HISTOLOGICAL BRAIN IMAGING SUPER-RESOLUTION WITH FREQUENCY-GUIDED DIFFUSION MODELS

Giovanni Casari, Federico Bolelli ✉, Costantino Grana

University of Modena and Reggio Emilia, Italy

{name.surname}@unimore.it

ABSTRACT

High-resolution histological imaging provides essential detail for quantitative brain modeling, yet acquiring whole-brain data at micrometer scale remains technically and economically challenging. This work introduces *Brain-SR*, a diffusion-based super-resolution framework designed to reconstruct high-resolution cortical sections from low-resolution BigBrain data. Building upon the InvSR paradigm, our method performs resolution enhancement in the latent space of a pretrained variational autoencoder, guided by a task-specific noise-predictor network. A key contribution is a frequency-domain supervision term that compares the magnitude spectra of predicted and target patches, enforcing spectral consistency while remaining robust to local misalignments. Quantitative evaluations demonstrate that *Brain-SR* achieves substantial improvements in LPIPS (-27%) and FID (-58%) compared to baseline diffusion Super-Resolution, while spectral analysis confirms accurate recovery of the frequency distribution. The resulting reconstructions preserve neuronal structures consistent with high-resolution references, offering a practical step toward large-scale, morphologically faithful brain histology reconstruction. The code is publicly available to support reproducibility: https://github.com/AImageLab-zip/Brain-SR.

***Index Terms*—** Super-Resolution, Brain Histology, Medical Imaging, BigBrain Project, Frequency-Domain Loss

1. INTRODUCTION

The quantitative modeling of the human brain increasingly relies on high-resolution histological data to capture the cellular organization underlying neural function and dysfunction. Recent large-scale initiatives, such as the BigBrain project [1] and the full-scale scaffold model of the human hippocampus [2], have demonstrated that micrometer-scale reconstructions of neuronal morphology are essential to enable realistic simulations of cortical and hippocampal circuits. These models require accurate maps of neuronal somata and local cytoarchitecture, but the acquisition of full-brain histological data at micrometer resolution remains technically and economically prohibitive. Even for well-studied regions such as the hippocampal CA1, only a limited subset of sections has been scanned at 1 µm per pixel [3], leaving most of the brain available only at coarser 20 µm resolution.

Deep learning-based Super-Resolution (SR) methods have emerged as a powerful alternative to bridge this gap. Early convolutional models such as SRCNN [4] and EDSR [5] established the feasibility of end-to-end SR, while GAN-based approaches like ESRGAN [6, 7] enhanced perceptual realism through adversarial and feature-space supervision. More recently, transformer-based architectures such as SwinIR [8] have further improved structural consistency by capturing long-range dependencies. In parallel, diffusion models have become a compelling alternative to GANs: SR3 [9] demonstrated their potential for generating photorealistic details, and InvSR [10] extended this concept with a learnable noise-predictor to adapt pretrained diffusion models.

Recent advances have extended SR to digital pathology, where it is used to reduce scanning time and storage while preserving diagnostically relevant microstructures. CNN- and GAN-based approaches have shown that low-magnification WSI tiles can be upsampled to recover cellular detail [11, 12] and even support blinded diagnostic assessment [13]. More recently, diffusion-based models [14, 15] have incorporated pathology-aware priors, improving fidelity across multiple tissue types, including brain samples. However, these methods lack explicit frequency-level supervision and remain sensitive to stain variability and domain shifts.

In this paper, we introduce *Brain-SR*, a diffusion-based super-resolution framework specifically designed for histological brain images. Our approach builds upon the InvSR paradigm [10], where a pretrained diffusion model is guided by a learned noise-predictor network to perform resolution enhancement in the latent space. We target the reconstruction of HR-like BigBrain sections from Low-Resolution (LR) scans, enabling downstream analyses such as neuronal density estimation and morphological pattern mapping required by scaffold models [2]. Such reconstructions support statistical analyses of neuron distribution, cortical layering, and tissue organization, providing morphologically consistent detail while avoiding pixel-level interpretative bias.

✉ Corresponding author: federico.bolelli@unimore.it.

Paper Contribution. The main contributions of this paper can be summarized as follows:

- We introduce *Brain-SR*, a diffusion-based super-resolution framework tailored for histological brain imaging, extending diffusion inversion to this domain.
- We propose a patch-wise Fourier supervision (*FFT-loss*) that enforces spectral consistency and stabilizes training, improving the preservation of fine morphological details over conventional pixel-wise losses.
- We provide an extensive quantitative, perceptual, and frequency-domain evaluation, including ablation studies, demonstrating accurate spectral reconstruction and superior perceptual quality compared to baseline SR models.

Together, these contributions enable *Brain-SR* to generate high-resolution, morphologically coherent reconstructions that support reliable quantitative analysis of neuronal density, cortical layering, and broader cytoarchitectural organization.

2. MATERIALS AND METHODS

Datasets. We employed data from the BigBrain project [1], a publicly available three-dimensional histological reconstruction of a complete human brain at 20 µm resolution. Each section, acquired through serial microtomy and digitized at gigapixel scale, is associated with an affine transformation mapping its coordinates into a common reference frame. In 2022, a subset of 145 sections was re-scanned at 1 µm per pixel using whole-slide imaging [3], providing the High-Resolution (HR) reference used in this work.

To construct paired training data, each LR section was aligned to its HR counterpart using the provided affine transformations. Direct reconstruction of the original 1 µm signal from 20 µm LR images is both computationally prohibitive and ill-posed due to the substantial frequency gap between the LR and HR domains. Therefore, HR sections were Gaussian-smoothed (σ=2) and downsampled to an effective 5 µm resolution to obtain a spectrally compatible target. Aligned LR and downsampled HR images were then partitioned into overlapping patches: 128×128 pixels for LR and the corresponding 512×512 regions for HR, consistent with the 4× scale factor. Patches containing more than 90% white background were discarded to exclude low-information regions.

This procedure yielded nearly 500,000 paired samples. A 5-fold cross-validation scheme was adopted, with each fold using a 70%/30% train–test split. No fixed test set was used; all sections appeared in the test partition across folds, ensuring a comprehensive and unbiased evaluation.

Model Overview. The proposed framework builds upon the *InvSR* architecture [10], a diffusion-based approach designed for efficient super-resolution through latent-space inversion. Instead of retraining an entire diffusion backbone, InvSR introduces a dedicated *Noise Predictor* (NP) network that estimates the residual high-frequency information missing from the low-resolution input. This component allows the pretrained diffusion model to be reused without modification, substantially reducing computational cost.

Given an input LR image x_{LR}, it is first bicubically upscaled to the HR size and encoded into the latent space of a VAE, producing a compact representation $z_{LR} \in \mathbb{R}^{4\times64\times64}$ that corresponds to an $8\times$ spatial reduction. The NP network predicts a latent noise map ϵ, which is scaled by a time-dependent coefficient σ_t derived from the diffusion scheduler. The noisy latent representation is then obtained as:

$$z_t = z_{LR} + \sigma_t\, \mathcal{S}(\epsilon), \quad \epsilon = \mathrm{NP}(x_{LR}), \tag{1}$$

where $\mathcal{S}(\cdot)$ denotes the sampling of the predicted noise.

This intermediate state z_t serves as the initialization for the reverse diffusion process [16], performed in a single-step regime following SD-Turbo [17]. Instead of running a full denoising trajectory, the pretrained diffusion model applies one denoising update to refine z_t into a clean latent estimate $\hat{z}_0$, which is then decoded by the VAE decoder to obtain the final high-resolution reconstruction $\hat{x}_{HR}$. This modular design enables controllable detail enhancement while maintaining global anatomical consistency. All supervision losses are computed directly in the latent space, significantly accelerating training while preserving the fidelity of the reconstructed high-resolution structures.

Loss Functions. The total training objective combines four complementary terms,

$$\mathcal{L}_{\text{tot}} = \lambda_{\text{L2}}\mathcal{L}_{L2} + \lambda_{\text{LPIPS}}\mathcal{L}_{\text{LPIPS}} + \lambda_{\text{FFT}}\mathcal{L}_{\text{FFT}} + \lambda_{\text{ADV}}\mathcal{L}_{\text{ADV}}, \tag{2}$$

where $\mathcal{L}_{L2}$ enforces pixel fidelity, $\mathcal{L}_{\text{LPIPS}}$ ensures perceptual similarity [18], and $\mathcal{L}_{\text{ADV}}$ promotes realistic textures via adversarial training. The most relevant component is the frequency-domain loss $\mathcal{L}_{\text{FFT}}$, inspired by [19], which compares the magnitude of Fourier spectra between predicted and target images. Each latent image is divided into small overlapping patches (typically $p \in \{4, 8, 16\}$ latent pixels), and for each patch x_p the 2D transform $\mathcal{F}(\cdot)$ is applied:

$$\mathcal{L}_{\text{FFT}} = \frac{1}{N}\sum_{i=1}^{N} \Big\| \, |\mathcal{F}(x_p^{(i)})| - |\mathcal{F}(\hat{x}_p^{(i)})| \, \Big\|_1 . \tag{3}$$

By operating on spectral magnitudes rather than spatial intensities, this loss encourages the model to reproduce the correct frequency distribution of histological textures, stabilizing training and improving the recovery of fine structural details. While the phase component of the Fourier spectrum carries most of the spatial structure, it is highly sensitive to even minimal misalignments [20] and thus unsuitable as a stable supervision signal, particularly in histological brain images characterized by numerous small, high-contrast cellular structures distributed over largely homogeneous backgrounds (Fig. 1). Restricting the comparison to magnitude spectra—computed locally over overlapping patches—provides a more robust and

Table 1: Training configurations and corresponding quantitative results for the different *Brain-SR* model variants. Each setup combines pixel-wise, perceptual, adversarial, and frequency-domain losses with different weights λ, showing the progressive transition from spatial to frequency-guided supervision. The column FFT_{conf} indicates the patch sizes p (in latent space) used for computing the FFT loss. Best results are highlighted in bold. * refers to metrics computed with respect to the latent-reconstructed HR images instead of the original HR reference.

Model	λ_{L2}	λ_{ADV}	λ_{LPIPS}	λ_{FFT}	FFT_{conf}	L2 ↓	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
Brain-SR-b1	1.0	0.05	1.5	–	–	**0.015**	**19.98**	0.303	0.550	139.4
Brain-SR-b2	1.0	0.10	2.0	–	–	0.017	19.40	**0.322**	0.412	98.7
Brain-SR-fft	0.3	0.08	0.7	0.2	{16,8}	0.019	18.90	0.319	0.401	83.2
Brain-SR-p16	–	0.08	0.7	0.4	{16,8}	0.024	17.86	0.280	0.412	63.7
Brain-SR-p8	–	0.08	0.7	0.4	{8,4}	0.023	18.06	0.282	0.412	62.0
Brain-SR	–	0.08	0.7	0.4	{4,2}	0.021	18.32	0.296	**0.398**	**57.5**
*Brain-SR**	–	*0.08*	*0.7*	*0.4*	*{4,2}*	*0.020*	*18.67*	*0.317*	*0.386*	*33.4*

spatially aware constraint: it enforces the correct frequency energy distribution within each region, while the complementary pixel-wise, perceptual, and adversarial losses implicitly promote phase alignment and structural coherence.

As supported by the experimental results presented in Sec. 3, this combination achieves both spectral consistency and visual fidelity in the reconstructed images.

Training Setup. *Brain-SR* models were trained for 10^5 iterations on three NVIDIA RTX A5000 GPUs, taking about three days per run. Optimization used Adam with a cosine learning-rate schedule from 2×10^{-5}, batch size 24, and EMA rate 0.99. Early variants (*Brain-SR-b1,b2*) employed slightly different hyperparameters during an initial tuning phase to ensure stable convergence, while all later models used the finalized setup for fair comparison. Loss weights and FFT patch configurations are listed in Tab. 1. The diffusion backbone and the latent VAE were initialized from pretrained SD-Turbo [17] weights and kept frozen throughout training.

3. EXPERIMENTAL RESULTS

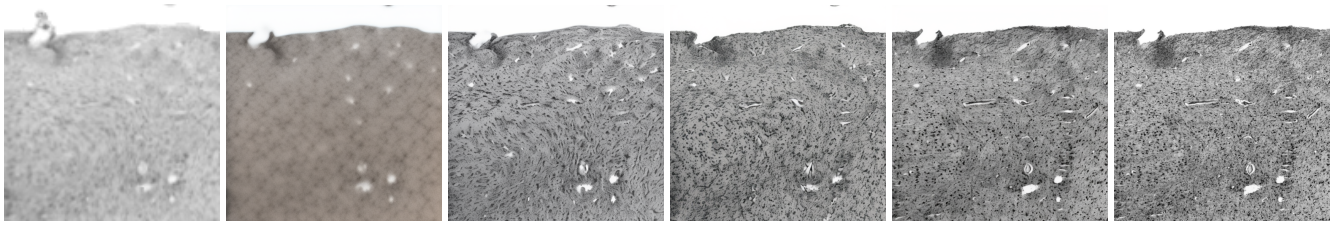
To evaluate the contribution of each component, multiple models were trained with different combinations and weights of these losses, as summarized in Tab. 1.

Quantitative Results. To assess the effect of the proposed frequency-guided training, we compared a series of models trained with different loss configurations, progressively replacing the pixel-wise $\mathcal{L}_2$ term with the FFT loss. Tab. 1 summarizes the results on the held-out test set. Models trained predominantly with $\mathcal{L}_2$ (*Brain-SR-b1* and *Brain-SR-b2*) achieved the best numerical values for L2 and PSNR, while those optimized with the FFT loss (*Brain-SR-fft* and below) obtained superior perceptual scores (LPIPS and FID). This confirms that pixel-based metrics are poorly correlated with visual quality in histological data, where small spatial misalignments strongly penalize $\mathcal{L}_2$ despite perceptually equivalent results. The final *Brain-SR* model, trained without $\mathcal{L}_2$ terms, reached the lowest LPIPS and FID, indicating a closer alignment to the HR distribution and improved perceptual realism. The proposed model *Brain-SR* reduces LPIPS by 27% and FID by 58% compared to the baseline model *Brain-SR-b1* that obtains the best L2 and PSNR values.

As shown in Fig. 1, the proposed method successfully reconstructs cell boundaries and fiber-like textures that are completely absent in the LR input, resulting in a substantial visual improvement and anatomically coherent structures. Nevertheless, when compared to the HR reference, minor discrepancies and residual patterns can still be observed, indicating that further refinement is required to achieve fully consistent reconstructions. This result represents an important step toward the generation of high-resolution histological data suitable for detailed structural brain analysis.

Frequency-Domain Analysis. To assess the impact of the FFT supervision, we also computed the Radial Power Spectrum (RPS) of HR, LR, and SR images. The RPS quantifies the energy associated with each spatial frequency by averaging the power of all Fourier components having the same frequency magnitude, regardless of their orientation [21, 22]. This provides an intuitive view of how different frequency components contribute to the overall final images, enabling a direct and quantitative comparison across image sets.

As shown in Fig. 2, the LR spectrum exhibits a marked loss of high-frequency energy compared to HR data, reflecting the absence of fine details. When the HR images are reconstructed through the VAE, the resulting *HR-Latent* spectrum shows a slight attenuation of the highest frequencies, a consequence of latent compression. Since all training losses are computed in the latent space, this *HR-Latent* curve effectively becomes the spectral reference. Across progressively refined configurations, the reconstructed spectra of the *Brain-SR* variants gradually approach this reference distribution, with the final model almost perfectly matching the *HR-Latent* curve, confirming that it faithfully reproduces the frequency distribution in the latent space. Residual differences in phase and frequency orientation account for the minor perceptual variations visible in Fig. 1 between *Brain-SR* and *HR-Latent*.



LR | *Brain-SR-b1* | *Brain-SR-b2* | *Brain-SR* | HR-Latent | HR

Fig. 1: Qualitative comparison of patches from the test set. Progressive adjustment of loss weights and hyperparameters leads to increasingly detailed reconstructions, with the final model best reproducing cellular structures consistent with the HR image.

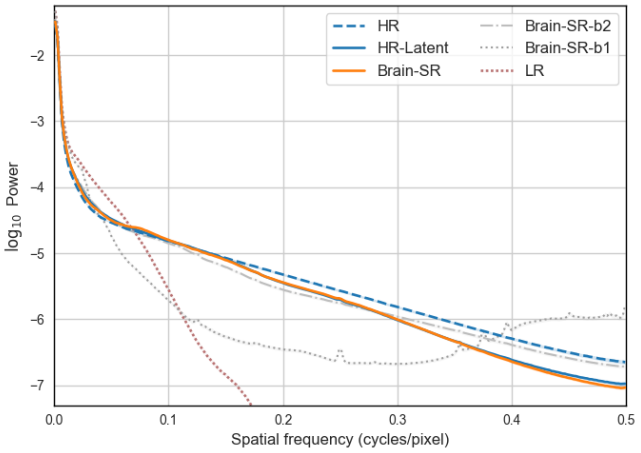


Fig. 2: Radial Power Spectrum (RPS) comparison between LR, HR, latent-reconstructed HR, and SR images. Each curve represents the distribution of image energy across spatial frequencies, where higher frequencies correspond to finer structural details. The LR spectrum exhibits a strong loss of high-frequency energy, whereas progressively refined models align more closely with the reference distribution. The final *Brain-SR* closely matches the *HR-Latent* spectrum, confirming accurate recovery of the frequency distribution encoded in the latent representation. *Best viewed in color.*

These findings show that frequency-domain supervision effectively constrains spectral content, while complementary losses ensure spatial and perceptual coherence.

Ablation and Latent-Space Trade-Offs. Two complementary analyses examined the effects of patch size and latent-space compression on model performance.

First, varying the patch size used in $\mathcal{L}_{\text{FFT}}$ demonstrated, as shown in Tab. 1, a consistent improvement in perceptual metrics when reducing p from 16 to 4 latent pixels (corresponding to approximately 128–32 spatial pixels), confirming that smaller patches enhance sensitivity to local details. This improvement arises because smaller patches impose stronger spatial constraints, by enforcing consistency over more localized regions, they provide richer spatial information and guide the model towards sharper and coherent reconstructions.

Second, we evaluated the impact of latent-space compression by comparing HR ground truths with their reconstructions obtained solely through the VAE encoder–decoder. As reported in the last row of Tab. 1, where metrics for *Brain-SR* are computed with respect to the latent-reconstructed images, nearly all values improve, with the largest gains observed in perceptual measures. This behavior highlights the slight difference between the spectral and perceptual distributions of the latent and original HR domains, as the model is trained to reproduce the characteristics of the former. Consistent evidence emerges from the RPS curves in Fig. 2, where the HR-latent spectrum shows a slight deviation from the original HR curve, particularly in the higher frequencies. This spectral shift is reflected visually in Fig. 1, where the HR-Latent images exhibit a slightly sparser neuronal density and a loss of very fine structural details compared to HR. Despite these differences, the latent-space formulation remains compact and spectrally consistent, providing substantial computational savings while maintaining perceptual fidelity.

4. CONCLUSION

This work presented *Brain-SR*, a diffusion-based super-resolution framework tailored for histological brain imaging. By integrating frequency-domain supervision into a latent diffusion architecture, the proposed approach achieves perceptually and spectrally consistent reconstructions of cortical tissue from low-resolution BigBrain data. The introduction of the FFT-based loss proved effective in stabilizing training and restoring the frequency energy distribution characteristic of high-resolution histological textures. Experimental results demonstrated substantial improvements in perceptual quality (LPIPS and FID) and strong alignment with the spectral profile of the latent-reconstructed HR images, confirming that *Brain-SR* faithfully captures the frequency content most relevant for structural interpretation.

These results represent a first step toward overcoming the current limitations in full-brain histological acquisitions, which remain infeasible at micron resolution. Future work will focus on extending the approach to multi-slice and 3D consistency, incorporating higher-resolution reference data, and exploring adaptive frequency weighting to enhance ultra-fine detail recovery. Ultimately, this framework aims to enable morphologically accurate, high-resolution brain reconstructions for large-scale structural analyses.

5. ACKNOWLEDGMENT

This project has received funding from the University of Modena and Reggio Emilia and Fondazione di Modena, through the FAR 2024 (CUP E93C24002080007) and FARD-2024 funds (Fondo di Ateneo per la Ricerca).

6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by [1,3]. Ethical approval was not required, as confirmed by the license attached to the open-access data.

7. REFERENCES

[1] K. Amunts et al., "BigBrain: An Ultrahigh-Resolution 3D Human Brain Model," *Science*, vol. 340, no. 6139, pp. 1472 – 1475, 2013.

[2] Daniela Gandolfi et al., "Full-scale scaffold model of the human hippocampus CA1 area," *Nature Computational Science*, vol. 3, no. 3, pp. 264–276, Mar 2023.

[3] Christian Schiffer et al., "Selected 1 micron scans of BigBrain histological sections (v1.0)," EBRAINS, 2022.

[4] Chao Dong et al., "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 38, no. 2, pp. 295–307, 2015.

[5] Bee Lim et al., "Enhanced Deep Residual Networks for Single Image Super-Resolution," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 136–144.

[6] Xintao Wang et al., "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," in *European Conference on Computer Vision Workshops (ECCVW)*, 2018.

[7] Xintao Wang et al., "Real-esrgan: Training real-world blind super-resolution with pure synthetic data," in *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*.

[8] Jingyun Liang et al., "SwinIR: Image Restoration Using Swin Transformer," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 1833–1844.

[9] Chitwan Saharia et al., "Image Super-Resolution via Iterative Refinement," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 45, no. 4, pp. 4713–4726, 2022.

[10] Zongsheng Yue et al., "Arbitrary-steps Image Super-resolution via Diffusion Inversion," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

[11] Mehdi Afshari et al., "Single patch super-resolution of histopathology whole slide images: a comparative study," *Journal of Medical Imaging*, vol. 10, no. 1, pp. 017501–017501, 2023.

[12] The Cancer Genome Atlas Research Network, "The cancer genome atlas pan-cancer analysis project," *Nature Genetics*, vol. 45, no. 10, pp. 1113–1120, 2013.

[13] Bin Li et al., "Single image super-resolution for whole slide image using convolutional neural networks and self-supervised color normalization," *Med Image Anal*, vol. 68, pp. 101938, 2020.

[14] Abdullah et al., "High-resolution histopathology whole slide image generation using wavelet diffusion model," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2025.

[15] Xuan Xu et al., "Histo-diffusion: A diffusion super-resolution method for digital pathology with comprehensive quality assessment," *arXiv preprint arXiv:2408.15218*, 2024.

[16] Jonathan Ho et al., "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 6840–6851, 2020.

[17] Axel Sauer et al., "Adversarial Diffusion Distillation," in *European Conference on Computer Vision (ECCV)*, 2024, pp. 87–103.

[18] Richard Zhang et al., "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 586–595.

[19] Dario Fuoli et al., "Fourier Space Losses for Efficient Perceptual Image Super-Resolution," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[20] Alan V. Oppenheim and Cram, "Discrete-time signal processing : Alan v. oppenheim, 3rd edition," 2011.

[21] Robert A Ulichney, "Dithering with blue noise," *Proceedings of the IEEE*, vol. 76, no. 1, pp. 56–79, 1988.

[22] A Spector et al., "Statistical Models for Interpreting Aeromagnetic Data," *Geophysics*, vol. 35, no. 2, pp. 293–302, 1970.