

논문 2024-61-2-12

오토인코더와 다중 레이블 분류기의 사전학습을 통한 GAN 학습 안정성 향상 및 생성 영상 제어

(Improvement of Training Stability and Controllability for GAN via
Reusing Pre-trained Autoencoder and Multi-label Classifier)

김 지 수*, 지 준**, 오 희 석**

(Jisoo Kim, Jun Ji, and Heeseok Oh[©])

요 약

적대적 생성 네트워크(Generative Adversarial Network)는 모든 픽셀에 대해 생성망과 판별망이 적대적으로 동작해야 하는 특성상 고해상도 이미지를 생성하기 어렵다는 문제와, 적대적 훈련 과정에 의해 데이터셋 분포 학습이 불안정하다는 문제를 동시에 가지고 있다. 데이터셋 분포 학습 과정에서 영상 제어를 통한 결과물을 생성하기 위해 다양한 기존 연구들이 생성망의 입력인 잠재벡터에 조건을 결합하는 방식으로 영상을 제어한다. 그러나 불안정한 GAN의 특성상 결합한 조건에 부합하는 고품질의 이미지를 생성하는데 한계가 있다. 비전 분야에서 기 학습된 특징 매핑 신경망을 도입해 하위 작업에 전이 학습하여 사용하는 접근은 잘 알려진 학습의 효과를 높여주는 방법이다. 본 논문에서는 전이 학습을 응용하여 생성자의 표현력을 향상시키고 동시에 학습의 안정성을 도모하고자 하였으며, 생성망의 입력인 잠재벡터에 조건을 결합한 조건 벡터에 부합하는 고해상도 이미지를 생성하고자 하였다. 영상을 저차원의 은닉 공간으로의 임베딩 후 다시 복원하는 일련의 과정을 인코더-디코더를 통해 사전 훈련하였고, 사전 훈련 과정에서 저차원의 은닉 공간으로 임베딩된 벡터를 분류기를 통해 예측한 레이블과 입력 레이블의 오차를 줄여나가는 방향으로 학습시켰다. 고차원의 영상 공간으로 복원하는 네트워크를 생성망의 초기 가중치로 설정하였으며, 저차원의 은닉 공간으로 임베딩하는 네트워크를 생성망 출력을 입력으로 받는 네트워크에 고정 가중치로 설정하여 최종적으로 향상된 학습 안정성과 컨디셔닝 제어를 가능하게 하는 모델인 PSC-GAN (Pre-weighted Stabilized Conditioning Generative Adversarial Network)을 제안함으로써, 동일 조건에서 기존 GAN 학습 시 비교군 기준 대비 정량적으로 10% 가까운 성능 향상이 이루어짐을 확인하였다.

Abstract

Generative Adversarial Networks (GANs) are generative models that face the challenge of sampling high-resolution images due to the generator's hostility towards the discriminator in a pixel-wise manner, as well as the instability of the dataset distribution estimation arising from adversarial training. To generate images that satisfy certain conditions, previous studies have embedded conditions into latent vectors, which are then fed into the generator. However, due to GANs' training instability, there is a limit to generating images having plausible quality with controlled conditions. One solution is to utilize a pre-trained feature mapper through transfer learning to increase the efficiency of learning. In this paper, we apply transfer learning to improve the fidelity of the generated image and promote training stability, ultimately generating high-resolution images that meet the combined condition and latent representation. We pre-trained the encoder-decoder for embedding an image into a low-dimensional latent space and reconstructing it into a spatial domain, learning the latent space by reducing the errors between the predictions and input labels through the classifier. The weights of the generator were initialized by the pre-trained decoder that reconstructs the image context, while an encoder that embeds an image into latent space was set as a frozen classifier that receives the generator's output. As a result, we verified that retraining the PSC-GAN (Pre-weighted Stabilized Conditioning Generative Adversarial Network) with improved training stability and controllability under the same conditions led to about 10% quantitative improvement compared to previous studies.

Keywords : Generative adversarial network, Training stability, Controllable generation, Transfer learning

*비회원, **정회원, 한성대학교 AI응용학과(Department of Applied AI, Hansung University)

© Corresponding Author(E-mail : ohhs@hansung.ac.kr)

※ 본 연구는 한성대학교 교내학술연구비 지원과제임.

Received ; June 28, 2023

Revised ; October 23, 2023

Accepted ; November 1, 2023

I. 서론

적대적 생성 네트워크(Generative Adversarial Network; GAN)^[1]은 생성망(generator)과 판별망(discriminator)으로 구성된 두 개의 신경망이 적대적으로 동작하며 학습된다. 생성망은 임의의 잠재벡터로부터 새로운 영상을 생성하고 판별망은 생성 영상의 진위 여부를 판별하도록 게임이론에서의 최선 대응 전략을 학습하며, 종단에는 두 신경망이 내쉬 평형에 도달하는 과정에서 주어진 데이터셋에 대한 우도(likelihood) 추정이 최대화됨으로써 묵시적(implicit)으로 신경망으로써 학습된 데이터 분포 추정 모델에서의 새로운 영상 샘플링을 가능하게 한다. 하지만 모든 픽셀 차원에 대해 생성망과 판별망이 적대적으로 동작해야 하는 특성 상 고해상도 이미지를 생성하기 어렵다는 문제와, 더불어 적대적 훈련에 따른 데이터셋 분포 학습 과정 자체가 불안정하다는 문제를 동시에 가지고 있다^[2, 3]. 본 논문에서는 이를 해결하고, 나아가 조건 결합에 따른 생성 영상의 조절 또한 수행할 수 있도록 오토인코더와 잠재 공간에서의 분류기 사전학습을 통한 강인한 GAN 학습 전략을 제안한다. 지도학습 기반 사전학습된 신경망의 전이 학습은 보편적으로 널리 이용되는 방법이다^[4, 5]. 본 논문에서는 단순히 하위 작업 적용을 위한 일반화된 표현학습을 넘어, 오토인코더의 인코더, 생성 모델의 생성자, 잠재벡터에 대한 추가 네트워크인 분류기 신경망을 통해 고차원 데이터셋에 대한 GAN의 학습 안정성도 및 조건 입력에 따른 결과 영상 제어까지 고려함으로써 생성망에 특화된 형태로 사전학습을 구성하였음에 차별성을 갖는다. 제안하는 기술은 생성 모델의 생성자를 오토인코더 형태로 사전 학습함으로써 고해상도 이미지 생성 및 안정적인 학습이 가능하도록 구성하였으며, 신경망 구조 변경을 통해 사전학습 과정 중 다중 스케일 특징이 인코더에서 디코더로 보다 효과적으로 전달될 수 있도록 구성하였다. 또한 생성 모델의 생성 결과를 인코더를 통과하는 신경망으로 제어 가능하게 설계함으로써 원하는 조건의 이미지 생성이 가능하도록 하였다. 제안하는 PSC-GAN(Pre-weighted Stabilized Conditioning Generative Adversarial Network)은 초기 가중치와 고정 가중치를 동시에 사용함으로써 기존 생성 모델보다 강한 학습 안정성과 결과에 대한 컨디셔닝 제어를 가능하게 하는데 강점을 지닌다.

본 논문에서는 실험을 통해 제안하는 기술이 결과적으로 조건 특징들에 대해서 효과적으로 학습할 수 있어

조건에 부합되는 우수한 품질의 생성이 수행된 결과를 정성적으로 확인할 수 있었고, 이를 이미지 생성 및 품질 정량 지표로 검증한 결과, 기존 기존 기술 대비 10% 성능 향상을 달성함을 보였다.

II. 관련 연구

1. 조건 제어 이미지 생성 모델

비조건적 적대적 생성 모델에서는 생성되는 이미지를 원하는 형태로 제어하기가 난해하며, 일반적으로는 사전 분포에서 랜덤하게 추출된 은닉 벡터 z 를 입력으로 암시적이게 생성망이 해당 벡터에 대응하는 결과 이미지를 출력한다. Conditional GAN(cGAN)^[6]을 시작으로 명시적인 조건을 생성 과정에 주입함으로써 사용자가 원하는 형태의 결과를 생성을 유도하도록 조절 가능한 기법들이 다양하게 소개되어왔으며, 여기서의 조건은 클래스 레이블을 활용한 분류 정규화(regularization)을 주로 활용한다. Dual-Discriminator cGAN^[7]에서는 조건 정보를 추가함과 동시에 두 개의 판별자를 사용해보다 강건하게 생성 이미지를 조절 가능한 학습 방법을 제안하였다. 생성망은 합성된 이미지를 생성함에 있어 합성 전 이미지 두 장을 조건으로 활용하며 각각의 판별자에게 두 이미지의 진위 여부를 가릴 수 있도록 손실함수를 구성함으로써 모델이 학습된다. Y. Choi 등은 멀티도메인 변환 관점에서 둘 이상의 도메인 간 스타일 변환을 하나의 생성망으로 가능하도록 도메인 분류기를 활용한 StarGAN을 제안하였다^[8]. 즉, 이미지와 타겟 도메인 정보를 조건으로 입력하여 원하는 도메인으로의 변환을 수행하며, 이 과정에서 판별망은 올바른 타겟 도메인으로의 변환이 이루어졌는지에 대한 분류 또한 수행한다. 이 타겟 도메인 정보를 추가할 때, 마스크벡터를 활용함으로써 어떤 데이터셋인지를 명시해 주고 결론적으로는 서로 다른 데이터셋을 구분할 수 있게 된다. StarGAN v2^[9]는 각 도메인에 대한 동일한 변형만을 수행하는 것이 아닌, 스타일 매핑과 인코딩을 통해 도메인에 특화된 스타일 코드를 추출하여 이를 조건으로 반영함으로써 이미지 생성 뿐만이 아닌 타겟 도메인으로의 스타일 전이가 가능하도록 구성하였다. 이후로는 이미지의 구조 정보를 유지함과 동시에 품질(fidelity)의 향상을 위한 단계별 컨디셔닝 기법에 관한 논의가 주를 이루었다. Style-based GAN^[10]에서는 사전 분포에서의 은닉 벡터 z 를 곧바로 입력으로 활용하는 것이 아닌, 데이터셋의 얽혀있는(entangle) 스타일 은닉 공간

을 효과적으로 매핑 가능한 신경망을 별도로 구성함으로써 W 공간으로 변환하고, 여기에서 추출한 스타일 벡터를 Adaptive Instance Normalization(AdaIN)을 통해 원하는 단계에 컨디셔닝을 하는 방안을 제안하였다.

2. 사전학습과 신경망 가중치 재사용

생성 모델은 이미지를 생성하는 학습 과정에서 판별자의 기울기 소실이나 수렴이 잘되지 않는 불안정성 문제가 발생한다^[9]. 생성 모델 분야에서는 GAN의 이러한 불안정성 문제를 해결하기 위해서 학습 과정에 있어 사전학습된 가중치를 재사용함으로써 학습의 안정성을 도모함과 동시에 생성 결과물의 품질을 향상시키는 다양한 연구들을 진행하였다. Larsen 등은 Variational Autoencoder(VAE)와 GAN을 결합한 VAEGAN을 제안하였다^[11]. VAE가 판별자와 대립하며 학습되며, 시각적 품질 관점에서 향상된 결과물의 생성이 가능해졌다. 판별자가 VAE의 추상적인 재구성 오류로부터 학습한 영상 속성을 활용하여 판별한다. 이 때, VAE와 GAN은 동시에 학습되며 이는 본 논문에서 네트워크를 학습하기 위해 결합하여 만든 손실함수와 관련한 모든 네트워크 매개 변수들을 업데이트하지 않아도 되기 때문에 가능하다. 나아가 판별자가 0으로 수렴하는 모드 붕괴(mode collapse) 문제를 해결할 수 있게 되었다. 이를 통해 학습한 생성 모델이 보다 좋은 영상을 생성함을 입증하였다. Swapping Autoencoder^[12]는 두 개의 오토인코더가 인코딩하는 저차원의 벡터를 일부 교환하는 형태이며, 텍스트 교환을 통한 맥락 특성과 스타일의 분리가 가능하도록 구성되었다. 이미지 조작을 수행하기 위해서 해당 논문은 오토인코더 구조를 차용하였으며 영상 생성의 과정에서 인코더-디코더의 학습된 가중치를 교환하여 향상된 조작의 결과가 나오는 것을 확인하였다. Unbalanced GANs^[13]은 적대적 생성망의 생성자를 사전학습된 VAE를 사용하는 방식을 제안하였다. 이를 통해 생성자의 안정적 학습과 판별자의 빠른 수렴을 억제함을 증명하였으며, 모드 붕괴 문제에 대응할 수 있음을 보였다. 나아가 GAN가 비교하였을 때, Unbalanced GANs이 학습의 빠른 수렴과 향상된 영상의 품질을 생성하는 것을 확인하였다. Transferring GANs^[14]은 생성 모델에 도메인 적용을 하기 위하여 사전학습 네트워크를 사용하여 타겟 도메인을 적용하여 이미지를 생성한다. 하지만 특정 도메인에만 유리하게 이미지 생성이 되는 한계점이 존재한다.

III. PSC-GAN

그림 1에서는 본 연구에서 제안하는 PSC-GAN의 두 단계 생성망 학습 프레임워크를 전체적으로 도식화하였다. 첫 번째 학습 단계에서는 오토인코더 구조를 갖는 신경망 E 와 G 가 고해상도 영상 데이터셋의 매니폴드 구성을 목표로 각각 인코딩 및 디코딩 능력을 학습함과 동시에, 신경망 C 는 E 를 통해 투영된 잠재벡터 z 를 입력으로 받아 N_c 개의 클래스로 영상을 분류할 수 있도록 학습된다. 두 번째 학습 단계에서는 디코더 G 를 다시 한 번 판별망 D 와 적대적 생성망 학습함에 따라 모드붕괴를 방지함과 동시에 안정적인 적대적 학습이 가능하도록 한다. 더불어 G 를 통해 생성된 이미지는 사전학습된 E 를 통해 잠재벡터로 매핑되고 C 를 통해 N_c 개 클래스로 분류된다. 사전학습된 E 와 C 를 재사용함에 따라 영상 생성에 있어 다중 레이블을 조건으로 부여할 시 입력 잠재벡터 z 로부터의 결과 영상이 무작위로 생성되는 것이 아닌, 컨디셔닝에 따른 타겟 도메인 영상이 생성되도록 제어한다.

1. 특징분리를 포함하는 오토인코더 구조 사전 학습

그림 2는 일반적인 GAN 학습 시의 불안정성에 대한 문제점을 예시로써 보여준다. 학습에 활용한 256×256 크기 영상 데이터셋에 대해 Deep Convolutional GAN(DCGAN)^[15]의 경우 고차원 영상 분포 학습에서는 생성망-판별망이 내쉬 평형에 도달하기 어려움에 따라 사용자의 면밀한 하이퍼파라미터 조절이 요구되며^[16], 그림 2의 경우는 동일한 세팅에서 38번의 학습 시도 끝에 영상 생성에 성공한 일례를 나타내었다. 또한, 해당 경우에도 모델 수렴이 빠르게 이루어지지 않으며 그림 2에서 볼 수 있듯이 많은 수의 학습 반복(10 에폭 이후) 생성 결과가 데이터셋 분포에 따른 영상 형태를 나타냄을 확인할 수 있다. 이처럼 GAN의 학습은 생성망-판별망 학습의 균형과 관련한 신경망의 구조, 최적화 방법 및 하이퍼파라미터 세팅의 표준 기준이 없고 그 과정이 필연적으로 불안정함에 따라, 본 논문에서는 오토인코더를 통해 잠재공간과 고해상도 영상 도메인 간 매핑 관계를 사전학습하고 신경망의 가중치를 재사용함으로써 학습 안정성을 향상시켰다.

디코더 G 는 잠재벡터 z 로부터 영상을 생성하는 영상으로의 복원력을 사전학습을 통해 학습하게 된다. 이는 인코더 E 의 입력 영상 x 와 G 의 복원 영상 $\hat{x}$ 간 재복원 손실함수 L_{rec} 를 통하여 N_b 개 미니배치(b : 배치 사

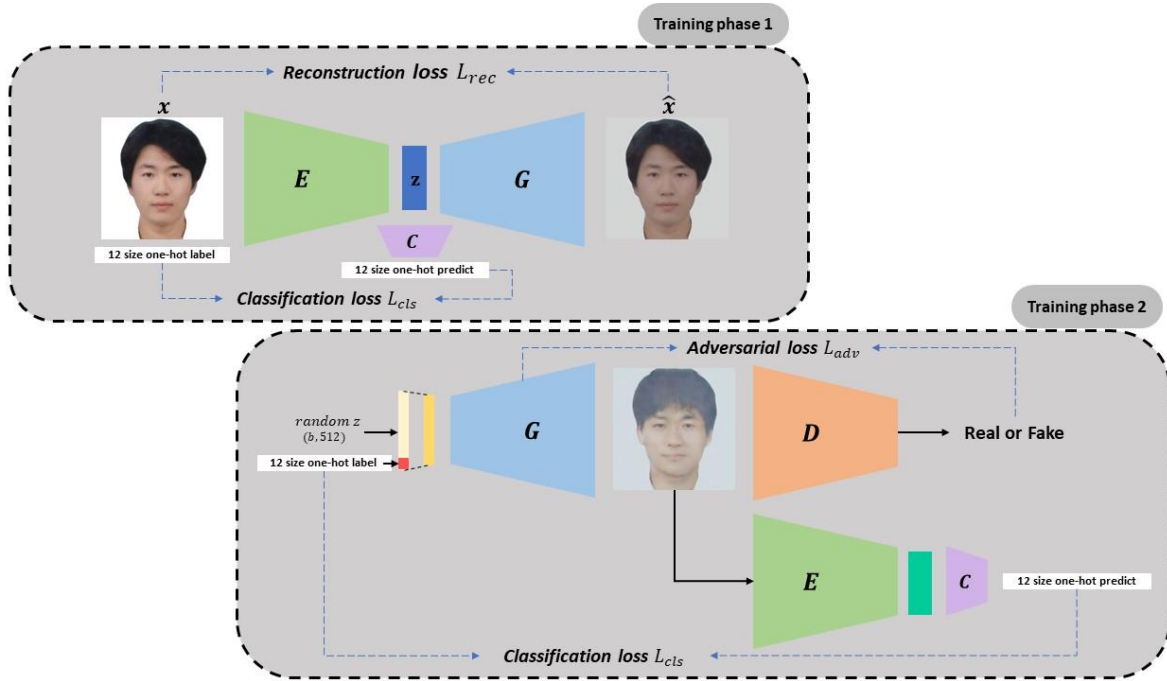


그림 1. PSC-GAN의 프레임워크: 두 단계 학습을 통한 적대적 생성망의 학습 안정성 향상과 조건 제어 영상 생성 기법
Fig. 1. Overall Framework of PSC-GAN: Enhancing the stability of training and conditioning in image generation techniques of GAN through a two-stage learning approach.



그림 2. 256×256 크기 영상에 대한 GAN 학습: 0, 5, 10, 30 에폭 생성 결과
Fig. 2. GAN training for 256×256 size images: results of 0, 5, 10 and 30 epochs.

이즈) 영상들에 대해 두 신경망이 자가학습하는 과정에서 G 는 저차원 잠재벡터에서 고차원 영상을 생성하는 도메인 변환함수로서의 가중치들을 학습하게 된다. 재복원 손실함수 L_{rec} 는 수식 (1)와 같이 사용하였다.

$$L_{rec} = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2. \quad (1)$$

즉, GAN 학습 단계에서 판별망 D 가 갖는 상대적으로 강인한 이진 분류 능력을 극복할 수 있도록 적절한 초기 가중치를 부여함으로써 G 의 원활한 학습을 유도함과 동시에, 픽셀 수준에서 보다 높은 품질을 가질 수 있도록 영상 생성을 위한 잠재벡터의 변환 방법을 일차적으로 학습한다.

인코더 E 는 영상으로부터 복원을 고려한 저차원 벡터로의 투영방법을 학습하고 잠재공간 내 데이터셋의 매니폴드를 구성한다. 추가로, 본 논문에서는 영상의 복원뿐만이 아닌 컨디셔닝에 따른 생성 영상 제어를 위해 분류기 C 를 도입하였다. 잠재벡터 z 를 입력으로 클래스를 예측하는 C 의 학습을 통해 E 가 인코딩하는 잠재벡터는 영상에 대한 전반적인 표현만을 학습하는 것이 아니라, 각 클래스에 내재하는 분리된 특징을 학습함으로써 추후 컨디셔닝에 따른 생성 제어가 가능하도록 유도한다. 본 논문에서는 조건 레이블 결합을 통한 컨디셔닝을 통해 원하는 특징을 포함하는 영상을 생성하는 것이 목적이므로, 하나의 예측된 레이블 결과 안에 타겟 도메인 조건을 모두 포함하는 다중 레이블 분류가 가능한 신경망 C 를 학습한다. 따라서 목적함수는 단순 클래스 분류를 위한 교차엔트로피가 아닌 다중 레이블 정보를 담은 실제 값 y 와 C 를 통해 예측된 다중 레이블 $\hat{y}=C(z)$ 간 다중 레이블 소프트 마진 손실 함수를 사용

하였으며 수식 (2)와 같이 사용하였다.

$$L_{ml}(\hat{y}, y) = -\frac{1}{N_c} \sum_j y_j \log((1 + \exp(-\hat{y}_j))^{-1}) + (1 - y_j) \log\left(\frac{\exp(-\hat{y}_j)}{1 + \exp(-\hat{y}_j)}\right). \quad (2)$$

레이블 인덱스 $j \in \{0, 1, \dots, N_c - 1\}$ 에 대해 원-핫 형태의 y 가 아닌 다중 레이블 값에 대해 1을 가지며 각 레이블에 대한 예측 확률 $\hat{y}_j \in \{0, 1\}$ 을 모두 높인다.

이와 같이 신경망 E 는 C 를 통해 N_c 개 레이블에 대한 표현을 함께 학습함으로써 레이블 조건의 컨디셔닝을 통한 생성 결과의 제어가 가능하도록 구현하며, 컨디셔닝을 반영한 제어 가능 생성 모델 학습의 구체적인 방법은 다음 절에서 다룬다. 사전 학습된 G 만을 생성망으로써 재사용하는 것이 아닌, 특징매핑으로의 E 와 다중 레이블 C 의 재사용을 위해 전체 신경망 구조를 구성하였고 이를 그림 3에 나타내었다.

제안하는 신경망 구조에서는 입력 영상의 순전파가 진행됨에 따라 인코더와 디코더 내부의 각 레이어가 추출하는 서로 다른 스케일의 특징맵들을 마지막 컨볼루션 레이어의 특징맵 스케일로 변형하여 채널 합을 해줌으로써 다중 스케일 특징을 반영해 저차원 공간으로 투영하고 또한 이를 활용해 영상을 복원할 수 있도록 하는 멀티-합 잔차 연결^[17] 기법을 도입하였다. E 를 통

해 인코딩된 z 를 입력받아 영상으로 재복원하는 G 와, 해당 잠재벡터를 다중 레이블로 분류하는 C 를 활용해 특징을 분리할 수 있도록 첫 번째 사전학습을 위한 목적함수를 구성하였으며 수식 (3)과 같이 사용하였다.

$$L_{pre} = L_{rec} + \lambda_1 L_{ml}. \quad (3)$$

여기서 λ_1 는 두 손실함수 텀 간 균형을 위한 페널티이며, 본 논문에서는 실험적으로 1을 이용하였다.

오토인코더 구조의 사전학습에서 사용하는 손실 함수를 정리하면 다음과 같다.

- L_{rec} : 입력 영상 x 와 출력 영상 $\hat{x}$ 의 손실을 줄이는 재복원 손실 함수
- L_{ml} : 다중 레이블에 대한 소프트 마진 손실 함수
- L_{pre} : 재복원 손실과 다중 레이블 소프트 마진 손실을 가중합 한 통합 손실 함수

2. 사전학습 신경망을 활용한 컨디셔닝 생성 모델

조건을 벡터화하여 입력 잠재벡터와의 결합을 통해 생성 영상을 제어하는 방법은 컨디셔닝의 고전적인 접근이며, cGAN^[4]에서는 생성망과 판별망 각각에 레이블 정보를 추가로 부여함으로써 원하는 도메인으로의 영상 생성을 유도한 바 있다. 하지만 입력 영상이 고해상도일 경우에는 조건까지 주어진 고차원 결합확률분포의 우도(likelihood) 최대화가 난해해짐에 따라 컨디셔

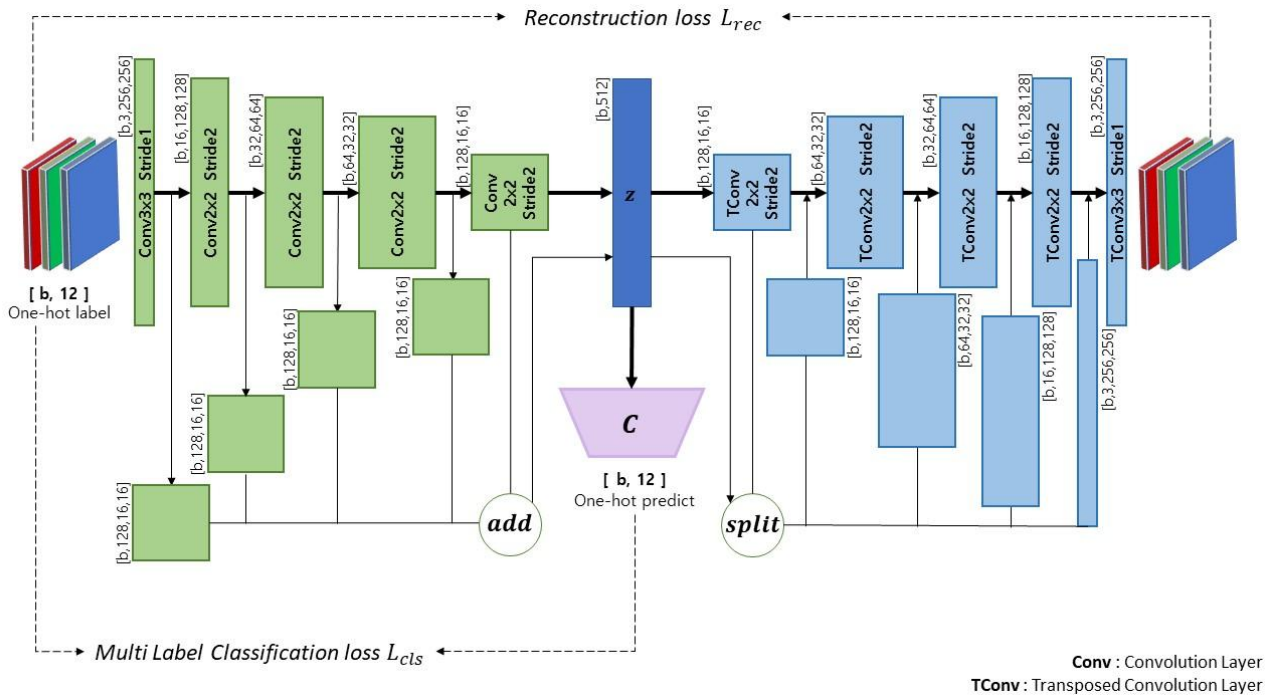


그림 3. PSC-GAN의 첫 번째 학습 단계에서 활용한 멀티-합 잔차 연결 오토인코더 구조

Fig. 3. Multi-add residual autoencoder architecture utilized in the first training stage of PSC-GAN.

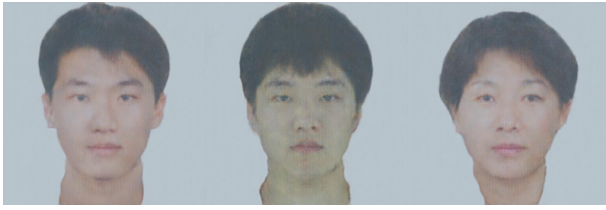


그림 4. ['여성', '웨이브형', '쌍꺼풀'] 조건 단순 결합 학습을 통해 학습한 모델 생성 영상

Fig. 4. Generated images of a model trained through ['Female', 'Wavy hair', 'Double eyelid'] conditional concatenation learning.

닝에 따른 생성망 학습이 쉽게 이루어지지 않으며, 그림 4에서 보이는 바와 같이 결과 영상 또한 조건을 반영하지 못한채 데이터셋 분포 근사 모델의 무작위 샘플링을 통해 원하는 이미지와는 전혀 다른 결과물이 생성되는 경향을 보인다. 또한 기존의 결합 조건을 이용한 생성 방법에서는 단일 도메인에 대해 동작하였으며 다중 도메인 조건 반영을 위해서는 별도의 마스크벡터가 추가적으로 요구되었다^[6].

본 논문에서는 추가적인 마스크 벡터 없이 조건 벡터 c 에 타겟으로 하는 다중 도메인 레이블을 모두 포함하여 생성망 입력인 잠재벡터 z 와 결합함으로써 생성 결과를 제어한다. 이를 위해 첫 번째 학습 단계를 통해 사전학습된 신경망 E 와 C 를 재사용하여 다시 한 번 생성 결과의 예측 레이블 오차를 줄여나가는 방식으로 모델 학습을 진행하였다. 두 번째 학습 단계에서는 사전학습된 E 와 C 의 가중치를 고정된 E' 와 C' 를 복사해 사용함으로써 N_c 개 레이블을 반영한 매핑 및 특징 추출, 그리고 다중 레이블 분류 능력을 그대로 활용한 다. GAN 학습을 위해 판별망 D 를 두 번째 학습단계에서 새롭게 학습하며, 사전학습된 디코더 G 는 적대적 학습을 통해 높은 품질을 갖는 영상 생성망으로써의 능력을 학습한다. 그림 5는 생성망 G 를 위한 PSC-GAN의 두 번째 학습 방법과 전체 모델 구조를 나타낸다.

먼저 다중 레이블 정보를 포함하는 조건 c 와 잠재벡터 z 를 결합하고 단일 전연결층을 통해 G 의 입력 차원으로 매핑한다. 해당 벡터 입력에서 G 를 통해 영상이 생성되며, 판별망 D 는 실제 훈련 데이터셋 샘플 영상인 x 와 생성영상의 비교를 통해 적대적으로 동작한다. 조건 c 를 포함하는 적대적 학습의 목적함수 L_{adv} 는 (4)와 같으며, 이진교차엔트로피를 이용하였다.

$$L_{adv} = \mathbf{E}_x [\log D(x)] + \mathbf{E}_{z,c} [1 - \log D(G(z,c))]. \quad (4)$$

이때 비록 조건 c 에 따른 컨디셔닝 생성을 위해 E' 와 C' 를 활용하여 분류된 예측 레이블과 실제 다중 레이블 간 차이를 최소화하는 방향으로 G 를 학습시키며, 이를 위한 목적함수를 도입하였다. 목적함수 L_{cls} 를 수식 (5)와 같이 사용하였다.

$$L_{cls} = \mathbf{E}_{z,c} [-\log C^f(c | E^f(G(z,c)))]. \quad (5)$$

(5)는 다중 도메인에 대한 특징 추출과 레이블 분류를 E' 와 C' 가 사전 학습하였으므로 N_c 개 레이블에 대한 고유 특징을 분리해 G 의 학습에 반영함으로써 G 가 내재적으로 학습하는 잠재공간에서 영상 분포로의 매핑 역시 조건 입력 c 에 따라 달라지게 하는 형태의 기능을 유도하기 위함이다. 해당 목적함수를 위한 손실 계산은 (2)와 동일하게 구성하였다. 결과적으로, 추론 시에는 조건입력 c 에 따른 결과 영상 제어 및 생성이 가능하도록 하며, 해당 단계에서 E' 와 C' 의 가중치는 고정된 채 기울기만을 역전파한다. 이에, 두 번째 학습 단계에서는 적대적 학습과 다중 분류 목적함수의 합을 최종 목적식으로 수식 (6)과 같이 사용하였다.

$$L = L_{adv} + \lambda_2 L_{cls}. \quad (6)$$

λ_2 는 조건 제어를 위한 분류 손실 함수에 대한 페널티이며 실험적으로 1로 설정하였다.

사전학습 신경망을 활용한 컨디셔닝 생성 학습에서 사용하는 손실 함수를 정리하면 다음과 같다.

- L_{adv} : 생성자와 판별자의 적대적 학습에 사용되는 이진교차엔트로피 손실 함수
- L_{cls} : 잠재벡터 z 에 결합된 조건 c 와 사전학습 신경망을 예측된 다중 레이블에 대한 소프트 마진 손실 함수
- L : 이진교차엔트로피 손실과 다중 레이블 소프트 마진 손실을 가중합 한 통합 손실 함수

IV. 실험 결과

1. 활용 데이터셋

본 기술을 시각적으로 검증하기 위해 특징 레이블들이 명시적으로 잘 나타나 있는 몽타주 데이터셋^[18]을 활용하여 실험을 수행하였다. 몽타주 데이터셋은 가상 인물 이미지와 인물에 대한 설명문에 기반한 몽타주 스케치 등으로 구축된 데이터셋이다. 이에 안면 특성이 레이블에 명시적으로 반영되어 있으며 성별과 얼굴, 눈,

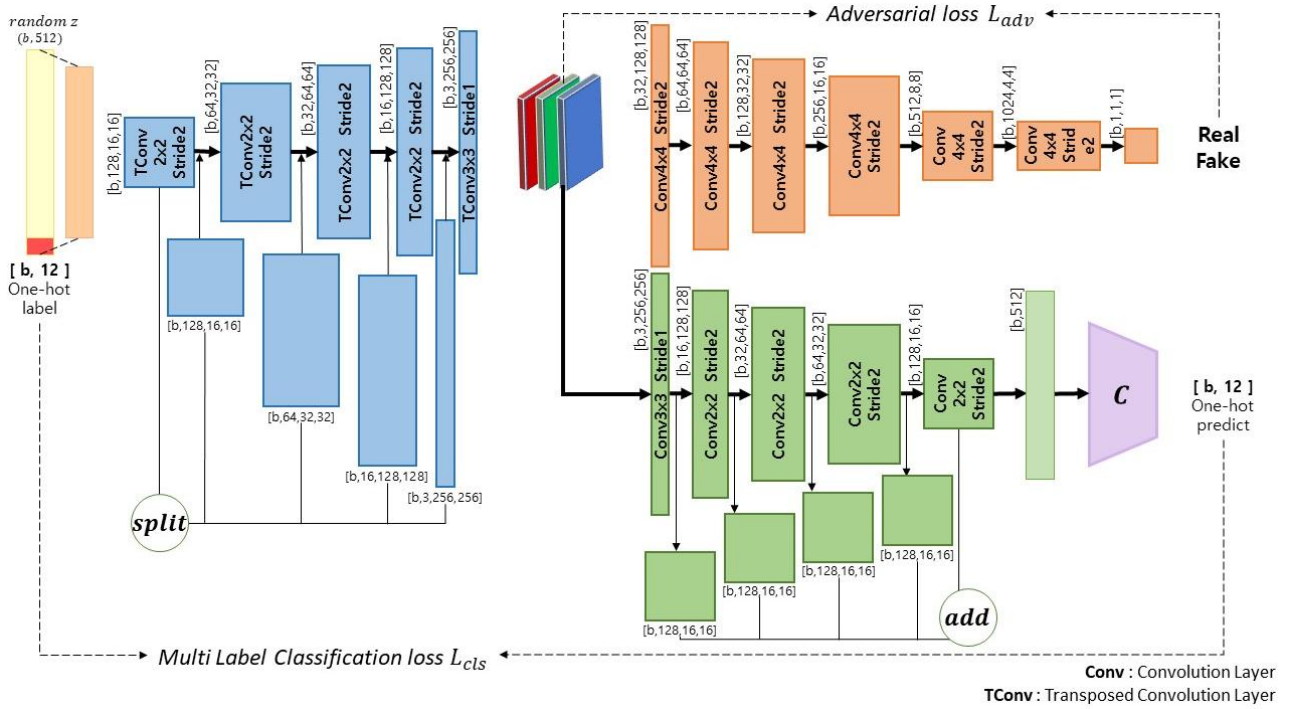


그림 5. PSC-GAN의 두 번째 학습 단계에서 활용한 생성망 구조

Fig. 5. GAN framework fine-tuned in the second training stage of PSC-GAN.

머리 스타일 등에 대한 특징들이 레이블로 포함되어 있다. 본 실험에서는 성별과 머리 그리고 눈 스타일을 사용한 컨디셔닝을 검증하였다. 성별은 '남성', '여성'으로 머리 스타일은 '곱슬머리', '직모(생머리)', '스포츠형', '탈모형', '웨이브형' 마지막으로 눈 스타일은 '쌍꺼풀', '속쌍꺼풀', '왼쪽 비대칭', '외꺼풀', '오른쪽 비대칭'으로 성별, 머리, 눈 스타일 3개의 복수 클래스를 탐지하여 분류가 된다. 따라서 하나의 레이블 길이는 총 성별 2, 머리 스타일 5, 눈 스타일 5의 합인 12이며 그 중 각 클래스에 대하여 정답인 3개의 값은 1로 정답이 아닌 값들은 0으로 표기 된다.

2. 오토인코더 사전학습을 이용한 생성 모델 결과

컨볼루션 연산을 이용한 vanilla GAN 학습을 통한 생성 결과와 제안하는 방법을 통해 생성된 결과를 그림 2와 6에 각각 나타내었다. PSC-GAN을 통해 생성된 이미지는 사전 학습된 G 의 초기 가중치를 사용하여 생성 모델이 재학습됨으로써 초기 단계부터 적대적 학습에서의 안정적인 모델 학습이 이루어지는 것과 동시에, 학습의 빠른 수렴을 긍정적으로 확인할 수 있었다.

제안하는 모델의 우수성을 수치로 입증하기 위해 본 논문에서는 객관적인 이미지 성능 평가 지표를 사용하였다. 표 1은 일반적인 생성 모델인 GAN과 PSC-GAN

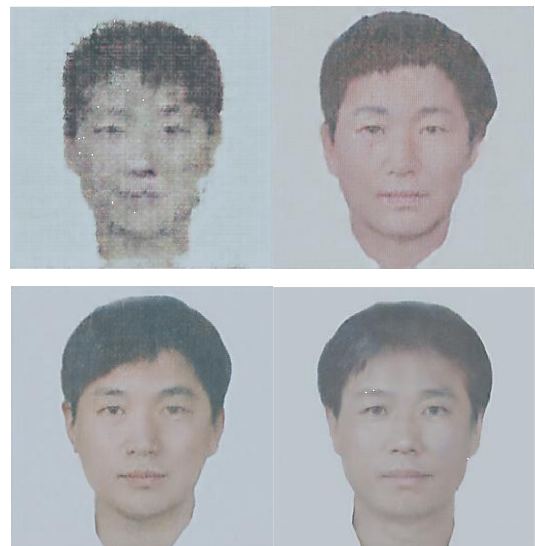


그림 6. 256×256 크기 영상에 대한 사전학습된 디코더를 사용한 GAN 학습: 0, 5, 10, 30 에폭 생성 결과

Fig. 6. GAN training using pre-trained decoder for 256×256 size images: Generated results for 0, 5, 10 and 30 epochs.

의 제안 기술 중 하나인 사전학습 G 를 활용한 생성 모델인 GAN의 생성된 결과를 비교한 측정 결과이다.

정량적 수치로써 제안하는 모델의 우수성을 입증하고자 생성 모델의 성능을 평가하는 지표인 Inception Score(IS)^[19], Fréchet Inception Distance(FID)^[20],

Learned Perceptual Image Patch Similarity (LPIPS)^[21]를 사용하였다. IS는 Inception 모델을 기반으로 생성 결과의 다양성을 반영하는 이미지 품질 평가 지표이다. FID는 참조 이미지의 분포와 생성된 영상의 분포에 대한 Wasserstein-2 거리를 측정함으로써 품질을 계산하는 지표이다. 마지막으로 LPIPS는 분류 네트워크가 추출하는 특징 공간에서의 거리를 계산함으로써 인지적 화질을 표현하는 평가 지표이다.

표 1. GAN과 사전학습 디코더를 활용한 GAN 모델 생성 결과의 IS, FID, LPIPS 측정

Table 1. IS, FID, and LPIPS measurement of vanilla GAN and proposed model employing pre-trained decoder.

	IS ↑	FID ↓	LPIPS ↓
Vanilla GAN	2.47	188.08	0.60
Proposed GAN employing pre-trained decoder	2.69	172.71	0.51

베이스라인 모델인 vanilla GAN은 사전학습 디코더인 G 의 효과적인 초기화가 없는 형태로 가장 일반적인 GAN 구조이다. 이를 사전학습된 디코더를 G 로써 활용하여 재학습을 수행한 제안 기술과 비교할 때, 제안하는 오토인코더를 통한 사전학습 방법이 높은 IS 및 낮은 FID를 가짐으로써 생성 결과의 품질과 다양성을 보장한

다는 것을 확인할 수 있다. 또한 베이스라인 모델 대비 낮은 LPIPS를 나타냄으로써, 시각 인지적으로 보다 향상된 결과물을 생성해냄을 객관적으로 증명할 수 있었다. 정성적/정량적 결과로써 제안하는 오토인코더를 활용한 사전학습과 디코더의 생성망 재사용을 통해 높은 학습 안정성을 지닌 생성 모델 구축이 가능함을 보였다.

3. 사전학습 신경망 활용 컨디셔닝 생성

제안하는 사전학습 잠재벡터에서의 분류기를 통한 컨디셔닝 기법의 우수성을 확인하고자 애블레이션 실험 (ablation study)을 위해 사전학습 E 와 C 를 통해 컨디셔닝을 제어하는 방법의 생성 결과를 그림 7과 같이 사전학습 신경망 E 와 C 를 사용하지 않고 랜덤 벡터 z 에 조건 레이블을 결합한 학습 방법과 비교하였다 (이하 애블레이션). PSC-GAN과 공정한 비교의 애블레이션 실험을 위해 50가지 레이블 조합 중 4가지 조합을 실험적으로 뽑아 학습을 진행하였다. 기 컨디셔닝에 사용한 조합을 Set 라고 명명하였으며 각 Set 는 다음과 같다.

표 2. 애블레이션 실험을 위해 활용한 4가지 조합

Table 2. 4 Label combinations for ablation study.

	Set_1	Set_2	Set_3	Set_4
Label combination	[여성, 웨이브형, 쌍꺼풀]	[여성, 직모, 외꺼풀]	[남성, 곱슬머리, 쌍꺼풀]	[남성, 직모, 외꺼풀]

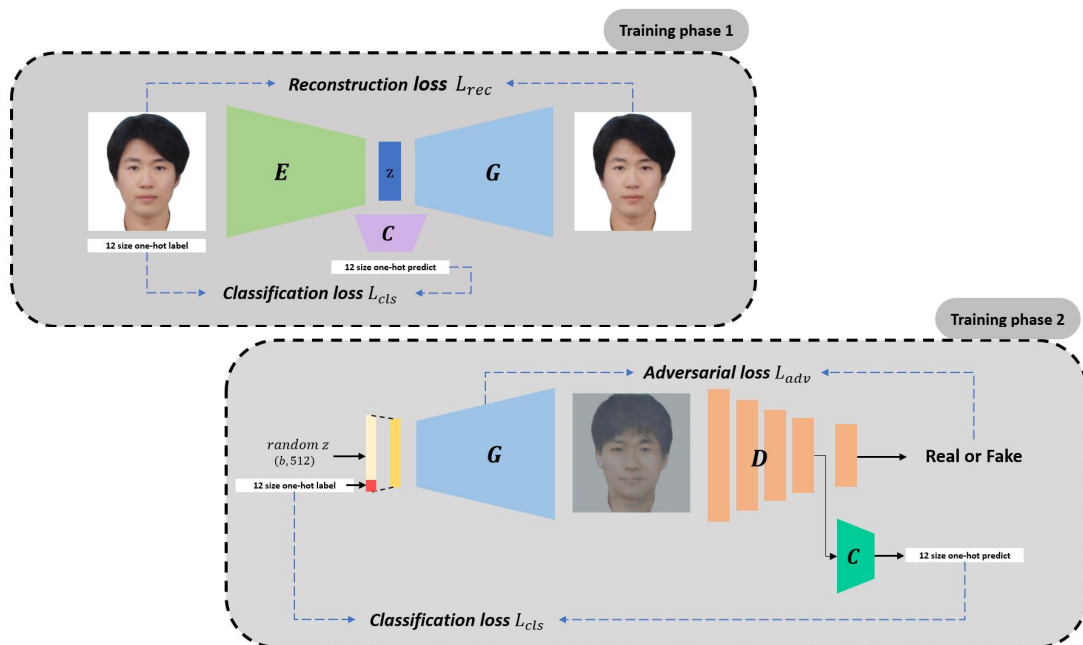


그림 7. PSC-GAN에서 사용된 사전학습 E 와 C 를 사용하지 않은 애블레이션

Fig. 7. Training framework of ablation which eliminates pre-trained E and C from the PSC-GAN.



그림 8. Set_1 (여성, 웨이브형, 쌍꺼풀)에 대한 PSC-GAN(위)과 애블레이션(아래) 생성 결과

Fig. 8. Generated results by PSC-GAN (top) and ablation (bottom) for Set_1 .



그림 9. Set_2 (여성, 직모, 외꺼풀)에 대한 PSC-GAN(위)과 애블레이션(아래) 생성 결과

Fig. 9. Generated results by PSC-GAN (top) and ablation (bottom) for Set_2 .

그림 8부터 11까지 제안하는 PSC-GAN과 애블레이션의 Set_1 - Set_4 컨디셔닝에 따른 동일 학습 에폭에 대해 생성된 이미지들을 나타내었다. 각각의 그림에서 위 행의 4개 이미지는 본 논문에서 제안하는 사전학습된 E 와 C 를 통해 조건을 제어함으로써 생성된 결과이다. 주어진 Set 에 부합되게 이미지가 생성됨을 확인할

수 있다. 반면, 그림 8에서 보이듯 애블레이션 생성 결과(아래)에서 제안 모델과는 달리 성별이 결합 조건과는 상이한 이미지가 많으며 눈 스타일 또한 다르게 생성된 것을 확인할 수 있다. 하지만 PSC-GAN의 결과는 조건을 결합한대로 원하는 이미지가 생성된 것을 확인할 수 있으며, 보다 뚜렷하고 세밀한 얼굴 영역에서

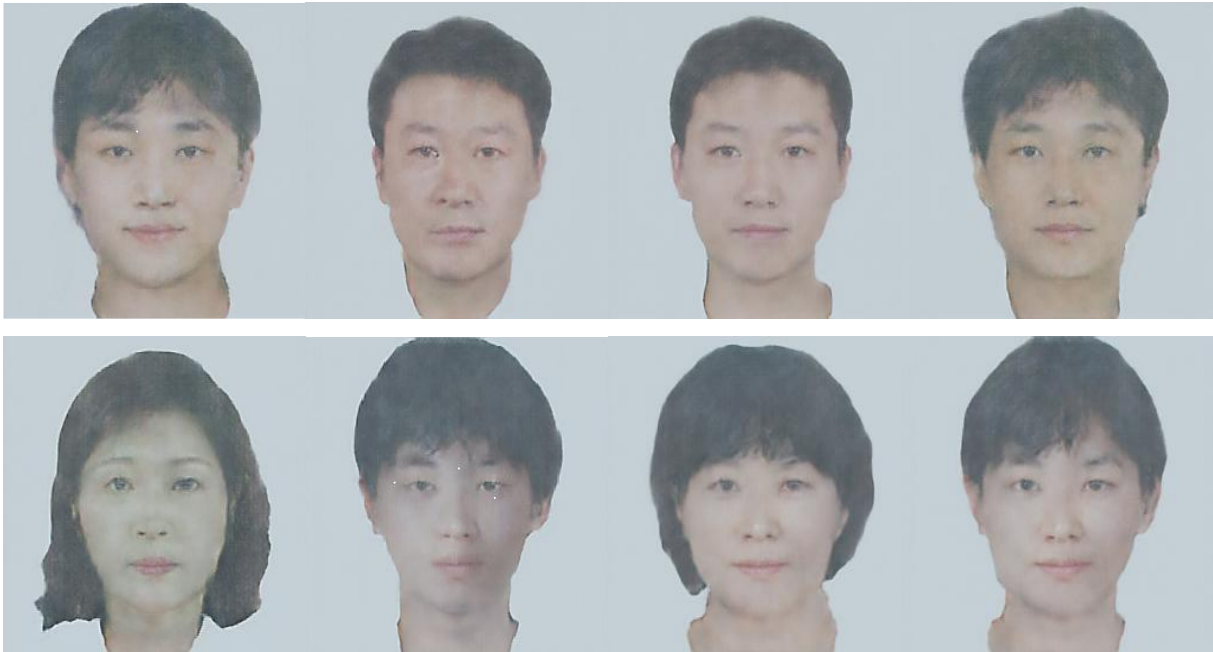


그림 10. Set_3 (남성, 곱슬머리, 쌍커풀)에 대한 PSC-GAN(위)과 애블레이션(아래) 생성 결과
Fig. 10. Generated results by PSC-GAN (top) and ablation (bottom) for Set_3 .



그림 11. Set_4 (남성, 직모, 외꺼풀)에 대한 PSC-GAN(위)과 애블레이션(아래) 생성 결과
Fig. 11. Generated results by PSC-GAN (top) and ablation (bottom) for Set_4 .

표현력이 향상되어 생성된 것을 확인 가능하다. 마찬가지로 그림 9에서 역시 제안 모델과는 달리 애블레이션 결과물의 색상 표현과 더불어 머리 스타일과 입 모양 등이 불완전하게 생성된 것을 토대로, 불안정한 학습에 따른 저품질의 데이터분포추정 및 샘플링이 수행된 것으로 판단할 수 있다. 그림 10와 11 또한 PSC-GAN과

애블레이션 비교를 통해 제안하는 방법이 조건을 결합하여 준 대로 모델의 학습을 제어하여 원하는 스타일이 반영된 고해상도의 결과물을 생성해낸 것을 확인 가능하다.

조건 결합의 우수성을 입증하기 위해 결합한 조건 Set 에 대한 애블레이션 결과의 예측 레이블과

PSC-GAN 결과의 예측 레이블의 정확도(accuracy)를 표 1과 같이 구분하였고, 생성 이미지의 품질 평가를 위한 FID와 LPIPS를 성능 평가 지표로 활용하였다. 정확도는 입력단에서 결합하는 조건 레이블 *Set*와 생성된 이미지의 예측 레이블의 정답 비율로써, 생성된 이미지가 의도대로 생성되었는지 평가하기 위한 분류 성능 지표를 활용된다. 표 3은 제안 모델인 PSC-GAN과 애블레이션의 결과에 대한 측정 수치를 나타내었다. 정량적 수치인 정확도는 높을수록 FID, LPIPS는 낮을수록 우수하다. 그러므로 제안 모델인 PSC-GAN이 애블레이션 대비 의도된 높은 품질의 결과물을 생성해낸다.

표 3. 애블레이션 실험의 생성 결과와 PSC-GAN의 생성 결과의 accuracy, FID, LPIPS 측정

Table 3. Accuracy, FID, and LPIPS measurements of the ablation study and of PSC-GAN generation results.

	Accuracy ↑		FID↓		LPIPS↓	
	Ablation	PSC-GAN	Ablation	PSC-GAN	Ablation	PSC-GAN
<i>Set</i> ₁	0.54	0.77	124.03	120.06	0.43	0.42
<i>Set</i> ₂	0.58	0.77	107.38	101.91	0.46	0.46
<i>Set</i> ₃	0.67	0.76	103.64	86.74	0.44	0.43
<i>Set</i> ₄	0.64	0.76	114.55	97.90	0.43	0.41

PSC-GAN이 애블레이션 대비 낮은 FID를 가짐으로써 제안 모델이 참조 이미지의 분포를 잘 따라서 참조 이미지와 비슷하지만 다른 이미지들을 생성해낸 것을 알 수 있다. 또한 PSC-GAN이 애블레이션 대비 0~1%의 낮은 LPIPS를 가짐으로써^[22, 23], 시각적으로 그럴듯한 결과물을 생성해낸 것을 확인할 수 있다. 마지막으로 컨디셔닝의 척도로 정확도를 보면 알 수 있듯, 애블레이션은 거의 절반 정도를 상회하는 수준의 조건 부여 이미지를 생성하지 못하는 것을 확인할 수 있었지만, PSC-GAN은 애블레이션의 정량적 수치와 0.3 가까이 차이가 나는 정확도를 보임으로써 조건에 부합하는 이미지를 생성해낸 것을 확인할 수 있다.

V. 결 론

본 논문에서는 GAN이 갖는 학습의 불안정성을 극복하고 컨디셔닝을 통한 생성 영상의 제어를 원활히 하기 위한 두 단계의 생성 모델 학습 기법에 대한 연구를 수행하였다. 먼저 오토인코더 구조로 사전학습된 디코더

를 생성망 학습에 재사용함으로써 안정적 학습으로 유도하였다. 또한, 이 과정에서 사전학습된 인코더가 별도 분류기를 통해 다중 레이블에 대한 특징을 분리해 매핑하는 함수 기능을 갖도록 학습됨에 따라 GAN의 적대적 학습 시, 해당 정보를 재사용함으로써 조건을 반영한 잠재벡터가 무작위 영상으로 생성되는 것이 아닌 컨디셔닝에 따라 결과물을 제어할 수 있음을 증명하였다. 실험을 통해 제안 방법으로 학습한 GAN 결과가 기존 기술 대비 정성적으로 우수한 품질의 영상을 도출한 것을 확인할 수 있었으며, 컨디셔닝에 충실한 영상을 생성해낸 것을 정량적으로 보였다. 본 연구에서 제안한 일련의 학습 방법은 GAN 뿐만이 아닌 다른 여타 생성 모델 학습에도 적용이 가능하며, 추후 다중 레이블 분류기 기반의 컨디셔닝 기법을 다중 모달리티로 제어되는 확산 프로세스 생성 모델의 학습에 반영할 수 있다.

REFERENCES

- [1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, "Generative Adversarial Networks," arXiv preprint arXiv:1406.2661, Jun. 2014.
- [2] Divya Saxena, Jiannong Cao. "Generative Adversarial Networks (GANs): Challenges, Solutions, and Future Directions," Association for Computing Machinery Computing Surveys (ACM), vol 54, pp 1-42. 2020.
- [3] Youssef Kossale, Mohammed Airaj, Aziz Darouichi, "Mode Collapse in Generative Adversarial Networks: An Overview," in International Conference on Optimization and Applications (ICOA), 2022.
- [4] T. L. Paine, P. Khorrami, W. Han, and T. S. Huang, "An analysis of unsupervised pre-training in light of recent advances," International Conference on Learning Representations (ICLR), 2015.
- [5] G. E. Hinton, S. Osindero, Y-W. Teh, "A fast learning algorithm for deep belief nets," Neural Computation, vol. 18, no. 7, pp. 1527-1554, 2006.
- [6] Mehdi Mirza, Simon Osindero. "Conditional Generative Adversarial Nets," arXiv preprint arXiv:1411.1784, Nov. 2014.
- [7] Jiayi Ma, Han Xu, et al., "DDcGAN: A Dual-Discriminator Conditional Generative

- Adversarial Network for Multi-Resolution Image Fusion,” IEEE Trans. on Image Process (TIP), vol 29. March. 2020
- [8] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, Jaegul Choo. “StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation,” Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR), pp. 8789-8797. 2018.
- [9] Yunjey Choi, Youngjung Uh, Jaejun Yoo, Jung-Woo Ha. “StarGAN v2: Diverse Image Synthesis for Multiple Domains,” Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR), pp. 8188-8197. 2020.
- [10] T. Karras, S. Laine and T. Aila, “A style-based generator architecture for generative adversarial networks,” Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR), pp. 4401-4410. 2019.
- [11] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, Ole Winther. “Autoencoding beyond pixels using a learned similarity metric,” Proceedings of Machine Learning Research (PMLR), Feb. 2016.
- [12] Taesung Park, Jun-Yan Zhu, Oliver Wang, Jingwan Lu, et al. “Swapping Autoencoder for Deep Image Manipulation,” Neural Information Processing Systems (NeurIPS), 2020.
- [13] Hyungrok Ham, Tae Joon Jun, Daeyoung Kim. “Unbalanced GANs: Pre-training the Generator of Generative Adversarial Network using Variational Autoencoder,” arXiv preprint arXiv:2002.02112, Sept. 2020
- [14] Yaxing Wang, Chenshen Wu, Luis Herranz, Joost van de Weijer, Abel Gonzalez-Garci, et al. “Transferring GANs: generating images from limited data,” Proceedings of the European Conference on Computer Vision (ECCV), pp. 218-234. 2018.
- [15] Alec Radford, Luke Metz, Soumith Chintala. “Unsupervised representation learning with deep convolutional generative adversarial networks” arXiv preprint arXiv:1511.06434, Nov. 2015
- [16] Xudong Mao, Qing Li, Haoran Xie, Raymond Y.K. Lau, Zhen Wang, Stephen Paul Smolley. “Least Squares Generative Adversarial Networks,” Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2794-2802, 2017.
- [17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. “Deep Residual Learning for Image Recognition,” Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778. 2016.
- [18] “Persona-based virtual character montage data,” <https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSetSn=618>, 2021.
- [19] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen and X. Chen, “Improved Techniques for Training GANs,” Advances in neural information processing systems (NIPS), 2016.
- [20] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler and S Hochreiter, “Gans Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium,” Advances in neural information processing systems (NIPS), 2017.
- [21] R. Zhang, P. Isola, A. A. Efros, E. Shechtman and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR), pp. 586-595, 2018.
- [22] Seung Ho Park, Young Su Moon, Nam Ik Cho, “Perception-Oriented Single Image Super-Resolution Using Optimal Objective Estimation,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1725-1735, 2023.
- [23] Håkon Hukkelås, Morten Smebye, Rudolf Mester, Frank Lindseth, “Realistic Full-Body Anonymization With Surface-Guided GANs,” Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 1430-1440, 2023.

 저 자 소 개



김 지 수(비회원)
2022년 한성대학교 패션학부
학사 졸업.
2023년 한성대학교 일반대학원
AI응용학과 석사 재학.

<주관심분야: 영상처리, 이미지생성, 컴퓨터비전>



지 준(정회원)
1989년 서울대학교 자원공학과
석사 졸업.
1995년 Stanford University,
Geophysics 박사 졸업.
1996년 3DGeo Inc. 선임연구원.
1997년~현재 한성대학교
AI응용학과 교수.

<주관심분야: 인공지능, 딥러닝, 머신러닝>



오 희 석(정회원)
2017년 연세대학교
전기전자공학과
박사 졸업.
2017년~2017년 삼성전자 DMC
연구소 책임연구원.
2017년~2020년
한국전자통신연구원
선임연구원.

2020년~현재 한성대학교 AI응용학과 조교수.
<주관심분야: 영상처리, 컴퓨터비전, 혼합현실,
심층생성모델 등>