Supplementary Material for "Revisiting Dual-Encoder Fusion and Optimal Transport Regularization in PET-CT Tumor Segmentation: A Matched-Control Study under Extended Training"

Man Li, Yongwei Tang, Xiaoqing Zhao, Yangyang Zhao, and Meng Wang

TABLE S-I
TRANSPORT ESTIMATORS COMPARED ON FOLD 0 UNDER THE IDENTICAL CONVERGED PROTOCOL. VALUES ARE PER-CASE DICE AVERAGED OVER THE 16 CASES OF FOLD 0. THE THREE TRANSPORT ARMS WERE TRAINED IN THE SAME BATCH AND ARE THEREFORE MUTUALLY COMPARABLE; THE NO-OT REFERENCE IS THE CONFIGURATION USED FOR THE MAIN COMPARISON AND COMES FROM AN EARLIER TRAINING BATCH.

Transport term	Fold-0 Dice
None (reference)	0.6406
Sorted-sample estimator (this work)	0.6224
Sliced Wasserstein, 32 projections	0.6385
Entropic OT, Sinkhorn iterations	0.6192
Fold-level sd of the converged protocol	0.075–0.081

TABLE S-II
PER-FOLD CANCER-TYPE COMPOSITION OF THE CROSS-VALIDATION SPLIT. FOLDS ARE CASE-LEVEL RANDOM SPLITS AND ARE NOT STRATIFIED.

Fold	Lung	Lymphoma	Melanoma	Total
0	6	1	9	16
1	6	6	3	15
2	3	6	6	15
3	4	8	3	15
4	8	4	2	14
Total	27	25	23	75

This document contains material that supports the main text but is not needed for its argument: a single-fold comparison of three transport estimators (Section S-I), and the per-fold cancer-type composition of the cross-validation split (Section S-II). Everything here is reported at the same fold-0 and 200-epoch protocol as the main text, and no result in this document carries a superiority claim.

S-I. Comparison of Transport Estimators at a Single Fold

Section III of the main text specifies the sampling-and-sorting estimator used for the transport term, gives its cost, and states that no accuracy guarantee is claimed for it. This section reports the comparison that was run in support of that statement.

All four configurations are the dual encoder of Section III, trained on fold 0 under the identical 200-epoch protocol, with the same optimiser, learning-rate schedule, augmentation pipeline and sampling of patches. The only quantity that differs is the transport term itself: no transport term at all; the sorted-sample estimator of Section III; sliced Wasserstein over 32 random one-dimensional projections; and log-domain entropic optimal transport with a squared-Euclidean cost solved by Sinkhorn iterations. Each configuration was trained once.

The four values span 0.021 Dice. The fold-level standard deviation of the converged protocol is 0.075 to 0.081 for the configurations in Table I of the main text, roughly four times that span. This comparison therefore cannot rank the three transports, and we do not rank them. What it can do is screen for gross estimator failure: an estimator that destroyed the training signal would place its arm outside the fold-to-fold spread, and none of the three does.

Two further limits apply. First, the comparison is single-fold, on fold 0, which contains 16 cases of which only one is lymphoma (Section S-II). Second, the no-OT reference and the three transport arms do not come from the same training batch: the reference is above the sorted-sample estimator by 0.0182, above the entropic-OT estimator by 0.0214 and above the sliced-Wasserstein estimator by 0.0021, while two independent runs of the same architecture in this study differ by 0.0101 Dice on fold 0. The two larger gaps are therefore no more than about twice a difference that is known to arise from re-training alone, and none of the three can be attributed to the transport term. We report this as a screening result only, and we make no claim about the relative accuracy of the transport estimators.

S-II. Per-Fold Cancer-Type Composition

The five folds used throughout this study are case-level random splits of the 75 cases. They are **not** stratified by cancer type, and the manuscript states this rather than asserting otherwise. Table S-II gives the composition of each fold.

Two consequences of this composition are load-bearing in the main text. Fold 0 contains a single lymphoma case, and it is the fold on which every pilot ablation and every MAE pre-training arm was run, so no fold-0 per-cancer contrast can carry a significance claim. More generally, the per-cancer stratum sizes (23 to 27 cases) are small relative to the fold-to-fold dispersion reported beside them in Table II of the main text, which is why that table reports dispersion rather than significance.