

# ACL-GAN: 새로운 loss 를 사용하여 하이퍼 파라미터 탐색속도와 학습속도를 향상시킨 영상변환 GAN

조정익, 윤경로  
건국대학교

jeongik.jo.01@gmail.com, yoonk@konkuk.ac.kr

## ACL-GAN: Image-to-Image translation GAN with enhanced learning and hyper-parameter searching speed using new loss function

JeongIk Cho, Kyoungro Yoon  
Konkuk University

### 요 약

Image-to-image 변환에서 인상적인 성능을 보이는 StarGAN 은 모델의 성능에 중요한 영향을 끼치는 adversarial weight, classification weight, reconstruction weight 라는 세가지 하이퍼파라미터의 결정을 전제로 하고 있다. 본 연구에서는 이 중 conditional GAN loss 인 adversarial loss 와 classification loss 를 대체할 수 있는 attribute loss 를 제안함으로써, adversarial weight 와 classification weight 를 최적화하는 데 걸리는 시간을 attribute weight 의 최적화에 걸리는 시간으로 대체하여 하이퍼파라미터 탐색에 걸리는 시간을 획기적으로 줄일 수 있게 하였다. 제안하는 attribute loss 는 각 특징당 GAN 을 만들 때 각 GAN 의 loss 의 합으로, 이 GAN 들은 hidden layer 를 공유하기 때문에 연산량의 증가를 거의 가져오지 않는다. 또한 reconstruction loss 를 단순화시켜 연산량을 줄인 simplified content loss 를 제안한다. StarGAN 의 reconstruction loss 는 generator 를 2 번 통과하지만 simplified content loss 는 1 번만 통과하기 때문에 연산량이 줄어든다. 또한 이미지 Framing 을 통해 배경의 왜곡을 방지하고, 양방향 성장을 통해 학습 속도를 향상시킨 아키텍처를 제안한다.

### 1. 서론

최근의 연구들은 원본 영상의 특징들을 이용하여 새로운 영상을 생성하는 GAN 에 많은 관심을 보이고 있으며, 그 중 한 응용으로 사람의 얼굴을 원하는 특징을 갖도록 변화시키는 연구가 진행되었다. 그 대표적인 경우인 StarGAN [1]은 WGAN-GP [2]의 adversarial loss, CycleGAN [3]의 reconstruction loss, conditional GAN [4]의 classification loss 를 사용하여 원본 사람의 얼굴을 원하는 특징을 갖도록 바꾼다. 그러나 이 경우, adversarial loss, reconstruction loss, 그리고 classification loss 의 세가지 손실함수를 조합하여 사용하기 때문에, 이 세가지 loss 들의 가중치, 즉 adversarial weight, reconstruction weight, classification weight 의 세가지 하이퍼파라미터의 결정이 그 성능에 중요한 영향을 미치게 된다. 따라서 생성된 GAN 의 용도가 변경되는 경우 이 세가지 하이퍼파라미터의 결정을 위한 실험 작업이 네트워크의 훈련 전, 또는 훈련 도중에 큰 비중을 갖게 되고 그 시간도 적지 않게 차지 하게 된다. 이러한 단점을 해결하기 위하여 본 논문에서는 adversarial loss 와 classification loss 를 대체할 수 있는 attribute loss 를 제안하여 결정하여야 하는 하이퍼파라미터의 수를 하나로 줄여, 성능의 손실 없이, 최적화된 loss 들의 가중치를 구하는 작업이 매우 간단하게 변화되었다.

### 2. 데이터 분포에 따른 각 loss 의 역할

StarGAN [1]에서 사용하는 Discriminator 의 loss, Generator 의 loss, 실제 이미지의 classification loss, Generator 가 생성한 가짜 이미지의 classification loss 는 각각 다음과 같다.

$$L_D = L_{adv}^d + \lambda_{cls} L_{cls}^r$$

$$L_G = L_{adv}^g + \lambda_{cls} L_{cls}^g$$

$$L_{cls}^r = E_{x, att \sim P_r(x, att)}[-\log(D_{cls}(att|x))]$$

$$L_{cls}^g = E_{x', att' \sim P_g(x', att')}[-\log(D_{cls}(att'|x'))]$$

이 때,  $x, att \sim P_r(x, att)$ 에서  $x$  는 실제 데이터를 의미하고  $att$  는 실제 데이터가 가지고 있는 attribute 를 나타내는 이진 벡터이다.  $x', att' \sim P_g(x', att')$ 에서  $x'$  은 generator 가 생성한 데이터를 의미하고  $att'$  은  $x'$  을 생성할 때 latent vector 와 함께 generator 에 입력된 이진 벡터이다.

conditional GAN 에서 adversarial loss 는 잘 알려진 LSGAN [5] 또는 WGAN-GP[2]와 같은 잘 알려진 loss 를 사용할 경우 생성된 데이터 분포와 실제 데이터 분포가 멀어도 의미 있는 gradient 를 생성하여 학습이 잘 된다. 하지만 cross-entropy 를 사용하는 conditional GAN 의 classification loss 의 경우 cross-entropy 는 오직 KL-divergence 만 측정하기 때문에 학습 초기에 의미있는 gradient 를 생성하기 힘들다.

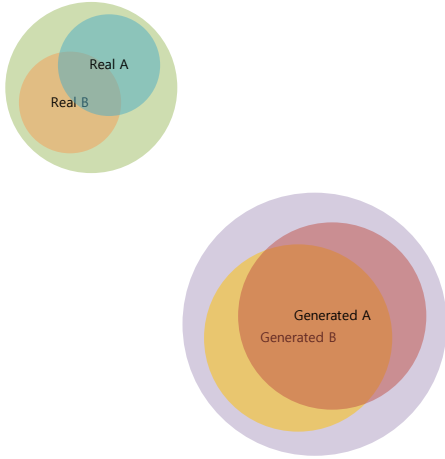


그림 1. 학습 초기의 데이터 분포

위의 그림 1 에서 Real A 는 attribute A 를 가진 실제 데이터 분포이고, Generated A 는 attribute A 를 가지도록 generator 가 생성한 분포이다. 학습 초기에는 Real A 와 Generated A 의 분포가 서로 멀리 떨어져 있어 classification loss 는 의미 있는 gradient 를 생성하지 못한다. LSGAN, WGAN-GP 등의 경우와 같이 두 분포의 거리가 멀어도 의미 있는 adversarial loss 만이 의미 있는 gradient 를 생성하여 학습이 진행된다.

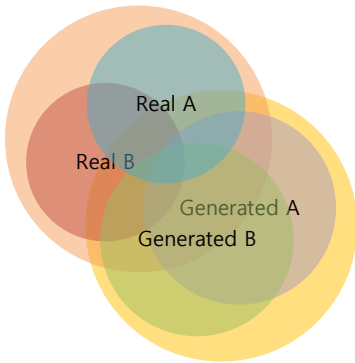


그림 2. 학습이 어느정도 진행된 후의 데이터 분포

그림 2 와 같이 학습이 어느정도 진행되어 Real A 와 Generated A, Real B 와 Generated B 가 어느정도 겹쳤을 때 classification loss 가 작동하여 conditional GAN 이 학습되게 된다.

### 3. Attribute loss 의 제안

이렇게 학습 초기에 classification loss 가 의미 있는 gradient 를 생성하지 못하는 문제를 해결하기 위해 attribute loss 를 제안한다. Attribute loss 는 그림 3 과 같이 각 Attribute 별로 GAN 을 만들 때, 각 GAN 들의 loss 의 합이다. Attribute A 를 학습하는 GAN A 의 경우 generator A 는 입력으로 latent vector 와 함께 activation value 를

받는다. activation value 로 1 을 받았을 때는 생성한 데이터로 discriminator A 를 속이려고 하고, 0 을 받았을 때는 don't care 한다(학습하지 않는다). Discriminator A 는 attribute A 를 가진 real data 에 대해서만 진짜라고 판별하고, attribute A 를 갖지 않은 real data 에 대해서는 don't care 한다. 또한 generator A 가 activation value 로 1 을 받아 생성한 데이터에 대하여는 가짜라고 판단하고, 0 을 받아 생성한 데이터에 대하여는 don't care 한다.

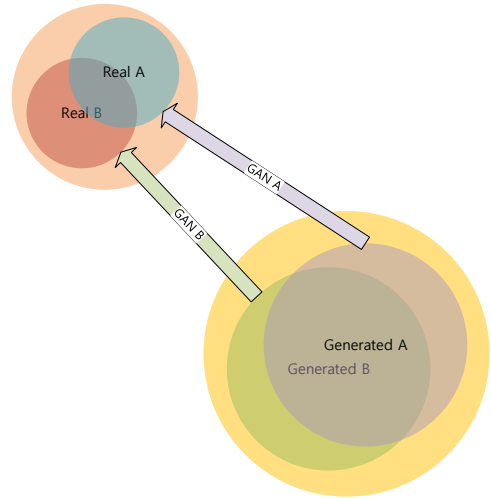


그림 3. Attribute loss

attribute loss 는 다음과 같이 정의 할 수 있다.

$$L_{att}^d = \sum_c^{att} L_c^d$$

$$L_{att}^g = \sum_c^{att} L_c^g$$

$$L_c^d = E_{x,c \sim P_r(x,c)} [(D_c(x) - 1)^2] + E_{x' \sim P_{G_c}(x',1)} [D_c(x')^2]$$

$$L_c^g = E_{x \sim P_z(x)} [(D_c(G_c(x, 1)) - 1)^2]$$

위 수식은 LSGAN 의 adversarial loss 를 쓰는 attribute loss 의 예시이다.  $L_c^d$  는 attribute c 를 학습하는 GAN 의 discriminator loss 이다.  $L_c^g$  는 attribute c 를 학습하는 GAN 의 generator loss 이다.  $x, c \sim P_r(x, c)$  에서  $x$  는 attribute c 를 가진 real data 이다.  $x' \sim P_{G_c}(x', 1)$  에서  $x'$  은 attribute c 를 학습하는 generator 인  $G_c$  가 latent vector 와 함께 activation value 로 1 을 입력 받아 생성한 데이터이다.  $x \sim P_z(x)$  에서  $x$  는 임의의 latent vector 이다. 각 GAN 들은 모든 hidden layer 를 공유하므로 하나의 모델처럼 생각할 수 있다. 이 경우 수식을 다음과 같이 변형할 수 있다.

$$L_{att}^d = E_{x, att \sim P_r(x, att)} [(D(x) - 1)^2 \cdot att] + E_{x', att' \sim P_g(x', att')} [D(x')^2 \cdot att']$$

$$L_{att}^g = E_{x', att' \sim P_g(x', att')} [(D(x') - 1)^2 \cdot att']$$

$x, att \sim P_r(x, att)$  에서  $att$  는  $x$  의 attribute 를 나타낸 이진 벡터이다.  $x', att' \sim P_g(x', att')$  에서  $att'$  은  $x'$  을 만들기 위해 generator 에 latent vector 와 함께 입력된

이진 벡터이다. ‘1’은 내적을 의미한다. attribute A 를 가진 데이터를 생성하려고 하고, 0 을 받았을 때는 discriminator A 는 attribute A 를 가진 real data 에 대해서만 진짜라고 판단하고, attribute A 를 가지지 않은 real data 에 대해서는 don't care 한다. 같은 원리로 WGAN-GP 와 같이 다른 adversarial loss 를 사용하는 attribute loss 도 만들 수 있다.

각 GAN 이 LSGAN 나 WGAN-GP 등의 loss 를 사용할 경우 학습 초기에도 의미 있는 gradient 를 생성할 수 있다. 각 GAN 은 모든 hidden layer 를 공유하기 때문에 하나의 GAN 을 사용하는 StarGAN 등과 비교하는 경우, 학습 연산량의 차이는 거의 없다. 또한 Attribute loss 의 경우 adversarial loss 와 classification loss 를 대체하므로 conditional GAN 의 하이퍼 파라미터 1 개가 줄어들기 때문에 하이퍼 파라미터 탐색에 걸리는 시간을 줄일 수 있다.

#### 4. 결론

본 논문에서는 Image-to-Image 변환 GAN 에서, adversarial loss, reconstruction loss 와 classification loss 의 세가지 loss 를 조합하는 대신, adversarial loss 와 classification loss 를 대체할 수 있는 attribute loss 를 제안하여, 결정하여야 하는 하이퍼파라미터의 수를 하나로 줄여, 최적화된 loss 들의 가중치, 즉, 최적화된 하이퍼파라미터를 구하는 작업이 매우 간단하게 변화되었다. 추가로 영상의 배경을 고정시키는 image framing 과 마스크 기법을 적용하여, 빠른 시간에 기존의 연구들 보다 우수한 결과를 얻을 수 있었다. 그림 4 는 좌측의 영상을 갈색머리, 입을 약간 벌리고, 웃는 모습으로 변환한 결과이다. 그림 4 는 좌측부터 원본영상, 변환된 영상, 변환을 위하여 생성된 마스크 영상, 마스크된 원본과 결합될 생성된 영상 세그먼트를 나타낸다.

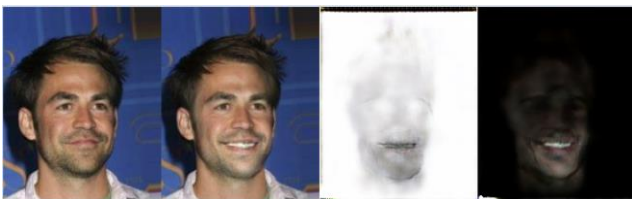


그림 4. 영상 변환 결과

#### 감사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학지원사업의 연구 결과로 수행되었음(No.2018-0-00213, SW 중심대학(건국대학교))

#### 참 고 문 헌

[1] 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, The IEEE

Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8789-8797, June 2018,

[2] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, Aaron Courville, Improved Training of Wasserstein GANs, Advances in Neural Information Processing Systems 30, pp. 5767~5777, 2017

[3] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, The IEEE International Conference on Computer Vision (ICCV), pp. 2223~2232, Oct. 2017

[4] Mehdi Mirza, Simon Osindero, Conditional Generative Adversarial Nets, ArXiv, 2014, available at <https://arxiv.org/abs/1411.1784>

[5] Xudong Mao, Qing Li, Haoran Xie, Raymond Y.K. Lau, Zhen Wang, Stephen Paul Smolley, Least Squares Generative Adversarial Networks, The IEEE International Conference on Computer Vision (ICCV), pp. 2794~2802, Oct., 2017