Our method is divided into 3 steps as a whole. First process the data, then extract the features of each sequence through the feature extraction network, and finally optimize the results.

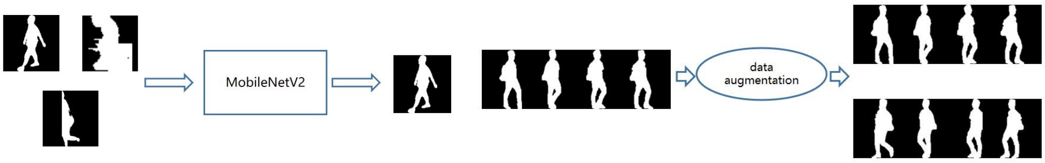

Fig 1

1. Data Process

It is observed that the training data and the test data contain low-quality pictures. So we manually selected 1103 pictures (including 212 low-quality pictures) as the training set to train a binary classification network MobileNetV2 [1]. Use the classification network to process the data and remove the pictures that are predicted to be of low quality, as shown on the left side of Figure 1.

The data set contains gait in different directions. In order to increase the robustness of the model, we use data augmentation (flip the images in each sequence horizontally). In this way, the posture of walking from left to right before is walking from right to left, as shown on the right side of Figure 1. At the same time the data set has also been expanded by 2 times.

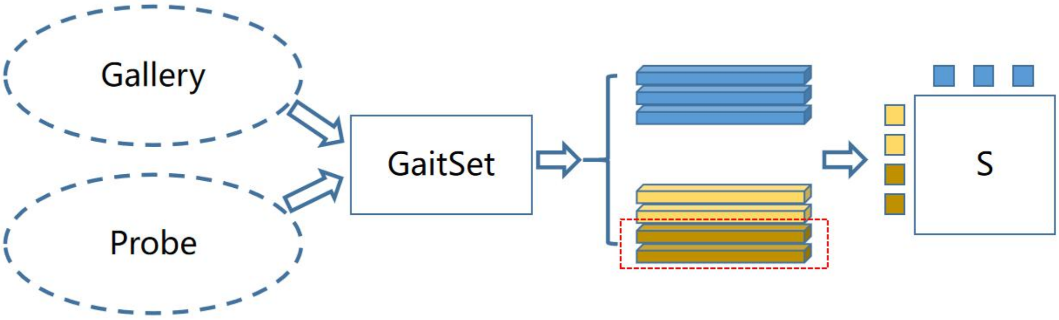

Fig 2

2. Feature Extraction

In the selection of the feature extraction network, considering the removal of low-quality images in the first step, the temporal will be destroyed to a certain extent, so we chose GaitSet [2] as the feature extraction network. At the same time, the public GaitSet training model file[3] on the OUMVLP dataset is used as the network initialization. Fine-tune 11w rounds on the competition data set.

In the feature extraction stage as shown in Figure 2, each gait sequence passes through the feature extraction network to obtain a feature vector of 15872 length. Use the feature vector to calculate the similarity matrix, where the row represents the sequence in the probe, the column represents the sequence in the gallery, and the metric uses Euclidean distance. Only the probe is enhanced by the sequence(red box in Figure 2), so that a similarity matrix S[2*P][G] will be obtained. P and G represent the number of sequences in the probe and the gallery, respectively.

3. Results Optimize

On the basis of the similarity matrix S, the k-reciprocal encoding[4] in pedestrian re-identification is used. After reordering, S’[2*P][G] is obtained. For the sequence in the probe, the closest one among all the galleries is taken as the predicted label. Due to the data enhancement of the probe, there will be 2 predictions for each sequence, and the one with the smallest distance is taken as the prediction label.

Experiment

The data used in the following experiments have been cleaned. All the experiments are based on the last one. The results are shown in Table 1.

	score
baseline(GaitSet)	54.6
+train data augmentation	57.4
+pre_trained	66.4
+k-reciprocal	76.8
+probe data augmentation	79.9

Table 1

Reference

[1]Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.

[2] Chao H, He Y, Zhang J, et al. Gaitset: Regarding gait as a set for cross-view gait recognition[C]//Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 8126-8133.

[3] https://github.com/AbnerHqC/GaitSet

[4] Zhong Z, Zheng L, Cao D, et al. Re-ranking person re-identification with k-reciprocal encoding[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1318-1327.