Journal of Artificial Intelligence Practice (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2026.090111
ISSN 2371-8412 Vol. 9 Num. 1

Accelerated Cardiac MRI T1 Mapping Reconstruction Using Gated Linear Attention Network

Chenqi Liu

School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai, China

Keywords: Cardiac MRI, T1 Mapping, Deep Unrolling, Gated Linear Attention

Abstract: Quantitative cardiac magnetic resonance (CMR) imaging often relies on high undersampling techniques to reduce scan time and suppress motion artifacts. However, extremely high acceleration rates inevitably lead to severe image aliasing and the loss of anatomical details. Existing reconstruction methods based on deep unrolling models face a dual challenge when processing dynamic sequences: severe error accumulation in early iterations and massive computational overhead for high-dimensional spatio-temporal feature extraction. To address these issues, this paper proposes an efficient and lightweight CMR reconstruction network (EGLAVarNet). First, to tackle the "cold start" problem in early iterations, a Content-Aware Structured Initialization (CASI) module is designed. This module leverages explicit low-rank physical priors to dynamically synthesize the initial state of the memory matrix, effectively suppressing the interference of high-rank random artifacts. Second, a Gated Linear Attention (GLA) unit is introduced in the core denoising stage. By utilizing the associative property of matrix multiplication to extract global structural features, this module reduces the time complexity of attention computation to a linear scale of $O(N)$ without requiring spatial downsampling. Experimental results on the public CMRxRecon dataset demonstrate that under extremely high acceleration rates (e.g., 8x and 10x), the reconstruction accuracy of EGLAVarNet significantly outperforms existing state-of-the-art methods. Compared to the baseline model, the proposed network not only visually improves the structural fidelity of myocardial boundaries but also reduces parameter count and computational overhead by approximately 50%. It achieves an optimal balance between reconstruction accuracy and model complexity, providing a highly robust and low-deployment-cost new solution for the rapid clinical imaging of dynamic cardiac sequences.

1. Introduction

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide. CMR imaging is widely recognized as the "gold standard" for the non-invasive assessment of cardiac structure and function. However, conventional CMR primarily relies on relative image contrast, making its signal intensity highly susceptible to variations in scanner hardware and acquisition parameters. To overcome this limitation, quantitative MRI (qMRI) has emerged as a crucial technical branch of CMR. By employing specific pulse sequence designs and signal modeling, qMRI extracts absolute

parameters that reflect the intrinsic physical properties of the myocardium, thereby significantly enhancing the objectivity of clinical diagnoses. As a prominent qMRI technique rooted in the development of rapid relaxometry acquisition and quantitative modeling methods [1, 2, 3], T_1 mapping enables the precise measurement of native longitudinal relaxation time (native T_1) and extracellular volume (ECV) fraction. These quantitative metrics are highly sensitive to microscopic myocardial pathologies, such as edema and diffuse fibrosis. Consequently, acquiring high-precision T_1 parametric maps holds profound clinical value for the early risk stratification and prognostic evaluation of cardiac diseases.

Despite its significant clinical utility, mainstream T_1 mapping sequences (e.g., [3]) require prolonged breath-holds, rendering them highly susceptible to myocardial motion artifacts. Therefore, the introduction of k-space undersampling techniques is imperative to accelerate the acquisition process. However, conventional reconstruction methods encounter severe bottlenecks at high acceleration rates. Parallel imaging (PI) methods such as SENSE and GRAPPA [4, 5] suffer from pronounced noise amplification and limited robustness under aggressive acceleration, while compressed sensing based approaches [6] often result in the loss of high-frequency details and entail prohibitively long non-linear iterative computation times. In recent years, deep learning (DL) has been widely applied to MR image reconstruction, ranging from direct manifold-learning based recovery [7] to variational network based reconstruction frameworks [8]. Early "black-box" direct reconstruction approaches lack adherence to the forward physical acquisition process, making them prone to generating artificial hallucinations. To address this, physics-driven [9] models have emerged. These methods (e.g., [10]) deeply integrate physical constraints with network optimization, enabling end-to-end learning while preserving physical fidelity. When dealing with high-dimensional dynamic data across multiple temporal phases, spatio-temporal modeling becomes essential [11]. The [12] further incorporates recurrent structures (e.g., ConvGRU). By utilizing hidden states to propagate spatio-temporal features across unrolled iterations, [12] has established itself as the current mainstream state-of-the-art solution for dynamic cardiac sequence reconstruction.

The main contributions of this study are summarized as follows:

(1) Proposed the Content-Aware Structured Initialization (CASI) module: By introducing explicit low-rank physical priors, this module dynamically synthesizes the initial state of the global memory matrix using a predefined low-rank vector bank. This mechanism helps to suppress the interference of high-rank random noise during early iterations, providing a stabilized initial state for the network and alleviating the error accumulation commonly observed in traditional models.

(2) Designed the Gated Linear Attention (GLA) denoising unit: It leverages the associative property of matrix multiplication to compress spatial features into a channel-wise covariance representation, enabling the extraction of global structural features without the need for spatial downsampling. This module reduces the time complexity of attention computation to a linear scale of $\mathcal{O}(N)$ and facilitates the efficient update and propagation of memory features across time steps via a gating mechanism.

(3) Constructed a lightweight dynamic cardiac reconstruction network (EGLAVarNet): By integrating the CASI module and the GLA denoising unit into a physics-driven unrolling framework, the proposed network significantly improves the structural fidelity of myocardial details at high acceleration rates (e.g., 8x) while strictly controlling computational overhead. This provides a reliable methodological reference for the rapid quantitative imaging of dynamic cardiac sequences.

2. Method

2.1 Overall Architecture

Physics-driven unrolling models achieve an effective integration of solution-space constraints and data-driven learning by unfolding iterative optimization algorithms into cascaded neural network modules. In the reconstruction architecture proposed in this study, the forward physical operator is first defined as $A = P \cdot \mathcal{F} \cdot S$, where S denotes the multi-channel coil sensitivity operator, $\mathcal{F}$ represents the 2D Fast Fourier Transform (FFT), and P is the undersampling mask simulating hardware acquisition. Accordingly, its adjoint operator, i.e., the backward physical operator, is defined as $A^H = S^H \cdot \mathcal{F}^{-1} \cdot P^H$, where S^H performs the coherent combination of multi-channel signals using the Root Sum of Squares (RSS) algorithm.

As illustrated in Figure 1, the network consists of T cascaded iterative modules, referred to as time steps. At the initialization stage of the iteration, the system retains the initial undersampled k-space data k_0 as the physical measurement baseline. Simultaneously, k_0 is mapped to the image domain using the backward operator to generate a zero-filled image X_{ZF}. Subsequently, X_{ZF} is fed into the CASI module to extract physical priors and generate the initial global memory matrix M_0. In subsequent cascaded iterations, this matrix is continuously propagated and updated across time steps in the form of M_t.

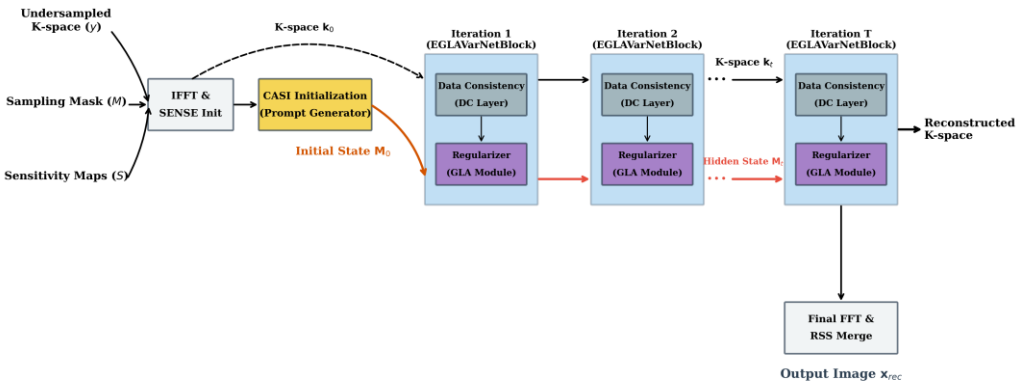


Figure 1: Overview of EGLAVarNet architecture.

At the t-th time step, where $t \in [1, T]$, the network executes a cyclic operation of "backward mapping - feature denoising - forward projection - data consistency calibration":

First, the network receives the k-space state k_{t-1} from the previous time step and maps it to the initial spatial image x_t via the backward physical operator:

$$x_t = S^H \cdot \mathcal{F}^{-1} \cdot P^H \cdot k_{t-1} \tag{1}$$

Next, x_t and the global memory matrix M_{t-1} propagated from the previous time step are jointly fed into the Gated Linear Attention denoising unit (GLARecurrentUnit). Employing a residual connection to optimize gradient propagation, this module outputs the denoised image $x_{denoised}$ and the updated memory matrix M_t:

$$x_{denoised} = x_t + \text{GLARecurrentUnit}(x_t, M_{t-1}) \tag{2}$$

Then, the denoised image is re-projected back to the k-space via the forward physical operator, generating a prior regularization term k_{prior}, which represents the frequency-domain distribution of

the ideal image predicted by the network:

$$k_{prior} = P \cdot \mathcal{F} \cdot S \cdot x_{denoised} \tag{3}$$

Finally, in the data consistency (DC) update stage, this step essentially acts mathematically as a single gradient descent optimization step on the data fidelity objective function. The network calculates the residual between the current state and the original measurement k_0 to obtain the physical constraint gradient, and combines it with the prior term k_{prior} to complete the final update of the k-space state:

$$k_t = k_{t-1} - \eta_t \cdot P \cdot (k_{t-1} - k_0) + k_{prior} \tag{4}$$

Here, η_t is a learnable parameter specifically assigned to the t-th cascaded module, acting equivalently as an adaptive step size in gradient descent. This parameter forcibly pulls back the network's prediction to ensure the physical fidelity of the reconstruction results to the true measurement data. The aforementioned process is iteratively executed T times, progressively approximating the optimal high-fidelity reconstruction solution within the deep unrolling architecture.

2.2 Gated Linear Attention (GLA) Denoising Unit

The computational complexity of the standard self-attention mechanism grows quadratically ($\mathcal{O}(N^2)$) with spatial resolution, making it highly prone to out-of-memory (OOM) errors when processing full-resolution medical images. Furthermore, it lacks a mechanism to propagate physical priors across time steps. To address these limitations, this study designs the GLA denoising unit. This module leverages the associative property of matrix multiplication to bypass the computationally expensive Softmax operation, following the general motivation of efficient linear-complexity attention mechanisms [13]. By maintaining a fixed-dimensional global memory matrix, it enables the storage and retrieval of physical priors with linear complexity.

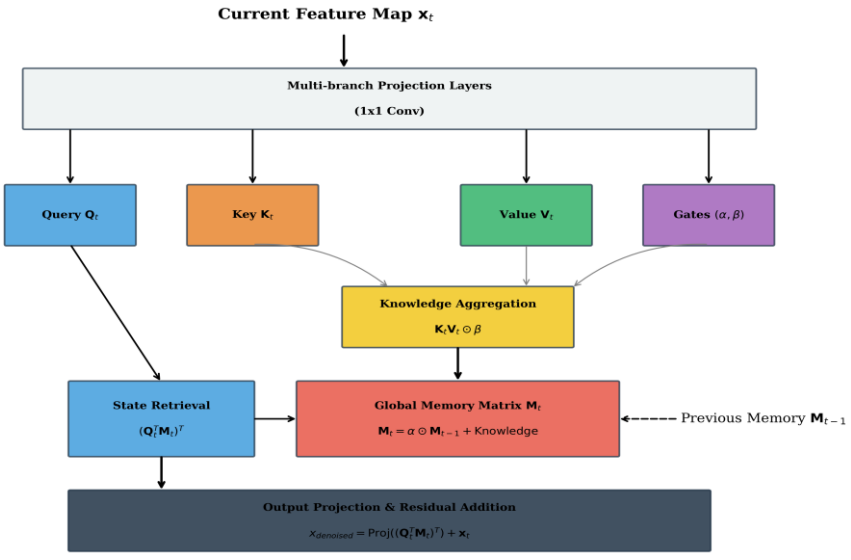


Figure 2: Internal architecture and memory matrix update flow of GLA module.

As illustrated in Figure 2, at the t-th time step, the input feature $\boldsymbol{x}_t \in \mathbb{R}^{C\times(H\times W)}$ first undergoes a linear projection to generate the query matrix $\boldsymbol{Q}_t$, key matrix $\boldsymbol{K}_t$, value matrix $\boldsymbol{V}_t$, alongside the

channel-wise forget gate α and input gate β:

$$\boldsymbol{Q}_t, \boldsymbol{K}_t, \boldsymbol{V}_t, \alpha, \beta = \text{LinearProjection}(\boldsymbol{x}_t) \tag{5}$$

Here, the gating parameters $\alpha, \beta \in [0,1]^C$ control the decay of historical memory and the injection of new knowledge at the current step, respectively.

During the knowledge aggregation stage, the module computes the cross-correlation between $\boldsymbol{K}_t$ and $\boldsymbol{V}_t^T$, compressing the spatial-dimensional features into a channel-wise covariance matrix representation. Integrated with the input gate β, the newborn knowledge matrix for the current time step is generated:

$$\text{Knowledge} = (\beta \odot \boldsymbol{K}_t)\boldsymbol{V}_t^T \tag{6}$$

This operation reduces the feature dimension from $C \times (HW)$ to $C \times C$, precisely extracting the global structural prior of the current image.

Subsequently, by combining the global memory matrix $\boldsymbol{M}_{t-1}$ from the previous time step with the forget gate α, the model executes a cross-time-step update of the memory bank:

$$\boldsymbol{M}_t = \alpha \odot \boldsymbol{M}_{t-1} + \text{Knowledge} \tag{7}$$

The updated $\boldsymbol{M}_t \in \mathbb{R}^{C \times C}$ serves as a cross-time-step memory dictionary. The forget gate adaptively erases residual artifact responses from historical iterations, while the input gate writes deterministic anatomical features as new knowledge.

Finally, in the state retrieval and residual fusion stage, the network utilizes $\boldsymbol{Q}_t$ to initiate a targeted retrieval from the global memory bank $\boldsymbol{M}_t$, extracting the enhanced spatial prior. Through a linear projection (Proj) and residual fusion with the original input $\boldsymbol{x}_t$, it outputs the denoised feature map $x_{denoised}$:

$$x_{denoised} = \text{Proj}(\boldsymbol{M}_t^T \boldsymbol{Q}_t) + \boldsymbol{x}_t \tag{8}$$

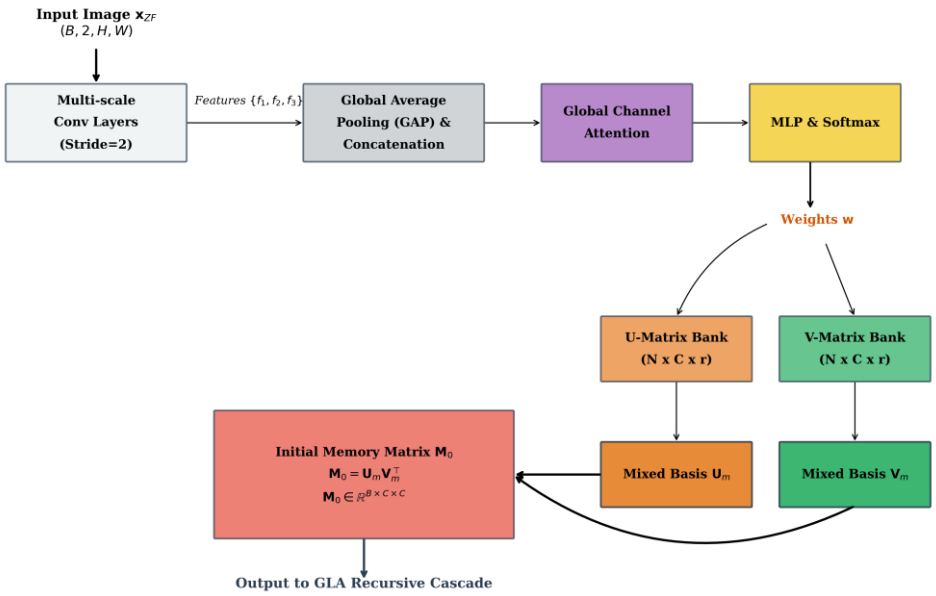
In summary, the GLA module extracts global features at a full-resolution scale without requiring spatial downsampling. It optimizes the time complexity of the attention mechanism to a linear level of $\mathcal{O}(N)$, while restricting the spatial complexity of the memory matrix to a constant level of $\mathcal{O}(1)$. This explicit feature evolution mechanism enables the network to effectively utilize global anatomical consensus to guide local detail denoising, thereby improving the structural fidelity of the model under high undersampling rates.

2.3 Content-Aware Structured Initialization (CASI) Module

In traditional deep unrolling networks, hidden states are typically initialized as zero vectors or mapped statically using simple convolutions. Under high undersampling rates, this approach makes the network highly prone to fitting high-frequency noise during early iterations, triggering the so-called "cold start" error accumulation. To alleviate this issue, this study designs the CASI module. This module introduces explicit low-rank physical priors, mapping the content features of the input image to the initial state of the global memory matrix.

As illustrated in Figure 3, the CASI module first feeds the backward-reconstructed initial image $\boldsymbol{x}_{ZF}$ into three-layer cascaded downsampling convolutions to extract multi-scale spatial features f_1, f_2, f_3. Subsequently, these spatial features are compressed via Global Average Pooling (GAP) and concatenated across scales to form a high-dimensional context descriptor $\boldsymbol{v}_{all}$:

$$\boldsymbol{v}_{all} = \text{Concat}\big(\text{GAP}(f_1), \text{GAP}(f_2), \text{GAP}(f_3)\big) \tag{9}$$



Cross-scale Adaptive State Initialization (CASI) with Explicit Low-rank Constraints

Figure 3: Cross-scale adaptive state initialization with explicit low-rank constraints.

To adapt to undersampling artifacts with various distribution patterns, the module introduces a cross-scale channel attention mechanism to perform weighted calibration on $\boldsymbol{v}_{all}$, and utilizes a Multi-Layer Perceptron (MLP) to predict the dynamic combination weights $\boldsymbol{w}$ tailored to the current input content:

$$\boldsymbol{w} = \text{softmax}\left(\text{MLP}\left(\text{Attention}(\boldsymbol{v}_{all})\right)\right) \tag{10}$$

In MR physics, genuine anatomical image sequences exhibit high correlation across spatial and contrast dimensions, inherently demonstrating low-rankness. Conversely, undersampling artifacts and measurement noise typically manifest as high-rank random interferences. Based on this physical principle, the CASI module refrains from directly regressing a full-rank memory matrix. Instead, it internally maintains two sets of learnable low-rank vector banks, $\boldsymbol{U}_{bank}$ and $\boldsymbol{V}_{bank}$, each containing N bases, with their low-rank dimension r being significantly smaller than the channel number C ($r \ll C$). According to the predicted weights $\boldsymbol{w}$, the model dynamically synthesizes the mixed basis matrices $\boldsymbol{U}_m$ and $\boldsymbol{V}_m$ customized for the current image:

$$\boldsymbol{U}_m = \sum_{i=1}^{N} w_i \, \boldsymbol{U}_{bank,i}, \quad \boldsymbol{V}_m = \sum_{i=1}^{N} w_i \, \boldsymbol{V}_{bank,i} \tag{11}$$

Finally, the initial global memory matrix $\boldsymbol{M}_0$ is synthesized via matrix multiplication:

$$\boldsymbol{M}_0 = \boldsymbol{U}_m \boldsymbol{V}_m^T \tag{12}$$

Mathematically, this formulation strictly ensures that the rank of the generated $\boldsymbol{M}_0$ does not exceed the predefined threshold r. This explicit low-rank construction operation essentially serves as a feature filter. It forces the network to eliminate high-rank random interferences through a low-rank truncation mechanism prior to the iterations, retaining only the robust anatomical skeleton information. This mechanism provides a noise-resistant "warm start" state for the subsequent GLA recurrent cascaded steps, effectively improving the convergence trajectory of the network under

extremely high acceleration rates.

3. Experiment

3.1 Experimental Setup

3.1.1 Dataset and Implementation Details

This study is based on the public [14] challenge dataset. Image acquisition was performed using a Siemens 3T MRI scanner with a [3] sequence and a 4-1-3-1-2 sampling scheme to acquire 9 short-axis images at different inversion times (TIs). Typical scanning parameters were as follows: field of view (FOV) 340×340 mm^2, spatial resolution 1.5×1.5 mm^2, and slice thickness 5.0 mm. The experiments retrospectively simulated 1D Cartesian undersampling along the phase-encoding direction, retaining 24 auto-calibration signal (ACS) lines at the center of k-space. The undersampling acceleration factors were set to 4x, 8x, and 10x, respectively. The dataset was strictly partitioned according to the official standards into a training set (120 subjects), a validation set (56 subjects), and an independent test set (108 subjects). The model was implemented using the PyTorch framework and trained on two NVIDIA RTX 3090 GPUs.

3.1.2 Loss Function and Training Strategy

Network optimization was performed using the Adam optimizer with an initial learning rate of 3×10^{-4}, which was dynamically scheduled via a step decay strategy. The batch size was set to 1. To balance frequency-domain physical fidelity and spatial-domain anatomical details, a joint loss function was employed for end-to-end optimization:

$$L_{total} = \lambda_1 L_{kspace} + \lambda_2 L_{l1} + \lambda_3 L_{ssim} + \lambda_4 L_{hfen} \tag{13}$$

where L_{kspace} denotes the frequency-domain normalized absolute error, L_{l1} is the image-domain absolute error, L_{ssim} represents the structural similarity loss, and L_{hfen} stands for the high-frequency error norm (HFEN) extracted based on the Laplacian of Gaussian (LoG) operator [15]. The weights of the respective loss terms were empirically set to $\lambda_1 = 3$ and $\lambda_2 = \lambda_3 = \lambda_4 = 1$.

3.1.3 Evaluation Metrics

To objectively and quantitatively evaluate the reconstruction quality, Normalized Mean Square Error (NMSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM) were utilized as the core evaluation metrics. NMSE reflects the global numerical deviation of the reconstructed image from the reference image; PSNR focuses on evaluating the model's capability to suppress aliasing artifacts and random noise; and SSIM is used to measure the structural fidelity of myocardial boundary contours and local tissues.

3.2 Experimental Results and Analysis

3.2.1 Quantitative Evaluation

To verify the effectiveness of the proposed network, this study compares EGLAVarNet with representative unrolling networks, [12] and VSharp dynamic, under three acceleration rates (R=4, 8, and 10). The quantitative results are presented in Table 1 EGLAVarNet achieves the optimal performance across all evaluation metrics. At R=4, the proposed model attains a PSNR of 49.32 dB and an SSIM of 0.9892, yielding an improvement of approximately 1.5 dB over the baseline [12].

Even under the extreme undersampling scenario at R=10, its PSNR is maintained at 41.66 dB, significantly outperforming VSharp dynamic (38.85 dB).

Table 1: Quantitative comparison of reconstruction results among different models.

AccFactor	Model Variants	PSNR (dB) ↑	SSIM ↑	NMSE ↓
R=4	Recurrent VarNet (Baseline)	47.85 ± 1.45	0.9851 ± 0.0039	0.0046 ± 0.0020
R=4	GLAVarNet (+ GLA)	48.35 ± 1.49	0.9869 ± 0.0034	0.0041 ± 0.0018
R=4	**EGLAVarNet (+ GLA & CASI)**	**49.32 ± 1.52**	**0.9892 ± 0.0028**	**0.0033 ± 0.0016**
R=8	Recurrent VarNet (Baseline)	43.46 ± 1.34	0.9727 ± 0.0057	0.0124 ± 0.0041
R=8	GLAVarNet (+ GLA)	42.64 ± 1.33	0.9720 ± 0.0057	0.0149 ± 0.0045
R=8	**EGLAVarNet (+ GLA & CASI)**	**43.73 ± 1.38**	**0.9773 ± 0.0049**	**0.0117 ± 0.0038**
R=10	Recurrent VarNet (Baseline)	41.54 ± 1.27	0.9672 ± 0.0064	0.0192 ± 0.0056
R=10	GLAVarNet (+ GLA)	40.84 ± 1.26	0.9659 ± 0.0066	0.0225 ± 0.0059
R=10	**EGLAVarNet (+ GLA & CASI)**	**41.66 ± 1.31**	**0.9704 ± 0.0059**	**0.0187 ± 0.0051**

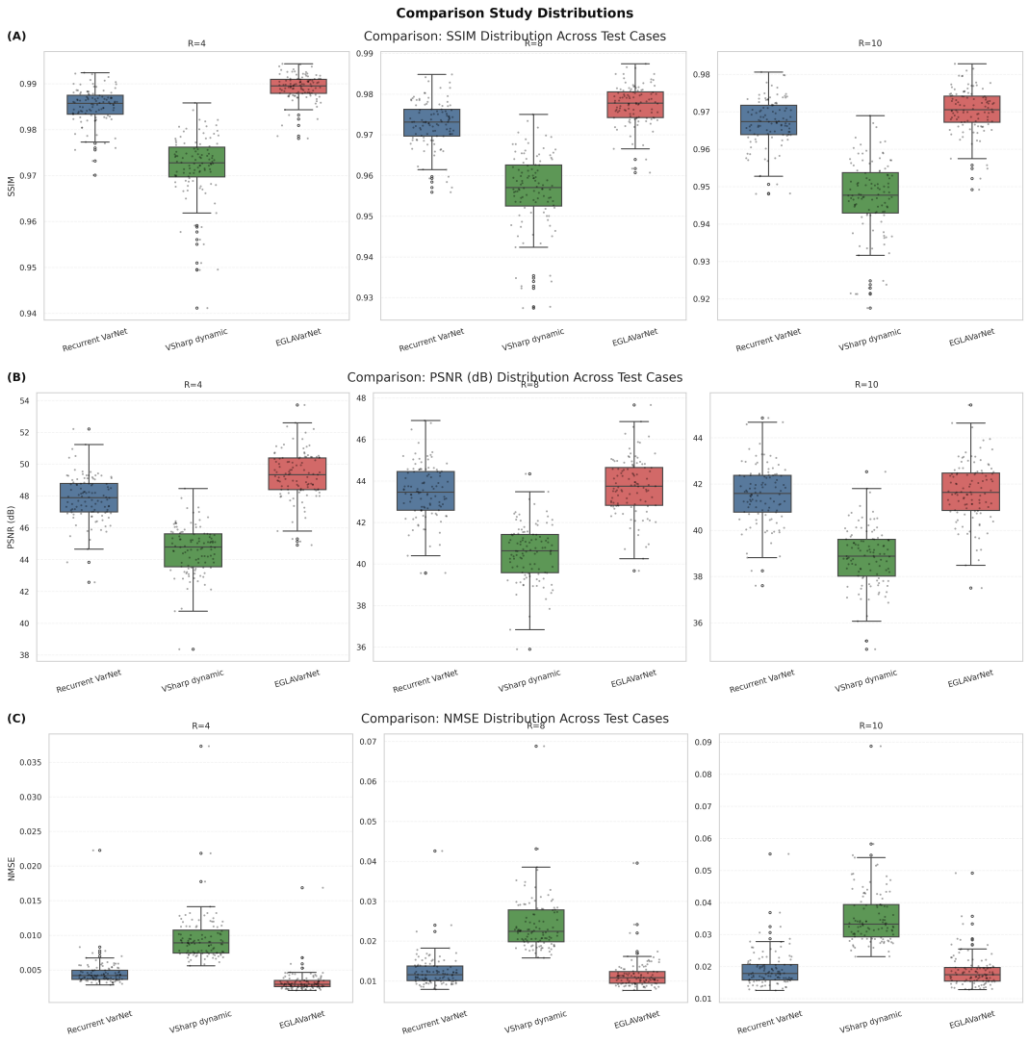


Figure 4: Boxplots of test results for each model across different metrics and acceleration rates.

To further evaluate the generalization stability of the models across different test samples, Figure 4 plots the boxplots of the evaluation metrics for each model. The data distribution indicates that the metrics for EGLAVarNet are more concentrated than those of the comparative methods, exhibiting a smaller variance and a significant reduction in abnormal outliers. In contrast, VSharp dynamic

produces a massive amount of underperforming outliers. This is primarily attributed to its bulky 3D convolutional architecture, which incurs an extremely high GPU memory footprint. Under constrained hardware conditions, its training must be executed using mixed precision, inevitably leading to numerical precision loss and the degradation of high-frequency details.

3.2.2 Qualitative Visual Analysis

Figure 5 and Figure 6 present the progressive acceleration results of the proposed model and visual comparisons among multiple models, respectively. At R=4, all unrolling networks can generate images with high fidelity. However, as the acceleration rate increases, the ill-posedness of the reconstruction problem intensifies significantly, and traditional methods are prone to generating obvious aliasing artifacts at tissue boundaries.

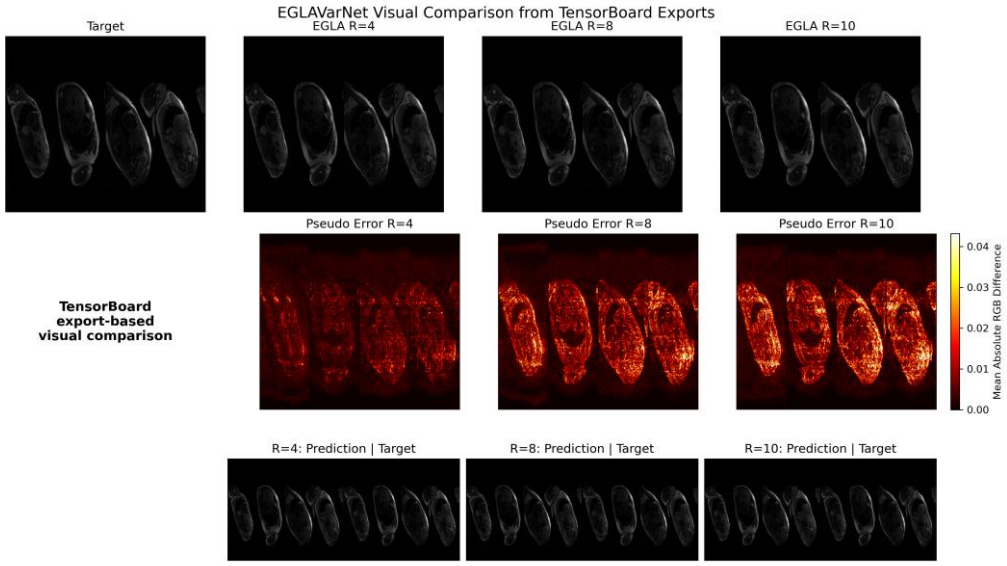


Figure 5: Reconstruction results and error heatmaps of EGLAVarNet under different acceleration rates.

Observing the pseudo-color error maps, thanks to the low-rank prior constraints provided by the CASI module and the global feature extraction of the GLA module, EGLAVarNet effectively suppresses the propagation of large-scale aliasing artifacts even under a high acceleration of R=8, maintaining high sharpness in the cardiac borders and cavity structures. Under the extreme acceleration of R=10, although minor local streak artifacts appear in the image, EGLAVarNet does not suffer from the deformation or collapse of macroscopic topological structures. These visual results are consistent with the aforementioned quantitative analysis, demonstrating that the proposed network possesses strong artifact-resistance robustness and anatomical structural fidelity under conditions of severe data deprivation.

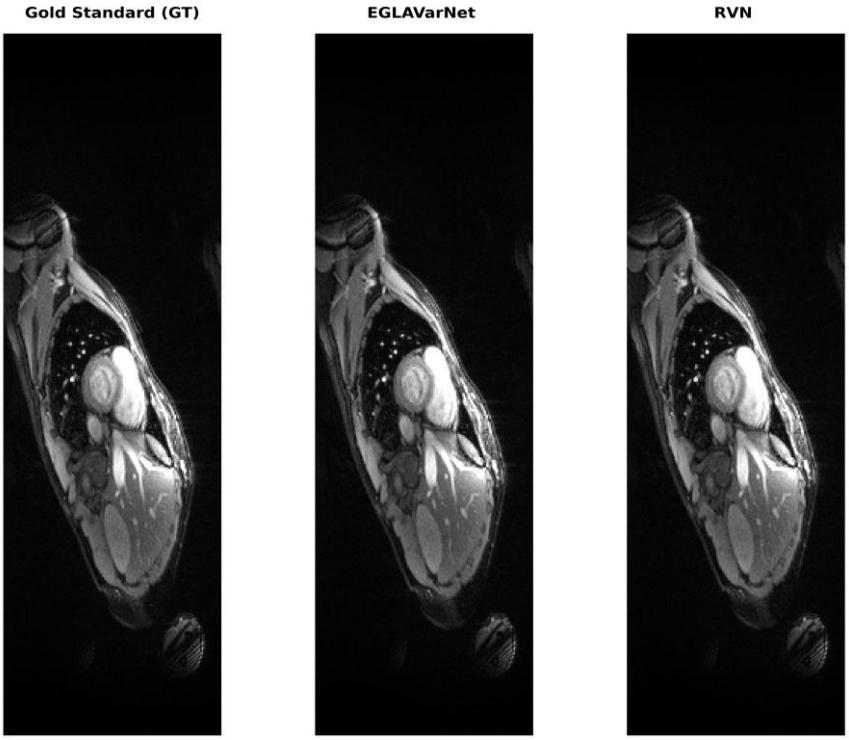


Figure 6: Visual comparison of reconstruction results among EGLAVarNet, [12], and Ground Truth (GT).

3.3 Model Complexity Analysis

In the practical clinical application of MR reconstruction, the computational overhead and parameter scale of an algorithm are critical factors determining its deployment feasibility. Table 2 lists the parameter counts, computational operations (in MACs), and model weight sizes of the compared models. The data indicate that the proposed EGLAVarNet possesses a significant lightweight advantage, with a parameter count of only 3.28 M and a computational cost of 236.58 GMACs. Compared to the baseline [12], EGLAVarNet incorporates the CASI and GLA modules while simultaneously reducing both the parameter count and computational overhead by approximately 50

Table 2: Comparison of parameter counts, computational costs, and weight sizes among different models.

Model Name	Params ↓	FLOPs ↓	Size ↓
VSharp dynamic (Mixed-precision)	180.81 M	2.58 TMacs	723.22 MB
Recurrent VarNet (Baseline)	6.87 M	506.46 GMacs	27.48 MB
EGLAVarNet (Proposed)	**3.28 M**	**236.58 GMacs**	**12.99 MB**

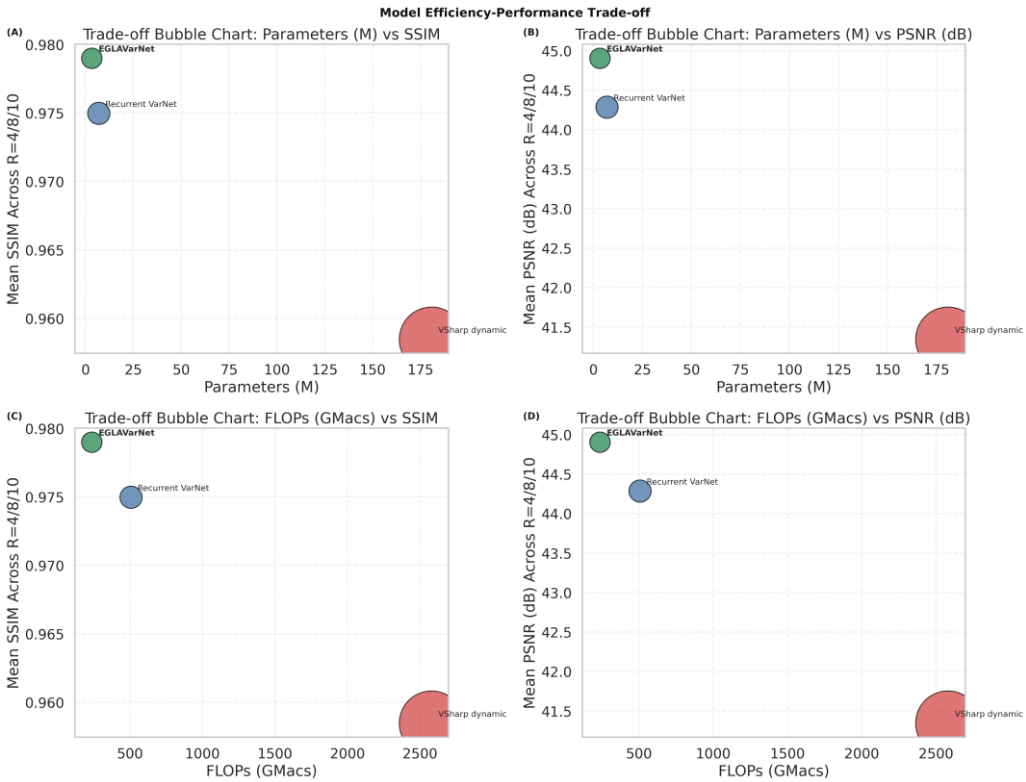


Figure 7: Pareto bubble charts of the trade-off between different metrics and computational costs for various models.

To intuitively evaluate the comprehensive trade-off between computational overhead and performance, Figure 7 plots the Pareto bubble charts of reconstruction accuracy (PSNR and SSIM) versus computational cost for each model under different acceleration rates. The coordinates of the bubbles in the figure reflect the relationship between performance and computational power, while the area of the bubbles represents the size of the model weights. The distribution results demonstrate that EGLAVarNet is situated at the Pareto frontier (i.e., the top-left corner) across all quadrants, with the smallest bubble area. This result confirms that EGLAVarNet does not rely on the expansion of parameter scale to boost accuracy. Instead, through an efficient architectural design, it breaks through existing performance bottlenecks while significantly reducing the computational burden, achieving an optimal balance between model lightweighting and the high fidelity of anatomical structures.

3.4 Ablation Study and Analysis

To verify the effectiveness of the core components in the proposed network, a step-wise ablation study was designed using [12] as the baseline model: the GLA denoising unit was sequentially introduced to form GLAVarNet, followed by the further incorporation of the CASI low-rank initialization module to construct the final EGLAVarNet. Table 3 lists the quantitative evaluation metrics for each model under different acceleration rates.

At an acceleration rate of R=4, introducing the GLA module yields noticeable performance gains, with the PSNR of GLAVarNet improving by approximately 0.5 dB compared to the baseline. This indicates that when artifacts are relatively mild, the GLA module can effectively extract global contextual features to assist reconstruction. However, as the acceleration rate increases to R=8 and R=10, the performance of GLAVarNet exhibits degradation. For instance, at R=10, its PSNR drops

from the baseline's 41.54 dB to 40.84 dB. This phenomenon aligns with the aforementioned hypothesis: under high undersampling rates, the initial zero-filled image contains severe high-rank aliasing artifacts. Lacking the guidance of physical priors, the global memory matrix of GLA is highly susceptible to noise interference during early iterations, trapping the network in a "cold start" dilemma and leading to error accumulation.

Table 3: Quantitative evaluation metrics of different model variants under various acceleration rates.

AccFactor	Model Variants	PSNR (dB) ↑	SSIM ↑	NMSE ↓
R=4	Recurrent VarNet (Baseline)	47.85 ± 1.45	0.9851 ± 0.0039	0.0046 ± 0.0020
R=4	GLAVarNet (+ GLA)	48.35 ± 1.49	0.9869 ± 0.0034	0.0041 ± 0.0018
R=4	**EGLAVarNet (+ GLA & CASI)**	**49.32 ± 1.52**	**0.9892 ± 0.0028**	**0.0033 ± 0.0016**
R=8	Recurrent VarNet (Baseline)	43.46 ± 1.34	0.9727 ± 0.0057	0.0124 ± 0.0041
R=8	GLAVarNet (+ GLA)	42.64 ± 1.33	0.9720 ± 0.0057	0.0149 ± 0.0045
R=8	**EGLAVarNet (+ GLA & CASI)**	**43.73 ± 1.38**	**0.9773 ± 0.0049**	**0.0117 ± 0.0038**
R=10	Recurrent VarNet (Baseline)	41.54 ± 1.27	0.9672 ± 0.0064	0.0192 ± 0.0056
R=10	GLAVarNet (+ GLA)	40.84 ± 1.26	0.9659 ± 0.0066	0.0225 ± 0.0059
R=10	**EGLAVarNet (+ GLA & CASI)**	**41.66 ± 1.31**	**0.9704 ± 0.0059**	**0.0187 ± 0.0051**

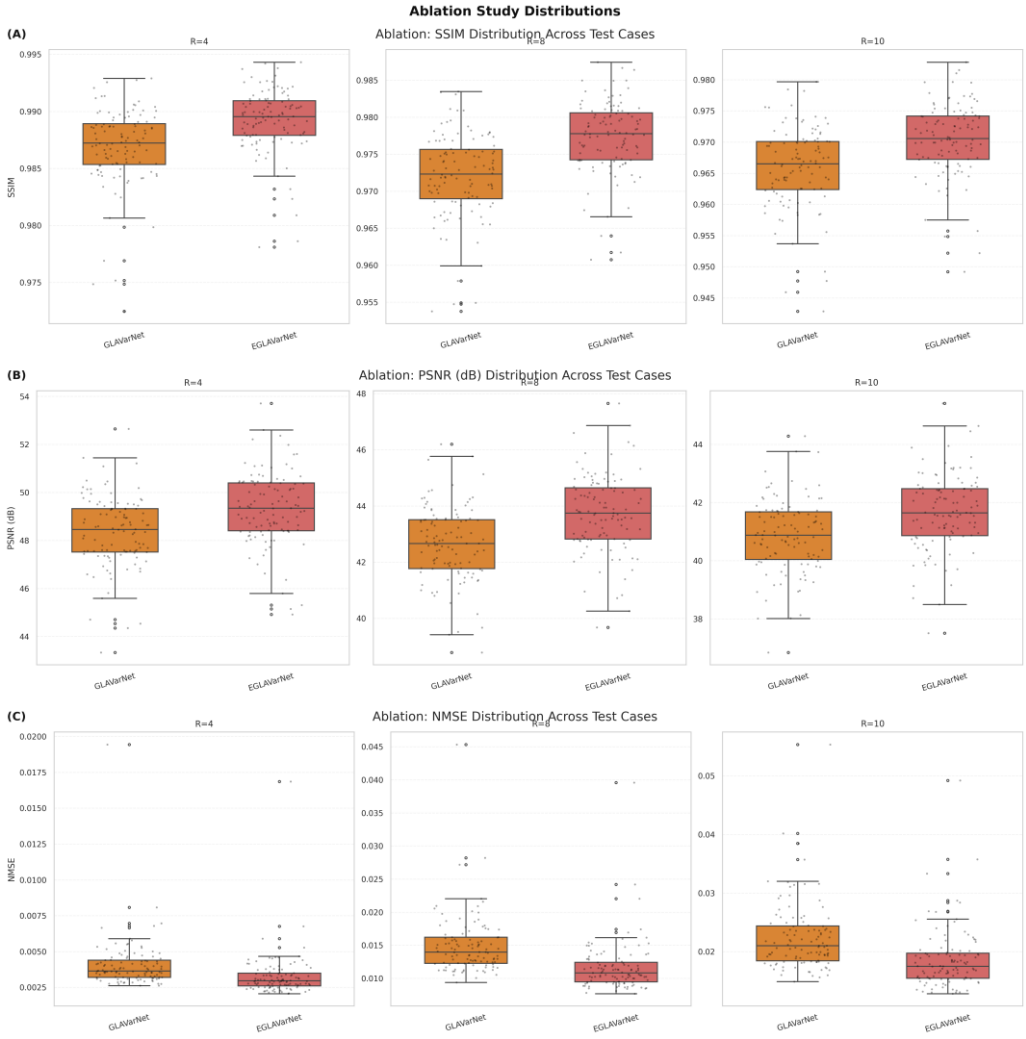


Figure 8: Boxplots of quantitative comparisons between GLAVarNet and EGLAVarNet.

Upon further incorporating the CASI module, the complete architecture (EGLAVarNet) achieves optimal performance across all acceleration rates. At R=10, its PSNR increases to 41.66 dB, and the

NMSE decreases to 0.0187. This demonstrates that the CASI module utilizes low-rank anatomical priors to effectively suppress the interference of high-rank artifacts in early iterations. It provides a relatively robust initial state for the feature evolution of the GLA module, thereby verifying the necessity of their synergistic interaction.

Figure 8 presents the boxplot distributions of the quantitative metrics. In the high-acceleration tests, after introducing the CASI module, the boxplots of EGLAVarNet shift upwards overall, and the medians of all core metrics are improved. This systematic improvement in data distribution implies that the introduction of the CASI module is not merely meant to repair individual extreme failure cases; rather, by providing a more accurate starting point for global optimization, it universally elevates the reconstruction lower bound and overall reliability of the model when dealing with complex data deprivation.

4. Conclusion

To address the massive computational overhead and early error accumulation issues faced by deep unrolling networks when processing highly accelerated CMR data, this paper proposes a lightweight physics-driven reconstruction network, EGLAVarNet. The network introduces the GLA mechanism, which optimizes the time complexity of spatial attention computation to a linear scale, enabling the propagation of global structural features across time steps while effectively controlling the parameter scale. Concurrently, the CASI module is designed to utilize explicit low-rank physical priors to suppress high-rank aliasing artifacts during the initial iterations, providing a noise-resistant and stable initial optimization state for the network.

Qualitative and quantitative experimental results demonstrate that under extremely high acceleration rates (e.g., 8x and 10x), EGLAVarNet can effectively suppress aliasing artifacts and preserve high-frequency anatomical details. Compared to existing state-of-the-art methods, the proposed model achieves a superior balance between reconstruction accuracy and model complexity, reaching Pareto optimality while significantly reducing parameter count and computational overhead. This study effectively overcomes the ill-posedness caused by extreme data deprivation, providing an efficient, reliable, and lightweight solution for the rapid clinical quantitative imaging of dynamic cardiac sequences.

References

[1] Look, D.C. and Locker, D.R. (1970) Time Saving in Measurement of NMR and EPR Relaxation Times. Review of Scientific Instruments, 41(2), 250–251. doi:10.1063/1.1684482.

[2] Deichmann, R. and Haase, A. (1992) Quantification of T1 Values by SNAPSHOT-FLASH NMR Imaging. Journal of Magnetic Resonance (1969), 96(3), 608–612. doi:10.1016/0022-2364(92)90347-A.

[3] Messroghli, D.R., Radjenovic, A., Kozerke, S., Higgins, D.M., Sivananthan, M.U. and Ridgway, J.P. (2004) Modified Look-Locker Inversion Recovery (MOLLI) for High-Resolution T1 Mapping of the Heart. Magnetic Resonance in Medicine, 52(1), 141–146. doi:10.1002/mrm.20110.

[4] Pruessmann, K.P., Weiger, M., Börnert, P. and Boesiger, P. (2001) Advances in Sensitivity Encoding with Arbitrary k-Space Trajectories. Magnetic Resonance in Medicine, 46(4), 638–651. doi:10.1002/mrm.1241.

[5] Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B. and Haase, A. (2002) Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA). Magnetic Resonance in Medicine, 47(6), 1202–1210. doi:10.1002/mrm.10171.

[6] Lustig, M., Donoho, D. and Pauly, J.M. (2007) Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging. Magnetic Resonance in Medicine, 58(6), 1182–1195. doi:10.1002/mrm.21391.

[7] Zhu, B., Liu, J.Z., Cauley, S.F., Rosen, B.R. and Rosen, M.S. (2018) Image Reconstruction by Domain-Transform Manifold Learning. Nature, 555(7697), 487–492. doi:10.1038/nature25988.

[8] Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T. and Knoll, F. (2018) Learning a Variational Network for Reconstruction of Accelerated MRI Data. Magnetic Resonance in Medicine, 79(6), 3055–3071. doi:10.1002/mrm.26977.

[9] Shlezinger, N., Segarra, S., Zhang, Y., Avrahami, D., Davidov, Z., Routtenberg, T. and Eldar, Y.C. (2025) Deep Unfolding: Recent Developments, Theory, and Design Guidelines. arXiv preprint arXiv:2512.03768 (2025). doi:10.48550/arXiv.2512.03768.
[10] Sriram, A., Zbontar, J., Murrell, T., Defazio, A., Zitnick, C.L., Yakubova, N., Knoll, F. and Johnson, P. (2020) End-to-End Variational Networks for Accelerated MRI Reconstruction, in: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D. and Joskowicz, L. (Eds.), Medical Image Computing and Computer Assisted Intervention–MICCAI 2020, Springer International Publishing, 2020, pp. 64–73. doi:10.1007/978-3-030-59713-9_7.
[11] Liang, Z.-P. (2007) SPATIOTEMPORAL IMAGINGWITH PARTIALLY SEPARABLE FUNCTIONS, in: 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2007, pp. 988–991. doi:10.1109/ISBI.2007.357020.
[12] Yiasemis, G., Sonke, J.-J., Sánchez, C. and Teuwen, J. (2022) Recurrent Variational Network: A Deep Learning Inverse Problem Solver Applied to the Task of Accelerated MRI Reconstruction. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 732–741.
[13] Shen, Z., Zhang, M., Zhao, H., Yi, S. and Li, H. (2018) Efficient attention: Attention with linear complexities. arXiv preprint arXiv:1812.01243 (2018). doi:10.48550/arXiv.1812.01243.
[14] Zhao, Y., Zhang, Y. and Tao, Q. (2024) Relaxometry Guided Quantitative Cardiac Magnetic Resonance Image Reconstruction, in: Camara, O., Puyol-Antón, E., Sermesant, M., Suinesiaputra, A., Tao, Q., Wang, C. and Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. Regular and CMRxRecon Challenge Papers, Springer Nature Switzerland, 2024, pp. 349–358. doi:10.1007/978-3-031-52448-6_33.
[15] Adamson, P.M., Desai, A.D., Dominic, J., Varma, M., Bluethgen, C., Wood, J.P., Syed, A.B., Boutin, R.D., Stevens, K.J., Vasanawala, S., Pauly, J.M., Gunel, B. and Chaudhari, A.S. (2025) Using Deep Feature Distances for Evaluating the Perceptual Quality of MR Image Reconstructions. Magnetic Resonance in Medicine, 94(1), 317–330. doi:10.1002/mrm.30437.