Deep Learning Network를 이용한 Video Codec에서 휘도성분 노이즈 제거

김양우 이영렬

세종대학교

ywkim@sju.ac.kr yllee@sejong.ac.kr

Luma Noise Reduction using Deep Learning Network in Video Codec

Kim, Yang-Woo Lee, Yung-Lyul

Sejong University

요약

VVC(Versatile Video Coding)는 YUV 입력 영상에 대하여 Luma 성분과 Chroma 성분에 대하여 각각 다른 최적의 방법으로 블록분할 후 해당 블록에 대해서 화면 내 예측 또는 화면 간 예측을 수행하고, 예측영상과 원본영상의 차이를 변환, 양자화하여 압축한다. 이 과정에서 복원영상에는 블록화 노이즈, 링잉 노이즈, 블러링 노이즈 발생한다. 본 논문에서는 인코더에서 원본영상과 복원영상의 잔차신호에 대한 MAE(Mean Absolute Error)를 추가정보로 전송하여 이 추가정보와 복원영상을 이용하여 Deep Learning 기반의 신경망 네트워크로 영상의 품질을 높이는 방법을 제안한다. 복원영상의 노이즈를 감소시키기 위하여 영상을 32×32블록의 임의로 분할하고, DenseNet기반의 UNet 구조로 네트워크를 구성하였다.

1. 서론

VVC(Versatile Video Coding)은 차세대 표준 비디오 코딩으로 ITU-T VCEG와 ISO/IEC MPEG이 JEVT(Joint Video Exploration Team)을 2015년 10월에 결성하고 2018년 4월부터 HEVC(High Efficiency Video Coding)이후의 비디오 코딩 표준화를 목표로 시작하였다.

VVC는 영상을 128×128 크기의 CTU(Coding Tree Unit)으로 분할하고, 이를 다시 최적의 블록으로 분할하여 화면 내 예측의 경우 주변 복원화소, 화면 간 예측의 경우 이전 영상의 복원화소를 이용하여 예측블록을 생성한다. 이후 예측영상과 원본영상의 차이인 잔차신호를 DCT-2 등으로 변환하고 양자화를 거치고 엔트로피 코딩으로 압축하여 디코더에 전송한다. 이 과정에서 잔차신호에 양자화로 인한 고주파수 신호 누락으로 인해 그림 1과 같이 Deblocking Arifact, Rining Artifact, Blur Artifact등의 노이즈가 발생한다. VVC에서는 이러한 노이즈를 복원하고 주관적 화질을 높이기 위하여 복원영상에 대하여 인루프필터를 적용한다. VVC의 인루프필터는 Deblocking Filter, Sample Adaptive offset, Adaptive Loop Filtering로 이루어져 있다.

최근 영상처리, 컴퓨터 비전 분야에서 기존의 연구에 Deep Learning을 이용하려는 연구들이 진행되었다. 특히 이러한 Deep Learning 기법 중 Convolutional Neural Network(CNN) 은 영상의 특성을 이용하여 Convolution layer를 중첩하여 영상의 특징들을 이용하여 뛰어난 성능을 보인다.

그림 1. VVC에서 발생하는 여러 가지 종류의 노이즈

2. CNN기반의 인루프필터

2.1 입력데이터 구축

네트워크를 학습시키기 위하여 Trainning set과 Test set을 구축하기 위하여 Trainning set을 위하여 22개의 비디오 시퀀스 60f/s 10초 분량, Test set을 위하여 6개의 비디오 시퀀스 60f/s 10초분량을 VVC의 Reference Software인 VTM 4.0으로 8 Frame씩 건너뛰면서 화면 내 예측 모드로 코딩하였다. 이후 그림 2와 같이 원본영상과 복원영상을 32x32 크기로 분할하여 각각 학습의 Ground Truth와 Input으로 구성하였다. VTM의 Quantization Parameter은 37로 고정하였다. 해당 방법으로 Trainning Set 6만개, Test Set 6천개를 구성하였다.

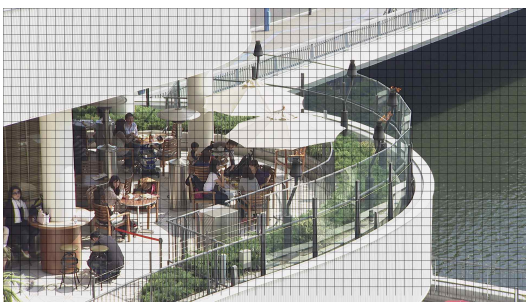

그림 2. Sequence 분할의 예시

이후 원본영상과 복원영상의 MAE(Mean Absolute Error)를 계산하여 이를 표1과 같이 8단계로 분류하였다. 분류된 MAE를 네트워크의 Input으로 구성하기 위하여 복원영상의 크기(32×32)와 같은 크기의

8개의 추가 채널을 만들고, 그림 2와 같이 해당 MAE로 분류된 채널만 해당 채널을 1로 채우고, 나머지 채널을 0으로 채우는 One-hot Encoding 방법으로 구성하였다.

이는 다른 이미지 디노이징 기법들과 다르게 Video Coding은 원본 영상에 대한 정보를 인코더에서 디코더로 전송 할 수 있기에 가능한

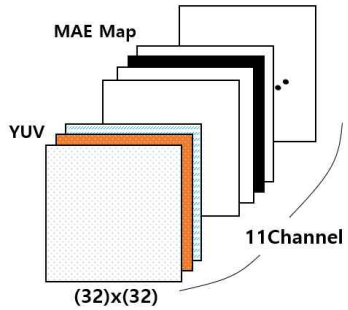

그림 3. Input Data 구성의 예시

방법이다.

2.2 제안하는 전체 네트워크 구조

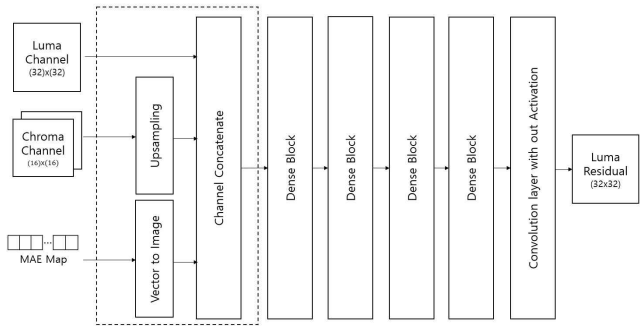

그림 4. 제안하는 방법의 전체 네트워크 구조

제안하는 네트워크 구조는 복원영상 YUV(4:2:0)에 대하여 색차성분을 Upsampling하고, Luma 성분의 원본영상에 대한 MAE를 8단계로 분류하여 MAE Map을 구성한다. 이후 MAE Map을 Luma Channel의 이미지의 크기에 맞게 Upsampling하고, 채널 단위로 Concatenate 하여 최종 입력데이터를 구성한다. 이후 Dense Network[1]의 Dense Block구조로 Convolution layer를 구성하여 최종 Output으로 Luma 성분에 대한 Residual 출력하도록 Network를 구성한다.

3. 실험결과

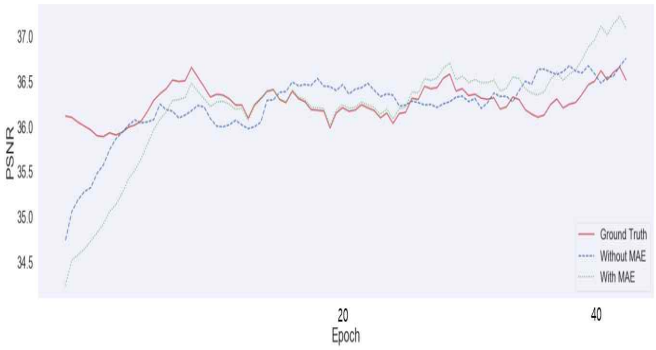

그림 5. Trainning 과정에서 PSNR 변화 추이

본 논문은 Deep Learning기반으로 원본영상과 복원영상의 MAE를 이용하여 네트워크 성능을 올리기 위한 구조를 제안하였다. 해당 구조를 검증하기 위하여 MAE가 없이 복원영상만으로 원본영상을 복원하는 구조의 네트워크와 성능 비교를 하였으며 실험결과 TestSet에서 복원영상에 대하여 0.21dB의 PSNR 향상과 MAE가 없는 복원영상에 대하여 0.12dB의 PSNR 향상을 보였다.

4. 결론

제안하는 32×32블록에 대하여 원본영상과 복원영상의 MAE와 Deep Learning Network를 이용하여 복원영상의 노이즈를 제거하는 방법은 기존 VVC 4.0의 복원영상 대비 0.21dB의 PSNR 향상을 보인다.

감사의 글

이 논문은 일부 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No. 2017-0-00072, 초실감 테라미디어를 위한 AV 부호화 및 LF 미디어 원천기술 개발)

참고문헌

[1] Huang Gao, Liu Zhuang, Laurens van der Maaten, Q. Weinberger Kilian, "Densely connected convolutional networks", 2017 IEEE Conference on Computer Vision and Pattern Recognition CVPR, pp. 2261-2269, 2017.

[2] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm, T. Wiegand, "High efficiency video coding (HEVC) text specification draft 7", document JCTVC-I1003, Jul. 2012 K. D. Hong and K. J. Lim, "A study on image understanding," IEEE Trans. Image Processing, vol. 3, no. 2, pp. 1-10, 2007.

[3] Zhang, K., Zuo, W., Chen, Y., et al.: 'Beyond a Gaussian denoiser: residual learning of deep Cnn for image denoising', IEEE Trans. Image Process., 2017, 26, (7), pp. 3142 - 3155.