특집논문 (Special Paper)
방송공학회논문지 제23권 제3호, 2018년 5월 (JBE Vol. 23, No. 3, May 2018)
https://doi.org/10.5909/JBE.2018.23.3.383
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

딥 러닝 기반의 이미지와 비디오 압축 기술 분석

조 승 현[a)‡], 김 연 희[a)], 임 웅[a)], 김 휘 용[a)], 최 진 수[a)]

A Technical Analysis on Deep Learning based Image and Video Compression

Seunghyun Cho[a)‡], Younhee Kim[a)], Woong Lim[a)], Hui Yong Kim[a)], and Jin Soo Choi[a)]

요 약

본 논문에서는 최근 활발히 연구되고 있는 딥 러닝 기반의 이미지와 비디오 압축 기술에 대해 살펴본다. 딥 러닝 기반의 이미지 압축 기술은 심층 신경망에 압축 대상 이미지를 입력하고 반복적 또는 일괄적 방식으로 은닉 벡터를 추출하여 부호화한다. 이미지 압축 효율을 높이기 위해 심층 신경망은 복원 이미지의 화질은 높이면서 부호화된 은닉 벡터가 보다 적은 비트로 표현될 수 있도록 학습된다. 이러한 기술들은 특히 저 비트율에서 기존의 이미지 압축 기술에 비해 뛰어난 화질의 이미지를 생성할 수 있다. 한편, 딥 러닝 기반의 비디오 압축 기술은 압축 대상 비디오를 직접 입력하여 처리하기 보다는 기존 비디오 코덱의 압축 툴 성능을 개선하는 접근법을 취하고 있다. 본 논문에서 소개하는 심층 신경망 기술들은 최신 비디오 코덱의 인루프 필터를 대체하거나 추가적인 후처리 필터로 사용되어 복원 영상의 화질 개선을 통해 압축 효율을 향상시킨다. 마찬가지로, 화면 내 예측 및 부호화에 적용된 심층 신경망 기술들은 기존 화면 내 예측 툴과 함께 사용되어 예측 정확도를 높이거나 새로운 화면 내 부호화 과정을 추가함으로써 압축 효율을 향상 시킨다.

Abstract

In this paper, we investigate image and video compression techniques based on deep learning which are actively studied recently. The deep learning based image compression technique inputs an image to be compressed in the deep neural network and extracts the latent vector recurrently or all at once and encodes it. In order to increase the image compression efficiency, the neural network is learned so that the encoded latent vector can be expressed with fewer bits while the quality of the reconstructed image is enhanced. These techniques can produce images of superior quality, especially at low bit rates compared to conventional image compression techniques. On the other hand, deep learning based video compression technology takes an approach to improve performance of the coding tools employed for existing video codecs rather than directly input and process the video to be compressed. The deep neural network technologies introduced in this paper replace the in-loop filter of the latest video codec or are used as an additional post-processing filter to improve the compression efficiency by improving the quality of the reconstructed image. Likewise, deep neural network techniques applied to intra prediction and encoding are used together with the existing intra prediction tool to improve the compression efficiency by increasing the prediction accuracy or adding a new intra coding process.

Keyword : Machine learning, Deep learning, Image, Video, Compression

Ⅰ. 서 론

최근 들어, 오브젝트 검출(detection)/인식(recognition)/추적(tracking), 분할(segmentation) 등의 컴퓨터 비전 분야와 초 해상도(super resolution), 왜곡 제거(artifact reduction) 등의 영상처리 분야에서 기계 학습 기반의 기술들은 눈부신 발전을 거듭하며 기존 알고리즘들을 뛰어 넘는 성능을 보이고 있다. 이는, 기계 학습 중에서도 딥 러닝(deep learning) 기술의 발전에 따른 것으로, 심층 신경망(deep neural network)에서의 학습과 관련된 난제들이 해결되었을 뿐만 아니라 학습 효율성을 높이기 위한 다양한 방법론들이 제시되고 있으며, 이와 함께 기술 구현을 위한 하드웨어 및 소프트웨어 환경이 비약적으로 발전하고 있기 때문이다. 본 논문에서는 딥 러닝에 기반하여 이미지와 비디오의 압축 효율을 개선하고자 하는 최신 연구들에 대해 분석하고 발전방향에 대해 전망해보고자 한다.

1990년대 또는 그 이전에도, 이미지 압축의 효율성을 높이기 위해 기계 학습을 적용하고자 했던 시도들이 있었으며, 이러한 시도들을 체계적으로 분석한 연구 논문[1]에서 가능한 접근 방법을 세가지 범주로 분류하였다. 첫째, 이미지 압축을 위해 신경망(neural network)을 직접 이용하는 방법은 신경망에 압축 대상 이미지를 입력하고 이를 통해 획득한 은닉 층(hidden layer)의 뉴런 값을 부호화한다. 둘째, 이미지 압축 요소 기술을 신경망으로 구현하는 방법은 이미 존재하여 이미지 압축에 활용되고 있는 기술을 신경망으로 대체한다. 셋째, 신경망을 기반으로 하는 새로운 이미지 압축 방법은 하나 이상의 신경망과 함께 연동되는 기능 블록들로 구성된 이미지 압축 코덱을 제공한다. 이러한 분류가 상당히 오래 전에 이뤄졌으며 이미지 압축에 국한된 것일 지라도, 딥 러닝을 이용하여 높은 이미지 및 비디오 압축 효율을 달성하고자 하는 최근의 많은 연구들의 분류에도 여전히 유효하다.

본 논문의 Ⅱ장에서는 압축 대상 이미지를 CNN(convolutional neural network) 기반 오토인코더(autoencoder)에 입력하여 추출한 은닉 벡터(latent vector)를 부호화하는 기술들 – 첫번째 범주[1] – 을 소개한다. Ⅲ장에서는 기존의 비디오 코덱 구조를 바꾸지 않고 일부 압축 툴을 신경망으로 대체하거나 신경망 기반의 새로운 압축 툴을 추가하려는 시도들 – 두번째 범주[1] – 에 대해 분석한다. 한편, 보다 심화된 형태라 할 수 있는 세번째 범주[1]에 해당되는 시도들은 아직 발견하지 못하였다. 마지막으로 Ⅳ장에서는 앞선 장들에서 살펴본 기술에 대한 요약과 함께 관련 분야의 발전방향과 가능성에 대해 전망해본다.

Ⅱ. 딥 러닝 기반의 이미지 압축 기술

딥 러닝 기반의 이미지 압축 기술[2]-[6]은 그림 1에 나타낸 것과 같은 합성곱 오토인코더(convolutional autoen- coder)의 일반적인 구조를 갖는다. 인코더 신경망은 입력된 압축 대상 이미지로부터 특징을 추출하여 차원이 축소된 은닉 벡터(latent vector)를 생성하고, 디코더 신경망은 다시 은닉 벡터로부터 이미지를 복원한다. 그러나, 일반적으로 은닉 벡터의 차원 축소만으로는 정보의 감소량이 크지 않아 그대로 압축에 활용할 수 없다. 따라서, 은닉 벡터의 효율적인 양자화(quantization)와 엔트로피 부호화(entropy encoding)를 위한 방법이 필요하며 이를 통해 최종 비트스트림(bitstream)을 생성하게 된다. 본 논문에서는 최신 딥 러닝 기반 이미지 압축 기술들을 비트스트림 생성 방식에 따라 반복 생성 방식과 일괄 생성 방식으로 구분하여 소개한다.

1. 비트스트림 반복 생성 방식

딥 러닝 기반 이미지 압축 기술 중 비트스트림 반복 생성

a) 한국전자통신연구원 실감AV연구그룹(Realistic AV Research Group, Electronics and Telecommunications Research Institute)
‡ Corresponding Author : 조승현(Seunghyun Cho)
E-mail: shcho@etri.re.kr
Tel: +82-42-860-6274
ORCID: https://orcid.org/0000-0003-1985-4420
※ 이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 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 Promotion(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 received April 9, 2018; Revised April 13, 2018; Accepted April 13, 2018.

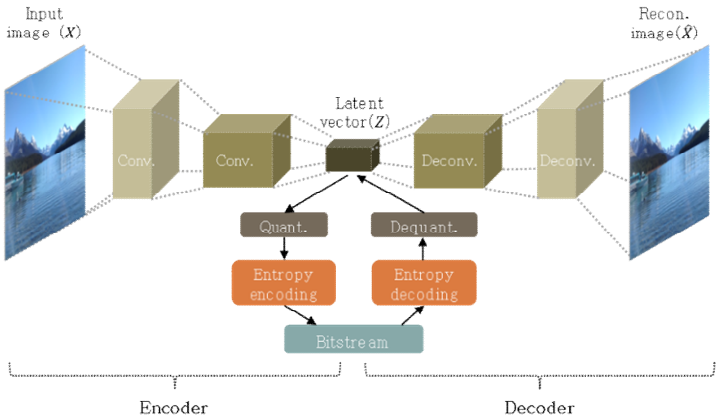

그림 1. 합성곱 오토인코더 기반 이미지 압축 개념
Fig. 1. The concept of image compression based on the convolutional autoencoder

방식[2]-[4]은 기존의 점층적 이미지 코딩(progressive coding)과 유사한 기술로, 잔차 오토인코더(residual autoencoder) 구조에 기반하고 있다. 부호화를 수행할 때 처음에는 원 영상을 입력하여 비트스트림을 생성하고, 그 이후부터는 출력과 원 영상과의 차이 신호(residual)를 입력하기를 반복하면서 비트스트림을 추가 생성한다. 이렇게 반복적으로 부호화를 수행할 때마다 매회 동일한 비트량이 누적되며 복원 영상의 화질은 점차 높아진다. 이러한 특징으로 인해, 비트스트림 반복 생성 방식은 화질에 따른 가변 비트율 코덱 설계가 가능한 장점과 함께 연산의 반복 수행으로 인해 부/복호화 복잡도가 필연적으로 증가하는 단점을 갖는다.

잔차 오토인코더의 은닉 층(hidden layer)을 완전 연결(fully-connected) 방식으로 구성하는 것 보다 합성곱(convolution)으로 구성하였을 때 우수한 성능을 보인다. 완전 연결 RNN 잔차 오토인코더 방법은 JPEG과 거의 같은 수준의 압축 성능을 보이고, 합성곱 RNN 잔차 오토인코더 방법은 JPEG보다 약 10% 이상의 비트 감축 성능을 보인다[2]. 압축 성능뿐만 아니라 복원 영상의 인지 화질도 우수한 것으로 발표되었다. 기존 압축 기술에서 나타나는 블록킹(blocking)이나 블러링(blurring)이 줄어들고, 컬러 샘플링(4:2:0)으로 인한 컬러 스미어링(smearing) 현상도 감소한다[2].

한편, CNN – 이미지 신호의 공간적 분포로부터 얻게 되는 특징 맵(feature map)만을 고려 – 기반의 오토인코더를 반복적으로 사용하여 잔차 오토인코더를 구성하는 것보다 시간상으로 앞서 수행한 결과를 뒤따르는 반복수행에서 참고할 수 있는 RNN(recurrent neural network)을 적용하면 압축 성능이 높일 수 있다[2]. RNN은 순차적으로 정보를 처리할 때 이전 입력이 다음 입력에 대한 신경망의 출력에 영향을 주도록 고안된 신경망이다. 그림 2에 RNN 잔차 오토인코더의 동작을 도식화하였다. RNN을 적용한 잔차 오토인코더는 이전의 입력 – 원본 또는 잔차 이미지 – 을

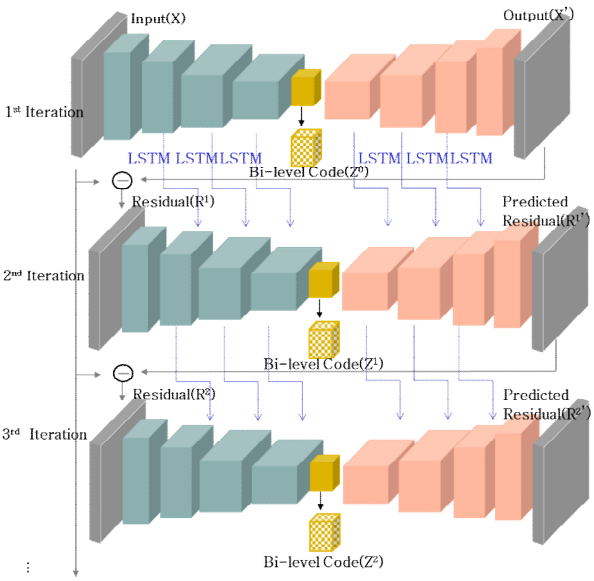

그림 2. 이미지 압축을 위한 RNN 잔차 오토인코더
Fig. 2. RNN-based residual autoencoder for image compression

처리한 결과가 반복 수행에 따른 다음 입력의 신경망 출력에 영향을 준다. RNN의 일종인 LSTM(long short-term memory)과 GRU(gated recurrent units)를 비교하였을 때, RNN 종류에 따른 성능 차이는 그리 크지 않은 것으로 보고되었다[3].

가장 최근의 연구[4]에 따르면, RNN 잔차 오토인코더 압축 성능을 추가로 개선하기 위해 세 가지 기술이 고안되었다. 첫 번째, 과거의 학습 정보가 현재의 학습에 영향을 미치는 RNN의 성능을 개선하기 위해 1회 차 오토인코더를 반복 수행하여 RNN 파라미터를 안정화하는 기술이다. 즉 1회 차의 인코더와 디코더를 동일 입력에 대해 반복 수행하여 높은 화질의 복원 영상을 얻은 후 그 이후의 잔차 오토인코더를 수행하여 MS-SSIM(multi-scale structural similarity)[10] 기준 약 0.47 AUC db 성능 향상을 얻었다. 두 번째, 학습 최적화에 적용하는 오차(loss)에 인지 화질 측정 단위를 반영하였다. 이는 성능 평가에 사용하는 지표와 유사한 지표를 최적화에 적용한 방법으로 MS-SSIM 기준 약 0.36 AUC db 성능 향상을 얻었다. 세 번째, 잔차 오토인코더 반복 횟수를 입력 영상에 따라 적응적으로 달리하여 비트량을 최적화하였다. 세 가지 기술을 추가한 합성곱 RNN 잔차 오토인코더를 이용한 이미지 압축 기술[3]은 MS-SSIM 화질 기준 JPEG 대비 약 40%, BPG 420 대비 약 10% 비트량 감축 성능을 얻었다.

2. 비트스트림 일괄 생성 방식

딥 러닝 기반의 이미지 압축 방법들 중 비트스트림 일괄 생성 방식[5][6]은 이미지 부호화의 결과로 출력되는 은닉 벡터를 표현하는 데에 필요한 비트량과 복호화의 결과로 출력되는 복원 영상의 왜곡을 동시에 최적화하는 방향으로 네트워크를 학습시키는 것이다. 본 절에서 소개할 두 기술[5][6]은 CNN 기반 오토인코더의 형태를 갖는 것들로써, 기본적인 구조는 서로 유사하지만 비트량의 조절을 위해 적용되는 양자화 방법 및 손실 함수(loss function)의 정의 등에서 구별된다.

첫번째 기술[5]은 오토인코더를 이용한 이미지의 손실 압축 방법을 제안하며 전체적인 구조는 그림 3과 같이 CNN 기반 오토인코더의 형태를 갖는다. 다만 부/복호화 각각의 은닉층 중간에 두 개의 잔차신호 연결(residual connection)이 존재하며, 부호화 결과로 출력된 은닉 벡터에 대해 반올림 연산을 적용하는 양자화 단계를 거친다. 따라서 부/복호화 단계에서는 잔차신호에 대한 훈련을 통하여 은닉 벡터를 생성하며, 양자화 단계를 거친 은닉 벡터를 입력으로 하여 복호화 과정을 거쳐 복원 이미지를 얻을 수 있다. 이는 입력 이미지 x, 부호화 함수 f, 대괄호로 표현되는 양자화 연산, 복호화 함수 g를 이용하여 다음의 식 (1)을 최적화하는 방식으로 네트워크를 훈련할 수 있다.

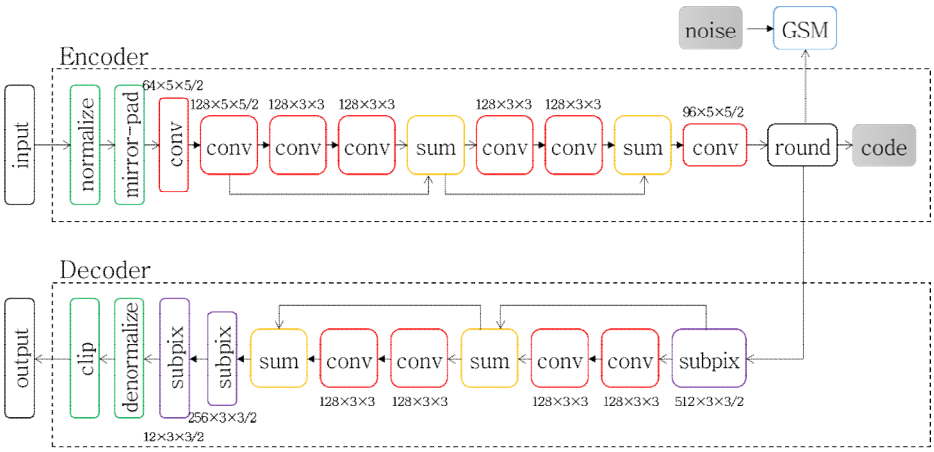

그림 3. 오토인코더 기반의 이미지 손실 압축 방법의 블록다이어그램 ('C×K×K/N"의 형식, C는 합성곱 필터 수, K×K는 필터의 크기, N은 필터 간격 또는 업샘플링 팩터)

Fig. 3. block-diagram of lossy image compression with compressive autoencoders (notation 'C×K×K/N" with C, the number of K×K sized filters and N, the stride or up-sampling factor)

$$-\log_2 Q([f(x)]) + \beta \cdot d(x, g([f(x)])) \quad (1)$$

식 (1)의 log 부는 입력 이미지 x에 대한 부호화 결과인 은닉 벡터를 양자화한 후, 이를 표현하는 데 필요한 비트량을 의미하며, d는 입력 이미지와 복호화된 이미지 간의 왜곡 (distortion)을 의미한다. β는 비트량과 왜곡 간의 트레이드오프(tradeoff)를 조절하는 값이다. 제안된 방법을 이용하여 네트워크를 훈련시키기 위해서는 미분 가능한 손실함수를 사용하여 역전파법(back propagation)을 사용한다. 그러나, 반올림 함수로 정의되는 양자화 함수는 이산 영역에서 표현되는 미분 불가능한 함수이므로 정수로 표현되는 각 양자화 심볼 값에 균등 가산 잡음(uniform additive noise)을 그림 3과 같이 더해주고, 이를 근사화하여 미분 가능한 형태로 바꾸어 사용한다. 이러한 훈련 과정 중 입력 이미지의 부호화, 은닉 벡터의 양자화, 복호화 과정으로 구성되는 전방향 패스(forward pass)에는 미분 불가능한 반올림 함수를 이용하여 양자화를 수행하고, 네트워크의 파라미터들을 업데이트하는 역방향 패스(backward pass)에는 미분 가능한 형태로 근사화된 양자화 함수를 사용한다. 이러한 과정에서 은닉 벡터 계수들의 분포를 Gaussian scale mixture(GSM)를 사용하여 모델링하고, 그 확률 모델을 기반으로 산술 부호화(arithmetic coding)를 적용했을 때 발생될 부호화 비트율(bitrate)을 추정한다.

두번째 기술인 실시간 적응적 이미지 압축[6]은 구조적인 측면에서 그림 4와 같이 오토인코더의 형태를 갖는다. 그림 4에서 특징 추출(feature extraction)과 특징 합성(syn- thesis from features)으로 표현되는 부/복호화 네트워크는 그림 3과 달리 내부적으로 그림 5처럼 입력 이미지를 다양한 공간 스케일로 나누어 각각의 스케일 및 서로 다른 스케일 간의 정보를 공유하도록 디자인 되었다. 그림 4에서 은닉 벡터는 양자화 과정을 거쳐 비트 평면 공간에서 적응적 산술 부호화되어 비트스트림을 생성한다. 이에 추가적으로 적응적 산술 부호화의 성능을 향상시키기 위하여 양자화된 특징 추출 결과의 각 요소별 값의 크기와 현재 위치를 기준으로 공간적으로 이웃한 주변 값 간의 차를 이용하여 발생될 비트 열의 길이를 정규화(그림4의 adaptive codelength regularization)한다. 그림 4에 판별자 손실(discriminator loss)로 나타낸 것처럼, 실시간 적응적 이미지 압축의 또 다른 특징은 복호화된 이미지가 입력 이미지와 같이 현실적인 특징을 갖도록 네트워크를 훈련시키기 위해 GAN(generative adversarial network)[9]의 일부 요소인 판별자 네트워크를 차용했다는 것이다. 이는 제안한 부/복호화 네트워크의 출력과 입력 이미지 간의 부호화 왜곡 측면뿐만 아니라, 현실감의 측면 또한 고려하는 것으로 이해할 수 있다. 실시간 적응적 이미지 압축은JPEG과 JPEG 2000에 비해서는 약 2.5배, WebP[7]에 비해서는 약 2배, 그리고 HEVC 화면 내 부호화 기술에 기반한 BPG[8]에 비해서는 약 1.7배의 압축률 향상을 얻었다고 발표되었다.

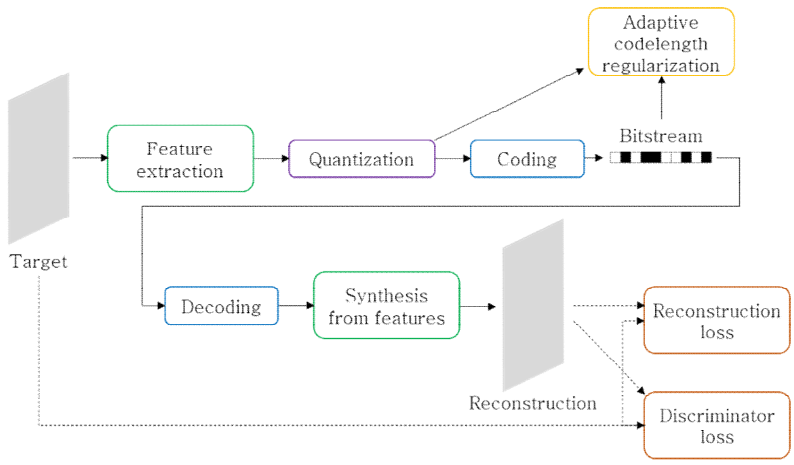

그림 4. 오토인코더 기반의 실시간 적응적 이미지 압축 방법의 블록 다이어그램
Fig. 4. A block-diagram of real-time adaptive image compression

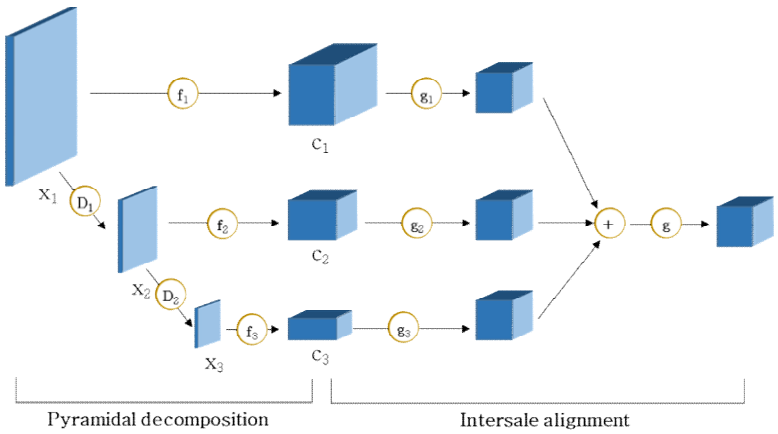

그림 5. 3-계층 스케일 공간으로의 입력 이미지 특징 추출 파이프라인 구조
Fig. 5. A pipeline structure of feature extraction from 3-scale image space

앞서 소개한 비트스트림 반복 생성 방식[2]~[4]이 심층 신경망의 반복 구동으로 인한 부/복호화 복잡도 증가를 수반하는 반면, 비트스트림 일괄 생성 방식[5][6]은 한번의 심층 신경망 구동으로 부/복호화가 완료되므로 상대적으로 복잡도가 낮다고 할 수 있다. 특히, 실시간 적응적 이미지 압축[6]의 경우, GPU에서 수행할 경우 부/복호화 시간이 각각 8.6, 9.9ms 수준인 것으로 보고되었다.

Ⅲ. 딥 러닝을 적용한 비디오 압축 툴 개선 기술

이번 장에서는 딥 러닝 기술을 적용하여 기존 비디오 코덱 압축 툴의 성능 개선을 시도한 연구들을 살펴본다. 이러한 접근 방법들은 비디오 코덱의 구조를 유지한 채, 기존의 특정 압축 툴을 대체하거나 기존 모드에 새로운 모드를 추가하기 위해 학습된 심층 신경망을 사용한다.

1. 인루프(in-loop) 및 후처리(post processing) 필터

인루프 및 후처리 필터는 딥 러닝의 적용이 가장 활발한 분야로, 기존 비디오 코덱의 복원 화질을 향상시켜 압축 효율 향상의 효과를 꾀한다[15]~[18]. 이 분야의 연구가 효과를 거두고 있는 이유는 이미지 초 해상도, 왜곡 제거(artifact reduction, 예: deblur, deblock 등), 잡음 제거(noise reduction) 측면에서 탁월한 성능을 보이는 딥 러닝 기반의 이미지 화질 개선 기술들[19]~[22] 에서 찾을 수 있다. 특히, 신경망을 통한 JPEG 이미지 부호화 왜곡(compression or coding artifact) 제거[19]~[21]에는 주목할만한 기법들이 사용되었다. 첫째, 부호화 왜곡을 포함한 영상을 신경망에 입력하고 원본에 가까운 영상을 출력하도록 신경망을 학습시킨다[19]~[21]. 둘째, 잔여 학습(residual learning)을 통해 신경망이 입력 영상과 원본 영상의 차이를 생성하도록 학습 시킨다[20][21]. 셋째, 배치 정규화(batch normalization)을 선택적으로 사용하여 학습 효율을 높인다[21]. 이러한 일반적 구조를 바탕으로, 딥 러닝 기반의 비디오 인루프 및 후처리 필터 분야에는 비디오 코덱의 특징을 활용하여 필터 성능을 더욱 높이려는 기법들이 발표되고 있다.

학습된 심층 신경망으로 HEVC(high efficiency video coding)[11]의 인루프 필터, 즉 디블록(deblock)[13]과 SAO (sample adaptive offset)[14]를 대체한 연구들[15]~[17]의 특징을 살펴본다. 이 연구들은 HEVC 인루프 필터 대신 사용되지만 화면 내 부호화(intra-frame coding)에만 적용하였기 때문에 사실상 후처리 필터로 간주 될 수 있다. 먼저, VRCNN(variable-filter-size residual learning convolutional neural network)[15]은 HEVC의 다양한 크기의 변환(transform) 및 양자화에 의한 왜곡을 효과적으로 제거하기 위해 서로 다른 크기의 CNN 커널(kernel)을 사용하였다. 또한, VRCNN에서는 신경망의 출력에 입력 영상을 더하여 복원 영상을 얻고, 복원 영상과 원본 영상 간의 L2 손실(loss)을

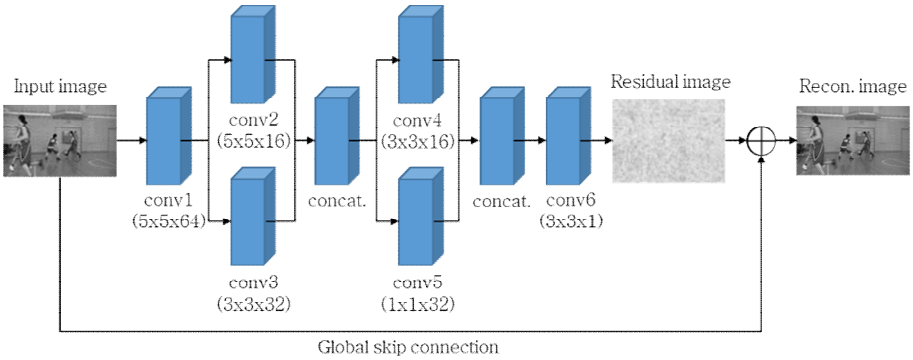

그림 6. VRCNN[15]의 구조와 입출력
Fig. 6. The structure of VRCNN[15] with its input and output

최소화하도록 신경망을 학습 즉, 잔여 학습시키며 배치 정규화는 사용하지 않았다. 그림 6에 VRCNN의 구조와 입출력을 도식화 하였다.

한편, 보다 최근에 발표된 MMS-net(multi-modal/multi-scale convolutional neural network)[16]은 CNN 커널 크기를 변화시키는 대신 입력 영상을 서로 다른 스케일에서 처리하는 네트워크, 예를 들면 original-scale과 half-scale network를 함께 사용한다. MMS-net의 또다른 특징이자 가장 두드러진 특징은 왜곡을 포함한 영상과 더불어 HEVC CU(coding unit) 및 TU(transform unit)의 분할 정보를 네트워크에 입력한다는 점이다. 이는 HEVC의 부호화 왜곡이 주로 CU 및 TU의 경계 영역에서 발생한다는 사실을 이용하여 네트워크가 해당 영역을 효과적으로 처리 할 수 있도록 위치를 알려주는 것이다. MMS-net은 VRCNN의 잔여 학습을 위해 사용된 네트워크 연결 – global skip connection – 뿐만 아니라 신경망 내부에 지역적인 skip connection을 갖는 잔여 블록(residual block)을 사용하여 네트워크를 구성하였으며 잔여 블록 내에서 convolution 연산에 앞서 배치 정규화를 수행하였다.

앞선 연구들과는 달리, DCAD(deep CNN-based auto decoder)[17]은 HEVC의 인루프 필터를 거친 복원 영상을 후처리 하여 화질을 개선한다. DCAD는 잔여 학습을 위해 VRCNN과 같은 방식으로 global skip connection을 구성하였으며 배치 정규화는 사용하지 않았다. DCAD는 3x3 convolution 커널을 사용하였으며 입출력 층을 제외하고 총 10층의 깊이로 디자인 되었다. DCAD는 각 QP별 신경망을 학습시키기 위해, 높은 QP에서 학습한 네트워크 파라미터를 낮은 QP를 위한 네트워크의 파라미터 초기값으로 사용하는 방식의 전이 학습(transfer learning)을 사용하였다. DCAD 뿐만 아니라, VRCNN, MMS-net도 마찬가지로 신경망 학습에 휘도(luminance) 신호만을 사용하였으며 색차(chrominance) 신호의 처리에는 동일한 네트워크 파라미터를 사용하였다.

앞서 살펴본 VRCNN, MMS-net, DCAD의 성능 비교를

표 1. 신경망 기반 인루프 및 후처리 필터의 BD-rate(Y) 결과(%) 비교
(AR-CNN[19]과 VDSR[22]의 성능은 참조논문[15]에서 발췌하였음)
Table 1. Comparisons of BD-rate (Y) results (%) for neural network-based in-loop and post filters

	AI					LP	LB	RA
	AR-CNN[19]	VDSR[22]	VRCNN[15]	MMS-net[16] (M=15)	DCAD[17]	DCAD[17]	DCAD[17]	DCAD[17]
Class A	0.9	-2.8	-3.5	-	-	-	-	-
Class B	1.0	-2.7	-3.3	-	-3.4	-5.3	-3.9	-4.6
Class C	-0.6	-4.1	-5.0	-9.3	-4.6	-4.8	-4.1	-4.3
Class D	-0.8	-4.4	-5.4	-7.7	-5.2	-5.2	-4.4	-4.4
Class E	0.4	-5.7	-6.5	-	-7.8	-11.8	-10.4	-10.1
Overall	0.2	-3.8	-4.6	-8.5	-5.0	-6.4	-5.3	-5.5

위해, 표 1에 HEVC와 비교한 BD-rate(Y) 결과를 나타내었으며, 기존의 CNN 기반 부호화 왜곡 제거기술 (AR-CNN[19])과 초해상도 기술(VDSR[22])과도 함께 비교하였다. HEVC AI(all intra) 조건에서, VRCNN은 전체 실험 영상에 대해 평균 4.6%, MMS-net은 class C, D 시퀀스에 대해 평균 8.5%, DCAD는 class A를 제외한 시퀀스에 대해 평균 5.0%의 BD-rate(Y) 이득이 있었다. 유일하게 후처리 – HEVC 인루프 필터의 대체가 아닌 – 기술로 명시한 DCAD는 HEVC LP(low-delay P), LB(low-delay B), RA (random access) 조건에서 class A를 제외한 시퀀스에 대해 각각 평균 6.4%, 5.3%, 5.5%의 BD-rate(Y) 이득을 보고하였다. 앞서 언급했듯이, VRCNN, MMS-net이 HEVC의 디블록 및 SAO 필터를 대신 사용되었지만 AI 조건에서만 활용되었기에 표 1에 나타낸 결과만으로는 진정한 의미의 인루프 필터 – 코딩 루프 안에 존재하며 결과 영상이 참조되는 – 로써의 효용성이 입증되었다고 보기 어렵다.

가장 최근에는, ITU-T VCEG(Q6/16)과 ISO/IEC MPEG (JTC 1/SC 29/WG 11)이 공동 구성한 JVET(joint video exploration team)에 의해 표준화가 시작되고 있는 FVC(future video coding)에 심층 신경망 기반 필터 기술 CNNF (convolutional neural network filter)[18]가 제안되었다. 그림 7은 FVC의 참조 소프트웨어 모델인 JEM[23]에 CNNF를 적용하였을 때 화면 내 복호화 구조가 어떻게 달라지는 지를 비교하여 보여준다. 그림 7에서 알 수 있듯이, CNNF는 FVC의 화면 내 부/복호화 과정에서 bilateral filter(BF), deblocking filter(DF), SAO를 대체한다. CNNF는 복잡도를 줄이기 위해 7개의 합성곱 층을 포함한 총 10층으로 구성되었으며, 앞서 살펴본 VRCNN, MMS-net과 동일한 방식의 잔여 학습을 사용하고 배치 정규화는 사용하지 않는다. 또한, CNNF는 CPU, GPU 등과 같은 이종의 디바이스에 대한 범용성을 고려하여, 부동 소수점(floating point) 연산을 사용하는 대신 신경망 내 모든 층의 출력은 16, 가중치(weight)와 바이어스(bias)는 각각 8, 32의 비트 폭(bit width)을 사용하여 고정 소수점(fixed point) 연산을 구현하였다. CNNF는 HEVC 적용 기술들과는 달리 luma와 chroma 각각에 대한 학습 데이터를 사용하여 신경망을 학습시켰다. CNNF는 AI 조건에서 전체 실험 영상에 대해 평균 3.57%(Y), 6.17%(U), 7.06%(V)의 BD-rate 이득이 있는 것으로 발표되었으며, 일부 실험 영상만 평가된 RA, LDB, LDP조건에서는 BD-rate 이득이 크게 감소하는 것으로 발표되었다. 한편, AI 조건에서 CNNF를 CPU+GPU 환경에 구동할 경우, 복호화 시간을 기준으로 한 복잡도 증가는 약 1.65배인 반면, CPU 환경에 구동하면 복잡도가 약 128.87배로 크게 증가한다고 보고되었다.

2. 화면 내 예측 및 부호화

화면 내 예측은 비디오 신호의 공간적 중복성을 이용하기 위해 복원된 주변 샘플로부터 부호화 대상 블록을 예측하며 시간적 중복성을 이용하는 화면 간 예측에 비해 상대적으로 예측 정확도가 높지 않다. IPFCN(fully connected network for intra prediction)[24]과 IPCNN(intra prediction

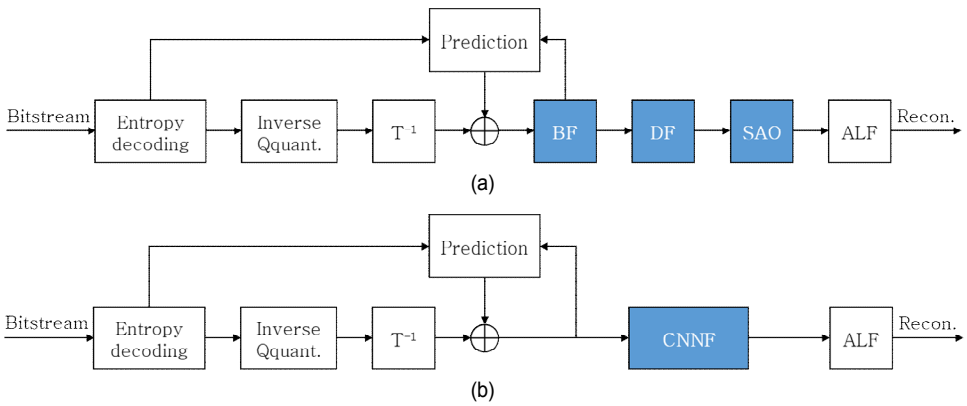

그림 7. (a) JEM과 (b) CNNF[18]의 화면 내 복호화 구조 비교
Fig. 7. Comparison of intra decoding scheme between (a) JEM and (b) CNNF[18]

CNN)[25]은 심층 신경망을 통해 화면 내 예측의 정확도를 높이는 시도를 하였다. 그림 8에 IPFCN의 구조와 입출력을 보였다. 그림 8에 나타낸 것처럼, IPFCN은 네트워크의 각 층의 노드가 이웃 층의 모든 노드에 완전히 연결된 형태로 구성되었다. IPFCN은 8x8 블록의 화면 내 예측을 위한 참조 블록을 입력 받아 예측 블록을 생성하도록 학습된다. 한편, IPCNN은 네트워크의 각 층이 합성곱에 기반하여 연결된 일반적인 CNN 구조 이며, 32x32, 16x16, 8x8 블록 크기를 지원한다. 또한 IPCNN은 판별자 네트워크를 추가하여 예측 블록의 영상 품질을 높이고자 하였다. IPFCN과 IPCNN은 각각 HEVC AI 조건에서 전체 실험 영상(class A ~ E)에 대해 각각 0.92%(Y), 1.33%(U), 1.42%(V)와 1.02%(Y), 0.65%(U), 0.56%(V)의 BD-rate 이득을 보였다.

그림 9에 구조를 나타낸 CNN 기반 블록 다운/업 샘플링 부호화[26]는 앞서 살펴본 IPFCN, IPCNN과 달리, HEVC의 화면 내 부호화 구조[12] 변경을 통해 심층 신경망을 이용한 압축 효율 향상을 시도하였다. 제안된 기술은 그림 9처럼 입력 영상에 대한 다운/업 샘플링 기반 화면 내 부호화 – 다운 샘플링하여 부호화하고 복호화 과정에서 복원 영상을 업샘플링 – 수행 여부를 CTU(coding tree unit) 단위로 결정한다. 입력 영상의 다운 샘플링에는 기존의 초 해상도 기술[27]을 개선한 CNNIF(CNN based interpolation filter)[26]와 HEVC 화면 간 예측에 사용되는 DCTIF(discrete cosine transform interpolation filter)[11]가 선택적으로 사용되었다. 영상의 다운/업 샘플링이 CTU 단위로 수행되기 때문에, CNNIF의 zero padding과 DCTIF의 참조 샘플 부족으로 인

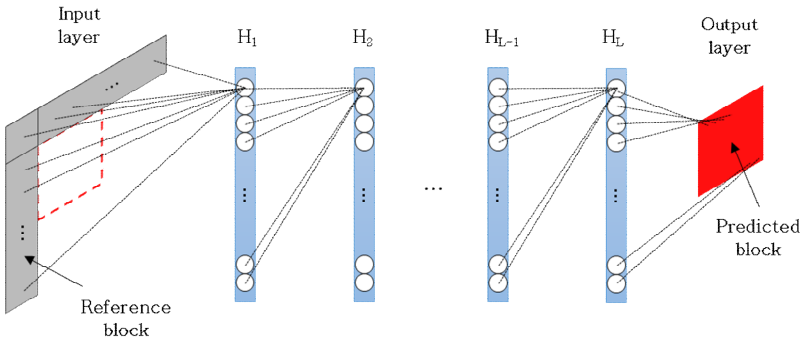

그림 8. 완전 연결 방식에 기반한 IPFCN[24] 구조와 입출력
Fig. 8. The structure and its in/output of IPFCN[24] based on fully connected layers

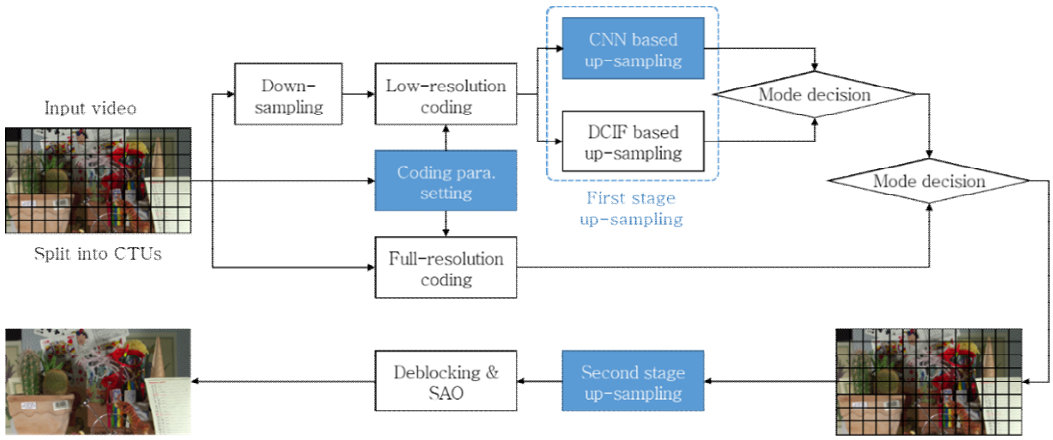

그림 9. CNN 기반 블록 다운/업 샘플링 부호화[26] 구조
Fig 9. CNN-based block down/up-sampling coding[26] scheme

해 복원 영상의 CTU 경계에 인접한 샘플에서 화질 열화가 발생하며, 이를 보완하기 위해 주변 CTU 상의 샘플을 활용한 추가적인 업 샘플링(그림 9의 second stage up-sampling)을 수행한다. 발표에 따르면, 제안된 방식을 적용하였을 때 HEVC AI 조건에서 전체 실험 영상에 대해 평균 5.5%(Y), 6.0%(U), 2.2%(V), UHD 영상에 대해서는 이보다 큰 9.0%(Y), 1.6%(U), 3.2%(V)의 BD-rate 이득이 측정되었다.

Ⅳ. 결론 및 전망

본 논문에서는 딥 러닝 기반의 이미지와 비디오 압축 최신 기술을 소개하고 기술 간의 공통점과 차이점을 설명하였다. 먼저, 딥 러닝 기반의 이미지 압축 기술은 합성곱 오토인코더의 은닉 벡터를 부호화하며, 비트스트림을 반복적으로 생성하는 경우 RNN 잔차 오토인코더가 활용되었다. 심층 신경망의 학습은 복원 이미지의 화질은 높이고 발생되는 비트량은 낮추는 방향으로 이뤄지는데, 화질 측정에는 주로 MS-SSIM을 사용하고 PSNR(peak signal-to-noise ratio)을 보조적으로 사용하는 경우가 많았다. 이런 새로운 방식들을 기존의 JPEG, WebP 등과 비교하면 저 비트율에서의 복원 이미지 화질이 크게 개선되는 장점이 있다. 이미지 압축 성능을 더욱 개선 시키기 위해서는 은닉 벡터 분포 모델링의 정확도 개선, 또는 입력 영상의 특징에 따른 적응적 은닉 벡터 분포 모델링의 연구가 필요할 것으로 관측된다.

다음으로, 딥 러닝 기반의 비디오 압축 기술은 기존 코덱 압축 툴을 대체하거나 수정하여 압축 효율을 개선한다. 인루프 및 후처리 필터 심층 신경망을 HEVC 또는 FVC에 적용하면 복원 영상의 화질 개선으로 상당한 BD-rate 이득이 있었다. 현재까지의 연구는 대부분 결과 영상이 참조되지 않는 형태였으며, 향후 참조 영상의 화질 개선을 위한 연구가 필요할 것으로 생각된다. 화면 내 예측 심층 신경망은 새로운 예측 모드로 추가하여 사용할 수 있으며 예측 정확도 개선으로 비교적 소폭의 BD-rate 이득을 얻을 수 있다. 심층 신경망 및 HEVC DCTIF 기반 다운/업샘플링을 활용한 화면 내 부호화는 초 해상도 기술을 선택적으로 사용하여 BD-rate 이득을 달성했다.

이 밖에도, 딥 러닝 기반 이미지와 비디오 압축 기술의 발전을 위해서는 인지 화질 측정 및 예측 기술을 위한 연구가 필요하다. 인지 화질 측정과 예측을 위해 계산된 특징 또는 이를 구현한 신경망은 이미지 또는 비디오 압축을 위한 신경망의 학습에 활용될 수 있다. 한편, 본 논문에서는 심층 신경망 사용에 따른 연산 복잡도 증가에 대해서는 논의하지 않았다. 그러나, 기존의 이미지 및 비디오 압축 표준의 발전 과정에서 기술의 제안과 평가에 있어 중요한 기준이 되고 있는 만큼, 심층 신경망 기술 활용을 위해서는 연산 복잡도를 낮추기 위한 노력이 병행되어야 할 것이다.

참 고 문 헌 (References)

[1] J. Jiang, “Image compression with neural networks,” Signal Processing: Image Communication Vol.14, No.9, pp.737-760, July 1999.

[2] G. Toderici, S. M. O’Malley, S. J. Hwang, D. Vincent, D. Minnen, S. Baluja, M. Covell, and R. Sukthankar, “Variable rate image compression with recurrent neural networks,” Proceeding of International Conference on Learning Representations, San Juan, Puerto Rico, May 2016.

[3] G. Toderici, D. Vincent, N. Johnston, S. J. Hwang, D. Minnen, J. Shor, and M. Covell, “Full Resolution Image Compression with Recurrent Neural Networks,” Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, pp.5435-5443, July 2017.

[4] N. Johnston, D. Vincent, D. Minnen, M. Covell, S. Singh, T. Chinen, S. J. Hwang, J. Shor, and G. Toderici, “Improved Lossy Image Compression with Priming and Spatially Adaptive Bit Rates for Recurrent Networks,” https://arxiv.org/abs/1703.10114 (Submitted on Mar 29, 2017)

[5] L. Theis, W. Shi, A. Cunningham, and F. Huszar, “Lossy Image Compression with Compressive Autoencoders,” Proceeding of International Conference on Learning Representations, Toulon, France, April 2017.

[6] O. Rippel and L. Bourdev, “Real-Time Adaptive Image Compression,” Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, PMLR 70:2922-2930, Aug. 2017.

[7] WebP - A new image format for the Web, https://developers.google.com/speed/webp/BPG image format, http://bellard.org/bpg

[8] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets”, Proceeding of Neural Information Processing Systems, Montreal, Canada, Dec. 2014.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Transactions on Image Processing, Vol.13, No.4, pp.600 – 612, April

2004.
[10] ITU-T and ISO/IEC JTC 1, “High Efficiency video coding,” ITU-T Recommendation H.265 and ISO/IEC 23008-2 (MPEG-H Part 2), Third edition: April 2015.
[11] J. Lainema, F. Bossen, W.-J. Han, J. Min, and K. Ugur, “Intra coding of the HEVC standard,” IEEE Trans. on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1792-1801, 2012.
[12] A. Norkin, G. Bjøntegaard, A. Fuldseth, M. Narroschke, M. Ikeda, K. Andersson, M. Zhou, and G. V. der Auwera, “HEVC Deblocking Filter,” IEEE Trans. on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1746-1754, 2012.
[13] C.-M. Fu, E. Alshina, A. Alshin, Y.-W. Huang, C.-Y. Chen, and C.-Y. Tsai, C.-W. Hsu, S.-M. Lei, J.-H. Park, and W.-J. Han, “Sample Adaptive Offset in the HEVC Standard,” IEEE Trans. on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1755 - 1764, 2012.
[14] Y. Dai, D. Liu, and F. Wu, “A Convolutional Neural Network Approach for Post-Processing in HEVC Intra Coding,” Proceeding of the 23rd International Conference on Multimedia Modeling, Reykjavik, Iceland, pp.28-39, Jan. 2017.
[15] J. Kang, S. Kim, and K. M. Lee, “Multi-modal Multi-scale Convolutional Neural Network based In-loop Filter Design for Next Generation Video Codec,” Proceeding of IEEE International Conference on Image Processing, Beijing, China, pp.16-30, Sept. 2017.
[16] T. Wang, M. Chen, and H. Chao, “A Novel Deep Learning-Based Method of Improving Coding Efficiency from the Decoder-end for HEVC,” Proceeding of Data Compression Conference, Snowbird, USA pp.410-419, April 2017.
[17] L. Zhou, X. Song, J. Yao, L. Wang, and F. Chen, “Convolution Neural Network Filter (CNNF) for Intra Frame,” JVET-I0022, Joint Video Exploration Team of ISO/IEC and ITU-T, Gwangju, Korea, Jan. 2018.
[18] C. Dong, Y. Deng, C. C. Loy, and X. Tan, “Compression Artifacts Reduction by a Deep Convolutional Network,” Proceeding of IEEE International Conference on Computer Vision, Santiago, Chile, pp.576-584, Dec. 2015.
[19] P. Svoboda, M.Hradis, D.Barina, and P.Zemcik, “Compression Artifacts Removal Using Convolutional Neural Networks,” Journal of WSCG, Vol.24, No.2, pp.63-72, 2016.
[20] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising,” IEEE Transactions on Image Processing, Vol.26, No.7, pp.3142-3155, 2017.
[21] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, pp.1646-1654, June 2016.
[22] JEM7.0, https://jvet.hhi.fraunhofer.de/svn/svn_HMJEMSoftware/branches/HM-16.6-JEM-7.0-dev/.
[23] J. Li, B. Li, J. Xu, and R. Xiong, “Intra Prediction Using Fully Connected Network for Video Coding,” Proceeding of IEEE International Conference on Image Processing, Beijing, China, pp.1-5, Sept. 2017.
[24] S. Cho, J. Lee, W. Lim, Y. Kim, J. Seok, H. Y. Kim, and J. Choi, “HEVC Intra Prediction through Convolutional Neural Network,” 30th Workshop on Image Processing and Image Understanding, Jeju, Korea, Feb. 2018.
[25] Y. Li, D. Liu, H. Li, L. Li, F. Wu, H. Zhang, and H. Yang, “Convolutional Neural Network-Based Block Up-sampling for Intra Frame Coding,” IEEE Transactions on Circuits and Systems for Video Technology, (Early Access), July 2017.
[26] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in European Conference on Computer Vision, pp.184 – 199, Springer, 2014.

저 자 소 개

조 승 현

- 2003년 8월 : 경북대학교 전자전기공학부 학사
- 2006년 2월 : 한국과학기술원 전기및전자공학과 석사
- 2015년 8월 : 한국과학기술원 전기및전자공학과 박사
- 2006년 6월 ~ 현재 : 한국전자통신연구원 실감AV연구그룹 책임연구원
- ORCID : https://orcid.org/0000-0003-1985-4420
- 주관심분야 : 영상처리/압축, 컴퓨터 비전, 기계 학습, 디지털 회로설계

저 자 소 개

김 연 희

- 2000년 : 아주대학교 정보및컴퓨터공학과 학사
- 2002년 : 아주대학교 정보및컴퓨터공학과 석사
- 2009년 : George Mason University Computer Science 박사
- 2009년 ~ 현재 : 한국전자통신연구원 실감AV연구그룹 선임연구원
- ORCID : http://orcid.org/0000-0003-0658-6762
- 주관심분야 : 영상처리/압축, 컴퓨터 비전, 정보 은닉, 실감미디어 서비스

임 웅

- 2008년 : 광운대학교 컴퓨터공학과 학사
- 2008년 3월 ~ 2010년 2월 : 광운대학교 컴퓨터공학과 석사
- 2010년 3월 ~ 2016년 : 광운대학교 컴퓨터공학과 박사
- 2012년 2월 ~ 2013년 : Simon Fraser Univ. 방문연구원
- 2016년 4월 ~ 현재 : 한국전자통신연구원 실감AV연구그룹 연구원
- ORCID : https://orcid.org/0000-0002-1772-0683
- 주관심분야 : 영상처리/압축, 기계 학습, 컴퓨터 비전

김 휘 용

- 2004년 2월 : 한국과학기술원 전기및전자공학과 박사
- 2003년 8월 ~ 2005년 10월 : (주)애드팍테크놀러지 기술연구소 멀티미디어 팀장
- 2006년 9월 ~ 2010년 8월 : 과학기술연합대학원대학교 겸임교수
- 2013년 9월 ~ 2014년 8월 : Univ. of Southern California 방문연구원
- 2005년 11월 ~ 현재 : 한국전자통신연구원 미디어연구본부 실감AV연구그룹장
- https://orcid.org/0000-0001-7308-133X
- 주관심분야 : 영상처리/압축, 컴퓨터 비전, UHD/HDR/3D/VR, MPEG

최 진 수

- 1990년 2월 : 경북대학교 전자공학과 공학사
- 1992년 2월 : 경북대학교 전자공학과 공학석사
- 1996년 2월 : 경북대학교 전자공학과 공학박사
- 1996년 5월 ~ 현재 : 한국전자통신연구원 책임연구원
- ORCID : https://orcid.org/0000-0003-4297-5327
- 주관심분야 : 영상처리/압축, UHDTV 방송, 3DTV 방송