MetaGait: Learning to Learn an Omni Sample Adaptive Representation for Gait Recognition

Huanzhang Dou[1], Pengyi Zhang[1], Wei Su[1],
Yunlong Yu[2⋆][0000−0002−0294−2099], and Xi Li[1,3,4⋆][0000−0003−3023−1662]

[1] College of Computer Science & Technology, Zhejiang University
[2] College of Information Science & Electronic Engineering, Zhejiang University
[3] Shanghai Institute for Advanced Study, Zhejiang University
[4] Shanghai AI Laboratory

A Network Architecture

Proposed MetaGait framework is mainly composed of a feature extractor, a Meta Triple Attention module, and a Meta Temporal Pooling module. As shown in Tab. 1, the architecture of MetaGait is presented, including the channels, kernel size, stride, padding, part, and the shape of output. The feature extractor is following [5], which removes LTA [5] and the temporal aggregation process.

Table 1: The architecture of MetaGait. In_C, Out_C, Kernel, and part represent the input channel, output channel, kernel size, and the part number, respectively.

Layer	In_C	Out_C	Kernel	Stride	Pad	Part	Output
Input	1	–	–	–	–	–	$B \times 1 \times T \times H \times W$
Conv0	1	32	(3,3,3)	(1,1,1)	(1,1,1)	–	$B \times 32 \times T \times H \times W$
Avg Pool0	–	–	(3,1,1)	(3,1,1)	–	–	$B \times 32 \times T/3 \times H \times W$
GLConv1	32	64	(3,3,3)	(1,1,1)	(1,1,1)	8	$B \times 64 \times T/3 \times H \times W$
Avg Pool1	–	–	(1,2,2)	(1,2,2)	–	–	$B \times 64 \times T/3 \times H/2 \times W/2$
GLConv2	64	128	(3,3,3)	(1,1,1)	(1,1,1)	8	$B \times 128 \times T/3 \times H/2 \times W/2$
GLConv3	128	128	(3,3,3)	(1,1,1)	(1,1,1)	8	$B \times 128 \times T/3 \times H \times W/2$
MTA	128	128	–	–	–	–	$B \times 128 \times T/3 \times H \times W/2$
MTP	128	128	–	–	–	–	$B \times 128 \times H \times W/2$

B Detailed Performance under Data-limited Scenario

As shown in Fig.5 of the main manuscript, MetaGait outperforms state-of-the-art methods by a considerable margin. To further validate the cross-view and cross-condition ability of MetaGait under the data-limited scenario, we show the detailed performance on CASIA-B in Tab. 2, which indicates that MetaGait improves the performance by a large margin in most viewpoints and conditions.

⋆ Co-corresponding authors.

Table 2: The performance comparison between MetaGait and other methods.

Gallery NM #1-4			0° − 180°											Mean
Probe			0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	
ST	NM	CNN-LB [6]	54.8	–	–	77.8	–	64.9	–	76.1	–	–	–	–
		GaitSet [1]	64.6	83.3	90.4	86.5	80.2	75.5	80.3	86.0	87.1	81.4	59.6	79.5
		MT3D [4]	71.9	83.9	90.9	90.1	81.1	75.6	82.1	89.0	91.1	86.3	69.2	82.8
		GaitGL [5]	77.0	87.8	93.9	92.7	83.9	78.7	84.7	91.5	92.5	89.3	74.4	86.0
		3DLocal [3]	71.2	86.2	93.2	91.4	83.5	77.9	84.4	90.3	92.4	86.6	68.4	84.1
		MetaGait	**79.5**	**90.7**	**95.8**	**95.0**	**85.8**	**81.0**	**87.1**	**93.6**	**94.7**	**92.8**	**77.5**	**88.5**
	BG	GaitSet	55.8	70.5	76.9	75.5	69.7	63.4	68.0	75.8	76.2	70.7	52.5	68.6
		MT3D	64.5	76.7	82.8	82.8	73.2	66.9	74.0	81.9	84.8	80.2	63.0	74.0
		GaitGL	68.1	81.2	87.7	84.9	76.3	70.5	76.1	84.5	87.0	83.6	65.0	78.6
		3DLocal	66.4	78.8	84.9	84.5	75.5	68.8	75.8	83.8	86.9	82.4	65.0	76.0
		MetaGait	**68.8**	**82.5**	**88.6**	**85.5**	**78.2**	**72.5**	**78.7**	**86.4**	**88.3**	**85.3**	**66.8**	**80.2**
	CL	GaitSet	29.4	43.1	49.5	48.7	42.3	40.3	44.9	47.4	43.0	35.7	25.6	40.9
		MT3D	46.6	61.6	66.5	63.3	57.4	52.1	58.1	58.9	58.5	57.4	41.9	56.6
		GaitGL	46.9	58.7	66.6	65.4	58.3	54.1	59.5	62.7	61.3	57.1	40.6	57.4
		3DLocal	**51.0**	**65.6**	**70.5**	67.3	61.0	56.1	62.0	63.3	62.9	**61.5**	45.9	60.6
		MetaGait	50.1	62.7	69.2	**69.4**	**62.2**	**67.9**	**63.4**	**66.2**	**63.9**	60.0	**46.6**	**61.9**
MT	NM	AE [7]	49.3	61.5	64.4	63.6	63.7	58.1	59.9	66.5	64.8	56.9	44.0	59.3
		MGAN [2]	54.9	65.9	72.1	74.8	71.1	65.7	70.0	75.6	76.2	68.6	53.8	68.1
		GaitSet	86.8	95.2	98.0	94.5	91.5	89.1	91.1	95.0	97.4	93.7	80.2	92.0
		MT3D	91.9	96.4	98.5	95.7	93.8	90.8	93.9	97.3	97.9	95.0	86.8	94.4
		GaitGL	93.9	97.6	98.8	97.3	95.2	92.7	95.6	98.1	98.5	96.5	91.2	95.9
		MetaGait	**94.6**	**98.0**	**99.4**	**98.3**	**96.6**	**93.8**	**96.7**	**98.5**	**98.8**	**97.7**	**92.6**	**96.8**
	BG	AE	29.8	37.7	39.2	40.5	43.8	37.5	43.0	42.7	36.3	30.6	28.5	37.2
		MGAN	48.5	58.5	59.7	58.0	53.7	49.8	54.0	51.3	59.5	55.9	43.1	54.7
		GaitSet	79.9	89.8	91.2	86.7	81.6	76.7	81.0	88.2	90.3	88.5	73.0	84.3
		MT3D	86.7	92.9	94.9	92.8	88.5	82.5	87.5	92.5	95.3	92.9	81.2	89.8
		GaitGL	88.5	95.1	95.9	94.2	91.5	85.4	89.0	95.4	97.4	94.3	86.3	92.1
		MetaGait	**90.1**	**97.3**	**97.7**	**96.1**	**93.0**	**87.7**	**90.8**	**97.2**	**98.5**	**95.9**	**89.1**	**94.0**
	CL	AE	18.7	21.0	25.0	25.1	25.0	26.3	28.7	30.0	23.6	23.4	19.0	24.2
		MGAN	23.1	34.5	36.3	33.3	32.9	32.7	34.2	37.6	33.7	26.7	21.0	31.5
		GaitSet	52.0	66.0	72.8	69.3	63.1	61.2	63.5	66.5	67.5	60.0	45.9	62.5
		MT3D	67.5	81.0	85.0	80.6	75.9	69.8	76.8	81.0	80.8	73.8	59.0	75.6
		GaitGL	70.7	83.2	87.1	84.7	78.2	71.3	78.0	83.7	83.6	77.1	63.1	78.3
		MetaGait	**76.5**	**88.5**	**92.3**	**89.5**	**82.4**	**76.7**	**83.8**	**87.4**	**89.5**	**84.1**	**68.4**	**83.5**

References

1. Chao, H., He, Y., Zhang, J., Feng, J.: GaitSet: Regarding gait as a set for cross-view gait recognition. In: AAAI (2019)
2. He, Y., Zhang, J., Shan, H., Wang, L.: Multi-task gans for view-specific feature learning in gait recognition. IEEE Trans. Inf. Forensics Secur. **14**(1), 102–113 (2018)
3. Huang, Z., Xue, D., Shen, X., Tian, X., Li, H., Huang, J., Hua, X.S.: 3d local convolutional neural networks for gait recognition. In: Int. Conf. Comput. Vis. pp. 14920–14929 (October 2021)
4. Lin, B., Zhang, S., Bao, F.: Gait recognition with multiple-temporal-scale 3d convolutional neural network. In: ACM Int. Conf. Multimedia. pp. 3054–3062 (2020)
5. Lin, B., Zhang, S., Yu, X.: Gait recognition via effective global-local feature representation and local temporal aggregation. In: Int. Conf. Comput. Vis. pp. 14648–14656 (October 2021)
6. Wu, Z., Huang, Y., Wang, L., Wang, X., Tan, T.: A comprehensive study on cross-view gait based human identification with deep cnns. IEEE Trans. Pattern Anal. Mach. Intell. **39**(2), 209–226 (2016)
7. Yu, S., Wang, Q., Shen, L., Huang, Y.: View invariant gait recognition using only one uniform model. In: Int. Conf. Pattern Recog. pp. 889–894 (2016)