

Thinking with Looped Flows

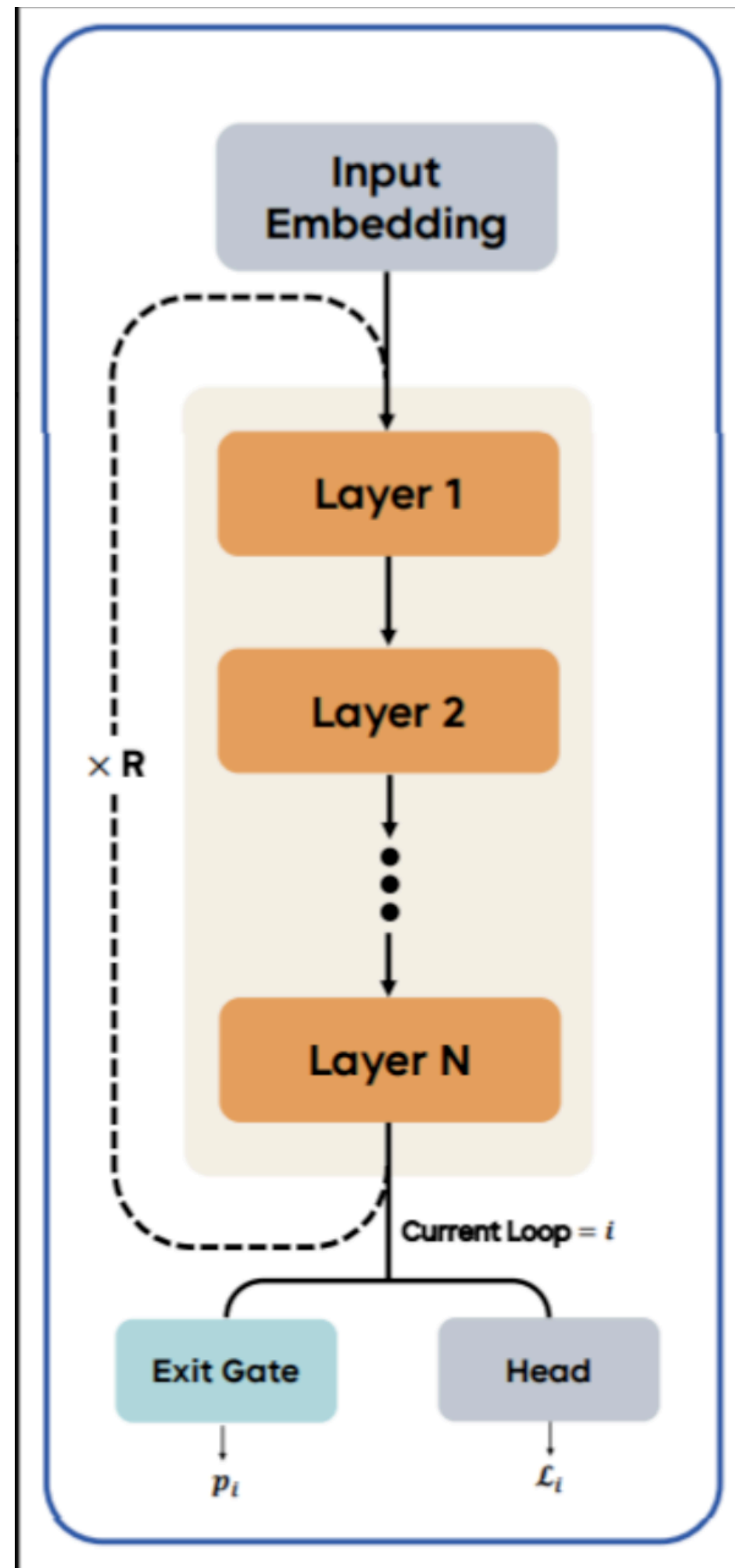
Released on September 10, 2026

Srinitish Srinivasan

Motivation

- Transformers have fixed compute depth
- Loops reuse the same transformer block
- Latent reasoning. Intermediate computation happens in the latent space instead of chain of thought tokens.
- $z_{k+1} = F_{\theta}(z_k, c)$

Looped Models



$$\mathcal{L}_{\text{loop}}(f, g) := \mathbb{E}_{\mathbf{c}, \mathbf{x}} \left[\sum_i \text{CE}(\hat{\mathbf{x}}_i, \mathbf{x}) \right], \quad \hat{\mathbf{x}}_i = g(\mathbf{z}_i), \quad \mathbf{z}_i = f(\text{sg}(\mathbf{z}_{i-1}); \mathbf{c})$$

-> There is no learning signal for past state!!

—> We need some useful information coming from the previous state .

—> Also, how do we know that $f(h_{t+1})$ is better than $f(h_t)$?

—> Also, the current setup is deterministic.

Flow Matching

$$\mathbf{x}_1 \sim p_1(\cdot \mid c),$$

$$I_t := (1 - t)\mathbf{x}_0 + t\mathbf{x}_1,$$

$$\dot{\mathbf{x}}_t = b_t(\mathbf{x}_t; c),$$

$$\mathbf{x}_{t_{i+1}} = \mathbf{x}_{t_i} + (t_{i+1} - t_i)\hat{b}_{t_i}(\mathbf{x}_{t_i}; c),$$

$$\bullet \quad D_t(\mathbf{x}; c) := \mathbb{E} [\mathbf{x}_1 \mid I_t = \mathbf{x}, c],$$

$$b_t(\mathbf{x}; c) = \frac{D_t(\mathbf{x}; c) - \mathbf{x}}{1 - t}.$$

$$\mathbf{x}_0 \sim p_0 = \mathcal{N}(0, \sigma^2 I),$$

$$I_t \sim p_t(\cdot \mid c),$$

$$t \in [0, 1],$$

$$0 = t_0 < \dots < t_n = 1,$$

Training

Algorithm 1 Training

Input: Steps k , time sampler μ , initial state $\mathbf{z}_0$

```
1: repeat  
2:    $(\mathbf{c}, \mathbf{x}_1) \sim p_{\text{data}}; (t_0, \dots, t_k) \sim \mu$   
3:    $\mathbf{x}_0 \sim p_0; \mathbf{z}_{t_0} \leftarrow \mathbf{z}_0$   
4:   for  $i = 0, \dots, k - 1$  do  
5:      $I_{t_i} \leftarrow (1 - t_i)\mathbf{x}_0 + t_i\mathbf{x}_1$   
6:      $(\hat{\mathbf{x}}_{t_i}, \mathbf{z}_{t_{i+1}}) \leftarrow \hat{D}_{t_i}(I_{t_i}, \text{sg}(\mathbf{z}_{t_i}); \mathbf{c})$   
7:      $\hat{\mathbf{q}} \leftarrow q(\mathbf{z}_{t_{i+1}}); \mathbf{q} \leftarrow \mathbf{1}(\mathbf{x}_1 = \text{round}(\hat{\mathbf{x}}_{t_i}))$   
8:     Gradient step with  $\text{CE}(\hat{\mathbf{x}}_{t_i}, \mathbf{x}_1) + \lambda \text{BCE}(\hat{\mathbf{q}}, \mathbf{q})$   
9:     if  $\hat{\mathbf{q}} > 1/2$  then break  
10:  end for  
11: until converged  
12: return  $\hat{D}$ 
```

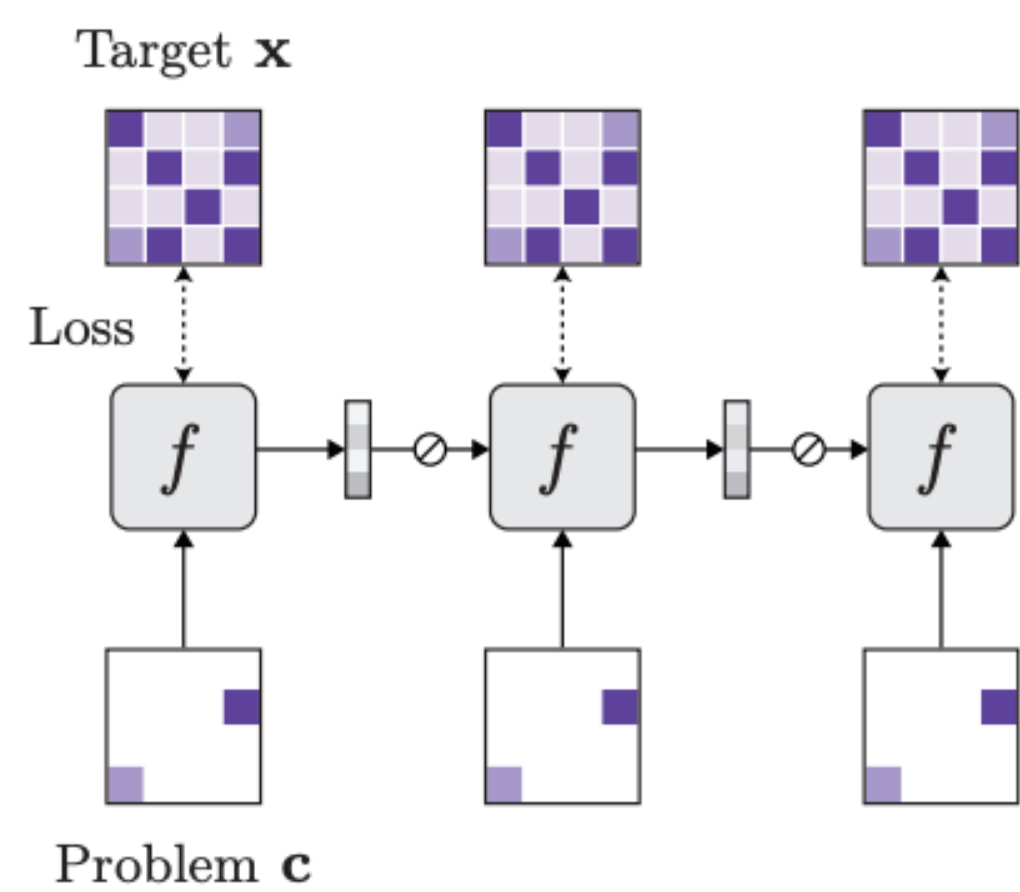
$$\mathcal{L}_{\text{LF}}(\hat{D}) := \mathbb{E}_{\mathbf{c}, \mathbf{x}_0, \mathbf{x}_1} \mathbb{E}_{(t_0, \dots, t_k) \sim \mu} \left[\sum_{i=0}^{k-1} \text{CE}(\hat{\mathbf{x}}_{t_i}, \mathbf{x}_1) \right], \quad (\hat{\mathbf{x}}_{t_i}, \mathbf{z}_{t_{i+1}}) = \hat{D}_{t_i}(I_{t_i}, \text{sg}(\mathbf{z}_{t_i}); \mathbf{c}),$$

$$\mathcal{L}_{\text{ACT}}(\hat{D}, q) := \mathbb{E}_{\mathbf{c}, \mathbf{x}_0, \mathbf{x}_1} \mathbb{E}_{(t_0, \dots, t_k) \sim \mu} \left[\sum_{i=0}^{k-1} \lambda \cdot \text{BCE}(\hat{\mathbf{q}}, \mathbf{1}(\mathbf{x}_1 = \text{round}(\hat{\mathbf{x}}_{t_i}))) \right], \quad \hat{\mathbf{q}} = q(\mathbf{z}_{t_{i+1}}).$$

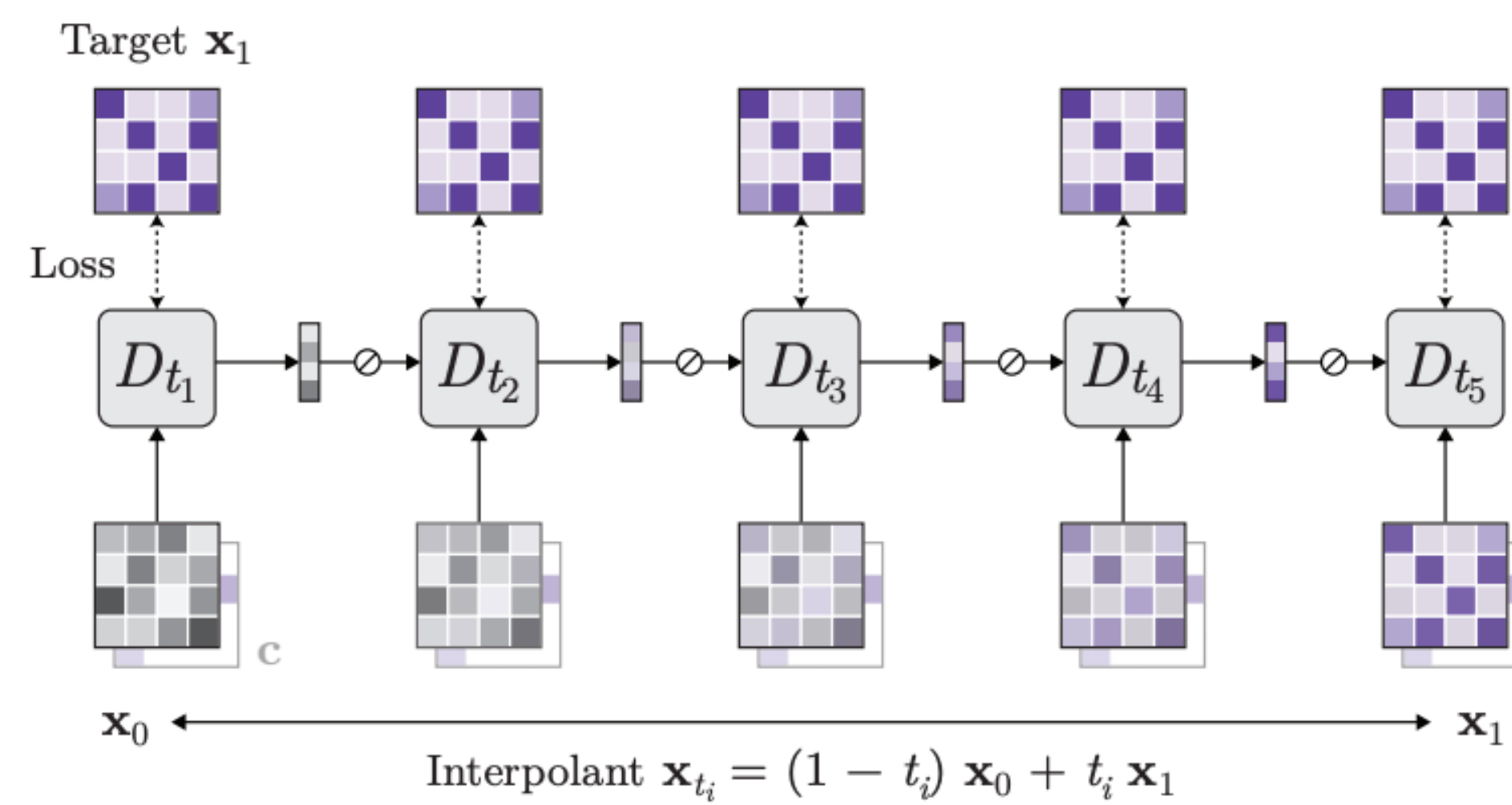
-> Adaptive computation time. As we go through the steps, accuracy may saturate and provide no meaningful training signal. The term decides whether to ignore steps after accuracy saturates.

$\text{---}\bigcirc\text{---}$ Stop-gradient
  Recurrent state $\mathbf{z}$

Truncated BPTT



Looped flows



Inference

- Integrating across a temporal grid $0 = t_0 < \dots < t_n = 1$
- Forward Euler method
- $x_{t_{i+1}} = x_{t_i} + (t_{i+1} - t_i)\hat{b}_{t_i}(x_{t_i}; c)$

Results

Method	# Params	Sudoku-Extreme	Maze-Hard	ARC-AGI-1	ARC-AGI-2
Direct prediction (Wang et al., 2025)	27M	0.0	0.0	21.0	0.0
Looped transformer (Baek et al., 2026)	7M	61.3	–	–	–
HRM (Wang et al., 2025)	27M	55.0	74.5	40.3	5.0
TRM (Jolicoeur-Martineau, 2025)	5–7M	87.4	85.3	44.6	7.8
FPRM (Movahedi et al., 2026)	7M	94.2	87.0	47.5	6.2
GRAM (Baek et al., 2026)	10M	–	–	52.0	11.1
Looped flows	5–7M	97.9 ± 0.4	86.7 ± 1.1	58.8 ± 1.8	12.2 ± 1.9
<i>Inference-time ensembling</i>					
PTRM (Sghaier et al., 2026)	5–7M	98.8 (100)	86.7 (100)	–	9.7 (25)
EqR (Huang et al., 2026)	5M	99.8 (128)	–	–	–
GRAM (Baek et al., 2026)	10M	97.0 (20)	–	–	–
Looped flows	5–7M	99.3 ± 0.2 (5)	86.9 ± 1.5 (5)	59.5 ± 1.9 (5)	12.2 ± 0.9 (5)

Results for Multi Solution reasoning

Table 2 Test results on multi-solution reasoning. We report solution accuracy (%) for N-Queens and constraint violations for Graph Coloring, with coverage (%) of distinct valid solutions from 20 inferences. Baseline results are from [Baek et al. \(2026\)](#). For looped flows, we run three repeated tests.

Method	# Params	N-Queens				Graph Coloring			
		8 × 8		10 × 10		8-vertex		10-vertex	
		Accuracy ↑	Coverage ↑	Accuracy ↑	Coverage ↑	Conflicts ↓	Coverage ↑	Conflicts ↓	Coverage ↑
Direct prediction (8 layers)	27M	40.4±1.1	13.7±1.1	13.6±0.5	1.6±0.2	179.3±4.0	19.9±0.2	198.7±5.0	6.7±0.1
Direct prediction (32 layers)	100M	40.2±1.3	13.6±1.1	13.1±0.4	1.6±0.2	174.0±18.0	19.1±1.7	227.7±34.5	6.5±1.9
Looped transformer	7M	68.4±3.7	23.6±1.9	50.0±7.6	6.2±3.2	136.0±16.1	20.5±1.5	157.3±9.0	7.2±0.7
HRM	27M	78.7±2.9	26.7±1.3	37.4±0.3	4.7±0.1	109.7±1.5	21.8±0.3	164.3±21.6	8.9±1.7
TRM	7M	66.8±5.7	36.1±22.5	17.5±11.2	2.0±1.3	109.3±3.1	22.3±0.6	170.7±17.9	6.8±0.3
AR transformer	10.6M	96.3±1.0	84.8±0.8	90.0±2.2	53.2±0.8	19.0±11.3	83.0±0.7	61.3±8.3	40.0±0.3
MDLM	12.6M	96.1±1.5	87.2±0.6	74.3±6.6	47.4±2.2	2.7±0.6	84.5±4.0	12.0±7.0	48.2±1.4
GRAM	10M	99.7±0.3	90.3±1.9	89.7±2.7	57.5±3.4	2.7±2.1	85.8±0.5	3.3±1.5	51.3±2.8
Looped flows	7M	99.9±0.1	91.4±0.4	94.4±0.7	61.5±0.1	0.7±0.6	89.4±0.7	1.0±1.0	55.2±0.3

Table 3 Training-time ablation of flow-related components. We report pass@2 accuracy (%).

Configuration	ARC-AGI-1	ARC-AGI-2
Looped flow	58.8 ± 1.8	12.2 ± 1.9
w/o time conditioning	56.4	9.9
w/o interpolant	51.5	9.9
w/o time & interpolant	43.6	5.0
w/o decreasing noise	51.6	9.9
w/o noise sharing	56.4	10.8

Table 4 Inference-time ablation: accuracy (pass@2 for ARC), in %.

Dataset	ODE $\gamma = 0$	SDE $\gamma = 5$
Sudoku	97.6 ± 0.5	97.9 ± 0.4
Maze	86.2 ± 0.7	86.4 ± 1.3
ARC-AGI-1	57.5 ± 2.9	58.8 ± 1.8
ARC-AGI-2	11.3 ± 1.5	11.8 ± 1.5

Table 5 Inference-time ablation on multi-solution tasks; coverage (%).

Dataset	ODE $\gamma = 0$	SDE $\gamma = 5$
NQ8	91.4 ± 0.4	91.4 ± 0.4
NQ10	54.7 ± 0.6	61.5 ± 0.1
GC8	88.4 ± 0.7	89.4 ± 0.7
GC10	54.0 ± 1.0	55.2 ± 0.3

Takeaways

- The paper introduced looped reasoning flows
- Improves recurrent reasoning through local denoising objectives at multiple noise levels
- Strong performance across benchmarks.