Research paper

Gait-based age group classification with adaptive Graph Neural Network

Timilehin B. Aderinola c, Tee Connie a,*, Thian Song Ong a, Andrew Beng Jin Teoh b, Michael Kah Ong Goh c

a Faculty of Information Science and Technology, Multimedia University, Malacca, Malaysia
b School of Electrical and Electronic Engineering, College of Engineering, Yonsei University, Seoul, South Korea
c School of Computer Science, University College Dublin, Ireland

ARTICLE INFO

Dataset link: https://github.com/timiderinola/mmu_gag

Keywords:
Age classification
Feature extraction
Graph neural network

ABSTRACT

Deep learning techniques have recently been utilized for model-free age-associated gait feature extraction. However, acquiring model-free gait demands accurate pre-processing such as background subtraction, which is non-trivial in unconstrained environments. On the other hand, model-based gait can be obtained without background subtraction and is less affected by covariates. For model-based gait-based age group classification problems, present works rely solely on handcrafted features, where feature extraction is tedious and requires domain expertise. This paper proposes a deep learning approach to extract age-associated features from model-based gait for age group classification. Specifically, we first develop an unconstrained gait dataset called Multimedia University Gait Age and Gender dataset (MMU GAG). Next, the body joint coordinates are determined via pose estimation algorithms and represented as compact gait graphs via a novel part aggregation scheme. Then, a Part-Adaptive Residual Graph Convolutional Neural Network (PairGCN) is designed for age-associated feature learning. Experiments suggest that PairGCN features are far more informative than handcrafted features, yielding up to 99% accuracy for classifying subjects as a child, adult, or senior in the MMU GAG dataset. These results suggest the feasibility of deploying Artificial Intelligence-enabled solutions for access control, surveillance, and law enforcement in unconstrained environments.

1. Introduction

Although the historical origin of age measurement is uncertain, it became precise with the standardization of time and date. Humans can roughly estimate age from physical appearance when it is impossible to obtain an individual’s chronological age. However, relying on eyeballing techniques for age prediction in mission-critical applications is not feasible. This motivates research in machine-based age estimation or age group classification based on physical attributes. Application areas include but are not limited to access control, surveillance, law enforcement, and human–computer interaction (Islam et al., 2021).

The two main steps in age group classification are age-associated feature extraction and classifier design. The main challenge is extracting robust age-associated features from physical attributes such as face or gait (Liao et al., 2018). In particular, although it is possible to obtain gait from a distance with video acquisition devices, gait features are affected by covariates such as viewpoint, apparel, walking speed, walking slope, and footwear (Rani and Kumar, 2023). In unconstrained environments, gait can only be acquired using model-based or model-free vision-based techniques. Model-based approaches measure body part parameters such as lengths and range of motion via skeleton representations. On the other hand, model-free approaches extract features from silhouette images obtained after background subtraction. Though more computationally expensive, the model-based approaches are more robust to external covariates such as clothing and carried objects, thanks to advanced pose estimation techniques (Aderinola et al., 2021b).

Deep learning has been widely used for age-associated feature extraction based on model-free gait descriptors such as gait energy images (Berksan, 2019; Lau and Chan, 2023; Russel and Selvaraj, 2021; Xu et al., 2021b; Kiichat et al., 2022; Xu et al., 2021a; Zhang et al., 2018). However, acquiring model-free gait requires accurate background subtraction, which is non-trivial in unconstrained environments. On the other hand, model-based gait can be obtained without background subtraction and is less affected by covariates. However, most model-based gait approaches for age estimation and age group classification (Aderinola et al., 2021a; Begg et al., 2005; Yang and Wang, 2015; Hediyeh et al., 2013; Hema and Pitta, 2019; Punnyani et al., 2018; Yoo and Kwon, 2017; Zhang et al., 2010) merely utilize handcrafted features, which requires domain expertise.

* Corresponding author.
E-mail address: tee.connie@mmu.edu.my (T. Connie).

https://doi.org/10.1016/j.engappai.2024.109081
Received 26 August 2022; Received in revised form 4 March 2024; Accepted 30 July 2024
Available online 12 August 2024
0952-1976/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).
Recent emerging studies on Graph Neural Networks (GNNs) generalize Deep Neural Networks (DNNs) to data that cannot be represented in Euclidean space (Bronstein et al., 2017; Zhao et al., 2020; Teepe et al., 2021; Sheng and Li, 2021). Furthermore, with recent advances in human pose estimation techniques, it is now possible to obtain joint coordinates automatically and accurately. Hence, gait can be modeled with joint coordinates as a graph, such as joints as nodes and bones as edges. This allows representation learning on model-based gait with GNNs.

This paper proposes a dedicated Graph Convolutional Network (GCN), Part-Adaptive Residual Graph Convolutional Neural Network (PairGCN), to extract age-associated features from model-based gait in unconstrained environments, which has not been attempted before. The PairGCN can automatically adjust its edge weights based on the dominant body parts in the input part graph. More specifically, PairGCN learns which body parts are essential in classifying subjects as a child, adult, or senior by taking advantage of our part aggregation scheme.

We also introduce the gait in the wild (GITW) as the first dataset in the Multimedia University Gait Age and Gender datasets (MMU GAG).\(^1\) GITW is naturally diverse with respect to age, race, gender, walking surface, shoe, and dressing. We impose no restrictions based on the scene, age, gender, or ethnicity, except that we restrict the collection to video recordings of healthy individuals walking at a normal pace. Table 1 compares GITW with similar existing gait datasets collected in unconstrained environments. Accordingly, GITW contains more subjects, covers more age groups, and is more naturally diverse in terms of dressing, walking surface, gender, and recording time.

Our main contributions are itemized below.

1. We study age group classification problems using model-based gait features determined by the deep learning approach, which is the first of its kind.

2. A Part-Adaptive Residual GCN (PairGCN) is proposed for automatic extraction of age-associated gait features in video sequences, which can incorporate age-associated part information from model-based gait.

3. A gait age and gender dataset, Gait in the Wild (GITW) is collected, pre-processed, analyzed, and released for age group classification in unconstrained environments.

The overall structure of this study is shown in Fig. 1. The rest of this paper is organized as follows. First, we present a brief overview of related works on age classification in Section 2. Then, in Section 3, we describe the collection and preprocessing of the dataset used in this study. Although this work is focused on automatic feature extraction, we show how manually selected features can be extracted using pose estimation in Section 4. Then, Section 5 describes how automatic feature extraction is achieved. Finally, experiments and results are discussed in Section 6, and conclusions and recommendations are given in Section 7.

---

**Table 1**

| Dataset               | #  | Age Groups* | Scene/Time | Covariates   |
|-----------------------|----|-------------|------------|--------------|
| SOTON (Shutler et al., 2004) | 115 | –           | In, out/Day| –            |
| CASIA-B (Yu et al., 2006)   | 124 | adult       | In/Day     | CO, DR       |
| USF 1 (Phillips et al., 2002) | 74  | adult       | Out/Day    | V, TR, SH, CO, T. |
| USF 2 (Sarkar et al., 2005) | 85  | adult       | Out/Day    | V, TR, SH, CO, T. |
| Outdoor-Gait (Song et al., 2019) | 138 | –           | Various/Day| CL, scene, CO. |
| GITW (ours)            | 439 | child, adult, senior | Arbitrary | Naturally diverse. |

# Number of subjects; In: Indoor; Out: Outdoor; V: view; CO: Carried Object; DR: dressing; SH: shoe; T: time; TR: terrain/walking surface.

* \(\text{child} = [1–17] ; \text{adult} = [18–64] ; \text{senior} = [64–] \text{years}.\)

---

**Table 2**

| Descriptor | Strengths | Limitations |
|------------|-----------|-------------|
| Face (local) | Face shape and geometry contain age-associated information | Requires manual feature selection, which requires domain expertise. |
| Face (holistic) | Considers both geometry and texture. | Performance degrades due to poor image quality or occlusion. |
| Model-free gait | Lightweight and require little computational power. Most datasets are model-free. | Sensitive to carried objects and apparel. Requires background subtraction. |
| Model-based gait | Robust to carried objects and apparel. | Current techniques require manual feature selection. Few public datasets. |
| Sensor-based gait | Inertial sensors are low-cost, ubiquitous and precise. | Sensors must be carried by the subjects. |

---

2. Related works

In this section, we present a brief review of related works on age classification using gait and other biometric features. We also discuss some recent approaches that use gait graphs for person and action recognition using gait. We summarize the strengths and limitations of different descriptors used for age classification in Table 2.

2.1. Age classification and estimation problem using biometric features

The face is the most widely used biometric modality for age estimation and age group classification and has been deployed for public
use (Burt, 2020). Face features are broadly classified as local or holistic. Holistic-based features are obtained from the raw pixel data of faces (Han, 2020; Liao et al., 2018; Cao et al., 2020; Akbari et al., 2021; Xie and Pun, 2020; Sharma et al., 2022; Dagher and Barbarà, 2021). Local face features are obtained from the geometry of facial landmarks like eyes, mouths, and noses (Rahman et al., 2020; Rizwan et al., 2022). Some approaches to age group classification also use a fusion of local and holistic face features (Zhang et al., 2018). The main limitation of using face features is that performance depends on face image quality.

Other biometric features used for human age group classification include voice (Kwasny and Hemmerling, 2021; Si et al., 2022) and ear images (Moolla et al., 2021; Yaman et al., 2019). However, these modalities are limited in application as data is not easily obtained in unconstrained environments. For example, voice signals in the wild are either missing or distorted by noise, face or ear images in the wild could be partially or fully occluded.

2.2. Age prediction using gait features

Age-associated gait features can be obtained using wearable sensors (sensor-based) or from videos (vision-based). The vision-based approach could be further sub-classified as model-free or model-based. Model-free gait features are based on subjects’ shape and appearance, and extracting them involves background subtraction and silhouette generation. The most common model-free gait descriptor is the Gait Energy Image (GEI) (Han and Bhanu, 2006). Since GEIs are represented as grayscale images, deep learning has been widely leveraged to extract age-associated features from GEIs (Zhu et al., 2020; Berksan, 2019; Lau and Chan, 2023; Ruzzel and Selvaraj, 2021; Kitchat et al., 2022; Song et al., 2021; Zhang et al., 2022). In addition, some other works perform manual feature extraction from GEIs for age estimation or age group classification (Mansouri et al., 2018; Nabila et al., 2018). However, since obtaining silhouette images requires accurate background subtraction, obtaining model-free gait age features in the wild is infeasible.

Unlike model-free gait, model-based gait age features such as length of body segments, body ratios, and joint angles/range of motion can be manually extracted in the wild. These features have been classified as biological, kinematic, or hybrid (Aderinola et al., 2021b). Biological features are obtained from measuring body parts, such as limb length and head-to-body ratio. While biological features are sufficient to differentiate children from adults (Aderinola et al., 2021a), more fine-grained age group classification requires kinematic features such as gait speed and joint angles (Begg et al., 2005; Yang and Wang, 2015; Zhang et al., 2010). However, several model-based approaches to age group classification use biological and kinematic features (Aderinola et al., 2021a; Hema and Pitta, 2019; Punyani et al., 2018; Yoo and Kwon, 2017). The main limitation of these approaches is that manual feature extraction is tedious and requires domain expertise.

2.3. Gait graphs for person and action recognition

Until recently, model-based gait had been suitable for conventional machine learning techniques only. However, with the recent advances in human pose estimation, it is now possible to obtain joint coordinates automatically and with high accuracy. Hence, pose-based gait analysis is receiving increasing attention. Certain studies, such as (Liao et al., 2017), have leveraged Deep Neural Networks (DNNs) for gait recognition on model-based gait features obtained via state-of-the-art pose estimation algorithms like AlphaPose (Fang et al., 2022). Recent works such as Li et al. (2020), Liao et al. (2020), Teepe et al. (2021) have automatically leveraged graph neural networks to learn gait and action recognition features. However, they are not transferred for age group classification problems.

3. Data collection

In this section, we describe the video data collection and preprocessing steps. We also describe how motion data is obtained from the videos via pose estimation.

3.1. Dataset description

The gait in the wild (GITW) dataset contains video recordings of healthy individuals walking at a normal pace in unconstrained environments. It was collected from public domain online videos (see Fig. 2). It contains 439 subjects with one clip per subject. Since the videos were collected online, the views and covariate conditions are unconstrained. The subjects were labeled with the apparent gender and age group — child, adult, or senior. The dataset has 155 children, 177 adults, and 107 seniors.

3.2. Data preparation

Fig. 3 shows the overall process followed in data preparation. Video pre-processing, pose estimation, and coordinate smoothing are discussed in this section, while body graph construction and body parts aggregation are discussed in Section 5.2.

3.2.1. Video preprocessing

As shown in Fig. 4, GITW videos are segmented manually using a linear video editor2 to ensure only one subject in each video. In addition, each video is segmented based on the walking direction of the participants — left, right, front, or back.

3.2.2. Pose estimation

Pose estimation is performed using AlphaPose (Fang et al., 2022) on a Windows 10 desktop computer with 8 GB of RAM, 8 GB GPU, and a core i5 Intel processor. For each subject with J body keypoints in a video with F frames, the pose estimation algorithm outputs a sequence of tuples \((x_i^t, y_i^t, c_i^t)\) \(i = 1, \ldots, J; t = 1, \ldots, F\), where \((x_i^t, y_i^t)\) represent the body keypoints coordinates, and \(c_i^t = [0, 1]\) is the confidence score for joint \(i\) in frame \(t\).

Data Augmentation. Since videos recorded by participants are of arbitrary lengths, \(F\) is standardized to 50 frames. Hence, sequences longer than 50 frames were divided into separate frames. However, clipped sequences shorter than 25 frames were discarded as too short. For further augmentation, pose sequences were reversed, simulating the backward movement of subjects.

3.2.3. Coordinates smoothing

The pose estimation output can be regarded as a multivariate time series, requiring some smoothing. However, smoothing with conventional filters could result in losing relevant information. Hence, we propose z-score smoothing, which replaces sudden spikes in the coordinate values with the average of the two adjacent coordinate values. Z-score smoothing works on the assumption that the motion time series data of subjects walking at a normal pace will have no sudden spikes. In the absence of sudden spikes, the data is left unchanged. After z-score smoothing, pairs of adjacent frames are averaged for dimensionality reduction, such that a sequence containing 50 frames is reduced to 25.

Let \(\mu\) be the mean of \(d \in \mathbb{R}^2\), where \(d\) is the univariate movement time series for a body keypoint across \(F\) frames. The frame-to-frame standard deviation is:

\[
\sigma^2 = (1/F) \sum_{i=1}^{F} (d_i - \mu)^2. \quad \sigma^2 \in \mathbb{R} \tag{1}
\]

---

2 Wondershare Filmora 9, https://filmora.wondershare.com
The absolute z-scores for each element in the sequence are obtained as

$$z_i = \frac{|d_i - \mu|}{\sigma} \quad i = 1, \ldots, F; \quad z \in \mathbb{R}^F$$ (2)

A smoothened sequence $d_s \in \mathbb{R}^F$ is then obtained according to the rule:

$$\delta_i = \begin{cases} d_i, & z_i \leq \tau, \\ (d_{i-1} + d_{i+1})/2, & \text{otherwise} \end{cases}$$ (3)

where $\tau$ is a threshold value defined as $\tau = (1/F) \sum_{i=1}^{F} z_i$. The pseudocode algorithm for z-score smoothing is given in Algorithm 1. Fig. 5 illustrates the effects of z-score smoothing on a very noisy neck keypoint sequence.

**Algorithm 1: Pseudocode Procedure for Z-Score Smoothing**

- **Data:** $d$
- **Result:** $d_s$
1. initialization;
2. $\mu \leftarrow MEAN(d)$;
3. $\text{std} \leftarrow \text{STANDARDDEVIATION}(d)$;
4. $z \leftarrow \text{ABS}((d - \mu)/\text{std})$;
5. threshold $\leftarrow MEAN(z)$;
6. INITIALIZE i = 0; list $d_s$;
7. while $i < \text{LENGTH}(d)$ do
   8. if $d_i > \text{threshold}$ then
      9. $d_s_i \leftarrow MEAN(d_{i+1}, d_{i-1})$
   10. else
      11. $d_s_i \leftarrow d_i$
   12. end
   13. INCREMENT $i$;
14. end
15. RETURN $d_s$;

4. **Handcrafted gender and age-associated features extraction**

In this section, we extract handcrafted gender and age-associated gait features from the GITW dataset for analysis and comparison with automatically extracted features. Let $u = \{x_i', y_i'\}$ and $v = \{x_{i+1}', y_{i+1}'\}$ be the coordinates of two adjacent keypoints in frame $t$, the Euclidean distance, $D_E$ between $u$ and $v$ is given by:

$$D_E(u, v) = [(x_i' - x_{i+1}')^2 + (y_i' - y_{i+1}')^2]^{1/2}$$ (4)

Hence, as shown in Table 3, subjects’ gait features are obtained from the keypoint coordinates obtained via pose estimation. The statistics of
from gait graphs. First, we discuss how we construct spatiotemporal
5. Automatic age feature extraction

handcrafted features obtained from GITW are summarized in Table 4,

Fig. 4. Video with Multiple Subjects (Left). Cropped Videos (Right). GITW videos were segmented to get the clips of interest. For example, if the subject appears only in the last
20 s, all the other parts of the video are discarded. Videos with multiple subjects were cropped into different videos, ensuring that each had no more than one subject.

5.1. Preliminary: Body graph construction

From the output of pose estimation described in Section 3.2.2, the
body is modeled as a connected graph, \( \text{body} = \{\text{joints}, \text{bones}\} \), where the
detected joints are the nodes (the black dots in Fig. 7), and the bones
are edges. First, we construct the body graph for a single frame. Then,
using the constructed body graph, we construct a spatiotemporal body
graph to represent the motion across all frames.

We represent the body graph in a single frame structurally with an
adjacency matrix \( A_{\text{body}} \in \mathbb{R}^{J \times J} \), and feature-wise with \( X_{\text{body}} \in \mathbb{R}^{J \times 3} \),
where \( J = |\text{joints}| \).

The spatiotemporal body graph across all frames consists of several
body graphs, and it has spatial edges \( E_{\text{spatial}} \) and temporal edges
\( E_{\text{temporal}} \). \( E_{\text{spatial}} \) is the set of all edges of the constituent body graphs.
\( E_{\text{temporal}} \) consists of the connections between nodes of the constituent
body graphs to their counterparts in adjacent frames. An example of
\( E_{\text{temporal}} \) is the connection between the right ankle in frame 1 and the
right ankle in frame 2. Hence, we represent the walking sequence of
each subject across \( F \) frames as a spatiotemporal graph \( G = (V, E) \).
\( V = \{v_j\} = \{1, \ldots, F\} \) and \( J = |J| \) is the set of all keypoints detected
across all video frames for a subject, while \( E = (E_{\text{spatial}}, E_{\text{temporal}}) \).

Suppose \( K = |E_{\text{spatial}}| \) is the total number of bones detected across
\( N \) frames, and \( L = |E_{\text{temporal}}| \). Since each pair of adjacent frames has \( J \)
temporal connections — one for each joint, hence \( L = J(F - 2) \) and the
total number of edges, \( |E| = K + L \). Each spatiotemporal body graph
has a node feature matrix \( X \in \mathbb{R}^{|V| \times 3} \), where \( |V| \) is the number of nodes across all video frames. Structurally, the body graphs are represented by a square adjacency matrix \( A \in \mathbb{Z}^{|E| \times |E|} \) that captures both spatial
and local connections of body keypoint coordinates. The spatiotemporal
body graph is illustrated in Fig. 8.

5.2. Body parts aggregation

Human motion is not regarded as a movement of joint collections
but as an interdependent movement of body parts, which we call
aggregation of joints. For instance, with reference to Table 3, the
spine is an essential determinant of upper body posture, consisting of
the head, neck, and central hip. The arms, consisting of the shoulder,
肘, and wrist, are vital in balance in walking, and the arm’s motion
can discriminate subjects as a child, adult, or senior (Fig. 6). The body’s
center of mass is at the hip, which comprises the left, central, and
right hips. The legs are arguably the most critical parts of locomotion.
Besides being the center of pressure, dynamic features such as gait
speed, step length, and cadence are determined by leg movement,
and these are crucial in more fine-grained classification. Therefore,
we propose a part aggregation scheme that groups the body graph into four
parts: spine, arm, hip, and leg, each made up of three joints.

As illustrated in Fig. 9, each part in the part graph is associated
with four features based on the raw coordinates aggregation: (a) range
of motion (ROM) \( \theta \) at the central node, (b) total length \( L \) of the part, (c)
swing \( S \), the distance between the left and right sides, and (d) distance
\( D \), measured as the distance in pixels covered per frame. Suppose \( b \) is
### Table 4
Gait gender and age-associated handcrafted features extracted from GITW dataset.

| Feature      | Female          | Male            |
|--------------|-----------------|-----------------|
|              | child #=70      | adult #=77      | senior #=44   | child #=85 | adult #=100 | senior #=63 |
| HB           | 0.20 ± 0.05     | 0.15 ± 0.02     | 0.15 ± 0.02   | 0.20 ± 0.04 | 0.14 ± 0.02 | 0.14 ± 0.03  |
| HG           | 0.50 ± 0.06     | 0.5 ± 0.04      | 0.52 ± 0.04   | 0.49 ± 0.05 | 0.53 ± 0.03 | 0.53 ± 0.05  |
| LG           | 0.42 ± 0.06     | 0.49 ± 0.04     | 0.46 ± 0.03   | 0.43 ± 0.05 | 0.48 ± 0.03 | 0.48 ± 0.03  |
| SL           | 0.43 ± 0.15     | 0.33 ± 0.11     | 0.29 ± 0.09   | 0.44 ± 0.13 | 0.37 ± 0.10 | 0.32 ± 0.10  |
| CD           | 52.59 ± 24.46   | 54.08 ± 24.65   | 47.01 ± 25.91 | 53.75 ± 26.13 | 65.24 ± 28.48 | 45.69 ± 28.12 |
| GS           | 23.07 ± 13.02   | 19.14 ± 12.38   | 14.39 ± 11.38 | 24.43 ± 13.89 | 25.66 ± 15.02 | 15.76 ± 13.52 |

#: number of subjects; HB: Head-to-body ratio; HG: hand-to-ground distance; LG: lower limb length; SL: step length; CD: cadence (steps per minute); GS: gait speed (cycles per minute). Mean values are given ± standard deviation.

---

**Fig. 6.** GITW Subjects’ Average Features by Age Group. Children in the GITW dataset are distinguishable by their greater head-to-body ratios and have a smaller hand-to-ground distance than the other classes. The smaller hand-to-ground distance of children follows directly from the shortness of their lower limbs compared to adults and seniors. Seniors in the dataset have the slowest gait speed.

**Fig. 7.** Body graph construction in a single frame. Each keypoint is represented with a row in the body features matrix while the body structure is represented with an adjacency matrix.

---

the central node with $a$ and $c$ as its adjacent nodes. From $\vec{ba}$ and $\vec{bc}$, $\theta$ and $L$ are obtained as:

$$L = |\vec{ba}| + |\vec{bc}|$$

(5)

$$\theta = \cos^{-1}((\vec{ba} \cdot \vec{bc})/|\vec{ba}||\vec{bc}|)$$

(6)

Suppose that in frame $i$, the central node $b_i$ for a part is at coordinate $(x_{b_i}, y_{b_i})$ in the 2D plane, $D$ is obtained as the Euclidean distance between $b_i$ and $b_{i+1}$:

$$D(b_i, b_{i+1}) = [(x_{b_i} - x_{b_{i+1}})^2 + (y_{b_i} - y_{b_{i+1}})^2]^{1/2}$$

(7)

Each part has two extremes: head and mid-hip for the spine; left and right each for the arm, leg, and hip. Let $p_j$ and $q_j$ be the extremes for an arbitrary part in frame $i$. The swing $S_i$ for each frame part is defined as:

$$S_i(p_j, q_j) = [(x_{p_j} - x_{q_j})^2 + (y_{p_j} - y_{q_j})^2]^{1/2}$$

(8)
Hence, the part graph is defined as a spatiotemporal graph $G_{\text{part}} = \{V_{\text{part}}, E_{\text{part}}\}$, as described in Fig. 10. Each part graph has a node feature matrix $X_{\text{part}} \in \mathbb{R}^{\vert V_{\text{part}} \vert \times 4}$, where $\vert V_{\text{part}} \vert = 4F$ represents the part nodes across all video frames, each with four features described above. $E_{\text{part}}$ is the set of all spatial edges and temporal edges. Hence, with three spatial edges per frame and four part nodes, $\vert E_{\text{part}} \vert = 3 + 4(F - 2)$. Structurally, the part graphs are represented by a square adjacency matrix $A_{\text{part}} \in \mathbb{Z}^{\vert E_{\text{part}} \vert \times \vert E_{\text{part}} \vert}$.

5.3. Part-adaptive residual graph convolutional neural network

5.3.1. Preliminary

This section presents the Part-Adaptive Residual Graph Convolutional Neural Network (PairGCN) with adaptive edge weights. PairGCN takes part graphs (or body graphs) as input and learns discriminative features for children, adults, and seniors. PairGCN uses gait features only and includes an enhancement that learns edge weights for the part graph adjacency matrix in the first layer. The proposed PairGCN is based on Morris et al.’s GCN operator (Morris et al., 2019), given as:

$$x_j^{l+1} = \theta_1 x_j^l + \theta_2 \sum_{j \in \mathcal{N}(i)} w_{ij} \cdot x_j^l$$

(9)

where, $x_j^l$ denotes the embedding of node $i$ (target node) in layer $l$; $\mathcal{N}(i)$ is the set of neighbors of node $i$; $x_j^l$ denotes the embedding of node $j$ (source node) in layer $l$; $\theta_1$ and $\theta_2$ are learnable parameters; and $w_{ij}$ is the edge weight from source node $j$ to target node $i$.

In conventional GCN, $w_{ij}$ is often set to a constant value for all edges. This often suffices in domains where the main discriminative features of the graph are the node features and graph structure. Examples include action recognition and gait recognition (Li et al., 2020; Teepe et al., 2021). Although extracting age-associated features from a body skeleton graph also requires the body structure and joint positions represented as node features, the connectivity of the human body contains no age-associated information. The weights associated with bones and body parts and other dynamic information such as range of motion...
and speed must also be adequately modeled to extract age-associated features.

Therefore, the proposed PairGCN is modified from the GCN operator in Eq. (9) by introducing an adaptive edge weight \( \theta_w \in \mathbb{R}^{|E_{part}|} \), such that

\[
x^{l+1}_i = \theta_1 x^l_i + \theta_2 \sum_{j \in N(i)} \theta_{w_{i,j}} x^l_j
\]  

(10)

How PairGCN captures age-associated features based on body parts is discussed in Section 6.1.

5.3.2. Part-adaptive GCN

As illustrated in Fig. 11, the first layer of PairGCN performs the essential GCN operation according to Eq. (9). For the first layer, it makes use of a constant \( \theta_w \), initialized as 1 for each edge, and node embeddings are obtained for each graph as an aggregation of node features and their 1-hop neighborhood. Then, after the first layer, \( \theta_w \) is updated based on the node embeddings obtained in the first layer as

\[
\theta'_w = \text{MEANPOOL}(\text{GCN}(x^l)) \theta_w
\]  

(11)

where \( \text{MEANPOOL}(\cdot) \) denotes the graph-wise mean of the first layer node embeddings, \( \text{GCN}(\cdot) \) refers to the forward pass performed by GCN and \( x^l \) is the node embedding in layer \( l, l = 1, \ldots, L \). Since edges are formed with 1-hop neighbors, edge weights \( \theta_w \) are updated based on first layer node embeddings. Subsequent layers then make use of the updated edge weight in obtaining node embeddings by operating:

\[
x^{l+1}_i = \theta_1 x^l_i + \theta_2 \sum_{j \in N(i)} \theta'_{w_{i,j}} x^l_j
\]  

(12)

The overall architecture of PairGCN with \( N \) residual blocks is given in Fig. 12. With an added fully connected layer, PairGCN is trained with the guidance of age group labels.

5.3.3. Loss function

Data preparation steps such as augmentation and segmentation based on view introduce class imbalance in the training dataset. In PairGCN, two approaches are used to solve the class imbalance problem. Firstly, weighted random subsampling is used, such that under-represented classes are oversampled in each training batch. Secondly, PairGCN uses a class-balanced cross-entropy loss that weights the total loss based on each class’s effective number of samples. To get the weighted losses, class weights are obtained based on the effective number of samples (Cui et al., 2019) per class which is given as

\[
E_i = (1 - \beta^i)/(1 - \beta)
\]  

(13)
where $c$ = total number of samples of a class; $\beta = (N - 1)/N$; $N$ = total number of samples. Then, the class-balanced loss $L_{cb}$ is defined as:

$$L_{cb}(p, y) = 1/E_y(L(p, y))$$

where $y$ is the class label and $p$ is the softmax probability. For this use case, $L(p, y)$ is the cross-entropy loss function for $C$ classes given as:

$$L_i(p, y) = (1 - \beta)/(1 - \beta^c) \sum_{i=1}^{C} y_i \log p_i$$

### 5.3.4. Inference

During inference, PairGCN is used as a feature extractor for unlabeled body or part graphs as input, for which it outputs a feature vector that describes the age-associated gait features in the graph. Due to age’s ordinal and hierarchical nature, features obtained from PairGCN are best suited for hierarchical age group classification with tree-based classifiers such as Random Forest. However, any classifier could be fitted on the gait embedding obtained from PairGCN and used for final classification.

### 6. Experiments and results

In this paper, GITW is used to evaluate the proposed method since it contains samples for child, adult, and senior classes. The dataset is divided into two main views with respect to the camera, namely, the front view ($0^\circ$–$45^\circ$) and the side view ($45^\circ$–$90^\circ$). Training and inference were performed on a Windows 11 PC with 8 GB RAM, 2 GB GPU, and a core i7 Intel processor. During the training phase, each network is trained with the guidance of the age groups. In the inference stage, PairGCN is used as a feature extractor for unlabeled gait offers more age-associated features for adults and seniors, associated features, especially for children. In comparison, the front view of gait offers more age-associated features for adults and seniors, as shown by the higher F1-score for both classes in the ($45^\circ$–$90^\circ$) view. For all views, using Random Forest gives better results.

#### 6.1. An analysis of adaptive weight distributions and body parts

The spatiotemporal part graph obtained via parts aggregation (Fig. 9) consists of spatial edges $E_{spatial}$ and temporal edges $E_{temporal}$. The spatial edges are $spine$–$arm$, $spine$–$hip$, and $hip$–$leg$, while the temporal edges are $spine$–$spine$, $arm$–$arm$, $hip$–$hip$, and $leg$–$leg$. For each edge in the part graph, the edge formed by part $j$ and $i$ is weighted as $\theta^j_{wi} \in [0, 1]$. These weights are averaged for each age group (Fig. 13). The $spine$–$arm$ and $arm$–$arm$ connections are important in discriminating between children and other age groups. This is as expected since children are easily distinguishable by their arm swings. However, the arm movement of seniors is often more cautious since they are prone to falls. On the other hand, $spine$–$hip$ and $hip$–$leg$ connections are assigned greater weights in distinguishing seniors. This follows since seniors have a greater range of motion at the hips and lower extremity joints. The equally weighted $leg$–$leg$ connection indicates the universal importance of dynamic features such as gait speed, step length, and cadence, which depend on the temporal movement of the lower limbs.

#### 6.2. Ablation studies

This section demonstrates the ablation studies on the performance of PairGCN architecture, adaptive edge weights, body parts aggregation, coordinate smoothing, and view regarding age classification accuracy and F1 scores on the GITW dataset.

### A. PairGCN vs. Baseline GCN architecture

GCN architectures are often highly task-specific. For comparison, a baseline GCN architecture (Fig. 14) is constructed according to the basic architecture proposed by You et al. (2020). As shown in Table 8, PairGCN achieves the highest overall accuracy on the GITW dataset in the $45^\circ$–$90^\circ$ view. As expected, the Baseline GCN and PairGCN both learn more age-associated features in the side view of gait. This is expected since most age-associated features such as arm swing, posture,
Table 5
List of variables and hyperparameters used in experiments.

| Variable/Description | BaselineGCN | SequenceGCN | PairGCN |
|----------------------|-------------|-------------|---------|
| batch                | 64          | 64          | 64      |
| C                    | 3           | 3           | 3       |
| F                    | 50          | 50          | 50      |
| f_in                  | 3           | 3           | 4       |
| f_out                 | 64          | 64          | 64      |
| n                    | 15          | 15          | 4       |
| |E|_body                  | N/A         | N/A         | N/A     |
| |E|_part                  | N/A         | N/A         | 342     |
| N                    | 0.001       | 0.001       | 0.001   |
| Learning rate         | 200         | 200         | 200     |
| Non-linearity         | LeakyReLU(negative_slope = 0.01) |

B. Adaptive edge weighting

The performance of PairGCN with adaptive edge weights is compared to its performance with constant edge weights on the 45°–90° view of the GITW dataset. PairGCN without adaptive edge weights shows better F1 Scores for children than the older age groups in Table 9. This suggests that it becomes increasingly difficult to discriminate against older age groups without adaptive edge weighting.

C. Part aggregation vs. Raw coordinates

The performance of part aggregation is compared to raw coordinates for age group classification, using the Baseline GCN and PairGCN on the 45°–90° view of the GITW dataset. The two methods are compared in model performance (Tables 8 and 9). Part aggregation improves the performance of both architectures.

D. Coordinate smoothing

As shown in Table 9, the F1 scores suggest that without smoothing, PairGCN finds it increasingly difficult to learn features associated with age groups as the age increases. While it is straightforward to learn features associated with children (such as their body size) even with noisy coordinates, the features associated with adults and seniors are more subtle and sensitive to noise.

6.3. The importance of temporal gait information

In sequential data and time-series analysis, recurrent neural networks (RNNs) such as Long-Short Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997) have achieved outstanding results in sequential classification and prediction. While it may seem more natural to model gait as a sequence and adopt RNNs, this section presents a comparative analysis between spatiotemporal and sequence-based gait for age group classification. A Sequence GCN with a Graph Gated Recurrent Unit (GraphGRU) architecture is constructed for this analysis, as illustrated in Fig. 15.
Table 7
Summary of PairGCN results on GITW dataset.

| View           | Classifier      | F1 scores | Accuracy |
|----------------|-----------------|-----------|----------|
|                |                 | child     | adult    | senior   |       |
| Front view (0°–45°) | PairGCN FC layer | 0.95      | 0.83     | 0.88     | 88%   |
|                | Random Forest   | 0.95      | 0.96     | 0.97     | 96%   |
| Side view (45°–90°) | PairGCN FC layer | 0.95      | 0.90     | 0.82     | 90%   |
|                | Random Forest   | 0.99      | 0.99     | 0.99     | 99%   |

FC layer: fully connected layer.

is represented as a sequence of body graphs, \( G_{seq} = \{G_b, \ldots, G_b\} \). SequenceGCN is trained using \( G_{seq} \) and \( A_b \) with the guidance of age group labels.

The results obtained by SequenceGCN (Table 10) suggest that the network cannot learn age-associated features from gait modeled as a sequence. Several reasons could be given for this observation. First, although gait is observed from the walking sequence of subjects, gait is more of a cycle than a sequence. However, sequence-based methods such as SequenceGCN will consider each frame in the context of previous frames while failing to consider the temporal connections

Fig. 14. Baseline GCN Architecture. During training, Baseline GCN is trained using age group labels. During inference, it is used as a feature extractor.
between the frames. Hence, modeling gait as a spatiotemporal graph is more suitable for age group classification. In contrast, modeling gait as a sequence of body graphs is perhaps more suited to naturally sequence-based tasks such as human action recognition.

6.4. Comparison with existing methods

PairGCN is proposed to extract age-associated model-based gait features for age group classification. However, at the time of this writing, existing model-based gait techniques for age group classification make use of handcrafted features and perform classification with conventional classifiers such as Random Forest (Aderinola et al., 2021a; Hema et al., 2019), Support Vector Machines (Chuen et al., 2015; Begg et al., 2005; Hema and Pitta, 2019; Yoo and Kwon, 2017), and K-Nearest Neighbor (Yang and Wang, 2015; Hediyeh et al., 2013).

This section compares features extracted using PairGCN with the performance of handcrafted features, using SVM, KNN, and RF as classifiers. The best-performing settings of the classifiers are chosen for comparison. The results in Table 11 show that compared to manually extracted features, PairGCN features contain richer age-associated features. It is also observed that the accuracy can be further improved with classifiers such as SVM, KNN, or Random Forests.

6.5. Discussion

We have proposed and evaluated PairGCN via several experiments on the front and side views of the GITW dataset. We showed that using the side view of gait offers 3% more accuracy because most of the dynamic features of gait can be observed on the side view (Table 7). A comparison of PairGCN to the BaselineGCN also shows that PairGCN’s adaptive weight learning reduces the error rate by up to 5%. Moreover, the proposed part aggregation scheme enables faster and better performance for both architectures even without adaptive edge weights.

Body coordinate features are noisy due to false detections and multiple detections of joints during pose estimation. Hence, we performed experiments to observe the effects of smoothing. The experiments suggest that the age-associated features of the elderly are subtle and more sensitive to noise than children, as children can easily be detected by their smaller body sizes and larger head-to-body ratios (Table 9).

Since gait is measured from the walking sequence of humans, it may seem more natural to use sequence-based architectures for gait age prediction. Hence, we examined this experimentally with a SequenceGCN, which was unable to learn age-associated features from gait (Table 10). This suggests that it is best to model gait as a large spatiotemporal graph to capture the inter-frame temporal information, which is crucial in age discrimination. This finding also corroborates this observation that several spatiotemporal gait features are measured from the inter-frame connections of joints.

7. Conclusion and future research

In this paper, we have proposed a deep learning approach to extract age-associated features from model-based gait for age group classification, which is the first of its kind. In particular, we proposed a Part-Adaptive Residual GCN (PairGCN) for automatic extraction of age-associated gait features in video sequences. PairGCN outperformed manual feature selection methods and a Gated Recurrent Unit architecture on a dataset collected in unconstrained environments, achieving up to 99% accuracy in classifying subjects as a child, adult, or senior. These results show the feasibility of age classification in live surveillance. To aid further related research, the dataset collected for gait-based age and gender classification in unconstrained environments is being made publicly available.

There are several limitations in this study. Firstly, the proposed model is sensitive to view change. Though it achieves high accuracy on the side view of gait, the performance degrades on the front view. However, the viewpoint challenge is common to all gait-related tasks. Domain adaptation has been applied in tasks like character recognition, enabling a model to perform comparably well on a dataset slightly different from the dataset it was trained on. Future research can achieve view-invariant gait-based age group classification by using domain adaptation.

In addition, PairGCN validation losses are relatively high (around 20%) due to the noisy nature of coordinates and the variability of age-associated gait features. For example, a child might have a gait like an...
adult and vice-versa. PairGCN is also prone to overfitting on the gait graphs when 200 epochs are exceeded. Hence, features extracted using PairGCN are best used with classifiers more robust to the irregularities in age-associated features.

To address the limitations above, future research will focus on collecting more data to enable more fine-grained age prediction while exploring domain adaption for view-invariant age group classification.

### CRediT authorship contribution statement

**Timilehin B. Aderinola:** Formal analysis, Investigation, Writing – original draft. **Tee Connie:** Conceptualization, Data curation, Supervision, Writing – review & editing. **Thian Song Ong:** Data curation, Supervision, Writing – review & editing. **Andrew Beng Jin Teoh:** Resources, Writing – review & editing. **Michael Kah Ong Goh:** Validation, Writing – review & editing.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

The data is available at [https://github.com/timiderinola/mmu_gag](https://github.com/timiderinola/mmu_gag).

### Acknowledgments

This work was supported in part by research grants from Fundamental Research Grant Scheme (FRGS) (FRGS/1/2020/ICT02/MMU/02/5) and Multimedia University IR Fund (MMUI/220026).

### References

Aderinola, T.B., Connie, T., Ong, T.S., Goh, K.O.M., 2021a. Automatic extraction of spatio-temporal gait features for age group classification. In: Garg, L., Sharma, H., Goyal, S.B., Singh, A. (Eds.), Proceedings of International Conference on Information and Communication Technologies. Springer, pp. 71–78. 

Aderinola, T.B., Connie, T., Ong, T.S., Yau, W.C., Teoh, A.B.J., 2021b. Learning age from gait: A survey. IEEE Access 9, 100352–100368. [http://dx.doi.org/10.1109/access.2021.3095477](http://dx.doi.org/10.1109/access.2021.3095477).

Akbari, A., Awaiss, M., Fatemifar, S., Khalid, S.S., Kittler, J., 2021. A novel ground metric for optimal transport-based chronological age estimation. IEEE Trans. Cybern. 52 (10), 9966–9999.

Begg, R., Palaniswami, M., Owen, B., 2005. Support vector machines for automated gait classification. IEEE Trans. Biomed. Eng. 52 (5), 828–838. [http://dx.doi.org/10.1109/TBME.2005.845241](http://dx.doi.org/10.1109/TBME.2005.845241).

Berksan, M., 2019. Gender Recognition and Age Estimation Based on Human Gait (Ph.D. thesis). Bilkent Üniversitesi Fen Bilimleri Enstitüsü.

Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P., 2017. Geometric deep learning: Going beyond euclidean data. IEEE Signal Process. Mag. 34 (4), 18–42. [http://dx.doi.org/10.1109/MSP.2017.2693418](http://dx.doi.org/10.1109/MSP.2017.2693418).

Burt, C., 2020. Biometrics could help gaming industry with cybercrime, age verification as market evolves | biometric update.

Cao, W., Mirjalili, V., Raschka, S., 2020. Rank consistent ordinal regression for neural networks with application to age estimation. Pattern Recognit. Lett. 140, 325–331.

Chuen, B.K.Y., Connie, T., Song, O.T., Goh, M., 2015. A preliminary study of gait-based age estimation techniques. In: 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA). pp. 800–806. [http://dx.doi.org/10.1109/APSIPA.2015.7415382](http://dx.doi.org/10.1109/APSIPA.2015.7415382).

Cui, Y., Jia, M., Lin, T.Y., Song, Y., Belongie, S., 2019. Class-balanced loss based on effective number of samples. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [http://dx.doi.org/10.1109/cvpr.2019.00941](http://dx.doi.org/10.1109/cvpr.2019.00941).

Dagheh, I., Barbara, D., 2021. Facial age estimation using pre-trained CNN and transfer learning. Multimedia Tools Appl. 80, 20369–20380.

Fang, H.-S., Li, J., Tang, H., Xu, C., Zhu, H., Xiu, Y., Li, Y.-L., Lu, C., 2022. Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. IEEE Trans. Pattern Anal. Mach. Intell.

Han, S., 2020. Age estimation from face images based on deep learning. In: 2020 International Conference on Computing and Data Science. CDS, IEEE, pp. 288–292. [http://dx.doi.org/10.1109/TPAMI.2006.38](http://dx.doi.org/10.1109/TPAMI.2006.38).

Hediyeh, H., Sayed, T., Zaki, M.H., 2013. Use of spatiotemporal parameters of gait for automated classification of pedestrian gender and age. Transp. Res. Rec. 2393 (1), 31–40. [http://dx.doi.org/10.3141/2393-04](http://dx.doi.org/10.3141/2393-04).

Hema, M., Babulak, K., Balaji, N., 2019. Gait based human age classification using random forest classifier. i-manager’s J. Pattern Recognit. 6 (2), 1–7.

Hema, M., Pitta, S., 2019. Human age classification based on gait parameters using a gait energy image projection model. In: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). pp. 1163–1168. [http://dx.doi.org/10.1109/ICOEI.2019.8862788](http://dx.doi.org/10.1109/ICOEI.2019.8862788).

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Islam, T.U., Awasthi, L.K., Garg, U., 2021. Gender and age estimation from gait: A review. In: International Conference on Innovative Computing and Communications: Proceedings of ICICC 2020, Volume 2. Springer, pp. 947–962.

Kitchat, K., Limcharoen, P., Khamsemanan, N., Nattee, C., 2022. Human age estimation from multi-angle gait silhouettes with convolutional neural networks. Thai J. Math. 18 (3), 1227–1238.

Kwanny, D., Hemmerling, D., 2021. Gender and age estimation methods based on speech using deep neural networks. Sensors 21 (14). [http://dx.doi.org/10.3390/s21144785](http://dx.doi.org/10.3390/s21144785).

Lau, L., Chan, K., 2023. Tree structure convolutional neural networks for gait-based gender and age classification. Multimedia Tools Appl. 82 (2), 2145–2164.

Li, N., Zhao, X., Ma, C., 2020. JointAgeItA model-based gait recognition method based on gait graph convolutional networks and joints relationship pyramid mapping. arXiv:2005.08625, [cs, eess].

Liao, R., Gao, C., Garcia, E.B., Yu, S., Huang, Y., 2017. Pose-based temporal-spatial network (PSFN) for gait recognition with carrying and clothing variations. In: Zhou, J., Wang, Y., Sun, Z., Xu, Y., Shen, L., Feng, J., Shan, S., Qiao, Y., Guo, Z., Yu, S. (Eds.), Biometric Recognition. Springer International Publishing, pp. 474–483. [http://dx.doi.org/10.1007/978-3-319-69923-3_51](http://dx.doi.org/10.1007/978-3-319-69923-3_51).

Liao, H., Yan, Y., Dai, W., Pan, F., 2018. Age estimation of face images based on CNN and divide-and-rule strategy. Math. Probl. Eng..

Lau, L., Chan, K., 2023. Tree structure convolutional neural networks for gait-based gender and age classification. Multimedia Tools Appl. 82 (2), 2145–2164.

Li, N., Zhao, X., Ma, C., 2020. JointAgeItA model-based gait recognition method based on gait graph convolutional networks and joints relationship pyramid mapping. arXiv:2005.08625, [cs, eess].

Liao, R., Cao, C., Garcia, E.B., Yu, S., Huang, Y., 2017. Pose-based temporal-spatial network (PSFN) for gait recognition with carrying and clothing variations. In: Zhou, J., Wang, Y., Sun, Z., Xu, Y., Shen, L., Feng, J., Shan, S., Qiao, Y., Guo, Z., Yu, S. (Eds.), Biometric Recognition. Springer International Publishing, pp. 474–483. [http://dx.doi.org/10.1007/978-3-319-69923-3_51](http://dx.doi.org/10.1007/978-3-319-69923-3_51).

Liao, H., Yan, Y., Dai, W., Pan, F., 2018. Age estimation of face images based on CNN and divide-and-rule strategy. Math. Probl. Eng.,
