

주파수 영역을 활용한 GAN

이채은¹, 정성훈²

¹한성대학교 AI 응용학과 석사과정

²한성대학교 AI 응용학과 교수

rjh2436@hansung.ac.kr, shjung@hansung.ac.kr

GAN using Frequency Domain

Chae-Eun Lee¹, Sung Hoon Jung²

¹Dept. of Applied Artificial Intelligence, Hansung University

²Dept. of Applied Artificial Intelligence, Hansung University

요 약

GAN은 이미지 생성모델로서 이미지 공간에서 좋은 결과를 보여왔다. 우리는 이러한 GAN의 능력을 더욱 향상하기 위하여 본 연구에서 주파수 영역에서 이미지를 학습하고 생성하는 새로운 방법을 제안한다. 이를 위하여 먼저 학습데이터를 2D FFT로 주파수 영역으로 변환한 후 변환된 학습데이터를 GAN이 학습하게 한다. 학습 후에 GAN은 새로운 이미지를 생성하며 생성된 이미지를 2D IFFT하여 이미지 공간으로 변환한다. 이렇게 주파수 영역에서 이미지를 생성하는 방법은 이미지 공간에서 생성하는 방법보다 다양한 장점이 있다. 생성된 이미지의 품질을 평가하기 위하여 4개 데이터 셋에 4개의 평가지표를 사용하여 평가한 결과 주파수 영역에서 생성한 이미지가 IS, P&R, D&C 측면에서 더 좋은 것으로 평가되었다.

1. 서론

Generative Adversarial Network(GAN)은 생성자(Generator)와 판별자(Discriminator)를 이용하여 경쟁적으로 이미지를 생성하고 판별하는 방식으로 학습한다[1]. 이러한 GAN은 이미지 생성뿐 아니라 이미지 변환, 해상도 향상 등 다양한 이미지 처리 분야에도 활용되고 있다[2].

이미지는 일반적으로 공간 (x, y) 상에서 2차원 픽셀 그리드로 표현된다. 이러한 2차원 이미지의 공간 정보를 차원 별로 주파수 성분으로 분해하여 주파수 영역(Frequency Domain)으로 나타낼 수 있다. 주파수 성분으로 분해된 정보는 위치별 주파수 정보를 포함하고 있어서 이미지의 주요한 구조와 특징을 제공한다. 그러므로 이미지 공간에서의 정보보다 더 다양한 정보를 제공할 수 있다[3-5].

본 논문에서 우리는 이러한 분석을 토대로 GAN이 이미지 공간의 이미지를 학습하여 새로운 이미지를 만드는 것이 아니라 주파수 공간의 이미지를 학습하여 새로운 이미지를 생성하는 방식을 제안한다. 이 방법은 이미지 공간에서 학습하는 방법보다 다양한 장점과 활용도를 가질 수 있을 것으로 판단된다. 이에 본 논문에서는 4개의 데이터셋에 다양한 실험

을 통하여 제안한 방법의 성능을 평가하고 효과를 분석한다.

제안방법의 성능을 실험하기 위하여 LSGAN을 선택하였다. 실험 데이터셋으로는 비교적 단순한 데이터셋 부터 복잡한 데이터셋으로 MNIST, Fashion MNIST, SVHN, MNIST-M을 사용하였다. 각 실험에서 생성된 이미지의 품질을 FID, IS, D&C, P&R의 4가지 평가지표를 사용하여 평가하고 학습 소요 시간을 비교하였다. 평가 결과 본 논문에서 제안한 방법이 거의 모든 평가지표 측면에서 더 좋은 결과를 보였다.

2. 관련 연구

이미지 공간보다 이미지의 주파수 영역에서의 분석이나 실험이 더 유용하다는 많은 연구가 있다. 논문 [3]의 연구에서는 기계학습 분야에서 주파수 영역을 이용한 학습의 장점을 언급한다. 주파수 영역에 포함된 풍부한 정보가 모델의 이미지 학습을 도울 뿐 아니라 이미지 정보를 더 잘 보존하고 정확도를 향상한다는 연구 결과를 보였다.

논문 [4]의 연구에서는 이산 푸리에 변환(Fourier Transformation)이 딥러닝의 컨볼루션 계산에서 상당한 속도를 제공한다는 연구 결과를 보여주며, 다

양한 CNN 모델에서 이산 푸리에 변환을 수행하여 모델을 학습시키면 더 빠른 속도로 높은 정확도에 모델이 수렴하는 것을 보여준다.

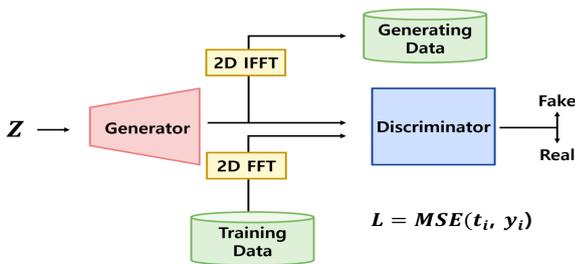
논문 [5]에서는 시간 영역(Time Domain)에서와 주파수 영역 각각의 데이터로 EMG 신호를 분류하는 모델을 구현하였는데, 주파수 영역 데이터를 이용한 모델이 신호 분류에서 가장 높은 성능을 나타내었음을 보여준다.

이 밖에도 주파수 영역에서는 주파수 대역에 따라 이미지의 정보 표현을 더욱 세밀하게 다룰 수 있다[6]. 또한, 이미지를 필터링하여 이미지의 잡음 제거 및 경계 검출 등을 통해 이미지 향상이 가능하다[7]. 마지막으로 이미지 경계 검출을 위한 Sobel 필터를 사용할 수 있는데, 이러한 필터링은 주파수 영역에서 컨볼루션을 사용하여 더욱 빠르게 처리할 수 있다[8].

이에 본 논문에서는 주파수 영역을 생성모델 GAN에 적용시킬 수 있는 방법을 제안하였다.

3. 제안방법

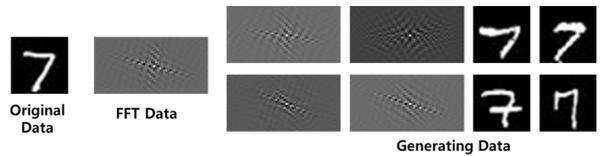
그림 1은 본 논문에서 우리가 제안한 주파수 영역을 활용한 GAN의 학습 및 생성과정을 나타낸다. 그림에서 보듯이 모든 학습데이터를 2D FFT를 사용하여 주파수 영역으로 변환한다. 주파수 영역으로 변환된 데이터는 이미지 크기와 동일한 크기로 진폭(amplitude)과 위상(phase)정보를 갖는 배열을 얻을 수 있다. 그러므로, MNIST와 같은 28x28 데이터는 진폭 28x28 데이터와 위상 28x28 데이터로 표현되며 이를 합쳐서 28x56 데이터로 저장된다. SVHN과 같은 컬러 데이터의 경우에는 RGB 각각의 채널별로 주파수 데이터로 변환된다. 주파수 영역으로 변환된 학습데이터를 학습한 후 새로운 주파수 영역의 데이터를 생성한다. 생성된 데이터를 2D IFFT 하면 이미지 영역의 데이터를 얻을 수 있다.



(그림 1) 주파수 영역을 활용한 학습방법

지금까지 개발된 대부분의 GAN을 실험에 사용할 수 있으나 우리는 그중에 LSGAN을 선택하였다. 그 이유는 LSGAN 이후에 개발된 모델들은 특정 목적에 특화된 것이 많아서 이미지 영역과 주파수 영역에서의 효과를 비교하기에는 LSGAN이 가장 적절하다고 판단하였다.

실험 데이터셋으로는 기존에 많이 사용되는 MNIST, Fashion MNIST, SVHN, MNIST-M 데이터셋을 사용하였다. 이미지 영역에서의 결과와 주파수 영역에서의 실험결과를 비교하기 위하여 GAN 모델의 성능 평가 지표인 FID, IS, P&R, D&C를 사용하였다.



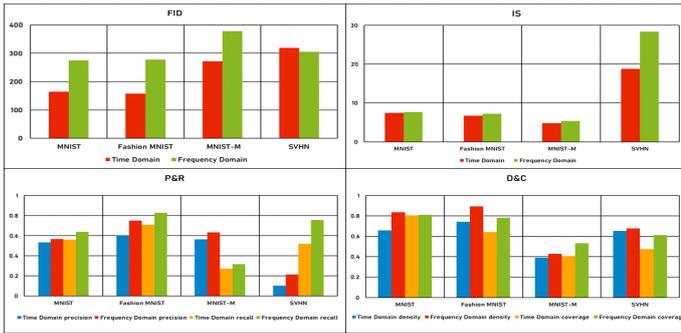
(그림 2) 주파수 영역을 활용하는 학습 과정 이미지

4. 실험결과

그림 2는 주파수 영역에서 이미지를 학습하고 생성하는 하나의 예를 보여준다. 먼저, 이미지 영역의 데이터는 2D FFT로 주파수 영역의 데이터로 변환된다. 이때 주파수 영역으로 변환하면 진폭과 위상 정보가 출력되어 기존 이미지 데이터 크기의 2배가 된다. 그림 2에서 FFT Data가 이에 해당하는데 진폭과 위상정보를 인터리빙하여 합친 결과이다. 이렇게 얻어진 주파수 영역 이미지를 GAN이 학습한 후 새로운 주파수 영역 이미지를 생성한다. 생성된 이미지를 2D IFFT 하기 전에 인터리빙 방법으로 진폭과 위상을 분리하기 때문에 이미지 공간으로 변환된 결과를 보면 새로운 이미지가 잘 생성된 것을 볼 수 있다.

MNIST, Fashion MNIST, SVHN, MNIST-M 데이터셋으로 실험하고 FID, IS, P&R, D&C로 평가한 결과는 그림 3과 같다. 실험에 사용한 학습 파라미터는 <표 1>과 같다. 실험은 모델의 초기 가중치를 바꿔가면서 5번의 실험결과를 평균 낸 것이다.

그림 3을 보면 FID 실험에서 MNIST, Fashion MNIST, MNIST-M의 결과를 제외하면 이외에 모든 다른 실험에서는 주파수 영역에서 GAN을 적용한 것이 더 좋은 성능을 보임을 알 수 있다.



| Metric | Data | Domain Value | | |
|--------|---------------|---------------|-----------|-------|
| | | Time | Frequency | |
| FID | MNIST | 164.777 | 275.03 | |
| | Fashion MNIST | 157.891 | 277.359 | |
| | SVHN | 318.971 | 305.628 | |
| | MNIST-M | 271.717 | 378.686 | |
| IS | MNIST | 7.45 | 7.67 | |
| | Fashion MNIST | 6.7 | 7.2 | |
| | SVHN | 18.8 | 28.3 | |
| | MNIST-M | 4.8 | 5.4 | |
| D & C | Density | MNIST | 0.697 | 0.836 |
| | | Fashion MNIST | 0.742 | 0.793 |
| | | SVHN | 0.618 | 0.697 |
| | | MNIST-M | 0.389 | 0.428 |
| | Coverage | MNIST | 0.805 | 0.811 |
| | | Fashion MNIST | 0.641 | 0.789 |
| SVHN | | 0.568 | 0.812 | |
| P & R | Precision | MNIST | 0.533 | 0.566 |
| | | Fashion MNIST | 0.602 | 0.75 |
| | | SVHN | 0.103 | 0.633 |
| | | MNIST-M | 0.563 | 0.214 |
| | Recall | MNIST | 0.561 | 0.638 |
| | | Fashion MNIST | 0.709 | 0.826 |
| | | SVHN | 0.521 | 0.754 |
| | | MNIST-M | 0.274 | 0.318 |

(그림 3) 주파수 영역과 이미지 영역에서의 평가지표

<표 1> 이미지 영역과 주파수 영역에서의 학습 파라미터

| Parameter | Value | Parameter | Value |
|------------|-------|---------------|--------|
| batch size | 128 | learning rate | 0.0002 |
| epochs | 1500 | optimizer | Adam |

본 논문에서 제안한 방법을 독립적으로 사용해도 되지만 이미지 공간 판별자와 같이 사용하면 효과가 더욱 확실히 나타날 것으로 판단된다.

5. 결론

본 논문에서는 GAN 학습에 사용되는 이미지 데이터에 2D FFT를 적용시켜 주파수 영역으로 이미지를 변환하여 학습하고 주파수 영역에서 데이터를

생성하도록 하는 학습방법을 제안하였다. 생성된 데이터는 2D IFFT를 통하여 이미지 영역 데이터로 변환하여 사용할 수 있다. 4개의 데이터셋에 실험하여 이미지 공간에서의 학습한 것과 비교한 결과 IS, P&R, D&C 평가지표 상으로 더 좋은 성능을 보였다. FID 실험에서 MNIST, Fashion MNIST, MNIST-M 데이터셋에서는 좋지 못했는데 향후 그 이유를 좀 더 명확히 분석할 예정이다. 본 논문에서의 방법은 다양한 방법으로 기존 방법과 같이 사용되어 성능을 높이는데 사용될 수 있다. 우리는 향후 이러한 방법에 대하여 추가로 연구할 예정이다.

참고문헌

- [1] Creswell, Antonia, et al. "Generative adversarial networks: An overview." IEEE signal processing magazine 35.1 (2018): 53-65.
- [2] Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).
- [3] Xu, Kai, et al. "Learning in the frequency domain." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
- [4] Rippel, Oren, Jasper Snoek, and Ryan P. Adams. "Spectral representations for convolutional neural networks." Advances in neural information processing systems 28 (2015).
- [5] Too, J. et al. "Classification of EMG signal based on time domain and frequency domain features." International Journal of Human and Technology Interaction (IJHaTI) 1.1 (2017): 25-30.
- [6] Shannon, Claude Elwood. "A mathematical theory of communication." ACM SIGMOBILE mobile computing and communications review 5.1 (2001): 3-55.
- [7] Sroubek, Filip, and Jan Flusser. "Multichannel blind iterative image restoration." IEEE Transactions on Image Processing 12.9 (2003): 1094-1106.
- [8] Marr, David, and Ellen Hildreth. "Theory of edge detection." Proceedings of the Royal Society of London. Series B. Biological Sciences 207.1167 (1980): 187-217.