문맥적응적 신경망 기반 화면내 예측의 트리 구조 반영 학습기법 분석

문기화, 허승정, 박도현, 김재곤
한국항공대학교
{ghmoon, tmdwjd07, dhpark}@kau.kr, jgkim@kau.ac.kr

Analysis of Training Method Using Tree Structure for Context Adaptive Neural Network-Based Intra Prediction

Gihwa Moon, Seung-Jeong Heo, Dohyeon Park, and Jae-Gon Kim
Korea Aerospace University

요 약

최근, 딥러닝 및 인공신경망 기술의 발전으로 비디오 부호화 분야에서도 인공지능을 이용한 요소 기술에 대한 연구가 활발이 진행되고 있다. 본 논문에서는 주변 참조샘플로부터 문맥정보를 이용하여 현재블록을 예측하는 CNN 기반의 화면내 예측 모델을 구현하고, 비디오 부호화의 블록 분할 구조를 반영한 학습 기법에 따른 부호화 성능을 분석한다. 실험결과 HM(HEVC Test Model)에 구현한 문맥적응적 신경망 기반 예측 모델에서 트리 분할 구조를 반영한 학습이 HM16.19 대비 0.35% BD-rate 부호화 성능 향상을 보였다.

1. 서론

최근 딥러닝 기술은 하드웨어의 발전과 함께 영상 분류, 영상 인식, 미디어 압축, 자연어 처리 등 다양한 분야에 적용되고 있으며 뛰어난 성능을 보이고 있다. 비디오 부호화 분야에서도 인공신경망 기술을 적용한 부호화 기술 개발이 활발히 진행되고 있다. 특히, 영상의 공간적 특성 및 참조샘플의 부족으로 인해 일반적인 화면내 예측은 성능 향상의 한계에 다다랐으며, 이를 극복할 수 있는 딥러닝 기반의 화면내 예측 기술의 중요성이 부각되고 있다[1]. 또한, VVC(Versatile Video Coding) 표준 완료 이후 JVET(Joint Video Experts Team)에서는 VVC 를 확장할 수 있는 신경망 기반의 비디오 부호화 기술의 잠재성을 확인하기 위한 NNVC(Neural Network based Video Coding) AhG(Ad-hoc Group)을 두고 관련 기술을 연구, 발전시키고 있다[2], [3].

본 논문은 JVET NNVC 에서 논의하고 있는 화면내 예측 모델로 예측 블록에 인접한 주변 복원 블록들을 입력하여 CNN(Convolutional Neural Network) 기반의 화면내 예측을 수행하는 문맥적응적(context-adaptive) 신경망 기반 화면내 예측 네트워크를 이용하여, 비디오 부호화의 블록 분할 구조를 반영하는 학습 기법이 비디오 부호화 성능에 미치는 영향을 분석한다[4].

2. 문맥적응적 신경망 기반 화면내 예측 모델 학습

(1) 문맥적응적 화면내 예측 네트워크

문맥적응적 모델은 현재 블록의 크기에 따라 학습되는 모델의 구조를 다르게 한다. 현재 블록의 크기가 $min(h,w) > 8$을 만족하는 경우 그림 1 의 n_a, n_l이 각각 4, 4 가 되고, 그림 1 과 같이 두개의 입력 참조 샘플($\boldsymbol{X}_0$, $\boldsymbol{X}_1$)이 입력되어서 예측블록을 생성한다. ${f_{h,w}}^0(.;{\boldsymbol{\theta}^0}_{h,w})$, ${f_{h,w}}^1(.;{\boldsymbol{\theta}^1}_{h,w})$는 CL (Convolutional Layer) 기반으로 구성되며, ${f_{h,w}}^f(.;{\boldsymbol{\theta}^f}_{h,w})$ 는 FCL(Fully-

Connected Layer) 기반으로 구성된다. 예측블록의 크기가 $\min(h,w) \leq 8$ 경우는 상단 부분과 왼쪽 부분의 참조샘플을 하나의 참조샘플로 만든 후, 해당 참조샘플을 입력하는 FCL 기반 네트워크를 통해 예측블록을 생성한다. 이때, 각각의 모델에 사용되는 입력 참조샘플의 라인 수는 4 로 고정한다.

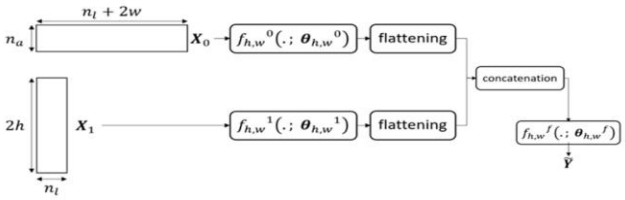

그림 1. 문맥적응적 신경망 기반 화면내 예측 네트워크[4]

(2) 트리 분할 구조 반영 학습

문맥적응적 신경망 기반 화면내 예측 모델을 비디오 부호화의 블록 분할을 반영하여 학습한다. 즉, 학습 과정에서 128x128 크기의 블록에 대해 다양한 블록 크기 및 이에 따른 예측 네트워크를 이용하여 예측블록을 생성하고, 재귀적 트리 구조로 예측 성능을 비교하여 최적의 블록 분할을 유도한다. 선택된 분할 깊이에 해당하는 데이터를 이용하여 블록 크기에 맞는 CA(Context-Adaptive) 신경망 화면내 예측 모델을 학습한다. COCO 데이터셋 및 Adam 최적화기를 이용하여 모델을 학습하였다.

4. 실험결과

학습한 네트워크를 HEVC(High Efficiency Video Coding)의 참조 소프트웨어인 HM16.19 의 추가적인 화면내 예측 모드로 구현하여 부호화 성능을 확인하였다[5]. 실험결과, 트리구조를 반영한 CA 신경망 기반 화면내 예측 모델이 HM16.19 대비 AI(All Intra) 부호화 환경에서 Y 와 Cr 각각 0.35%, 0.07% BD-rate 성능 향상을 보였다.

표 1. 제안 기법의 성능 (AI, over HM16.19)

Class	Y	U	V
Over HM16.19			
ClassA	-0.30%	0.16%	0.08%
Class B	-0.32%	0.00%	-0.32%
Class C	-0.26%	-0.24%	-0.54%
Class D	-0.46%	0.49%	0.88%
Class E	-0.45%	-0.10%	-0.58%
Overall	**-0.35%**	**0.07%**	**-0.07%**

5. 결론

본 논문에서는 문맥적응적 신경망 기반 화면내 예측 모델을 트리 분할 구조를 반영한 학습을 수행한 경우 그 부호화 성능을 분석하였다. 제안하는 모델 및 학습 기법을 HM16.19 위에 적용하였으며, 0.35%의 BD-rate 이득을 확인하였다.

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 원천기술 개발)

참 고 문 헌(References)

[1] “Use cases and requirements for Deep Neural Networks based Video Coding,” ISO/IEC JTC 1/SC 29/WG 2, N22, Oct. 2020.

[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] T. Dumas, A. Roumy and C. Guillemot, “Context Adaptive Neural Network Based Prediction for Image Compression,” IEEE Trans. on image proc., Vol. 29, 2020.

[5] High Efficiency Video Coding, Version 1, Rec. ITU-T H.265, ISO/IEC 23008-2, Jan. 2013.