

## 완전연결계층 기반의 다중 모델을 이용한 화면내 예측

김민재, 문기화, 박도현, \*권형진, 김재곤

한국항공대학교, \*한국전자통신연구원

{alswo12745, ghmoon, dhpark}@kau.kr, kwonjin@etri.re.kr, jgkim@kau.ac.kr

### Intra Prediction Using Multiple Models Based on Fully Connected Layer

Minjae Kim, Gihwa Moon, Dohyeon Park, \*Hyoungjin Kwon, and Jae-Gon Kim

Korea Aerospace University, \*ETRI

#### 요 약

딥러닝 기술과 하드웨어의 발전으로 다양한 분야에서 인공지능경망과 관련한 연구가 활발히 진행되고 있다. 비디오 코덱 부분에서도 딥러닝 기술을 적용하는 부호화 기술이 많이 연구되고 있다. 본 논문은 최근 완료된 VVC에 채택된 신경망 기반의 기술인 MIP(Matrix Weighted Intra Prediction)를 확장하여 보다 깊은 계층의 모델로 학습된 새로운 화면내 예측 모델을 제안한다. 기존 VVC의 MIP의 성능과 비교하기 위하여 기존 MIP 모델과 제안하는 다중 완전연결계층(Fully Connected Layer) 화면내 예측 모델을 HEVC(High Efficiency Video Coding)에 적용하여 그 성능을 비교하였다. 실험결과 제안기법은 VVC MIP 대비 0.08 BD-rate 성능 향상을 보였다.

#### 1. 서론

최근 MPEG 과 VCEG 공동으로 구성된 JVET(Joint Video Experts Team)은 다양한 비디오 서비스에 적합하며 HEVC(High Efficiency Video Coding) 대비 월등히 개선된 압축 성능을 갖는 VVC(Versatile Video Coding) 표준화를 2020 년 7 월 완료하였다[1], [2].

JVET 은 비디오 코덱의 성능의 향상을 위해 MIP(Matrix Weighted Intra Prediction) 및 LFNST(Low Frequency Non-Separable Transform)와 같은 학습된 신경망 모델을 이용한 부호화 기술들을 채택하였다. 또한 VVC 를 확장할 수 있는 신경망 기반의 비디오 부호화 기술의 잠재성을 확인하기 위한 AhG11(Ad-hoc Group11)을 구성하고 관련 기술을 탐색하고 있다[3].

본 논문은 VVC 에 채택되어 있는 MIP 의 다중 예측 모델을

참조하여 보다 깊은 네트워크를 갖는 다중 예측 모델을 학습하였으며, 그 부호화 성능을 측정하였다.

#### 2. VVC MIP 및 완전연결계층 기반 다중 모델

##### (1) VVC MIP

VVC MIP 는 참조샘플 감소, MIP 예측 및 예측샘플 보간의 순서로 크게 3 단계로 구성된다. 참조샘플 감소 및 예측샘플 보간은 MIP 의 복잡도 감소를 위한 단계이며 실제 예측샘플은 MIP 예측단계에서 생성된다. MIP 예측을 위한 모델은한 개 1 계층의 완전연결계층(Fully Connected Layer: FLC)을 갖는 복수개의 네트워크로 구성되며 8x8 크기의 블록에 대해 총 8 개의 예측 네트워크를 갖는다.

## (2) 완전연결계층 기반 네트워크

본 논문에서는 그림 1 와 같이 예측하고자 하는 블록의 주변 참조샘플을 4 계층의 완전 연결계층에 입력하여 예측 블록을 생성한다. 본 논문에서 사용되는 참조 샘플 라인의 수는 4이다.

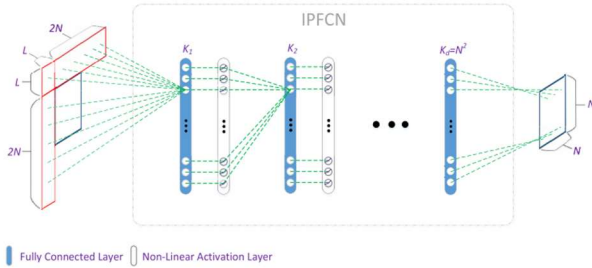


그림 1. 여러 계층의 완전연결계층 기반 네트워크[4]

그림 1 와 같은 네트워크 구조 갖는 복수개의 모델을 학습할 때 사용하는 손실함수  $L(S, \hat{k})$ 는 다음과 같다[4].

$$L(S, \hat{k}) = loss_{pred}(\hat{k}) + loss_{bit}(\hat{k}) + loss_{mode}(\hat{k}) \quad (1)$$

$$\hat{k} = \underset{k}{\operatorname{argmin}} loss_{pred}(k) \quad (2)$$

$$loss_{pred}(k) = |c_k| + \alpha g(\beta |c_k| - \gamma) \quad (3)$$

$$c_k = T \cdot (o - p_k) \quad (4)$$

$p_k$ 는  $k$  모드에서의 예측 샘플,  $T$ 는 2D DCT-2 변환 행렬이며  $c_k$ 는  $k$  모드에서 변환 계수이다.  $g$ 는 로지스틱 함수를 의미한다( $g(x) = 1/(1 + e^{-x})$ ).  $\hat{k}$ 는  $loss_{pred}$  기준으로 최적의 비용을 갖는 모드를 의미하며  $\hat{k}$ 에 대한  $loss_{pred}$ , 모드 예측 모델 결과에 대한 크로스 엔트로피 손실인  $loss_{mode}$ , 그리고 실제 모드를 전송할 때 발생하는 비트수인  $loss_{bit}$  더해서 최종 손실함수로 사용한다.

본 논문에서는 8 개의 완전연결계층 기반 네트워크 구성하고 학습하였다. 학습은 8x8 블록에 대해서 진행하였으며 COCO 데이터셋 및 Adam 최적화 방법을 이용하였다[5].

## 3. 실험결과

본 논문에서 제안하는 완전연결계층 기반 다중 모델의 비디오 부호화 성능을 평가하기 위해 제안기법과 VVC MIP 를 HEVC 의 추가적인 화면내 예측 모드로 구현하였다. HEVC 참조 소프트웨어인 HM16.19 를 8x8 블록으로 제한하여 VVC MIP 를 구현한 후, AI(All Intra) 부호화 환경에서 HEVC 대비 0.29% BD-rate 성능향상을 보였다. VVC MIP 대비 제안기법은 AI(All Intra) 부호화 환경에서 Y, Cb, Cr 에 대해서 각각 0.08, 0.06, 0.04 의 BD-rate 성능 향상을 보였다.

표 1. 제안기법의 성능(JVET CTC, All Intra)

Class	Y	U	V
<b>Over HM16.19 with VVC MIP</b>			
Class B	-0.15%	-0.10%	-0.05%
Class C	-0.05%	-0.06%	-0.03%
Class D	-0.03%	-0.01%	-0.03%
<b>Overall</b>	<b>-0.08%</b>	<b>-0.06%</b>	<b>-0.04%</b>

## 4. 결론

본 논문에서는 VVC 의 MIP 보다 깊은 네트워크 구조를 갖는 완전연결계층 기반 다중 모델이 기존 VVC MIP 대비 비디오 부호화 성능을 향상함을 확인하였다. 실험에서는 제안 다중 모델과 VVC MIP 를 HEVC 에 구현하였으며, 실험결과, 8x8 블록에 대해 VVC MIP 대비 Y, Cb, Cr 에 대해서 각각 0.08, 0.06, 0.04 BD-rate 성능 향상을 보였다. 또한, 다양한 블록 크기에 대해 수행할 수 있도록 모델을 확장하여 비디오 부호화의 블록 분할을 반영하는 추가적인 연구가 필요하다.

## Acknowledgement

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2017-0-00072, 초실감 테라미디어를 위한 AV 부호화 및 LF 미디어 원천 기술 개발).

## 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] P. Helle, J. Pfaff, M. Schäfer, R. Rischke, H. Schwarz, D. Marpe, and T. Wiegand, "Intra Picture Prediction for Video Coding with Neural Networks," In Proc. Data Compression Conference 2019, IEEE, Mar. 2019.
- [5] T. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. Zitnick, and P. Dollár, "Microsoft COCO: Common Objects in Context," 2015, arXiv:1405.0312.