Novelty Assessment Report

Paper: Diagnosing and Improving Diffusion Models by Estimating Optimal Loss Value

PDF URL: https://openreview.net/pdf?id=X7JfjLKKLQ

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-27

Abstract

Diffusion models have achieved remarkable success in generative modeling. Despite more stable training, the loss of diffusion models is not indicative of absolute data-fitting quality, since its optimal value is typically not zero but unknown, leading to the confusion between large optimal loss and insufficient model capacity. In this work, we advocate the need to estimate the optimal loss value for diagnosing and improving diffusion models. We first derive the optimal loss in closed form under a unified formulation of diffusion models, and develop effective estimators for it, including a stochastic variant scalable to large datasets with proper control of variance and bias. With this tool, we unlock the inherent metric for diagnosing training quality of representative diffusion model variants, and develop a more performant training schedule based on the optimal loss. Moreover, using models with 120M to 1.5B parameters, we find that the power law is better demonstrated after subtracting the optimal loss from the actual training loss, suggesting a more principled setting for investigating the scaling law for diffusion models.

Core Task Landscape

This paper addresses: **Estimating Optimal Loss Value for Diffusion Models**

A total of **50 papers** were analyzed and organized into a taxonomy with **28 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Loss Function Design and Optimization**
- **Likelihood-Based Training and Variational Bounds**
- **Guidance and Controllable Generation**
- **Preference Optimization and Alignment**
- **Model Compression and Efficient Training**
- **Sampling and Generation Strategies**
- **Theoretical Analysis and Foundations**
- **Discrete and Structured Data Diffusion**
- **Application-Specific Diffusion Models**
- **Privacy, Security, and Model Behavior Analysis**
- ... and 1 more categories

Complete Taxonomy Tree

- Estimating Optimal Loss Value for Diffusion Models Survey Taxonomy
- Loss Function Design and Optimization
  - Weighted and Adaptive Loss Formulations (3 papers)
  - [7] Analyzing and improving the training dynamics of diffusion models (Tero Karras, 2024) View paper
  - [15] Scaling laws for diffusion transformers (Liang Zhengyang, 2024) View paper
  - [16] Perception Prioritized Training of Diffusion Models (Jooyoung Choi, 2022) View paper
  - Reconstruction-Generation Trade-off Optimization (2 papers)
  - [1] Reconstruction vs. generation: Taming optimization dilemma in latent diffusion models (Jingfeng Yao, 2025) View paper
  - [9] Palette: Image-to-image diffusion models (Saharia, 2022) View paper
  - Contrastive and Perceptual Loss Integration (3 papers)
  - [14] A bearing fault data augmentation method based on hybrid-diversity loss diffusion model and parameter transfer (Yuan Wei, 2025) View paper
  - [21] Zero-shot contrastive loss for text-guided diffusion image style transfer (Serin Yang, 2023) View paper
  - [50] SSL: A Self-similarity Loss for Improving Generative Image Super-resolution (Du Chen, 2024) View paper
  - Domain-Specific and Task-Adapted Losses (2 papers)
  - [10] Forget-me-not: Learning to forget in text-to-image diffusion models (Zhang Gong, 2024) View paper
  - [39] Fine-tuning diffusion model to generate new kite designs for the revitalization and innovation of intangible cultural heritage (Yaqin Zhou, 2025) View paper
- Likelihood-Based Training and Variational Bounds
  - Variational Diffusion Frameworks (2 papers)
  - [2] Variational diffusion models (Kingma, 2021) View paper
  - [12] Variational SchrÃ¶dinger Diffusion Models (Deng Wei, 2024) View paper
  - Maximum Likelihood and Score Matching Connections (2 papers)
  - [6] Maximum likelihood training of score-based diffusion models (Song Yang, 2021) View paper
  - [32] Likelihood-based diffusion language models (Gulrajani, 2023) View paper

- Density Estimation and Theoretical Guarantees (2 papers)
- [43] Diffusion models are minimax optimal distribution estimators (Oko, 2023) View paper
- [49] A score-based density formula, with applications in diffusion generative models (Li Gen, 2024) View paper

- Guidance and Controllable Generation
  - Classifier and Discriminator Guidance (2 papers)
  - [11] Improving Discriminator Guidance in Diffusion Models (Verine, 2025) View paper
  - [13] Learn to guide your diffusion model (Galashov, 2025) View paper
  - Loss-Based and Training-Free Guidance (2 papers)
  - [28] Understanding and improving training-free loss-based diffusion guidance (Dongqi Han, 2024) View paper
  - [47] Loss-guided diffusion models for plug-and-play controllable generation (J Song, 2023) View paper
  - Conditional and Attribute-Specific Control (3 papers)
  - [18] Target-aware video diffusion models (Kim Taek-Soo, 2025) View paper
  - [41] Linguistic Binding in Diffusion Models: Enhancing Attribute Correspondence through Attention Map Alignment (Rassin, 2023) View paper
  - [42] SVGDreamer: Text Guided SVG Generation with Diffusion Model (Ximing Xing, 2023) View paper
- Preference Optimization and Alignment
  - Direct Preference Optimization for Diffusion (2 papers)
  - [38] D3PO: Preference-Based Alignment of Discrete Diffusion Models (Umberto Borso, 2025) View paper
  - [48] HuViDPO:Enhancing Video Generation through Direct Preference Optimization for Human-Centric Alignment (Lifan Jiang, 2025) View paper
  - Reinforcement Learning-Based Alignment (3 papers)
  - [17] Find: Fine-tuning initial noise distribution with policy optimization for diffusion models (Changgu Chen, 2024) View paper
  - [35] Diffusion Policies as an Expressive Policy Class for Offline Reinforcement Learning (Wang, 2022) View paper
  - [40] Fine-Tuning Text-to-Speech Diffusion Models Using Reinforcement Learning with Human Feedback (Chen Jingyi, 2025) View paper
- Model Compression and Efficient Training
  - Knowledge Distillation and Model Pruning (3 papers)
  - [23] Diffusion probabilistic model made slim (Xing-yi Yang, 2023) View paper
  - [26] Distilling diffusion models into conditional gans (Minguk Kang, 2024) View paper
  - [45] Progressive Knowledge Distillation Of Stable Diffusion XL Using Layer Level Loss (Paul, 2024) View paper
  - Domain Transfer and Fine-Tuning Efficiency (1 papers)
  - [22] DogFit: Domain-guided Fine-tuning for Efficient Transfer Learning of Diffusion Models (Shateri, 2025) View paper
- Sampling and Generation Strategies
  - Few-Step and Direct Denoising Methods (2 papers)
  - [4] Monocular depth estimation using diffusion models (Saxena, 2023) View paper
  - [25] Directly Denoising Diffusion Models (Zhang Dan, 2024) View paper
  - Residual and Alternative Denoising Frameworks (1 papers)
  - [5] Residual denoising diffusion models (Jiawei Liu, 2024) View paper
- Theoretical Analysis and Foundations
  - Unified Perspectives and Formulations (1 papers)
  - [36] Understanding diffusion models: A unified perspective (Luo, 2022) View paper
  - Singularities and Mathematical Properties (1 papers)
  - [46] Mathematical Analysis of Singularities in the Diffusion Model Under the Submanifold Assumption (Yubin Lu, 2023) View paper
  - Scaling Laws and Optimization Dynamics (1 papers)
  - [30] Exploring the Frontiers of Softmax: Provable Optimization, Applications in Diffusion Model, and Beyond (Gu, 2024) View paper
- Discrete and Structured Data Diffusion
  - Discrete State-Space Diffusion (1 papers)
  - [33] Structured Denoising Diffusion Models in Discrete State-Spaces (Austin, 2021) View paper
- Application-Specific Diffusion Models
  - Human Motion and Pose Generation (2 papers)
  - [3] Human motion diffusion model (Tevet, 2022) View paper
  - [29] Grpose: Learning graph relations for human image generation with pose priors (Di, 2025) View paper
  - Object Detection and Network Optimization (3 papers)
  - [19] Diffusiondet: Diffusion model for object detection (Chen, 2023) View paper
  - [20] A new prediction strategy for dynamic multi-objective optimization using diffusion model (Ke Tang, 2025) View paper
  - [31] DiffSG: A generative solver for network optimization with diffusion model (Ruihuai Liang, 2025) View paper
  - Safety-Critical and Constrained Control (1 papers)
  - [27] From Uncertain to Safe: Conformal Adaptation of Diffusion Models for Safe PDE Control (Peiyan Hu, 2025) View paper
- Privacy, Security, and Model Behavior Analysis
  - Memorization and Generalization Analysis (1 papers)
  - [8] On the edge of memorization in diffusion models (Buchanan, 2025) View paper
  - Membership Inference and Model Inversion (2 papers)
  - [34] Loss and likelihood based membership inference of diffusion models (Hailong Hu, 2023) View paper
  - [37] Model inversion attacks through target-specific conditional diffusion models (Li, 2024) View paper
  - Class Imbalance and Distribution Overlap (1 papers)
  - [44] Training class-imbalanced diffusion model via overlap optimization (Qi Lu, 2024) View paper
- Foundational Models and Core Methodologies
  - Core Diffusion Model Frameworks (1 papers)
  - [24] Denoising diffusion probabilistic models (Ho, 2020) View paper
  - Optimal Loss Estimation and Diagnostics ★ (1 papers)
  - [0] Diagnosing and Improving Diffusion Models by Estimating Optimal Loss Value (Anon et al., 2026) View paper

Narrative

Core task: Estimating optimal loss value for diffusion models. The field of diffusion models has grown into a rich ecosystem organized around several major themes. At the foundational level, researchers explore likelihood-based training and variational bounds (e.g., Variational Diffusion[2], Maximum Likelihood Score[6]) alongside theoretical analyses that clarify the mathematical underpinnings of score-based generative processes. Loss function design and optimization form another central pillar, addressing how to best train these models through improved objectives and weighting schemes. Guidance and controllable generation branches focus on steering outputs toward desired attributes, while preference optimization and alignment methods (such as D3PO[38], HuViDPO[48]) adapt diffusion models to human feedback. Parallel efforts in model compression and efficient training (e.g., Diffusion Model Slim[23], Progressive Knowledge Distillation[45]) aim to reduce computational costs, and sampling strategies explore faster or higher-quality generation paths. Application-specific branches span domains from human motion (Human Motion Diffusion[3]) to depth estimation (Monocular Depth Diffusion[4]), and discrete or structured data diffusion extends the framework beyond continuous spaces. Privacy and security analyses (Membership Inference Diffusion[34], Model Inversion Attacks[37]) round out the landscape by examining model behavior and vulnerabilities.

Within this diverse taxonomy, a particularly active line of work centers on understanding training dynamics and diagnosing model performance. Training Dynamics Diffusion[7] and Edge of Memorization[8] investigate how models learn and when they begin to overfit, while Reconstruction vs Generation[1] examines the trade-offs between faithful data reconstruction and creative sample diversity. Optimal Loss Diffusion[0] sits squarely in this diagnostic cluster, focusing on estimating the theoretically best achievable loss to benchmark training progress and identify when further optimization yields diminishing returns. This emphasis on loss estimation complements nearby efforts like Residual Denoising[5], which refines the denoising objective itself, and contrasts with works that prioritize architectural scaling (Scaling Laws DiT[15]) or perceptual quality (Perception Prioritized Training[16]). By providing a principled target for loss values, Optimal Loss Diffusion[0] offers a lens through which practitioners can assess whether their models are approaching fundamental limits or still have room for improvement.

Related Works in Same Category

No sibling papers were found in the same taxonomy leaf. A taxonomy-subtopic-level comparison will be produced instead.

Taxonomy-Level Summary

The original leaf focuses on diagnostic and optimization methods that estimate optimal loss values to assess and improve diffusion model training, while the sibling subtopic covers foundational DDPM formulations. Both relate to core diffusion model theory, but the original leaf addresses meta-level training analysis (estimating what loss should be achieved) rather than the base model architecture or loss function definition itself.

Similarities: - Both are foundational/theoretical aspects of diffusion models rather than application-specific adaptations - Both involve understanding loss functions in the context of denoising diffusion models - Both contribute to the theoretical understanding of how diffusion models learn

Differences: - The original leaf focuses on estimating optimal/expected loss values for diagnostic purposes, while the sibling covers the base DDPM formulation and architecture - The original leaf is about training quality assessment and schedule improvement, whereas the sibling defines the core probabilistic framework - The original leaf is a meta-analysis tool (analyzing what loss should be), while the sibling is the primary model definition (what the loss is)

Suggested Search Directions: - Methods for computing theoretical lower bounds on diffusion model loss - Techniques for diagnosing training issues by comparing actual vs optimal loss trajectories - Approaches for using optimal loss estimates to design better noise schedules or training curricula

Sibling Subtopics

- **Core Diffusion Model Frameworks** (leaves: 1, papers: 1)
- Scope: Original DDPM and foundational denoising diffusion probabilistic model formulations.
- Exclude: Variants, extensions, or application-specific adaptations belong to other categories.

Contributions Analysis

Overall novelty summary. The paper proposes closed-form derivations and practical estimators for the optimal loss value in diffusion models, enabling practitioners to diagnose training quality by comparing actual loss to theoretical minima. Within the taxonomy, it occupies a unique leaf ('Optimal Loss Estimation and Diagnostics') under 'Foundational Models and Core Methodologies,' with no sibling papers in that leaf. This positioning suggests the work addresses a relatively sparse research direction—while the broader field contains 50 papers across 28 leaf nodes, this specific focus on optimal loss estimation as a diagnostic tool appears underexplored compared to more crowded areas like guidance methods or loss function design.

The taxonomy reveals that neighboring research directions concentrate on training dynamics (e.g., 'Edge of Memorization,' 'Reconstruction vs Generation') and theoretical foundations (e.g., 'Unified Perspectives,' 'Scaling Laws'). The paper's emphasis on deriving optimal loss values connects it to likelihood-based training frameworks and theoretical analyses, yet diverges by focusing on practical diagnostics rather than pure mathematical properties or convergence guarantees. Its proposed training schedule improvements also touch on optimization dynamics, bridging foundational theory with empirical training practices. The taxonomy's scope notes clarify that general loss function design or training dynamics without optimal loss estimation belong elsewhere, reinforcing this work's distinct positioning.

Among the three contributions analyzed, the literature search examined 19 candidates total, finding refutable prior work for each. The closed-form derivation and estimators (10 candidates examined, 1 refutable) appear most novel, though one candidate provides overlapping methodology. The training schedule design (5 candidates, 1 refutable) and modified scaling law formulation (4 candidates, 2 refutable) face more substantial prior work, with the scaling law contribution encountering two potentially overlapping papers. These statistics reflect a limited semantic search scope, not an exhaustive survey, so the presence of refutable candidates indicates some methodological overlap within the examined subset rather than definitive lack of novelty.

Given the limited search scope of 19 candidates, the analysis suggests moderate novelty: the optimal loss estimation framework occupies a sparse taxonomy leaf, but each contribution encounters at least one overlapping candidate among those examined. The work's integration of closed-form theory, practical estimators, and scaling law refinements may offer value through synthesis and application, even if individual components have partial precedents. A broader literature review would be needed to assess whether the 4 refutable pairs represent isolated overlaps or systematic prior coverage.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Closed-form derivation and practical estimators for diffusion model optimal loss

Description: The authors derive a closed-form expression for the optimal loss value of diffusion models and develop practical estimators, including a scalable stochastic estimator (cDOL) that controls variance and bias for large datasets. This enables measuring absolute data-fitting quality rather than only relative quality.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Constrained diffusion models via dual training

URL: View paper

Brief Assessment

Constrained Dual Training[56] focuses on constrained optimization for diffusion models with distribution constraints, not on deriving closed-form optimal loss expressions or developing practical estimators for measuring absolute data-fitting quality.

2. NODI: Out-Of-Distribution Detection with Noise from Diffusion

URL: View paper

Brief Assessment

NODI[58] focuses on out-of-distribution detection using diffusion models for noise prediction, not on deriving closed-form optimal loss expressions or developing practical estimators for measuring data-fitting quality in diffusion model training.

3. Explicit Flow Matching: On The Theory of Flow Matching Algorithms with Applications

URL: View paper

Brief Assessment

Explicit Flow Matching[61] focuses on deriving tractable loss functions for flow matching models to reduce training variance, not on estimating optimal loss values for measuring absolute data-fitting quality in diffusion models.

4. Diagnosing and Improving Diffusion Models by Estimating the Optimal Loss Value

URL: View paper

Prior Art Analysis

Optimal Loss Estimation[52] demonstrates that the closed-form expression for optimal loss and practical estimators (including the scalable stochastic estimator cDOL) were already published. The candidate paper presents identical mathematical formulations, including the same closed-form expression (Theorem 1), the same estimator equations (eqs. 7-11), and the same theoretical framework for measuring absolute data-fitting quality. The candidate explicitly derives the optimal loss as j(x0) t ∗ = ep(x0)‖x0‖2 - ep(xt)[ep(x0|xt)[x0]]^2, develops the same three-tier estimator hierarchy (naive, DOL, cDOL), and provides the same theoretical justification (Theorem 2) for the corrected estimator's consistency properties.

Evidence

Evidence 1 - **Rationale**: This demonstrates that Optimal Loss Estimation[52] already published the scalable cDOL estimator with variance-bias control that the original paper claims as a novel contribution. - **Original**: we develop estimators for the optimal loss based on the expression. for large datasets, we design a scalable estimator based on dataset sub-sampling, with a delicate design to properly balance variance and bias. - **Candidate**: we introduce a simple correction by down-weighting such pairs with a coefficient c, and call it the corrected dol (cdol) estimator: ˆbcdol t := 1 m x ˜m∈[m] p l∈[l]‚l=l ˜m x(l) 0 kt(x(˜m) t , x(l) 0) +1 c x(l ˜m) 0 kt(x(˜m) t , x(l ˜m) 0) p l′∈[l]‚l′=l ˜m kt(x(˜m) t , x(l′) 0) +1 c kt(x(˜m)...

Evidence 2 - **Rationale**: This shows that Optimal Loss Estimation[52] already published the theoretical justification for the cDOL estimator's consistency properties, including the same mathematical statement about expectation equivalence and consistency. - **Original**: theorem 2. the ˆbcdol t estimator with subset sizel has the same expectation as the ˆbsnis t estimator with subset sizel-1whenm→ ∞, c→ ∞, hence is a consistent estimator. - **Candidate**: theorem 2. the ˆbcdol t estimator with subset sizel has the same expectation as the ˆbsnis t estimator with subset size l - 1 when m → ∞, c→ ∞, hence is a consistent estimator.

Evidence 3 - **Rationale**: This pair shows nearly identical problem formulation and motivation, demonstrating that Optimal Loss Estimation[52] already established the conceptual framework for measuring absolute versus relative data-fitting quality. - **Original**: the diffusion loss only reflects therelativedata-fitting quality for monitoring training process or comparing models under the same setting, while remains obscure for measuring theabsolutefit to the training data. it is due to that the optimal loss of diffusion model,i.e., the lowest possible loss v... - **Candidate**: the diffusion loss only reflects the relative data-fitting quality for monitoring training process or comparing models under the same setting, while remains obscure for measuring the absolute fit to the training data. it is due to that the optimal loss of diffusion model, i.e., the lowest possible l...

5. Seeds: Exponential sde solvers for fast high-quality sampling from diffusion models

URL: View paper

Brief Assessment

SEEDS[62] focuses on accelerating sampling from diffusion models through exponential SDE solvers, not on deriving closed-form optimal loss expressions or developing practical estimators for measuring absolute data-fitting quality.

6. Diffusion Models: A Mathematical Introduction

URL: View paper

Brief Assessment

Mathematical Introduction Diffusion[55] is a tutorial paper focused on deriving diffusion model mathematics from first principles. It does not address optimal loss estimation or develop practical estimators for measuring absolute data-fitting quality in diffusion models.

7. Data-driven machine learning approach based on physics-informed neural network for population balance model

URL: View paper

Brief Assessment

Physics Informed PBM[59] addresses population balance models using physics-informed neural networks for particulate systems, not diffusion generative models or their optimal loss estimation.

8. Optimizing diffusion models for joint trajectory prediction and controllable generation

URL: View paper

Brief Assessment

Joint Trajectory Prediction[60] focuses on optimizing diffusion models for trajectory prediction in autonomous driving, not on deriving closed-form optimal loss expressions or developing practical estimators for measuring absolute data-fitting quality in diffusion models.

9. Variational diffusion models

URL: View paper

Brief Assessment

Variational Diffusion[2] focuses on deriving a variational lower bound (VLB) for diffusion models and optimizing the noise schedule, not on deriving closed-form expressions for the optimal loss value or developing practical estimators like cDOL for measuring absolute data-fitting quality.

---

10. CURE: Concept Unlearning via Orthogonal Representation Editing in Diffusion Models

URL: View paper

Brief Assessment

CURE[57] focuses on concept unlearning in text-to-image diffusion models through orthogonal representation editing in cross-attention weights, not on deriving optimal loss values or developing estimators for measuring data-fitting quality in diffusion models.

Contribution 2: Optimal-loss-based training schedule design for diffusion models

Description: The authors propose a new training schedule that uses the gap between actual and optimal loss to determine loss weights and noise schedules. This approach improves FID scores by 2%-25% across multiple datasets and diffusion model variants.

This contribution was assessed against **5 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Diagnosing and Improving Diffusion Models by Estimating the Optimal Loss Value

URL: View paper

Prior Art Analysis

Optimal Loss Estimation[52] demonstrates that the optimal-loss-based training schedule design was already published. The candidate paper presents the same methodology: using the gap between actual and optimal loss to determine loss weights ($w\sigma = a/j*\sigma$) and adaptive noise schedules ($p(\sigma) \propto w\sigma(j\sigma(\theta) - j*\sigma)$). The candidate reports identical experimental improvements on the same datasets (2%-14% on CIFAR-10, 7%-25% on ImageNet-64, 9% on ImageNet-256) and provides the same theoretical justification that the loss gap, not the loss itself, reflects data-fitting insufficiency.

Evidence

Evidence 1 - **Rationale**: This pair shows that Optimal Loss Estimation[52] already published the optimal-loss-based training schedule with identical performance improvements on the same datasets, refuting the novelty claim. - **Original**: from the analysis, we designed a principled training schedule for diffusion models, based on the gap between the actual loss and the optimal loss. our training schedule improves the fid by 2%-14%(for edm (karras et al., 2022) / fm (lipman et al., 2023)) on cifar-10,7%-25%(for edm / fm) on imagenet-6... - **Candidate**: from the analysis, we designed a principled training schedule for diffusion models, based on the gap between the actual loss and the optimal loss. we find that our training schedule improves generation performance in fid by 2%-14% (for edm [24] / fm [33]) on cifar-10, 7%-25% (for edm / fm) on imagen...

Evidence 2 - **Rationale**: This demonstrates that Optimal Loss Estimation[52] already published the specific loss weight formulation based on optimal loss that the original paper claims as novel. - **Original**: the loss weight.w σ calibrates the error resolution across different noise scales. for this, the optimal lossj ⋆ σ provides a perfect reference scale for the loss at each diffusion stepσ, sow σ =a/j ⋆ σ with a scale factorais a natural choice to align the loss at variousσto the same scale. - **Candidate**: the loss weight. wσ calibrates the error resolution across different noise scales. for this, the optimal loss j∗ σ provides a perfect reference scale for the loss at each diffusion step σ, so wσ = a/j∗ σ with a scale factor a is a natural choice to align the loss at various σ to the same scale.

Evidence 3 - **Rationale**: This shows that Optimal Loss Estimation[52] already published the adaptive noise schedule formulation based on the loss gap, including the same mathematical expression. - **Original**: the noise schedule.p(σ)allocates the optimization frequency to each noise level. a desiredp(σ)should favor noise steps on which the optimization task has not yet been done well, which can be measured by the difference from wσjσ(θ)tow σj⋆ σ. this provides a principled measure for optimization insuffi... - **Candidate**: the noise schedule. p(σ) allocates the optimization intensity to each noise level. a desired p(σ) should favor noise steps on which the optimization task has not yet been done well, which is to minimize wσjσ(θ) down to wσj⋆ σ. therefore, the weight-calibrated loss gap provides a principled measure f...

Evidence 4 - **Rationale**: This pair shows that Optimal Loss Estimation[52] already established the theoretical justification for using the loss gap as the basis for training schedule design. - **Original**: here, we argue that analyzing thegapbetween the loss and the optimal loss would be a more principled approach, since it is the gap but not the loss itself that reflects the data-fitting insufficiency and the potential for improvement. - **Candidate**: here, we argue that analyzing the gap between the loss and the optimal loss would be a more principled approach, since it is the gap but not the loss itself that reflects the data-fitting insufficiency and the potential for improvement.

---

2. Towards Faster Training of Diffusion Models: An Inspiration of A Consistency Phenomenon

URL: View paper

Brief Assessment

Consistency Phenomenon Training[53] focuses on curriculum learning based on timestep difficulty and momentum decay strategies, not on using the gap between actual and optimal loss to determine training schedules. The approaches are fundamentally different in their design principles.

---

3. Infergrad: Improving Diffusion Models for Vocoder by Considering Inference in Training

URL: View paper

Brief Assessment

Infergrad[54] focuses on incorporating inference schedules into training for vocoder tasks, not on using optimal loss gaps to design training schedules for general diffusion models.

---

4. Analyzing and improving the training dynamics of diffusion models

URL: View paper

Brief Assessment

Training Dynamics Diffusion[7] focuses on architectural modifications to preserve activation/weight magnitudes and does not use optimal loss estimation for training schedule design. Their approach addresses training dynamics through magnitude-preserving layers rather than loss gap analysis.

---

5. Adaptive time-stepping schedules for diffusion models

URL: View paper

Brief Assessment

Adaptive Time Stepping[51] focuses on adjusting discretization points during the sampling/generation process to minimize convergence bounds, not on training schedule design using loss gaps as proposed in the original paper.

Contribution 3: Modified scaling law formulation using optimal loss offset

Description: The authors propose modifying the neural scaling law for diffusion models by subtracting the optimal loss as an offset, showing that this formulation better satisfies the power law relationship between model size and performance.

This contribution was assessed against **4 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Sdxl: Improving latent diffusion models for high-resolution image synthesis

URL: View paper

Brief Assessment

SDXL[64] focuses on architectural improvements for latent diffusion models in image synthesis, not on scaling law formulations or optimal loss estimation for diffusion models.

---

2. Towards precise scaling laws for video diffusion transformers

URL: View paper

Brief Assessment

Video DiT Scaling[63] focuses on scaling laws for video diffusion transformers by modeling optimal hyperparameters (batch size and learning rate) rather than addressing optimal loss offset in the loss formulation itself. The candidate's contribution is about hyperparameter optimization for model size selection, not about modifying the loss function by subtracting optimal loss values.

---

3. Diagnosing and Improving Diffusion Models by Estimating the Optimal Loss Value

URL: View paper

Prior Art Analysis

Optimal Loss Estimation[52] demonstrates that the modified scaling law formulation with optimal loss offset was already published. The candidate paper presents the identical modified power law j(f) - j∗ = βfα (equation 13), conducts the same experiments on ImageNet-64 and ImageNet-512 with EDM2 models (120M-1.5B parameters), and reports the same finding that subtracting optimal loss improves correlation coefficients (ρ = 0.94 vs 0.82 for single noise scale, ρ = 0.9917 for total loss). The candidate even provides the same fitted scaling law equation j(f) = 0.3675f^-0.014 + 0.015.

Evidence

Evidence 1 - **Rationale**: This pair shows that Optimal Loss Estimation[52] already published the modified scaling law approach with the same experimental setup (EDM2 models, 120M-1.5B parameters, ImageNet-64/512) and the same finding about improved power law satisfaction. - **Original**: we challenge the conventional formulation to study neural scaling law for diffusion models. we propose using the loss gap as the measure for data-fitting quality. using state-of-the-art diffusion models (karras et al., 2024) in various sizes from 120m to 1.5b on both imagenet-64 and imagenet-512, we... - **Candidate**: we challenge the conventional formulation to study neural scaling law for diffusion models. we propose using the loss gap as the measure for data-fitting quality. using state-of-the-art diffusion models [25] in various sizes from 120m to 1.5b on both imagenet-64 and imagenet-512, we find that our mo...

Evidence 2 - **Rationale**: This demonstrates that Optimal Loss Estimation[52] already published the exact modified power law formulation (equation 13) that the original paper claims as novel. - **Original**: the specialty of a scaling law study for diffusion model is that, as the optimal loss sets a non-zero lower bound of the training loss, not all the loss value inj(f)can be reduced with the increase off, questioning the form which converges to zero asf→ ∞. instead, the following modified power law is... - **Candidate**: the specialty of a scaling law study for diffusion model is that, as the optimal loss sets a non-zero lower bound of the training loss, not all the loss value in j(f) can be reduced with the increase of f, questioning the form which converges to zero as f → ∞. instead, the following modified power l...

Evidence 3 - **Rationale**: This shows that Optimal Loss Estimation[52] already published the specific quantitative results (correlation coefficients and fitted equation) that demonstrate the improved power law satisfaction with optimal loss offset. - **Original**: we can observe that in the modified version, the envelope is indeed closer to a line, and the improved correlation coefficientρ= 0.94(vs.0.82) validates this quantitatively. for the total loss, we use the optimized adaptive loss weight by edm2 (karras et al., 2024). the result is shown in fig. 4(c),... - **Candidate**: we can observe that in the modified version, the envelope is indeed closer to a line, and the improved correlation coefficient ρ = 0.94 (vs. 0.82) validates this quantitatively. for the total loss, we use the optimized adaptive loss weight by edm2 [25]. the result is shown in fig. 3(c), which achiev...

---

4. Broken neural scaling laws

URL: View paper

Prior Art Analysis

Broken Neural Scaling[65] demonstrates that subtracting an offset (specifically using a 'smoothly broken power law' functional form) from neural scaling laws was already proposed and validated across diverse tasks before the original paper's submission. The candidate paper shows that their functional form y = a + (bx^-c0) * product_terms accurately models scaling behavior by incorporating an offset constant 'a' that represents the performance limit, which is conceptually equivalent to subtracting optimal loss. This work was published at ICLR 2023 and extensively validated across vision, language, and diffusion models, directly overlapping with the original paper's core contribution of using optimal loss offset for better power law satisfaction.

Evidence

Evidence 1 - **Rationale**: Both papers claim to improve power law modeling for diffusion models and other neural networks by modifying the standard scaling law formulation. The candidate was published at ICLR 2023, establishing prior work on this approach. - **Original**: we propose using the loss gap as the measure for data-fitting quality. using state-of-the-art diffusion models (karras et al., 2024) in various sizes from 120m to 1.5b on both imagenet-64 and imagenet-512, we find that our modification leads to better satisfaction of the power law. - **Candidate**: we present a smoothly broken power law functional form (referred to by us as a broken neural scaling law(bnsl)) that accurately models and extrapolates the scaling behaviors of deep neural networks... this set includes large-scale vision, language, audio, video, diffusion, generative modeling

Evidence 2 - **Rationale**: The candidate's constant 'a' serves the same conceptual role as the original paper's optimal loss offset - both represent an irreducible baseline that must be subtracted to properly model the power law relationship. - **Original**: we challenge the conventional formulation to study neural scaling law for diffusion models. we propose using the loss gap as the measure for data-fitting quality. - **Candidate**: constant a represents the limit as to how far the value of y (performance evaluation metric) can be reduced (or maximized) even if x (the quantity being scaled) goes to infinity.

Evidence 3 - **Rationale**: Both papers validate their modified scaling law approaches on diffusion models, with the candidate demonstrating this capability in 2023, before the original paper's claimed contribution. - **Original**: using models with 120m to 1.5b parameters, we find that the power law is better demonstrated after subtracting the optimal loss from the actual training loss, suggesting a more principled setting for investigating the scaling law for diffusion models. - **Candidate**: in fig. 9, bnsl accurately extrapolates the scaling behavior of diffusion generative image models.

Evidence 4 - **Rationale**: The mathematical formulation in the candidate explicitly includes an offset term 'a' that serves the same purpose as the optimal loss offset in the original paper - both improve power law fit by accounting for an irreducible baseline. - **Original**:

moreover, using models with 120m to 1.5b parameters, we find that the power law is better demonstrated after subtracting the optimal loss from the actual training loss - **Candidate**: the general functional form of a broken neural scaling law (bnsl) is given as follows: y = a + (bx-c0) * product_terms... constant a represents the limit as to how far the value of y (performance evaluation metric) can be reduced

Appendix: Text Similarity Detection

Textual similarity detection checked 17 papers and found 3 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Diagnosing and Improving Diffusion Models by Estimating the Optimal Loss Value

Detected in: Contribution: contribution_1, Contribution: contribution_2, Contribution: contribution_3

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] Diagnosing and Improving Diffusion Models by Estimating Optimal Loss Value View paper
- [1] Reconstruction vs. generation: Taming optimization dilemma in latent diffusion models View paper
- [2] Variational diffusion models View paper
- [3] Human motion diffusion model View paper
- [4] Monocular depth estimation using diffusion models View paper
- [5] Residual denoising diffusion models View paper
- [6] Maximum likelihood training of score-based diffusion models View paper
- [7] Analyzing and improving the training dynamics of diffusion models View paper
- [8] On the edge of memorization in diffusion models View paper
- [9] Palette: Image-to-image diffusion models View paper
- [10] Forget-me-not: Learning to forget in text-to-image diffusion models View paper
- [11] Improving Discriminator Guidance in Diffusion Models View paper
- [12] Variational SchrÃ¶dinger Diffusion Models View paper
- [13] Learn to guide your diffusion model View paper
- [14] A bearing fault data augmentation method based on hybrid-diversity loss diffusion model and parameter transfer View paper
- [15] Scaling laws for diffusion transformers View paper
- [16] Perception Prioritized Training of Diffusion Models View paper
- [17] Find: Fine-tuning initial noise distribution with policy optimization for diffusion models View paper
- [18] Target-aware video diffusion models View paper
- [19] Diffusiondet: Diffusion model for object detection View paper
- [20] A new prediction strategy for dynamic multi-objective optimization using diffusion model View paper
- [21] Zero-shot contrastive loss for text-guided diffusion image style transfer View paper
- [22] DogFit: Domain-guided Fine-tuning for Efficient Transfer Learning of Diffusion Models View paper
- [23] Diffusion probabilistic model made slim View paper
- [24] Denoising diffusion probabilistic models View paper
- [25] Directly Denoising Diffusion Models View paper
- [26] Distilling diffusion models into conditional gans View paper
- [27] From Uncertain to Safe: Conformal Adaptation of Diffusion Models for Safe PDE Control View paper
- [28] Understanding and improving training-free loss-based diffusion guidance View paper
- [29] Grpose: Learning graph relations for human image generation with pose priors View paper
- [30] Exploring the Frontiers of Softmax: Provable Optimization, Applications in Diffusion Model, and Beyond View paper
- [31] DiffSG: A generative solver for network optimization with diffusion model View paper
- [32] Likelihood-based diffusion language models View paper
- [33] Structured Denoising Diffusion Models in Discrete State-Spaces View paper
- [34] Loss and likelihood based membership inference of diffusion models View paper
- [35] Diffusion Policies as an Expressive Policy Class for Offline Reinforcement Learning View paper
- [36] Understanding diffusion models: A unified perspective View paper
- [37] Model inversion attacks through target-specific conditional diffusion models View paper
- [38] D3PO: Preference-Based Alignment of Discrete Diffusion Models View paper
- [39] Fine-tuning diffusion model to generate new kite designs for the revitalization and innovation of intangible cultural heritage View paper
- [40] Fine-Tuning Text-to-Speech Diffusion Models Using Reinforcement Learning with Human Feedback View paper
- [41] Linguistic Binding in Diffusion Models: Enhancing Attribute Correspondence through Attention Map Alignment View paper
- [42] SVGDreamer: Text Guided SVG Generation with Diffusion Model View paper
- [43] Diffusion models are minimax optimal distribution estimators View paper
- [44] Training class-imbalanced diffusion model via overlap optimization View paper
- [45] Progressive Knowledge Distillation Of Stable Diffusion XL Using Layer Level Loss View paper
- [46] Mathematical Analysis of Singularities in the Diffusion Model Under the Submanifold Assumption View paper
- [47] Loss-guided diffusion models for plug-and-play controllable generation View paper
- [48] HuViDPO:Enhancing Video Generation through Direct Preference Optimization for Human-Centric Alignment View paper
- [49] A score-based density formula, with applications in diffusion generative models View paper
- [50] SSL: A Self-similarity Loss for Improving Generative Image Super-resolution View paper
- [51] Adaptive time-stepping schedules for diffusion models View paper
- [52] Diagnosing and Improving Diffusion Models by Estimating the Optimal Loss Value View paper
- [53] Towards Faster Training of Diffusion Models: An Inspiration of A Consistency Phenomenon View paper
- [54] Infergrad: Improving Diffusion Models for Vocoder by Considering Inference in Training View paper
- [55] Diffusion Models: A Mathematical Introduction View paper

- [56] Constrained diffusion models via dual training View paper
- [57] CURE: Concept Unlearning via Orthogonal Representation Editing in Diffusion Models View paper
- [58] NODI: Out-Of-Distribution Detection with Noise from Diffusion View paper
- [59] Data-driven machine learning approach based on physics-informed neural network for population balance model View paper
- [60] Optimizing diffusion models for joint trajectory prediction and controllable generation View paper
- [61] Explicit Flow Matching: On The Theory of Flow Matching Algorithms with Applications View paper
- [62] Seeds: Exponential sde solvers for fast high-quality sampling from diffusion models View paper
- [63] Towards precise scaling laws for video diffusion transformers View paper
- [64] Sdxl: Improving latent diffusion models for high-resolution image synthesis View paper
- [65] Broken neural scaling laws View paper