

CNN 기반의 VVC 인-루프 필터 설계

문현철, 김재곤
한국항공대학교

hcmoon@kau.kr, jgkim@kau.ac.kr

CNN Based In-loop Filter in Versatile Video Coding (VVC)

Hyeonchul Moon and Jae-Gon Kim
Korea Aerospace University

요 약

본 논문에서는 새로이 시작된 비디오 압축 표준인 VVC(Versatile Video Coding)의 인-루프(in-loop) 필터링을 위한 CNN 구조를 제안한다. 제안하는 CNN 구조는 복호화된 영상을 입력으로 하고 원본 영상과 복호화된 영상의 오차를 손실함수로 사용하여 학습을 진행한다. 또한, 비디오 부호화에서의 다양한 크기의 CU(Coding Unit)를 고려한 다양한 크기의 컨볼루션 필터를 사용하여 특징을 추출하는 구조에 기반하고 있다. 실험을 통하여 제안한 CNN 기반의 필터링이 VVC의 시험모델인 VTM(VVC Test Model)의 인-루프 필터링의 성능을 개선할 수 있음을 확인하였다.

1. 서론

기존 HEVC의 2배의 압축 성능을 목표로 한 초고효율의 새로운 비디오 부호화 표준인 VVC(Versatile Video Coding) 표준화가 시작되었다. VVC 후보 기술로 새로운 부호화 통과 함께 화면내 예측, 화면간 예측, 인-루프필터 [2], [3], 변환 부호화 등에 딥러닝을 적용한 기술들도 제안되었다 [1] - [3].

딥러닝 기술들은 영상인식/분류, 객체검출 등 다양한 분야에 뛰어난 성능을 보여주고 있으며, 그 중에 CNN(Convolutional Neural Network) 알고리즘은 필터링 기능과 영상복원에 특화된 특징을 가지고 있다. 이 알고리즘을 응용한 SRCNN(Super-Resolution CNN) 구조는 저해상도의 영상을 고해상도로 복원한다. SRCNN은 기존의 영상처리 기반의 SR 기법보다 우수한 성능을 가지며 다양한 응용에서 기본 구조로 채택되고 있다 [4], [5]. 또한 기존의 SRCNN 구조를 개선하여 CNN 계층을 더 깊게 구성한 VDSR(Very Deep Super-Resolution) [6], 영상 압축의 아티팩트(artifact)를 줄이는 ARCNN(Artifact Reduction CNN) [7], 다수의 필터를 적용하도록 구성한 VRCNN(Variable-filter-size Residual-learning CNN) [8] 등이 있다.

본 논문에서는 필터링 기능과 잡음 제거에 용이한 CNN의 특징을 고려하여 비디오 압축의 블록 경계에서의 아티팩트와 양자화 잡음을 제거하기 위한 인-루프 필터에 적용하는 방안을 제시한다.

본 논문의 구성은 다음과 같다. 2절에서는 본 논문에서 제안하는 인-루프 필터에 적용할 CNN 구조를 기술하고, 3절에서는 기존의 기법과 제안하는 기법의 실험결과 및 성능비교를 제시한다. 마지막으로, 4절에서는 결론을 맺는다.

2. 제안하는 CNN 구조

제안하는 CNN 기반의 인-루프 필터의 구조는 SRCNN 및 VRCNN 구조를 기반으로 한다. 그림 1의 제안하는 구조의 입력영상은 복호화된 영상을 입력으로 한다. 학습과정에서는 출력영상을 압축전의 원본영상과 입력영상의 차이인 잔차영상을 출력하게 유도하고, 최종 출력영상은 입력영상과 잔차영상을 더한 영상으로 정의한다. 이 최종 출력영상과 입력영상의 평균최소자승오차(MSE: Mean Squared Error)를 손실함수(loss function)로 학습을 진행한다. 제안한 구조의 CNN 기반의 인-루프 필터를 적용할 때는 출력영상과 입력영상을 더한 최종 결과영상을 얻는다.

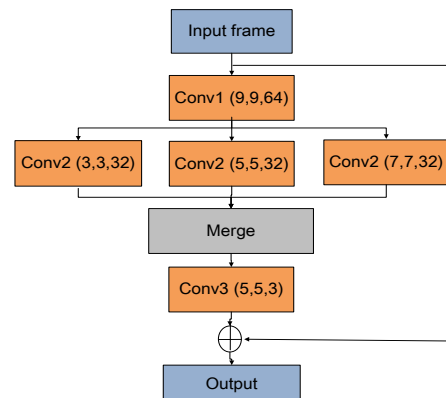


그림 1. 제안하는 CNN 구조

그림 1의 제안하는 CNN 구조는 식 (1) ~ (3)으로 표현된다. 괄호의 숫자는 (필터사이즈, 필터사이즈, 특징맵의 수)를 의미하며, Conv1, Conv2 층은 Relu 활성화 함수를 사용한다. 식 (1)에서 W_i 와 B_i 는 각각 i 계층에서의 가중치와 Biases 를 의미한다. 즉, 한 층의 출력 값은 전 층의 출력 값과 현재 컨볼루션

선 계층의 가중치를 곱한 값에 현재 층의 Biases 를 더한 값이 되며, 이를 활성화 함수(ReLu)를 통과한 값이 다음 계층의 입력 값으로 된다. 식 (2)는 그림 1 의 Merge 층을 나타낸 것으로, Conv2 층의 출력 값이 각각 3, 5, 7 의 필터 크기를 사용하는 Conv2 층의 평균 값으로 설정됨을 나타낸 것이다. 이와 같이 제안한 CNN 구조는 다양한 컨볼루션 필터 크기를 사용하고 이를 다시 합치는 방식이다. 식 (3)은 마지막 층의 출력 Y 에 대한 식이며, 이 때는 활성화 함수 Relu 를 사용하지 않고 최종 결과물인 잔차영상을 생성한다.

$$F_i = \max(0, W_i * F_{i-1} + B_i), i = 1, 2 \quad (1)$$

$$F_2 = Average(F_{2,3} + F_{2,5} + F_{2,7}) \quad (2)$$

$$Y = W_3 * F_2 + B_3 \quad (3)$$

3. 실험 결과

본 논문에서 제안하는 CNN 기반의 인-루프 필터의 성능을 확인하기 위한 실험은 All Intra 모드에서 6 개의 시퀀스 BasketballDrill, BQMall, PartyScene, BlowingBubbles, BQSquare 그리고 PeopleOnstreet 를 학습 데이터를 사용했으며, 다양한 QP 의 값(27, 32, 37)으로 부호화한 시퀀스를 혼합하여 학습을 진행하였다. 테스트 시퀀스로는 테스트 시퀀스와 겹치지 않도록 BasketballDrill, BQMall 의 각각의 1~100 번째 프레임 을 사용하였다. 실험결과 는 VVC(Versatile Video Coding)의 참조 SW 코덱인 VTM-1.0 의 결과를 기준으로 비교하였다.

표 1. 제안 기법의 성능평균 PSNR

방법	PSNR(dB)
인-루프 적용 전	32.74
인-루프 적용 후	33.10
VRCNN [5]	33.32
제안하는 기법 (CNN)	33.34

표 1 의 실험결과 는 각 경우에서의 실험에 사용한 4 개의 QP 22, 27, 32, 37 에서의 평균 PSNR 을 나타낸 것으로, 제안하는 CNN 기법은 VCC 의 인-루프 필터 대비 평균 0.24dB 가 향상됨을 확인할 수 있었으며, 필터 크기를 다양하게 적용할 수 있는 기존의 VRCNN 과 비교했을 때 약 0.02dB 의 향상을 확인할 수 있다.

표 2. 각 QP 값에 따른 제안한 기법 성능(PSNR 이득)

QP	제안 기법
22	+0.03
27	+0.17
32	+0.22
37	+0.35

표 2 의 결과와 같이 높은 QP 에서 상대적으로 제안 기법의 성능 개선이 큰 것을 확인할 수 있다.

그림 2 는 기존의 기법과 제안된 기법의 주관적 화질을 비교한 것으로 CNN 기반의 제안 기법이 보다 개선된 화질을 얻을 수 있음을 확인할 수 있다.

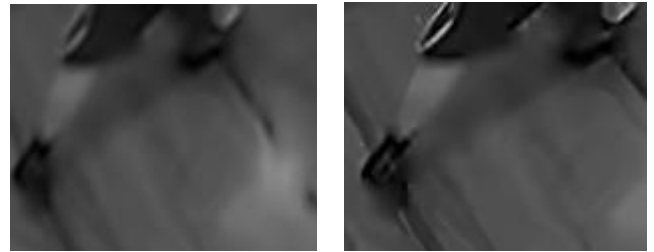


그림 2. 주관적 화질 비교(BasketballDrill, 6 번째 프레임)

4. 결론

본 논문에서는 CNN 기반의 VVC 인-루프 필터를 제시하였다. 다양한 크기의 CU 를 고려하여 다양한 크기의 필터를 포함한 CNN 구조를 사용하였다. 실험결과 기존의 VCC 의 인-루프 필터와 비교했을 때 0.24 dB PSNR 이득을 얻을 수 있었고, 주관적 화질 비교에서도 두드러진 화질 개선을 확인할 수 있었다.

제안 기법은 다양한 QP 에서 성능개선을 보이지만 낮은 QP 에 대해서는 상대적으로 개선 정도가 낮으며 이에 대한 개선이 필요하다. 또한, BD-Rate 의 성과와 보다 다양한 시퀀스에서의 실험을 통한 성능 확인과 Random Access 및 Low Delay 의 부호화 모드에서의 성능을 확인할 예정이다. 또한, 향후 JVET 에 제안된 VVC 의 인-루프 필터와 성능 비교를 진행할 예정이다.

감사의 글

이 논문은 본 연구는 산업통상자원부 국가표준기술원에서 시행한 국가표준기술력향상사업[10084981, 인공지능 기반의 패턴인식 기술 국제표준화 개발]의 지원을 받아 수행된 연구임.

참고 문헌

- [1] S. Liu et al, "JVET AHG report: Neural Networks in Video Coding," JVET document, JVET-J0008, Apr. 2018.
- [2] C. W. Hsu, et al, "Description of SDR video coding technology proposal by MediaTek," JVET document, JVET-J0018, Apr. 2018.
- [3] Zhou et al, "Convolutional Neural Network Filter (CNNF) for intra frame," JVET document, JVET-I0022, Apr. 2018.
- [4] W. S. Park and M. Kim, "CNN-based in-loop filtering for coding efficiency improvement," In. Proc. IVMSIP 2016, July 2016.
- [5] C. Dong, C. C. Loy, K. He, K., and X. Tang, "Image super-resolution using deep convolutional networks," IEEE Trans. Pattern Analysis Machine Intelligence, vol. 38, no. 2, Feb. 2016.
- [6] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," In Proc. CVPR 2016, Jan. 2016.
- [7] Y. Dai, D. Liu, and F. Wu, "A Convolutional neural network approach for post-processing in HEVC intra coding," arXiv:1703.03502, 2017.
- [8] Ke YU, et al. "Deep convolution networks for compression artifacts reduction," arXiv:1608.02778, 2016.