arXiv:2508.10359v1 [cs.CV] 14 Aug 2025

AtomDiffuser: Time-Aware Degradation Modeling for Drift and Beam Damage in STEM Imaging

Hao Wang[1], Hongkui Zheng[2], Kai He[2], and Abolfazl Razi[1*]
[1]Clemson University
[2]University of California, Irvine
{hao9, arazi}@clemson.edu, {kai.he, zhongkui}@uci.edu

Abstract

Scanning transmission electron microscopy (STEM) plays a critical role in modern materials science, enabling direct imaging of atomic structures and their evolution under external interferences. However, interpreting time-resolved STEM data remains challenging due to two entangled degradation effects: spatial drift caused by mechanical and thermal instabilities, and beam-induced signal loss resulting from radiation damage. These factors distort both geometry and intensity in complex, temporally correlated ways, making it difficult for existing methods to explicitly separate their effects or model material dynamics at atomic resolution. In this work, we present AtomDiffuser, a time-aware degradation modeling framework that disentangles sample drift and radiometric attenuation by predicting an affine transformation and a spatially varying decay map between any two STEM frames. Unlike traditional denoising or registration pipelines, our method leverages degradation as a physically heuristic, temporally conditioned process, enabling interpretable structural evolutions across time. Trained on synthetic degradation processes, AtomDiffuser also generalizes well to real-world cryo-STEM data. It further supports high-resolution degradation inference and drift alignment, offering tools for visualizing and quantifying degradation patterns that correlate with radiation-induced atomic instabilities. Our source code and sample datasets are available at: https://arazi2.github.io/aisends.github.io/project/Atom

1. Introduction

Understanding how materials evolve at the atomic scale is central to modern materials science, with implications ranging from battery durability to semiconductor reliability [35]. Many key properties, such as conductivity, mechanical strength, and chemical stability, are governed by atomic arrangements and their transformations over time [25, 32]. Scanning transmission electron microscopy (STEM), transmission electron microscopy (TEM), and cryogenic variants (cryo-STEM/TEM) enable visualization of individual atomic columns and structural changes under controlled environments [17, 35]. While these imaging modalities enable direct observation of atomic-scale degradation, their application to beam-sensitive materials remains challenging, as even low-dose electron exposure can lead to atomic displacement, defect formation, or irreversible chemical changes [6, 12, 31]. For instance, the garnet-type solid electrolyte $\mathbf{Li_7La_3Zr_2O_{12}}$ (LLZO) is a promising candidate for solid-state lithium batteries [36]. However, its crystal lattice is prone to beam-induced degradation even under cryogenic temperatures, and tools for quantitatively analyzing such degradation pathways remain limited due to its current way of imaging.

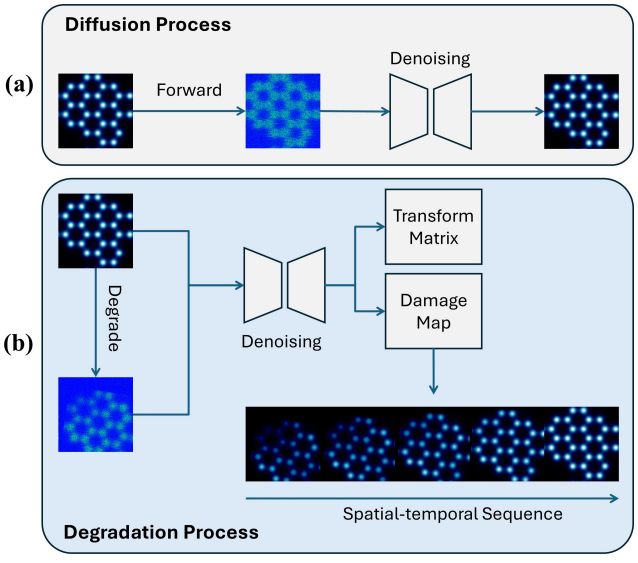

Figure 1. Differences between the diffusion process [24] and the degradation process [1].

*Corresponding author

During atomic-resolution STEM imaging, two degradation mechanisms: atomic drift and beam-induced damage, intertwined, pose fundamental challenges to accurate structural interpretation [10]. Atomic drift arises from specimen charging, thermal expansion, or mechanical instabilities in the stage, leading to subpixel displacements of atomic columns across sequential frames. Although drift does not alter signal intensity, it distorts geometric structure and may mimic dynamic phenomena [10]. In contrast, beam-induced damage stems from inelastic or elastic energy transfer, causing irreversible structural alterations such as atom ejection, radiolytic bond scission, or elemental loss [5]. These effects manifest as intensity reduction, blurring, or structural disappearance. Since drift and damage often co-occur and evolve gradually during imaging, their entangled manifestations complicate the analysis of dynamic processes, particularly in beam-sensitive materials. Disentangling these two mechanisms is essential for reliable interpretation of material behavior under electron irradiation [20].

As beam-induced damage is readily quantifiable in the absence of motion, correcting for drift becomes the primary challenge in enabling such direct assessment. Traditionally, the field has addressed drift using either hardware-level solutions, such as revolving scan paths[26] or spiral scanning [27], or post-acquisition correction algorithms based on scan orientation [20]. These methods improve alignment and image quality but do not provide interpretable or predictive models of atomic motion itself. More recently, automated tracking systems[8] and deep learning–based image restoration models[9, 14] have been applied to high-resolution STEM data, achieving impressive results in noise reduction and scan artifact correction. However, the drift effects addressed in these approaches are more akin to scan artifacts or optical distortions within a single frame, rather than true atomic motion across frames.

Unlike scan distortion or minor drift commonly addressed in single-frame correction methods, the geometric evolution in our STEM sequences reflects sample-driven atomic motion accumulated over long exposures [36]. These sequences are temporally sparse, with frames acquired 1.6–2.0 seconds apart, which is sufficient for beam-induced forces to cause substantial and spatially nonuniform displacements, often exceeding tens or hundreds of pixels. Moreover, because damage and drift co-occur, image features may vanish, deform, or shift in ways that are physically irreversible. This poses a fundamental modeling challenge: there is no stable reference frame, no consistent pixel-to-pixel correspondence, and no purely geometric transformation that can recover the original signal.

To address these challenges, we present *AtomDiffuser*, a diffusion-variant spatiotemporal degradation model that disentangles atomic drift and beam-induced damage from sequential STEM image pairs. As shown in Figure 1, given a reference frame, a target frame, and a time-step signal, the model predicts a global affine transformation for drift and an atomic-level attenuation map for beam damage. AtomDiffuser is trained on synthetic sequences that simulate time-dependent physical degradation, enabling it to learn the spatiotemporal dynamics of atomic evolution via a time-conditioned U-Net. In evaluations of damage assessment, AtomDiffuser significantly outperforms the existing CNN method [36]. Applied to synthetic and cryo-STEM data of LLZO, it enables estimation of atomic drift and prediction of spatially resolved, interpretable beam-induced signal degradation.

Our main contributions are as follows:

- We introduce a time-conditioned lightweight deep learning model based on U-Net, together with a synthetic data generator that simulates realistic drift, stochastic decay, and acquisition noise during STEM imaging, reflecting key sources of temporal degradation that limit the interpretability of atomic-resolution experiments on beam-sensitive materials.
- We propose a degradation learning framework that formulates inter-frame atomic motion as a global affine transformation, enabling explicit and interpretable drift estimation across arbitrary time intervals, beyond conventional flow or registration methods, which is critical for compensating specimen instabilities in long-exposure or time-resolved microscopy studies.
- We model beam-induced degradation as a differentiable, pixel-wise attenuation process, allowing fine-grained prediction of atomic damage aligned with physical decay mechanisms, providing a step toward spatially resolved damage mapping that could inform dose management and acquisition strategies in in-situ STEM experiments.

2. Related Work

2.1. Atomic Feature Extraction from STEM Images

Early attempts at atom localization were severely affected by drift distortion, which was addressed through RevSTEM [26], a method that estimates and corrects sample drift using rotational scan series without requiring prior structural knowledge. Subsequent classical pipelines incorporated a model-free correction method based on orthogonal scan pairs to address nonlinear scan distortions[20]. Meanwhile, a specialized scan strategy is applied to reduce distortion, such as the spiral scanning approach, which mitigates flyback and nonlinear artifacts by replacing raster trajectories with smooth, continuous spiral paths [27]. To enable high-precision quantification of atomic fluctuations from time-resolved TEM videos, an automated image processing pipeline was developed that combines template matching, subpixel centroid localization, and geometric modeling to track atom positions and structural evolution in catalyst

nanoparticles [8]. With the advent of deep learning, convolutional networks have become the de facto tools for predicting pixel-wise atomic column probability maps under low-dose conditions [13]. These models substantially improve the recall of weak-contrast atoms and support downstream tasks such as lattice reconstruction, defect counting, and strain mapping [4]. Nevertheless, most existing networks assume static scenes. When consecutive frames are compared, the entangled effects of beam-induced attenuation and specimen drift are typically addressed through post-hoc intensity differencing. As a result, they cannot produce interpretable, parameterized estimates of the two degradation mechanisms that drive structural evolution.

Recent efforts have also addressed atomic-level distortion using statistical modeling techniques, such as Gaussian modeling[37] and Gaussian process regression [23]. Deep learning has been applied to suppress scan-induced artifacts [9, 14], reconstruct atomic structures from noisy observations [7, 11], and track atom centers with high precision. Industrial-ready toolkits such as AtomAI and AtomSegNet have further enabled automated atomic structure reconstruction and large-scale analysis of STEM data [13, 40]. Beyond reconstruction, deep models have been used to simulate complex physical processes, including graphene healing dynamics [6], defect detection in van der Waals materials [31], and recovery of molecular 3D motion from electron microscopy videos [33]. To the best of our knowledge, no prior work explicitly models atomic drift and attenuation as coupled, learnable degradation processes across frames.

2.2. Diffusion Modeling

Denoising diffusion probabilistic models (DDPMs) have recently emerged as the dominant paradigm for high-quality image synthesis across a wide range of computer vision tasks [24]. Compared to classical Generative Adversarial Networks (GANs), diffusion models offer greater stability and flexibility, particularly in conditional generation settings where guidance signals such as text prompts or semantic maps can be incorporated [3, 19]. These capabilities have enabled applications in digital art [34], content creation [18, 38], and industrial data augmentation [29], where semantic control and visual fidelity are critical.

In the materials science domain, the EMDiffuse family [15] first demonstrated that fine-tuning a generic DDPM on as little as a single noisy–clean image pair can outperform CNN-based denoisers on cryo-EM and cryo-STEM data, while additionally providing per-pixel uncertainty estimates. Its super-resolution variant, EMDiffuse-r [16], synthesizes high-resolution 3D volumes from anisotropic stacks, enabling isotropic reconstruction without hardware modification. More recently, SARDiffuse [22] introduced a diffusion-based approach for correcting lens-aberration blur in uncorrected STEM images, achieving routine sub-Ångström resolution and accurate atomic localization. Other diffusion models have been applied to reconstruct crystal structures from noisy data [21] and to simulate irradiation-induced defect cascades in solids [12], further highlighting the versatility of probabilistic diffusion frameworks for materials research.

Despite the superior performance of DDPM in physics modeling of atomic images, current diffusion models for EM/STEM heavily focus on correcting atom distortion rather than inter-frame drift [2], and prefer to analyze damage only on single images instead of temporally. Works that can observe material evolution over time while explicitly disentangle geometric drift from radiometric attenuation are still rare.

3. Methodology

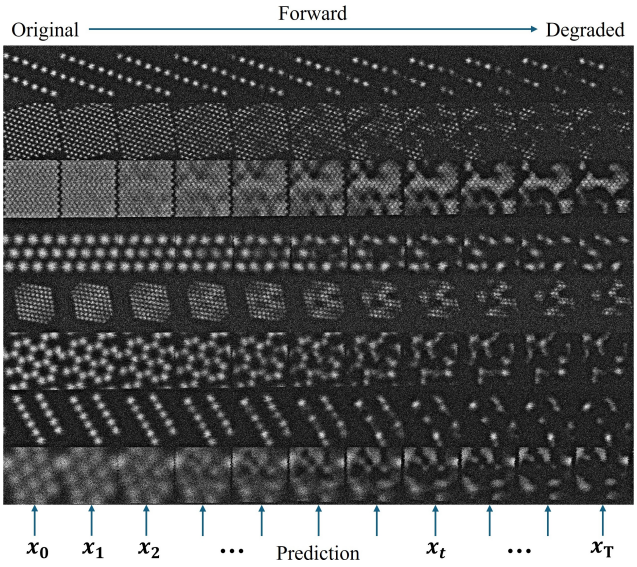

Figure 2. Degradation modeling. From left to right, the STEM samples are degraded over time.

3.1. Degradation Modeling

End-to-end prediction of atomic changes in STEM sequences, such as directly regressing x_t from x_0, is often unreliable due to the presence of noise, sparsity, and large-scale rigid motion [35]. To address this, we adopt a degradation modeling approach inspired by diffusion models, enabling continuous and physically interpretable modeling of structural evolution.

While classical diffusion models have shown strong performance in generative and denoising tasks by learning to reverse a stochastic corruption process [16], they are less suited for modeling physically driven phenomena such as drift or beam damage. In DDPM, for example, the forward diffusion is defined by progressively adding Gaussian noise

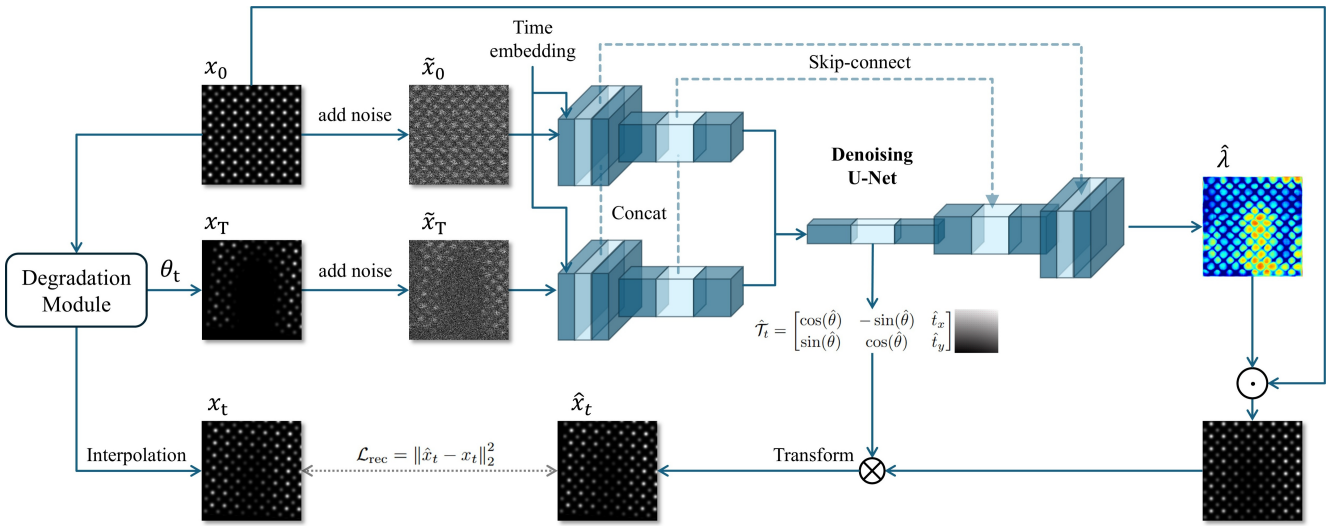

Figure 3. Framework of the proposed AtomDiffuser.

to x_0:

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t}\, \epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}), \tag{1}$$

where $\bar{\alpha}_t$ denotes the cumulative product of the noise schedule, and the noise factor ϵ is sampled from a standard Gaussian distribution $\mathcal{N}$. However, this formulation is limited in its ability to incorporate interpretable, structured degradation signals such as rigid motion, as DDPMs are primarily designed for noise removal rather than physical factor estimation.

In contrast, deterministic degradation models, such as Cold Diffusion [1], replace the stochastic noise process with a physically interpretable degradation process:

$$x_t = \mathcal{D}(x_0, t), \tag{2}$$

where $\mathcal{D}$ may represent operations such as blur, masking, or spatial warping, as shown in Figure 2. This formulation preserves temporal consistency and structural semantics while aligning more naturally with real-world degradation patterns, allowing the degraded state x_t to be computed directly from any time step t, which is consistent with the classical DDPM.

Inspired by this degradation modeling, we define a physically structured degradation function that models the evolution of STEM images over time as a combination of atomic drift and signal attenuation:

$$x_t = \mathcal{D}(x_0, t) = \Phi(\lambda_t \odot x_0;\ \mathcal{T}_t), \tag{3}$$

where Φ applies a spatial transformation ($\mathcal{T}_t \in \mathbb{R}^{2\times 3}$) to account for atomic drift that maps from x_0 to x_t, and $\lambda_t \in [0,1]^{H\times W}$ denote a pixel-wise attenuation map to model beam-induced signal loss, where each value can be interpreted as the relative survival coefficient of the corresponding atomic column over time under cumulative electron exposure. Given a clean reference image x_0, our goal is to generate the degraded state x_t, following the theory from the Cold-Diffusion, as illustrated in Figure 1.

3.2. Restoration

Unlike Cold-Diffusion models, our objective is not to reconstruct the clean image x_0, but rather to recover the latent physical factors, specifically, atomic drift and signal decay, that cause x_0 to degrade into x_t:

$$\hat{\lambda}_t, (\hat{\theta}, \hat{t}_x, \hat{t}_y) = \{(x_0, t;\ x_T), \tag{4}$$

where $(\hat{\theta}, \hat{t}_x, \hat{t}_y)$ denote the predicted parameters of in-plane rotation and translation, and $\lambda \in [0,1]^{H\times W}$ denote the predicted decay map.

This fundamental distinction motivates a different conditioning strategy: we use x_T as the structural prompt, and supervise the prediction of intermediate degradation states x_t by learning to infer the parameters that explain the evolution from x_0 to x_t.

Given the intermediate parameters $(\hat{\lambda}_t, \hat{\mathcal{T}}_t)$, we are able to reconstruct the degraded state x_t by applying two successive transformations:

$$x'_t = \hat{\lambda}_t \odot x_0 \tag{5}$$

$$\hat{x}_t = \Phi(x'_t;\ \hat{\mathcal{T}}_t), \tag{6}$$

where first, a spatially varying attenuation map $\hat{\lambda}_t$ to model signal decay, and second, an affine transformation $\hat{\mathcal{T}}_t \in \mathbb{R}^{2\times 3}$ to model global drift, where

$$\hat{\mathcal{T}}_t = \begin{bmatrix} \cos(\hat{\theta}) & -\sin(\hat{\theta}) & \hat{t}_x \\ \sin(\hat{\theta}) & \cos(\hat{\theta}) & \hat{t}_y \end{bmatrix}. \tag{7}$$

This decay-before-drift ordering respects the physical chronology of STEM degradation, where beam-induced damage occurs during exposure, and slow-stage drift affects the entire frame afterward, as shown in Figure 3.

3.3. Training Objective

To generate intermediate training samples, we design both the decay map and the drift transformation to be temporally inferred, as shown in Figure 4. We start by creating a final decay field λ_T using non-linear (Perlin) noise, and compute the intermediate decay as $\lambda_t = \lambda_T \cdot \left(\frac{t}{T}\right)$. The affine drift matrix is similarly defined by interpolating from a final transformation $\mathcal{T}_T$, giving $\mathcal{T}_t = \mathcal{T}_T \cdot \left(\frac{t}{T}\right)$. This setup allows us to simulate degradation at any time step t, with in-batch variation that improves training stability and reduces overfitting.

To better reflect real-world STEM conditions, we add synthetic noise after applying decay and drift, including Poisson noise, scan-line jitter, and readout noise:

$$\tilde{x} = x + \epsilon_{\text{noise}}. \tag{8}$$

Given noisy inputs $(\tilde{x}_0, \tilde{x}_T)$ and a time index t, the model is trained to predict the decay map $\hat{\lambda}_t$ and affine transformation $\hat{\mathcal{T}}_t$, which are then used to generate the intermediate prediction $\hat{x}_t$ (see Equation 5).

The training objective is a simple L2 reconstruction loss between the predicted and ground-truth degraded frames:

$$\mathcal{L}_{\text{rec}} = \|\hat{x}_t - x_t\|_2^2 . \tag{9}$$

This setup enables dense supervision and disentangled learning of physical degradation factors, while supporting interpolation or extrapolation across arbitrary time steps t. It is well-suited for temporally-aware applications such as future frame prediction or real-time correction in sparse STEM sequences. The overall training pipeline is illustrated in Figure 4.

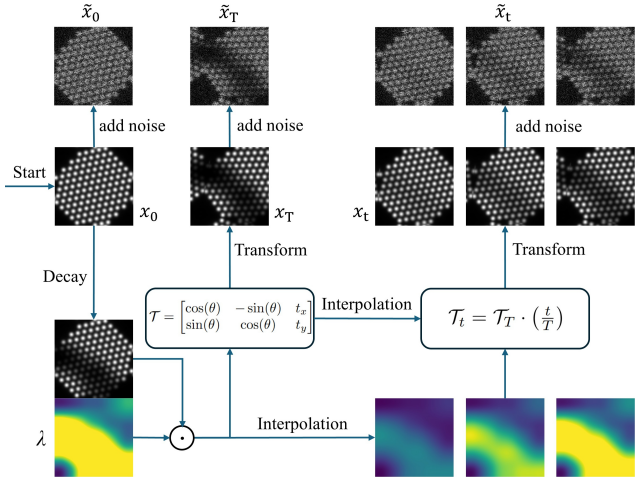

Figure 4. Training process. For each iteration, we start with a random atom map x_0, then use the generated decay map λ and affine matrix $\mathcal{T}$ to obtain the final state x_T. Finally, obtain x_t the interpolation functions.

3.4. Network Architecture and Training Strategy

The network of AtomDiffuser adopts a dual-stream U-Net backbone, with shared encoders processing the clean input x_0 and its degraded counterpart x_T to extract comparable multi-scale features. At the bottleneck, features are fused and modulated by a time embedding $\gamma(t)$ to enable temporal conditioning. A fully connected head predicts the affine transformation $\hat{\mathcal{T}}_t$, while the decoder uses dual skip connections to reconstruct a per-pixel decay map $\hat{\lambda}_t \in [0,1]^{H\times W}$. This architecture integrates spatial structure, temporal context, and global motion into a unified and interpretable framework, with outputs aligned to the input resolution for high-fidelity reconstruction.

Inspired by the data generation strategy in TEMImageNet [13], we sample raw atom maps (no-background-noise version) from the TEMImageNet dataset (14,364 samples in total) and apply our proposed degradation process dynamically during training to avoid duplication. For each sample, the clean image is used as x_0, and a degraded version is generated as x_T. The intermediate frame x_t is synthesized via temporal interpolation, as described in Section 3.3 and illustrated in Figure 2. Our network is implemented in PyTorch and trained on a single NVIDIA RTX 3090 GPU. We use a batch size of 32 and input resolution of $1 \times 256 \times 256$. The model is trained for 200 epochs using the AdamW optimizer with an initial learning rate of 0.0001, decayed progressively with a cosine scheduler to support fine-tuning during the later training stages [30, 39]. Training completes in 5.83 hours under our setup.

4. Results

4.1. Atomic Damage Assessment

To evaluate the model's ability to estimate beam-induced signal degradation independently of spatial drift, we perform a controlled experiment using a fixed synthetic atomic image. We generate degradation-only sequences by applying progressive attenuation while introducing one of three noise types: (1) Gaussian noise with structured black hole artifacts [36], (2) low-frequency Perlin noise [28], and (3) random noise mimicking scan jitter and long-range distortions, as illustrated in Figure 5. Each sequence consists of ten frames with linearly increasing damage intensity, ranging from 0 (undamaged) to 0.9 (severely degraded), and no geometric transformation is applied. This setting isolates the signal decay process, allowing us to assess the model's capacity to infer spatially varying attenuation under different degradation conditions, as visualized in Figure 5(e)–(g).

Formally, given a predicted decay map $\hat{\lambda}_t$, we define the scalar damage intensity as

$$\bar{\Delta}_t = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} \left(1 - \hat{\lambda}_t(i,j)\right), \tag{10}$$

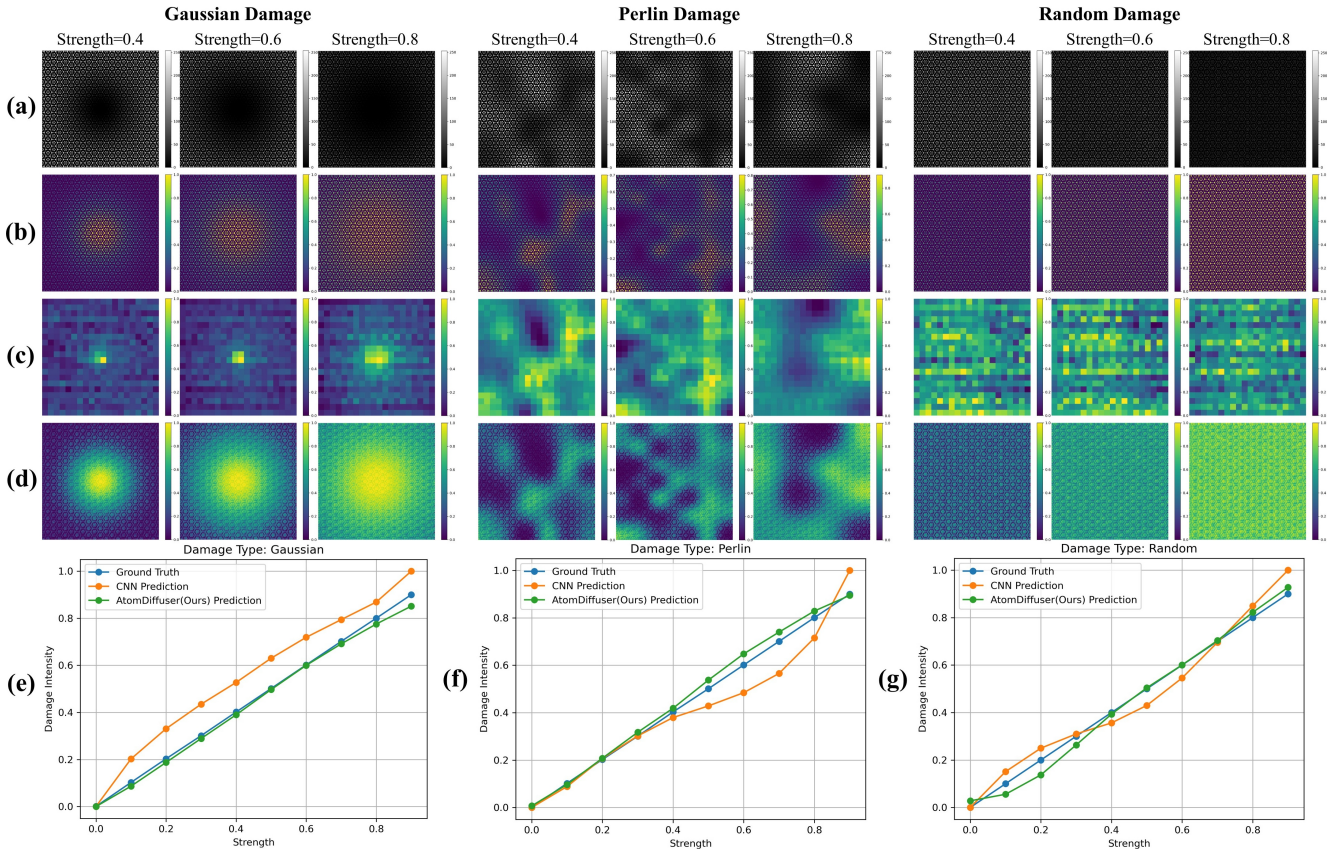

Figure 5. Damage Assessment. (a) are the ground-truth samples, (b) are the ground-truth damage maps, (c) are the predicted damage maps with CNN method [36], (d) are the predicted damage maps with the propsoed method, and (e)(f)(g) are the plotted damage intensity curves with different damage types: Gaussian black hole, Perlin non-linear masks, and Random noise.

which quantifies the overall proportion of signal loss at time step t. The predicted intensity $\bar{\Delta}_t$ is then compared against the ground-truth value Δ_t across the full temporal sequence. We compare our model against a CNN-based baseline by plotting the predicted damage intensity curves alongside the ground-truth trajectory, as shown in Figure 5(e)–(g). Evaluation metrics include mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R^2). Additionally, we report the variance of the prediction error (Var.), computed as the sample variance of the difference between predicted and ground-truth values over time.

To assess the spatial quality of the predicted decay maps, we conduct side-by-side comparisons with a CNN-based baseline [36]. While the CNN performs well under the Perlin setting, our model produces full-resolution predictions aligned with the input image dimensions, enabling fine-grained localization of damage regions, as shown in Figure 5(d). In the Gaussian and Random settings, our method yields more coherent attenuation boundaries and better matches the ground-truth degradation patterns. In contrast, the CNN baseline struggles to generalize under irregular or spatially heterogeneous noise, likely due to resolution loss from downsampling and limited robustness to complex distortions.

	MAE	MSE	RMSE	R^2	Var.
Gaussian					
CNN	0.0994	0.0113	0.1065	0.8620	0.0177
Ours	**0.0147**	**0.0004**	**0.0198**	**0.9952**	**0.0002**
Perlin					
CNN	0.0573	0.0059	0.0767	0.9285	0.0061
Ours	**0.0193**	**0.0006**	**0.0239**	**0.9931**	**0.0003**
Random					
CNN	0.0635	0.0074	0.0859	0.9106	0.0025
Ours	**0.0236**	**0.0009**	**0.0306**	**0.9886**	**0.0004**

Table 1. Benchmark of Simulated Atomic Damage

To evaluate scalar performance, we test our model on a fixed synthetic sample with known degradation profiles. As shown in Table 1, our model achieves an MSE as low as 0.0004 and an R^2 of 0.9952. The predicted intensity closely follows the ground-truth linear trend, demonstrating high accuracy in estimating the degree of beam-induced damage (Figure 5(e)–(g)). In contrast, the CNN-based baseline [36], which treats the task as generic feature-to-feature regression from x_0 to x_t, exhibits greater variance in cumulative dam-

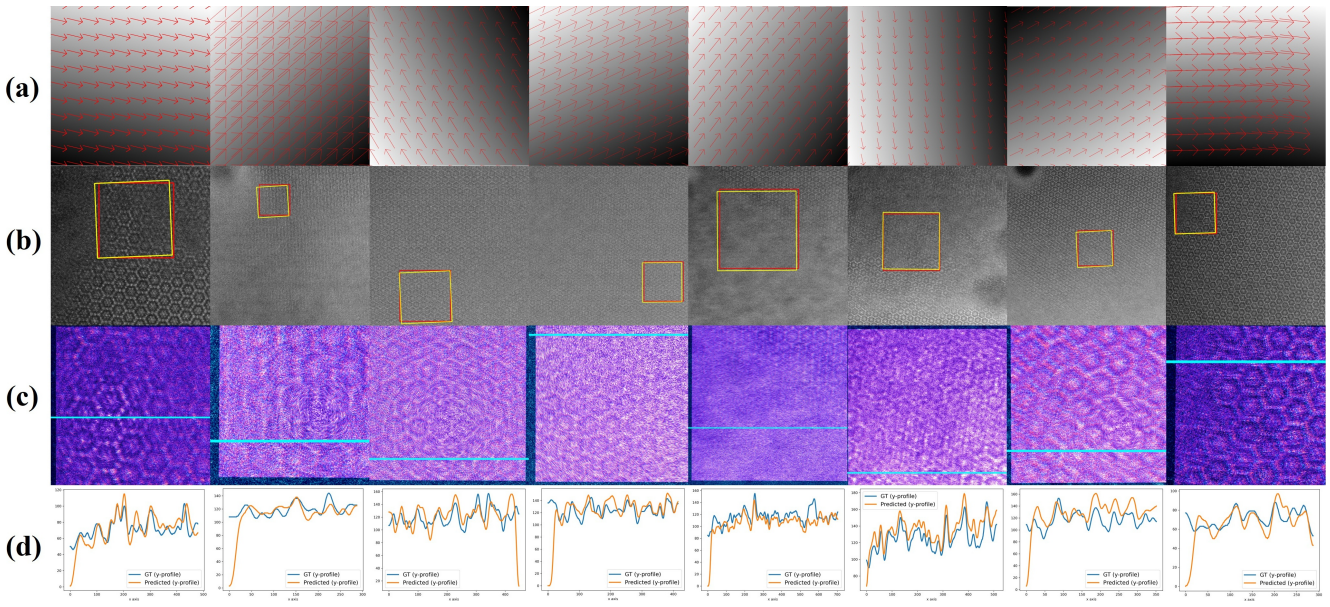

Figure 6. Drift test on STEM samples. (a) are the flow maps of the predicted motion directions, (b) are the STEM samples and its generated two frames (x_0 and x_T), (c) are the overlaps of predicted $\hat{x}_T$ and x_T, where horizontal line indicates the selected location for side-plot, and (d) are the plots of side-views of selected axis from (c).

age estimates. Notably, AtomDiffuser remains robust even under random noise, where the CNN baseline shows significant deviation from the ground-truth trajectory.

4.2. Sample Drift Correction

To quantitatively evaluate the model's ability to recover atomic drift, we conduct a controlled ablation experiment that isolates affine motion without introducing signal decay. Specifically, we disable attenuation ($\lambda_t = 1$ everywhere) and test the model's capacity to predict geometric transformations in real STEM images under synthetic drift conditions.

Given a real high-resolution STEM image x_0, we randomly extract subregions and apply a simulated affine transformation matrix $\mathcal{T} \in \mathbb{R}^{2\times3}$ to generate a corresponding drifted image x_T:

$$x_T = \Phi(x_0;\ \mathcal{T}), \tag{11}$$

where Φ denotes an image warping operator parameterized by the ground truth transformation $\mathcal{T}$.

For each test, we use our trained model to predict the drift matrix $\hat{\mathcal{T}}$ that maps x_0 to x_T. The predicted matrix is then compared against the ground-truth drift matrix $\mathcal{T}$. We report two evaluation metrics:

Drift error: the sum of absolute pixel-wise translation errors:

$$\text{Drift Err.} = |\hat{t}_x - t_x| + |\hat{t}_y - t_y|, \tag{12}$$

where $(\hat{t}_x, \hat{t}_y)$ are the translation components of the predicted matrix $\hat{\mathcal{T}}$, and (t_x, t_y) are the corresponding components of the ground-truth affine matrix $\mathcal{T}$.

Rotation error: the absolute angular difference (in degrees) between the predicted and ground-truth rotation angles:

$$\text{Rotation Err.} = |\hat{\theta} - \theta|, \tag{13}$$

where $\hat{\theta}$ and θ denote the rotation components of $\hat{\mathcal{T}}$ and $\mathcal{T}$, respectively.

Unlike the synthetic degradation setting used during training, this benchmark is constructed entirely from real STEM samples, with single-frame drift derived from geometric analysis. To visualize the performance, we sampled several test results from each STEM image and plotted detailed comparisons. As shown in Figure 6, the image pairs (x_0, x_T) are extracted from the original STEM images, and the proposed AtomDiffuser is used to predict the transformation between x_0 and x_T. The warped version of x_T is then overlaid onto x_0 (Figure 6(c)).

To further assess alignment quality, we visualize a side-profile view (Figure 6(d)). Specifically, we select a random location from the overlapped region in Figure 6(c) and plot the pixel intensity values along a chosen axis. A 1D Gaussian filter is applied to smooth the profiles for improved visualization and interpretability. As shown in Figure 6(d), the drift compensation results in well-aligned intensity peaks between the two signals. In other words, the closer the curves align, the better the model performs in estimating and correcting atomic drift.

As shown in Table 2, the proposed AtomDiffuser successfully estimates affine transformations across a range of synthetic drift and rotation settings applied to real data. The

Rotation Setting (°)	Drift Setting (px)	Drift Err.	Rotation Err.
5	5	2.4271	0.6032
10	10	3.0344	0.9085
15	15	3.5103	1.7986
15	25	4.7775	1.5405
15	50	6.4971	1.3608

Table 2. Drifting Benchmark for STEM Samples

model demonstrates high robustness under moderate motion (e.g., 15 px), achieving low drift errors (2.4–3.5 px) and accurate rotation recovery (below 2°). Importantly, the model maintains stable performance even under asymmetric or unbalanced configurations, such as 15° rotation with 25 px or 50 px drift. In these harder cases, the drift error increases only modestly (to 4.8 and 6.5 px respectively), while rotation accuracy remains largely unaffected. This suggests that the model captures the underlying spatial dynamics without overfitting to balanced motion scales, and generalizes well beyond the training distribution.

4.3. Degradation Inference & Performance

To demonstrate the full potential of AtomDiffuser—the ability to continuously infer between frames and explore material characteristics by observing beam damage progression without the interference of drift, we apply the model to two consecutive frames selected from a high-resolution cryo-STEM dataset of the garnet-type solid electrolyte $\mathbf{Li_7La_3Zr_2O_{12}}$ (LLZO) [36]. Specifically, given two frames x_0 and x_T, the model generates intermediate degradation states, including motion $\mathcal{T}_t$ and decay factor λ_t, by varying the time input $t \in (0, T)$. Notably, the target frame x_T exhibits both atomic drift and beam-induced damage relative to x_0, indicating the two frames are both misaligned and structurally dissimilar.

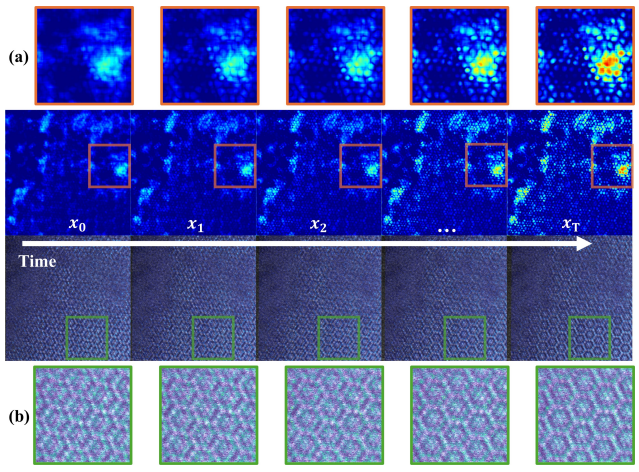

Figure 7. Degradation inference test. For each iteration, we apply x_0, x_T, and t as inputs, then increase t to generate an x_t, subsequently obtained sequences.

As illustrated in Figure 7(a), the inferred frames exhibit plausible decay progression with smooth transitions in both spatial and intensity domains. Meanwhile, the two frames become gradually aligned as time approaches T, suggesting a recovery of subpixel displacements and damage-aware registration, as shown in Figure 7(b). Overall, the results produced by our method reveal a smooth and interpretable trajectory of structural change that has rarely been demonstrated in existing deep learning approaches.

More importantly, despite the training process being relatively complex and resource-intensive, the model remains lightweight and portable. As shown in Table 3, the input size is flexible and supports resolutions up to 2048×2048 with minimum inference time. This makes the model significantly more practical for deployment across various experimental conditions.

	Inference Time (per sample)	Params (M)	Input Size	Size on disk
CNN Method	1.017	11.69	224×224	44.59 MB
AtomDiffuser (Ours)	**0.002**	**2.51**	**2048×2048**	**9.98** MB

Table 3. Comparison of Model Parameters

5. Conclusion

Inspired by recent advances in degradation modeling, AtomDiffuser rethinks atomic-scale imaging by modeling degradation as a structured, time-evolving physical process. By disentangling drift and beam-induced damage through a unified, interpretable framework, it enables the observation of structural evolution across sparse STEM sequences. Evaluated on both synthetic and cryo-STEM data, AtomDiffuser demonstrates robust performance in recovering atomic motion and damage localization. Beyond its performance, this work introduces a new paradigm for degradation modeling in micro-level imaging by leveraging physical priors as conditional signals, which opens new possibilities for studying progressive material evolution in a frame-by-frame technique. We envision that this capability will support future investigations into time-dependent beam effects, enable virtual time-lapse reconstruction from sparse acquisitions, and provide new insights into dynamic material behavior at the atomic scale.

Acknowledgments

The authors acknowledge the support by the National Science Foundation under Award No. 2239598, as well as the use of facilities and instrumentation at the University of California Irvine Materials Research Institute, which is supported in part by the National Science Foundation Materials Research Science and Engineering Center program through the University of California Irvine Center for Complex and Active Materials (DMR-2011967).

References

[1] Arpit Bansal, Eitan Borgnia, Hong-Min Chu, Jie Li, Hamid Kazemi, Furong Huang, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Cold diffusion: Inverting arbitrary image transforms without noise. *Advances in Neural Information Processing Systems*, 36:41259–41282, 2023. 1, 4

[2] Christian K Belardi, Chia-Hao Lee, Yingheng Wang, Justin Lovelace, Kilian Q Weinberger, David A Muller, and Carla P Gomes. Improving multislice electron ptychography with a generative prior. *arXiv preprint arXiv:2507.17800*, 2025. 3

[3] Jooyoung Choi, Yunjey Choi, Yunji Kim, Junho Kim, and Sungroh Yoon. Custom-edit: Text-guided image editing with customized diffusion models. *arXiv preprint arXiv:2305.15779*, 2023. 3

[4] Bappaditya Dey, Vic De Ridder, Victor Blanco, Sandip Halder, and Bartel Van Waeyenberge. Addressing class imbalance and data limitations in advanced node semiconductor defect inspection: A generative approach for sem images. In *2024 International Symposium ELMAR*, pages 141–148. IEEE, 2024. 3

[5] RF Egerton, P Li, and M Malac. Radiation damage in the tem and sem. *Micron*, 35(6):399–409, 2004. 2

[6] Ayana Ghosh, Maxim Ziatdinov, Ondrej Dyck, Bobby G Sumpter, and Sergei V Kalinin. Bridging microscopy with molecular dynamics and quantum simulations: an atomai based pipeline. *npj Computational Materials*, 8(1):74, 2022. 1, 3

[7] Bintao He, Fa Zhang, Huanshui Zhang, and Renmin Han. A hybrid frequency-spatial domain model for sparse image reconstruction in scanning transmission electron microscopy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2682–2691, 2021. 3

[8] Zahra Hussaini, Pin Ann Lin, Bharath Natarajan, Wenhui Zhu, and Renu Sharma. Determination of atomic positions from time resolved high resolution transmission electron microscopy images. *Ultramicroscopy*, 186:139–145, 2018. 2, 3

[9] Shiro Ihara, Hikaru Saito, Mizumo Yoshinaga, Lavakumar Avala, and Mitsuhiro Murayama. Deep learning-based noise filtering toward millisecond order imaging by using scanning transmission electron microscopy. *Scientific reports*, 12(1): 13462, 2022. 2, 3

[10] Sergei V Kalinin, Maxim Ziatdinov, Jacob Hinkle, Stephen Jesse, Ayana Ghosh, Kyle P Kelley, Andrew R Lupini, Bobby G Sumpter, and Rama K Vasudevan. Automated and autonomous experiments in electron and scanning probe microscopy. *ACS nano*, 15(8):12604–12627, 2021. 2

[11] Chia-Hao Lee, Abid Khan, Bryan Clark, and Pinshane Y Huang. Using cyclegans to generate realistic stem images for machine learning and atom-by-atom analysis on the million-atom scale. *Microscopy and Microanalysis*, 31 (Supplement_1):ozaf048–1074, 2025. 3

[12] Ruihao Liao, Ke Xu, Yifan Liu, Zibo Gao, Shuo Jin, Linyun Liang, and Guang-Hong Lu. Fast prediction of irradiation-induced cascade defects using denoising diffusion probabilistic model. *Nuclear Materials and Energy*, 41:101805, 2024. 1, 3

[13] Ruoqian Lin, Rui Zhang, Chunyang Wang, Xiao-Qing Yang, and Huolin L Xin. Temimagenet training library and atomsegnet deep-learning models for high-precision atom segmentation, localization, denoising, and deblurring of atomic-resolution images. *Scientific reports*, 11(1):5386, 2021. 3, 5

[14] I Lobato, T Friedrich, and S Van Aert. Deep convolutional neural networks to restore single-shot electron microscopy images. *npj Computational Materials*, 10(1):10, 2024. 2, 3

[15] Chixiang Lu, Kai Chen, Heng Qiu, Xiaojun Chen, Gu Chen, Xiaojuan Qi, and Haibo Jiang. Emdiffuse: a diffusion-based deep learning method augmenting ultrastructural imaging and volume electron microscopy. *bioRxiv*, pages 2023–07, 2023. 3

[16] Chixiang Lu, Kai Chen, Heng Qiu, Xiaojun Chen, Gu Chen, Xiaojuan Qi, and Haibo Jiang. Diffusion-based deep learning method for augmenting ultrastructural imaging and volume electron microscopy. *Nature Communications*, 15(1): 4677, 2024. 3

[17] Nabila Masud, Jaydeep Rade, Md Hasibul Hasan Hasib, Adarsh Krishnamurthy, and Anwesha Sarkar. Machine learning approaches for improving atomic force microscopy instrumentation and data analytics. *Frontiers in Physics*, 12: 1347648, 2024. 1

[18] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*, 2021. 3

[19] Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Null-text inversion for editing real images using guided diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6038–6047, 2023. 3

[20] Colin Ophus, Jim Ciston, and Chris T Nelson. Correcting nonlinear drift distortion of scanning probe and scanning transmission electron microscopies from image pairs with orthogonal scan directions. *Ultramicroscopy*, 162:1–9, 2016. 2

[21] Teerachote Pakornchote, Natthaphon Choomphon-Anomakhun, Sorrjit Arrerut, Chayanon Atthapak, Sakarn Khamkaeo, Thiparat Chotibut, and Thiti Bovornratanaraks. Diffusion probabilistic models enhance variational autoencoder for crystal structure generative modeling. *Scientific Reports*, 14(1):1275, 2024. 3

[22] Zanlin Qiu, Yuan Meng, Junxian Li, Yanhui Hong, Ning Li, Xiaocang Han, Yu Liang, Wing Ni Cheng, Guolin Ke, Linfeng Zhang, et al. Deep learning for sub-ångström resolution imaging in uncorrected scanning transmission electron microscope. *National Science Review*, page nwaf235, 2025. 3

[23] Kevin M Roccapriore, Nicole Creange, Maxim Ziatdinov, and Sergei V Kalinin. Identification and correction of temporal and spatial distortions in scanning transmission electron microscopy. *Ultramicroscopy*, 229:113337, 2021. 3

[24] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 1, 3

[25] Alireza Sadri, Timothy C Petersen, Emmanuel WC Terzoudis-Lumsden, Bryan D Esser, Joanne Etheridge, and Scott D Findlay. Unsupervised deep denoising for four-dimensional scanning transmission electron microscopy. *npj Computational Materials*, 10(1):243, 2024. 1

[26] Xiahan Sang and James M LeBeau. Revolving scanning transmission electron microscopy: Correcting sample drift distortion without prior knowledge. *Ultramicroscopy*, 138: 28–35, 2014. 2

[27] Xiahan Sang, Andrew R Lupini, Jilai Ding, Sergei V Kalinin, Stephen Jesse, and Raymond R Unocic. Precision controlled atomic resolution scanning transmission electron microscopy using spiral scan pathways. *Scientific Reports*, 7 (1):43585, 2017. 2

[28] Hao Wang, Sayed Pedram Haeri Boroujeni, Xiwen Chen, Ashish Bastola, Huayu Li, Wenhui Zhu, and Abolfazl Razi. Flame diffuser: Wildfire image synthesis using mask guided diffusion. In *2024 IEEE International Conference on Big Data (BigData)*, pages 6171–6179. IEEE, 2024. 5

[29] Hao Wang, Xiwen Chen, Ashish Bastola, Jiayou Qin, and Abolfazl Razi. Diffusion prism: Enhancing diversity and morphology consistency in mask-to-image diffusion. In *Proceedings of the Winter Conference on Applications of Computer Vision (WACV) Workshops*, pages 228–237, 2025. 3

[30] Hao Wang, Wenhui Zhu, Xuanzhao Dong, Yanxi Chen, Xin Li, Peijie Qiu, Xiwen Chen, Vamsi Krishna Vasa, Yujian Xiong, Oana M Dumitrascu, et al. Many-mobilenet: Multi-model augmentation for robust retinal disease classification. In *MICCAI Challenge on Ultra-Widefield Fundus Imaging for Diabetic Retinopathy*, pages 144–154. Springer, 2025. 5

[31] Mads Weile, Sergii Grytsiuk, Aubrey Penn, Daniel G Chica, Xavier Roy, Kseniia Mosina, Zdenek Sofer, Jakob Schiøtz, Stig Helveg, Malte Rösner, et al. Defect complexes in crsbr revealed through electron microscopy and deep learning. *Physical Review X*, 15(2):021080, 2025. 1, 3

[32] Rui Xiao, Yanzhu Zhang, and Mi Li. Automated high-throughput atomic force microscopy single-cell nanomechanical assay enabled by deep learning-based optical image recognition. *Nano Letters*, 24(39):12323–12332, 2024. 1

[33] Enze Ye, Yuhang Wang, Hong Zhang, Yiqin Gao, Huan Wang, and He Sun. Recovering a molecule's 3d dynamics from liquid-phase electron microscopy movies. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10767–10777, 2023. 3

[34] Xinxi Zhang, Song Wen, Ligong Han, Felix Juefei-Xu, Akash Srivastava, Junzhou Huang, Vladimir Pavlovic, Hao Wang, Molei Tao, and Dimitris Metaxas. Soda: Spectral orthogonal decomposition adaptation for diffusion models. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 4665–4682. IEEE, 2025. 3

[35] Hongkui Zheng, Xiner Lu, and Kai He. In situ transmission electron microscopy and artificial intelligence enabled data analytics for energy materials. *Journal of Energy Chemistry*, 68:454–493, 2022. 1, 3

[36] Hongkui Zheng, Xiwen Chen, Abolfazl Razi, and Kai He. Convolutional neural networks for evaluation of sequential beam damage of beam-sensitive solid electrolytes, 2024. 1, 2, 5, 6, 8

[37] Hongkui Zheng, Hao Wang, Xiwen Chen, Abolfazl Razi, and Kai He. Gaussian differential assessment of sequential stem radiation damage in beam-sensitive materials. *Microscopy and Microanalysis*, 31(Supplement_1):ozaf048–1065, 2025. 3

[38] Linqi Zhou, Aaron Lou, Samar Khanna, and Stefano Ermon. Denoising diffusion bridge models. *arXiv preprint arXiv:2309.16948*, 2023. 3

[39] Wenhui Zhu, Peijie Qiu, Xiwen Chen, Xin Li, Natasha Lepore, Oana M Dumitrascu, and Yalin Wang. nnmobilenet: Rethinking cnn for retinopathy research. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2285–2294, 2024. 5

[40] Maxim Ziatdinov, Ayana Ghosh, Chun Yin Wong, and Sergei V Kalinin. Atomai framework for deep learning analysis of image and spectroscopy data in electron and scanning probe microscopy. *Nature Machine Intelligence*, 4(12): 1101–1112, 2022. 3