

특집논문 (Special Paper)

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

<https://doi.org/10.5909/JBE.2022.27.4.527>

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

딥러닝 기반 컨텐츠 적응적 영상 압축 기술 동향

한 창 우^{a)}, 김 흥 일^{b)}, 강 현 구^{a)}, 권 형 진^{c)}, 임 성 창^{c)}, 정 승 원^{a)†}

Survey on Deep learning-based Content-adaptive Video Compression Techniques

Changwoo Han^{a)}, Hongil Kim^{b)}, Hyun-ku Kang^{a)}, Hyoungjin Kwon^{c)}, Sung-Chang Lim^{c)},
and Seung-Won Jung^{a)†}

요 약

멀티미디어 컨텐츠의 수요와 공급이 증가함에 따라 전 세계의 인터넷 트래픽이 증가하는 가운데 이를 완화하기 위해 여러 표준화 그룹에서는 더 효율적인 압축 표준을 제정하는데 노력을 기울이고 있다. 이러한 노력 중 압축 표준에 딥러닝 기술을 도입하고자 하는 연구들이 활발히 진행되고 있다. 그러나 딥러닝 기반 압축 기술은 학습 데이터와 특성이 다른 영상을 압축할 때 압축 효율이 저하되는 문제를 갖는다. 이를 해결하기 위해 컨텐츠에 적응적으로 딥러닝 기술을 도입하는 시도들이 있었다. 본 논문에서는 이들을 크게 코덱 정보 사용, 모델 선택, 추가 정보 전송의 세 가지로 나누어 살펴보고자 한다.

Abstract

As multimedia contents demand and supply increase, internet traffic around the world increases. Several standardization groups are striving to establish more efficient compression standards to mitigate the problem. In particular, research to introduce deep learning technology into compression standards is actively underway. Despite the fact that deep learning-based technologies show high performance, they suffer from the domain gap problem when test video sequences have different characteristics of training video sequences. To this end, several methods have been made to introduce content-adaptive deep video compression. In this paper, we will look into these methods by three aspects: codec information-aware methods, model selection methods, and information signaling methods.

Keyword : Contents adaptive filtering, Deep-learning, In-loop filtering, Post-processing, Video compression

a) 고려대학교 전기전자공학과 (Department of Electrical Engineering, Korea University)

b) 동국대학교 전기전자공학부 (School of Electronics and Electrical Engineering, Dongguk University)

c) 한국전자통신연구원 통신미디어연구소 (Telecommunications & Media Research Laboratory, Electronics and Telecommunications Research Institute)

† Corresponding Author : 정승원(Seung-Won Jung)

E-mail: swjung83@korea.ac.kr

Tel: +82-2-2231-4853

ORCID: <https://orcid.org/0000-0002-0319-4467>

※ 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2017-0-00072, 초실감 테라미디어를 위한 AV 부호화 및 LF 미디어 원천기술 개발).

※ This work was supported by Institute for Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00072, Development of Audio/Video Coding and Light Field Media Fundamental Technologies for Ultra Realistic Tera-media).

· Manuscript May 19, 2022; Revised July 4, 2022; Accepted July 4, 2022.

I. 서 론

스마트폰 및 PC 등 개인 통신기기의 보급과 멀티미디어 콘텐츠의 수요와 공급이 증가함에 따라 전 세계의 인터넷 트래픽에서 영상 신호의 트래픽이 차지하는 비율이 큰 폭으로 증가하였다. 전 세계의 영상 트래픽의 비율은 2015년 68%에서 2020년 82%까지 급증하였는데^[1], 이는 4K에 이르는 고해상도의 동영상 소비 확대와 COVID-19 바이러스의 세계적 유행으로 인한 동영상 콘텐츠 소비가 확대가 된 것이 원인이다.

지속적으로 증대되는 영상 트래픽의 부담을 줄이기 위해 여러 표준화 그룹에서 더욱 효율적인 영상 압축 표준을 제정하는데 심혈을 기울이고 있다. 대표적으로 Joint Picture Experts Group (JPEG)의 정지영상 압축 표준인 JPEG 2000^[2] 및 ITU-T VCEG, ISO/IEC MPEG에서 만든 협력연구 기구인 JVET의 동영상 압축 표준인 Versatile Video Coding (VVC)^[3] 등이 있다. 이러한 압축 표준들은 원본 신호와 예측 신호의 잔차 신호를 주파수 변환 뒤 양자화하여 영상 신호의 시간적, 공간적 중복성을 없앴고 신호를 엔트로피 코딩을 통해 통계적 중복성을 제거하였다. 이로써 원본 영상에 비해 훨씬 적은 비트로 높은 품질의 영상을 저장할 수 있게 되었다.

최근 딥러닝 기술의 발전에 따라 압축 표준에 딥러닝 기술을 적용하여 성능을 향상시키는 연구들이 진행되고 있다. 특히 기존 압축 표준의 구성 기술을 딥러닝을 통해 개선하고자 하는 모델들이 많이 제안되고 있다. 딥러닝 기반 기술들은 기존 압축 표준에 비해 일반적으로 높은 부호화 성능

을 보이나, 단일 모델로는 다양한 특성의 비디오를 효율적으로 압축하는데 한계가 존재한다. 이는 일반적으로 딥러닝 모델은 학습 데이터의 전반적인 특성을 학습하기 때문에 모든 데이터에 대해 한 가지 모델로 같은 연산을 수행하게 되면 입력 영상의 특성에 따라 화질을 향상시키기 힘들기 때문이다. 이러한 문제를 해결하기 위해 영상 후처리/루프 내 필터 분야에서 딥러닝 모델의 성능을 더욱 향상시켜 입력 영상에 적응적으로 필터링할 수 있는 모델에 대한 연구가 진행되고 있다. 본 논문에서는 딥러닝 기반 영상 전/후 처리를 통해 기존 압축 표준의 압축률을 높이는 연구 중 영상 적응적 필터링 기술들의 최신 동향에 대해 살펴보고자 한다. 2장에서는 딥러닝 기반 필터가 코덱의 정보를 활용하는 기술들을 알아보고, 3장에서는 입력 영상에 적응적인 모델 선택 방식 기술들을 알아본다. 4장에서는 추가 정보 전송 기술들을 살펴본다. 마지막으로 5장에서 결론을 맺는다.

II. 코덱 정보 활용

비디오 압축 표준에서 영상 압축은 블록 단위로 진행되므로 영상 복원 시에 블록 경계 부근에 열화가 발생한다. 또한 양자화를 거치면서 영상의 고주파 정보 손실에 의한 ringing 현상이 발생한다. 이뿐만 아니라, 일반적으로 영상을 압축할 때 quantization parameter (QP)가 클수록 비디오 프레임의 왜곡 현상이 심해진다. 이를 해결하기 위해 de-blocking filter (DBF), sample adaptive offset (SAO) 등의

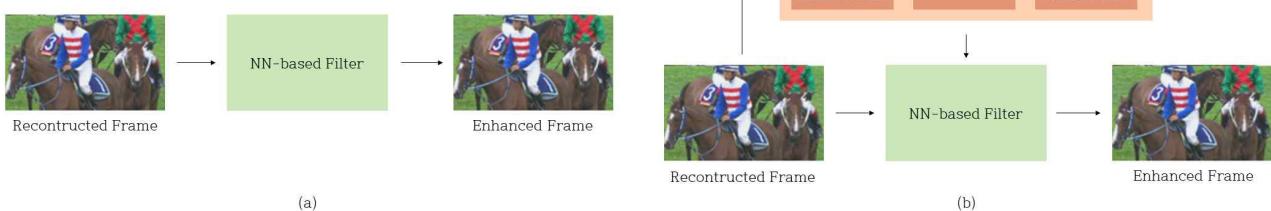


그림 1. 일반적인 딥러닝 기반 필터링과 컨텐츠 적응적 필터링 구조 (a) 일반적인 딥러닝 기반 필터링; (b) 컨텐츠 적응적 필터링

Fig. 1. Deep learning-based general filtering and contents adaptive filtering structure (a) General filtering; (b) Contents adaptive filtering

기술들이 활용되었지만 필터링 성능에 한계가 존재하여 convolutional neural network (CNN) 기반의 필터들이 제안되었다. 하지만 기존의 CNN을 기반으로 한 루프 내 필터는 재구성된 프레임의 화질을 향상시키기 위해 모델의 복잡도가 크게 증가하여 연산량이 과도하게 증가하는 문제점이 발생한다. 딥러닝 기반 필터의 모델 복잡도를 감소시키고 복원 성능을 향상시키기 위해서, 필터링 과정에 도움이 되는 정보들을 추가 입력으로 사용한 연구들이 제안되었다. 이후에 소개할 코덱의 정보를 추가로 활용하는 모델은 네트워크의 복잡도 증가를 예방하고 화질 복원에 더 향상된 결과를 보인다.

이번 장에서 소개하는 코덱 정보를 추가로 사용하는 딥러닝 기반 필터링 기술들은 다양한 코덱 정보들을 활용하는데, 크게 1) Residual 정보를 추가 입력으로 사용하는 기술, 2) 여러 QP 옵션을 적용할 수 있는 기술, 3) QP 정보와 블록 정보를 추가로 활용하는 기술로 나누어 정리하였다.

1. Residual 정보 추가 활용

블록 단위 코딩은 하드웨어에 친화적이고 병렬화에 유리하기 때문에 정지영상 뿐만 아니라 동영상 압축에도 사용된다. 하지만 블록 단위로 동영상을 압축함에 따라, 복원된 프레임 내에서 분할된 블록 경계를 따라 블로킹 열화가 발생한다. 이 때 예측 영상과 원본 영상 간의 차이인 잔차 영상이 블로킹 열화를 줄이는데 도움이 될 수 있다.

비디오 프레임이 블록 단위로 나뉠 때 변환 및 양자화로 인해 블록의 중앙보다 경계에서 열화가 발생하게 된다. 그림 2에서 볼 수 있듯이 동영상 압축시 발생하는 예측 영상과 원본 영상 간의 차이인 잔차 영상은 변환 및 양자화 과정에서 많은 정보를 손실함에도 불구하고 원본 영상의 텍스쳐를 보존하고 있음을 확인할 수 있다. 따라서 Residual-reconstruction-based CNN (RRNet)^[4]의 저자들은 잔차 영상이 블록 분할에 대한 좋은 정보를 가지고 있어 프레임 내 경계와 텍스쳐 영역의 재구성에 도움이 된다고 주장하였으며 이러한 특성에 주목하여 딥러닝 모델이 잔차 영상과 복원된 영상을 동시에 입력으로 받아 두 영상의 특성을 포착하여 압축된 영상의 화질을 향상시키는 방식을 제안했다.

그림 3(a)의 in-loop based on low-complexity CNN using residuals (CNMF-R)^[5] 또한 잔차 영상을 추가 입력으로 활용했다. CNMF-R의 저자들은 딥러닝 네트워크에 영상 압축의 다양한 블록 분할 환경을 고려하여 생성된 잔차 영상을 추가로 입력하면 화질 향상 뿐만 아니라 모델의 경량화에도 도움이 될 수 있다고 주장했다. 블록 파티션 정보와 프레임 텍스쳐 정보를 가진 잔차 영상을 추가로 입력하게 되면 모델 학습 시에 유용한 정보가 사용되어 필터링 성능 향상에 도움이 될 수 있다. 또한, 제안된 방식으로 학습한 모델은 노이즈가 존재하는 재구성된 프레임만을 입력 받는 기존 네트워크^[6]보다 화질 향상 성능이 증가하며 복잡도가 약 30% 감소했다.



그림 2. 좌: 원본 영상, 우: 잔차 영상^[4]
Fig. 2. Left: Original frame, Right: Residual frame^[4]

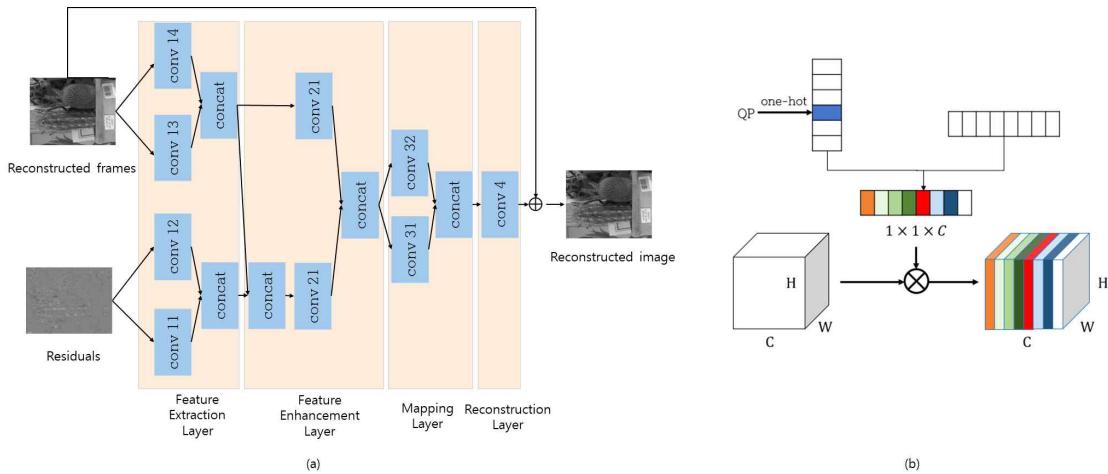


그림 3. (a) CNNF-R 네트워크 구조^[5]; (b) QP attention module 구조^[7]
Fig. 3. (a) CNNF-R network structure^[5]; (b) QP attention module structure^[7]

2. QP 정보 추가 활용

동영상을 압축할 때 인-루프 필터는 압축 열화를 감소시키고 객관적/주관적 화질을 높이기 위해 사용된다. 압축 열화는 주로 양자화 과정에서 발생하는데, QP에 따라 압축 노이즈가 다르게 발생한다. QP 값이 클수록 더 큰 양자화 오차가 발생하며, 복원 영상의 화질 열화도 더 심해진다. 발생하는 노이즈의 정도가 다르기 때문에, 기존의 네트워크들은 각 QP 별로 각기 다른 CNN 모델을 학습시켜야 했다. 이를 해결하기 위해 QP 별로 최적의 모델을 할당하여 학습하는 방식이나, 기존의 QP를 블록이나 프레임 크기로 확장하여 단순히 추가해주는 방식이 제안되었지만 여전히 학습에 여러 모델이 필요하거나, 오히려 성능이 떨어지는 등의 문제점이 잔존했다.

QP attention module (QPAM)^[7]의 저자들은 그림 3(b)와 같이 입력 특성 맵에 QP 값에 따라 다른 채널 별 attention 을 적용하는 모듈을 제안했다. 이 방식은 QP에 따라 압축 정도를 판별하고 채널 축에 유의미한 특징을 강조하여 I-frame, P-frame, B-frame 같이 다양한 프레임 종류들과 다양한 QP에 적응적으로 필터링을 수행하도록 했다. 또한 1 차원 QP attention map을 얻어 하나의 CNN 모델이 여러 QP 값을 다룰 수 있도록 설계했다. 한편, JVET-V0101^[8]은 QP 정보를 다른 방식으로 활용하는 방법을 제안하였는데 QP 정보를 각 단계마다 fully connected layer를 거쳐 추가

해주는 방식을 사용하여 성능을 향상시켰다.

3. QP 정보와 블록 정보 추가 활용

Attention-based dual-scale CNN (ADCNN)^[9]의 저자들은 그림 4(a)처럼 QP 정보를 딥러닝 모델에 입력하고 추가로 블록 분할 정보까지 입력하여 압축이 진행되는 코덱 내부의 정보를 네트워크가 충분히 활용할 수 있는 구조를 제안했다. 다른 QP 값에 따라 다른 정도의 복원을 수행하기 위해 현재 휘도 혹은 색차 신호의 정규화된 QP 값을 전달하는 QPmap을 구성하였으며, coding unit map의 경계 정보를 함께 입력해 복원과 attention이 필요한 지역을 알려주었다.

한편, 영상 압축에 자주 사용되는 YUV420 포맷은 휘도 (luma, Y) 변화가 색차 (chroma, UV)의 변화보다 넓은 범위를 갖는다. 이에 따라 원본 색차 신호는 밝기 신호보다 1/4 크기로 저장된다. 그림 4(b)에서 볼 수 있듯이 YUV420 포맷의 색차 신호는 휘도 신호에 비해 그 편차가 적고 휘도 신호와 비슷한 값을 가진다. ADCNN은 이러한 휘도 신호와 색차 신호의 특성을 활용하기 위해 딥러닝 모델이 attention 구조를 통해 공간 정보와 채널 정보에서 추출한 특징지도 (feature map)를 융합하여 각 성분이 서로의 특성을 활용하여 더 좋은 화질의 영상을 복원할 수 있도록 하였다.

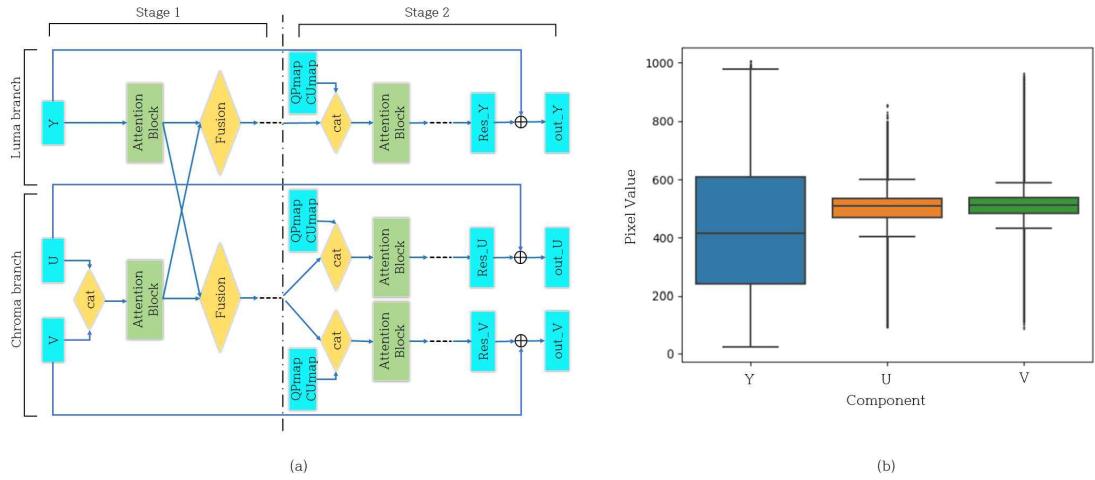


그림 4. (a) ADCNN 네트워크 구조^[9]; (b) 휘도(Y), 색차(U, V) 성분의 동적 영역(dynamic range) 도표^[9]
Fig. 4. (a) ADCNN network structure^[9]; (b) Luma(Y), Chrominance(U, V) components' dynamic range char^[9]

III. 모델 선택

일반적으로 딥러닝 모델의 파라미터가 많을수록 더 높은 필터링 수행 능력을 보여줄 수 있게 된다. 따라서 성능을 높이기 위해 많은 파라미터를 갖는 큰 모델이 필요하고, 이는 수행시간의 증가로 이어진다. 이를 극복하기 위해서 특성이 같은 영상들을 모아서 같은 특성을 갖는 영상들에 대해서만 모델을 학습시키면 해당 영상들에 대해서는 큰 모델을 사용하지 않아도 높은 성능을 낼 수 있다. 이에 따라 성능을 높이며 모델 크기를 줄이기 위해서는 영상을 일정한 특성에 따라 그룹을 나눈 후 각 그룹마다 별도의 모델을 학습시키고, 입력 영상을 필터링할 최적의 모델을 선택하는 모델 선택 방법을 사용할 수 있다.

모델 선택 방식은 다음과 같다. 우선 데이터를 일정한 특성에 따라 분류한 뒤에 분류한 데이터마다 별도의 모델을 학습하여 모델들이 모인 모델 집합을 구성한다. 그리고 테스트시에 적절한 방법으로 입력 영상의 특성이 어떤 데이터 분류에 속하는지 찾은 후 모델 집합에서 해당 데이터에 특화된 모델을 사용한다.

모델 선택은 딥러닝 기반 영상 압축 기술에서 널리 쓰이고 있는 방법인데, QP마다 영상의 열화 정도가 다르기에 특정 QP에 특화된 모델을 사용하는 것이 부호화 성능 향상에 효과적이기 때문이다. 그러나 이 장에서는 QP에 따른

모델 선택 뿐 아니라 같은 QP에서도 다른 모델을 선택하는 방식 또한 소개한다. 모델 선택 방식을 1) 코덱의 정보를 활용한 모델 선택, 2) rate-distortion cost (RD cost)에 기반한 모델 선택, 3) 딥러닝 모델을 활용한 모델 선택으로 나누어 정리하였다.

1. 코덱 정보 활용

코덱에는 영상에 대한 많은 정보가 저장되어 있는데, 이러한 정보들은 영상이 어떻게 부호화되었는지를 저장하여 복호화기에 전달하는 역할을 하고 있다. 이때 영상이 부호화된 모드에 따라 열화된 영상의 특성이 달라지게 되고, 같은 방식으로 부호화된 프레임들은 같은 특성을 공유하게 된다. 동영상 압축의 이러한 특성에 따라서, 이번에 소개할 기술은 코덱의 정보를 이용하여 데이터를 분류하고 해당 데이터들에 특화된 모델 집합을 구성하여 모델을 선택하는 방식이다.

Random access (RA)와 같이 화면 간 예측을 수행하는 설정에서는 화면 내 예측을 하는 I 프레임과 화면 간 예측을 하는 B 프레임은 예측 프레임 생성 방식 및 화질 등 여러 부분에서 차이가 있다. 이러한 특성을 구분하여 학습시키기 위해 dense Inception network (DIA_Net)^[10]의 저자들은 I 프레임 모델과 B 프레임 모델을 분리하여 학습하였다.



그림 5. 모델 선택 기준에 따른 모델 선택 구조도 (a) 선택 네트워크를 활용한 모델 선택 방식; (b) 모델 인덱스 전송을 통한 모델 선택 방식
Fig. 5. Model selection methods: (a) Selection network; (b) Transmission model indexing.

I 프레임 모델은 all intra (AI) 설정의 프레임과 RA 설정의 I 프레임을 사용하여 학습하였고 B 프레임 모델은 RA의 B 프레임을 사용하여 학습하였다. 학습된 모델을 테스트할 때는 입력 영상의 부호화 모드에 따라 I 프레임 혹은 B 프레임 모델을 선택하여 필터링을 수행했다.

2. 선택 네트워크 활용

앞서 코덱의 정보를 활용하여 모델 선택을 하는 방법을 알아보았다. 그러나 모델을 코덱 정보만으로 분류하기에는 한계가 존재한다. 비트스트림의 헤더에는 다양한 정보들이 존재하는데 대부분은 블록 분할 구조 정보와 같이 세분화된 정보들은 모델의 열화의 특성과 관계가 있으나 이러한 정보들은 대부분 직접적으로 열화의 특성을 알려주지는 않는다. 따라서 이번 장에서는 이를 보완하기 위해 그림 5(a)와 같이 입력 영상을 가장 잘 필터링하는 모델을 선택하는 선택 네트워크를 제안한 기술을 소개한다.

JVET-X0126^[11]에서는 모델 선택 네트워크를 학습시켜 열화의 정도에 따라 모델을 선택하게 하였다. 우선 특정 QP로 압축된 영상들에 대해 모델을 학습시키고 학습 데이터셋의 패치들이 어떤 모델에서 가장 높은 성능을 보이는지를 라벨링하였다. 이때 QP에 따라 모델들이 학습되었기 때문에 일반적으로 열화가 적은 패치는 낮은 QP로 학습된 모델이 라벨링되게 된다. 이에 따라 각 패치가 어떤 라벨을 가지는지 모델 선택 네트워크를 학습시키면 입력 영상의 열화의 정도를 구분할 수 있다. 열화의 정도는 QP에 영향을 많이 받지만 높은 QP가 높은 열화를 보장하지 않기 때문에 높은 QP로 부호화된 영상이라도 낮은 QP로 학습된 모델이 더 높은 성능을 보이는 경우가 있으므로 패치마다

적절한 모델을 선택할 필요가 있다. 제안 기술에서는 앞서 학습한 모델 선택 네트워크를 사용하여 가장 적절한 QP 모델을 선택하여 입력 영상에 필터링을 수행하였다.

Jia, Chuanmin, et al^[12] 저자들도 유사하게 QP별 모델들을 사용하여 모델 집합을 구성하고 학습 패치에 라벨링하여 모델 선택 네트워크를 학습시켰다. 그뿐 아니라 모델 선택 네트워크를 라벨에 맞게 학습하고, 모델 선택 네트워크의 결과를 통해 모델 집합의 모델들을 재학습하여 성능을 향상시켰다.

3. 모델 인덱스 전송

부호화기는 원본 영상을 알고 있기 때문에 어떤 모델이 가장 좋은 복원 성능을 보여주는지 알고 있다. 그렇기에 부호화기가 각 모델의 성능을 비교하여 모델을 선택하면 이상적으로 모델을 선택하는 것이 가능하다. 그러므로 그림 5(b)와 같이 부호화기가 어떤 모델을 선택하였는지에 대한 인덱스를 복호화기에 전송하면 복호화기에서도 같은 모델을 선택할 수 있다.

iDAM^[13], JVET-X0065^[14], JVET-X0054^[15]의 저자들은 여러 QP로 학습된 모델 중 최적의 모델을 선택하는 방식을 사용하였다. 동영상 압축에서는 QP에 따라 영상 화질에 크게 차이가 나기 때문에 대부분의 딥러닝 기반 인-루프 필터 모델들은 QP별로 다른 모델을 학습한다. QP가 낮을수록 모델이 영상 복원을 적게 하므로 QP에 따라 나눈 모델들이 필터링의 강도에 따라 나뉘졌다고 할 수 있다. 이런 점을 이용하여 더 원본과 유사하게 복원하는 모델을 RD cost를 기준으로 선택하고, 선택 정보를 복호화기에 전송하는 방식을 채택하였다. 선택 정보의 오버헤드를 줄이기 위해

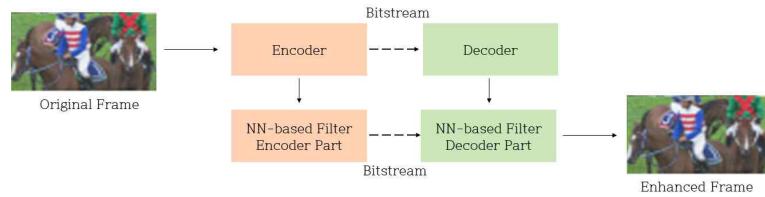


그림 6. 추가 정보 전송 방식의 구조도
 Fig. 6. Additional information transfer methods structure

iDAM, JVET-X0065, JVET-X0054에서는 모델 집합에 세 가지 모델만을 포함시켜 선택하게 하였다.

Adaptive-switching network (ASN)^[16]는 학습에 사용될 패치들을 랜덤으로 군집화한 후에 각 군집에 대해 별도의 모델을 학습시켰다. 이후 모델에 맞춰 학습 패치들을 다시 군집화하고 각 군집마다 모델을 재학습시켰다. 다양한 군집화 방법에 대하여 ablation study를 수행하였으며 원본과 개선 이미지의 차이 값에 discrete cosine transform을 적용한 후 t-SNE^[17]를 통해 차원축소를 하고 난 뒤에 군집화한 것이 가장 좋은 성능을 보였다. 테스트 환경에서는 입력 영상을 가장 높은 화질로 필터링하는 모델을 선택하였다.

IV. 추가 정보 전송

3장에서 설명한 바와 같이 단순한 압축 영상만으로는 제한된 크기의 모델로 성능을 향상시키기에 한계가 존재한다. 이를 개선하기 위해 2장과 같이 코덱의 정보를 활용하거나 3장과 같이 데이터를 분류하여 분류된 데이터에 특화된 모델을 만드는 등의 연구들이 수행되었다. 이번 장에서는 추가적인 정보를 전송하여 영상에 특화된 정보를 활용하는

기술을 소개한다.

영상을 압축하는 부호화기는 압축 전의 원본 영상과 압축 후의 열화된 영상을 둘 다 가지고 있으며 복원에 사용된 모델 또한 가지고 있다. 이를 활용하면 그림 6과 같이 모델이 복원 성능을 향상시키기 위해서 필요한 정보를 만들어 전송하여 디코더에서 활용하는 것이 가능하다. 추가 정보를 전송함에 따라 전송해야 할 비트량은 늘지만 그 이상의 화질 향상을 기대할 수 있다면 보내는 비트량 대비 압축 효율은 증가할 수 있다.

추가로 전송할 정보는 모델에 따라 자유롭게 선택할 수 있지만, 현재까지 제안된 방식을 1) 모델 업데이트 정보 전송, 2) 모델에 의한 오프셋 전송, 3) 합 가중치 전송로 분류하여 정리하였다.

1. 모델 파라미터 업데이트

딥러닝에서 모델을 특정 데이터에만 지속적으로 학습하게 되면 그 학습 데이터에 대해서는 성능이 높아지게 되지만 그 외의 데이터에 대해서는 성능이 감소하게 되어 범용성이 떨어지는 과적합 현상이 일어난다. 이는 학습 데이터 외의 데이터에서도 좋은 성능을 보여 범용성을 가져야 하

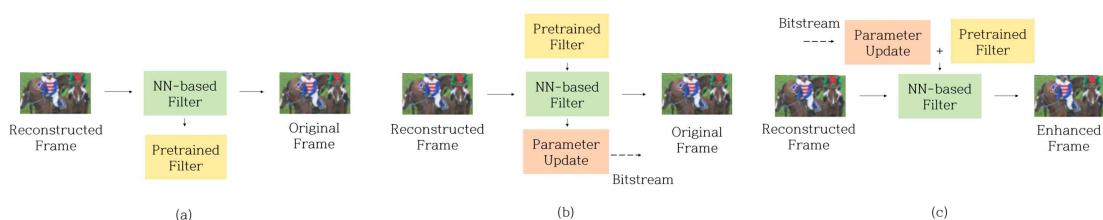


그림 7. 모델 파라미터 업데이트 방식의 과정 (a) 모델을 학습하는 과정; (b) 부호화기에서 영상에 모델을 재학습 후 업데이트 정보 전송 과정; (c) 복호화기에서 모델 업데이트하여 사용하는 과정

Fig. 7. Processes for updating model parameters methods (a) Pretrain model; (b) Finetune model to sequence in encoder; (c) Update and use model in decoder

는 일반적인 문제에서는 피해야 하는 현상이다.

그러나 이번에 소개하는 기술은 반대로 과적합 현상에서 학습 데이터에 대해서는 성능이 높아지는 점을 이용하여 영상 적응적으로 필터링하는 기술이다. 그림 7은 이 기술의 과정을 나타낸 그림이다. 그림 7(b)에서 영상 부호화 중에 딥러닝 모델의 출력 영상과 원본 영상의 차이를 손실함수로 이용하여 모델을 학습시킨 뒤에 파라미터의 업데이트 정보를 전송하고, 그림 7(c)에서 업데이트 정보를 사용하여 복호화기에서도 부호화기에서 학습시킨 것과 동일한 모델이 되도록 파라미터를 업데이트하여 성능을 향상시킨다. 일반적으로 모델의 파라미터를 전부 전송하는 것은 전송 비트량이 너무 높아지기에 비효율적이다. 따라서 파라미터 중 일부만 학습하고 전송하는 방식을 사용한다.

Lam, Yat-Hong, et al^[18], JVET-Y0059^[19]의 저자들은 모델의 노이즈 제거 등의 필터링 정보는 이차원 커널(kernel)에 저장되어 있을 것이라고 가정하였고 이에 따라 기저

(bias)만 업데이트하는 방식을 사용하였다. 먼저 학습 데이터셋에 사전 학습된 모델을 만들고 테스트 시에 입력 영상과 원본 영상의 차이를 손실 함수로 사용하여 사전 학습된 모델의 기저만 학습시키고 업데이트 정보를 전송하였다. 이에 따라 복호화기에서 부호화기에서 학습한 모델처럼 업데이트하여 성능을 향상시켰다.

JVET-Z0082^[20]의 저자들은 앞선 기저 업데이트 방식보다 개선된 업데이트 방식을 제안하였다. 제안 기술에서는 각 CNN의 출력에 multiplier를 곱하는 방식을 사용한다. 사전 학습시에는 multiplier를 1로 고정하여 학습한 뒤 커널 및 기저 파라미터를 복호화기와 공유한다. 그리고 테스트 환경에서 각 영상에 재학습시에 커널 및 기저 파라미터를 고정하고 multiplier만 학습하여 업데이트 정보만 전송하여 복호화기에서도 multiplier를 같이 사용하는 방식을 사용한다.

표 1. 소개된 기술들의 컨텐츠 적응적 필터링 방법

Table 1. Contents adaptive filtering methods in this paper

| Paper/ Contribution | Method | Method Details | |
|--------------------------------------|--|--|--|
| RRNet ^[4] | Encoder information exploitation | Information source | Residual image |
| CNNF-R ^[5] | | | Residual image |
| QPAM ^[7] | | | QP |
| JVET-V0101 ^[8] | | | QP |
| ADCNN ^[9] | | | QP, Block partitioning |
| DIA_Net ^[10] | Model selection | Selection criterion/ Model candidates | I/B Frame/ I/B models |
| JVET-X0126 ^[11] | | | Model selection network/ QP models |
| Jia, Chuanmin, et al ^[12] | | | Model selection network/ Finetuned QP models |
| iDAM ^[13] | | | Best model's index signaling/ QP models |
| JVET-X0065 ^[14] | | | Best model's index signaling/ QP models |
| JVET-X0054 ^[15] | | | Best model's index signaling/ QP models |
| ASN ^[16] | | | Best model's index signaling/ Multi-group data models |
| Lam, Yat-Hong, et al ^[18] | Additional information signaling | Signaling information | Bias updates |
| JVET-Y0059 ^[19] | | | Bias updates |
| JVET-Z0082 ^[20] | | | Multiplier parameter |
| ECN and OEN ^[21] | | | Model offsets |
| GNR ^[22] | | | Sum of weights |
| Revisiting ^[23] | | | Sum of weights |

2. 모델에 의한 오프셋 전송

앞서 소개한 모델 파라미터 업데이트 방식은 영상에 맞춰 업데이트시킨 파라미터를 전송하는 방식이다. 그러나 이러한 방식은 모델의 사전 학습 단계에서 모델을 재학습 시켜 사용할 것이라는 점이 반영되어 있지 않아 이상적인 방식이라고는 하기 어렵다.

이번에 소개하는 방식은 모델이 자체적으로 복호화기에 전송할 정보를 생성한다. 이때 사용하는 정보 또한 모델이 만들기 때문에 보내는 정보 또한 학습을 통해 성능을 최대한 높일 수 있도록 최적화될 수 있다.

Error classification network (ECN) and offset estimation network (OEN)^[21]의 저자는 딥러닝 기반 필터와 오프셋 생성 네트워크를 설계하였다. 오프셋 생성 네트워크는 원본 영상만 열화된 영상을 사용하여 숫자 몇 개의 오프셋을 생성하고 해당 오프셋을 딥러닝 기반 필터에서 채널 attention으로 적용시켜 향상된 이미지를 생성하게 하였다. 실제 테스트 단계에서는 부호화기에서 생성된 오프셋을 전송받아 복호화기에서는 오프셋 생성 과정 없이 딥러닝 기반 필터에 오프셋을 적용만하여 사용한다.

3. 합 가중치 전송

합 가중치 전송 방식은 모델의 출력을 다수의 채널로 구성하고 각 채널의 신호를 선형 결합하여 최종 출력 영상을 구성하는 방식으로, 이 때 선형 결합에 사용하는 가중치를 추가 정보로 전송한다. 이 방식은 우선 학습시 모델의 출력과 원본 영상의 차이가 최소가 되도록 최소자승법을 사용하여 얻어진 가중치를 각 채널 출력에 곱하여 최종 출력 영상을 만들어 학습시키고 테스트 시에는 마찬가지로 부호화기에서 최소자승법을 통해 가중치를 만들고 이를 전송한 뒤에 복호화기에서 가중치를 곱하여 최종 출력을 낸다.

Guided CNN Restoration (GNR)^[22]와 Bordes, Philippe, et al^[23]의 저자들은 앞 문단에서 설명한 합 가중치 전송 방식을 사용하였는데 최종 출력의 각 채널의 신호가 서로 다른 특성을 갖는다. Bordes, Philippe, et al^[23] 저자들은 기존 인-루프 필터중 하나인 sample adaptive offset (SAO)

가 입력 영상에 따라 오프셋을 결정하듯이 여러 개의 모델을 학습시켜 각 모델의 결과를 선형 결합하는 오프셋(선형 결합 가중치)을 결정하여 필터링을 수행했다. 제안된 오프셋을 계산할 때는 GNR과 유사하게 최소자승법을 통해 모델의 선형 결합 가중치를 구하여 복호화기에 전송했다.

V. 전망과 결론

본 논문에서는 딥러닝 기반 영상 적응적 압축의 기술 동향에 대해 알아보았다. 통신 기술의 발전으로 통신의 대역폭이 날로 커지고 있지만, 영상 콘텐츠에 대한 트래픽은 대역폭이 증가하는 것 이상의 폭발적인 증가세를 보이기 때문에 더 효율적인 영상 압축 알고리즘의 필요성이 그 어느 때 보다 절실히다. 한편, 컴퓨팅 파워의 증가와 AI 기술의 발전이 딥러닝에 대한 관심과 수요를 확대시켰고, 컴퓨터 비전 등 다양한 분야에서 딥러닝이 활용되고 있다. 이에 따라 영상 압축 분야에서도 딥러닝을 도입하여 기준 압축 표준을 뛰어넘는 성능을 보이고 있으나 딥러닝 기반 기술들은 너무 큰 수행시간을 요구하는 문제들이 있다.

이에 수행시간을 줄이며 성능을 높이기 위해 딥러닝 기반 영상 적응적 압축 기술들의 연구가 활발히 진행되고 있다. 이러한 기술들은 코덱 정보 활용, 모델 선택, 추가 정보 전송을 통해 성능을 높이면서도 복잡도를 낮추고자 한다. 그러나 이러한 노력에도 불구하고 아직 딥러닝 기반 기술들은 너무 큰 수행시간을 요구하기에 영상 압축에 활용되기 위해서는 더욱 발전이 필요하다. 후속 연구들이 영상 적응적 기술들을 좀 더 발전시켜 이러한 문제들을 해결한다면 딥러닝 기반 압축 기술들이 상용화되어 사회에 기여할 수 있으리라 기대한다.

참 고 문 헌 (References)

- [1] Cisco, Cisco Annual Internet Report (2018-2023) White Paper, Mar. 2020.
- [2] A. Skodras, C. Christopoulos and T. Ebrahimi, "The JPEG 2000 still image compression standard," Signal Processing Magazine, Vol.18., No.5, pp 36-58, 2001.
doi: <https://doi.org/10.1109/79.952804>
- [3] B. Bross, J. Chen, S. Liu and Y.-K. Wang, "Versatile video coding

- (Draft 10)," JVET-S2001, Jul. 2020.
- [4] Wei Jia, et al "Residual-guided In-loop Filter Using Convolution Neural Network," ACM Trans. Multimedia Comput. Communications, and Applications, 2021
doi: <https://doi.org/10.1145/3460820>
- [5] Li, Daowen, and Lu Yu. "An in-loop filter based on low-complexity CNN using residuals in intra video coding," IEEE International Symposium on Circuits And Systems 2019.
doi: <https://doi.org/10.1109/ISCAS.2019.8702443>
- [6] Dai, Yuanjing, Dong Liu, and Feng Wu. "A convolutional neural network approach for post-processing in HEVC intra coding," International Conference on Multimedia Modeling, Springer, Cham, pp. 28-39, 2017.
doi: https://doi.org/10.1007/978-3-319-51811-4_3
- [7] Huang, Zhijie, et al. "An efficient QP variable convolutional neural network based in-loop filter for intra coding." IEEE Data Compression Conference, pp. 33-42, 2021.
doi: <https://doi.org/10.1109/dcc50243.2021.00011>
- [8] Y. Li, L. Zhang, K. Zhang, "Conditional in-loop filter with parameter selection", JVET-V0101, Apr. 2021.
- [9] Wang, Ming-Ze, et al. "Attention-based dual-scale CNN in-loop filter for versatile video coding," IEEE Access, Vol.7, pp. 145214-145226, 2019.
doi: <https://doi.org/10.1109/access.2019.2944473>
- [10] Xu, Xiaoyu, et al. "Dense inception attention neural network for in-loop filter," IEEE Picture Coding Symposium, pp. 1-5, 2019.
doi: <https://doi.org/10.1109/pcs48520.2019.8954499>
- [11] Z. Dai, et al, "AHG11: Neural network-nased adaptive model selection for CNN in-loop filtering", JVET-X0126, Oct. 2021.
- [12] Jia, Chuanmin, et al. "Content-aware convolutional neural network for in-loop filtering in high efficiency Video coding," IEEE Transactions on Image Processing, Vol.28, No.7, 2019.
doi: <https://doi.org/10.1109/tip.2019.2896489>
- [13] Li, Yue, Li Zhang, and Kai Zhang. "IDAM: Iteratively trained deep in-loop filter with adaptive model selection," ACM Transaction on Multimedia Computing, Communications, and Application, 2022.
doi: <https://doi.org/10.1145/3529107>
- [14] Y. Li, K. Zhang, and L. Zhang. "EE1-1.2: Test on deep in-loop filter with adaptive model selection and external attention," JVET-X0065, Oct, 2021.
- [15] L. Wang, X. Xu, and S. Liu, "AHG11: Neural network based in-loop filter with adaptive model selection," JVET-X0054, Oct. 2021.
- [16] W. Lin, et al. "Partition-aware adaptive switching neural networks for post-processing In HEVC," IEEE Transactions on Multimedia, Vol.22, No.11, pp. 2749-2763, 2019.
doi: <https://doi.org/10.1109/tmm.2019.2962310>
- [17] L. van Der Maaten, and G. Hinton. "Visualizing data using t-SNE," Journal of Machine Learning Research, Vol.9, No.11, 2008.
- [18] Lam, Yat-Hong, et al. "Efficient adaptation of neural network filter for video compression." Adaptive Model Selection," ACM International Conference on Multimedia, pp. 358-366, 2020.
doi: <https://doi.org/10.1145/3394171.3413536>
- [19] M. Santamaría, et al. "AHG11: Hannuksela, Content-adaptive post-processing filter," JVET-Y0059, Jan. 2022.
- [20] M. Santamaría, et al. "AHG11: Content-adaptive neural network post-filte," JVET-Z0082, Apr. 2022.
- [21] Lee, So Yoon, et al. "Offset-based in-loop filtering with a deep network in HEVC," IEEE Access, Vol.8, pp. 213958-213967, 2020.
doi: <https://doi.org/10.1109/access.2020.3040751>
- [22] Kong, Lingyi, et al. "Guided CNN restoration with explicitly signaled linear combination," IEEE International Conference on Image Processing, pp. 3379-3383, 2020.
doi: <https://doi.org/10.1109/icip40778.2020.9190807>
- [23] Bordes, Philippe, et al. "Revisiting the sample adaptive offset post-filter of VVC with neural-networks," IEEE Picture Coding Symposium, pp. 1-5, 2021.
doi: <https://doi.org/10.1109/pcs50896.2021.9477457>

저자 소개

한창우

- 
- 2022년 2월 : 고려대학교 전기전자공학부 학사
 - 2022년 3월 ~ 현재 : 고려대학교 전기전자공학과 석사과정
 - ORCID : <https://orcid.org/0000-0003-4689-2000>
 - 주관심분야 : 영상 처리, 영상 압축

저자 소개

김 홍 일



- 2015년 3월 ~ 현재 : 동국대학교 전기전자공학부 학사과정
- ORCID : <https://orcid.org/0000-0003-2540-0126>
- 주관심분야 : 영상 처리, 영상 압축

강 현 구



- 2021년 2월 : 동국대학교 멀티미디어공학과 학사
- 2021년 3월 ~ 현재 : 고려대학교 전기전자공학과 석사과정
- ORCID : <https://orcid.org/0000-0002-6556-374X>
- 주관심분야 : 영상 처리, 영상 압축

권 형 진



- 1997년 2월 : KAIST 전기 및 전자공학과 학사
- 2001년 2월 : KAIST 전기 및 전자공학과 석사
- 2001년 3월 ~ 현재 : 한국전자통신연구원 통신미디어연구소 책임연구원
- ORCID : <https://orcid.org/0000-0002-2888-3154>
- 주관심분야 : 영상 신호처리, 영상 압축, 기계학습

임 성 창



- 2006년 2월 : 세종대학교 컴퓨터공학과 학사
- 2008년 2월 : 세종대학교 컴퓨터공학과 석사
- 2022년 2월 : 세종대학교 컴퓨터공학과 박사
- 2008년 3월 ~ 현재 : 한국전자통신연구원 통신미디어연구소 책임연구원
- ORCID : <https://orcid.org/0000-0002-0110-7489>
- 주관심분야 : 영상 압축, 영상 처리

정 승 원



- 2005년 8월 : 고려대학교 전기전자공학과 학사
- 2011년 2월 : 고려대학교 전기전자공학과 박사
- 2012년 3월 ~ 2014년 2월 : 삼성 종합기술원 전문연구원
- 2014년 3월 ~ 2020년 2월 : 동국대학교 조교수
- 2020년 3월 ~ 2020년 8월 : 고려대학교 조교수
- 2020년 9월 ~ 현재 : 고려대학교 전기전자공학과 부교수
- ORCID : <https://orcid.org/0000-0002-0319-4467>
- 주관심분야 : 영상 처리, 컴퓨터 비전, 영상 압축