7 March 2025 ·

Qeios

Peer Review

Review of: "Enhancing Sample Generation of Diffusion Models using Noise Level Correction"

Teresa Klatzer[1]

1. University of Edinburgh, United Kingdom

TLDR; The approach seems novel and competitive, but the presentation has major weaknesses, detailed below. I suggest a major revision.

List of comments:

Introduction

- The role of the distance to the data manifold in convergence and sampling is unclear. You write "...is crucial for convergence and accurate sampling" but it is not detailed what sampling means in that context.
- The abbreviation EDM is never explicitly explained.
- Qualitative results should appear in the experimental section, not just in the introduction and appendix.
- The bullet point list at the end of the introduction is repetitive and should be merged.

Background

- Lacks clarity & definitions.
- Many claims remain vague. For example, section 2.1 implies that a data point lies in a low-dimensional manifold but lacks precise context (specifically when referring to a generated image, this manifold can be very high-dimensional, in the context of a diffusion model as opposed to the result of a PCA).
- After Eq. 2, you refer to a "noise vector" which is inconsistently defined across the document; \epsilon under Eq. 3; then \nabla dist_K(x_t) just above Figure 2; then as \epsilon_\theta under Eq. (19).

doi.org/10.32388/BLWWQ6

- \epsilon with varying subscripts is redefined on several occasions, which is confusing; e.g., as Gaussian noise and also as the learned denoiser.
- It remains unclear how “randomly sampled triplets (...) are used to update the denoiser′s parameters” within a gradient descent procedure (under Eq.(4)).
- Section 2.2: “Several methods have been proposed to … while reducing the number of iterations” - this part lacks citations and specificity.
- Contradiction: It is claimed that diffusion models enforce data consistency, but by design, they do not (see [1], Sec 4.1).
- The paper refers to the ″true distance of noisy samples to the manifold″ without clarifying how this is measured without explicit knowledge of the manifold. Later, it is stated that the true distance is infeasible to compute—this should be clarified earlier.
- Confusion regarding the manifold hypothesis: If K is a manifold of dimension d; d<<n; but then you define x_0 \in K \in R^n, resulting in d==n? (under Eq. (11)).
- Eq. (1) contains a typo (extra norm signs).

Methods

- When explaining the network architecture:
  - Figure 3: The role of the diffusion model is unclear; where is it applied?
  - The interaction between the pretrained denoiser and the diffusion model is not well explained.
  - The scaling factor \lambda and how it expands the input/output space of the noise level correction network requires clarification.
- Typo under Eq. (19) - “process” seems incorrect.
- Section 3.3 is unclear in its explanation of how the method integrates with the existing diffusion methods; experimental conclusions are mixed with the method description, and references are missing.
- Section 3.4 lacks a clear baseline for comparison; terms like “further enhance” and “also improve performance” need explicit references.
- Eq. (20): x_(k) remains unspecified - is it the same as x_(k+1)? k is running from 0 to K_max according to the text.
- Under Eq.(21): experimental details appear within the methodology section, making it harder to follow.

Experiments

- The introduction of the "10-step DDIM" in section 4.1 is abrupt - this should be established earlier. Why is 10 a good choice, and which variable does the number refer to?
- Reproducibility concerns: Experimental details are mostly in the Appendix, but some details should be in the main text. E.g., for deblurring with a Gaussian kernel - which noise level, and which width of the Gaussian kernel? in Section 4.3
- In Tables 1 and 2: The meaning of boldface numbers is unclear.
- Many sections are not self-contained - abbreviations are introduced too late, making it difficult to track methods.
- The "consistency metric" appears to be a standard L1 data-consistency term, not a new metric.
- Tables 3 and 4 are too dense; can the information be condensed?
- Section 4.4 - The benefit of using a lookup table is unclear, and there are contradictions regarding when it is effective. The final claim about the NLC network's minimal computational overhead raises the question of whether the lookup table is necessary at all.
- The additional overhead from correction networks vs. a standard diffusion model should be thoroughly explained in the main text.

Conclusion:

The paper presents a promising approach but lacks clarity and rigor. Key terms should be defined consistently, contradictions resolved, and experimental details made more explicit in the main text. Tables and figures should be refined for clarity, and claims should be better substantiated with references and precise explanations.

[1] Feng, B. T., Smith, J., Rubinstein, M., ... & Freeman, W. T. (2023). Score-based diffusion models as principled priors for inverse imaging. In ICCV (pp. 10520-10531).

Declarations

Potential competing interests: No potential competing interests to declare.