Lightweight Model on Super-Resolution Image for Complete Model Copyright Protection

Bingyi Xie, Honghui Xu, Daehee Seo, and Zhipeng Cai, *Senior Member, IEEE*

***Abstract*—Deep learning-based techniques are broadly used in a variety of applications, which exhibit superior performance compared to traditional methods. One of mainstream topics in computer vision is the image super-resolution task. In recent deep learning neural networks, the number of parameters in each convolution layer has been increasing along with more layers and more feature maps, resulting in better image image super-resolution performance. However, this raises a problem in that all these neural networks require a significant amount of time and computational resource to train. It is not feasible to implement massive neural networks into these devices that have limited computational resources. Meanwhile, it is not a trivial thing to think about the complete model copyright protection. Therefore, there is a demand to find smaller networks that can perform well while achieving the protection of the original model's copyright. To address this problem, this paper proposes a lightweight model to replace the original complete model for image super-resolution. Finally, comprehensive experiments are conducted on multiple datasets to demonstrate the superiority of the proposed approach in generating super-resolution images even using lightweight neural network.**

I. INTRODUCTION

There are many big models for researches [1], [2], [3] that are open to public use and modification. However, there are copyright and privacy issues when we use and release such big and powerful neural network models. Other researchers can freely use and modify the model for their own purposes, and all information about the model is exposed to the public. To address the problem of copyright and privacy, we can build lightweight models that inherit from the big models. We can use lightweight techniques to modify the original network and release the lightweight models to the public. The lightweight models can achieve high performance compared to the original models while still being relatively small in size. Therefore, other scholars and organizations can use the lightweight models for their applications or research without any privacy or copyright problems.

The challenge of generating a high-resolution image from a low-resolution image is called super-resolution (SR). With the rapid development of computer vision, SR has received significant attention [4], [5]. For high-quality video transmission, super-resolution can rebuild old and low-resolution videos into new high-resolution videos [6]. In the surveillance area [7], [8], [9], due to the size of the camera sensor, the video and image quality can become an issue when replaying the camera footage. Super-resolution images can help improve the lower quality to a high standard resolution for better usage. Another essential research field that involves image super-resolution is the gaming industry [10]. Both NVIDIA [11] and AMD [12] have proposed their own image super-resolution technology to improve the quality and performance of the gaming experience, even on low-computation machines.

Deep convolutional neural networks have shown excellent performance on image super-resolution tasks. However, traditional CNNs require a massive number of parameters and high computation power to achieve high accuracy. For instance, ResNet-50 [2] has 23.9 million parameters, and VGG-16 [1] has an astounding 134.7 million parameters. It takes approximately 15 hours to train IMAGENET [13] using state-of-the-art GPUs from NVIDIA. Nonetheless, in our daily lives, many mobile devices such as cellphones and laptops do not have the resources to perform large deep neural networks. Over the recent years, many lightweight methods have been proposed to reduce the number of parameters and computation power required by traditional CNNs, while maintaining high accuracy compared to large deep neural networks. Examples of such methods include knowledge distillation [14], model compression [15], and ShuffleNet [16]. Inspired by these lightweight designs, in this paper, we propose our lightweight model specifically designed for image super-resolution while accomplishing the complete model copyright protection.

Our contributions are summarized as follows:

- We design a lightweight model is designed using depthwise and separable convolution techniques, which enables it to perform image super resolution tasks while utilizing only a fraction of parameters compared to other deep neural network models.
- To generate the super-resolution image, we employ Generative Adversarial Networks (GANs) in our proposed model. In order to further enhance the quality of the generated image, we replace the traditional Mean Squared Error (MSE) loss function with Wasserstein Distance, which has been shown to produce better results in GAN-based models.
- Through extensive experiments on real datasets, we demonstrate that our proposed lightweight image super-resolution model can achieve high performance compared to other state-of-the-art models with significantly fewer parameters.

The rest of this paper is organized as follows. The related works are briefly summarized in Section II. We elaborate the details of our lightweight analysis and our image super-

Corresponding Author: Zhipeng Cai
B. Xie, H. Xu and Z. Cai are with the Department of Computer Science, Georgia State University, Atlanta, GA, USA. D. Seo is with the Department of Computer Science, Sangmyung University, Seoul, South Korea
E-mail: {bxie2, hxu16}@student.gsu.edu, zcai@gsu.edu, daehseo@smu.ac.kr

resolution algorithm in Section III and Section IV, respectively. Then we conduct experiments on real datasets and analyze all the results in Section V. Finally, we end up with a conclusion in Section VI.

II. Related Works

The most recent works on image super-resolution and lightweight neural network models are categorized into the following section.

A. Image Super-Resolution

There has been extensive research on image super resolution. Traditional mathematical solutions were used to increase the resolution of images. For example, there is a well-known method called bicubic interpolation [17], [18]. The basic idea of bicubic interpolation is to upscale the image by updating each pixel with the nearest 16 pixels in the original image to calculate the new pixel, where each pixel has a different weight that is constantly updated to achieve the best result. Research has shown that bicubic interpolation can have better performance than nearest neighbor and bilinear methods.

With the development of convolutional neural networks, There has been significant research on using convolutional neural networks for image super-resolution. In 2014, [19] proposed the first use of a deep learning neural network for upscaling images (SRCNN). The authors first used bicubic interpolation to upscale the image to match the original size, then extracted patches and features from the low-resolution image. Moreover, non-linearly mapped the low-resolution feature maps to high-resolution feature maps and reconstructed the high-resolution image from these maps. In 2016, [20] proposed an improved approach based on their previous work with SRCNN to achieve faster image super-resolution, in which they eliminated the bicubic interpolation pre-processing step to achieve faster speed and used smaller kernels to reduce computation. The newly proposed model achieved better results with faster computation time. In the same year, [21] proposed the use of Generative Adversarial Networks (GAN) to recover high-resolution images, where the authors used ResNet to generate high-resolution images, and the discriminator compared the original image and generated image to update the model until two images could not be distinguished by the neural network.

B. Lightweight Model

Deep learning neural network (DNN) has been proven to be one of the most popular methods in computer science. With the progress of accuracy in DNN models [22], [23], the number of parameters in the network has also increased by a large amount. Therefore, lightweight models have been proposed to shrink the CNN model and reduce computation power and time [24]. In 2017, Google proposed MobileNets[25], which is an efficient CNN model for various computer vision tasks. MobileNets use depthwise separable convolution instead of traditional kernel multiplication [26], [27]. Specifically, they split the convolution into two parts: depthwise convolution and pointwise convolution. In this way, the calculation is addition instead of multiplication, which saves time and resources by a significant margin. In the same year, Zhang et al. [16] proposed a lightweight network called ShuffleNet, which is a very small and efficient CNN model that works on mobile devices. ShuffleNet uses group convolution, divides the input vector and kernels into several small groups, conducts convolution computation separately to reduce the parameters, and shuffles the channel to enhance the performance results. In 2020, Han et al. [28] proposed GhostNet. For GhostNet, the authors discovered that many similar feature maps produced by the convolution process could be generated by some cheap operations. To be specific, they first generate some intrinsic feature maps with normal convolution operation, and then use these intrinsic feature maps to generate the ghost feature maps with cheap addition operation. Consequently, GhostNet can achieve similar performance compared to the normal convolution process while saving time and computation consumption.

Our proposed mechanism aims to achieve high performance while using significantly fewer parameters compared to other state-of-the-art image super-resolution models. To achieve this, we combine the advantages of image super-resolution and lightweight method to build our proposed model for image super-resolution tasks. In the end, extensive experiments on real datasets will be conducted to demonstrate the effectiveness of our approach.

III. Lightweight Analysis

Deep learning has succeeded in many industrial and research fields. These neural networks are powerful in many different tasks, but they require tremendous computing power to train the models. However, strong computers with powerful capabilities are not always available in realistic scenarios. With the fast development of mobile internet, there is a rising demand to perform tasks such as image classification and segmentation on smaller devices with low computation power and resources. These tasks need to deliver great performance despite the limitations. To address this issue, MobileNets that is a much smaller neural network was proposed to be implemented on low-computation devices. This proposed lightweight network is different from the original convolutional networks as it uses a depthwise separable convolution network, which is composed of two parts, including depthwise convolution and pointwise convolution.

In traditional convolution layer, there is an input $D_F * D_F * M$ with M channels of input size F, and it combines with filter $D_K * D_K * N$ to produce the output feature map $D_G * D_G * N$, D_G is the width and height of a output feature map, D_K is the size of kernel, and N is the number of output layer or channel. The output feature map of a traditional convolution layer with stride one and padding have the computational cost:

$$D_F * D_F * M * N * D_K * D_K \tag{1}$$

And the parameter number of a standard convolution layer will be:

$$D_K * D_K * N * M \tag{2}$$

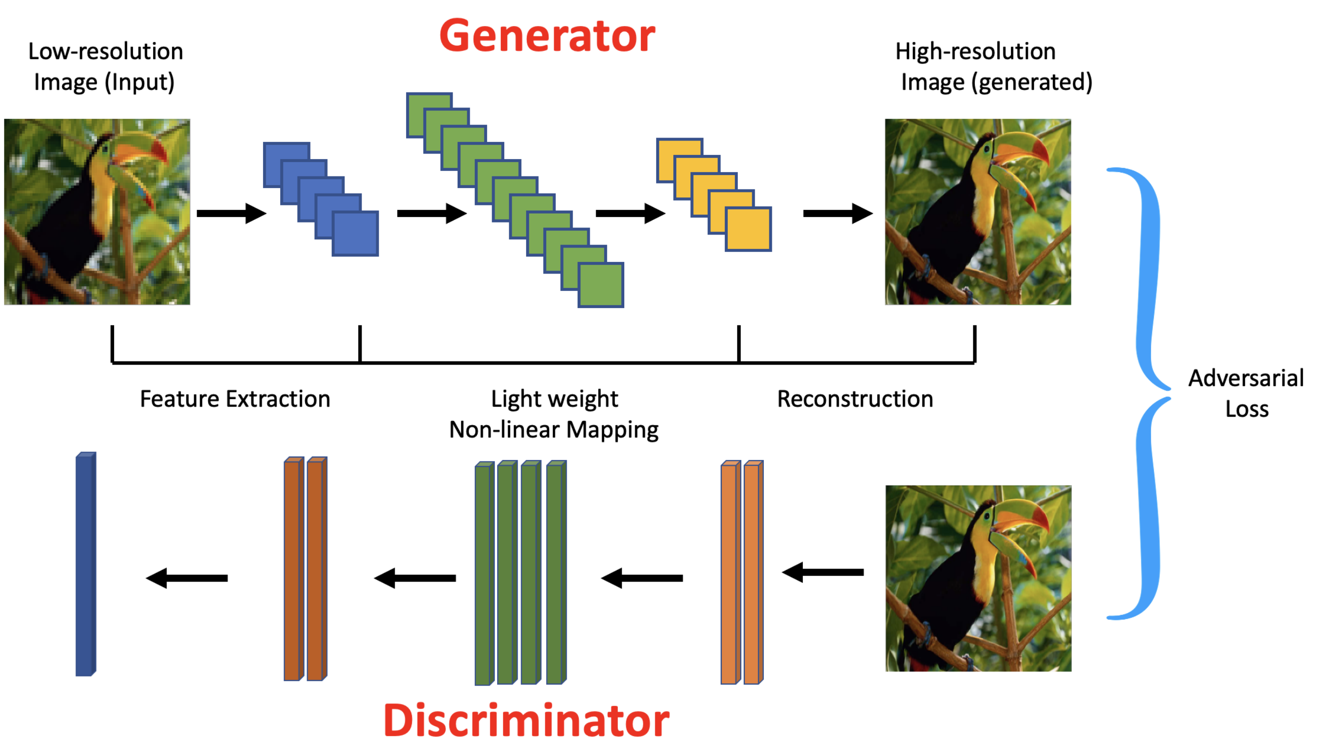


Fig. 1: The architecture of the proposed lightweight super-resolution mechanism.

The computational cost and number of parameters highly depend on the kernel size D_K and the number of layers N. However, standard convolution mainly uses multiplication with a large number of filters, making the parameters increasingly large for some neural networks.

In contrast, MobileNets divides the traditional convolution into two separate parts. The first part is depthwise convolution, where only one single filter is applied for each input channel during the convolution process. The second part of MobileNets is the pointwise convolution, which applies a $1 * 1$ filter to combine the outputs of the depthwise convolution. The depthwise separable convolution splits traditional convolution into two layers: one layer for filtering and an additional separate layer for combining. This convolution can drastically reduce computation and model size.

The first layer of depthwise separable convolution is the depthwise convolution. This layer applies a single filter for every input channel. After the convolution, there are the same number of feature maps as the input channel with size $D_F * D_F * M$. The depthwise convolution is very efficient compared to traditional convolution, but it only filters the input channel. Therefore, there is a second layer to combine the output of depthwise convolution to create new feature maps, which is the pointwise convolution. This layer applies a simple $1 * 1$ convolution to the output of the depthwise layer and combines it with a linear operation.

With the same input and kernel size of standard convolution $D_F * D_F * M$ and $D_K * D_K$, the depthwise separable convolution has a computation cost:

$$D_F * D_F * M * D_K * D_K + M * N * D_F * D_F \tag{3}$$

Also the parameter amount of the depthwise separable convolution is:

$$D_K * D_K * N + N * M \tag{4}$$

By comparing with the standard convolution and depthwise convolution, the reduction ratio in computational cost is shown in the following equation:

$$\begin{aligned} &\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} \\ =&\frac{1}{N} + \frac{1}{(D_K)^2} \end{aligned} \tag{5}$$

We use MobileNet's idea to design our lightweight image super-resolution model, in which we use a $3 * 3$ kernel in our depthwise separable convolution. Therefore, according to Eq. (5), we can obtain between 8 to 9 times less computation cost than traditional convolution, as well as a smaller amount of parameters in one convolution layer.

IV. Methodology

In this section, we will introduce the technical details of the proposed lightweight image super-resolution model and explain how to generate high-resolution images using this mechanism. Our super-resolution model is built on neural networks by combining the advantages of lightweight neural network methods and generative adversarial network methods, which have excellent ability to upscale low-resolution images while maintaining low computation costs. In addition, the design idea of Generative Adversarial Networks (GANs) [29] is adopted for generating high resolution images. A GAN model consists of two "adversarial" modules, including a generator and a discriminator, in which these two adversarial modules compete with each other in a min-max game to update themselves alternatively, where the generator tries to generate fake data to fool the discriminator, while the discriminator tries to discriminate whether its input is real or fake. The framework of this image super-resolution generative mechanism is shown in Fig. 1, where the lightweight image generation network P works as the generator in GANs and the critic network C works as the discriminator in GANs.

The input of the generator network P is the low resolution image x. During the generation process, the low resolution sample x is passed through and a high resolution image $P_\theta(x)$ is generated, where θ represents the parameter of the generative network. Next, both the generated image $P_\theta(x)$ and the original image y are inputs to the discriminator network C_ϕ, where ϕ denotes the parameters of the discriminator network. Then, both generative network parameters θ and discriminator network parameters ϕ can be optimized to minimize the adversarial training loss that is described below.

Adversarial Loss: Instead of using the traditional loss function based on Mean Squared Error MSE), we use Wasserstein Distance [30] to calculate the adversarial loss for helping train a more robust generative network. The Wasserstein Distance is defined as

$$W\left(\mathbb{P}_r, \mathbb{P}_\theta\right)=\sup_{\|f\|_L \leq 1} \mathbb{E}_{x \sim \mathbb{P}_r}[f(x)]-\mathbb{E}_{x \sim \mathbb{P}_\theta}[f(x)] \quad (6)$$

where θ is the parameter drawn from the generative network, and ϕ is the parameter from discriminator network. The Wasserstein distance is one step ahead of the earth move distance, the effort to move one distribution to another distribution. In our method, we want to min max the distance between generated high resolution image and original image. Then, our adversarial loss L_{adv} is defined as the Internal Wasserstein Distance between original image y and $P_\theta(x)$:

$$L_{adv}=W(y, P_\theta(x)). \quad (7)$$

Besides the adversarial loss, to generate high resolution images closer to the original images, we one additional loss, pixel loss, to facilitate training of the image generation mechanism.

Pixel Loss: Pixel loss, defined in Eq. (8), is the distance between original sample y and generated high resolution sample $P_\theta(x)$. Taking into account pixel loss L_{pix} into the training process can help smooth the image and mitigate the adversarial effect.

$$L_{pix}=\left\|P_\theta\left(x\right)-x\right\|_2 \quad (8)$$

L_2 norm is used to calculate the pixel loss when we require that the average pixel difference is below a threshold.

The overall loss of our mechanism is defined in Eq. (9) by integrating the aforementioned three loss functions, including L_{adv} in Eq. (7), L_{pix}, and Eq. (8)

$$L_{P_\theta}=\alpha * L_{adv}+\beta * L_{pix}, \quad (9)$$

where α, and β are the weights of three loss terms. We train the super-resolution mechanism by minimizing the overall loss and training procedure of the generative mechanism can be summarized as Algorithm 1.

Algorithm 1: Super-Resolution Mechanism

Input: Low-resolution image x, original image y, lightweight image super-resolution network P_θ, critic network C_ϕ, and the number of iteration T.

Output: P_θ

for $i=1$ to T **do**
  Forward pass x to the generative network P_θ
  Forward pass $P_\theta(x')$ to discriminate network C_ϕ
  Calculate overall loss L_{P_θ} as Eq. (9)
  Optimize θ and ϕ by minimize L_{P_θ}
end for
return P_θ

V. Experiments and Results

In this section, we first introduce our experimental settings, and then carry out comprehensive experiments to quantitatively and qualitatively compare our super-resolution method with the state-of-the-art models.

A. Experiment Settings

Our experiments were implemented and tested on an Ubuntu 16 operating system with a Tesla V100 GPU. First, we used a convolution network to increase the image channels four times and simply upscaled the low-resolution images. Then, to demonstrate the performance capability of our model, we trained and tested it using various datasets in different categories, including five benchmark datasets: Set5, Set14, General100, BSDS100, and Manga109.

1) Baseline: We compare our proposed super-resolution lightweight model with four different baseline mechanisms. First, the traditional Bicubic [18] mechanism simply uses mathematical techniques to upscale the images. To further justify the effectiveness of our super-resolution lightweight method, we use three mainstream convolutional neural network models for comparison. SRCNN [19] is the first machine learning model to solve the image super-resolution problem. And SRResNet [21] uses the well-known Resnet deep learning neural network to increase the performance of generating high-resolution images. Additionally, SRGAN [21] is another the state-of-the-art mechanism in the deep learning field for generating and recovering high-resolution images.

2) Metrics: In this paper, we use two different metrics to measure the performance of transforming low-resolution images to high-resolution images.

The first metric used is Peak Signal-to-Noise Ratio, also known as PSNR. PSNR [31] is one of the most common metrics for measuring the performance of two different resolution images. It is the ratio of the mean square error (MSE) of two images and the maximum pixel value in the original image. The higher the PSNR value, the less distortion in the image and the closer it is to the original image.

The second metric we use is the structural similarity index (SSIM). SSIM[32] is used to measure the similarity of two images based on three aspects: brightness, contrast, and structure. The SSIM score ranges between [0, 1], where a score of

1 means two images are exactly the same with no difference. Therefore, a higher SSIM score closer to 1 indicates better performance on super-resolution models.

B. Quantitative Evaluation

Image Super-Resolution task: We generate high-resolution images with an upscale ratio of 4, which means we quadruple the original image resolution to achieve the super-resolution task. We compare our results with four different image-super-resolution baseline models. Table I contains only 5 kinds of images in this dataset, and every model achieves great performance. The lowest one is the original Bicubic model, which only uses pure mathematical methods to perform the image super-resolution. Therefore, in both PSNR and SSIM metrics, it has relatively low performance compared to other neural network models. In terms of other machine learning image super-resolution models, our lightweight method achieves the second-highest performance in the SSIM metric but falls behind in the PSNR metric. Specifically, compared to the highest model, SRGAN, the performance of our lightweight model is only 8.3% lower in PSNR, and only 0.1% lower in terms of SSIM. From Table II-Table V, we can obtain the same conclusion that the Bicubic method has the lowest performance among all the models, and our lightweight method has a comparable performance with other SOTA models.

TABLE I: Performance on the Set5 Dataset

	Bicubic	SRCNN	SRResNet	SRGAN	**Ours**
PSNR	23.116	30.26	28.66	30.8	28.24
SSIM	0.697	0.861	0.801	0.874	0.864

TABLE II: Performance on the Set14 Dataset

	Bicubic	SRCNN	SRResNet	SRGAN	**Ours**
PSNR	21.84	29.54	26.74	27.7	26.63
SSIM	0.598	0.703	0.723	0.762	0.752

TABLE III: Performance on the General100 Dataset

	Bicubic	SRCNN	SRResNet	SRGAN	**Ours**
PSNR	23.157	27.96	26.97	29.8	28.04
SSIM	0.696	0.785	0.751	0.840	0.827

TABLE IV: Performance on the BSDS100 Dataset

	Bicubic	SRCNN	SRResNet	SRGAN	**Ours**
PSNR	22.72	29.603	25.89	26.95	26.39
SSIM	0.585	0.671	0.672	0.720	0.713

Model Parameters: As mentioned in Section III, we proposed a lightweight convolutional neural network built into the GAN model. We replaced all traditional convolution calculations with depthwise separable convolutions to reduce model parameters and run time. In deep neural networks, more parameters and layers often bring better overall performance. In our experiments, this situation is the same: the SRCNN and SRGAN contain many more parameters compared to our lightweight model and achieve slightly better results. However, our lightweight image super-resolution model only uses 1/8 of the parameters used in SRGAN and still obtain comparable performance. In Table VI, we can find that our lightweight model has significantly fewer parameters compared to SRResNet and SRGAN. Overall, our lightweight image super-resolution model can use lower cost to achieve a great super-resolution performance.

TABLE V: Performance on the Manga109 Dataset

	Bicubic	SRCNN	SRResNet	SRGAN	**Ours**
PSNR	20.194	28.473	23.51	28.81	27.15
SSIM	0.689	0.779	0.673	0.884	0.870

TABLE VI: Parameters of Image Super-Resolution Models

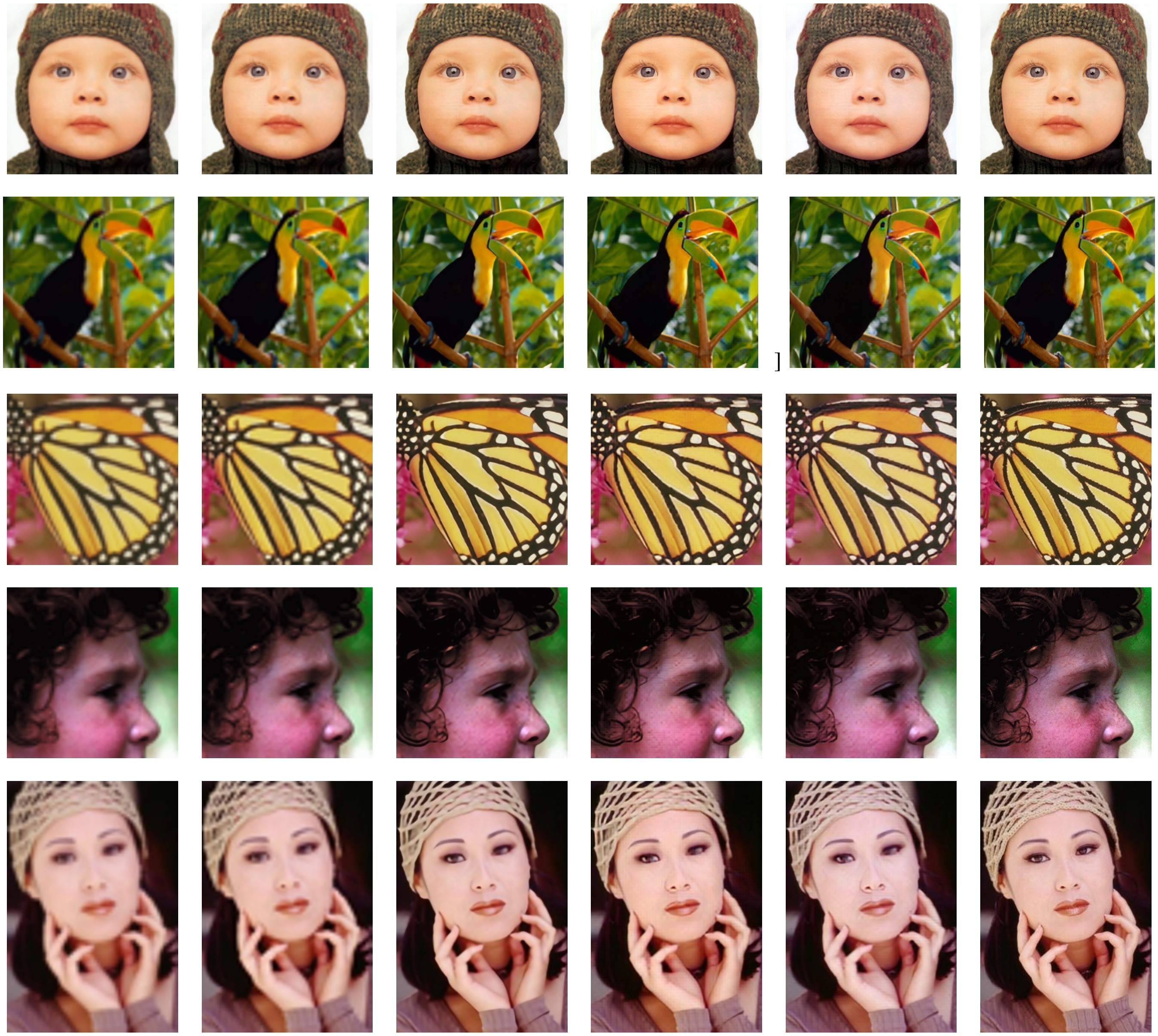
Models	Bicubic	SRCNN	SRResNet	SRGAN	**Ours**
Parameter		67552	7.1m	5.2m	0.58m

C. Qualitative Evaluation

Qualitative evaluation aims to compare the quality of super-resolution images through intuitive observations. With fewer parameters compared to other models, we selected five sample images from the testing dataset to evaluate the quality of the super-resolution image results. For each sample, an original low-resolution image, the corresponding super-resolution images in the baseline mechanisms, and a groundtruth super-resolution image are listed in Fig. 2. Take the baby images as examples, compared with the original image and Bicubic method, the generated image from our lightweight mechanism is more clear. Also, our generated super-resolution image is as clear as the images generated by other SOTA models. Ultimately, our generated super-resolution image in Fig. 2(e) can not be easily distinguish from Fig. 2(f). This result indicates that our mechanism has great image super-resolution performance while maintaining low parameter numbers and computation time and cost. Similar conclusions can be drawn from the comparison results of other image samples, shown in Fig. 2.

VI. Conclusion

This paper proposes a lightweight image super-resolution mechanism by integrating a depthwise separable convolution method with a generative adversarial network. To achieve better image quality, we use the Wasserstein distance to calculate the adversarial training loss in the formulation of the overall loss function in our proposed model. In the end, we conduct comprehensive quantitative and qualitative evaluations



(a) Low Resolution (b) Bicubic (c) SRResNet (d) SRGAN (e) **Ours** (f) High Resolution

Fig. 2: Qualitative Evaluation on Super-Resolution Image

to demonstrate the effectiveness of our proposed mechanism. The quantitative evaluation shows that our mechanism achieves competitive performance in terms of PSNR and SSIM metrics, and even outperforms some state-of-the-art methods in certain datasets and metrics. And the qualitative evaluation shows that our mechanism can generate high-quality super-resolution images with more details and less distortion. Overall speaking, our proposed lightweight mechanism can achieve great performance while maintaining lower computational cost and resource consumption, which is very suitable for practical applications that require real-time or on-device super-resolution image processing and satisfies the requirement of the complete model copyright protection.

Acknowledgment

This work was partly supported by the National Science Foundation of U.S. (1704287, 1829674, 1912753, and 2011845).

References

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[4] Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3365–3387, 2020.

[5] S. Anwar, S. Khan, and N. Barnes, “A deep journey into super-resolution: A survey,” *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–34, 2020.

[6] A. Kappeler, S. Yoo, Q. Dai, and A. K. Katsaggelos, “Video super-resolution with convolutional neural networks,” *IEEE transactions on computational imaging*, vol. 2, no. 2, pp. 109–122, 2016.

[7] L. Zhang, H. Zhang, H. Shen, and P. Li, “A super-resolution reconstruction algorithm for surveillance images,” *Signal Processing*, vol. 90, no. 3, pp. 848–859, 2010.

[8] Y. Pang, J. Cao, J. Wang, and J. Han, “Jcs-net: Joint classification and super-resolution network for small-scale pedestrian detection in surveillance images,” *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 12, pp. 3322–3331, 2019.

[9] P. Rasti, T. Uiboupin, S. Escalera, and G. Anbarjafari, “Convolutional neural network super resolution for face recognition in surveillance monitoring,” in *Articulated Motion and Deformable Objects: 9th International Conference, AMDO 2016, Palma de Mallorca, Spain, July 13-15, 2016, Proceedings 9*. Springer, 2016, pp. 175–184.

[10] A. Watson, “Deep learning techniques for super-resolution in video games,” *arXiv preprint arXiv:2012.09810*, 2020.

[11] E. Liu, “Dlss 2.0-image reconstruction for real-time rendering with deep learning,” in *GPU Technology Conference (GTC)*, 2020.

[12] J. Gu, H. Cai, C. Dong, R. Zhang, Y. Zhang, W. Yang, and C. Yuan, “Super-resolution by predicting offsets: An ultra-efficient super-resolution network for rasterized images,” in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX*. Springer, 2022, pp. 583–598.

[13] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[14] J. Gou, B. Yu, S. J. Maybank, and D. Tao, “Knowledge distillation: A survey,” *International Journal of Computer Vision*, vol. 129, pp. 1789–1819, 2021.

[15] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, “A survey of model compression and acceleration for deep neural networks,” *arXiv preprint arXiv:1710.09282*, 2017.

[16] X. Zhang, X. Zhou, M. Lin, and J. Sun, “Shufflenet: An extremely efficient convolutional neural network for mobile devices,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6848–6856.

[17] H. Chang, D.-Y. Yeung, and Y. Xiong, “Super-resolution through neighbor embedding,” in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1. IEEE, 2004, pp. I–I.

[18] A. Lukin, A. S. Krylov, and A. Nasonov, “Image interpolation by super-resolution,” in *Proceedings of GraphiCon*, vol. 2006, no. Citeseer. Citeseer, 2006, pp. 239–242.

[19] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[20] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. Springer, 2016, pp. 391–407.

[21] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[22] L. Beyer, O. J. Hénaff, A. Kolesnikov, X. Zhai, and A. v. d. Oord, “Are we done with imagenet?” *arXiv preprint arXiv:2006.07159*, 2020.

[23] D. Tsipras, S. Santurkar, L. Engstrom, A. Ilyas, and A. Madry, “From imagenet to image classification: Contextualizing progress on benchmarks,” in *International Conference on Machine Learning*. PMLR, 2020, pp. 9625–9635.

[24] M. Denil, B. Shakibi, L. Dinh, M. Ranzato, and N. De Freitas, “Predicting parameters in deep learning,” *Advances in neural information processing systems*, vol. 26, 2013.

[25] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017.

[26] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[27] L. Kaiser, A. N. Gomez, and F. Chollet, “Depthwise separable convolutions for neural machine translation,” *arXiv preprint arXiv:1706.03059*, 2017.

[28] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, “Ghostnet: More features from cheap operations,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1580–1589.

[29] Z. Cai, Z. Xiong, H. Xu, P. Wang, W. Li, and Y. Pan, “Generative adversarial networks: A survey toward private and secure applications,” *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–38, 2021.

[30] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, “Improved training of wasserstein gans,” *Advances in neural information processing systems*, vol. 30, 2017.

[31] S. Winkler and P. Mohandas, “The evolution of video quality measurement: From psnr to hybrid metrics,” *IEEE transactions on Broadcasting*, vol. 54, no. 3, pp. 660–668, 2008.

[32] U. Sara, M. Akter, and M. S. Uddin, “Image quality assessment through fsim, ssim, mse and psnr—a comparative study,” *Journal of Computer and Communications*, vol. 7, no. 3, pp. 8–18, 2019.