Journal of Digital Imaging (2023) 36:240–249
https://doi.org/10.1007/s10278-022-00686-9

Inter-Slice Resolution Improvement Using Convolutional Neural Network with Orbital Bone Edge-Aware in Facial CT Images

Hee Rim Yun[1] · Min Jin Lee[1] · Helen Hong[1] · Kyu Won Shim[2]

Received: 10 August 2021 / Revised: 13 July 2022 / Accepted: 14 July 2022 / Published online: 22 August 2022

Abstract

The 3D modeling of orbital bones in facial CT images is essential to provide a customized implant for reconstructions of orbit and related structures during surgery. However, 3D models of the orbital bone show an aliasing effect and disconnected thin bone in the inter-slice direction because the slice thickness is two to three times larger than the pixel spacing. To improve the inter-slice resolution of facial CT images, we propose a method based on a 2D convolutional neural network (CNN) that uses the spatial information on the sagittal and axial planes and the orbital bone edge-aware (OBE) loss. First, intermediate slices are generated on the sagittal plane. Second, the generated intermediate slices are transformed to an axial image, which is then compared with the original axial image. To generate intermediate slices with an accurate orbital bone structure, the OBE loss considering the orbital bone structure on the sagittal and axial planes is used. To improve the perceptual quality of the generated intermediate slices, the feature map difference loss is additionally used on the axial plane. In the experiment, the proposed method showed the best performance among bilinear and bicubic interpolations, 3D SRGAN, and a 2D CNN-based method. Experimental results confirmed that the proposed method can generate intermediate slices with clear edges of thin bones as well as cortical bones on both the sagittal and the axial plane.

Keywords Facial CT · Orbital bone · Thin bone · Inter-slice resolution · Convolutional neural network

Introduction

The orbit is a socket of the skull with a four-sided pyramid structure that protects the eyes and blood vessels. As shown in Fig. 1a, the orbit is composed of the orbital roof, the orbital lateral wall, the orbital medial wall, and the orbital floor. In particular, the thin bones of the orbital medial wall and the orbital floor are easily fractured, even with a weak impact. In order to reconstruct the orbit surgically, it is necessary to create a customized implant by reconstructing the orbital bone into a 3D model [1]. However, as shown in Fig. 1b, the 3D model of the orbital bone shows an aliasing effect and disconnected thin bone in the inter-slice (z-axis) direction, as the slice thickness of the CT image is two to three times larger than the pixel spacing. Therefore, it is necessary to improve the inter-slice resolution of facial CT images before reconstructing the orbital bone into a 3D model.

Interpolation is a conventional method used to improve the inter-slice resolution, and bilinear interpolation [2] and bicubic interpolation [3] are the most commonly used forms with two-dimensional images. Bilinear interpolation is a resampling method that estimates a new pixel value using the distance-weighted average of the four nearest pixel values. Bicubic interpolation is a resampling method that estimates new pixel values using a cubic spline with the weighted sum of the sixteen nearest pixel values. These methods are fast and convenient but often generate over-smoothed boundaries.

Recently, deep-learning-based methods such as the convolutional neural network (CNN) or generative adversarial network (GAN) have been proposed to improve the inter-slice

✉ Helen Hong
hlhong@swu.ac.kr

✉ Kyu Won Shim
shimkyuwon@yuhs.ac

[1] Department of Software Convergence, Seoul Women's University, 621 Hwarang-ro, Nowon-gu, Seoul, Republic of Korea

[2] Department of Pediatric Neurosurgery, Craniofacial Reforming and Reconstruction Clinic, Yonsei University College of Medicine, Severance Children's Hospital, 50 Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea

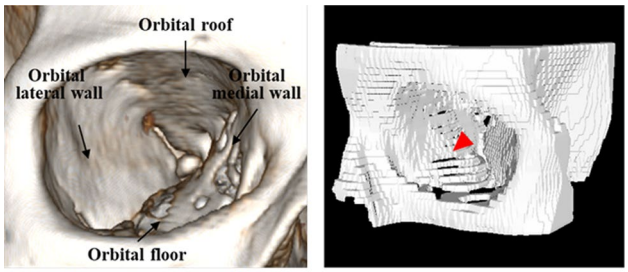

Fig. 1 Characteristics of the orbital bone: **a** structure of orbit, **b** 3D-reconstructed model of the orbital bone from CT images with a thick slice thickness. The red arrow indicates the disconnection of the orbital floor

resolutions of medical images. In the CNN-based method, Bae et al. [4] proposed a 2.5D network based on a fully residual CNN to improve the inter-slice resolution of chest CT images and used it to detect lung nodules. The 2.5D network uses consecutive coronal images as input to consider the spatial correlation between the coronal and sagittal planes. Georgescu et al. [5] proposed a 2D CNN-based network to improve the inter-slice resolution of brain MRI. The proposed 2D CNN takes coronal images as input. The network consists of two convolutional blocks. The first convolutional block is for generating images with improved inter-slice resolutions, and the second convolutional block is for refining the detail differences between the generated image and the original image. Jurek et al. [6] proposed a 2D CNN-based method to improve the inter-slice resolution of brain MRI and utilized it for vessel radius estimations. Yun et al. [7] proposed Orbital Bone-Super Resolution (OB-SR) based on a 2D CNN with sagittal Thin-Bone Structure Aware (TSA) loss to improve the inter-slice resolution of head-and-neck CT images. These 2D CNN-based methods produce images of much higher quality than interpolation-based methods. However, because these methods only use spatial information on one plane, such as the coronal or sagittal plane, the boundaries are jagged on the axial plane of the generated intermediate slices. The basic idea of the GAN is to train the generator to produce images with improved inter-slice resolutions while simultaneously training the discriminator to distinguish a given image as either original or generated. Li et al. [8] proposed 3DSRGAN based on a super-resolution GAN (SRGAN) to improve the inter-slice resolution of T2 FLAIR head MRI. Delannoy et al. [9] proposed segSRGAN to improve the inter-slice resolution of neonatal brain MRI and to segment the brain in the generated image. van der Ouderaa et al. [10] applied a reversible GAN (RevGAN) to improve the inter-slice resolutions of chest CT images. Kudo et al. [11] proposed the Virtual Thin Slice (VTS) method based on a conditional GAN (cGAN) to improve the inter-slice resolutions of CT images of various body parts. These methods typically use the mean squared error (MSE) or mean absolute error (MAE) as the loss function. Given that the MSE and MAE are defined based on the difference in the intensity values in pixels, they cannot restore fine details such as edges [12]. Several GAN-based methods have been proposed to generate images with clear edges using the gradient difference as a loss function of the generator. Liu et al. [13] proposed Edge-Enhanced SRGAN (EE-SRGAN), a 2D network that uses a hybrid loss function, which is a combination of the MSE loss and the edge loss, to improve the inter-slice resolutions of brain MRI. Chai et al. [14] proposed a 2D network, Edge-Guided GAN (EG-GAN), to improve the inter-slice resolutions of brain MRI. The input of EG-GAN is a coronal image with the intermediate slices being blank rows. EG-GAN estimates a gradient map of the blank rows and then estimates the intensity values of the blank rows based on the estimated gradient map. Although these methods generate images with sharper boundaries than methods that use only MSE or MAE as the loss function, they are insufficient when used to consider small objects such as thin bones, as they minimize the gradient difference in the entire image.

In this paper, we propose a method for improving the inter-slice resolution to reduce the aliasing effect of the 3D-reconstructed orbital bone. To improve the inter-slice resolutions of facial CT images, intermediate slices are generated using spatial information of the sagittal plane. To generate intermediate slices with an accurate orbital bone structure on the axial and sagittal planes, the generated intermediate slices are transformed to the axial planes and compared with the original axial images. To generate intermediate slices with clear edges of thin bones as well as the cortical bone, an orbital bone label mask is used to consider the information of the orbital bone structure. To improve the perceptual quality of the generated intermediate slices on the axial plane, the feature map difference is used as an additional loss function.

Materials

Our study was approved by the Institutional Review Board of Severance Hospital, Yonsei University College of Medicine, Seoul, Korea (IRB number: 4–2016-0603). The dataset used in this paper included facial CT images of 355 patients, consisting of 217 men and 138 women, aged 13 to 83 years with a mean age of 34.4 years. The CT images were acquired from six different CT scanners, including those by GE Medical Systems (Revolution EVO, Revolution CT), Siemens (Sensation 64, Somatom Definition Flash, Definition AS +), and Philips (Philips iCT 256 scanner), with settings of 100 kVp and 100 ~ 300 mAs. The CT images were reconstructed using the following seven kernels: STANDARD, H30f, H31f, H31s, H40s, UB, and ULTRA. Each CT

image had a matrix size of 512 × 512 pixels with in-plane resolutions ranging from 0.4 to 0.619 mm and with a slice thickness of 1 mm. The dataset was randomly divided into 228 datasets for training, 56 datasets for validation, and 71 datasets for testing.

Methods

Overview

The proposed method, shown in Fig. 2, consists of two major stages: intermediate slice generation on the sagittal plane and orbital bone quality improvement on the axial plane. In the first stage, intermediate slices with clear edges of the orbital bone are generated using consecutive sagittal CT images and orbital bone label masks. In the second stage, to refine the edges of the orbital bone, the generated intermediate slices are transformed to the axial plane image and compared with the original axial image in the orbital bone area.

Data Preprocessing

The intensity values of trabecular and thin bone are distributed in the range of 50 ~ 200 HU, the intensity values of soft tissue are distributed in the range of – 200 ~ 100 HU, and the intensity values of cortical bone are mostly over 1000 HU. To highlight the orbital bone and surrounding tissue regions in the facial CT image, the intensity range of [– 200 HU, 400 HU], which includes the intensity values of thin bone, surrounding soft tissues, cortical bone, and trabecular bone, is rescaled to the gray scale [0, 255]. Intensity rescaling is performed using the following equation:

$$I_{new} = \left(\frac{I - I_{min}}{I_{max} - I_{min}} \right) \times 255 \quad (1)$$

where I is the intensity value of the CT image before intensity rescaling and I_{min} and I_{max} are – 200 HU and 400 HU, representing the minimum and maximum intensity values of the range to be rescaled, respectively.

In order to consider only the orbital bone area in the facial CT image, the region of interest (ROI) is defined as 172 × 205 × 48 in size, which is the maximum size to account for the orbital bones of all patients. To generate the input image of the proposed network, the CT images are downsampled to half the size of the original image in the inter-slice direction. To consider as well the information of orbital bone area, an orbital bone label mask with a 1-pixel dilation level is also used as input.

Intermediate Slice Generation in Sagittal Plane

When using typical loss functions such as MSE or MAE, which account for intensity differences, the edges of thin bones may be blurred in the generated intermediate slices because thin bones have low intensity values, similar to those of the surrounding soft tissues. Therefore, to consider a thin bone with a small area, the orbital bone edge-aware (OBE) loss is

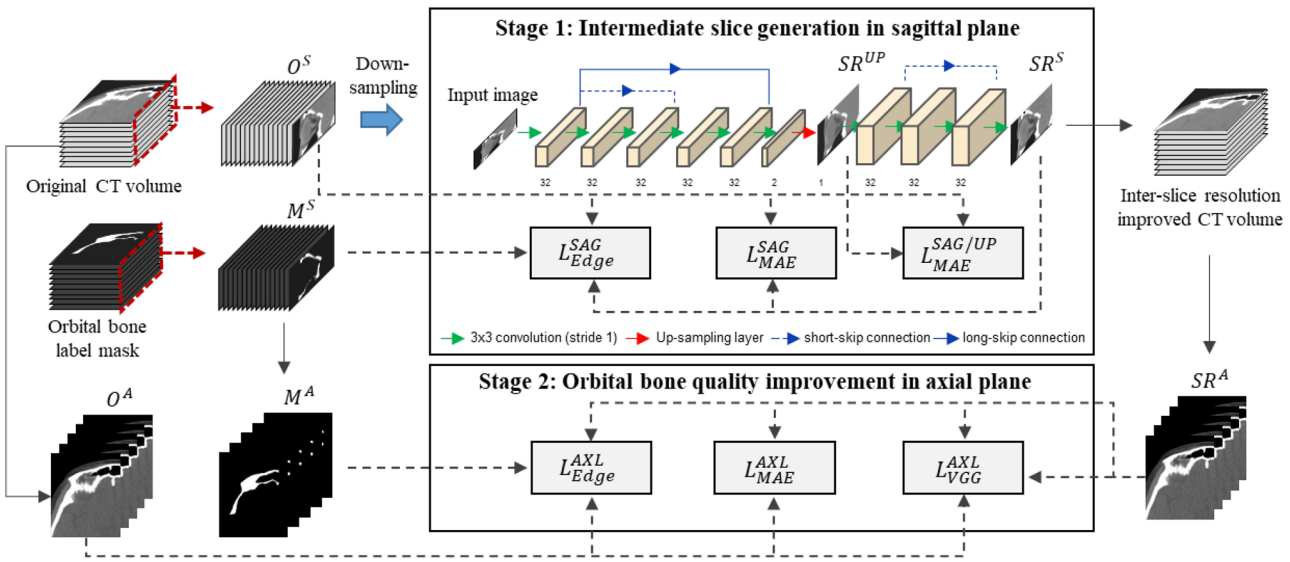

Fig. 2 Architecture of the proposed network. Green and red arrows indicate the convolutional and up-sampling layers, respectively. The blue arrow with the solid line indicates a long-skip connection and the blue arrows with the dashed lines indicate short-skip connections. The numbers below the output after the convolutional layer (green arrow) indicate the number of feature maps in each convolution layer

used together with the orbital bone label mask to minimize the gradient differences of the orbital bone.

A 2D CNN consists of six convolutional layers for feature extraction, an up-sampling layer to generate the intermediate slices, and four convolutional layers for image detail refinement. To extract the feature maps, convolutional layers with a 3×3 kernel size are used, with zero-padding applied to keep the size of each feature map equal to the input image. In the first six convolutional layers, each convolutional layer from the first to fifth extracts 32 feature maps, and the sixth convolutional layer extracts two feature maps. To maintain information from low-level feature maps such as textures and gradients, a short skip connection and a long skip connection add the feature map of the first convolutional layer to the feature maps of the third and fifth convolutional layers, respectively. After the sixth convolutional layer is completed, an up-sampling layer using sub-pixel convolution [15] is completed to double the resolution in the inter-slice direction. To improve the differences that still remain in the details between the original image and the output image of the up-sampling layer, four additional convolutional layers using a 3×3 kernel size are completed. The first to third convolutional layers generate 32 feature maps, and the last convolutional layer outputs the final image of one channel.

The sagittal loss function consists of two MAE losses and the OBE loss. The MAE loss is calculated the difference in intensity between the original image and the generated image using Eq. (2). To refine the image detail, another MAE loss value is calculated as the intensity difference between the original image and the output image of the up-sampling layer using Eq. (3).

$$L_{MAE}^{SAG} = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}\left|SR_{i,j}^{S} - O_{i,j}^{S}\right| \tag{2}$$

$$L_{MAE}^{SAG/UP} = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}\left|SR_{i,j}^{UP} - O_{i,j}^{S}\right| \tag{3}$$

In these equations, W and H are respectively the width and height of the original sagittal image and i and j are likewise the indices of the pixel location. SR^{UP} is the output image of the up-sampling layer, SR^{S} is the final output image of the last convolutional layer, and O^{S} is the original sagittal image.

To minimize the gradient difference only in the orbital bone area, the orbital bone label mask is used as an indicator of the orbital bone area. The OBE loss is calculated using Eq. (4).

$$L_{Edge}^{SAG} = \sum_{i=1}^{W}\sum_{j=1}^{H}\left|E(SR^{S})_{i,j} - E(O^{s})_{i,j}\right| * M_{i,j}^{S} \tag{4}$$

where W and H are correspondingly the width and height of the original sagittal image, i and j are the indices of the pixel location, SR^{S} is the final output image of the last convolutional layer, and O^{S} is the original sagittal image. Additionally, $E(SR^{S})$ and $E(O^{S})$ are respectively the edge maps extracted from SR^{S} and O^{S} using Sobel operator and M^{S} is the orbital bone label mask of the sagittal image, where the orbital bone area is 1 and the background is 0.

Orbital Bone Quality Improvement in Axial Plane

As shown in Fig. 3, intermediate slices generated using only the spatial information of the sagittal plane have problems such as jagged edges of cortical bone and disconnected edges of thin bone. To solve these problems, the generated intermediate slices are refined using additional axial loss functions to consider the structural information of the orbital bone on the axial plane.

The axial loss function consists of the axial MAE loss, the axial OBE loss, and the feature map difference loss. The axial MAE loss is calculated using Eq. (5). To generate intermediate slices with clear edges of the orbital bone on the axial plane, the axial OBE loss is calculated using Eq. (6). To refine the fine details and edges of the generated

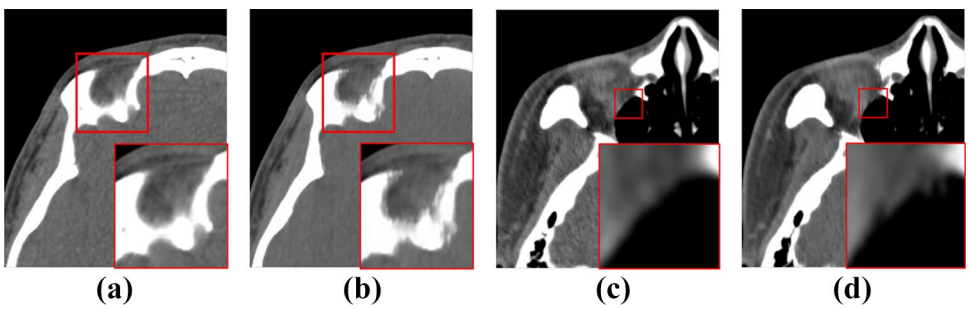

Fig. 3 Examples of blurry edges and disconnections of the orbital bone caused by generating intermediate slices using only spatial information from the sagittal plane: **a** original slice of cortical bone, **b** blurred edges of the cortical bone in the intermediate slice, **c** original slice of the thin bone of the orbital floor, and **d** disconnected edges of the thin bone of the orbital floor

intermediate slices, the feature map difference loss is calculated as the difference from the feature map extracted by a pre-trained VGG16 [16] network using Eq. (7). The VGG16 network is a classification model trained using the ImageNet dataset [17]. The feature maps extracted from the VGG16 network contain high-level image feature representations inferred from the deep CNN; minimizing the differences in the feature maps helps to generate images that are perceptually close to the original image.

$$L_{MAE}^{AXL} = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}\left|SR_{i,j}^{A} - O_{i,j}^{A}\right| \tag{5}$$

$$L_{Edge}^{AXL} = \sum_{i=1}^{W}\sum_{j=1}^{H}\left\{\left|E(SR^{A})_{i,j} - E(O^{A})_{i,j}\right| * M_{i,j}^{A}\right\} \tag{6}$$

$$L_{VGG}^{AXL} = \sum_{i=1}^{W_{x,y}}\sum_{j=1}^{H_{x,y}}\left\{\phi_{x,y}\left(SR^{A}\right)_{i,j} - \phi_{x,y}\left(O^{A}\right)_{i,j}\right\}^{2} \tag{7}$$

Here, SR^A is the generated intermediate slice, O^A is the original axial image, i and j are the pixel positions, and W and H are the width and height of the original axial image, respectively. M^A the orbital bone label mask of the axial image, where the orbital bone area is 1 and the background is 0. $\phi_{x,y}$ indicates the feature map obtained by the x th convolutional layer before the y-th maxpooling layer within the VGG16 network. $W_{x,y}$ and $H_{x,y}$ are width and height of the feature map $\phi_{x,y}$, and x and y are set to 3.

Results

We trained the proposed network with the Adam optimizer by setting $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The learning rate was set to 10^{-3}. The proposed network was implemented in Python 3.6.9 and Tensorflow and trained using four GeForce GTX 1080Ti GPUs.

To validate the performance of the proposed method, the generated images were evaluated qualitatively and quantitatively. For the comparison methods, bilinear interpolation, bicubic interpolation, 3D SRGAN [18], and the 2D CNN-based method proposed by Georgescu et al. [5] were used. For the accurate performance evaluation, the evaluation ROI was divided into the three evaluation ROIs of the whole orbital bone, the thin bone of the orbital medial wall, and the thin bone of the orbital floor (Fig. 4). The evaluation ROI of the whole orbital bone was defined for each patient as the maximum size to include the whole orbital bone label mask, and the evaluation ROIs of the orbital medial wall and the orbital floor are manually defined for each patient to include the orbital medial wall and the orbital floor between the eyeball and nasal cavity. For a quantitative evaluation, the generated images were compared to the original images using three evaluation metrics: the peak signal-to-noise ratio (PSNR) [19], the structural similarity index (SSIM) [20], and the visual information fidelity (VIF) [21]. The PSNR is measured for the intensity difference based on MSE (Eq. (8)) and is calculated as the ratio between the maximum intensity value of the original image and the intensity difference using Eq. (9). The higher the PSNR value is, the closer it is to the original image. The SSIM measured structural differences and captured the similarities between the generated image and the original image by considering the intensity, contrast, and structure using Eq. (10). The SSIM value is between 0 and 1, where a value of 1 means that the original image and the generated image are identical. The VIF was measured for perceptual differences such as noise and blur; to reflect the human perception, it uses human visual system (HVS) based on a Gaussian scale mixture model. The VIF is calculated as Eq. (11) and the value is between 0 and 1, where a value of 1 means that the original image and the generated image are identical.

$$MSE = \frac{1}{PQ}\sum_{i=1}^{P}\sum_{j=1}^{Q}(x_{i,j} - y_{i,j})^{2} \tag{8}$$

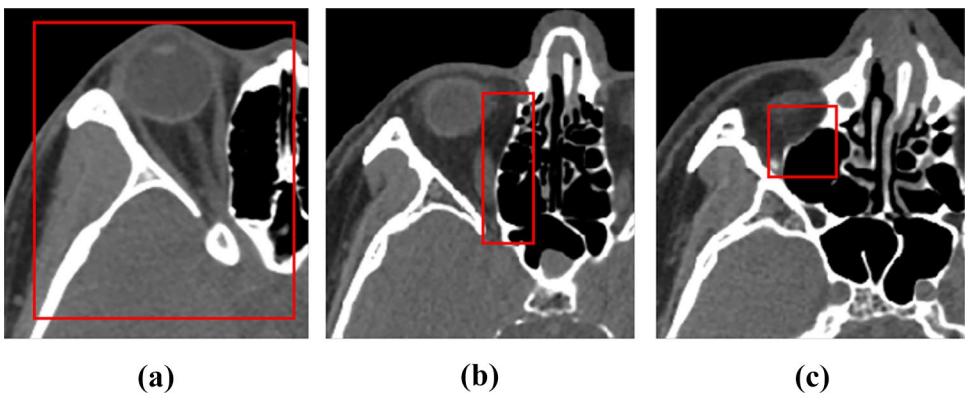

Fig. 4 Three evaluation ROIs for the inter-slice resolution improvement of the orbital bone in facial CT images: **a** evaluation ROI of the whole orbital bone, **b** the orbital medial wall, and **c** the orbital floor

$$PSNR = 10 \cdot log_{10}(\frac{MAX_I^2}{MSE}) \tag{9}$$

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{10}$$

$$VIF = \frac{\sum_{j \in subbands} I(\vec{y}^{N,j}; \vec{X}^{N,j} | s^{N,j})}{\sum_{j \in subbands} I(\vec{y}^{N,j}; \vec{Y}^{N,j} | s^{N,j})} \tag{11}$$

In these equations, x is the image with the improved inter-slice resolution and y is the original image, i and j are the pixel positions, P and Q are the width and height of the orginal image, MAX_I^2 is the square of the maximum intensity value, μ_x and μ_y are the mean pixel values of x and y, σ_x^2 and σ_y^2 are the variance values of x and y, σ_{xy} is the covariance between x and y, and C_1 and C_2 are variables to stabilize the division with a weak denominator. X and Y are the outputs of HVS when the inputs are x and y, considered as the visual signal. $\vec{y}^{N,j}$ is the result of the wavelet decomposition of y; j denotes the indices of the subbands in $\vec{y}^{N,j}$, and N represents the indices of the element in subband j. $I(\vec{y}^{N,j}; \vec{X}^{N,j} | s^{N,j})$ is the mutual information of y and X, which is considered as the information extracted by the brain when the generated image is observed by a human. Likewise, $I(\vec{y}^{N,j}; \vec{Y}^{N,j} | s^{N,j})$ is the mutual information of y and Y, which is considered as the information extracted by the brain when the original image is observed by a human. $s^{N,j}$ is the HVS model parameter.

Figure 5 shows the qualitative evaluation results of sagittal images with improved inter-slice resolutions using the comparative methods and the proposed network. With the use of the bilinear and bicubic interpolation methods, the aliasing effect remained in the sagittal images. For the 3D SRGAN method, the aliasing effect was reduced in the sagittal image, but the edges are blurred compared to those in the original image. With the 2D CNN-based method, the resulting images are similar to the original image and the edges are clearer than in the 3D SRGAN case. When using the proposed network, the resulting images are similar to the original images and show higher contrast on the edges of the thin bone compared to the 2D CNN-based method.

Figure 6 shows the intermediate slice images generated using the comparative methods and the proposed network. The first to third rows represent the cortical bone, the thin bone of the orbital medial wall, and the thin bone of the orbital floor, respectively. When using the bilinear and bicubic interpolation methods, the edges of the orbital bone are unclear and different from those in the original image. With the 3D SRGAN method, the generated intermediate slice images are more structurally similar to the original images than the images generated by the interpolation methods but still show blurry edges of the orbital bone. For the 2D CNN-based method, the edges are clearer than those by the 3D SRGAN method but are still jagged for the cortical bone and show a disconnection of the edge of the thin bone. In the case of the proposed network, the generated intermediate slice images have clear edges for both the cortical and thin bone, similar to the original images, without any disconnections at the edges of the orbital bone.

Table 1 shows the results of the quantitative evaluation of images with the improved inter-slice resolutions generated by the proposed network and by the comparison methods. In the whole orbital bone, the PSNR, SSIM, and VIF of the proposed network showed the best performance, which were estimated as 33.56, 0.9883, and 0.7428, respectively. Compared to 3D SRGAN, the proposed method shows corresponding improvements of 24.7%, 2.9%, and 39% for the PSNR, SSIM, and VIF. The percentage increase is especially high in the VIF case, as the 3D SRGAN method generated blurry images. Compared to the 2D CNN-based method, the proposed method showed corresponding improvements

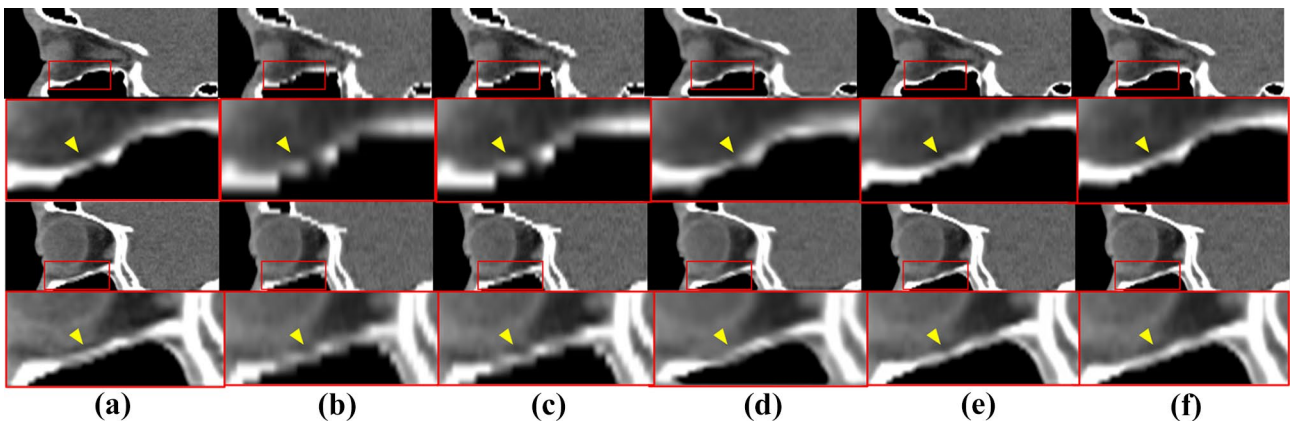

Fig. 5 Qualitative evaluation results of the inter-slice resolution improvement on the sagittal planes: **a** original image, **b** bilinear interpolation, **c** bicubic interpolation, **d** 3D SRGAN [16], **e** 2D CNN [5], and **f** the proposed network

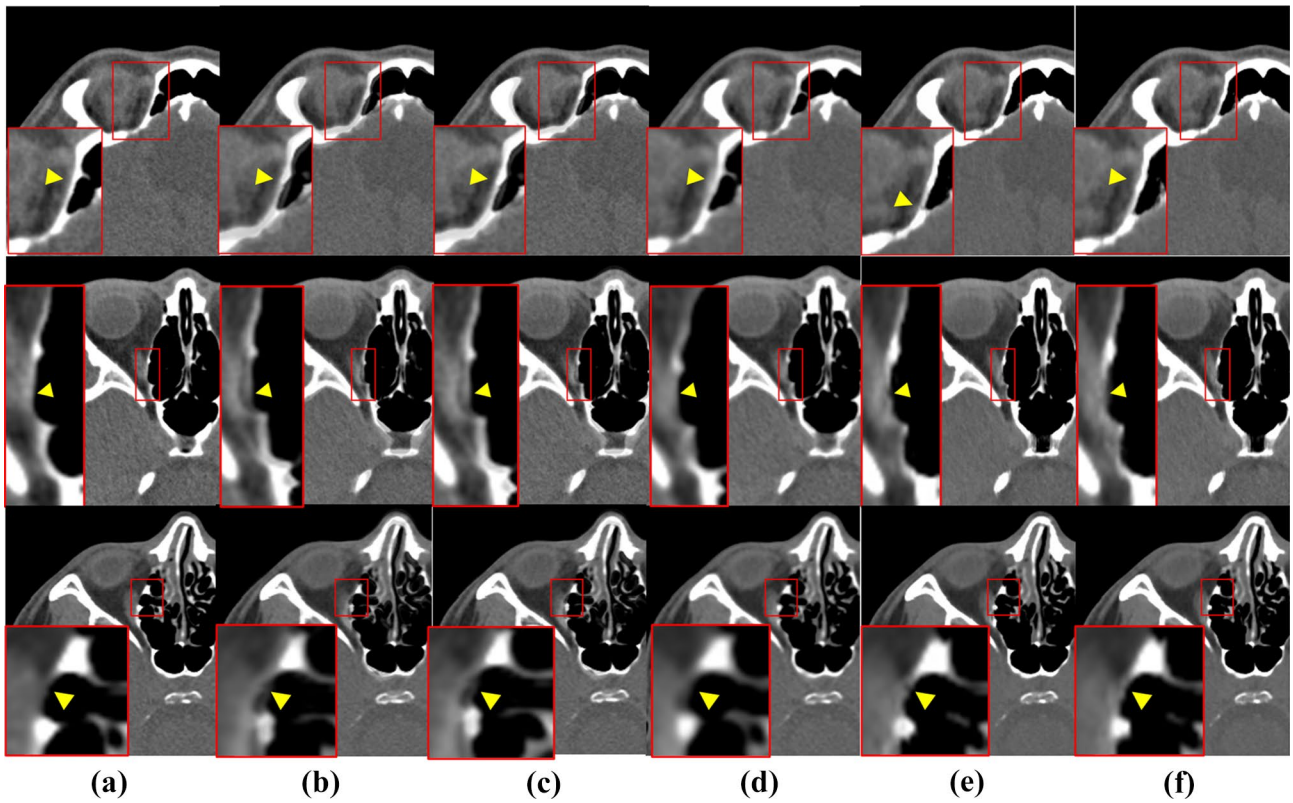

Fig. 6 Qualitative evaluation results of inter-slice resolution improvement on the axial planes: **a** original image, **b** bilinear interpolation, **c** bicubic interpolation, **d** 3D SRGAN [16], **e** 2D CNN [5], and **f** the proposed network

of 1.5%, 0.1%, and 1.2% for the PSNR, SSIM, and VIF. In the thin bone of the orbital medial wall, the PSNR, SSIM, and VIF outcomes by the proposed network showed the best performance. Compared to the 3D SRGAN method, the proposed method was better by 19.8%, 2.8%, and 23.6% in terms of the PSNR, SSIM, and VIF. Compared to the 2D CNN-based method, the proposed method was better by 5.3%, 0.1%, and 0.9% for the PSNR, SSIM, and VIF, exceeding the percentage increase of PSNR for the whole orbital bone. In the thin bone of the orbital floor, the proposed method showed the best performance, and the PSNR, SSIM, and VIF outcomes were higher than those of the orbital medial wall. Compared to the 3D SRGAN method, the proposed network showed improvements of 23.8%, 2.8%, and 37.9% for the PSNR, SSIM, and VIF. Compared to the 2D CNN-based method, the proposed network was better by 1.7% and 1.2% for the PSNR and VIF, exceeding the percentage increase of VIF in the orbital medial wall. In the comparison between the 2D CNN-based method and the proposed network, the percentage increases for PSNR in thin bone were higher than those in the whole orbital bone. These outcomes demonstrate that the proposed network can effectively generate intermediate slice images with clear edges of thin bone areas as well as the cortical bone area.

Discussion

In this paper, we proposed a method to improve the inter-slice resolution of facial CT images by considering the structural information of the orbital bone on the sagittal and axial planes. The contribution of the proposed method can be summarized in two ways. First, the proposed method generates an intermediate slice using spatial information not only on the sagittal plane but also on the axial plane. Typical 2D CNN-based methods use spatial information from one plane, such as the sagittal or coronal plane, to infer the intermediate slice; accordingly, the axial plane of the intermediate slice has jagged or blurry edges. To rectify this, the proposed method uses an additional axial loss function to minimize the image difference between the generated intermediate slice and the original axial image. Second, the proposed method uses a combination of the OBE loss and the feature map difference loss to generate intermediate slices with orbital bones that are perceptually similar to the original image. Using the typical gradient difference loss, the edges of the thin bones may not be significantly considered because they have a small area and low intensity values similar to the surrounding tissues. The OBE loss can greatly account for the edges of thin bones given its use of an orbital

Table 1 Quantitative evaluation results of the images with improved inter-slice resolution

Method	Whole orbital bone			Thin bone of the orbital medial wall			Thin bone of the orbital floor		
	PSNR	SSIM	VIF	PSNR	SSIM	VIF	PSNR	SSIM	VIF
Bilinear interpolation	24.81 ± 3.59	0.9486 ± 0.0233	0.4898 ± 0.0901	27.60 ± 7.18	0.9473 ± 0.0508	0.5282 ± 0.2351	26.33 ± 5.08	0.9541 ± 0.0389	0.4847 ± 0.1522
Bicubic interpolation	24.85 ± 3.63	0.9503 ± 0.0228	0.4912 ± 0.0903	27.63 ± 7.40	0.9479 ± 0.0494	0.5295 ± 0.2432	26.49 ± 5.20	0.9549 ± 0.0382	0.4874 ± 0.1512
3D SRGAN [16]	26.91 ± 2.66	0.9604 ± 0.0181	0.5341 ± 0.0673	27.93 ± 6.81	0.9529 ± 0.0451	0.5502 ± 0.2597	27.62 ± 4.67	0.9618 ± 0.0328	0.5159 ± 0.1430
2D CNN [5]	33.08 ± 5.18	0.9876 ± 0.0081	0.7341 ± 0.0963	31.79 ± 7.50	0.9793 ± 0.0212	0.6740 ± 0.1666	33.63 ± 6.39	0.9887 ± 0.0127	0.7033 ± 0.1193
Proposed network	**33.56 ± 5.12**	**0.9883 ± 0.0078**	**0.7428 ± 0.0946**	**33.46 ± 10.33**	**0.9798 ± 0.0205**	**0.6803 ± 0.1662**	**34.20 ± 6.79**	**0.9891 ± 0.0125**	**0.7116 ± 0.1195**

The highest values were denoted in bold

bone label mask to minimize the gradient difference only in the orbital bone area. The feature map difference loss is additionally used on the axial plane to improve the perceptual quality of the generated intermediate slices. The feature map difference loss function provides structural information of the images, which is difficult to consider when using pixel-wise losses such as the MAE or MSE only as the loss functions. By minimizing the difference in the feature map containing the feature representation of the orbital bone in the CT images, it helps to consider the structural information of the thin bone on the axial plane and to generate an intermediate slice with an accurate orbital thin bone structure that is perceptually similar to the original image.

In the experiment conducted here, we evaluated facial CT images generated from down-sampled facial CT images using PSNR, SSIM, and VIF as evaluation metrics in three evaluation ROIs, in this case, the whole orbital bone, the thin bone of the orbital medial wall, and the thin bone of the orbital floor. As shown in Table 1, the proposed method achieved the highest score in terms of the PSNR, SSIM, and VIF and showed the best result for the three evaluation ROIs. Compared with bilinear and bicubic interpolation and the 3D SRGAN method, the proposed method showed improvement of 51.7% and 51.2% and 39.1%, respectively, for the VIF of the whole orbital bone, an outcome significantly higher than the percentage increases for the PSNR and SSIM metrics. This confirms that the proposed method generates CT images with an accurate orbital bone structure and clear edges of the orbital bones, as VIF captures perceptual differences such as noise and blur. Compared to the 2D CNN-based method, the proposed method showed improvements of 5.3% and 1.7% in the thin bone of the orbital medial wall and the thin bone of the orbital floor, respectively, in terms of the PSNR. These percentage increases of the PSNR were higher than those of the SSIM and VIF due to the very small area of the thin bone, and the numerical difference is greater in terms of the pixel-wise difference than the structural or perceptual difference. These results confirm that the proposed method can effectively generate intermediate slices with accurate structures and clear edges of thin bones as well as cortical bones.

Although the proposed method achieved the best score in the quantitative assessment, the numerical difference between the PSNR, SSIM, and VIF metrics may seem insignificant compared to the 2D CNN-based method. Figure 7 shows the results of a qualitative evaluation of intermediate slices with enlarged thin bone areas. This figure shows that the proposed method generated intermediate slices with smooth thin bone edges, whereas the 2D CNN-based method generated intermediate slices with jagged or disconnected edges of thin bones. This confirms that the proposed method is effective when used to generate intermediate slices with clear thin bone edges as it accounts for the structure of the orbital bone on the axial plane.

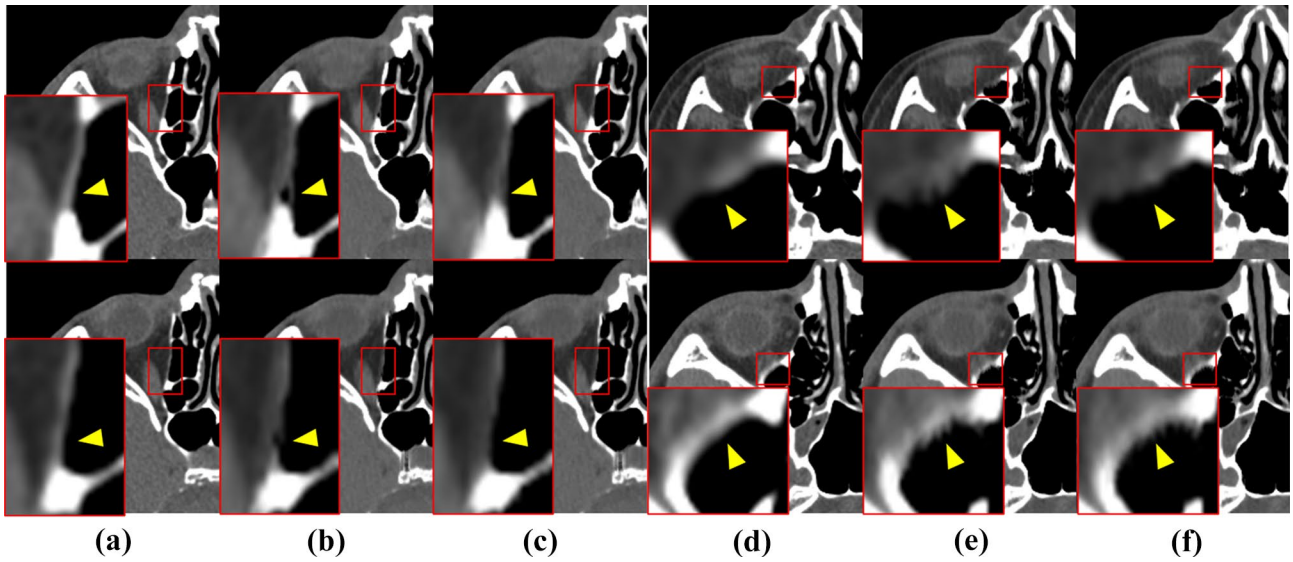

Fig. 7 Qualitative evaluation results of intermediate slice generation in the thin bones of the orbital medial wall and orbital floor: **a**–**c** represent the thin bone of the orbital medial wall and **d**–**f** represent the thin bone of the orbital floor. **a**, **d** original image, **b**, **e** 2D CNN [5], and **c**, **f** proposed network

Conclusion

In this paper, we proposed a method to improve the inter-slice resolution of facial CT images by considering the structural information of the orbital bone on the sagittal and axial planes. To consider the spatial information of the sagittal and axial planes, the proposed method generates intermediate slices on the sagittal plane and the generated intermediate slices are then transformed to the axial plane to compare them with the original axial images. To generate CT images with an accurate orbital bone structure, the OBE loss using an orbital bone label mask is used on the sagittal and axial planes to minimize the gradient difference only in the orbital bone area. To improve the perceptual quality of intermediate slices on the axial plane, the feature map difference loss is additionally used. The experimental results indicate that the proposed method outperformed in quantitative and qualitative evaluations among several compared methods. It is confirmed that the proposed method can reconstruct an accurate 3D model for orbital surgery caused by acute orbital trauma or orbital fractures by solving the problems of the aliasing effect and thin bone disconnection in facial CT images with thick slice thickness. Our method can therefore be utilized to provide an accurate customized bone plate for surgical reconstructions of fractured orbital thin bones.

Funding This research was partly supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (No. 2017R1D1A1B03034927) and was partly supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Health & Welfare) (Project Number: 9991007552) and was partly supported by a grant (21174MFDS224) from Ministry of Food and Drug Safety in 2021.

Declarations

Conflict of Interest The authors declare no competing interests.

References

1. Burnstine MA: Clinical recommendations for repair of orbital facial fractures. Current opinion in ophthalmology 14(5): 236-240, 2003
2. Kirkland EJ: Bilinear Interpolation. Advanced Computing in Electron Microscopy. https://doi.org/10.1007/978-1-4419-6533-2_12, 2010
3. Keys R: Cubic convolution interpolation for digital image processing. IEEE transactions on acoustics, speech, and signal processing 29(6): 1153-1160, 1981
4. Bae W, Lee S, Park G, Park H., Jung KH: Residual CNN-based image super-resolution for CT slice thickness reduction using paired CT scans: preliminary validation study. Medical Imaging with Deep Learning, 2018
5. Georgescu, MI., Ionescu RT, Verga N: Convolutional neural networks with intermediate loss for 3D super-resolution of CT and MRI scans. IEEE Access 8: 49112-49124, 2020
6. Jurek J, Kociński M, Materka A, Elgalal M, Majos A: CNN-based superresolution reconstruction of 3D MR images using thick-slice scans. Biocybernetics and Biomedical Engineering 40(1): 111-125, 2020
7. Yun, H. R., Lee, M. J., Hong, H., Shim, K. W., & Jeon, J: Improvement of inter-slice resolution based on 2D CNN with thin bone structure-aware on head-and-neck CT images. International Society for Optics and Photonics, 2021.

8. Li Z, Wang Y, Yu J: Reconstruction of thin-slice medical images using generative adversarial network. International Workshop on Machine Learning in Medical Imaging, 2017
9. Delannoy Q, Pham CH, Cazorla C, Tor-Díez C, Dollé G, Meunier H, Rousseau F, et al: SegSRGAN: Super-resolution and segmentation using generative adversarial networks—Application to neonatal brain MRI. Computers in Biology and Medicine 120: 103755, 2020
10. van der Ouderaa TF, Worrall DE, van Ginneken B: Chest CT Super-resolution and Domain-adaptation using Memory-efficient 3D Reversible GANs. arXiv preprint arXiv:1908.00295. 2019
11. Kudo A, Kitamura Y, Li Y, Iizuka S, Simo-Serra E: Virtual thin slice: 3D conditional GAN-based super-resolution for CT slice interval. International Workshop on Machine Learning for Medical Image Reconstruction pp.91–100, 2019
12. Ledig C, Theis L, Husz´ar F., Caballero J, Cunningham A, Acosta A, Aitken A P, Tejani A, Totz J, Wang Z, et al: Photo-realistic single image super-resolution using a generative adversarial network. CVPR vol 2: p 4, 2017
13. Liu J, Chen F, Wang X, Liao H: An edge enhanced srgan for mri super resolution in slice-selection direction. Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy pp. 12–20, 2019
14. Chai Y, Xu B, Zhang K, Lepore N, Wood JC: Mri restoration using edge-guided adversarial learning. IEEE Access 8: 83858-83870, 2020
15. Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R, Wang, Z, et al: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. Proceedings of the IEEE conference on computer vision and pattern recognition pp. 1874–1883, 2016
16. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014
17. Deng J, Dong W, Socher R, Li L J, Li K, Fei-Fei L: Imagenet: a large-scale hierarchical image database. CVPR09, 2009
18. Sánchez I, Vilaplana V: Brain MRI super-resolution using 3D generative adversarial networks. arXiv preprint arXiv:1812.11440, 2018
19. Huynh-Thu Q, Ghanbari M: Scope of validity of PSNR in image/video quality assessment. Electronics letters 44(13): 800-801, 2008
20. Wang Z., Bovik AC, Sheikh HR, Simoncelli EP: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13(4): 600-612, 2004
21. Sheikh HR, Bovik AC: Image information and visual quality. IEEE Transactions on image processing 15(2): 430-444, 2006