Beyond Cavity Constraints: A Systematic Evaluation of Spatial Priors in Cascaded Left Atrial Scar Segmentation

Hao Wen[1][0000−0002−8856−0883] and Jingsu Kang[2(✉)][0000−0003−4263−1495]

[1] College of Science, China Agricultural University, Beijing, China
[2] Tianjin Medical University, Tianjin, China
kangjingsu@tmu.edu.cn

Abstract. Automatic left atrial (LA) scar segmentation from late gadolinium enhancement MRI (LGE-MRI) is challenging due to extreme class imbalance, thin atrial walls, and multi-center domain shift. Prior work commonly adopts a two-phase pipeline where scar predictions are spatially constrained using cavity-derived information. This paper systematically evaluates whether such constraints are beneficial on the Comprehensive Analysis of Real-world medical images (CARE) 2026 Left Atrium track. We find that both hard and soft cavity-derived constraints are detrimental: hard dilation clipping suppresses valid scar predictions, while soft signed distance map (SDM) inputs improve training performance but generalize poorly. A properly sized 6-stage U-Net achieves near-optimal performance, with further increases in model complexity bringing only marginal gains. Beyond spatial constraints, a Gaussian spatial weighting loss derived from scar ground truth improves segmentation, whereas a cavity-wall-derived weighting of identical form brings no benefit, indicating that the benefit of a spatial prior depends on what anatomical structure it is derived from. Built on the nnU-Net framework, our method achieves competitive results on the CARE 2026 benchmark using only the challenge-provided training data, without external datasets or pre-trained models.

Keywords: Left atrial scar segmentation · LGE-MRI · Spatial constraints · Gaussian spatial weighting · nnU-Net.

1 Introduction

Late gadolinium enhancement magnetic resonance imaging (LGE-MRI) enables non-invasive visualization of left atrial (LA) scar tissue, the fibrotic substrate that sustains atrial fibrillation (AF) and independently predicts post-ablation recurrence [9]. Automatic, accurate scar quantification from LGE-MRI is therefore essential for risk stratification and personalized ablation planning. However, this task remains extremely challenging: the atrial wall is only 1–3 mm thick, scar occupies fewer than 5% of LA voxels on average, image quality varies substantially across scanners and centers, and scar patterns are highly heterogeneous in both size and shape [8].

These challenges have been the focus of several benchmarking efforts, including the Left Atrial Scar Quantification and Segmentation (LAScarQS) 2022 and LAScarQS++ 2024 challenges [15,10,1,12,14]. The current Comprehensive Analysis of Real-world medical images (CARE) 2026 Left Atrium track [17,16,7,2] provides over 200 multi-center LGE-MRIs and 300 computed tomography (CT) volumes across four clinical centers.

A large proportion of prior methods for LA scar segmentation adopt a two-phase coarse-to-fine pipeline: a first network localizes the LA cavity, and a second network predicts scar within the cavity region. These methods share an implicit assumption that scar predictions should be spatially constrained using cavity-derived information. Such constraints manifest in diverse forms, including shape attention [7], signed distance maps (SDMs) of the cavity boundary as additional Phase 2 inputs [15], wall region-of-interest masks [10,6], and SDM-based soft priors for loss weighting [5]. Anatomically, this assumption is questionable. The LA cavity label marks the blood-pool *lumen*; scar tissue resides in the *myocardial wall* that surrounds the lumen, a structurally different compartment. The majority of ground-truth scar voxels lie outside the ground-truth LA cavity; even after 5 mm dilation of the cavity, 1.1% of scar remains outside the dilated cavity (Fig. 1). No prior work has systematically tested whether cavity-derived spatial constraints (hard or soft) are beneficial for LA scar segmentation.

We conduct a systematic evaluation of spatial constraints in cascaded LA scar segmentation, examining both hard post-processing constraints and soft training-time priors. Prior work has noted that dilation clipping may be inaccurate [15]; we quantify both hard and soft forms of this constraint. Removing hard dilation clipping improves G-DSC, suggesting that such constraints are imprecise and at times counterproductive. Soft cavity-derived constraints have a different failure pattern: augmenting Phase 2 input with a cavity signed distance map improves training G-DSC but degrades validation performance, pointing to overfitting on imperfect Phase 1 cavity predictions rather than anatomical exclusion. In contrast, a Gaussian spatial weighting loss derived from scar ground truth rather than cavity geometry improves segmentation by focusing the loss on scar-adjacent regions without imposing cavity-derived priors. A cavity-wall-derived Gaussian of identical mathematical form brings no improvement. The anatomical origin of the weighting, rather than its mathematical formulation, is what drives the improvement. We further identify a hidden pitfall for multi-center deployment: training at a single-center fixed spacing means the model may not encounter interpolation artifacts during training, yet such artifacts arise when processing images at different resolutions during testing. A mixed-spacing inference strategy addresses this without retraining. Our method builds on the nnU-Net framework [3] and validates these design choices on the CARE 2026 benchmark.

2 Methods

2.1 Dataset and Evaluation

The CARE 2026 Left Atrium track provides LGE-MRI and CT data from four clinical centers [17,16,7,2]: Center A (Utah NAMIC-CARMA), Center B (BIDMC, Boston), Center C (KCL, London), and Center D (Fuzhou Univ. Affiliated Provincial Hospital). Table 1 summarizes the dataset statistics; slash-separated values denote Task 1/Task 2 splits (for Center A) or labeled/unlabeled splits (for Center D). The challenge comprises three tasks. Task 1 (LA scar quantification) provides 60 LGE-MRIs with cavity and scar annotations from Center A for training, 10 cases for validation, and 24 cases for testing. Task 2 (LA cavity segmentation) provides 130 cavity-only LGE-MRIs from Center A plus the 60 Task 1 cases, totaling 190 training volumes; validation includes 10 cases from Center A and 10 from Center C; testing includes 14 cases from Center A, 20 from Center B, and 10 from Center C, introducing a cross-center domain shift. Task 3 (CT multi-structure segmentation) provides 150 CTs from Center D, of which only 50 are labeled, with 20 validation cases and 130 test cases.

Table 1. Data across the 4 clinical centers. Center B provides only test data; labels for validation and test are withheld by the organizers.

Center	Modality	Task	# Train	# Val	# Test	Spacing (mm)
A	LGE-MRI	1, 2	60 / 130	10 / 10	24 / 14	0.625×0.625×2.5[†]
B	LGE-MRI	2	✗	✗	20	1.4×1.4×1.4[‡]
C	LGE-MRI	2	✗	10	10	1.0×1.0×1.0[§]
D	CT	3	50 / 100	20	130	0.3–0.8 (in-plane) × 0.5 (z)[¶]

[†] 100 of 130 Task 2 training cases; the rest 30 cases have 1.0×1.0×1.0 mm spacing.
[‡] Test data only; spacing from the challenge documentation.
[§] Measured from validation data (640×640×88 voxels); differs from the official documentation (1.3×1.3×4.0 mm).
[¶] In-plane resolution varies across scans; z-spacing is 0.5 mm for all.

Task 1 uses the generalized Dice score (G-DSC) [8] as the primary metric, defined as G-DSC $= 2\sum_c w_c|P_c \cap G_c| \,/\, \sum_c w_c(|P_c| + |G_c|)$ with class weights $w_c = 1/|V_c|^2$, where $|V_c|$ is the ground-truth volume of class c. This weighting gives the rare scar class orders of magnitude more influence than background, making extreme class imbalance the central challenge. Accuracy (ACC) and sensitivity (SEN) are reported as secondary metrics. Task 2 and 3 are evaluated with the Dice similarity coefficient (DSC) and Hausdorff distance (HD), where higher DSC and lower HD reflect better segmentation quality.

2.2 Two-Phase Pipeline for LA Scar Segmentation

We adopt a coarse-to-fine two-phase pipeline for LGE-MRI, illustrated in Fig. 2. Phase 1 performs binary LA cavity segmentation, producing a cavity mask from

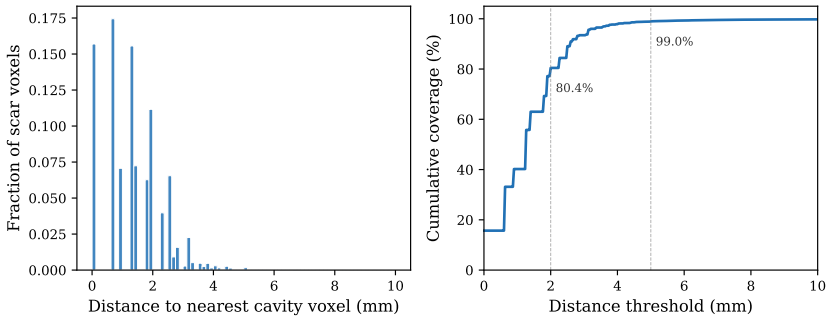

Fig. 1. Scar-to-cavity distance distribution across the 60 Task 1 training cases. **Left:** histogram of voxel-wise distances. **Right:** cumulative coverage as a function of distance threshold. 5 mm dilation covers 98.9 % of scar.

which the LA centroid is computed. A physical region of $160 \times 160 \times 110$ mm (corresponding to $256 \times 256 \times 44$ voxels at the Center A training spacing of $0.625 \times 0.625 \times 2.5$ mm) centered on the LA centroid is extracted and fed to Phase 2 for multi-class segmentation of cavity and scar. The scar class is extracted and placed back into native image space.

We compare three strategies for handling the spacing mismatch between training and inference. In the first, the input volume is resampled to the Center A training grid ($576 \times 576 \times 44$ voxels at $0.625 \times 0.625 \times 2.5$ mm) before Phase 1; both phases then operate at this fixed spacing. In the second, both phases run at the original image spacing; the Phase 2 crop extent ($160 \times 160 \times 110$ mm) is converted to native voxels by dividing by the image spacing. The third strategy, used as our default, runs Phase 1 at native resolution and Phase 2 at the Center A training spacing. This addresses a subtle issue for multi-center deployment: training Phase 2 on single-center, single-spacing data makes nnU-Net's internal resampling nearly an identity operation during training. When validation images arrive at different native spacings (e.g., 1.0 mm isotropic for Task 1 validation), a purely native Phase 2 would trigger substantial spline interpolation never encountered during optimization. Keeping Phase 1 at native resolution (where cavity segmentation is robust to spacing variation) while keeping Phase 2 at its training distribution avoids this degradation without retraining.

Phase 1 is trained on 190 LGE-MRIs (60 Task 1 + 130 Task 2) for binary cavity segmentation. Phase 2 is trained on 60 Task 1 cases for multi-class (cavity + scar) segmentation. Both phases use the 6-stage Plain U-Net backbone [3] (30.7M parameters) with feature dimensions [32, 64, 128, 256, 320, 320], instance normalization, LeakyReLU activations, and deep supervision at every decoder level. The initial encoder stride of [1,1,1] preserves the through-plane resolution. For the backbone ablation (Section 4), we also train a 6-stage Residual U-Net [3], which replaces plain convolutions with residual blocks, uses progressive stage depths (1, 3, 4, 6, 6, 6 blocks per stage; 101.7M parameters). We restrict our study to convolutional neural network (CNN)-based U-Net archi-

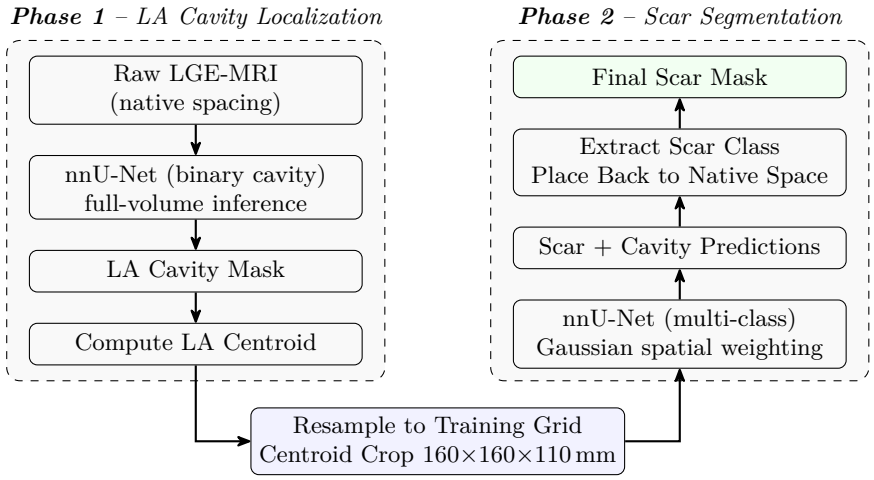

Fig. 2. Two-phase coarse-to-fine pipeline for LA scar segmentation. Phase 1 localizes the LA cavity at native resolution; Phase 2 segments scar from a centroid crop at the Center A training spacing. The crop extent corresponds to $160 \times 160 \times 110$ mm at the Center A training spacing.

tectures, as recent large-scale benchmarking has shown that properly configured CNN U-Nets within the nnU-Net framework match or exceed Transformer-based and Mamba-based approaches when validated under controlled conditions [4].

2.3 Gaussian Spatial Weighting Loss

Scar occupies fewer than 5% of LA voxels in the $256 \times 256 \times 44$ Phase 2 crop, creating extreme foreground-background imbalance. To address this, we replace the standard cross-entropy term in the nnU-Net loss (Dice + cross-entropy) with a spatially weighted cross-entropy:

$$\mathcal{L} = \mathcal{L}_{\text{Dice}} + \mathcal{L}_{\text{CE}}^{\text{weighted}}. \tag{1}$$

The voxel-wise weight map $w(\mathbf{x})$ is derived from the distance to the nearest ground-truth scar voxel: $w(\mathbf{x}) = 1 + w_0 \cdot \exp\left(-\frac{d(\mathbf{x})^2}{2\sigma^2}\right)$, where $d(\mathbf{x})$ is the Euclidean distance (in mm) from voxel $\mathbf{x}$ to the nearest scar ground-truth voxel, $w_0 = 5$, and $\sigma = 2$ mm in the in-plane direction (≈ 3.2 voxels at the Center A training resolution of 0.625 mm; applied isotropically in voxel space). The distance transform is computed efficiently on GPU via separable 3D Gaussian convolution of the scar ground-truth binary mask, avoiding CPU-based distance transform bottlenecks during training.

The key design choice is the *source* of the spatial prior. This weighting is derived from scar ground truth (the target's own distribution), not from cavity geometry. To test whether the source matters, we also evaluate a cavity-wall-derived variant: the same Gaussian form, but with $d(\mathbf{x})$ computed as the distance

to the nearest cavity-wall voxel (obtained by Gaussian blur of the cavity mask, excluding the lumen interior). This ablation isolates the effect of prior source while holding the mathematical form constant.

3 Implementation

Software. All models are built on nnU-Net v2 (2.8.0) [3] with Python 3.12 and PyTorch 2.8. Training uses stochastic gradient descent (initial learning rate 0.01, Nesterov momentum 0.99, weight decay 3×10^{-5}) with a polynomial learning rate schedule (power 0.9) for 1000 epochs, following the standard nnU-Net training recipe. The batch size is 2. Input images are normalized per volume via Z-score normalization. Many prior works on atrial segmentation adopt multidimensional contrast limited adaptive histogram equalization (MCLAHE) [11] for preprocessing [13], but we found that replacing nnU-Net's built-in normalization with MCLAHE degrades performance, and therefore retain the default approach. Patch size for Phase 2 is $32 \times 224 \times 224$ (z, y, x), with sliding-window inference (stride 0.5) at test time. Data augmentation follows nnU-Net defaults: random rotations, scaling, elastic deformations, gamma correction, and mirroring. Test-time augmentation (TTA) uses 8-fold flip averaging across all $\{x, y, z\}$ axis combinations. All results use 5-fold cross-validation ensembles.

Hardware. All experiments run on Ubuntu 22.04 (Intel Xeon Platinum 8470Q, CUDA 12.8) with a single NVIDIA GeForce RTX 5090 (32 GB) on a cloud GPU instance.

4 Experiments and Results

We evaluate on the 60 Task 1 training cases from Center A and the 10 Task 1 validation cases from the CARE 2026 Left Atrium track. Validation metrics are obtained by submitting predictions to the challenge evaluation platform, as ground-truth validation labels are not released to participants. The held-out test set is reserved for final evaluation (Section 4.3) and is not consulted during ablation studies. In all experiments, Phase 2 receives the centroid from Phase 1 predictions rather than from ground-truth cavity labels, so that training-set metrics reflect end-to-end performance under realistic conditions. Unless noted otherwise, training-set metrics use 5-fold cross-validation without TTA, and validation results include 8-fold flip TTA.

4.1 Spatial Constraints and Model Capacity

The Multi-class Bi-Atrial Segmentation (MBAS) 2024 benchmark [13] focused on cavity wall segmentation, a thin structure occupying a small fraction of the image volume, and found that lightweight CNN-based architectures remain among the strongest baselines for such tasks on limited data. As discussed in Section 2, we

restrict our study to CNN-based architectures [4]. Following these insights, we began exploration with a 4-stage U-Net (6.9M parameters). All experiments in this subsection use the default nnU-Net loss (Dice + cross-entropy); the Gaussian weighting loss is introduced in Section 4.2. Table 2 traces the exploration path. Validation metrics are obtained by submitting predictions to the challenge platform; ✗ indicates that no validation submission was made for this configuration, and this symbol is used with the same meaning throughout. The Time column reports end-to-end inference time per case (5-fold ensemble, full pipeline); see Section 3 for hardware details.

Table 2. Effect of spatial constraints and model capacity.

	Architecture	Train G-DSC	Val G-DSC	Time (s)
Baseline	4-stage U-Net	0.3463	0.2189	8
+ Cavity SDM	4-stage U-Net	0.3573	0.1882	11
+ 5 mm cavity dilation	4-stage U-Net	0.3309	✗	8
+ Model capacity	6-stage Plain U-Net	**0.6333**	**0.4743**	36
+ Cavity SDM	6-stage Plain U-Net	0.6224	✗	36

Soft constraints. We first tested whether providing cavity geometry as an additional input channel helps Phase 2. Starting from the 4-stage baseline trained with MRI and scar annotations, we added a cavity signed distance map (SDM) as a second input channel. Training-set G-DSC increased slightly ($0.3463 \rightarrow 0.3573$), but validation G-DSC dropped from 0.2189 to 0.1882. This pattern (training improvement with validation degradation) suggests overfitting: the model learns to exploit cavity geometry from the Phase 1 prediction, but Phase 1 cavity masks are imperfect, and this noise propagates to scar predictions at test time. At the final 6-stage capacity, the same SDM input reduced training G-DSC to 0.6224, indicating that this soft constraint remains counterproductive.

Hard constraints. We then tested cavity dilation as a hard post-processing constraint on the 4-stage model. Applying 5 mm dilation, which covers 98.9% of ground-truth scar voxels (Fig. 1), reduced training G-DSC from 0.3463 to 0.3309, driven primarily by a drop in sensitivity. This indicates that even a near-complete anatomical coverage threshold suppresses valid scar predictions, as expected from the observation that scar and cavity lumen are disjoint compartments.

Model capacity. Scaling from the 4-stage to the 6-stage Plain U-Net ($6.9\text{M} \rightarrow 30.7\text{M}$ parameters), a well-established strong baseline for 3D medical image segmentation [4], improved training G-DSC from 0.3463 to 0.6333 and validation G-DSC from 0.2189 to 0.4743, yielding the largest performance gain.

4.2 Loss, Backbone Refinements, and Inference

All subsequent experiments use neither cavity SDM input nor dilation-based post-processing. With the 6-stage Plain U-Net established as the base configuration, we investigated refinements: loss design, backbone variant, and inference strategy. Table 3 summarizes these results. The baseline uses Gaussian (scar) loss, multi-class (cavity + scar) training, mixed-spacing pipeline, and 5-fold ensemble; training-set metrics unless Val G-DSC is reported.

Table 3. Ablation of loss design, training target, backbone, and inference refinements.

	Experiment	Train G-DSC	Val G-DSC
Loss	Gaussian (scar) → DC+CE	0.6631 → 0.6333	**0.4791** → 0.4743
	Gaussian (scar) → Gaussian (cavity-wall)	0.6631 → 0.6629	**0.4791** → 0.4767
Target	multi-class (cavity + scar) → scar-only	0.6631 → 0.6317	✗
Backbone	Plain → Residual (mixed-spacing)	0.6631 → **0.6682**	**0.4791** → 0.4736
	Plain → Residual (same-spacing)	0.6631 → **0.6682**	0.4411 → 0.4324
Inference	mixed-spacing → same-spacing (Plain)	0.6631	**0.4791** → 0.4411
	mixed-spacing → same-spacing (Residual)	**0.6682**	0.4736 → 0.4324

Loss design. The baseline configuration uses Gaussian spatial weighting derived from scar ground truth (Equation 1), achieving training G-DSC 0.6631. Replacing this with the standard nnU-Net loss (Dice + cross-entropy, DC+CE) reduces training G-DSC from 0.6631 to 0.6333 (−4.5%) and validation G-DSC from 0.4791 to 0.4743 (−1.0%). Up-weighting scar-adjacent regions provides a clear benefit. Replacing the scar-derived weighting with a cavity-wall-derived Gaussian of identical form (same w_0 and σ, distance computed to the cavity wall instead of scar) produces a negligible change (0.6629 vs. 0.6631, −0.03%). The source of a spatial prior thus determines its utility: the scar-derived prior helps, but a cavity-derived prior of identical mathematical form does not.

Phase 2 is trained to segment both cavity and scar (multi-class), rather than scar alone. Removing the cavity target and training on scar only reduces G-DSC from 0.6631 to 0.6317 (−4.7%). A reasonable explanation is that multi-class training allows the model to learn the spatial relationship between cavity and scar implicitly through the shared representation, rather than relying on externally imposed constraints or priors.

Backbone. Replacing the 6-stage Plain U-Net with a Residual U-Net improves training G-DSC from 0.6631 to 0.6682 (+0.8%). However, this training-set gain does not translate to validation: the Residual U-Net achieves a lower validation G-DSC of 0.4736 (−1.1% vs. Plain U-Net's 0.4791) under mixed-spacing inference, and the gap persists under same-spacing inference (0.4411 → 0.4324, −2.0%). This is consistent with the finding of Isensee et al. [4] that architectural

refinements beyond a well-configured Plain U-Net yield diminishing returns in 3D medical image segmentation.

Inference strategy. Training Phase 2 at a fixed Center A spacing creates a hidden mismatch: nnU-Net's internal resampling is essentially an identity operation during training, so the model rarely encounters interpolation artifacts. At test time, images from different centers arrive at different native spacings, triggering interpolation artifacts unseen during optimization. Reverting from mixed-spacing (Phase 1 native, Phase 2 at training spacing) to same-spacing inference (both phases at native resolution) reduces validation G-DSC from 0.4791 to 0.4411 (-7.9%). The Residual U-Net shows an even larger drop under the same switch ($0.4736 \rightarrow 0.4324$, -8.7%), suggesting that increased model complexity amplifies sensitivity to spacing mismatches. This highlights a practical concern for multi-center deployment: when training data comes from a single center, the model can overfit to that center's acquisition grid, and careful handling of the train–test spacing gap is essential.

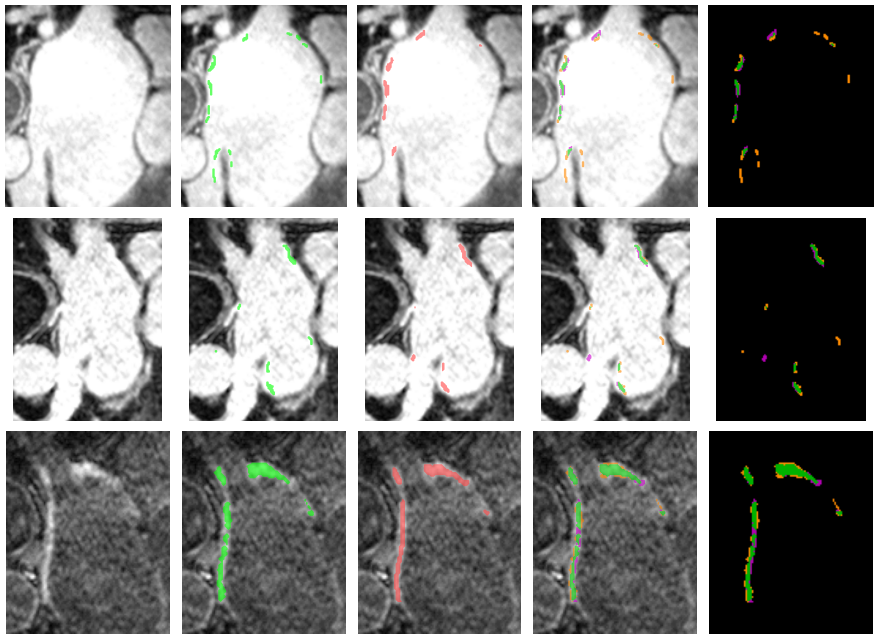

Fig. 3. Qualitative results of the baseline configuration on three training cases. Columns, left to right: **(A)** input image; **(B)** image with ground-truth scar overlay (green); **(C)** image with predicted scar overlay (red); **(D)** difference map on image; **(E)** difference map alone. Difference maps use green (true positive), purple (false positive), and orange (false negative).

We also tested 8-fold flip TTA, which provided a marginal improvement consistent with the known modest benefit of TTA for well-regularized nnU-Net models. Given the 8× inference cost, the practical value of TTA for this task is limited.

Fig. 3 shows qualitative results of the baseline configuration on three training cases. Rows from top to bottom correspond to the worst, median, and best cases by G-DSC. The worst case (top) exhibits substantial false negatives in regions with weak LGE enhancement, where scar-to-background contrast is inherently low. The best case (bottom) captures scar extent accurately. This is as expected: large contiguous scar regions are segmented reliably, whereas small scattered scar patches are harder to detect.

4.3 Test Set Results

Table 4 reports results on the official test set, evaluated by the challenge organizers. All tasks use nnU-Net with 5-fold ensemble and TTA. Task 1 uses the baseline configuration from Table 3 (mixed-spacing pipeline); Task 2 uses native-spacing inference. The test set is larger than the validation set and, for Task 2, introduces unseen centers.

Table 4. Validation and test set results.

Task	Modality	Metric	Val	Test	Δ
1 (LA scar quantification)	LGE-MRI	G-DSC	0.4791	0.4298	−0.0493
		ACC	0.7360	0.7087	−0.0273
		SEN	0.4722	0.4176	−0.0546
2 (LA cavity segmentation)	LGE-MRI	DSC	0.8871	0.8285	−0.0586
		HD (mm)	17.59	20.27	+2.68
3 (multi-structure segmentation)	CT	DSC	0.9563	0.9543	−0.0020
		HD (mm)	12.45	9.72	**−2.73**

Task 3, evaluated entirely on the same center as training (Center D), shows virtually identical performance between validation and test (DSC drop of 0.002, HD improvement of 2.73 mm), demonstrating that the nnU-Net recipe generalizes well when training and deployment data share the same acquisition characteristics. Task 2 exhibits the largest drop (DSC −0.059, HD +2.68 mm), aligning with the fact that 30 of 44 test cases (68 %) come from Centers B and C, whose scanners and acquisition protocols differ from the Center A training data. This cross-center degradation reflects the spacing-mismatch concern discussed in Section 2: the model, trained exclusively on Center A data, encounters interpolation artifacts at unseen centers that it never saw during optimization. Task 1 shows a moderate drop (G-DSC −0.049), attributable to both the larger test set (24 vs. 10 validation cases) and error propagation from Phase 1 cavity segmentation.

5 Conclusion

A systematic evaluation of spatial constraints for cascaded LA scar segmentation on the CARE 2026 Left Atrium track reveals that both hard and soft cavity-derived priors are counterproductive. The anatomical root cause is that scar occupies the myocardial wall, not the cavity lumen, so constraints derived from the lumen inherently exclude valid scar tissue. Removing these constraints and adopting a properly sized and configured U-Net with a scar-derived Gaussian weighting loss and mixed-spacing inference yields competitive performance; a cavity-wall-derived weighting of identical form brings no benefit, confirming that the anatomical origin of a spatial prior matters more than its mathematical formulation. All results were obtained using only the challenge-provided training data. The main limitation is that training data came from a single center; cross-center domain shift remains a challenge, as evidenced by the performance gap on unseen scanners in the Task 2 test set. Addressing this gap through data augmentation, test-time adaptation, or multi-center training is an important direction for future work.

Acknowledgments. This study was funded by the National Natural Science Foundation of China (grant number 12471290).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Arega, T.W., Bricq, S., Meriaudeau, F.: Using Polynomial Loss and Uncertainty Information for Robust Left Atrial and Scar Quantification and Segmentation. In: Left Atrial and Scar Quantification and Segmentation. p. 133–144. Springer Nature Switzerland (2023). https://doi.org/10.1007/978-3-031-31778-1_13
2. Gao, S., Zhou, H., Gao, Y., Zhuang, X.: BayeSeg: Bayesian Modeling for Medical Image Segmentation with Interpretable Generalizability. Medical Image Analysis **89**, 102889 (10 2023). https://doi.org/10.1016/j.media.2023.102889
3. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. Nature Methods **18**(2), 203–211 (12 2020). https://doi.org/10.1038/s41592-020-01008-z
4. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jäger, P.F.: nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. p. 488–498. Springer Nature Switzerland (2024). https://doi.org/10.1007/978-3-031-72114-4_47
5. Kundu, B., Linte, C.A.: A Two Stage Pipeline for Left Atrial Wall Constrained Scar Segmentation and Localization from LGE-MR Images. arXiv preprint arXiv:2604.27101 (2026). https://doi.org/10.48550/arXiv.2604.27101
6. Lefebvre, A.L., Yamamoto, C.A.P., Shade, J.K., Bradley, R.P., Yu, R.A., Ali, R.L., Popescu, D.M., Prakosa, A., Kholmovski, E., Trayanova, N.A.: LASSNet: A Four

Steps Deep Neural Network for Left Atrial Segmentation and Scar Quantification. In: Left Atrial and Scar Quantification and Segmentation. p. 1–15. Springer Nature Switzerland (2023). https://doi.org/10.1007/978-3-031-31778-1_1
7. Li, L., Zimmer, V.A., Schnabel, J.A., Zhuang, X.: AtrialJSQnet: A New Framework for Joint Segmentation and Quantification of Left Atrium and Scars Incorporating Spatial and Shape Information. Medical Image Analysis **76**, 102303 (2 2022). https://doi.org/10.1016/j.media.2021.102303
8. Li, L., Zimmer, V.A., Schnabel, J.A., Zhuang, X.: Medical Image Analysis on Left Atrial LGE MRI for Atrial Fibrillation Studies: A Review. Medical Image Analysis **77**, 102360 (4 2022). https://doi.org/10.1016/j.media.2022.102360
9. Marrouche, N.F., Wilber, D., Hindricks, G., Jais, P., Akoum, N., Marchlinski, F., Kholmovski, E., Burgon, N., Hu, Nanand Mont, L., Deneke, T., Duytschaever, M., Neumann, T., Mansour, M., Mahnkopf, C., Herweg, B., Daoud, E., Wissner, E., Bansmann, P., Brachmann, J.: Association of Atrial Tissue Fibrosis Identified by Delayed Enhancement MRI and Atrial Fibrillation Catheter Ablation: The DECAAF Study. JAMA **311**(5), 498 (2 2014). https://doi.org/10.1001/jama.2014.3
10. Ogbomo-Harmitt, S., Grzelak, J., Qureshi, A., King, A.P., Aslanidi, O.: TESSLA: Two-Stage Ensemble Scar Segmentation for the Left Atrium. In: Zhuang, X., Li, L., Wang, S., Wu, F. (eds.) Left Atrial and Scar Quantification and Segmentation. p. 106–114. Springer Nature Switzerland (2023). https://doi.org/10.1007/978-3-031-31778-1_10
11. Stimper, V., Bauer, S., Ernstorfer, R., Scholkopf, B., Xian, R.P.: Multidimensional Contrast Limited Adaptive Histogram Equalization. IEEE Access **7**, 165437–165447 (2019). https://doi.org/10.1109/access.2019.2952899
12. Thaler, F., Štern, D., Plank, G., Urschler, M.: LA-CaRe-CNN: Cascading Refinement CNN for Left Atrial Scar Segmentation. In: Comprehensive Analysis and Computing of Real-World Medical Images. p. 180–191. Springer Nature Switzerland (2025). https://doi.org/10.1007/978-3-031-87009-5_18
13. Xu, F., Kennelly, J., Zolotarev, A.M., Roney, C., Nohel, M., Ulrich, C., Anenberg, B., Chang, P., On, Y.H., Varela, M., Thesing, C., Banerjee, A., Almar-Munoz, E., Tiefenthaler, M., Merino-Caviedes, S., Nnadozie, E.C., Qayyum, A., Mazher, M., Anwaar, W., Xue, W., Wen, H., Kang, J., Beveridge, L., Zhang, Leand Gunawardhana, M., Kiuchi, K., Stiles, M.K., Zhao, J.: MBAS2024: A Large-Scale Benchmark for Multi-Class Bi-Atrial Segmentation in Multi-Center Contrast-Enhanced MRIs. Medical Image Analysis **113**, 104203 (9 2026). https://doi.org/10.1016/j.media.2026.104203
14. Zhang, Y., Cheng, H., Li, D., Pan, L.: Left Atrial Scar Segmentation and Quantification Using Residual CBAM-EAM Attention UNet for LGE MRI. In: Comprehensive Analysis and Computing of Real-World Medical Images. p. 149–157. Springer Nature Switzerland (2025). https://doi.org/10.1007/978-3-031-87009-5_15
15. Zhang, Y., Meng, Y., Zheng, Y.: Automatically Segment the Left Atrium and Scars from LGE-MRIs Using a Boundary-Focused nnU-Net. In: Left Atrial and Scar Quantification and Segmentation. p. 49–59. Springer Nature Switzerland (2023). https://doi.org/10.1007/978-3-031-31778-1_5
16. Zhuang, X.: Multivariate Mixture Model for Myocardial Segmentation Combining Multi-Source Images. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(12), 2933–2946 (12 2019). https://doi.org/10.1109/tpami.2018.2869576
17. Zhuang, X., Shen, J.: Multi-Scale Patch and Multi-Modality Atlases for Whole Heart Segmentation of MRI. Medical Image Analysis **31**, 77–87 (7 2016). https://doi.org/10.1016/j.media.2016.02.006