Temporal Score Analysis for Understanding and Correcting Diffusion Artifacts

Supplementary Material

Overview

This is the appendix for "Temporal Score Analysis for Understanding and Correcting Diffusion Artifacts". Tab. 3 summarizes the abbreviations and symbols used in the paper. This appendix is organized as follows:

- Section 6 presents additional implementation details, including the Pseudo-code of Adapted SARGD and the LLaVA prediction prompt generation method.
- Section 7 presents an additional ablation study, including artifact persistence across NFE, harmless analysis, and more quantitative analysis.
- Section 8 presents additional qualitative results, including more visualization of abnormal score dynamics (Fig. 15, Fig. 16), and corrected samples (Fig. 13, Fig. 14).

Table 3. List of abbreviations and symbols used in the paper

	Meaning
Abbreviation	
ASCED	Abnormal Score Correction for Enhancing Diffusion
TTC	Trajectory-aware Targeted Correction
DM	Diffusion Model
DDPM	Denoising Diffusion Probabilistic Model
NFE	Number of Function Evaluations
LMM	Large Multi-Modal Model
MAD	Median Absolute Deviation
Symbol	
T_a	Artifact emerge step
T_c	Artifact correction step
T_c^*	Latest viable correction step
T_d	Artifact detection starting step
Ω	Spatial location of an image
Ω_t^a	Artifact region at t
Ω^a	Accumulated artifact region
$\mathcal{S}$	Score Bank
τ	Adaptive abnormal score dynamic threshold
γ	Perturbation intensity
x_0	Final output Image from Diffusion Model
x_0'	Predicted final output
x_t	Intermediate state at t
$w(t)$	Temporal weighting function
$s_\theta(\cdot)$	Score network
$\epsilon_\theta(\cdot)$	Noise network
T	Total time-steps
$\beta_1, \ldots, \beta_T$	Variance schedule
α_t	$1 - \beta_t$
$\bar{\alpha}_t$	$\prod_{s=1}^{t} \alpha_s$

6. Implementation Details

6.1. Adapted SARGD

Since SARGD [49] was originally designed for super-resolution where the final output (low resolution) is available, we adapt it to our scenario by using the predicted clean image (x_0') at T_d as guidance instead of the real Low-Resolution (LR) image. The rest of the correction process follows the original SARGD implementation, including artifact detection and refinement, but operates within our identified correction window (T_c to T_d). The complete algorithm is provided in Algorithm 2, where the red mark-out text indicates removed steps from the original SARGD, and the green text shows our adaptations.

6.2. LLaVA Prediction Prompts

To generate effective prompts for LLaVA's [22] artifact detection, we first manually collected examples of images with artifacts. These examples were presented to GPT-4 [1] for prompt synthesis. We repeated this process 50 times, each time with different image combinations, generating 50 distinct prompts. The final evaluation used the best-performing prompt (No. 5) based on detection accuracy. For reproducibility, we provide below the complete set of prompts used in Tab. 2:

1. "Assess if there are any visible flaws in this image that a person could easily detect, like irregular shapes, unexpected color variations, blurred regions, or any other clear image disruptions. Answer with 'yes' or 'no'."
2. "Does this image contain any significant artifacts that distort the natural appearance, such as unexpected color patches, blurring, or pixelation? Please reply 'yes' or 'no'."
3. "Are there any obvious flaws in this image, such as large blurry areas, severe distortions, or color errors? Respond with 'yes' or 'no' only."
4. "Can you identify any glaring visual defects in this image that would be immediately noticeable to a human viewer? Reply with just 'yes' or 'no'."
5. "Determine if this image shows any noticeable defects or artifacts that would be easily seen by a human, including shape distortions, color issues, blurring, or pixelation in areas where it should be smooth. Please reply 'yes' or 'no'."
6. "Does this image have any obvious visual artifacts such as severe blurring, distortion, or unrealistic colors that would make it appear unnatural or of poor quality? Answer 'yes' or 'no'."

Algorithm 2 Adapted Self-Adaptive Reality-Guided Diffusion (SARGD) Pseudo-code

Input: ~~LR image $\boldsymbol{I}_{LR}$~~, and total diffusion steps T
Load: Encoder $\mathcal{E}$, artifact detector $\mathcal{A}$ and LR decoder $\mathcal{D}$
◀ **Step 1: Initialization** ▷ Removed as the final output x_0 is not accessible
~~Upscale LR image as $up(\boldsymbol{I}_{LR})$~~
~~Encode the upsampled image as $\boldsymbol{x} = \mathcal{E}(up(\boldsymbol{I}_{LR}))$~~
~~Initialize the $\boldsymbol{x}$ as a realistic latent $\boldsymbol{x}_r$ and set it as guidance~~
~~Compute the realty score of the realistic latent $\boldsymbol{s}_r$~~
◀ **Step 2: Sampling**
for t = T, ..., 1 **do**
    Sample $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ if $t > 1$, else $\boldsymbol{\epsilon} = 0$
    Computer the latent variable at the current step $\boldsymbol{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta\left(\boldsymbol{x}_t, \boldsymbol{x}, t\right)\right) + \sigma_\theta\left(\boldsymbol{x}_t, t\right)\boldsymbol{\epsilon}$
    if $t = T_d$ **then** ▷ We use the predicted x_0' to estimate the realty score
        Set predicted x_0' as guidance (using Eq. (5)
        Computer an estimated realty score of the realistic latent $\boldsymbol{s}_r$
    else if $T_c <= t < T_d$ **then** ▷ Align SARGD correction timing with ours
        Detect artifacts of the current latent $E_A = \mathcal{A}\left(\mathcal{D}\left(\boldsymbol{x}_{t-1}\right)\right)$ ▷ Following steps remain the same
        Refine the latent $\boldsymbol{x}_{t-1} = \boldsymbol{x}_{t-1} \times (1 - E_A) + \boldsymbol{x}_r \times E_A$
        Decode the refined latent into an image $\mathbf{I}_r = \mathcal{D}\left(\boldsymbol{x}_{t-1}\right)$
        Generate the current binary reality map $M_R = \mathcal{R}\left(\mathbf{I}_r\right)$
        Calculate the current reality score $s_r^{t-1} = \mathcal{S}\left(M_R\right)$
        Encode the current realistic latent $\boldsymbol{x}_r^{t-1} = \mathcal{E}\left(\mathbf{I}_r\right)$
        Update the guidance $\boldsymbol{x}_r = \mathcal{G}\left(\boldsymbol{x}_r, \boldsymbol{x}_r^{t-1}\right)$ if $\boldsymbol{s}_r^{t-1} > \boldsymbol{s}_r$
        Update the reality score $\boldsymbol{s}_r = \boldsymbol{s}_r^{t-1}$ if $\boldsymbol{s}_r^{t-1} > \boldsymbol{s}_r$
return the artifact-free **SR** $\boldsymbol{L}_{HR} = \mathcal{D}(\boldsymbol{x}_0)$

7. "Is the quality of this image significantly impaired by visual defects like large areas of pixelation, color mismatches, or misplaced objects? Please respond with 'yes' or 'no'."
8. "Can you identify any prominent visual issues in this image, such as incorrect color rendering, noticeable noise, severe blurring, or any elements that appear misplaced or distorted? Answer with 'yes' or 'no'."
9. "Does this image contain any obvious visual flaws that significantly degrade its quality, such as large blurry sections, strange artifacts, or clearly incorrect proportions of objects? Answer 'yes' or 'no'."
10. "Is there any obvious visual artifact in this image, like a hand growing out of a face, unrealistic color transitions, or large areas of texture inconsistency that make the image appear fake or unnatural? Please respond 'yes' or 'no'."
11. "Determine if this image has any clear visual artifacts that affect its appearance, such as distorted shapes, wrong color patches, excessive noise, or objects that are clearly in the wrong place. Reply with 'yes' or 'no'."
12. "Is the visual quality of this image compromised by obvious flaws, including but not limited to severe blurring, incorrect object placement, or large areas of unrealistic colors? Respond with 'yes' or 'no'."
13. "Examine the image for any significant visual issues, like pronounced noise, pixelation, unrealistic colors, or misaligned elements that affect the overall image quality. Please answer 'yes' or 'no'."
14. "Does this image exhibit any major visual artifacts that a human observer would immediately notice, such as large blurs, odd color patterns, or misplaced elements? Answer only with 'yes' or 'no'."
15. "Assess if there are any visible and distracting visual artifacts in this image, such as large unnatural blurs, obvious pixelation, incorrect object shapes, or areas of incorrect coloring. Reply with 'yes' or 'no'."
16. "Does this image contain any major visual artifacts that significantly degrade its quality? Answer only 'yes' or 'no'."
17. "Are there any obvious flaws in this image, such as large blurry areas, severe distortions, or color errors? Respond with 'yes' or 'no' only."
18. "Can you identify any glaring visual defects in this image that would be immediately noticeable to a human viewer? Reply with just 'yes' or 'no'."
19. "Does this image exhibit any significant visual anomalies like body parts in unnatural positions or severe pixelation? Answer 'yes' or 'no'."
20. "Is the overall quality of this image notably poor due to

visible artifacts or distortions? Provide only a 'yes' or 'no' response."

21. "Are there any major visual imperfections in this image that make it look unrealistic or poorly generated? Reply with 'yes' or 'no'."
22. "Does this image contain any obvious flaws that would make you question its authenticity or quality? Answer only with 'yes' or 'no'."
23. "Can you spot any significant visual errors in this image, such as misplaced facial features or unnatural textures? Respond with just 'yes' or 'no'."
24. "Is there any clear evidence of poor image generation or editing in this picture, like inconsistent lighting or impossible anatomy? Reply 'yes' or 'no'."
25. "Does this image exhibit any significant visual anomalies like body parts in unnatural positions or severe pixelation? Answer 'yes' or 'no'."
26. "Is the overall quality of this image notably poor due to visible artifacts or distortions? Provide only a 'yes' or 'no' response."
27. "Are there any major visual imperfections in this image that make it look unrealistic or poorly generated? Reply with 'yes' or 'no'."
28. "Does this image contain any obvious flaws that would make you question its authenticity or quality? Answer only with 'yes' or 'no'."
29. "Would you consider this image to be of low quality due to noticeable visual artifacts or errors? Answer with only 'yes' or 'no'."
30. "Does this image appear to be of normal quality, without any obvious visual artifacts such as blurring, distortion, or unnatural colors? Answer 'yes' if it appears normal, 'no' if there are visible issues."
31. "Is this image free of any significant visual defects like pixelation, color mismatches, or misplaced objects? Respond with 'yes' if there are no issues, or 'no' if such artifacts are present."
32. "Can you confirm that this image has no prominent visual issues, such as incorrect color rendering, noticeable noise, severe blurring, or misplaced elements? Answer 'yes' if there are no issues, and 'no' if there are."
33. "Can you spot any significant visual errors in this image, such as misplaced facial features or unnatural textures? Respond with just 'yes' or 'no'."
34. "Does this image lack any obvious visual flaws that would significantly degrade its quality, such as blurry sections, artifacts, or incorrect object proportions? Answer 'yes' for no flaws, 'no' if flaws are present."
35. "Is the image free from visual artifacts like hands growing out of faces, unrealistic color transitions, or texture inconsistencies that make the image look unnatural? Reply 'yes' if the image is clear, or 'no' if artifacts are present."
36. "Determine whether this image has any major visual artifacts affecting its appearance, such as distorted shapes, incorrect colors, excessive noise, or misaligned elements. Reply 'yes' if the image looks normal, or 'no' if such issues exist."
37. "Is the visual quality of this image high, with no obvious flaws like severe blurring, misplaced objects, or large patches of unrealistic colors? Respond 'yes' if the quality is good, 'no' if issues are found."
38. "Evaluate this image for any significant visual issues, such as noise, pixelation, unrealistic colors, or misaligned elements. Respond 'yes' if no issues are found, 'no' if any artifacts are present."
39. "Does this image have any noticeable visual artifacts that a human observer would immediately recognize, such as large blurs, odd color patterns, or misplaced elements? Answer 'yes' if there are no artifacts, or 'no' if artifacts are present."
40. "Is this image clear of any visible and distracting visual artifacts, such as large blurs, obvious pixelation, incorrect object shapes, or wrong coloring? Reply 'yes' if the image is free of artifacts, 'no' if artifacts are visible."
41. "Are there any jarring inconsistencies or unnatural elements in this image that detract from its realism? Answer 'yes' or 'no'."
42. "Does this image show any signs of poor rendering, such as incomplete objects or abrupt transitions? Respond with only 'yes' or 'no'."
43. "Can you detect any major issues with perspective or proportions in this image that make it look artificial? Reply with 'yes' or 'no'."
44. "Are there any noticeable problems with the lighting or shadows in this image that seem unrealistic? Answer only 'yes' or 'no'."
45. "Does this image contain any elements that appear to be unnaturally distorted or warped? Provide a 'yes' or 'no' response."
46. "Can you identify any significant issues with the texture or surface details in this image that look artificial? Reply with just 'yes' or 'no'."
47. "Are there any obvious problems with the edges or outlines of objects in this image, such as jagged lines or haloing? Answer 'yes' or 'no'."
48. "Does this image exhibit any clear signs of overprocessing or artificial enhancement that degrade its quality? Respond with 'yes' or 'no' only."
49. "Can you spot any major inconsistencies in the style or appearance of different parts of this image? Reply with 'yes' or 'no'."
50. "Are there any glaring issues with the color balance or saturation in this image that make it look unnatural? Answer only with 'yes' or 'no'."

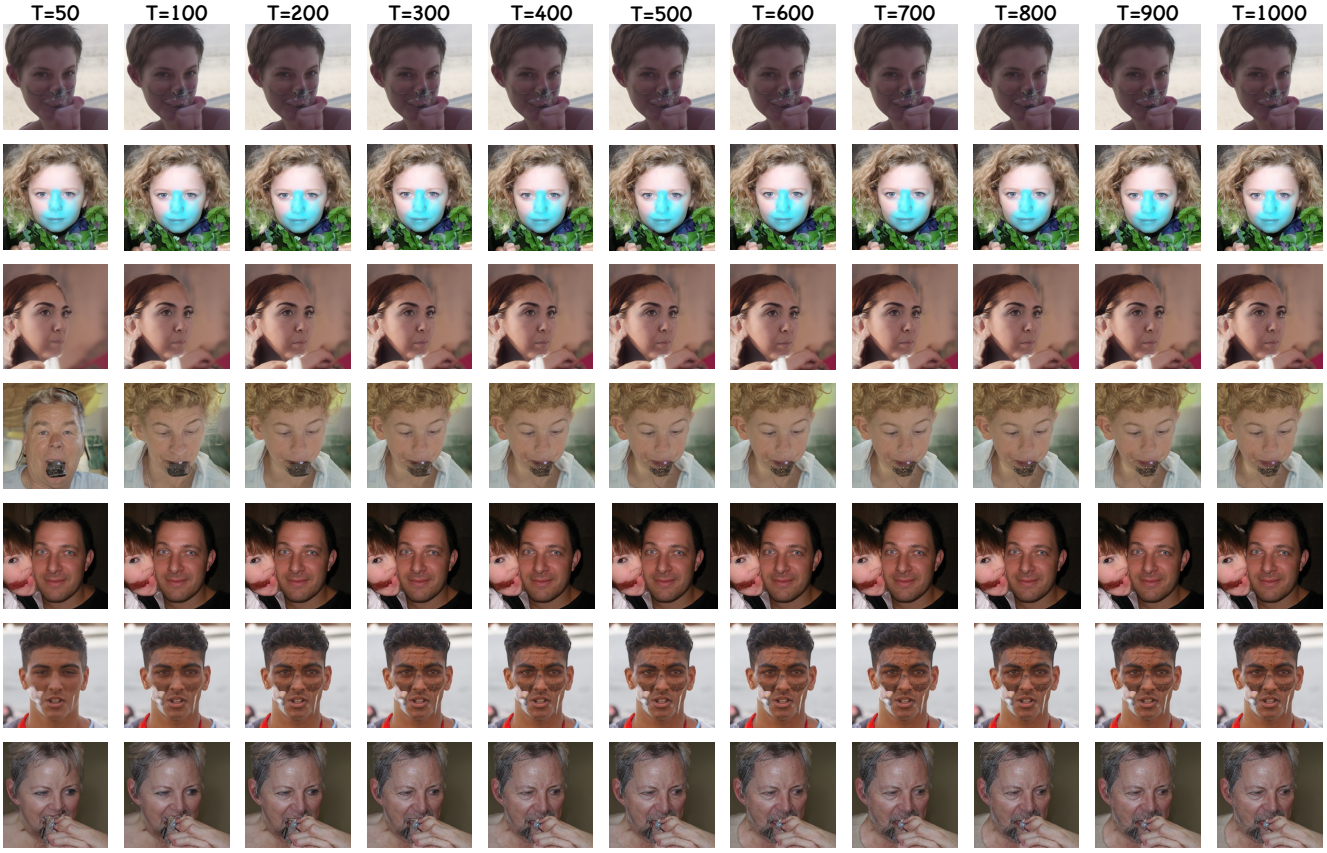

Figure 10. Visual artifacts persist across different numbers of sampling steps (NFE (T) from 50 to 1000 (original)) using the same random seed. While non-artifact regions show minor evolution in details, artifact regions remain virtually unchanged, demonstrating that artifacts stem from disrupted score dynamics rather than insufficient sampling granularity.

7. Additional Analysis

7.1. Artifact Persistence Across NFE

Our main experiments use DDIM with NFE$=25$ for efficiency. To evaluate the potential effects of sampling granularity, we tested increasing NFE by up to 1000 steps (original). As shown in Fig. 10, while a higher NFE allows more iterations for pixel evolution, leading to changes in overall image composition, the artifact regions remain visually unchanged. This observation supports our score trap analysis (Sec. 3.4): surrounding pixels continue to evolve with more sampling steps, but the trapped regions maintain their patterns, demonstrating that these areas indeed stop updating due to disrupted score dynamics rather than insufficient sampling steps.

7.2. Harmlessness in Non-Artifact Regions

To understand why our correction method maintains semantic coherence while enabling controlled diversity in non-artifact regions, we need to examine both the local score dynamics and the fundamental properties of diffusion models. In normal regions, pixels maintain coupled evolution through the score function as described in Eq. (7), where each pixel evolves in coordination with its neighborhood context. When our method introduces controlled perturbations in these regions, two mechanisms work in concert to preserve image integrity.

First, the coupled score evolution pattern remains intact, as these regions maintain normal dynamics without entering score traps. This coupling naturally guides the perturbed pixels to evolve in harmony with their surroundings. Second, and more fundamentally, diffusion models are inherently equipped to handle noise through their denoising objective:

$$\arg\min_{\boldsymbol{\theta}} D_{\mathrm{KL}}\left(q\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0\right) \mid p_{\boldsymbol{\theta}}\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t\right)\right) \quad (11)$$

$$=\arg\min_{\boldsymbol{\theta}} \frac{1}{2\sigma_q^2(t)} \frac{\bar{\alpha}_{t-1}\left(1-\alpha_t\right)^2}{\left(1-\bar{\alpha}_t\right)^2}\left[\left\|\hat{x}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right)-\boldsymbol{x}_0\right\|_2^2\right] \quad (12)$$

where $\hat{x}_\theta(\cdot)$ predicts x_0 directly [19]. Following [25], this objective can be rewritten in terms of Signal-to-Noise Ratio (SNR):

$$\arg\min_{\boldsymbol{\theta}} \frac{1}{2}(\mathrm{SNR}(t-1)-\mathrm{SNR}(t))\left[\left\|\hat{\boldsymbol{x}}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right)-\boldsymbol{x}_0\right\|_2^2\right] \quad (13)$$

where $\mathrm{SNR}(t)=\frac{\bar{\alpha}_t}{1-\bar{\alpha}_t}$. This formulation reveals that the diffusion process naturally increases SNR during denoising,

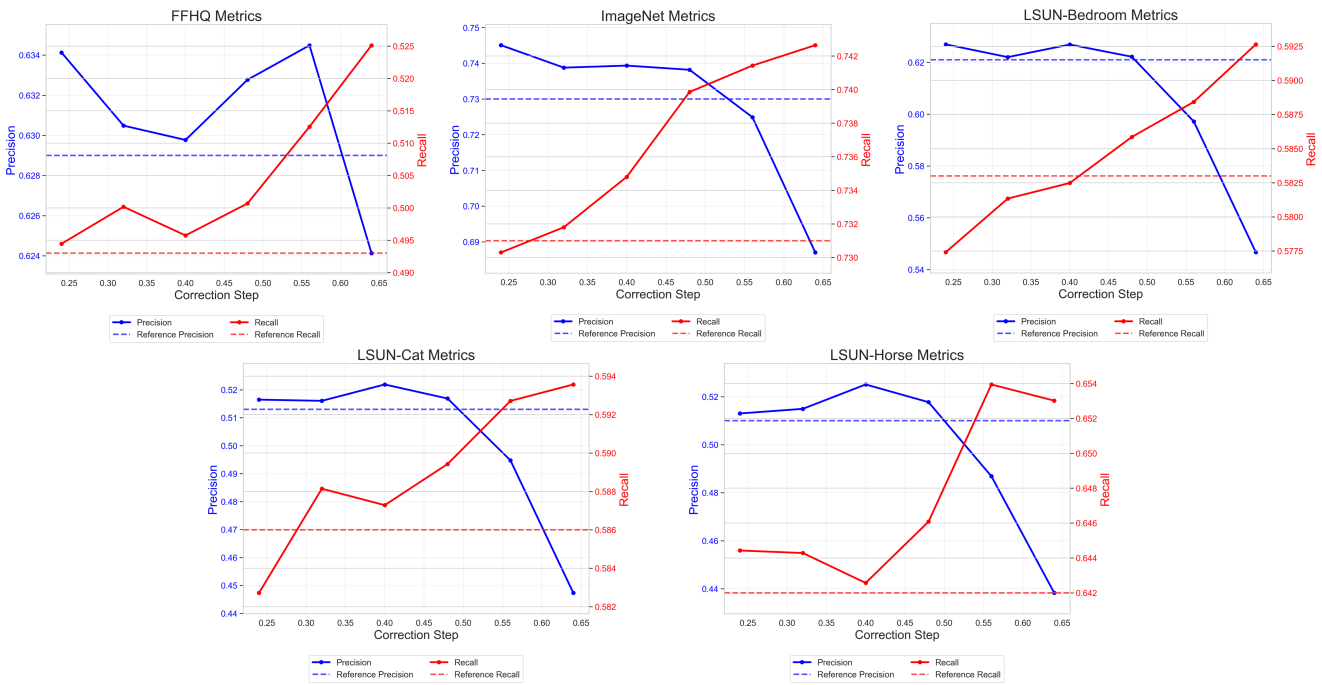

Figure 11. Impact of correction timestep T_c on artifact removal performance for FFHQ, ImageNet, LSUN bedrooms, horses, and cats. For each dataset, the blue and red solid lines show the Precision (fidelity) and Recall (diversity), respectively, while the corresponding dashed lines indicate the baseline performance of the original diffusion model.

ensuring controlled perturbations are effectively processed while maintaining semantic structure through coupled score evolution. Additional visual examples of perturbation effects in non-artifact regions are provided in Fig. 12.

7.3. Analysis of Global Correction Application

Given our perturbations maintain semantic coherence and introduce controlled diversity in non-artifact regions, a natural question arises: Why not extend these perturbations to the entire image regardless of artifact detection? When applying correction globally, each image would essentially undergo a "second" generation process. Since the underlying diffusion model has an inherent probability of generating artifacts, this universal application would maintain the same artifact rate rather than reduce it. Therefore, introducing perturbations selectively based on artifact detection is necessary, avoiding unnecessary variations in well-formed regions while preserving the diversity benefits where needed.

7.4. Individual Precision and Recall

Following the timing analysis in the main text (Sec. 4.4), we present detailed performance curves for each dataset in Fig. 11. The results reveal different patterns across datasets: while most datasets exhibit a clear performance drop after T_c^*, FFHQ shows a more gradual degradation. This difference can be attributed to the complexity of facial features, which allows for more flexible refinement compared to other domains. Notably, all datasets maintain performance above their respective baselines (shown in dashed lines) when corrections are applied before T_c^*, demonstrating the robustness of the identified threshold. The consistent pattern of optimal performance near $T_c^* \approx 0.48$ in various datasets validates the generality of this timing criterion for diffusion artifact correction.

8. Additional Experiment Results

8.1. More Abnormal Score Dynamics Visualization

Extended from the representative cases in the main paper (Fig. 4), Fig. 15 and Fig. 16 show additional qualitative analysis of score dynamics in artifact regions. These examples consistently demonstrate the characteristic abnormal score patterns: sharp variations in score changes displayed in activation maps and the distinct acceleration-deceleration curves in artifact regions compared to normal areas.

8.2. More Corrected Samples

Following the qualitative analysis (Fig. 5) in the main paper, we provide additional correction results (Fig. 13 and Fig. 14) across different datasets to demonstrate the consistent performance of the proposed method. These examples further illustrate the effectiveness of trajectory-aware target correction (ours) in preserving local details while removing artifacts.

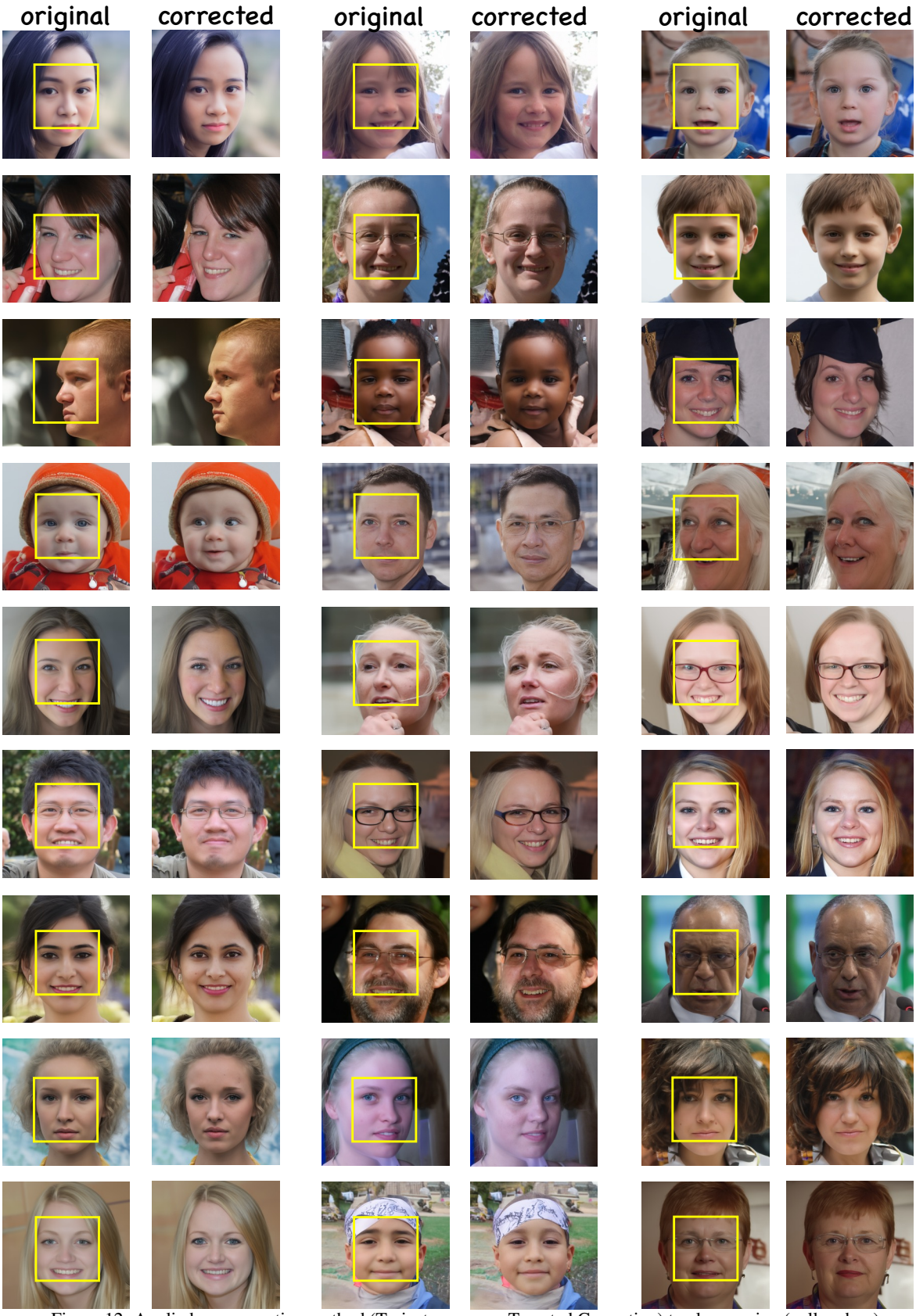

Figure 12. Applied our correction method (Trajectory-aware Targeted Correction) to clean region (yellow box).

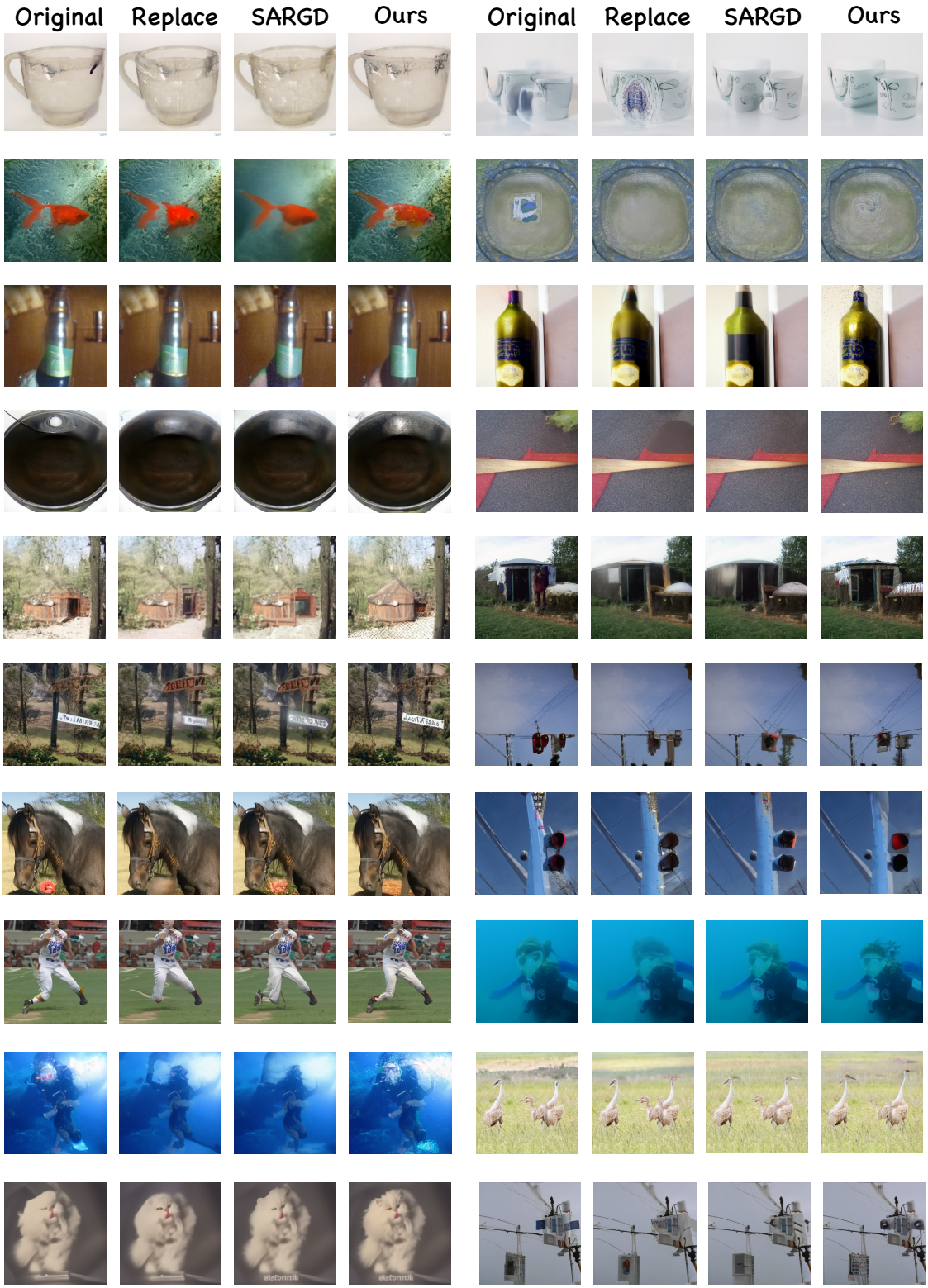

Figure 13. Additional qualitative comparison of artifact correction methods, following the similar format as Fig. 5 in the main text.

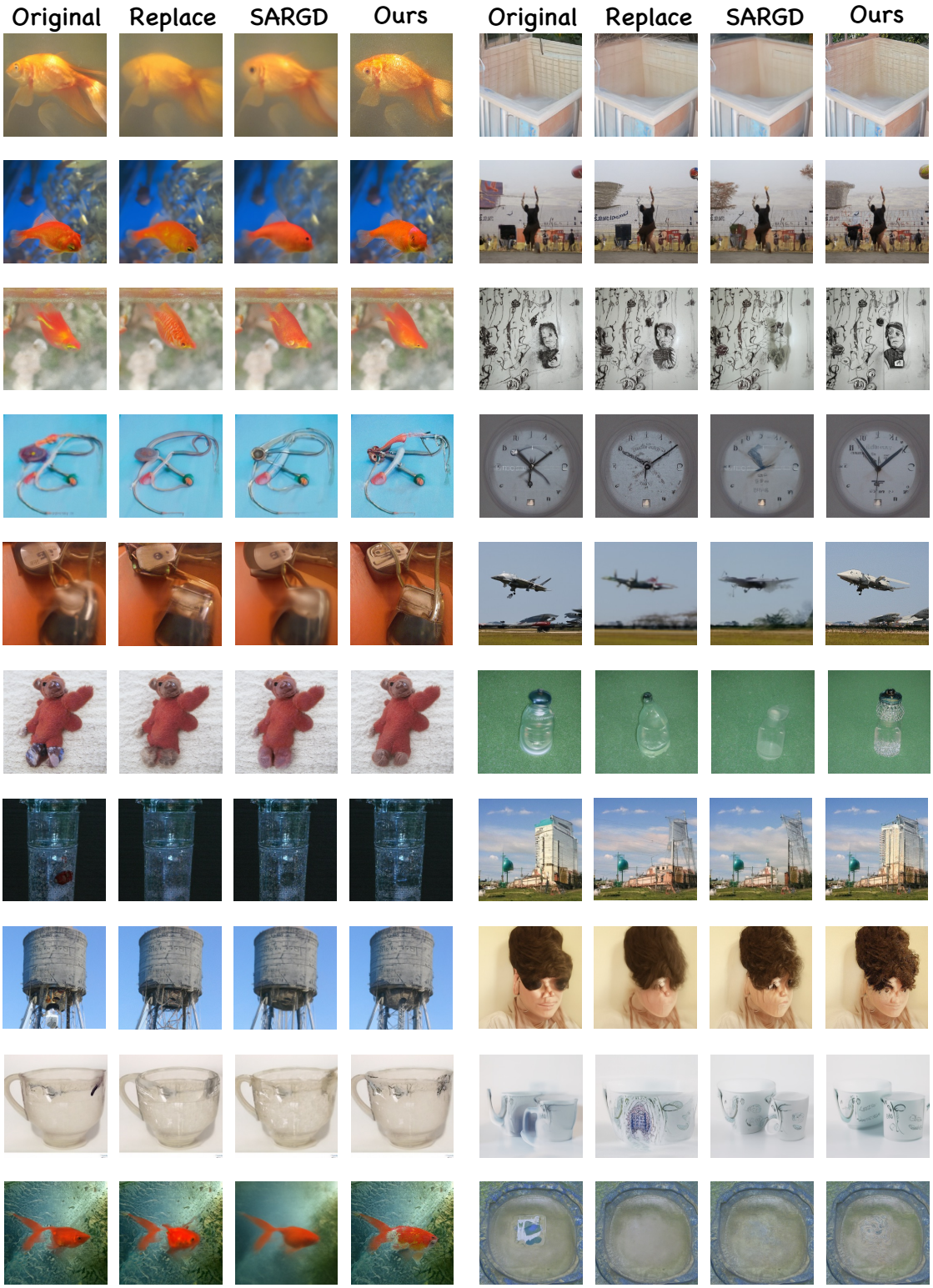

Figure 14. Additional qualitative comparison of artifact correction methods, following the similar format as Fig. 5 in the main text.

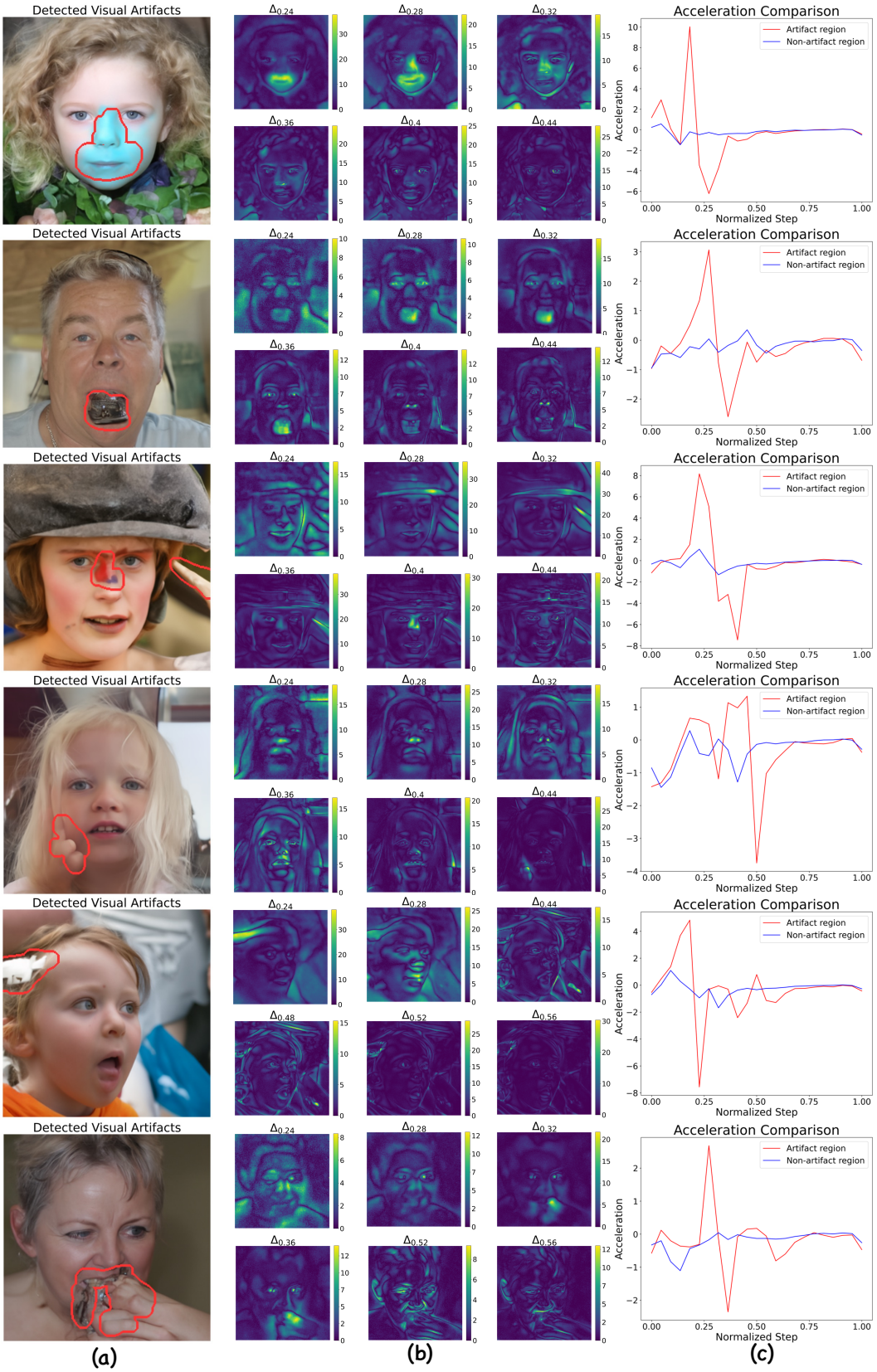

Figure 15. Extended visualization of abnormal score dynamics and visual artifact detection with more examples, following the analysis shown in Fig. 4 of the main text. The same patterns of score acceleration and deceleration in artifact regions are consistently observed across different cases.

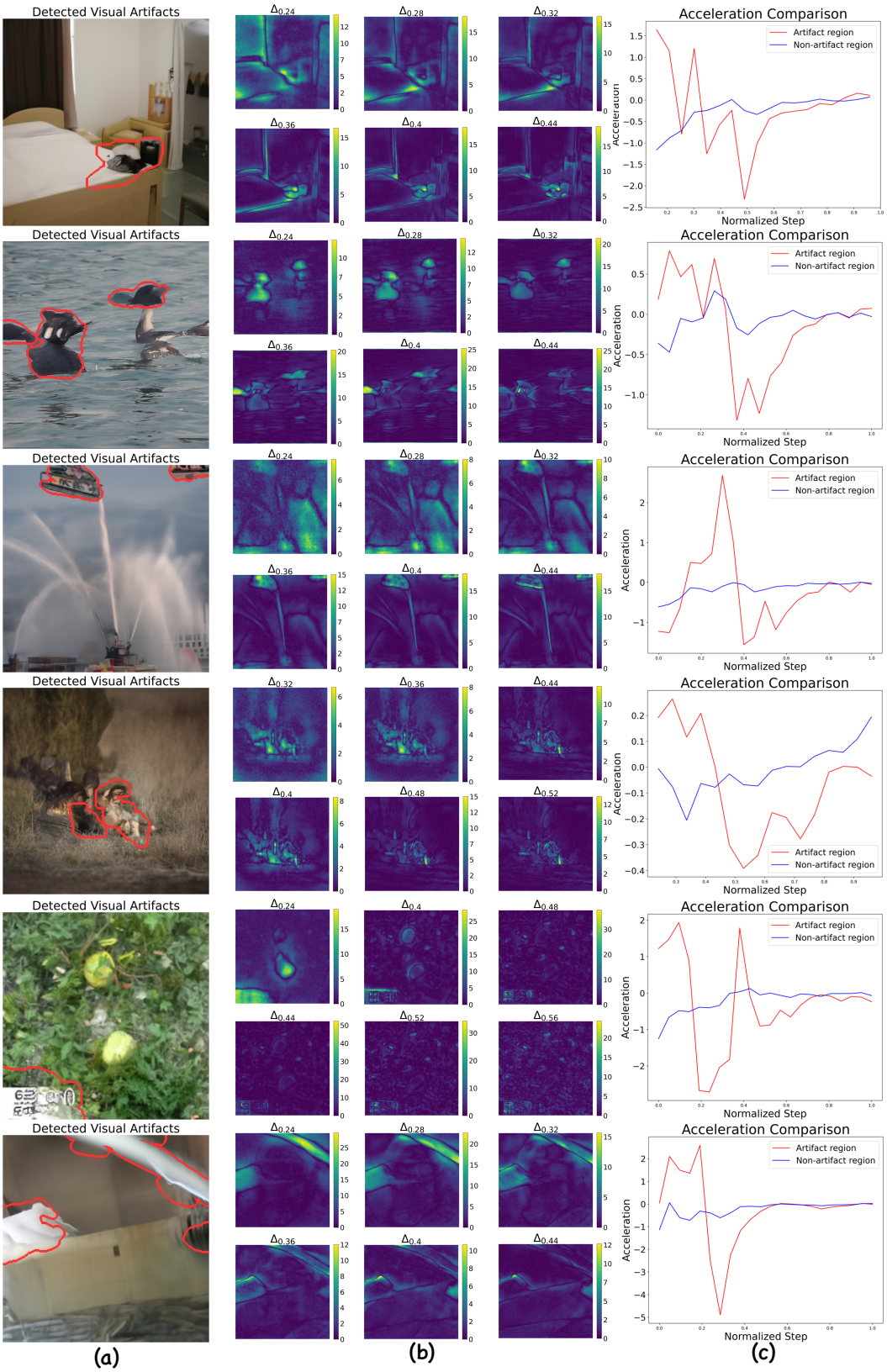

Figure 16. Extended visualization of abnormal score dynamics and visual artifact detection with more examples, following the analysis shown in Fig. 4 of the main text. The same patterns of score acceleration and deceleration in artifact regions are consistently observed across different cases.