https://doi.org/10.2352/ISSN.2470-1173.2021.11.HVEI-155

Scrambling Parameter Generation to Improve Perceptual Information Hiding

Koki Madono[1,2]***, Masayuki Tanaka***[2,3]***, Masaki Onishi***[2]***, Tetsuji Ogawa***[1,2]
[1] ***Department of Communications and Computer Engineering, Waseda University; Tokyo, Japan***
[2] ***The National Institute of Advanced Industrial Science and Technology; Tokyo, Japan***
[3] ***Tokyo Institute of Technology; Tokyo, Japan***

Abstract

The present study proposes the method to improve the perceptual information hiding in image scramble approaches. Image scramble approaches have been used to overcome the privacy issues on the cloud-based machine learning approach. The performance of image scramble approaches are depending on the scramble parameters; because it decides the performance of perceptual information hiding. However, in existing image scramble approaches, the performance by scrambling parameters has not been quantitatively evaluated. This may be led to show private information in public. To overcome this issue, a suitable metric is investigated to hide PIH, and then scrambling parameter generation is proposed to combine image scramble approaches. Experimental comparisons using several image quality assessment metrics show that Learned Perceptual Image Patch Similarity (LPIPS) is suitable for PIH. Also, the proposed scrambling parameter generation is experimentally confirmed effective to hide PIH while keeping the classification performance.

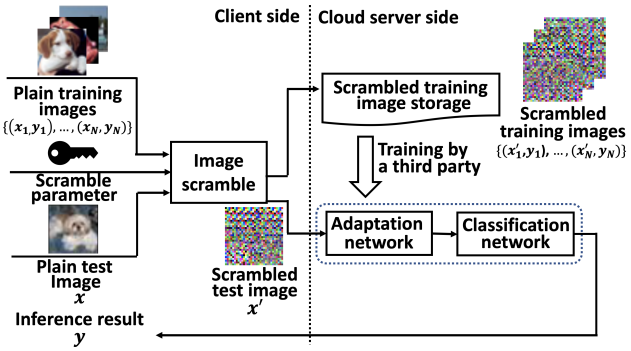


Figure 1. *Image scramble approach. Plain training images are scrambled with a scrambling parameter to sent to the cloud server. On the server-side, a third party constructs a classification model with the adaptation network. In the test phase, the inference result of a scrambled image is sent back to a user.*

Introduction

Recently, the machine learning is used as a high-performance tool by a wide range of users. Concurrently with these demands, cloud-based services such as Google Cloud [1] and Microsoft Azure [2] have attracted much attention because these services enable the computationally expensive algorithms in practice. The service users can easily use a large computational resource with those cloud-based services.

However, these easily-accessible cloud-based machine learning services has a security issue. For example, during development of a new network, the training data may be accessed by someone else on the cloud server if any appropriate privacy preserving algorithm are not in place. In the testing phase, the uploaded test data is open access at least by the service provider if intended. Even if the user uploads the data through secure communication, the privacy issue remains because the plain data is fed to the machine learning network for inference. It means that the cloud service provider can easily access the contents of the data uploaded by the user.

To overcome this security issue, image scramble approach [3, 4, 5, 6] and homomorphic encryption [7, 8, 9, 10, 9, 11, 12] have been proposed. In this paper, we focus on the image scramble approach to overcome the security issue of an image classification task because the homomorphic encryption requires huge computational power and memory despite its mathematical rigorousness.

Now, we consider two types of image information; perceptual information and non-perceptual information. We define the perceptual information is the information which human can perceptually understand. The non-perceptual information is the information which human can not perceptually understand, but machines may be possibly able to recognize. In the image scramble approach as shown in Fig. 1, a plain image is scrambled to hide the perceptual information. The image scramble can hide the perceptual information while the scrambled image still keeps the non-perceptual information. One of the keys to the image scramble approach is an adaptation network. The adaptation network is put before the classification network to improve classification accuracy. In other words, the adaptation network can extract the non-perceptual information.

A perceptual information hiding (PIH) is performed by the image scramble. The image scramble usually has a scrambling parameter. The image scramble with different scrambling parameters yields a different scrambled image. If one knows the scrambling parameter, the scrambled images can be easily restored to the plain image. Therefore, the scrambling parameter is sometimes used as a simple key. It is empirically known that the scrambling parameter affects the PIH performance. However, quantitative analysis for the scambling parameter and the PIH performance has not been addressed on the image scramble approach in the previous studies [3, 4, 5, 6].

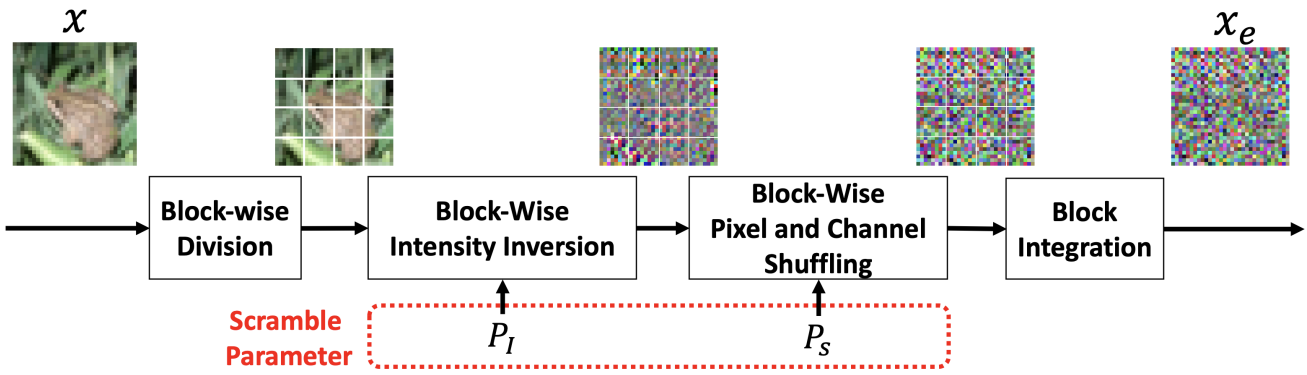


Figure 2. *Learnable image encryption [3] for image scramble, where $\{P_I, P_S\}$ is the scramble parameter.*

Therefore, we propose an effective scrambling parameter generation to improve PIH because the parameter is emperically known as a key of the PIH performance. Our metric comparison shows the Learned Perceptual Image Patch Similarity (LPIPS) [13] is better to evaluate the PIH performance, than any other metrics evaluated. Then, we propose an effective scrambling parameter generation to improve PIH based on the LPIPS metric. We demonstrate that the proposed approach can improve PIH while keeping the classification accuracy.

Related Works

Image Scramble Approach

As mentioned in Introduction, the image scramble approaches [3, 4, 5, 6] have been proposed to overcome the issue of machine learning with the untrusted cloud service. The image scramble includes several operations. Figure 2 shows the image scramble pipeline in the learnable image encryption [3]. The input image is divided into blocks. Then, the intensity values specified by the parameter P_I are inverted. In each block, the intensity location is shuffled with the parameter P_S. Then, the output scrambled image can be obtained by integrating the blocks.

Here, we consider the block size is fixed. Then, parameters of the scramble operation are $\{P_I, P_S\}$. We refer those parameters to scrambling parameter. If one knows the scrambling parameter, one can restore the scrambled image to the plain image. Then, the scrambling parameter can be used as a simple key. It is empirically known that the PIH performance depends on the scramble parameter. However, in the exisiting algorithm, the performance is only validated by the human. To the best of our knowledge, there is no research to quantitatively analyze the PIH performance.

Image Quality Assessment Metrics

An image quality assessment is an important research field in image processing. There are two types of image quality assessment metrics; a full-reference metric and a non-reference metric. The full-reference metric requires the true image for the image quality assessment, while the non-reference metric does not require the true image. RMSE (root mean square error) and PSNR (Peak Signal-to-Noise Ratio) are known as a classical full-reference metric. SSIM (Structural Similarity) become popular [14, 15]. Recently, the full-reference metrics which evaluate the difference in the feature domain of deep network have been proposed [16, 17, 13]. The non-reference metrics [18, 19, 20] basically evaluate the naturalness of the target image. The above image quality assessment metrics have been proposed to evaluate the goodness of the image. In this paper, we try to find a suitable metric to evaluate the PIH performance.

Metric Selection for Perceptual Information Hiding

This section describes a detail of LPIPS and a metric comparison to show that LPIPS is a suitable metric for perceptual information hiding (PIH). It aims at verifying a suitable metric for PIH on the proposed scramble parameter generation. A detail of LPIPS is first explained, the metric comparison is, then, conducted using nine metric candidates.

LPIPS (Learned Percptual Image Patch Similarity)

To evaluate the perceptual information hiding, we experimentally confirm the LPIPS (Learned Perceptual Image Patch Similarity) [13] is a suitable metric. LPIPS is a network-based image similarity metric to estimate the similarity between the two input images, which was developed to imitate the human perception-based image patch similarity [13]. First, feature maps are extracted from input images with a pre-trained model such as VGG [21], SqueezeNet [22] or AlexNet [23]. Then, a discriminator or a regressor is trained to estimate the human perception-based patch similarity. For that training, they use Berkeley-Adobe Perceptual Patch Similar-ity (BAPPS) Dataset which includes image patch triplets and associated human perception [13]. In this paper, we used pre-trained LPIPS with AlexNet.

Metric Comparison

Before conducting experiments for our proposed scrambling parameter generation, we first evaluated nine metric candidates to demonstrate that LPIPS is the most suitable one for perceptual information hiding. The candidates are:

- **MSE**: Mean Squared Error
- **PSNR**: Peak-to-Signal Noise Ratio
- **SSIM**: Structure Similarity Index Measure [14]
- **BRISQUE**: Blind/Referenceless Image Spatial Quality Evaluator [18]
- **NIQE**: Natural Image Quality Evaluator [19]

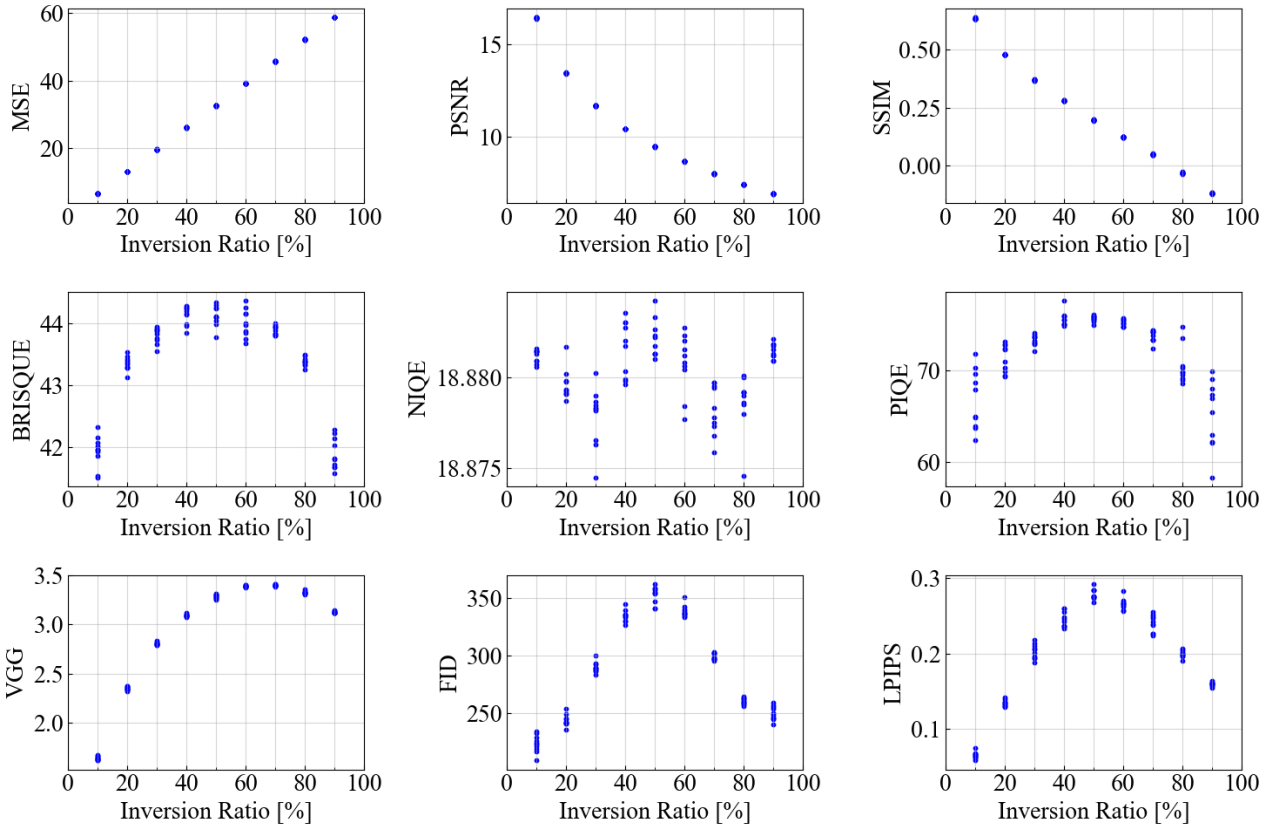


Figure 3. *Different metrics scores and corresponding intensity inversion ratio.*

- **PIQE**: Perception based Image Quality Evaluator [20]
- **VGG**: MSE in VGG's feature domain [17]
- **FID**: Fréchet Inception Distance [16]
- **LPIPS**: Learned Perceptual Image Patch Similarity [13]

A pixel-based image scramble method [6] was used to evaluate the candidate nine metrics because it can control the PIH performance by the single parameter of the inversion ratio. In the pixel-based image scramble method, the image scramble is performed by inverting some intensity values. The scrambling parameter determines combination of channels and intensities to be inverted. Moreover, the PIH performance can be controlled by changing the inversion ratio. We assumed 50% of the inversion ratio has a higher PIH because the scrambling parameter has been coventionally set to the same value of their inversion ratio. It is expected that the PIH performance is the best at the inversion ratio of 50% and becomes worse at a smaller or larger inversion ratio than 50%. Note that humans can recognize the intensity-inverted images. In this sense, the scrambled image with the inversion ratio closer to 100% shows a lower PIH performance. We used the CIFAR-10 dataset [24]. 50,000 images from training dataset were scrambled by the pixel-based image scramble method with different scramble parameters changing the inversion ratio. The mean value of each metric is plotted in Fig. 3, where the experiments were iterated ten times for each inversion ratio. From comparisons in Fig. 3, we can find that BRISQUE, FID, and LPIPS have more suitable properties for PIH metric, which have the highest value at 50% inversion ratio and lower as closer to 0% or 100% of the inversion ratio. LPIPS has the sharpest shape compared with BRISQUE and FID. Then, in this paper, we used the LPIPS as PIH metric.

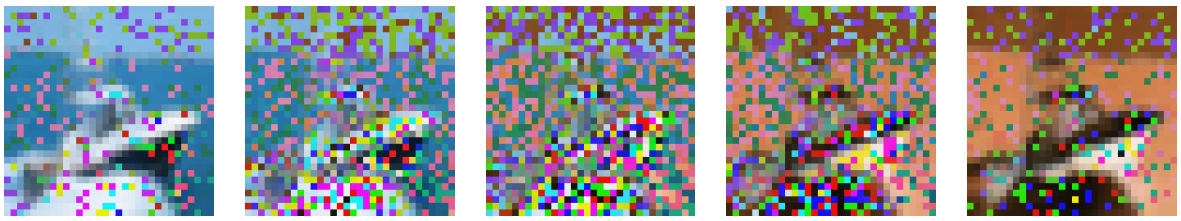
Algorithm 1 Proposed scrambling parameter generation

Input: Images $\{\boldsymbol{x}_i\}$, Number of trials N

Randamly generate the scrambling parameter $\boldsymbol{P}_M$
$\rho_M \leftarrow \rho(\boldsymbol{P}_M; \{\boldsymbol{x}_i\})$
for $j = 2$ to N **do**
    Randamly generate the scrambling parameter $\boldsymbol{P}_j$
    $\rho_j \leftarrow \rho(\boldsymbol{P}_j; \{\boldsymbol{x}_i\})$
    if $\rho_M < \rho_j$ **then**
        $\rho_M \leftarrow \rho_j$
        $\boldsymbol{P}_M \leftarrow \boldsymbol{P}_j$
Return the scrambling parameter $\boldsymbol{P}_M$

Table 1 shows examples of scrambled images with differnt inversion ratios and the associated LPIPS values. One can find the scrambled image at 50% of the inversion ratio is the most difficult to recognize. The LPIPS values correspond to those difficulties.

Table 1. Visualization of scramble images with different intensity inversion ratio (IIR) and corresponding LPIPS score.

IIR	10%	30%	50%	70%	90%
					
LPIPS	0.0865	0.2432	0.3809	0.3597	0.2540

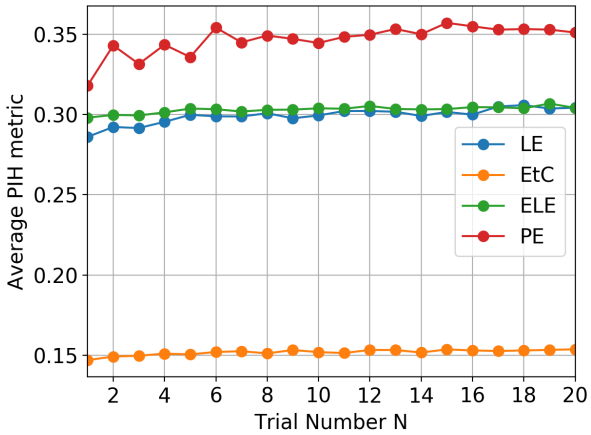


Figure 4. *Average PIH metric at different number of scrambling parameter selections.*

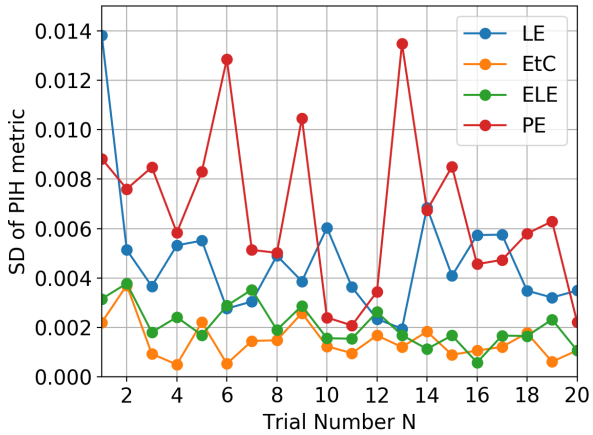


Figure 5. *Standard deviation (SD) of PIH metric at different number of scrambling parameter selections.*

Proposed Scrambling Parameter Generation Considering PIH

In the existing scramble approaches [3, 4, 5, 6], the scrambling parameter is just randomly generated although the PIH performance depends on the scramble parameter. The randomly generated parameter may possibliy result in a poor PIH performance. Therefore, we proposed an algorithm to generate a reliable scrambling parameter considering the PIH performance. Our proposed scrambling parameter generation is based on the PIH metric in the previous section. Algorithm 1 shows a pseudo-code of the proposed scrambling parameter generation. We provide a set of images to be used in PIH metric evaluation and the maximum trial number. Then, the algorithm provides us the scrambling parameter with the highest PIH metric. The proposed method can be applied to any kind of scramble approaches. In algorithm 1, we evaluated the PIH metric by calculating the LIPIS value between the plain image and the scrambled image. Here, the PIH metric ρ based on LPIPS can be expressed as

$$\rho(\boldsymbol{P};\{\boldsymbol{x}_i\}) \quad = \quad \frac{1}{M}\sum_{i=1}^{M}\mathrm{LPIPS}(\boldsymbol{x}_i,\boldsymbol{f}(\boldsymbol{x}_i;\boldsymbol{P}))\,, \tag{1}$$

where $\boldsymbol{P}$ is the scramble parameter, $\{\boldsymbol{x}_i\}$ is a set of images, M is the number of images, $\boldsymbol{f}(\cdot;\boldsymbol{P})$ is the image scramble operation with the scrambling parameter $\boldsymbol{P}$, and LPIPS represents the LPIPS network.

Experiments

The experiments were conducted on the training dataset of CIFAR-10 to evaluate the proposed scrambling parameter generation in a realistic setting with the following four image scrambling approaches:

- **PE**: Pixel-base image Encryption [6]
- **EtC**: Encryption-then-compression [25]
- **LE**: Learnable encryption [3]
- **ELE**: Extended learnable encryption [5]

These scrambling approaches are mainly used for privacy-preserving machine learning which aims to understand by machine, but not by human.

The reproduction code is publicly avaliable online [1].

PIH Improvement

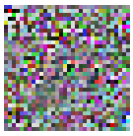
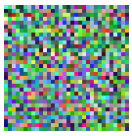
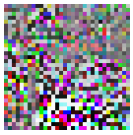
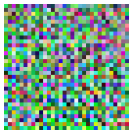
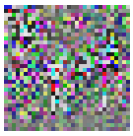
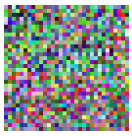
The proposed scrambling parameter generation were computed in the four exsisting image scramble approaches: PE, EtC, LE, and ELE. We iterated ten times for each trial number. Then,

[1] The code will be avaliable after acceptance.

Table 2. Mean and standard deviation (SD) of classification accuracy and PIH metric of the original four image scramble approaches and those with the proposed scrambling parameter generation.

	Original				Proposed scrambling parameter generation			
	ELE	EtC	PE	LE	ELE	EtC	PE	LE
Acc. (mean)	69.49	79.42	93.40	93.75	71.38	78.47	93.58	93.05
Acc. (SD)	1.6205	2.9198	0.3114	0.2620	3.1268	3.3404	0.1201	0.8164
PIH metric (mean)	0.3000	0.1482	0.3161	0.2880	0.2999	0.1513	0.3420	0.3016
PIH metric (SD)	0.0026	0.0033	0.0038	0.0091	0.0030	0.0014	0.0075	0.0046

Table 3. Predicted class, corresponding posterior probability and LPIPS by PE and LE with the proposed scrambling parameter generation.

	Plain image	PE (proposed)	LE (proposed)
			
pred. class	frog	frog	frog
probability	0.9902	0.9993	0.9978
LPIPS	0	0.3701	0.2804
			
pred. class	automobile	automobile	automobile
probability	0.9971	0.9986	0.9999
LPIPS	0	0.3874	0.3553
			
pred. class	horse	horse	horse
probability	0.9994	0.9997	0.9992
LPIPS	0	0.4070	0.3130

the average and the standard deviation of the PIH metrics were evaluated. Figure 4 shows the average PIH metrics. Comparing the original approach, the PIH metric is slightly improved by the proposed scrambling parameter generation in terms of the average value, except ELE algorithm. The original ELE already has a very high PIH metric value. As discussed in the next section, the ELE algorithm has lower classification accuracy. It implies that there is a trade-off relationship between classification accuracy and PIH performance.

Figure 5 show the standard deviation of the PIH metic. This result demonstrates the proposed scrambling parameter generation effectively decreases the standard deviation of the PIH metric while the PIH performance of the PE algorithm is significantly unstable. The PIH performance depends on the trial number. In this paper, we set 20 for the trial number from Fig. 5.

Classification Accuracy and PIH

For the classification network, we used the shakedrop network [26]. We also adopted the adaptation network before the classification network. The adaptation network is proposed in each image scramble approach [3, 4, 5, 6]. The datasets used for the evaluation were CIFAR-10. The mini-batch size was 64 during training and testing. The SGD with the Nesterov was used as the optimizer, where the momentum was 0.9. The network was trained with 100 epochs of iterations. The learning rate was scheduled as 0.1 for 0-to-50 epochs, 0.01 for 50-to-75 epochs, and 0.001 for 75-to-100 epochs.

Table 2 shows the mean and standard deviation of classification accuracy and PIH metric of the original four image scramble approaches and those with the proposed scrambling parameter generation. The mean accuracies of The ELE and EtC of both the original and with the proposed scrambling parameter generation are lower than those of PE and LE. Then, we focus on PE and LE. The proposed scrambling parameter generation improves the PIH performance of the PE and LE, while the accuracies of them are comparable. The standard deviations of the PIH of PE and LE are decreased by the proposed scrambling parameter generation. It means that the proposed scrambling parameter generation can constantly generate the good PIH performance of the scrambling parameter.

Table 3 shows examples of the scrambled images and associated LIPIS values of the PE and LE algorithms with the proposed scrambling parameter generation. The proposed PE and LE of predicted class and corresponding class probability are same as plain images. However, it is very difficult for the human to recognize the scrambled images in Table 3.

Conclusion

In this paper, we have investigated a suitable metric to evaluate the perceptual information hiding (PIH) on the scramble approach. Then, we have shown the LPIPS metric has better properties to evaluate the PIH performance. Based on the PIH metric, expeirmental comparisons demonstrate that the proposed scrambling parameter generation can improve the PIH performance.

Furthermore, we have experimentally verified that the image scramble approach can provide a reasonablly good classification performance with the proposed scrambling parameter generation in order to hide the perceptual information of the image.

References

[1] Alphabet Inc., “Google cloud..” https://cloud.google.com.

[2] Microsoft Corp., “Microsoft azure..” https://azure.microsoft.com.

[3] M. Tanaka, “Learnable image encryption,” *2018 IEEE In-*

ternational Conference on Consumer Electronics-Taiwan (ICCE-TW), pp. 1–2, 2018.

[4] T. Chuman, W. Sirichotedumrong, and H. Kiya, "Encryption-then-compression systems using grayscale-based image encryption for jpeg images," *IEEE Transactions on Information Forensics and Security*, vol. 14, pp. 1515–1525, 2019.

[5] K. Madono, M. Tanaka, M. Onishi, and T. Ogawa, "Block-wise scrambled image recognition using adaptation network," *ArXiv*, vol. abs/2001.07761, 2020.

[6] W. Sirichotedumrong, T. Maekawa, Y. Kinoshita, and H. Kiya, "Privacy-preserving deep neural networks with pixel-based image encryption considering data augmentation in the encrypted domain," *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 674–678, 2019.

[7] S. Singh, Y.-S. Jeong, and J. H. Park, "A survey on cloud computing security: Issues, threats, and solutions," *J. Network and Computer Applications*, vol. 75, pp. 200–222, 2016.

[8] Z. Shan, K. Ren, M. Blanton, and C. Wang, "Practical secure computation outsourcing: A survey," *ACM Comput. Surv.*, vol. 51, pp. 31:1–31:40, 2018.

[9] T. van Elsloo, G. Patrini, and H. Ivey-Law, "Sealion: a framework for neural network inference on encrypted data," *ArXiv*, vol. abs/1904.12840, 2019.

[10] R. Gilad-Bachrach, N. Dowlin, K. Laine, K. E. Lauter, M. Naehrig, and J. R. Wernsing, "Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy," in *ICML*, 2016.

[11] P. Xie, B. Wu, and G. Sun, "Bayhenn: Combining bayesian deep learning and homomorphic encryption for secure dnn inference," in *IJCAI*, 2019.

[12] Q. Lou, B. Feng, G. C. Fox, and L. Jiang, "Glyph: Fast and accurately training deep neural networks on encrypted data," *ArXiv*, vol. abs/1911.07101, 2019.

[13] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 2004.

[15] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, vol. 2, pp. 1398–1402 Vol.2, 2003.

[16] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *NIPS*, 2017.

[17] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *ECCV*, 2016.

[18] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, pp. 4695–4708, 2012.

[19] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, pp. 209–212, 2013.

[20] D. Temel, M. Prabhushankar, and G. Al-Regib, "Unique: Unsupervised image quality estimation," *IEEE Signal Processing Letters*, vol. 23, pp. 1414–1418, 2016.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2015.

[22] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and ¡1mb model size," *ArXiv*, vol. abs/1602.07360, 2017.

[23] A. Krizhevsky, "One weird trick for parallelizing convolutional neural networks," *ArXiv*, vol. abs/1404.5997, 2014.

[24] A. Krizhevsky, "Learning multiple layers of features from tiny images," 2009.

[25] T. Chuman, W. Sirichotedumrong, and H. Kiya, "Encryption-then-compression systems using grayscale-based image encryption for jpeg images," *IEEE Transactions on Information Forensics and Security*, vol. 14, pp. 1515–1525, 2018.

[26] Y. Yamada, M. Iwamura, T. Akiba, and K. Kise, "Shakedrop regularization for deep residual learning," vol. abs/1802.02375, 2018.

Author Biography

Koki Madono received the B.S. degree in Fundamental Science and Engineering from Waseda University, Tokyo, Japan, in 2019. He is currently pursuing the master's degree with the Department of Fundamental Science and Engineering at Waseda University. He has been a Research Assistant at the National Institute of Advanced Industrial Science and Technology (AIST) since 2019. His research interests include artificial intelligence, machine learning, image processing, privacy, and security.

Masayuki Tanaka received his bachelor and master degrees in control engineering and Ph.D. degree from Tokyo Institute of Technology in 1998, 2000, and 2003. He joined Agilent Technology in 2003. He was a Research Scientist at Tokyo Institute of Technology since 2004 to 2008. Since 2008, He has been an Associate Professor at the Graduate School of Science and Engineering, Tokyo Institute of Technology. He was a Visiting Scholar with Department of Psychology, Stanford University, CA, USA.

Masaki Onishi received the M.Eng. and Dr.Eng. degrees fron the Osaka Prefecture University in 1999 and 2002, respectively. From 2002 to 2006, he was a research scientist at the Bio-Mimetic Control Research Center, RIKEN. Since 2006, he has been a research scientist at the National Institute of Advanced Industrial Science and Technology (AIST). His research interests are computer vision, video surveillance, and human-robot interaction.

Tetsuji Ogawa received the B.S., M.S., and Ph.D. degrees in electrical engineering from Waseda University, Tokyo, Japan, in 2000, 2002, and 2005, respectively. He was a Research Associate from 2004 to 2007, and a Visiting Lecturer in 2007 at Waseda University. From 2007 to 2012, he was an Assistant Professor at Waseda Institute for Advanced Study. Since 2012, he has been an Associate Professor at Waseda University and the Egypt–Japan University of Science and Technology. He was a Visiting Scholar in the Center for Language and Speech Processing, Johns Hopkins University, Baltimore, MD, USA, from June to September in

2012 and from June to August in 2013. He was a Visiting Scholar in Speech Processing Group and a Faculty of information technology at the Brno University of Technology, Czech Republic, from June to July in 2014 and May to August in 2015. His research interests include stochastic modeling for pattern recognition, speech enhancement, and speech and speaker recognition.

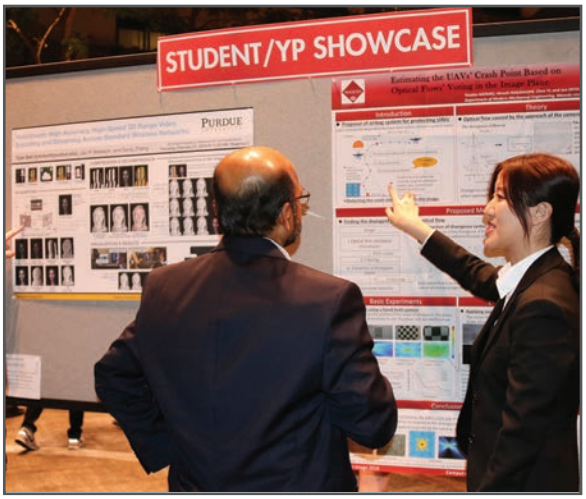
JOIN US AT THE NEXT EI!

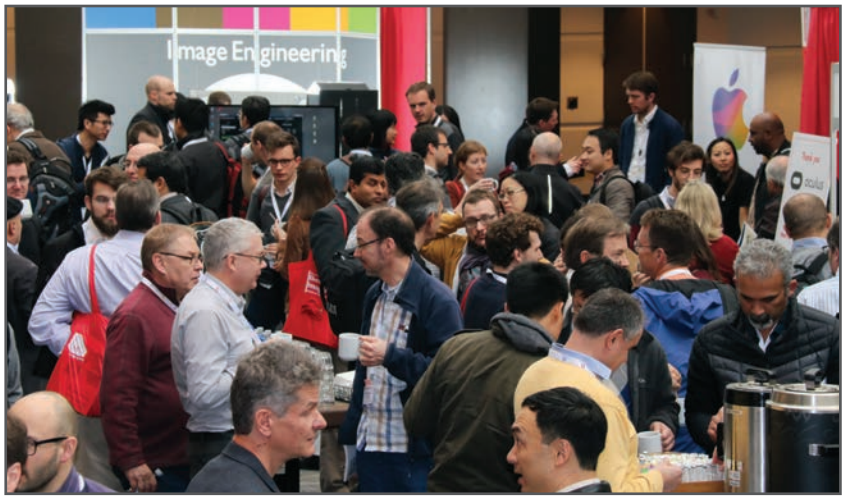
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org