Occluded Gait Recognition with Mixture of Experts: An Action Detection Perspective

Panjian Huang [1,3*] Yunjie Peng [2,4*] Saihui Hou [1,3†] Chunshui Cao [3] Xu Liu [3] Zhiqiang He [2,4] Yongzhen Huang [1,3†]

[1]Beijing Normal University [2]Beihang University [3]Watrix.AI [3]Lenovo Research

Introduction

Gait recognition aims to recognize human from a distance through the unique walking patterns under occlusion, cross-view, and cross-clothing scenerios. This work introduces an innovative perspective regarding a gait sequence as a composition of actions and employs a Mixture of Experts to extract accurate actions for addressing occluded gait recognition.

1. **Challenges.** Extensive occlusions in real-world scenarios pose challenges to gait recognition due to missing and noisy information, as well as body misalignment in position and scale.

Figure 1. Occlusion Issues.

2. **Perspective.** Regarding a gait sequence as a composition of actions, action detection based Mixture of Experts allows information integration between holistic and occluded actions.

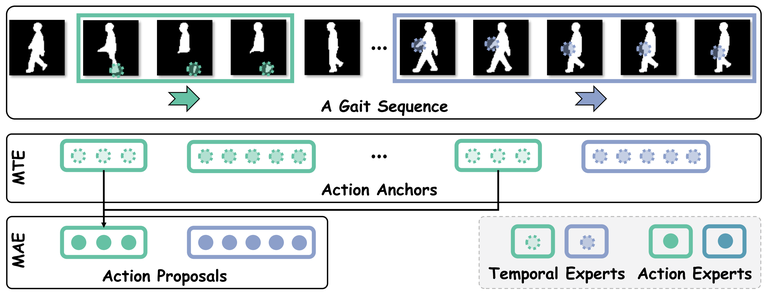

Figure 2. Action Detection Based Mixture of Experts. Each temporal expert focuses on one body region with individual temporal size, constructing action anchors. Each action expert integrates similar action anchors from different gait cycles, constructing one action proposal.

Contributions

1. To address occlusion challenges, we introduce an action detection perspective where an Action Detection Based Mixture of Experts (GaitMoE) structures a gait sequence as a composition of action.
2. To qualify and quantify occlusion issues, we build a novel Occluded Gait recognition benchmark (OccGait), including diverse occlusion scenarios and explicit annotations of occlusion types.
3. To evaluate effectiveness and robustness, extensive experimental results on OccGait, OccCASIA-B, Gait3D, and GREW demonstrate that our method significantly outperforms other state-of-the-art methods.

Methodology

Figure 3. The overview of GaitMoE. The input gait sequence is constructed into action anchors by MTE, then action proposals are generated by MAE for identification.

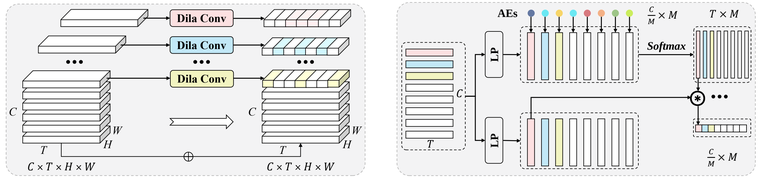

Figure 4. (Left) Mixture of Temporal Experts. Dila Conv (DC) represents Dilated Convolution, predefining action anchors with different dilated ratios. (Right) Mixture of Action Experts. LP denotes the linear projection. Similar action anchors adaptively integrate into action proposals.

1. **Action Anchor Stage.** MTE predefines various sizes of temporal experts to construct action anchors, which are dilated convolutions with different dilated ratios for corresponding channel segments.
2. **Action Proposal Stage.** MAE adaptively constructs action proposals from action anchors by action experts, which are learnable prototypes.
3. **Identification.** Action detection as a proxy task with gait recognition is an endto-end joint training only with ID labels.

OccGait

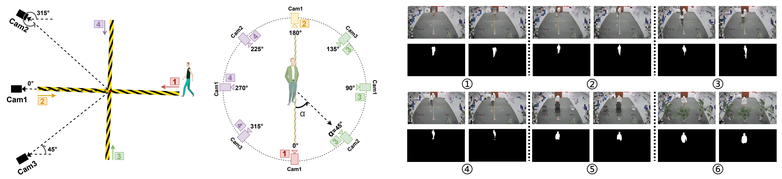

Figure 5. (Left) The layout of data collection. (Right) The 3 types of occlusion scenarios: Carrying Occlusion, Crowd Occlusion and Static Occlusion

1. **Diverse Occlusion Scenarios.** Each subject has 4 different types of occlusion situations.
2. **Explicit Occlusion Types.** OccGait provides explicit occlusion types for each gait sequence.

EXPERIMENT

Methods	OccGait				OccCASIA-B			
	NM	CA	CR	ST	NO	CO	SO	DO
GaitSet	84.7	69.3	74.2	74.2	80.2	66.9	73.5	70.4
GaitPart	82.6	62.4	66.3	71.9	82.0	68.1	74.3	67.2
GaitGL	87.5	70.6	71.9	74.0	86.4	72.3	78.7	74.6
STOR	88.3	76.3	77.1	80.6	88.0	77.9	82.3	83.2
GaitBase	86.0	75.4	78.5	80.3	83.9	74.4	78.3	76.3
GaitMoE-T(Ours)	**91.4**	**82.1**	**79.9**	**84.7**	**89.5**	**80.5**	**83.9**	**85.3**

Table 1. The Rank-1 accuracy (%) on OccGait and OccCASIA-B.

Method	Venue	Gait3D		GREW	
		Rank-1	mAP	Rank-1	Rank-5
GaitSet	AAAI19	36.7	30.0	46.3	63.6
GaitPart	CVPR20	28.2	47.6	44.0	60.7
GaitGL	ICCV21	29.7	22.3	47.3	63.6
SMPLGait	CVPR22	46.3	37.2	-	-
MTSGait	MM22	48.7	37.6	55.3	71.3
GaitBase	CVPR23	64.6	-	60.1	-
DANet	CVPR23	48.0	-	-	-
GaitGCI	CVPR23	50.3	39.5	68.5	80.8
DyGait	ICCV23	66.3	56.4	71.4	83.2
HSTL	ICCV23	61.3	55.5	62.7	76.6
GaitMoE-T(Ours)	ECCV24	**71.3**	**62.5**	**74.4**	**84.9**
GaitMoE-B(Ours)	ECCV24	**73.7**	**66.2**	**79.6**	**89.1**

Table 2. The Rank-1 accuracy (%) and mAP (%) on Gait3D and GREW.

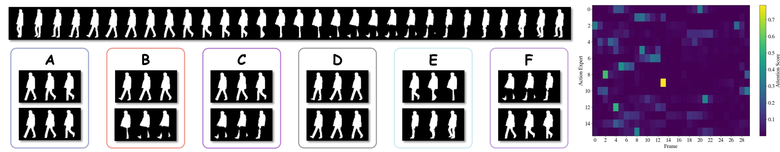

Figure 6. The visualization of action composition.

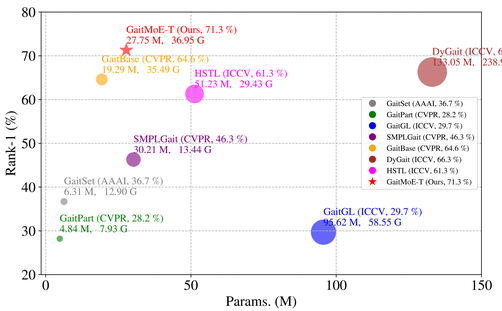

Figure 7. The accuracy and efficiency. Rank-1 (%), Param. (M) and FLOPs. (G) on Gait3D.