This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Supplementary Material
MoSAIC: Motion and Supervision-Aware Inference for Myocardial Scar and Edema Segmentation from Multi-Sequence and Cine CMR

Shucheng Liu[1], Mingcheng Hu[2], Liangqiao Gui[3], Yineng Chen[3], Peng Huang[4], Hui Guo[5], Shu Hu[6], Xin Wang[3], and Hongtu Zhu[7(✉)]

[1] Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[2] Department of Statistics, The Chinese University of Hong Kong, Hong Kong SAR, China
[3] Department of Epidemiology and Biostatistics, University at Albany, State University of New York, Albany, NY, USA
[4] School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, China
[5] Department of Mathematics, University at Albany, State University of New York, Albany, NY, USA
[6] School of Applied and Creative Computing, Purdue University, West Lafayette, IN, USA
[7] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
htzhu@email.unc.edu

This document collects the derivations referenced from the main paper: the thin plate spline alignment field (Sec. S1), the gated fusion rule across modalities (Sec. S2), the auxiliary training objectives (Sec. S3), the component ablation study (Sec. S4), and an extended discussion (Sec. S5).

S1 Thin Plate Spline Alignment Field

With target control points $\{p_k\}$ on a regular $g \times g$ lattice spanning $[-r, r]^2$ and predicted offsets $\theta_k^{(m)}$, the thin plate spline [1] is the interpolant of $\{p_k \mapsto \theta_k^{(m)}\}$ with minimal bending energy:

$$\mathcal{T}_\theta(q) \;=\; a_0 + A\,q \;+\; \sum_{k=1}^{K} w_k\, U(\|q - p_k\|)\,, \qquad U(d) = d^2 \log d, \tag{1}$$

where (w, a_0, A) solve the standard TPS linear system. That system depends only on $\{p_k\}$, which are fixed, so it is inverted once and cached. The final linear layer of the localization head $h^{(m)}$ is initialized to zero with its bias set to the target lattice, giving a stable starting warp that is updated by the task loss.

S2 Gated Fusion Across Modalities

The Stage 2a decoder fuses the aligned pyramids at level ℓ as

$$a_\ell = \tfrac{1}{2}\left[\sigma\Big(\hat{f}_\ell^{(\mathrm{C0})}\Big) + \sigma\Big(\hat{f}_\ell^{(\mathrm{T2})}\Big)\right], \tag{2}$$

$$d_\ell = \Phi_\ell\Big(\Big[\,\mathrm{up}(d_{\ell+1}),\;\; f_\ell^{(\mathrm{LGE})} \odot a_\ell,\;\; \sigma\Big(\hat{f}_\ell^{(\mathrm{C0})}\Big),\;\; \sigma\Big(\hat{f}_\ell^{(\mathrm{T2})}\Big)\Big]\Big), \tag{3}$$

initialized from the concatenated bottlenecks $d_4 = [f_4^{(\mathrm{LGE})}, \hat{f}_4^{(\mathrm{C0})}, \hat{f}_4^{(\mathrm{T2})}]$. When an auxiliary sequence is absent, its masked feature pyramid is zero, so its sigmoid gate channel is fixed at $\sigma(0) = \frac{1}{2}$. This is a consequence of the gate parameterization, not an identity gate; the absent sequence contributes no feature map derived from images. The Stage 1 prior gates every decoder level, $d_\ell \leftarrow d_\ell \odot \sigma(\mathrm{pool}(P))$.

S3 Auxiliary Training Objectives

Myocardium consistency across sequences. Three auxiliary heads $g^{(m)}$ decode a myocardium mask from each sequence's own features. For the available sequences $A = \{m : m_m = 1\}$, each head is supervised against the myocardium label and every available pair in $\mathcal{Q} = \{(m, m') : m < m',\, m, m' \in A\}$ is driven into agreement:

$$\mathcal{L}_{\mathrm{cons}} = \frac{\mathcal{L}_{\mathrm{sup}} + \mathcal{L}_{\mathrm{pair}}}{N}, \qquad N = |A| + |\mathcal{Q}|. \tag{4}$$

Here $\mathcal{L}_{\mathrm{sup}}$ sums the myocardium Dice terms specific to each sequence, and $\mathcal{L}_{\mathrm{pair}}$ sums the pairwise Dice agreement terms.

Stage 2b edema expert. The reference pyramid is formed as

$$e_\ell^{\mathrm{ref}} = \max\big(E_\ell^{\mathrm{LGE}}([\mathbf{I}_{\mathrm{LGE}}, M]),\; E_\ell^{\mathrm{C0}}([\mathbf{I}_{\mathrm{C0}}, M])\big), \qquad e_\ell^{\mathrm{T2}} = E_\ell^{\mathrm{T2}}([\mathbf{I}_{\mathrm{T2}}, M]), \tag{5}$$

with M a dilated cardiac mask from Stage 1.

S4 Ablation Study

All tested removals reduce the mean scar Dice, although the magnitude is not uniform. The largest drops occur when removing the spatial prior gate (-0.032) or the Stage 1 anatomy cascade (-0.030). We therefore interpret the gain as cumulative rather than being driven by a single module. For CineMyoPS, replacing the multi-phase input with a single ED frame while keeping the remaining settings unchanged decreases scar Dice from 0.457 to 0.420.

Table 2 varies the sequences available to the multisequence branch. The full LGE+C0+T2 setting improves over LGE alone by 0.055 Dice, and the largest incremental gain comes from adding T2 to LGE+C0 (+0.048). For the cine branch, where LGE is not available by definition, the multi-phase model was compared with an ED-only control while keeping the remaining settings unchanged; temporal aggregation improves cine scar Dice from 0.420 ± 0.035 to 0.457 ± 0.038.

Table 1. Component ablation on scar segmentation, five-fold Dice (mean ± s.d.). Every variant is retrained from scratch on the same folds. Δ is reported relative to the full framework of the same cohort.

Variant	Scar Dice ↑	Δ
MyoPS (multisequence)		
MoSAIC (full)	$0.668_{\pm 0.022}$	—
w/o spatial prior gate	$0.636_{\pm 0.024}$	−0.032
w/o Stage 1 anatomy cascade	$0.638_{\pm 0.029}$	−0.030
w/o cross-sequence consistency	$0.650_{\pm 0.028}$	−0.018
early fusion (single shared encoder)	$0.652_{\pm 0.026}$	−0.016
w/o supervision masking	$0.658_{\pm 0.015}$	−0.010
CineMyoPS (cine)		
MoSAIC (full, multi-phase)	$0.457_{\pm 0.038}$	—
single ED frame (T=1)	$0.420_{\pm 0.035}$	−0.037

Table 2. MyoPS scar Dice under reduced sequence availability, using the same five-fold splits (mean ± s.d.).

Sequences available	Scar Dice ↑
LGE + C0 + T2 (full)	$0.668_{\pm 0.022}$
LGE + C0	$0.620_{\pm 0.028}$
LGE only	$0.613_{\pm 0.032}$

S5 Extended Discussion

Why edema lags the comparator. MoSAIC improves scar but not edema relative to nnU-Net (0.5983 versus 0.6691). Two properties of this cohort plausibly contribute. First, edema supervision is sparse and center-limited: only 80 of the 220 multisequence training cases carry edema annotations, and all of them come from Centers B and C (Table 1 of the main paper), so the edema expert sees roughly a third of the cases and none of the acquisition variation present in the remaining centers. Second, pure edema is not predicted directly but recovered as $Z \setminus \widehat{\text{scar}}$, so scar errors propagate into the edema map, whereas a single-stage baseline predicts the two classes jointly. We therefore read the present results as evidence that MoSAIC is effective for scar under heterogeneous supervision, and that its edema pathway does not yet justify its additional structure at this supervision density.

Generalization of the cine branch. Internal five-fold evaluation holds out patients from the same two CineMyoPS centers, whereas official validation scores a separate cohort; the drop from 0.457 to 0.2069 therefore measures cross-distribution transfer rather than held-out performance. With only 64 cine cases from two centers, the branch has limited exposure to inter-center variation, although the present experiments do not isolate this as the sole cause. The ED-only ablation (Suppl. S4) shows that temporal aggregation helps inside the training

distribution (0.457 versus 0.420); it does not establish that the motion representation transfers across scanners and frame rates. Cross-center validation is the necessary next step before the cine branch can be recommended.

Cost of the multi-stage design. MoSAIC introduces several components for a comparatively small cohort, and the benefit is not uniform: official-validation Dice improves for MyoPS scar (0.6258 → 0.6965) and CineMyoPS scar (0.1816 → 0.2069) but decreases for edema. No single ablated component accounts for the scar gain on its own (Suppl. S4), which is consistent with a cumulative effect but also means the design cannot be reduced to one decisive mechanism on this data. Establishing which components remain necessary at larger scale requires more data and more controlled comparisons than a challenge-sized cohort and a workshop-length paper allow.

References

1. Bookstein, F.L.: Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence **11**(6), 567–585 (1989)