

JPEG AI의 부호화 프레임워크들의 분석 및 활용 사례에 대한 소개

□ 한승진, 김영섭 / 단국대학교

요약

이미지 압축은 이미지 및 영상처리에서 주요한 역할을 하며, 자율주행, 클라우드, 영상 송출 등의 분야에서 빅데이터를 처리해야 하는 수요가 늘어남에 따라 지속적인 연구가 진행 중이다. 그 중심에는 딥러닝(deep learning)의 발전이 자리잡고 있으며, 심층 신경망(deep neural network)을 효과적으로 학습하는 알고리즘들을 적용한 논문들은 기존 압축 포맷인 JPEG, JPEG 2000, MPEG 등의 압축 성능을 뛰어넘는 결과를 보여주고 있다. 이에 따라 JPEG AI는 딥러닝 기반 학습 이미지 압축의 표준을 제정하는 일을 진행 중이다. 본 기고에서는 JPEG AI가 표준화하고자 하는 기술과 JPEG AI에 제안한 압축 프레임워크들을 분석하고, 활용 사례들을 소개하여 JPEG AI 기반 학습 이미지 압축 모델의 동향에 대해 알아보려고 한다.

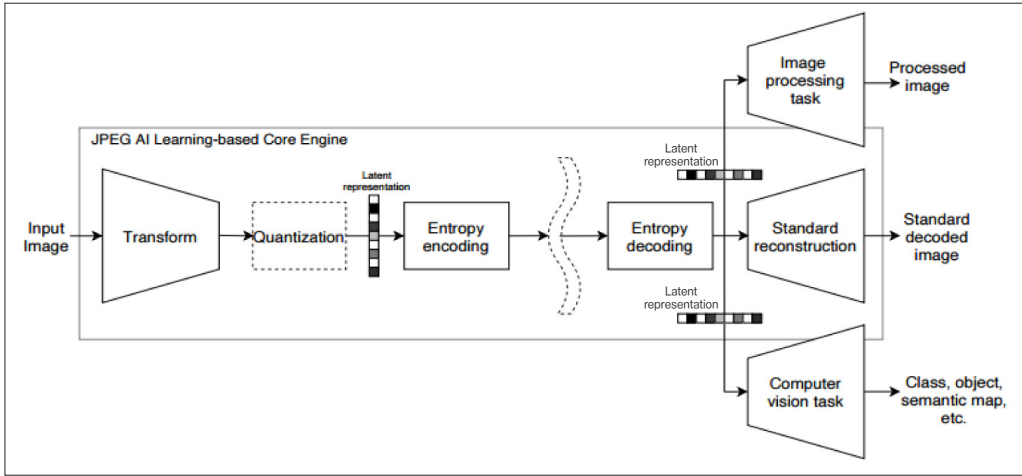
I. 서론

전 세계적으로 생성되는 데이터의 총량이 해가 갈수록

급격하게 늘어나고 있다. 2025년까지 전 세계 데이터 생산 총량은 약 163ZB까지 늘어난다는 전망이다[1]. 이에 따라 광대한 데이터를 어떻게 처리할 것인가?에 대한 답이 요구되고 있다. 따라서 이미지의 데이터 사이즈를 압축하여 저장 및 처리하는 이미지 압축 기술은 매우 필수적이라고 볼 수 있다. JPEG AI는 딥러닝 기반 학습 이미지 압축을 통해, 이미지 데이터 크기를 줄이고 동시에 처리속도 또한 향상시키는 표준 모듈들을 구축 중에 있다.

JPEG AI는 딥러닝 기반 학습 이미지 압축을 통해 크게 세 가지 과제들을 해결하고자 한다. 이미지 부호화, 이미지 처리, 컴퓨터 비전 과제가 그것들이다. <그림 1>은 러닝베이스 JPEG AI 프레임워크를 보여준다. 이 프레임워크에서 세 가지 과제들 중 이미지 처리와 컴퓨터 비전 과제는 딥러닝 기반 학습 이미지 압축기술을 이용하면 기존 코덱에서는 압축된 이미지를 따로 복호화 하

※ 이 논문은 연구재단 중견연구자 과제 NRF-2020R1A2C200917의 지원으로 제출합니다.



<그림 1> JPEG AI 딥러닝 기반 이미지 압축 프레임워크

는 과정을 거쳐야 하는 것과 달리, 부호화 이후 압축된 (잠재 벡터의) 상태로 적용할 수 있다는 장점이 있다. JPEG AI는 이미지 처리, 이미지 부호화, 컴퓨터 비전 세 가지 과제를 표준화하고자 한다. 평가 테스트셋으로는 21개의 이미지를 사용하였다.

본 기고에서는 JPEG AI가 표준화하고자 하는 기술들과 활용 사례들을[2] 소개하고, 각 기업 중에 특히 화웨이, 바이트댄스가 제안한 압축 프레임워크를 분석하여 딥러닝 기반 학습 이미지 압축 모델의 동향에 대해 알아보하고자 한다.

II. JPEG AI 과제 개요

JPEG AI 과제의 목적은 학습 가능한 차세대 이미지 압축 코덱(Learning based image compression codec)을 개발하는 것이다. JPEG AI 코덱은 일반적인 이미지에 대해 기존 코덱보다 향상된 압축률과 비트 스트림에서의 효율적인 이미지 처리를 지원해야 한다. <그림 1>에 제시하였듯, 과제는 이미지 처리, 이미지 부호화, 컴

퓨터 비전 세 가지 과제를 포함한다. 참여한 여러 기업들 중 10개의 기업 및 대학팀(Huawei[3], Tencent[4], Bytedance[5], NYCU-PUT[6])에서 제안서 및 코드를 제출하였으며 이 기고에서는 Huawei[3], Bytedance[5]가 제안한 코덱을 집중적으로 서술하였다.

1. 이미지 처리

JPEG AI의 모듈에 따르면, 압축된 비트 스트림은 다수의 목적을 가진다. 이미지 처리 과제는 의미있는 이미지 정보를 얻기 위해 이미지를 수정하거나 품질을 향상시키는 과제를 포함한다. 다음은 이미지 처리 과제들이다.

- 초해상도(Super-resolution)
- 노이즈 제거(Denoising)
- 색보정(Color correction)
- 인페인팅(Inpainting)

압축 도메인 초해상도는 JPEG AI의 딥러닝 기반 이

이미지 코덱의 비트 스트림에서 곧바로 해상도를 높인 이미지를 얻는 것을 목표로 한다. 평가 지표는 기존 업샘플링 기법과 DNN 기반 초해상도 기법을 비교하는 것으로 한다. 이때, 표준을 위한 조건은 압축률은 조정 가능해야 하고, 높은 압축률이어야 하며, 변환된 이미지는 사람이 봤을 때 높은 효율과 성능의 이미지이어야 한다. 압축 도메인 노이즈 제거는 복호화 과정 중, JPEG AI의 덤러닝 기반 이미지 코덱의 압축된 비트 스트림에서 곧바로 노이즈를 제거하는 것을 목표로 한다. 이러한 목표를 위해 학습 가능한 디코더가 요구된다. 표준화 작업을 위해 노이즈 제거 Pre-trained 모델인 FFDNet이 제공되었고 파이썬의 사이킷-런 라이브러리를 통해 노이즈를 구현하였다. 압축 도메인 색보정은 복호화 과정 중, JPEG AI의 덤러닝 기반 이미지 코덱의 압축된 비트 스트림에서 곧바로 색보정을 지원한다. 압축 도메인 인페인팅은 복호화 과정 중, JPEG AI의 덤러닝 기반 이미지 코덱의 압축된 비트 스트림에서 곧바로 훼손된 이미지 영역을 복원하는 것을 목표로 한다.

2. 이미지 부호화

이미지 부호화는 압축을 통해 얻은 잠재 벡터로부터 이미지를 부호화하는 과제이다. 이미지 부호화를 포함하여 단일 코덱에서 다수의 과제를 수행할 수 있어야 한다. 즉, 이미지 처리와 컴퓨터 비전 과제에 사용된 인코더와 디코더가 이미지 부호화 과제에 사용된다.

3. 컴퓨터 비전

다음은 JPEG AI 컴퓨터 비전 과제들이다. 컴퓨터 비전 과제는 분류(Classification) 과제이며, 복호화된 이미지가 아닌 부호화된 잠재 벡터에서 분류하는 과제이다. 학습을 위해서 1300K장의 이미지넷 학습 데이터

셋을 사용하며, 50K장의 이미지넷 평가 테스트셋을 사용한다. 과제 참여 팀들은 Torchvision의 Resnet-50 Pretrained 모델을 사용하였다.

- 이미지 분류(Object Classification)
- 이미지 검출(Object Detection)
- 의미론적 분할(Semantic Segmentation)
- 행동 검출 및 인식(Event Detection and Action Recognition)
- 얼굴 검출 및 인식(Face Detection and Recognition)

압축 도메인 이미지 분류는 복호화 과정 중, JPEG AI의 덤러닝 기반 이미지 코덱의 압축된 비트 스트림에서 곧바로 학습에 사용된 레이블에 따라 객체를 분류한다. 압축 도메인 이미지 검출은 복호화 과정 중, JPEG AI의 덤러닝 기반 이미지 코덱의 압축된 비트 스트림에서 학습에 사용된 레이블에 따라 객체를 검출한다. 압축 도메인 의미론적 분할은 복호화 과정 중, JPEG AI의 덤러닝 기반 이미지 코덱의 압축된 비트 스트림에서 곧바로 타겟 객체를 픽셀 단위로 분할하는 의미론적 분할을 진행한다. 압축 도메인 이미지 행동/얼굴 검출 및 인식은 복호화 과정 중, JPEG AI의 덤러닝 기반 이미지 코덱의 압축된 비트 스트림에서 학습에 사용된 레이블에 따라 타겟 객체의 특정 행동 및 얼굴을 검출한다.

III. 기업들의 JPEG AI 과제 참여 현황 및 제안서

JPEG AI의 표준화 과제에 참여한 기업들의 현황을 살펴본다. <그림 2>를 통해 각 기업이 참여한 과제의 종류를 알 수 있다.

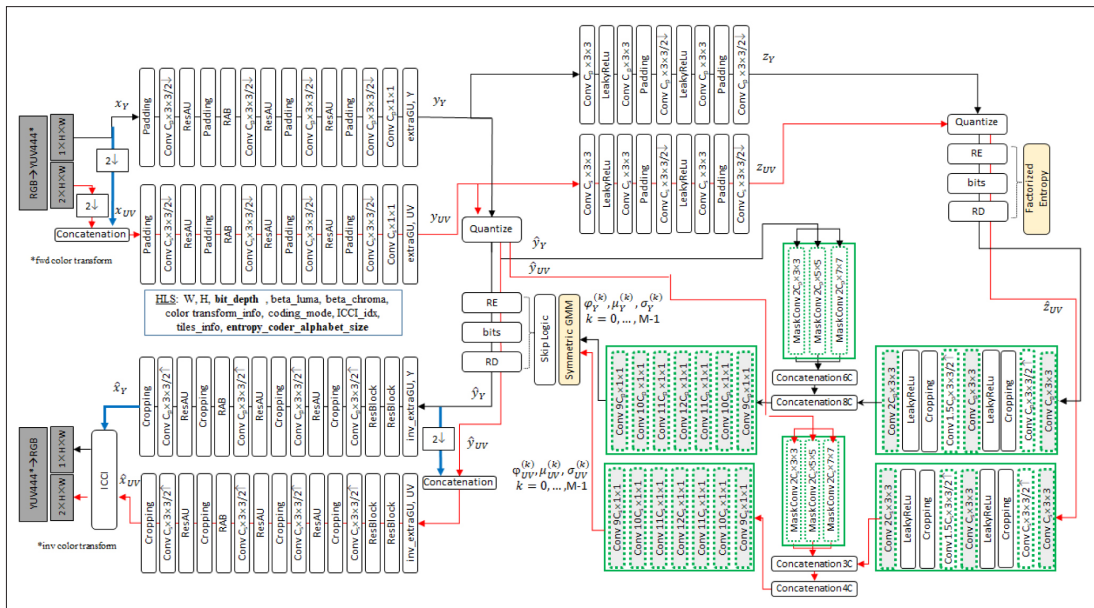
Team/codec Name	Team Affiliation	Std. Recons.	SuperResol.	Denoising	Image Class
Learning-based Image Coding Based on Combined Decoding and Denoising	EPFL				
Device agnostic learnable image coding using primary component extraction and conditional coding	Huawei Technologies (TEAM14)				
Neural Image Coding via Stacked Transformer and Convolution	Nanjing University, Purdue University (TEAM22)				
Augmented Normalizing Flow-based Image Compression	National Yang Ming Chiao Tung University, Poznan University of Technology (TEAM15)				
Learning-based image compression	Tencent America, Wuhan University (TEAM19)				
Task-driven end-to-end image compression	Tencent America, Wuhan University (TEAM13)				
IMCL-High Efficient Neural Image Coding	University of Science and Technology of China (TEAM17)				
Learned Image Compression with Discretized Gaussian-Laplacian-Logistic Mixture Model and Concatenated Residual Modules	Xi'an Jiaotong University, Simon Fraser University, University of Science and Technology of China, Google (TEAM12)				
Learned Image Compression using Long and Short Attention Modules	Xidian University, OPPO (TEAM21)				
Bytedance E2E Codec: subjective	Bytedance (TEAM16)				
Bytedance E2E Codec: objective	Bytedance (TEAM24)				

<그림 2> 기업 및 대학별 표준화 과제 참여표

1. 화웨이[3]

중국의 화웨이는 이미지 부호화와 이미지 분류 작업을 진행하였다. 제안된 설계는 가변비트 전송이 가능

하며 장치간 상호 운용성, 색상 구성 요소의 분리 기능을 포함한다. 또한 JPEG AI에서 선정된 테스트셋을 기준으로, 32.3%의 비트-왜곡을 이득을 얻었다. 화웨이가 제안한 전체 코덱의 구조를 <그림 3>에 제시하였다.



<그림 3> 화웨이의 딥러닝 이미지 압축 코덱

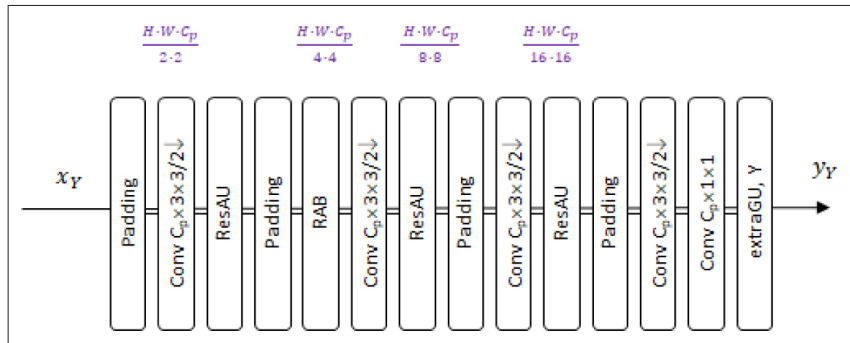
〈그림 11〉에서는 테스트셋의 이미지를 대상으로 VVC와 비교한 이미지 부호화 결과를 제시하였다.

표준화를 위한 JPEG AI의 7가지 평가 지표는 인간의 시각을 기반으로 정해졌다. 대다수의 평가 지표는 YUV 색 공간에서 계산이 되며, 특히 MS-SSIM과 같은 지표는 오직 광도 요소만을 계산한다. 따라서, 화웨이가 제안한 이미지 코덱은 한 가지 색 요소가 가장 많은 정보를 포함하며, 인간의 시각에 가장 많은 영향을 준다는 가정에 설계되었다. 색 변환은 [7]에서 참조되어 설계되었다. YUV 변환은 TUV BT.709 표준을 따른다.

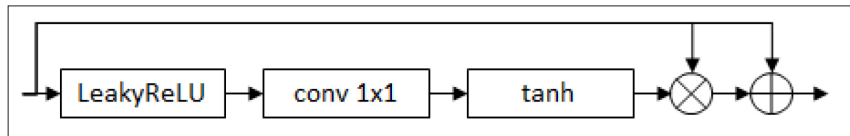
1) Analysis transform(부호화 과정)

〈그림 4〉는 부호화 과정의 주요 요소를 보여준다.

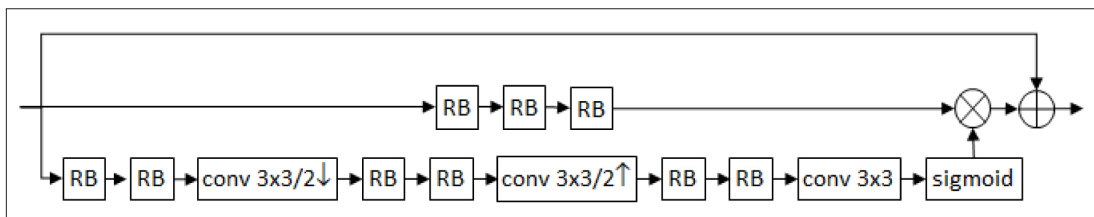
부호화 과정은 패딩 기술의 적용 이후에 네 개의 연속된 컨볼루션(Stride 2의 다운샘플링)을 포함한다. 이 과정은 Y요소에 대해서만 진행이 되며, 다른 UV요소에 대해서는 따로 똑같은 부호화 과정을 거쳐 진행하며 ICCI 모듈에서 통합된다. 컨볼루션 레이어 사이에 위치한 ResAU의 경우 요소별 비선형적 연산(Element-wise nonlinear operation)으로, 두 비선형 함수 사이에 1×1 컨볼루션이 포함된 형태이다. 이는 이미지의 어떠한 부분이 더욱 압축되어야 하는지에 대한 다운샘플링된 중요도 맵을 추출하는데 기여한다. 또한 한 개의 잔차 비지역 집중 블록(Residual non-local attention block)이 사용된다. RNAB 모듈은 잔차 블록들의 시퀀스로 이루어져 있다. 〈그림 5〉는 ResAU 블록의 구성



〈그림 4〉 분석 변환 단계의 주요 요소

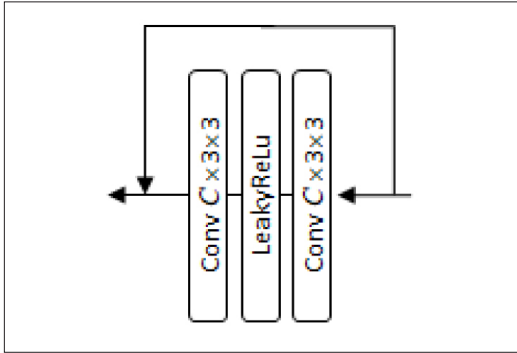


〈그림 5〉 ResAU 블록



〈그림 6〉 RNAB 블록

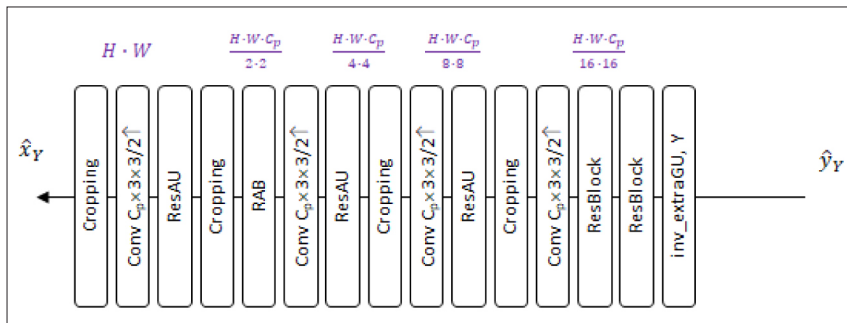
을, <그림 6>은 RNAB 블록을 보여주고, <그림 7>은 RNAB 블록을 이루는 잔차 블록을 보여준다.



<그림 7> 잔차 블록

2) Synthesis transform (복호화 과정)

<그림 8>은 복호화 과정의 주요 요소들을 보여준다. 복호화 과정은 부호화 과정의 역으로써, 네 개의 연속된 컨볼루션(Stride 2의 업샘플링 과정)을 포함한다. <그림 8>을 보면, 합성 변환 레이어의 출력의 행렬 크기를 확인할 수 있다(보라색). 두 개의 잔차 블록은 역이득 유닛(Inverse Gain Unit) 뒤에 위치해 있는 것을 확인할 수 있다. RNAB 모듈은 뒤에서 두 번째와 마지막 컨볼루션 레이어 사이에 위치한 것을 확인할 수 있다. ResAU 활성화 함수는 복호화 과정의 것과 동일하다.



<그림 8> 합성 변환 단의 주요 요소

3) 양자화

부호화 과정은 부동 소수점에서 작동한다. YUV에 대한 부호화의 출력은 float32 타입의 데이터이고 비트 스트림으로 쓰여지기 전에 정수 형태로 양자화된다. 출력의 비트 깊이는 분석단의 출력에 따라 설정된 엔트로피 부호화기의 비트 수를 따르는데, 512비트부터 2048비트까지 분포한다. 이러한 과정은 높은 확률로 존재하는 이미지 정보에는 적으면서도 적당한 비트 길이를 할당하여 정확도를 잃지 않도록 하며, 적은 확률로 존재하는 이미지 정보에는 많은 비트를 할당하지 않도록 한다.

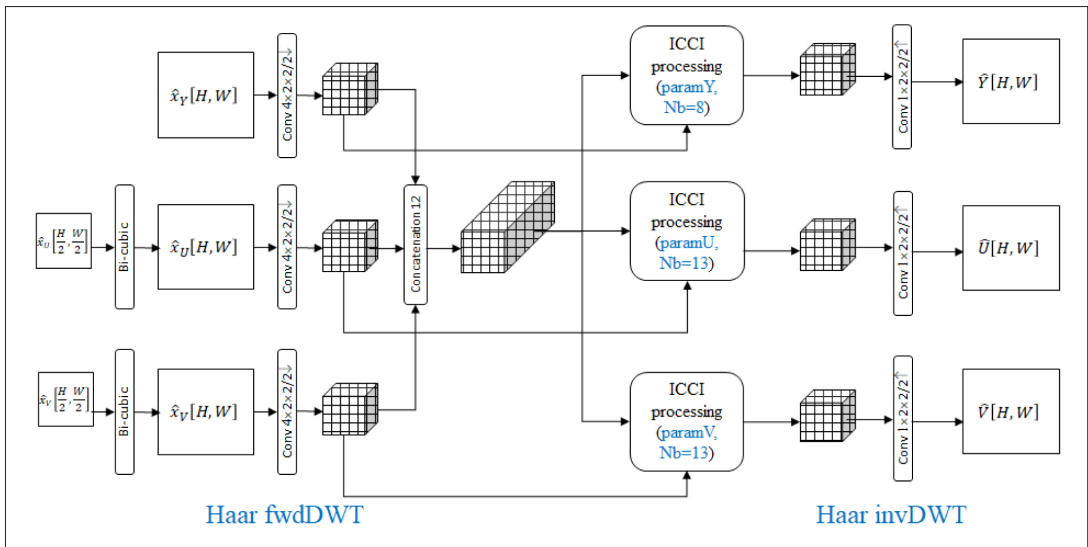
4) ICCI (Inter Channel Correlation Information filter sub-network)

ICCI 모듈은 차원 재구성 과정, 컨볼루션 과정, ICCI 처리 모듈, DWT (Discrete Wavelet Transform) 변환 과정으로 이루어진다. 쓰이는 DWT는 Haar 웨이블릿 변환으로써, Forward DWT와 Inverse DWT로 나뉜다. 앞서 언급된 것과 같이 색 요소들은 RGB에서 YUV로 색 변환된 이후 Y와 UV 영역으로 나뉘어 처리된다. 이때, Y요소를 제외한 UV 요소들은 부호화 과정에서 추가적인 차원 축소 과정을 거치게 되므로, 이후 Y요소와 함께 복호화를 진행하기 위해서는 UV영역에서 별도의 차원 확장 과정이 필요하다. ICCI 보조 네트워크

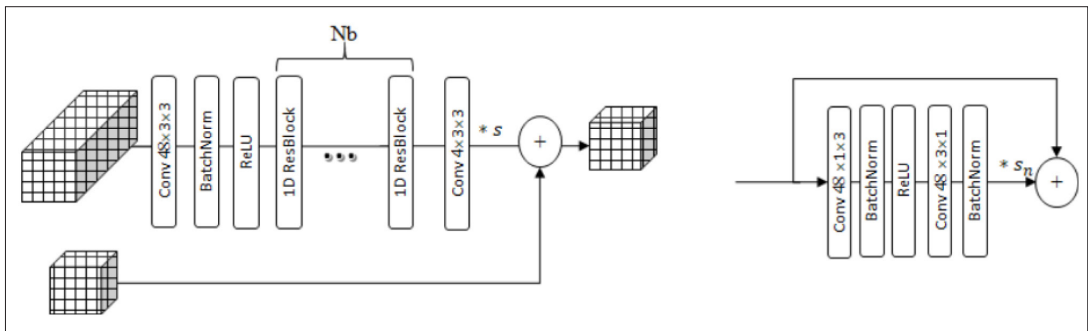
모듈은 $\hat{x}_{UV}(H/2, (W/2))$ 과 $\hat{x}_Y[H; W]$ 를 입력으로 받아 \hat{x}_{UV} 요소는 bi-cubic 보간법을 통해 차원 확장이 수행되며 \hat{x}_{YUV} 요소는 컨볼루션(Stride 2의 다운샘플링) 레이어와 Forward DWT 과정을 지난다. ICCI 네트워크는 차원 재구성 과정의 출력으로써 $\hat{x}_{YUV}(H/2, W, 12)$ 의 연결된 행렬을 얻고 Forward DWT 과정을 통해 네트워크는 $\hat{x}_X(H/2, W, 4)(X=Y$ 또는 $X=U$ 또는 $X=V)$ 행렬을 얻는다. <그림 9>에서 세 가지 색 요소가 합쳐진 행렬과 Forward DWT를 거친 행렬이 ICCI 처리 모

듈에 향하는 것을 확인할 수 있다. ICCI 처리 모듈은 각 입력을 받아 1차원 잔차 컨볼루션 블록의 수에 해당하는 Nb의 값에 따라 컨볼루션 단을 지난다. <그림 10>을 보면 ICCI 처리 모듈과 1차원 잔차 컨볼루션 블록을 확인할 수 있다.

<그림 11>은 이미지 부호화 과정의 테스트셋 결과 비교를 보여준다. 좌측은 화웨이가 만든 결과이고, 우측은 VVC를 이용해서 만든 결과이다. 그림에서처럼 화웨이에서 만든 결과가 VVC를 이용해서 만든 결과보다 약간



<그림 9> ICCI 모듈



<그림 10> ICCI 처리 모듈(좌), 1차원 잔차 컨볼루션 블록(우)



<그림 11> 이미지 부호화 과제의 테스트셋 결과 비교(좌: 화웨이 팀14, 우: VVC, bpp:0.12bpp)

더 좋은 것을 확인할 수 있다.

간을 단축시켰다.

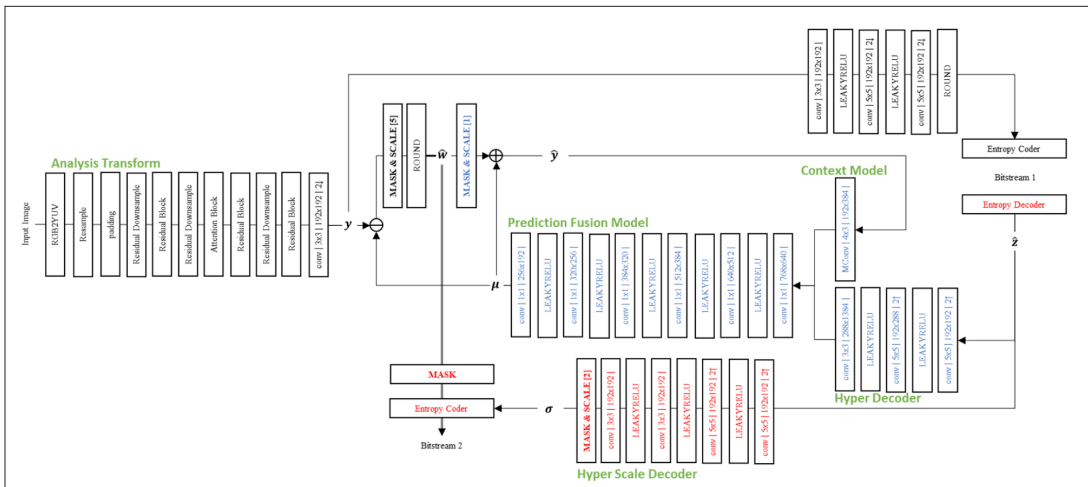
2. 바이트댄스[5]

중국의 바이트댄스에서 2개의 팀이 이미지 부호화 과제에 제안하였다. 테스트셋을 기준으로, 각 팀은 VVC를 이용한 비디오 부호화 방법에 비해 -17.9%와 -29.9%의 평균 비트-왜곡률 이득을 얻었다. 또한, 두 팀 모두 GPU 환경에서 VVC와 HEVC보다 복호화에 걸리는 시

1) Analysis transform (부호화 과정)

부호화 과정은 네 가지 단계로 나뉜다. <그림 12>는 바이트댄스의 딥러닝 이미지 압축 부호화기를 보여준다.

- 분석 변환: 합성 변환에 대응되는 변환이다. 입력 이미지로부터 잠재 벡터 샘플 y 를 얻는다.



<그림 12> 바이트댄스의 딥러닝 이미지 압축 부호화기

- 하이퍼 분석 변환: 하이퍼 합성 변환에 대응되는 변환이다. 엔트로피 코딩을 거쳐 <그림 12>의 비트 스트림 1을 생성한다.
- 잔차 샘플 생성: 복호화단의 잠재 샘플 예측에 대응되는 모듈이다. 예측 샘플 μ 을 얻기 위해 진행된다. 이후 잠재 벡터 샘플 y 와의 차이를 통해 잔차 샘플을 얻고, 반올림하여 양자화된 잔차 샘플 \hat{w} 을 얻는다.
- 엔트로피 부호화: 엔트로피 복호화에 대응되는 변환이다. 하이퍼 분석 변환을 통해 얻은 표준 편차 σ 를 이용하여 양자화된 잔차 샘플 \hat{w} 을 엔트로피 부호화한다.

2) Synthesis transform (복호화 과정)

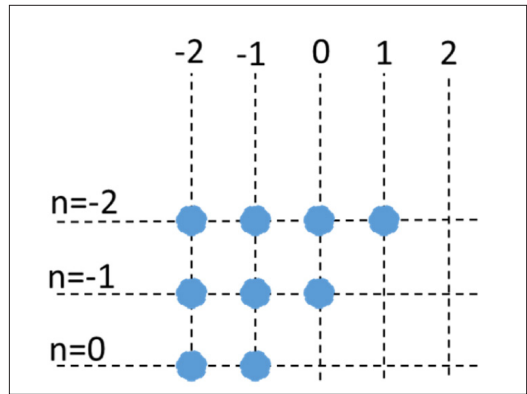
복호화 과정은 세 가지 단계로 나뉜다. <그림 13>은 바이트댄스의 딥러닝 이미지 압축 복호화기를 보여준다.

- 엔트로피 복호화 처리: 양자화된 하이퍼 잠재벡터 \hat{z} 과 양자화된 잔차 잠재벡터 \hat{w} 을 얻는다.
- 잠재 샘플 예측: 앞서 추출한 \hat{z} 과 \hat{w} 로부터 잠재 샘플 \hat{y} 을 예측한다.
- 합성 변환: 잠재 샘플 예측 과정에서 예측한 \hat{y} 으로부터 이미지를 복원한다.

3) Wavefront parallel processing

(파면 병렬 프로세싱)

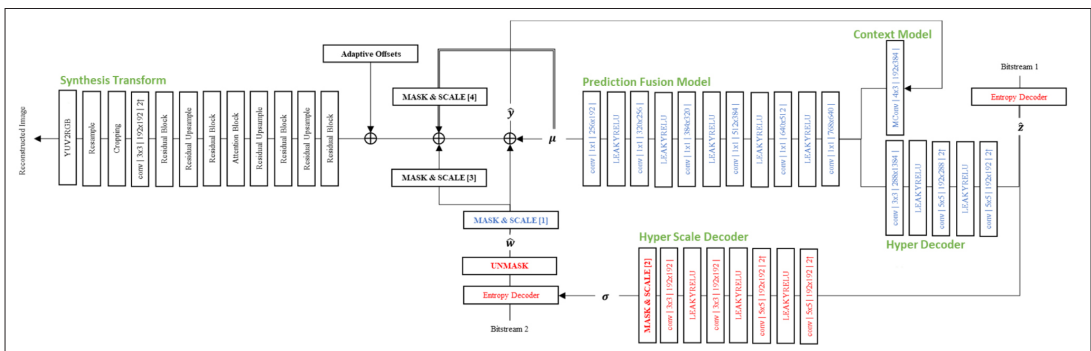
파면 병렬 프로세싱은 병렬 연산이 가능한 GPU의 활용도를 높이기 위한 방법으로, 부호화 과정에 따른 컨볼루션 연산을 병렬적으로 처리할 수 있게 한다. <그림 14>는 마스킹된 컨볼루션 커널을 보여주는데, 이때 예측한 샘플은 (0,0)좌표에서 제공된다. 커널은 한 샘플의 지연으로 샘플 행과 병렬로 샘플 행을 처리할 수 있는 방식으로 설계되었다.



<그림 14> 마스킹된 컨볼루션 커널

4) YUV 색 포맷 사용

부호화/복호화 과정에서, JPEG AI는 입력 이미지와



<그림 13> 바이트댄스의 딥러닝 이미지 압축 복호화기

복호화된 이미지 모두 8비트 또는 16비트 이미지로 설계 제한 조건을 설정하였다. 따라서, 모델의 학습 과정에서 입력 이미지는 BT. 709 표준[7]에 따라 YUV 색 공간으로 변환되어 사용되었다. YUV 색 공간을 훈련에 사용함으로써 Y영역에서의 학습 손실을 줄였다.

5) Adaptive quantization (적응형 양자화 기술)

적응형 양자화 기술을 통해 디코더에서 실수 픽셀 요소들은 스케일링 기술과 마스킹 기술을 통해 강조할 픽

셀과 양자화 간격을 결정하여 양자화한다. 아래 수식들을 따라 구현된다.

$w[c,i,j]$ 는 양자화되기 전의 잔차 벡터 샘플이며, “thr”, “scale”, “flag”는 비트 스트림에 적용되는 파라미터들이다. 2와 3에 참조된 마스크는 1에서 정해진 값이다. <그림 15>는[8] 이미지 부호화 과제 테스트셋의 결과로써 VVC표준과 유사하거나 다소 향상된 부호화 결과를 볼 수 있다.

1. 각 잠재 샘플 벡터에 대해 다음 조건에 따라 마스킹된다.

$$\text{mask}[c, i, j] = \begin{cases} \text{True} & \text{if } \sigma[c, i, j] > \text{thr}, \text{flag} = \text{True} \\ \text{False} & \text{if } \sigma[c, i, j] < \text{thr}, \text{flag} = \text{False} \\ \text{False} & \text{otherwise} \end{cases}$$

2. 1에서 정해진 $\text{mask}[c,i,j]$ 값에 따라 스케일링 연산이 양자화된 잔차 샘플들과 가우시안 표준편차 샘플들에 적용된다.

$$\sigma[c, i, j] = \begin{cases} \sigma[c, i, j] \times \text{scale} & \text{if } \text{mask}[c, i, j] = \text{True} \\ \sigma[c, i, j] & \text{otherwise} \end{cases}$$

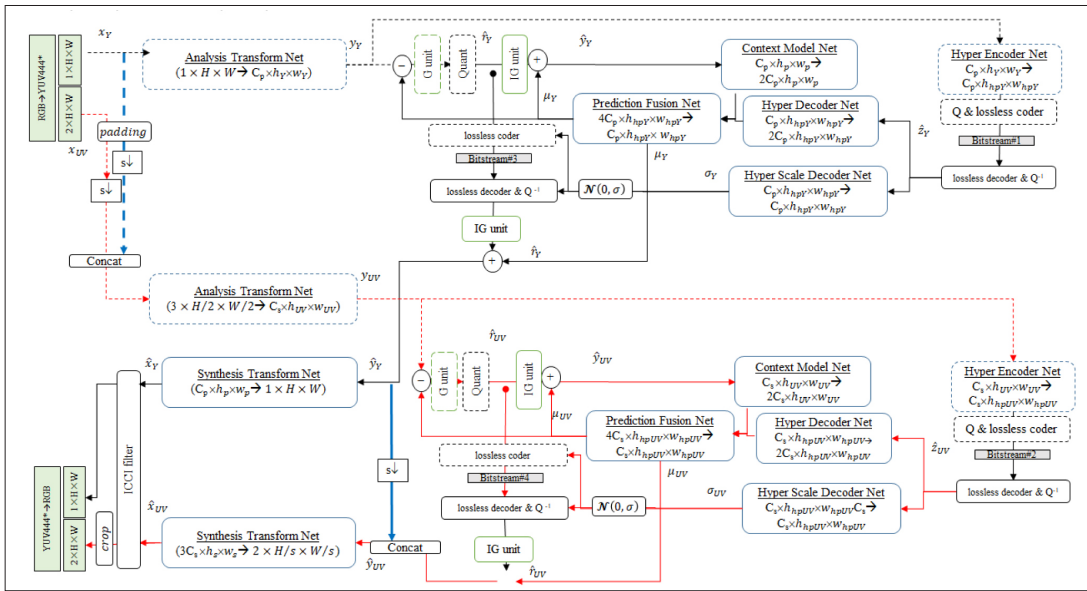
$$\hat{w}[c, i, j] = \begin{cases} \hat{w}[c, i, j] \times \text{scale} & \text{if } \text{mask}[c, i, j] \text{ is equal to True} \\ \hat{w}[c, i, j] & \text{otherwise} \end{cases}$$

3. 부호화단에서 추가적으로 다음 수식을 따라 적응형 양자화를 진행한다.

$$w[c, i, j] = \begin{cases} \hat{w}[c, i, j] / \text{scale} & \text{if } \text{mask}[c, i, j] \text{ is equal to True} \\ \hat{w}[c, i, j] & \text{otherwise} \end{cases}$$



<그림 15> 이미지 부호화 과제 테스트셋의 결과 비교(왼쪽: 팀 16, 중간: 팀 24, 오른쪽: VVC표준, 0.25bpp)



<그림 16> VMuC 프레임워크

3. VMuC (Verification Model under Consideration)[9]

앞서 소개한 두 기업의 코덱을 참고하여 JPEG AI에서 만든 VM (Verification Model)을 소개하고, 딥러닝 학습 과정을 소개한다. 현재 완성되지 않아서 VMuC로 명명하고 있다. <그림 16>은 VMuC의 프레임워크를 보여준다.

1) 훈련 데이터셋 전처리

JPEG AI의 5264개 데이터셋이 사용되었다. 학습 속도의 향상을 위해, 데이터셋을 122502개의 패치로 나누었다. 우선 데이터셋 이미지의 크기가 1024×1024 보다 크면 자르지 않았고, 그보다 클 시 1024×1024 의 크기로 잘라 사용하였다. 이후 패치 이미지의 밀도를 향상시키기 위해 1024×1024 패치는 512×512 의 크기로 다운샘플링되었고, 최종적으로 320×320 크기의 패치로 잘라 학습 과정에 사용되었다.

2) 손실함수

최적화를 위한 손실함수는 아래 수식과 같다.

- (1) $Loss = bpp + \beta \cdot distortion$
- (2) $Loss = bpp + \beta \cdot (MSE + MSSIM/2)$
- (3) $Loss = bpp$

비트를 bpp 는 샘플 수에 따라 정규화되었다. β 는 조절 가능하고, 왜곡 정도를 조절하는 변수이다. 각 모델마다 다르게 적용되었다.

3) 학습 과정

딥러닝 모델의 학습 과정은 4단계로 나뉘어 진행되었다. 첫 단계는 가중치 파라미터를 랜덤 난수로 설정하여 학습을 시작하는 단계로, epoch은 64회로 설정된다. β 는 각 모델마다 0.002, 0.007, 0.015, 0.05의 네 가지 경우 중 하나로 사용되었으며 고정되어 학습하였다. 손실함수는 3.2의 (1) Loss를 사용했다. 두 번째 단계는

epoch이 12회이며 β 는 0.002, 0.007, 0.015, 0.05 중 랜덤하게 선택하여 학습하였다. 손실함수는 (1) *Loss*를 사용하였다. 세 번째 단계는 epoch이 48회이며 β 는 0.002, 0.007, 0.015, 0.05 중 랜덤하게 선택하여 학습하였다. 손실함수는 (2) *Loss*를 사용하였다. 네 번째 단계는 epoch이 20회이며 엔트로피 부분만을 학습하기 때문에 (3) *Loss*를 사용하였다. 모든 과정에서 학습률은 0.001로 사용되었다.

4. 평가방법[8]

정량적 평가방법으로는 MS-SSIM, SSIM, VMAF, PSNR 등을 사용하였다. SSIM (Scale Structure Similarity)은 색 요소를 기준으로 원본 이미지와 왜곡된 복원 이미지 사이의 품질 차이를 측정한다. 다음은 SSIM의 수식이다. x 는 원본 이미지, y 는 복원 이미지이다.

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{(\mu_x)^2 + (\mu_y)^2 + C_1}$$

PSNR (Peak Signal-to-noise ratio)은 신호의 최대 전력에 대한 잡음의 전력이다. 압축된 이미지의 화질에 대한 손실 정보를 평가하기 위한 목적으로 사용된다. 다음은 PSNR의 수식이다. s 는 채널의 최댓값과 최솟값의 차이, MSE는 원본 이미지와 복원 이미지와의 평균 제곱오차이다.

$$\text{PSNR} = 10\log \frac{s^2}{\text{MSE}}$$

MS-SSIM (Multi-Scale Structure Similarity)도 SSIM과 마찬가지로 원본 이미지와 왜곡된 이미지 사이의 품질 차이를 측정한다. MS-SSIM은 SSIM에서 이미지 해상도와 같은 요소들을 추가하여 단일 규모 방식인 SSIM보다 더 다양한 측면에서의 평가를 가능하게 하였다. MS-SSIM은

VMuC 모델의 손실함수의 요소로도 사용되었다.

VMAF (Video Multimethod Assessment Fusion)는 넷플릭스에 의해 만들어진 평가지표로써, 압축과 차원 재구성에 의한 노이즈에 대한 영향을 측정한다. VMAF는 여러 품질 평가 알고리즘을 기계학습의 회귀 알고리즘인 서포트 벡터 머신에 적용하여 다양한 측면에서의 품질 평가를 가능하게 해준다. VMAF의 연산에 사용되는 요소는 VIF (Visual Information Fidelity), DLM (Detail Loss Metric), Motion이다.

IV. 활용 사례

1. 클라우드 서비스

여러 온라인 스토리지 서비스의 개발과 함께, 많은 이미지가 클라우드에 저장되고 있다. Tencent, Microsoft, Facebook, Google과 같은 회사들은 수십만의 이미지를 저장하고 있고 동시에 엄청난 규모의 저장 공간, 전력, 대역폭을 필요로하므로, 표준화를 위해 주요한 요구사항은 다음과 같다.

- 조정 가능한 손실 압축률/품질
- 높은 압축효율
- 이미지 처리 Task들을 위한 효율적인 압축 형태
- 인식 가능한 품질 최적화
- 사생활 보호

2. 영상보안 시스템

영상보안 시스템은 여러 물체들을 감시 및 검출하기 위해 설치되어 있다. 의심스러운 활동의 검출부터, 포렌식 데이터 증거 제공 및 지능형 제어가 가능한 시스템과

지 다양하다. 이러한 감시 카메라들은 처리를 요하거나, 분석되거나 압축되어야 하는 동영상 데이터를 다량 생성하므로, 표준화를 위한 주요한 요구사항은 다음과 같다.

- 조정 가능한 손실 압축률/품질
- 높은 압축효율
- 컴퓨터 비전 과제들을 위한 효율적인 압축 형태
- 낮은 부호화 복잡성
- 랜덤 공간 탐지(특히 고해상도 카메라를 위해)
- 사생활 보호

3. 실시간 미디어 송출 및 처리

실시간 송출을 위한 동영상 데이터의 양은 급속도로 증가하였다. 실시간 송출 데이터는 전문적인 서비스인 온라인 강의나 실시간 컨퍼런스 뿐만 아니라 비디오 게임 라이브 송출 혹은 소규모의 개인 비디오까지 다양하다. 그러한 동영상 데이터는 부적절한 콘텐츠를 검출하거나 미디어법을 어기는 영상을 포함할 수 있고, 얼굴 및 감정, 제스처와 같은 추가적인 정보들도 제공할 수 있기 때문에 분석되어야 한다. 또한, 컴퓨터 비전 과제들도 이러한 실시간 미디어에 적용될 수 있으므로, 표준화를 위한 주요한 요구사항은 다음과 같다.

- 조정 가능한 손실 압축률
- 높은 압축효율
- 컴퓨터 비전 과제들을 위한 효율적인 압축 형태
- 신속한 부호화 및 복호화
- 사생활 보호

4. 자율주행

자율주행 자동차 뿐만 아니라 드론과 같이 자율주행

이 가능한 장치들은 분석을 요하는 동영상 데이터들을 대량 생산하며, 직접적인 처리를 요한다. 따라서, 효율적으로 데이터를 저장 및 전송하는 시스템이 필요하다. 예를 들어, 자율운행 드론들은 수십 기가바이트의 고해상도 이미지 데이터를 생산하므로 전송 및 저장 기기에 따른 한계점이 존재한다. 또한, 교통사고 분석과 교통 상황 모니터링과 같이 직접적인 처리가 필요한 컴퓨터 비전 과제들은 객체 검출(Object Detection), 의미론적 분할(Semantic segmentation)을 수행하므로 데이터의 저장과 전송이 더욱 중요하다. 표준화를 위한 주요한 요구사항은 다음과 같다.

- 비트율이 조정가능한 손실 압축 기술
- 높은 압축효율
- 신속한 처리가 가능한 코덱
- 컴퓨터 비전 과제들을 위한 효율적인 압축 형태
- 이미지 처리 과제들을 위한 효율적인 압축 형태
- 비손실 압축이 가능한 부호화
- 사생활 보호

V. 결론[10]

JPEG AI에서는 다인종, 다연령층으로 구성된 평가단 274명의 정성적 평가 정보를 수치로 환산한 DMOS (Differential Mean Opinion Scores) 점수를 통해 10개 팀의 결과를 분석하였다. 여러 이미지를 대상으로 진행한 평가 실험을 통해 얻은 평가단의 신뢰구간은 95%이다. <그림 17>에 각 평가 이미지에 대해 비교한 그래프를 제시하였으며, HEVC, JPEG 2000을 비교 기준에 포함하였다. 이를 통해 다음과 같은 결론이 도출되었다.

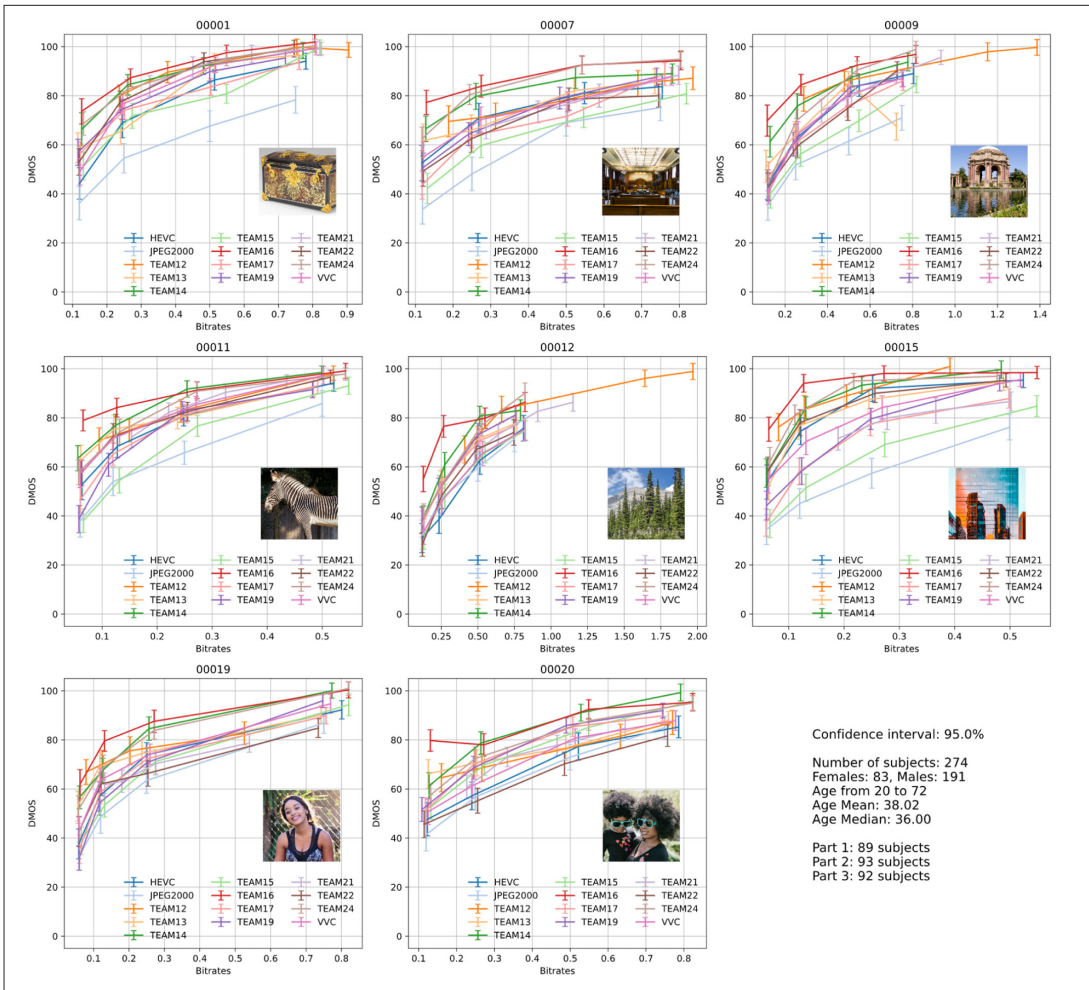
- JPEG 2000은 가장 낮은 압축률을 보여준다. 최근

의 기술은 기존 압축 포맷을 뛰어넘는 결과를 보여 주고 있다. 다만, 테스트셋 00020 이미지의 경우 팀 22보다 향상된 압축률을 보여주었으며 00012 이미지의 경우 HEVC Intra 기술을 적용한 이미지보다 JPEG 2000 기술을 적용한 이미지가 더 높은 압축률을 보여주었다.

- VVC Intra 기술은 HEVC Intra 기술을 적용했을 때보다 더 높은 압축률을 보여주었다. 다만,

00015 이미지의 경우 HEVC 코덱을 사용했을 때 VVC 코덱보다 더 높은 압축률을 보여주었다. 이는 화면 문맥 기술(Screen Content tool)이 적용된 결과인 것으로 보인다.

- 대다수의 경우 약 4개의 팀(팀 12,14,16,24)들은 VVC 코덱보다 더 나은 성능을 보여준다. 많은 특징을 포함하는 고주파 이미지 00020의 경우 8개의 팀이 VVC 코덱을 뛰어넘는 것으로 보아 학습 가능



<그림 17> 팀별 평가 이미지에 대한 비트율-DMOS 성능 비교 그래프

한 코덱이 고주파 이미지에 대해 좋은 성능을 낸다는 것을 확인할 수 있었다. 또한, 00001 이미지에 대해서는 DMOS가 80인 구간에서 VVC 코덱보다 33%의 비트감소율을 확인할 수 있었다.

- 팀 16의 코덱은 대다수의 경우 다른 팀들보다 향상된 코덱 성능을 보여주었다. 특히, 낮은 비트율로 압축한 경우에 그러하였다. 팀 14의 경우에도 상당히 높은 성능을 보여주었는데, 특히 이미지 00020, 00007, 00011 등에서 그러한 경향을 보였다. 팀 24도 팀 14에 버금가는 높은 성능을 보여주었다.

본 기고에서는 JPEG AI에 제안한 팀들의 압축 프레임워크의 분석 및 활용 사례에 대해 알아보았다. 해가 갈수록 폭발적으로 데이터가 생산되고 있어, 데이터를

더욱 효율적으로 압축하여 저장하는 기술은 필수적이다. 한편, 딥러닝 알고리즘의 발전과 GPU 성능의 향상은 이미지 압축 분야에 대한 새로운 수요를 확대시켰다. 이에 기존 고효율 압축, 고성능 복원이라는 일반적인 연구 목표를 넘어 향상된 처리 속도와 이미지 처리 과제에 대한 광범위한 활용 사례 또한 활발히 연구되어가고 있다. 딥러닝을 이용한 이미지 압축은 기존의 딥러닝을 사용하지 않은 코덱과 비교하였을 때 압축률과 처리 속도면에서 향상된 성능을 보여주고 있음을 확인할 수 있었다. 다만, 아직 표준화 연구가 진행 중이고 사회의 수요에 맞게 상용화되지는 않은 상황이다. 따라서 압축 기술의 꾸준한 연구와 함께 앞서 소개한 세 가지 과제들의 발전에 발맞춰 압축 기술이 적용된다면 더욱 상용화되어 사회의 편익에 기여할 수 있을 것이다.

참고 문헌

- [1] IT 데일리, “2025년 전 세계 데이터총량, 지금의 10배인 163ZB 규모”, <http://www.itdaily.kr/news/articleView.html?idxno=82830>
- [2] ISO/IEC JTC1/SC29/WG1 N100094, “JPEG AI Use Cases and Requirements,” 94th Meeting, Online, January 2022.
- [3] ISO/IEC JTC 1/SC29/WG1 M96016, “Presentation of the Huawei response to the JPEG AI Call for Proposals,” 96th JPEG Meeting, Online, July 2022.
- [4] ISO/IEC JTC 1/SC29/WG1 M96051, “Tencent response to the JPEG AI Call for Proposals: Learning-based image compression,” 96th JPEG Meeting, Online, July 2022.
- [5] ISO/IEC JTC 1/SC29/WG1 M96053, “Bytedance’s response to the JPEG AI Call for Proposals,” 96th JPEG Meeting, Online, July 2022.
- [6] ISO/IEC JTC 1/SC29/WG1 M96054, “NYCU-PUT Response to the JPEG AI Call for Proposals,” 96th JPEG Meeting, Online, July 2022.
- [7] Tiansheng Guo, Jing Wang, Ze Cui, Yihui Feng, Yunying Ge, Bo Bai, Huawei Technologies, Beijing, China, Variable Rate Image Compression With Content Adaptive Optimization
- [8] ISO/IEC JTC 1/SC29/WG1 N100106, “JPEG AI Common Training and Test Conditions,” 94th JPEG Meeting, Online, January 2022
- [9] ISO/IEC JTC 1/SC29/WG1 M96016, “Description of the JPEG AI Verification Model under Consideration and associated software integration procedure,” 97th JPEG Meeting, Online, October 2022
- [10] ISO/IEC JTC 1/SC29/WG1 N100095, “Final Call for Proposals for JPEG AI”, 94th JPEG Meeting, Online, January 2022.

필자 소개



한 승 진

- 2020년 ~ : 단국대학교 전자전기공학과 학사과정
- 주관심분야 : 영상 처리, 영상 압축, 딥러닝



김 영 섭

- 1984년 : 한양대학교 학사
- 1991년 : 남가주대학교 석사
- 2001년 : 랜셀러폴리테크닉 대학교 박사
- 2003년 ~ 현재 : 단국대학교 전자전기공학부 교수
- 주관심분야 : 영상 처리, JPEG/MPEG 표준, 증강현실과 가상현실 처리, 딥러닝