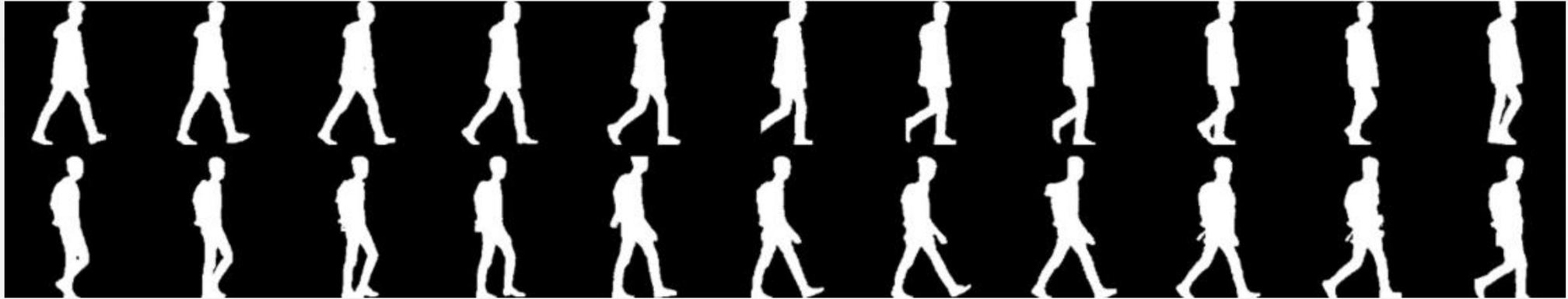


Multi-grid Spatial and Temporal Feature Fusion for Human Identification at a Distance



Panfeng Zhang^{1#}, Zhiqiang Song^{1#}, XiangleiXing^{1*}

¹Harbin Engineering University, School of Intelligent Systems
Science and Engineering

Speaker:Zhiqiang Song

★ Result

User	Accuracy
Beibeilin	63.0
brl	54.1
panfengzhang(ours)	53.4
ctsu-ca	51.5

★ **The main difficulties**

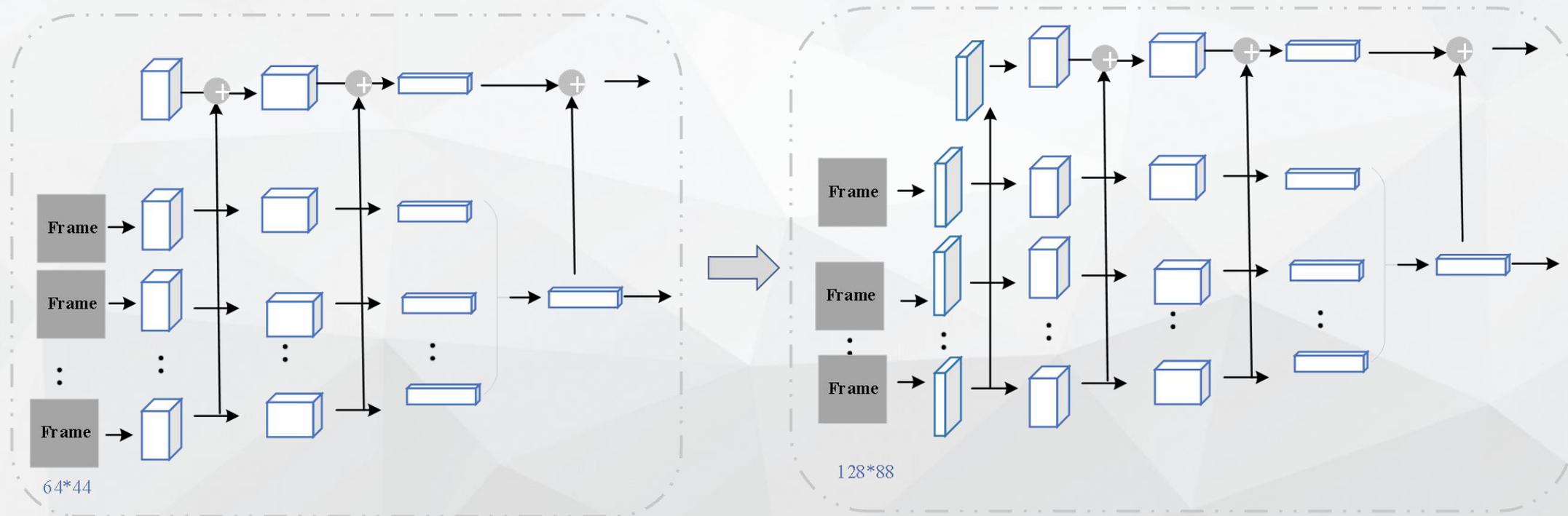
- ★ The given data sets have great complexity cause gait silhouettes from diverse camera's viewpoint and different clothes conditions are mixed in one gait sequence,even ignoring the oder information .
- ★ Each person has only one gait sequence in the gallery set, but there are numbers of unlabeled gait sequences in the probe set.

★ GaitSet

- ★ First, each silhouette is sent into CNN to to extract frame-level features independently.
- ★ Second, an operation called Set Pooling is used to aggregate frame-level features into a single set-level feature this operation is applied on high-level feature maps instead of the original silhouettes, it can preserve more spatial and temporal information.
- ★ Third, a structure called Horizontal Pyramid Mapping is used to map the set-level feature into a more discriminative space to obtain the final representation.

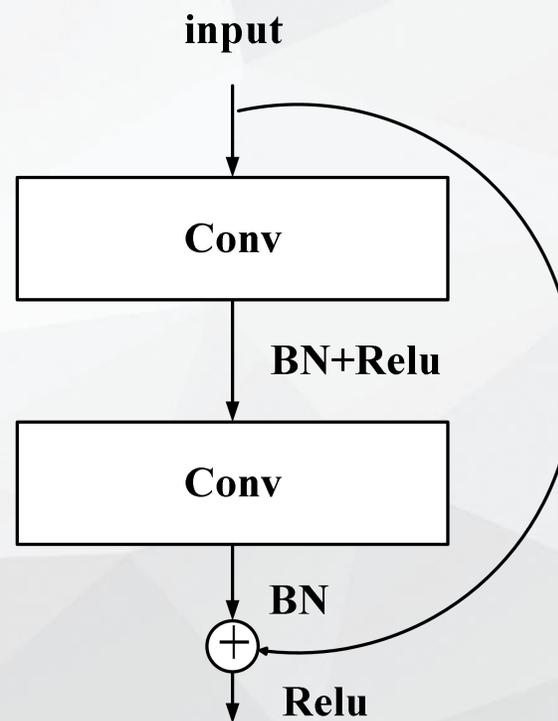
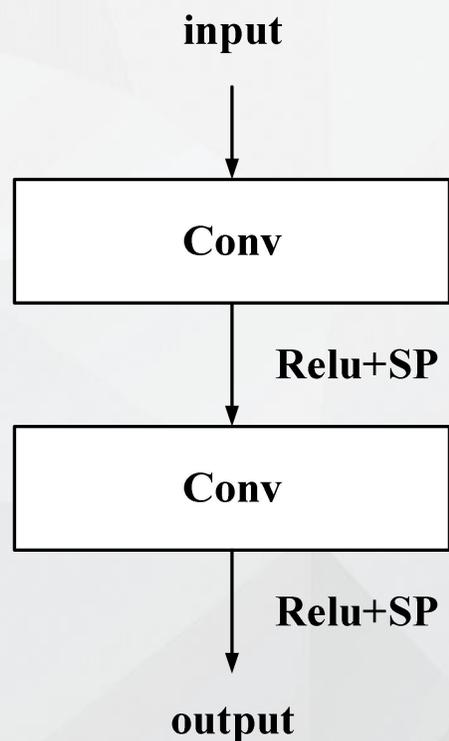
★ Our method

Considering the complexity of the data set, we believe that deepening the backbone network properly will obtain deeper level spatiotemporal information and achieve higher recognition accuracy. This will be justified by the next experiment.

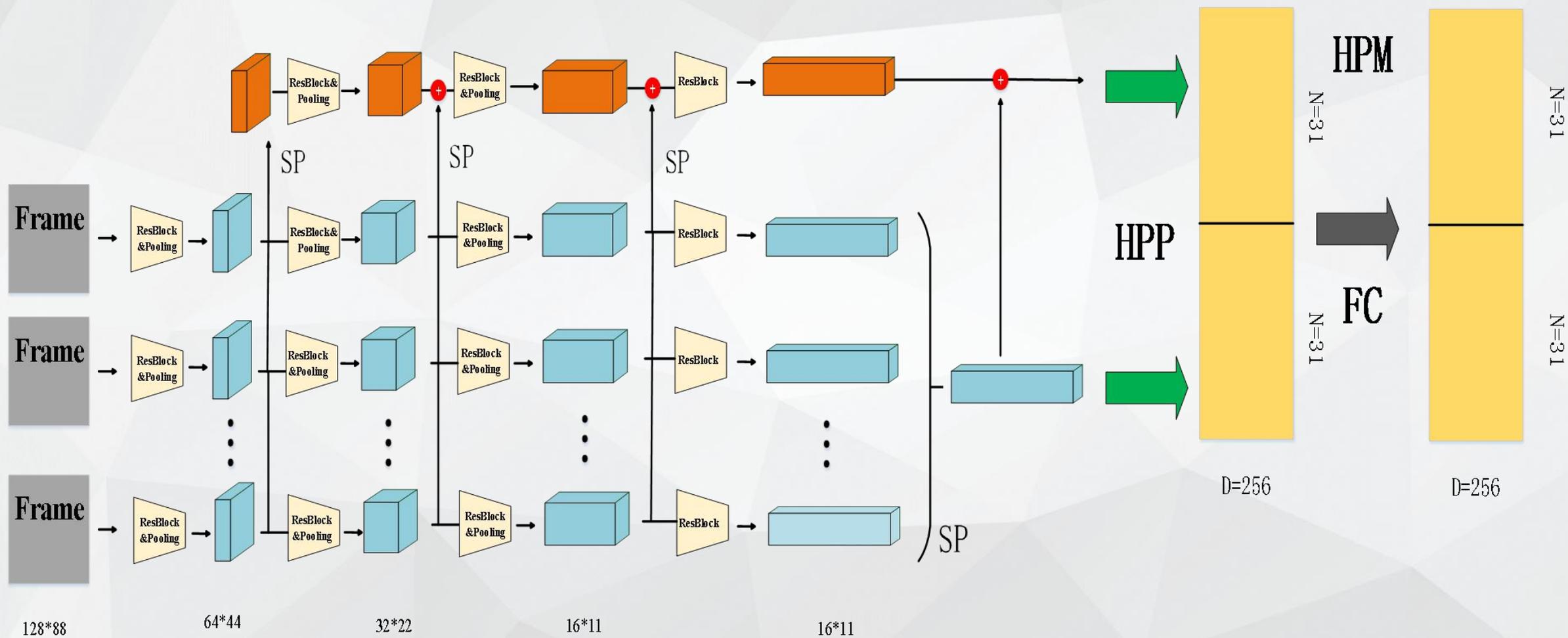


★ Our method

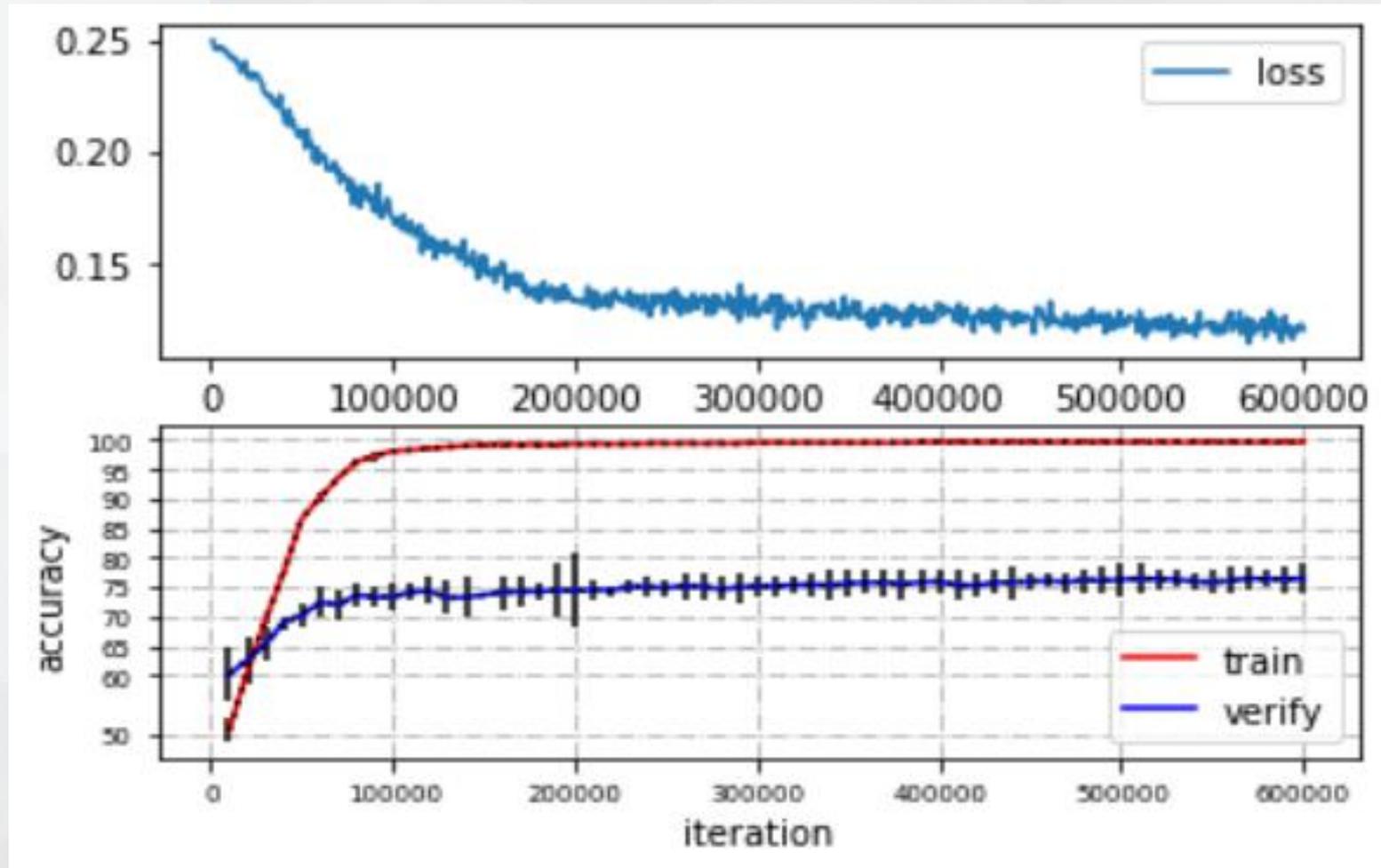
Generally, the deeper the network structure is, the more information we can obtain. To train the deeper network easier and more stable, and to work on the smooth solution space, we employ the ResNet as our backbone.



★ The frame work of our method

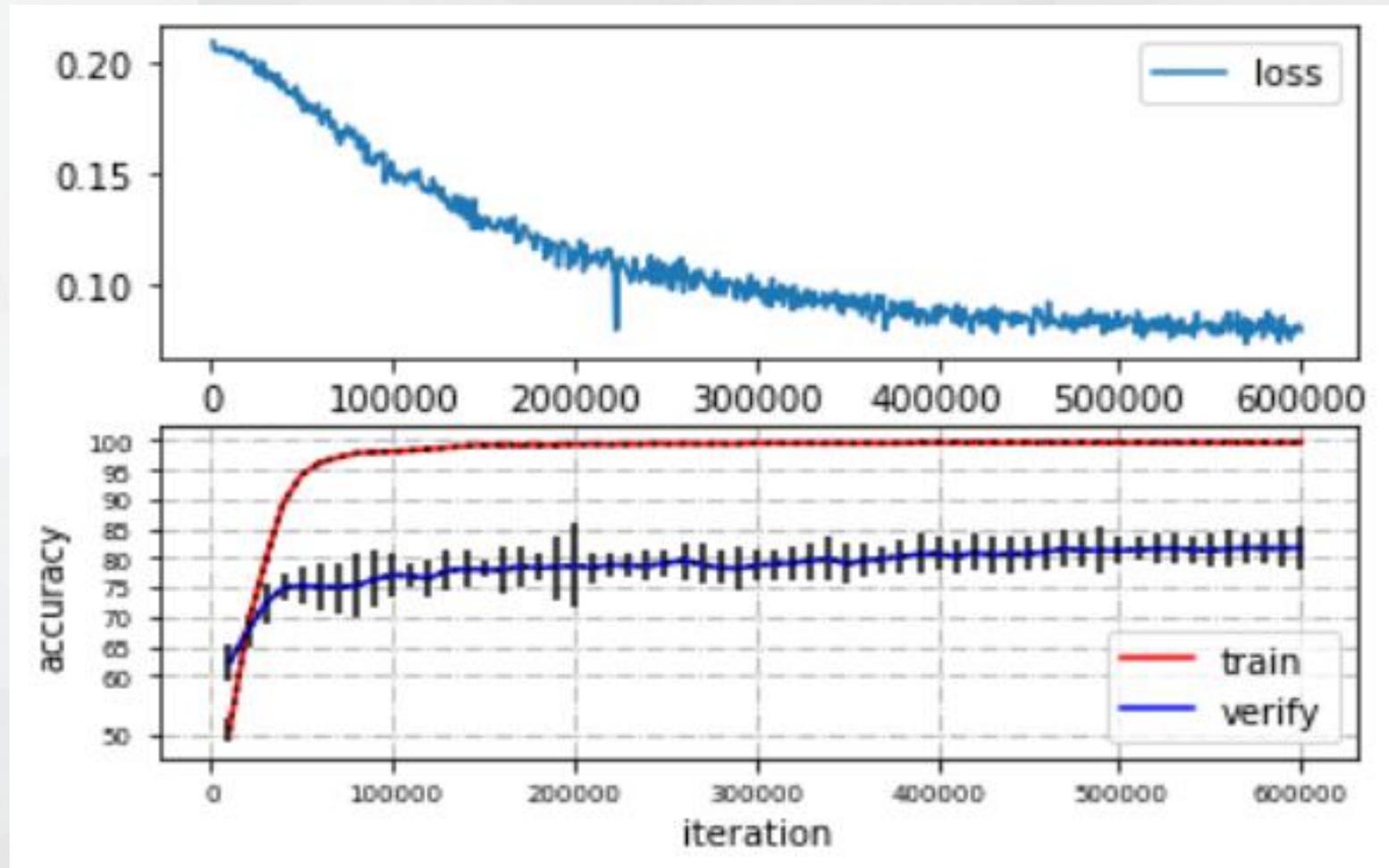


★ Experiment



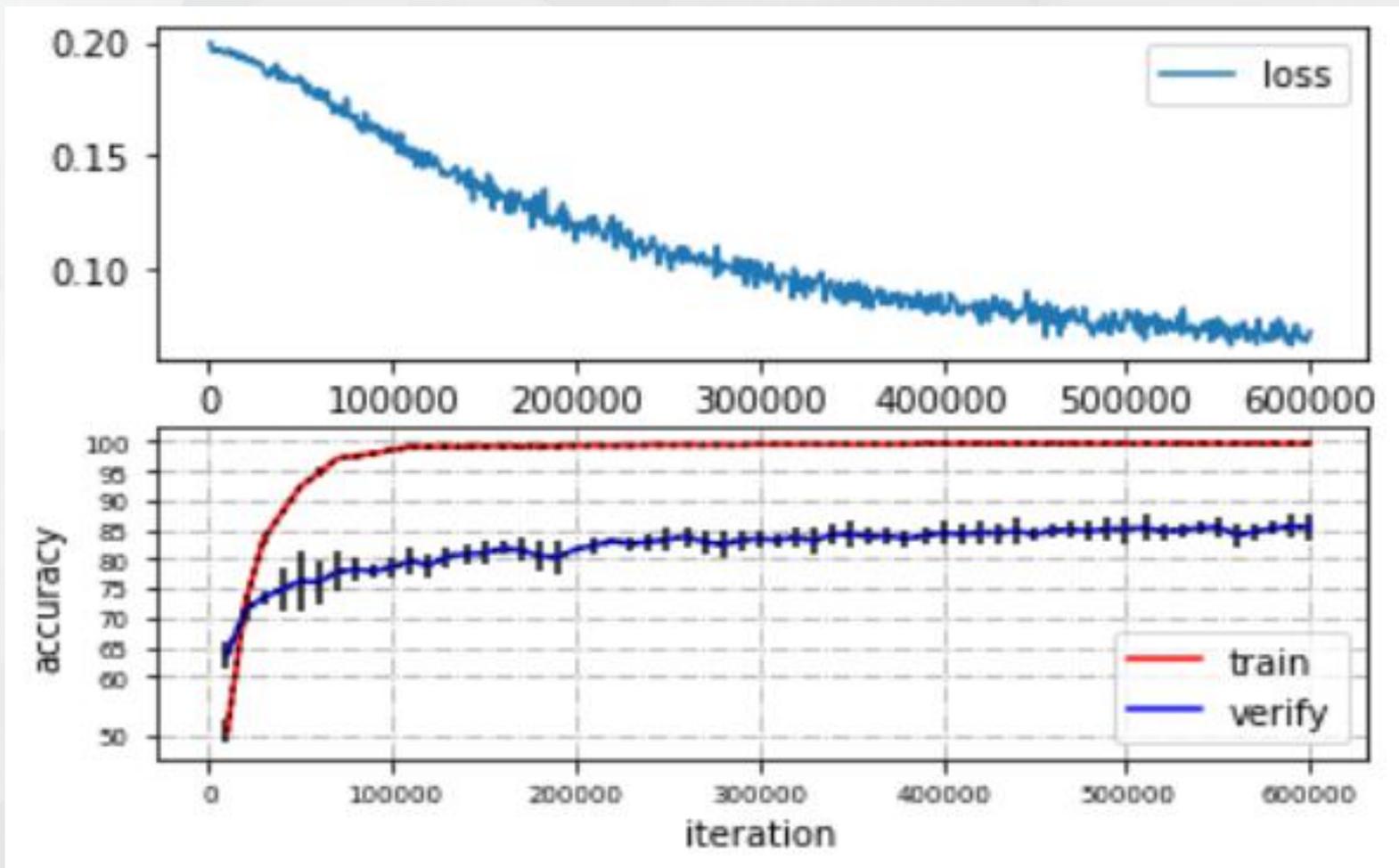
GaitSet. The result of the GaitSet model without modification.

★ Experiment



High-resolution data flow. Improving the GaitSet method with high-resolution data 128*128.

★ Experiment



Our method: ResNet backbone+ High-resolution data flow.

★ Experiment

Method	Iterations	Accuracy
GaitSet	450000 times	V:75.3% T:49.2%
GaitSet_high-resolution	450000 times	V:80.6% T:51.2%
Ours	450000 times	V:85.1% T:53.2%
GaitSet	600000 times	V:76.4% T:49.9%
GaitSet_high-resolution	600000 times	V:81.8% T:51.6%
Ours	600000 times	V:85.4% T:53.4%