

PEGAN : Training Generative Adversarial Networks with Pre-trained Ensemble Generator

(Abstract) 이 논문에서는 모드 붕괴 문제를 해결하기 위해 사전학습된 생성자를 가중치 기반 앙상블 기법을 사용해 결합한 새로운 생성적 적대 신경망 모델인 PEGAN(Pre-trained Ensemble Generative Adversarial Networks)을 제안한다. PEGAN은 단일 생성자가 특정 모드에만 집중하여 모드 붕괴가 발생하는 전통적인 GAN의 문제점을 해결하기 위하여 각 클래스별 모드를 사전학습한 여러 생성자들에 가중치를 곱하여 결합함으로써, 모델이 전체적인 데이터 모드를 잘 포괄하도록 하여 문제를 해결했다. MNIST 데이터 학습 실험결과, PEGAN은 고전적인 GAN모델들보다 생성 이미지의 품질과 다양성을 평가하는 평가지표인 Inception Score, Frechet Inception Distance 에서 매우 우수한 성능을 보였다.

1. 서론

오늘날, 생성적 적대 신경망(Generative Adversarial Networks, GAN)은 환경, 안전, 의학 등의 다양한 분야에서 우리 생활 속에 자리 잡은 생성 모델로, 그 활용 범위는 지속적으로 확대되고 있다[6]. GAN을 훈련하는 과정에서 발생하는 주요한 문제로, 생성 데이터의 다양성이 줄어드는 모드 붕괴(mode collapse) 현상이 존재한다. 이는 생성자의 학습 과정 중, 다양한 데이터를 생성하지 못하고, 특정 모드(mode)에 집중하는 현상으로, 생성 데이터의 다양성이 크게 감소해 모델의 성능이 낮아지는 문제이다[5]. 이러한 모드 붕괴 현상은 GAN 알고리즘이 풀어나가야 할 주요 과제 중 하나이다. 이에, 최근 10년간 모드 붕괴를 해소하기 위하여 판별자의 업데이트 과정을 확장하여 학습의 안정성을 높인 Unrolled GAN[3], 저해상도 이미지부터 점진적으로 고해상도 이미지를 점진적으로 생성함으로써 안정성을 높인 Progressive GAN[9]등의 방법이 제안되고 있다. 이런 GAN의 연구 흐름에 부응하여, 본 논문에서는 모드 붕괴 현상을 해결하기 위해 각 클래스 별로 사전학습(Pre-training)된 생성 모델을 앙상블(Ensemble)기법을 활용하여 결합하여 최종 모델을 구축하고, 이 모델을 미세조정(Fine-tuning)하여 모드 붕괴 현상을 해소하는 Pre-trained Ensemble Generative Adversarial Networks(PEGAN)을 제안하고자 한다.

2. 본문 내용

Table 1. PEGAN 학습루프

Algorithm 1. PEGAN training loop
Input : 훈련 데이터셋 $X = \{x_i\}_{i=1}^K$
Output : 학습된 생성자 $G_f(\cdot; \theta)$ 와 판별자 $D(\cdot; \phi)$
1: 초기화
• 사전 학습된 생성자 $\{G_k(\cdot; \phi_k)\}_{k=1}^K$ 의 파라미터 ϕ 를 변경 불가능하도록 불러와서 생성자 G 에 결합해 구축한다.
• 생성자 G 의 파라미터 θ 와 판별자 D 의 파라미터 ϕ 를 초기화한다.
2: $t \leftarrow 0$
3: while the objective not converge and $t < \max_iter$ do :
4: 훈련 데이터 X 에서 미니배치 X_t 를 샘플링한다.
5: 확률분포 P_z 에서 노이즈 벡터 $z \in \mathbb{R}^d$ 를 샘플링한다.
6: $z' = \text{Dense}(K \times d)(z)$
7: $z' = [z_1, z_2, z_3 \dots z_K]$, $\{z_k\}_{k=1}^K \in \mathbb{R}^d$
8: $\{g_k = G_k(z_k)\}_{k=1}^K$
9: $\{W_k = \sigma(\text{Dense}(1)(z_k))\}_{k=1}^K$, $\sigma = \text{sigmoid}(x)$
10: $\{g_{weighted,k} = W_k \times g_k\}_{k=1}^K$
11: $G_f(\cdot; \theta) = \text{Layers}(\text{Concatenate}(\{g_{weighted,k}\}_{k=1}^K))$
12: $D_{real} = D(X_t; \phi)$
13: $D_{fake} = D(z; \phi)$
14: $L_D = -\mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] - \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G_f(z)))]$
15: $L_{similar} = \mathbb{E}_{z_1, z_2 \sim p_z(z)}[\text{Sim}(G_f(z_1), G_f(z_2))]$
16: $L_G = -\mathbb{E}_{z \sim p_z(z)}[\log D(G_f(z))] + \lambda \cdot L_{similar}$
17: Update $D(\cdot; \phi)$: $\phi \leftarrow \phi - \eta_D \nabla_{\phi} L_D$
18: Update $W_k \leftarrow W_k - \eta_G \nabla_{W_k} L_{similar}$
19: Update $G_f(\cdot; \theta)$: $\theta \leftarrow \theta - \eta_G \nabla_{\theta} L_G$
20: $t \leftarrow t + 1$
21: end while

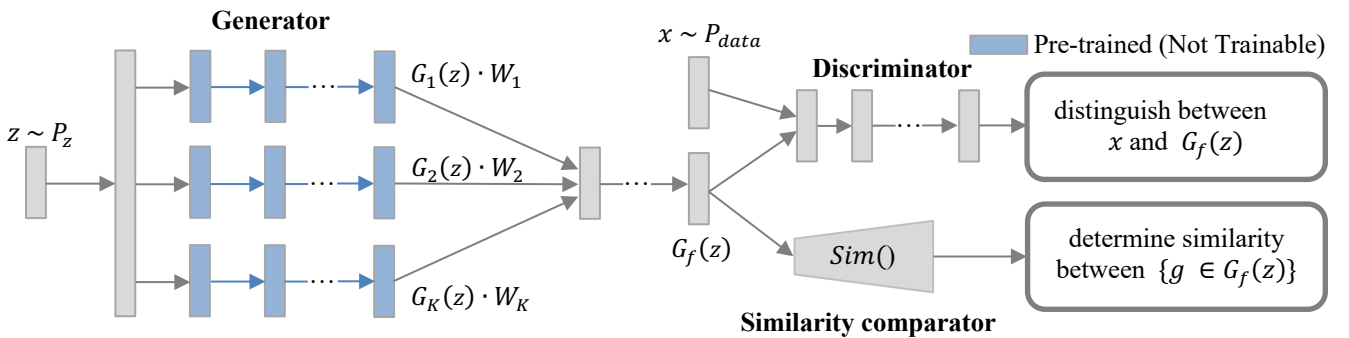


Fig. 1. PEGAN의 구조

PEGAN은 GAN의 모드 붕괴 현상을 해소하기 위하여 제안되었다. 따라서, PEGAN의 생성자는 Fig. 1과 Table.1에 설명된 것처럼 입력 노이즈 벡터 $z \in \mathbb{R}^d$ 를 받아들이며, 이 벡터는 Dense 레이어를 통해 $K \times d$ 차원의 벡터 z' 으로 확장된다. 이 확장된 벡터 z' 는 K 개의 개별 벡터 $\{z_k\}_{k=1}^K \in \mathbb{R}^d$ 로 분할되며, 각 벡터는 사전학습된 K 개의 생성자 $\{G_k(\cdot; \varphi_k)\}_{k=1}^K$ 에 입력된다. 이 사전 학습된 생성자들은 각각 다른 모드의 특성을 학습한 상태로 고정되어 있으며, 학습 중 업데이트되지 않는다. 각 생성자가 출력하는 벡터 $\{g_k\}_{k=1}^K$ 는 각각 가중치 $\{W_k\}_{k=1}^K$ 에 의해 곱해져 결합된 후, 최종 생성자 G_f 의 출력이 된다.

PEGAN의 이러한 구조는 단일 생성자가 데이터의 특정 모드만을 학습하는 기존의 GAN 구조에서 벗어나, 여러 생성자가 협력하여 전체적인 데이터의 모드를 잘 포괄하도록 한다. 가중치 $\{W_k\}_{k=1}^K$ 는 학습 과정에서 업데이트되며, 이를 통해 생성자 출력이 적절히 결합되고, 생성된 샘플의 다양성이 유지되도록 한다. 이러한 구조는 단일 생성자가 특정 모드에만 집중하여 모드 붕괴가 발생하는 전통적인 GAN의 문제점을 완화하는 데 기여한다.

PEGAN의 학습 과정에서 생성자 G 와 판별자 D 는 두 플레이어 간의 min-max 게임을 통해 최적화된다. PEGAN의 전체 목적 함수(1)는 정통적인 GAN[1]과 비슷하지만, 생성된 샘플 간의 유사성을 측정하는 마지막 항을 추가해 유사성을 줄이는 방향으로 학습되어 각 생성자의 출력에 균형을 잡는다. 이 과정을 통해, 모드 붕괴 문제를 해결하려고 한다.

$$\min G \max D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G_f(z)))] - \mathbb{E}_{z_1, z_2 \sim p_z(z)} [Sim(G_f(z_1), G_f(z_2))] \quad (1)$$

따라서, 다양한 데이터를 생성할 수 있도록 유사성 손실 $L_{similar}$ (2)을 도입하여 생성자의 가중치 $\{W_k\}_{k=1}^K$ 가 유사성 손실 함수 $L_{similar}$ 에 의한 그레디언트 업데이트 규칙(3)을 따르도록 구성한다.

$$L_{similar} = \mathbb{E}_{z_1, z_2 \sim p_z(z)} [Sim(G_f(z_1), G_f(z_2))] \quad (2)$$

$$W_k \leftarrow W_k - \eta_G \nabla_{W_k} L_{similar} \quad (3)$$

수식 (2)에서 유사성 함수 $Sim(G_1, G_2)$ 는 MSE(Mean Squared Error), SSIM(Structural Similarity Index), Cosine Similarity와 같은 다양한 방법으로 정의될 수 있다. 이를 통해 생성된 샘플 간의 다양성을 조정하여 PEGAN을 구성하면 모드붕괴 문제를 완화할 수 있다.

Table 2. MNIST 데이터셋 학습 실험 평가지표

Model	Inception Score [7]	Frechet Inception Distance [8]
GAN [1]	6.728	9.467
DCGAN [2]	8.674	7.215
PEGAN (MSE)	8.462	7.649
PEGAN (SSIM)	8.857	6.947
PEGAN (Cosine Similarity)	7.826	7.293
PEGAN (Wasserstein distance[4])	9.239	6.723

Table 2.는 150epoch, 100batch size, 0.0005 learning rate로 학습한 GAN 모델의 모델의 Inception Score(IS), Frechet Inception Distance(FID)를 나타낸 것이다. 실험 결과, PEGAN은 고전적인 GAN보다 다양한 이미지를 생성하며, $Sim(G_f(z_1), G_f(z_2))$ 함수를 SSIM과 Wasserstein distance로 유사도 기반 손실 함수를 구축했을 때 효과적으로 다양한 클래스의 이미지를 생성한다.

3. 결론

이 논문에서는 모드 붕괴 문제를 해결하기 위하여 사전학습된 여러 생성자를 가중치 기반 앙상블 기법을 통해 결합한 PEGAN 모델을 제안하였다. PEGAN은 기존 GAN 및 DCGAN 대비 IS, FID 측면에서 우수한 성능을 보였으며, 유사성 손실 함수로 SSIM과 Wasserstein distance를 활용하여 가중치를 업데이트할 경우 다양한 데이터를 높은 품질로 생성함을 확인하였다. 본 연구는 고전적인 GAN의 성능을 개선하여 모드 붕괴 문제를 해결하기 위한 새로운 방법을 제시하였으며, 향후 다양한 분야에서 PEGAN을 확장하고 응용할 수 있을 것으로 기대된다.

참고문헌

- [1] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. *Generative adversarial nets*. *arXiv preprint*, arXiv:1406.2661 [stat.ML], 1–9 (2014)
- [2] Radford, A., Metz, L., & Chintala, S. *Unsupervised representation learning with deep convolutional generative adversarial networks*. *arXiv preprint*, arXiv:1511.06434 [cs.LG] (2016)
- [3] Metz, L., Sohl-Dickstein, J., & Poole, B. *Unrolled generative adversarial networks*. *Proceedings of the International Conference on Learning Representations (ICLR)*, (2017) *arXiv preprint*, arXiv:1611.02163 [cs.LG]
- [4] Arjovsky, M., Chintala, S., & Bottou, L. *Wasserstein GAN*. *arXiv preprint*, arXiv:1701.07875 [stat.ML] (2017)
- [5] Kossale, Y., Airaj, M., & Darouichi, A. *Mode collapse in generative adversarial networks: An overview*. *2022 8th International Conference on Optimization and Applications (ICOA)*, 1–6. IEEE (2022)
<https://doi.org/10.1109/ICOA55659.2022.9934291>
- [6] 이용오. "고장 진단 및 예측에서 생성적 적대 신경망의 활용 방안". *한국신뢰성학회 2022 춘계학술대회 [초록집]*, 한국신뢰성학회 학술대회논문집, p. 69 (2022)
- [7] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. *Improved techniques for training GANs*. *Advances in Neural Information Processing Systems*, **29**, 2234–2242 (2016)
<https://arxiv.org/abs/1606.03498>
- [8] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. *GANs trained by a two time-scale update rule converge to a local Nash equilibrium*. *Advances in Neural Information Processing Systems*, **30**, 6626–6637 (2017)
- [9] Karras, T., Aila, T., Laine, S., & Lehtinen, J. *Progressive growing of GANs for improved quality, stability, and variation*. *arXiv preprint*, arXiv:1710.10196 [cs.NE] (2018)

PEGAN : Training Generative Adversarial Networks with Pre-trained Ensemble Generator

Hankuk Academy of Foreign Studies

Natural Science Track

MinseoKang

2024.12.

Email: minseo_0606@naver.com

Github: [minecode0606](https://github.com/minecode0606)