Latent Constraints: Generating Conditionally from Unconditional Generative Models

Chavda Haarit Ravindrakumar (23110077)

Introduction

Unconditional models like GANs and VAEs are largely used for generating data samples. These generated outputs are random and we do not have control over the attributes. In practice, however, we require data with specific properties. Conditional GANs address this by allowing control over the generated data. But retraining existing models is very expensive. Therefore, we aim to leverage the latent space distribution of a pre-trained model to generate conditional samples.

Summarizing the problem statement: Use pre-trained unconditional generative models to generate conditional samples.

Preliminary Work

Explored the paper "Latent Constraints" which aims to do the same and partially replicated its approach. The remainder of the findings and observations in this report are based on that paper.

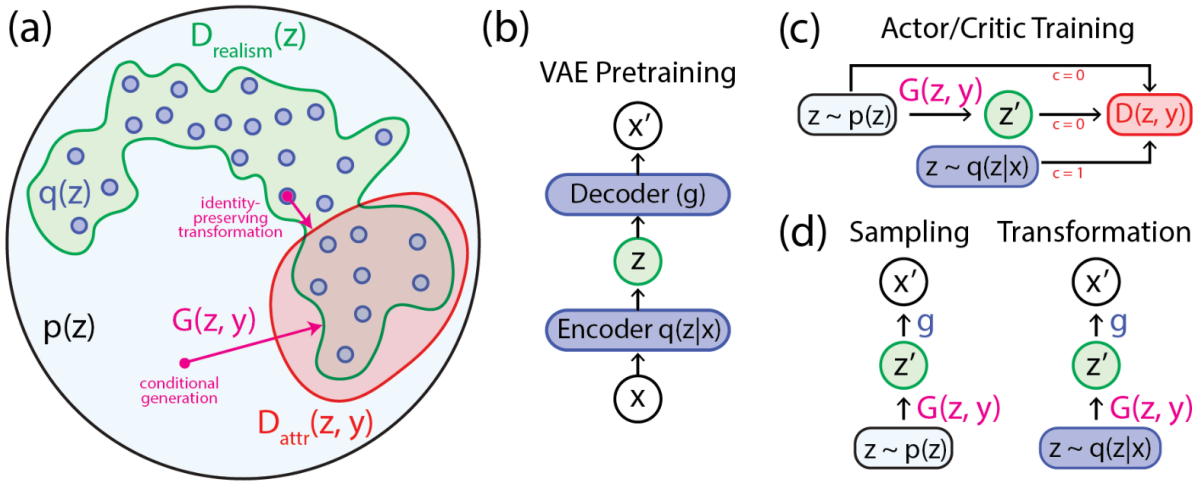

Figure 1: (a) Diagram of latent constraints for a VAE. We use one critic D_{attr} to predict which regions of the latent space will generate outputs with desired attributes, and another critic D_{realism} to predict which regions have high mass under the marginal posterior, $q(z)$, of the training data. (b) We begin by pretraining a standard VAE, with an emphasis on achieving good reconstructions. (c) To train the actor-critic pair we use constraint-satisfaction labels, c, to train D to discriminate between encodings of actual data, $z \sim q(z|x)$, versus latent vectors $z \sim p(z)$ sampled from the prior or transformed prior samples $G(z \sim p(z), y)$. Similar to a Conditional GAN, both G and D operate on a concatenation of z and a binary attribute vector, y, allowing G to learn conditional mappings in latent space. If G is an optimizer, a separate attribute discriminator, D_{attr} is trained and the latent vector is optimized to reduce the cost of both D_{attr} and D_{realism}. (d) To sample from the intersection of these regions, we use either gradient-based optimization or an amortized generator, G, to shift latent samples from either the prior ($z \sim p(z)$, sampling) or from the data ($z \sim q(z|x)$, transformation).

Figure 1. Diagram for illustrating the latent space structure

Concept (Derived from the paper)

Combining two key constraints: a realism constraint, which ensures that generated samples resemble real data by aligning with the posterior distribution of the training data, and an attribute constraint, which enforces specific desired attributes in the generated samples. The paper uses an actor-critic architecture, where the actor maps random latent vectors sampled from a prior distribution to regions of the latent space that satisfy both constraints, and the critic evaluates whether these mappings meet the realism and attribute requirements. This process can be implemented using gradient-based optimization or an amortized actor network, avoiding the need to retrain the generative model. The approach is computationally efficient and versatile, supporting tasks like conditional sampling, identity-preserving transformations, and zero-shot conditional generation even in the absence of labeled data. By leveraging these latent constraints, the framework bridges the gap between reconstruction quality and sample quality while maintaining diversity in outputs.

The Actor Loss Function:

$$L_{c=1}(z) = -\log(D(z))$$
$$L_{c=0}(z) = -(1 - \log(D(z)))$$

$$L_D(z) = \mathbb{E}_{z\sim q(z|x)}[L_{c=1}(z)] + \mathbb{E}_{z\sim p(z)}[L_{c=0}(z)] + \mathbb{E}_{z\sim G(p(z))}[L_{c=0}(z)]$$

where:

$$\begin{aligned} \mathbb{E}_{z\sim q(z|x)}[L_{c=1}(z)] &\quad \text{(BCE b/w z and marginal posterior)} \\ \mathbb{E}_{z\sim p(z)}[L_{c=0}(z)] &\quad \text{(BCE b/w z and prior)} \\ \mathbb{E}_{z\sim G(p(z))}[L_{c=0}(z)] &\quad \text{(BCE b/w z and generated distribution)} \end{aligned}$$

The Critic Loss Function:

$$L_G(z) = \mathbb{E}_{z\sim p(z)}\left[L_{c=1}(G(z)) + \lambda_{\text{dist}} L_{\text{dist}}(G(z), z)\right]$$

where the distance loss function is defined as:

$$L_{\text{dist}}(z', z) = \frac{1}{\bar{\sigma}_z^2} \log\left(1 + (z' - z)^2\right)$$

with the averaged variance:

$$\bar{\sigma}_z = \frac{1}{N}\sum_n \sigma_z(x_n) \quad \text{(Averaged over all samples)}$$

Experiment on CelebA Dataset

As this is replicated from the paper, the model architecture and loss function are the same as those in the paper.

First, an unconditional VAE is trained on the CelebA dataset. The latent space dimension of this

model is 1024. Then an actor-critic pair is trained on the latent space to learn the latent vectors corresponding to specific attributes. The input for the generator (the actor) is $1024 + 10 = 1034$, where 1024 values come from the latent vectors and the one-hot encoded attributes are padded to each latent vector.

It is evident from the results that when we pad the attributes vector to the latent vector of an image and pass it through the actor-critic pair, it produces another latent vector corresponding to the attributes, which will reconstruct the image with the required attributes.

In this experiment, the identity of the original image is maintained by adding a regularisation term to the loss function. Therefore, it appears that the same bald person has worn eyeglasses as shown in the last row.

Figure 2. With Identity Preservation

A subset of this experiment was to train without identity preservation, which would generate images from the same reconstructed image but where the identity is not retained, although the image still satisfies both the realism and attribute constraints. This is shown in the following figure.

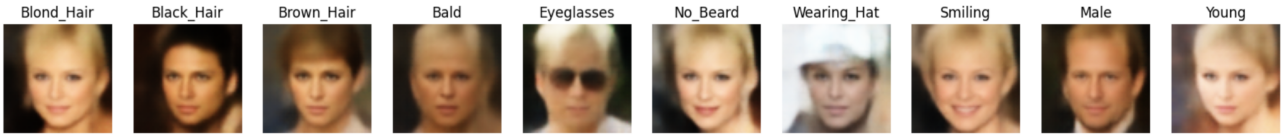

Figure 3. Identity Not Preserved