LiDAR-based Person Re-identification

Supplementary Material

A. LReID

We provide detailed instructions about our data acquisition system and scenes to supplement the data collection process of LReID.

A.1. Data Acquisition System

We collect LReID with 4 acquisition nodes, each including a Livox Mid-100 LiDAR and an industrial camera, as shown in Fig. 1. The acquisition nodes are portable and powered by their carried power supplies. The Livox Mid-100 LiDAR continuously scans the scene at about 30,000 points per frame for a horizontal view of 100° and a vertical view of 40° toward the front, providing accurate depth information. The Livox Mid-100 LiDAR is a low-cost option, priced at around $1,400 [1], which expands the practical applications of LiDAR-based person ReID.

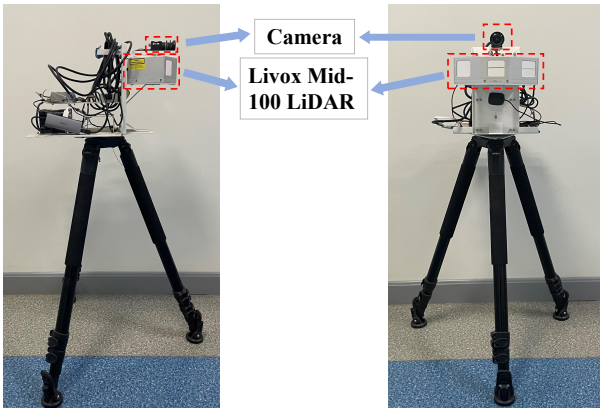


Figure 1. The acquisition node.

A.2. Data Acquisition Scenes

With the aim of improving the generalization and robustness of models trained on LReID, as well as expanding the data coverage, we collect LReID in various outdoor scenes: a crossroad and a square, capturing different time periods and weather conditions, as shown in Fig. 2.

B. Feature Visualization

To visually demonstrate the superiority of our proposed ReID3D, we use t-SNE to visualize the feature distributions of LReID test set extracted by different methods.

B.1. Comparison with Camera-based Method

For the purpose of comparing the disparities between different modalities, we visualize the feature distributions obtained from TCLNet [**?**] that achieves the highest accuracy among camera-based methods and ReID3D without pre-training (B-ReID3D), as shown in Fig. 3 and Fig. 4. We can observe that: (1) It is difficult to distinguish pedestrians in low light condition for TCLNet, with their features entangled in the feature space. However, their features extracted by B-ReID3D are dispersed in the feature space, like the features of pedestrians under normal light. (2) TCLNet is prone to confusion between individuals with similar appearances, as illustrated by the four examples in Fig. 3. In contrast, their features extracted by B-ReID3D display excellent discriminability. This is because camera-based methods primarily focus on learning the visual appearance of pedestrians, while ReID3D is able to acquire intrinsic features, such as body shape and gait. (3) B-ReID3D also exhibits mistakes in distinguishing certain pedestrians, which are easily discernible for camera-based methods. Therefore, methods based on the two modalities have complementarities.

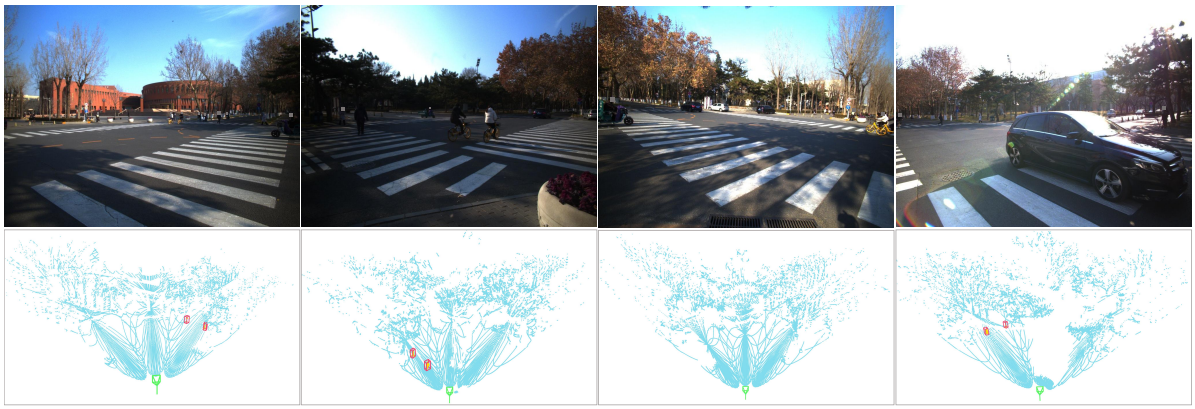
B.2. Effect of Pre-training

In order to demonstrate the effectiveness of pre-training with multiple tasks, we visualize the feature distribution obtained from ReID3D, as shown in Fig. 5. Compared to Fig. 4, the feature distributions extracted by ReID3D and B-ReID3D exhibit overall similarity. However, ReID3D demonstrates better discriminative ability when dealing with individuals with similar features in local regions of the feature space. We attribute this to the fact that LReID-sync improves the diversity of the actual training data, and the pre-training with multiple tasks enhances GCEE in learning features of human body shape.

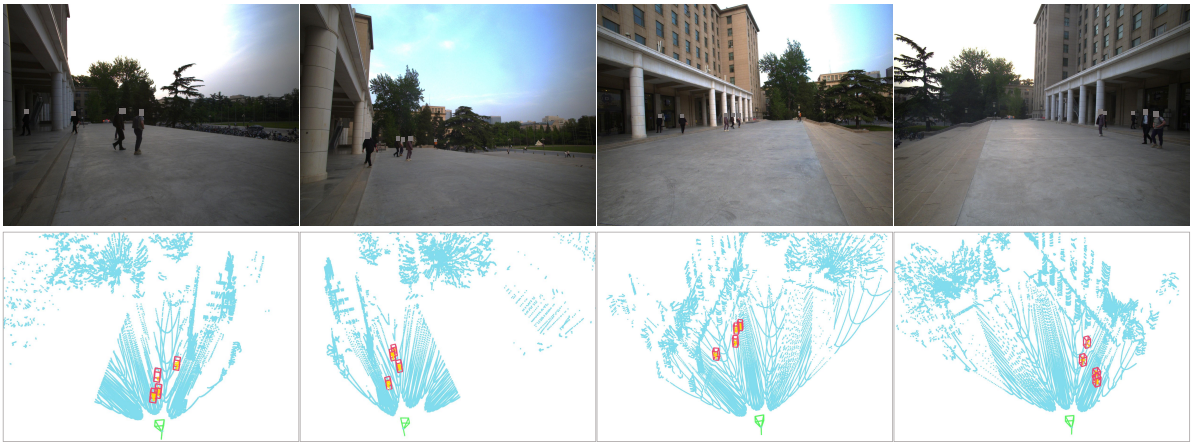
B.3. Comparison with Point Transformer

To compare different point cloud encoders, we replace GCEE in B-ReID3D with Point Transformer [**?**] and visualize its feature distribution, as shown in Fig. 6. Compared to Fig. 4, the features extracted by Point Transformer from samples of the same pedestrians are not effectively aggregated, and there are more scattered feature points distributed in the feature space. Additionally, the feature distances from different pedestrians lack sufficient distinctiveness. Therefore, Point Transformer exhibits a lesser capability in extracting features of pedestrian point clouds compared to GCEE.

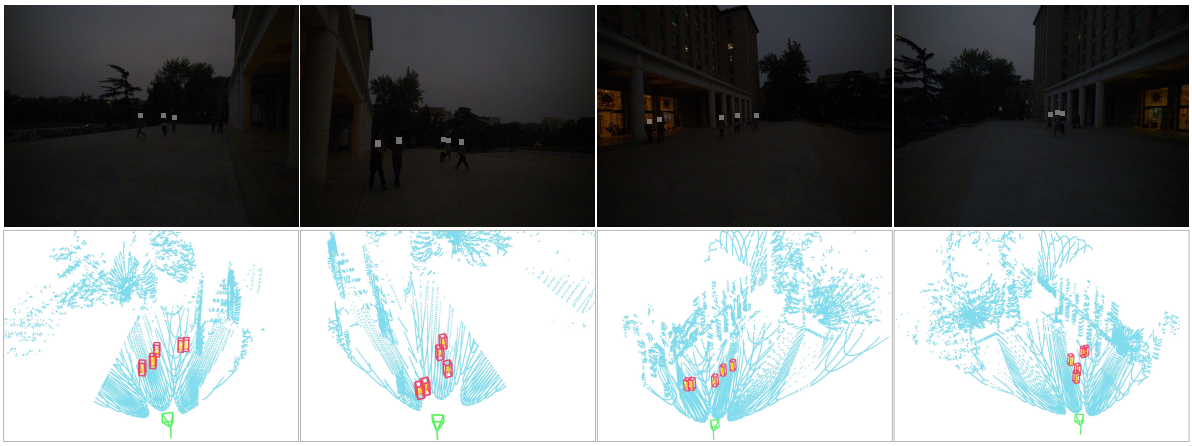
[1] Purchase link



(a) Crossroads scene.



(b) Square scene in normal light condition.



(c) Square scene in low light condition.

Figure 2. Data acquisition samples in different scenes and lighting conditions.

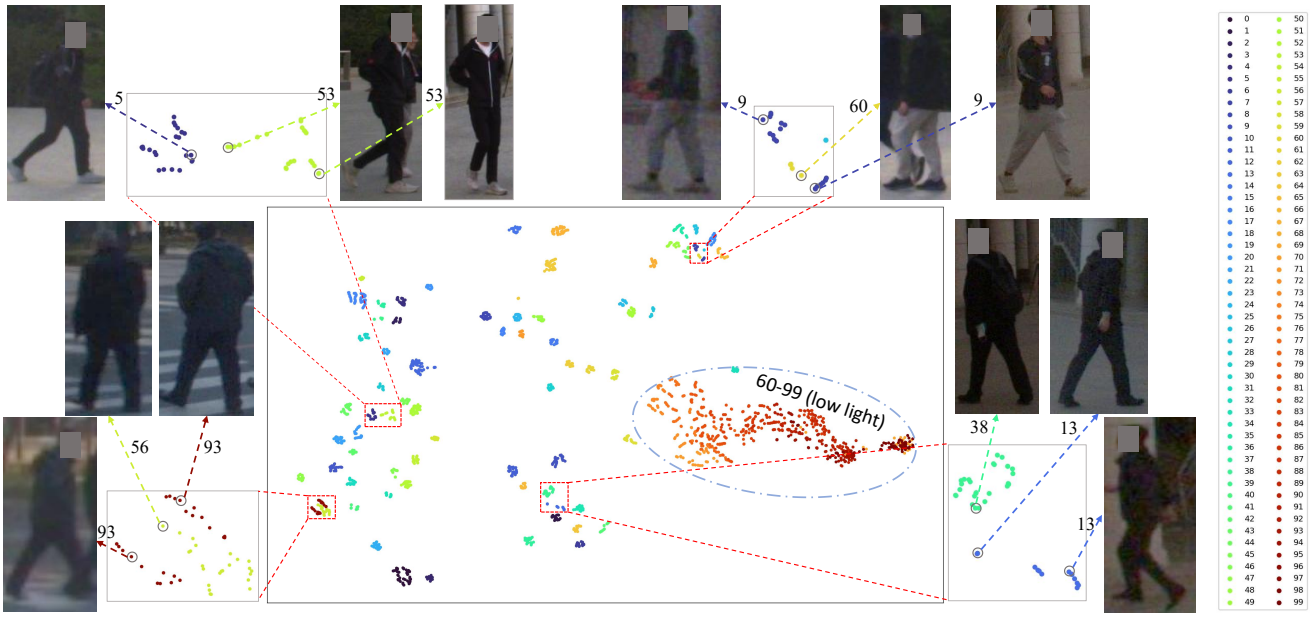


Figure 3. The feature distribution of LReID test set extracted by TCLNet [**?**]. Each point corresponds to a feature vector output by TCLNet and distinct colors represent different individuals in LReID test set. Each image is selected from the corresponding sample of the pedestrian. It is challenging to distinguish the features of pedestrians with similar appearances and those under low light.

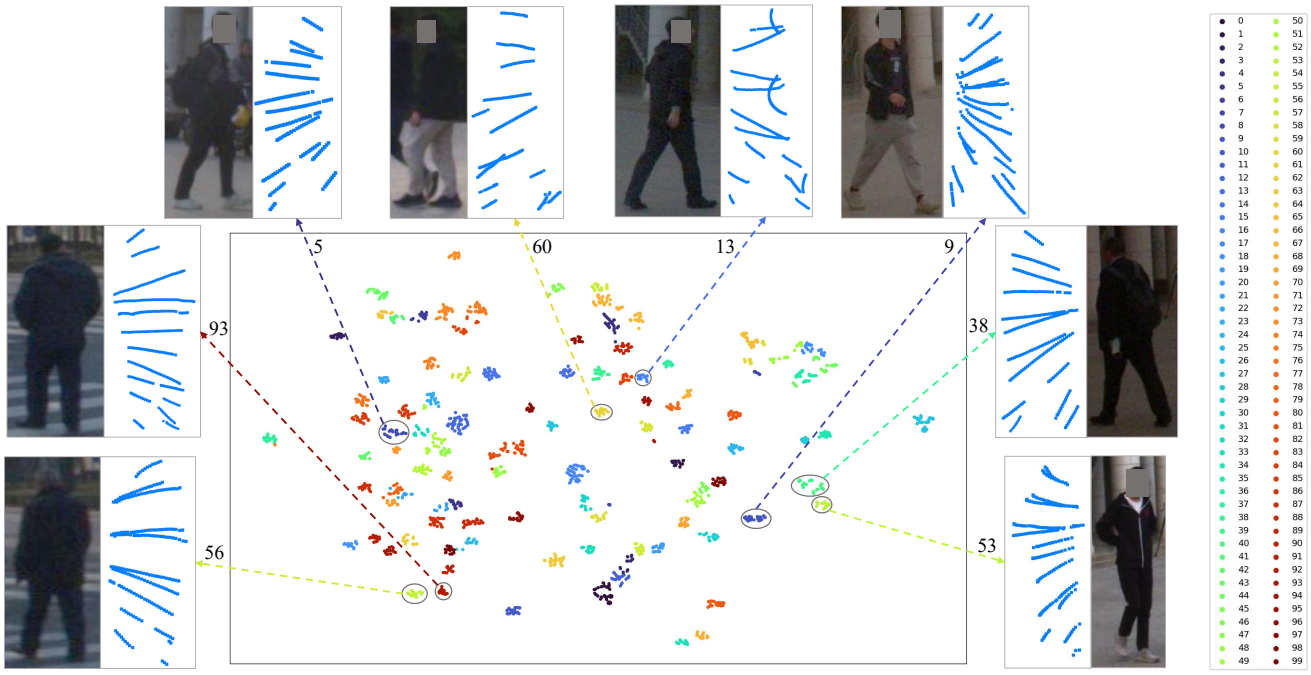


Figure 4. The feature distributions of LReID test set extracted by ReID3D without pre-training (B-ReID3D). The images and corresponding point clouds are from the same pedestrians in Fig. 3. The features of pedestrians with similar appearances and those under low light are dispersed in the feature space, making them distinctive.

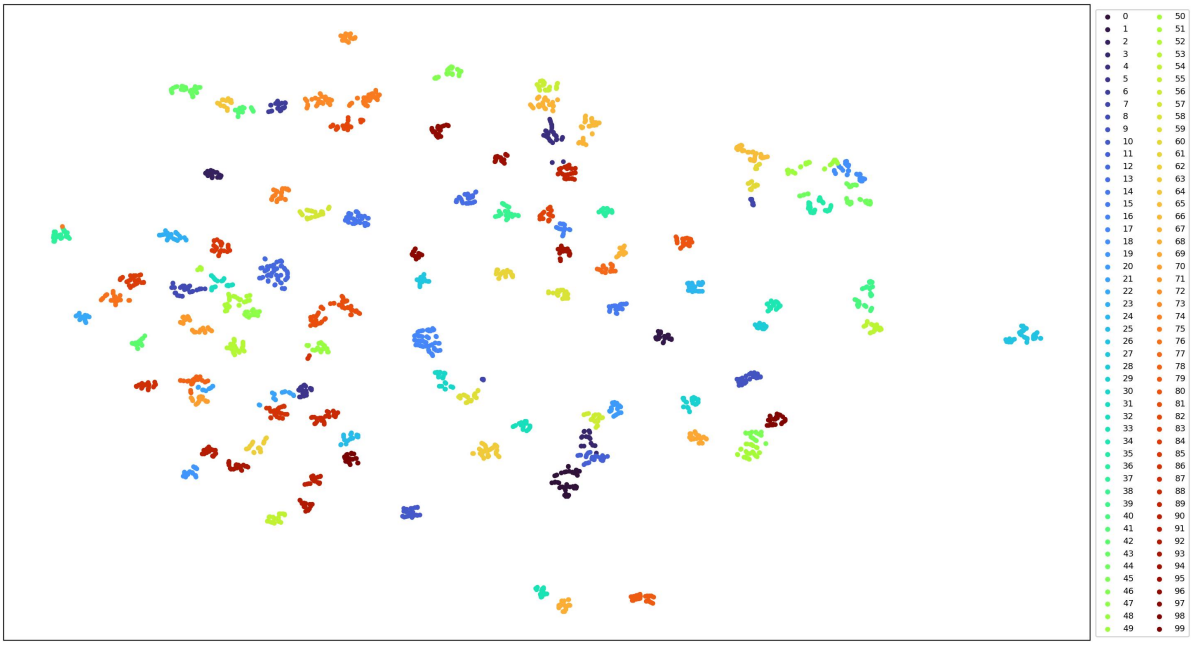


Figure 5. The feature distribution of LReID test set extracted by ReID3D.

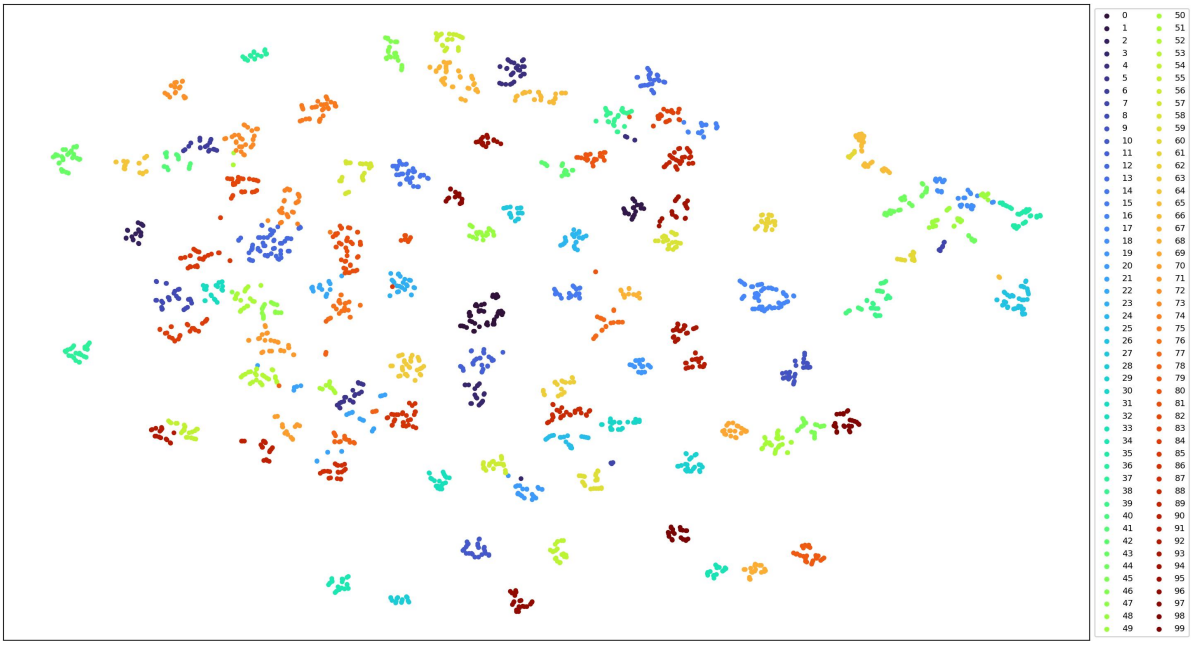


Figure 6. The feature distribution of LReID test set extracted by Point Transformer [**?**] with the same temporal fusion module.