text to image R@5 (dataset: Sydney) baseline s-clip soft-pl 35 hard-pl basic-al probvlm 30 recall 25 J 20 -15 -10 -0.2 0.4 0.8 1.0 0.0 0.05 0.1 Label ratio