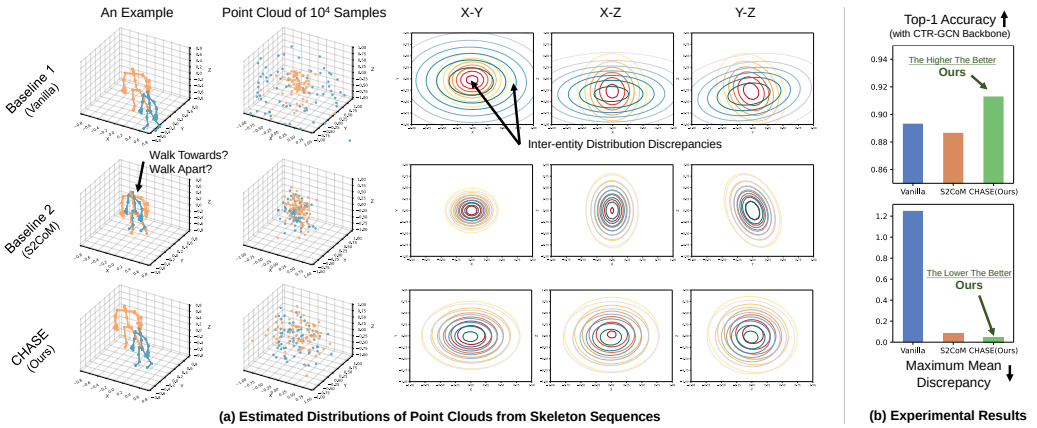

Figure 1: **Inter-entity distribution discrepancies in multi-entity action recognition task.** (a) We delineate three distinct settings: *Vanilla* (a common practice), *S2CoM* (an intuitive baseline approach), and *CHASE* (our proposed method). Column 2 illustrates spatiotemporal point clouds defined by the skeletons over 10^4 sequences. Column 3-5 depict the projections of estimated distributions of these point clouds onto the x-y, z-x, and y-z planes. These projections reveal significant inter-entity distribution discrepancies when using *Vanilla*. (b) The discrepancies observed in *Vanilla* introduce bias into backbone models, leading to suboptimal optimization and poor performance. Although *S2CoM* can reduce these discrepancies, it makes the classifiers produce wrong predictions due to a complete loss of inter-entity information. With the lowest inter-entity discrepancy, our method unbiases the subsequent backbone to get the highest accuracy, underscoring its efficacy.

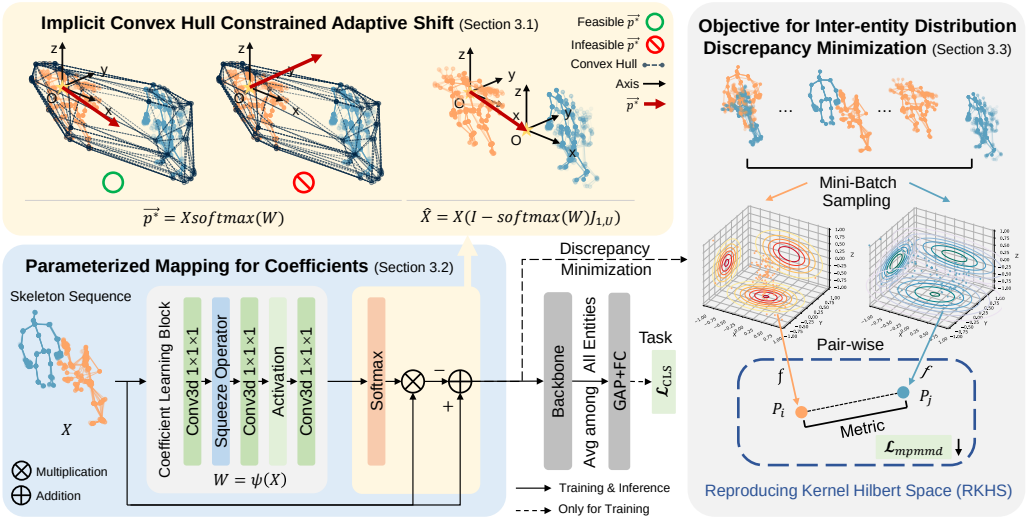

Figure 2: **The overall framework of the proposed CHASE for multi-entity action recognition.** Taking a multi-entity action skeleton sequence as input, CHASE executes an implicit convex hull constrained adaptive shift with the Coefficient Learning Block, implemented as a lightweight backbone wrapper. CHASE also collects pair-wise shifted skeletons within mini-batches, effectively alleviating inter-entity distribution discrepancies by introducing an additional objective.