Enhanced Spatial and Multi-Scale Temporal Feature Modeling for Human Identification at a Distance

Haijun Xiong[1], Xiaohu Huang[1], and Bin Feng[1]

[1]*Hauzhong University of Science and Technology*

Introduction

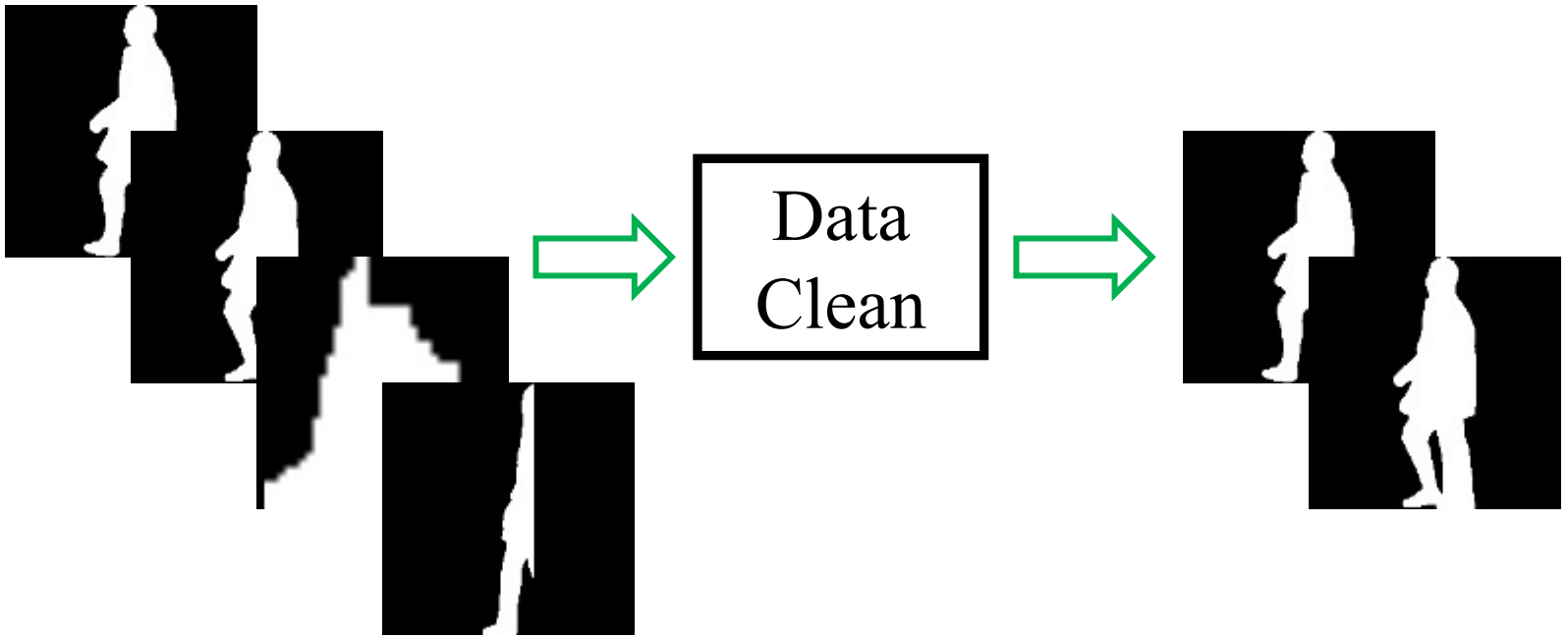
- Gait is a unique biometric feature that can be recognized at a distance, unlike face or iris, which does not depend on the cooperation of subjects. However, the accuracy of gait recognition is affected by external factors such as dressing, walking speed, carrying and camera viewpoint.

Preprocessing

- Data Cleaning
  - To avoid using low quality images, a simple image-filtering strategy by considering the ratio of the foreground was used to filter out low-quality images.

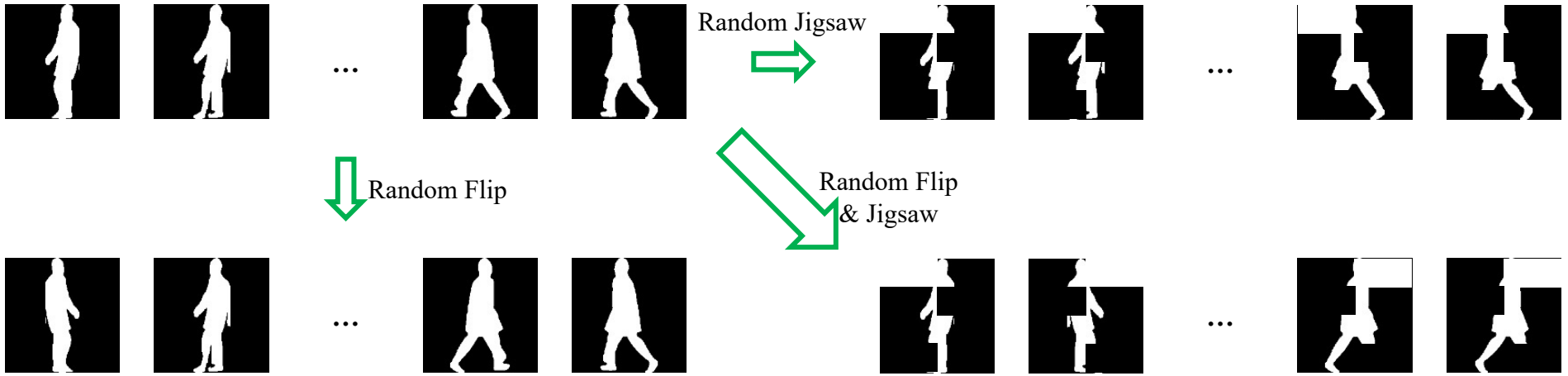
- Data Augmentation
  - We use the different data augmentation methods during training and testing. The train data augmentation methods include random horizontal flipping and random jigsaw, while the test data are only processed with random horizontal flipping.

Data Cleaning



The ratio of the foreground pixels of each image in each sequence was calculated, then to sort the images by the ratio of the foreground of each image in each sequence. The images with their ratios out of the range [0.85, 1.15] of the median will be removed.

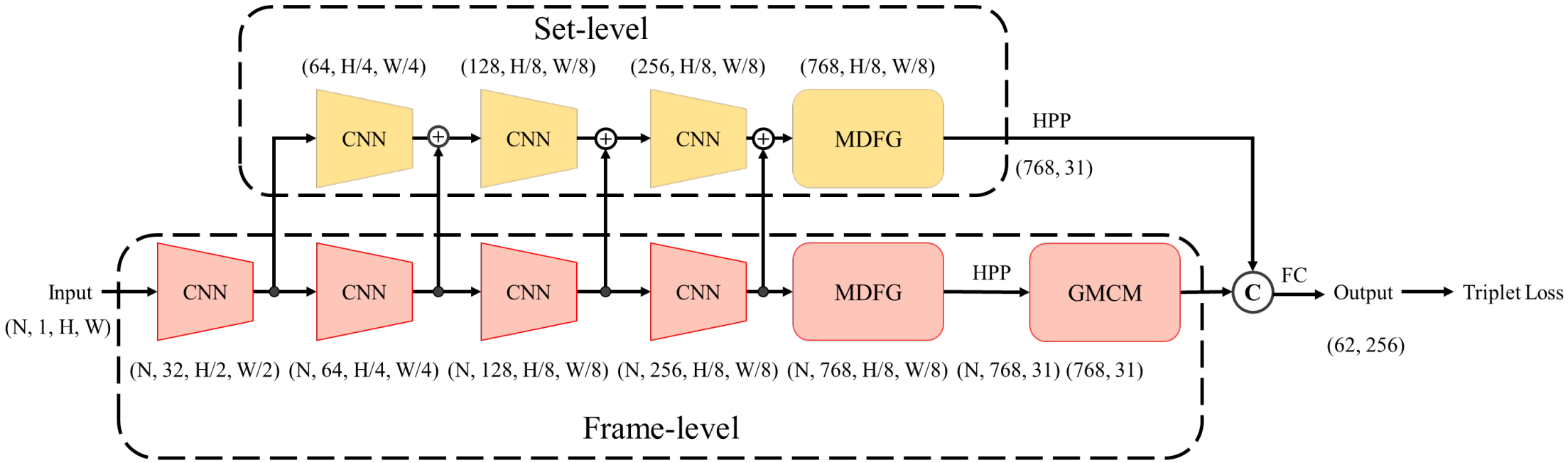
Data Augmentation



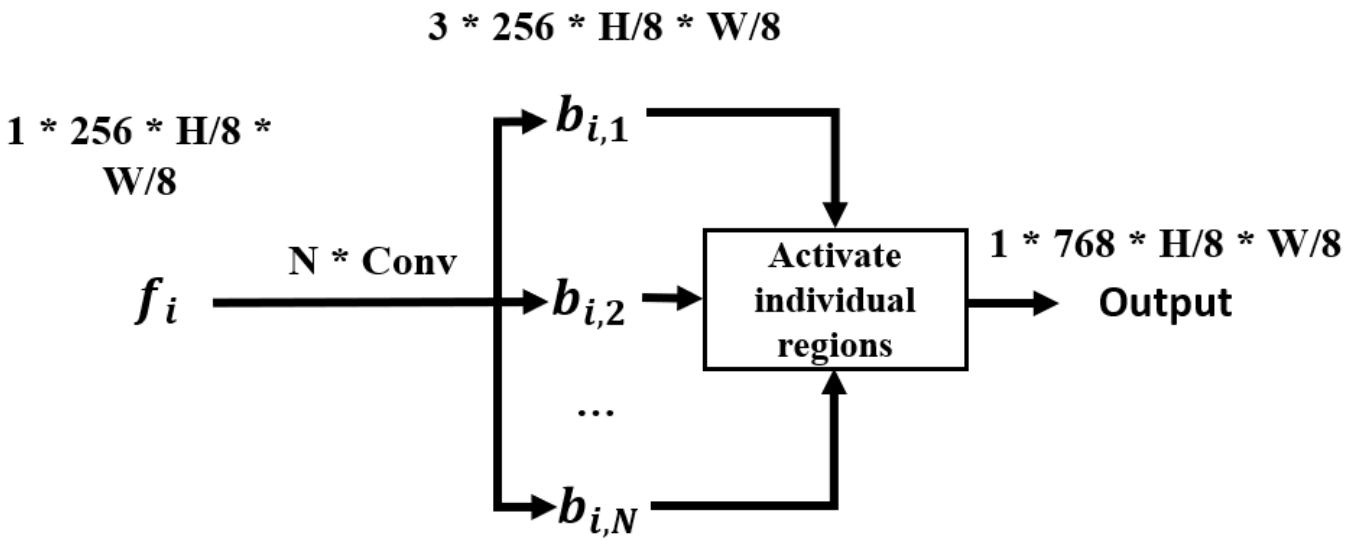
In the training stage, random horizontal flipping and jigsaw were used together.

In the testing stage, only random horizontal flipping was used.

Network


Set-level
(64, H/4, W/4)
(128, H/8, W/8)
(256, H/8, W/8)
(768, H/8, W/8)
CNN
CNN
CNN
MDFG
HPP
(768, 31)
Input
(N, 1, H, W)
CNN
CNN
CNN
CNN
MDFG
HPP
GMCM
C
FC
Output
Triplet Loss
(62, 256)
(N, 32, H/2, W/2)
(N, 64, H/4, W/4)
(N, 128, H/8, W/8)
(N, 256, H/8, W/8)
(N, 768, H/8, W/8)
(N, 768, 31)
(768, 31)
Frame-level

MDFG



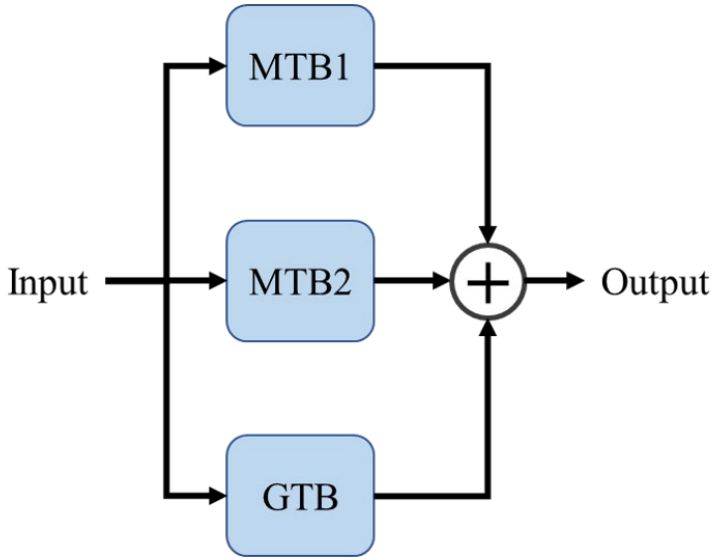
$$A_{i,j} = \frac{1}{1+\exp(-H_{i,j} - \sigma_{i,j})}$$

$$L_{oap}^{i} = \frac{1}{N}\sum_{x,y}\left(A_{i,1} \odot A_{i,2} \odot \cdots \odot A_{i,N}\right)$$

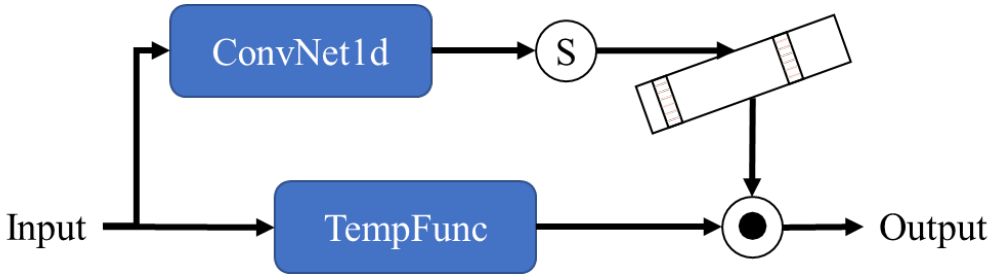
$$L_{oap} = \frac{1}{S}\sum_{i=1}^{S} L_{oap}^{i}$$

MDFG is employed in both Set-level and Frame-level to generate visual clues in diverse regions for fine-grained feature learning.

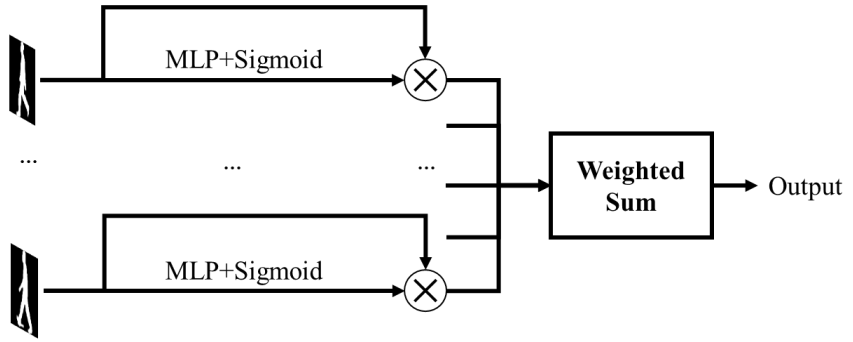
GMCM



GMCM



MTB



GTB

MTB aims to map the frame-level part-informed feature vectors into the micromotion feature vectors, and GTB is designed to map the frame-level global-informed feature vectors into the feature vectors.

Results On CASIA-B

Gallery		0°~180°											Ours	GaitSet
Probe		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°		
ST (24)	NM#5-6	78.94	88.94	93.89	92.17	86.61	81.06	85.76	91.01	92.93	89.09	73.48	86.716	79.5
	BG#1-2	72.27	81.46	87.63	85.41	78.48	72.78	77.02	83.96	86.44	80.36	66.75	79.324	68.6
	CL#1-2	41.26	49.60	58.28	56.55	52.27	48.44	50.66	53.91	50.66	46.72	35.91	49.477	40.9
MT (62)	NM#5-6	93.44	98.93	98.52	98.03	95.16	92.54	94.75	98.20	99.26	97.46	89.43	95.976	92.0
	BG#1-2	89.02	93.93	95.08	93.14	89.26	85.41	89.10	93.69	96.06	95.37	82.46	91.138	84.3
	CL#1-2	68.44	75.25	80.33	75.66	66.31	64.84	67.62	69.59	75.25	70.90	61.31	70.499	62.5
LT (74)	NM#5-6	96.90	99.90	99.70	98.80	97.30	95.50	97.70	99.50	99.70	99.20	94.30	98.045	95.0
	BG#1-2	93.50	95.40	96.10	95.25	91.70	87.60	90.10	94.60	97.90	97.58	89.00	93.521	87.2
	CL#1-2	73.10	81.40	84.80	78.90	73.20	71.90	71.60	72.90	77.50	76.60	69.10	75.545	70.4

Experiments

Rank	User	Accuracy
1	BeibeiLin	83.9%
2	dtdtdt	79.9%
3	Haijunxiong	71.3%
4	AlexFor	66.8%
5	YaoJun	66.6%
6	ZhangYuxuan	63.7%

华中科技大学
HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
明德厚学 求是创新
WUHAN CHINA

Thanks