

## JRDB-Act Metric and Evaluation

We evaluate three sub-tasks on JRDB-Act namely, individual action, social group and social activity detection on the key-frames. Key-frames are sampled every 1 second (every 15 frames starting from 15th frame) in JRDB-Act. We utilize the widely used Mean Average Precision (MAP) on the key-frames of test-set using detected bounding boxes. At inference, for each detected bounding box in a key-frame, a model should predict a set of individual action labels and a social group ID. A set of social activity labels can be inferred for that box by utilizing its social group ID and the predicted individual action labels of that group’s members. The inferred social activity labels for all the bounding boxes of a social group, would be the common individual action labels of the members of that group, the action labels repeated by two or more than two persons in the same group. If the group has a single member, the social activity would be identical to the box’s individual action label.

To clarify the evaluation strategy of the three sub-tasks, we show an example in Fig. 1. On the left we show the ground-truth scenario with 4 social groups indicated by different colors. Each box has a number of action labels indicated by  $A_i$ . On the right, we show the predicted scenario.

For the individual action detection task, the reported mAP considers the true positive cases as (box1, A1), (box2, A4), (box2, A5), (box3, A3), (box3, A7), (box5, A6), (box5, A8), (box6, A6) and the false negative cases as (box1, A7), (box2, A3), (box4, A2) and the false positive cases as (box3, A8), (box7, A6).

For the social grouping task, the reported AP considers true positive cases as boxes 2,3,5,6, false negative case as box4 and false positive cases as boxes 1,7. mAP is also reported as the average AP of groups with 1, 2, 3, 4 and more than 5 members and is used as the main metric for the social group detection challenge.

For the social activity task, the inferred ground-truth

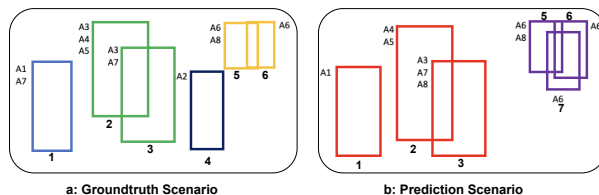


Figure 1: An example of the ground-truth and prediction scenarios for the three sub-tasks of individual action, social group and social activity detection. Matched ground-truth and detected bounding boxes are shown with similar number. The set of actions for each box is indicated by  $A_i$  next to it.

social activity labels for the blue group is A1, A7, for the green group is A3, for the navy group is A2 and for the yellow group is A6. Social activity predictions are inferred by utilizing the predicted social groups and individual actions. For the red group we consider no predicted social activity label for boxes 1, 2, 3, for the purple group A6 is considered as the predicted social activity label for boxes 5, 6, 7. mAP1 takes into account the social activity label as well as the group membership for each box and is used as the main metric for the social activity detection challenge. Thus, it considers true positive cases as (box5, A6), (box6, A6) and false negative cases as (box4, A2), (box1, A1), (box1, A7), (box2, A3), (box3, A3) and false positive cases as (box7, A6). mAP2 is also calculated similar to the individual action detection task.

For all the sub-tasks a detailed result is reported per category and per sequence. In order to incorporate the annotated difficulty level for each annotated label in JRDB-Act, we exclude the labels tagged as difficult or impossible and their corresponding predicted label from the evaluation in each sub-task. Precisely, for mAP in individual action detection task, AP and mAP in social group detection and mAP2 in social activity detection tasks, ground-

truth labels tagged as difficult or impossible and their corresponding predicted labels are excluded from the evaluation. For mAP1 in social activity detection task, the excluded ground-truth social activity labels are either tagged as difficult or impossible or their corresponding box's social group is tagged as difficult or impossible.