다중 스케일 특징 융합 모듈을 통한 종단 간 학습기반 공간적 스케일러블 영상 압축

*신주연 **강제원
전자전기공학과 스마트팩토리융합전공 +
이화여자대학교
*wndus2158g@ewhain.net **jewonk@ewha.ac.kr

End-to-End Learning-based Spatial Scalable Image Compression with Multi-scale Feature Fusion Module

*Shin, Juyeon **Kang, Jewon
Electronic and Electrical Department and Graduate Program in Smart Factory,
Ewha Womans University

요약

최근 기존의 영상 압축 파이프라인 대신 신경망의 종단 간 학습을 통해 압축을 수행하는 알고리즘의 연구가 활발히 진행되고 있다. 본 논문은 종단 간 학습 기반 공간적 스케일러블 압축 기술을 제안한다. 보다 구체적으로 본 논문은 신경망의 각 계층에서 하위 계층의 학습된 특징 (feature)을 융합하여 상위 계층으로 전달하는 다중 스케일 특징 융합 (multi-scale feature fusion) 모듈을 도입해 상위 계층이 더욱 풍부한 특징 정보를 학습하고 계층 사이의 특징 중복성을 더욱 잘 제거할 수 있도록 한다. 기존 방법 대비 향상 계층(enhancement layer)에서 1.37%의 BD-rate가 향상된 결과를 볼 수 있다.

1. 서론

최근 기존 영상 압축 파이프라인 대신 신경망의 종단 간 학습을 통해 압축을 수행하는 알고리즘의 연구가 활발하게 진행되고 있다. 기존 압축 기술은 여러 단계로 구성된 내부 모듈 사이 합동 최적화가 어려운 문제가 있는 반면, 종단 간 학습기반 영상 압축 기술은 이러한 문제를 극복하며 우수한 성능을 제공하고 있다.

공간적 스케일러블 영상 압축 기술은 네트워크의 대역폭 및 디바이스 사양의 제약에 따라 영상의 공간적 품질을 선별적으로 복원하도록 비트스트림을 선별적으로 생성한다 [1,2]. 종단 간 학습기반 영상 압축 기술에서 기존 연구[3,4]는 공간적 스케일러블 기능을 제공하기 위하여 계층 사이 예측을 수행하고, 남은 잔차 신호를 학습하였다. 그러나 이러한 잔차 신호는 상위 계층에서의 예측 정확도를 고려하지 않아 부호화 성능이 저하되는 문제가 있다.

본 논문에서는 영상 신호의 특징을 계층적으로 학습하는 과정에서 상위 계층이 하위 계층의 중간 특징 (feature)을 사용할 수 있도록 다중 스케일 특징 융합 (multi-scale feature fusion) 모듈을 도입하는 것을 제안한다. 다중 스케일 특징 융합 모듈을 통해 향상 계층 (enhancement layer)는 더욱 풍부해진 정보를 가질 뿐 아니라 참조 계층 사이 특징 정보 간 연관성이 커져 보다 우수한 부호화 성능을 제공할 수 있게 된다.

This work is supported by the NRF grant funded by MSIT (No.NRF-2022R1A2C4002052).

2. 제안 방법

가. 모델 전체 구조

그림 1은 본 논문이 제안하는 전체 모델 구조이다. 기존 연구[3]에선 잠재 특징 수준의 잔차를 압축하는 방법뿐만 아니라 하위 계층의 복원된 특징 맵을 향상 계층의 복원 특징 맵을 연결한 뒤 복호화를 하여 향상된 성능을 보였다. 그러나, 하위 계층의 복원된 특징 맵을 재사용하는 방식은 디코더의 복잡도를 크게 늘린다는 한계점이 있다. 본 논문에선 [3]의 잠재 특징의 잔차를 압축하는 방식을 채택하여 제안 모델을 발전시켰다.

그림1에서 각 계층 l은 부호화 과정을 통해 입력 영상 x_l에 대한 특징 맵 y_l을 생성한다. 이때, i번째 향상 계층 l_{ei}은 하위 계층의 특징을 선별적으로 참고하는 다중 스케일 특징 융합 모듈을 통해 압축에 우수한 특징을 갖게 된다. 계층별로 특징 맵이 생성된 후엔, 향상 계층 l_{ei}은 하위 계층의 특징을 예측하는 모듈을 통해 잔차 특징 맵 u_{ei}를 만든다. 특

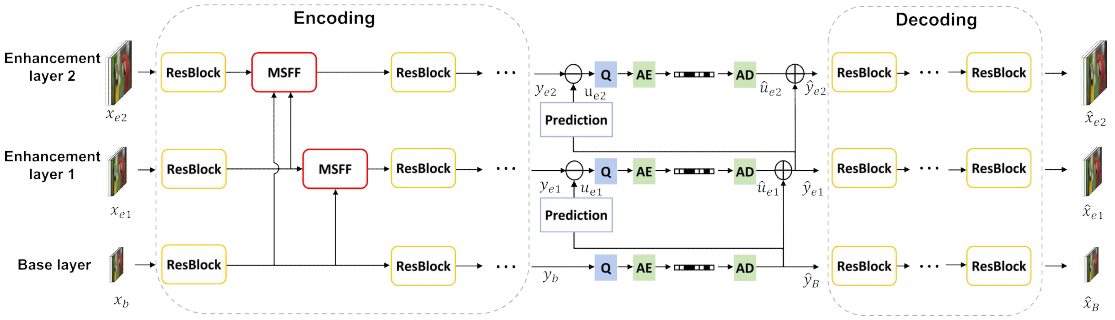

그림 1 전체 모델 구조

징 맵을 양자화한 $\hat{y}_b$, $\hat{u}_{ei}$은 비트 스트림 R_l로 추출하기 위하여 산술 부호화가 적용된다. 산술 복호화된 특징은 다시 영상 도메인으로 복원된다.

모델 전체를 학습하기 위한 목적함수는 다음과 같다. 기본 계층과 향상 계층의 학습 목적함수는 왜곡 D_l과 비트 레이트 R_l 의 최적화 문제로 정의하였다. 왜곡과 비트 레이트 비용 사이의 관계를 표현하는 라그랑주 승수는 λ_l로 표기한다. 최종 손실은 기본 계층의 손실 L_B와 Z개의 향상 계층의 손실 L_e의 합으로 정의하여 모델을 학습했다.

$$L_B = D_B + \lambda_B R_B, \quad (1)$$

$$L_{ei} = D_{ei} + \lambda_{ei} R_{ei}, \quad (2)$$

$$L = L_B + \sum_{i=1}^{Z} L_{ei}. \quad (3)$$

이때 원본 영상 x_l과 복원된 영상 $\hat{x}_l$ 사이에 발생하는 왜곡 D_l은 평균 제곱 오차(Mean squared error)로 정의한다. 비트 레이트 R_l는 양자화된 특징 맵의 엔트로피로 정의한다. i번째 향상 계층의 비트 레이트는 $i-1$번째 하위 계층까지의 비트 레이트와 i번째 잔차 특징 맵 $\hat{u}_{ei}$의 비트 레이트를 더한 값으로 정한다.

$$D_l = MSE(x_l, \hat{x}_l), \quad (4)$$

$$R_B = -E_{\hat{y}_B}[\log_2(\hat{y}_B)], \quad (5)$$

$$R_{ei} = -E_{\hat{u}_{ei}}[\log_2(\hat{u}_{ei})] + \sum_{t=1}^{i-1} R_{et} + R_B. \quad (6)$$

나. 다중 스케일 특징 융합 모듈

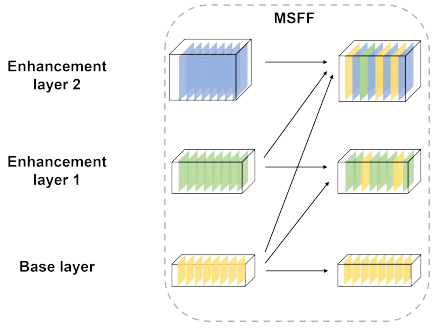

그림 2. 다중 스케일 특징 융합 모듈

그림 2에서 부호화 중 하위 계층의 특징을 선별적으로 참고하는 다중 스케일 특징 융합 모듈을 볼 수 있다. 이 모듈은 상위 계층에서 영상을 부호화할 때, 하위 계층에서 압축 효율에 중요한 정보를 주는 특징 맵의 채널별로 참고하여 융합하도록 한다. 영상 도메인에서 특징 도메인으로 변환 과정 중에 압축에 중요한 정보를 가진 일부 채널을 하위 계층에서 관측하게 되면, 향상 계층은 더욱 풍부한 표현을 갖게 되고, 기본 계층의 특징과 부분적으로 유사하게 되어 예측 과정에서 중복성이 잘 제거된다. 실험 결과에서 향상 계층 압축 효율이 향상됨을 볼 수 있다.

3. 실험 결과 분석

GeForce RTX 3090 12G 2개를 사용하여 pytorch와 pytorch library인 compressAI[5]에서 실험 환경을 구성했다. 학습 데이터셋은 Flickr2K 데이터셋, 테스트 셋은 Kodak 데이터셋을 사용하였다. 기본 계층와 향상 계층의 입력으로 각각 128x128, 256x256 패치 8개로 넣었다. 또한, RD 곡선의 한 포인트를 결정짓는 하이퍼 파라미터는 표1의 lambda 값과 특징 맵 개수에서 볼 수 있다.

특징 맵 개수	Lambda $\lambda_B = \lambda_{e1}$
128	0.0016
128	0.0032
192	0.015
192	0.045

표1. RD 곡선의 한 포인트를 결정짓는 특징 맵 개수와 Lambda 값

그림 3-4에선 다중 스케일 특징 융합 모듈과 예측 모듈이 있는 모델, 다중 스케일 특징 융합 모듈을 넣지 않고 예측 모듈만 있는 모델, 그리고 예측 모듈이 없는 Simulcast 종단 간 학습 모델 총 3가지의 RD curve 결과를 볼 수 있다. 그림 3-4와 표 2에서 볼 수 있듯이, 본 논문이 제안하는 다중 스케일 특징 융합 모듈이 있는 모델이 다른 모델 성능에 비해 좋은 것을 확인할 수 있다.

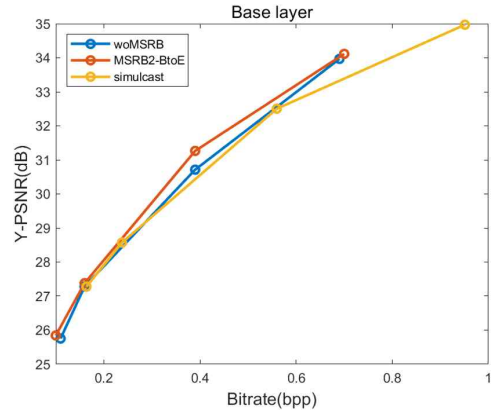

그림 3 기본 계층에서의 RD 곡선 비교

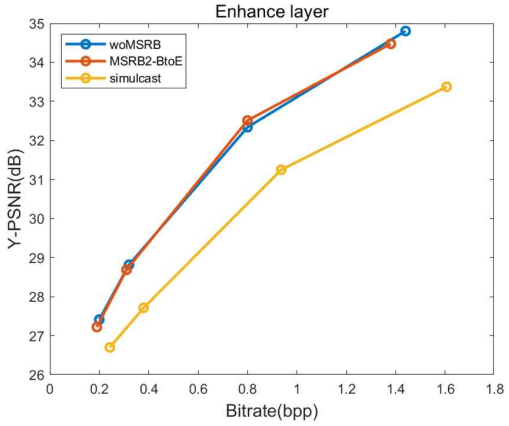

그림 4 향상 계층에서의 RD 곡선 비교

	기본 계층		향상 계층 1	
제안 모델 vs	BD-PSNR (dB)	BD-Rate (%)	BD-PSNR (dB)	BD-Rate (%)
wo MSRB [3]	0.321	-7.390	0.055	-1.372
Simulcast	0.452	-9.685	2.532	-48.587

표 4 BD-PSNR and BD-Rate

다중 스케일 특징 융합 모듈이 있는 모델은 기본 계층의 정보를 참고하여 향상 계층의 정보를 압축하기 때문에, 향상 계층이 풍부해진 정보를 가지게 되며, 그와 동시에 두 계층 간에 특징 맵의 유사도가 높아졌기 때문에 하위 계층의 특징을 상위 계층으로 예측하여 잔차 특징 맵을 구할 때 적은 비트스트림으로 좋은 화질을 제공할 수 있음을 확인할 수 있다.

4. 결론

본 논문은 네트워크의 대역폭 및 디바이스 사양의 제약에 따라 공간적으로 확장할 수 있는 종단 간 학습기반 영상 압축 기술의 효율을 더 높이기 위해, 변환 방식을 개선한 모델을 제안한다. 제안 모델은 변환 과정 중 계층 간의 중간 특징 맵을 융합하는 Multi-scale feature fusion 모듈을 도입해 계층 간 feature 중복성을 충분히 제거하여, 논문[3] 대비 기본 계층에서 7.39%, 향상 계층에서 1.37%의 BD-rate를 증대시켰다. 후속 연구에서는 향상 계층의 압축 효율을 증가시키기 위해 기본 계층의 학습을 멈추고 향상 계층을 미세 조정할 경우의 실험을 진행하고자 한다. 또한, 공간적일 뿐만 아닌 화질 확장 가능 압축 모델의 방향성을 모색하고자 한다.

참고 문헌

[1] H. Schwarz, D. Marpe and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, pp. 1103-1120, Sept. 2007, doi: 10.1109/TCSVT.2007.905532.

[2] J. M. Boyce, Y. Ye, J. Chen and A. K. Ramasubramonian, "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 20-34, Jan. 2016, doi: 10.1109/TCSVT.2015.2461951.

[3] Y. Mei, L. Li, Z. Li and F. Li, "Learning-Based Scalable Image Compression with Latent-Feature Reuse and Prediction," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2021.3114548.

[4] C. Jia, Z. Liu, Y. Wang, S. Ma and W. Gao, "Layered Image Compression Using Scalable Auto-Encoder," 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2019, pp. 431-436, doi: 10.1109/MIPR.2019.00087.

[5] Bégaint, Jean, et al. "Compressai: a pytorch library and evaluation platform for end-to-end compression research." arXiv preprint arXiv:2011.03029 (2020).