像缺失的现象,本文将采用方形掩模对图像目标区域进行遮掩。同时对三个数据集的原始数据进行训练。学习率设置为  $2 \times 10^{-4}$  和  $\beta = 0.05$ 。权衡参数设置为  $\alpha_d = 0.1, \alpha_c = 0.1, \alpha_r = 1, \alpha_R = 0.001$ 。将本文所提 EFIF 算法与 SH (Shift-net)<sup>[3]</sup>, GLCI<sup>[15]</sup>, CSA<sup>[22]</sup>三个代表性算法模型进行对比。实验的硬件配置为:CPU为 Intel(R) Core(TM) i7-8700K@3.70 GHz, GPU为 RTX 2080 Ti。内存为 64 G。运行环境为 Python3.7,模型采用 PyTorch 库编写,操作系统为 Ubuntu-16.04。

### 4.2 定性评价实验

#### 4.2.1 Places2 数据集实验结果

Places2 数据集是由麻省理工大学负责维护的场景图像数据集。数据集涉及 400 多个场景,共包含 1 千万张图片。该数据集是图像修复领域广为应用的数据集之一。本文从数据集中选取大量的建筑、风景图片,用于训练和实验。

在 Places2 数据集中的仿真结果如图 6 所示。最左边的一列为经过掩模处理后的输入图像,最右边一列是原始图像,中间分别为 SH, GLCI, CSA 和 EFIF 修复模型的修复结果。实验(a)~(c)为建筑

图像的修复,实验(d)~(e)为风景图像修复。从实验结果可以得出本文提出的图像修复结果比其他修复模型的修复结果更能有效的修复目标的特征。同时在纹理重建方面也更加合理。



图 6 现有修复模型与本文修复模型在 Place2 数据集中比较,从右到左依次为输入图像,SH、GLCI、CSA、粗修复、EFIF 及原图

Fig. 6 Comparison of the existing restoration model and the restoration model in this paper in the Place2 database, from right to left, the input images, SH, GLCI, CSA, roughly repaired image, EFIF, and the original image

#### 4.2.2 RUIE 数据集实验结果

RUIE 数据集是由大连理工大学一立命馆大学国际信息与软件学院提出的一个针对水下图像研究的数据集,具有图像数据量大,图像场景、色彩多样,检测目标丰富等特点。其常用于水下目标的检测与识别、水下图像的增强与复原。本文从 RUIE 数据集选取大量的水下雕塑,汽车残骸等进行仿真实验。与 SH, GLCI 和 CSA 图像

修复模型作对比,其仿真结果如图 7 所示。

从仿真结果来看实验(a)、实验(b)、实验(c)表示轻微模糊的场景实验。实验(d)、实验(e)和实验(f)表示严重模糊的场景实验。实验(b)中,SH 模型的修复结果存在着目标丢失的现象。SH, GLCI 和 CSA 修复模型在实验(a)、实验(b)、实验(e)、实验(f)的修复中均存在着修复区域模糊,并不能有效地修复缺失区域的纹理。实验

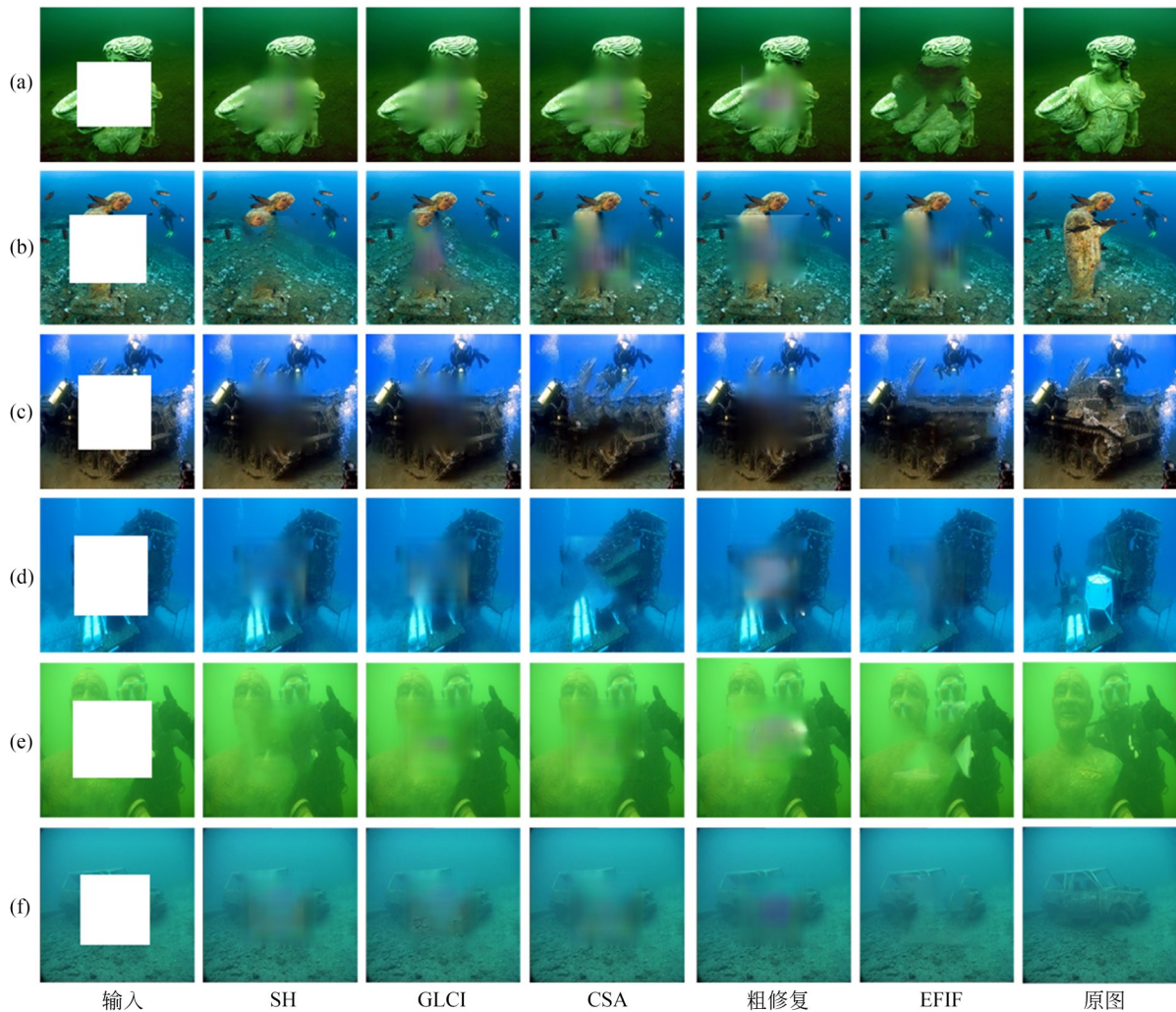


图 7 现有修复模型与本文修复模型在 RUIE 数据集中比较,从右到左依次为输入图像,SH、GLCI、CSA、粗修复、EFIF,及原图

Fig. 7 Comparison of the existing restoration model and the restoration model in this paper in the RUIE database, from right to left, the input images, SH, GLCI, CSA, roughly repaired image, EFIF, and the original image

(b)、实验(d)、实验(e)对应粗修复图像的修复区域纹理模糊现象。本文所提出的修复结果能够有效修复缺失区域特征,且能够生成合理的图像纹理。如实验(c)、实验(e)所示修复的结果与原图十分接近。

#### 4. 2. 3 Underwater Target 数据集实验结果

Underwater Target 数据集是本文研究团队针对水下图像所建立的数据集。主要用于水下复杂场景中特征提取、目标检测与识别、场景理解,为自主式水下航行器(Autonomous Underwater Vehicle, AUV)实现自主航行能力、目标识别与跟踪、自主规划路线和协同控制提供重要保障。在水下图像处理领域的研究中有着重要意

义。其中包括鱼雷、潜艇、蛙人、AUV 等类别。本文从 Underwater Target 数据集中选取大量的鱼雷、潜艇、AUV 图像进行实验,与 SH、GLCI 和 CSA 图像修复模型对比。如图 8 所示。从仿真结果来看实验(a)、实验(b)为不同场景下的鱼雷修复。其中实验(a)表清晰的场景,实验(b)表模糊的场景,实验(c)表示模糊的场景 AUV 的修复实验。实验(d)、实验(e)分别表示清晰、模糊场景下的潜艇的修复。实验(b)中,SH, GLCI 模型修复的结果存在特征丢失现象。实验(a)、实验(d)中,粗修复图像存在修复区域纹理模糊的现象。本文模型能够更好的修复残缺区域,同时注重修复区域的纹理,如实验(b)、实验(c)本文修

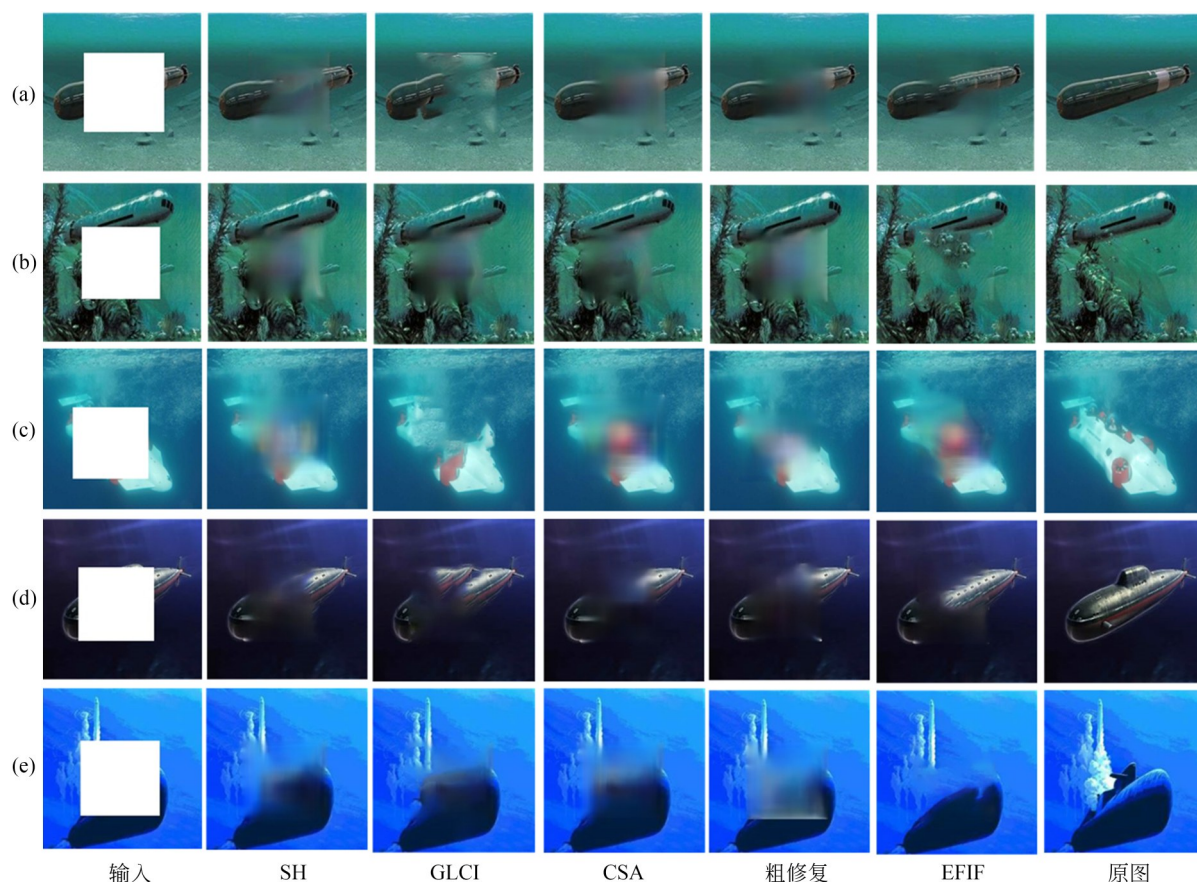


图 8 现有修复模型与本文修复模型在 Underwater Target 数据集中比较, 从右到左依次为输入图像, SH、GLCI、CSA、粗修复、EFIF, 及原图

Fig. 8 Comparison of the existing restoration model and the restoration model in this paper in the Underwater Target database, from right to left, the input images, SH, GLCI, CSA, roughly repaired image, EFIF, and the original image

复结果接近于原图像。从而说明了本文模型更侧重于结构和纹理的修复能力。

### 4.3 定量评价实验

#### 4.3.1 定量评价指标

由于个体的差异、喜好等主观因素的影响, 对实验结果的评价会在一定程度上存在片面性。为了获取更准确的修复结果的质量评价。在上述式(22)修复网络的损失函数单一定量评价指标的基础上, 本文引入峰值信噪比 PSNR<sup>[33]</sup>和结构相似性 SSIM<sup>[34]</sup>两个评价指标, 对修复结果进行更加客观的定量评价。峰值信噪比是通过两张图像对应的像素点的误差评价图像的质量。其值越大表示图像的修复结果越好, 如下式所示:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|x(i, j) - y(i, j)\|^2, \quad (24)$$

$$PSNR = 10 \cdot \log_{10} \left( \frac{MAX_x^2}{MSE} \right), \quad (25)$$

其中:  $m, n$  表示图像尺寸的大小,  $MSE$  表示张图像之间的均方差。  $MAX_x^2$  表示图像中所取的最大值。

SSIM 指标从图像的亮度、对比度以及结构信息衡量图像的结构相似性, 从而评价图像的失真程度, 其值越大说明失真越小, 修复图像越接近于原始图像, 如下式所示:

$$SSIM = \frac{(2\sigma_x\sigma_y + C_1)(2\delta_{xy} + C_2)}{(\sigma_x^2 + \sigma_y^2 + C_1)(\delta_x^2 + \delta_y^2 + C_2)}, \quad (26)$$

其中:  $\sigma_x$  和  $\sigma_y$  分别为原图像和修复图像的平均值,  $\sigma_x^2$  和  $\sigma_y^2$  分别表示原图像和修复图像的方差,  $\delta_{xy}$  表示协方差,  $C_1, C_2$  为用来维持稳定的常数。

#### 4.3.2 定量评价实验结果

本文使用 PSNR, SSIM 和  $L_1$  损失作为定量

评价的衡量指标,其中 PSNR 和 SSIM 能客观的反映模型修复残缺图像的性能。 $L_r$ 用来测量修复图像和真实图像间的距离。本文修复模型与对比修复模型在 Place2、RUIE 和 Underwater Target 数据集的 PSNR,SSIM 和  $L_r$  损失值,如表 1 所示。从表中可以看出本文修复模型在 Place2、RUIE 和 Underwater Target 数据集取得了最优结果,PSNR 值和 SSIM 值是最大的, $L_r$  损失是最低的。本文模型图像修复结果的 PSNR 值最高为 35.98,相比 CSA 算法提高了 2.17%。SSIM 值最高为 0.983,相比提高了 1.08%。 $L_r$  损失最低为 0.71,相对降低了 6.96%。

Underwater Target 数据集主要用于模糊、复杂水下场景中危险性目标的识别与重构。其特征提取难度极大,对修复算法有效性的挑战最为显著,本文将图 8 每个仿真结果的 PSNR、SSIM 用柱状图的形式直观的展现出来,用以更加客观的展现所提修复模型面对复杂环境的修复效果,如图 9 所示 Underwater Targe 数据集中本文所提算法 PSNR 最高为 26.75,相比 CSA 算法提高了 1.44。SSIM 最高为 0.935,相比 CSA 算法提高了 0.006。从而可以得出所提算法的修复结果一致优于现有对比模型。

表 1 Place2、RUIE 和 Underwater Target 数据集各修复结果的客观数据

Tab. 1 Objective data for restoration results in Place2, RUIE and Underwater Target database

数据集	掩码率	PSNR					SSIM					$L_r$ (%)			
		SH	GLCI	CSA	粗修 复	EFIF	SH	GL- CL	CSA	粗修 复	EFIF	SH	GLCI	CSA	EFIF
Place2	10%–20%	31.34	31.27	33.36	31.03	<b>34.87</b>	0.906	0.911	<b>0.989</b>	0.905	0.987	1.02	1.00	0.78	<b>0.71</b>
	20%–30%	30.24	31.79	32.85	30.08	<b>33.12</b>	0.893	0.897	0.982	0.894	<b>0.983</b>	1.57	1.31	0.98	<b>0.94</b>
	30%–40%	21.96	24.17	25.16	21.35	<b>24.51</b>	0.886	0.891	0.926	0.887	<b>0.931</b>	2.74	2.36	1.29	<b>1.23</b>
	40%–50%	20.39	22.18	24.51	20.40	<b>24.84</b>	0.832	0.853	0.873	0.829	<b>0.894</b>	3.67	3.24	2.85	<b>2.46</b>
RUIE	10%–20%	31.45	33.72	35.84	31.53	<b>35.98</b>	0.929	0.936	0.953	0.925	<b>0.978</b>	1.55	1.47	1.18	<b>1.06</b>
	20%–30%	30.75	33.46	34.97	28.95	<b>35.73</b>	0.919	0.925	0.942	0.917	<b>0.971</b>	1.94	1.73	1.56	<b>1.47</b>
	30%–40%	27.84	27.76	28.06	27.46	<b>30.93</b>	0.897	0.902	0.916	0.898	<b>0.923</b>	2.95	2.86	2.72	<b>2.55</b>
	40%–50%	21.97	23.19	23.97	21.99	<b>24.45</b>	0.846	0.874	0.899	0.841	<b>0.905</b>	3.77	3.78	3.24	<b>3.16</b>
Underwater Target	10%–20%	24.83	25.01	25.42	24.67	<b>27.01</b>	0.913	0.928	0.936	0.915	<b>0.949</b>	1.65	1.44	1.11	<b>1.09</b>
	20%–30%	24.36	24.72	25.31	24.21	<b>26.75</b>	0.904	0.916	0.929	0.899	<b>0.935</b>	2.04	1.96	1.54	<b>1.44</b>
	30%–40%	20.48	21.14	22.07	20.83	<b>22.35</b>	0.884	0.890	0.907	0.881	<b>0.916</b>	3.27	2.98	2.73	<b>2.49</b>
	40%–50%	15.39	15.04	16.42	15.46	<b>16.91</b>	0.862	0.876	0.890	0.868	<b>0.891</b>	4.41	4.04	3.27	<b>3.03</b>

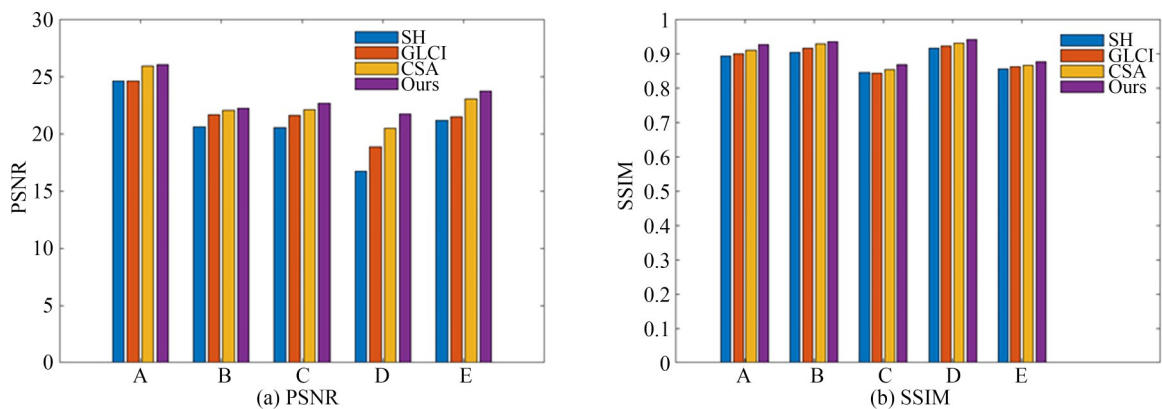


图 9 Underwater Target 数据集中本文修复模型和对比修复模型修复结果 PSNR、SSIM 值的柱状图

Fig. 9 Histograms of PSNR and SSIM values for restoration results from this restoration model and comparative restoration models in the Underwater Targe dataset.

## 5 结 论

本文提出一种基于场内外特征融合的残缺图像精细修复方法。首先,通过改进的DMN+算法实现与目标相关外部知识和图像特征的融合;其次,将融合后的特征输入到带有梯度惩罚约束条件的对抗生成网络生成粗修复图像;最后,利用相关特征连贯层对粗修复图进一步优化,获取纹理清晰、特征连贯的精细修复图像。

### 参考文献:

- [1] 李志丹, 苟慧玲, 程吉祥, 等. 结合梯度特征与色彩一致性的图像修复[J]. 光学精密工程, 2019, 27(1): 251-259.  
LI ZH D, GOU H L, CHENG J X, *et al.* Image inpainting using gradient features and color consistency[J]. *Opt. Precision Eng.*, 2019, 27(1): 251-259. (in Chinese)
- [2] V V V, A S R, I M V, *et al.* Video inpainting of complex scenes based on local statistical model[J]. *Electronic Imaging*, 2016, 2016(15): 1-6.
- [3] YAN Z Y, LI X M, LI M, *et al.* *Shift-net: Image Inpainting Via Deep Feature Rearrangement* [M]. Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [4] XIONG W, YU J H, LIN Z, *et al.* Foreground-aware image inpainting[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 15-20, 2019, Long Beach, CA, USA. *IEEE*, 2019: 5833-5841.
- [5] NAZERI K, NG E, JOSEPH T, *et al.* EdgeConnect: generative image inpainting with adversarial edge learning [EB/OL]. 2019: *arXiv: 1901.00212* [cs. CV]. <https://arxiv.org/abs/1901.00212>
- [6] EFROS A A, LEUNG T K. Texture synthesis by non-parametric sampling [C]. *Proceedings of the Seventh IEEE International Conference on Computer Vision*. September 20-27, 1999, Kerkyra, Greece. *IEEE*, 1999: 1033-1038.
- [7] BERTALMIO M, VESE L, SAPIRO G, *et al.* Simultaneous structure and texture image inpainting [J]. *IEEE Transactions on Image Processing*, 2003, 12(8): 882-889.
- [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] LE MEUR O, GAUTIER J, GUILLEMOT C. Exemplar-based inpainting based on local geometry [C]. 2011 *18th IEEE International Conference on Image Processing*. September 11-14, 2011, Brussels, Belgium. *IEEE*, 2011: 3401-3404.
- [10] 金炜, 王文龙, 符冉迪, 等. 联合块匹配与稀疏表示的卫星云图修复[J]. 光学精密工程, 2014, 22(7): 1886-1895.  
JIN W, WANG W L, FU R D, *et al.* Satellite cloud image inpainting based on patch matching and sparse representation [J]. *Opt. Precision Eng.*, 2014, 22(7): 1886-1895. (in Chinese)
- [11] CHAN T F, SHEN J H. Nontexture inpainting by curvature-driven diffusions [J]. *Journal of Visual Communication and Image Representation*, 2001, 12(4): 436-449.
- [12] CHAN T F, SHEN J H, VESE L. Variational PDE models in image processing [R]. Defense Technical Information Center, 2002.
- [13] 朱为, 李国辉. 基于自动结构延伸的图像修补方法[J]. 自动化学报, 2009, 35(8): 1041-1047.  
ZHU W, LI G H. Image completion based on automatic structure propagation [J]. *Acta Automatica Sinica*, 2009, 35(8): 1041-1047. (in Chinese)
- [14] 李志丹, 苟慧玲, 程吉祥, 等. 结合梯度特征与色彩一致性的图像修复[J]. 光学精密工程, 2019, 27(1): 251-259.  
LI ZH D, GOU H L, CHENG J X, *et al.* Image inpainting using gradient features and color consistency [J]. *Opt. Precision Eng.*, 2019, 27(1): 251-259. (in Chinese)
- [15] IIZUKA S, SIMO-SERRA E, ISHIKAWA H.

- Globally and locally consistent image completion [J]. *ACM Transactions on Graphics*, 2017, 36(4): 1-14.
- [16] DEKEL T, GAN C, KRISHNAN D, *et al.* Sparse, smart contours to represent and edit images [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. *IEEE*, 2018: 3511-3520.
- [17] GOODFELLOW J, POUGET-ABADIE J, MIRZA M, *et al.* Generative adversarial nets [C]. In *Advances in Neural Information Processing Systems, Montreal, Canada, AIPS*, 2014: 2672-2680.
- [18] GU J X, ZHAO H D, LIN Z, *et al.* Scene graph generation with external knowledge and image reconstruction [C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 15-20, 2019, Long Beach, CA, USA. *IEEE*, 2019: 1969-1978.
- [19] SAINBAYAR S, SZLAM A, JASON W, *et al.* End-To-End memory networks [C]. *International Conference on Neural Information Processing Systems*, 2017: 2440-2448.
- [20] CAIMING X, STEPHEN M, RICHARD S, *et al.* Dynamic memory networks for visual and textual question answering [C]. *International Conference on Machine Learning*, 2016: 2397-2406.
- [21] ZHENG C X, CHAM T J, CAI J F. Pluralistic image completion [C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 15-20, 2019, Long Beach, CA, USA. *IEEE*, 2019: 1438-1447.
- [22] LIU H Y, JIANG B, XIAO Y, *et al.* Coherent semantic attention for image inpainting [C]. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 27 - November 2, 2019, Seoul, Korea (South). *IEEE*, 2019: 4169-4178.
- [23] YU J H, LIN Z, YANG J M, *et al.* Generative image inpainting with contextual attention [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. *IEEE*, 2018: 5505-5514.
- [24] WU Q, SHEN C H, WANG P, *et al.* Image captioning and visual question answering based on attributes and external knowledge [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(6): 1367-1381.
- [25] KUMAR A, IRSOY O, ONDRUSKA P, *et al.* Ask me anything: dynamic memory networks for natural language processing [C]. *International Conference on Machine Learning, New York*, 2015, pp. 2068-2078.
- [26] BARNES C E, SHECHTMAN A, GOLDMAN D B, *et al.* Patchmatch: a randomized correspondence algorithm for structural image editing [J]. *Transactions on Graphics*, 2009, 28(3): 1-11.
- [27] PATHAK D, KRÄHENBÜHL P, DONAHUE J, *et al.* Context encoders: feature learning by inpainting [C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 27-30, 2016, Las Vegas, NV, USA. *IEEE*, 2016: 2536-2544.
- [28] YANG C, LU X, LIN Z, *et al.* High-resolution image inpainting using multi-scale neural patch synthesis [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017, Honolulu, HI, USA. *IEEE*, 2017: 4076-4084.
- [29] RUSSAKOVSKY O, DENG J, SU H, *et al.* ImageNet large scale visual recognition challenge [J]. *International Journal of Computer Vision*, 2015, 115(3): 211-252.
- [30] KAMALYAN A G, SIMONYAN V V. On the number of solutions of a certain type of one-dimensional pseudodifferential equations in the Sobolev-Slobodetski space [J]. *Journal of Shellfish Research*, 2002, 21(1): 201-210.
- [31] LIU Z W, LUO P, WANG X G, *et al.* Deep learning face attributes in the wild [C]. 2015 *IEEE International Conference on Computer Vision (ICCV)*. December 7-13, 2015, Santiago, Chile. *IEEE*, 2015: 3730-3738.
- [32] LIU R S, FAN X, ZHU M, *et al.* Real-world underwater enhancement: challenges, benchmarks, and solutions under natural light [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(12): 4861-4875.
- [33] WANG Y J, LI J H, LU Y, *et al.* Image quality evaluation based on image weighted separating block peak signal to noise ratio [C]. *International Conference on Neural Networks and Signal Pro-*

cessing, 2003. *Proceedings of the 2003. December 14-17, 2003, Nanjing, China. IEEE*, 2003: 994-997.

- [34] TUNCEL E, FERHATOSMANOGLU H, ROSE K. VQ-index: an index structure for similar-

ity searching in multimedia databases[C]. *Proceedings of the tenth ACM international conference on Multimedia-MULTIMEDIA '02. December 1-6, 2002. Juan-les-Pins, France. New York: ACM Press*, 2002: 543-552.

#### 作者简介:



徐 涛(1981—),男,河南新乡人,博士,讲师,硕士研究生导师,2017年于北京工业大学获得博士学位,主要从事智能机器人环境感知、机器视觉等方面的研究。E-mail: xutao@hist.edu.cn



周纪勇(1995—),男,河南漯河人,硕士研究生,2019年于绥化学院获得学士学位,现于河南科技学院攻读硕士研究生,研究方向为图像处理。E-mail: zhoujiyong2020@126.com