

<심사용 논문>

훈련 안정성을 위한 병렬 구조의 생성자 기반 적대적 생성 네트워크

*박 범 찬, **도 인 실, *양 단 아

*한국성서대학교 컴퓨터소프트웨어학과 학생

**이화여자대학교 사이버보안학과 부교수

*한국성서대학교 컴퓨터소프트웨어학과 조교수

Generative Adversarial Networks based on Parallel Structured Generators for Training Stability

*Bumchan Park, **Inshil Doh, *Dana Yang

*Undergraduate Student, Dept of Computer Software, Korea Bible University,
Republic of Korea

**Associate Professor, Dept of Cyber Security, Ewha Womans University,
Republic of Korea

*Assistant Professor, Dept of Computer Software, Korea Bible University,
Republic of Korea

● 투고분야 : 인공지능 / 머신러닝 / 딥러닝

● Corresponding Author : Dana Yang

- Address : (01757) Korea Bible University, 32, Dongil-ro 214-gil, Nowon-gu,
Seoul, Republic of Korea

- TEL : +82-2-950-5487

- FAX : +82-70-4275-0167

- E-mail : dana1112@bible.ac.kr

● This work was supported by the National Research Foundation of Korea(NRF)
grant funded by the Korea government (MSIT) (NRF-2023R1A2C1005712)

훈련 안정성을 위한 병렬 구조의 생성자 기반 적대적 생성 네트워크

Generative Adversarial Networks based on Parallel Structured Generators for Training Stability

Abstract

Over the past few years, Generative Adversarial Network(GAN) has experienced significant growth in various applications as a generative model. However, common challenges in GANs, such as overfitting, instability in the training process, convergence issues, and mode collapse, need to be addressed. To mitigate these problems, this paper proposes a novel GAN model that applies dual-parallelized generators. This study designs new methodologies by inputting three data to the discriminator and updating the average of the loss values. Experimental results indicate that the proposed model demonstrates an ideal convergence graph compared to the traditional model, reducing the loss gap between the generator and discriminator by approximately 40%. The results also show an enhancement in the quality of generated data. This demonstrates that the model achieves stability during the learning process and effectively overcomes the aforementioned challenges.

Keywords: Parallel Structure Generator, Average Loss, SNGAN, Training Stability

1. 서론

적대적 생성 신경망(Generative Adversarial Networks, GAN)은 최근 몇 년간 생성 모델로서의 이미지 생성, 영상 생성, 텍스트 생성 등의 생성 분야를 통해 예술, 의료, 자율주행 차량, 게임 개발 등 다양한 분야에 응용되고 있다. GAN은 생성자(Generator)와 판별자(Discriminator) 두 신경망을 경쟁시키는 구조로, 생성자는 실제 데이터와 구별하기 어려운 거짓 데이터를 생성하려고 하며 판별자는 이를 실제 데이터와 구분하는 역할을 한다[1]. 이러한 상호작용을 통해 학습이 지속됨에 따라서 생성 모델은 더 좋은 성능을 보일 수 있다. GAN의 가장 핵심은 실제와 구분하기 어려운 데이터를 생성하기에 용이하다는 점이다[2]. 생성자와 판별자 간의 경쟁을 통해 생성자가 생성하는 데이터의 성능이 점점 좋아지고 이에 따라 다양한 유형의 데이터에 적용할 수 있게 된다. 특히 의료 분야에서 이러한 데이터 생성이 널리 사용된다[3][4]. GAN이 높은 성능을 가진다면 질병의 초기 징후를 탐지하는 잠재력을 높이며 의료 영상 해석을 개선하고 진단을 보조하는 데 도움을 준다[5][6]. 따라서 GAN의 성능을 높이기 위해서는 생성된 데이터의 품질 향상을 유도하는 것이 중요하다[7].

하지만 일반적으로 GAN에서 판별 모델의 분류 성능은 주어진 학습 데이터의 양에 비례하기 때문에 만약 모델의 규모에 비해 적은 양의 학습 데이터를 사용한다면 과적합(Overfitting)이 발생할 위험이 있다. 또한 두 개의 신경망이 서로 상호 경쟁을 거치는 과정에서 모델이 수렴하기까지 불

안정한 상태를 보이는 경향이 있다. 생성자와 판별자 간의 경쟁 학습이 균형을 이루지 못할 때 불안정하게 학습되거나 한쪽으로 편향될 수 있으며, 이는 모드 붕괴(Mode Collapse)나 수렴 문제로 이어질 수 있다[8]. 특히 이러한 문제는 의료 분야에서 매우 치명적이다. 모델의 불안정성으로 인해 실제 임상 환경에서의 모델 성능이 저하될 수 있으며 의료진의 판단에 혼란을 주거나 잘못된 진단을 유도할 위험이 있다[9]. 결론적으로 GAN에서는 빠른 학습과 고품질 데이터 생성에 대한 부분들도 중요하지만 과적합을 방지하면서 생성자와 판별자 간의 균등한 학습을 위한 효율적이고 안정적인 학습이 먼저 선행되어야 한다[10].

본 논문에서는 동일한 생성 모델 두 개를 병렬 구조로 정의한 Parallel Structured Generators로 정의한다. 해당 구조에서 각 생성자 모델은 서로 독립되지만 유사한 거짓 데이터를 생성한다. 판별자는 이러한 두 개 데이터와 실제 데이터까지 총 세 개의 데이터 판별을 통해 두 생성 모델의 평균 손실을 추가한다. 이러한 방식을 통해 훈련 과정의 불안정성을 방지하면서도 기존의 방법론들에 비해 품질 개선까지 이루는 방안을 제안한다. 본 논문의 유효성을 평가하기 위해 기존의 기술들과의 손실(Loss) 및 정확도(Accuracy)에 대한 지표와 생성한 이미지를 비교하였으며, 실험 결과 제안한 방법이 기존의 방법론에 비하여 성능이 우수함을 증명한다.

본 논문의 구성은 다음과 같다. 2장에서 본 논문과 관련된 연구인 SNGAN(Spectral Normalization Generative Adversarial Networks)에 대하여 정의하며 3장에서 본 연구에서 제안하는 기술을 설명한다. 4장에서는 제안한 기술로 실험한 결과를 기존의 기술들과 객관적인 지표로 비교하여 성능을 평가하고 5장에서 결론을 맺는다.

2. Spectral Normalization GAN(SNGAN)

GAN은 본질적으로 두 개의 신경망 즉, 생성자와 판별자로 구성된 경쟁적 학습을 기반으로 동작한다. 생성자는 실제 데이터와 거짓 데이터를 생성하기 위해, 판별자는 생성자가 생성하는 거짓 데이터를 판별하기 위해 서로 경쟁하며 학습을 한다. 이러한 구조는 다양한 분야에서 효과적으로 데이터를 생성하는데 사용되어 왔지만 훈련 과정의 불안정성과 모드 붕괴, 과적합 문제가 지속적으로 제기되어 왔다[11][12].

앞선 문제를 해결하기 위해 Miyato는 2018년에 판별자 네트워크의 각 레이어에 스펙트럴 정규화(Spectral Normalization, SN)를 적용한 SNGAN을 제안하였다[13]. 이 접근법은 기존의 GAN 구조를 기반으로 판별자의 각 레이어에 SN을 적용하며 각 레이어의 가중치 행렬의 스펙트럼을 활용하여 행렬의 가장 큰 특이값을 사용하여 가중치를 정규화한다. Fig. 1을 통해 SNGAN의 구조를 살펴보면 생성자(Gen)의 구조는 전형적인 GAN의 생성자와 유사하다. 랜덤한 노이즈 벡터(z)를 입력으로 받아 이를 업 샘플링(up-sampling)하고 합성곱(convolution) 연산을 거치며 최종적으로 데이터의 분포와 거짓 데이터(Gen(z))를 출력한다. SNGAN에서 x 는 실제 데이터를 나타내며 거짓 데이터와 함께 판별자에 의해 진위여부가 판별된다. 이때 판별자는 SN을 통해 가중치가 적용된다. 판별자 함수의 립시츠 상수(Lipschitz constant)를 제한함으로써 판별자는 더 안정적으로 기울기를 전파하며 학습 동안 발생할 수 있는 기울기 소실(Gradient Vanishing)을 방지한다[13]. 판별자는 입력 데이터의 실제(real)와 거짓(fake)을 나타내는 이진 출력을 통해 손실 함수(Loss)를 거친다. Gen(z)을 실제 데이터로 판단할 경우 거짓(False)이 되어 판별자에 손실값을 갱신하고 역전파 과정을 거친다. x 를 real로 판단할 경우 참(True)이 되어 Gen에 손실값을 갱신하고 역전파를 거치며 성능 개선을 한다.

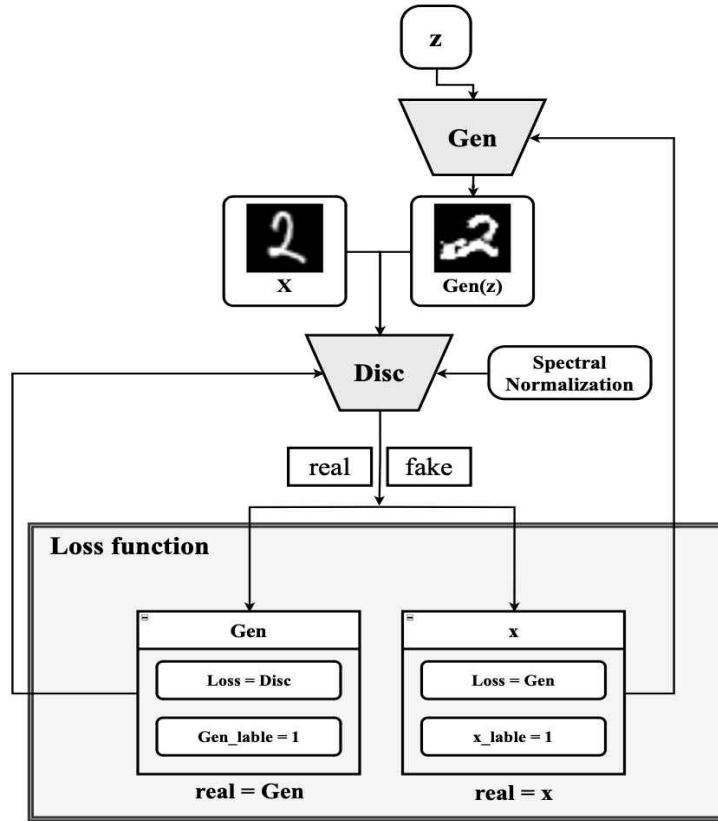


Fig. 1. Overview for Spectral Normalization GAN.

SNGAN의 핵심적인 요소인 SN은 각 합성곱 또는 완전 연결 레이어의 가중치에 대한 스펙트럴 노름(Spectral Norm)을 계산함으로써 그 가중치의 립시츠 상수를 제한한다. 해당 접근 방식은 이론적으로 판별자의 무한한 변화를 방지하고 결과적으로 GAN의 훈련을 더 안정적으로 만들어준다. 하지만 이러한 스펙트럴 연산은 추가적인 자원을 요구하며 특히 대규모 신경망에서 그 부담이 증가한다[14]. SN의 실행은 가중치 행렬의 가장 큰 특이값을 반복적으로 계산하는 과정을 포함한다. 이는 파워 이터레이션 알고리즘(Power Iteration Algorithm) 또는 그 변형을 사용하여 수행되며 네트워크의 각 레이어에서 이러한 연산을 수행해야 하므로 계산 비용이 기하급수적으로 증가한다[15]. 대규모의 데이터 셋과 복잡한 모델에 대해 신속한 훈련을 요구하는 경우 이러한 계산 오버헤드는 확장성과 효율성을 제한하는 원인이 된다. 뿐만 아니라 SNGAN의 구현은 이러한 정규화 계산이 훈련 과정에서 반복적으로 수행되어야 한다는 점을 고려할 때 연산량에 한계가 있다. 또한 모델의 크기가 클수록 Spectral Norm을 계산하기 위해 더 많은 Power Iteration 단계가 필요하게 되고 이는 전체 훈련 시간에 영향을 미친다.

결론적으로, SNGAN에서 SN을 활용하는 것은 훈련의 안정성을 보장하지만 계산 비용이라는 문제점을 동반한다. 학습의 불안정성과 과적합 문제 또한 완전히 개선되었다고 보기엔 무리가 있다. 이러한 종합적인 문제를 해결하기 위해 지속적인 연구가 필요하다. 3장에서는 이를 보완하기 위해 본 논문에서 제안한 Parallel Structured Generator GAN에 대해 설명한다.

3. Parallel Structured Generator GAN(PSG-GAN)

본 논문에서는 과적합을 방지하기 위한 방법론으로 SNGAN 알고리즘의 한계점에 따라서 SN 없이 훈련의 불안정성과 과적합 문제를 해결하기 위한 방법을 제안한다. 해당 알고리즘은 생성자 병렬처리를 통해 생성한 두 데이터를 판별자의 Loss에 평균을 내는 방식으로 GAN에서 발생하는 문제를 보완할 수 있다.

3.1 Architecture

앞서 언급했듯이 GAN에서 발생하는 여러 문제점들을 해결하기 위해 본 논문에서는 두 생성자 네트워크 병렬 처리를 제안한다. 이러한 GAN의 방법론을 PSG-GAN으로 명시한다. PSG가 제안하는 시스템 구조를 Fig. 2를 통해 살펴보면 성능이 같은 독립된 생성자를 두 개 형성한다(Gen1, Gen2). 두 생성자는 서로 다른 독립된 거짓 데이터(Gen1(z), Gen2(z))를 생성한다. 이렇게 생성된 두 데이터는 실제 데이터 x와 함께 판별자(Disc)에 들어가게 된다. 기존 GAN의 판별자는 두 데이터를 받아 real과 fake에 대한 이진 분류를 하는 반면 제안하는 시스템에서는 두 생성자가 출력하는 거짓 데이터와 실제 데이터 x까지 총 세 개의 데이터가 판별자의 분류기에 진입한다. 세 데이터 중 판별자가 실제로 판별하는 데이터는 하나뿐이고 나머지 데이터는 분류기가 거짓으로 분류한다.

제안한 PSG 방식을 사용한다면 스펙트럴 정규화를 하지 않더라도 큰 계산 비용없이 판별자가 생성자보다 훨씬 빨리 학습하는 문제점을 개선할 수 있다. 따라서 해당 방식은 앞서 설명한 GAN의 고질적인 문제점인 모드 붕괴 현상과 과적합 문제를 해결하기에 매우 효과적이다.

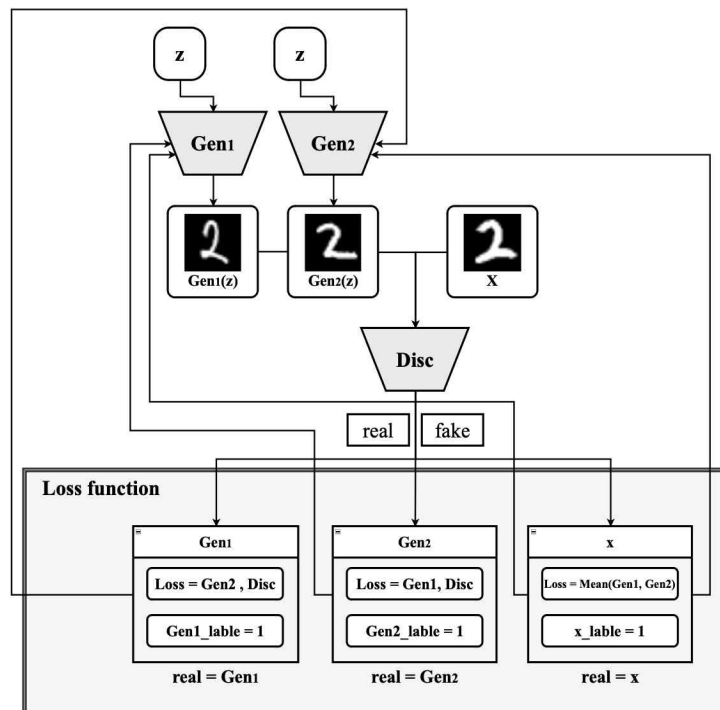


Fig. 2. Overview for Parallel Structure Generator GAN.

3.2 Average Loss

PSG는 판별자에 진입하는 데이터의 총 개수가 3개로 기존의 GAN과는 다르기 때문에 손실을 계산하는 방식이 기존 GAN의 방식과는 다르다. PSG 모델의 방법론으로 인해 판별자 함수에 앞서 언급한 세 종류의 데이터를 넘김에 따라 거짓으로 분류된 데이터가 두 개 생성되며 손실값 갱신도 총 두 번 발생한다. 판별자 분류에 따른 손실값 갱신에 대한 경우의 수는 다음과 같다.

Algorithm 1: Update loss value

Data: Discriminator classifications
Result: Updated loss values for *Gen1* and *Gen2*

```

if the discriminator classifies Gen1 as real then
    | Update the loss value for Gen2 and Disc
else
    | if the discriminator classifies Gen2 as real then
    | | Update the loss value for Gen1 and Disc
    | else
    | | if the discriminator classifies x as real then
    | | | Update the average loss value for Gen1 and Gen2

```

이러한 평균 손실 계산 방식을 통해 Gen1과 Gen2가 모두 균형적인 학습을 유도하여 손실 함수의 변동성을 줄임으로써 학습의 안정성을 개선한다. 이는 손실 함수가 각 학습 단계에서 발생하는 오차를 더 일관되게 반영함으로써 이루어진다. 안정화된 학습 과정은 모델이 데이터의 다양한 특성을 효과적으로 학습하고 이를 바탕으로 더 정확한 예측을 수행할 수 있게 한다. 결과적으로, 이는 생성자와 판별자 간의 수렴하는 속도를 증진시켜 생성 데이터가 높은 품질의 출력물을 생성하는 데 기여한다.

이와 같이, PSG 모델은 두 개의 독립된 생성자를 사용함으로써 판별자의 과도한 학습을 조절하고 모드 붕괴 및 과적합 문제를 완화한다. 또한 평균 손실 계산 방식은 각 생성자가 발생시키는 오차를 균형 있게 반영하여 손실 함수의 변동성을 줄이고 이를 통해 전체 모델의 학습 과정을 안정화시킨다. 이러한 접근은 생성자와 판별자 간의 학습 속도 차이를 줄이면서 높은 품질의 이미지를 생성할 수 있다.

4. 실험 결과

4장에서는 본 논문에서 제안하는 기술인 PSG-GAN과 SNGAN 방법론 간의 다양한 평가 지표와 성능 비교를 통해서 PSG-GAN이 학습의 불안정성과 과적합에 유리하고 품질이 향상됨을 증명한다.

4.1 실험 환경

본 논문에서는 Table. 1과 같이 SN과 PSG 모델의 효율적인 훈련과 정밀한 시뮬레이션을 위해 Windows 10 운영 체제와 AMD Ryzen CPU를 사용한다. 또한, 복잡한 그래픽 처리와 딥러닝 연산에 필요한 GPU와 대규모 데이터셋 처리를 위한 RAM이 시뮬레이션 환경으로 포함된다.

Table. 1. information for simulation

OS	Windows 10 Pro
CPU	AMD Ryzen Threadripper 3960X 24-Core
GPU	NVIDIA GeForce RTX 3080
RAM	128GB

4.2 에포크 당 손실값 비교

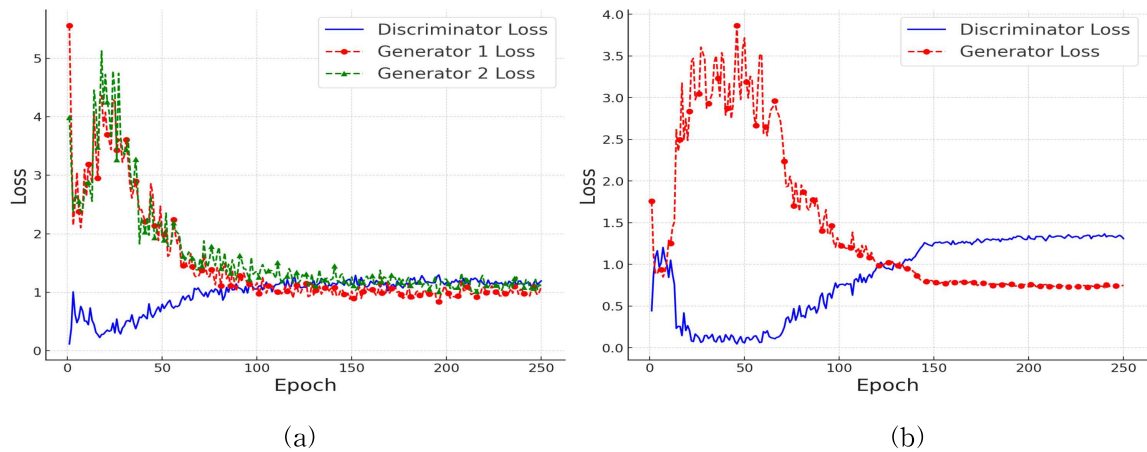


Fig. 3. Loss comparison graph per epoch (a) PSG-GAN, (b) SNGAN

Fig. 3 그래프를 통해서 PSG 모델이 SN 모델에 비해서 학습 과정 동안 손실값이 크게 증폭되거나 급격한 변화 없이 안정적으로 감소하는 것을 볼 수 있다. 이는 학습이 안정적으로 진행되고 있음을 의미하며, 과적합의 위험을 감소시키는 것을 나타낸다. 또한 Fig. 3의 (b)의 SNGAN은 에포크가 약 125일 때 수렴하는 반면, Fig. 3의 (a)의 PSG-GAN은 에포크가 약 75일 때 먼저 수렴하는 것을 볼 수 있다. 이를 통해 PSG-GAN이 약 10% 먼저 수렴했으며, 수렴 후에도 생성자와 판별자 간의 균형에서도 SN 알고리즘에 비해서 두 개의 생성자와 판별자의 손실값이 서로 균형을 이루며 학습하는 것을 볼 수 있다. 이는 두 개의 생성자가 상호 작용하며 학습이 원활하게 진행되고 있음을 나타낸다. 생성자와 판별자 간의 균형은 GAN의 안정적인 학습에 중요하며 과적합 방지에도 우수함을 보였다.

4.3 에포크 당 정확도 비교

Fig. 4는 PSG와 SN 모델에서 각 판별자의 에포크 당 정확도 그래프를 나타낸다. 기존의 GAN에서는 학습이 잘 이루어지고 있다면 판별자의 정확도는 0.5 근처에 수렴해야 한다. 반면에 PSG-GAN 알고리즘은 판별자가 실제와 거짓을 구별하는 데이터가 세 종류 들어가기 때문에 0.33 근처에 수렴해야 한다. 수렴이 시작되는 에포크의 지점이 빠를수록 유사도가 높은 거짓 데이터를 빠른 속도로 만들 수 있다. 그래프를 살펴보면 PSG-GAN Convergence선이 SNGAN Convergence선에 비해 10% 먼저 수렴하게 되는 결과를 볼 수 있다. 이는 데이터 생성 속도와 학습 안정화에 효과적임을 보여준다.

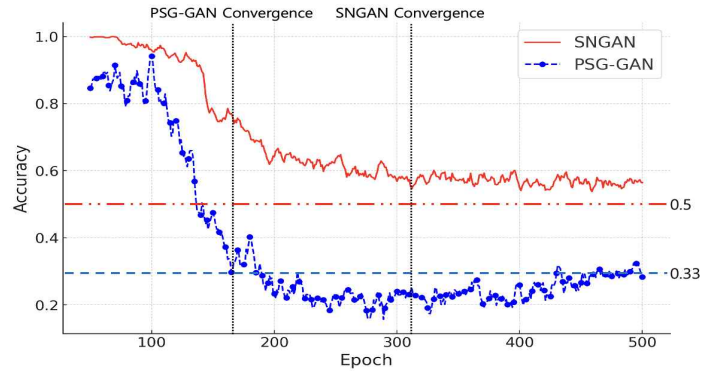


Fig. 4. Accuracy comparison graph per epoch

4.4 MNIST 이미지 비교

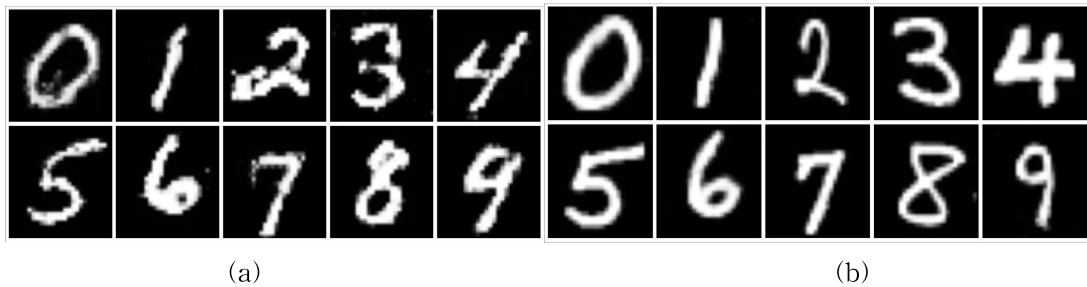


Fig. 5. Comparison MNIST Dataset (a) SNGAN, (b) PSG-GAN

Fig. 5의 생성된 이미지를 통해 SNGAN과 PSG-GAN의 이미지 품질을 알 수 있다. SNGAN으로 생성된 이미지들은 상대적으로 경계가 불분명하고 일관성이 떨어지는 경향을 보인다. 반면 PSG-GAN이 생성한 숫자 이미지는 구조적인 명확성과 더불어 연속적인 모서리 및 형태의 선명도가 두드러졌으며 전반적인 이미지의 질적인 측면에서도 높게 평가할 수 있다. 따라서 제안된 PSG-GAN이 SNGAN에 비해 숫자 이미지 생성에 있어서 더 나은 품질을 달성하였음을 보인다.

5. 결론

GAN이 가지고 있는 학습의 안정성 문제는 지난 몇 년간 지속적으로 지적되었다. 본 논문에서는 동일한 성능의 두 생성자가 병렬 처리를 통해서 학습할 때의 동등한 평균 손실 계산을 통해 생성자와 판별자의 균등한 학습을 유도하여 효율적이고 안정적인 학습을 가능하게 한다. 이러한 알고리즘은 기존의 통용되는 방법론들과 비교했을 때 학습 과정에서의 불안정성과 과적합 방지에 훨씬 효율적이며 생성 데이터의 품질도 향상되었음을 실험을 통해 확인하였다. 이러한 새로운 패러다임을 통해 앞으로 정확하고 세밀한 작업이 필요한 의료분야, 빠른 객체의 변화에 대한 인식을 필요로 하는 자율주행 차량 등에서 적절한 방향성을 제시할 수 있을 것이다. 향후 연구로는 학습의 수렴속도에 따른 데이터의 품질과 수렴의 균형을 적절히 고려하여 더욱 안정적인 시스템을 구축하고자 한다.

REFERENCE

- [1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. "Generative Adversarial Nets," *Advances in Neural Information Processing Systems*, Vol. 2014, pp. 2672-2680, 2014.
- [2] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. "Generative Adversarial Networks: An Overview," *IEEE Signal Processing Magazine*, Vol. 35, no. 1, pp. 53-65, 2018.
- [3] Yi, X., Walia, E., and Babyn, P. "Generative Adversarial Network in Medical Imaging: A Review," *Medical Image Analysis*, Vol. 58, no. 101552, 2019.
- [4] Kazemini, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N., Albarqouni, S., and Mukhopadhyay, A. "GANs for Medical Image Analysis," *Artificial Intelligence in Medicine*, Vol. 109, no. 101938, 2020.
- [5] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. "Improved Techniques for Training GANs," *Advances in Neural Information Processing Systems*, pp. 2234-2242, 2016.
- [6] Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., and Greenspan, H. "Synthetic Data Augmentation using GAN for Improved Liver Lesion Classification," In *IEEE 15th International Symposium on Biomedical Imaging*, pp. 289-293, 2018.
- [7] Alqahtani, H., Kavakli-Thorne, M., and Kumar, G. "GAN-based Synthetic Medical Image Augmentation for Increased CNN Performance in Liver Lesion Classification," *Computer Methods and Programs in Biomedicine*, Vol. 200, no. 105823, 2021.
- [8] Arora, S., and Zhang, Y. "Do GANs Actually Learn the Distribution? An Empirical Study," *arXiv preprint arXiv:1706.08224*, 2017.
- [9] Brock, A., Donahue, J., and Simonyan, K. "Large Scale GAN Training for High Fidelity Natural Image Synthesis," *arXiv preprint arXiv:1809.11096*, 2018.
- [10] Mirza, M., and Osindero, S. "Conditional Generative Adversarial Nets," *arXiv preprint arXiv:1411.1784*, 2014.
- [11] Arjovsky, M., Chintala, S., and Bottou, L. "Wasserstein GAN," *arXiv preprint arXiv:1701.07875*, 2017.
- [12] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. "Improved Training of Wasserstein GANs," *arXiv preprint arXiv:1704.00028*, 2017.
- [13] Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. "Spectral Normalization for Generative Adversarial Networks," *arXiv preprint arXiv:1802.05957*, 2018.
- [14] Kurach, K., Lucic, M., Zhai, X., Michalski, M., and Gelly, S. "A Large-Scale Study on Regularization and Normalization in GANs," *arXiv preprint arXiv:1807.04720*, 2019.
- [15] Miyato, T., and Koyama, M. "cGANs with Projection Discriminator," *arXiv preprint arXiv:1802.05637*, 2018.