Joint EM Image Super-Resolution and Segmentation with Semantic and Structural Priors

Jiateng Shou *Haoyuan Shi* *Zhiwei Xiong*

University of Science and Technology of China

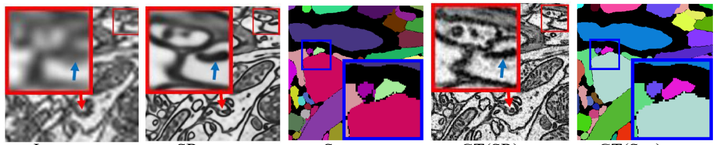
https://github.com/jtshou/EMSRSeg

NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

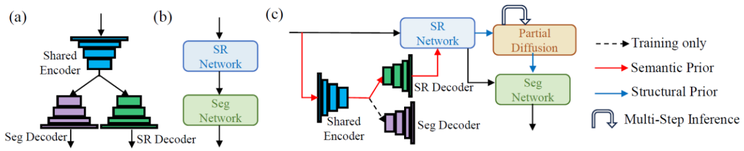
Background on EM Image Super-Resolution and Segmentation

- EM image super-resolution reconstructs high-resolution images from fast, low-resolution scans to accelerate EM imaging.
- Neuron segmentation delineates individual neurons in EM volumes for downstream connectivity analysis.



Motivation: Coupling SR and Segmentation

- Existing EMSR methods focus on pixel-level reconstruction while ignoring their interaction with downstream segmentation.
- Shared-encoder frameworks suffer from structural mismatches between SR and EM segmentation networks and difficulty inaccurately upsampling affinity-based predictions.
- In cascaded frameworks, SR lacks semantic guidance from segmentation, while segmentation does not account for discrepancies between super-resolved and real HR features.



(a) Architecture with a shared encoder and task-specific decoders. (b) Cascaded SR–segmentation framework. (c) Our framework.

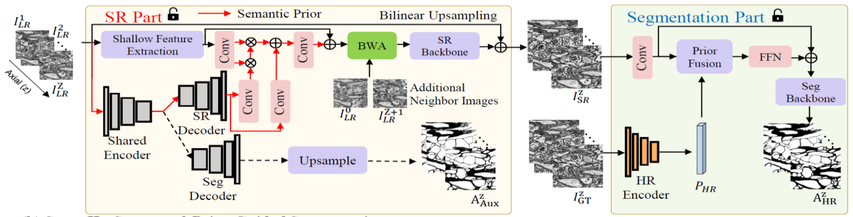
Main Contribution

- We propose the first unified framework for EM image super-resolution and segmentation.
- For SR, we introduce an auxiliary segmentation branch to provide a semantic prior, and design a BWA module to exploit cross-section structural continuity.
- For segmentation, we introduce a lightweight partial diffusion model that denoises from a partially noised SR embedding, producing a structural prior that closely approximates real HR features.
- Extensive experiments across multiple datasets and degradation settings demonstrate the effectiveness of our framework.

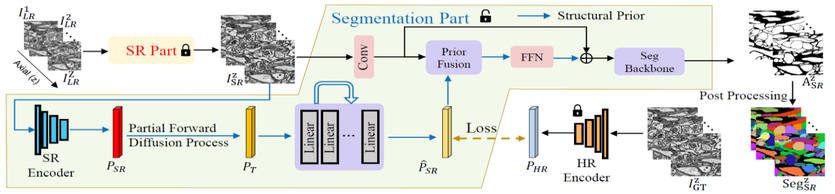
Method

Overview of framework

(a) Stage I: Semantic-Prior-Guided Super-Resolution and HR-Prior-Guided Segmentation



(b) Stage II : Structural-Prior-Guided Segmentation



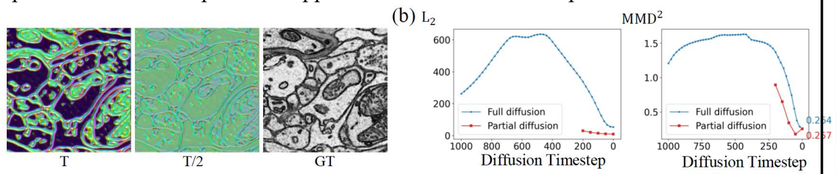
Stage I introduces a semantic prior for SR and an HR image prior for segmentation through joint training. Stage II replaces the HR prior with a structural prior from a partial diffusion model while fixing the SR part.

Semantic Prior

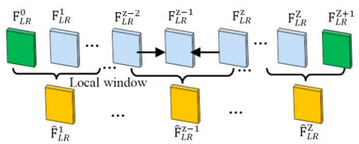
- We introduce an auxiliary segmentation branch to learn asemantic prior and inject it into the SR network through attention, preserving the original SR backbone.

Structrual Prior

- A lightweight partial diffusion model denoises a partially noised SR embedding to produce a structural prior that approximates the learned HR prior.



(a) Partial diffusion: structural cues at initializationand progressive boundary refinement through denoising. (b) Closer alignment with the target noisy distributionthan full diffusion throughout denoising.



BWA Module

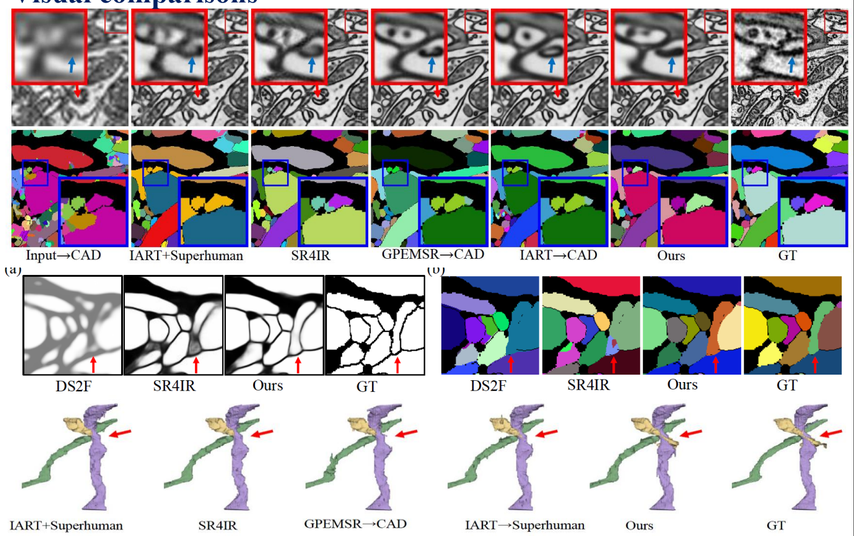
- BWA aligns features to the center section of each overlapping window in a single step and locally fuses the aligned features.

Evaluation

Quantitative comparisons

Methods		CREMI A SR (8×)				CREMI B SR (8×) + noise ($\sigma \in [10, 25]$)				CREMI C SR (16×)			
Super-resolution	Segmentation	PSNR↑	SSIM↑	VOI↓	ARAND↓	PSNR↑	SSIM↑	VOI↓	ARAND↓	PSNR↑	SSIM↑	VOI↓	ARAND↓
BasicVSR++	Superhuman	30.77	0.8237	0.7741	0.0915	24.93	0.6671	3.0169	0.4729	22.82	0.5591	3.2039	0.3287
	CAD			0.7670	0.1165			2.7715	0.2746			4.0470	0.4630
GPEMSR	Superhuman	30.89	0.8225	0.7582	0.0909	24.66	0.6512	2.6800	0.1920	22.75	0.5505	3.0550	0.2824
	CAD			0.7560	0.1123			2.9752	0.2940			3.9181	0.4378
IART	Superhuman	31.03	0.8281	0.7595	0.0903	24.88	0.6572	4.0798	0.6654	22.62	0.5491	3.8690	0.4108
	CAD			0.7516	0.1114			3.3360	0.3761			5.0567	0.6191
IART + Superhuman		30.20	0.8063	0.7152	0.1073	24.19	0.5945	1.8364	0.1660	22.20	0.4918	3.0379	0.2850
DS2F		25.10	0.5821	1.8280	0.4646	23.46	0.4999	3.0766	0.2954	21.21	0.5079	5.4758	0.7306
SR4IR		30.52	0.8154	0.7270	0.1126	24.32	0.6164	1.7061	0.1610	21.89	0.4777	2.7879	0.2509
Ours		**31.13**	**0.8318**	**0.6717**	**0.0893**	**25.13**	**0.6724**	**1.6332**	**0.1038**	**23.01**	**0.5645**	**2.4881**	**0.2083**

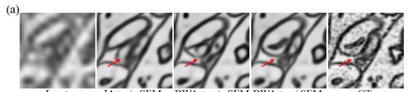
Visual comparisons



Effectiveness of proposed module

Methods	Steps	Wafer 4 VOI↓	Wafer 4 ARAND↓	Wafer 36-2 VOI↓	Wafer 36-2 ARAND↓
w/o diff	-	1.5770	0.4431	0.6546	0.1038
MLP	-	1.5490	0.3785	0.6177	0.1056
Full	50	1.3623	0.1751	0.6013	0.0951
Truncated	4	1.4684	0.1496	0.6175	**0.0933**
Truncated	10	1.3794	0.1452	0.6369	0.1316
Partial	1	1.4644	0.2376	0.6347	0.1140
Partial	2	1.4409	0.1610	0.6168	0.1015
Partial	4	1.3726	**0.1230**	**0.5971**	0.0934
Partial	8	**1.3496**	0.1657	0.6198	0.1097
Real	-	1.2762	0.1275	0.5766	0.0896

Backbone	Alignment	SEM	Wafer 4 PSNR↑	Wafer 4 SSIM↑	Wafer 36-2 PSNR↑	Wafer 36-2 SSIM↑
CNN	IA	✗	17.50	0.4162	17.57	0.4831
CNN	BWA	✗	17.81	0.4421	18.04	0.5234
CNN	BWA	✓	**17.99**	**0.4466**	**18.28**	**0.5327**
Trans	BWA	✗	17.91	0.4416	18.14	0.5208
Trans	BWA	✓	**18.06**	**0.4557**	**18.38**	**0.5420**



- BWA: Better SR with single-step alignment.
- Semantic prior: Better SR across both backbones.
- Structural prior: Better segmentation than no-prior and MLP baselines.
- Partial-4: Comparable quality to Full-50 with 12.5× fewer denoising steps.