UNPAIRED ALL-IN-ONE WEATHER RESTORATION VIA STRUCTURED STOCHASTIC UPDATES IN SCHRÖDINGER BRIDGES

Anonymous authors
Paper under double-blind review

ABSTRACT

Unpaired all-in-one weather image restoration requires a single generative model to learn restoration mappings for multiple degradations from non-corresponding degraded and clean images. Unpaired Neural Schrödinger Bridge progressively constructs restoration paths by predicting clean endpoints at different bridge positions. However, its original training estimates gradients using only one complete stochastic realization at a randomly sampled bridge position, without integrating conditional gradient information from multiple stochastic realizations in the joint parameter space of the generator and the trainable feature network. To this end, we propose Structured Stochastic Update Schrödinger Bridge (SSUSB). Conditional Gradient Ensembling (CGE) estimates joint gradients of the generator and feature network from multiple complete stochastic realizations and constructs a consensus using their mean. Bridge Stratified Coupling (BSC) further implements joint sampling without replacement while preserving the uniform marginal distribution of the bridge position for each view. Consensus Tangent-space Projection (CTP) uses the multi-view consensus to calibrate a prespecified candidate gradient. In a frozen Adam-derived metric, it replaces the candidate's axial component with the consensus while preserving its orthogonal component. Extensive experiments demonstrate that SSUSB can effectively improve unpaired all-in-one weather image restoration performance.

1 INTRODUCTION

Real-world outdoor vision faces diverse and complex degradations, motivating all-in-one image restoration to handle multiple degradation types within a single model. Early studies explore all-in-one weather restoration through architecture search Li et al. (2020) and shared Transformers Valanarasu et al. (2022), while subsequent methods introduce prompt learning Potlapalli et al. (2023), frequency modulation Cui et al. (2025), diffusion-based modeling (Cheng et al., 2025; Tang et al., 2026; Zheng et al., 2024), degradation-aware modulation, visual in-context learning Rajagopalan & Patel (2025) and prompt guidance Wen et al. (2025) to improve adaptability to diverse degradations. Despite these advances, most existing methods still rely on paired data or explicit degradation supervision Liu et al. (2024); Li et al. (2022a); Liu et al. (2023). However, real-world weather and imaging conditions make strictly aligned degraded-clean pairs difficult and costly to obtain Fang et al. (2026). Therefore, we investigate whether an all-in-one restoration model can be learned using only a collection of images with diverse degradations and an unaligned collection of clean images.

To reduce reliance on paired data, existing studies model degraded-to-clean transport through endpoint mappings or stochastic paths. RCOT Tang et al. (2024) incorporates degradation residuals into the transport cost and restoration mapping, while DA-RCOT Tang et al. (2025) extends residual conditioning to multiple degradations using multiscale representations. These methods use intermediate predictions to construct residual conditions, without explicitly modeling a stochastic restoration process through a common path measure. In contrast, Schrödinger bridges formulate transport over stochastic paths relative to a reference diffusion under endpoint distribution constraints. Along these paths, conditional transitions progressively evolve the bridge state while the generator refines its clean endpoint predictions. UNSB Kim et al. (2024) uses the Markov decomposition and self-similarity of Schrödinger bridges to predict clean endpoints at different bridge positions and

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

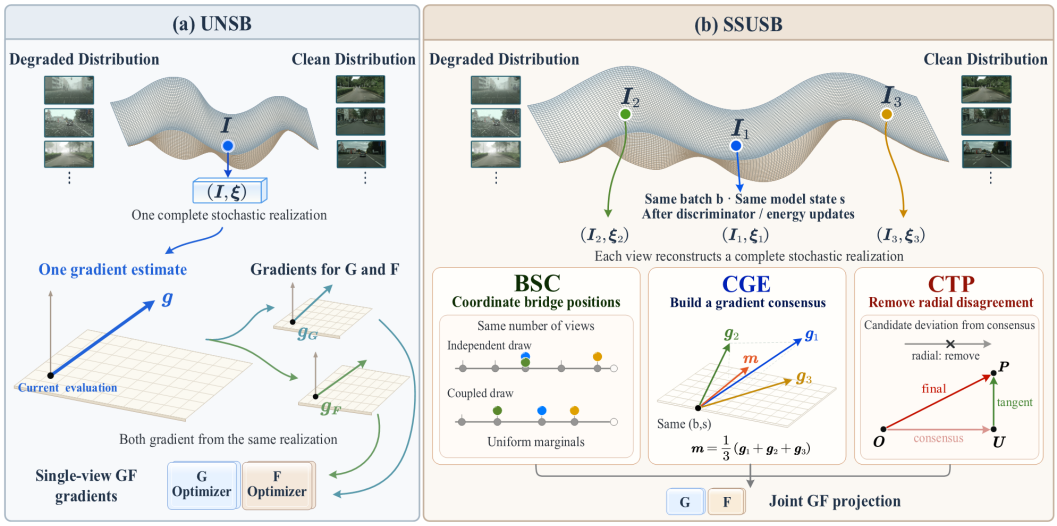


Figure 1: Stochastic updates in UNSB and SSUSB. (a) UNSB computes GF gradients from one complete stochastic realization. (b) SSUSB builds a multi-view consensus at a fixed batch and model state, then corrects a candidate using a shared GF projection coefficient.

recursively sample subsequent states. DDSB Yi et al. (2025) further introduces degradation-aware transport and dynamic consistency constraints to preserve degradation information throughout this process. These studies provide a path-space foundation for modeling shared restoration dynamics across multiple weather degradations under a unified time coordinate.

Existing bridge-based methods emphasize restoration paths and their constraints. We focus on how stochastic realizations inform updates of the shared generator (G) and trainable feature network (F). As shown in Figure 1(a), UNSB computes gradients for both networks from the same loss using one complete stochastic realization at a randomly selected bridge position. The resulting gradient estimate varies with bridge-position selection and internal randomness. Gradients obtained across iterations generally correspond to different model states. Therefore, we evaluate multiple complete realizations at a fixed batch and model state and investigate how their gradients can jointly inform each GF update.

To address this issue, we propose Structured Stochastic Update Schrödinger Bridge (SSUSB), as shown in Figure 1(b). After updating the discriminator (D) and energy network (E), CGE averages joint GF gradients from multiple complete realizations sharing the same batch and model state. BSC samples bridge positions without replacement, improving coverage at a fixed view budget while preserving uniform per-view marginals. CTP uses the resulting consensus to calibrate a prespecified candidate gradient in a frozen Adam-derived metric. It matches the candidate's component along the consensus axis to that of the consensus while preserving its orthogonal component. A single projection coefficient computed from both GF blocks couples their corrections before the modified gradients are passed to the respective optimizers.

Our main contributions are as follows.

1. We propose Structured Stochastic Update Schrödinger Bridge (SSUSB), which forms a joint gradient consensus for G and F from multiple complete stochastic realizations. These realizations share the fixed state after discriminator and energy network updates and inform a single joint parameter update.
2. We design bridge stratified coupling (BSC) to sample distinct bridge positions uniformly without replacement, improving coverage at a fixed view budget while preserving uniform per-view marginals.
3. We introduce CTP, which makes the smallest modification to a prespecified joint GF gradient under a consensus-axis constraint in a frozen Adam-derived metric. A single projection coefficient computed from both parameter blocks couples their corrections while preserving the candidate's orthogonal component.

4. Experiments on six restoration benchmarks and three unseen datasets demonstrate improved restoration quality and cross-dataset generalization relative to the evaluated unpaired baselines.

2 Related Work

Schrödinger Bridge. Schrödinger bridges minimize path-space relative entropy to a reference process subject to endpoint marginal constraints Léonard (2014). DSB De Bortoli et al. (2021) applies the framework to score-based generation through iterative proportional fitting, while DSBM Shi et al. (2023) learns transport dynamics through iterative Markovian fitting. FSBM Theodoropoulos et al. (2025) incorporates feedback guidance from limited paired samples. TSBM Noble et al. (2026) extends bridge matching to reference processes with time-dependent potentials and introduces trajectory variance reduction. These studies provide a foundation for modeling stochastic paths between degraded and clean domains for unpaired restoration.

Unpaired Image Translation. CycleGAN Zhu et al. (2017) constrains mappings through cycle consistency, while CUT Park et al. (2020) and DCLGAN Han et al. (2021) preserve content correspondence through unidirectional and bidirectional contrastive learning, respectively. Santa Xie et al. (2023) introduces shortest-path regularization, while DN Ho et al. (2025) reduces artifacts in high-resolution tiled inference through pixel-level statistics estimation. Contrastive-SDE Kotyada et al. (2025) uses time-dependent contrastive guidance to preserve domain-invariant features during inference with a pretrained stochastic differential equation model. These methods provide distribution and content constraints for learning mappings without paired supervision.

Unpaired Image Restoration. Compared with general translation across domains, unpaired restoration emphasizes balancing degradation removal and scene structure preservation. D^4 Yang et al. (2022) augments training through haze density and depth decomposition, while ODCR Wang et al. (2024) separates degradation from content using orthogonal feature disentanglement and weighted contrastive constraints. Within the bridge framework, DehazeSB Lan et al. (2025) combines regularization for detail preservation with vision-language guidance to dehaze real images. DRDD Lin et al. (2026) combines paired mapping learning with denoising on unpaired target data, while FACT Zhu et al. (2026) uses pretrained priors and optimization during inference for zero-shot unified restoration. Unlike these methods, SSUSB focuses on designing a stochastic update mechanism under bridge objectives.

Stochastic Gradient Optimization. Stochastic gradient optimization improves gradient estimates through mini-batching Bottou et al. (2018), control variates Johnson & Zhang (2013) and stratified sampling Zhao & Zhang (2014), while PCGrad Yu et al. (2020) and CAGrad Liu et al. (2021) coordinate multi-task gradients through directional adjustments. We focus on unpaired bridge-based training, combining gradient ensembling at the same state, bridge-position sampling without replacement and consensus projection under a frozen metric in a single joint update of G and F.

3 Preliminaries

Schrödinger Bridge. Given the degraded distribution p_{d} and the clean distribution p_{c}, we seek a shared model that restores images to the clean domain while preserving scene content. Images are represented as vectors in $\mathbb{R}^d$. Let $\Omega = C([0,1],\mathbb{R}^d)$ be the continuous path space and $\mathcal{P}(\Omega)$ its set of probability measures. The Schrödinger bridge minimizes relative entropy to a reference path measure under endpoint constraints

$$\mathbb{Q}^\star \in \underset{\mathbb{Q}\in\mathcal{P}(\Omega)}{\arg\min}\, \mathrm{KL}(\mathbb{Q}\,\|\,\mathbb{R}^\tau)\,, \qquad \mathbb{Q}_0 = p_{\mathrm{d}}, \quad \mathbb{Q}_1 = p_{\mathrm{c}}, \tag{1}$$

where $\mathbb{Q}$ is a candidate path measure, $\mathbb{Q}^\star$ is the optimal path measure and $\mathbb{R}^\tau$ is the reference path measure induced by Brownian motion with initial distribution p_{d} and diffusion variance parameter $\tau > 0$.

Conditional Endpoint Prediction and Bridge Sampling. For discrete bridge times $0 = t_0 < \cdots < t_L = 1$, let x_i denote the image state at t_i, where $i \in \{0,\ldots,L-1\}$ indexes bridge positions.

Following UNSB's endpoint-prediction parameterization, G predicts a clean endpoint from x_i, the position condition and a random latent variable,

$$\widehat{x}_{c|i} = G_{\theta_G}(x_i, i, z_i), \qquad z_i \sim p_z, \tag{2}$$

where θ_G denotes the generator parameters. Different positions share the same set of generator parameters and both x_i and $\widehat{x}_{c|i}$ lie in image space. Given the predicted endpoint, the next state is sampled from the Brownian bridge transition,

$$x_{i+1} = (1-\alpha_i)x_i + \alpha_i \widehat{x}_{c|i} + \sigma_i \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, I_d). \tag{3}$$

where $\alpha_i = \frac{t_{i+1}-t_i}{1-t_i}$, $\sigma_i^2 = \tau \alpha_i (1-\alpha_i)(1-t_i)$. The latent variable and bridge noise are resampled at each step. Starting from the degraded input x_0, alternating endpoint prediction and state transition constructs a restoration sequence. When $i = L-1$, $\alpha_i = 1$ and $\sigma_i = 0$, the terminal output equals the final endpoint prediction.

Joint Gradient Observations at Fixed State. Let θ_F denote the trainable feature network parameters and define the joint parameter vector

$$\Theta = \mathrm{concat}(\theta_G, \theta_F) \in \mathbb{R}^p, \qquad p = p_G + p_F. \tag{4}$$

Given an unpaired batch b, all views evaluate gradients at the same reference state s, with joint parameters Θ_s. The state s includes all network parameters, relevant buffers, forward-pass rules and the optimizer history before the update. At the bridge position i, ξ collects bridge sampling noise, generative latent variables, feature sampling and other non-time randomness. Each complete stochastic realization constructs the bridge state, predicts the clean endpoint and computes the loss following the base model. Let $\Upsilon = \mathrm{concat}(\vartheta_G, \vartheta_F)$ be the joint differentiation variable. The local stochastic loss for a complete stochastic realization is defined as

$$\ell_{GF}(\Upsilon; b, i, \xi, s) = \lambda_{\mathrm{adv}} \ell_{\mathrm{adv}} + \lambda_{\mathrm{SB}} \ell_{\mathrm{SB}} + \lambda_{\mathrm{reg}} \ell_{\mathrm{reg}}. \tag{5}$$

where ℓ_{adv}, ℓ_{SB} and ℓ_{reg} denote the adversarial loss, bridge-based transport constraint and content-preserving regularization term, respectively. The joint gradient of a single complete stochastic realization is defined as

$$g_i(\xi; b, s) = \nabla_\Upsilon \ell_{GF}(\Upsilon; b, i, \xi, s)|_{\Upsilon=\Theta_s} = \mathrm{concat}(g_{G,i}(\xi; b, s), g_{F,i}(\xi; b, s)). \tag{6}$$

where $g_{G,i}$ and $g_{F,i}$ are the gradients of the same loss with respect to G and F parameters, respectively. The discriminator and energy network parameters remain fixed, while gradients propagate through both networks to the generator.

4 METHOD

SSUSB combines multi-view information at a fixed reference state to construct a joint gradient for G and F. As shown in Figure 2, CGE builds a gradient consensus, BSC coordinates bridge-position sampling and CTP forms the update direction.

4.1 CONDITIONAL GRADIENT ENSEMBLING

The gradient observation of a single complete stochastic realization is jointly affected by bridge-position selection and internal randomness, whereas observations across iterations are also accompanied by changes in model parameters. To reduce the estimation fluctuations of a single observation, CGE reconstructs multiple complete stochastic realizations before a joint update of G and F and forms a consensus by averaging the joint gradients. We fix the current unpaired batch b and reference state s. Let $\nu_i(\cdot \mid b, s)$ denote the distribution of non-time randomness at the bridge position i. Under UNSB's uniform sampling of discrete bridge positions, a complete stochastic realization is sampled as $I \mid b, s \sim \mathrm{Unif}\{0, \ldots, L-1\}$, $\xi \mid I = i, b, s \sim \nu_i(\cdot \mid b, s)$. L is the number of training bridge positions, I is the sampled index, Unif is the discrete uniform distribution and ξ is the non-time randomness. Based on the stochastic gradient $g_i(\xi; b, s)$, we define the bridge-position conditional expected gradient and its average gradient target under the uniform bridge-position distribution as

$$\mu_i(b, s) = \mathbb{E}_{\xi \sim \nu_i(\cdot|b,s)}[g_i(\xi; b, s)], \qquad \bar{\mu}(b, s) = \frac{1}{L} \sum_{i=0}^{L-1} \mu_i(b, s). \tag{7}$$

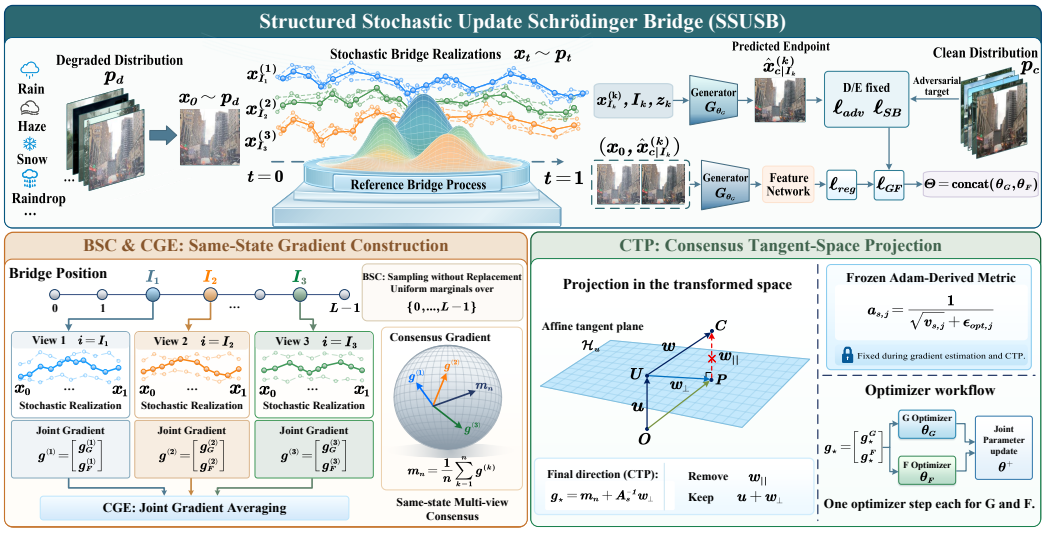


Figure 2: Overview of SSUSB. After D and E updates, CGE averages joint gradients from complete realizations sharing a batch and model state. BSC reduces repeated positions while preserving uniform marginals. CTP combines consensus with only the prespecified candidate's tangential disagreement under a frozen Adam-derived metric for one joint GF update.

For a predetermined integer $n \geq 1$ and fixed b, s, we independently sample n complete stochastic realizations and define

$$g^{(k)} = g_{I_k}(\xi_k; b, s), \qquad m_n = \frac{1}{n}\sum_{k=1}^{n} g^{(k)}, \tag{8}$$

where $g^{(k)}$ is the joint G and F gradient of realization k. m_n is the empirical consensus used for subsequent direction construction. Independence applies across realizations conditional on b, s, random factors within each realization need not be independent. To analyze finite-sample estimation error, we assume the stochastic gradient has a finite second moment given b, s and define its conditional covariance as

$$\Sigma = \mathbb{E}\Big[(g^{(1)} - \bar{\mu})(g^{(1)} - \bar{\mu})^{\mathsf{T}} \,\Big|\, b, s\Big] \tag{9}$$

By the conditional law of total covariance, we obtain

$$\Sigma = \underbrace{\frac{1}{L}\sum_{i=0}^{L-1} \mathbb{E}_{\xi\sim\nu_i}\left[(g_i - \mu_i)(g_i - \mu_i)^{\mathsf{T}}\right]}_{\Sigma_{\text{within}}} + \underbrace{\frac{1}{L}\sum_{i=0}^{L-1}(\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^{\mathsf{T}}}_{\Sigma_{\text{between}}}, \tag{10}$$

where Σ_{within} captures gradient variability conditional on bridge position, while Σ_{between} captures variation in conditional mean gradients across positions.

Proposition 1 (Conditional Error Reduction via Independent Ensembling)

Fix b, s. Suppose the views are conditionally independent and identically distributed under the sampling law defined above, with $\mathbb{E}[\|g^{(1)}\|_2^2 \mid b, s] < \infty$. Then the CGE consensus satisfies

$$\mathbb{E}[m_n \mid b, s] = \bar{\mu}(b, s), \qquad \mathrm{Cov}(m_n \mid b, s) = \frac{1}{n}\Sigma, \tag{11}$$

$$\mathbb{E}\big[\|m_n - \bar{\mu}\|_2^2 \,\big|\, b, s\big] = \frac{\mathrm{tr}(\Sigma_{\text{within}}) + \mathrm{tr}(\Sigma_{\text{between}})}{n}. \tag{12}$$

Therefore, under fixed b, s and the same sampling law, the independent n-view consensus has the same conditional expectation as a single observation, and its conditional mean squared estimation error is $\frac{1}{n}$ that of a single observation. The reduction is strict when $n > 1$ and $\mathrm{tr}(\Sigma) > 0$. See Appendix A.1 for the proof.

4.2 BRIDGE STRATIFIED COUPLING

Independent ensembling reduces conditional estimation error, but sampling with replacement may repeat bridge positions. Since variation in conditional mean gradients across positions also contributes to this error, BSC coordinates position sampling at a fixed view budget. Each view's uniform marginal and the conditional estimation target remain unchanged. Fix b, s, an integer $L \geq 2$ and a predetermined view count $n \in \{1, \ldots, L\}$. Treating each bridge position as a stratum, BSC uniformly samples n ordered positions without replacement. For $(i_1, \ldots, i_n) \in \{0, \ldots, L-1\}^n$, the joint distribution is

$$\mathbb{P}_{\mathrm{BSC}}(I_1 = i_1, \ldots, I_n = i_n \,|\, b, s) = \begin{cases} \dfrac{(L-n)!}{L!}, & i_k \neq i_r, \quad \forall\, 1 \leq k < r \leq n, \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

This distribution can take the first n entries of a uniformly random permutation of all bridge positions. Each view k has marginal $\mathbb{P}_{\mathrm{BSC}}(I_k = i \mid b, s) = \frac{1}{L}, i \in \{0, \ldots, L-1\}$. Given all sampled positions, each view independently draws $\xi_k \sim \nu_{I_k}(\cdot \mid b, s)$, with its conditional distribution depending only on its own position.

Proposition 2 (Conditional Error Reduction through Uniform Sampling without Replacement) Fix b, s, integers $L \geq 2$ and $1 \leq n \leq L$. Under BSC and the conditional sampling rule above, assume the joint gradient has a finite conditional second moment. Then m_n is conditionally unbiased for $\bar{\mu}(b, s)$ and

$$\mathrm{Cov}_{\mathrm{BSC}}(m_n \mid b, s) = \frac{1}{n}\Sigma_{\mathrm{within}} + \frac{L-n}{n(L-1)}\Sigma_{\mathrm{between}}. \tag{14}$$

IID denotes independent sampling with replacement. Under the same b, s, n and within-position distributions,

$$\mathrm{Cov}_{\mathrm{IID}}(m_n \mid b, s) - \mathrm{Cov}_{\mathrm{BSC}}(m_n \mid b, s) = \frac{n-1}{n(L-1)}\Sigma_{\mathrm{between}} \succeq 0. \tag{15}$$

By conditional unbiasedness, taking traces gives the corresponding reduction in conditional mean squared error, which is strict when $n > 1$ and $\mathrm{tr}(\Sigma_{\mathrm{between}}) > 0$. See Appendix A.3 for the proof.

4.3 CONSENSUS TANGENT-SPACE PROJECTION

CGE and BSC improve conditional mean gradient estimation. CTP minimally modifies a candidate gradient to match the consensus's axial component in a frozen Adam-derived metric. Fix (b, s) and prespecify the first view in the random generation order as the candidate, with deviation $d_n = g^{(1)} - m_n$. Define the frozen positive diagonal transformation $A_s = \mathrm{diag}(a_{s,1}, \ldots, a_{s,p})$ using the pre-update Adam second-moment buffers of G and F, where

$$a_{s,j} = \begin{cases} \dfrac{1}{\sqrt{v_{s,j}} + \epsilon_{\mathrm{opt},j}}, & j \in \mathcal{B}_s, \\ 1, & j \notin \mathcal{B}_s, \end{cases} \qquad j = 1, \ldots, p, \tag{16}$$

where $\mathcal{B}_s$ indexes coordinates with initialized buffers, $v_{s,j} \geq 0$ is the pre-update gradient second-moment estimate without bias correction and $\epsilon_{\mathrm{opt},j} > 0$ belongs to the corresponding optimizer group. During view evaluation and direction construction, A_s stays fixed, inducing $\langle a, b \rangle_s = (A_s a)^{\mathsf{T}}(A_s b)$. Set $u = A_s m_n$ and $w = A_s d_n$. When $u \neq 0$, it defines the consensus axis. Otherwise, radial filtering is skipped. Define $P_{\parallel}(u) = uu^{\mathsf{T}}/\|u\|_2^2$ for $u \neq 0$ and $P_{\parallel}(0) = 0$. The disagreement decomposes as $w_{\parallel} = P_{\parallel}(u)w$ and $w_{\perp} = (I_p - P_{\parallel}(u))w$, where I_p is the identity matrix. The projection uses a single coefficient computed jointly from both the G and F parameter blocks. CTP removes $w_{\parallel}$ and combines $w_{\perp}$ with the consensus,

$$g_{\star} = m_n + A_s^{-1} w_{\perp}. \tag{17}$$

4.4 TRAINING OBJECTIVE AND OVERALL UPDATE

SSUSB retains UNSB's corrected backbone and losses, modifying only the joint GF update. After D and E updates for an unpaired batch b, we fix the reference state s and construct the frozen transformation A_s. For predetermined $n \in \{1, \ldots, L\}$, BSC samples ordered bridge positions uniformly

Table 1: Quantitative comparison of SSUSB with other unpaired image restoration methods. The top three results are highlighted as **best**, second and third.

Method	FoggyCityscapes		LowLightTrafficData		RainCityscapes		RainDS-syn		RSCityscapes		SnowTrafficData	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
CycleGAN	20.18	0.864	14.16	0.708	21.19	0.781	15.53	0.594	19.69	0.738	19.84	0.764
CUT	22.86	0.899	17.92	0.747	22.46	0.808	19.24	0.666	21.08	0.764	21.37	0.789
DCLGAN	23.41	0.907	18.64	0.759	22.93	0.817	17.81	0.624	21.63	0.774	21.92	0.798
D^4	24.56	0.925	20.81	0.791	23.68	0.829	18.73	0.637	22.16	0.786	22.64	0.809
Santa	24.13	0.918	19.37	0.776	23.42	0.826	18.46	0.633	21.94	0.782	22.41	0.806
UNSB	25.02	0.929	20.24	0.786	24.16	0.842	19.12	0.642	22.57	0.796	23.05	0.822
ODCR	25.83	0.939	21.36	0.809	24.51	0.851	19.67	0.651	22.91	0.802	23.78	0.832
DN	25.47	0.946	22.05	0.821	24.38	0.848	19.94	0.656	23.24	0.809	23.52	0.837
RCOT	26.12	0.952	22.48	0.828	24.87	0.859	20.38	0.661	23.48	0.813	24.13	0.843
DA-RCOT	26.67	0.957	23.17	0.841	25.56	0.868	21.06	0.674	23.97	0.821	24.56	0.852
DDSB	27.06	0.963	24.02	0.852	25.81	0.878	21.71	0.668	24.36	0.829	25.04	0.861
SSUSB	**28.34**	**0.977**	**25.88**	**0.873**	**26.64**	**0.892**	**22.94**	**0.685**	**25.21**	**0.842**	**26.12**	**0.876**

Table 2: Perceptual quality comparison of SSUSB with other unpaired image restoration methods. The top three values are highlighted as **best**, second, and third.

Method	FoggyCityscapes		LowLightTrafficData		RainCityscapes		RainDS-syn		RSCityscapes		SnowTrafficData	
	LPIPS ↓	NIQE ↓	LPIPS ↓	NIQE ↓	LPIPS ↓	NIQE ↓	LPIPS ↓	NIQE ↓	LPIPS ↓	NIQE ↓	LPIPS ↓	NIQE ↓
CycleGAN	0.142	4.87	0.170	5.04	0.167	4.75	0.223	5.59	0.186	5.26	0.194	4.82
CUT	0.119	4.62	0.146	4.78	0.139	4.57	0.179	5.24	0.163	5.07	0.162	4.63
DCLGAN	0.112	4.54	0.138	4.67	0.133	4.49	0.202	5.41	0.154	4.98	0.151	4.57
D^4	0.098	4.38	0.119	4.48	0.121	4.36	0.193	5.34	0.146	4.89	0.138	4.48
Santa	0.103	4.43	0.127	4.56	0.116	4.41	0.197	5.38	0.149	4.94	0.142	4.53
UNSB	0.088	4.27	0.124	4.42	0.108	4.25	0.185	5.17	0.137	4.81	0.129	4.41
ODCR	0.079	4.18	0.113	4.31	0.101	4.17	0.173	5.05	0.129	4.73	0.114	4.28
DN	0.083	4.21	0.105	4.23	0.097	4.12	0.168	4.98	0.123	4.69	0.119	4.32
RCOT	0.074	4.13	0.101	4.17	0.094	4.07	0.161	4.89	0.117	4.61	0.104	4.21
DA-RCOT	0.069	4.07	0.094	4.05	0.089	3.94	0.144	4.78	0.111	4.56	0.085	4.14
DDSB	0.062	3.97	0.086	3.96	0.083	4.02	0.152	4.69	0.103	4.49	0.079	4.07
SSUSB	**0.054**	**3.85**	**0.071**	**3.82**	**0.074**	**3.77**	**0.124**	**4.54**	**0.095**	**4.41**	**0.067**	**3.96**

without replacement and draws non-time randomness independently across views conditional on these positions. CGE averages the joint gradients of these complete realizations to obtain m_n. CTP uses the prespecified first view gradient $g^{(1)}$, removes its radial disagreement under the frozen metric, and combines the retained tangential component with m_n to form $g_\star$. We split $g_\star$ into G and F parameter blocks and pass each to its optimizer for one update. Inference retains UNSB's conditional endpoint prediction and recursive bridge-state sampling.

5 Experiments

5.1 Implementation Details

Parameter Settings and Training Strategy. We implement the model in PyTorch and run all experiments on four NVIDIA GeForce RTX 5090 GPUs with CUDA 12.8. Following UNSB's architecture and losses, we use Adam Kingma & Ba (2015). We train for 400 epochs using 256×256 images and a batch size of 1. The learning rate is 2×10^{-4} for the first 200 epochs and decays linearly to 0 over the remaining 200. Training and inference both use a five-step discrete bridge with a diffusion variance of 0.01. Unless otherwise stated, we conduct our experiments with $n = 3$ as a representative configuration. All loss weights are set to 1. Degraded and clean training images from the six datasets form a mixed degradation domain and a unified clean domain, respectively. We independently sample from both domains for unpaired joint training. Each compared method jointly trains one model across all degradations, without degradation labels or task-specific fine-tuning.

Datasets and Evaluation Metrics. We evaluate the restoration performance of SSUSB on six public datasets, including FoggyCityscapes Sakaridis et al. (2018), LowLightTrafficData Wen et al. (2025), RainCityscapes Hu et al. (2019), RainDS-syn Quan et al. (2021), RSCityscapes Wen et al. (2024) and SnowTrafficData Chen et al. (2023). We use four image quality evaluation metrics, including PSNR, SSIM Wang et al. (2004), LPIPS Zhang et al. (2018) and NIQE Mittal et al. (2013).

5.2 Comparison with Existing Methods

To comprehensively evaluate SSUSB, we compare it with eleven representative methods. As shown in Tables 1 and 2, SSUSB achieves the best results on all four metrics. Compared with UNSB,

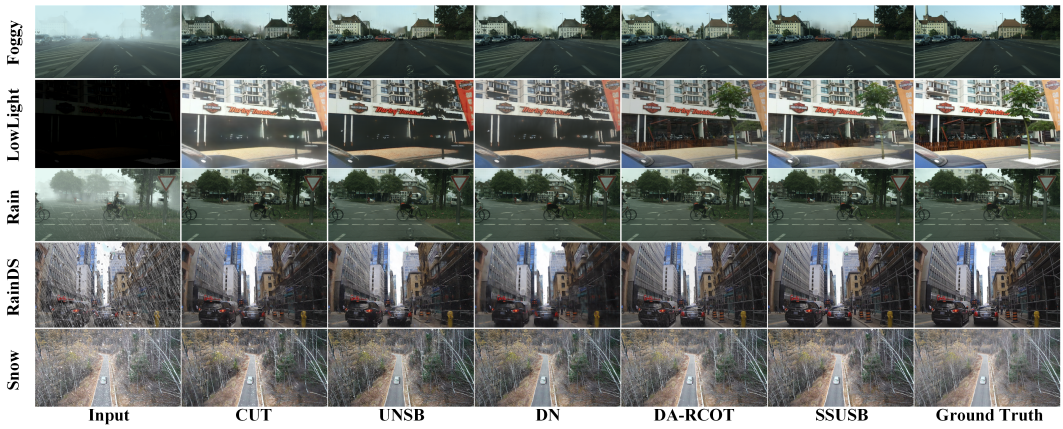


Figure 3: Qualitative comparison of all-in-one image restoration results across multiple datasets.

Table 3: Cross-dataset comparison on three datasets. Inference time is measured on 256×256 images using one GPU.

Method	RealRain-1k-L		NH-HAZE		LOL-v1		Overhead	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	Para.(M)	Time(ms)
CycleGAN	17.85	0.784	12.76	0.652	16.49	0.738	11.38	9.45
CUT	18.63	0.799	13.48	0.675	17.12	0.761	11.38	9.42
DCLGAN	18.91	0.806	13.71	0.683	17.46	0.779	11.38	9.47
D^4	19.27	0.817	14.62	0.712	17.85	0.793	10.70	23.84
Santa	19.16	0.821	14.29	0.698	17.63	0.785	11.43	88.26
UNSB	19.74	0.827	14.86	0.721	18.07	0.801	14.06	81.53
ODCR	20.34	0.842	15.58	0.734	18.42	0.813	11.38	9.74
DN	20.18	0.838	15.41	0.740	18.76	0.822	11.40	84.32
RCOT	20.73	0.850	15.93	0.747	19.11	0.835	14.20	746.21
DA-RCOT	21.61	0.865	16.34	0.756	19.98	0.844	49.80	186.51
DDSB	21.24	0.858	16.81	0.763	19.74	0.853	14.68	80.98
SSUSB	**22.49**	**0.876**	**17.69**	**0.771**	**20.55**	**0.862**	14.06	81.48

Table 4: Impact of views n.

n	PSNR↑	SSIM↑	Relative Time↓
1	23.06	0.814	1.00×
2	23.48	0.820	1.95×
3	23.97	0.827	2.91×
4	24.23	0.832	3.85×
5	24.51	0.838	4.79×

Table 5: Ablation on the projection scope of CTP.

Setting	PSNR↑	SSIM↑
G	25.05	0.844
F	24.82	0.841
G/F	25.49	0.851
GF	**25.86**	**0.858**

SSUSB improves PSNR by up to 5.64 dB with unchanged architecture and losses, supporting our joint gradient update strategy. SSUSB reduces LPIPS by up to 0.028 and 0.023 compared with DDSB and DA-RCOT, respectively, indicating improved perceptual quality. As shown in Figure 3, UNSB shows limited detail recovery in dark regions, while DA-RCOT leaves raindrops and artifacts. SSUSB better suppresses degradations while preserving details and natural colors.

5.3 Cross-Dataset Generalization Analysis

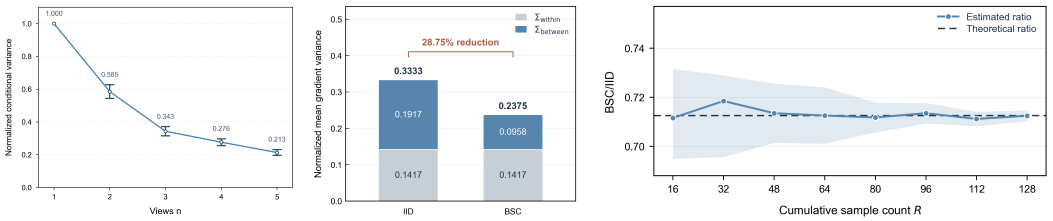
We evaluate generalization on three unseen datasets, including RealRain-1k-L Li et al. (2022b), NH-HAZE Ancuti et al. (2020) and LOL-v1 Wei et al. (2018). All methods use the weights from the main experiments without retraining or fine-tuning. As shown in Table 3, SSUSB achieves the best PSNR and SSIM on all three datasets. It improves PSNR over UNSB by up to 2.83 dB with the same parameter count and comparable inference time. Compared with DA-RCOT, it improves SSIM by up to 0.018 while reducing inference time by 105.03 ms per image.

5.4 Ablation Experiments

We conduct ablation experiments in this section. The results of all ablation experiments are the average metrics across the six datasets.

Effect of the Number of Views. Table 4 evaluates the CGE view count n with BSC and CTP disabled and update counts fixed. Unlike UNSB, the $n = 1$ baseline resamples after the discriminator and energy-network updates. Three views improve average PSNR by 0.91 dB over one. Moving from three to five views adds 0.54 dB but increases training time by 64.6%.

Conditional Variance Reduction Effect of CGE. To assess CGE's gradient estimation stability, we sample 32 independent groups of stochastic realizations for each view count at a fixed model state and batch. We estimate conditional variance using the sample covariance trace of the mean joint gradients across groups, normalize it by the single-view result, and repeat the procedure with five different sampling seeds. In Figure 4(a), normalized variance decreases by 65.7% from 1.000 at



(a) (b) (c)

Figure 4: (a) CGE variance versus view count. (b) Variance decomposition under IID and BSC. (c) Estimated and theoretical BSC/IID variance ratios.

Table 6: Ablation on the components and projection metric of CTP.

m	$w_{\parallel}$	$w_{\perp}$	Projection Metric	PSNR↑	SSIM↑
✓	—	—	—	24.65	0.839
✓	✓	—	Adam	24.36	0.833
✓	—	✓	Adam	**25.86**	**0.858**
✓	—	✓	Euclidean	25.37	0.849
✓	✓	✓	—	23.06	0.814

Table 7: Progressive ablation of the SSUSB components.

Setting	PSNR↑	SSIM↑
UNSB	22.36	0.803
Baseline	23.06	0.814
CGE	23.97	0.827
CGE + BSC	24.65	0.839
CGE + BSC + CTP	**25.86**	**0.858**

$n = 1$ to 0.343 at $n = 3$ and further falls to 0.213 at $n = 5$. These results indicate that independent gradient ensembling improves consensus estimation stability.

Gradient Variance and Decomposition of BSC. With the model state and batch fixed, we compare IID and BSC at $L = 5$ and $n = 3$, normalizing variance by the single-view result. The theoretical decomposition in Figure 4(b) leaves the variance contribution within positions unchanged and halves that between positions, reducing normalized total variance by 28.75% from 0.3333 to 0.2375. In Figure 4(c), the BSC/IID variance ratio remains near the predicted 0.7125 as R increases from 16 to 128, reaching 0.71246 at $R = 128$. These results support BSC's reduction of gradient fluctuations caused by differences between bridge positions at the same view budget.

Disagreement Component Selection and Projection Metric in CTP. Table 6 shows that CTP improves average PSNR and SSIM by 1.21 dB and 0.019 over the consensus alone. Retaining both disagreement components recovers the candidate gradient, CTP improves the corresponding scores by 2.80 dB and 0.044. These results support the selected component configuration. The frozen Adam-derived metric further improves PSNR and SSIM by 0.49 dB and 0.009 over the Euclidean metric.

Effect of Joint GF Projection in CTP. With CGE and BSC fixed, we examine the effects of the scope of CTP and the organization of projection on restoration performance. As shown in Table 5, projecting the G and F gradients separately outperforms projecting either alone. Joint GF projection further improves PSNR by 0.37 dB and SSIM by 0.007 over separate projection, supporting CTP's use of the joint gradient geometry.

Contributions of the Components. Table 7 evaluates CGE, BSC, and CTP added successively to a baseline using single-view resampling after discriminator and energy network updates. CGE with three views improves restoration, BSC adds gains at the same view budget, and CTP further improves performance by removing radial disagreement while retaining tangential components. Full SSUSB improves average PSNR and SSIM by 2.80 dB and 0.044 over the baseline.

6 Conclusion

In this work, we propose SSUSB to improve joint updates for unpaired all-in-one weather restoration using multiple gradient observations at the same model state. CGE and BSC build gradient consensus and improve bridge-position coverage, while CTP removes radial disagreement and retains tangential components under a frozen Adam-derived metric. Experiments demonstrate improved restoration across multiple degradations and the tested unseen datasets without changing the base architecture, losses or inference process. These results support improving unpaired bridge restoration through stochastic gradient organization. Future work will explore adaptive view allocation and efficient gradient estimation to balance restoration quality and training cost.

AI USE STATEMENT

In this work, we used generative AI tools to improve the clarity, grammar, and readability of the manuscript, assist with identifying and fixing bugs in our research code, and support literature searches by suggesting relevant papers. The authors reviewed all AI-assisted contributions. We checked the revised text to ensure that it accurately reflected our intended scientific meaning, reviewed and tested the suggested code corrections, and verified the references and associated descriptions against the original publications. The authors take full responsibility for the final manuscript, implementation, and reported results, including all content produced with the assistance of generative AI tools.

ETHICS STATEMENT

This work advances unpaired all-in-one weather image restoration, with potential benefits for outdoor visual perception and intelligent transportation under adverse weather conditions. As with other image generation and editing technologies, the method could potentially be misused to produce misleading visual content. We encourage responsible use and appropriate safeguards to prevent malicious applications.

REPRODUCIBILITY STATEMENT

To support reproducibility, Section 4 describes the components of SSUSB and the overall training update. Section 5.1 specifies the datasets, evaluation metrics, training hyperparameters, and computational hardware. Sections 5.3 and 5.4 describe the cross-dataset evaluation and ablation settings, respectively. The assumptions and derivations supporting our theoretical results are provided in Section 4 and Appendix A. Additional quantitative and qualitative results are presented in Appendix B. Our source code is provided in the supplementary materials.

REFERENCES

Codruta O. Ancuti, Cosmin Ancuti, and Radu Timofte. NH-HAZE: An Image Dehazing Benchmark with Non-Homogeneous Hazy and Haze-Free Images. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1798–1805, 2020. doi: 10.1109/cvprw50498.2020.00230. URL https://doi.org/10.1109/cvprw50498.2020.00230.

Léon Bottou, Frank E. Curtis, and Jorge Nocedal. Optimization Methods for Large-Scale Machine Learning. *SIAM Review*, 60(2):223–311, 2018. doi: 10.1137/16m1080173. URL https://doi.org/10.1137/16m1080173.

Haoyu Chen, Jingjing Ren, Jinjin Gu, Hongtao Wu, Xuequan Lu, Haoming Cai, and Lei Zhu. Snow Removal in Video: A New Dataset and A Novel Method. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13165–13176, 2023. doi: 10.1109/iccv51070.2023.01215. URL https://doi.org/10.1109/iccv51070.2023.01215.

Zihan Cheng, Liangtai Zhou, Dian Chen, Ni Tang, Xiaotong Luo, and Yanyun Qu. UniLDiff: Unlocking the Power of Diffusion Priors for All-in-One Image Restoration. *arXiv preprint arXiv:2507.23685*, 2025. doi: 10.48550/arXiv.2507.23685. URL https://arxiv.org/abs/2507.23685.

Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. AdaIR: Adaptive All-in-One Image Restoration via Frequency Mining and Modulation. In *International Conference on Learning Representations*, pp. 101306–101327, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/fb44a668c2d4bc984e9d6ca261262cbb-Paper-Conference.pdf.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling. In *Advances in Neural Information Processing Systems*, volume 34, pp. 17695–17709. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/940392f5f32a7ade1cc201767cf83e31-Paper.pdf.

Wenxuan Fang, Jiangwei Weng, Jianjun Qian, Jian Yang, and Jun Li. WeatherCycle: Unpaired Multi-Weather Restoration via Color Space Decoupled Cycle Learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 36(5):6705–6717, 2026. doi: 10.1109/tcsvt.2026.3651329. URL https://doi.org/10.1109/tcsvt.2026.3651329.

Junlin Han, Mehrdad Shoeiby, Lars Petersson, and Mohammad Ali Armin. Dual Contrastive Learning for Unsupervised Image-to-Image Translation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 746–755, 2021. doi: 10.1109/cvprw53098.2021.00084. URL https://doi.org/10.1109/cvprw53098.2021.00084.

Ming-Yang Ho, Che-Ming Wu, Min-Sheng Wu, and Yufeng Jane Tseng. Every Pixel Has Its Moments: Ultra-High-Resolution Unpaired Image-to-Image Translation via Dense Normalization. In *Computer Vision – ECCV 2024*, volume 15103 of *Lecture Notes in Computer Science*, pp. 312–328. Springer Nature Switzerland, 2025. doi: 10.1007/978-3-031-72995-9_18. URL https://doi.org/10.1007/978-3-031-72995-9_18.

Xiaowei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. Depth-Attentional Features for Single-Image Rain Removal. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8014–8023, 2019. doi: 10.1109/cvpr.2019.00821. URL https://doi.org/10.1109/cvpr.2019.00821.

Rie Johnson and Tong Zhang. Accelerating Stochastic Gradient Descent using Predictive Variance Reduction. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/file/ac1dd209cbcc5e5d1c6e28598e8cbbe8-Paper.pdf.

Beomsu Kim, Gihyun Kwon, Kwanyoung Kim, and Jong Chul Ye. Unpaired Image-to-Image Translation via Neural Schrödinger Bridge. In *International Conference on Learning Representations*, pp. 19312–19331, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/5491280797f3192b895bce84eb83df8d-Paper-Conference.pdf.

Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*, 2015. URL https://arxiv.org/abs/1412.6980.

Venkata Narendra Kotyada, Revanth Eranki, and Nagesh Bhattu Sristy. Contrastive-SDE: Guiding Stochastic Differential Equations with Contrastive Learning for Unpaired Image-to-Image Translation. *arXiv preprint arXiv:2510.03821*, 2025. doi: 10.48550/arXiv.2510.03821. URL https://arxiv.org/abs/2510.03821.

Yunwei Lan, Zhigao Cui, Xin Luo, Chang Liu, Nian Wang, Menglin Zhang, Yanzhao Su, and Dong Liu. When Schrödinger Bridge Meets Real-World Image Dehazing with Unpaired Training. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8756–8765, 2025. doi: 10.1109/iccv51701.2025.00819. URL https://doi.org/10.1109/iccv51701.2025.00819.

Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014. doi: 10.3934/dcds.2014.34.1533. URL https://doi.org/10.3934/dcds.2014.34.1533.

Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-In-One Image Restoration for Unknown Corruption. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17431–17441, 2022a. doi: 10.1109/cvpr52688.2022.01693. URL https://doi.org/10.1109/cvpr52688.2022.01693.

Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3172–3182. IEEE, 2020.

Wei Li, Qiming Zhang, Jing Zhang, Zhen Huang, Xinmei Tian, and Dacheng Tao. Toward Real-world Single Image Deraining: A New Benchmark and Beyond. *arXiv preprint arXiv:2206.05514*, 2022b. doi: 10.48550/arXiv.2206.05514. URL https://arxiv.org/abs/2206.05514.

Ziyue Lin, Jiahe Hou, Hongyu Xia, Xinrui Xie, Feifei Wang, Yuyin Zhou, Wei Wang, Jiawei Liu, and Liangqiong Qu. Decoupled Residual Denoising Diffusion Models for Unified and Data Efficient Image-to-Image Translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 35967–35977, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Lin_Decoupled_Residual_Denoising_Diffusion_Models_for_Unified_and_Data_Efficient_CVPR_2026_paper.html.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-Averse Gradient Descent for Multi-task learning. In *Advances in Neural Information Processing Systems*, volume 34, pp. 18878–18890. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/9d27fdf2477ffbff837d73ef7ae23db9-Paper.pdf.

Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos Theodorou, Weili Nie, and Anima Anandkumar. I^2SB: Image-to-Image Schrödinger Bridge. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 22042–22062. PMLR, 2023. URL https://proceedings.mlr.press/v202/liu23ai.html.

Xiaohui Liu, Zhilu Zhang, Xiaohe Wu, Chaoyu Feng, Xiaotao Wang, Lei Lei, and Wangmeng Zuo. Learning Real-World Image De-weathering with Imperfect Supervision. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 3738–3746, 2024. doi: 10.1609/aaai.v38i4.28164. URL https://doi.org/10.1609/aaai.v38i4.28164.

Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013. doi: 10.1109/lsp.2012.2227726. URL https://doi.org/10.1109/lsp.2012.2227726.

Maxence Noble, Marie Scheid, Yazid Janati, Eric Moulines, and Alain Durmus. Twisted Schrödinger Bridge Matching. *arXiv preprint arXiv:2607.16987*, 2026. doi: 10.48550/arXiv.2607.16987. URL https://arxiv.org/abs/2607.16987.

Taesung Park, Alexei A. Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive Learning for Unpaired Image-to-Image Translation. In *Computer Vision – ECCV 2020*, pp. 319–345. Springer International Publishing, 2020. doi: 10.1007/978-3-030-58545-7_19. URL https://doi.org/10.1007/978-3-030-58545-7_19.

Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in neural information processing systems*, 36:71275–71293, 2023.

Ruijie Quan, Xin Yu, Yuanzhi Liang, and Yi Yang. Removing Raindrops and Rain Streaks in One Go. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9143–9152, 2021. doi: 10.1109/cvpr46437.2021.00903. URL https://doi.org/10.1109/cvpr46437.2021.00903.

Sudarshan Rajagopalan and Vishal M. Patel. AWRaCLe: All-Weather Image Restoration Using Visual In-Context Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 6675–6683, 2025. doi: 10.1609/aaai.v39i6.32716. URL https://doi.org/10.1609/aaai.v39i6.32716.

Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic Foggy Scene Understanding with Synthetic Data. *International Journal of Computer Vision*, 126(9):973–992, 2018. doi: 10.1007/s11263-018-1072-8. URL https://doi.org/10.1007/s11263-018-1072-8.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion Schrödinger Bridge Matching. In *Advances in Neural Information Processing Systems*, volume 36, pp. 62183–62223. Curran Associates, Inc., 2023. doi: 10.52202/075280-2717. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/c428adf74782c2092d254329b6b02482-Paper-Conference.pdf.

Ni Tang, Xiaotong Luo, Zihan Cheng, Liangtai Zhou, Dongxiao Zhang, and Yanyun Qu. Diffusion Once and Done: Degradation-Aware LoRA for All-in-One Image Restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 9448–9456, 2026. doi: 10.1609/aaai.v40i11.37905. URL https://doi.org/10.1609/aaai.v40i11.37905.

Xiaole Tang, Xin Hu, Xiang Gu, and Jian Sun. Residual-Conditioned Optimal Transport: Towards Structure-Preserving Unpaired and Paired Image Restoration. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 47757–47777. PMLR, 2024. URL https://proceedings.mlr.press/v235/tang24d.html.

Xiaole Tang, Xiang Gu, Xiaoyi He, Xin Hu, and Jian Sun. Degradation-Aware Residual-Conditioned Optimal Transport for Unified Image Restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(8):6764–6779, 2025. doi: 10.1109/tpami.2025.3562211. URL https://doi.org/10.1109/tpami.2025.3562211.

Panagiotis Theodoropoulos, Nikolaos Komianos, Vincent Pacelli, Guan-Horng Liu, and Evangelos Theodorou. Feedback Schrödinger Bridge Matching. In *International Conference on Learning Representations*, pp. 58939–58967, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/94398c15080cca93180416a52989949a-Paper-Conference.pdf.

Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M. Patel. TransWeather: Transformer-Based Restoration of Images Degraded by Adverse Weather Conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2353–2363, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Valanarasu_TransWeather_Transformer-Based_Restoration_of_Images_Degraded_by_Adverse_Weather_Conditions_CVPR_2022_paper.html.

Zhongze Wang, Haitao Zhao, Jingchao Peng, Lujian Yao, and Kaijie Zhao. ODCR: Orthogonal Decoupling Contrastive Regularization for Unpaired Image Dehazing. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25479–25489, 2024. doi: 10.1109/cvpr52733.2024.02407. URL https://doi.org/10.1109/cvpr52733.2024.02407.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. doi: 10.1109/tip.2003.819861. URL https://doi.org/10.1109/tip.2003.819861.

Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep Retinex Decomposition for Low-Light Enhancement. In *British Machine Vision Conference*, 2018. URL https://daooshee.github.io/BMVC2018website/.

Yuanbo Wen, Tao Gao, Kaihao Zhang, Peng Cheng, and Ting Chen. Restoring vision in rain-by-snow weather with simple attention-based sampling cross-hierarchy Transformer. *Pattern Recognition*, 156:110743, 2024. doi: 10.1016/j.patcog.2024.110743. URL https://doi.org/10.1016/j.patcog.2024.110743.

Yuanbo Wen, Tao Gao, Jing Zhang, Ziqi Li, and Ting Chen. Multi-axis Prompt and Multi-dimension Fusion Network for All-in-one Weather-degraded Image Restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 8323–8331, 2025. doi: 10.1609/aaai.v39i8.32898. URL https://doi.org/10.1609/aaai.v39i8.32898.

Shaoan Xie, Yanwu Xu, Mingming Gong, and Kun Zhang. Unpaired Image-to-Image Translation with Shortest Path Regularization. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10177–10187, 2023. doi: 10.1109/cvpr52729.2023.00981. URL https://doi.org/10.1109/cvpr52729.2023.00981.

Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiaojie Guo, and Dacheng Tao. Self-augmented Unpaired Image Dehazing via Density and Depth Decomposition. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2027–2036, 2022. doi: 10.1109/cvpr52688.2022.00208. URL https://doi.org/10.1109/cvpr52688.2022.00208.

Jingjun Yi, Qi Bi, Hao Zheng, Huimin Huang, Yixian Shen, Haolan Zhan, Wei Ji, Yawen Huang, Yuexiang Li, Xian Wu, and Yefeng Zheng. Degradation-Aware Dynamic Schrödinger Bridge for Unpaired Image Restoration. In *Advances in Neural Information Processing Systems*, volume 38, pp. 2415–2439. Curran Associates, Inc., 2025. doi: 10.52202/085713-0086. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/039c30e9af8039fbd1b58da9d04f38e9-Paper-Conference.pdf.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient Surgery for Multi-Task Learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 5824–5836. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/3fe78a8acf5fda99de95303940a2420c-Paper.pdf.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018. doi: 10.1109/cvpr.2018.00068. URL https://doi.org/10.1109/cvpr.2018.00068.

Peilin Zhao and Tong Zhang. Accelerating Minibatch Stochastic Gradient Descent using Stratified Sampling. *arXiv preprint arXiv:1405.3080*, 2014. doi: 10.48550/arXiv.1405.3080. URL https://arxiv.org/abs/1405.3080.

Dian Zheng, Xiao-Ming Wu, Shuzhou Yang, Jian Zhang, Jian-Fang Hu, and Wei-Shi Zheng. Selective Hourglass Mapping for Universal Image Restoration Based on Diffusion Model. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25445–25455, 2024. doi: 10.1109/cvpr52733.2024.02404. URL https://doi.org/10.1109/cvpr52733.2024.02404.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, 2017. doi: 10.1109/iccv.2017.244. URL https://doi.org/10.1109/iccv.2017.244.

Zhidong Zhu, Bangshu Xiong, Shuzhen Yu, Jinhao Zhu, Zhibo Rao, Qiaofeng Ou, Xing Li, and Wei Li. FACT: Frequency adaptive consistency tuning for efficient zero-shot unified image restoration. *Neurocomputing*, 695:134056, 2026. doi: 10.1016/j.neucom.2026.134056. URL https://doi.org/10.1016/j.neucom.2026.134056.

A PROOFS

A.1 PROOF OF PROPOSITION 1

Fix b, s and omit this conditioning from all expectations and covariances below. The finite second moment assumption ensures their existence. By uniform sampling of bridge positions and the law of total expectation, we have

$$\begin{aligned}\mathbb{E}[g^{(k)}] &= \sum_{i=0}^{L-1} \mathbb{P}(I_k = i \mid b, s)\, \mathbb{E}[g^{(k)} \mid I_k = i, b, s] \\ &= \frac{1}{L}\sum_{i=0}^{L-1} \mu_i = \bar{\mu}.\end{aligned} \tag{18}$$

Therefore,

$$\mathbb{E}[m_n] = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[g^{(k)}] = \bar{\mu}. \tag{19}$$

For each bridge position i, we have

$$g_i - \bar{\mu} = (g_i - \mu_i) + (\mu_i - \bar{\mu}). \tag{20}$$

Since

$$\mathbb{E}_{\xi\sim\nu_i}[g_i - \mu_i] = 0, \tag{21}$$

both cross terms obtained by expanding the outer product have zero expectation. Therefore,

$$\begin{aligned}\Sigma &= \frac{1}{L}\sum_{i=0}^{L-1} \mathbb{E}_{\xi\sim\nu_i}\big[(g_i - \bar{\mu})(g_i - \bar{\mu})^{\mathsf{T}}\big] \\ &= \frac{1}{L}\sum_{i=0}^{L-1} \mathbb{E}_{\xi\sim\nu_i}\big[(g_i - \mu_i)(g_i - \mu_i)^{\mathsf{T}}\big] + \frac{1}{L}\sum_{i=0}^{L-1}(\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^{\mathsf{T}} \\ &= \Sigma_{\text{within}} + \Sigma_{\text{between}}.\end{aligned} \tag{22}$$

Let $e_k = g^{(k)} - \bar{\mu}$. Given b, s, the e_k are independent and have zero mean. Therefore, for $k \neq r$,

$$\mathbb{E}[e_k e_r^{\mathsf{T}}] = \mathbb{E}[e_k]\mathbb{E}[e_r]^{\mathsf{T}} = 0. \tag{23}$$

Combining this with the fact that the views are identically distributed,

$$\begin{aligned}\mathrm{Cov}(m_n) &= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{r=1}^{n} \mathbb{E}[e_k e_r^{\mathsf{T}}] \\ &= \frac{1}{n^2}\sum_{k=1}^{n} \Sigma \\ &= \frac{\Sigma}{n}.\end{aligned} \tag{24}$$

By the unbiasedness in Eq. (19),

$$\begin{aligned}\mathbb{E}\big[\|m_n - \bar{\mu}\|_2^2\big] &= \mathrm{tr}(\mathrm{Cov}(m_n)) \\ &= \frac{\mathrm{tr}(\Sigma)}{n} \\ &= \frac{\mathrm{tr}(\Sigma_{\text{within}}) + \mathrm{tr}(\Sigma_{\text{between}})}{n}.\end{aligned} \tag{25}$$

The mean squared error of a single observation is $\mathrm{tr}(\Sigma)$. Thus, when $n > 1$ and this value is positive, the mean squared error after ensembling is strictly reduced.

A.2 Relationship Between the Conditional Objective and the Loss Gradient

Fix b, s, for each bridge position and stochastic realization, freeze the outputs of the base model's stop-gradient branches at their values evaluated at Θ_s, including recursively constructed bridge states and key features in the contrastive loss. Use Υ as the differentiation variable for the remaining active branches to define the local stochastic loss $\ell_{GF}(\Upsilon; b, i, \xi, s)$. Under the frozen sampling distribution, define the local conditional objective as

$$\mathcal{J}_{b,s}(\Upsilon) = \frac{1}{L}\sum_{i=0}^{L-1} \mathbb{E}_{\xi\sim\nu_i(\cdot|b,s)}[\ell_{GF}(\Upsilon; b, i, \xi, s)]\,. \tag{26}$$

Neither the sampling distribution nor the above frozen quantities change with Υ. For each position i, assume that the loss at the reference parameters is integrable

$$\mathbb{E}_{\xi\sim\nu_i}[|\ell_{GF}(\Theta_s; b, i, \xi, s)|] < \infty. \tag{27}$$

Further assume that there exists an open neighborhood of Θ_s, U and a nonnegative integrable function $H_i(\xi)$ such that, for ν_i-distributed ξ, with probability one the loss at Θ_s is differentiable with respect to Υ and for all $\Upsilon, \Upsilon' \in U$,

$$|\ell_{GF}(\Upsilon; b, i, \xi, s) - \ell_{GF}(\Upsilon'; b, i, \xi, s)| \le H_i(\xi)\|\Upsilon - \Upsilon'\|_2, \qquad \mathbb{E}_{\xi\sim\nu_i}[H_i(\xi)] < \infty. \tag{28}$$

Eq. (28) ensures that the objective is finite near the reference parameters and bounds the difference quotients of the loss. The gradient norm at the reference parameters does not exceed $H_i(\xi)$, so the remainder of the first-order expansion, divided by the norm of the parameter increment, is bounded by $2H_i(\xi)$. By differentiability and the dominated convergence theorem, the expected remainder is of smaller order than the norm of the parameter increment. Therefore, $\mathcal{J}_{b,s}$ is differentiable at Θ_s and

$$\nabla_\Upsilon \mathcal{J}_{b,s}(\Upsilon)|_{\Upsilon=\Theta_s} = \frac{1}{L}\sum_{i=0}^{L-1} \mathbb{E}_{\xi\sim\nu_i}[g_i(\xi; b, s)] = \bar{\mu}(b, s). \tag{29}$$

Therefore, CGE estimates the joint gradient of a local conditional objective consistent with the stop-gradient rules of the base model, rather than computing total derivatives through the generator-dependent sampling distribution or the truncated computational branches.

A.3 Proof of Proposition 2

Fix b and s. From L positions, selecting n distinct ordered positions yields $\frac{L!}{(L-n)!}$ possible outcomes, to which BSC assigns equal probability. Fix any view k and position i, the number of ordered outcomes satisfying $I_k = i$ is $\frac{(L-1)!}{(L-n)!}$. Therefore,

$$\mathbb{P}_{\mathrm{BSC}}(I_k = i) = \frac{(L-1)!}{(L-n)!}\cdot\frac{(L-n)!}{L!} = \frac{1}{L}. \tag{30}$$

Combining this with the unchanged sampling distribution within each position gives

$$\begin{aligned}\mathbb{E}_{\mathrm{BSC}}[g^{(k)}] &= \sum_{i=0}^{L-1} \mathbb{P}_{\mathrm{BSC}}(I_k = i)\mathbb{E}_{\xi\sim\nu_i}[g_i] \\ &= \frac{1}{L}\sum_{i=0}^{L-1}\mu_i = \bar{\mu}.\end{aligned} \tag{31}$$

By the linearity of expectation,

$$\mathbb{E}_{\mathrm{BSC}}[m_n] = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}_{\mathrm{BSC}}[g^{(k)}] = \bar{\mu}. \tag{32}$$

Therefore, although coupling without replacement changes the dependencies between views, it does not change the conditional expectation of the empirical gradient consensus. In addition, the complete stochastic realization of each view has the same marginal distribution as under independent sampling, so

$$\mathrm{Cov}_{\mathrm{BSC}}(g^{(k)}) = \Sigma = \Sigma_{\mathrm{within}} + \Sigma_{\mathrm{between}}. \tag{33}$$

When $n \geq 2$, for any distinct views $k \neq r$ and distinct positions $i \neq j$, the number of ordered outcomes satisfying $I_k = i, I_r = j$ is $\frac{(L-2)!}{(L-n)!}$. Therefore,

$$\mathbb{P}_{\mathrm{BSC}}(I_k = i, I_r = j) = \begin{cases} \dfrac{1}{L(L-1)}, & i \neq j, \\ 0, & i = j. \end{cases} \tag{34}$$

To simplify the notation, let

$$a_i = \mu_i - \bar{\mu}, \qquad i = 0, \ldots, L-1. \tag{35}$$

By the definition of $\bar{\mu}$,

$$\sum_{i=0}^{L-1} a_i = 0, \qquad \frac{1}{L}\sum_{i=0}^{L-1} a_i a_i^{\mathsf{T}} = \Sigma_{\mathrm{between}}. \tag{36}$$

Given all bridge positions, the non-time random variables of the views are mutually independent and their respective distributions are determined only by the corresponding positions and the fixed conditions b, s. Therefore, for $i \neq j$,

$$\mathbb{E}_{\mathrm{BSC}}\Big[(g^{(k)} - \bar{\mu})(g^{(r)} - \bar{\mu})^{\mathsf{T}} \,\Big|\, I_k = i, I_r = j\Big] = (\mu_i - \bar{\mu})(\mu_j - \bar{\mu})^{\mathsf{T}} = a_i a_j^{\mathsf{T}}. \tag{37}$$

Taking the expectation over the joint distribution of bridge positions yields

$$\begin{aligned} \mathrm{Cov}_{\mathrm{BSC}}\Big(g^{(k)}, g^{(r)}\Big) &= \frac{1}{L(L-1)} \sum_{\substack{i,j=0 \\ i \neq j}}^{L-1} a_i a_j^{\mathsf{T}} \\ &= \frac{1}{L(L-1)} \left[\left(\sum_{i=0}^{L-1} a_i\right) \left(\sum_{j=0}^{L-1} a_j\right)^{\mathsf{T}} - \sum_{i=0}^{L-1} a_i a_i^{\mathsf{T}} \right] \\ &= -\frac{1}{L-1}\Sigma_{\mathrm{between}}. \end{aligned} \tag{38}$$

Substituting the covariance of an individual view and the cross-covariances between distinct views gives

$$\begin{aligned} \mathrm{Cov}_{\mathrm{BSC}}(m_n) &= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{r=1}^{n} \mathrm{Cov}_{\mathrm{BSC}}\Big(g^{(k)}, g^{(r)}\Big) \\ &= \frac{1}{n^2}\left[n\Sigma - \frac{n(n-1)}{L-1}\Sigma_{\mathrm{between}}\right] \\ &= \frac{1}{n}\Sigma_{\mathrm{within}} + \left[\frac{1}{n} - \frac{n-1}{n(L-1)}\right]\Sigma_{\mathrm{between}} \\ &= \frac{1}{n}\Sigma_{\mathrm{within}} + \frac{L-n}{n(L-1)}\Sigma_{\mathrm{between}}. \end{aligned} \tag{39}$$

For $n = 1$, there are no cross-covariance terms and Eq. (39) also holds, reducing to the covariance Σ of a single observation. According to Proposition 1,

$$\mathrm{Cov}_{\mathrm{IID}}(m_n) = \frac{1}{n}\Sigma_{\mathrm{within}} + \frac{1}{n}\Sigma_{\mathrm{between}}. \tag{40}$$

Therefore,

$$\mathrm{Cov}_{\mathrm{IID}}(m_n) - \mathrm{Cov}_{\mathrm{BSC}}(m_n) = \frac{n-1}{n(L-1)}\Sigma_{\mathrm{between}} \succeq 0, \tag{41}$$

where positive semidefiniteness follows from $\Sigma_{\mathrm{between}} \succeq 0$. By the conditional unbiasedness of both estimators,

$$\begin{aligned} \mathbb{E}_{\mathrm{BSC}}\big[\|m_n - \bar{\mu}\|_2^2\big] &= \mathrm{tr}(\mathrm{Cov}_{\mathrm{BSC}}(m_n)) \\ &= \frac{1}{n}\,\mathrm{tr}(\Sigma_{\mathrm{within}}) + \frac{L-n}{n(L-1)}\,\mathrm{tr}(\Sigma_{\mathrm{between}})\,. \end{aligned} \tag{42}$$

Subtracting Eq. (42) from the mean squared error under independent sampling gives

$$\mathbb{E}_{\mathrm{IID}}\left[\|m_n - \bar{\mu}\|_2^2\right] - \mathbb{E}_{\mathrm{BSC}}\left[\|m_n - \bar{\mu}\|_2^2\right] = \frac{n-1}{n(L-1)}\,\mathrm{tr}(\Sigma_{\mathrm{between}})\,. \tag{43}$$

Finally,

$$\mathrm{tr}(\Sigma_{\mathrm{between}}) = \frac{1}{L}\sum_{i=0}^{L-1}\|\mu_i - \bar{\mu}\|_2^2. \tag{44}$$

Therefore, the trace is positive if and only if the conditional mean gradients across bridge positions are not all identical. When this condition holds and $n > 1$, the conditional mean squared error is strictly reduced.

A.4 Position Coverage with Finitely Many Views

To characterize position coverage separately, let K_n denote the number of distinct bridge positions directly used for gradient evaluation by the n views,

$$K_n = |\{I_1, \ldots, I_n\}|\,. \tag{45}$$

Under uniform sampling without replacement, the n views select distinct bridge positions, so

$$K_n = n \qquad \text{almost surely.} \tag{46}$$

Under independent sampling with replacement, for any position i, the probability that it is not selected in any of the n samples is

$$\mathbb{P}_{\mathrm{IID}}(I_1 \neq i, \ldots, I_n \neq i) = \left(1 - \frac{1}{L}\right)^n. \tag{47}$$

Expressing the number of distinct positions covered using indicator functions,

$$K_n = \sum_{i=0}^{L-1} \mathbf{1}\{i \in \{I_1, \ldots, I_n\}\}\,, \tag{48}$$

gives

$$\mathbb{E}_{\mathrm{IID}}[K_n \mid b, s] = L\left[1 - \left(1 - \frac{1}{L}\right)^n\right]. \tag{49}$$

When $2 \leq n \leq L$, independent sampling produces repeated positions with positive probability and always satisfies $K_n \leq n$. Therefore,

$$\mathbb{E}_{\mathrm{IID}}[K_n \mid b, s] < n = \mathbb{E}_{\mathrm{BSC}}[K_n \mid b, s]. \tag{50}$$

A.5 Geometric Properties of Consensus Tangent-Space Projection

Fix b,s and a predetermined integer $n \geq 1$. Given all currently realized view gradients $g^{(1)}, \ldots, g^{(n)}$ retain the definitions in the main text

$$m_n = \frac{1}{n}\sum_{k=1}^{n} g^{(k)}, \qquad d_n = g^{(1)} - m_n, \qquad u = A_s m_n, \qquad w = A_s d_n. \tag{51}$$

Where A_s is a frozen, finite and invertible positive diagonal matrix. The following conclusions hold for the given set of gradients and do not depend on independence or exchangeability between views. When $u \neq 0$, the radial projection matrix defined in the main text satisfies

$$P_{\parallel}(u)^{\mathsf{T}} = P_{\parallel}(u), \qquad P_{\parallel}(u)^2 = P_{\parallel}(u). \tag{52}$$

Therefore, $P_{\parallel}(u)$ is the Euclidean orthogonal projection onto $\mathrm{span}(u)$ and its complementary projection $I_p - P_{\parallel}(u)$ corresponds to the linear space

$$\mathcal{T}_u = \{z \in \mathbb{R}^p : u^{\mathsf{T}} z = 0\}. \tag{53}$$

Where the tangent space refers to the orthogonal complement of the consensus in the transformed coordinates and does not involve the geometry of the level sets of the restoration loss. By the properties of orthogonal projections,

$$w = w_{\parallel} + w_{\perp}, \qquad w_{\parallel}^{\mathsf{T}} w_{\perp} = 0, \qquad u^{\mathsf{T}} w_{\perp} = 0. \tag{54}$$

According to $g^{(1)} = m_n + d_n$, we have

$$A_s g^{(1)} = u + w_{\parallel} + w_{\perp}, \qquad A_s g_\star = u + w_{\perp}. \tag{55}$$

This yields

$$\begin{aligned} u^{\mathsf{T}} A_s g_\star &= \|u\|_2^2, \\ (I_p - P_{\parallel}(u)) A_s g_\star &= (I_p - P_{\parallel}(u)) A_s g^{(1)} = w_{\perp}. \end{aligned} \tag{56}$$

Define the axial constraint set

$$\mathcal{H}_u = \left\{ z \in \mathbb{R}^p : u^{\mathsf{T}} z = \|u\|_2^2 \right\}. \tag{57}$$

When $u \neq 0$, this set is an affine hyperplane passing through the consensus point u and $\mathcal{H}_u = u + \mathcal{T}_u$. By Eq. (56), $A_s g_\star \in \mathcal{H}_u$. For any $z \in \mathcal{H}_u$, we have

$$u^{\mathsf{T}}(z - A_s g_\star) = 0. \tag{58}$$

On the other hand, by Eq. (55),

$$A_s g^{(1)} - A_s g_\star = w_{\parallel}, \tag{59}$$

and $w_{\parallel}$ is parallel to u. Therefore, $z - A_s g_\star$ is orthogonal to $w_{\parallel}$ and hence

$$\|z - A_s g^{(1)}\|_2^2 = \|z - A_s g_\star\|_2^2 + \|w_{\parallel}\|_2^2. \tag{60}$$

The right-hand side of Eq. (60) attains its minimum only when $z = A_s g_\star$. Since A_s is invertible, we obtain the unique solution

$$g_\star = \underset{h \in \mathbb{R}^p : u^{\mathsf{T}} A_s h = \|u\|_2^2}{\operatorname{argmin}} \frac{1}{2} \|A_s(h - g^{(1)})\|_2^2. \tag{61}$$

Therefore, $g_\star$ is the unique direction closest to the candidate gradient in the frozen metric under this axial constraint. When $u = 0$, the invertibility of A_s implies $m_n = 0$. Following the convention in the main text, $P_{\parallel}(u) = 0$, so $g_\star = g^{(1)}$. In this case, $\mathcal{H}_u = \mathbb{R}^p$ and Eq. (61) still holds, but there is no geometric interpretation involving a nonzero consensus axis. Finally, by Eq. (55) and orthogonality,

$$\|A_s g_\star\|_2^2 = \|u\|_2^2 + \|w_{\perp}\|_2^2. \tag{62}$$

Therefore, preserving the axial component does not imply preserving the overall direction norm or the actual Adam update step size.

B EXPERIMENT

B.1 Cross-Dataset Generalization Analysis on Perceptual Metrics

To evaluate SSUSB's generalization to unseen degradation distributions, we test on RealRain-1k-L, NH-HAZE and LOL-v1, none of which is used for training. All methods use the model weights from the main experiments without retraining or fine-tuning on the target datasets. As shown in Table 8, SSUSB achieves the lowest LPIPS and NIQE on all three unseen datasets. Compared with UNSB, SSUSB reduces LPIPS and NIQE by up to 0.062 and 0.60, respectively. Compared with DA-RCOT, the maximum reductions are 0.024 and 0.23, respectively. The results indicate that SSUSB reduces perceptual differences from reference images and improves natural image quality scores on the tested unseen datasets without target adaptation. Figure 5 presents qualitative comparisons of different methods on the three unseen datasets. CUT and UNSB leave residual rain streaks and haze and provide insufficient brightness recovery in low-light scenes. DN produces relatively dark dehazing results, making some structures in shadowed regions difficult to distinguish. Although DA-RCOT improves scene visibility, it does not fully suppress nonuniform haze. In comparison, SSUSB more effectively reduces rain and haze obscuration and improves brightness, showing better adaptability to the tested shifts in degradation distributions.

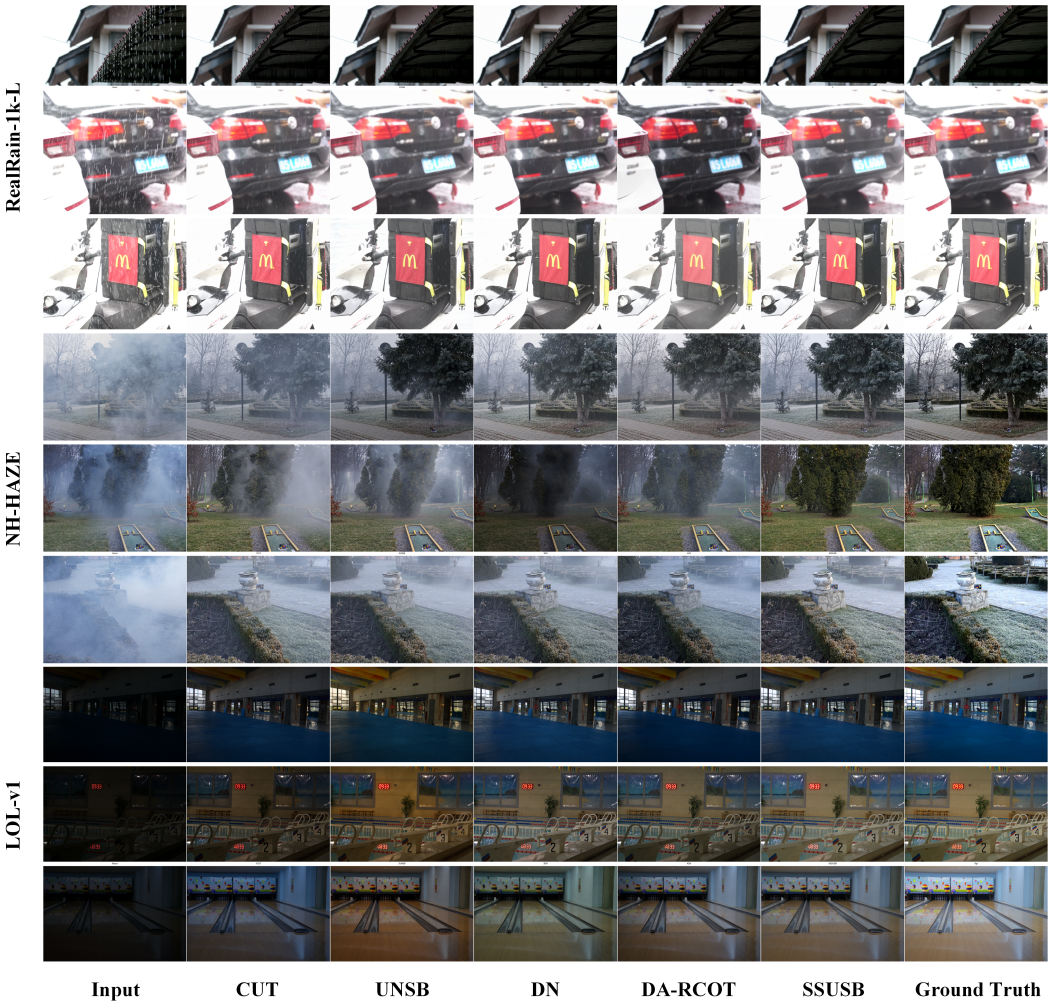


Figure 5: Qualitative comparisons on three unseen datasets.

Table 8: Cross-dataset generalization results on three datasets using two perceptual quality metrics.

Method	RealRain-1k-L		NH-HAZE		LOL-v1	
	LPIPS↓	NIQE↓	LPIPS↓	NIQE↓	LPIPS↓	NIQE↓
CycleGAN	0.201	5.38	0.265	5.82	0.207	5.55
CUT	0.191	5.25	0.247	5.69	0.197	5.41
DCLGAN	0.188	5.21	0.241	5.61	0.191	5.31
D^4	0.183	5.12	0.217	5.39	0.184	5.20
Santa	0.185	5.09	0.226	5.48	0.188	5.24
UNSB	0.176	5.02	0.210	5.33	0.180	5.13
ODCR	0.168	4.91	0.190	5.20	0.173	5.04
DN	0.171	4.96	0.195	5.14	0.166	4.94
RCOT	0.162	4.86	0.182	5.05	0.158	4.86
DA-RCOT	0.149	4.69	0.172	4.96	0.146	4.78
DDSB	0.154	4.78	0.159	4.84	0.150	4.71
SSUSB	**0.139**	**4.56**	**0.148**	**4.73**	**0.137**	**4.62**

B.2 Effect of the CGE View Count n on Perceptual Metrics

Table 9 examines how the number of views affects perceptual quality in CGE, which ensembles gradients at the same model state. With BSC and CTP disabled and a fixed update count, three views reduce average LPIPS and NIQE by 0.011 and 0.11, compared with one view. The gains indicate that integrating gradients from multiple complete stochastic realizations helps reduce perceptual differences and improve natural image quality scores. Both metrics continue to decrease as the view count increases. Increasing from three to five views further reduces LPIPS and NIQE by 0.009 and 0.09, respectively, but increases training time by 64.6%. Although five views achieve the best

Table 9: Impact of the number of views n in CGE on perceptual quality and relative training time.

n	LPIPS↓	NIQE↓	Relative Time↓
1	0.118	4.46	1.00×
2	0.114	4.41	1.95×
3	0.107	4.35	2.91×
4	0.105	4.30	3.85×
5	0.098	4.26	4.79×

Table 10: Ablation on the components and projection metric of CTP on perceptual quality.

m	$w_{\parallel}$	$w_{\perp}$	Projection Metric	LPIPS↓	NIQE↓
✓	—	—	—	0.096	4.24
✓	✓	—	Adam	0.103	4.27
✓	—	✓	Adam	**0.081**	**4.06**
✓	—	✓	Euclidean	0.088	4.15
✓	✓	✓	—	0.118	4.46

Table 11: Ablation on the projection scope of CTP on perceptual quality.

Setting	LPIPS↓	NIQE↓
G	0.092	4.19
F	0.096	4.22
G/F	0.086	4.11
GF	**0.081**	**4.06**

Table 12: Progressive ablation of the SSUSB components on perceptual quality.

Setting	LPIPS↓	NIQE↓
UNSB	0.129	4.56
Baseline	0.118	4.46
CGE	0.107	4.35
CGE + BSC	0.096	4.24
CGE + BSC + CTP	**0.081**	**4.06**

perceptual scores among the tested configurations, the additional gains come at a higher training cost. Therefore, we adopt $n = 3$ as the default view count for subsequent experiments.

B.3 Effect of CTP Disagreement Component Selection and Projection Metrics on Perceptual Metrics

As shown in Table 10, CTP removes radial disagreement while retaining tangential disagreement. Compared with retaining both components, it reduces average LPIPS by 0.037 and NIQE by 0.40. These reductions indicate that radial filtering helps reduce perceptual differences between restored and reference images and improves naturalness scores. Compared with using only the consensus mean, retaining tangential disagreement further lowers LPIPS by 0.015 and NIQE by 0.18. It indicates that discarding all disagreement is not optimal under the current experimental setting. Retaining only radial disagreement raises LPIPS by 0.007 and NIQE by 0.03 relative to using the consensus mean alone. These comparisons support an update rule that builds on consensus, removes radial disagreement and retains tangential components. With the same component configuration, the frozen Adam metric reduces LPIPS by 0.007 and NIQE by 0.09 compared with the Euclidean metric. These results favor the Adam-derived metric over Euclidean projection in this setting.

B.4 Effect of Joint GF Projection in CTP Evaluated on Perceptual Metrics

With CGE and BSC fixed, we examine how CTP's projection scope and organization affect perceptual quality. As shown in Table 11, separate projection for G and F achieves an average LPIPS of 0.086 and NIQE of 4.11, outperforming projection for either network alone. It suggests that processing both gradient blocks helps improve perceptual quality. Joint GF projection further lowers LPIPS to 0.081 and NIQE to 4.06, reducing LPIPS by 0.005 and NIQE by 0.05 compared with separate projection. Under the current experimental setting, these results suggest that constructing the projection direction using the overall geometry of both gradient blocks helps reduce perceptual differences between restored and reference images and improves naturalness scores. These findings support CTP's joint GF projection design.

B.5 Component Contributions Evaluated on Perceptual Metrics

To evaluate each component's contribution to perceptual quality, we use single-view resampling after the discriminator and energy network updates as the baseline. We introduce CGE, BSC and CTP in sequence. As shown in Table 12, CGE with three views reduces average LPIPS by 0.011 and NIQE by 0.11. Under the same view budget, BSC further reduces LPIPS by 0.011 and NIQE by 0.11, supporting the contribution of the bridge position sampling design to perceptual quality. CTP removes radial disagreement while retaining tangential components, further reducing LPIPS by 0.015 and NIQE by 0.18. Compared with the baseline, the complete SSUSB reduces average LPIPS by 0.037 and NIQE by 0.40. These results support the incremental gains from the three components and indicate that the update strategy helps reduce perceptual differences between restored and reference images while improving naturalness scores.