CardioMamba: Multi-Frame Temporally Coherent Cine MRI Super-Resolution via State-Space Modeling

Simin Mirzaei*, Maharshi Panchal, Him Wai Ng, Howard Yin, Nasim Hajati, and Panos Nasiopoulos

Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada
siminmirzaei@ece.ubc.ca, mp2001@student.ubc.ca, michael.ng@ubc.ca, howard33@student.ubc.ca, nasim.hajati@ubc.ca, panosn@ece.ubc.ca

Abstract. Dynamic 2D+t cine cardiac MRI is central to the assessment of ventricular function and myocardial motion, yet acquisition constraints such as limited breath-hold duration and accelerated imaging often limit spatial resolution. Learning-based super-resolution (SR) can enhance spatial detail after acquisition; however, most methods process frames independently or rely on pairwise temporal referencing, which can lead to inter-frame flicker and boundary instability. We propose CardioMamba, a temporally coherent SR model that leverages a StereoMamba-inspired state-space architecture with correspondence-aware fusion to jointly exploit information from preceding, center, and following low-resolution frames for accurate reconstruction of the central high-resolution frame. We train and evaluate our model on the public Sunnybrook Cardiac Data (SCD), comparing it with single-frame SR, two-frame SR, and video SR methods. Results show that CardioMamba achieves superior performance at $\times 2$ and $\times 3$, with the strongest reconstruction quality, competitive temporal consistency, and substantially fewer parameters than most competing methods. At $\times 4$, our model remains competitive despite the increased task difficulty while maintaining computational efficiency. CardioMamba provides a lightweight and scalable solution for cardiac MRI SR, offering an efficient trade-off among visual quality, temporal consistency, and computational cost for potential clinical use.

Keywords: Cardiac Cine MRI · Super-Resolution · CardioMamba · State-Space Models · Temporal Consistency · Deep Learning.

1 Introduction

Cine cardiac MRI is the clinical gold standard for assessing ventricular function, ejection fraction, and myocardial strain [1]. However, its spatial resolution is constrained by breath-hold duration, SNR trade-offs, and accelerated acquisition methods [2–6]. In dynamic 2D+t cine MRI, spatial and temporal resolution are

* Corresponding author.

tightly coupled, meaning that faster acquisitions often reduce spatial detail [7–10]. Although existing reconstruction methods can mitigate these limitations, they may amplify noise and artifacts, reducing the reliability of ejection fraction estimates [11–14]. Therefore, recovering spatial detail from low-resolution (LR) data while preserving temporal consistency remains challenging.

Learning-based Super-Resolution (SR) has emerged as a promising post-processing solution for recovering high-frequency details without extending the acquisition time. Early Single-Image SR (SISR) methods achieve strong reconstruction performance on individual frames, with EDSR leveraging deep residual learning [15] and RCAN further improving feature representation through channel attention [16]. Diffusion-based approaches such as Res-SRDiff improve reconstruction fidelity through residual diffusion refinement [17], but process individual slices independently and do not explicitly model temporal relationships. More recently, Mamba-based restoration models have enabled long-range spatial modeling. MambaIRv2 uses selective state-space modeling for non-local feature interactions [18], but remains frame-wise and does not exploit temporal relationships. These approaches therefore do not explicitly model temporal dependencies across cine frames, motivating Video SR (VSR) methods that incorporate information from neighboring frames. For example, RBPN [19] uses previously reconstructed frames through iterative back-projection and feature refinement to propagate temporal information, while BasicVSR propagates features bidirectionally using optical-flow-guided alignment and aggregation [20]. However, reliance on frame alignment and previously reconstructed frames can be problematic for complex, non-rigid myocardial motion. A complementary line of research focuses on feature sharing between correlated inputs. StereoMamba combines Visual State-Space Modules (VSSMs) with a Stereo Bidirectional Cross-Attention Module (SBCAM) to enable symmetric information exchange between paired inputs [21]. Similarly, CVHSSR alternates hierarchical intra-view feature extraction with cross-view interaction to strengthen correspondence between two input streams [22]. Applied to adjacent cine frames, these methods allow neighboring frames to inform one another but remain limited to pairwise interactions.

Despite recent advances, existing SR methods remain suboptimal for cine cardiac MRI. Frame-wise approaches ignore temporal dependencies, flow-based methods struggle with complex cardiac motion, diffusion models do not ensure temporal consistency, and pairwise fusion methods cannot fully exploit available temporal context. Moreover, no existing approach combines Mamba-based long-range spatial modeling with multi-frame temporal fusion for cine MRI. Addressing this gap requires a framework that efficiently recovers spatial detail while preserving temporal coherence.

In this paper, we propose CardioMamba, a temporally coherent SR framework for cine cardiac MRI. CardioMamba jointly processes a temporal triplet consisting of the preceding, center, and following LR frames through an early-fusion strategy and enriches the resulting representation using Mamba-based residual state-space blocks. The network reconstructs only the center High-Resolution (HR) frame while exploiting complementary information from both

temporal directions. We evaluate CardioMamba on the Sunnybrook Cardiac Data (SCD) [23] under different upscaling factors and compare it against representative SISR, pairwise fusion, and VSR approaches. Performance is assessed using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Temporal Consistency Error (TCE), and parameter count. Experiments demonstrate that CardioMamba achieves strong overall performance across all upscaling factors, consistently balancing reconstruction quality, temporal consistency, and parameter efficiency.

The remainder of this paper is organized as follows: Section 2 presents the proposed architecture, Section 3 evaluates the results, and Section 4 concludes the paper.

2 Methodology

In this section, we present the proposed CardioMamba architecture and its main components. We also describe the architectural variants evaluated to study the impact of different temporal fusion strategies.

2.1 Dataset Preparation

All experiments are conducted on the publicly available SCD dataset [23], which contains short-axis 2D cine cardiac MRI sequences spanning the full cardiac cycle. Each scan is treated as a 2D+t spatiotemporal sequence. The original images serve as HR targets, while LR inputs are generated using bicubic and Lanczos interpolation. For temporal modeling, a sliding window of three consecutive frames forms training triplets (x_{i-1}, x_i, x_{i+1}). Dataset splitting is performed at the subject level to prevent patient overlap between training and testing sets.

2.2 Proposed Architecture and Loss Regime

The primary training objective is the pixel-wise Charbonnier loss function, which is mathematically equivalent to a smoothed L_1 loss. To investigate the impact of the number of input frames and the fusion strategy, we evaluate four architectural variants. Variant 1 adapts the pairwise SR architecture of StereoMamba [21] to process two adjacent frames using temporal feature-level fusion. Variant 2 uses the same pairwise setting but introduces early fusion, in which the two input frames are concatenated before feature extraction. Variant 3 extends StereoMamba to enable tri-temporal fusion across three frames using temporal feature-level fusion. Variant 4 extends the early-fusion strategy to the three-frame input triplet by concatenating the frames before feature extraction. Variant 4 corresponds to the proposed CardioMamba architecture, illustrated in Fig. 1. As shown, CardioMamba takes a temporal triplet of LR frames, which are concatenated and processed by a shared shallow convolutional encoder. The resulting unified features are refined by a StereoMamba-inspired temporal SB-CAM module and a stack of Residual State-Space Groups (RSSGs). Finally, a

sub-pixel convolution layer reconstructs the HR center frame. The main components are described below:

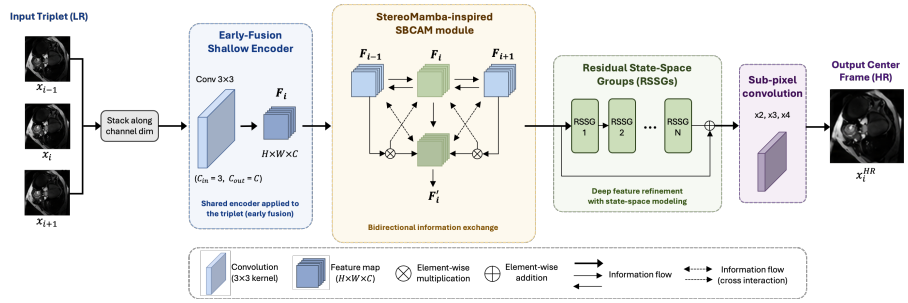

Fig. 1. Overall architecture of CardioMamba (Variant 4). A triplet of consecutive low-resolution frames is early fused by a shared shallow encoder. The SBCAM module is subsequently applied to the fused representation, followed by RSSGs for feature refinement and sub-pixel convolution to reconstruct the high-resolution center frame.

Early-Fusion Shallow Encoder A shallow convolutional encoder is used to extract low-level feature representations from the input frames. For Variant 1, the encoder is applied independently to each frame in the input pair, producing shallow features F_{i-1} and F_i. For Variant 2, an early-fusion strategy first concatenates the two input frames along the channel dimension, after which they are jointly processed by a shared convolutional encoder to produce a unified pairwise feature representation. For Variant 3, the encoder is applied independently to each frame in the input triplet, producing F_{i-1}, F_i, and F_{i+1}. In contrast, Variant 4 employs an early-fusion strategy in which the three neighboring frames are first concatenated along the channel dimension and then jointly processed by a shared convolutional encoder. This early fusion enables the encoder to extract a unified feature representation from the combined information of all three frames. The resulting fused representation is subsequently fed into the SBCAM block for further refinement.

Temporal Cross-Attention Module We adapt the SBCAM design from StereoMamba to a temporal setting, enabling interaction between the center-frame features and their neighboring temporal context. In Variant 1, SBCAM is applied to the two frame representations in a pairwise configuration. Variant 2 applies SBCAM to the jointly encoded representation produced by early fusion. In Variant 3, the center-frame feature interacts with both neighboring frame features, while direct interaction between the neighboring frames is omitted based on preliminary experiments. In Variant 4 (CardioMamba), SBCAM further refines the unified three-frame representation produced by early fusion, enabling selective integration of the available temporal context.

Residual State-Space Refinement The temporally enriched feature representations in Variants 2 and 4 (CardioMamba) are refined using a stack of RSSGs, adapted from the StereoMamba architecture [21]. These modules enhance feature representations while capturing long-range spatial dependencies through structured state-space modeling. In Variants 1 and 3, the feature representation from each input frame is processed independently using its own stack of RSSGs, consistent with the feature-level fusion design. RSSGs enable efficient global context aggregation while maintaining linear computational complexity.

Upsampling and Reconstruction Finally, a sub-pixel upsampling and reconstruction head generates the HR output of the center frame at scale factors of $\times 2$, $\times 3$, or $\times 4$ for Variants 2 and 4 (CardioMamba). In Variants 1 and 3, the reconstruction head produces two and three adjacent HR outputs corresponding to the input frames, respectively.

3 Experimental Results and Evaluation

This section presents a comprehensive evaluation of the proposed CardioMamba framework through ablation studies, qualitative and quantitative assessments, as well as statistical analyses.

3.1 Ablation Study

We conducted an ablation study by varying the number of input frames and the fusion strategy. Variants 1 and 2 use two input frames with feature-level and early fusion, respectively, while Variants 3 and 4 use three input frames with feature-level and early fusion, respectively. In addition, MambaIRv2 is included as a single-frame baseline to assess the benefit of temporal information. This design isolates the effects of fusion strategy and temporal context.

As shown in Table 1, Variant 4 (CardioMamba) achieves the highest PSNR and SSIM across all scale factors, reaching 40.180 dB/0.978, 35.460 dB/0.945, and 29.634 dB/0.875 at $\times 2$, $\times 3$, and $\times 4$, respectively. The results show that increasing the temporal context from one to two frames and from two to three frames improves both PSNR and SSIM, as observed by the consistent improvement from MambaIRv2 to the pairwise variants and from the pairwise to the three-frame variants. Furthermore, adopting early fusion improves performance in both the pairwise and three-frame settings, as demonstrated by the improvements of Variant 2 over Variant 1 and Variant 4 over Variant 3, respectively. These results demonstrate the benefits of incorporating additional temporal information and early fusion for MRI super-resolution.

3.2 Qualitative Comparison with State-of-the-Art Methods

Qualitative comparisons of CardioMamba with RCAN, the best-performing existing non-Mamba method, and StereoMamba, the best Mamba-based method,

Table 1. Ablation study of the number of input frames and fusion strategy at different upsampling scales. Best results are shown in bold.

Architectures	**PSNR/SSIM**		
	×2	×3	×4
MambaIRv2 (single frame) [18]	31.134/0.910	27.818/0.851	25.919/0.802
1 (pairwise feature-level fusion)	37.911/0.880	32.753/0.864	28.341/0.803
2 (pairwise early fusion)	37.998/0.931	32.993/0.922	28.837/0.867
3 (three-frame feature-level fusion)	38.220/0.963	34.550/0.926	28.957/0.872
4 (Ours: three-frame early fusion)	**40.180/0.978**	**35.460/0.945**	**29.634/0.875**

at a scale factor of ×2 are shown in Fig. 2. For visual assessment, we selected a representative short-axis cardiac MRI frame containing fine anatomical structures and intensity variations, providing a challenging case for evaluating detail preservation and structural continuity. As shown in Fig. 2, differences between the reconstructed images can be difficult to notice visually, particularly when the overall reconstruction quality is high. Therefore, enlarged views are provided to better highlight subtle differences. In particular, the thin white structure indicated by the yellow arrows appears partially missing and over-smoothed in the competing methods, whereas CardioMamba better preserves this structure and its local sharpness, resulting in a more faithful reconstruction. Preservation of such fine anatomical structures is important for assessing cardiac morphology, and the improved structural fidelity of CardioMamba provides encouraging evidence of its clinical relevance.

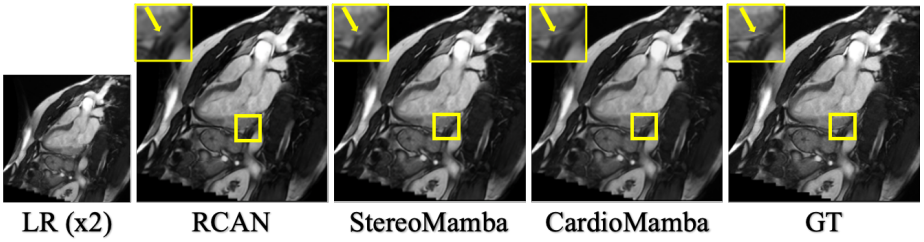

Fig. 2. Qualitative comparison of CardioMamba with the best-performing state-of-the-art methods at ×2 upsampling.

3.3 Quantitative Comparison with State-of-the-Art Methods

The quantitative comparison of CardioMamba with state-of-the-art methods is presented in Table 2 in terms of PSNR, SSIM, TCE, and model complexity. TCE measures temporal consistency error by quantifying discrepancies in temporal variations between consecutive reconstructed MRI frames and their corresponding reference frames; lower values indicate better temporal consistency.

At ×2, CardioMamba achieves the highest PSNR (40.180 ± 2.593 dB) and SSIM (0.978±0.009), while maintaining a low TCE of 0.995±0.002. Although CVHSSR achieves a lower TCE (0.853 ± 0.351), CardioMamba exhibits a substantially lower standard deviation, indicating more consistent temporal reconstruction. CardioMamba also achieves these results with only 15.175M parameters, substantially fewer than EDSR and comparable to RCAN and StereoMamba.

At ×3, CardioMamba achieves the highest PSNR (35.460 ± 2.330 dB) and SSIM (0.945±0.019), while also obtaining the lowest TCE (1.250±0.002) among the evaluated methods. At ×4, CardioMamba remains competitive, achieving higher PSNR and SSIM than RCAN, MambaIRv2, and StereoMamba, although CVHSSR and BasicVSR achieve higher PSNR and SSIM. CardioMamba also achieves a competitive TCE of 2.729 ± 0.003 while maintaining approximately 15.3M parameters, demonstrating a favorable balance between reconstruction quality, temporal consistency, and model complexity.

Table 2. Quantitative comparison of CardioMamba and state-of-the-art methods at ×2, ×3, and ×4 upsampling scales in terms of PSNR, SSIM, and Temporal Consistency Error (TCE), reported as mean ± standard deviation, along with model complexity in terms of the number of parameters (M).

Model	Scale	PSNR ± std (dB)	SSIM ± std	TCE ± std	Params (M)
EDSR [15]	×2	37.455 ± 2.421	0.920 ± 0.026	1.159 ± 0.502	40.720
RCAN [16]	×2	39.116 ± 2.352	0.956 ± 0.026	1.155 ± 0.477	15.445
RBPN [19]	×2	–	–	–	–
BasicVSR [20]	×2	–	–	–	–
CVHSSR [22]	×2	38.735 ± 2.460	0.959 ± 0.029	**0.853 ± 0.351**	2.216
MambaIRv2 [18]	×2	31.134 ± 3.981	0.910 ± 0.061	4.119 ± 1.997	22.970
StereoMamba [21]	×2	37.911 ± 2.872	0.880 ± 0.094	1.275 ± 0.617	15.531
CardioMamba	×2	**40.180 ± 2.593**	**0.978 ± 0.009**	0.995 ± 0.002	15.175
EDSR [15]	×3	32.529 ± 2.612	0.890 ± 0.096	1.303 ± 0.491	43.671
RCAN [16]	×3	34.761 ± 2.424	0.944 ± 0.019	1.317 ± 0.608	15.629
RBPN [19]	×3	33.172 ± 2.102	0.909 ± 0.083	3.820 ± 1.042	11.338
BasicVSR [20]	×3	33.226 ± 2.337	0.911 ± 0.092	1.392 ± 0.547	6.325
CVHSSR [22]	×3	–	–	–	–
MambaIRv2 [18]	×3	27.818 ± 3.501	0.851 ± 0.078	5.302 ± 2.561	23.150
StereoMamba [21]	×3	32.753 ± 2.376	0.864 ± 0.089	1.703 ± 0.759	15.745
CardioMamba	×3	**35.460 ± 2.330**	**0.945 ± 0.019**	**1.250 ± 0.002**	15.360
EDSR [15]	×4	30.094 ± 2.103	0.860 ± 0.084	1.386 ± 0.528	43.081
RCAN [16]	×4	26.350 ± 3.496	0.818 ± 0.086	6.025 ± 3.078	15.592
RBPN [19]	×4	31.429 ± 2.092	0.881 ± 0.074	1.381 ± 0.933	12.751
BasicVSR [20]	×4	31.742 ± 2.193	0.880 ± 0.043	1.499 ± 0.784	6.291
CVHSSR [22]	×4	**32.450 ± 1.801**	**0.907 ± 0.028**	**0.936 ± 0.403**	2.237
MambaIRv2 [18]	×4	25.919 ± 3.499	0.802 ± 0.092	9.876 ± 3.318	23.120
StereoMamba [21]	×4	28.341 ± 2.725	0.803 ± 0.074	3.759 ± 0.802	15.812
CardioMamba	×4	29.634 ± 1.997	0.875 ± 0.041	2.729 ± 0.003	15.320

3.4 Statistical Analysis

The statistical significance analysis in Fig. 3 examines pairwise differences between CardioMamba and competing methods at ×2 for PSNR, SSIM, and TCE.

Two-sided Welch's t-tests are used, with Holm correction applied for multiple comparisons and adjusted $p < 0.05$ considered statistically significant. CardioMamba shows statistically significant differences from all competing methods across all three metrics ($p < 10^{-4}$). The significantly higher PSNR and SSIM values indicate improved reconstruction fidelity and structural preservation, while the significant differences in TCE demonstrate superior temporal consistency.

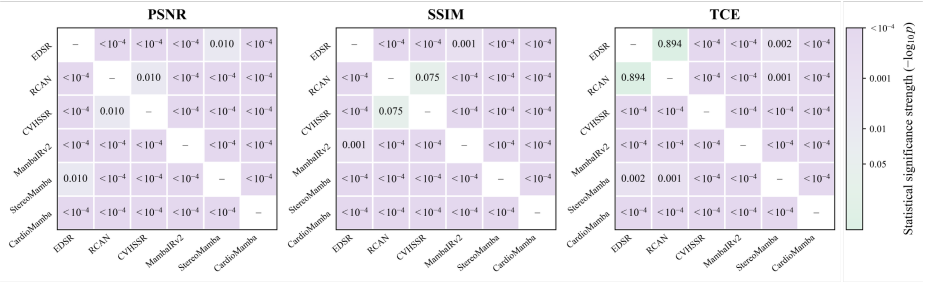

Fig. 3. Statistical significance analysis of CardioMamba against competing methods for PSNR, SSIM, and TCE. Values of $p < 0.05$ indicate statistically significant differences.

4 Conclusion

In this work, we proposed CardioMamba, a lightweight and temporally coherent super-resolution framework for dynamic 2D+t cine cardiac MRI. CardioMamba employs an early-fusion strategy to jointly process a temporal triplet of low-resolution frames through a shared shallow encoder, producing a unified representation that is further refined by Mamba-based residual state-space groups. The network reconstructs the center high-resolution frame while leveraging spatiotemporal information from both temporal directions. Unlike existing methods, CardioMamba explicitly exploits multi-frame temporal dependencies while maintaining computational efficiency. Experimental results demonstrate that CardioMamba achieves the best PSNR and SSIM at $\times 2$ and $\times 3$, with strong temporal consistency and substantially fewer parameters than most competing methods. At $\times 4$, despite severe degradation, CardioMamba remains competitive in reconstruction quality while maintaining favorable model complexity.

Future work will extend CardioMamba to jointly perform spatial and temporal upsampling, using its spatiotemporal representations to not only enhance the spatial resolution of cardiac MRI frames but also synthesize intermediate or missing frames for improved temporal resolution. In addition, we will investigate the impact of CardioMamba super-resolution on ventricular segmentation and functional parameter estimation to assess its clinical utility.

Disclosure of Interests. The authors declare that they have no competing interests.

References

1. Seetharam, K., Lerakis, S.: Cardiac magnetic resonance imaging: the future is bright. F1000Research **8**, 1636 (2019)
2. Petersen, S.E., Aung, N., Sanghvi, M.M., Zemrak, F., Fung, K., Paiva, J.M., Francis, J.M., Khanji, M.Y., Lukaschuk, E., Lee, A.M., Carapella, V.: Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort. Journal of Cardiovascular Magnetic Resonance **19**(1), 18 (2017)
3. Pennell, D.J.: Cardiovascular magnetic resonance. Circulation **121**(5), 692–705 (2010)
4. Schulz-Menger, J., Bluemke, D.A., Bremerich, J., Flamm, S.D., Fogel, M.A., Friedrich, M.G., Kim, R.J., von Knobelsdorff-Brenkenhoff, F., Kramer, C.M., Pennell, D.J., Plein, S.: Standardized image interpretation and post processing in cardiovascular magnetic resonance: Society for Cardiovascular Magnetic Resonance (SCMR) board of trustees task force on standardized post processing. Journal of Cardiovascular Magnetic Resonance **15**(1), 35 (2013)
5. Kramer, C.M., Barkhausen, J., Bucciarelli-Ducci, C., Flamm, S.D., Kim, R.J., Nagel, E.: Standardized cardiovascular magnetic resonance imaging (CMR) protocols: 2020 update. Journal of Cardiovascular Magnetic Resonance **22**(1), 17 (2020)
6. Saeed, M., Van, T.A., Krug, R., Hetts, S.W., Wilson, M.W.: Cardiac MR imaging: current status and future direction. Cardiovascular Diagnosis and Therapy **5**(4), 290 (2015)
7. Pruessmann, K.P.: Parallel imaging at high field strength: synergies and joint potential. Topics in Magnetic Resonance Imaging **15**(4), 237–244 (2004)
8. 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 **47**(6), 1202–1210 (2002)
9. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: The application of compressed sensing for rapid MR imaging. Magnetic Resonance in Medicine **58**(6), 1182–1195 (2007)
10. Oscanoa, J.A., Middione, M.J., Alkan, C., Yurt, M., Loecher, M., Vasanawala, S.S., Ennis, D.B.: Deep learning-based reconstruction for cardiac MRI: a review. Bioengineering **10**(3), 334 (2023)
11. Wang, N., Liu, M., Wang, W., Zhao, Y., Li, J., Ren, X., Yu, D., Liu, A., Hou, X., Song, Q.: Accelerated Reduced Field of View T2-Weighted Imaging of Pancreaticobiliary Disorders Using Deep Learning-Based Reconstruction: Reduction of Acquisition Time and Improvement of Image Quality. Canadian Association of Radiologists Journal (2026)
12. 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. Magnetic Resonance in Medicine **79**(6), 3055–3071 (2018)
13. Aggarwal, H.K., Mani, M.P., Jacob, M.: MoDL-MUSSELS: model-based deep learning for multishot sensitivity-encoded diffusion MRI. IEEE Transactions on Medical Imaging **39**(4), 1268–1277 (2019)
14. Yaman, B., Hosseini, S.A.H., Moeller, S., Ellermann, J., Uğurbil, K., Akçakaya, M.: Self-supervised physics-based deep learning MRI reconstruction without fully-sampled data. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 921–925. IEEE (2020)

15. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1132–1140. IEEE (2017)
16. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: European Conference on Computer Vision (ECCV), pp. 294–310. Springer, Cham (2018)
17. Safari, M., Wang, S., Eidex, Z., Li, Q., Qiu, R.L.J., Middlebrooks, E.H., Yu, D.S., Yang, X.: MRI super-resolution reconstruction using efficient diffusion probabilistic model with residual shifting. Physics in Medicine & Biology **70**(12), 125008 (2025)
18. Guo, H., Guo, Y., Zha, Y., Zhang, Y., Li, W., Dai, T., Xia, S.T., Li, Y.: MambaIRv2: Attentive state space restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 28124–28133 (2025)
19. Haris, M., Shakhnarovich, G., Ukita, N.: Recurrent back-projection network for video super-resolution. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3892–3901. IEEE (2019)
20. Chan, K.C.K., Wang, X., Yu, K., Dong, C., Loy, C.C.: BasicVSR: the search for essential components in video super-resolution and beyond. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4945–4954. IEEE (2021)
21. Ma, Z., Tohidypour, H.R., Nasiopoulos, P., Leung, V.C.: StereoMamba: Enhancing stereo image super-resolution with structured state space models and bi-directional cross attention. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1–5 (2025)
22. Zou, W., Gao, H., Chen, L., Zhang, Y., Jiang, M., Yu, Z., Tan, M.: Cross-view hierarchy network for stereo image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1396–1405 (2023)
23. Cardiac Atlas Project: Sunnybrook Cardiac Data (2009). Available at: https://www.cardiacatlas.org/sunnybrook-cardiac-data/, last accessed 2026/06/12