

특집논문 (Special Paper)

방송공학회논문지 제27권 제4호, 2022년 7월 (JBE Vol.27, No.4, July 2022)

<https://doi.org/10.5909/JBE.2022.27.4.477>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

# VVC 화면 내 예측에서의 딥러닝 기반 예측 블록 개선을 통한 부호화 효율 향상 기법

정 혜 선<sup>a)</sup>, 강 제 원<sup>a)\*</sup>

## Accurate Prediction of VVC Intra-coded Block using Convolutional Neural Network

Hye-Sun Jeong<sup>a)</sup> and Je-Won Kang<sup>a)\*</sup>

### 요 약

본 논문에서는 컨볼루션 신경망 네트워크를 이용하여 VVC 화면 내 예측으로 얻은 예측 블록을 개선하여 잔차 신호를 보다 줄이는 화면 내 예측 방법을 제안한다. 기존의 화면 내 예측 방법은 일부 고정 규칙을 기반으로 주변의 재구성된 참조 샘플로부터 예측 블록을 생성하므로 복잡한 콘텐츠의 예측 블록을 생성하기 어렵다는 한계가 있다. 또한, 참조 샘플로 이용할 수 있는 정보의 양이 시간적 주변 정보에 비해 적기 때문에 화면 간 예측보다 낮은 부호화 성능을 가진다. 본 연구에서는 앞서 언급한 문제를 해결하기 위해 기존의 비디오 부호화 과정의 화면 내 예측을 통해 생성되는 예측 블록에 CNN을 적용하여 원본 블록과 예측 블록의 차분 신호를 줄이는 화면 내 예측 방법을 제안한다. 부호기에서는 제안 알고리즘의 활성 여부를 나타내는 플래그가 함께 부호화된다. 제안하는 화면 내 예측 방법은 최신 비디오 압축 표준인 Versatile Video Coding의 참조 모델인 VTM version 10.0<sup>[1]</sup> 대비 휘도 성분에 대하여 향상된 압축 성능을 제공한다.

### Abstract

In this paper, we propose a novel intra-prediction method using convolutional neural network (CNN) to improve a quality of a predicted block in VVC. The proposed algorithm goes through a two-step procedure. First, an input prediction block is generated using one of the VVC intra-prediction modes. Second, the prediction block is further refined through a CNN model, by inputting the prediction block itself and reconstructed reference samples in the boundary. The proposed algorithm outputs a refined block to reduce residual signals and enhance coding efficiency, which is enabled by a CU-level flag. Experimental results demonstrate that the proposed method achieves improved rate-distortion performance as compared a VVC reference software, I.e., VTM version 10.0.

Keyword : Intra Prediction, Convolutional Neural Network, Versatile Video Coding

## 1. 서론

최근 4K 및 8K 고해상도 비디오의 시청 수요가 늘어남에 따라 전체 네트워크 트래픽에서 비디오 데이터의 용량이 지속해서 증가하고 있다<sup>[2]</sup>. 비대면 환경에서의 업무 및 소셜 네트워킹 빈도가 많아지며 보다 높은 숫자의 화소 수, 비디오 프레임율, 명암비, 색 영역 등에서 더 높은 품질의 비디오 제작 수요가 늘어나고 있다. 이와 같이 최근의 고품질 비디오 제작환경에서 영상 데이터의 용량은 더 커지고 있고 시장에서 영상 데이터의 저장 및 전송을 위한 비디오 압축 기술이 점차 중요하게 여겨지고 있다.

비디오 압축에 관한 국제표준 그룹인 MPEG(Moving Picture Experts Group)에서는 이런 문제를 해결하기 위한 기술개발 및 표준화를 진행해 오고 있다<sup>[3]</sup>. 현재 가장 최신의 국제 비디오 압축 표준은 VVC(Versatile Video Coding)/H.266<sup>[4]</sup>이다. VVC 표준에서 비디오의 부호화 과정은 기존 압축 표준과 같이 영상을 적절한 블록으로 분할하고, 분할된 블록은 시간적/공간적으로 재구성된 주변 블록의 화소들을 참조 샘플로 사용하여 예측 블록을 생성하는 과정을 거친다. 그리고 생성된 예측 블록과 원본 영상의 차이인 차분 신호를 변환 및 양자화하여 전송한다. VVC는 기존 여러 비디오 압축 기술을 고도화함과 동시에 새로운 부호화 기술을 채택하여 HEVC(High Efficiency Video Coding)/H.265 대비 약 35%의 높은 부호화 효율을 보인다<sup>[5]</sup>.

비디오 압축에서 예측 기법은 부호화 효율에 큰 영향을 준다. 한 장의 프레임 안에서 예측 부호화를 수행하는 방식

을 화면 내 예측이라고 하며, 현재 블록 주변의 공간적으로 재구성된 샘플들을 활용하여 예측을 수행한다. VVC에서는 DC 및 Planar의 무방향성 예측 모드와 65개의 방향성 예측 모드 및 WAP(Wide-Angle Intra Prediction)을 이용한다. 이외에도, MRLP(Multiple Reference Line Intra Prediction), CCLM(Cross-Component Linear Model), PDPC(Position Dependent Intra Prediction Combination), ISP(Intra Sub-partitions), MIP(Matrix-based Intra Prediction)의 화면 내 예측 기술<sup>[6]</sup>이 적용되면서 이전 HEVC에 비해 더 세밀한 예측이 가능해졌다. 하지만 여전히 화면 내 예측 방법은 일부 고정된 규칙을 활용하여 예측 블록을 생성하므로 복잡한 콘텐츠의 예측 블록을 생성하기 어렵고, 참조 샘플로 이용할 수 있는 정보의 양이 시간적 주변 정보에 비해 적기 때문에 낮은 부호화 성능을 가진다.

따라서 본 논문에서는 컨볼루션 신경망 네트워크(convolutional neural network, CNN)를 이용하여 예측 블록을 개선하는 화면 내 예측 방법을 제안한다. 제안하는 알고리즘은 CNN을 이용하여 기존 VVC의 화면 내 예측 방법으로 생성된 예측 블록과 주변의 재구성된 참조 샘플을 입력으로 사용하여 현재 예측 블록을 개선함으로써 차분 신호를 줄여 비디오 부호화 효율을 증가시키는 것을 목표로 한다. 기존의 신경망을 이용하여 화면 내 예측을 수행하는 기술<sup>[7-11]</sup>의 경우는 주변 블록 샘플을 이용하여 예측 블록을 생성하였지만 주변 샘플과 부호화 블록 내부의 샘플 간의 거리 때문에 정확한 샘플을 생성하지 못하는 문제가 있었다. 그러나 제안 알고리즘에서는 첫 번째 단계에서 기존의 VVC 예측으로 예측 샘플을 생성하고 두 번째 단계에서 예측 샘플을 원본에 근사하도록 개선하는 방식을 제안한다. 본 논문은 본 연구진의 선행 연구<sup>[24,25]</sup> 대비 화면 내 예측 샘플을 효과적으로 학습하기 위한 기술을 제안한다. 제안 기술에 의하면 VVC 대비하여 우수한 부호화 효율을 제공한다.

본 논문은 다음과 같이 구성한다. 2장에서는 기존의 신경망 기반의 화면 내 예측 방법들에 대한 선행연구를 소개한다. 3장에서는 제안하는 CNN 기반 화면 내 예측 방법 알고리즘과 모델 학습 방식을 설명한다. 그리고 4장에서는 모델 학습에 관해 설명하고, 5장에서는 실험 결과를 분석한다. 마지막으로 6장에서는 결론을 맺는다.

a) 이화여자대학교 전자전기공학전공 스마트팩토리 프로그램(Department of Electronic and Electrical Engineering and Graduate Program in Smart Factory, Ewha Womans University)

‡ Corresponding Author : 강재원(Je-Won Kang)

E-mail: jewonk@ewha.ac.kr

Tel: +82-2-3277-2347

ORCID: <https://orcid.org/0000-0002-1637-9479>

※ This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. 2020-0-00920, Development of Ultra High Resolution Unstructured Plenoptic Video Storage/Compression/Streaming Technology for Medium to Large Space, 50%), and the ITRC(Information Technology Research Center) support program(IITP-2022-2020-0-01460, 50%) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

• Manuscript April 18, 2022; Revised June 27, 2022; Accepted June 27, 2022.

## II. 기존 연구

최근 딥러닝은 컴퓨터 비전, 자연어 처리 등의 여러 가지 분야에서 큰 성능 향상을 가져왔다. 이에 따라 비디오 부호화에 인공지능망을 비롯한 머신 러닝 기술을 적용하려는 시도들이 다수 있었다<sup>[17-22]</sup>. 화면 내 예측 방법에도 인공지능망 기술을 적용하는 기존의 연구가 있었다. [7]은 기존의 화면 내 예측이 복잡한 콘텐츠를 처리하기에 적합하지 않음을 해결하기 위해 완전 연결 계층(fully connected layer)로 이루어진 신경망 네트워크 기반의 화면 내 예측을 제안하였다. 제안하는 신경망 네트워크는 주변의 재구성된 참조 샘플들을 입력으로 하여 예측 블록을 생성한다. [7]에서 제안하는 방법은 HEVC에서 기존에 활용하던 화면 내 예측 모드 중 하나로 추가되어 동작하게 설계되었다. 이와 유사하게 주변의 재구성된 참조 샘플을 신경망 네트워크를 이용하여 예측 블록을 생성하는 화면 내 예측 방법을 VVC에 적용한 연구로는 [8]이 있다. [8]은 블록의 위쪽과 왼쪽에 있는 인접 참조 샘플로부터 현재 블록과 LFNST 변환 방식을 예측한다. 블록 사이즈에 따라 각각의 신경망 네트워크가 존재한다. 기본적으로 완전 연결 계층을 이용한 네트워크 구조를 가지고 있고, 블록의 사이즈가 큰 경우에는 컨볼루션 계층을 추가한 네트워크를 사용하였다. 이는 VVC의 화면 내 예측 모드 중 하나로 추가되어 동작한다. [9]는 완전 연결 계층에서 다차원 정보를 손실하는 문제를 해결하기 위하여 주변 참조 샘플로부터 정교한 예측 블록을 생성하기 위해 RNN(recurrent neural network) 기반의 네트워크 구조를 활용하였다. 이후에는 [10]과 [11]로 더 정교한 예측 블록을 생성하기 위해 모델의 구조를 확장하였다. 최종적으로 [11]은 픽셀 간의 상관 관계를 탐색함으로써 점진적 방식으로 예측 신호를 생성하는 progressive spatial recurrent neural network (PS-RNN)이라는 RNN 기반의 모델을 제안하였다.

CNN 기반의 구조를 활용하는 화면 내 예측 방식으로는 [12]가 있다. [12]는 HEVC의 예측 블록과 그 주변의 참조 샘플을 모두 네트워크의 입력으로 사용하여 예측 블록의 정확도를 올리는 MSCNN(Multi-scale Convolutional Neural Network)를 제안하였다. 서로 다른 크기의 특징 맵을 활용하기 위한 Multi-scale feature extraction network와

최종적으로 개선된 예측 블록을 생성하는 서브 넷으로 이루어져 있다. 이 밖에도 Inpainting 네트워크를 이용하여 예측 블록을 생성하고자 하는 연구<sup>[13]</sup>, 기존의 모든 화면 내 예측 방법을 Conditional Autoencoder를 통해서 양자화된 latent vector를 보내도록 대체하는 연구<sup>[14]</sup>도 있다.

신경망 혹은 딥러닝 네트워크를 이용하여 색차 성분의 예측 블록을 생성하는 연구도 있다<sup>[15-16]</sup>. 이미 재구성된 휘도 성분의 공간적 정보를 활용할 수 있는 네트워크를 사용하여 색차 성분의 예측 성능을 개선한다. 이는 VVC의 색차 성분 화면 내 예측 모드 중 하나로 추가된다.

앞서 비디오 부호화의 화면 내 예측을 개선하기 위한 연구들을 소개하였다. 대부분의 연구들은 인공 신경망 혹은 딥러닝 네트워크의 입력으로 현재 코딩하려는 블록 주변의 재구성된 참조 샘플만을 이용하여 예측 블록을 생성한다. 하지만 이는 기존의 화면 내 예측 기술을 통해 원본과 비슷하게 생성된 예측된 블록을 활용하지 않는다는 한계가 있다. 예를 들면, 기존의 화면 내 예측의 방향성 모드를 통해 생성된 예측 블록을 방향성을 기반으로 예측되었기 때문에 현재 블록의 대략적인 구조적 정보를 나타낼 수 있음에도 불구하고 이를 사용하지 않는다.

따라서 본 논문에서는 현재 블록의 주변 참조 샘플과 함께 VVC의 기존 화면 내 예측을 통해 생성된 예측 블록의 정보를 함께 활용하여 예측 블록 개선함으로써 원본 블록과의 차분 신호를 줄여 화면 내 예측의 부호화 효율을 증가시키는 것을 목표로 한다. 그리고 다양한 훈련 기법에 따른 부호화 성능을 보인다.

## III. 제안하는 알고리즘

본 논문은 화면 내 예측을 통해 생성된 예측 블록을 CNN을 이용하여 개선하는 화면 내 예측 방법을 제안한다. 제안 알고리즘에서 CNN은 기존의 화면 내 예측을 통해 생성된 예측 블록을 원본 블록에 근사하도록 개선한다. 이렇게 예측 블록과 원본 블록 사이의 차분 신호를 줄임으로써 화면 내 예측의 코딩 효율을 증가시킬 수 있다.

그림 1은 그러한 과정을 보인다. 그림 1에서 맨 왼쪽의 블록은 기존의 VVC 모드로 예측된 예측 블록이다. CNN을

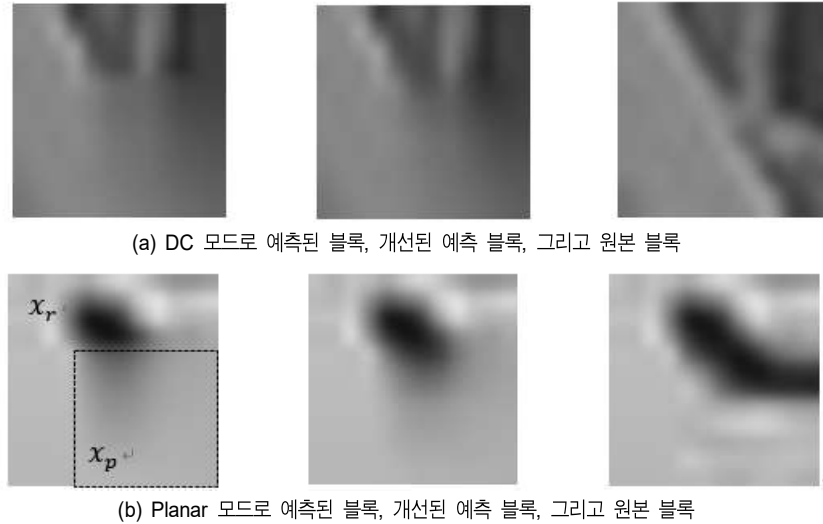


그림 1. CNN을 이용한 예측 블록 개선 결과 예시

Fig. 1. The refined prediction block results of proposed convolutional neural network

통해 개선된 예측 블록 (가운데 블록)은 원본인 오른쪽 블록을 예측하는 데보다 효과적이다.

### 1. 예측블록개선 신경망

예측 블록을 개선하는 CNN 모델을  $f$ 라고 할 때, 이는 식 (1)과 같이 표현될 수 있다.

$$x_p' = f(x_p, x_r) \quad (1)$$

CNN 모델  $f$ 는 VVC 화면 내 예측으로 생성된 예측 블록  $x_p$ 과 주변의 재구성된 참조 샘플  $x_r$ 이 입력으로 들어가 최종적으로 원본 블록에 가능한 한 근사하도록 개선된 예측 블록  $x_p'$ 을 출력하는 함수이다.  $x_p$ 과  $x_r$ 은 그림 1에 표현되어 있다.

그림 2는 예측 블록을 개선하기 위해 사용되는 네트워크의 구조로 함수  $f$ 를 묘사한다. 제안 알고리즘에서  $f$ 는 3개의 컨볼루션 계층 (convolution layer)과 2개의 전치 컨볼루션 계층 (transposed convolution layer)으로 이루어져 있다. 맨 마지막 계층을 제외한 각각의 컨볼루션 계층과 전치 컨볼루션 계층 뒤에는 활성화 함수로 ReLU가 존재한다. 그림 2에서 보이는 바와 같이 VVC의 화면 내 예측을 통해 생성

된 예측을 통해 생성된  $h \times w$  사이즈의 예측 블록  $x_p$ 과 주변의 참조 샘플들을 포함한  $(h+r) \times (w+r)$  사이즈의 블록이 네트워크의 입력으로 사용된다. 이때  $r$ 이 의미하는 바는 현재 예측하려는 블록의 위쪽과 왼쪽의 재구성된 참조 샘플의 라인 수이다. 본 실험에서는  $h$ 와  $w$ 는 8로,  $r$ 은 4이다.

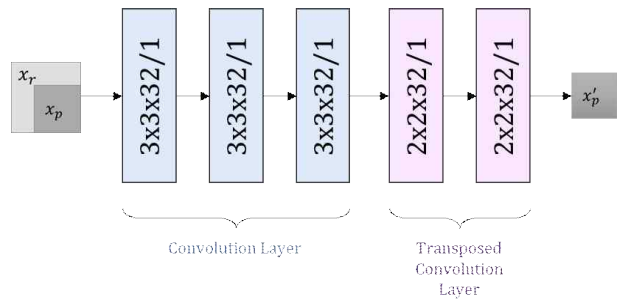


그림 2. 예측 블록 개선을 위한 CNN의 구조

Fig. 2. Convolutional Neural Network architecture for prediction block refinement

그림 2에서 3개의 컨볼루션 계층은 3x3 사이즈의 커널 사이즈를 가지고 있으며, 커널의 개수는 32이다. 이렇게 컨볼루션 계층으로부터 나온 특징 맵은 다시 2개의 전치 컨볼루션 계층을 거치면서 최종적으로  $h \times w$  사이즈의 개선된 예측 블록이 된다. 2개의 전치 컨볼루션 계층은 2x2 사이즈

의 커널 사이즈를 가진다. 첫번째 전치 컨볼루션 계층은 32개의 커널을 가지며, 개선된 예측 블록을 출력하는 마지막 전치 컨볼루션은 계층은 1개의 커널을 가진다. 이런 방식으로 기존 VVC 화면 내 예측을 통해 생성된 예측 블록과 주변의 재구성된 참조 샘플을 이용하여 예측 블록을 개선한다.

위의 컨볼루션 신경망 네트워크는 VVC의 방향성 모드에 적응적으로 학습하여 예측 블록 개선을 수행할 수 있다. 모드에 적응적이게 학습된 네트워크 모델은 다음과 같이 표현할 수 있다.

$$x'_p = f_m(x_p, x_r) \quad (2)$$

식 2에서  $m$ 은 VVC의 화면 내 예측 모드를 의미한다.  $x_p$ 의 해당 모드  $m$ 에 따라 서로 다른 모델  $f_m$ 이 적용된다. 제안 연구에서는 VVC에서 제공하는 총 80개의 화면 내 예측 모드를 7가지로 분류하여 모델을 구성했다. 보다 구체적인 내용은 IV-2에 기술한다.

기존의 비디오 부호화 과정에서 부호기는 비트 수 대비 현재 예측하려는 블록의 원본과 가장 유사한 예측 블록을 생성하는 화면 내 예측 모드를 선택한다. 그중 가장 많이 사용되는 방향성 모드는 방향성을 기반으로 주변 참조 샘플로부터 예측 블록을 생성한다. 따라서 같은 방향성 모드를 가지는 블록들은 서로 비슷한 구조적 정보를 가지고 있을 확률이 높다. 따라서 기존 VVC의 화면 내 예측 방법 중 방향성 모드로 예측된 블록들을 방향성에 따라 분류하여 컨볼루션 신경망 네트워크를 학습하고, 식 2와 같이 모드에 적응적인 네트워크를 사용함으로써 예측 블록을 개선하는 것이다.

## 2. VVC 참조 모델에서의 구현

그림 3은 VVC 부호기 안에서 제안하는 알고리즘의 동작 과정을 나타낸다. 본래 비디오 부호화 과정에서 화면 내 예측 모드의 선택은 다음과 같은 방식을 거친다.

우선 모든 화면 내 예측 모드에 대해 Hadamard 변환 비용을 계산하여 최종적으로 선택될 확률이 높은 모드 그룹으로 후보 목록을 구성한다. 다음으로 후보 목록에 포함된 모드에 대해 비트율-왜곡 최적화 (rate-distortion optimization, RDO)를 통해서 최적의 화면 내 예측 모드를 선

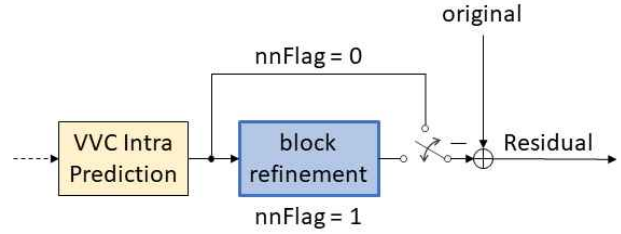


그림 3. VVC 부호기에서 제안하는 신경망 기반 화면 내 예측 방법의 동작 과정<sup>[24,25]</sup>

Fig. 3. The diagram of proposed neural network based intra prediction in VVC encoder<sup>[24,25]</sup>

택한다. 제안하는 CNN 기반 예측 블록 개선 알고리즘은 복잡도 감소를 위하여 모든 화면 내 압축 모드에 적용되는 것이 아니라 앞서 설명하는 과정 중 두 번째의 RDO 후보 목록에 있는 모드에만 적용된다.

제안하는 알고리즘은 VVC 참조 모델에 통합되어 RDO를 통해 활성화 여부가 결정된다. 즉, 부호기 측에서는 현재 화면 내 예측된 블록에 예측블록개선 신경망을 적용한 경우의 비트율-왜곡 비용이 적용하지 않았을 경우보다 작으면 제안 알고리즘을 활성화하는 플래그가 1로 설정이 되고 반대의 경우에는 비활성화된다. 제안하는 알고리즘의 활성화 여부를 결정하는 플래그는 CU(Coding Unit) 레벨에서 동작하며, 비트스트림에 함께 부호화 된다. 복호기에서는 현재 CU에 대해 해당 플래그 상태에 따라 예측 블록 개선 네트워크의 활성화 여부를 알 수 있으며, 플래그가 활성화되어 있으면 그 CU에 신경망 네트워크를 적용하여 예측 블록을 개선하는 일을 수행한다. 제안하는 알고리즘은 현재 블록의 휘도 성분에만 적용되었다.

## IV. 화면 내 예측 블록 개선 신경망 학습 기법

### 1. 훈련 데이터 세트 및 학습 구성

네트워크의 학습에 사용된 학습 데이터셋은 DIV2K<sup>[23]</sup>을 활용하였다. 학습 데이터셋은 다음과 같은 과정을 통해 만들어진 것이다. 일단, DIV2K 데이터셋의 이미지를 YUV 4:2:0 형식으로 변환하고 이를 VTM10.0을 이용하여 양자화 파

라메터 (QP) 값 22, 27, 32, 37의 서로 다른 양자화 세기마다 부호화하여 비트스트림 파일을 얻는다. 이렇게 만들어진 비트스트림 파일을 복호기로 복원 시에 주변 참조 샘플과 예측 블록으로 이루어진 블록을 출력하여 학습 데이터셋으로 이용하였다. 학습에 사용되는 DIV2K 데이터셋은 앞으로 성능 평가에 사용하게 될 테스트 비디오 세트에 포함되지 않는다.

위에서 구성된 모든 학습 블록들은 8x8 사이즈이고, 너무 단순하거나 복잡한 텍스처를 가진 블록을 학습 과정에서 제외하기 위해 다음과 같은 필터링 과정을 거친다<sup>[23]</sup>.

$$0.1 < \frac{MSE}{\overline{MSE}} < 2 \quad (3)$$

식 3에서  $MSE$ 는 원래 블록과 VVC 화면 내 예측을 통해 생성된 블록 사이의 평균 제곱 오차(Mean Square Error, MSE)를 나타내며,  $\overline{MSE}$ 는 이미지 안의 블록들의 평균 MSE를 의미한다. 본 논문에서는 식 3의 범주에 속하는 블록들만 학습에 사용된다. 너무 복잡한 블록의 경우 네트워크가 개선하기 어렵고, 평활한 블록의 경우 기존의 VVC의 화면 내 예측을 통해 쉽게 예측할 수 있기 때문이다. 이렇듯 데이터셋을 필터링 함으로써 네트워크 훈련에 일관성을 줄 수 있다.

네트워크의 Optimizer는 Adam optimizer를 사용하였다. Learning rate는 1e-4로 설정하였고, 훈련이 진행됨에 따라 지수 함수로 줄어든다. 네트워크는 각각의 QP 값을 (22, 27, 32, 37)에 따라 학습되었고, 휘도 성분에 대해서만 학습을 진행하였다.

## 2. 방향성 모드에 따른 네트워크 학습

VVC의 방향성 모드는 크게 방향성에 따라 Planar (모드 0), DC (모드 1), 수평 방향 모드 (모드 14~22), 수직 방향 모드 (모드 46~54), 우상향 대각 모드 (모드 55이상), 좌상향 대각 (모드 23~45), 좌하향 (모드 22이하) 총 7가지로 분류할 수 있다.

본 논문에서는 VVC의 방향성 모드에 적응적으로 네트워크를 학습하여 예측 블록의 개선을 수행할 수 있다. 이때, 위에서 분류한 7개의 방향성 모드에 따라 각 모드를 그룹으로 묶어 학습을 진행한다. 이때, 블록의 방향성 특징을 유지하기 위해서 전처리 과정을 수행한다. 그림 4는 수직 방향 모드로 예측된 블록의 전처리 과정 예시를 보여준다. 수직 방향 모드로 예측된 블록을 상하 반전하고, 시계 방향으로 90도 회전함으로써 수평 방향 모드로 예측된 블록과 동일한 형태를 가지도록 만드는 과정이다. 따라서 본 논문에서는 모드 적응적인 네트워크 학습 및 적용 시에 그림 4와 같은 전처리 과정을 거쳐 네트워크의 입력으로 사용한다. 그리고 출력으로 나온 블록은 다시 원상태로 복원하는 과정을 거친다. 다른 그룹들에 대해서도 같은 방식을 적용한다. 이렇게 학습을 진행하면 입력 예측 블록의 방향성 특성을 유지하면서 학습 모델의 개수를 7개에서 5개로 감소시킬 수 있다. 또한, 전체 데이터셋을 방향성 예측 모드에 따라 나누면서 학습 샘플의 개수가 줄어들게 되는데 그림 4와 같은 전처리를 통해 두 개의 모드를 합침으로써 CNN 훈련 시의 학습 샘플 개수를 늘릴 수 있다.

## V. 실험 결과 및 분석

본 장에서는 제안하는 알고리즘에 대한 성능을 보인다. 제안하는 알고리즘은 VVC 참조 소프트웨어인 VTM-10.0

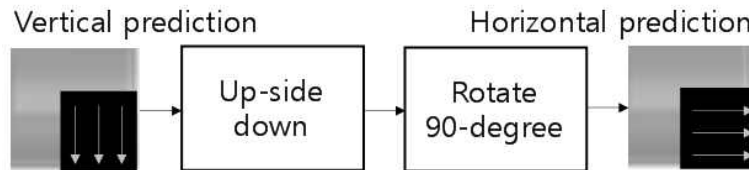


그림 4. 수직 방향 예측 블록의 전처리를 통한 수평 방향 예측 블록으로의 변환 예시  
Fig. 4. Example of conversion to a horizontal prediction block through preprocessing of a vertical prediction block

에 구현하였다. 실험에서 AI(All Intra) 부호화 환경에서 ISP(intra sub-partitions)와 MIP(matrix-weighted intra prediction) 기능을 비활성화한 상태에서 QP 22, 27, 32, 37에 대하여 BD-rate 결과를 보인다. 참조 소프트웨어에서 역시 해당 기술을 비활성화하고 anchor로 사용하였다. 또한 제안 기술은 8x8 블록으로 제한하여 구현하였다. 테스트 시퀀스는 MPEG에서 제공하는 JCT-VC 시퀀스 Class B, C, D, E를 사용하였다. 본 논문에서는 휘도 성분에 대해서만 제안하는 알고리즘을 적용하였으므로 휘도 성분에 대한 BD-rate 결과를 분석한다.

표 1은 제안하는 방향성 적응적인 CNN 기반 예측 블록 개선 화면 내 예측 방법을 anchor와 비교하였을 때의 BD-rate 결과이다. 화면 내 예측 모드의 방향성을 고려하여 학습한 방식을 모드 적응형으로 명명한다. 반대로 방향을 구분하지 않고 모드에 상관이 없이 CNN을 활용하여 학습시킨 결과를 모드 일반형으로 명명하여 부호화 효율을 비교하였다. 표 1에서 보듯이 모드 적응형으로 학습한 방식은 휘도 성분의 부호화 효율로 평균 약 -1.18%의 BD-rate

감소를 보인다. 방향성에 상관이 없이 제안 모델을 학습하였을 경우에는 평균 약 -1.13% 이었다. 모드 일반형의 경우 최대 부호화 효율은 “BasketBallDrill” 시퀀스에서 -2.48%의 효율을 보였지만 일부 시퀀스에서 손실이 발생한 경우도 관찰되었다. 반면에 모드 적응형의 경우 최대 부호화 효율은 RaceHorses에서 -2.00% 였지만 모든 시퀀스에서 손실이 발생한 경우는 없었다. 표1의 실험에서는 8x8 블록으로 제한하여 구현하였지만, 16x16 및 32x32 블록에 대해 적용하고 같은 구성에서 실험하였을 때에는 -0.4%의 부호화 효율을 보였다. 또한 MIP와 ISP를 활성화 하여 같은 조건에서 비교하였을 때는 -0.3%의 부호화 효율을 보였다.

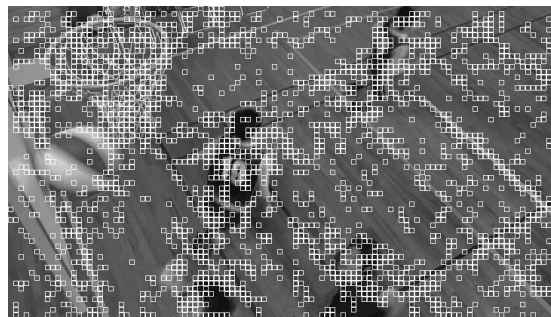
그림 5는 제안하는 신경망 기반의 화면 내 예측 기술이 적용된 CU의 시각적인 분포를 보여준다. 그림 5에 보이는 바와 같이 각 프레임에 대하여 복잡한 콘텐츠를 가진 영역 혹은 경계선 부분에서 제안하는 알고리즘이 적용된 블록이 많음을 볼 수 있다. 이를 통해서 컨볼루션 신경망 네트워크를 이용하여 예측 블록을 개선하는 알고리즘이 보다 복잡

표 1. 제안하는 알고리즘의 부호화 효율 비교 실험  
Table 1. Rate-distortion performance of the proposed algorithm

Classes	Sequence	BD-Rate (Y)	
		Mode-dependent BD-Rate (Y)	Mode-independent BD-Rate (Y)
Class B	Kimono	-0.41%	0.27%
	ParkScene	-1.08%	-1.27%
	Cactus	-1.01%	-0.90%
	BasketballDrive	0.51%	0.18%
	BQTerrace	-1.33%	-1.67%
	Average	-0.67%	-0.68%
Class C	BasektballDrill	-1.68%	-2.48%
	BQMall	-1.78%	-1.53%
	PartyScene	-1.71%	-1.73%
	RaceHorsesC	-1.05%	-1.30%
	Average	-1.55%	-1.76%
Class D	BasketballPass	-0.75%	-0.29%
	BQSquare	-1.87%	-2.03%
	BlowingBubbles	-1.86%	-1.36%
	RaceHorses	-2.00%	-2.25%
	Average	-1.62%	-1.48%
Class E	FourPeople	-0.68%	-0.85%
	Johnny	-1.15%	-0.14%
	KristenAndSara	-0.85%	-0.84%
	Average	-0.89%	-0.61%
Overall Average		-1.18%	-1.13%



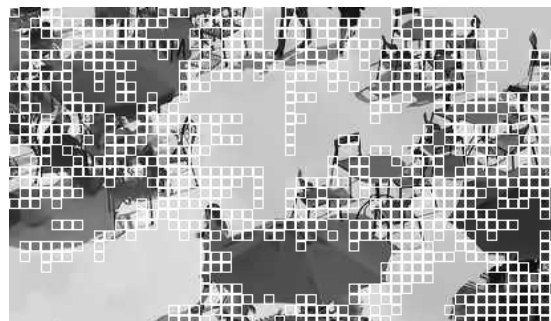
(a) Cactus coded with QP22



(b) BasketballDrill coded with QP27



(c) FourPeople coded with QP32



(d) BQSquare coded with QP37

그림 5. 각 시퀀스에 대해 제안하는 알고리즘이 적용된 CU의 분포

Fig. 5. The distribution of CUs predicted by proposed algorithm

한 콘텐츠에서 효과적으로 동작하고 있음을 시각적으로 확인할 수 있다. 전체 시퀀스에 대해서 블록의 평균 선택 비율은 15.5%였다. BasketballDrive에서는 모드 선택 비율이 8.8%로 전체 시퀀스에서 가장 낮은 비율을 보였고 표1에서 보이는 바와 같이 부가 정보로 인해 부호화 손실이 발견되었다. RaceHorses에서는 27.7%로 가장 높은 모드 선택 비율을 보였으며 이 때 2%의 부호화 효율을 보였다. 제안 알고리즘의 부호기 및 복호기에서의 시간 복잡도는 CUDA를 이용한 GPU 환경에서 연산하였을 때 부호기 복잡도는 250% 증가, 복호기 복잡도는 364% 증가하였다. 부호기 측에서는 늘어난 모드의 숫자로, 복호기 측에서는 추가적인 신경망 연산으로 복잡도가 증가하였다.

## VI. 결 론

본 논문은 컨볼루션 신경망 네트워크를 이용하여 예측 블록을 개선하는 화면 내 예측 기법을 제안한다. 컨볼루션

신경망 네트워크는 기존의 VVC의 화면 내 예측 모드를 통해 생성된 예측 블록과 주변 참조 샘플을 이용하여 예측 블록을 개선함으로써 원본 블록과의 차분 신호를 줄여 비디오 부호화 성능을 향상시킬 수 있다. 제안하는 알고리즘은 VVC 휘도 성분의 화면 내 예측 모드 중 하나로 추가되어 동작한다. 실험의 결과, VTM-10.0 대비 휘도 성분에 대해 약 1.18%의 BD-rate를 개선하였다. 이는 예측 블록을 인공 신경망을 이용하여 개선함으로써 화면 내 예측 성능을 향상시킬 수 있다는 점에서 의의를 가진다.

## 참 고 문 헌 (References)

- [1] Versatile Video Coding Test Model (VTM) 10.0 : [https://vegit.hhi.fraunhofer.de/jvet/VVCSsoftware\\_VTM-10.0/tree/VTM-10.0](https://vegit.hhi.fraunhofer.de/jvet/VVCSsoftware_VTM-10.0/tree/VTM-10.0)
- [2] T. Barnett, et al. "Cisco visual networking index (vni) complete forecast update, 2017 - 2022." Americas/EMEAR Cisco Knowledge Network (CKN) Presentation (2018).
- [3] Moving Picture Experts Group (MPEG) : <https://mpeg.chiariglione>.

- org
- [4] JVET of ITU-T and ISO/IEC, "Versatile Video Coding (Draft 10)", Documents JEVTS2001, June 2020.
- [5] B. Bross et al., "Overview of the Versatile Video Coding (VVC) Standard and its Applications" IEEE Transactions on Circuits and Systems on Video Technologies, vol. 31, no. 10, 2021.  
doi: <https://doi.org/10.1109/TCSVT.2021.3101953>
- [6] J. Pfaff et al., "Intra Prediction and Mode Coding in VVC", IEEE Transactions on Circuits and Systems on Video Technologies, vol. 31, no. 10, 2021.  
doi: <https://doi.org/10.1109/TCSVT.2021.3072430>
- [7] J. Li, et al. "Intra prediction using fully connected network for video coding." 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017.  
doi: <https://doi.org/10.1109/ICIP.2017.8296231>
- [8] JVET of ITU-T and ISO/IEC, "AHG11 Neural Network-based Intra Prediction with Transform selection in VVC", Document JVET-T0073, Oct. 2020.
- [9] Y. Hu, et al. "Enhanced intra prediction with recurrent neural network in video coding." 2018 Data Compression Conference. IEEE, 2018.  
doi: <https://doi.org/https://doi.org/10.1109/DCC.2018.00066>
- [10] Y. Hu, et al. "Optimized spatial recurrent network for intra prediction in video coding." 2018 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2018.  
doi: <https://doi.org/10.1109/VCIP.2018.8698658>
- [11] Y. Hu, et al. "Progressive spatial recurrent neural network for intra prediction." IEEE Transactions on Multimedia 21.12 (2019): 3024-3037.  
doi: <https://doi.org/10.1109/TMM.2019.2920603>
- [12] Y. Wang, et al. "Multi-scale convolutional neural network-based intra prediction for video coding." IEEE Transactions on Circuits and Systems for Video Technology 30.7 (2019): 1803-1815.  
doi: <https://doi.org/10.1109/TCSVT.2019.2934681>
- [13] L. Zhu, et al. "Generative adversarial network-based intra prediction for video coding." IEEE transactions on multimedia 22.1 (2019): 45-58.  
doi: <https://doi.org/10.1109/TMM.2019.2924591>
- [14] F. Brand, S. Jürgen, and K. André. "Intra frame prediction for video coding using a conditional autoencoder approach." 2019 Picture Coding Symposium (PCS). IEEE, 2019.  
doi: <https://doi.org/10.1109/PCS48520.2019.8954546>
- [15] M. G. Blanch, et al. "Chroma intra prediction with attention-based CNN architectures." 2020 IEEE International Conference on Image Processing (ICIP). IEEE, 2020.  
doi: <http://dx.doi.org/10.48550/arXiv.2006.15349>
- [16] JVET of ITU-T and ISO/IEC, "AHG11: Neural Network based cross-component Prediction model", Document JVET-W0111, July 2021
- [17] Sang-hyo Park and Je-Won Kang, "Fast Multi-type Tree Partitioning for Versatile Video Coding Using a Lightweight Neural Network", IEEE Transactions on Multimedia, 2021.  
doi: <https://doi.org/10.1109/TMM.2020.3042062>
- [18] Jyung-Kyung Lee, Nayoung Kim, Seunghyun Cho, and Je-Won Kang, "Deep Video Prediction Network Based Inter-Frame Coding in HEVC," IEEE Access, 2020.  
doi: <https://doi.org/10.1109/ACCESS.2020.2993566>
- [19] Sookyung Ryu and Je-Won Kang, "Machine Learning-Based Fast Angular Prediction Mode Decision Technique in Video Coding," IEEE Transaction on Image Processing, Nov. 2018.  
doi: <https://doi.org/10.1109/TIP.2018.2857404>
- [20] Je-Won Kang, Soo-Kyung Ryu, Na-Young Kim, Minjoo Kang, "Efficient Residual DPCM using an L-1 Robust Linear Prediction in Screen Content Video Coding," IEEE Transaction on Multimedia, vol. 18, no. 10, pp.2054-2065, Oct. 2016.  
doi: <https://doi.org/10.1109/TMM.2016.2595259>
- [21] Jyung-Kyung Lee, Nayoung Kim, and Je-Won Kang, "Rate-distortion optimized temporal segmentation using reinforcement learning for video coding," APSIPA, 2021.
- [22] Je-Won Kang, Gabbouj, M., and Jay Kuo, C. C. "Sparse/DCT (S/DCT) two-layered representation of prediction residuals for video coding", IEEE Transactions on Image Processing, 22(7), 2711-2722.  
doi: <https://doi.org/10.1109/TIP.2013.2256917>
- [23] Radu Timofte, Eirikur Agustsson, Luc Van Gool, MingHsuan Yang, and Lei Zhang, "Ntire 2017 challenge on single image super-resolution: Methods and results," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 114 - 125.  
doi: <https://doi.org/10.1109/cvprw.2017.150>
- [24] H. Jeong, Je-Won Kang "Improvements of intra-prediction in VVC", Summer conference in the Korean Society of Broadcast and Media Engineers, 2022
- [25] H. Jeong, Je-Won Kang, "Intra prediction through block refinement", 34<sup>th</sup> Image processing and understanding, 2022.

---

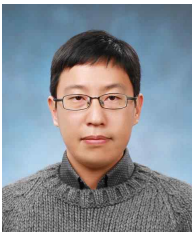
저 자 소 개

---



정 혜 선

- 2018년 8월 : 이화여자대학교 전자공학과 (학사)
- 2022년 2월 : 이화여자대학교 전자전기공학과 (석사)
- ORCID : <https://orcid.org/0000-0002-3148-1906>
- 주관심분야 : Video Coding, Image/Video Processing, Deep Learning



강 제 원

- 2006년 2월 : 서울대학교 전기컴퓨터공학부 (학사)
- 2008년 2월 : 서울대학교 전기컴퓨터공학부 (석사)
- 2012년 5월 : University of Southern California (박사)
- 2012년 8월 ~ 2014년 2월 : Qualcomm (Multimedia RnD and Standard team)
- 2014년 3월 ~ 2019년 3월 : 이화여자대학교 전자공학과 조교수
- 2019년 3월 ~ 현재 : 이화여자대학교 전자전기공학과 부교수
- ORCID : <http://orcid.org/0000-0002-1637-9479>
- 주관심분야 : Video coding, 3D Video, Image/Video Processing, Machine Learning