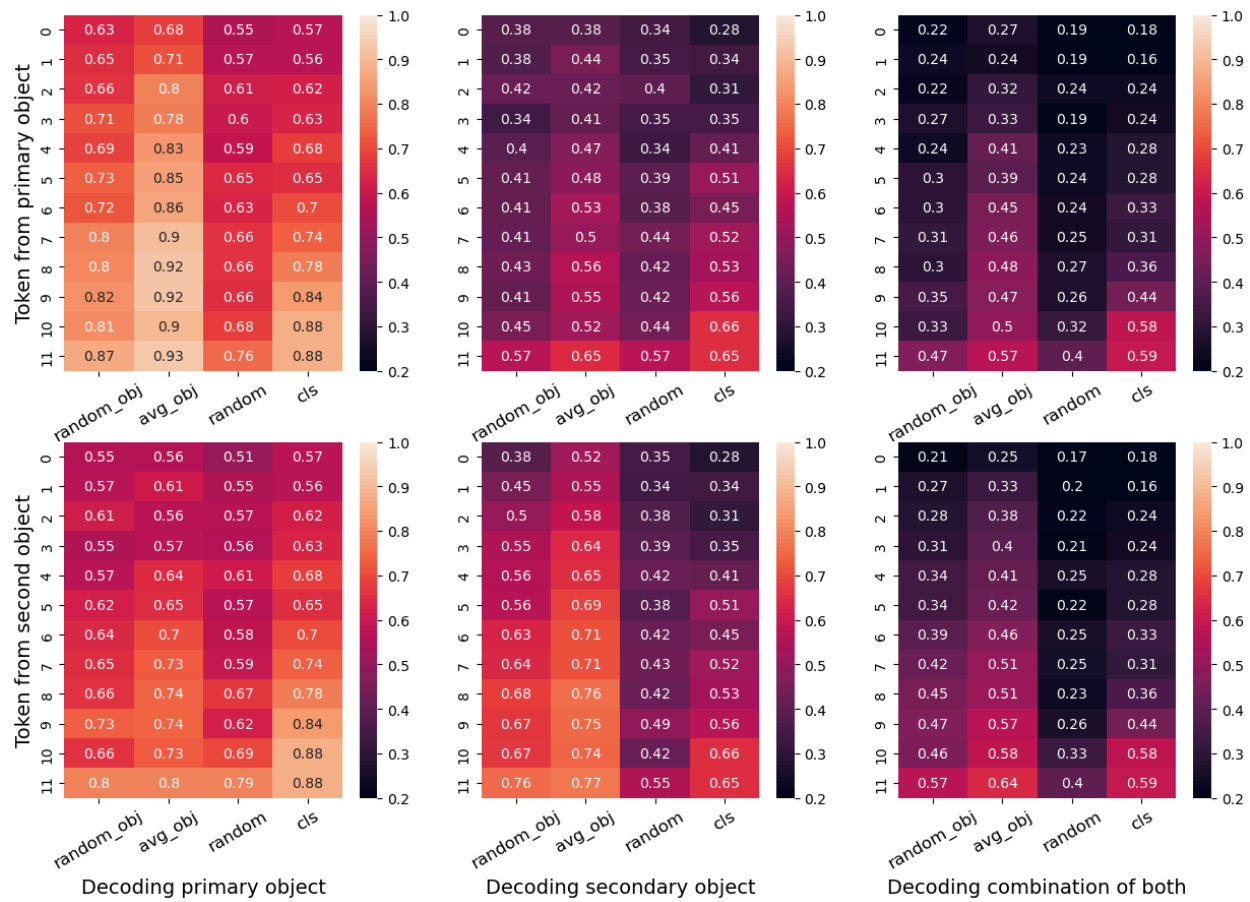


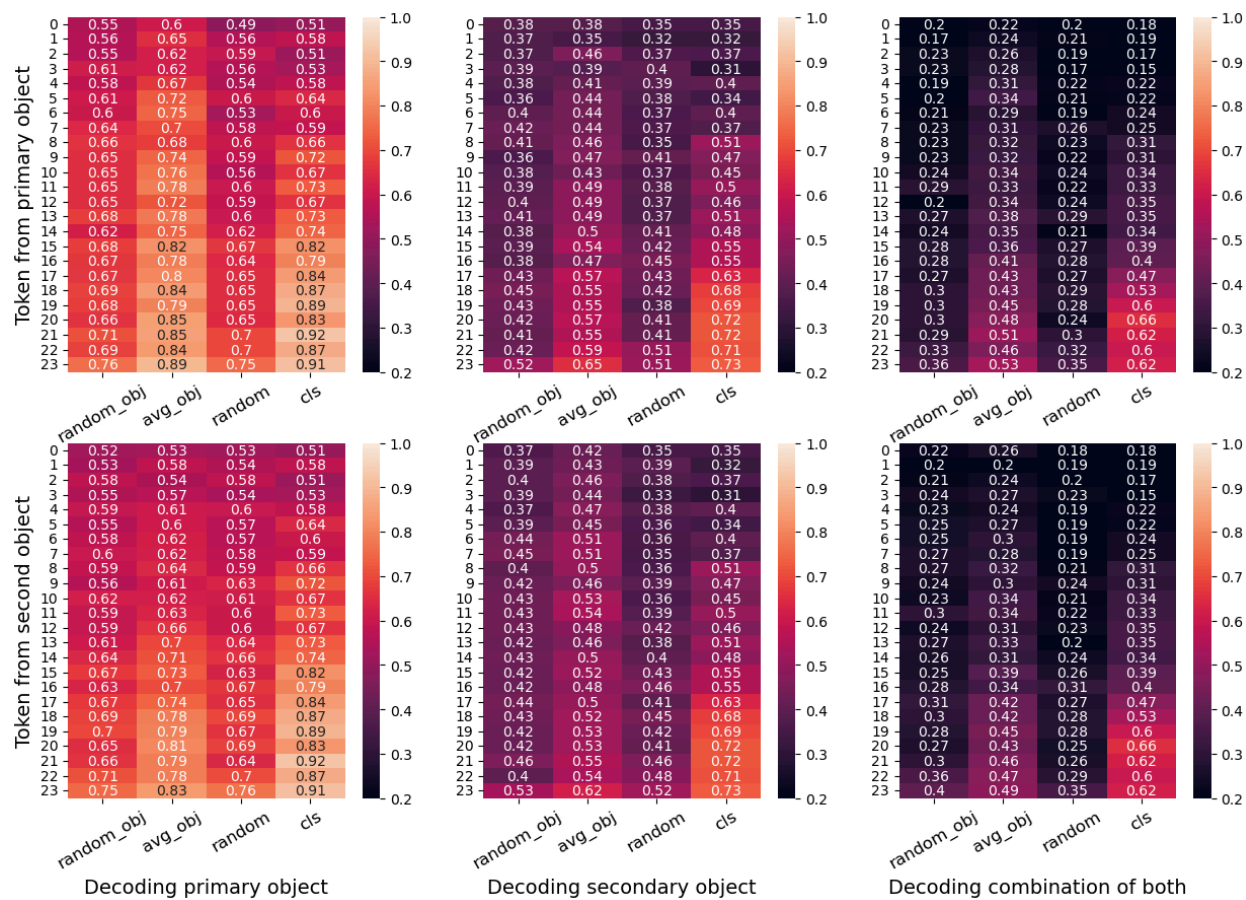
Supplementary material

The detailed plot for the paired object decoding task for each model is given below:

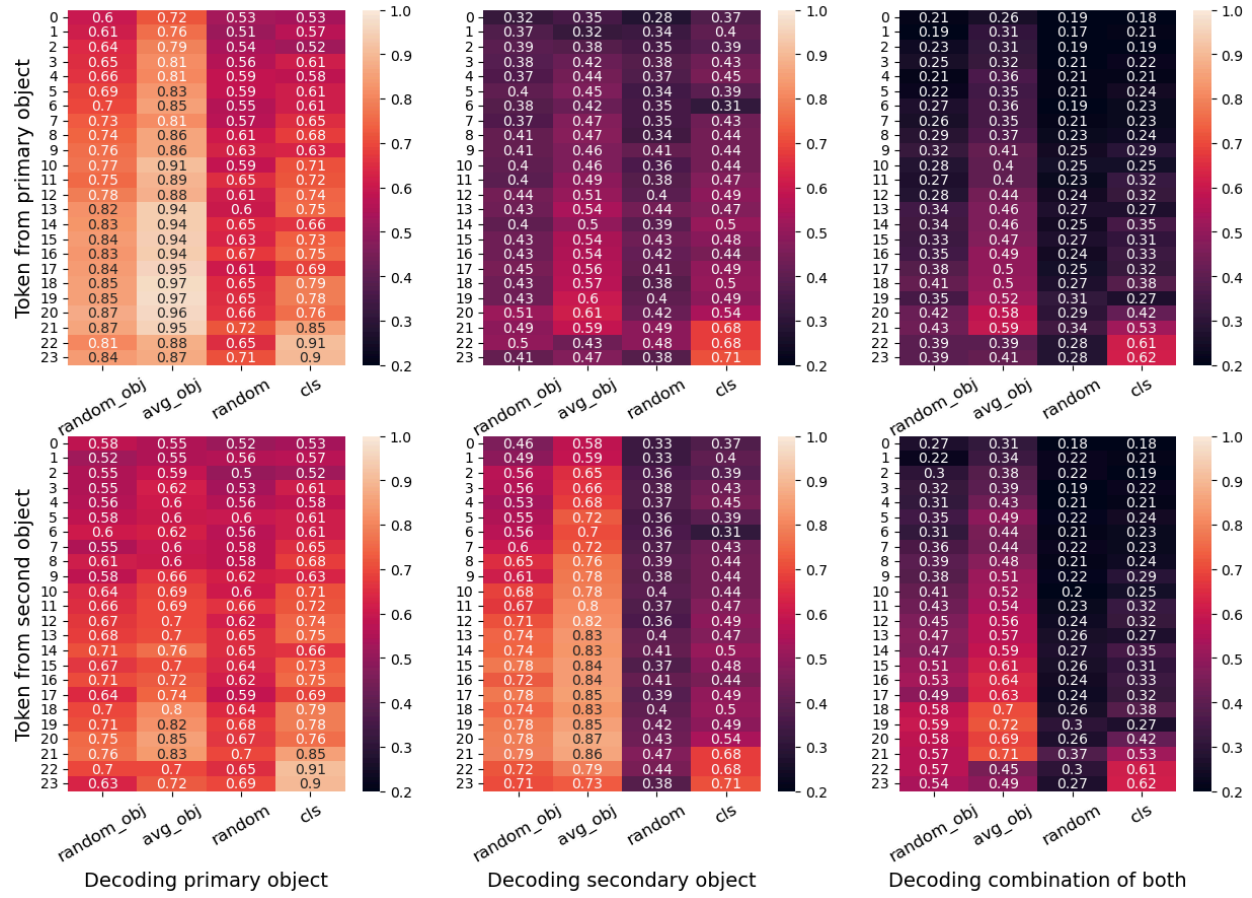
CLIP



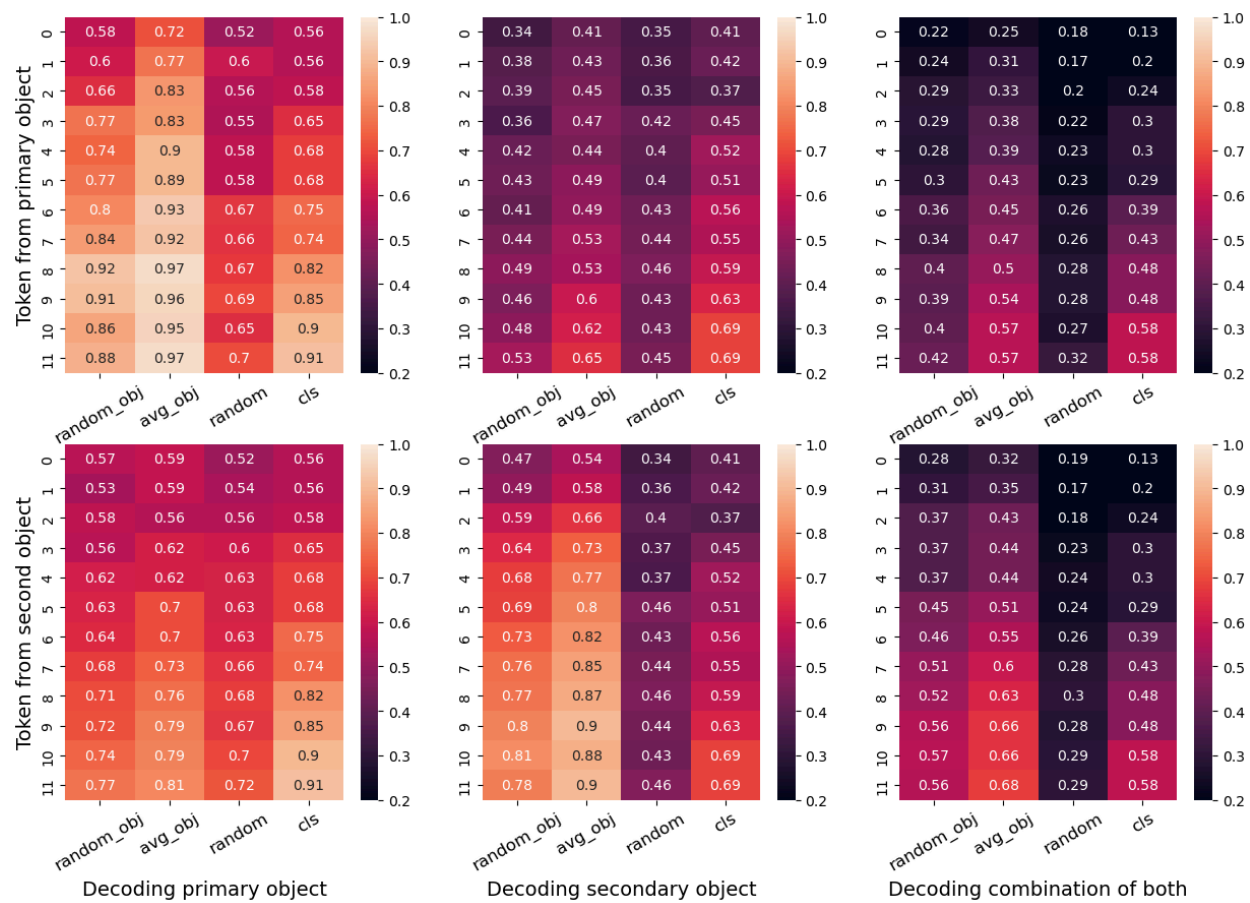
CLIP-Large



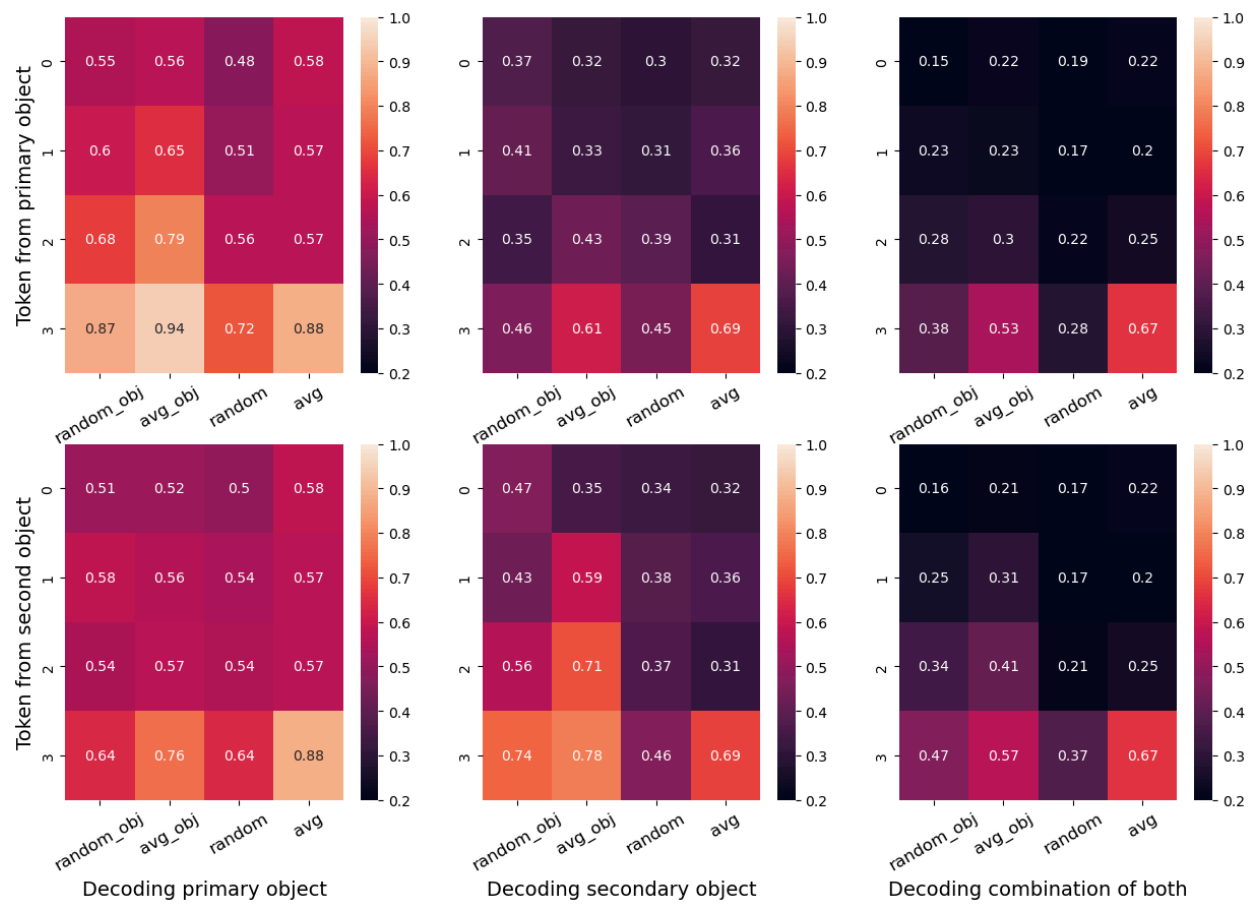
BLIP-Large



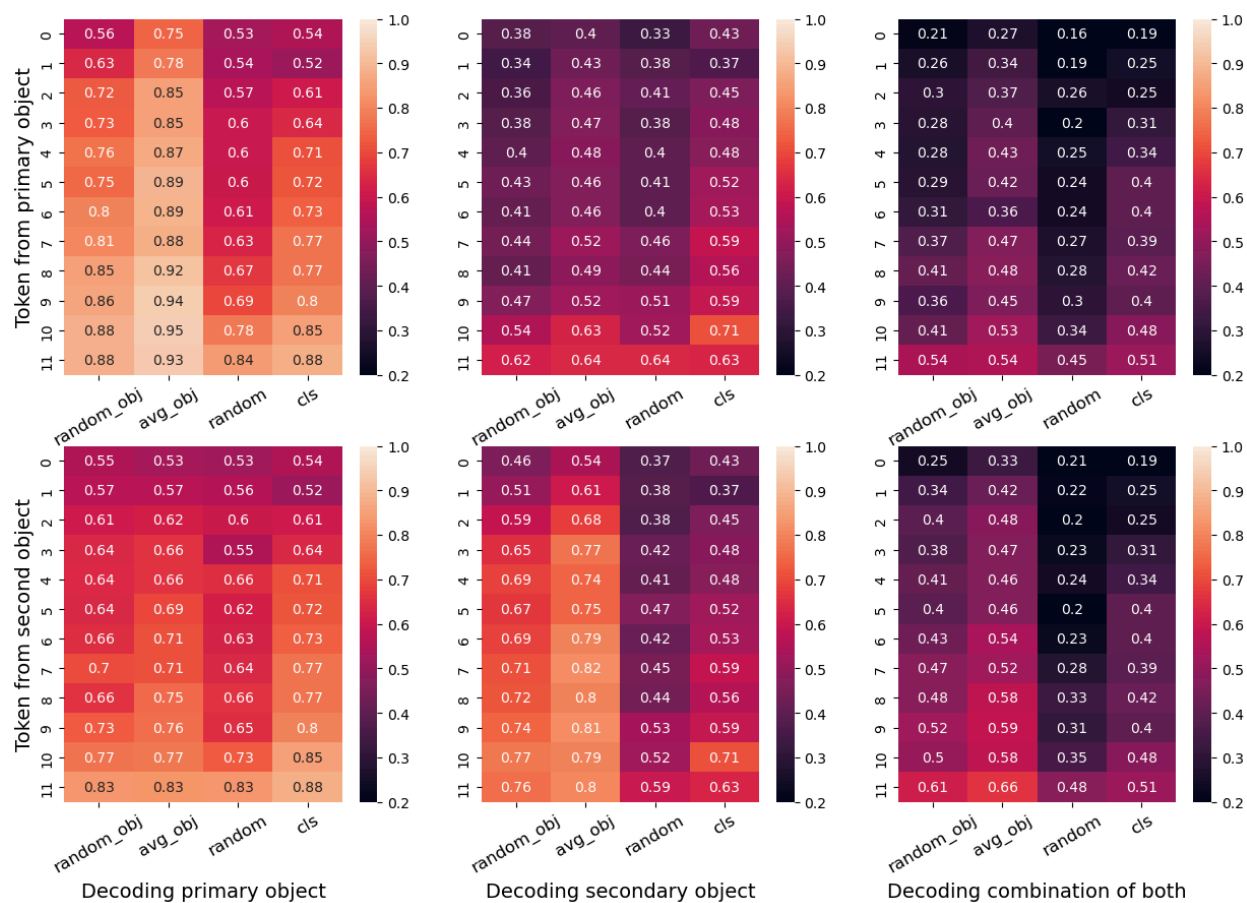
FLAVA



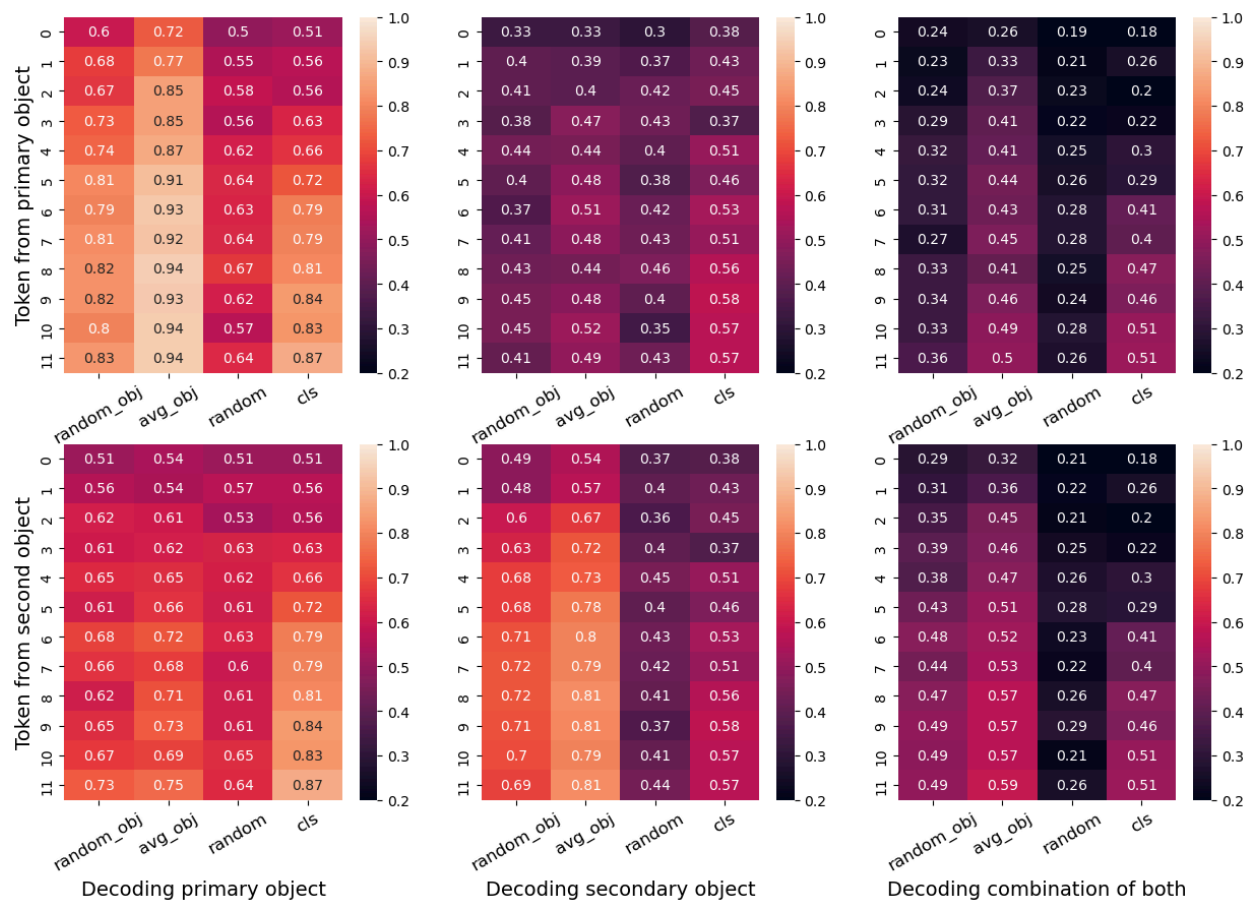
CLIP Renet X4



ViT (Imagenet 21k)



DINO



DINO2

