SCIENCE CHINA
Information Sciences

• LETTER •

August 2025, Vol. 68, Iss. 8, 189102:1–189102:2
https://doi.org/10.1007/s11432-024-4493-4

Orpaint: a zero-shot inpainting model for oracle bone inscription rubbings with visual mamba block

Zijie MENG[1†], Yuanze ZENG[2†], Xiang CHANG[3], Tianshuo XU[4], Fei CHAO[2,3*], Xixin CAO[1*], Changjing SHANG[3] & Qiang SHEN[3]

[1]*School of Software and Microelectronics, Peking University, Beijing 102206, China*
[2]*School of Informatics, Xiamen University, Xiamen 361005, China*
[3]*Faculty of Business and Physical Sciences, Aberystwyth University, Aberystwyth SY23 3DB, UK*
[4]*AI Thrust, INFO Hub, Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511400, China*

Received 14 July 2024/Revised 14 December 2024/Accepted 23 March 2025/Published online 8 July 2025

Citation Meng Z J, Zeng Y Z, Chang X, et al. Orpaint: a zero-shot inpainting model for oracle bone inscription rubbings with visual mamba block. Sci China Inf Sci, 2025, 68(8): 189102, https://doi.org/10.1007/s11432-024-4493-4

Oracle bone inscriptions, originating in the second millennium before common era (BCE), embody profound historical and cultural significance by revealing the evolution of ancient societal structures, religious beliefs, and the origins of modern Chinese characters [1]; however, their restoration, a critical step in Oracle bone studies, is hindered by severe degradation of rubbing images caused by natural erosion, microbial damage, and anthropogenic factors [2].

Traditional restoration methods, reliant on manual intervention and constrained to single-character restoration, lack generalizability and scalability [3], while modern neural network-based approaches, particularly those utilizing U-Net architectures with multi-head self-attention (MSA) [4], suffer from quadratic computational complexity, rendering them impractical for high-resolution image restoration due to excessive time and resource consumption. Given the trade-offs between computational efficiency and the effective receptive field in existing methods, it is imperative to develop an end-to-end, training-free, low-cost model specifically designed for Oracle bone inpainting, capable of efficient large-scale restoration without sacrificing restoration quality.

Therefore, we propose Orpaint, an end-to-end zero-shot image restoration framework based on a diffusion model, designed to efficiently restore degraded Oracle bone inscription rubbings. Orpaint leverages the reverse generative capabilities of diffusion models, which emulate the degradation process by progressively adding noise and then reversing the process to reconstruct the structural features and texture details of the original image, achieving high-quality, diverse, and generalizable restorations. Central to its design is the integration of the visual state space (VSS) block [5], enhanced by the efficient 2D scanning (ES2D) mechanism, which replaces the traditional U-Net architecture based on MSA. The contributions of this work are twofold. (1) Orpaint eliminates the need for manual intervention or pretraining, demonstrating the ability to learn feature distributions directly from large Oracle datasets and repair undeciphered inscriptions, thereby exhibiting emergent capabilities. (2) By incorporating the ES2D-based VSS block, Orpaint achieves significant reductions in time and computational costs compared to MSA-based methods, establishing itself as a highly efficient and cost-effective solution for Oracle bone inpainting.

Method. Orpaint introduces an advanced zero-shot inpainting framework tailored for Oracle bone inscription rubbings, employing a sequence of stacked Orpaint blocks. The structure of an Orpaint block, applied to a single denoising time step, is illustrated in Figure 1(a). The framework operates within the latent space over T time steps, starting with a noise sample $x_T \sim \mathcal{N}(0, I)$ and progressively denoising it. The known pixels, represented as $(1-m) \odot x_t^{\text{known}}$, and the unknown pixels, defined as $m \odot x_t^{\text{unknown}}$, are iteratively processed to reconstruct the complete image. To incorporate the resampling strategy and "jump length" j, the output of a single Orpaint block at time step t is defined as

$$x_{t-1} = (1-m) \odot x_{t-1}^{\text{known}} + m \odot \hat{x}_{t-1}^{\text{unknown}}, \qquad (1)$$

$$\hat{x}_{t-1}^{\text{unknown}} = \text{DDPM}(x_t, j), \qquad (2)$$

where x_{t-1}^{known} is sampled from the observed regions $m \odot x_0$, while $\hat{x}_{t-1}^{\text{unknown}}$ is generated from the reverse diffusion process; in addition, DDPM(x_t, j) denotes the denoising diffusion probabilistic model with a configurable jump length j, which determines the number of time steps to jump forward in the diffusion process. The resampling operation adjusts the diffusion process by forward diffusing x_{t-j} back to x_t and then reapplying reverse diffusion, progressively harmonizing the generated and known regions by leveraging

* Corresponding author (email: fchao@xmu.edu.cn, cxx@ss.pku.edu.cn)
† These authors contributed equally to this work.

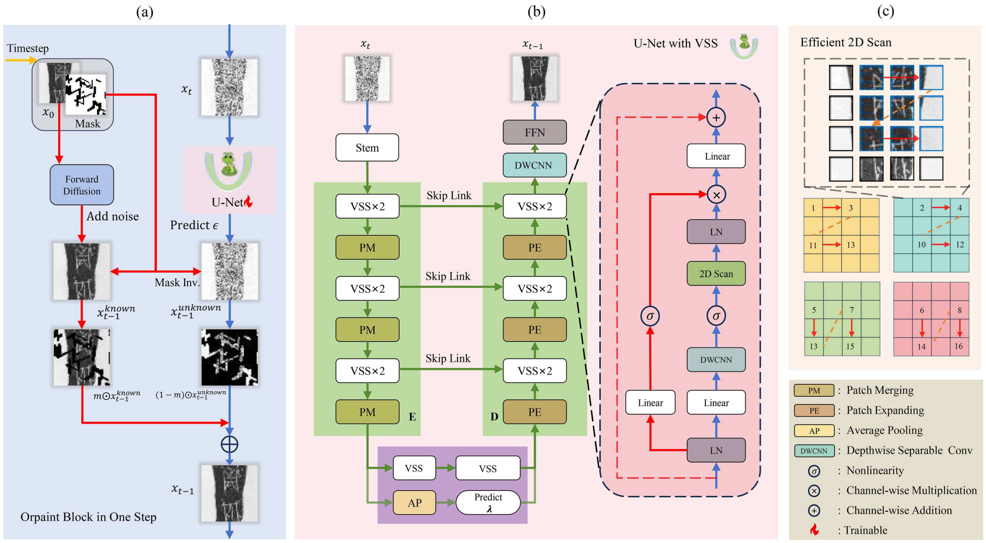

Figure 1 (Color online) Overview architecture of Orpaint. (a) Design details of the Orpaint block, where a single block performs one denoising step; (b) improved U-Net using VSS blocks with ES2D scanning mechanism, replacing traditional attention modules; (c) illustration of the ES2D scanning mechanism.

the DDPM's ability to naturally produce consistent structures. This forward-backward mechanism corrects semantic inconsistencies, refines boundaries, and enhances restoration quality through multiple iterations.

The core innovation of Orpaint lies in the ES2D mechanism (Figure 1(c) top) integrated into the VSS block within a U-Net architecture, replacing traditional MSA (Figure 1(b)). ES2D optimizes spatial scanning by adopting a sparse leap sampling strategy with a configurable stride p, reducing the time complexity from $O(N)$ to $O(N/p^2)$ while preserving contextual and directional information. Given an input feature map $X \in \mathbb{R}^{C\times H\times W}$, ES2D partitions spatial dimensions into subsets $\{\Omega_i\}_{i=1}^{4}$ and updates directional features using learnable parameters $\{\Theta_k\}$. This process, involving scan, update, and merge steps, is defined as

$$y(t) = \sum_{k=1}^{4} C_t\big(A_t h_k(t) + (B_t + \Theta_{k,t})x(t)\big) \odot z(t), \qquad (3)$$

where $h_k(t)$ denotes hidden states for direction k, A_t, B_t, and C_t are state-space parameters, and $z(t)$ is a gating signal. Integrated into the VSS block, ES2D allows efficient long-range dependency modeling with fewer processed patches, significantly enhancing the computational efficiency of hierarchical feature extraction in both encoder and decoder stages of U-Net.

Experiments and results. Three experiments were conducted to evaluate the inpainting performance, component effectiveness, and generalization capability of Orpaint. First, a comparative experiment was performed to assess inpainting quality and resource consumption under various mask types. Compared to the latest state-of-the-art zero-shot methods, Orpaint demonstrated superior fidelity and diversity, particularly for large-area masks, while achieving reduced computational complexity and faster inference. Second, an ablation study validated the effectiveness of the VSS block and ES2D scanning mechanism. For resolution expansion, Orpaint exhibited linear growth in computational complexity with increasing image size, maintaining stable inpainting quality, and outperforming in scalability at higher resolutions (e.g., 1024×1024). For 2D scanning modes, ES2D significantly reduced time complexity compared to other advanced methods while preserving competitive inpainting quality. Lastly, a generalization experiment demonstrated Orpaint's adaptability to other datasets, such as ImageNet, producing high-quality inpainting results with sufficient pertaining. Detailed experimental configurations and results are provided in Appendix C.

Discussion and conclusion. We present Orpaint, a zero-shot model for inpainting Oracle bone inscription rubbing images, built on a diffusion framework enhanced by VSS blocks and the ES2D mechanism integrated into the U-Net denoising network. Orpaint demonstrates exceptional fidelity, adaptability to complex masks, and superior parameter efficiency, sampling speed, and scalability compared to alternative architectures, showcasing the potential of leveraging advanced architectural innovations for Oracle bone inscription conservation, and inspiring further scholarly contributions to this field.

Acknowledgements This work was supported by Key Program of the National Natural Science Foundation of China Joint Fund (Grant No. U23A20383).

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Chang X, Chao F, Shang C, et al. Sundial-GAN: a cascade generative adversarial networks framework for deciphering oracle bone inscriptions. In: Proceedings of ACM International Conference on Multimedia, 2022. 1195–1203
2 Assael Y, Sommerschield T, Shillingford B, et al. Restoring and attributing ancient texts using deep neural networks. Nature, 2022, 603: 280–283
3 Fu X, Yang Z, Zeng Z, et al. Improvement of oracle bone inscription recognition accuracy: a deep learning perspective. ISPRS Int J Geo-Inform, 2022, 11: 45
4 Lugmayr A, Danelljan M, Romero A, et al. RePaint: inpainting using denoising diffusion probabilistic models. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. 11461–11471
5 Liu Y, Tian Y, Zhao Y, et al. VMamba: visual state space model. 2024. ArXiv:2401.10166