E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TN919.8　　**文献标识码**:A　　**文章编号**:1006-8961(2024)07-1948-12

论文引用格式:Liu J X, Zhou Y, Lin K, Yin H B and Tang X H. 2024. Progressive iteration network for hole filling in virtual view rendering. Journal of Image and Graphics, 29(07):1948-1959(刘家希, 周洋, 林坤, 殷海兵, 唐向宏. 2024. 面向虚拟视点绘制空洞填充的渐进式迭代网络. 中国图象图形学报, 29(07):1948-1959)[DOI:10.11834/jig.230290]

面向虚拟视点绘制空洞填充的渐进式迭代网络

刘家希,周洋*,林坤,殷海兵,唐向宏

杭州电子科技大学通信工程学院,杭州　310018

摘　要:**目的**　基于深度图像的绘制(depth image based rendering,DIBR)是合成虚拟视点图像的关键技术,但在绘制过程中虚拟视图会出现裂纹和空洞问题。针对传统算法导致大面积空洞区域像素混叠和模糊的问题,将深度学习模型应用于虚拟视点绘制空洞填充领域,提出了面向虚拟视点绘制空洞填充的渐进式迭代网络。**方法**　首先,使用部分卷积对大面积空洞进行渐进修复。然后采用U-Net网络作为主干对空洞区域进行编解码操作,同时嵌入知识一致注意力模块加强网络对有效特征的利用。接着通过加权合并方法来融合每次渐进式迭代生成的特征图,保护早期特征不被破坏。最后结合上下文特征传播损失提高网络匹配过程中的鲁棒性。**结果**　在微软实验室提供的2个多视点3D(three-dimension)视频序列以及4个3D-HEVC(3D high efficiency video coding)序列上进行定量与定性评估实验,以峰值信噪比(peak signal-to-noise ratio,PSNR)和结构相似性(structural similarity,SSIM)作为指标。实验结果表明,本文算法在主观和客观上均优于已有方法。相比于性能第2的模型,在Ballet、Breakdancers、Lovebird1和Poznan_Street数据集上,本文算法的PSNR提升了1.302 dB、1.728 dB、0.068 dB和0.766 dB,SSIM提升了0.007、0.002、0.002和0.033;在Newspaper和Kendo数据集中,PSNR提升了0.418 dB和0.793 dB,SSIM提升了0.011和0.007。同时进行消融实验验证了本文方法的有效性。**结论**　本文提出的渐进式迭代网络模型,解决了虚拟视点绘制空洞填充领域中传统算法过程烦琐和前景纹理渗透严重的问题,取得了极具竞争力的填充结果。

关键词:虚拟视点绘制;空洞填充;注意力;特征提取;多视点视频加深度

Progressive iteration network for hole filling in virtual view rendering

Liu Jiaxi, Zhou Yang*, Lin Kun, Yin Haibing, Tang Xianghong

School of Communication Engineering, Hangzhou Dianzi University, Hangzhou　310018, China

Abstract: **Objective** Depth image-based rendering (DIBR) makes full use of the depth information in a reference image and can combine color image and depth information organically, which is faster and less complex than the general rendering method. Therefore, DIBR is selected by ISO as the primary virtual view rendering technology in 3D multimedia video. The principal challenge associated with virtual view rendering technology is the 3D warping of the reference view, which leads to exposure of the background that was previously obstructed by the foreground. As a result, certain areas appear as holes in the virtual view due to the absence of pixel values. The search for an effective solution to address missing regions in the rendered view image is a critical challenge in virtual view rendering technology. The traditional algorithms mainly fill the holes based on the space-domain consistency and time-domain consistency methods. Filtering can effectively remove

收稿日期:2023-05-30;**修回日期**:2023-10-17;**预印本日期**:2023-10-23

*通信作者:周洋　zhouyang@hdu.edu.cn

基金项目:浙江省自然科学基金项目(LY21F020021);浙江省尖兵领雁计划项目(2022C01068)

Supported by: Natural Science Foundation of Zhejiang Province, China(LY21F020021); "Pioneer" and "Leading Goose" R&D Program of Zhejiang Province(2022C01068)

the cracks and some of the holes but cannot handle the large-area holes. The patch-based method can fill large-area holes, but the process is tedious, the amount of data is too large, and the accuracy of searching for the best matching patch is not high, which may lead to the texture belonging to the foreground being incorrectly filled to the hole area belonging to the background. Based on the time-domain consistency method, a model is developed to reconstruct the vacant part of the background using various models, and the foreground part is repositioned to the virtual viewpoint location to reduce the computational complexity and increase the adaptability to the scene. However, the moving camera scene contains both stationary and moving objects, which easily causes some parts of the foreground to be modeled as the background, resulting in the mixing of foreground and background pixels. Therefore, a deep learning model is applied to the field of hole filling in virtual view rendering, and a progressive iterative network for hole filling in virtual view rendering is proposed to address the problem of traditional algorithms leading to pixel blending and blurring in large hole regions. **Method** In this study, a progressive iterative network based on convolutional neural network is built. The network model mainly consists of a knowledge consistent attention module, a contextual feature propagation loss module, and a weighted merging module. First, partial convolutions are used in the initial stage of the network for progressive repair of large area holes. The partial convolutions are operated using only the valid pixels in the hole region, and the updated masks are retained throughout the iterations until they are reduced and updated in the next iteration, which is beneficial to the extraction of shallow valid features. Then, the U-Net network is used as the backbone to codify and decode the empty regions and cascade the shallow and deep information by introducing skip connections to tackle the problem of missing information. To select effective features in the network, we embed a knowledge consistent attention module. One benefit of this attention module is that it measures the attention score by weighing the current score with the score obtained from the previous iteration, which establishes the correlation between the front and back frame patches and effectively avoids the problem of foreground and background pixel blending in the traditional algorithm. The contextual feature propagation loss module is also used in a progressive iterative network with an attention module. This module plays a complementary role to the knowledge consistent attention module, reducing the difference between the reconstructed images in the encoder and decoder and enhancing the robustness of the network matching process. In addition, it allows for the creation of semantically consistent patches to fill in background holes by utilizing auxiliary images as guidance. Furthermore, we employ a pre-trained Visual Geometry Group (VGG-16) feature extractor to facilitate the joint guidance of our model using L1 loss, perceptual loss, style loss, and smoothing loss, ultimately enhancing the resemblance between reference and target views. Lastly, the feature maps produced in each successive iteration are integrated via a weighted merging approach. This process involves the development of an adaptive map through the learning process. Specifically, through the concatenation of soft weight maps and the output feature maps of adaptive merging, the method provides an adaptive map that preserves original feature information with soft weight map assistance and protects early features from corruption, thus preventing gradient erosion. **Result** The experiments were quantitatively and qualitatively evaluated on multi-view 3D video sequences provided by Microsoft Labs and four 3D high efficiency video coding (3D-HEVC) sequences. Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) metrics were used to measure the algorithm's performance, and a set of hole masks suitable for virtual view rendering were collected for training. Our experimental results demonstrate that our model yields the most reasonable images in terms of subjective perceptual quality. Furthermore, compared with the model with the second-highest performance, our model outperforms in terms of PSNR and SSIM, improving 1.302 dB, 1.728 dB, 0.068 dB, and 0.766 dB, and 0.007, 0.002, 0.002, 0.002, and 0.033 on the Ballet, Breakdancers, Lovebird1, and Poznan_Street datasets, respectively. Meanwhile, compared with the deep learning model, the PSNR and SSIM increased by 0.418 dB and 0.793 dB, and 0.011 and 0.007, respectively, in the Newspaper and Kendo datasets. In addition, we conducted a series of ablation experiments to verify the effectiveness of each module in our model, including the knowledge consistent attention module, the contextual feature propagation loss module, the weighted merging module, and the number of iterations. **Conclusion** In this study, we apply deep learning to the field of hole filling in virtual view rendering. Our proposed progressive iterative network model was validated through experimental demonstration. We observed that our model performs exceptionally well in terms of avoiding tedious processes and minimizing foreground texture infiltration, ultimately leading to superior filling outcomes. However, our model exhibits some limitations. While it can focus on effective texture features, its overall efficiency still

requires further improvement. Moreover, depth maps associated with 3D video sequences can be utilized as a guide, enabling the convolutional neural network to comprehend more intricate structural aspects and enhancing the model's overall performance. In future research, we may consider merging frame interpolation and inpainting techniques to concentrate on the motion-related information of objects over time.

Key words: virtual view rendering; hole-filling; attention; feature extraction; multi-view video plus depth

0 引 言

自由视点视频、立体电视和虚拟现实等新兴的三维(three-dimension,3D)多媒体视觉服务能给用户带来沉浸式和交互式的视觉体验,越来越受到人们的关注和喜爱(Tanimoto 等,2011;Ye 等,2013)。但是表征交互式3D视频需要大量的视点信息,由于采集成本和带宽的限制,在实际应用中只能采集和传输有限个视点的场景。目前,多视点加深度(multi-view plus depth,MVD)编码格式是压缩3D视频和自由视点视频(Luo 和 Zhu,2017)的主流格式(Lin 等,2018),它在解码端通过基于深度图像的绘制(depth image based rendering,DIBR)(Zhu 和 Li,2016)技术绘制出需要的虚拟视点来弥补视点数目的不足。然而在绘制虚拟视点时,不同视点间由于存在前后景遮挡等问题会导致绘制图像中存在空洞、裂纹等缺失区域,这严重影响了虚拟视图的视觉质量。因此,如何有效解决绘制视点图像中的缺失区域问题是虚拟视点绘制技术中的困难和挑战(王旭 等,2023)。

传统的合成视点空洞填充方法主要分为两类:基于空域一致性的填充和基于时域一致性的填充。但这些方法通常依赖于深度图的准确性。Liu 等人(2016)提出一种基于聚类的视点交叉滤波来优化解码端的深度图质量。Chen 等人(2020)设计了自适应多模态残差网络来提升绘制端的深度图质量。目前,基于空域一致性的填充方法主要包括使用滤波器和基于块(补丁)匹配的方法。Lee 和 Effendi(2011)采用自适应边缘定向平滑滤波器对深度图进行平滑处理。Zhu 等人(2019)使用形态学算子检测前景边缘,并应用非对称高斯滤波器平滑过渡区域。但这类方法面向大面积空洞的填充时会产生模糊现象,导致合成图像出现几何失真。Criminisi 等人(2004)针对大面积空洞先提出了一种等照度线驱动的匹配优先权计算方法,然后在源区域搜索匹配块填充到空洞中。Ahn 和 Kim(2013)通过在数据项中使用鲁棒结构张量和新的置信项来填充空洞区域。Wang(2016)根据图像像素的校准参数确定图像绘制的方向,再以最小 RGB(red green blue)之和为参考标准的匹配块进行空洞填充。梁海涛等人(2019)通过在优先级计算和最佳匹配块搜索过程中增加深度信息提高空洞填充算法的精确度。基于补丁的方法从单帧信息中搜索相似补丁进行填充,容易赋予前背景相同的权重,会误引入前景纹理来填充背景空洞。基于时域一致性的方法提出利用背景模型来构建背景的空洞区域。Sun 等人(2012)基于可切换高斯模型的在线背景方法,以降低计算复杂度。Yao 等人(2014)使用高斯混合模型(Gaussian mixture model,GMM)从几个连续的视频帧和深度图离线构建稳定的视频背景。Luo 和 Zhu(2017)通过去除深度图的前景对象来构建背景,并用重建的背景图来填补虚拟视图。Han 等人(2018)使用多阈值 Otsu 方法将深度图像分割成多层并进行分层三维变形。Smirnov 等人(2019)采用阈值分割提取前景目标对背景层进行补偿。Luo 等人(2020)在去除前景对象的基础上改进了背景重建的方法。然而,当场景包含静止和运动物体,容易造成部分前景被建模为背景,因此基于时域一致性的方法会导致前背景像素混叠以及伪影现象。综上,基于补丁匹配的方法和基于时域一致性的填充方法都存在难以克服的缺点,且计算复杂度高。

现今,深度学习网络在视频图像处理领域表现优异。Yu 等人(2018)利用上下文注意力机制增强空间特征提取能力,但没有考虑时间信息。Li 等人(2020)专注于时域信息的捕捉,但未强化特征融合。Quan 等人(2022)基于编—解码器网络进行三阶段修复,但缺少对特征提取的约束性。鉴于上述问题,本文将基于 U-Net 的网络应用于虚拟视点绘制中,采用端到端的方式对绘制空洞进行填充。首先使用部分卷积代替普通卷积缩小绘制生成的大空洞面积,然后将知识一致注意力模块(knowledge consistent attention,KCA)和上下文特征传播损失(contex-

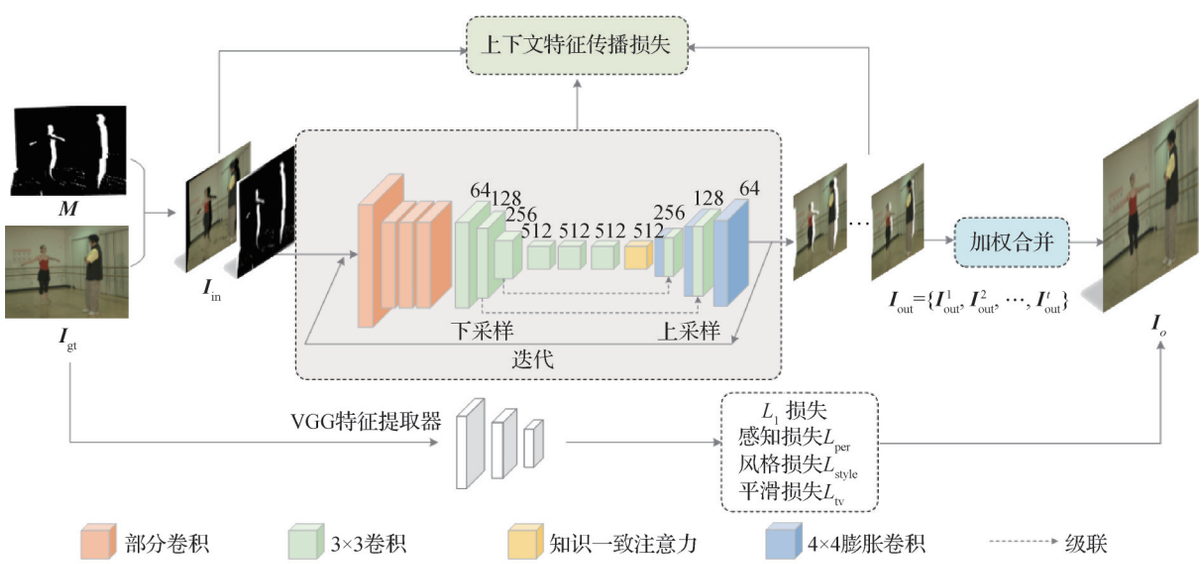
tual feature propagation loss，CFP loss）嵌入 U-Net 网络架构中解决传统方法中基于样本块填充的像素混叠问题，最后通过加权合并方法融合每次渐进式迭代生成的绘制特征图。本文贡献主要概括为以下 3 点：1）将卷积神经网络应用到虚拟视点绘制空洞填充领域，并制作了面向虚拟视点绘制的空洞掩膜数据集；2）基于 U-Net 网络架构，设计了新的上下文特征传播损失模块，缩小重建图像在编码器和解码器之间的差异；3）提出加权合并方法来融合迭代生成的特征图，保护网络早期的特征信息并有效克服梯度消失。

1　渐进式迭代网络模型

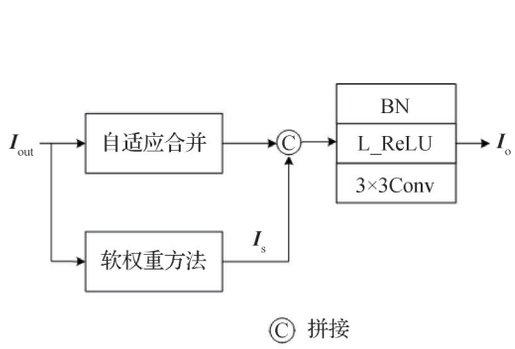
本文设计了一种面向虚拟视点绘制空洞填充的渐进式迭代网络（progressive iteration network，PINet）。模型包括加权合并模块、上下文特征传播损失模块和知识一致注意力模块，网络的整体架构如图 1（a）所示。本文网络模型创新点在于：1）在 RFR（recurrent feature reasoning）（Li 等，2020）网络中使用上下文特征传播损失，增强对匹配内容的约束；2）在 RFR 网络自适应合并的基础上，提出软权重方法进行辅助，即本文的加权合并模块。

1.1　网络总体架构

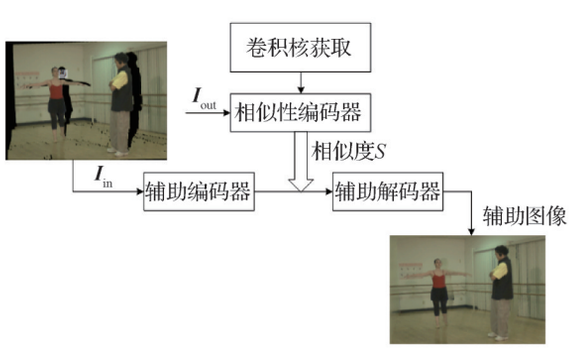
首先，在网络初始阶段，引入 4 个 7 × 7 部分卷积（Li 等，2020）进行局部空洞识别，同时使用批归一化和激活函数加速模型收敛。部分卷积只使用空洞区域中的有效像素进行运算，更新后的掩膜在整个迭代过程中被保留，直到下一次迭代时被缩小更新，有利于浅层有效特征的提取。编码器先通过 3 个步长为 2 的 3 × 3 卷积进行下采样来捕获浅层语义信息；再使用 3 个步长为 1 的 3 × 3 卷积保持通道数不变来



（a）主要框架图



（b）加权合并



（c）上下文特征传播损失

图 1　渐进式迭代网络的整体架构

Fig. 1　Architecture of progressive iteration network

（（a）main frame diagram；（b）weighted merge；（c）contextual feature propagation loss）

提取深层特征，并利用知识一致注意力模块（Li等，2020）对有效特征进行赋值，其优势在于注意力分数是通过当前注意力分数和上一次迭代的分数的加权和来衡量，使前后帧补丁间具有相关性。解码器采用3个步长为2的4×4膨胀卷积来进行上采样重构图像，同时将浅层特征和深层特征进行级联，扩大模型的感受野。接着网络使用上下文特征传播损失（Zeng等，2021）来提高特征匹配过程中的鲁棒性，通过在深层特征中引入保证网络准确获得深层语义特征，进而产生语义一致的内容；然后将第1次输出的图像再次送入网络进行渐进式迭代处理，直到空洞区域被填充。最后，使用加权合并方法来融合渐进式迭代生成的特征图组$\boldsymbol{I}_{\text{out}}$，通过拼接操作对特征信息进行保留，使用激活函数（Leaky-ReLU，L_ReLU）和批归一化（batch normalization，BN）避免梯度消失。该文还使用一个预训练的VGG-16（Visual Geometry Group）特征提取器，联合L_1损失、感知损失、风格损失和平滑损失对模型进行约束，提高了参考图像和目标图像间的相似性。

1.2 加权合并模块

当渐进式迭代次数达到设定的阈值时，空洞区域填充完成，阈值设定见2.5.3小节。但如果直接使用此时输出的特征图，则会存在梯度消失和中间生成特征丢失问题。如果采用平均合并和自适应合并，则早期输出的重建图像中的缺失区域会影响最终输出图像的质量。

为了解决上述问题，本文提出一种加权合并的方法（图2（b））来融合每次渐进式迭代生成的特征图。自适应合并方法通过分析掩膜生成权重，而无需学习过程，如图2（a）所示，首先划分掩膜特征图$\boldsymbol{f}_{\text{mask}}$，具体为

$$\boldsymbol{f}_{\text{mask}}=\begin{cases}\boldsymbol{I}_{m_0} & t=0\\ \boldsymbol{I}_{m_t}-\boldsymbol{I}_{m_{t-1}} & \text{其他}\end{cases} \tag{1}$$

式中，t是迭代次数，$\boldsymbol{I}_{m_t}$和$\boldsymbol{I}_{m_{t-1}}$分别是第t次和第$t-1$次掩膜（mask）迭代更新后的特征图。本文为每次渐进式迭代生成的输出图像$\boldsymbol{I}_{\text{out}}^t$构造一个权重映射$\boldsymbol{W}_j^t$，具体为

$$\boldsymbol{W}_j^t=\sum_{j=0}^{N_1}\boldsymbol{f}_{\text{mask}}\times\frac{1}{1+\text{e}^{-\frac{t+1}{n-j}}} \tag{2}$$

式中，j是本次待修复的图像块数目，N_1是待修复的图像块总数目，$\boldsymbol{W}_j^t$表示第t次迭代输出图像$\boldsymbol{I}_{\text{out}}^t$对应的第$j$个修复区域的权重映射，$n$为本次修复过程中生成的图像块总数量。

然后，使用softmax函数归一化得到权重映射值w_j。$\boldsymbol{I}_o$由第t次迭代输出特征图$\boldsymbol{I}_{\text{out}}^t$在位置$(x, y)$的加权和组成，即

$$\boldsymbol{I}_o=\sum_{j=0}^{n}w_j\times\delta\left(\frac{v_{x,y}^t-\mu_{x,y}}{\sqrt{\sigma_{x,y}^2+\varepsilon}}\right) \tag{3}$$

$$\mu_{x,y}=\frac{1}{H\times W}\sum_{x=1}^{H}\sum_{y=1}^{W}v_{x,y}^t \tag{4}$$

$$\sigma_{x,y}=\sqrt{\frac{1}{H\times W}\sum_{x=1}^{H}\sum_{y=1}^{W}\left(v_{x,y}^t-\mu_{x,y}\right)^2} \tag{5}$$

式中，$\mu_{x,y}$，$\sigma_{x,y}$是特征值$v_{x,y}^t$的均值和标准差，ε是误差值，防止标准差为0的情况，$v_{x,y}^t$表示第t次迭代输出图像$\boldsymbol{I}_{\text{out}}^t$的特征值，$H$和$W$分别表示第$t$次迭代输出图像$\boldsymbol{I}_{\text{out}}^t$的高度和宽度。$\delta$表示Leaky-ReLU激活函数。

加权合并是由软权重映射和自适应合并的输出特征图拼接（concat）后经过学习过程得到的自适应图，拼接操作可以在不破坏原始特征图的情况下整体地保留特征的信息。软权重方法的细节图如图2（b）所示，本文将$\boldsymbol{I}_{\text{out}}$和输入特征映射$\boldsymbol{I}_{\text{in}}$连接起来，以获得一个软权重映射$\boldsymbol{W}_{\text{s}}$，具体为

$$\boldsymbol{W}_{\text{s}}=\sigma\left(Conv\left(\left[\boldsymbol{I}_{\text{in}},\boldsymbol{I}_{\text{out}}\right]\right)\right)\times(\boldsymbol{1}-\boldsymbol{M})+\boldsymbol{M} \tag{6}$$

式中，σ是sigmoid激活函数，$\boldsymbol{M}$是二值化图，$\boldsymbol{0}$表示空洞区域，$\boldsymbol{1}$表示有效区域。软权重得到的特征图的值I_{s}表示为

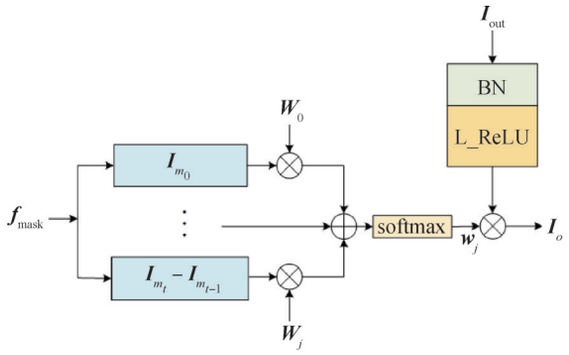
$$I_{\text{s}}=\sum_{j=0}^{n}w_{\text{s}}\times\delta\left(\frac{v_{x,y}^t-\mu_{x,y}}{\sqrt{\sigma_{x,y}^2+\varepsilon}}\right) \tag{7}$$

式中，δ，$v_{x,y}^t$，$\mu_{x,y}$，$\sigma_{x,y}$和ε变量同式（3）。

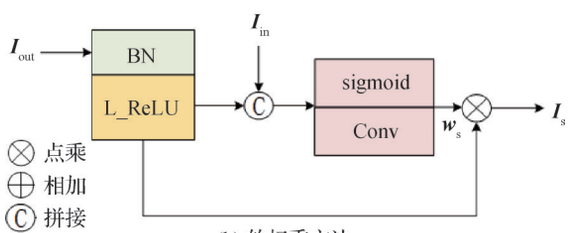
1.3 上下文特征传播损失模块

受图像修复中的CR损失（contexutal reconstruction loss）（Zeng等，2021）启发，本文设计了上下文特征传播损失模块。不同于CR损失的是，CR损失用于无注意力的精网络阶段，类似于DeepFillv2（Yu等，2019）中的架构去掉上下文注意力（contextual attention，CA）（Yu等，2018），本文用于有注意力的渐进式迭代网络，与知识一致注意力模块起到相辅相成的作用，它并不直接参与空洞的生成，仅在学习更好特征的训练过程中起作用。

图1（c）为上下文特征传播损失的流程图。相



(a) 自适应合并方法



(b) 软权重方法

图2　加权合并细节

Fig. 2　The detail of weighted merge

((a) adaptive merge; (b) soft weight)

似性编码器将渐进式迭代生成的输出图像$\boldsymbol{I}_{\text{out}}$作为输入，辅助编码器将输入特征$\boldsymbol{I}_{\text{in}}$作为输入，在辅助编码器的特征通过解码器之前，通过相似性编码卷积计算生成图像块和非空洞区域图像块的相似度$\boldsymbol{S}$，其中卷积核获取操作是通过提取自身的补丁来处理背景特征图，大小设置为步长为2的4×4卷积核，避免造成棋盘效应。然后将相似度$\boldsymbol{S}$送入辅助解码器，通过辅助解码器反卷积重建得到辅助图像。在辅助图像的指引下，使得背景空洞能够生成具有语义一致性的填充块。

不同于本文使用的其他损失项，上下文特征传播损失定义为辅助图像的空洞填充损失，可以看做是辅助图像块局部损失的总和，即

$$L_{\text{CFP}} = \sum_i l_i(\boldsymbol{u}_i) \tag{8}$$

式中，i表示迭代输出的$\boldsymbol{I}_{\text{out}}$中相似度$\boldsymbol{S}$最高的图像块索引，$\boldsymbol{u}_i$表示相似度$\boldsymbol{S}$最高的辅助图像块，$l_i(\cdot)$是第$i$个辅助图像块的局部空洞损失。

1.4　损失函数

在损失函数设计中，本文融合了L_1损失、感知损失(Johnson等，2016)、风格损失(Gatys等，2016)、平滑损失(Liu等，2018)，以及上下文特征传播损失来提升背景空洞区域和已知区域的语义一致性。

L_1损失是由非空洞区域的损失L_{valid}和空洞区域的损失L_{hole}组成，即

$$L_1 = \lambda_1 L_{\text{valid}} + \lambda_2 L_{\text{hole}} \tag{9}$$

式中，λ_1和λ_2分别表示L_{valid}和L_{hole}的权重，且

$$L_{\text{valid}} = \frac{1}{H \times W \times C} \left\| \boldsymbol{M} \odot (\boldsymbol{I}_{\text{out}}^t - \boldsymbol{I}_{\text{gt}}) \right\|_1 \tag{10}$$

$$L_{\text{hole}} = \frac{1}{H \times W \times C} \left\| (\boldsymbol{1} - \boldsymbol{M}) \odot \left(\boldsymbol{I}_{\text{out}}^t - \boldsymbol{I}_{\text{gt}}\right) \right\|_1 \tag{11}$$

式中，$\boldsymbol{M}$如式(4)定义，$\odot$表示点乘。$\boldsymbol{I}_{\text{gt}}$表示原始参考视图，$\boldsymbol{I}_{\text{out}}^t$表示第$t$次渐进式迭代网络输出的预测图，$C$、$H$、$W$分别是通道数、特征补丁的高度和宽度，分别为$C = 256$，$H = 32$，$W = 32$。根据Liu等人(2018)的设计将权重参数设为$\lambda_1 = 1$，$\lambda_2 = 6$。

感知损失用来增强填充图像和真实图像间高级特征结构的相似性，其定义为

$$L_{\text{per}} = \sum_{m=1}^{N_2} \frac{\left\| \boldsymbol{\psi}_m\left(\boldsymbol{I}_{\text{gt}}\right) - \boldsymbol{\psi}_m\left(\boldsymbol{I}_o\right) \right\|_1}{H_m W_m C_m} \tag{12}$$

式中，$\boldsymbol{I}_o$是输出结果图，由$\boldsymbol{I}_{\text{out}}^t$中的空洞生成像素和$\boldsymbol{I}_{\text{gt}}$中的非空洞像素组成；$\boldsymbol{\psi}_m\left(\boldsymbol{I}_{\text{gt}}\right)$和$\boldsymbol{\psi}_m\left(\boldsymbol{I}_o\right)$分别表示$\boldsymbol{I}_{\text{gt}}$和$\boldsymbol{I}_o$经过预训练网络VGG-16第$m$个池化层后输出的特征($m = 1, 2, 3$)；$H_m$，$W_m$，$C_m$表示提取到的第$m$特征图的高度、宽度和通道数；$N_2$是池化层的数目。

风格损失用来补偿感知损失不能有效保持填充区域与周边区域的风格一致性问题，定义为

$$L_{\text{style}} = \sum_{m=1}^{N_2} \frac{1}{C_m \times C_m} \left\| \frac{1}{H_m W_m C_m} \left(\boldsymbol{\psi}_m^{\text{gram}}\left(\boldsymbol{I}_{\text{gt}}\right) - \boldsymbol{\psi}_m^{\text{gram}}\left(\boldsymbol{I}_o\right) \right) \right\|_1 \tag{13}$$

式中，$\boldsymbol{\psi}_m^{\text{gram}}$是通过计算特征图的格拉姆矩阵(Gram)来获得网络中每层特征的相似性，即$\boldsymbol{\psi}_m^{\text{gram}} = \boldsymbol{\psi}_m \boldsymbol{\psi}_m^{\text{T}}$。

平滑损失用来保持填充后图像的光滑性，定义为

$$L_{\text{tv}} = \sum_{(i,j) \in \boldsymbol{R}, (i,j+1) \in \boldsymbol{R}} \frac{\left\| P_{i,j+1} - P_{i,j} \right\|_1}{N_c} + \sum_{(i,j) \in \boldsymbol{R}, (i+1,j) \in \boldsymbol{R}} \frac{\left\| P_{i+1,j} - P_{i,j} \right\|_1}{N_c} \tag{14}$$

式中，$P_{i,j}$表示$\boldsymbol{I}_o$中的一个像素点，$P_{i,j+1}$和$P_{i+1,j}$分别表示$P_{i,j}$垂直方向和水平方向的相邻像素点。$\boldsymbol{R}$是空洞区域中像素为1的膨胀区域，N_c是$\boldsymbol{I}_o$的图像块总数量。

综上所述，再结合1.4节的上下文特征传播损失L_{CFP}，总损失函数可表达为

$$L_{\text{total}} = L_1 + \lambda_{\text{per}} L_{\text{per}} + \lambda_{\text{style}} L_{\text{style}} + \lambda_{\text{tv}} L_{\text{tv}} + \lambda_{\text{CFP}} L_{\text{CFP}} \tag{15}$$

式中，λ_{per}，λ_{style}，λ_{tv}，λ_{CFP}分别是感知损失、风格损失、平滑损失和上下文特征传播损失的权重参数，其大小是根据Liu等人(2018)和Zeng等人(2021)方法，通过实验得到：$\lambda_{\text{per}} = 0.05$，$\lambda_{\text{style}} = 120$，$\lambda_{\text{tv}} = 0.1$，$\lambda_{\text{CFP}} = 0.5$。

2 实验结果与分析

为方便与先前方法对比，本文也采用先前方法中的测试序列，即微软实验室提供的3D视频序列(Luo等，2020)、韩国ETRI(Electrouics and Telecommunications Research Institute)研究院提供的Lovebird1序列和波兰波兹南理工大学提供的Poznan_Street序列，额外增加日本名古屋大学提供的Kendo序列和韩国GIST(Gwangju Institute of Science and Technology)研究院提供的Newspaper序列来评估渐进式迭代网络模型。本文使用自制的空洞掩膜集对模型进行训练和测试，数据集详情见表1。

表1 数据集详情

Table 1 The detail about datasets

数据集	分辨率/像素	训练集:测试集	修复帧数	总帧数
Ballet	1 024×768	9:1	100	800
Breakdancers	1 024×768	9:1	100	800
Lovebird1	1 024×768	9:1	100	3 600
Newspaper	1 024×768	9:1	100	2 700
Kendo	1 024×768	9:1	100	2 800
Poznan_Street	1 920×1 088	9:1	100	2 500

2.1 实验参数设置

本文使用Adam优化器对模型进行优化，优化器的参数设置为$\beta_1 = 0.5$，$\beta_2 = 0.999$。初始时，以0.000 1的学习率训练模型，然后将学习速率设置为0.000 01来对模型进行微调。本文模型以端到端的方式训练，batchsize设置为8，并在自制的空洞掩膜集上进行训练和测试。为防止出现显卡存储空间不足，在训练时将分辨率为1 024×768像素的序列集和对应的空洞掩膜图裁剪成256×256像素大小作为模型输入，分辨率为1 920×1 088像素的Poznan_Street序列集裁剪为240×272像素。所有模型均在11 G NVIDIA RTX2080Ti GPU上进行训练，网络构建基于PyTorch深度学习框架。

2.2 空洞掩膜集

已有的面向图像修复的掩膜是采用规则矩形掩膜或涂抹产生的不规则掩膜(Liu等，2018)，这与绘制图像的空洞分布特征是存在差别的。虚拟视点绘制过程中产生的空洞是因为使用DIBR合成虚拟视图的过程中需要执行三维图像变换(3D image warping)，使得原先被前景所遮挡的背景在合成视点下暴露出来，从而产生空洞等问题。如图3所示，3D视频虚拟视点绘制空洞面积较大，主要分布在图像的边界以及前景物体的右侧。本文根据3D视频虚拟视点绘制的需求制作了48 000幅空洞掩膜，通过执行裁剪、水平和镜像翻转操作从4 000幅原始像素的空洞掩膜中随机采样1幅掩膜来扩充空洞掩膜数据集。其中选取9 600幅用于训练，1 200幅用于测试。用于训练和测试的空洞掩膜和序列图大小均为256×256像素。

2.3 定性实验

本文通过各序列的参考视点和虚拟视点的位置关系来命名绘制结果。比如对于Ballet序列，以Cam4作为参考视点，绘制Cam3视点处的虚拟视图，得到的结果命名为BA43。本文方法与VSRS(view synthesis reference software)(Wegner等，2013)、Ahn(Ahn和Kim，2013)、Luo(Luo等，2020)、RFR(Li等，2020)和LGNet(local and global refinement subnetworks)(Quan等，2022)共5种方法在自制的空洞掩膜集上进行对比。VSRS是MPEG(motion picture expert group)视图合成参考软件。Ahn方法先对绘制图和深度图进行后处理，再在深度图引导下进行空洞填充。Luo方法先通过前景提取、运动补偿模块获取背景绘制图和深度图，再由修复模块填充空洞。为验证本文模型的性能，加入图像修复中的主流模型进行对比，RFR和LGNet都是基于深度学习的图像修复模型。

图4展示了BA54第89帧、BR56第16帧、LB85第93帧和PS35第99帧的实验结果，图5展示了深度学习模型在Newspaper第95帧和Kendo第76帧的实验结果。VSRS方法由于采用行遍历算法填补空洞，往往会产生前景纹理的渗透和不真实的填充结果，如图4(c)所示，空洞部分区域错误地将前景纹

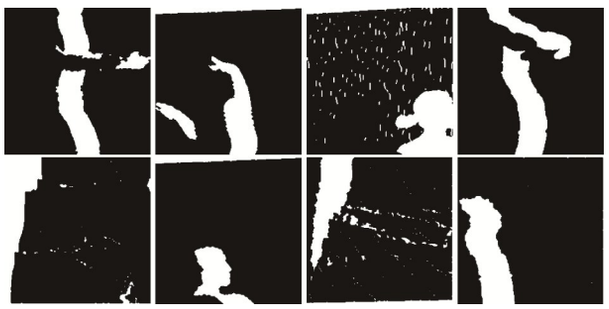


图 3　部分空洞掩膜图

Fig. 3　The part of hole mask maps

理填充到背景区域，导致人物出现伪轮廓，前景错误扩散现象比较严重，并且左边墙面出现不真实的纹理。从图 4(d)可以看出，Ahn 方法在整体视觉上没有严重的错误纹理，但由于块匹配过程中没有前后帧关联性，导致左边墙面上的海报填充错误，男舞者背景区域的窗帘细节纹理与周围环境不一致，并且出现了明显的块效应。Luo 方法在背景重建中采用前景提取和运动补偿有效防止前景纹理的模糊效果和伪影，并针对虚拟视点绘制的大面积空洞生成结构合理的结果，但对于纹理复杂的区域有一定难度，如图 4(e)最左边的海报以及男舞者突兀的背景纹理，同时芭蕾女舞者腿边的背景区域出现局部错误的纹理。基于深度学习的图像修复模型(RFR 和 LGNet)对于纹理复杂的结构有较好的修复能力，如图 4(f)和图 4(g)，但待修补的背景边界处存在鬼影。从图 4(h)和图 5(e)可知，本文方法对于虚拟视点绘制的大面积空洞填充效果要优于其他方法，在渐进式迭代网络中引入 KCA 模块和 CFP 损失使得整体的结构具有连贯性，对于纹理复杂区域也更具一致性，更接近参考视图。

(a) 参考视图　(b) 空洞区域　(c) VSRS　(d) Ahn　(e) Luo　(f) RFR　(g) LGNet　(h) PINet(本文)

图 4　传统方法和深度学习方法在 4 个序列上的定性比较

Fig. 4　Qualitative comparison of traditional methods and deep-learning methods on four sequences

((a) reference view; (b) hole region; (c) VSRS; (d) Ahn; (e) Luo; (f) RFR; (g) LGNet; (h) ours)

(a) 参考视图　(b) 空洞区域　(c) RFR　(d) LGNet　(e) PINet(本文)

图 5　深度学习方法在 Newspaper 和 Kendo 序列上的定性比较

Fig. 5　Qualitative comparison of deep-learning methods on Newspaper and Kendo sequences

((a) reference view; (b) hole region; (c) RFR; (d) LGNet; (e) ours)

2.4 定量实验

为客观评价本文模型的空洞填充结果，分别使用峰值信噪比（peak signal-to-noise ratio，PSNR）和结构相似性（structural similarity，SSIM）作为评价标准。表2展示本文方法在6个测试序列生成的所有目标视图的PSNR和SSIM值。PSNR是基于对应像素点间的误差。SSIM是从结构、亮度和对比度3方面衡量参考图像和目标图像相似度的指标。PSNR和SSIM数值越大，图像质量越好。由表2可知，本文方法与传统方法和深度学习模型相比，PSNR和SSIM评价指标在Ballet序列集上分别提升1.2%～21.2%和0.4%～3.8%，在Breakdancers序列集上分别提升了0.2%～3.6%和0.1%～4.3%，在Lovebird1序列集上分别提升了0.04%～2.4%和0.1%～5.02%，在Poznan_Street序列集上分别提升了2.9%～5.5%和4.1%～8.5%；在Newspaper和Kendo序列集上，本文方法与性能第2的前沿深度学习模型相比，其PSNR和SSIM分别提升了1.76%、2.63%和1.38%、0.79%，体现了本文方法的优越性。

表2 不同序列集上空洞填充结果对比

Table 2 Comparison of hole-filling results on different sequence sets

方法	BA54		BA43		BR56		BR57		LB84	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
VSRS (Wegner等，2008)	27.331	0.865	27.522	0.862	29.536	0.789	27.413	0.762	20.890	0.676
Ahn (Ahn和Kim，2013)	30.819	0.872	27.615	0.869	29.617	0.790	28.013	0.770	21.037	0.677
Luo (Luo等，2020)	31.594	0.880	27.996	0.870	29.551	0.790	28.1772	0.772	21.280	0.701
RFR (Li等，2020)	32.126	0.889	28.457	0.870	30.065	0.796	28.285	0.788	21.301	0.703
LGNet (Quan等，2022)	32.731	0.891	29.989	0.872	30.188	0.798	28.393	0.793	21.338	0.708
PINet (本文)	**33.151**	**0.898**	**31.291**	**0.876**	**30.614**	**0.799**	**28.457**	**0.795**	**21.406**	**0.710**

方法	LB85		PS35		NP46		KD31	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
VSRS (Wegner等，2008)	22.512	0.711	25.270	0.768	–	–	–	–
Ahn (Ahn和Kim，2013)	22.606	0.728	25.386	0.771	–	–	–	–
Luo (Luo等，2020)	22.656	0.729	25.541	0.775	–	–	–	–
RFR (Li等，2020)	22.889	0.735	25.877	0.793	23.423	0.798	29.416	0.884
LGNet (Quan等，2022)	22.984	0.740	25.909	0.800	23.715	0.800	30.102	0.889
PINet (本文)	**22.994**	**0.741**	**26.675**	**0.833**	**24.133**	**0.811**	**30.895**	**0.896**

注：加粗字体表示各列最优结果。“–”表示没有对比数据。

2.5 消融实验

2.5.1 知识一致注意力的有效性

为验证KCA的有效性，对比了未使用注意力、添加CA模块以及KCA模块后的网络，在BA54和BR56序列上进行实验测试。如表3所示，本文引入的KCA模块比深度学习领域中较为经典的CA模块更精准地捕获特征。图6中未添加注意力的图像缺少细节纹理，添加KCA模块后更接近参考视图。

表3 知识一致注意力模块消融实验

Table 3 Ablation experiments of KCA

方法	BA54序列集		BR56序列集	
	PSNR/dB	SSIM	PSNR/dB	SSIM
未使用	32.285	0.893	29.490	0.791
CA	32.239	0.892	29.438	0.790
KCA	**33.151**	**0.898**	**30.614**	**0.799**

注：加粗字体表示各列最优结果。

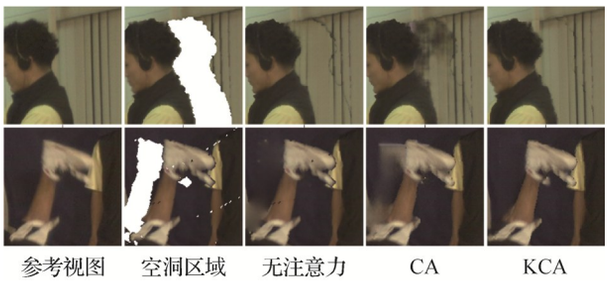


图 6　知识一致注意力的消融实验对比图

Fig. 6　Comparison of ablation experiments of KCA

2.5.2　加权合并的有效性

为证明本文提出的加权合并的有效性，图 7 展示不同特征合并方法的比较。如图 7 所示，如果只使用最后一个特征图作为输出进行合并，会导致纹理模糊和内容不充分。从图 7 可以看出，平均合并和自适应合并虽然有纹理细节，但还是存在纹理模糊和扩散的现象。因此本实验证明在软权重方法的辅助下可以有效保留网络早期特征。

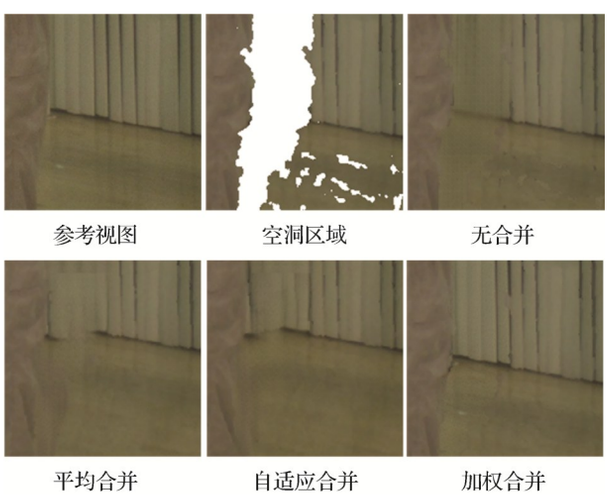


图 7　加权合并的消融实验对比图

Fig. 7　Comparison of ablation experiments of weighted merge

2.5.3　迭代次数的有效性

通过实验证明，迭代次数设置为 6 完成对空洞区域的填充并得到最优结果。为说明不同迭代次数的影响，本文在空洞面积较大的 Ballet 序列上进行实验。表 4 给出了不同迭代次数对应的结果。结果表明，并不是层数越深，模型性能越好。对于较小和较大的迭代次数，都没有达到网络的最佳性能。

2.5.4　上下文特征传播损失的有效性

为验证本文的 CFP 损失模块的有效性，同时能与 KCA 模块起到相辅相成的作用，表 5 展示了基线模型（baseline）、只添加 KCA 模块、只添加 CFP 损失模块的模型和本文方法在 BA54 序列和 BR56 序列上的比较，Baseline 表示本文网络不加 KCA 模块和 CFP 损失模块。可以看出，虽然 CFP 损失只起到影响网络的作用，但本文方法在 BA54 序列上 PSNR 和 SSIM 分别比单独加入 KCA 模块高出 2.03% 和 0.39%，可见本文方法的优越性。从图 8 也可知，本文方法同时引入 KCA 模块和 CFP 损失能够更好地提取语义信息，增强目标视图和参考视图的视觉一致性。

表 4　迭代次数在 Ballet 序列上的影响

Table 4　The influence of the number of iterations on Ballet sequence

迭代次数	PSNR/dB	SSIM
4	32.128	0.887
6	**33.151**	**0.898**
8	33.086	0.895

注：加粗字体表示各列最优结果。

表 5　上下文特征传播损失的消融实验

Table 5　Ablation experiments of CFP loss

方法	BA54 序列集		BR56 序列集	
	PSNR/dB	SSIM	PSNR/dB	SSIM
Baseline	30.517	0.882	28.346	0.789
Baseline+KCA	32.491	0.894	29.515	0.793
Baseline+CFP	32.285	0.893	29.490	0.791
本文	**33.151**	**0.898**	**30.614**	**0.799**

注：加粗字体表示各列最优结果。

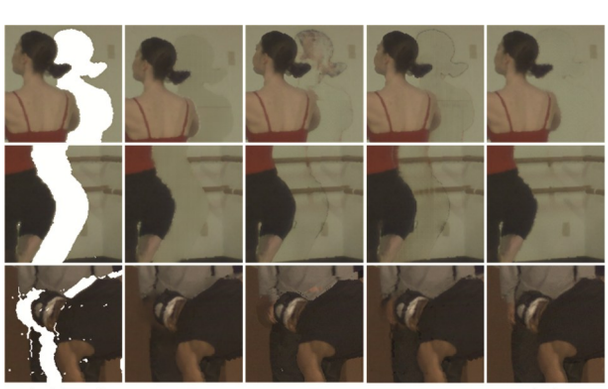


图 8　上下文特征传播损失的消融实验对比图

Fig. 8　Comparison of ablation experiments of CFP loss

3 结 论

针对虚拟视点绘制中产生的大面积空洞问题，本文提出了一种渐进式迭代网络来填充绘制图像缺失区域。该网络结构采用深度学习的编码—解码架构，同时引入跳跃连接，通过浅层特征和深层特征的级联，减少了信息丢失的问题。采用知识一致注意力模块对有效特征进行关注，并且在深层特征中引入上下文特征传播损失模块，使得网络对特征匹配进行约束。最后使用加权合并方法保护早期生成的特征。此外，针对虚拟视点绘制空洞填充的需求，本文自制了适用于虚拟视点绘制的空洞掩膜集。实验结果表明，与对比方法相比，本文网络取得了更优的结果。

本文方法也存在一定的局限性。1）在网络模型设计上，受限于物理设备，没有尝试使用具有多头注意力的Transformer模型，由于Transformer模型具有较强的特征表达能力，可以得到更佳的性能。2）本文方法没有加入3D视频序列的相关深度图，如果给予深度图的引导，网络能够更好地理解图像的相关信息，并习得更准确的填充结构。后续工作将以深度图为结构信息，增强空间信息的使用。

参考文献（References）

Ahn L and Kim C. 2013. A novel depth-based virtual view synthesis method for free viewpoint video. IEEE Transactions on Broadcasting, 59(4): 614-626 [DOI: 10.1109/TBC.2013.2281658]

Chen S Q, Liu Q and Yang Y. 2020. Adaptive multi-modality residual network for compression distorted multi-view depth video enhancement. IEEE Access, 8: 97072-97081 [DOI: 10.1109/ACCESS.2020.2996258]

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]

Gatys L A, Ecker A S and Bethge M. 2016. Image style transfer using convolutional neural networks//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2414-2423 [DOI: 10.1109/CVPR.2016.265]

Han D X, Chen H, Tu C H and Xu Y Y. 2018. View synthesis using foreground object extraction for disparity control and image inpainting. Journal of Visual Communication and Image Representation, 56: 287-295 [DOI: 10.1016/j.jvcir.2018.10.004]

Johnson J, Alahi A and Li F F. 2016. Perceptual losses for real-time style transfer and super-resolution//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 694-711 [DOI: 10.1007/978-3-319-46475-6_43]

Lee P J and Effendi. 2011. Nongeometric distortion smoothing approach for depth map preprocessing. IEEE Transactions on Multimedia, 13(2): 246-254 [DOI: 10.1109/TMM.2010.2100372]

Li J Y, Wang N, Zhang L F, Du B and Tao D C. 2020. Recurrent feature reasoning for image inpainting//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 7757-7765 [DOI: 10.1109/CVPR42600.2020.00778]

Liang H T, Chen X D, Xu H Y, Ren S Y, Wang Y and Cai H Y. 2019. Virtual view rendering based on depth map preprocessing and image inpainting. Journal of Computer-Aided Design and Computer Graphics, 31(8): 1278-1285 (梁海涛, 陈晓冬, 徐怀远, 任思宇, 汪毅, 蔡怀宇. 2019. 基于深度图预处理和图像修复的虚拟视点绘制. 计算机辅助设计与图形学学报, 31(8): 1278-1285) [DOI: 10.3724/SP.J.1089.2019.17541]

Lin C Y, Zhao Y, Xiao J M and Tillo T. 2018. Region-based multiple description coding for multiview video plus depth video. IEEE Transactions on Multimedia, 20(5): 1209-1223 [DOI: 10.1109/TMM.2017.2766043]

Liu G L, Reda F A, Shih K J, Wang T C, Tao A and Catanzaro B. 2018. Image inpainting for irregular holes using partial convolutions//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 89-105 [DOI: 10.1007/978-3-030-01252-6_6]

Liu Z, Liu Q, Yang Y, Liu Y C, Jiang G Y and Yu M. 2016. Cluster-based cross-view filtering for compressed multi-view depth maps//Proceedings of IEEE Visual Communications and Image Processing Conference. Chengdu, China: IEEE: 1-4 [DOI: 10.1109/VCIP.2016.7805550]

Luo G B and Zhu Y S. 2017. Foreground removal approach for hole filling in 3D video and FVV synthesis. IEEE Transactions on Circuits and Systems for Video Technology, 27(10): 2118-2131 [DOI: 10.1109/TCSVT.2016.2583978]

Luo G B, Zhu Y S, Weng Z Y and Li Z T. 2020. A disocclusion inpainting framework for depth-based view synthesis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(6): 1289-1302 [DOI: 10.1109/TPAMI.2019.2899837]

Quan W Z, Zhang R S, Zhang Y, Li Z F, Wang J and Yan D M. 2022. Image inpainting with local and global refinement. IEEE Transactions on Image Processing, 31: 2405-2420 [DOI: 10.1109/TIP.2022.3152624]

Smirnov S, Battisti F and Gotchev A P. 2019. Layered approach for improving the quality of free-viewpoint depth-image-based rendering images. Journal of Electronic Imaging, 28(1): #013049 [DOI:

10.1117/1.JEI.28.1.013049]

Sun W X, Au O C, Xu L F, Li Y J and Hu W. 2012. Novel temporal domain hole filling based on background modeling for view synthesis//Proceedings of the 19th IEEE International Conference on Image Processing. Orlando, USA: IEEE: 2721-2724 [DOI: 10.1109/ICIP.2012.6467461]

Tanimoto M, Tehrani M P, Fujii T and Yendo T. 2011. Free-viewpoint TV. IEEE Signal Processing Magazine, 28(1): 67-76 [DOI: 10.1109/MSP.2010.939077]

Wang S M. 2016. An unidirectional criminisi algorithm for DIBR-synthesized images//Proceedings of the 2nd IEEE International Conference on Computer and Communications. Chengdu, China: IEEE: 574-578 [DOI: 10.1109/CompComm.2016.7924766]

Wang X, Liu Q, Peng Z J, Hou J H, Yuan H, Zhao T S, Qin Y, Wu K J, Liu W Y and Yang Y. 2023. Research progress of six degree of freedom (6DoF) video technology. Journal of Image and Graphics, 28(6): 1863-1890 (王旭, 刘琼, 彭宗举, 侯军辉, 元辉, 赵铁松, 秦熠, 吴科君, 刘文予, 杨铀. 2023. 6DoF 视频技术研究进展. 中国图象图形学报, 28(6): 1863-1890) [DOI: 10.11834/jig.230025]

Wegner K, Stankiewicz O, Tanimoto M and Domański M. 2013. Enhanced view synthesis reference software (VSRS) for free-viewpoint television. ISO/IEC JTC1/SC29/WG11 MPEG2013/M31520

Yao C, Tillo T, Zhao Y, Xiao J M, Bai H H and Lin C Y. 2014. Depth map driven hole filling algorithm exploiting temporal correlation information. IEEE Transactions on Broadcasting, 60(2): 394-404 [DOI: 10.1109/TBC.2014.2321671]

Ye G Z, Liu Y B, Deng Y, Hasler N, Ji X Y, Dai Q H and Theobalt C. 2013. Free-viewpoint video of human actors using multiple handheld kinects. IEEE Transactions on Cybernetics, 43(5): 1370-1382 [DOI: 10.1109/TCYB.2013.2272321]

Yu J H, Lin Z, Yang J M, Shen X H, Lu X and Huang T. 2019. Free-Form image inpainting with gated convolution//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 4470-4479 [DOI: 10.1109/ICCV.2019.00457]

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]

Zeng Y, Lin Z, Lu H C and Patel V M. 2021. CR-fill: generative image inpainting with auxiliary contextual reconstruction//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 14144-14153 [DOI: 10.1109/ICCV48922.2021.01390]

Zhu C and Li S. 2016. Depth image based view synthesis: new insights and perspectives on hole generation and filling. IEEE Transactions on Broadcasting, 62(1): 82-93 [DOI: 10.1109/TBC.2015.2475697]

Zhu S P, Xu H and Yan L N. 2019. An improved depth image based virtual view synthesis method for interactive 3D video. IEEE Access, 7: 115171-115180 [DOI: 10.1109/ACCESS.2019.2935021]

作者简介

刘家希，女，硕士研究生，主要研究方向为虚拟视点合成技术。E-mail：211080050@hdu.edu.cn

周洋，通信作者，男，副教授，硕士生导师，主要研究方向为三维视频编码。E-mail：zhouyang@hdu.edu.cn

林坤，男，硕士研究生，主要研究方向为虚拟视点合成技术。E-mail：201080008@hdu.edu.cn

殷海兵，男，教授，主要研究方向为视频编码。E-mail：yhb@hdu.edu.cn

唐向宏，男，教授，主要研究方向为图像处理与传输。E-mail：tangxh@hdu.edu.cn