From Low-Quality to High-Quality: Generating Structure-Preserved Atomic-Scale HRTEM images via an Enhanced CycleGAN

Abstract High-quality high-resolution transmission electron microscopy (HRTEM) images are essential for linking materials structure–property relationships at the atomic scale. However, capturing dynamic processes with high temporal resolution inevitably leads to severely degraded HRTEM images under low-dose imaging conditions, which substantially limits accurate structural analysis. In this paper, we develop a structure-preserving HRTEM restoration framework that enhances low-quality HRTEM images with blurred or incomplete atomic arrangements using a generative deep learning approach while maintaining physical fidelity. Specifically, we propose HRTEM-GAN, a cycle-consistent generative framework that operates under unpaired training conditions and performs patch-level distribution modeling between high- and low-quality image domains, while explicitly incorporating frequency-domain constraints to preserve atomic-scale structural fidelity. This design enables effective restoration of low-dose HRTEM images, yielding structurally coherent atomic lattices that are fully suitable for subsequent recognition and quantitative analysis. The proposed method is validated on a real experimental dataset acquired during in situ imaging of Au catalysts under CO oxidation conditions. Compared with representative methods, HRTEM GAN achieves substantial improvements in image restoration quality and consistently enhances downstream atomic column recognition performance. These results demonstrate the potential of the proposed framework to facilitate reliable atomic-scale analysis in HRTEM studies.

Keywords HRTEM, Atomic-resolution image restoration, Unpaired image-to-image translation, Generative adversarial networks.

1 Introduction

Understanding and controlling material properties through atomic-scale structure has long been a central pursuit in materials science, as atomic arrangements

fundamentally govern material functionality and performance [1–4]**错误!未找到引用源。**. High-resolution transmission electron microscopy (HRTEM), therefore, plays a pivotal role in reliable materials characterization [5–7]**错误!未找到引用源。**; however, realistic experimental constraints often result in severely degraded images, which in turn hinder accurate identification of atomic sites and subsequent structural analysis [4,8]. For instance, in catalytic systems, chemical reactions are often accompanied by rapid atomic-scale structural rearrangements occurring on timescales of tens of milliseconds [9–11]**错误!未找到引用源。**. Capturing image sequences at such high temporal resolution inevitably yields images that are strongly affected by shot noise, further complicating atomic-scale feature extraction and statistical analysis. It is generally impractical to suppress noise by increasing the incident electron beam intensity, as high-dose exposure to high-energy electrons can induce damage to the material. This motivates the need to recover high-quality (LQ) atomic-resolution information from degraded low-quality (LQ) HRTEM images while preserving physical fidelity [12,13].

In recent years, a variety of computational methods, including conventional image processing algorithms [14–16]**错误!未找到引用源。错误!未找到引用源。** and deep learning techniques [17–19]**错误!未找到引用源。**, have been applied to mitigate the inherent limitations of microscopy imaging and improve image quality. Conventional rule-based methods, such as Wiener filtering, bilateral filtering, and BM3D, have been widely used for noise suppression and basic image enhancement [16,20]. While effective in attenuating random noise, these methods rely on fixed assumptions and limited local statistics, making it difficult to robustly handle complex backgrounds and severe image degradation. Recently, deep learning (DL)-based methods, leveraging their strong non-linear modeling capacity, have been extensively explored for electron microscopy image processing, primarily targeting low-level image enhancement tasks such as denoising, super-resolution, and generation [17].

Representative DL-based denoising approaches for TEM typically rely on supervised learning with synthetic or simulated training data [18,19,21]. For example, Lin et al. [18] introduced the AtomSegNet framework, in which convolutional encoder–decoder networks are trained on physics-informed simulated images. Mohan et al. [12] further proposed a simulation-based denoising (SBD) framework, demonstrating that CNNs trained on carefully designed forward models can effectively denoise low-SNR TEM images and generalize to real experimental data, particularly when large receptive fields are used to capture non-local atomic periodicities. More recently, frequency-aware enhancement strategies have been explored, incorporating spatial–frequency interactions to better exploit the periodic nature of atomic arrangements. Li et al. [22] proposed a framework that integrates a spatial-frequency interaction network with noise calibration-based data synthesis to model frequency-domain information. Nevertheless, most existing methods still struggle under severely degraded imaging conditions. Importantly, most methods rely on synthetic images to construct paired samples for supervision, while overlooking the rich and realistic texture characteristics present in real high-quality experimental images.

To alleviate the dependence on paired supervision, unpaired image-to-image translation frameworks have been actively studied [23–26]**错误!未找到引用源。错误!未找到引用源。**. In this context, approaches based on the Cycle-Consistent Generative Adversarial Network (CycleGAN) [26–28] have attracted increasing attention as a pioneering framework capable of translating images between two domains without requiring paired training samples. Building upon this paradigm, Quan et al. [29] proposed an asymmetrically cyclic adversarial network for unpaired denoising in electron microscopy, which extends the CycleGAN framework to learn forward and inverse noise mappings for artifact suppression without paired clean targets. However, the method still relies primarily on cycle-consistency for structural preservation and lacks explicit atomic-scale geometric or physical constraints; their whole-image-level modeling with weak implicit constraints limits controllability at the atomic scale,

frequently resulting in inaccurate atomic reconstruction and inconsistent atomic positions.

In this paper, we propose HRTEM-GAN, a CycleGAN-based framework designed for structure-preserved image restoration in atomic-resolution HRTEM imaging. Instead of performing whole-image translation, HRTEM-GAN adopts patch-level modeling, which reduces the effective receptive scope of generation and allows the network to focus on fine-grained atomic structures, alleviating incomplete atomic recovery commonly observed in full-image generation. To further improve the modeling of long-range atomic correlations, a Vision Transformer (ViT) [30] module is incorporated at the bottleneck of the generator, enabling global dependency modeling beyond local convolutional neighborhoods. In addition, we explicitly introduce frequency-domain modeling and constraints to exploit the intrinsic spectral characteristics of HRTEM images. By enforcing spatial-frequency consistency via a dedicated frequency branch and frequency-aware losses, the proposed framework balances image quality enhancement and atomic-scale structural fidelity. We have performed extensive quantitative and qualitative evaluations on a real experimental dataset, demonstrating that HRTEM-GAN outperforms existing methods of TEM image quality enhancement and atomic structure preservation. Evaluations using AtomSegNet further show improved atomic recognition accuracy and stability, underscoring the value of the proposed framework for reliable quantitative atomic-scale analysis.

2 Experimental

2.1 Pipeline overview and dataset preparation

Our framework comprises two stages arranged in a sequential pipeline (see Fig. 1). Firstly, the acquired HRTEM images are divided into high- and low-quality groups (unpaired) and cropped into patches that are then classified into foreground (nanocrystal area) and background (support area). Further details of this preprocessing are provided in Supplementary Note I (online). Secondly, the proposed HRTEM-GAN model is

trained exclusively on unpaired high- and low-quality foreground patches, enabling the network to focus on learning atomic patterns. We employ CycleGAN to establish bidirectional mappings between the high- and low-quality domains, using two generators and two discriminators constrained by cycle consistency. We next detail our HRTEM-GAN (Section 2.2), frequency-enhanced feature interaction (Section 2.3), and implementation details (Section 2.4).

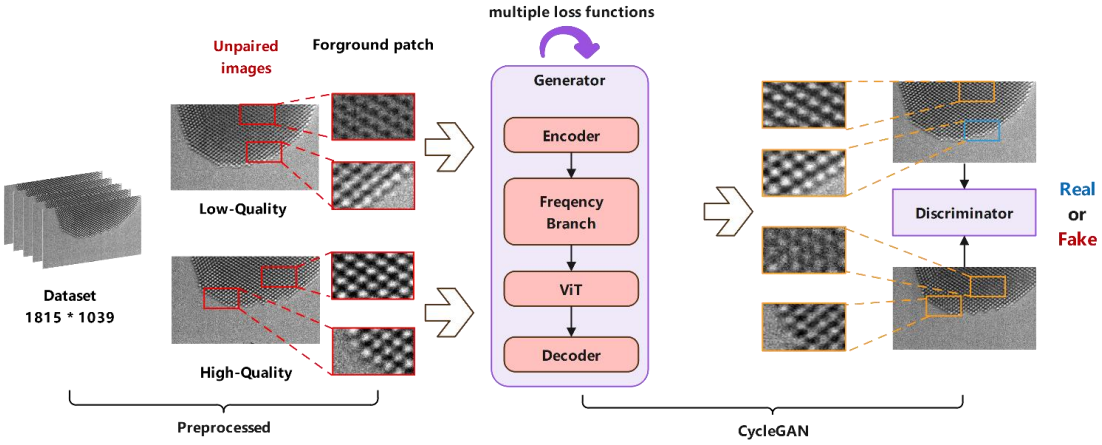

Figure 1. Overall pipeline.

Our dataset was obtained from in situ aberration-corrected transmission electron microscopy experiments, capturing surface step-site dynamics of Au catalysts during CO oxidation at room temperature. Due to low electron dose and high temporal resolution, the acquired HRTEM images exhibit significant quality variations. The images were therefore screened based on spatial resolution, atomic-column clarity, and lattice integrity, and subsequently divided into high-quality and low-quality subsets that are intrinsically unpaired and non-aligned, forming an unpaired training dataset.

2.2 HRTEM-GAN

To address the challenge of obtaining perfectly aligned pairs of noisy experimental images and clean ground-truth images in HRTEM experiments — which renders standard supervised learning methods (e.g., Pix2Pix [31]) inapplicable—we propose HRTEM-GAN (see Fig. 2(a)). It is based on CycleGAN and learns robust cross-domain mappings using unpaired data. In particular, the cycle-consistency constraint

encourages an image translated from the source domain to the target domain and back to remain faithful to the input, which is essential for preserving the geometric arrangement of atomic columns during enhancement. Here, we consider unpaired translation between the low-quality and high-quality HRTEM domains, denoted as LQ and HQ, respectively. We train two generators $G_{\mathrm{H}} \equiv G_{\mathrm{LQ}\to\mathrm{HQ}}$ and $G_{\mathrm{L}} \equiv G_{\mathrm{HQ}\to\mathrm{LQ}}$, together with discriminators D_{H} and D_{L}. The base objective is formulated as

$$\mathcal{L}_{\mathrm{base}} = \mathcal{L}_{\mathrm{GAN}}(G_{\mathrm{H}}, D_{\mathrm{H}}) + \mathcal{L}_{\mathrm{GAN}}(G_{\mathrm{L}}, D_{\mathrm{L}}) + \lambda_{\mathrm{cyc}}\mathcal{L}_{\mathrm{cyc}}(G_{\mathrm{H}}, G_{\mathrm{L}}) + \lambda_{\mathrm{idt}}\mathcal{L}_{\mathrm{idt}}(G_{\mathrm{H}}, G_{\mathrm{L}}) , \tag{9}$$

where $\mathcal{L}_{\mathrm{GAN}}$, $\mathcal{L}_{\mathrm{cyc}}$, and $\mathcal{L}_{\mathrm{idt}}$ are defined in Supplementary Note III.

Nevertheless, standard CycleGAN architectures often struggle to fully resolve the high-frequency periodic patterns intrinsic to atomic lattices. To address this limitation, HRTEM-GAN introduces two major modifications.

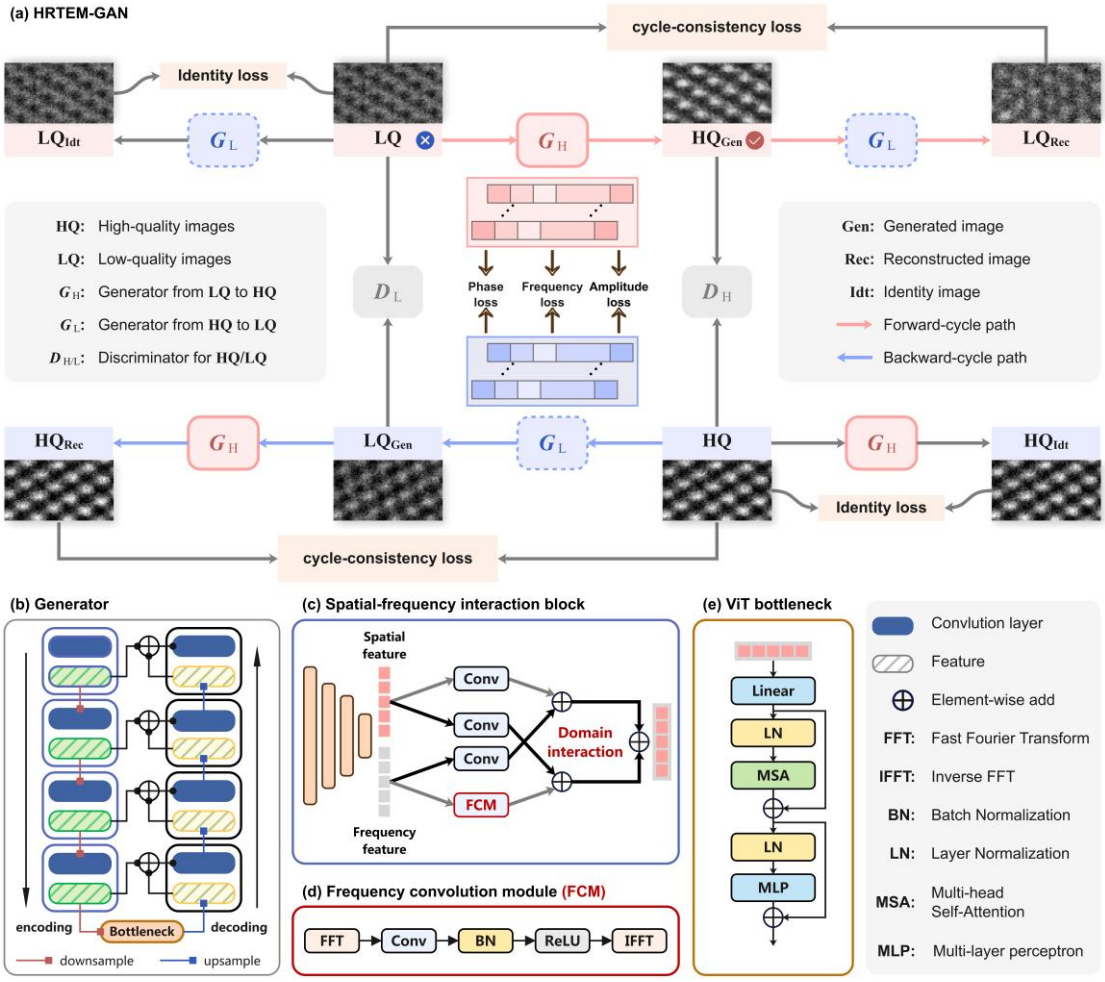

Figure 2. Schematic diagram of the proposed HRTEM-GAN. (a) Overall HRTEM-GAN framework. **(b)** Enhanced UNet-ViT generator. **(c)** Spatial-frequency interaction block. **(d)** Frequency convolution module (FCM). **(e)** ViT bottleneck.

Firstly, we impose feature-level alignment on the fused bottleneck representations of the two generators G_{H} and G_{L} to encourage the generated samples to approximate the feature characteristics of the target domain. Let z_H and z_L denote the corresponding bottleneck features. Feature consistency is enforced from two complementary aspects: frequency-domain alignment and distributional alignment.

Frequency-domain alignment. We apply the fast Fourier transform (FFT) to the bottleneck features and penalize the discrepancy between their magnitude spectrum, thereby encouraging the translated results to preserve consistent spectral statistics and reduce unrealistic high-frequency artifacts. Formally, let $\mathcal{F}(\cdot)$ denote the FFT and $|\cdot|$ its magnitude spectrum. The frequency-domain alignment loss is defined as

$$\mathcal{L}_{\mathrm{FFT}} = \mathbb{E}_{z_{\mathrm{H}},\, z_{\mathrm{L}}}[\,\|\,|\mathcal{F}(z_{\mathrm{H}}(x))| - |\mathcal{F}(z_{\mathrm{L}}(y))|\,\|_1]\,, \tag{9}$$

where $\|\cdot\|_1$ denotes the element-wise ℓ_1 norm.

Distributional alignment via the Characteristic Function (CF). In addition, we introduce the Characteristic Function (CF) [32] as a robust measure to match feature-space distributions between the two domains. The CF yields complex-valued responses whose real and imaginary parts encode cosine and sine components, naturally supporting an amplitude–phase decomposition. The amplitude characterizes how strongly and broadly the feature distribution responds to each probing direction, which correlates with the coverage of fine-grained textures and local contrast variations. The phase, by contrast, is more sensitive to structural shifts and thus penalizes misalignment that would manifest as lattice distortion or atomic-position drift. Accordingly, we enforce joint amplitude and phase consistency between the bottleneck features of G_{H} and G_{L}. This motivates our design to enforce joint amplitude and phase consistency between the two domains:

$$\mathcal{L}_{\mathrm{CF}} = \|\,|\,\mathrm{CF}(z_{\mathrm{H}})\,| - |\,\mathrm{CF}(z_{\mathrm{L}})\,|\,\|_1 + \lambda_{\phi}(1 - \cos(\angle\mathrm{CF}(z_{\mathrm{H}}) - \angle\mathrm{CF}(z_{\mathrm{L}})))\,, \tag{9}$$

where $|\cdot|$ and $\angle(\cdot)$ denote the magnitude and phase of the complex-valued CF output, respectively, and λ_{ϕ} balances the phase term. Details of the CF construction are deferred to Supplementary Note V.

In summary, the total loss of our training is:

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{base}} + \lambda_{\mathrm{CF}}\mathcal{L}_{\mathrm{CF}} + \lambda_{\mathrm{FFT}}\mathcal{L}_{\mathrm{FFT}}\,, \tag{9}$$

where λ_{CF} and λ_{FFT} are weighting hyperparameters that balance the contributions of the CF constraint and the frequency-domain consistency constraint to the overall training objective.

Secondly, our HRTEM-GAN moves beyond CycleGAN whose conventional U-Net [33] generator by integrating a dual-branch architecture in each encoder layer (see Fig. 2(c)). Specifically, considering that spectral features naturally encode global

periodicity, we introduce an independent frequency branch alongside the spatial branch to preserve the intrinsic long-range order and fine-grained details of crystal lattices (detailed in the following section). To further enhance the generator's ability to recover lattice-level structural regularity, a ViT [30] module is incorporated at the bottleneck of the dual-branch encoder–decoder architecture. While convolutional operations effectively capture local atomic textures, their limited receptive field constrains the modeling of long-range periodicity. Addressing this, the fused spatial-frequency features at the bottleneck are flattened into a token sequence, and two-dimensional positional encodings are added to preserve the original lattice geometry. The token sequence is then processed by several stacked Multi-head Self-Attention (MSA) layers, enabling the network to learn global dependencies between arbitrary spatial locations and to represent crystal lattices with coherent long-range ordering. The globally contextualized features are subsequently reshaped back into a two-dimensional feature map and propagated into the decoder through skip connections. By embedding global self-attention at the structural bottleneck, this hybrid design enforces global lattice consistency while retaining high-fidelity local atomic details provided by the convolutional and frequency branches.

2.3 Frequency-enhanced feature interaction

Complementing the independent frequency branch in each layer of the encoder, the input HRTEM image is processed by a standard convolutional layer to obtain primary features, which are then split along the channel dimension to initialize parallel spatial and frequency streams. After that, we employ the spatial-frequency interaction block [22] to progressively enhance both feature streams. As illustrated in Fig. 2(d), convolutional modules are used to capture local geometric structures and atomic neighborhoods, while frequency convolution modules (FCM) emphasize periodic lattice patterns and high-frequency structural details by leveraging Fourier transform.

To prevent semantic drift and ensure that both representations evolve in a mutually consistent manner, a cross-branch information exchange mechanism is incorporated within each interaction block. Specifically, the updated features from one branch are

passed through an additional convolutional layer and added to the counterpart branch, enabling reciprocal refinement. This design ensures that spatial features are continuously informed by global periodic cues, while frequency-domain features remain anchored to localized structural geometry.

After the interaction stages, the enhanced spatial–frequency features are fused into a unified representation and propagated through the decoder to reconstruct the output image. This integrated design significantly improves the generator's capacity to restore atomic lattice periodicity, preserve sharp high-frequency details, and maintain local structural fidelity under unpaired training conditions. The restored image patches are finally reassembled into full-resolution frames, yielding high-quality HRTEM images with globally coherent and physically meaningful atomic arrangements.

2.4 Implementation details

Our model is trained on unpaired data. All experiments were implemented in PyTorch and trained on an NVIDIA 4090 GPU with CUDA acceleration. The networks were optimized using the Adam optimizer with an initial learning rate of 2×10^{-4} and $\beta_1 = 0.5$, where the learning rate was kept constant for the first 150 epochs and then linearly decayed to zero over the following 20 epochs. An unaligned dataset setting with an high-to-low translation direction was adopted, and all input images were resized to 286×286 and randomly cropped to 256×256 during training, with a batch size of 128. The model followed the standard CycleGAN configuration with 64 base feature channels for both generator and discriminator, employed a PatchGAN discriminator, and was trained using the least-squares GAN objective, together with an image buffer of size 50 to stabilize training. Note that although HRTEM-GAN is trained with bidirectional mappings between LQ and HQ domains, practical deployment primarily uses the LQ→HQ branch.

3 Results and discussion

3.1 Evaluation protocols

To comprehensively evaluate the proposed method, we conducted both image generation quality assessment and recognition performance evaluation. For image generation, quantitative metrics were adopted to measure the distribution consistency between generated images and real images, complemented by qualitative feature-space visualization. Specifically, Fréchet Inception Distance (FID) [34] and Kullback-Leibler (KL) divergence [35] were used for quantitative evaluation, while t-SNE visualization [36] was employed for qualitative analysis. For recognition, standard pixel-level metrics were adopted to assess the accuracy and overlap between predicted masks and ground-truth annotations, including Precision, Dice coefficient, and Intersection over Union (IoU). The formulas for computing the evaluation metrics are provided in Supplementary Note III.

As comparison baselines, we compared HRTEM-GAN with both conventional denoising algorithms and recent learning-based restoration methods. Specifically, we included Wiener filtering and ABSF as representative classical frequency-/filtering-based approaches, and BM3D as a widely used patch-based denoiser. In addition, we evaluated three learning-based alternatives, i.e., AtomSegNet [18], SBD [12], and SFIN [22], which have been adopted in prior work for improving TEM image quality and downstream atomic recognition. To assess recognition performance in a controlled manner, we used the same segmentation network (AtomSegNet) to predict atomic masks from the restored images produced by each restoration method, and then computed pixel-level metrics (Precision, Dice, and IoU) against the ground-truth annotations.

3.2 Quantitative results of restoration

We first quantify the distribution-level fidelity of restored images to the HQ domain using feature-based metrics, aiming to evaluate whether a restoration method genuinely narrows the gap between LQ inputs and HQ references in the learned

representation space. Specifically, we report the FID and KL divergence computed between restored foreground patches and real HQ patches. The quantitative comparison across different methods is summarized in Table 1.

Table 1. Quantitative comparison of restored results produced by our method and existing approaches. "KL" denotes KL divergence. The best and second-best results are highlighted in **bold** and underlined, respectively.

Metric	Conventional methods			Learning-based methods			
	Wiener	ABSF	BM3D	AtomSegNet	SBD	SFIN	Ours
FID ↓	215.825	223.188	132.504	172.224	<u>169.088</u>	211.103	**33.622**
KL ↓	1.530	1.320	0.424	<u>0.326</u>	0.402	0.348	**0.092**

Among the conventional denoising methods, BM3D achieves the strongest performance, with an FID of 132.504 and a KL divergence of 0.424, while Wiener and ABSF exhibit substantially larger distribution gaps ($\mathrm{FID} > 215$ and $\mathrm{KL} \geq 1.320$). The learning-based counterparts (AtomSegNet, SBD, and SFIN) reduce KL divergence to a moderate level (0.326–0.402) but still yield relatively high FID values (169.088–211.103), indicating that the restored outputs remain far from the target HQ distribution in feature space. In contrast, our HRTEM-GAN achieves the lowest FID (33.622) and KL divergence (0.092), outperforming all competing approaches by a large margin and demonstrating substantially improved distribution alignment and restoration fidelity.

3.3 Qualitative results of restoration

3.3.1 Patch-level qualitative comparison

We first inspect atomic-level detail recovery through local magnification. As shown in Fig. 3, two regions are cropped from each HRTEM image restored by BM3D, AtomSegNet, SBD, SFIN, and our proposed method. It is observed that BM3D and other deep learning-based methods are effective in noise reduction but fail to recover the underlying atomic structures.

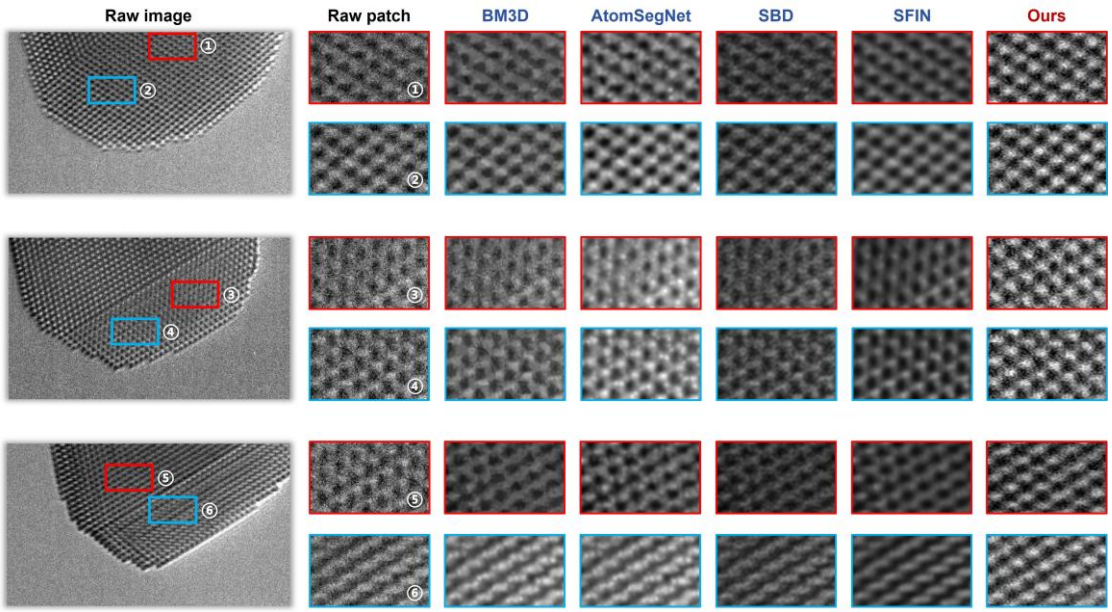

Figure 3. Qualitative comparison of patch-level restored results produced by our method and existing approaches.

Two representative cases further highlight the difference. In region (3), where the raw observation is severely blurred and atomic contrast is barely discernible, other methods mainly suppress noise but still fail to reveal stable lattice periodicity. In contrast, our HRTEM-GAN recovers clearer atomic-point contrast with sharper boundaries, making the underlying lattice pattern more distinguishable. In region (6), adjacent atomic columns appear partially merged due to imaging degradations; competing methods tend to preserve the adhesion after denoising, whereas HRTEM-GAN better separates the fused atomic points and restores more consistent inter-column boundaries. Similar trends are observed across the remaining regions.

3.3.2 Image-level qualitative comparison

To evaluate overall restoration quality beyond local patches, we compare full-image reconstructions produced by HRTEM-GAN and competing approaches. Following the patch-based inference setting, we adopt an overlapping sliding-window strategy to aggregate patch-wise predictions into a complete image: each restored patch is mapped back to its spatial location, and pixel-wise accumulation followed by averaging is applied within overlapping areas. Afterward, redundant regions introduced

by boundary padding are cropped, and the reconstructed image is resized to the original resolution to obtain the final restored output. Figure 4 presents representative image-level results across four test images.

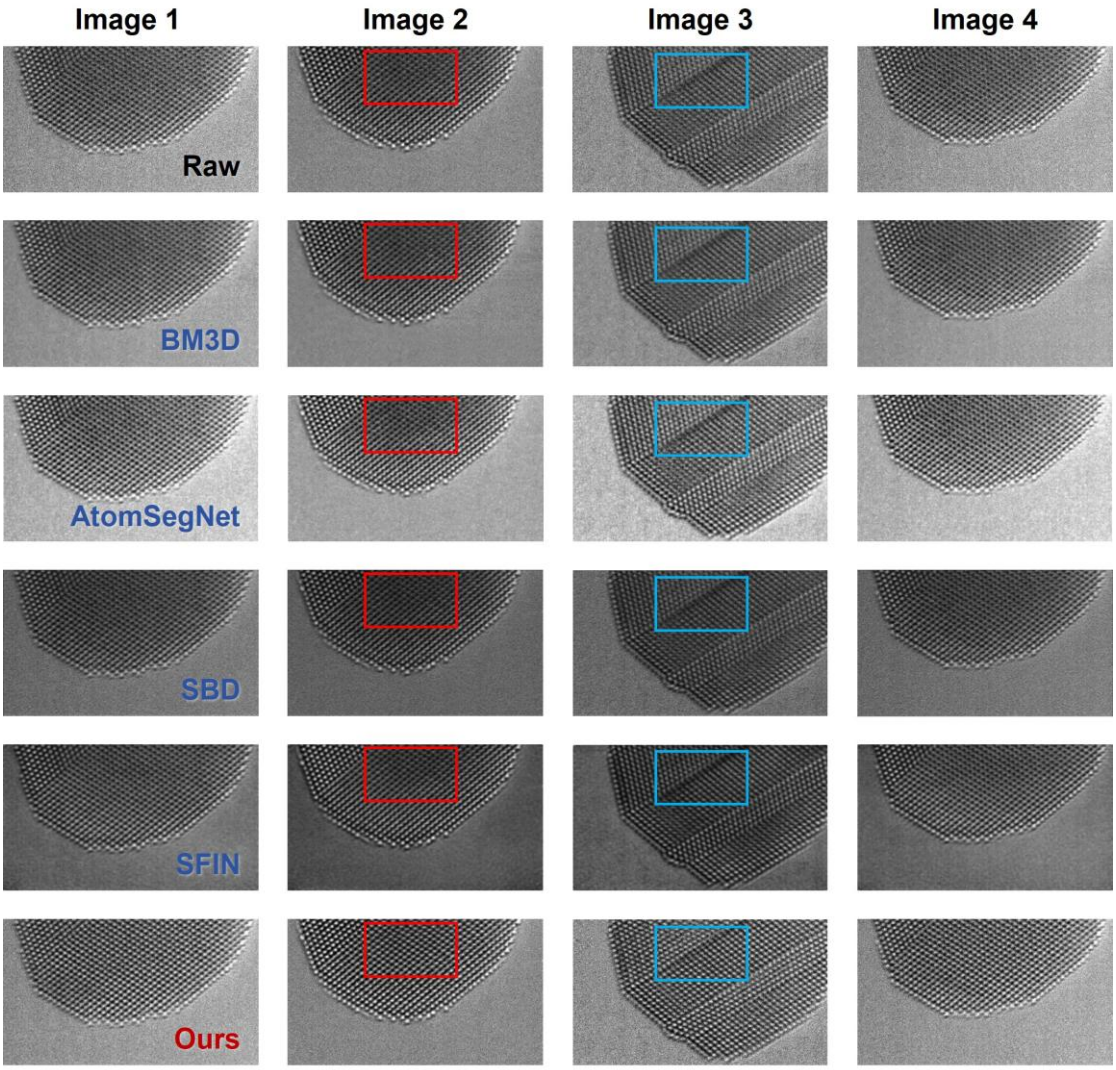

Figure 4. Qualitative comparison of image-level restored results produced by our method and existing approaches. Columns correspond to different test images, and rows correspond to different restoration methods (method names are indicated on the left). The red and blue bounding boxes mark representative regions in Image 2 and Image 3.

As shown in Figure 4, BM3D and the learning-based methods (AtomSegNet, SBD, and SFIN) reduce noise to some extent, but their outputs remain largely smoothing-dominated, with attenuated atomic contrast and locally unstable lattice textures. This

limitation is particularly evident in Image 2 and Image 3, where the highlighted regions emphasize challenging structures. In Image 2 (red box), the competing methods tend to over-smooth the lattice fringes, leading to weakened peak–valley contrast and less well-defined atomic columns in the interior lattice region. In contrast, HRTEM-GAN preserves clearer atomic-point contrast and more regular lattice periodicity within the boxed area, maintaining sharper boundaries between adjacent columns while suppressing background noise. In Image 3 (blue box), the specimen exhibits a twin-boundary-like (or stacking-fault-like) interfacial contrast, where the lattice orientation and fringe continuity change across a narrow region. Such defect- or interface-related patterns are particularly challenging under low SNR, as they require preserving not only periodic lattice fringes but also the coherence across the interface. The other methods tend to over-smooth this area and partially wash out the interfacial contrast, resulting in blurred textures and locally disrupted periodicity. Oppositely, HRTEM-GAN better maintains the interface-induced contrast variation while preserving the surrounding lattice order, yielding a more coherent reconstruction with clearer atomic features in the highlighted region. Overall, these observations indicate that the proposed method goes beyond generic denoising and yields restorations that are visually closer to high-quality TEM observations.

3.3.3 Spatial-frequency consistency analysis

Beyond spatial-domain appearance, a key indicator of physically meaningful HRTEM restoration is whether the method recovers the characteristic lattice periodicity in the frequency domain. We compute the 2D FFT log-magnitude after applying a two-dimensional Hann window and visualize it with the zero-frequency component centered. This process suppresses boundary-induced leakage and yields more stable reciprocal-space patterns for assessing lattice-related high-frequency content.

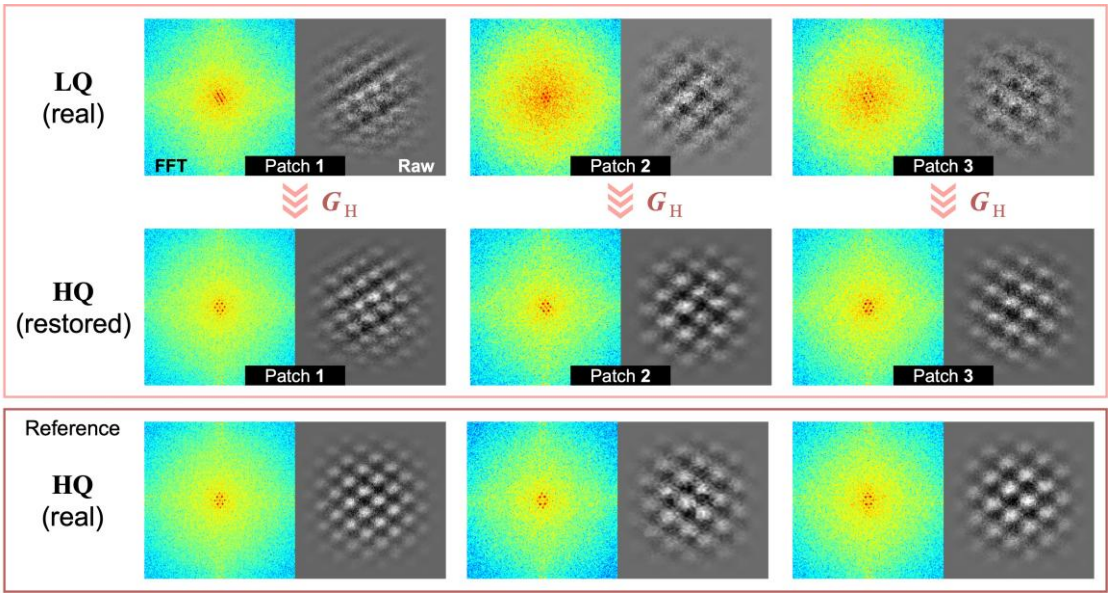

Figure 5. Spatial-frequency comparison of restoration results. The bottom row provides representative real HQ patches as a reference for the expected reciprocal-space patterns (unpaired with the LQ inputs).

Under this spatial-frequency inspection (see Fig. 5), LQ experimental patches (the first row) exhibit attenuated and diffuse high-frequency responses, where reciprocal-lattice signatures are weak and partially blurred, consistent with diminished fringe contrast in the spatial domain. In contrast, the outputs restored by G_H (the second row) show clearer atomic-column modulation and sharper lattice fringes, accompanied by more discernible Bragg-like peaks (or reciprocal-lattice spots) and reduced spectral smearing in the log-magnitude representation. These observations suggest that our HRTEM-GAN improves image quality by reconstructing lattice periodicity rather than merely amplifying contrast. Representative HQ real patches are additionally provided as a reference (not paired with the corresponding LQ inputs) to illustrate typical reciprocal-space patterns of high-quality images, which are qualitatively better matched by the restored results.

3.3.4 t-SNE visualization analysis

To examine whether restored images are not only visually improved but also feature-wise aligned with the high-quality (HQ) domain, we visualize the deep-feature

embeddings of foreground patches using t-SNE in Fig. 6, where red circles denote real HQ patches and blue triangles denote the corresponding patches obtained after restoration.

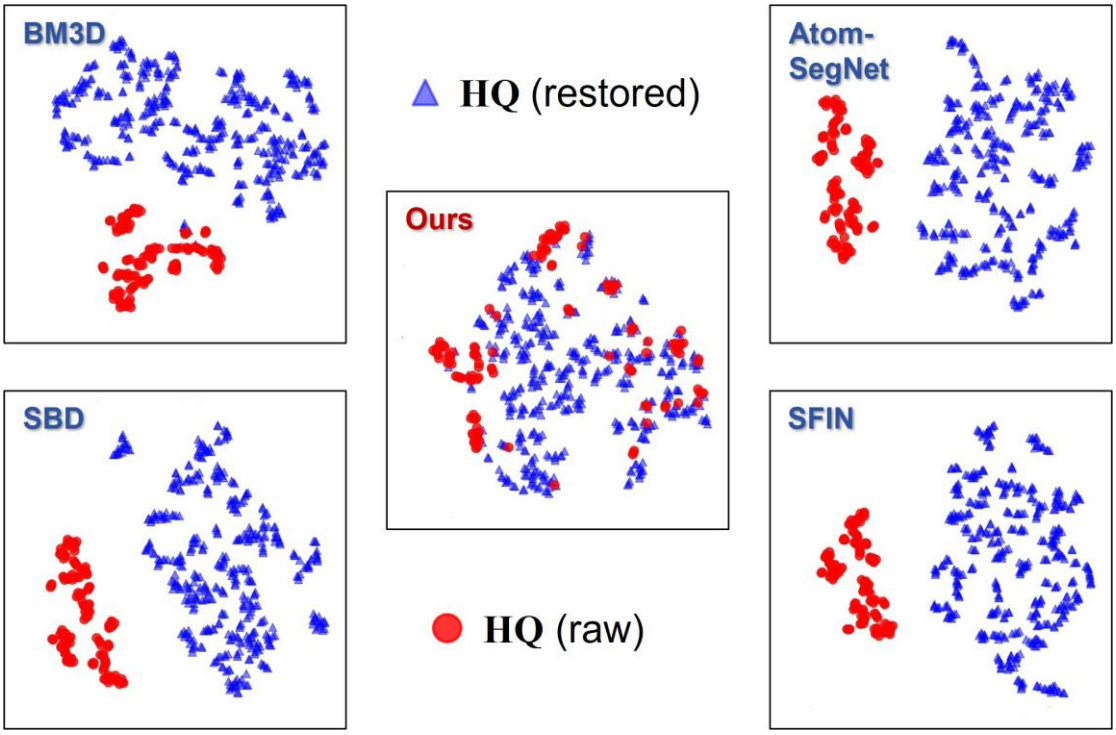

Figure 6. t-SNE visualization of deep-feature embeddings for atomic patches under different restoration methods. Red circles denote real HQ (raw) patches, and blue triangles denote the corresponding patches obtained after restoration.

After applying BM3D and the learning-based methods (AtomSegNet, SBD, and SFIN), the restored features remain noticeably detached from the HQ cluster and often become more scattered, suggesting that these methods mainly perform local denoising/enhancement without effectively mapping the low-quality inputs toward the HQ feature distribution. In contrast, the embeddings produced by HRTEM-GAN show substantially increased overlap with the HQ points and reduced inter-domain discrepancy, indicating that the proposed model better recovers high-frequency structural characteristics and overall statistical properties consistent with the HQ domain, thereby achieving more effective distribution-level alignment.

3.4 Performance of atomic recognition

To further evaluate how well different restoration models preserve atomic-level structures, we apply the trained AtomSegNet to segment atomic columns on the restored outputs of representative DL method (SBD and SFIN) and quantify structural fidelity using segmentation-based metrics (Precision, Dice, and IoU) against the ground-truth labels, and additionally provide qualitative comparisons of the predicted masks to visually assess atomic-column completeness, separability, and boundary consistency under challenging imaging conditions.

Table 2. Quantitative comparison of segmentation performance under different restoration methods.

Method	Precision	Dice ↑	IoU ↑
SBD	0.468	0.548	0.386
SFIN	0.692	<u>0.691</u>	<u>0.533</u>
Ours	**0.806**	**0.731**	**0.578**

Table 2 summarizes the segmentation accuracy on the restored images. Compared with SBD and SFIN, our method yields consistent gains across all metrics, indicating improved atomic-column structural fidelity after restoration. Specifically, our method achieves a Precision of 0.806, which is higher than SFIN (0.692) and SBD (0.468), suggesting fewer false positives and better separability between adjacent atomic columns. Likewise, the Dice score increases to 0.731 (vs. 0.691 for SFIN and 0.548 for SBD), and IoU improves to 0.578 (vs. 0.533 for SFIN and 0.386 for SBD), reflecting better mask overlap and more complete atomic-column recovery. These quantitative improvements are consistent with the qualitative comparison below.

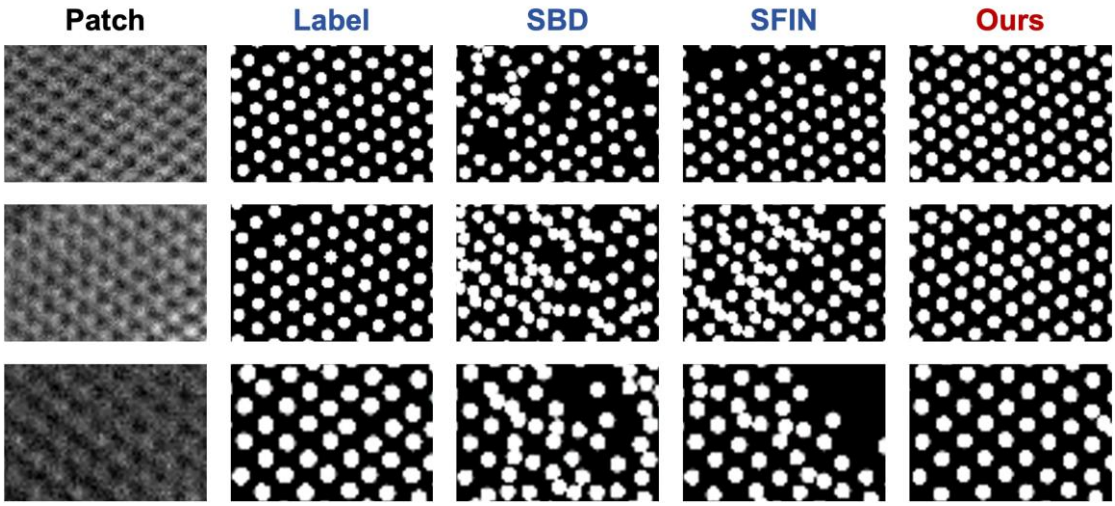

Figure 7. Qualitative comparison of atomic-column segmentation results.

As shown in Fig. 7, although SBD and SFIN improve the segmentation quality in certain areas, their results still suffer from issues such as atomic region merging, positional shifts, and disruption of lattice structure. In contrast, our method consistently produces more accurate and coherent segmentation results across all test patches. The segmentation results generated by our method exhibit uniformly distributed atomic regions and well-preserved lattice periodicity, maintaining complete and regular atomic structures.

Overall, the segmentation results demonstrates that HRTEM-GAN outperforms other learning-based restoration methods. The images restored by HRTEM-GAN enable the atomic segmentation network to achieve more accurate and stable atomic identification, indicating superior structural fidelity in atomic-resolution images. This advantage provides a more reliable image foundation for subsequent atomic modeling and materials analysis.

4 Conclusion

In this paper, we have developed an HRTEM-image restoration method (HRTEM-GAN) built on the generative method, CycleGAN. Targeted at the atomic-resolution scenario, HRTEM-GAN restores low-quality HRTEM observations into high-quality counterparts while preserving critical structure information, without requiring strictly

paired low-/high-quality training data. We combine spatial restoration with frequency-consistent supervision and align the restored features with real high-quality features to enhance the reality. The ViT-based design at the bottleneck helps capture long-range lattice regularity and improves structural faithfulness in complex regions. Extensive experiments demonstrate that the restored outputs exhibit improved visual fidelity and closer distributional alignment with real high-quality images in both pixel space and feature space. Importantly, the enhanced image quality consistently improves downstream atomic-column segmentation performance, yielding higher Precision, Dice, and IoU compared with recent learning-based restoration baselines.

In future work, we will explore more condition-aware and physics-guided constraints [37,38] (e.g., incorporating imaging parameter priors or forward models) as well as broader multi-domain training to improve robustness. We also plan to investigate lightweight deployment strategies to facilitate practical integration into routine HRTEM workflows. Overall, HRTEM-GAN provides a practical and effective pathway for restoring atomic-resolution HRTEM images and enhancing subsequent quantitative analysis.

Acknowledgments This work is supported by the National Major Science and Technology Projects of China (2025ZD0620003) and the National Natural Science Foundation of China under Grant 00000000, 00000000, and 00000000.

Compliance with ethics guidelines

Conflict of interest

All authors declare that they have no conflict of interest.

References

[1] Sun LG, Wu G, Wang Q, et al. Nanostructural metallic materials: Structures and mechanical properties. Mater Today 2020;38:114–35.

[2] Liu Y, Niu C, Wang Z, et al. Machine learning in materials genome initiative: A review. J Mater Sci Technol 2020;57:113–22.

[3] Morgan D, Jacobs R. Opportunities and Challenges for Machine Learning in Materials Science. Annu Rev Mater Res 2020;50:71–103.

[4] Liu Y, Zhao T, Ju W, et al. Materials discovery and design using machine learning. J Materiomics 2017;3:159–77.

[5] Liu G, Shih AJ, Deng H, et al. Site-specific reactivity of stepped Pt surfaces driven by stress release. Nature 2024;626:1005–10.

[6] Crozier PA, Leibovich M, Haluai P, et al. Visualizing nanoparticle surface dynamics and instabilities enabled by deep denoising. Science 2025;387:949–54.

[7] Ziletti A, Kumar D, Scheffler M, et al. Insightful classification of crystal structures using deep learning. Nat Commun 2018;9:2775.

[8] Ziatdinov M, Dyck O, Maksov A, et al. Deep Learning of Atomically Resolved Scanning Transmission Electron Microscopy Images: Chemical Identification and Tracking Local Transformations. ACS Nano 2017;11:12742–52.

[9] Ge M, Su F, Zhao Z, et al. Deep learning analysis on microscopic imaging in materials science. Mater Today Nano 2020;11:100087.

[10] Yang S-H, Choi W, Cho BW, et al. Deep Learning-Assisted Quantification of Atomic Dopants and Defects in 2D Materials. Adv Sci 2021;8:2101099.

[11] Botifoll M, Pinto-Huguet I, Arbiol J. Machine learning in electron microscopy for advanced nanocharacterization: current developments, available tools and future outlook. Nanoscale Horiz 2022;7:1427–77.

[12] Mohan S, Manzorro R, Vincent JL, et al. Deep denoising for scientific discovery: A case study in electron microscopy. IEEE Trans Comput Imaging 2022;8:585–97.

[13] He R, Cheng R, Lyu X, et al. Efficient online training for zero-shot time-lapse microscopy denoising and super-resolution. Proceedings of the AAAI Conference on Artificial Intelligence 2025;39:3419–27.

[14] Kilaas R. Optimal and near-optimal filters in high-resolution electron microscopy. Journal of Microscopy 1998;190:45–51.

[15] Mevenkamp N, Binev P, Dahmen W, et al. Poisson noise removal from high-resolution STEM images based on periodic block matching. Adv Struct Chem Imag 2015;1:3.

[16] Dabov K, Foi A, Katkovnik V, et al. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. IEEE Transactions on Image Processing 2007;16:2080–95.

[17] Choudhary K, DeCost B, Chen C, et al. Recent advances and applications of deep learning methods in materials science. Npj Comput Mater 2022;8:1–26.

[18] Lin R, Zhang R, Wang C, et al. TEMImageNet training library and AtomSegNet deep-learning models for high-precision atom segmentation, localization, denoising, and deblurring of atomic-resolution images. Sci Rep 2021;11:5386.

[19] Lobato I, Friedrich T, Van Aert S. Deep convolutional neural networks to restore single-shot electron microscopy images. Npj Comput Mater 2024;10:1–19.

[20] Kushwaha HS, Tanwar S, Rathore KS, et al. De-noising Filters for TEM (Transmission Electron Microscopy) Image of Nanomaterials. 2012 Second International Conference on Advanced Computing & Communication Technologies, 2012, p. 276–81.

[21] Kazimi B, Ruzaeva K, Sandfeld S. Self-Supervised Learning with Generative Adversarial Networks for Electron Microscopy. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024, p. 71–81.

[22] Li H, Wu Z, Shao R, et al. Noise calibration and spatial-frequency interactive network for STEM image enhancement. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), 2025, p. 21287–96.

[23] Lee H-Y, Tseng H-Y, Huang J-B, et al. Diverse Image-to-Image Translation via Disentangled Representations. Proceedings of the European Conference on Computer Vision (ECCV), 2018, p. 35–51.

[24] Huang X, Liu M-Y, Belongie S, et al. Multimodal Unsupervised Image-to-image Translation. Proceedings of the European Conference on Computer Vision (ECCV), 2018, p. 172–89.

[25] Nizan O, Tal A. Breaking the Cycle - Colleagues Are All You Need, 2020, p. 7860–9.

[26] Zhao Y, Wu R, Dong H. Unpaired Image-to-Image Translation Using Adversarial Consistency Loss. Proceedings of the European Conference on Computer Vision (ECCV), 2020, p. 800–15.

[27] Khan A, Lee C-H, Huang PY, et al. Leveraging generative adversarial networks to create realistic scanning transmission electron microscopy images. Npj Comput Mater 2023;9:1–9.

[28] Eliasson H, Lothian A, Surin I, et al. Precise Size Determination of Supported Catalyst Nanoparticles via Generative AI and Scanning Transmission Electron Microscopy. Small Methods 2025;9:2401108.

[29] Quan TM, Hildebrand DGC, Lee K, et al. Removing Imaging Artifacts in Electron Microscopy using an Asymmetrically Cyclic Adversarial Network without Paired Training Data. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, p. 3804–13.

[30] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. International Conference on Learning Representations, 2021.

[31] Isola P, Zhu J-Y, Zhou T, et al. Image-To-Image Translation With Conditional

Adversarial Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, p. 1125–34.

[32] Wang S, Yang Y, Liu Z, et al. Dataset distillation with neural characteristic function: a minmax perspective. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), 2025, p. 25570–80.

[33] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical image computing and computer-assisted intervention – MICCAI 2015, Cham: Springer International Publishing; 2015, p. 234–41.

[34] Heusel M, Ramsauer H, Unterthiner T, et al. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc.; 2017.

[35] Perez-Cruz F. Kullback-Leibler divergence estimation of continuous distributions. 2008 IEEE International Symposium on Information Theory, 2008, p. 1666–70.

[36] Maaten L van der, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res 2008;9:2579–605.

[37] Lu C, Chen K, Qiu H, et al. Diffusion-based deep learning method for augmenting ultrastructural imaging and volume electron microscopy. Nat Commun 2024;15:4677.

[38] Wang S, Liu X, Li Y, et al. A deep learning-based stripe self-correction method for stitched microscopic images. Nat Commun 2023;14:5393.

Supplementary Material for

From Low-Quality to High-Quality: Generating Structure-Preserved Atomic-Scale HRTEM images via an Enhanced CycleGAN

Supplementary Note I: Image quality classification and reconstruction

Image quality classification and reconstruction pipeline. Atomic-scale microscopy video streams were acquired using transmission electron microscopy (TEM), from which individual frames were manually screened according to image quality. Twenty high-quality images and twenty low-quality images were selected to construct an unpaired dataset. The original images, with a resolution of 1936×1112, were uniformly resized to 256×256 prior to training and subsequently divided into non-overlapping patches of size 32×32. For each patch, foreground/background pseudo-labels were automatically generated using multiple unsupervised clustering strategies, including clustering based on grayscale distribution features, statistical features such as mean intensity and intensity standard deviation, as well as frequency-domain and structural statistical features. The results from different clustering methods were integrated via a voting mechanism to obtain robust foreground/background labels. In addition, each patch inherited a high-quality or low-quality domain label from its source image. Patch samples annotated with both patch-level and image-level labels were then used to jointly learn reconstruction and classification tasks at the patch level. During training, a U-Net based encoder-decoder architecture was employed as the backbone network, in which the encoder consists of five hierarchical convolutional blocks with downsampling operations to progressively extract multi-scale features, while the decoder symmetrically restores spatial resolution through upsampling combined with skip connections, ultimately producing a single-channel reconstructed output.

Training objectives and optimization. The proposed framework was trained in a multi-task manner, jointly optimizing image reconstruction and classification objectives at the patch level. For reconstruction, the network was supervised to recover

the grayscale intensity of the input patch, where the RGB input was converted to a single-channel grayscale image using a linear combination of color channels. To emphasize the restoration of structurally meaningful regions, the reconstruction loss was computed only on patches labeled as foreground, while background patches were excluded from reconstruction supervision. The reconstruction objective was formulated using the mean squared error (MSE) loss between the reconstructed output and the corresponding grayscale input. In addition, two classification objectives were introduced: a patch-level foreground/background classification loss, which encourages the network to distinguish structural regions from background regions, and an image-level quality classification loss, which predicts whether a patch originates from a high-quality or low-quality image domain. These objectives were combined into a weighted loss function, where the reconstruction loss was assigned a higher weight to prioritize structural fidelity, and the classification losses served as auxiliary constraints to guide feature learning. The network was optimized using the Adam optimizer with an initial learning rate of 1×10-3. Training was performed with a batch size of 16 for 50 epochs. During optimization, the total loss was minimized as a weighted sum of the reconstruction loss, the patch-level classification loss, and the image-level classification loss. Model selection was based on the validation loss, and the network parameters yielding the lowest validation loss were retained for inference.

Supplementary Note II: Atomic region segmentation

Training Pipeline. The foreground segmentation model was trained using a fully supervised learning pipeline based on conventional U-Net architecture. The training dataset consists of 105 atomic-resolution images, for which the corresponding foreground masks were manually annotated using LabelMe to delineate atomic regions of interest. Input images and their associated masks were paired on a per-file basis to ensure strict one-to-one correspondence between images and labels. During training, identical random geometric augmentations were synchronously applied to each image-mask pair to preserve spatial alignment. Specifically, center cropping and random rotations were performed on both the input images and their corresponding masks,

while appearance-based intensity augmentations were applied exclusively to the input images. After preprocessing and normalization, the input images were fed into the U-Net model, which outputs a single-channel prediction map representing pixel-wise foreground likelihood. For stable model training and evaluation, the dataset was randomly divided into training and validation subsets, and mini-batch optimization was carried out using a standard data-loading strategy. During validation, the predicted probability maps were resized back to the original image resolution for visualization and further analysis.

Training Objectives and Optimization. The model was trained to perform binary foreground segmentation using a composite loss function that jointly enforces pixel-wise classification accuracy and region-level overlap consistency. Specifically, the training objective combines a binary cross-entropy loss with logits and a Dice loss computed on sigmoid-activated predictions, thereby balancing local discrimination capability with global foreground shape consistency. Model parameters were optimized using the Adam optimizer, with learning rate scheduling applied throughout training to regulate the optimization process. Training was conducted for a fixed number of epochs, with model performance evaluated on a held-out validation set after each epoch. Quantitative evaluation on the validation set demonstrates that the proposed segmentation model achieves a Dice coefficient of 0.9837, indicating highly accurate delineation of atomic regions.

Supplementary Note III: CycleGAN loss formulation

In CycleGAN, the discriminators are updated by back-propagating the loss corresponding to failures in distinguishing real and translated images, commonly referred to as the generative adversarial loss (GAN loss). In this work, we define two image domains, where domain A corresponds to the low-quality HRTEM images and domain B represents the high-quality images:

	$\mathcal{L}_{disc,A} = \mathbb{E}_{x\sim B}\ell_{GAN}\left(\mathcal{D}_A\left(\mathcal{G}_{B\to A}(x)\right),0\right) + \mathbb{E}_{x\sim A}\ell_{GAN}(\mathcal{D}_A(x),1)\,,$	(1)

here, ℓ_{GAN} denotes the adversarial loss used in CycleGAN, which can take different forms depending on the GAN variant, such as binary cross-entropy, least-squares loss, or the Wasserstein objective. The labels 0 and 1 indicate fake and real images, respectively, when applicable. The generators are updated by back-propagating loss from three sources: GAN loss, cycle-consistency loss, and identity-consistency loss. Using $G_{A\to B}$ as an example:

	$\mathcal{L}_{GAN,A} = \mathbb{E}_{x\sim A}\ell_{GAN}\big(\mathcal{D}_A\big(\mathcal{G}_{A\to B}(x)\big), 1\big),$	(2)
	$\mathcal{L}_{cyc,A} = \mathbb{E}_{x\sim A}\ell_{reg}\big(\mathcal{G}_{B\to A}\big(\mathcal{G}_{A\to B}(x)\big), x\big),$	(3)
	$\mathcal{L}_{idt,A} = \mathbb{E}_{x\sim A}\ell_{reg}(\mathcal{G}_{B\to A}(x), x).$	(4)

And,

	$\mathcal{L}_{gen,A\to B} = \mathcal{L}_{GAN,A} + \lambda_{cyc}\mathcal{L}_{cyc,A} + \lambda_{idt}\mathcal{L}_{idt,A},$	(5)
	$\mathcal{L}_{gen,B\to A} = \mathcal{L}_{GAN,B} + \lambda_{cyc}\mathcal{L}_{cyc,B} + \lambda_{idt}\mathcal{L}_{idt,B},$	(6)
	$\mathcal{L}_{CycleGAN} = \mathcal{L}_{disc,A} + \mathcal{L}_{disc,B} + \mathcal{L}_{gen,A\to B} + \mathcal{L}_{gen,B\to A},$	(7)

here, ℓ_{reg} can be any regression loss function, and λ_{cyc} and λ_{idt} are combination coefficients.

Supplementary Note IV: Characteristic-function construction

Let $z_H \in \mathbb{R}^{C\times H\times W}$ and $z_L \in \mathbb{R}^{C\times H\times W}$ denote the fused bottleneck feature maps produced by G_H and G_L, respectively. We interpret each spatial location as a C-dimensional feature vector and form two empirical feature sets:

$$\mathcal{U}_H = \{\mathbf{u}_n^H\}_{n=1}^N, \mathcal{U}_L = \{\mathbf{u}_n^L\}_{n=1}^N, N = H\cdot W,$$

where $\mathbf{u}_n^H \in \mathbb{R}^C$ (resp. $\mathbf{u}_n^L$) is obtained by flattening z_H (resp. z_L) over spatial dimensions. In practice, to reduce computation and improve robustness, we optionally subsample N' vectors uniformly from the N locations (or from multiple images within a mini-batch) and apply channel-wise normalization to stabilize the scale of features.

Given a set of probing directions $\{\mathbf{t}_m\}_{m=1}^{M} \subset \mathbb{R}^C$ (shared by both domains), the empirical characteristic function of a feature set $\mathcal{U}$ is computed as

$$\hat{\varphi}(\mathbf{t}_m; \mathcal{U}) = \frac{1}{|\mathcal{U}|} \sum_{\mathbf{u} \in \mathcal{U}} \exp\ (i\, \mathbf{t}_m^{\top} \mathbf{u}), m = 1, \ldots, M\ ,$$

which yields complex-valued responses. We use the same $\{\mathbf{t}_m\}$ for both $\mathcal{U}_H$ and $\mathcal{U}_L$ to ensure a consistent comparison. In our implementation, $\{\mathbf{t}_m\}$ can be sampled once and kept fixed (e.g., i.i.d. from a zero-mean Gaussian with a controlled scale, or uniformly on the unit sphere followed by a fixed radius), which provides a stable set of "probes" for distribution matching.

We adopt an amplitude–phase decomposition of the CF response:

$$A(\mathbf{t}_m) = |\, \hat{\varphi}(\mathbf{t}_m)\, |, \Phi(\mathbf{t}_m) = \angle \hat{\varphi}(\mathbf{t}_m),$$

where $|\cdot|$ and $\angle(\cdot)$ denote the magnitude and phase (computed via $\operatorname{atan2}(\Im(\cdot), \Re(\cdot))$), respectively. The CF loss enforces both magnitude (coverage) and phase (alignment) consistency between domains:

$$\mathcal{L}_{\mathrm{CF}} = \frac{1}{M} \sum_{m=1}^{M} (|\, A_H(\mathbf{t}_m) - A_L(\mathbf{t}_m)\, | + \lambda_{\phi} (1 - \cos\ (\Phi_H(\mathbf{t}_m) - \Phi_L(\mathbf{t}_m)))),$$

where $A_H(\mathbf{t}_m) = |\, \hat{\varphi}(\mathbf{t}_m; \mathcal{U}_H)\, |$ and $A_L(\mathbf{t}_m) = |\, \hat{\varphi}(\mathbf{t}_m; \mathcal{U}_L)\, |$ (similarly for Φ_H, Φ_L). The phase term uses the circular distance $1 - \cos\ (\Delta\Phi)$ to respect the 2π-periodicity of angles and to avoid discontinuities. Unless otherwise stated, both terms are averaged over the M probing directions and over mini-batches.

Practical note: we set small constants (e.g., ϵ) where needed for numerical stability when computing phases, and we compute $\mathcal{L}_{\mathrm{CF}}$ on mini-batch features so that it serves as a stochastic approximation of dataset-level distribution matching.

Formally, the characteristic function of the feature distribution is defined as follows:

$$\Phi_x(t) = \mathbb{E}[\cos(\langle t, x \rangle)] + j\mathbb{E}[\sin(\langle t, x \rangle)]\ , \tag{8}$$

where x denotes the feature representation, t is the frequency argument, and the cosine term represents the amplitude component while the sine term captures the phase component of the feature distribution.

Supplementary Note V: Evaluation metrics

FID computes the distance between two multivariate Gaussian distributions fitted to deep feature representations of real and generated images, defined as:

$$\mathrm{FID} = \|\mu_r - \mu_g\|_2^2 + \mathrm{Tr}(\sum_r + \sum_g - 2(\sum_r \sum_g)^{1/2}), \tag{10}$$

where $(\mu_r, \sum_r)$ and $(\mu_g, \sum_g)$ denote the mean and covariance of features extracted from real and generated images, respectively. A lower FID indicates closer feature distributions.

KL divergence is employed to further quantify the difference between the two distributions, expressed as:

$$D_{\mathrm{KL}}(P\|Q) = \sum_x P(x)\log\frac{P(x)}{Q(x)}, \tag{11}$$

where P and Q represent the feature distributions of real and generated samples. Smaller KL values imply higher distributional consistency.

t-SNE visualization for projecting high-dimensional features into a low-dimensional space, enabling an intuitive comparison of feature clustering between real and generated images.

To assess recognition accuracy, we employed Precision, Dice coefficient, and Intersection over Union (IoU) as evaluation metrics. Precision measures the proportion of correctly predicted foreground pixels and is defined as:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{12}$$

where TP and FP denote the numbers of true positive and false positive pixels, respectively. The Dice coefficient evaluates the overlap between predicted masks and ground truth, formulated as:

$$\mathrm{Dice} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{13}$$

while IoU is defined as:

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{14}$$

with FN representing false negatives. Higher values of Precision, Dice, and IoU indicate better recognition performance.