IEEE International Joint Conference on Biometrics

GaitRef: Gait Recognition with Refined Sequential Skeletons

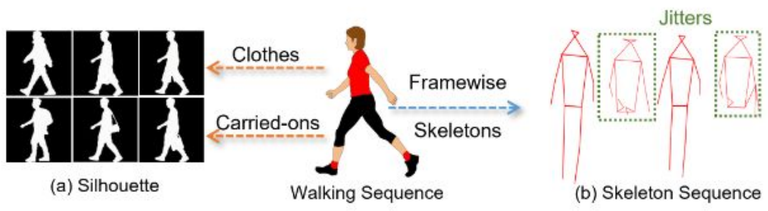
Haidong Zhu*, Wanrong Zheng*, Zhaoheng Zheng and Ram Nevatia

University of Southern California

USC Viterbi School of Engineering, Department of Computer Science

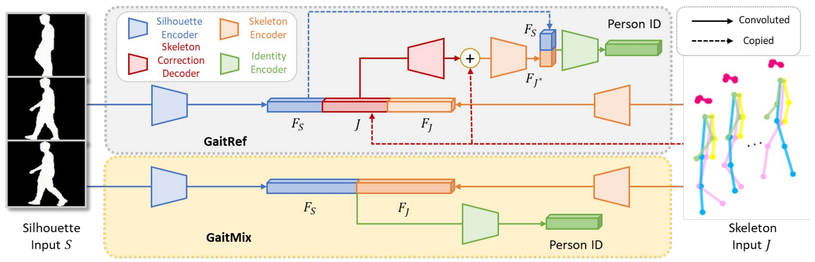
Paper | Code

INTRODUCTION



- Gait recognition identifies individuals based on their walking patterns that can be observed at distance.
- Skeleton and silhouettes are the two most widely used modalities. Appearance variances in 2-D silhouette images, such as carried-on objects, and clothing, make the task challenging, while skeletons are usually noisy with jitters.
- We combine silhouette and skeletons, and use the temporal consistency of silhouette to refine skeletons for gait recognition. We show state-of-the art methods on multiple datasets.

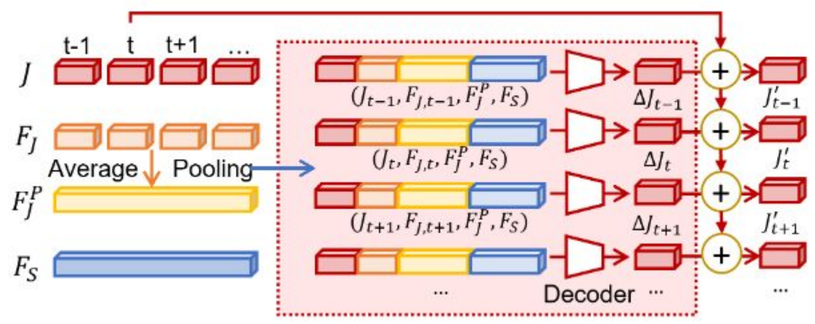
METHODOLOGY



- We propose two methods for combining silhouette and skeletons in a video sequenc.
  - **GaitMix**
    - Aggregates features of silhouette and skeleton together for identification.
    - End-to-end training, works as a baseline.
  - **GaitRef**
    - Refines skeleton sequence with temporal information encoded from silhouettes.
    - Aggregates features of refined skeletons with silhouette features for identification.
    - End-to-end training for correction network and other feature encoders.
- For encoders and decoders we used:
  - Silhoeutte encoder - we use different encoders based on our datasets. We use SMPLGait for Gait3D and GaitGL for other datasets.
  - Skeleton encoder - ST-GCN.
  - Skeleton correction decoder - ST-GCN.
  - Identity encoder - FC layers
- Training
  - Triplet and classification loss

$$L = \lambda_1 L_{triplet} + \lambda_2 L_{cls}$$

NETWORK DETAILS



- For skeleton correction network
  - Reversed ST-GCN
  - Model input - four different representations
    - Input joints J
    - Feature extracted per frame F_J
    - Average of skeleton features F_J^P
    - Silhouette features F_S
  - Model output - joint correction
    - Joint modification per frame ΔJ_t
    - Added on skeleton of each frame
  - The skeleton encoder after the correction network is shared with original encoder
    - Ensure the corrected skeletons are in the same domain as input skeletons.

DATASETS

- CASIA-B:
  - 124 subjects with 110 videos for each subject.
- OUMVLP
  - 10,307 subjects with normal walking .
- Gait3D
  - 4,000 identities with 25,309 sequences.
  - Dataset for in-the-wild case.
- GREW
  - 26,345 identities with 128,671 sequences.
  - Dataset for in-the-wild case.

RESULTS

- CASIA-B (Rank-1 accuracy):

Method	NM	BG	CL
GaitGL	97.3	94.4	83.5
CSTL	97.8	93.6	84.2
ModelGait	97.9	93.1	77.6
GaitMix	97.7	95.2	85.8
GaitRef	**98.1**	**95.9**	**88.0**

- OUMVLP (Rank-1 accuracy, 64 x 44):

Method	GLN	GaitGL	MvModelG.
Accuracy	89.2	89.6	89.7
Method	CSTL	GaitMix	GaitRef
Accuracy	**90.2**	89.9	**90.2**

RESULTS

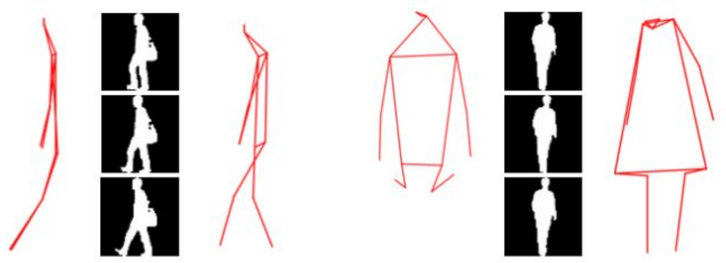
- Gait3D (64 x 44)

Method	Rank 1	Rank 5	mAP	mINP
GaitGL	29.70	48.50	22.29	13.26
OpenGait	42.90	63.90	35.19	20.83
CSTL	11.70	19.20	5.59	2.59
SMPLGait	46.30	64.50	37.16	22.23
GaitMix	45.80	65.60	36.74	22.09
GaitRef	**49.00**	**69.30**	**40.69**	**25.26**

- GREW

Method	Rank 1	Rank 5	Rank 10	Rank 20
GaitSet	46.3	63.6	70.3	76.8
GaitPart	44.0	60.7	67.4	73.5
CSTL	50.6	65.9	71.9	76.9
GaitGL	51.4	67.5	72.8	77.3
GaitMix	52.4	67.4	72.9	77.2
GaitRef	**53.0**	**67.9**	**73.0**	**77.5**

CORRECTED SKELETONS



- A successful and a failure example
- From left to right, we have original skeletons, silhouette of the nearby timestamp and corrected skeletons from skeleton correction network.
- For both cases, the ID prediction is correct after the skeleton correction, while the original prediction is wrong with its baseline.

REFERENCE

[1] Chao, Hanqing, Yiwei He, Junping Zhang, and Jianfeng Feng. "Gaitset: Regarding gait as a set for cross-view gait recognition." *AAAI. 2019.*

[2] Fan, Chao, Yunjie Peng, Chunshui Cao, Xu Liu, Saihui Hou, Jiannan Chi, Yongzhen Huang, Qing Li, and Zhiqiang He. "Gaitpart: Temporal part-based model for gait recognition." *CVPR 2020.*

[3] Hou, Saihui, Chunshui Cao, Xu Liu, and Yongzhen Huang. "Gait lateral network: Learning discriminative and compact representations for gait recognition." *ECCV 2020.*

[4] Lin, Beibei, Shunli Zhang, and Xin Yu. "Gait recognition via effective global-local feature representation and local temporal aggregation." *ICCV 2021.*

[5] Huang, Xiaohu, Duowang Zhu, Hao Wang, Xinggang Wang, Bo Yang, Botao He, Wenyu Liu, and Bin Feng. "Context-sensitive temporal feature learning for gait recognition." *ICCV* 2021.

[6] Li, Xiang, Yasushi Makihara, Chi Xu, Yasushi Yagi, Shiqi Yu, and Mingwu Ren. "End-to-end model-based gait recognition." *ACCV* 2020.