DREAM: Diffusion Rectification and Estimation-Adaptive Models

1 Overview

DREAM (Diffusion Rectification and Estimation-Adaptive Models) enhances diffusion-based models for image super-resolution by improving the training-sampling alignment, leading to better performance. It offers significant improvements in training speed and image quality with minimal code modifications. DREAM features two components: diffusion rectification, which adjusts training to reflect the sampling process, and estimation adaptation, which balances perception against distortion. When applied to image super-resolution (SR), DREAM adeptly navigates the tradeoff between minimizing distortion and preserving high image quality. Experiments demonstrate DREAM's superiority over standard diffusion-based SR methods, showing a 2 to 3× faster training convergence and a 10 to 20× reduction in sampling steps to achieve comparable results. DREAM aims to encourage a reconsideration of the training methods used for diffusion models.

The source code is available on GitHub: DREAM GitHub Repository.

2 Dataset

The folder `celebahq_16_128.zip` in the above-mentioned GitHub repository contains images for the CelebA-HQ dataset, with the numbers 16 and 128 referring to the image resolutions in pixels. Below are the contents of the folder:

- `hr_128`: High-resolution images (128 × 128 pixels).
- `lr_16`: Low-resolution images (16 × 16 pixels).
- `sr_16_128`: Super-resolution images, where the model upscales the 16 × 16 images to 128 × 128 pixels.

3 Compute Required

The DREAM project for Diffusion Rectification and Estimation-Adaptive Models is designed to handle high-computation tasks, particularly for image super-resolution (SISR) using diffusion models. These tasks require significant computational resources, especially when training on large datasets like CelebA-HQ or DIV2K and performing high-resolution super-resolution (e.g., upscaling 16 × 16 images to 128 × 128). Below are resources typically required for the DREAM project:

- **GPUs**: Requires 1–2 NVIDIA RTX A5000 (24 GB) or equivalent for efficient training.
- **System RAM**: At least 64 GB of RAM is recommended for handling large datasets.
- **VRAM**: Each GPU should have at least 24 GB VRAM for memory-intensive diffusion models.
- **Storage**: Requires at least 500 GB of disk space for datasets, model checkpoints, and outputs.
- **Software**: Uses PyTorch, CUDA, and cuDNN for deep learning and GPU acceleration.
- **Training Time**: Training typically takes several days, with DREAM reducing time by 2–3×.
- **Power & Cooling**: Adequate power supply and cooling are necessary to support high-performance GPUs.

4 DREAM Training Strategy

The DREAM training strategy consists of two core components that aim to bridge the training-sampling discrepancy seen in diffusion models. This discrepancy arises because the model operates differently during training (using ground-truth data) and during inference (sampling from its own previous outputs). DREAM addresses this by introducing Diffusion Rectification and Estimation Adaptation to improve both training efficiency and image quality:

- **Diffusion Rectification**: Adjusts the training process by adding a forward pass that allows the model to use its own predictions during training. This helps align training with the sampling process, reducing errors caused by discrepancies between training (with ground-truth data) and sampling (with model-generated predictions).
- **Estimation Adaptation**: Balances the trade-off between minimizing distortion and preserving perceptual quality by adaptively incorporating ground-truth information. This ensures that the model focuses more on the ground truth at early steps and relies on its predictions at later stages.

DREAM accelerates training convergence by 2-3× and reduces sampling steps by 10-20× while achieving superior image super-resolution results. This strategy requires minimal changes to existing diffusion model code and can be applied to various super-resolution tasks.