Received: 24 August 2018 | Revised: 19 December 2018 | Accepted: 13 February 2019

DOI: 10.4218/etrij.2018-0473

ORIGINAL ARTICLE

WILEY ETRI Journal

Constrained adversarial loss for generative adversarial network-based faithful image restoration

Dong-Wook Kim[1] | **Jae-Ryun Chung**[1] | **Jongho Kim**[2] | **Dae Yeol Lee**[2] | **Se Yoon Jeong**[2] | **Seung-Won Jung**[1]

[1]Department of Multimedia Engineering, Dongguk University, Seoul, Rep. of Korea

[2]Broadcasting and Media Research Laboratory, Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea

Correspondence
Seung-Won Jung, Department of Multimedia Engineering, Dongguk University, Seoul, Rep. of Korea.
Email: swjung83@gmail.com

Funding information
This research was supported by the Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00072, Development of Audio/Video Coding and Light-Field Media Fundamental Technologies for Ultra-Realistic Tera-media). This work was supported by the Dongguk University Research Fund of 2014.

Generative adversarial networks (GAN) have been successfully used in many image restoration tasks, including image denoising, super-resolution, and compression artifact reduction. By fully exploiting its characteristics, state-of-the-art image restoration techniques can be used to generate images with photorealistic details. However, there are many applications that require faithful rather than visually appealing image reconstruction, such as medical imaging, surveillance, and video coding. We found that previous GAN-training methods that used a loss function in the form of a weighted sum of fidelity and adversarial loss fails to reduce fidelity loss. This results in non-negligible degradation of the objective image quality, including peak signal-to-noise ratio. Our approach is to alternate between fidelity and adversarial loss in a way that the minimization of adversarial loss does not deteriorate the fidelity. Experimental results on compression-artifact reduction and super-resolution tasks show that the proposed method can perform faithful and photorealistic image restoration.

KEYWORDS
compression artifact reduction, deep learning, generative adversarial network, image restoration

1 | INTRODUCTION

Almost all state-of-the-art image restoration techniques are based on deep convolutional neural networks (CNN), including denoising [1], super-resolution [2], deblurring [3], and compression-artifact reduction [4]. With image restoration using CNNs, the training data can be readily obtained by applying degradations to the original distortion-free images. The restoration network can then be trained to generate images close to the originals, given degraded images. Research progress has occurred mainly in the adoption of recent network architectures developed for other computer-vision tasks, such as object classification and detection. Image-specific loss functions [5] and problem-specific network modifications [1–3] have also been attempted for better image restoration performance.

The revolutionary idea of the generative adversarial network (GAN) [6] has shown extraordinary capability for generating images from random noise signals. The generative network of the GAN is used to generate images resembling real images, whereas the discriminative network of the GAN is used to distinguish between the generated and real images. In particular, deep convolutional GAN (DCGAN) [7] and its variants have been successful with many image generation tasks. When GAN is applied to

image restoration, the generative network must generate not only realistic images but also faithful ones close to the original distortion-free versions. To satisfy these two difficult constraints, previous loss functions of the generative network typically include fidelity loss (i.e., the measure of the difference between the generated and original images) and adversarial loss (i.e., the measure of how well the generative network fools the discriminative network).

We empirically found that the above two loss terms could not be effectively minimized together when a loss function was defined as a weighted sum of the two loss terms. This was confirmed by [4], where the compression artifact-reduced images had even lower peak signal-to-noise ratios (PSNR) than naive Joint Photographic Experts Group (JPEG) images, or by [2], where super-resolved images sometimes showed lower PSNRs than bicubic-upsampled images. Nevertheless, objective quality metrics, such as PSNR, do not highly correlate with human perception of image quality. Thus, GAN-based methods [2,4] can be more desirable for favorable image reconstruction. However, there remain many applications that require faithful rather than visually appealing image reconstruction, such as medical imaging and surveillance. Furthermore, in-loop filtering, adopted in most video coding standards, should generate images with quality close to the original ones for better compression of subsequent video frames [8].

In this paper, we propose to alternate between the two loss functions (i.e., fidelity and adversarial loss functions) during generative network training. Instead of simple alternation, we devise a criterion for the network parameter update such that the minimization of the adversarial loss function does not exert a negative influence on the minimization of the fidelity loss. Experimental results on compression artifact reduction verify that the proposed training scheme effectively reduces both loss functions, thus enabling faithful and photorealistic image restoration.

2 | RELATED WORK

In this section, we review the representative GAN-based image restoration methods: super-resolution GAN (SRGAN) [2] and artifact reduction GAN (ARGAN) [4]. Both methods share a similar network architecture comprising convolutional layers with small-sized kernels, batch normalization layers, and skip connections. Here, we focus only on the loss function of the generative network.

Let X, Y, and $\hat{X}$ denote the original distortion-free image, degraded image, and restored image, respectively. The restored image is generated by the generative network, G, which takes the degraded image as an input (i.e., $\hat{X} = G(Y)$). The loss function of the generative network of the SRGAN is defined as follows:

$$l_G^{\mathrm{SR}} = l_{\mathrm{fid}}(X,\hat{X}) + \lambda \times l_{\mathrm{adv}}(\hat{X}), \tag{1}$$

where l_{fid} and l_{adv} denote the fidelity loss and the adversarial loss, respectively. In SRGAN, two different options are provided for the fidelity loss: mean-square error (MSE) and the visual geometry group (VGG) loss [9]. The MSE, the most widely used difference measure in image restoration, is defined as follows:

$$l_{\mathrm{fid}}^{\mathrm{MSE}}(X,\hat{X}) = \frac{1}{|\Psi|}\sum_{(i,j)\in\Psi}\left(X(i,j)-\hat{X}(i,j)\right)^2, \tag{2}$$

where $X(i, j)$ represents a pixel value at position (i, j), and Ψ is a set of pixel coordinates. $|\Psi|$ is the cardinality of Ψ, which is equivalent to the number of pixels in X. However, the MSE correlates poorly with human perception of image quality and often results in images that lack high-frequency details. Motivated by [10], the VGG loss evaluates the feature similarity rather than image similarity. Using the pretrained VGG network, the VGG loss is measured as follows:

$$l_{\mathrm{fid}}^{\mathrm{VGG}}(X,\hat{X}) = \frac{1}{|\Upsilon|}\sum_{(i,j,k)\in\Upsilon}\left(\phi_X(i,j,k)-\phi_{\hat{X}}(i,j,k)\right)^2, \tag{3}$$

where ϕ_X and $\phi_{\hat{X}}$ are the VGG feature maps of X and $\hat{X}$, respectively. (i, j, k) is an element index of the feature map, and Υ is a set of indices. In particular, the 19-layer VGGNet is used for feature extraction [9].

Following the original GAN [6], l_{adv} is defined as follows:

$$l_{\mathrm{adv}}(\hat{X}) = -\log\left(D(\hat{X})\right), \tag{4}$$

where D indicates the discriminative network. Thus, the adversarial loss decreases as the generative network successfully fools the discriminative network. To train the discriminative network, the following loss function is used.

$$l_D(X,\hat{X}) = -\left[\log\left(D(X)\right)+\log\left(1-D(\hat{X})\right)\right]. \tag{5}$$

The discriminative network is thus trained to distinguish between the generated and real images.

Table 1 of [2] shows that the highest mean opinion score was obtained when $l_{\mathrm{fid}} = l_{\mathrm{fid}}^{\mathrm{VGG}}$ and $\lambda = 10^{-3}$ (defined as SRGAN-VGG). However, the objective image quality of SRGAN-VGG was significantly lower than the case when $l_{\mathrm{fid}} = l_{\mathrm{fid}}^{\mathrm{MSE}}$ and $\lambda = 0$ (defined as SRResNet-MSE). For example, at an upsampling ratio of 4, the PSNR of SRGAN-VGG was about 2-dB lower than that of SRResNet-MSE for the Set14 dataset [11]. Whereas SRGAN-VGG could effectively generate images with realistic details, it tended to fail in faithful reconstruction of image details. In terms of PSNR or structural similarity (SSIM) [12], SRGAN was comparable only to simple bicubic interpolation.

From the viewpoint of the loss function, ARGAN is very similar to SRGAN. The application-specific JPEG loss term, $l_{\mathrm{fid}}^{\mathrm{JPEG}}$ [4], is included, and $l_{\mathrm{fid}}^{\mathrm{VGG}}$ is combined with $l_{\mathrm{fid}}^{\mathrm{JPEG}}$ using another weighting parameter. The loss function of the

discriminative network is the same as (4). Likewise, using the result of SRGAN, ARGAN reconstructs images with abundant textures but fails in faithful image reconstruction. Note that the compression artifact-reduced images obtained by ARGAN have even lower PSNRs than naive JPEG images [4].

In summary, both SRGAN and ARGAN show impressive image restoration performance in terms of subjective image quality. However, very low performance in terms of the PSNR or SSIM can restrict their applicability. We conjecture that the fidelity loss plays only a subsidiary role to the adversarial loss when the loss function is defined. See (1). In this paper, we attempt to find a better way to jointly optimize the two loss terms.

3 | PROPOSED METHOD

We attempt to change the training procedure of the GAN's generative network for faithful image restoration. As a case study, we apply the proposed training scheme to a state-of-the-art compression artifact reduction method (i.e., ARGAN [4]) and a super-resolution method (i.e., SRGAN [2]). We then explain the proposed method using ARGAN, because the same methodology can be readily applied to SRGAN. To this end, we describe the difficulty of optimizing the loss function of ARGAN. Figure 1 shows the results of ARGAN with the following the generative network loss function and different λ values.

$$l_G = l_{\text{fid}}(X, \hat{X}) + \lambda \times l_{\text{adv}}(\hat{X}). \tag{6}$$

In this example, $l_{\text{fid}}^{\text{MSE}}$, in (2), is applied for fidelity loss. This term is the most widely used fidelity loss in image restoration. $l_{\text{fid}}^{\text{JPEG}}$ is excluded, because it is only applicable in JPEG-compressed images. As expected, the highest PSNR is obtained when $\lambda = 0$ (i.e., when the GAN is disabled). However, the reconstructed image, as shown in Figure 1C, lacks fine details and rich textures, compared to the original image shown in Figure 1A. As λ increases, the PSNR decreases, but rich textures increase. Thus, it seems

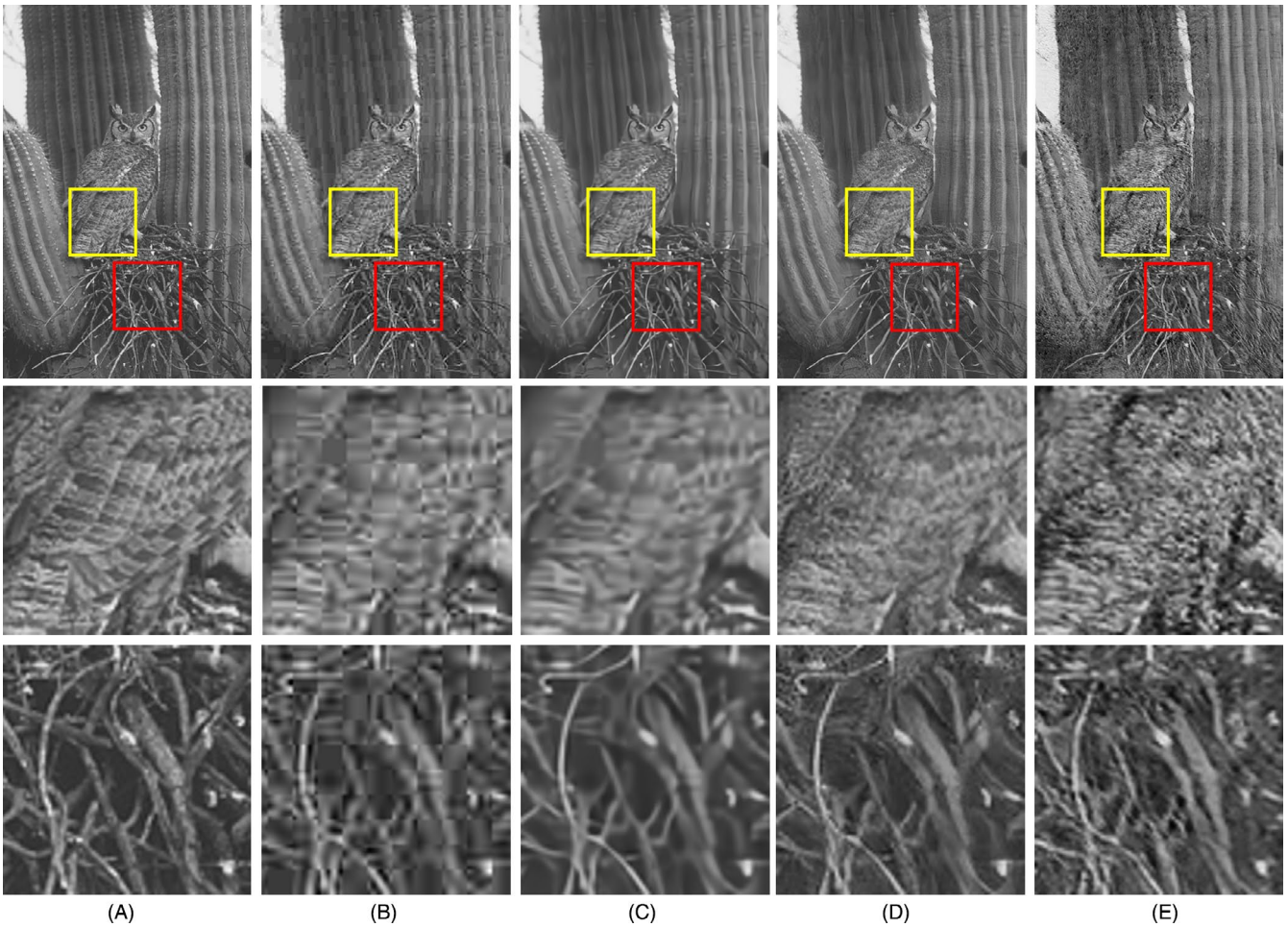

FIGURE 1 Comparison under a quality factor of 10 on BSDS500: (A) Original; (B) JPEG (PSNR: 25.83 dB); and (C)–(E) reconstruction result using (6) with $\lambda = 0$ (PSNR: 27.05 dB), $\lambda = 10^{-3}$ (PSNR: 25.47 dB), and $\lambda = 10^{-1}$ (PSNR: 20.91 dB), respectively. The second and third rows are the image patches corresponding to the yellow and red rectangles in the top row

inevitable to renounce the objective image quality such as the PSNR in favor of acquiring images with rich textures.

Considering the aforementioned difficulty, we propose to use two separate loss functions as follows:

$$l_G^1(\theta) \triangleq l_{\text{fid}}(X, \hat{X}) = l_{\text{fid}}(X, G_\theta(Y)), \tag{7}$$

$$l_G^2(\theta) \triangleq l_{\text{adv}}(\hat{X}) = l_{\text{adv}}(G_\theta(Y)). \tag{8}$$

Note that the first loss function, l_G^1, reflects only the fidelity loss, and the second loss function, l_G^2, reflects only the adversarial loss. Let G_θ represent the generative network parameterized by θ. The set of parameters, θ, can then be updated as follows:

$$\theta_1 = f(l_G^1(\theta)), \tag{9}$$

$$\theta_2 = f(l_G^2(\theta)), \tag{10}$$

where θ_1 and θ_2 denote the sets of parameters being updated to decrease the first and second loss functions, respectively. A general function, f, refines the parameters according to the gradient descent principle. We chose f as Adam, with a learning rate of 10^{-4} and a momentum of 0.5, for our experiment.

An interesting observation was made in [5], where the switch between two loss functions was found to be effective in minimizing either of the loss functions. In particular, it was shown that the MSE could be further decreased when the network that converges by the MSE-loss is retrained using the mean absolute difference as a new loss function. This is understandable, because a new criterion can help the network move out of a local minimum. Motivated by [5], we also attempt to alternate between fidelity and adversarial loss. The alternation can be applied at the early stage of training or after the network is converged by one loss function. However, we empirically found that the switch to another loss function after the network has converged by one loss function [5] is not helpful. We conjecture that this is because the fidelity and adversarial loss functions pursue largely deviated solutions. Thus, a new criterion could not help the network move out of the local minimum.

Algorithm 1 describes our training procedure of ARGAN. Unconditional alternation between the fidelity and adversarial loss functions during the training of the generative network is not necessarily different from the optimization using (6). Therefore, putting a bigger emphasis on the fidelity loss for faithful image restoration, we define a new criterion for the parameter update as follows:

$$\theta' = \begin{cases} \theta_2, & \text{if } l_G^1(\theta_2) < l_G^1(\theta_1) \\ \theta_1, & \text{otherwise} \end{cases}, \tag{11}$$

where θ' is the parameter set obtained after the update. θ_2 is a temporary parameter set acquired to reduce the adversarial loss, using (10). Thus, we accept θ_2 if the update to reduce the adversarial loss *accidentally* reduces the fidelity loss as well. This constrained parameter update scheme, denoted as cARGAN, can explicitly prevent the adversarial loss from exerting a negative influence to the fidelity loss. The above training scheme can also be applied to SRGAN in a straightforward manner. Thus, we denote the SRGAN with our proposed training scheme, cSRGAN.

Algorithm 1 Proposed minibatch training of ARGAN.

```
for number of training iterations do
    for k steps do
        Update the discriminative network by (5)
    end for
    Update the generative network by (9)
    Temporarily update the generative network by (10)
    Accept or reject the above update according to (11)
end for
```

4 | EXPERIMENTS

In this section, we evaluate the effectiveness of the proposed training scheme for compression artifact reduction and super-resolution tasks, primarily using ARGAN [4] and SRGAN [2]. Note that ARGAN and SRGAN were chosen only for the verification of the proposed constrained parameter update scheme. Our training scheme can be applied to other GAN-based image restoration methods [14–16].

4.1 | Compression artifact reduction

4.1.1 | Dataset and parameter settings

In our experiments, Microsoft COCO 2017 training images [17] were used for training, and the BSDS500 dataset [18] was used for validation. An image patch of the size 48×48

TABLE 1 Comparisons to the state-of-the-art methods on the BSDS500 dataset

Quality	Method	PSNR (dB)	SSIM
10	JPEG	27.80	0.7875
	TNRD	29.16	0.8225
	DDCN	29.59	0.8381
	$\text{ARGAN}_{\text{base}}$	29.43	0.8309
	$\text{ARGAN}_{\text{VGG}}$	27.06	0.7440
20	JPEG	30.05	0.8671
	TNRD	31.41	0.8889
	DDCN	31.88	0.8996
	$\text{ARGAN}_{\text{base}}$	31.73	0.8953
	$\text{ARGAN}_{\text{VGG}}$	29.23	0.8420

TABLE 2 Comparisons of ARGAN training methods

		$ARGAN_{base}$		$ARGAN_{VGG}$		cARGAN	
Quality	**Fidelity loss**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
10	MSE	29.43	0.8309	N/A	N/A	29.10	0.8143
	VGG	N/A	N/A	27.06	0.7440	27.36	0.7786
20	MSE	31.73	0.8953	N/A	N/A	31.15	0.8753
	VGG	N/A	N/A	29.23	0.8420	29.50	0.8548

FIGURE 2 Comparison of different ARGAN training schemes under quality factor 10 on BSDS500: (A) JPEG (PSNR: 26.09 dB (top), 25.09 dB (bottom)); (B) $ARGAN_{base}$ (PSNR: 27.48 dB, 26.18 dB); (C) $ARGAN_{VGG}$ (PSNR: 25.32 dB, 23.67 dB); and (D) cARGAN (PSNR: 27.21 dB, 25.86 dB). For each test image, the second and third rows are the image patches corresponding to the yellow and red rectangles in the top row. These are best viewed with zoom

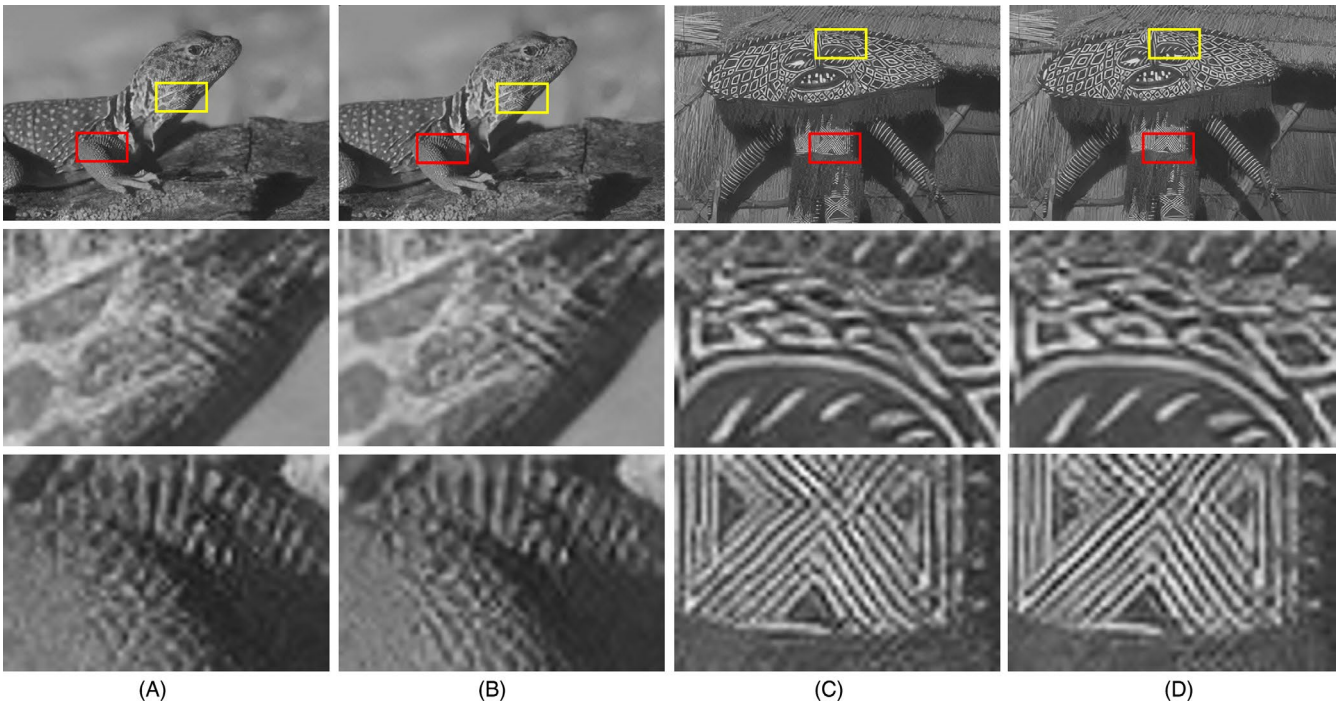

FIGURE 3 Comparison between $\text{ARGAN}_{\text{VGG}}$ and cARGAN at a similar objective image quality for the quality factor of 10 on BSDS500: (A) Result of $\text{ARGAN}_{\text{VGG}}$ (PSNR: 26.74 dB); (B) result of cARGAN (PSNR: 26.75 dB); (C) Result of $\text{ARGAN}_{\text{VGG}}$ (PSNR: 22.81 dB); and (D) result of cARGAN (PSNR: 22.84 dB). For each test image, the second and third rows are the image patches corresponding to the yellow and red rectangles in the top row. These are best viewed with zoom

and a random noise patch of the size 24 × 24 were used as input for network training [4]. The batch size was set to 16. The MATLAB JPEG encoder was used to obtain JPEG compressed images with quality factors of 10 and 20. The image restoration network was built for each quality factor. We followed the same parameter settings and obtained about the same performance presented in [4]. The number of steps needed to apply to the discriminator, k, was set as 1.

4.1.2 | Evaluation of the proposed method

Table 1 compares state-of-the-art deep learning-based compression artifact reduction techniques, including trainable nonlinear reaction diffusion [14], dual-domain CNNs (DDCN) [15], and ARGAN [4]. Likewise with [4], $\text{ARGAN}_{\text{base}}$ represents the ARGAN with $\lambda = 0$ in (6). Thus, $\text{ARGAN}_{\text{base}}$ trains only the generative network to minimize the MSE. Owing to its advanced architecture, $\text{ARGAN}_{\text{base}}$ exhibits state-of-the-art performance in terms of the PSNR and the SSIM.[1] However, when the adversarial loss is included, and the VGG loss is used as the fidelity loss (denoted as $\text{ARGAN}_{\text{VGG}}$), the generated images become more favorable [4]. However, the PSNR and the SSIM drop significantly, as shown in Table 1. A network with performance comparable to $\text{ARGAN}_{\text{base}}$, in terms of objective image quality, and $\text{ARGAN}_{\text{VGG}}$, in terms of subjective image quality, will be the most desirable. $\text{ARGAN}_{\text{VGG}}$ is not the same as the original ARGAN [15], because the JPEG-specific loss term [15] was not applied.

Table 2 compares the ARGANs trained using different loss functions of the generative network. $\text{ARGAN}_{\text{base}}$ showed the highest PSNR and SSIM compared to other methods, because only the fidelity loss function was minimized. As discussed, $\text{ARGAN}_{\text{VGG}}$ resulted in the lowest PSNR and SSIM. Owing to the explicit guidance of the adversarial loss, fidelity loss is not severely affected by adversarial loss. Therefore, cARGAN, with MSE-loss, resulted in only slightly lower PSNRs, compared to $\text{ARGAN}_{\text{base}}$.

Figure 2 compares the results for the JPEG images compressed with the quality factor of 10. Here, the results obtained by the four different training methods are compared. As can be seen, $\text{ARGAN}_{\text{base}}$ produced the most blurry results and the other methods generated images with at least more details and textures, compared to $\text{ARGAN}_{\text{base}}$, owing to the use of adversarial loss. However, $\text{ARGAN}_{\text{VGG}}$ sometimes overemphasize the textures, as shown in Figure 2C, resulting lower objective image quality. Figure 2D shows that the synthesized textures of cARGAN are not only visually appealing, but that they also closely resemble the real textures of the original images.

[1] In our implementation, $\text{ARGAN}_{\text{base}}$ showed about 0.1-dB lower PSNRs than the scores presented in Table 2 of [4].

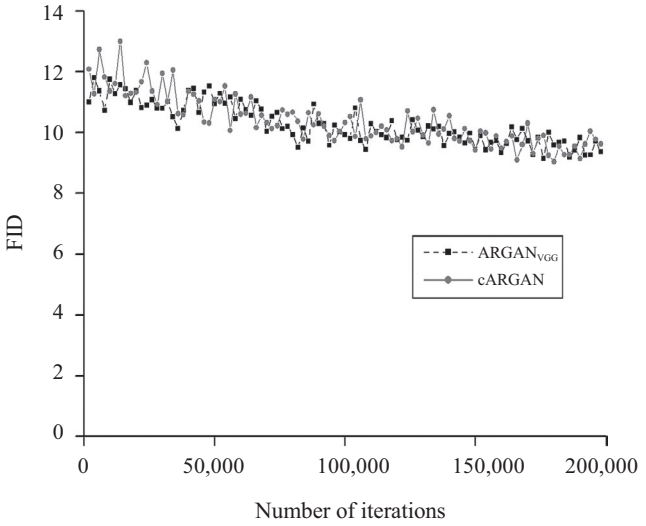

FIGURE 4 FID plot for ARGAN$_{VGG}$ and cARGAN with VGG loss

Because we have mainly attempted to trade off adversarial loss against fidelity loss, one may wonder whether similar performance can be achieved by fine-tuning λ in (6). To this end, we fine-tuned λ of ARGAN$_{VGG}$, such that the obtained objective scores became similar to those of cARGAN. Figure 3 compares the results of the JPEG images, compressed with the quality factor of 10. It can be seen that cARGAN, with VGG loss, produced images with more faithful textures than the fine-tuned version of ARGAN$_{VGG}$ at a similar level of objective quality score. Thus, the proposed method did not simply undermine the adversarial loss, like the fine-tuned version of ARGAN$_{VGG}$ did, it also found a solution that better balanced fidelity loss and adversarial loss.

Stability is an important issue of GAN [19–21]. The original ARGAN [4] tackled this issue by pretraining the generative network using the fidelity loss only. Then, it refined the pretrained network for actual GAN training. Because we changed the training mechanism of ARGAN, one may wonder whether the proposed method worsens the stability of GAN training. To this end, we measured the Fréchet inception distance (FID) [22], a widely used performance measure of GANs. Figure 4 shows FID scores against training iterations for ARGAN$_{VGG}$ and cARGAN with the VGG loss. Specifically, Microsoft COCO 2017 validation images [17] were compressed by JPEG with a quality factor of 10. These compressed images were then processed by ARGAN$_{VGG}$ and cARGAN, respectively, using the networks obtained for every training epoch. The FID scores were finally measured using the original image set and each restored image set. The result shows that both methods performed similarly in terms of the smoothness of the FID curve.

4.2 | Super-resolution

4.2.1 | Dataset and parameter settings

To implement SRGAN [2], a random sample of 350K images from the Microsoft COCO 2017 training database [17] and the same number of random samples from the ImageNet training database [23] were used for training. An image patch of size 96 × 96 was used as input for the network training. The batch size was set as 16. The low-resolution images were obtained by downsampling the original images using a downsampling factor of 4 and a bicubic kernel.

4.2.2 | Evaluation of the proposed method

Here, we show the generalizability of the proposed training scheme to super-resolution. Likewise, with ARGAN$_{base}$, SRGAN$_{base}$ represents the SRGAN with l_{fid}^{MSE} as the fidelity loss and $\lambda = 0$ (defined as SRResNet-MSE in [2]). SRGAN$_{VGG}$ denotes the original SRGAN [2]. Table 3 compares the performance of the SRGANs on Set5 , Set14 [11], and BSD100 [18].

TABLE 3 Comparisons of the training methods of SRGAN

		SRGAN$_{base}$		SRGAN$_{VGG}$		cSRGAN	
Upsampling ratio	**Fidelity loss**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Set5							
4	MSE	32.27	0.8994	N/A	N/A	31.07	0.8619
	VGG	N/A	N/A	29.74	0.8387	29.81	0.8395
Set14							
4	MSE	28.57	0.7814	N/A	N/A	27.54	0.7379
	VGG	N/A	N/A	26.21	0.6968	26.30	0.6952
BSD100							
4	MSE	27.66	0.7388	N/A	N/A	27.06	0.7086
	VGG	N/A	N/A	25.65	0.6531	25.69	0.6546

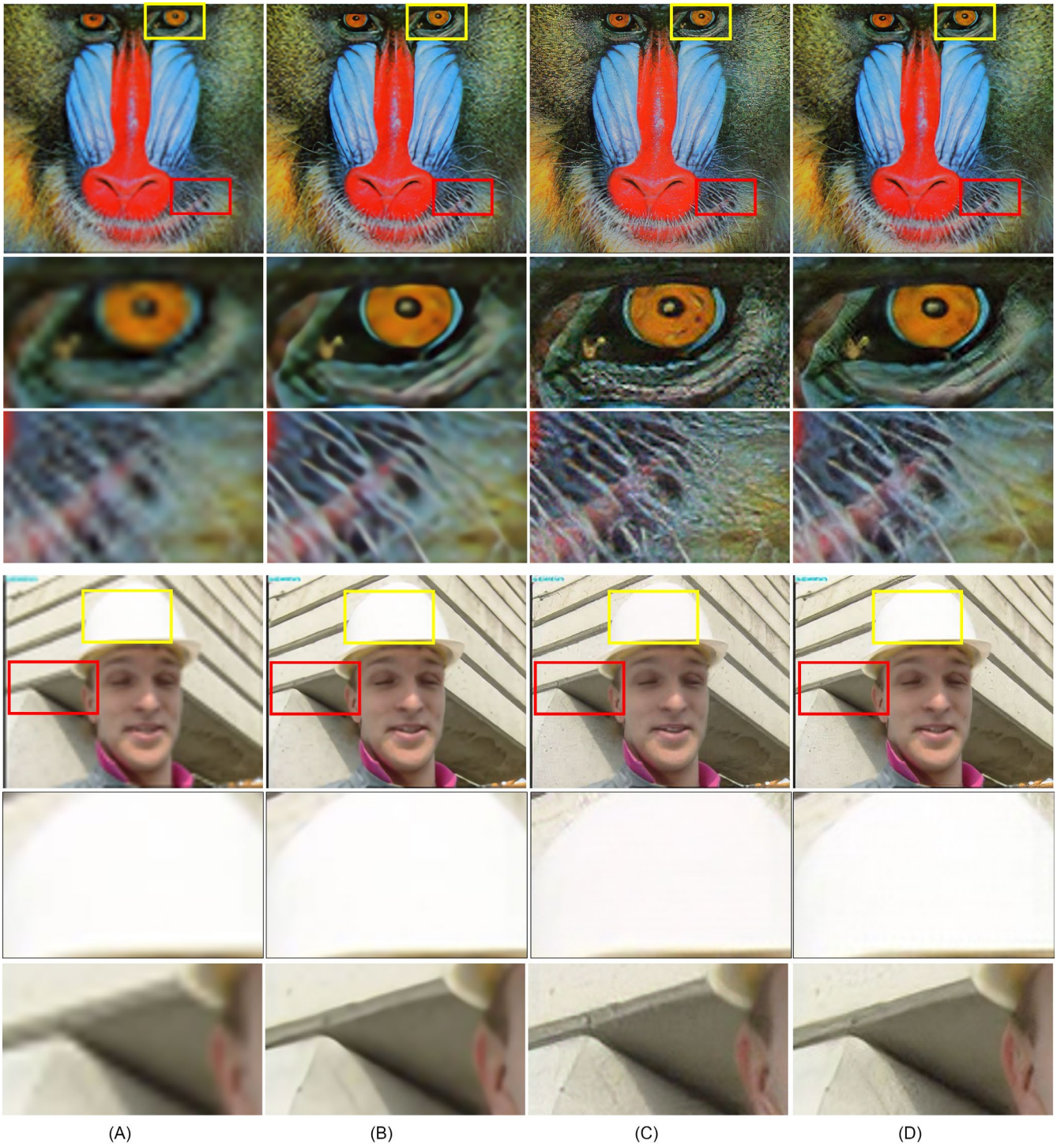

FIGURE 5 Comparison of different SRGAN training schemes under an upsampling factor of 4 on Set14: (A) Bicubic (PSNR: 22.44 dB (top), 29.38 dB (bottom)); (B) $\text{SRGAN}_{\text{base}}$ (PSNR: 23.01 dB, 33.88 dB); (C) SRGAN_{VGG} (PSNR: 20.45 dB, 31.48 dB); and (D) cSRGAN (PSNR: 22.60 dB, 32.74 dB). For each test image, the second and third rows are the image patches corresponding to the yellow and red rectangles in the top row. These are best viewed with zoom

As expected, $\text{SRGAN}_{\text{base}}$ showed the highest PSNR[2] and SSIM, compared to other methods. However, it produced blurry images, as shown in Figure 5B. Additionally, $\text{SRGAN}_{\text{VGG}}$ yielded images with photorealistic textures. Using the proposed training scheme, cSRGAN, with the MSE-loss, resulted in slightly lower PSNRs and SSIMs than $\text{SRGAN}_{\text{base}}$, and cSRGAN with the VGG loss resulted in comparable or slightly higher PSNRs and SSIMs than $\text{SRGAN}_{\text{VGG}}$. The upsampling

[2]In our implementation, $\text{SRGAN}_{\text{base}}$ showed about 0.1–0.2-dB higher PSNRs than the scores presented in Table 2 of [2].

FIGURE 6 Comparison between $\text{SRGAN}_{\text{VGG}}$ and cSRGAN at the similar objective image quality for the upsampling factor of 4 on Set14: (A) Result of $\text{SRGAN}_{\text{VGG}}$ (PSNR: 21.67 dB); and (B) Result of cSRGAN (PSNR: 21.56 dB). For each test image, the second and third rows are the image patches corresponding to the yellow and red rectangles in the top row. These are best viewed with zoom

results shown in Figure 5D show that cSRGAN, with the MSE-loss, produced photorealistic and faithful image textures. Moreover, the magnified image regions corresponding to the yellow rectangles in Figure 5 indicate that the proposed method does not make artificial textures in textureless regions.

We also fine-tuned λ of $\text{SRGAN}_{\text{VGG}}$, such that the obtained objective scores became similar to the scores of cSRGAN. Figure 6 shows that cSRGAN, with VGG loss, produced images with more faithful textures than the fine-tuned version of $\text{SRGAN}_{\text{VGG}}$ at a similar level of objective quality score. This is consistent with the result we found for the compression artifact reduction task.

As in Figure 7, we measured FID scores against training iterations for $\text{SRGAN}_{\text{VGG}}$ and cSRGAN with VGG loss. Microsoft COCO 2017 validation images [17] were downsampled by a factor of 4 using bicubic downsampling. These low-resolution images were then super-resolved with $\text{SRGAN}_{\text{VGG}}$ and cSRGAN, using the networks obtained for every training epoch. The result shows that both methods performed similarly in terms of smoothness of the FID curve. This again verifies that the proposed training method did not particularly worsen the stability of GAN training.

More results and supplementary materials are available at https://github.com/spnova12/ARGAN_FIR.

5 | CONCLUSIONS

GAN has been successfully applied in many image restoration applications. However, previous GAN-based image restoration methods tended to overemphasize adversarial loss. This resulted in non-negligible degradation of the objective image quality. In this study, we attempted to broaden the applicability of GAN-based image restoration by finding a better trade-off between

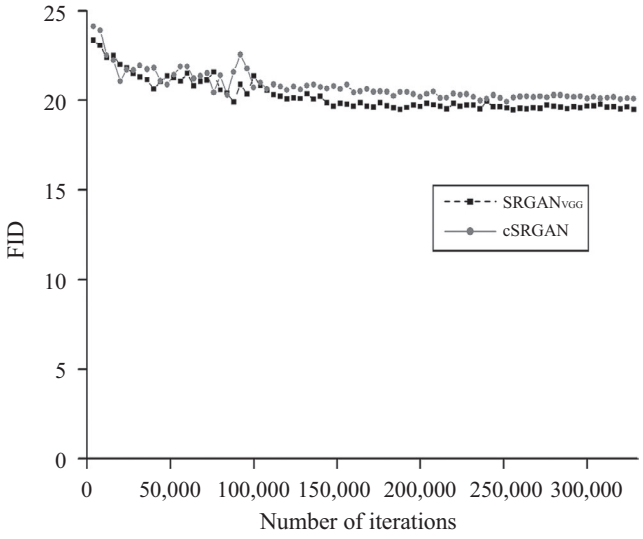

FIGURE 7 FID plot for $\text{SRGAN}_{\text{VGG}}$ and cSRGAN with VGG loss

fidelity and adversarial loss. In particular, a novel training scheme that explicitly constrained the adversarial loss, considering its impact on fidelity loss, was presented. Experimental results on compression artifact reduction and super-resolution tasks demonstrated that the proposed training scheme can enable photorealistic and faithful image restoration.

ORCID

Seung-Won Jung http://orcid.org/0000-0002-0319-4467

REFERENCES

1. K. Zhang et al., *Beyond a Gaussian denoiser: Residual learning of deep cnn for image denoising*, IEEE Trans. Image Process. **26** (2017), no. 7, 3142–3155.
2. C. Ledig et al., *Photo-realistic single image super-resolution using a generative adversarial network*, IEEE Conf. Comput. Vision Pattern Recogn. (CVPR), Honolulu, HI, USA, July 2017, pp. 105–114.
3. S. Nah, T. H. Kim, and K. M. Lee, *Deep multi-scale convolutional neural network for dynamic scene deblurring*, IEEE Conf. Comput. Vision Pattern Recogn. (CVPR), Honolulu, HI, USA, July 2017, pp. 257–256.
4. J. Guo and H. Chao, *One-to-many network for visually pleasing compression artifacts reduction*, IEEE Conf. Comput. Vision Pattern Recogn. (CVPR), Honolulu, HI, USA, July 2017, pp.4687–4876.
5. H. Zhao et al., *Loss functions for image restoration with neural networks*, IEEE Trans. Comput. Imag. **3** (2017), no. 1, 47–57.
6. I. J. Goodfellow et al., *Generative adversarial nets*, Proc. Adv. Neural Inform. Process. Syst. (NIPS), Montreal, Canada, Dec. 2014, pp. 2672–2680.
7. A. Radford, L. Metz, and S. Chintala, *Unsupervised representation learning with deep convolutional generative adversarial networks*, Proc. Int. Conf. Learning Representations (ICLR), San Juan, Puerto Rico, May 2–4, 2016.
8. W.-S. Park and M. Kim, *CNN-based in-loop filtering for coding efficiency improvement*, Proc. IEEE Image, Video, Multidimensional Signal Process. Workshop (IVMSP), Bordeaux, France, July 2016, pp. 1–5.
9. K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, Proc. Int. Conf. Learning Representations (ICLR), San Diego, CA, USA, May 7–9, 2015.
10. J. Johnson, A. Alahi, and L. Fei-Fei, *Perceptual losses for real-time style transfer and super-resolution*, Proc. Eur. Conf. Comput. Vision (ECCV), Amsterdam, Netherlands, Oct. 11–14, 2016.
11. R. Zeyde, M. Elad, and M. Protter, *On single image scale-up using sparse-representations*, Proc. Int. Conf. Curves Surfaces, Avignon, France, June 24–30, 2010, pp. 711–730.
12. Z. Wang et al., *Image quality assessment: from error visibility to structural similarity*, IEEE Trans. Image Process. **13** (2004), no. 4, 600–612.
13. D. Kingma and J. Ba, *Adam: A method for stochastic optimization*, Proc. Int. Conf. Learning Representations (ICLR), San Diego, CA, USA, May 7–9, 2015.
14. Y. Chen, W. Yu, and T. Pock, *On learning optimized reaction diffusion processes for effective image restoration*, Proc. IEEE Conf. Comput. Vision Pattern Recogn. (CVPR), Boston, MA, USA, June 7–12, 2015, pp. 5261–5269.
15. J. Guo and H. Chao, *Building dual-domain representations for compression artifacts reduction*, Proc. Eur. Conf. Comput. Vision (ECCV), Amsterdam, Netherlands, Oct. 11–14, 2016, pp. 628–644.
16. L. Galteri et al., *Deep generative adversarial compression artifact removal*, Proceedings of IEEE Int. Conf. Comput. Vision (ICCV), Venice, Italy, Oct. 2017, pp. 4836–4845.
17. T.-Y. Lin et al., *Microsoft COCO: Common objects in context*, Proc. Eur. Conf. Comput. Vision (ECCV), Zurich, Switzerland, Sept. 6–12, 2014, pp. 740–755.
18. P. Arbelaez et al., *Contour detection and hierarchical image segmentation*, IEEE Trans. Pattern Anal. Mach. Intell. **33** (2011), no. 5, 898–916.
19. N. Kodali et al., *On convergence and stability of GANs*, arXiv preprint arXiv:1705.07215, Dec. 2017.
20. K. Shmelkov, C. Schmid, and K. Alahari, *How good is my GAN?*, Proc. Eur. Conf. Comput. Vision (ECCV), Munich, Germany, Sept. 2018.
21. K. Kurach et al., *The GAN landscape: Losses, architectures, regularization, and normalization*, arXiv preprint arXiv:1807.04720, Oct. 2018.
22. M. Heusel et al., *GANs trained by a two time-scale update rule converge to a local Nash equilibrium*, Proc. Adv. Neural Inform. Process. Syst. (NIPS), Long Beach, CA, USA, Dec. 2017, pp. 6629–6640.
23. J. Deng et al., *ImageNet: A large-scale hierarchical image database*, Proc. IEEE Conf. Comput. Vision Pattern Recogn. (CVPR), Miami, FL, USA, June 2009, pp. 248–255.
24. M. Bevilacqua et al., *Low-complexity single-image super-resolution based on nonnegative neighbor embedding*, Proc. British Mach. Vision Conf. (BMVC), Surrey, UK, Sept. 2012, pp. 1–10.
25. ITU. *Rec. ITU-R BT.500-13, Methodology for the subjective assessment of the quality of television pictures*, ITU, 2012.

AUTHOR BIOGRAPHIES

Dong-Wook Kim received the BS degree in Multimedia Engineering from Dongguk University, Seoul, Rep. of Korea, in 2017. He is currently working on a master's degree. His research interests include image reconstruction, image colorization, and generative adversarial networks.

Jae-Ryun Chung received the BS degree in Multimedia Engineering from Dongguk University, Seoul, Rep. of Korea, in 2018. He is currently a master's degree student in multimedia engineering at the same University. His current research interests include deep learning-based artifacts reduction and computer vision applications.

Jongho Kim received the BS degree in Control and Computer Engineering from Korea Maritime University, Busan, Rep. of Korea, in 2005 and the MS degree from the University of Science and Technology, Daejeon, Rep. of Korea, in 2007. Since 2008, he has been a senior researcher at Electronics and Telecommunications Research Institute in Daejeon, Rep. of Korea. His research interests include video processing, video coding, and machine learning-based compression.

Dae Yeol Lee received the BS and MEng degrees in Electrical and Computer Engineering from Cornell University, Cornell University, NY, USA, in 2012 and 2013, respectively. Since 2013, he has been a researcher at Electronics and Telecommunications Research Institute (ETRI) in Daejeon, Rep. of Korea. In 2018, he joins the University of Texas at Austin, USA, as a PhD student. His research interests include image/video quality assessment and enhancement, immersive media, machine learning, and computer vision.

Seyoon Jeong received the BS and MS degrees in Electronic Engineering from Inha University, Seoul, Rep. of Korea, in 1995 and 1997, respectively, and a PhD degree in Electrical and Electronic Engineering from Korea Advanced Institute of Science and Technology , Daejeon, Rep. of Korea, in 2014. In 2014, he attended the Media Communications Laboratory at the University of Southern California , Los Angeles, CA, USA, as a Visiting Scholar. His research interests include video quality assessment, HDR video processing, perceptual, and machine learning-based video coding.

Seung-Won Jung received the BS and PhD degrees in Electrical Engineering from Korea University, Seoul, Rep. of Korea, in 2005 and 2011, respectively, where he was a research professor with the Research Institute of Information and Communication Technology at Korea University from 2011 to 2012. He was a research scientist with the Samsung Advanced Institute of Technology, Yongin, Rep. of Korea, from 2012 to 2014. He is currently an assistant professor with the Department of Multimedia Engineering, Dongguk University, Seoul, Rep. of Korea. He has authored over 50 peer-reviewed articles in international journals. His current research interests include deep learning and computer-vision algorithms.