

Wasserstein 거리 척도 기반 SRGAN을 이용한 위성 영상 해상도 향상

황지언, 유초시, 신요안*

승실대학교 전자정보공학부

(*교신저자)

hwangjieon@soongsil.ac.kr, csyu@soongsil.ac.kr, yashin@ssu.ac.kr

요약

본 논문에서는 기존 SRGAN (Super Resolution Generative Adversarial Network)에서 Wasserstein 거리 척도를 이용하여 위성 영상의 해상도를 더욱 향상하는 방안을 제안한다. GAN의 단점인 모드 붕괴 현상을 개선하기 위해 Wasserstein 거리 및 Gradient Penalty를 적용하였으며, 평균 제곱 오차 손실을 추가하여 픽셀 단위로 비교해 최적화하였다. 또한 입력과 출력 영상의 유사성이 존재하므로 배치 정규화를 제거하여 약 40%의 메모리 사용량을 줄였다. 실험 결과를 통해 본 논문에서 제안된 딥러닝 기법 및 손실을 적용하여 좋은 품질의 초고해상도 위성 영상을 생성할 수 있음을 보였다.

I. 서론

초고해상도 (Super Resolution) 영상 연구는 위성 영상, 의료 영상, 감시 카메라 등 다양한 분야에서 활발히 진행되고 있다. 그중 단일 영상의 초고해상도 변환 목표는 저해상도 영상을 입력으로 받아 딥러닝과 같은 알고리즘을 통해 고해상도 영상을 만드는 것이다. 이를 위해, 최근에는 GAN (Generative Adversarial Network)[1] 딥러닝 알고리즘 기반의 SRGAN (Super Resolution GAN)[2] 기법을 활용한 초고해상도 영상 생성 방법에 대한 연구도 진행되고 있으며, 참고문헌 [3]에서는 위성 영상 해상도 향상을 위해 SRGAN 기법을 적용하였다.

본 논문에서는 기존 GAN의 단점을 개선한 Wasserstein GAN (WGAN)[4]을 활용하여 보다 안정적으로 학습할 수 있고, Gradient Penalty[5]와 평균 제곱 오차 (Mean Square Error; MSE) 손실을 추가하여 해상도를 향상하는 방안을 제안한다.

II. 본론

1. 제안 손실 함수

SRGAN[2]은 아래 식 (1)에서와 같이 생성기 (Generator) 네트워크 G 와 판별기 (Discriminator) 네트워크 D 가 서로 경쟁하며 최소-최대 문제를 최적화하는 방식으로 학습한다.

$$\min_G \max_D E_{I^{HR} \sim P_r(I^{HR})} [\log D(I^{HR})] - E_{I^{LR} \sim P_g(I^{LR})} [\log(1 - D(G(I^{LR})))] \quad (1)$$

(P_r : I^{HR} 의 확률분포, P_g : I^{LR} 의 확률분포)

여기서 생성기 G 는 저해상도 영상을 입력으로 받아 고해상도 영상 I^{SR} 을 생성하며, 판별기 D 는 원본 고해상도 영상과 I^{SR} 을 판별하게 된다. 그러나, 실제 학습 시 판별기가 학습됨에 따라 생성기 네트워크는 반대로 성능이 저하되는 불안정한 현상, 즉 모드 붕괴 (Mode Collapse) 현상이 발생한다. 이는 생성기 네트워크가 다양한 샘플을 만들지 못하여 매번 반복된 샘플의 출력이 발생하는 현상이다. 이를 해결하기 위해 여러 연구가 진행 되었으며, 그중 WGAN에서는 확률분포 P_r 와 P_g 간의 Wasserstein 거리가 연속적일 때 이런 해결할 수 있으며[4], 이 때 위의 식 (1)이 아래와 같이 조정된다.

$$\min_G \max_D E_{I^{HR} \sim P_r(I^{HR})} [D(I^{HR})] - E_{I^{LR} \sim P_g(I^{LR})} [D(G(I^{LR}))]. \quad (2)$$

참고문헌 [4]에 따르면, Wasserstein 거리가 연속적이려면 Lipschitz 함수의 연속성을 유지해야한다. 따라서 Lipschitz 연속 조건을 만족하도록 가중치를 $[-0.01, 0.01]$ 범위 안으로 강제로 클리핑을 한다. 하지만, 이 경우 지나치게 간략하게 수렴하므로 왜곡이 발생하며 학습 속도도 감소하게 된다. 이를 개선하기 위해, 클리핑 대신 Gradient Penalty를 사용하여 1에서 크게 벗어날 때 Penalty를 부과하여 Lipschitz 함수의 연속성을 안정적으로 유지할 수 있게 된다[5]. 따라서, 이를 반영하면 판별기의 손실은 다음과 같이 표현된다.

$$L_D = E_{I^{LR} \sim P_g(I^{LR})} [D(G(I^{LR}))] - E_{I^{HR} \sim P_r(I^{HR})} [D(I^{HR})] + \lambda_{gp} E_{I^{SR} \sim P_{gp}(I^{SR})} [(\|\nabla_{I^{SR}} D(I^{SR})\|_2 - 1)^2]. \quad (3)$$

($\nabla_{I^{SR}}$: I^{SR} 의 기울기, λ_{gp} : Gradient Penalty의 가중치 비율)

한편 생성기의 손실은, 기존의 SRGAN[2]에서 제안된 바와 같이 CNN (Convolutional Neural Network) 네트워크 중 하나인 VGG19 네트워크를 대표적인 영상 데이터 세트인 ImageNet으로 미리 학습한 가중치를 사용한다. 이를 지각 손실 (Perceptual Loss)이라 하며, 다음과 같이 표현된다.

$$L_{vgg} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} [\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(I^{LR})_{x,y}]^2. \quad (4)$$

($W_{i,j}$: Width, $H_{i,j}$: Height, $\phi_{i,j}$: j 번째 Convolution Layer의 Feature Map)

위 식에서 I^{HR} 와 I^{LR} 의 각 Feature Map을 L_2 거리로 비교하여 지각 유사성을 측정한다. 본 논문에서는 여기에 덧붙여 초고해상도 영상 처리시 일반적으로 사용하는 MSE 손실을 추가하여 픽셀 단위로 비교하여 최적화한다. 이들을 반영한 생성기의 모든 손실은 다음과 같이 표현된다.

$$L_G = \lambda_G E_{I^{LR} \sim P_g(I^{LR})} [D(G(I^{LR}))] + \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H (I_{x,y}^{HR} - I_{x,y}^{LR})^2 + \lambda_{vgg} L_{vgg}. \quad (5)$$

(λ_G : 생성기 손실의 가중치 비율, λ_{vgg} : VGG 손실의 가중치 비율)

2. 제안 딥러닝 네트워크 구조

제안 딥러닝 네트워크 구조는 SRGAN[2]의 구조를 기반을 두었다. 그림 1에서 생성기는 안정적으로 학습을 할 수 있도록 잔차 블록

(Residual Block) 16개와 Skip Connection 구조로 이루어져 있다. 본 논문에서는 입력과 출력 영상의 유사성이 이미 존재하므로 정규화를 할 필요가 없기에, 기존 구조[2]에서 배치 정규화는 제거하여[6] 약 40%의 메모리 사용량을 줄였다.

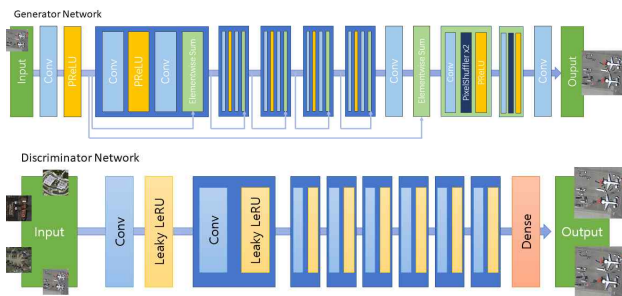


그림 1. 제안된 SRGAN 구조

Fig. 1. Proposed SRGAN structure

III. 실험 결과

본 논문의 실험을 위한 데이터 세트는 에어버스사에서 제공하는 <https://www.intelligence-airbusds.com/>의 공개 위성 영상 데이터를 참조하였으며, 학습 데이터를 488장, 테스트 데이터를 122장을 사용하였다. 각 영상에 대해서 384×384 크기의 패치로 나누어 입력 데이터로 사용하였다. 저해상도 영상은 96×96 크기이며, Bicubic 기법을 적용하여 4배 크기로 Down Scaling하였다. 그림 2는 총 2000 Epoch 학습의 결과이며, 다른 초고해상도 변환 기법과 성능을 비교하였다. 이 때 평가 방법으로 PSNR (Peak Signal-to-Noise Ratio)와 SSIM (Structural Similarity Index Measure)를 사용하였으며, 실험 결과는 표 1에 정리하였다. 이들 결과로부터, 우리가 제안한 WGAN 기반의 SRGAN 기법을 통해 변환된 초고해상도 영상이 기존의 SRGAN보다 향상된 것을 확인할 수 있다.

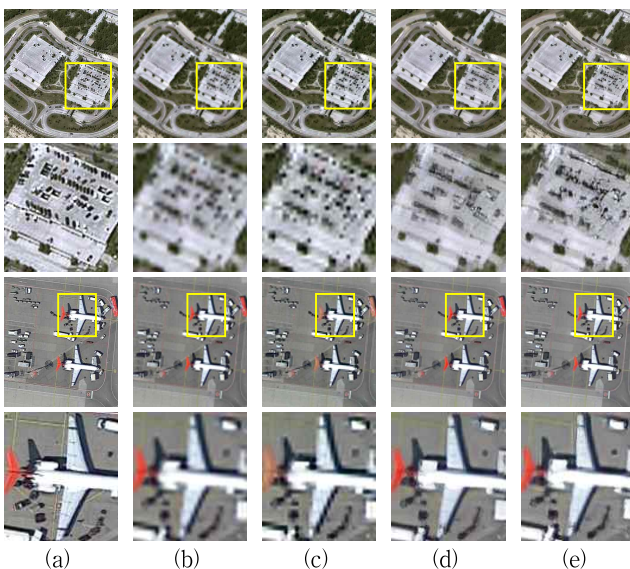


그림 2. 실험 결과

Fig. 2. Experiment results

(a) Ground truth, (b) Bicubic, (c) SRCNN(7), (d) SRGAN, (e) Proposed

IV. 결론

본 논문에서는 향상된 초고해상도 영상을 얻기 위하여 변형된 SRGAN을 적용하였다. GAN의 단점인 모드 붕괴 현상을 개선하기 위해 Wasserstein 거리 및 Gradient Penalty를 적용하였으며, MSE 손실을 추가하여 픽셀 간 측정을 추가하였다. 실험 결과, 기존 기법에 비교해 향상된 초고해상도 영상을 얻을 수 있음을 확인하였다.

표 1. 성능 비교

Table 1. Performance comparison

	Bicubic	SRCNN	SRGAN	Proposed
PSNR	19.1982	19.6860	20.1466	20.3707
SSIM	0.5624	0.5969	0.6115	0.6222
PSNR	21.2244	21.8164	22.6600	22.8154
SSIM	0.7002	0.6780	0.7221	0.7287

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2020-2018-0-01424)

참고 문헌

- [1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Proc. NIPS 2014*, pp. 2672-2680, Montreal, Canada, Dec. 2014.
- [2] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *Proc. IEEE CVPR 2017*, pp. 4681-4690, Honolulu, USA, July 2017.
- [3] 유초시, 이상훈, 신요안, "위성 영상 해상도 향상을 위한 SRGAN 적용 방안," *2020년도 한국통신학회 동계종합학술발표회 논문집*, pp. 202-203, 2020년 2월.
- [4] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," *Proc. ICML 2017*, pp. 214-223, Sydney, Australia, Aug. 2017.
- [5] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," *Proc. NIPS 2017*, pp. 5767-5777, Long Beach, USA, Dec. 2017.
- [6] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," *Proc. IEEE CVPR 2017*, pp. 1132-1140, Honolulu, USA, July 2017.
- [7] C. Dong, C. C. Loy, K. He, and S. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. & Machine Intell.*, vol. 38, no. 2, pp. 295-307, Feb. 2016.