ESRGAN의 성능 향상을 위한 판별자 설계 공간 재검토에 관한 연구

박성욱[1], 김준영[1], 박준[1], 정세훈[2], 심춘보[1]
[1]순천대학교 IT-Bio융합시스템전공
[2]순천대학교 컴퓨터공학과
411050@scnu.ac.kr, shjung@scnu.ac.kr, cbsim@scnu.ac.kr

A Research on Re-examining Discriminator Design Space for Performance Improvement of ESRGAN

Sung-Wook Park[1], Jun-Yeong Kim[1], Jun Park[1], Se-Hoon Jung[2], Chun-Bo Sim[1]
[1]Interdisciplinary Program in IT-Bio Convergence System, Sunchon National University
[2]Dept. of Computer Engineering, Sunchon National University

요 약

초해상은 저해상도의 영상을 고해상도 영상으로 합성하는 기술이다. 이 기술에 딥러닝이 적용되어, 2014년에는 SRCNN(Super Resolution Convolutional Neural Network) 모델이 발표됐다. 이후에는 SRCAE(Super Resolution Convolutional Autoencoders)와 GAN(Generative Adversarial Networks)을 기반으로 한 SRGAN(Super Resolution Generative Adversarial Networks) 등, SRCNN의 성능을 능가하는 모델들이 발표됐다. ESRGAN(Enhanced Super Resolution Generative Adversarial Networks)은 SRGAN 모델의 성능을 개선했지만, 완벽한 성능을 내지 못하는 문제점이 있다. 이에 본 논문에서는 판별자(Discriminator) 구조를 변경하여 ESRGAN의 성능을 개선한다. 실험 결과, 제안하는 모델이 ESRGAN보다 더 높은 성능을 보일 것으로 기대된다.

1. 서론

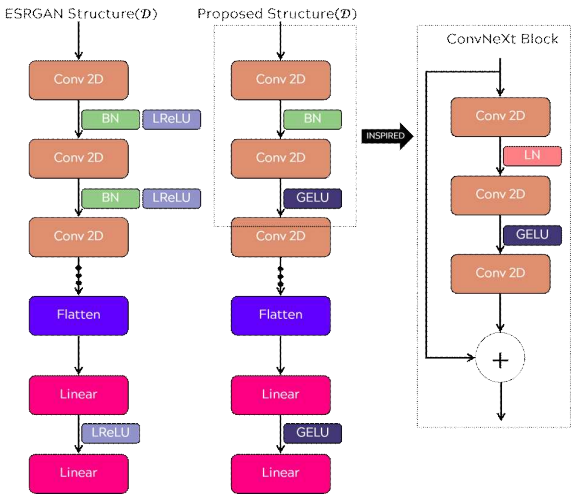
딥러닝(Deep Learning)은 영상의 해상도를 높이는 초해상(Super Resolution, SR) 기술에도 응용된다. SR이란 해상도가 낮은 영상을 해상도가 높은 영상으로 합성하는 기술이다.

딥러닝을 이용한 SR의 실현은 2014년에 발표된 SRCNN(Super Resolution Convolutional Neural Network) 모델로부터 됐다. 3층의 매우 단순한 모델이지만 기존 방법을 능가하는 성능을 보였다. 이후 SRCAE(Super Resolution Convolutional Autoencoders), 합성된 영상을 실제 영상과 통계적으로 구분되지 않도록 강제하는 GAN(Generative Adversarial Networks) 기반의 SRGAN(Super Resolution Generative Adversarial Networks) 등 SRCNN의 성능을 웃도는 모델들이 발표됐다. 얼마 지나지 않아 SRGAN 모델의 성능을 개선한 ESRGAN(Enhanced Super Resolution Generative Adversarial Networks) 모델도 발표됐지만, 아직 완벽한 성능을 내진 못한다[1].

이에 본 논문에서는 ESRGAN의 성능을 개선하기 위해 판별자(Discriminator, *D*) 구조를 변경하여 실험해보고, 기존 모델과 성능을 비교한다. 성능지표로는 최대 신호 대 잡음 비(Peak Signal to Noise Ratio, PSNR)를 이용한다.

2. 제안하는 모델

ESRGAN은 3개의 신경망으로 구성돼 있다. 잔차(Residual) 모듈을 이용하는 생성자(Generator, *G*), *D*, 사전 훈련된 VGG(Visual Geometry Group)-19 신경망이다. ESRGAN에 의해 복구된 영상 품질을 더욱 향상하기 위해 *D*의 구조를 ConvNeXt 모델에 영감받아 그림 1과 같이 변경한다[2].



(그림 1) Block designs for an ESRGAN, a Proposed,

and a ConvNeXt

G와 D가 깊은 신경망이고, 훈련 데이터 세트와 테스트 데이터 세트의 통계가 상이하면 아티팩트(Artifact) 발생위험이 생기며 이는 곧 일반화 능력 저하까지 귀결될 수 있다. 따라서 안정적인 훈련 환경을 제공하기 위해 최적인 정규화 층을 추가한다. 정규화 층을 추가하면 일반화 능력과 SR 및 디블러링(Deblurring) 작업에서의 성능이 더욱 향상될 것으로 기대된다. 활성 함수는 가우시안 오차 선형 유닛(Gaussian Error Linear Units)을 이용한다.

ESRGAN의 기본 구조는 유지하되 새로운 블록을 제안했다. SR에서 꼭 층을 늘리는 것만이 좋은 것은 아니라는 관찰에 기반을 두어 제안한 블록은 ESRGAN의 블록보다 얕고, 덜 복잡하다. 제안하는 모델은 지각 손실(Perceptual Loss)을 이용하여 합성된 영상을 더 자연스럽고, 세부 사항을 더 예술적으로 보이게 한다. 지각 손실(l^{SR})은 콘텐츠 손실(l_X^{SR})과 적대적 손실(l_{Gen}^{SR})의 가중 합계로 정의되며 식 (1)과 같다.

$$l^{SR} = l_X^{SR} + 10^{-3} \times l_{Gen}^{SR} \qquad \text{식 (1)}$$

식 (1)의 첫 번째 항은 사전 훈련된 VGG-19가 합성한 특징맵(Feature Map)을 이용해 획득한 콘텐츠 손실(Content Loss)이다. 이는 수학적으로 재구성된 영상의 특징맵과 기존 고해상도 참조 영상 사이의 유클리드 거리(Euclidean Distance)다. 식 (1)의 두 번째 항은 적대적 손실(Adversarial Loss)이다. 이 항은 G가 만든 영상이 D를 속일 수 있게 하고자 설계된 항이다.

3. 실험 및 성능평가

3.1 데이터 세트 구성

훈련에는 2,048 너비의 DIV(DIVerse)2K 데이터 세트를 이용한다. 텍스처(Texture)가 풍부한 데이터 세트를 훈련에 이용하면 G가 더욱 자연스러운 영상을 합성하기 때문이다.

테스트에는 잘 알려진 벤치마크(Benchmark) 데이터 세트인 Set5, Set14, BSD(Berkeley Segmentation Dataset)100 및 Urban100을 이용한다.

3.2 평가 지표

일반적으로 SR에서는 성능평가 지표로 PSNR을 이용한다. PSNR의 사전적 의미는 정지 영상, 동영상의 손실 압축에서 화질 손실 정보를 수치로 표현한 값으로 영상에 워터마크를 삽입할 때 원본 영상과 워터마크(Watermark)를 삽입한 영상의 차이를 표현하는 수치다. 단위는 데시벨(Decibel)이며 식 (3)으로 정의할 수 있고, MSE(Mean Square Error)는 식 (2)로 계산할 수 있다.

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \qquad \text{식 (2)}$$

$$PSNR = 10 \times \log_{10}\left(\frac{MAX^2}{MSE}\right) \qquad \text{식 (3)}$$

식 (2)의 MSE는 잡음이 없는 $m \times n$ 원본 영상 I 와 잡음이 있는 근사 K 영상으로 정의할 수 있다. 즉, 식 (2)는 원본 영상과 출력 영상의 평균 제곱 오차다. 식 (3)의 MAX는 원본이 가질 수 있는 최댓값이다. PSNR은 원본과 출력이 유사할수록 값이 커진다.

학습 중 검증 데이터의 PSNR이 최고점이면 모델을 저장하는 모델 체크포인트(Model Checkpoint) 알고리즘을 적용하고, 최적 학습 에폭(Epoch)을 자동으로 탐색하게 설정한다.

4. 결론

본 논문에서는 ESRGAN과 제안하는 모델의 SR 성능을 실험하고 비교한다. 실험 결과 흐릿하던 입력 영상의 윤곽은 예측 영상에서 뚜렷해지고, 실제 영상과 더 가까워질 것으로 기대된다. 육안으로 식별했을 때 ESRGAN과 제안하는 모델의 성능은 큰 차이가 없을 수 있지만, 최적의 PSNR은 제안하는 모델이 더 높을 것으로 사료된다.

본 논문의 분석 결과는 시간과 비용을 절약하는데 크고 작은 도움이 될 것으로 판단되며 실험에 이용된 모델들은 하이퍼파라미터 값 정밀 조정을 통해 더 높은 성능을 얻을 수 있을 것으로 보인다. 또한, 제안하는 모델은 고속 처리가 가능하여 동영상 같은 실시간 처리에도 적합할 것으로 기대된다. 향후 연구 과제로 SR의 성능을 지금보다 향상할 수 있는 기술 개발이 필요할 것으로 사료된다.

Acknowledgment

This work was supported by the BK21 plus program through the National Research Foundation (NRF) funded by the Ministry of Education of Korea(5199990214660).

참고문헌

[1] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, et al., “Esrgan: Enhanced super-resolution generative adversarial networks,” *In Proceedings of the European conference on computer vision workshops*, 2018.

[2] Z. Liu, H. Mao, C. Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, “A ConvNet for the 2020s,” *arXiv*, arXiv:2201.03545, 2022.