https://doi.org/10.2352/ISSN.2470-1173.2020.10.IPAS-028

CNN-based Classification of Degraded Images

Kazuki Endo [a]***, Masayuki Tanaka*** [a,b]***, and Masatoshi Okutomi*** [a]
[a] ***Tokyo Institute of Technology, Tokyo, Japan***
[b] ***National Institute of Advanced Industrial Science and Technology, Tokyo, Japan***

Abstract

Classification of degraded images is very important in practice because images are usually degraded by compression, noise, blurring, etc. Nevertheless, most of the research in image classification only focuses on clean images without any degradation. Some papers have already proposed deep convolutional neural networks composed of an image restoration network and a classification network to classify degraded images. This paper proposes an alternative approach in which we use a degraded image and an additional degradation parameter for classification. The proposed classification network has two inputs which are the degraded image and the degradation parameter. The estimation network of degradation parameters is also incorporated if degradation parameters of degraded images are unknown. The experimental results showed that the proposed method outperforms a straightforward approach where the classification network is trained with degraded images only.

Introduction

Image classification based on a deep convolutional neural network (CNN) has been investigated in these years [1, 2, 3, 4, 5, 6]. Most of the research in image classification only focuses on clean images without any degradation. However, in practice, we need to consider image degradation, such as compression, noise, and blurring. Classification of degraded images has been studied in several papers [7, 8, 9, 10]. A classification network trained only with annotated clean images does not have enough ability to classify degraded images. A straightforward approach to overcome this problem is to train a network with degraded images, as shown in Fig. 1-(a), where the structure of the network is the same as before. This training procedure can be considered as a kind of data augmentation.

Most of the research in classifying degraded images has used an image restoration network as the pre-process of a classification network [7, 8]. The restoration network outputs a restored image from a degraded image. Then, the restored image is input into the classification network. On the other hand, several papers have indicated that privileged information or side information can improve the classification performance [11, 12, 13].

In this paper, we propose a method to classify degraded images involving degradation parameters. Degraded images usually have some degradation parameters against their clean images. Therefore, it is a natural extension for the classification network of degraded images to use the degradation parameter as an additional input, as shown in Fig. 1-(b). Although degradation parameters might be provided with degraded images in some cases, they are usually unknown in practice. Therefore, we incorporate the estimation of degradation parameters, as shown in Fig. 1-(c).

This paper mainly focuses on the JPEG distortion as a typical example of image degradation because the JPEG compression is the de-facto standard for image compression. The degradation parameter of the JPEG distortion is a quality factor, where the lower the quality factor is, the lower the image quality, but the smaller its file size. For confirming the effectiveness of using a degradation parameter as an additional input to the classification network, we first confirm the performance of the estimation network for degradation parameters. Then, we show classification performance is improved by the classification network which has a degraded image and an estimated degradation parameter as its inputs.

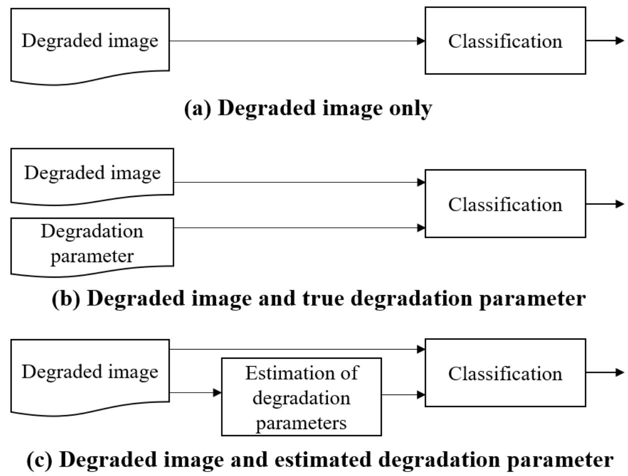


Figure 1. *Comparison for classification networks of degraded images. (a), (b), and (c) are categorized in terms of the input of the classification network as described in each caption.*

Our proposed method is applicable not only to the JPEG distortion but also other image degradation. We also show the effectiveness of the proposed method for additive Gaussian noise.

Related Works

CNN-based classification of images

Many CNN-based networks of image classification have been proposed [1, 2, 3, 4, 5, 6]. For classification of degraded images, CNN-based approaches have been studied in several papers [7, 8, 9, 10]. A common approach of some studies is to use images restored by a super-resolution network or a restoration network [7, 8]. The proposed method in this paper is an alternative approach to cope with degraded images and degradation parameters.

In this paper, the JPEG distortion is focused as an example of image degradation. There are some reports about the classifi-

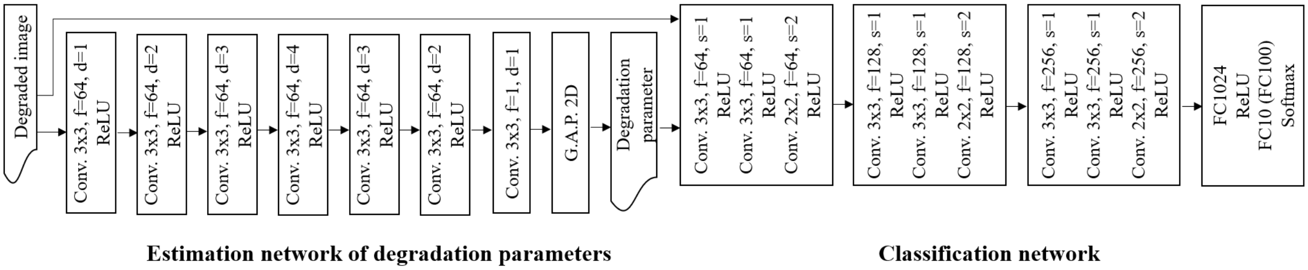


Figure 2. *Network architecture, where 3x3 or 2x2 represent the filter size, f is the dimension of feature map, d is the dilation rate, and s is the stride. G.A.P. denotes "Global Average Pooling".*

cation of the JPEG image. Das *et al.* have shown that the JPEG compression is effective in eliminating adversarial noise [14]. Gueguen *et al.* have proposed the network whose input is the coefficient of discrete cosine transformation (DCT) in YC_bC_r color space instead of the RGB intensity [15]. The key point of the proposed method is to use the JPEG quality factor as a degradation parameter.

Estimation of degradation parameters

Estimation of degradation parameters is to estimate unknown degradation parameters from a single degraded image. Estimation of degradation parameters and its application have been proposed in [16, 17]. Uchida *et al.* have focused on the JPEG distortion as an image degradation [16]. They have proposed a CNN to detect the JPEG compression and estimate JPEG quality factors. In this paper, we use the existing network to estimate degradation parameters [16].

Degradation parameters are additional information for the classification network in the proposed method. Learning with privileged information or side information has been investigated in machine learning and deep learning area [11, 12, 13, 18]. Vapnik *et al.* have used privileged information for learning support vector machines. They added poetic descriptions of MNIST as privileged information and improved the classification accuracy of MNIST recognition [11]. Hoffman *et al.* have used depth data as side information and constructed a hallucination network [13]. In this paper, we show that degradation parameters contribute to improving the classification performance of degraded images.

Proposed Method

Proposed network

Here, the CNN-based classification of degraded images is proposed. The proposed network consists of two networks as shown in Fig. 2; an estimation network of degradation parameters and a classification network. The input of the proposed network is a degraded image only. The key point is that the classification network in the proposed network has two inputs. One is a degraded image, and another is a degradation parameter. The degradation parameter is estimated by the estimation network. When degradation parameters are known, the estimation network in the proposed network is omitted, as shown in Fig.1-(b). And then, the known parameters are directly used for the input of the classification network.

The estimation of degradation parameters has almost the same network architecture proposed in [16, 19], where a batch normalization [20] is omitted for simplicity. Degradation parameters can be estimated for each pixel of an input image. Then, the pixel-wise degradation parameters are averaged spatially before inputting into the classification network. The loss function for the training of the estimation network is the mean square error (MSE) between correct degradation parameters and estimated parameters. The optimizer was Adamax [21] in experiments.

The architecture of the classification network is a VGG-like network [1], where the architecture includes the spatial dropout [3] and the convolution pooling [2] instead of the max pooling. The loss function is cross-entropy between correct labels and predicted labels. Adamax [21] was also used to minimize the loss function in experiments.

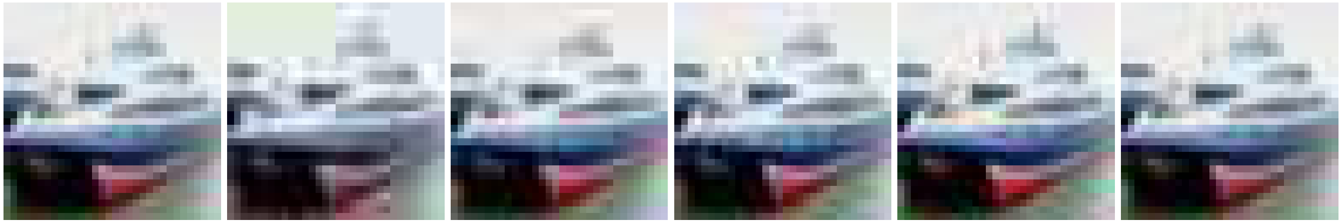
This paper uses a CNN-based network for the estimation of degradation parameters. However, the estimation method for degradation parameters is not limited to the CNN-based network.

Training procedure

The estimation network of degradation parameters and the classification network are separately trained. Especially, the estimation network of degradation parameters needs to be trained before training the classification network. The training of the classification network requires annotated images, while the estimation network of degradation parameters can be trained without annotation. The training procedure of the networks is explained for the estimation network of degradation parameters and then, for the classification network.

To train the estimation network of degradation parameters, we collect a huge number of clean images without annotation. Then, we apply the degradation to the clean images with some degradation parameters. Degradation parameters can be easily controlled by ourselves. Now, the estimation network of degradation parameters can be trained with pairs composed of degraded images and degradation parameters.

Both degraded images with annotation and estimated degradation parameters are necessary for the training of the classification network. Degraded images with annotation can be generated from clean images with annotation by applying the degradation. On the other hand, degradation parameters can be estimated by the pre-trained network to estimate degradation parameters. Now, the classification network can be trained with degraded images and estimated degradation parameters. The pre-trained network to estimate degradation parameters is not trained further during the training of the classification network. In the case of known degradation parameters, the classification network can be just trained



Original Q.F.=20 Q.F.=40 Q.F.=60 Q.F.=80 Q.F.=100

Figure 3. *JPEG compression effect for the CIFAR-10 image, where JPEG compression was applied to the original image with quality factors shown below each image.*

Table 1. Combinations of networks and training data in the case of the JPEG distortion.

Networks	Image only		Image and degradation parameter
Images	Original	JPEG	JPEG
Degradation parameters	-	-	Estimated JPEG quality factor
Notations	Org	Jpg	Jpg and Est. Q.F.

with annotated images and known degradation parameters.

Experiments

Experiments were focused on the JPEG image and the JPEG quality factor as an example of degraded images and degradation parameters. Figure 3 shows the CIFAR-10 images compressed with different JPEG quality factors. The JPEG quality factor controls the quality of JPEG images. Therefore, adding the JPEG quality factor to the input of a classification network is expected to improve classification performance. To the best of our knowledge, this is the first time to introduce JPEG quality factors for the classification of JPEG images. We also confirm the effectiveness of the proposed method in the case of Gaussian noisy images and their noise levels.

Datasets and data augmentation

To train the estimation network of JPEG quality factors, we used datasets of Yang91 [22], Urban100 [23], and General100 [24]. 64×64 sized patches were extracted from each image. Then, data augmentation using transpose, horizontal, and vertical flips was applied to the patches. Finally, we compressed the patches by the JPEG compression, and then generated pairs of JPEG distorted images and JPEG quality factors. The quality factors were randomly sampled from 1 to 100[1].

The CIFAR datasets [25] were used for the training of the classification network. Data augmentation was applied for the CIFAR images; zoom, shearing, horizontal flip, rotation, vertical, and horizontal shifts. Then, we applied the JPEG compression to each image, where the JPEG quality factor was randomly sampled from 1 to 100. We called the original CIFAR images without any degradation as "original CIFAR" for simplicity. In the same way, we called the JPEG compressed CIFAR images as "JPEG CIFAR." The reproduction code is available online[2].

[1]Note that the detailed JPEG compression algorithm depends on the library. We used a Python Image Library(PIL) for the JPEG compression.

[2]http://www.ok.sc.e.titech.ac.jp/res/CNNIR/IRDI/

Interval mean accuracy

The classification performance of degraded images strongly depends on degradation parameters. Therefore, we introduce an interval mean accuracy to evaluate the classification performance of images degraded with different degradation parameters.

The interval mean accuracy is defined by:

$$\overline{Acc}\left(Q_l, Q_u\right) \overset{def}{=} \frac{\sum_{i=Q_l}^{Q_u} Acc\left(\mathrm{D}\left(\mathbf{X}, q=i\right), \mathbf{Y}\right)}{Q_u - Q_l + 1}, \tag{1}$$

where Q_l and Q_u denote degradation parameters $\left(Q_l < Q_u\right)$, $\mathbf{X}$ denotes clean images without any degradation, $\mathrm{D}\left(\cdot, q\right)$ is a degradation operator with a degradation parameter q, $\mathbf{Y}$ denotes true labels for $\mathbf{X}$, and Acc represents an accuracy. The accuracy means a ratio dividing the number of predicted class labels, which coincide with correct class labels, by the number of all test samples.

Classification network and training data

We compare two types of classification network which have different inputs as summarized in Table 1. One is a classification network which has an image only for the input. There are two different training procedures for training data; original clean images and JPEG images. We denote these two procedures "Org" and "Jpg," respectively.

The other type of classification network, of which inputs are both a degraded image and a degradation parameter, is trained with JPEG images and estimated JPEG quality factors. We denote the network "Jpg and Est. Q.F."

Estimation performance of JPEG quality factors

First, we confirm the estimation performance of JPEG quality factors. For the validation, test images of the CIFAR-10 were used. We compressed the test images with different quality factors. Then, quality factors were estimated from the compressed test images by the estimation network of JPEG quality factors. Figure 4 and Table 2 show the mean and the standard deviation of estimated JPEG quality factors. The results show that the performance of the estimation network for JPEG quality factors seems to be almost the same level as that of Uchida *et al.* [16]. Thus, the

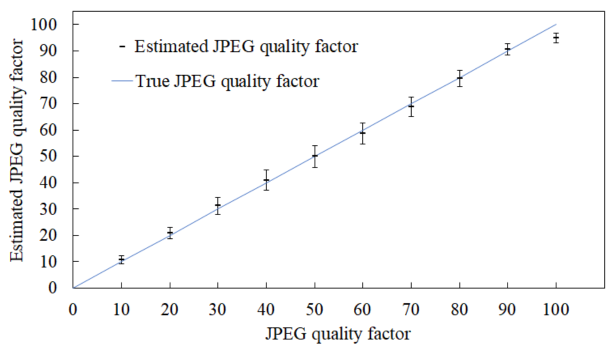


Figure 4. *Estimation performance of JPEG quality factors. For estimated JPEG quality factors, the mean and the standard deviation are plotted.*

Table 2. Estimation performance of JPEG quality factors.

Truth	20	40	60	80	100
Mean	20.84	40.93	58.58	79.53	94.91
S.D.	2.28	3.91	4.07	3.07	1.91

Table 3. Comparison for the interval mean accuracy of "JPEG CIFAR-10."

Networks	Org	Jpg	Jpg and Est. Q.F.
$\overline{Acc}(1,20)$	0.431	0.724	*0.744*
$\overline{Acc}(21,40)$	0.700	0.844	*0.849*
$\overline{Acc}(41,60)$	0.763	0.857	*0.864*
$\overline{Acc}(61,80)$	0.799	0.866	*0.872*
$\overline{Acc}(81,100)$	0.861	0.874	*0.880*

estimation performance of JPEG quality factors seems to be good enough for the experiments.

Estimated JPEG quality factor for the input of classification network

We confirm the classification performance by using the estimated JPEG quality factor for the input of the classification network. Table 3 summarizes the interval mean accuracy of "JPEG CIFAR-10." Figure 5 shows the accuracy plotted against the quality factor. The network trained with "original CIFAR-10" exhibits lower performance, *i.e.* the quality factor is below about 95, because the network is not trained with JPEG images. "Jpg and Est. Q.F.," which is the proposed network, exhibits better performance than "Jpg" for almost all quality factors. The results show that the classification performance is improved by using the estimated quality factor for the input of the classification network.

JPEG compression of CIFAR-100 dataset

We also evaluate the proposed network for the CIFAR-100 dataset in the case of the JPEG distortion. Table 4 shows the interval mean accuracy of the proposed network and the other networks to compare. Figure 6 shows the accuracy of them. "Jpg and Est. Q.F." exhibits almost better performance than "Jpg." The tendency of the CIFAR-100 dataset case is almost the same as that of the CIFAR-10 dataset case.

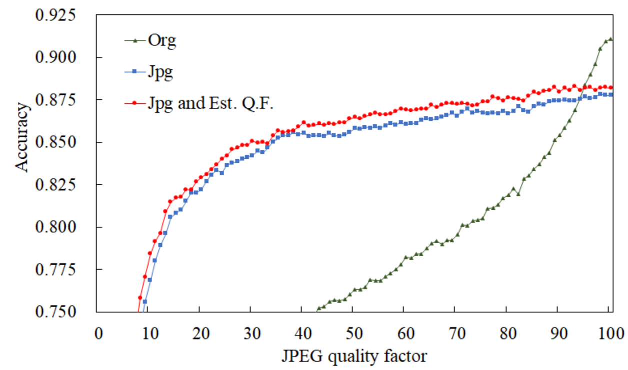


Figure 5. *Comparison for the accuracy of "JPEG CIFAR-10."*

Table 4. Comparison for the interval mean accuracy of "JPEG CIFAR-100."

Networks	Org	Jpg	Jpg and Est. Q.F.
$\overline{Acc}(1,20)$	0.202	0.448	*0.454*
$\overline{Acc}(21,40)$	0.407	0.561	*0.563*
$\overline{Acc}(41,60)$	0.465	0.577	*0.581*
$\overline{Acc}(61,80)$	0.504	0.583	*0.588*
$\overline{Acc}(81,100)$	0.582	0.591	*0.596*

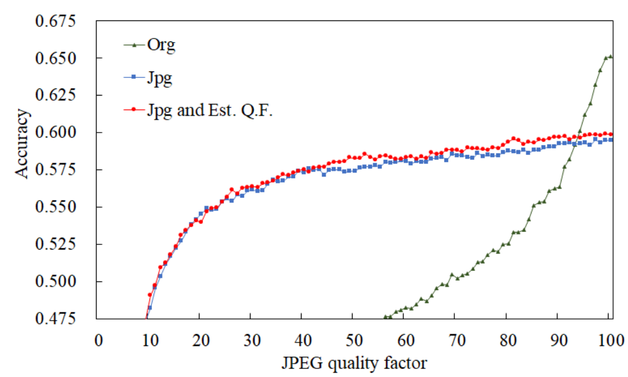


Figure 6. *Comparison for the accuracy of "JPEG CIFAR-100."*

Application to additive Gaussian noise

We have mainly discussed the JPEG distortion as an example of image degradation. However, the proposed method is not limited to the JPEG distortion. Here, we apply the proposed method to additive Gaussian noise.

In the case of additive Gaussian noise, we used the same network architecture as in the case of the JPEG distortion. Noisy images for the input were synthesized by adding Gaussian noise, where the noise level was changing from 0 to 50 for the 8-bit image. The proposed network for additive Gaussian noise was trained in the same manner as for that of the JPEG distortion.

Figure 7 and Table 5 show the performance of the estimation network for each Gaussian noise level. The performance of it seems to be good enough.

We also evaluate three different networks with the CIFAR-10 dataset. Table 6 is the summary of interval mean accuracy. Figure 8 shows the accuracy plotted against the noise level. The classification performance is also improved by using the estimated noise level for the input of the classification network.

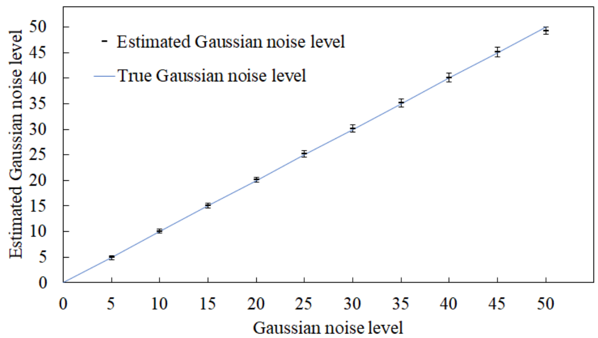


Figure 7. Estimation performance of Gaussian noise level. For the estimated Gaussian noise level, the mean and the standard deviation are plotted.

Table 5. Estimation performance of Gaussian noise level.

Truth	10	20	30	40	50
Mean	10.03	20.11	30.16	40.11	49.26
S.D.	0.42	0.54	0.69	0.89	0.70

Table 6. Comparison for the interval mean accuracy of Gaussian noisy CIFAR-10.

Networks	Org	Noisy	Noisy and Est. N.L.
$\overline{Acc}(0,10)$	0.796	0.851	*0.882*
$\overline{Acc}(11,20)$	0.328	0.844	*0.872*
$\overline{Acc}(21,30)$	0.138	0.829	*0.856*
$\overline{Acc}(31,40)$	0.106	0.809	*0.834*
$\overline{Acc}(41,50)$	0.101	0.785	*0.810*

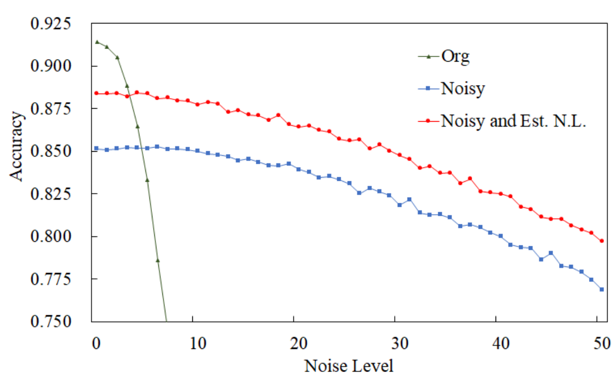


Figure 8. Comparison for the accuracy of Gaussian noisy CIFAR-10.

Therefore, we experimentally confirmed that the proposed method is effective for not only the JPEG distortion but also additive Gaussian noise.

Conclusion

This paper has proposed the CNN-based classification of degraded images. The key point is to input additional information about image degradation into the classification network of the proposed network. First, we have confirmed the performance of the estimation network of degradation parameters. Then, we have experimentally shown the additional use of the estimated degradation parameter can improve the classification performance. Experimental comparisons confirmed that the proposed method could improve classification performance for the JPEG distortion and additive Gaussian noise. For only high-quality images, the results also found that the network trained with only original clean images showed better accuracy than the proposed network. Further investigation is needed for this point. As one of the practical extensions, a mixture of some distortions can also be considered in the proposed method. This extension is our future work.

References

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.

[2] J.T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving simplicity: the all convolutional net," in *International Conference on Learning Representations Workshop Track*, 2015.

[3] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 648–656.

[4] M. Tanaka, "Weighted sigmoid gate unit for an activation function of deep neural network," *arXiv:1810.01829v1*, 2018.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European Conference on Computer Vision*, 2016, pp. 630–645.

[7] D. Cai, K. Chen, Y. Qian, and J.K. Kämäräinen, "Convolutional low-resolution fine-grained classification," *Parttern Recognition Letters(in press)*, 2017.

[8] A. Vest, M. Jamal, and B. Gong, "Very low resolution image classification," *http://crcv.ucf.edu/REU/2017/Vest/Final_paper.pdf*, 2017.

[9] X. Peng, J. Hoffman, S.X. Yu, and K. Saenko, "Fine-to-coarse knowledge transfer for low-res image classification," in *IEEE International Conference on Image Processing*, 2016.

[10] Y. Pei, Y. Huang, Q. Zou, H. Zang, X. Zhang, and S. Wang, "Effects of image degradations to cnn-based image classification," *arXiv:1810.05552*, 2018.

[11] V. Vapnik and A. Vashist, "A new learning paradigm: learning using privileged information," *Neural Networks*, vol. 22, pp. 544–557, 2009.

[12] R. Jonschkowski, S. Hfer, and O. Brock, "Patterns for learning with side information," *arXiv:1511.06429*, 2016.

[13] J. Hoffman, S. Gupta, and T. Darrell, "Learning with side information through modality hallucination," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[14] N. Das, M. Shanbhogue, S.T. Chen, F. Hohman, L. Chen, M.E. Kounavis, and D.H. Chau, "Keeping the bad guys out: protecting and vaccinating deep learning with jpeg compression," *arXiv:1705.02900v1*, 2017.

[15] L. Gueguen, A. Sergeev, B. Kadlec, R. Liu, and J. Yosinski, "Faster neural networks straight from jpeg," in *Neural Information Processing Systems*, 2018.

[16] K. Uchida, M. Tanaka, and M. Okutomi, "Pixelwise jpeg compression detection and quality factor estimation based on convolutional neural network," in *Proceedings of IS&T International Symposium*

on Electronic Imaging, 2019.

[17] X. Liu, M. Tanaka, and M. Okutomi, "Single-image noise level estimation for blind denoising," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5226–5237, December 2013.

[18] D. P. Kingma, D. J. Rezende, S. Mohamed, and M. Welling, "Semi-supervised learning with deep generative models," in *Neural Information Processing Systems*, 2014.

[19] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep cnn denoiser prior for image restoration," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[20] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015.

[21] D.P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *International Conference on Learning Representations*, 2015.

[22] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *IEEE Conference on Computer Vision*, 2008.

[23] J.B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *IEEE Conference on Computer Vision and Pattern Recongnition*, 2015, pp. 5197–5206.

[24] C. Dong, C.C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European Conference on Computer Vision*, 2016.

[25] A. Krizhevsky, "Learning multiple layers of features from tiny images," M.S. thesis, Department of Computer Science, University of Tronto, 2009.

Author Biography

Kazuki Endo *received a bachelors degree in mathematics and a masters degree in industrial engineering and management from Tokyo Institute of Technology, Tokyo, Japan, in 1997 and 1999, respectively. He joined Industrial Bank of Japan, Ltd., which is current Mizuho Bank, Ltd., Tokyo, Japan, in 1999. He was in charge of the market business, especially for developing derivatives pricing algorithms, systems, and trading interest-rate derivatives. Since 2014, he has been in charge of developing web services at Topology Inc., Tokyo, Japan. Since 2019, he has belonged to the Ph.D. course at Tokyo Institute of Technology, Tokyo, Japan.*

Masayuki Tanaka *received his bachelors and masters 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 from 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.*

Masatoshi Okutomi *received a B.Eng. degree from the Department of Mathematical Engineering and Information Physics, the University of Tokyo, Japan, in 1981 and an M.Eng. degree from the Department of Control Engineering, Tokyo Institute of Technology, Japan, in 1983. He joined Canon Research Center, Canon Inc., Tokyo, Japan, in 1983. From 1987 to 1990, he was a visiting research scientist in the School of Computer Science at Carnegie Mellon University, USA. In 1993, he received a D.Eng. degree for his research on stereo vision from Tokyo Institute of Technology. Since 1994, he has been with Tokyo Institute of Technology, where he is currently a professor in the Department of Systems and Control Engineering, the School of Engineering.*