

Appendix

A Dataset Preprocessing

Table 1: Description of the pre-processing and Augmentation details performed on the HECKTOR Dataset.

Augmentations	Axis Probability		
Orientation	PLS	-	-
CT/PET Concatenation	1	-	-
Random Crop	-	0.5	96x96x96
Random Flip	x,y,z	0.2	-
Rotate by 90 (up to 3x)	x,y	0.2	-

B Dimensionality reduction

Table 2: we explore different dimensionality reduction techniques while passing $2N \times D$ in the encoder. We are faced by the need to keep the dimensions of the input of the decoder part as $N \times d$, hence we conduct one experiment where take alternative tokens from both modalities, one where we take tokens coming only from CT, and another one where we only pass the CT tokens to the decoder. The latter one shows better dice scores, consequently we adopt this dimensionality reduction for all the subsequent experiments.

	MDA	Mix CT only	PET only
Tumor Dice	0.47	0.63	0.30
Lymph Dice	0.20	0.57	0.06
Avg Dice	0.49	0.62	0.27