특징 추출기에 따른 SRGAN의 초해상 성능 분석

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

Super Resolution Performance Analysis of GAN according to Feature Extractor

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

초해상이란 해상도가 낮은 영상을 해상도가 높은 영상으로 합성하는 기술이다. 딥러닝은 영상의 해상도를 높이는 초해상 기술에도 응용되며 실현은 2014년에 발표된 SRCNN(Super Resolution Convolutional Neural Network) 모델로부터 시작됐다. 이후 오토인코더(Autoencoders) 구조로는 SRCAE(Super Resolution Convolutional Autoencoders), 합성된 영상을 실제 영상과 통계적으로 구분되지 않도록 강제하는 GAN(Generative Adversarial Networks) 구조로는 SRGAN(Super Resolution Generative Adversarial Networks) 모델이 발표됐다. 모두 SRCNN의 성능을 웃도는 모델들이나 그중 가장 높은 성능을 끌어내는 SRGAN 조차 아직 완벽한 성능을 내진 못한다. 본 논문에서는 SRGAN의 성능을 개선하기 위해 사전 훈련된 특징 추출기(Pre-trained Feature Extractor) VGG(Visual Geometry Group)-19 모델을 변경하고, 기존 모델과 성능을 비교한다. 실험 결과, VGG-19 모델보다 윤곽이 뚜렷하고, 실제 영상과 더 가까운 영상을 합성할 수 있는 모델을 발견할 수 있을 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.

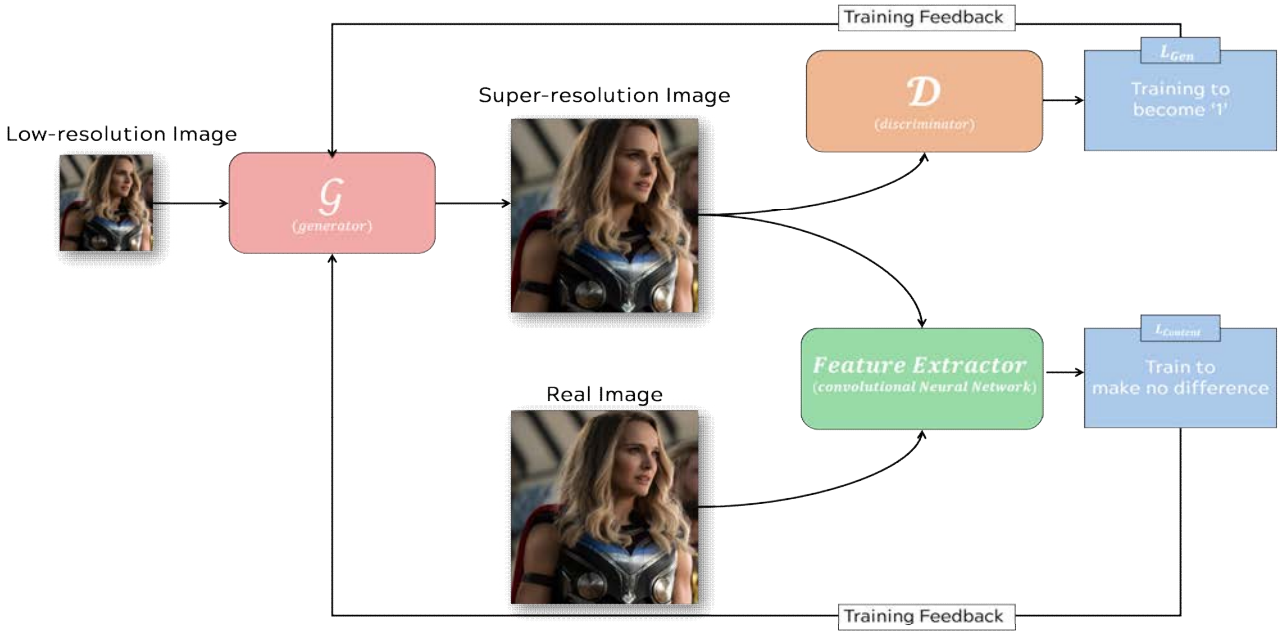
딥러닝을 이용한 SR의 실현은 2014년에 발표된 SRCNN(Super Resolution Convolutional Neural Network) 모델로부터 시작됐다[1]. 이후 오토인코더(Autoencoders) 구조로는 SRCAE(Super Resolution Convolutional Autoencoders), 합성된 영상을 실제 영상과 통계적으로 구분되지 않도록 강제하는 GAN(Generative Adversarial Networks) 구조로는 SRGAN(Super Resolution Generative Adversarial Networks) 모델이 발표됐다[2-4]. 모두 SRCNN의 성능을 웃도는 모델들이나 그중 가장 높은 성능을 끌어내는 SRGAN 조차 아직 완벽한 성능을 내진 못한다.

SRGAN은 그림 1과 같이 3개의 신경망으로 이뤄진다. 잔차(Residual) 모듈을 이용하는 생성자(Generator, *G*), 판별자(Discriminator, *D*), 사전 훈련된 특징 추출기(Pre-trained Feature Extractor) VGG(Visual Geometry Group)-19 모델이다[5]. 본 논문에서는 SRGAN의 성능을 개선하기 위해 특징 추출기를 변경하여 실험해보고, 기존 모델과 성능을 비교한다. 성능지표로는 최대 신호 대 노이즈 비(Peak Signal to Noise Ratio, PSNR)를 이용한다.

2. 이용할 모델 구조와 방법

·Generator: *G*는 크게 그림 2a와 같이 특징을 추출하는 기본 블록과 영상 크기를 확대하는 업샘플링(Upsampling) 레이어로 구성돼 있다.

그림 2b *G*의 기본 블록은 ResNet과 유사하여 컨벌루션(Convolution), 배치 정규화(Batch Normalization), 활성 함수(Activation Function) 층을 통과한다. 컨벌루션 층을 통과한 특징 맵은 입력 텐서(Tensor)와 더해 출력한다. 활성 함수는 PReLU(Parametric Rectified Linear Unit)를 이용한다. PReLU는 LeakyReLU(Leaky Rectified Linear Unit)와 유사한 활성 함수다. LeakyReLU는 0 이하의 값에서 고정된 값의 기울기를 갖지만 PReLU는 0 이하의 값에서 갖는 기울기를 학습할 수 있다. 기본 블록대로 층을 배치하고, 순전파(Forward Propagation)를 정의한다. 마지막으로 입력 텐서와 컨벌루션 층을 거친 텐서를 더한다. 기본 블록을 통과하면 영상을 업샘플링 하여 크기를 확대한다.



(그림 1) Architecture and the principle of SRGAN(Super Resolution Generative Adversarial Networks)

그림 2b에서 입력 텐서는 컨벌루션 층을 한 번 통과한다. 두 번째로 *G*의 기본 블록을 세 번 통과하고, 컨벌루션 층과 배치 정규화 층을 통과한다. 세 번째로 첫 번째 컨벌루션 층의 결과와 컨벌루션 블록을 거친 결과를 추가한다. 마지막으로 이전에 구현한 업샘플링 블록과 컨벌루션 층을 통과한다. 마지막 컨벌루션 층에 활성 함수가 붙지 않는 까닭은 SRGAN의 결과가 곧 각 픽셀의 값이 되기 때문이다. 활성 함수로 인해 값이 변하면 정보도 같이 변하기 때문에 컨벌루션 층의 결과를 정제 없이 이용한다.

·Discriminator: 그림 2c *D*의 구조는 일반적인 CNN과 유사하다. 컨벌루션 층을 통과한 뒤, 분류를 위해 영상을 1차원으로 변환한다. 마지막으로 시그모이드(Sigmoid) 함수가 활성 함수인 분류기(Classifier) 다층 퍼셉트론(Multilayer Perceptron, MLP)으로 이진 분류의 결과를 출력한다.

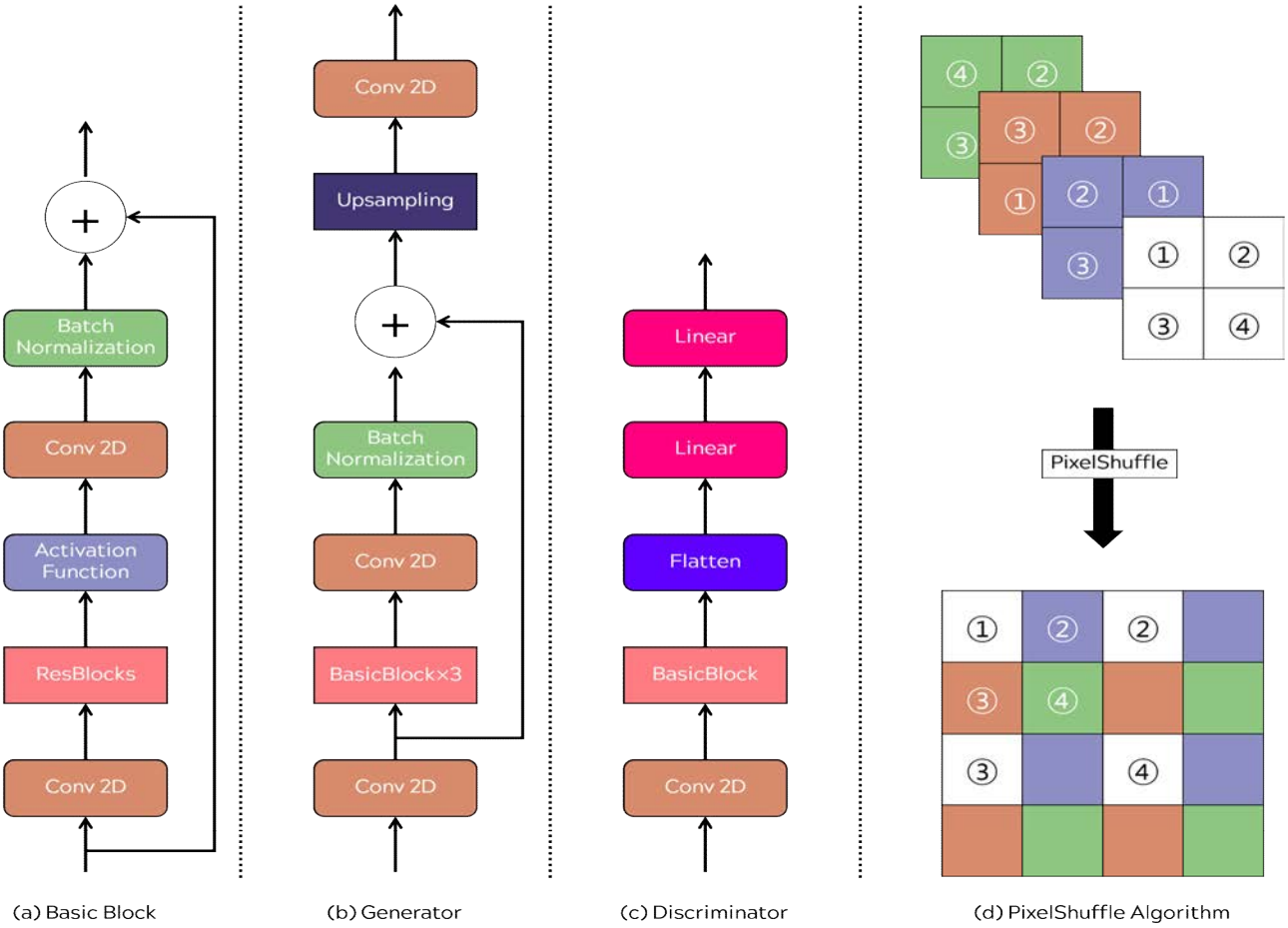
*D*는 영상 크기를 확대하는데 픽셀 셔플(Pixel Shuffle) 알고리즘을 이용한다. 그림 2d의 픽셀 셔플은 가중치를 갖지 않으며, 영상의 채널을 섞어 크기를 확대하는 알고리즘이다. 조정 가능한 가중치를 갖지 않기 때문에 곧장 픽셀 셔플을 이용하면 특징 맵으로부터 정보를 복원할 수 없다. 따라서 픽셀 셔플 전 컨벌루션 연산을 이용해 특징을 추출해야 한다. 그림 2d 상단의 ①~④는 하단의 ①~④ 위치로 이동한다. 상단 ②~④는 하단 ②~④로 이동한다. 마찬가지로 ②를 기준으로 뒤에 붙은 픽셀들은 ①에 인접한 ②~④처럼 ② 주변으로 이동한다.

·Feature Extractor: SRGAN은 *D* 이외에도 특징 추출기가 추출하는 특징을 고려해야 한다. 특징 추출기로 VGG-19 모델을 이용한다. VGG 모델의 특성상, 영상 크기를 축소하는 풀링(Pooling) 연산이 내재해 있으므로, 9개 층만을 이용한다. 그 이상의 층을 이용하면 영상의 최대 크기보다 작게 되어 오류가 발생한다. SRGAN에 의해 복구된 영상 품질을 더욱 향상하기 위해 이용할 특징 추출기는 ConvNeXt, DenseNet, EfficientNetV2, Inception V3, RegNet이다[6].

·Model Training: SRGAN의 *G*는 두 가지 손실을 이용한다. GAN 손실과 콘텐츠 손실이다. GAN 손실은 *D*가 *G*가 합성한 영상을 진짜라고 인식하게 하기 위한 손실 함수다. 콘텐츠 손실은 *G*가 합성한 영상이 진짜 영상과 유사해지도록 하는 손실 함수로써 진짜 영상과 *G*가 합성한 영상을 특징 추출기에 입력한 뒤, 출력된 두 값의 차이를 나타낸다. 둘의 특징이 유사해지도록 L1 손실을 이용해 콘텐츠 손실을 계산한다. L1 손실은 두 텐서 차이의 절댓값이다. L2 손실은 평균 제곱 오차(Mean Square Error, MSE)처럼 두 값의 차이를 제곱한다. 따라서 L2 손실은 1보다 큰 오차를 확대하고, 1보다 작은 오차는 줄여주는 효과가 있다. 단, 일반적으로 영상을 0과 1 사이의 값으로 스케일링(Scaling)하면 모든 픽셀의 값이 0과 1 사이의 값이 된다. 그러므로 L2 손실을 이용하면 오차가 줄어들게 된다. 이런 이유로 SRGAN은 L1 손실을 이용한다.

3. 결론

본 논문에서는 특징 추출기에 따른 SRGAN의 SR 성능을 실험하고 비교한다. 실험 결과 VGG-19 모델보다 윤곽이 뚜렷하고, 실제 영상과 더 가까운 합성 영상을 생성할 수 있는 모델을 발견할 수 있을 것으로 기대된다. 육안으로 식별했을 때 VGG-19 모델과 발견한 모델의 성능은 큰 차이가 없을 수 있지만, 최적의 PSNR은 발견한 모델이 더 높을 것으로 사료된다.



(그림 2) Generator, discriminator structure and upsampling algorithm of SRGAN

본 논문의 분석 결과는 시간과 비용을 절약하는데 크고 작은 도움이 될 것으로 기대되며 향후 SRGAN의 응용 모델도 제안한 방법을 이용하면 더 높은 성능을 얻을 수 있을 것으로 사료된다.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2020R1I1A3054843) and 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] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE transactions on pattern analysis and machine intelligence*, Vol. 38, No. 2, pp. 295-307, 2015.

[2] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections," *arXiv*, arXiv:1606.08921, 2016.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., "Generative adversarial networks," *Communications of the ACM*, Vol. 63, No. 11, pp. 139-144, 2020.

[4] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, et al., "Photo-realistic single image super-resolution using a generative adversarial network," *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4681-4690, 2017.

[5] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv*, arXiv:1409.1556, 2014.

[6] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11976-11986, 2022.