TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214
DOI: 10.26599/TST.2023.9010082
Volume xx, Number x, xxxxxxx 20xx

Lightweight Super-Resolution Model for Complete Model Copyright Protection

Bingyi Xie, Honghui Xu, YongJoon Joe, Daehee Seo and Zhipeng Cai*

Abstract: Deep learning-based techniques are broadly used in various applications, which exhibit superior performance compared to traditional methods. One of the 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 increased along with more layers and feature maps, resulting in better image super-resolution performance. In today's era, numerous service providers offer super-resolution services to users, providing them with remarkable convenience. However, the availability of open-source super-resolution services exposes service providers to the risk of copyright infringement, as the complete model could be vulnerable to leakage. Therefore, safeguarding the copyright of the complete model is a non-trivial concern. To tackle this issue, this paper presents a lightweight model as a substitute for the original complete model in image super-resolution. This research has identified smaller networks that can deliver impressive performance while protecting the original model's copyright. 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 networks.

Key words: lightweight; copyright protection; adversarial learning, image super-resolution

1 Introduction

There are many big models and big data analytic researches[1–4] that are open to public use and modification. However, there are copyright and privacy issues[5–8] when we use and release such prominent and powerful neural network models. Other researchers can freely use and modify the model for their purposes, and all information about the model is exposed to the public[9–12]. Some scholars use knowledge distillation strategies, such as model compression, and semi-supervised knowledge distillation, to address the problem. The idea of knowledge distillation is to use the teacher network as a reference, the students learning from this teacher network can achieve comparable performance as the teacher while preventing the teacher network from copyright leakage. Inspired by this idea, we propose another method to address the copyright problem. Specifically, we build a lightweight model that inherits from the original complete model, which helps hide the hyperparameter setting and the model structure in the teacher network. Then, we can use lightweight techniques to modify the original network and release the lightweight models to the public. In other words, other scholars and organizations can use lightweight models for their applications or research without any privacy or copyright problems.

The challenge of generating a high-resolution image

• B. Xie, H. Xu, and Z. Cai* are with the Department of Computer Science, Georgia State University, Atlanta, GA, USA, 30303. E-mails: bxie2@student.gsu.edu; hxu16@student.gsu.edu; zcai@gsu.edu.
• Y. Joe is with LSWare Inc., Director, Republic of Korea, 08504. E-mail: research@yongjoon.net.
• D. Seo is with the College of Intelligence Information Engineering, Sangmyung University, Republic of Korea, 03016. E-mail: daehseo@smu.ac.kr.
* To whom correspondence should be addressed.
Manuscript received: 2023-Aug-02; accepted: 2023-Aug-04

from a low-resolution image is called super-resolution. With the rapid development of computer vision, super-resolution has received significant attention[13, 14]. For high-quality video transmission, super-resolution can rebuild old and low-resolution videos into new high-resolution videos[15]. In the surveillance area[16–18], the video and image quality can become an issue when replaying the camera footage due to the size of the camera sensor. 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[19]. Both NVIDIA[20] and AMD[21] 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 Convolutional Neural Networks (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[22] using state-of-the-art GPUs from NVIDIA. Nonetheless, in our daily lives, many mobile devices such as cell phones 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[23], model compression[24], and ShuffleNet[25]. Inspired by these lightweight designs, in this paper, we propose our copyright-protected model specifically designed for image super-resolution while accomplishing complete model copyright protection.

Our contributions are summarized as follows:

- We have created a lightweight model using depth-wise and separable convolution techniques for image super-resolution, which performs well while utilizing significantly fewer parameters compared to other deep neural network models, thereby protecting the copyright of the original model.
- To improve the performance of super-resolution, we employ a Generative Adversarial Network (GAN) 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 2. We elaborate on the details of our lightweight analysis and our image super-resolution algorithm in Section 3 and Section 4, respectively. Then we conduct experiments on real datasets and analyze all the results in Section 5. Finally, we end up with a conclusion in Section 7.

2 Related Works

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

2.1 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[26, 27]. 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 2015, Dong et al. [28] 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, they [29] 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, Ledig et al. [30] proposed the use of Generative Adversarial Networks (GANs) 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.

2.2 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[31,32], 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[33]. In 2017, Google proposed MobileNets[34], which is an efficient CNN model for various computer vision tasks. MobileNets use depthwise separable convolution instead of traditional kernel multiplication[35,36]. 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.[25] 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.[37] proposed GhostNet. For GhostNet, the authors discovered that some cheap operations could generate many similar feature maps produced by the convolution process. Specifically, 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 leverages the concept of lightweight model design to effectively protect the copyright of the original super-resolution model. We accomplish this by combining the strengths of an image super-resolution model with a lightweight approach, resulting in our innovative model for image super-resolution tasks. Furthermore, we will conduct thorough experiments on real datasets to demonstrate the efficacy and effectiveness of our approach.

3 Lightweight Design

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 the 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 which 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 the 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 an output feature map, D_K is the size of the 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 has the computational cost:

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

The parameter number of a standard convolution layer will be:

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

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 many 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:

$$D_F * D_F * M * D_K * D_K + M * N * D_F * D_F \quad (3)$$

Also, the parameter amount of the depthwise separable convolution is:

$$D_K * D_K * N + N * M \quad (4)$$

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

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

Our lightweight image super-resolution model is designed based on the principles of MobileNet, incorporating a $3 * 3$ kernel in our depthwise separable convolution. By doing so, this approach also results in a reduced number of parameters in each convolution layer as indicated by Eq. (5). By utilizing fewer parameters, we can effectively protect the copyright of the parameters in the original image super-resolution model.

4 Copyright-Protected Super Resolution

In this section, we will introduce the technical details of the proposed copyright-protected 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 an excellent ability to upscale low-resolution images while protecting the copyright of the original super-resolution model. In addition, the design idea of GAN[38–40] 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 copyright-protected image super-resolution mechanism is shown in Fig. 1, where the copyright-protected generation network P works as the generator in GAN, and the critic network C works as the discriminator in GAN.

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 the 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[41] to calculate the adversarial loss to help 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's movement distance, the effort to move one distribution to another distribution. In our method, we want to min-max the distance between the generated high-resolution image and the original image. Then, our adversarial loss L_{adv} is defined as the Internal Wasserstein Distance between original image y and

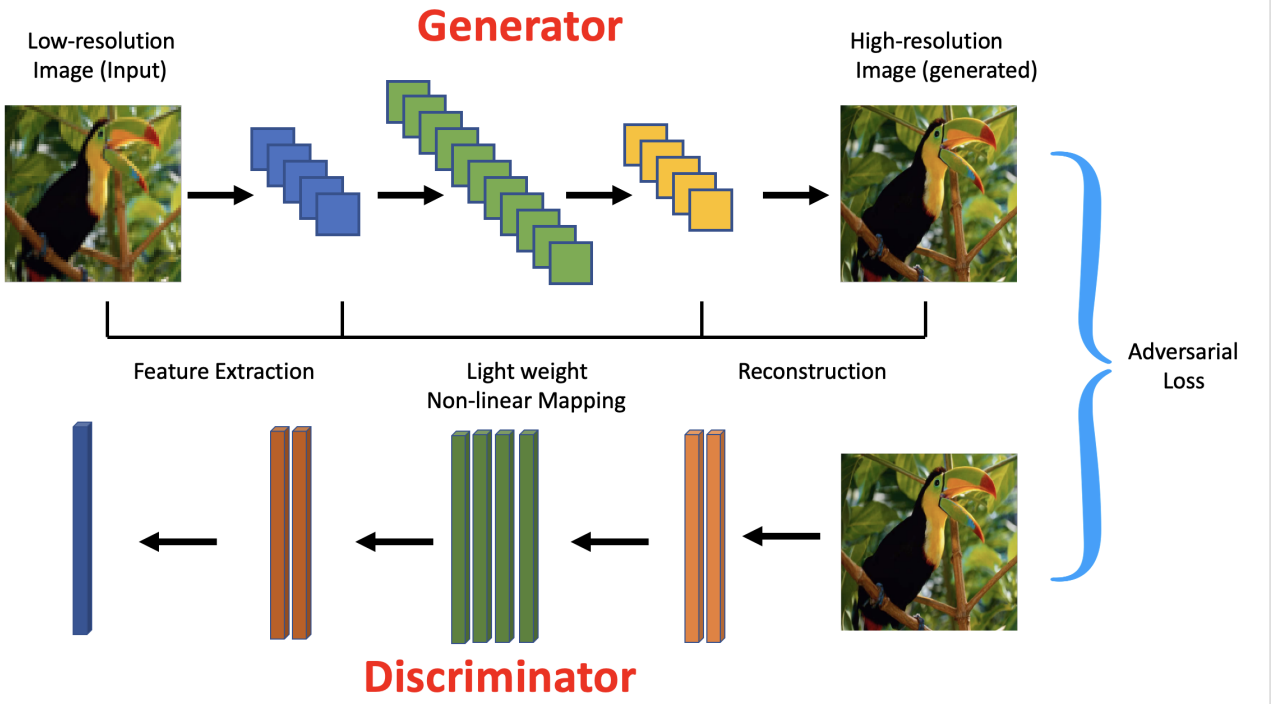


Fig. 1 The architecture of the proposed copyright-protected super-resolution mechanism.

$P_\theta(x)$:

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

Besides the adversarial loss, to generate high-resolution images closer to the original images, we have 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 the 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.

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

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)

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

where α, and β are the weights of two loss terms. We train the super-resolution mechanism by minimizing the overall loss and the training procedure of the generative mechanism can be summarized as Algorithm 1. All notations used in this paper are listed in Table 1.

Algorithm 1 Super-Resolution Mechanism

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

Output: P_θ

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

5 Experiments and Results

In this section, we introduce our experimental settings, carry out comprehensive experiments to quantitatively and qualitatively compare our super-resolution method with the state-of-the-art models, and illustrate the effectiveness of the proposed copyright-protected super-resolution model.

5.1 Experiment Settings

Our experiments were implemented and tested on an Ubuntu 16 with a Tesla V100 GPU. More details of the experimental setting are described below.

Table 1 List of notations used in this paper

Symbol	Description
(D_F)	input size
(M)	input channel
(D_K)	kernel size
(D_G)	output feature map size
(N)	output channel
(θ)	generative network parameter
(P_θ)	copyright-protected super-resolution network
(C_ϕ)	critic network

5.1.1 Datasets

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. **Set5** dataset contains five typical images for the image super-resolution model which are "baby", "bird", "butterfly", "head", and "woman". **Set14** dataset is another commonly used image dataset to test the performance of image super-resolution models, which is an extended version of the Set5 dataset from 5 images to 14 images including more variation and diversity. **General100** dataset contains 100 BMP format images with no compression. The size of the 100 images ranges from $131 * 112$(small) to $710 * 704$ (large). **BSDS100** is a dataset created by Berkeley University and used frequently for segmentation and image super-resolution, which is the testing set of the Berkeley segmentation dataset BSD300 and contains 100 test images composed of a large variety of images ranging from natural images to object-specific such as plants, people, food, etc. **Manga109** dataset has been compiled by the Aizawa Yamasaki Matsui Laboratory at the University of Tokyo, which consists of 109 manga volumes drawn by professional manga artists in Japan and it is popularly used for image super-resolution.

5.1.2 Hyperparameter Settings

First, we implement an upscale factor of 4, which means increasing the resolution of images 4 times. We set batch size as 2, which means that we group every two images as a single batch goes into our model to process the images, and we set the learning rate for our model as $1e^-4$ for training. We train our proposed model with 10 epochs to get the optimized results. For Each dataset training, we use 200 images training images, 200 images validation images, and 100 testing images.

Table 2 Parameters of Image Super-Resolution Models

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

Table 3 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 4 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

5.1.3 Baseline

We compare our proposed copyright-protected super-resolution model with four different baseline mechanisms. First, the traditional Bicubic[26] 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[28] is the first machine-learning model to solve the image super-resolution problem. SRResNet[29] uses the well-known Resnet deep learning neural network to increase the performance of generating high-resolution images. Additionally, SRGAN[30] is another state-of-the-art mechanism in the deep learning field for generating and recovering high-resolution images.

5.1.4 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 the Peak Signal-to-Noise Ratio, also known as PSNR. PSNR [42,43] 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.

Table 5 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 6 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

The second metric we use is the structural similarity index (SSIM). SSIM[44] 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.

5.2 Ablation Study

We conduct experiments with different epochs. Normally, higher iterations mean higher accuracy. In this work, we desire to set a proper epoch in our model that can not only bring us relatively high performance but also keep the training time and cost low. In Fig.2(a) and Fig.2(b), we can see that our lightweight image super-resolution model with PSNR and SSIM metric by varying three different epochs, 10, 20, and 30. From these results, it can be found that the performance metrics do not increase consistently when the epochs are over 20. Considering the time cost to train with more epochs, we choose 10 epochs in this paper.

Moreover, we conduct experiments to gradually increase the lightweight neural network layers in our GAN-based model to investigate the relationship between the number of lightweight layers and image super-resolution performance. In Fig.3(a) and Fig.3(b), we test our proposed copyright-protected model with different lightweight layers. 0 means that we do not use any lightweight layer, and we add one lightweight layer into our model each time to test the PSNR and SSIM performance. Through this ablation study, the PSNR and SSIM of image super-resolution decrease with the increase of lightweight layers. To sum up, the more lightweight the model is, the more performance the model will have. However, we can obtain a comparable performance of the baseline when using lightweight layers as little as possible.

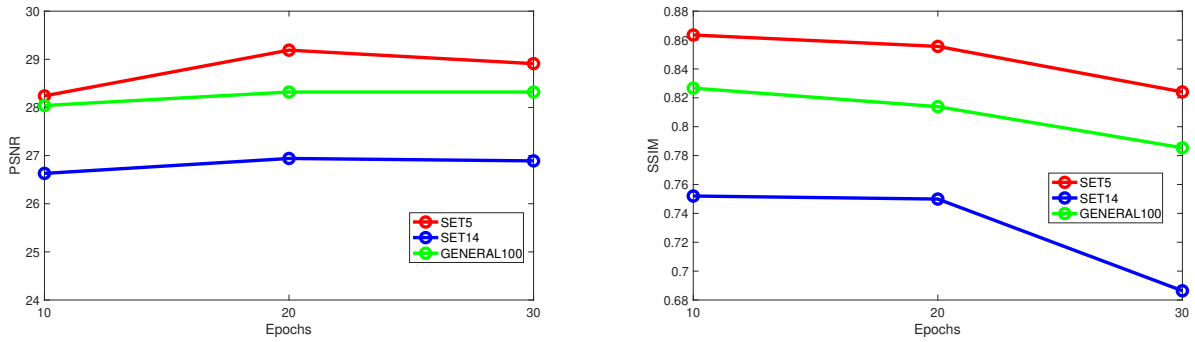
Table 7 Performance on the Manga109 Dataset

	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

5.3 Quantitative Evaluation

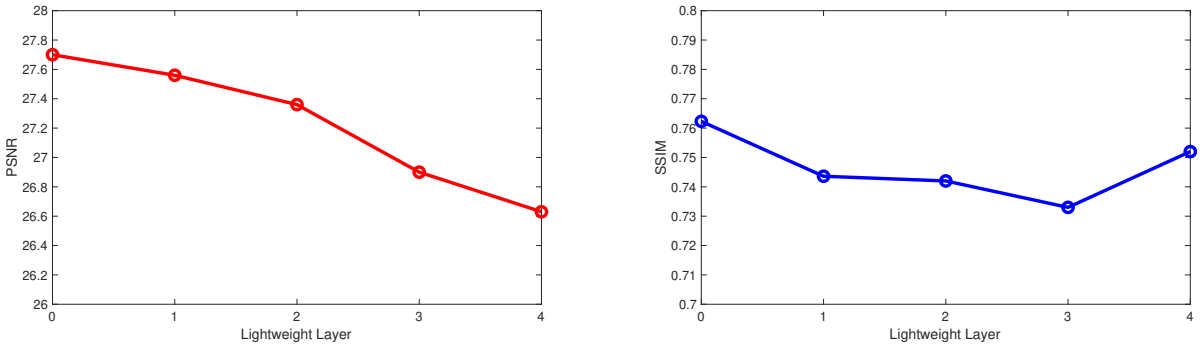
Effectiveness of Copyright Protection: As mentioned in Section 3, we integrate a lightweight convolutional neural network built with a GAN model. Specifically, we replaced all traditional convolution calculations with depthwise separable convolutions to reduce model parameters and run time. In deep neural 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 copyright-protected super-resolution model only uses 1/8 of the parameters used in SRGAN and still obtains comparable performance. In Table 2, we can find that our proposed model has significantly fewer parameters compared to SRResNet and SRGAN such that the copyright of the original super-resolution model can be protected.

Image Super-Resolution Performance: 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 3 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 proposed copyright-protected 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 proposed copyright-protected model is only 8.3% lower in PSNR, and only 0.1% lower in terms of SSIM. From Table 4-Table 7, we can obtain the same



(a) Impact of Epochs on PSNR (b) Impact of Epochs on SSIM

Fig. 2 Impact of Epochs on Our Proposed Copyright-Protected Model



(a) Impact of the lightweight layers on PSNR (b) Impact of the lightweight layers on SSIM

Fig. 3 Impact of the Lightweight Layers on Performance of Our Proposed Model

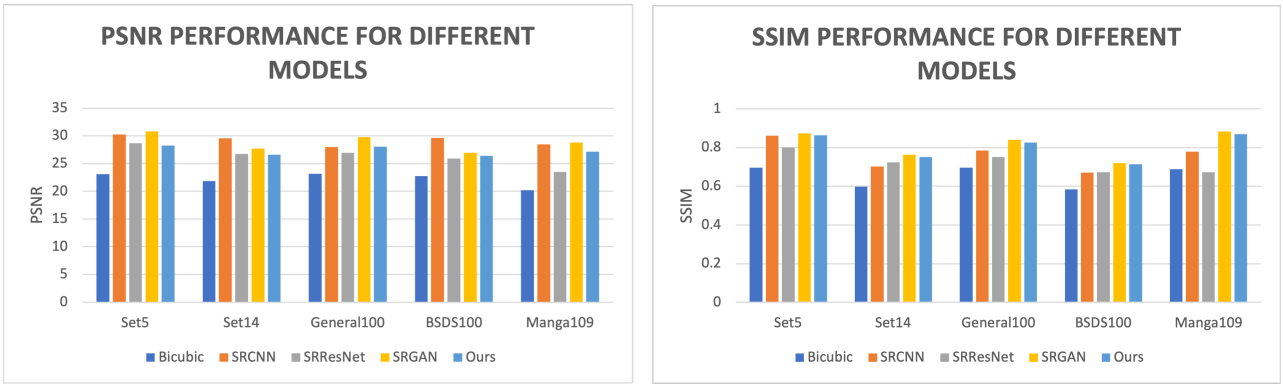
conclusion that the Bicubic method has the lowest performance among all the models, and our proposed method has a comparable performance with other SOTA models.

What's more, we use Fig. 4 to more explicitly demonstrate the effectiveness of our proposed copyright-protection model by comparing it with other baseline models. From Fig. 4(a), we can see that our model has comparable performance with SRCNN, SRResNet, and SRGAN on five different datasets. In some scenarios, our proposed model can bypass the SRResNet model even by only using $1/8$ of the parameter amount. Besides, by observing Fig. 4(b), the performance of our model overperforms the SRCNN and SRResNet models for all five testing datasets, which is only lower than the SRGAN model due to the utilization of lightweight architecture in our proposed model.

Moreover, we draw Fig. 5 to show the average performance for both PSNR and SSIM metrics to further illustrate our proposed model's effectiveness. These average metrics demonstrate that our model can achieve relatively high performance, and reach a comparable performance compared with the state-of-the-art neural network models even using our designed lightweight architecture. In other words, these results prove that our model can achieve a higher super-resolution performance while protecting the original architecture of models (*i.e.*, using a lightweight architecture). One thing we need to point out is that the performance for SSIM metrics is slightly better compared to the PSNR metric with other baseline models. This situation is caused by the calculation of these two metrics. The PSNR is defined by MSE, which is calculated by every pixel difference between two images. While, the computation of SSIM not only considers the image pixel difference but also takes into account the brightness, contrast, and structure of images. In all, SSIM is better than PSNR due to its more comprehensive information used in the similarity measurement.

5.4 Qualitative Evaluation

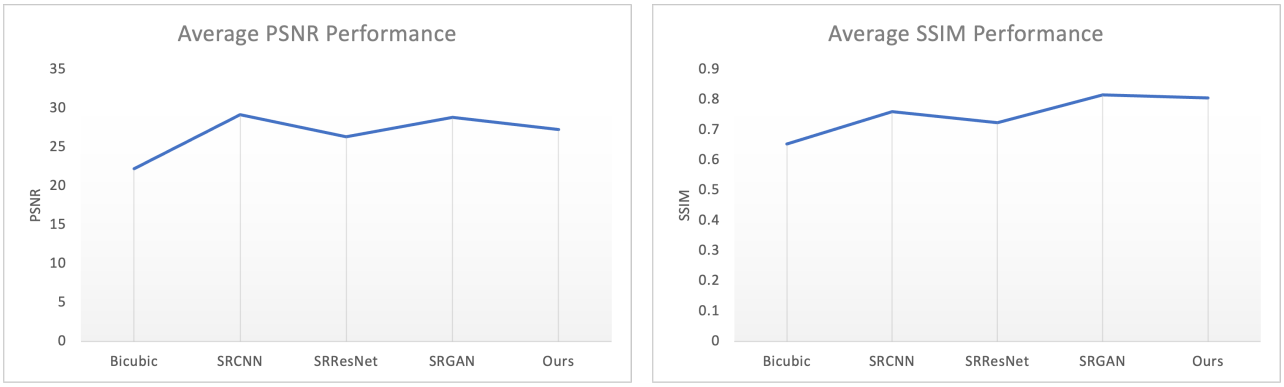
The qualitative evaluation aims to compare the quality of super-resolution images through intuitive observations. With fewer parameters than other models,



(a) PSNR Results on Different Datasets

(b) SSIM Results on Different Datasets

Fig. 4 Performance Comparison Results (Baselines v.s. Ours)



(a) Average PSNR Results on Different Datasets

(b) Average SSIM Results on Different Datasets

Fig. 5 Average Performance Comparison Results (Baselines v.s. Ours)

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

Fig. 6 Qualitative Evaluation on Super-Resolution Image

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 ground truth super-resolution image are listed in Fig. 6. Take the baby images as examples, compared with the original image and the Bicubic method, the generated image from our lightweight mechanism is more apparent. 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. 6(e) can not be easily distinguished from Fig. 6(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. 6.

6 Future Discussion

(1) Through our ablation study, we proved that increasing lightweight layers results in lower accuracy in two different metrics, which indicates that there is a trade-off between model performance and the number of parameters. Therefore, how to balance the performance of the model and the privacy protection degree of the model is also a promising research direction.

(2) From the performance results, we can conclude that our copyright-protected model has different performance results on different datasets. It is important to investigate this situation from the viewpoint of improving model generalization to make sure that our proposed model can consistently have the same performance when applied to different datasets.

(3) Although our copyright-protected model can achieve relatively high performance on image super-resolution tasks and has been demonstrated that our model can upscale the image from a more economical perspective. With the similarity between image processing and video processing, we believe our model can be implemented with a video with more complex and continuous events. Because Video is made by multiple images within a short time, it is possible to super-resolution a complete video by upscaling each frame.

(4) In our copyright protection study, the copyright issue is an important subject of how to protect the original neural network while keeping the performance of super-resolution at an acceptable rate. Our lightweight model is a successful example to show the lightweight network can be feasible to process multiple tasks without leaking any information for the original network. The model distillation topic which investigates the relationship between the teacher network and the student network can be studied in the future to achieve the copyright protection of the original model.

(5) In this work, we successfully achieve the copyright protection of the deep learning model by manually changing the original architecture to a more lightweight model architecture. The neural Architecture Search (NAS) technique may be another promising way to accomplish the copyright protection of the deep learning model by automatically finding a different model architecture while maintaining the model's performance.

7 Conclusion

This paper proposes a copyright-protected super-resolution model 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 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, our proposed copyright-protected mechanism can achieve great performance with 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 complete model copyright protection.

Acknowledgment

This research was supported by SW Copyright Ecosystem R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports, and Tourism in 2023. Project Name: Development of Large-Scale Software License Verification Technology by Cloud Service Utilization and Construction Type (No. RS-2023-00224818).

References

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

[2] He, K., Zhang, X., Ren, S., & Sun, J., Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, (2016).

[3] Krizhevsky, A., Sutskever, I., & Hinton, G. E., Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90, (2017).

[4] Nti, I. K., Quarcoo, J. A., Aning, J., & Fosu, G. K.. A mini-review of machine learning in big data analytics: Applications, challenges, and prospects. Big Data Mining and Analytics, 5(2), 81-97 (2022).

[5] Zhang, K., Tian, Z., Cai, Z., & Seo, D., Link-privacy preserving graph embedding data publication with adversarial learning. Tsinghua Science and Technology, 27(2), 244-256 (2022).

[6] Li, K., Tian, L., Zheng, X., & Hui, B., Plausible Heterogeneous Graph k-Anonymization for Social Networks. Tsinghua Science and Technology, 27(6), 912-924 (2022).

[7] Cai, Z., He, Z, Guan, X., & Li, Y., Collective Data-Sanitization for Preventing Sensitive Information Inference Attacks in Social Networks. IEEE Transactions on Dependable and Secure Computing. 15(4): 577-590 (2018)

[8] Huang, Y., Li, Y., & Cai, Z., Security and Privacy in Metaverse: A Comprehensive Survey Journal of Big Data Mining and Analytics (BDMA). 6(2): 234-247, (2023).

[9] Cai, Z. & Zheng X., A Private and Efficient Mechanism for Data Uploading in Smart Cyber-Physical Systems. IEEE Transactions on Network Science and Engineering (TNSE). 7(2): 766-775, (2020).

[10] Cai, Z., Zheng, X., Wang, J., & He, Z., Private Data Trading Towards Range Counting Queries in Internet of Things. IEEE Transactions on Mobile Computing (TMC). Accepted.

[11] Zheng, X., & Cai, Z., Privacy-Preserved Data Sharing towards Multiple Parties in Industrial IoTs. IEEE Journal on Selected Areas in Communications (JSAC). 38(5): 968-979 (2020).

[12] Liang Y., Cai, Z., Yu, J., Han, Q., & Li, Y., Deep Learning Based Inference of Private Information Using Embedded Sensors in Smart Devices. IEEE Network Magazine. 32(4): 8-14 (2018).

[13] Wang, Z., Chen, J., & Hoi, S. C., Deep learning for image super-resolution: A survey. IEEE Transactions on pattern analysis and machine intelligence, 43(10), 3365-3387, (2020).

[14] Anwar, S., Khan, S., & Barnes, N., A deep journey into super-resolution: A survey. ACM Computing Surveys (CSUR), 53(3), 1-34, (2020).

[15] Kappeler, A., Yoo, S., Dai, Q., & Katsaggelos, A. K., Video super-resolution with convolutional neural networks. IEEE Transactions on Computational Imaging, 2(2), 109-122, (2016).

[16] Zhang, L., Zhang, H., Shen, H., & Li, P., A super-resolution reconstruction algorithm for surveillance images. Signal Processing, 90(3), 848-859, (2010).

[17] Pang, Y., Cao, J., Wang, J., & Han, J., JCS-Net: Joint classification and super-resolution network for small-scale pedestrian detection in surveillance images. IEEE Transactions on Information Forensics and Security, 14(12), 3322-3331 (2019).

[18] Rasti, P., Uiboupin, T., Escalera, S., & Anbarjafari, G., 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, pp. 175-184 (2016).

[19] Watson, A., Deep learning techniques for super-resolution in video games. arXiv preprint arXiv:2012.09810, (2020).

[20] Liu, E., DLSS 2.0-Image reconstruction for real-time rendering with deep learning. In GPU Technology Conference (GTC) (2020).

[21] Gu, J., Cai, H., Dong, C., Zhang, R., Zhang, Y., Yang, W., & Yuan, C., 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).

[22] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L., Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255, (2009).

[23] Gou, J., Yu, B., Maybank, S. J., & Tao, D., Knowledge distillation: A survey. International Journal of Computer Vision, 129, 1789-1819, (2021).

[24] Cheng, Y., Wang, D., Zhou, P., & Zhang, T., A survey of model compression and acceleration for deep neural networks. arXiv preprint arXiv:1710.09282, (2017).

[25] Zhang, X., Zhou, X., Lin, M., & Sun, J., Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6848-6856, (2018).

[26] Chang, H., Yeung, D. Y., & Xiong, Y., Super-resolution through neighbor embedding. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. I-I, (2004).

[27] Lukin, A., Krylov, A. S., & Nasonov, A., Image interpolation by super-resolution. In Proceedings of GraphiCon, Vol. 2006, No. Citeseer, pp. 239-242, (2006).

[28] Dong, C., Loy, C. C., He, K., & Tang, X., Image super-resolution using deep convolutional networks. IEEE Transactions on pattern analysis and machine intelligence, 38(2), 295-307, (2015).

[29] Dong, C., Loy, C. C., & Tang, X., 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, pp. 391-407, (2016).

[30] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W., Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4681-4690, (2017).

[31] Beyer, L., Hénaff, O. J., Kolesnikov, A., Zhai, X., & Oord, A. V. D., Are we done with imagenet?. arXiv preprint arXiv:2006.07159, (2020).

[32] Tsipras, D., Santurkar, S., Engstrom, L., Ilyas, A., & Madry, A., From imagenet to image classification: Contextualizing progress on benchmarks. In International Conference on Machine Learning, pp. 9625-9635, (2020).

[33] Denil, M., Shakibi, B., Dinh, L., Ranzato, M. A., & De Freitas, N., Predicting parameters in deep learning. Advances in neural information processing systems, 26, (2013).

[34] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H., Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, (2017).

[35] Chollet, F., Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251-125, (2017).

[36] Kaiser, L., Gomez, A. N., & Chollet, F., Depthwise separable convolutions for neural machine translation. arXiv preprint arXiv:1706.03059, (2017).

[37] Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C., Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 1580-1589, (2020).

[38] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y., Generative adversarial networks. Communications of the ACM, 63(11), 139-144, (2020).

[39] Wang, R., Jiang, B., Yang, C., Li, Q., & Zhang, B.. MAGAN: Unsupervised low-light image enhancement guided by mixed-attention. Big Data Mining and Analytics, 5(2), 110-119 (2022).

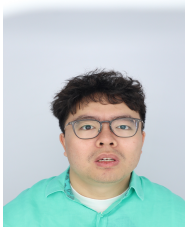
[40] Cai, Z., Xiong, Z., Xu, H., Wang, P., Li, W., & Pan, Y., Generative adversarial networks: A survey toward private and secure applications. ACM Computing Surveys (CSUR), 54(6), 1-38, (2021).

[41] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C., Improved training of wasserstein gans. Advances in neural information processing systems, 30, (2017).

[42] Winkler, S., & Mohandas, P., The evolution of video quality measurement: From PSNR to hybrid metrics. IEEE transactions on Broadcasting, 54(3), 660-668 (2008).

[43] Huynh-Thu, Q., & Ghanbari, M., Scope of validity of PSNR in image/video quality assessment. Electronics letters, 44(13), 800-801 (2008).

[44] Sara, U., Akter, M., & Uddin, M. S., Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. Journal of Computer and Communications, 7(3), 8-18 (2019).



Bingyi Xie received a B.S. degree in economics from Indiana University of Bloomington, Bloomington, IN, USA in 2017, and an M.S. degree in project management from Northeastern University, Boston, MA, USA in 2019. He is currently pursuing a Ph.D. degree with the Department of Computer Science, at Georgia State University, Atlanta, GA, USA. His research interests include privacy preservation, computer vision, and machine learning.

Honghui Xu is a Ph.D. student in the Department of Computer Science at Georgia State University (GSU). He received a Bachelor's degree from the University of Electronic Science and Technology of China (UESTC) in 2019. His research focuses on machine learning and deep learning, including the fundamental theory of machine learning, the applications of deep learning in the computer vision field, and the topic of privacy-preserving machine learning.

YongJoon Joe was born in Seoul, South Korea, on March 23rd, 1987. He received his B.S. and M.S. degrees in information science from Kyushu University, Japan, in 2011 and 2013, respectively. Since 2016, he has been working at LSware Inc until the present. His research interests include game theory, copyright, blockchain, and concurrent/parallel simulation computing.

Daehee Seo (Member, IEEE) received a B.S. degree in electronic and electrical engineering from Dongshin University, Naju, South Korea, in February 2001. He also obtained an M.S. degree in computer science and engineering and a Ph.D. degree in computer science from Soonchunhyang University, South Korea, in February 2003 and February 2006, respectively. Currently, he serves as an Assistant Professor at the Faculty of College of Intelligence

Information Engineering, Sangmyung University (SMU), Seoul, South Korea.

Zhipeng Cai received his M.S. degree in 2004 and Ph.D. degree in 2008, both in the Department of Computing Science at the University of Alberta, and his B.S. degree from Beijing Institute of Technology. He is currently a Professor in the Department of Computer Science at Georgia State University, and also an Affiliate Professor in the Department of Computer Information Systems at Robinson College of Business, as well as an Associate Director at INSPIRE Center. Dr. Cai's research expertise lies in the areas of Resource Management and scheduling, High-Performance Computing, Privacy, Networking, and Big Data. Dr. Cai is the recipient of an NSF CAREER Award.