SYN-OPD: CONFLICT-AWARE ON-POLICY DISTILLATION FOR GENERATIVE VIDEO SUPER-RESOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Generative video super-resolution (VSR) has recently benefited from powerful video generative priors, enabling realistic detail synthesis beyond conventional reconstruction-based methods. However, a fundamental fidelity–generation dilemma remains: improving perceptual realism often comes at the expense of faithful reconstruction, while optimizing fidelity tends to suppress generative details. On-policy distillation (OPD) provides dense trajectory-level supervision that reduces unguided exploration, while multi-teacher variants enable capability integration across specialized experts. However, when fidelity-oriented and generation-oriented experts supervise the same restoration state, their optimization directions may conflict. This raises two coupled questions: **which expert should dominate at each noise level, and how should conflicting directions be reconciled?** To address these challenges, we propose **Syn-OPD**, a **conflict-aware multi-teacher OPD framework** that consolidates fidelity and generative capabilities into a single VSR model. We first specialize two complementary teachers from the same initialization: a **Fidelity Expert** for structure-preserving reconstruction and a **Generative Expert** for realistic detail synthesis. During distillation, hybrid SDE–ODE rollouts provide student-visited states for dense same-state supervision. Syn-OPD then resolves expert competition progressively: timestep-aware gating determines which expert should dominate, while asymmetric gradient decoupling preserves the preferred direction and removes only the conflicting auxiliary component. Extensive experiments on six benchmarks demonstrate consistent gains in both perceptual quality and reconstruction fidelity, with our 1.3B model even outperforming larger VSR models on perceptual metrics.

1 INTRODUCTION

Video super-resolution (VSR) plays an increasingly important role in practical applications such as smartphone photography, live streaming, gaming, and digital entertainment. Given degraded low-resolution (LR) videos affected by limited imaging quality and compression artifacts, VSR aims to reconstruct high-quality videos while faithfully preserving the underlying scene content and temporal consistency. Early VSR methods primarily improve reconstruction through recurrent propagation, alignment, and Transformer architectures (Fuoli et al., 2019; Chan et al., 2021; 2022b; Cao et al., 2021). More recently, advances in diffusion-based video generation have demonstrated strong generative priors for modeling realistic visual content (Blattmann et al., 2023; Chen et al., 2024; 2025a; Guo et al., 2025), motivating generative VSR approaches that leverage such priors to synthesize richer textures and improve perceptual quality (Chen et al., 2025b). Despite their promising visual quality, a fundamental challenge remains: ***how can generative VSR recover perceptually rich and realistic details while faithfully preserving the structures and content supported by the LR observation?*** Achieving a favorable balance between reconstruction fidelity and generative perceptual quality therefore remains a central problem in generative VSR.

Recently, on-policy distillation (OPD) has emerged as a promising post-training paradigm that mitigates the limitations of reward-based methods such as Group Relative Policy Optimization (GRPO) (Shao et al., 2024). Whereas GRPO-style optimization relies on sparse outcome-level scalar feedback, OPD trains the student on its own trajectories with dense teacher supervision, providing more direct, trajectory-level learning signals throughout the generation process (Agarwal et al., 2024; Zhang et al., 2026; Tan & Hong, 2026). Building upon this paradigm, recent multi-

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.

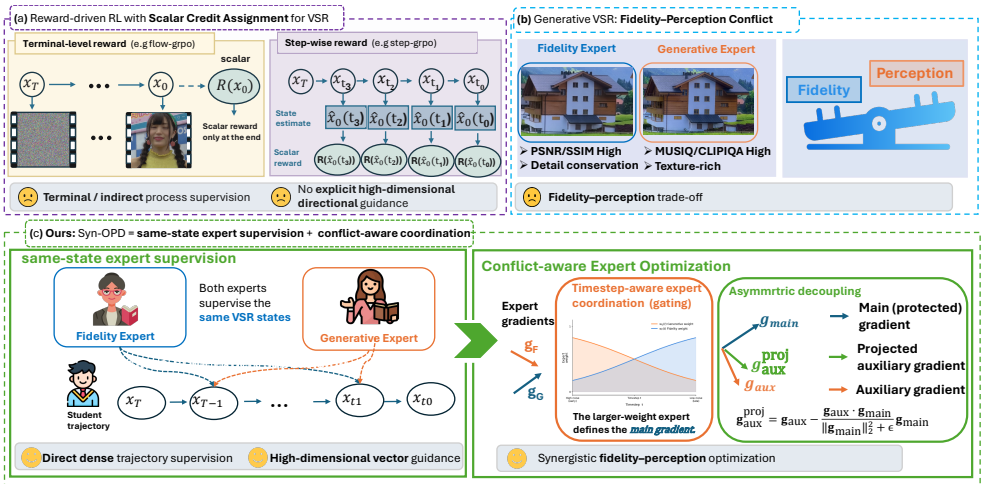


Figure 1: **Motivation of Syn-OPD. (a)** Reward-driven RL provides scalar feedback with limited directional guidance. **(b)** Generative VSR suffers from fidelity–perception conflicts. **(c)** Syn-OPD combines dense same-state dual-expert supervision with synergistic expert optimization, where **timestep-aware gating** establishes the preferred expert direction at each noise level and **asymmetric gradient decoupling** resolves conflicts while protecting that direction.

teacher OPD methods further attempt to consolidate multiple specialized capabilities into a single student (Li et al., 2026; Fang et al., 2026; Ma et al., 2026). These methods typically operate in a multi-task setting, where the current task determines which specialized teacher supervises the student, e.g., routing coding samples to a coding teacher and mathematical problems to a mathematics teacher. In generative VSR, however, fidelity and generative quality are not separable tasks, but two competing objectives that must be optimized simultaneously for the same restoration sample and even the same on-policy state. Therefore, task-based teacher routing cannot directly resolve the conflict between these objectives. This raises a pivotal question: *Can OPD effectively distill multiple potentially conflicting teachers into a single-task student, enabling fidelity and generative quality to improve synergistically rather than at the expense of each other?*

To address this question, we introduce **Syn-OPD**, a **conflict-aware on-policy distillation framework for jointly transferring fidelity and generative capabilities into a single VSR student**, as shown in Fig.1. To the best of our knowledge, Syn-OPD is the first to study conflict-aware multi-teacher OPD for generative VSR under a *single-task, dual-objective* setting, where multiple teachers must simultaneously guide the same restoration sample rather than being separated by task-specific routing. Syn-OPD follows a two-stage training strategy. We first specialize two complementary teachers from the same VSR initialization: a **Fidelity Expert**, optimized toward faithful reconstruction of structures and LR-supported content, and a **Generative Expert**, optimized toward realistic and perceptually rich detail synthesis. Building upon these specialized experts, Syn-OPD performs a three-step orchestration: (1) a hybrid SDE–ODE rollout explores states visited by the current student; (2) both frozen experts provide dense trajectory-level supervision at the same on-policy state; and (3) timestep-aware coordination with asymmetric gradient decoupling suppresses destructive interference between their competing directions. This design enables fidelity and generative quality to evolve synergistically within a single student. Extensive experiments demonstrate that Syn-OPD consistently improves perceptual quality while maintaining or improving reconstruction fidelity. In summary, our contributions are three-fold:

- We identify the **same-state fidelity–generation conflict** in generative VSR and introduce **Syn-OPD**, a conflict-aware multi-teacher OPD framework. To the best of our knowledge, we are the first to investigate multi-teacher OPD in a **single-task, dual-objective** VSR setting, where fidelity and generation must be jointly optimized on the same sample and on-policy state.
- We introduce a **synergistic expert optimization** mechanism that enables conflicting Fidelity and Generative Experts to be distilled collaboratively. It first establishes timestep-dependent expert priorities and then resolves conflicts asymmetrically by preserving the preferred direction and projecting only the conflicting auxiliary component.

- Extensive experiments on six benchmarks demonstrate consistent gains in both fidelity- and perception-oriented metrics. Quantitative results and attention visualization verify the complementary behaviors of the two experts, while ablation study and gradient-conflict analysis reveal substantial expert interference and validate the proposed coordination.

2 Related Work

Generative VSR and Reward-based Post-Training. Conventional VSR methods mainly exploit temporal propagation, alignment, and Transformer architectures (Chan et al., 2021; 2022b; Cao et al., 2021). Recent advances in generative priors have driven diffusion-based VSR methods toward improved perceptual quality and realistic detail synthesis (Zhou et al., 2024; Xie et al., 2025; Xu et al., 2025; Chen et al., 2025b; Sun et al., 2025; Wang et al., 2026a; Zhuang et al., 2026; Li et al., 2025). Meanwhile, reward-based post-training has shown promise for aligning generative models with desired perceptual objectives (Liu et al., 2025). However, scalar rewards provide only global preference signals and may struggle with conflicting objectives such as fidelity and perception. Syn-OPD complements scalar rewards with dense vector-field supervision, providing direct optimization directions and reducing inefficient exploration in the high-dimensional restoration space.

On-Policy Distillation. OPD provides dense supervision on student-generated trajectories, reducing the distribution mismatch of conventional offline distillation (Agarwal et al., 2024; Gu et al., 2024; Ko et al., 2024; Wang et al., 2026c; Chen et al., 2026; Lu & Lab, 2025; Zhou et al., 2026). Recent extensions explore multi-teacher capability integration in diffusion and flow models (Li et al., 2026; Ma et al., 2026). Existing multi-teacher OPD primarily considers multi-task or multi-domain settings, where different samples are routed to specialized teachers. Generative VSR instead requires fidelity and generation to be optimized simultaneously for the same restoration state. Syn-OPD therefore jointly distills two conflicting experts at the same on-policy state and coordinates their directions through timestep-aware gating and asymmetric gradient decoupling.

3 Method

3.1 Motivation: Fidelity–Generation Conflict in Unified SFT and GRPO

Our goal is to jointly improve reconstruction fidelity and generative capability in VSR, enabling more realistic detail synthesis while faithfully preserving the underlying content. A straightforward solution is to jointly optimize reconstruction- and generation-oriented objectives within a unified SFT framework. Given the HR target x and the restored output $\hat{x}$, we combine pixel reconstruction, LPIPS, flow matching, frequency distribution loss (FDL), and variational score distillation (VSD).

To better preserve high-frequency statistics, we employ FDL following the spectral regularization used in diffusion Transformer SR (Wang et al., 2026b):

$$\mathcal{L}_{\mathrm{FDL}}(x,\hat{x}) = \mathrm{SW}(A_\Phi(x), A_\Phi(\hat{x})) + \lambda_{\mathrm{phase}}\mathrm{SW}(P_\Phi(x), P_\Phi(\hat{x}))\,, \tag{1}$$

where $A_\Phi(\cdot)$ and $P_\Phi(\cdot)$ denote the frequency-domain amplitude and phase, respectively, and SW denotes the Sliced Wasserstein distance. Throughout the paper, λ_* denotes the corresponding weight. Let $z_0 = E_{\mathrm{VAE}}(x_{\mathrm{HQ}})$, $z_1 \sim \mathcal{N}(0, I)$, and $z_{\mathrm{LR}} = \mathrm{LR_Proj}(x_{\mathrm{LR}})$ denote the HQ latent, Gaussian source, and LR condition, respectively. We define $z_t = (1-t)z_0 + tz_1, t \sim \mathcal{U}(0,1)$, and optimize

$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E}_{z_0,z_1,t}\left[\|v_\theta(z_t, t, z_{\mathrm{LR}}) - (z_1 - z_0)\|_2^2\right]. \tag{2}$$

We further employ latent-space VSD regularization following OSEDiff (Wu et al., 2024) to encourage realistic detail generation under the pretrained generative prior. The SFT objective is

$$\mathcal{L}_{\mathrm{SFT}} = \lambda_{\mathrm{pix}}\|\hat{x} - x\|_1 + \lambda_{\mathrm{lpips}}\mathcal{L}_{\mathrm{LPIPS}}(\hat{x}, x) + \lambda_{\mathrm{FM}}\mathcal{L}_{\mathrm{FM}} + \lambda_{\mathrm{FDL}}\mathcal{L}_{\mathrm{FDL}} + \lambda_{\mathrm{VSD}}\mathcal{L}_{\mathrm{VSD}}. \tag{3}$$

However, we observe that these objectives favor different restoration solutions. Increasing the contribution of reconstruction-oriented losses improves fidelity metrics but tends to suppress perceptual quality. Conversely, emphasizing flow matching and VSD strengthens realistic detail generation but can degrade pixel-level fidelity. Such fidelity–perception tension has also been observed in diffusion-based SR (Luo et al., 2024), and is further confirmed by our ablation study.

Reward-based post-training with GRPO can steer the model toward a desired preference, but still exhibits a fidelity–generation trade-off when optimizing competing rewards. Moreover, scalar rewards provide only global preference signals, offering limited directional guidance in the high-dimensional VSR solution space. As shown in our ablation study, vanilla GRPO improves the targeted perceptual preference but can sacrifice reconstruction fidelity, exhibiting a similar fidelity–generation seesaw.

These observations suggest that simply balancing SFT losses or scalar rewards forces a single model onto a fidelity–generation trade-off, where improving one objective often comes at the expense of the other. This raises a central question: ***instead of forcing a single model to compromise between competing fidelity and generation objectives through loss or reward balancing, can we first preserve their respective strengths in two specialized experts and then transfer both capabilities into a single model?*** This motivates our fidelity–generative expert specialization and the subsequent Syn-OPD framework, which consolidates the respective strengths of both experts into a unified student.

3.2 Fidelity–Generative Expert Specialization

We obtain the two experts through objective-specific post-training from the same VSR initialization. For the Fidelity Expert, we increase the weights of reconstruction-oriented objectives, including pixel, LPIPS, and FDL losses during SFT, and correspondingly emphasize fidelity-related rewards during GRPO; for the Generative Expert, we instead place greater emphasis on flow-matching/VSD objectives and perceptual-quality rewards to encourage realistic detail synthesis. As shown in Tab. 1, the two experts exhibit clearly different optimization preferences. Teacher-F improves PSNR from 21.73 to 22.15 and LPIPS from 0.3810 to 0.3309, whereas Teacher-G substantially improves MUSIQ from 67.47 to 74.04 and CLIPIQA from 0.6021 to 0.7156. These results confirm their complementary fidelity- and generation-oriented capabilities. We further visualize their attention patterns in Fig. 2 under strictly controlled conditions, using the same input, noise realization, timestep, and intermediate state. The Fidelity Expert places relatively more attention on structurally stable background regions, while the Generative Expert responds more strongly to fine-detail regions such as animal fur (see Fig. A2 in Appendix). This qualitative difference is consistent with their quantitative specialization, suggesting that the two experts learn distinct restoration preferences. The remaining challenge is therefore how to transfer both expert directions into a single student without destructive interference, which motivates our Syn-OPD training.

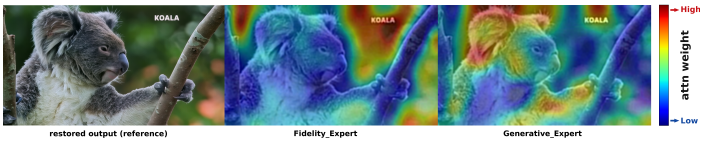


Figure 2: Attention maps of the Fidelity and Generative Experts.

Table 1: Quantitative comparison of the FlashVSR initialization and the two specialized experts. All models use the same initialization and LoRA capacity. The best results are highlighted in red, and the second-best results are underlined in blue.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	NIQE ↓	MUSIQ ↑	CLIPIQA ↑	FloLPIPS ↓
FlashVSR	21.7265	0.5946	0.3810	4.0951	67.4659	0.6021	0.3858
Fidelity Expert	22.1511	0.6064	0.3309	3.9053	64.9561	0.4756	0.3388
Generative Expert	21.1854	0.5327	0.3971	3.7806	74.0381	0.7156	0.3806

3.3 Syn-OPD: Synergistic Multi-Expert On-Policy Distillation

Given the two complementary experts, we introduce **Syn-OPD** to simultaneously transfer their fidelity- and generation-oriented capabilities into a single student, with the goal of maintaining strong performance on both objectives. However, as discussed above, the two experts may provide conflicting optimization directions, and directly aggregating their supervision can cause destructive interference. To address this issue, we introduce synergistic expert optimization, which consists of two complementary mechanisms: timestep-aware expert prioritization and asymmetric gradient decoupling. **The former** determines which expert should contribute more strongly at each noise level, while **the latter** resolves local directional conflicts by protecting the timestep-preferred expert and removing only the conflicting component of the auxiliary direction. As illustrated in Fig. 3, Syn-OPD consists of three key components: **(i) on-policy student rollout**, **(ii) same-state dual-expert**

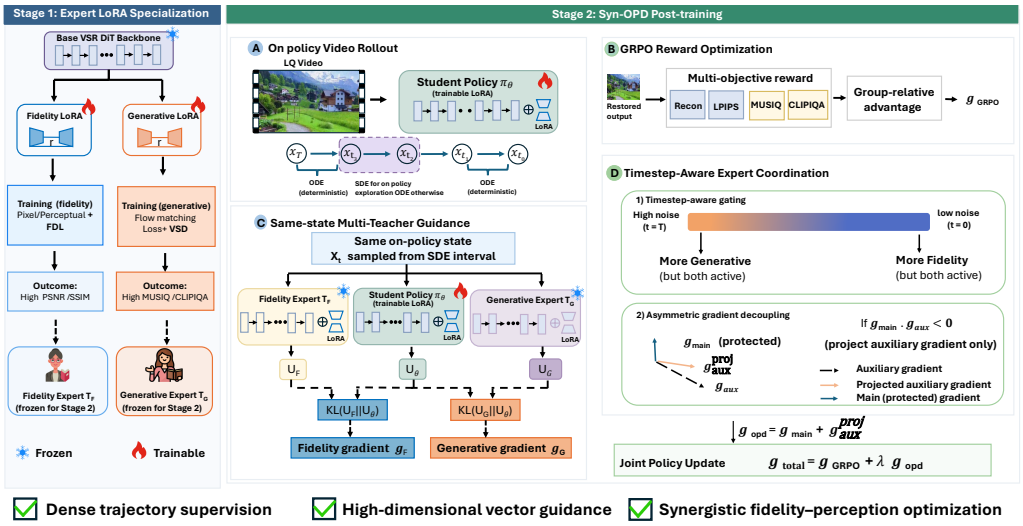


Figure 3: **Overview of the proposed Syn-OPD framework. Stage 1** specializes two complementary LoRA experts for fidelity and generation. **Stage 2** performs hybrid SDE–ODE on-policy rollout with GRPO guidance and dense dual-expert supervision. Syn-OPD coordinates expert directions via timestep-aware gating and asymmetric gradient decoupling.

supervision, and **(iii) synergistic expert optimization**. Together, they provide global reward guidance, dense trajectory supervision, and conflict-aware coordination, respectively.

On-Policy Student Rollout. Rather than training only on interpolated states constructed from the HQ latent and Gaussian noise, we let the current student generate its own restoration trajectories and optimize on the states it actually visits. To balance stochastic exploration with stable restoration, we adopt a hybrid SDE–ODE rollout, applying SDE within a randomly sampled interval $\mathcal{S}$ and ODE over the remainder of the trajectory,

$$\mathrm{d}z_t = \begin{cases} v_\theta(z_t, t, z_{\mathrm{LR}})\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t, & t \in \mathcal{S}, \\ v_\theta(z_t, t, z_{\mathrm{LR}})\,\mathrm{d}t, & t \notin \mathcal{S}. \end{cases} \tag{4}$$

Here, $v_\theta(z_t, t, z_{\mathrm{LR}})$ denotes the conditional flow velocity predicted by the current student, σ_t controls the exploration strength, and W_t is a standard Wiener process.

The resulting restoration is evaluated using a composite reward that balances reconstruction fidelity and perceptual quality:

$$\mathcal{R}(\hat{x}, x_{\mathrm{HQ}}) = -\lambda_{\mathrm{recon}}\mathcal{L}_{\mathrm{recon}} - \lambda_{\mathrm{lpips}}\mathcal{L}_{\mathrm{lpips}} - \lambda_{\mathrm{fdl}}\mathcal{L}_{\mathrm{fdl}} + \lambda_{\mathrm{iqa}}\mathcal{R}_{\mathrm{IQA}}(\hat{x}). \tag{5}$$

where $\mathcal{R}_{\mathrm{IQA}}(\hat{x}) = \frac{1}{2}\left[S_{\mathrm{CLIPIQA}}(\hat{x}) + s_{\mathrm{musiq}} S_{\mathrm{MUSIQ}}(\hat{x})\right]$. For each condition, we compute the group-relative advantage over G sampled candidate trajectories.

$$\hat{A}_i = \frac{R_i - \mu_R}{\sigma_R + \epsilon}, \qquad R_i = \mathcal{R}(\hat{x}_i, x), \tag{6}$$

where μ_R and σ_R denote the group reward statistics. The clipped GRPO objective yields the global policy gradient $\mathbf{g}_{\mathrm{GRPO}}$. It indicates *which* restoration is preferred, but provides limited guidance on how intermediate states should evolve.

Same-State Dual-Expert Supervision. To provide direct trajectory-level guidance, we further distill both frozen experts on the states actually visited by the student. Following the on-policy formulation of Flow-OPD (Fang et al., 2026), for each sampled state x_t from the stochastic rollout interval, the student, Fidelity Expert, and Generative Expert are queried using exactly the same (x_t, t, c):

$$v_\theta = v_\theta(x_t, t, c), \qquad v_F = v_{\mathcal{T}_F}(x_t, t, c), \qquad v_G = v_{\mathcal{T}_G}(x_t, t, c). \tag{7}$$

The expert-specific OPD objectives are defined as

$$\mathcal{L}_k^{\mathrm{OPD}} = \mathbb{E}_{x_t \sim \rho_t^\theta,\, t \in \mathcal{S}}\left[\frac{1}{2}\kappa(t)\,\|v_\theta(x_t, t, c) - \mathrm{SG}(v_k(x_t, t, c))\|_2^2\right], \quad k \in \{F, G\}, \tag{8}$$

where ρ_t^θ denotes the student's on-policy state distribution, $\kappa(t)$ is the timestep-dependent weighting term, and $\mathrm{SG}(\cdot)$ denotes stop-gradient. The corresponding expert gradients are

$$\mathbf{g}_F = \nabla_\theta \mathcal{L}_F^{\mathrm{OPD}}, \qquad \mathbf{g}_G = \nabla_\theta \mathcal{L}_G^{\mathrm{OPD}}. \tag{9}$$

Unlike task routing, both experts supervise the *same* VSR state, since fidelity and generation must be optimized jointly. Thus, $\mathbf{g}_F$ and $\mathbf{g}_G$ provide dense directional guidance toward complementary expert behaviors while exposing their local conflicts.

Synergistic Expert Optimization. Our analysis shows that the Fidelity and Generative Experts provide complementary but potentially conflicting supervision at the same on-policy state. Such gradient interference is common in joint multi-objective optimization (Yu et al., 2020; Liu et al., 2021), and we **observe increasingly frequent conflicts between the two expert gradients during training** (Fig. 5). Empirically, the Generative Expert is more effective at high-noise stages, while the Fidelity Expert is more beneficial at low-noise stages (Lin et al., 2025; He et al., 2026).Motivated by these observations, we introduce a **synergistic expert optimization** mechanism that first establishes timestep-dependent expert priorities and then resolves conflicting directions asymmetrically.

Timestep-Aware Expert Prioritization. Let $\tilde{t}$ denote the shifted diffusion timestep used by the underlying flow scheduler. We define timestep-aware expert coordination, as shown in Fig.1

$$w_G(\tilde{t}) = \sigma\left(\frac{\tilde{t}-\tau}{\gamma}\right), \qquad w_F(\tilde{t}) = 1 - w_G(\tilde{t}), \tag{10}$$

where $\sigma(\cdot)$ is the sigmoid function, τ denotes the gating threshold and γ controls the transition smoothness. Thus, the Generative Expert receives a larger weight at high-noise timesteps, whereas the Fidelity Expert is emphasized at low-noise timesteps. The gated expert gradients are

$$\tilde{\mathbf{g}}_F = w_F(\tilde{t})\mathbf{g}_F, \qquad \tilde{\mathbf{g}}_G = w_G(\tilde{t})\mathbf{g}_G. \tag{11}$$

Thus, both experts remain active throughout the trajectory, while their relative importance varies smoothly with the noise level. The expert with the larger gate weight is treated as the preferred main direction $\mathbf{g}_{\mathrm{main}}$, while the other serves as the auxiliary direction $\mathbf{g}_{\mathrm{aux}}$.

Asymmetric Gradient Decoupling. Timestep-aware prioritization determines which expert should be emphasized, but does not eliminate local conflicts between their directions. When $\mathbf{g}_{\mathrm{main}}^\top \mathbf{g}_{\mathrm{aux}} < 0$, direct aggregation causes partial cancellation. We therefore preserve the preferred direction and remove only the conflicting component of the auxiliary gradient:

$$\mathbf{g}_{\mathrm{aux}}^{\mathrm{proj}} = \begin{cases} \mathbf{g}_{\mathrm{aux}} - \dfrac{\mathbf{g}_{\mathrm{aux}}^\top \mathbf{g}_{\mathrm{main}}}{\|\mathbf{g}_{\mathrm{main}}\|_2^2 + \epsilon}\mathbf{g}_{\mathrm{main}}, & \mathbf{g}_{\mathrm{main}}^\top \mathbf{g}_{\mathrm{aux}} < 0, \\ \mathbf{g}_{\mathrm{aux}}, & \text{otherwise.} \end{cases} \tag{12}$$

The resulting synergistic OPD direction is

$$\mathbf{g}_{\mathrm{OPD}} = \mathbf{g}_{\mathrm{main}} + \mathbf{g}_{\mathrm{aux}}^{\mathrm{proj}}. \tag{13}$$

Unlike symmetric conflict-resolution schemes that treat both objectives equally, our decoupling is conditioned on the **timestep-dependent expert priority. We preserve the preferred direction and remove only the conflicting component of the auxiliary gradient**, thus retaining complementary information without suppressing the expert that is more relevant to the current noise level.

Joint Policy Update. Finally, we combine the gradient of the negative clipped GRPO objective with the conflict-decoupled expert gradient, $\mathbf{g}_{\mathrm{GRPO}} = \nabla_\theta \mathcal{L}_{\mathrm{GRPO}}$. The student is then updated by

$$\mathbf{g}_{\mathrm{total}} = \mathbf{g}_{\mathrm{GRPO}} + \lambda_{\mathrm{OPD}}\mathbf{g}_{\mathrm{OPD}}, \qquad \theta \leftarrow \theta - \eta \mathbf{g}_{\mathrm{total}}. \tag{14}$$

Overall, Syn-OPD integrates **reward-guided on-policy exploration, dense dual-expert supervision, and conflict-aware expert coordination** in a unified framework. Together, these components improve fidelity and generative quality while mitigating their optimization conflict.

4 Experiments

4.1 Implementation Details

Settings. We adopt FlashVSR (Zhuang et al., 2026), built upon Wan2.1-1.3B (Wan et al., 2025), as our base model and initialize all experiments from its pretrained checkpoint. As a one-step

Flow Matching VSR model, FlashVSR provides an efficient initialization. During post-training, we use multi-step on-policy rollouts to obtain intermediate states for dense supervision, this training-time rollout discretization is independent of the number of inference steps used at test time. We further optimize its LoRA parameters with a hybrid flow-matching and pixel-space reconstruction objective, improving both one-step and multi-step restoration performance (Huang et al., 2026). The fidelity and generative experts are trained independently and then frozen during Syn-OPD training, where only the student LoRA is updated. Training videos are collected from UltraVideo (Xue et al., 2025) and HQ-VSR (Chen et al., 2025b), filtered by NIQE and LAION-based scores (Schuhmann et al., 2022), and degraded using the RealBasicVSR pipeline (Chan et al., 2022a). Unless otherwise specified, all experiments share the same initialization, training data, and LoRA parameter budget.

Benchmarks. We evaluate on five full-reference VSR benchmarks, YouHQ40 (Zhou et al., 2024), REDS (Nah et al., 2019), SPMCS (Tao et al., 2017), UDM10 (Yi et al., 2019), and MVSR4x (Wang et al., 2023b), together with the real-world VideoLQ dataset (Chan et al., 2022a). For the five datasets with ground truth, we report both full-reference and no-reference perceptual metrics; for VideoLQ, we report only no-reference metrics. We also conducted a human preference study.

Baselines and Evaluation Metrics. We compare Syn-OPD with Upscale-A-Video (Zhou et al., 2024), STAR (Xie et al., 2025), SeedVR2-3B (Wang et al., 2026a), SwiftVR (Yan et al., 2026), and FlashVSR (Zhuang et al., 2026), with FlashVSR serving as the direct baseline. We report PSNR and SSIM (Wang et al., 2004) for fidelity, LPIPS (Zhang et al., 2018) for perceptual distance, and NIQE, MUSIQ (Ke et al., 2021), and CLIPIQA (Wang et al., 2023a) for perceptual quality.



Figure 4: Visualization results of video super-resolution on Benchmarks. (**Zoom in for best view**)

4.2 Comparison with existing methods

Quantitative Comparisons. We compare Syn-OPD with representative multi-step and one-step VSR methods. Tab. 2 reports results across six benchmarks. Compared with the one-step FlashVSR baseline, Syn-OPD consistently improves MUSIQ and CLIPIQA on all six datasets and PSNR on all five full-reference benchmarks. Despite using only a 1.3B backbone, Syn-OPD also surpasses the larger SeedVR2-3B and SwiftVR-5B models in MUSIQ and CLIPIQA across all six benchmarks, demonstrating the effectiveness of our post-training strategy without increasing model capacity. Meanwhile, LPIPS and NIQE remain competitive across the datasets. Overall, Syn-OPD achieves consistent improvements in perceptual quality and PSNR over FlashVSR while remaining competitive on the other reconstruction and perceptual metrics. At inference time, the two experts are discarded, and Syn-OPD retains the same inference architecture and efficiency as FlashVSR.

Qualitative Comparisons. We present qualitative comparisons in Fig. 4 and Fig. A1 in the Appendix. For clarity, the LR input and our restoration are shown in a diagonally split overview on the left, followed by enlarged crops for comparison. Syn-OPD better preserves structural details, such as building edges and facial structures, while producing richer, more natural high-frequency textures in challenging regions, including animal fur and vegetation. In particular, the second row shows that Syn-OPD recovers finer textures with clearer structures and more realistic details than competing methods. These observations are consistent with the quantitative gains in perceptual quality. A blind human preference study is presented in Sec. A.6 in the Appendix.

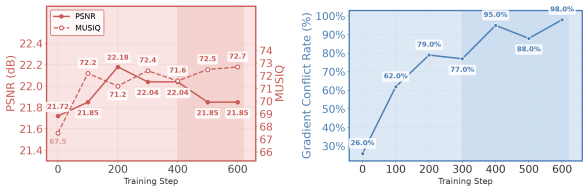
Table 2: Quantitative comparison with representative VSR methods across six benchmarks.

Dataset	Metric	Upscale-A-Video (0.86B, 30 steps)	STAR (1.8B, 15 steps)	SeedVR2-3B (3B, 1 step)	SwiftVR (5B, 1 step)	FlashVSR (1.3B, 1 step)	Ours (1.3B, 1 step)
MVSR4x	PSNR ↑	22.5886	21.5146	22.3892	22.31	22.208	22.432
	SSIM ↑	0.7516	0.7524	0.7295	0.7186	0.7274	0.7355
	LPIPS ↓	0.3881	0.4446	0.4026	0.3712	0.3409	0.3534
	NIQE ↓	7.8685	8.7121	6.1808	5.2265	5.2944	5.6062
	MUSIQ ↑	35.2405	32.0273	39.2930	66.22	68.748	69.411
	CLIPIQA ↑	0.2999	0.2679	0.3768	0.5565	0.5791	0.6537
REDS4x	PSNR ↑	22.3701	21.2093	21.6727	21.49	21.197	21.321
	SSIM ↑	0.5834	0.6000	0.5867	0.5853	0.5643	0.5694
	LPIPS ↓	0.5435	0.3868	0.3604	0.3106	0.2821	0.2779
	NIQE ↓	6.5326	4.2952	3.5870	3.1897	2.9854	3.1343
	MUSIQ ↑	27.3046	47.7379	47.1614	53.80	58.261	59.157
	CLIPIQA ↑	0.1652	0.2608	0.2487	0.2667	0.3340	0.3476
SPMCS	PSNR ↑	23.5920	19.5500	22.3949	22.31	22.496	22.594
	SSIM ↑	0.6405	0.5800	0.6694	0.5641	0.6300	0.6281
	LPIPS ↓	0.3617	0.3208	0.2270	0.2812	0.2024	0.2171
	NIQE ↓	5.8741	4.3516	3.7921	3.4078	3.6054	3.7783
	MUSIQ ↑	54.7610	67.5984	69.2841	70.64	71.818	72.961
	CLIPIQA ↑	0.3100	0.4886	0.5995	0.4738	0.5916	0.6597
UDM10	PSNR ↑	26.7340	23.0824	25.5356	25.51	25.122	25.368
	SSIM ↑	0.8030	0.7416	0.7871	0.7649	0.7713	0.7867
	LPIPS ↓	0.3156	0.2508	0.1971	0.2618	0.2034	0.1984
	NIQE ↓	6.4251	4.7197	3.4749	3.7941	3.6963	3.8038
	MUSIQ ↑	40.0894	65.6386	65.7276	66.35	69.544	69.834
	CLIPIQA ↑	0.2238	0.4755	0.5464	0.4790	0.5507	0.5672
YouHQ40	PSNR ↑	22.3573	21.3897	20.6524	22.07	21.776	22.169
	SSIM ↑	0.6101	0.6150	0.5555	0.5849	0.5904	0.6053
	LPIPS ↓	0.5225	0.5632	0.4502	0.3545	0.3269	0.3300
	NIQE ↓	7.2284	8.1821	4.3041	3.4678	3.5231	3.8637
	MUSIQ ↑	30.1147	26.6087	50.7820	69.60	69.537	69.688
	CLIPIQA ↑	0.2207	0.1929	0.3150	0.5948	0.5953	0.5964
VideoLQ	NIQE ↓	6.5857	5.6728	3.9053	4.1025	3.9529	3.9178
	MUSIQ ↑	26.9315	37.1708	57.5547	55.61	57.477	57.7759
	CLIPIQA ↑	0.1927	0.2690	0.3053	0.3765	0.4849	0.4937

5 Ablation Study

All ablation studies are conducted on YouHQ40. The main benchmark comparison in Tab. 2 focuses on the ultra-fast **1-step inference** setting, whereas the ablation study are conducted under a **4-step inference** setting using FlashVSR as the baseline. The longer trajectory exposes more intermediate states, enabling a finer-grained analysis of timestep-aware expert coordination. Unless otherwise specified, all ablation variants use identical training and inference settings for fair comparison.

Training Dynamics and Gradient Conflict Analysis. Fig. 5 further analyzes the training dynamics of Syn-OPD. As training progresses, MUSIQ increases from 67.46 to 72.75, while PSNR also improves over the initial baseline, indicating substantial perceptual gains without sacrificing reconstruction fidelity. Meanwhile, the conflict rate between the fidelity and generative expert gradients increases from 26% in the first 100 steps to over 80% in the later stages. At the early stage, the student is still far from both experts, and their gradients share more common improvement directions; as training progresses toward the fidelity–perception trade-off frontier, their local optimization preferences gradually diverge, resulting in more frequent gradient conflicts. The increasingly frequent



(a) PSNR & MUSIQ vs. Training Step. (b) Two Expert Gradient Conflict Rate.

Figure 5: Training dynamics and gradient conflict analysis of Syn-OPD.

activation of gradient projection therefore highlights the necessity of conflict-aware gradient decoupling during later-stage optimization.

Effect of Timestep-Aware Gating and Gradient Decoupling. Tab. 3 analyzes the two components of synergistic expert optimization: timestep-aware expert prioritization and asymmetric gradient decoupling. Direct Multi-Teacher OPD (MT-OPD) directly aggregates the Fidelity and Generative Expert gradients without explicit coordination, allowing conflicting directions to partially cancel. Adding timestep-aware gating prioritizes the Generative Expert at high-noise stages and the Fidelity Expert at low-noise stages, while MT-OPD + Inv. Gating reverses this priority. Although inverse gating slightly improves PSNR and LPIPS, it noticeably degrades MUSIQ and CLIPIQA, supporting the proposed timestep-dependent prioritization. Building on this priority, Syn-OPD further applies asymmetric gradient decoupling, preserving the preferred direction while removing only the conflicting auxiliary component. This further improves CLIPIQA while maintaining comparable PSNR, LPIPS, and MUSIQ, supporting the complementary roles of prioritization and conflict resolution.

Table 3: Ablation of Synergistic Expert Optimization on the YouHQ40 benchmark. Results are reported as mean ± standard deviation over checkpoints sampled after training stabilizes.

Method	PSNR ↑	LPIPS ↓	MUSIQ ↑	CLIPIQA ↑
FlashVSR-4step	21.7265	0.3810	67.4659	0.6021
Direct MT-OPD	22.03±0.18	0.3745±0.0144	71.26±1.17	0.6420±0.0374
MT-OPD + Gating	22.03±0.13	0.3694±0.0098	71.77±0.65	0.6554±0.0162
MT-OPD + Inv. Gating	22.13±0.16	0.3631±0.0140	70.64±0.84	0.6237±0.0198
Syn-OPD	22.07±0.14	0.3696±0.0100	71.74±0.64	0.6718±0.0162

Why Dense Expert Distillation? Tab. 4(a) compares different post-training strategies from the same initialization. Importantly, Vanilla GRPO uses the same multi-objective reward as Syn-OPD, yet it improves perceptual quality at the cost of reconstruction fidelity, exhibiting a clear fidelity–perception seesaw. In contrast, Syn-OPD improves both PSNR and perceptual metrics. This controlled comparison indicates that the balanced improvement cannot be attributed to reward design alone; dense same-state expert supervision and conflict-aware coordination provide additional directional guidance beyond scalar reward optimization.

Table 4: Ablation on YouHQ40. (a) Comparison of post-training strategies, where Vanilla GRPO and Syn-OPD use the same multi-objective reward. (b) Validation of specialized expert distillation.

(a) Post-training strategies.

Method	PSNR ↑	LPIPS ↓	MUSIQ ↑	CLIPIQA ↑
FlashVSR-4step	21.7265	0.3810	67.4659	0.6021
Data-Aug. SFT	22.0625	0.3625	68.5245	0.5689
Vanilla GRPO	21.4410	0.3579	69.9263	0.6001
Syn-OPD	22.0712	0.3696	71.6847	0.6651

(b) Specialized expert distillation.

Method	PSNR ↑	LPIPS ↓	MUSIQ ↑	CLIPIQA ↑
FlashVSR-4step	21.7265	0.3810	67.4659	0.6021
F-OPD	22.3494	0.3761	65.7179	0.4878
G-OPD	21.5417	0.4063	73.6102	0.7046
Syn-OPD	22.0712	0.3696	71.6847	0.6651

Effectiveness of Specialized Experts. Tab. 4(b) validates the complementary behaviors of the two specialized experts. Fidelity expert OPD (F-OPD) distills solely from the Fidelity Expert, whereas Generative expert OPD (G-OPD) uses only the Generative Expert. F-OPD achieves the highest PSNR, while G-OPD substantially improves MUSIQ and CLIPIQA, demonstrating specialization toward reconstruction fidelity and perceptual generation, respectively. By jointly distilling and coordinating these expert signals, Syn-OPD achieves a more balanced fidelity–perception trade-off.

6 Conclusion

We present **Syn-OPD**, a conflict-aware on-policy distillation framework for generative VSR that jointly improves reconstruction fidelity and generative perceptual quality. By combining specialized fidelity and generative experts with same-state dense supervision, timestep-aware coordination, and asymmetric gradient decoupling, Syn-OPD effectively mitigates destructive interference between competing optimization objectives. Extensive experiments across six benchmarks demonstrate consistent perceptual improvements while maintaining or improving reconstruction fidelity, highlighting the effectiveness of Syn-OPD for balanced generative VSR.

AI USE STATEMENT

Generative AI tools were used to assist with language polishing, manuscript organization, LaTeX formatting, and the refinement of figure layouts and captions. They were also used to suggest potentially relevant literature, which was subsequently checked by the authors against the original sources. Generative AI was not used to generate experimental data or results, nor to make unverified scientific claims on behalf of the authors. All AI-assisted text, figures, and references were manually reviewed and revised by the authors for correctness, originality, and consistency with the actual methodology and experiments. We take full responsibility for the final content of this work, including any text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

Our work focuses on improving generative video super-resolution for visual restoration and enhancement. The main experiments use publicly available research datasets and pretrained models and do not involve the collection of private personal data. We additionally conduct a small-scale human preference evaluation with Six participants to assess the generated video results. Participation is voluntary, and no personally identifiable information is collected or reported.

REPRODUCIBILITY STATEMENT

We provide detailed descriptions of the model architecture, expert specialization, Syn-OPD training procedure, optimization objectives, datasets, evaluation protocols, and hyperparameters in the main paper and appendix. All ablation settings use controlled initialization, training data, and trainable parameter budgets to facilitate fair comparison. We plan to release the implementation and necessary configuration files upon acceptance to support reproduction and further research.

REFERENCES

Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos Garea, Matthieu Geist, and Olivier Bachem. On-policy distillation of language models: Learning from self-generated mistakes. In *International Conference on Learning Representations*, volume 2024, pp. 21246–21263, 2024.

Andreas Blattmann, Tim Dockhorn, Sumith Kulal, Daniel Mendelevitch, Maciej Kilian, Dominik Lorenz, Yam Levi, Zion English, Vikram Voleti, Adam Letts, et al. Stable video diffusion: Scaling latent video diffusion models to large datasets. *arXiv preprint arXiv:2311.15127*, 2023.

Jiezhang Cao, Yawei Li, Kai Zhang, and Luc Van Gool. Video super-resolution transformer. *CoRR*, abs/2106.06847, 2021. URL https://arxiv.org/abs/2106.06847.

Kelvin C. K. Chan, Shangchen Zhou, Xiangyu Xu, and Chen Change Loy. Investigating tradeoffs in real-world video super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5962–5971, 2022a.

Kelvin CK Chan, Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Basicvsr: The search for essential components in video super-resolution and beyond. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4947–4956, 2021.

Kelvin CK Chan, Shangchen Zhou, Xiangyu Xu, and Chen Change Loy. Basicvsr++: Improving video super-resolution with enhanced propagation and alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5972–5981, 2022b.

Boyuan Chen, Diego Martí Monso, Yilun Du, Max Simchowitz, Russ Tedrake, and Vincent Sitzmann. Diffusion forcing: Next-token prediction meets full-sequence diffusion. In *Advances in Neural Information Processing Systems*, volume 37, pp. 24081–24125, 2024.

Guibin Chen, Dixuan Lin, Jiangping Yang, Chunze Lin, Junchen Zhu, Mingyuan Fan, Hao Zhang, Sheng Chen, Zheng Chen, Chengcheng Ma, et al. Skyreels-v2: Infinite-length film generative model. *arXiv preprint arXiv:2504.13074*, 2025a.

Tianlei Chen, Jiao Ou, Ziyuan Liu, Ruiming Tang, Jian Liang, and Han Li. Counteraction-aware multi-teacher on-policy distillation for general capability recovery with domain preservation. *arXiv preprint arXiv:2605.27115*, 2026.

Zheng Chen, Zichen Zou, Kewei Zhang, Xiongfei Su, Xin Yuan, Yong Guo, and Yulun Zhang. Dove: Efficient one-step diffusion model for real-world video super-resolution. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 85218–85237. Curran Associates, Inc., 2025b. doi: 10.52202/085713-2855. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/7b04ec5f2b89d7f601382c422dfe07af-Paper-Conference.pdf.

Zhen Fang, Wenxuan Huang, Yu Zeng, Yiming Zhao, Shuang Chen, Kaituo Feng, Yunlong Lin, Lin Chen, Zehui Chen, Shaosheng Cao, et al. Flow-opd: On-policy distillation for flow matching models. *arXiv preprint arXiv:2605.08063*, 2026.

Dario Fuoli, Shuhang Gu, and Radu Timofte. Efficient video super-resolution through recurrent latent space propagation. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 3476–3485, 2019.

Yuxian Gu, Li Dong, Furu Wei, and Minlie Huang. Minillm: Knowledge distillation of large language models. In *International Conference on Learning Representations*, volume 2024, pp. 32694–32717, 2024.

Yuwei Guo, Ceyuan Yang, Ziyan Yang, Zhibei Ma, Zhijie Lin, Zhenheng Yang, Dahua Lin, and Lu Jiang. Long context tuning for video generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 17281–17291, 2025.

Haodong He, Yuan Gao, Weizhong Zhang, and Gui-Song Xia. Timestep-aware block masking for efficient diffusion model inference. *arXiv preprint arXiv:2603.19939*, 2026.

Shuwei Huang, Tianyao Luo, Jicheng Liu, and Pan Zhou. RFMSR: Residual flow matching for image super-resolution. *arXiv preprint arXiv:2607.12753*, 2026.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. MUSIQ: Multi-scale image quality transformer. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5128–5137, 2021. doi: 10.1109/ICCV48922.2021.00510.

Jongwoo Ko, Sungnyun Kim, Tianyi Chen, and Se-Young Yun. Distillm: Towards streamlined distillation for large language models. *arXiv preprint arXiv:2402.03898*, 2024.

Quanhao Li, Junqiu Yu, Kaixun Jiang, Yujie Wei, Zhen Xing, Pandeng Li, Ruihang Chu, Shiwei Zhang, Yu Liu, and Zuxuan Wu. DiffusionOPD: A unified perspective of on-policy distillation in diffusion models. *arXiv preprint arXiv:2605.15055*, 2026.

Xiaohui Li, Yihao Liu, Shuo Cao, Ziyan Chen, Shaobin Zhuang, Xiangyu Chen, Yinan He, Yi Wang, and Yu Qiao. Diffvsr: Revealing an effective recipe for taming robust video super-resolution against complex degradations. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15319–15328. IEEE, 2025.

Qinwei Lin, Xiaopeng Sun, Yu Gao, Yujie Zhong, Zheng Zhao, Dengjie Li, and Haoqian Wang. TASR: Timestep-aware diffusion model for image super-resolution. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 10034–10043, 2025.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *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.

Jie Liu, Gongye Liu, Jiajun Liang, Yangguang Li, Jiaheng Liu, Xintao Wang, Pengfei Wan, Di ZHANG, and Wanli Ouyang. Flow-grpo: Training flow matching models via online rl. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 40783–40818. Curran Associates, Inc., 2025. doi: 10.52202/085713-1362. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/3a10c46572628d58cb44fb705f25cbbf-Paper-Conference.pdf.

Kevin Lu and Thinking Machines Lab. On-policy distillation. *Thinking Machines Lab: Connectionism*, 2025. doi: 10.64434/tml.20251026. https://thinkingmachines.ai/blog/on-policy-distillation.

Xiaotong Luo, Yuan Xie, Yanyun Qu, and Yun Fu. Skipdiff: Adaptive skip diffusion model for high-fidelity perceptual image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 4017–4025, 2024.

Wenhan Ma, Jianyu Wei, Liang Zhao, Hailin Zhang, Bangjun Xiao, Lei Li, Qibin Yang, Bofei Gao, Yudong Wang, Rang Li, et al. MOPD: Multi-teacher on-policy distillation for capability integration in llm post-training. *arXiv preprint arXiv:2606.30406*, 2026.

Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1996–2005. IEEE, 2019.

Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, et al. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in neural information processing systems*, 35:25278–25294, 2022.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Yujing Sun, Lingchen Sun, Shuaizheng Liu, Rongyuan Wu, Zhengqiang Zhang, and Lei Zhang. One-step diffusion for detail-rich and temporally consistent video super-resolution. In *Advances in Neural Information Processing Systems*, 2025.

Zhiquan Tan and Yinrong Hong. Self-supervised on-policy distillation for reasoning language models. *arXiv preprint arXiv:2605.17497*, 2026.

Xin Tao, Hongyun Gao, Renjie Liao, Jue Wang, and Jiaya Jia. Detail-revealing deep video super-resolution. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4482–4490. IEEE, 2017.

Team Wan, Ang Wang, Baole Ai, Bin Wen, Chaojie Mao, Chen-Wei Xie, Di Chen, Feiwu Yu, Haiming Zhao, Jianxiao Yang, et al. Wan: Open and advanced large-scale video generative models. *arXiv preprint arXiv:2503.20314*, 2025.

Jianyi Wang, Kelvin C. K. Chan, and Chen Change Loy. Exploring CLIP for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 2555–2563, 2023a. doi: 10.1609/aaai.v37i2.25353.

Jianyi Wang, Shanchuan Lin, Zhijie Lin, Yuxi Ren, Meng Wei, Zongsheng Yue, Shangchen Zhou, Hao Chen, Yang Zhao, Ceyuan Yang, et al. SeedVR2: One-step video restoration via diffusion adversarial post-training. In *International Conference on Learning Representations*, volume 2026, pp. 40987–41009, 2026a.

Jingkai Wang, Yixin Tang, Jue Gong, Jiatong Li, Shu Li, Libo Liu, Jianliang Lan, Yutong Liu, and Yulun Zhang. Spectral and trajectory regularization for diffusion transformer super-resolution. *arXiv preprint arXiv:2603.06275*, 2026b.

Ruohao Wang, Xiaohui Liu, Zhilu Zhang, Xiaohe Wu, Chun-Mei Feng, Lei Zhang, and Wangmeng Zuo. Benchmark dataset and effective inter-frame alignment for real-world video super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1168–1177, 2023b.

Yuran Wang, Zekun Wang, Bohan Zeng, Ruixu Zhang, Wenxuan Liu, Liu Yang, Yifan Dai, Yang Shi, Bozhou Li, Chengzhuo Tong, Daili Hua, Yuanxing Zhang, and Wentao Zhang. Flux-OPD: On-policy distillation with evolving contexts, 2026c. URL https://arxiv.org/abs/2607.28022.

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.

Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. *Advances in Neural Information Processing Systems*, 37: 92529–92553, 2024.

Rui Xie, Yinhong Liu, Penghao Zhou, Chen Zhao, Jun Zhou, Kai Zhang, Zhenyu Zhang, Jian Yang, Zhenheng Yang, and Ying Tai. STAR: Spatial-temporal augmentation with text-to-video models for real-world video super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 17108–17118, 2025.

Yiran Xu, Taesung Park, Richard Zhang, Yang Zhou, Eli Shechtman, Feng Liu, Jia-Bin Huang, and Difan Liu. Videogigagan: Towards detail-rich video super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2139–2149, 2025.

Zhucun Xue, Jiangning Zhang, Teng Hu, Haoyang He, Yinan Chen, Yuxuan Cai, Yabiao Wang, Chengjie Wang, Yong Liu, Xiangtai Li, et al. Ultravideo: High-quality uhd video dataset with comprehensive captions. *arXiv preprint arXiv:2506.13691*, 2025.

Jiaqi Yan, Xiangyu Chen, Xinlin Zhong, Haibin Huang, Chi Zhang, Jie Liu, Jiantao Zhou, and Xuelong Li. SwiftVR: Real-time one-step generative video restoration. *arXiv preprint arXiv:2606.09516*, 2026.

Peng Yi, Zhongyuan Wang, Kui Jiang, Junjun Jiang, and Jiayi Ma. Progressive fusion video super-resolution network via exploiting non-local spatio-temporal correlations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3106–3115, 2019.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *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.

Ke Zhang, Yunjie Tian, Dongdi Zhao, Yijiang Li, Yuanye Liu, Vishal M Patel, and Di Fu. On-policy distillation with best-of-n teacher rollout selection. *arXiv preprint arXiv:2605.09725*, 2026.

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.

Shangchen Zhou, Peiqing Yang, Jianyi Wang, Yihang Luo, and Chen Change Loy. Upscale-a-video: Temporal-consistent diffusion model for real-world video super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2535–2545, 2024.

Wei Zhou, Xiongwei Zhu, Zelin Xu, Bo Dong, Lixue Gong, Yongyuan Liang, Meng Chu, Leigang Qu, Lingdong Kong, Wei Liu, and Tat-Seng Chua. DanceOPD: On-policy generative field distillation, 2026. URL https://arxiv.org/abs/2606.27377.

Junhao Zhuang, Shi Guo, Xin Cai, Xiaohui Li, Yihao Liu, Chun Yuan, and Tianfan Xue. FlashVSR: Towards real-time diffusion-based streaming video super resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 43482–43493, 2026.

A APPENDIX

A.1 DISTINCTION FROM RELATED WORK

Beyond conventional multi-teacher OPD. Recent multi-teacher OPD methods, such as DiffusionOPD (Li et al., 2026) and MOPD (Ma et al., 2026), explore how multiple specialized teachers can be integrated into a single student. These methods mainly focus on capability integration across different tasks, domains, or teacher specializations. Syn-OPD instead considers a single-task setting in which the Fidelity and Generative Experts simultaneously supervise the same restoration state while optimizing competing objectives. The central difficulty is therefore not only *which knowledge to distill*, but also *how to coordinate conflicting expert directions along the restoration trajectory*.

Beyond generic gradient conflict resolution. Gradient-based multi-objective methods such as PCGrad (Yu et al., 2020) and CAGrad (Liu et al., 2021) mitigate interference between competing objectives by modifying or balancing conflicting gradients. Our setting differs in that expert importance is explicitly *timestep-dependent*: the Generative Expert is preferred at high-noise stages, whereas the Fidelity Expert becomes more important toward low noise. Syn-OPD therefore does not apply gradient projection independently. It first establishes a timestep-dependent expert priority through gating, and then conditions conflict resolution on this priority by preserving the preferred direction and removing only the conflicting component of the auxiliary gradient.

Coupled synergistic optimization. The components of Syn-OPD are thus functionally coupled rather than independently combined. Same-state OPD exposes the local disagreement between complementary experts; timestep-aware gating determines *which expert should dominate*; and asymmetric gradient decoupling determines *how the conflict should be resolved* under that priority. This forms a unified priority-aware coordination mechanism for multi-expert OPD rather than a direct combination of conventional multi-teacher distillation and generic gradient surgery.

A.2 MORE IMPLEMENTATION DETAILS

We provide detailed training configurations for the Fidelity Expert, Generative Expert, and Syn-OPD student. All models are built upon the same FlashVSR/Wan2.1-1.3B backbone, with the pretrained backbone frozen and only LoRA parameters optimized. Both experts use LoRA rank 96 and alpha 64, but adopt different SFT objectives and GRPO reward preferences to encourage fidelity- and generation-oriented specialization, respectively. During Syn-OPD training, the two expert LoRAs are frozen and only the student LoRA is updated. The heavier degradation used for the Generative Expert is intended to encourage greater reliance on the generative prior and promote richer high-frequency detail synthesis. For each rollout, we uniformly sample the starting position of a contiguous stochastic interval $\mathcal{S}$ covering 25% of the trajectory; the remaining steps are integrated deterministically with the ODE. For the OPD objective in Eq. (8), we use the timestep-dependent weighting $\kappa(t) = \frac{\Delta t}{2}\left(\frac{\sigma_t(1-t)}{2t} + \frac{1}{\sigma_t}\right)^2$, where Δt denotes the rollout step size and σ_t is the noise scale at timestep t. λ_{warp} weights an optical-flow-based temporal consistency loss.

Table A1: SFT training configurations for Fidelity and Generative Experts.

Configuration	Fidelity Expert	Generative Expert
LoRA rank / alpha	96 / 64	96 / 64
Frames	25	25
Crop size	384×384	384×384
λ_{recon}	1.20	0.40
λ_{LPIPS}	1.80	0.40
λ_{FM}	0.60	1.00
λ_{warp}	0.10	0.0
λ_{FDL}	0.06	0.0
λ_{VSD}	0.90	0.90
VSD start step	None	2500
Degradation mode	Standard	Heavy degradation
Top-k ratio	12	6
Local range	12	6

Table A2: GRPO training configurations for Fidelity and Generative Experts.

Configuration	Fidelity Expert	Generative Expert
LoRA rank / alpha	96 / 64	96 / 64
Learning rate	3×10^{-5}	3×10^{-5}
Frames	21	21
Crop size	1024×1024	1024×1024
Inference steps	20	20
Generations per prompt	6	6
Batches per epoch	2	2
Inner epochs	2	2
SDE ratio	0.25	0.25
Sigma shift	5.0	5.0
Clip range	1×10^{-3}	1×10^{-3}
Advantage clip	5.0	5.0
Gradient accumulation	2	1
Top-k ratio	6	6
Local range	12	12
λ_{recon}	3.5	0.5
λ_{LPIPS}	1.6	0.5
λ_{FDL}	0.1	0.3
λ_{IQA}	0.4	2.5
KL scale	0.03	0.03

Table A3: Training configuration of Syn-OPD.

Configuration	Value
LoRA rank / alpha	96 / 64
Learning rate	6×10^{-5}
Frames	21
Crop size	1024×1024
On-policy rollout steps	20
Generations per prompt	8
SDE ratio	0.25
Sigma shift	5.0
Inner epochs	2
Clip range	3×10^{-3}
Advantage clip	5.0
Gradient accumulation	1
Top-k ratio	6
Local range	12
OPD KL scale	0.07
Soft gating	Enabled
Gate threshold τ	650/1000
Gate temperature γ	360/1000
λ_{recon}	0.5
λ_{LPIPS}	0.5
λ_{FDL}	0.3
λ_{IQA}	2.5

A.3 One-Step Initialization and Inference-Step Analysis

One-Step Initialization. Our framework is initialized from the pretrained FlashVSR model (Zhuang et al., 2026), which is built upon Wan2.1-1.3B (Wan et al., 2025) and natively supports one-step VSR inference. Therefore, unlike approaches that obtain few-step or one-step

generation through an additional distillation stage, Syn-OPD directly inherits the one-step capability of the pretrained backbone. All subsequent expert specialization and Syn-OPD training are performed through LoRA adaptation, while the pretrained backbone remains frozen.

Hybrid Flow and Reconstruction Supervision. To preserve the native one-step capability while further improving restoration quality, we optimize the trainable LoRA parameters with both flow-space and reconstruction-space supervision. Specifically, in addition to the flow-matching objective, we directly supervise the predicted clean output $\hat{x}_0$ in pixel space:

$$\mathcal{L}_{\text{hybrid}} = \lambda_{\text{FM}}\mathcal{L}_{\text{FM}} + \lambda_{\text{rec}}\mathcal{L}_{\text{rec}}(\hat{x}_0, x), \tag{15}$$

where $\mathcal{L}_{\text{FM}}$ constrains the learned velocity field and $\mathcal{L}_{\text{rec}}$ directly encourages the predicted clean sample to recover the ground-truth high-resolution frame. The two objectives provide complementary supervision: flow matching preserves a well-formed restoration trajectory, whereas direct x_0 supervision constrains the final reconstruction in pixel space. Such joint supervision is particularly useful for few-step restoration, since prediction errors cannot be gradually corrected through a long denoising trajectory. This observation is also consistent with recent studies on flow-matching restoration models (Huang et al., 2026).

Importantly, this training does not introduce an additional step-distillation procedure. We distinguish the **training rollout steps** used to construct on-policy trajectories from the **test-time inference steps**. During Syn-OPD training, we use a 20-step rollout to expose intermediate states for dense expert supervision, while the trained student retains the native one-step inference capability inherited from FlashVSR. Instead, the one-step capability is inherited from FlashVSR and further adapted through the hybrid training objective. Meanwhile, because the underlying Flow Matching formulation still defines a continuous restoration trajectory, the trained model can also be evaluated with additional inference steps. This allows us to study whether Syn-OPD remains effective beyond its native one-step setting.

Table A4: Comparison of FlashVSR and Syn-OPD under different inference steps on YouHQ40. For each inference setting, the better result is highlighted in red.

Metric	25 Steps		4 Steps		1 Step	
	FlashVSR	Ours	FlashVSR	Ours	FlashVSR	Ours
PSNR ↑	21.7171	21.6149	21.7265	22.0712	21.776	22.169
SSIM ↑	0.5832	0.5775	0.5966	0.5980	0.5904	0.6053
LPIPS ↓	0.3692	0.3573	0.3810	0.3696	0.3269	0.3300
NIQE ↓	3.8035	3.6585	3.8240	4.2171	3.5231	3.8637
MUSIQ ↑	68.9683	71.9079	67.4659	71.6847	69.537	69.688
CLIPIQA ↑	0.5937	0.6720	0.6021	0.6651	0.5953	0.5964

Effect of Inference Steps. Tab. A4 compares FlashVSR and Syn-OPD under one-, four-, and 25-step inference. Syn-OPD remains effective across different sampling budgets, indicating that the proposed post-training does not destroy the native few-step capability of the backbone. Under one-step inference, Syn-OPD improves PSNR from 21.776 to 22.169 and SSIM from 0.5904 to 0.6053, while also slightly improving MUSIQ and CLIPIQA. This result is particularly important because no additional one-step distillation is introduced during our training.

With four inference steps, Syn-OPD achieves a stronger overall balance, improving PSNR, SSIM, LPIPS, MUSIQ, and CLIPIQA over the corresponding FlashVSR baseline. When increasing the sampling budget to 25 steps, the perceptual benefit becomes more pronounced: MUSIQ improves from 68.97 to 71.91 and CLIPIQA from 0.5937 to 0.6720, together with improved LPIPS and NIQE. The slight decrease in PSNR and SSIM at 25 steps suggests that longer trajectories increasingly favor perceptual refinement, consistent with the fidelity–perception trade-off studied in the main paper.

Overall, these results show that Syn-OPD is not tied to a particular inference schedule. It preserves the strong one-step capability inherited from FlashVSR, remains effective under the four-step setting used for our diagnostic ablations, and can further exploit additional sampling steps to enhance

perceptual quality. We attribute this flexibility to the combination of the one-step Flow Matching initialization and hybrid flow/reconstruction supervision, which jointly preserve trajectory modeling and direct clean-output reconstruction.

A.4 More Visualations

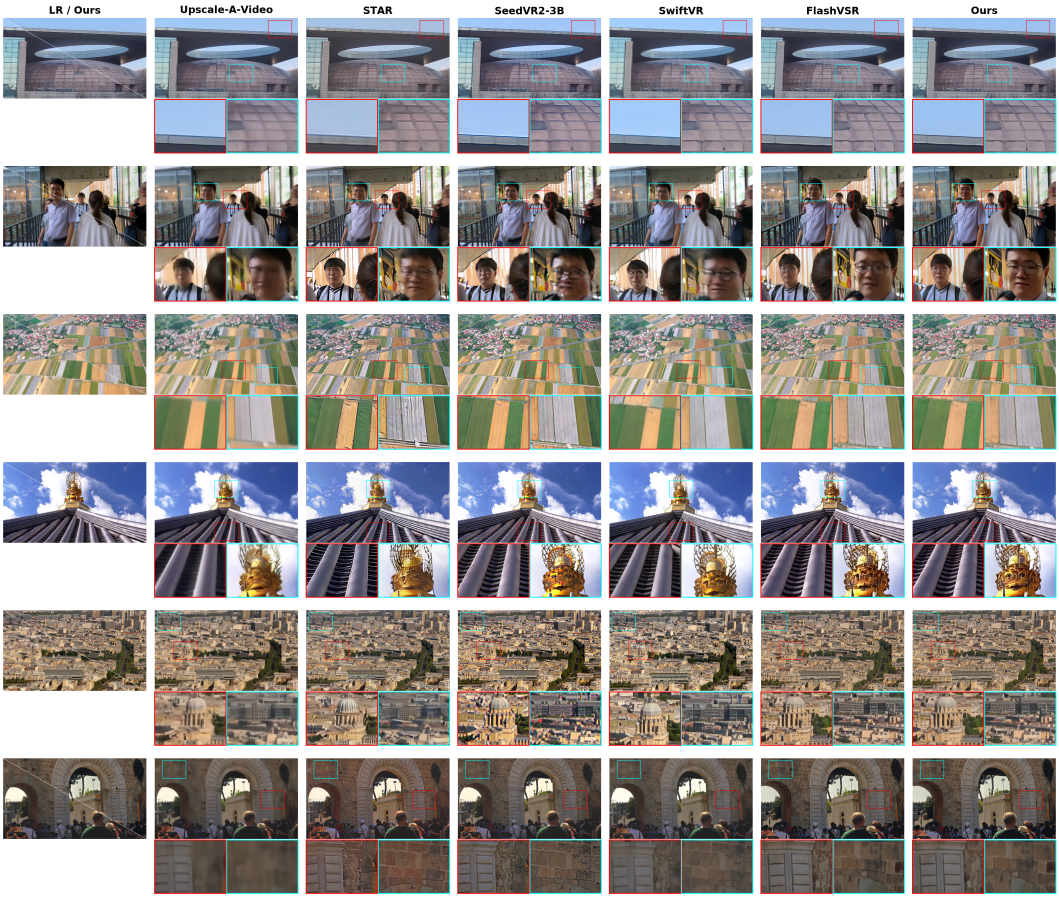


Figure A1: Visualization results of video super-resolution on Benchmarks. (**Zoom in for best view**)

A.5 Detailed Analysis of Two Experts

We further analyze the Fidelity and Generative Experts from both qualitative and model-internal perspectives. Although both experts are initialized from the same pretrained VSR model, their different training preferences lead to clearly different restoration behaviors, intermediate responses, and depth-wise adaptation patterns.



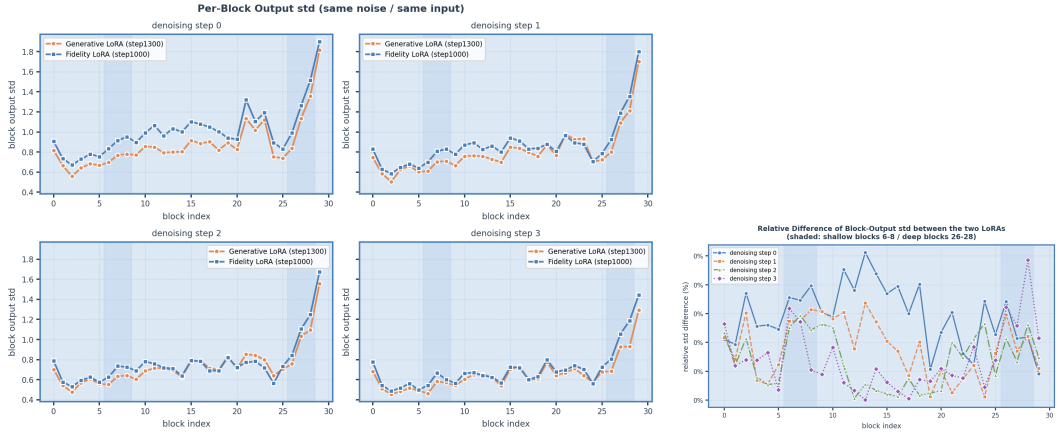
(a) Fidelity Expert. (b) Generative Expert.

Figure A2: **Qualitative comparison of the specialized experts.** The Fidelity Expert produces smoother and more conservative restoration, while the Generative Expert synthesizes substantially richer fine-scale textures.

A.5.1 Qualitative Expert Specialization

Fig. A2 qualitatively demonstrates the different restoration preferences learned by the two experts. The Fidelity Expert produces a smoother result and tends to preserve the appearance supported by the LR observation. In contrast, the Generative Expert synthesizes much richer high-frequency details, particularly on the animal skin. This behavior is consistent with their intended specialization: the Fidelity Expert favors faithful reconstruction, whereas the Generative Expert relies more strongly on the learned generative prior for realistic detail synthesis. Together with the quantitative results in the main paper, this comparison further confirms their complementary restoration behaviors.

A.5.2 Intermediate Feature Response Analysis



(a) Per-block output standard deviation across denoising steps. (b) Relative difference between the two experts.

Figure A3: **Intermediate feature responses of the two experts.** The Fidelity Expert generally maintains larger feature variations, while the discrepancy between the two experts varies substantially across network depth. We evaluate the restoration trajectory using four inference steps, denoted as step 0, step 1, step 2, and step 3, corresponding to sequential denoising stages.

We next compare the intermediate block outputs under the same input, noise, and denoising conditions. As shown in Fig. A3a, the Fidelity Expert generally exhibits higher output standard deviation across most blocks and denoising steps. Since the standard deviation reflects the dispersion and dynamic range of intermediate features, this suggests that the Fidelity Expert preserves stronger input-dependent variation throughout the restoration trajectory. By contrast, the Generative Expert shows relatively more concentrated feature responses, which is consistent with a stronger reliance on learned generative priors.

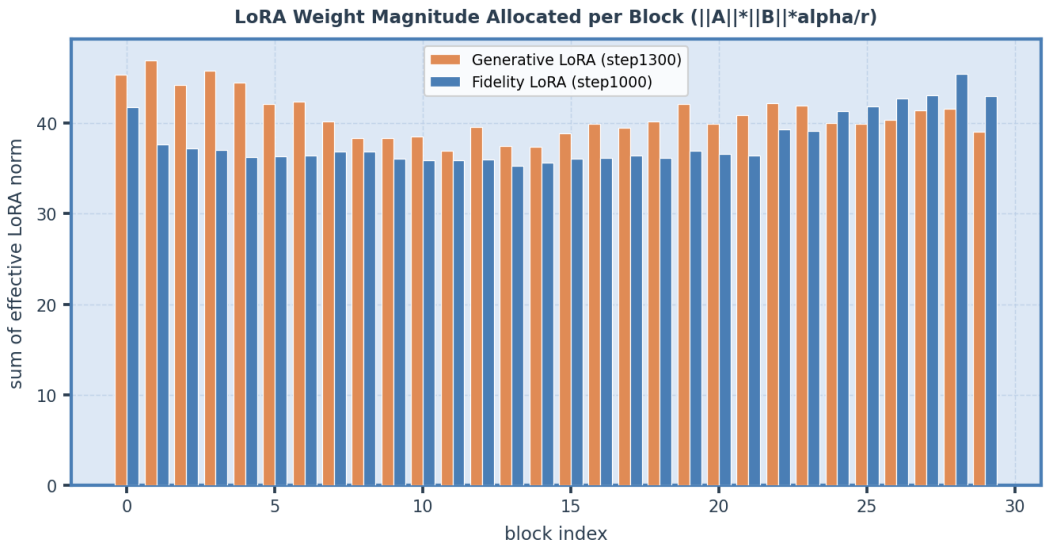


Figure A4: **Depth-wise LoRA adaptation of the two experts.** The Generative Expert allocates stronger adaptation to earlier blocks, whereas the Fidelity Expert becomes relatively stronger in later blocks.

The discrepancy is also highly depth-dependent. As shown in Fig. A3b, the difference between the two experts is not uniformly distributed over the network, but becomes more pronounced in several specific shallow and deep regions. This suggests that the two experts do not simply differ by a global feature-scale change, but develop distinct internal processing patterns at different network depths. We note that output standard deviation is an activation statistic rather than a direct measure of texture or structure; nevertheless, it provides complementary evidence for the specialization observed in the restored outputs.

A.5.3 Depth-wise LoRA Adaptation

Finally, we examine how the two LoRA experts modify different depths of the pretrained DiT backbone. As shown in Fig. A4, the Generative Expert exhibits larger LoRA magnitudes in the early blocks, indicating stronger adaptation in the front part of the network. Toward later blocks, the difference gradually decreases and the Fidelity Expert becomes relatively stronger in several deep blocks. This reveals distinct depth-wise adaptation patterns: the Generative Expert concentrates more adaptation capacity in earlier feature transformations, whereas the Fidelity Expert distributes its adaptation more evenly and places relatively greater emphasis on later layers.

Importantly, LoRA magnitude indicates where parameter adaptation is allocated, rather than directly proving the semantic function of each block. However, together with the feature-response analysis in Fig. A3 and the qualitative comparison in Fig. A2, these results consistently show that the two experts develop distinct internal representations despite sharing the same pretrained backbone.

A.6 Temporal Consistency and Human Preference Evaluation

Temporal Consistency. To further examine whether the enhanced generative details introduce temporal instability, we evaluate Syn-OPD and the direct FlashVSR baseline on YouHQ40 using FloLPIPS, together with the spatial metrics reported in the main comparison. Both methods are evaluated under the same 1-step inference setting.

Syn-OPD reduces FloLPIPS from 0.3672 to 0.3211, indicating improved temporal consistency over the FlashVSR baseline. Meanwhile, it achieves higher PSNR, SSIM, MUSIQ, and CLIPIQA, with comparable LPIPS. These results suggest that the improved generative quality does not come at the cost of increased temporal instability.

Table A5: Spatial and temporal comparison with FlashVSR on YouHQ40. Lower LPIPS, NIQE, and FloLPIPS are better, while higher values are better for the remaining metrics.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	NIQE ↓	MUSIQ ↑	CLIPIQA ↑	FloLPIPS ↓
FlashVSR	21.776	0.5904	**0.3269**	**3.5231**	69.537	0.5953	0.3672
Syn-OPD	**22.169**	**0.6053**	0.3300	3.8637	**69.688**	**0.5964**	**0.3211**

Human Preference Evaluation. To complement the automatic metrics, we conduct a blind human preference study on YouHQ40, comparing Syn-OPD with FlashVSR. For each video clip, the two restored results are displayed side by side with randomized left–right ordering, and the method identities are hidden from the participants.

Six participants independently evaluate each pair according to three criteria: **overall visual quality**, **content fidelity**, and **temporal consistency**. For each criterion, participants select Syn-OPD, FlashVSR, or "Tie / No clear preference." We aggregate all clip–participant judgments and report the resulting preference ratios in Tab. A6.

Table A6: Blind human preference comparison between Syn-OPD and FlashVSR on YouHQ40. Six participants independently evaluate each video pair. Each entry reports the percentage of judgments preferring Syn-OPD, FlashVSR, or indicating no clear preference.

Criterion	Syn-OPD ↑	FlashVSR	Tie
Overall Visual Quality	**63.3**%	10.8%	25.8%
Content Fidelity	**67.5**%	11.6%	20.9%
Temporal Consistency	**70.8**%	7.5%	21.7%

The results show a consistent preference for Syn-OPD across all three criteria. In particular, Syn-OPD receives 70.8% of the judgments for temporal consistency, complementing the FloLPIPS evaluation and providing perceptual evidence that the enhanced generative details do not introduce obvious temporal instability.