SHORT-WINDOW DYNAMICS FOR CONFIGURATION SCREENING IN CONTROLLED DIFFUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Inference-time control often requires searching over intervention operators, strengths, source weights, and temporal schedules. Running a complete denoising trajectory for every candidate quickly becomes expensive. We present SHORTWIN, a training-free screening method that probes only a short reverse-denoising window in each active control stage. From these local trajectories, SHORTWIN measures control response, propagation gain, directional conflict with the native denoising update, and relative intervention magnitude. For tasks with explicit source constraints, these generic diagnostics can be paired with a lightweight consistency check. The measurements are summarized at the stage level, aggregated across a multi-stage schedule, and used to decide which configurations continue to full sampling. On a dual-source AFHQ-Dog testbed, we collect 109,080 seven-step traces from 5,454 configurations. Among 3,780 effective two-stage configurations, the screen rejects 18.67% of sample–configuration pairs overall and 60.30% for Laplacian high-frequency control. On 90 fully sampled cases, accepted and rejected groups show clear separation in source-consistency metrics; the screening margin correlates with reference SSIM ($\rho = 0.645$) and reference low-frequency distance ($\rho = -0.724$). The results show that short local trajectories contain useful information for reducing controlled-diffusion configuration search.

1 INTRODUCTION

Diffusion models produce images through a sequence of denoising updates (Ho et al., 2020; Nichol and Dhariwal, 2021; Song et al., 2021). Many control methods modify these updates at inference time so that a pretrained model can follow a reference image, an external objective, or a spatial condition (Choi et al., 2021; Meng et al., 2022; Bansal et al., 2023; Mo et al., 2024). In practice, the controller itself introduces a search problem. Its operator, strength, source weights, and active interval can all affect the output, and guidance behavior can change substantially across noise levels (Kynkäänniemi et al., 2024). A useful configuration may therefore be one point in a large schedule space.

The usual way to compare configurations is to complete the full denoising process for each one and score the final images. This gives direct evidence, but it spends the same amount of sampling compute on promising and poor candidates. The early part of a controlled trajectory already reveals several useful behaviors. A perturbation can be too weak to change the clean-image estimate, strong enough to dominate the native update, or oriented against the direction in which the sampler is moving. These effects are local and can be measured before the trajectory is complete.

Recent early-exit work in text-to-image diffusion studies whether a partially generated *sample* should continue. Probe-Select and Diffusion Probe predict final quality from early denoiser features (Guo et al., 2026; Huang et al., 2026), while CFG-Rejection uses statistics collected along classifier-free-guidance trajectories (Wang et al., 2026). We study a different search unit: the control configuration applied to a fixed input. The aim is to screen operator–strength–schedule choices from their interaction with the native denoising dynamics.

SHORTWIN implements this idea with short, stage-wise probes. For each candidate configuration π, it runs K local reverse steps in every active control stage and records four quantities: local response, propagation gain, directional conflict, and relative intervention magnitude. In the dual-source image-translation setting used here, we also measure how well the predicted clean image remains compatible

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.

Table 1: Scope relative to recent early-assessment methods. "Learned" denotes an additional learned assessment model.

Method	Screened unit	Primary signal	Learned	Target
Probe-Select (Guo et al., 2026)	seed/sample	denoiser activations	yes	final quality
Diffusion Probe (Huang et al., 2026)	sample/workflow	attention statistics	yes	final quality
CFG-Rejection (Wang et al., 2026)	sample	CFG path statistic	no	sample filtering
SHORTWIN	config–input pair	intervention dynamics	no	config screening

with the two source images in low-frequency space. Stage summaries are combined into schedule-level statistics and a screening margin. Configurations that satisfy the screening criteria continue to full sampling; the rest are removed from the search. Figure 1 gives an overview.

We evaluate the method on 5,454 configurations and 20 paired inputs, producing 109,080 short-window records. The experiments show a strong operator effect: effective Gaussian and Fourier low-pass controls are retained in 90.39% and 82.67% of cases, whereas Laplacian controls are retained in 39.70%. Generic local dynamics capture many of the unstable Laplacian cases, while low-frequency consistency measurements are more informative for smooth controls that remain visually plausible but drift away from the intended sources. Full-sampling validation further shows that the screening margin tracks reference consistency over completed trajectories. Together, these results support short-window dynamics as a practical signal for configuration search.

2 Related Work

Inference-time diffusion control. ILVR constrains unconditional diffusion by matching low-frequency content from a reference image (Choi et al., 2021), and SDEdit introduces noise into a guide before denoising it toward the model distribution (Meng et al., 2022). Universal Guidance applies gradients from external objectives without task-specific retraining (Bansal et al., 2023), while FreeControl uses pretrained diffusion features for spatial control (Mo et al., 2024). Learned control modules such as ControlNet, T2I-Adapter, and IP-Adapter provide additional conditioning pathways (Zhang et al., 2023; Mou et al., 2024; Ye et al., 2023). These methods focus on how control is injected. Our study focuses on how to choose among multiple settings of an inference-time controller.

Time-dependent guidance and inference-time search. The effect of guidance varies over the denoising trajectory. Restricting guidance to a middle interval can improve sample and distribution quality (Kynkäänniemi et al., 2024), which makes the active schedule an important configuration variable. Recent work also frames inference-time scaling as search over diffusion decisions or trajectories (Zhang et al., 2026). SHORTWIN complements this direction by providing a local signal for deciding which configurations deserve a complete trajectory.

Early assessment and sample filtering. Probe-Select learns to predict final text-to-image quality from early denoiser activations (Guo et al., 2026). Diffusion Probe uses early attention statistics for a related purpose (Huang et al., 2026). CFG-Rejection filters samples using classifier-free-guidance path statistics without a learned reward model (Wang et al., 2026). Table 1 summarizes the scope of these methods. SHORTWIN uses the explicit control intervention u_t and its local interaction with the base sampler, so its screened unit is a configuration–input pair.

3 Problem Formulation

Let $F_\theta(\mathbf{x}_t, t)$ denote the base sampling update from state $\mathbf{x}_t$. A controlled sampler adds an intervention $u_t(\pi)$ determined by configuration

$$\pi = \{\mathcal{O}, \boldsymbol{\lambda}, \mathcal{T}, \mathcal{W}\}, \qquad \mathbf{x}^{\pi}_{t-1} = F_\theta(\mathbf{x}_t, t) + u_t(\pi), \tag{1}$$

where $\mathcal{O}$ is the intervention operator, $\boldsymbol{\lambda}$ contains its strength parameters, $\mathcal{T}$ specifies the active windows, and $\mathcal{W}$ contains source weights. The effective intervention u_t may come from additive, difference, or replacement-style control.

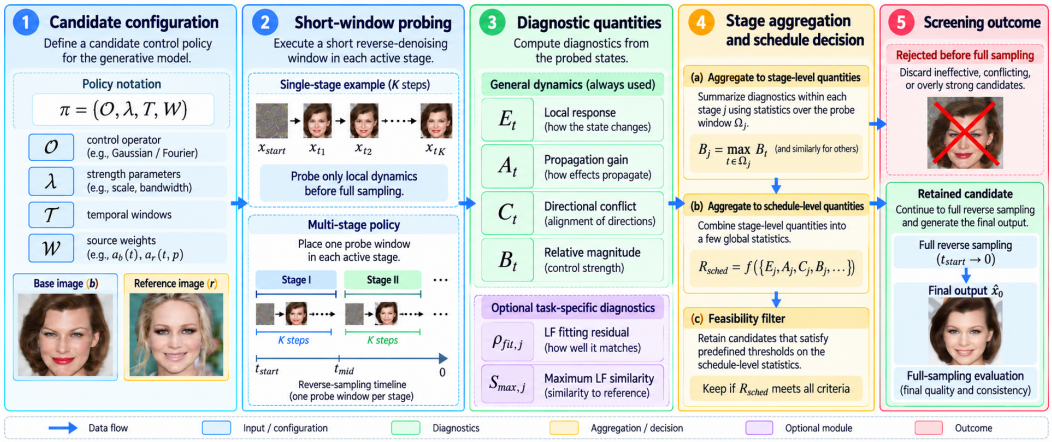

Figure 1: Overview of SHORTWIN. A candidate control policy is probed in short reverse-denoising windows. Local dynamics and optional task-specific consistency measurements are summarized within each stage, aggregated across the schedule, and used to filter configurations before full sampling.

For a candidate set Π, exhaustive search completes every trajectory and evaluates a utility $U(\pi)$ from the final output. Short-window screening instead computes a decision $s(\pi) \in \{0, 1\}$ from local probes and keeps $\Pi_{\text{keep}} = \{\pi : s(\pi) = 1\}$. The screen is useful when it reduces the number of full trajectories while preserving configurations that perform well under the downstream evaluation.

4 SHORT-WINDOW CONFIGURATION SCREENING

4.1 SHORT-WINDOW PROBING

For every active stage j, we place a probe window Ω_j containing K consecutive reverse-sampling steps. We use $K = 7$ throughout the experiments. Let $P_\theta(\mathbf{x}_t, t)$ be the denoiser prediction of the clean image $\hat{\mathbf{x}}_0$. At a controlled step $t \in \Omega_j$, define

$$\mathbf{v}_t = \frac{u_t}{\|u_t\|_2 + \varepsilon}, \qquad q_t = \langle P_\theta(\mathbf{x}_t, t), \mathbf{v}_t \rangle. \tag{2}$$

The probe follows the same sampler and controller as the full trajectory; it simply stops after the local measurements have been collected.

4.2 LOCAL DIAGNOSTIC QUANTITIES

The four generic diagnostics describe different aspects of the interaction between the intervention and the sampler.

Local response E_t. We measure the sensitivity of the current clean-image prediction along the intervention direction:

$$E_t = \frac{\|\nabla_{\mathbf{x}_t} q_t\|_1}{\|u_t\|_2 + \varepsilon}. \tag{3}$$

A larger response means that the current denoiser prediction is more sensitive to the applied control direction.

Propagation gain A_t. Let $\mathbf{x}^0_{t-1} = F_\theta(\mathbf{x}_t, t)$ and $\mathbf{x}^\pi_{t-1} = \mathbf{x}^0_{t-1} + u_t$. We compare the next clean-image predictions:

$$A_t = \frac{\|P_\theta(\mathbf{x}^\pi_{t-1}, t-1) - P_\theta(\mathbf{x}^0_{t-1}, t-1)\|_2}{\|u_t\|_2 + \varepsilon}. \tag{4}$$

This quantity measures how strongly the intervention survives into the next denoising step.

Directional conflict C_t. Let $d_t = F_\theta(\mathbf{x}_t, t) - \mathbf{x}_t$ be the native sampling displacement. We record antagonistic alignment between the intervention and the base update:

$$C_t = \max\left(-\frac{\langle u_t, d_t\rangle}{\|u_t\|_2\|d_t\|_2 + \varepsilon}, 0\right). \tag{5}$$

Large C_t indicates that the controller pushes against the native denoising direction.

Relative magnitude B_t. The intervention strength relative to the model update is

$$B_t = \frac{\|u_t\|_2}{\|d_t\|_2 + \varepsilon}. \tag{6}$$

The step-level diagnostics are averaged within each probe window,

$$\bar{Z}_j = \frac{1}{|\Omega_j|}\sum_{t\in\Omega_j} Z_t, \qquad Z \in \{E, A, C, B\}, \tag{7}$$

and the first three are combined into a stage response score

$$R_j = \sigma\big(\beta_0 + \beta_1 \log(\bar{E}_j + \varepsilon) + \beta_2 \log(\bar{A}_j + \varepsilon) - \beta_3 \bar{C}_j\big). \tag{8}$$

We use $(\beta_0, \beta_1, \beta_2, \beta_3) = (0, 1, 1, 3)$. The relative-magnitude statistic $\bar{B}_j$ is kept separate because it directly measures whether control has become large compared with the native update.

4.3 Task-specific source consistency

In dual-source translation, a configuration should also preserve information from the designated sources. At the end of probe window Ω_j, let $L(\cdot)$ be a fixed low-pass operator and let $\mathbf{b}$ and $\mathbf{r}$ denote the base and reference images. We fit the clean prediction with a convex low-frequency mixture,

$$\boldsymbol{\alpha}^* = \arg\min_{\boldsymbol{\alpha}\in\Delta^2} \|L(P_\theta(\mathbf{x}, t)) - \alpha_b L(\mathbf{b}) - \alpha_r L(\mathbf{r})\|_2^2, \tag{9}$$

where $\Delta^2 = \{\boldsymbol{\alpha} : \alpha_b, \alpha_r \geq 0, \alpha_b + \alpha_r = 1\}$. We record the normalized fitting residual $\rho_{\mathrm{fit},j}$ and the maximum source affinity $S_{\max,j}$, defined from negative low-frequency ℓ_2 distance. These two measurements provide a task-level check alongside the generic intervention dynamics.

4.4 Schedule aggregation and screening

For a multi-stage configuration, stage summaries are aggregated conservatively:

$$R_{\mathrm{sched}} = \min_j R_j, \quad B_{\mathrm{sched}} = \max_j \bar{B}_j, \quad \rho_{\mathrm{sched}} = \max_j \rho_{\mathrm{fit},j}, \quad S_{\max,\mathrm{sched}} = \min_j S_{\max,j}. \tag{10}$$

The configuration passes when

$$\begin{aligned}\mathrm{Pass}(\pi) = [R_{\mathrm{sched}} \geq \tau] \wedge [B_{\mathrm{sched}} \leq \kappa_B] \\ \wedge [\rho_{\mathrm{sched}} \leq \kappa_{\mathrm{fit}}] \wedge [S_{\max,\mathrm{sched}} \geq \kappa_S].\end{aligned} \tag{11}$$

We also define a signed margin from the nearest schedule-level boundary,

$$M(\pi) = \min\left\{\frac{R_{\mathrm{sched}} - \tau}{\tau}, \frac{\kappa_B - B_{\mathrm{sched}}}{\kappa_B}, \frac{\kappa_{\mathrm{fit}} - \rho_{\mathrm{sched}}}{\kappa_{\mathrm{fit}}}, \frac{S_{\max,\mathrm{sched}} - \kappa_S}{|\kappa_S|}\right\}. \tag{12}$$

Positive values satisfy every criterion; larger values indicate greater distance from the closest screening boundary. We use this margin to rank retained candidates and to study its relation to completed-sample metrics.

5 Experiments

5.1 Testbed and configuration sweep

We use AFHQ-Dog (Choi et al., 2020) at 256×256 with a pretrained learned-variance diffusion model. The original diffusion process has 1,000 training steps and is respaced to 100 reverse-sampling steps

Table 2: Main values in the configuration sweep. Operator-specific parameters apply only to the corresponding branch.

Variable	Values
stage boundaries	$t_{\text{mid}} \in \{80, 60, 40, 20\}$, $t_{\text{end}} \in \{20, 10\}$
first-stage Gaussian	$\sigma_1 \in \{12, 16, 24, 32\}$, $\lambda_1 \in \{0.05, 0.10, 0.20\}$
second-stage operator	Fourier low-pass, Gaussian, Laplacian
second-stage strength	$\lambda_2 \in \{0, 0.002, 0.005, 0.01, 0.02, 0.04\}$
operator parameters	Fourier cutoff $\{0.1, 0.2, 0.3, 0.4\}$; Gaussian $\sigma_2 \in \{2, 4, 8, 12\}$
controlled source	base or reference
ILVR settings	downsample $\{8, 32, 128\}$; ref. weight $\{0.25, 0.5, 0.75\}$; base weight $\{0, 0.25, 0.5\}$

for inference. Each configuration is evaluated on 20 paired base/reference inputs. The sweep covers three families: one-stage difference injection, one-stage ILVR-style low-frequency replacement, and two-stage difference injection. In the two-stage family, Gaussian low-frequency control is used first, followed by Fourier low-pass, Gaussian, or Laplacian control in a later stage.

The full sweep contains 5,454 configurations: 216 one-stage difference, 54 one-stage replacement, and 5,184 two-stage difference configurations. Across 20 paired inputs this gives 109,080 short-window records. The two-stage grid includes settings with an inactive second stage ($\lambda_2 = 0$) or a degenerate late interval ($t_{\text{mid}} \leq t_{\text{end}}$). Removing these leaves 3,780 effective two-stage configurations and 75,600 records.

5.2 Screening setup and implementation

All experiments use $K = 7$, $\tau = 0.5$, $\kappa_B = 0.5$, $\kappa_{\text{fit}} = 1.5$, and $\kappa_S = -200$. The activity threshold is 10^{-4}. Probe locations follow the active stage; representative two-stage runs start the early and late probes at respaced indices 92 and 29. The code averages E_t, A_t, C_t, and B_t inside each window, then applies the schedule-level min/max aggregation in Eq. (10). E_t is evaluated with a vector–Jacobian product, A_t uses clean predictions from controlled and uncontrolled next states, and the low-frequency source fit is projected onto the two-source simplex.

5.3 Full-sampling validation

We fully sample 90 selected cases: 30 clear passes, 30 clear rejections, 20 boundary cases arranged as 10 matched pass/fail configuration pairs, and 10 inactive-control negative controls. This design gives coverage of the center and boundary of the screening rule. For each final output $\mathbf{y}$, we compute SSIM (Wang et al., 2004) to the base and reference images, low-frequency ℓ_1 distance to each source, and cosine similarity to the reference color histogram. Clear pass/reject groups are compared with two-sided Mann–Whitney U tests. Spearman correlation relates $M(\pi)$ to full-sampling metrics over the 80 active cases, and the 10 boundary pairs are compared with two-sided Wilcoxon signed-rank tests.

6 Results

6.1 Screening behavior across operators

Table 3 reports sample-level acceptance rates. The effective two-stage subset accepts 81.33% of records, so 18.67% are removed before full sampling. The operator choice has a large effect: Gaussian and Fourier low-pass controls pass at 90.39% and 82.67%, while Laplacian control passes at 39.70%. The one-stage experiments show the same pattern, with only 0.21% of Laplacian records passing. High-frequency intervention is more likely to become large relative to the native update or to oppose its direction, which is reflected by B, C, and the response score R.

The generic and task-specific measurements contribute differently across operators. Over the effective two-stage subset, the generic R/B conditions reject 5.65% of records and the low-frequency fit/affinity conditions reject 18.13%; together they reject 18.67%. For Gaussian and Fourier low-pass, the generic conditions reject none of the tested records, while the complete screen rejects 9.61% and

Table 3: Short-window acceptance over the configuration sweep. Two-stage operator rows use effective configurations with $\lambda_2 > 0$ and $t_{\mathrm{mid}} > t_{\mathrm{end}}$.

Family	operator	configs	records	accept (%)
one-stage diff.	all	216	4,320	49.28
one-stage replace	ILVR	54	1,080	68.70
two-stage diff.	all scanned	5,184	103,680	71.80
two-stage effective	Gaussian	1,680	33,600	90.39
two-stage effective	Fourier LPF	1,680	33,600	82.67
two-stage effective	Laplacian	420	8,400	39.70
two-stage effective	**all**	**3,780**	**75,600**	**81.33**

Table 4: Full-sampling validation. Values are mean±std for 30 clear pass and 30 clear reject cases; p_U is two-sided Mann–Whitney. Spearman ρ_s uses 80 active samples.

Metric	pass	reject	p_U	ρ_s	p_ρ
base SSIM ↑	$.382 \pm .341$	$.058 \pm .191$	1.87×10^{-5}	.450	2.81×10^{-5}
ref. SSIM ↑	$.608 \pm .311$	$.113 \pm .204$	2.03×10^{-7}	.645	1.05×10^{-10}
base LF L_1 ↓	$.196 \pm .096$	$.274 \pm .094$	1.12×10^{-2}	-.189	9.34×10^{-2}
ref. LF L_1 ↓	$.127 \pm .085$	$.304 \pm .079$	1.10×10^{-8}	-.724	3.29×10^{-14}
ref. hist. cos. ↑	$.665 \pm .212$	$.412 \pm .194$	2.13×10^{-5}	.461	1.72×10^{-5}

17.33%. For Laplacian, the generic conditions reject 50.85%, and the complete screen rejects 60.30%. The split is intuitive: the generic dynamics expose aggressive high-frequency control, while source consistency catches smoother configurations whose local trajectory remains stable but moves away from the desired source content.

6.2 Threshold sensitivity

Changing the response threshold τ from 0.3 to 0.8, with the other thresholds fixed, changes effective two-stage acceptance only from 81.56% to 81.13%. At $\tau = 0.5$, the individual failure rates are 5.65% for R, 5.40% for B, 12.13% for the low-frequency fit, and 9.28% for source affinity; the conditions overlap. These numbers show that the current operating point is driven more by the low-frequency consistency terms and the operator-dependent magnitude/conflict behavior than by small changes in the response threshold.

6.3 Relation to completed-sample consistency

Table 4 compares the 30 clear pass and 30 clear reject cases after full sampling. All five source-consistency metrics separate the two groups. Reference SSIM averages 0.608 for accepted cases and 0.113 for rejected cases, while reference low-frequency distance changes from 0.127 to 0.304. The continuous screening margin also tracks these measurements over the 80 active cases: $\rho = 0.645$ for reference SSIM and $\rho = -0.724$ for reference low-frequency distance. Base low-frequency distance has a weaker correlation ($\rho = -0.189$, $p = 0.093$), which indicates that the margin is more closely tied to some aspects of source consistency than others.

The boundary pairs show a smaller but consistent effect. The passing configuration has higher base and reference SSIM in 8 of 10 matched pairs; the Wilcoxon p values are 0.0488 and 0.0273. Low-frequency and histogram differences are smaller near the boundary. All 10 inactive-control cases fail the screen, as expected from their negligible intervention response.

Artifact-oriented image statistics give additional context. High-frequency absolute response is 0.039 ± 0.054 for the clear-pass group and 0.288 ± 0.569 for the clear-reject group ($p = 0.0224$); gradient absolute response is 0.0158 ± 0.0171 versus 0.0843 ± 0.1511 ($p = 0.00831$). Saturation, entropy, and the high-frequency-to-gradient ratio remain similar across the two groups. The screen therefore separates configurations through source consistency and intervention behavior more clearly than through generic image statistics alone.

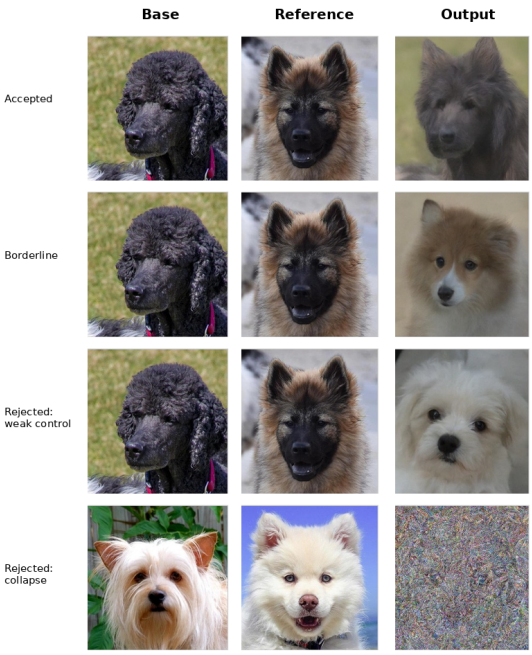

Figure 2: Representative completed cases. Columns are base image, reference image, and final output. Rows show a clear pass, a boundary case, a visually plausible rejection with weak source correspondence, and a rejected Laplacian configuration with high-frequency collapse.

6.4 Representative outputs

Figure 2 shows four completed examples spanning the screening range. The accepted low-pass configuration preserves source information and remains visually stable. The boundary example produces a plausible output with a smaller margin. One rejected configuration remains visually natural but loses source correspondence, while the rejected Laplacian case shows severe high-frequency distortion. These examples match the quantitative split between source-consistency failures and unstable high-frequency control.

6.5 Cost of screening

Let C_F be the cost of one full trajectory, C_P the cost of all short-window probes for a candidate, and r the retained fraction. The relative cost of screening followed by full sampling of retained candidates is

$$\frac{C_{\text{screen}}}{C_{\text{exhaustive}}} = \frac{C_P + rC_F}{C_F} = \frac{C_P}{C_F} + r. \tag{13}$$

A configuration family benefits computationally when its rejection fraction exceeds the probe-cost ratio C_P/C_F. The current implementation uses extra clean-prediction evaluations for A_t and a vector–Jacobian product for E_t, so probe cost is larger than the cost of the same number of ordinary denoising steps. Table 5 converts the observed retention rates into break-even probe budgets. For example, Laplacian screening remains favorable whenever the probe costs less than 60.30% of a full trajectory, while Gaussian screening requires a substantially cheaper probe.

7 Further Analysis

7.1 Diagnostic decomposition

Table 6 separates the generic dynamics from the dual-source consistency checks. The difference between the two groups is especially clear for Gaussian and Fourier low-pass control. Their short

Table 5: Break-even probe budget implied by the observed two-stage retention rates. The last column is the largest C_P/C_F for which Eq. (13) is below exhaustive sampling cost.

Operator	retained (%)	break-even C_P/C_F (%)
Gaussian	90.39	9.61
Fourier LPF	82.67	17.33
Laplacian	39.70	60.30
All effective	81.33	18.67

Table 6: Failure decomposition for effective two-stage configurations. Entries are rejection rates; conditions within a column may overlap across individual criteria.

Operator	generic R/B	source consistency	full screen
Gaussian	0.00	9.61	9.61
Fourier LPF	0.00	17.33	17.33
Laplacian	50.85	55.37	60.30
All effective	5.65	18.13	18.67

trajectories are locally stable under the generic R/B criteria, yet 9.61% and 17.33% of records are removed once source consistency is considered. Laplacian behaves differently: half of its records already fail the generic dynamics, and the source checks raise the total rejection rate to 60.30%. This split explains why the same screen can handle both visibly unstable control and visually plausible source drift.

The two diagnostic groups also suggest a useful ordering for implementation. Generic dynamics can be computed for any controller whose effective intervention u_t is observable. Task-specific checks can then be added when the application supplies a meaningful notion of consistency, such as source preservation in image translation. This keeps the generic screen reusable while allowing the final decision rule to reflect the task being optimized.

7.2 Boundary behavior and ranking

The signed margin $M(\pi)$ is useful beyond the binary pass/fail decision because it measures how close a configuration is to its nearest violated criterion. In the clear pass and reject groups, the large margin separation is reflected by substantial differences in reference SSIM and low-frequency distance. Near the boundary, the effect becomes smaller: among the 10 matched pairs, the passing member has higher base and reference SSIM in 8 pairs, while low-frequency and histogram differences are less consistent. This pattern supports a simple search strategy: use the binary screen to remove clear failures, then use the margin to prioritize the retained set for full evaluation.

The 10 inactive-control cases provide a complementary sanity check. Their negligible intervention produces weak local response, and all 10 are filtered before completion. At the other extreme, rejected Laplacian configurations often have large relative magnitude or directional conflict. The screen therefore covers both ends of the useful-control range: interventions that do too little and interventions that disrupt the native update.

7.3 Calibration for a new search space

For a new backbone or controller, the same diagnostics can be used with a small calibration set. A practical procedure is to fully sample a subset of configurations, choose thresholds that retain the desired fraction of satisfactory outputs, and then freeze those thresholds for the remaining inputs. The resulting evaluation can report three quantities together: the fraction of candidates removed, the fraction of satisfactory configurations retained, and the best downstream utility reached under a fixed full-sampling budget. This connects the local screening rule directly to the goal of configuration search.

Probe length can be calibrated in the same way. Shorter windows reduce C_P but provide fewer observations per stage; longer windows give more stable summaries but consume more of the

trajectory. Because the four generic diagnostics are defined step by step, K can be varied without changing the method. The seven-step setting used here provides a concrete operating point for the current 100-step sampler and can serve as the reference when comparing cheaper or longer probes.

8 Discussion

The experiments reveal two complementary roles for short-window measurements. First, the generic dynamics are effective when a configuration fails through a local mechanism that is visible immediately, such as a large or antagonistic high-frequency intervention. This behavior is prominent for Laplacian control. Second, smooth low-frequency operators can follow stable local dynamics while gradually drifting away from the intended source. In that regime, the task-specific low-frequency fit and affinity terms provide the stronger signal. The resulting screen is therefore best viewed as a small set of interpretable checks whose relative importance depends on the controller.

The current thresholds are fixed operating points for the AFHQ-Dog dual-source setting. Applying the same diagnostics to another backbone, sampler, or control family requires calibration of the schedule-level thresholds, but the diagnostic definitions remain unchanged. A practical calibration procedure is to choose thresholds on a set of inputs with full-sampling labels, freeze them, and then measure candidate retention and downstream search quality on disjoint inputs. This also makes it possible to compare different probe lengths or cheaper approximations of E_t and A_t under the same compute budget.

The full-sampling study is stratified across clear, boundary, and inactive cases, which is useful for studying how the screen behaves around its decision surface. A larger randomly sampled validation set would additionally support population-level estimates of good-candidate retention. Extending the experiments to latent text-to-image diffusion and adapter-based controllers is another direct next step, especially because those settings offer larger configuration spaces and different types of intervention.

9 Conclusion

We introduced ShortWin, a training-free method for screening controlled-diffusion configurations from short local denoising trajectories. The method measures response, propagation, directional conflict, and relative intervention magnitude, combines these signals across control stages, and can incorporate a task-specific consistency check when source constraints are available. A 109,080-record sweep shows clear operator-dependent screening behavior, and full-sampling validation links the screening margin to source consistency in completed outputs. The results indicate that local control dynamics can reduce the number of configurations that need full evaluation and provide an interpretable basis for configuration search.

Reproducibility Statement

The paper specifies the probe length, diagnostic equations, thresholds, aggregation rule, configuration grid, validation split, and statistical tests used for the reported results. The source package includes the diffusion implementation used to compute the probes. Reported numerical values were recomputed from the supplied experiment logs before manuscript generation; private file paths and source-image identifiers are omitted from the submission package to preserve anonymity.

AI Use Statement

Generative AI tools were used for manuscript organization, literature discovery, language editing, and consistency checks across the experiment logs, code, and manuscript. They were not used to generate the reported experimental outputs. The authors verified the claims, citations, numerical results, and submitted artifacts.

REFERENCES

Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 843–852, 2023.

Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. ILVR: Conditioning method for denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14367–14376, 2021.

Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. StarGAN v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8188–8197, 2020.

Huanlei Guo, Hongxin Wei, and Bingyi Jing. Toward early quality assessment of text-to-image diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 38410–38419, 2026.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.

Bukun Huang, Benlei Cui, Zhizeng Ye, Xuemei Dong, Tuo Chen, Hui Xue, Dingkang Yang, Longtao Huang, Haiwen Hong, and Jingqun Tang. Diffusion probe: Generated image result prediction using cnn probes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 35926–35935, 2026.

Tuomas Kynkäänniemi, Miika Aittala, Tero Karras, Samuli Laine, Timo Aila, and Jaakko Lehtinen. Applying guidance in a limited interval improves sample and distribution quality in diffusion models. *arXiv preprint arXiv:2404.07724*, 2024.

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided image synthesis and editing with stochastic differential equations. In *International Conference on Learning Representations*, 2022.

Sicheng Mo, Fangzhou Mu, Kuan Heng Lin, Yanli Liu, Bochen Guan, Yin Li, and Bolei Zhou. FreeControl: Training-free spatial control of any text-to-image diffusion model with any condition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7465–7475, 2024.

Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhang, Zhongang Qi, and Ying Shan. T2I-Adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 4296–4304, 2024.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8162–8171, 2021.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.

Sixian Wang, Zhiwei Tang, and Tsung-Hui Chang. Diffusion sampling path tells more: An efficient plug-and-play strategy for sample filtering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 36030–36039, 2026.

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.

Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. IP-Adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721*, 2023.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023.

XiangCheng Zhang, Haowei Lin, Haotian Ye, James Y. Zou, Jianzhu Ma, Yitao Liang, and Yilun Du. Inference-time scaling of diffusion models through classical search. In *International Conference on Learning Representations*, 2026.