Image Compression and Restoration Using Deep Learning Considering Spatial Frequency Characteristics of the Visual System

Naoki Tada[1], Keita Hirai[1], Takahiko Horiuchi[1]

horiuchi@faculty.chiba-u.jp
[1]Graduate School of Science and Engineering, Chiba University, Yayoi-cho 1-33, Inage-ku, Chiba, 263-8522, Japan
Keywords: Image coding, JPEG, U-Net, CNN, Visual system

ABSTRACT

This paper proposes image compression and restoration techniques that consider the visual characteristics of humans with respect to the spatial frequency. The method reduces information with low visual sensitivity, encodes and decodes by JPEG, and restores the high-quality image using U-Net. The feasibility of the method was verified experimentally.

1 Introduction

Image data compression has been studied for extensively, and lossless compression methods such as PNG encoding and lossy compression methods such as JPEG encoding have been proposed. These methods have been used widely according to their purpose. In recent years, not only these algorithm-based image compression methods but many new approaches based on deep learning have been proposed to improve their performance [1-7]. Most of them use autoencoders and combine them with tools such as the variational autoencoder and recurrent neural networks. The effectiveness of such methods in terms of both accuracy and computing speed has been verified.

However, such approaches based on deep learning were designed to adapt the network architecture and loss function in the encoding and decoding processes. In other words, these methods incorporate lossless compression in deep learning, and there are few image compression methods based on deep learning that reduce information related to human visual characteristics the way JPEG does.

This paper proposes image compression and restoration methods based on deep learning, considering the frequency characteristics of the visual system. It is well-known that human visual systems have band-pass characteristics, that is, they are sensitive to a specific band in the frequency domain and can be expressed by a contrast sensitivity function. Considering these characteristics, frequency components other than the high-sensitivity band of the original image are reduced. Then the image restoration process uses U-Net [8], which is a convolutional neural network, to achieve a visually high-quality image restoration from JPEG-decoded images.

2 Image Compression and Restoration Methods

Figure 1 shows a flowchart of the proposed method. The proposed method consists of three steps.

1. Information reduction considering visual characteristics
2. Encoding / decoding using JPEG
3. Image restoration using U-Net

The details of each method are explained in the following sections.

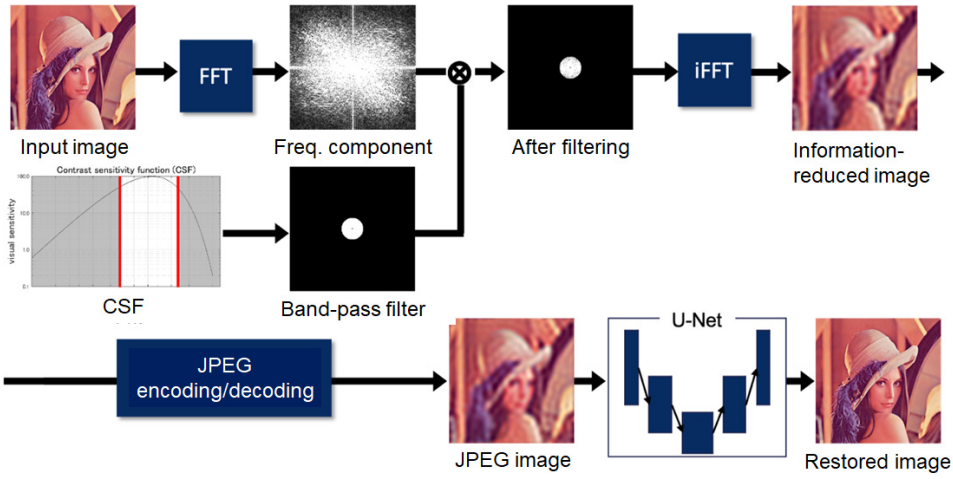


Fig. 1 Procedure of the proposed method

2.1 Information Reduction

Human visual systems have band-pass characteristics for luminance components that are highly sensitive to specific frequencies. Figure 2 shows the contrast sensitivity function (CSF) for the luminance component. The CSF used in this study is a model proposed by Daly [9]. As shown in Fig. 2, human vision has a sensitivity peak at 2 to 5 cycles/deg (cpd). Therefore, information was reduced by maintaining the frequency band between 2 and 5 cpd and reducing the areas with frequencies frequencies outside this band. In JPEG coding, information is reduced considering the visual sensitivity characteristics for human spatial frequencies; however, in JPEG compression, low compression processing is performed for the entire low-frequency band, rather than only for the peak sensitivity range. In contrast, the proposed information reduction method has been designed to retain information only in the peak band and delete the low-frequency band outside this range.

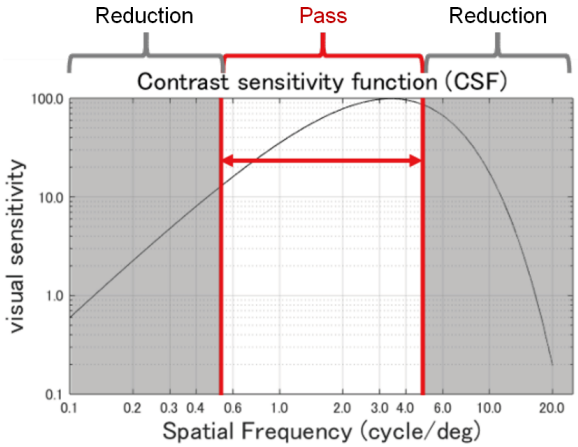


Fig. 2 Information reduction based on CSF

The images used in this study are RGB color images. The color space is converted from RGB to YCbCr and decomposed into each of the luminance component Y, the chrominance components Cb, and Cr. After that, different filtering processing is performed for each component. Each component after processing is combined, and finally the color space is converted from YCbCr to RGB. Figure 3 shows the overall procedure. The different processing for each YCbCr component is due to the difference in human visual characteristics with respect to spatial frequency between the luminance and chrominance components. According to the low-pass characteristics of CSF for chrominance components, information was reduced by leaving the frequency band around 0 to 3.9 cpd.

2.2 Image Encoding and Decoding

In this study, JPEG coding was used as the image coding and decoding method to reduce the data size while maintaining the information-reduced image. In JPEG coding, an image compressed by reducing components in the specific frequency domain is converted into a spatial frequency domain using the DCT transform. Next, it is converted into a bit string by quantization, zigzag scan, run-length encoding, and Huffman encoding. The image can be decoded from the bit string using reverse processing. The information is further reduced and deteriorated during the quantization process. In this study, the quality factor was adjusted experimentally.

2.3 Image Restoration

We used U-Net [8] to restore high-quality images from compressed images with reduced information. In U-Net, the input and output images are low-quality compressed images and high-quality restored images, respectively. The restored image is designed to match the original image before compression.

Figure 4 shows a model diagram of the U-Net. A convolutional neural network usually extracts features from an image using a convolutional layer and a pooling layer, but this network has the drawback that it loses features in a local region. The U-Net is characterized by the ability to retain local area information by a skip connection between convolution and deconvolution, as shown in Fig. 4. U-Net is a network originally designed for segmentation tasks; however, in this study, we applied it to image conversion tasks.

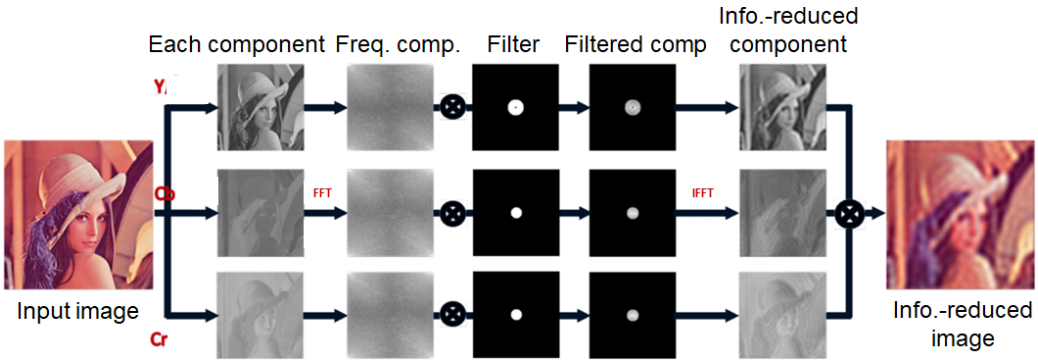


Fig.3 Procedure of information reduction

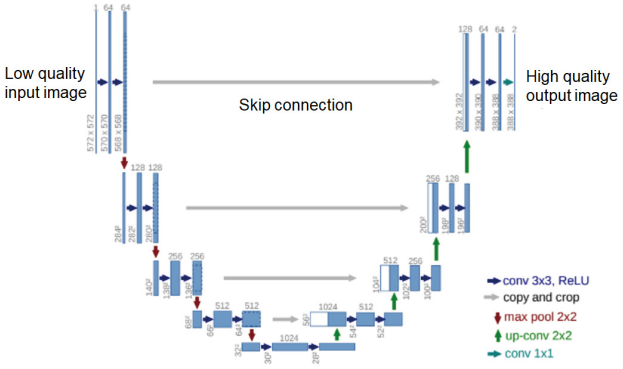


Fig. 4 Model diagram of the U-Net

For the construction of the network, 14,154 images from the SUN database [10] were used as training images. This dataset is not categorized and contains various types of images, such as landscape and portrait images. The training images were resized to 256 × 256 pixels and compressed using the proposed method to create low-quality images for input. The training and validation images were divided at a ratio of 9:1; thus, the number of training images was 12,739, and the number of validation images was 1,415. The network was trained with Adam optimization, and the mean squared error was used for the loss function. It had a learning rate of 0.0001 and a batch size of 48.

3 Experiment

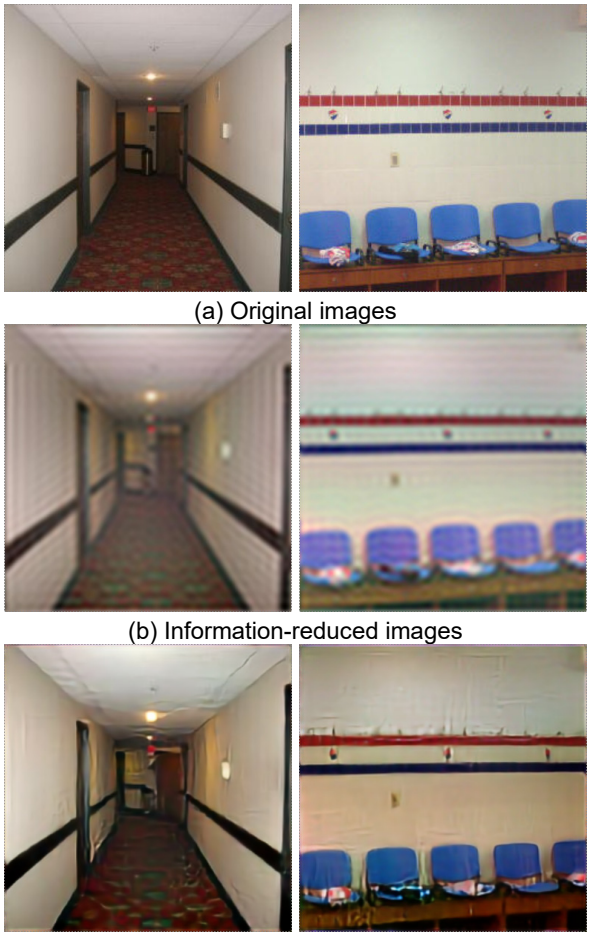
The effectiveness of the proposed method is verified by a subjective evaluation experiment and objective image quality evaluation indices.

The effectiveness of the proposed method was verified via subjective evaluation and objective image-quality evaluation indices.

Figure 5 illustrates two examples of experimental results. Figures 5(a), (b), and (c) show the original images, the information-reduced images, and the restored images, respectively. As shown in Fig. 5(b), the information-reduced images retain the structure of the image; however, periodic artifacts can be observed, and details have been lost. In contrast, as shown in Fig. 5(c), the proposed algorithm can restore the details of the images from the images in Fig. 5 (b). From a signal processing standpoint, Figs. 5(a) and (c) are not identical, but they are visually similar and well-restored.

In the proposed method, JPEG coding was performed in Step 2 to reduce the amount of data. It was necessary to maintain the coding efficiency of JPEG as a constant in order to verify the performance of the proposed method. The peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) were calculated before and after coding by changing the "Quality" parameter of JPEG. Consequently, when the quality parameter was 65, the difference between the images could not be visually recognized (PSNR> 40 and SSIM> 0.98). Therefore, in subsequent experiments, the quality of JPEG was fixed at 65. Under this condition, the bit per pixel (BPP) was 0.085.

A comparison was made with the normal JPEG coding results. To unify the data size of JPEG for comparison with the proposed method, we compared our methods with two qualities of JPEG: 20 (BPP = 0.081) and 25 (BPP = 0.99).



(a) Original images

(b) Information-reduced images

(c) Restored images

Fig. 5 Experimental results

3.1 Subjective Evaluation

Four images were compared for subjective evaluation: the image before restoration, the restored image, and the two JPEG compressed images (quality = 20, 25). The original image was used as a reference image; it was compared with each test image, and the observer responded on a scale of 1 to 6 for image quality. There were 10 observers, each with normal color vision, and the images were displayed on a calibrated display (Eizo ColorEdge CG-221 BK) with a viewing angle of 4.9° for each image. Forty evaluation images were randomly selected from 1,415 images for each observer.

To compare the visual evaluation of the images by

each method, the scores of each observer were ranked in descending order, and the average ranking of 10 people for each method was calculated (Table 1). As shown in the table, the results reproduced by the proposed method were visually more similar to the original image than the JPEG images were under the same BPP, and there was a significant difference in the 5% significance level.

Table 1 Subjective evaluation results (average ranking by 10 observers)

Before restoration	Proposed restoration	JPEG (quality=20)	JPEG (quality=25)
2.6	1.5	2.7	2.0

3.2 Objective Evaluation

Because there is no standard visually appropriate evaluation index, PSNR and SSIM, which are normally used for image quality evaluation, were calculated. Table 2 shows the average values for the 1,415 images. JPEG showed better results for each index, which means that the results of the indices used did not reflect the visual evaluation results discussed in subsection 3.1.

Table 2 Objective evaluation results

(a) PSNR

Before restoration	Proposed restoration	JPEG (quality=20)	JPEG (quality=25)
20.4	20.5	27.6	28.4

(b) SSIM

Before restoration	Proposed restoration	JPEG (quality=20)	JPEG (quality=25)
0.48	0.49	0.81	0.84

4 Conclusions

In this paper, we proposed an image compression method consisting of the following three steps: (1) image information reduction considering visual characteristics with respect to human spatial frequency, (2) JPEG encoding/decoding, and (3) image restoration using U-Net deep learning. Through subjective evaluation experiments, we confirmed that the proposed method can produce a reproduction that is significantly more similar to the original image than the existing JPEG coding for the same data size. However, indices such as PSNR and SSIM yielded different results from subjective evaluations, highlighting the need for evaluation indices that can model human perception.

For our framework, JPEG was used as the encoding/decoding method, and U-Net was used for image restoration. We did not verify whether these methods are optimal in the framework of the proposed method, and it is possible to use different coding methods and deep learning methods. Determining the optimum method at each step is a potential task for future research on this topic.

References

[1] R. Salakhutdinov and G. Hinton, “Semantic Hashing,” Int. J. Approx. Reason. Vol. 50, No. 7, pp. 969-978 (2009).

[2] A. Krizhevsky and G. Hinton, “Using Very Deep Autoencoders for Content-Based Image Retrieval,” ESANN, (2011).

[3] Y. Ollivier, “Auto-encoders: reconstruction versus compression,” arXiv:1403.7752v2 (2015).

[4] K. Gregor et al., “Towards Conceptual Compression,” Proc. NIPS (2016).

[5] J. Ballé, V. Laparra and E.P. Simoncelli, “End-to-end optimization of nonlinear transform codes for perceptual quality,” Proc. Picture Coding Symposium (2016).

[6] G. Toderici et al., “Full Resolution Image Compression with Recurrent Neural Networks,” Proc. CVPR, Vol. 1, pp. 5435-5443 (2017).

[7] L. Theis et al., “Lossy image compression with compressive autoencoders,” arXiv:1703.00395v1 (2017).

[8] O. Ronneberger, P. Fischer and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,”, Proc. MICCAI, pp.234-241 (2015).

[9] S.J. Daly, “Engineering observations from spatiovelocity and spatiotemporal visual models,” Proc. SPIE, Vol. 3299, pp. 180-191 (1998).

[10] J. Xiao et al., “SUN Database: Large-scale Scene Recognition from Abbey to Zoo,” Proc. CVPR (2010).