



Clinical evaluation of super-resolution for brain MRI images based on generative adversarial networks



Yasuhiko Terada ^{a,*}, Tomoki Miyasaka ^a, Ai Nakao ^a, Satoshi Funayama ^b, Shintaro Ichikawa ^c, Tomohiro Takamura ^d, Daiki Tamada ^b, Hiroyuki Morisaka ^b, Hiroshi Onishi ^b

^a Institute of Applied Physics, University of Tsukuba, Tsukuba, Japan

^b Department of Radiology, University of Yamanashi, Chuo, Japan

^c Department of Radiology, Hamamatsu University School of Medicine, Hamamatsu, Japan

^d Department of Radiology, Shizuoka General Hospital, Shizuoka, Japan

ARTICLE INFO

Keywords:

Deep neural networks
Super-resolution
Clinical evaluation
Generative adversarial networks
Brain MRI

ABSTRACT

In magnetic resonance imaging (MRI), reducing long scan times is an urgent issue that could be addressed with super-resolution (SR) techniques. Most of the SR networks using deep neural networks (DNNs) have been evaluated only based on numeric metrics, and the image restoration quality for individual lesions is not evaluated sufficiently. Here, we evaluated the performances of different types of SR networks using DNNs for brain MRI, in terms of diagnostic performance and image quality. We focused on comparing the performance between generative adversarial networks (GANs) and non-GAN networks. There was a trade-off in such restoration quality between GAN- and non-GAN-based SRs, with the GANs being more accurate in restoring images of anatomical structures but less accurate in restoring those of lesions; non-GANs showed the opposite tendency. The non-GAN SRs were preferable in terms of diagnostic performance and image quality. This result suggested that the evaluation of DNN performance for lesions might be changed drastically by adding a clinical evaluation perspective. The dependence of network architecture on network performance obtained in this study will provide guidance for future development of SR DNN for medical images.

1. Introduction

In many clinical and research applications, reducing long scan times for magnetic resonance imaging (MRI) is an urgent issue. If these can be shortened, numerous benefits are expected, including increased patient throughput, reduced motion artifacts, and the ability to add other sequences within the same examination time. The scan time for conventional Cartesian MRI sampling is proportional to the number of phase encodings, and one way to shorten the scan time is to use super-resolution (SR) analysis, in which the number of phase encodings is reduced to acquire data at a lower resolution (LR) with a smaller matrix size, coupled with post-processing to restore the original high-resolution (HR) image.

SR techniques have been studied for a long time in computed imaging [1]. Classical methods such as bicubic and b-spline interpolation are often used because of their simplicity, but they tend to produce blurred edges and blocking artifacts. The problem of SR is now often modeled as an inverse problem and is solved by an optimization

algorithm using regularization terms (such as low rank constraint and total variation) [2–5].

SR methods using deep learning (DL) are claimed to have achieved “state-of-the-art” performance [6–8]. The deep neural network (DNN) architecture for SR has been developed primarily based on super-resolution tasks for natural images. The initial simple network for SR consisted of multiple layers of simple stacked convolutional neural networks (CNNs) called SRCNN [9]. The architecture of SR subsequently advanced from simple SRCNN networks to residual-based networks, a design in which a large number of “residual blocks” using residual connections are connected in series to avoid the degradation problem and allow training large CNNs with high accuracy. SR using residual-based networks includes VDSR [10], which is based on deep residuals, and DRCN [11], DRRN [12], DLRRN [13], and cross-domain heterogenous network [14] which have recursive structures. Improvements were also proposed in the method of upscaling the image, and instead of using interpolation such as bicubic operations at the beginning of the network, a new method was devised to include a layer in the

* Corresponding author. Institute of Applied Physics, University of Tsukuba, Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8573, Japan.
E-mail address: terada.yasuhiko.fu@u.tsukuba.ac.jp (Y. Terada).

network that performs the upsampling operation (e.g., ESPCN [15] and FSRCNN [16]). Later, EDSR [17] and LapSRN [18], which were SR networks that incorporated this new upsampling scheme into the resnet, were developed. SRdenseNet [19] and RDNs [20], which are SR networks that extend the simple resnet structure by increasing the number of skipped connections and devising new connection methods, were also proposed for SR tasks. More recently, autoencoder-based SR approaches that learn latent space representations from input LR images, has also been shown to provide state-of-the-art performance with reduced computational complexity [21]. Tian et al. presented an enhanced super-resolution group CNN (ESRGCNN) with a shallow architecture by fully fusing deep and wide channel features to extract more accurate low-frequency information [22]. Other recent studies, such as Cascading Residual network (CARN) [23], CFSRCNN [24], CADUF [25], and DRLN [26], are based on cascaded CNN architectures to form coarse to fine methods for single-image SR. Lan et al. proposed cascading and enhanced residual networks that employ CARN with multiple locally shared groups to effectively extract image features, enhanced residual networks to extract long-range spatial features, and multiscale block to obtain feature representations of input images at different scales [27]. Attention-based learning is another powerful SR network architecture which is being used frequently these days. It is based on the idea of inserting attention blocks into the model to extract some important features, rather than considering all channels and spatial locations equally important for the SR task. Examples of attention-based SR networks include channel attention (RCAN) [28], second-order channel attention (SPCA) [29,30], and spatial feature extraction local and non-local attention [31,32]. Multi-scale SR models are another SR networks architecture that has recently emerged, where different representations are made on multiple paths, later combining these representations to improve performance [14,26,33–39]. SR models that use the attention mechanism to reduce network weight have also been proposed. A dense, lightweight network called MADNet [40] learns more powerful multi-scale feature representations and feature correlations. Jiang et al. proposed a lightweight SR network that utilizes a sparse optimization method and a sparse-aware attention module [41]. Tan et al. proposed an efficient SR network via Self-Calibrated Feature Fuse (SCFFN), by constructing the self-calibrated feature fuse block using an attention mechanism [42]. Yang et al. reported an SR model that incorporates a convolutional kernel attention mechanism and a non-linear perceptual multi-scale network [39]. The wavelet transform (WT), which uses high-frequency sub-bands to represent texture and low-frequency sub-bands to represent global structural information, is also used in SR in a hybrid form with other learning methods. Examples include multilevel wavelet CNN (MWNCC) [43,44], wavelet-based residual attention network (WRAN) [45], and wavelet-based dual recursive network [46].

Generative adversarial network (GAN) is a generative model that train two networks competitively—generators and discriminators—to synthesize a “better looking” image by introducing perceptual loss as a loss function. In recent years, a number of SR networks have been reported using GANs that outperform traditional CNN networks. Examples include SRGAN [47], which combines a generator using a residual network and a discriminator using a CNN; ESRGAN [48], an extension of SRGAN; pix2pix [49], a contrast transform; and pix2pixHD [50], an improvement of pix2pix.

These DLs for SR have been applied to medical images and have attracted much attention. For example, SRCNN [9], modified U-net [51], SRGAN [52], and GAN-CIRCLE [53] are used for super-resolution of CT images, while SRCNN [54,55], residual-based networks [56], Dual U-Net [57], and GAN [58–63] are used for MR images. For 3D MR images, self super resolution (SSR), a network that estimates a high-resolution 3D image from an in-plane high-resolution image and a through-plane lower-resolution image, has also been proposed [64].

There are two categories for DL-based SR of MR images according to the number of input sources involved in the process: single-contrast SR

(SCSR) and multi-contrast SR (MCSR). SCSR uses a single contrast image as input, whereas MCSR uses numerous contrast images as input. It has also been reported that MCSR outperforms SCSR [65].

Many of the current DNNs for image processing and restoration of medical images focus on improving the performance to recover frequently appearing images accurately: namely based on normal anatomical structures. However, the location, size, and contrast of lesions in clinical images vary widely, and the numbers of lesions in training datasets are small compared with the amount of healthy tissue. Therefore, there is a high risk of lesions disappearing (false negatives) or appearing as artifacts (false positives) in DL-based image processing and restoration, and DL might be unstable for small structural changes such as lesions in the brain. This issue has attracted increasing attention in recent years, as the number of DL applications in medical imaging has increased. Thus, Antun et al. [66] pointed out this instability in DL-based image reconstruction: instabilities with respect to small structural changes (for example a brain image with or without a small tumor). They also provided tools for investigating these instabilities. Gottschling et al. [67] predicted theoretically that this instability is persistent and difficult to remove in DL-based reconstruction problems. As Antun et al. showed [66], these instabilities are likely to vary significantly depending on the network's architecture. Therefore, it is important to find suitable DLs that reduce instability as much as possible. However, most of the DNNs to date have been evaluated only based on numeric metrics such as peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), which measure the average accuracy of the entire image, and the image restoration quality for individual lesions is not evaluated sufficiently. This is partly because lesion assessment is not automatic and is a time-consuming and labor-intensive task compared with the calculation of numerical indices. This is also because many studies use public datasets, which either do not contain lesions or, if they do, the numbers of lesions are relatively small or they are not annotated.

The purpose of this study was to compare the SR performance between GAN and non-GAN networks from the clinical evaluation perspective. We collected brain imaging datasets containing lesions from routine clinical examinations and built SR DNNs to investigate which network architecture might be the most relevant clinically. To assess the clinical performance of the SR network, we evaluated the restoration quality of images of anatomical structures and lesions. Because a previous study suggested that MCSR can outperform SCSR [65], we also compared such networks.

The main contribution of this work is to evaluate the performance of different brain SR-DNNs from a clinical evaluation perspective. We show that there is a trade-off between GANs and non-GAN-based SRs, with GANs being more accurate in restoring anatomical structures but less accurate in restoring lesions. Our results suggest that the network performance might be changed if clinical evaluation is considered.

This paper starts with an introduction to DL-based SR of MR images, followed by a section about methods of SR of brain MRI and their clinical evaluation. Afterwards, the results, discussions, and conclusions are presented.

2. Methods

2.1. Study design principles

Our study focused on evaluating the performance and stability of DL-based SR imaging according to a realistic scenario representative of actual clinical imaging examinations. Accordingly, we designed this study according to the following principles.

- (i) We used SR DNNs to shorten the overall scan time for three contrast image formats commonly acquired in routine examinations: T_1 -weighted, T_2 -weighted, and fluid-attenuated inversion recovery (FLAIR) images.

- (ii) The scan times for input LR images were set as short as possible within the range that could be set easily by the user on the console of a clinical MRI scanner.
- (iii) Among the three contrast images, FLAIR and T_1 -weighted images were chosen as target images to be super-resolved. We chose FLAIR because it is essential for the diagnosis of neurological diseases and is the best for checking any lesion recovery. We also chose T_1 -weighted images because the relevant acceleration factor (AF) was the highest in the clinical MRI scanner used in this study, and the overall time was greatly reduced. The AFs for the FLAIR and T_1 -weighted images were chosen to be the maximum values that could be set by our clinical scanner, respectively, and this is the reason why the AF differed significantly between the FLAIR and T_1 -weighted images. T_2 -weighted images were not subjected to SR because undersampling all contrast images and using them as input images would significantly degrade SR performance.
- (iv) We aimed to perform a prospective study using datasets from clinical examinations.
- (iv) We aimed to focus on evaluating the quality of restoration of lesion images in addition to frequently appearing anatomical structures.

2.2. Dataset

This study was approved by our local Institutional Review Board. Following the policy described above, all HR and LR images were acquired in the same routine clinical session and used as the training and testing datasets. A GE 3T system (SIGNA Premier; General Electric Co., Boston, MA, USA) was used for data acquisition. As HR images, FLAIR and T_1 -and T_2 -weighted images (defined here as FLAIR, T1W, and T2W, respectively) were acquired using standard protocols used in clinical practice, and the FLAIR and T1W images were used as ground truth (GT) SR images from FLAIR and T1W, respectively. FLAIR and T1W images with a reduced number of phase encodings were also acquired as LR images in the same imaging session (FLAIR-LR and T1W-LR, respectively). According to the study design stated in (ii) above, the number of phase encodings for the LR images was set to the minimum number that the user could set on the console. The imaging conditions are summarized in Table 1. Here, the AFs were defined as the proportion of the number of encodings of HR and LR images. An example of MR images in dataset was shown in Supplementary Fig. S1.

In all, 3687 slice images were acquired for 145 cases, 3083 slice images for 122 cases were used for training and validation, and 604 slice images for 23 cases were used for clinical evaluation. Patient characteristics of the clinical evaluation cohort were investigated on the medical record system. The total data acquisition time for the HR images (FLAIR, T1W, T2W) was 152 s. The acquisition times were 32 s for FLAIR-LR and 15 s for T1W-LR. For clinical diagnosis, three T1W, T2W, and FLAIR images each with high spatial resolution were needed. Therefore, when FLAIR-SCSR was realized (and T1W and T2W images were not super-resolved), FLAIR-LR, T1W, and T2W images needed to be acquired over 120 s (78.9% of the total acquisition time of the three HR images). Similarly, when T1W-SCSR was realized (and FLAIR and T2W images were not super-resolved), FLAIR, T1W-LR, and T2W images were

acquired over 116 s (76.3% as above). When MCSR was realized, FLAIR-LR, T1W-LR, and T2W images were acquired over 84 s (55.3% as above).

2.3. Network

Fig. 1 shows an overall workflow diagram of the SR tasks. Table 2 summarizes the DNNs used for SR imaging used in this study. We prepared seven networks to determine the superiority or inferiority between GAN and non-GAN and between SCSR and MCSR (Figs. 2–5). We prepared two GAN-based networks: pix2pix and SRGAN. The pix2pix approach [49] uses U-Net as the generator and CNN as the discriminator. The SRGAN [47] uses SRResNet as the generator and CNN as the discriminator. As a non-GAN approach, we prepared a network in which the discriminator was removed from each GAN network. U-Nets, ResNets, and CNNs used in these networks are baseline analyses that are often used in image-to-image translation tasks including medical or natural image processing and reconstruction. MC-2stage (Fig. 5) was used as a two-stage GAN in which SCSR images were trained with SC-SRGAN at the first stage and MCSR images were trained with MC-pix2pix at the second stage. This was introduced with the expectation that it would show the highest performance.

2.4. Network implementation

The entire training process was implemented in standard Python (v3.6; Python Software Foundation, Wilmington, DE, USA). The networks were designed using the Keras package [68] running TensorFlow [69] as backend. The networks were trained and tested on a 64-bit Windows 10 system with a Threadripper 1900X CPU, 128 GB RAM, and four NVIDIA GeForce GTX 1080Ti graphic processing units (GPUs).

We modified the source codes of pix2pix and SRGAN provided online (<https://github.com/eriklindernoren/Keras-GAN>) to train the models on our dataset. The network parameters were set to be as similar as possible to the values in the original papers [47,49]. For MC-U-Net and MC-pix2pix (Fig. 4), three input images (FLAIR-LR, T1W-LR, and T2W, $512 \times 512 \times 1$ pixels (two-dimensional, 2-D images with one channel)) were concatenated along the channel direction. The loss function for SC-pix2pix/MC-pix2pix consisted of a content loss, l_{MAE} and adversarial loss, l_{MSE} . The content loss used was a mean absolute error (MAE) loss, which accounts for the L1 distance between the output of the generator network and the GT image. The adversarial loss used was a mean square error (MSE) loss on the output of the discriminator network. Then, the total loss function was calculated as $l_{pix2pix} = l_{MAE} + 10^{-2}l_{MSE}$. For SC-U-Net/MC-U-Net, the discriminator was removed from SC-pix2pix/MC-pix2pix, and the loss function was given by $l_{U-Net} = l_{MAE}$.

For SC-SRResNet and SC-SRGAN (Fig. 3), the UpSampling 2-D layer used in the downloaded source codes was replaced with sub-pixel convolution layers, as used in the original networks [47]. Similarly, the activation layers in the residual blocks were replaced with the PReLU layer. Because of limited GPU memory, the input images (512×512 pixels) for SRResNet and SRGAN were down-sized to 256×256 pixels using the OpenCV library. The loss function of SC-SRGAN, l_{SRGAN} , consisted of a content loss, l_{MSE} , and an adversarial loss, l_{GEN} . The content loss used was an MSE loss, which accounts for the L2 distance between the output of the generator network and the GT image. The

Table 1

Acquisition parameters for brain MRI datasets. FLAIR: fluid-attenuated inversion recovery, T1W: T_1 -weighted images, T2W: T_2 -weighted images, LR: low-resolution.

	Repetition time [ms]	Echo time [ms]	Echo train length	Matrix	pixel size [mm]	Reduction factor (encoding lines)	Acquisition Time [s]
T1W	550	9	3	384x192	0.573x1.16	1	51
T2W	5360	103	20	384x256	0.573x0.859	–	37
FLAIR	8000	143	20	416x192	0.529x1.16	1	64
T1W-LR	550	9	3	384x16	0.573x13.8	0.083 (1/12)	15
FLAIR-LR	8000	143	20	416x78	0.529x2.82	0.41 (1/2.5)	32

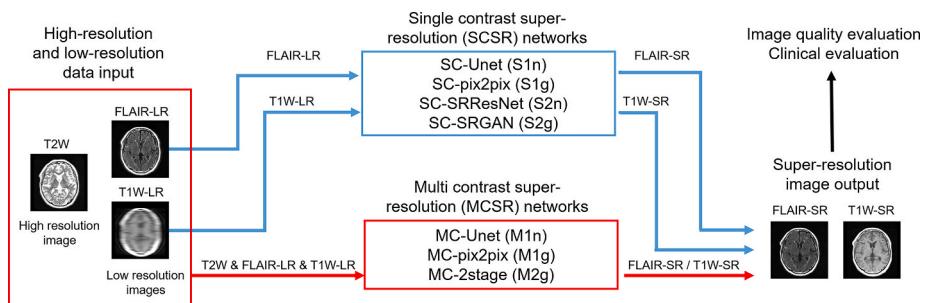


Fig. 1. Overall workflow diagram of super-resolution tasks.

Table 2
Deep neural networks used for super-resolution imaging.

Network	SC/MC	generator	discriminator	number of parameters	input	3-case evaluation [FLAIR only]	all-case evaluation
S1n	SC-U-Net	SCSR	U-Net	–	41,837,185	FLAIR-LR or T1W-LR	
S1g	SC-pix2pix	SCSR	U-Net	CNN	47,369,987	FLAIR-LR or T1W-LR	
S2n	SC-SRResNet	SCSR	SRResNet	–	1,386,369	FLAIR-LR or T1W-LR	x
S2g	SC-SRGAN	SCSR	SRResNet	CNN	11,818,627	FLAIR-LR or T1W-LR	x
M1n	MC-U-Net	MCSR	U-Net	–	41,839,233	FLAIR-LR/T1W-LR/T2W	x
M1g	MC-pix2pix	MCSR	U-Net	CNN	47,376,131	FLAIR-LR/T1W-LR/T2W	x
M2g	MC-2stage	MCSR	SC-SRGAN/MC-pix2pix		FLAIR-SR/T1W-LR/T2W or FLAIR-LR/T1W-SR/T2W	x	x

SCSR, single-contrast super-resolution; MCSR, multi-contrast super-resolution. FLAIR: fluid-attenuated inversion recovery, T1W: T_1 -weighted images, T2W: T_2 -weighted images, GT: Ground truth, LR: low-resolution, S2n: SC-SRResNet, S2g: SC-SRGAN, M1n: MC-U-Net, M1g: MC-pix2pix, M2g: MC-2stage.

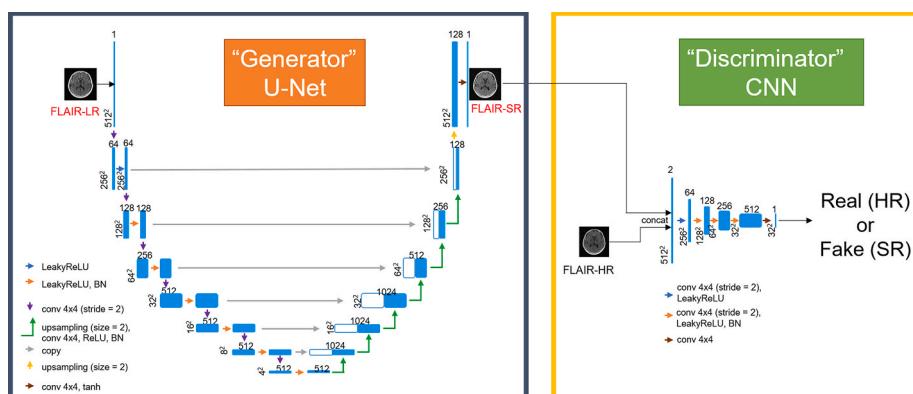


Fig. 2. Networks for SC-U-Net and SC-pix2pix.

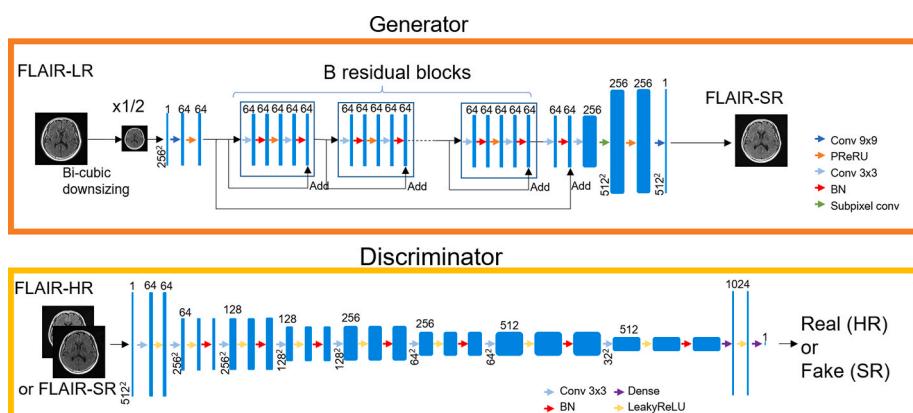


Fig. 3. Networks for SC-SRResNet and SC-SRGAN.

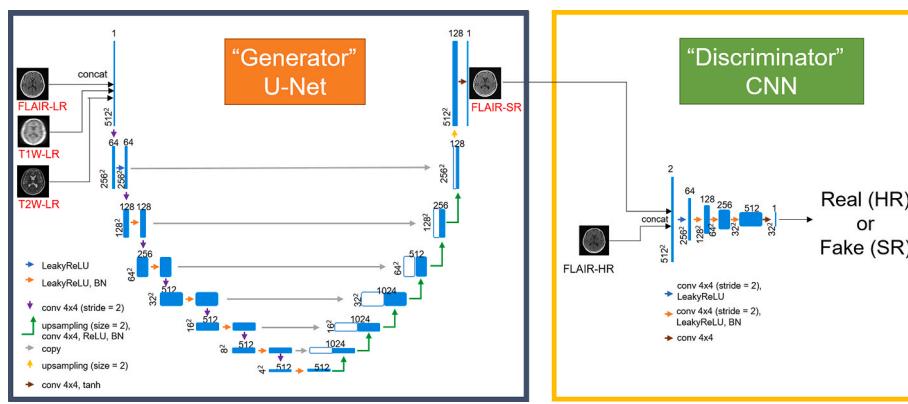


Fig. 4. Networks for MC-U-Net and MC-pix2pix.

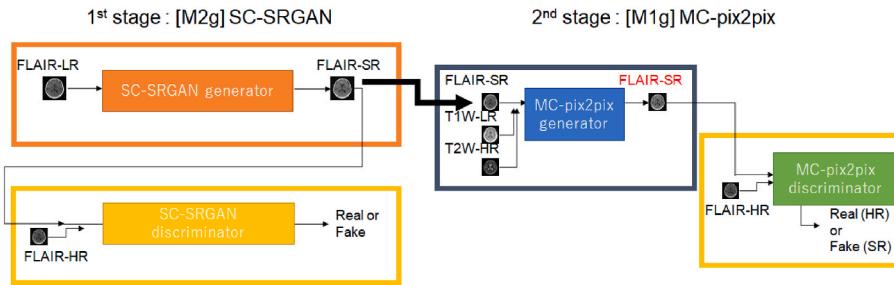


Fig. 5. Networks for MC-2stage.

adversarial loss used was a binary-cross entropy loss on the output of the discriminator network. Then, the total loss function for SC-SRGAN was calculated as $l_{\text{SC-SRGAN}} = l_{\text{MSE}} + 10^{-3}l_{\text{GEN}}$. For SC-SRResNet, the discriminator was removed from SC-SRGAN, and the loss function was given by $l_{\text{SC-SRResNet}} = l_{\text{MSE}}$.

The hyperparameter settings were listed in Table 3. All the networks were trained for 1000 epochs using the Adam solver with a learning rate of 0.0002 and a batch size of 12. Adam and learning rate were determined according to the original pix2pix paper [49]. Data augmentation with a rotation angle range of 3°, zoom range of 0.05, and horizontal and vertical flips was used during training. The values for data augmentation were determined appropriately to be within the range of image deformations expected in MRI images acquired during clinical examinations. For GANs, the generator and discriminator networks were trained jointly by alternating gradient optimization.

2.5. Numerical evaluation

Quantitative image quality comparisons between the GT and SR images were performed using the PSNR and SSIM values. Here, PSNR was defined as:

$$\text{PSNR} = 20 \log_{10} \frac{I_{\max}}{\text{RMSE}},$$

Table 3
Hyperparameter settings for training all the networks.

Hyperparameter	Values
Number of epochs	1000
Solver	Adam
Learning rate	0.0002
Batch size	12
Data augmentation	Rotation angle range of 3°, zoom range of 0.05, and horizontal and vertical flips

where RMSE is the root mean square error between the GT and SR images, and I_{\max} is the maximum intensity of the GT image. SSIM was defined as:

$$\text{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\mu_x^2 + \mu_y^2 + C_2)},$$

where μ_x and σ_x are the mean and variance of the GT image, μ_y and σ_y are the mean and variance of the SR image, respectively, and σ_{xy} is the covariance of the GT and SR images. C_1 and C_2 are included to avoid instability and we specifically chose that $C_1 = 0.01$ and $C_2 = 0.03$. Similar images have higher PSNRs and higher SSIMs (range 0–1). The mean PSNR and SSIM values over test images were calculated for 23 cases.

2.6. Clinical evaluation

The most frequent diagnosis of the 23 cases for the clinical evaluation (mean 61.0 [range, 20–88]; male: female, 12 : 11) was aneurysm (n

Table 4
Characteristics of clinical evaluation cohort.

Patient characteristics	Values
Age (mean, range)	61.0 (20–88)
Sex (male: female)	12:11
Diagnosis	Aneurysm Pituitary adenoma Screening Post-operation of meningioma Arterial stenosis Old cerebral infarction Dural arteriovenous fistula Cerebral hemangioma Intraorbital tumor Post-operation of cerebral hemangioblastoma Post-operation cerebral glioblastoma
	5 3 3 3 2 1 1 1 1 1 1

= 5). The details of patient characteristics are shown in Table 4. All test brain images were evaluated clinically in two stages in terms of diagnostic performance and image quality. At the first stage, test images of three cases for five FLAIR networks (Table 2) were evaluated by a radiologist (Rater A) with 6 years of experience. The evaluation was performed with no blinding; the rater evaluated the SR images in comparison with the corresponding GT (FLAIR-GT or T1W-GT) images. The contrast values of the corticomedullary region, conspicuity of the corticomedullary junction, conspicuity of the basal ganglia, conspicuity of the intracranial arteries, and sharpness of ventricular edge were graded on a 5-point scale (1, unacceptable; 2, poor; 3, acceptable; 4, good; and 5, excellent), and the image artifacts and overall image quality were graded on a 3-point scale (1, not diagnostic; 2, suboptimal; and 3, acceptable). The numbers of false negatives (disappearing lesions) and false positives (generated pseudolesions) were counted by the radiologist. At the second stage, test images of all 23 cases for two FLAIR and two T1W networks (Table 2) were evaluated clinically by Rater A and a radiologist (Rater B) with 15 years of experience according to the same criteria as in the first-stage evaluation. Unlike that approach, only the GAN results for MCSR images (M1g) and non-GAN results for MCSR images (M1n) were evaluated. The means of the two scores were calculated as the mean-opinion score (MOS) for each assessment item. Rater B recorded the size of each lesion appearing on the FLAIR-GT images. The raters also noted comments on noticeable image artifacts. The inter-observer agreement was assessed by calculating Kendall's coefficient of concordance with ties (W); the scale for W for inter-observer agreement was as follows: <0.20 = poor, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, and 0.81–1.00 = near perfect. The metrics used to evaluate the restoration quality of lesion images included sensitivity and positive predictive value (PPV). The

sensitivity was calculated as $TP/(TP + FN)$, and the PPV was calculated as $TP/(TP + FP)$, where TP, FN, and FP stand for true positive, false negative, and false positive values, respectively.

The results of the clinical evaluation and the sensitivity and PPV were tested for significance of the difference by using paired Student t tests ($P < 0.05$).

2.7. Instability testing against small structural changes in images

According to a published method [66], we evaluated the instability of the network with respect to small structural changes for both training and test datasets. Only the FLAIR-SR networks were used for this evaluation. The white letters “Can U see it ●” were added to input images (FLAIR, and T2W-LR) used for evaluation.

3. Results

3.1. Tumor size and numbers

Most of the lesions appearing in the FLAIR-GT images were hyperintense defects in the brain’s white matter. Some postoperative changes and evidence of old strokes were also observed. There was also one case of multiple sclerosis. There were 1188 FLAIR lesions in the 23 patients, with a mean \pm standard deviation (SD) of 51.7 ± 54.4 lesions per patient. The mean and SD of FLAIR lesion sizes were $31.9 \pm 130 \text{ mm}^2$, respectively, with a maximum of 1930 mm^2 and a minimum of 1 mm^2 . There were 154 T1W lesions, with a mean \pm SD of 6.7 ± 6.1 per patient.

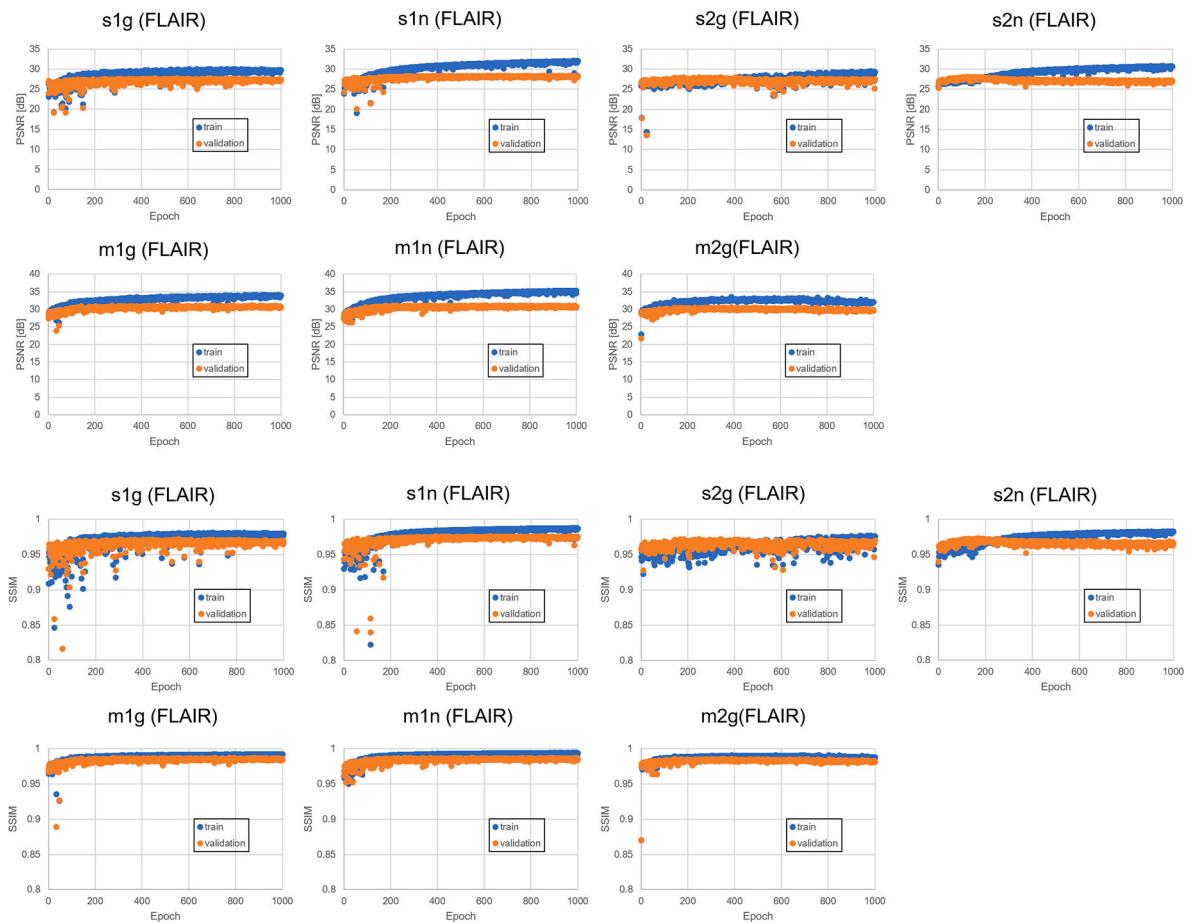


Fig. 6. PSNR and SSIM vs epoch for training and testing for FLAIR.

3.2. Numerical evaluation

Figs. 6 and 7 show PSNR and SSIM vs epoch for all the networks for both training and testing. Table 5 summarizes the mean PSNR and SSIM of FLAIR- and T1W-SR images. For both FLAIR-SR and T1W-SR, the three MCSRs (M1n/M1g/two-stage GAN (M2g)) showed higher PSNR and SSIM values than the four SCSRs (S1n/S1g/S2n/S2g), and the differences were larger in T1W-SR than in FLAIR-SR images. There were insignificant differences in the PSNR and SSIM values between the non-GAN and GAN images.

3.3. Clinical evaluation of FLAIR-SR images

Fig. 8 shows the results of the first-stage clinical evaluation. The numbers of disappeared lesions (false negatives) and false lesions (false positives) per patient for MCSRs (M1n/M1g/M2g) were lower than those for SCSRs (S2n/S2g), and those for non-GANs were lower than those for GANs (S2n vs S2g; M1n vs M1g) (Fig. 8(a)). Among the three MCSRs (M1n/M1g/M2g), M2g had the most disappearing and false lesions.

The restoration quality of the anatomical structures (Fig. 8(b)) showed a different tendency from the restoration quality of lesions. Although there was some variation among the five evaluation items, S2n was generally rated the lowest, and S2g, M1n, M1g, and M2g were rated higher in that order; MCSRs (M1n/M1g/M2g) were higher than SCSRs (S2n/S2g), GAN images tended to have higher ratings than non-GAN images (S2n vs S2g, M1n vs M1g), and M2g had the highest ratings.

Regarding the artifacts (see Fig. 8(c)), the two SCSRs (S2n/S2g) were rated almost the same, whereas the non-GAN image (M1n) was rated higher than GANs (M1g/M2g) among the three MCSRs, and M2g was rated the lowest. There were no significant differences in the overall

Table 5

Mean PSNR and SSIM values for SR images. FLAIR: fluid-attenuated inversion recovery, T1W: T₁-weighted images, T2W: T₂-weighted images, GT: Ground truth, LR: low-resolution, S2n: SC-SRResNet, S2g: SC-SRGAN, M1n: MC-U-Net, M1g: MC-pix2pix, M2g: MC-2stage.

Network	FLAIR-SR		T1W-SR		
	PSNR [dB]	SSIM	PSNR [dB]	SSIM	
S1n	SC-U-Net	35.7	0.968	35.2	0.918
S1g	SC-pix2pix	35.0	0.963	34.6	0.910
S2n	SC-SRResNet	34.9	0.962	34.3	0.905
S2g	SC-SRGAN	34.7	0.961	35.0	0.921
M1n	MC-U-Net	36.5	0.976	37.9	0.954
M1g	MC-pix2pix	36.3	0.975	38.1	0.957
M2g	MC-2stage	36.2	0.974	37.3	0.947

evaluation.

Fig. 9 shows the results of the MOS of the second-stage clinical evaluation for all 23 cases. Clinical evaluation at this stage showed a similar trend to that at the first stage. When comparing non-GAN and GAN images, the non-GAN approach was superior in terms of the lesion's image restoration quality (Fig. 9(a)), whereas GAN was superior in terms of anatomical structure restoration quality (Fig. 9(b)). The non-GAN approach outperformed GAN in terms of artifacts and overall assessment (Fig. 9(c)). There were statistical differences ($P < 0.05$) between GAN and non-GAN on most evaluation items.

3.4. Clinical evaluation of T1W-SR images

The results of the clinical evaluation of T1W SR for all 23 patients are shown in Fig. 10. Each evaluation item showed a trend similar to that of the second-stage evaluation of FLAIR-SR; non-GAN was superior to GAN

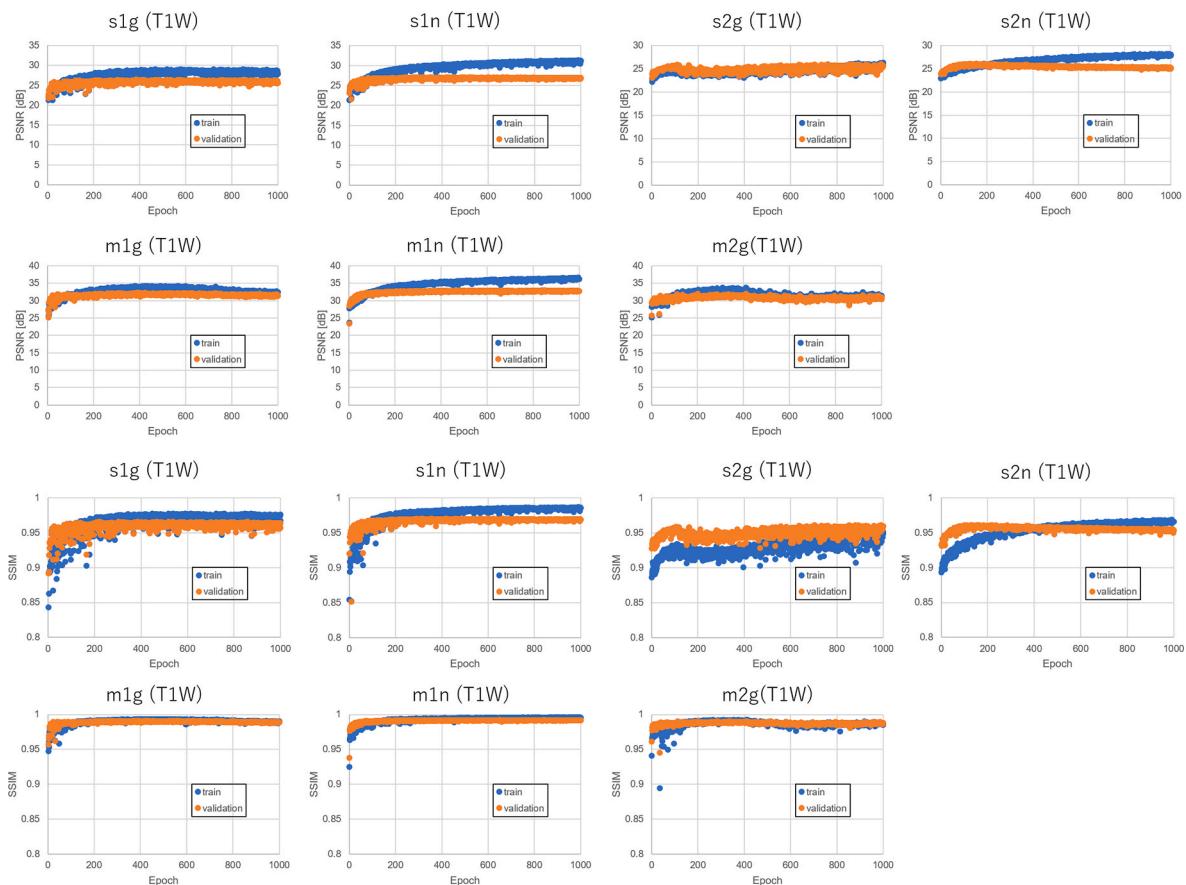


Fig. 7. PSNR and SSIM vs epoch for training and testing for T1W.

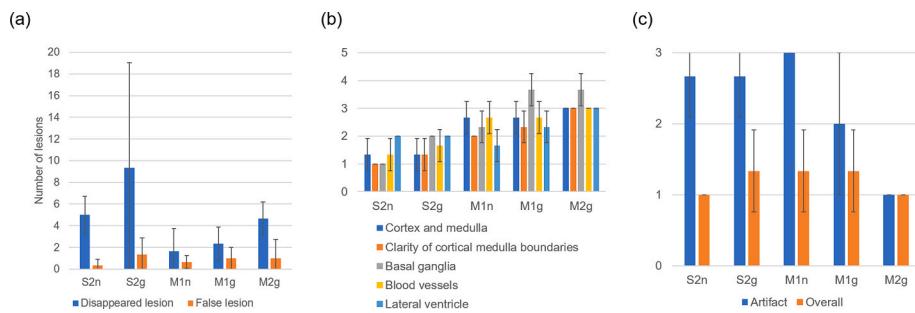


Fig. 8. First-stage clinical evaluation (Rater A, FLAIR). The error bars indicate standard deviations. S2n: SC-SRResNet, S2g: SC-SRGAN, M1n: MC-U-Net, M1g: MC-pix2pix, M2g: MC-2stage.

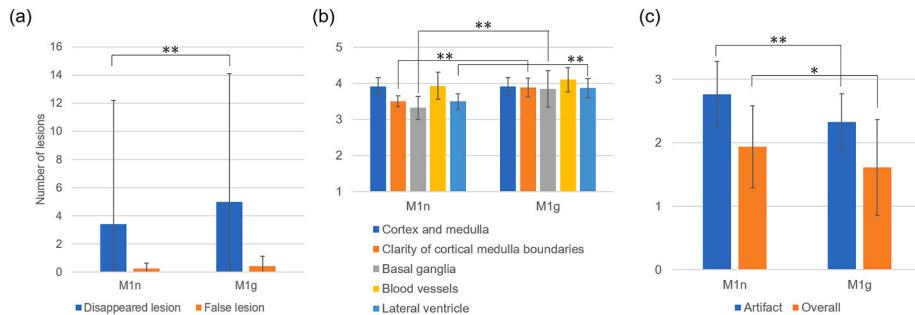


Fig. 9. Second-stage clinical evaluation (MOS, FLAIR). The error bars indicate standard deviations. There were statistical differences (*, $P < 0.05$ and **, $P < 0.01$) between GAN and non-GAN on most evaluation items. FLAIR: fluid-attenuated inversion recovery, MOS: mean-opinion score, M1n: MC-U-Net, M1g: MC-pix2pix.

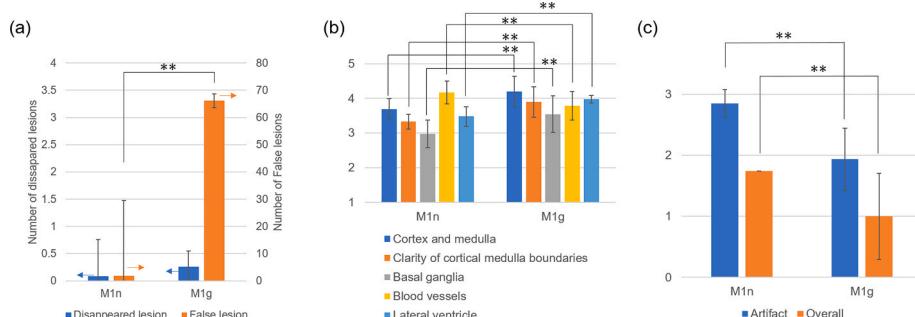


Fig. 10. Second-stage clinical evaluation (MOS, T1W). The error bars indicate standard deviations. The arrows in the figure are drawn to clarify the correspondence between the axes and bars. There were statistical differences (**, $P < 0.01$) between GAN and non-GAN on all evaluation items except the number of disappeared lesions. MOS: mean-opinion score, T1W: T_1 -weighted images, M1n: MC-U-Net, M1g: MC-pix2pix.

in terms of lesion image restoration quality (Fig. 10(a)), artifacts, and overall evaluation (Fig. 10(c)), whereas GAN was superior to non-GAN in terms of the anatomical structure image restoration quality (Fig. 10(b)). The difference between T1W-SR and FLAIR-SR was that the number of false lesions in M1g was much higher for T1W-SR. There were statistical differences ($P < 0.01$) between GAN and non-GAN on all evaluation items except the number of disappeared lesions.

3.5. Inter-observer agreement for clinical evaluation

Table 6 summarizes W values for each clinical assessment. The inter-observer agreement ranged from substantial ($W = 0.50$) to near perfect ($W = 0.90$). In particular, the overall evaluation showed substantial agreement for FLAIR ($W = 0.68$, $P = 0.052$) and near perfect agreement for T1W ($W = 0.90$, $P < 0.01$).

Table 6

Kendall's coefficient of concordance with ties (W) for each clinical assessment. FLAIR: fluid-attenuated inversion recovery, T1W: T_1 -weighted images.

		Cortex and medulla	Clarity of cortical medulla boundaries	Basal ganglia	Blood vessels	Lateral ventricle	Artifact	Overall
FLAIR	Kendall W	0.50	0.50	0.73	<0.05	0.50	0.59	0.68
	P value	0.47	0.47					
T1W	Kendall W	0.60	0.66	0.63	0.50	0.50	0.59	0.90
	P value	0.17	0.07					

3.6. Sensitivity and PPV

The sensitivity and PPV results of the FLAIR- and T1W-SR approaches are shown in Fig. 11. In FLAIR-SR, the sensitivity of M1n was significantly higher than M1g ($P < 0.01$), and the PPV values of M1n were almost as high as those of M1n. For T1W-SR, both M1n and M1g showed similarly high sensitivity values. The PPV was significantly lower for T1W-M1g ($P < 0.01$) because the number of false positives was much higher.

3.7. Example of FLAIR-SR images

Examples of FLAIR-SR images are shown in Fig. 12. In these examples, the trends observed in the clinical assessment are well illustrated. The white-matter lesions indicated by the arrows in Fig. 12 appeared with hyperintense signals on FLAIR-GT and T2W images, but almost disappeared on FLAIR-LR and could not be differentiated on T1W-LR. Comparing the MCSRs and SCSRs (S1n vs M1n, S1g vs M1g), which had the same network architecture except for inputs; these lesions could be differentiated on MCSRs, whereas they were almost invisible in SCSRs. Comparing non-GANs and GANs (S1n vs S1g, S2n vs S2g, and M1n vs M1g), those lesions were more clearly visible in non-GAN images. The clinical evaluation revealed that such white-matter lesions were most visible in the non-GAN-based MCSR (M1n), and a similar trend can be observed in Fig. 12.

The image restoration quality of anatomical structures also showed similar trends to the clinical evaluation. The cortical medulla and corticomedullary border were more clearly delineated in MCSRs than in SCSRs (S1n vs M1n, S1g vs M1g). They were more clearly differentiated for GANs than for non-GANs (S1n vs S1g, S2n vs S2g, M1n vs M1g/M2g), contrary to the lesion image restorations. The texture of the basal ganglia in the output images of GANs (S1g, S2g, M1g, and M2g) was close to that of the GT image.

3.8. Instability test against small structural changes in FLAIR-SR images

The results of the instability test for FLAIR-SR images in the training and test datasets are shown in Supplementary Figs. S2 and S3, respectively. Similar features were observed in these tests. Overall, the restoration quality of the added letters was similar to that of the lesions: in the FLAIR-LR input image, the letters were not very clear, but the output image of M1n showed the clearest letters. In the output images of non-GANs (S1n, S2n, M1n), the letters were blurred but their shapes were not significantly corrupted. On the other hand, in the output images of GANs (S1g, S2g, M1g, M2g), the letters were not blurred, but their shapes were greatly corrupted.

3.9. Examples of T1W-SR images

Examples of T1W-SR images are shown in Fig. 13. The lesion

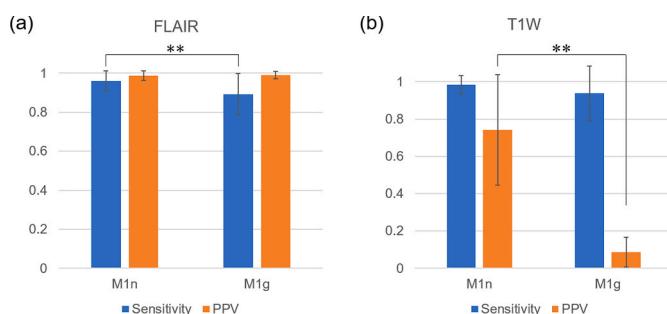


Fig. 11. Sensitivity and positive predictive value (PPV). The error bars indicate standard deviations. ** $P < 0.01$. FLAIR: fluid-attenuated inversion recovery, T1W: T_1 -weighted images, M1n: MC-U-Net, M1g: MC-pix2pix.

indicated by the arrow in this figure had a high signal in the T1W-GT, FLAIR-LR, and T2W input images, but was barely delineated in the T1W-LR input image. This lesion was also barely delineated in the SCSR output images (S1n, S1g, S2n, and S2g) where only the T1W-LR image was used as input. In contrast, this lesion was delineated in the MCSR output images (M1n, M1g, and M2g), which also used the FLAIR-LR and T2W images as input. On the other hand, the lesion was slightly obscured in the M2g output image.

Supplementary Fig. S4 shows an example of false positives and false negatives. The lesion marked with a circle in the figure had a low signal on T1W-GT images, almost disappeared in the FLAIR-LR and T1W-LR input images and appeared with high signal intensity on the T2W input images. This lesion was absent in the SCSR output images (S1n, S1g, S2n, and S2g) but was depicted as low signals in the MCSR output images (M1n, M1g, M2g). The lesion marked by the arrow in Supplementary Fig. S4 had a high signal intensity only in the T2W input image. This lesion was not seen in the T1W-GT or SCSR output images but appeared falsely in the MCSR output images. In addition, the M1g images showed false lesions that were not depicted in the T1W-GT or other input images (white circles in Supplementary Fig. S4). Many such false lesions were observed only in the M1g images, hence explaining the large number of false positives shown in Fig. 10.

4. Discussion

4.1. Main contributions of this work

The main contribution of this study is our clinical evaluation of the baseline SR DNNs of clinical brain MR images, providing guidelines for network design. In particular, we focused on the image restoration quality for lesions, which has not been emphasized so far, and compared the performance between GAN and non-GAN architectures. According to our clinical evaluation by the two expert raters, GANs were superior to non-GAN approaches in restoring the images of anatomical structures but not for lesions, and overall the non-GANs were superior to GAN approaches.

4.2. Non-GAN vs GAN techniques

Our comparison between non-GANs and GANs showed interesting results. There was a trade-off between the restoration quality of lesion and anatomical structure images, with non-GANs being superior in lesion image restoration and GANs being superior in anatomical structure image restoration. In the overall clinical evaluation, non-GANs outperformed GANs because the raters insisted on the accuracy of lesion image restoration. Anatomical brain structures are to some extent universal and independent of the patient, but each lesion is patient-dependent, with small structural changes such that its size and location are diverse. This result suggests that the GAN approach tends to be less tolerant to small structural changes that occur less frequently than in non-GAN images.

The non-GAN approach used here was a generator-only network that used MSE as an evaluation function, just like ordinary CNNs. The GAN, on the other hand, was designed to create synthetic images that deceived the discriminator by adding a discriminator network as a feedback system and training it competitively with the generator. Recent developments in GAN-based DNNs show the possibility for image generation with more realistic textures. Indeed, in this study, the GANs gave better visual impressions of anatomical features than the non-GANs. Brain features such as the cortex, medulla, and basal ganglia were considered major features that did not change much between patients and accounted for a high percentage of the dataset images. In contrast, small structural changes, such as white-matter lesions, were considered minor features that did not account for a large percentage of the dataset images, and thus tended to be ignored (i.e., unstable image restoration).

The observed instability against lesion image restoration could be

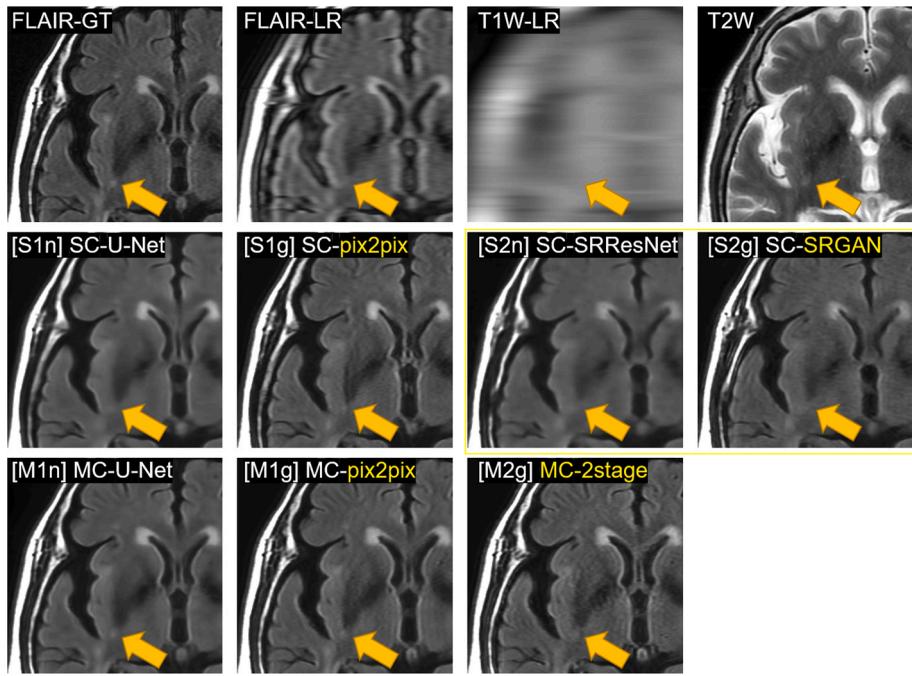


Fig. 12. Example of FLAIR SR. FLAIR: fluid-attenuated inversion recovery, T1W: T_1 -weighted images, T2W: T_2 -weighted images, GT: Ground truth, LR: low-resolution, S2n: SC-SRResNet, S2g: SC-SRGAN, M1n: MC-U-Net, M1g: MC-pix2pix, M2g: MC-2stage.

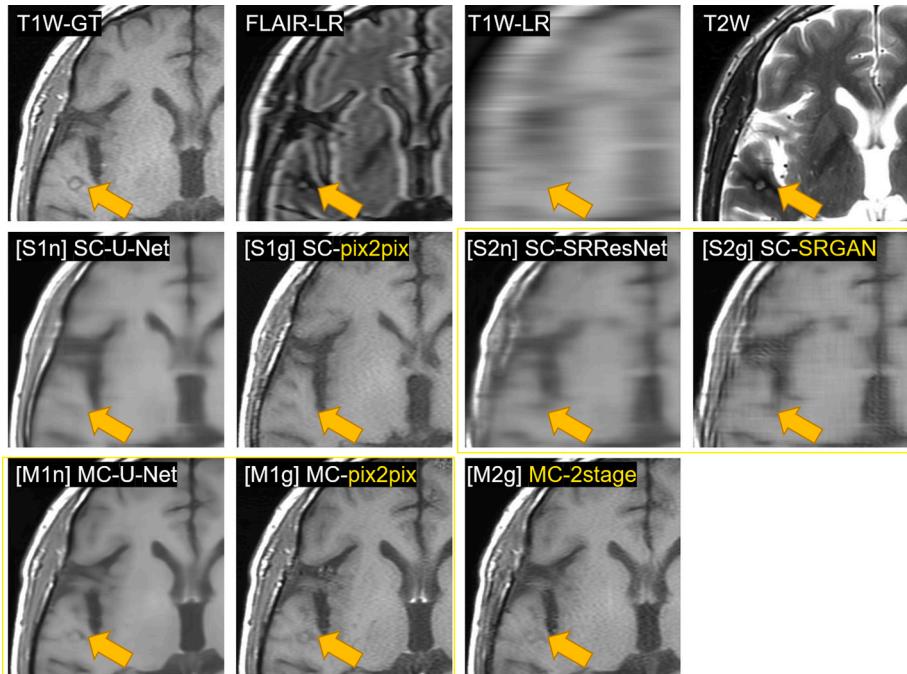


Fig. 13. Example of T1W SR. FLAIR: fluid-attenuated inversion recovery, T1W: T_1 -weighted images, T2W: T_2 -weighted images, GT: Ground truth, LR: low-resolution, S2n: SC-SRResNet, S2g: SC-SRGAN, M1n: MC-U-Net, M1g: MC-pix2pix, M2g: MC-2stage.

similar to the well-known concept of overfitting in machine learning. Overfitting occurs when a network performs well on a training set but poorly on a test set. However, whereas the general concept of overfitting implies a decrease in restoration performance for the entire test image, here, we assume a decrease in the restoration performance in the limited area around the lesion. The imbalance between a small number of lesions compared with healthy tissues in the training dataset might increase the risk of overfitting.

This observed instability might also be related to the

overperformance of DL for image reconstruction. The overperformance can occur for both training and test datasets, and thus is not the same as the concept of overfitting. The instability and performance of DL have been investigated theoretically for DL-based image reconstruction that solves the “inverse problem” [67]. This revealed that it is difficult to remove network instability against small perturbations in the true image or its measurements because of a lack of null space (kernel) awareness; the difference between different tiny perturbations in the input tends to lie close to the kernel of the measurement matrix, and instabilities that

will give rise to false positives and negatives can easily occur. These instabilities are related to the performance of the DL, and there is a delicate balance between the two. This might also be true for the image-to-image SR networks in this study. By analogy, the smaller the difference between the GT and input LR images containing false positives or negatives, the more likely it is that the lesion image restoration will fail. If the performance of the DL is too high, instability can increase. In the case of GAN, the recovery of frequently appearing anatomical structures was improved, and the performance of the SR networks was improved, but at the cost of increased instability for lesion imaging. The instability test (Supplementary Figs. S2 and S3) showed similar trends for training and test datasets, which could support this idea.

In the output T1W images of the MCSR, there were many cases of false positives (Supplementary Fig. S4). Especially in the MC-pix2pix approach, pseudolesions (circled in white in Supplementary Fig. S4) that were not present in the three input images appeared. For lesions that were barely depicted in the T1W-LR images, the T1W-SR network was almost the same as an image synthesis using T2W and FLAIR-LR images as inputs. Although a GAN such as pix2pix is often used as an image synthesis network, the results of this study suggest that great care should be taken against the appearance of such false lesions when implementing image synthesis using GANs.

4.3. Tips for improving the quality of lesion image restoration

There could be several ways to improve the stability of the DL for small structural changes such as lesions. The first method may be to integrate sparse regularization terms in a loss function, as is done in classical, non-DL compressed sensing. Gottschling et al. showed that this works well for DLs solving inverse problems, but at the same time setting regularization parameters is challenging [67], and this could be an issue to be addressed in the future.

Another approach is to increase the resolution of LR images to make it easier to differentiate between small lesions. The DL network could be more stable when lesions are more differentiated in the input images. We designed this study to reduce the scan time as much as possible within the range that can be easily set up on a clinical machine. Indeed, the size of the FLAIR-LR image was 0.529 mm × 2.82 mm, and only 191 lesions (16% of the totals) of this size or smaller were included in the 23 test cases. In addition, in some cases, small tumors could not be identified because they were covered by truncation artifacts. Therefore, increasing the number of phase encodings in the LR images and depicting small lesions—even roughly—in the LR images would be effective in improving network instability. Furthermore, the restoration of lesion images might be improved using data augmentation. Following the above discussion, if the diversity of lesions and the ratio of their occurrence in the dataset images can be increased artificially, kernel awareness will increase and instability can be improved. GAN-based, tumor-specific data augmentation [70–73] could be useful.

4.4. Efficacy of instability testing for the evaluation of lesion image restoration

The results of the instability test to measure the restoration quality of the lesion images were qualitatively consistent with the trend of clinical evaluation by the expert raters, suggesting that such instability testing can be used for the clinical evaluation of image-to-image SR networks. Therefore, it is likely that the use of such instability testing can reduce the burden on the raters to some extent. However, in the case of MCSR, because the lesion's contrast differed between three input images, it was difficult to adjust the contrast of the letters aimed at mimicking the lesion, and we did not perform the instability test on T1W images. Searching for the appropriate contrast of the letters used in MCSR is a future issue.

4.5. Benefits and limitations of the study design

In this study, we evaluated the performance of several SR DNNs clinically in terms of their ability to delineate both lesions and anatomical structures. To this end, we designed the study using the four policies described earlier. The training and testing image datasets used in this study were all acquired during routine clinical examinations and included a variety of lesions. Although this type of study is often conducted retrospectively, here we used a prospective study. The results of this study are useful for implementation and development in a clinical setting, which is an advantage.

On the other hand, there were several limitations in our study design. First, the reduction factor was fixed according to the policy described in (ii). Because the restoration of lesions and structures strongly depends on the reduction factor, it will be necessary to investigate different reduction factors in the future. Second, because of the policy listed in (ii), the reduction factors of FLAIR and T1W were unbalanced, and the reduction factor of T1W was extremely large. Therefore, the T1W-LR images contained little information on anatomical structures or lesions. Increasing the reduction factor for T1W would improve the network performance. Third, the networks used here were not up-to-date. However, they have been used widely in most modern networks, and it should be possible to utilize the knowledge gained in this study in their implementation and development. Forth, no optimization/hyper-parameter tuning was performed in this study. This is ideally necessary, but for simplicity, we used values close to the default values of the original papers [49]. Finally, the ability to detect lesions in SR images is highly dependent on the size of the lesion relative to the restored spatial resolution. Hence, the analysis should ideally be divided into groups of lesions of different sizes.

5. Conclusions

In this study, we focused on image restoration quality of brain lesions in DL-based SR and clinically evaluated the performance for different SR DNNs with GAN and non-GAN architectures. There was a trade-off in performance between GANs and non-GANs, with GANs being more accurate in depicting anatomical structures but less accurate in depicting lesions. This trade-off might have arisen from the diversity and rarity of brain lesions in training images. These results suggest that the evaluation of network performance could be changed if clinical evaluation of lesions is considered, although this aspect has been neglected in previous studies. The dependence of network architecture and dataset on network performance obtained in this study will provide guidance for future SR network development.

Conflict of interests disclosure

The authors declare that they have no conflicts of interest.

Acknowledgments

We thank Dr. Utaroh Motosugi, Department of Radiology at Kofu-Kyoritsu Hospital, for his contribution to the early stage of this study and his fruitful comments on the results.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.imu.2022.101030>.

References

- [1] Park SC, Park MK, Kang MG. Super-resolution image reconstruction: a technical overview. IEEE Acoust Speech Signal Process Newslett 2003;20:21–36. <https://doi.org/10.1109/MSP.2003.1203207>.

- [2] Manjón JV, Coupé P, Buades A, Fonov V, Louis Collins D, Robles M. Non-local MRI upsampling. *Med Image Anal* 2010;14:784–92. <https://doi.org/10.1016/j.media.2010.05.010>.
- [3] Tourbier S, Bresson X, Hagmann P, Thiran J-P, Meuli R, Cuadra MB. An efficient total variation algorithm for super-resolution in fetal brain MRI with adaptive regularization. *Neuroimage* 2015;118:584–97. <https://doi.org/10.1016/j.neuroimage.2015.06.018>.
- [4] Shi F, Cheng J, Wang L, Yap P-T, Shen D. LRTV: MR image super-resolution with low-rank and total variation regularizations. *IEEE Trans Med Imag* 2015;34: 2459–66. <https://doi.org/10.1109/TMI.2015.2437894>.
- [5] Kuklisova-Murgasova M, Quaghebeur G, Rutherford MA, Hajnal JV, Schnabel JA. Reconstruction of fetal brain MRI with intensity matching and complete outlier removal. *Med Image Anal* 2012;16:1550–64. <https://doi.org/10.1016/j.media.2012.07.004>.
- [6] Yang W, Zhang X, Tian Y, Wang W, Xue J-H. Deep learning for single image super-resolution: a brief Review. *IEEE Trans Multimed* 2019;21:3106–21. <https://doi.org/10.1109/TMM.2019.2919431>.
- [7] Bashir SMA, Wang Y, Khan M. A comprehensive Review of deep learning-based single image super-resolution. *PeerJ Comput. Sci.* 2021;7:e621. <https://doi.org/10.7717/peerj.cs.621>.
- [8] Li Y, Sixou B, Peyrin F. A Review of the deep learning methods for medical images super resolution problems. *IRBM* 2021;42:120–33. <https://doi.org/10.1016/j.irbm.2020.08.004>.
- [9] Dong C, Loy CC, He K, Tang X. Learning a deep convolutional network for image super-resolution. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. *Comput. Vis. – eccv 2014*. Cham: Springer International Publishing; 2014. p. 184–99. https://doi.org/10.1007/978-3-319-10593-2_13.
- [10] Kim J, Lee JK, Lee KM. Accurate image super-resolution using very deep convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016. p. 1646–54.
- [11] Kim J, Lee JK, Lee KM. Deeply-recursive convolutional network for image super-resolution. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016. p. 1637–45.
- [12] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2017. p. 3147–55.
- [13] Tan C, Wang L, Cheng S. Image super-resolution via dual-level recurrent residual networks. *Sensors* 2022;22:3058. <https://doi.org/10.3390/s22083058>.
- [14] Ji L, Zhu Q, Zhang Y, Yin J, Wei R, Xiao J, et al. Cross-domain heterogeneous residual network for single image super-resolution. *Neural Netw Off J Int Neural Netw Soc* 2022;149:84–94. <https://doi.org/10.1016/j.neunet.2022.02.008>.
- [15] Shi W, Caballero J, Huszar F, Totz J, Aitken AP, Bishop R, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016. p. 1874–83.
- [16] Dong C, Loy CC, Tang X. Accelerating the super-resolution convolutional neural network. In: Leibe B, Matas J, Sebe N, Welling M, editors. *Comput. Vis. – eccv 2016*. Cham: Springer International Publishing; 2016. p. 391–407. https://doi.org/10.1007/978-3-319-46475-6_25.
- [17] Lim B, Son S, Kim H, Nah S, Mu Lee K. Enhanced deep residual networks for single image super-resolution. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2017. p. 136–44.
- [18] Lai W-S, Huang J-B, Ahuja N, Yang M-H. Deep laplacian Pyramid networks for fast and accurate super-resolution. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2017. p. 624–32.
- [19] Tong T, Li G, Liu X, Gao Q. Image super-resolution using dense skip connections. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2017. p. 4799–807.
- [20] Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y. Residual dense network for image super-resolution. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*; 2018. p. 2472–81.
- [21] Andrew J, Mhatesh TSR, Sebastin RD, Sagayam KM, Eunice J, Pomplun M, Dang H. Super-resolution reconstruction of brain magnetic resonance images via lightweight autoencoder. *Inform Med Unlocked* 2021;26:100713. <https://doi.org/10.1016/j.imu.2021.100713>.
- [22] Tian C, Yuan Y, Zhang S, Lin CW, Zuo W, Zhang D. Image super-resolution with an enhanced group convolutional neural network. *Neural Network* 2022;153:373–85. <https://doi.org/10.1016/j.neunet.2022.06.009>.
- [23] Ahn N, Kang B, Sohn K-A. Fast, accurate, and lightweight super-resolution with cascading residual network. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. *Comput. Vis. – eccv 2018*. Cham: Springer International Publishing; 2018. p. 256–72. https://doi.org/10.1007/978-3-030-01249-6_16.
- [24] Tian C, Xu Y, Zuo W, Zhang B, Fei L, Lin C-W. Coarse-to-Fine CNN for image super-resolution. *IEEE Trans Multimed* 2021;23:1489–502. <https://doi.org/10.1109/TMM.2020.2999182>.
- [25] Lopez-Tapia S, de la Blanca NP. Fast and robust cascade model for multiple degradation single image super-resolution. *IEEE Trans Image Process Publ IEEE Signal Process Soc* 2021;30:4747–59. <https://doi.org/10.1109/TIP.2021.3074821>.
- [26] Anwar S, Barnes N. Densely residual laplacian super-resolution. *IEEE Trans Pattern Anal Mach Intell* 2022;44:1192–204. <https://doi.org/10.1109/TPAMI.2020.3021088>.
- [27] Lan R, Sun L, Liu Z, Lu H, Su Z, Pang C, et al. Cascading and enhanced residual networks for accurate single-image super-resolution. *IEEE Trans Cybern* 2021;51: 115–25. <https://doi.org/10.1109/TCYB.2019.2952710>.
- [28] Zhang Y, Li K, Li K, Wang L, Zhong B, Fu Y. Image super-resolution using very deep residual channel attention networks. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. *Computer vision–ECCV 2018*. *ECCV 2018. Lecture notes in computer science*, vol. 11211. Cham: Springer; 2018. <https://doi.org/10.48550/arXiv.1807.02758>.
- [29] Dai T, Cai J, Zhang Y, Xia S-T, Zhang L. Second-order attention network for single image super-resolution. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*; 2019. p. 11065–74.
- [30] Yang Y, Qi Y. Image super-resolution via channel attention and spatial graph convolutional network. *Pattern Recogn* 2021;112:107798. <https://doi.org/10.1016/j.patcog.2020.107798>.
- [31] Zhang Y, Li K, Li K, Zhong B, Fu Y. Residual non-local attention networks for image restoration. In: *Proceedings of the 7th international conference on learning representations, ICLR 2019*; 2019. <https://doi.org/10.48550/arXiv.1903.10082>.
- [32] Chen Y, Liu L, Phonevilay V, Gu K, Xia R, Xie J, et al. Image super-resolution reconstruction based on feature map attention mechanism. *Appl Intell* 2021;51: 4367–80. <https://doi.org/10.1007/s10489-020-02116-1>.
- [33] Gao S, Zhuang X. Multi-scale deep neural networks for real image super-resolution. In: *Proceedings of the IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshop CVPRW*; 2019. <https://doi.org/10.1109/CVPRW.2019.000252>. 2006–13.
- [34] Hu Y, Gao X, Li J, Huang Y, Wang H. Single image super-resolution via cascaded multi-scale cross network. 2018. <https://doi.org/10.48550/arXiv.1802.08808>. arXiv preprint.
- [35] Li J, Fang F, Li J, Mei K, Zhang G. MDCN: multi-scale dense cross network for image super-resolution. *IEEE Trans Circ Syst Video Technol* 2020;31:2547–61. <https://doi.org/10.1109/TCSVT.2020.3027732>.
- [36] Li J, Fang F, Mei K, Zhang G. Multi-scale residual network for image super-resolution. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. *Comput. Vis. – ECCV 2018*, vol. 11212. Cham: Springer International Publishing; 2018. p. 527–42. https://doi.org/10.1007/978-3-030-01237-3_32.
- [37] Qin J, Huang Y, Wen W. Multi-scale feature fusion residual network for Single Image Super-Resolution. *Neurocomputing* 2020;379:334–42. <https://doi.org/10.1016/j.neucom.2019.10.076>.
- [38] Fan Y, Yu J, Liu D, Huang TS. Scale-wise convolution for image restoration. *Proc AAAI Conf Artif Intell* 2020;34:10770–7_07.
- [39] Yang A, Li L, Wang J, Ji Z, Pang Y, Cao J, et al. Non-linear perceptual multi-scale network for single image super-resolution. *Neural Netw.* 2022;152:201–11. <https://doi.org/10.1016/j.neunet.2022.04.020>.
- [40] Lan R, Sun L, Liu Z, Lu H, Pang C, Luo X. MADNet: a fast and lightweight network for single-image super resolution. *IEEE Trans Cybern* 2021;51:1443–53. <https://doi.org/10.1109/TCYB.2020.2970104>.
- [41] Jiang X, Wang N, Xie J, Xia X, Yang X, Gao X. Learning lightweight super-resolution networks with weight pruning. *Neural Netw.* 2021;144:21–32. <https://doi.org/10.1016/j.neunet.2021.08.002>.
- [42] Tan C, Cheng S, Wang L. Efficient image super-resolution via self-calibrated feature fuse. *Sensors* 2022;22:329. <https://doi.org/10.3390/s22010329>.
- [43] Liu P, Zhang H, Zhang K, Lin L, Zuo W. Multi-level wavelet-CNN for image restoration. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*; 2018. p. 773–82.
- [44] Achieving super-resolution remote sensing images via the wavelet transform combined with the recursive res-net. *IEEE Trans Geosci Rem Sens* 2019;57: 3512–27. <https://doi.org/10.1109/TGRS.2018.2885506>.
- [45] Xue S, Qiu W, Liu F, Jin X. Wavelet-based residual attention network for image super-resolution. *Neurocomputing* 2020;382:116–26. <https://doi.org/10.1016/j.neucom.2019.11.044>.
- [46] Xin J, Li J, Jiang X, Wang N, Huang H, Gao X. Wavelet-based dual recursive network for image super-resolution. *IEEE Transact Neural Networks Learn Syst* 2022;33:707–20. <https://doi.org/10.1109/TNNLS.2020.3028688>.
- [47] Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2017. p. 4681–90.
- [48] Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, et al. ESRGAN: enhanced super-resolution generative adversarial networks. In: Leal-Taixé L, Roth S, editors. *Comput. Vis. – ECCV 2018 workshop*. Cham: Springer International Publishing; 2019. p. 63–79. https://doi.org/10.1007/978-3-030-11021-5_5.
- [49] Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-Image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2017. p. 5967–76.
- [50] Wang T-C, Liu M-Y, Zhu J-Y, Tao A, Kautz J, Catanzaro B. High-resolution image synthesis and semantic manipulation with conditional GANs. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*; 2018. p. 8798–807.
- [51] Park J, Hwang D, Kim KY, Kang SK, Kim YK, Lee JS. Computed tomography super-resolution using deep convolutional neural network. *Phys Med Ampmathsemicol Biol* 2018;63:145011. <https://doi.org/10.1088/1361-6560/aacd44>.
- [52] Mansoor A, Vongkovit T, Linguraru MG. Adversarial approach to diagnostic quality volumetric image enhancement. In: *Proceedings of the IEEE 15th int. Symp. Biomed. Imaging. ISBI 2018*; 2018. p. 353–6.
- [53] You C, Li G, Zhang Y, Zhang X, Shan H, Li M, et al. CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE). *IEEE Trans Med Imag* 2020;39:188–203. <https://doi.org/10.1109/TMI.2019.2922960>.
- [54] Pham C-H, Ducournau A, Fablet R, Rousseau F. Brain MRI super-resolution using deep 3D convolutional networks. In: *Proceedings of the IEEE 14th int. Symp. Biomed. Imaging. ISBI 2017*; 2017. p. 197–200.
- [55] de Leeuw den Bouter ML, Ippolito G, O'Reilly TPA, Remis RF, van Gijzen MB, Webb AG. Deep learning-based single image super-resolution for low-field MR

- brain images. *Sci Rep* 2022;12:6362. <https://doi.org/10.1038/s41598-022-10298-6>.
- [56] Oktay O, Bai W, Lee M, Guerrero R, Kamnitsas K, Caballero J, de Marvao A, Cook S, O'Regan D, Rueckert D. Multi-input cardiac image super-resolution using convolutional neural networks. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W, editors. *Medical image computing and computer-assisted intervention - MICCAI 2016*. Lecture notes in computer science, vol. 9902. Cham: Springer; 2016. https://doi.org/10.1007/978-3-319-46726-9_29.
- [57] Qiu D, Cheng Y, Wang X. Dual U-Net residual networks for cardiac magnetic resonance images super-resolution. *Comput Methods Progr Biomed* 2022;218:106707. <https://doi.org/10.1016/j.cmpb.2022.106707>.
- [58] Sanchez I, Vilaplana V. Brain MRI super-resolution using 3D generative adversarial networks. 2018. <https://doi.org/10.48550/arXiv.1812.11440>. arXiv preprint.
- [59] Zhao M, Wei Y, Wong KKL. A Generative Adversarial Network technique for high-quality super-resolution reconstruction of cardiac magnetic resonance images. *Magn Reson Imaging* 2022;85:153–60. <https://doi.org/10.1016/j.mri.2021.10.033>.
- [60] Zhu J, Tan C, Yang J, Yang G, Lio' P. Arbitrary scale super-resolution for medical images. *Int J Neural Syst* 2021;31:2150037. <https://doi.org/10.1142/S0129065721500374>.
- [61] Ahmad W, Ali H, Shah Z, Azmat S. A new generative adversarial network for medical images super resolution. *Sci Rep* 2022;12:9533. <https://doi.org/10.1038/s41598-022-13658-4>.
- [62] Sui Y, Afacan O, Jaimes C, Gholipour A, Warfield SK. Scan-specific generative neural network for MRI super-resolution reconstruction. *IEEE Trans Med Imag* 2022;41:1383–99. <https://doi.org/10.1109/TMI.2022.3142610>.
- [63] Zhang K, Hu H, Philbrick K, Conte GM, Sobek JD, Rouzrokh P, et al. SOUP-GAN: super-resolution MRI using generative adversarial networks. *Tomography* 2022;8:905–19. <https://doi.org/10.3390/tomography8020073>.
- [64] Zhao C, Carass A, Dewey BE, Woo J, Oh J, Calabresi PA, et al. A deep learning based anti-aliasing self super-resolution algorithm for MRI. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Med. Image comput. Comput. Assist. Interv. – MICCAI 2018*, Lecture notes in computer science, vol. 11070. Cham: Springer; 2018. https://doi.org/10.1007/978-3-030-00928-1_12.
- [65] Zeng K, Zheng H, Cai C, Yang Y, Zhang K, Chen Z. Simultaneous single- and multi-contrast super-resolution for brain MRI images based on a convolutional neural network. *Comput Biol Med* 2018;99:133–41. <https://doi.org/10.1016/j.combiomed.2018.06.010>.
- [66] Antun V, Renna F, Poon C, Adcock B, Hansen AC. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proc Natl Acad Sci U S A* 2020;117. <https://doi.org/10.1073/pnas.1907377117>. 30088–95.
- [67] Gottschling NM, Antun V, Adcock B, Hansen AC. The troublesome kernel: why deep learning for inverse problems is typically unstable. 2020. arXiv preprint, <https://arxiv.org/abs/2001.01258>.
- [68] Chollet, et al. Keras; 2015. <https://keras.io> [accessed 21 July 2022].
- [69] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. 2016. arXiv preprint, <https://arxiv.org/abs/1603.04467>.
- [70] Sandfort V, Yan K, Pickhardt PJ, Summers RM. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Sci Rep* 2019;9:16884. <https://doi.org/10.1038/s41598-019-52737-x>.
- [71] Bowles C, Chen L, Guerrero R, Bentley P, Gunn R, Hammers A, et al. GAN augmentation: augmenting training data using generative adversarial networks. 2018. arXiv preprint, <https://arxiv.org/abs/1810.10863>.
- [72] Mok TCW, Chung ACS. Learning data augmentation for brain tumor segmentation with coarse-to-fine generative adversarial networks. In: Crimi A, Bakas S, Kuifje H, Keyvan F, Reyes M, van Walsum T, editors. *BrainLes: Glioma, multiple sclerosis, stroke and traumatic brain injuries*. BrainLes 2018. Lecture notes in computer science, vol. 11383. Cham: Springer; 2018. https://doi.org/10.1007/978-3-030-11723-8_7.
- [73] Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* 2018;321:321–31. <https://doi.org/10.1016/j.neucom.2018.09.013>.