基于多尺度特征融合的轻量化人脸图像修复算法

赵晓，赵子怡，杨晨
（陕西科技大学电子信息与人工智能学院，陕西 西安 710021）

摘　要：针对当前遮挡的人脸图像修复中修复图像质量差和模型参数量大的问题，提出了一种基于多尺度特征融合的改进U-Net的轻量化人脸图像修复模型——LM-UNET。首先，使用深度可分离卷积替换原有卷积，增强模型对不同通道和上下文信息的特征表达能力，实现模型轻量化；其次，在跳跃连接中设计了多尺度特征注意力融合模块，充分融合不同尺度特征的信息，内嵌残差块减少特征间语义差距，提高模型修复准确率；最后，引入了位置注意力模块，增强人脸图像的显著信息，提升模型对人脸位置像素信息的有效提取能力。在基于CK+数据集生成的遮挡人脸数据集MFD上对该算法进行训练、验证和测试，修复后的图像的峰值信噪比（PSNR）达到30.49 dB，结构相似性（SSIM）达到96.85%，与其他模型的对比实验结果表明，该模型对存在遮挡的人脸修复图像质量和视觉效果更好。

关键词：图像修复；人脸图像；深度可分离卷积；多尺度特征注意力融合；位置注意力
中图分类号：TP391
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2024183

Lightweight face image restoration algorithm based on multi-scale feature fusion

ZHAO Xiao, ZHAO Ziyi, YANG Chen
School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi’an 710021, China

Abstract: Aiming at the problems of poor quality of restored images and large number of model parameters in the current occluded face image restoration, a lightweight face image restoration model based on multi-scale feature fusion with improved U-Net, LM-UNET, was proposed. Firstly, the original convolution was replaced by a depthwise separable convolution to enhance the feature expression ability of the model for different channels and contextual information. Secondly, a multi-scale feature attention fusion module was designed in the jump connection to fully fuse the information of different scale features, and the embedded residual block reduced the semantic gap between features to improve the repair accuracy of the model. Finally, a positional attention module was introduced to enhance the salient

收稿日期：2024−04−17；**修回日期：**2024−06−20
通信作者：赵子怡，875194085@qq.com
基金项目：国家自然科学基金资助项目（No.61971272，No.61601271）
Foundation Items: The National Natural Science Foundation of China(No.61971272, No.61601271)

information of the face image, and improve the model's effective extraction ability of facial positional pixel information of the model. The algorithm was trained, validated and tested on the occluded face dataset MFD generated based on the CK+ dataset, and the PSNR of the repaired image reached 30.49 dB and SSIM reached 96.85%. The experimental results of comparing the model with the other models show that the model has better image quality and visual effect for restoration of the face in the presence of occlusion.

Key words: image restoration, face image, depthwise separable convolution, multi-scale feature attention fusion, positional attention

0　引言

人脸图像是计算机视觉领域中常见的信息载体，承载着丰富的身份特征信息。图像修复就是在图像遮挡或退化等情况下，基于已有的像素信息补充未知的像素信息，还原图像中的真实信息。传统人脸图像修复模型基于图像样本相似度、结构纹理一致性等思想，结合数学、物理理论构建算法模型修复小区域破损图像，主要分为基于偏微分方程类和基于样本类两类修复方法。Ballester 等[1]基于偏微分方程提出的 VFGL 模型，使用灰度级、向量场的联合插值和流体动力学思想修复简单图像，但不能修复复杂纹理。Criminisi 等[2]基于样本类提出的 Criminisi 模型，复制结构纹理进行小区域修复，但计算相似度函数不稳定，速度慢。曹燕等[3]基于加权优先级和分类匹配的图像修复模型，改善修复顺序和最佳匹配块的选择问题，但计算量较大。传统方法对复杂纹理和高级语义要求更高的人脸图像修复没有良好的修复效果。

随着深度学习的不断发展，研究者们使用深度学习模型进行图像修复，通过深度学习捕捉更高层次的语义信息和更加丰富的特征信息有效补充缺失信息，达到很好的图像复原效果。基于 GAN 网络模型的图像修复[4-5]是根据博弈对抗思想提出的预测生成模型，被应用于图像修复领域且取得较好的效果，但该类模型存在图像修复不稳定、图像伪影和计算消耗资源大的问题。基于 Transformer 网络模型的图像修复[6-7]通过自注意力机制解决了卷积层只能获取局部感受野的不足，实现较大缺失区域的图像修复，但该类方法参数量大且难以修复图像中的较小视觉区域。U-Net 是一种经典的编码器-解码器结构的全卷积网络[8]，其特殊的跳跃连接结构，可以在特征提取和上采样过程中学习并拼接不同尺度的特征信息，进而重建出合理的图像纹理的结构。Liu 等[9]设计了一种连贯语义注意（coherent semantic attention，CSA）层，可以保留图像上下文结构，且能有效学习缺失区域特征间的语义相关性，但其难以学习图像缺失区域和已知区域间的对应关系，会导致修复图像出现伪影现象，该模型反复计算样本间相似度，参数量大。Zeng 等[10]在 U-Net 中使用金字塔注意力转移网络从深到浅逐层填充图像的缺失区域，但修复图像出现纹理伪影现象。罗仕胜等[11]将人脸特征点预测与 U-Net 结构结合，增加人脸关键点信息，但由于未有效利用不同尺度特征信息而存在图像模糊现象。Qin 等[12]提出的 MSA-Net 模型在 U-Net 结构中引入了多尺度注意力单元加强捕获不同感受野的深层特征能力，却存在语义混乱问题。Wang 等[13]使用动态选择机制和可迁移卷积动态选择空间卷积位置，但修复图像存在结构扭曲的情况。Liao 等[14]使用 U-Net 结构联合语义注意传播模块获取图像远距离语义相关性，但边界信息不一致导致修复图像模糊。

针对图像修复过程中存在语义混乱、对不同尺度特征的感知表达存在不足导致图像模糊和模型参数量大的问题，本文提出一种改进 U-Net 的

基于多尺度特征融合的轻量化人脸图像修复模型——LM-UNET。首先，利用深度可分离卷积（depthwise separable convolution，DWSC）替换原有卷积，提升模型通道与上下文特征的表达能力，有效提高了修复图像的质量且实现模型轻量化；其次，在跳跃连接中，设计了多尺度特征注意力融合（multi-scale feature attention fusion，MFAF）模块，加强不同尺度特征的感知表达能力，内部增加残差块丰富语义特征，减少特征间语义差距，有效提高了算法模型的修复精度；最后，引入了位置注意力模块（positional attention mode，PAM），增强人脸图像的显著信息，提升算法模型对人脸关键位置像素信息的有效提取能力。

1 方法

1.1 LM-UNET模型整体架构

LM-UNET整体结构如图1所示，该网络结构以U-Net为主干网络，主要分为编码器、解码器和跳跃连接3部分。编码器由5层DWSC模块、4层Down Sampling和2层PAM构成，随感受野增大提取丰富浅层信息，增强不同通道的特征交互并实现模型轻量化，添加PAM层加深模型对图像关键位置信息的关注与提取。解码器由4层DWSC模块、4层Up Sampling和4层特征拼接构成，提取深层信息的同时融合浅层信息。通过在跳跃连接中添加MFAF模块，将不同尺度特征充分融合，内嵌残差块加深浅层特征信息，减少特征间语义差距，使模型获得更好的修复效果。

1.2 深度可分离卷积模块

U-Net的特征提取主干网络使用3×3卷积学习原始特征，由于传统卷积层中卷积核是静态的，在提取特征时因为缺少空间和通道特征信息的交互，降低了模型的特征提取能力，图像修复效果不佳。在现有U-Net网络模型的改进研究中，常常通过更改模块结构或者添加模块结构等操作来提高模型的准确率，但是模型参数量的增加提高了计算成本。为此，本文提出采用深度可分离卷积替代U-Net原有卷积，一方面利用深度可分离卷积中的点卷积增强不同通道之间的信息交互，提升模型的特征表达能力，另一方面使用深度可分离卷积实现轻量化，减少模型的计算成本。

深度可分离卷积属于轻量化卷积，相比于普通卷积能够大大减少参数量和计算成本，可以实现网络模型轻量化，解决了网络模型计算量大的问题且在特征提取方面取得了良好的效果[15]。深度可分离卷积对有M个特征通道的特征图1，先通过M个3×3卷积核令其关注每一个通道内的特征信息，获得具有丰富特征信息的单通道特征图2，再通过1×1×M的点卷积有效利用不同通道中相同空间位置上的特征信息，获得不同通道信息融合的特征图3。所提的深度可分离卷积模块如图2所示。图像经过编码器、解码器共9层的

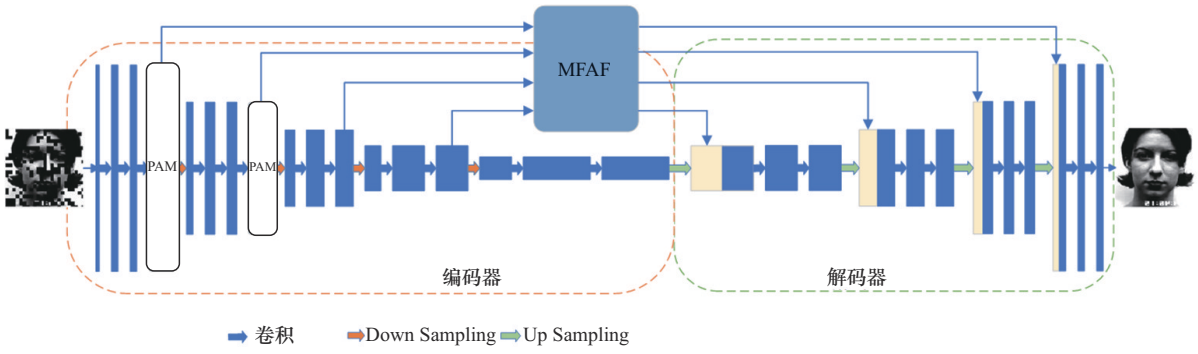


图1 LM-UNET整体结构

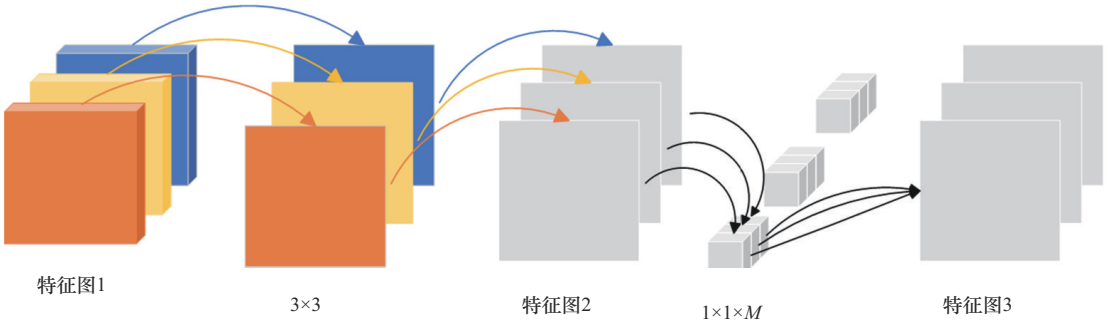


图2　深度可分离卷积模块

深度可分离卷积模块，优化后的图像特征向后传递进行卷积，提供了更有效的特征表示，增强主干网络对图像特征提取的精准度。

1.3　多尺度特征注意力融合模块

单一的卷积方式使浅层信息和深层信息间缺少不同尺度特征信息的交互，随着卷积的不断加深，浅层信息和深层信息之间的语义信息的差距越来越大，导致对图像细节特征信息的感知能力降低。为增加不同尺度特征信息交互和减小特征语义差距，本文提出在跳跃连接部分增添多尺度特征注意力融合模块。MFAF 中的 4 个残差迭代注意力特征融合（res-iterative attention feature fusion，RIAFF）模块由下向上递进式融合，将下层 RIAFF 的融合特征卷积后与上一层的 RIAFF 进行融合，分别向上和向后传递进行特征拼接，充分融合不同尺度的特征信息，并为解码器提供更加丰富的特征信息，保证特征信息的完整性，增强模型的感知表征能力。多尺度特征注意力融合模块如图 3 所示，残差特征注意力融合模块如图 4 所示。

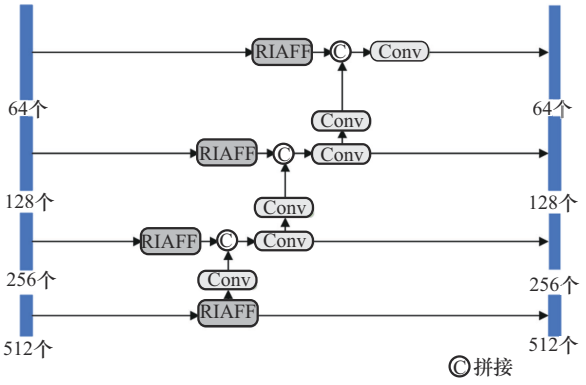


图3　多尺度特征注意力融合模块

RIAFF 由残差块（residual block）、迭代注意力特征融合模块（iterative attention feature fusion，IAFF）和残差连接构成。RIAFF 模块实现

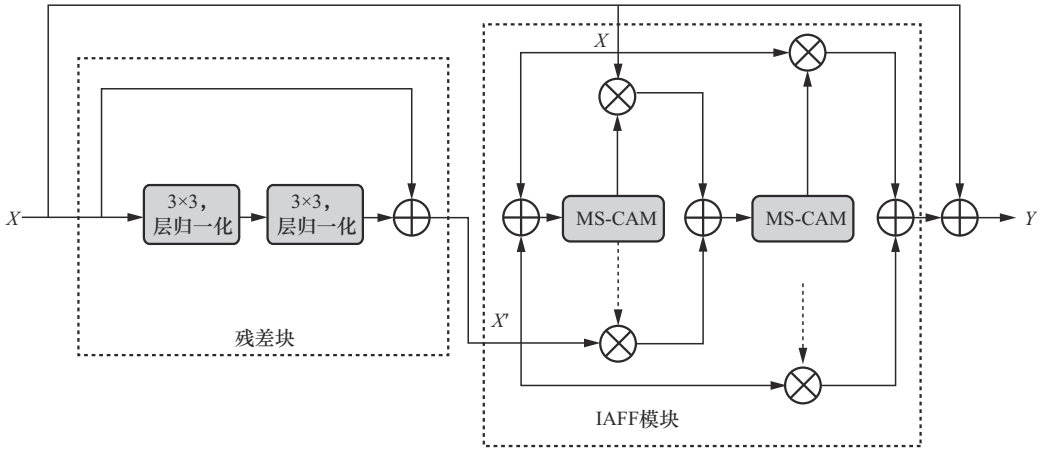


图4　残差特征注意力融合模块

步骤如下：

（1）将输入特征X通过残差块得到残差特征X'；

（2）将残差特征X'和输入特征X同时输入IAFF模块，通过内部的2层多尺度通道注意模块（multi-scate channel attention module，MS-CAM）为残差特征X'和输入特征X分配注意力权重并融合；

（3）将残差特征多尺度融合后的特征再次与原输入特征X进行残差连接以加深浅层特征信息，获得最后的融合特征Y。

RIAFF模块内部引入残差块与残差连接可以加深浅层特征，减少特征语义差距同时防止过拟合，以便在解码器的上采样中获得充足的特征信息，修复更加真实的图像。在残差块后添加IAFF模块用来关注通道信息和进行特征融合，IAFF使用2层MS-CAM对输入的2个不同特征进行2次分配注意力权重融合。MS-CAM通过添加全局池化层的左分支获取全局通道注意力权重，通过右分支来提取局部通道注意力权重，结合残差跳跃连接，可以更好地融合语义和尺度不一致的特征。多尺度通道注意模块如图5所示。

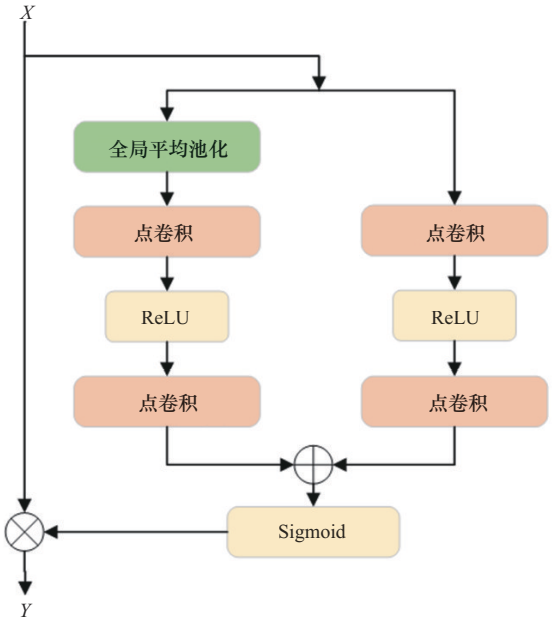


图5　多尺度通道注意模块

1.4　位置注意力模块

眼睛、鼻子和嘴等在人脸中具有相对的位置信息，这些位置信息可作为实现人脸图像修复的重要因素。当前模型忽略了图像中重要的位置信息，导致修复图像的细节处理不足，易受背景信息的干扰，考虑位置信息会随着卷积层次的不断加深而逐渐减弱，本文在编码器前2层卷积后引入位置注意力模块（positional attention module，PAM），利用上下文信息推断目标不同区域的重要性并根据其重要性分配相应权重，加强特征图的关键显著信息。位置注意力模块如图6所示。

位置注意力机制产生位置增强特征图分为3个步骤。首先，特征图A经过3个由BN和ReLU

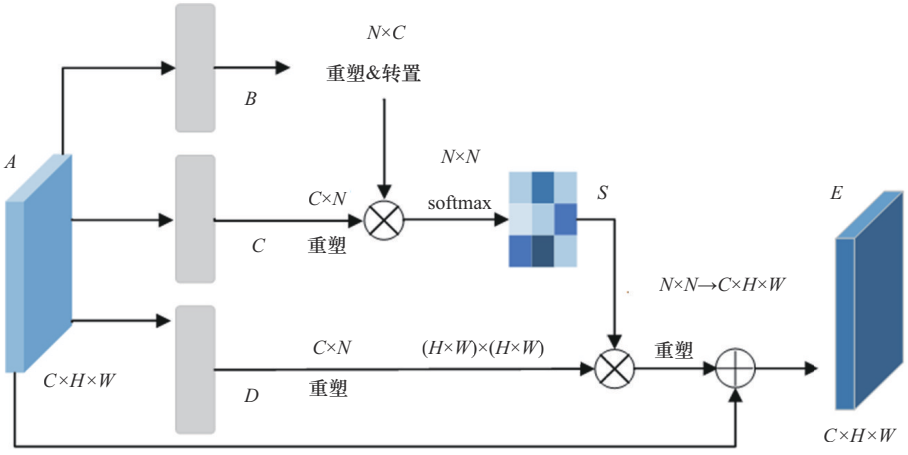


图6　位置注意力模块

构成卷积层得到B、C、D 3个特征图，通过重塑操作后将B与C进行转置相乘，生成一个空间注意力矩阵S，该矩阵可以对特征图中任意两个像素之间的空间关系进行计算；其次，将注意力机制矩阵S与通过重塑操作后的D进行矩阵乘法运算再重塑回原来大小；最后，将结果与A相应像素进行求和，得到可以反映位置信息新的特征图E。位置注意力的计算式如式（1）、式（2）所示。

$$S_{ji}=\frac{\exp(B_i\cdot C_j)}{\sum_{i=1}^{N}\exp(B_i\cdot C_j)} \tag{1}$$

其中，S_{ji}表示位置i对位置j的影响，B_i表示B中的元素，C_j表示C中的元素。

$$E_j=\alpha\sum_{i=1}^{N}(S_{ji}\cdot D_i)+A_j \tag{2}$$

其中，E_j表示E中的元素，α是尺度因子，D_i表示D中的元素，A_j表示A中的元素。

2　实验与分析

2.1　实验数据集与实验环境

本文在常用的公共真实人脸表情数据集CK+（Cohn-Kanade+）中选择部分图像采用随机水平翻转，随机旋转5°进行扩充，并调整输入图像大小为256×256（单位为像素），使用5×5的遮挡块对图像进行10%~50%的随机遮挡，生成模拟遮挡的数据集称作MFD数据集。该数据集共有13 900张人脸表情图像，其中10 548张图像作为训练集数据、3 352张图像作为验证集数据。图像遮挡效果如图7所示，图7显示了相对于原图10%~50%的5张不同遮挡效果。

本实验主要硬件环境配置为64 GB内存，Intel Core i9-12900K处理器，主频3.19 GHz；GPU为NVIDIA GeForce RTX 3080；操作系统为Windows 10。在实验部分，深度学习框架选用Pytorch1.11.0，开发语言版本是Python 3.8。实验参数设置是：Batch Size设置为8，学习率设置为0.01，迭代轮数为200，优化器为随机梯度下降，损失函数为L1损失函数。

2.2　定量评估指标

本文使用图像修复方面最常用的峰值信噪比（peak signal-to-noise ratio，PSNR）和结构相似性（structural similarity，SSIM）两种客观评价指标进行评估。

PSNR是一种像素级的信噪比，PSNR值越高，表明修复效果越逼真，用来衡量图像的修复质量。PSNR的计算式如式（3）所示：

$$\mathrm{PSNR}=10\lg\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right) \tag{3}$$

其中，MAX_I为图片可能的最大像素值，MSE为均方误差，MSE的计算式如式（4）所示：

$$\mathrm{MSE}=\frac{1}{N}\sum_{i=1}^{N}(x_i-y_i)^2 \tag{4}$$

其中，x表示生成图像，y表示原始图像，N表示图像像素点总量，i表示图像像素点变量。

SSIM是一种重要的图像相似度指数，用来度量修复图像与原来图像之间的相似性程度，综合考虑了图像亮度、对比度和结构信息3个因素，从而能更准确反映修复图像质量。SSIM值越接近1，修复效果越逼真，修复的图像质量越高。SSIM的计算式如式（5）~式（8）所示：

$$l(x,y)=\frac{2\mu_x\mu_y+\mathrm{c}_1}{\mu_x^2+\mu_y^2+\mathrm{c}_1} \tag{5}$$

原图　10%　20%　30%　40%　50%

图7　图像遮挡效果

$$c(x,y)=\frac{2\delta_x\delta_y+c_2}{\delta_x^2+\delta_y^2+c_2} \tag{6}$$

$$s(x,y)=\frac{\delta_{xy}+c_3}{\delta_x\delta_y+c_3} \tag{7}$$

$$\mathrm{SSIM}=l(x,y)^{\alpha}\cdot c(x,y)^{\beta}\cdot s(x,y)^{\gamma}=\frac{(2\mu_x\mu_y+c_1)(2\delta_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\delta_x^2+\delta_y^2+c_2)} \tag{8}$$

其中，$l(x,y)$表示亮度，$c(x,y)$表示对比度，$s(x,y)$表示结构，α、β、γ分别表示亮度、对比度、结构三者在SSIM的占比参数，通常都为1，c_1、c_2、c_3均表示常量，μ_x表示平均值，δ_x表示标准差，δ_{xy}表示x、y的协方差。

2.3 消融实验

为了验证所提网络模型LM-UNET各个模块的有效性，本节依次对每一个模块进行消融实验，其中Baseline表示原始U-Net模型，MFAF表示多尺度特征注意力融合模块；DWSC表示深度可分离卷积模块，PAM表示位置注意力模块。该实验在MFD数据集上验证，在网络模型中逐步添加模块或者修改相应的模块进行消融实验，PSNR、SSIM与epoch变化曲线如图8所示，图8展示了在图像复原指标PSNR和SSIM上随训练轮次的变化曲线，消融实验结果见表1。

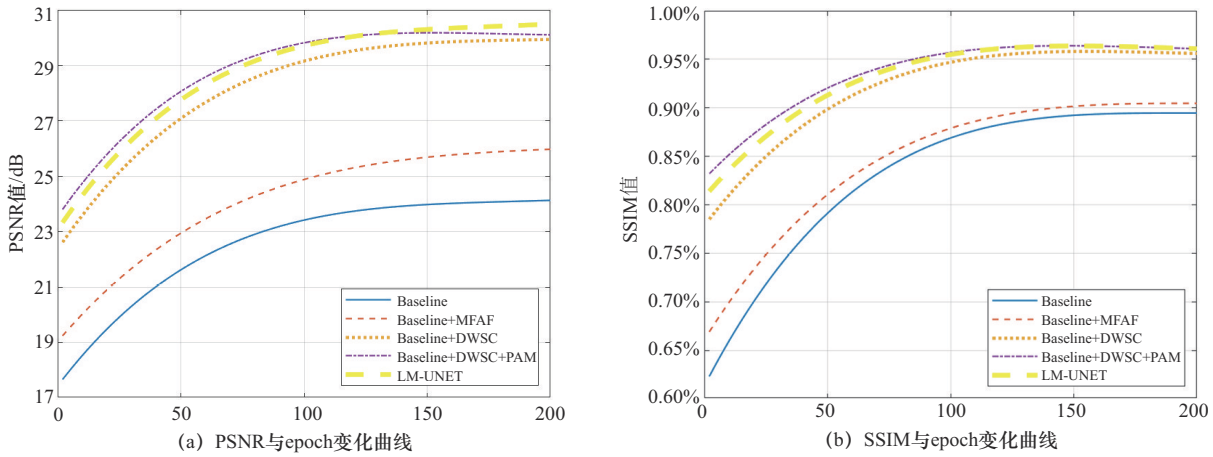
由图8和表1可知，在同一数据集上，随着不同模块的增加使用，LM-UNET模型的PSNR和SSIM值相对原始U-Net模型都有提升。添加MFAF模块较原始U-Net模型PSNR提升了1.01 dB，SSIM提升2.05%，但参数量提升6.98×10^6；添加DWSC替换原来卷积块，较原始U-Net模型PSNR提升了5.24 dB，SSIM提升6.29%，参数量出现大幅下降；使用DWSC和PAM模块，较原始U-Net模型PSNR提升了5.46 dB，SSIM提升6.60%，仍保持较少的参数量；同时使用MFAF、PAM和DWSC模块，LM-UNET较原始U-Net模型PSNR提升了5.92 dB，SSIM提升6.89%，Params降低了19.84×10^6。

表1 消融实验结果

Baseline	MFAF	DWSC	PAM	PSNR/dB	SSIM	Params/10^6
+	-	-	-	24.57	89.96%	32.08
+	+	-	-	25.58	92.01%	39.06
+	-	+	-	29.81	96.25%	5.23
+	-	+	+	30.03	96.56%	5.26
+	+	+	+	**30.49**	**96.85%**	**12.24**

为更好体现各个模块的有效性，对每次实验修复的面部图像进行保存。部分图像修复样例见表2。

由表2可知，原始U-Net模型的图像复原效果有明显的棋盘效应，修复效果模糊不清。在添加MFAF模块后棋盘效应减退，图像亮度有明显提升，但是存在部分细节缺失。在更改为DWSC



(a) PSNR与epoch变化曲线　(b) SSIM与epoch变化曲线

图8 PSNR、SSIM与epoch变化曲线

表2　部分图像修复样例

Baseline	MFAF	DWSC	PAM	样例1	样例2
+	-	-	-		
+	+	-	-		
+	-	+	-		
+	-	+	+		
+	+	+	+		

卷积之后，图像整体画质明显提升，但图像的关键细节仍存在模糊。增加关键细节信息补充，添加PAM模块后关注显著细节特征，模型修复的图像弥补了细节模糊的不足，人脸轮廓边缘更为清晰。最后，在完整模型的修复样例中，可以明显看出修复后的图像在亮度、对比度及相似度上体现较好的效果。

以上实验结果证明了本文所提每个模块在本文模型中的有效性。首先，使用多尺度特征注意力融合模块可为模型在上采样过程中提供丰富的语义信息，提升模型性能；其次，使用深度可分离卷积替换原有卷积，一方面增加了模型深度和提升模型的特征提取表征能力，另一方面降低了模型的参数量；最后，在一定程度上使用位置注意力模块，可动态加强特征图中的显著位置信息，提升模型处理细节模糊的复原效果。因此，使用本文所提模块，可使LM-UNET模型以较少的参数量明显提高修复图像质量。

2.4　对比实验

为了验证所提网络模型LM-UNET对遮挡人脸图像的修复的有效性和准确性，将LM-UNET与GC[16]、MADF[17]、RFR-NET[18]、PEN-NET[10]、AOT-GAN[19]、EIGC[20]等6种模型在同一数据集上进行定性和定量对比实验，从PSNR、SSIM和Params三方面进行客观评价，PANR和SSIM值均越大越好，Params越低越好，对比实验评价结果见表3。

表3　对比实验评价结果

模型名称	PSNR/dB	SSIM	Params/10^6
GC	25.57	89.96%	38.69
MADF	29.77	96.98%	324.83
RFR-NET	28.23	94.15%	119.14
PEN-NET	26.10	91.30%	39.03
AOT-GAN	27.24	90.62%	87.72
EIGC	28.44	94.74%	26.77
LM-UNET	30.49	96.85%	12.24

由表3可知，本文所提网络模型LM-UNET的PSNR值均超过其他6种模型，SSIM值仅次于MADF，且仅有12.24×10^6的参数量。MADF模型与本文所提模型相比SSIM值高0.13%，但PSNR值低0.72 dB，且该模型参数量很大。本文所提模型与RFR-NET和EIGC模型相比PSNR分别高2.26 dB和2.05 dB，SSIM分别高2.70%和2.11%，Params分别低106.9×10^6和14.53×10^6。从以上分析可知本文提出的图像修复模型已取得较好的修复效果。

人脸图像修复效果对比如图9所示，图9展示了10%~50%不同遮挡程度下不同模型的人脸图像修复效果，所提模型LM-UNET的修复效果明显好于其他模型。GC模型生成的图像画质较为模糊，质量低下，分辨率较低；AOT-GAN模型生成的图像存在局部纹理模糊，第3行的眼睛存在白色色块，第4行的牙齿和第5行的眼睛都存在模糊；PEN-NET和EIGC模型生成的图像存在纹理伪影和边界区域修复不清晰；MADF模型生成图像较好，但依然存在图像细节模糊，个别图像会出现颜色提亮且计算时间长。本文所提网络模型LM-UNET生成的图像相较其他网络模型

图9　人脸图像修复效果对比

纹理细节修复效果更清晰，修复视觉效果更好，图像修复质量更高，且使用轻量化卷积模块进一步降低了计算成本。

3　结束语

本文提出一种改进U-Net的多尺度特征融合的轻量化人脸图像修复模型——LM-UNET。利用深度可分离卷积模块替代原有卷积结构，增强模型通道间和上下文特征的表达能力，减少网络模型的参数量，实现轻量化。在残差跳跃过程中，添加了多尺度特征注意力融合模块，实现不同尺度特征信息充分融合，内部增加残差块，减少浅层特征与深层特征的语义差距。最后通过引入位置注意力模块，增强人脸图像的显著信息，提升模型对人脸位置像素信息的有效提取能力。实验证明，本文所提网络模型在人脸图像复原任务中取得较好的修复效果。目前深度学习的人脸图像修复模型可取得较好的修复性能，但大多使用的数据集图像是基于欧洲人脸的图像，而欧洲人脸较亚洲人脸更为立体。因此，如何基于亚洲人脸遮挡图像进行修复，并将修复融于人脸表情识别处理中，提高人脸表情识别的准确率，仍是值得深入研究的一个方向。

参考文献：

[1] BALLESTER C, BERTALMIO M, CASELLES V, et al. Filling-in by joint interpolation of vector fields and gray levels[J]. IEEE Transactions on Image Processing, 2001, 10(8): 1200-1211.

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

[3] 曹燕, 金炜, 符冉迪. 基于加权优先级和分类匹配的图像修复方法[J]. 电信科学, 2017, 33(4): 94-100.
CAO Y, JIN W, FU R D. Image restoration method based on weighted prioritization and classification matching[J]. Telecommunication Science, 2017, 33(4): 94-100.

[4] 李悦, 钱亚冠, 关晓惠, 等. 面向人脸识别的口罩区域修复算法[J]. 电信科学, 2021, 37(8): 66-76.
LI Y, QIAN Y G, GUAN X H, et al. A mask region repair algorithm for face recognition[J]. Telecommunication Science, 2021, 37(8): 66-76.

[5] 陈北京, 王鹏, 喻乐延, 等. 注意力融合双流特征的局部GAN生成人脸检测算法[J]. 东南大学学报（自然科学版）, 2023, 53(3): 543-551.
CHEN B J, WANG P, YU L Y, et al. Localized GAN-generated face detection algorithm for attention-fused dual-stream features[J]. Journal of Southeast University (Natural Science Edition), 2023, 53(3): 543-551.

[6] 卢奇. 局部信息缺失的RGB-D人脸修复及识别[D]. 成都: 西南交通大学, 2022.

LU Q. RGB-D face restoration and recognition with localized information loss[D]. Chengdu: Southwest Jiaotong University, 2022.

[7] ZHENG C, CHAMT J, CAI J. TFill: image completion via a transformer-basedarchitecture[J].2021.arXiv:2104.00845,2021.

[8] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234-241.

[9] LIU H Y, JIANG B, XIAO Y, et al. Coherent semantic attention for image inpainting[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 4170-4179.

[10] ZENG Y H, FU J L, CHAO H Y, et al. Learning pyramid context encoder network for high-quality image inpainting[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 1486-1494.

[11] 罗仕胜, 陈明举, 陈柳,等. 基于面部特征点的人脸图像修复网络[J]. 中国科技论文, 2021, 16(7): 729-734, 742.
LUO S S, CHEN M J, CHEN L, et al. Face image restoration network based on facial feature points[J]. Chinese Science and Technology Paper, 2021, 16(7): 729-734, 742.

[12] QIN J, BAI H, ZHAO Y. Multi-scale attention network for image inpainting[J]. Computer Vision and Image Understanding, 2021, 204: 103155.

[13] WANG N, ZHANG Y, ZHANG L. Dynamic selection network for image inpainting[J]. IEEE Transactions on Image Processing, 2021, 30: 1784-1798.

[14] LIAO L, XIAO J, WANG Z, et al. Image inpainting guided by coherence priors of semantics and textures[C]//Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2021: 6539-6548.

[15] GUO Y, LI Y, WANG L, et al. Depthwise convolution is all you need for learning multiple visual domains[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019: 8368-8375.

[16] YU J, LIN Z, YANG J, et al. Free-form image inpainting with gated convolution[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 4470-4479.

[17] ZHU M, HE D, LI X, et al. Image inpainting by end-to-end cascaded refinement with mask awareness[J]. IEEE Transactions on Image Processing, 2021, 30: 4855-4866.

[18] LI J, WANG N, ZHANG L, et al. Recurrent feature reasoning for image inpainting[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 7757-7765.

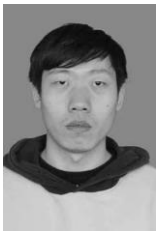
[19] ZENG Y H, FU J L, CHAO H Y, et al. Aggregated contextual transformations for high-resolution image inpainting[J]. IEEE Transactions on Visualization and Computer Graphics, 2023, 29 (7): 3266-3280.

[20] WANG F P, LI W L, LIU Y, et al. Face inpainting algorithm combining edge information with gated convolution[J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(1): 150-162.

[作者简介]

赵晓（1978− ），女，博士，陕西科技大学电子信息与人工智能学院副教授，主要研究方向为数字图像处理、模式识别等。

赵子怡（1999− ），女，陕西科技大学硕士生，主要研究方向为图像修复、图像识别。



杨晨（1997− ），男，陕西科技大学硕士生，主要研究方向为图像处理、图像识别。