Supplementary Material for "Gait Recognition in the Wild with Dense 3D Representations and A Benchmark"

Jinkai Zheng[1*] Xinchen Liu[2†] Wu Liu[2†] Lingxiao He[2] Chenggang Yan[1] Tao Mei[2]
[1]Hangzhou Dianzi University, Hangzhou, China [2]Explore Academy of JD.com, Beijing, China
{zhengjinkai3, cgyan}@hdu.edu.cn, {liuxinchen1, liuwu1, helingxiao3, tmei}@jd.com

Abstract

In this supplementary material, we first introduce more details of the proposed SMPLGait framework. Then we provide additional experimental results for 1) analysis of the influence of the sequence length during inference, 2) the cross-domain evaluation of the State-Of-The-Art (SOTA) methods on the widely used datasets and our Gait3D dataset, and 3) exemplar results of the SMPLGait framework on the Gait3D dataset. At last, we discuss the potential negative impact of our work.

1. Details of the SMPLGait Framework

1.1. Details of Network Structure

The Silhouette Learning Network (SLN) consists of six convolutional layers, and each convolutional layer is followed by LeakyReLU whose negative slope is equal to 0.01. The structure of SLN is inspired by GaitSet [1] and OpenGait [1]. The detailed parameters are listed in Table 1.

The 3D Spatial Transformation Network (3D-STN) consists of several Fully Connected (FC) layers with dropout, Batch Normalization (BN) layers, and ReLU activation functions. The details of 3D-STN are listed in Table 2. Note that the h and w listed in the table are the height and width of the feature map from SLN. For convenient computation, we set $h = w = max(h, w)$ in our experiments.

Layers	Kernel #	Kernel Size	Stride	Padding
Conv1	64	5×5	1	2
LeakyReLU (0.01)	-	-	-	-
Conv2	64	3×3	1	1
LeakyReLU (0.01)	-	-	-	-
Max Pooling	-	2×2	2	0
Conv3	128	3×3	1	1
LeakyReLU (0.01)	-	-	-	-
Conv4	128	3×3	1	1
LeakyReLU (0.01)	-	-	-	-
Max Pooling	-	2×2	2	0
Conv5	256	3×3	1	1
LeakyReLU (0.01)	-	-	-	-
Conv6	256	3×3	1	1
LeakyReLU (0.01)	-	-	-	-

Table 1. Details of the SLN Network.

Layers	Neuron #	Dropout Rate
FC1	128	0.0
BN1	-	-
ReLU	-	-
FC2	256	0.2
BN2	-	-
ReLU	-	-
FC3	$h \times w$	0.2
BN3	-	-
ReLU	-	-

Table 2. Details of the 3D-STN Network. (Input Size: 88×128).

2. Additional Experimental Results

In this section, we first analyze the effect of sequence length on accuracy during inference. Then, we conduct the cross-domain experiments to reflect the domain gap between existing datasets and our Gait3D dataset. At last, we provide some exemplar results of our SMPLGait framework for gait recognition to qualitatively demonstrate the effectiveness of our method.

2.1. Effect of the Lengths of Test Sequences

This subsection analyzes the influence of the length of testing sequences on the accuracy of gait recognition. We sample 10% ∼ 100% frames with the 10% increment from the sequences during testing. The plots are demonstrated in Figure 1. From the results, we can observe that the accuracy is improved with the increasing frames of the sequence.

*This work was done when Jinkai Zheng was an intern at Explore Academy of JD.com.
†Corresponding author.
[1]https://github.com/ShiqiYu/OpenGait

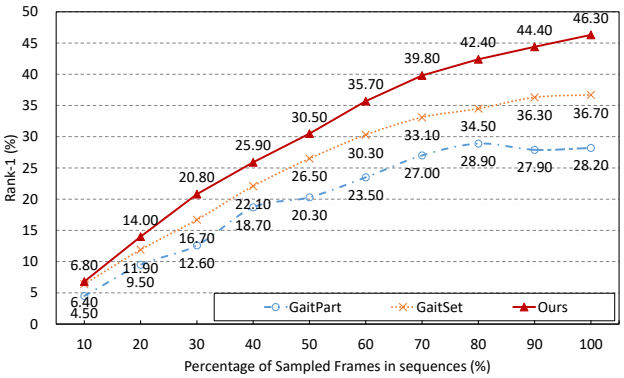

Figure 1. The effect of the lengths of test sequences.

Therefore, in real practice, we need to make a trade-off between accuracy and efficiency.

2.2. Cross-domain Experiments

In this subsection, we analyze the domain gap between our Gait3D dataset and existing widely used datasets including CASIA-B [3], OU-LP [2] and the recently released GREW [4]. Because existing datasets do not provide 3D representations, we only adopt the 2D silhouettes for gait representations in our experiments. We adopt the GaitSet [1] for the cross-domain experiments since it is the SOTA model-free method on the Gait3D dataset.

2.2.1 Evaluation Protocol

In the cross-domain experiments, we train the GaitSet model on the training set of one source dataset and evaluation the trained model on the testing set of one target dataset. For CASIA-B, OU-LP, and Gait3D, we use the official train/test split for training and testing. Because the test set of GREW [4] is not released publicly for evaluation, we randomly sample 1,000 IDs from the training set of GREW for evaluation. For the sampled 1,000 IDs, we further randomly select one sequence from each ID to build the query set with 1,000 sequences, while the rest of the sequences become the gallery set with 4,095 sequences. We utilize Rank-1 (R-1), Rank-5 (R-5), and mAP as the evaluation metrics.

2.2.2 Main Results

The results of the cross-domain evaluation are listed in Table 3. From the results, we first find that the GaitSet models trained on the in-the-lab datasets, i.e., CASIA-B [3] and OU-LP [2] obtain very poor results, only 6.90% and 6.10% Rank-1. This reflects that there is a huge domain gap between the in-the-lab research and the in-the-wild application.

Next, we observe that the model trained on GREW then tested on Gait3D only obtains 16.50% Rank-1, while the model trained on Gait3D then tested on GREW achieves much higher Rank-1, i.e., 43.86%. It is worth noting that the training set of GREW has 20,000 IDs, while our Gait3D only contains 3,000 IDs. This demonstrates the significant domain gap between our Gait3D and GREW although the two datasets are both collected in the wild. Moreover, it indicates that the model trained on Gait3D has a more powerful capability of generalization than the one trained on GREW.

Source	Target	R-1 (%)	R-5 (%)	mAP (%)
CASIA-B [3]	Gait3D	6.90	14.60	4.64
OU-LP [2]		6.10	12.40	4.42
GREW [4]		16.50	31.10	11.71
Gait3D	CASIA-B [3]	66.71	71.59	33.88
	OU-LP [2]	97.84	99.38	68.06
	GREW [4]	43.86	60.89	28.06

Table 3. Results of cross-domain experiments. The method is trained on each source dataset and directly tested on the target datasets.

At last, we can see that the GaitSet trained on Gait3D achieves competitive accuracy on in-the-lab datasets, i.e., Rank-1 of 66.71% on CASIA-B and 97.84% on OU-LP which is close to the model trained on OU-LP (99.89%) in our implementation). These results further prove that there is a huge gap between the in-the-lab dataset and in-the-wild application, while our Gait3D enables the model to learn more generalized gait representations.

2.3. Exemplar Results of SMPLGait

Figure 2 - 5 provides several exemplar results of our SMPLGait framework on the Gait3D dataset. The top two rows with **blue** bounding boxes are the silhouette sequence and 3D Mesh sequence of the query, respectively. The rows following the query are the top-5 gallery sequences ranked by their similarities to the query sequence. The results with **green** bounding boxes are the correctly matched sequences, while those with **red** bounding boxes are wrong results.

From Figure 2 - 4, we can observe that the 3D representations can work well in multi-viewpoint, occluded, and multi-person cases. By this means, 3D meshes can provide more information about shapes, poses, and viewpoints of human bodies, which can help improve the accuracy of gait recognition in real-world scenes.

Figure 5 illustrates a bad case in which the top-3 persons with similar clothes and shapes seriously interfere with the matching results. This indicates that very similar clothing and shapes of persons are one of the main challenges of gait recognition.

3. Discussion

Limitations. Although we proposed a 3D gait recognition framework, its performance still has a large space to

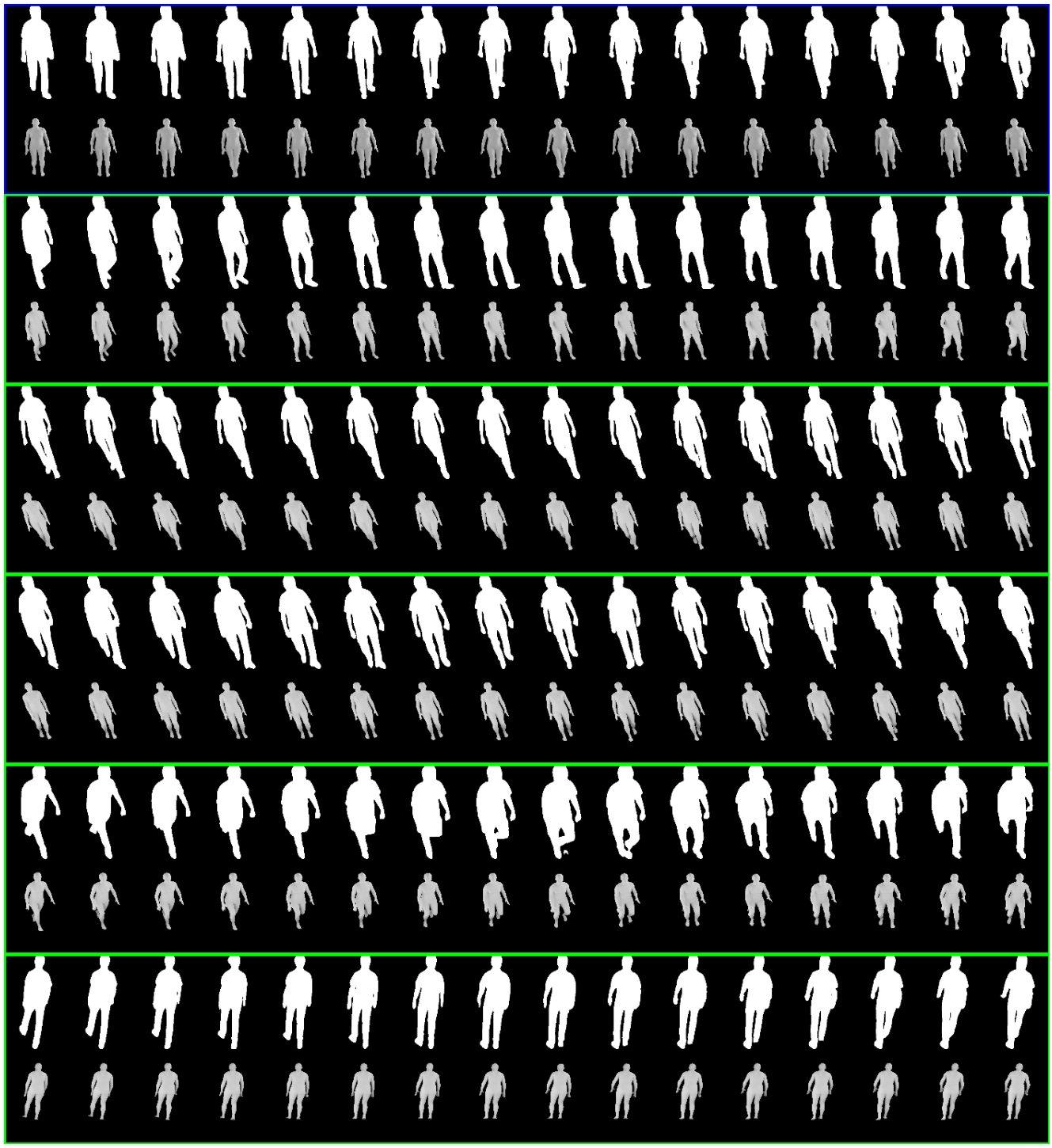

Figure 2. Exemplar results of SMPLGait on the Gait3D. 16 consecutive frames are sampled from each sequence for visualization. This case shows that our method obtains good results when the samples are high-quality. (Best viewed in color.)

improve for practical applications. In addition, we only exploit a few frames of gait sequences, e.g., 30 frames, in our framework. The temporal dynamics are not fully explored for gait recognition in the wild.

Potential Negative Impact. The potential negative outcomes mainly come from the fact that, with the large-scale deployment of the urban monitoring network, if the accuracy of gait recognition in the real scenarios is greatly

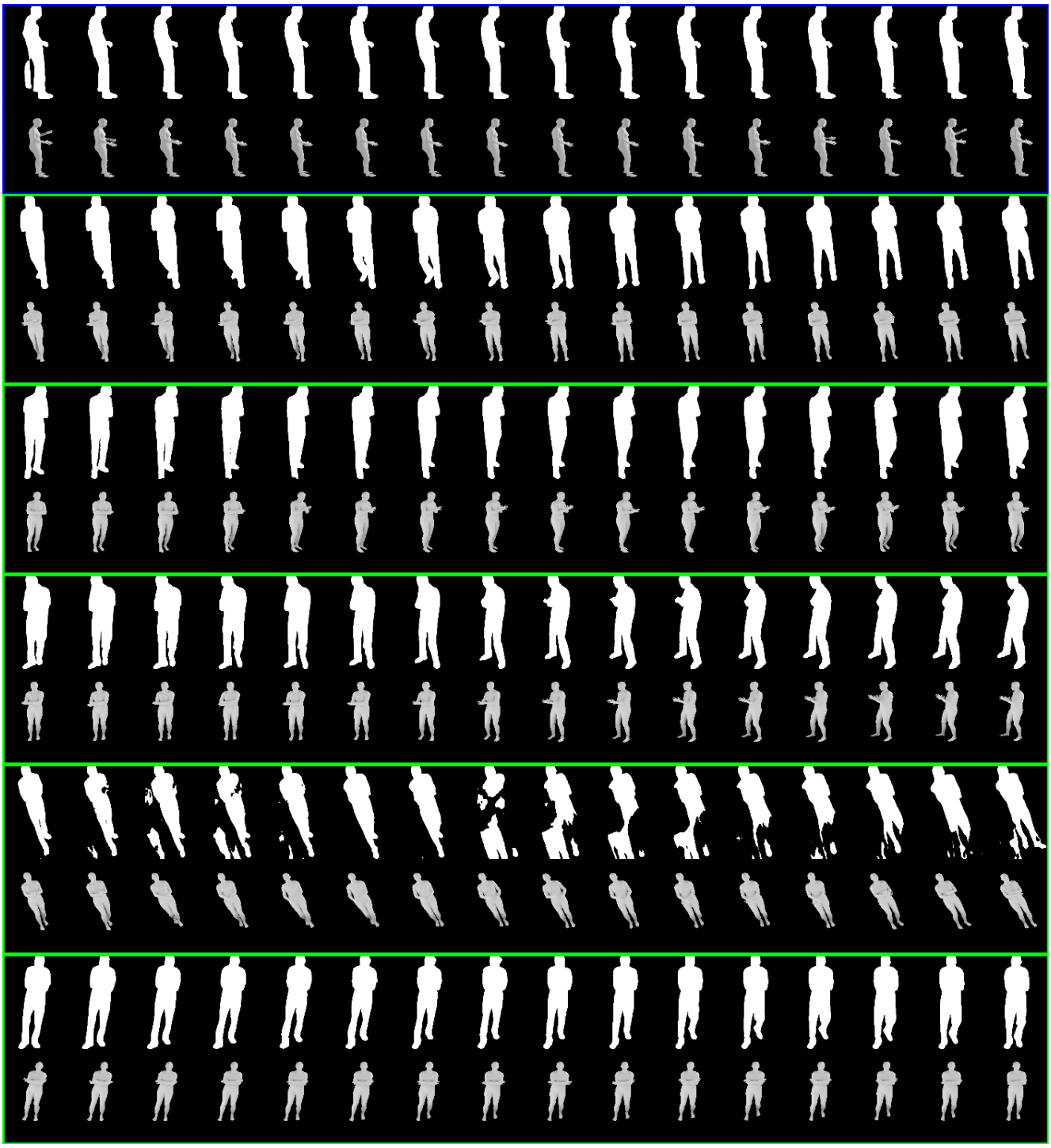

Figure 3. Exemplar results of SMPLGait on the Gait3D. This example reflects that our method can work well when part of the person is occluded. (Best viewed in color.)

improved in the future, it may cause some privacy and security issues. To minimize these risks, our Gait3D dataset will be distributed only for research purposes via the case-by-case application with a strict license. In summary, we will try our best to protect the privacy of the subjects. We hope that the development of gait recognition can help to create a better human society, as well as better services for human beings, such as assisting the police to solve crimes, looking for lost people, and so on.

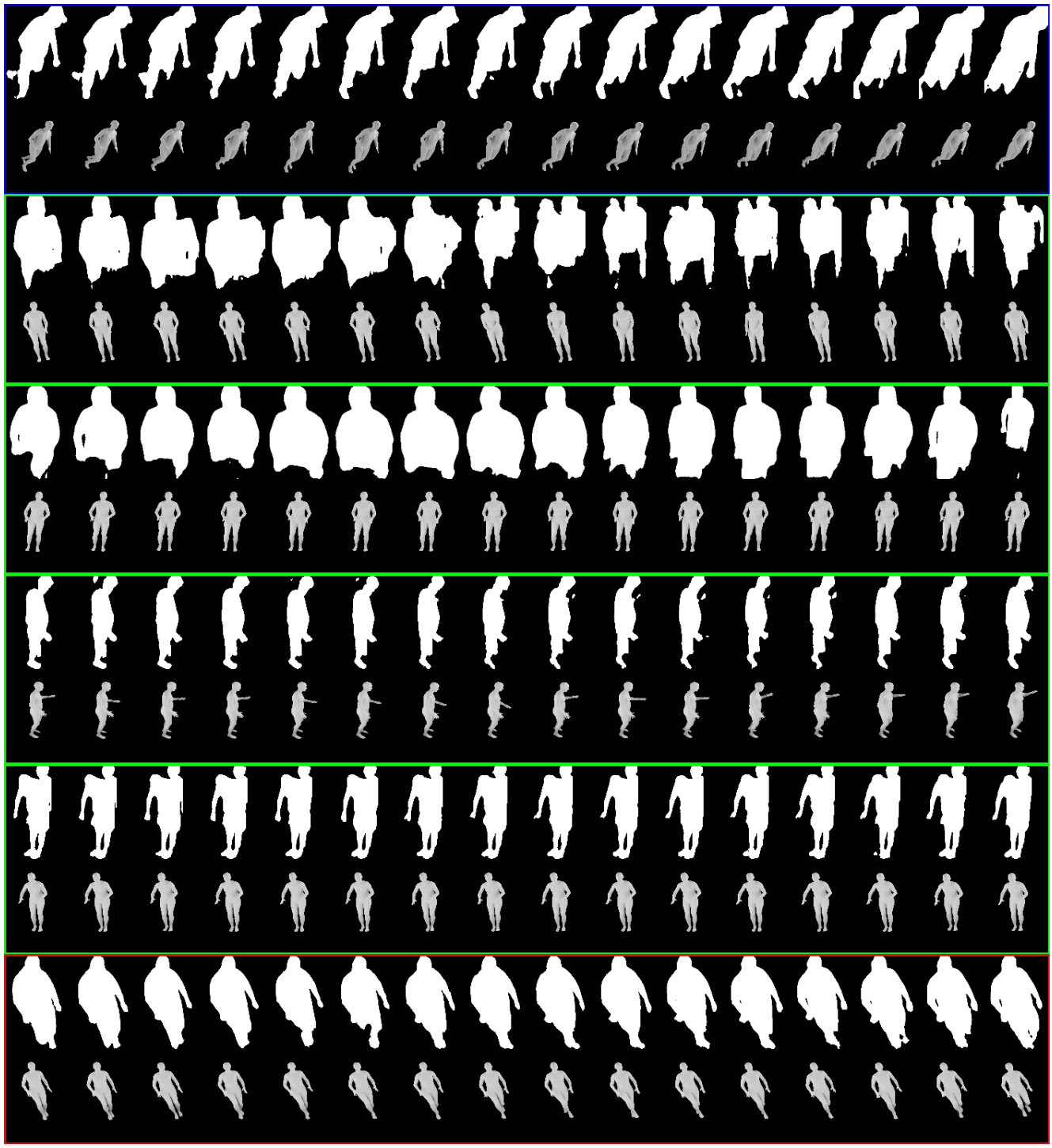

Figure 4. Exemplar results of SMPLGait on the Gait3D. This example shows that even when the silhouettes are of low quality, the 3D meshes can help the model obtain correct results. (Best viewed in color.)

References

[1] Hanqing Chao, Yiwei He, Junping Zhang, and Jianfeng Feng. GaitSet: Regarding gait as a set for cross-view gait recognition. In *AAAI*, pages 8126–8133, 2019. 1, 2

[2] Haruyuki Iwama, Mayu Okumura, Yasushi Makihara, and Yasushi Yagi. The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE TIFS*, 7(5):1511–1521, 2012. 2

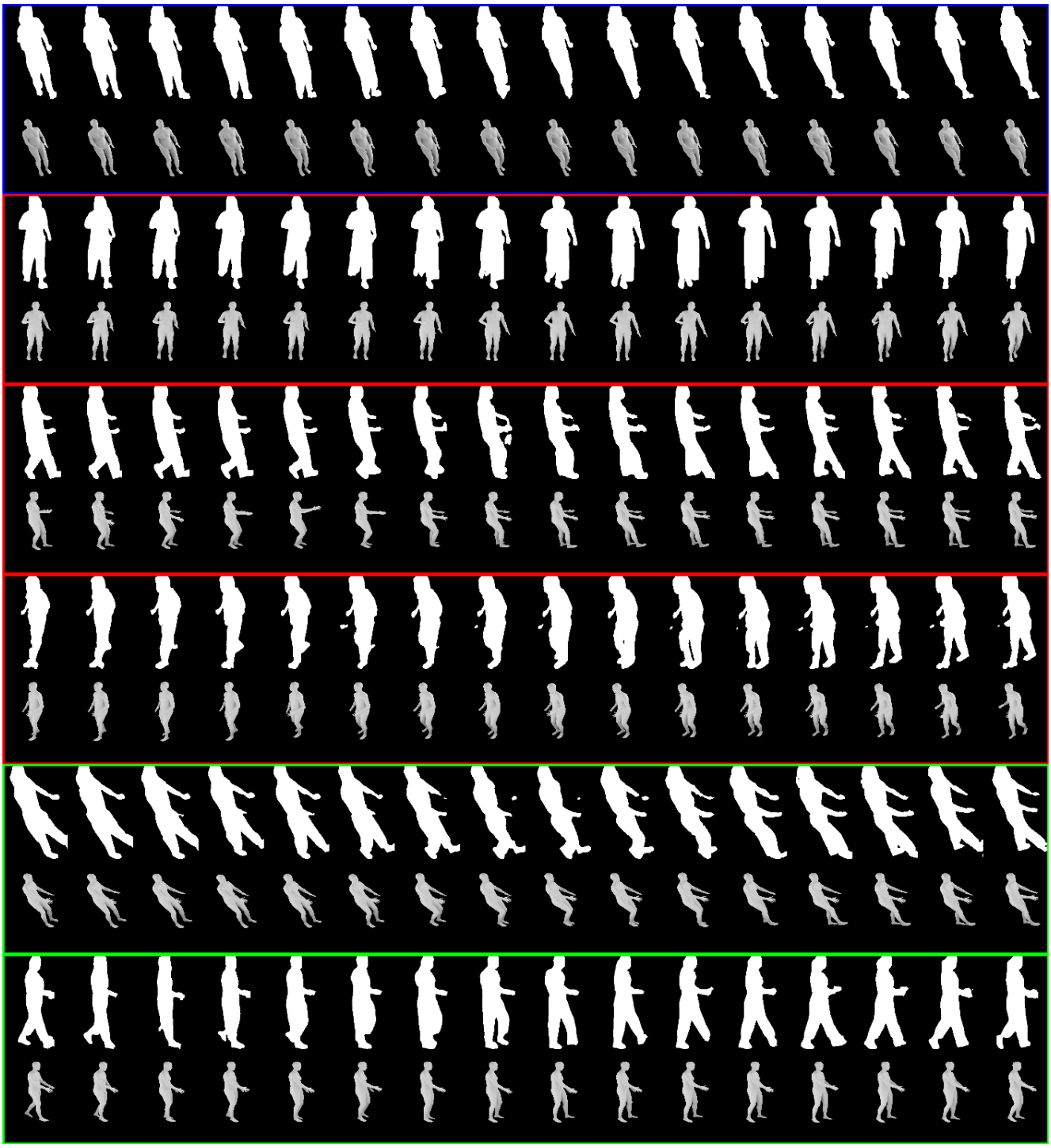

Figure 5. A bad case of SMPLGait on the Gait3D. The top-3 results contain persons wearing very similar clothes to the query, which is a very challenging condition of gait recognition. (Best viewed in color.)

[3] Shiqi Yu, Daoliang Tan, and Tieniu Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *ICPR*, pages 441–444, 2006. 2

[4] Zheng Zhu, Xianda Guo, Tian Yang, Junjie Huang, Jiankang Deng, Guan Huang, Dalong Du, Jiwen Lu, and Jie Zhou. Gait recognition in the wild: A benchmark. In *ICCV*, pages 14789–14799, 2021. 2