Diffusion-supplemented Implicit Layers: Operator Smoothing for more Robust Implicit Solvers

Bader Rasheed[1] **Dinislam Gabitov**[2] **Anastasia Antsiferova**[2 3] **Dmitriy S. Vatolin**[2 3]
[1]Research Center of the Artificial Intelligence Institute, Innopolis University, Innopolis, Russia
[2]Laboratory of Innovative Technologies for Processing Video Content, Innopolis University, Innopolis, Russia [3]MSU AI Institute, Moscow, Russia
`{b.rasheed,d.gabitov}@innopolis.university`
`{aantsiferova,dmitriy}@graphics.cs.msu.ru`

Abstract

Implicit networks compute hidden states as fixed points. When the implicit map is poorly conditioned, solvers slow or fail. We propose *Diffusion-Supplemented Implicit Layers* (DSIL): insert a few denoising steps on the latent before each evaluation of the map. Under standard Lipschitz assumptions in a common metric, this preconditioning reduces the effective Lipschitz constant of the composed map, yielding stronger contraction; with a true proximal denoiser the contraction factor is explicitly tunable by the step size. On CIFAR-10 with a SODEF head, DSIL provides modest robustness gains without adversarial training. DSIL is architecture-agnostic and complements existing stabilization methods.

1 Introduction

Implicit neural networks - including Neural ODEs, deep equilibrium models (DEQs) and SODEF - define hidden states through an implicit function rather than via explicit multi-layer composition. This paradigm affords continuous-depth representations and reduces memory cost but introduces algorithmic challenges: one must ensure that a solution exists, that it is stable with respect to perturbations and that the numerical solver converges in a reasonable number of iterations. A common requirement for convergence is that the implicit map $T : \mathbb{R}^d \to \mathbb{R}^d$ be *contractive* or more generally *averaged* [1]. However, implicit networks in practice may exhibit large Lipschitz constants or near-degenerate Jacobians, leading to slow or divergent fixed-point iterations [2]. In adversarial settings, high sensitivity to perturbations further undermines reliability [3]. Our work aims to improve contraction and stability of implicit networks by leveraging recent advances in diffusion modeling. For the current step of the research, we position this study as *proof-of-concept*: our focus is solver conditioning and transparent limitations rather than state-of-the-art robustness.

Contributions. We introduce *Diffusion-Supplemented Implicit Layers* (DSIL): a few denoising (reverse-diffusion) steps applied to features before each evaluation of an implicit map T. Our contributions are:

- **Operator view.** We model the denoiser as a resolvent $D_\sigma = (\mathrm{Id} + \sigma A)^{-1}$; it is firmly nonexpansive and, if A is μ-strongly monotone, $\mathrm{Lip}(D_\sigma) \le (1 + \sigma\mu)^{-1}$.
- **Contraction bound.** For an L-Lipschitz implicit map T, the composition $T \circ D_\sigma$ satisfies $\mathrm{Lip}(T \circ D_\sigma) \le L/(1 + \sigma\mu)$. When $L/(1 + \sigma\mu) < 1$, standard fixed-point iterations enjoy linear convergence.
- **Spectral intuition.** Jacobian factorization $J_{T\circ D_\sigma} = J_T(\cdot)\, J_{D_\sigma}$ shows diffusion acts as a low-pass smoother that shrinks high-frequency modes and improves conditioning.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Constrained Optimization for Machine Learning.

- **Proof-of-concept.** On CIFAR-10 with a SODEF head, DSIL reduces solver iterations and yields small robustness gains without adversarial training, at modest overhead.

2 Background

Metric Lipschitzness and resolvents. Let $\|u\|_M := \sqrt{u^\top M u}$ for $M \succ 0$. A map $T : \mathbb{R}^d \to \mathbb{R}^d$ is L–Lipschitz in $\|\cdot\|_M$ if $\|Tz - Tz'\|_M \le L\|z - z'\|_M$; it is *nonexpansive* if $L \le 1$ and *contractive* if $L < 1$. For an operator A and stepsize $\sigma > 0$, its resolvent $J_{\sigma A} = (\mathrm{Id} + \sigma A)^{-1}$ is single–valued and *firmly nonexpansive* when A is monotone; if A is μ–strongly monotone, $J_{\sigma A}$ is $(1+\sigma\mu)^{-1}$–Lipschitz in $\|\cdot\|_M$ [4]. Proximal operators $\mathrm{prox}_{\sigma R} = (\mathrm{Id} + \sigma\partial R)^{-1}$ (for closed, proper, convex R) are resolvents.

Implicit layers and fixed points. DEQs [5] and related implicit architectures compute $z^\star$ from $z^\star = T(z^\star, x)$ instead of composing explicit layers; Neural ODEs integrate $\dot{z}(t) = f(z(t), x)$; SODEF adds Lyapunov–stable equilibria [6]. If T is contractive (in some metric), Banach's theorem ensures existence, uniqueness, and *linear* convergence of $z_{k+1} = T(z_k)$. For broader nonexpansive/averaged settings, Krasnosel'skii–Mann iterations can converge under step–size conditions [1]. In practice, Anderson acceleration is used to speed up DEQ-based implicit models.

Diffusion/denoising as operators. Score–based diffusion trains $s_\theta(x,t) \approx \nabla_x \log p_t(x)$ and samples via a reverse SDE/ODE [7]; a discretized reverse step acts as a denoising operator. Plug–and–play methods replace a proximal map with a learned denoiser [8]. We consider two realizations for D_σ: (i) a *proximal/resolvent* giving formal nonexpansiveness and explicit Lipschitz constants; (ii) a *learned* denoiser with *enforced/empirical* Lipschitz bound (e.g., spectral normalization), acknowledging generic denoisers need not be resolvents [9]. In both cases, the same metric $\|\cdot\|_M$ is used to assess Lipschitzness of T and D_σ.

3 Method & Theory: Diffusion Preconditioning for Implicit Layers

Setup. Given $T(\cdot, x; \theta)$ defining DEQ: $z^\star = T(z^\star, x; \theta)$ or ODE: $\dot{z}(t) = T(z(t), x; \theta)$, we insert k denoising steps D_σ on the latent before each evaluation of T:

$$z \leftarrow D_\sigma^{(k)}(z) := \underbrace{D_\sigma \circ \cdots \circ D_\sigma}_{k \text{ times}}(z), \qquad z \leftarrow T(z, x; \theta),$$

and solve the fixed point with Anderson acceleration or calculate an integral with numerical integration. Backpropagation uses the implicit function theorem [5]. The extra cost is k calls to D_σ per solver iteration; empirically $k \le 3$.

Assumptions (minimal, verifiable). Let $\|u\|_M = \sqrt{u^\top M u}$ with $M \succ 0$. We assume:

A1 T is L–Lipschitz in $\|\cdot\|_M$.

A2 D_σ is κ–Lipschitz in the same metric with $\kappa \le 1$.

*How to meet **A2**:* **(A) Proximal:** $D_\sigma = (\mathrm{Id} + \sigma A)^{-1}$ with A μ–strongly monotone gives $\kappa = (1 + \sigma\mu)^{-1} < 1$ [4]. **(B) Learned:** enforce/measure $\kappa \le 1$ (e.g., spectral normalization); we report empirical $\hat{\kappa}$.

Core guarantee (composition contraction). Let $T_\sigma := T \circ D_\sigma^{(k)}$. Since $D_\sigma^{(k)}$ is κ^k–Lipschitz in $\|\cdot\|_M$, [Sufficient condition] T_σ is $L\kappa^k$–Lipschitz in $\|\cdot\|_M$. If $L\kappa^k < 1$, then T_σ has a unique fixed point and $z_{t+1} = T_\sigma(z_t)$ converges linearly with factor $L\kappa^k$. $\mathrm{Lip}(T \circ D_\sigma^{(k)}) \le L\,\kappa^k$ by submultiplicativity; apply Banach's fixed–point theorem.

Iteration complexity and cost trade–off. With $z^\star$ the fixed point of T_σ, $\|z_t - z^\star\|_M \le (L\kappa^k)^t \|z_0 - z^\star\|_M$; to reach $\|z_t - z^\star\|_M \le \varepsilon$ it suffices that $t \ge \log(\varepsilon / \|z_0 - z^\star\|_M) / \log(L\kappa^k)$. DSIL reduces solver iterations (smaller $L\kappa^k$) while adding k denoiser calls per iteration; we choose the smallest (k, σ) achieving a clear contraction.

Table 1: Proof-of-concept robustness on CIFAR-10 ($\varepsilon = 8/255$). Bold and underline highlights best and second best performance on each experiment respectively. Diffusions are trained without adversarial data.

Model	**Clean**	**FGSM**	**PGD**	**AutoAttack**
ResNet32	**91.3**	12.1	0.35	0.00
SODEF	<u>85.7</u>	37.3	20.5	0.05
SODEF + DiffODE (ours)	84.9	**42.6**	**27.1**	0.72
SODEF + Diff (ours)	85.5	<u>38.7</u>	20.9	<u>2.02</u>
SODEF + DiffODE w/ DS (ours)	85.5	36.3	17.3	0.43
SODEF + Diff w/ DS (ours)	85.2	38.2	<u>21.2</u>	**2.49**

Scope and caveats. *IFT/conditioning.* Shrinking $\mathrm{Lip}(T_\sigma)$ typically correlates with a smaller $\rho(J_{T_\sigma})$ and improved IFT conditioning, but it does not on its own guarantee $(I - J_{T_\sigma}(z^\star))$ invertibility if an eigenvalue is 1. *Interleaved variants.* Interleaving D_σ *inside* solver updates (DiffODE) changes the effective operator being iterated and falls outside Prop. 3; we evaluate it empirically. *Heuristics.* Spectral "low–pass" and small–σ expansions aid intuition and are placed in the Appendix; we do not rely on them for guarantees.

4 Experiments

Our empirical goal is merely to illustrate the theoretical claims; extensive tuning or adversarial training is beyond our scope. We adopt the SODEF architecture from [6] on CIFAR-10. The baseline uses a ResNet-32 backbone with a SODEF head. We insert diffusion preconditioning either immediately before the implicit head (**Diff**) or interleaved inside the implicit function (**DiffODE**). For diffusion, we use a three-step discrete reverse process with step size $\sigma = 0.02$ as a denoising score network. Diffusion denoiser is selected as small 3 layer mlp network with hidden size 128. We also test the Drift towards Stability (DS, check Appendix A) with $\lambda = 0.02$. Hyper-parameters are borrowed from [6]; first, we train SODEF in 3 stages, then an additional stage is used to train the diffusion network for 100 epochs with learning rate 10^{-2} using Adam optimizer. At inference, we measure the model on clean examples and evaluate robustness using FGSM, PGD (step size $2/255$, four iterations), and AutoAttack with $\varepsilon = 8/255$ as recommended for reliable evaluation [3].

4.1 Robustness results

Table 1 summarises clean accuracy and robustness under various attacks. While clean accuracy drops slightly when diffusion is applied, robustness improves modestly: for FGSM attacks the accuracy increases from 37.3% for SODEF to 42.6% for SODEF+DiffODE; for PGD attacks the accuracy increases from 20.5% for SODEF to 27.1% for SODEF+DiffODE, and AutoAttack accuracy improves from nearly zero to 2.49% when using diffusion with DS. These numbers are small because no adversarial training is used; nonetheless, they indicate that diffusion smoothing yields larger attraction basins and hinders gradient-based attacks. Runtime overhead is approximately $1.5\times$ for Diff and $4\times$ for DiffODE comparing to baseline SODEF.

5 Discussion

Our theoretical results establish that diffusion preconditioning lowers an *upper bound* on contraction and *can* accelerate convergence of implicit layers under reasonable assumptions. The trade-off is an additional computational overhead proportional to the number of diffusion steps. Our experiments, though limited in scale, support the theory: robustness improves slightly without adversarial training. Limitations include the idealised assumptions on diffusion being firmly nonexpansive and the small scale of experiments. Future work should explore learned few-step samplers, adaptive diffusion schedules, integration with adversarial training and applications to stiff ODEs and large-scale models.

6 Limitations

Our analysis relies on a resolvent/strong monotonicity model for the denoiser; learned diffusion steps may only approximate this. Our empirical validation is currently limited to a single architecture on CIFAR-10. The generalizability of DSIL to other implicit models and larger-scale problems remains an open question. Resource overhead from denoisers should be considered. We also do not compare against monDEQ or 1-Lipschitz heads; DSIL is intended to be orthogonal and modular, which we leave to future work.

7 Reproducibility statement

We describe all architectural details, training schedules and hyper-parameters used in our experiments. The SODEF baseline follows [6]. We train the diffusion network for four stages with learning rate 10^{-2} on ResNet-32 features. We evaluate robustness using FGSM, PGD with step size $2/255$ and four iterations, and AutoAttack with $\varepsilon = 8/255$ [3]. Source code and pre-trained models will be released upon acceptance to ensure full reproducibility.

8 Broader impact and ethics

Implicit networks have the potential to improve robustness and memory efficiency in machine learning. Our work proposes using diffusion models as an operator-level preconditioner, which could enhance reliability when training on unreliable data. DSIL is simple to integrate, improves conditioning, and is complementary to adversarial training. However, diffusion networks are computationally expensive. Additionally, our method leverages generative models that could inadvertently memorise sensitive information. Properly anonymising training data and following responsible AI practices remain critical.

9 Computing resources

All models were trained on GPU; all experiments require less than 4GB VRAM and 8GB RAM.

10 NeurIPS checklist

We follow the NeurIPS 2025 requirements: the paper is anonymised, uses the official style file, and the main text does not exceed four pages. The appendix contains extra details and proofs. We discuss limitations and potential societal impacts. We will release code and models. Our evaluation uses recommended robustness baselines such as AutoAttack [3].

Acknowledgment. The research was supported by the Ministry of Economic Development of the Russian Federation (agreement No. 139-10-2025-034 dd. 19.06.2025, IGK 000000C313925P4D0002)

References

[1] H. H. Bauschke, P. L. Combettes, Correction to: convex analysis and monotone operator theory in hilbert spaces, in: Convex analysis and monotone operator theory in Hilbert spaces, Springer, 2020, pp. C1–C4.

[2] M. Revay, R. Wang, I. R. Manchester, Lipschitz bounded equilibrium networks, arXiv preprint arXiv:2010.01732 (2020).

[3] F. Croce, M. Hein, Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks, in: International conference on machine learning, PMLR, 2020, pp. 2206–2216.

[4] N. Parikh, S. Boyd, et al., Proximal algorithms, Foundations and trends® in Optimization 1 (3) (2014) 127–239.

[5] S. Bai, J. Z. Kolter, V. Koltun, Deep equilibrium models, Advances in Neural Information Processing Systems 32 (2019).

[6] Q. Kang, Y. Song, Q. Ding, W. P. Tay, Stable neural ode with lyapunov-stable equilibrium points for defending against adversarial attacks, Advances in Neural Information Processing Systems 34 (2021) 14925–14937.

[7] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, B. Poole, Score-based generative modeling through stochastic differential equations, International Conference on Learning Representations (2021).

[8] S. V. Venkatakrishnan, C. A. Bouman, B. Wohlberg, Plug-and-play priors for model based reconstruction, in: 2013 IEEE global conference on signal and information processing, IEEE, 2013, pp. 945–948.

[9] P. Salmon, J.-M. Morel, B. Sander, Image denoising by non-local means: A review, SIAM Journal on Imaging Sciences 6 (3) (2013) 944–978.

A Heuristics and Extensions

Local contraction of learned denoisers (heuristic). For a learned step of the form $x \mapsto x + \eta\, s_\theta(x,t)$ (reverse-diffusion/denoise), a first-order expansion gives $J_{D_\sigma}(x) \approx I + \eta\, \nabla s_\theta(x,t)$. Local contraction in a metric $\|\cdot\|_M$ holds if $\|I + \eta\, \nabla s_\theta(x,t)\|_M \leq 1$ in a neighbourhood, e.g., when the symmetric part of ∇s_θ is negative semidefinite on average and η is small. This motivates an *empirical* shrinkage factor (sometimes written $q(\sigma)$), but it is *not* a global Lipschitz bound and may fail outside that neighbourhood.

Averagedness and KM iterations (with conditions). Our main text relies only on contraction via Lipschitz constants. If one wishes to use Krasnosel'skii–Mann (KM) theory, composition requires extra structure [1]. Two useful special cases (not used in our guarantees) are:

1. **Commuting averaged maps.** If T is α-averaged and D_σ is firmly nonexpansive (hence $1/2$-averaged) *and* they commute (or satisfy suitable cocoercivity/compatibility), then $T \circ D_\sigma$ remains averaged, enabling KM with $\mathcal{O}(1/k)$ rates.
2. **Forward–backward form.** If $T = \mathrm{Id} - \tau\nabla f$ with f convex, L-smooth and $\tau \in (0, 2/L)$, and $D_\sigma = \mathrm{prox}_{\sigma g}$ for convex g, then $T \circ D_\sigma$ matches a forward–backward operator, which is averaged under standard step sizes [1].

Outside such conditions we do *not* claim averagedness of $T \circ D_\sigma$.

Interleaving diffusion within the solver (DiffODE). Interleaving D_σ inside solver updates produces a *non-stationary* iteration whose effective operator changes with t. General contraction/averagedness guarantees do not directly apply. Convergence may still hold under non-stationary fixed-point theory when each iterate uses averaged maps with parameters uniformly bounded < 1 and a common fixed point; verifying these conditions is problem-specific [1]. We therefore report DiffODE results as empirical.

Estimating and enforcing Lipschitz constants (practice). For the learned denoiser path, we (i) *enforce* $\kappa \leq 1$ via spectral normalization / 1-Lipschitz architectures, and (ii) *report* an empirical $\hat{\kappa}$:

- *Metric:* choose M (e.g., diagonal or layerwise) and estimate norms in $\|\cdot\|_M$.
- *Power iteration:* estimate $\sup_{\|v\|_M=1} \|J_{D_\sigma}(z)v\|_M$ by jvp/vjp on random z; aggregate (max/quantile) over a validation batch.
- *For T:* similarly estimate L (or a high quantile) to assess $L\kappa^k$ and relate it to observed solver iterations.

These diagnostics connect the theoretical factor $L\kappa^k$ to practice.

Adversarial evaluation with stochastic denoisers. If D_σ is stochastic (e.g., reverse diffusion with noise), evaluation should either (i) fix the random seed during attacks, or (ii) use an expectation-over-transforms (EOT) attack to avoid gradient masking. We follow this in our evaluation setup.

Runtime accounting and break-even. Let t_0 be mean solver iterations (baseline) and t_{ds} with DSIL, each iteration costing c_T for T and c_D per denoise call. Baseline cost: $t_0\, c_T$. DSIL cost: $t_{\mathrm{ds}}\,(c_T + k\, c_D)$. DSIL is faster when

$$t_{\mathrm{ds}}/t_0 \;<\; \frac{c_T}{c_T + k\, c_D}.$$

This clarifies how small k and light D_σ must be to realize speedups alongside improved conditioning.

Spectral intuition (non-claim). Empirically, $J_{T_\sigma} = J_T(\cdot)\, J_{D_\sigma}$ often shows reduced estimated spectral radius and damped "high-frequency" modes in latent space, aligning with fewer solver iterations. We present this as *intuition and evidence*, not as a general theorem.

Drift towards a Stable point (DS) - conditional local descent. Let V be L_V-smooth and μ-strongly convex in a neighbourhood of a class equilibrium $\hat{z}_c$ in the M-metric. Consider $D_\sigma^{\mathrm{DS}}(z) = D_\sigma(z) - \lambda\, \nabla_M V(z)$. If D_σ is firmly nonexpansive in $\|\cdot\|_M$ and $\lambda \in (0, 2\mu/L_V)$, then, locally,

$$V\big(D_\sigma^{\mathrm{DS}}(z)\big) - V(z) \;\le\; -c\,\lambda\, \|\nabla_M V(z)\|_{M^{-1}}^2 + \mathcal{O}(\lambda^2),$$

for some $c > 0$. This is a *variant* requiring a stronger structure than the main contraction result; we treat it as an empirical ablation.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Our abstract and introduction clearly state our theory and contribution. The experiments Section 4 support our claim.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We have clearly stated limitations of our method in Section 6.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Our paper provided all proofs and states assumptions on our theory.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: All details provided in Section 7.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Data and code will be made publicly available once anonymization is no longer required.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: All experiment details provided in Secton 4.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: We have limited time and resources to make multiple runs on the experiments. Further research on this topic will be conducted with a statistical significance report.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Resources needed to reproduce experiment are stated in Section 9.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This research conforms with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: See Section 8.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: Our models are trained on open-access data and have no risk for misuse.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All used datasets, model variants and theory provided are credited in the paper.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: Not applicable.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: Not applicable.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: Not applicable.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: Not applicable.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.