From Undersampling K-space to Individual Anatomy Modeling: Comprehensive Head Digital Twins Construction via Super-Resolution Reconstruction and Segmentation

Wei Jia[a]

a. School of Biomedical Engineering and State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, China

A B S T R A C T

Objective: This work aims to bridge the gap between theoretical knowledge of magnetic resonance imaging (MRI) and its practical application in accelerated imaging and personalized healthcare. Specifically, we replicate the MRI accelerated scanning and reconstruction workflow learned in class, and innovatively extend this to the construction of high-resolution head digital twins (HDTs) to address the current lack of focus on the head region in digital twin development and providing a solution for individually high-quality anatomical modeling.

Methods: We begin by exploring the impact of various undersampling patterns on both the acceleration and iFFT reconstruction quality of MRI images, gaining practical insights into the trade-offs between scan time and image fidelity. We then leverage these insights to develop an end-to-end deep learning framework for super-resolution MRI reconstruction, which directly inpaints undersampled k-space data into high-resolution images, mitigating aliasing artifacts and enhancing clarity by utilizing dual image and kspace domain knowledge. Subsequently, using these reconstructed super-resolution images, we train a universal segmentation model of key craniofacial and central nervous system to construct high-resolution head digital twins (HDTs).

Results: Undersampling experiments revealed the trade-offs between scan time and image quality. The CNN-based reconstruction achieved a PSNR of 31.82 and an SSIM of 0.84. The HDTs demonstrated high segmentation accuracy, achieving a Dice score of 0.91.

Conclusions: This end-to-end workflow, from undersampled scanning to high-resolution image display, not only solidify our understanding of MRI principles and their practical implications, but also contribute novel solutions to the challenges of accelerated MRI, CNN-based super resolution reconstruction, ultimately culminating in the successful construction of high-resolution head digital twins.

Clinical significance: This work contributes to the advancement of HDT technology, making it more accessible and reliable for clinical practice. The framework supports precise anatomical modeling and surgical planning in digital dentistry, and opens new avenues for exploring brain-craniofacial relationships and their implications for neuromuscular disorders, thus offering a significant step towards personalized healthcare.

Keywords *: MRI acceleration, K-space undersampling pattern, CNN-based super-resolution reconstruction, Head digital twins construction, Personalized healthcare, Hands-on experience of MRI workflow*

1.Introduction

1.1 Accelerated MRI: Challenges and Opportunities

1.1.1 The Challenge of Slow MRI Acquisition

Magnetic resonance imaging (MRI) is a dominating clinical imaging modality, renowned for its ability to provide high-resolution structural and functional information without the use of ionizing radiation. However, MRI inherently relies on the sequential sampling of k-space data, an inherent slow process constrained by hardware limitations and physiological factors, such as the limited switching speed of gradient systems, specific absorption rate (SAR) thresholds and peripheral nerve stimulation (PNS). These intrinsic constraints significantly hinder efforts to accelerate data acquisition. Consequently, acquiring high-resolution MRI images often requires prolonged imaging times.

However, prolonged scan times remain a critical bottleneck in MRI, leading to increased susceptibility to motion artifacts, reduced efficiency in data collection, and difficulties in imaging uncooperative patients [1-2] , such as elderly individuals with cardiopulmonary conditions or neonates. As a result, the extended acquisition duration raises costs and limits the broader adoption of MRI in clinical practice.

1.1.2 K-space Undersampling for Accelerated Imaging

Past decades have witnessed many approaches to address the problem and accelerate the MRI process such as parallel imaging. [3] A more popular choice is to sample only a subset of the k- space points and use reconstruction models to fill the rest or to directly get a fully sampled image. Such methods reduce the time needed to repeat the scanning and are more flexible since the acceleration rate can be adjusted according to the sampling ratio, balancing the factor of image quality and speed. The relationship between acceleration ratio (R) and the sampled point is demonstrated by

$$R = \frac{Number\ of\ k-space\ points\ in\ fully\ sampled\ scan}{Number\ of\ k-space\ points\ in\ accelerated\ scan}$$

While undersampling accelerates data acquisition, it violates the Nyquist-Shannon sampling theorem, leading to the visual manifestation of aliasing artifacts in the reconstructed images. In essence, the Nyquist-Shannon sampling theorem dictates that, to accurately reconstruct a signal, the sampling rate must be at least twice the highest frequency component of the signal (the Nyquist rate). [4] When k-space is undersampled, aliasing arises from the superposition of spectral components, creating spurious patterns that manifest as streaking or distortions in the reconstructed image. [5] In MRI, such artifacts degrade image fidelity and can obscure diagnostically relevant details. Advanced reconstruction techniques aim to mitigate these issues by leveraging prior information or imposing constraints on the reconstruction process.

1.2 Advances in CNN Reconstruction for Undersampled MRI

The aim of reconstruction is to fill the missing k-space data caused by undersampling, guided by the sampling pattern and prior knowledge of the k-space distribution, while ensuring consistency with the observed data. [6] Traditional methods, including Compressed Sensing (CS) and dictionary learning-based image reconstruction, have laid the foundation for accelerated MRI imaging.

Compressed Sensing (CS) is a signal reconstruction technique based on the theory of signal sparsity, aimed at reconstructing the original signal using a small number of random samples. The core idea is that, although a signal may not be sparse in its original domain, it can often exhibit sparsity in certain transform domains (such as wavelet or Fourier transforms). [7] Therefore, by performing downsampling on the signal and applying nonlinear optimization techniques, CS allows for complete signal reconstruction at a rate lower

than the Nyquist sampling rate. This method requires the signal to have a sparse representation in some transform domain, and the sampling scheme must be incoherent with the sparse representation's transform domain. However, in practical applications, many signals are only approximately sparse, and this assumption can lead to a loss in reconstruction accuracy. Moreover, the choice of regularization parameters in compressed sensing has a significant impact on the quality of the reconstruction. [8]

Another dominating traditional reconstruction method is dictionary learning-based image reconstruction, which restores images by learning their sparse representations. The core idea is that an image can be effectively represented by a sparse linear combination of atoms in a specific dictionary. The dictionary, consisting of basis vectors, captures the local structure of images. [9] By learning the dictionary from a large set of images and using sparse coding, the original image can be recovered from noisy or low-quality data. Unlike traditional methods, dictionary learning generates an adaptive dictionary based on data rather than relying on fixed transforms. This method is widely used in image denoising, super-resolution, and compressed sensing, improving reconstruction accuracy. However, it faces challenges such as long training times and high computational complexity. [10]

To address the limitations of traditional methods, recent advancements in deep learning have revolutionized MRI reconstruction. Convolutional neural networks (CNNs) have emerged as powerful tools capable of capturing complex image features and patterns in the data. [11] These models excel in capturing complex image features and have demonstrated superior performance in reconstructing high-quality images from undersampled k-space data. Moreover, CNNs allow for end-to-end solutions that integrate domain-specific knowledge, such as k-space consistency, to further enhance reconstruction fidelity.

In our project, we reproduced a classical MRI reconstruction method based on CNN deep learning. [12] Our model employs a cascade paradigm, providing an end-to-end solution that reconstructs fully sampled images from undersampled k-space. Additionally, we leveraged the characteristics of MRI images by incorporating a k-space consistency layer to enhance the quality of the reconstruction in the image domain. Compared to traditional machine learning and dictionary learning-based reconstruction methods, our model achieves the exceptional performance in terms of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) on test set.

1.3 Constructing Head Digital Twins from Reconstructed Super-Resolution Images

Digital Twins have emerged as a transformative technology in healthcare, offering unprecedented opportunities for personalized treatment optimization. [13] However, existing attempts at constructing digital twins often lack focus on the head region (Wasserthal, J., et al., 2023), which overlooks the central to our thought and action. [14] What's more, high-resolution head digital twins (HDTs) are particularly crucial for digital dentistry, which enable clinicians to simulate and optimize treatments before actual intervention. This can potentially reduce complications and improve recovery outcomes.

However, the potential of HDTs is currently hampered by three major challenges: 1) the quality of medical imaging data, 2) the complexity of anatomical labeling process and 3) the robust model foundation for universal segmentation. To address these challenges, we present a comprehensive solution combining CNN-based reconstruction method to provide super-resolution image, an innovative three-stage labeling engine and a robust universal segmentation model. The labeling engine progressively advances from expert manual annotation, through semi-automatic refinement, to fully automatic segmentation, ensuring both accuracy and efficiency in dataset creation. As for universal segmentation model, nn-UNet is a robust framework that adaptively performs preprocessing and data augmentation based on the statistical spacing and intensity distributions of the training data, while also incorporating MRI modality priors. [15] Furthermore, it automatically determines hyperparameters such as learning rate and batch size based on the

data's fingerprint, making it a benchmark in the Medical Segmentation Decathlon. [16]

1.4 Aim and scope

This work aims to bridge the gap between theoretical knowledge of MRI learned in class and its practical application in accelerated imaging and personalized healthcare. We embark on this endeavor by first exploring the effects of various undersampling patterns on MRI image acquisition and reconstruction, allowing us to directly apply and validate the principles of k-space sampling learned in the classroom. This exploration provides a hands-on understanding of the trade-offs between scan time and image quality, a core concept in MRI. Building upon this practical foundation, we then develop an end-to-end deep learning framework for super-resolution MRI reconstruction. This framework, which utilizes image and k-space domain knowledge, demonstrates our ability to translate theoretical understanding into a tangible solution for mitigating aliasing artifacts and enhancing image clarity. Furthermore, we demonstrate the construction of high-resolution head digital twins (HDTs) using these reconstructed super-resolution images. This process directly addresses the current lack of focus on the head region in digital twin development, showcasing our innovative approach to creating high-quality, individualized anatomical models. This end-to-end workflow, from undersampled scanning to high-definition image display, not only solidify our understanding of MRI principles and their practical implications, but also contribute novel solutions to the challenges of accelerated MRI, CNN-based super resolution reconstruction, ultimately culminating in the successful construction of high-resolution head digital twins.

2.Method

2.1Down-sample Acceleration Experiment

2.1.1 Image quality evaluation metric: PSNR & SSIM

Commonly employed metrics for evaluating image quality include the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM). These metrics are described in detail below.

The Peak Signal-to-Noise Ratio (PSNR) is a widely adopted metric for assessing the quality of reconstructed images. It is calculated using the following formula:

$$PSNR = 10 \cdot log_{10}\left(\frac{R^2}{MSE}\right)$$

where R represents the maximum possible pixel value of the image (e.g., $R = 255$ for 8-bit images), and MSE denotes the Mean Squared Error between the reconstructed and original images. PSNR values are expressed in decibels (dB), with higher values generally indicating superior image quality.

The Structural Similarity Index (SSIM) is introduced as a more perceptually relevant metric. SSIM is computed as:

$$SSIM = \frac{(2\mu x\mu y + C1)(2\sigma xy + C2)}{(\mu x^2 + \mu y^2 + C1)(\sigma x^2 + \sigma y^2 + C_2)}$$

where x and y represent two corresponding image patches; μ_x and μ_y are the mean pixel values of patches x and y, respectively; σ_{x^2} and σ_{y^2} are the variances of the patches; σ_{xy} is the covariance between the patches; and C_1 and C_2 are small constants introduced to stabilize the division when dealing with weak denominators. SSIM focuses on the structural and perceptual information of the image, including the

preservation of edges and textures, which makes it more aligned with human visual perception. SSIM is a perceptual metric that quantifies the similarity between two images in terms of brightness, contrast, and structure, with values ranging from 0 to 1. As SSIM compares local regions of two images, it requires a ground truth as a reference.

PSNR and SSIM exhibit distinct strengths and weaknesses. PSNR, while computationally simple, provides a general measure of image quality based on pixel-level differences. However, its reliance on pixel-level differences often leads to a poor correlation with human perception. SSIM, on the other hand, is a perceptual metric that measures the similarity between two images in terms of brightness, contrast, and structure, with values ranging from 0 to 1. It requires a ground truth reference due to its comparison of local image regions. Its strong correlation with human perception makes it a better indicator of image quality in our view. Given these differences, image quality should be assessed by analyzing both metrics in conjunction to obtain a more comprehensive understanding of image fidelity.

2.1.2 Downsample patterns exploration : Six different types of under-sample masks

In our undersampling experiments, we referred to existing literature and explored six different k-space sampling patterns to investigate the impact of varying undersampling rates on reconstructed image quality. [17] These sampling patterns were designed to simulate different data acquisition strategies and to examine their effects on image reconstruction performance. Fig. 1. illustrates the different downsample masks.

- **Random Line:** Random horizontal scan lines were selected for sampling. This pattern simulates the random loss of scan lines during data acquisition.
- **Compressed Sensing:** Compressive sensing was employed, sampling the k-space with random noise points. This pattern aims to explore the potential of compressive sensing methods for accelerated MRI.
- **Cartesian:** A Cartesian sampling pattern was used, with a normal distribution of sampling density centered on the horizontal midline. This pattern retains more information in the central region, primarily low-frequency information, simulating the importance of the central region in traditional MRI scans.
- **Upper Half Sampling:** Only the upper half of the k-space was sampled. This pattern simulates scenarios where only partial k-space data is acquired in certain specific scanning situations.
- **Radial:** A radial sampling pattern was used, with points distributed radially outward from the center of k-space. This pattern provides good coverage of k-space and has some resistance to motion artifacts.
- **Spiral:** A spiral sampling pattern was used, with points distributed spirally outward from the center of k-space. This pattern effectively fills k-space and offers high sampling efficiency.

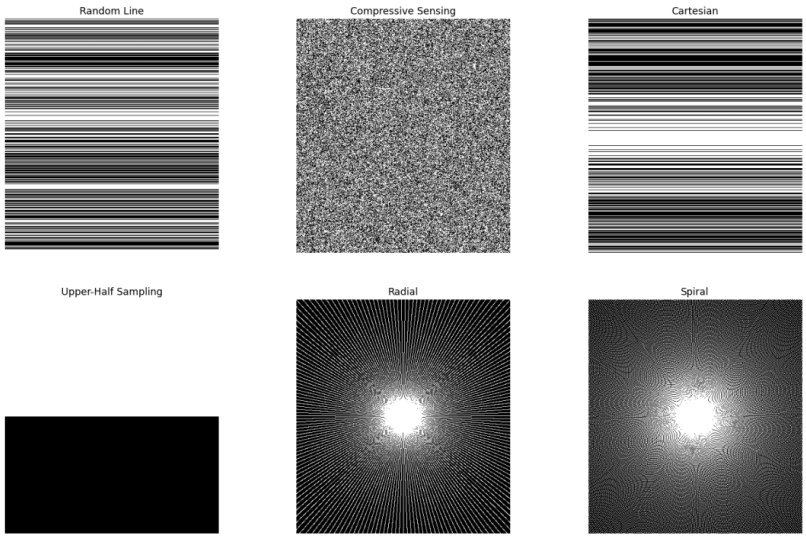

Fig. 1. Six different downsample masks

2.2 CNN-based MRI super resolution reconstruction

2.2.1 Problem formulation

Inpainting is an image processing task aimed at filling in the missing or damaged parts of an image while preserving the harmony and consistency of the observed areas. [18] Specifically, given an image with certain regions damaged or missing, the goal is to use the surrounding known information (such as the image's context, texture, color, etc.) to generate plausible content that completes the missing parts, ensuring that the restored image appears natural and realistic without disrupting the overall structure and style of the original image. The proposed super-resolution MRI image reconstruction algorithm we reproduce leverages the spatial redundancy between the slices in the 3D volumetric model for data observation. So, we define our reconstruction problem as an inpainting problem. Specifically, the mathematical formulation of the inpainting reconstruction problem is demonstrated as follows:

$$\boldsymbol{y} = F_u x + e$$

Here $F_u \in C^{M\times N}$ is an undersampled Fourier encoding matrix and $e \in C^M$ is acquisition noise modelled as additive white Gaussian (AWG) noise. In the case of Cartesian acquisition, we have $F_u = MF$ where $F \in C^{N\times N}$ applies two-dimensional Discrete Fourier Transform (DFT) to each frame in the sequence and $M \in C^{M\times N}$ is an undersampling mask selecting lines in *k*-space to be sampled for each frame.

2.2.2.1 DC Layer

$$\mathbf{s}_{\text{rec}}(j) = \begin{cases} \mathbf{s}_{\text{cnn}}(j) & \text{if } j \notin \Omega \\ \dfrac{\mathbf{s}_{\text{cnn}}(j) + \lambda \mathbf{s}_0(j)}{1+\lambda} & \text{if } j \in \Omega \end{cases}$$

The role of the DC layer is to maintain data consistency by performing a weighted average of the observed values in the k-space(known data) and the values predicted by the CNN, ensuring that the harmony of the data is not disrupted. This process helps preserve the quality and accuracy of the reconstructed image. The formula is shown above. λ in the formula is a hyper-parameter that can be set or trained.

2.2.2.1 Deep Cascade of CNN

Boosting is an ensemble learning method, and its core idea is to combine multiple weak classifiers to form a strong classifier, thereby enhancing the model's predictive performance and reducing both bias and variance. [19] In traditional machine learning, the strong performance of Boosting has been well-established, particularly in scenarios where features need to be manually extracted. Although Convolutional Neural Networks (CNN) already possess strong spatio-temporal predictive capabilities, we still aim to incorporate the Boosting concept into deep learning model training to further improve the model's performance. To achieve this, we introduce a cascade method within CNN, transforming the training process into an end-to-end optimization process. This approach not only optimizes the training procedure but also boosts the overall performance of the model. The whole reconstruction pipeline can be optimised from training, as seen in Fig. 2.

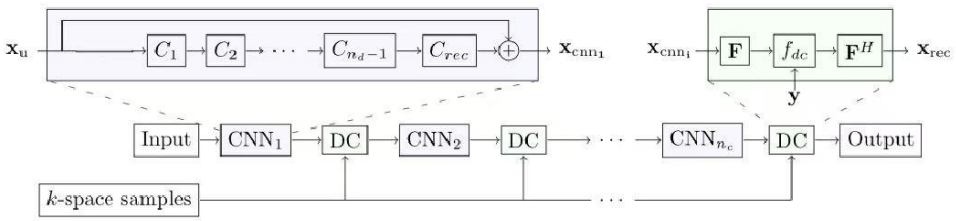

Fig. 2. Data consistency layer and deep cascade CNN for reconstruction

2.2.2 Model training process

We used 42 sets of 3T data for training, dividing the data into 36 training sets, 2 validation sets, and 4 test sets. Based on this data, we performed iterative training on our model. The learning rate for the deep learning network was set to 0.0001, and Mean Squared Error (MSE) was used as the loss function for model training, with 100 epochs (Fig. 3.). The convergence of the model is demonstrated by the MSE convergence observed on both the training and validation sets.

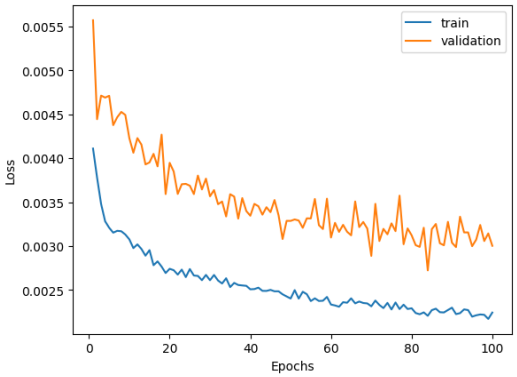

Fig. 3. Training process until convergence

2.3 Head digital twins construction

Ideally, the construction of our digital twin model would have utilized images reconstructed from the previous stage. However, given that this head digital twin model was developed from scratch and the training dataset was limited, we opted to perform annotation and segmentation on high-resolution images acquired using a 5T United Imaging MR system. This approach was chosen to minimize the influence of noise of image during network training. The comprehensive approach for HDT construction focuses on both development of a novel data labelling engine and the training of universal segmentation model.

For data collection, we have selected proton density-weighted (PDw) MR imaging for its superior soft tissue contrast. PDw MR imaging has become the gold standard for maxillofacial diagnostics due to its ability to provide detailed visualization of structures such as the joint disc. [20] We utilize a 5T uMR Jupiter system, which offers higher resolution than the conventional 3T MRI and provides comparable detail of fine structures (e.g., vascular branches) to 7T MRI. [21] Our dataset comprises scans from twenty participants, with their images acquired at 0.3 mm^3 spatial resolution and dimensions of 796 × 928 × 560. This configuration ensures high-resolution imaging and sufficient field of view to capture key anatomical structures spanning the maxillofacial region and brain.

To annotate the data, we have developed a three-stage labelling engine to create a comprehensive, high-quality dataset. Our dataset encompasses 12 key anatomical structures across multiple regions: skeletal structures (skull, mandible, cervical vertebra), joint components (temporomandibular joint disc and cervical articular discs), muscles (temporalis, masseter, medial pterygoid, and lateral pterygoid), and other structures (eye, upper airway and spinal cord). The labelling engine operates in three sequential stages. In the first expert manual stage, annotators use interactive tools (e.g., ITK-SNAP) to establish initial high-quality labels for key anatomical regions. This is followed by a second semi-automatic stage, where a trained nnU-Net model is used to generate preliminary predictions. These predictions are then reviewed and refined by experts, significantly reducing annotation time while maintaining accuracy. In the finally full automatic stage, the refined model generates segmentation masks autonomously, with no human intervention. This systematic approach ensures annotation accuracy and consistency while enabling efficient dataset construction.

To achieve universal segmentation of key craniofacial and central nervous system (CNS) structures, we trained and fine-tuned two distinct models. First, we employed a supervised nn-UNet model, leveraging previously annotated labels, to segment eight craniofacial anatomical structures and the spinal cord belonging to the CNS. Here, we utilized nnUNet as a robust segmentation foundation, training it in a supervised manner with meticulously annotated labels as the gold standard, enabling the prediction of critical craniofacial and spinal cord CNS structures. Second, for detailed segmentation of the brain within the CNS, we fine-tuned a model pre-trained on a large-scale dataset by IDEA Lab, yielding three brain masks. [22-23] By fusing the predictions from these two models, we constructed individual comprehensive head digital twins encompassing the patient's craniofacial region and CNS (spinal cord and brain). The overall workflow is shown in Fig.4 . , comprising data acquisition, labelling engine and model training

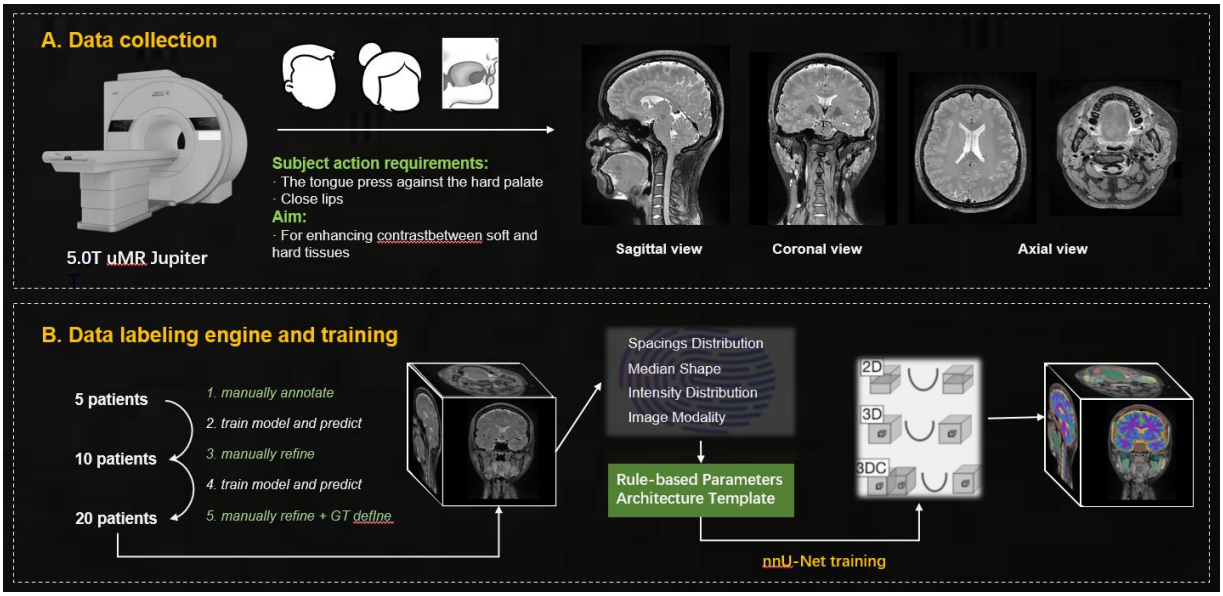

Fig.4. Our proposed full-stack Head digital twins construction framework: (A) we first utilized the 5.0T uMR Jupiter system to acquire high-resolution MR PDw images; (B) Next, we used a data engine to annotate the image iteratively; (C) Finally, the head structures segmentation model is trained;

3.Results

3.1 Down-sample Acceleration Experiment

3.1.1 UI image : Downsample ratio, Scanning time and Image quality

We scanned three images with acceleration factors of 2x, 4x, and 8x. Establishing the 2x acceleration images as the gold standard, we calculated the PSNR and SSIM metrics for United Imaging's 4x and 8x acceleration images. The acceleration ratio , scanning time and image quality are demonstrated in Table 1 and Fig.5.

Table 1 | UI image experiment

Acceleration ratio		**2x**	**4x**	**8x**
Time/min		15	11	7
Image quality	PSNR	GT	27.23	0.9922
	SSIM	GT	26.28	0.9909

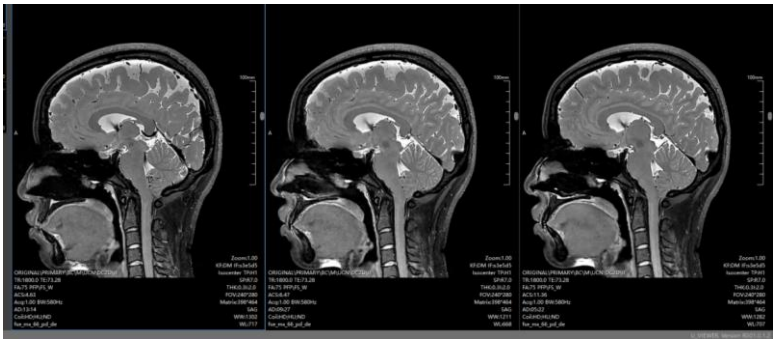

Fig.5. From left to right: GT, 4*acc, 8*acc

3.1.2 Six Down-sample Mask and Image Quality Experiment

As elaborated in Section 2.1.2, we tested six types of undersampling masks to investigate the theoretical impact of different k-space undersampling patterns on image quality, comparing them to United Imaging's undersampled (i.e., accelerated) images using PSNR and SSIM as evaluation metrics. The experimental results are presented in Table 2, and the visualizations of the different undersampling patterns are shown in Fig 6. and Fig 7. It can be observed that, regardless of the mask used, the image quality of the reconstructed images obtained after applying FFT and iFFT transformations is not as prominent as that of United Imaging. This indicates that United Imaging employs an advanced super-resolution image reconstruction algorithm to fill in the missing k-space samples.

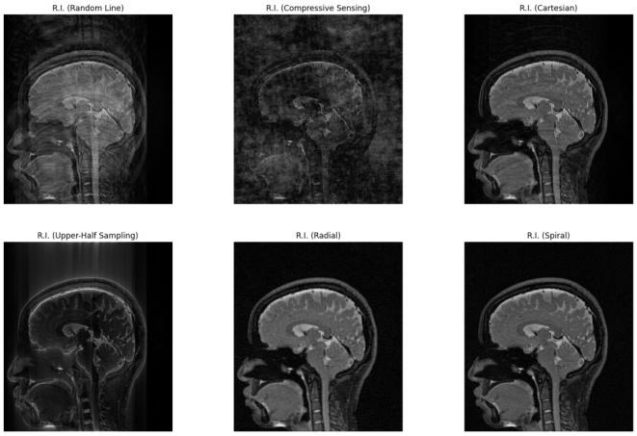

Fig 6. Our experiment of different downsample mask with accelerate rate = 4

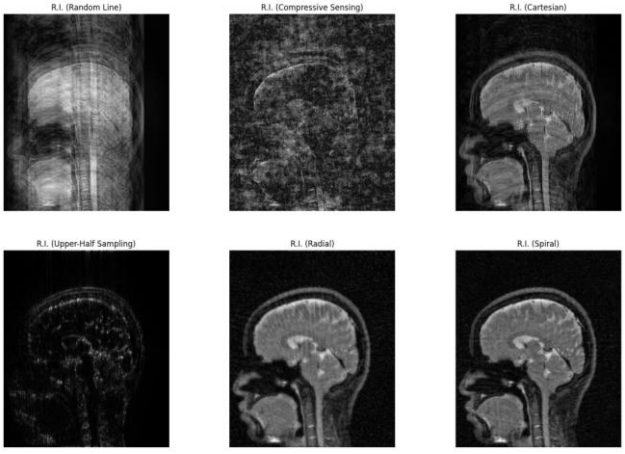

Fig 7. Our experiment of different downsample mask with accelerate rate = 8

Table 2 | Downsample experiment with different masks

3D	PSNR		SSIM	
Acceleration ratio	**4x**	**8x**	**4x**	**8x**
UI	**26.28**	**27.23**	**0.9909**	**0.9922**
random line mask	20.80	23.23	0.3003	0.4984
compressive sensing mask	20.47	22.60	0.1672	0.3277
cartesian mask	29.04	32.73	0.6740	0.7987
upper half sampling mask	18.68	21.35	0.1768	0.4579
radial mask	29.30	31.72	0.5276	0.6609
spiral mask	30.18	33.04	0.5630	0.7038

3.2 The super resolution reconstruction and feasibility of cascade

Convolutional Neural Networks (CNNs) are inherently powerful deep learning models, capable of automatically extracting features from data and performing pattern recognition. However, a single CNN model may not always achieve optimal performance on complex tasks. The cascade method combines multiple layers of models, where each layer improves upon the previous one, gradually reducing errors and enhancing performance. This process is similar to the idea in Boosting, where weak classifiers gradually learn and optimize the misclassified samples.

A cascade CNN can be viewed as a step-by-step refinement process, where each layer of the CNN is trained based on the output of the previous layer. The model progressively focuses on and corrects the errors of the previous layer, optimizing the output at each stage to ultimately improve the overall prediction accuracy. This is akin to the Boosting strategy of "focusing on the errors of the previous model." In Boosting, weak classifiers assign more weight to misclassified samples, while in the cascade structure of CNNs, each layer serves as a "supplement" to the previous layer, gradually reducing errors and enhancing feature extraction capabilities.

Furthermore, the cascade method enables gradual optimization and reinforcement of the feature extraction process at each layer. Each layer does not simply process a fixed input but continuously improves the feature representation from the previous layer. This step-by-step optimization approach allows CNNs to progressively reduce errors, especially when dealing with complex feature data. This iterative refinement helps improve the model's accuracy and robustness. For the effectiveness verification, the detailed ablation study is demonstrated in Fig 6. and Table 3.

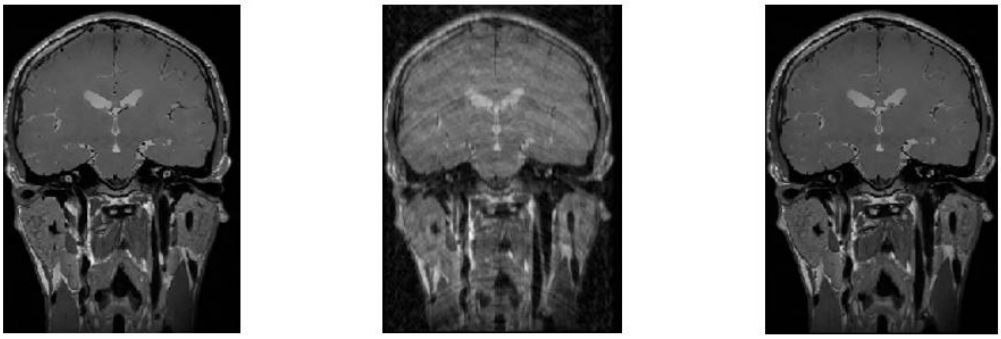

Fig 8. From left to right: GT, downsampled image, and reconstructed image of 5 cascaded CNNs.

Table 3 | Cascade methodology ablation study

# cascades	PSNR(mean)	SSIM(mean)
0	23.88375554	0.7678294
3	25.32230053	0.82147963
5	28.82096367	0.89147963

3.3 HDT accuracy evaluation and visualization

3.3.1 Dice Evaluation metric

The Dice coefficient (Dice Similarity Coefficient, DSC), also known as the Dice similarity index, is a metric used to measure the similarity between two samples, commonly used in image segmentation tasks to evaluate the overlap between a model's predicted segmentation and the ground truth. It essentially measures the similarity between two regions, with a value ranging from 0 to 1, where 1 indicates perfect overlap and 0 indicates no overlap. Therefore, a higher Dice coefficient is better. The calculation formula is:

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|}$$

where |A ∩ B| represents the number of overlapping pixels between the predicted region and the ground truth region, |A| represents the number of pixels in the ground truth region, and |B| represents the number of pixels in the predicted region.

3.3.2 Dice evaluation and visualization

On the test set, the model achieved an average Dice coefficient of 0.91 for all tissues. Based on visual results, not only does the model comprehensively segmented soft tissues, for which MRI has inherent imaging advantages, such as the temporalis muscle, masseter muscle, medial pterygoid muscle, lateral pterygoid muscle, and eyeballs. It also segmented bony structures including the skull, mandible, and articular discs. Furthermore, for the spinal central nervous system, our model segmented the vertebrae, articular discs, and spinal cord. For the brain, we segmented three sets of masks: the first is general tissue, divided into gray matter, white matter, and cerebrospinal fluid; the second is 52 brain regions, which facilitates the construction of individualized brain atlases; and the third is the left and right hemispheres of the brain, which can be used to analyze the function and connections of the left and right brain. The Dice statistics for each segmentation accuracy are presented in Table 4. The visualization results of the HDTs are shown in Fig.9.

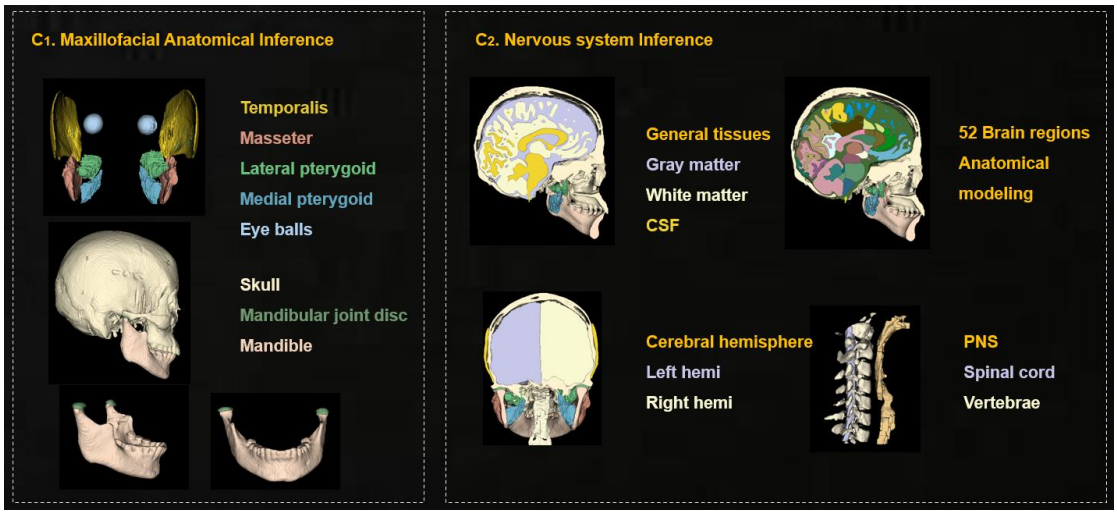

Fig 9. Test set inference output visualization

Table 4 | Test set inference dice statistics

Category		Class	Dice
Maxillofacial	**Soft Tissues**	temporalis muscle	0.8547
		masseter muscle	0.9699
		medial pterygoid muscle	0.9018
		lateral pterygoid muscle	0.9779
		eyeballs	0.8969
	Hard Tissues	skull	0.9698
		mandible	0.9641
		articular discs	0.8314
	Others	upper airway	0.9083
CNS	**Spinal**	vertebrae	0.8788
		articular discs	0.8570
		spinal cord	0.9573
	Brain	General tissues (CSF, GM, WM)	N\A
		Cerebral hemisphere	N\A
		52 Brain regions	N\A

4.Discussion

4.1 Performance analysis of undersampling masks based on k-Space principles

Using the fundamental principles of k-space, we evaluated the impact of different undersampling masks on image quality. This analysis involves applying various masks to k-space data, followed by an inverse fast Fourier transform (iFFT) to reconstruct the image in the spatial domain. The central region of k-space, rich in low-spatial-frequency information, governs the overall image structure and broad features, while the peripheral regions, containing high-spatial-frequency information, are critical for capturing fine details such as edges and textures.

Based on the fundamental principles of k-space, we can analyze the variations in image quality resulting from the application of different undersampling masks in k-space, followed by a direct iFFT to the image domain. The central region of k-space predominantly contains low-spatial-frequency information, which dictates the overall image structure and broad features, while the peripheral regions contain high-spatial-frequency information, crucial for defining fine details such as edges and texture. In accordance with this principle, the Cartesian mask, by retaining low-frequency data around the horizontal midline, plays a vital role in preserving image structure and sharpness, thus exhibiting the highest SSIM of 0.7987 and a favourable PSNR of 32.73. Masks such as Spiral and Radial, which attempt to capture both central and peripheral k-space data points, demonstrate superior performance compared to masks focused solely on the central region or random sampling. Conversely, masks like Random Line and Compressed Sensing, which do not adhere to a structured or uniform pattern, lead to the loss of essential image data, resulting in inferior PSNR and SSIM values. These masks yield highly incomplete or inconsistent data preservation, consequently degrading the overall image quality. A systematic summary of the masks, grounded in k-space principles, is presented in Table 5.

Table 5 | Analysis of undersampling masks based on k-Space principles

Mask Type	K-Space Distribution	Analysis
Spiral Mask	Spiral pattern sampling from center outward	The spiral pattern ensures that both low and high-frequency information is captured in a controlled manner. This leads to relatively better retention of image structure and detail.
Cartesian Mask	Concentrates on low-frequency center	Retaining the central low-frequency part of the k-space (mostly responsible for the basic image structure) helps in preserving the major image features, resulting in higher SSIM and PSNR.
Radial Mask	Radial sampling from center outward	Retaining low-frequency components near the center helps preserve large-scale features, but the pattern is less efficient in capturing higher-frequency components, reducing fine details.
Random Line Mask	Random horizontal lines across k-space	By randomly sampling k-space, this mask may miss crucial high- or low-frequency information, leading to loss of both global and fine details. The random selection introduces inconsistencies in image retention.
Upper Half Sampling Mask	Retains only upper half of k-space	This mask retains only a portion of the low-frequency components, significantly affecting the overall image sharpness and structure, which explains the poor performance in both SSIM and PSNR.
Compressed Sensing Mask	Random noise sampling throughout k-space	The random selection of points in k-space leads to a highly inconsistent sampling, which results in the loss of both low- and high-frequency information, resulting in the lowest SSIM and PSNR values.

4.2 The potential limitation of our reconstruction algorithm and ways for improvements

The reconstruction algorithm we have reproduced presents several potential limitations, primarily related to the size and diversity of the dataset. First, the dataset is relatively small, which may limit the model's learning capability and generalization ability. Although we used 42 sets of 3T data for training, the performance of deep learning models is often directly influenced by the size of the dataset. A smaller dataset may lead to overfitting, which in turn restricts the model's performance in real-world applications. To address this issue, future work could consider incorporating more data samples or utilizing data augmentation techniques to expand the current dataset.

Second, the training dataset contains only one modality of MRI imaging, i.e. PDWI. While this modality provides valuable information, medical imaging in practical applications often involves multiple modalities (such as T1-weighted, T2-weighted, contrast-enhanced, etc.), each of which offers different biological information. The complementarity between these modalities can help enhance the model's predictive performance. Since our training dataset relies solely on a single modality, the model's generalization ability may be limited, particularly when faced with data from other modalities. To improve the model's generalization, future research could consider expanding the dataset to include images from different modalities for joint training.

To address these limitations, several improvements can be proposed. First, transfer learning presents a viable solution to the challenges of limited dataset size and modality diversity. [24] By pretraining the model on large, multimodal datasets, the model can learn generalized features that improve its performance when fine-tuned for specific tasks. This approach has proven effective in medical imaging, particularly in scenarios with limited data availability. Second, data augmentation offers an opportunity to increase dataset diversity

and size by generating varied training samples through techniques such as rotation, translation, flipping, and cropping. [25] Furthermore, advanced methods like Generative Adversarial Networks (GANs) can be employed to synthesize realistic and diverse training samples, including those from rare or hard-to-acquire modalities. By leveraging these strategies, the reconstruction algorithm's performance and generalizability can be substantially enhanced, making it more suitable for clinical and research applications.

4.3 Refining Current Digital Twins: A Focus on Comprehensive Head Modeling

In recent years, the application of universal segmentation for constructing digital twins has gained increasing traction. [26] This approach offers distinct advantages over traditional expert-driven models. Specifically, universal segmentation, through comprehensive annotations, facilitates a holistic understanding of the anatomical landscape, capturing the spatial relationships and dependencies between diverse structures. [27] This enables more accurate contextual segmentation, where the relative positions of structures aid in delineating boundaries and reducing ambiguity. Furthermore, universal segmentation allows for a more systematic learning of tissue-specific representations, such as those of hard tissues, soft tissues, and cartilage, rather than focusing on isolated structures. [28] Expert models, often limited by their narrow annotation scope, typically lack this crucial inter-structural context and tissue-specific understanding, leading to less robust segmentation performance, particularly in cases of anatomical variations or complex pathologies. Consequently, universal segmentation, by leveraging both inter-structural relationships and systematic tissue representation learning, provides more reliable and precise segmentation outcomes, thereby enhancing the fidelity and utility of digital twins.

The construction of digital twins (DTs) has garnered significant attention across various biomedical domains; however, current efforts often focus on specific organs or systems. For instance, Obaid, D. R., et al. developed HeartNavigator, a DT of the heart designed for pre-procedural planning in complex cardiac interventions such as transcatheter aortic valve replacement (TAVR). [29] Similarly, the DT of the lungs gained prominence during the COVID-19 pandemic, with initiatives like Project BreathEasy utilizing lung DTs of COVID-19 patients to optimize ventilator resource allocation and improve patient outcomes. [30] Furthermore, the Living Brain project exemplifies the application of DTs in tracking the progression of neurodegenerative diseases. [31] An overview of existing types of DTs healthcare is provided in Table 6. These projects, while impactful, represent a fragmented approach towards the long-term goal of creating a comprehensive DT of the entire human body. While some researchers, such as Wasserthal et al., [32] have made strides in whole-body segmentation, identifying over 110 tissue classes, their focus on the head region remains limited, often only including the brain. This is a critical oversight, given that the head, encompassing the brain and craniofacial structures, serves as the central hub for cognition, perception, and action. The limited focus on the head region in existing whole-body DT efforts underscores the need for more comprehensive approaches that can accurately model the intricate anatomical details of this crucial area.

To address these limitations and advance the field, we developed a robust segmentation model, trained in a supervised manner using high-resolution images and comprehensive annotations. This model enables the precise segmentation of 12 key maxillofacial anatomical structures, along with the spinal central nervous system, and 52 distinct brain regions, thereby facilitating the construction of individualized, high-fidelity digital twins.

Table 6 | Various types of DTs in healthcare [33-39]

Physical Entity	Entity	Mechanism	Endpoint
Lung	Lexma	Runs simulations of blood and oxygen flow	Predict ventilation requirements
Heart	Dassault, Medtronic, Boston Scientific, FDA;	Simulates the structure and physiologic function of the heart	Customization and optimization of cardiac devices
	Siemens Healthineers	Simulates the structure and physiologic function of the heart	Cardiac resynchronization
	Heart Navigator	simulated TAVR implantations with different aor- tic prosthesis	Surgical planning
Spine	Ahmadian et al.	Predict Vertebral Fracture after Stereotactic Body Radiotherapy	Optimal radiation plan to minimize treatment side effects
Alzheimer's disease	Unlearnai	Predicting the individual outcome in neurological diseases	DT of controls of clinical trial and ultimately clinical interventions
Breast lesions	VICTRE trial	Image based virtual patients comparing digital mammography to tomosynthesis	Determine which imaging tool is better at detecting breast lesions

4.4 Clinical significance of digital twins

The clinical significance of this work is multifaceted. First, our digital twin approach offers personalized precision treatment by leveraging patient-specific anatomical data, which optimizes surgical planning, monitoring, and recovery assessment to improve clinical outcomes. [40] Furthermore, it establishes a framework for exploring the intricate relationships between brain and craniofacial functions, potentially advancing therapies for neuromuscular-related craniofacial deficits through the integration of brain-computer interfaces. [41] This detailed imaging and modeling allows for a deeper understanding of conditions such as craniofacial syndromes, traumatic injuries, and developmental anomalies, while also facilitating the development of targeted therapies for neuromuscular conditions affecting speech, mastication, and facial expressions. The resulting personalized anatomical models enhance the accuracy of diagnoses and interventions, particularly in understanding complex anatomical variations and planning custom treatments. Moreover, advanced visualization tools support surgical planning by providing a clear visualization of the operative site, reducing risks, shortening surgery times, and improving postoperative outcomes in procedures such as craniofacial reconstruction, tumor resections, and neurosurgeries. Finally, these tools enhance doctor-patient communication by allowing physicians to more effectively explain diagnoses and treatment plans, thereby improving patient understanding and trust.

5.Conclusion

This work integrates principles of magnetic resonance imaging (MRI) with advanced deep learning techniques to address challenges in accelerated MRI and personalized healthcare. We begin by exploring the impact of various undersampling patterns on both the acceleration and reconstruction quality of MRI images, gaining practical insights into the trade-offs between scan time and image fidelity. We then leverage these insights to develop an end-to-end deep learning framework for super-resolution MRI reconstruction, which directly transforms undersampled k-space data into high-resolution images, mitigating aliasing artifacts and enhancing clarity by utilizing dual domain knowledge. Subsequently, using these reconstructed super-resolution images, we construct high-resolution head digital twins (HDTs), addressing the current lack of focus on the head region in digital twin development and providing a solution for individually high-quality anatomical modeling. This process not only replicates the workflow of MRI instruments—from undersampled scanning to algorithmic reconstruction and high-definition image display—thereby reinforcing our understanding of MRI principles, but also yields a universal segmentation model for HDT creation. The significance of our work extends beyond these technical contributions, as our framework supports precise anatomical modeling and surgical planning in digital dentistry, and opens new avenues for exploring brain-craniofacial relationships and their implications for neuromuscular disorders. Ultimately, our approach represents a significant step towards making HDT technology more accessible and reliable for clinical practice.

6.Reference

[1] M. Lustig, D. Donoho, J.M. Pauly, Sparse MRI: the application of compressed sensing for rapid MR imaging, Magn. Reson. Med. 58 (6) (2007) 1182–1195.

[2] J. Liu, Y. Pan, M. Li, Z. Chen, L. Tang, C. Lu, J. Wang, Applications of deep learning to MRI images: a survey, Big Data Min. Anal. 1 (1) (2018) 1–18, doi: 10.26599/BDMA.2018.9020001.

[3] X. Zhang, D. Guo, Y. Huang, Y. Chen, L. Wang, F. Huang, Q. Xu, X. Qu, Image reconstruction with low-rankness and self-consistency of k-space data in parallel MRI, Med. Image Anal. 63 (2020) 101687, doi: 10.1016/j.media.2020.101687.

[4] Wang, H. (2023). Compressed sensing: Theory and applications. In Journal of Physics: Conference Series (Vol. 2419, No. 1, p. 012042). IOP Publishing.

[5] Hoff, M., J. Andre, and B. Stewart. "Artifacts in magnetic resonance imaging." Image Principles, Neck, and the Brain (2016): 165-190.

[6] Singh, D., Monga, A., de Moura, H. L., Zhang, X., Zibetti, M. V., & Regatte, R. R. (2023). Emerging trends in fast MRI using deep-learning reconstruction on undersampled k-space data: a systematic review. *Bioengineering*, *10*(9), 1012.

[7] Gilbert, A. C., Indyk, P., Iwen, M., & Schmidt, L. (2014). Recent developments in the sparse Fourier transform: A compressed Fourier transform for big data. *IEEE Signal Processing Magazine*, *31*(5), 91-100.

[8] Mahdaoui, A. E., Ouahabi, A., & Moulay, M. S. (2022). Image denoising using a compressive sensing approach based on regularization constraints. *Sensors*, *22*(6), 2199.

[9] Tošić, I., & Frossard, P. (2010). Dictionary learning for stereo image representation. *IEEE Transactions on Image Processing*, *20*(4), 921-934.

[10] Ravishankar S, Bresler Y. MR image reconstruction from highly undersampled k-space data by dictionary learning. IEEE Trans Med Imaging. 2011 May;30(5):1028-41. doi: 10.1109/TMI.2010.2090538. Epub 2010 Nov 1. PMID: 21047708.

[11] Hossain, M. B., Shinde, R. K., Oh, S., Kwon, K. C., & Kim, N. (2024). A systematic review and identification of the challenges of deep learning techniques for undersampled magnetic resonance image reconstruction. *Sensors*, *24*(3), 753.
[12] Schlemper J, Caballero J, Hajnal J V, et al. A deep cascade of convolutional neural networks for dynamic MR image reconstruction[J]. IEEE transactions on Medical Imaging, 2017, 37(2): 491-503.
[13] Katsoulakis, E., Wang, Q., Wu, H., et al. "Digital twins for health: a scoping review." NPJ Digital Medicine 7.1 (2024): 77.
[14] Wasserthal, J., Breit, H. C., Meyer, M. T., Pradella, M., Hinck, D., Sauter, A. W., ... & Segeroth, M. (2023). TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. Radiology: Artificial Intelligence, 5(5).
[15] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, *18*(2), 203-211.
[16] Isensee, F., Jäger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2019). Automated design of deep learning methods for biomedical image segmentation. *arxiv preprint arxiv:1904.08128*.
[17] Dong, Q., Liu, Y., **ao, J., & Pang, Y. (2024). EEUR-Net: End-to-End Optimization of Under-Sampling and Reconstruction Network for 3D Magnetic Resonance Imaging. *Electronics*, *13*(2), 277.
[18] Jam, J., Kendrick, C., Walker, K., Drouard, V., Hsu, J. G. S., & Yap, M. H. (2021). A comprehensive review of past and present image inpainting methods. *Computer vision and image understanding*, *203*, 103147.
[19] Ferreira, A. J., & Figueiredo, M. A. (2012). Boosting algorithms: A review of methods, theory, and applications. *Ensemble machine learning: Methods and applications*, 35-85.
[20] Derwich M, Mitus-Kenig M, Pawlowska E. Interdisciplinary approach to the temporomandibular joint osteoarthritis—review of the literature[J]. Medicina, 2020, 56(5): 225.
[21] Shi Z, Zhao X, Zhu S, et al. Time-of-flight intracranial MRA at 3 T versus 5 T versus 7 T: visualization of distal small cerebral arteries[J]. Radiology, 2023, 306(1): 207-217.
[22] Wang, L., Nie, D., Li, G., Puybareau, É., Dolz, J., Zhang, Q., ... & Shen, D. (2019). Benchmark on automatic six-month-old infant brain segmentation algorithms: the iSeg-2017 challenge. *IEEE transactions on medical imaging*, *38*(9), 2219-2230.
[23] Shi, F., Yap, P. T., Fan, Y., Gilmore, J. H., Lin, W., & Shen, D. (2010). Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation. *Neuroimage*, *51*(2), 684-693.
[24] Rahate, A., Walambe, R., Ramanna, S., & Kotecha, K. (2022). Multimodal co-learning: Challenges, applications with datasets, recent advances and future directions. *Information Fusion*, *81*, 203-239.
[25] Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, *16*, 100258.
[26] Semanjski, I. C. (2023). *Smart Urban Mobility: Transport Planning in the Age of Big Data and Digital Twins*. Elsevier.
[27] Marinov, Z., Jäger, P. F., Egger, J., Kleesiek, J., & Stiefelhagen, R. (2024). Deep interactive segmentation of medical images: A systematic review and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*.
[28] Abila, E., Buljan, I., Zheng, Y., Veres, T., Weng, Z., Nackenhorst, M. C., ... & Rendeiro, A. F. (2024). Tissue clocks derived from histological signatures of biological aging enable tissue-specific aging predictions from blood. *bioRxiv*, 2024-11.
[29] Obaid, D. R., Smith, D., Gilbert, M., Ashraf, S. & Chase, A. Computer simulated “Virtual TAVR” to guide TAVR in the presence of a previous Starr-Edwards mitral prosthesis. J. Cardiovasc. Comput. Tomogr. 13, 38–40 (2019).
[30] Onscale, L. Digital Twins of Lungs to Improve covid-19 patient outcomes, https://onscale.com/blog/project-breatheasy-digital-twins-of-lungs-toimprove-covid-19-patients-outcomes/

[31] Systems, D. The Living Brain Project. (2022).
[32] Wasserthal, J., Breit, H.-C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D., Cyriac, J., Yang, S., Bach, M., Segeroth, M., 2023. TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. Radiology: Artificial Intelligence. https://doi.org/10.1148/ryai.230024
[33] Lexma. Project Breath Easy Lexma Digital Twin Lung, http://projectbreatheasy.org/ (2023).
[34] Baillargeon, B., Rebelo, N., Fox, D. D., Taylor, R. L. & Kuhl, E. The Living Heart Project: A robust and integrative simulator for human heart function. Eur. J. Mech. A Solids 48,38–47 (2014)
[35] Healthineers, S. The Value of Digital Twin Technology, https://www.siemens-healthineers.com/en-us/services/valuepartnerships/asset-center/white-papers-articles/value-of-digitaltwin-technology.
[36] Obaid, D. R., Smith, D., Gilbert, M., Ashraf, S. & Chase, A. Computer simulated "Virtual TAVR" to guide TAVR in the presence of a previous Starr-Edwards mitral prosthesis. J. Cardiovasc. Comput. Tomogr. 13, 38–40 (2019).
[37] Ahmadian, H. et al. A digital twin for simulating the vertebroplasty procedure and its impact on mechanical stability of vertebra in cancer patients. Int J. Numer. Method. Biomed. Eng. 38, e3600 (2022).
[38] Fisher, C. K. et al. Machine learning for comprehensive forecasting of Alzheimer's Disease progression. Sci. Rep. 9, 13622 (2019).
[39] Badano, A. et al. Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial. JAMA Netw. Open 1, e185474 (2018).
[40] Zhao, F., Wu, Y., Hu, M., Chang, C. W., Liu, R., Qiu, R., & Yang, X. (2024). Current Progress of Digital Twin Construction Using Medical Imaging. arxiv preprint arxiv:2411.08173.
[41] O'Higgins, P., Fitton, L. C., Phillips, R., Shi, J., Liu, J., Gröning, F., ... & Fagan, M. J. (2012). Virtual functional morphology: novel approaches to the study of craniofacial form and function. Evolutionary Biology, 39, 521-535.