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

This supplement reports per-organ and per-configuration results condensed in the main text. All values are computed on the same splits and seeds described in Section 3. Tables 3 and 4 give the per-organ breakdown behind the Level 1 and Level 2 aggregates in Table 1. Localization error is lowest for kidneys and highest for colon and intestine across every configuration. On permutation, frozen MLP heads fall to near zero on the esophagus while a Transformer head on the same features recovers it, which is the per-organ form of the pooling effect discussed in Section 4. Table 5 gives the full per-configuration zero-shot transfer to TotalSegmentator, condensed to ranges in the main text; symmetry stays near chance for all 13 configurations while relative position and orientation vary across configurations rather than degrading uniformly.

Table 3. Per-organ localization error (MED, px, mean over three seeds) on WORD. Lower is better. Kidneys are easiest; colon and intestine hardest across configurations.

Model	Liver	Spln	K-L	K-R	Stom	GB	Esph	Panc	Duod	Colon	Intes	Adrn
Trained from scratch												
CNN-MLP	8.3	8.0	7.2	7.1	10.3	9.1	8.6	9.1	8.1	11.5	10.9	9.5
CNN-Trans	10.7	10.6	8.3	8.8	14.1	11.4	9.0	11.9	9.6	16.8	15.0	12.1
ViT-MLP	11.6	10.6	8.5	8.4	14.3	11.3	7.3	11.1	10.0	17.6	15.3	11.2
ViT-Trans	11.2	10.2	8.0	8.3	13.8	10.7	10.1	10.8	9.6	17.6	16.1	11.0
ViT-Comb	11.1	10.0	8.2	8.2	13.5	11.6	8.5	10.9	9.6	17.2	15.1	11.0
Frozen foundation models												
RadDINO-MLP	8.9	8.8	7.5	7.4	11.3	9.4	6.9	9.7	8.5	14.9	11.8	10.4
RadDINO-Trans	8.8	8.5	7.3	7.3	10.6	9.0	7.0	9.5	8.6	14.2	11.6	9.7
BiomedCLIP-MLP	9.8	9.6	8.1	7.7	11.8	10.4	8.3	10.0	9.0	15.1	12.5	11.1
BiomedCLIP-Trans	9.0	9.0	7.7	7.4	11.2	9.9	8.9	9.5	8.4	14.2	12.1	10.8
SAM-Med2D-MLP	8.9	9.2	7.7	8.0	11.8	9.5	8.0	9.7	8.8	14.7	12.3	10.9
SAM-Med2D-Trans	9.5	9.4	8.0	8.2	11.9	9.7	7.4	10.2	9.3	14.0	12.0	10.5
Finetuned foundation models												
RadDINO-MLP†	10.5	10.2	8.4	8.4	13.2	11.3	11.2	11.2	9.7	17.1	14.5	11.2
RadDINO-Trans†	11.2	10.5	8.4	8.2	14.2	11.0	8.2	10.9	9.8	17.4	15.5	11.6

Table 4. Per-organ permutation accuracy (patch accuracy, %, mean over three seeds) on WORD. Higher is better. Frozen MLP heads collapse on variable organs, most sharply the esophagus, while a Transformer head on the same features recovers them.

Model	Liver	Spln	K-L	K-R	Stom	GB	Esph	Panc	Duod	Colon	Intes	Adrn
Trained from scratch												
CNN-MLP	98.5	96.6	97.2	97.5	92.6	85.3	53.8	88.7	88.1	94.9	87.3	92.4
CNN-Trans	98.3	94.3	96.2	95.9	87.8	70.2	10.3	79.0	78.4	93.3	74.9	87.7
ViT-MLP	98.1	95.4	96.8	97.3	84.9	81.8	30.8	84.1	82.8	89.9	78.1	92.6
ViT-Trans	99.2	97.4	98.7	98.9	92.5	88.2	79.5	94.9	90.5	96.9	87.7	96.8
ViT-Comb	97.0	93.5	95.4	96.2	78.4	73.7	48.7	81.6	79.9	85.2	75.1	87.7
Frozen foundation models												
RadDINO-MLP	92.6	83.7	73.9	70.9	75.1	55.3	12.8	62.9	61.6	82.3	60.0	85.4
RadDINO-Trans	98.0	94.1	93.7	94.4	91.3	79.8	82.1	88.2	83.2	95.1	84.0	93.2
BiomedCLIP-MLP	57.4	46.5	48.8	50.3	37.7	23.1	0.0	33.6	22.9	46.6	25.9	40.1
BiomedCLIP-Trans	98.4	95.0	95.6	95.1	90.0	79.4	74.4	92.1	84.7	96.4	85.9	93.5
SAM-Med2D-MLP	84.1	61.6	80.8	77.6	36.4	23.6	2.6	27.3	30.0	69.9	50.0	60.8
SAM-Med2D-Trans	98.4	93.9	96.4	97.2	87.6	82.4	53.8	87.0	82.6	93.5	82.9	93.5
Finetuned foundation models												
RadDINO-MLP†	98.9	97.2	97.9	98.5	92.5	87.6	59.0	90.6	89.5	95.8	88.7	94.4
RadDINO-Trans†	99.6	97.7	98.5	99.0	96.2	92.5	82.0	96.8	93.9	98.5	92.5	96.2

Table 5. Full zero-shot transfer to TotalSegmentator, Level 3 probes (%, mean over three seeds; RadDINO-Trans† two seeds). No retraining. Symmetry stays near its 33.3% chance level for every configuration while other probes shift.

Model	RelPos	Orient	Symm	Size	Dist
Trained from scratch					
CNN-MLP	68.8	9.1	30.9	38.4	48.2
CNN-Trans	67.5	9.3	31.4	34.3	49.1
ViT-MLP	67.8	14.8	31.8	32.7	46.9
ViT-Trans	59.6	18.9	29.7	34.7	54.2
ViT-Comb	68.9	8.3	31.1	31.5	49.1
Frozen foundation models					
RadDINO-MLP	50.0	20.1	33.9	38.2	44.3
RadDINO-Trans	50.8	16.4	33.8	31.2	54.0
BiomedCLIP-MLP	69.8	12.8	30.7	37.2	52.7
BiomedCLIP-Trans	68.9	12.1	31.4	36.2	53.4
SAM-Med2D-MLP	68.7	21.5	34.9	32.3	43.7
SAM-Med2D-Trans	60.3	22.1	27.3	31.4	57.6
Finetuned foundation models					
RadDINO-MLP†	48.7	21.1	31.4	33.0	48.9
RadDINO-Trans†	50.3	20.2	34.0	36.0	50.9
Chance	50.0	20.0	33.3	33.3	50.0