

특징 공간 기반 중심 제약 손실 함수를 활용한 GAN의 품질 및 다양성 향상

서상원, 박현철*

국립한국교통대학교

llfkstep9811@aut.ac.kr, *hc.park@ut.ac.kr

Feature Space-Based Center Constraint Loss for Improving Quality and Diversity in GANs

Seo Sang Won, Park Hyun Cheol*

Korea National University of Transportation Univ.

요약

본 논문은 적대적 생성 신경망(Generative Adversarial Networks, GAN)의 이미지 품질과 다양성을 동시에 향상시키기 위한 새로운 제약 기반 손실 함수를 제안한다. 제안하는 방법은 학습 데이터셋을 사전 학습된 신경망을 통해 특징 공간으로 임베딩한 뒤, 해당 공간에서의 중심점을 계산하고, 각 특징 벡터와 중심점 간의 평균 거리를 참조 거리로 정의한다. 이후 학습 과정에서 생성자가 생성한 이미지 또한 동일한 특징 공간에 임베딩되며, 생성 이미지의 특징 벡터가 중심점으로부터 참조 거리보다 벗어나는 경우, 이에 대한 패널티를 손실 함수에 추가한다. 이 제약은 생성자가 실제 데이터의 분포 범위 내에서 이미지를 생성하도록 유도함으로써, 품질과 다양성의 균형을 효과적으로 유지한다. CelebA 데이터셋을 기반으로 한 실험 결과, 제안하는 방법은 FID 및 LPIPS 지표에서 기존 방식 대비 우수한 성능을 보인다.

I. 서론

적대적 생성 신경망은 높은 품질의 이미지를 생성할 수 있는 대표적인 생성 모델로, 다양한 컴퓨터 비전 분야에서 활발히 활용되고 있다[1]. 그러나 GAN은 학습의 불안정성과 함께, 이미지의 다양성이 부족해지는 모드 붕괴 문제, 그리고 생성 이미지의 품질 저하와 같은 한계를 여전히 안고 있다.

이러한 문제를 해결하기 위한 기존 연구 중 하나로, 낮은 해상도에서 시작하여 점진적으로 고해상도로 확장해 나가며 훈련의 안정성과 이미지 품질을 향상시키는 점진적 성장 기법이 제안된 바 있다 [2]. 해당 방법은 일정 수준의 성능 향상을 보였으나, 네트워크 구조의 단계적 변경 및 해상도 전환을 위한 추가 로직 설계가 요구되어 모델의 구현 복잡도 및 계산 비용이 증가하는 한계를 가진다.

본 논문에서는 위와 같은 한계를 보완하고자, 특징 공간에서의 중심점 기반 제약 손실 함수를 새롭게 제안한다. 제안된 방법은 먼저 학습 데이터셋을 사전 학습된 VGG 네트워크를 통해 특징 공간으로 임베딩한 후, 해당 공간에서 데이터의 중심점과 각 특징 벡터 간의 평균 거리를 계산하여 이를 참조 거리로 정의한다. 이후 학습 과정에서 생성된 이미지 역시 동일한 특징 공간에 임베딩되며, 생성 이미지의 특징 벡터가 중심점으로부터 이 참조 거리보다 크게 벗어나는 경우, 이에 대한 패널티를 손실 항에 추가한다.

이러한 제약은 생성자가 실제 데이터 분포의 중심을 벗어나지 않는 범위 내에서 이미지를 생성하도록 유도함으로써, 이미지 품질 저하 없이 다양성 손실도 방지하는 효과를 기대할 수 있다. 제안한 방법은 구조 변경이나 별도의 네트워크 추가 없이도 손실 함수 수준에서 간단하게 적용 가능하다는 점에서 효율적인 대안이 될 수 있다.

II. 본론

2.1 특징 공간에서의 중심점과 참조거리 계산

사전 학습된 VGG 네트워크를 통해 학습 데이터셋의 각 이미지 x_i 를

특징 벡터 $f_i = \Phi(x_i) \in R^d$ 로 임베딩한다. 이후 전체 특징 벡터들의 평균 벡터를 중심점 c 로 정의한다.

$$c = \frac{1}{N} \sum_{i=1}^N f_i \quad (1)$$

그리고 각 특징 벡터들과 중심점 간의 평균 거리, 즉 참조 거리는 다음과 같이 정의된다:

$$d_{ref} = \frac{1}{N} \sum_{i=1}^N \|f_i - c\|_2 \quad (2)$$

중심점과 참조 거리는 학습 전에 계산 후 고정되며, 생성 이미지 평가 시 기준으로 사용된다.

2.2 생성 이미지에 대한 중심 제약 손실

학습 중 생성자가 생성한 샘플 $\hat{x} = G(z_j)$ 에 대해서도 동일한 방식으로 특징 벡터 $\hat{f}_j = \Phi(\hat{x}_j)$ 를 추출한다. 이때, 생성된 특징이 중심점으로부터 참조 거리보다 멀리 떨어진 경우, 손실 항을 통해 패널티를 부여한다.

$$L_{centroid} = \frac{1}{M} \sum_{j=1}^M \max(0, \|\hat{f}_j - c\|_2 - d_{ref}) \quad (3)$$

이 손실은 생성 이미지의 특징 벡터를 데이터 분포의 중심점을 기준으로 정의된 참조 거리 범위를 벗어나지 않도록 유도한다.

2.3 최종 생성자 손실 함수

기존의 적대적 손실과 함께, 본 논문에서 제안하는 중심 기반 제약 손실을 결합하여 생성자의 최종 손실은 다음과 같이 정의된다:

$$L_G = L_{adv} + \lambda_{centroid} \cdot L_{centroid} \quad (4)$$

생성자 손실에 새롭게 추가된 $L_{centroid}$ 를 최소화함으로써, 생성된 이미

지의 분포가 실제 데이터의 분포를 더욱 잘 근사하도록 유도할 수 있다.

Algorithm 1 Training

```

1: repeat
2:    $\hat{x} \leftarrow G(z)$ 
3:    $\mathcal{L}_D \leftarrow \text{BCE}(D(x), 1) + \text{BCE}(D(\hat{x}), 0)$ 
4:    $\hat{f} \leftarrow \Phi(\hat{x})$ ,  $m \leftarrow \text{batch\_size}$ 
5:    $\mathcal{L}_{\text{centroid}} \leftarrow \frac{1}{m} \sum_{j=1}^m \max(0, \|\hat{f}_j - c\|_2 - d_{\text{ref}})$ 
6:    $\mathcal{L}_G \leftarrow \text{BCE}(D(\hat{x}), 1) + \lambda \cdot \mathcal{L}_{\text{centroid}}$ 
7: until converged

```

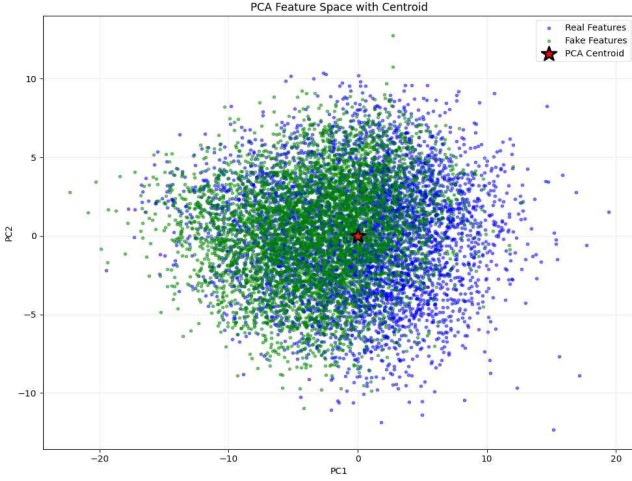


그림. 1 임베딩된 특징들과 그리고 중심점

그림 1은 실제 이미지와 생성 이미지를 특징 공간에 임베딩한 결과를 보여준다. 두 분포는 공통된 중심점을 기준으로 유사한 형태로 퍼져 있으며, 생성 이미지의 특징 벡터들이 실제 이미지의 특징 분포와 잘 정렬되어 있음을 확인할 수 있다. 이는 본 논문에서 제안한 중심 기반 제약 손실 함수가 생성 이미지의 특징 분포를 실제 데이터의 중심점과 일치시키는 데 효과적임을 시각적으로 입증한다.

III. 실험

3.1 실험 설정

본 논문의 실험은 CelebA 데이터셋을 활용하여 진행했으며, 생성 모델은 DCGAN[3] 구조를 기반으로 한다. 특징 추출에는 사전 학습된 VGG-16 네트워크의 16번째 계층까지 통과한 출력을 사용하였다. 제안한 중심 기반 제약 손실의 가중치 계수는 0.1로 설정한다. 모델의 성능 평가는 FID와 LPIPS 지표를 통해 수행했으며, FID는 50k, LPIPS는 5k 쌍을 비교하여 계산한다.

3.2 정량적 지표와 정성적 평가

표. 1 CelebA 64x64

	FID	LPIPS
DCGAN[3]	32.1361	0.2918
Ours	30.9519	0.2981

표. 1은 기존 DCGAN[3]과 본 논문에서 제안한 중심 기반 제약 손실 함수를 적용한 모델 간의 정량적 성능을 비교한 결과를 나타낸다. 제안한 기법은 DCGAN[3] 대비 FID가 낮고, LPIPS는 높은 값을 기록한다. 이는 중심 기반 제약이 생성 이미지의 특징 분포를 실제 데이터의 중심 영역에 보다 가깝게 유도함으로써, 품질 저하 없이 시각적으로 구분 가능한 이미지를 다양하게 생성하는 데 일정 수준 기여했음을 의미한다. 두 지표 간 균형 개선이 확인되었다는 점에서, 본 손실 함수가 기존 구조에 단순하게

적용되면서도 이미지의 품질과 다양성 간의 균형을 조절할 수 있는 유효한 전략이 될 수 있음을 시사한다. 다만, 전체적인 성능 개선 폭은 제한적이며, 향후 다양한 조건에서의 추가 실험이 필요하다.



그림. 2 (Ours) CelebA 64x64 생성 이미지

그림. 2는 제안한 특징 공간 기반 중심 제약 손실 함수를 적용한 DCGAN[3] 모델이 생성한 이미지들을 보여준다.

IV. 결론

본 논문에서는 적대적 생성 신경망의 이미지 품질 저하 및 다양성 감소 문제를 완화하기 위한 새로운 중심 기반 제약 손실 함수를 제안한다. 제안한 방법은 학습 데이터셋을 특징 공간에 임베딩하고, 그 중심점과 참조 거리를 기반으로 생성 이미지의 특징 분포가 실제 데이터 분포로부터 과도하게 벗어나지 않도록 유도한다. 이는 별도의 네트워크 구조 변경 없이 손실 함수 수준에서 간단히 적용 가능하며, 학습 안정성과 분포 정합성을 동시에 향상시킬 수 있다는 장점이 있다.

향후 연구로는 해당 제약 조건을 적응적으로 조절하는 방식이나, 다양한 특징 공간 및 생성 모델에의 확장을 통해 일반화 가능성을 검토할 수 있을 것이다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2024-00338504).

참 고 문 헌

- [1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, Proc. NeurIPS '14, pp. 1 - 9, Dec. 2014.
- [2] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," *International Conference on Learning Representations*, Proc. ICLR '18, p. 1 - 26, Apr. 2018.
- [3] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *International Conference on Learning Representations*, Proc. ICLR '16, pp. 1 - 16, May 2016.