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

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

[1] School of Artificial Intelligence, Beijing Normal University
[2] School of Computer Science and Technology, Beihang University
[3] WATRIX.AI
[4] AI Lab, Lenovo Research

## 1 Supplementary Material

The supplementary material includes:

- Privacy protection and authorization.
- Ablation study of MTE on Gait3D [4].
- Detailed architectures of GaitMoE for OccGait, OccCASIA-B [3], Gait3D [4] and GREW [5].

### 1.1 Ethical Statement

In this work, we strictly follow ECCV ethical research standards and laws. During OccGait collection, we obtain authorization from all subjects who are informed for academic data collection in advance. OccGait, as well as other databases obtained from publicly available sources in our study only use with silhouettes for anonymization to protect personal privacy. Privacy is also the highest priority in our research. We hope to balance AI development and privacy protection.

### 1.2 Ablation Study

We provide a detailed ablation study on Gait3D to validate the design of MTE. Tab. 1 shows that a smaller S or larger K degrades original information, and a larger S or smaller K restricts dynamic information.

### 1.3 Network Architecture

We give further details of GaitMoE on OccGait, OccCASIA-B, Gait3D and GREW. GaitMoE on OccGait and OccCASIA-B shown in Tab. 2, we take Conv Block (*i.e.*, 2D CNNs) as C2D Block, and set channels to [64, 128, 256]. GaitMoE-T on Gait3D and GREW shown in Tab. 3, we take Residual Block [1,2] as C2D Block, and set channels to [64, 128, 256, 512] for better convergence. For GaitMoE-M, Residual Block2 and Residual Block3 are stacked four times.

⋆ Equal contribution
† Corresponding author

**Table 1:** Comparisons on Gait3D.

| Analysis on (S, K) in MTE (Rank-1 (%) on Gait3D) | | | | |
|---|---|---|---|---|
| (4, 4) | (8, 4) | (16, 4) | (8, 1) | (8, 8) |
| 70.8 | **71.3** | 69.9 | 68.4 | 69.7 |

**Table 2:** GaitMoE on OccGait, OccCASIA-B. Conv Block consists of two convolution layers with a stride of 1. Dila denotes dilated ratio. MTE divides the channels into 8 segments, with dilated ratios of 1, 2, 3, and 4 assigned to four of them, while the remaining four channels undergo an identity mapping.

| OccGait & OccCASIA-B | | |
|---|---|---|
| Layers | Output size | Kernel |
| Conv Block1 | (T, 64, 64, 44) | (1, 3, 3, 64)<br>(1, 3, 3, 64) |
| MTE1 | (T, 64, 64, 44) | (3, 1, 1, 8)<br>(Dila=1, 2, 3, 4) |
| Conv Block2 | (T, 128, 64, 44) | (1, 3, 3, 128)<br>(1, 3, 3, 128) |
| MTE2 | (T, 128, 64, 44) | (3, 1, 1, 16)<br>(Dila=1, 2, 3, 4) |
| Conv Block3 | (T, 256, 64, 44) | (1, 3, 3, 256)<br>(1, 3, 3, 256) |
| MTE3 | (T, 256, 64, 44) | (3, 1, 1, 32)<br>(Dila=1, 2, 3, 4) |
| HP | (T, 256, 64, 1) | Horizontal Pooling |
| MAE | (1, 256, 64, 1) | Linear=(256, 256)<br>Expert=(16, 16) |
| Head | (1, 256, 64, 1) | 64 Separate FCs<br>BNNecks |

**Table 3:** GaitMoE-T on Gait3D, GREW. Residual Block is the basic block in ResNet [1, 2]. Down-sampling is performed by Residual Block2 and Residual Block3 with a stride of 2. Dila denotes dilated ratio. MTE divides the channels into 8 segments, with dilated ratios of 1, 2, 3, and 4 assigned to four of them, while the remaining four channels undergo an identity mapping.

| Gait3D & GREW | | |
|---|---|---|
| Layers | Output size | Kernel |
| Conv0 | (T, 64, 64, 44) | (1, 3, 3, 64) |
| Residual Block1 | (T, 64, 64, 44) | (1, 3, 3, 64)<br>(1, 3, 3, 64) |
| MTE1 | (T, 64, 64, 44) | (3, 1, 1, 8)<br>(Dila=1, 2, 3, 4) |
| Residual Block2 | (T, 128, 32, 22) | (1, 3, 3, 128)<br>(1, 3, 3, 128) |
| MTE2 | (T, 128, 32, 22) | (3, 1, 1, 16)<br>(Dila=1, 2, 3, 4) |
| Residual Block3 | (T, 256, 16, 11) | (1, 3, 3, 256)<br>(1, 3, 3, 256) |
| MTE3 | (T, 256, 16, 11) | (3, 1, 1, 32)<br>(Dila=1, 2, 3, 4) |
| Residual Block4 | (T, 512, 16, 11) | (1, 3, 3, 512)<br>(1, 3, 3, 512) |
| MTE4 | (T, 512, 16, 11) | (3, 1, 1, 64)<br>(Dila=1, 2, 3, 4) |
| HP | (T, 512, 16, 1) | Horizontal Pooling |
| MAE | (1, 512, 16, 1) | Linear=(512, 512)<br>Expert=(16, 32) |
| Head | (1, 256, 16, 1) | 16 Separate FCs<br>BNNecks |

## References


1. Fan, C., Liang, J., Shen, C., Hou, S., Huang, Y., Yu, S.: Opengait: Revisiting gait recognition towards better practicality. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9707–9716 (2023) 1, 2
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016) 1, 2
3. Peng, Y., Cao, C., He, Z.: Occluded gait recognition. In: 2023 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2023) 1
4. Zheng, J., Liu, X., Liu, W., He, L., Yan, C., Mei, T.: Gait recognition in the wild with dense 3d representations and a benchmark. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20228–20237 (2022) 1
5. Zhu, Z., Guo, X., Yang, T., Huang, J., Deng, J., Huang, G., Du, D., Lu, J., Zhou, J.: Gait recognition in the wild: A benchmark. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 14789–14799 (2021) 1