Temporal Proposal Module for Human Identification at a Distance

Tao Ding[1], Yuchao Yang[2], Shuiwang Li[3], and Qijun Zhao[4]

[1] Sichuan University, 18308460613@163.com
[2] Sichuan University, yuchao_yang@126.com
[3] Sichuan University, lishuiwang0721@163.com
[4] Sichuan University, qjzhao@scu.edu.cn

1 Team details

Team name: BRiLiant
Affiliation: National Key Laboratory of Fundamental Science on Synthetic Vision, College of Computer Science, Sichuan University
Final evalution score: 54.1%

2 Dataset and preprocessing

The training set contains 500 subjects with 10 video sequences for each one. The test set contains 514 subjects which are different from the training set. In the test set, the gallery includes one sequence of each subject, and the probe set consists of the rest sequences. The dataset contains views and walking conditions, such as walking with bag and wearing coat and so on. Compared with the CASIA-B dataset, this dataset is more challenging because it contains sequences that subjects may stop when they are walking. In addition, there are quite a lot images of low quality, including images without subject, images with multi-subjects and unclear silhouette. These problems may be caused by the errors of human body detection and segmentation algorithms.
To deal with this, we trained the MobileNetV2[1] as a classifier to select high quality images from an input sequence in both training and test phases. Specifically, from the training dataset we mannually selected 1103 images, out of which there are 212 images of low quality. Fig. 1 is an illustration of this process. Since MobileNetV2 is a lightweight model, this process is very fast. Finally, the number of competition dataset was reduced by 147430.

3 Method

Our model is based on the GaitSet[2] model. In view of that input is the silhouette image of human body and, more importantly, that the gait sequence may be discontinuous after low quality images are excluded, we believethat the GaitSet model, which takes input images as a set rather than a sequence, fits in with the

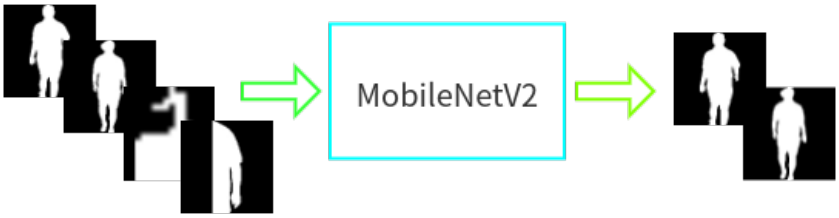

Fig. 1. Illustration of the proposed process for selecting images of high quality from input sequences.

task very well. In preprocessing, we align the images so that each subject is at the image center and then resize the images to 64x64. We use a larger number of feature channels in the backbone of the GaitSet model to enlarge the model capacity for better feature representation and more powerful learning ability. During training, 30 frames are randomly selected from each input sequence and the model is updated by back propagating the triplet loss.

4 Other tries

The attention mechanism in the set pooling (SP) module of the GaitSet model occupies large amount of GPU memory but with less than 1% gain of recognition accuracy. So the attention mechanism of GaitSet is not used in our model. We used the CASIA-B dataset to enlarge our training data but found little improvement. In our experiment, the best number of frames as input is 30. We also tried the GaitPart[3] model, but the best recognition rate the model could get on the competition dataset is 38%, which is obviously worse than that of the GaitSet model. The experimental results are summarized in Table 1.

Table 1. Tried experiments conducted on the competition dataset. GaitSet ({32, 64, 128}) denotes the original GaitSet model with ({32, 64, 128}) being the channel numbers of the first three layers. And 'MobileNetV2 Selection' is the module used to select images of high quality from input sequences.

Method	Rank-1 Accuracy
GaitSet ({32, 64, 128})	41.043%
GaitSet ({32, 64, 128})	50.348%
GaitSet ({32, 64, 128})+SP Attention	50.522%
GaitSet ({32, 64, 128})+MobileNetV2 Selection	54.143%

5 Other details

1) Language and learning framework
   Python and Pytorch.
2) Hardware used and complexity of the method
   GPU: Titan X (12G) 2
   FLOPS: 216.3 GMac
   Params: 26.3 M
   Training time:24hours

6 References

References

1. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. (2018) IEEE/CVF Conference on Computer Vision and Pattern Recognition.
2. Chao, H., He, Y., Zhang, J., Feng, J.: Gaitset: Regarding gait as a set for cross-view gait recognition. (2019) Proceedings of the AAAI Conference on Artificial Intelligence.
3. Fan, C., Peng, Y., Cao, C., Liu, X., He, Z.: Gaitpart: Temporal part-based model for gait recognition. (2020) IEEE/CVF Conference on Computer Vision and Pattern Recognition.