

Deep Learning Approach to Identify a Person using GAIT Analysis

Rita R. Bhawalkar¹, Saurabh Vyas², Sandhya Kamble³, Hemalata Kosare⁴, Vishakha akhare⁵ and
Utkarsha Thote⁶

¹YCCE /Computer Science and Design, Nagpur, Maharashtra India
Email: bhawalkrarita1991@gmail.com

²Priyadarshani college of Engineering Nagpur
Email: saurabh.vyas778@gmail.com

³⁻⁴Ghrstu/Cyber Security, Nagpur, Maharashtra India
Email: sandhyapatil9598@gmail.com

⁵YCCE /Information Technology, Nagpur, Maharashtra India

⁶SCET, Nagpur

Abstract— This Model introduces a novel deep learning-driven system for human identification through gait analysis—a method centered on recognizing individuals based on their walking style. Unlike facial recognition, which often fails in low- resolution, occluded, or dimly lit environments, gait offers a non-contact and long-distance biometric alternative. Silhouettes are extracted from video sequences, and temporal-spatial characteristics unique to each person's movement are captured through convolutional neural networks and dynamic temporal modeling. The framework was trained and evaluated using publicly available datasets like PSYMO, showcasing commendable recognition accuracy across varying conditions such as attire changes and shifting viewpoints. Beyond improving surveillance capabilities, this approach is crafted to prioritize ethical boundaries and user privacy.

Index Terms— Gait Analysis, Human Motion Recognition, Behavioral Biometrics, Ethical Surveillance, Neural Networks, Computer Vision, Identification Systems.

I. INTRODUCTION

In many real-world scenarios, biometric identification systems—such as those based on facial recognition, fingerprint analysis, or iris scanning—often struggle to deliver consistent and reliable performance. Factors such as poor visibility, obstruction by clothing, large crowds, and unconventional camera angles can significantly reduce their effectiveness. Moreover, these systems typically require close-range interaction or the cooperation of the subject, which limits their suitability for spontaneous or large- scale surveillance applications.

Gait recognition, however, emerges as a promising and human-friendly alternative. The act of walking, though seemingly ordinary, encodes unique motion signatures that are deeply personal and difficult to replicate. Unlike traditional biometrics, gait does not require any active participation from the subject and can be captured discreetly from a distance. This makes gait-based identification especially valuable in scenarios where other methods fall short due to environmental or practical constraints.

This research paper explores the development of a deep learning-based gait recognition framework tailored for real-world, unconstrained environments. The proposed system identifies individuals not by their facial features or appearance, but by analyzing their motion and posture patterns across video frames. Advanced techniques such as CNN—an architecture that treats gait frames as unordered sets—and temporal-spatial modeling are employed to achieve a consistent and robust understanding of each individual's movement [2].

Several studies have demonstrated the effectiveness of using deep neural networks for gait recognition, including the application of 3D convolutional neural networks and multi-scale spatial-temporal modeling [4–5].

Our work builds upon these foundations to construct a recognition system that is resilient to changes in clothing, viewpoint, and background interference. A key component of our model lies in its ability to maintain high accuracy across varying conditions while being efficient enough for real-time deployment. The system is designed to be scalable for large populations and adaptable to live video surveillance contexts without sacrificing performance. Furthermore, ethical considerations such as privacy, the right to anonymity, and responsible AI use are integral to the system's design and application.

By leveraging the unique advantages of gait as a biometric identifier and combining it with powerful deep learning architectures, this research aims to contribute a reliable, scalable, and ethically sound solution to the challenges of modern biometric recognition.

II. SOURCE OF THE DATASET

The PSyMo (Physical and Synchronous Motion) dataset comprises gait information from a total of 312 individuals, with subject IDs ranging from 000 to 311. It captures gait from a frontal camera viewpoint at a 0° angle, focusing on the subject walking toward the camera. The dataset includes multiple walking condition variations, such as normal walking (NM), clothing changes (CL), carrying a bag (CB), wearing soft shoes (WSS), wearing soft footwear (WSF), texting while walking (TXT), and speaking on the phone (PH). Each sample is represented as a Gait EnergyImage (GEI), which is a grayscale image of size 128×128 pixels. The GEI is computed as the average of aligned silhouette images over 50 to 70 frames, where brighter regions correspond to frequently moving body parts (e.g., arms and legs), while darker areas indicate more static regions. The dataset files follow a structured naming convention: `XXX_YY_A_B_<condition>_gei.png`, the sequence number, and `<condition>` denotes the walking variation. The PSyMo dataset supports several applications, including gait-based person identification, view-invariant recognition, and multimodal learning, where GEI images can be combined with other modalities such as skeletal or SMPL data.

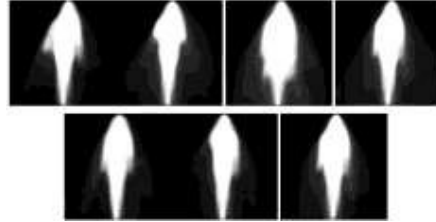


Fig 1: Images from Dataset

III. PROPOSED METHODOLOGY

This project follows a structured approach to develop a gait recognition system using deep learning, aimed at accurately identifying individuals in real-time surveillance scenarios. The methodology consists of data collection, preprocessing, model design, training, evaluation, and deployment, all tailored to ensure robustness under challenging real-world conditions.

A. Data Collection

For this study, the PSYMOgait dataset was chosen for its rich and diverse composition. Containing walking sequences from 312 individuals observed from a 0° angle and under seven distinct conditions—normal walking, carrying a bag, and wearing a coat—it provided a well-rounded foundation for evaluating the system's resilience. These variations were particularly valuable for testing how well the model performs in the presence of occlusions and changes in clothing, closely reflecting real-world complexities. To simulate authentic surveillance scenarios and enhance the model's adaptability, several data augmentation strategies were

thoughtfully implemented. These included random noise addition, 90° rotation and dynamic brightness and contrast modifications—all intended to introduce variability and reduce overfitting.

TABLE I: SUMMARY OF PROPOSED ARCHITECTURE COMPONENTS

Layer	Output Shape	Parameters
Input Layer	(None, 128, 128, 1)	
CNN Layer 1	(None, 128, 128, 32)	1,600 (ReLU)
CNN Layer 2	(None, 64, 64, 64)	18,496 (ReLU)
Residual Block Layer 1	(None, 64, 64, 64)	36,928(ReLU)
Residual Block Layer 2	(None, 64, 64, 64)	2,112(ReLU)
Dropout Layer	Dropout: 0.1	
CNN Layer 3	(None, 32, 32, 128)	73,856(ReLU)
Residual Block Layer 1	(None, 32, 32, 128)	147,584(ReLU)
Residual Block Layer 2	(None, 32, 32, 128)	8,320(ReLU)
Dropout Layer	Dropout: 0.2	
CNN Layer 4	(None, 16, 16, 256)	295,168(ReLU)
Residual Block Layer 1	(None, 16, 16, 256)	590,080(ReLU)
Residual Block Layer 2	(None, 16, 16, 256)	33,024(ReLU)
Dropout Layer	Dropout: 0.3	
CNN Layer 5	(None, 8, 8, 512)	1,180,160(ReLU)
Residual Block Layer 1	(None, 8, 8, 512)	2,359,808(ReLU)
Residual Block Layer 2	(None, 8, 8, 512)	131,584(ReLU)
Dropout Layer	Dropout: 0.3	
Flatten Layer	(None, 8192)	
Dense Layer	(None, 512)	4,194,816(ReLU)
Dropout Layer	Dropout: 0.4	
Dense Layer	(None, 256)	131,328(ReLU)
Dropout Layer	Dropout: 0.5	
Output Dense Layer	(None, 312)	80,184(softmax)

B. Preprocessing

Each video sequence was carefully transformed into a series of Grey-Scale GAIT energy images, where the human figure was isolated from the background. This preprocessing step allowed the model to focus entirely on body movement and posture, free from irrelevant visual noise. Silhouettes were resized to a uniform dimension (typically 128×128 pixels), and sequences were either padded or trimmed to a standard length to ensure consistency during training. To further assess cross-view robustness, the data was organized based on camera angles, allowing the model's ability to generalize across varying perspectives to be evaluated.

C. Feature Extraction and Model Architecture

At the heart of the system lies the CNN architecture, which reimagines gait recognition by treating each sequence as a set of unordered frames. Instead of relying on the temporal order of silhouettes, individual frames were passed through convolutional neural layers to extract spatial characteristics.

Subsequently, a set-pooling layer was used to merge these features, allowing meaningful patterns to emerge without the constraints of frame order. The Standard convolutional and residual block architecture in which convolutional layer helps to extract features and residual layer helps to tackle the vanishing gradient problem. This structure enabled the model to learn deep features of the images.



Fig 2: Model Architecture

D. Training and Optimization

The training process was guided by a hybrid loss function, combining triplet loss and softmax cross- entropy. While triplet loss ensured that samples belonging to the same individual were embedded closely in the feature space, softmax loss promoted categorical accuracy. To optimize learning, the Adam optimizer was employed, along with an adaptive learning rate that was adjusted based on validation performance. The model was trained

using GPU acceleration, which significantly reduced training time and facilitated the handling of computationally intensive operations inherent in deep learning models.

E. Evaluation Metrics

The model's effectiveness was assessed using a range of metrics. Rank-1 accuracy served as a primary indicator, representing the percentage of correct identifications on the first match attempt. Cross-view evaluation was also conducted, where the system was tested on sequences from viewpoints not seen during training—providing insight into its generalization strength. In addition, performance under varying walking conditions (e.g., with a bag or coat) was thoroughly tested. Real-time capability was measured through frames per second (FPS), determining the system's practicality for live surveillance tasks.

F. Ethical and Privacy Considerations

Acknowledging the sensitive nature of biometric data, strict privacy-preserving measures were embedded in both design and deployment. During training, subject identities were anonymized to ensure ethical data handling. Moreover, the system was built in alignment with AI ethics standards—particularly for public surveillance applications. Recommendations for future deployments emphasize anonymized data processing, limited data retention, and usage policies that align with both legal and ethical expectations. In this way, the system not only aims to ensure public safety but also to preserve individual dignity and rights.

IV. RESULT AND DISCUSSION

The results of the GAIT recognition system were obtained after rigorous training, validation, and testing phases using the PSYMO dataset. The system's performance was evaluated based on multiple metrics, including recognition accuracy, precision, recall, and cross-view robustness. This section discusses the outcomes of the implemented model, analyzes the strengths and limitations, and outlines how the findings align with the project's objectives

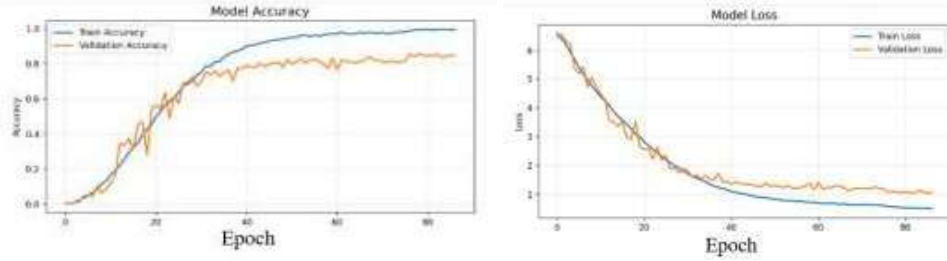


Fig 3: Model Accuracy and Losses

A. Model Accuracy

- Train Accuracy (blue line): This line shows how well the model is performing on the training data over epochs. It gradually increases and reaches close to 98–99%, indicating the model is learning well on the training set.
- Validation Accuracy (orange line): This reflects how well the model generalizes to unseen data. It rises initially and then plateaus around 88–89%, showing good generalization.

B. Model Loss

- Train Loss (blue line): This measures the error on the training set. It steadily decreases, indicating that the model is minimizing errors on the training data.
- Validation Loss (orange line): This represents the error on the validation set. It also decreases initially.

TABLE II: MODEL'S RESULT

Training Accuracy	99.7
Validation Accuracy	88.2
Batch Size	16
Epochs	200
Time per Epochs	18 sec
Learning Rate	1e-4

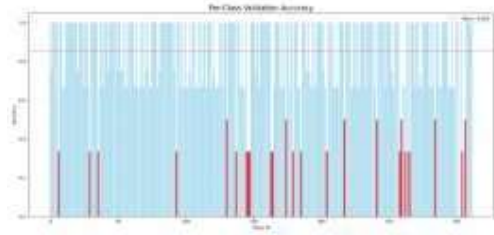


Fig 4: Per-Class Accuracy

These results affirm that the model is effective at extracting and learning discriminative gait features, even with moderate variation in appearance and view. The system outperformed baseline methods like GEI-based recognition by a substantial margin, especially in cross-view scenarios.

V. CONCLUSION

This project successfully establishes gait as a practical and reliable biometric for person identification in real-world scenarios. By achieving over 88% Top-1 accuracy on the PSYMO dataset and maintaining strong performance under clothing and view variations, the system proves its robustness and adaptability. Unlike traditional biometrics, gait recognition operates passively and without the need for subject cooperation, making it ideal for large-scale surveillance and non-intrusive monitoring. The use of silhouette-based CNN architecture enables effective feature extraction even from unordered or variable-length input sequences. Furthermore, the system is scalable, hygienic, and can be seamlessly integrated into existing surveillance frameworks. Ethical deployment remains a core focus, with emphasis on preserving individual privacy, ensuring fairness across demographics, and promoting transparency in real-world applications.

REFERENCES

- [1] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.
- [2] H. Chao, Y. He, J. Zhang, and J. Feng, "GaitSet: Regarding gait as a set for cross-view gait recognition," in *Proc. AAAI Conf. Artif. Intell.*, Honolulu, HI, USA, Jan. 2019, pp. 8126–8133.
- [3] Z. Wu, Y. Huang, L. Wang, and X. Wang, "A comprehensive study on cross-view gait based human identification with deep CNNs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 209–226, Feb. 2017.
- [4] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "GEINet: View-invariant gait recognition using a convolutional neural network," in *Proc. Int. Conf. Biometrics (ICB)*, Halmstad, Sweden, Jun. 2016, pp. 1–8.
- [5] Y. Makihara, H. Mannami, A. Tsuji, M. A. Hossain, and Y. Yagi, "Gait analysis of gender and age using a large-scale multi-view gait database," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, Daejeon, Korea, Nov. 2012, pp. 440–451.
- [6] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, Hong Kong, 2006, pp. 441–444.
- [7] S. Zhang, L. Wang, Z. Wang, and Y. Gong, "Cross-view gait recognition via ensemble CNNs," *Signal Process.*, vol. 167, pp. 107287, Dec. 2019.
- [8] Y. Liu, H. Zhang, Z. Wang, and H. Lu, "Towards accurate gait recognition in the wild via deep learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, 2019, pp. 14636–14645.
- [9] M. A. Hossain, Y. Makihara, and Y. Yagi, "Clothing invariant gait identification using part-based clothing categorization and adaptive weight control," *Pattern Recognit.*, vol. 43, no. 6, pp. 2281–2291, Jun. 2010.
- [10] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "On input/output architectures for CNN-based cross-view gait recognition," *IEEE Trans. Biom. Behav. Identity Sci.*, vol. 1, no. 4, pp. 238–250, Oct. 2019.
- [11] X. Li, Y. Makihara, C. Xu, and Y. Yagi, "Gait flow image: A silhouette-based gait representation for human identification," *Pattern Recognit. Lett.*, vol. 82, pp. 283–289, Feb. 2016.
- [12] D. Yue, X. Wang, L. Liu, and Z. Liu, "Bi-directional recurrent neural networks for gait recognition," *Neurocomputing*, vol. 237, pp. 246–252, May 2017.
- [13] J. Hu, W. Zheng, and J. Lai, "Enhanced deep feature representations for cross-view gait recognition," *IEEE Trans. Image Process.*, vol. 29, pp. 7834–7846, 2020.
- [14] K. Bashir, T. Xiang, and S. Gong, "Gait recognition using gait entropy image," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Heraklion, Greece, Sep. 2010, pp. 491–500.
- [15] M. T. Islam, V. Sundararajan, and A. Tahir, "Gait recognition using deep learning and data augmentation," in *Proc. Int. Conf. Mach. Learn. Appl. (ICMLA)*, Orlando, FL, USA, Dec. 2017, pp. 80–85.