

一种基于快速图像去模糊的视觉室内定位方法

吕梦妍¹ 郭新新¹ 张坤鹏¹ 张立晔¹

LYU Mengyan GUO Xinxin ZHANG Kunpeng ZHANG Liye

摘要

在图像拍摄过程中, 相机抖动、物体运动等因素导致图像产生模糊, 进而使得模糊图像的特征点减少、匹配效果较差, 最终对室内定位精度产生负面影响。为解决这一问题, 提出一种基于快速图像模糊识别和模糊去除的视觉室内定位方法。首先, 采用编解码器结构学习上下文信息, 随后与高分辨率分支结合以保留空间细节, 最终生成高质量的清晰的图像, 消除图像模糊对定位精度的影响。接着, 通过图像检索、特征点提取和匹配的方式计算查询相机的位置。实验结果表明, 所提出的方法在 GoPro 数据集上峰值信噪比 (PSNR) 为 31.64, 结构相似性指数 (SSIM) 为 0.949, 平均去模糊处理时间为 24 ms; 去模糊处理后, 85% 的室内查询图片定位误差小于 0.5 m。基于快速图像去模糊的视觉室内定位方法能够有效恢复图像细节, 提高室内视觉定位的精度。

关键词

视觉定位; 图像去模糊; 图像检索; 五点法; 室内定位; 位置估计

doi: 10.3969/j.issn.1672-9528.2024.04.003

0 引言

近年来, 随着信息技术和智能移动设备的发展, 室内定位需求日益增加。基于视觉的室内定位具有稳定性持久、成本低廉等优势, 逐渐成为研究热点。

室内视觉定位系统通常分为离线阶段和在线阶段。离线阶段在已知位置捕获图像数据, 构建离线数据库, 在线阶段对用户查询进行图像检索、匹配以及位置求解得到用户位置。然而在图像获取过程中, 物体快速运动、光线的复杂性以及相机抖动等因素可能会导致拍摄的图像质量低、失真严重, 这对后续的定位任务造成严重影响。查询图片模糊往往导致定位失败。因此, 对于数据库图像模糊处理, 对查询图像进行图像模糊判断、图像模糊去除, 能够有效提高图像的视觉质量, 进一步提升视觉室内定位方法的鲁棒性。图像去模糊研究对于室内定位具有十分重要的意义。

1 相关工作

1.1 图像去模糊

图像去模糊任务是从一张模糊的图像中复原出清晰的图像, 并尽可能真实地复原图像细节。根据复原任务中的模

糊核是否已知, 可将其分为盲去模糊和非盲去模糊两类。早期的研究主要是非盲去模糊任务, 即在已知模糊核的情况下使图像清晰化。但在实际应用中, 大多数情况下图像的模糊核是未知的。盲去模糊方法对模糊核进行估计, 核心是确定参数, 在找到最优参数后, 估计模糊核。文献 [1-2] 通过引入生成对抗网络, 先后提出 DeBlurGAN 和 DeBlurGANv2 两种网络模型, 借用生成对抗网络恢复高质量的清晰图像。文献 [3] 利用不同尺度的模糊图像同时进行高频制导去模糊处理。

1.2 视觉室内定位

基于视觉的室内定位方法是指智能设备利用图像信息感知自身位置^[4]。视觉定位方法包括基于结构和基于回归的方法。基于结构的定位方法利用查询图像的局部特征与 3D 模型匹配来建立对应关系, 去除异常值后得到查询图像的相机位置。文献 [5] 将场景的表面分割成地标块预测。基于回归的定位利用神经网络回归器得到查询图像的相机姿态。文献 [6] 使用稀疏特征描述符来回归相机姿态。文献 [7] 利用多任务联合损失函数, 计算图像位置。

然而, 现有的图像去模糊方法的性能和实时性并不能很好地满足室内定位任务需求。因此, 本文提出了一种基于快速图像去模糊的室内定位方法: 对查询图像首先进行快速图像模糊判别、模糊去除, 然后检索得到粗略位置, 最后精准计算室内位置。

1. 山东理工大学计算机科学与技术学院 山东淄博 255000
[基金项目] 国家自然科学基金青年基金: 基于图像辅助的 WiFi 室内定位系统研究, No. 62001272; 山东省自然科学基金: 基于超混沌同步控制的移动机器人全覆盖路径规划研究, No.ZR2023MF015

2 基于图像去模糊的视觉室内定位方法

基于快速图像去模糊的室内定位方法如图1。离线阶段，由手机摄像头从目标场景中收集图像，通过预训练的神经网络模型提取图像全局特征，构建离线特征数据库。在线阶段，对用户查询首先进行模糊判断，若模糊则模糊去除，若为清晰图像进行检索，检索得到最为相似的若干张数据库图片，然后对于每一对查询-数据库图像，利用极线约束、基于五点法计算本质矩阵 E ，分解 E 得到相对位置，进而恢复查询的绝对位置，即室内用户位置。

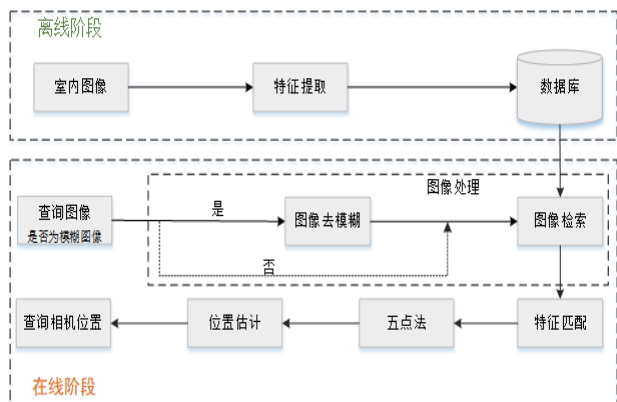


图1 视觉室内定位框架

2.1 基于双分支结构的图像去模糊网络

2.1.1 图像模糊

图像去模糊旨在将模糊的输入图像恢复成潜在清晰的图像。一般认为图像的退化过程具有空间不变性，图像模糊的数字模型表示为：

$$I_B = K * I_S + N \quad (1)$$

式中： I_B 为模糊图像， I_S 为原始清晰图像， K 为模糊核， $*$ 为卷积操作， N 为添加的噪声。

2.1.2 图像模糊识别

本文利用清晰图像和模糊图像的梯度差异，提出了一种图像模糊识别方法：

$$B = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \left(\frac{(g_x(i,j))^2 + (g_y(i,j))^2}{2} \right)^{1/2} \times \frac{G_{num}}{N} \quad (2)$$

式中： H 表示图像的高度， W 表示图像的宽度， $g_x(i,j) = |f(i+1,j) - f(i,j)|$ 为图像在 X 方向上的梯度映射， $g_y(i,j) = |f(i,j+1) - f(i,j)|$ 为图像在 Y 方向上的梯度映射， G_{num} 为 X 方向梯度映射和 Y 方向梯度映射中非零梯度值的个数之和， N 为 X 方向和 Y 方向梯度图中梯度值的个数之和。

根据式(2)计算查询图像的模糊识别数值 B ，若 B 大于给定阈值 T ，则认为查询为清晰图像，不需要进行图像复原。否则，查询为模糊图像，需要进行模糊去除。

2.1.3 网络模型

图像去模糊网络如图2，模型由两部分组成，首先使用一个单编码器-双解码器架构学习上下文信息，然后与保留空间细节的高分辨率分支结合，得到清晰图像。

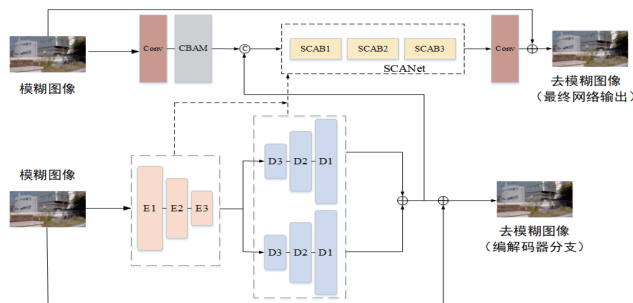


图2 双分支结构图像去模糊网络

通过一个编码器-两个解码器子网络 F 从模糊图像中估计残差图像，其表示为：

$$\hat{r} = D_1(z; \Theta_{D1}) + D_2(z; \Theta_{D2}) \quad (3)$$

式中： D_1 和 D_2 是两个独立的具有相同结构的解码器网络， Θ_{D1} 和 Θ_{D2} 为 D_1 和 D_2 的网络参数， z 为参数 Θ_e 的编码器 e 编码后的图像特征：

$$z = \varepsilon(I_B, \Theta_e) \quad (4)$$

根据线性跨度理论，通过两个独立的解码器对编码特征中的模糊退化信息进行强制分割，即一个解码器沿主轴方向产生模糊残差，另一个解码器产生互补残差，完成线性重构，输出以最大解空间。

为了避免无监督下发生信息解耦现象导致的网络性能降低，利用分离解码器的特性，将两个解码器的网络参数共享为：

$$\hat{r} = D_1(z; \Theta_d) + D_2(z; \Theta_{rd}) \quad (5)$$

式中： Θ_{rd} 是 Θ_d 逆时针旋转90度后的网络参数， D_1 为主解码器。

引入保留空间细节的高分辨率子网络，与编解码器分支相结合。卷积注意模块的结构如图3(a)，包括通道和空间注意两部分。

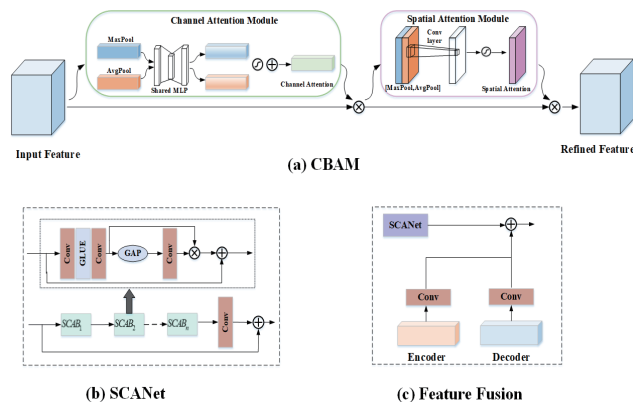


图3 CBAM、SCANet体系结构以及编解码器分支与SCANet特征融合图

通道注意模块, 通过平均池化和最大池化对特征图的空间信息进行聚合, 生成两个不同的空间上下文描述符: F_{avg}^c 和 F_{max}^c 。然后, 这两个描述符被转发到一个共享网络, 生成通道注意力图 $M_c \in R^{C \times 1 \times 1}$ 。该共享网络由多层感知器和一个隐藏层组成, 隐藏激活为 $R^{R/r \times 1 \times 1}$, R 为缩减比。经共享网络, 元素求和合并输出特征向量:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (6)$$

式中: σ 为 Sigmoid 函数, 多层感知机 MLP 共享权重。

空间注意模块, 利用特征的空间关系生成空间注意力图。对于平均池化和最大池化后生成的 2D 地图: $F_{avg}^s \in R^{1 \times H \times W}$ 和 $F_{max}^s \in R^{1 \times H \times W}$, 经过标准卷积层进行连接和卷积, 生成二维空间注意力图 $M_s(F) \in R^{H \times W}$ 。空间注意力可表示为:

$$M_s(F) = \sigma(f([AvgPool(F); MaxPool(F)])) \quad (7)$$

输入的中间层特征 $F \in R^{C \times H \times W}$, 经卷积块注意模块 CBAM (convolutional block attention module) 依次得到一维通道注意力图 $M_c \in R^{C \times 1 \times 1}$ 和二维空间注意力图 $M_s \in R^{1 \times H \times W}$ 。CBAM 过程表示如下:

$$F_1 = M_c(F) \otimes F \quad (8)$$

$$F_2 = M_s(F_1) \otimes F_1$$

式中: $\otimes$ 表示逐元素乘法, F_2 为最终的输出。

通道注意子网络由若干个简化的通道注意块组成的, 以保证图像细节信息。其如图 3 (b) 所示, 输入特征 F_i 将通道注意块得到输出 F_o , 定义如下:

$$\begin{aligned} F_o &= F_i + F_{i1} \otimes \text{Conv}(GAP(F_{i1})) \\ F_{i1} &= \text{Conv}(GELU(\text{Conv}(F_i))) \end{aligned} \quad (9)$$

式中: GAP 表示全局平均池化, Conv 表示卷积。

如图 3 (c) 所示, 将 CBAM 得到的输出 F_2 和解码器的输出进行融合, 特征融合后输入通道注意子网络。特征融合信息 FF_i 定义如下:

$$\begin{aligned} FF_i &= CABG_i(FF_{i-1}) + X_{Di} + X_{Ei} \\ FF_0 &= [F_2, X_{D3}] \end{aligned} \quad (10)$$

式中: X_{Ei} 为第 i 级编码器输出, X_{Di} 为第 i 级两解码器输出之和 ($i=1,2,3$)。

2.1.4 损失函数

双分支图像去模糊模型预测残差图像, 以端到端的方式优化, 损失函数定义如下:

$$\begin{aligned} L &= \sum_{S=1}^2 [L_{char}(X_S, X_{gt}) + \lambda L_{edge}(X_S, X_{gt})] \\ L_{char} &= \sqrt{\|X_S - X_{gt}\|^2 + \varepsilon^2} \\ L_{edge} &= \sqrt{\|\Delta(X_S) - \Delta(X_{gt})\|^2 + \varepsilon^2} \end{aligned} \quad (11)$$

式中: X_S 由 S 分支预测的残差图加上输入的模糊图像 I_B 得到,

X_{gt} 为清晰图像的真值, L_{char} 为 Charbonnier 损失^[8], ε 的大小为 10^{-3} , L_{edge} 为边缘损失, Δ 为拉普拉斯算子, λ 的大小设为 0.5。

2.2 位置估计

通过图像检索返回与查询图像相似度分数最高的若干张数据库图像, 根据核线约束, 通过图像间特征点匹配计算得出位置信息。

2.2.1 相对位置计算

归一化平面的匹配点对 $x_1 = [u_1, v_1, 1]^T$ 和 $x_2 = [u_2, v_2, 1]^T$, 依据对极几何约束计算如下:

$$[u_1 u_2, u_1 v_2, u_1, v_1 u_2, v_1 v_2, v_1, u_2, v_2, 1] E = 0 \quad (12)$$

同样表示其他的点对, 利用奇异值分解可以求出目标 - 数据库图像间的相对位置 R 、 t 。

2.2.2 绝对位置计算

根据两图像对 (I_q, I_i) 、 (I_q, I_j) 估计的查询的绝对旋转, 由两条射线 l_1 、 l_2 的交点计算查询图像坐标:

$$l_1 = c_{l_1} + \lambda_1 R_{l_1}^T R_i t_i \quad (13)$$

$$l_2 = c_{l_2} + \lambda_2 R_{l_2}^T R_j t_j \quad (14)$$

式中: $\lambda_1, \lambda_2 \in R$ 定义了沿射线的点位置。只有当三个摄影机中心不共线时, 才会定义三角测量的结果。 $c_{l_1} = -R_{l_1}^T t_1$ 是相机中心的全局坐标。在实验中, 使用查询到的三个数据库图像计算最终姿势。

通过两个图像对 (I_q, I_i) 、 (I_q, I_j) 得到一个绝对位置的假设 (R_{l_q}, t_{l_q}) , 对任意查询 - 数据库图像对 (I_q, I_m) 的四个可能的位置, 找到 $R_m R_{l_m}$ 的值最接近 R_{l_q} 的 R_m , 从 I_q 到 I_m 预测的相对平移:

$$t_{pre} = R_{l_m} (c_{l_q} - c_{l_m}) \quad (15)$$

$$\alpha = \cos^{-1} \left(\frac{t_m^T t_{pre}}{\|t_m\|_2 \|t_{pre}\|_2} \right) \quad (16)$$

式 (16) 为阈值 α 的定义, 当预测的平移方向之间的角度小于给定的阈值 α 时, 计为一个 inlier。统计所有图像对中该位置假设的 inlier, 选择 inlier 最多的假设作为最终的输出。

3 实验结果与分析

本文提出的双分支结构图像去模糊算法在 GoPro 中进行训练。在 GoPro-test^[9] 和 RealBlur-J^[10] 数据集上测试, 实验结果表明, 本文提出的方法在精度和运行时间上都获得了具有竞争力的结果。

3.1 图像去模糊性能评价

本文根据峰值信噪比 (PSNR) 和结构相似 (SSIM) 两个评价指标与其他图像去模糊算法比较。表 1 为 GoPro 数据

集和 RealBlur-J 数据集的测试结果, 在 GoPro 数据集上进行训练, 直接将训练模型应用到 RealBlur-J 数据集以评估对真实图像泛化性。

表 1 去模糊定性结果

	GoPro		RealBlur-J	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
DSSN	30.69	0.939	28.41	0.845
MSCAN	31.24	0.945	28.82	0.862
SimpleNe ^[13]	31.52	0.950	26.95	0.864
Our	31.64	0.949	28.87	0.886

从表 1 中可以看出, 本文提出的基于双分支结构的图像去模糊网络表现良好。在 GoPro 数据集进行训练, 直接迁移应用到 RealBlur-J 数据集时也获得较为突出的结果。此外, 本文方法不仅在图像质量评价方面取得较好效果, 模型大小和平均运行时间方面也极具竞争力, 如表 2, 双分支结构的图像去模糊模型为 8.2 MB, 平均处理时间为 24 ms。

表 2 模型大小和平均运行时间的比较

方法	模型大小 /MB	运行时间 /ms
DSSN	25.2	30
MSCAN	7.5	140
SimpleNe	25.1	376
Our	8.2	24

图 4 是室内图片去模糊效果, 其中 (a) 是输入的模糊图像, (b) 是 DSSN^[11] 算法, (c) 是 MSCAN^[12] 算法效果, (d) 是本文提出方法。总体来说, 本文提出的模型恢复图像比其他模型更清晰, 更接近真实图像。

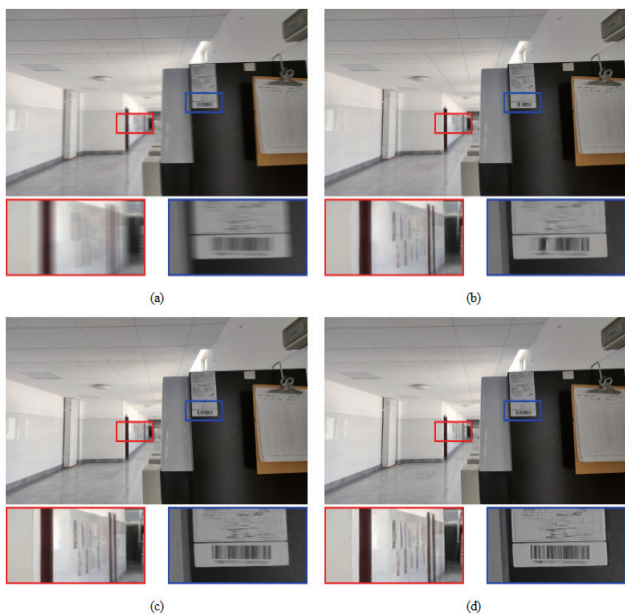


图 4 室内定位数据集去模糊效果

3.2 室内定位结果集性能评价

室内定位数据库是在山东理工大学 9 号教学楼第 4 层所拍摄的走廊图像。为了评估所提出方法的定位精度, 在实验过程中以相等的间隔对室内地面进行网格化, 每个网格间距为 0.5 m。将查询相机放置在三脚架上, 在拍摄查询图像时将三脚架放置在网格交点处, 在采集查询图像的同时记录查询相机的地面实况位置。

在基于视觉的室内定位任务中, 计算的相机姿态与地面实况姿态进行比较, 测量估计结果与地面真值的接近程度。具体来说, 位置误差被表示为查询图像估计的位置和记录的真值之间的以 m 为单位的欧式距离, 具体表示为:

$$t_{abs_err} = \|t_{abs_gt} - t_{abs_pre}\|_2 \quad (17)$$

另外, 本文分析了教学楼走廊场景下去模糊算法对室内定位性能的影响。从图 5 中可以看到, 模糊图像中提取的特征点数量明显少于去模糊图像。当使用模糊图像进行室内定位时, 无法实现图像之间的精确匹配, 定位精度大大降低。在特征点较少、场景信息较差的走廊场景中, 往往无法实现图像匹配, 导致定位失败。

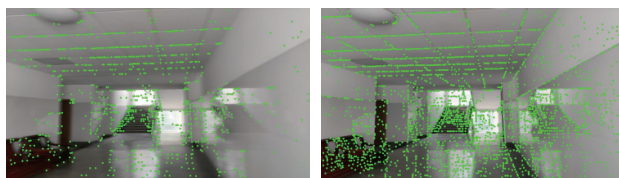


图 5 走廊模糊图像与去模糊图像特征提取的性能比较

实验结果表明, 基于图像去模糊的室内定位方法可以实现更好的定位效果, 其平均误差与最大误差都远远优于模糊数据。图 6 直观显示了去模糊算法对走廊数据集定位精度的影响: 使用模糊查询图计算室内位置时, 室内数据集中的平均定位误差为 0.39 m, 最大定位误差大于 1.5 m; 而使用去模糊后的室内图片进行位置估计时, 平均定位误差仅为 0.21 m。同时, 本文提出的图像去模糊网络模型较小, 运行速度较快, 满足室内环境的定位任务需求。

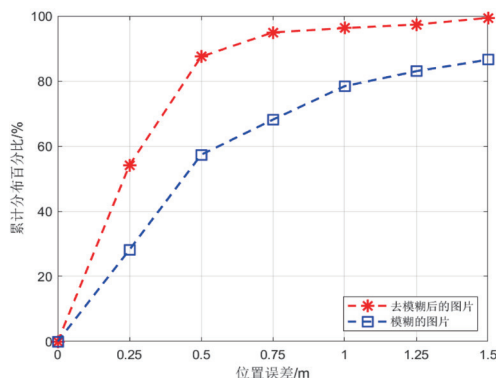


图 6 室内定位误差

4 结语

为解决相机抖动等因素造成的图像模糊问题对室内定位产生的不利影响,本文提出了一种基于高效图像去模糊的视觉室内定位方法。图像去模糊网络利用双分支结构逐步恢复模糊图像的上下文信息和空间细节,利用参数共享方案,减少网络参数的数量。首先对于查询图像进行模糊判断,若为模糊图像,则进行快速模糊处理、图像恢复后再进行图像检索;若为清晰图像,则直接进行图像检索过程。然后,对于查询图像和数据库中的相似图像进行特征提取匹配,利用本质矩阵求解得到查询相机的位置信息。实验结果验证了该方法的有效性,经模糊处理后的室内查询中,85%的室内查询图片定位误差小于0.5 m。

参考文献:

- [1] KUPYN O, BUDZAN V, MYKHAILYCH M, et al. Deblurgan: blind motion deblurring using conditional adversarial networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8183-8192.
- [2] KUPYN O, MARTYNIUK T, WU J, et al. Deblurgan-v2: deblurring (orders-of-magnitude) faster and better[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 8878-8887.
- [3] LIU Y, FANG F, WANG T, et al. Multi-scale grid network for image deblurring with high-frequency guidance[J]. IEEE transactions on multimedia, 2021, 24: 2890-2901.
- [4] 蔡旭东, 王永才, 白雪薇, 等. 基于先验地图的视觉重定位方法综述 [J]. 软件学报, 2023: 1-35.
- [5] HUANG Z, ZHOU H, LI Y, et al. Vs-net: voting with segmentation for visual localization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 6101-6111.
- [6] BACH T B, DINH T T, LEE J H. FeatLoc: absolute pose regressor for indoor 2D sparse features with simplistic view synthesizing[J]. ISPRS journal of photogrammetry and remote sensing, 2022, 189: 50-62.
- [7] 董思强, 邓年茂. 基于几何约束孪生卷积网络的相机6DOF定位研究 [J]. 系统工程与电子技术, 2023, 45(6): 1828-1835.
- [8] CHARBONNIER P, BLANC-FERAUD L, AUBERT G, et al. Two deterministic half-quadratic regularization algorithms for computed imaging[C]//Proceedings of 1st International Conference on Image Processing. Piscataway: IEEE, 1994, 2: 168-172.
- [9] NAH S, HYUN KIM T, MU LEE K. Deep multi-scale convolutional neural network for dynamic scene deblurring[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 3883-3891.
- [10] RIM J, LEE H, WON J, et al. Real-world blur dataset for learning and benchmarking deblurring algorithms[C]//Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XXV 16. Cham: Springer, 2020: 184-201.
- [11] LIM S, KIM J, KIM W. Deep spectral-spatial network for single image deblurring[J]. IEEE signal processing letters, 2020, 27: 835-839.
- [12] WAN S, TANG S, XIE X, et al. Deep convolutional-neural-network-based channel attention for single image dynamic scene blind deblurring[J]. IEEE transactions on circuits and systems for video technology, 2020, 31(8): 2994-3009.
- [13] LI J, TAN W, YAN B. Perceptual variousness motion deblurring with light global context refinement[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 4116-4125.

【作者简介】

吕梦妍(1999—), 女, 山东济宁人, 硕士研究生, 研究方向: 图像处理、视觉定位。

郭新新(1999—), 女, 山东滨州人, 硕士研究生, 研究方向: 同步定位与建图。

张坤鹏(1998—), 男, 山东淄博人, 硕士研究生, 研究方向: 视频理解。

张立晔(1986—), 通信作者(email: zhangliye@sdu.edu), 男, 山东济南人, 博士, 副教授, 研究方向: 机器学习、图像处理、室内定位。

(收稿日期: 2024-01-18)