# Wavelet Based Image Coding via Image Component Prediction Using Neural Networks

Takuma Takezawa and Yukihiko Yamashita

***Abstract*— In the process of wavelet based image coding, it is possible to enhance the performance by applying prediction. However, it is difficult to apply the prediction using a decoded image to the 2D DWT which is used in JPEG2000 because the decoded pixels are apart from pixels which should be predicted. Therefore, not images but DWT coefficients have been predicted. To solve this problem, predictive coding is applied for one-dimensional transform part in 2D DWT. Zhou and Yamashita proposed to use half-pixel line segment matching for the prediction of wavelet based image coding with prediction. In this research, convolutional neural networks are used as the predictor which estimates a pair of target pixels from the values of pixels which have already been decoded and adjacent to the target row. It helps to reduce the redundancy by sending the error between the real value and its predicted value. We also show its advantage by experimental results.**

***Index Terms*—Wavelet image coding, discrete wavelet transform, predictive coding, neural networks, super resolution.**

## I. Introduction

Recently, by the development of multimedia terminals, a large amount of information is produced every day. It causes the problems such as the shortage of memories. Thus, the methods to reduce the data redundancies have been investigated. For example, image compression technique such as JPEG and JPEG2000 has been developed.

JPEG is one of the most popular methods for image compression. In the coding process of JPEG, image is split into blocks of $8 \times 8$ pixels and each block is transformed by the discrete cosine transform (DCT) [1]. It causes a problem called block distortions that visual degradation is observed in its decoded image. In the case of block based transform algorithms, the prediction is also applied to blocks.

As a method to solve the problem of block distortions, a discrete wavelet transform (DWT) is used in the wavelet-based image coding technique such as JPEG2000 [2], [3]. However, in the wavelet based image coding, since the signals which should be predicted are apart from the pixels which can be used for prediction, the prediction is applied to DWT coefficients not to the image. Christopher *et al.* [4] used neural networks as a predictor of DWT coefficients. Since the wavelet decomposition is linear, nonlinear dependencies among DWT coefficients remains. Therefore, by predicting and eliminating such nonlinear dependencies using nonlinear predictor, the data size can be reduced. Because image have many features, its prediction will be effective comparing to the prediction of coefficients and we can improve coding efficiency by using the decoded part of an image. To solve this problem, Zhou and Yamashita proposed to use a block matching method for the prediction [5]. In the paper, they predict the image components on each row by using LL components and the image component which have already been decoded. In this case, DWT is performed row by row. Therefore, a segment of the target row is predicted by block matching to decoded row adjacent to the target rows.

In this research, we propose to apply the artificial neural networks to predict a target row by rows in the decoded image. CNNs are trained as a predictor which can estimate the values of a pair of target pixels on the target row from the values of its corresponding adjacent decoded pixels. The prediction accuracy is expected to improve because compared to the conventional method, which used decoded high frequency coefficients to predict the target high frequency coefficient, the decoded image right above on the target row and the low frequency components on the target row can be used to predict the pair of target pixels. This type of prediction can be considered as the super resolution of a row where high resolution image is given adjacent to the row. CNNs are trained offline by using several training images and rescaled as an extracted block. And then the trained predictor is implemented in set partitioning in hierarchical trees (SPIHT) encoder [6]. The coding experiment was conducted to introduce the result of coding efficiency and evaluate the usefulness of this research.

This paper is organized follows: Section II refers to the related works. Section III explains wavelet based image coding schemes and neural networks as a prediction technique. Section IV describes the proposed method. Section V shows the experimental results and Section VI summarizes our conclusion for this research.

## II. Related Works

Antonini, *et al.* [7] describes about image coding using wavelet transform and developed a compression method associated with a wavelet transform and a vector quantization coding scheme. The proposed scheme is taking into account the features both in the space and frequency domains. The wavelet transform is well adopted both obtaining a set of biorthogonal subclasses of images, and a progressive architecture which is presented to minimize the cost which the receiver has to reorganize the image.

Christopher, *et al.* [4] explored the use of neural networks to predict wavelet coefficients for image compression. Since the wavelet decomposition is linear, nonlinear dependencies

Manuscript received July 25, 2019; revised September 19, 2020.

The authors are with Tokyo Institute of Technology, School of Environment and Society, Meguro-ku, Tokyo 1528552 Japan (e-mail: takezawa.t.aa@m.titech.ac.jp, yamasita@tse.ens.titech.ac.jp).

doi: 10.18178/ijmlc.2021.11.2.1026

among wavelet coefficients remains. Therefore, they proposed to utilize neural networks as the predictor for such nonlinear dependencies. The paper presented wavelet image compression technique with the predictor which estimate target wavelet high frequency coefficients that are far different from images its corresponding coefficients which have been already reconstructed.

Zhou and Yamashita [5] proposed a framework of image coding by using DWT decomposition with prediction of lines by block matching. In the conventional 2D DWT decomposition, the prediction using images seems to be difficult since the pixels to be predicted are apart from the pixels that are decoded. To solve this problem, they predicted the image components respectively by using the low frequency image components and their image components that have already been decoded. Since these image components are adjacent to the target row, the prediction accuracy expected to improve.

## III. Image Coding

### A. Wavelet Based Image Coding

Wavelet-based image coding is one of the image compression methods using DWT.

DWT is a method to analyze signals, divides them into different frequency components, and subsample them. As a result, low and high frequency coefficients of DWT are obtained. The amount of data can be reduced by encoding the coefficients of DWT.

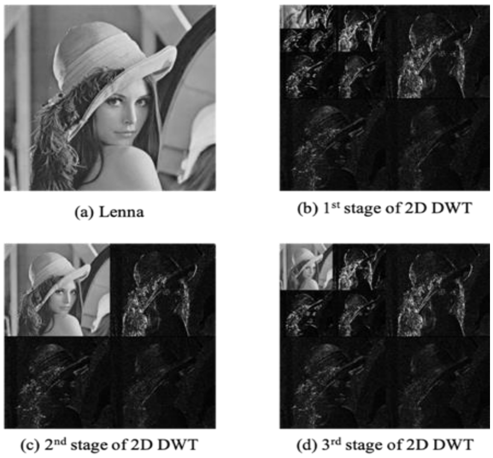


Fig. 1. 2D DWT sub-band decomposition.

One-dimensional DWT can be extended to two dimensions (2D DWT). Since the image data can be thought as a 2D signal of pixel value, 2D DWT is used for the transformation for digital images. A digital image can be represented by 2D array $X = [x_{m,n}]$, with *m* rows and *n* columns. Note that both *m* and *n* is an integer. Firstly, the low frequency coefficients LI and the high frequency coefficients HI can be obtained by performing 1D DWT vertically to the original image. Then repeatedly 1D DWT is preformed horizontally to obtain DWT coefficients LL, LH, HL, HH [8]. Since the low frequency coefficients represent approximate part of the input image, LL sub-bands can be used as the input of further 2D DWT to obtain the multi stage of DWT sub-band decomposition. Fig. 1 shows the three stages of 2D DWT decomposition of a standard image Lenna.

### B. Predictive Coding

Predictive coding is a compression method used for text and image data. It focuses on the correlations between adjacent signals. The first information theoretic treatment of compressing data through the use of prediction was done by Elias, who called the technique, the predictive coding [9]. In this research, a feedforward neural network consists of a single convolution layer and a single fully-connected hidden layer is used to obtain the optimized mapping (prediction function) from the input to its corresponding targets.

### C. Feedforward Neural Networks

Feedforward neural networks is a learning system inspired by the biological neuron structure in natural animals' brains. The goal of a feedforward network is to approximate a function $f^*$ [10]. A feedforward network defines a mapping $y = f(x;\theta)$ and learns the values of the parameters $\theta$ that result in the best function approximation.

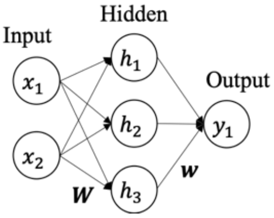


Fig. 2. Simple feedforward neural networks with three layers.

Convolutional neural networks, or CNNs, are a specialized kind of neural networks for processing data that have a known grid-like topology. For example, image data can be thought of as a 2D grid of pixels. The name "convolutional neural network" indicates that the network employs a mathematical operation called convolution.

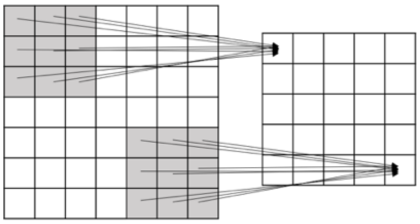


Fig. 3. Convolutional operation.

Since the convolutional operation can work as feature filters for its input, CNNs have an advantage especially in image processing because it can refer to the features of images in comparison with feedforward neural networks.

## IV. Proposed Method

### A. Coding Scheme

Here, we describe the coding scheme using the DWT decomposition with prediction in detail. The fundamental process of encoding is as follows:

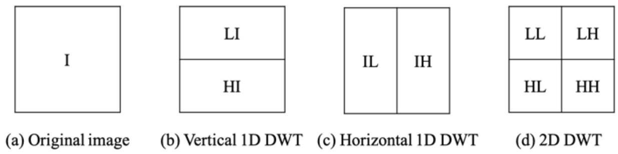
- Restore the low frequency component of target row from low frequency coefficients.
- Predict the target row by using the reference rows and the low frequency component of target row.

- Transform the predicted row by DWT and obtain high frequency coefficients.
- Output the quantized difference between the true high frequency coefficients and the predicted ones.

### B. One-dimensional DWT for Prediction

To obtain a pixel value by the inverse DWT for decoding, several coefficients around the pixel are needed. Therefore, the prediction with the 2D DWT decomposition using the decoded image has been considered difficult because we have to predict the values of pixels that are apart from the decoded pixels by not less than several pixels.

To solve this problem, Zhou *et al.* proposed to apply prediction into the 1D DWT part in the 2D DWT decomposition. Different with 2D DWT decomposition, horizontal and vertical transforms are performed separately in 1D DWT, but not at the same time. Fig. 4 shows the 1D DWT decomposition and its coefficients. For instance, the input image, I, is decomposed vertically and horizontally into LI and HI. Here, L represents for low and H represents for high frequency coefficients. Then, LI and HI are decomposed horizontally into LL, LH, HL, and HH. Comparing to LH and HL, LI and IL are considered to be images.



(a) Original image (b) Vertical 1D DWT (c) Horizontal 1D DWT (d) 2D DWT

Fig. 4. Coefficients of 1D DWT decomposition.

In order to explain the aim of using 1D DWT for prediction, we define some notation for convenience. Here, let *M* be the times of decomposition, $C_{XXm}$(XX = LL, LH, HL, HH, LI, HI, IL, IH, and $m$=1, 2, …, $M$) denotes the coefficients in *XXm*. And $C_{LL0}$ denotes the original image. In general image coding, we send $C_{LLM}, C_{LHM}, C_{HLM}, C_{HHM}, C_{LH(M\text{-}1)}, C_{HL(M\text{-}1)}, \dots, C_{HH1}$. Then, decoding is done by reconstructing $C_{LL(m\text{-}1)}$ from $C_{LLm}, C_{LHm}, C_{HLm}, C_{HHm}$ for $m$=1, 2, …, $M$.

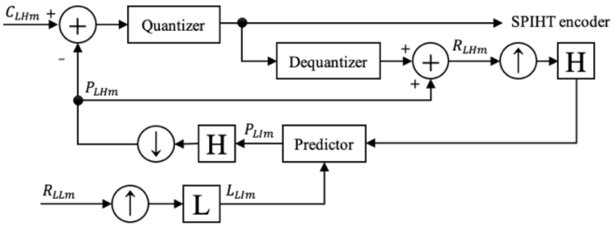


Fig. 5. Block diagram of prediction process in encoding.

Consider that $C_{LIm}$ is decomposed into $C_{LLm}$ and $C_{LHm}$in 1D DWT. The decomposition is done row by row from top to bottom in $C_{LIm}$. $C_{LLm}$ and the rows in $C_{LIm}$ above the target row can be decoded. If it is possible to predict the target row in $C_{LIm}$, the predicted value in $C_{LHm}$ in the row can be obtained by DWT. Since the target row is adjacent to the above one, the target pixel and its corresponding reference pixels which can be used for prediction are adjacent. Therefore, the predictive coding is available for 1D DWT and the data size can be reduced by sending the quantized difference between the real target value and its predicted value, not the quantized value of $C_{LHm}$. In this method, the improvement in prediction accuracy is expected because compared to the conventional methods which predict the high frequency coefficients from high frequency coefficients, the decoded image is used to predict the decoded image itself.

### C. Prediction Using Neural Networks

In this research, convolutional neural networks are used as a predictor for predictive coding.

In order to predict the target pixels on the target row, the extracted blocks are set. The extracted block is shown in Fig. 6.

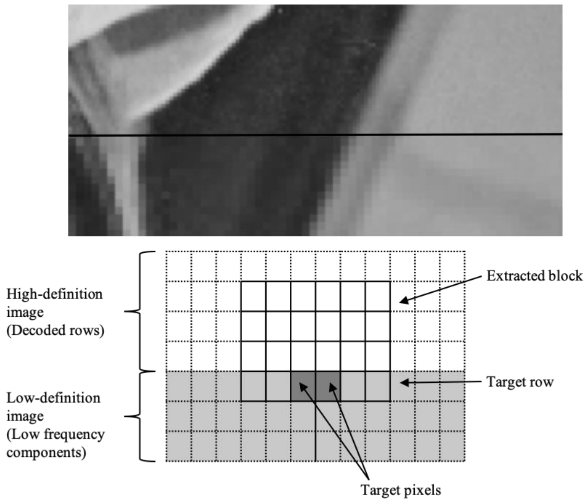

High-definition image (Decoded rows)
Low-definition image (Low frequency components)
Extracted block
Target row
Target pixels

Fig. 6. Block extraction with size of 6×4.

In the process of signal decomposition by DWT, the fundamental component has both of up-sampling and down-sampling with the factor of two. Therefore, the target pixels of low frequency image components at even and odd coordinates should be extracted and gathered separately so that they have different features. To solve this problem, we propose to choose two pixels in the middle of the target row as a target pixel pair. Here, some notations are defined for convenience:

- $b_x$: the width of the extracted block
- $b_y$: the height of the extracted block

In this research, $b_x$ and $b_y$ are set as 6 and 4 respectively. Rows from 1 to $b_y$ - 1 of the block are extracted from the decoded image components of $C_{LIm}$, which can be considered as reconstructed image (High-definition image). And the last row is extracted from that of up-sampled and filtered by using only low frequency coefficients (Low-definition image). To predict the value of the pair of target pixels, we use neural networks. Since this prediction can be considered as super-resolution problem in case that the high definition image on the right above of the target row is obtained, CNNs are used because convolutional layers can work as a feature detector in images.

## V. Experimental Results

### A. Image Data

In this experiment, we used 8 standard images with different size (Fig. 7(a) ~ (h)) as training data for neural network, Lenna (Fig. 7(i)) as the evaluation data for the training of CNNs, and 4 standard images with size of $512\times512$ (Fig. 8) as the test data.

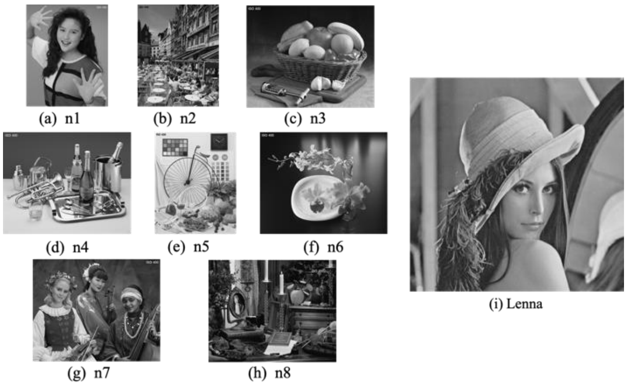


Fig. 7. Training and evaluation data.

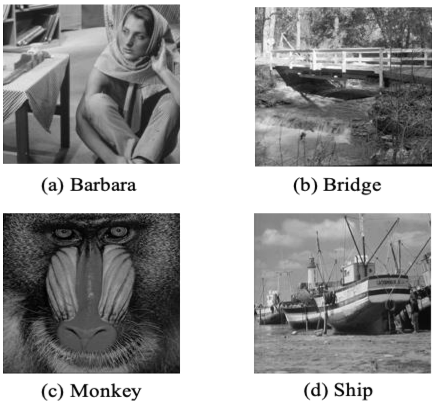


Fig. 8. Test data.

## B. Network Structure

We used CNNs consisted of one convolutional layer and one hidden layer. All the training data is decomposed to the levels of five by 2D DWT, and a network is trained for each wavelet level. As the input data, each value in the extracted block with size of $6\times4$ was used. For the convolution part, we chose to use 64 kernels with size of $3\times3$. To train the network, Adam [11] was used as optimization method, and the learning rate was set to 0.001.

## C. Evaluation Methods

To evaluate the prediction efficiency and the coding efficiency, the root mean squared error (RMSE) of a predicted image and the rate distortion curve (bitrate vs. PSNR of a decoded image) are used respectively.
RMSE can be given by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i^2 - \hat{y}_i^2)}\ , \qquad (3)$$

Here, $y_i$ and $\widehat{y_i}$ represent the target value and the predicted value respectively. Since it represents an error between $y_i$ and $\widehat{y_i}$, the smaller RMSE leads better efficient of prediction.

PSNR, or Peak Signal-to-Noise Ratio, is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation [12]. PSNR (in dB) is defined as:

$$PSNR = 10\log_{10}\left(\frac{MAX_I^{\,2}}{MSE}\right) = 20\log_{10}\left(\frac{MAX_I}{RMSE}\right), \qquad (4)$$

Here, $MAX_I$ is the maximum possible pixel value of the image. In this research, since only the gray scale images are used, $MAX_I$ takes 255. And $MSE$ is the mean squared error. Generally, the higher PSNR is, the better coding efficiency is. Therefore, the higher PSNR at the same bit-rate means a better coding efficiency.

## D. Prediction Efficiency and Coding Results

Table I and Fig. 9 show the result of RMSE of evaluation data "Lenna" for the predicted image by neural networks. We used from n1 to n8 as training data and set the number of kernels in convolution layer from 8 to 128 to evaluate which size of network can obtain the best prediction result.

TABLE I: RMSE FOR DIFFERENT NUMBER OF KERNELS

| Number of kernels | RMSE |
|---|---|
| 8 | 3.222 |
| 16 | 3.087 |
| 32 | 3.467 |
| 64 | 3.244 |
| 128 | 3.444 |

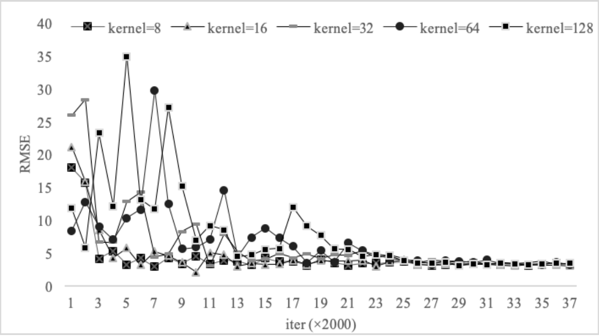
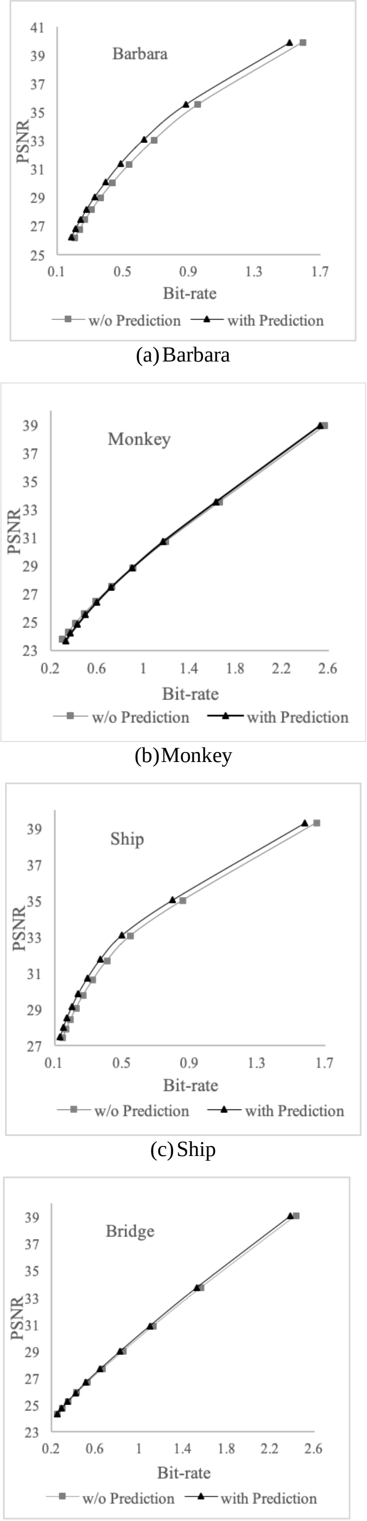


Fig. 9. RMSE for different number of kernels.

According to the results, since the RMSE of each kernel number does not change dramatically, the coding efficiencies for the number of kernels seems to be close. However, when we conducted the coding experiments on each number of kernels, the predictor with 64 kernels has the best result of compression ratio. Therefore, we chose to use 64 kernels for the convolutional layer.

Fig. 10 and Table II show the coding results. To obtain the rate distortion curve, we changed the quantization factor q. Among the four test images, the proposal method outperformed the conventional one (without prediction). According to the results, the compression ratio varied in each test image. For example, in cases of "Barbara" and "Ship", the bit-rate is reduced around 9% in the rate distortion curve. However, in the other cases, the compression ratio is not as large as "Barbara" or "Ship". The reason can be considered that "Monkey" or "Bridge" has more complex texture such as animal hair or leaves. And since those features seemed not to be contained so many in the leaning images, the prediction could not work as well as the cases of "Barbara" and "Ship". This result is about the same as that of Christopher, *et al.* [4], which predict the efficiency directly, but still the super-resolution using prediction can improve the coding efficiency.



(a) Barbara

(b) Monkey

(c) Ship

(d) Bridge

Fig. 10. Rate distortion curve.

TABLE II: COMPRESSION RATIO FOR EACH TEST IMAGE

| Test images | Compression ratio (%) |
|---|---|
| Barbara | 9.015 |
| Monkey | 1.854 |
| Ship | 9.017 |
| Bridge | 2.963 |

## VI. CONCLUSION

In this paper, the neural network is used as a predictor for the wavelet based image coding. In the framework of this prediction, the prediction can be considered as the super resolution problem since the pair of target pixels was predicted from its adjacent pixels in high-definition image and low-definition image on the same row of the targets. The network is trained with 8 standard images for each wavelet level off-line and the trained network is used both encoding and decoding schemes. The compression test is conducted for 4 test images.

According to the results, the coding efficiencies improved in every test image. Especially in cases of “Ship” and “Barbara”, the proposed method succeeded in reducing the data sizes by prediction more than those of other cases. The reason for this result can be considered that the prediction can work on the edges in the image, and the predictor could not learn the features contained especially in “Monkey” and “Bridge” enough from the learning images.

It is possible to improve the coding efficiency by a more precise prediction. And to improve the prediction accuracy, it is necessary to optimize the network architecture. Since the learning data is contained with high-definition image (decoded image) and low-definition image (target row), the features in which each region has might be different. Therefore, the simple CNNs may not work as well, and it may be possible to learn well by using the different architecture of neural network for each region. Furthermore, as the results show, the compression ratio will depend on the features on the learning data. Therefore, the coding efficiency can improve by using more leaning data which contain different types of features in it.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

Takezawa proposed, optimized, and implemented the neural network for the prediction, conducted experiments, and analyzed the results. Yamashita proposed the framework of wavelet image coding with prediction and implemented it except the neural network.

## ACKNOWLEDGEMENT

A part of this work was supported by JSPS KAKENHI Grant Numbers17H01760.

## REFERENCES

[1] W. B. Pennebacker and J. L. Mitchell, “JPEG still image compression standard,” *Van Nostrand Reinhold*, New York, 1992.

[2] H. Kobayashi, and E. R. Bahl, “Image data compression by predictive coding,” *IMB J. Res. Dev.*, vol.18, pp. 164, 1974.

[3] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelet and Wavelet Transfroms: A Primer*, Englewood Cliffs: Prentice Hall, NJ: Upper Saddle River, 1997.

[4] C. J. C. Burges, P. Y. Simard, and H. S. Malver, “Improving wavelet image compression with neural networks,” *Microsoft Research Tech. Rep.*, MSR-TR-2001-47, pp. 1-18, 2001.

[5] S. Zhou and Y. Yamashita, “Image coding using discrete wavelet transform decomposition with prediction of half-pixel line Segment matching,” in *Proc. Picture Coding Symposium of Japan*, 2013, no. 27.

[6] A. Said and W. A. Pearlman, “A new, fast, and efficient image codec based on set partitioning in hierarchical trees,” *IEEE Transactions on Circuits and System for Video Technology*, vol. 6, no. 3, pp. 243-250, 1996.

[7] M. Antonini, M. Barland, P. Mathieu, and I. Daubechie, “Image coding using wavelet transform,” *IEEE Trans. on Image Processing*, vol. 1, pp. 205-220, 1992.

[8] A. Goel, “Discrete wavelet transform (DWT) with two channel filter bank and decoding in image texture analysis,” *International Journal of Science and Research*, vol. 3, pp. 391-397, 2014.

[9] Y. Huang and R. P. N. Rao, “Predictive coding,” *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 2, pp. 580-593, 2011.

[10] I. Goodfellow, Y. Bengio, and A. Courvile, *Deep Learning*, MA: MIT Press, Cambridge, 2017.

[11] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” in *Proc. International Conference of Learning Representations*, arXiv:1412.6980, 2015.

[12] Q. Huynh-Thu and M. Ghanbari, “Scope of validity of PSNR in image/video quality assessment,” *Electronics Letters*, vol. 44, pp. 800-801, 2008.

**Takuma Takezawa** was born in 1995 in Shibuya-Ku, Tokyo, Japan. He received the B.E. from Tokyo Institute of Technology in 2018. He is majoring in computer science.

He is now in the second year of a master’s degree in Tokyo Institute of Technology, School of Environment and Society, Department of Transdisciplinary Science and Engineering, Global Engineering for Development Environment and Society Graduate Major. From 2017, he researches on Image coding by using deep learning at Tokyo Institute of Technology. His research interests in deep learning, neural networks, machine learning, and image processing.

**Yukihiko Yamashita** was born in 1960 in Kawasaki-Shi, Kanagawa Prefecture, Japan. He received the B.E., the M.E., and the Dr. Eng. degrees from Tokyo Institute of Technology in Tokyo Japan in 1983, 1985, and 1993, respectively.

From 198 to 1988, he was a researcher in the Japan Atomic Energy Research Institute. From 1988 to 1989, he was a researcher in the ISAC corporation. In 1989 he joined the faculty of the Tokyo Institute of Technology, where he is now an associate professor of the School of Environment and Society. His research interests in image processing, pattern recognition, and machine learning.

Prof. Yamashita is a member of the Institute of Electrical and Electronics Engineers (IEEE), the Institute of Electronics, Information, and Communication Engineers (IEICE) of Japan, Information Processing Society of Japan (IPSJ). He was editor-in-chief of IEICE Transactions of Information and Systems from 2017 to 2019. He received a Paper Award in 1993 from IEICE.