

적응적 크기 조정을 이용한 블록 기반 신경망 이미지 부호화

박민정, 김영웅, *김동현, *임성창, 김휘용*

경희대학교, *ETRI

{mindyeoi, duddnd7575, hykim.v}@khu.ac.kr

*{kimddng, sclim}@etri.re.kr

Neural Image Compression using Block based Adaptive Resizing

Min Jeong Park Yeongwoong Kim *Donghyun Kim *Sung Chang Lim Hui Yong Kim*

Kyung Hee University, *ETRI

+: 교신저자(Corresponding author)

요 약

본 논문에서는 최근 연구되고 있는 신경망 이미지 부호화(NNIC: Neural Network based Image Coding)를 위한 적응적 크기 조정을 이용한 블록 기반 신경망 이미지 부호화 알고리즘을 제안한다. 제안 방법은 이미지를 여러 개의 $2N \times 2N$ 블록으로 분할한 후 분할된 각 블록에 대해 두 가지 크기 조정 모드 중 하나로 부호화를 수행한다. 첫번째 모드는 $2N \times 2N$ 블록을 구성하는 4 개의 $N \times N$ 블록을 각각 NNIC 인코더의 입력으로 사용하는 모드 1(크기 미조정 모드)이며, 두번째 모드는 $2N \times 2N$ 블록을 하나의 $N \times N$ 블록으로 다운 스케일링하여 NNIC 입력으로 사용하는 모드 2(크기 조정 모드)이다. 모드 결정은 비트율-왜곡 비용(Rate-distortion Cost)이 더 적도록 이루어진다.

블록 기반 부호화와 제안 알고리즘을 비교하면, BDBR 은 약 -1.75%, BDSNR 은 약 0.073dB 으로 제안 알고리즘에서 성능 향상이 나타났고, 픽처 부호화와 제안 알고리즘을 비교하면 BDBR 은 약 0.57%, BDSNR 은 -0.029dB 로 픽처 부호화와 거의 유사한 성능을 보인다는 것을 확인할 수 있다.

1. 서론

정보통신기술과 하드웨어의 급격한 발전으로 인해 최근에는 연구소 및 기업 뿐만 아니라 일반 사용자들도 대용량 이미지 데이터를 다루고 있다. 초고화질 이미지에 대한 수요가 증가함에 따라 고성능 이미지 부호화 기술도 중요한 위치를 차지하고 있다. 또한 GPU 의 획기적인 성능 향상으로 근래 딥러닝 분야가 큰 발전을 이루며 다양한 딥러닝 모델이 발표되고 있다. 딥러닝 모델은 많은 분야에 적용되고 있지만 특히 고성능의 이미지 부호화 기술을 위해서도 활용되고 있다.

신경망 이미지 부호화(NNIC: Neural Network based Image Coding)는 엔트로피를 최소화하는 방향으로 연구되어 이미지 부호화 성능을 획기적으로 개선하고 있다.[4][5][6][7]. 하지만 픽처 전체를 압축 및 복원하기 때문에 단말에서 이를

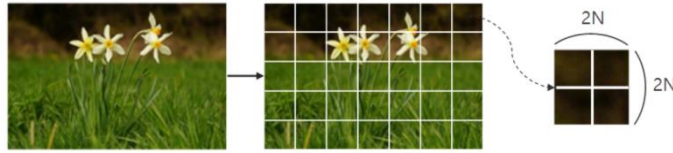
활용하기에는 어려움이 있다. 따라서 픽처를 블록 기반으로 처리하여 압축 및 복원할 필요가 있다.

따라서 본 논문에서는 픽처를 블록 기반으로 처리하여 단말에서도 신경망 이미지 부호화가 용이하도록 한다. 하지만 이 방법만으로는 성능 저하가 예상되므로, 두 가지 압축 모드 중 유효-왜곡 비용이 최소화되는 모드를 선택하는 적응적 크기 조정을 적용하는 방법을 제안한다. 이를 통해 블록 기반 이미지 처리의 장점을 활용하며 단점도 보완할 수 있다.

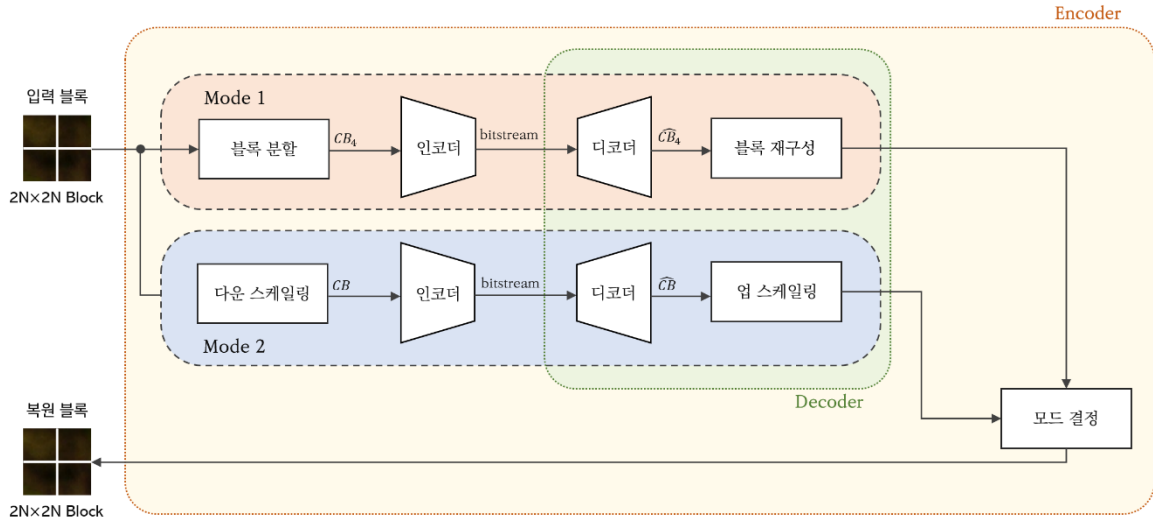
2. 관련 연구

2.1. 블록 기반 부호화

H.264/AVC, HEVC, VVC 등 최근 상용화된 비디오 코덱은 압축 효율과 병렬처리 등을 위해 하나의 픽처를 계층적으로



[그림 1] 픽처 내 분할된 2N × 2N 블록



[그림 2] 적응적 크기 조정이 적용된 인코더 블록도

분할하여 처리하는 블록 기반 부호화를 한다[8][9][10]. 최근 신경망 이미지 부호화 연구에도 블록 기반 부호화를 적용하는 연구가 진행되었는데, 인접 블록 간의 관계를 활용하는 화면 내 예측을 수행하며 효과적인 정보 예측을 달성함으로써 VVC 보다 더 높은 성능을 달성하였다[11].

2.2. 모드 결정법

모드 결정법은 HEVC 와 H.264/AVC 화면 내 예측 과정에서도 사용한다[10][12]. HEVC 에서는 16×16, 32 ×32, 64×64 크기의 부호화 단위(CTU: Coding Tree Unit)을 설정하고 각각의 CTU 들을 다양한 크기의 부호화 단위(CU: Coding Unit)로 나눈다. 각 CU 에서는 부호화를 위한 다양한 크기의 예측 단위(PU: Prediction Unit) 열는데, 이때 35 가지의 예측 모드를 사용하여 예측을 수행한다. H.264/AVC 에서는 9 가지의 예측 모드를 사용한다.

2.3. 신경망 이미지 부호화 (NNIC: Neural Network based Image Coding)

신경망 이미지 부호화는 2016 년 구글의 심층 신경망 기반의

이미지 압축 기술 연구를 기점으로 지속적으로 발전하고 있다[1]. 초기에는 오토 인코더 기반의 연구가 이루어졌지만 기존 이미지 코덱과 비교하여 낮은 성능을 보였다. 이후 엔트로피 모델을 이용한 울-왜곡 최적화 기법을 이용하여 입력 픽처를 낮은 엔트로피 코드 데이터로 변환하는 기술이 연구되었고[2][3], 엔트로피를 최소화하는 다양한 알고리즘이 개발되어 이미지 부호화 성능을 획기적으로 개선하였다[4][5][6][7]. 성능 평가와 울-왜곡 최적화를 위하여 RD Cost 로 손실함수를 정의하고 이를 통해 성능을 평가한다. 식은 다음과 같다.

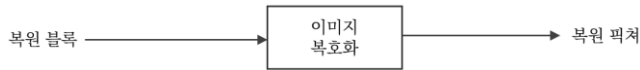
$$Loss = Rate + \lambda \times Distortion$$

[식 1] 손실함수 식

λ 는 화질 수준을 나타내는 파라미터이고, $Rate$ 는 비트량을 나타내며 $Distortion$ 은 원본 픽처와 복원 픽처의 차이를 의미한다.

3. 적응적 크기 조정을 이용한 블록 기반 신경망 이미지 부호화

본 논문에서 제시하는 방법은 다음과 같다. [그림 1]과 같이 픽처를 2N×2N 블록으로 나눈 뒤 인코더에서 해당 블록을 두 가



[그림 3] 적응적 크기 조정이 적용된 디코더 블록도

지 모드로 부호화 및 복호화 한다.

모드 1에서는 $2N \times 2N$ 블록을 4개의 $N \times N$ 서브 블록으로 분할하여 CB_4 를 만든다. 즉, CB_4 는 4개의 $N \times N$ 서브 블록을 나타낸다. 각각의 $N \times N$ 서브 블록은 인코더의 입력이 되고, 출력으로 나온 비트스트림(bitstream)은 디코더의 입력이 되어 복원된 4개의 $N \times N$ 서브 블록인 \widehat{CB}_4 를 출력한다. \widehat{CB}_4 을 통해 복원된 $2N \times 2N$ 블록을 재구성하고, 복원된 $2N \times 2N$ 블록과 부호화 전 $2N \times 2N$ 을 비교하여 [식 1]의 손실함수 식을 통해 손실을 계산한다.

모드 2에서는 $2N \times 2N$ 블록을 쌍 3차 보간법(bicubic interpolation)을 통해 $N \times N$ 블록으로 다운 스케일링 한다. 이와 같이 $N \times N$ 으로 다운 스케일링된 블록을 CB 라고 하자. CB 는 인코더의 입력이 되어 출력으로 비트스트림을 생성하고, 이 비트스트림은 다시 디코더의 입력이 되어 복원된 $N \times N$ 블록 \widehat{CB} 을 생성한다. \widehat{CB} 은 쌍 3차 보간법을 통해 $2N \times 2N$ 블록으로 업 스케일링 된다. 이후 $2N \times 2N$ 블록으로 업 스케일링 된 블록과 부호화 전의 $2N \times 2N$ 을 비교하여 모드 1의 마지막 단계와 같이 [식 1]의 손실함수 식을 통해 손실을 계산한다.

각 모드에서의 손실값을 비교하여 손실이 적은 모드를 선택하고, 해당 모드로 복원된 블록을 디코더로 보낸다. 인코더에 사용된 인코더와 디코더는 신경망 이미지 부호화 및 복호화 모델이며, 위 과정은 [그림 2]에 블록도로 나타나 있다.

복원된 블록들을 통해 이미지 복호화 과정을 거쳐 최종적으로 복원 픽처를 생성한다. 이 과정은 [그림 3]에 블록도로 나타나 있다.

4. 실험 방법 및 결과

4.1. 실험 방법

성능 평가에는 Kodak24[13] 데이터셋이 사용되었다. 신경망 이미지 부호화 모델은 Minnen[15]의 Mean-Scale Hyperprior 모델을 사용했다. 해당 모델은 CompressAI[14]에서 제공하였는데, Vimeo90K[16], xue2019video[17] 데이터셋을 임의로 256×256 패치로 추출하여 400 만에서 500 만 step 학습된 사전 학습 모델이다.

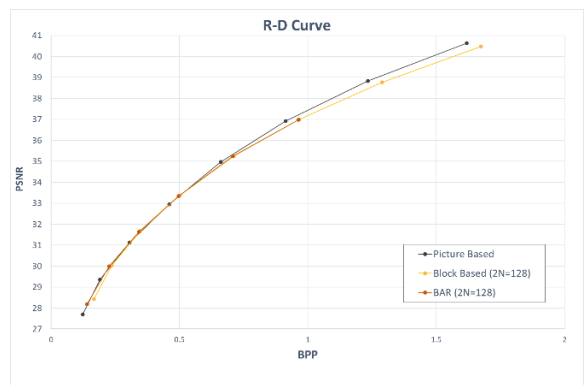
픽처를 압축하는 알고리즘(이하 Picture Based)과 픽처를 $2N \times 2N(2N=128)$ 블록 기반으로 처리하여 부호화 하는 알고리즘(이하 Block Based), 그리고 제안 방식인 적응적 크기 조절을 이용한 $2N \times 2N(2N=128)$ 블록 기반 부호화 알고리즘(이하 BAR)을

비교한다. BAR 은 3절에서 제시한 적응적 크기 조절을 이용한 블록 기반 신경망 이미지 부호화 모델을 적용한 알고리즘이다. 모드 2는 모드 1 보다 2 단계 더 높은 화질 수준을 갖는 신경망 이미지 부호화 모델을 사용하는데, 이는 각 모드가 유사한 수준의 PSNR 을 보이도록 한 후 손실값을 비교하기 위함이다.

성능 평가는 BPP와 PSNR 을 화질 수준 1~6에 대해 비교하여 R-D Curve와 BDBR, BDPSNR 을 통해 진행한다. BAR 의 경우 BPP 계산 시 모드를 선택하는 bit 값도 고려하였다.

4.2. 실험 결과

아래 [그림 4]는 Picture Based, Block Based, BAR 의 성능을 R-D Curve 를 통해 비교하여 나타낸 것이다.



[그림 4] Picture Based, Block Based, BAR 을 비교한 R-D Curve

위 그림을 통해 Block Based 가 Picture Based 보다 모든 BPP 에 대해 성능이 낮다는 것을 확인할 수 있다. 하지만 제안 알고리즘인 BAR 의 경우, 낮은 BPP 에 대해서 Picture Based 보다 더 높거나 유사한 성능을 보이는 것을 확인할 수 있다.

Block Based 와 BAR 을 비교하면, BDBR 은 약 -1.75%, BDSNR 은 약 0.073dB 으로 제안 알고리즘인 BAR 에서 성능 향상이 나타났고, Picture Based 와 BAR 을 비교하면 BDBR 은 약 0.57%, BDSNR 은 -0.029dB 로 Picture Based 와 거의 유사한 성능을 보인다는 것을 확인할 수 있다.

5. 결론

본 논문에서는 적응적 크기 조절을 이용한 블록 기반 신경망 부호화 알고리즘을 제시하였다. 제시한 알고리즘은 픽처를 블록

기반으로 처리함으로써 단말에서도 신경망 이미지 부호화가 용이하도록 하면서도, 적응적 크기 조정을 통해 부호화 효율을 픽처 자체를 처리하는 것과 유사하도록 성능 향상을 이루었다.

앞으로 신경망 이미지 부호화 모델과 초해상화 모델의 합동 훈련[18]을 통해 부호화 모델과 모드 2 의 업 스케일링 방식을 개선한다면 더 큰 성능 향상 이룰 수 있을 것으로 기대한다.

ACKNOWLEDGEMENT

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

참고문헌

[1] J. Ballé, D. Minnen, S. Singh, S.-J. Hwang, and N. Johnston. Variational image compression with a scale hyperprior. In the 6th Int. Conf. on Learning Representations, 2018.

[2] D. Minnen, J. Ballé, and G. Toderici. Joint autoregressive and hierarchical priors for learned image compression. In Advances in Neural Information Processing Systems, 2018.

[3] J. Lee, S. Cho, and S.-K. Beack, Context-adaptive entropy model for end-to-end optimized image compression. In the 7th Int. Conf. on Learning Representations, 2019.

[4] J. Lee, S. Cho, and M. Kim. A hybrid architecture of jointly learning image compression and quality enhancement with improved entropy minimization. arXiv preprint arXiv:1912.12817, 2019.

[5] B. Bross, W. Han, J. Ohm, G. Sullivan, Y. Wang, and T. Wiegand, "High efficiency video coding (HEVC) text specification draft 10(for FDIS & Consent)," JCTVC-L1003, 12th JCT-VC Meeting, 2013(1).

[6] Lee, Jong-Seok, et al. "Overview of VVC." Broadcasting and Media Magazine 24.4 (2019): 10-25.

[7] T. Wiegand, G. J. Sullivan, G. Bjontegarrd, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. Circuits Syst. Video Technol., pp.560-576, 2003(7).

[8] Y. Wu, X. Li, Z. Zhang, X. Jin and Z. Chen, "Learned Block-based Hybrid Image Compression," in IEEE Transactions on Circuits and Systems for Video Technology, doi: 10.1109/TCSVT.2021.3119660.

[9] 황수진(Soo-Jin Hwang), and 호요성(Yo-Sung Ho). "H.264/AVC 의 비트율 왜곡값과 시간 상관도를 이용한 고속 모드 결정 방법." 한국방송공학회 학술발표대회 논문집 2011.7 (2011): 348-351.

[10] 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. In the 4th Int. Conf. on Learning Representations, 2016.

[11] J. Ballé, V. Laparra, and E.-P. Simoncelli. End-to-end optimized image compression. In the 5th Int. Conf. on Learning Representations, 2017.

[12] L. Theis, W. Shi, A. Cunningham, and F. Huszár. Lossy image compression with compressive autoencoders. In the 5th Int. Conf. on Learning Representations, 2017.

[13] R. Franzen. (1999). Kodak Lossless True Color Image Suite. [Online]. Available: <http://r0k.us/graphics/kodak>

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

[15] Minnen, David, Johannes Ballé, and George D. Toderici. "Joint autoregressive and hierarchical priors for learned image compression." Advances in neural information processing systems 31 (2018).

[16] Tianfan Xue, Baian Chen, Jiajun Wu, Donglai Wei, and William T Freeman. Video enhancement with task-oriented flow. International Journal of Computer Vision, 127(8):1106-1125, 2019.

[17] Liu, Ce and Sun and Deqing. "On Bayesian adaptive video super resolution." IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(2):346-360, 2013.

[18] 조현동, 김영웅, 차준영, 김동현, 임성창, 김휘용. "신경망 이미지 부호화 모델과 초해상화 모델의 합동훈련" 2022 년 한국방송미디어공학회 하계학술대회