ONE STEP DIFFUSION MODEL USING DISTILLATION TECHNIQUES

**1. Diffusion Process**

In the diffusion process, data is gradually noised over multiple steps, usually modeled as a Markov chain. Starting from clean data, noise is added step-by-step until the data becomes pure noise.

**2. Reverse Process**

The reverse process is typically the generation phase, where the model learns to denoise the data, step-by-step, to revert from pure noise back to the original data distribution. This process can be very slow since it often requires many steps to generate a single sample.

**3. Distillation Technique**

Distillation is a technique where a complex model (teacher) is used to train a simpler model (student) to mimic the behavior of the teacher. In the context of diffusion models:

* **Teacher Model**: The original multi-step diffusion model.
* **Student Model**: The distilled one-step diffusion model.

**Steps to Create a One-Step Diffusion Model Using Distillation:**

1. **Train the Teacher Model**:
   * Train a traditional diffusion model that performs gradual denoising over many steps.
2. **Generate Intermediate Data**:
   * Use the teacher model to generate intermediate noisy data at various steps. This creates a dataset of (noisy, clean) pairs.
3. **Distill the Knowledge**:
   * Train the student model to learn the mapping from noisy data (from any step) to clean data directly in one step. This can be done using supervised learning where the input is the noisy data, and the output is the corresponding clean data.
4. **Loss Function**:
   * The loss function can be a simple L2 loss (mean squared error) between the predicted clean data and the actual clean data. Other perceptual losses or GAN-based losses can also be used to improve quality.
5. **Optimization**:
   * Use standard optimization techniques (e.g., gradient descent) to train the student model.

Benefits

* Speed
* Efficiency

Challenges

* Quality
* Training complexity

**3. Mathematical Representation**

Let x0x\_0x0​ represent the original clean data. The noising process can be represented as: where:

* xtx\_txt​ is the noisy data at step ttt.
* αt\alpha\_tαt​ is a schedule of noise levels decreasing from 1 to 0 as ttt increases.
* ϵ\epsilonϵ is random noise, typically Gaussian noise.

**4. Reverse Process**

The reverse process aims to revert xtx\_txt​ back to x0x\_0x0​ (the clean data). This is learned through training on many such noisy and clean pairs: x^0=fθ(xt,t)\hat{x}\_0 = f\_\theta(x\_t, t)x^0​=fθ​(xt​,t) where fθf\_\thetafθ​ is the model with parameters θ\thetaθ that predicts the clean data from noisy data.

TEXT TO IMAGE CONVERSION

* Text Encoding(Text embedding(BERT, GPT, CLIP)->>>conditioning vector)
* Diffusion Process: start with a high resolution image and gradually add noise and hence produce a really noisy image
* Denoising : using the conditioning vector as a guide gradually denoise the image step by step and simultaneously train the model, conditioned by the text vector. This process is called reversing.
* Train the model: datset(a large set of images with text descriptions).perform both forward training and reverse training.
* One step diffusion using distillation: train the student model to map the extremely noisy images to high resolution images, with the help of muti step teacher model.

**Practical Implementations**

Several state-of-the-art models and research have explored text-to-image generation using diffusion models:

* **DALL·E 2**: Uses a diffusion prior where images are generated from text descriptions.
* **Imagen**: A diffusion model that achieves high fidelity in text-to-image generation.
* **Stable Diffusion**: A latent diffusion model that generates images conditioned on text.

REFERENCE PAPERS

1. <https://openaccess.thecvf.com/content/CVPR2024/papers/Yin_One-step_Diffusion_with_Distribution_Matching_Distillation_CVPR_2024_paper.pdf>

* Diffusion with distribution
* Minimise KL-divergence(Kullback-Leibler): difference between the predictions of the obtained model and the expected one.
* 2.62 FID on ImageNet 64⇥64 and 11.49 FID on zero-shot COCO-30k,
* FID – Frechet Inception Distance, lower the better: measures the distance between the two multivariate guassian models, essentially the predicted one and the real one.

#### Acceleration Approaches:

1. **Fast Diffusion Samplers**: These reduce the number of sampling steps but often at the cost of performance.
2. **Diffusion Distillation**: This method frames the acceleration as knowledge distillation, where a student model learns to perform the multi-step process in one step. Previous approaches like Luhman et al. and DSNO have high computational costs due to full denoising trajectory calculations. ProgressiveDistillation (PD), InstaFlow, and Consistency Distillation (CD) offer alternative methods but still have challenges.

* Combining the stability of GAN training with regression losses, our method achieves high realism on complex datasets like LAION. Unlike GAN-based methods, our approach specializes in diffusion distillation, introducing regression loss and demonstrating state-of-the-art results for text-to-image tasks.
* By leveraging lower-rank matrices, LoRA offers a more efficient and cost-effective approach to model adaptation, significantly reducing the trainable parameters and GPU memory requirements, thus enabling faster training and memory efficiency.

While diffusion models can learn complex distributions, sampling requires a compu-

tationally expensive iterative process. Existing distillation methods enable efﬁcient

sampling, but have notable limitations, such as performance degradation with very

few sampling steps, reliance on training data access, or mode-seeking optimization

that may fail to capture the full distribution. We propose EM Distillation (EMD), a

maximum likelihood-based approach that distills a diffusion model to a one-step

generator model with minimal loss of perceptual quality. Our approach is derived

through the lens of Expectation-Maximization (EM), where the generator parame-

ters are updated using samples from the joint distribution of the diffusion teacher

prior and inferred generator latents. We develop a reparametrized sampling scheme

and a noise cancellation technique that together stabilizes the distillation process.

We further reveal an interesting connection of our method with existing methods

that minimize mode-seeking KL. EMD outperforms existing one-step generative

methods in terms of FID scores on ImageNet-64 and ImageNet-128, and compares

favorably with prior work on distilling text-to-image diffusion models

While diffusion models can learn complex distributions, sampling requires a compu-

tationally expensive iterative process. Existing distillation methods enable efﬁcient

sampling, but have notable limitations, such as performance degradation with very

few sampling steps, reliance on training data access, or mode-seeking optimization

that may fail to capture the full distribution. We propose EM Distillation (EMD), a

maximum likelihood-based approach that distills a diffusion model to a one-step

generator model with minimal loss of perceptual quality. Our approach is derived

through the lens of Expectation-Maximization (EM), where the generator parame-

ters are updated using samples from the joint distribution of the diffusion teacher

prior and inferred generator latents. We develop a reparametrized sampling scheme

and a noise cancellation technique that together stabilizes the distillation process.

We further reveal an interesting connection of our method with existing methods

that minimize mode-seeking KL. EMD outperforms existing one-step generative

methods in terms of FID scores on ImageNet-64 and ImageNet-128, and compares

favorably with prior work on distilling text-to-image diffusion models

<https://www.researchgate.net/publication/380907314_EM_Distillation_for_One-step_Diffusion_Models>

* While diffusion models can learn complex distributions, sampling requires a compu-tationally expensive iterative process. Existing distillation methods enable efﬁcientsampling, but have notable limitations, such as performance degradation with veryfew sampling steps, reliance on training data access, or mode-seeking optimizationthat may fail to capture the full distribution. We propose EM Distillation (EMD), amaximum likelihood-based approach that distills a diffusion model to a one-stepgenerator model with minimal loss of perceptual quality. Our approach is derivedthrough the lens of Expectation-Maximization (EM), where the generator parame-ters are updated using samples from the joint distribution of the diffusion teacherprior and inferred generator latents. We develop a reparametrized sampling schemeand a noise cancellation technique that together stabilizes the distillation process.We further reveal an interesting connection of our method with existing methodsthat minimize mode-seeking KL. EMD outperforms existing one-step generativemethods in terms of FID scores on ImageNet-64 and ImageNet-128, and comparesfavorably with prior work on distilling text-to-image diffusion models

This research paper introduces "EM Distillation (EMD)," a novel method for efficient diffusion model distillation. Diffusion models have revolutionized high-quality generative tasks across various data types by transforming complex distributions into Gaussian distributions through a sequence of intermediate distributions. However, sampling from these models typically requires multiple evaluations, making real-time generation challenging. To address this, the authors propose EMD, which minimizes an approximation of the mode-covering divergence between a pre-trained diffusion model (teacher) and a latent-variable student model, allowing efficient generation in just one step.

**Main Contributions:**

1. **EM Distillation (EMD) Framework:**
   * The proposed method uses an Expectation-Maximization (EM) framework where the E-step involves Monte Carlo sampling to estimate learning gradients, and the M-step updates the student model through gradient ascent.
   * To address the high cost of direct sampling from the teacher model, the authors introduce a modified Markov Chain Monte Carlo (MCMC) sampling scheme that jointly updates data and latent variable pairs initialized from the student samples. This approach accelerates convergence and simplifies hyperparameter tuning.
2. **Langevin Dynamics and Noise Cancellation:**
   * Langevin dynamics is used for MCMC sampling, but the process is reparameterized to ensure uniform step sizes across noise levels, improving performance.
   * The authors identify and remove accumulated noise during the sampling process to enhance learning efficiency and gradient estimation accuracy.
3. **Empirical Validation:**
   * The EMD method demonstrates superior performance in one-step generation tasks, outperforming existing approaches in both ImageNet conditional generation and text-to-image generation using Stable Diffusion models.
   * EMD achieves state-of-the-art FID scores on ImageNet-64 and ImageNet-128 datasets.

**Technical Insights:**

1. **Diffusion Models and Score Matching:**
   * The paper reviews the fundamentals of diffusion models, which progressively denoise data to recover the original distribution, and score matching, which estimates the score function of the noisy data distribution using neural networks.
2. **Maximum Likelihood and Expectation-Maximization:**
   * The EM framework's application to latent variable models is discussed, emphasizing the use of MCMC techniques to approximate posterior distributions and optimize model parameters.
3. **Variational Score Distillation (VSD) and Diff-Instruct:**
   * The authors highlight the connections between their method and existing distillation techniques like VSD and Diff-Instruct, showing that EMD can interpolate between mode-seeking and mode-covering divergences through different sampling schemes.

**Algorithmic Implementation:**

* The paper presents detailed algorithms for EMD, including the joint MCMC sampling process and the optimization steps for both the generator and score networks.