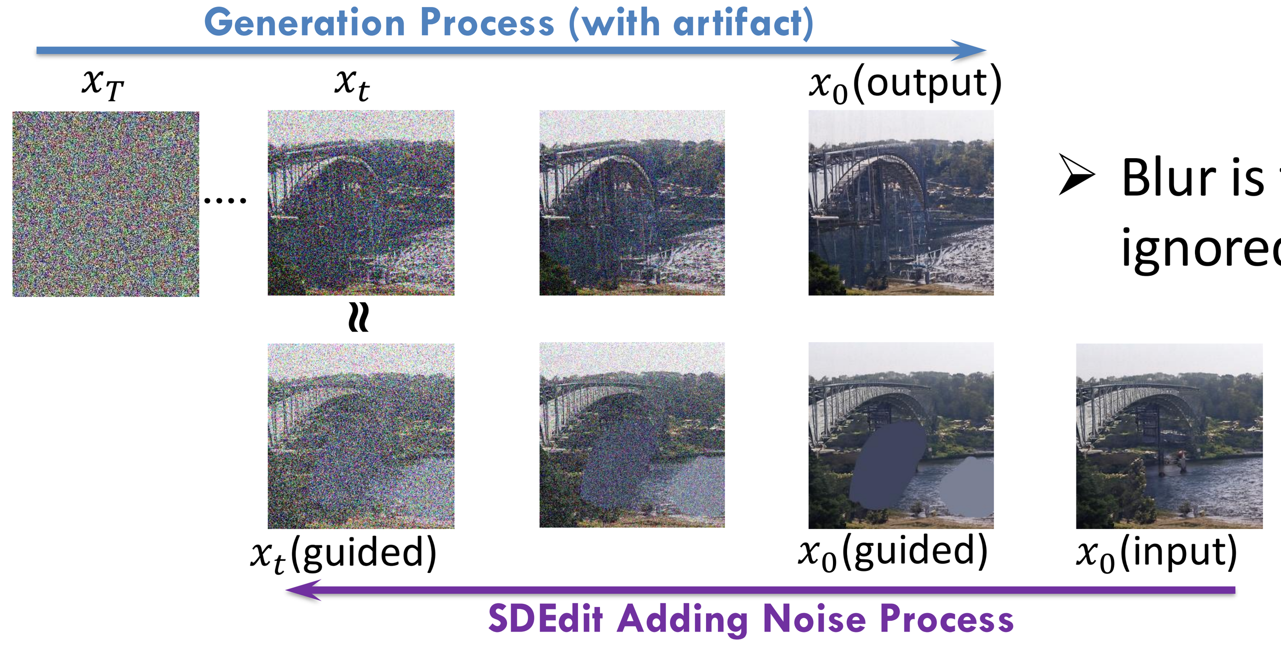
Temporal Score Analysis for Understanding and Correcting Diffusion Artifacts

Project Page

Yu Cao Zengqun Zhao Ioannis Patras Shaogang Gong
Queen Mary University of London
{yu.cao; zengqun.zhao; i.patras; s.gong}@qmul.ac.uk

Introduction

▪ Motivation: Why do diffusion models generate artifacts?



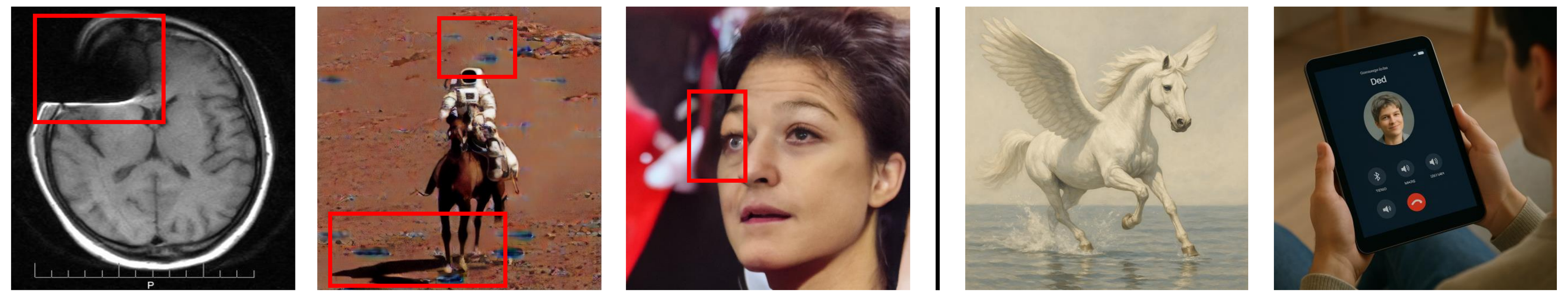
➢ Blur is fixed; artifact is ignored — why?

▪ Rethinking Generation Dynamics

➢ Uncertainty estimation methods rely on spatial cues, whilst overlooking how generation evolves over time. (e.g., BayesDiff)

➢ Post-generation filtering via pretrained classifiers enables targeted correction, whilst requiring extra steps and suffering from domain shift issues. (e.g., SARGD)

Problem Definition



Visual Artifact (left): manifest as local irregularities or distortions in a generated image, such as blurred patches, unnatural textures, broken structures.

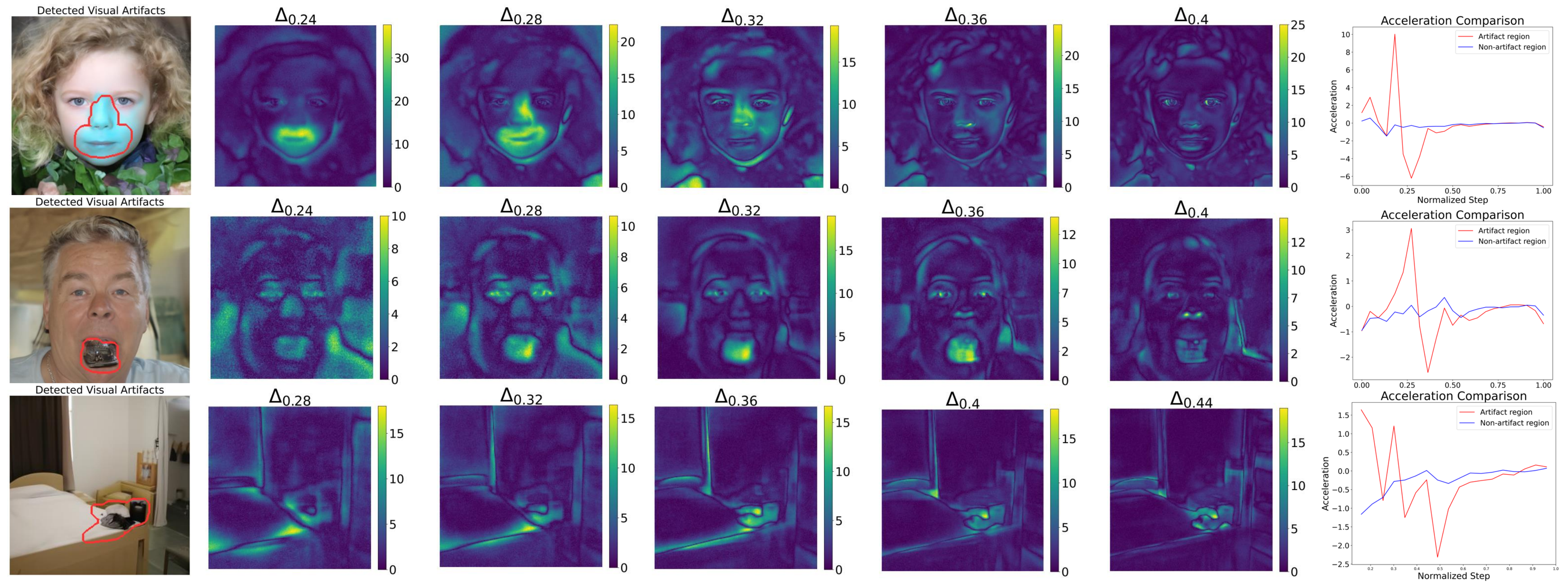
Hallucination (right): refer to semantically generating incoherent content, such as extra limbs, misplaced objects or counterfactuals.

Research Question: Can we design a training-free pipeline that seamlessly integrates artifact detection and correction into the generation process?

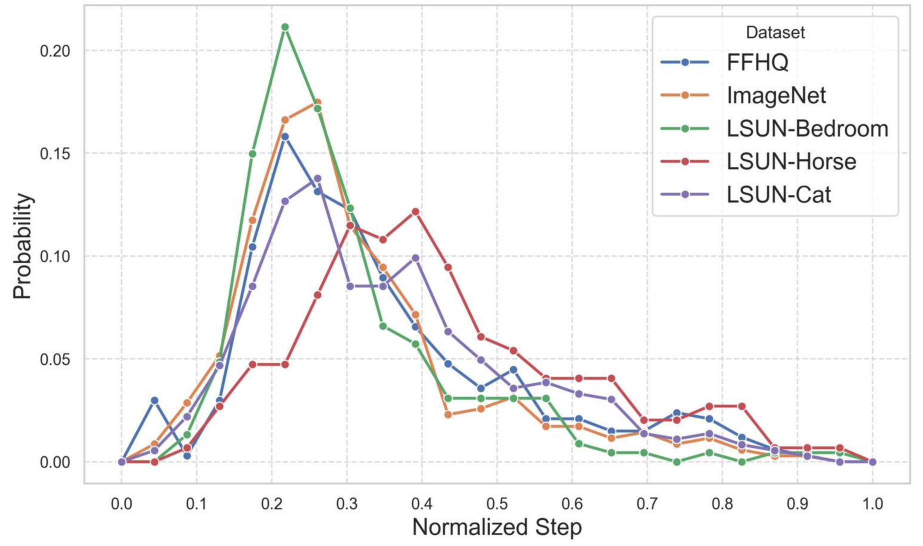
Methods

▪ Empirical Studies

➢ **Detection by Anomalous Score Dynamic**



➢ **The frequency of abnormal scores at each time step**



$\boxed{1}$ **Monitor Score Dynamics**

During $T_d \rightarrow T_c$, the score dynamics $\boldsymbol{s}_\theta(\boldsymbol{x}_t, t)$ are recorded to analyse evolution trends.

$\boxed{2}$ **Detect Visual Artifacts**

At $t = T_c$, the accumulated dynamics $\boldsymbol{s}_\theta$ are analysed. Anomalous regions Ω^a are detected based on the temporal variation:

$$\Omega^a = \left\{(i,j) || \Delta \left(w(k) \cdot s_\theta \left(x_k^{i,j}, k \right) \right) | > \tau \right\}$$

where τ is adaptively determined by:

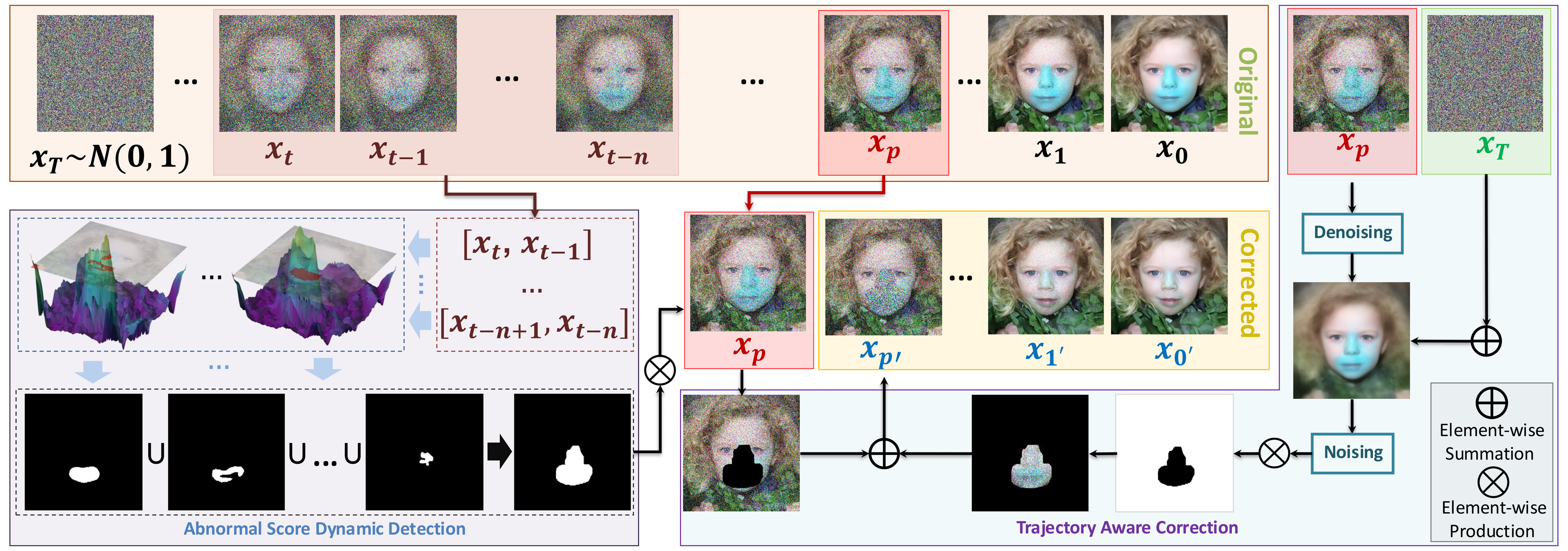
$$\tau = \max \left(\mathrm{MAD} \left(\Delta \left(w(k) \cdot s_\theta \right) \right), \mathrm{mean}(\mathcal{S}) \right)$$

$\boxed{3}$ **Targeted Correction**

For detected artifact regions Ω^a, a trajectory-aware correction is applied:

$$\boldsymbol{x}_t = \boldsymbol{x}_t \cdot \mathbb{1}_{\bar{\Omega}^a} + \left(\sqrt{\bar{\alpha}_t} \hat{\boldsymbol{x}}_0(t) + \sqrt{1 - \bar{\alpha}_t} \epsilon \right) \cdot \gamma(t) \xi \mathbb{1}_{\Omega^a}$$

▪ Framework



Experiments

▪ Quantitative Comparisons to Existing Methods

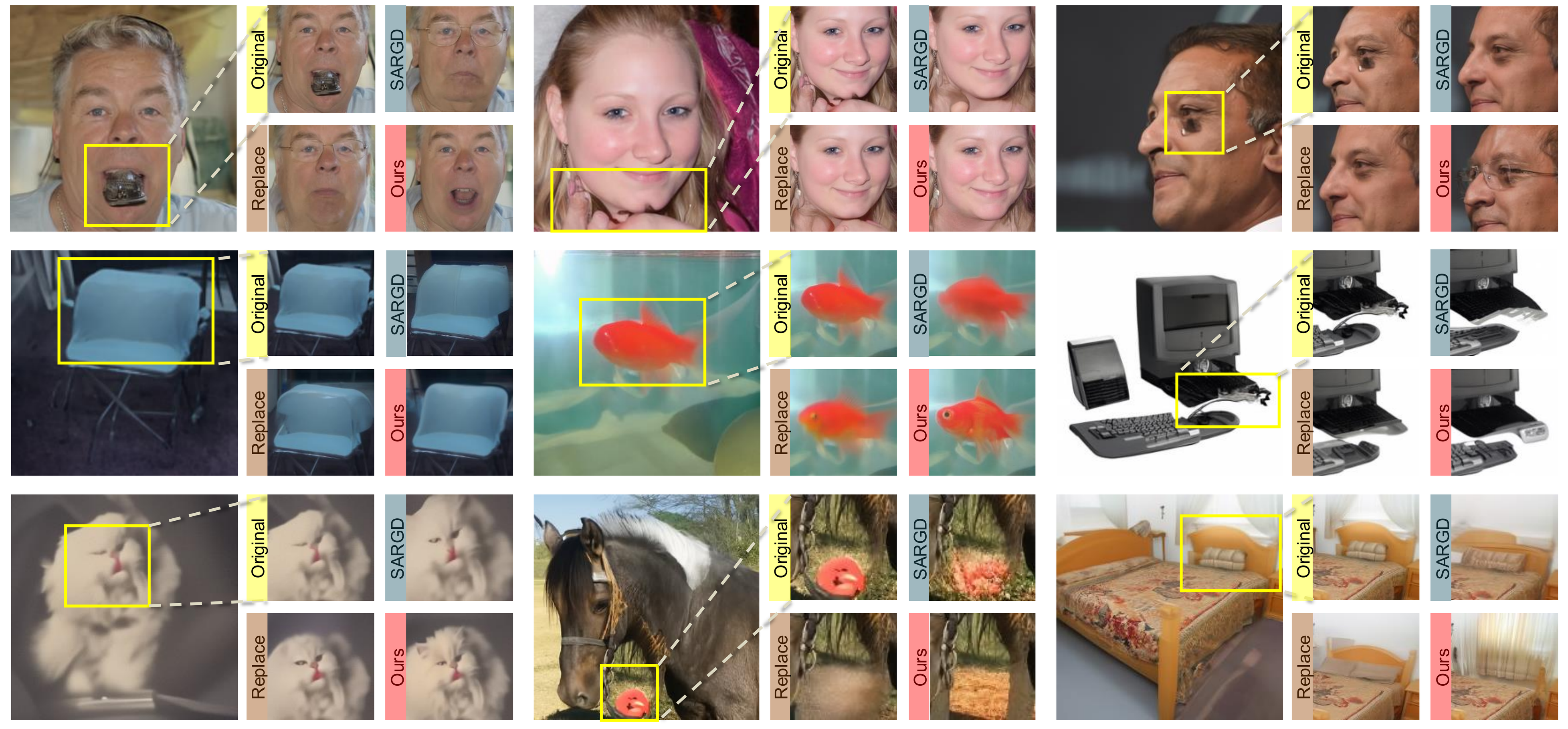
➢ Visual Artifact Detection Accuracy Comparison

Method	Type	FFHQ	ImageNet	Bedroom	Cat	Horse
PAL	Sup	51.4%	69.2%	52.4%	69.8%	60.9%
LLaVA	ZS	63.1%	91.1%	75.9%	59.5%	72.2%
Ours	UnS	56.7% (-6.4)	67.7% (-1.5)	65.0% (-10.9)	68.3% (-1.5)	70.3% (-1.9)

➢ Real-Time Correction Performance Comparison

Methods	Type	FFHQ[17]			ImageNet[10]			LSUN-Cat[40]			LSUN-Horse[40]			LSUN-Bedroom[40]		
		FID ↓	Pre. ↑	Rec. ↑	FID ↓	Pre. ↑	Rec. ↑	FID ↓	Pre. ↑	Rec. ↑	FID ↓	Pre. ↑	Rec. ↑	FID ↓	Pre. ↑	Rec. ↑
Original [34]	UnS	36.69	0.629	0.493	14.68	0.739	0.734	22.17	0.513	0.586	29.36	0.510	0.642	12.96	0.627	0.583
State Replace	UnS	37.09	0.635	0.495	14.61	**0.743**	0.733	22.79	0.510	0.587	30.36	0.502	0.642	12.95	0.628	0.574
Score Clipping	UnS	36.36	0.630	0.498	14.58	0.742	**0.736**	22.12	**0.515**	0.585	29.26	0.511	0.642	12.92	0.627	**0.585**
BayesDiff [20]	UnS	36.99	0.632	0.491	14.53	**0.743**	0.730	22.50	0.513	0.585	28.70	0.518	0.634	12.88	0.625	0.569
SARGD [49]	Sup	38.37	**0.637**	0.464	15.34	0.731	0.727	22.65	**0.523**	0.570	30.02	0.510	0.621	13.82	**0.639**	0.554
PAL [43] + TTC	Sup	**36.35**	0.624	**0.500**	**14.01**	0.731	**0.747**	**21.83**	0.514	**0.588**	**28.68**	**0.519**	**0.646**	**12.71**	**0.629**	0.579
ASCED (Ours)	UnS	**36.28**	**0.637**	**0.503**	**14.41**	**0.750**	0.735	**21.91**	**0.515**	**0.593**	**27.66**	**0.521**	**0.652**	**12.53**	0.628	**0.590**

▪ Qualitative Analysis of Correction Methods



▪ Ablation Study

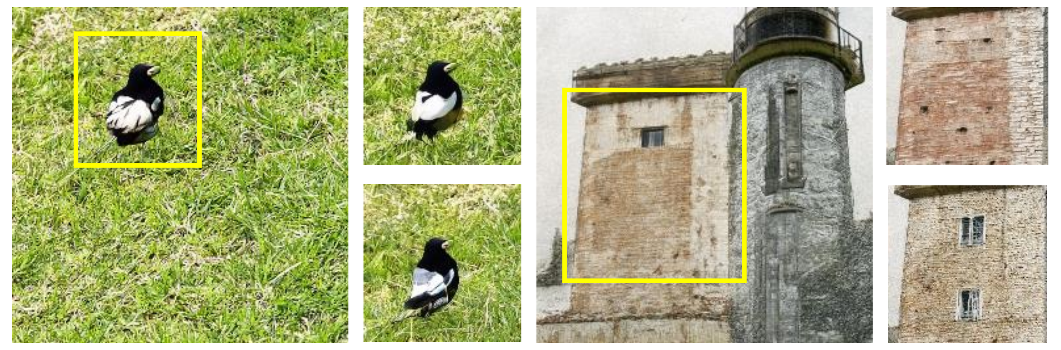


Figure 6. Applied our correction method to clean regions.

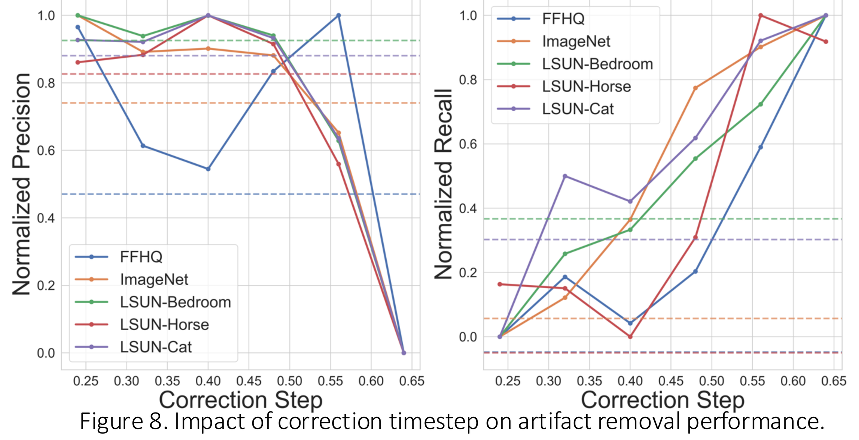


Figure 8. Impact of correction timestep on artifact removal performance.