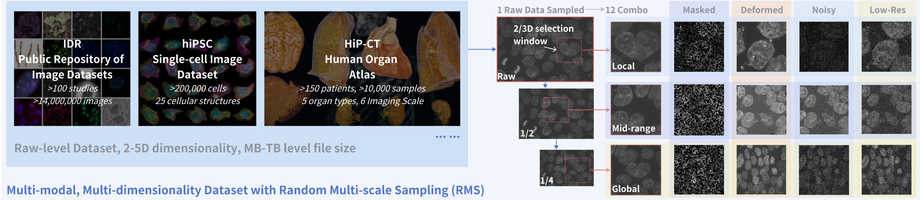
Orochi: Versatile Biomedical Image Processor

Gaole Dai *[1], Chenghao Zhou *[1,2], Yu Zhou *[3], Rongyu Zhang [1], Yuan Zhang [1], Chengkai Hou [1], Tiejun Huang [1], Jianxu Chen †[,3], Shanghang Zhang †[,1]
[1] State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University
[2] Academy for Advanced Interdisciplinary Studies, Peking University
[3] Leibniz-Institut für Analytische Wissenschaften – ISAS – e.V.
* Equal Contribution † Equal Supervision

Dataset

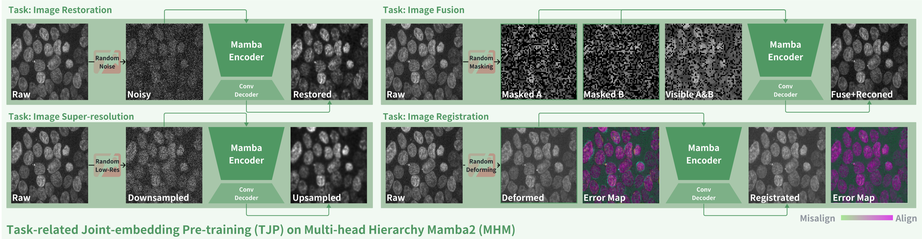
We collected unlabeled data from **over 100 studies** and preprocessed the data into a format suitable for model training. This dataset includes **multi-scale and multi-modal** images derived from diverse studies.



Multi-modal, Multi-dimensionality Dataset with Random Multi-scale Sampling (RMS)

Method

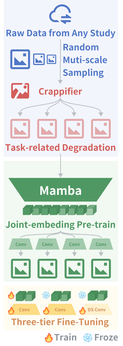
The overall concept is straightforward: we extend the Masked Image Modeling (MIM) degradation strategy employed in MAE and I-JEPA to **multiple task-specific degradation operations**—including masking, deformation, downsampling, and noise addition—and utilize a **unified encoder** paired with a lightweight decoder to reconstruct the original image.



Task-related Joint-embedding Pre-training (TJP) on Multi-head Hierarchy Mamba2 (MHM)

Performance Overview

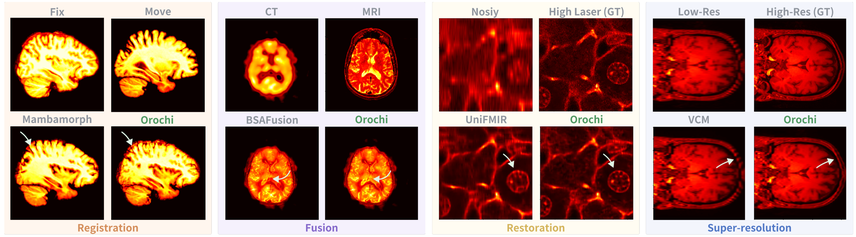
Here, we provide an overview of the fine-tuned results obtained using our pre-trained encoder. As a **versatile model**, it demonstrates strong generalization capabilities across various biomedical image processing tasks and achieves competitive performance compared to **recent state-of-the-art methods in each domain**.



	Registration	Fusion	Restoration	Super-Resolution
Dataset	OASIS	VIFB	CARE	HBA
ConvexAdam (TMI 2024)	✓ Dice 81.20	✗	✗	✗
BSAFusion (AAAI 2025)	✓ Dice 70.19	✓ Qabf 0.39	✗	✗
UniFMIR (Nature Methods 2024)	✗	✗	✓ PSNR 27.40	✓ PSNR 26.02
VCM (WACV 2025)	✗	✗	✗	✓ PSNR 27.53
Orochi (Light)	✓ Dice 79.61	✓ Qabf 0.34	✓ PSNR 29.88	✓ PSNR 32.56
Orochi (Best)	✓ Dice 83.62	✓ Qabf 0.41	✓ PSNR 29.88	✓ PSNR 33.63

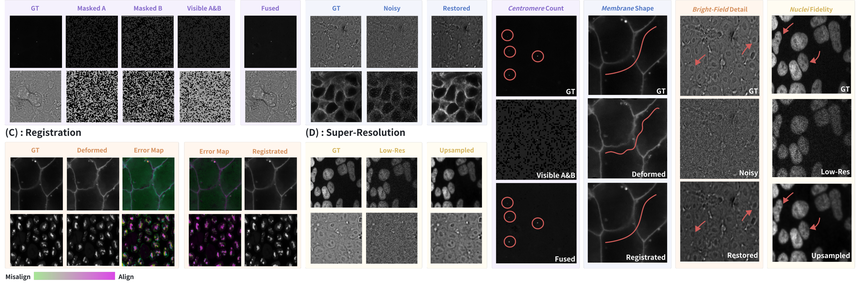
Qualitative Evaluation

Upon examining the visualization, it is evident that our model demonstrates strong capability in handling **subtle changes** and **fine-grained**, low-level degradations.



Ablation Study

We conducted an ablation study on the trade-off between fine-tuning cost and model performance, revealing that **full fine-tuning is not always required to achieve optimal results**. We present several adaptation strategies, including full fine-tuning, replacement with a dense decoder, and replacement with a lightweight decoder. For small datasets, parameter-efficient fine-tuning (PEFT) represents a highly effective alternative.



Zero-shot Ability

Here, we present several case studies demonstrating the zero-shot capabilities of our model. While we are confident in the validity of results on in-domain data, the observed task-related performance degradation still indicates a **more direct connection to downstream tasks**.