블록정보를 이용한 CNN기반 인 루프 필터

*김양우 **이영렬

세종대학교

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

CNN-based In-loop Filtering Using Block Information

*Yangwoo Kim **Yung-lyul Lee

Sejong University

요약

VVC(Versatile Video Coding)는 입력 YUV영상을 CTU(Coding Tree Unit)으로 분할하고, 다시 이를 QTBTTT(Quad Tree, Binary Tree, Ternery Tree)로 최적의 블록으로 분할하고 각각의 블록을 공간적, 시간적 정보를 이용하여 예측하고 예측블록과 원본블록의 차분신호를 변환, 양자화를 통해 전송한다. 이를 위해 여러가지 인코딩정보가 디코더에 전송되며 이를 이용하여 디코더는 인코더와 똑같은 순서로 영상을 복원 할 수 있다. 본 논문에서는 이러한 VVC 인코더에서 반드시 전송하는 정보를 추가적으로 이용하여 딥러닝 기반의 Convolutional Neural Netwrok로 영상의 압축률 및 화질개선 하는 방법을 제안한다.

1. 서론

VVC(Versatile Video Coding)은 차세대 비디오 표준으로 ISO/IEC MPEG과 ITU-T VCEG이 JVET(Joint Video Exploration)을 2015년 10월에 결성하고 2018년 4월부터 HEVC(High Efficiency Video Coding)[1]을 잇는 비디오 표준화 코딩을 목표로 시작하였다.

VVC는 영상을 CTU(Coding Tree Unit)으로 분할하고 이를 QTBTTT(Quad Tree, Binary Tree Ternery Tree) 블록으로 분할한다. 이후 화면 내 예측, 화면 간 예측의 모드에 따라서 각각의 블록을 공간적, 시간적 정보를 이용하여 예측하고 예측신호와 원본신호의 차분신호를 만들어 이를 DCT-2, DST-7등의 변환 후 양자화를 통해 압축하여 디코더에 전송한다. 디코더는 이러한 블록분할 정보, 예측모드에 관한 정보, 변환방법에 관한 정보, 양자화 수준에 관한 정보 등 무수히 많은 정보를 토대로 신호를 복원한다.

한편 복원된 영상은 여전히 양자화로 인한 에러가 있고, 원본영상과 많은 차이가 있다. 이러한 복원된 영상의 에러는 이후 화면 간 예측에서도 여전히 남아 있어서 전체 비디오의 압축 성능을 떨어트리는 원인이 되고 또한 주관적 화질 또한 낮아진다. 따라서 HEVC와 VVC는 인 루프 필터 과정을 통해 객관적, 주관적 화질을 올리고, 화면 간 예측에서 사용되는 참조픽쳐에 대한 에러를 줄임으로써 압축효율을 향상시킨다.

최근 영상처리, 컴퓨터비전 분야에서 딥러닝은 매우 성공적인 방법이다. Classification, Segmentation, Object Detection등의 많은 세부분야에서 성공적인 성능 향상을 보이고 있으며, 또한 Image Deblur, Denoise등의 Image Enhancement 분야에서도 독보적인 성능 향상을 보이고 있다. 본 논문에서는 이러한 딥러닝 기반의 네트워크 학습을 이용하여 비디오 압축 에러를 효과적으로 제거하는 방법을 제안한다.

2. CNN기반의 인루프 필터

2.1 VVC에서의 여러가지 블록정보

VVC에서 CTU 혹은 CU(Coding Unit)은 여러가지 인코딩, 디코딩 Flag를 가지고 있다. QP(Quantization Parameter)는 해당 블럭을 어느정도의 양자화 수준으로 압축 할 것 인지를 지칭한다. 예측모드는 해당 블럭을 화면 내 예측 또는 화면 간 예측을 사용할 때 화면 내 예측의 경우 0~66번의 모드 중에 몇 번 째 모드를 사용했는지를 지칭하고, 화면 간 예측의 경우 Merge, Affine 등의 모드 인덱스를 의미한다. Depth란 CTU에서 한번씩 QTBTTT중의 하나로 분할 될때 마다 블럭의 Depth는 한 단계씩 올라가게 된다. 블럭의 차분신호를 어떤 방식으로 변환할 것 인지도 블럭마다 인코더에서 디코더로 전송된다. VVC에서는 DCT-2, DST-7, DST-8의 3가지 방법으로 수평 수직 방향 각각 변환을 수행한다.

2.2 네트워크 입력 구성

일반적으로 이러한 Image Denoise를 딥러닝 기반의 모델로 수행할때 Convolutional Neural Net(CNN)중에서 모든 layer가 Convolution 연산으로 이루어진 Fully Convolutional Network(FCN)을 사용하는 것이 효과적인 것으로 알려져 있다. 이러한 FCN은 MLP(Multi-Layer Perceptron)등과 같은 Input과 Ouput의 크기가 지정되어있는 모델과 다르게 Input과 Output의 크기를 비교적 자유롭게 사용 할 수 있다. 이러한 FCN 특성을 이용하여 네트워크의 Traning은 비교적 작은 Sub-Image(Patch)에서 진행하고, Test는 다른 크기의 이미지에서 사용 할 수 있다. FCN에 블록정보를 효과적으로 입력하기 위하여, 블록의 가로세로 크기와 같은 크기의 마스크를 만들고 해당 블

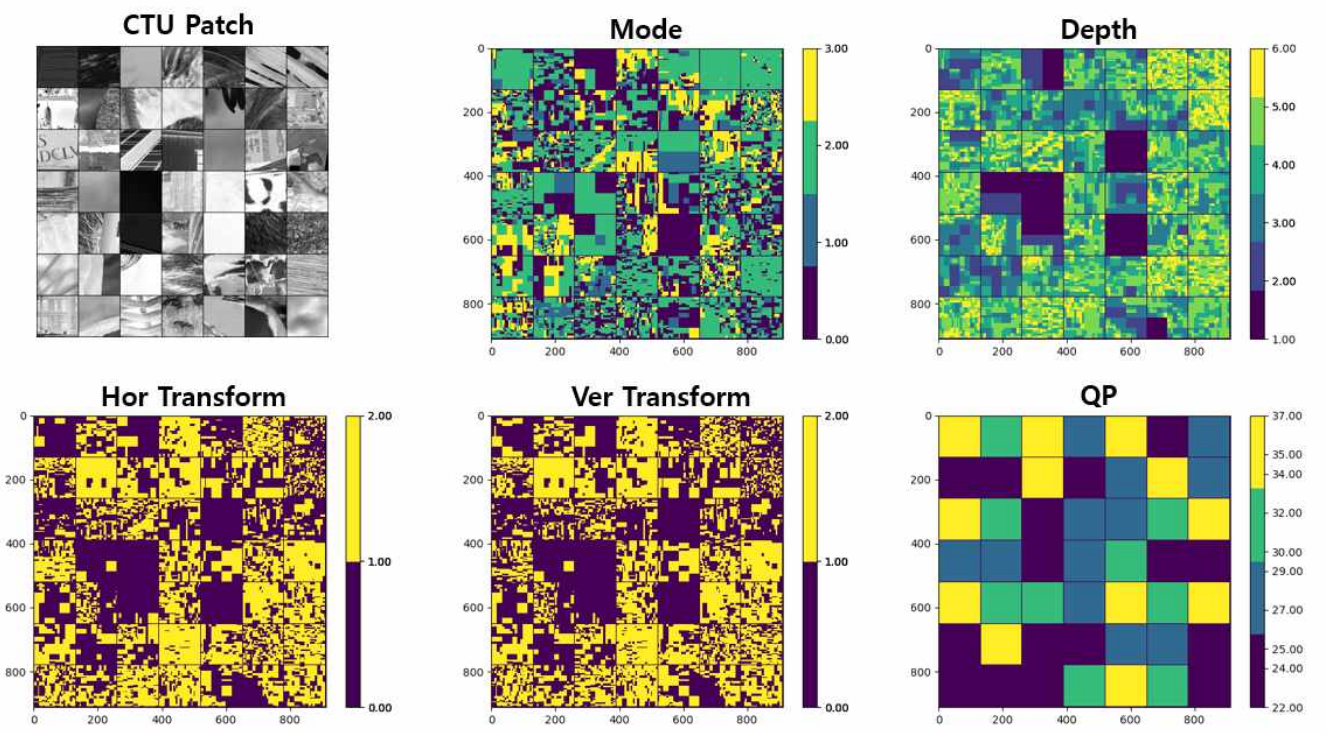

그림 1. CTU 임의 Patch별 블록 정보의 예시.

록의 픽셀값마다 해당하는 정보를 넣어 원래 YUV 픽셀 정보와 채널 단위로 연결하였다. 그림 2와 같이 QP, Mode, Vertical transform mode, Horizontal mode, Depth map의 정보를 YUV 블록과 똑같은 크기로 만들어서 채널 단위로 연결 하였다.

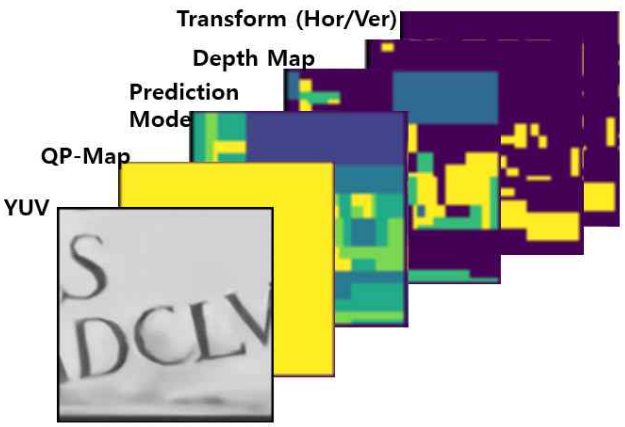

그림 2. 네트워크 입력의 예시

2.3 네트워크 구조

네트워크 구조는 복원영상 YUV 4:2:0에 대하여 색차성분을 Upsampling하고 각각의 블록정보벡터들을 YUV와 같은 크기의 Mask로 구성하여 YUV영상과 채널 단위로 연결하여 입력을 만든다. 그 이후 Residual Dense Block[2]구조로 FCN을 연결하여 입력 복원영상과 원본영상과의 차이를 출력하도록 구성하였다.

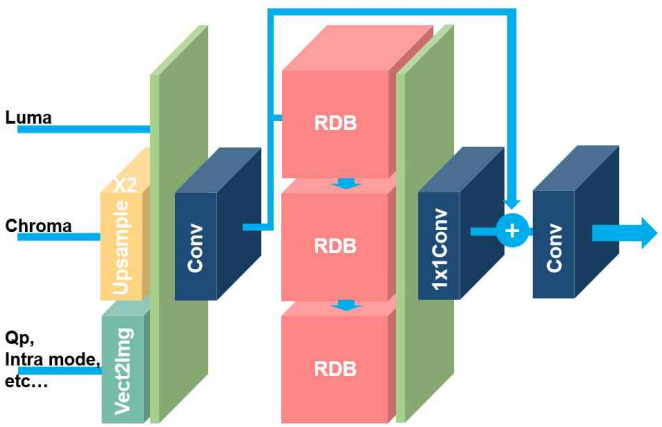

그림 3. 전체 네트워크 구조.

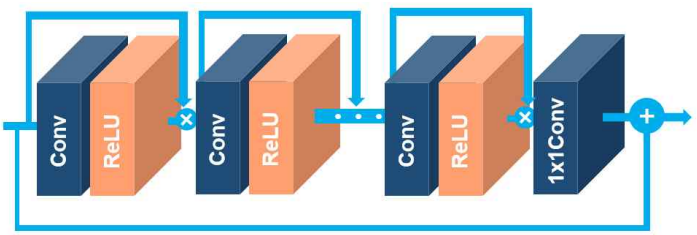

그림 4. Residual Dense Block의 구조.

3. 실험결과

VVC 인루프 필터에서 ALF 이후에 위 모듈을 CTU단위로 On/Off flag를 전송하여 사용하였다. VVC에서 블럭 정보 입력을 이용

한 네트워크는 VTM6.1 대비 4.23%의 성능 향상을 보인다.

4. 결론

본 논문은 기존의 디코더에서도 이용가능한 VVC의 블럭단위의 정보를 이용하여 기존의 FCN구조의 딥러닝 기반 네트워크의 성능 향상을 제안한다. 기존 VTM 6.1 대비 4.23%의 성능 향상을 보인다.

감사의 글

이 논문은 일부 2019년도 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] 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.
[2]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.