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: Stage-Aligned Knowledge Distillation for Fast and Memory-Efficient 4D Flow MRI Reconstruction

Yuliang Xiao[1,2(✉)], Kian Anvari Hamedani[1,2], Zach Vavasour[1,2], Simon J. Graham[1,2], and Mark Chiew[1,2]

[1] Physical Sciences Platform, Sunnybrook Research Institute, Toronto, ON, Canada
[2] Department of Medical Biophysics, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

1 Detailed Complete-Case Inference Benchmark

We denote an eight-stage FlowVN by S8 and a 16-stage FlowVN by S16; knowledge distillation is abbreviated as KD. Table S1 provides the stepwise S16 execution-path decomposition underlying the complete-case benchmark summarized in main Table 2. The benchmark uses a fixed S16 checkpoint and one complete four-encoding case at acceleration $R = 10$ on an RTX 3090 graphics processing unit (GPU). Wall time includes cold compilation, loading, preprocessing, model construction and execution, host transfer, denormalization, grouping, and saving. Peak memory is measured in gibibytes (GiB).

Table S1. Detailed checkpoint-backed complete-case inference benchmark. Arrows indicate the favorable direction. Speed is relative to the eager reference, and relative L_2 compares each saved complex output with that reference.

(a) Runtime decomposition

Execution path	Wall (s) ↓	Forward (s) ↓	Load (s) ↓	Speed ↑
Eager reference	105.341	74.291	25.916	1.00×
Base stack	46.883	33.802	8.303	2.25×
Base + GPU adjoint	39.112	32.342	2.819	2.69×
Base + transpose-as-conv	41.436	33.217	4.381	2.54×
Extended combined path	38.554	31.778	2.805	2.73×

(b) Device memory and saved-output agreement

Execution path	Peak (GiB) ↓	Relative L_2 ↓
Eager reference	13.72	reference
Base stack	3.44	2.76×10^{-6}
Base + GPU adjoint	3.44	2.66×10^{-6}
Base + transpose-as-conv	3.44	2.90×10^{-6}
Extended combined path	3.44	2.75×10^{-6}

The base stack combines vectorized FlowVN interpolation, selective activation and regularizer compilation, overlapped loading, omission of unused ground-truth preprocessing, and convolution-library autotuning. The two extensions move adjoint preprocessing to the GPU and replace transposed convolution with an algebraically equivalent convolution using channel exchange and spatially flipped kernels. All accelerated paths reduce peak device memory from 13.72 to 3.44 GiB and agree with eager saved outputs to relative L_2 differences below 3×10^{-6}. These measurements are specific to the reported case, checkpoint, and hardware.

2 Primary-Effect Uncertainty

We report six metrics: normalized complex L_1, normalized root-mean-square error (nRMSE), structural similarity index measure (SSIM), relative velocity error (RelErr), angular error (AngErr, degrees), and VENC-scaled velocity-vector RMSE (VENC RMSE, cm/s). Confidence interval (CI) denotes the 10,000-sample case-bootstrap interval after averaging paired effects across the two training seeds. Table S2 reports absolute intervals and paired percentage changes; these are distinct estimands. All four reference-arm VENC revalidations passed the same 80-group completeness and provenance audit used for the other ablation arms.

Table S2. Absolute uncertainty and paired percentage change for the primary full-KD-minus-supervised-S8 comparison. Arrows indicate the favorable direction.

Metric	Absolute Δ (95% CI)	Paired change
Normalized L_1 ↓	-0.0022 ([-0.0043, +0.0005])	-1.69%
nRMSE ↓	-0.0012 ([-0.0018, -0.0007])	-1.91%
SSIM ↑	+0.0037 ([+0.0023, +0.0054])	+0.46%
RelErr ↓	-0.0310 ([-0.0535, -0.0112])	-2.90%
AngErr (°) ↓	-0.492 ([-0.747, -0.220])	-1.31%
VENC RMSE (cm/s) ↓	-0.589 ([-0.964, -0.211])	-2.48%

3 Nested Component Ablation

All arms share the S8 architecture, S16-derived initialization, data order, optimizer, schedule, precision, and epoch-index-9 checkpoint rule. Contrasts add final output, trajectory, and update matching in order; they are conditional increments, not factorial main effects. Tables S3 and S4 report means and ordered contrasts. VENC effects come from protocol-matched physical revalidation and are not inferred from phase-domain metrics.

Table S3. Two-seed means for the nested ablation; arrows indicate the favorable direction. AngErr is in degrees and VENC RMSE in cm/s.

Arm	Norm. L_1 ↓	nRMSE ↓	SSIM ↑	RelErr ↓	AngErr ↓	VENC RMSE ↓
Supervised	0.2114	0.0722	0.9000	0.7832	49.667	23.677
Output KD	0.2111	0.0722	0.8998	0.7839	49.676	23.694
+Trajectory	0.2024	0.0697	0.9054	0.7457	49.221	22.931
Full KD	0.2092	0.0710	0.9037	0.7521	49.175	23.087

Table S4. Ordered nested contrasts across six metrics. Effects are arm minus control; brackets give 95% case-bootstrap CIs after within-case averaging over accelerations and two seeds. VENC RMSE is in cm/s.

Added term	Metric	Control → arm	Δ [95% CI]
Final output	Norm. L_1 ↓	0.2114 → 0.2111	-0.0004 [-0.0006, -0.0002]
Final output	nRMSE ↓	0.0722 → 0.0722	-0.000003 [-0.000082, +0.000071]
Final output	SSIM ↑	0.9000 → 0.8998	-0.0001 [-0.0002, -0.0001]
Final output	RelErr ↓	0.7832 → 0.7839	+0.0007 [-0.0005, +0.0019]
Final output	AngErr (°) ↓	49.667 → 49.676	+0.009 [-0.008, +0.026]
Final output	VENC RMSE ↓	23.677 → 23.694	+0.017 [-0.003, +0.037]
Trajectory	Norm. L_1 ↓	0.2111 → 0.2024	-0.0086 [-0.0110, -0.0066]
Trajectory	nRMSE ↓	0.0722 → 0.0697	-0.0025 [-0.0033, -0.0018]
Trajectory	SSIM ↑	0.8998 → 0.9054	+0.0056 [+0.0038, +0.0075]
Trajectory	RelErr ↓	0.7839 → 0.7457	-0.0382 [-0.0552, -0.0227]
Trajectory	AngErr (°) ↓	49.676 → 49.221	-0.455 [-0.687, -0.221]
Trajectory	VENC RMSE ↓	23.694 → 22.931	-0.763 [-1.061, -0.485]
Update	Norm. L_1 ↓	0.2024 → 0.2092	+0.0068 [+0.0031, +0.0114]
Update	nRMSE ↓	0.0697 → 0.0710	+0.0013 [+0.0006, +0.0020]
Update	SSIM ↑	0.9054 → 0.9037	-0.0017 [-0.0032, -0.0004]
Update	RelErr ↓	0.7457 → 0.7521	+0.0065 [-0.0026, +0.0159]
Update	AngErr (°) ↓	49.221 → 49.175	-0.046 [-0.216, +0.130]
Update	VENC RMSE ↓	22.931 → 23.087	+0.156 [-0.048, +0.384]

4 Acceleration-Specific Ablation and VENC-Scaled Velocity RMSE

Table S5 gives every arm, acceleration, and all six metrics, including raw VENC RMSE in cm/s. Table S6 follows the uncertainty format of Table S2, reporting an absolute paired effect and 95% CI for every metric, ordered component, and acceleration. Thus S5 contains raw means, whereas S6 contains matched effects and uncertainty. The supervised and full-KD values use accepted re-evaluations of their frozen epoch-index-9 checkpoints; training and checkpoint selection were unchanged.

Table S5. Raw two-seed arm means at every acceleration for all six metrics; arrows indicate the favorable direction. VENC RMSE is in cm/s. Comparisons are descriptive and unadjusted for multiplicity.

R Arm	Norm. L_1 ↓	nRMSE ↓	SSIM ↑	RelErr ↓	AngErr (°) ↓	VENC RMSE ↓
10 Supervised	0.1186	0.0347	0.9674	0.3342	31.277	11.580
10 Output KD	0.1186	0.0347	0.9674	0.3342	31.294	11.583
10 +Trajectory	0.1179	0.0343	0.9669	0.3281	31.063	11.507
10 Full KD	0.1139	0.0334	0.9681	0.3247	30.411	11.373
20 Supervised	0.1513	0.0499	0.9430	0.4778	39.925	15.948
20 Output KD	0.1511	0.0499	0.9429	0.4779	39.960	15.962
20 +Trajectory	0.1500	0.0498	0.9428	0.4748	39.846	15.914
20 Full KD	0.1511	0.0493	0.9435	0.4716	39.437	15.886
30 Supervised	0.2027	0.0690	0.9100	0.6821	52.037	22.614
30 Output KD	0.2024	0.0690	0.9099	0.6826	52.045	22.626
30 +Trajectory	0.1950	0.0674	0.9134	0.6641	51.215	21.996
30 Full KD	0.2042	0.0688	0.9119	0.6699	51.626	22.220
40 Supervised	0.2559	0.0879	0.8730	0.9326	60.552	29.014
40 Output KD	0.2550	0.0878	0.8729	0.9333	60.559	29.036
40 +Trajectory	0.2408	0.0837	0.8825	0.8869	59.706	27.957
40 Full KD	0.2536	0.0871	0.8778	0.9016	60.000	28.342
50 Supervised	0.3287	0.1195	0.8064	1.4893	64.543	39.227
50 Output KD	0.3282	0.1196	0.8060	1.4914	64.522	39.263
50 +Trajectory	0.3084	0.1133	0.8215	1.3745	64.274	37.281
50 Full KD	0.3234	0.1162	0.8170	1.3929	64.399	37.615

Table S6: Acceleration-specific ordered effects for all six metrics. Each cell is absolute Δ [95% case-bootstrap CI], where Δ is the component-containing arm minus control after averaging the two seeds. VENC RMSE is in cm/s; comparisons are descriptive and unadjusted for multiplicity.

R	Metric	Absolute Δ [95% CI]
Final-output matching		
10	Norm. L_1 ↓	+0.000023 [-0.000022, +0.000068]
	nRMSE ↓	-0.000014 [-0.000027, -0.000002]
	SSIM ↑	-0.000014 [-0.000035, +0.000006]
	RelErr ↓	-0.000011 [-0.000198, +0.000199]
	AngErr (°) ↓	+0.017 [+0.003, +0.032]
	VENC RMSE (cm/s) ↓	+0.003 [-0.002, +0.009]

Continued on next page

Table S6 (continued)

R	Metric	Absolute Δ [95% CI]
20	Norm. L_1 ↓	-0.000134 [-0.000235, -0.000048]
	nRMSE ↓	+0.000000 [-0.000029, +0.000034]
	SSIM ↑	-0.000032 [-0.000076, +0.000010]
	RelErr ↓	+0.000151 [-0.000142, +0.000463]
	AngErr (°) ↓	+0.035 [+0.002, +0.073]
	VENC RMSE (cm/s) ↓	+0.014 [+0.001, +0.031]
30	Norm. L_1 ↓	-0.000298 [-0.000458, -0.000161]
	nRMSE ↓	-0.000006 [-0.000064, +0.000063]
	SSIM ↑	-0.000114 [-0.000211, -0.000023]
	RelErr ↓	+0.000500 [-0.000392, +0.001365]
	AngErr (°) ↓	+0.008 [-0.008, +0.024]
	VENC RMSE (cm/s) ↓	+0.012 [-0.008, +0.032]
40	Norm. L_1 ↓	-0.000835 [-0.001367, -0.000398]
	nRMSE ↓	-0.000090 [-0.000309, +0.000098]
	SSIM ↑	-0.000102 [-0.000272, +0.000059]
	RelErr ↓	+0.000723 [-0.000790, +0.002334]
	AngErr (°) ↓	+0.007 [-0.045, +0.064]
	VENC RMSE (cm/s) ↓	+0.022 [-0.007, +0.055]
50	Norm. L_1 ↓	-0.000508 [-0.000794, -0.000250]
	nRMSE ↓	+0.000095 [-0.000104, +0.000280]
	SSIM ↑	-0.000407 [-0.000729, -0.000115]
	RelErr ↓	+0.002090 [-0.003290, +0.007400]
	AngErr (°) ↓	-0.021 [-0.110, +0.049]
	VENC RMSE (cm/s) ↓	+0.036 [-0.060, +0.124]
Trajectory matching		
10	Norm. L_1 ↓	-0.000793 [-0.002390, +0.001077]
	nRMSE ↓	-0.000404 [-0.000726, -0.000033]
	SSIM ↑	-0.000482 [-0.001179, +0.000080]
	RelErr ↓	-0.006084 [-0.011456, +0.001006]
	AngErr (°) ↓	-0.231 [-0.456, +0.017]
	VENC RMSE (cm/s) ↓	-0.076 [-0.226, +0.127]
20	Norm. L_1 ↓	-0.001093 [-0.002564, +0.000678]
	nRMSE ↓	-0.000090 [-0.000559, +0.000385]
	SSIM ↑	-0.000078 [-0.000631, +0.000477]
	RelErr ↓	-0.003101 [-0.008358, +0.002589]
	AngErr (°) ↓	-0.114 [-0.366, +0.155]
	VENC RMSE (cm/s) ↓	-0.048 [-0.211, +0.122]
30	Norm. L_1 ↓	-0.007387 [-0.010215, -0.005085]
	nRMSE ↓	-0.001595 [-0.002367, -0.000843]
	SSIM ↑	+0.003487 [+0.002405, +0.004497]
	RelErr ↓	-0.018447 [-0.033176, -0.004637]
	AngErr (°) ↓	-0.830 [-1.066, -0.583]
	VENC RMSE (cm/s) ↓	-0.630 [-0.928, -0.353]

Continued on next page

Table S6 (continued)

R	Metric	Absolute Δ [95% CI]
40	Norm. L_1 ↓	-0.014190 [-0.019533, -0.009617]
	nRMSE ↓	-0.004091 [-0.005795, -0.002591]
	SSIM ↑	+0.009575 [+0.006341, +0.013106]
	RelErr ↓	-0.046397 [-0.073696, -0.021210]
	AngErr (°) ↓	-0.853 [-1.320, -0.359]
	VENC RMSE (cm/s) ↓	-1.079 [-1.564, -0.607]
50	Norm. L_1 ↓	-0.019773 [-0.026465, -0.013971]
	nRMSE ↓	-0.006307 [-0.008398, -0.004505]
	SSIM ↑	+0.015456 [+0.010156, +0.021165]
	RelErr ↓	-0.116866 [-0.177158, -0.063257]
	AngErr (°) ↓	-0.248 [-0.844, +0.328]
	VENC RMSE (cm/s) ↓	-1.982 [-2.922, -1.111]
Update matching		
10	Norm. L_1 ↓	-0.003948 [-0.004585, -0.003256]
	nRMSE ↓	-0.000885 [-0.001257, -0.000533]
	SSIM ↑	+0.001179 [+0.000712, +0.001657]
	RelErr ↓	-0.003432 [-0.007512, +0.001063]
	AngErr (°) ↓	-0.652 [-1.002, -0.317]
	VENC RMSE (cm/s) ↓	-0.134 [-0.256, +0.010]
20	Norm. L_1 ↓	+0.001062 [-0.001221, +0.004165]
	nRMSE ↓	-0.000435 [-0.000998, +0.000055]
	SSIM ↑	+0.000653 [-0.000427, +0.001454]
	RelErr ↓	-0.003219 [-0.011300, +0.005332]
	AngErr (°) ↓	-0.408 [-0.693, -0.071]
	VENC RMSE (cm/s) ↓	-0.028 [-0.250, +0.252]
30	Norm. L_1 ↓	+0.009194 [+0.004283, +0.015182]
	nRMSE ↓	+0.001397 [+0.000690, +0.002171]
	SSIM ↑	-0.001486 [-0.002715, -0.000368]
	RelErr ↓	+0.005772 [-0.004332, +0.016820]
	AngErr (°) ↓	+0.411 [+0.143, +0.673]
	VENC RMSE (cm/s) ↓	+0.224 [-0.142, +0.549]
40	Norm. L_1 ↓	+0.012727 [+0.006974, +0.019480]
	nRMSE ↓	+0.003434 [+0.002027, +0.005032]
	SSIM ↑	-0.004607 [-0.007392, -0.002141]
	RelErr ↓	+0.014700 [-0.005785, +0.036500]
	AngErr (°) ↓	+0.294 [-0.139, +0.742]
	VENC RMSE (cm/s) ↓	+0.385 [+0.021, +0.747]
50	Norm. L_1 ↓	+0.015035 [+0.008612, +0.023350]
	nRMSE ↓	+0.002861 [+0.001262, +0.004391]
	SSIM ↑	-0.004478 [-0.007936, -0.000752]
	RelErr ↓	+0.018431 [-0.003262, +0.040028]
	AngErr (°) ↓	+0.125 [-0.305, +0.565]
	VENC RMSE (cm/s) ↓	+0.334 [-0.122, +0.777]

5 Acceleration-Specific Full-KD Effects

Table S7 reports the distinct joint-recipe comparison, full-KD S8 versus supervised S8, by acceleration. Each cell is the mean within-case relative effect after averaging the two matched training seeds. This differs from the raw arm means in S5 and the ordered component effects in S6. The comparisons are descriptive, unadjusted for multiplicity, and are not separate efficacy claims.

Table S7. Joint-recipe effect by acceleration: full-KD S8 versus supervised S8. Cells are mean within-case relative changes after averaging the two seeds. Negative is favorable for errors; positive is favorable for SSIM. Asterisks mark favorable absolute paired bootstrap intervals excluding zero.

R	Norm. L1 ↓	nRMSE ↓	SSIM ↑	RelErr ↓	AngErr ↓	VENC ↓
10	−4.79%*	−4.34%*	+0.07%	−3.44%	−3.25%*	−3.02%
20	−0.86%	−1.32%*	+0.05%	−1.49%*	−1.41%*	−1.20%
30	+0.20%	−0.42%	+0.21%*	−1.82%*	−0.82%*	−1.85%*
40	−1.18%*	−0.87%	+0.57%*	−3.20%	−0.86%*	−2.73%*
50	−1.81%*	−2.62%*	+1.40%*	−4.53%*	−0.23%	−3.60%*

6 Organizer Task 2 External Validation

The three submissions were frozen before organizer scores were returned. The fixed-S8 comparison is full-KD versus supervised training. Comparisons with the organizer FlowVN reference, identified as S10 from the released implementation, are not stage matched. Tables S8 and S9 show the evaluator-supplied overall and center-level values. The evaluator did not expose per-volume numerical errors, so we do not construct a case-specific external-results table.

Table S8. Organizer Regular Task 2 validation on 32 volumes. The released FlowVN implementation identifies the quality reference as S10. Scores retain the evaluator's displayed precision; lower Complex Difference Error is better.

Submission	Stages	Error	Difference from FlowVN	Better than FlowVN
S16 teacher	16	0.085	−0.033	32/32
Supervised S8	8	0.129	+0.011	0/32
Full-KD S8	8	0.127	+0.010	0/32

Table S9. Evaluator-supplied organizer Task 2 validation Complex Difference Error by acquisition center. Parenthetical counts are validation volumes (17, 8, 2, 2, and 3; 32 total). Lower Complex Difference Error is better; per-volume numerical errors were not exposed.

Submission	C007 (17)	C008 (8)	C010 (2)	C011 (2)	C012 (3)
S16 teacher	0.085	0.051	0.153	0.115	0.105
Supervised S8	0.125	0.098	0.198	0.164	0.162
Full-KD S8	0.124	0.093	0.199	0.164	0.163

7 Control-Selected Qualitative Example

Figure S1 shows the original control-selected example. It is descriptive and was selected using the supervised S8 result, not the KD improvement.

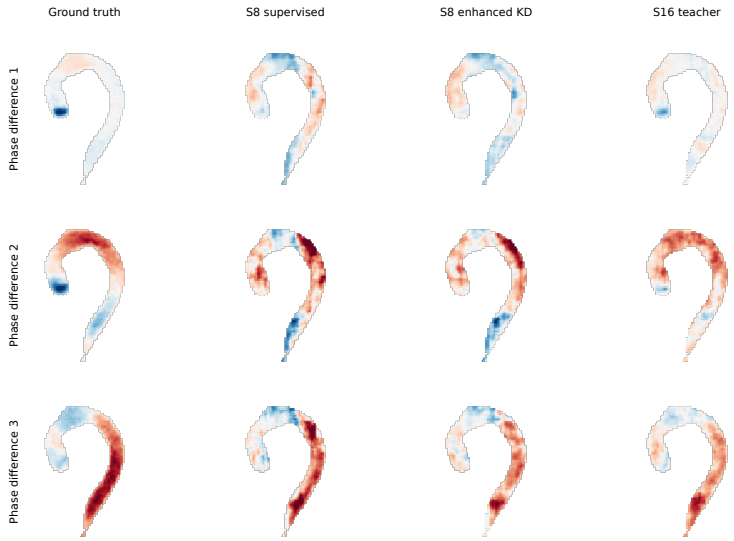


Fig. S1. Control-selected $R = 40$ case nearest the supervised-S8 median nRMSE. Rows show three phase-difference maps at the center slice and peak ground-truth velocity-energy frame with shared ground-truth scales. The matched S8 pair uses seed 12345.