Towards Secured and Transparent Artificial Intelligence Technologies in Hierarchical Computing Networks

Yitu Wang and Takayuki Nakachi

Abstract

Researchers at NTT Network Innovation Laboratories have recently been focusing on the interdiscipline of transparent artificial intelligence (AI) technologies and hierarchical computing networks. A hierarchically distributed computing structure not only improves the quality of computation but also creates an extra degree of diversity for algorithm refinement. Sparse coding, an important transparent AI technique, is finding application in this new domain. We propose in this article a secure sparse coding scheme that enables computing directly on cipher-texts. We also demonstrate its application to image compression and face recognition in edge and cloud networks.

Keywords: secure sparse coding, edge and cloud, image compression, face recognition

1. Introduction

We are witnessing a sense of excitement in the research community and a frenzy in the media regarding advances in artificial intelligence (AI). Remarkable progress has been made in a variety of AI tasks such as image classification and speech comprehension by making use of deep neural networks [1]. However, we need to be able to fully trust the algorithmic prescriptions before we can readily accept and apply them in practice. For this reason, interpretability and explainability are two essential ingredients in AI algorithm design.

Sparse coding was inspired by the sparsity mechanism of nature [2] and has received considerable attention as a representative transparent AI technique. For example, the sparsity mechanism exists in the human vision system. A learning algorithm that attempts to find sparse linear codes for natural scenes will develop a complete family of localized, oriented, bandpass receptive fields. It has proved to be an extraordinarily powerful solution in a wide range of application fields, especially in signal processing, image processing, machine learning, and computer vision [3]. In addition, sparse coding demonstrates strong potential in practical implementations because it is sufficiently flexible to capture much of the variation in real datasets and provides insights into the features extracted from the dataset.

For the purposes of practical application and providing further support to real-time services, the computational complexity involved should not be overlooked. These complex and well-engineered AI approaches require substantial effort in parameter tuning, and they pose exigent requirements on computation capability, which cannot be easily satisfied by solely relying on devices due to their limited resources, for example, limited computation capability and memory size.

A cloud built on top of a datacenter, which seamlessly integrates storage and computation, could be an ideal platform for implementing the aforementioned algorithms. It, however, faces significant challenges in data collection and service distribution over the network, given that the devices in service are globally and remotely distributed. One promising solution to address these limitations is to make use of the hierarchically distributed computing structure consisting of the edge, cloud, and devices, as shown in **Fig. 1** [4]. This configuration makes it possible to

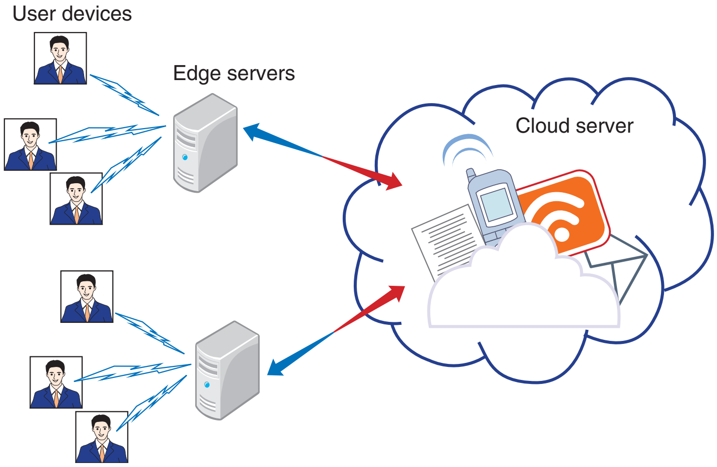


Fig. 1. Edge and cloud network.

not only substantially reduce the tension between computation-intensive applications and resource-limited devices, but also to completely avoid the long latency incurred in the information exchange between devices and the cloud in wide area networks [4]. In addition, the relative uniqueness of the information available from various devices in service prompts the algorithms to capture different patterns along the system dynamics, and in turn, creates an extra degree of diversity. Therefore, transparent AI techniques are being applied in this new domain to reduce computation demands at each device, while further enhancing the performance by exploiting the multi-device diversity.

To exploit more dimensions of edge and cloud resources for purposes other than just fulfilling computation demands, we allow the cloud to produce a joint computing result based on information obtained from each device and try to investigate what the fundamental benefit is of exploiting the multi-device diversity. However, this could lead to serious privacy concerns, as the private information could be collected and misused without permission by the third party. Commonly deployed encrypting algorithms such as advanced encryption standard and secure hash algorithm provide the capability of security, but they cannot render the designed algorithms valid; that is, computing cannot be carried out only with the encrypted data. Even though algorithms such as homomorphic encryption and secure multi-party computation enable computing on cipher-texts, they are faced with the curse of dimensionality regarding the size of data and thus incur significant computational complexity.

In this article, we report a secure sparse coding scheme with low complexity based on random unitary transform, which enables sparse modeling based algorithms to directly compute on the encrypted data. Moving one step ahead, we further demonstrate its application to image compression and face recognition in edge and cloud networks and show the superiority of the proposed framework through simulation results.

2. Secure sparse coding

As illustrated in **Fig. 2**, an observed signal set $\boldsymbol{Y}$ can be represented as the linear combination of only a few atoms from the dictionary $\boldsymbol{D}$. The core sparse representation problem is defined as the quest to find the sparsest possible representation $\boldsymbol{X}$. Due to the underdetermined nature of $\boldsymbol{D}$, this linear system offers in general many infinitely possible solutions, and among these we seek the one with the fewest nonzeros. This problem is known to be NP (nondeterministic polynomial time)-hard with a reduction to NP-complete subset selection problems in combinatorial optimization.

Alternatively, it is possible to find an approximate solution by taking the following two steps.

1) Dictionary training

Given a training set $\boldsymbol{Y}_{train}$, learning a reconstructive dictionary with K atoms for obtaining the sparsest representation can be accomplished by solving the following optimization problem,

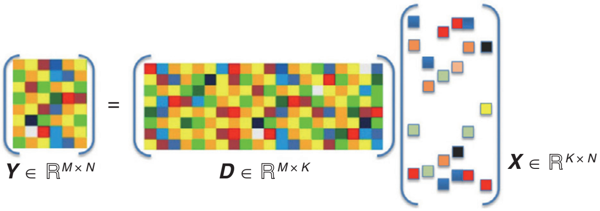

$Y \in \mathbb{R}^{M \times N}$
=
$D \in \mathbb{R}^{M \times K}$
$X \in \mathbb{R}^{K \times N}$

Fig. 2. Sparse coding.

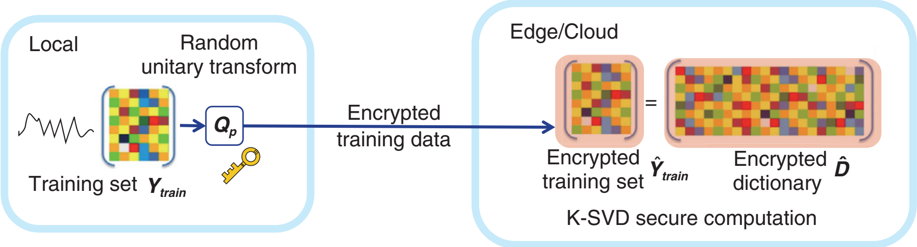


Fig. 3. Dictionary training based on encrypted data.

$$\arg\min_{D} \|Y_{train} - DX\|_2^2, \ s.t., \ \|x_i\|_0 \leq \in, \forall \mathrm{i} \in Y_{train}, \quad (1)$$

where D is the learned dictionary, X is the sparse representation, and $\in$ is the sparsity constraint factor. We can apply K-SVD to find an approximate solution.

2) Sparse representation

Given a testing sample $y \in Y$, a sparse representation $x \in X$ based on the trained dictionary D can be calculated by

$$\arg\min_{x} \|y - Dx\|_2^2, \ s.t., \ \|x\|_0 \leq \in. \quad (2)$$

The above optimization problem can be efficiently solved using orthogonal matching pursuit (OMP).

To address the security issue, we adopt random unitary transform, which not only proves to be effective for biometric template protection, but also has the desired low computational complexity for application in scenarios with a large cipher-text size [5]. Any vector v encrypted by random unitary matrix Q_p with private key p can be expressed as $\hat{v} = Q_p \times v$, where $\hat{v}$ is the encrypted vector and Q_p satisfies $Q_p^* \times Q_p = I$, where $(\cdot)^*$ and I respectively represent the Hermitian transpose and the identity matrix. Gram-Schmidt orthogonalization can be adopted for generating Q_p. This encryption technique has been proved to be robust in terms of brute-face attacks, diversity, and irreversibility.

The process to extract the feature dictionary from the encrypted data is depicted in **Fig. 3**. The user device first encrypts its training data locally, which are then transmitted to the edge server in close proximity via wireless channels. Then the dictionary is trained directly using the encrypted data at the edge/cloud.

We proved in a previous study [6] that the relationship between the dictionary $\hat{D}$ trained from the encrypted data $\hat{X}$, and D trained from the original data X satisfies $\hat{D} = Q_p \times D$.

The process to obtain the secure sparse representation is shown in **Fig. 4**. The user device first encrypts its testing data locally. After receiving the encrypted data, the edge server calculates its sparse representation using the trained dictionary.

We have proved that the sparse representation $\hat{x}$ of the encrypted data $\hat{y}$ based on the encrypted dictionary $\hat{D}$ is identical to x of the original data y based on the dictionary D, that is, $\hat{x} = x$ [7].

3. Application to image compression

The traditional method for secured image transmission is based on compression-then-encryption (CtE) systems. In CtE systems, images are uploaded to social networking service (SNS) providers by users with the assumption that the entire process can be

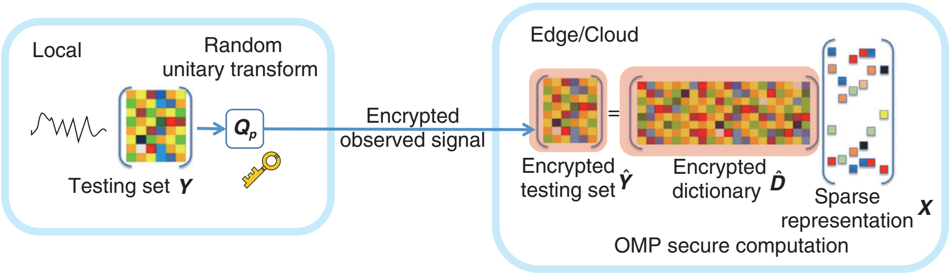

Local
Random unitary transform
Testing set Y
Q_p
Encrypted observed signal
Edge/Cloud
=
Encrypted testing set $\hat{Y}$
Encrypted dictionary $\hat{D}$
Sparse representation X
OMP secure computation

Fig. 4. Secure sparse representation.

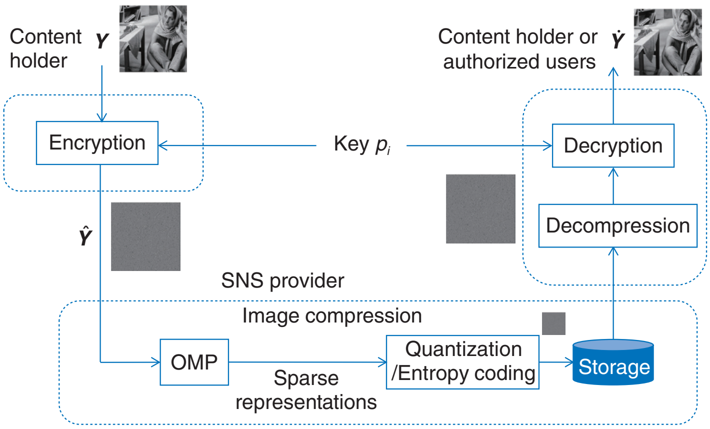


Fig. 5. EtC system using secure sparse coding.

trusted, but the privacy of the uploaded images cannot actually be controlled by the users. Therefore, there are serious concerns about the privacy of those uploaded images because SNS providers simply take full control of this process.

Encryption-then-compression (EtC) systems have been proposed to securely transmit images through an untrusted channel provider. EtC systems enable us to protect unencrypted images from the SNS providers because the encrypted images can only be recovered by authorized users, while enabling recompression by the providers. This approach supports compressing images on the cloud while keeping the image data secure.

The effectiveness of sparse coding in image compression has been reported. One study [8] showed that rate-distortion based sparse coding outperforms JPEG[*1] and JPEG 2000 up to 6+ dB and 2+ dB, respectively. An EtC system using the proposed secured sparse coding for image archives and sharing in SNSs is illustrated in **Fig. 5**. We obtain the sparse representations by feeding the encrypted dictionary $\hat{\boldsymbol{D}}$ and the encrypted image $\hat{\boldsymbol{Y}}$ into the secure OMP computation. The proposed algorithm can easily control the number of sparse representations without decrypting the encrypted images. To this end, the rate-distortion tradeoff can be easily controlled without decrypting the encrypted images.

As shown in **Figs. 6** and **7**, it is difficult to recognize the original image and the dictionary from the encrypted ones, and it would be computationally expensive to obtain the original image and dictionary from the encrypted ones without knowledge of the private key. The authorized user can recover the image $\dot{\boldsymbol{Y}} = \boldsymbol{Q}_p^* \hat{\boldsymbol{D}} \hat{\boldsymbol{X}}$ based on the encrypted dictionary $\hat{\boldsymbol{D}}$ and its sparse representation $\hat{\boldsymbol{x}}$, as shown in **Fig. 8(a)**. The image cannot be recovered by an unauthorized user, as shown in **Fig. 8(b)**.

Moreover, the trained dictionary also demonstrates

[*1] JPEG: JPEG is a standard format for image compression developed by the Joint Photographic Experts Group.

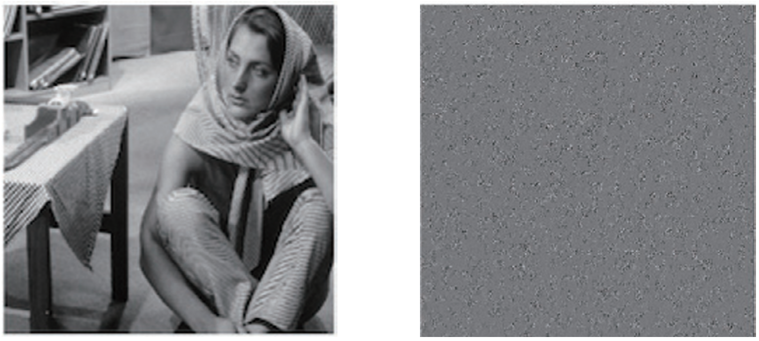


Fig. 6.   Image before (left) and after (right) encryption.

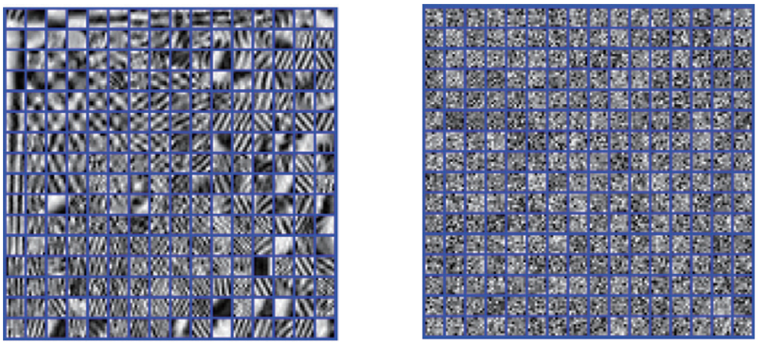
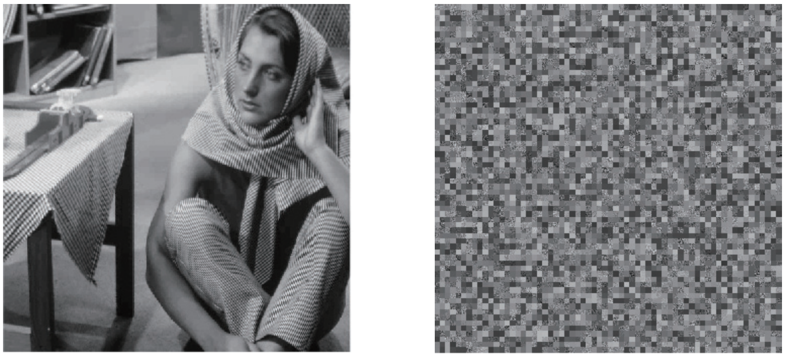


Fig. 7.   Dictionary before (left) and after (right) encryption.



(a) Authorized user (PSNR=39.28)

(b) Unauthorized user (PSNR=10.40)

PSNR: peak signal-to-noise ratio

Fig. 8.   Recovered images.

a representative capability. The rate-distortion performance (average sparsity ratio vs. decoded/decrypted image quality peak signal-to-noise ratio (dB)) when compared with overcomplete discrete cosine transform

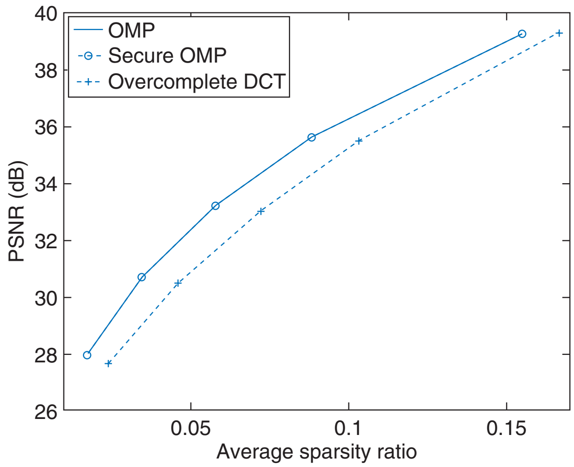


Fig. 9.   Rate-distortion performance.

(DCT) is plotted in **Fig. 9**. The average sparsity ratio is defined as the ratio of the number of sparse representations to the number of atoms in the dictionary. It can be seen that the proposed secure sparse coding (secure OMP) can represent the image with fewer atoms than overcomplete DCT. Furthermore, it is confirmed that the proposed secure sparse coding yields the same results as the unencrypted version of sparse coding (OMP) [9].

4. Application to face recognition

Face recognition has been a prominent biometric technique for identity authentication in a wide range of areas and applications, for example, public security and virtual reality. While the integration of face recognition and the edge/cloud network generates an extra degree of freedom for performance enhancement, significant concerns have been raised about privacy, as such biometric information could be misused without permission.

Our objective is to construct a secured framework to reduce the computation demands at each device, while taking advantage of this benefit to produce a more accurate face recognition result. To this end, we preserve privacy by deploying secure sparse coding, which enables dictionaries/recognition results to be trained/drawn from the encrypted images. We further prove both theoretically and through simulation that such encryption will not affect the accuracy of face recognition. To fully utilize the multi-device diversity, we extract deeper features in an intermediate space, which is expanded according to the dictionaries from each device, and perform classification in this new feature space to combat noise and modeling errors. This approach is demonstrated to achieve higher correctness of predictability through simulation results.

In the ensemble training process, as shown in **Fig. 10**, we jointly train a discriminative dictionary and classifier parameter based on the encrypted data at each edge server. Then we extract the decision templates for each class of individuals to be recognized at the remote cloud in order to efficiently combat noise and modeling errors.

In the recognition process, as shown in **Fig. 11**, we devise a pairwise similarity measurement, based on which we compare the current decision profile for a testing sample with each of the formulated decision templates. The closest match will produce the classification result.

We investigate the performance of the proposed framework by simulation. The performance improvement achieved by exploiting the multi-device diversity through ensemble learning is shown in **Fig. 12**. The performance improvement is significant when the number of devices is large due to the extra degree of freedom. In addition, we verified that by adopting random unitary transform, the result of face recognition is not affected, which proves that the proposed framework operates on a secured plane without any performance degradation.

We also show in Fig. 12 a performance comparison with a deep learning based algorithm, in which a 5-layer convolutional neural network was adopted and principal component analysis (PCA) was deployed to learn filter kernels in order to extract more discriminative features. Even though the deep learning based algorithm outperforms the proposed framework by 0.7% in terms of recognition accuracy, it requires 67% more dictionary training samples for each class. Significantly, when there are 30 dictionary training samples for each class, which is the most common setting when evaluating the performance on the dataset we used, the proposed framework dominates by over 4%. When there are only 10 training samples per class, which is reasonable in real-world settings due to the scarcity of fine-grained and manually labeled data, the proposed framework dominates by over 12%.

The performance comparison in terms of computational complexity is indicated in **Table 1**. Even though the deep learning based algorithm adopts: 1) PCA instead of a stochastic gradient descent to learn filter kernels and 2) a hashing method to simplify the

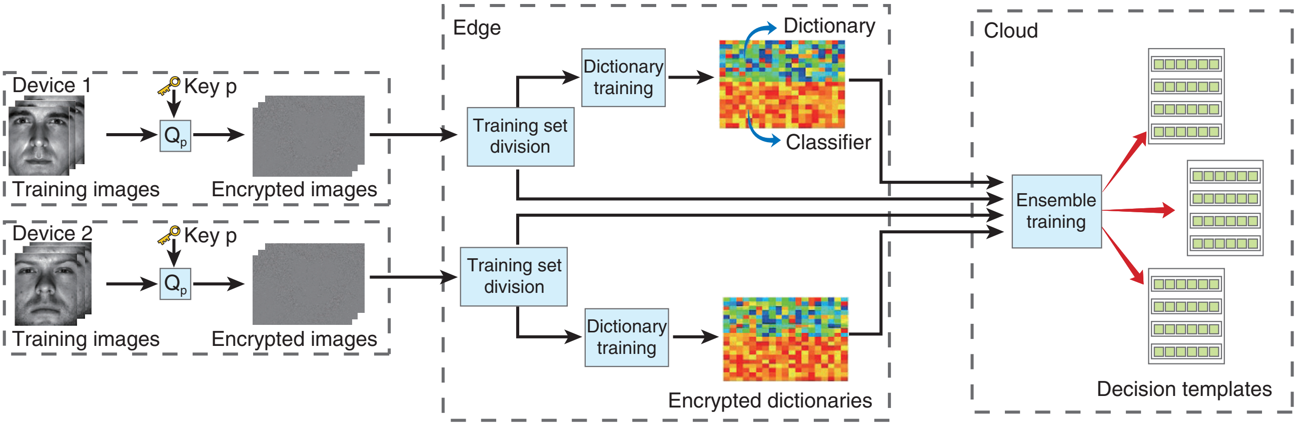


Fig. 10. Ensemble training.

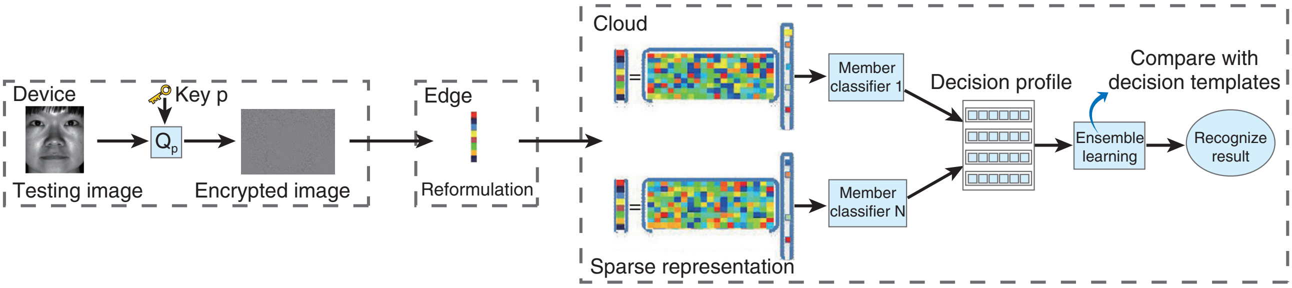


Fig. 11. Recognition process.

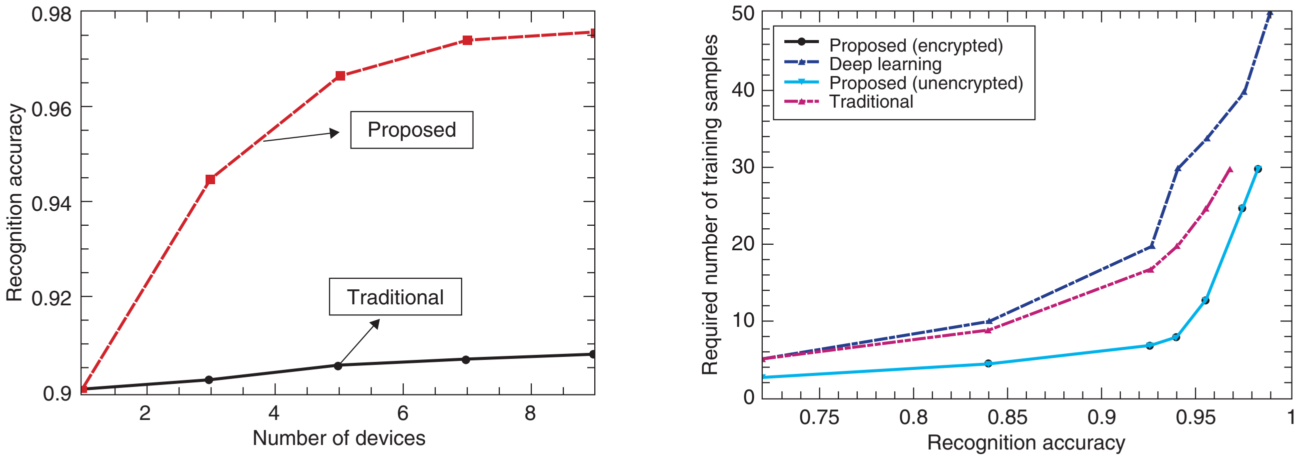


Fig. 12. Multi-device diversity and performance comparison.

nonlinear processing layer in order to reduce the computational complexity, it still requires a very long training time under such a small database. The proposed algorithm is extremely fast in terms of testing

Table 1. Comparison of running time.

	Training time (s)	Testing time (s)
Proposed	7.29	1.64×10^{-3}
Deep learning	5780	1.20
Traditional	4.84	1×10^{-4}

time, which makes it possible to support real-time face recognition applications.

5. Conclusion and future work

We have conducted basic research in the interdiscipline of sparse coding and networking from a security perspective. Specifically, we propose a secure sparse coding scheme with low complexity and demonstrate its application to image compression and face recognition for both preserving privacy and enhancing performance. We plan to further investigate the integration of sparse coding and networking in areas such as online traffic prediction and anomaly detection in order to extend our scientific contribution. Moreover, we are looking for opportunities for practical implementation to support secured real-time multimedia processing related applications, for example, secured face recognition in surveillance cameras, in order to demonstrate its commercial value as well as practical significance.

References

[1] Y. Goyal, A. Mohapatra, D. Parikh, and D. Batra, “Towards Transparent AI Systems: Interpreting Visual Question Answering Models,” arXiv:1608.08974, 2016.

[2] B. A. Olshausen and D. J. Field, “Emergence of Simple-cell Receptive Field Properties by Learning a Sparse Code for Natural Images,” Nature, Vol. 381, pp. 607–609, 1996.

[3] Y. Xu, Z. Li, J. Yang, and D. Zhang, “A Survey of Dictionary Learning Algorithms for Face Recognition,” IEEE Access, Vol. 5, pp. 8502–8514, 2017.

[4] S. Teerapittayanon, B. McDanel, and H. T. Kung, “Distributed Deep Neural Networks over the Cloud, the Edge and End Devices,” Proc. of the 37th IEEE International Conference on Distributed Computing Systems (ICDCS 2017), pp. 328–339, Atlanta, GA, USA, June 2017.

[5] Y. Mao, J. Zhang, S. Song, and K. Letaief, “Stochastic Joint Radio and Computational Resource Management for Multi-user Mobile-edge Computing Systems,” IEEE Trans. Wireless Commun., Vol. 16, No. 9, pp. 5994–6009, 2017.

[6] Y. Saito, I. Nakamura, S. Shiota, and H. Kiya, “An Efficient Random Unitary Matrix for Biometric Template Protection,” Proc. of Joint 8th International Conference on Soft Computing and Intelligent Systems (SCIS 2016) and 17th International Symposium on Advanced Intelligent Systems (ISIS 2016), pp. 366–370, Sapporo, Japan, Aug. 2016.

[7] T. Nakachi, Y. Bandoh, and H. Kiya, “Secure Dictionary Learning for Sparse Representation,” The 27th European Signal Processing Conference (EUSIPCO 2019), Coruña, Spain, Sept. 2019, accepted.

[8] J. Zepeda, C. Guillemot, and E. Kijak, “Image Compression Using Sparse Representations and the Iteration-tuned and Aligned Dictionary,” IEEE J. Sel. Topics Signal Process., Vol. 5, No. 5, pp. 1061–1073, 2011.

[9] T. Nakachi and H. Kiya, “Practical Secure OMP Computation and Its Application to Image Modeling,” Proc. of the 2018 International Conference on Information Hiding and Image Processing (IHIP 2018), pp. 25–29, Manchester, UK, Sept. 2018.

Yitu Wang

Research Engineer, NTT Network Innovation Laboratories.

He received a B.S. and Ph.D. from Zhejiang University, Hangzhou, China, in 2013 and 2018. From August to November 2014, he was a visiting scholar with the University of Paris-Sud, Orsay, France. He joined NTT Network Innovation Laboratories in 2019. His research interests are mainly focused on communication networks and statistical data processing.

Takayuki Nakachi

Senior Research Engineer, Supervisor, NTT Network Innovation Laboratories.

He received a Ph.D. in electrical engineering from Keio University, Tokyo, in 1997. Since joining NTT, he has been researching super-high-definition image/video coding and media transport technologies. In 2006–2007, he was a visiting scientist at Stanford University, USA. His current research interests include communication science, information theory, and signal processing. He received the 26th TELECOM System Technology award in 2010, the 6th Paper Award of the Journal of Signal Processing in 2012, and the Best Paper Award of the Institute of Electrical and Electronics Engineers (IEEE) International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) 2015. He is a member of IEEE and the Institute of Electronics, Information and Communication Engineers (IEICE).