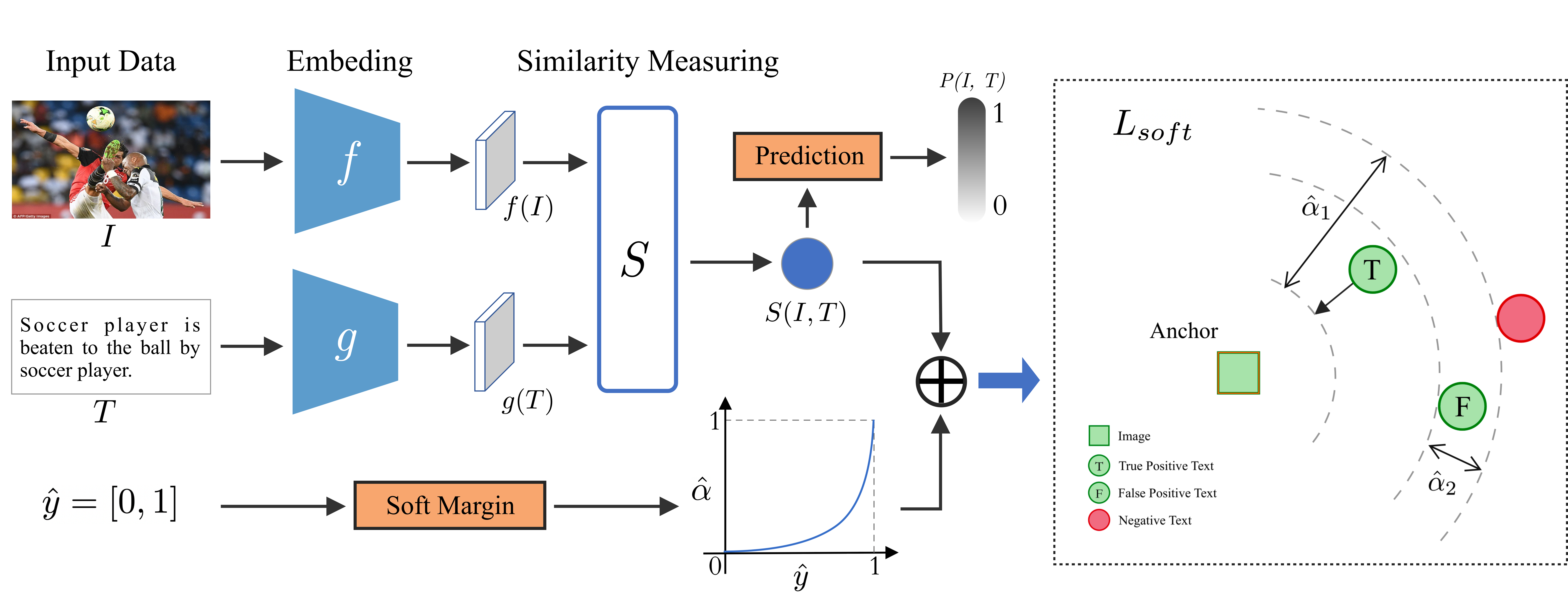


(a) Training pipeline



(b) Robust image-text matching network (A and B)