Multi Attention Neural Network for Digital Rock CT Images Super-Resolution

Zhihao Xing[1], Jun Yao[1], Lei Liu[1], and Hai Sun[1]

[1]China University of Petroleum (East China)

December 8, 2022

Abstract

High-quality digital rock images are essential for subsequent high-precision numerical simulations. But limited by the imaging capability of computed tomography (CT), high resolution digital rock images with wide imaging field of view (FOV) cannot be acquired simultaneously. To cope with this constraint, we propose a novel Multi Attention Super-Resolution Neural Network (MASR) that enhances the resolution of images with wide FOV. Considering that textures and edges are more crucial in digital rocks, MASR introduces the component attention mechanism of Component Divide-and-Conquer Super-Resolution (CDCSR) model. By redesigning the hourglass network with spatial and channel attention mechanisms, proposing a spatial attention-based mask module, and optimizing the component attention mask calculation process, MASR delivers higher information utilization with fewer parameters and faster training than CDCSR. And we optimize the depth of MASR to trade off speed and super-resolution quality. Furthermore, we retrained several state-of-the-art models. Through quantitative evaluations and qualitative visualizations, it is verified that MASR can recover sharper edges while removing noise, and obtain digital rock images with superior quality and reliability. The pixelwise relative errors of MASR reconstructions are reduced by 15% to 26% over bicubic interpolation method. Our codes are publicly available at https://github.com/MHDXing/MASR-for-Digital-Rock-Images.

1

Multi Attention Neural Network for Digital Rock CT Images Super-Resolution

3 **Zhihao Xing[1], Jun Yao[1], Lei Liu[1] and Hai Sun[1]**

4 [1]Research Centre of Multiphase Flow in Porous Media, China University of Petroleum (East
5 China), Huangdao District, Qingdao, China.

6 Corresponding author: Jun Yao (rcogfr_upc@126.com)

7 **Key Points:**

8 • A Multi Attention Neural Network model is proposed to enhance the resolution of digital
9 rock CT images.

10 • Based on the component attention model, the proposed model incorporates channel and
11 spatial attention mechanisms to achieve higher performance with fewer parameters.

12 • The proposed model can rely on low resolution images to recover sharp details and edges
13 while suppressing noise, breaking through hardware limitations to boost digital rock
14 quality.

15

Abstract

High-quality digital rock images are essential for subsequent high-precision numerical simulations. But limited by the imaging capability of computed tomography (CT), high resolution digital rock images with wide imaging field of view (FOV) cannot be acquired simultaneously. To cope with this constraint, we propose a novel Multi Attention Super-Resolution Neural Network (MASR) that enhances the resolution of images with wide FOV. Considering that textures and edges are more crucial in digital rocks, MASR introduces the component attention mechanism of Component Divide-and-Conquer Super-Resolution (CDCSR) model. By redesigning the hourglass network with spatial and channel attention mechanisms, proposing a spatial attention-based mask module, and optimizing the component attention mask calculation process, MASR delivers higher information utilization with fewer parameters and faster training than CDCSR. And we optimize the depth of MASR to trade off speed and super-resolution quality. Furthermore, we retrained several state-of-the-art models. Through quantitative evaluations and qualitative visualizations, it is verified that MASR can recover sharper edges while removing noise, and obtain digital rock images with superior quality and reliability. The pixelwise relative errors of MASR reconstructions are reduced by 15% to 26% over bicubic interpolation method. Our codes are publicly available at https://github.com/MHDXing/MASR-for-Digital-Rock-Images.

1 Introduction

In recent years, digital rock technology has been playing an increasingly important role in oil and gas development, as it enables the study of pore morphology and network topology at the micro and nano scale (Yao et al., 2005). Moreover, it can be flexibly combined with numerical simulations to analyze the petrophysical and flow properties of rocks (Liu et al., 2018; Mostaghimi et al., 2013; Y Wang et al., 2018). X-ray computed tomography (CT) is the most direct, efficient and extensively used method to obtain three-dimensional (3D) digital rock images (Chung et al., 2019; Iglauer and Lebedev, 2018; Oluwadebi et al., 2019). In addition, CT does not destroy the rock during imaging, which ensures that the rock can be subsequently used for other experiments (Wildenschild and Sheppard, 2013).

A qualified rock CT image should satisfy two requirements simultaneously: sufficient resolution and enough field of view (FOV) (Y Wang, 2018). In practical terms, however, these two conditions are in conflict with each other due to the imaging capability of the device. High resolution (HR) CT images are required to resolve minute structural features of rocks for subsequent simulations, yet the FOV of these images is usually not wide enough to characterize the heterogeneity of rocks at multiple scales (Li et al., 2017; Y Wang, 2018).

Image super-resolution (SR) reconstruction is a method of recovering HR images from low resolution (LR) images (Z Wang et al., 2021). SR is a highly viable and effective method that can be used to enhance the resolution of digital rock images as much as possible while obtaining a large enough imaging FOV to surpass the limitations of physical imaging hardware (seen in Figure 1). Deep learning-based SR algorithms have become mainstream in recent years with their outstanding performance, outlawing previous traditional classical algorithms, including bicubic interpolation, iterative back-projection (Tekalp et al., 1992), neighborhood embedding method (Rahiman and George, 2017), sparse representation (Yang et al., 2010), etc.

112 pixels / 1.38mm	28 pixels / 0.345mm	448 pixels / 1.38mm	112 pixels / 0.345mm	112 pixels / 0.345mm
LR image	**LR patch**	**SR image**	**SR patch**	**HR image**
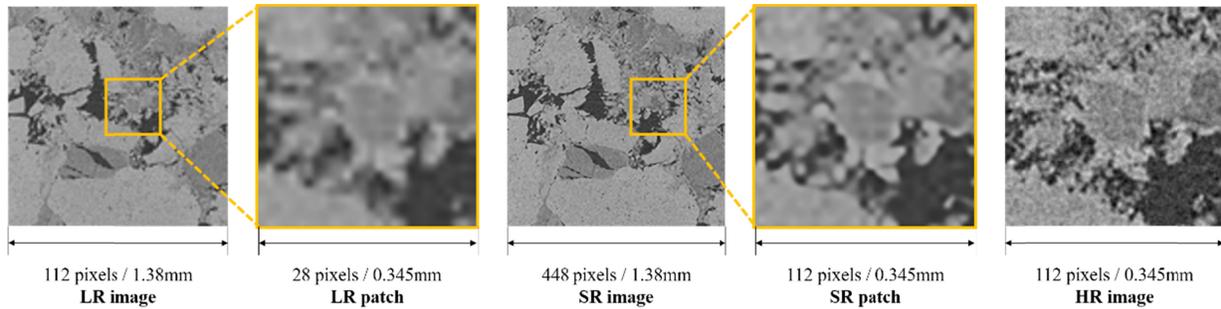
Figure 1. Low resolution (LR), super-resolution (SR) and high resolution (HR) digital rock images. Limited by digital rock imaging hardware, either only wide FOV LR images or narrow FOV HR images can be acquired. The SR model can enhance the resolution of LR images, trade-off FOV and resolution. At ×4 SR, the reconstructed image has a 4× FOV of the HR image (16× the imaging area).

Dong et al. (Dong et al., 2014) first applied deep learning to image SR and proposed the Super-Resolution Convolutional Neural Network (SRCNN), which has higher quality and faster reconstruction compared to the traditional sparse-coding-based SR with optimal performance at that time. Based on SRCNN, Wang et al. (Y Wang et al., 2019) proposed 3DSRCNN to realize 3D image SR of rock samples. But the structure of bicubic interpolation upsampling on SRCNN is computationally complex and amplifies the noise effect. Therefore, Dong et al. (Dong et al., 2016) then proposed the Fast SRCNN (FSRCNN), which greatly speeds up SR using post-upsampling of deconvolution layer.

Enhanced Deep Super-Resolution Network (EDSR) removes unnecessary modules from SRResNet (Ledig et al., 2017), allowing the training procedure to be more stable and the network to be stacked deeper. Wide Activation Super-Resolution (WDSR) improves SRResNet in another way, it reduces the depth and expands the width (Yu et al., 2018). SRResNet, EDSR, and WDSR are used on the digital rock SR. The rock images reconstructed by these deep learning based algorithms not only remove LR noise but also recover sharp edges, which is significantly better than traditional methods (Y D Wang et al., 2019).

In order to generate texture features that are more natural and closer to HR images, some Generative Adversarial Network (GAN) (Goodfellow et al., 2020) based models are used for SR tasks, such as Super-Resolution Generative Adversarial Network (SRGAN) (Ledig et al., 2017), Enhanced SRGAN (ESRGAN) (X Wang et al., 2018), etc. Wang et al. (Y D Wang et al., 2020) performed SR reconstruction of 2D and 3D digital rock images using ESRGAN, and the excellent performance of ESGAN was verified by subsequent segmentation and simulation tasks.

The addition of attention mechanisms is another network design idea to improve the SR performance of neural networks. With the Residual in Residual (RIR) module and channel attention mechanism, Residual Channel Attention Network (RCAN) (Y Zhang et al., 2018) is capable of exploiting inter-channel features and reducing the learning difficulty. Component Divide-and-Conquer Super-Resolution (CDCSR) (Wei et al., 2020) proposed the component attention mechanism, which allows the model to focus more on restoring textures and details. Accurate resolution of edges and textures of the various rock components (pores, fractures, minerals, etc.) is essential to improve the accuracy of subsequent tasks (Y D Wang et al., 2020). Therefore, the component attention mechanism is more suitable for digital rock images SR tasks.

94   In order to recover higher quality images using fewer parameters, we propose a novel
95 Multi Attention Super-Resolution Neural Network (MASR) for the characteristics of digital rock
96 images. MASR combines component, channel and spatial attention mechanisms to enhance
97 feature extraction and improve information utilization. In addition, we propose a spatial
98 attention-based component mask module to assist MASR in focusing on textures and details and
99 improving performance. Thirdly, we optimize the component mask calculation process and
100 investigate the effect of MASR depth on SR reconstruction performance in order to obtain higher
101 quality images with as short training time as possible. Finally, because of the domain gap
102 between digital rock CT images and photographs, we retrain EDSR, RCAN and CDCSR using
103 digital rock CT images and compare them with MASR to verify that MASR has state-of-the-art
104 performance.

105   In the following sections of this paper, Section. 2 introduces several advanced super-
106 resolution model architectures and describes the principle of MASR. Section. 3 explores the
107 appropriate network depth for MASR and evaluates the performance of MASR against other
108 state-of-the-art models. Section. 4 provides the conclusions of this study and future research
109 work.

2 Deep learning-based SR models

111   The study of this paper belongs to Single-Image Super-Resolution (SISR), and SISR
112 refers to reconstructing a HR image from a single LR image. SISR is an ill-posed problem, since
113 one LR image may correspond to multiple HR images (K Zhang et al., 2015). Deep learning-
114 based SR models try to learn a function $\mathbf{F}$ (I^{LR}) through diverse structured neural networks that
115 can obtain a SR image I^{SR} as close as possible to HR image I^{HR} based on an input LR image
116 I^{LR}.

117   All deep learning-based SR models in this paper are Convolutional Neural Networks
118 (CNNs), a type of deep, feedforward networks (LeCun et al., 2015) containing at least one
119 convolutional layer. These CNN SR models can be considered as two parts: on the one hand,
120 various combinations of neural network layers (primarily the activated convolutional layers) with
121 diverse structures extract features from LR images. On the other hand, the upsampler maps and
122 scales the features to the same size as HR images.

123   A general convolutional layer usually requires the definition of two parameters, width F
124 and kernel size $k \times k$. After this convolution layer operation, c feature maps will be convolved by
125 F group filters to F feature maps, where the shape of each filter is $k \times k \times c$. Aiming to develop
126 complex representations, the feature maps output by the convolution layer are usually activated
127 by a nonlinear function. The most simple and popular nonlinear activation function is Rectified
128 Linear Unit (ReLU), with $\alpha=0$ in Equation 1. If α is small and constant, Equation 1 denotes
129 Leaky ReLU. And if α is a learnable parameter, Equation 1 is Parametric ReLU (PReLU).
130 Sigmoid is a smoother activation function, as in Equation 2.

131
$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha x, & \text{if } x \le 0. \end{cases} \tag{1}$$

132
$$f(x) = 1/(1+e^{-x}) \tag{2}$$

133   Currently in upsamplers, sub-pixel convolution layers (Shi et al., 2016) are more widely
134 applied than deconvolution layers (Dong et al., 2016) due to their higher computational

135 efficiency, larger receptive field, and fewer checkerboard artifacts in the generated image. Sub-
136 pixel convolution layer achieves upsampling by convolving to add output channels and then
137 reshaping them. In this layer, assuming the scaling factor is s, an input tensor of size $h×w×c$ will
138 initially be convolved into an output of size $h×w×cs^2$. Then a shuffle operation is performed and
139 the tensor is reshaped to the size of $sh×sw×c$.

2.1 EDSR

141 Based on the SRResNet (Ledig et al., 2017), EDSR (Lim et al., 2017) removes the batch
142 normalization layers (Nah et al., 2017), which greatly reduces GPU memory consumption.
143 Therefore, with the same computational resources, EDSR can boost the number of network
144 layers (depth) to capture richer information and enhance the SR performance. Simply increasing
145 the depth brings the problem of vanishing/exploding gradients and makes the model training
146 more difficult. EDSR effectively prevents these from taking advantage of residual learning (He
147 et al., 2016), i.e., retaining long and short skip connections in SRResNet.


148
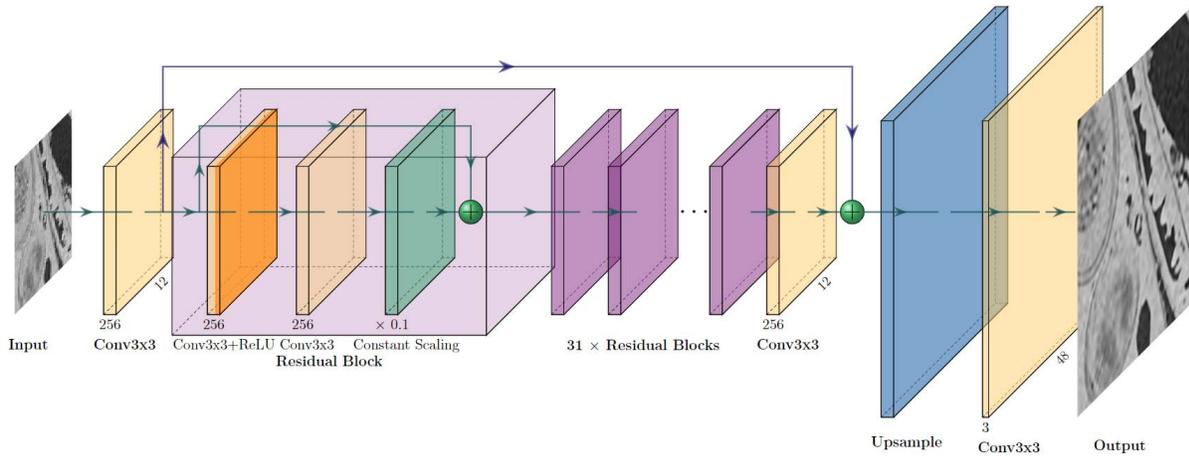
Figure 2. Architecture of the EDSR. The constant scaling layers in residual blocks scale the
features by a certain multiplicity, which can greatly stabilize the training process when the model
is large.

152 The structure of EDSR (as shown in Figure 2) can be summarized in three parts：
153 1) **Width enhancement**. The 3-channel I^{LR} is output by the first convolutional layer as a 256-
154 channel X.
155 2) **Residual group.** X is mapped in order through 32 residual blocks and a convolutional layer as
156 residual $\mathbf{H}(X)$. Then X adds directly to $\mathbf{H}(X)$ via a long skip connection to obtain $\hat{X} = \mathbf{H}(X) + X$.
157 This structure of residual learning is simple yet very effective and provides faster convergence at
158 the early stage (He et al., 2016).
159 3) **Upsampler.** $\hat{X}$ is scaled up to the same size as I^{HR} by the sub-pixel convolution layers and
160 synthesized into a 3-channel SR image I^{SR} by the last convolution layer of $F=3$.

161 In EDSR, the width of all convolutional layers except those in the upsampler is 256
162 ($F=256$). The kernel size of all convolutional layers is 3×3. Each residual block comes with a
163 short skip connection and consists of a convolutional layer with ReLU activation, another
164 convolutional layer, and a constant scaling layer in sequence.

165 2.2 RCAN

166

167
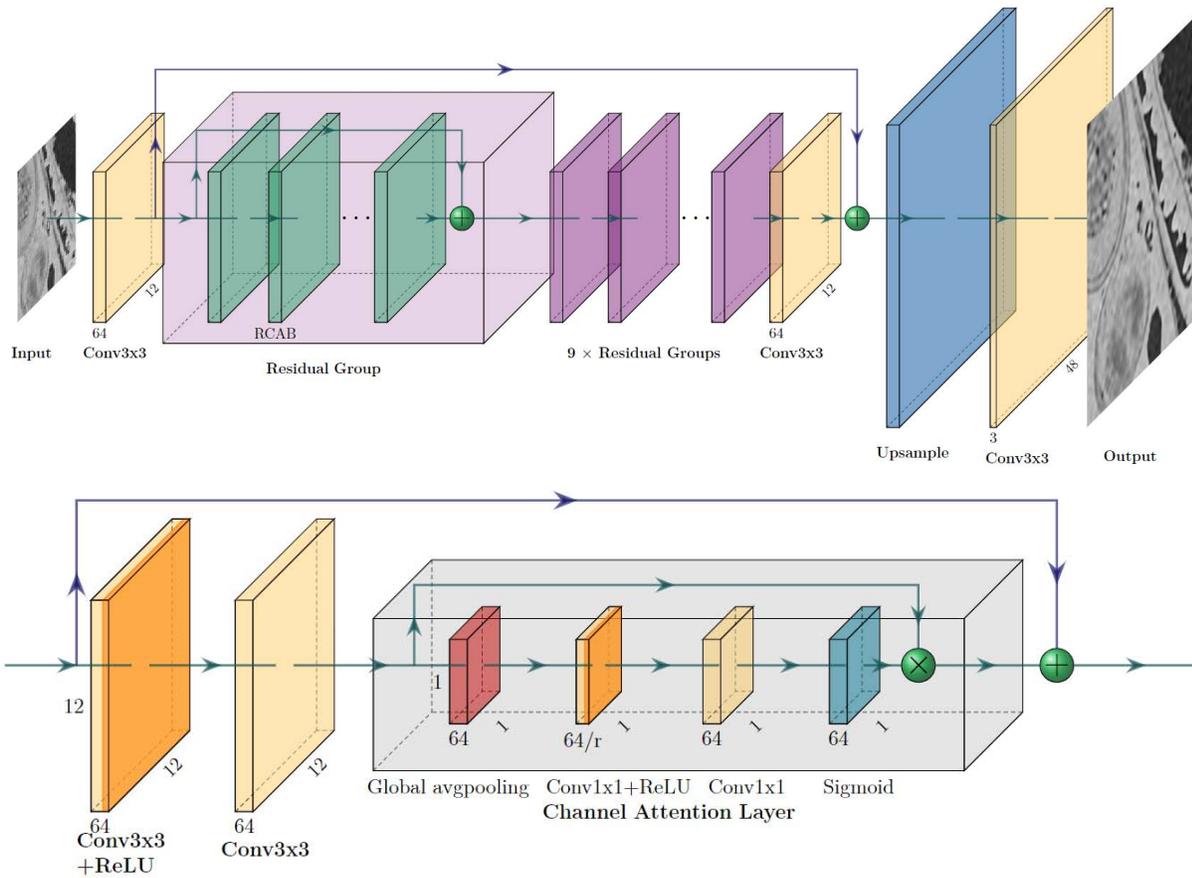
Figure 3. Top: Architecture of the RCAN. Bottom: Architecture of the Residual Channel Attention Blocks (RCABs) in RCAN. In the channel attention layer, feature maps are pooled into 1×1 elements, which are actually learnable and assigned weights to different feature maps. By channel-wise multiplication, the channel attention layer highlights the more valuable feature maps.

As shown in Figure3, the main structure of the RCAN is similar to the EDSR and also has long and short skip connections, which allows the network to bypass abundant low-frequency information and concentrate recovery of high-frequency information (Y Zhang et al., 2018). RCAN's parameters are greatly reduced attributed to its channel attention (CA) mechanism, which is implemented with residual channel attention blocks (RCABs). RCAN utilizes the sub-pixel convolution upsampling and 10 residual groups, with 20 RCABs in each residual group. In RCAN, F=64 and kernel size is 3×3.

A RCAB has a short skip connection, consisting of two convolution layers with ReLU in between and a CA layer. CA layer will process Y of shape $h×w×c$ into c 1×1 elements by global average pooling. The c elements go through a convolution layer with ReLU, another convolution layer and a Sigmoid activation function, and are producted channel-wise with Y to assign different weights to each channel. As a result, RCAB captures information about the interdependencies between different channels and highlights the more valuable features.

186   2.3 MASR

187   2.3.1 Component attention mechanism

188   An image can be decomposed into three components: flat, edges and corners. Flat regions
189   have almost constant pixel values, edges can be regarded as the boundary of different flat
190   regions, and multiple edges interweave into corners (Wei et al., 2020). Generally, flat regions
191   occupy most of the pixels of an image, but they are the least difficult to SR reconstruct, and the
192   main losses are corners that represent details and textures. If the three components are treated
193   homogeneously, the model will be overfitting to recover the easily reconstructed components.
194   CDCSR assigns different attention to different components and drives the model to recover
195   details and edges. In digital rock images SR, we prefer to get sharp edges, clear textures and rich
196   details. The component attention mechanism is exactly right for this task.

197

198
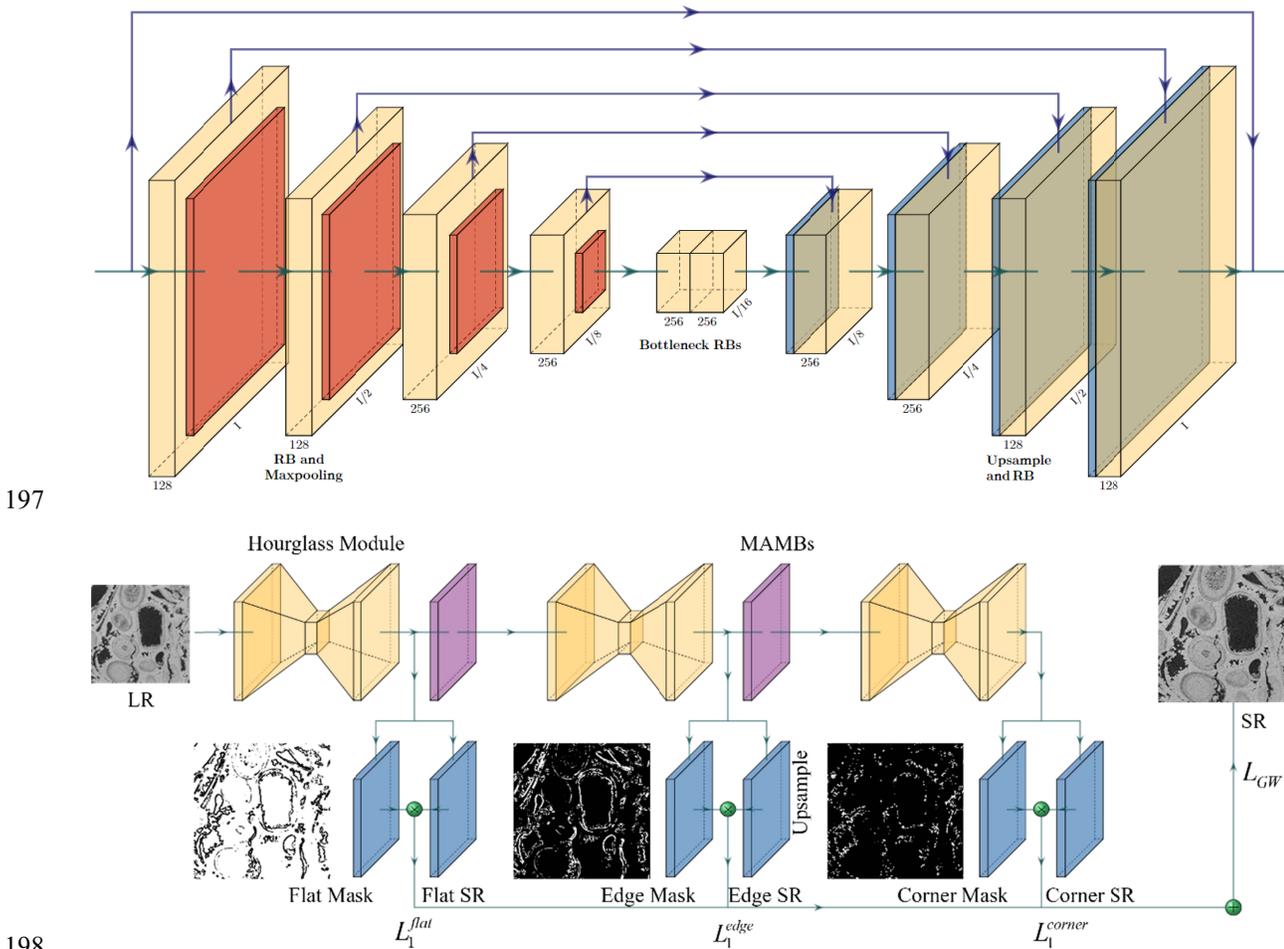
199   **Figure 4**. Top: Architecture of the hourglass module in MASR and CDCSR. Bottom:
200   Architecture of the MASR. Hourglass modules are divided equally into three groups, dealing
201   with flat, edge and corner components. If the number of hourglass modules is not divisible by 3,
202   round down and assign the excess modules to the corner component.

203   Inspired by CDCSR, we propose a novel Multi Attention Super-Resolution Neural
204   Network (MASR). As shown in Figure 4, the backbone network of MASR is several stacked

205 hourglass modules, which are based on CDCSR. An hourglass module can be seen as an
206 encoder-decoder that captures features at different scales. In the encoder, a feature map of shape
207 $h\times w$ is downsampled by the four maximum pooling layers to the size of $h/2^4\times w/2^4$, and each
208 pooling layer is preceded by a residual block (RB). The feature is then fed into the decoder by
209 two RBs. The decoder performs four times nearest neighbor interpolation to restore the feature to
210 the original size of $h\times w$. At the corresponding scale, there is a skip connection between the
211 decoder and the encoder.

212 　　MASR divided the hourglass modules into three component-attentive blocks (CABs),
213 handling flat, edges and corners, respectively. Each CAB incorporates two nearest neighbor
214 interpolation upsamplers. One generates the intermediate SR image I_i, and the other generates a
215 component prediction mask M_i. At the pixel corresponding to I_i, the value of the mask is the
216 probability of component i, where i denotes flat, edge or corner component. And the output of
217 CAB is the element-wise product of M_i and I_i. MASR merges the SR results of the three
218 components to form the final SR image, which can be expressed as

219
$$I^{SR} = I_{flat} \otimes M_{flat} \oplus I_{edge} \otimes M_{edge} \oplus I_{corner} \otimes M_{corner} \tag{3}$$

220 where $\oplus$ and $\otimes$ denote element-wise addition and multiplication.

221 　　In the training stage, giving different weights to each CAB achieves different attention to
222 the three components. MASR uses an Intermediate Supervision (IS) strategy, i.e., the CAB
223 outputs the SR results directly without further input to the subsequent network. IS drives the
224 CAB to focus on the recovery of a particular component, improving SR performance and
225 accelerating convergence.

226 　　2.3.2 Channel and spatial attention mechanism

227 　　To reduce the parameters and further improve the network SR performance, MASR
228 chose the strategy of decreasing the network width and increasing the network depth. We
229 redesign the structure of RB in CDCSR and embed Multi-path Adaptive Modulation Block
230 (MAMB) into MASR to exploit inter-channel and spatial information of feature maps (Kim et
231 al., 2020). The structure of RB and MAMB is illustrated in Figure 5. The activation function in
232 RB is LeakyReLU, and the role of the convolution layer with kernel size 1×1 is to adjust the
233 number of channels. When the number of channels of RB input and output are the same, the
234 identity layer indicates a skip connection, otherwise it performs convolution to match the
235 channels of other path output.

236 　　Since SR aims to recover high-frequency information such as textures and details, and
237 variance is a frequency-related indicator, MAMB adopts global variance pooling to calculate the
238 variance of each feature map. Using stacked convolutional layers to further extract features of
239 the variance, MAMB implements the channel attention mechanism.

240 　　Each feature map has a different texture meaning, and features vary spatially within each
241 channel. For example, some channels require complex filters to extract high-frequency
242 information such as edges and details, while others require simple filters to extract homogeneous
243 flat components representing low-frequency information. In purchase to preserve the
244 characteristics of each channel and extract the spatial information within the channel, MAMB
245 performs independent convolution for each channel, i.e., depth-wise convolution (Howard et al.,
246 2017). MAMB achieves both channel and spatial attention mechanisms through a multi-path
247 attention layer, which is expressed as follows

248
$$\bar{Z} = Z \otimes Sigmoid[\boldsymbol{F}_{var}(Z) \oplus \boldsymbol{F}_{CA}(\boldsymbol{F}_{var}(Z)) \oplus \boldsymbol{F}_{SA}(Z)] \tag{4}$$

249 where Z denotes the feature maps input to attention layer, $\boldsymbol{F}_{var}$, $\boldsymbol{F}_{CA}$ and $\boldsymbol{F}_{SA}$ represent global
250 variance pooling, channel attentional convolution, and depth-wise convolution, respectively,
251 and $\oplus$ and $\otimes$ denote element-wise addition and multiplication.
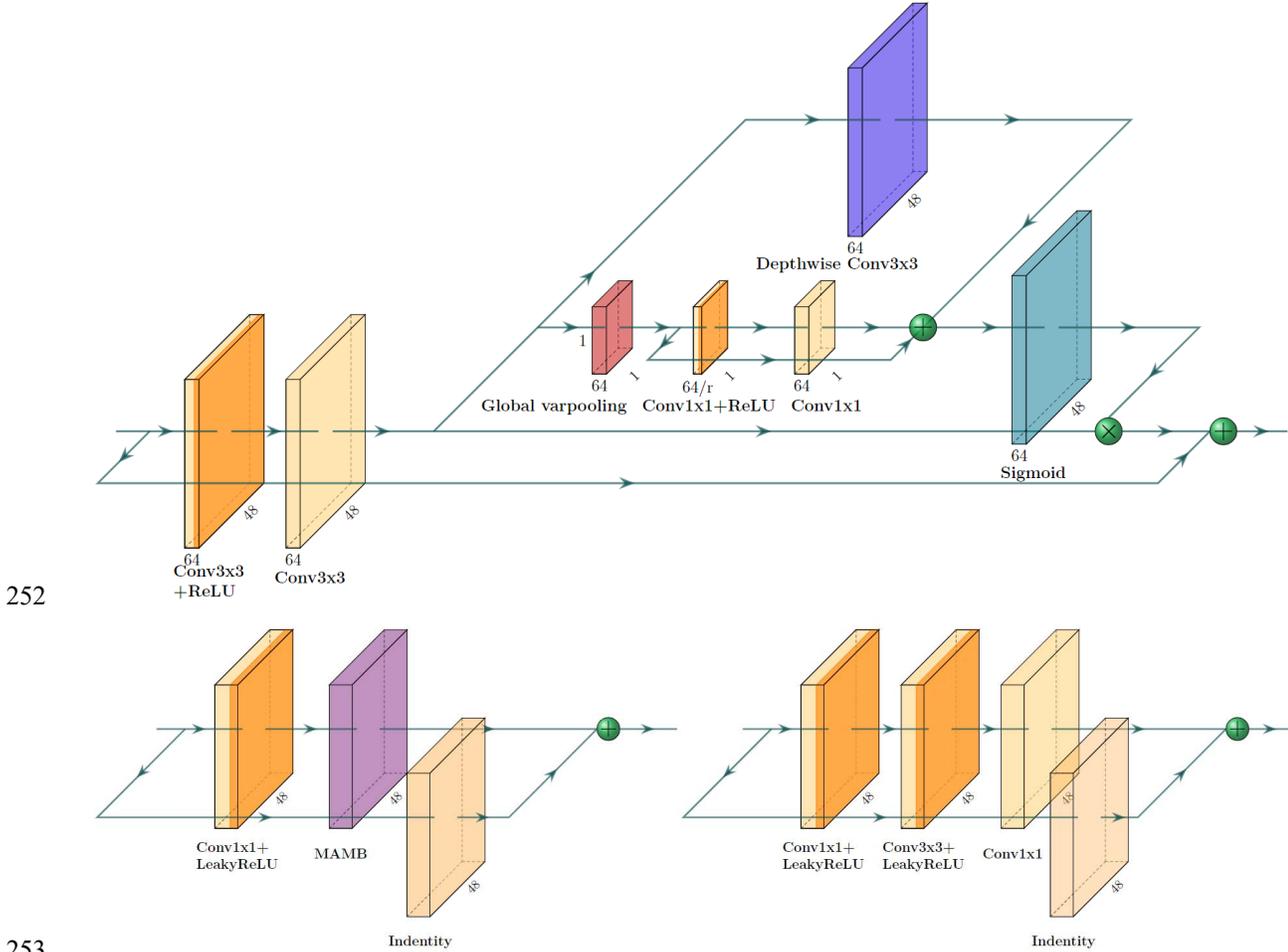
252

253

254 **Figure 5**. Top: Architecture of the MAMB. Bottom left: Architecture of the Residual Blocks
255 (RBs) in MASR. Bottom right: Architecture of the Residual Blocks (RBs) in CDCSR.

256 **Width reduction.** The addition of the multi attention mechanism improves the utilization
257 of information and allows MAMB to reduce the width of the hourglass module. The widths of
258 the four pairs RBs in CDCSR hourglass module are 128, 128, 256 and 256, while in MASR they
259 are set to 96, 96, 128 and 128.

260 **Depth enhancement.** Two Residual Inception Blocks (RIBs) are connected between
261 hourglass modules in CDCSR. As shown in Figure 6, RIBs have a parallel cascade structure and
262 concatenate feature maps produced by filters of different sizes (Szegedy et al., 2017). To
263 improve the depth of CABs and fully utilize IS strategy, MASR replaces RIBs with MAMB
264 group having a long skip connection.

265
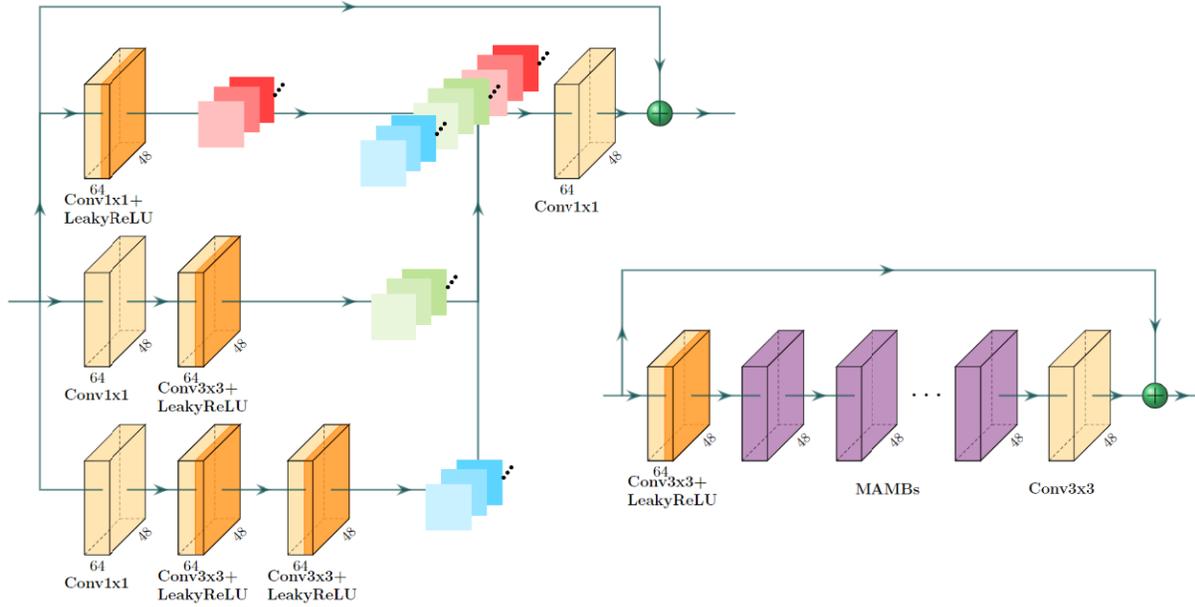
Figure 6. Left: Architecture of the RIBs in CDCSR. Right: Architecture of the RIBs in MASR.

2.3.3 Spatial attention-based mask

268
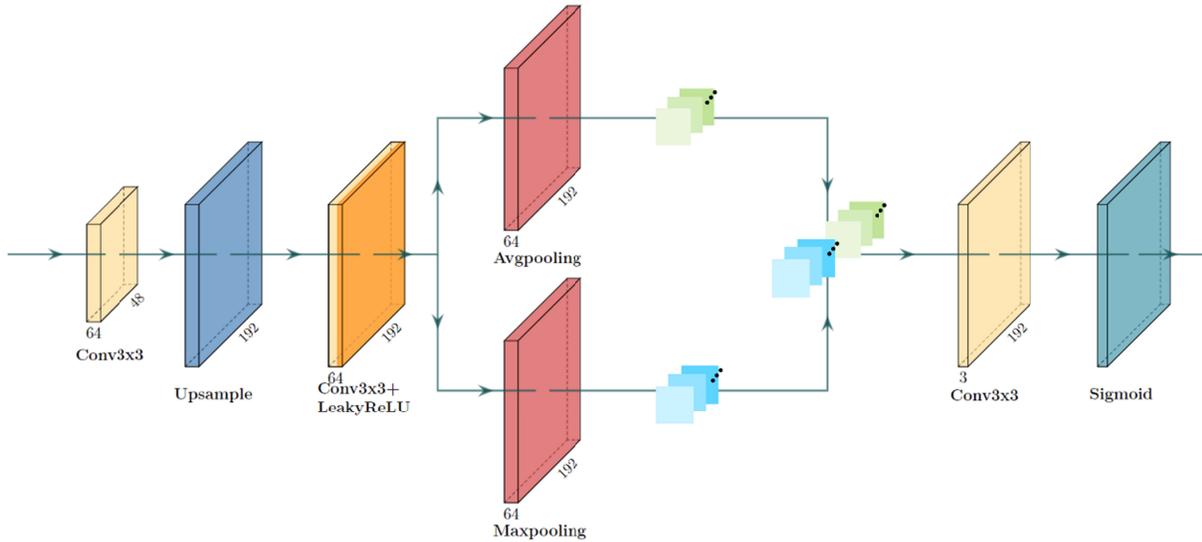
Figure 7. Architecture of proposed Spatial Attention-based Mask (SAM).

We propose a Spatial Attention-based Mask (SAM) that provides higher performance making the application of component attention mechanism more efficient. In Figure 7, the structure of CDCSR mask is the same as the upsampler that generates the intermediate SR results. SAM adds multi-path maximum pooling and average pooling layers and concatenates the feature maps. Finally, Sigmoid activates the convolution result and outputs a mask with values in the 0-1 interval.

276   As in Equation 5, CDCSR normalizes the values of masks with the Softmax function,
277   which magnifies the value of a particular mask, making a pixel value in the final result overly
278   dependent on the intermediate SR result of a particular CAB. SAM directly activates the mask to
279   remove the Softmax, which ensures that the intermediate results of CAB complement each other
280   and strengthen the connection between CABs.

281
$$Softmax(M_i) = e^{M_i} \Big/ \sum_j e^{M_j} \tag{5}$$

282   where M_i denotes the input mask, M_j is flat, edge or corner mask.

283   2.3.4 Loss functions

284   In SR tasks, loss functions are used to calculate image reconstruction error and guide the
285   model optimization (Z Wang et al., 2021). In earlier times, deep learning-based SR models
286   usually chose the pixel-wise L2 loss or mean squared error (MSE). But L2 loss penalizes larger
287   errors and tolerates smaller errors, which causes the SR results to be too smooth (Z Wang et al.,
288   2021). Therefore, EDSR and RCAN employ the pixel-wise L1 loss that is more conducive to
289   improving model performance. L1 loss and L2 loss are calculated as

290
$$L_1(I^{SR}, I^{HR}) = \frac{1}{hwc} \sum_{i,j,k} \left| I^{SR}_{i,j,k} - I^{HR}_{i,j,k} \right| \tag{6}$$

291
$$L_2(I^{SR}, I^{HR}) = \frac{1}{hwc} \sum_{i,j,k} (I^{SR}_{i,j,k} - I^{HR}_{i,j,k})^2 \tag{7}$$

292   where i, j, k denote the pixel in row i and column j on channel k, and h, w, c are the height, width
293   and number of channels of the evaluated images, respectively.

294   To generate sharper images, CDCSR proposes a Gradient-Weighted (GW) loss, is
295   defined as

296
$$\begin{cases} L_{GW} = L_1(D_{GW} \otimes I^{SR}, D_{GW} \otimes I^{HR}) \\ D_{GW} = (1+\alpha \left| G^{SR}_x - G^{HR}_x \right|)(1+\alpha \left| G^{SR}_y - G^{HR}_y \right|) \end{cases} \tag{8}$$

297   where $\left| G^{SR}_x - G^{HR}_x \right|$, $\left| G^{SR}_y - G^{HR}_y \right|$ represent gradient difference maps between SR and HR in the

298   horizontal and vertical directions, α is a scalar weight, in this paper, $\alpha = 4$, $\oplus$ and $\otimes$ denote
299   element-wise addition and multiplication.

300   CDCSR and MASR use IS strategy, in addition to GW loss, we also need to calculate the
301   reconstruction error of SR results for each component, we use L1 loss as the IS criterion, and the
302   total loss is

303
$$L = L_{GW} + \sum_i \alpha_i L_1(I^{SR}_i \otimes M^{HR}_i, I^{HR}_i \otimes M^{HR}_i) \tag{9}$$

304   where I^{SR}_i represents the SR result for flat, edge or corner components, and M^{HR}_i is the
305   corresponding mask of I^{SR}_i, α_i is the component attention weight, the weights of the flat, edge
306   and corner are 1, 2, 5 in our experiments, $\otimes$ denotes element-wise multiplication.

307 **3 Experiments**

308   3.1 Dataset

309   All deep learning-based SR models in this paper are trained on the DeepRock-SR 2D
310 dataset (Da Wang et al., 2019), which contains 4000 HR digital rock CT images each of
311 sandstone, carbonate and coal at 500×500 pixels. And image resolution ranges from 2.7 to 25
312 μm. The dataset is split into training, validation and test sets with 8:1:1 ratio. The degradation of
313 HR images to LR images in the real world is very complex and unknown. To simulate the real
314 situation as much as possible, all LR images are generated by ×4 downsampling from HR images
315 with random kernels (box, triangle, lanczos2, or lanczos3), i.e., the size of LR images is 125×125
316 pixels.

317   3.2 SR quality measurements

318   Peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) are the most
319 widely used evaluation criteria in the SR reconstruction field (Z Wang et al., 2021). PSNR is
320 defined as

$$321 \qquad PSNR = 10 \times \log_{10}(\frac{L^2}{MSE}) \qquad (10)$$

322 where MSE is the mean squared error, or the pixel-wise L2 loss, as shown in Equation 6, L
323 indicates the maximum value of pixels in the image, usually L=255. The larger the value of
324 PNSR, the better the quality of SR reconstruction.

325   SSIM is proposed taking into account the human visual system, based on independent
326 comparisons of image luminance, contrast, and structures. For an image I, the luminance μ_I and
327 contrast σ_I are estimated as the mean and standard deviation of the image intensity,
328 respectively. Given two images x and y, SSIM is calculated as

$$329 \qquad SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \qquad (11)$$

330 where σ_{xy} is the covariance of x and y, $C_1 = (k_1 L)^2$ and $C_2 = (k_2 L)^2$ are constants used to maintain

331 stability, L is the dynamic range of pixel values, k_1=0.01 and k_2=0.03. The closer SSIM is to 1,
332 the more similar x and y are.

333   3.3 Experimental settings

334   The experiments are conducted on a high-performance computing cluster node containing
335 forty Intel(R) Xeon(R) Gold 5218R CPUs @ 2.10GHz, two NVIDIA Tesla V100 32GB GPUs
336 and 376GB RAM. The software environment consists of Red Hat Enterprise Linux Server
337 release 7.8 (Maipo) OS, CUDA 11.0, and the deep learning framework Pytorch 1.8.

338 **Table 1.** Hyperparameter settings for different SR models.

Model	Patch size	Batch size	Initial learning rate	Decay epoch	Optimizer
EDSR	12×12		1×10^{-4}		Adam
RCAN	12×12	32	1×10^{-4}	160,200,230	β_1=0.9
CDCSR	48×48		2×10^{-4}		
MASR	48×48		2×10^{-4}		β_2=0.999

339   For training, MASR uses patches of size 48×48 cropped from random positions on LR
340 images as input, with the corresponding HR patches as ground truth. MASR is trained using the
341 Adam optimizer with exponential decay rates set to $\beta_1=0.9$ and $\beta_2=0.999$. The learning rate is
342 initialized to 2×10^{-4}, and halved at the {160, 200, 230}-th epoch. And training lasts 250 epochs
343 with batch size of 32. The hyperparameter settings of other models are summarized in Table 1.

344   3.4 Experimental results

345   3.4.1 Network Depth

346   In general, the deeper the model the better it is at extracting complex features, but it also
347 means that the model has more parameters and consumes more memory and time. Therefore, we
348 provide experimental evaluations to determine the appropriate depth of MASR. The network
349 depth is a hyperparameter, so the experimental evaluation of it is performed on the validation set.

350 **Table 2.** Performance comparison of the number of hourglass modules in MASR for training
351 250 epochs. Bold indicates optimal performance and underline indicates suboptimal
352 performance.

Model	Hourglass modules	PSNR (dB)	SSIM	Training time (hours)	Number of parameters
	3	33.5242	0.7366	12.6	13.5M
MA	4	33.5999	0.7366	15.0	18.1M
SR	6	**33.6145**	<u>0.7368</u>	19.8	27.3M
	8	**33.6145**	**0.7369**	25.3	36.5M

353   **Evaluation on hourglass modules.** MASR is constructed by connecting hourglass
354 modules in series, and the number of hourglass modules determines the depth of MASR. For
355 setting the number of MAMBs between two hourglass modules to 16, the experimental
356 evaluation on the number of hourglass modules is shown in Table 2. There is a significant
357 improvement in SR performance of the model when the number of hourglass modules is
358 increased from 3 to 6, but the performance improvement of the model is weak when the number
359 of hourglass modules is increased to 8. In addition, the computational complexity and the
360 number of parameters increase almost linearly with the model depth. If the number of hourglass
361 modules increases from 3 to 8, the training time becomes nearly 2 times longer. Thus, with a
362 trade-off between model performance and speed, the number of hourglass modules in MASR is
363 set to 6.

364 **Table 3.** Performance comparison of the number of MAMBs in MASR for training 250 epochs.
365 Bold indicates optimal performance and underline indicates suboptimal performance.

Model	MAMBs	PSNR(dB)	SSIM	Training time (hours)	Number of parameters
	8	33.5770	<u>0.7371</u>	16.5	24.3M
	12	33.6081	<u>0.7371</u>	18.1	25.8M
MASR	16	33.6145	0.7368	19.8	27.3M
	20	**33.6210**	**0.7372**	21.5	28.8M
	24	<u>33.6196</u>	0.7370	23.3	30.3M

366   **Evaluation on MAMBs.** Another major factor affecting MASR depth is the number of
367 MAMBs placed between every two hourglass modules. Setting up 6 hourglass modules, the
368 effect of the number of MAMBs on the model performance is shown in Table 3. When the
369 number of MAMBs is increased from 8 to 20, the model performance is consistently enhanced
370 bringing a 0.044dB PNSR improvement. On the contrary, when the number of MAMBs turns to

371 20, the model performance slightly decreases. This is because MASR learns special features on a
372 limited training set and overfitting occurs, resulting in poor generalization on the validation set.
373 Hence, the appropriate amount of MAMB is selected as 20.

374   3.4.2 Comparisons with state-of-the-art models

375   We compare MASR with other state-of-the-art SR algorithms, including EDSR, RCAN,
376 and CDCSR. The features between the digital rock images and the photographs differ
377 significantly, i.e., there is a domain gap between the data. Therefore, the pretrained models on
378 the photo should not be used directly for digital rock images SR, and these models need to be
379 retrained for comparison.

380 **Table 4.** Comparison of the number of parameters and the training time consumed by training
381 250 epochs between different models.

Model	Training time (hours)	Number of parameters
EDSR	3.1	43.1M
RCAN	10.5	16.5M
CDCSR	26.6	39.9M
CDCSR (Optimized)	11.8	39.9M
MASR	21.5	28.8M

382   The training time and the number of parameters of the models in this paper are listed in
383 Table 4. Compared with EDSR, the parameters of RCAN are greatly reduced, but introducing
384 the attention mechanism increases computational complexity and slows down the training speed
385 to 0.3 times. CDCSR takes the Harris Corner detection method (Harris and Stephens, 1988) to
386 compute component masks once per iteration, which makes CDCSR even less efficient. We
387 optimize the CDCSR algorithm by saving the component masks before training so that there is
388 no need to repeat the computation during training process. The optimized CDCSR is more than
389 twice as efficient, and even after adding multiple attention mechanisms (i.e., MASR) training is
390 still faster than the original CDCSR.

391 **Table 5.** Performance comparison of different models on digital rock images test sets. Bold
392 indicates optimal performance and underline indicates suboptimal performance.

Model	Coal		Sandstone		Carbonate	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
Bicubic	41.5398	0.9443	27.3421	0.593	24.4485	0.4901
EDSR	44.3677	0.9591	29.9944	0.6684	25.7138	0.5663
RCAN	<u>44.4417</u>	<u>0.9594</u>	<u>30.0368</u>	0.6731	25.7838	0.5750
CDCSR	44.2302	0.9585	30.0235	<u>0.6736</u>	<u>25.8305</u>	<u>0.5780</u>
MASR	**44.5157**	**0.9597**	**30.1527**	**0.6749**	**25.8675**	**0.5792**

393   Considering the different difficulties in SR recovery for various types of digital rock
394 images, we verified the SR performance of each model on separate test sets for coal, carbonate
395 and sandstone. The results of quantitative comparison are shown in Table 5 and Figure 8. The
396 multiple attention mechanism improves the information utilization, so MASR achieves the best
397 and most stable performance on different types of digital rock images with 72% of the CDCSR
398 parametric number. Compared with the suboptimal model, the average PSNR of MASR

399   improves by 0.074 dB, 0.1159 dB, 0.037 dB, and the average SSIM improves by 0.0003, 0.0013,
400   and 0.0012, for coal, sandstone, and carbonate test sets, respectively. All deep learning-based SR
401   models are remarkably superior to the traditional bicubic interpolation method. On coal,
402   sandstone and carbonate images, the pixelwise relative errors of MASR reconstructions are
403   reduced by 20%, 26% and 15% over bicubic interpolation.
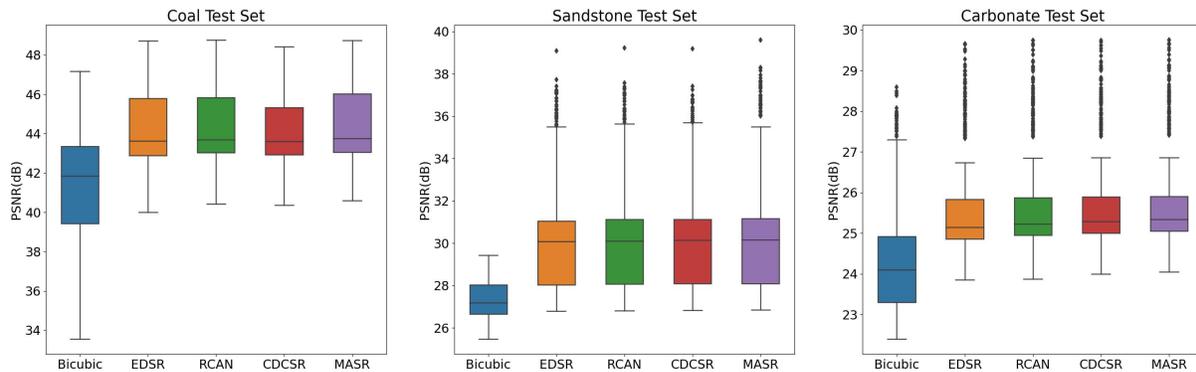
404

405   **Figure 8**. Boxplots of the average PSNR of EDSR, RCAN, CDCSR and MASR on coal,
406   sandstone and carbonate test sets.

407   Figure 9 and Figure 10 visualize the SR results for coal and sandstone images. Coal and
408   sandstone images have simple textures, hence various deep learning-based SR algorithms are
409   able to recover high-quality features, and the performance gap between them does not seem to be
410   as large as assessed by objective metrics. Nevertheless, MASR recovers sharper pore edges and
411   more consistent details with ground truth. It is observed from Figure 11 that the performance
412   superiority of MASR is more prominent for carbonate rock images with more complex texture
413   and noise interference. Additionally, it is demonstrated by subjective evaluation that the deep
414   learning-based SR algorithm not only recovers sharp edges and details, but also has natural
415   smoothness to remove noise, which is exactly the SR result we desire for digital rock images.
416

417   **4 Conclusions**

418   We propose a MASR model to exceed hardware limitation and acquire CT digital rock
419   images with wide FOV and HR. By redesigning the hourglass network and proposing a spatial
420   attention-based mask, MASR integrates component, channel, and spatial attention mechanisms.
421   To avoid overfitting and to trade-off SR accuracy and training speed, we explore the appropriate
422   network depth for MASR through experimental evaluations, including the number of hourglass
423   modules and MAMBs. The subjective and objective evaluations on coal, sandstone and
424   carbonate images verify that MASR has higher SR reconstruction accuracy than other advanced
425   SR models. And MASR recovers sharper edges and more accurate textures while removing
426   noise. We optimize the process of calculating the masks and introduce multiple attention
427   mechanisms to enhance the ability for feature extraction, hence MASR consumes less time and
428   memory in the training stage than CDCSR.

429   If MASR is directly extended to SR of 3D digital rocks, the parameters and training time
430   of the model will become unacceptable. In the future study, we will further optimize the
431   efficiency of deep learning-based models to achieve feasible 3D SR reconstructions.

432

433
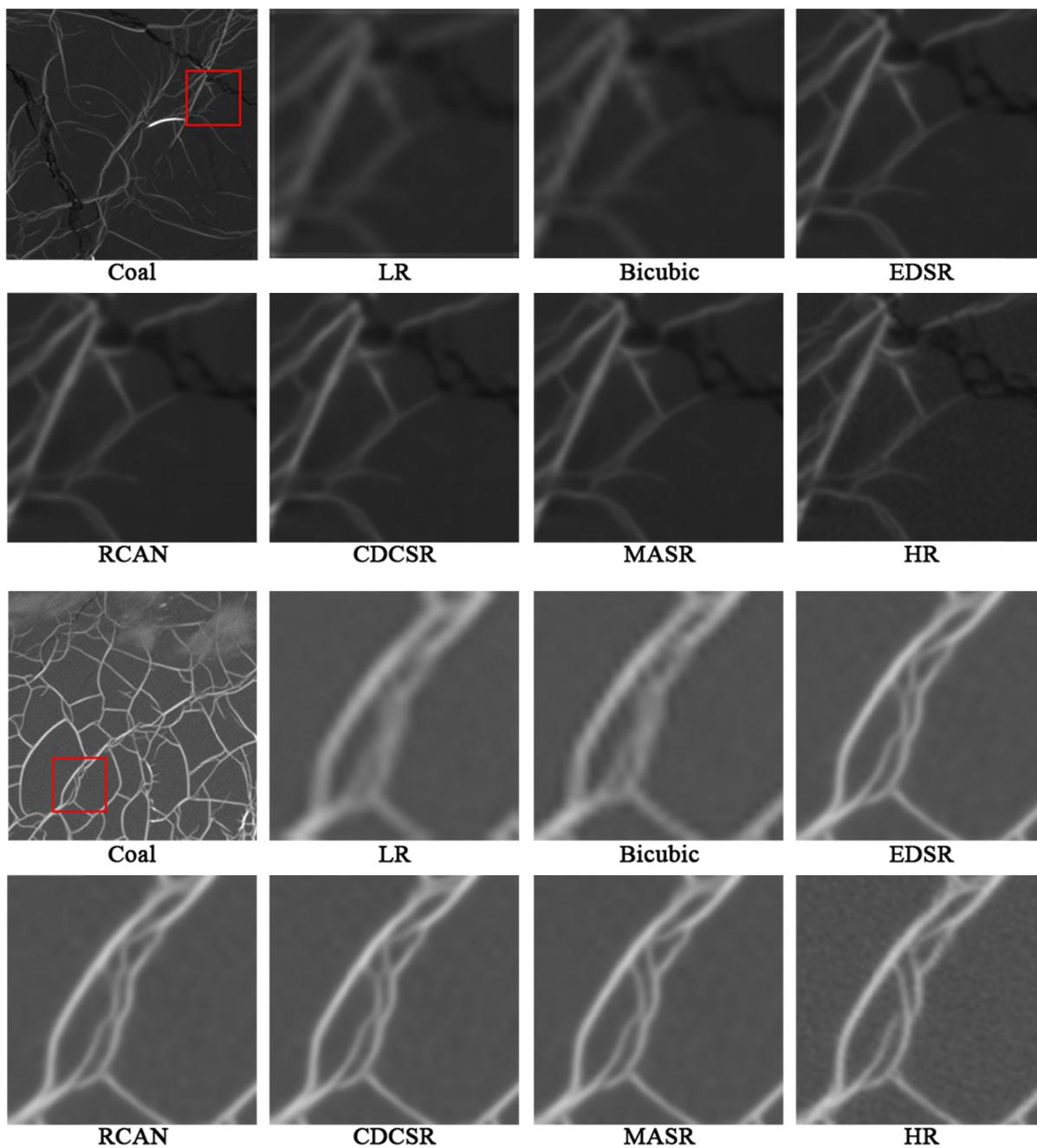
Figure 9. Qualitative comparison of our model with other works at ×4 super-resolution on the coal test set. Compared to other state-of-the-art models, our models recover more accurate textures and sharper pore edges.

437

438
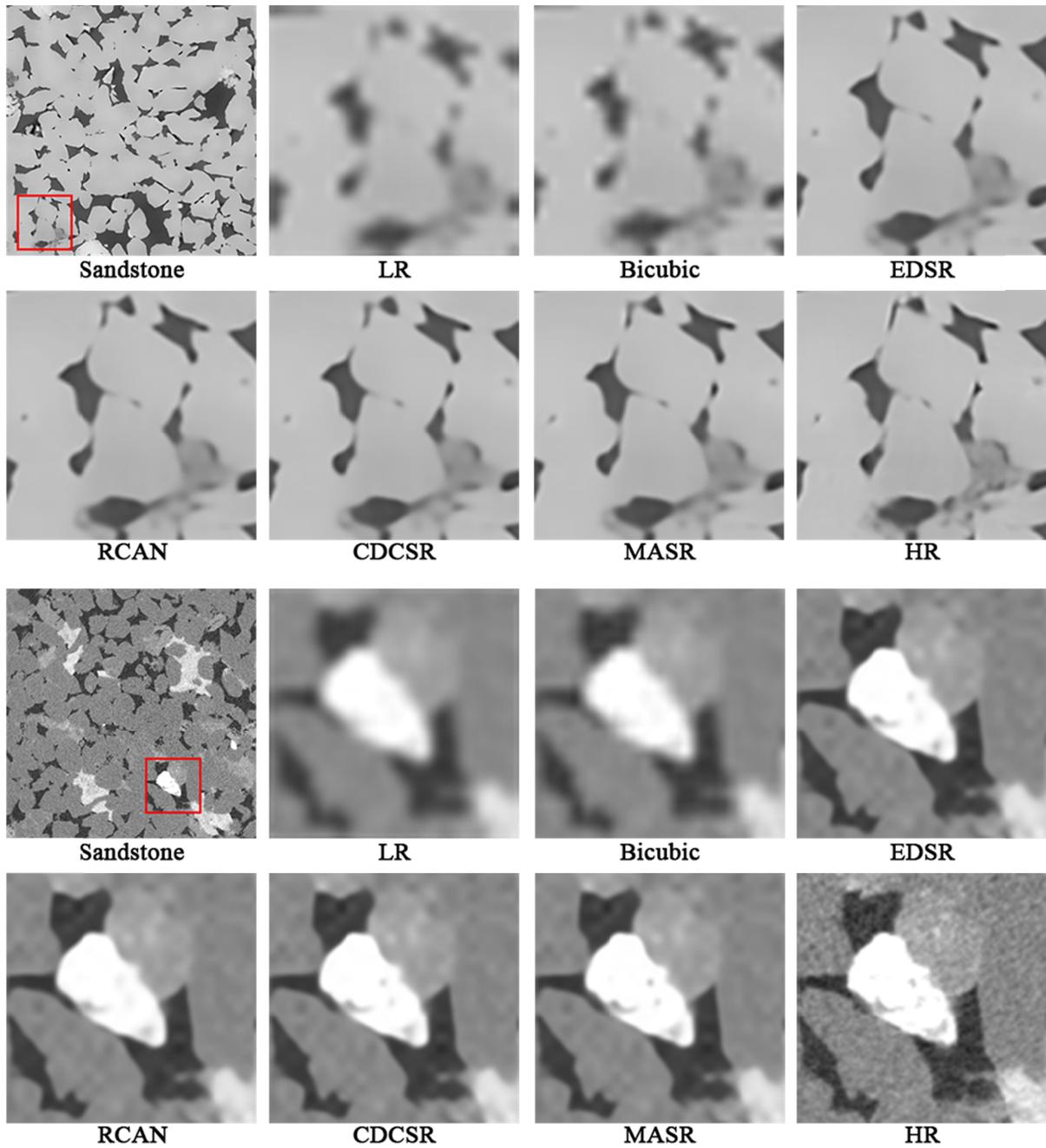
Figure 10. Qualitative comparison of our model with other works at ×4 super-resolution on the sandstone test set. Compared with other works, our model recovers the pore edges more sharply and resolves the features more clearly.

442

443
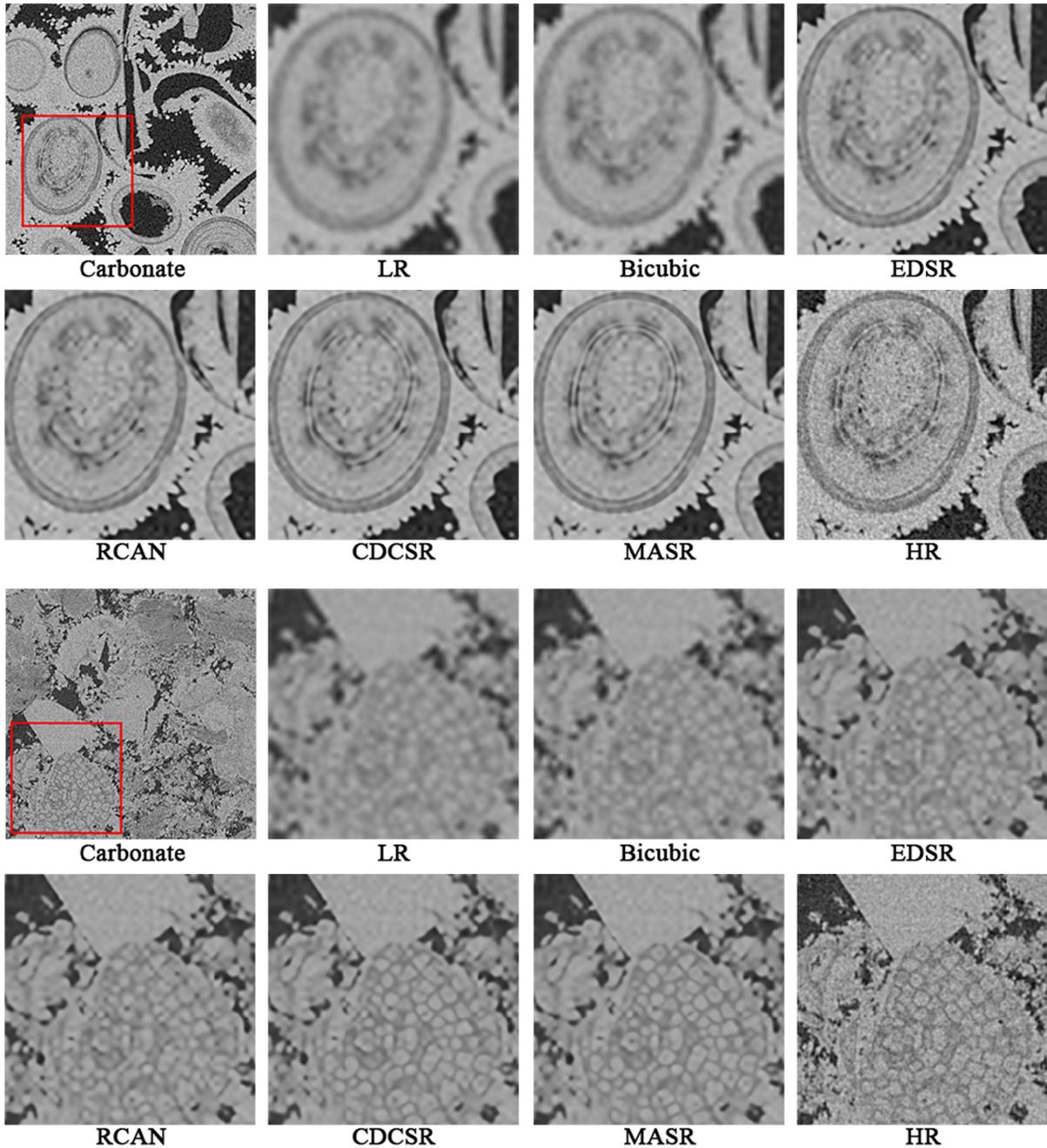
Figure 11. Qualitative comparison of our model with other works at ×4 super-resolution on the carbonate test set. Deep learning-based SR models are naturally smooth and able to remove noise. Benefiting from this, MASR recovers textures that are even more prominent than HR images.

452 **Data Availability Statement**

453 Our codes are publicly available at https://github.com/MHDXing/MASR-for-Digital-Rock-

454 Images. The dataset for this research is derived from the open source project DeepRock-SR (Da

455 Wang et al., 2019) via https://digitalrocks-dev.tacc.utexas.edu/projects/215.

456 **References**

457 Chung, T., Y. D. Wang, R. T. Armstrong, and P. Mostaghimi (2019), Approximating

458 permeability of microcomputed-tomography images using elliptic flow equations, *SPE Journal*,

459 *24*(03), 1154-1163.

460 Da Wang, Y., P. Mostaghimi, and R. Armstrong (2019), A diverse super resolution dataset of

461 sandstone, carbonate, and coal (deeprock-sr), edited.

462 Dong, C., C. C. Loy, and X. Tang (2016), Accelerating the super-resolution convolutional neural

463 network, paper presented at European conference on computer vision, Springer.

464 Dong, C., C. C. Loy, K. He, and X. Tang (2014), Learning a deep convolutional network for

465 image super-resolution, paper presented at European conference on computer vision, Springer.

466 Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and

467 Y. Bengio (2020), Generative adversarial networks, *Communications of the ACM*, *63*(11), 139-

468 144.

469 Harris, C., and M. Stephens (1988), A combined corner and edge detector, paper presented at

470 Alvey vision conference, Manchester, UK.

471   He, K., X. Zhang, S. Ren, and J. Sun (2016), Deep residual learning for image recognition, paper

472   presented at Proceedings of the IEEE conference on computer vision and pattern recognition.

473   Howard, A. G., M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H.

474   Adam (2017), Mobilenets: Efficient convolutional neural networks for mobile vision

475   applications, *arXiv preprint arXiv:1704.04861*.

476   Iglauer, S., and M. Lebedev (2018), High pressure-elevated temperature x-ray micro-computed

477   tomography for subsurface applications, *Advances in Colloid and Interface Science*, *256*, 393-

478   410.

479   Kim, J.-H., J.-H. Choi, M. Cheon, and J.-S. Lee (2020), MAMNet: Multi-path adaptive

480   modulation network for image super-resolution, *Neurocomputing*, *402*, 38-49.

481   LeCun, Y., Y. Bengio, and G. Hinton (2015), Deep learning, *nature*, *521*(7553), 436-444.

482   Ledig, C., L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J.

483   Totz, and Z. Wang (2017), Photo-realistic single image super-resolution using a generative

484   adversarial network, paper presented at Proceedings of the IEEE conference on computer vision

485   and pattern recognition.

486   Li, Z., Q. Teng, X. He, G. Yue, and Z. Wang (2017), Sparse representation-based volumetric

487   super-resolution algorithm for 3D CT images of reservoir rocks, *Journal of Applied Geophysics*,

488   *144*, 69-77.

489   Lim, B., S. Son, H. Kim, S. Nah, and K. Mu Lee (2017), Enhanced deep residual networks for

490   single image super-resolution, paper presented at Proceedings of the IEEE conference on

491   computer vision and pattern recognition workshops.

492   Liu, L., J. Yao, H. Sun, Zhang，Lei, and Yang，Yongfei (2018), Reconstruction of digital rock

493   considering micro-fracture based on multi-point statistics, *Chin Sci Bull*, *63*(30), 3146-3157.

494 Mostaghimi, P., M. J. Blunt, and B. Bijeljic (2013), Computations of absolute permeability on

495 micro-CT images, *Mathematical Geosciences*, *45*(1), 103-125.

496 Nah, S., T. Hyun Kim, and K. Mu Lee (2017), Deep multi-scale convolutional neural network

497 for dynamic scene deblurring, paper presented at Proceedings of the IEEE conference on

498 computer vision and pattern recognition.

499 Oluwadebi, A. G., K. G. Taylor, and L. Ma (2019), A case study on 3D characterisation of pore

500 structure in a tight sandstone gas reservoir: the Collyhurst Sandstone, East Irish Sea Basin,

501 northern England, *Journal of Natural Gas Science and Engineering*, *68*, 102917.

502 Rahiman, V. A., and S. N. George (2017), Single image super resolution using neighbor

503 embedding and statistical prediction model, *Computers & Electrical Engineering*, *62*, 281-292.

504 Shi, W., J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang

505 (2016), Real-time single image and video super-resolution using an efficient sub-pixel

506 convolutional neural network, paper presented at Proceedings of the IEEE conference on

507 computer vision and pattern recognition.

508 Szegedy, C., S. Ioffe, V. Vanhoucke, and A. A. Alemi (2017), Inception-v4, inception-resnet and

509 the impact of residual connections on learning, paper presented at Thirty-first AAAI conference

510 on artificial intelligence.

511 Tekalp, A. M., M. K. Ozkan, and M. I. Sezan (1992), High-resolution image reconstruction from

512 lower-resolution image sequences and space-varying image restoration, paper presented at

513 [Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and

514 Signal Processing, IEEE.

515   Wang, X., K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy (2018), Esrgan:

516   Enhanced super-resolution generative adversarial networks, paper presented at Proceedings of

517   the European conference on computer vision (ECCV) workshops.

518   Wang, Y. (2018), Reconstruction of high-resolution pore structures from micro-tomography

519   images, UNSW Sydney.

520   Wang, Y., S. S. Rahman, and C. H. Arns (2018), Super resolution reconstruction of μ-CT image

521   of rock sample using neighbour embedding algorithm, *Physica A: Statistical Mechanics and its*

522   *Applications*, *493*, 177-188.

523   Wang, Y., Q. Teng, X. He, J. Feng, and T. Zhang (2019), CT-image of rock samples super

524   resolution using 3D convolutional neural network, *Computers & Geosciences*, *133*.

525   Wang, Y. D., R. T. Armstrong, and P. Mostaghimi (2019), Enhancing Resolution of Digital

526   Rock Images with Super Resolution Convolutional Neural Networks, *Journal of Petroleum*

527   *Science and Engineering*, *182*.

528   Wang, Y. D., R. T. Armstrong, and P. Mostaghimi (2020), Boosting Resolution and Recovering

529   Texture of 2D and 3D Micro-CT Images with Deep Learning, *Water Resources Research*, *56*(1).

530   Wang, Z., J. Chen, and S. C. H. Hoi (2021), Deep Learning for Image Super-Resolution: A

531   Survey, *IEEE Trans Pattern Anal Mach Intell*, *43*(10), 3365-3387.

532   Wei, P., Z. Xie, H. Lu, Z. Zhan, Q. Ye, W. Zuo, and L. Lin (2020), Component divide-and-

533   conquer for real-world image super-resolution, paper presented at European Conference on

534   Computer Vision, Springer.

535   Wildenschild, D., and A. P. Sheppard (2013), X-ray imaging and analysis techniques for

536   quantifying pore-scale structure and processes in subsurface porous medium systems, *Advances*

537   *in Water resources*, *51*, 217-246.

538 Yang, J., J. Wright, T. S. Huang, and Y. Ma (2010), Image super-resolution via sparse

539 representation, *IEEE transactions on image processing*, *19*(11), 2861-2873.

540 Yao, J., X. Zhao, Y. Yi, and J. Tao (2005), The current situation and prospect on digital core

541 technology, *Petroleum Geology and Recovery Eficiency*, *12*(6), 52-54.

542 Yu, J., Y. Fan, J. Yang, N. Xu, Z. Wang, X. Wang, and T. Huang (2018), Wide activation for

543 efficient and accurate image super-resolution, *arXiv preprint arXiv:1808.08718*.

544 Zhang, K., D. Tao, X. Gao, X. Li, and Z. Xiong (2015), Learning multiple linear mappings for

545 efficient single image super-resolution, *IEEE Transactions on Image Processing*, *24*(3), 846-861.

546 Zhang, Y., K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu (2018), Image super-resolution using

547 very deep residual channel attention networks, paper presented at Proceedings of the European

548 conference on computer vision (ECCV).

549

Supporting Information for

Multi Attention Neural Network for Digital Rock CT Images Super-Resolution

Zhihao Xing[1], Jun Yao[1], Lei Liu[1] and Hai Sun[1]

[1]Research Centre of Multiphase Flow in Porous Media, China University of Petroleum (East China), Huangdao District, Qingdao, China.

Contents of this file

Text S1
Figures S1 to S9

Introduction

- Experiments and algorithm details;
- Sample slices of training dataset;
- Data augmentation method in this research;
- The training log of the super-resolution models;
- Subplots of Figure 8 in the main article (Figures S4 to S6);
- Additional qualitative comparison on the coal test set and their difference maps;
- Additional qualitative comparison on the sandstone test set and their difference maps;
- Additional qualitative comparison on the carbonate test set and their difference maps.

Text S1.

Our codes are publicly available at https://github.com/MHDXing/MASR-for-Digital-Rock-Images. All experiments and algorithm details are in this open source project.

Sheared Coal   Naturally Fractured Coal   Leopard Sandstone   Glidehauser Sandstone   Savonnieres Carbonate   Massangis Carbonate
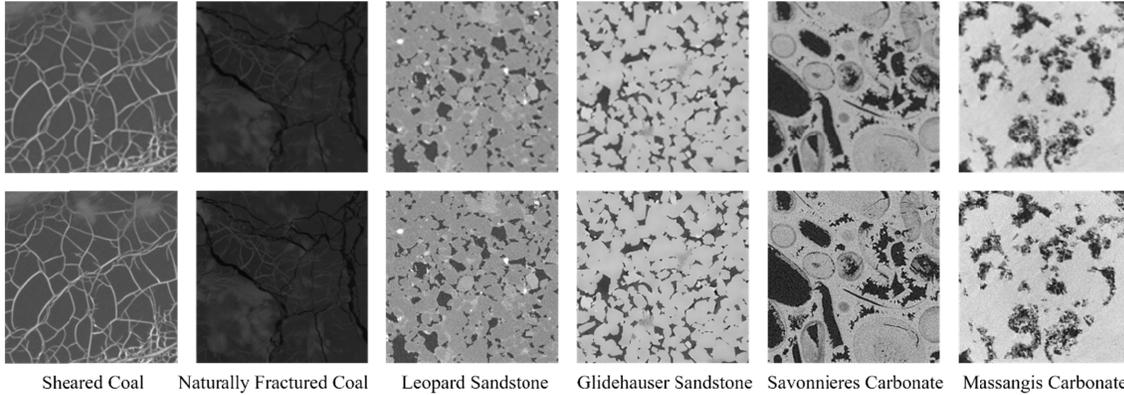
Figure S1. Sample slices of training dataset. First row: Low resolution images, i.e., inputs to the model. Second row: Corresponding high resolution images, i.e. ground truth. In the training stage, the model calculates the loss between the output and the ground truth, and optimizes the parameters by back propagation algorithm to make the output as close as possible to the ground truth

No operation   Rotate 90 degrees   Flip horizontally   Flip vertically

Flip horizontal +Flip vertically   Rotate 90 degrees + Flip vertically   Rotate 90 degrees +Flip horizontally   Flip horizontally+ Rotate 90 degrees+Flip vertically
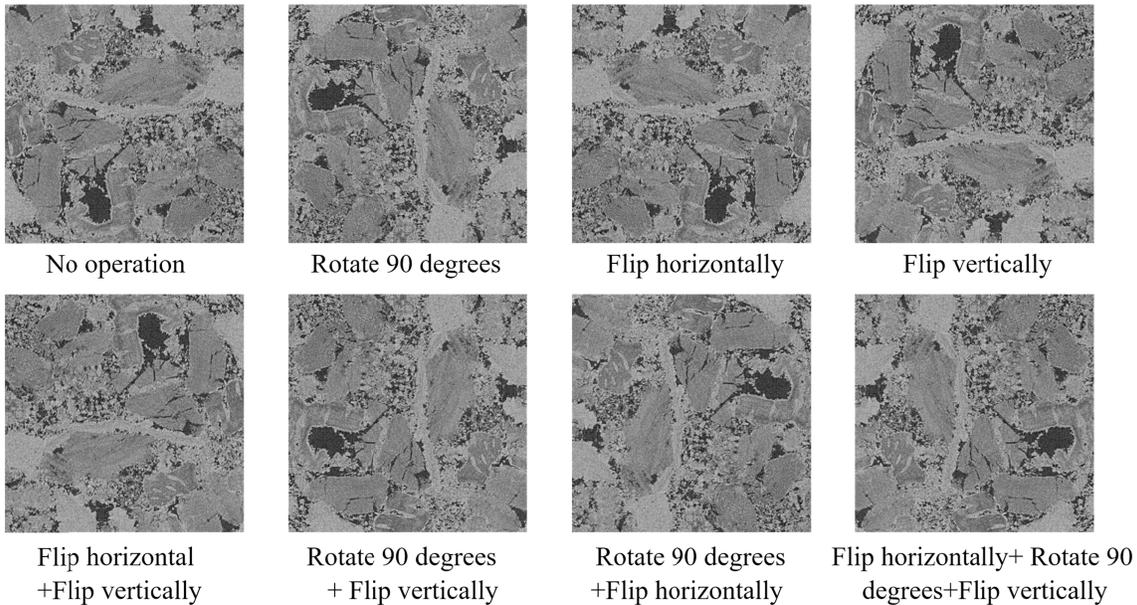
Figure S2. Data augmentation method in this research. In the training stage, the inputs are randomly flipped horizontally, flipped vertically or rotated to enhance the generalization of the model.

41
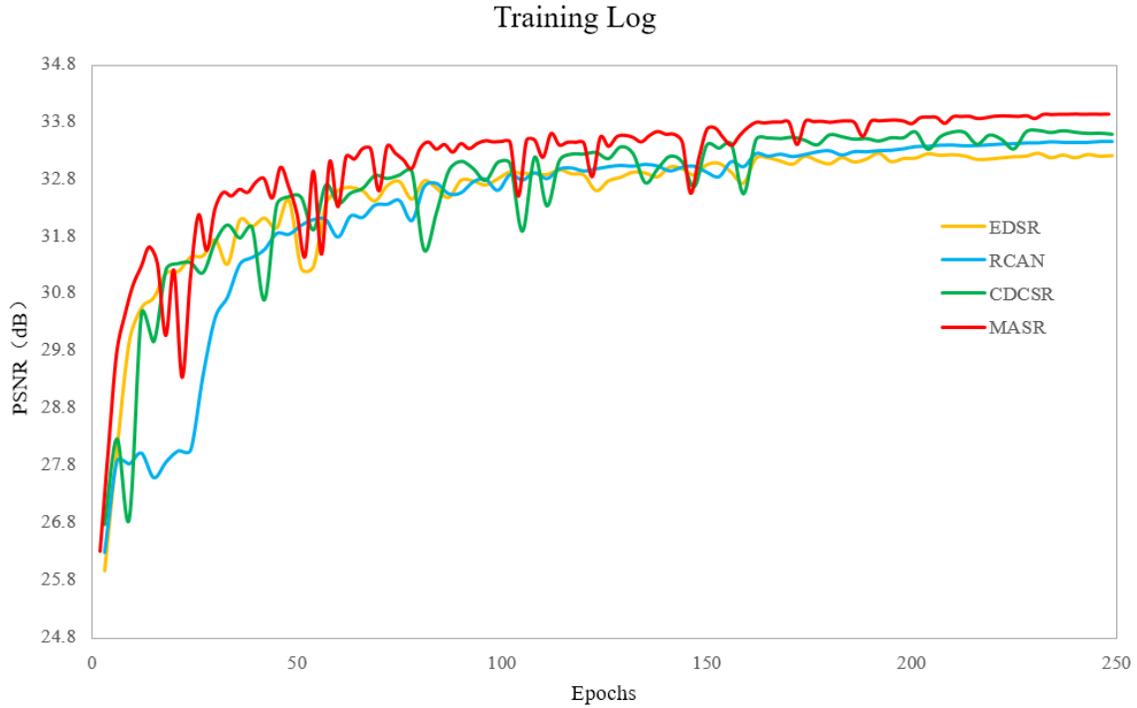
Figure S3. The average Peak Signal-to-Noise Ratio (PSNR) of the super-resolution models on the validation set during training. Our proposed model (MASR) converges faster and better.

44

45
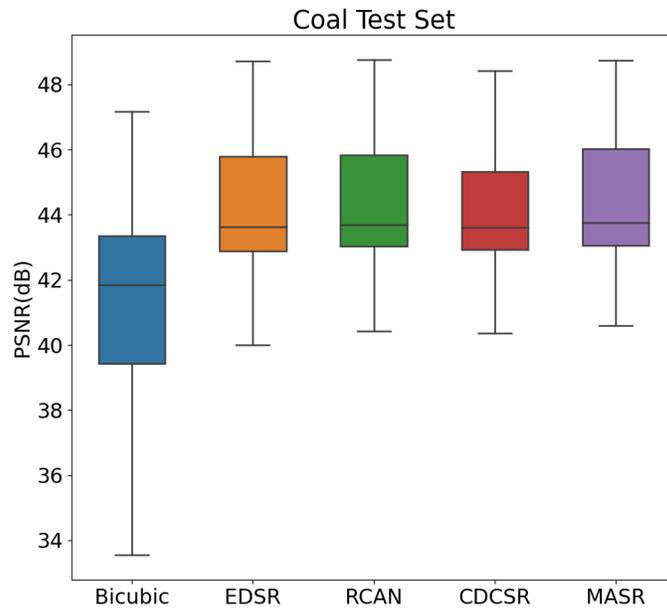
Figure S4. Boxplot of the average PSNR of EDSR, RCAN, CDCSR and MASR on coal test set. The subplot of Figure 8 in the main article.
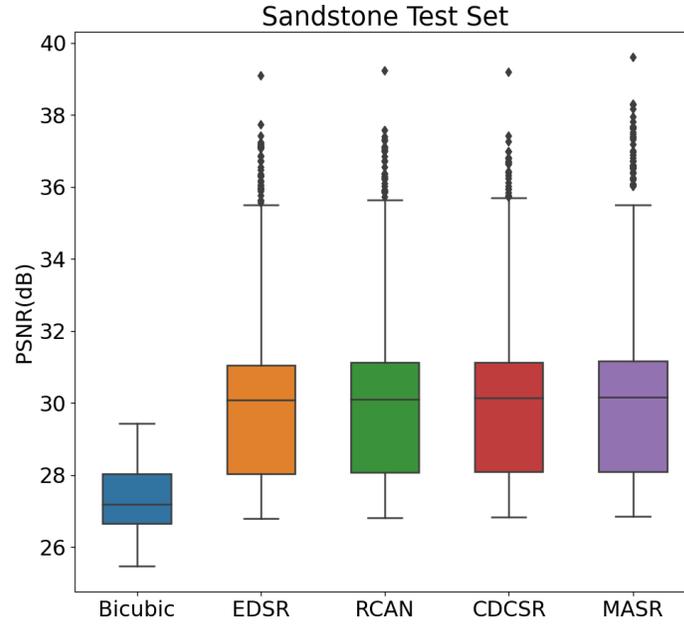
Figure S5. Boxplot of the average PSNR of EDSR, RCAN, CDCSR and MASR on sandstone test set. The subplot of Figure 8 in the main article.


Figure S5. Boxplot of the average PSNR of EDSR, RCAN, CDCSR and MASR on sandstone test set. The subplot of Figure 8 in the main article.
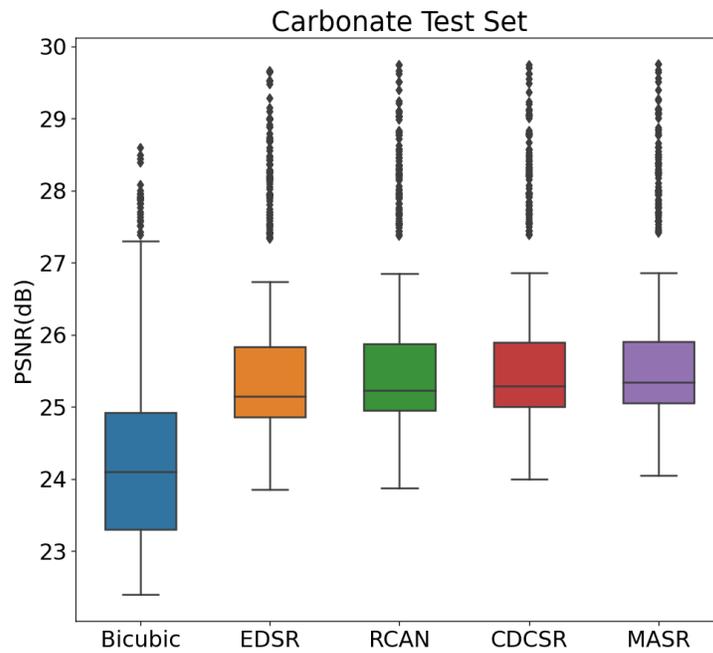
Figure S6. Boxplot of the average PSNR of EDSR, RCAN, CDCSR and MASR on carbonate test set. The subplot of Figure 8 in the main article.
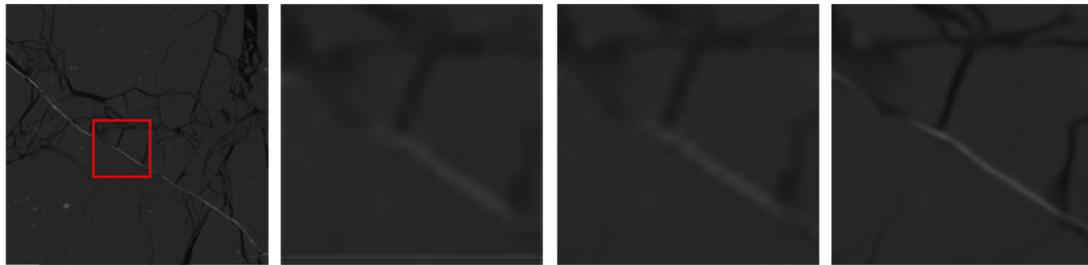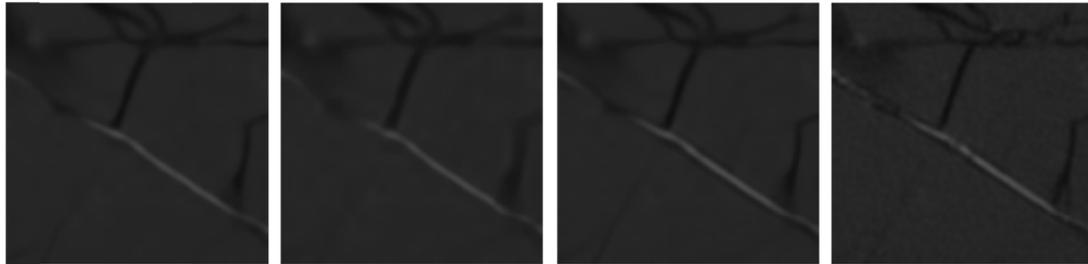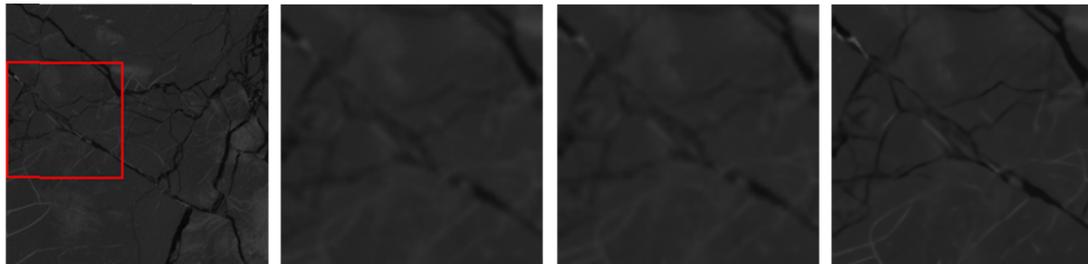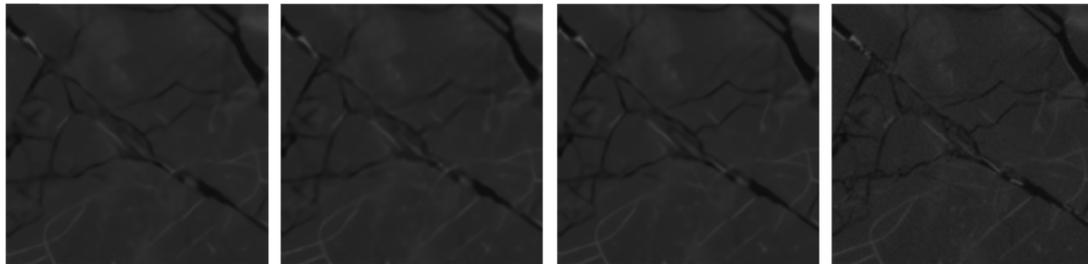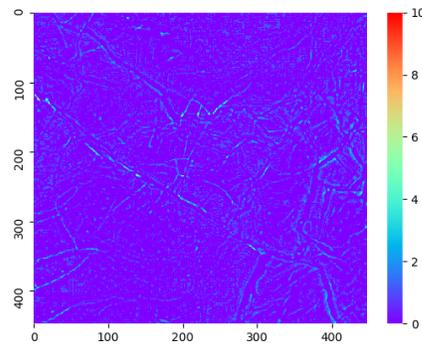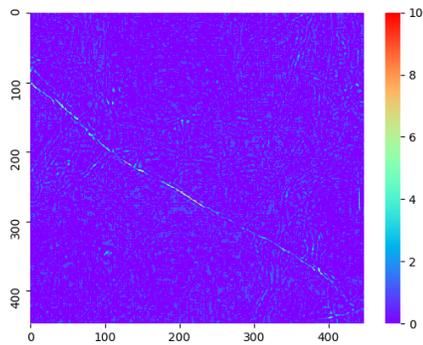
Figure S7. Additional qualitative comparison of our model with other works at ×4 SR on the coal test set. Bottom: Difference maps of MASR and RCAN SR results in the above two sets of images (MASR minus RCAN). The difference maps show that the edge recovered by MASR is sharper.
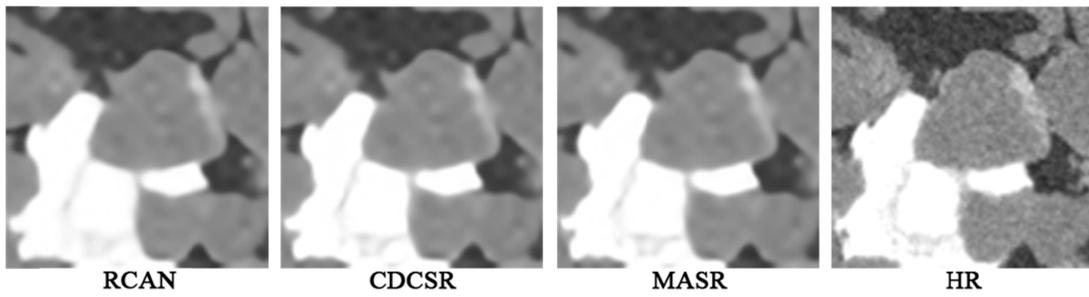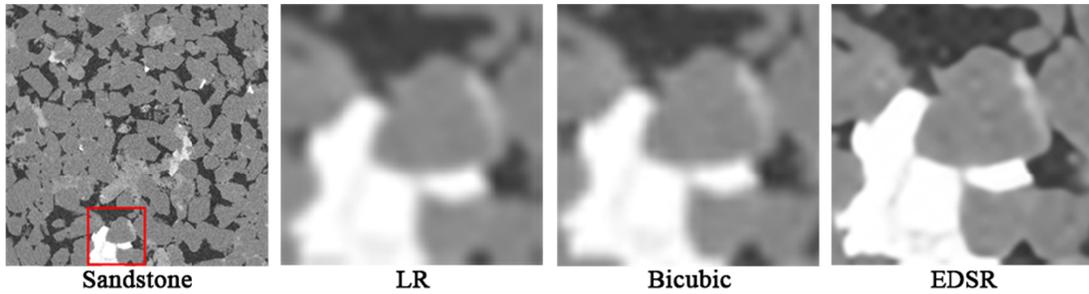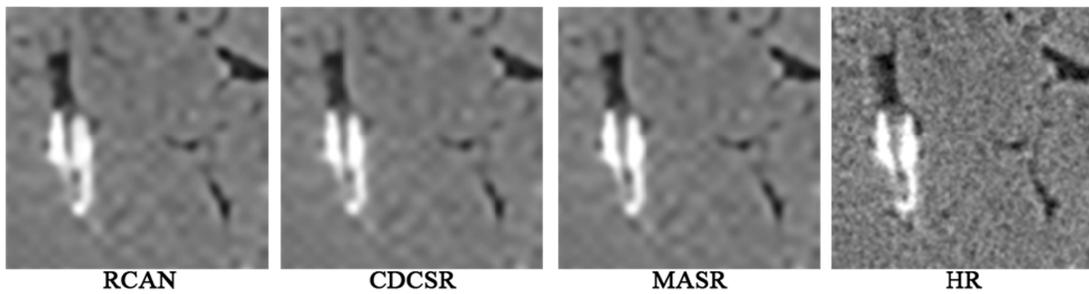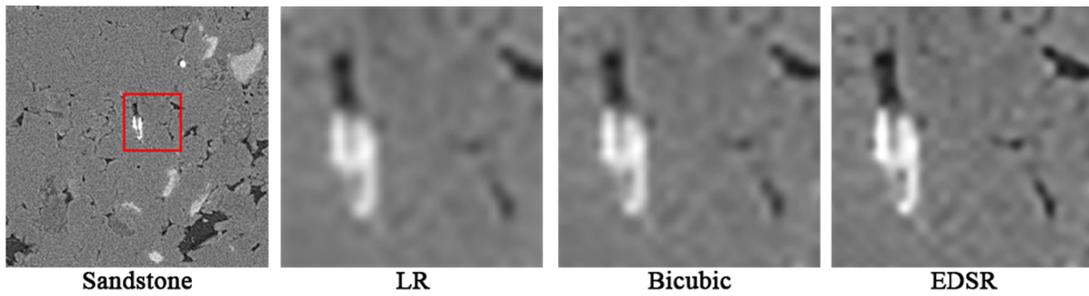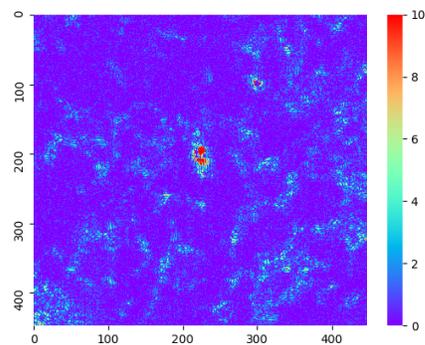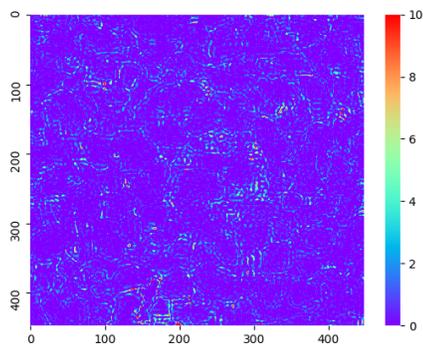
Figure S8. Additional qualitative comparison of our model with other works at ×4 SR on the sandstone test set. Bottom: Difference maps of MASR and RCAN SR results in the above two sets of images (MASR minus RCAN). The difference maps show that the edge recovered by MASR is sharper.
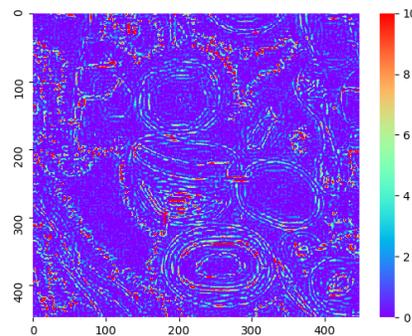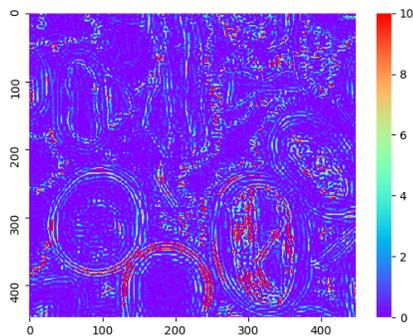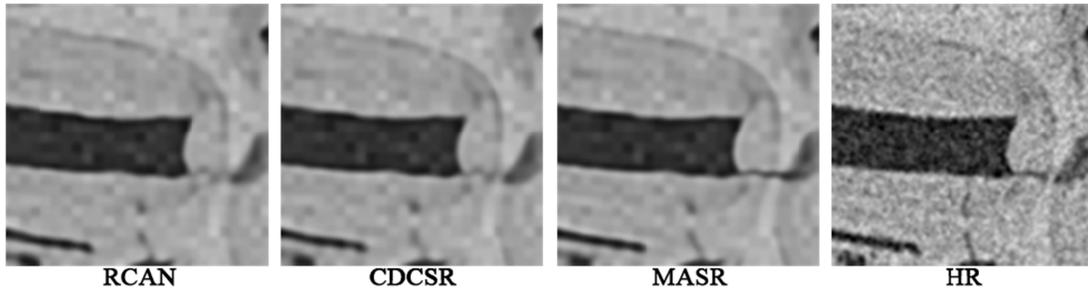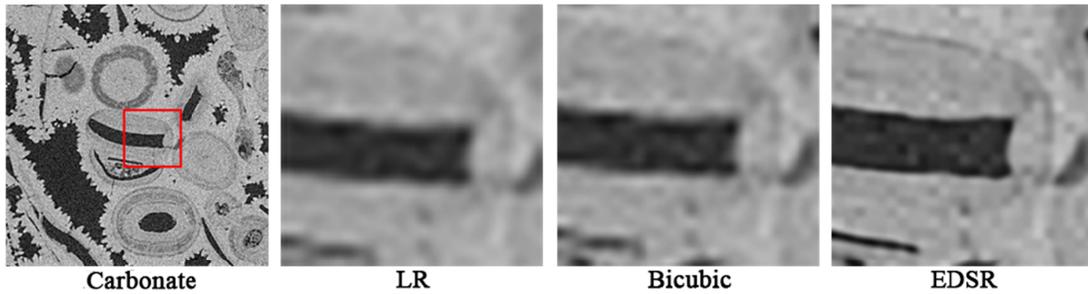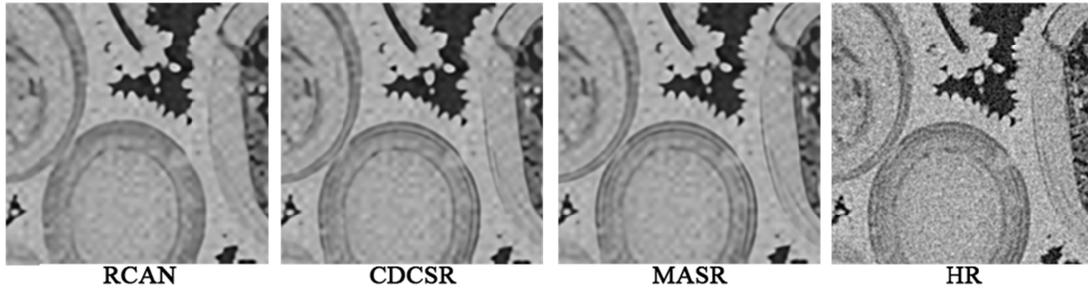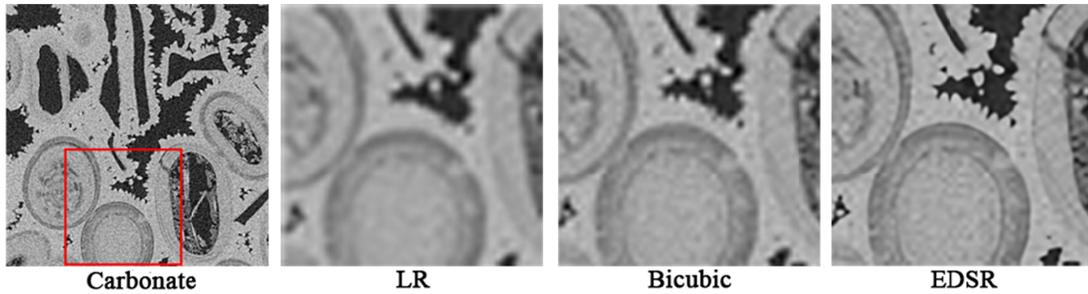
Figure S9. T Additional qualitative comparison of our model with other works at ×4 SR on the carbonate test set. Bottom: Difference maps of MASR and RCAN SR results in the above two sets of images (MASR minus RCAN). The difference maps show that the edge recovered by MASR is sharper.