Compression artifacts reduction by improved generative adversarial networks

Zengshun Zhao[1,2,3*] , Qian Sun[1], Haoran Yang[1], Heng Qiao[2], Zhigang Wang[4] and Dapeng Oliver Wu[2]

Abstract

In this paper, we propose an improved generative adversarial network (GAN) for image compression artifacts reduction task (artifacts reduction by GANs, ARGAN). The lossy compression leads to quite complicated compression artifacts, especially blocking artifacts and ringing effects. To handle this problem, we choose generative adversarial networks as an effective solution to reduce diverse compression artifacts. The structure of "U-NET" style is adopted as the generative network in the GAN. A discriminator network is designed in a convolutional manner to differentiate the restored images from the ground truth distribution. This approach can help improve the performance because the adversarial loss aggressively encourages the output image to be close to the distribution of the ground truth. Our method not only learns an end-to-end mapping from input degraded image to corresponding restored image, but also learns a loss function to train this mapping. Benefit from the improved GANs, we can achieve desired results without hand-engineering the loss functions. The experiments show that our method achieves better performance than the state-of-the-art methods.

Keywords: GANs, CNN, Compression artifacts, JPEG compression

1 Introduction

Image restoration technology has become one of the most important applications in computer vision and computer graphics and attracted increasing attention in the field of digital image processing, such as image haze removal [1], image super-resolution [2–4], image deblur [5, 6], and image understanding [7]. Image compression artifacts reduction aims at recovering a sharp image from the degraded image which is formed by JPEG compression or other causes. JPEG compression is a kind of lossy compression method that uses inaccurate approximations for representing the encoded content. Although JPEG compression is very common in our daily life, it may lead to quite complicated compression artifacts, especially blocking artifacts and ringing effects which not only decrease the perceptual visual quality, but also introduce obstruction to other low-level image processing routines.

In this paper, we use a deep learning-based approach for image compression artifacts reduction. More

specifically, we propose a principled and efficient generative adversarial network (GAN) for this task. We denote the proposed networks as artifacts reduction by GANs (ARGAN) which was inspired from the GANs [8]. Similar to the standard GANs, ARGAN also consists of two feed-forward convolutional neural networks (CNNs), the generative network G and the discriminative network D. The purpose of the generative network G is to generate reasonable results from the input degraded images. The goal of the discriminative network D is to discover the discrepancy between the generated image and the corresponding ground-truth image. Our proposed method differs from the existing traditional [9] or other deep learning-based approaches [10]. The traditional approaches need to extract the features of the images manually. The deep learning-based approaches are usually based on CNN. We are the first to use (GANs) for image compression artifacts reduction.

There are two main contributions in our work:

(1) We are the first to use an end-to-end generative adversarial network (GAN) for image compression artifacts reduction. The experiments show that our method achieves better performance than the state-

* Correspondence: zhaozengshun@163.com
[1]College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, People's Republic of China
[2]Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA
Full list of author information is available at the end of the article

of-the-art methods [9–11]. In this paper, we focus on the restoration of the luminance channel (in YCrCb space) as in [10], and the network is specially designed for this task.

(2) We demonstrate that generative adversarial networks are useful in the image compression artifacts reduction task and can achieve better quality than the traditional or other deep learning-based methods. Our method directly learns an end-to-end mapping which can effectively estimate the reasonable results from input degraded images and make the restored image more real.

2 Related works

Various methods have been proposed to reduce image compression artifacts. Early works, such as sparsity-based image restoration approaches [12, 13], are proposed to produce sharpened images. Now in the field of image compression artifacts reduction, the existing popular methods can be roughly divided into two categories including deblocking-oriented and deep learning-based approaches. The goal of the deblocking-oriented algorithms is to eliminate ringing and blocking artifacts. Shape-adaptive discrete cosine transform (SA-DCT) [9] is widely considered as the state-of-the-art deblocking-oriented algorithm, but like many other deblocking-oriented algorithms, it may produce vague effects and is not able to maintain the sharp edges of the original images. Recently, Wu [14] proposed wavelet transform for blocking artifact reduction based on Meyer algorithm.

Neural networks and deep learning currently provide the best solutions to many problems [10, 11, 15–17]. In recent years, deep learning has been increasingly improved in its ability to provide accurate recognition and prediction. The image compression artifacts reduction technology thus made a breakthrough making use of the recent progresses in deep learning. Dong et al. [10] applied deep learning to the task of image restoration, where a simple model consisting of four convolution layers was designed for image compression artifacts reduction. With the improvement in GPU performance and optimization algorithms, researchers started to train larger and deeper neural networks.

Wang et al. [11] proposed a Deep Dual-Domain based fast restoration model to remove artifacts of JPEG-compressed images. It leverages the large learning capacity of deep networks. Extensive experiments verify the superiority of the proposed D3 model over several state-of-the-art methods. Soon after, Wang et al. [15] proposed an intensity-guided CNN (IG-Net) model, which learns an end-to-end mapping between the intensity image and distorted depth map to the uncompressed depth map.

Convolutional neural networks are one of the most important methods in deep learning, which are widely used in the field of computer vision. In recent years, the convolutional neural networks (CNNs) are trained in the supervised manner for various image-related tasks, such as object detection [18, 19] By penalizing the discrepancy between the output image and ground-truth image, optimal CNNs can be trained to discover the mapping from the input image to the reasonable output image. These various CNN models mainly differ in the network construction and loss function design. One of the most straightforward methods is to pixel-wisely evaluate the output images [2, 3, 10], e.g., using L2 (or L1) norm to calculate the distance between the output and ground-truth images in the pixel space. However, this method may generate blurred results which could make the output images look unsatisfactory.

Fortunately, there are a large body of successful applications based on generative adversarial networks (GANs) (e.g., SRGAN [4], DCGAN [20], Pix2Pix [21]) since Goodfellow et al. [8] first officially proposed GANs in 2014. GANs perform an adversarial process alternating between identifying and faking, and the generative adversarial losses are formulated to evaluate the discrepancy between the generated distribution and the real data distribution. A lot of researches show that generative adversarial losses are beneficial for generating more "realistic" images. Inspired by the success of generative adversarial networks (GANs) on image-to-image translation [21, 22], we designed an efficient GAN network for compression artifacts reduction. In this paper, we show that the proposed networks are effective on our image task. Experiments show that our method outperforms current state-of-the-art methods [9–11] both perceptually and quantitatively.

3 Method

In this section, we will introduce the proposed generative adversarial networks for image compression artifacts reduction. First, the generative adversarial losses of ARGAN are described. Then, an overview of the proposed method and the details of our networks are illustrated.

3.1 Generative adversarial loss

GAN-based models have been widely used in learning generative model due to their success in image generation. GAN was proposed to solve the disadvantages of other generative models. Instead of maximizing the possibility, GAN introduces the theory of adversarial learning between the generator and the discriminator. This adversarial process gives GAN obvious advantages over the other generative models. Moreover, GAN can sample the generated data in a simple way unlike other models in which the sampling is computationally slow and not accurate. For these advantages, GAN gained our

attention, and this is the original intention for us to use the framework of GAN. We therefore adapt the GAN learning strategy to tackle the problem of image compression artifacts. More specifically, the proposed ARGAN consists of two feed-forward convolutional neural networks (CNNs): the generative network G and the discriminative network D. The reason why we use CNN is that it can greatly stabilize GAN training. ARGAN suggests an architecture guideline in which the generator is composed of a CNN and a transposed CNN, and the discriminator is composed of a CNN with an output dimension 1. Batch normalization, rectified linear unit (ReLU) and leaky rectified linear unit (LeakyReLU) activation functions are utilized for the generator and the discriminator to help stabilize the GAN training. The purpose of the generative network G is to generate reasonable result $G(x)$ from input image x. Meanwhile, each input image x has a corresponding ground-truth image y. $G(x)$ is encouraged to have the same data distribution with the ground-truth image y. The goal of the discriminative network D is to discover the discrepancy between the data distribution of generated image and the corresponding ground-truth image. G and D compete with each other to achieve their respective purposes, thus generating the term adversarial. The generative adversarial loss can be expressed as:

$$\min_G \ \max_D E_y[\ \log D(y)] + E_x[\ \log(1-D(G(x)))] \quad (1)$$

The loss function is a binary cross entropy function that is commonly used in binary classification problems. Where in Eq. 1, x is the input degraded image which has a corresponding ground-truth image y. G tries to minimize the loss, whereas the adversarial D tries to maximize it.

Some recent works have found that it is desirable to mix the generative adversarial loss with other traditional loss, such as L1 [21] or L2 [23] distance. We also consider this trick, but here, L2 distance is adopted rather than L1 distance because L2 distance encourages G to explore the mapping from the input image to its ground truth and therefore make images more realistic:

$$L_{L2}(G) = E_{x,y}\big[\|y-G(x)\|_2^2\big] \quad (2)$$

Above all, the loss function of generative network L_G and the loss function of discriminative network L_D are formally defined as:

$$L_G = \ \log(1-D(G(x))) + \lambda(y-G(x))^2 \quad (3)$$

$$L_D = - \ \log(D(y)) - \log(1-D(G(x))) \quad (4)$$

The purpose of the discriminative network D is to distinguish real or fake data. From D's perspective, if a sample comes from real data, D will maximize its output.

While, if a sample comes from G, D will minimize its output. Thus, the overall aim is to minimize Eq. 4. Simultaneously, G wants to confuse D, so it tries to maximize D's output when a fake sample is presented to D, that is, to minimize Eq. 3 where x is the input-degraded image whose corresponding ground-truth image is y and λ is the hyper parameter.

3.2 Architecture of networks
The architecture of the proposed ARGAN is based on two deep convolutional neural networks, namely the generative network G and discriminative network D, whose combined efforts aim at obtaining a sharp image for a given input image. Figure 1 shows the architecture of the proposed ARGAN.

3.2.1 Generative network
The generative network G is designed for generating a sharp image by reducing the image compression artifacts given the input image. The structure of generative network is inspired by the configuration of "U-Net" [24] which is an encoder-decoder with skip connections between mirrored layers in the encoder and decoder stacks. This kind of structure allows the same size of the input and output image, and the local and global information of the image can be taken into account at the same time. This is the reason why we adopted the structure of "U-NET" as the generative network in the ARGAN. The network G firstly encodes the input image into high-dimensional representation, utilizing a stack of convolution-batch normalization-LeakyReLU layers, and then the rest of deconvolution-batch normalization-ReLU layers will decode the output image. The details of the generative network G are demonstrated in the Table 1.

3.2.2 Discriminative network
The discriminative network D is proposed to compute the discrepancy between the data distribution of the generated images and the ground-truth images. The output of the discriminative network D represents the possibility that the input image comes from the real-world dataset (true) rather than from the generative network (fake). All the convolution layers use LeakyReLU activations, with the exception of the final layer, which adopts a sigmoid activation. The details of the network D are listed in the Table 2.

4 Results and discussion experiments
4.1 Experiment settings
We use the VOC2012 dataset [25], which includes 16,700 images, as our training set. We trained our network models (in 1,500,000 iterations with a batch size of 64) on one NVIDIA GTX970 GPU using Pytorch. The
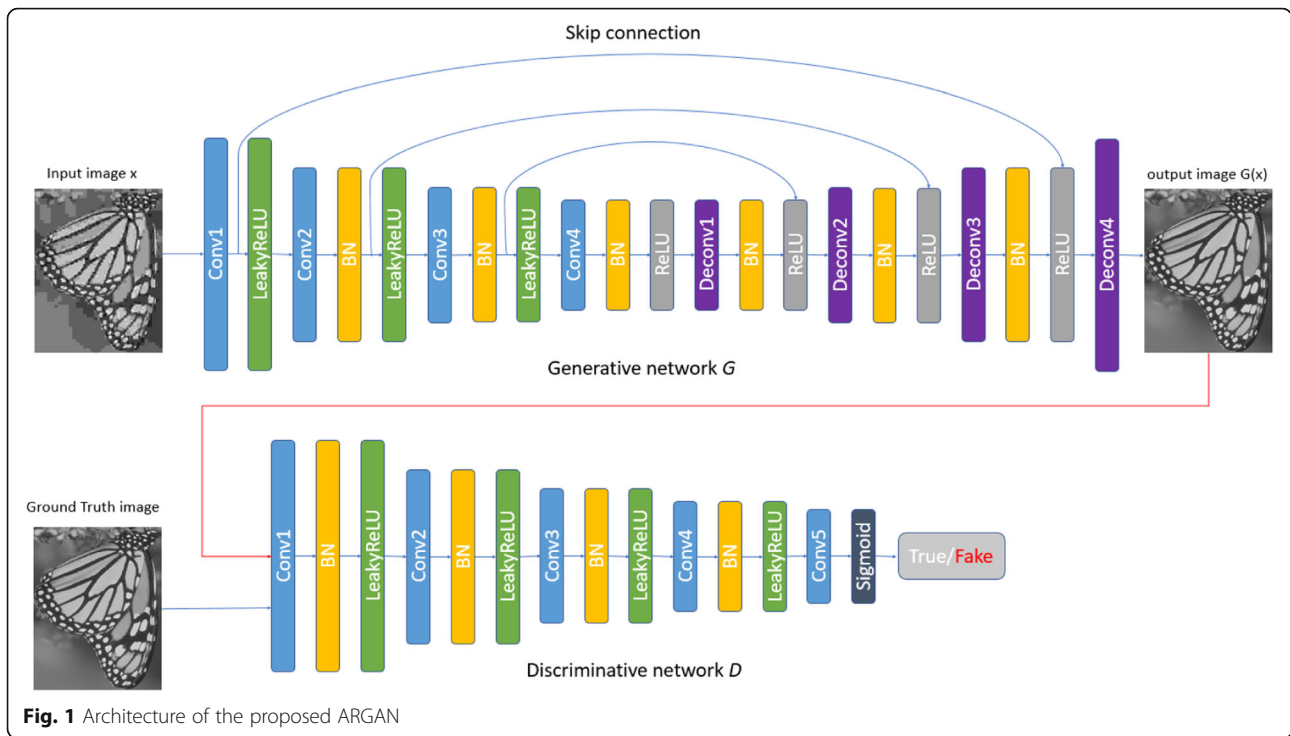
Fig. 1 Architecture of the proposed ARGAN

weights of the networks are trained from scratch. Training images are random cropped into 64×64 sub-images during each batch. In the training phase, we follow [21] and use an Adam solver with a learning rate of 0.0002 and an initial momentum of 0.5. After one update of the discriminative network D, the generative network G will also be updated one time.

We compare our method with several state-of-the-art algorithms, including the deblocking oriented method SA-DCT [9] and the deep model ARCNN [10], D3 [11],

on restoring JPEG-compressed images. Following [8], we adopt standard JPEG compression and set JPEG quality $q = 10, 20, 30, 40$ (from low quality to high quality) in JPEG encoder. With the test datasets LIVE1 [26] and the 5 test images in [9], the PSNR, SSIM [27] measurements are applied for quality assessment. In this paper, our method is only applied on the luminance channel as same as [10] (Y channel in YcbCr color space), and PSNR is evaluated on the Y channel. The PSNR can be defined as:

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i, j) - K(i, j)\|^2 \tag{5}$$

Table 1 The details of the generative network G

Generative network G
Input: *x*, size = 64×64
[Conv 1]   Conv (3, 3, 64), stride = 2, padding = 1; LeakyReLU;
[Conv 2]   Conv (4, 4, 128), stride = 2, padding = 1; BatchNorm; LeakyReLU;
[Conv 3]   Conv (4, 4, 256), stride = 2, padding = 1; BatchNorm; LeakyReLU;
[Conv 4]   Conv (4, 4, 512), stride = 2, padding = 1; BatchNorm; ReLU;
[DeConv1]   DeConv (4, 4, 256), stride = 2, padding = 1; BatchNorm; ReLU;
[DeConv2]   DeConv (4, 4, 128), stride = 2, padding = 1; BatchNorm; ReLU;
[DeConv3]   DeConv (4, 4, 64), stride = 2, padding = 1; BatchNorm; ReLU;
[DeConv4]   DeConv (4, 4, 3), stride = 2, padding = 1
Output: *G(x)*, size = 64×64

Table 2 The details of the discriminative network D

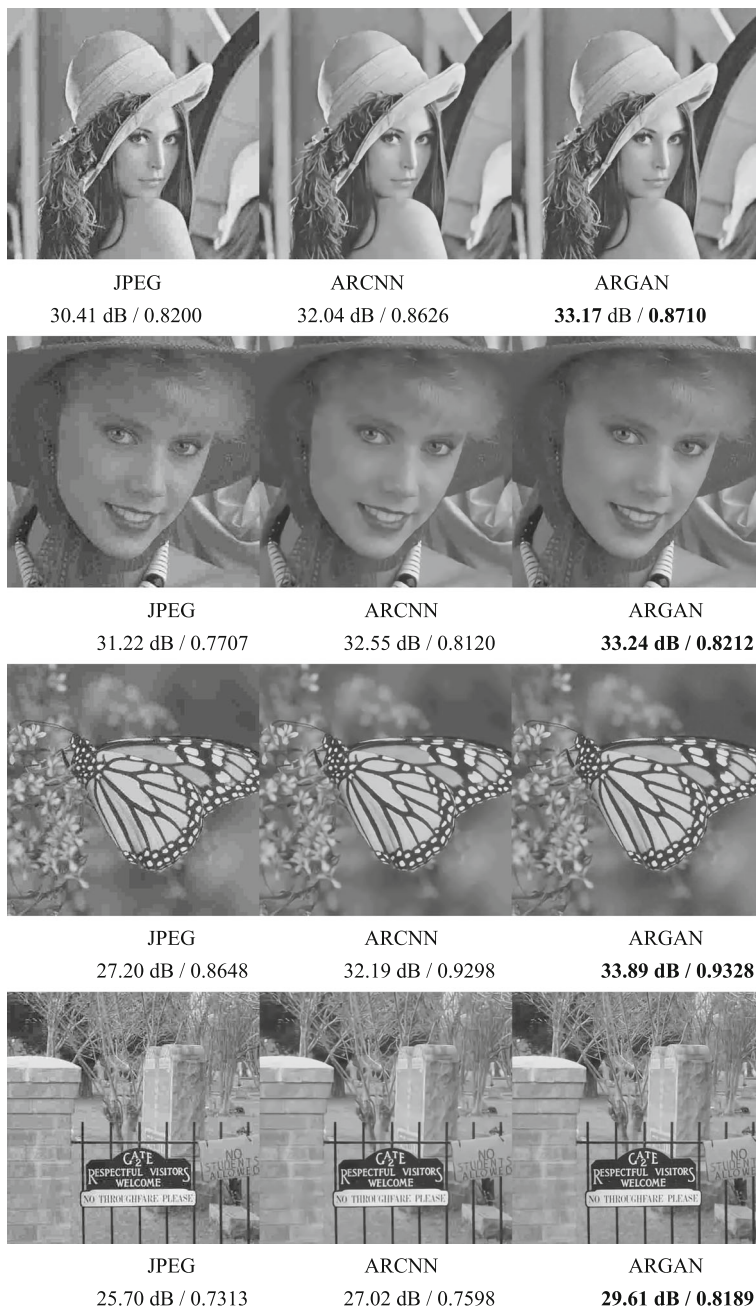| Discriminative network D |
| --- |
| Input: G(x) or ground truth, size = 64×64 |
| [Conv 1]   Conv (4, 4, 64), stride = 2, padding = 1; BatchNorm; LeakyReLU; |
| [Conv 2]   Conv (4, 4, 128), stride = 2, padding = 1; BatchNorm; LeakyReLU; |
| [Conv 3]   Conv (4, 4, 256), stride = 2, padding = 1; BatchNorm; LeakyReLU; |
| [Conv 4]   Conv (4, 4, 512), stride = 2, padding = 1; BatchNorm; LeakyReLU; |
| [Conv 5]   Conv (4, 4, 1), stride = 1, padding = 0; Sigmoid |
| Output: size = 1×1 |

Fig. 2 The results of the ARCNN and ARGAN

$$
\begin{aligned}
\text{PSNR} &= 10 \cdot \log_{10}\left(\frac{\text{MAX}_I^2}{\text{MSE}}\right) \\
&= 20 \cdot \log_{10}\left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}}\right)
\end{aligned}
\tag{6}
$$

where I is the ground-truth image and K is the restored image. The size of the I and K are both $m \times n$, and the MAX_I is the gray level of the image. Generally speaking, the better the image quality, the larger the value of PSNR.

The SSIM can be defined as:

$$
\text{SSIM}(I, K) = \frac{(2\mu_I\mu_K + c_1)(\sigma_{\text{IK}} + c_2)}{\left(\mu_I^2 + \mu_K^2 + c_1\right)\left(\sigma_I^2 + \sigma_K^2 + c_2\right)}
\tag{7}
$$

where I is the ground-truth image and K is the restored image and μ_I and μ_K are the average value of I and K, respectively. σ_I and σ_K are the standard deviation of I and K, respectively. σ_{IK} is the covariance, and c_1 and c_2 are

Table 3 The average results of PSNR (dB), SSIM on the LIVE1 dataset [27]

	Quality	JPEG	SA-DCT [9]	ARCNN [10]	D3 [11]	ARGAN
PSNR	$q = 10$	27.77	28.65	29.13	29.96	31.25
	$q = 20$	30.07	30.81	31.40	32.21	33.41
	$q = 30$	31.41	32.08	32.69	33.28	34.81
	$q = 40$	32.35	32.99	33.63	34.75	35.73
SSIM	$q = 10$	0.7905	0.8093	0.8232	0.8233	0.8544
	$q = 20$	0.8683	0.8781	0.8886	0.8903	0.9044
	$q = 30$	0.9000	0.9078	0.9166	0.9207	0.9263
	$q = 40$	0.9173	0.9240	0.9306	0.9296	0.9388

constants. When $I = K$, the SSIM usually approximately to be 1.

4.2 Intuitive visual comparison

Figure 2 gives the comparative results of our method and the state-of-the-art algorithm ARCNN [10]. As shown in Fig. 2, our method obtains the highest image recovery quality ($q = 10$) and the best PSNR and SSIM scores.

4.3 Quantitative comparisons

In order to prove the efficiency of our proposed method, we compare ARGAN with three methods: SA-DCT [9], which is widely considered as the state-of-the-art deblocking-oriented algorithm; ARCNN [10], an efficient deep learning-based method for image compression artifacts reduction; D3 (Deep Dual-Domain based fast restoration) model [11]. Using the exactly the same dataset, we directly use the results of ARCNN in the original paper. We trained and tested D3 model according to their paper [11]. As shown in Table 3, our method always yields the highest scores. The results show that the results of the proposed ARGAN are superior to the other algorithms. We have also conducted an evaluation on five test images used in [9]. The results also show that ARGAN achieves the highest performance. The results are listed in Table 4.

Table 4 The average results of PSNR (dB), SSIM on the five test images [9]

	Quality	JPEG	SA-DCT [9]	ARCNN [10]	D3 [11]	ARGAN
PSNR	$q = 10$	27.82	28.88	29.04	29.65	31.06
	$q = 20$	30.12	30.92	31.16	32.12	33.06
	$q = 30$	31.48	0.8844	32.52	33.29	34.28
	$q = 40$	32.43	0.9011	33.34	34.36	35.05
SSIM	$q = 10$	0.7800	0.8071	0.8111	0.8279	0.8401
	$q = 20$	0.8541	0.8914	0.8967	0.8974	0.9016
	$q = 30$	0.8844	0.8914	0.8967	0.8981	0.9155
	$q = 40$	0.9011	0.9055	0.9101	0.9192	0.9257

5 Discussion

We have presented corresponding experimental results in the above section. The proposed method is compared with several state of the arts, e.g., JPEG, SA-DCT [9], and the deep model ARCNN [10], D3 (Deep Dual-Domain based fast restoration) model [11]. We can see that the proposed method is extremely effective in dealing with various compression artifacts.

Subjectively speaking, the performance improvement owes to the following aspects. Firstly, we modified the generative model to be the "U-Net" [24] instead of the standard CNN within the inner structure of GANs. Accordingly, utilizing a stack of convolution-batch normalization-LeakyReLU layers makes the model more effective.

This paper provides the research theory for the image compression artifacts reduction by modified GANs. Even with complicated working conditions on the scene, the performance is improved consistently and greatly. Moreover, we designed a customized network to further improve the PSNR and SSIM. In the further work, we hope to reduce the computational load and increase the efficiency of the model. A structure-optimized GANs can improved neural network and may help us to solve this problem. This issue will be our future work.

6 Conclusion

In this paper, image compression artifacts reduction is achieved by generative adversarial networks, and we make sufficient comparisons with SA-DCT [9], ARCNN [10], and D3 [11], respectively. The results show that the proposed ARGAN is effective in removing various compression artifacts. The detail information maintains better, making the images look clearer.

Abbreviations
ARGAN: Artifacts reduction by GANs; CNNs: Convolutional neural networks; GANs: Generative adversarial networks; PSNR: Peak signal-to-noise ratio; SSIM: Structural similarity index

Availability of data and materials
The dataset used during the current study is VOC2012 dataset [25] and are available online or from the corresponding author on reasonable request.

Authors' contributions
ZZ was a major contributor in writing the manuscript. And he analyzed and interpreted the entire framework of GANS with the help of ZW and DW. QS, HY, and HQ performed the coding and experiments. All authors read and approved the final manuscript.

Author details
[1]College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, People's Republic of China. [2]Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA. [3]School of Control Science and Engineering, Shandong University, Jinan 250061, People's Republic of China. [4]Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, People's Republic of China.

References
1. K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341–2353 (2011)
2. C. Dong, C.C. Loy, K. He, et al., Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2014)
3. C. Dong, C.L. Chen, X. Tang, in *European Conference on Computer Vision*. Accelerating the super-resolution convolutional neural network (Springer, Cham, 2016a), pp. 391–407
4. Ledig C, Wang Z, Shi W, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network 2016 arXiv.org.
5. J. Sun, W. Cao, Z. Xu, et al., *Learning a Convolutional Neural Network for Non-Uniform Motion Blur Removal* (IEEE Conference on Computer Vision & Pattern Recognition, 2015), pp. 769–777
6. C.J. Schuler, M. Hirsch, S. Harmeling, et al., Learning to deblur. IEEE Trans. Pattern Anal. Mach. Intell. **38**(7), 1439–1451 (2016)
7. C. Yan, L. Li, C. Zhang, et al., Cross-modality bridging and knowledge transferring for image understanding. IEEE Trans. Multimedia (2019) https://doi.org/10.1109/TMM.2019.2903448
8. I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., Generative adversarial networks. Adv. Neural Inf. Proces. Syst. **3**, 2672–2680 (2014)
9. A. Foi, V. Katkovnik, K. Egiazarian, Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images. IEEE Trans. Image Process. **16**(5), 1395 (2007)
10. C. Dong, Y. Deng, C.L. Chen, et al., in *IEEE International Conference on Computer Vision*. Compression artifacts reduction by a deep convolutional network (IEEE International Conference on Computer Vision, 2016), pp. 576–584
11. Z. Wang, L. Ding, S. Chang, et al., in *IEEE Conference on Computer Vision & Pattern Recognition*. D3: Deep dual-domain based fast restoration of JPEG-compressed images (2016)
12. H. Chang, M.K. Ng, T. Zeng, Reducing artifacts in JPEG decompression via a learned dictionary. IEEE Trans. Signal Process. **62**(3), 718–728 (2014)
13. R. Rothe, R. Timofte, L. Van, in *IEEE International Conference on Image Processing*. Efficient regression priors for reducing image compression artifacts (IEEE International Conference on Image Processing, 2015), pp. 769–777
14. M.T. Wu, Wavelet transform based on Meyer algorithm for image edge and blocking artifact reduction. Inf. Sci. **474**, 125–135 (2019)
15. X. Wang, P. Zhang, Y. Zhang, et al., Deep Intensity Guidance Based Compression Artifacts Reduction for Depth Map. J. Vis. Commun. Image Represent. **57**, 234-242 (2018)
16. R. Shan, Z.S. Zhao, P.F. Chen, W.J. Liu, S.Y. Xiao, Y.H. Hou, Z. Wang, Network modeling and assessment of ecosystem health by a multi-population swarm optimized neural network ensemble. Appl. Sci. **6**, 175 (2016) doi:10.3390
17. Z.-G. Wang, Z.-S. Zhao, C.-S. Zhang, Incremental multiple instance outlier detection. Neural Comput. & Applic. **26**(4), 957–968 (2015)
18. K. He, G. Gkioxari, P. Dollar, et al., Mask R-CNN. IEEE Trans. Pattern Anal. Mach. Intell. (2017) https://doi.org/10.1109/TPAMI.2018.2844175
19. Y. Chenggang, X. Hongtao, C. Jianjun, et al., An effective Uyghur text detector for complex background images. IEEE Trans. Multimedia. **20**(12), 3389-3398 (2018) https://doi.org/10.1109/TMM.2018.2838320
20. A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks. Comput. Sci. (2015) https://arxiv.org/pdf/1511.06434.pdf
21. P. Isola, J. Zhu, T. Zhou, A.A. Efros, *Image-to-image translation with conditional adversarial networks, 2017* (IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, 2017), pp. 5967–5976
22. J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, in *IEEE International Conference on Computer Vision (ICCV)*. Unpaired image-to-image translation using cycle-consistent adversarial networks (2017)
23. D. Pathak, P. Krähenbühl, J. Donahue, et al., *Context encoders: Feature learning by Inpainting* (IEEE Conference on Computer Vision and Pattern Recognition, 2016), pp. 2536–2544
24. O. Ronneberger, P. Fischer, T. Brox, in *International Conference On Medical Image Computing And Computer-Assisted Intervention*. U-net: Convolutional networks for biomedical image segmentation (Springer, Cham, 2015), pp. 234–241
25. M. Everingham, L.V. Gool, C.K.I. Williams, et al., The Pascal Visual Object Classes (VOC) challenge. Int. J. Comput. Vis. **88**(2), 303–338 (2010)
26. Z. Wang, A.C. Bovik, H.R. Sheikh, et al., Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
27. H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik, *Live Image Quality Assessment Database Release 2* (2005)