

## 신경망 이미지 부호화 모델과 초해상화 모델의 합동훈련

조현동, 김영웅, 차준영, \*김동현, \*임성창, 김휘용‡

경희대학교, \*한국전자통신연구원(ETRI)

{gusehd1113, duddnd7575, c jy8922, hykim.v} @khu.ac.kr

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

‡: 교신저자(Corresponding Author)

Joint Training of Neural Image Compression  
and Super Resolution ModelHyun Dong Cho, YeongWoong Kim, Junyeong Cha, \*DongHyun Kim, \*Sung Chang Lim, Hui Yong Kim<sup>†</sup>  
Kyunghee University, \*ETRI

## 요약

인터넷의 발전으로 수많은 이미지와 비디오를 손쉽게 이용할 수 있게 되었다. 이미지와 비디오 데이터의 양이 기하급수적으로 증가함에 따라, JPEG, HEVC, VVC 등 이미지와 비디오를 효율적으로 저장하기 위한 부호화 기술들이 등장했다. 최근에는 인공지능 경망을 활용한 학습 기반 모델이 발전함에 따라, 이를 활용한 이미지 및 비디오 압축 기술에 관한 연구가 빠르게 진행되고 있다. NNIC (Neural Network based Image Coding)는 이러한 학습 가능한 인공지능 기반 이미지 부호화 기술을 의미한다.

본 논문에서는 NNIC 모델과 인공지능 기반의 초해상화(Super Resolution) 모델을 합동훈련하여 기존 NNIC 모델보다 더 높은 성능을 보일 수 있는 방법을 제시한다. 먼저 NNIC 인코더(Encoder)에 이미지를 입력하기 전 다운 스케일링(Down Scaling)으로 쌍삼차보간법을 사용하여 이미지의 화소를 줄인 후 부호화(Encoding)한다. NNIC 디코더(Decoder)를 통해 부호화된 이미지를 복호화(Decoding)하고 업 스케일링으로 초해상화를 통해 복호화된 이미지를 원본 이미지로 복원한다. 이때 NNIC 모델과 초해상화 모델을 합동훈련한다.

결과적으로 낮은 비트량에서 더 높은 성능을 볼 수 있는 가능성을 보였다. 또한 합동훈련을 함으로써 전체 성능의 향상을 보아 학습 시간을 늘리고, 압축 잡음을 위한 초해상화 모델을 사용한다면 기존의 NNIC 보다 나은 성능을 보일 수 있는 가능성을 시사한다.

## 1. 서론

인터넷의 발전으로 비디오와 이미지 데이터의 양이 방대해졌다. 방대한 데이터를 효율적으로 저장하기 위해 JPEG, HEVC<sup>[1]</sup>, VVC<sup>[2]</sup> 등 다양한 부호화 기술이 등장했다. 최근에는 인공지능 경망을 활용한 학습 기반 모델이 발전함에 따라, 이를 활용한 NNIC (Neural Network based Image Coding) 기술에 관한 연구가 빠르게 진행되고 있다.<sup>[3][4][5][6]</sup> 실제로 NNIC를 활용한 기술들은 기존 기술을 뛰어넘는 성능을 보이고 있다. 일반적으로 NNIC 모델은 오토인코더(Auto-Encoder) 기반의 부호화 및 복호화 신경망과 학습 가능한 엔트로피 모델(Entropy Model)을 통해 구현된다. NNIC 모델의 목적은 비트 대비 화질을 최적화되도록 학습하는 것이고, 이를 율-왜곡화(Rate-Distortion) 최적화라고 부른다.

한편 가장 흔하게 볼 수 있는 JPEG 압축 이미지에는 압축함으로써 생기는 블록 아티팩트, 링잉 아티팩트, 블러링 등 잡음이 존재한다. 이는 특히 낮은 비트량을 사용하여 압축할 때 두드러진다. 그래서 낮은 비트량을 사용하여 이미지를 JPEG으로 압축 시 이미지를 다운 스케일링하여 JPEG 알고리즘에 입력하고 복원 후 업 스케일링을 하면 원래 이미지를 JPEG 알고리즘으로 바로 압축했을 때 보다 시각적으로 유리함이 잘 알

려져 있다.<sup>[7]</sup>

두 가지 배경에서 착안하여 NNIC 모델의 입출력에 다운 스케일링(Down Scaling)과 업 스케일링(Up Scaling)을 추가할 것을 제안한다. 이는 두 가지 가정을 전제로 실험하였다. 첫째로 이미지를 다운 스케일링하여 NNIC 인코더에 입력한다면 낮은 비트량으로 화질을 구현할 수 있다. 둘째로 공간적 중복도가 높은 단순한 이미지 또는 낮은 비트량을 사용하여 압축 및 복원하는 이미지는 다운 스케일링 및 업 스케일링 시 원본 복원이 잘되기 때문에 이러한 특성이 강할 것이라고 가정했다. 따라서 낮은 비트량에서의 성능 향상을 예상했다. 또한 업 스케일링 기법으로 초해상화 기법을 NNIC 모델에 합동훈련하여 최적화를 한다면 성능 폭이 전체적으로 향상될 것이라고 보았다.

본 논문에서 제안하는 구조는 다음과 같다. 먼저 다운 스케일링을 통해 원본 이미지의 해상도를 줄인 후 NNIC 모델의 입력으로 사용한다. NNIC 모델은 해당 입력을 부호화 및 복호화하여 다운 스케일링된 이미지를 출력한다. 마지막으로 초해상화를 통해 원본 이미지와 동일한 크기가 되도록 업 스케일링을 진행한다. 이때 NNIC 모델과 초해상화 모델이 최적화 되도록 합동훈련한다.

## 2. 관련 연구

### 2.1. 초해상화 알고리즘

기술 발전과 사용자의 요구에 따라 고해상도의 영상을 위한 초해상화 알고리즘(Super Resolution)이 등장하였다. 먼저 단순히 주변의 값을 복사하는 최근접 이웃 보간법(Nearest Neighbor Interpolation), 선형적으로 값을 연산하는 선형 보간법(Bilinear Interpolation), 다항식을 이용하여 값을 연산하는 쌍삼차보간법(Bicubic Interpolation) 등 여러 보간법이 등장했다. 그 중 쌍삼차보간법은 1개의 점을 추정하기 위해 16개의 주변의 값을 3차 함수를 이용하여 값을 연산하는 방법이며, 최근접 이웃 보간법이나 선형 보간법에 비해 부드럽게 화질을 표현할 수 있다. 한편, 인공지능경망을 활용한 기술들이 발전하면서 초해상화 알고리즘 분야에서도 인공지능경망을 활용하기 시작했다. 합성곱 신경망(CNN)을 처음으로 초해상화 알고리즘 분야에 적용한 SRCNN<sup>[8]</sup>이 필두로 SRCNN을 개선한 FSRCNN<sup>[9]</sup>, ESPCN<sup>[10]</sup>, 잔차 학습법을 활용한 VDSR<sup>[11]</sup>, EDSR<sup>[12]</sup>, 또 GAN을 활용한 SRGAN<sup>[13]</sup> 등 여러 인공지능경망 기반의 초해상화 알고리즘이 등장하였다. 보간법은 정해져 있는 값으로 복사, 연산하는 반면 인공지능경망 기반의 초해상화 알고리즘은 학습을 통해 최적의 값을 출력해낸다. 또한 여러 겹의 합성곱 신경망을 통해 이미지의 비선형적 특성을 잘 분석할 수 있다.

본 논문에서는 다운 스케일링 기법으로 쌍삼차보간법을 사용하였다. 업 스케일링 기법으로는 코드의 접근이 쉽고 사전 모델을 제공하는 EDSR 모델을 사용하였다.

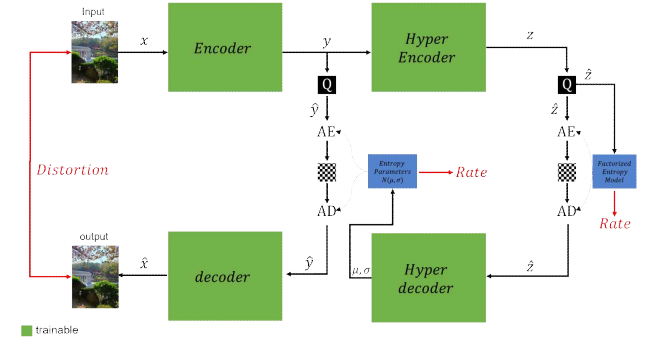
### 2.2. NNIC (Neural Network based Image Coding)

최근 인공지능경망 기반의 압축 기술인 NNIC가 다양한 방식으로 발전하고 있다. 그 중 본 논문에서는 Minnen이 제안한 NNIC를 사용한다.[4] [그림1]은 논문에서 사용한 모델의 구조이다. [4]는 오토인코더 형태의 모델로서 인코더와 디코더로 구성되어 있다. 입력 이미지  $x$ 는 인코더를 거쳐 은닉 벡터  $y$ 로 변환된다. 은닉 벡터  $y$ 는 또 다른 은닉 벡터  $z$ 를 출력하는 하이퍼 인코더(Hyper Encoder)로 입력된다. 엔트로피 모델을 활용하여 은닉 벡터  $z$ 를 채널마다 확률 분포를 추정하며 비트량을 계산한다.(Factorized Entropy Model) 부호화 및 복호화 후 하이퍼 디코더(Hyper Decoder)로 입력돼 평균과 표준편차 값을 출력한다. 한편, 은닉 벡터  $z$ 에서 얻은 평균과 표준편차 값으로 엔트로피 모델을 활용하여 은닉 벡터  $y$ 의 각 화소 마다 확률분포를 추정하여 비트량을 계산한다.(Entropy Parameters) 그리고 부호화 및 복호화를 한 뒤 디코더를 거쳐 이미지를 복원한다. 복원된 이미지와 입력 이미지의 차이를 구하여 입력 이미지에 근사하게 복원되도록 학습한다.

[그림1]은 앞선 설명에 대한 그림이다. 입력 이미지와 출력 이미지의 차이(Distortion)와 은닉벡터  $y$ 와  $z$ 에서 엔트로피 모델을 이용하여 계산한 비트량(Rate)을 두 가지 목적함수로 정의하여 인코더와 디코더를 학습한다. 두 목적함수에서 비트량을 줄이면 화질이 저하되고, 비트량을 늘리면 화질이 상승되는 관계를 보인다. 따라서 [그림2]와 같이 파라미터  $\lambda$ 를 통해 두 목적함수를 정의한다.

[4]의 논문에서는 화소 간 중복성을 고려한 Context Model을 추가하였지만, Context Model에서 은닉 벡터들 간에 연속적인 중복성이 존재하여 병렬처리가 어렵고 이에 따라 계산 시간이 급증하기 때문에 본 논

문에서는 Context Model을 제외하고 학습했다.

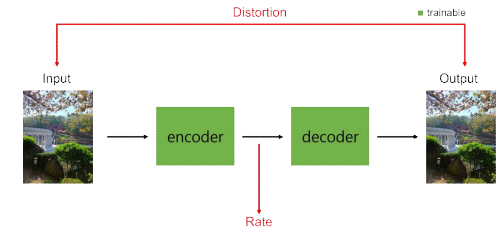


[그림1] NNIC 모델

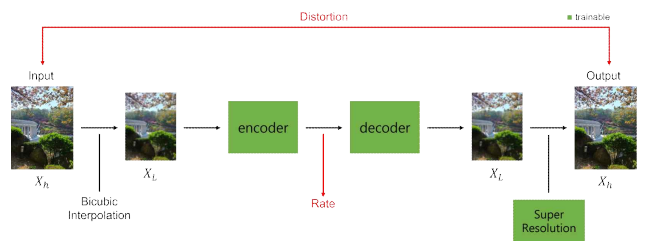
$$Loss = Rate + \lambda \times Distortion$$

[그림2] NNIC 목적함수

## 3. 제안



[그림3] NNIC 모델(요약)



[그림4] 제안 모델

### 3.1. NNIC 모델에서의 쌍삼차보간법 과 초해상화 기법

먼저 [그림3]의 NNIC 모델의 입력과 출력이 쌍삼차보간법과 초해상화 기법을 사용하는 것을 제안한다.[그림4 참고] 입력 이미지  $X_h$  ( $H \times W$ )를 쌍삼차보간법으로  $X_L$  ( $H/2 \times W/2$ )로 다운 스케일링 하여 인코더에 입력함으로써 낮은 비트량으로 부호화할 수 있다. 엔트로피 모델을 이용하여 비트량을 계산한 후 복호화를 하여  $X_L$  ( $H/2 \times W/2$ )를 출력한다. 원본 이미지로 복원하기 위하여 업 스케일링 기법으로 초해상화 기법 중 EDSR 모델을 이용하여  $X_h$  ( $H \times W$ )의 이미지 크기로 복원한다.

### 3.2. NNIC 모델과 초해상화 모델 합동훈련(Joint Training)

EDSR은 NNIC 모델에 최적화 되지 않았기 때문에 최적화를 위해 [그림4]의 인코더, 디코더, EDSR의 합동훈련을 제안한다. 그 결과 합동훈련을 하지 않았을 때와 비교하여 전체적으로 성능 향상을 보였다.

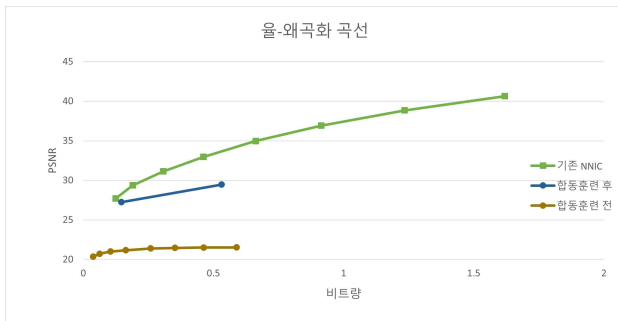
#### 4. 실험

##### 4.1. 데이터셋

본 논문에서는 OpenImage14를 사용했다. 약 9백만개의 이미지 중 가로, 세로 크기가 모두 1000화소 이상인 것 중 3만개의 이미지를  $384 \times 384$ 로 잘라 학습하였다. 초반 약 100만 iteration까지는 배치 사이즈는 4, 옵티마이저로 Adam을 사용했고 학습률은  $1e-4$ 로 설정했다. 이후 학습률을  $5e-5$ , 배치사이즈 8로 설정하고 모델 성능의 개선이 없을 시, 학습률의 0.5로 줄이는 ReduceLROnPlateau를 이용하여 옵티마이저를 조절했고, 학습률이  $1e-6$  미만이 된다면 학습을 멈추도록 설정하였다. 입력 이미지와 출력 이미지의 차이는 MSE로 계산하여 PSNR으로 평가하였고, 비트량은 Entropy로 계산하였다.

테스트 셋으로는 Kodak2415를 사용했다. 입력 이미지와 출력 이미지의 차이는 MSE로 계산하여 PSNR로 평가하였다. 또 비트량은 Entropy로 계산하였다.

##### 4.2. 실험결과



[그림5] 울-왜곡화 곡선

[그림5]는 울-왜곡화 곡선을 나타낸 것이다. 가로축은 비트량이고, 세로축은 PSNR을 나타낸다.

기존 NNIC는 Compressai에서 제공하는 [4]의 사전 학습된 모델로 평가한 곡선이다. 합동훈련 전은 NNIC와 EDSR 모두 사전 학습된 모델로 평가한 곡선이다. 합동훈련 후는 사전 학습된 NNIC와 EDSR을 합동 훈련하여 두 개의  $\lambda(2,6)$ 에 대해 평가한 곡선이다.

합동훈련 전에 비해 합동훈련 후에 성능 향상을 볼 수 있다. 하지만 기존 NNIC에 비해서는 성능이 떨어진다. 이는 학습하는 데 장시간이 걸리기 때문에 충분히 학습하지 못했기 때문이다. 학습을 충분한 시간동안 하고 EDSR이 아닌 압축 잡음을 위한 초해상화 모델을 사용한다면 성능이 더욱 향상될 것이다.

#### 5. 결론 및 향후 과제

본 논문에서는 NNIC 모델과 초해상화 모델의 합동훈련을 통해 기존 NNIC 모델 보다 더 높은 성능 향상을 볼 수 있는 방법을 제안했다. 해당 방법은 이미지를 다운 스케일링하여 NNIC 인코더에 입력함으로써 비트

량을 줄여 낮은 비트량에서 성능 향상을 기대했다. 또한 NNIC 모델과 초해상화 모델을 합동훈련함으로써 전체적인 성능 향상을 보였다. 본 모델을 학습하는 데에는 장시간이 걸리기 때문에 본 논문에서는 가능성만을 제시했다. 학습 시간을 충분히 늘리고 학습률을 조절한다면 충분한 성능 향상의 여지가 있다.

한편 압축함으로써 생기는 블록 아티팩트, 블러링 등 잡음 제거를 하면서 동시에 고해상도의 영상을 위한 인공신경망 기반의 초해상화 알고리즘 분야에 대한 연구도 진행되고 있다. 기존의 초해상화 알고리즘은 압축함으로써 생기는 잡음을 효과적으로 제거하지 못해 잡음을 제거하는 구조를 더한 다양한 초해상화 알고리즘이 발전하였다.[16][17][18] EDSR이 아닌 압축 잡음 제거를 위한 초해상화 모델을 더한다면 큰 성능의 향상을 보일 수 있는 여지가 있다. 따라서 향후 본 연구를 지속해서 개선할 것이다.

또한 [19]의 연구에서는 이미지를 블록 기반으로 나누어 NNIC의 인코더로 입력을 하여 NNIC의 성능을 높였다. 본 논문과 더불어 학습한다면 큰 성능 향상을 이룰 것으로 기대한다.

#### 6. 감사의 글

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

#### 7. 참고문헌

- [1] Gary J Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the high efficiency video coding (hevc) standard," IEEE Transactions on circuits and systems for video technology, 2012.
- [2] Jens-Rainer Ohm and Gary J Sullivan, "Versatile video coding&#8211;towards the next generation of video compression," In Picture Coding Symposium, volume 2018, 2018.
- [3] Johannes Balle, Valero Laparra, Eero P. Simoncelli, "END-TO-END OPTIMIZED IMAGE COMPRESSION", In International Conference on Learning Representations, 2017.
- [4] David Minnen, Johannes Ballé, George Toderici, "Joint Autoregressive and Hierarchical Priors for Learned Image Compression" Neural Information Processing Systems. 2018
- [5] Johannes Balle, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston, "Variational image compression with a scale hyperprior," In International Conference on Learning Representations, 2018.
- [6] Jooyoung Lee, Seunghyun Cho, Seung-Kwon Beack, "CONTEXT-ADAPTIVE ENTROPY MODEL FOR END-TO-END OPTIMIZED IMAGE COMPRESSION" In International Conference on Learning Representations, 2019.
- [7] . A. M. Bruckstein, M. Elad and R. Kimmel, "Down-scaling for better transform compression," in IEEE Transactions on Image Processing, vol. 12, no. 9, pp. 1132-1144, Sept. 2003, doi: 10.1109/TIP.2003.816023.
- [8] Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang "Image Super-Resolution Using Deep Convolutional Networks"

TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 38, NO. 2, FEBRUARY 2016

[9]Chao Dong, Chen Change Loy, Xiaoou Tang “Accelerating the Super-Resolution Convolutional Neural Network”, ECCV 2016 pp 391-407

[10] Wenzhe Shi, Jose Caballero, Ferenc Huszar, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, Zehan Wang, “Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1874-1883

[11] Jiwon Kim, Jung Kwon Lee, Kyoung Mu Lee. “Accurate Image Super-Resolution Using Very Deep Convolutional Networks”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1646-1654

[12] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, Kyoung Mu Lee. “Enhanced Deep Residual Networks for Single Image Super-Resolution”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2017, pp. 136-144

[13]Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4681-4690”

[16] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Mallocci, A. Kolesnikov, T. Duerig, and V. Ferrari. “The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale.” IJCV, 2020.

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

[16] Chao Dong, Yubin Deng, Chen Change Loy, Xiaoou Tang, “Compression Artifacts Reduction by a Deep Convolutional Network”, Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 576-584

[17]Honggang Chen, Xiaohai He, Chao Ren, Linbo Qing, Qizhi Teng “CISRDCNN: Super-resolution of compressed images using deep convolutional neural networks”, Neurocomputing Volume 285, 12 April 2018, Pages 204-219

[18] Hongming Luo, Fei Zhou, Guangsen Liao, and Guoping Qiu, “Super-resolving Compressed Images via Parallel and Series Integration of Artifact Reduction and Resolution Enhancement” arXiv:2103.01698

[19] 박민정, 김영웅, 김동현, 임성창, 김휘용, “적응적 크기 조정을 이용한 블록 기반 신경망 이미지 부호화” 방송미디어공학회, 2022