MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

DNN-based Simultaneous Screen-to-Camera and Screen-to-Eye Communications

Fujihashi, Takuya; Koike-Akino, Toshiaki; Watanabe, Takashi; Orlik, Philip V.

TR2019-132 December 09, 2019

Abstract

Simultaneous screen-to-camera and screen-to-eye communications, i.e., watermarking, have been proposed in visible light communications. The main purpose of such communications is to provide many data bits for camera devices and visual information for human eyes by using a common displayed image. To this end, the existing studies leverage the capability discrepancy and distinctive features between the human vision system and camera devices. However, the existing techniques mainly require high refresh rates in both screen and camera devices to achieve better throughput while keeping high visual quality. In this paper, we propose a novel transmission scheme for efficient simultaneous screen-to-camera and screento-eye communications without a need of high refresh rates. Specifically, we use deep convolutional neural networks (DCNN)-based watermark encoder and decoder to embed many bits into high-quality images, and then to maximize throughput from the bit-embedded image. With end-to-end adversarial learning, the encoder networks learn a mapping function to embed digital data into an original image based on a perceptual loss function while the decoder networks also learn a mapping function from the bitembedded image to the data bits based on a cross-entropy loss function. From the evaluations, we show that the proposed watermark encoding and decoding networks yield high throughput from the bit-embedded images compared with a simple DCNNbased watermarking. In addition, the bit-embedded images on the screen achieve high quality for human perception.

IEEE Global Communications Conference (GLOBECOM)

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

DNN-based Simultaneous Screen-to-Camera and Screen-to-Eye Communications

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*—Simultaneous screen-to-camera and screen-to-eye communications, i.e., watermarking, have been proposed in visible light communications. The main purpose of such communications is to provide many data bits for camera devices and visual information for human eyes by using a common displayed image. To this end, the existing studies leverage the capability discrepancy and distinctive features between the human vision system and camera devices. However, the existing techniques mainly require high refresh rates in both screen and camera devices to achieve better throughput while keeping high visual quality. In this paper, we propose a novel transmission scheme for efficient simultaneous screen-to-camera and screen-to-eye communications without a need of high refresh rates. Specifically, we use deep convolutional neural networks (DCNN)-based watermark encoder and decoder to embed many bits into high-quality images, and then to maximize throughput from the bit-embedded image. With end-to-end adversarial learning, the encoder networks learn a mapping function to embed digital data into an original image based on a perceptual loss function while the decoder networks also learn a mapping function from the bit-embedded image to the data bits based on a cross-entropy loss function. From the evaluations, we show that the proposed watermark encoding and decoding networks yield high throughput from the bit-embedded images compared with a simple DCNN-based watermarking. In addition, the bit-embedded images on the screen achieve high quality for human perception.**

I. Introduction

Visible light communications (VLC) [1], [2] have emerged as promising complementary technologies to conventional radio-frequency (RF) wireless communications. Screen-camera communications [3]–[6] are such VLC technologies, where digital data can be transmitted via image signals from a screen to a camera. For screen-camera communications, digital bits are encoded in the screen image on devices, e.g., laptop computers and smartphones. A receiver equipped with camera image sensors captures the screen to decode the digital information. Screen-camera communications can be used for various wireless applications, such as inter/intra vehicle communications [7], near field communications [8], [9], and augmented reality (AR) [10].

The existing studies on screen-camera communications can be classified into two types: 1) high-speed and reliable short-range communications and 2) simultaneous screen-to-camera and screen-to-eye communications. In former studies, PixNet [11] is a pioneer work of high-speed and reliable communication for screen-camera links. PixNet uses orthogonal discrete multi-tone (DMT) for single-color channel transmission. In Smartphone Visible Light Communication (SVLC) [12], they realize screen-camera communications using smartphones as a sender and a receiver. Analogously, COlor Barcode stReaming for smArtphones (COBRA) [13] designed a two-dimensional (2D) color barcode for screen-camera communications.

In the latter studies, they realize concurrent screen-to-camera and screen-to-eye communications, i.e., watermarking, using screen and camera for mainly AR applications. Specifically, the sender embeds transmission bits into images and then displays the bit-embedded images on the screen. The receiver captures the displayed images by using camera sensors. As a result, the displayed images provide visual information for human eyes in addition to digital information for camera sensors. A major challenge of the concurrent screen-to-camera and screen-to-eye communications is to simultaneously achieve unobtrusive visual communication for human and higher throughput for camera sensors by using the common bit-embedded images.

For this purpose, the existing studies mainly exploit the capability discrepancy and distinctive features in the temporal domain between human vision systems and camera devices. InFrame++ [3] is a pioneer work of simultaneous screen-to-camera and screen-to-eye communications. They propose complementary frames to exploit flicker fusion property of human vision system. HiLight [14] conveys data bits through the pixel translucency change within a time window. Uber-in-Light [15] also conveys data bits through intensity change over red, green, and blue (RGB) color channels. ImplicitCode [16] uses grayscale images and combines InFrame++ and HiLight to adjust invisibility and throughput. TextureCode [17] realizes video texture-aware adaptive embedding to improve invisibility. ChromaCode [18] also realizes adaptive bit embedding in uniform color space for achieving better throughput. Although the above-mentioned studies achieve unobtrusive and high-rate communications by exploiting the difference between human vision system and camera device, [3], [15]–[18] require even high-rate screen and camera devices and [14] requires an additional channel for the communications.

In this study, we aim at high throughput in simultaneous screen-to-camera and screen-to-eye communications even in low-rate screen and camera devices without additional channel requirement. For this purpose, we propose deep convolutional neural network [19]–[21] (DCNN)-based watermark encoder and decoder to embed data bits into the images to simultaneously realize high throughput and visual quality. Specifically,

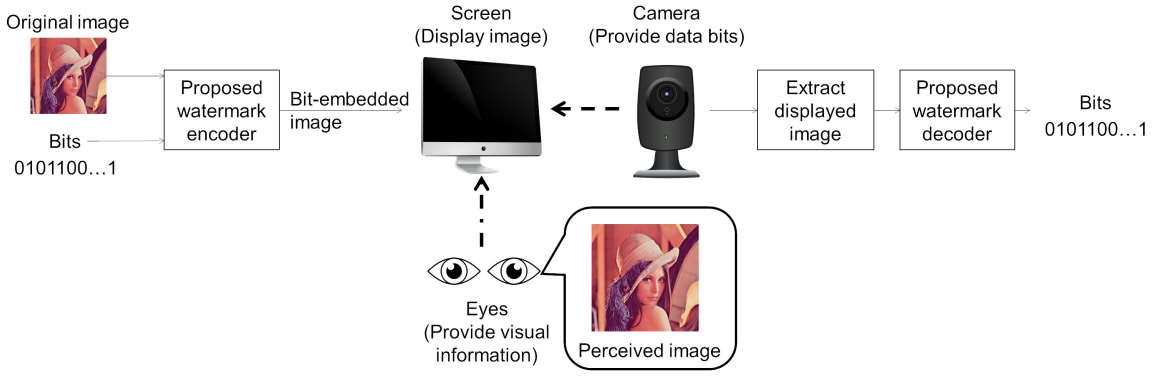


Fig. 1. Proposed watermark encoder and decoder, based on DCNN end-to-end design, for concurrent screen-to-camera and screen-to-eye communications.

the proposed scheme integrates DCNN-based image reconstruction, specifically, recently proposed deep image prior (DIP) [22], into simultaneous screen-to-camera and screen-to-eye communications to reliably deliver a large number of bits by using the bit-embedded images with unnoticeable visual quality distortion. The DIP consists of DCNNs to find linear and nonlinear effects for reconstructing clean images even when a lot of bits are embedded into the images. To embed many bits into high-quality images, the proposed scheme learns the weights of the encoding and decoding networks based on loss gradients obtained from the original images and transmission bits, whose datasets were synthetically generated from watermark simulations offline. The trained DCNN weights are later used by the receiver to decode bits from newly bit-embedded images. There was no study addressing the impact of the DCNN-based approach in simultaneous screen-to-camera and screen-to-eye communications, to the best of authors' knowledge.

Evaluation results show that the proposed scheme reliably sends 98.6% of data bits from the bit-embedded image with the visual quality of 0.96 in terms of structural similarity (SSIM). In addition, we discuss the impacts of parameter regularization in the loss function and number of learning epochs on achievable throughput and visual quality of the proposed scheme.

II. DCNN-based Concurrent Screen-to-Camera and Screen-to-Eye Communications

A. Overview

The purpose of our study is to simultaneously provide many digital bits for camera devices and clean visual information for human eyes from the same bit-embedded images. Fig. 1 shows the schematic of our proposed scheme. We use a pair of screen and camera as the sender and receiver. Note that there are three major differences between RF wireless communications. First, input values for the screen, i.e., pixel luminance values, should not be complex-valued numbers. Second, the input values are two dimensional (2D) in the spatial domain. Third, the pixel luminance values typically range over finite non-negative floats between 0 and 1.

The sender first watermarks $H \times W$ data bits into the images with the resolution of $H \times W$ pixels with three color channels by using the proposed DCNN-based watermark encoder. The proposed DCNN-based watermark encoder generates the bit-embedded images to display them on the screen. At the receiver, the displayed images are captured by camera sensors and human eyes. The displayed images provide data bits for camera sensors while visual information for human eyes. We then feed the images captured by the camera sensors into the proposed DCNN-based watermark decoder to decode the data bits from the captured images. This framework is known as end-to-end learning, which jointly optimizes encoder and decoder DCNNs.

B. Watermark Embedding Networks

The proposed watermark encoder embeds data bits to images and then generates high-quality bit-embedded images by using DIP-based image reconstruction [22]. The proposed watermark decoder then decodes the transmission bits from the captured and bit-embedded images. Specifically, the proposed watermark encoder consists of convolutional neural networks with skip connections as shown in the left side of Fig. 2, where the ith original image $p_i \in \mathbb{R}^{3\times W\times H}$ and ith data bits $b_i \in \{0,1\}^{W\times H}$ are fed into the encoder to generate the bit-embedded image $f_{\theta_e}(p_i, b_i) \in \mathbb{R}^{3\times W\times H}$ with θ_e being the weights of the encoding networks. The proposed watermark decoder also consists of convolutional neural networks as shown in the right side of Fig. 2, where the bit-embedded image $f_{\theta_e}(p_i, b_i) = \hat{p}_i \in \mathbb{R}^{3\times W\times H}$ is fed into the decoder to retrieve the transmission bit $g_{\theta_d}(\hat{p}_i) \in \{0,1\}^{W\times H}$ with θ_d being the weights of the decoding networks. Both watermark encoding and decoding networks have three functionalities for embedding many bits without visual quality degradation: feature encoding, skip-connections, and feature decoding [22].

The feature encoding can be divided into two parts: the first part contains convolution, down-sampling, batch normalization, and leaky rectified linear unit (ReLU) layers while the second part contains convolution, batch normalization, and leaky ReLU layers. The convolutional layers reduce the spatial resolution of the feature maps. For the down-sampling, we use

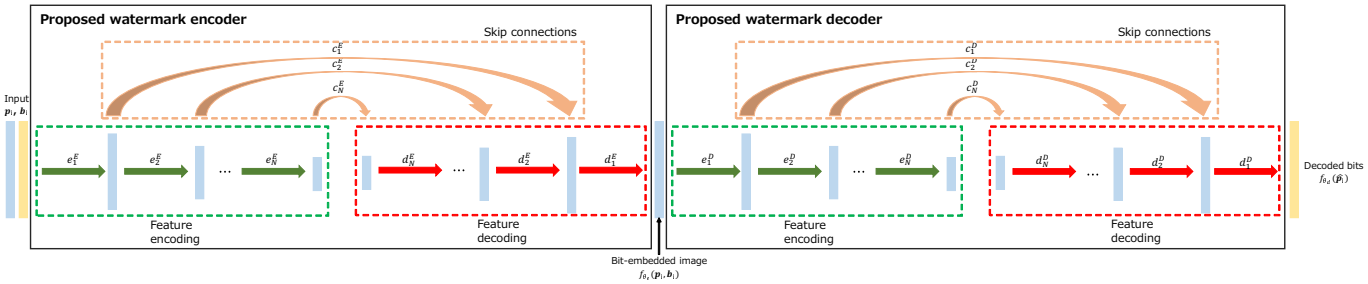


Fig. 2. Proposed DIP-based watermark encoder and decoder networks.

the striding function implemented within convolution modules. The batch normalization standardizes the input values to balance the values in order to reduce the effect of outliers. The leaky ReLU is used for tuning the nonlinearity.

The skip connection is used to propagate the loss gradients due to the watermark from the output to the input to quickly reach the optimal allocation of all the network weights likewise residual networks. Specifically, this part uses convolution, batch normalization, and leaky ReLU layers to predict residuals, which has been verified to be more robust.

The feature decoding has two parts: the first part contains batch normalization, convolution, batch normalization, and leaky ReLU layers while the second part has batch normalization, convolution, and leaky ReLU layers. Finally, the decoding part takes up-sampling operation based on the bilinear upsampling to increase the spatial resolution for the reconstruction of the bit-embedded image/transmission bits.

C. Weight Learning

The proposed watermark encoder networks can embed many bits into an image with high visual quality while the proposed watermark decoder networks decode many data bits from the image based on the weights between the network components. To learn better weights for high-performance watermark encoding and decoding, image and transmission bit datasets are generated offline via Monte–Carlo simulations. Specifically, all potential distortions due to bit embedding and decoding are synthetically analyzed by both proposed watermark encoder and decoder in the off-line learning phase. By using synthetic datasets for the pairs of original image and data bits, the proposed scheme can learn better network weights of the watermark encoder and decoder at the same time for high-quality bit-embedded image and reliable data communication.

We note that the proposed watermark encoder and decoder need to account for two adversarial objective functions simultaneously; i) minimizing visual quality distortion by the encoder networks, and ii) maximizing the reliability of digital bit retrieval at the decoder networks. To this end, we utilize a parameter regularizer, which is an additional term in the loss function of image quality, for our loss function. Specifically, our loss function ℓ_{total} with a regularizer is defined as:

$$\ell_{\text{total}} = \ell_{\text{ssim}} + \lambda\, \ell_{\text{binary}}, \tag{1}$$

where ℓ_{ssim} is the perceptual loss function of image quality in terms of SSIM, ℓ_{binary} is a regularization term that imposes a restriction on data communication reliability, i.e., cross-entropy, and λ is an adjustable parameter. The loss function of SSIM is defined as:

$$\ell_{\text{ssim}} = -\text{ssim}\big(f_{\theta_\text{e}}(p_i, b_i), p_i\big), \tag{2}$$

where p_i denotes the ith pair of original image patch and $f_{\theta_\text{e}}(p_i, b)$ denotes the bit-embedded image patch obtained from the proposed watermark encoder with the DCNN weight set of θ_e. It is known that the SSIM metric is more relevant for perceptual image quality than the conventional mean-square-error loss function. On the other hand, the regularization term ℓ_{binary} can be considered to achieve high-reliable data retrieval in a binary classification problem. We define the loss function ℓ_{binary} by using the following binary cross-entropy:

$$\ell_{\text{binary}} = -\frac{1}{WH}\sum_{j=1}^{WH}(b_j \log_2(\hat{b}_j) + (1-b_j)\log_2(1-\hat{b}_j)), \tag{3}$$

where $\hat{b}_j = \sigma(g_{\theta_\text{d}}(\hat{p}_j))$ is the soft-decision decoder output for the jth bit and $\sigma(\cdot)$ is the sigmoid function:

$$\sigma(x) = \frac{1}{1+\exp(-x)}. \tag{4}$$

By training both encoding and decoding network weights θ_e and θ_d using the loss function of SSIM [23] and binary cross entropy, the proposed scheme can generate clean bit-embedded images while many bits can be transmitted from the images. We use a stochastic gradient descent algorithm based on adaptive momentum (Adam) optimizer [24] for weight learning in the proposed watermark encoding/decoding networks.

III. Performance Evaluation

A. Simulation Settings

Metric: We evaluate the performance of the proposed scheme in terms of SSIM and throughput. SSIM can predict the perceived quality of the bit-embedded image. Larger values of SSIM close to 1 indicates higher perceptual similarity between original and bit-embedded images. In view of throughput, we define the achievable throughput as follows.

$$R = WH \cdot \mathcal{I}(b_i; \hat{b}_i), \tag{5}$$

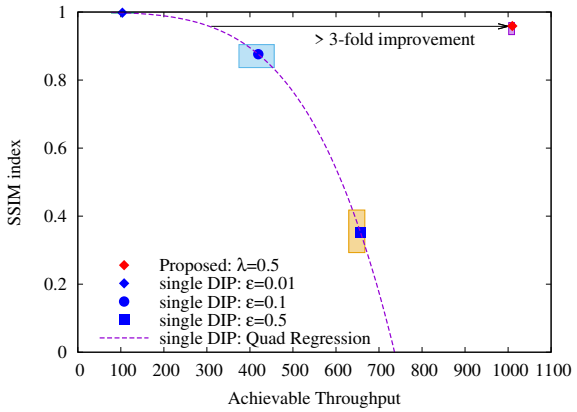


Fig. 3. SSIM index and throughput performance of the single-end DIP and proposed end-to-end schemes.

where R denotes throughput (bits/image), WH is the total number of transmitted bits in one bit-embedded image, and $\mathcal{I}(b_i;\hat{b}_i)$ is an empirical mutual information between transmission data bits b_i and decoded data bits $\hat{b}_i$. Here, the lower bound of the mutual information is calculated directly from the DCNN outputs via cross-entropy loss as follows:

$$\mathcal{I}(b_i;\hat{b}_i) = 1 - \ell_{\text{binary}}. \tag{6}$$

Test Images: We use the benchmark dataset, namely, CIFAR-100 [25] for evaluations. CIFAR-100 consists of multiple training images and testing images with the resolution of 32×32 pixels and 100 classes. The training images are used for learning the network weights while the testing images are used for comparison in terms of achievable throughput and visual quality. We consider 50,000 training images and 100 testing images for evaluations of the proposed watermark method.

B. Baseline Performance

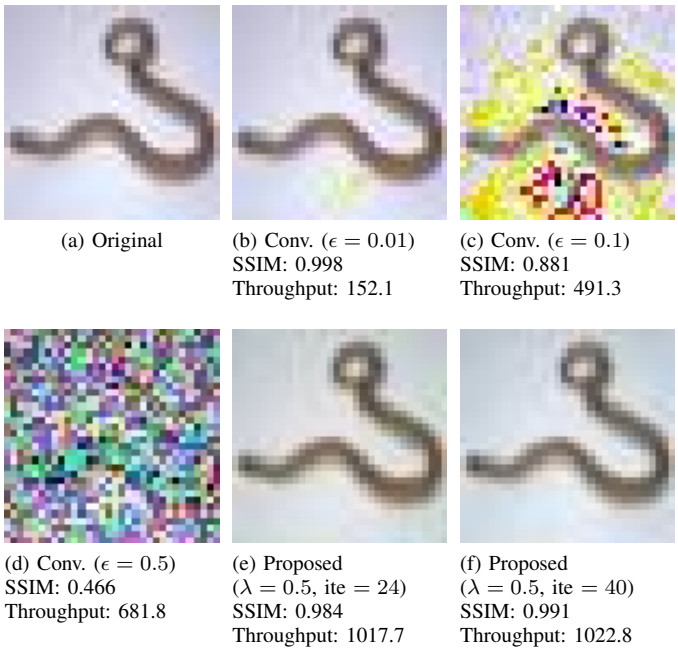
We first evaluate the throughput and visual quality of reference schemes to discuss the baseline performance of the proposed scheme. In order to demonstrate the benefit of end-to-end learning, we consider a single-end learning method to compare; specifically, DIP-based decoder is trained to retrieve digital bits under a conventional superposition watermarking. This scheme, namely, **single DIP**, linearly embeds data bits into a plain image using a small perturbation parameter ϵ as follows:

$$\hat{p}_i = p_i + \epsilon\, b_i, \tag{7}$$

Specifically, the sender transmits the superposition image and the receiver takes DIP-based watermark decoder to reconstruct data bits from the superposition image.

Fig. 3 shows the throughput and visual quality of the single DIP and proposed schemes at the epochs of 24. The results give us the following key observations:

- The proposed scheme simultaneously achieves better visual quality in bit-embedded images and higher achievable throughput compared with the single DIP schemes irrespective of ϵ values.



(a) Original

(b) Conv. ($\epsilon = 0.01$)
SSIM: 0.998
Throughput: 152.1

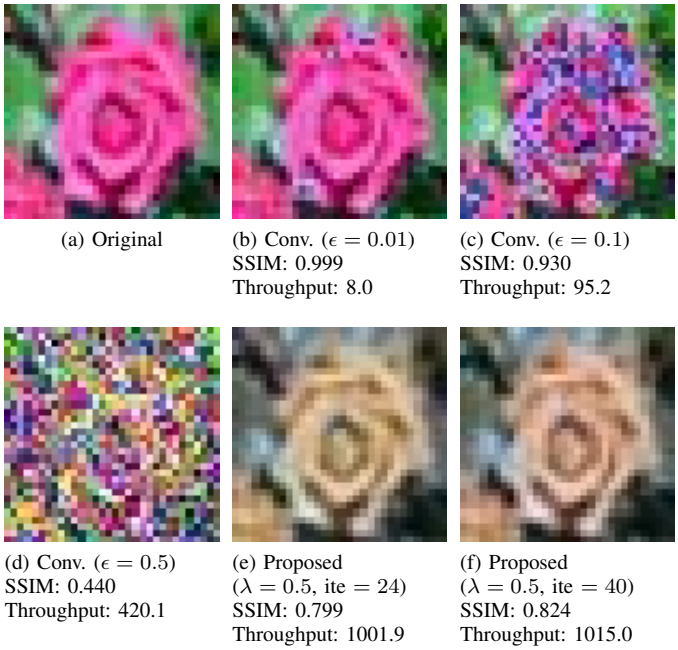
(c) Conv. ($\epsilon = 0.1$)
SSIM: 0.881
Throughput: 491.3

(d) Conv. ($\epsilon = 0.5$)
SSIM: 0.466
Throughput: 681.8

(e) Proposed
($\lambda = 0.5$, ite $= 24$)
SSIM: 0.984
Throughput: 1017.7

(f) Proposed
($\lambda = 0.5$, ite $= 40$)
SSIM: 0.991
Throughput: 1022.8

Fig. 4. Snapshots of CIFAR-100 in conventional and proposed watermarking schemes, having the best visual quality.



(a) Original

(b) Conv. ($\epsilon = 0.01$)
SSIM: 0.999
Throughput: 8.0

(c) Conv. ($\epsilon = 0.1$)
SSIM: 0.930
Throughput: 95.2

(d) Conv. ($\epsilon = 0.5$)
SSIM: 0.440
Throughput: 420.1

(e) Proposed
($\lambda = 0.5$, ite $= 24$)
SSIM: 0.799
Throughput: 1001.9

(f) Proposed
($\lambda = 0.5$, ite $= 40$)
SSIM: 0.824
Throughput: 1015.0

Fig. 5. Snapshots of CIFAR-100 in conventional and proposed watermarking schemes, having the worst visual quality.

- The single DIP scheme improves the throughput as the value of ϵ increases by superposing many data bits into original images while it significantly degrades the visual quality in return.
- The proposed scheme prevents visual quality degradation even with high throughput by using DIP-based image

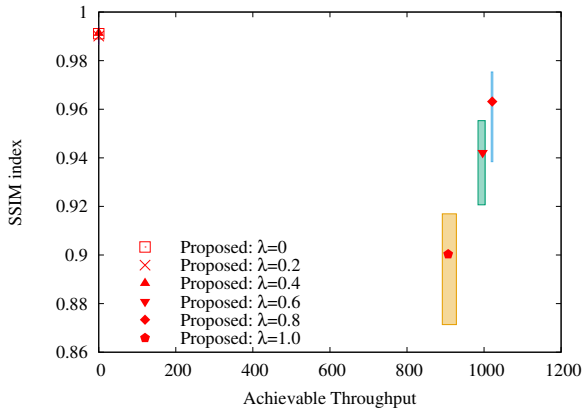


Fig. 6. SSIM index and throughput performance of the proposed scheme under the different adjustable parameters λ.

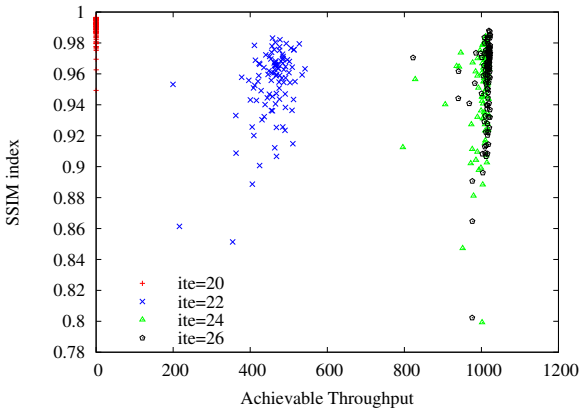
reconstruction for end-to-end learning in both watermark encoding and decoding.

For example, the proposed scheme with $\lambda = 0.5$ achieves 1.54 and 2.41 times higher throughput than the single DIP schemes with ϵ of 0.5 and 0.1, respectively. At the same time, the proposed scheme improves visual quality up to 0.607 and 0.083 in SSIM index compared with the single DIP scheme with ϵ of 0.5 and 0.1, respectively. When we compare with a quadratic regression envelope of single DIP across variable ϵ, the proposed method achieves greater than three-fold improvement in throughput at the same visual quality.

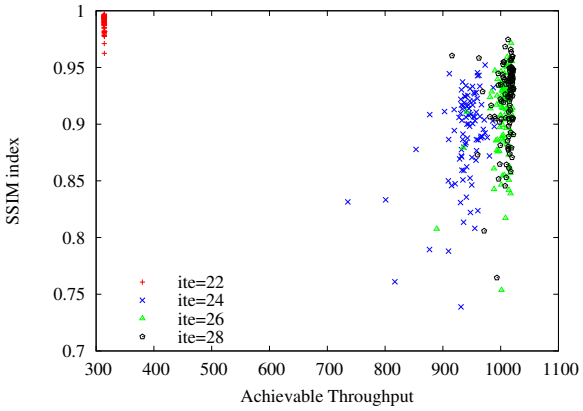
Figs. 4 and 5 show snapshots of the single DIP and proposed schemes with different λ and ϵ values to discuss visual quality of bit-embedded image across the reference schemes. Here, we consider two cases of the proposed schemes at a different learning epoch of 24 and 40. Fig. 4 shows the snapshots of a CIFAR-100 image, whose visual quality was best in the proposed scheme while Fig. 5 depicts the snapshots of another image whose quality was worst among the test datasets.

In Figs. 4(c) and (d), the color information of the bit-embedded image in the single DIP scheme is severely distorted to superpose data bits into the image. In Fig. 4(e), the proposed scheme with a small number of epochs degrades the color information in the background. On the other hand, as shown in Fig. 4(f), the proposed scheme with a large number of epochs realizes improved visual quality closer to the original image shown in Fig. 4(a).

In Figs. 5(c) and (d), the color information of the bit-embedded images in the single-end DIP schemes is also highly distorted due to linear-superposition watermarking of data bits. In addition, the proposed schemes with small and large numbers of learning epochs do not reconstruct the color information of the flower. Nonetheless, the proposed method can embed higher amount of data bits while a moderate SSIM index is maintained. We left how to reconstruct color information irrespective of the bit-embedded image structure as the future work.



(a) Regularizer $\lambda = 0.5$



(b) Regularizer $\lambda = 1.0$

Fig. 7. SSIM index and throughput of the proposed scheme under different number of learning epochs.

C. *Effect of Parameter Regularization*

The visual quality and throughput of the proposed scheme depend on the value of parameter regularization, i.e., λ. In this section, we discuss the impact of parameter regularization in more details.

Fig. 6 shows the visual quality and throughput performance of the proposed scheme with the different values of λ. We can see that the proposed schemes with a large λ value can reliably send data bits using the bit-embedded image. Specifically, the maximum throughput from a single bit-embedded image is 1024 bits while the achievable throughput of the proposed scheme with λ values of 1.0 and 0.8 is 906.8 bits/image and 1021.3 bits/image, respectively. On the other hand, the visual quality of the bit-embedded image is lower than the proposed scheme with a small value of λ. When the value of λ is from 0 to 0.4, the visual quality of the bit-embedded images was nearly identical to the original images. For these cases, the achievable throughput was nearly zero. To achieve higher throughput and visual quality simultaneously, it was found that the proposed scheme should use intermediate λ value, e.g., 0.6, for the parameter regularization.

D. Effect of Learning Epochs

To find better weights of watermark encoding and decoding networks, the stochastic gradient optimization is taken place over training datasets along a certain number of epoch iterations. In this section, we discuss the impact of learning epochs on the performance of the proposed scheme.

Figs. 7(a) and 7(b) show the throughput and visual quality of the proposed scheme at λ values of 0.5 and 1.0, respectively, under the different number of epochs. When the number of epochs is less than 22 and 24 at λ values of 0.5 and 1.0, respectively, bit-embedded images obtained from the proposed scheme have high visual quality while they suffer from low throughput. After the number of epochs of 22 and 24 at λ values of 0.5 and 1.0, the visual quality of the bit-embedded images becomes lower. At the same time, the proposed scheme improves the throughput from the images because the watermarking encoder and decoder may find better weights for reliable communications. Finally, at the number of epochs of 26 and 28 in the λ values of 0.5 and 1.0, the average throughput and visual quality performance of the proposed scheme are better than those of the proposed scheme at the number of epochs of 24 and 26, respectively.

IV. Conclusion

In this paper, we proposed a novel scheme for simultaneous screen-to-camera and screen-to-eye communications by exploiting an adversarial autoencoder framework for end-to-end deep learning. The proposed scheme uses DCNN-based watermark encoder and decoder networks to embed high amount of bits into an image without causing perceptual distortion and to reliably deliver data bits from the bit-embedded image at the same time. Specifically, the proposed scheme uses a loss function with parameter regularizer to account for two objective functions; to minimize the perceptual loss for high visual quality and to maximize data throughput. Evaluation results showed that the proposed watermark networks can generate high-quality bit-embedded images to send many bits. While we focused on CIFAR-100 datasets for proof-of-concept analysis in this paper, more rigorous validation over high-resolution images and videos remains as a future study.

Acknowledgment

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

References

[1] H. Burchardt, N. Serafimovski, D. Tsonev, S. Videv, and H. Haas, "VLC: Beyond point-to-point communication," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 98–105, 2014.

[2] E. Curry, D. Borah, and J. M. Hinojo, "Optimal symbol set design for generalized spatial modulations in MIMO VLC systems," in *IEEE GLOBECOM*, 2016, pp. 1–7.

[3] A. Wang, Z. Li, C. Peng, G. Fang, and B. Zeng, "InFrame++: Achieve simultaneous screen-human viewing and hidden screen-camera communication," in *ACM MobiSys*, 2015, pp. 181–195.

[4] V. Nguyen, Y. Tang, A. Ashok, M. Gruteser, K. Dana, W. Hu, and E. Wengrowski, "High-rate flicker-free screen-camera communication with spatially adaptive embedding," in *IEEE INFOCOM*, 2016, pp. 1–9.

[5] M. Izz, Z. Li, H. Liu, Y. Chen, and F. Li, "Uber-in-light: Unobtrusive visible light communication leveraging complementary color channel," in *IEEE INFOCOM*, 2016, pp. 1–9.

[6] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. Orlik, "Nonlinear equalization with deep learning for multi-purpose visual MIMO communications," in *IEEE ICC*, 2018, pp. 1–6.

[7] T. Yamazato, I. Takai, H. Okasa, T. Fujii, T. Yendo, S. Arai, M. Andoh, T. Harada, K. Yasutomi, K. Kagawa, and S. Kawahito, "Image-sensor-based visible light communication for automotive applications," *IEEE Communication Magazine*, vol. 52, no. 7, pp. 88–97, 2014.

[8] A. Wang, S. Ma, C. Hu, J. Huai, C. Peng, and G. Shen, "Enhancing reliability to boost the throughput over screen-camera links," in *ACM Annual International Conference on Mobile Computing and Networking*, 2014, pp. 41–52.

[9] W. Hu, H. Gu, and Q. Pu, "Lightsync: Unsynchronized visual communication over screen-camera links," in *ACM Annual International Conference on Mobile Computing and Networking*, 2013, pp. 15–26.

[10] T. W. Kan, C. H. Teng, and W. S. Chou, "Applying QR code in augmented reality applications," in *ACM International Conference on Virtual Reality Continuum and its Applications in Industry*, 2009, pp. 253–257.

[11] S. Perli, N. Ahmad, and D. Katabi, "PixNet: Interference-free wireless links using LCD-camera pairs," in *ACM Annual International Conference on Mobile Computing and Networking*, 2010, pp. 137–148.

[12] R. Boubezari, H. L. Minh, Z. Ghasemlooy, and A. Bouridane, "Smartphone camera based visible light communication," *Journal of Lightwave Technology*, vol. 34, no. 17, pp. 4120–4126, 2016.

[13] T. Hao, R. Zhou, and G. Xing, "COBRA: Color barcode streaming for smartphone systems," in *ACM International Conference on Mobile Systems, Applications, and Services*, 2012, pp. 85–98.

[14] T. Li, C. An, X. Xiao, A. T. Campbell, and X. Zhou, "Real-time screen-camera communication bihind any scene," in *ACM MobiSys*, 2015, pp. 1–15.

[15] M. Izz, Z. Li, H. Liu, Y. Chen, and F. Li, "Uber-in-light: Unobtrusive visible light communication leveraging complementary color channel," in *IEEE International Conference on Computer Communications*, 2016, pp. 1–9.

[16] S. Shi, L. Chen, W. Hu, and M. Gruteser, "Reading between lines: High-rate, non-intrusive visual codes within regular videos via implicitcode," in *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, pp. 157–168.

[17] V. Nguyen, Y. Tang, A. Ashok, M. Gruteser, K. Dana, W. Hu, E. Wengrowski, and N. Mandayam, "High–rate flicker–free screen–camera communication with spatially adaptive embedding," in *IEEE International Conference on Computer Communications*, 2016, pp. 1–9.

[18] K. Zhang, C. Wu, C. Yang, Y. Zhao, K. Huang, C. Peng, Y. Liu, and Z. Yang, "ChromaCode: A fully imperceptible screen-camera communication system," in *ACM Annual International Conference on Mobile Computing and Networking*, 2018, pp. 575–590.

[19] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[20] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1256–1272, 2017.

[21] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *IEEE International Conference on Computer Vision (ICCV)*, 2018, pp. 4549–4557.

[22] U. Dmitry, A. Vedaldi, and V. Lempitsky, "Deep image prior," 2017.

[23] P.-H. Su, 2017. [Online]. Available: https://github.com/Po-Hsun-Su/pytorch-ssim

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

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