Developing of a multi-angle and varied walking conditions dataset for human gait recognition

Jose Misael Burruel-Zazueta [a], Hector Rodriguez-Rangel [a], Vicenç Puig Cayuela [b], Manuel Alejandro Medrano-Diaz [a], Sofia Isabel Fernandez-Gregorio [c], Luis Alberto Morales-Rosales [d,*]

[a] *Division de Estudios de Posgrado e Investigacion, Instituto Tecnologico de Culiacan, Juan de Dios Batiz 310 pte, Culiacan, 80220, Sinaloa, Mexico*
[b] *Institut de Robótica i Informática Industrial (CSIC-UPC), Carrer LLorens Artigas 4-6, Barcelona, 08028, Spain*
[c] *Facultad de Estadística e Informática, Universidad Veracruzana, Av. Xalapa, Obrero Campesina, Xalapa-Enriquez, 91020, Veracruz, Mexico*
[d] *Facultad de ingenieria civil, CONAHCYT-Universidad Michoacana de San Nicolás de Hidalgo, Avenida Francisco J. Múgica S/N Ciudad Universitaria, Morelia, 58060, Michoacan, Mexico*

ARTICLE INFO

Keywords:
Deep learning
Self-supervised vision transformer
Human gait recognition
Walking conditions dataset

ABSTRACT

The field of Human Gait Recognition (HGR) leverages unique walking patterns for non-invasive, discreet biometric identification. This study highlights the importance of comprehensive datasets in the development, testing, and validation of HGR algorithms. While current datasets, such as those from Carnegie Mellon University Motion of Body (CMU MoBo), Southampton (SOTON), Chinese Academic of Science Institute of Automation (CASIA B), and Osaka University Institute of Scientific Industrial Research (OU-ISIR), have advanced in scale and complexity, they often lack diversity and comprehensive sample representation. To address this gap, we introduce the TecNM Gait-DS dataset (Tecnologico Nacional de Mexico), specifically designed for Latin American populations, featuring 13 viewing angles and five walking variations. Utilizing a Self-Supervised Vision Transformer DINO (Deeper Into Neural Networks) model for view angle classification, our evaluation demonstrates significant improvements in classification accuracy. This dataset not only enhances sample diversity but also supports the development of more robust HGR systems. Our results underscore the potential for improved accuracy and ethical considerations in HGR, advocating for ongoing refinement of datasets to achieve optimal performance and societal acceptance.

1. Introduction

Human Gait Recognition (HGR) is a biometric technology that identifies individuals based on their unique walking patterns or gait (Gafurov, 2007). HGR, a burgeoning field in biometrics, has garnered substantial attention due to its non-invasive nature and wide-ranging applications in identity verification, surveillance, and forensic analysis (Wang and Yan, 2020; Al Kork et al., 2017). The cornerstone of successful HGR systems lies in the availability of comprehensive and representative datasets that facilitate algorithm development, testing, and validation.

The main advantage is the difficulty of spoofing due to its roots in each person's unique physical structure and biomechanics (Hadid et al., 2012). In addition, its less intrusive nature and independence from external conditions make it applicable to diverse situations (Rani and Kumar, 2023). However, HGR also has disadvantages, including susceptibility to a person's gait variations due to factors such as injuries, aging, mood, clothing, or objects carried at the time (Rani and Kumar, 2023; Jawed et al., 2018).

The most widely used techniques in the field of HGR are based on deep learning. This is because deep learning has generated a large number of solutions to classification problems by increasing the accuracy (Seng et al., 2024).

Researchers in the field of HGR have classified this recognition technique into two distinct approaches (image-based). The model-based approach involves extracting specific behavioral features from the subject, such as limb flexion angles, torso and head movements, and combinations of these features (Bouchrika, 2018). On the other hand, the appearance-based approach captures complete subject silhouettes and relies on various image processing and recognition strategies based on these silhouettes. Among these strategies, the most common in the current state of the art is recognition through GEIs, which consolidate all silhouettes into a single image for feature extraction and classification (Bouchrika, 2018; Han and Bhanu, 2006).

The quality and diversity of datasets are crucial for HGR algorithms (Gong et al., 2023; Mazurek and Wielgosz, 2023). Effective

* Corresponding author.
E-mail address: lamorales@conacyt.mx (L.A. Morales Rosales).

datasets include varied gait patterns that reflect real-world complexities, thereby aiding the development of robust algorithms. However, creating such datasets poses challenges, including gathering diverse subjects, capturing data under varied conditions, and addressing ethical considerations.

Significant datasets like SOTON (Shutler et al., 2004), CASIA B (Yu et al., 2006), and OU-ISIR (Iwama et al., 2012; Takemura et al., 2018) have advanced the field by providing extensive video collections and silhouette sequences. Yet, these datasets often lack model-based extracted features such as key points or body part angles. There is also a need for datasets that represent a wider range of ethnicities and walking variants to improve the generalization of HGR systems.

TecNM Gait-DS is introduced as a novel dataset for HGR. It provides an alternative to the existing datasets mentioned above for training, evaluating, and validating HGR systems based on indoor scenarios. It considers subjects with Latin traits and introduces two walking variants that were not previously considered in the generalization of the systems in conjunction with other datasets. The TecNM Gait-DS contributions are listed as follows:

- It includes 124 Mexican participants and offers 13 viewing angles, simulating various surveillance scenarios. TecNM Gait-DS stands out with its five distinct walking variations. We introduce complexity and realism to gait recognition experiments by including two new different scenarios: carrying a backpack and a box compared with the state-of-the-art datasets on HGR. TecNM Gait-DS dataset comprises a total of 22,568 videos, making it a valuable resource for algorithm development and evaluation.
- TecNM Gait-DS provides a rich set of features extracted from the videos, including skeletons tracked through 15 key points, 19 angles of body part inclination and flexion, silhouettes extracted frame by frame, and GEIs derived from silhouettes. These features are meticulously extracted using state-of-the-art tools such as Densepose (Güler et al., 2018) for silhouettes and Detectron2 (Wu et al., 2019) for key points, ensuring precision and accuracy in feature representation.
- An analysis of how camera angles and extracted characteristics of the scenarios included in TecNM Gait-DS have improved the performance of person identification by using Transformer DINO VIT. The comprehensive evaluation of the Transformer DINO VIT (Caron et al., 2021) model was conducted with five sets of energy images corresponding to different walking variations. Each evaluation used two images per subject for a total of 3224 images per walking variation. This approach allowed a detailed understanding of the impact of variability in walking variation on the classification of viewing angles. The obtained results comprehensively evaluate the extracted features and the trained model under various conditions, providing detailed insight into its performance and robustness.
- Finally, we provide a Github repository[1] where all the information related to this dataset, a format for requesting videos, silhouettes, energy images, silhouettes, algorithms used for feature extraction, energy image generation, data cleaning, and experiments can be found.

2. Related work

In the landscape of contemporary research, the significance of datasets as foundational pillars of knowledge acquisition and innovation cannot be overstated. Datasets, serving as repositories of structured and organized information, underpin many scientific and practical endeavors across diverse domains. These digital reservoirs of data facilitate the retrieval, analysis, and interpretation of information, driving discoveries and enabling the development of cutting-edge methodologies.

Gait recognition research hinges on the availability of high-quality datasets that accurately capture the intricacies of human walking patterns. These datasets play a pivotal role in developing and evaluating gait recognition algorithms, allowing researchers to explore diverse scenarios and challenges. In this section, we delve into some of the prominent gait recognition datasets that have significantly contributed to the advancement of the field.

Within this section, a comprehensive overview unfolds, featuring five distinct gait datasets, each contributing to the evolving landscape of gait recognition research. The chronological journey begins with the CMU Motion of Body (MoBo) dataset, a pioneering resource that laid the foundation for gait biometrics. Subsequently, the SOTON Gait dataset, followed by the CASIA B Gait dataset, is explored, each offering unique insights and challenges in gait analysis. Then, the OU ISIR Gait dataset and the OU ISIR Multi-View Large Population (MVLP) dataset are examined, shedding light on their contributions to the field and setting the stage for an in-depth analysis of gait-based biometric methodologies. Finally, two current databases developed in a wild environment called GREW and Gait3D, the proposed benchmarks are needed to train and evaluate the gait recognizer in the wild. Table 1 summarizes the state of art comparison; in the following subsection, we describe each dataset in a detailed manner and present their discussion.

2.1. CMU Motion of Body (MoBo)

The CMU Motion of Body (MoBo) dataset (Gross, 2001) was meticulously curated to address the evolving landscape of biometric identification. Unlike conventional biometric modalities, such as fingerprint or facial recognition, the dataset zeroes in on the distinct and nuanced patterns associated with human gait. Consequently, it offers a unique platform for investigating the feasibility of using gait as a biometric identifier.

The dataset comprises 25 subjects who participated in four distinct walking activities on a treadmill. To capture comprehensive insights into the nuances of human gait, data from six strategically positioned cameras were synchronized, ensuring a multifaceted view of the subjects' movements. This approach enables the examination of gait from various angles and perspectives, reflecting real-world scenarios.

Noteworthy data characteristics include an 11-second sequence length for each activity, an image resolution with a pixel height of 500, and a frame rate of 30 frames per second. These specifications are pivotal in enabling fine-grained analysis of gait patterns.

2.2. SOTON

The authors (Shutler et al., 2004) address the importance of having a comprehensive and diverse dataset for gait analysis and recognition research. They proposed creating a large-scale dataset encompassing a wide range of gait sequences, providing researchers with a valuable resource for algorithm development and evaluation.

While specific details about the dataset are not mentioned in the provided information, it can be inferred that their dataset was designed to contain extensive sequences of human gait data. These sequences capture the walking patterns of individuals from different perspectives.

The motivation behind developing this large sequence-based human gait dataset was to facilitate advancements in gait analysis and recognition algorithms. By providing researchers access to a substantial and diverse collection of gait sequences, the authors aimed to stimulate research in this field and foster the development of more accurate and robust gait recognition techniques.

The paper explains the process of creating the dataset, which involved recruiting 115 participants, data collection using specialized cameras, and data preprocessing steps. The authors also outline the

[1] https://github.com/MisaelZazueta/TecNM_Gait-DS.

Table 1
Summary of work related to the generation of databases for the HGR.

Name	# Subjects	# Videos per subject	# Total videos	Environment	Views (angles)	Walking variations
CMU MoBo	25	24	600	Indoor	6	Three walking speeds, carrying a ball (all on a treadmill)
SOTON	115	–	2128	Indoor & Outdoor	2	Normal walking on a treadmill
CASIA B	124	110	13 640	Indoor	11	Normal walking, carrying a bag, wearing a coat
OU-ISIR	4007	–	31 368	Outdoor	4	Normal walking
OU-ISIR-MVLP	10 307	–	259 013	Indoor	14	Normal walking
GREW	26 345	–	128 671	Wild	–	Undefined
Gait3D	4 000	–	25 309	Wild	–	Undefined

protocols and guidelines to ensure consistency and quality across the dataset. The authors recorded 2128 video sequences with two cameras (frontal and oblique), indoors and outdoors.

Additionally, the authors presented some experimental results or findings derived from the dataset to demonstrate its potential applications.

2.3. CASIA-B

The CASIA-B dataset (Yu et al., 2006) for gait recognition was developed by the Chinese Academy of Sciences' Institute of Automation (CASIA). The dataset was specifically designed to advance research in gait recognition algorithms and their applications.

The creation of the CASIA-B dataset involved recording gait data from a group of individuals under controlled conditions. The dataset creators recruited 124 subjects to participate in the data collection process. The subjects were carefully selected to represent variations in age, gender, and body characteristics, ensuring a diverse range of gait patterns.

The gait data collection process occurred on a specialized walkway with multiple synchronized cameras. The cameras were positioned at different angles to capture the subjects' gait from various viewpoints. The 11 angles considered for the recordings were 0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 172° and 180°. Each subject performed multiple walking trials, with each trial capturing a sequence of consecutive video frames showcasing their gait.

The dataset creators provided specific instructions to the subjects regarding walking conditions to ensure consistency and accuracy. This included walking at a natural pace, wearing particular types of clothing, and carrying a bag. By introducing these variations, the dataset aimed to simulate real-world conditions and assess the robustness of gait recognition algorithms. They recorded six video batches for normal walking and two for wearing a coat and carrying a bag. Each subject registered 110 videos, so there are 13 640 videos in this dataset.

The recorded gait data was then carefully processed and annotated to extract relevant information for analysis. Preprocessing steps typically involved segmenting the gait cycles from the video sequences and aligning them to a common temporal reference. This alignment helped to normalize the walking speed and facilitate accurate comparisons between different gait patterns.

The CASIA-B dataset offers a substantial collection of gait sequences for research purposes. Each subject contributes multiple gait cycles, resulting in a comprehensive dataset with a significant number of gait samples. The dataset provides a diverse representation of human walking patterns, considering variations in age, gender, and body characteristics.

In addition to the well-known CASIA B, there are also the CASIA-A (Wang et al.) and CASIA-C (Tan et al.) datasets. Dataset A (formerly NLPR Gait Database), created on December 10, 2001, includes 20 individuals, each with 12 image sequences in three directions (parallel, 45 degrees, and 90 degrees to the image plane), totaling 19,139 images and occupying 2.2 GB. The sequences vary in length, ranging from 37 to 127 images. On the other hand, Dataset C, collected between July and August 2005 using an infrared camera, covers 153 subjects and contemplates four walking conditions: normal walking, slow walking, fast walking, and normal walking with a bag. This dataset is characterized by being captured at night, which adds a unique dimension to the analysis of walking under low light conditions.

Researchers and practitioners have utilized the CASIA-B dataset for various purposes, including developing and evaluating gait recognition algorithms. By training and testing algorithms on this dataset, researchers can assess the performance of their models in recognizing individuals based on their unique walking patterns. The dataset's diversity and size enable comprehensive analysis and evaluation of gait recognition systems.

In summary, the CASIA-B dataset for gait recognition was created by recording and annotating gait data from 124 subjects. The dataset offers a diverse range of gait patterns, capturing variations in age, gender, and body characteristics. Researchers have used this dataset to advance the development and evaluation of gait recognition algorithms, contributing to the progress of gait recognition technology.

2.4. OU-ISIR

The dataset aims to address the need for a large-scale and diverse collection of gait data for evaluating gait recognition algorithms. It consists of a significant number of gait sequences captured from a large population of individuals.

The work of Iwama et al. (2012) provides insights into the creation and characteristics of the OU-ISIR gait dataset. It describes the data collection process, which involved recording gait sequences using specialized cameras or sensors. The dataset encompasses gait data from a diverse range of individuals, including variations in age, gender, and other demographic factors.

The dataset includes multiple gait cycles for each subject, enabling a comprehensive analysis of individual walking patterns. It may also contain annotations or metadata associated with each gait sequence, such as age, gender, and other relevant information. This dataset contains 31 368 video sequences obtained from 4007 different participants. The authors set up four cameras for recording, with an inclination of 55°, 65°, 75° and 85°. All videos were recorded outdoors and in normal walking conditions.

Furthermore, the OU-ISIR evaluation was focused on training and testing gait recognition models on the dataset and assessing their performance in accurately recognizing individuals based on their gait patterns.

OU-ISIR gait dataset provides a large-scale collection of gait sequences from a diverse population. It serves as a valuable resource for researchers working in the field of gait recognition, enabling algorithm development, performance evaluation, and further advancements in this domain.

2.5. OU-ISIR MVLP

The primary focus of Takemura et al. (2018) is to address the challenges of gait recognition across different camera views. The authors aim to provide a comprehensive dataset that enables the evaluation and development of gait recognition algorithms capable of handling variations in camera viewpoints.

The OU-ISIR MVLP dataset introduced by Takemura et al. (2018) is a large-scale collection of gait sequences captured from multiple camera views. It includes gait data from a significant population representing a diverse range of age groups and demographics.

The authors described the data collection process, involving capturing gait sequences using multiple synchronized cameras. These cameras were positioned to cover different viewpoints and angles, simulating real-world scenarios where individuals were captured from varying perspectives. This dataset is the largest ever recorded, with 259 013 video sequences from 10 307 participants. Although they only used a normal walk, the participants were recorded at 14 angles.

The dataset was created to facilitate the evaluation of gait recognition algorithms under cross-view scenarios. It enables researchers to train and test their algorithms on gait data captured from different camera views, assessing their ability to recognize individuals even when the viewpoints vary accurately.

The authors presented performance evaluation results, showcasing the effectiveness of gait recognition algorithms on the multi-view large population gait dataset. They discuss the recognition accuracy, robustness to viewpoint variations, and potential challenges encountered in cross-view gait recognition tasks.

Overall, they introduced a multi-view large population gait dataset and emphasized the importance of evaluating gait recognition algorithms in cross-view scenarios. Their work provided a valuable resource to researchers for developing and testing algorithms that can handle variations in camera viewpoints, advancing the field of gait recognition and its application in real-world scenarios.

2.6. GREW

In recent advancements within the field of gait recognition, the GREW (Gait REcognition in the Wild) dataset stands out as a significant contribution (Zhu et al., 2021), addressing the limitations of existing datasets that are predominantly captured in controlled environments. Introduced by Zhang et al. GREW is constructed from natural videos and encompasses a vast collection of 26,000 identities and 128,000 sequences, along with a distractor set of over 233,000 sequences. This dataset offers diverse and practical view variations, making it exceptionally suited for real-world applications of gait recognition. The GREW benchmark is notable for its comprehensive manual annotations and the inclusion of various challenging factors found in natural settings. The study by Zhang et al. demonstrates that the GREW dataset is indispensable for training and evaluating high-performance gait recognition systems in the wild, highlighting the potential for significant improvements in current state-of-the-art methods. Additionally, GREW can be an effective pretraining resource for controlled gait recognition tasks, underscoring its utility and relevance in advancing the field.

2.7. Gait3D

Recent advancements in gait recognition have traditionally focused on 2D representations, such as silhouettes or skeletons, often constrained to controlled environments. Addressing the inherent limitations of 2D projections, which lose critical information regarding viewpoint, shape, and dynamics, a novel approach has been proposed by Zheng et al. in 2022. This paper introduces SMPLGait, a framework leveraging dense 3D representations through the 3D Skinned Multi-Person Linear (SMPL) model for gait recognition in the wild. SMPLGait integrates two meticulously designed branches: one for extracting appearance features from silhouettes and another for learning 3D viewpoints and shapes from the SMPL model. To support this framework, the authors present Gait3D, the first large-scale 3D representation-based gait recognition dataset, comprising 4000 subjects and over 25,000 sequences captured by 39 cameras in an unconstrained indoor environment. Gait3D provides dense 3D information on body shape, viewpoint, and dynamics by recovering 3D SMPL models from video frames. Comprehensive comparisons with existing methods underscore the superior performance of SMPLGait and highlight the significant potential of 3D representations in advancing gait recognition technology for real-world applications.

In summary, the generation of gait recognition datasets has evolved significantly, covering a wide spectrum of environments and capturing conditions. From controlled indoor settings, such as in the CASIA B and CMU MoBo datasets, to more varied and expansive environments, like the outdoor settings of the OU-ISIR dataset and the real-world scenarios of the GREW and Gait3D datasets, each dataset contributes unique attributes and challenges to the field of HGR. However, our work specifically addresses the need for a controlled dataset with well-defined angles, a fixed quantity of videos per subject, and explicit subject authorization. This approach ensures consistency and reliability in data collection, facilitating the development of robust and reproducible gait recognition algorithms. By maintaining stringent controls over the experimental setup, our dataset aims to bridge the gap between controlled and uncontrolled environments, providing a valuable resource for advancing both theoretical research and practical applications in gait recognition.

2.8. Related work discussion

To address the research gaps in the related work, it is essential to highlight how the TecNM Gait-DS dataset advances the field of gait recognition, especially compared to existing datasets. One significant gap in previous research is the limited representation of real-world complexity in the available datasets. Many well-established datasets, such as CASIA B, SOTON, and OU-ISIR, predominantly focus on controlled indoor or outdoor environments with basic walking conditions. While they capture multiple angles and some walking variations, they often lack the inclusion of practical scenarios that mirror real-life situations, such as carrying objects. The TecNM Gait-DS addresses this gap by introducing two new scenarios: walking while carrying a backpack and carrying a box. This enhancement provides a more realistic challenge for gait recognition algorithms, simulating actual surveillance environments, which prior datasets did not fully address.

Current datasets often do not comprehensively integrate or leverage state-of-the-art feature extraction tools. TecNM Gait-DS uses advanced tools like DensePose and Detectron2 to extract key features such as skeletons, angles of body part inclinations, and silhouettes. This meticulous feature extraction provides high precision and enables more robust algorithm evaluation, which is not typically seen in other gait datasets.

Many datasets require a formal process to access, and few provide open-source tools for replicating experiments or extracting features. TecNM Gait-DS overcomes this limitation by offering an openly accessible GitHub repository containing all necessary data, algorithms, and tools, enabling broader research and encouraging reproducibility, which is often a challenge in current gait recognition research.

Several studies have contributed significantly to the field of gait recognition, leveraging both model-based and appearance-based approaches, along with state-of-the-art datasets. For instance, the CASIA-B and OU-ISIR datasets have been widely used in appearance-based gait recognition approaches such as, Liao et al. (2020), Lin et al. (2020) and Hou et al. (2020), providing large-scale, multi-view video sequences to improve recognition accuracy under various viewing angles and occlusions. Model-based approaches, such as those using 3D skeleton models, have been effectively applied to the CASIA-B and OU-ISIR

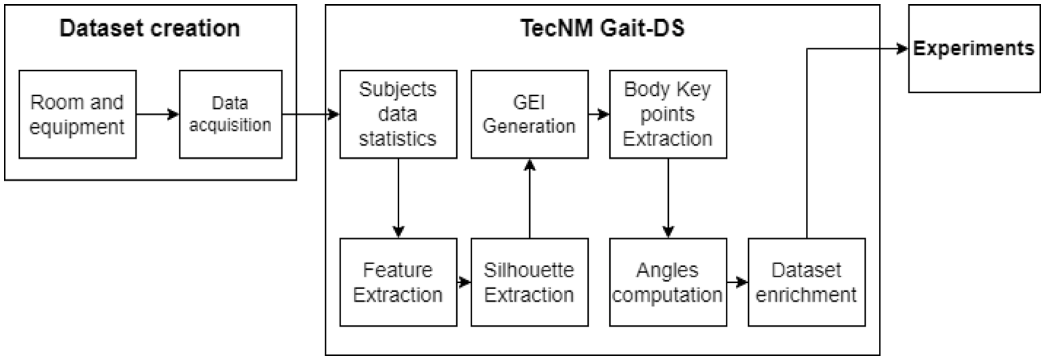

Fig. 1. Workflow methodology.

MVLP datasets, which focus on extracting structural information about human gait, e.g., An et al. (2020), Li et al. (2020a) and Jun et al. (2020). Additionally, several hybrid approaches combining both models and appearance features, such as Li et al. (2020b) and Zhang et al. (2022, 2019), have been implemented using the CASIA-B dataset, improving recognition in environments with background noise or low resolution.

3. Methodology

In this section, we describe the systematic approach to develop and validate the TecNM Gait-DS dataset along with the experiments conducted. The methodology encompasses three primary phases: (1) the creation of the dataset, (2) a detailed description of the TecNM Gait-DS, and (3) the experimentation process using the collected data. The workflow, shown in Fig. 1, begins with dataset creation, which includes recording video sequences of gait from different subjects and processing these videos to extract meaningful information such as silhouettes, energy images, and key points for gait recognition. This process ensures that the data capture a wide range of real-world walking conditions and variations in subjects' movements. The resulting dataset is then used to train and evaluate human gait recognition (HGR) models, ensuring robustness across diverse scenarios.

Next, the TecNM Gait Dataset subsection explains the dataset's structure, including the number of subjects, viewing angles, and specific walking conditions, like carrying a backpack or a box. This dataset is enriched by offering comprehensive information on each subject's walk, including extracted angles, body points, and gait energy images (GEIs), enabling both model-based and appearance-based approaches to be evaluated.

Finally, the Experiments subsection details the process of designing and running experiments to validate the dataset's utility. We tested the dataset using advanced deep learning techniques, such as a transformer-based model and measured the model's performance across several metrics. The experimental setup is designed to explore how well different features (such as angles and silhouettes) can contribute to accurate person identification, further enriching our understanding of gait as a biometric identifier. The results from these experiments offer valuable insights into the advantages and limitations of the TecNM Gait-DS and guide future research in this area.

3.1. Dataset creation

The recording of the dataset was carried out at the *Tecnologico Nacional de Mexico campus Culiacan* in a controlled environment in a conference room equipped for this activity. The Fig. 1 illustrates the workflow used to generate the TecNM Gait-DS dataset and how it can be used for the identification person problem by implementing DINO VIT algorithm. We begin by preparing the equipment and room with controlled conditions, as detailed in Section 3.1.1. We obtained consent from all participants in the study to form the TecNM Gait-DS database. Then, the collection of the videos of the five different walks began from this point. The conditions of this stage are detailed in Section 3.1.2. The verification of the captured data focused on observing that (1) the participants walk in the middle of the dotted line that is prepared to guide them in their walk, (2) that all the cameras observed the participants without delay in their capture, (3) that the lighting was captured in the same way by all cameras, and (4) that the videos from all cameras per participant had the same length and frame rate. Once the process of collecting and storing the videos was completed, see Section 3.2, the analysis of each scenario was carried out to determine the performance of identifying a person using the DINO VIT algorithm. This analysis is detailed in Section 3.3.

3.1.1. Room and equipment

The space used for recording the dataset consists of a room approximately 20 m long, 15 m wide, and 3 m high, illuminated with artificial light, and two large windows to allow the entry of natural light. In this room, an 8-meter-long corridor was defined for the participants' walk, which was surrounded by seven cameras positioned at 5 m each from the center of the corridor, each camera separated by 15 degrees of inclination as shown in Fig. 2.

3.1.2. Data acquisition

The methodology used to acquire the gait dataset is described as follows.

- Participant Recruitment: Recruit diverse participants willing to contribute to the dataset. Ensure that participants sign an image use permission form, providing consent for their data to be used for research purposes.
- Setup Design: Set up an 8-m walking corridor in a controlled environment with artificial lighting and no occlusions. Install seven cameras along the corridor, each positioned at a distance of 5 m from the center of the corridor. Use one CPU and two GPUs to record the gait sequences simultaneously from all seven cameras.
- Camera Calibration: Calibrate the cameras to ensure accurate synchronization and alignment. This step is crucial to maintain consistency and enable seamless integration of the data captured by different cameras.
- Data Collection: Instruct each participant to perform gait sequences in the walking corridor. Capture the gait data from each participant under different angles and walking variations. Each participant walked 28 times (twice per batch) through the corridor to generate 14 batches of videos (vb), six for normal walking and two for the other four walking variants.
- Angles (α): Conduct two takes for each participants' walk in 13 different angles. In the first take (from left to right), capture gait sequences at angles of 0°, 15°, 30°, 45°, 60°, 75°, and 90°. In the second take (from right to left), capture gait sequences at angles of 105°, 120°, 135°, 150°, 165°, and 180° as shown in Fig. 3. Adjust the position of the participant accordingly for each angle.

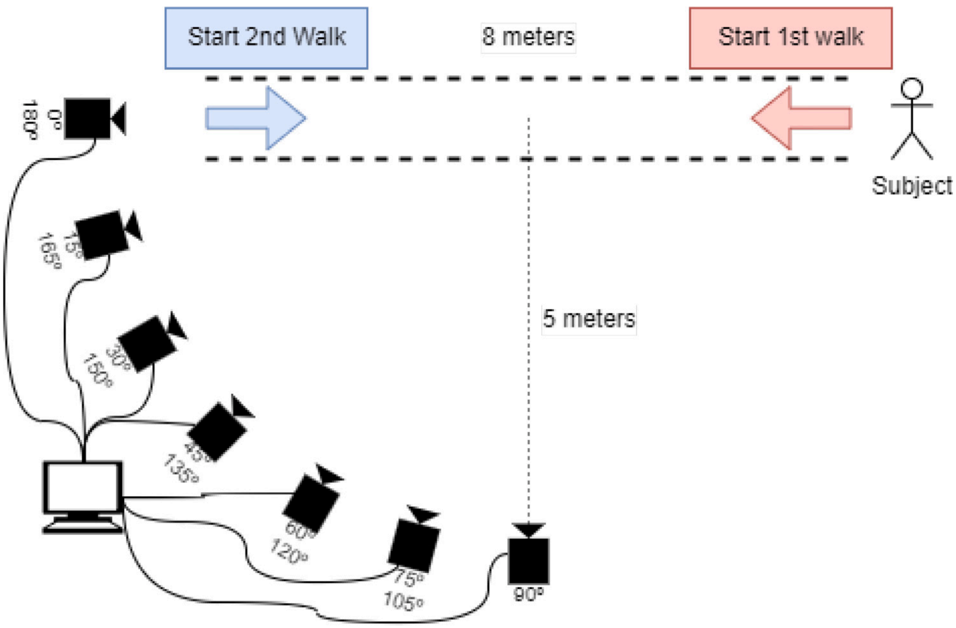

Fig. 2. Dataset recording scenario.

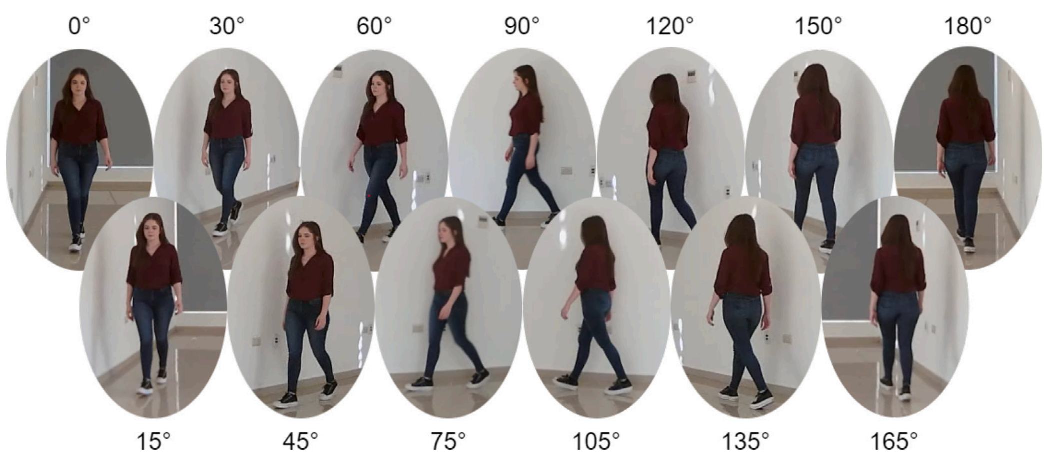

Fig. 3. Walking views angles in normal walking variation.

- Walking variations (*wv*): Instruct participants to walk naturally for the baseline walking variation. Additionally, capture gait sequences while participants carry a bag (*bg*), carry a backpack (bp), carry a box (*cx*), and wear a coat (*cl*) as shown in Fig. 4. This variation in walking variations adds diversity to the dataset.
- Metadata Collection: Collect additional metadata for each participant, including age and height. Ensure that participants provide this information accurately.
- Data Storage and Organization: Store the recorded gait sequences along with the associated metadata in a structured manner. Use appropriate file formats and naming conventions to ensure ease of access and retrieval.
- Privacy and Ethics: Ensure that all data collection procedures adhere to privacy and ethical guidelines. Protect the identity and personal information of participants according to legal requirements and ethical standards.

Following the above methodology, a gait dataset can be created with an 8-meter walking corridor, a controlled environment with seven cameras in specific positions to capture 13 angles for each participant, various walking variations, and associated participant metadata. This dataset will be valuable for gait recognition research and algorithm development.

3.2. Tecnm gait dataset

After implementing the above methodology, we organize the TecNM Gait Dataset as follows:

1. Subjects: We included 124 subjects in the dataset. Each subject is represented by a dedicated folder labeled with a number to protect their identity.
   (a) Walking Variants: Each subject's folder contains five subfolders representing different walking variants.
      - Normal walking (nm)
      - Carrying a backpack (bp)
      - Carrying a bag (*bg*)
      - Carrying a box (*cx*)
      - Wearing a coat (*cl*)

Fig. 4. Additional walking variations bg, bp, cx, and cl.

i. Videos: Each walking variant folder contains the batches videos. In this case α is 13, representing the angles considered.

- In the nm folder, there are 6 * α = 78 videos.
- In the other walking variant folders, there are 2 * α = 26 videos each.
- Each video corresponds to a combination of a specific walking variant, walk instance, and angle.
- The "Normal walking" folder has videos from 6 batches and 13 different angles.
- The other walking variant folders have videos from 2 batches and 13 different angles.
- The name of each video corresponds with the following nomenclature: $sub - wv - vb - \alpha$. Where sub is the subject number (from 001 to 124), wv is the walking variant (nm, bg, bp, cl, cx), vb is the video batch number (01 to 06 in nm and 01 to 02 in the other wv), and α is the corresponding angle.

ii. Extracted Features: Each video has a dedicated folder (with the same name of the video) inside the corresponding walking variant folder. Inside these video folders, there are subfolders for extracted features. These feature folders are divided into two main categories: (1) dense and (2) skeletons.

A. dense: The "dense" folder contains two subfolders:

- silhouettes: the "silhouettes" folder contains all the silhouettes extracted from each frame of the video.
- gei: the "gei" folder contains a GEI generated from all previously extracted silhouettes.

B. skeletons: Inside the "skeletons" folder, there are two subfolders.

- angles: contains CSV files with angle measurements derived from the skeleton data of each video frame.
- keypoints: contains CSV files with key point coordinates extracted from each video frame's skeleton.

This hierarchical structure allows for organized storage and retrieval of video data, as shown in Fig. 5, along with various types of extracted features for analysis.

3.2.1. *Subjects data statistics*

The global and individual statistics of the dataset by subject are the following:

- Global statistics:
  - 124 registered subjects
  - 22 568 video sequences
  - 62.67 h of video (10 s average length each video)
  - 36 GB of videos
  - 44.6% women
  - 55.4% men
  - 162.45 cm women height average (range from 148 cm to 180 cm)
  - 170.64 cm men height average (range from 162 cm to 197 cm)
- Individual statistics:
  - 182 video sequences (30 frames per second each)
  - 13 angles
  - 5 walking variants
  - 28 walks

3.2.2. *Feature extraction*

The proposed approach integrates advanced computer vision techniques, including DensePose from Detectron2 (Güler et al., 2018) for silhouette extraction and Detectron2 body keypoints detection (Wu et al., 2019) for obtaining detailed body pose information. The extracted features encompass both spatial and kinematic aspects of human movement, enhancing the discriminative power of the gait recognition system.

3.2.3. *Silhouette extraction*

The initial step in our feature extraction pipeline involves the application of DensePose from Detectron2 (Güler et al., 2018). This algorithm was employed to meticulously extract silhouettes of individuals from each walking video within our extensive dataset comprising 22,568 samples. The resulting silhouettes are a foundational representation capturing the spatial characteristics of the subjects' walking patterns. Fig. 6 shows an example of a silhouette extraction process.

In our methodology, we leverage a state-of-the-art model for dense human pose estimation to extract robust features from our gait recognition dataset. This model establishes dense correspondences between an RGB image and a surface-based representation of the human body. The primary objective of dense human pose estimation is to map each pixel in an image to a corresponding point on a standardized human body model.

The DensePose model was trained using a dataset of 50,000 individuals from the COCO dataset, which includes comprehensive annotations that map 2D image coordinates to 3D surface points on the human body. The annotation pipeline employed for this purpose was highly efficient, enabling the generation of dense correspondences despite

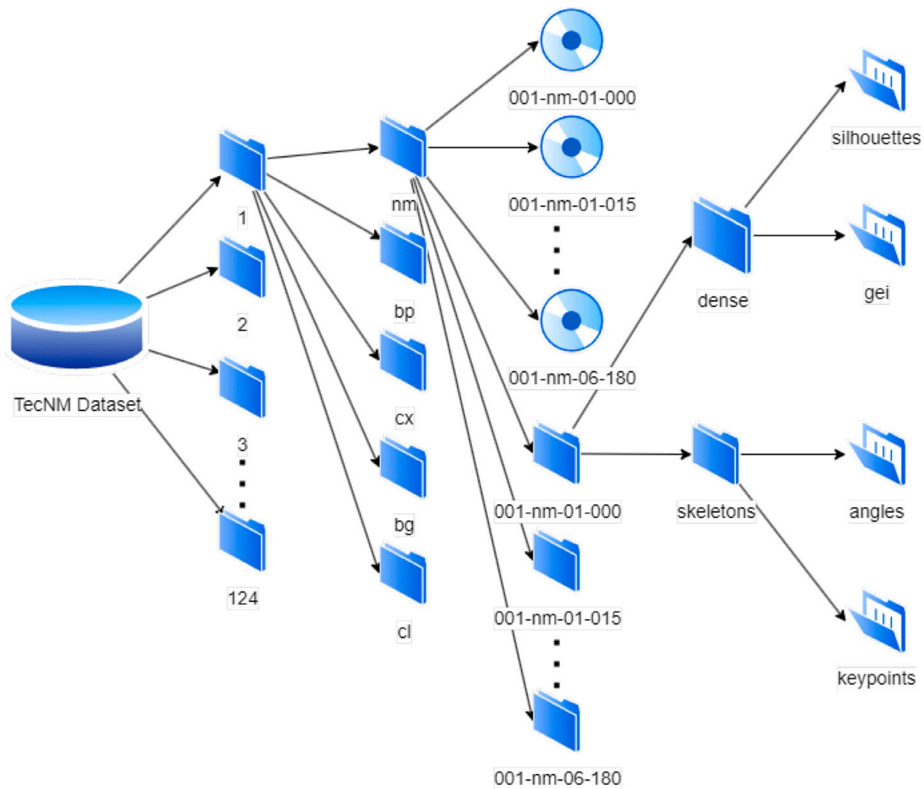

Fig. 5. Dataset hierarchical structure.

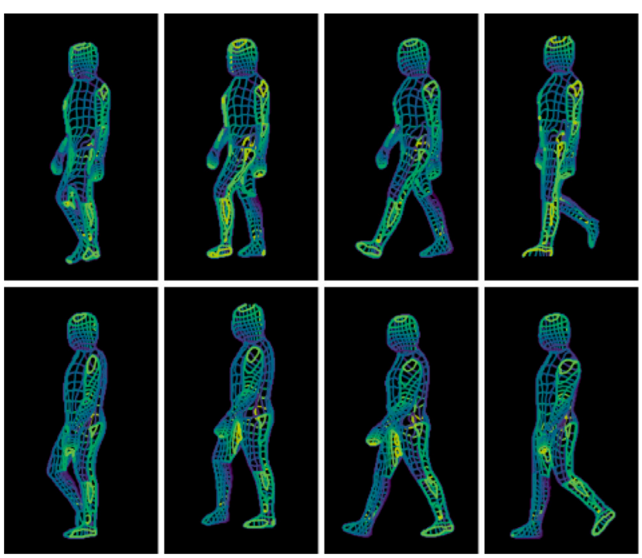

Fig. 6. Silhouettes extraction sequence from a video.

challenges such as background noise, occlusions, and variations in scale.

The architecture of the DensePose model combines fully-convolutional networks and region-based models. Through extensive experimentation, region-based models were found to outperform fully-convolutional networks, especially in scenarios with complex backgrounds and occlusions. To enhance accuracy further, a cascading approach was adopted, resulting in a highly accurate system capable of processing multiple frames per second on a single GPU.

A detailed architecture of the DensePose model used for feature extraction is illustrated in Fig. 7. This figure outlines the various

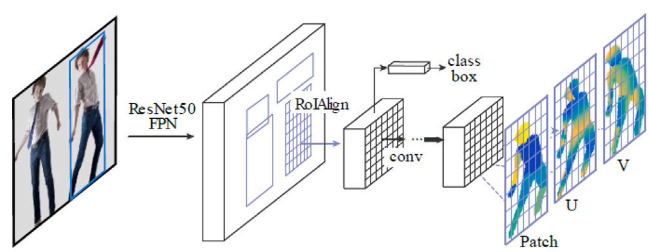

Fig. 7. DensePose from Detectron2 architecture implemented by Güler et al. (2018).

components of the model, including the input RGB image, the dense correspondence mapping, and the final feature representation utilized in our gait recognition framework

3.2.4. Gait energy image generation

Building upon the extracted silhouettes, we generated GEIs to encode the temporal dynamics of the walking sequences. This step enhances our dataset's discriminative power by incorporating shape and motion information. GEIs provide a comprehensive spatial–temporal representation of the gait dynamics essential for subsequent analysis. Fig. 8 shows a GEI example.

- Energy Calculation: For each processed silhouette region R_i', a binary mask M_i is created (Eq. (1)):

$$M_i(x, y) = \begin{cases} 255, & \text{if } R_i'(x, y) > 0 \\ 0, & \text{other case} \end{cases} \tag{1}$$

Here (x, y) denotes the pixel coordinates, and the condition $R_i'(x, y) > 0$ ensures that any pixel belonging to the subject (with values greater than zero) is assigned a value of 255 in the binary mask, representing the foreground. Conversely, all other pixels are set to zero, representing the background. This binary

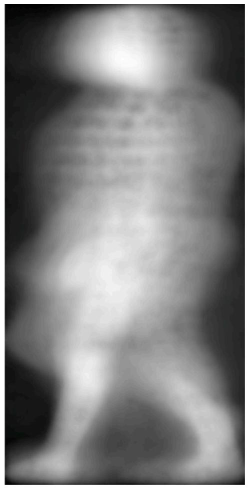

Fig. 8. GEI obtained from silhouettes.

thresholding operation is similar to the process performed by the `cv2.threshold` function (Itseez, 2015), ensuring that only the subject's silhouette is retained for further processing in the GEI generation. The resulting binary image highlights the regions of interest (the walking subject) while discarding the background, facilitating the construction of a clear and informative GEI for gait recognition tasks.
The energy of the silhouette E_i (Eq. (2)) is calculated using the distance transform:

$$E_i(x,y) = \sqrt{\sum_{(x',y')\in M_i} (x-x')^2 + (y-y')^2} \tag{2}$$

This equation computes the Euclidean distance between the point (x, y) and all points (x', y') on the binary mask M_i. In this context, $E_i(x, y)$ represents the energy at the specific location (x, y) of the silhouette, with the distance between foreground pixels contributing to the calculation of energy in the region.
To ensure the energy values are manageable and normalized, a logarithmic transformation is applied to the calculated energy using Eq. (3):

$$E_i = \log(E_i + 1) \tag{3}$$

This transformation helps scale the energy values, ensuring that large distances do not disproportionately influence the energy image construction.

- Energy Image Construction: An empty energy image I_{energy} is initialized with the same dimensions as the processed silhouette regions. For each processed silhouette region R_i', the energy E_i is added to the corresponding region in I_{energy} as follows (Eq. (4)):

$$I_{\text{energy}} += E_i \tag{4}$$

This iterative addition of energy values from each silhouette region accumulates the energy contributions into the final energy image I_{energy}, which encodes the temporal dynamics and spatial features of the subject's gait. This resulting image serves as a robust input for further gait analysis tasks.

3.2.5. Body key points extraction

We extracted 15 key points per frame from each walking video employing Detectron2's body key points detection (Wu et al., 2019). These key points detail the subject's body pose throughout the walking cycle, offering insights into the articulation of limbs and body segments as shown in Fig. 9.

The 15 key points are: wrists, elbows, shoulders and mid-shoulder, head (nose), three (left, right and mid) hip points, knees, and ankles.

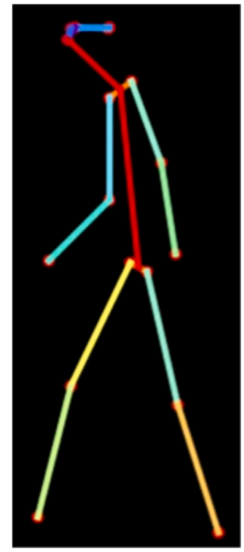

Fig. 9. Body key points representation.

3.2.6. Angles computation

From the extracted body key points, we computed 19 inclination and flexion angles using Pythagorean principles using Eq. (5) for inclination (floor referenced) and Eq. (6) for flexion.

$$\text{degrees} = \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \times \frac{180}{\pi} \tag{5}$$

where y_2 and y_1 are the ordinates, and x_2 and x_1 are the abscesses of the two considered points (head and mid-shoulder to calculate head inclination for example).

$$\text{degrees} = \arctan\left(\frac{c_2 - b_2}{c_1 - b_1}\right) - \arctan\left(\frac{a_2 - b_2}{a_1 - b_1}\right) \times \frac{180}{\pi} \tag{6}$$

where a_2, b_2, and c_2 are the ordinates, and a_1, b_1, and c_1 are the abscesses of the three considered points (wrist, ankle, and shoulder to calculate the ankle flexion for example).

These angles quantify the relative orientations and positions of different body segments, offering a comprehensive kinematic representation of the individual's gait dynamics. The inclination angles calculated are head, upper arms (shoulder to elbow), lower arms (elbow to shoulder), torso, upper legs (hip to knee), and lower legs (knee to ankle). Finally, the flexion angles calculated are neck, armpits, elbows, hip (left and right), and knees.

3.2.7. Dataset enrichment

The resulting feature set was applied to enrich our gait recognition dataset. Including diverse features contributes to a more comprehensive understanding of individual gait patterns, improving the potential for accurate and reliable recognition.

Our feature extraction process, encompassing silhouette extraction, GEI generation, and body key points-based angle computation, offers a holistic approach to gait representation. The effectiveness of our approach is validated through its application to a large-scale dataset, demonstrating its potential for real-world gait recognition applications.

3.3. Experiments

In this section, we describe the experiments performed to evaluate the quality of the extracted features, in this case, silhouettes, using the Detectron2 DensePose technique (Güler et al., 2018). From these silhouettes, GEIs were generated for each video of the TecNM Gait Dataset, comprising 22,568 videos of people walking. This quality assessment of the extracted features was performed in a context where the videos are divided into 124 subjects, 13 viewing angles, and five walking variations. The primary objective of these experiments is to evaluate the efficacy of the silhouettes and GEIs generated and to understand how they impact classification accuracy when recognizing a person due to viewing angle and walking variation.

The energy images generated were used to train a transformer model, specifically, the DINO VIT model developed by Facebook (Caron et al., 2021). During the training phase, we utilized 13 classes, each representing a distinct viewing angle. These classes were constructed by combining four GEIs from each subject, resulting in a total of 6448 GEIs. This dataset was divided into 80% for training and 20% for testing, ensuring a robust evaluation of the model's performance.

This data distribution and configuration (train-test-validation) was also used to generate experiments concerning the data obtained in the form of bending angles and inclination. For these experiments, images were generated from the frame-by-frame angle data matrices ($19xn$), where 19 is the number of angles extracted and n is the number of frames per video. Each pixel of the image corresponds to a normalized spectrum from 0 to 255 in grayscale of each angle calculated in each frame.

The DINO (Distillation with No Labels) model is particularly effective when combined with Vision Transformers. Unlike conventional supervised learning methods, DINO utilizes a self-supervised framework that capitalizes on the inherent capabilities of ViTs to understand and segment images semantically. This self-supervised approach ensures that the ViT features encapsulate explicit information about the semantic segmentation of an image, which is not as prominent in supervised ViTs or convnets.

During the training phase, the DINO model was trained on a large dataset without any labels, effectively learning to distill knowledge from the data itself. This process is akin to self-distillation, where the model refines its own predictions iteratively to improve accuracy. Notably, the DINO model has demonstrated remarkable performance, achieving an 80.1% top-1 accuracy on ImageNet in linear evaluation using a ViT-Base architecture.

The choice of these particular GEIs during training was made to capture a comprehensive representation of the variability in viewing angles and subjects, allowing the model to learn and adapt to different conditions.

This rigorous and comprehensive approach to training prepared the model for the task of viewing angle classification with a large amount of training data, which increased the likelihood that the model could make accurate predictions in various scenarios.

Once the model was trained, we evaluated its performance in five different evaluations, one for each walking variation, using two GEIs from each subject. This resulted in a total of 2324 GEIs used in each evaluation. The choice of two GEIs per subject allowed for a representative evaluation of the model in each walking variation and its ability to classify viewing angles.

Each evaluation focused on understanding how walking variation affects the model's ability to make accurate predictions. Subjects' energy images, captured during normal walking, were compared with those of other styles, providing valuable information on the impact of variations in walking variation on classification accuracy.

The five confusion matrices (Figs. 10 to 14) from these evaluations allowed a detailed analysis of the model's ability to adapt to different walking variations. They provided valuable information on the strengths and limitations of the model considering five scenarios(experiments). These evaluations contributed to a more complete understanding of the impact of extracted features and training on the model's ability to classify viewing angles in diverse walking variations.

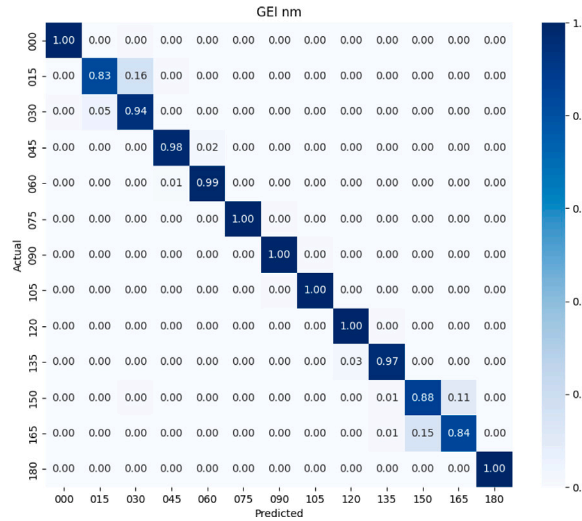

Fig. 10. Experiment A. Confusion matrix resulting from the GEI assessment in normal walking (nm).

4. Results

The results provide a comprehensive analysis of the performance of the DINO ViT-B/16 Transformer model in classifying walk cycle energy (GEI) images under various conditions. Each experiment, simulating different walking scenarios, revealed varied acceptance rates across different viewing angles. The model demonstrated robustness under normal walking conditions but showed sensitivity to external factors such as carrying a bag, wearing a coat, or carrying a box, with angles 015, 150, and 165 consistently exhibiting lower acceptance rates, suggesting areas for potential improvement. Figs. 10 to 14 present the classification outcomes for five walking variations: normal walking (nm, Fig. 10), carrying a bag (bg, Fig. 11), wearing a coat (cl, Fig. 12), carrying a backpack (bp, Fig. 13), and carrying a box (cx, Fig. 14). The confusion matrices in these figures illustrate the correct classification rates by comparing predicted classes to actual ones, ideally showing a diagonal line with values close to 1, indicating high effectiveness in the evaluation process.

4.1. Model sensitivity

4.1.1. Experiment A

In the normal walking experiment, the model demonstrated a remarkable ability to correctly classify most viewing angles, as shown in Fig. 10. However, a significant decrease in acceptance rate was observed for angles 015, 150, and 165. Upon closer analysis, it can be inferred that angle 015, representing a side view, may present an additional challenge due to variability in the subjects' lateral posture. The decrease in angles 150 and 165 may be due to variations in the subject's frontal posture, indicating a possible sensitivity of the model to subtle changes in body orientation.

4.1.2. Experiment B

The introduction of a bag during the walking negatively affected the acceptance rate, especially for angles 015, 150, and 165, as shown in Fig. 11. This suggests that the presence of external objects, such as a bag, can generate significant changes in the subject's appearance and movement, affecting the model's ability to generalize to these untrained conditions.

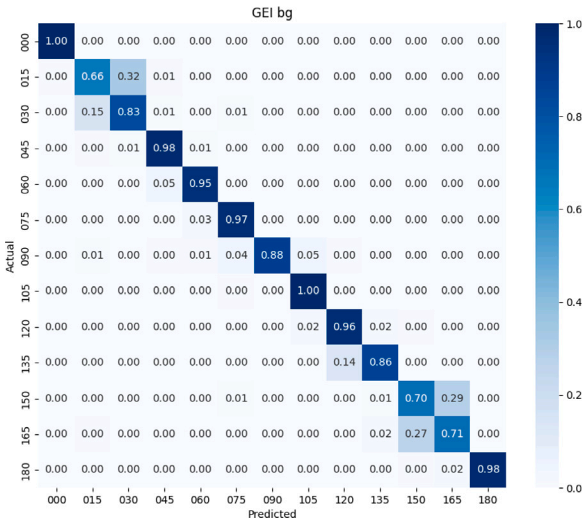

Fig. 11. Experiment B. Confusion matrix resulting from the GEI assessment in carrying a bag (bg).

Fig. 12. Experiment C. Confusion matrix resulting from the GEI assessment in wearing a coat (cl).

Fig. 13. Experiment D. Confusion matrix resulting from the GEI assessment in carrying a backpack (bp).

Fig. 14. Experiment E. Confusion matrix resulting from the GEI assessment in carrying a box (cx).

4.1.3. *Experiment C*

Similar to experiment B, the acceptance rate decreased for angles 015, 150, and 165; again, the sensitivity of the model is affected due to wearing a coat, as shown in Fig. 12. This highlights the importance of considering variations in load and clothing when training models for walking image classification tasks.

4.1.4. *Experiment D*

In this case, the loading of a backpack primarily affects angles 150, 165, and 015, indicating a consistency in the model's sensitivity to these specific conditions, as shown in Fig. 13. The high acceptance rate for other angles suggests that the model can adapt to variations in loading.

4.1.5. *Experiment E*

In Fig. 13, the acceptance rate decreases considerably for angles 030, 150, and 015, revealing a particular sensitivity to the act of carrying a box, as shown in Fig. 14. These results indicate that physical loading may introduce more complex variations in walking, challenging the model's ability to generalize especially in this experiment at an angle of 30 degrees.

Despite the model's sensitivity to specific conditions, such as carrying a bag, wearing a coat, or carrying a box, the model's overall performance in classifying walking images remains robust. The results highlight the importance of carefully considering specific training conditions to improve model generalization. Although certain angles and conditions presented challenges, the overall robustness of the model suggests that it could benefit from specific training strategies and architecture adjustments to address these specific variations. This analysis provides valuable guidelines for future research and improvements in the model's ability to adapt to diverse and challenging conditions.

4.2. *Cross validation*

To evaluate the robustness and generalization of the extracted gait characteristics, we used the DINO ViT-B/16 model. TecNM Gait-DS dataset, consisting of 124 subjects, was divided into several subsets following a consistent strategy. The subset of 25 subjects included 1300 energy images (GEI) divided into 80% for training and 20% for

Table 2
Cross validation experiments with different subsets sizes of TecNM Gait-DS.

Total subjects	Accuracy				
	nm	bg	cl	bp	cx
25	0.979	0.895	0.963	0.92	0.852
50	0.965	0.899	0.924	0.909	0.841
75	0.978	0.928	0.944	0.947	0.878
100	0.966	0.915	0.908	0.897	0.841
124	0.956	0.902	0.883	0.882	0.822

testing, and 650 additional images for validation testing in each gait variant: normal (nm), carrying a bag (bg), wearing a coat (cl), carrying a backpack (bp), and carrying a box (cx). The subset of 50 subjects included 2600 images divided equally, with 1300 images for validation in each variant. The subset of 75 subjects comprised 3900 images, with 1950 intended for validation in each variant. For the subset of 100 subjects, 5200 images were used for training and testing, and 2600 for validation in each variant.

The Table 2 shows the accuracies obtained in different data sets and walking conditions:

- Normal Walking (nm): the accuracy remains high in all sets, ranging from 0.956 to 0.979. This suggests that the model is extremely effective in recognizing walking under normal conditions, regardless of the number of subjects
- Carrying a Bag (bg): Accuracies range between 0.895 and 0.928, highlighting the model's ability to handle the variability introduced by carrying a bag. A slight decrease compared to normal gait is expected due to alterations in the walking pattern. Carrying a Bag (bg): Accuracies range between 0.895 and 0.928, highlighting the model's ability to handle the variability introduced by carrying a bag. Compared to normal gait, a slight decrease is expected due to alterations in the walking pattern.
- Wearing a Coat (cl): Accuracies are in the range of 0.883 to 0.963, indicating that the model maintains robust performance even with changes in clothing that may affect the subject's silhouette.
- Carrying a Backpack (bp): Results range from 0.882 to 0.947, demonstrating that the model can effectively adapt to variations in walking caused by carrying a backpack.
- Carrying a Box (cx): The accuracy is slightly lower, with values between 0.822 and 0.878. This reflects the additional challenge of walking when carrying a box, although the model still shows acceptable performance.

5. Discussion

TecNM Gait-DS stands out for including two scenarios for identifying a person through their way of walking: carrying a backpack and carrying a box. These scenarios add complexity to the moment of person recognition when considering the angle of placement of the cameras. On the other hand, TecNM Gait-DS contains a total of 22,568 videos that allow them to be classified through 13 classes. These videos and angles (classes) allow for a detailed analysis while maintaining a controlled capture environment for indoor applications, highlighting the behaviors that arise when transporting materials in office environments. From the analysis of the angles made in the different scenarios, it can be observed that in the vast majority of angles, regardless of their scenario (carrying a box, wearing a coat), there is an accuracy above 90%. However, recordings taken at angles 15°, 30°, 150°, and 165° show a lower performance regardless of the scenario, whereas the scenario of carrying a box shows a significantly inferior performance.

Among the observation differences of the individuals, the TecNM Gait-DS captures 13 different viewing angles, which exceeds the six angles of CMU MoBo, the two of SOTON, and the 11 of CASIA B.

Table 3
CCR averages from CASIA B and TecNM Gait-DS experiments.

	Experiments CCR avg				
	A (nm)	B (bg)	C (cl)	D (bp)	E (cx)
CASIA B	0.997	0.289	0.678	–	–
TecNM Gait-DS	0.956	0.902	0.883	0.882	0.822

By capturing the walk from multiple angles, TecNM Gait-DS allows a more complete and detailed analysis of body movements, which is crucial for biometric applications and biomechanical research. It also identifies which angles and conditions (carrying a box, wearing a coat, etc.) where the patterns or characteristics of the individual obtained are not favorable for their identification, allowing the exploration of camera combinations at different angles. This behavior is shown in the confusion matrices in Figs. 10 to 14.

The TecNM Gait-DS offers 22,568 video clips with an average length of 10 s per video at 30 frames per second, facilitating a nuanced exploration of individual-level gait variations, considering the two scenarios that differentiate our database. We remark that the video length was not specified for CASIA B, SOTON, OU ISIR, and OU ISIR MVLP. The frame rate for CASIA B, SOTON, and OU ISIR MVLP is 25fps; only OU ISIR is considered 30fps.

TecNM Gait-DS introduces a broader spectrum of conditions in a single database compared to the rest of the works [CASIA-B, OU ISIR, OU ISIR MVLP, MOBO, CMU, GaIt3d, GREW], including carrying a backpack and a box, common actions in controlled environments. When analyzing TecNM Gait-DS it was identified that for all scenarios, cameras with angles 15°, 30°, 150°, and 165° decrease the precision to identify the subject by HGR, so for the placement of the cameras in real scenarios, it could be favored by considering the angles 0°, 45°, 60°, 75°, 90°, 105°, 120°, 135°, and 180° since the patterns or characteristics of the person are better captured. This knowledge of the walking conditions helps guide research to improve the robustness of the models and algorithms developed, ensuring that they are applicable in real-world scenarios where walking conditions can vary significantly in addition to the placement of the cameras.

The databases (CASIA B, OU ISIR, OU ISIR MVLP, SOTON) are made up of eastern participants. In the case of TecNM Gait-DS, its focus is on collecting Latin American people. This aspect allows for consideration of racial biases that have sometimes been reported in studies such as (Hill et al., Gonzales et al.). The methodology for capturing people from multiple viewing angles allows for generating 182 videos per subject, considering five scenarios from 13 different angles to observe variations in walking. Thus, these camera arrangements will allow for the future biomechanical analysis of walking (Mu et al., Alaqtash et al.), biometric security systems (Arshad et al.) or the detection of problems during a clinical evaluation of walking (Devanne et al.).

CASIA B dataset study conducted in 2006, the evaluation of the gait recognition algorithm was based on correct classification rates (CCR) using three sets of experiments with normal, carrying a bag, and wearing a coat walking variations (A, B, and C). The CCR averages obtained were 0.977, 0.289, and 0.678 for experiments A, B, and C, respectively, as shown in Table 3.

TecNM Gait-DS, included feature extraction from silhouettes and energy image generation (GEI) using a Facebook DINO ViT-B/16 model. The results obtained indicate a CCR average of 0.956, 0.902 and 0.883 for experiments A, B and C, respectively. In addition, experiments (D and E) were introduced with CCR averages of 0.882 and 0.822, related to walking carrying a backpack and a box, respectively, as shown in Table 3.

The comparison is made directly with the CASIA B dataset, as the similarity between the two datasets is closer than the other alternatives (e.g., OU-ISIR-MVLP, which has over 250,000 videos but only one walk variant). CASIA B and TecNM Gait-DS share three walk variations, but the TecNM Gait-DS dataset incorporates two additional variations.

Table 4
Comparison of the main human gait datasets.

Name	# Subjects	# Videos per subject	# Total videos	Environment	Views (angles)	Walking variations
CMU MoBo	25	24	600	Indoor	6	Three walking speeds, carrying a ball (all on a treadmill)
SOTON	115	–	2128	Indoor & Outdoor	2	Normal walking on a treadmill
CASIA B	124	110	13 640	Indoor	11	Normal walking, carrying a bag, wearing a coat
OU-ISIR	4007	–	31 368	Outdoor	4	Normal walking
OU-ISIR-MVLP	10 307	–	259 013	Indoor	14	Normal walking
GREW	26 345	–	128 671	Wild	–	Undefined
Gait3d	4000	–	25 309	Wild	–	Undefined
TecNM Gait-DS (Ours)	124	182	22 568	Indoor	13	Normal walking, carrying a bag, carrying a backpack, carrying a box, wearing a coat

Including additional walking variation in TecNM Gait-DS expands the situations evaluated, reflecting a more complete understanding of the capabilities of the transformer model, especially in more complex and varied contexts present in TecNM Gait-DS. Despite the difference in years between the two studies, this comparison highlights significant advances in the field of gait recognition, highlighting the impact of technological and methodological evolution over time.

Table 4 presents a comprehensive overview of six prominent gait recognition datasets, shedding light on their key characteristics, including the number of subjects, videos per subject, total videos, environmental conditions, viewpoints, and walking variations. This discussion analyzes these datasets' significance and potential impact on gait recognition research.

TecNM Gait-DS surpasses CMU MoBo, considering subject count (124 vs. 25) and video quantity per subject (182 vs. 24), presenting a more extensive and diverse dataset. Featuring 22,568 total videos in contrast to CMU MoBo's 600, TecNM Gait-DS provides a significantly larger collection for detailed analysis. The shared indoor environment between both datasets fosters standardization.

With 13 view angles, TecNM Gait-DS stands out, offering a richer perspective than CMU MoBo's six angles. Moreover, TecNM Gait-DS captures various gait situations, including carrying objects and wearing a coat, enhancing the dataset's versatility.

TecNM Gait-DS encompasses 124 subjects, slightly surpassing SOTON, which comprises 115 subjects—though the disparity is statistically insignificant. Notably, TecNM Gait-DS exhibits a marginally larger subject pool. Each subject in TecNM Gait-DS is associated with 182 videos, while specific information on the video count per subject in SOTON is unavailable, hindering a precise quantitative comparison. This lack of detailed data in SOTON complicates the assessment in this particular aspect. In terms of total videos, TecNM Gait-DS boasts 22,568 videos, significantly exceeding SOTON's total of 2128 videos. This substantial difference establishes TecNM Gait-DS as the superior dataset in total video content.

SOTON's data collection contains indoor and outdoor environments, whereas TecNM Gait-DS concentrates solely on indoor settings. The relevance of this distinction depends on the specific requirements of the intended application. Notably, TecNM Gait-DS's provision of 13 view angles is a detail absent in the SOTON dataset. The more significant number of view angles in TecNM Gait-DS contributes to a more comprehensive and detailed perspective.

The TecNM Gait-DS dataset boasts 182 videos per subject, surpassing the CASIA B dataset, which contains 110 videos per subject. This disparity in video quantity per subject within TecNM Gait-DS facilitates a more nuanced exploration of gait variations at the individual level. TecNM Gait-DS comprises 22,568 videos, whereas CASIA B includes 13,640 videos. TecNM Gait-DS markedly excels in the overall number of videos, signifying a substantial resource for comprehensive gait analysis.

Moreover, TecNM Gait-DS provides a more extensive array of perspectives with 13 view angles compared to CASIA B's 11 view angles. This feature of TecNM Gait-DS offers a slightly more comprehensive view, capturing data from additional angles and enhancing the dataset's potential for in-depth analysis.

TecNM Gait-DS encompasses 124 subjects, presenting a notable contrast to the substantially larger datasets of OU ISIR with 4007 subjects and OU ISIR MVLP boasting 10,307 subjects. In terms of subject volume, TecNM Gait-DS is markedly smaller compared to both OU ISIR and OU ISIR MVLP. Analyzing the total video count, TecNM Gait-DS comprises 22,568 videos, positioning itself in an intermediary capacity between OU ISIR with 31,368 videos and OU ISIR MVLP with an extensive 259,013 videos. This intermediate positioning underscores the dataset's unique characteristics.

Distinguishing itself, TecNM Gait-DS incorporates diverse gait variations, including scenarios such as carrying a bag, backpack, box and wearing coat. This comprehensive representation of gait conditions within TecNM Gait-DS holds significant utility for applications demanding a broader spectrum of gait diversity. In comparison, OU ISIR MVLP, with its considerably larger subject and video counts, stands out as an advantageous resource, especially for tasks requiring substantial data volumes for training machine learning models.

To evaluate the robustness of our model, we performed a cross-validation using the full dataset (TecNM Gait-DS) and three subdatasets (25, 50, 75, 100 subjects).

We note that, in general, model accuracy is high across all data sets and walking conditions. Slight decreases in accuracy with increasing number of subjects and variations in gait conditions are expected and acceptable within the context of gait recognition. These trends suggest that the model is sufficiently robust and generalizable.

Results with 124 subjects show accuracies that are comparable to those obtained with smaller sets, reaffirming that the model can handle a large number of subjects without losing accuracy significantly. This is crucial, as the diversity and quantity of training data contribute to the model's ability to better generalize to new, unseen data.

The cross-validation experiments confirm that the number of 124 subjects is adequate to train a robust and accurate model in various walking conditions. The high accuracy in different conditions and the ability of the model to handle a large number of subjects underline the accuracy and generalizability of the DINO ViT-B/16 model. These results validate the use of TecNM Gait-DS as a rich and diverse database, ideal for the development of advanced gait recognition algorithms.

In summary, TecNM Gait-DS's comprehensive methodology, marked by increased capture angles, a higher number of subject-specific videos, and diverse gait variations, yields a rich and varied dataset with potential applications spanning biomechanics research (Mu et al., 2010; Alaqtash et al., 2011), biometric system development (Arshad et al., 2022), and clinical gait analysis (Hill et al., 2020; Gonzales et al., 2020).

6. Conclusion

The TecNM Gait-DS human gait database significantly contributes to the field of HGR research with its practical applications. It features 22,568 video sequences captured from 124 subjects and spanning 13 different viewing angles. One of the key merits of TecNM Gait-DS is its comprehensive data structure, enriched with a broad source of information, including the location of key body points, bending and leaning angles during gait, as well as silhouettes and walking energy derived from the videos. Additionally, the database introduces practical scenarios such as carrying a backpack and a box, making it highly relevant for real-world applications. These features make TecNM Gait-DS a valuable resource for both model-based and appearance-based gait recognition methods.

Likewise, each video was processed to extract the relevant characteristics of the walk cycle in the two main HGR approaches (model-based and appearance-based). For the model-based approach, the Dectectron2 (Wu et al., 2019) methodology was used, and for the appearance-based approach, DensePose (Güler et al., 2018) was used. From the silhouettes extracted by DensePose, we formed an HGR by video processing.

While fundamental datasets exist in human gait research, such as CMU MoBo, SOTON, OU ISIR, OU ISIR MVLP, and the widely used CASIA B, it is essential to recognize the need for increased diversity in gait variants, viewing angles, and amount of data. TecNM Gait-DS addresses this challenge by offering more diverse walking variants and viewing angles, particularly in scenarios like carrying objects, which are less explored in other databases. We analyzed how camera angles and extracted characteristics of the scenarios included in TecNM Gait-DS have improved the performance of person identification using DINO ViT-B/16. The results indicated a CCR average of 0.956, 0.902, and 0.883 for experiments with normal walking, carrying a bag, and wearing a Coat, respectively. In addition, experiments carrying a backpack and carrying a box were introduced with CCR averages of 0.882 and 0.822, respectively. These new scenarios show that for person identification, there is a need for new algorithms robust to occlusions or that work with partial information, which is related to the angle of observations to extract the gait characteristics. Besides, we observe that the angles with better performance for the five scenarios analyzed are: 0°, 45°, 60°, 75°, 90°, 105°, 120°, 135°, and 180°.

However, despite these contributions, there are limitations. The TecNM Gait-DS database primarily focuses on indoor environments, which may limit its generalizability to outdoor or uncontrolled settings. Additionally, while the dataset includes 124 subjects, larger datasets such as OU ISIR MVLP contain even more subjects and more diverse environmental conditions; hence, we will focus on expanding the number of subjects and scenarios in future versions of our dataset to address these limitations.

In summary, TecNM Gait-DS presents a new set of videos collected from Latin American people. This provides new examples for person identifications since existing databases such as CASIA B, OU ISIR, OU ISIR MVLP, and SOTON are focused on eastern participants. We provide the dataset and the algorithms used in https://github.com/MisaelZazueta/TecNM_Gait-DS.

In future works, we will focus on improving the person identification task by using, comparing, and merging the TecNM Gait-DS with other datasets. This will allow us to address some of the limitations mentioned and extend the scope of existing research, enhancing the robustness of algorithms to different conditions and environments. Additionally, we will focus on developing new algorithms capable of handling occlusions and partial information, which are critical for real-world applications. The potential for future research and development in this area is vast, leading and motivating us to continue our work in this field.

This work was processed by Grammarly to improve writing, translation and identify areas for improvement.

CRediT authorship contribution statement

Jose Misael Burruel-Zazueta: Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Data curation. **Hector Rodriguez-Rangel:** Writing – review & editing, Supervision, Project administration. **Vicenç Puig Cayuela:** Writing – review & editing, Formal analysis, Conceptualization. **Manuel Alejandro Medrano-Diaz:** Writing – review & editing, Software. **Sofia Isabel Fernandez-Gregorio:** Writing – review & editing. **Luis Alberto Morales-Rosales:** Writing – review & editing, Formal analysis, Conceptualization.

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

Data will be made available on request.

Acknowledgments

We thank the Mexican National Council of Humanities, Science, and Technology (CONAHCYT) for the funding to carry out this work through scholarships 968524, 1003615, 478122, and research project 613.

References

Al Kork, S.K., Gowthami, I., Savatier, X., Beyrouthy, T., Korbane, J.A., Roshdi, S., 2017. Biometric database for human gait recognition using wearable sensors and a smartphone. In: 2017 2nd International Conference on Bio-Engineering for Smart Technologies (BioSMART). IEEE, pp. 1–4.

Alaqtash, M., Sarkodie-Gyan, T., Yu, H., Fuentes, O., Brower, R., Abdelgawad, A., 2011. Automatic classification of pathological gait patterns using ground reaction forces and machine learning algorithms. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 453–457. http://dx.doi.org/10.1109/IEMBS.2011.6090063.

An, W., Yu, S., Makihara, Y., Wu, X., Xu, C., Yu, Y., Liao, R., Yagi, Y., 2020. Performance evaluation of model-based gait on multi-view very large population database with pose sequences. IEEE Trans. Biometr. Behav. Identity Sci. 2 (4), 421–430. http://dx.doi.org/10.1109/TBIOM.2020.3008862.

Arshad, H., Khan, M.A., Sharif, M.I., Yasmin, M., Tavares, J.M.R., Zhang, Y.-D., Satapathy, S.C., 2022. A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition. Expert Syst. 39 (7), e12541.

Bouchrika, I., 2018. A survey of using biometrics for smart visual surveillance: Gait recognition. In: Karampelas, P., Bourlai, T. (Eds.), Surveillance in Action: Technologies for Civilian, Military and Cyber Surveillance. Springer International Publishing, Cham, ISBN: 978-3-319-68533-5, pp. 3–23. http://dx.doi.org/10.1007/978-3-319-68533-5_1.

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A., 2021. Emerging properties in self-supervised vision transformers. CoRR abs/2104.14294 arXiv:2104.14294 URL https://arxiv.org/abs/2104.14294.

Devanne, M., Wannous, H., Daoudi, M., Berretti, S., Del Bimbo, A., Pala, P., 2016. Learning shape variations of motion trajectories for gait analysis. In: 2016 23rd International Conference on Pattern Recognition. ICPR, pp. 895–900. http://dx.doi.org/10.1109/ICPR.2016.7899749.

Gafurov, D., 2007. A survey of biometric gait recognition: Approaches, security and challenges. In: Annual Norwegian Computer Science Conference. Annual Norwegian Computer Science Conference Norway, pp. 19–21.

Gong, Y., Liu, G., Xue, Y., Li, R., Meng, L., 2023. A survey on dataset quality in machine learning. Inf. Softw. Technol. (ISSN: 0950-5849) 162, 107268. http://dx.doi.org/10.1016/j.infsof.2023.107268, URL https://www.sciencedirect.com/science/article/pii/S0950584923001222.

Gonzales, M.M., Wang, C.-P., Quiben, M., MacCarthy, D., Seshadri, S., Jacob, M., Hazuda, H., 2020. Joint trajectories of cognition and gait speed in mexican American and European American older adults: The san antonio longitudinal study of aging. Int. J. Geriatr. Psychiatry 35 (8), 897–906.

Gross, R., 2001. The CMU Motion of Body (Mobo) Database. Carnegie Mellon University, The Robotics Institute.

Güler, R.A., Neverova, N., Kokkinos, I., 2018. Densepose: Dense human pose estimation in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7297–7306.

Hadid, A., Ghahramani, M., Kellokumpu, V., Pietikäinen, M., Bustard, J., Nixon, M., 2012. Can gait biometrics be Spoofed? In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). pp. 3280–3283.

Han, J., Bhanu, B., 2006. Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. 28 (2), 316–322. http://dx.doi.org/10.1109/TPAMI.2006.38.

Hill, C.N., Reed, W., Schmitt, D., Sands, L.P., Queen, R.M., 2020. Racial differences in gait mechanics. J. Biomech. (ISSN: 0021-9290) 112, 110070. http://dx.doi.org/10.1016/j.jbiomech.2020.110070, URL https://www.sciencedirect.com/science/article/pii/S0021929020304942.

Hou, S., Cao, C., Liu, X., Huang, Y., 2020. Gait lateral network: Learning discriminative and compact representations for gait recognition. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (Eds.), Computer Vision – ECCV 2020. Springer International Publishing, Cham, ISBN: 978-3-030-58545-7, pp. 382–398.

Itseez, 2015. Open source computer vision library. https://github.com/itseez/opencv.

Iwama, H., Okumura, M., Makihara, Y., Yagi, Y., 2012. The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Trans. Inf. Forensics Secur. 7 (5), 1511–1521. http://dx.doi.org/10.1109/TIFS.2012.2204253.

Jawed, B., Khalifa, O.O., Newaj Bhuiyan, S.S., 2018. Human gait recognition system. In: 2018 7th International Conference on Computer and Communication Engineering. ICCCE, pp. 89–92. http://dx.doi.org/10.1109/ICCCE.2018.8539245.

Jun, K., Lee, D.-W., Lee, K., Lee, S., Kim, M.S., 2020. Feature extraction using an RNN autoencoder for skeleton-based abnormal gait recognition. IEEE Access 8, 19196–19207. http://dx.doi.org/10.1109/ACCESS.2020.2967845.

Li, X., Makihara, Y., Xu, C., Yagi, Y., Yu, S., Ren, M., 2020b. End-to-end model-based gait recognition. In: Proceedings of the Asian Conference on Computer Vision. ACCV.

Li, N., Zhao, X., Ma, C., 2020a. JointsGait:A model-based gait recognition method based on gait graph convolutional networks and joints relationship pyramid mapping. arXiv:2005.08625 URL https://arxiv.org/abs/2005.08625.

Liao, R., An, W., Yu, S., Li, Z., Huang, Y., 2020. Dense-view GEIs Set: View space covering for gait recognition based on dense-view GAN. In: 2020 IEEE International Joint Conference on Biometrics. IJCB, pp. 1–9. http://dx.doi.org/10.1109/IJCB48548.2020.9304910.

Lin, B., Zhang, S., Bao, F., 2020. Gait recognition with multiple-temporal-scale 3D convolutional neural network. In: Proceedings of the 28th ACM International Conference on Multimedia. MM '20, Association for Computing Machinery, New York, NY, USA, ISBN: 9781450379885, pp. 3054–3062. http://dx.doi.org/10.1145/3394171.3413861.

Mazurek, S., Wielgosz, M., 2023. Assessing dataset quality through decision tree characteristics in autoencoder-processed spaces. arXiv preprint arXiv:2306.15392.

Mu, T., Pataky, T.C., Findlow, A.H., Aung, M.S.H., Goulermas, J.Y., 2010. Automated nonlinear feature generation and classification of foot pressure lesions. IEEE Trans. Inf. Technol. Biomed. 14 (2), 418–424. http://dx.doi.org/10.1109/TITB.2009.2028338.

Rani, V., Kumar, M., 2023. Human gait recognition: A systematic review. Multimedia Tools Appl. 1–35.

Seng, M.T.C., bt Rashidi, N.H., et al., 2024. Enhanced brain tumor classification using modified ResNet50 architecture. Borneo J. Sci. Technol. 6 (1), 11–18.

Shutler, J.D., Grant, M.G., Nixon, M.S., Carter, J.N., 2004. On a large sequence-based human gait database. In: Applications and Science in Soft Computing. Springer, pp. 339–346.

Takemura, N., Makihara, Y., Muramatsu, D., Echigo, T., Yagi, Y., 2018. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. IPSJ Trans. Comput. Vis. Appl. 10, 1–14.

Tan, D., Huang, K., Yu, S., Tan, T., 2006. Efficient night gait recognition based on template matching. In: 18th International Conference on Pattern Recognition. ICPR'06, Vol. 3, pp. 1000–1003. http://dx.doi.org/10.1109/ICPR.2006.478.

Wang, L., Tan, T., Ning, H., Hu, W., 2003. Silhouette analysis-based gait recognition for human identification. IEEE Trans. Pattern Anal. Mach. Intell. 25 (12), 1505–1518. http://dx.doi.org/10.1109/TPAMI.2003.1251144.

Wang, X., Yan, W.Q., 2020. Human gait recognition based on frame-by-frame gait energy images and convolutional long short-term memory. Int. J. Neural Syst. 30 (01), 1950027.

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., Girshick, R., 2019. Detectron2. https://github.com/facebookresearch/detectron2.

Yu, S., Tan, D., Tan, T., 2006. 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, pp. 441–444.

Zhang, Z., Tran, L., Liu, F., Liu, X., 2022. On learning disentangled representations for gait recognition. IEEE Trans. Pattern Anal. Mach. Intell. 44 (1), 345–360. http://dx.doi.org/10.1109/TPAMI.2020.2998790.

Zhang, Z., Tran, L., Yin, X., Atoum, Y., Liu, X., Wan, J., Wang, N., 2019. Gait recognition via disentangled representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR.

Zheng, J., Liu, X., Liu, W., He, L., Yan, C., Mei, T., 2022. 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. CVPR, pp. 20228–20237.

Zhu, Z., Guo, X., Yang, T., Huang, J., Deng, J., Huang, G., Du, D., Lu, J., Zhou, J., 2021. Gait recognition in the wild: A benchmark. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. ICCV, pp. 14789–14799.