NUS School of Computing Summer Research Internship 2024 Report

Domain Adaptation of Gait Recognition Algorithms

Submitted by

Fatima Nuzhat

Summer Research Intern

Media Research Lab (Computer Vision)

Department of Computer Science

School of Computing

National University of Singapore

Advisor

Terence Sim

Associate Professor

Contents

1 Abstract **1**

2 Introduction **2**

2.1 Project Scope 3

2.2 Objectives 3

3 Internship Experience **5**

3.1 Phase 1: Background Study 5

3.2 Phase 2: Novel Triplet Loss Implementation 6

3.3 Phase 3: Experimentation 8

3.4 Learnings & Challenges 9

4 Conclusion **10**

References **10**

Chapter 1

Abstract

During my 10-week internship under the guidance of Dr. Terence Sim and Dr. Sanka Rasnayaka, I focused on improving the generalizability of deep learning models for gait recognition. While existing models performed well on the datasets they were trained on, their accuracy dropped significantly when applied to larger datasets. My goal was to explore ways to enhance their adaptability to new data.

In the initial weeks, I studied the newly introduced CCGR dataset and a triplet model developed by Habib et al., which utilizes a triplet loss— a common approach in recognition tasks. Inspired by this model, we aimed to create a novel loss function that not only considers the subject's silhouette but also factors in other characteristics of their gait, such as viewing angle and variations in clothing. This led to the development of a three-way loss function that was subsequently implemented and tested in the following weeks. The results of these training attempts are presented in this report.

I am deeply grateful for the opportunity to work under the mentorship of Dr. Sim and Dr. Rasnayaka. Their guidance, along with the experience gained during this project, has significantly broadened my understanding of deep learning and gait recognition. I also want to express my sincere thanks to the NUS School of Computing (SOC) Summer Research Internship (SRI) for providing such a supportive and inspiring research environment. This internship has not only honed my technical skills but also inspired me to continue exploring advancements in this field.

Chapter 2

Introduction

Gait recognition, the process of identifying individuals based on their walking patterns, has gained increasing attention within the fields of computer vision and biometric authentication. Unlike other biometric techniques such as facial recognition or fingerprint scanning, gait recognition has the unique advantage of being able to operate at a distance, even when the subject's identity is partially obscured or their facial features are not clearly visible [1]. This makes it particularly useful in areas such as surveillance, security [2], and healthcare monitoring [3].

However, despite its potential, the generalizability of gait recognition models remains a significant challenge. Many existing models [4, 5, 6, **?**] perform exceptionally well on specific datasets [7, 8] but struggle to maintain accuracy when faced with larger or more diverse datasets [9]. Real-world scenarios often introduce factors such as varying camera angles, changes in clothing, and the dynamic nature of human motion, all of which can drastically affect the performance of these models.

During my 10-week internship, I worked to address this issue of generalizability. The core focus of my project was to enhance the capability of deep learning models to adapt to new and larger datasets, without suffering a significant drop in accuracy. Building on a well-established recognition framework that uses triplet loss [10], I sought to develop a more comprehensive loss function. This approach was designed to improve the model's resilience to changes in gait characteristics, such as the angle of observation and variations in clothing, making it more robust in real-world applications.

2.1 Project Scope

The scope of my internship was broad, covering several critical aspects of gait recognition. The overarching goal was to explore and develop techniques that would allow deep learning models to better generalize across different datasets. This involved extensive research into the current state-of-the-art models [4, 5, ?] and loss functions, particularly those that employ triplet loss mechanisms for recognition tasks.

One of the key challenges addressed during the internship was the limitations of traditional silhouette-based gait recognition models. To overcome this, my work involved the development of a novel triplet loss function that integrates multiple covariates. This loss function was designed to simultaneously consider silhouette, angle, and clothing differences, providing a more holistic representation of the subject's gait.

My responsibilities extended beyond theoretical research. A significant part of the internship involved implementing and testing the proposed loss function using the CCGR dataset, which is characterized by its diversity in covariates [9]. The experiments I conducted focused on measuring the performance of the new loss function in comparison to existing methods, with the goal of demonstrating its ability to improve model generalizability. Additionally, I took part in analyzing the results and refining the model to enhance its adaptability to various gait recognition scenarios.

In essence, the scope of this internship was not limited to understanding and solving a single technical issue but also encompassed addressing practical challenges that arise in real-world applications of gait recognition technology. The insights and methods developed throughout this project contribute toward building more reliable and flexible models that can be deployed in a wider range of settings, potentially leading to more secure and efficient biometric systems.

2.2 Objectives

The primary objectives of this internship were focused on improving the generalization capability of deep learning models for gait recognition. These objectives were structured to address both theoretical and practical challenges in the field. The specific goals of the internship were as follows:

- Enhance Model Generalization Develop methods to improve the adaptability of existing gait recognition models, particularly in scenarios involving larger and more diverse datasets, to ensure consistent accuracy across varying data conditions.

- Study the CCGR Dataset Analyze the newly introduced CCGR dataset to understand its structure and characteristics, with a specific focus on how its diversity (e.g., changes in viewing angles and clothing) can challenge existing models.

- Implement a Novel Triplet Loss Function Design and implement a new triplet loss function that incorporates additional gait-related factors such as silhouette, viewing angle, and clothing conditions, aiming to enhance the robustness of gait recognition models.

- Experimentation and Evaluation Conduct experiments using the CCGR dataset to evaluate the performance of the proposed loss function. Compare the results with baseline models to measure improvements in generalization and recognition accuracy.

Chapter 3

Internship Experience

The 10-week internship provided a comprehensive and hands-on experience in the field of deep learning and gait recognition. Throughout the duration, my work progressed through three key phases: background study, algorithm development and low-level testing, and experimentation and result analysis. Each phase presented unique challenges and learning opportunities, from familiarizing myself with the fundamentals of gait recognition and the CCGR dataset to developing and refining the new triplet loss function. These phases together laid the foundation for a deeper understanding of both the theoretical and practical aspects of model generalization in gait recognition.

3.1 Phase 1: Background Study

Recent advancements in gait recognition have led to the development of several deep learning-based frameworks aimed at improving the accuracy and generalizability of gait recognition models. OpenGait is one such open-source framework designed for the efficient implementation and evaluation of various gait recognition algorithms [11]. Its flexibility allows for the integration of diverse architectures and loss functions, making it a valuable resource for researchers working to enhance model performance. As a result, it has become widely adopted in the gait recognition community for benchmarking and experimentation.

BigGait [12] and DeepGait [6] have also contributed significantly to the field, each pushing the boundaries of how deep learning models can extract and interpret gait features. BigGait focuses on scaling up gait recognition models by introducing large-scale datasets and more complex network architectures, addressing the need for handling diverse data at scale. DeepGait, on the other hand, emphasizes the development of deeper networks specifically tailored for learning rich gait representations. Both models achieve high accuracy on standard datasets, but their performance often declines when exposed to covariate shifts, such as changes in viewing angle, clothing, or background conditions, and also mixed covariates [9].

The research conducted by Habib et al. delves into one of the primary challenges in gait recognition: the domain gap caused by varying viewing angles [13]. This research highlights how most gait recognition models are biased toward specific viewpoints, which limits their performance in real-world scenarios where camera angles are often unpredictable. The authors propose a unsupervised domain adaptation technique (GOUDA) to mitigate this issue, allowing models to adapt more effectively across different viewpoints without requiring extensive labeled data.

Addressing covariate variations is a key focus in Cross-Covariate Gait Recognition, which introduces the CCGR dataset [9]. This dataset includes a wide range of covariates such as changes in clothing, carrying conditions, and viewing angles, presenting a more realistic benchmark for evaluating the adaptability of gait recognition models. The paper emphasizes the importance of designing models and loss functions that can be generalized across these covariates, highlighting the limitations of traditional silhouette-based approaches. The CCGR dataset, therefore, has become a crucial resource for testing how well models can handle these real-world challenges.

Together, these works provided the foundation for my project, which aimed to improve the generalization of gait recognition models by accounting for covariates such as viewing angle and clothing. By leveraging insights from these papers, I developed a novel triplet loss function that incorporates multiple aspects of gait, moving beyond silhouette-based recognition to enhance performance across varied data conditions.

3.2 Phase 2: Novel Triplet Loss Implementation

In this project, a novel triplet loss function was developed to improve the generalization capability of gait recognition models. Unlike traditional triplet loss [10, 13], which only considers the distance between positive and negative samples based on their identity, our formulation incorporates covariate variations such as viewing angle and clothing conditions. The overall loss function, $\mathcal{L}$, is defined as a combination of three separate triplet losses:

$$\mathcal{L} = \beta \cdot \mathcal{L}_b + \alpha \cdot \mathcal{L}_v + \mathcal{L}_c$$

where $\mathcal{L}_b$, $\mathcal{L}_v$, and $\mathcal{L}_c$ represent the triplet losses based on variations in body silhouette, viewing angle, and covariates (e.g., clothing) respectively. The parameters α and β are hyperparameters set to balance the contributions of each loss term, initially tuned to $\alpha = 0.01$ and $\beta = 0.001$.

Covariate Triplet Loss ($\mathcal{L}_c$):

The covariate-based triplet loss, $\mathcal{L}_c$, focuses on variations in clothing and accessories. It is defined such that:

- The positive sample (p) has the same identity as the anchor (a) but a different covariate, while sharing the same viewing angle.
- The negative sample (n) has a different identity from the anchor but the same covariate and viewing angle.

This formulation helps the model learn to distinguish between individuals based on covariate variations, improving its robustness to such changes.

Triplet Loss ($\mathcal{L}_b$):

The silhouette-based triplet loss, $\mathcal{L}_b$, is designed to handle changes in both covariate and viewing angle. Here:

- The positive sample (p) has the same identity as the anchor but differs in both covariate and viewing angle.
- The negative sample (n) has a different identity but shares the same covariate and viewing angle as the anchor.

This loss term encourages the model to focus on identifying the overall structure and movement of a person's gait, regardless of external factors like clothing or camera angle.

Viewing Angle Triplet Loss ($\mathcal{L}_v$):

Finally, the viewing angle-based triplet loss, $\mathcal{L}_v$, addresses challenges posed by changes in the viewing angle. In this case:

- The positive sample (p) has the same identity and covariate as the anchor but a different viewing angle.
- The negative sample (n) has a different identity but the same covariate and viewing angle as the anchor.

This term helps the model become invariant to viewing angle changes, enabling it to generalize across different camera perspectives.

By combining these three triplet losses, the model is better equipped to handle the various covariate and viewing angle challenges common in real-world gait recognition tasks.

For each of the triplet losses $\mathcal{L}_b$, $\mathcal{L}_v$, and $\mathcal{L}_c$, we can define the loss as follows:

Body Silhouette Triplet Loss ($\mathcal{L}_b$)

$$\mathcal{L}_b = \sum_{i=1}^{N} \left[\max_p \|f(x_i^a) - f(x_i^p)\|_2^2 - \min_n \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$

Here, the hardest positive example is selected by maximizing the distance between the anchor a and positive p, and the hardest negative is selected by minimizing the distance between the anchor a and negative n.

Viewing Angle Triplet Loss ($\mathcal{L}_v$)

$$\mathcal{L}_v = \sum_{i=1}^{N} \left[\max_p \|f(x_i^a) - f(x_i^p)\|_2^2 - \min_n \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$

Again, for the viewing angle-based loss, we select the hardest positive p with the maximum distance from the anchor a (with the same identity but a different view), and the hardest negative n (different identity but same view).

Covariate Triplet Loss ($\mathcal{L}_c$)

$$\mathcal{L}_c = \sum_{i=1}^{N} \left[\max_p \|f(x_i^a) - f(x_i^p)\|_2^2 - \min_n \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$

For the covariate-based loss, the hardest positive is defined as the one with the same identity but a different covariate, and the hardest negative is the one with a different identity but the same covariate.

3.3 Phase 3: Experimentation

In the final phase of the internship, the newly developed triplet loss function was tested extensively on the CCGR dataset to evaluate its effectiveness. Through multiple rounds of training and testing, we observed some improvements in the model's ability to generalize across various covariate conditions, such as changes in clothing and viewing angle. While the preliminary results were promising, showing an enhancement in performance compared to baseline methods, the next step is to conduct a more comprehensive evaluation. This will involve comparing our model's performance with state-of-the-art gait recognition methods across several key benchmarks and datasets. The aim is to demonstrate the generalizability of the new loss function in diverse settings. Although this comparison is ongoing, it marks a significant milestone in our work, and this is where the internship comes to a close.

3.4 Learnings & Challenges

- **Understanding Complex Loss Functions:** One of the primary learning points was deepening my understanding of how triplet loss functions operate, particularly in the context of gait recognition, and how incorporating covariates can significantly impact model performance.
- **Working with Large Datasets:** Managing and preprocessing a dataset as extensive as the CCGR was a challenge, but it taught me valuable skills in handling real-world data with significant variability.
- **Balancing Multiple Factors in Model Design:** Learning to balance various covariates like viewing angle and clothing while designing the loss function was intellectually challenging, as it required both theoretical knowledge and practical experimentation to get right.
- **Hyperparameter Tuning:** Tuning hyperparameters such as α and β for the triplet loss function required meticulous testing and a strong understanding of the model's behavior under different configurations.
- **Model Generalization:** A major focus of this internship was improving the generalization of models and understanding the limitations of existing models was crucial in formulating our new approach.
- **Experimentation and Iteration:** Testing different versions of the loss function involved continuous experimentation and iteration, requiring a lot of patience and problem-solving to handle unexpected results.
- **Research and Comparative Analysis:** As we move into the phase of comparing our model's performance with other state-of-the-art approaches, learning how to critically analyze and benchmark our work against the broader field has been an ongoing challenge.
- **Time Management:** Given the limited duration of the internship, time management was critical. Balancing research, development, and experimentation required careful planning to ensure all goals were met.

Chapter 4

Conclusion

This internship has been a highly enriching experience, providing me with the opportunity to delve deep into the field of gait recognition and tackle some of its key challenges. By working on improving the generalization of deep learning models, I gained a strong understanding of both the theoretical and practical aspects of model design, from handling covariate variations to implementing a novel triplet loss function. The results achieved thus far are promising, with some improvements observed in handling diverse gait conditions. However, there is still much work to be done, especially in conducting comprehensive comparisons with state-of-the-art methods to fully validate the efficacy of our approach.

Overall, this internship has been invaluable in expanding my technical knowledge, problem-solving skills, and research capabilities. It has equipped me with the tools and mindset necessary to approach complex real-world problems with confidence. I am deeply grateful for the mentorship I received and for the support from the NUS School of Computing and the SRI team, who provided me with an ideal environment for learning and growth. As I conclude this chapter, I look forward to continuing my work in this exciting and evolving field, building upon the foundations laid during this internship.

References

[1] M. S. Nixon and J. N. Carter, "Automatic recognition by gait," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 2013–2024, 2006.

[2] Y. Ran, Q. Zheng, R. Chellappa, and T. M. Strat, "Applications of a simple characterization of human gait in surveillance," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 4, pp. 1009–1020, 2010.

[3] F. Sun, W. Zang, R. Gravina, G. Fortino, and Y. Li, "Gait-based identification for elderly users in wearable healthcare systems," *Information fusion*, vol. 53, pp. 134–144, 2020.

[4] H. Chao, Y. He, J. Zhang, and J. Feng, "Gaitset: Regarding gait as a set for cross-view gait recognition," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 8126–8133.

[5] B. Lin, S. Zhang, and X. Yu, "Gait recognition via effective global-local feature representation and local temporal aggregation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 14 648–14 656.

[6] C. Fan, S. Hou, Y. Huang, and S. Yu, "Exploring deep models for practical gait recognition," *arXiv preprint arXiv:2303.03301*, 2023.

[7] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 4. IEEE, 2006, pp. 441–444.

[8] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition," *IPSJ transactions on Computer Vision and Applications*, vol. 10, pp. 1–14, 2018.

[9] S. Zou, C. Fan, J. Xiong, C. Shen, S. Yu, and J. Tang, “Cross-covariate gait recognition: A benchmark,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 7, 2024, pp. 7855–7863.

[10] A. Hermans, L. Beyer, and B. Leibe, “In defense of the triplet loss for person re-identification,” *arXiv preprint arXiv:1703.07737*, 2017.

[11] C. Fan, J. Liang, C. Shen, S. Hou, Y. Huang, and S. Yu, “Opengait: Revisiting gait recognition towards better practicality,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 9707–9716.

[12] D. Ye, C. Fan, J. Ma, X. Liu, and S. Yu, “Biggait: Learning gait representation you want by large vision models,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 200–210.

[13] G. Habib, N. Barzilay, O. Shimshi, R. Ben-Ari, and N. Darshan, “Watch where you head: A view-biased domain gap in gait recognition and unsupervised adaptation,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 6109–6119.