

# Reducing JPEG Artifacts on the Coefficient and Spatial Domain

Minhyeong Kim(2070013) Jihye Lee(2071040)

June 2023

## Abstract

기존의 JPEG 압축 과정에서 생기는 이미지 artifact를 줄이기 위해 HR-CNN과 DCSC 모델을 제안한다. 제안된 모델은 decoding 및 artifact 억제를 통합한 end-to-end fully convolutional network으로 구성되어 있다. HR-CNN(Heterogeneous Residual Convolutional Neural Network)은 효과적인 artifact 억제를 위해 복잡한 과정을 단순화하는 특징을 가지고 있으며, DCSC(Deep Convolutional Spatial Coding) 모델은 spatial domain에서 artifact를 억제한다. 실험결과, 제안된 모델은 JPEG 압축으로 인해 발생하는 artifact를 효과적으로 억제하고 이미지 품질을 개선하는데 성공했다. 이러한 모델은 영상 압축 및 통신 분야에서 유용하게 사용될 수 있다.

**Index Terms** Jpeg, decoding, artifacts suppression  
[https://github.com/penguin1109/2023\\_DIP\\_Team\\_Project](https://github.com/penguin1109/2023_DIP_Team_Project)

## 1 Introduction

JPEG 압축 기술은 국제적으로 표준화된 방식으로, DCT 변환 방식을 기반으로 하여 우수한 영상 재생성을 제공하면서도 이미지의 화질을 최대한 보존할 수 있는 방식이다. 이러한 특성으로 인해 JPEG은 영상 압축 및 통신 분야에서 많이 사용되고 있다. JPEG은 높은 압축률을 제공하면서도 이미지의 품질을 상당히 유지할 수 있는데, 이는 낮은 주파수 성분을 보존하면서도 높은 주파수 성분을 상대적으로 정확도를 감소시킴으로써 압축하는 방식이다. 이러한 과정에서 artifact가 발생할 수 있다.

JPEG 코드를 기반으로 한 artifact를 억제하기 위해 새로운 heterogeneous residual convolutional neural network (HR-CNN)을 제안한다. 이 모델은 decoding과 artifact 억제를 하나의 end-to-end fully convolutional network로 결합하여 작동 효율성을 향상시키고 메모리 사용량을 줄이는 등의 이점을 가지고 있다. HR-CNN은 복잡한 과정을 단순화하면서도

artifact 억제를 효과적으로 수행하는 것이 특징이다. 추가로 DCSC 모델을 이용하여 spatial domain에서 artifact를 억제해보았다.

## 2 Design Objectives and Requirements

JPEG 압축의 주요한 문제 중 하나는 압축 과정에서 발생하는 artifact이다. JPEG은 이미지를 압축할 때 높은 주파수 성분을 보다 낮은 정밀도로 표현하여 압축률을 높이는데, 이로 인해 이미지의 세부 정보가 손실되고 artifact가 발생할 수 있다. 주로 macro blocking artifacts 문제를 해결하고자 한다. 이에 대해 네트워크를 채널 분할, decoding 및 매크로 블록킹 artifact 억제, 이미지 디테일 sharpening으로 구성하여 압축 과정에서 발생하는 artifact를 최소화하고자 한다.

원본 이미지를 JPEG 압축 프로토콜을 통해 인코딩한 후, 주파수 도메인 이미지를 기반으로 원래 이미지와 최대한 유사한 이미지로 복구하는 것이 목표이다. 원본 이미지의 크기를 128x128로 조정한 후 가로와 세로를 각각 8개의 블록으로 분할한다. 이후 각 블록에 대해 JPEG 압축 프로토콜을 적용하여 복원한 이미지를 얻게 된다.

다음은 JPEG으로 압축된 이미지에서 원본 이미지에 유사한 reconstructed image를 만들기 위한 수행방법이다. 이미지의 JPEG 압축으로 인해 생길 수 있는 문제점을 해결하기 위해 다음과 같은 접근 방식을 사용한다. 네트워크는 크게 채널 분할, decoding 및 macro blocking artifact 억제, 이미지 디테일 sharpening으로 구성되어 있다. 먼저 채널 분할을 수행한다. 이를 통해 각 채널이 가진 특성을 독립적으로 학습할 수 있다. GPU를 활용하여 계산의 효율성을 높이고 연산량을 줄일 수 있다. 채널분할을 통해 일반화 능력은 더욱 향상된다.

JPEG compression을 수행하는 과정에서 quantizing으로 인해 발생할 수 있는 macro blocking artifact를 억제하기 위해 decoding과 suppression 단계 이전에 원래 크기로 확장하는 데에 transposed con-

volution을 사용한다. 주파수 도메인의 샘플을 enrich하기 위해 spectral space sample을 추출한다. original JPEG decoding은 각 8x8 block에 inverse DCT를 적용하는 방법을 사용한다. 하지만 이렇게 하면 다른 spectral channel의 uncorrelated nature로 인해서 robust하다. 따라서 주파수 도메인의 JPEG 코드에서 시작하여 채널별로 spectral space sample을 추출한다. 단계별로 추출된 스펙트럼 샘플은 독립적으로 decoding되고 공간 도메인에서 결합된다.

다음은 Image detail sharpness enhancement 단계이다. Image detail sharpening은 이미지의 macro blocking artifact를 억제한 것 외에도 세부 정보를 복원하여 이미지의 선명도를 향상시키는 역할을 한다. 이를 위해 residual connection과 cascaded convolution을 사용하며, activation function로는 ReLU를 사용하였다. 이와 같은 단계를 거쳐 최종적인 모델을 도출하였다.

### 3 Methods

#### 3.1 HR-CNN

##### 3.1.1 Network Architecture

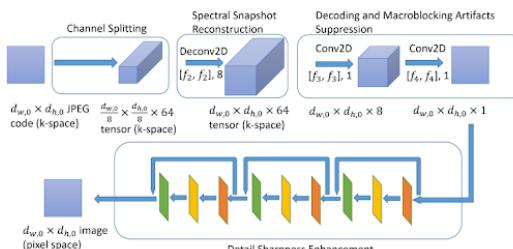


Fig. 1: Structural illustration of the heterogeneous residual convolutional neural network.

Figure 1: Network architecture

#### 3.2 DCSC

##### 3.2.1 Network Architecture



Figure 2: Network architecture

## 4 Training Details

### 1. Dataset

DIV2K에서 제공하는 train, valid set로 나뉘어진 HR(High Resolution) 이미지를 사용하였다. 기존에 데이터셋의 목적은 2x, 3x, 4x 단계의 downgrade operator을 사용하여 만들어진 저해상도 이미지를 고해상도 이미지로 바꾸는 것이고, 따라서 이미지의 해상도를 증가함에 따라 불가피하게 “비어있는 픽셀”을 채워 넣는 과정에서 최대한 원본과 비슷하게 만드는 task를 수행하기 위함이다.

하지만 우리는 고해상도로 바꾸는 것이 아닌 JPEG encoding 과정의 quantizing으로 인해서 발생하는 artifact 등을 JPEG 복원 후에 최소화하는 것이 최종적인 목적이다. 따라서 DIV2K 데이터셋의 HR데이터만을 사용하였다. Coefficient domain에서부터 원본 이미지를 만드는 HR-CNN의 학습을 위해서는 직접 구현한 JPEG encoding을 사용하였으며, spatial domain에서부터 원본 이미지를 만드는 DCSC의 학습을 위해서는 PIL 라이브러리에서 제공하는 JPEG 변환을 사용하여 input dataset을 구축하였다.

### 2. HR-CNN Details

HR-CNN 모델을 학습 시키는 과정에서는 128x128 크기의 이미지를 입력으로 사용하게 되는데, DIV2K의 원본 HR 영상은 2048x1024, 1096x2004 등으로 크기가 다양했기 때문에 128x128의 크기로 random crop을 하였다. DCT를 적용한 coefficient domain에서 제일 많이 쓰이는 8x8 quantization matrix를 사용하여서 quantizing 까지 적용하여 high frequency value들이 사라진, JPEG encoding 과정에서 정보가 소실된 k-space code를 만들어 모델이 reconstructed image를 생성하도록 하였다. 또한, DIV2K는 모두 RGB 영상만을 포함하지만, 본 실험에서는 gray scale로 바꿔 학습 시켰다.

SGD optimizer을 사용하였으며, learning rate는 1e-3, batch size는 64, 그리고 momentum과 weight decay는 각각 0.5, 0.005로 총 50 epoch을 학습 시켰다.

Loss function은 MSE를 사용하다가 과하게 output이 어두워지는 경향이 있음을 확인하여 MSE Loss에 Content Loss를 추가적으로 적용하여 0.5, 0.5의 weight를 각각의 loss function에 적용하여 더한 손실함수로 네트워크를 학습 시켰

으며, 그 결과 MSE만 사용하였을 때보다 훨씬 개선된 결과를 보임을 확인하였다.

### 3. DCSC Details

DCSC 모델을 학습 시키는 과정에서는  $80 \times 80$  크기의 이미지를 입력으로 사용하게 되고, 앞선 HR-CNN과 동일하게 DIV2K의 HR영상에 random crop을 하였다. 이때에는 PIL 라이브러리를 사용해서 JPEG 압축 후 decoding 되어서 JPEG artifact가 존재하는 영상을 네트워크의 입력으로 넣어 주었다. RGB 영상인 DIV2K를 YCrCb space로 변환한 뒤에 밝기 정보를 포함하고 있는 Y 채널의 영상만을 취하여 HR-CNN과 동일하게 gray scale로 바꾸어 학습을 진행하였다. 이때 다양한 JPEG quality value에서도 robust하게 본 method가 동작할 수 있음을 보이기 위하여 학습 데이터에서 10, 20, 30의 다양한 JPEG quality value를 통해 압축한 영상을 사용하였다.

Adam optimizer을 사용하였으며, learning rate는  $1e-4$ , batch size는 8, 총 30 epoch을 학습 시켰다.

Loss function은 MSE의 과도한 blurring을 유도한다는 특징으로 인해서 MAE Loss를 선택하였다.

## 5 Experiments and Results

### 1. JPEG Encoding

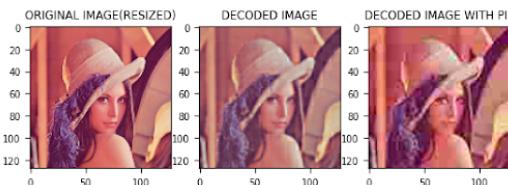


Figure 3: JPEG Encoding

### 2. HR-CNN

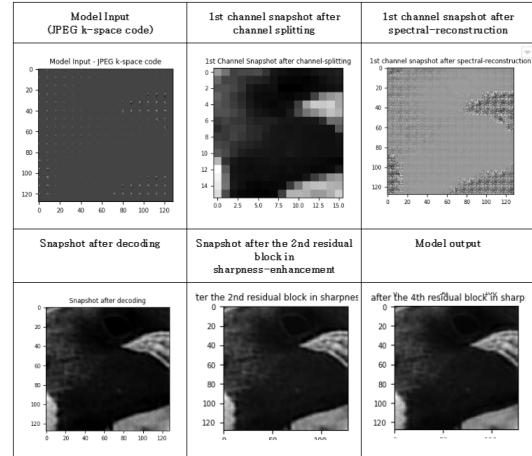


Figure 4: HR-CNN

### 3. DCSC

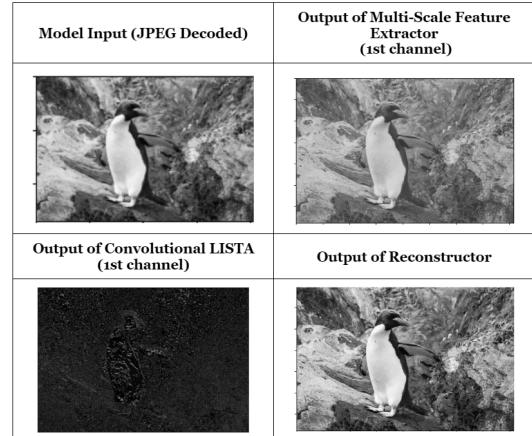


Figure 5: DCSC

### 4. PSNR Results

PSNR 결과는 skimage 라이브러리의 `peak_signal_noise_ratio`를 사용하여 모든 방법에 대해서 동일하게 계산하였다. 데이터셋은 DIV2K의 valid dataset을 사용하였으며, 총 100장의 이미지 모두에 대해서 PSNR의 평균으로 결과를 비교하였다.

	Vanilla JPEG	HR-CNN	DCSC
Average PSNR	<b>28.85601</b>	24.5279	30.3359
Sample 1			
Sample 2			
Sample 3			

Figure 6: PSNR Results

## 6 Conclusion and Future work

본 연구에서는 JPEG artifact 억제를 위한 HR-CNN 모델과 DCSC 모델을 제안하였다. 제안된 모델은 decoding과 artifact 억제를 한번에 처리하는 end-to-end processing을 하여 계산 비용과 복잡성을 줄이고 효율적인 처리를 가능하게 한다. 실험 결과에서 우수한 성능을 보였다. 하지만 학습을 함에 따라 전체적으로 픽셀 값들이 어두워지는 경향을 보였다. 이는 현재 사용하고 있는 ‘MSE loss’가 현재 방법에는 적합하지 않을 수도 있음을 의미한다. 또한, PSNR 역시 MSE loss를 기반으로 계산되기에 MSE 값이 작아지면 이미지 품질이 향상되었다고 판단하는 경향이 있다. 그러나 PSNR은 인간이 시각적으로 느끼는 품질 차이를 정확히 반영하지는 않는다. 따라서 PSNR 값이 높게 나와도 사람이 보기에는 PSNR이 낮게 나온 이미지가 화질이 더 좋아보일 수 있다는 점을 유의해야한다. HR-CNN에서, 이미지를 128x128 크기의 patch로 잘라서 처리한 후 이어 붙일 때 모서리 경계 부분에서 zero-padding으로 인해 흰색 artifact가 발생하는 문제가 있었다. 이를 해결하기 위해 padding mode를 ‘replicate’로 변경하였다. 이를 통해 모서리 부분의 patch가 잘려나가는 대신, 가장 가까운 pixel의 intensity를 복제하여 적용하여 자연스러운 이미지 연결이 가능하게 하였다.

특정한 Q-factor에서도 JPEG 압축은 이미지의 주

파수 성분에 따라 모든 디지털 이미지에 동일한 양의 정보 손실이 일어나지 않는다. 복잡한 구조로 이루어진 이미지의 경우, JPEG 압축은 현저한 artifact를 유발한다. 이는 원본 이미지가 높은 주파수 성분을 가지고 있기에 압축 후에 정확도가 감소하기 때문이다. 반대로, 세부사항이 없는 이미지의 경우, JPEG 압축은 동일한 수준의 artifact를 유발하지 않는다. 이 경우, 이미지를 재구성하더라도 PSNR의 큰 증가를 보이진 않는다.

MSE로 학습시킬 때보다 더 좋은 결과를 얻기 위해서 다음과 같은 손실 함수 사용을 제안한다.

### 1. MAE(Mean Absolute Error)

MAE는 절댓값 오차의 평균을 계산하는 방식으로, 더 큰 오차에 민감하게 반응한다.

### 2. Feature Loss(=Perceptual Loss)

사전 학습된 VGGNet, ResNet 등과 같은 단순한 CNN 모델이 예측한 reconstructed image와 original image 사이의 feature map 또는 latent vector 간의 차이를 줄이는 방식으로 loss function을 정의할 수 있다. 이는 이미지의 perceptual feature를 보존하도록 학습하는 것을 목표로 한다.

### 3. Gram Matrix Loss

style loss는 서로 다른 feature들 간의 correlation인 ‘style’을 나타내는 loss function이다. 각 layer의 channel 수에 따라 생성되는 feature map을 matrix multiplication을 통해 Gram Matrix를 계산한다. Gram Matrix는 reconstructed image와 original image의 Gram Matrix 간의 차이를 줄이는 것을 목표로 한다. 이는 같은 모델의 동일한 layer의 동일한 channel이 같은 feature에 집중하는 경향을 활용한다. Gram Matrix의 크기는 layer의 channel 수에 따라 결정되며, 해당 행렬의 내적을 계산하여 얻을 수 있다. 1번째 layer의 channel 수가 C라면, Gram Matrix의 크기는  $C \times C$ 가 된다.

이러한 대체 loss function을 사용하여 학습하면 MSE보다 우수한 결과를 얻을 수 있을 것이다. data와 network의 특성에 따라 성능이 달라질 수 있으므로 실험과 비교를 통해 가장 적합한 loss function을 정하는 것을 권장한다.

## 7 Reference

- [1] 신인경 외, “JPEG2000의 이미지 압축시 Noise 제거에 관한 연구” 한국지식정보기술학회 논문지,

vol.5, no.2, pp.91-96, 2010

- [2] Jun Niu, "End-to-End JPEG Decoding and Artifacts Suppression Using Heterogeneous Residual Convolutional Neural Network", arXiv:2007.00639v1 [eess.IV], 2020
- [3] Xueyang Fu, "JPEG Artifacts Reduction via Deep Convolutional Sparse Coding", arXiv:10.1109/ICCV.2019.00259, 2019