DNN-based Overhead Reduction for High-Quality Soft Delivery

Fujihashi, T.; Koike-Akino, T.; Watanabe, T.; Orlik, P.V.

TR2019-133 December 09, 2019

Abstract

Soft delivery, i.e., analog transmission, has been proposed to provide graceful video/image quality even in unstable wireless channels. However, existing analog schemes require a significant amount of metadata for power allocation and decoding operations. It causes large overheads and quality degradation due to rate and power losses. Although the amount of overheads can be reduced by introducing Gaussian Markov random field (GMRF) model, the model mismatch can degrade reconstruction quality. In this paper, we propose a novel analog transmission scheme to simultaneously reduce the overheads and yield better reconstruction quality. The proposed scheme uses a deep neural network (DNN) for metadata compression and decompression. Specifically, the metadata is compressed into few variables using the proposed DNN-based metadata encoder before transmission. The variables are then transmitted and decompressed at the receiver for high-quality video/image reconstruction. Evaluations using test images demonstrate that our proposed scheme reduces overheads by 80.0 % with 11.2 dB improvement of reconstruction quality compared to the existing analog transmission schemes.

IEEE Global Communications Conference (GLOBECOM)

DNN-based Overhead Reduction for High-Quality Soft Delivery

Takuya Fujihashi[†*], Toshiaki Koike-Akino[†], Takashi Watanabe[*], and Philip V. Orlik[†]
[†]Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA
[*]Graduate School of Information and Science, Osaka University, Osaka, Japan

***Abstract*—Soft delivery, i.e., analog transmission, has been proposed to provide graceful video/image quality even in unstable wireless channels. However, existing analog schemes require a significant amount of metadata for power allocation and decoding operations. It causes large overheads and quality degradation due to rate and power losses. Although the amount of overheads can be reduced by introducing Gaussian Markov random field (GMRF) model, the model mismatch can degrade reconstruction quality. In this paper, we propose a novel analog transmission scheme to simultaneously reduce the overheads and yield better reconstruction quality. The proposed scheme uses a deep neural network (DNN) for metadata compression and decompression. Specifically, the metadata is compressed into few variables using the proposed DNN-based metadata encoder before transmission. The variables are then transmitted and decompressed at the receiver for high-quality video/image reconstruction. Evaluations using test images demonstrate that our proposed scheme reduces overheads by 80.0 % with 11.2 dB improvement of reconstruction quality compared to the existing analog transmission schemes.**

I. Introduction

Multimedia delivery, i.e., image and video delivery, is one of the major applications in the wireless environment — according to Cisco visual networking index studies, around 82 % of the world's mobile data traffic will be video contents by 2022 [1]. In conventional multimedia streaming, the digital compression and digital wireless transmission are carried out in sequence [2]–[4]. For example, in wireless digital video delivery, the video compression part uses H.264/Advanced Video Coding (AVC) [5] or H.265/High-Efficiency Video Coding (HEVC) [6] standards to generate a compressed bit stream using quantization and entropy coding. The wireless transmission part uses channel coding and a digital modulation scheme to reliably transmit the encoded bit stream.

However, the digital-based conventional schemes have the following problems due to the unreliable wireless channel. First, the encoded bit stream is highly vulnerable to bit errors. When the channel's signal-to-noise ratio (SNR) falls under a certain threshold, the image/video quality drops significantly. This phenomenon is referred to as the cliff effect. Second, the image/video quality does not gracefully improve even when the wireless channel quality is improved. Finally, quantization is a lossy process, whose distortion cannot be recovered at the receiver. Some studies [7], [8] have been proposed to mitigate the cliff effect in the digital transmission by introducing scalable source coding and scalable channel coding. However, in these studies, the cliff effect is converted into the so-called staircase effect [9]. In the staircase effect, the multimedia quality discontinuously improves as the wireless channel quality improves.

To overcome the above-mentioned problems, analog transmission schemes [10]–[14] have been proposed for wireless image/video delivery. For example, SoftCast [10] directly transmits linearly-transformed video signals over a lossy channel and allocates power to the signals to maximize the received quality, instead of using digital compression and digital modulation. In contrast to the conventional digital scheme, the received video quality of SoftCast can be gracefully improved according to the wireless channel quality. FoveaCast [11] extends SoftCast to wireless image delivery to reduce perceptual redundancy by considering user's foveated point. Specifically, FoveaCast adaptively assigns transmission power to linear-transformed image signals based on the foveated point to simultaneously realize graceful quality improvement and user's interest-aware quality adaptation.

However, the performance of soft delivery schemes depends strongly on the chunk size. In both SoftCast and FoveaCast, a sender allocates transmission power to the image/video signals such that the receiver noise can be minimized. The power allocation is based on the power of each linearly-transformed signal. Hence, the sender needs to transmit the power information of all the image/video signals to decode the signals at the receiver. The transmission of this metadata causes large overhead, resulting in image/video quality degradation due to power and rate loss. To reduce metadata overhead, both schemes divide the signals into multiple chunks and transmit a smaller number of metadata corresponding to each chunk. In turn, the chunk division may degrade performance due to improper power allocation, in particular when a large chunk size is used for lower overhead.

To reduce the amount of overheads, the existing studies [13], [14] use a fitting function based on Gaussian Markov random field (GMRF) [15], [16] model to approximate the power information of linearly-transformed video signals only with few parameters. Specifically, [13] realizes overhead reduction in soft delivery of single-view video using a GMRF-based fitting function. They demonstrated that the fitting function brings significant overhead reduction since the power information can be fit by the fitting function with four parameters. [14] extends the fitting function to free viewpoint soft video delivery. Although the free viewpoint video signals are five-dimensional (5D) video signals, the fitting function based on first-order 5D-GMRF can approximate the corresponding power information

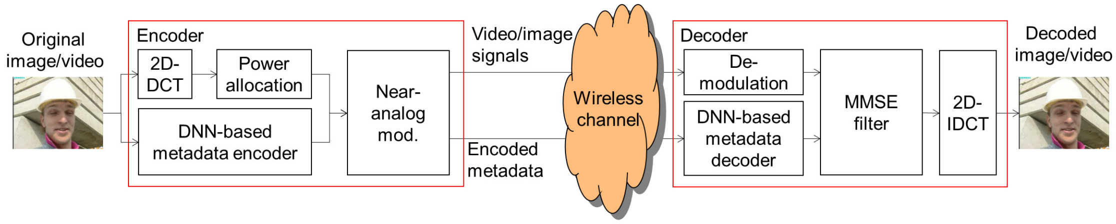


Fig. 1. Overview of the proposed soft delivery, employing DNN-based metadata encoder/decoder.

using nine parameters. Although the existing schemes yield low metadata overhead, they still remain degraded reconstruction quality since the fitting functions suffer model mismatch for real single-view/multi-view signals.

In this paper, we propose a novel analog scheme to achieve better reconstruction quality under nearly zero overhead requirement. To obtain the power values of linear-transformed signals without transmitting large-overhead metadata, the proposed scheme uses a deep neural network (DNN) employing denoising auto-encoders [17] for metadata encoding and decoding. Specifically, the proposed metadata encoder obtains a few variables from the pixel values of all the images before transmission and the proposed decoder decodes the accurate power information from the received variables for proper power allocation. For better reconstruction quality, the weights in both DNN-based metadata encoder and decoder are trained beforehand based on synthetic datasets generated through offline Monte–Carlo simulations. Evaluations using test images show that the proposed scheme improves reconstruction quality by 11.2 dB with 80.0 % reduction in the metadata overhead compared with the existing analog scheme employing the GMRF model [13].

Our contribution is three-fold: 1) we introduce denoising auto-encoders for metadata compression/decompression, employing new loss function to maximize image quality in the presence of power mismatch, 2) we verify that the power information can be reliably shared by the DNN-based metadata decoder even in the presence of noise, and 3) we demonstrate that only one variable metadata is sufficient to realize high-quality image reconstruction for soft delivery over wireless channels.

II. Soft Delivery with DNN-based Metadata Compression

The objectives of our proposed scheme are 1) to achieve reconstruction quality that gracefully improves according to the wireless channel quality and 2) to reduce the amount of metadata. Fig. 1 shows the schematic of our proposed scheme. The encoder first performs two-dimensional discrete cosine transform (2D-DCT) operation on the original images. At the same time, according to the pixel values of the original images, the DNN-based metadata encoder compresses the pixel values into several variables, i.e., metadata. The DCT coefficients are then scaled according to the metadata and analog-modulated for transmissions. In addition, the metadata is also scaled and analog-modulated in prior to transmission. Finally, the encoder sends the analog-modulated image symbols and metadata symbols to the receiver over a wireless channel with additive white Gaussian noise (AWGN).

At the receiver side, the decoder first reconstructs the power information from the received metadata symbols by using DNN-based metadata decoder. The reconstructed power values are used for minimum mean-square error (MMSE) filter. The DCT coefficients are then obtained from the received analog-modulated image symbols through the use of MMSE filtering. The reconstructed pixel values can be reconstructed by taking inverse 2D-DCT operation for the filtered DCT coefficients.

A. Encoder

The encoder first performs 2D-DCT operation on each original image to obtain the corresponding DCT coefficients. In addition, the original images are encoded into variables by using DNN-based metadata encoder. The DCT coefficients and encoded variables are mapped to I (in-phase) and Q (quadrature) components after the following power allocation.

Let x_i denote the ith analog-modulated symbol of the DCT coefficients/encoded variables. Each analog-modulated symbol is scaled by g_i for noise reduction:

$$x_i = g_i \cdot s_i. \tag{1}$$

Here, s_i is the ith DCT coefficient/encoded variable and g_i is the scale factor which determines the power allocation. The transmitter performs optimal power control by selecting g_i to achieve the highest reconstruction quality. Specifically, the best g_i is obtained by minimizing the mean-square error (MSE) under the power constraint with total power budget P as follows:

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

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

where $\mathbb{E}[\cdot]$ denotes expectation, $\hat{s}_i$ is an estimate of the transmitted symbol, λ_i is the power of ith DCT coefficient/encoded variable, N is the number of DCT coefficients/encoded variables, and σ^2 is a receiver noise variance. The near-optimal solution is expressed as

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

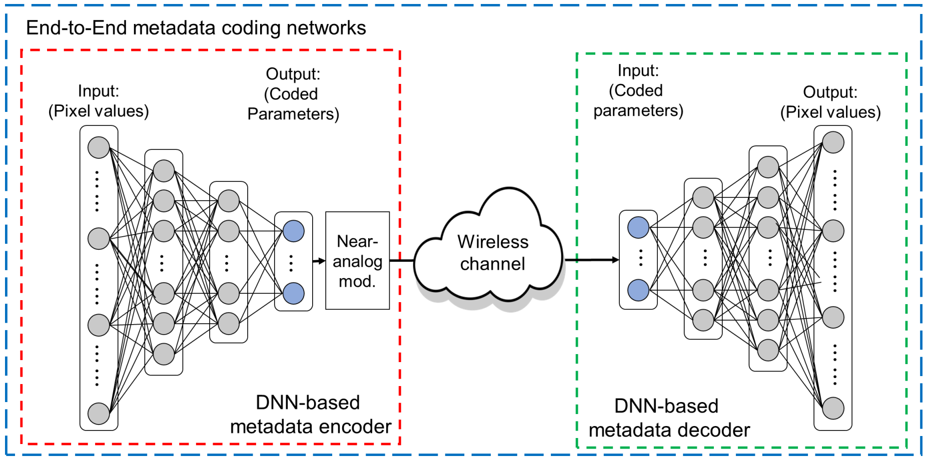


Fig. 2. End-to-End DNN-based metadata coding/decoding networks, analogous to denoising auto-encoders [17].

B. Decoder

After transmission over the wireless channel, each symbol of the DCT coefficients and encoded variables at the receiver can be modeled as follows:

$$y_i = x_i + n_i, \tag{5}$$

where y_i is the ith received symbol and n_i is an effective noise having a variance of σ^2. The receiver first obtains the encoded variables from the received symbols. The variables are then decoded to the power information of the DCT coefficients λ'_i by using DNN-based metadata decoder. The receiver extracts DCT coefficients from I and Q components, and reconstructs the coefficients using the reconstructed power information and MMSE filter [10] as follows:

$$\hat{s}_i = \frac{g'_i \lambda'_i}{g'^2_i \lambda'_i + \sigma^2} \cdot y_i, \tag{6}$$

where g'_i is the power allocation estimate by replacing λ_i with λ'_i in (4). The decoder then obtains corresponding image by taking the inverse 2D-DCT for the filter output $\hat{s}_i$.

C. Deep Neural Network-based Overhead Reduction

In order for the receiver to carry out proper MMSE filtering in (6), the sender needs to transmit the power information of all coefficients λ_i as metadata such that $g'_i = g_i$ and $\lambda'_i = \lambda_i$. Although it yields the best reconstruction quality, the amount of metadata can be significantly large. For example, when the sender transmits a color image with the resolution of 176×144, the sender needs to transmit metadata for all DCT coefficients of $176 \times 144 \times 3 = 76{,}032$ to the receiver. Such a high overhead can impose rate and power losses in practice. To reduce the overheads, the existing methods divide coefficients into chunks and carry out power allocation and MMSE filter for each chunk. However, overheads are still high and the chunk division causes performance degradation due to improper power allocation. When the chunk is a size of 44×36 pixels, 96 metadata are still required every images. To further reduce the overheads, the existing study used first-order GMRF-based fitting function to approximate the power values λ_i from few parameters. Specifically, the sender needs to transmit five metadata for every images. Although the fitting function can decrease the amount of metadata, it suffers a low image quality due to a fitting error ($\lambda'_i \neq \lambda_i$).

To reduce the metadata overheads with keeping better reconstruction quality, the proposed scheme uses DNN-based metadata coding networks to reconstruct clean images from few variables for a variety of images. Fig. 2 shows the proposed end-to-end DNN-based metadata encoder and decoder. Both proposed metadata encoder and decoder networks use multi-layer perceptron (MLP). Specifically, the proposed encoder network encodes the pixel values into m variables. Here, the number of neurons at the input and output layers is $3 \times H \times W$ and m, respectively, where H and W are the number of pixels in horizontal and vertical domains. For the kth hidden layer, the number of neurons is $2^{N_{\text{hidden}}-k} \cdot m$ where N_{hidden} is the number of hidden layers. The m metadata variables are analog-modulated and transmitted over wireless channels.

The proposed decoder reconstructs the pixel values from noisy m variables, and then obtains the corresponding power values λ'_i from the reconstructed pixel values. Specifically, the number of neurons in input and output layers of the proposed decoder are m and $3 \times H \times W$, respectively. For the kth hidden layer, the number of neurons is $2^k \cdot m$. Our scheme uses the reconstructed power information λ'_i for the MMSE filter (6).

In this case, the compression and reconstruction performance of the encoding and decoding networks depend on the pre-trained weights in both networks. To learn better weights, noisy datasets are generated offline via Monte–Carlo simulations. Specifically, all potential distortions due to additive noise, 2D-DCT, and MMSE filter are synthetically analyzed by DNN-based encoder and decoder in off-line learning phase. By using synthetic datasets for the pairs of original and reconstructed image signals at specific channel models, the proposed scheme can learn better network weights depending on wireless channel quality. For stochastic gradient optimization, the proposed scheme uses the following loss function of soft delivery's MSE:

$$l_{\mathrm{MSE}} = \mathbb{E}\left[(s_i - \hat{s}_i)^2\right] = \mathbb{E}\left[\frac{\lambda_i + \lambda'_i\sqrt{\lambda'_i}\frac{1}{\alpha'\sigma^2}}{(1+\sqrt{\lambda'_i}\frac{1}{\alpha'\sigma^2})^2}\right], \quad (7)$$

where $\alpha' = 1/\sum\sqrt{\lambda'_i}$ is a normalization factor.

It should be noted that the proposed DNN method is similar to denoising auto-encoders [17] while the MSE loss function is not conventional because we do not directly minimize auto-associative errors between input and output signals. Because the modified MSE loss function takes the presence of AWGN noise and power mismatch $\lambda'_i \neq \lambda_i$ into account, learning the network weights under the proposed loss function (7) can reconstruct clean image signals for soft image delivery. We note that the proposed scheme uses adaptive momentum (ADAM) optimizer [18] for joint weight learning of metadata coding/decoding networks.

III. Performance Evaluation

A. Simulation Settings

Metric: We evaluate the reconstruction quality of the reference schemes in terms of MSE between the original images and the reconstructed images sent over wireless AWGN channels.

Test Images: We used the benchmark dataset, namely, CIFAR-100 [19] for evaluations. CIFAR-100 consists of multiple training images and testing images with 100 classes. The training images are used for learning the network weights while the testing images are used for comparison in terms of image and visual quality. We consider 50,000 training images and 100 testing images for evaluations of the proposed metadata coding networks.

Amount of Metadata: As we mentioned in Sec. II-C, the proposed scheme sends analog-modulated m variables for each image. We first evaluate the baseline performance of the proposed scheme with $m = 1$ variable, and then discuss an impact of the number of variables on the reconstruction quality in Sec. III-C. SoftCast [10] transmits mean and variance as metadata variables for each chunk. GMRF-based scheme [13] sends five parameters for each image.

B. Reconstruction Quality

We first evaluate the image reconstruction quality of the proposed and four existing schemes in different wireless channel quality. For the comparison, we measure the proposed scheme, GMRF-based soft delivery scheme [13], and SoftCast [10] schemes with different chunk sizes: 1×1, 2×2, and 32×32 pixels. The corresponding amount of metadata in SoftCast and GMRF-based soft delivery schemes become 3072, 1536, 6, and 5 for each image, respectively. GMRF-based soft delivery and SoftCast schemes directly map linear-transformed 2D-DCT coefficients on the I and Q components for image delivery to prevent cliff effect and gracefully improve image quality according to wireless channel quality.

Fig. 3 shows the reconstructed image quality of the proposed and existing soft delivery schemes as a function of wireless

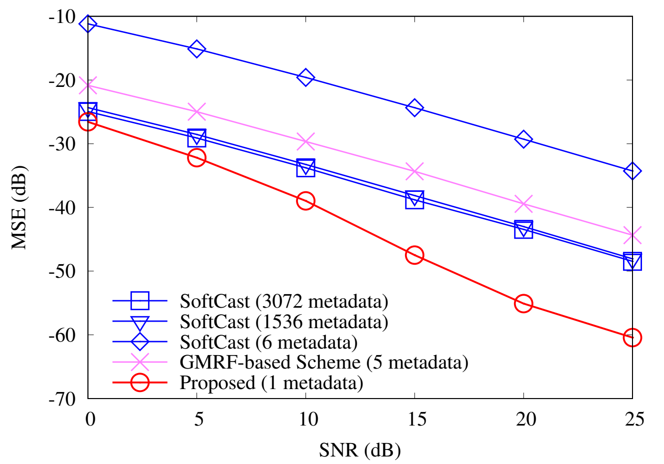


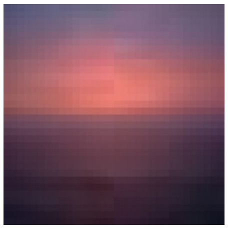
Fig. 3. MSE performance of the proposed and existing soft delivery schemes as a function of wireless channel SNR.

channel SNRs. Here, the amount of metadata in the proposed scheme is one variable. From the evaluation results, we can see the following key observations:

- The proposed scheme realizes graceful quality improvement with the improvement of wireless channel quality and yields better reconstruction quality compared with the existing soft delivery schemes.
- Even though CIFAR-100 has 100 different classes, i.e., large-class data set, the proposed DNN-based encoder can represent the metadata with only one variable.
- Although SoftCast with chunk size of 1×1 pixel needs to send all the power values of the DCT coefficients as metadata to realize graceful quality improvement, the proposed scheme achieves better and graceful reconstruction quality even with one metadata.
- SoftCast with chunk size of 32×32 pixels can reduce the overheads. However, it has a low reconstruction quality irrespective of wireless channel SNRs due to improper power allocation.
- GMRF-based soft delivery scheme reduces the overheads and realizes better reconstruction quality compared with SoftCast with a large chunk size, i.e., 32×32 pixels. On the other hand, it still has a low reconstruction quality compared with the proposed scheme due to a fitting error.

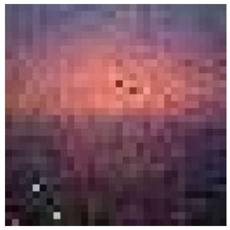
Especially, it is demonstrated that the MSE performance of the proposed scheme is better than SoftCast scheme with a chunk size of 1×1 pixel irrespective of wireless channel SNRs. This may be because the proposed method directly maximizes the image quality with the modified MSE loss function in (7), whereas SoftCast schemes assign sub-optimal power without taking the mismatch of the original and reconstructed power information into account. Since the proposed scheme optimizes the weights of the DNN-based metadata coding networks under the consideration of the mismatch of the power information, it can keep better MSE performance.

For example, the proposed scheme achieves quality im-

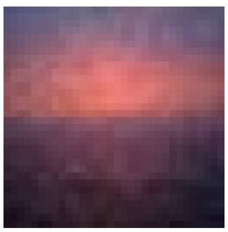


(a) Original

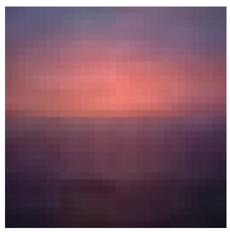
(b) SoftCast (1×1 chunk)
SNR: 0 dB
MSE: −24.9 dB



(c) Proposed
SNR: 0 dB
MSE: −26.6 dB



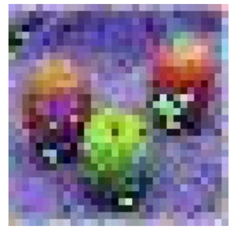
(d) SoftCast (1×1 chunk)
SNR: 10 dB
MSE: −33.8 dB



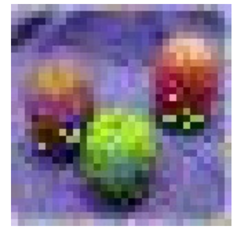
(e) Proposed
SNR: 10 dB
MSE: −39.0 dB

Fig. 4. Snapshot of a reconstructed image from CIFAR-100 at SNRs of 0 dB and 10 dB.

(a) Original



(b) SoftCast (1×1 chunk)
SNR: 0 dB
MSE: −24.9 dB



(c) Proposed
SNR: 0 dB
MSE: −26.6 dB

(d) SoftCast (1×1 chunk)
SNR: 10 dB
MSE: 33.8 dB

(e) Proposed
SNR: 10 dB
MSE: −39.0 dB

Fig. 5. Another snapshot of the reconstructed image from CIFAR-100 at SNRs of 0 dB and 10 dB.

provement by 11.2 dB, 7.0 dB, 7.6 dB, and 21.2 dB on average over GMRF-based soft delivery scheme, SoftCast with chunk size of 1×1 pixel, 2×2 pixels, and 32×32 pixels across wireless channel SNRs of 0 to 25 dB with 80.0%, 99.97%, 99.93%, and 83.33% of overhead reduction, respectively.

To discuss reconstructed visual quality of each reference scheme, Figs. 4 and 5 show snapshots at different wireless channel SNRs. For example, in Figs. 4(b) and (d), the reconstructed images are severely distorted even after an optimized MMSE filter. The proposed scheme in Figs. 4(c) and (e) can effectively denoise the images in Figs. 4(b) and (d) by using the power values of all the DCT coefficients obtained from the proposed DNN-based metadata decoder.

C. Effect of Amount of Metadata

We then discuss the reconstructed image quality with different amount of overheads in the proposed scheme. Fig. 6 shows the reconstructed image quality of the existing soft delivery and proposed schemes with different amount of metadata overheads as a function of wireless channel SNRs. Here, we consider three proposed schemes with different amount of metadata m: one, two, and four variables. We can see that the proposed schemes achieve the better reconstruction quality in low wireless SNR regimes irrespective of the overheads. In addition, the proposed scheme with four variables offers only a little improvement in the reconstructed image quality compared with the proposed scheme with one variable across the wireless channel SNRs of 0 to 10 dB. It is hence verified that the proposed metadata compression based on DNN can realize nearly zero overhead for soft image delivery.

We also observe that the proposed scheme with four variables can degrade the reconstruction quality at high wireless SNR regimes. Fig. 7 shows the learning curves of testing images for the proposed DNN-based metadata coding networks with different number of variables. We plot the best loss function, i.e., MSE, obtained from testing images in each epoch. From the learning curves, the proposed scheme with one variable realizes a low MSE after 70 epochs. It is demonstrated that the proposed scheme with four variables still has a high MSE even in 100 epochs. In order to improve the learning convergence, another DNN architecture such as convolutional counterparts can be considered. Detail analysis and discussion for another DNN architecture will be left as future works.

IV. CONCLUSION

This paper proposed a low overhead analog transmission scheme, employing DNN-based metadata encoder/decoder.

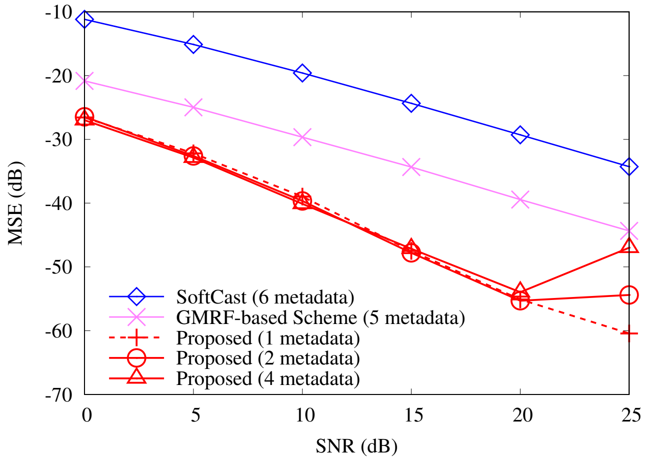


Fig. 6. MSE performance of the proposed and existing soft delivery schemes with different amount of metadata m as a function of wireless channel SNR.

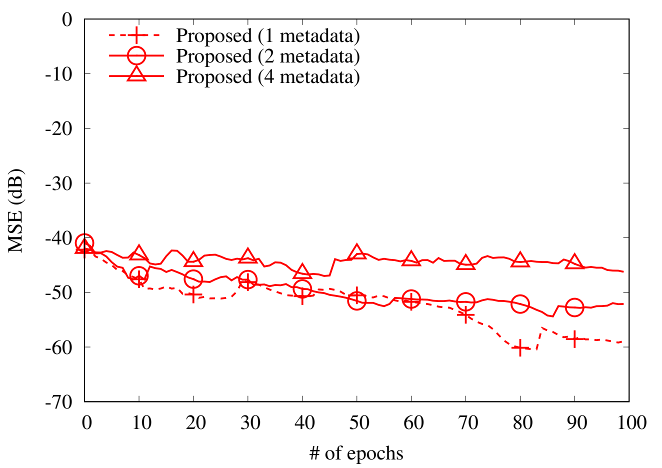


Fig. 7. Learning curves of the proposed schemes with different amount of metadata m as a number of epochs at wireless channel SNR of 25 dB.

Specifically, the proposed scheme encodes the pixel values of original images into a few variables using pre-trained DNN-based metadata encoder and the variables are decoded by the pre-trained DNN-based metadata decoder. To improve the image quality over noisy wireless channels, we derived a new loss function which directly accounts for the presence of noise and power information errors. Performance evaluations show that our proposed scheme achieves graceful and higher reconstruction quality compared to existing analog schemes with the improvement of wireless channel quality. In addition, the proposed scheme is found to realize one variable metadata, which significantly reduces the required amount of overheads. This reduction saves transmission power and may result in further quality improvement especially for band-limited wireless channels.

Acknowledgment

This work was partly supported by JSPS KAKENHI Grant Number 17K12672 and The Telecommunications Advancement Foundation.

References

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update 2017-2022," Nov. 2018.

[2] T. Stockhammer, H. Jenkac, and G. Kuhn, "Streaming video over variable bit-rate wireless channels," *IEEE Transactions on Multimedia*, vol. 6, no. 2, pp. 268–277, 2004.

[3] Z. Guo, Y. Wang, E. Erkip, and S. Panwar, "Wireless video multicast with cooperative and incremental transmission of parity packets," *IEEE Transactions on Multimedia*, vol. 17, no. 8, pp. 1135–1346, 2015.

[4] S. Almowuena, M. M. Rahman, C. H. Hsu, A. A. Hassan, and M. Hafeeda, "Energy-aware and bandwidth-efficient hybrid video streaming over mobile networks," *IEEE Transactions on Multimedia*, vol. 18, no. 1, pp. 102–115, 2016.

[5] W. Thomas, S. G. J, B. Gisle, and L. Ajay, "Overview of the H. 264/AVC video coding standard," *IEEE Transactions of Circuirts And Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[6] D. Grois, D. Marpe, A. Mulayoff, B. Itzhaky, and O. Hadar, "Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders," in *IEEE PCS*, 2013, pp. 394–397.

[7] M. Li, Z. Chen, and Y. P. Tan, "Scalable resource allocation for SVC video streaming over multiuser MIMO-OFDM networks," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1519–1531, 2013.

[8] M. M. Ghandi and M. Ghanbari, "Layered H.264 video transmission with hierarchical QAM," *Journal of Visual Communication Image representation*, vol. 17, no. 2, pp. 451–466, 2006.

[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] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in *ACM Annual International Conference on Mobile Computing and Networking*, 2011, pp. 289–300.

[11] J. Shen, L. Yu, L. Li, and H. Li, "Foveation based wireless soft image delivery," *IEEE Transactions on Multimedia*, vol. PP, no. 99, pp. 1–13, 2018.

[12] X. L. Liu, W. Hu, C. Luo, Q. Pu, F. Wu, and Y. Zhang, "ParCast+: Parallel video unicast in MIMO-OFDM WLANs," *IEEE Transactions on Multimedia*, vol. 16, no. 7, pp. 2038–2051, 2014.

[13] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "High-quality soft video delivery with gmrf-based overhead reduction," *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 473–483, 2018.

[14] ——, "FreeCast: Graceful free-viewpoint video delivery," *IEEE Transactions on Multimedia*, vol. PP, no. 99, pp. 1–11, 2019.

[15] H. Rue and H. Leonhard, *Gaussian Markov random fields: theory and applications*. CRC Press, 2005.

[16] C. Zhang and D. Florencio, "Analyzing the optimality of predictive transform coding using graph-based models," *IEEE Signal Processing Letters*, vol. 20, no. 1, pp. 106–109, 2013.

[17] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 1096–1103.

[18] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[19] A. Krizhevsky, V. Nair, and G. Hinton, "Cifar-100 (canadian institute for advanced research)." [Online]. Available: http://www.cs.toronto.edu/ kriz/cifar.html