# Subjective Quality Driven Image Encoding Method Using Image Completion

Shota Orihashi, Shinobu Kudo, Ryuichi Tanida, Hideaki Kimata
NTT Media Intelligence Laboratories, NTT Corporation, Japan
E-mail: shota.orihashi.bt@hco.ntt.co.jp

***Abstract*—This paper presents a still image coding method using deep learning-based image completion. Deep learning-based image completion can restore skipped areas of images in high quality. When we introduce image completion to image coding, it is possible to reduce the coded-bit amount compared with normative codec-based methods by replacing complex areas, such as textures, with simple signal values at the encoder and completing them at the decoder. However, there is no method for automatically detecting the skipped areas at the encoder because we cannot evaluate the quality of completion by objectively comparing the signal difference between the original and completed images. To resolve this issue, we propose an image quality estimation model without referencing original images. Our key idea is to obtain the model by adversarial training with a completion network assuming that original images have higher quality than the completed ones. We also propose a detection algorithm for skipped areas. Our algorithm detects skipped areas by giving priority to complex areas where a large coded-bit amount is required for the normative codec-based method to increase coding efficiency. The proposed method reduced the coded-bit amount by 25% compared with an HEVC-based method while maintaining the subjective quality for particular images.**

## I. Introduction

Normative video coding standards have been established by predicting and transforming images. They predict images from original input images by utilizing the redundancy that occurs in the spatial and temporal directions. The informational amount is then reduced by sending only residual images transformed from the pixel domain to the frequency domain. This architecture is adopted in H.265/HEVC [1] and will be used in the next-generation standard VVC [2]. For still images, the file format to encode using HEVC is also defined and widely used as HEIF [3].

Methods based on normative coding do not perform well when they are applied to images that include complex textures, such as trees and water surfaces. When images including such textures are encoded, prediction efficiency becomes lower than that obtained when encoding simple images. As a result, the informational amount for residual images increases and the image quality degrades since encoding is performed within the target bitrate.

One method to address this problem is generating images at a decoder to maintain subjective image quality. The decoder counterpart (i.e., the encoder) edits original images to reduce the informational amount they contain. This is done especially for complex areas or areas where there is no need to reproduce by exact pixel value, such as texture regions. For example, texture synthesis-based methods [4–7] use texture synthesis at the decoder to reduce the coded-bit amount while maintaining subjective quality at the texture region. Similarly, seam carving-based methods [8, 9] use seam carving at the encoder and reconstructed seams at the decoder utilizing interpolation. Although both methods can reduce the coded-bit amount, their available areas are limited to uniform textures or simple areas, respectively.

To generate varied areas, deep learning-based image completion [10–14] has been investigated recently. For example, Pathak et al. [10] proposed an image completion method that utilizes a convolutional neural network (CNN) by applying the generative adversarial network (GAN) framework [15]. Iizuka et al. [11] improved the quality of completed images by using two discriminators that only focus on local or global components.

Deep learning-based image completion enables high-quality output for various images. By replacing complex areas of the input images with particular easy-to-encode signal values at the encoder and completing them at the decoder, it is possible to reduce the coded-bit amount while maintaining subjective quality compared with normative codec-based methods. However, there are areas where completion does not perform well, including large areas. Therefore, we need to evaluate the quality of the completed images, but the quality cannot be evaluated using the signal difference between the original and completed images, as used in the encoder of normative codec-based methods, because completed images significantly differ from their original images. Because of this issue, a method for detecting the skipped areas to be completed for efficient encoding has not yet been established.

In this paper, we propose a still image encoding method with an image-coding framework for executing image completion at the decoder. Our method automatically detects skipped areas and optimizes the bitrate and subjective quality of the completed image by using two techniques. First, to overcome the issue of quality evaluation, we introduce a scoring model called the image quality estimation (IQE) model. Our idea is to train the IQE model by adversarial training with a completion network assuming that original images have higher quality than the completed ones. Second, to reduce the coded-bit amount efficiently, we apply a detection algorithm for skipped areas. Our algorithm detects skipped areas focusing on complex areas that are defined by coded-bit amount when encoding with the HEVC-based method. Our experimental

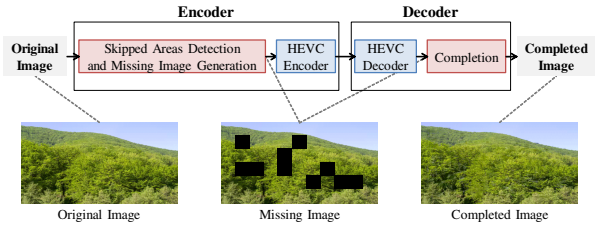



Fig. 1. Image-coding framework with image completion.

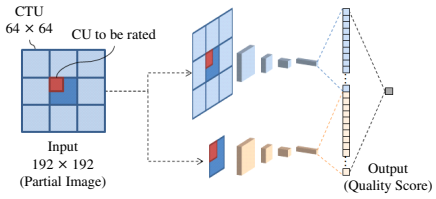



Fig. 2. Image quality estimation model.

results demonstrate that the proposed method reduces the coded-bit amount by 25% compared with using an HEVC-based method while maintaining the same perceptual quality for particular images.

In Section 2 of this paper, we briefly explain GAN-based image completion. In Section 3, we describe the proposed method. In Section 4, we present the details of the simulation evaluation and results. We conclude in Section 5 with a brief summary.

## II. GAN-Based Image Completion

In this section, we briefly go over GAN-based image completion. Image completion is effective in restoring the skipped areas of images, and recently, deep learning approaches [10–14] based on the GAN framework [15] have provided high-quality completed images. In these approaches, a CNN is trained for completion (completion network). The input of the completion network is an image with skipped areas (missing image) and the output is the completed image. Binary masks that indicate skipped areas to combine the missing and completed images are also used.

When these approaches train a completion network, they use another CNN that discriminates the original and completed images (discriminator). Suppose that completion network $C(x, M)$ processes completion from the original image $x$ and mask $M$, which indicates skipped areas. Discriminator $D(x)$ estimates the probability that input image $x$ is original. In this scenario, the objective of adversarial training is to solve the following min-max optimization problem:

$$\min_C \max_D \mathbb{E}\left[\alpha L(x, M) + \log(1.0 - D(x)) + \log D(C(x, M))\right]. \quad (1)$$

Here, $L(x, M)$ is the mean squared error (MSE) of completed image $C(x, M)$ and original image $x$ represented as

$$L(x, M) = \|C(x, M) - x\|^2, \quad (2)$$

and $\alpha$ is a parameter that balances the MSE and adversarial loss for the completion network. By training a completion network and discriminator alternately while varying $x$ and $M$, the completion network generates natural completed images that the discriminator cannot distinguish.

## III. Proposed Method

### A. Image-Coding Framework with Image Completion

We first describe the image-coding framework of the proposed method, which applies image completion. An overview is shown in Fig. 1. At the encoder, skipped areas to be completed at the decoder are first detected (the detection algorithm is detailed in the next section). Then, a missing image is generated by alternating the pixel values of skipped areas with easy-to-encode values such as the average values of each area. An encoder of HEVC encodes the missing image and sends it as a bit stream. The positions of skipped areas are also sent as side information. At the decoder, the missing image is decoded and skipped areas are completed. We assume that image completion is done with a CNN using a GAN-based training technique (such as [11]).

### B. Detection Algorithm for Skipped Areas

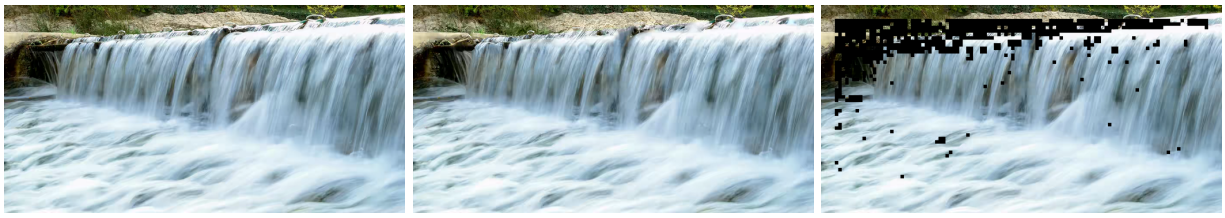
Our encoding method detects skipped areas in the image by optimizing the coded-bit amount and subjective quality while focusing on hard-to-encode areas for HEVC. Skipped areas are detected on the basis of HEVC coding units (CUs) to ensure that missing images are efficiently encoded by HEVC.

In the detection algorithm, the original input image is first encoded using an HEVC encoder. In this first encoding, the structure of the CU partition, the coded-bit amount for each CU, and the encoded image are stored. Then, the evaluating order of CUs is determined by descending order of the coded-bit amount for each CU. Giving priority to CUs with a large coded-bit amount allows the proposed method to preferentially complete hard-to-encode areas for HEVC.

By setting the target CU according to the determined order, evaluating whether the target CU is skipped is conducted repeatedly. During this evaluation, the quality score $Q(p)$ of the partial image $p$ that includes the CU to be rated and the surrounding encoded image is measured using the IQE model (described in the next section). The target CU is determined as a skipped area when the following requirements are satisfied:

- **Reasonability to skip the target CU:** We evaluate reasonability to skip the target CU by considering the quality score and the coded-bit amount. Let $p^t_{Comp}$ be a partial image including the target CU that was skipped and completed using the completion network, and $p^t$ be a partial image including the target CU without skipping. Also, let $R$ be the reduced number of bits for the target CU when the target CU is skipped. The first requirement

(a) Test 1: $-24.6\%$



(b) Test 2: $-27.6\%$

(c) Test 3: $-23.3\%$

Fig. 3. Result images with reduced coded-bit amount. Left: decoded images with base-HEVC method. Center: completed images with proposed method. Right: missing images with proposed method.

TABLE I
COMPARISON OF CODED-BIT AMOUNT, OBJECTIVE SCORES, AND MOS.

| | Test 1 | | | | Test 2 | | | | Test 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coded Bits | PSNR | MS-SSIM | MOS | Coded Bits | PSNR | MS-SSIM | MOS | Coded Bits | PSNR | MS-SSIM | MOS |
| Base-HEVC | 941,520 | 30.48 | 0.97 | 3.60 | 259,016 | 33.98 | 0.96 | 2.60 | 599,944 | 31.15 | 0.97 | 3.33 |
| Comparison-HEVC | 666,424 | 30.11 | 0.95 | 3.00 | 192,864 | 33.41 | 0.95 | 2.33 | 433,208 | 30.64 | 0.95 | 3.00 |
| Proposed | 710,318 | 29.78 | 0.76 | **3.47** | 187,536 | 33.33 | 0.91 | **2.60** | 459,866 | 30.44 | 0.86 | **3.07** |

to determine the target CU as a skipped area is to satisfy the following inequality:

$$Q(p^t_{Comp}) + \lambda R > Q(p^t), \tag{3}$$

where $\lambda$ is based on the compression rate and set in advance.

- **No significant decrease of completion quality on surrounding skipped areas:** If the target CU is determined as a skipped area, the quality score of its surrounding CUs that are already determined to be a skipped area may decrease. Therefore, we re-evaluate whether each surrounding CU's quality score is decreased. Let $p^s_{Comp}$ be a partial image including the surrounding CU that was skipped and completed together with the target CU. Also, let $p^s$ be a partial image including the surrounding CU that was skipped and completed by itself. The second requirement to determine the target CU as a skipped area is to satisfy the following inequality for all surrounding CUs belonging to the skipped area:

$$Q(p^s_{Comp}) > \mu Q(p^s), \tag{4}$$

where $\mu$ is a parameter that defines an acceptable score decrease.

### C. Image Quality Estimation Model

When we measure image quality with the proposed method, it is not appropriate to use a pixel-based difference, such as the MSE, due to drastic image changes caused by completion. The encoding method therefore measures image quality by using a trained IQE model without referencing original images.

Figure 2 shows such a model developed on the basis of a CNN. The input is a partial image $p$ that includes one CU to be rated in the central coding tree unit (CTU) and surrounding image. In the same way as for the discriminator Iizuka et al. reported [11], the input is separated into global and local parts. The IQE model outputs the estimated quality score $Q(p)$.

The IQE model is trained using the GAN framework with the completion network. By assuming the original images have higher quality than the completed ones, the proposed method trains the IQE model by replacing the discriminator (described in Section 2) with the IQE model.

## IV. Experiments

We conducted simulation experiments on the proposed method. We first compared the coded-bit amount and subjective image quality of the proposed method and HEVC-based methods, namely, base-HEVC and comparison-HEVC methods. Then we evaluated the effectiveness of the proposed encoding techniques.

### A. Experimental Conditions

We obtained experimented images in the public domain from Flickr [16] with resolutions of 1856 $\times$ 960 pixels. We set the minimum size of the skipped area to 16 $\times$ 16 pixels. The skipped areas' pixel values were replaced by the encoder with average values of each area, based on the coding efficiency of a preliminary experiment. We used HM-16.0 [17] for HEVC encoding. HEVC applies a constraint quantization parameter (QP) and all-intra mode. The QP for the base-HEVC method and proposed method was set to 37. We encoded the comparison-HEVC method and the proposed method to reduce the coded-bit amount by around 25% compared to the base-HEVC method. We selected $\lambda$ for the proposed method and QP for the comparison-HEVC method to satisfy the target coded-bit amount. Parameter $\mu$ for the proposed method was set to 0.95. The deblocking filter was disabled in the experiments. We applied the proposed method only for luminance components and HEVC was used to encode all chroma components. Binary flags for each CU, which indicate whether the CU was skipped, were sent as side information and counted in the coded-bit amount of the proposed method.

Architectures of the IQE model and completion network were based on the discriminator and completion network of [11], respectively. For the completion network, we introduced the architecture of U-Net [18] to improve the quality of completed images. For training the completion network and the IQE model, we used 830,000 192 $\times$ 192 patches made from the DIV2K dataset [19]. The completion network trained 40 epochs, while the IQE model trained 35. The optimizer was AdaDelta [20] and the batch size was 16.

Subjective evaluation was also conducted using a five-point absolute category rating, where the image quality was scored in five levels for each image independently. The number of participants was 15, and other settings followed [21].

### B. Results

Figure 3 shows the images obtained with the base-HEVC method, and completed and missing images obtained with the proposed method indicating reduced coded-bit amount. Skipped areas are drawn in black for the missing images. Note that while the base-HEVC method and missing images were both encoded by setting QP to 37, the coded-bit amount decreased by around 25% with the proposed method. The results lead us to conclude that desired completion results were obtained, as we could not recognize which images the proposed method completed at a glance.

(a) MS-SSIM (b) Raster scanning (c) Proposed

Fig. 4. Effect of IQE model and coded-bit-based evaluating order.

Table 1 lists the coded-bit amount, peak signal-to-noise ratio (PSNR), multi-scale structural similarity (MS-SSIM) [22], and mean opinion score (MOS) of the subjective evaluation for tested images. Note that the PSNR and MS-SSIM are not appropriate to evaluate the proposed method since completion quality cannot be measured with them. As Table 1 shows, the MOSs of the proposed method were higher than those of the comparison-HEVC method for the same bitrate. Significantly, for Test 2 the proposed method's MOSs equaled those of the base-HEVC method. However, for Test 3 they were worse than those for other tests. Some viewers said the discontinuity between waterfall and trees was noticeable. One possible solution would be to introduce an image segmentation technique not to detect skipped areas on object boundaries.

We also evaluated the effectiveness of the proposed method by applying the IQE model with coded-bit-based evaluating order. Figure 4 shows a comparison of completed images at an equal bitrate encoded using MS-SSIM instead of the IQE model to measure image quality, evaluating CUs to detect skipped areas by raster scanning order instead of coded-bit-based order, and using the proposed method. Although MS-SSIM provides close to subjective image quality, it also includes noticeable noise, as shown in the figure. Also, since the raster scanning order-based method requires large skipped areas to encode within the target bitrate, we obtained poor image quality. However, the proposed method enabled us to reduce unwanted noise.

## V. Conclusion

We have proposed an image-encoding method for executing image completion at the decoder. Our method detects skipped areas to be completed using a cost function with a trained scoring model and processes the order in which coding efficiency increases. With this method, we could reduce the coded-bit amount by 25% while maintaining subjective quality for particular images.

## REFERENCES


[1] Recommendation ITU-T H.265 | ISO/IEC 23008-2, "High Efficiency Video Coding," Apr. 2013.

[2] B. Bross, J. Chen, S. Liu, and Y. K. Wang, "Versatile Video Coding (Draft 9)," JVET-R2001, Apr. 2020.

[3] ISO/IEC 23008-12, "Image File Format," Aug. 2015.

[4] Y. Shi, Y. Hou, B. Yin, and W. Ding, "Image Coding Approach based on Image Decomposition," *Picture Coding Symposium (PCS),* pp. 534–537, Dec. 2010.

[5] F. Racape, S. Lefort, D. Thoreau, M. Babel, and O. Deforges, "Characterization and Adaptive Texture Synthesis-based Compression Scheme," *European Signal Processing Conference (EUSIPCO),* pp. 6–10, Aug. 2011.

[6] U. S. Thakur and B. Ray, "Image Coding using Parametric Texture Synthesis," *IEEE International Workshop on Multimedia Signal Processing (MMSP),* Sep. 2016.

[7] B. Wandt, T. Laude, Y. Liu, B. Rosenhahn, and J. Ostermann, "Extending HEVC using Texture Synthesis using Detail-aware Image Decomposition," *Picture Coding Symposium (PCS),* pp. 144–148, Jun. 2018.

[8] Y. Tanaka, M. Hasegawa, and S. Kato, "Image Coding using Concentration and Dilution based on Seam Carving with Hierarchical Search," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* pp. 1322–1325, Mar. 2010.

[9] M. Decombas, F. Dufaux, E. Renan, B. Pequet-Popesu, and F. Capman, "Improved Seam Carving for Semantic Video Coding," *IEEE International Workshop on Multimedia Signal Processing (MMSP),* pp. 53–58, Sep. 2012.

[10] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context Encoders: Feature Learning by Inpainting," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 2536–2544, Jun. 2016.

[11] S. Iizuka, E. Simoserra, and H. Ishikawa, "Globally and Locally Consistent Image Completion," *ACM Transactions on Graphics,* vol. 36, no. 4, Jul. 2017.

[12] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative Image Inpainting with Contetual Attention," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 5505–5514, Jun. 2018.

[13] G. Liu, F. A. Reda, K. J. Shih, T. C. Wang, A. Tao, and B. Catanzaro, "Image Inpainting for Irregular Holes using Partial Convolutions," *European Conference on Computer Vision (ECCV),* pp. 85–100, Sep. 2018.

[14] C. Zheng, T. J. Cham, and C. Jianfei, "Pluralistic Image Completion," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 1438–1447, Jun. 2019.

[15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," *Advances in Neural Information Processing Systems (NIPS),* pp. 2672–2680, Dec. 2014.

[16] "Flickr," https://www.flickr.com/.

[17] "HEVC reference software," https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI),* vol. 9351, pp. 234–241, Oct. 2015.

[19] E. Agustsson and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops,* Jul. 2017.

[20] M. Zeiler, "ADADELTA: An Adaptive Learning Rate Method," *arXiv preprint,* arXiv: 1212.5701, Dec. 2012.

[21] Recommendation ITU-T P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications," Apr. 2008.

[22] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale Structural Similarity for Image Quality Assessment," *IEEE Asilomar Conference on Signals, Systems, and Computers (ACSSC),* vol. 2, pp. 1398–1402, Nov. 2003.