Context-Sensitive Temporal Feature Learning for Gait Recognition (*Supplementary Material*)

Xiaohu Huang*[1], Duowang Zhu*[1], Hao Wang[1], Xinggang Wang[1], Bo Yang[2], Botao He[2], Wenyu Liu[1], and Bin Feng†[1]

[1]Huazhong University of Science and Technology
[2]Wuhan FiberHome Digital Technology Co., Ltd

1. Supplementary Material

This supplementary material includes:

- Investigation on the impact of different part division numbers.
- Comparison of different spatial selection strategies.
- Some qualitative results of salient spatial feature learning.

1.1. Ablation Study on Part Division Numbers

In order to investigate the effects of the part division number, *i.e.*, K, in this paper, we conduct 5 experiments with K of 1, 2, 4, 8, 16, and 32, where 32 is the largest number since the output feature dimension is 32×22. As shown in Fig.1, we notice that the accuracy improves continually with the increasing of number of parts, which indicates that more fine-grained division provides richer clues for spatio-temporal modeling, thus satisfies the diverse motion expression of different body parts. Therefore, we set $K = 32$ in this paper.

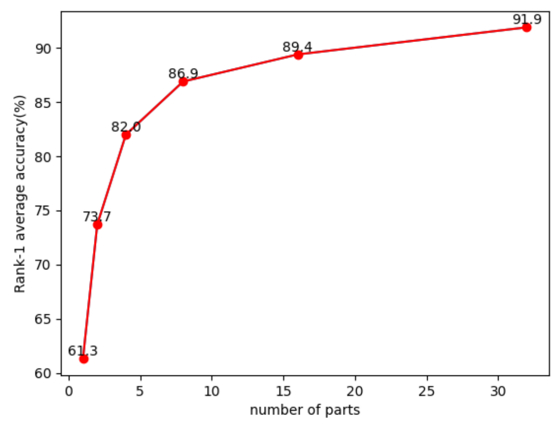


Figure 1. Study on the impact of different part division numbers on CASIA-B [1] in terms of averaged rank-1 accuracy under NM, BG and CL conditions.

1.2. Ablation Study on Different Spatial Selection Strategies

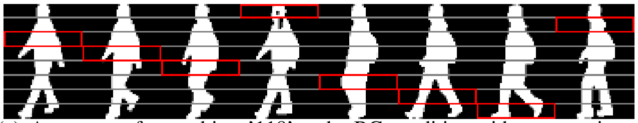
In order to investigate the effectiveness of our spatial learning module (SSFL) for supplementing corrupted spatial features, we conduct two more experiments for comparison: 1) we replace SSFL with a random frame selection to demonstrate that not each frame has good spatial features. 2) We set the number of selected parts as 1 in SSFL. In this situation, SSFL turns to be a frame-level feature selection instead of part-level feature selection.

As shown in Tab.1, we notice that: SSFL outperforms the other two strategies, which proves the spatial learning capability of our method. On one hand, random frame selection is probably incapable of obtaining high quality spatial features due to the randomness. On the other hand, although frame-level spatial selection achieves better performance than random frame selection, it still limits the diverse discriminative expression of local parts, especially considering the occlusion of motion and change of camera viewpoints. Compared to the above strategies, our SSFL extracts spatial clues in a fine-grained manner and utilizes the in-

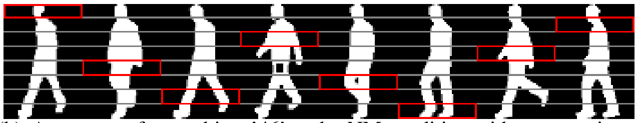
Table 1. Comparisons of spatial selection strategies on CASIA-B [1] in terms of averaged rank-1 accuracy.

Methods	Rank-1 Accuracy			
	NM	BG	CL	Mean
random frame	97.4	92.4	76.7	88.8
SSFL (frame-level)	97.7	92.7	80.1	90.2
SSFL	**97.8**	**93.6**	**84.2**	**91.9**

*Contributed equally.
†Corresponding author: fengbin@hust.edu.cn.



(a) A sequence from subject '118' under BG condition with camera viewpoint of 72 degrees.



(b) A sequence from subject '46' under NM condition with camera viewpoint of 54 degrees.

(c) A sequence from subject '46' under NM condition with camera viewpoint of 36 degrees.

(d) A sequence from subject '55' under BG condition with camera viewpoint of 126 degrees.

(e) A sequence from subject '72' under NM condition with camera viewpoint of 18 degrees.

(f) A sequence from subject '101' under NM condition with camera viewpoint of 126 degrees.

Figure 2. Illustration of spatial salient feature learning. The red boxes indicate selected parts.

herent motion characteristics to leverage rich visual clues across the sequence.

1.3. Qualitative Results

Here, we provide 6 more examples to further verify the effectiveness of SSFL module under complex situations in Fig. 2, where we set the number of selected parts as 8 in SSFL for better visualization. As we perceive, SSFL generally tends to select distinct parts without occlusion under clothing-changing and multi-view scenarios, which proves the robustness of our approach against complex variations.

References

[1] Shiqi Yu, Daoliang Tan, and Tieniu Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. *ICPR*, 4:441–444, 2006. 1