CRUNet-MR-Univ: A Foundation Model for Diverse Cardiac MRI Reconstruction

Donghang Lyu[1], Marius Staring[1], Hildo J. Lamb[1], and Mariya Doneva[2]

[1] Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands
d.lyu@lumc.nl
[2] Philips Innovative Technologies, Hamburg, Germany

Abstract. In recent years, deep learning has attracted increasing attention in the field of Cardiac MRI (CMR) reconstruction due to its superior performance over traditional methods, particularly in handling higher acceleration factors, highlighting its potential for real-world clinical applications. However, current deep learning methods remain limited in generalizability. CMR scans exhibit wide variability in image contrast, sampling patterns, scanner vendors, anatomical structures, and disease types. Most existing models are designed to handle only a single or narrow subset of these variations, leading to performance degradation when faced with distribution shifts. Therefore, it is beneficial to develop a unified model capable of generalizing across diverse CMR scenarios. To this end, we propose CRUNet-MR-Univ, a foundation model that leverages spatio-temporal correlations and prompt-based priors to effectively handle the full diversity of CMR scans. Our approach consistently outperforms baseline methods across a wide range of settings, highlighting its effectiveness and promise.

Keywords: Diverse CMR Reconstruction · Foundation Model · Prompt

1 Introduction

Cardiac MRI (CMR) is widely used in clinical practice for assessing cardiovascular function, offering high-resolution images and excellent soft tissue contrast. To shorten long acquisition times and reduce breath-hold discomfort, undersampling is commonly used to accelerate scanning. CMR reconstruction then restores the image from the undersampled k-space data, which involves reducing artifacts and noise. Compared to traditional methods like Parallel Imaging [5,14] and Compressed Sensing [4], deep learning methods [15,21,19,8,20] for CMR reconstruction are gaining attention for their stronger performance at higher acceleration factors.

Despite these advancements, a lot of deep learning approaches remain constrained to specific scenarios, largely due to the use of highly specialized training data. This results in performance degradation when applied to data with different distributions, which is common in practice, where variations span image contrast, sampling trajectories, scanner vendors, anatomical structures, and disease types, etc. Given these diversities, training a separate model for each specific scenario is impractical, underscoring the need for a reconstruction foundation model that generalizes across diverse CMR settings. Since the introduction of the GPT series [16,17,3], foundation models have gained traction due to their ability to learn from large, diverse datasets and generalize across tasks. In the medical domain, numerous foundation models have emerged, such as MedSAM models [9,10] for medical image segmentation and vision-language models [6,18] for report generation and visual question answering. Although recent efforts from the CMRxRecon 2024 challenge [20] have aimed to address variability across contrast, acceleration factor, and sampling pattern, they still fall short of capturing the full diversity of real-world CMR scans.

In this work, we propose **CRUNet-MR-Univ**, a foundation model designed for diverse CMR reconstruction that combines an unrolled architecture with Convolutional Recurrent U-Net (CRUNet) model and prompt-based priors to enhance generalization. Recognizing the inherent temporal dimension in most CMR scans, our model leverages rich spatio-temporal information across the entire sequence. Unlike CRNN-MRI [15], which employs a basic convolutional recurrent design, CRUNet integrates bidirectional recurrence into a U-Net by splitting it into two unidirectional units with opposite directions, placed separately in the encoder and decoder. This enables continuous spatio-temporal feature extraction. Additionally, inspired by previous methods such as PromptMR [21], PCP-UNet [24], and UPCMR [7], we incorporate both learnable and text-based prompts to encode diverse CMR scan attributes and help improve robustness. After training and evaluating on the CMRxRecon2025 dataset[3] , which includes data from multiple medical centers, scanner vendors, field strengths, disease types, image contrasts, sampling trajectories, and acceleration factors, CRUNet-MR-Univ demonstrates stronger performance over the other baseline methods.

2 Methodology

2.1 CRUNet-MR-Univ

Overall, CRUNet-MR-Univ adopts an unrolled network design due to the iterative nature of MRI reconstruction, as illustrated in Figure 1. From a global perspective, the model takes two inputs: an undersampled multi-coil k-space and its corresponding sampling mask. Firstly, the k-space data is transformed into the image domain using an inverse Fourier transform. A coil-combined image is then generated via the Root Sum of Squares (RSS) operation across coils, which serves as the input to the following cascade block. Simultaneously, a temporally averaged autocalibration signal (ACS) region, derived based on the mask, is fed into a Sensitivity Maps Estimator (SME) block, adapted from PromptMR [21], which estimates coil sensitivity maps (CSMs). Furthermore, to promote effective information flow across cascades, a Cascaded Feature Aggregation (CFA)

[3] `https://www.synapse.org/Synapse:syn59814210/wiki/631023`

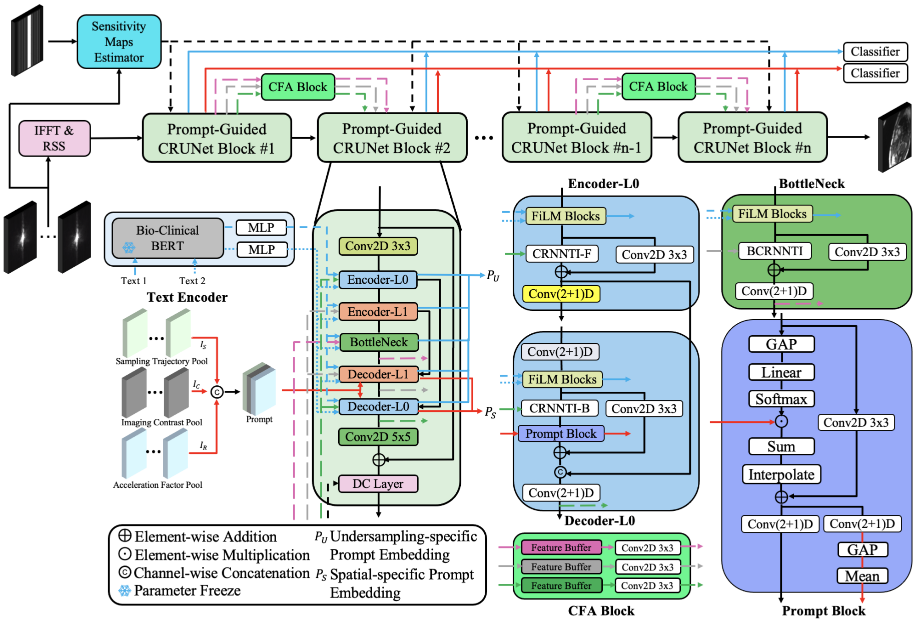

Fig. 1. Overview of the CRUNet-MR-Univ model. The bottom section details the structure of each cascade block and its core components. Each cascade block contains a CRUNet model, with pink, gray, and green long-dashed lines showing the flow of hidden state features within the CRNNTI block at each level. Blue dotted and dashed lines represent text prompt inputs, while blue and red solid lines denote the flow of output undersampling-specific and spatial-specific prompt embeddings, respectively. The black dashed lines correspond to the input of the estimated CSM.

block is introduced to aggregate all preceding feature maps to guide the convolutional recurrent modules in subsequent cascade. Following the prompt design in UPCMR [7], each cascade block also incorporates two kinds of prompts: an undersampling-specific prompt P_U and a spatial-specific prompt P_S. These prompts interact with image features to generate joint embeddings, which are concatenated across cascades and used for classifications to condition the reconstruction process. Specifically, separate MLP-based classifiers are assigned to each prompt type, predicting imaging contrast, sampling trajectory, and acceleration factor. This design encourages each prompt to better capture its associated contextual information.

Prompt-guided CRUNet Block CRUNet builds upon the CRNN-MRI [15] model, a simple yet effective network for cine MRI reconstruction that leverages the strong spatio-temporal correlations within the sequence. CRNN-MRI incorporates two types of convolutional recurrent units: Bidirectional Convolutional Recurrent Units evolving over Time and Iterations (BCRNNTI), and Convolutional Recurrent Units evolving over Iterations (CRNNI). The former enables information propagation both across the temporal sequence and between cascade blocks, while the latter focuses solely on iterative refinement across cascade blocks. However, in its original design, only a single BCRNNTI block is placed at the beginning of each cascade block, which limits the continuous extraction of spatio-temporal features throughout the cine sequence. This discontinuity can hinder the model's ability to capture internal spatio-temporal features effectively. Therefore, CRUNet enhances the architecture by splitting a BCRNNTI block into two CRNNTI blocks with forward and backward propagation directions (i.e., CRNNTI-F and CRNNTI-B) and placing them in the encoder and decoder parts of U-Net structure, respectively. Their outputs are fused via skip connections, effectively forming an enhanced BCRNNTI unit that captures temporal information from neighboring frames while integrating both low-level and high-level spatial features. As shown in Figure 1, each CRUNet follows a two-level U-Net structure with two convolutional layers at each end for adjusting the channel number, which is fixed at 64 throughout the model. The two intermediate levels share a similar design, except that Conv(2+1)D layers are used for downsampling and upsampling in the first level. A full BCRNNTI block is placed at the bottleneck to keep extracting spatio-temporal features within the sequence. Furthermore, to enhance spatial feature extraction via a larger receptive field, we apply dilation factors of (1, 2, 4) to the two levels and bottleneck of CRUNet. These are used in both the CRNNTI units and Conv(2+1)D layers.

At each level of the encoder and decoder, we incorporate two types of prompts, each providing a distinct perspective. To leverage richer prior information, we design text-based prompts consisting of two components: one encoding scanner-specific information (vendor, model, field strength), and the other capturing CMR acquisition details (contrast, sampling trajectory, acceleration factor). These are formatted using two templates: (1) *"{vendor} {model} MRI scanner at {field strength} field strength"* and (2) *"MRI scan of {contrast}, sampled using {sampling trajectory} trajectory with an acceleration factor of {acceleration factor}"*. The textual prompts are processed by a frozen Bio-Clinical BERT [1], pretrained on a large-scale clinical corpora, and subsequently refined via MLPs to match the target dimensionality. For each encoder and decoder block, the resulting prompt embeddings are integrated using two FiLM blocks [12]. Each FiLM block generates modulation parameters, weight $W_P \in \mathbb{R}^{B \times T \times C}$ and bias embeddings $B_P \in \mathbb{R}^{B \times T \times C}$, by taking the concatenation of prompt and feature embeddings as input. Here, B denotes the batch size, T the number of frames, and C the number of channels. These parameters modulate the feature representations, enabling dynamic conditioning on prior information. It is worth noting that, in the second FiLM block, W_P and B_P are summed and averaged across the temporal dimension to obtain an updated undersampling-specific prompt embedding.

In the decoder part, we introduce a PromptBlock with a learnable spatial-specific prompt $P_S \in \mathbb{R}^{3 \times C \times H \times W}$, where H and W denote the initial spatial size. P_S is formed by concatenating three learnable embeddings, each selected from a prompt pool based on prior information: sampling trajectory (I_S), imaging contrast (I_C), and acceleration factor (I_R). Following the prompt design in

PromptIR [13], a weight embedding is generated from the input feature map to modulate the first dimension of P_S via a weighted sum, effectively fusing prior information. The resulting prompt is then interpolated and added to the input feature map, enhancing it with the integrated priors. An updated spatial-specific prompt embedding is further obtained through global average pooling followed by temporal averaging.

CFA Block In the CRNNTI block, information is propagated across cascade blocks by treating the output feature map from the previous cascade as an additional input for the current one. However, as the network becomes deeper, information from earlier cascade blocks tends to highly degrade or vanish, despite potentially containing useful context for the current cascade. To address this, a Cascaded Feature Aggregation (CFA) block is introduced. Given the j-th level of the i-th cascade block, we maintain a feature buffer that stores all previous feature maps at that level, denoted as $F_{j,0}, \ldots, F_{j,i-1}$. These feature maps are concatenated along the channel dimension and passed through a convolutional layer to produce an updated feature map, effectively integrating information from all preceding cascade blocks. Notably, CRUNet comprises two levels and a bottleneck, yielding three feature buffers in the CFA block, one for each of them, with dilation factors of convolution layers matching those used in CRUNet.

2.2 Loss Function

The overall loss function is composed of two parts: reconstruction loss $\mathcal{L}_{rec}$ and classification loss $\mathcal{L}_{cls}$. The $\mathcal{L}_{cls}$ is the sum of the cross-entropy losses for the contrast class, sampling trajectory class and acceleration factor class, while the $\mathcal{L}_{rec}$ is the weighted sum of L1, MSE and SSIM loss terms, defined as follows:

$$\mathcal{L}_{rec} = \lambda_{l1} \| |I_{rec}| - |I_{gnd}| \|_1 + \lambda_{l2} \| |I_{rec}| - |I_{gnd}| \|_2^2 + \lambda_{ssim}(1 - \mathrm{SSIM}(|I_{rec}|, |I_{gnd}|)), \tag{1}$$

where I_{rec} denotes the reconstructed CMR image sequence and I_{gnd} represents the ground-truth sequence. All loss terms are computed using the absolute value of the images. We set $\lambda_{l1} = \lambda_{l2} = 0.5$ and $\lambda_{ssim} = 1$. Finally, the overall loss function represents as follows:

$$\mathcal{L} = \lambda_{cls}\mathcal{L}_{cls} + \mathcal{L}_{rec}. \tag{2}$$

Since classification serves as an auxiliary task primarily for guiding prompt tuning, which is much less important than the main reconstruction objective, we assign it a small weight of $\lambda_{cls} = 0.025$.

3 Experimental Setup

3.1 Dataset and Task Description

The CMRxRecon2025 challenge aggregates data from over 5 medical centers and more than 10 MRI scanners from GE, Philips, Siemens, and United Imaging, including 1.5T and 3.0T scans. Furthermore, the dataset spans multiple MRI modalities and sequences: bSSFP is used for cine, phase-contrast (PC), and tagging sequences; FLASH is employed for mapping and dark-blood imaging; and TSE is utilized for T2-weighted imaging. Likewise, the dataset encompasses a variety of cardiac diseases. The dataset comprises multi-parametric CMR imaging from 600 subjects, divided into 200 training, 100 validation, and 300 testing cases. Consequently, the challenge focuses on developing a foundation model that generalizes well to unseen data from different medical centers and across diverse cardiovascular diseases. In this paper, we focus on the first task, evaluating model generalization across multiple centers.

Within the training dataset, three acceleration factors (8×, 16×, 24×) and three sampling patterns (uniform Cartesian, Gaussian Cartesian, pseudo-radial) are provided. Notably, Gaussian Cartesian and pseudo-radial trajectories employ temporal interleaving, whereas uniform Cartesian does not. The ACS region comprises the central 20 lines for Cartesian trajectories and a 20×20 central area for pseudo-radial sampling.

3.2 Implementation Details

The training of CRUNet-MR-Univ was conducted in two stages. In the first stage, we verified the effectiveness of key components (i.e., the CFA block and prompt modules). In the second stage, we further unlocked the model's potential by adjusting training settings and employing a curriculum learning strategy.

The First Stage Before training, several preprocessing steps were applied. To accommodate CRUNet's requirement for temporal input, modalities without a time dimension (e.g., black-blood, T1w, T2w) were expanded into single-frame sequences. Although CRUNet supports variable frame counts, some samples with excessive frames (e.g., 54) hindered recurrent operations and highly increased computational cost. Therefore, for cases with more than 12 frames, we randomly selected 12 continuous frames for training. Furthermore, to address the imbalance in the number of training samples across modalities, we enforced uniform sampling in the data loader, ensuring roughly equal exposure per modality in each training epoch. For data normalization, we transformed the multi-coil k-space data into the image domain, normalized it by dividing by the maximum absolute value, and then converted it back to the k-space domain.

The models were implemented in PyTorch 2.0.0 and trained on an NVIDIA A100 GPU with 80GB memory. To optimize GPU usage, we employed mixed precision training [11]. To enable faster evaluation and reduce overall training time, all unrolled methods used 6 cascade blocks and were trained for 60 epochs, with 6,000 samples selected per epoch. Batch size was set to 1. We used the AdamW optimizer with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$, an initial learning rate of 2×10^{-4} and a weight decay of 0.1. The learning rate was reduced by a factor of 0.9 every two epochs, with a minimum threshold of 2×10^{-5}. In addition to CRUNet-MR-Univ, we evaluated two additional related baseline methods for comparison. The first is CRNN-MRI [15], and the second adopts the

CRUNet architecture but replaces all CRNNTI blocks with standard Conv3D blocks (i.e. UNet-MR). Notably, both models did not incorporate prompts.

The Second Stage We modified the preprocessing procedure to improve inference on longer sequences and reduce training overhead, as randomly selecting 12 consecutive frames in our earlier setup hindered reconstruction performance of longer sequences. Inspired by previous studies [21,23], we adopted a strategy of using five consecutive frames as input while focusing the reconstruction on the middle frame. For cases with fewer than five frames, model was configured to output the entire sequence directly. This approach enhances inference by reconstructing each frame through the exploitation of spatio-temporal correlations with neighboring frames, while simultaneously reducing GPU memory consumption, thereby enabling the use of more cascade blocks in CRUNet-MR-Univ during training.

Another implementation change is the training strategy: we adopted curriculum learning [2] to enable progressive, step-wise learning in CRUNet-MR-Univ. The details are as follows:

1. Initialize the model with 6 cascade blocks and train for 40 epochs with an acceleration factor of $\{8\}$.
2. Add 4 new blocks to the model (total 10 blocks), train for 40 epochs with an acceleration factor of $\{8, 16\}$, with the sampling probabilities of $\{0.2, 0.8\}$.
3. Add 2 new blocks to the model (total 12 blocks), train for 32 epochs with an acceleration factor of $\{8, 16, 24\}$, with the sampling probabilities of $\{0.1, 0.1, 0.8\}$.
4. Train the complete model for 13 epochs using all the acceleration factors with the equal sampling probability.

For the first three steps, each epoch was trained with 6000 samples, while the final step used 16000 samples. The first three steps employed a cosine-annealing scheduler with warm-up. Steps one and two used 6 warm-up epochs with learning rates of 2×10^{-4} and 1×10^{-4}, and a minimum learning rate of 1×10^{-5}. Step three used 5 warm-up epochs, a learning rate of 5×10^{-5}, and a minimum learning rate of 1×10^{-6}. In the final step, the initial learning rate was set to 8×10^{-5} and reduced by a factor of 0.4 every two epochs, then it was set to 1×10^{-7} in the last epoch.

For evaluation metrics, peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and normalized mean squared error (NMSE) were chosen, computed on the cropped central region of each validation case via submission to the challenge website.

4 Results

The organizers have evaluated several traditional methods such as SENSE, GRAPPA, and zero-filled (ZF) reconstruction. Then Table 1 summarizes the overall performance comparison, covering both baseline methods and ablation studies. Our

Table 1. Comparison of CRUNet-MR-Univ (proposed) with baseline models on the center-cropped validation set. S1 refers to the first stage, while S2 represents the second stage. Best results are shown in bold.

Methods	PSNR ↑	SSIM ↑	NMSE ↓
ZF	21.765	0.584	0.113
SENSE	23.648	0.587	0.132
GRAPPA	24.433	0.639	0.084
CRNN-MRI	25.900	0.730	0.076
UNet-MR	25.229	0.705	0.087
CRUNet-MR-Univ w/o CFA & Prompts (S1)	26.209	0.749	0.064
CRUNet-MR-Univ w/o Prompts (S1)	26.440	0.752	0.061
CRUNet-MR-Univ (S1)	26.484	0.755	0.063
CRUNet-MR-Univ (S2)	**28.232**	**0.809**	**0.04**

proposed CRUNet-MR-Univ consistently outperforms related baselines on the small center-cropped region. The ablation results further confirm the positive contributions of the CFA blocks and prompt components to overall performance. Comparing the two training stages of CRUNet-MR-Univ reveals that introducing additional cascade blocks in combination with a curriculum learning strategy provides clear benefits. Moreover, adopting the strategy of reconstructing the middle frame from five input frames enhances consistency across reconstructed sequences during inference. Finally, Table 2 reports the detailed performance of CRUNet-MR-Univ at each medical center for both training stages (S1 and S2).

Table 2. Quantitative multi-center performance evaluation of CRUNet-MR-Univ across two training stages (S1 and S2). Best results are highlighted in bold.

		CRUNet-MR-Univ (S1)		CRUNet-MR-Univ (S2)	
Center	Vendor	SSIM↑	PSNR↑	SSIM↑	PSNR↑
C001	UIH-3.0T	0.737	25.80	**0.801**	**27.86**
C002	Siemens-3.0T	0.668	23.90	**0.727**	**25.20**
	UIH-3.0T	0.727	24.86	**0.785**	**27.07**
C003	UIH-3.0T	0.783	27.50	**0.823**	**29.32**
C004	Siemens-1.5T	0.710	25.24	**0.769**	**27.01**
C005	GE-1.5T	0.808	28.62	**0.855**	**30.21**
	Siemens-3.0T	0.758	26.42	**0.820**	**28.23**
C006	Siemens-3.0T	0.759	27.02	**0.816**	**28.55**
	UIH-3.0T	0.814	27.84	**0.861**	**29.71**
C008	GE-1.5T	0.788	27.63	**0.837**	**29.14**
Overall Mean		0.755	26.48	**0.809**	**28.23**

Figure 2 presents qualitative results of CRUNet-MR-Univ (S2) on randomly selected validation cases, covering multiple contrasts at an acceleration factor of 24 across three sampling trajectories. Overall, the reconstructed results demon-

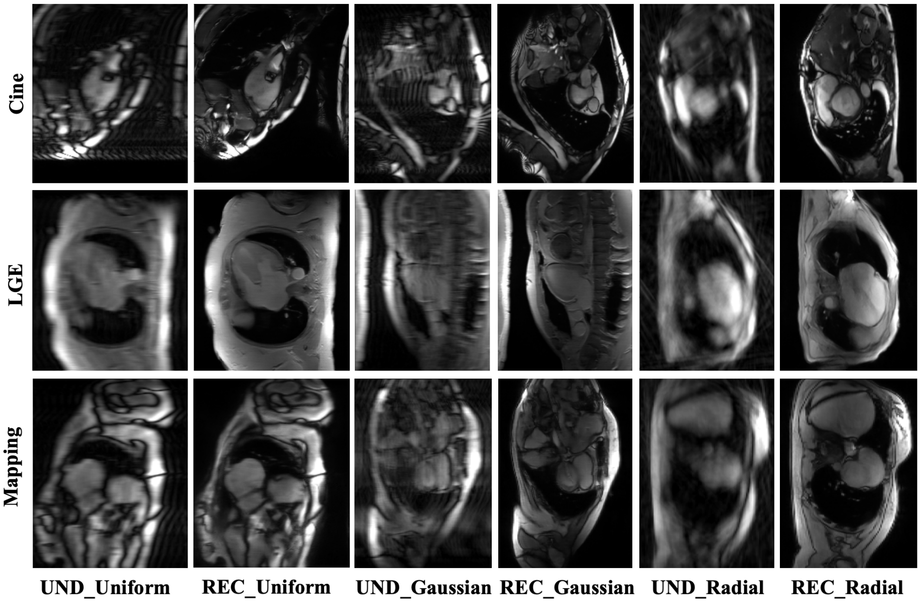

Fig. 2. Visualizations of CRUNet-MR-Univ (proposed, S2) reconstruction results for three contrasts under three k-space trajectories at an acceleration factor of 24. Here, these cases are from validation set and lack ground truth references. 'REC' indicates the reconstructed images, and 'UND' is the original undersampled inputs.

strate effective suppression of aliasing artifacts and significant reduction of blurriness. However, for some certain cases, fine cardiac structures remain partially blurred or insufficiently detailed, indicating that the current model still has room for improvement at high acceleration factor.

5 Discussion and Conclusion

As shown in Table 1, 2 and Figure 2, CRUNet-MR-Univ outperforms baseline methods under identical conditions, effectively removing artifacts and recovering fine details across diverse validation scenarios. Introduced components and training strategies have also been shown to positively impact overall reconstruction performance. However, the current performance of CRUNet-MR-Univ on cropped cardiac regions still lags behind the top-ranked methods on the leaderboard.

Although CRUNet-MR-Univ is trained with additional cascade blocks and more epochs in the second stage, potential limitations in the training process may still affect its performance. Our initial thought is to train the model with more epochs and dynamic learning rate, achieved by setting a relatively small number of training samples per epoch. However, as noted in previous studies [21,22,23], extensive epochs are not strictly necessary; rather, ensuring a sufficient number of training samples in each epoch is more important, which aligns with the training principles of foundation models. Another limitation might be the current loss function, which ignores k-space frequency information and relies solely on global image magnitude. Moreover, different imaging contrasts exhibit distinct characteristics, which may also require tailored combinations of loss terms. Designing an improved loss function could allow the model to account for a broader range of reconstruction characteristics. Then, although we propose an effective combination of CRNN operations and the U-Net structure to better exploit strong spatio-temporal correlations for reconstruction, the model may still be limited at high acceleration factors due to the restricted receptive field inherent in convolutional operations. In contrast, operations such as applying channel attention along the temporal-channel dimension may offer greater benefits at high acceleration factors by leveraging spatio-temporal features within a global receptive field. Therefore, these aspects can be explored in future studies to assess their impact on the overall performance of CRUNet-MR-Univ.

In this work, we propose CRUNet-MR-Univ, a foundation model for CMR reconstruction across diverse conditions. By integrating CRUNet modules, CFA blocks, prompt-based priors, and further refining training process, CRUNet-MR-Univ achieves strong performance and generalization, including on data from unseen medical centers. While some limitations still remain in the current training approach, the model offers strong potential for further improvement.

Acknowledgement. This work is part of the project ROBUST: Trustworthy AI-based Systems for Sustainable Growth with project number KICH3.LTP.20.006, which is (partly) financed by the Dutch Research Council (NWO), Philips Research, and the Dutch Ministry of Economic Affairs and Climate Policy (EZK) under the program LTP KIC 2020-2023.

References

1. Alsentzer, E., Murphy, J.R., Boag, W., Weng, W.H., Jin, D., Naumann, T., McDermott, M.: Publicly available clinical bert embeddings. arXiv preprint arXiv:1904.03323 (2019)
2. Bengio, Y., Louradour, J., Collobert, R., Weston, J.: Curriculum learning. In: Proceedings of the 26th annual international conference on machine learning. pp. 41–48 (2009)
3. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al.: Language models are few-shot learners. Advances in neural information processing systems **33**, 1877–1901 (2020)
4. Donoho, D.L.: Compressed sensing. IEEE Transactions on information theory **52**(4), 1289–1306 (2006)
5. Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A.: Generalized autocalibrating partially parallel acquisitions (grappa). Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **47**(6), 1202–1210 (2002)
6. Li, C., Wong, C., Zhang, S., Usuyama, N., Liu, H., Yang, J., Naumann, T., Poon, H., Gao, J.: Llava-med: Training a large language-and-vision assistant for

biomedicine in one day. Advances in Neural Information Processing Systems **36**, 28541–28564 (2023)

7. Lyu, D., Rao, C., Staring, M., van Osch, M.J., Doneva, M., Lamb, H.J., Pezzotti, N.: Upcmr: A universal prompt-guided model for random sampling cardiac mri reconstruction. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 453–463. Springer (2024)
8. Lyu, J., Qin, C., Wang, S., Wang, F., Li, Y., Wang, Z., Guo, K., Ouyang, C., Tänzer, M., Liu, M., et al.: The state-of-the-art in cardiac mri reconstruction: Results of the cmrxrecon challenge in miccai 2023. Medical Image Analysis **101**, 103485 (2025)
9. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)
10. Ma, J., Li, F., Kim, S., Asakereh, R., Le, B.H., Nguyen-Vu, D.K., Pfefferle, A., Wei, M., Gao, R., Lyu, D., et al.: Efficient medsams: Segment anything in medical images on laptop. arXiv preprint arXiv:2412.16085 (2024)
11. Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., Ginsburg, B., Houston, M., Kuchaiev, O., Venkatesh, G., et al.: Mixed precision training. ICLR (2018)
12. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
13. Potlapalli, V., Zamir, S.W., Khan, S.H., Shahbaz Khan, F.: Promptir: Prompting for all-in-one image restoration. Advances in Neural Information Processing Systems **36**, 71275–71293 (2023)
14. Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P.: Sense: sensitivity encoding for fast mri. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **42**(5), 952–962 (1999)
15. Qin, C., Schlemper, J., Caballero, J., Price, A.N., Hajnal, J.V., Rueckert, D.: Convolutional recurrent neural networks for dynamic mr image reconstruction. IEEE transactions on medical imaging **38**(1), 280–290 (2018)
16. Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al.: Improving language understanding by generative pre-training (2018)
17. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al.: Language models are unsupervised multitask learners. OpenAI blog **1**(8), 9 (2019)
18. Said, E.T., Soufiane, A.E.A., Jamal, E.T.: Medifics: Model calling enhanced vlm for medical vqa. In: 2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS). pp. 1–6. IEEE (2024)
19. Schlemper, J., Caballero, J., Hajnal, J.V., Price, A.N., Rueckert, D.: A deep cascade of convolutional neural networks for dynamic mr image reconstruction. IEEE transactions on Medical Imaging **37**(2), 491–503 (2017)
20. Wang, F., Wang, Z., Li, Y., Lyu, J., Qin, C., Wang, S., Guo, K., Sun, M., Huang, M., Zhang, H., et al.: Towards universal learning-based model for cardiac image reconstruction: Summary of the cmrxrecon2024 challenge. arXiv preprint arXiv:2503.03971 (2025)
21. Xin, B., Ye, M., Axel, L., Metaxas, D.N.: Fill the k-space and refine the image: Prompting for dynamic and multi-contrast mri reconstruction. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 261–273. Springer (2023)
22. Xin, B., Ye, M., Axel, L., Metaxas, D.N.: Enhanced deep unrolled models applied to the cmrxrecon2024 challenge. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 289–300. Springer (2024)

23. Xu, R., Özer, C., Oksuz, I.: Hypercmr: Enhanced multi-contrast cmr reconstruction with eagle loss. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 152–163. Springer (2024)
24. Zhang, C., Loecher, M., Alkan, C., Yurt, M., Vasanawala, S.S., Ennis, D.B.: On the foundation model for cardiac mri reconstruction. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 226–235. Springer (2024)