Introducing Dynamic Latent Scale GAN for GAN Inversion

Jeongik Cho

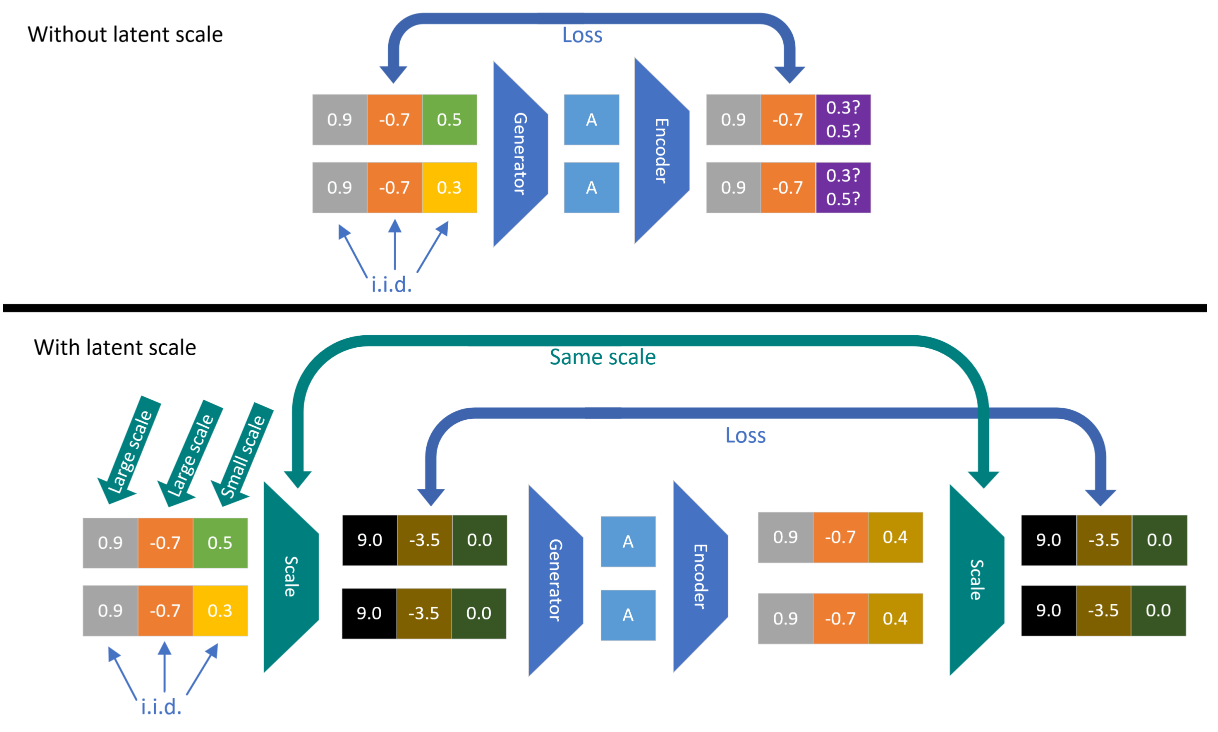
A generative model maps a simple, easy-to-sample latent distribution into a complex data distribution. Inverting this process—referred to as generative model inversion—involves mapping observed data back into the latent space, which enables representation learning and can support tasks such as data preprocessing and multimodal modeling.

The most basic way to perform generative model inversion is to train an encoder that inverts the generative model using a mean squared error (MSE) loss. In this case, the loss used for the encoder is as follows:

$$L_{enc} = \text{avg}\left((Z - E(G(Z)))^2\right)$$

In the above equation, E and G denote the encoder and generator, respectively; Z is the latent variable.

In this study, we propose Dynamic Latent Scale GAN (DLSGAN), an unsupervised and architecture-agnostic algorithm that facilitates generator inversion. The core idea is to scale the latent distribution in each dimension based on the encoder's output variance, enabling effective training under the assumption that latent dimensions are independent.



The figure above illustrates the core concept of DLSGAN. When the entropy of the latent distribution is significantly higher than that of the data distribution, the generator tends to ignore or underutilize parts of the latent variables. Although this does not necessarily degrade the quality of generated samples, it hinders the ability to invert the generator.

To address this issue, DLSGAN introduces dimensional scaling of the latent distribution. Specifically, the encoder is trained using a mean squared error loss, assuming that each latent dimension is independent. The scale applied to each latent dimension is proportional to the output variance of the corresponding encoder dimension. As a result, dimensions from which the encoder can extract information more reliably are assigned larger scales, while those with less reliable information are scaled down. The following equations summarize this mechanism:

$$v \approx E(G(Z \circ s))^2$$

$$s = \sqrt{d_z} \cdot \frac{\sqrt{v}}{\|\sqrt{v}\|_2}$$

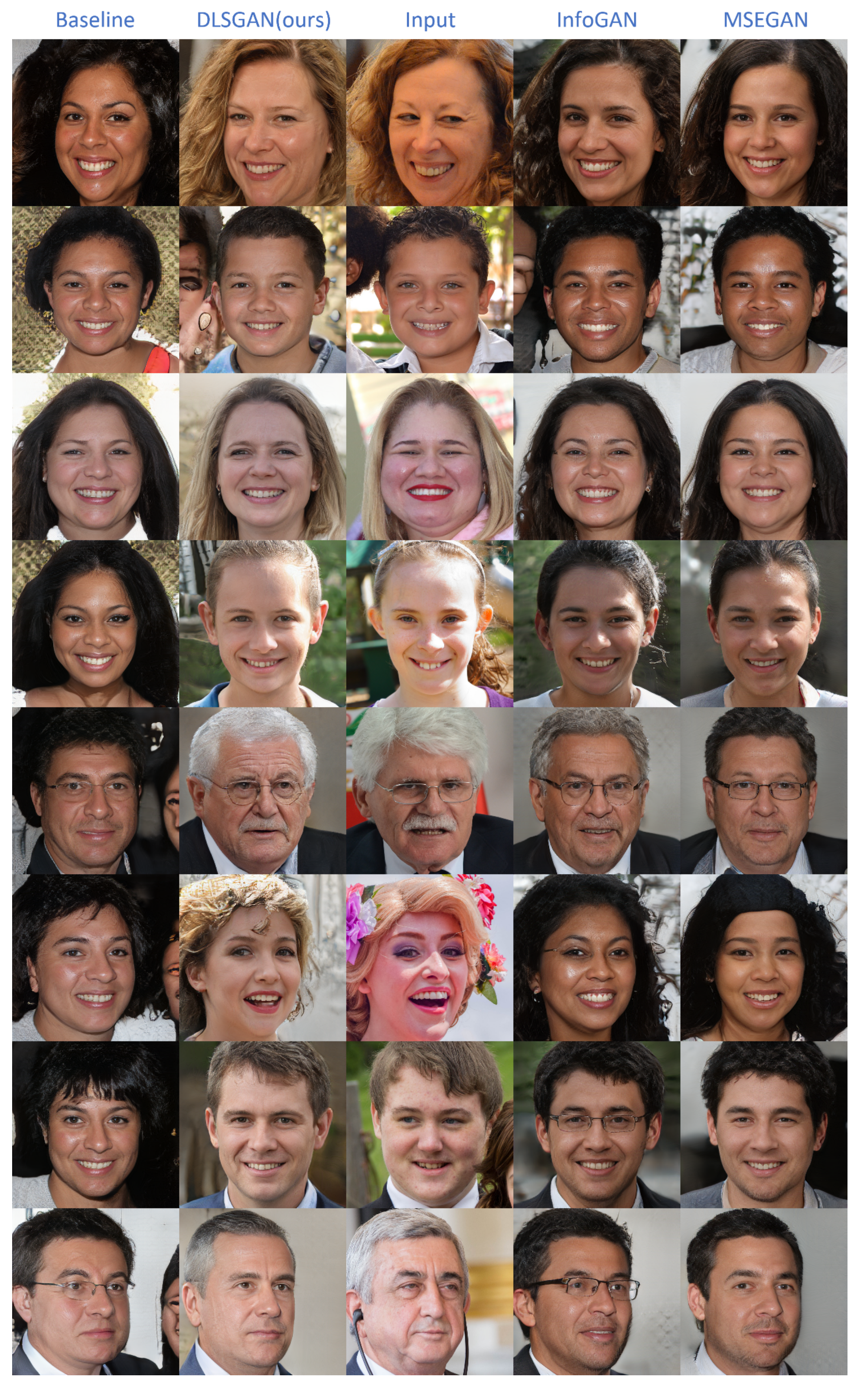
$$L_{enc} = \text{avg}\left(((Z - E(G(Z \circ s))) \circ s)^2\right)$$

$$L_d = L^d_{adv} + \lambda_{enc} L_{enc}$$

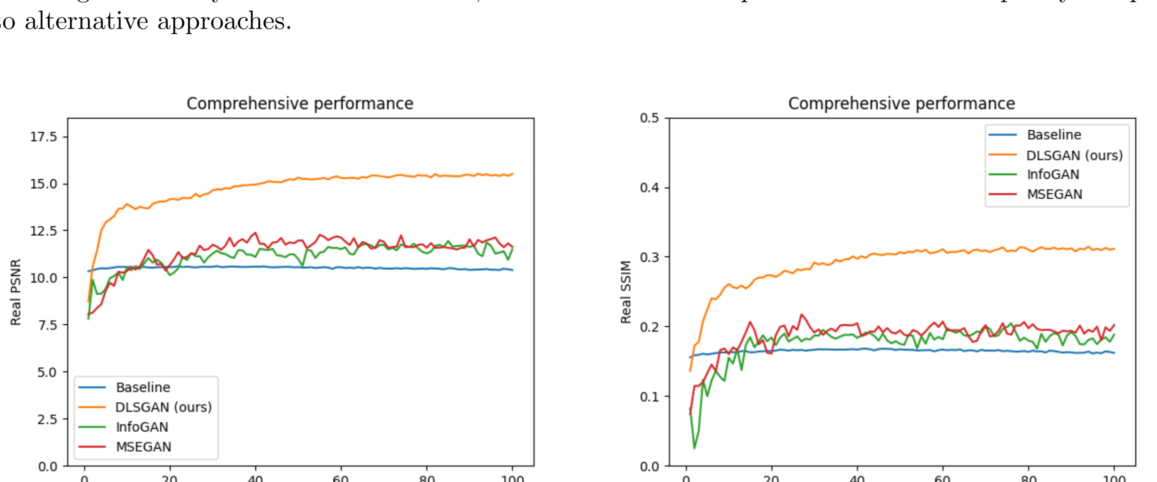
$$L_g = L^g_{adv} + \lambda_{enc} L_{enc}$$

In the above equations, v is the variance vector of the encoder outputs; and s is the scaling vector for the latent distribution. The element-wise operator $\circ$ denotes element-wise product. The scaling vector s is normalized to be proportional to $\sqrt{v}$, thereby adapting the latent space to the encoder's representational capacity.

The figure below compares image reconstruction results obtained by DLSGAN and other baseline methods:



The center column displays the input images, while the second column shows the reconstruction results generated by DLSGAN. As shown, DLSGAN achieves superior reconstruction quality compared to alternative approaches.



This figure presents the reconstruction performance on real images. Among the compared methods, DLSGAN achieves the highest performance in quantitative evaluations, demonstrating its effectiveness in generator inversion.

The full implementation is available on GitHub: `https://github.com/jeongik-jo/DLSGAN`