

# Closed-Form Factorization of Latent Semantics in GANs

Seungmin Lee (profile2697@gmail.com; 2020-20866),

Dept. of Electrical and Computer Engineering, Seoul National University

## 1. Introduction

The latent space of Generative Adversarial Network (GAN) has a rich set of interpretable directions that we can use to edit synthesized images. However, previous methods to find the interpretable directions require human annotations on a collection of synthesized images. In this paper, the authors propose a closed-form algorithm that identifies the semantic directions without using human annotations. More specifically, the proposed method discovers semantics by only using the weights of a pre-trained generator.

## 2. Preliminaries: Manipulating Generator in GAN Latent Space

A generator  $G(\cdot)$  takes a  $d$ -dimensional latent vector  $\mathbf{z}$  from the latent space  $\mathcal{Z} \in \mathbf{R}^d$  and produces an image  $\mathbf{I} = G(\mathbf{z})$ . The authors focus on the first layer of the generator ( $G_1: \mathbf{R}^d \rightarrow \mathbf{R}^m$ ) since it directly acts on the latent space. Like most many GANs have done, the authors assume  $G_1$  is an affine transformation:

$$\mathbf{y} := G_1(\mathbf{z}) = \mathbf{W}\mathbf{z} + \mathbf{b},$$

where  $\mathbf{W} \in \mathbf{R}^{m \times d}$  and  $\mathbf{b} \in \mathbf{R}^m$  denote the weights and bias, respectively.

## 3. Method

## 4. Results

## 5. Personal Note

## References