International Journal of Machine Learning and Cybernetics (2024) 15:149–159
https://doi.org/10.1007/s13042-023-01806-9

ORIGINAL ARTICLE

Pathological image super-resolution using mix-attention generative adversarial network

Zhineng Chen[1,2] · Jing Wang[3] · Caiyan Jia[3] · Xiongjun Ye[4]

Received: 4 August 2022 / Accepted: 16 February 2023 / Published online: 3 March 2023

Abstract

Image super-resolution (SR) is a fundamental research task in low-level vision. Recently it has been applied to digital pathology to build transformations from low-resolution (LR) to super-resolved high-resolution (HR) images, which benefits pathological image sharing, storage, management, etc. However, existing studies on pathological image SR are mostly carried out on simulated dataset. It cannot fully reveal the challenge of real-world SR. Meanwhile, these studies rarely investigate SR models from a pathological-tailored perspective. This paper aims to promote studies on pathological image SR from the two aspects. Firstly, we construct PathImgSR, a dataset containing real-captured paired LR-HR pathological images by leveraging the progressively imaging property of pathological images. Second, we develop MASRGAN, a GAN-based mix-attention network to implement the SR. It devises a mix-attention block that is featured by modeling the channel and spatial attentions in parallel. Therefore it better captures the discriminative feature from pathological images spatially and channel-wisely. Furthermore, by formulating the learning processing in an adversarial learning manner, it also improves the subjective perception quality of the reconstructed HR image. Experiments on PathImgSR demonstrate that MASRGAN outperforms popular CNN-based and GAN-based SR methods in both quantitative metrics and visual subjective perception.

Keywords Super-resolution · Pathological image · Attention mechanism · Generative adversarial networks

1 Introduction

Image super-resolution (SR) is a longstanding computer vision task that aims to faithfully reconstruct a high-resolution (HR) image from its low-resolution (LR) counterpart. It is a basic building block in a wide range of applications, such as computational photography [1, 2], video surveillance [3, 4] and medical imaging [5–8]. In particular, it has been applied to digital pathology to enhance the spatial resolution of the scanned pathological image, which has shown promising prospects recently in recovering the HR image and assisting the clinical diagnosis [9–11]. It is regarded as a feasible means to compensate for the missing details that are captured only by expensive high-quality whole-slide scanners. The scanner is costly and not readily available for a large number of hospitals and other medical-related institutions.

Image SR is of great importance to digital pathology. To be specific, in order to identify the suspect lesion regions and make a decision, pathologists are required to inspect the tissue section across magnifications. To enable such a digital diagnosis procedure, whole slide imaging (WSI)

✉ Caiyan Jia
cyjia@bjtu.edu.cn

Zhineng Chen
zhinchen@fudan.edu.cn

Jing Wang
19120405@bjtu.edu.cn

Xiongjun Ye
yexiongjun@cicams.ac.cn

[1] School of Computer Science, Fudan University, Shanghai 200438, China

[2] Shanghai Qi Zhi Institute, Shanghai 200230, China

[3] School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

[4] Department of Urology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

techniques are invented. Typically, a whole slide scanner (e.g., Aperio AT2 Digital Pathology Scanner) is employed to scan the tissue section. It generates several resolution-gradually enlarged images describing the tissue section at different magnifications, e.g., 5×, 10×, 20×, etc. The image at 40× (around 0.25 um/pixel), the finest magnification that describes submicron-level tissue appearance, usually results in an ultra-high resolution image of tens of thousands, e.g., 50000 × 50000. Such an image contains information vital for disease assessment, e.g., nuclear morphology and distributions. However, it requires advanced and expensive optical hardware to acquire, which not only involves a long scanning time, but also the device price is not affordable by many primary hospitals. Moreover, it yields a considerably large storage space, e.g., several gigabytes for a WSI image [12]. Both are obstacles limiting the popularity of digital pathology in the daily clinical workflow. Meanwhile, hindering the use of image analysis and machine learning techniques to assess the pathological image, i.e., computer-aided pathological diagnosis. It is highly demanded in clinical that the WSI image could be smaller, and obtained faster while still preserving similar image quality [13].

Image SR exhibits promising prospects for alleviating this dilemma. It employs an SR algorithm to compensate for the missing details and enhance the image resolution, which can be viewed as a complementary means to further extend the limit of optical hardware. It has the potential of revolutionizing pathological diagnosis by first scanning an LR image with widely available low-cost scanners. The image is smaller (typically 4 or 16 times smaller) thus scanning much faster and requiring much less storage. Then, the inspection and diagnosis are conducted on the on-time super-resolved HR image, with no or only minimum influence on diagnostic quality. Under this pipeline, the quality of the reconstructed HR image becomes a key factor. While the SR task is heavily investigated in natural images, it is now well recognized that the leading methods are convolutional neural network (CNN)-based [14–17] and generative adversarial network (GAN)-based ones [18–20]. Following these studies, there are some works dedicated to the pathological image SR, either using existing CNN models [21] or devising new protocols that are claimed to be tailored to the characteristic of pathological images [10, 22]. Through these studies, it is observed that the generated HR pathological images are helpful in clinical [11, 23].

Despite great progress made, it is observed that most existing studies on SR for pathological image encounter two problems. First, it suffers from the unrealistic training data assumption [24, 25]. This is a common problem in image SR. CNN-based models always require paired LR-HR images for modeling training. However, it is difficult to acquire a large number of real-world paired natural images. As an alternative, most methods are trained on simulated data. For example, the LR image is generated by a simple manipulation such as bicubic downsampling. Such a trivial and uniform assumption clearly ignores the fact that real-world degradation is diverse and complex. As a consequence, there is no guarantee that the trained model would also work well in real-world scenarios. However, the existing SR studies of pathological images mostly follow this assumption. We argue that pathological images are different from natural images. They are captured by dedicated micro-cameras at extremely limited receptive fields. Therefore it is less likely to conform with simple and uniform degradation models such as bicubic interpolation. Meanwhile, the WSI image is obtained by re-scanning the tissue section recurrently using different focus depths. It contains images real-captured at different magnifications, naturally forming real paired LR-HR data. As illustrated in Fig. 1, image (a), (b) and (c) and ground-truth (40×), bicubic downsampled image, and ground-truth sampled at a smaller magnification (10×). It is seen that there is a perceptible difference

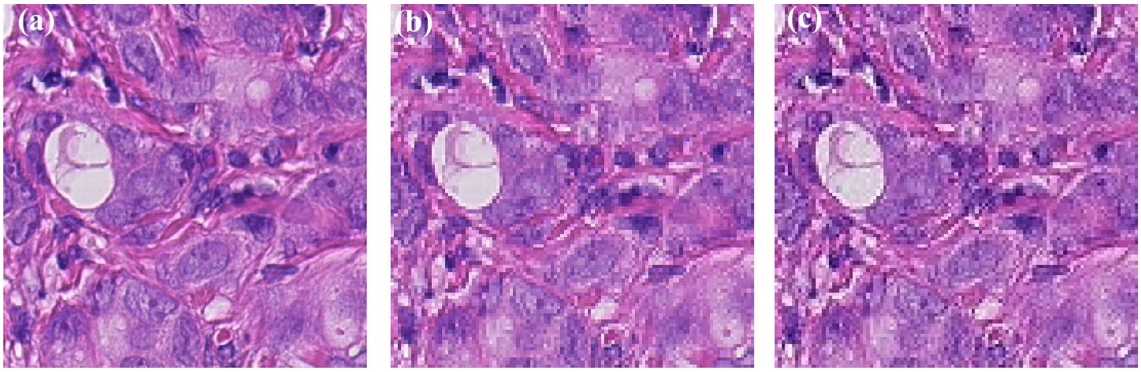

Fig. 1 Pathological image visual effect comparison. Image **a** depicts a pathological image region at magnification 40×, where the image is obtained from the Camelyon16 dataset. Image **b** is obtained by shrinking image **a** by 4 times and then enlarged. Image **c** is directly cropped from the corresponding 10× image and enlarged. Both image **b** and **c** undergo a ×4 bicubic interpolation. It is seen that the real-captured image has visible difference with the interpolated one, corresponding to different degradation process

between the simulated and real-captured LR images (image (b) and (c)). The real-captured one shows more obvious texture details compared to the interpolated one, mainly due to the characteristics of the micro-camera imaging. Therefore, it is not surprising that models trained on simulated image pairs are sub-optimal in real-world applications.

Second, the pathological image is a microscopic-level imaging. It contains a large proportion of recurrent and homogeneous regions, e.g., cells, stroma, etc. This structural knowledge provides additional clues to benefit the SR. For example, exploring the correlation between similar image regions to assist the SR. Moreover, existing studies have already shown that it was useful to explore spatial attention [26, 27] and channel attention [16, 28], which are regarded as useful means to capture the aforementioned clue. SR is likely to benefit if both kinds of attention are considered simultaneously. However, they have not been jointly explored in the context of SR for pathological images.

Motivated by the observation above, in this paper we aim to conduct two tasks to facilitate the studies on pathological image SR. First, we construct PathImgSR, a new pathological image dataset containing real-world paired LR-HR images. It is created from the famous public benchmark Camelyon16 [29], where image patches at different magnifications are sampled and registered to build a realistic benchmark for pathological image SR. We believe our exploration not only provides a readily available dataset for the community, but also offers a guideline for similar dataset construction, e.g., real-world SR dataset for other kinds of microscopy images. Second, we propose MASRGAN, a GAN-based mix-attention network to implement the SR task. It devises a novel mix-attention block (MAB) that first splits the extracted CNN features into several groups from the channel dimension. Then, both spatial and channel attention mechanisms are applied to each group in parallel. By merging the generated features, MAB successfully explores the importance of different feature channels and image regions. Therefore, it better captures the discriminative features that are helpful to the SR. Moreover, the adversarial learning mechanism is incorporated to further enhance the visual subjective perception of the super-resolved HR image. Accordingly, the generators with and without considering GAN are denoted as MASRNet and MASRGAN, respectively. Extensive experiments are conducted to validate the effectiveness of MASRGAN. It is shown that MASRGAN outperforms several popular CNN-based and GAN-based SR networks, while the ablation studies suggest that both channel and spatial attention takes effect. Contributions of this paper are summarized as:

- We construct a new pathological image SR dataset composed of real-captured paired LR-HR images. It better reflects real-world challenges and thus with a larger positive effect on clinical practice.
- MASRGAN is developed for pathological image SR. It aggregates the discriminative features from both spatial and channel dimensions to better contribute to the SR task.
- We carry out extensive experiments to demonstrate the effectiveness of MASRGAN, where the rationality of the designed network is verified, while performance gains are observed when compared with popular SR models.

2 Related work

2.1 Image super-resolution

Image SR is a fundamental low-level vision task that aims to recover realistic details from the LR image and obtain the HR image. With the rapid development of deep learning, CNN-based SR methods have made significant progress in this field. SRCNN proposed by Dong et al. [14] applied CNN to the SR for the first time. It established an end-to-end generation from LR to HR images. Later, residual structure and recursive supervision were further employed to reduce the training difficulty and obtain models with better SR performance, e.g., EDSR [30] and DRRN [31]. The above methods mainly focused on minimizing the pixel difference between the super-resolved image and HR images. They got SR results with high PSNR in general, but tended to be suboptimal in terms of edge and texture reconstruction, thus with slightly inferior subjective metrics. In the year 2017, Ledig et al. [18] first applied GAN to image SR. They designed SRGAN with a perceptual loss to ensure that the generated images are more natural and realistic. More importantly, GAN incorporated realistic-guided model learning from a different perspective, thus significantly improving the subjective perception of generated images. The task was further promoted with the incorporation of novel mechanisms or structures such as attention and graph convolutional networks. Typical works include: RCAN [16] first applied the attention mechanism to image SR. IGNN [32] utilized the graph convolution network for SR. Ma et al. [33] devised DCANet that enabled information flow among adjacent attention blocks. It achieved superior performance with a little additional computational overhead. Mei et al. [34] introduced non-local sparse attention mechanism to SR. These models got impressive results on simulated datasets, i.e., the LR image is not real-captured. It is obtained from the corresponding HR image by image downsampling. This assumption barely portrays the complex image degradation, thus exposing the problem that the SR performance on real-world applications is greatly compromised.

In order to achieve a more realistic SR, researchers paid attention to modeling the degradation process of real-world images, e.g., ensuring simulated LR images with similar noise distribution, blur kernel, etc., as the real degraded images. For example, Zhang et al. [24] proposed SRMD that simultaneously fed the fuzzy kernel, noise, and LR images into the network. It yielded an SR model that is applicable to a wide range of degraded images. Similar methods included RealSR [25], DSGAN [35], DASR [36], etc. Due to the significant information loss in LR images, it is difficult to further improve the SR accuracy through the image's own information. SR methods taking into account reference images also received attention. Remarkable efforts in this field include CrossNet [37], TTSR [38], MASA [39], etc. They mostly showed advantages in recovering the missing details.

Recently, image SR techniques have also been applied to digital pathology. Mukherjee et al. [22] proposed a pathological image SR method based on recurrent neural networks. However, the depth of the network is limited, which in turn limits its performance. Starting from SRGAN, different variants [9, 40, 41] were proposed to construct the pathological image SR models and gained satisfactory performance. Chen et al. [10] developed a method that jointly took into account SR clues from both image and wavelet domains. Li et al. [9] proposed a multi-scale CNN to perform SR in a progressive way. In spite of this, efforts have actively promoted the SR technique for pathological images and it has shown promising prospects in assisting clinical practice. We argue that this can be further boosted by building models based on more realistic data, as well as designing more advanced SR networks.

2.2 Attention mechanism

The attention mechanism has been widely used in computer vision tasks. Channel and spatial are two major attention mechanisms in the literature. Hu et al. [28] proposed SENet to re-weight the feature channels, where the importance of critical channels is promoted while less relevant ones are suppressed. Woo et al. [42] proposed the CBAM module that sequentially combines spatial and channel attention. It achieved better results, but also has an increased computational overhead. Similar channel and spatial attention variants include ECANet [43] and SGE [26]. The attention mechanism was also considered in image SR. Zhang et al. [16] combined the channel attention with the residual structure to deepen the network depth. It achieved good SR results. Liu et al. [27] proposed RFANet that introduced a spatial attention enhancement module for image SR. It guides the network paying more attention to key spatial regions in feature maps. The above as well as other practices demonstrated that the attention mechanism could highlight important features while enhancing the feature usage. However, existing studies are mostly evaluated on natural images. There is still room for further studying the attention mechanism to benefit the pathological image SR.

2.3 Generative adversarial networks

Recently, GAN has made remarkable progress in many image generation tasks including image SR. It has shown its advantage in resolving over-smoothing problems encountered by CNN based SR methods. With the advent of SRGAN [18], Wang et al. further proposed ESRGAN [44] and SFTGAN [45] to rectify the unnatural artifacts and improved the visual perceptual quality of generated HR images. In the year 2020, SPSR [46] introduced an independent gradient branch to guide a structural-aware image recovery. It generates images with sharper edges and textures while also ensuring high quantitative metrics. It is widely believed that the incorporation of GAN helps in reconstructing realistic HR images, which is important for the clinical purpose of pathology images. So we are interested in evaluating the applicability of GAN on real-world paired LR-HR images, and exploring the appropriate way of combining it with the devised network.

3 Methods

We propose MASRGAN, a mix-attention generative adversarial network for pathological image SR. It consists of a generator network termed as MASRNet, and a PatchGAN-based discriminator network. These modules are elaborated as follows.

3.1 Generator network

The generator network takes an LR image as input and outputs the corresponding HR image. A novel network architecture, termed MASRNet, is designed for this purpose. As shown in Fig. 2, it contains a series of newly devised residual overlap blocks (ROBs) for feature extraction, followed by an upsampling-based reconstruction. In MASRNet, residual connections within and between ROB blocks are considered, which encourages gradient flow at multiple granularities.

Let I_{LR} and I_{SR} denote the input LR image and output super-resolved image, respectively. The network first experience a 3×3 convolution, which extract feature F_0 from the LR image.

$$F_0 = H_{Conv}(I_{LR}) \tag{1}$$

where $H_{Conv}(\cdot)$ describes this convolution and the obtained F_0 is further feed into the first ROB.

Then, M ROBs are stacked sequentially for comprehensive feature extraction. This process can be expressed as:

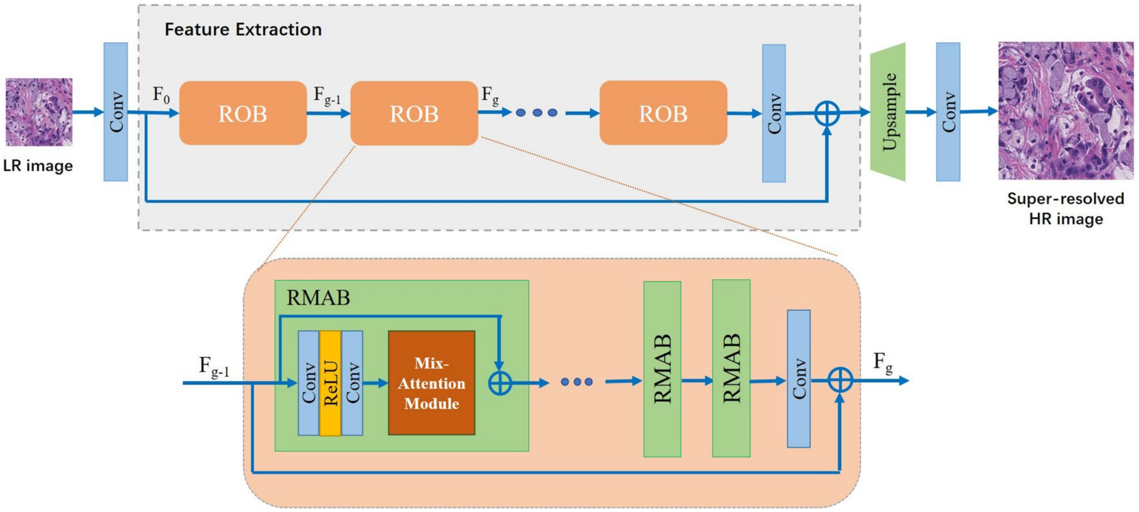

Fig. 2 Structure of the MASRNet. The LR image experiences a series of proposed ROBs for deep feature extraction, and then upsampled. In each ROB, several RMABs are stacked for mix-attention based feature extraction

$$F_g = H_{ROB}(F_{g-1}), g = 1, 2, \ldots, M \tag{2}$$

where F_{g-1} and F_g denote the input and output of the gth ROB, and $H_{ROB}(\cdot)$ denotes the proposed ROB operation that will be described later.

After that, the reconstructed image I_{SR} is obtained by:

$$I_{SR} = H_{Conv}(H_{up}(F_0 + H_{ROB}(F_M))) \tag{3}$$

where $H_{up}(\cdot)$ is the upsampling function, e.g., the deconvolution operation.

Existing studies such as EDSR [31] and SRGAN [16] show that the residual structure plays a significant role in improving SR accuracy. This is explained by the fact that LR images contain low-frequency information and SR tasks mostly focus on recovering missing high-frequency details. The residual structure enables the network to bypass the low-frequency information and thus guides the feature extraction backbone to emphasize extracting high-frequency features. To better implement the objective, as depicted in Fig. 2, we devise a network containing K residual mix-attention blocks (RMABs), where an RMAB has two convolutional layers with a ReLU activation in the middle, and a mix-attention module described below.

3.2 Mix-attention module

As the aforementioned, pathological image has a large portion of recurrent and homogeneous regions. The SR task could be further benefited if this structural knowledge could be well utilized. We argue that the knowledge could be explored from two aspects. First, some knowledge (e.g., edges and texture) could be captured by certain feature channels. Second, it also has special and recurrent visual patterns in spatial, e.g., nuclear morphology. Therefore, we devise a novel mix-attention module that combines the channel and spatial attention in parallel, as illustrated in Fig. 3.

As can be seen, the mix-attention module first splits the feature channels into several groups. For each group, channel attention and spatial attention are both applied. The former uses global average pooling (GAP) followed by two fully connected layers. It then undergoes a sigmoid activation to generate a normalized 1D score vector, which identifies the importance of every channel. Meanwhile, the spatial attention experiences a convolution and a sigmoid activation, generating a normalized 2D score matrix, which highlights the significance of corresponding feature regions. The two kinds of attention are respectively applied to the feature maps in parallel, generating two differently weighted feature maps that are merged together. This manipulation enhances the discriminative of feature maps for the group processed. In the following, the split groups are aggregated and a feature cube of the same dimension as the input feature maps is generated. Finally, we shuffle the channels to increase the feature diversity, such that the next RMAB block can further explore the feature from a different perspective.

The mix-attention module has several attractive properties. First, it is different from existing channel-spatial combined solutions (e.g., CBAM) in that it explores different attention mechanisms in parallel. As contrast, CBAM applies the two kinds of attention sequentially. Thus, the mix-attention module effectively alleviates the distraction between the two kinds of attention. Second, it employs the feature grouping and shuffling strategy. As demonstrated by [47], it could increase the feature diversity with the fixed computational cost budget.

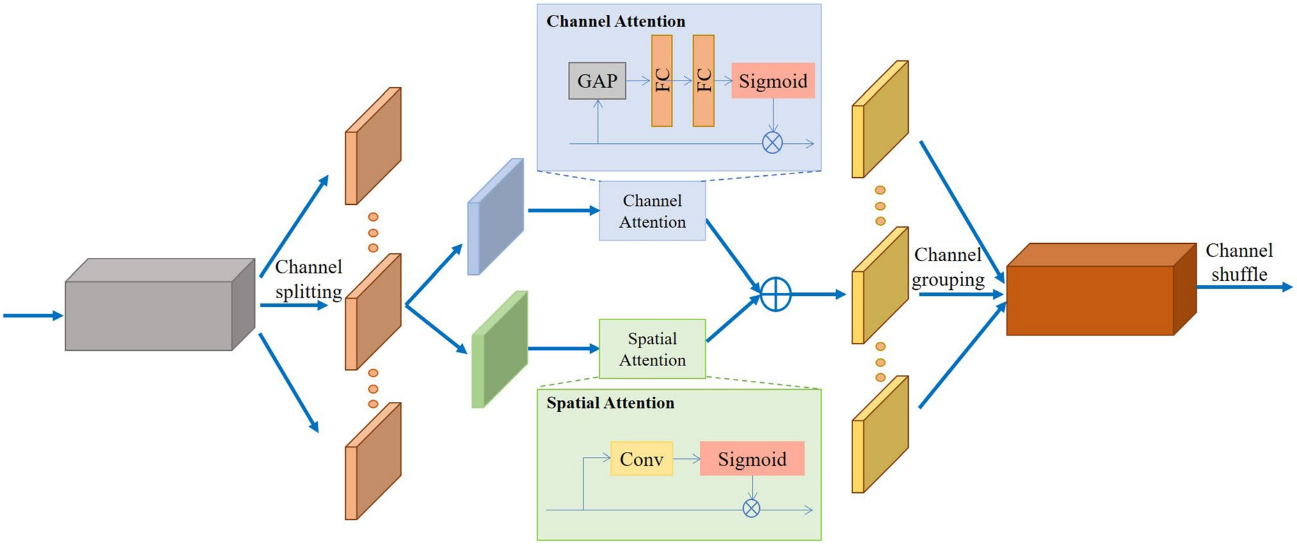

Fig. 3 Structure of the mix-attention module. It is featured by utilizing channel splitting and grouping, paralleled channel and spatial attention, and channel shuffle

3.3 Discriminator network

The discriminator network serves the role of identifying the unrealistic details from the reconstructed image, so as to further improve its subjective quality. To this end, we proposed to use PatchGAN, a discriminator network proposed by Isola et al. [48] that performs the *True* or *False* evaluation based on aggregated evidence from local image regions. Specifically, PatchGAN is a fully convolutional network. When using it, the image is first split into $N \times N$ grids each corresponding to a fixed-sized image region. Then, the regions are fed into PatchGAN one-by-one, where a confidence score is generated for each region. Finally, the scores are averaged to derive an evaluation score that describes the degree of realism of the whole image.

3.4 Loss functions

We introduce the loss functions being taken into account as follows.

MAE loss. Most image SR methods use pixel-based Mean Square Error (MSE) loss or Mean Absolute Error (MAE) loss to measure the difference between the reconstructed image and ground-truth. It yields SR models performing well when evaluated by PSNR. Recent studies have shown that MAE loss can achieve better performance and convergence speed in image SR. Therefore, MAE loss is chosen, which is defined as:

$$L_{mae} = \sum_{I_{SR} \in \mathbf{T}} \mathbb{E}_{I_{SR}} \| G(I_{LR}) - I_{HR} \|_1 \tag{4}$$

where G is the generator network. Note that the term should be summed over all training images, i.e., the training set $\mathbf{T}$.

Perceptual loss. It aims to improve the image quality from the content understanding and perception level. Similar to [16, 44], the feature is extracted from a pre-trained VGG network. The loss minimizes the Euclidean distance between features obtained from the ground-truth and the super-resolved image, which is calculated as follows.

$$L_{per} = \sum_{I_{SR} \in \mathbf{T}} \mathbb{E}_{I_{SR}} \| \emptyset(G(I_{LR})) - \emptyset(I_{HR}) \| \tag{5}$$

where $\emptyset(\cdot)$ is the second convolution (after activation) before the second max-pooling layer within the VGG19 network.

GAN loss. Since GAN is employed, it encourages our generator to favor solutions that show more realistic perception details by incorporating the adversarial learning mechanism to fool the discriminator. As a result, the loss functions of the generator network G is given as:

$$L_{adv} = -\sum_{I_{SR} \in \mathbf{T}} \mathbb{E}_{I_{SR}} [log(D(G(I_{LR})))] \tag{6}$$

Generator loss. Then, the overall loss function of our method is defined as

$$L_G = \alpha L_{mae} + \beta L_{adv} + L_{per} \tag{7}$$

where α, β are two terms to balance the weights of L_{mae} and L_{adv}, respectively.

4 Experiments

4.1 PathImgSR dataset

As described, existing pathological image SR studies mostly experimented on simulated datasets. It ignores the instinctive nature of WSI images, where real-captured images at different magnifications are available. Therefore, we construct PathImgSR, a new pathological image SR dataset. It is derived from the publicly available benchmark Camelyon16 [29], whose training and testing sets include 270 and 129 WSI images sampled from normal and breast cancer patients, respectively. The following scheme is designed to acquire paired LR-HR images.

- **Candidate patch selection.** This step aims at selecting a set of representative image patch coordinates as candidates. Noticing Li et al. [49] sampled 400,000 representative image patches of size 768 × 768 from Camelyon16 (magnification 40×, 200,000 tumor and 200,000 normal). We directly sample the centroid coordinates of these samples.
- **Patch determination.** Randomly pick a candidate sample, sampling a 1050 × 1050 image region at magnification 40X according to its centroid coordinate. Then judge whether it satisfies: over 80% of the image falls in the tissue foreground, and over 80% of the image falls in the tumor or normal region. We keep the patch if both are satisfied, otherwise it is discarded. This process is repeated until 5000 tumor and 5000 normal patches are selected from the training set, while 1000 tumor and 1000 normal patches are selected from the test set.
- **LR-HR pairs acquisition.** Mapping each centroid coordinate to magnification 10×, cropping a patch of size 256 × 256. The patch is enlarged to 1024 × 1024 by bicubic interpolation. Then, a pixel-level registration is performed to find the best-matched offsets that the enlarged 1024 × 1024 patch within the 1050 × 1050 image region. This alignment is designed to address subtle offsets caused by camera movement during slide scanning. After that, the best-matched 1024 × 1024 region at magnification 40× is selected as the HR image while the 256 × 256 one at 10× is the LR image.

With the operations above, we obtain the PathImgSR dataset containing 12,000 LR-HR image pairs. It supports pathological image SR of ×4. The ratio of training and test images is 5:1 and the ratio of tumor and normal patches is 1:1.

4.2 Training details

To train the SR models, the Adam optimizer is employed with an initial learning rate of 10^{-4}, which was halved every 2000 iterations. The batch size is 16. No data augmentation is employed. For balancing the three loss terms, α and β in Eq. 7 are empirically set to 0.01 and 0.005, respectively. For the generator MASRNet, the number of ROBs (i.e., M) is set to 10 and the number of RMABs (i.e., K) is set to 20. All the models are trained on the 10,000 LR-HR training pairs and tested on the 2000 pairs. PSNR, SSIM and perceptual index (PI) are used as evaluation metrics and their average values on the test images are reported. PSNR and SSIM are two widely used objective evaluation metrics. The PSNR is calculated on the luminance channel on YCbCr color space. PI is the same as used in the PIRM-SR challenge [50]. It emphasizes the subjective perception quality given an image. The lower the PI score, the better human perception is indicated. All the models are trained on one Nvidia TITAN Xp GPU using the PyTorch framework.

4.3 Results and comparisons

To validate the effectiveness of the proposed MASRNet and MASRGAN, we compare it with several popular CNN-based and GAN-based SR methods. Their quantitative metrics are given in Table 1, where the best performance with respect to each metric is marked in bold. For a fair comparison, the results are grouped into two sets. One is CNN-based models while the other is GAN-based models. Correspondingly, MASRNet and MASRGAN are included in the two sets, representing the proposed methods for the two sets, respectively.

We first investigate CNN-based comparisons. MASRNet is compared with RCAN [16], IGNN [32] and DCANet [33]. All three methods establish attention-based SR in different ways. As can be seen, MASRNet outperforms them in all the evaluated metrics. It indicates that the proposed mix-attention mechanism better captures the instinct characteristics required by the pathological image SR task.

Then, MASRGAN and several GAN-based SR methods are compared. Their PSNR and SSIM drop

Table 1 Results of different pathological image SR methods on PathImgSR

Methods	PSNR/dB↑	SSIM↑	PI↓
RCAN [16]	29.38	0.8032	6.12
IGNN [32]	29.45	0.8041	5.98
DCANet [33]	29.51	0.8052	5.72
MASRNet	**29.52**	**0.8066**	5.01
QEGAN [9]	24.19	0.6308	6.08
WA-SRGAN [41]	24.25	0.6361	5.63
ESRGAN [44]	24.54	0.6706	5.89
SPSR [46]	24.58	0.6724	4.92
MASRGAN	25.33	0.6730	**4.03**

considerably compared to the CNN-based ones, but their PI scores improve. For example, MASRGAN further reduces the PI score by nearly 1 point compared to MASRNet. The observation is in line with the observations in natural image SR. The incorporation of GAN could improve the subjective quality of the generated image but often leads to worse PSNR and SSIM. When inspecting the GAN-based methods, it can be seen from Fig. 4 that the three GAN-based methods are all good at recovering detailed textures. MASRGAN generates the super-resolved image that is closest to the HR image, where contours of cell nucleus are more naturally depicted. It also achieves the best quantitative metrics.

It is observed that both PSNR and SSIM of the GAN-based methods has dropped a lot. For example, PSNR has by 4–5 points when compared to the CNN-based methods. The relatively large reduction comes from two aspects. First, pathological images contain rich edges and textures. Small variations in these patterns are insensitive to humans, but are sensitive to PSNR. As a result, the performance gap, although significant, not exceeds human expectations so human do not think the GAN-based ones are worse. Second, the GAN-based methods are observed to have a slight color drift, thus resulting in a larger PSNR drop. We can see from Fig. 4 that the GAN-based solutions can generate faithful textural patterns but their colors are affected by the adversarial learning process and show perceptible color drift, i.e., a little bit redder. This problem has not been mentioned by existing studies before. We argue that it is mainly caused by uneven slide staining. In other words, pathological images have similar textural characteristics, but they exhibit large variants in stained color (e.g., dark or light) because of the different duration of staining. We will continue to explore ways that better separate the textural and color, and thus eliminate the undesired color variants.

In Fig. 5, another illustrative example containing both quantitative and qualitative results of seven methods is provided. Despite MASRNet with slightly better detail recovery performance, the images generated by CNN-based methods exhibit over-smoothed textures in general, e.g., blurring nuclear contours. As for the GAN-based methods, ESRGAN and SPSR generate sharper edges and textures but also contain unnatural artifacts. On the contrary, MASRGAN utilizes the channel attention that better separates the high-frequency and low-frequency signals. As shown in ESRGAN and MASRGAN, the artifacts are largely suppressed due to the incorporation of the channel attention. Meanwhile, the spatial attention also takes effect in encouraging different image regions to learn from each other, thus generating more natural cell nucleus contours and stroma details. Therefore, MASRGAN better surpasses the artifacts, thus receiving better objective and subjective metrics. Note that it exhibits texture patterns very similar to real-captured pathological images (see Fig. 1a). The results also verify the effectiveness of the proposed mix-attention mechanism.

4.4 Ablation study

To validate the effectiveness of the mix-attention mechanism and the devised network, we conduct controlled experiments as follows.

We first assess the use of different attention mechanisms. Concretely, none, channel, spatial and mix-attention are all evaluated and Table 2 gives their performances. The results are obtained by removing the corresponding attention blocks within MASRNet. It can be seen that considering either of the two attention mechanisms leads to a significant performance improvement. Channel and spatial attention perform similarly when solely taken into account. By simultaneously considering the two kinds of attention, the mix-attention gains further

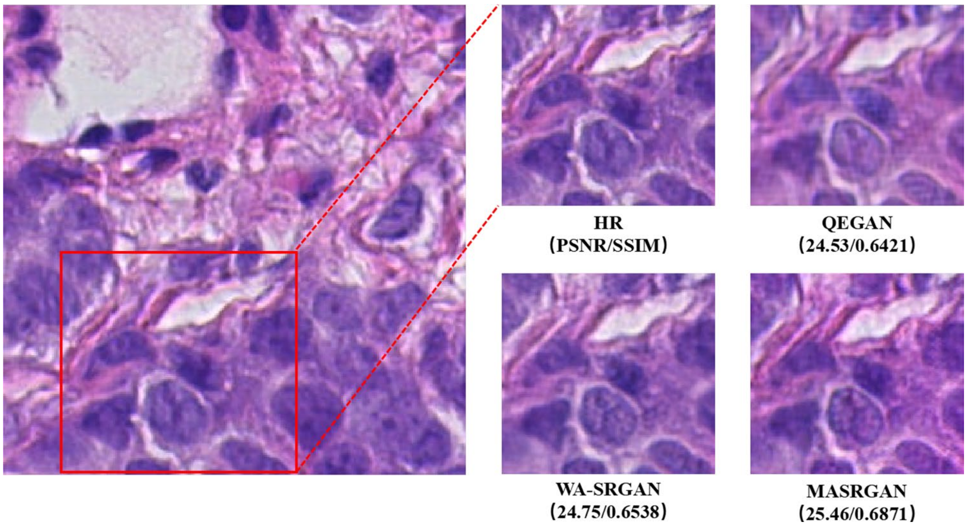

Fig. 4 Qualitative results of different GAN-based methods on pathological image SR

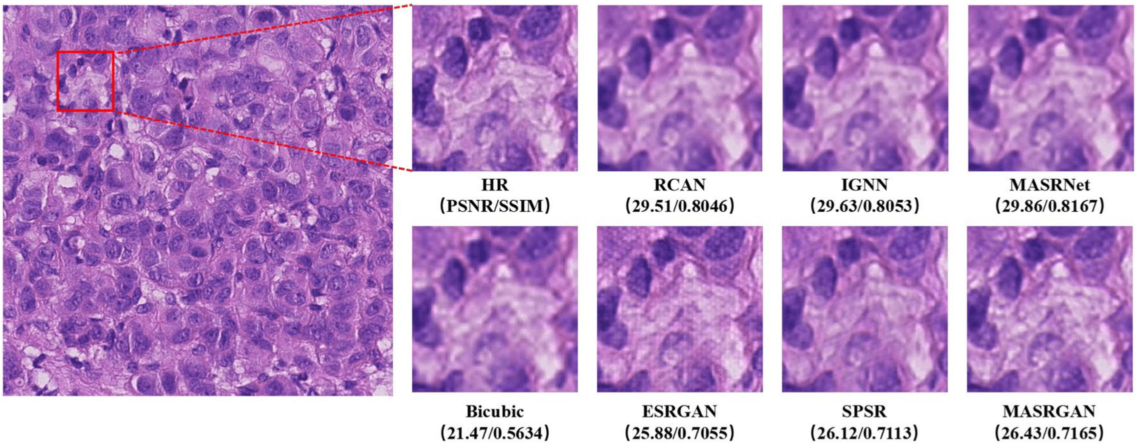

Fig. 5 Qualitative comparisons of different methods on pathological image SR

Table 2 SR performance of different attention mechanisms

Methods	Channel attention	Spatial attention	PSNR/dB↑	SSIM↑	PI↓
1			27.57	0.7767	6.70
2	✓		29.34	0.8028	6.15
3		✓	29.23	0.8003	6.23
4	✓	✓	**29.52**	**0.8066**	**5.01**

Table 3 SR performance of different discriminators

Methods	PSNR/dB↑	SSIM↑	PI↓
MASRGAN-VGG16	25.16	0.6712	4.37
MASRGAN-VGG128	25.27	0.6721	4.28
MASRGAN-PatchGAN	**25.33**	**0.6730**	**4.03**

improvement especially on the PI score. It indicates that the two kinds of attention are complementary to each other. The result clearly verifies the effectiveness of the proposed mix-attention.

Then, we assess the effectiveness of using PatchGAN as the discriminator network. As a comparison, VGG16 and VGG128 are taken into account. They are adopted by natural image SR models SRGAN [18] and RealSR [25], respectively. Both evaluate the image fidelity at the whole image level. As shown in Table 3, MASRGAN-PatchGAN presents better quantitative metrics. The result is attributed to the characteristic of pathological images, which are mainly composed of recurrent and homogeneous cellular-level appearance such that less emphasis is placed on global-level semantics. Therefore, PatchGAN is an appropriate choice.

5 Conclusions

This paper targets developing image SR datasets and techniques to better accommodate the pathological image SR task, which are crucial components of pathological slide digitization. To this end, we have constructed PathImgSR, a dataset containing over 12,000 paired LR-HR pathological images. It is obtained from real-captured images thus forming a benchmark that better reveals the challenges of real-world pathological image SR. Meanwhile, we have proposed MASRGAN, a mix-attention generative adversarial network for pathological image SR, where a mix-attention module is devised to utilize the channel and spatial attention in parallel. Experiments on PathImgSR basically validate the effectiveness of MASRGAN, where performance improvements are observed when compared to popular CNN-based and GAN-based SR methods. We also notice that for MASRGAN generated images there is a small color drift. Therefore, future studies include developing methods that well overcome this problem. Moreover, we are also interested in employing the proposed method to deal with pathological images scanned by different scanners to verify the generalization ability, and applied it to enhance the resolution of other kinds of microscopy images.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grants 62172103, 61876016 and 62076007.

Data availability The PathImgSR dataset comes from Camelyon16, a public benchmark at https://camelyon16.grand-challenge.org/. We will make coordinates of the sampled images public available, while gives the link to the sampled images according to request.

References

1. Umer RM, Micheloni C (2021) Rbsricnn: Raw burst super-resolution through iterative convolutional neural network. arXiv preprint arXiv:2110.13217
2. Chen Z, Ai S, Jia C (2019) Structure-aware deep learning for product image classification. ACM Trans Multimed Comput, Commun, Appl (TOMM) 15(1s):1–20
3. Deshmukh AB, Usha Rani N (2019) Fractional-grey wolf optimizer-based kernel weighted regression model for multi-view face video super resolution. Int J Mach Learn Cybern 10(5):859–877
4. Zuxuan W, Li H, Xiong C, Jiang Y-G, Davis LS (2022) A dynamic frame selection framework for fast video recognition. IEEE Trans Pattern Anal Mach Intell 44(4):1699–1711
5. Li Y, Sixou B, Peyrin F (2021) A review of the deep learning methods for medical images super resolution problems. IRBM 42(2):120–133
6. Liu C, Xie H, Zhang Y (2020) Self-supervised attention mechanism for pediatric bone age assessment with efficient weak annotation. IEEE Trans Med Imaging 40(10):2685–2697
7. Jingyuan X, Xie H, Liu C, Yang F, Zhang S, Chen X, Zhang Y (2021) Hip landmark detection with dependency mining in ultrasound image. IEEE Trans Med Imaging 40(12):3762–3774
8. Liu C, Xie H, Zhang S, Mao Z, Sun J, Zhang Y (2020) Misshapen pelvis landmark detection with local-global feature learning for diagnosing developmental dysplasia of the hip. IEEE Trans Med Imaging 39(12):3944–3954
9. Upadhyay U, Awate SP (2019) A mixed-supervision multilevel gan framework for image quality enhancement. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, p 556–564
10. Chen Z, Guo X, Yang C, Ibragimov B, Yuan Y (2020) Joint spatial-wavelet dual-stream network for super-resolution. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, p 184–193
11. Li B, Keikhosravi A, Loeffler AG, Eliceiri KW (2021) Single image super-resolution for whole slide image using convolutional neural networks and self-supervised color normalization. Med Image Anal 68:101938
12. Huisman A, Looijen A, van den Brink SM, van Diest PJ (2010) Creation of a fully digital pathology slide archive by high-volume tissue slide scanning. Human Pathol 41(5):751–757
13. Ghaznavi F, Evans A, Madabhushi A, Feldman M (2013) Digital imaging in pathology: whole-slide imaging and beyond. Annu Rev Pathol 8(1):331–359
14. Dong C, Loy CC, He K, Tang X (2015) Image super-resolution using deep convolutional networks. IEEE Trans Pattern Anal Mach Intell 38(2):295–307
15. Kim J, Lee JK, Lee KM (2016) Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p 1637–1645
16. Zhang Y, Li K, Li K, Wang L, Zhong B, Fu Y (2018) Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European conference on computer vision (ECCV), p 286–301
17. Lee W, Lee J, Kim D, Ham B (2020) Learning with privileged information for efficient image super-resolution. In: European Conference on Computer Vision. Springer, p 465–482
18. Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, et al (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p 4681–4690
19. Zhang W, Liu Y, Dong C, Qiao Y (2019) Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, p 3096–3105
20. Mahapatra D, Bozorgtabar B, Garnavi R (2019) Image super-resolution using progressive generative adversarial networks for medical image analysis. Comput Med Imaging Graph 71:30–39
21. Deng Y, Feng M, Jiang Y, Zhou Y, Qing H, Xiang F, Wang Y, Bao J, Bu H (2020) Development of pathological super-resolution images using artificial intelligence based on whole slide image
22. Mukherjee L, Bui HD, Keikhosravi A, Loeffler A, Eliceiri KW (2019) Super-resolution recurrent convolutional neural networks for learning with multi-resolution whole slide images. J Biomed Opt 24(12):126003
23. Qiao C, Li D, Guo Y, Liu C, Jiang T, Dai Q, Li D (2021) Evaluation and development of deep neural networks for image super-resolution in optical microscopy. Nat Methods 18(2):194–202
24. Zhang K, Zuo W, Zhang L (2018) Learning a single convolutional super-resolution network for multiple degradations. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p 3262–3271
25. Cai J, Zeng H, Yong H, Cao Z, Zhang L (2019) Toward real-world single image super-resolution: A new benchmark and a new model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 3086–3095
26. Li X, Hu X, Yang J (2019) Spatial group-wise enhance: Improving semantic feature learning in convolutional networks. arXiv preprint arXiv:1905.09646,
27. Liu J, Zhang W, Tang Y, Tang J, Wu G (2020) Residual feature aggregation network for image super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, p 2359–2368
28. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p 7132–7141
29. Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, Van Der Jeroen AWM, Laak MH, Manson Quirine F, Balkenhol M et al (2000) Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 318
30. Lim B, Son S, Kim H, Nah S, Mu Lee K (2017) Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, p 136–144
31. Tai Y, Yang J, Liu X (2017) Image super-resolution via deep recursive residual network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p 3147–3155
32. Zhou S, Zhang J, Zuo W, Loy CC (2020) Cross-scale internal graph neural network for image super-resolution. Adv Neural Inf Process Syst 33:3499–3509
33. Ma X, Guo J, Tang S, Qiao Z, Chen Q, Yang Q, Fu S (2020) Dcanet: Learning connected attentions for convolutional neural networks. arXiv preprint arXiv:2007.05099
34. Mei Y, Fan Y, Zhou Y (2021) Image super-resolution with non-local sparse attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, p 3517–3526
35. Fritsche M, Gu S, Timofte R (2019) Frequency separation for real-world super-resolution. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). IEEE, p 3599–3608
36. Wang L, Wang Y, Dong X, Xu Q, Yang J, An W, Guo Y (2021) Unsupervised degradation representation learning for blind super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, p 10581–10590
37. Zheng H, Ji M, Wang H, Liu Y, Fang L (2018) Crossnet: an end-to-end reference-based super resolution network using cross-scale

warping. In: Proceedings of the European conference on computer vision (ECCV), p 88–104
38. Yang F, Yang H, Fu J, Lu H, Guo B (2020) Learning texture transformer network for image super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, p 5791–5800
39. Lu L, Li W, Tao X, Lu J, Jia J (2021) Masa-sr: matching acceleration and spatial adaptation for reference-based image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, p 6368–6377
40. Zhang H, Fang C, Xie X, Yang Y, Mei W, Jin D, Fei P (2019) High-throughput, high-resolution deep learning microscopy based on registration-free generative adversarial network. Biomed Opt Express 10(3):1044–1063
41. Shahidi F (2021) Breast cancer histopathology image super-resolution using wide-attention gan with improved Wasserstein gradient penalty and perceptual loss. IEEE Access 9:32795–32809
42. Woo S, Park J, Lee J-Y, Kweon IS (2018) Cbam: convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV), p 3–19
43. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q (2020) Eca-net: Efficient channel attention for deep convolutional neural networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), p 11531–11539
44. Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Qiao Y, Loy CC (2018) Esrgan: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European conference on computer vision (ECCV) workshops
45. Wang X, Yu K, Dong C, Loy CC (2018) Recovering realistic texture in image super-resolution by deep spatial feature transform. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p 606–615
46. Ma C, Rao Y, Cheng Y, Chen C, Lu J, Zhou J (2020) Structure-preserving super resolution with gradient guidance. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, p 7769–7778
47. Xie S, Girshick R, Dollár P, Tu Z, He K (2017) Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p 1492–1500
48. Isola P, Zhu J-Y, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p 1125–1134
49. Li Y, Ping W (2018) Cancer metastasis detection with neural conditional random field. arXiv preprint arXiv:1806.07064
50. Blau Y, Mechrez R, Timofte R, Michaeli T, Zelnik-Manor L (2018) The pirm challenge on perceptual super resolution

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.