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 for Curriculum Learning and ROI-Aware Supervision for Highly Accelerated 4D Flow MRI Reconstruction

1 Validation Metrics

The validation metrics are calculated using the same definitions as the official challenge metrics. Validation is performed by submitting the reconstructed results to the official server; no separate local validation set is used.

Let Ω be the evaluation ROI, x and $\hat{x}$ be the reference and reconstructed magnitude images, and $\mathbf{v}$ and $\hat{\mathbf{v}}$ be the reference and reconstructed velocity vectors. The metric formulas used in the revised description are

$$\mathrm{SSIM}(x,\hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)}, \tag{S1}$$

$$\mathrm{nRMSE} = \frac{\sqrt{\sum_{i\in\Omega}(\hat{x}_i - x_i)^2}}{\sqrt{\sum_{i\in\Omega} x_i^2} + \epsilon}, \tag{S2}$$

$$\mathrm{RelErr} = \frac{\sqrt{\sum_{i\in\Omega}(\|\hat{\mathbf{v}}_i\|_2 - \|\mathbf{v}_i\|_2)^2}}{\sqrt{\sum_{i\in\Omega}\|\mathbf{v}_i\|_2^2} + \epsilon}, \tag{S3}$$

$$\mathrm{AngErr} = \frac{1}{|\Omega|}\sum_{i\in\Omega} \arccos\left(\frac{\hat{\mathbf{v}}_i^{\mathsf{T}}\mathbf{v}_i}{\|\hat{\mathbf{v}}_i\|_2\|\mathbf{v}_i\|_2 + \epsilon}\right). \tag{S4}$$

Here, μ, σ, and $\sigma_{x\hat{x}}$ denote the local means, standard deviations, and covariance used by SSIM, respectively, and ϵ is a numerical stabilizer. ComplexDiffErr is reported according to the CMRx4DFlow2026 Task 2 server definition.

2 Statistical Analysis

The following table reports the results of the two-sided paired t test report. The analysis compares the combined method with Original FlowVN using the five server-reported center-level results ($df = 4$). The mean difference is defined as combined method minus Original FlowVN.

Table S1. Center-level paired analysis for the combined method versus Original FlowVN (five centers, two-sided paired t-test).

Metric	Mean difference	95% CI	p value
SSIM	+0.004264	[−0.000135, +0.008664]	0.0546
nRMSE	-0.004812	[−0.007664, -0.001960]	0.0094
RelErr	-0.041836	[−0.070588, -0.013083]	0.0156
AngErr	-2.270997	[−4.016756, -0.525238]	0.0225
ComplexDiffErr	-0.010770	[−0.018000, -0.003540]	0.0144

3 Data Preparation and Augmentation

The frequency-encoding (FE) dimension is inverse Fourier transformed before training data are extracted. Five adjacent FE slices are selected for each training sample, and the same slice range is used for the coil-sensitivity map and ROI mask. The coil-sensitivity maps are normalized voxelwise as

$$c_j(\mathbf{r}) \leftarrow \frac{c_j(\mathbf{r})}{\sqrt{\sum_l |c_l(\mathbf{r})|^2}}, \tag{S5}$$

where j indexes the coil. The zero-filled image, fully sampled target, and masked k-space are normalized by the same sample-specific factor,

$$\nu = \frac{\|f\|_2}{\| \, |f| > 0 \, \|_2}. \tag{S6}$$

During training, one velocity encoding is selected at random for each sample. A temporal block is selected with cyclic indexing. For the selected acceleration factor, the implementation generates a k–t Gaussian undersampling mask online over the two phase-encoding directions; the FE direction remains fully sampled. The mask generator uses a central radius of 0.5 in both directions, Gaussian widths of one fifth of the corresponding matrix dimensions, a minimum-distance factor of 3, and a repetition-decay factor of 0.5. No additional geometric or intensity augmentation is applied.

4 ROI-Mask Processing

The ROI mask is extracted using the same FE-slice interval as the corresponding training image, rearranged to the image axes, converted to Boolean type, and expanded as required for the loss calculation. The ROI mask is used to construct the loss only; it is not concatenated with, or otherwise supplied as, a network input. With binary mask m_i, the voxel weight is

$$w_i = m_i + \eta(1 - m_i), \qquad \eta = 0.1. \tag{S7}$$

Table S2. Observed ROI fractions in the training data.

Statistic	Value
Training cases	138
Five-slice crops	4,208
Mean ROI fraction per crop	3.037836%
Median ROI fraction per crop	1.832274%
Interquartile range	0.119579–5.180564%
Crops without ROI voxels	826/4,208 (19.629%)

For reconstruction $\hat{x}$ and target x, the implemented objective is

$$\mathcal{L}_{\text{complex}} = \frac{\sum_i w_i |\hat{x}_i - x_i|}{\sum_i w_i}, \tag{S8}$$

$$\mathcal{L}_{\text{magnitude}} = \frac{\sum_i w_i \big||\hat{x}_i| - |x_i|\big|}{\sum_i w_i}, \tag{S9}$$

$$\mathcal{L} = 1.0\,\mathcal{L}_{\text{complex}} + 0.3\,\mathcal{L}_{\text{magnitude}}. \tag{S10}$$

The loss is evaluated over the complete five-slice training crop.

4.1 Observed ROI Fraction

We measured the ROI fraction from all 138 training masks using the same five-slice FE cropping procedure as the training data loader. This produced 4,208 training crops.

For an ROI fraction p, the fraction of the total voxel weight assigned to the ROI is

$$C_{\text{ROI}}(p) = \frac{p}{p + \eta(1-p)}. \tag{S11}$$

With $\eta = 0.1$, the mean crop-level effective contributions were 19.692% for the ROI and 80.308% for the background. Thus, the tenfold per-voxel weight emphasizes the aortic region while preserving substantial background supervision. Crops without ROI voxels reduce to an ordinary full-field loss because all voxel weights are equal within those crops.

The reconstruction network does not use an ROI mask as an input at inference. The supplied segmentation is applied only after reconstruction to generate the ROI-masked challenge submission. The limitation that robustness to imperfect ROI masks has not been evaluated is stated in the revised Conclusion.

5 Baseline Sampling, Curriculum, and Padding

The Original FlowVN baseline samples the acceleration factor uniformly from $\{10, 20, 30, 40, 50\}$. The curriculum implementation uses the rate pools $\{10, 20\}$, $\{10, 20, 30\}$, and $\{10, 20, 30, 40, 50\}$, switching at epochs 10 and 20. Within each active pool, the acceleration factor is randomly sampled. The implementation uses $\eta = 0.1$, a complex-loss weight of 1.0, and a magnitude-loss weight of 0.3.

Table S3. Expected training exposure at each acceleration factor.

Training strategy	R10	R20	R30	R40	R50
Uniformly mixed FlowVN	20.00%	20.00%	20.00%	20.00%	20.00%
Proposed curriculum	24.33%	24.33%	19.33%	16.00%	16.00%

5.1 Hyperparameter Selection and Training Exposure

The background weight $\eta = 0.1$, complex-loss weight 1.0, and magnitude-loss weight 0.3 were fixed before the reported ablations and were not selected by a systematic grid search. The value $\eta = 0.1$ gives each ROI voxel ten times the weight of a background voxel while retaining nonzero background supervision. The complex-domain loss remains the primary objective, whereas the magnitude-domain term acts as a weaker auxiliary constraint. The curriculum milestones at epochs 10 and 20 provide two short introductory phases, followed by 80 epochs using all acceleration factors.

All variants used the same dataset, batch size, number of epochs, and total number of optimizer updates. However, the curriculum changes the distribution of updates across acceleration factors.

The exact realized draw counts were not recorded, so Table S3 reports the expected proportions implied by the sampling schedule. The current ablation therefore evaluates the combined effect of progressive ordering and exposure reweighting; it does not isolate a pure ordering effect. An exposure-matched shuffled-order control remains future work.

The acceleration-dependent functions $\lambda_r^{(k)}(R)$ and $\lambda_d^{(k)}(R)$ are inherited from FlowVN. For each convolutional or transposed-convolutional operation, the implementation applies zero padding of $(D_{\text{size}} - 1)/2$ in the FE-slice direction and $(K-1)/2$ in the two in-plane directions, where D_{size} and K are the corresponding kernel sizes. The loss is computed over all five output slices rather than only the central slice.

6 Center-Level Validation Results

The center-level validation results for Original FlowVN, the two single-component variants, and the complete method are summarized in Table S4.

The complete method improved upon Original FlowVN across all reported metrics at each of the five validation centers. It achieved the lowest nRMSE and ComplexDiffErr at every center, while the best SSIM, RelErr, and AngErr results were shared between the complete method and the single-component variants. These findings indicate that the aggregate improvements were consistently observed across different acquisition centers.

Table S4. Center-level validation metrics for the ablation study. Higher SSIM and lower nRMSE, RelErr, AngErr, and ComplexDiffErr indicate better performance. Best values are shown in bold.

Method	Center	$SSIM_{adj}$ ↑	$nRMSE_{adj}$ ↓	$RelErr_{adj}$ ↓	$AngErr_{adj}$ ↓	$ComplexDiffErr_{adj}$ ↓
Original FlowVN	Center07	0.930440	0.058966	0.531844	39.342014	0.108312
	Center08	0.955827	0.038709	0.412496	34.056143	0.078778
	Center10	0.892293	0.077113	0.454388	39.837996	0.183772
	Center11	0.925060	0.081844	0.582311	47.848292	0.142812
	Center12	0.914712	0.059902	0.572069	42.499855	0.146534
ROI loss only	Center07	**0.932795**	0.057456	0.512026	**37.666073**	0.105317
	Center08	0.958487	0.037491	**0.401379**	**33.286418**	0.076844
	Center10	0.895883	0.075179	0.428733	37.679420	0.175917
	Center11	**0.926686**	0.079563	0.552223	45.519034	0.138731
	Center12	0.920365	0.057062	0.531959	40.963510	0.139070
Curriculum only	Center07	0.929900	0.057998	0.509874	39.116267	0.105996
	Center08	0.958225	0.035391	0.426649	35.088513	0.074733
	Center10	0.893568	0.075126	0.413353	37.651477	0.173379
	Center11	0.923654	0.079405	0.540987	46.634328	0.139285
	Center12	0.916525	0.056463	0.557270	42.008483	0.138286
Ours	Center07	0.931695	**0.057059**	**0.503052**	37.897893	**0.104243**
	Center08	**0.960322**	**0.034042**	0.403769	33.840658	**0.071029**
	Center10	**0.896155**	**0.073569**	**0.396466**	**36.558680**	**0.169318**
	Center11	0.926679	**0.075796**	**0.516281**	**38.956063**	**0.134012**
	Center12	**0.924803**	**0.052007**	**0.524362**	**38.956020**	**0.127755**

Table S5. Distribution of the 32 validation cases across centers and acceleration factors.

Center	R10	R20	R30	R40	R50	Total
Center007	4	3	3	4	3	17
Center008	3	2	0	1	2	8
Center010	0	0	2	0	0	2
Center011	0	0	1	0	1	2
Center012	0	2	0	1	0	3
All	7	7	6	6	6	32

7 Validation Distribution Across Centers and Acceleration Factors

Each validation case contained one acceleration factor, and one shared checkpoint per method was used for all acceleration factors.

The overall distribution was nearly balanced across R10–R50. However, the official server provided aggregate and center-level metrics rather than per-case scores. Separate reference-based metrics for each acceleration factor could therefore not be calculated. The center-level results document evaluation coverage but are not substitutes for a per-acceleration performance analysis, because center, scanner, field strength, and acceleration factor are confounded.