Enhancing Generative Modeling with Hybrid VAE-Diffusion Architectures

Subtitle: Joint Training, Adaptive Conditioning, and Latent Refinement

Project Overview

This project investigates hybrid VAE-Diffusion models, merging the VAE's efficient latent representations with the iterative refinement capabilities of diffusion models. Begin by replicating the DiffuseVAE baseline model, then implement and analyze extensions such as joint training and adaptive conditioning. Bonus challenges are available for those interested in exploring adaptive conditioning layers and latent diffusion techniques.

Introduction

The significance of hybrid VAE-Diffusion models is paramount in the evolving landscape of generative modeling. These models combine the strengths of VAEs in generating compressed latent representations and the superior output quality of diffusion models, making them a promising area of study for advancing generative techniques.

1 Stage 1: Baseline Replication of DiffuseVAE

Objective

Replicate the DiffuseVAE model to grasp the principles of two-stage hybrid VAE-Diffusion modeling.

Description of DiffuseVAE

DiffuseVAE operates in two stages:

1. A VAE generates a low-quality or coarse sample.
2. A diffusion model refines this output into a high-quality image.

This approach enables the VAE to learn compressed representations while the diffusion model corrects imperfections, yielding high-quality results but at an increased computational cost during the diffusion phase.

Implementation Tasks

- **Implement VAE and Diffusion Models:** Train a VAE for initial samples and a diffusion model for refinement.
- **Dataset:** Utilize CIFAR-10 and CelebA-64.

Evaluation Criteria

- **Quantitative:** Compare FID and IS metrics against established benchmarks to validate baseline performance.
- **Qualitative:** Visually inspect outputs for coherence, sharpness, and refinement levels.
- **Efficiency:** Measure sampling time and the number of diffusion steps.

Practical Tips

- **Debugging:** If outputs appear overly smooth, examine the VAE's latent space variance; high variance may indicate underfitting.
- **Computational Load:** Start with smaller sample batches to expedite training and testing.

Expected Output

Documented baseline results of the DiffuseVAE model, including FID, IS, and sampling time metrics for both CIFAR-10 and CelebA-64 datasets, serving as a reference for later extensions.

2 Stage 2: Extension 1 - Joint End-to-End Training

Objective

Experiment with joint training of the VAE and diffusion model to synchronize latent space and refinement outputs.

Description

Joint training aligns the VAE and diffusion processes, potentially enhancing initial sample quality and reducing refinement demands. Begin with a basic weighted sum of the VAE's ELBO and the diffusion model's log-likelihood, then explore more intricate hybrid loss functions as confidence builds.

Implementation Tasks

- **Implement Joint Training:** Start with a basic weighted sum, adjusting weights incrementally to balance loss terms.
  - **Guidance on Weight Tuning:** Initiate with a low weight for the diffusion model's loss and increase as necessary; if instability arises, boost the diffusion model's weight relative to the VAE.
- **Incremental Loss Experimentation:**
  - (Optional) Explore complex hybrid loss functions, such as adaptive combinations or annealing weights.

Evaluation Criteria

- **Quantitative:** Compare FID, IS, and sampling times to the baseline.
- **Latent Space Alignment:** Visualize latent space interpolations to assess improvements in smoothness and coherence from joint training.
- **Qualitative:** Examine reductions in artifacting or smoothing effects in refined outputs.

Practical Tips

- **Stability:** Use small learning rates initially to prevent training instability.
- **Loss Weighting:** Incremental adjustments help maintain stability; advanced hybrid losses are optional.

Expected Output

Documentation on the impacts of joint training on sample quality and generation efficiency, with a comparison of basic and exploratory hybrid losses.

3 Stage 3: Extension 2 - Adaptive Conditioning with Flow-Based Prior

Objective

Implement adaptive conditioning in the diffusion model to adjust refinement based on the quality of VAE outputs, utilizing a flow-based prior to enable a more flexible latent space.

Description

The fixed conditioning in DiffuseVAE does not account for variability in VAE output quality, necessitating differing refinement levels in the diffusion phase. Adaptive conditioning dynamically adjusts based on output quality indicators, with latent space variance serving as the primary measure of uncertainty.

Adaptive Conditioning Possibilities

- **Uncertainty-Based Conditioning:**
  - Use the variance σ^2 from the VAE's latent distribution $q_\phi(z|x)$ as an uncertainty indicator, guiding refinement levels.
- **(Bonus Challenge) Adaptive Conditioning Layer:**
  - Implement a normalizing flow (e.g., RealNVP or Glow) as an adaptive conditioning layer to modify VAE outputs dynamically based on uncertainty, enhancing the diffusion model's refinement quality.

Implementation Tasks

- **Implement Flow-Based Prior in VAE:** Replace the Gaussian prior with a flow-based prior[1] to enhance the expressive capacity of the latent space and improve initial sample quality.
- **Implement Adaptive Conditioning:** Utilize latent space variance to dynamically modulate conditioning strength.

Evaluation Criteria

- **Quantitative:** Compare FID, IS, and output consistency across different conditioning strategies.
- **Variance Sensitivity:** Assess how adaptive conditioning responds to uncertain VAE outputs, noting the quality of refinements.
- **Efficiency:** Record sampling times, highlighting reductions attributable to improved initial sample quality from the flow-based prior.

Practical Tips

- **Debugging:** If adaptive conditioning does not yield quality improvements, check the scale of latent variance; excessively large values may indicate poor initialization.
- **Visualization:** Include before-and-after images of conditioned diffusion outputs to illustrate the effects of adaptive conditioning.

[1] For more details, visit `https://jmtomczak.github.io/blog/7/7_priors.html`.

Expected Output

Analysis of adaptive conditioning effectiveness, comparing various conditioning strategies and emphasizing improvements in quality, robustness, and efficiency.

4 Bonus Challenge: Latent-Diffusion Hybrid

Objective

Implement latent-space diffusion to enhance efficiency while maintaining sample quality.

Description

Conducting diffusion in the VAE's latent space reduces computational costs while preserving high-fidelity outputs, streamlining the diffusion process by leveraging a lower-dimensional latent space.

Implementation Tasks

- **Implement Latent Space Diffusion:** Apply the diffusion model within the VAE's latent space, using the VAE decoder to convert refined latents back to image space.

Evaluation Criteria

- **Quantitative:** Record FID, IS, and sample diversity for latent vs. pixel-space diffusion.
- **Efficiency:** Report reductions in sampling time and computational resources.

Practical Tips

- **Efficiency Measurement:** Use profiling tools (e.g., PyTorch Profiler) to monitor memory usage and computation in both latent and pixel spaces.
- **Interpretation:** Report efficiency gains in relative terms (e.g., percentage reductions) for clearer insights.

Expected Output

A performance comparison demonstrating the computational advantages and quality trade-offs of latent diffusion relative to pixel diffusion.

5 Final Evaluation and Analysis

Objective

Synthesize results from all stages to systematically evaluate each model extension.

Structured Evaluation Rubric

- **Guiding Questions:**
  - Quality: "Does joint training improve FID? By how much?"
  - Efficiency: "How does adaptive conditioning affect sampling time?"
  - Robustness: "How does each model handle uncertain or noisy VAE outputs?"
  - Latent Space Coherence: "Is the latent space smoother or more interpretable with joint training?"

Evaluation Criteria

- **Quantitative and Qualitative:** Measure FID, IS, and sample diversity.
- **Comparative Insight:** Use structured questions to guide the analysis for each model variant.

Practical Tips for Analysis

- **Clear Visualizations:** Include comparison tables or charts to illustrate relative improvements.
- **Interpretation:** Focus on specific findings rather than overly complex analyses; concise answers to rubric questions will enhance clarity.

Appendix: Evaluation Metrics Overview

FID (Fréchet Inception Distance)

- **Pros:** Measures distributional similarity between generated and real samples, reflecting overall quality.
- **Cons:** Sensitive to batch size; potentially misleading with small sample sizes.
- **Pitfalls:** Low FID does not guarantee individual image quality but indicates distributional alignment.

Inception Score (IS)

- **Pros:** Balances quality and diversity of generated samples.
- **Cons:** Biased toward classes in ImageNet; may overestimate quality with noisy samples.
- **Pitfalls:** Can yield high scores even if some generated samples are flawed, so interpret alongside FID.

Qualitative Assessments

Focus on coherence (object consistency), sharpness (detail clarity), and structural integrity (how well objects retain structure across samples).

Additional Guidance

- **Data Preprocessing:** Ensure to follow specific preprocessing steps for CIFAR-10 and CelebA-64, including normalization and resizing. Refer to relevant libraries for assistance with preprocessing tasks.
- **Documentation Format:** The final report should be organized into sections such as "Introduction," "Methodology," "Results," "Discussion," and "Conclusion" to ensure comprehensive documentation of findings.
- **Critical Reflection:** In the final evaluation stage, include a reflective section discussing lessons learned, challenges encountered, and potential future improvements. This can deepen critical thinking and understanding of the subject matter.

Table 1: Grading Rubric

Criterion	**Excellent (5)**	**Good (4)**	**Satisfactory (3)**	**Needs Improvement (2)**	**Unsatisfactory (1)**
Stage 1: Baseline Replication	Accurate replication of DiffuseVAE; comprehensive understanding demonstrated.	Minor errors in replication; good understanding shown.	Basic replication; some gaps in understanding.	Significant errors; lacks clarity in understanding.	Failure to replicate; minimal effort shown.
Stage 2: Joint Training Implementation	Clear implementation of joint training; effective tuning of weights.	Implementation mostly correct; some tuning issues.	Basic implementation; significant tuning needed.	Poor implementation; unclear understanding of joint training.	No implementation; fails to grasp the concept.
Stage 3: Adaptive Conditioning	Innovative application of adaptive conditioning; effective use of flow-based prior.	Good implementation; minor issues with adaptive conditioning.	Basic application; lacks depth and clarity.	Poorly executed; significant misunderstandings.	No attempt or serious flaws in understanding.
Evaluation and Results	Thorough and insightful evaluation addressing all guiding questions; clear metrics analysis (e.g., FID and IS).	Good evaluation; addresses most guiding questions with relevant metrics analysis.	Basic evaluation; some guiding questions answered; metrics analysis is superficial.	Inadequate evaluation; few guiding questions addressed; limited metrics analysis.	Minimal evaluation; fails to address guiding questions or present metrics.
Documentation and Presentation	Well-organized and clear documentation; follows specified format precisely.	Organized documentation; minor formatting issues.	Basic documentation; lacks organization or clarity.	Poorly organized; significant formatting errors.	Minimal documentation; difficult to follow.
Critical Reflection	Thoughtful reflections on challenges and learnings; identifies potential improvements.	Good reflections; some insights on challenges.	Basic reflections; limited insight into improvements.	Minimal reflection; lacks depth in understanding challenges.	No reflection or serious flaws in analysis.

References

[1] Pandey, Kushagra, Mukherjee, Avideep, Rai, Piyush, and Kumar, Abhishek. *DiffuseVAE: Efficient, controllable and high-fidelity generation from low-dimensional latents.* arXiv preprint arXiv:2201.00308, 2022.