

# Generative Adversarial Networks 의 응용 현황

김동욱, 김세송, 정승원  
동국대학교 멀티미디어공학과  
e-mail : kimdongwook@dongguk.edu

## Applications of Generative Adversarial Networks

Dong-Wook Kim, Sesong Kim, and Seung-Won Jung  
Dept. of Multimedia Engineering, Dongguk University

### 요 약

Generative adversarial networks (GAN)에 대한 간략하게 설명하고, MNIST (숫자 손 글씨 데이터 셋)를 이용한 간단한 실험을 통해 GAN 구조 구조의 이해를 돕는다. 그리고 GAN 이 어떻게 응용이 되고있는지 다양한 논문들을 통해 살펴본다. 본 고에서는 GAN 논문들을 크게 이미지 스타일 변경, 3D 오브젝트 추정, 손상된 이미지 복원, 언어의 시각화, 기타 등으로 분류하였다.

### 1. 서론

“내가 다시 창조할 수 없는 것은 내가 이해한 것이 아니다.” - 물리학자 리처드 파인만

기존의 인공지능경망을 학습시키는 방법으로는 개와 고양이를 구분하고, 음성을 인식을 할 수는 있었지만 개를 직접 그리거나 음성을 만들지는 못했다. 그런데 GAN 은 이미지 등을 직접 만들 수 있다. 기존에 오토 인코더(Autoencoder)나 RBM(Restricted Boltzmann machine) 등이 비지도학습을 통해 이미지를 생성할 수 있었지만 GAN 은 기존의 기술보다 큰 성과를 보여 주고있다. GAN 은 몬트리올 대학의 Ian Goodfellow 가 2014 년에 NIPS 에 제출한 논문이다. 특히 Nips 2016 에서 가장 뜨거운 키워드로 GAN 이 주목되면서 큰 이슈가 되었다. 페이스북의 Yann LeCun 이 지난 10 년간 있었던 머신 러닝 연구 중 가장 재미있는 아이디어라고 평했던 GAN 에 대해 본 고에서는 그 원리에 대한 간단한 설명과 현재 어떻게 응용되고 연구되고 있는지 서술하고자 한다.

### 2. 본론

GAN 은 이름 그대로 경쟁적으로 무엇인가 생성하는 네트워크이다. GAN 은 두개의 네트워크로 구성된다. 하나는 무엇인가(이미지 등)을 생성하는 네트워크이다. Generator 라고 한다. 다른 하나는 생성된 무엇인가(이미지 등)가 제대로 생성되었는지 판별하는 네트워크이다. Discriminator 라고 한다. Goodfellow 의 GAN 논문에서는 Generator 을 위조지폐 범에 비유하고, Discriminator 는 위조지폐 범을 잡는 경찰에 비유하였다. Generator 은 최대한 진짜 같은 위조지폐를 만든다. 반면 Discriminator 은 지폐가 진짜인지 가짜

인지 판별한다. 이때 위조지폐 범과 경찰이 서로 경쟁적으로 학습을 하게 되면 최종적으로 위조지폐범은 진짜 같은 위조 지폐를 만들 수 있게 된다. 그림 (1) 은 GAN 에 대한 간단한 실험 결과이다. MNIST 를 학습을 시켜 Generator 가 진짜 같은 손 글씨를 생성할 수 있게 하였다. MNIST 는 구조가 간단하여 간단한 fully connected layer 만 사용해도 학습이 가능하다. 그림 (1)은 Generator 와 Discriminator 에 대한 예시이다. (a)는 Generator 이다. Input layer 에 크기가 10 이 Noise 를 넣어주었다. Hidden Layer 의 크기는 256 으로 해주었다. 결과로 28 \* 28 크기의 Fake image 가 생성된다. (b)는 Discriminator 이다. Input layer 에 28 \* 28 크기의 Fake image 나 Real image 를 넣어주었다. Hidden layer 의 크기는 Generator 와 같이 256 으로 해주었다. 입력이 Fake image 면 0 을, 입력이 Real image 면 1 을 출력하도록 학습한다. (c) 는 MNIST 60000 장 이미지에 대해 200 번 학습 시킨 후 생성한 결과 이미지 10 개이다.

GAN 자체의 이론은 굉장히 단순하지만, Generator 와 Discriminator 의 구조를 CNN 과 Batch normalization 등을 사용하여 더욱 복잡하게 만들 수 있다. 그림 (2)는 DCGAN 을 사용하여 생성한 얼굴 이미지들이다. DCGAN 은 기본적인 구조는 위의 그림 (1)의 이미지와 같지만 Generator 와 Discriminator 의 구조가 좀더 깊고 복잡하다. GAN 은 얼굴 외에도 학습의 대상과 방법에 따라 다양한 결과물을 생성할 수 있다. GAN 을 어떻게 하면 더 효과적으로 학습을 시켜서 더 완벽한 결과를 만들 수 있을 지에 대한 연구도 중요하지만 응용을 어떻게 할 것인지도 그만큼 중요하다. 아래에서는 GAN 이 지금까지 어떤 식으로 응용되고 있는지 서술한다.

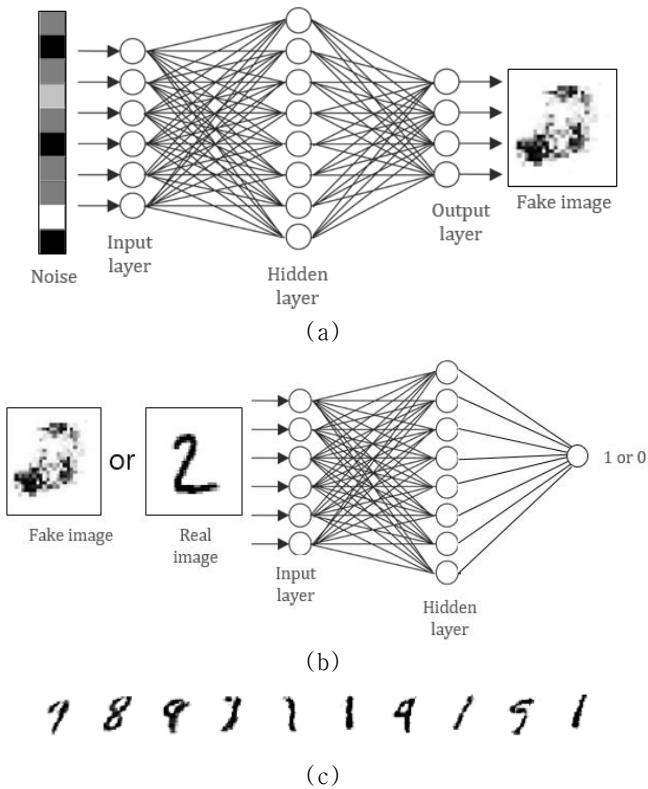


그림 (1) (a) Generator, (b) Discriminator, (c) 결과 이미지



그림 (2) DCGAN 을 사용하여 생성한 얼굴 이미지

### 2.1. 이미지의 스타일 변경

GAN 을 이용하여 입력 이미지 스타일을 바꾸는 연구는 GAN 이 이슈가 되는데 큰 역할을 한 분야이다.[1] 입력한 이미지에 추가로 다른 이미지를 넣으면 추가 이미지의 스타일대로 입력 이미지를 바꿔준다. 이와 비슷하게, 흑백 이미지를 넣으면 칼라 이미지로 색칠 해주거나[2], 스케치한 그림을 넣으면 색칠을 해주는 것[3], 실제 이미지를 캐릭터처럼 바꿔주는 것 또는 반대로 캐릭터 이미지를 실제 이미지처럼 바꿔주는 것 [4] 등이 있다.

### 2.2. 3D 오브젝트 추정

2D 또는 2.5D 이미지를 토대로 3D 오브젝트의 모양을 추정하는 것도 가능하다. 2D 또는 2.5D 이미지의 가구를 입력으로 받아 3D 의 모델로 만들어주거나[5], 손의 일부를 보고 손가락 전체 모양을 예측하는 것 [6], 또는 사람 이미지에서 뼈대 모양을 추정해서 팔과 다리의 위치를 추정하는 것[7] 등이 있다.

### 2.3. 손상된 이미지 복원

이미지의 손상된 부분을 복원하는 것은 없는 부분을 생성하는 것이기 때문에 GAN 을 이용하기에 적합하다. 저화질의 이미지를 고화질로 바꿔주는 것[8], jpeg 등의 압축을 통해 생긴 손상을 복원하는 것[9, 10] 또는 비가 내리는 풍경 이미지에서 비를 제거해주는 것[11] 등이 있다.

### 2.4. 언어의 시각화

문장을 읽고 문장에 알맞은 상황을 그림으로 그려주는 것[12] 또는 그림으로부터 상황을 설명하는 문장으로 만들어 주는 것[13]이 있다.

### 2.5. 기타

입력 이미지에서 Saliency map 을 추정하는 것이 있다[14]. 두개의 비슷한 이미지를 아주 그럴 듯 하게 합성해 주는 것이 있다[15]. DCGAN 이 얼굴을 생성하듯 지형 데이터를 생성한다[16]. CT 촬영 이미지를 보정하거나 이미지 안의 암 등을 찾는다 [17]. 아주 짧은 동영상상을 생성하거나[18], 비디오의 다음 프레임을 예측해서 생성한다[19].

## 3. 결론

지금까지 GAN 이 무엇인지, 어떻게 응용되고 있는지를 살펴보았다. 본 고를 통하여 GAN 이 다양한 분야에서 성과를 낼 수 있기를 기대한다.

## 참고문헌

- [1] Li, C., & Wand, M. (2016, October). Precomputed real-time texture synthesis with markovian generative adversarial networks. In European Conference on Computer Vision (pp. 702-716). Springer International Publishing.
- [2] Cao, Y., Zhou, Z., Zhang, W., & Yu, Y. (2017). Unsupervised Diverse Colorization via Generative Adversarial Networks. arXiv preprint arXiv:1702.06674.
- [3] Liu, Y., Qin, Z., Luo, Z., & Wang, H. (2017). Auto-painter: Cartoon Image Generation from Sketch by Using Conditional Generative Adversarial Networks. arXiv preprint arXiv:1705.01908.
- [4] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint arXiv:1703.10593.
- [5] Yang, B., Wen, H., Wang, S., Clark, R., Markham, A., & Trigoni, N. (2017). 3D Object Reconstruction from a Single Depth View with Adversarial Learning. arXiv preprint arXiv:1708.07969.
- [6] Wan, C., Probst, T., Van Gool, L., & Yao, A. Crossing Nets: Combining GANs and VAEs with a Shared Latent Space for Hand Pose Estimation.
- [7] Chen, Y., Shen, C., Wei, X. S., Liu, L., & Yang, J. (2017). Adversarial PoseNet: A Structure-aware Convolutional Network for Human Pose Estimation.

- arXiv preprint arXiv:1705.00389.
- [8] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2016). Photo-realistic single image super-resolution using a generative adversarial network. arXiv preprint arXiv:1609.04802.
  - [9] Galteri, L., Seidenari, L., Bertini, M., & Del Bimbo, A. (2017). Deep Generative Adversarial Compression Artifact Removal. arXiv preprint arXiv:1704.02518.
  - [10] Guo, J., & Chao, H. (2016). One-to-Many Network for Visually Pleasing Compression Artifacts Reduction. arXiv preprint arXiv:1611.04994.
  - [11] Zhang, H., Sindagi, V., & Patel, V. M. (2017). Image De-raining Using a Conditional Generative Adversarial Network. arXiv preprint arXiv:1701.05957.
  - [12] Zhang, H., Xu, T., Li, H., Zhang, S., Huang, X., Wang, X., & Metaxas, D. (2016). Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. arXiv preprint arXiv:1612.03242.
  - [13] Liang, X., Hu, Z., Zhang, H., Gan, C., & Xing, E. P. (2017). Recurrent Topic-Transition GAN for Visual Paragraph Generation. arXiv preprint arXiv:1703.07022.
  - [14] Pan, H., & Jiang, H. (2017). Supervised Adversarial Networks for Image Saliency Detection. arXiv preprint arXiv:1704.07242.
  - [15] Wu, H., Zheng, S., Zhang, J., and Huang, K. 2017. Gp-gan: Towards realistic high-resolution image blending. arXiv:1703.07195
  - [16] Beckham, C., & Pal, C. (2017). A step towards procedural terrain generation with GANs. arXiv preprint arXiv:1707.03383.
  - [17] Yi, X., & Babyn, P. (2017). Sharpness-aware Low dose CT denoising using conditional generative adversarial network. arXiv preprint arXiv:1708.06453.
  - [18] Tulyakov, S., Liu, M. Y., Yang, X., & Kautz, J. (2017). MoCoGAN: Decomposing Motion and Content for Video Generation. *arXiv preprint arXiv:1707.04993*.
  - [19] Liang, X., Lee, L., Dai, W., & Xing, E. P. (2017). Dual Motion GAN for Future-Flow Embedded Video Prediction. arXiv preprint arXiv:1708.00284.