졸업논문 소개

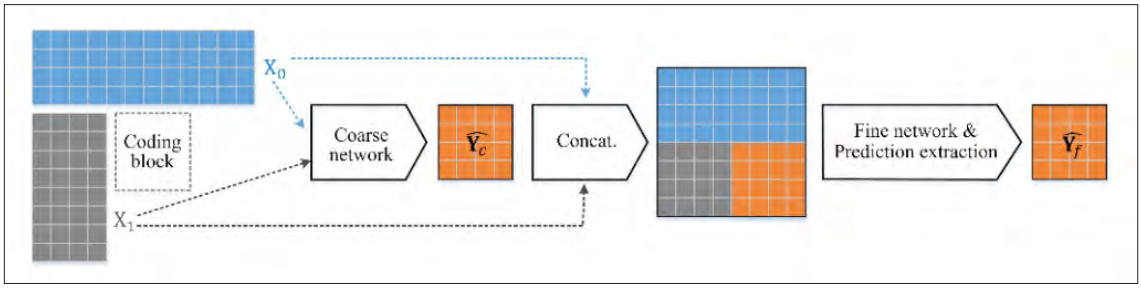
VVC를 위한 Coarse-to-Fine 신경망 기반의 화면 내 예측

박도현 / 한국항공대학교 Media Communication Lab

최신 비디오 부호화 표준인 VVC는 동일한 비디오 품질에서 HEVC와 비교하여 약 50% 비트율 감소를 제공하며, 8K 초고해상도, HDR(High Dynamic Range) 및 360도 동영상을 포함한 다양한 유형의 비디오를 효과적으로 수용할 수 있도록 설계되었다. 딥러닝 기술의 발전에 따라서 기계 학습 기반의 비디오 부호화 기술들은 VVC 개발 과정에서 매우 효율적인 부호화 성능을 보여주었으며 MIP(Matrix-based Intra Prediction) 및 LFNST(Low Fequencey Non Separable Transform) 등과 같은 기술로 VVC에 채택되었다. JVET은 VVC의 표준화 완료 이후 신경망 기반 비디오 부호화(NNVC: Neural Network-based Video Coding)를 위한 탐색 실험(EE: Experimental Experiment)을 시작하여 이러한 부호화 기술들의 성능과 비디오 부호화 표준 기술로서의 적합성을 평가하고 탐색하고 있다.

한편 화면 내 예측은 공간적 종속성을 줄이기 위해 미리 정의된 단순한 방향성을 갖는 예측 모드에 의존하며 VVC 또한 HEVC에서 확장된 방향성을 갖는 화면 내 예측을 제공하지만 여전히 복잡한 공간적 특징이나 복수의 방향성을 포함하는 영상을 효과적으로 부호화하는데 한계가 있다. 또한 화면 내 예측 성능과 모드 부호화를 위한 비트 오버헤드 사이의 적절한 균형을 유지하기 위해 수행 가능한 화면 내 예측 모드의 수는 제한된다. 따라서, 이러한 화면 내 예측의 문제들을 해결하기 위해 다양하고 복잡한 특징을 가진 부호화 블록에 대해 효율적으로 화면 내 예측을 수행할 수 있는 신경망 기반의 화면 내 예측에 대한 연구가 활발히 이루어지고 있다.

본 논문에서는 〈그림 1〉과 같이 Coarse-to-Fine 네트워크 구조를 갖는 신경망 기반의 화면 내 예측을 제안한다. Coarse-to-Fine 네트워크와 같은 두 단계의 네트워크 구조는 특정 목적에 맞게 훈련될 수 있는 분리된 네트워크를 사용하기 때문에 다양하고 안정적인 예측을 가능하게 하고 이미 영상 인페인팅(inpainting) 분야에서 널리 사용되고 있다. 그러나 이러한 네트워크의 구조는 계산 복잡성으로 인해 재귀적 분할 최적화를 수행하는 비디오 코덱에 적합하지 않을 수 있다. 따라서, 기존의 신경망 기반 화면 내 예측에 대한 연구들은 모두 단일 단계 네트워크에 중점을 둔다. 본 논문에서는 복잡도를 증가시키지 않으면서 두 단계의 네트워크 구조의 이점을 활용하기 위해 Coarse-to-Fine 개념만을 포함하는 단순한 네트워크 구



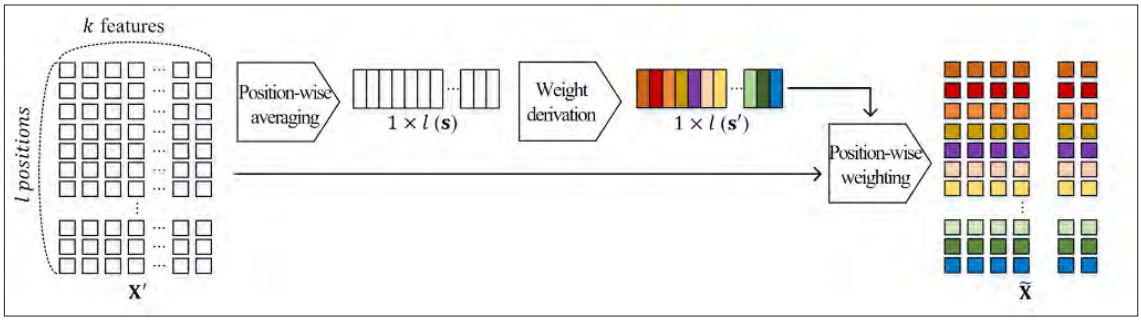
<그림 1> 제안 Coase-to-Fine 네트워크 구조

조를 제안한다.

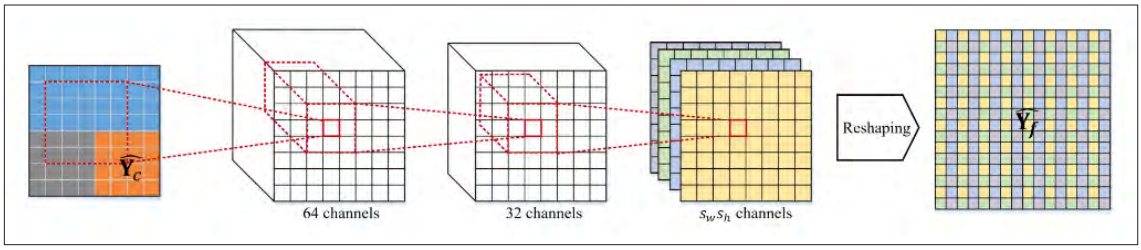
기존 연구들의 화면 내 예측 네트워크 구조는 참조 샘플과 예측 샘플을 직접 연결하기 때문에 네트워크가 참조 샘플의 위치 특성과 문맥에 따른 중요성을 반영하기 어려운 한계가 있다. 본 논문에서 제안하는 Coarse 네트워크는 이러한 문제를 해결하기 위해 참조 샘플의 특징(feature)을 위치와 공간적 문맥을 고려하여 특징의 스케일을 조절함으로써 중요한 특징을 잘 반영하는 화면 내 예측을 수행한다. 즉, 제안하는 Coarse 네트워크는 특징 재보정과 예측 두 가지 부분으로 구성된다. 〈그림 2〉와 같이 특징 재보정에서는 위치와 특성에 따라 참조 샘플의 중요성이 블록 단위로 나누어 결정되며 예측 부분에서는 위치별 가중치가 적용된 특성을 입력으로 사용하여 예측 샘플을 생성한다.

또한, 완전연결계층 기반의 화면 내 예측에서는 예측 샘플이 부호화 블록 내의 위치에 따라 개별적으로 생성되기 때문에 예측 샘플과 주변 블록 간의 연결성이 떨어질 수 있으며 이는 부호화 아티팩트(artifact)를 야기할 수 있다. 따라서, 본 논문에서는 〈그림 3〉과 같이 간단한 CNN을 기반으로 한 Fine 네트워크를 추가적으로 구성하여 Coarse 네트워크의 예측 정확도와 예측 블록과 인접 참조 샘플과의 연결성을 향상한다.

뿐만 아니라 블록 분할을 고려한 데이터셋 구축 방법과 변환 및 양자화를 기반으로 한 손실함수를 제안하여 비디오 코덱에서 네트워크의 사용성을 더욱 향상시켰으며 제안된 네트워크는 다양한 블록 크기에 대해 개별적으로 훈련되어 VVC와 같은 블록 기반 하이브리드 비디오 코덱을 효율적으로 지원할 수 있게 하였다. 뿐만 아니라 하나의



<그림 2> 제안 Coarse 네트워크의 특징 재보정



<그림 3> 제안 Fine 네트워크 구조

Coarse 네트워크와 블록 크기별로 특정한 Upscale 네트워크를 사용하여 다양한 블록 크기에 대한 네트워크를 저장하기 위한 메모리 오버헤드를 효율적으로 줄였다.

제안 기법의 부호화 성능을 평가하기 위해 제안하는 화면 내 예측 모드를 VTM(VVC Test Model) 11.0에 추가적인 화면 내 예측 모드로 통합하였다. 실험 결과, 제안 기법은 VTM 11.0과 비교하여 휘도 성분에 대해 평균 1.31%의 BD-rate(Bjøntegaard delta-rate) 절감을 보였으며, 이전 최신 연구와 비교하여 별다른 부호화 복잡도 증가 없이 평균 0.47%의 BD-rate 절감을 보였다.

박 도 현

- 2016년 2월 : 국립한밭대학교 멀티미디어공학과 학사
- 2018년 2월 : 한국항공대학교 전자정보공학과 석사
- 2023년 2월 : 한국항공대학교 전자정보공학과 박사
- 2023년 2월 ~ 현재 : 퀄컴코리아 시니어엔지니어
- 주관심분야 : 비디오 코덱, 딥러닝