

PathSRGAN: Multi-supervised super-resolution for cytopathological images using generative adversarial network

Jiabo Ma, Student *Member, IEEE*, Jingya Yu, Student *Member, IEEE*, Sibbo Liu, Student *Member, IEEE*, Li Chen, Xu Li, Jie Feng, Zhixing Chen, Shaoqun Zeng, Xiuli Liu, Shenghua Cheng

Abstract— In the cytopathology screening of cervical cancer, high-resolution digital cytopathological slides are critical for the interpretation of lesion cells. However, the acquisition of high-resolution digital slides requires high-end imaging equipment and long scanning time. In the study, we propose a GAN-based progressive multi-supervised super-resolution model called PathSRGAN (pathology super-resolution GAN) to learn the mapping of real low-resolution and high-resolution cytopathological images. With respect to the characteristics of cytopathological images, we design a new two-stage generator architecture with two supervision terms. The generator of the first stage corresponds to a densely-connected U-Net and achieves 4× to 10× super resolution. The generator of the second stage corresponds to a residual-in-residual DenseBlock and achieves 10× to 20× super resolution. The designed generator alleviates the difficulty in learning the mapping from 4× images to 20× images caused by the great numerical aperture difference and generates high quality high-resolution images. We conduct a series of comparison experiments and demonstrate the superiority of PathSRGAN to mainstream CNN-based and GAN-based super-resolution methods in cytopathological images. Simultaneously, the reconstructed high-resolution images by PathSRGAN improve the accuracy of computer-aided diagnosis tasks effectively. It is anticipated that the study will help increase the penetration rate of cytopathology screening in remote and impoverished areas that lack high-end imaging equipment.

Index Terms— Cervical cancer, cytopathological images, generative adversarial learning, super resolution

I. INTRODUCTION

Cervical cancer is the third most common cancer among women worldwide. Approximately, 560 thousand women were diagnosed with cervical cancer each year and 310 thousand of them died of the disease [1]. Therefore, cervical cancer screening is critical for early detection and death prevention death. Papanicolaou test and liquid-based cytology test (LCT) are the main methods of cervical cancer screening. Since the traditional manual screening is subjective and time-consuming, computer-aided diagnosis is gradually developed [2], [3].

Given the development of LCT and scanning system, the acquisition of a large number of digital slides has become a reality. Digital slides are convenient to transfer, store and analyze. These advantages have promoted the development of computer-aided diagnosis systems. High-resolution digital slides with more details provide a good diagnostic basis for pathologists and computer-aided algorithms. However, imaging high-resolution digital slides requires expensive equipment, long scanning time and a high amount of storage space when compared to imaging low-resolution digital slides. To solve this problem, single-image super-resolution (SISR) can be used to convert low-resolution slides to high-resolution slides. The super-resolution microscopy technique has large field of view and fast scanning speed. At the same time, it can improve image quality, reduce storage space, speed up transmission and provide more details to improve diagnostic accuracy. In the field of medical images, super resolution is used to improve the contrast and image quality of computed tomography (CT), magnetic resonance imaging (MRI), ultrasound and other medical images, thereby improving the effect of automatic detection and segmentation [4]-[6]. In the field of microscopy images, super resolution is utilized to improve clarity of biological structures in histopathological slides and fluorescence microscopy images [7], [8]. However, there are few studies addressing the aforementioned challenges in the field of cytopathological slide images.

SISR is an important computer vision problem with a long history. Early super-resolution methods include interpolation methods such as bicubic interpolation and Lanczos interpolation [9]. Although the methods are fast, reconstructed high-resolution images by interpolation are limited. To further improve the super-resolution effect, learning-based super-resolution methods were developed, including statistical image prior [10], [11], neighbor embedding methods [13], [14], random forest [15], dictionary learning, and sparse representation [16]-[19]. Recently, deep learning-based super-resolution technology has developed rapidly. From the convolutional neural network

Copyright (c) 2019 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

S. Cheng and X. Liu both are corresponding authors.

J. Ma and J. Yu contributed equally to this work.

J. Ma, J. Yu, S. Liu, X. Li, J. Feng, Z. Chen, S. Zeng, X. Liu and S. Cheng are with Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics-Huazhong University of Science and Technology,

Wuhan, Hubei 430074, China and MoE Key Laboratory for Biomedical Photonics, School of Engineering Sciences, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China. (email: majiabo@hust.edu.cn; jingyayu@hust.edu.cn; liam.liusibo@gmail.com; goodlixu@163.com; jifeng@hust.edu.cn; zhixing_chen0123@163.com; sqzeng@mail.hust.edu.cn; xliu@mail.hust.edu.cn; chengshen@hust.edu.cn)

L. Chen is with Department of Clinical Laboratory, Tongji Hospital, Huazhong University of Science and Technology, Wuhan, Hubei 430030, China. (email: chenliisme@126.com)

(CNN) based methods [20]-[25] to generative adversarial network (GAN) based methods [26]-[33], extant studies expected to obtain better super-resolution effect via designing network structures, loss functions, learning strategies and other aspects.

In most super-resolution models, low-resolution (LR) images are generated via the degradation of corresponding high-resolution (HR) images. Degraded LR images are essentially identical to HR images in terms of optical resolution and depth of field (DOF) except pixel size. Additionally, degraded LR and HR images are already registered. However, real LR slide images tend to be more complex. Real LR images exhibit worse optical resolution and greater DOF than degraded LR images. The real LR slides are more blurred and include more content. Thus, super resolution of the real LR slide images is difficult when compared with degraded LR slide images. Hence, super-resolution models for degraded images are often difficult to be applied on real LR images. Figure 1 shows the challenge. Furthermore, strict registration of LR and HR images is required, and this adds difficulties to the super resolution of cytopathological slide images.

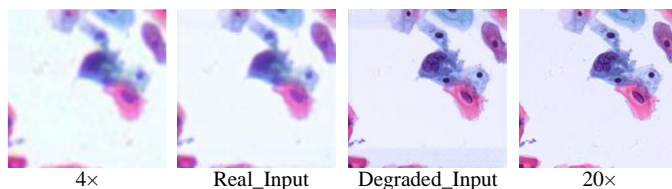


Fig. 1. Degradation model performance on a real image. 4 \times corresponds to the target image that we wish to process, and 20 \times denotes the corresponding HR image. The SRGAN model is trained on degraded images. Real_Input and Degraded_Input are reconstructed 20 \times images by inputting 4 \times image and degraded image, respectively.

To address the aforementioned challenges, some super-resolution microscopy methods were proposed [33], [34]. These methods use registered LR-HR image pairs to train generative models and perform well on microscopy images. However, the methods may face challenges for high scale factor super resolution on cytopathology slides with big numerical aperture (NA) difference (4 \times /0.1-NA, 20 \times /0.75-NA). Here, we propose a GAN-based progressive multi-supervised super-resolution model called PathSRGAN (pathology super-resolution GAN) to learn the mapping of real LR and HR cytopathology slide images. With respect to the characteristics of cytopathological images, we design a new two-stage generator structure with two supervision terms. The generator of the first stage corresponds to a densely-connected U-Net and achieves 4 \times to 10 \times super resolution by supervision of real 10 \times images. The generator of the second stage corresponds to a residual-in-residual DenseBlock and achieves 10 \times to 20 \times super resolution by supervision of 20 \times images. The designed generator alleviates the difficulty in learning the mapping from 4 \times images to 20 \times images caused by the great NA difference and generates high quality HR images. Additionally, we develop a two-step registration method with sub pixel accuracy, which is important for PathSRGAN. We conduct a series of comparison experiments and demonstrate the superiority of PathSRGAN to mainstream CNN-based and GAN-based super resolution methods in cytopathology images. Simultaneously, the reconstructed HR images by PathSRGAN

effectively improve the accuracy of computer-aided diagnosis tasks.

The main contributions of our study are as follows. Firstly, we propose a progressive multi-supervised super-resolution method for cytopathological slide images. The method is proven to be effective and superior to current mainstream super-resolution methods. Secondly, we design a new two-stage generator with two supervision terms for the characteristics of cytopathological images. The performance of the new generator exceeds that of the classic U-Net and SRResNet generators. Additionally, the generator outputs reconstructed images with two different resolutions. Thirdly, we demonstrate that the reconstructed HR images obtained via the proposed method improve the accuracy of various computer-aided diagnostic tasks. Thus, the proposed method effectively reduces the cost of data acquisition and storage. This is important for cervical cancer screening in remote and impoverished areas that lack high-end instruments.

II. RELATED WORKS

It is known that SISR is always a very active research field in computer vision. Extant studies attempted a variety of methods to optimize super-resolution effects and made significant progress. The related works in this section mainly focus on the super-resolution methods based on CNN [20]-[25] and GAN [26]-[33], which are similar to our study.

A. CNN-based Super Resolution

Traditional sparse coding based super-resolution methods are proven to be effective although they depend on feature engineering. Dong et al. [20] first proposed using CNN to learn non-linear mapping in image space to achieve super-resolution (SRCNN). The method surpasses the bicubic interpolation method and sparse coding-based methods without any feature engineering. Kim et al. [21] proposed VDSR via modifying the reconstruction network architecture of SRCNN. In the method, super resolution is obtained via deeper convolutional networks with larger receptive field when compared to SRCNN. Lim et al. [25] proposed an enhanced deep super-resolution network (EDSR) via removing unnecessary modules, such as batch normalization layers to improve computational efficiency, and developing a new multi-scale deep super-resolution system (MDSR) to reconstruct HR images of different sizes in a single model. LapSRN [22] and CARN [23] modified network architecture as well as VDSR to make network deeper and wider. LapSRN used the Laplacian pyramid to predict high-frequency residuals in a coarse-to-fine manner, thereby decreasing artifacts and computation complexity. CARN used a cascaded residual network to achieve efficient information flow and gradient propagation, thereby alleviating the problem of gradient disappearance and achieving better results with fewer parameters.

CNN-based super-resolution methods are end-to-end mappings between LR images and HR images via hidden layers and achieve better super resolution than traditional methods. However, the minimization of the L1 or L2 loss in CNN-based methods tends to generate smoother images that lack high frequencies.

B. GAN-based Super Resolution

To produce more realistic texture details of the reconstructed images, Ledig et al. [27] proposed SRGAN. SRGAN used adversarial loss to improve high frequency information in reconstructed images and content loss motivated by perceptual similarity to achieve realistic texture reconstruction. To obtain better performance, ESRGAN [35] modified SRGAN including removing batch normalization, using a residual-in-residual DenseBlock as generator and employing RaGAN [36] as discriminator. ProGanSR [30] adopted progressive up-sampling architecture, progressive training methods, and the asymmetrical pyramid network. The strategies enable the model to obtain high reconstruction quality and superior computing speed in super-resolution tasks with high upscaling factors.

All the aforementioned methods correspond to supervised super-resolution methods, i.e., needing registered HR images as supervision. Thus, Yuan et al. [31] proposed Cycle-in-Cycle GAN using two CycleGAN [32] structures to achieve unsupervised image super-resolution. A CycleGAN maps the LR space to noise-free LR space. Another CycleGAN further completes the super-resolution work.

GAN-based super-resolution methods recover the texture details better by using adversarial loss than CNN-based methods. However, most of them focus on natural images, and they are trained on degraded images. The super-resolution models based on degraded images face challenges in performing well on real biomedical LR images because the point spread function of real LR images is unknown. Wang [33] used a registration method to obtain paired real LR and HR images and achieved well super-resolution on fluorescence microscopy images using GAN with a U-Net generator. Wang's work indicates that learning a direct mapping from real LR images to HR images is feasible. However, with respect to super resolution from $4\times/0.1\text{-NA}$ objective lens cytopathological images to $20\times/0.75\text{-NA}$ images, the differences in DOF and optical resolution significantly exceed that of Wang's $10\times/0.4\text{-NA}$ to $20\times/0.75\text{-NA}$.

C. Differences with Other Works

To address the challenge, we developed a multi-supervised super-resolution method for cytopathological slide images using GAN. We designed an image-to-image generative model to directly learn the mapping from real LR to HR cytopathological images. Most of the aforementioned super-resolution methods learn mappings based on degraded images. In the terms of method architecture, we designed a new two-stage generator with two supervisions and two discriminators. The generator of the first stage corresponds to a densely-connected U-Net with DenseBlocks, achieving $4\times/0.1\text{-NA}$ to $10\times/0.25\text{-NA}$ super resolution. The generator of the second stage is a residual-in-residual DenseBlock, which achieves $10\times/0.25\text{-NA}$ to $20\times/0.75\text{-NA}$ super resolution. SRGAN and Wang's method are both one-step upscaling with one supervision. Our method's generator is different from SRGAN's SRResNet and Wang's U-Net with residual blocks. ProGanSR used a progressive multi-supervised pyramidal architecture for high scale factor reconstruction. The difference between our method and ProGanSR except generator architecture is that our method directly reconstructs images of

different resolutions while ProGanSR learns interpolation functions at different resolutions. In terms of the loss functions, our method introduced cytopathology perceptual loss apart from common mean square error (MSE) loss and adversarial loss.

III. METHODS

A. Problem Setting

The aim of the study is to obtain reconstructed medium-resolution images I_R^{MR} and high-resolution I_R^{HR} from input low-resolution images I^{LR} . To facilitate the explanation of PathSR-GAN, we term the module that outputs I_R^{MR} as $G_1: I^{LR} \rightarrow I_R^{MR}$ and the module that outputs I_R^{HR} as $G_2: I_R^{MR} \rightarrow I_R^{HR}$. Similarly, discriminators corresponding to G_1 and G_2 are referred to as D_1 and D_2 , respectively. Besides, we design a registration method to obtain matched LR, MR, and HR images.

B. Image Registration

Since our super resolution (SR) method is supervised, strictly registered images are necessary. However, slide images obtained from different objective lenses are rarely fully registered. To solve the problem, we propose a two-step registration scheme based on correlation. The two-step registration strategy is developed for two reasons. Firstly, the local offsets of two slides are inconsistent, thus it is difficult to obtain an accurate global offset to register two slides. Secondly, a significant amount of computation is involved in directly performing local registration for every patch owing to the large search range. In short, we coarsely register two digital slides to obtain the relative coarse global offset to narrow the search range for further registration. Additionally, local registration is used to improve registration accuracy. We use $4\times$ to $10\times$ as an example to illustrate the registration method. The registration method for $10\times$ and $20\times$ is similar. Notably, cytopathological slides of different lenses only exhibit non-global offsets in the vertical and horizontal directions without distortion or rotation.

Coarse Registration

Coarse registration aids in narrowing the search range, and Figure 2 shows the coarse registration. First, we randomly take n patches (P_{4-1} to P_{4-n}) on $4\times$ slide S_4 . The size of P_{4-1} is determined based on the size of S_4 . Generally, increasing the size of P_{4-1} causes the registration result to be more stable and reliable although the amount of calculation increases. We map P_{4-1} to $10\times$ slide S_{10} , set redundancy as Δr and consider the region P_{10-1} corresponding to P_{4-1} . The redundancy Δr exhibits a decisive effect on the registration. The matching may fail if Δr is excessively small. If Δr is excessively large, it increases the amount of calculation. Subsequently, we slide P_{4-1} on P_{10-1} and obtain the offset $\Delta x_1, \Delta y_1$ of the two patches when the correlation is maximum. We repeat the same operation for the other selected patches. All the relative offsets obtained are averaged as $E_{\Delta x}, E_{\Delta y}$. The average value is considered as the coarse global offset between S_4 and S_{10} , which is the result of coarse matching. Additionally, we interpolate $4\times$ slides to $5\times$ to avoid float numbers in patch size. Unless otherwise specified, the $4\times$ slides refer to the interpolated $5\times$ slides in this study.

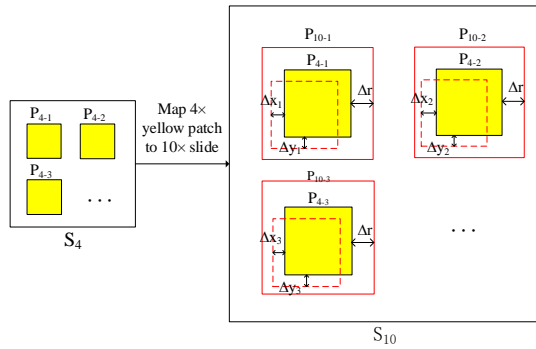


Fig. 2. Coarse registration of 4 \times slide and 10 \times slide. P_{s-x} denotes the x^{th} patch on slide S . In the figure, the yellow box is a sliding template, the red solid line frame denotes the sliding area, and the red dashed line frame denotes the matched area corresponding to the yellow box.

Fine Registration

After coarse registration, a less accurate global offset between S_4 and S_{10} is obtained. We use the offset to decrease the search range and conduct local registration to obtain accurate paired images. The registration accuracy significantly affects the subsequent super resolution. Hence, we control registration accuracy by setting correlation threshold T of matched images. As shown in Figure 3, with respect to a patch P_4 on 4 \times slides, we consider a corresponding patch P_{10} on 10 \times slides with ΔR redundancy. Hence, ΔR is determined by coarse registration accuracy. If the coarse matching accuracy is high, ΔR is appropriately decreased to increase the registration speed. If the registration accuracy is poor, then we increase ΔR to ensure the matching effect. We slide P_4 on P_{10} and consider the matching area when the highest correlation is observed. If the correlation coefficient exceeds threshold T , then the registration is successful and a pair of images is obtained. Otherwise, the matching result is discarded. Additionally, we require patches to exhibit sufficient foreground pixels because the correlation of blank images is not reliable.

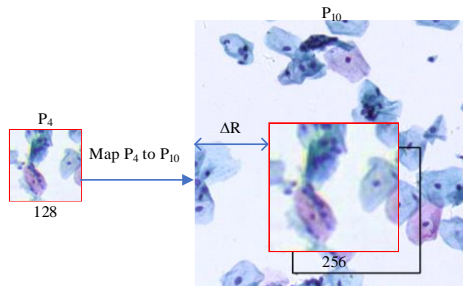


Fig. 3. Fine registration of 4 \times patch and 10 \times patch where P_4 denotes a randomly selected patch of 4 \times slide to be registered, and P_{10} denotes the area corresponding to P_4 on the 10 \times slide with small redundancy based on the coarse registration result. The red box in P_{10} corresponds to the mapped P_4 , and the black box denotes the target registration result.

C. Super-Resolution Network

In our study, we design a two-stage multi-supervised super-resolution method for real cytopathological images. The flow chart of the method is shown in Figure 4. The first stage consists of a generator G_1 and a discriminator D_1 to generate 10 \times images. We introduce real 10 \times images as a supervision to alleviate the image content jump caused by the large DOF difference and obtain 10 \times style generated images. The second stage consists of

generator G_2 and discriminator D_2 , which further reconstruct generated 10 \times images to 20 \times images.

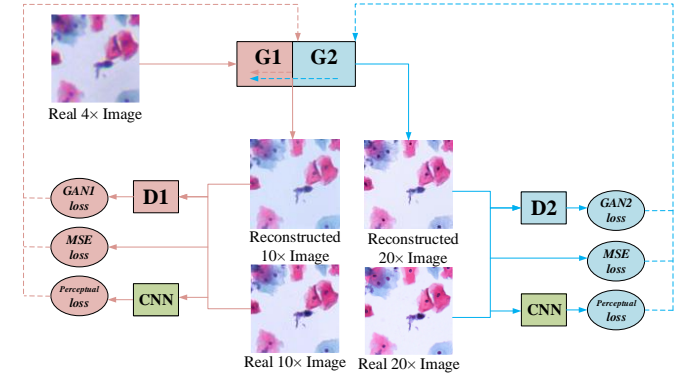


Fig. 4. The two-stage multi-supervised super-resolution model. The first stage consists of G_1 and D_1 , which are used to generate 10 \times images. The second stage consists of G_2 and D_2 , which are used to generate 20 \times images. The loss function for both stages includes MSE loss, perceptual loss and adversarial loss. The solid line represents forward propagation and the dashed line represents back propagation.

LR to MR Reconstruction

The LR to MR reconstruction process corresponds to the first stage of the proposed approach. In this stage, we design a U-Net with DenseBlocks as our generator G_1 as shown in Figure 5. The input of G_1 is a 4 \times /0.1-NA image, and its quality is extremely poor owing to low optical resolution and large DOF. To obtain better reconstruction, we use DenseBlock [37] instead of multiple stacked convolutions in the native U-Net. Specifically, DownBlock in G_1 consists of a 3×3 convolution and two DenseBlocks. UpBlock in G_1 consists of two 3×3 convolutions and two DenseBlocks. The 3×3 convolutions are used to adjust channel number and integrate information. The first and last convolutions of G_1 correspond to 9×9 filters with a large receptive field. Additionally, G_1 includes two outputs, namely generated 10 \times images and feature maps. The output feature maps correspond to the input of the second stage generator G_2 . The designed structure of G_1 exhibits rich feature representation and can integrate information of different scales. Thus, the enhanced representation ability of G_1 helps learn the difficult super-resolution reconstruction from 4 \times images to 10 \times images.

In the field of image generation, the standard discriminator is widely used. The standard discriminator outputs a probability value to determine whether the entire image is real or fake. In the super-resolution process, local details are important. Inspired by Pixel2Pixel, we use a PatchGAN discriminator [38] as opposed to the standard discriminator. The discriminator determines as to whether $N \times N$ patches are real or fake and then averages all responses to obtain a final discrimination result. At this stage, we use 8×8 patches. We use a simple full convolution model with several max-pooling layers as our discriminator D_1 .

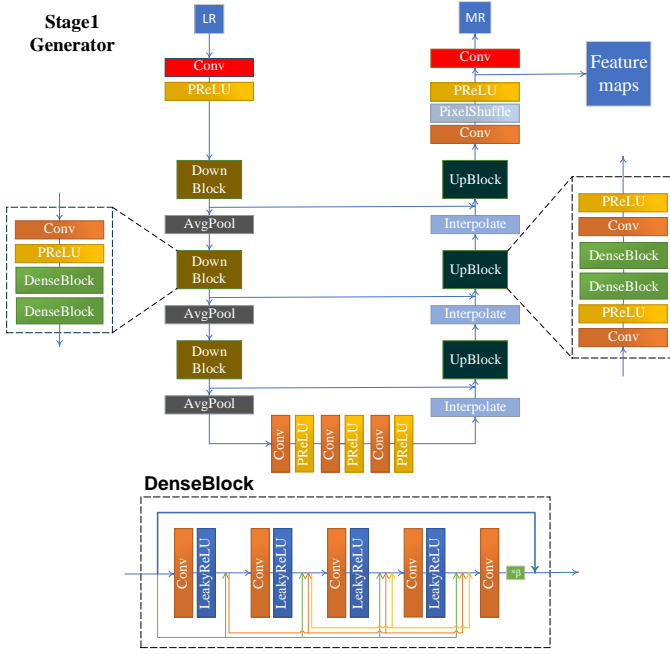


Fig. 5. Structure of stage 1 generator G_1 . The data at the arrows in the U-Net is integrated in a stacked manner. The first and last convolutions of G_1 correspond to 9×9 filters. All other convolutions correspond to 3×3 filters. In the DenseBlock, the data at the arrows is integrated using addition and β is set to 0.5.

Our goal is to train the generator G_1 that outputs a corresponding MR image Y for a given LR image X . To achieve the goal, we train G_1 with parameters θ_G . For a pair of training samples (X, Y) , θ_G should be satisfied as follows:

$$\theta_G^* = \underset{\theta_G}{\operatorname{argmin}} \frac{1}{N} \sum_{n=1}^N L(G_{\theta_G}(X^{(n)}), Y^{(n)}) \quad (1)$$

Here, N denotes the number of samples, $X^{(n)}$ denotes the n^{th} input LR image, and $Y^{(n)}$ denotes the corresponding MR image. Additionally, L denotes a function that measures the difference between two images. To obtain a more comprehensive model on various evaluation indicators, we use the following three loss functions.

MSE Loss

Super-resolution reconstruction tasks typically use MSE or mean absolute error (MAE) to optimize the model, and these loss functions tend to increase the Peak Signal-to-noise Ratio (PSNR) of the reconstructed images. It is widely known that MSE or MAE loss excessively blurs the generated image. However, it is difficult for the model to converge when the pixel-by-pixel MSE or MAE loss function is not used. By convention, we also use the MSE loss function that is defined as follows:

$$I_{\text{MSE}}^{S1} = \frac{1}{N \times C \times H \times W} \sum_{n=1}^N \sum_{c=1}^C \sum_{h=1}^H \sum_{w=1}^W (G_{\theta_G}(X^{(n)})_{c,h,w} - Y^{(n)}_{c,h,w})^2 \quad (2)$$

Here, C , H , and W denote the size of reconstructed images, N denotes the number of training images, $X^{(n)}$ denotes the n^{th} input LR image, $Y^{(n)}$ denotes the corresponding MR image, and G_{θ_G} denotes the generator with parameters θ_G .

Perceptual Loss

For cytopathological images, it is important to maintain the semantic consistency of reconstructed images. We add perceptual loss to ensure the consistency of high-level semantic information and improve the visual quality of images. In contrast to the common VGG loss, we use the cytopathological data to fine-tune the VGG model based on ImageNet weights as opposed to directly using the ImageNet pre-trained VGG weights. The VGG model can be considered as a fine-grained model, which exhibits a very different response to lesion cells and normal cells, while the model trained on ImageNet does not exhibit this type of large difference. We consider the last layer feature maps of the VGG16 model as the comparison object. We then define the following perceptual losses:

$$L_f^{S1} = \frac{1}{N \times C_f \times H_f \times W_f} \sum_{n=1}^N \sum_{c=1}^{C_f} \sum_{h=1}^{H_f} \sum_{w=1}^{W_f} (F(G_{\theta_G}(X^{(n)}))_{c,h,w} - F(Y^{(n)})_{c,h,w})^2 \quad (3)$$

Here, F denotes all convolution modules of the VGG16, $F(Y)$ outputs the feature maps of image Y , and C_f , H_f , and W_f denote the sizes of the feature maps. The details of re-training VGG16 can be found in the *supplementary material S1*.

GAN Loss

With respect to cytopathological images, the visual quality of images obtained by only using the aforementioned loss functions is still not good enough. Hence, we introduce adversarial loss to improve visual quality. The adversarial loss encourages generator G to generate a realistic image to deceive the discriminator D , and D distinguishes the difference between the image generated by G and the real image to the maximum possible extent. The generator and discriminator oppose each other and improve each other. In the process of D and G opposing each other, the image generated by G becomes increasingly more realistic. The loss of the generator is defined as follows:

$$L_G^{S1} = -\frac{1}{N \times H_D \times W_D} \sum_{n=1}^N \sum_{h=1}^{H_D} \sum_{w=1}^{W_D} \log(D_{\theta_D}(G_{\theta_G}(X^{(n)}))_{h,w}) \quad (4)$$

D_{θ_D} is a discriminator with parameter θ_D , and H_D and W_D denote the size of the probability map output by the discriminator. The loss of the discriminator is defined as follows:

$$L_D^{S1} = -\frac{1}{N \times H_D \times W_D} \sum_{n=1}^N \sum_{h=1}^{H_D} \sum_{w=1}^{W_D} (\log(D_{\theta_D}(Y^{(n)})_{h,w}) + \log(1 - D_{\theta_D}(G_{\theta_G}(X^{(n)}))_{h,w})) \quad (5)$$

As previously mentioned, we use three loss functions to ensure that stage 1 exhibits good visual effects. The magnitudes of the three loss functions are inconsistent, and thus we use α , β to balance the loss functions. Finally, the total loss function of the generator is given as follows:

$$L^{S1} = L_{\text{MSE}}^{S1} + \alpha I_f^{S1} + \beta L_G^{S1} \quad (6)$$

The selection of α and β is detailed in the experimental section.

MR to HR Reconstruction

This part corresponds to the second stage G_2 of our proposed method that reconstructs HR images from MR feature maps generated by G_1 . Generally, the difference in content between MR and HR images is smaller than that between LR and MR images. Based on this fact, we design a shallower network for

G_2 , namely a residual-in-residual DenseBlock inspired by ESRGAN. As shown in Figure 6, G_2 consists of two Dense Blocks, a cross-layer skip connection, a PixelShuffle layer, and several convolutions. In the DenseBlock, we use LeakyReLU activation with a coefficient of 0.1. Specifically, we use feature maps of stage 1 as the input of G_2 as opposed to the reconstructed MR images of stage 1. The advantage of using feature maps is that the feature maps with high-level semantic information decrease the depth of the second stage model while ensuring the effect. Simultaneously, it decreases the effect of the reconstructed MR images on subsequent G_2 reconstruction. We will demonstrate the difference with experiments.

In the second stage, we still use the patch discriminator. With respect to the loss functions of the second stage, it also remains consistent with the first stage. When training the two-stage model, we first train G_1 , then train G_2 with freezing G_1 , and finally collectively fine tune G_1 and G_2 . Details of the training will be given in the experimental section.

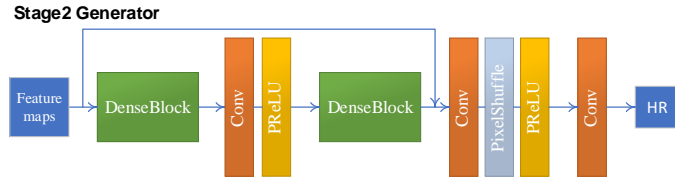


Fig. 6. Architecture of the stage 2 generator G_2 . DenseBlock is consistent with Figure 5, and information is integrated via a stacking manner.

IV. EXPERIMENTS

In this section, we quantitatively analyzed the differences between our model and other classic super-resolution models. Subsequently, we analyzed our designed generator via an ablation study. Additionally, we also demonstrated that our super-resolution reconstructed images positively aid in different computer-aided diagnostic tasks.

A. Data Description

We were unable to obtain a publicly available cytopathological slide dataset for image super resolution. In this study, 142 liquid-based cytology test slides were obtained from the department of clinical laboratory, Tongji Hospital, Huazhong University of Science and Technology. These slides were digitized to whole slide images (WSIs) via a slide scanning system with $4\times/0.1$ -NA, $10\times/0.25$ -NA and $20\times/0.75$ -NA objective lenses as experimental datasets. The resolutions of the $20\times$, $10\times$, and $4\times$ objective lens corresponded to 0.293, 0.586, and 1.465 $\mu\text{m}/\text{pixel}$, respectively. We used 118 slides as the training set and the remaining 24 as the test set. We used the aforementioned registration method to obtain matched images in three different resolutions. With respect to the $20\times$ images, we selected patches with a size of 512×512 as the training sample. With respect to $10\times$ image, the image size corresponded to 256×256 . With respect to the $4\times$ image, we interpolated $4\times$ to $5\times$ for convenience, and the $5\times$ image size corresponded to 128×128 . The details of data are shown in Table I.

TABLE I
SUPER-RESOLUTION DATA

	Train	Test
Slides	118	24
Patches	145200	29300

B. Registration Results

To measure the registration accuracy, we quantitatively tested the effect of the registration. Based on the aforementioned registration method, in the registration process of $4\times$ - $10\times$ and $4\times$ - $20\times$, 5k images were extracted to test the registration effect. The registered paired image was magnified 10 times, and the template matching method was applied again to calculate the offset. The offset was then divided by the magnification to obtain the registration accuracy. With respect to the $4\times$ - $10\times$ and $4\times$ - $20\times$ images, values of mean and standard deviation of fine registration offsets are shown in Table II.

As shown in Table II, the accuracies of the $4\times$ - $10\times$ and $4\times$ - $20\times$ registration reached the sub-pixel level. The $4\times$ - $20\times$ registration accuracy was slightly worse than the $4\times$ - $10\times$ registration accuracy, and stability was also poorer. The performance was reasonable because the difference between $4\times$ and $20\times$ images was relatively higher. Although the $4\times$ - $20\times$ registration was less stable, the standard deviation was still less than one. Overall, the registration effect satisfied super resolution model requirements.

TABLE II
VALUES OF MEAN AND STANDARD DEVIATION OF FINE REGISTRATION
OFFSETS FOR $4\times$ - $10\times$ AND $4\times$ - $20\times$ IMAGES

	Ex	Ey	Sx	Sy
$4\times$ - $10\times$	0.37	0.38	0.27	0.27
$4\times$ - $20\times$	0.38	0.39	0.72	0.82

Ex and Ey represent the average offset errors in horizontal and vertical directions respectively, which are used to measure the error of registration methods. Sx and Sy represent the standard deviation of errors in horizontal and vertical directions respectively.

C. Super Resolution

Training Details

This section details the setup of the hyper-parameters and the model training process. The strategy involved first training G_1 , then training G_2 and finally collectively fine tuning G_1 and G_2 . In this study, all models were optimized via the Adam optimizer with parameters β_1 of 0.9 and β_2 of 0.999. The generator and discriminator learning rates were set to $1e-4$, and each epoch (one iteration of entire train set) halved the learning rate. With respect to the balance of the various loss functions, we set loss weights α to $6e-3$ and β to $1e-3$. First, we set the parameters of G_1 as learnable. At this time, only G_1 was trained. Thus, G_1 generated high quality images after approximately 6 epochs. Subsequently, we fixed all the parameters of G_1 , set the parameters of G_2 as learnable, and trained G_2 . A high quality $20\times$ image was obtained by iterating approximately 6 epochs. Finally, we finetuned G_1 and G_2 with a learning rate of $1e-5$. Additionally, the study was implemented with a PyTorch framework [39] in the Ubuntu 18.10 operating system equipped with 4 Nvidia Tesla P40 graphics cards.

Comparisons with Other Super-resolution Methods

To demonstrate the effect of the proposed method, we compared our method with mainstream SR methods including VDSR, EDSR, SRGAN, ProGanSR and Wang's method on the cytopathological dataset. All the methods were re-trained on the same cytopathological dataset. In accordance with convention, we used PSNR and Structural Similarity Index Measure (SSIM) to evaluate results. However, both the two indicators were not consistent with the visual perception of the human eyes. Therefore, we additionally added Mean Opinion Score (MOS) [27] to measure the super-resolution reconstruction visual effect. In the MOS test, five raters independently scored 1k patches via a variety of different methods. The score was divided into 5 levels with 5 denoting the best visual effect and 1 denoting the worst visual effect. The evaluation criteria of the image mainly included the degree of image clarity and natural degree of image texture. Super resolution quantitative indicators are shown in Table III. PSNR and SSIM were computed in the test set of 29300 patches as shown in Table I (evaluated on RGB space). MOS was evaluated in 1k patches that were randomly selected from the test set.

As shown in Table III, VDSR and EDSR exhibit high PSNR and SSIM. However, the visual effect of these CNN-based methods is poor, and thus the MOS evaluation is low. Conversely, with respect to GAN-based methods including SRGAN, Wang's and ProGanSR, the MOS evaluation is significantly higher although the PSNR and SSIM are lower. This indicates that the images generated based on GAN exhibit a good visual effect. Our method is higher in PSNR and SSIM than the three GAN-based methods and display the highest MOS score. The statistical results indicate that the proposed method is superior to the current mainstream methods in terms of the cytopathological slide image super resolution. We explained why CNN-based methods have high PSNR and SSIM in the *supplementary material S2*.

TABLE III
COMPARISONS OF THE PROPOSED METHOD WITH OTHER METHODS

	Bicubic	VDSR	EDSR	SRGAN	Wang's	ProGanSR	Our_20×	Our_10×	HR_20×
PSNR	25.25	27.60	28.13	27.03	26.75	25.33	26.92	29.71	-
SSIM	0.82	0.85	0.86	0.82	0.79	0.80	0.88	0.91	-
MOS	1.05	1.36	1.73	2.45	3.36	2.43	3.80	-	4.42

Figure 7 shows the reconstructed HR images obtained via the proposed method and other methods. The images generated by our method are clear and display a natural texture. When compared to CNN-based VDSR and EDSR, although the PSNR of the proposed method is lower, the images generated by our method are clearer. The reconstructed HR images of the two methods are completely blurred. When compared to GAN-based SRGAN, ProGanSR and Wang's, the texture generated by our method is more natural and exhibits fewer artifacts. Additionally, the reconstructed HR images of SRGAN and ProGanSR are locally blurred. Furthermore, our method simultaneously generates high-quality 10× images and 20× images, which is important for different precision tasks.

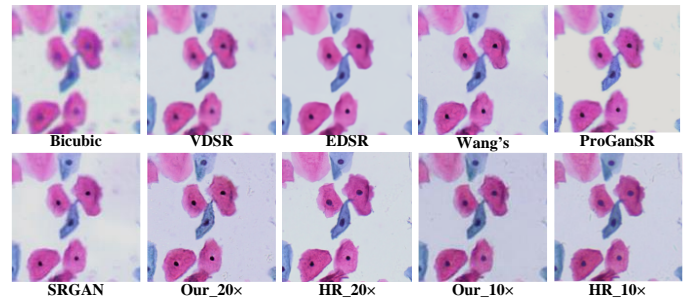


Fig. 7. Comparisons of the proposed method and other methods. HR_20× and HR_10× denote the ground truth. To facilitate easy comparisons, our_10× and HR_10× are interpolated to the same size as HR_20×.

Ablation Studies

To examine the effects of the various parts of the proposed method, we gradually changed the composition of our model and then studied the influence of the modified part. We mainly investigated intermediate supervision, generator structure, and input of the second stage G_2 .

TABLE IV
QUANTITATIVE RESULTS OF ABLATION STUDIES

		PSNR	SSIM	MOS
Intermediate supervision	NO	26.86	0.80	3.4
	YES	26.92	0.88	3.8
Input of G_2	Generated MR images	25.71	0.78	2.0
	Feature maps	26.92	0.88	3.8
Generator architecture	Experiment 1	26.15	0.88	3.0
	Experiment 2	27.23	0.80	2.9
	Experiment 3	26.52	0.78	2.7
	Designed network	26.92	0.88	3.8

Intermediate Supervision

To examine the effect of intermediate supervision in the proposed method, we removed intermediate supervision and maintained all other parameters as identical. We trained a new model without intermediate supervision, and it indicated that intermediate supervision is useful. The statistical results in test set are shown in Table IV. This is because the addition of intermediate supervision gradually decreased the differences in optical resolution and DOF of 4× and 20× images when compared to learning a direct mapping. Furthermore, another advantage of adding intermediate supervision is to generate realistic 10× images.

Designed Architecture

To investigate the effect of the proposed generator architecture on the cytopathological images, we have done three experiments. In the experiment 1, we replaced G_1 and G_2 with a two-stage network composed of a SRResNet of 15 residual blocks and a SRResNet of 5 residual blocks. In the experiment 2, we replaced G_1 with a SRResNet of 15 residual blocks, but kept G_2 unchanged. In the experiment 3, we kept G_1 unchanged, but replaced G_2 with a SRResNet of 5 residual blocks. We compared the reconstructed images of these three experiments and our designed network. The reconstructed images and statistical results are shown in Table IV and Figure 8. In addition to the superiority in SSIM and MOS, the texture of images reconstructed by

our method is more natural and the background is cleaner. The images generated by Experiment 1 and Experiment 2 have some ripple. The images generated by Experiment 3 are blur and have serious fake texture. These results prove the superiority of our designed generator architecture.

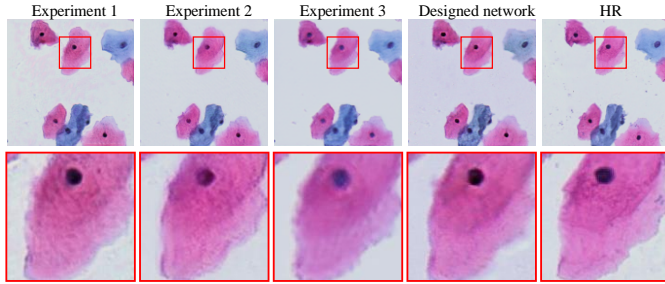


Fig. 8. Images reconstructed by different networks. Experiment 1 replaces G_1 and G_2 with a two-stage model consisting of a SRResNet of 15 residual blocks and a SRResNet of 5 residual blocks, respectively. Experiment 2 replaces G_1 with a SRResNet of 15 residual blocks but keeps G_2 unchanged. Experiment 3 keeps G_1 unchanged but replaces G_2 with a SRResNet of 5 residual blocks. HR is high-resolution images.

Feature Maps as Input

To investigate the impact of the different inputs of the second stage, we tested the generated $10\times$ images and the feature maps as inputs. Figure 9 and Table IV show the reconstructed images and their differences in PSNR, SSIM and MOS. The three indexes when taking the feature maps as input are evidently higher. The output $20\times$ images of G_2 when taking the generated $10\times$ images as input exhibit severe distortion and evident water wave texture. This phenomenon is mainly caused by two factors. Firstly, the information in the generated $10\times$ images is less than that in the feature maps. Secondly, G_2 has limited ability to extract features from images since its relatively shallow architecture.

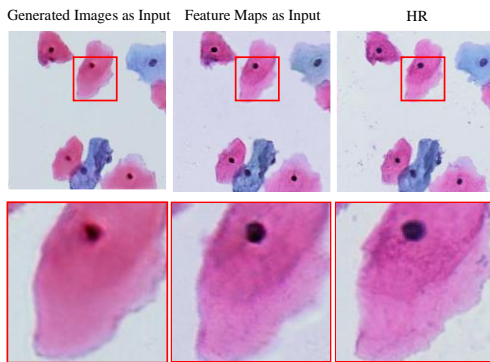


Fig. 9. Effect of different inputs in the second stage. The left is taking generated $10\times$ images as input, the middle is taking the feature maps as input, and the right is the ground truth.

D. Super Resolution of WSIs

In this section, we performed super-resolution reconstruction on the $4\times$ WSI. The model is unable to handle large-size images, and thus it is necessary to split and concatenate the images while processing WSI. In the split phase, the redundancy corresponded to 64 pixels. Subsequently, each split patch was input into the model to obtain a high-resolution patch. Finally, the reconstructed high-resolution patches were concatenated. When concatenating, the redundant regions of adjacent patches were weighted and summed based on the distance to the center line.

Then the high resolution WSI was finally obtained. Figure 10 shows the super-resolution reconstruction result of a $4\times$ WSI. Super resolution for WSIs is important for practical applications of our method including visualization of low-resolution WSIs, cytopathologist interpretation, and computer-aided recognition algorithms.

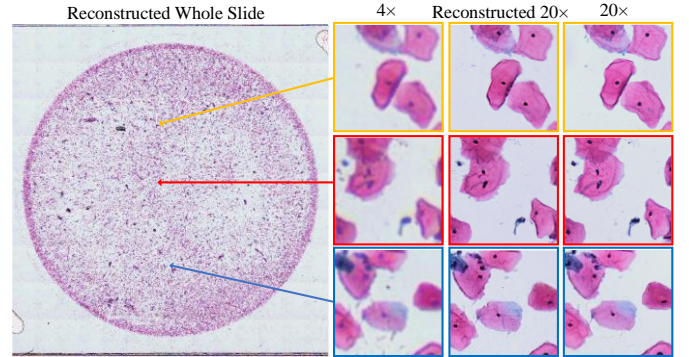


Fig. 10. Whole slide super resolution. The images in the left correspond to $4\times$, the images in the middle are reconstructed via the proposed method, and the images in the right correspond to the ground truth $20\times$.

E. Task Performance

HR cytopathology images obtain better results in various interpretation tasks. Generally, a preliminary screening needs $10\times/0.25\text{-NA}$ images and an accurate diagnosis requires $20\times/0.75\text{-NA}$ or higher resolution images. However, HR image acquisition entails high hardware cost. The above experimental results indicate that PathSRGAN significantly enhances the visual effect of LR images. To verify whether it aids in machine interpretation, we conducted the following experiments. Notably, these resolution parameters are applicable to cytopathology interpretation rather than other microscopy applications.

Lesion Classification Task

On $10\times$ images, we trained a lesion classification model ResNet50 to determine whether a lesion occurs. We first loaded pre-trained weights based on ImageNet datasets. Subsequently, we fine-tuned it on the cytopathological slides to obtain a stable and reliable classification model. It is widely known that the CNN model is sensitive to pixel-level perturbations. Although the reconstructed super-resolution images perform well in terms of human eyes, they significantly affect the accuracy of the classification model. To solve the problem, we added a task loss to adjust the generator such that the generated images conformed to the data distribution required by the classification model. First, we input LR images to the generator G_1 , then input reconstructed $10\times$ images to the lesion classification model, and finally used the task model's prediction error to adjust G_1 . It is noted that the lesion classification model is not updated. The classification task loss function is defined as follows:

$$L_{\text{task}} = -\frac{1}{N} \sum_{n=1}^N y^{(n)} \log \left(c \left(G_{\theta_{G_1}}(X^{(n)}) \right) \right) + (1 - y^{(n)}) \log \left(1 - c \left(G_{\theta_{G_1}}(X^{(n)}) \right) \right) \quad (7)$$

Here, $X^{(n)}$ denotes the input $4\times$ image, C denotes the pre-trained classification model described above, and $y^{(n)}$ denotes the task label of the corresponding $10\times$ image. Additionally, θ_{G_1} indicates that L_{task} only optimizes G_1 .

During training classification task model, we used 118 slides

(74 positive and 44 negative) as training set, 24 (14 positive and 10 negative) slides as test sets. The positive cells in slides were annotated with QuPath [40], which is an open source digital pathological slide analysis software. The data we used in the lesion classification task is shown in Table V. More data and training details can be found in the *supplementary material S3*.

We designed five experiments to illustrate that our super-resolution model help improve task accuracy. The first is a classifier trained on 4× images, which should be the lower limit of the task accuracy. The second is a classifier trained on 10× images, which should be the upper limit of the task accuracy. The third is to use the interpolation 10× images (bicubic interpolation of 4× images) as the input to the second model. The fourth is a classifier trained on interpolation 10× images (ToI). The fifth is to use the super-resolution reconstructed 10× images as the input to the second model. Experiment results in Table VI illustrate that the improvement of classification accuracies derives from the super-resolution reconstruction instead of interpolation. In the task experiment, the generator was adjusted by the task loss. Although the task accuracy will be greatly improved, the visual quality of the generated image may be affected. To explore the impact of the task loss on image visual quality and task accuracy, we designed ablation studies and found that the task accuracy and the generated image quality is a compromise when given a fixed pre-trained task model. The compromise derives from the difference between the required data distribution of a fixed task model and that of the super-resolution generative model. Experimental details can be found in the *supplementary material S4*.

TABLE V
DATA OF THE LESION CLASSIFICATION MODEL

	Train		Test	
	Negative	Positive	Negative	Positive
Slides	44	74	10	14
Patches	80365	82047	3225	3225

TABLE VI
LESION CLASSIFICATION TASK PERFORMANCE UNDER DIFFERENT IMAGES

	4×	Bicubic	ToI	SR	10×
Accuracy	0.892	0.843	0.907	0.936	0.954

Nucleus Segmentation Task

On 20× images, we trained a nucleus segmentation U-Net network. In a manner like the lesion classification task, we used the error of the U-Net segmentation results to adjust the generator G_2 , and thus the generated data distribution satisfied the requirements of the segmentation model. The loss function of the nucleus segmentation task is as follows:

$$L_{\text{Stask}} = -\frac{1}{N} \sum_{n=1}^N \sum_{h=1}^H \sum_{w=1}^W y_{h,w}^{[n]} \log \left(S \left(G_{\theta_{G_2}}(X^{[n]}) \right)_{h,w} \right) + (1 - y_{h,w}^{[n]}) \log \left(1 - S \left(G_{\theta_{G_2}}(X^{[n]}) \right)_{h,w} \right) \quad (8)$$

Here, $X^{[n]}$ denotes the input LR image, S denotes the pre-trained segmentation network described above, and $y^{[n]}$ denotes the corresponding HR image nucleus mask. Additionally,

θ_{G_2} indicates that L_{Stask} only optimizes G_2 , and H and W denote the size of masks. The data we used in segmentation task is show in Table VII. More data and training details can be found in the *supplementary material S3*.

Like the lesion classification task, we designed five experiments. The results shown in Table VIII demonstrate that simple interpolation does not improve the segmentation effect efficiently while the super resolution reconstruction significantly improves the segmentation effect. When compared with the classification task, the effect of segmentation task is not very significant. There are several reasons. First, 4× and 20× images exhibit a large difference in DOF, and thus the content of the image itself is different. For example, a few nuclei appear in 4× images and do not exist in 20× images owing to severe defocusing. Second, in deep-stained or cell-crowded 4× images, the nucleus is almost indistinguishable even after super-resolution. The classification task is relatively robust to the situation, and thus the promotion was more evident. In addition, we compared nucleus segmentation task accuracies of reconstructed 20× images by different SR methods in *supplementary material S5*.

TABLE VII
DATA OF THE NUCLEUS SEGMENTATION TASK

	Train	Test
Slides	118	24
Patches	100000	5000

TABLE VIII
NUCLEUS SEGMENTATION TASK PERFORMANCE UNDER DIFFERENT IMAGES

	4×	Bicubic	ToI	SR	20×
Average IoU	0.27	0.37	0.42	0.59	0.83

V. CONCLUSION

HR digital cytopathological slides are widely used visual analysis object in cervical cancer screening. However, the acquisition of HR digital slides entails high-end imaging equipment and long scanning time. To address the challenge, the present study proposes a novel GAN-based two-stage multi-supervised super-resolution model to learn the mapping of real LR and HR cytopathological images. A new two-stage generator is designed. The generator of the first stage corresponds to a densely-connected U-Net with DenseBlocks and achieves 4×/0.1-NA to 10×/0.25-NA super resolution. The generator of the second stage corresponds to a residual-in-residual DenseBlock and achieves 10×/0.25-NA to 20×/0.75-NA super resolution. The designed generator alleviates the problems caused by the differences in optical resolution and DOF between LR and HR images. We perform a series of comparison experiments and demonstrate that the proposed model is superior to mainstream CNN-based and GAN-based methods in super resolution of cytopathological slide images. Additionally, we examine the effects of the various parts of the proposed method via gradually changing the composition of our model and comparing the effects of the modified model. Three aspects including intermediate supervision, generator structure and input of the second stage G_2 are investigated and demonstrate the rationality of the proposed method architecture. Finally, we also

demonstrate that the reconstructed HR images via the proposed method effectively improve computer-aided diagnosis task accuracies.

There are also a few disadvantages of the proposed approach. First, image pairs of different resolutions are required to train the model, and this adds difficulties in terms of obtaining training data. Second, although we demonstrate the validation of our methods on hundreds of WSIs, a certain distance exists from actual applications owing to the diverse staining styles of different hospitals and different imaging scanners. Further, generalizing our method for data of diverse staining styles require collecting more image pairs, which is expensive. To address the weakness, we will conduct researches of unsupervised domain adaption.

In future, we will combine the proposed method with computer-aided diagnosis algorithms and mobile tiny scanner to increase the penetration of cervical cancer screening in remote areas that lack high-end imaging equipment and cytopathologists.

ACKNOWLEDGMENT

We thank the Optical Bioimaging Core Facility of WNLO HUST for the support in data acquisition. We thank Dr. Tingwei Quan of HUST for the discussion.

REFERENCES

- [1] L. Bruni *et al.*, "Human Papillomavirus and Related Diseases in the World," *ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre)*, Jan. 2019.
- [2] H. Shin *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285-1298, May 2016.
- [3] H. R. Roth *et al.*, "Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1170-1181, May 2016.
- [4] C. You *et al.*, "CT Super-resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE)," *IEEE Transactions on Medical Imaging*, vol. 39, no. 1, pp. 188-203, Jan. 2020.
- [5] F. Shi, J. Cheng, L. Wang, P. Yap, and D. Shen, "LRTV: MR Image Super-Resolution with Low-Rank and Total Variation Regularizations," *IEEE Transactions on Medical Imaging*, vol. 34, no. 12, pp. 2459-2466, Dec. 2015.
- [6] K. Christensen-Jeffries, R. J. Browning, M. Tang, C. Dunsby, and R. J. Eckersley, "In Vivo Acoustic Super-Resolution and Super-Resolved Velocity Mapping Using Microbubbles," *IEEE Transactions on Medical Imaging*, vol. 34, no. 2, pp. 433-440, Feb. 2015.
- [7] H. Zhang *et al.*, "High-throughput, high-resolution deep learning microscopy based on registration-free generative adversarial network," *Biomed. Opt. Express*, vol. 10, no. 3, pp. 1044-1063, Mar. 2019.
- [8] L. Mukherjee, A. Keikhosravi, D. Bui, and K. W. Eliceiri, "Convolutional neural networks for whole slide image super resolution," *Biomed. Opt. Express*, vol. 9, no. 11, pp. 5368-5386, Nov. 2018.
- [9] C. E. Duchon, "Lanczos Filtering in one and two dimensions," *J. Appl. Meteor.*, vol. 18, pp. 1016-1022, 1979.
- [10] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-Based Super-Resolution," *CGA*, vol. 22, pp. 56-65, Mar. 2002.
- [11] J. Sun *et al.*, "Image super-resolution using gradient profile prior," in *CVPR*, pp. 1-8, June 2008.
- [12] K. I. Kim and Y. Kwon, "Single-Image Super-Resolution Using Sparse Regression and Natural Image Prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1127-1133, June 2010.
- [13] H. Chang *et al.*, "Super-resolution through neighbor embedding," in *CVPR*, Washington, DC, USA, 2004, pp. I-I.
- [14] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *ICCV*, Kyoto, Japan, 2009, pp. 349-356.
- [15] S. Schulter, C. Leistner, and H. Bischof, "Fast and accurate image upscaling with super-resolution forests," in *CVPR*, Boston, MA, 2015, pp. 3791-3799.
- [16] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image Super-Resolution Via Sparse Representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861-2873, Nov. 2010.
- [17] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *International conference on curves and surfaces*, Berlin, Heidelberg, 2010.
- [18] S. Wang, L. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *CVPR*, Providence, RI, 2012, pp. 2216-2223.
- [19] R. Timofte, V. De, and L. V. Gool, "Anchored Neighborhood Regression for Fast Example-Based Super-Resolution," in *ICCV*, Sydney, NSW, 2013, pp. 1920-1927.
- [20] C. Dong, *et al.*, "Learning a deep convolutional network for image super-resolution," *European conference on computer vision*. Springer, Cham, 2014, pp. 184-199.
- [21] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," in *CVPR*, Las Vegas, NV, 2016, pp. 1646-1654.
- [22] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution," in *CVPR*, Honolulu, HI, 2017, pp. 5835-5843.
- [23] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight Super-Resolution with Cascading Residual Network," in *ECCV*, Munich, Germany, 2018, pp. 252-268.
- [24] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *ECCV*, Amsterdam, Netherlands, 2016, pp. 694-711.
- [25] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," in *CVPRW*, Honolulu, HI, 2017, pp. 1132-1140.
- [26] I. Goodfellow *et al.*, "Generative adversarial nets," in *NIPS*, Montreal, Canada, 2014, pp. 2672-2680.
- [27] C. Ledig *et al.*, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *CVPR*, Honolulu, HI, 2017, pp. 105-114.
- [28] M. S. M. Sajjadi, B. Schölkopf, and M. Hirsch, "EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis," in *ICCV*, Venice, 2017, pp. 4501-4510.
- [29] A. Bulat and G. Tzimiropoulos, "Super-FAN: Integrated Facial Landmark Localization and Super-Resolution of Real-World Low Resolution Faces in Arbitrary Poses with GANs," in *CVPR*, Salt Lake City, UT, 2018, pp. 109-117.
- [30] Y. Wang *et al.*, "A Fully Progressive Approach to Single-Image Super-Resolution," in *CVPRW*, Salt Lake City, UT, 2018, pp. 977-97709.
- [31] Y. Yuan *et al.*, "Unsupervised Image Super-Resolution Using Cycle-in-Cycle Generative Adversarial Networks," in *CVPRW*, Salt Lake City, UT, 2018, pp. 814-81409.
- [32] J. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *ICCV*, Venice, 2017, pp. 2242-2251.
- [33] H. Wang *et al.*, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nat. Methods*, vol. 16, pp. 103-110, 2019.
- [34] Y. Rivenson *et al.*, "Deep Learning Microscopy," in *Optica*, vol. 4, Issue 11, pp. 1437-1443, 2017.
- [35] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *ECCV*, Munich, Germany, 2018, pp. 0-0.
- [36] J. Martineau, "The relativistic discriminator: a key element missing from standard gan," *arXiv preprint arXiv: 1807.00734* (2018).
- [37] G. Huang *et al.*, "Densely connected convolutional networks," in *CVPR*, Honolulu, HI, 2017, pp. 4700-4708.
- [38] P. Isola *et al.*, "Image-to-image translation with conditional adversarial networks," in *CVPR*, Honolulu, HI, 2017, pp. 1125-1134.
- [39] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 2019, pp. 8024-8035.
- [40] P. Bankhead *et al.*, "QuPath: Open source software for digital pathology image analysis," *Scientific reports*, 7(1), 16878, 2017.