CMAMRNet: A Contextual Mask-Aware Network Enhancing Mural Restoration Through Comprehensive Mask Guidance

Yingtie Lei, Fanghai Yi, Yihang Dong, Weihuang Liu, Xiaofeng Zhang, Zimeng Li, Chi-Man Pun, Xuhang Chen*

University of Macau

Abstract

Murals, as invaluable cultural artifacts, face continuous deterioration from environmental factors and human activities. Digital restoration of murals faces unique challenges due to their complex degradation patterns and the critical need to preserve artistic authenticity. Existing learning-based methods struggle with maintaining consistent mask guidance throughout their networks, leading to insufficient focus on damaged regions and compromised restoration quality. We propose CMAMRNet, a Contextual Mask-Aware Mural Restoration Network that addresses these limitations through comprehensive mask guidance and multi-scale feature extraction. Our framework introduces two key components: (1) the Mask-Aware Up/Down-Sampler (MAUDS), which ensures consistent mask sensitivity across resolution scales through dedicated channel-wise feature selection and mask-guided feature fusion; and (2) the Co-Feature Aggregator (CFA), operating at both the highest and lowest resolutions to extract complementary features for capturing fine textures and global structures in degraded regions. Experimental results on benchmark datasets demonstrate that CMAMRNet outperforms state-of-the-art methods, effectively preserving both structural integrity and artistic details in restored murals. The code is available at https://github.com/CXH-Research/CMAMRNet.

Methods

(1) We propose CMAMRNet, a novel mask-aware framework for mural restoration that maintains comprehensive mask guidance throughout the processing pipeline while effectively capturing multiscale contextual features through transformer architectures.
(2) We propose MAUDS that integrates mask information through dedicated channel operations, ensuring consistent mask sensitivity across resolution scales and preventing diminishing mask influence in deep networks.
(3) We introduce CFA operating at highest and lowest resolutions, employing parallel CFFB and SFFB modules to extract complementary features that capture both fine-grained textures and global structures in degraded regions.

Results

Table 1:Quantitative results of comparative experiments with state-of-the-art methods. The best results are shown in **bold** and the second best results are underlined.

Method	MuralDH				Dunhuang			
	PSNR ↑	SSIM ↑	MAE ↓	LPIPS ↓	PSNR ↑	SSIM ↑	MAE ↓	LPIPS ↓
RN	14.3577	0.2712	42.0932	0.9877	17.9298	0.5401	23.9331	0.7688
DeepFillV2	14.4970	0.2734	41.5719	1.0131	16.2900	0.5002	30.8472	0.8756
HFill	14.5221	0.2741	41.5619	1.0232	17.0136	0.5058	26.3423	0.9000
CoordFill	23.6762	0.5706	12.2239	0.3160	20.4858	0.4134	16.1557	0.5367
SyFormer	24.5005	0.6328	11.3667	0.6102	26.0647	0.7156	8.8938	0.4659
LaMa	28.4603	0.9195	6.6463	0.0861	29.5888	0.8346	5.8470	0.1067
EdgeConnect	29.9298	0.9427	5.4444	0.0788	30.7807	0.8519	4.9383	0.1045
T-Former	30.9527	0.9564	4.5003	0.0729	31.5000	0.8608	4.6571	0.0854
DWTNet	31.0443	0.9565	4.4492	0.0726	31.5097	0.8609	4.6609	0.0855
Misf	34.1878	0.9635	1.8124	0.0555	32.1570	0.8701	4.4660	0.0691
DFNet	34.6129	0.9641	1.3644	0.0577	31.0897	0.8549	4.7918	0.0936
HINT	35.4833	0.9663	1.2337	0.0542	32.2421	0.8695	4.3776	0.0706
Ours	**36.2565**	**0.9683**	**1.1454**	**0.0497**	**33.0738**	**0.8944**	**3.9708**	**0.0598**

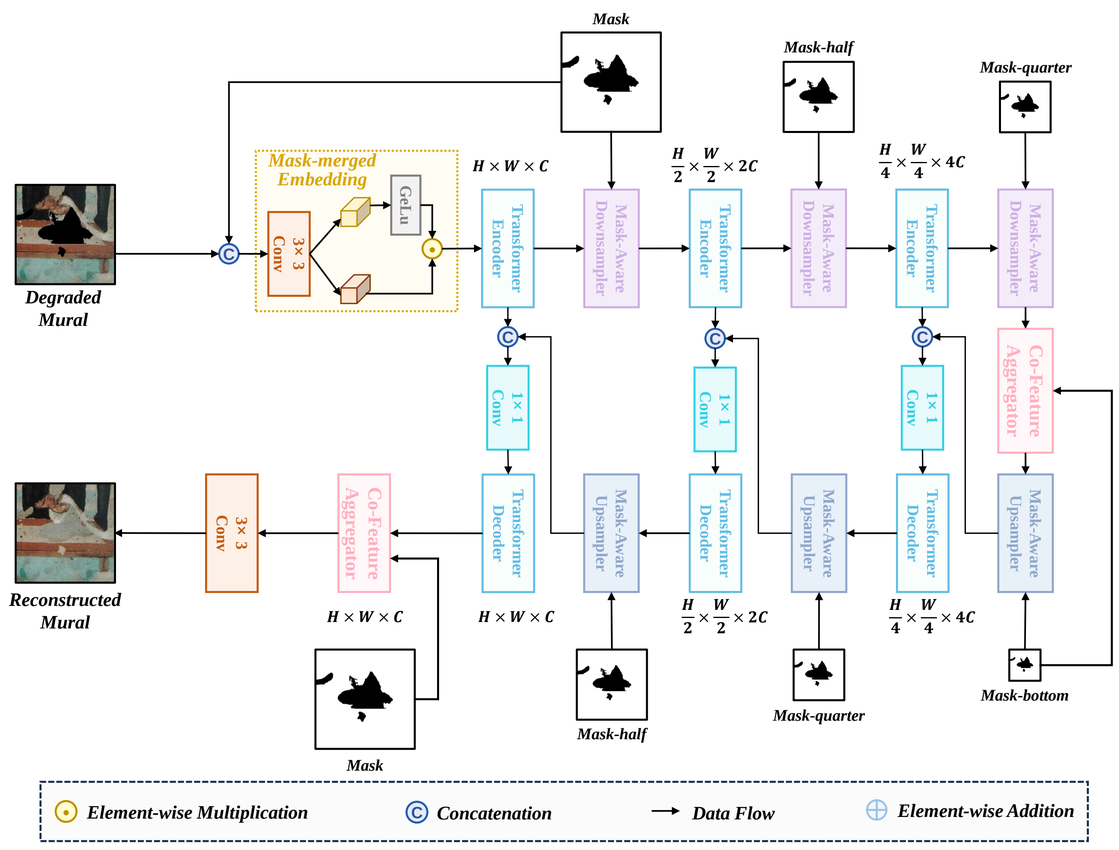

Figure 1:Architecture overview of the proposed CMAMRNet. The framework is built upon a U-shaped backbone with Transformer-based encoders and decoders for hierarchical feature extraction. It introduces two key innovations: (1) Mask-Aware Up/Down-Sampler (MAUDS) modules that preserve mask information across resolution changes via selective channel processing and mask-guided feature integration, and (2) Co-Feature Aggregator (CFA) modules operating at both highest and lowest resolutions to capture comprehensive features of fine details and global structures for holistic restoration quality.

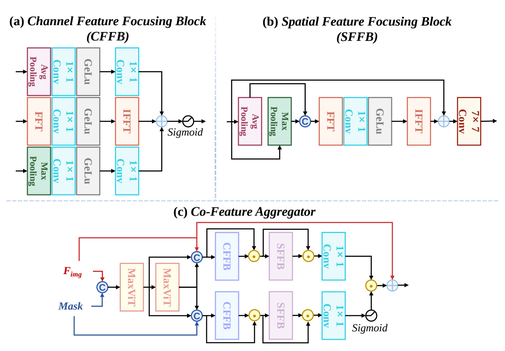

Figure 2:Architecture of: (a) the Channel Feature Focusing Block (CFFB), (b) the Spatial Feature Focusing Block (SFFB), and (c) the Co-Feature Aggregator (CFA) structure with integrated focusing blocks.

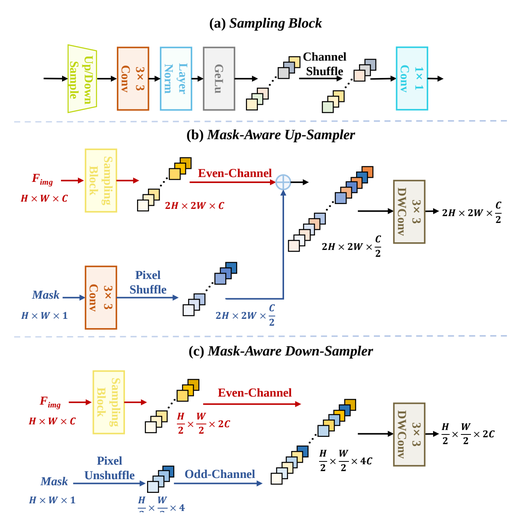

Figure 3:Overview of Mask-Aware Up/Down-Sampler (MAUDS): (a) Sampling Block, (b) Mask-Aware Down-Sampler (MADS), and (c) Mask-Aware Up-Sampler (MAUS) for mask-guided feature transitions across different resolution scales.

Acknowledgement

This work was supported in part by the Science and Technology Development Fund, Macau SAR, under Grant 0141/2023/RIA2 and 0193/2023/RIA3, and the University of Macau under Grant MYRG-GRG2024-00065-FST-UMDF, in part by the Shenzhen Polytechnic University Research Fund (Grant No. 6025310023K).