Received: February 28, 2024. Revised: May 2, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Enhancing the Accuracy of 3D Brain MRI Reconstruction via Super Resolution by Utilizing Residual Variational Autoencoder

Muhammad Ibadurrahman Arrasyid Supriyanto[1] Riyanarto Sarno[1]*
Chastine Fatichah[1] Supriyanto[2]

[1]Department of Informatics, Institut Teknologi Sepuluh November, Surabaya, Indonesia
[2]Faculty of Mathematics and Natural Sciences, Universitas Mulawarman, Samarinda, Indonesia
* Corresponding author's Email: riyanarto@its.ac.id

Abstract: To improve the quality of a medical image, it is essential to employ equipment that is equipped with a higher-intensity magnetic field. This poses unique challenges for healthcare providers, particularly those who have limited resources and an urgent need for efficiency. Interpolation is the most efficient technique for transforming low-resolution photos into high-resolution images through a straightforward calculation. The interpolation technique yields images with reduced clarity, especially along seamless boundaries, leading to the omission of crucial details. This study suggests employing the residual variational autoencoder model for the purpose of reconstructing high-resolution images. The model comprises three components: an encoder, a decoder, and a latent space. According to the test findings, the suggested model performs better than state-of-the-art techniques already in use, including interpolation, multi-level densely connected super-resolution networks, and variational autoencoders. The evaluation of Structural Similarity and Peak signal-to-noise ratio metrics reveals a significant improvement, with an approximate 5–10% rise in Structural Similarity and a 10-15% increase in Peak signal-to-noise ratio compared to the state-of-the-art.

Keywords: Magnetic resonance imaging, Deep learning, Reconstruction, Enhancing, Image processing.

1. Introduction

Medical magnetic resonance imaging (MRI) is a very accurate method of obtaining detailed and in-depth images of the body's organs, bones, and tissues. This examination is performed as a diagnostic tool for doctors. MRI [1, 2] is used in neurology to provide a detailed image of the brain, spine, and blood vessels which can be visualized in three sections: axial, coronal, and sagittal. However, the low image resolution has an impact on the diagnostic process performed by medical personnel [3].

In order to achieve superior medical images, it is necessary to employ costlier gear, acquire samples for a longer period of time, and utilize hardware with a more powerful magnetic field. This poses novel difficulties for healthcare professionals, particularly those with few resources and pressing requirements. Interpolation is the most efficient technique for reconstructing high-resolution images from low-resolution ones using straightforward calculations. Interpolation is the ongoing calculation of the average value of discrete samples. Interpolation [4] is a commonly employed technique in digital image processing to increase, decrease, and correct spatial distortions in images. Efficient interpolation methods are essential because of the abundant data included in digital photographs.

Numerous methodologies for picture interpolation have been suggested in recent decades. The use of these techniques is essential to the advancement of algorithms for analyzing and visualizing medical data in both two and three dimensions [5]. The work conducted by[4]introduced a way for creating 3D image data in the medical field. This method involves utilizing trilinear interpolation, data scaling techniques, and

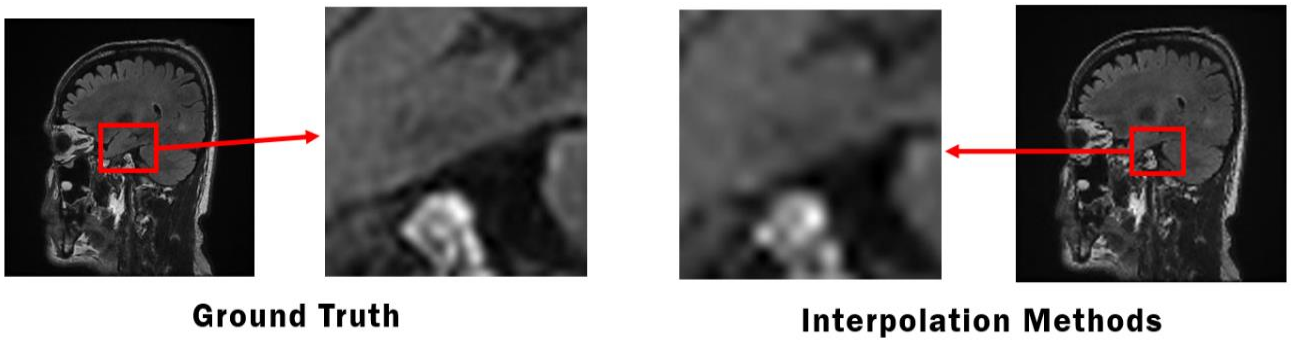

Figure. 1 Comparison of Interpolation Results and Ground Truth on blurred areas

annotation projections to incorporate a large amount of Digital Imaging and Communications in Medicine (DICOM) format information into the images [6-8]. Given the substantial dimensions of 3D picture data and the

constraints of current technology, there is a need for a technique to adjust the size of 3D images. The study's findings indicate that the proposed method is capable of projecting annotations onto scaled images while effectively managing a variety of DICOM image data.

Despite this, the interpolation method [7] makes images less clear, especially in the smoother areas, which means important data is lost, as shown in Fig. 1. Consequently, there is a need for a technique to restore photos from low to high resolution. A study undertaken by [9] utilized technological advancements, particularly in the field of machine learning, to successfully reconstruct high-resolution images from low-resolution inputs. This work aims to address the persistent challenge of reconstructing low-resolution 3D medical images into high-resolution 3D medical images by drawing upon many studies conducted on this subject.

To enhance the quality of reconstructed 3D images from interpolation, researchers propose using a model called residual variational autoencoder. This model aims to address the issue of blurry and less detailed image structures resulting from the interpolation process. By utilizing the residual variational autoencoder, the researchers aim to improve the clarity and detail of the reconstructed images. This approach offers a promising solution to the challenges associated with 3D image reconstruction, ensuring that the resulting images accurately represent the underlying structures with enhanced resolution and sharpness.

The paper is divided into key sections: Introduction (Section 1) provides background, problem statement, and objectives. Methodology (Section 2) outlines the approach, including the use of Residual Variational Autoencoder. Section 3 covers Data Explanation, detailing datasets and preprocessing. Implementation and Results (Section 4) discusses practical application, results, and analysis, demonstrating the efficacy of the proposed method in enhancing 3D brain MRI reconstruction accuracy via super resolution.

The contributions made include:

1. The proposed method aims to enhance the accuracy of reconstructing three-dimensional brain MRI images.
2. This research utilizes super resolution techniques to enhance the resolution of MRI images. Increasing image resolution allows for clearer visualization of small details in brain structures
3. The proposed method was outperformed by previous research methods in reconstructing low-resolution 3D medical images into high-resolution 3D medical images.

2. Methodology

2.1 Related work

Advancements in neuroimaging research, particularly focused on brain magnetic resonance imaging (MRI) scans, have introduced innovative techniques to enhance image processing. For instance, [17] proposed a multi-level densely connected super-resolution network (mDCSRN), trained using GAN methodology, for processing 3D data. This model, utilized in the study, offers rapid and realistic image reconstruction from the nifti dataset, demonstrating promising results with a patch dimension of (64,64,64) [18]. Additionally, researchers [19] presented an algorithm for automatic detection of the anterior and posterior commissure in MRI data, requiring no extensive training and compatible with T1-W and T2-W MRIs.

However, despite these advancements, challenges persist in resizing 3D images due to technology constraints. In response, [5] developed a method utilizing trilinear interpolation, 3D image

scaling, and annotation projection to create 3D image data from DICOM format medical images. While this approach effectively applies annotations and handles diverse DICOM image data, it suffers from reduced image clarity, especially along seamless boundaries, leading to significant detail loss. To address this, [13] introduced a super-resolution technique using an autoencoder model to generate high-resolution images from low-resolution inputs. Additionally, [20] devised a Super Resolution Generative Adversarial Network model (SRGAN) that effectively reconstructed images, outperforming standard super-resolution models in human visual perception despite lower PSNR values. These studies collectively highlight ongoing efforts to improve image reconstruction and address the challenges in neuroimaging research.

2.2 Variational autoencoder

The method proposed in this study is a development of the variational autoencoder method in this study [10, 11]. The proposed Residual Variational Super Resolution Autoencoder (ResVae) method also consists of encoder, decoder, and latent space layers. Similar to previous research, the encoder layer is tasked with extracting features from LR and GT images, resulting in a latent space that is randomly sampled to obtain the mean and variance. Reparameterization is performed for backpropagation, allowing feature estimation when reconstructed to original image dimensions in the decoder layer. The encoder layer represents the posterior approximation $q_\varphi(z|x)$, the decoder layer represents the generative $p_\theta(x,z)$.

The posterior approximation is represented as a multivariate gaussian with a diagonal covariance matrix. Multi-Layer Perceptron (MLP) is used to calculate the parameters of the gaussian distribution using x as input, characterized by nonlinear and linear functions. The MLP is described by two nonlinear functions, μ_φ and σ_φ, which correspond to the mean and standard deviation vectors, respectively, and two linear functions, μ_φ and σ_φ, which map the mean and standard deviation vectors from x.

$$q_\emptyset(z|Y) = \mathcal{N}(z; \mu_\emptyset(Y), \sigma_\emptyset(Y), I) \quad (1)$$

The generative model $p_\theta(x,z)$ assumes that $p(z)$ is determined by Gaussian multivariate units, specifically $p(z) = N(0,I)$. If we assume that the functions $\sigma\theta$ and $\mu\theta$ are nonlinear, mapping the standard deviation vector and mean to z, respectively, we can conclude that

$$p_\theta(\hat{Y}|z) = \mathcal{N}(\hat{Y}; \mu_\theta(z), \sigma_\theta(z), I) \quad (2)$$

We can identify why this model is named an autoencoder based on its network architecture.

$$Y \xrightarrow{q_\emptyset(z|Y)} z \xrightarrow{p_\theta(Y|z)} \hat{Y} \quad (3)$$

The encoder $q_\emptyset$ probabilistically translates the input Y to the code z, which is then probabilistically translated back into the input space by the decoder $p_\emptyset$. We extract samples $z^i, l = 1 \ldots . l$ from the conditional distribution $q_\emptyset(z|Y)$. In order to reduce the variability of our gradient estimation, we employ the noise distribution of a randomly picked epsilon from the Gaussian multivariate unit, denoted as , i.e., $p(\epsilon) \sim N(0,1)$,to carry out a reparameterization operation on the multivariate Gaussian distribution $q_\theta(z|Y)$. The reparameterization of q_φ can be expressed in the following manner:

$$z^{(l)} = \mu_\emptyset(Y) + \varepsilon \odot \sigma_\emptyset(Y) \quad (4)$$

Where

$$\epsilon^{(l)} \sim N(0,1) \quad (5)$$

Backpropagation is done by calculating the gradient by adding up the mean and variance values and then multiplying it by the random epsilon value, which epsilon in this study is stochastic.

2.3 Stacking DICOM

The acquired MRI medical picture is a per-slice image in the DICOM file format. To generate a three-dimensional image, it is necessary to arrange the data in a stacked manner. The DICOM stacking performed in this study is based on a prior study [8] that requires the stacking to be done according to the slice order specified in the DICOM metadata.

In Fig. 2 above, the value z is the number of pixels per row x and the number of pixels per

```
(0020, 1041) Slice Location                 DS: '-62.201228293506'
(0028, 0000) Group Length                   UL: 162
(0028, 0002) Samples per Pixel              US: 1
(0028, 0004) Photometric Interpretation     CS: 'MONOCHROME2'
(0028, 0010) Rows                           US: 256
(0028, 0011) Columns                        US: 240
(0028, 0030) Pixel Spacing                  DS: [1.0546875, 1.0546875]
(0028, 0100) Bits Allocated                 US: 16
(0028, 0101) Bits Stored                    US: 12
(0028, 0102) High Bit                       US: 11
(0028, 0103) Pixel Representation           US: 0
(0028, 0106) Smallest Image Pixel Value     US: 0
(0028, 0107) Largest Image Pixel Value      US: 1037
(0028, 1050) Window Center                  DS: '374.0'
(0028, 1051) Window Width                   DS: '832.0'
```

Figure. 2 DICOM Metadata

column y were represented by the number of slices. the DICOM tag (0028, 0010) and the DICOM tag(0028, 0011). Once the values of $x, y,$ and z are known, a 3D representation is created by stacking the medical images. The spacing between slices refers to the DICOM tag (0028, 0030) which is the spacing between slices, while the order of the slice positions refers to the DICOM tag (0020, 1041) slice location.

2.4 Patching data

Patching is employed to mitigate the issue of constrained memory. The researchers utilize only 16 GB of RAM to perform the patching process on each 3D dataset, using the methodology employed in a prior study [12]. To calculate the number of batches and the total path formed, it can be seen using the following equation (6).

$$batch_{x,y,z} = \frac{x_{dim}, y_{dim}, z_{dim}}{Split} \quad (6)$$

Where the dimension value is the initial dimension value of the image to be patched and the split is the desired dimension per patch value. If the batch value is float, it must be padded on z_{dim} as much as the value of x_{dim} or y_{dim}. After finding the next batch value, the calculation of the number of patches in each dataset is carried out using the following equation (7):

$$\sum patch = x_{batch} \times y_{batch} \times z_{batch} \quad (7)$$

Where the batch value on each dimension (x, y, z) is obtained from the previous equation.

2.5 Loss

In the research that will be carried out, loss calculation is denoted as l^{SR} which consists of the sum of MSE loss and reconstruction loss.

$$l^{SR} = l_{mse} + (l_{bce} + l_{KL}) \quad (8)$$

The MSE loss value is obtained using the equation for 3D reconstruction according to a study [13] :

$$MSE = \frac{1}{n}\left(\hat{Y}_i - Y_i\right)^2 \quad (9)$$

Where $\hat{Y}_i$ is the reconstructed image with dimensions (x, y, z) while Y_i is the ground truth image. Meanwhile, the reconstruction loss obtained in equation (11, 12) which is referenced from a study [14] using cross entropy is as follows:

$$loss\left(\hat{Y}, Y_i\right) = \{l_1 \ldots \ldots . l_n\} \quad (10)$$

$$l_n = -w_n\left(Y_i \cdot \log \hat{Y} + (1 - Y_i) \cdot \log\left(1 - \hat{Y}\right)\right) \quad (11)$$

Where $\hat{Y}_i$ is a reconstructed image with dimensions (x, y, z), while Y_i is the ground truth image that the target Y_i must be a number between 0 and 1. Kullback-Leibler divergence loss is a statistical method used to calculate the distance between data distributions so that it can be written in the following notation:

$$KL\left(\hat{Y}, Y\right) = Y \times \log\frac{Y}{\hat{Y}} = Y \times \left(\log Y - \log \hat{Y}\right) \quad (12)$$

Where Y and $\hat{Y}$ are the same dimensions and Y is a ground truth and $\hat{Y}$ is output.

2.6 Evaluation metric

The testing protocol involves taking photos from both ADNI and local datasets to compare the results of implementing a model on 3D data to generate high-resolution medical images. Assessment metrics like structural similarity and PSNR are used to evaluate image quality [13]. Better quality is indicated by a higher PSNR number. The SSIM metric, proposed by [15], evaluates the decline in reconstructed image quality, with values closer to 1 indicating a strong resemblance between the high-resolution MRI image and the reconstructed super-resolution MRI image.

3. Data

3.1 ADNI dataset

The study utilized brain MRI data from ADNI [16], a biomedical research organization, to study brain-related regions. The data was formatted in DICOM, with each point assigned unique $x, y,$ and z coordinates, allowing for detailed analysis. The dataset consists of 92 data points and is classified into multiple groups depending on diagnosis of Alzheimer's disease. These groups consist of specific numbers of individuals within certain age ranges and gender distributions. The first group includes six males aged 55 to 57 years. The second group consists of 43 cognitively normal individuals aged 55 to 88 years, with 13 males and 30 females.

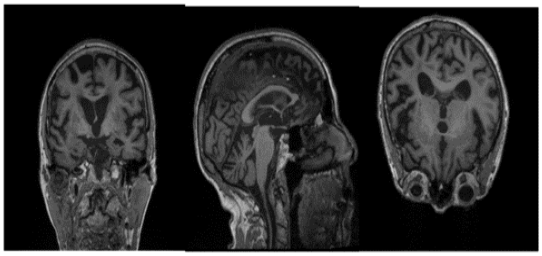

Figure. 3 Visualization of ADNI dataset from left coronal, sagittal and axial

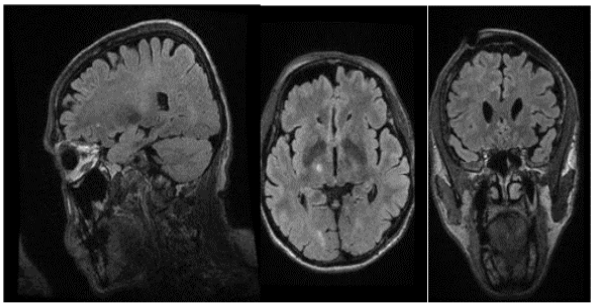

Figure. 4 Visualization of ADNI dataset from left axial, coronal and sagittal

The third group is composed of five individuals aged 73 to 89 years with early mild cognitive impairment, including four males and one female. The fourth group is made up of six individuals aged 72 to 83 years with late mild cognitive impairment, evenly split between males and females. The fifth group consists of 27 individuals aged 55 to 88 years with mild cognitive impairment, including six males and 21 females. The final group includes five individuals aged 73 to 80 years with significant memory concerns, with three males and two females.

Fig. 3 is a 3D visualization of the ADNI dataset, whose structural and functional parts were scanned using the 3T siemens machine performed by ADNI on all datasets used. The protocol configuration used is for structural using T1-MPRAGE, voxel size of $1.2\ x\ 1.1, 1.1$ mm3, image matrix of $(256,256)$ and number of slices of 170 and flip angle of 9°.

3.2 Local dataset

In addition to using the ADNI dataset, the researchers added local datasets like in Fig. 4. A total of 90 pieces were obtained from several hospitals in Surabaya, East Java Province, Indonesia.

This dataset is in the form of DICOM and was obtained using a 1.5 Tesla device. In this study, two categories of brain MRI data were used with different configurations including 92 ADNI datasets obtained using 3 Tesla devices and the other 90 local datasets obtained using 1.5 Tesla devices. The total number of datasets used in the study can be calculated using the following equation:

$$\sum dataset = (\sum dataset_{Adni} + \sum dataset_{lokal}) \times patch \quad (13)$$

Where $\sum dataset_{Adni}$ is the number of ADNI datasets used and $\sum dataset_{lokal}$ is the number of local datasets used. The patch is the number of patches in each dataset that have been patched according to equation (7).

4. Implementation and result

4.1 Training and parameter

The experiment started with low-resolution data from the patching dataset approach and was carried out on Google Colab with a Tesla K80 GPU and 16 GB RAM. The data was in the format of $(256,256,256)$ dimensions. Afterwards, the dataset was partitioned into segments of $(32,32,32)$, resulting in eight batches for each dimension (x, y, z). The process of dividing resulted in a grand total of 512 patches. Based on the calculation using the equation provided above, the batch and number of patches are as follows:

$$batch = \frac{256,256,256}{32} = 8,8,8 \quad (14)$$

$$\sum patch = 8 \times 8 \times 8 = 512 \quad (15)$$

The total number of datasets used in the study can be calculated by the following equation (16) :

$$\sum dataset = (92 + 90) \times 512 = 93.184 \quad (16)$$

As a result, a total of $93,184$ datasets were acquired, each having dimensions of $(32,32,32)$. The dataset was divided into two segments for this research 70% for training and 30% for testing. Consequently, the training dataset comprised $65,229$ instances $(70\%\ of\ 93,184)$, while the testing dataset consisted of $27,955$ instances $(30\%\ of\ 93,184)$. For the down-sampling and up-sampling procedures in this study, the trilinear interpolation technique was utilized. The initial scaling factor for down-sampling was 2, and the latent dimensions employed in this study were set at a constant value of 128. For the training phase, an Adam optimizer was utilized, with an initial learning rate of $1e-3$. The training procedure lasted for 100 epochs.

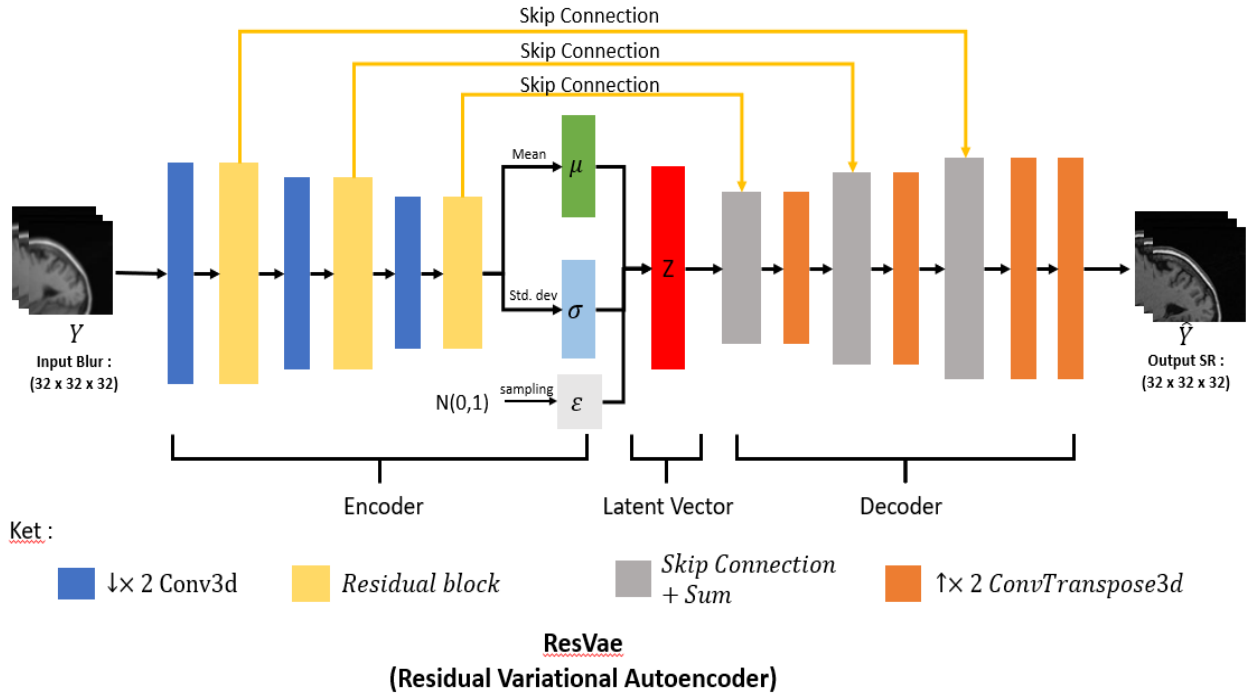

Figure. 5 Visualization of the proposed neural network architecture residual variational autoencoder

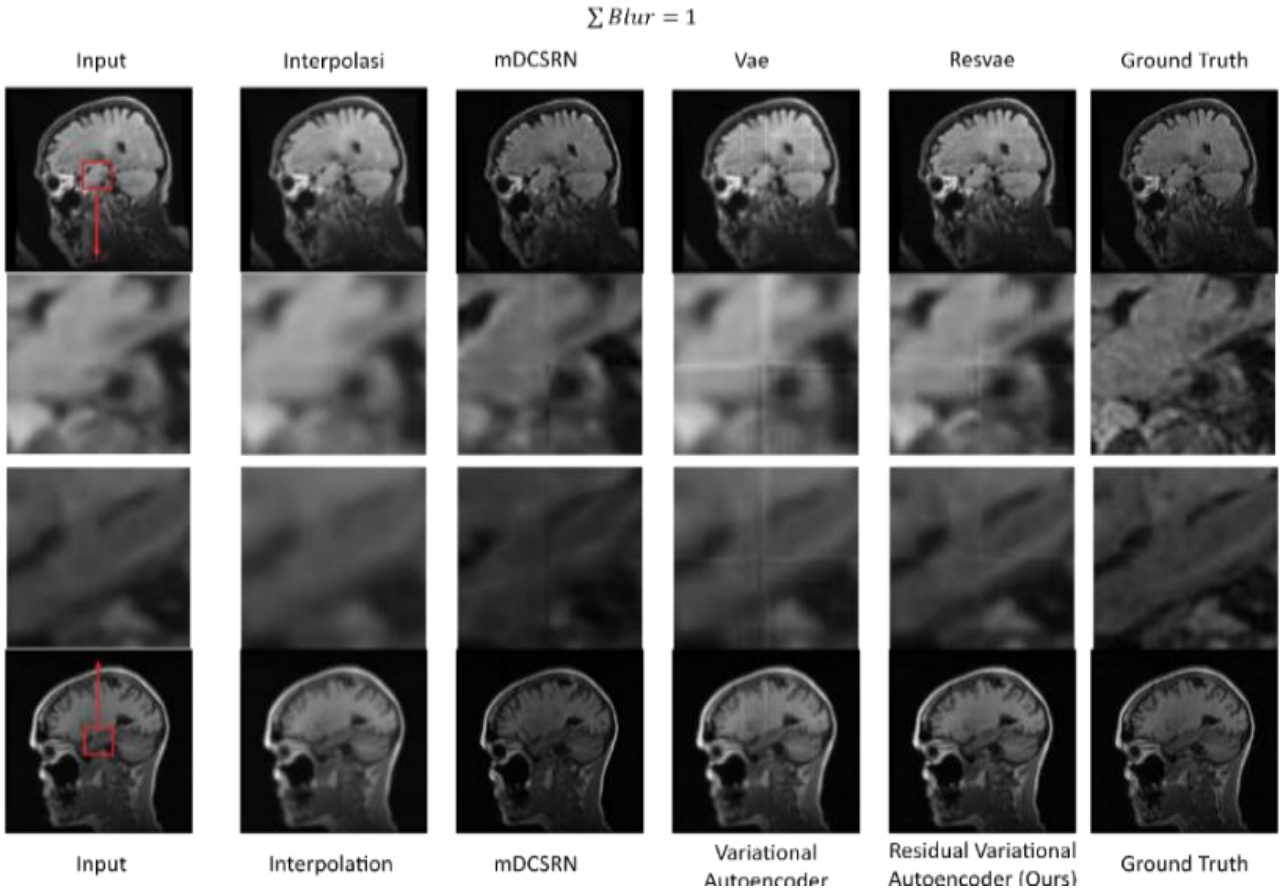

Figure 6. Comparison of the reconstruction results of the residual variational autoencoder model with other models with a blur value of sigma 1 (top) local dataset, (bottom) ADNI dataset

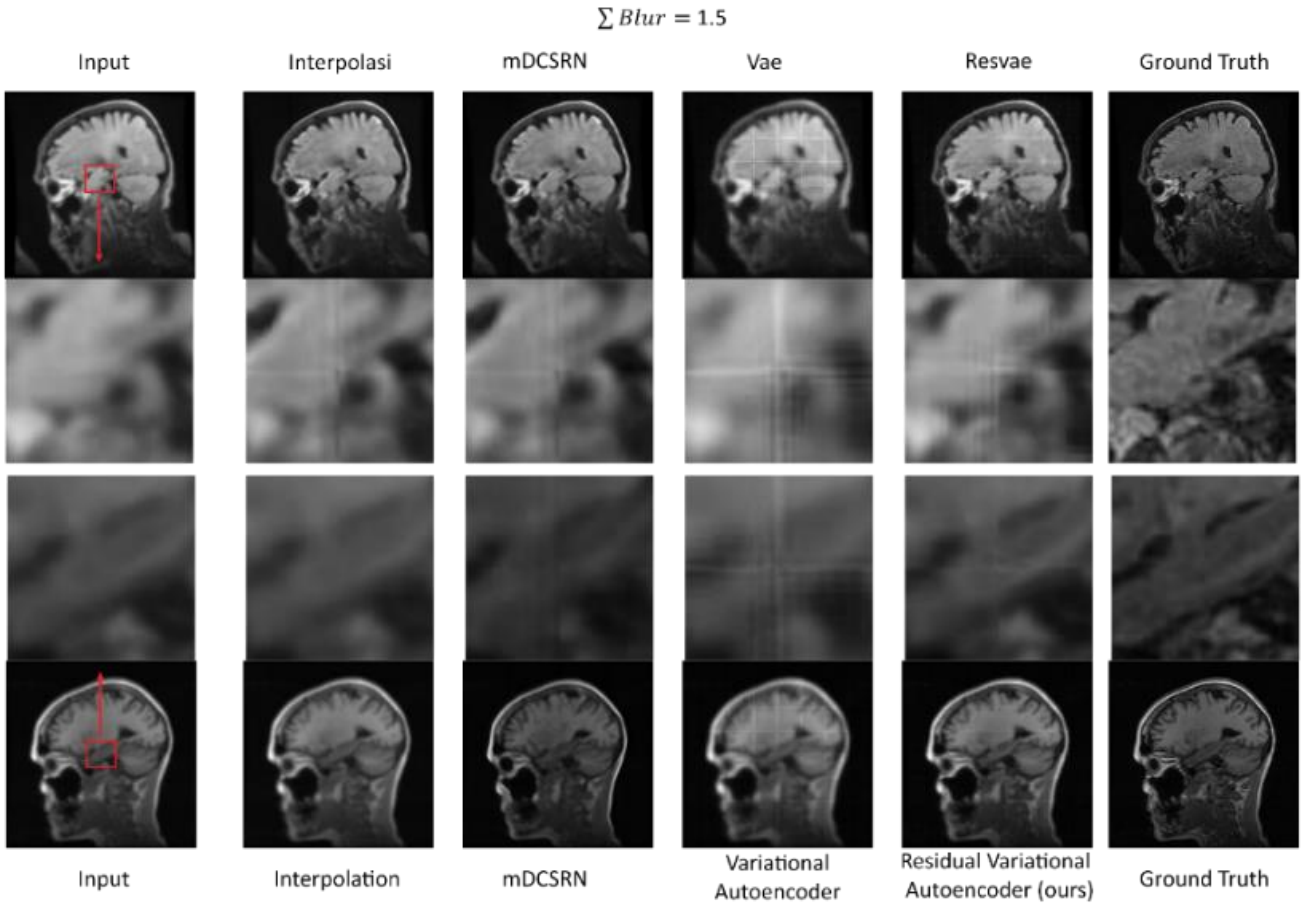

Figure. 7 Comparison of the reconstruction results of the residual variational autoencoder model with other models with a blur value of sigma 1.5 (top) local dataset, (bottom) ADNI dataset

4.2 Architecture residual variational autoencoder

Fig. 5 shows the design of the neural network architecture of the super-resolution residual variational autoencoder for the case of 3D magnetic resonance imaging of the brain. This architecture consists of an encoder layer, a z layer, and a decoder. The input of this neural network is a 3D low-resolution image that has been patched. The difference between a residual variational autoencoder and an existing variational autoencoder is the use of a residual model for feature extraction and the upsample process.

The encoder layer depicted in Fig. 8 is designed to decrease the dimensions of the input by utilizing convolution. This layer is composed of Residual Network blocks [17], with each block including four convolutions using normalized and leaky ReLU (rectified linear unit) activation, followed by a skip connection. The convolutions employ a (3,3,3) kernel and a stride of 2, with the initial number of filters set at [16, 32, 64]. The employed normalization technique was Group Normalization (GN) [18], which demonstrates superior performance compared to batch normalization when the batch size is limited.

In this scenario, the value 1 is employed because to the significant impact on memory usage caused by a high batch norm value. The spatial dimensions of the final encoder layer are (64,3,3,3), which is 8 times smaller than the input image.

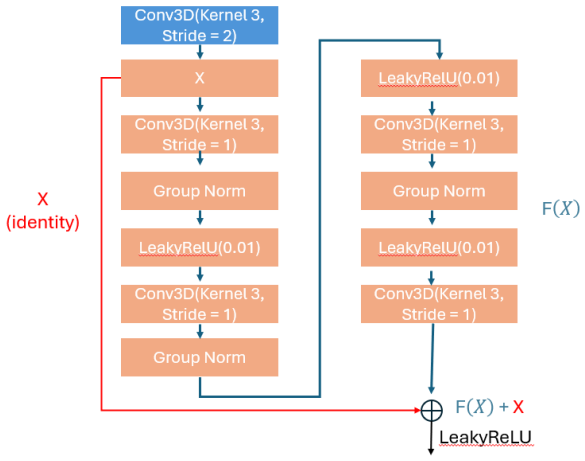

Figure. 8 Residual Neural Network Blocks in Layer Encoder

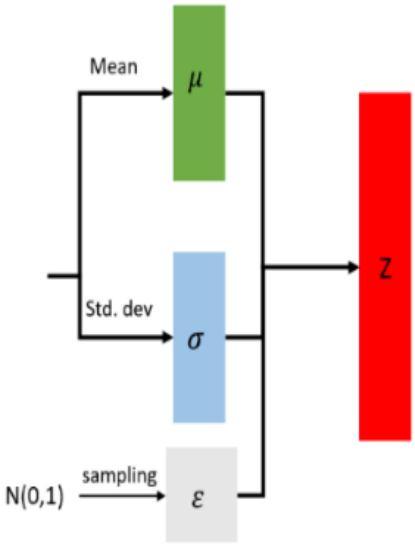

Figure. 9 Mean and Variance in Latent Vector Layers

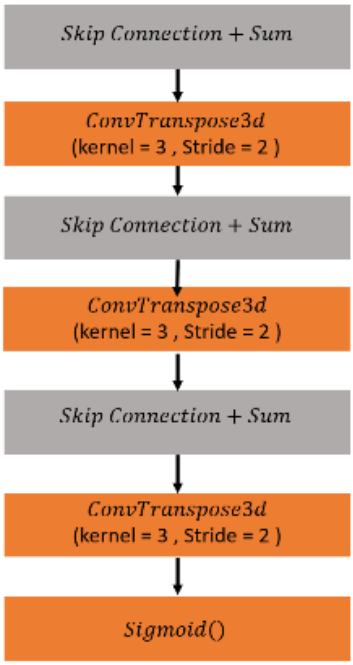

Figure. 10 Residual Neural Network Blocks in Layer Decoder

The z layer, seen in Fig. 9, is a latent variable that arises from the process of feature extraction performed by the decoder layer $q_{\emptyset}(z|Y)$. Subsequently, this layer is partitioned into z, which comprises a dataset composed of a stochastic sampling of the average and variability, utilizing a latent space value of 128. The latent variable contains the attributes of the feature extraction process. The aim of this study is to determine the mean and variance values that align with the ground truth image. Nevertheless, the decoder layer is unable to carry out backpropagation due to its reliance on integral computations for the purpose of maximizing distribution. The reparameterization approach outlined in equation (4) to compute the z values that were sampled from the mean and variance.

The layer decoder seen in Fig. 10 is designed to carry out the process of increasing the resolution of the latent space. The decoder's structure closely resembles that of the encoder, with progressively larger block dimensions at each spatial level. The decoder level commences with a skip connection, followed by a 3D convolution transposition that decreases the number of features using a kernel size of (3,3,3) and a stride of 2. The researchers opted using convolution transpose instead of trilinear upsampling due to the inclusion of skip connections and adds, which necessitated an adjustment to the weight of the leaky ReLU (rectified linear unit) activation function. The ultimate decoder layer possesses identical dimensions to the input image.

4.3 Reconstruction using residual variational autoencoder

To evaluate the suggested residual variational autoencoder model. The researchers conducted a comparison of the image reconstruction outcomes, which are illustrated in Fig. 7 and 8. The study produced super-resolution images using several comparison methods, and the results are shown here. The data utilized in this investigation consists of photos with low resolution. The study consists of two distinct stages: sigma 1 and sigma 1.5. An increase in the sigma value in this example results in a reduction in the image intensity, causing the image to appear blurred.

Fig. 7 demonstrates that the autoencoder residual variational model outperforms other approaches, such as interpolation, the multi-level densely connected super-resolution network (mDCSRN) model, and the variational autoencoder, in generating higher-quality images on both the ADNI and local datasets when using Sigma 1. The residual variational autoencoder and mDCSRN models enhance image sharpness by reducing the blur effect, resulting in clear visibility of intricate areas of the brain. This is evidenced by the elevated PSNR values between the two models. However, the mDCSRN model demonstrated a considerable decline in reconstruction outcomes on the ADNI datasets, as seen by a substantial decrease in SSIM values compared to the reconstructions of local data images. This decrease causes the image reconstruction result to be darker.

Table 1. The comparison of SSIM and PSNR metrics evaluation results is conducted on ADNI datasets.

Methods	Sigma	Dataset ADNI	
		SSIM	PSNR
Interpolation [4]	1	0.87	22.89
	1.5	0.83	21.83
mDCSRN [19]	1	0.75	19.20
	1.5	0.80	22.31
Variational Autoencoder [10]	1	0.88	23.19
	1.5	0.84	22.17
Residual Variational Autoencoder (Ours)	1	**0.93**	**26.03**
	1.5	**0.89**	**24.22**

The outcome of the experiments depicted in Fig. 8 demonstrates that, despite increasing the sigma value, the residual variational model autoencoder remains proficient in reconstructing accurately. Upon thorough examination of the reconstructed residual variational model of the autoencoder, distinct characteristics of the brain organ are still discernible in comparison to alternative approaches. This is the point where the most significant degradation in image quality takes place within the variational autoencoder model. If there is a strong resemblance between two images that are conceptually distinct in the autoencoding variational model, a bias will arise, resulting in significant blurring of the image.

The evaluation results of the research conducted as shown in Table 1 compare four different super-resolution methods: interpolation, mDCSRN, variational autoencoder, and the proposed residual variational autoencoder method using two datasets, namely ADNI and local datasets. The proposed technique on the ADNI dataset gets a structural similarity evaluation metric value of (0.93) for sigma 1. The value is 6.1% greater than the ssim value generated by the variational autoencoder method (0.88) and 19.3% greater than the mDCSRN method, which only attained a value of (0.75). Compared to the interpolation method, the proposed method shows a 2.6% increase in SSIM value with respect to the (0.87) interpolation strategy. In addition, the proposed strategy demonstrates a PSNR value of (26.03 dB), which exceeds the performance of the variational autoencoder method by 11.1% (23.19 dB) and the mDCSRN method by 27.5% (19.20 dB). Concerning the interpolation technique, the proposed method achieves a PSNR value of (22.89 dB), exhibiting an improvement of 11.3% compared to the PSNR value of the interpolation method.

Moreover, the ADNI dataset has a sigma value of 1.5. The proposed approach gets an evaluation score of (0.89) for the SSIM metric, outperforming the variational autoencoder method by 5.6% (0.84) and the mDCSRN method by 10.7% (0.80). The interpolation strategy has a value of (0.83), which is 3.8% higher than the ssim value of the proposed method. The proposed strategy obtains a PSNR value of (24.22dB), beating the PSNR values of the variational autoencoder method (22.17dB) by 8.5% and the mDCSRN method (22.31dB) by 4.5%. Concerning the interpolation approach, the proposed method exhibits a PSNR value that surpasses the interpolation method by 8.5%, reaching (21.83dB).

After that, when used on separate datasets, as seen in Table 2, especially the local dataset with a sigma value of 1, the proposed approach obtains an SSIM evaluation metric value of (0.90). The value beats the variational autoencoder method by 7.1%, with a value of (0.83), and exceeds the mDCSRN method by 11.3%, with a value of (0.79). Regarding the interpolation approach, the value is (0.81), showing an 8.7% rise in comparison to the SSIM value of the proposed method. The proposed

Tabel 2. The comparison of SSIM and PSNR metrics evaluation results is conducted on local datasets.

Methods	Sigma	Local Datasets	
		SSIM	PSNR
Interpolation [4]	1	0.81	20.37
	1.5	0.75	18.45
mDCSRN [19]	1	0.79	22.52
	1.5	0.66	15.73
Variational Autoencoder [10]	1	0.83	20.35
	1.5	0.77	18.09
Residual Variational Autoencoder (Ours)	1	**0.90**	**24.01**
	1.5	**0.85**	**21.67**

approach attains a PSNR of (24.01dB), outperforming the variational autoencoder technique by 14.9% (20.35 dB) and the mDCSRN technique by 7.8% (22.52 dB). Concerning the interpolation technique, the proposed method achieves a PSNR value of 20.37 dB, surpassing the PSNR value of the interpolation method by 14.5%. Moreover, the dataset that is accessible locally is similarly relevant to Sigma 1.5. The proposed strategy produces an SSIM value of (0.85), surpassing the SSIM value of (0.77) obtained by the variational autoencoder method by 9.3% and surpassing the SSIM value of (0.66) produced by the mDCSRN method by 22.3%. In relation to the interpolation approach, the proposed method exhibits an SSIM value that is 11.7% higher than the interpolation method when the value is set to (0.75). Furthermore, the proposed strategy exhibits a PSNR value of (21.67dB), surpassing the variational autoencoder method by 16.4% with a PSNR value of (18.09dB) and the mDCSRN method by 36.9% with a PSNR value of (15.73dB). Concerning the interpolation approach, the proposed method demonstrates a PSNR value that is 18.2% higher than that of the interpolation method.

In summary, our study showcases the superior performance of the proposed residual variational autoencoder method compared to other advanced techniques, including variational autoencoder, mDCSRN, and interpolation. Notably, when sigma was set to 1 and 1.5, both the ADNI and local datasets exhibited significantly higher SSIM and PSNR values. On average, our method demonstrates a remarkable 5–10% enhancement in SSIM and a notable 10–15% improvement in PSNR compared to the current state-of-the-art approach. These findings underscore the potential of our proposed method as a highly effective solution for achieving superior super-resolution results. Our research suggests that this approach holds considerable promise in advancing the field of image processing and could lead to substantial improvements in various practical applications.

5. Conclusion

In summary, the research introduces the residual variational autoencoder model, comprising an encoder, decoder, and latent space, for reconstructing high-resolution images. The encoder identifies low-resolution image features, while reparameterization facilitates backpropagation and feature estimation.

The results of this research, indicate that the proposed residual variational autoencoder method outperforms existing state-of-the-art techniques, including variational autoencoder, mDCSRN, and interpolation, in terms of both structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) metrics. When evaluated on the ADNI dataset with a sigma value of 1, the proposed method achieves a significantly higher SSIM value of 0.93, representing a 6.1% improvement over variational autoencoder and a 19.3% improvement over mDCSRN. Similarly, the PSNR value of 26.03 dB for the proposed method outperforms both variational autoencoder and mDCSRN by 11.1% and 27.5%, respectively. Furthermore, on the local dataset with a sigma value of 1, the proposed method demonstrates SSIM and PSNR values of 0.90 and 24.01 dB, respectively, surpassing variational autoencoder and mDCSRN by significant margins. The evaluation of Structural Similarity and Peak signal-to-noise ratio metrics reveals a significant improvement, with an approximate 5–10% rise in Structural Similarity and a 10-15% increase in Peak signal-to-noise ratio compared to the state-of-the-art. These findings underscore the scientific contribution of the proposed method, highlighting its effectiveness in achieving superior super-resolution results compared to existing approaches.

Future studies are poised to leverage artificial intelligence generative models for super-resolution image reconstruction, particularly in the realm of 3D MRI medical images.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

This research was made possible thanks to contributions from previous studies. Conceptualization, Riyanarto Sarno; methodology, Chastine Fatichah; software, writing—original draft preparation, and editing, visualization, Muhammad Ibadurrahman Arrasyid Supriyanto; validation, Supriyanto; supervision, Riyanarto Sarno, Chastine Fatichah, and Supriyanto.

Acknowledgments

The funding for this research was provided by the Asian Development Bank through the HETI ADB Project, Institut Teknologi Sepuluh Nopember (ITS) through the Penelitian Flagship Program, the Indonesian Ministry of Education and Culture through the Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) Program, and Institut Teknologi

Sepuluh Nopember (ITS) through the Publication Writing and IPR Incentive Program (PPHKI).

References

[1] D. Sunaryono, J. Siswantoro, R. Sarno, R. I. Susilo, and S. I. Sabilla, "Epilepsy Detection using Combination DWT and Convolutional Neural Networks Based on Electroencephalogram", In: *Proc. of 2023 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, pp. 411–4162023, doi: 10.1109/ISITIA59021.2023.10221031.

[2] A. Fajar, R. Sarno, C. Fatichah, R. I. Susilo, and G. Pangestu, "Cyclical Learning Rate Optimization on Deep Learning Model for Brain Tumor Segmentation", *IEEE Access*, Vol. 11, pp. 119802–119810, 2023, doi: 10.1109/ACCESS.2023.3326475.

[3] Q. Lyu, H. Shan, and G. Wang, "MRI Super-Resolution With Ensemble Learning and Complementary Priors", *IEEE Trans Comput Imaging*, Vol. 6, pp. 615–624, 2020, doi: 10.1109/TCI.2020.2964201.

[4] W. Q. Shiaa, A. B. Abdulghafour, and O. H. Hassoon, "A Novel Method Based on Interpolation for Accurate 3D Reconstruction from CT Images", *International Journal of Intelligent Engineering and Systems*, Vol. 16, No. 2, pp. 506–516, 2023, doi: 10.22266/ijies2023.0430.41.

[5] A. Goshtasby, D. A. Turner, and L. V. Ackerman, "Matching of tomographic slices for interpolation", *IEEE Trans Med Imaging*, Vol. 11, No. 4, pp. 507–516, 1992, doi: 10.1109/42.192686.

[6] M. I. A. Supriyanto, R. Sarno, C. Fatichah, and A. Fajar, "A Comparison Between Interpolation Method and Neural Network Approach in 3D Digital Imaging and Communications in Medicine", In: *Proc. of 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, IEEE, pp. 869–873, 2023, doi: 10.1109/ICCoSITE57641.2023.10127803.

[7] M. I. A. Supriyanto, A. Fajar, R. Sarno, C. Fatichah, A. Fahmi, S. A. Utomo, and F. Notopuro, "Slice Reconstruction on 3D Medical Image using Optical Flow Approach", In: *Proc. of 2021 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob)*, IEEE, pp. 242–246, 2021, doi: 10.1109/APWiMob51111.2021.9435238.

[8] A. Fajar, R. Sarno, C. Fatichah, and A. Fahmi, "Reconstructing and resizing 3D images from DICOM files", *Journal of King Saud University - Computer and Information Sciences*, 2020, doi: 10.1016/j.jksuci.2020.12.004.

[9] C. C. and T. X. Dong Chao and Loy, "Accelerating the Super-Resolution Convolutional Neural Network", In: *Proc. of Computer Vision – ECCV 2016*, pp. 391–407, 2016.

[10] H. V. Guleria, A. M. Luqmani, H. D. Kothari, P. Phukan, S. Patil, P. Pareek, K. Kotecha, A. Abraham, and L. A. Gabralla, "Enhancing the Breast Histopathology Image Analysis for Cancer Detection Using Variational Autoencoder", *Int J Environ Res Public Health*, Vol. 20, No. 5, 2023, doi: 10.3390/ijerph20054244.

[11] Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, and L. Carin, "Variational Autoencoder for Deep Learning of Images, Labels and Captions", In: *Proc. of Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016*, Barcelona, Spain, pp. 2352–2360, 2016. [Online]. Available: https://proceedings.neurips.cc/paper/2016/hash/eb86d510361fc23b59f18c1bc9802cc6-Abstract.html

[12] I. Sanchez and V. Vilaplana, "Brain MRI super-resolution using 3D generative adversarial networks", *Medical Imaging with Deep Learning*, 2018. [Online]. Available: https://openreview.net/forum?id=rJevSbniM

[13] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network", In: *Proc. of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos,

CA, USA, pp. 105–114, 2017, doi: 10.1109/CVPR.2017.19.

[14] Z. Zhang and M. Sabuncu, "Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels", In: *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., Curran Associates, Inc., 2018. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2018/file/f2925f97bc13ad2852a7a551802feea0-Paper.pdf

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp. 600–612, 2004, doi: 10.1109/TIP.2003.819861.

[16] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and M. W. Weiner, "Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization", *Neurology*, Vol. 74, No. 3, pp. 201–209, 2010, doi: 10.1212/WNL.0b013e3181cb3e25.

[17] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution", In: *Proc. of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, pp. 1132–1140, 2017, doi: 10.1109/CVPRW.2017.151.

[18] K. Wu Yuxin and He, "Group Normalization", In: *Computer Vision – ECCV 2018*, M. and S. C. and W. Y. Ferrari Vittorio and Hebert, Ed., Cham: Springer International Publishing, pp. 3–19, 2018.

[19] F. and C. A. G. and X. Y. and Z. Z. and L. D. Chen Yuhua and Shi, "Efficient and Accurate MRI Super-Resolution Using a Generative Adversarial Network and 3D Multi-level Densely Connected Network", In: *Proc. of Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pp. 91–99, 2018.