electronics

MDPI

Article

Transformer-Based Integrated Framework for Joint Reconstruction and Segmentation in Accelerated Knee MRI

Hongki Lim

Department of Electronic Engineering, Inha University, Incheon 22212, Republic of Korea; hklim@inha.ac.kr

Abstract: Magnetic Resonance Imaging (MRI) reconstruction and segmentation are crucial for medical diagnostics and treatment planning. Despite advances, achieving high performance in both tasks remains challenging, especially in the context of accelerated MRI acquisition. Motivated by this challenge, the objective of this study is to develop an integrated approach for MRI image reconstruction and segmentation specifically tailored for accelerated acquisition scenarios. The proposed method unifies these tasks by incorporating segmentation feedback into an iterative reconstruction algorithm and using a transformer-based encoder–decoder architecture. This architecture consists of a shared encoder and task-specific decoders, and employs a feature distillation process between the decoders. The proposed model is evaluated on the Stanford Knee MRI with Multi-Task Evaluation (SKM-TEA) dataset against established methods such as SegNetMRI and IDSLR-Seg. The results show improvements in the PSNR, SSIM, Dice, and Hausdorff distance metrics. An ablation study confirms the contribution of feature distillation and segmentation feedback to the performance gains. The advancements demonstrated in this study have the potential to impact clinical practice by facilitating more accurate diagnosis and better-informed treatment plans.

Keywords: MRI; reconstruction; segmentation; vision transformer

Citation: Lim, H. Transformer-Based Integrated Framework for Joint Reconstruction and Segmentation in Accelerated Knee MRI. *Electronics* **2023**, *12*, 4434. https://doi.org/10.3390/electronics12214434

Academic Editor: Juan M. Corchado

Received: 5 September 2023
Revised: 5 October 2023
Accepted: 26 October 2023
Published: 27 October 2023

1. Introduction

The superior soft tissue contrast of magnetic resonance imaging (MRI) makes it an invaluable diagnostic instrument across a wide variety of diseases [1]. However, prolonged acquisition times can lead to patient discomfort, reduced throughput, and introduction of motion artifacts. Therefore, reducing the duration of MRI scans has become a pressing area of research [2].

One promising approach to achieve rapid MRI acquisition is the use of compressed sensing (CS) techniques [3]. These techniques violate the Nyquist–Shannon sampling theorem by undersampling and collecting fewer measurements than is conventionally required to reconstruct diagnostic-quality images. Despite their efficiency, they contradict the standard sampling theory, leading to aliasing artifacts. To mitigate this, researchers have incorporated additional a priori knowledge [4]. Recently, machine learning (ML) techniques have been integrated into the image reconstruction process [2]. The application of ML has led to the development of algorithms that can reconstruct high-quality images from sparsely sampled MRI data, significantly accelerating MRI scans and reducing acquisition time [5–7]. Reconstructed images often serve as a means to derive clinically relevant parameters through postprocessing steps such as segmentation and tissue characterization [8]. ML tools excel in this task, even automating dense image labeling tasks to match expert variability [6,9,10].

While many machine learning-based segmentation algorithms operate under the assumption of receiving a "clean" image, they do not necessarily account for the challenges posed by undersampled MRI scenarios [11]. Though traditional approaches to medical imaging often treat acquisition, reconstruction, and segmentation as distinct phases, it is widely understood within the research community that these stages are interrelated and

can significantly impact one another. Nonetheless, the emphasis has often been on each individual stage, sometimes overlooking the cumulative effects [8]. Specifically, when prioritizing faster imaging speeds without proper image restoration, there is a risk of introducing residual aliasing and blurring, which can subsequently lead to errors in segmentation [12]. Figure 1 illustrates how a segmentation network trained on fully sampled (clean) data can yield misleading segmentation masks for undersampled (noisy) data. Thus, there is a need for efficient automated approaches that can simultaneously reconstruct MRI data and accurately segment the region of interest (ROI) [13]. However, the lack of segmentation datasets with k-space data required for MRI reconstruction poses a significant challenge in the development of joint reconstruction–segmentation algorithms [12].

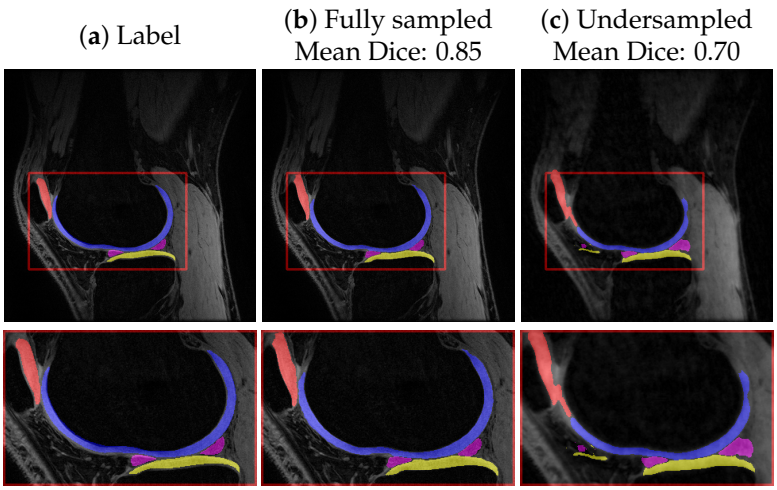

Figure 1. A demonstration of unsuccessful generalization to undersampled data when a knee MRI segmentation network (the pretrained network provided by the authors of [6]) is trained using fully sampled data. In the absence of image restoration before segmentation, the prediction becomes unreliable, leading to a 15% reduction in the mean Dice score for this specific slice. (**a**) Segmentation label and (**b**,**c**) segmentation prediction based on fully sampled data and undersampled data, respectively.

Despite limited public datasets, several studies have addressed the joint MRI reconstruction and segmentation problem. SegMRI [8] uses CS and Gaussian mixture model segmentation on patch-based dictionaries for sparse image representation, thereby enhancing image reconstruction and segmentation. SegNetMRI [11] uses an iterative framework involving a data fidelity unit and UNet [14]-based denoising and segmentation networks. Each iteration shares an encoder between segmentation and denoising with unique decoders. The segmentation decoder is reused across iterations, while denoisers use different encoders–decoders across iterations. Multiple segmentation results are merged using a 1×1 convolution at the end of iterations. FR-Net [13] presents a deep learning approach that includes a reconstruction network derived by unrolling the Fast Iterative Shrinkage–Thresholding Algorithm (FISTA) [15]. This is followed by a segmentation network that operates independently without sharing parameters with the reconstruction component. Lastly, IDSLR-SEG [12] introduces a framework for joint calibrationless Parallel MRI (PMRI) reconstruction and segmentation based on unrolling an iterative re-weighted least squares algorithm to minimize a CLEAR cost function [16] for calibrationless PMRI reconstruction [17]. The denoising network shared across all iterations and the segmentation network both use a shared encoder and are trained end-to-end using a few-shot learning strategy.

In summary, while existing models offer a foundational approach through end-to-end joint training and a shared encoder for dual tasks, they do not completely harness the latest progress in multitask learning. The method introduced in this study aims to enhance this by creating a more synergistic framework for simultaneous knee MRI reconstruction and segmentation. In this paper, we introduce an innovative approach

that takes undersampled k-space data as input and outputs both reconstructed knee MRI images and ROI segmentation.

Leveraging recent multitask learning developments, a framework is proposed for joint segmentation and reconstruction. The primary contributions of this paper are shown below:

1. Segmentation-Integrated Unrolled Reconstruction: we proposes a unique cost function for unrolling the reconstruction algorithm that integrates segmentation results into the reconstruction process.
2. Enhanced Encoder–Decoder Architecture: this paper employs a Swin Transformer [18,19]-based encoder–decoder architecture for multitask denoising and segmentation. A shared attention mechanism is implemented wherein Query and Key vectors in the self-attention module of the task-specific decoder are computed using the shared encoder's output.
3. Feature Distillation in Multitask Decoders: the proposed model introduces integration of features between decoders through a distillation process by applying spatial attention features from each task that are then incorporated into the other task's decoder.

2. Materials and Methods

2.1. Background

2.1.1. Compressed Sensing in MRI

Reconstructing MRI images can be formulated as an optimization problem [20,21], as presented in (1):

$$\hat{x} = \arg\min_x f(x) + \beta \mathsf{R}(x), \tag{1}$$

where x represents the reconstructed images, $f(x) = \frac{1}{2}\|Ax - y\|_2^2$ is the data fidelity term, y is the data acquired in the k-space domain, and A stands for the imaging model, which includes coil sensitivity profile maps, a Fourier transformation operation, and data subsampling. The regularization function $\mathsf{R}(x)$ and its corresponding parameter β constrain the problem when datasets are highly subsampled [21], preventing ill-posed conditions where multiple solutions might fulfill (1).

One way to solve (1) is to use a two-step iterative process alternating between gradient descent and a proximal operation, as provided in Equations (2) and (3) [22]:

$$x^{(n')} = x^{(n)} - t\nabla f(x^{(n)}) \tag{2}$$

$$x^{(n+1)} = \mathbf{prox}_{\beta \mathsf{R}}(x^{(n')}). \tag{3}$$

In these equations, n represents the n-th iteration, t is a scalar that indicates the gradient's step size, and $\mathbf{prox}_{\beta \mathsf{R}}(\cdot)$ denotes the proximal operator of R.

A promising approach to developing efficient reconstruction algorithms is to use a data-driven method to learn optimal trainable parameters in regularization functions. Here, the proximal step is often replaced with a deep neural network, which directly learns a parameterized form of the regularization function [23,24]. Consequently, the proximal update in (3) is redefined [21,25] as follows:

$$x^{(n+1)} = N_\theta(x^{(n')}), \tag{4}$$

where N_θ is the neural network and θ denotes learnable parameters, which can either be shared or distinct across iterations. The iterative process provided by (2) and (4) is "unrolled" into a model U_θ, which is then trained by minimizing the loss function $\min_\theta \sum_i L(U_\theta(y_i, A_i), x_i)$ [21,25], where $L(\cdot)$ measures the distance between its inputs and x_i represents the ith ground-truth example.

2.1.2. Transformers in Medical Imaging

Integration of attention mechanisms [26] into architectures influenced by Convolutional Neural Networks (CNNs) [27] has been a significant focus within the computer vision community. Consequently, it has led to the development of "Vision Transformer" (ViT) models [28]. Their popularity has grown due to their capacity to encode long-range dependencies and generate effective feature representations [29].

ViT models have demonstrated considerable potential in MRI restoration and analysis tasks. In the domain of MRI restoration, notable work includes that by Feng et al. [30], who have developed a cross-attention module capable of extracting and merging complementary features from auxiliary imaging modalities [29]. In the field of MRI analysis, a standout example is Swin UNETR [31], which has achieved leading performance in the Brain Tumor Segmentation (BraTS) 2021 challenge [32]. This model combines a Swin Transformer encoder with a CNN-based decoder. The Swin Transformer using a patch partition layer can create non-overlapping patches from the input data and construct windows for self-attention computations. These processed feature representations are then forwarded to a CNN decoder via skip connections at multiple resolutions [29].

2.1.3. Multi-Task Learning for Dense Predictions

Multi-Task Learning (MTL) [33] seeks to build generalized ML models able to generate all pertinent task outputs from a given input [34]. MTL can improve generalization capability with shared representation learning from multiple task-specific training signals [35]. MTL offers advantages over single-task learning such as increased inference speeds by avoiding repetitive feature calculations in shared layers [36], with potential performance improvements when tasks share information or are able to regularize each other [37].

Significant MTL work in pixel-level prediction tasks has led to innovations in network architecture and optimization techniques [34]. Optimization methods can maintain balance among tasks during training to prevent any single task's dominance. For instance, Kendall et al. [38] have quantified homoskedastic uncertainty to balance single-task losses, and GradNorm [39] is able to equalize task-specific gradients. On the architectural front, methods are able to leverage shared information among tasks. Approaches include hard parameter sharing [40,41] (a shared encoder branching into task-specific heads), soft parameter sharing [42,43] (individual task parameters with cross-task feature sharing), and designs that first predict tasks and then leverage these predictions to enhance task outputs [44]. For example, Xu et al. [45] have developed a multimodal distillation module that can distill information from initial predictions of other tasks using spatial attention and then incorporate it into the task of interest in order to effectively utilize intermediate predictions' complementary information.

2.2. Proposed Method

2.2.1. Incorporating Segmentation Feedback into the Reconstruction Cost Function

Improvement of reconstruction can inherently lead to enhanced segmentation; thus, in most joint MRI reconstruction and segmentation research, segmentation is typically executed subsequent to the reconstruction process. This approach has been employed in previous studies such as the study of Huang et al. [13], where the segmentation network was only employed after the reconstruction process. However, segmentation outcomes can reciprocally refine the reconstruction process. As such, a novel cost function is proposed in this study to facilitate an optimization algorithm that enables integration of segmentation results into the reconstruction process. This is achieved by appending a term $\mathsf{S}(x;\{w_k\})$ to (1):

$$\hat{x} = \arg\min_x f(x) + \beta \mathsf{R}(x) + \mu \mathsf{S}(x;\{w_k\}). \tag{5}$$

Here, μ adjusts the impact of the added term, the set $\{w_k\}$ constitutes a transformed segmentation mask, and $\mathsf{S}(x;\{w_k\})$ is a term indicating the relationship between the reconstructed image and the transformed segmentation mask.

To comprehend the role of $\mathsf{S}(x)$, consider $\mathsf{S}(x;\{\check{w}_k\}) = \frac{1}{2}\sum_{k=1}^{K}\|C_k x\|^2_{\check{W}_k} = \frac{1}{2}\sum_{k=1}^{K}(C_k x)^T \check{W}_k(C_k x)$, where $\check{W}_k = \mathrm{diag}\{\check{w}_k\}$, $\check{w}_k$ represents an indicator image for the ROI boundary extracted from the segmentation mask, and C_k denotes a finite differencing matrix in the x, y, or z directions; thus, in this case, $K = 3$. Each image update involves the gradient of $\mathsf{S}(x)$, defined as $\nabla\mathsf{S}(x) = \sum_{k=1}^{K} C_k^T \check{W}_k C_k x$, and is zeroed where $\check{w}_k$ is zero, indicating the boundary region. Consequently, the update considering $\nabla\mathsf{S}(x)$ encourages spatial smoothness outside the boundary region while limiting smoothing across boundaries.

Building on this understanding, a modification to (2) with $\mathsf{S}(x^{(n)};\{w_k^{(n)}\}) = \frac{1}{2}\sum_{k=1}^{K}\|c_k * x^{(n)}\|^2_{W_k^{(n)}}$ can be proposed, as shown below:

$$x^{(n')} = x^{(n)} - t\nabla f(x^{(n)}) - t'\nabla\mathsf{S}(x^{(n)};\{w_k^{(n)}\}), \tag{6}$$

where t' sets the step size for the added term and

$$\nabla\mathsf{S}(x^{(n)};\{w_k^{(n)}\}) = \sum_{k=1}^{K}\tilde{c}_k * \left(W_k^{(n)}\left(c_k * x^{(n)}\right)\right). \tag{7}$$

Here, $W_k^{(n)} = \mathrm{diag}\{w_k^{(n)}\}$, $w_k^{(n)} = \sum_{k'=1}^{K'} g(c'_k * m^{(n)})$, $g(\cdot)$ is an activation function for nonlinearity, $\tilde{c}_k$ is a flipped convolution kernel of c_k, and $m^{(n)}$ designates the ROI mask obtained from the segmentation network at the n^{th} iteration. Both $\{c_k\}$ and $\{c'_k\}$ denote K and K' sets of convolutional filters trained end-to-end alongside the denoising and segmentation network, thereby allowing the data to guide how the model utilizes the segmentation result for both tasks. Although $\mathsf{S}(x^{(n)};\{w_k^{(n)}\})$ is not considered as a part of the data consistency term, it is differentiable its gradient can be easily found. Therefore, the update for the $\mathsf{S}(x^{(n)};\{w_k^{(n)}\})$ term is included in (2) rather than in the proximal step. This update is denoted as the modified data consistency step in Figure 2a.

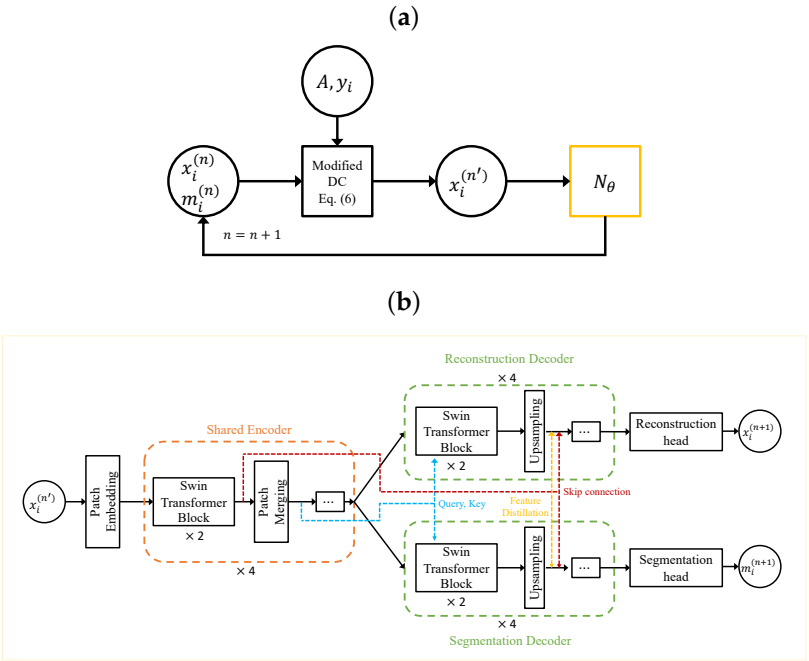

Figure 2. (**a**) High-level overview of the proposed method and (**b**) detailed block diagram of the joint denoising and segmentation network N_θ in (**a**). The proposed approach, rooted in an unrolled compressed sensing algorithm, iteratively updates both the image x and segmentation mask m. It incorporates a Swin Transformer-based encoder–decoder framework for MRI denoising and segmentation, which employs a shared attention mechanism and an innovative feature exchange process between decoders to leverage inter-task synergies and enhance performance.

2.2.2. Swin Transformer-Based Encoder–Decoder Approach

The proposed method leverages a Swin Transformer-based encoder–decoder architecture [46] for multitask [47] denoising and segmentation, as depicted in Figure 2b. The encoder, inspired by ResNet [48], includes a patch embedding process that adjusts the spatial resolution and channel dimension, followed by several Transformer blocks. This configuration produces a feature pyramid compatible with many vision task architectures. The Swin Transformer's design [18,19] of alternating window partitioning and shifted window partitioning is then applied.

The decoder module [46,47], influenced by CNN-based decoders, replaces convolutional layers with Swin Transformer blocks. The four-stage decoder comprises two Swin Transformer blocks per stage, with an upsampling layer in between to double the spatial resolution and halve the channel dimension. Here, a "stage" refers to a set of multiple Transformer blocks that process data at the same spatial resolution. In contrast to [47], the encoder features are fed to the decoder via a skip connection at the same resolution.

A shared attention mechanism [47] is employed as well. In standard Vision Transformer self-attention, each multi-head self-attention layer independently creates its Query, Key, and Value vectors using only its own input. These vectors are used to calculate an attention score, with the Query and Key vectors determining the score and the Value vector generating a weighted sum to form the self-attention output. In contrast, the shared attention mechanism modifies this process within the last Transformer block of the decoder at each stage. Specifically, it computes Query and Key vectors from the shared encoder's output that corresponds to the same spatial resolution, while the Value vector is derived from the preceding decoder stage for the specific task, ensuring task-specific outputs. This configuration mirrors the decoder in the original Transformer design [26]. By incorporating shared encoded features into the computation of the Query and Key vectors, this shared attention mechanism enhances each task's ability to utilize cross-task relationships and dependencies.

2.2.3. Feature Sharing and Distillation across Multitask Decoders

The proposed model, as outlined in Section 2.2.2, utilizes an encoder–decoder framework with a shared encoder and two distinct decoders for MRI reconstruction and segmentation tasks. This design with separate paths for each task can facilitate task-specific feature processing.

A key aspect of the proposed architecture involves sharing and integration of features between decoders. This design draws inspiration from a previous study [45]. However, unlike the previous study, the proposed approach applies a distillation process to intermediate features of decoders rather than applying it to initial predictions. Before expanding the spatial dimension, features from each task are subjected to a spatial attention process and then incorporated into the other task's decoder:

$$I_r^{i+1} = O_r^i + \sigma(W_{s,r} O_s^i) \odot O_s^i \tag{8}$$

$$I_s^{i+1} = O_s^i + \sigma(W_{r,s} O_r^i) \odot O_r^i. \tag{9}$$

Here, I_r^{i+1} and I_s^{i+1} denote inputs to the $(i+1)$-th stage of the reconstruction and segmentation decoders. Each stage contains multiple transformation blocks processing vectors of identical dimensions, O_r^i and O_s^i symbolize outputs from the i-th stage of respective decoders, $\sigma(\cdot)$ denotes the sigmoid function, and $W_{s,r}$ and $W_{r,s}$ represent tunable parameters.

The term $\sigma(W_{s,r} O_s^i)$ presents a spatial mask applied to the segmentation decoder feature for the reconstruction decoder. This spatial attention mechanism enables the model to highlight critical spatial locations within feature maps through gating, thereby controlling the information flow between decoders. Following this refinement, these features are blended into the other task's decoder before expanding the spatial dimensions. This approach takes advantage of mutual benefits between reconstruction and segmentation tasks, potentially augmenting the overall effectiveness of the network.

3. Results

3.1. Dataset Details

3.1.1. SKM-TEA Dataset

The Stanford Knee MRI with Multi-Task Evaluation (SKM-TEA) dataset [6] offers a substantial pool of quantitative knee MRI (qMRI) scans, enabling evaluation of MRI reconstruction and analysis methods. The dataset comprises around 25,000 slices from 155 patients, including raw-data measurements, scanner-generated DICOM images, manual segmentation of four tissues (Patellar Cartilage, Femoral Cartilage, Tibial Cartilage, and Meniscus), and annotations for sixteen pathologies. We employed this dataset to benchmark the comparing methods and the proposed method.

In the provided dataset, an inverse Fourier transform is applied to the fully-sampled k-space in the readout direction to generate a hybridized k-space ($x \times k_y \times k_z$). Sensitivity maps for each 2D axial slice were estimated using JSENSE [49] and the fully-sampled k-space was reconstructed using SENSE [50], which then served as the target image for the reconstruction task. Dataset acquisition used double-echo steady-state (qDESS) MRI method, which provides two sets of 3D images (termed echoes—E1 and E2). In this study, we utilized only E1 data from the two available echoes. The majority of the data consisted of 8-channel coil. Data with 16-channel coil (8 out of 155 samples) were omitted due to GPU memory constraints. This study adhered to the training, validation, and test splits provided within the dataset.

3.1.2. Data Preprocessing

The reconstruction baseline method provided by authors of the SKM-TEA dataset used 2D k-space data in the axial direction ($k_y \times k_z$). However, through empirical findings, it was observed that training a segmentation network with axial slices posed a more significant challenge compared to the sagittal direction due to the comparatively sparse distribution of various tissue classes (such as tibial cartilage and meniscus). In light of these findings, 2D k-space data in the sagittal direction were used, which is consistent with the baseline segmentation methods of the SKM-TEA dataset. For this, ($k_x \times k_y \times k_z$) k-space data were produced by applying the Fourier transform in the readout direction, then an inverse Fourier transform along the z axis, resulting in a hybridized k-space ($k_x \times k_y \times z$). For undersampling, a 2D Poisson disc at an acceleration factors of 8 was utilized with the code provided by the authors of the SKM-TEA dataset. The undersampling mask was generated for the true acquisition region (512×416), then zero-padded to match with the kspace data size (512×512). During training, 10,000 precomputed undersampling masks were cached to ensure consistency across different training sessions. A fixed undersampling mask was generated for each scan in the test dataset for evaluation.

3.2. Baseline and Comparative Methods

As baseline, a 2D UNet trained for joint reconstruction and segmentation was employed utilizing an image-to-image approach. UNet features a shared encoder and two distinct decoders, with each decoder being dedicated to either reconstruction or segmentation. Moreover,the proposed method was compared with previously suggested methods for joint MRI reconstruction and segmentation, including SegNetMRI [11] and IDSLR-SEG [12]. These methods alternate between data consistency and denoising via neural networks, bearing close resemblance to the unrolled compressed sensing in (2)–(4). Considering the dataset's distinct challenge of multicoil Knee MRI reconstruction and segmentation, these methods were adapted and reimplemented, in the course of which reimplementations were based on unrolled compressed sensing, utilizing U-Net as the denoiser and the segmentation network as detailed in previous studies [11,12]. For IDSLR-SEG, the method was adapted into a calibrated approach considering the provision of sensitivity maps with the dataset. The main divergences between implementations of SegNetMRI and IDSLR-SEG, lie in whether to use a shared or distinct denoising encoder across iterations and whether segmentation occurs multiple times during the iterative process. The distinctions between

the models are summarized in Table 1. A visual comparison between SegNetMRI and IDSLR-SEG has been provided previously in [12].

Table 1. Comparative analysis of different models (DSNA: Denoising and Segmentation Network Architecture; SEDSN: Shared Encoder between Denoising and Segmentation Networks; SDI: Shared Denoiser across Iterations; MSPI: Multiple Segmentation Predictions across Iterations).

Model	DSNA	SEDSN	SDI	MSPI
SegNetMRI	U-Net	Yes	No	Yes
IDSLR-Seg	U-Net	Yes	Yes	No
Proposed	Transformer	Yes	Yes	Yes

3.3. Details on Implementation and Training

The following training specifications were applied to all methods mentioned in this section. The Pytorch [51] deep learning library was leveraged for training. The training objective combined a complex $\ell 1$ loss for the reconstruction task and a soft Dice loss for the segmentation task. Input k-space data were normalized using the same standard deviation value of the target 3D volume as in the reconstruction baseline method of SKM-TEA dataset. Training was conducted over 200 epochs using the AdamW optimizer with a weight decay of 0.05. The learning rate was adjusted using a custom CyclicLR scheduler, with the parameters set as follows: maximum learning rate of 0.0005, gamma value of 0.5, and step size up of 15. Gradient accumulation steps were used to achieve an effective batch size of 16 across all methods. For a fair comparison, the number of trainable parameters for all methods was set at approximately 40 million. All unrolling-based methods employed a total of four unrolling iterations. The best epoch was determined using the sum of PSNR and the mean Dice score multiplied by 40, and was evaluated using the validation split of the SKM-TEA dataset for all methods presented in this section.

In the proposed method, both the encoder and each decoder featured two Transformer blocks at each stage. The window size was set as 8 and the number of heads in each stage was set as 3, 6, 12, and 24, respectively. K and K' in (7) were set as 16. To enhance the training stability, two approaches were employed in the implementation of the method described in Section 2.2.1: First, the terms $c_k * x^{(n)}$ and $c'_k * m^{(n)}$ in (7) were calculated using a sequence of layers composed of a convolutional layer, followed by an instance normalization layer, and a PReLU activation function. Second, the multiplication involving $W_k^{(n)}$ in (7) was parameterized by concatenating $w_k^{(n)}$ and $c_k * x^{(n)}$, which was then followed by a similar sequence of layers as in the previous approach. For increased model flexibility, separate convolutional filters were used instead of employing the flipped version of c_k, denoted as $\tilde{c}_k$ in (7).

3.4. Results: Quantitative and Qualitative Evaluation

The test split of the SKM-TEA dataset (34 samples, excluding two 16-channel coil samples) was used for performance assessments. The reconstruction quality was quantitatively evaluated using the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM [52]) metrics. The segmentation performance was measured using the Dice similarity coefficients and 95% Hausdorff distance for each tissue class. The PyTorch-based MEDDLR framework [53] was employed for evaluations. These metrics were computed based on 3D volumes, with the mean and standard deviation values reported from 34 volumes.

In the reconstruction task, the proposed model generally outperformed other methods in both PSNR and SSIM metrics, as highlighted in Table 2. This numerical advantage was corroborated by the empirical observations. The images generated by the proposed model, as illustrated in Figure 3, displayed fewer or similar levels of errors when compared to those produced by alternative methods.

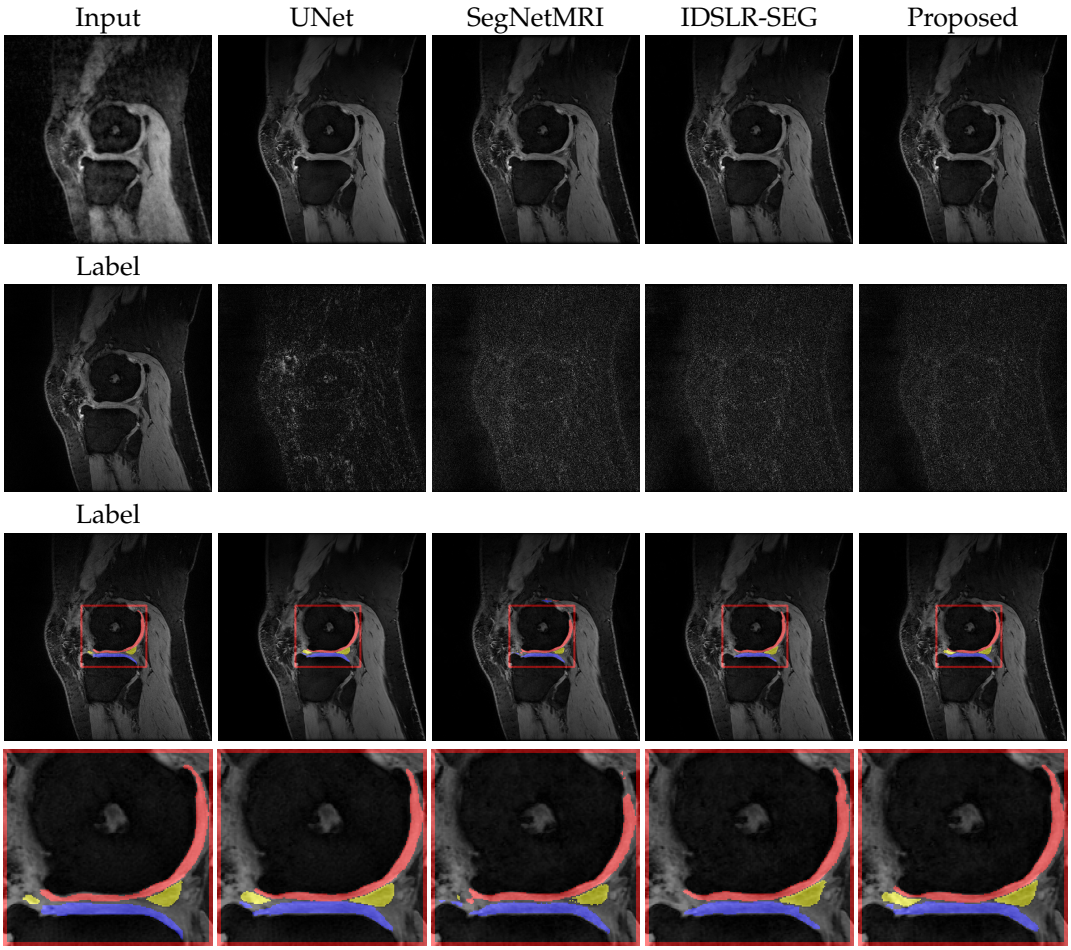

Figure 3. Comparative results for reconstruction and segmentation from 8× accelerated MRI data using various methods, including the proposed approach. The top row displays magnitude images that represent the reconstruction outcomes of each method. The second row shows the corresponding error images. The third and fourth rows overlay segmentation results on the reconstructed magnitude images obtained from the respective methods.

In the segmentation task, the proposed method yielded competitive Dice scores and demonstrated improvements in Hausdorff distances compared to other methods, as indicated by the data in Table 2. This performance is visually corroborated in Figure 3.

To sum up, the proposed model offers more accurate results in the reconstruction task and is on par with other models in terms of Dice scores in the segmentation task while improving the Hausdorff distances. The proposed architecture, which combines transformer-based encoders and decoders for an integrated approach to reconstruction and segmentation, contributed to these results.

Table 2. Comparison of different joint MRI reconstruction and segmentation methods using the peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), Dice score, and 95% Hausdorff distance (mean (standard deviation)).

Method	Reconstruction		Segmentation	
	PSNR (dB)	SSIM	DICE	Hausdorff (mm)
UNet	33.153 (0.977)	0.765 (0.027)	0.824 (0.053)	5.25 (3.27)
SegNetMRI	35.322 (1.025)	0.834 (0.022)	0.821 (0.051)	13.62 (17.37)
IDSLR-Seg	35.139 (1.002)	0.828 (0.022)	0.826 (0.056)	6.68 (8.42)
Proposed	35.550 (1.012)	0.834 (0.021)	0.825 (0.055)	4.63 (3.14)

4. Discussion

In this study, a novel approach has been introduced for integrated MRI image reconstruction and segmentation, challenging the traditional view of these tasks as separate entities. The proposed method synergistically combines these tasks, utilizing a combination of distillation applied to features between decoders and a reconstruction cost function guided by segmentation feedback. The proposed approach demonstrates advantages over existing techniques in terms of achieving improved image reconstruction and segmentation quality.

4.1. Discussion of Reconstruction Results

SegNetMRI and IDSLR-SEG produced reconstructed images of comparable quality, whereas U-Net removed many image details and introduced a higher level of blur, consistent with the numerical results shown in Table 2. These results suggest that design differences between SegNetMRI and IDSLR-SEG, as summarized in Table 1, do not significantly affect the final reconstruction outcome. However, the utilization of iterative methods was found to be a crucial factor in enhancing reconstruction performance. While IDSLR-Seg and SegNetMRI performed well, they were slightly outperformed by the proposed method. The proposed method showed the highest performance in terms of PSNR, with a value of 35.550, and was tied with SegNetMRI in SSIM with a score of 0.834.

4.2. Discussion of Segmentation Results

IDSLR-Seg led in terms of the DICE coefficient with a score of 0.826, closely followed by the proposed method at 0.825 (0.055) and UNet at 0.824. The proposed method excelled at minimizing the Hausdorff distance, achieving the lowest score of 4.63. The segmentation of meniscus pixels proved more challenging than that of other tissues. The SegNetMRI and IDSLR-SEG methods struggled to accurately segment the meniscus in the specific slice depicted in Figure 3. Contrary to the reconstruction results, design variations between SegNetMRI and IDSLR-SEG influenced their segmentation performance. Specifically, IDSLR-SEG outperformed SegNetMRI in terms of the Hausdorff distance, suggesting that that the use of distinct denoisers across iterations or the merging of multiple segmentation results using a 1×1 convolutional operation can adversely affect segmentation performance.

4.3. Ablation Study

To evaluate the influence of specific design elements on the performance of the proposed method in MRI reconstruction and segmentation tasks, an ablation study was performed. The analysis was mainly focused on two components: (i) feature distillation between task-specific decoders, and (ii) implementation of a reconstruction cost function guided by segmentation feedback.

To ascertain the individual contributions of feature distillation (FD) and segmentation feedback (SF), the proposed model was evaluated under three different conditions: without feature distillation and segmentation feedback, with only segmentation feedback, and with both. The results were then used to determine the effects of these features on the model's performance compared to a baseline model devoid of these components.

Table 3 presents the results of the ablation study. The data suggest that both FD and SF contributed to the enhanced performance of the proposed model. Specifically, When neither FD nor SF was applied, the PSNR was 35.524 dB, the SSIM was 0.836, the DICE was 0.826, and the Hausdorff distance was 4.702 mm. Implementing SF alone improved both the PSNR (35.554 dB) and SSIM (0.840), indicating enhanced reconstruction performance; however, it led to a slight degradation in the DICE (0.819) and an increase in the Hausdorff distance (5.225 mm), suggesting that the segmentation performance suffered. Incorporating both FD and SF yielded a PSNR of 35.550 dB and an SSIM of 0.834, maintaining strong reconstruction performance. Notably, the DICE score was almost identical to the baseline (0.825) and the Hausdorff distance improved to 4.63 mm, suggesting a balanced improvement across both

reconstruction and segmentation tasks. In summary, while the inclusion of segmentation feedback (SF) did improve reconstruction performance as measured by PSNR and SSIM, it slightly compromised the segmentation performance in terms of DICE and increased the Hausdorff distance. Incorporating feature distillation (FD) mitigated this sacrifice in segmentation performance and resulted in a balanced overall performance improvement.

Table 3. Ablation study results comparing the effectiveness of Feature Distillation (FD) and Segmentation Feedback (SF) on the peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), Dice score, and 95% Hausdorff distance (mean (standard deviation)).

Proposed Method		Reconstruction		Segmentation	
FD	**SF**	**PSNR (dB)**	**SSIM**	**DICE**	**Hausdorff (mm)**
No	No	35.524 (1.024)	0.836 (0.021)	0.826 (0.054)	4.70 (2.55)
No	Yes	35.554 (1.016)	0.840 (0.020)	0.819 (0.057)	5.23 (3.42)
Yes	Yes	35.550 (1.012)	0.834 (0.021)	0.825 (0.055)	4.63 (3.14)

4.4. Limitations

This study has certain limitations tied to the dataset used in the experiment. The method's training and validation were limited to a specific set of MRI images sourced from a single hospital and obtained using scanners from the same manufacturer. This might restrict the generalizability of this study's findings. Future studies should aim to validate the proposed model with more diverse datasets incorporating images from a variety of MRI machines and diverse patient populations.

Although our ablation study indicated improved performance due to the Transformer-based network architecture, feature distillation, and segmentation feedback,the individual contributions of these techniques deserve further exploration. Additionally, the computational cost of the proposed model might pose challenges in time-sensitive clinical settings.

5. Conclusions

The proposed model offers threefold benefits: enhanced accuracy in MRI reconstruction, precise tissue segmentation, and substantial time savings during both MRI acquisition and post-acquisition image analysis. Employing an acceleration factor of eight for undersampling theoretically reduces the MRI scan time by a factor of eight. Concurrent segmentation capabilities further streamline the diagnostic process, potentially saving additional time in clinical workflows. The model's proficiency in detail-rich reconstruction and precise tissue segmentation holds promise for earlier diagnosis and more timely treatments. These compelling results warrant further research to fully realize the model's potential in clinical settings.

Future work should concentrate on optimizing the algorithm to reduce computational time without sacrificing performance. Furthermore, future studies should investigate the scalability of the model [54] and assess the relationship between the number of parameters and performance. Insight into this relationship could inform the development of more efficient model architectures. Future investigations should explore pretraining strategies that make use of extensive datasets. Studying the effects of large-scale pretraining on the model's performance could reveal new avenues for improving MRI reconstruction and segmentation.

Funding: This work was supported in part by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (2022R1F1A1069055), in part by the MSIT (Ministry of Science and ICT), Korea, under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2023-RS-2023-00259678) supervised by the IITP (Institute for Information and Communications Technology Planning and Evaluation), in part by a Korean Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korean Government (MOTIE) (RS-2023-00243974, Graduate School of Digital-based Sustainable Energy Process Innovation

Convergence), in part by an Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korean government (MSIT) (No.RS-2022-00155915, Artificial Intelligence Convergence Innovation Human Resources Development (Inha University)), and in part by an Inha University Research Grant.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found at: https://github.com/StanfordMIMI/skm-tea (accessed on 6 July 2023).

Acknowledgments: The authors would like to acknowledge Arjun Desai (Stanford) for his invaluable feedback regarding the utilization of the SKM-TEA dataset.

Conflicts of Interest: The author declares no conflict of interest. The funders had no role in the design of the study, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

MRI	Magnetic Resonance Imaging
SKM-TEA	Stanford Knee MRI with Multi-Task Evaluation
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index
ML	Machine Learning
ROI	Region of Interest
FISTA	Fast Iterative Shrinkage–Thresholding Algorithm
PMRI	Parallel MRI
CLEAR	Calibration-free Locally low-rank EncourAging Reconstruction
Swin	Shifted Windows
ViT	Vision Transformer
BraTS	Brain Tumor Segmentation
MTL	Multi-Task Learning
CNN	Convolutional Neural Network
ReLU	Rectified Linear Unit
qMRI	Quantitative MRI
DICOM	Digital Imaging and Communications in Medicine
SENSE	Sensitivity Encoding
JSENSE	Joint Image Reconstruction and Sensitivity Estimation in SENSE
qDESS	Double-Echo Steady-State
GPU	Graphics Processing Unit
2D	Two-Dimensional
3D	Three-Dimensional
DSNA	Denoising and Segmentation Network Architecture
SEDSN	Shared Encoder between Denoising and Segmentation Networks
SDI	Shared Denoiser across Iterations
MSPI	Multiple Segmentation Predictions across Iterations
FD	Feature Distillation
SF	Segmentation Feedback

References

1. van Beek, E.J.; Kuhl, C.; Anzai, Y.; Desmond, P.; Ehman, R.L.; Gong, Q.; Gold, G.; Gulani, V.; Hall-Craggs, M.; Leiner, T.; et al. Value of MRI in medicine: More than just another test? *J. Magn. Reson. Imaging* **2019**, *49*, e14–e25. [CrossRef] [PubMed]
2. Zbontar, J.; Knoll, F.; Sriram, A.; Murrell, T.; Huang, Z.; Muckley, M.J.; Defazio, A.; Stern, R.; Johnson, P.; Bruno, M.; et al. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv* **2018**, arXiv:1811.08839.
3. Lustig, M.; Donoho, D.; Pauly, J.M. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med. Off. J. Int. Soc. Magn. Reson. Med.* **2007**, *58*, 1182–1195. [CrossRef] [PubMed]

4. Tibshirani, R. Regression shrinkage and selection via the LASSO. *J. R. Stat. Soc. Ser. Stat. Methodol.* **1996**, *58*, 267–288. [CrossRef]
5. Hammernik, K.; Klatzer, T.; Kobler, E.; Recht, M.P.; Sodickson, D.K.; Pock, T.; Knoll, F. Learning a variational network for reconstruction of accelerated MRI data. *Magn. Reson. Med.* **2018**, *79*, 3055–3071. [CrossRef] [PubMed]
6. Desai, A.D.; Schmidt, A.M.; Rubin, E.B.; Sandino, C.M.; Black, M.S.; Mazzoli, V.; Stevens, K.J.; Boutin, R.; Ré, C.; Gold, G.E.; et al. SKM-TEA: A dataset for accelerated MRI reconstruction with dense image labels for quantitative clinical evaluation. *arXiv* **2022**, arXiv:2203.06823.
7. Pal, A.; Rathi, Y. A review and experimental evaluation of deep learning methods for MRI reconstruction. *J. Mach. Learn. Biomed. Imaging* **2022**, *1*, 001. [CrossRef]
8. Caballero, J.; Bai, W.; Price, A.N.; Rueckert, D.; Hajnal, J.V. Application-driven MRI: Joint reconstruction and segmentation from undersampled MRI data. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014: 17th International Conference, Boston, MA, USA, 14–18 September 2014; Proceedings, Part I 17; Springer: Berlin/Heidelberg, Germany, 2014; pp. 106–113.
9. Bien, N.; Rajpurkar, P.; Ball, R.L.; Irvin, J.; Park, A.; Jones, E.; Bereket, M.; Patel, B.N.; Yeom, K.W.; Shpanskaya, K.; et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet. *PLoS Med.* **2018**, *15*, e1002699. [CrossRef]
10. Liu, Z.; Tong, L.; Chen, L.; Jiang, Z.; Zhou, F.; Zhang, Q.; Zhang, X.; Jin, Y.; Zhou, H. Deep learning based brain tumor segmentation: A survey. *Complex Intell. Syst.* **2023**, *9*, 1001–1026. [CrossRef]
11. Sun, L.; Fan, Z.; Ding, X.; Huang, Y.; Paisley, J. Joint CS-MRI reconstruction and segmentation with a unified deep network. In Proceedings of the Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, 2–7 June 2019; Proceedings 26; Springer: Berlin/Heidelberg, Germany, 2019; pp. 492–504.
12. Pramanik, A.; Jacob, M. Joint calibrationless reconstruction and segmentation of parallel MRI. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 437–453.
13. Huang, Q.; Yang, D.; Yi, J.; Axel, L.; Metaxas, D. FR-Net: Joint reconstruction and segmentation in compressed sensing cardiac MRI. In Proceedings of the Functional Imaging and Modeling of the Heart: 10th International Conference, FIMH 2019, Bordeaux, France, 6–8 June 2019; Proceedings 10; Springer: Berlin/Heidelberg, Germany, 2019; pp. 352–360.
14. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
15. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202. [CrossRef]
16. Trzasko, J.D.; Manduca, A. CLEAR: Calibration-free parallel imaging using locally low-rank encouraging reconstruction. *Proc. Int. Soc. Magn. Reson. Med.* **2012**, *517*.
17. Pramanik, A.; Aggarwal, H.K.; Jacob, M. Deep generalization of structured low-rank algorithms (Deep-SLR). *IEEE Trans. Med. Imaging* **2020**, *39*, 4186–4197. [CrossRef] [PubMed]
18. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
19. Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L.; et al. Swin transformer v2: Scaling up capacity and resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LO, USA, 21–24 June 2022; pp. 12009–12019.
20. Fessler, J.A. Optimization methods for magnetic resonance image reconstruction: Key models and optimization algorithms. *IEEE Signal Process. Mag.* **2020**, *37*, 33–40. [CrossRef] [PubMed]
21. Sandino, C.M.; Cheng, J.Y.; Chen, F.; Mardani, M.; Pauly, J.M.; Vasanawala, S.S. Compressed sensing: From research to clinical practice with deep neural networks: Shortening scan times for magnetic resonance imaging. *IEEE Signal Process. Mag.* **2020**, *37*, 117–127. [CrossRef]
22. Combettes, P.L.; Pesquet, J.C. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*; Springer: New York, NY, USA, 2011; pp. 185–212.
23. Mardani, M.; Sun, Q.; Donoho, D.; Papyan, V.; Monajemi, H.; Vasanawala, S.; Pauly, J. Neural proximal gradient descent for compressive imaging. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 9596–9606.
24. Aggarwal, H.K.; Mani, M.P.; Jacob, M. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Trans. Med. Imaging* **2018**, *38*, 394–405. [CrossRef]
25. Diamond, S.; Sitzmann, V.; Heide, F.; Wetzstein, G. Unrolled optimization with deep priors. *arXiv* **2017**, arXiv:1705.08041.
26. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
27. Ramachandran, P.; Parmar, N.; Vaswani, A.; Bello, I.; Levskaya, A.; Shlens, J. Stand-alone self-attention in vision models. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 68–80.

28. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
29. Shamshad, F.; Khan, S.; Zamir, S.W.; Khan, M.H.; Hayat, M.; Khan, F.S.; Fu, H. Transformers in medical imaging: A survey. *Med. Image Anal.* **2023**, *88*, 102802. [CrossRef] [PubMed]
30. Feng, C.M.; Yan, Y.; Chen, G.; Xu, Y.; Hu, Y.; Shao, L.; Fu, H. Multi-modal transformer for accelerated MR imaging. *IEEE Trans. Med. Imaging* **2022**, *42*, 2804–2816. [CrossRef] [PubMed]
31. Hatamizadeh, A.; Nath, V.; Tang, Y.; Yang, D.; Roth, H.R.; Xu, D. Swin unetr: Swin transformers for semantic segmentation of brain tumors in MRI images. In Proceedings of the International MICCAI Brainlesion Workshop, Virtual Event, 27 September 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 272–284.
32. Baid, U.; Ghodasara, S.; Mohan, S.; Bilello, M.; Calabrese, E.; Colak, E.; Farahani, K.; Kalpathy-Cramer, J.; Kitamura, F.C.; Pati, S.; et al. The RSNA-ASNR-MICCAI BRATS 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv* **2021**, arXiv:2107.02314.
33. Caruana, R. Multitask learning. *Mach. Learn.* **1997**, *28*, 41–75. [CrossRef]
34. Vandenhende, S.; Georgoulis, S.; Van Gansbeke, W.; Proesmans, M.; Dai, D.; Van Gool, L. Multi-task learning for dense prediction tasks: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3614–3633. [CrossRef]
35. Ruder, S. An overview of multi-task learning in deep neural networks. *arXiv* **2017**, arXiv:1706.05098.
36. Eigen, D.; Fergus, R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 2650–2658.
37. Standley, T.; Zamir, A.; Chen, D.; Guibas, L.; Malik, J.; Savarese, S. Which tasks should be learned together in multi-task learning? In Proceedings of the 37th International Conference on Machine Learning, Online, 13–18 July 2020; pp. 9120–9132.
38. Kendall, A.; Gal, Y.; Cipolla, R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018, pp. 7482–7491.
39. Chen, Z.; Badrinarayanan, V.; Lee, C.Y.; Rabinovich, A. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 794–803.
40. Lu, Y.; Kumar, A.; Zhai, S.; Cheng, Y.; Javidi, T.; Feris, R. Fully-adaptive feature sharing in multi-task networks with applications in person attribute classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5334–5343.
41. Guo, P.; Lee, C.Y.; Ulbricht, D. Learning to branch for multi-task learning. In Proceedings of the 37th International Conference on Machine Learning, Online, 13–18 July 2020; pp. 3854–3863.
42. Misra, I.; Shrivastava, A.; Gupta, A.; Hebert, M. Cross-stitch networks for multi-task learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 3994–4003.
43. Liu, S.; Johns, E.; Davison, A.J. End-to-end multi-task learning with attention. In Proceedings of the IEEE/CVF Conference On Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1871–1880.
44. Vandenhende, S.; Georgoulis, S.; Van Gool, L. MTI-net: Multi-scale task interaction networks for multi-task learning. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part IV 16; Springer: Berlin/Heidelberg, Germany, 2020; pp. 527–543.
45. Xu, D.; Ouyang, W.; Wang, X.; Sebe, N. Pad-net: Multi-tasks guided prediction-and-distillation network for simultaneous depth estimation and scene parsing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 675–684.
46. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 205–218.
47. Bhattacharjee, D.; Zhang, T.; Süsstrunk, S.; Salzmann, M. Mult: An end-to-end multitask learning transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12031–12041.
48. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
49. Ying, L.; Sheng, J. Joint image reconstruction and sensitivity estimation in SENSE (JSENSE). *Magn. Reson. Med. Off. J. Int. Soc. Magn. Reson. Med.* **2007**, *57*, 1196–1202. [CrossRef]
50. Pruessmann, K.P.; Weiger, M.; Scheidegger, M.B.; Boesiger, P. SENSE: Sensitivity encoding for fast MRI. *Magn. Reson. Med. Off. J. Int. Soc. Magn. Reson. Med.* **1999**, *42*, 952–962. [CrossRef]
51. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.P.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 8026.
52. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]

53. Desai, A.D.; Ozturkler, B.M.; Sandino, C.M.; Vasanawala, S.; Hargreaves, B.A.; Re, C.M.; Pauly, J.M.; Chaudhari, A.S. Noise2Recon: A Semi-Supervised Framework for Joint MRI Reconstruction and Denoising. *arXiv* **2021**, arXiv:2110.00075.
54. Kaplan, J.; McCandlish, S.; Henighan, T.; Brown, T.B.; Chess, B.; Child, R.; Gray, S.; Radford, A.; Wu, J.; Amodei, D. Scaling laws for neural language models. *arXiv* **2020**, arXiv:2001.08361.