Project Proposal

Team: Cem Beyenal, Mason Menser

Intentions

We intend to evaluate the effectiveness of Denoising Diffusion Implicit Models (DDIMs) in comparison to other generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs).

There are already existing diffusion-based models, such as the Denoising diffusion Probabilistic Models (DDPMs), however they are computationally expensive due to longer training and sampling times. DDIMs are a newer variant of this, and the goal is to speed up the training and sampling time, while maintaining the same standard of quality.

Methods

Our methods would involve using a GAN or VAE model to produce some form of images, (example for now could be CIFAR-10) and track evaluation metrics such as training time, sampling time, and sample quality. Once we have the results of the GAN/VAE model, we will replicate the experiment using a DDIM instead. We will have the same evaluation metrics for the DDIM as the GAN/VAE and compare them for our findings.

Data

We will use data to compare DDIMs with a GAN. We will look at several different metrics.

Metric	Explanation
Training Time	Efficiency
Sampling Time	Generation speed - time to produce images
Sample Quality	Fidelity of output - can use quantitative metrics like Frechet Inception Distance, Inception Score, or Kernel Inception Distance
Turing Test/Visual Comparision	Could ask participants to guess real vs generated samples or just quality of both samples, see if one model performs better

We also plan on using multiple different datasets. We will use the CIFAR-10 dataset to train and compare DDIM and GAN models. We also plan on using CelebA to compare the models with more realistic generation outputs like human faces. If need be and time permits, we can add more datasets to compare like MNIST (grayscale digits) and SVHN (house numbers).

Background

Relevant background would include knowledge of GANs and their goal to learn a data distribution so that they can produce new, realistic samples. They have been used in image generation since their birth in 2014 and achieve high-quality realistic data, however they are difficult to train due to limited diversity in training and the need for large training datasets. VAEs are typically more stable to train than GANs due to relying on variational inference and evidence of lower bound optimization rather than adversarial competition. However, VAEs often produce more detailed images than GANs due to the constraints of their probabilistic decoders.

We also need to understand DDIMs as background work. They're also a generative model that can create data/images, similar to GANs. DDIMs seek to strike a balance between GANs fast training times yet unstable output and DDPMs slow training time but high quality output.

To start, we will explore using Google Colab to train these models on our first dataset.

Tentative Plan

Date	Goal
10/15	Dataset Preparation
10/20	Train a DDIM & GAN on first dataset
10/25	Progress Report Due
10/30	Compare DDIM results in generating CIFAR-10 images with established GAN experiments
11/15	Train with different dataset & compare different models
12/1	Prepare findings
12/9	Final Report