

## VVC 를 위한 CNN 기반의 변환계수 적응적 화질개선 기법

김지훈, 임성균, 박도현, 김재곤

한국항공대학교

{dhpark, sglim}@kau.kr, jgkim@kau.ac.kr

## CNN-Based Transformed Coefficient-Adaptive Enhancement for VVC

Ji-Hun Kim, Sung-Gyun Lim, Dohyeon Park, and Jae-Gon Kim

Korea Aerospace University

## 요 약

최근 VVC(Versatile Video Coding) 표준 완료 이후 JVET(Joint Video Experts Team)은 NNVC(Neural Network-based Video Coding) AhG(Ad-hoc Group)을 구성하고 인공지능을 이용한 비디오 압축 기술들을 탐색하고 있다. 본 논문에서는 VVC 복원 영상의 DCT 계수를 기반으로 복원 영상을 분류하고, 분류된 각 클래스에 따라 적응적으로 CNN(Convolutional Neural Network) 기반의 화질 개선을 수행하는 VVC 후처리 기법을 제안한다. 실험결과, 제안기법은 AI(All Intra) 부호화 모드에서 1.23% BD-rate 이득을 보였다.

## 1. 서론

최근, HEVC(High Efficiency Video Coding) 대비 월등히 개선된 압축 성능을 갖는 VVC(Versatile Video Coding) 표준화가 완료되었다[1], [2]. JVET(Joint Video Experts Team)에서는 VVC 표준 완료 이후, VVC 를 확장할 수 있는 인공지능 기반의 비디오 부호화 기술의 잠재성 및 성능을 확인하기 위해 애드혹 그룹(AHG11: Neural network-based video coding)을 구성하고 EE(Exploration Experiments)를 설정하여 관련된 기술들을 탐색, 검증하고 있다[3], [4]. 특히, 딥러닝 기반의 인-루프(in-loop) 필터링 기술들에 대한 연구들이 활발히 수행되고 있다.

VVC 에 채택된 인-루프 필터인 ALF(Adaptive Loop Filter)는 영상을 4x4 블록으로 분할하고 각 블록에 대해 방향성 및 복잡도를 기반으로 최대 25 가지의 클래스로 분류하고 각 클래스에 해당하는 필터를 이용하여 화질을 개선한다[3]. 본 논문에서는 VVC 코덱으로 복원된 영상을 32x32 블록으로 분할하고 각 블록의 DCT(Discrete Cosine Transform) 계수를 기반으로 4 개의 클래스로 분류하고, 클래스에 따라 적응적으로

CNN(Convolutional Neural Network)을 적용하여 복원 영상의 화질을 개선하는 후처리 기법을 제안한다.

## 2. 제안하는 화질개선 기법

## (1) 변환 계수 기반 영상 분류

제안하는 기법의 영상 분류 및 화질 개선은 32x32 블록 단위로 수행한다. 영상 분류를 위해 VVC 코덱으로 복원된 영상을 32x32 로 분할하고 각 블록에 대해 2D-DCT 연산을 수행한다. 다음, 각 변환 계수의 위치에 따른 가중치를 두고, 그림 1 에서 보이는 4 개의 DCT 영역(Red, Orange, Yellow, Green 으로 표시된 4 개 영역) 별 변환 계수의 합이 가장 큰 영역에 대한 인덱스로 해당하는 블록의 클래스를 분류한다.

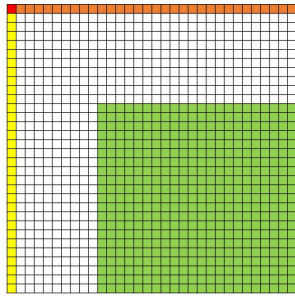


그림 1. DCT 변환 계수 분할 영역(32x32 블록)

### (2) CNN 기반 화질 개선 네트워크

제안하는 기법에서 사용하는 화질개선을 위한 CNN 기반의 네트워크의 구조는 그림 1 과 같다[5]. 제안하는 인공신경망은 총 10 개의 계층을 갖으며 휘도 성분에 대한 복원영상과 원본영상의 차이를 출력함으로써 화질개선을 수행한다. 제안하는 화질개선 기법을 수행하기 위해서 4 개의 클래스에 대한 분류와 4 개의 QP(22, 27, 32, 37)의 조합으로 동일한 네트워크 구조를 갖는 16 개의 모델을 이용한다. 학습 데이터는 BVI-DVC 데이터셋을 사용하였으며[6], MSE(Mean Square Error) 손실함수와 Adam 최적화 기법을 이용하여 각 모델을 학습하였다.

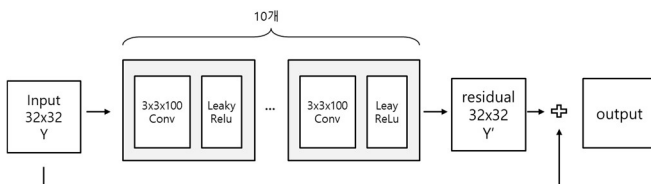


그림 2. CNN 기반 화질 개선 네트워크 구조

### 3. 실험결과

본 논문에서는 제안하는 화질 개선 기법을 VVC 참조 소프트웨어인 VTM11.0 을 이용하여 JVET CTC(Common Test Condition) 및 All Intra 환경에서 실험을 진행하였다. 표 1 은 VTM11.0 대비 제안기법에 대한 BD-rate 결과이다. 실험결과, 평균 1.23%의 성능 향상을 보였다.

표 1. 제안기법의 성능(JVET CTC, All Intra)

Class	Y	U	V
<b>Over VTM-11.0 in AI</b>			
Class A1	-0.19%	0.00%	0.00%
Class A2	-1.30%	0.00%	0.00%
Class B	-0.67%	0.00%	0.00%
Class C	-1.43%	0.00%	0.00%
Class D	-2.19%	0.00%	0.00%
<b>Overall</b>	<b>-1.23%</b>	<b>0.00%</b>	<b>0.00%</b>

### 5. 결론

본 논문에서는 32x32 블록 단위의 DCT 변환 계수를 이용한 분류 및 CNN 기반의 화질 개선 기법을 제시하였다. 제안하는 화질개선 기법을 VTM11.0 로 압축한 복원영상에 적용하였으며, 1.23%의 BD-rate 이득이 있음을 확인하였다. 제안기법은 비디오 부호화로 인해 열화된 영상을 그 특성에 따라 적응적으로 CNN 기반의 네트워크를 적용함으로써 보다 효율적인 복원 화질을 개선할 수 있음을 확인하였다.

### Acknowledgement

이 논문은 2017 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2017-0-00486, 360 VR 영상정보의 공간 기하학적 특성을 이용한 화질개선 및 트랜스코딩 기술개발)

### 참 고 문 헌(References)

- [1] High Efficiency Video Coding, Version 1, Rec. ITU-T H.265, ISO/IEC 23008-2, Jan. 2013.
- [2] Versatile Video Coding, ISO/IEC FDIS 23090-3, Jul. 2020.
- [3] S. Liu, E. Alshina, J. Pfaff, M. Wien, P. Wu and Y. Ye, "JVET AHG report: Neural-network-based video coding," Joint Video Experts Team of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-V0011, Apr. 2021.
- [4] A. Browne, T. Hashimoto, H. Jhu, D. Rusanovskyy, K. Kawamura and T. Zhou, "EE1: Summary Report on Exploration Experiments on Neural Network-based Video Coding," Joint Video Experts Team of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-V0023, Apr. 2021.
- [5] J. Chen, H. Wang, A. Kotra and M. Karczewicz, "AHG11: In-loop filtering with convolutional neural network and large activation," Joint Video Experts Team of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-U0104, Dec. 2020.
- [6] D. Ma, F. Zhang and D. Bull, "BVI-DVC: A training database for deep video compression," 2020, arXiv:2003.13552. (<https://data.bris.ac.uk/data/dataset/3hj4t64fkbrgn2ghwp9en4vhtn>)