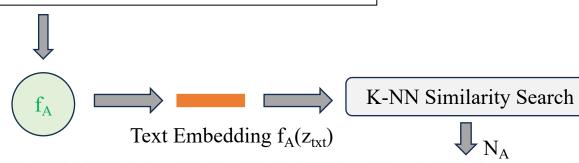
Target Image z_{img}



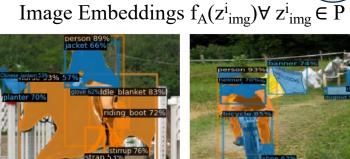
 f_A is trained with z

Detected: 10/11 Precision: 0.22 Recall: 0.91

Image Caption z_{txt}: See Junior riders compete at the hieghts level to improve their skills.



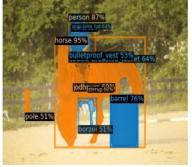


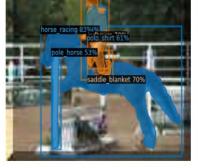




Public Image Set P













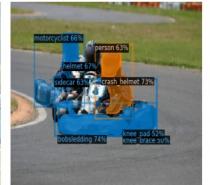
 f_{B}

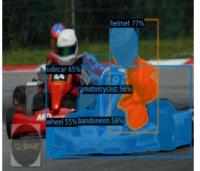
Text Embedding $f_B(z_{txt})$

 \bigvee N_B

Image Embeddings $f_B(z^i_{img}) \forall z^i_{img} \in P$











Precision: 0.08 Recall: 0.18