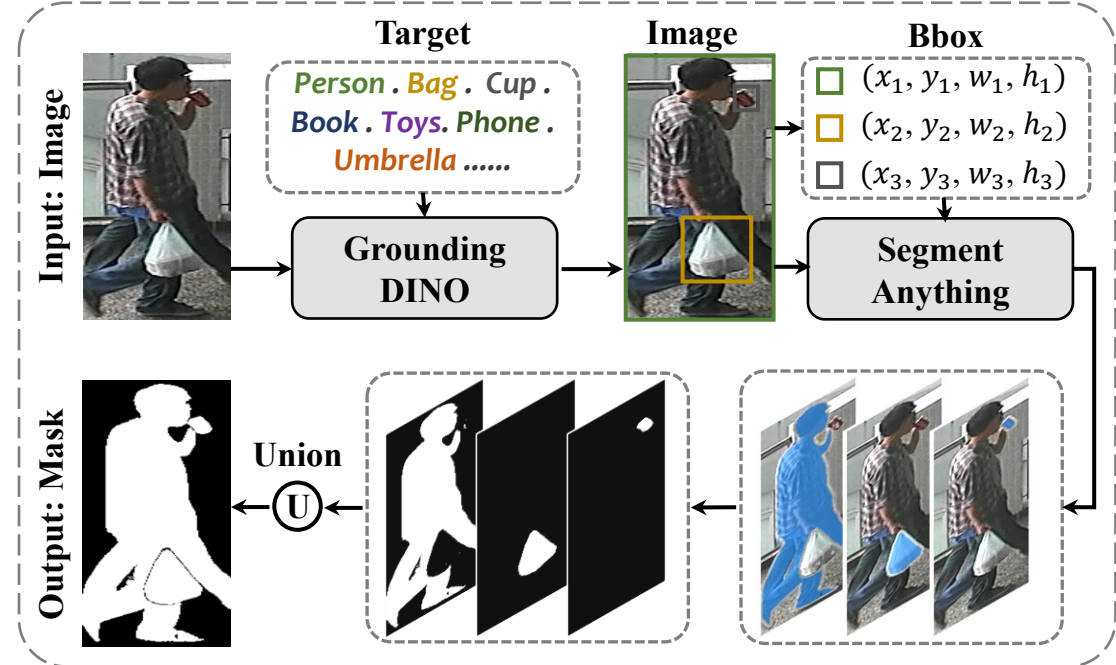
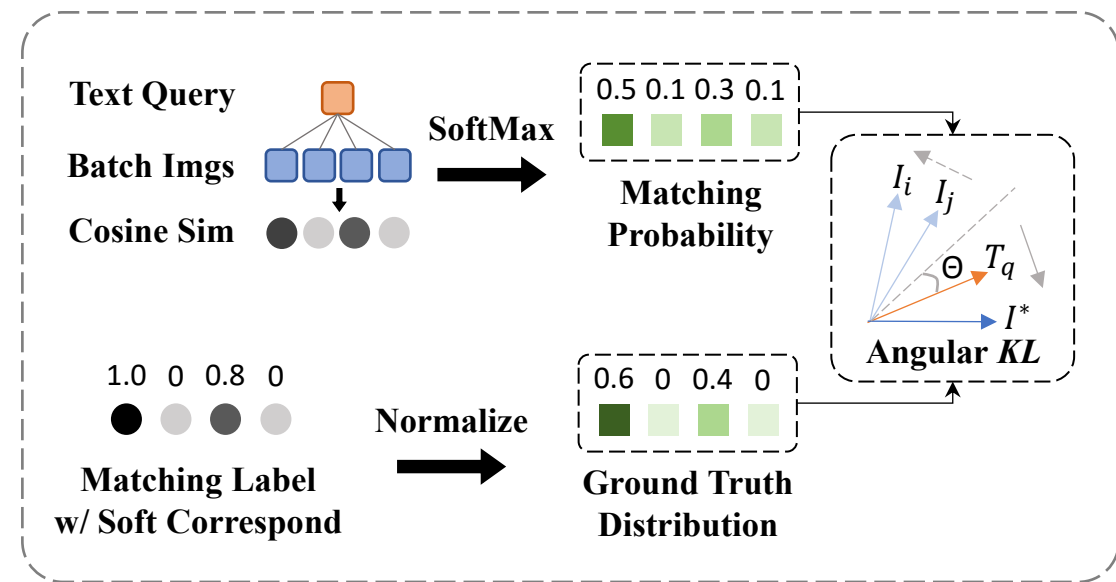


(a) Cross-Modal Object-Centric Learning



(b) Person Mask Generation Pipeline



(c) Soft Angular Distribution Matching