A Hybrid Data-Centric Framework for Thermal Multiple-Object Tracking with Complex Motion Patterns

Supplementary Material

The overall structure of the supplementary material is listed as follows:
▷Sec. A: Metric Details for Evaluation.
▷Sec. B: Training Progress.
▷Sec. C: Visualization Results.

A. Metric Details

In multi-object tracking, MOTA, IDF1, IDP, and IDR are related but emphasize different aspects of performance.
MOTA (Multiple Object Tracking Accuracy) Measures overall tracking errors over all frames: it penalizes missed detections, false positives, and ID switches in one aggregate score (closer to 1 or 100% is better).
IDP (Identification Precision) Precision in terms of identities:

$$\text{IDP} = \frac{\text{IDTP}}{\text{IDTP} + \text{IDFP}}$$

where IDTP are identity true positives (correctly identified detections) and IDFP are identity false positives (detections assigned to the wrong ID or spurious).
IDR (Identification Recall) Recall in terms of identities:

$$\text{IDR} = \frac{\text{IDTP}}{\text{IDTP} + \text{IDFN}}$$

where IDFN are identity false negatives (ground-truth identities that are not correctly tracked).
IDF1 (ID F1 score) Harmonic mean of IDP and IDR, summarizing identity preservation quality in a single number:

$$\text{IDF1} = \frac{2 \cdot \text{IDTP}}{2 \cdot \text{IDTP} + \text{IDFP} + \text{IDFN}}$$

High IDF1 indicates that the tracker both assigns IDs accurately (high IDP) and maintains them over as much of each trajectory as possible (high IDR).

B. Training Progress

For detector training, we evaluated multiple input resolutions (640, 960, 1280, 1600, 1920, and 2560) to determine an appropriate trade-off between target visibility and computational cost. Lower resolutions accelerate inference but may miss small or distant thermal pedestrians, while higher resolutions preserve finer object details at higher runtime and memory cost. Based on validation performance, 1600 and 1920 provided the most reliable detection accuracy, so we used these two resolutions in subsequent testing. For re-identification training, we evaluated TMOT, VT-MOT, and MTMMC both individually and in combination to analyze dataset contribution and cross-domain generalization. Training on a single dataset yielded more limited identity diversity, whereas combining all three datasets improved variation in viewpoint, scene context, and thermal appearance. Therefore, for the final Re-ID setting, we trained with

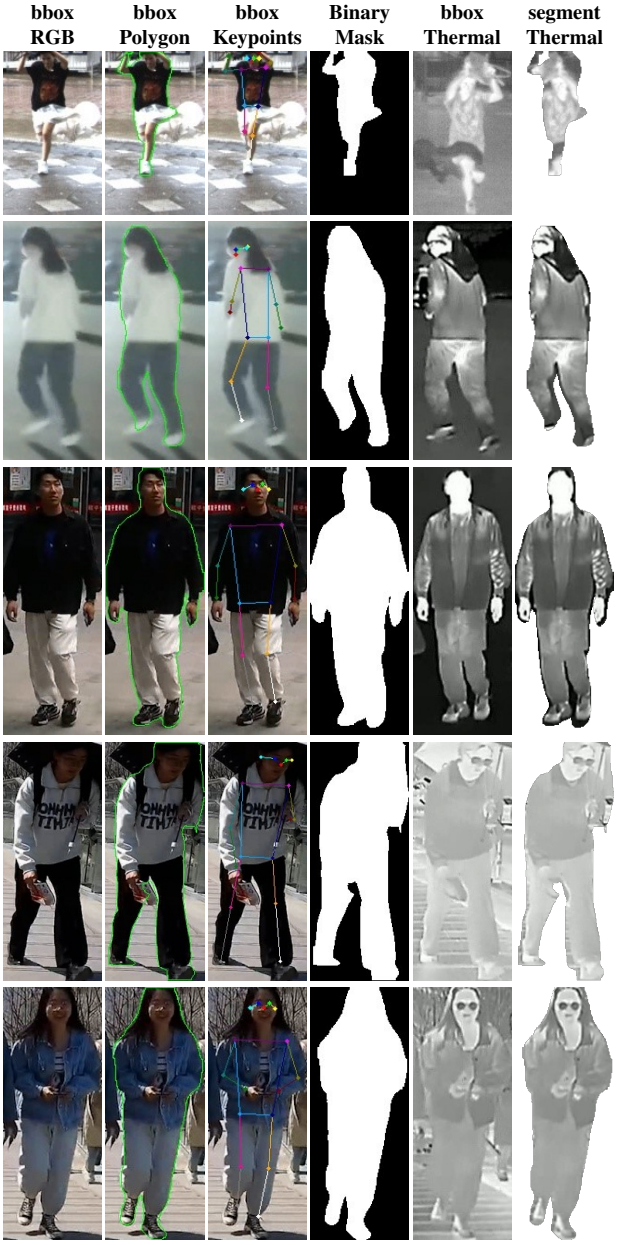

Figure 6. Visualization Result.

all three datasets to obtain stronger and more stable identity features.

C. Visualization Results

As shown in Fig. 6, the training dataset includes examples of cropping, segmentation, and keypoint detection.