

Motion artifact correction for MR images based on convolutional neural network*

ZHAO Bin^{1,2}, LIU Zhiyang^{1,2}, DING Shuxue^{1,2,3}, LIU Guohua^{1,2}, CAO Chen⁴, and WU Hong^{1,2**}

1. College of Electronic Information and Optical Engineering, Nankai University, Tianjin 300350, China

2. Tianjin Key Laboratory of Optoelectronic Sensor and Sensing Network Technology, Nankai University, Tianjin 300350, China

3. School of Artificial Intelligence, Guilin University of Electronic Technology, Guilin 541004, China

4. Department of Medical Imaging, Tianjin Huanhu Hospital, Tianjin 300350, China

(Received 21 May 2021; Revised 29 September 2021)

©Tianjin University of Technology 2022

Magnetic resonance imaging (MRI) is a common way to diagnose related diseases. However, the magnetic resonance (MR) images are easily defected by motion artifacts in their acquisition process, which affects the clinicians' diagnosis. In order to correct the motion artifacts of MR images, we propose a convolutional neural network (CNN)-based method to solve the problem. Our method achieves a mean peak signal-to-noise ratio (*PSNR*) of (35.212 ± 3.321) dB and a mean structural similarity (*SSIM*) of 0.974 ± 0.015 on the test set, which are better than those of the comparison methods.

Document code: A **Article ID:** 1673-1905(2022)01-0054-5

DOI <https://doi.org/10.1007/s11801-022-1084-z>

Since magnetic resonance imaging (MRI) is sensitive to the motion, there are motion artifacts in MRI acquisition. As reported in Ref.[1], about 10%—42% of brain examinations bring motion artifacts, which interfere with the clinician's diagnosis. The reasons leading to motion artifacts during brain MRI acquisition can be classified as sudden involuntary movements caused by sneezing, coughing, yawning, and semi-regular movements such as swallowing and blinking, and conscious motion of body parts^[2].

Nowadays, some methods have been proposed to prevent, mitigate or correct motion artifacts, one of which prevents motion artifacts by restricting head movement with cushions and head holders. Besides, fast single-shot pulse sequence, non-Cartesian k-space acquisition strategy, measuring the head pose in a real-time manner either in image space or k-space are used to obtain clean magnetic resonance (MR) images during their acquisition process^[3,4]. Since the factors influencing motion artifacts during MRI acquisition are more complex, there is no simple and effective general solution for this problem. In addition, the availability of the above mentioned motion artifact correction methods varies among MRI device manufacturers, which would hinder their application in clinical applications.

Deep learning with convolutional neural networks has also been applied to correct motion artifacts. For instance,

a recurrent neural network based method is proposed to reduce cardiac MRI motion artifacts and the multi-scale structures are used in the method to extract local and global features^[5]. TAMADA et al^[6] developed a method based on denoising convolutional neural network (DnCNN)^[7] for motion artifact reduction of liver, which significantly reduces the magnitude of the artifacts and blurring induced by respiratory motion. Nevertheless, these non-rigid motion artifacts are different from rigid motion artifacts on appearance. A motion correction network (MoCoNet)^[8] is proposed as a specific method for brain motion artifacts correction. However, it only adjusts the number of convolutional layer in U-Net and has limited capacity in motion artifacts correction. The deep residual network with densely connected multi-resolution block (DRN-DCMB)^[9] is another model recently proposed for brain motion artifacts reduction and achieves higher performance compared with the comparison methods. However, DRN-DCMB is trained on the small image patches that are randomly selected from the full size image, which leads to decrease conspicuity of small anatomic structures of the original full image. Meanwhile, using ℓ_2 loss function to train DRN-DCMB leads to omitting the underlying structure in the MR images. Therefore, in this paper, we propose a new network to correct motion artifacts, which is trained on the original full MR images and focuses on the whole correction.

* This work has been supported in part by the National Natural Science Foundation of China (Nos.61871239 and 62076077), and the Natural Science Foundation of Tianjin (No.20JCQNJC0125).

** E-mail: wuhong@nankai.edu.cn

Since the motion artifacts are caused unintentionally during MRI acquisition, it is impossible to deliberately make patients cooperate to obtain MR image pairs with and without motion artifacts. Therefore, it is necessary to simulate the motion artifacts on clean MR images.

Note that the head motion is rigid, and can be modeled as a combination of translational and rotational motions with six degrees of freedom. By defining a voxel coordinate of MR image as $V=(v_x, v_y, v_z)$, its coordinate after rigid motion can be modeled as

$$V = \begin{bmatrix} r_{xx} & r_{yx} & r_{zx} & \Delta x \\ r_{xy} & r_{yy} & r_{zy} & \Delta y \\ r_{xz} & r_{yz} & r_{zz} & \Delta z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ v_z \\ 1 \end{bmatrix}, \quad (1)$$

where the left matrix on the right hand side represents the rigid motion, which is recorded as T for convenience. The upper-left sub-matrix and the last column of T indicate the rotation and translation along the three axes, respectively.

Fourier theorem indicates that translation in the spatial domain results in phase errors in the k-space domain along the phase-encoding direction, while rotation in spatial domain results in identical rotation of the k-space data with the rotation axis through the origin. Given a clean MR image X , its corresponding MR image Y with motion artifacts can be calculated as

$$Y = f'[f(T * X)_{(\alpha, \beta)}], \quad (2)$$

where f and f' indicate the Fourier transform and inverse Fourier transform, respectively. α and β are the parameters of motion artifacts simulation. The center $\alpha\%$ of the k-space lines of X are kept intact to reserve the low frequency data that determined image contrast. $\beta\%$ of the remaining peripheral k-space lines are randomly selected for rigid motion transformation.

The architecture of our proposed network is illustrated in Fig.1. The network consists of four feature extraction modules (FEMs), which are used for feature extraction of the input MR images with motion artifacts and connected through the attention dense connection to reuse the former features. The feature-extraction module is proposed based on the U-Net^[10], while convolutional block and copy connection are replaced with attention paralleling block (APB) and attention residual block (ARB), respectively. The architectures of APB and ARB are shown in Fig.1(c) and Fig.1(d). The APB and the ARB can help our network to extract more features and further improve the correction results. This attention mechanism uses channel and spatial attention synergies^[11], firstly, input features are passed through the channel attention module in order to give more weight to the features that contribute to the task, and then the channel-weighted features are fed into the spatial attention module to obtain weighted features in the spatial dimension. Residual connectivity^[12] has two roles in the network proposed in this paper, in addition to facilitating gradient propagation

and preventing gradient disappearance and gradient explosion during training, it also allows reusing the previous features to facilitate the extraction of richer semantic information in the later layers. The activation function used in this paper is rectified linear unit (ReLU).

For each FEM, the parameters of each convolutional layer remain the same except for the input channel parameters. The numbers of input and output channels are labeled next to each module in Fig.1(b), where the number of input channels in the first module is replaced by x . According to the input and output channel parameters of FEM, the parameter settings of the entire network can be inferred.

In order to evaluate the correction results of motion artifacts, the peak signal-to-noise ratio ($PSNR$) and structural similarity ($SSIM$) are used as evaluation metrics in this paper. $PSNR$ is an objective standard for evaluating images. It is often used as a measurement method of signal reconstruction quality. It is used to measure the ratio of average energy between peak signal and background noise. Its unit is dB. The larger the value, the smaller the distortion. Given a set of images I and images O , $PSNR$ can be defined as

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right), \quad (3)$$

where MSE is the mean square error of the two images, and MAX_I is the maximum pixel value of I .

Since $PSNR$ evaluates the image quality based on the error between the corresponding pixels, it does not take into account the visual characteristics of human eye, that is, the human eye is more sensitive to contrast differences with lower spatial frequencies and more sensitive to differences in brightness contrast, which often results in inconsistencies between the evaluation results and people's subjective feelings. As a full-reference image quality evaluation metric, $SSIM$ measures the similarity of images from the three aspects of luminance, contrast and structure, and can be consistent with human visual perception on the whole. The definition of $SSIM$ is

$$SSIM(I, O) = \frac{(2\mu_I\mu_O + c_1)(2\sigma_{IO} + c_2)}{(\mu_I^2 + \mu_O^2 + c_1)(\sigma_I^2 + \sigma_O^2 + c_2)}, \quad (4)$$

where μ_I and σ_I^2 are the mean and variance of I , respectively, μ_O and σ_O^2 are the mean and variance of O respectively, σ_{IO} is the covariance of I and O , $c_1=(k_1L)^2$, $c_2=(k_2L)^2$, k_1 and k_2 are fixed values with 0.01 and 0.03, respectively, and L denotes the range of pixel values.

The experimental data includes 45 brain T1-weighted (T1W) MRI scans with matrix sizes of $336 \times 448 \times 56$, which were collected from a retrospective database of Tianjin Huanhu Hospital and anonymized prior to the use of researchers. The research has been approved by the institutional ethic committee. The experienced clinicians from Tianjin Huanhu Hospital confirmed that these T1W images have no motion artifacts. The images are randomly divided into training set, validation set and test set, with 27, 9 and 9 subjects, respectively.

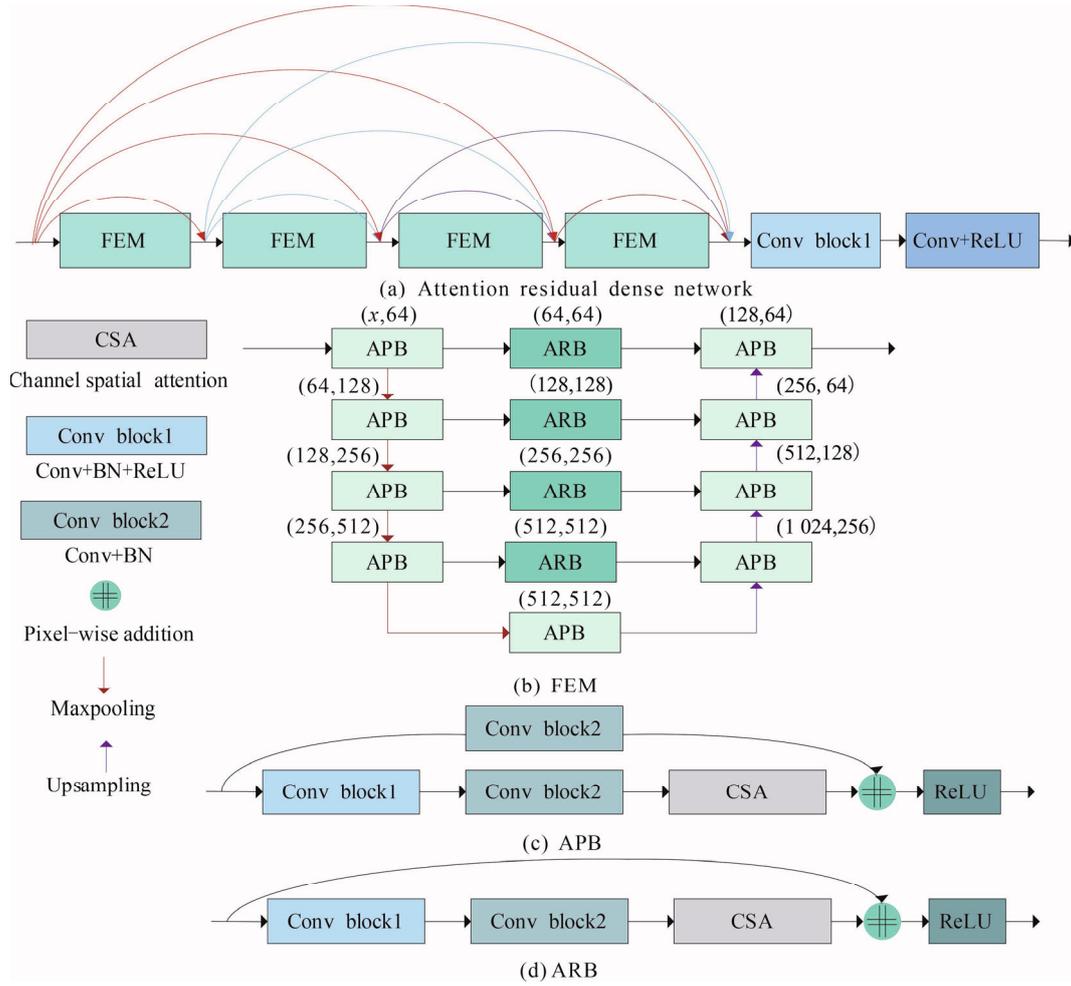


Fig.1 Attention residual dense network and module illustration

Each MR image is resampled to $192 \times 192 \times 56$ using linear interpolation before simulating motion artifacts. The intensities of the clean and corrupted images are normalized to a range of $[0,1]$ before entering the networks. Data augmentation techniques are adopted in this research. In particular, each input image is randomly rotated by a degree ranging from 1° to 360° , flipped vertically and horizontally on the fly, so as to augment the dataset and reduce memory footprint. Every augmentation method has a random number with $0-1$ uniform distribution. If the random number exceeds 0.5, the corresponding augmentation method is performed.

The parameters used in simulating motion artifacts are randomly generated with maximum translation of voxel along the x -axis, y -axis and z -axis are 8 mm, 8 mm and 4 mm, respectively, the maximum rotation angle is 9° , and α is set to 10. β is drawn from a uniform distribution within $[20, 60]$.

The networks are initialized using Kaiming's method^[13] and the optimizer is the Adam method^[14]. The initial learning rate is 10^{-3} . During training, the learning rate is scaled down by a factor of 0.1 if no progress is made for 15 epochs on validation loss, and the training stops after 30 epochs with no progress on the validation loss.

The experiments are performed on a computer with an Intel Core i7-6800K CPU, 64 GB RAM and Nvidia GeForce 1080Ti GPU with 11 GB memory. All networks are implemented in PyTorch.

Since the T1W image reflects the anatomical structure, motion artifact correction should focus on the luminance, contrast, and texture. Inspired by the work in image restoration^[15], we propose to use the sum of multi-scale structural similarity (MS -SSIM) loss and ℓ_1 loss to correct motion artifacts, where MS -SSIM loss preserves the texture and contrast in high-frequency regions, and ℓ_1 loss preserves the luminance. The loss used in the proposed method is defined as

$$\mathcal{L}(X, \hat{X}) = \lambda \cdot [1 - MS(X, \hat{X})] + (1 - \lambda) \cdot G_{\sigma_g} \cdot \|X - \hat{X}\|_1, \quad (5)$$

where X and \hat{X} are the clean MR image and the reconstructed MR images, respectively. MS denotes MS -SSIM loss. G_{σ_g} is the Gaussian coefficient. λ is a tradeoff coefficient between the MS -SSIM and the ℓ_1 loss. In our experiment, λ is set to be 0.84.

After the motion artifact simulation, the motion artifacts generated on the clean T1W images are shown in

Fig.2(b). These motion artifacts, which appear as blurs or ghosts, severely degrade the image quality and may interfere with the clinician's diagnosis. Tab.1 summarizes the quantified results between the images with the motion artifacts and the original clean images on the test set, which is denoted as MA. The *SSIM* is 0.635 ± 0.070 , which indicates that the motion artifacts bring obvious effects to the original image in terms of luminance, contrast and structure. Especially for image slices with complex structure, the influence of motion artifacts is more obvious. For instance, The *PSNR* and *SSIM* of the image slice in the first row of Fig.2(b) are 25.288 dB and 0.503, respectively, which are both lower than their average values of 27.507 dB and 0.635 on the test set.

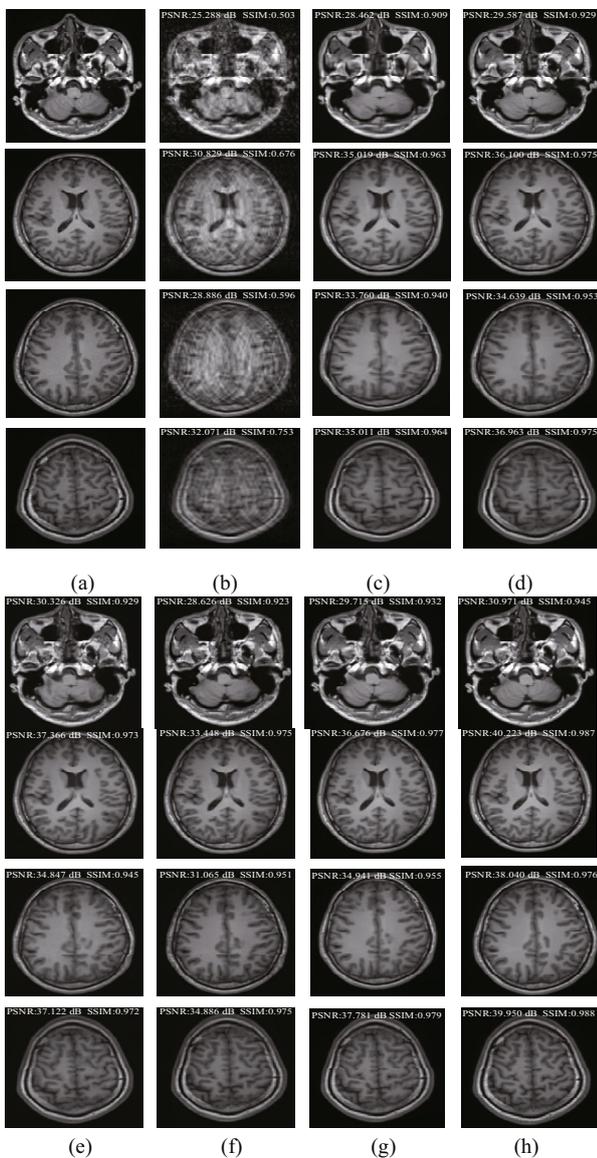


Fig.2 Visualized examples of the reconstruction performance: (a-b) Clean MR image and the corresponding MR images with motion artifacts, respectively; (c-h) Correction results of FCN-8s^[16], U-Net^[10], DnCNN^[7], DRN-DCMB^[9], MoCoNet^[8] and Ours, respectively

Tab.1 Quantitative evaluation results on the test set (The best results have been highlighted in bold.)

Method	<i>PSNR</i> (dB)	<i>SSIM</i>
MA	27.507±2.347	0.635±0.070
FCN-8s ^[16]	31.526±2.607	0.942±0.023
U-Net ^[10]	33.391±2.817	0.961±0.019
DnCNN ^[7]	33.580±3.032	0.954±0.022
DRN-DCMB ^[9]	29.852±2.912	0.955±0.022
MoCoNet ^[8]	33.627±2.905	0.964±0.018
Ours	35.212±3.321	0.974±0.015

For the sake of comparison, we also train and evaluate fully convolutional network (FCN)-8s^[16], U-Net^[10], DnCNN^[7], MoCoNet^[8] and DRN-DCMB^[9] using the same settings and loss function on our dataset. Fig.2 visualizes some slices of motion artifacts and their correction results. As Fig.2 shows, the input MR images contain severe motion artifacts after the rigid motion transformation. Although FCN-8s and U-Net are generally used for segmentation task, they present high performance on motion artifacts correction and obtain clean MR images. As a specified denoising network, DnCNN has achieved competitive results on the *PSNR*, which highlights its denoising ability, but it ignores some image information such as structure and obtain a lower *SSIM*. However, as a specified model for correcting motion artifacts, DRN-DCMB ignores some structure contrast. By replacing the two-layer convolution of U-Net with three-layer convolution, MoCoNet achieves close correction results to U-Net. In comparison, the proposed method substantially corrects the motion artifacts and produces cleaner MR images, and meanwhile the image contrast is maintained and subtle image details are preserved.

The quantitative evaluation results are summarized in Tab.1. As Tab.1 shows, benefit from multi-scale feature fusion, U-Net and MoCoNet achieve competitive correction results, which exceeds the comparison methods except our proposed method. Our proposed method, however, achieves the best correction results with a mean *PSNR* of (35.212 ± 3.321) dB and a mean *SSIM* of 0.974 ± 0.015 thanks to the utilization of the attention dense connection to reuse the former features and four feature extraction modules to extract more features.

Furthermore, we evaluate our proposed method on the MR images with real motion artifacts. Visualized examples of the MR image with real motion artifacts and the corresponding motion-corrected images are shown in Fig.3. Our proposed method substantially reduces the motion artifacts while maintaining the image contrast and details.

In this paper, we propose a CNN-based method to correct motion artifacts of MR images. The feature-extraction modules extract rich semantic information and the dense connection reuse more features. The experiments of motion artifacts simulation demonstrate

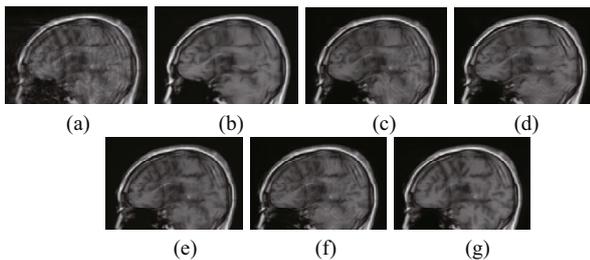


Fig.3 Visualized examples of the MR image with real motion artifacts and the corresponding motion-corrected images: (a) MR image with real motion artifacts; (b-g) Correction results of FCN-8s^[16], U-Net^[10], DnCNN^[7], DRN-DCMB^[9], MoCoNet^[8] and Ours, respectively

that our proposed method outperforms the comparison methods.

Statements and Declarations

The authors declare that there are no conflicts of interest related to this article.

References

- [1] ANDRE J B, BRESNAHAN B W, BASHA M M, et al. Toward quantifying the prevalence, severity, and cost associated with patient motion during clinical MR examinations[J]. *Journal of the American college of radiology*, 2015, 12(7): 689-695.
- [2] GODENSCHWGW F, KAGEBEIN U, STUCHT D, et al. Motion correction in MRI of the brain[J]. *Physics in medicine & biology*, 2016, 61(5): R32-R56.
- [3] VAN DER KOUWE A J W, BENNER T, DALE A M. Real-time rigid body motion correction and shimming using cloverleaf navigators[J]. *Magnetic resonance in medicine: an official journal of the international society for magnetic resonance in medicine*, 2006, 56(5): 1019-1032.
- [4] WELCH E B, MANDUCA A, GRIMM R C, et al. Spherical navigator echoes for full 3D rigid body motion measurement in MRI[J]. *Magnetic resonance in medicine: an official journal of the international society for magnetic resonance in medicine*, 2002, 47(1): 32-41.
- [5] LYU Q, SHAN H, XIE Y, et al. Cine cardiac MRI motion artifact reduction using a recurrent neural network[EB/OL]. (2020-06-23)[2021-09-20]. <https://arxiv.org/abs/2006.12700v1>.
- [6] TAMADA D, KROMREY M L, ICHIKAWA S, et al. Motion artifact reduction using a convolutional neural network for dynamic contrast enhanced MR imaging of the liver[J]. *Magnetic resonance in medical sciences*, 2020, 19(1): 64-76.
- [7] ZHANG K, ZUO W, CHEN Y, et al. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising[J]. *IEEE transactions on image processing*, 2017, 26(7): 3142-3155.
- [8] PAWAR K, CHEN Z, SHAH N J, et al. MoCoNet: motion correction in 3D MPRAGE images using a convolutional neural network approach[EB/OL]. (2018-07-29) [2021-09-20]. <https://arxiv.org/abs/1807.10831v1>.
- [9] LIU J, KOCAK M, SUPANICH M, et al. Motion artifacts reduction in brain MRI by means of a deep residual network with densely connected multi-resolution blocks (DRN-DCMB)[J]. *Magnetic resonance imaging*, 2020, 71: 69-79.
- [10] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]//*International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, October 5-9, 2015, Munich, Germany. Berlin, Heidelberg: Springer-Verlag, 2015: 234-241.
- [11] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the European Conference on Computer Vision (ECCV)*, September 8-14, 2018, Munich, Germany. Berlin, Heidelberg: Springer-Verlag, 2018: 3-19.
- [12] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [13] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: surpassing human-level performance on imagenet classification[C]//*Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 1026-1034.
- [14] KINGMA D P, BA J. Adam: a method for stochastic optimization[EB/OL]. (2017-06-30)[2021-09-20]. <https://arxiv.org/pdf/1412.6980.pdf>.
- [15] ZHAO H, GALLO O, FROSIO I, et al. Loss functions for image restoration with neural networks[J]. *IEEE transactions on computational imaging*, 2016, 3(1): 47-57.
- [16] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 3431-3440.