Supplementary Material: VeloSubspace

Weihao Cheng and Huihui Ye*

The School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China
weihao_cheng@hdu.edu.cn; yehuihui@hdu.edu.cn

1 Locked analysis protocol

P005 is the sole calibration case and P006 is the held-out evaluation case. No P006 image, metric, or target was used to select rank, blend, phase weight, iteration count, comparator settings, slice, cardiac phase, or display scale. The real-case slice is the P006 slice with maximum segmentation area; the cardiac phase maximizes the fully sampled frame-wise 95th-percentile speed on that slice. Repeated masks are sampling perturbations of one subject and are not analyzed as independent observations.

The hypotheses, endpoints, and analysis rules were recorded before evaluation. The standard Global LR-DC comparator uses a shared rank-12 projection, unit blend, three iterations, no phase regularization, and the same multi-coil update as the proposed method. The 14-iteration phase-disabled variant is evaluated at the three prespecified $R = 20$ masks (seeds 101–103).

2 Calibration trajectory

Table 1 reports every iteration of the locked joint-basis P005 calibration search. The objective is magnitude nRMSE plus velocity relative error plus angular error divided by 180°. Reaching the tested iteration boundary is disclosed as a specialization risk; the search was not extended after P006 evaluation.

On the separate P005 confirmation mask ($R = 40$, seed 29), the independent-basis three-iteration comparator achieved nRMSE 0.2014, SSIM 0.5408, relative error 0.7360, and angular error 61.41°. The frozen joint 14-iteration candidate achieved 0.1755, 0.6245, 0.6664, and 55.44°, respectively. This checks mask sensitivity but not subject-level generalization.

3 Per-mask held-out results

Table 2 reports paired VeloSubspace minus Global LR-DC changes. Negative is favorable except for SSIM, where positive is favorable. No population-level test is applied.

* Corresponding author

Table 1. P005 joint-basis calibration trajectory.

Iteration	nRMSE	RelErr	AngErr (°)	Objective
1	0.2116	0.6812	55.47	1.2009
2	0.1948	0.6630	53.48	1.1550
3	0.1852	0.6457	52.14	1.1206
4	0.1789	0.6316	51.16	1.0947
5	0.1742	0.6197	50.42	1.0741
6	0.1705	0.6097	49.84	1.0572
7	0.1675	0.6015	49.38	1.0433
8	0.1648	0.5945	49.03	1.0318
9	0.1625	0.5888	48.76	1.0222
10	0.1605	0.5839	48.53	1.0139
11	0.1586	0.5800	48.35	1.0073
12	0.1570	0.5765	48.20	1.0012
13	0.1554	0.5736	48.08	0.9961
14	0.1540	0.5708	47.98	0.9913

Acceleration-stratified and worst-case audit

Table 3 stratifies the paired changes by acceleration. All four mean changes are favorable at every rate. Across the 15 individual pairs, the least favorable changes are -0.0316 nRMSE, $+0.1388$ SSIM, -0.0500 velocity relative error, and $-6.10°$ angular error; hence no single mask reverses the comparator ordering on any organizer endpoint. This remains a descriptive sampling-robustness audit of one held-out subject, not a population-level inference.

Table 4 expresses the signed paired differences as relative effect sizes. VeloSubspace lowers nRMSE, velocity relative error, and angular error by 16.8%, 12.8%, and 14.3% relative to Global LR-DC, while increasing SSIM by 0.152. The smallest retained gain across the five acceleration strata is still 12.5% nRMSE reduction, 7.5% velocity relative-error reduction, 13.6% angular-error reduction, and 0.142 absolute SSIM. These are descriptive effect sizes, not population estimates.

Table 5 omits all three masks at one acceleration and recomputes the paired mean over the other 12 conditions. Every omission preserves a favorable direction for all four endpoints. Across the five omissions, mean changes range from -0.0404 to -0.0364 nRMSE, $+0.1489$ to $+0.1544$ SSIM, -0.0947 to -0.0808 velocity relative error, and $-8.96°$ to $-8.23°$ angular error. No single acceleration stratum therefore drives the aggregate direction. The analysis remains descriptive and comes from one held-out subject.

Iteration-matched phase ablation

Table 6 holds rank, blend, iteration count, seed, and data-consistency updates fixed and changes only circular relative-phase regularization across all five accelerations. SSIM and both velocity endpoints improve at every rate. Averaged over rates, velocity relative and angular errors decrease by 0.0476 and 4.58°; nRMSE

Table 2. Held-out changes: VeloSubspace minus Global LR-DC.

R	Seed	ΔnRMSE	ΔSSIM	ΔRelErr	ΔAngErr
10	101	-0.0475	+0.1560	-0.1227	-6.10
10	102	-0.0496	+0.1562	-0.0942	-6.70
10	103	-0.0481	+0.1562	-0.1085	-6.54
20	101	-0.0416	+0.1636	-0.1189	-8.10
20	102	-0.0428	+0.1662	-0.0895	-7.87
20	103	-0.0409	+0.1616	-0.1126	-8.41
30	101	-0.0373	+0.1551	-0.1069	-8.60
30	102	-0.0368	+0.1505	-0.0827	-9.64
30	103	-0.0363	+0.1471	-0.0796	-9.85
40	101	-0.0363	+0.1546	-0.0702	-10.05
40	102	-0.0336	+0.1396	-0.0907	-8.90
40	103	-0.0342	+0.1456	-0.0596	-9.09
50	101	-0.0322	+0.1419	-0.0555	-9.11
50	102	-0.0316	+0.1388	-0.0500	-8.86
50	103	-0.0332	+0.1446	-0.0536	-9.09

Table 3. Acceleration-stratified paired robustness: mean VeloSubspace minus Global LR-DC over three fixed masks. All displayed directions are favorable.

R	ΔnRMSE	ΔSSIM	ΔRelErr	ΔAngErr (°)
10	-0.0484	+0.1561	-0.1085	-6.45
20	-0.0418	+0.1638	-0.1070	-8.12
30	-0.0368	+0.1509	-0.0897	-9.36
40	-0.0347	+0.1466	-0.0735	-9.35
50	-0.0323	+0.1418	-0.0530	-9.02

is effectively unchanged (mean -0.00005), with improvements at three rates and increases of only 0.00005 and 0.00002 at the other two. The phase term acts most strongly on velocity while leaving magnitude essentially unchanged.

At $R = 20$, Table 7 covers all three masks. The phase-aware method improves all four organizer metrics on all three $R = 20$ masks. Averaged over these masks, changes are -0.00016 nRMSE, $+0.00180$ SSIM, -0.0605 velocity relative error, and $-4.75°$ angular error. All seven cross-rate/replication pairs improve SSIM and both velocity endpoints. These remain correlated reconstructions of one held-out subject and are reported descriptively.

4 Official FlowVN and regional diagnostics

The organizer checkpoint is run at seed 101 for every rate, giving a consistent learned-reference comparison across acceleration. Its single-mask-per-rate points have no error bars.

Table 4. Descriptive effect sizes of VeloSubspace relative to Global LR-DC. Error reductions are relative percentages; SSIM is an absolute gain. Each rate averages three fixed masks from the same held-out subject.

Scope	nRMSE red. (%)	SSIM gain	RelErr red. (%)	AngErr red. (%)
10	26.1	+0.156	18.4	13.6
20	18.6	+0.164	16.1	14.6
30	15.3	+0.151	12.9	15.1
40	13.9	+0.147	10.4	14.5
50	12.5	+0.142	7.5	13.6
Overall	16.8	+0.152	12.8	14.3

Table 5. Leave-one-acceleration-out sensitivity of the paired VeloSubspace minus Global LR-DC changes. Each row omits all three masks at the indicated rate and averages the remaining 12 conditions. Negative is favorable except for SSIM.

Omitted R	Pairs	ΔnRMSE	ΔSSIM	ΔRelErr	ΔAngErr (°)
10	12	-0.0364	+0.1508	-0.0808	-8.96
20	12	-0.0380	+0.1489	-0.0812	-8.54
30	12	-0.0393	+0.1521	-0.0855	-8.23
40	12	-0.0398	+0.1532	-0.0895	-8.24
50	12	-0.0404	+0.1544	-0.0947	-8.32

For regional fidelity, reference-relative phase is background-corrected and converted component-wise using $v = \phi\, 150/\pi$ cm/s. Mean-curve nRMSE compares the 25-point ROI mean-speed waveforms. Robust peak speed is the maximum across frames of the within-ROI 95th percentile. The jet set contains the reference top 5% speed voxels across all frames. The boundary set is the one-voxel inner shell of the organizer ROI. These endpoints are reconstruction diagnostics, not clinical validation of net flow or wall shear stress.

5 Frozen FlowVN measured-sample refinement

The official FlowVN prediction is refined only with the SENSE residual update in Eq. (4) of the main manuscript. P005 $R = 20$, seed 17 selects the step size and iteration count before P006: the grid contains steps $\{0.25, 0.5, 0.75, 1\}$, one to three updates, and a zero-update baseline. The minimum occurs at a unit step with three updates. Because this is a boundary optimum, the grid is not expanded after P006 is opened. The primary P006 $R = 20$ result is frozen first; the identical setting is then applied to the remaining rates.

At $R = 20$, the normalized acquired-sample residual falls from 0.307 to 0.073, a 76.1% relative reduction. Mean-speed-curve nRMSE changes from 0.259 to 0.205 and 95th-percentile-speed-curve nRMSE from 0.269 to 0.213. Top-5% velocity error improves from 0.727 to 0.722, whereas boundary error changes from

Table 6. Cross-acceleration iteration-matched phase ablation on P006, seed 101. Changes are phase enabled minus disabled; lower is favorable except for SSIM. Every prespecified rate is retained.

R	Δ nRMSE	Δ SSIM	Δ RelErr	Δ AngErr (°)
10	+0.00005	+0.00070	−0.0365	−2.48
20	−0.00007	+0.00154	−0.0667	−4.48
30	−0.00011	+0.00089	−0.0653	−4.49
40	+0.00002	+0.00029	−0.0378	−6.46
50	−0.00012	+0.00054	−0.0315	−4.99
Mean	−0.00005	+0.00079	−0.0476	−4.58

Table 7. Iteration-matched phase ablation on P006 at $R = 20$. Arrows denote phase disabled → enabled; all other settings are identical.

Seed	nRMSE ↓	SSIM ↑	RelErr ↓	AngErr (deg.) ↓
101	.18341→.18334	.63744→.63898	.61903→.55234	51.53→47.06
102	.17900→.17870	.65125→.65341	.59144→.55219	52.01→47.63
103	.18670→.18658	.62501→.62669	.63960→.56416	53.66→48.26

0.862 to 0.869. The update improves the organizer endpoints and temporal waveform fidelity, although regional behavior is not uniformly better.

6 Target-free cross-center deployment audit

The official Task S1 validation archive contains 40 Center006 aortic cases from a UIH 3 T uMR790 scanner, stratified into eight cases at each nominal acceleration. It supplies undersampled k-space, coil maps, sampling masks, segmentations, and metadata, but no fully sampled target. Before any outcome was inspected, we locked one case per rate as the lexicographically first subject: P010, P002, P001, P003, and P005 at $R = 10, 20, 30, 40, 50$. Frozen P005 parameters are used unchanged. Cases are processed sequentially with one CPU thread; no reconstructed volume is retained after endpoint extraction.

The confirmatory endpoints are completion with finite values and a positive reduction between the residual immediately before and after the final measured-sample update. All 5/5 cases satisfy both endpoints. Median final-step reduction is 4.07%, median final residual is 0.137, median wall time is 105.3 s, and the largest peak resident memory is 1.43 GiB. Cumulative zero-filled-to-final residual reduction is exploratory (median 74.6%, range 70.6–78.7%).

The external cases have only 6–9 cardiac frames. Frozen rank 12 therefore saturates to effective rank N_t in every case, making the temporal projection the identity. The experiment tests the circular-phase and data-consistency path under a new center/vendor forward model; it does not test external temporal

Table 8. Official FlowVN, P006 seed 101.

R	nRMSE	SSIM	RelErr	AngErr (°)
10	0.0785	0.9131	0.3964	34.11
20	0.1217	0.8774	0.4624	40.15
30	0.0847	0.8902	0.5290	45.12
40	0.1067	0.8110	0.6437	52.15
50	0.0987	0.8355	0.6049	50.69

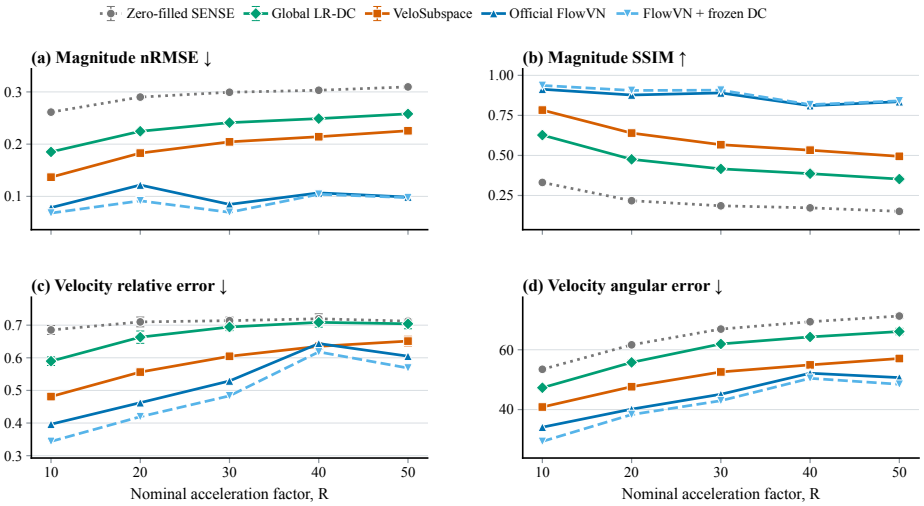

Fig. 1. All organizer metrics across acceleration. Error bars are standard deviations over three P006 masks for the training-free methods. FlowVN and FlowVN+DC use seed 101 and therefore have no error bars.

compression. Likewise, without a fully sampled target, residual reduction cannot be interpreted as external nRMSE, SSIM, velocity accuracy, or clinical validity.

Exploratory cross-subject replication

After the five-case audit was complete, we locked two additional official subjects using only rate, non-overlap, and compressed footprint: P026 at the canonical $R = 20$ rate and P022 at high acceleration $R = 40$. Both use the same frozen reconstruction and target-free endpoints. Both complete with finite outputs and positive final-step residual reduction. This replication is explicitly exploratory because the two rates were selected after the primary five-rate audit, although subjects were fixed before their outcomes.

Table 9. Regional velocity-fidelity diagnostics.

Method	Curve nRMSE	Peak bias (%)	Jet RelErr	Boundary RelErr
Zero-filled	0.2628	+3.5	0.9874	1.1697
Global LR-DC	0.2366	+2.6	0.8989	1.1209
LR-DC matched-14	0.1881	-1.4	0.8152	1.0865
VeloSubspace	0.2908	-23.8	0.7280	0.9232
FlowVN	0.2588	-31.1	0.7267	0.8621
FlowVN+DC	0.2054	-24.5	0.7218	0.8694

Table 10. Official FlowVN before and after the P005-frozen three-step acquired-sample update on P006 seed 101. Arrows denote base → refined.

R	nRMSE ↓	SSIM ↑	RelErr ↓	AngErr (°) ↓
10	0.079→0.068	0.913→0.937	0.396→0.344	34.11→29.35
20	0.122→0.092	0.877→0.906	0.462→0.419	40.15→38.33
30	0.085→0.070	0.890→0.907	0.529→0.483	45.12→42.96
40	0.107→0.104	0.811→0.817	0.644→0.618	52.15→50.45
50	0.099→0.098	0.836→0.841	0.605→0.569	50.69→48.47

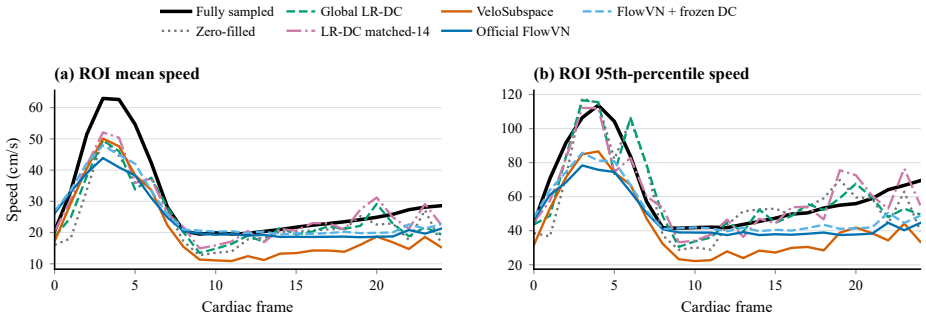

Fig. 2. P006 $R = 20$, seed-101 time-resolved ROI speed. Every curve is computed from the same organizer ROI and background-phase correction. These repeated frames are descriptive measurements of one subject, not independent samples.

Table 11. Prespecified target-free Center006/UIH deployment audit. Residuals are normalized acquired-sample residuals. The final-step change is pre-update to post-update; cumulative change is zero-filled to final.

Case	R	N_t	r_{eff}	ZF res.	Final res.	Final / cumulative Δ
P010	10	6	6	0.524	0.112	−4.59% / −78.67%
P002	20	6	6	0.492	0.145	−4.07% / −70.57%
P001	30	6	6	0.518	0.132	−2.91% / −74.60%
P003	40	9	9	0.631	0.151	−4.22% / −76.05%
P005	50	7	7	0.497	0.137	−3.66% / −72.43%

Table 12. Exploratory cross-subject replication. Subjects were selected by footprint and non-overlap before outcomes. Settings remain frozen.

Case	R	N_t	Final res.	Final-step Δ	Cumulative Δ	Time (s)
P026	20	7	0.135	−3.49%	−68.42%	70.2
P022	40	9	0.145	−3.32%	−72.12%	117.8

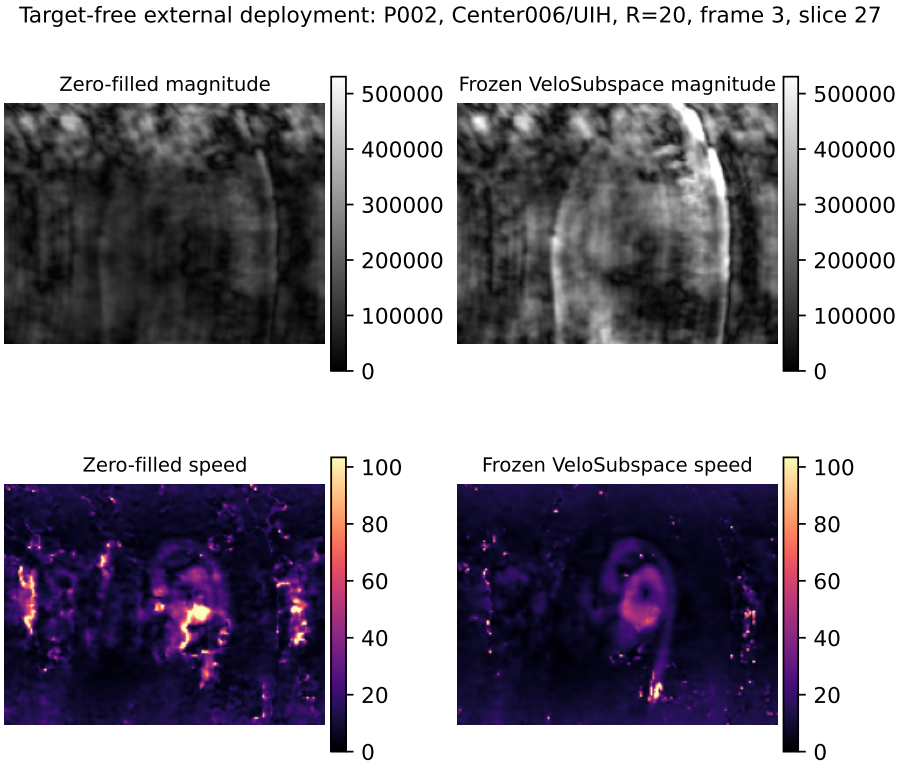

Fig. 3. Representative prespecified Center006/UIH case (P002, $R = 20$). Magnitude and speed are shown before and after the frozen pipeline. The figure documents target-free execution only; there is no fully sampled reference and no image-quality claim is inferred from visual appearance.

7 Data and compute provenance

The official demo release contains P005 and P006 and occupies 1.4 GB locally. The supplied Task S1 archive is 8.3 GiB compressed and contains 40 validation subjects. The locked five-case subset and two-case exploratory replication retain approximately 1.3 GiB in total. The full R1R2 release remains unused: its 17 split segments total roughly 271 GiB, beyond the local storage and retained-data budget. No external or simulated 4D Flow case is substituted for official challenge data.

Experiments use Python 3.9, NumPy 2.0, SciPy 1.13, PyTorch 2.5, and the organizer repository at the recorded commit. The two matched-14 ablation runs used one CPU thread, took 271.0 and 262.7 seconds, and remained below 1.27 GiB peak resident memory. Each run writes a partial CSV after every condition and can be resumed. The reproducibility archive contains locked protocols, scripts, environment metadata, per-mask metrics, regional definitions, calibration trajectories, and SHA-256 checksums.