

HEVC 비디오 영상 압축 왜곡 제거를 위한 Generative Adversarial Network 적용 기법

전진, *김문철
한국과학기술원

wlsheon@kaist.ac.kr, * mkimee@kaist.ac.kr

Generative Adversarial Network based CNN model for artifact reduction on HEVC-encoded video

Jin Jeon, *Munchurl Kim

Korea Advanced Institute of Science and Technology

요 약

본 논문에서는 비디오 영상 압축 왜곡 제거를 위해 Generative Adversarial Network (GAN)을 적용한 컨벌루션 뉴럴 네트워크 (CNN) 모델을 제안한다. GAN 모델의 생성 모델 (Generator)은 노이즈가 아닌 High Efficiency Video Coding (HEVC)로 압축된 영상을 입력 받은 뒤, 압축 왜곡이 제거된 영상을 출력하며, 분류 모델 (Discriminator)은 원본 영상과 압축된 영상을 입력 받은 뒤, 원본 영상과 압축 왜곡이 포함된 압축된 영상을 분류한다. 분류 모델은 5 개 층을 쌓은 컨벌루션 뉴럴 네트워크 구조를 사용하였고, 생성 모델은 5 개 층을 쌓은 SRCNN 구조와 VDSR 구조를 기반으로 한 두 개의 모델을 이용한 실험을 통해 얻은 결과를 비교하였다. 비디오 영상 압축 왜곡 제거 실험을 위해 원본 비디오 영상을 HEVC를 이용하여 2Mbps, 4Mbps로 압축된 영상을 사용하였으며, 압축된 영상 대비 왜곡이 제거된 영상을 얻을 수 있었다.

1. 서론

방송 기술의 발달로 인해 고해상도 방송이 시작되었으며, 그에 따라 고해상도 영상을 압축하고 전송하기 위한 기술들이 개발되고 있다. 최신 영상 압축 표준으로 High Efficiency Video Coding (HEVC) [3]는 고해상도 영상을 압축하기 위한 차세대 압축 기술이다. 영상을 압축하고 전송하는 과정에서 영상의 왜곡이 발생하기 때문에 원본 영상과 압축 영상과의 차이를 줄이기 위한 연구 또한 지속적으로 연구되고 있으며, 최근 활발하게 연구되고 있는 딥러닝 기법이 영상 압축 왜곡을 제거하기 위해 연구되고 있다 [5].

최근 가장 주목받고 있는 딥러닝 기법은 Generative Adversarial Network (GAN) 모델로, 생성모델과 분류모델이 경쟁적으로 학습하는 과정에서 기존 딥 컨벌루션 뉴럴 네트워크 (deep convolutional neural network) 기법보다 성능을 향상시키는 결과를 보여주고 있다 [1]. 생성모델은 입력으로 노이즈를 받아, 이미지를 생성하고, 분류 모델은 생성된 이미지와 원본 이미지를 분류하는 역할을 한다. 생성 모델은 분류 모델이 원본 이미지와 생성된 이미지를 분류할 수 없도록 원본 최대한 이미지와 가까운 이미지를 생성할 수 있도록 학습한다. 또한 VDSR [6] 모델은 초고해상도 시스템에서 제안된 방법으로 입력 이미지를 그대로 학습하지 않고, 출력과 입력의 차이만을 학습함으로써 빠른 학습 속도가 높은 성능을 보여주고 있다.

딥러닝을 이용한 압축 왜곡 제거 모델인 ARCNN [5]은 JPEG [4]을 이용한 이미지 압축에서 발생하는 왜곡을

제거하기 위해 SRCNN [2]구조를 기반으로 4 개 층의 컨벌루션 뉴럴 네트워크 구조를 사용하였다.

본 논문은 비디오 영상 압축 왜곡 제거를 위해 GAN 모델을 적용한 방법을 제안하며, 압축된 영상을 입력으로 받아 압축이 제거된 영상을 확인하였고, 향후 연구 과제를 도출한다.

2. GAN을 적용한 비디오 영상 압축 왜곡 제거 모델

비디오 영상 압축 왜곡 제거를 위해 GAN 적용하여 압축된 영상을 입력으로 받아 압축이 제거된 영상을 출력하는 모델은 SRCNN 구조와 VDSR 구조를 생성 모델로 사용하였다. 생성모델의 목적 함수는 다음과 같다.

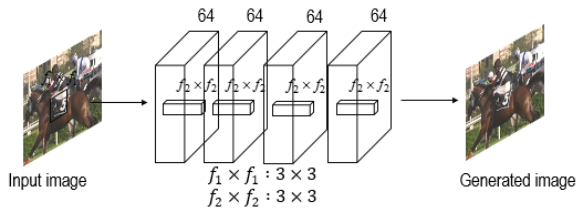
$$\min_G E_{I^{noisy} \sim P_G(I^{noisy})} [L_{MSE} + \alpha \times l_{GAN}] \quad (1)$$

생성모델은 압축 왜곡이 포함된 영상을 입력 받으며, 기존 GAN 모델의 loss에 MSE loss를 결합하여 이를 최소화하는 학습을 하며, 이때, α 는 0.0005를 사용하였다. GAN loss와 MSE loss는 다음과 같다.

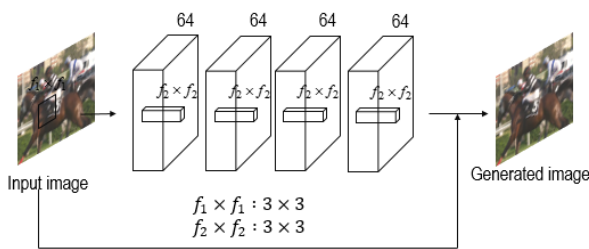
$$l_{GAN} = \sum_{n=1}^N -\log D(G(I^{noisy})) \quad (2)$$

$$l_{MSE} = \frac{1}{r^2WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{ori} - G(I^{noisy})_{x,y})^2 \quad (3)$$

SRCNN-GAN 모델은 입력 이미지를 네트워크를 통해 학습하여 출력하는 반면, VDSR 구조의 Residual-GAN 모델은 입력 이미지와 출력 이미지와의 차이만을 학습한다. SRCNN 기반의 생성모델과 Residual 기반의 생성모델의 구조는 다음 그림과 같다.



(a) SRCNN 기반의 생성 모델 구조



(b) Residual 기반의 생성 모델 구조

그림 1. 생성 모델 구조

분류 모델은 원본 이미지와 생성된 이미지를 분류하기 위한 모델로, 원본 이미지가 입력으로 들어오면 '1'을 출력하고, 생성된 이미지가 입력으로 들어오면 '0'을 출력하도록 학습하여 입력된 이미지에 따라 확률 값을 출력한다. 분류 모델의 구조는 다음과 같다.

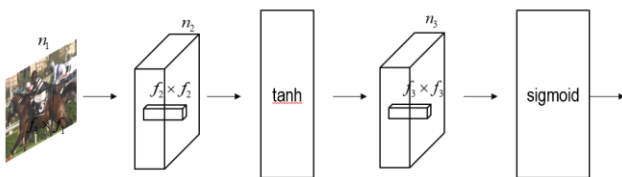


그림 2. 분류 모델 구조

3. 실험 결과

본 실험에서 사용한 데이터는 1920x1080, 30fps full HD 비디오 영상으로 HEVC 표준 소프트웨어를 사용하여 2Mbps, 4Mbps 로 압축한 총 50 장의 프레임을 사용하였다. HEVC 로 압축된 yuv 비디오 영상을 YCbCr 채널로 변환한 뒤, 휘도 성분, Y 채널을 32x32 패치 사이즈로 추출하여 학습하였다. 5 개의 층을 사용한 SRCNN 구조와 Residual 구조로 학습한 실험 결과는 다음과 같다.

표 1. 영상 PSNR 비교 (dB)

	Original	SRCNN-GAN	Residual-GAN
2Mbps 영상	34.21	32.79	34.62
4Mbps 영상	38.11	36.05	38.61

실험 결과, SRCNN 구조를 사용한 SRCNN-GAN 모델은 압축 영상보다 낮은 PSNR 값을 보여주었고, Residual-GAN 모델은 입력 영상 대비 높은 PSNR 값을 보여주었다. Residual-GAN

모델은 2Mbps 영상에 대해서 0.41dB, 4Mbps 영상에 대해서 0.5dB 높은 결과를 얻을 수 있었다. 그림 3 은 Residual-GAN 모델로 학습한 결과 이미지와 원본 이미지, 입력 이미지이다.

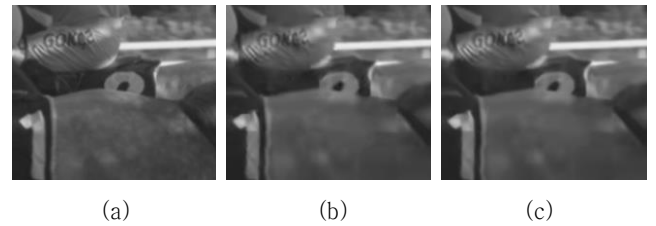


그림 3. Residual-GAN 모델로 학습한 결과 이미지. (a). 원본 이미지, (b). 입력 이미지, (c). 생성 이미지

그림 3 은 4Mbps 로 압축한 Full-HD, 1920x1080 해상도의 비디오 영상 한 프레임에서 일부분을 잘라낸 것으로, Residual-GAN 모델로 입력 이미지(b)의 압축 왜곡을 제거한 생성 이미지(c)를 보여준다. 앞서 제시한 표 1 의 결과에서 입력 영상 대비 0.5dB 가 증가한 성능을 보였지만, 시각적으로 큰 성능 향상을 확인하기 어려운 한계점을 보인다.

4. 결론 및 향후 연구 계획

본 논문에서는 비디오 영상 압축 왜곡 제거를 위해 SRCNN 구조와 Residual 구조를 GAN 모델에 적용하여 얻은 결과를 비교하였다. 제안 모델에 압축된 영상을 입력한 뒤, 압축이 제거된 영상을 얻을 수 있었다. 향후 연구 계획으로 기존에 사용한 SRCNN 구조와 Residual 구조를 개선한 새로운 모델을 제안하여 압축 왜곡이 제거 성능을 향상시키는 연구를 할 것이다.

Acknowledgement

이 논문은 2017 년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No. 2017R1A2A2A05001476)

참고문헌

- [1] Goodfellow, Ian, et al., "Generative adversarial nets." Advances in neural information processing systems, 2014.
- [2] Dong, Chao, et al., "Learning a deep convolutional network for image super-resolution." European Conference on Computer Vision. Springer International Publishing, 2014.
- [3] Sullivan, Gary J., et al., "Overview of the high efficiency video coding (HEVC) standard." IEEE Transactions on circuits and systems for video technology, pp. 1649-1668, 2012.
- [4] Wallace, Gregory K., "The JPEG still picture compression standard," IEEE transactions on consumer electronics, vol.38, no.1, 1992.
- [5] Dong, Chao, et al., "Compression artifacts reduction by a deep convolutional network." Proceedings of the IEEE International Conference on Computer Vision, 2015.
- [6] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.