Classifier-Free Diffusion Guidance

Author: Ho et al.

created by paper2slides

Uploaded to arXiv: 2022-07-26

https://arxiv.org/abs/2207.12598

Executive Summary

- Diffusion models have emerged as a strong generative model for image and audio synthesis.
- Classifier guidance improves sample quality at the cost of diversity.
- Introduced classifier-free guidance which removes the need for classifiers.
- Jointly train conditional and unconditional diffusion models.
- Achieves a trade-off between sample quality and diversity similar to classifier guidance.

Introduction: Diffusion Models

- Diffusion models generate data by gradually denoising from Gaussian noise.
- Competitive sample quality and likelihood scores for image and audio synthesis.
- Achieved state-of-the-art results on datasets like ImageNet.

Introduction: Challenge with Classifier Guidance

- Classifier guidance improves sample quality by combining classifier gradients.
- Requires an extra classifier trained on noisy data.
- Problematic as it resembles gradient-based adversarial attacks.
- Motivates exploring guidance without a separate classifier.

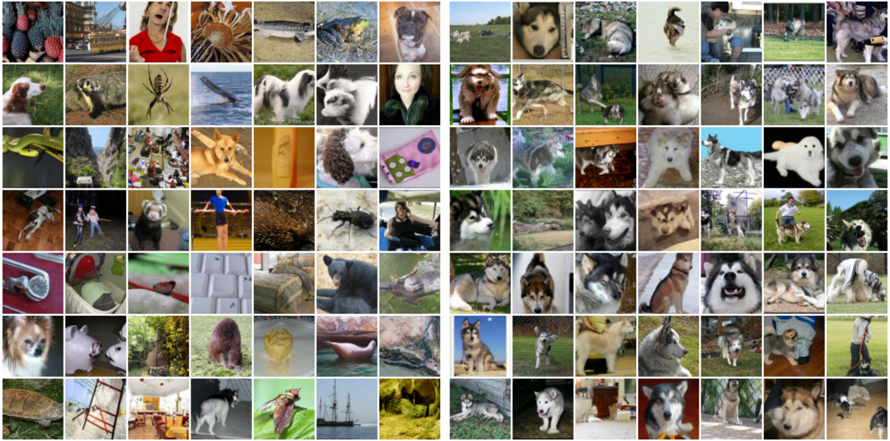

Figure: Classifier-free guidance on the malamute class for a 64×64 ImageNet diffusion model.

Proposed Method: Classifier-Free Guidance

- Train a joint model for conditional and unconditional diffusion.
- Conditional model: $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c})$.
- Unconditional model: $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda)$.
- Generate samples using a weighted sum of the scores:

$$\tilde{\boldsymbol{\epsilon}}_\theta(\mathbf{z}_\lambda, \mathbf{c}) = (1 + w)\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c}) - w\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda).$$

Algorithm: Joint Training

- Sample data $(\mathbf{x}, \mathbf{c}) \sim p(\mathbf{x}, \mathbf{c})$.
- Randomly discard conditioning $\mathbf{c} \leftarrow \varnothing$ with probability p_{uncond}.
- Train on corrupted data $\mathbf{z}_\lambda = \alpha_\lambda \mathbf{x} + \sigma_\lambda \boldsymbol{\epsilon}$.
- Optimization of denoising model:

$$\nabla_\theta \left\| \boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c}) - \boldsymbol{\epsilon} \right\|^2 .$$

Algorithm: Conditional Sampling

- Initialize $\mathbf{z}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
- Form the classifier-free guided score at each step:

$$\tilde{\boldsymbol{\epsilon}}_t = (1 + w)\boldsymbol{\epsilon}_\theta(\mathbf{z}_t, \mathbf{c}) - w\boldsymbol{\epsilon}_\theta(\mathbf{z}_t).$$

- Perform sampling step iterated for T timesteps.

Experiments: Classifier-Free Guidance

- Evaluated on 64$\times$64 and 128$\times$128 ImageNet generation.
- Range of guidance strengths $w \in \{0, 0.1, \ldots, 4\}$.
- Metrics: FID and Inception Scores.

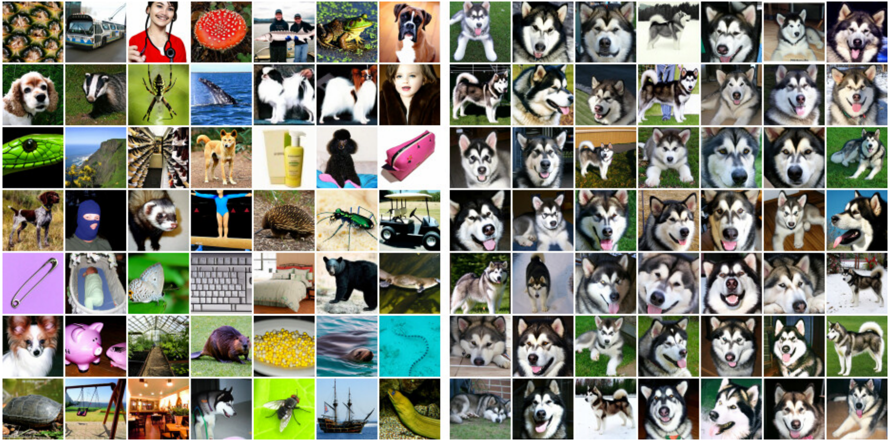

Figure: Classifier-free guidance with high strength demonstrates high fidelity samples with decreased diversity.

Results: 64×64 ImageNet

- Best FID: 1.55 with $w = 0.1$.
- Best IS: 260.2 with $w = 4.0$.
- Maintains competitive results compared to state-of-the-art methods.
- Provides a clear trade-off between sample fidelity and diversity.

Results: 128×128 ImageNet

- Best FID: 2.43 with $w = 0.3$.
- Best IS: 422.29 with $w = 4.0$.
- Outperforms previous methods with fewer parameters.
- Demonstrates state-of-the-art performance on large-scale datasets.

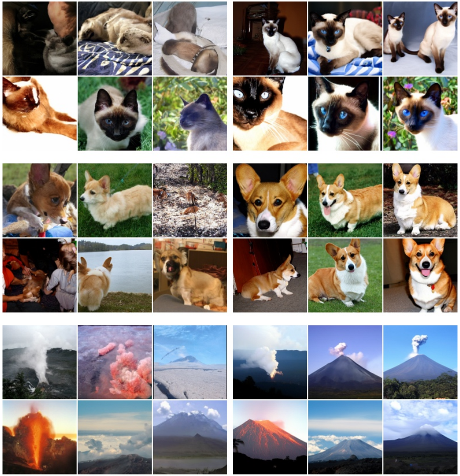

Figure: 128×128 ImageNet generation: Non-guided samples (left) vs. classifier-free guidance (right).

Conclusion

- Introduced classifier-free guidance for diffusion models.
- Removes dependency on extra classifiers during sampling.
- Accomplishes superior FID and IS trade-offs similar to classifier guidance.
- Future work: speed up sampling, explore applications to other data modalities.