특집논문 (Special Paper)
방송공학회논문지 제30권 제6호, 2025년 11월 (JBE Vol.30, No.6, November 2025)
https://doi.org/10.5909/JBE.2025.30.6.1041
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

토큰당 고정 비트 할당을 통한 프레임 압축기의 설계

이 혜 린[a], 허 정 윤[a], 송 성 운[a], 이 학 범[a], 서 영 호[b)‡]

Design of a Frame Compressor with Fixed-Bit Token Assignment

Heylin Lee[a], Jeong-Yun Heo[a], Seong-Un Song[a], Hak-Bum Lee[a], and Young-Ho Seo[b)‡]

요 약

본 논문에서는 VQ-VAE(Vector Quantized Variational AutoEncoder) 기반의 픽셀당 고정 비트 할당 방식을 적용하여, 정확한 4:1 고정 비트율을 보장하는 프레임 압축기를 제안한다. 본 연구는 압축 효율 경쟁을 지향하는 표준 비디오 코덱의 접근과 달리, 프레임 저장 및 전송 과정에서 메모리 대역폭을 정확하게 예측할 수 있는 하드웨어 친화적 구조를 목표로 한다. 제안하는 압축기는 인코더-벡터 양자화기-디코더 구조로 구성되며, residual 블록과 HardSwish 활성화 함수를 적용하여 하드웨어 구현을 고려한 설계를 하였다. 공간적 다운샘플링 없이 벡터 양자화기의 6비트 임베딩 차원 조정을 통해 정확한 4:1 압축비를 달성한다. HEVC-B 데이터셋을 사용한 실험에서 평균 PSNR 43.92dB와 BPP 6.0을 기록하여 목표 성능을 만족하였으며, 원본과 시각적으로 구별하기 어려운 수준의 복원 품질을 보여준다. 본 방법은 온칩 프레임 저장 및 파이프라인 데이터 전송 등 정확한 비트율 제어가 요구되는 응용 환경에서 효과적으로 활용될 수 있다.

Abstract

In this paper, we propose a frame compression codec that guarantees an exact 4:1 constant bit-rate (CBR) by applying a pixel-wise fixed-bit allocation strategy based on the Vector Quantized Variational AutoEncoder (VQ-VAE) Unlike standard video codecs that primarily aim to maximize compression efficiency, the proposed method focuses on a hardware-friendly architecture that enables precise prediction of memory bandwidth during on-chip frame storage and intra-pipeline data transmission. The proposed codec adopts an encoder‑vector-quantizer‑decoder structure, where Residual Blocks and the HardSwish activation function are incorporated to enhance hardware implementability. Without any spatial down-sampling, a strict 4:1 compression ratio is achieved by constraining the codebook index to 6 bits per pixel. Experimental results on the HEVC-B dataset demonstrate that our method achieves a mean PSNR of 43.92 dB at 6 bits per pixel, providing visually indistinguishable reconstruction quality from the original input. Owing to its predictable bit-rate and lightweight design, the proposed approach is suitable for applications where accurate bandwidth control is essential, such as on-chip frame buffering and pipeline-level data transmission.

Keyword : VQ-VAE, Constant Bit Rate Compression, Frame Compression, Residual Block, HardSwish

Ⅰ. 서 론

21세기 디지털 미디어 시대에 접어들면서 고화질 비디오 콘텐츠에 대한 수요가 급증하고 있으며, 모바일·엣지 환경에서도 안정적인 영상 처리가 요구되고 있다. 현재 널리 사용되고 있는 비디오 압축 표준들은 높은 압축 효율을 위해 가변 비트율(Variable Bit Rate, VBR) 방식으로 운용되는 경우가 많다[1]. 그러나 VBR 방식은 프레임별 비트량 편차가 커 네트워크 대역폭 예측과 임의 프레임 접근에 제약이 존재한다. 이러한 문제를 해결하기 위한 대안으로 고정 비트율(Constant Bit Rate, CBR) 압축 방식을 고려할 수 있다. 이와 관련하여, 고정 비트율 기반 설계의 필요성은 다양한 연구에서 확인되었다. 예를 들어 Mohsenian et al.(1999)은 단일 패스에서 프레임 상대 복잡도를 추정해 압축 파라미터를 조정함으로써 CBR 제약 하에 품질을 높이고, 방송/편집 환경에서 CBR 인코더가 멀티플렉싱·프리즈 프레임 등 실무 요구에 부합함을 보였다[2]. Sanz-Rodríguez et al.(2007)는 stored B-picture 기반 RC(Rate Control)와 톱니형 목표 버퍼 레벨을 도입해 버퍼 점유·목표 비트율 추종성을 개선, B-픽처 효율과 저지연 운용 사이의 트레이드오프를 완화하였다[3]. 이처럼 고정 비트율 설계는 네트워크/시스템 레벨에서 검증된 실용적 장점이 있음에도 불구하고, 표준 코덱과 동일한 철학으로 압축 효율 경쟁을 추구하는 방향에 집중되어 있다. 반면, 하드웨어 기반 영상 처리 파이프라인에서는 요구 조건이 다르다. 온칩 메모리 대역폭은 한정적이며, 프레임 비트량 변동이 버퍼 오버플로우·언더플로우를 야기하여 시스템 동작 안정성을 크게 저해할 수 있다. 따라서 하드웨어 시스템에서는 모든 픽셀에 동일한 비트를 부여함으로써 각 프레임의 전송 비트량을 사전에 보장하고, 파이프라인 단계 간 대역폭 및 버퍼 크기 설계를 단순화할 수 있다. 또한 엔트로피 디코딩과 같은 복잡한 처리를 배제하여 일관된 처리 속도를 유지할 수 있으며, 임의 프레임 접근성을 강화하고 시스템의 예측 가능성과 신뢰성을 크게 향상시킨다. 즉, 본 연구는 압축 효율 극대화를 목적으로 비트량을 가변적으로 조정하거나 QP를 변화시키는 기존 접근과 달리, 모든 픽셀에 대해 정확히 동일한 비트를 할당함으로써 시스템 안정성과 실시간 처리 보장을 최우선으로 하는 고정 비트율 기반 설계에 중점을 둔다.

최근 딥러닝 기술의 발전은 영상 압축 분야에도 큰 변화를 가져왔다. 딥러닝 기반 압축 방식은 데이터로부터 최적의 표현을 자동으로 학습하고, 특정 도메인/콘텐츠 유형에 특화된 모델을 훈련시킬 수 있다(의료·위성·게임 등). 그 중에서도 VQ-VAE는 연속적인 잠재 공간 대신 이산적인 코드북을 사용해 안정적인 학습과 정확한 비트 표현이 가능하다[4]. 엔트로피 코딩에 의존하지 않아 압축 구조를 간결하게 유지하며 목표 비트율을 직접 제어할 수 있어, 사전에 지정된 비트 예산을 준수해야 하는 고정 비트율 압축기로 자연스럽게 확장할 수 있다[5].

딥러닝 아키텍처 측면에서, ResNet은 잔여 연결을 통해 깊은 네트워크의 학습 안정성·표현력을 높였고[6], 압축에서도 RNN/Residual 설계를 통해 재구성 품질 향상과 아티팩트 저감을 보고하였다[7]. 활성화 함수 역시 성능과 구현 복잡도에 큰 영향을 준다. ReLU[8] 이후 Leaky-ReLU/ELU/Swish(SiLU) 등이 제안되었고[9,10] 모바일/엣지 환경에서는 ReLU6, Hard-Swish/Hardsigmoid[11] 등 하드웨어 친화적 근사 함수가 널리 활용된다. 정규화 또한 중요하다. 원래 Conv-BN-Activation 순서로 제안된 BN(Batch Normalization)[11]은 Pre-activation(BN-ReLU-Conv) 설계를 통해 더 깊은 네트워크에서 안정적 학습을 가능케 했고, 다양한 정규화/활성화 조합의 효과가 체계적으로 분석되었다. 압축 응용에서는 정보 손실로 인한 학습 불안정이 발생하기 쉬워, 정규화 선택/배치의 영향이 더 크며, 이는 일정 비트율 제약 하에 품질/안정성을 동시에 달성하려는 본 연구의 설계와 직결된다.

이러한 배경을 바탕으로, 본 논문에서는 모든 픽셀에 대

a) 광운대학교 전자융합공학과(Department of Electronic Convergence Engineering, Kwangwoon University)
b) 광운대학교 전자재료공학과(Department of Electronic Materials Engineering, Kwangwoon University)
‡ Corresponding Author : 서영호(Young-Ho Seo)
E-mail: yhseo@kw.ac.kr
Tel: +82-2-300-0263
ORCID: https://orcid.org/0000-0003-1046-395X
※이 논문의 결과 중 일부는 한국방송·미디어공학회 2025년 하계학술대회에서 발표한 바 있음
※본 연구는 2025년도 중소벤처기업부의 기술개발사업 지원에 의한 연구임 G21023159501. This work was supported by the Technology development Program(G21023159501) funded by the Ministry of SMEs and Startups(MSS, Korea)
· Manuscript November 7, 2025; Revised November 10, 2025; Accepted November 10. 2025.

해 정확히 동일한 비트를 할당하여 4:1 고정 비트율을 보장하는, VQ-VAE 기반 프레임 압축기를 제안한다. 제안 모델은 ResNet 기반 잔차 학습 구조를 적용하고 Conv-BN-Activation 순서를 채택하여 복원 성능과 학습 안정성을 동시에 향상시켰다. 활성화 함수는 딥러닝 영상 처리 분야에서 유효성이 검증된 SiLU(Sigmoid Linear Unit)를 기본으로 하되, 실제 하드웨어 구현을 고려한 Hard-Swish + Hardsigmoid 활성화 함수를 적용하여 연산 복잡도를 최소화하면서도 높은 복원 품질을 유지하도록 설계하였다. 또한 공간적 다운샘플링 없이 벡터 양자화기의 코드북 인덱스를 6비트로 고정하여 정확히 4:1 고정 비트율을 달성하였다. 그 결과, 얕은 네트워크 깊이와 50 epochs의 짧은 학습으로도 40dB 이상의 복원 품질을 확보하였으며, 프레임 저장 및 내부 전송 환경에서 요구되는 예측 가능한 데이터 처리 특성을 만족함을 보였다. 이는 전통 CBR 연구가 목표로 해온 지연·버퍼 안정성 요구와 딥러닝 기반 압축의 표현력·적응성을 유기적으로 결합한 접근으로 볼 수 있다.

본 논문의 구성은 다음과 같다. 제2장에서는 고정 비트율(CBR)과 딥러닝 기반 압축에 관한 관련 연구를 검토하고, 제3장에서는 제안하는 압축기의 구조를 상세히 기술한다. 제4장에서는 실험 결과를 통해 제안 기법의 성능을 검증한다. 마지막으로 제5장에서는 결론을 제시하고 향후 연구 방향을 논의한다.

II. 관련 연구

1. 고정 비트율 압축 기법

네트워크 대역폭·지연·패킷 손실 제약이 존재하는 실시간 멀티미디어 전송 환경에서는 고정 비트율(CBR) 운용이 전송의 예측 가능성과 지연 관리 측면에서 여전히 핵심적이다. 그러나 표준 비디오 코덱의 비트 생성 메커니즘은 장면 복잡도에 내재적으로 의존하므로, 실제 CBR 운용은 버퍼 모델과 레이트 컨트롤(RC)에 기반한 비트량 보정 기법을 필요로 한다. 이러한 방식에서는 뷰포인트 변화나 텍스처 복잡도 증가 시 QP(Quantization Parameter) 조정, 비트 재분배, GOP 구조 변화 등을 통해 목표 비트율을 지속적으로 추종한다.

MPEG-2 단일 패스 CBR 인코딩을 다룬 Mohsenian 등(1999)은, 부분적으로 인코딩된 스트림의 평균 복잡도 대비 개별 프레임의 상대 복잡도를 추정하여 단일 패스에서 압축 파라미터를 적응적으로 조정함으로써 CBR 제약 하에 품질을 향상시키는 방법을 제시하였다. 이들은 특히 방송·편집 환경에서 CBR 인코더의 매력을 강조하는데, 해당 맥락에서는 디스플레이와 정지(프리즈) 모드 모두에서 고충실도의 비디오 객체 표현이 지속적으로 요구되며, 일정 전송률 운용은 프로그램 멀티플렉싱과 스튜디오 내 시각 분석의 안정성에 직접적으로 기여한다(Mohsenian et al., 1999)[2].

슬라이스 기반 CBR 레이트 컨트롤을 제안한 Yao 등(2014)은, 표준 해상도 비디오에서 기존 코덱 시스템의 고유 지연(260ms 이상) 문제를 지적하고, 입력을 슬라이스로 세분하여 획득 지연을 축소한 뒤 슬라이스를 기본 단위로 코딩하는 구조를 제시하였다. 프레임/매크로블록 라인 단위의 비트 배분 전략을 통해 재구성 화질을 유지하면서도 시스템 지연을 약 100ms 수준으로 줄여 실시간 처리 요구를 만족함을 보였다(Yao et al., 2014). 이는 CBR 제약 하에서도 구조적(슬라이스) 설계와 RC 알고리즘의 결합으로 지연·버퍼 안정성·품질의 균형점을 실증한 사례다[12].

Stored B-picture 기반 저지연 RC를 제안한 Sanz-Rodríguez 등(2007)은, B-picture가 예측 효율을 높여 평균 비트율을 추가로 절감할 수 있으나, 프레임 재정렬 및 버퍼 평활화로 인한 지연 증가를 야기한다는 점을 지적한다. 이를 해결하기 위해 P와 Stored B-picture에 상이한 QP/MAD 모델을 두고, GOP 내 적절한 비트 배분을 유도하는 톱니형(saw-tooth) 목표 버퍼 레벨을 도입하여 버퍼 점유·목표 비트율 추종을 개선하였다. 결과적으로 참조 RC 대비 버퍼 안정성과 목표율 수렴성이 향상되었고, 대가로 경미한 화질 저하를 보고하였다(Sanz-Rodríguez et al., 2007). 즉, B-픽처의 효율과 저지연 운용 간 트레이드오프를 CBR 지향 RC로 완화한 연구다[3].

2. 딥러닝 기반 영상 압축

딥러닝 기반 영상 압축 연구는 2010년대 중반부터 본격

화되었다. Ballé 등은 변분 오토인코더(Variational Auto-Encoder)를 활용한 최초의 end-to-end 학습 가능한 영상 압축 시스템을 제안했다[5]. 이 연구는 기존 JPEG 대비 우수한 성능을 보여주며 딥러닝 압축의 가능성을 입증하였지만 posterior collapse 문제가 있었다. Van den Oord 등은 연속적인 잠재 공간의 posterior collapse 문제를 해결하기 위해 이산적 표현을 도입한 VQ-VAE를 제안했다[4]. 원래는 음성 합성과 이미지 생성을 목적으로 개발되었지만, 곧 압축 응용으로 확장되었다. 압축 분야에서는 상대적으로 최근에 ResNet 구조가 활용되기 시작했다. Toderici 등은 RNN 기반 압축 모델에서 잔여 연결을 통해 재구성 품질 향상을 달성했으며[7], Wang 등은 잔여 블록을 활용한 후처리 네트워크로 압축 아티팩트의 효과적인 제거가 가능함을 입증했다[14].

3. 활성화 함수와 하드웨어 최적화

활성화 함수는 딥러닝 모델의 성능을 결정하는 중요한 요소이다. ReLU[8] 도입 이후 Leaky ReLU, ELU, SiLU(Swish)[9,10] 등 다양한 변형들이 제안되었으며, 각각 고유한 장점을 통해 특정 응용에서 우수한 성능을 보였다. SiLU(Swish)는 부드러운 곡선 형태로 인해 기울기 흐름이 개선되어 깊은 네트워크에서 뛰어난 성능을 나타낸다[10]. 모바일 및 엣지 컴퓨팅의 중요성이 증대되면서 하드웨어 구현을 고려한 활성화 함수 연구가 활발해졌다. Howard 등은 MobileNet에서 복잡한 지수 함수를 선형 구간별 함수로 근사한 HardSwish와 Hardsigmoid를 개발하여 하드웨어 구현 비용을 크게 줄이면서도 성능을 유지하였다[11].

4. Batch Normalization 배치 순서

Batch Normalization의 위치는 딥러닝 모델 설계에서 지속적인 연구 주제이다. 원래 BN은 Ioffe와 Szegedy에 의해 Conv-BN-Activation 순서로 제안되었다[12]. 그러나 이후 연구들에서 다른 순서가 특정 상황에서 더 나은 성능을 보일 수 있음이 밝혀졌다. He 등은 ResNet의 후속 연구에서 BN-ReLU-Conv 순서(Pre-activation)를 제안하여 더 깊은 네트워크에서 안정적인 학습이 가능함을 보였다[15]. Xie 등은 다양한 정규화 기법과 활성화 함수의 조합을 체계적으로 분석했다[16]. 압축 응용에서는 정규화의 효과가 더욱 중요할 수 있다. 압축 과정에서 발생하는 정보 손실로 인해 학습이 불안정해질 수 있으며, 적절한 정규화는 이를 완화하는 데 도움이 된다.

Ⅲ. 제안 구조

1. 전체 아키텍처

본 논문에서 제안하는 4:1 고정 비트율 프레임 압축기는 Vector Quantization Variational AutoEncoder(VQ-VAE) 아키텍처를 기반으로 한다. VQ-VAE는 연속적인 잠재 공간 대신 이산적인 코드북을 사용하여 더 안정적인 학습과 효율적인 압축을 가능하게 하며[4], 특히 목표 비트율을 직접 제어할 수 있다는 장점을 가진다[5]. 제안하는 시스템은 인코더, 벡터 양자화기, 디코더로 구성되며, 공간적 다운샘플링 없이 벡터 양자화기의 코드북 인덱스를 6비트로 고정하여 정확한 4:1 압축비를 달성한다.

입력 RGB 프레임 $x \in \mathbb{R}^{H\times W\times 3}$에 대해 인코더 E_θ는 동일 해상도의 잠재 특징 $z \in \mathbb{R}^{H\times W\times C}$를 산출한다. 벡터 양자화기 Q는 각 위치 (h,w)의 연속 벡터 $z_{h,w}$를 코드북 $\mathcal{C} = \{e_k\}_{k=1}^{K}$의 인덱스 $i_{h,w} \in \{1,\dots,K\}$로 치환하여 인덱스 맵 $I \in \{1,\dots,K\}^{H\times W}$을 생성한다. 본 연구에서는 $K = 64$를 사용하므로 위치당 $\log_2 K = 6bit$가 필요하고 원본 영상이 픽셀당 24bit(RGB 8:8:8 구성)인 점을 고려하면, 제안 기법에서는 모든 픽셀에 정확히 6비트를 강제 할당함으로써 픽셀 단위에서 정확히 4:1의 비트 축소가 이루어진다. 이에 따라 프레임 전체의 출력 비트량 또한 항상 원본 대비 25%로 일정하게 유지되므로, 프레임 저장 및 전송 과정에서 정확한 고정 비트율을 보장할 수 있다. 디코더 $D_\emptyset$는 I를 임베딩으로 복원한 후 $\hat{x} = D_\emptyset(embed(I))$를 출력한다. 프레임당 비트 수는 $Bit_{frame} = H \times W \times \log_2 K$로 결정적이며 엔트로피 코딩에 의존하지 않으므로 비트스트림 길이가 입력 통계에 좌우되지 않고 프레임당 비트 수가 상수로 고정되어 지터, 버퍼 변동을 최소화하고 네트

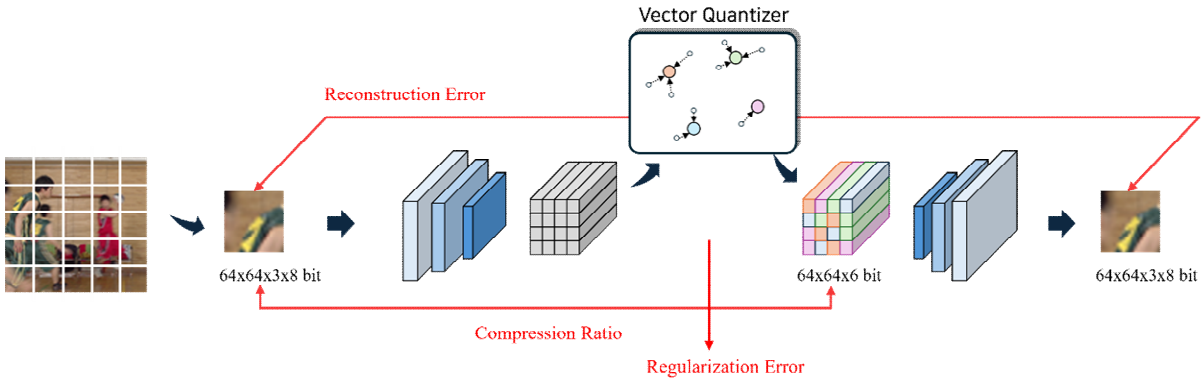


그림 1. 제안하는 전체 아키텍처
Fig. 1. Proposed architecture

워크 대역폭, 저장 용량, 지연 상한을 사전에 정확히 산출할 수 있다.

2. 개선된 Residual Block 구조

제안하는 압축기의 핵심 구성 요소는 개선된 Residual block이다. 잔여 학습 구조는 ResNet에서 처음 제안된 이후 딥러닝 모델의 훈련 안정성과 성능 향상에 중요한 역할을 해왔으며, 영상 처리 분야에서도 세밀한 디테일 보존에 기여하는 것으로 알려져 있다[6]. 본 연구에서는 기존의 Conv-BN-Activation 순서 대신 BN-Activation-Conv 순서를 채택하여 이른바 pre-activation 효과를 부여한다. 이는 정보 병목이 필연적인 압축 태스크에서 발생하는 학습 불안정과 기울기 소실 문제를 완화하고, VQ 단계로 전달되는 잠재 분포의 안정성을 높이기 위함이다.

구체적인 Residual block의 구조는 입력 특징 맵의 정규화를 통해 학습 안정성을 향상시키는 Batch Normalization, SiLU의 하드웨어 친화적 근사로서 높은 표현력과 구현 효율성을 동시에 제공하는 Hard swish[11] 활성화 함수, 그리고 특징 추출 및 변환을 수행하는 Convolution 연산의 순서로 구성된다. 이러한 pre-activation 구성은 두 가지 관점에서 압축 태스크와 정합적이다. 첫째, 프레임 압축은 본질적으로 정보 병목을 도입하기 때문에 과도한 비선형, 축소가 고주파 성분의 손실을 유발하기 쉽다. 제안 블록은 항등 경로를 통해 세부 구조를 보존하고 변환 경로는 의미 압축을 수행함으로써 정보 보존과 요약 사이의 균형을 유지한다. 둘째, 본 연구는 공간적 다운샘플링 없이 채널 차원 변환만으로 압축을 수행하므로 블록 수준의 안정적 변환이 곧 공간 정밀도 유지로 직결된다. 이는 후단 편집, 분석, 프레임 정렬성 요구가 높은 응용에서 유리하다. 또 본 블록 설계는 VQ 단계와의 상호작용 측면에서도 유리하다. VQ는 연속 잠재 z를 코드북 벡터에 최근접 할당하는 과정이며 이때 z의 스케일, 분포 안정성은 코드북 경쟁 균형과 시간적 일관성

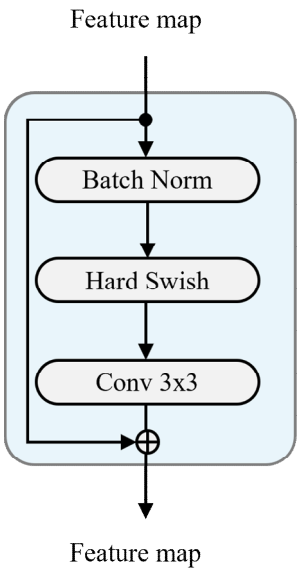


그림 2. 제안하는 residual block 구조
Fig. 2. Proposed residual block

을 좌우한다. 블록 선두의 BN은 z의 통계를 일정하게 유지하여 특정 코드의 과도한 점유로 인한 코드북 붕괴 경향을 억제한다. 학습 및 배포 효율성도 고려하였다. Hard Swish는 비선형부의 연산 복잡도를 낮추어 FPGA/엣지 환경에서 지연, 전력, 자원 제약을 만족시키는데 유리하며 프레임 단위 고정 비트율 전송과 결합되어 큐잉 지연과 버퍼 변동을 구조적으로 억제한다.

3. 인코더 구조

제안하는 압축기의 인코더는 입력 이미지를 벡터 양자화에 적합한 잠재 표현으로 변환하는 역할을 수행한다. 인코더는 초기 특징 추출부, 핵심 압축부, 최종 매핑부로 구성되어 점진적인 특징 압축과 표현 학습을 수행한다. 초기 특징 추출부는 단일 convolution 레이어로 구성되어 입력 RGB 이미지를 고차원 특징 공간으로 매핑한다. 이 과정에서 픽셀 단위의 원시 정보가 의미 있는 특징 표현으로 변환되며, 후속 압축 과정을 위한 기초 특징이 추출된다. 핵심 압축부에서는 3개의 개선된 residual 블록이 순차적으로 배치되어 특징의 압축과 추상화를 수행한다. 각 residual 블록은 앞서 설명한 BN-Hard_swish-Conv 구조를 따르며, 블록 간에는 배치 정규화와 Hard_swish 활성화 함수가 적용된 convolution 레이어가 삽입되어 안정적인 특징 변환을 보장한다. 이러한 구성을 통해 인코더는 원본 이미지의 중요한 의미 정보를 보존하면서도 효율적인 압축을 달성한다. 최종 매핑부는 압축된 특징을 벡터 양자화기에 적합한 형태로 변환하는 역할을 담당한다. Hard_swish 활성화 함수를 통해 최종 특징 변환이 수행되며, 이를 통해 생성된 잠재 표현은 벡터 양자화기로 전달되어 이산적인 코드북 벡터로 매핑된다. 전체 인코더 구조는 공간적 다운샘플링 없이 채널 차원에서의 특징 변환에 집중하여 세밀한 공간 정보를 보존하면서도 재구성에 필요한 핵심 특징만을 효율적으로 표현하도록 인코더를 설계하였다.

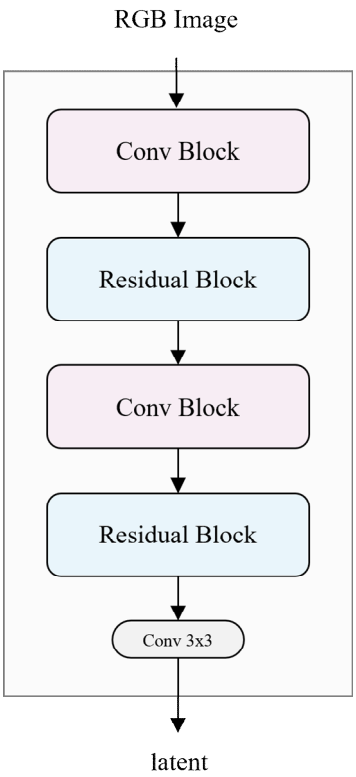


그림 3. 제안하는 인코더 구조
Fig. 3. Proposed encoder

4. 벡터 양자화기

벡터 양자화기는 인코더에서 출력된 연속적인 잠재 표현을 이산적인 코드북 벡터로 매핑하는 핵심 구성 요소이다. 제안하는 벡터 양자화기는 고정 비트율을 엄격히 보장하기 위해 크기 K=64의 임베딩 벡터로 구성된 코드북을 사용하여 코드북 인덱스는 $\log_2 K = 6$비트로 표현되므로 토큰당 전송 비트 수가 고정되어 비트율이 일정하게 유지된다. 각 임베딩 벡터는 256차원으로 설정하였다. 임베딩 차원은 표현력과 학습 ·메모리 비용 간의 고전적인 트레이드오프를 형성하며, 차원이 너무 작으면 재구성 품질이 저하되고 지나치게 크면 학습 ·메모리 비용이 급증하기 때문에 코드북 크기와 임베딩 차원의 균형을 적절히 조정해야 한다. 이와 관련해서는 기존 벡터 양자화 기반 이미지 표현 연구들에서 표현력과 연산 복잡도 간의 균형점으로 검증된 구성을 참조하였다[4]. 이를 통해 정확한 4:1 압축비를 달성하고 고정 비트율 압축의 핵심 요구사항인 일정한 압축 성능을 보장한다. 양자화 과정에서는 입력 특징과 코드북 벡터 간의 유클리드 거리를 계산하여 가장 가까운 벡터를 선택하는 방식을 사용한다. 거리 계산은 효율적인 행렬 연산을 통해

수행되며, 벡터 양자화 과정에서 발생하는 손실은 양자화 손실과 commitment 손실로 구성된다. Commitment cost는 0.25로 설정하여 인코더가 코드북에 맞춰 학습되도록 균형을 맞추며, 양자화 과정에서 발생하는 미분 불가능성 문제를 해결하기 위해 straight-through estimator를 사용하여 순전파에서는 양자화된 값을 사용하고 역전파에서는 기울기를 직접 전달함으로써 end-to-end 학습을 가능하게 한다. 코드북의 모든 벡터가 효율적으로 사용되도록 하기 위해 사용량 추적 메커니즘을 도입하였다. 코드북은 훈련 시 전체 시퀀스에 공통적으로 사용하며 연속으로 20회 이상 사용되지 않은 코드북 벡터는 자동으로 새로운 입력 특징으로 교체되어 코드북의 표현력을 유지하며, 이는 코드북 붕괴 문제를 효과적으로 방지하고 안정적인 학습을 보장한다. 이렇게 학습된 단일 코드북을 추론 시 전체 시퀀스에 공통적으로 사용하였다. 이러한 설계를 통해 제안하는 벡터 양자화기는 목표 압축 효율과 안정적인 학습 성능을 동시에 달성할 수 있다.

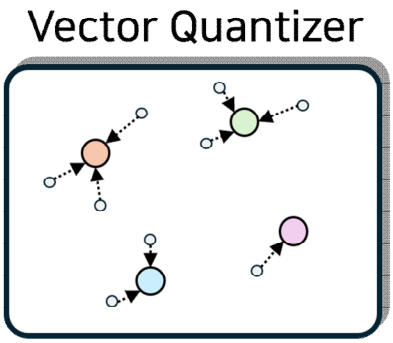


그림 4. 벡터 양자화기
Fig. 4. Vector Quantizer

5. 디코더 구조

디코더는 벡터 양자화된 잠재 표현을 원본 이미지로 복원하는 역할을 수행한다. 그림 5에 나타낸 바와 같이, 디코더는 인코더와 대칭적인 구조로 설계되어 압축된 정보만으로도 고품질 복원이 가능하도록 구성하였다. 초기 복원부는 단일 Conv 레이어로 구성되어 벡터 양자화기로부터 전달받은 코드북 인덱스를 연속적인 특징 맵으로 변환하고 양자화 과정에서 손실된 정보를 보완한다. 핵심 복원부에서는 3개의 residual 블록이 인코더와 역순으로 배치되어 점진적인 특징 복원을 수행한다. 각 residual 블록 사이에 삽입된 BN-Hard swish-Conv 레이어는 배치 정규화를 통한 안정적인 학습과 Hard swish 활성화 함수를 통한 비선형적 표현력을 동시에 확보한다. 이러한 구성은 압축 과정에서 손실된 세부 정보를 단계적으로 복원하며, 잔여 연결을 통해 중요한 특징 정보의 전파를 보장한다. 최종 복원부는 Hard Sigmoid 활성화 함수로 복원된 특징을 최종 이미지 픽셀 값으로 변환한다. 이러한 구조를 통해 디코더는 압축된 잠재 표현으로부터 원본과 시각적으로 구별하기 어려운 수준의 고품질 이미지 복원을 달성한다. 추론 시 디코더는 학습된 코드북을 모델 가중치에 포함된 상태로 사용한다. 따라서 복원에 필요한 모든 임베딩 벡터가 로컬에 존재하며, 전송되는 정보는 오직 코드 인덱스뿐이다. 이는 추가적인 코드북 전송이나 동적 업데이트 없이도 정확한 4:1 고정 비트율과 예측 가능한 메모리 사용량을 유지할 수 있다는 장점을 제공한다.

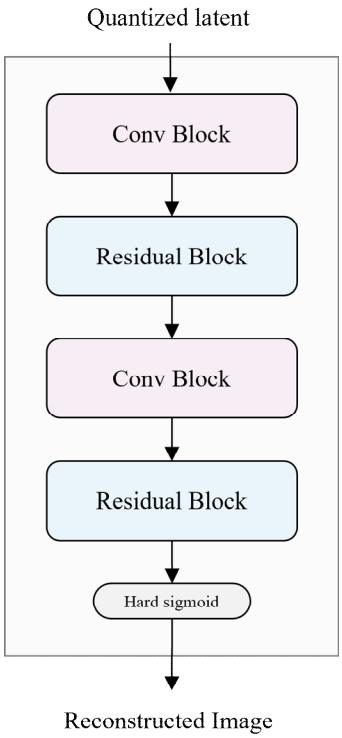


그림 5. 제안하는 디코더 구조
Fig. 5. Proposed decoder

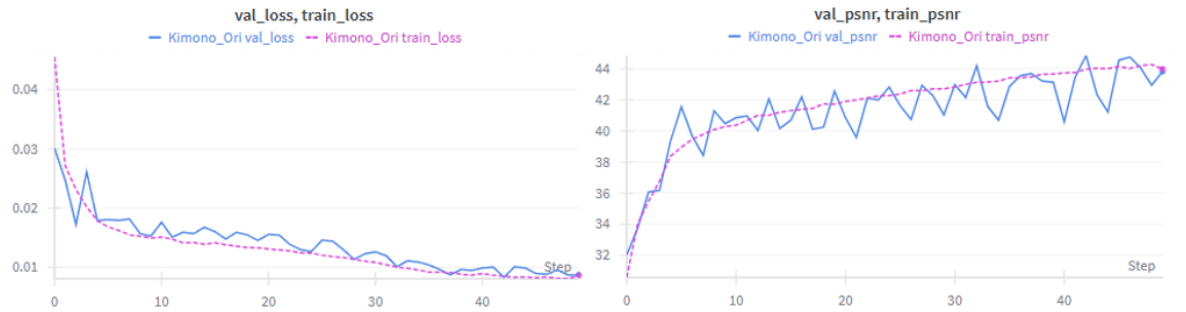
Ⅳ. 실험 결과

1. 실험 설정

제안된 VQ-VAE 아키텍처의 성능을 검증하기 위해 고해상도 이미지 데이터셋인 HEVC-B를 사용하여 실험을 수행하였다. 실험 데이터는 1920×1080 해상도의 이미지를 64×64 픽셀 크기의 세그먼트로 분할하여 총 120,000개의 세그먼트를 생성하였다. 학습은 NVIDIA GeForce RTX 4070Ti에서 배치 크기 64, AdamW 옵티마이저를 적용한 학습률 1e-4, 총 50 에폭으로 설정하여 진행하였다.

2. 학습 안정성 분석

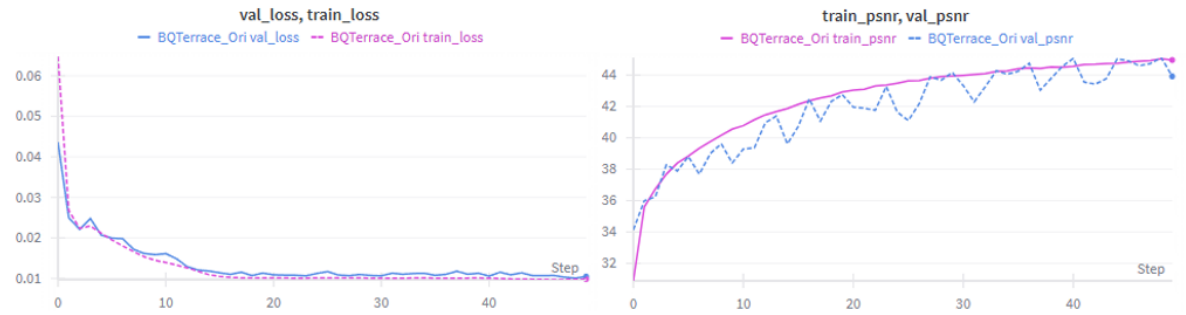
그림 6-(a)와 6-(b)에서 보는 바와 같이 학습 과정에서 train loss와 validation loss가 유사한 패턴으로 수렴하며 두 손실 간의 차이가 적게 나타났다. 이는 모델이 과적합 없이 안정적으로 학습되었음을 의미하며, 제안된 아키텍처의 일반화 성능이 우수함을 보여준다. 학습 곡선의 안정적인 수렴 패턴은 제안된 아키텍처의 핵심 구성 요소들이 네트워크 최적화에 효과적으로 기여했음을 시사한다. 인코더와 디코더에 각각 3개씩 배치된 residual 블록은 잔여 연결을 통해 깊은 네트워크에서 발생할 수 있는 기울기 소실 문제



(a) (b)

그림 6. (a) Kimono sequence validation PSNR과 training PSNR을 비교한 그래프 (b) Kimono sequence validation loss와 training loss를 비교한 그래프

Fig. 6. (a) Comparison graph of Kimono sequence validation PSNR and training PSNR (b) Comparison graph of Kimono sequence validation loss and training loss



(a) (b)

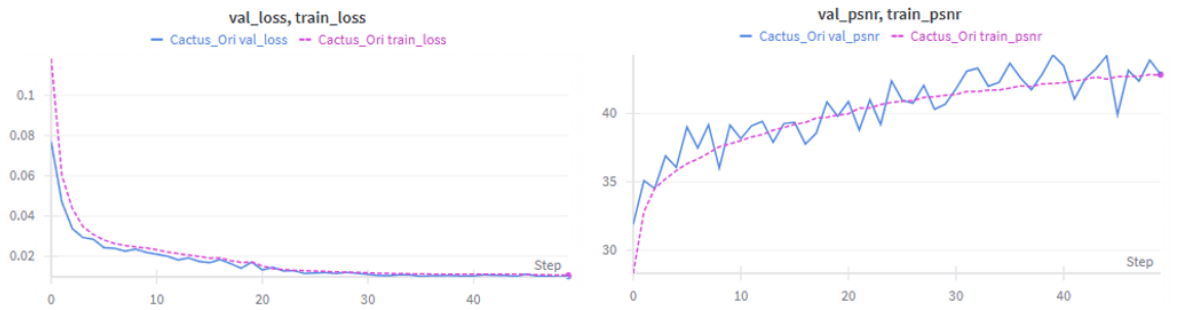
그림 7. (a) BQTerrace sequence validation PSNR과 training PSNR을 비교한 그래프 (b) BQTerrace sequence validation loss와 training loss를 비교한 그래프

Fig. 7. (a) Comparison graph of BQTerrace sequence validation PSNR and training PSNR (b) Comparison graph of BQTerrace sequence validation loss and training loss

(a) (b)

그림 8. (a) Basketball sequence validation PSNR과 training PSNR을 비교한 그래프 (b) Basketball sequence validation loss와 training loss를 비교한 그래프

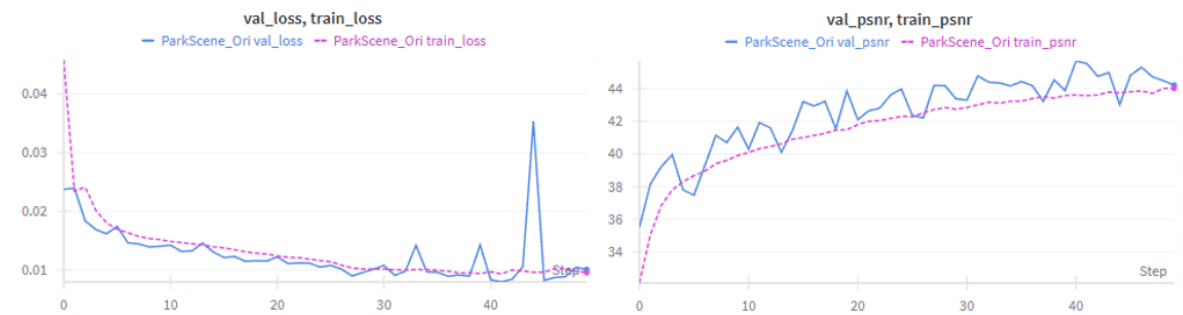
Fig. 8. (a) Comparison graph of Basketball sequence validation PSNR and training PSNR (b) Comparison graph of Basketball sequence validation loss and training loss



(a) (b)

그림 9. (a) Cactus sequence validation PSNR과 training PSNR을 비교한 그래프 (b) Cactus sequence validation loss와 training loss를 비교한 그래프

Fig. 9. (a) Comparison graph of Cactus sequence validation PSNR and training PSNR (b) Comparison graph of Cactus sequence validation loss and training loss



(a) (b)

그림 10. (a) ParkScene sequence validation PSNR과 training PSNR을 비교한 그래프 (b) ParkScene sequence validation loss와 training loss를 비교한 그래프

Fig. 10. (a) Comparison graph of ParkScene sequence validation PSNR and training PSNR (b) Comparison graph of ParkScene sequence validation loss and training loss

를 해결하여 안정적인 학습을 가능하게 했으며, 입력 특징과 변환된 특징을 직접 결합함으로써 정보 손실을 최소화했다. Conv-BN-Hard swish와 BN-Hard swish-Conv 순서로 구성된 배치 정규화는 각 레이어의 입력 분포를 안정화하여 압축 과정에서 발생하는 정보 손실로 인한 학습 불안정성을 효과적으로 해결했다. Hard swish 활성화 함수는 하드웨어 친화적 특성을 유지하면서도 부드러운 비선형성을 제공하여 복잡한 이미지 패턴의 세밀한 특징 표현을 가능하게 하여 전체적으로 안정적인 학습 환경을 조성했다.

3. 정량적 성능 평가

성능 지표 측면에서 학습 진행에 따른 단계별 성능 향상을 확인할 수 있었다. 학습 초기에는 양자화기와 디코더가 충분히 적응하지 못해 색 번짐과 윤곽 흐림이 두드러진다. 1 epoch 기준으로 성능은 HEVC-B의 모든 영상 시퀀스에 대해 평균 PSNR: 34.53dB, 평균 SSIM: 0.96 수준으로 관찰된다. 10 epoch부터는 pre-activation residual 구조가 안정적인 기울기 흐름을 제공하면서 엣지 선명도와 색 정확도가 뚜렷이 개선된다. 특히 미세 질감과 고명암 대비 경계에서의 ringing/banding이 현저히 감소한다. 장기간 미사용 엔트리를 재활성화하여 50 epoch이 가까워질수록 표현력 유지와 시간적 일관성을 확보한다. 잔존하던 저주파 색 편향과 고주파 과보정이 완화되며 세부 질감이 안정화된다. 최종 성능은 평균 PSNR: 43.92dB, 평균 SSIM: 0.99 수준으로 관찰된다.

4. 시각적 품질 평가

그림 8의 시각적 품질 평가 결과는 정량적 성능 평가 지

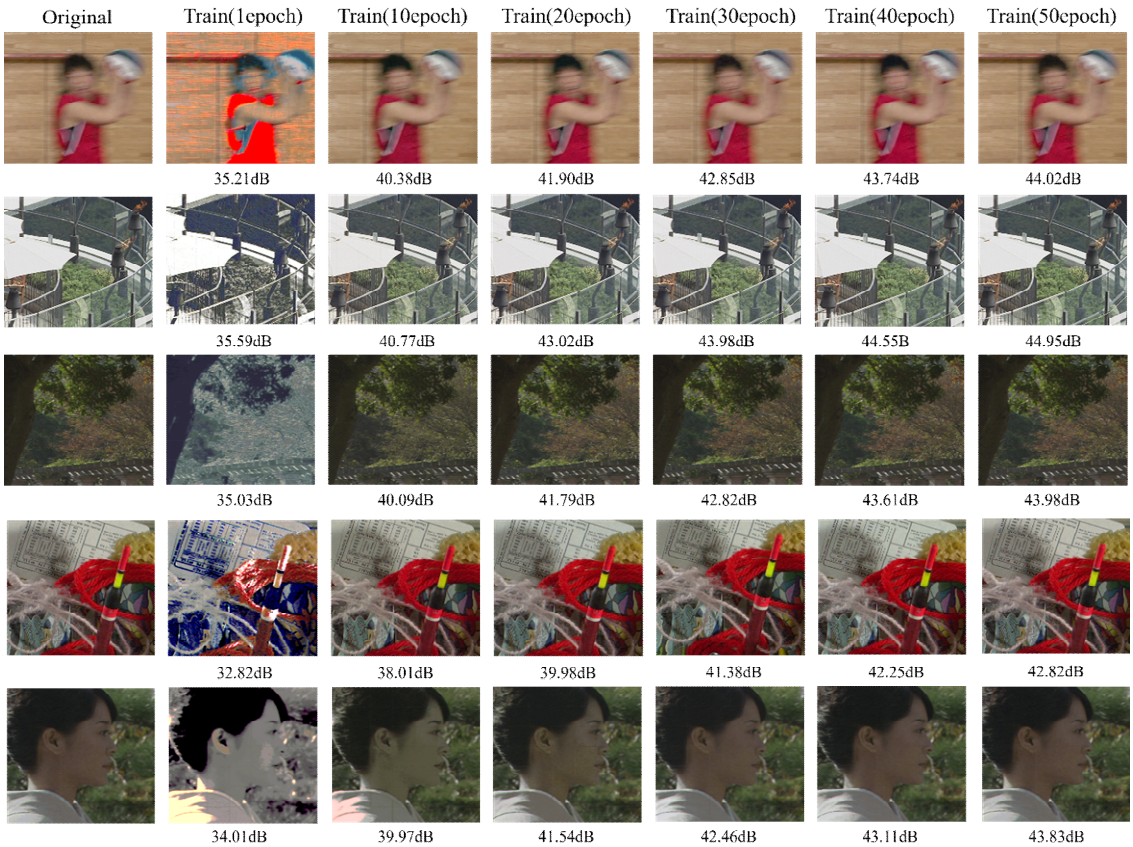


그림 11. 학습 결과
Fig. 11. Train result

표와 일관된 경향을 보인다. 먼저 학습, 추론 모두 40dB 이상의 PSNR 구간에서는 원본 대비 복원 영상의 차이를 관찰자 수준에서 식별하기 어렵다. 세부 구조 보존 측면에서, 실내 체육관의 목재 패턴, 카페 난간과 테라스의 미세 구조, 직물과 실타래의 반복 질감, 수목과 잎의 고주파 성분 등 복잡한 텍스처에서 과평활이나 노이즈 증폭 없이 미세 대비가 유지된다. 이는 다운샘플링 없이 채널 변환 중심으로 설계된 인코더와 코드북 활용 균등화 덕분에 고주파 성분의 선택적 표현력이 유지된 결과로 해석된다. 고명암 대비 경계에서는 ringing과 halo가 억제되어 자연스러운 edge 전이가 관찰된다. 특히 금속 난간과 같은 주기적 패턴에서 aliasing 징후가 두드러지지 않으며 이는 벡터 양자화 기반 표현이 블록 변환 기반에서 흔한 blocking 아티팩트를 야기하지 않기 때문이다. 색 재현 관점에서는 빨간 직물과 피부톤에서 색상 편향이나 chroma bleed가 두드러지지 않고 명암 변화에 다른 색상 이동이 제한적이다. 이는 학습 중 Hard-Swish 기반의 pre-activation residual block이 기울기 흐름을 안정화하여 채널 간 상관 구조를 과도하게 왜곡하지 않도록 작동한 결과와 합치한다.

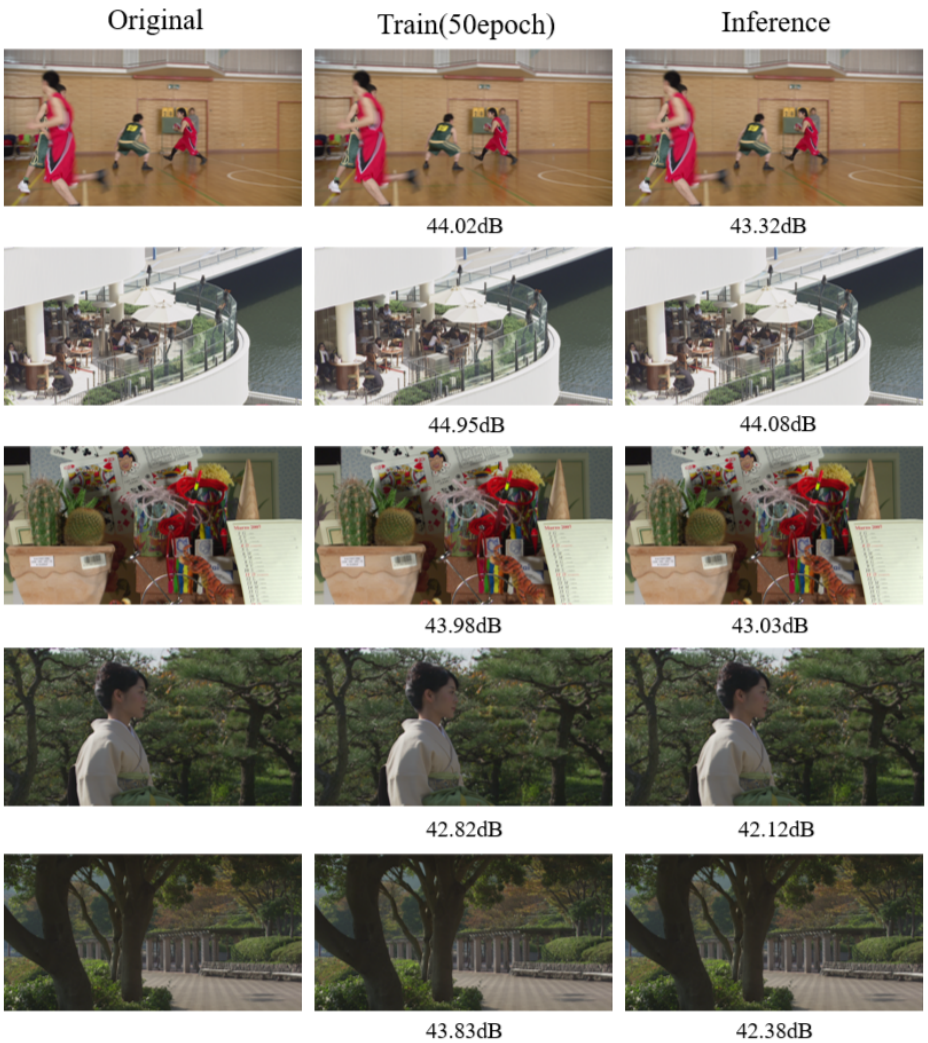


그림 12. 시각적 비교 결과
Fig. 12. Visual comparison of results

Ⅴ. 결 론

본 연구에서는 VQ-VAE 기반의 4:1 고정 비트 할당 방식 프레임 압축기를 제안하고 그 성능을 실험적으로 검증하였다. 제안된 방법은 인코더 – 벡터 양자화기 – 디코더 구조를 기반으로 하며, 공간적 다운샘플링 없이 코드북 인덱스를 6비트로 고정하여 모든 픽셀에 동일한 비트를 할당함으로써 정확한 4:1 고정 비트율을 보장한다. HEVC-B 데이터셋의 모든 영상 시퀀스에 대해 50 에폭 학습을 수행한 결과, 평균 PSNR 43.92dB를 달성하였으며, 이는 일반적으로 원본 대비 시각적으로 구별이 어려운 수준인 40dB 이상의 복원 품질을 안정적으로 확보했음을 의미한다. 또한 학습 과정에서 train loss와 validation loss가 안정적으로 수렴하는 양상을 보여, 제안된 아키텍처의 우수한 일반화 성능을 확인하였다. 복잡한 텍스처나 경계 영역에서도 자연스러운 복원 결과를 보이며, 실용적 적용 가능성 또한 입증하였다.

본 연구의 핵심 기여는 복잡한 레이트 컨트롤이나 엔트로피 코딩 없이 코드북 기반 임베딩 표현만으로 정해진 비트 예산을 정확히 준수한다는 점에 있다. 이를 통해 일정한 비트율이 요구되는 하드웨어 기반 프레임 저장 및 파이프라인 내부 데이터 전송 환경에서 높은 활용 가능성을 제시한다.

향후에는 본 연구의 하드웨어 친화적 설계 철학을 바탕으로 제안 구조를 FPGA 기반 실시간 구현으로 확장하여, 자원 소모(resource usage), 지연(latency), 전력 효율성(power efficiency) 등의 지표를 종합적으로 분석할 예정이다. 또한, 보다 높은 압축비에서도 복원 품질을 유지하기 위한 코드북 및 임베딩 구조 최적화, 그리고 다양한 해상도와 비디오 도메인에 대한 일반화 성능 확장 연구를 병행할 계획이다.

따라서 제안된 방법은 일정한 대역폭 사용량과 예측 가능한 저장 용량이 요구되는 실시간 스트리밍, 비디오 편집, 클라우드 기반 미디어 서비스 등의 응용 분야에서 효과적으로 활용될 수 있을 것으로 기대된다.

참 고 문 헌 (References)

[1] ITU-T Recommendation H.264, “Advanced video coding for generic audiovisual services,” 2003, https://www.itu.int/rec/T-REC-H.264-200305-S/en

[2] Mohsenian, N., Rajagopalan, R., and Gonzales, C. A., “Single-pass constant- and variable-bit-rate MPEG-2 video compression,” IBM Journal of Research and Development, vol. 43, no. 4, pp. 489 – 509, 1999, https://ieeexplore.ieee.org/document/5389228
doi: https://doi.org/10.1147/rd.434.0489

[3] Sanz-Rodríguez, S., et al., “A rate control algorithm for low-delay H.264 video coding with stored-B pictures,” in 2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. I-1153 – I-1156, 2007, https://ieeexplore.ieee.org/document/4217289
doi: https://doi.org/10.1109/ICASSP.2007.366117

[4] Van Den Oord, A., Vinyals, O., and Kavukcuoglu, K., “Neural discrete representation learning,” Advances in neural information processing systems, Vol. 30, 2017, https://arxiv.org/abs/1711.00937
doi: https://doi.org/10.48550/arXiv.1711.00937

[5] Ballé, J., Laparra, V., and Simoncelli, E. P., “End-to-end optimized image compression,” International Conference on Learning Representations (ICLR), 2017, https://arxiv.org/abs/1611.01704
doi: https://doi.org/10.48550/arXiv.1611.01704

[6] He, K., Zhang, X., Ren, S., and Sun, J., “Deep residual learning for image recognition,” Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 770-778, June 2016, https://ieeexplore.ieee.org/document/7780459
doi: https://doi.org/10.1109/CVPR.2016.90

[7] Toderici, G., Vincent, D., Johnston, N., Hwang, S. J., Minnen, D., Shor, J., and Covell, M., “Full resolution image compression with recurrent neural networks,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5435-5443, July 2017, https://ieeexplore.ieee.org/document/8100060
doi: https://doi.org/10.1109/CVPR.2017.577

[8] Nair, V. and Hinton, G. E., “Rectified linear units improve restricted boltzmann machines,” Proceedings of the 27th international conference on machine learning (ICML), pp. 807-814, 2010, https://www.cs.toronto.edu/~fritz/absps/reluICML.pdf

[9] Maas, A. L., Hannun, A. Y., and Ng, A. Y., “Rectifier nonlinearities improve neural network acoustic models,” Proceedings of ICML, Vol. 30, No. 1, p. 3, 2013, https://www.semanticscholar.org/paper/Rectifier-Nonlinearities-Improve-Neural-Network-Maas-Hannun/367f2c63a6f6a10b3b64b8729d601e69337ee3cc

[10] Elfwing, S., Uchibe, E., and Doya, K., “Sigmoid-weighted linear units for neural network function approximation in reinforcement learning,” Neural Networks, vol. 107, pp. 3-11, 2018, https://www.sciencedirect.com/science/article/pii/S0893608017302976
doi: https://doi.org/10.1016/j.neunet.2017.12.012

[11] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., and Adam, H., “Searching for MobileNetV3,” Proceedings of the IEEE/CVF

International Conference on Computer Vision (ICCV), pp. 1314-1324, October 2019, https://ieeexplore.ieee.org/document/9008835/
doi: https://doi.org/10.1109/ICCV.2019.00140

[12] Ioffe, S. and Szegedy, C., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," International conference on machine learning, pp. 448-456, 2015, https://proceedings.mlr.press/v37/ioffe15.html

[13] Yao, C., et al., "A rate control algorithm for low delay video coding," Open Cybernetics & Systemics Journal, vol. 8, pp. 773 - 778, 2014, https://benthamopenarchives.com/abstract.php?ArticleCode=TOCSJ-8-773
doi: https://doi.org/10.2174/1874110X01408010773

[14] Wang, Z., Liu, D., Yang, J., Han, W., and Huang, T., "Deep networks for image super-resolution with sparse prior," Proceedings of the IEEE international conference on computer vision, pp. 370-378, 2015, https://openaccess.thecvf.com/content_iccv_2015/html/Wang_Deep_Networks_for_ICCV_2015_paper.html
doi: https://doi.org/10.1109/ICCV.2015.50

[15] He, K., Zhang, X., Ren, S., and Sun, J., "Identity mappings in deep residual networks," European conference on computer vision, pp. 630-645, 2016, https://link.springer.com/chapter/10.1007/978-3-319-46493-0_38
doi: https://doi.org/10.1007/978-3-319-46493-0_38

[16] Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K., "Aggregated residual transformations for deep neural networks," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1492-1500, 2017, https://openaccess.thecvf.com/content_cvpr_2017/html/Xie_Aggregated_Residual_Transformations_CVPR_2017_paper.html
doi: https://doi.org/10.1109/CVPR.2017.634

저 자 소 개

이 혜 린

- 2022년 2월~ 현재 : 광운대학교 전자융합공학과 재학
- ORCID : https://orcid.org/0009-0008-3657-1565
- 주관심분야 : 하드웨어 설계, 딥러닝, 영상 신호처리

허 정 윤

- 2022년 2월~ 현재 : 광운대학교 전자융합공학과 재학
- ORCID : https://orcid.org/0009-0007-9629-980X
- 주관심분야 : 전자공학, 정보보안, 임베디드 시스템, 기계학습, 뉴럴 네트워크

송 성 운

- 2022년 2월~ 현재 : 광운대학교 전자융합공학과 재학
- ORCID : https://orcid.org/0009-0008-4437-1029
- 주관심분야 : 하드웨어 설계, 딥러닝, 영상 코덱

저 자 소 개

이 학 범

- 2024년 2월 : 광운대학교 전자재료공학과학과 졸업(공학사)
- 2024년 3월 ~ 현재 : 광운대학교 전자재료공학과 일반대학원(석사과정)
- ORCID : https://orcid.org/0000-0003-0721-4944
- 주관심분야 : 멀티뷰 카메라 캘리브레이션, 3D 인체 복원, 컴퓨터 비전, 딥러닝

서 영 호

- 1999년 2월 : 광운대학교 전자재료공학과 졸업(공학사)
- 2001년 2월 : 광운대학교 일반대학원 졸업(공학석사)
- 2004년 8월 : 광운대학교 일반대학원 졸업(공학박사)
- 2005년 9월 ~ 2008년 2월 : 한성대학교 조교수
- 2008년 3월 ~ 현재 : 광운대학교 전자재료공학과 교수
- ORCID : https://orcid.org/0000-0003-1046-395X
- 주관심분야 : 실감미디어, 2D/3D 영상 신호처리, SoC 설계, 디지털 홀로그램