Abstract #02502 ISMRM ABSTRACT

Cross-Cascade Feature Aggregation for Improved Spatio-Temporal Reconstruction in Cardiac Cine MRI

Primary: Acquisition & Reconstruction - Image Reconstruction: AI 〉 **Secondary:** Acquisition & Reconstruction - Subsampling: Parallel imaging, compressed sensing and low-rank modeling 〉 **Digital Poster** · 60 min | Deep Learning Meets k-space: New Frontiers in Reconstruction · Wednesday, May 13 at 02:40 PM 〉 **Keywords:** CARDIAC CINE MRI RECONSTRUCTION CROSS-CASCADE FEATURE AGGREGATION FEATURE BANK CONVOLUTIONAL RECURRENT NEURAL NETWORK SPATIOTEMPORAL INFORMATION

Donghang Lyu [1], **Marius Staring**[1], **Yiming Dong**[2], **Jochen Keupp**[3], **Hildo J Lamb**[4], **Mariya Doneva**[3]

[1]Department of Radiology, Leiden University Medical Center, Leiden, Netherlands
[2]Department of Radiology, C.J. Gorter MRI Center, Leiden University Medical Center, Netherlands
[3]Philips Innovative Technologies, Hamburg, Germany
[4]Department of Radiology, Cardio Vascular Imaging Group, Leiden University Medical Center, Leiden, Netherlands

Presenting Author: *Donghang Lyu (d.lyu@lumc.nl)*

Impact

This study explores how integrating features from multiple preceding cascade blocks in unrolled networks can improve cine MRI reconstruction, potentially enabling more accurate and temporally consistent cardiac imaging.

Synopsis

Motivation: Cascaded deep learning methods have attracted attention for their ability to further push the undersampling in cine MRI by better representations of the spatio-temporal correlations in the data.

Goals: To investigate the benefits of integrating features from multiple previous cascades, as earlier stages may contain complementary information that enhances the spatio-temporal representation of the current block.

Approach: A feature bank is introduced to store features from all preceding blocks, enabling their aggregation to strengthen the representation of the current block.

Results: Incorporating the proposed module further enhances the model's ability to exploit spatio-temporal information, leading to improved performance.

Introduction

Cardiac cine MRI plays a pivotal role in capturing the heart's dynamic motion, offering valuable insights into cardiac function and aiding the diagnosis of diverse cardiovascular diseases [1] . However, the prolonged acquisition time remains a major limitation, and undersampling is commonly applied to accelerate scans and improve patient comfort.

Deep learning–based reconstruction methods have increasingly been used to obtain high quality images from undersampled data by using neural networks as learned regularizers. Prior studies [2,3] have shown that effectively extracting spatio-temporal features in cine sequences can highly enhance the reconstruction. However, information in unrolled networks typically propagate only between adjacent blocks. Aggregating features from multiple preceding cascade blocks could leverage spatio-temporal information more effectively, as earlier stages may retain valuable cues to improve the current block's representations.

In this work, we introduce a feature bank that stores intermediate features and reuses them via a channel attention module. This approach builds on the CRUNet-MR model, which enhances spatio-temporal feature utilization for cine MRI reconstruction beyond conventional UNet [4] models. Overall, aggregating features from the feature bank improves reconstruction performance without substantially changing the model structure and increasing reconstruction time, demonstrating the benefits of leveraging features from previous blocks to better capture spatio-temporal information.

Method

CRUNet-MR is an unrolled network designed for cardiac cine MRI reconstruction. Within each CRUNet cascade block, convolutional recurrent operations are integrated into the UNet architecture. A bidirectional convolutional recurrent unit is decomposed into two unidirectional units with opposite information propagation directions, placed in the encoder and decoder, respectively. This design enables continuous extraction of temporal features across the sequence while effectively combining low- and high-level spatial representations, resulting in outputs with more powerful spatio-temporal representations.

Given i-th cascade block of CRUNet-MR, the overall process can be described as follows:

$$x_{i+1} = \mathrm{CRUNet}_{\theta,i}(H_i, F_i, x_i, y, S, M),$$

where x_i denotes the coil-combined image input to the i-th cascade block, y is the original multi-coil under-sampled k-space data, M the sampling mask, and S the coil sensitivity map. H_i denotes hidden states, while F_i refers to the intermediate feature maps from previous cascades. Then instead of using only the preceding feature maps, a feature bank stores all prior outputs. These are concatenated and refined via a channel attention module to form a comprehensive feature map integrating information from all cascades. The process is formulated as:

$$F_{i,j} = \mathrm{CA}_{\theta,j}\big(\mathrm{concat}(F_{0,j}, \ldots, F_{i-1,j})\big).$$

Here, $\mathrm{CA}_{\theta,j}$ represents the channel attention module for the j-th layer. The network architecture and further details are illustrated in Fig.1.

We used the public CMRxRecon2023 [5] dataset, which was acquired on a 3T MRI scanner (MAGNETOM Vida, Siemens Healthineers, Germany) from 300 healthy volunteers (120 training, 60 validation, 120 testing). Cardiac cine data were collected using a TrueFISP [6] readout with retrospective ECG-gated segmented acquisition. The dataset includes short-axis (SAX, 5~14 slices) and long-axis (LAX, 1 slice) views. Typical parameters: spatial resolution 1.5×1.5 mm², slice thickness 8 mm, TR/TE = 3.6/1.6 ms, and FOV= 340×300 mm² (LAX) or 340×340 mm² (SAX). Data were acquired during breath-holds (2 for LAX, 11 for SAX), using uniform Cartesian sampling beyond the 24-line ACS region, with consistent undersampling masks (R=4, 8) and a simulated R=24 mask following the same strategy.

The loss function combines k-space MSE with image-domain L1/MSE and SSIM terms. All deep learning models employ a consistent cascade number of 5. Training was performed for 144 epochs with a batch size of 1, using AdamW optimizer (lr=0.0003) and a cosine-annealing scheduler with 10 warm-up epochs (minimum lr=0.0001). We tested two feature bank fusion modules: a convolution layer and a channel attention module. Except traditional PSNR and SSIM metrics, DISTS [7] and HaarPSI [8] are employed to provide a more comprehensive evaluation from different perceptual perspectives. Image reconstruction was performed with the proposed method as well as the following methods: low rank plus sparse (L+S) [9], CRNN-MRI [2], PromptMR [3] and CRUNet-MR.

Results

Fig.1 shows the quantitative comparison with other benchmarks, while Fig.2 presents the qualitative results. Fig.3 provides a dynamic analysis of spatio-temporal feature exploitation, and Fig.4 illustrates the effect of varying the number of preceding cascade blocks for feature aggregation.

Discussion and conclusion

From the quantitative and qualitative evaluations, the proposed feature aggregation module further enhances performance without highly increasing the reconstruction time, especially at higher acceleration factors. It also facilitates the recovery of some cardiac details that are missed by the base model (CRUNet-MR). The ablation study further confirms the contribution of information aggregated from all preceding blocks. Moreover, this framework is general and can be extended to other unrolled networks in future work.

Acknowledgements

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. Kramer, C. M., Barkhausen, J., Bucciarelli-Ducci, C., Flamm, S. D., Kim, R. J., & Nagel, E. (2020). Standardized cardiovascular magnetic resonance imaging (CMR) protocols: 2020 update. Journal of Cardiovascular Magnetic Resonance, 22(1), 17.

2. Qin, C., Schlemper, J., Caballero, J., Price, A. N., Hajnal, J. V., & Rueckert, D. (2018). Convolutional recurrent neural networks for dynamic MR image reconstruction. IEEE transactions on medical imaging, 38(1), 280-290.

3. Xin, B., Ye, M., Axel, L., & Metaxas, D. N. (2023, October). 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). Cham: Springer Nature Switzerland.

4. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer; 2015. p 234–241.

5. Wang, C., Lyu, J., Wang, S., Qin, C., Guo, K., Zhang, X., ... & Qu, X. (2024). CMRxRecon: A publicly available k-space dataset and benchmark to advance deep learning for cardiac MRI. Scientific Data, 11(1), 687.

6. Oppelt, A., Graumann, R., Barfuss, H., Fischer, H., Hartl, W., & Schajor, W. (1986). FISP—a new fast MRI sequence. Electromedica, 54(1), 15-18.

7. Ding, K., Ma, K., Wang, S., & Simoncelli, E. P. (2020). Image quality assessment: Unifying structure and texture similarity. IEEE transactions on pattern analysis and machine intelligence, 44(5), 2567-2581.

8. Reisenhofer, R., Bosse, S., Kutyniok, G., & Wiegand, T. (2018). A Haar wavelet-based perceptual similarity index for image quality assessment. Signal Processing: Image Communication, 61, 33-43.

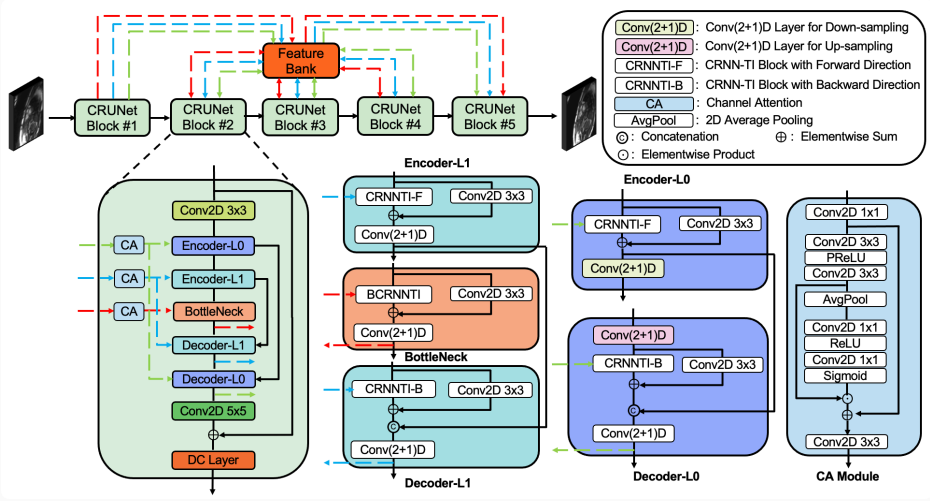
9. Otazo, R., Candes, E., & Sodickson, D. K. (2015). Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. Magnetic resonance in medicine, 73(3), 1125-1136.

Figures and Tables

View	R	Models	PSNR (dB) ↑	SSIM ↑	DISTS ↑	HaarPSI ↑	Inference Time (s) ↓
Multi-Coil LAX	4×	L+S	34.83 ± 3.73†	0.893 ± 0.065†	0.875 ± 0.038†	0.758 ± 0.115†	**0.411 ± 0.035**
		CRNN-MRI	44.17 ± 2.77†	0.983 ± 0.009†	0.954 ± 0.014†	0.953 ± 0.029†	0.915 ±0.159
		PromptMR	41.54 ± 4.72†	0.982 ± 0.017†	0.953 ± 0.015†	0.945 ± 0.054†	2.603 ±0.253
		CRUNet-MR	45.74 ± 2.96†	0.986 ± 0.009†	0.962 ± 0.014†	0.966 ± 0.028†	1.247 ±0.162
		CRUNet-MR w/ Conv	46.28 ± 2.97	**0.988 ± 0.008**†	0.965 ± 0.013†	0.970 ± 0.026	1.299 ±0.168
		CRUNet-MR w/ CA	**46.44 ± 2.98**	**0.988 ± 0.008**	**0.966 ± 0.013**	**0.971 ± 0.025**	1.484 ±0.167
	8×	L+S	29.38 ± 2.80†	0.806 ± 0.068†	0.812 ± 0.030†	0.575 ± 0.096†	**0.407 ± 0.034**
		CRNN-MRI	39.33 ± 2.42†	0.965 ± 0.013†	0.916 ± 0.013†	0.893 ± 0.044†	0.913 ±0.121
		PromptMR	38.92 ± 3.78†	0.970 ± 0.021†	0.924 ± 0.015†	0.903 ± 0.063†	2.599 ±0.254
		CRUNet-MR	40.97 ± 2.83†	0.971 ± 0.014†	0.929 ± 0.015†	0.918 ± 0.052†	1.245 ±0.142
		CRUNet-MR w/ Conv	41.51 ± 2.77†	0.973 ± 0.013†	0.934 ± 0.014	0.928 ± 0.047†	1.284 ±0.144
		CRUNet-MR w/ CA	**41.64 ± 2.86**	**0.974 ± 0.013**	**0.935 ± 0.015**	**0.930 ± 0.047**	1.486 ±0.171
	24×	L+S	27.62 ± 2.62†	0.789 ± 0.067†	0.792 ± 0.028†	0.514 ± 0.092†	**0.407 ± 0.034**
		CRNN-MRI	34.20 ± 2.00†	0.932 ± 0.019†	0.867 ± 0.013†	0.773 ± 0.055†	0.913 ±0.139
		PromptMR	34.80 ± 2.67†	**0.948 ± 0.023**	0.886 ± 0.014†	0.814 ± 0.069†	2.589 ±0.251
		CRUNet-MR	35.51 ± 2.32†	0.943 ± 0.020†	0.888 ± 0.014†	0.818 ± 0.063†	1.241 ±0.147
		CRUNet-MR w/ Conv	35.90 ± 2.29†	0.946 ± 0.019†	0.893 ± 0.013	0.828 ± 0.061†	1.291 ±0.157
		CRUNet-MR w/ CA	**35.99 ± 2.29**	0.947 ± 0.019	**0.894 ± 0.013**	**0.833 ± 0.061**	1.486 ±0.171
Multi-Coil SAX	4×	L+S	35.65 ± 3.81†	0.906 ± 0.058†	0.878 ± 0.039†	0.779 ± 0.113†	**0.560 ± 0.098**
		CRNN-MRI	44.48 ± 2.95†	0.983 ± 0.009†	0.954 ± 0.014†	0.952 ± 0.030†	1.266 ±0.221
		PromptMR	40.64 ± 5.05†	0.981 ± 0.020†	0.953 ± 0.018†	0.937 ± 0.062†	3.448 ±0.464
		CRUNet-MR	46.17 ± 3.10†	0.987 ± 0.008†	0.964 ± 0.013†	0.967 ± 0.025†	1.704 ± 0.279
		CRUNet-MR w/ Conv	46.64 ± 3.15†	**0.988 ± 0.008**†	**0.967 ± 0.013**†	0.971 ± 0.024†	1.756 ±0.281
		CRUNet-MR w/ CA	**46.73 ± 3.05**	**0.988 ± 0.007**	**0.967 ± 0.012**	**0.972 ± 0.022**	2.030 ±0.321
	8×	L+S	30.53 ± 2.72†	0.829 ± 0.062†	0.817 ± 0.032†	0.607 ± 0.094†	**0.560 ± 0.098**
		CRNN-MRI	39.97 ± 2.49†	0.965 ± 0.014†	0.916 ± 0.013†	0.892 ± 0.045†	1.249 ±0.213
		PromptMR	38.97 ± 4.27†	0.972 ± 0.022†	0.924 ± 0.017†	0.904 ± 0.067†	3.450 ±0.466
		CRUNet-MR	41.75 ± 2.86†	0.973 ± 0.013†	0.932 ± 0.013†	0.923 ± 0.045†	1.700 ±0.278
		CRUNet-MR w/ Conv	42.28 ± 2.95†	0.974 ± 0.013†	0.936 ± 0.014†	0.931 ± 0.044†	1.752 ±0.282
		CRUNet-MR w/ CA	**42.43 ± 2.83**	**0.975 ± 0.012**	**0.937 ± 0.013**	**0.934 ± 0.040**	2.027 ±0.326
	24×	L+S	28.40 ± 2.35†	0.807 ± 0.060†	0.797 ± 0.030†	0.527 ± 0.080†	**0.560 ± 0.098**
		CRNN-MRI	35.03 ± 2.00†	0.933 ± 0.019†	0.870 ± 0.013†	0.773 ± 0.050†	1.254 ±0.220
		PromptMR	35.38 ± 2.95†	**0.950 ± 0.024**	0.887 ± 0.015†	0.815 ± 0.066†	3.441 ±0.465
		CRUNet-MR	36.46 ± 2.14†	0.945 ± 0.018†	0.892 ± 0.012†	0.819 ± 0.052†	1.697 ±0.276
		CRUNet-MR w/ Conv	36.85 ± 2.21†	0.947 ± 0.018†	0.896 ± 0.012†	0.831 ± 0.052†	1.751 ±0.281
		CRUNet-MR w/ CA	**37.21 ± 2.19**	**0.950 ± 0.018**	**0.898 ± 0.013**	**0.842 ± 0.050**	2.026 ±0.319

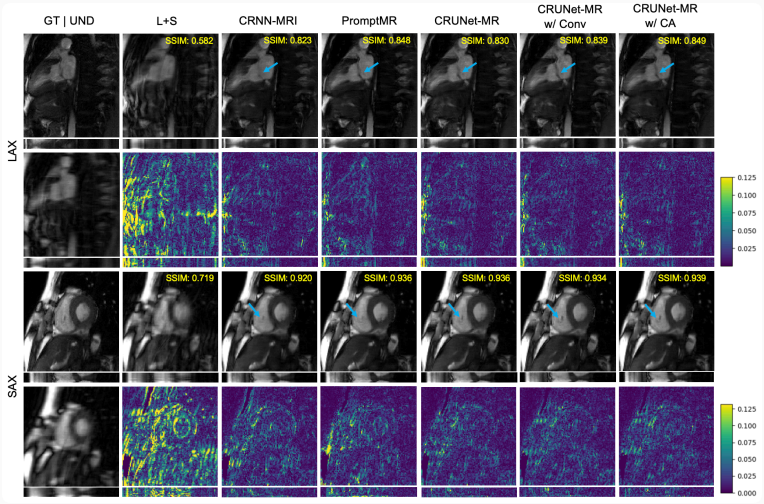
Scan for high-resolution version

Figure 1: *Quantitative comparison with benchmark models in terms of evaluation metrics and per-slice inference time, using the same hardware (NVIDIA RTX A6000, 48 GB). Best results are in bold. Two feature aggregation modules are tested: a convolutional layer (Conv) and a channel attention module (CA). The Wilcoxon signed-rank test with Bonferroni correction is applied; † ($p < 0.0125$) indicates statistical significance in pairwise comparisons against the proposed CRUNet-MR with the CA module.*



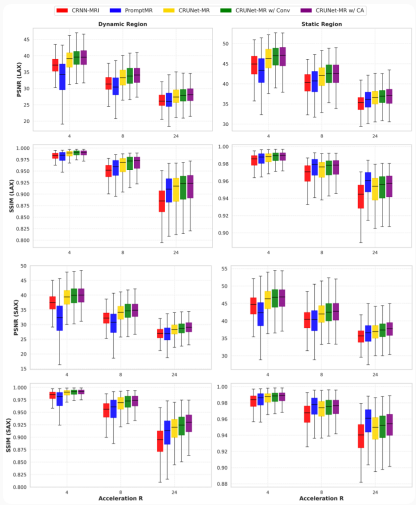
Scan for high-resolution version

Figure 2: *Architecture of the CRUNet-MR model integrating a feature bank and a feature aggregation module. Each CRUNet block adopts a three-level UNet structure with two convolutional recurrent units at each level, operating in forward (CRNNTI-F) and backward (CRNNTI-B) directions. Green, blue, and red dashed lines indicate the feature map flow at each level. Treating each level as a unit, three CA modules are employed to aggregate feature maps.*



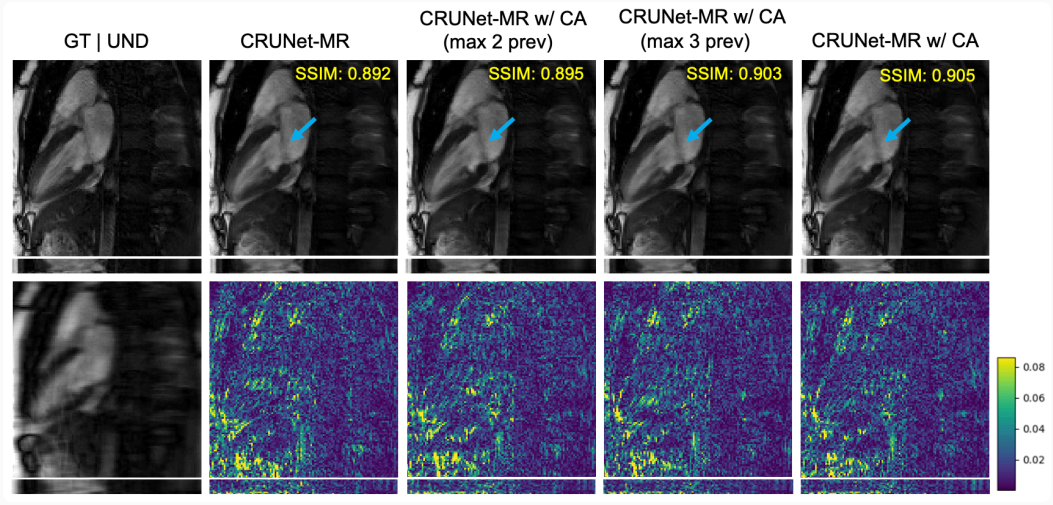
Scan for high-resolution version

Figure 3: *Qualitative evaluation on LAX and SAX views at R=8. Frames are cropped to highlight cardiac details, with SSIM (top-right) reflecting performance on the cropped cine sequence. GT and UND denote ground truth and undersampled images, respectively. The x–t view of the central line is shown at the bottom of each image. For models that aggregate features from preceding blocks, this design enhances detail recovery and sharpens features (blue arrows), further improving CRUNet-MR's performance.*



Scan for high-resolution version

Figure 4: *Dynamic analysis of test data across two views. By thresholding the temporal standard deviation, cine sequences are split into dynamic (mainly the heart) and static regions. Overall, CRUNet-MR already outperforms baselines on dynamic regions, demonstrating effective extraction of spatio-temporal information. Then feature aggregation from previous blocks further enhances performance in both regions, indicating improved spatio-temporal feature utilization.*



Scan for high-resolution version

Figure 5: *Qualitative evaluation of the ablation study on varying the number of preceding cascade blocks for feature aggregation on LAX view at R=8. The third and fourth columns correspond to aggregating up to 2 or 3 preceding features. Based on the SSIM values and visual quality in the cardiac region (indicated by blue arrows), aggregating features from all cascade blocks, including distant ones, further enhances reconstruction quality and recovery of fine cardiac details.*