

<https://doi.org/10.7236/JIIBC.2023.23.3.91>  
JIIBC 2023-3-13

## 이미지 생성을 위한 변동 자동 인코더 분산 제약

# Variational Auto Encoder Distributed Restrictions for Image Generation

김용길\*

Yong-Gil Kim\*

**요약** GAN(Generative Adversarial Networks)이 합성 이미지 생성 및 기타 다양한 응용 프로그램에 현재 사용되고 있지만, 생성 모델을 제어하기가 어렵다. 문제는 생성 모델의 잠재 공간에 있는데, 이미지 생성과 관련하여 입력된 잠재 코드를 받아 특정 텍스트 및 신호에 따라 지정된 대상 속성이 향상되도록 하고 다른 속성은 크게 영향을 받지 않도록 하기 위해서는 상당한 제약이 요구된다. 본 연구에서는 이미지 생성 및 조작과 관련하여 변동 자동 인코더의 잠재 벡터에 관해 특정 제약을 수반한 모델을 제안한다. 제안된 모델에 관해 TensorFlow의 변동 자동 인코더를 통해 실험한 결과 이미지의 생성 및 조작과 관련하여 비교적 우수한 성능을 갖는 것으로 확인된다.

**Abstract** Recent research shows that latent directions can be used to image process towards certain attributes. However, controlling the generation process of generative model is very difficult. Though the latent directions are used to image process for certain attributes, many restrictions are required to enhance the attributes received the latent vectors according to certain text and prompts and other attributes largely unaffected. This study presents a generative model having certain restriction to the latent vectors for image generation and manipulation. The suggested method requires only few minutes per manipulation, and the simulation results through Tensorflow Variational Auto-encoder show the effectiveness of the suggested approach with extensive results.

**Key Words** : Auto-encoder, Variational Auto-encoder, Latent space, Tensorflow

### 1. 서론

자동 인코더는 자율 학습 신경망으로 주로 데이터 압축 학습에 사용되며 본질적으로 항등함수를 학습한다<sup>[1]</sup>. 이미지 도메인에서 자동 인코더는 이미지를 입력으로 받는다. 시스템은 더 적은 비트를 사용하여 이를 재구성한다. 자동 인코더는 주성분 분석과 같은 차원 축소 알고리

즘과 유사한 개념을 가진다. 데이터의 필수 요소는 보존하고 중요하지 않은 부분은 필터링하는 잠재 공간을 만든다. 비선형성이 없는 하나의 층이 있는 자동 인코더는 주성분 분석으로 간주할 수 있다. 자동 인코더 네트워크에는 두 개의 블록이 있는데, 하나는 인코더이고, 다른 하나는 디코더이다. 먼저 인코더는 입력을 잠재 공간으로 알려진 더 적은 수의 비트로 압축하는 네트워크 일부

\*정회원, 조선이공대학교 컴퓨터보안과  
접수일자 2023년 4월 7일, 수정완료 2023년 5월 7일  
게재확정일자 2023년 6월 9일

Received: 7 April, 2023 / Revised: 7 May, 2023 /  
Accepted: 9 June, 2023

\*Corresponding Author: ygkim@cst.ac.kr

Dept. of computer security, Chosun college of science & technology, Korea

이며 때때로 병목 현상을 일으킨다. 디코더는 압축된 표현에서 입력 이미지를 재구성하는 네트워크 일부에 해당한다<sup>[2]</sup>.

VAE(Variational Auto-encoder)는 Variational Bayesian 및 그래픽 모델의 방법에서 영감을 받았으며, 베이저안 추론에 뿌리를 두고 있다<sup>[3]</sup>. 즉, 해당 분포에서 새로운 데이터를 추출하기 위해 데이터의 기본 확률 분포를 모델로 사용한다. VAE는 표준 자동 인코더와 구별되는 근본적으로 고유한 속성을 가지고 있으며, 생성 모델에 매우 유용하게 만드는 것이 바로 이 속성이다. 잠재 공간은 의도적으로 연속적이며 쉬운 무작위 추출 및 보간 방식을 허용한다. VAE는 잠재 벡터에 사전 확률 분포를 적용하는 생성 모델이다. 잠재 벡터는 특정 우선순위를 가진다. 공간의 한 지점에 이미지를 대응시키는 대신에 VAE 인코더는 이미지를 정규분포에 대응시킨다<sup>[4]</sup>. 데이터 집합은 본질적으로 정규분포와 유사한 분포를 따른다고 가정한다. VAE에서 정규분포를 따르도록 잠재 변수를 적용하는 것은 일반적이며 가장 잘 작동한다. VAE가 학습된 후 표준 정규 분포에서 간단히 표본 추출하여 새 이미지를 생성할 수도 있다. 최근 작업에서는 텍스트 캡션에서 이미지를 생성하고 있으며, 여러 단계로 이미지를 칠하는 데 주의를 기울이는 변형 반복 자동 인코더가 사용되고 있다<sup>[5]</sup>.

본 연구와 관련하여 VAE의 주요 목표는 배후 데이터 분포를 학습하여 해당 분포로부터 새로운 데이터 표본을 생성할 수 있도록 하는 것이다. 일반적으로 데이터 분포는 정규분포를 가정하지만, VAE로 재구성한 이미지의 시각 품질은 대체로 우수하다. 정규분포에는 변동 손실의 KL Divergence에 대한 분석적 평가, 효율적인 기울기 계산을 위한 매개변수 기법의 재적용과 같이 VAE 훈련에 필요한 여러 가지 속성이 있다. 특히, 인코더 분포와 배후 분포 간에 KL Divergence를 최소화하도록 VAE를 훈련하는 것이 중요하다. 본 연구에서는 이미지 생성과 관련하여 VAE 제약을 하는 형태와 표준 자동 인코더의 주의 사항을 극복하는 문제, VAE 손실 함수 및 VAE의 Reparametrization 기법을 나타내고, 특정 이미지 데이터 세트에서 TensorFlow의 인코더 및 디코더를 구현한다.

## II. 관련 연구

실제 데이터는 차원이 높으면 중복되는 경우가 많다.

이것은 계산 효율성에 대한 문제를 제기하며, 표현 모델을 방해한다. 비선형 차원 감소와 관련하여 매니폴드 학습은 고차원 데이터를 저차원 잠재 매니폴드에 투영하는 것을 목표로 하는 다양한 학습 방식을 나타낸다<sup>[6]</sup>. 특히 값 분해 및 주성분 분석과 같은 차원 감소에 사용되는 선형 분해 방법의 일반화로 이해할 수 있다. 차원 축소의 기본 가정은 고차원 데이터가 종종 원본 데이터의 내용을 표현하기에 충분한 저차원 임베딩을 갖는다는 것이다. 이미지 표현 문제에서 이 개념을 확장하면 이미지 데이터 세트의 내용을 설명하기에 충분해야 하는 더 낮은 차원 공간이 존재한다. 그러한 공간을 잠재 공간이라 한다. 이것은 데이터 세트의 모든 인스턴스가 근접해 있을 것으로 예상되는 고차원 이미지의 저차원 매니폴드이다<sup>[7]</sup>.

일반적으로 기계 학습 모델을 구축하려고 할 때 주어진 라벨이 있는 일련의 이미지에서 모델을 훈련하고 모델이 저지른 실수를 정량화하는 손실 함수를 최소화한다. 이것을 지도 학습이라 하는데, 각 인스턴스에 대한 라벨이 있는 기계 학습 알고리즘을 제공한다. 그러나 실제 응용 프로그램에서는 종종 이러한 라벨이 없거나 데이터 세트의 고유한 구조를 (예를 들어, 데이터 세트 인스턴스 간의 관계) 추출하려고 한다. 이를 위해서는 비지도/자기 지도 학습이라는 다른 접근 방식이 필요하다. Auto-encoder는 이러한 비지도 학습 방법 가운데 하나이다. 유사한 객체/이미지가 근접해 있는 경향이 있는 잠재 공간에 각 인스턴스를 투영하여 데이터 세트의 고유 구조를 포함한다. Auto-encoder의 일반적인 구조는 그림 1과 같다.

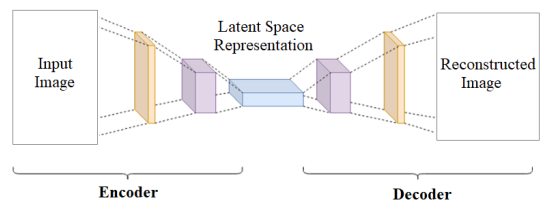


그림 1. 표준 자동 인코더 아키텍처  
Fig. 1. Standard Auto-encoder Architecture

생성 모델은 현대 기계 학습 개발자가 사용할 수 있는 방대한 양의 데이터를 활용한다. 이러한 모델은 데이터 세트에 인코딩된 정보보다 훨씬 적은 매개변수를 가지므로 데이터 분포의 압축된 표현을 학습해야 한다<sup>[8]</sup>. 생성 모델을 구성하는 다양한 방법이 있는데, 그중에서 가장 주목할만한 것은 GAN과 VAE이다. 그러나 이러한 모든

방법은 신경망을 사용하여 정보 병목 현상을 통해 데이터를 스트리밍한다. 여기서 데이터 분포는 계산 흐름의 최소 표현인 잠재 공간으로 인코딩된다. VAE에서 병목 현상은 일반적으로 인코더와 디코더 네트워크 블록 사이에 구성된다. 잠재 공간은 이상적으로는 데이터 분포에서 발견되는 의미 체계 및 공간 정보를 최소한으로 표현한 것이다<sup>[9]</sup>.

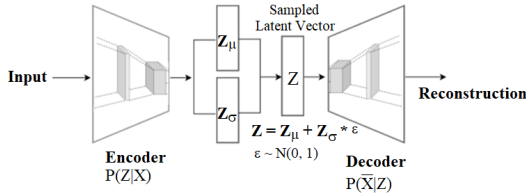


그림 2. 기본 변동 자동 인코더 아키텍처  
 Fig. 2. Basic Variational Auto-encoder Architecture

자동 인코더의 주요 강점은 보이지 않는 인스턴스를 처리해야 하는 데이터 공간의 추상 표현을 추출하는 능력에 있다<sup>[10]</sup>. 이는 잠재 공간을 사용하여 아직 보지 못한 새로운 이미지의 생성 가능성을 열어준다. 그러나 일반적인 자동 인코더 아키텍처는 잠재 공간을 통과하는 데 따른 많은 자유도를 허용하지 않는다. 이는 잠재 벡터 대신 잠재 분포를 학습하여 잠재 공간에서 보간 가능한 VAE에 의해 우회될 수 있다. 심층 생성 모델의 잠재 공간은 종종 새로운 데이터 표본을 생성하기 위해 표본 추출된다. 이 과정에서는 잠재 공간의 각 지점이 데이터 공간의 의미 있는 표본에 해당해야 한다. 따라서 전체 잠재 공간은 구조화되고 의미론적으로 의미가 있어야 한다. 그러나 표준 자동 인코더 네트워크에서는 잠재 공간의 정규화가 없다. 결과적으로 네트워크는 데이터 세트의 예제가 과대 적합되고 공간이 의미가 없는 잠재 표현을 학습하는 경향이 있다. VAE는 학습된 잠재 표현에 대한 정규화를 시행하기 위해 개발되었다. 인코딩된 분포는 정규분포로 선택되어 인코더가 정규분포 평균 및 공분산 행렬을 반환한다<sup>[11]</sup>.

### III. 잠재 공간 제약

자동 인코더에서는 입력 이미지  $X$ 가 주어졌을 때, 매개변수  $\theta$ 를 갖는 인코더  $G$ 는 고정된 잠재 벡터  $Z$ 로 입력을 대응시키도록 학습된다. 잠재 벡터  $Z$ 에서 매개변수  $\phi$ 를 갖는 디코더  $F$ 는 이미지를 재구성하는 방법을 학습

한다. 입력 이미지  $X$ 와 유사한 이미지  $\hat{X}$ 를 재구축하도록 학습된다. 이러한 목적을 달성하기 위해서 재구성 손실, 즉 다음과 같이 주어진 평균 제곱 오차를 최소화한다. 자동 인코더에서 잠재 공간  $Z$ 는 차원이 입력  $X$ 의 크기보다 작아서 병목 현상을 갖기 때문에, Encoder-Decoder 모델은 재구성 손실을 최소화하기 위해 공동으로 훈련된다. 주어진  $Z$ 에서 재구축된 출력  $\hat{X}$ 가 입력과 유사하게 보이게 되는 입력  $X$ 의 인코딩  $Z$ 를 학습한다. 그렇지만, 자동 인코더의 주요 문제는 잠재 공간 구조에 있었기 때문에 새로운 이미지를 생성하는데 능숙하지 않다. 인코딩된 벡터는 서로 다른 데이터 클래스에 해당하는 클러스터로 그룹화되며 클러스터 간에 큰 간격이 있다.

VAE는 잠재 벡터에 사전 확률 분포를 적용하는 생성 모델이다. 입력 이미지는 인코더에 삽입되는데, 그것은 훈련 중에 학습된 분포의 매개변수에 해당하는 두 잠재 변수  $Z_\mu$ 와  $Z_\sigma$ 를 출력한다. 어떠한 분포를 따르도록 강요하지 않는 잠재 공간을 직접 출력하는 대신에 VAE에서는 잠재 벡터  $Z$ 를 추출할 수 있도록 하는 두 개의 잠재 변수  $Z_\mu$ 와  $Z_\sigma$ 가 있다. 추출된 잠재 벡터  $Z$ 은 다변량 정규 분포  $G(Z_\mu, Z_\sigma)$ 에서 표본 추출되는 표본 추출 계층이라 할 수 있다. 여기서  $Z_\mu$ 와  $Z_\sigma$ 는 각각 평균과 분산이다. 추출된 벡터  $Z$ 을 디코더에 전송하여 예측된 이미지  $\hat{X}$ 를 얻는다. 잠재 벡터와 관련하여  $Z_\mu$ 가 0에 가깝고,  $Z_\sigma$ 는 1에 가까운 표준 정규 분포를 따르도록 훈련된다. VAE는 표준 자동 인코더에서와 같이 잠재 공간의 단일 지점 대신 표준 편차에 해당하는 크기로 평균값을 중심으로 특정 영역을 커버한다. 이렇게 하면 디코더가 더 많은 작업을 수행할 수 있다. 해당 영역의 모든 표본은 원래 입력과 매우 유사하다.

잠재 변수  $Z_\mu$ 와  $Z_\sigma$ 를 얻는 방식과 관련하여, 인코더에 convolution 계층이 있고 마지막 convolution 계층 출력이 벡터로 평면화된다고 가정한다. 평면화된 출력은 두 개의 분리된 조밀 계층  $Z_\mu$ 와  $Z_\sigma$ 에 (예를 들어, 각각  $N$ 개의 뉴런을 갖는) 공급된다.  $Z_\mu$ 와  $Z_\sigma$ 의 크기는 모두  $[N, 1]$ 이다. VAE 잠재 변수  $Z_\sigma$ 은 잠재 공간 차원과 상관관계가 없는 것으로 가정하며, 대각선 행렬은 폐쇄형이며 구현하기 쉽다. 따라서 공분산 행렬의 대각선 요소, 즉 분산 또는 표준 편차 요소만 고려한다. 따라서 단일 데이터에 대한 정규분포를 가정하는 기존 VAE 인코더는

잠재 매개변수  $Z_\mu$ 와  $Z_\sigma$ 을 생성한다. 각각의  $Z_\mu$ 와  $Z_\sigma$ 는  $N$ 개의 요소를 갖기 때문에 잠재 매개변수의 전체 개수는  $2N$ 이다.

VAE 목적 함수와 관련하여  $Z_\mu$ 와  $Z_\sigma$ 로부터 추출된 잠재 벡터  $Z$ 을 생성하기 위해서는 다음과 같이 인코더 네트워크를 제한할 수 있다.

① 표준 정규분포  $N(0,1)$

$$\textcircled{2} \text{ 재구성 오차 최소화: } \frac{1}{N} \sum_{i=1}^N (X_i - \hat{X}_i)^2 \quad (1)$$

손실 함수는 재구성 오차와 네트워크 목표를 표현하는데 사용되는 KL(Kullback-Leibler) Divergence 오차로 구성되며, 최종 손실은 두 손실의 가중 합이다. VAE의 전체 손실은 다음과 같다.

$$L(\phi, \theta, x) = \frac{1}{N} \sum_{i=1}^N (X_i - \hat{X}_i)^2 + KL[G(Z_\mu, Z_\sigma), N(0, 1)] \quad (2)$$

VAE의 재구성 손실은 Auto-encoder에서 사용한 손실인 MSE와 유사하다. 재구성 손실은 입력 이미지가 출력에서 재구성되도록 보장하며 그렇게 함으로써 손실은 본질적으로 VAE의 인코딩 및 디코딩을 효율적이고 의미 있게 만든다. VAE의 목표는 분포를 학습하는 것뿐만 아니라 훈련 데이터와 유사한 사실적인 이미지를 생성하는 것이다. 따라서 재구성 오차 함수가 필요하다. 재구성 손실은 다음과 같이 나타낼 수 있다.

$$L_{MSE}(\theta, \phi) = \frac{1}{N} \sum_{i=1}^N (x_i - f_\theta(g_\phi(x_i)))^2 \quad (3)$$

여기서  $\theta$ 와  $\phi$ 은 각각 인코더와 디코더의 매개변수이다.  $N$ 은 손실이 계산되는 데이터 세트 또는 미니 배치의 이미지 수이다.

VAE와 관련하여 본 연구에서는 이미지 인코딩이 고유하면서도 가능하면, 서로 가깝기를 원하므로 잠재 공간에서 더 완전한 보간을 통해 실제 표본과 유사하게 보이는 표본을 생성할 필요가 있다. 이와 관련하여 KL Divergence는 하나의 확률 분포가 두 번째 참조 확률 분포와 어떻게 다른지에 대한 척도이다. VAE에서 주요 목표는 배후 데이터 분포를 학습하여 해당 분포로부터 새로운 데이터 표본을 생성할 수 있도록 하는 것이다. 분포를 추정하려면 데이터가 Gaussian, Bernoulli 등과 같은 특정 분포에서 온다고 가정해야 한다<sup>[12]</sup>. 일반적으로 VAE에서 데이터 분포는 정규분포를 가정한다. 인코더 분포  $P_\phi(Z | X)$ 와  $P(Z)$  간에 KL Divergence

를 최소화하도록 VAE를 훈련한다.

$P(Z)$ 은 표준 정규분포로 가정된다. 인코더가 표준 정규분포로부터 멀어지는 인코딩  $Z$ 을 출력하는 경우 KL-divergence 손실은 더 많은 불이익을 부여한다. KL-divergence는 인코딩을 유지하는 정규화 역할을 한다. 즉, KL-divergence는 확률 분포 매개변수  $\mu$ 와  $\sigma$ 을 최적화하여, 표준 정규 분포에 가깝게 한다. 매개변수  $\mu$ 와  $\sigma$ 을 갖는 정규분포가 표준 정규 분포와 어느 정도 차이가 있는지 측정한다. KL-divergence를 계산하는 동안 매개변수  $\sigma^2$ 을 2보다 작도록 제약하면 잠재 공간이 매끄럽게 표현되고 수행 시간이 훨씬 빠르다. 이러한 제약은 다음과 같이 닫힌 형태의 KL-divergence를 갖는다.

$$L_{KL}[G(Z_\mu, Z_\sigma) | N(0, 1)] = -0.5 \sum_{i=1}^N (1 - Z_{\mu_i}^2 - Z_{\sigma_i}^2 + \log(Z_{\sigma_i}^2)) \quad (4)$$

특히,  $\sigma^2 < 2$ 의 제약 조건에서 로그 식의 다항식 근사를 사용하면 다음과 같은 하한을 얻을 수 있다.

$$L_{KL}[G(Z_\mu, Z_\sigma) | N(0, 1)] \geq 0.5 \sum_{i=1}^N (Z_{\mu_i}^2 + 0.5(Z_{\sigma_i}^2 - 1)^2) \quad (5)$$

여기서  $Z_\mu$ 와  $Z_\sigma^2$ 은 인코더 잠재 공간의 평균 및 분산 벡터이고, 그 합계는 잠재 공간의 모든 차원에 적용된다.

그림 3은 두 가지 계산 그래프로 상단은 원래 형태이고, 하단은 매개변수로 표현된 형태이다.  $Z_\mu$ 와  $Z_\sigma^2$ 은 네트워크가 학습하려는 매개변수이다. 사각형으로 표현된 노드는 입력 및 가중치 관련 정적 노드이고, 원으로 표현된 노드는 확률적 노드이다. 훈련 중에 이미지  $X$ 은 두 잠재 변수  $Z_\mu$ 와  $Z_\sigma^2$ 에 대응되고, 이미지  $\hat{X}$ 의 출력을 위해 디코더에 전송하여 두 잠재 변수들로부터 벡터  $Z$ 을 추출한다. 그렇지만, 이러한 확률적 추출 작업은 확률적 특성으로 기울기가 추출 계층을 통해 역전파 될 수 없어서  $Z$ 가 병목 현상을 일으키는 확률적 노드이다. 즉,  $Z_\mu$ 와  $Z_\sigma^2$ 은 학습이 가능하지 않다. 역전파는 기울기를 반복적으로 전달하고 체인 규칙을 적용하기 위해 노드가 결정적이어야 한다. 이러한 문제를 해결하기 위해 확률적 노드를 정적 노드로 변환하는 재 매개변수 기법이 필요하다<sup>[13][14]</sup>.  $Z_\mu$ 와  $Z_\sigma^2$  벡터가 전체 시스템의 확률적 성질을 유지하면서 네트워크의 학습 가능한 매개변수로

남을 수 있다. 구체적으로 추출 벡터  $Z$  대신에 표준 정규 분포에서 추출한 임의의 상수들의 합을 근사치로 사용한다. 새로운 표본 추출 작업은 다음과 같다. 여기서  $\otimes$  은 요소별 곱셈을 나타낸다.

$$Z = Z_{\mu} + Z_{\sigma}^2 \otimes \epsilon, \epsilon \sim N(0, 1) \quad (6)$$

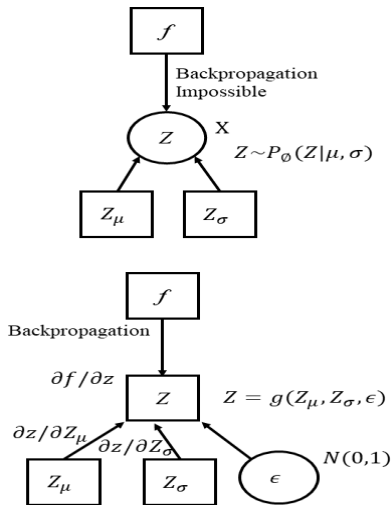


그림 3. 재 매개변수화  
 Fig. 3. Reparameterization Technique

#### IV. 시뮬레이션

본 실험에서 사용되는 데이터 세트는 구글의 cartoonset에서 가져온 것으로 데이터 세트는 10개의 작품 범주, 4개의 색상 범주 및 4개의 비율 범주로 다양하며,  $10^{13}$ 개 조합이 가능한 2D cartoon 이미지들의 무작위 집합이다. 무작위로 선택된 10k 및 100k 세트의 cartoon과 라벨이 지정된 속성을 제공한다. 여기서는 제안된 VAE 훈련을 위해 10k 이미지 세트를 사용한다. 데이터 세트 로드에는 지정된 디렉터리에서 데이터를 가져오는 tf\_keras 전처리 데이터 세트 모듈을 사용할 수 있다. 자율 학습에 해당하므로 필요한 인수로 image\_size [256, 256, 3], batch\_size = 128 및 label\_mode=None을 전달한다. 또한, Lambda 함수를 사용하여 모든 입력 이미지를 정규화하고 모델 훈련에 사용할 이미지를 얻는다. Lambda 함수에서는 정의된 전처리 계층을 전달한다.

Tensorflow를 사용한 인코더 네트워크는 크기 [-, 256, 256, 3]의 입력을 받아 5개의 Conv 블록을 진행한다. 각 블록은 Conv2D, BatchNorm 및 LeakyReLU

활성화 함수를 가지며, 각 블록에서 이미지는 down-sampling 된다. 최종 블록에서 잠재 변수들의 벡터가 정의되고 모델 출력은 sampling 네트워크에 공급된다. 여기에서 크기  $256 \times 256$ 의 컬러 이미지가 크기 200의 잠재 변수들의 벡터로 인코딩된다. 디코더 네트워크의 초기 블록은 4,096개 뉴런을 갖는 Dense 계층이다. 첫 4개의 Convolution 블록은 Conv2DTranspose, BatchNorm 및 LeakyReLU 활성화 함수로 구성된다. 마지막 블록은 이미지 정규화를 위해 sigmoid 함수를 갖는 Conv2DTranspose로 구성된다. 각 블록에서 이미지는 upsampling 된다. 네트워크 출력은 크기 [-, 256, 256, 3]의 tensor이다. 그림 4는 제안된 VAE 모델에 의한 cartoon 이미지의 재구성 결과이다. VAE는 cartoon 이미지의 재구성에 있어서 비교적 좋은 결과를 제공함을 확인할 수 있다.

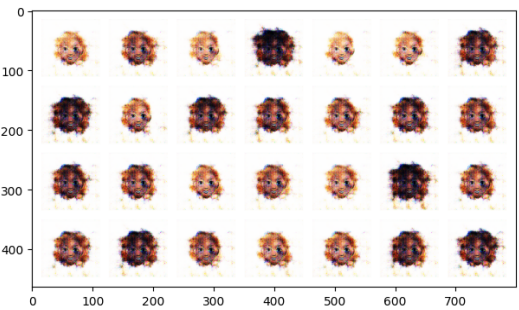


그림 4. Cartoon 이미지 재구성  
 Fig. 4. Cartoon Image Reconstruction

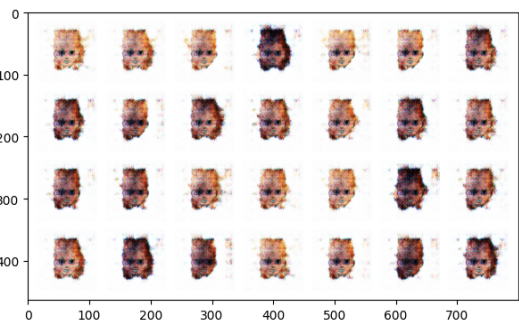


그림 5. 잠재 공간 추출에 의한 재구성  
 Fig. 5. Reconstruction through Latent Space Sampling

균일하게 추출된 잠재 벡터로 Cartoon 이미지를 재구성하고자 할 때 단순히 균등 분포 함수에 하한과 상한을 전달하여 200D의 잠재 공간에서 점을 균일하게 표본 추출할 수 없다. 그 이유는 모든 200D에 대해 이 작업을

수행함으로 상수의 값이 필요하다. 모든 5K 이미지에서 200D의 최솟값과 최댓값을 취하고 값이 0과 1 사이에 있는 크기  $10 \times 200$ 의 균등 행렬을 사용한다. 그런 다음에 잠재 공간의 최솟값과 최댓값 사이의 차이를 취하여 값의 크기를 조정한다. 마지막으로 크기 조정된 출력을 디코더에 전달하고 이미지를 생성한다. 그림 5는 재구성된 Cartoon 이미지이다. 제약을 사용하는 VAE의 디코더는 Cartoon 집합에 있는 것과 유사한 이미지를 아주 빠르게 생성한다. 생성된 이미지는 데이터 세트에 없는 것도 있지만 정규분포를 따른다. 데이터 세트에 없는 이미지의 예로 다른 머리 스타일과 머리 색깔을 가진 디코더에 의해 생성된 이미지가 있다. 또한 같은 Cartoon 이미지일지라도 데이터 집합 내에서 같은 머리 스타일 또는 머리 색깔을 갖지 않을 수 있지만 생성된 이미지가 다소간 현실적임을 확인할 수 있다.

## V. 결 론

GAN 모델의 문제 중 하나는 얽힌 잠재 표현이다. 얽힌 표현을 사용하면 데이터 분포가 반드시 입력 벡터  $z$ 를 추출하려는 정규분포를 따르지 않는다. 즉, 데이터 분포는 눈과 얼굴의 비율이 비현실적이 되는 영역을 나타내는 모서리를 갖게 된다. GAN의 생성 모델에 관해 잠재 공간 제약을 사용하면 입력 벡터를 중간 잠재 공간으로 인코딩하는 대응 네트워크를 추가하여 이를 더욱 개선할 수 있다. 그렇지만, 다양한 세부 수준을 제어하는 데 별도의 값이 사용되고 많은 시간이 소요된다. 본 연구에서 제안된 잠재 공간 제약 방식은 KL-divergence와 관련된 가중치를 줄이기 때문에, 얽힌 잠재 표현을 억제하여 현실적인 이미지를 생성하고 조작할 수 있도록 한다. 또한 훈련 시간이 더 빨라지고 훈련이 훨씬 더 안정적이다.

## References

- [1] H G. E. Hinton and R. R. Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks. *SCIENCE*, vol. 313, pp. 504-507, 2006. DOI: <https://doi.org/10.1126/science.1127647>
- [2] AJ. Yoo, H. Eom and Y. S. Choi. Image-To-Image Translation Using a Cross-Domain Auto-Encoder and Decoder. *Appl. Sci.* vol. 9, no. 22, 4780, 2019. DOI: <https://doi.org/10.3390/app9224780>
- [3] J.P. Cristovao, H. Nakada, Y. Tanimura, and H. Asoh. Generating In-Between Images Through Learned Latent Space Representation Using Variational Autoencoders. *IEEE Access*, vol. 8, pp. 149456-149467, 2020.. DOI: <https://doi.org/10.1109/ACCESS.2020.3016313>
- [4] AD. P. Kingma and M. Welling. Auto-Encoding Variational Bayes. 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014.. DOI: <https://doi.org/10.48550/arXiv.1312.6114>
- [5] J.C. Louizos and M. Welling. Multiplicative normalizing flows for variational Bayesian neural networks. *International Conference on Machine Learning*. 2218-2227, 2017..
- [6] JB. Kim, S. Shin, and H. Jung. Variational Autoencoder-Based Multiple Image Captioning Using a Caption Attention Map. *Appl. Sci.* Vol.9, No.13, 2699, 2019.. DOI: <https://doi.org/10.3390/app9132699>
- [7] SN. D. Lawrence, A unifying probabilistic perspective for spectral dimensionality reduction: insights and new models. *Journal of Machine Learning Research*. 13, 1609-1638, 2012.
- [8] Y.G Kim, K.I Moon. Hair Classification and Region Segmentation by Location Distribution and Graph Cutting. *The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)*. Vol.22, No.3, pp1-8 Jun.30.2022 DOI: <https://doi.org/10.7236/JIIBC.2022.22.3.1>
- [9] MJ. A. Lee and M. Verleysen, *Nonlinear Dimensionality Reduction*. Springer, 2007, ISBN 978-0-387-39350-6.. Y.G Kim, K.I Moon. Object Detection Based on Hellinger Distance IoU and Objectron Application. *The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)*. Vol.22, No2, pp.63-70 Apr.30. 2022 DOI: <https://doi.org/10.7236/JIIBC.2022.22.2.63>
- [10] AB. U. Dideriksen, K. Derosche, Z. H. Tan, iVAE-GAN: Identifiable VAE-GAN Models for Latent Representation Learning. *IEEE Access*, vol. 10, pp. 48405-48418, 03,May,2022. DOI: <https://doi.org/10.1109/ACCESS.2022.3172333>
- [11] J. Duda. Gaussian AutoEncoder arXiv:1811.04751v4 [cs.LG] 14 Jan 2019. DOI: <https://doi.org/10.48550/arXiv.1811.04751>
- [12] Y.G Kim, K.I Moon. Image Restoration Based on Inverse Order and Power Spectrum Density, *The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)*. Vol.16, No.2, pp.113~122, 2016. DOI: <https://doi.org/10.7236/JIIBC.2016.16.2.113>
- [13] Y.G Kim, Image Reconstruction Using Poisson Model Screened from Image Gradient. *The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)*. Vol.18, No 2, pp.117-123, 2018. DOI: <https://doi.org/10.7236/JIIBC.2018.18.2.117>

## 저 자 소 개

김 용 길(정회원)



- 1990년 : 호남대학교 전산통계학과 졸업(이학사)
  - 1992년 : 광주대학교 대학원 컴퓨터학과 졸업(공학석사)
  - 2021년 : 호남대학교 대학원 컴퓨터공학과 졸업(공학박사)
  - 2014년 ~ 현재 : 조선이공대학교 컴퓨터보안과 부교수
- 관심분야 : 지능시스템, 네트워크보안, 정보보호