

Score-Based Matching with Target Guidance for Cryo-EM Denoising

Xiaoqi Wu^{1†}, Xueying Zhan^{1†}, Wen Li¹, Junhao Wu², Xin Huang², Ke Ni^{3‡}, and Min Xu^{1*}

¹ Carnegie Mellon University, Pittsburgh PA 15213, USA
{xiaoqi, xueyingz, wenli3, mxu1}@andrew.cmu.edu

² Towson University, Towson MD 21252, USA
jwu17@students.towson.edu
xhuang@towson.edu

³ Independent Researcher
keniauthor@gmail.com

Abstract. Cryo-electron microscopy (cryo-EM) enables single-particle analysis of biological macromolecules under strict low-dose imaging conditions, but the resulting micrographs often exhibit extremely low signal-to-noise ratios and weak particle visibility. Image denoising is therefore an important preprocessing step for downstream cryo-EM analysis, including particle picking, 2D classification, and 3D reconstruction. Existing cryo-EM denoising methods are commonly trained with pixel-wise or Noise2Noise-style objectives, which can improve visual quality but do not explicitly account for structural consistency required by downstream analysis. In this work, we propose a score-based denoising framework for cryo-EM that learns the clean-data score to recover particle signals while better preserving structural information. Building on this formulation, we further introduce a target-guided variant that incorporates reference-density guidance to stabilize score learning under weak and ambiguous signal conditions. Rather than simply amplifying particle-like responses, our framework better suppresses structured low-frequency background, which improves particle-background separability for downstream analysis. Experiments on multiple cryo-EM datasets show that our score-based methods consistently improve downstream particle picking and produce more structure-consistent 3D reconstructions. Experiments on multiple cryo-EM datasets show that our methods improve downstream particle picking and produce more structure-consistent reconstructions. Our code is available at <https://github.com/xulabs/aitom>.

Keywords: Cryo-electron Microscopy · Image Denoising · Score-based Models

*Corresponding author.

†Xiaoqi Wu and Xueying Zhan contributed equally to this work.

‡Work was done prior to joining Amazon.

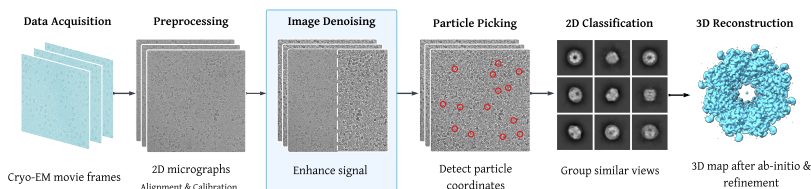


Fig. 1: Cryo-EM single-particle analysis pipeline, illustrating the main computational stages from raw cryo-EM micrographs to 3D reconstructed particle density map.

1 Introduction

Cryo-EM enables the reconstruction of 3D structures of biological macromolecules from large collections of noisy 2D projection images acquired under cryogenic conditions [13, 18, 19]. Due to strict low-dose imaging requirements, cryo-EM micrographs exhibit extremely low SNR and weak image contrast. In single-particle analysis (SPA), these micrographs are processed by a multi-stage pipeline including preprocessing, particle picking, 2D classification, and 3D reconstruction, where image-level errors propagate and affect downstream structural recovery [25]. Fig. 1 summarizes this workflow. Improving micrograph quality therefore remains critical for reliable cryo-EM analysis. Image denoising is particularly challenging because particle signals are weak, spatially small, and often indistinguishable from structured background noise, especially for small particles or low-contrast views where aggressive denoising may remove meaningful structural information. Unlike natural image restoration tasks, cryo-EM denoising lacks clean ground-truth supervision due to radiation damage constraints. Moreover, modern restoration methods increasingly rely on large-scale pretrained models and natural-image priors, which transfer poorly to cryo-EM data due to the substantial domain gap. Consequently, general-purpose vision denoisers often struggle to preserve weak particle structures under extreme noise conditions.

Early cryo-EM denoising mainly used classical signal-processing methods such as Gaussian filtering [9], but the gains are limited in the extreme low-SNR regime. Recent work has therefore shifted toward learning-based denoisers trained directly on cryo-EM data. Many existing methods adopt Noise2Noise (N2N)-style training, enabling self-supervised learning from paired noisy observations without clean ground truth [2, 15, 16]. While improving visual quality, these approaches rely on pixel-level regression objectives that do not explicitly enforce structural consistency, often leading to over-smoothing or structural distortion under weak particle signals. Methods like DRACO improve training stability through modified data splitting and masking strategies [22]. However, two challenges remain: regression objectives do not explicitly model noise statistics in low-SNR regimes, and improved denoising quality does not necessarily translate into better downstream reconstruction. This reveals a gap between image restoration objectives and structural recovery requirements.

Motivated by these challenges, we adopt denoising score matching (DSM) as the foundation of our approach, since score-based learning models image corruption through the underlying data distribution rather than direct pixel-wise regression, making it more suitable for preserving structural signals under heavy noise without clean supervision. We observe that the extremely low-SNR regime of cryo-EM is precisely the setting that score matching is designed to handle, yet this connection has not been explored in practice. However, under extremely low SNR, score estimation becomes ill-conditioned in the low-noise regime close to the clean signal, where weak structural cues provide limited optimization guidance. DSM alone therefore fails to recover reliable score directions. To address this issue, we then introduce target-guided score learning by incorporating prior particle-structure information in the form of a reference density map. The proposed guidance stabilizes score estimation near the clean signal manifold, providing directional cues under severe noise and improving preservation of fine particle details. To date, DSM has been applied almost exclusively to natural images, and target guidance has remained a largely theoretical construct with little demonstrated real-world application. Our contributions are summarized as follows:

- We identify cryo-EM micrograph denoising as a natural but previously unexplored application domain for score matching, and develop a target-guided score-based framework that addresses the ill-conditioning of score estimation under cryo-EM’s extreme low-SNR conditions without requiring clean ground-truth micrographs as supervision.
- We introduce similarity-guided annealing, an adaptive training scheme that balances target guidance and score matching to preserve particle structure without over-suppressing non-particle regions.
- Experiments across multiple cryo-EM datasets, including simulated benchmarks, show consistent gains in downstream tasks like particle picking and improved 3D reconstruction quality.

2 Related Work

Traditional Cryo-EM Denoising Methods. Early cryo-EM denoising methods relied on classical signal-processing techniques such as Gaussian and Wiener filtering [11]. They assume simplified noise models and handcrafted priors, which provide limited improvements under extremely low-SNR conditions and often fail to recover fine structural details required for downstream structural analysis.

Self-supervised Denoising with Noisy Observations. Deep learning methods have significantly advanced cryo-EM denoising. N2N [15] enables self-supervised learning from paired noisy observations without clean ground truth. Topaz-Denoise [2] adopts this paradigm using a U-Net architecture and demonstrates strong robustness across datasets. More recently, foundation-model-style pre-training has been explored for cryo-EM data, such as DRACO [22], which combines self-supervised pretraining with N2N-based learning. Despite their empirical success, these approaches rely on regression-based objectives that implicitly

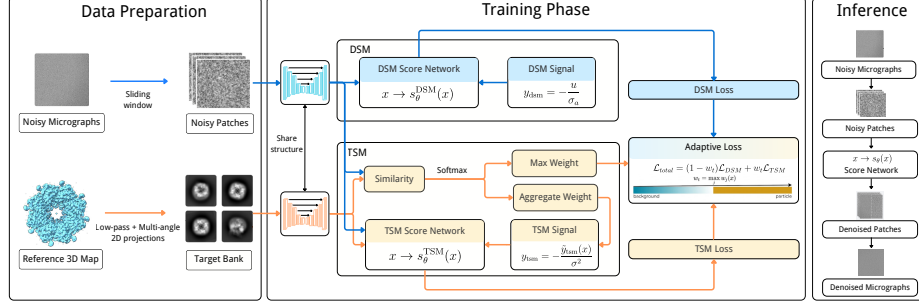


Fig. 2: Overview of the proposed target-guided score-based denoising framework.

learn signal statistics without explicitly enforcing structural detail preserving, which lead to structural distortion under weak particle signals.

Blind-spot and Pseudo-supervised Approaches. Blind-spot approaches such as Noise2Void [12] and Self2Self [21] enable denoising from single noisy observations by predicting masked pixels from local context. However, their locality assumption limits the modeling of global structural information, which is critical in cryo-EM images with weak and spatially correlated signals. Pseudo-supervised methods have also been explored; for example, NT2C [16] learns noise characteristics using simulated clean data. These synthetic supervisions introduce a statistical mismatch between simulated and experimental noise distributions, limiting generalization to real micrographs.

Score-based Models for Cryo-EM Denoising. Score-based generative models estimate data distributions through score learning [24]. DSM [23] provides a principled framework for modeling noise corruption without clean supervision. However, under the extremely low-SNR conditions of cryo-EM imaging, score estimation can become unstable near the clean signal regime, leading to loss of subtle structural details. Target Score Matching [4] extends DSM by incorporating task-preferred properties into score estimation. Our work builds on this line by introducing structure-guided score learning tailored to cryo-EM denoising.

3 Methodology

We first summarize the cryo-EM noise characteristics and then introduce our target-guided score-based denoising framework.

3.1 Cryo-EM Noise Pattern

Cryo-EM micrographs are acquired under dose-limited imaging, where the dominant noise sources are commonly categorized as: (i) *structural noise* (irreproducible background and conformational variability), (ii) *shot noise* (electron

counting statistics), and (iii) *detector/digitization noise* [1, 16]. We summarize these effects using an effective additive formulation:

$$X = Y + \varepsilon, \quad \varepsilon = \varepsilon_{\text{struct}} + \varepsilon_{\text{shot}} + \varepsilon_{\text{det}}. \quad (1)$$

Shot noise is signal-dependent, modeled as $X_i \sim \text{Poisson}(\alpha Y_i + b)$. Detector noise is typically approximated as zero-mean Gaussian. For tractability, we use a single isotropic Gaussian surrogate in both analysis and experiments.

3.2 Problem Definition

We study denoising of cryo-EM *micrographs*, i.e., large grayscale microscope images stored in scientific formats (e.g., MRC, TIFF). Since micrographs are typically several thousand pixels per side (e.g., 4096×4096), we perform learning on local patches. Let $x \in \mathbb{R}^{h \times w}$ be a noisy patch and $y \in \mathbb{R}^{h \times w}$ its (unobserved) clean counterpart. We vectorize patches and write $x, y \in \mathbb{R}^d$ with $d = h \cdot w$. In cryo-EM, clean ground truth y is generally unavailable.

Let $X \sim p_X$ and $Y \sim p_Y$ denote the noisy and clean patch-level random variables. Define the conditional score of the corruption model as $s_{X|Y}(x | y) := \nabla_x \log p_{X|Y}(x | y)$. The noisy score satisfies

$$\nabla_x \log p_X(x) = \int p_{Y|X}(y | x) s_{X|Y}(x | y) dy, \quad (2)$$

where $p_{Y|X}(y | x)$ denotes the posterior distribution of the latent clean patch given $X = x$. Intuitively, this identity expresses the score of the noisy-patch distribution p_X as a posterior expectation of the noise-model score under $p_{Y|X}$.

When the noise model is known and paired samples (X, Y) are available, the score $\nabla_x \log p_X(x)$ can be approximated by a score network $s_X^\theta(x)$ by minimizing

$$\ell(\theta) = \mathbb{E}_{X,Y} \|s_X^\theta(X) - s_{X|Y}(X | Y)\|^2, \quad (3)$$

which yields the DSM objective [24]. Even under additive Gaussian corruption $X = Y + N$ with $N \sim \mathcal{N}(0, \sigma_a^2 I)$, the score target becomes ill-conditioned as σ_a decreases. Specifically, $s_{X|Y}(x | y) = \nabla_x \log p_{X|Y}(x | y) = -(x - y)/\sigma_a^2$, so its scale grows as σ_a^{-2} . Consequently, the posterior variance satisfies

$$\sum_{i=1}^d \text{Var}_{Y|X}((s_{X|Y}(X | Y))_i) \sim d \sigma_a^{-2} \quad \text{as } \sigma_a \rightarrow 0, \quad (4)$$

leading to unstable score estimation in the low-noise regime relevant to cryo-EM. Although denoising should become easier as noise decreases, the DSM target becomes ill-conditioned as $\sigma_a \rightarrow 0$. This is critical for cryo-EM, where informative structures concentrate near the clean manifold; Fig. 2 summarizes our target-guided score-based denoising pipeline.

3.3 Optimization With Target Guidance

Motivated by the instability of denoising score matching in the low-noise regime, we reformulate score estimation by exploiting a key property of additive Gaussian corruption. When the corruption noise is sufficiently small, the score of the noisy observation approaches that of the clean signal, i.e., $s_X(x) = \nabla_x \log p_X(x) \approx \nabla_x \log p_Y(x)$. This observation enables incorporating *target-side* structural information to guide score estimation, instead of relying solely on the conditional noise-model score $s_{X|Y}(x | y) = \nabla_x \log p_{X|Y}(x | y)$. Following target score matching [4], the marginal score admits the approximation

$$\nabla_x \log p_X(x) \approx \mathbb{E}[\nabla_y \log p_Y(Y) | X = x]. \quad (5)$$

In practice, we approximate the clean-data score using a target-score network $s_Y^\phi(y)$ and train the noisy score estimator s_X^θ to match this target-guided estimate. Unlike DSM, whose target magnitude scales with σ_a^{-2} under Gaussian noise, the proposed formulation avoids explicit noise amplification and provides a substantially more stable training signal in the low-noise regime.

3.4 Target Matching

To approximate the target-side expectation in Sec. 3.3, we construct a bank of coarse structural targets representing object-consistent particle appearances. When the particle identity in SPA is known, publicly available 3D templates from EMDB [7] or PDB [6][§] can be used as reference volumes. When no external structural template is available, a coarse reference can alternatively be estimated directly from the data (e.g., via ab-initio reconstruction), serving the same role as a low-frequency structural prior. Since such references often contain high-frequency details not reliably observable in noisy cryo-EM micrographs, we apply low-pass filtering to retain only coarse structural components.

From the reference volume, we generate a bank of 2D target projections $\{\tilde{y}_j\}_{j=1}^m$ over multiple viewing directions, where each $\tilde{y}_j \in \mathbb{R}^d$ is a plausible particle observation. Let $\psi(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^q$ be a feature mapping and define $z_j = \psi(\tilde{y}_j) \in \mathbb{R}^q$. These features span a target subspace $\mathcal{V}_T = \text{span}\{z_1, \dots, z_m\} \subset \mathbb{R}^q$, with orthogonal projector B_T . Given a noisy patch $x \in \mathbb{R}^d$ and its feature representation $x' = \psi(x)$, we project it onto the target subspace as $\hat{x}' = B_T x'$. This projection preserves components consistent with the target manifold while suppressing orthogonal variations, providing a feature-space approximation of the target-consistent component required in Sec. 3.3.

3.5 Train with Target-Score Supervision

We use a surrogate target score as a supervision signal to train the noisy score network $s_\theta(x) \approx \nabla_x \log p_X(x)$. Consistent with the additive Gaussian corruption

[§]EMDB: <https://www.ebi.ac.uk/emdb/>; PDB: <https://www.rcsb.org/>.

model used during training, we introduce an isotropic Gaussian *surrogate score model* $p_Y^{\text{sur}}(y) = \mathcal{N}(0, \sigma^2 I)$ solely to define a closed-form target score field, where σ^2 is estimated once from the target bank (empirical Bayes), and targets are centered relative to the matched structural reference. This surrogate does not model the true clean-data distribution and is not used to generate targets; it only provides an analytic score field $\nabla_y \log p_Y^{\text{sur}}(y) = -y/\sigma^2$.

We train the noisy score network $s_\theta(x) \approx \nabla_x \log p_X(x)$ using a surrogate target score as supervision. Under the additive Gaussian corruption model, we introduce an isotropic Gaussian surrogate $p_Y^{\text{sur}}(y) = \mathcal{N}(0, \sigma^2 I)$ to define a closed-form target score field, where σ^2 is estimated once from the target bank (empirical Bayes) and targets are centered relative to the matched structural reference. This surrogate is used only for the analytic score $\nabla_y \log p_Y^{\text{sur}}(y) = -y/\sigma^2$, not to model the true clean-data distribution or generate targets.

Construct TSM pairs For $x \in \mathbb{R}^d$, we build an input-adaptive target to approximate the conditional expectation in Sec. 3.3. Given target features $\{z_j\}_{j=1}^m$ and the projected noisy feature $\hat{x}'$ (Sec. 3.4), we compute cosine similarities

$$a_j(x) = \text{sim}(\hat{x}', z_j) = \frac{\hat{x}'^\top z_j}{\|\hat{x}'\|_2 \|z_j\|_2}, \quad j = 1, \dots, m.$$

We convert them into weights via a temperature-scaled softmax,

$$w_j(x) = \frac{\exp(a_j(x)/\tau)}{\sum_{k=1}^m \exp(a_k(x)/\tau)}, \quad \sum_{j=1}^m w_j(x) = 1, \quad (6)$$

and form the customized target by weighted aggregation:

$$\tilde{y}_{\text{tsm}}(x) = \sum_{j=1}^m w_j(x) \tilde{y}_j. \quad (7)$$

This yields the training pair $(x, \tilde{y}_{\text{tsm}}(x))$ and can be viewed as a nonparametric estimate of the conditional target expectation in Sec. 3.3.

Rescaled objective under AR-DAE parameterization Replacing the conditional noise-model score with target-guided supervision introduces scale mismatch across noise levels. To stabilize optimization, we follow the AR-DAE pre-conditioning analysis in [4, 17, 24] and use the following per-sample rescaled loss:

$$\ell_{\text{TSM}}^{\text{res}}(x; \theta) := \frac{\sigma^2 \sigma_a^2 + \sigma^4}{\sigma^2} \left\| \frac{\sigma_a}{\sigma \sqrt{\sigma_a^2 + \sigma^2}} s_\theta\left(\frac{x}{\sqrt{\sigma^2 + \sigma_a^2}} + \sigma_a\right) + \frac{x}{\sigma^2 + \sigma_a^2} - \frac{\tilde{y}_{\text{tsm}}(x)}{\sigma^2} \right\|_2^2. \quad (8)$$

We optimize the expected objective $L_{\text{TSM}}^{\text{res}}(\theta) = \mathbb{E}_x[\ell_{\text{TSM}}^{\text{res}}(x; \theta)]$. This rescaling preserves target-score supervision while restoring scale compatibility with DSM, leading to more stable gradients, especially at low noise. As in standard score-based training, σ_a can be gradually annealed during optimization.

Relation to DSM Under additive Gaussian corruption,

$$x = y + \sigma_a u, \quad u \sim \mathcal{N}(0, I), \quad (9)$$

the noise-model score is $\nabla_x \log p_{X|Y}(x | y) = -\frac{x-y}{\sigma_a^2}$. The corresponding per-sample DSM loss is

$$\ell_{\text{DSM}}(x, u; \theta) = \left\| s_\theta(x + \sigma_a u) + \frac{u}{\sigma_a} \right\|_2^2 = \|\sigma_a s_\theta(x + \sigma_a u) + u\|_2^2, \quad (10)$$

where the second form corresponds to the AR-DAE residual parameterization. The expected DSM objective is

$$L_{\text{DSM}}(\theta) = \mathbb{E}_{x,u}[\ell_{\text{DSM}}(x, u; \theta)].$$

DSM and TSM estimate the same marginal score $\nabla_x \log p_X(x)$ but use different supervision signals. TSM can be viewed as a variance-reduced approximation to DSM, replacing the conditional noise-model score with a target-guided estimate under the small-noise assumption.

3.6 Adaptive Learning

DSM and TSM provide complementary supervision signals during training. DSM captures globally consistent image statistics dominated by background and low-SNR components, while TSM enhances particle-specific structural information through target guidance. The reliability of target-guided supervision varies across noisy observations. While DSM remains stable regardless of particle visibility, TSM depends on alignment between the noisy input and proxy structural targets. We therefore introduce an adaptive learning strategy that interpolates between DSM and TSM according to target-alignment confidence. Using the similarity weights $\{w_j(x)\}_{j=1}^m$ defined in Sec. 3.5, we measure alignment confidence as

$$w_t(x) := \max_{1 \leq j \leq m} w_j(x) \in [0, 1], \quad (11)$$

which reflects consistency between the noisy observation and at least one target projection under unknown viewing directions in SPA. The confidence weight $w_t(x)$ is computed without gradient propagation and used only as a sample-wise weighting factor. The final objective interpolates between TSM and DSM:

$$L_{\text{adp}}(\theta) = \mathbb{E}_x \left[w_t(x) \ell_{\text{TSM}}^{\text{res}}(x; \theta) + (1 - w_t(x)) \ell_{\text{DSM}}(x; \theta) \right]. \quad (12)$$

Samples with confident alignment emphasize target-guided learning, while uncertain inputs (non-particle regions) naturally fall back to DSM supervision, improving robustness under heterogeneous cryo-EM imaging conditions.

Algorithm 1 Target-Guided Score Training**Require:** Noisy micrographs; target bank $\{\tilde{y}_j\}_{j=1}^m$ **Ensure:** Trained score network s_θ

```

1: Estimate  $\sigma^2$  from target bank
2: Extract noisy patches  $\{x\}$ 
3: for  $t = 1$  to  $T$  do
4:   Set  $\sigma_a(t)$  and  $\lambda(t)$ 
5:   for all  $x$  do
6:     Compute  $\psi_\theta(x)$ 
7:     Compute similarities  $\{a_j(x)\}$ 
8:     Compute weights  $w_j(x)$ 
9:      $\tilde{y}_{\text{tsm}}(x) \leftarrow \sum_j w_j(x) \tilde{y}_j$ 
10:     $w_t(x) \leftarrow \max_j w_j(x)$ 
11:     $\ell_{\text{final}} \leftarrow \lambda(t)w_t(x)\ell_{\text{TSM}} + (1 - \lambda(t)w_t(x))\ell_{\text{DSM}}$ 
12:   end for
13:   Update  $\theta$  via SGD
14: end for
15: return  $s_\theta$ 

```

3.7 Training Procedure

Fig. 1 summarizes the overall target-guided training pipeline. Algorithm 1 provides the detailed procedure. During training, noisy cryo-EM micrographs are partitioned into patches using a sliding-window strategy. At iteration t , a noise level $\sigma_a(t)$ and a global guidance coefficient $\lambda(t)$ are set according to predefined schedules, with $\lambda(t) = 0$ during an initial warmup phase to stabilize early training. For each patch x , similarities to the target bank are computed to construct an input-adaptive proxy target $\tilde{y}_{\text{tsm}}(x)$ and a confidence score $w_t(x)$. The effective supervision weight is given by $\lambda(t)w_t(x)$, yielding the combined loss

$$\ell_{\text{final}} = \lambda(t)w_t(x)\ell_{\text{TSM}} + (1 - \lambda(t)w_t(x))\ell_{\text{DSM}}.$$

When alignment confidence is low or during early iterations, optimization relies primarily on DSM. As training progresses and alignment improves, target-guided supervision gradually dominates.

4 Experiments

We evaluate our method on multiple cryo-EM datasets against representative baselines. This section first introduces the experimental setup and then reports results on denoising, particle picking, and 3D reconstruction.

4.1 Experimental Setup

Datasets. We select cryo-EM datasets from Electron Microscopy Public Image Archive (EMPIAR)[¶] [10], including EMPIAR-10081 [14], EMPIAR-10289, and

[¶]EMPIAR: <https://www.ebi.ac.uk/empiar/>.

EMPIAR-10291 [5]. To further assess generalization beyond transport-protein data, we additionally include EMPIAR-10017 (Beta-galactosidase) and EMPIAR-10406 (70S ribosome). We also construct simulated particle stacks with known ground truth to verify the core denoising mechanism under controlled conditions. For all datasets, we adopt the CryoPPP benchmark [8], which provides standardized preprocessing micrographs (e.g., motion correction) and expert-validated particle picking annotations.

Baselines and backbones. We select representative and widely used baselines for cryo-EM image denoising, including classical signal processing approaches and SOTA learning-based methods. Specifically, we include classical **Gaussian** low-pass filter, **Topaz-Denoise**, and **DRACO**. These baselines cover both traditional filtering-based techniques and modern deep learning frameworks, providing a comprehensive comparison across different methodological paradigms.

Downstream cryo-EM processing. To assess the impact of denoising on structural reconstruction, we perform a standardized cryo-EM processing pipeline. Particle picking is conducted using the Topaz particle picker [3] integrated in cryoSPARC software [20], employing the default pretrained ResNet16 model (32 feature units) without additional fine-tuning. All subsequent cryo-EM data processing steps, including particle extraction, 2D classification, ab initio reconstruction, and refinement, are performed in cryoSPARC. For target construction, all main-paper results use a unified 10Å low-pass filter applied to the reference map unless otherwise stated. This unified pipeline ensures a fair and reproducible evaluation of downstream performance.

Evaluation metrics. Since clean ground-truth micrographs are unavailable in modern cryo-EM, image-level metrics such as PSNR and SSIM are not applicable. We therefore evaluate denoising performance through downstream cryo-EM tasks. For particle picking, we use a distance-based detection metric, where a prediction is considered correct if it matches a ground-truth particle within a predefined distance threshold under one-to-one matching. We report precision, recall, and F1 under both micro- and macro-averaging. Micro scores are computed globally by aggregating counts across all micrographs, while macro scores are averaged over per-micrograph metrics. Because macro-averaging weights every micrograph equally, it is more sensitive to rare or sparsely populated particle subsets than micro-averaging, which should be kept in mind when comparing the two protocols. For structural reconstruction, we report the resolution of the reconstructed 3D density map, estimated using the gold-standard Fourier Shell Correlation (FSC) 0.143 criterion.

4.2 Overall Performance

Denoising Performance The qualitative comparisons in Fig. 3 suggest that effective cryo-EM denoising enhances particle-relevant structure rather than

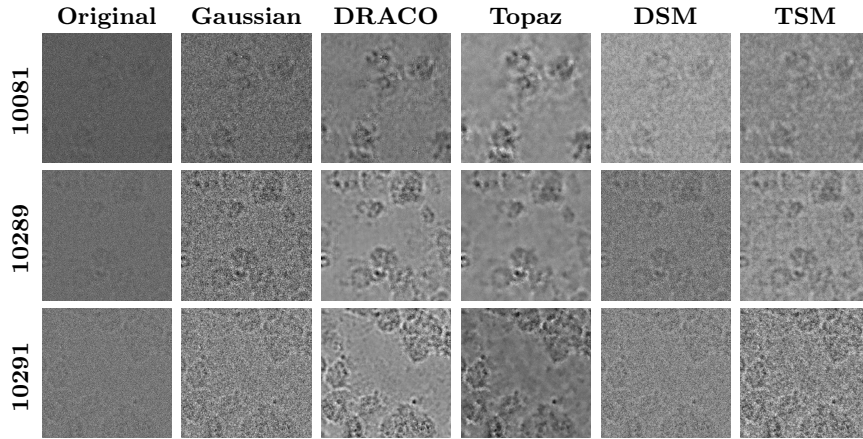


Fig. 3: Qualitative denoising comparisons on three cryo-EM datasets.

	EMPIAR-10291			EMPIAR-10289			EMPIAR-10081		
Method	P	R	F1	P	R	F1	P	R	F1
Original	0.617	0.646	0.631	0.352	0.287	0.316	0.546	0.738	0.627
Gaussian	0.610	0.675	0.641	0.352	0.287	0.316	0.546	0.739	0.628
DRACO	0.542	0.723	0.620	0.384	0.351	0.367	0.383	0.456	0.417
Topaz	0.539	0.710	0.613	0.384	0.350	0.366	0.389	0.459	0.421
DSM	0.601	0.694	0.644	0.360	0.304	0.330	0.567	0.755	0.648
TSM	0.590	0.721	0.649	0.382	0.358	0.370	0.577	0.729	0.644

Table 1: Micro-averaged particle picking performance comparison.

merely altering image appearance. Gaussian filtering provides only limited enhancement, leaving many weak particle patterns difficult to distinguish. DRACO and Topaz produce stronger responses, but these changes are accompanied by less stable local structure. By contrast, the score-based methods generate more particle-consistent patterns while better preserving the overall structure. TSM is particularly effective in this regard, yielding clearer particle visibility and more reliable cues for downstream particle picking. Such improved visual interpretability also facilitates manual inspection and particle annotation. These observations matches the following downstream picking results.

Particle Picking Performance Tables 1 and 2 report downstream particle picking performance under the distance-based (32px) evaluation protocol using both micro- and macro-averaged metrics. Compared with DRACO and Topaz, our score-based methods consistently achieve stronger overall performance across

Method	EMPIAR-10291	EMPIAR-10289	EMPIAR-10081
Original	0.623 ± 0.073	0.313 ± 0.048	0.610 ± 0.111
Gaussian	0.633 ± 0.071	0.313 ± 0.048	0.610 ± 0.111
DRACO	0.610 ± 0.082	0.359 ± 0.062	0.390 ± 0.132
Topaz	0.604 ± 0.081	0.360 ± 0.055	0.393 ± 0.136
DSM	0.636 ± 0.072	0.325 ± 0.047	0.626 ± 0.118
TSM	0.640 ± 0.076	0.364 ± 0.048	0.623 ± 0.115

Table 2: Macro-averaged F1 score (mean $\pm$ std across micrographs) under distance-based evaluation.

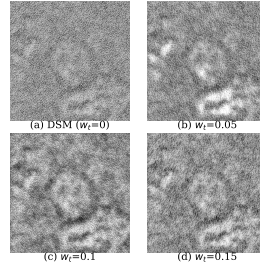


Fig. 4: Ablation of the target weight w_t on EMPIAR-10289.

datasets, indicating that effective cryo-EM denoising should improve particle-relevant structure rather than merely amplifying particle-like responses.

Within the score-based family, TSM achieves the best F1 on EMPIAR-10291 and EMPIAR-10289 under both micro and macro evaluation, while DSM performs best on EMPIAR-10081 and TSM remains highly competitive. This difference is consistent with our theoretical motivation: target-guided supervision provides a more stable training signal in the low-noise regime and is particularly beneficial when particle localization depends on subtle structural cues. On EMPIAR-10081, DSM achieves the highest F1 while TSM attains the highest precision, suggesting that both score-based variants improve downstream picking but favor different precision–recall operating points. Examining the precision and recall behavior more closely, target guidance is most beneficial when particles are weak but still retain projection patterns consistent with the reference. On the lower-SNR datasets EMPIAR-10289 and EMPIAR-10291, it mainly improves recall, indicating better detectability after denoising, whereas on EMPIAR-10081 where the signal is already clearer it mainly improves precision by reducing structure-inconsistent responses.

3D Reconstruction Performance The FSC curves for 3D reconstructions on EMPIAR-10291 and EMPIAR-10081 are shown in Fig. 6, with the corresponding density maps visualized in Fig. 5. Reconstruction resolution is evaluated using the gold-standard FSC 0.143 criterion, where a lower resolution value indicates better structural recovery. We do not include EMPIAR-10289 here because it targets the same macromolecular complex as EMPIAR-10291 but was collected under different microscope settings and experimental conditions. Overall, DSM and TSM achieve substantially better reconstruction quality than the other learned denoising baselines, particularly DRACO and Topaz. This trend is also consistent with the reconstructed density maps.

On EMPIAR-10081, DRACO leads to severe structural degradation, with substantial missing regions, while DSM and TSM preserve both the overall morphology and local structural details much more faithfully, and their reconstructions remain visually much closer to the GT maps. Although Gaussian filtering can also produce reasonable reconstructions, achieving such performance requires

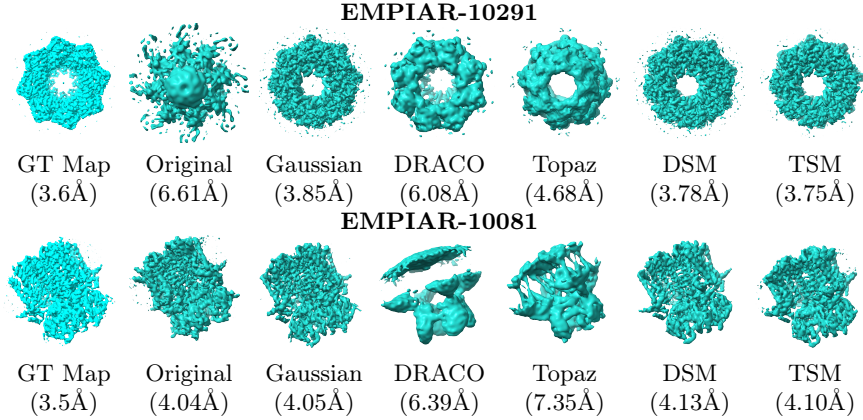


Fig. 5: 3D reconstruction visualizations on EMPIAR-10291 and EMPIAR-10081. Each row shows the ground truth (GT) and reconstructions from the original input and five compared methods. Reported values indicate reconstruction resolution (lower is better).

careful parameter tuning, which makes it less reliable in practice. We also note that small differences in the reported resolution values do not necessarily imply substantial differences in reconstruction quality. When the values are close (i.e., difference within $0.1\text{\AA}$), the practical difference is often limited and should be interpreted together with the FSC curves and the visualized density maps.

These observations are also aligned with the particle-picking results. On EMPIAR-10291, although the original micrographs already achieve reasonable overall picking performance, the remaining false positives in the picked particle set can still degrade downstream reconstruction and are consistent with the artifacts observed in the reconstructed map. In contrast, DSM and TSM yield cleaner reconstructions, suggesting that denoising is more effective when it improves particle detectability while reducing false positives and preserving particle-level structural information for downstream alignment and averaging.

4.3 Ablation Study & Sensitivity Analysis

Target weight w_t . We study the effect of the target weight w_t on EMPIAR-10289. Fig. 4 compares DSM-only ($w_t = 0$) with $w_t \in \{0.05, 0.1, 0.15\}$. Without target guidance (Fig. 4a), DSM mainly reduces noise while leaving particle structures poorly defined. Enabling TSM (Figs. 4b–d) consistently improves structural clarity across different values of w_t , producing sharper particle boundaries and reduced background responses. The improvement remains stable over a wide range of w_t , indicating low sensitivity to precise parameter tuning once target guidance is introduced. When w_t becomes excessively large (Fig. 4d), mild over-regularization begins to suppress fine structural variations. Overall, moderate values of w_t achieve the best balance between structural guidance and preservation of fine details.

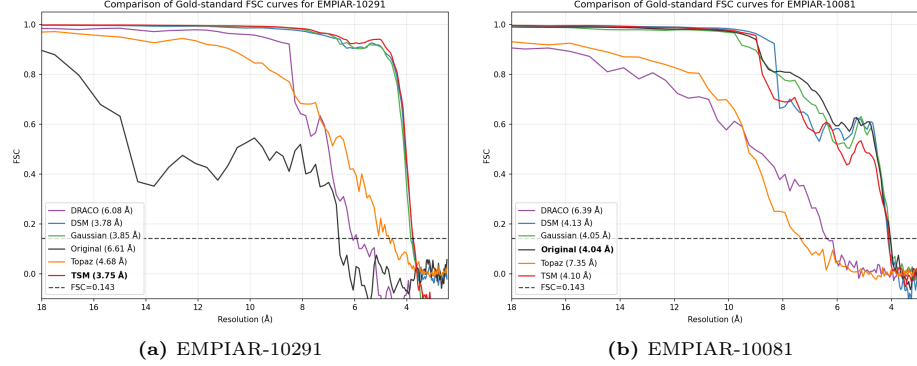


Fig. 6: FSC curves of 3D reconstructions comparison on EMPIAR-10291 and EMPIAR-10081. Resolution is determined at FSC 0.143; a right-shifted curve (higher spatial frequency at the threshold) indicates better resolution.

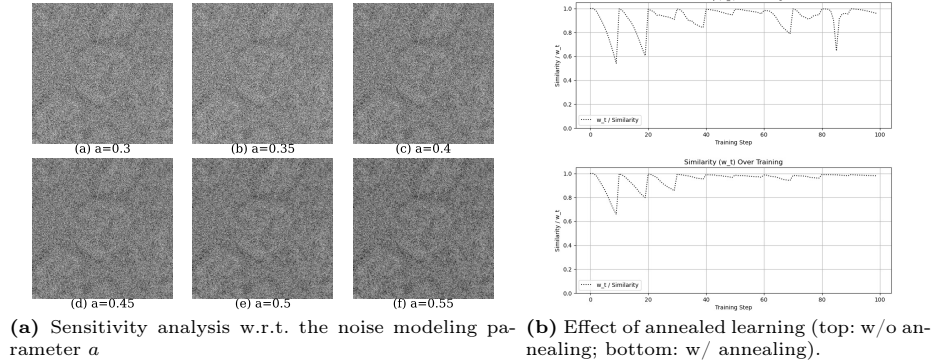


Fig. 7: Left: sensitivity analysis under different noise simulation parameters a . Right: comparison of training w/ and w/o annealed learning.

Global guidance coefficient $\lambda(t)$. We evaluate the effect of the global guidance coefficient $\lambda(t)$ on EMPIAR-10291, which controls the contribution of TSM during training. Fig. 7b compares training without annealing (top) and with annealed learning (bottom), where the similarity weight w_t between a fixed noisy input and its plausible targets is tracked across training. The encoder is updated every 10 epochs. Without annealing, TSM is applied from the beginning, when the encoder is still unstable; consequently, the projected noisy-target features are unreliable, and the similarity w_t exhibits noticeable oscillations after each feature update. These fluctuations indicate that early similarity estimates are not trustworthy, leading to unstable target guidance. In contrast, with annealed learning we set $\lambda(t) = 0$ in the first phase to optimize only DSM, allowing the encoder to learn stable feature representations before introducing target guidance. The coefficient $\lambda(t)$ is then increased linearly in the middle phase and

fixed to $\lambda(t) = 1$ in the final phase. As shown in Fig. 7b (bottom), this schedule significantly stabilizes the similarity curve, producing consistently high and less oscillatory w_t values. Overall, annealed learning improves the reliability of similarity estimation and yields more stable optimization dynamics by delaying target guidance until the encoder becomes sufficiently well-structured.

Sensitivity to noise modeling coefficient σ_a . Fig. 7a evaluates the sensitivity to the noise modeling coefficient a , which controls the noise scale σ_a . Qualitatively, the method produces consistent visual results across a wide range of a , and a moderate range ($a \in [0.4, 0.5]$) offers a favorable trade-off between noise suppression and structure detail preservation. Smaller a leaves more residual noise, whereas larger a slightly oversmooths fine structural details. Overall, the visual quality shows limited sensitivity to the choice of σ_a .

5 Conclusion

In conclusion, we introduced a target-guided score-based denoising framework for cryo-EM to address extreme noise and weak particle signals. Our design leverages score-based learning to preserve particle-relevant structure without overly suppressing non-particle regions, and further improves particle-region accuracy via target-informed guidance. Extensive experiments across multiple cryo-EM datasets show consistent downstream benefits: our method improves particle picking over classical signal-processing baselines, remains competitive with strong deep learning denoisers, and supports high-quality 3D reconstruction, as reflected by FSC trends and cleaner density maps.

Looking ahead, improving target construction under dataset-specific and heterogeneous noise conditions remains an important direction. While similarity-guided annealing provides stable guidance during training, richer target representations and more accurate noise modeling could further improve robustness on particularly challenging micrographs. Beyond cryo-EM, the proposed target-informed score-matching formulation is general and could extend to other low-SNR scientific imaging problems, including cryo-electron tomography, fluorescence microscopy, and other physics-driven inverse problems where structural preservation is critical.

Acknowledgement

This work was supported in part by U.S. NIH grant R35GM158094. This work used the Bridges-2 system at the Pittsburgh Supercomputing Center (PSC) through allocation cis250268p from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by U.S. National Science Foundation (NSF) grants #2138259, #2138286, #2138307, #2137603, and #2138296.

References

1. Baxter, W.T., Grassucci, R.A., Gao, H., Frank, J.: Determination of signal-to-noise ratios and spectral snrs in cryo-em low-dose imaging of molecules. *Journal of structural biology* **166**(2), 126–132 (2009)
2. Bepler, T., Kelley, K., Noble, A.J., Berger, B.: Topaz-denoise: general deep denoising models for cryoem and cryoet. *Nature communications* **11**(1), 5208 (2020)
3. Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A.J., Berger, B.: Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature methods* **16**(11), 1153–1160 (2019)
4. Bortoli, V.D., Hutchinson, M., Wirnsberger, P., Doucet, A.: Target score matching (2024)
5. Burendei, B., Shinozaki, R., Watanabe, M., Terada, T., Tani, K., Fujiyoshi, Y., Oshima, A.: Cryo-em structures of undocked innexin-6 hemichannels in phospholipids. *Science advances* **6**(7), eaax3157 (2020)
6. Burley, S.K., Berman, H.M., Kleywegt, G.J., Markley, J.L., Nakamura, H., Velankar, S.: Protein data bank (pdb): the single global macromolecular structure archive. *Protein crystallography: methods and protocols* pp. 627–641 (2017)
7. wwPDB Consortium, T.: Emdb—the electron microscopy data bank. *Nucleic acids research* **52**(D1), D456–D465 (2024)
8. Dhakal, A., Gyawali, R., Wang, L., Cheng, J.: A large expert-curated cryo-em image dataset for machine learning protein particle picking. *Scientific Data* **10** (2023). <https://doi.org/10.1038/s41597-023-02280-2>
9. Frank, J.: *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. Oxford University Press (2006)
10. Iudin, A., Korir, P.K., Somasundharam, S., Weyand, S., Cattavittello, C., Fonseca, N., Salih, O., Kleywegt, G.J., Patwardhan, A.: Empiar: the electron microscopy public image archive. *Nucleic Acids Research* **51**(D1), D1503–D1511 (2023)
11. Jiang, L., Zhu, B., Long, W., Xu, J., Wu, Y., Li, Y.W.: A review of denoising methods in single-particle cryo-em. *Micron* p. 103817 (2025)
12. Krull, A., Buchholz, T.O., Jug, F.: Noise2void: Learning denoising from single noisy images. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 2129–2137 (2019)
13. Kühlbrandt, W.: The resolution revolution. *Science* **343**(6178), 1443–1444 (2014)
14. Lee, C.H., MacKinnon, R.: Structures of the human hcn1 hyperpolarization-activated channel. *Cell* **168**(1), 111–120 (2017)
15. Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., Aila, T.: Noise2noise: Learning image restoration without clean data. In: *Proceedings of the International Conference on Machine Learning (ICML)*. pp. 2965–2974. PMLR (2018)
16. Li, H., Zhang, H., Wan, X., Yang, Z., Li, C., Li, J., Han, R., Zhu, P., Zhang, F.: Noise-transfer2clean: Denoising cryo-em images based on noise modeling and transfer. *Bioinformatics* **38**(7), 2022–2029 (2022)
17. Lim, J.H., Courville, A., Pal, C., Huang, C.W.: Ar-dae: Towards unbiased neural entropy gradient estimation. In: *Proceedings of the 37th International Conference on Machine Learning (ICML)*. pp. 6061–6071. PMLR (2020)
18. Milne, J.L., Borgnia, M.J., Bartesaghi, A., Tran, E.E., Earl, L.A., Schauder, D.M., Lengyel, J., Pierson, J., Patwardhan, A., Subramaniam, S.: Cryo-electron microscopy—a primer for the non-microscopist. *The FEBS Journal* **280**(1), 28–45 (2013)

19. Nogales, E.: The development of cryo-em into a mainstream structural biology technique. *Nature Methods* **13**(1), 24–27 (2016)
20. Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A.: cryosparc: algorithms for rapid unsupervised cryo-em structure determination. *Nature methods* **14**(3), 290–296 (2017)
21. Quan, Y., Chen, M., Pang, T., Ji, H.: Self2self with dropout: Learning self-supervised denoising from a single image. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 1887–1895 (2020)
22. Shen, Y., Dai, H., Chen, Q., Zeng, Y., Zhang, J., Pei, Y., Yu, J.: DRACO: A denoising-reconstruction autoencoder for cryo-em. In: *Proceedings of the Thirty-Eighth Annual Conference on Neural Information Processing Systems (NeurIPS)* (2024)
23. Vincent, P.: A connection between score matching and denoising autoencoders. *Neural computation* **23**(7), 1661–1674 (2011)
24. Xie, Y., Yuan, M., Dong, B., Li, Q.: Unsupervised image denoising with score function. In: *Proceedings of the Thirty-Seventh International Conference on Neural Information Processing Systems (NeurIPS)* (2023)
25. Zhang, J., Chen, Q., Zeng, Y., Gao, W., He, X., Liu, Z., Yu, J.: Cryogem: Physics-informed generative cryo-electron microscopy (2024)