Implicit Neural Representation-based Hybrid Digital-Analog Image Delivery

Akihiro Kuwabara*, Yutaro Osako†, Sorachi Kato†,
Takuya Fujihashi†, Toshiaki Koike-Akino‡, Takashi Watanabe†
*School of Engineering, Osaka University, Japan
†Graduate School of Information Science and Technology, Osaka University, Japan
‡Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA 02139, USA

***Abstract*—Implicit Neural Representation (INR) is an emerging technology for representing multimedia signals, such as RGB images, with small data size. A key issue in INR is the accuracy of high-frequency details in RGB images under the limited data size. Many studies in machine learning have discussed activation functions and coding to improve the accuracy. This paper aims at the same goal and proposes a novel communication-oriented solution for INR. To represent high-frequency details to the user, the proposed scheme exploits analog transmission for the residual signals between the original image and the decoded image derived from the trained INR. The integration of the INR and analog transmission provides high-frequency details to users with low traffic, and further improves the image quality as a function of the wireless channel quality between the transmitter and each user. Evaluations using an RGB image dataset show that the proposed scheme achieves better image quality than the existing INR-based image compression and standard image codecs under the same amount of traffic.**

I. Introduction

Representing multidimensional signals in a memory-efficient format is a key technology for transmitting high-resolution, high-dimensional multimedia signals with small data sizes. Implicit Neural Representation (INR) [1]–[4] is a novel memory-efficient format for representing multidimensional signals. INR consists of a small Multi-Layer Perceptron (MLP)-based Neural Network (NN) architecture, and then the perceptron overfits the target multimedia signal, such as an RGB image. The INRs require less data traffic than explicit representations such as pixels and 3D points. Recent studies [5], [6] used the INR for image compression and demonstrated large traffic reduction compared to traditional image coders.

One of the key issues in INR for image signals is to represent high-frequency details in each image using a small MLP-based architecture. A lack of high-frequency details will result in low image quality in the INR-based services. Recent studies in machine learning have developed sinusoidal encodings and activations [4], [7], [8] to reduce errors in high-frequency details. In this paper, we propose a novel communication-side approach to solve the issue of the INR. A key idea of the proposed scheme is to send coded residuals containing high-frequency details between the original and decoded pixel values obtained from the trained INR, in addition to sending the trained parameters of the INR, to compensate for the details. Specifically, the proposed scheme integrates digital and analog transmission, inspired by hybrid digital-analog (HDA) transmission [9], [10]. The digital transmission part sends the parameter set of the trained INR in a digital manner, while the analog transmission part encodes the residuals and sends them in an analog manner. Specifically, the proposed scheme computes and encodes residuals between the original pixel values and the decoded pixel values obtained from the trained INR, and transmits the coded residuals using analog modulation. By sending the coded residuals in an analog manner, the proposed scheme can send high-frequency details to the users but also improve the image quality according to the instantaneous wireless channel quality between the sender and each receiver. We note that the future machine-learning-side approaches for the INR will contribute to the quality improvement of our communication-side approach since they reduce the energy of the residuals for further quality enhancement [11].

Evaluations using a Kodak image dataset [12] show that the proposed scheme achieves better image quality than standard image codecs, such as JPEG, JPEG2000, and High-Efficiency Image File Format (HEIF), and the existing INR-based image compression, COIN, under the same amount of traffic. In addition, the improvement in wireless channel quality leads to an improvement in image quality by sending the coded residuals in an analog manner.

II. Related Work

A. Implicit Neural Representation

INR represents multimedia signals, such as images and 3D point clouds, not as explicit values, but as a continuous mapping function from spatial or spatio-temporal coordinates to some quantity of interest using a small NN architecture. A key issue in INR is the lack of accuracy in high-frequency details. To represent high-frequency details using a small NN architecture, SIREN in [2] argues that sinusoidal activations work better than ReLU networks because the sinusoidal activations are able to model multimedia signals contained in higher-order derivatives. Neural Radiance Fields (NeRF) in [4] proposed positional coding, and Tancik et al. [8] proposed positional coding in a Neural Tangent Kernel framework to address

979-8-3503-7099-7/24/$31.00 ©2024 IEEE

the same problem. Based on the results of the INR work, recent work has used INR for image compression [5], [6]. Unlike neural compression [13], [14] for RGB color values, INR-based image compression represents the same number of RGB color values using a small NN architecture. Therefore, it achieves traffic reduction compared to the existing image codecs. We note that the continuous fashion of the INR also has the potential for various tasks such as interpolation and super-resolution [15]–[17].

Our study is inspired by the advantage of INR, which can represent multimedia signals with low traffic. On the other hand, the existing solutions still require large traffic to represent high-frequency details compared to the existing codecs. We propose an analog-empowered solution to provide the details for users with low traffic. Specifically, the proposed scheme computes the residuals containing the details and transmits the coded residuals in an analog manner over wireless channels. The integration with the analog transmission improves the image quality even with a small increase in traffic, and achieves graceful image quality improvement according to the wireless channel quality.

B. Hybrid Digital-Analog Transmission

There are several HDA solutions for multimedia transmission [9], [10], [18] that take advantage of both digital-based and analog-based transmission schemes. HDA solutions typically use low-rate digital codecs for the images and analog coding for the residuals and sequentially/simultaneously transmit the digital-coded and analog-coded transmission symbols over wireless channels. The earlier studies [9] have designed HDA transmission for image and video signals and the recent studies [10], [18] extended HDA transmission for multi-view video plus depth and point cloud signals.

The proposed scheme is the first study to propose an INR-based HDA transmission scheme. Specifically, the digital transmission part sends the trained parameters of the INR to guarantee the basic image quality. The analog transmission part sends the coded residuals to complement the details of the image. The integration can provide many details to the users under almost the same traffic as the existing INR-based solution.

III. Proposed Scheme

A. Overview

Fig. 1 shows an overview of the proposed scheme. We consider a data set of an RGB image with MN pixels $\mathcal{D} = \{\boldsymbol{x}_i, f(\boldsymbol{x}_i)\}_i$, where $\boldsymbol{x}_i = [x_i, y_i]$ is i-th pixel coordinates and $f(\boldsymbol{x}_i) = [r_i, g_i, b_i]$ is the color values of the coordinates. Here, we consider the height and width of the RGB image to be M and N pixels, respectively. The proposed scheme consists of digital and analog transmitters and receivers. The digital part guarantees the baseline quality of the RGB image, and the analog part improves the image quality according to the instantaneous wireless channel quality. We assume that the transmitter and receiver share the same NN architecture of the INR in advance.

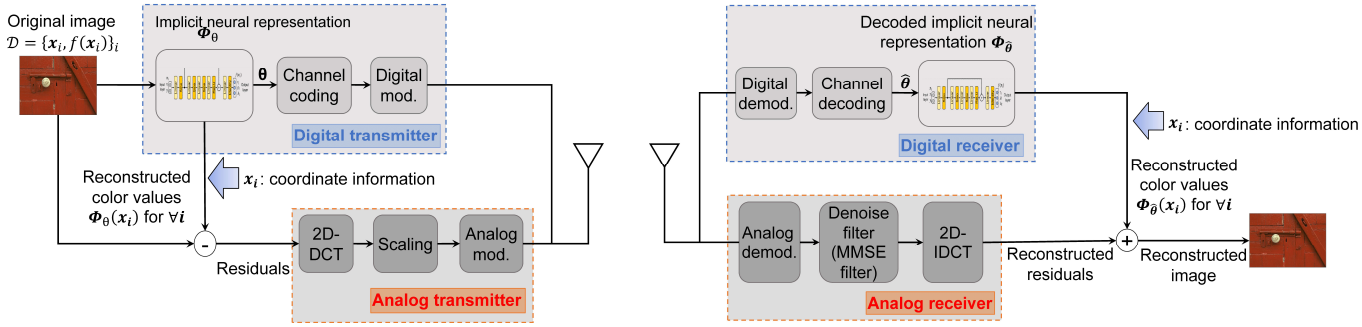
B. Transmitter

1) Digital Transmitter: The digital transmitter consists of the INR, the channel encoder, and the digital modulator. The INR Φ_θ: $\mathbb{R}^2 \to \mathbb{R}^3$ is based on the existing NN architecture for INR [2], [5] with skip connection with a set of parameters θ for fast convergence. The input to INR is the i-th pixel coordinates of RGB images $\boldsymbol{x}_i = [x_i, y_i]$. We set the number of linear layers to L and the number of neurons to W. In addition, the activation function for the front hidden layers is the sine function representing the high-frequency details, and the activation function for the last hidden layer is the ReLU function to avoid the gradient vanishing. The last layer uses an identity function to output the color values of the pixel coordinates $\Phi_\theta(\boldsymbol{x}_i) = [\hat{r}_i, \hat{g}_i, \hat{b}_i]$. The purpose of the INR is to reconstruct the pixel values of each pixel index as close to the original values as possible. This means that the parameter set of the INR θ should be optimized to minimize the error between the original color values $f(\boldsymbol{x}_i)$ and the reconstructed ones $\Phi_\theta(\boldsymbol{x}_i)$ over all pixel coordinates of the RGB image as follows:

$$\mathcal{L}_{\mathrm{MSE}}(\theta) = \sum_i \|\Phi_\theta(\boldsymbol{x}_i) - f(\boldsymbol{x}_i)\|^2. \tag{1}$$

The input of the channel coding is the binarized parameter set θ for error protection. The channel-coded bits are assigned to the transmission symbols using the digital modulation formats, e.g., Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying (QPSK), and Quadrature Amplitude Modulation (QAM). For example, the i-th modulated symbol of BPSK $s_i^{\langle \mathsf{d} \rangle}$ is formed by $s_i^{\langle \mathsf{d} \rangle} = b_i$, where $b_i \in \mathbb{X} = \{\pm 1\}$.

2) Analog Transmitter: The analog transmitter consists of an analog encoder, a scaler, and an analog modulator. To reduce the signal energy of the analog transmitter, the proposed scheme calculates the residuals $\boldsymbol{d}_i = (d_i^{(r)}, d_i^{(g)}, d_i^{(b)})$ from the original color values $f(\boldsymbol{x}_i)$ and the reconstructed ones derived from the INR $\Phi_{\hat{\theta}}(\boldsymbol{x}_i)$ with the set of decoded parameters $\hat{\theta}$ as: $\boldsymbol{d}_i = f(\boldsymbol{x}_i) - \Phi_{\hat{\theta}}(\boldsymbol{x}_i)$. After calculating the residuals for all pixel coordinates, the analog transmitter feeds the two-dimensional (2D) residuals of the color channels $\mathbf{D} = [\mathbf{D}^{(r)}, \mathbf{D}^{(g)}, \mathbf{D}^{(b)}] \in \mathbb{R}^{M \times N \times 3}$ to the analog encoder. Here $\mathbf{D}^{(\cdot)} \in \mathbb{R}^{M \times N}$ is the 2D residual for a color channel. The analog encoder is a full-frame 2D Discrete Cosine Transform (2D-DCT) to transform the residuals of each color channel into frequency representations $\mathbf{C} \in \mathbb{R}^{M \times N}$ and discards low-energy frequency representations for further energy compression. Unlike the digital transmitter, the analog transmitter skips the bit-to-symbol mapping for transmission. Instead, it maps frequency representations directly to transmission symbols, i.e., analog modulation. Analog modulation ensures that the quality of the transmitted residuals can be incrementally improved as the wireless channel quality improves. The simple



(a) Transmitter (b) Receiver

Fig. 1. Overview of the proposed scheme: (a) transmitter side and (b) receiver side.

mapping between frequency representations and transmission symbols results in low quality due to channel noise. The scaling operation unevenly assigns transmission power to the frequency representations before analog modulation for error protection.

Let $s_i^{\langle a \rangle}$ be the i-th analog-modulated symbol, $c_i \in \mathbf{C}$ be the i-th DCT coefficient, and g_i is the scaling factor for the i-th DCT coefficient. Here, the j-th analog-modulated symbol is formed by $s_j^{\langle a \rangle} = g_i \cdot c_i + \jmath g_{i+1} \cdot c_{i+1}$. A key issue for analog modulated symbols is to find the optimal scaling factor for each DCT coefficient to minimize the MSE under the transmit power budget as follows:

$$\min_{\{g_i\}} \quad \mathsf{MSE} = \mathbb{E}\left[(c_i - \hat{c}_i)^2\right] = \frac{1}{N}\sum_i^N \frac{\sigma^2 \lambda_i}{g_i^2 \lambda_i + \sigma^2}, \tag{2}$$

$$\text{s.t.} \quad \frac{1}{N}\sum_i^N g_i^2 \lambda_i = P, \tag{3}$$

where $\hat{c}_i$ is the estimate of the i-th DCT coefficient, P is the transmit power budget, $\lambda_i = |c_i|^2$ is the power of the i-th DCT coefficient, and σ^2 is the noise power in the wireless channel. By solving the MSE minimization problem, the optimized scaling factor g_i can be derived as follows:

$$g_i = m\lambda_i^{-1/4}, \qquad m = \sqrt{\frac{P}{\sum_j \lambda_j^{1/2}}}. \tag{4}$$

C. Receiver

The proposed scheme sequentially transmits digital and analog-modulated symbols over wireless channels. The wireless channel, denoted by η, takes the digital and analog modulated symbols $\boldsymbol{s}$ as input and produces the output as the received signal $\boldsymbol{y}$. The channel transfer function from the transmitter to the receiver can be modeled as $\boldsymbol{y} = \eta(\boldsymbol{s}) = \boldsymbol{s} + \boldsymbol{n}$, where $\boldsymbol{n} \sim \mathcal{CN}(0, \boldsymbol{I}\sigma^2)$ is a vector of additive white Gaussian noise (AWGN) with an average noise variance of σ^2, $\boldsymbol{I}$ is the identity matrix, and $\mathcal{CN}(a, b)$ is a complex Gaussian distribution with a mean of a and a variance of b.

1) Digital Receiver: The digital receiver consists of the digital demodulator, the channel decoder, and the decoded INR. The digital receiver first demodulates the digitally modulated symbols and decodes the channel-coded bits to obtain the set of decoded parameters $\hat{\theta}$ of the INR. The decoded parameter set $\hat{\theta}$ is then assigned to the INR at the receiver side. The digital receiver then feeds the pixel coordinates of the RGB image $\boldsymbol{x}_i$ to the decoded INR to reconstruct the color values $\Phi_{\hat{\theta}}(\boldsymbol{x}_i)$ of the pixel coordinates.

2) Analog Receiver: The analog receiver includes the analog demodulator, denoiser, and decoder. It considers analog modulated symbols as the received frequency representations and descales the frequency representations of each color channel $\hat{\mathbf{C}} \in \mathbb{R}^{M \times N}$ using a minimum MSE (MMSE) filter that minimizes the MSE between the original and descaled frequency representations. The MMSE filter provides an estimate of the DCT coefficient $\hat{c}_i \in \hat{\mathbf{C}}$ by exploiting the knowledge of the power information of the coefficient and the channel noise as follows:

$$\hat{c}_i = \frac{g_i \lambda_i}{g_i^2 \lambda_i + \sigma^2} \cdot y_i. \tag{5}$$

where $y_i \in \boldsymbol{y}$ is i-th received analog modulated symbol.

The analog receiver performs inverse 2D-DCT (2D-IDCT) for each color channel to transform the denoised frequency representations into the decoded residuals $\hat{\mathbf{D}} \in \mathbb{R}^{M \times N \times 3}$. Finally, the proposed scheme adds the reconstructed color values from the decoded INR $\Phi_{\hat{\theta}}(\boldsymbol{x}_i)$ to the reconstructed residuals $\hat{\boldsymbol{d}}_i$ to decode the final color values $\boldsymbol{r}_i$ as follows:

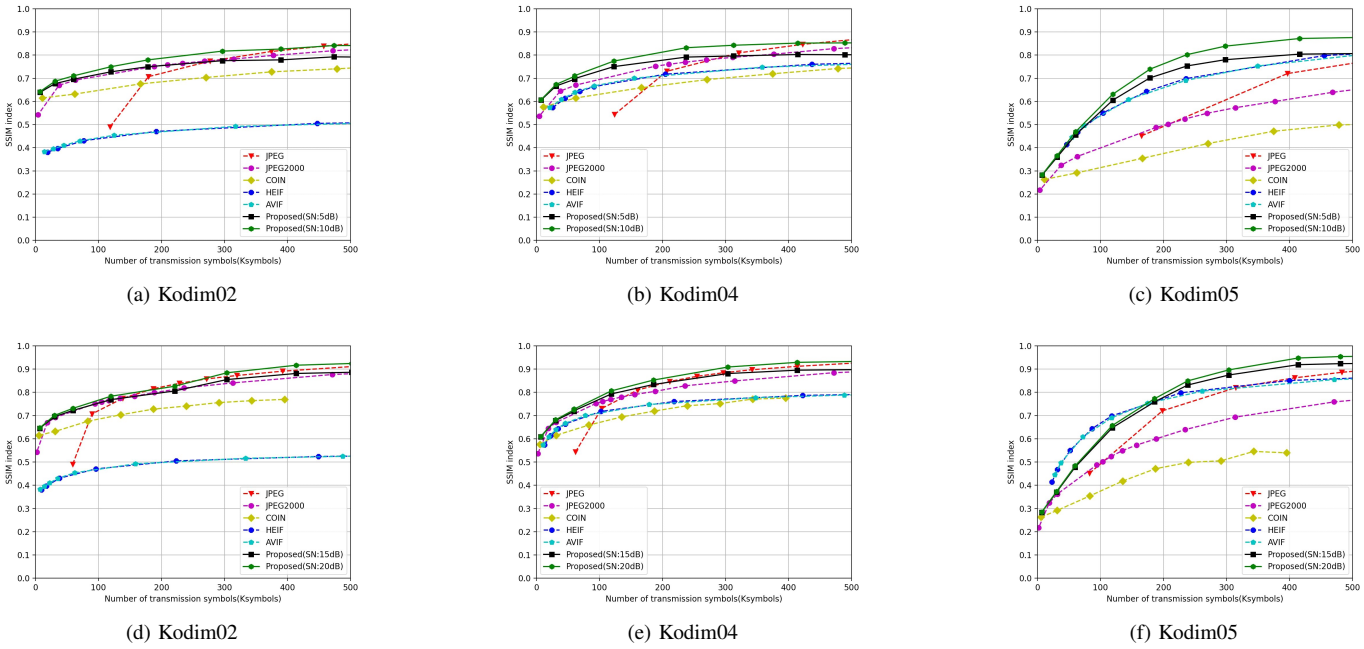
$$\boldsymbol{r}_i = \Phi_{\hat{\theta}}(\boldsymbol{x}_i) + \hat{\boldsymbol{d}}_i. \tag{6}$$

IV. Evaluation

A. Settings

Metric: We evaluate image quality in terms of the Structural Similarity Index (SSIM) [19]. SSIM can predict the perceived quality of the original and decoded images. Larger values of SSIM close to 1 indicate higher perceptual similarity between the original and decoded images.

Dataset: We perform experiments on the Kodak image dataset [12], which consists of 24 images of 768×512



(a) Kodim02 (b) Kodim04 (c) Kodim05

(d) Kodim02 (e) Kodim04 (f) Kodim05

Fig. 2. Rate distortion performance of the proposed and baseline schemes under modulation format of BPSK with/without $1/2$-rate convolution code. (a)-(c) BPSK with $1/2$-rate convolution code. (d)-(f) BPSK without convolution code.

pixels. We selected three images from the dataset, Kodim02, Kodim04, and Kodim05, to discuss the SSIM index and visual quality in different RGB images. Specifically, Kodim02 is a natural image, Kodim04 is a face image, and Kodim05 is a complicated image.

Baselines: We compare the proposed scheme with the INR-based compression scheme, COIN, and the standard image codecs, JPEG, JPEG2000, HEIF, and AV1 Image File Format (AVIF). We used Pillow 8.4.0 for JPEG and JPEG2000, pillow-heif 0.11.1 for HEIF, and pillow-avif-plugin 1.3.1 for AVIF implementations. We implement the proposed scheme and COIN in PyTorch and perform all experiments on a single RTX3080Ti GPU. We use the adaptive momentum (ADAM) optimizer for weight learning of the proposed scheme and COIN with a learning rate of 0.0002 for $50,000$ epochs. We set the number of linear layers L to 5 and the number of neurons W to 2 to 64 to discuss the image quality under the different amount of traffic.

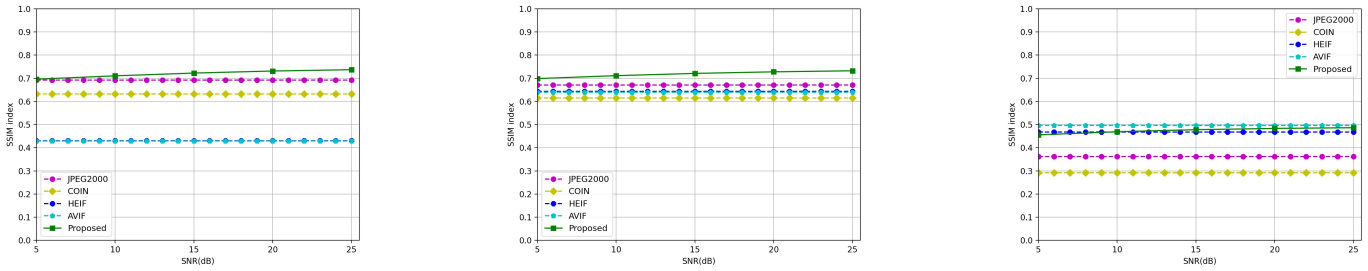
Wireless Channel Settings: The transmitted digital and analog symbols are impaired by an AWGN channel. We used scikit-commpy 0.8.0 for the wireless channel simulation. The digital transmitter of the proposed scheme and the baselines use the BPSK digital modulation format with/without a 1/2-rate convolutional code and a constraint length of 8. From preliminary evaluations, the receiver decodes the BPSK-modulated symbols with a $1/2$ convolutional code above the wireless channel SNR of 5 dB, and then the corresponding results are used for the performance comparison in low-quality wireless channels. In addition, the receiver decodes the BPSK-modulated symbols without a convolutional code above the wireless channel SNR of 14 dB, and then the results are used for the performance comparison in higher-quality wireless channels.

B. Rate Distortion Performance

We first discuss the rate-distortion performance of the proposed and baseline schemes under the different wireless channel environments. Figs. 2 (a) to (c) show the SSIM index of the proposed and baseline schemes as a function of the amount of image traffic in terms of the total number of transmitted symbols considering the modulation format of BPSK with a $1/2$ rate of the convolutional code at the wireless channel SNRs of 5 dB and 10 dB. We note that the performance of the proposed scheme depends on the instantaneous wireless channel quality. We consider the proposed schemes under the different wireless channel SNRs of 5 dB and 10 dB to discuss the effect of channel quality fluctuation on the rate-distortion performance.

The results of the evaluation reveal the following key observations:

- The proposed scheme achieves the best image quality in lower-traffic regimes. Specifically, the proposed scheme achieves the best performance up to 178.5 K transmitted symbols in Kodim02, 237.5K symbols in Kodim04, and 298.1K symbols in Kodim05,
- The proposed scheme gradually improves the image quality with the improvement of the wireless channel quality by sending the residuals in an analog manner,



(a) Kodim02 (b) Kodim04 (c) Kodim05

Fig. 3. SSIM index as a function of wireless channel SNRs in different RGB images at the number of transmitted symbols ranging from 60.5 Ksymbols to 76.6 Ksymbols. The detailed number of transmitted symbols in each scheme is shown in Fig. 4.

- In Kodim02 and Kodim04, JPEG and JPEG2000 achieve better image quality than the proposed scheme in larger traffic environments,
- HEIF and AVIF perform well in Kodim05, while they have low rate-distortion performance in Kodim02, and
- COIN achieves better image quality than other standard image codecs for a total number of transmitted symbols of 11.2 Ksymbols.

The above results suggest that the proposed scheme has the potential to provide high quality images to users in low-quality wireless channels.

Figs. 2 (d) and (f) show the SSIM index of the proposed and baseline schemes as a function of the amount of image traffic considering the modulation format of BPSK under the wireless channel SNRs of 15 dB and 20 dB. We also consider the proposed schemes under the wireless channel SNRs of 15 dB and 20 dB for the same purpose in Fig. 2 (a) to (c). From the results, there are several findings:

- The proposed scheme achieves the best image quality at the number of transmitted symbols of 5.9 Ksymbols in Kodim02 and Kodim04,
- The proposed scheme outperforms JPEG2000 at a large number of transmitted symbols and a wireless channel SNR of 15 dB in the same images,
- When the wireless channel quality improves to 20 dB, the proposed scheme achieves the best image quality, especially in the number of transmitted symbols above 303.7 K, and
- HEIF and AVIF achieve the best image quality with fewer symbols in Kodim05, while the proposed scheme overcomes them with many symbols.

C. Effect of Channel Quality Fluctuation

The above section demonstrated that the proposed scheme is well-performed in low-quality wireless channels and improves the rate-distortion performance with the improvement of the wireless channel quality during image delivery. In this section, we discuss the effect of channel quality improvement on image quality in detail.

Figs. 3 (a) to (c) show the SSIM index of the proposed and baseline schemes as a function of the instantaneous wireless channel SNRs, considering the modulation format of BPSK with a $1/2$ rate of the convolutional code at the number of transmitted symbols from 60.5 Ksymbols to 76.6 Ksymbols. In COIN and the standard image codecs, the image quality is constant regardless of the channel quality improvement. The image quality of COIN is limited due to the size of the NN architecture, and the receiver side does not recover quantization errors in the image codecs. In contrast, the proposed scheme gradually improves the image quality according to the instantaneous SNRs of the wireless channels by sending the residuals in an analog manner. For example, the proposed scheme improves 0.0352 on average from the wireless channel SNRs of 5 dB to 25 dB.

D. Visual Quality

Figs. 4 (a)–(u) show snapshots of the proposed scheme and baselines for different RGB images to discuss the visual quality. Here we consider the modulation format of BPSK with a $1/2$ rate convolutional code at wireless channel SNR of 5 dB and 10 dB and the number of transmitted symbols from 60.5 Ksymbols to 76.6 Ksymbols. We can see that the decoded images of JPEG2000 and COIN discard high-frequency details, those of HEIF and AVIF contain errors in some color channels, and the proposed scheme exhibits better visual quality compared to the other schemes by reconstructing the details.

V. Conclusion

Although INR has the potential to transmit high-quality RGB images with low traffic, the image quality is still limited due to the lack of high-frequency representations of the images. This paper proposes an INR-based HDA transmission scheme to overcome the quality limitation of INR-based solutions. In addition to transmitting the trained parameters of the INR in a digital manner, the proposed scheme also transmits the residuals in an analog manner. From the evaluations, the proposed scheme yields better image quality compared to the existing INR-based image compression under the same traffic

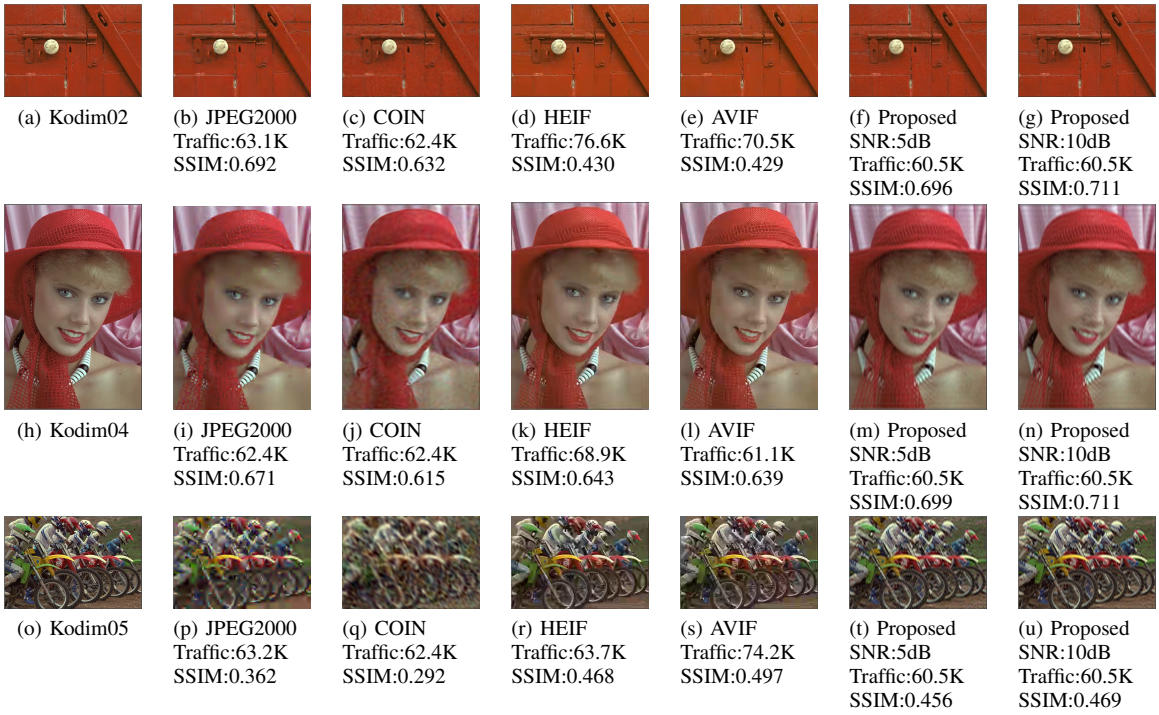


Fig. 4. Snapshots of baseline and proposed schemes in different RGB images at wireless channel SNR of 5 dB and 10 dB and the number of transmitted symbols of 60.5 Ksymbols up to 76.6 Ksymbols.

by sending the residuals in the analog part. In addition, the image quality of the proposed scheme can be improved according to the wireless channel quality between the transmitter and receiver.

REFERENCES

[1] S. Saito, T. Simon, J. Saragih, and H. Joo, "Pifuhd: multi-level pixel-aligned implicit function for high-resolution 3d human digitization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 84–93.

[2] V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020, pp. 1–12.

[3] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, "Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction," *Advances in Neural Information Processing Systems*, vol. 34, pp. 27 171–27 183, 2021.

[4] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[5] E. Dupont, A. Golinski, M. Alizadeh, Y. W. Teh, and A. D. an an, "COIN: COmpression with implicit neural representations," in *Neural Compression: From Information Theory to Applications – Workshop ICLR 2021*, 2021.

[6] "COIN++: neural compression across modalities," *Transactions on Machine Learning Research*, vol. 2022, no. 11, 2022.

[7] A. Gropp, L. Yariv, N. Haim, M. Atzmon, and Y. L. an an, "Implicit geometric regularization for learning shapes," in *Proceedings of Machine Learning and Systems 2020*, 2020, pp. 3569–3579.

[8] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7537–7547, 2020.

[9] L. Yu, H. Li, and W. Li, "Wireless scalable video coding using a hybrid digital-analog scheme," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 2, pp. 331–345, 2014.

[10] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "HoloCast+: hybrid digital-analog transmission for graceful point cloud delivery with graph fourier transform," *IEEE Transactions on Multimedia*, vol. 24, pp. 2179–2191, 2021.

[11] V. Prabhakaran, R. Puri, and K. Ramchandran, "Hybrid digital-analog codes for source-channel broadcast of Gaussian sources over Gaussian channels," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4573–4588, 2011.

[12] "The kodak color image dataset," 2022. [Online]. Available: https://r0k.us/graphics/kodak/

[13] D. Minnen, J. Ballé, and G. D. Toderici, "Joint autoregressive and hierarchical priors for learned image compression," *Advances in neural information processing systems*, vol. 31, 2018.

[14] G. Lu, X. Zhang, W. Ouyang, L. Chen, Z. Gao, and D. Xu, "An end-to-end learning framework for video compression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3292–3308, 2021.

[15] Q. Wu, Y. Li, Y. Sun, Y. Zhou, H. Wei, J. Yu, and Y. Z. an an, "An arbitrary scale Super-Resolution approach for 3D MR images via implicit neural representation," *IEEE J Biomed Health Inform*, vol. 27, no. 2, pp. 1004–1015, Feb. 2023.

[16] K. Zhang, D. Zhu, X. Min, and G. Z. an an, "Implicit neural representation learning for hyperspectral image Super-Resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–12, 2023.

[17] Q. Wu, Y. Li, L. Xu, R. Feng, H. Wei, Q. Yang, B. Yu, X. Liu, J. Yu, and Y. Z. an an, "IREM: High-Resolution magnetic resonance image reconstruction via implicit neural representation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, 2021, pp. 65–74.

[18] P. Li, F. Yang, J. Zhang, Y. Guan, A. Wang, and J. Liang, "Synthesis-distortion-aware hybrid digital analog transmission for 3D videos," *IEEE Access*, vol. 8, pp. 85 128–85 139, 2020.

[19] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Imge quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, apr 2004.