Research Article

PHASEY: A Contrastive Learning Approach for Enhanced Human Gait Phases Recognition

[1]Urvashi, [2]Deepak Kumar, [1]Vinay Kukreja and [1]Ayush Dogra

[1]Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India
[2]Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India

Article history
Received: 26-09-2024
Revised: 26-12-2024
Accepted: 25-02-2025

Corresponding Author:
Vinay Kukreja
Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India
Email:
onlyvinaykukreja@gmail.com

Abstract: Human gait has gained much attention in behavioral biometrics as it possesses unique and distinctive characteristics. Gait phases, which describe the different patterns of human walking, are significant for the analysis and understanding of movement in an individual. Hence, the identification of gait phases is important for the accurate determination and interpretation of walking patterns, ranging from healthcare and security to rehabilitation. This study aims to propose an efficient model, called Precision Human Gait Activity Segmentation for Gait Phases Recognition using YOLOv9 (PHASEY), and a contrastive learning method that localizes and recognizes the gait stance phase and swing phase more efficiently and correctly. The proposed PHASEY model localizes the walking Gait phase patterns and distinguishes movement patterns in each of the phases. It uses CSPDarknet 53 as its backbone, which is further trained to identify swing and stance gait phases using silhouette images. The PHASEY model has three prime components- backbone, neck, and head. There is feature extraction from the backbone, then, visualization of those features through Grad-CAM within the neck is provided. Lastly, the head unit is accountable for the gait phase classification. By training the CSPDarknet 53 in the PHASEY model, the accuracy, as well as Intersection over Union (IoU), and inference time were calculated with different epochs. The experimental results show that the model attained the highest accuracy of 0.9907 at the epoch value 50. After comparing the YOLO models, it was evident that YOLOv9 achieved the highest accuracy of 94.8%, with a Precision value of 93.1%, Recall 91.9% and IoU with 87.8%. By utilizing this real-time object detection model for determining the phases of the gait cycle, the approach demonstrated exceptional performance in both localization and classification across different subjects.

Keywords: Swing Gait Phase, Stance Gait Phase, Object Detection, You Only Look Once (YOLO), Pretrained Network

Introduction

Techniques involved in the identification of humans from behavioral or physiological characteristics comprise biometrics. Speech and facial features are the most used traits in biometric recognition. Facial recognition and fingerprint-based applications have been widely applied in medical diagnostics-related fields. However, an increasing modality of biometry is gait. Biometric recognition through gait occurs because the walking patterns of a person have established the possibility of identification in this domain (Huang *et al.*, 1999). Human gait analysis has proven to be an important field of study that offers an in-depth understanding of how people walk. Human gait describes how a person moves their arms or legs as they walk, drink, sit, leap, and do many other things. In other words, the walking gait patterns are broadly divided into two phases: the stance phase and the swing phase (Umberger, 2010). The duration for which the foot is in contact with the ground is known as the stance phase, and while the foot is in contact with the air is known as the swing phase. The stance and swing phases make a complete gait cycle. These phases comprise sub-phases further. Initial Contact, Loading Response, Mid Stance, and Terminal Stance fall under the stance phase, while Pre-Swing, Toe Off, Mid Swing, and Terminal Swing are the steps of the swing phase. Accurate recognition of the gait patterns helps in identifying abnormal walking patterns in patients suffering from any neurological disorder, like Parkinson's disease or any other medical injury. Moreover, for monitoring rehabilitation progress and optimizing the performance of sportspersons, gait recognition has played its part well. Several different methods have been proposed for human gait recognition. These methods can be broadly classified into two categories: Model-based and Motion-based techniques (Kusakunniran, 2020). In the model-based approach, model parameters represent the human body structure, which is fitted based on the extracted image feature. On the other hand, a compact representation is used to characterize a motion pattern of the human body without considering the underlying model structure. Researchers have worked on various techniques of gait recognition since the early period of the emergence of gait as a Biometric. Approaches of gait can be categorized broadly into two groups.

Model-Based Approach

In model-based approaches, the kinematics of joint angles are modeled while the subjects are walking. The approach started with extracting the skeletons and joints of the human body in each frame. Model-based techniques create a human body model and then extract its characteristics. Cunado *et al.* modeled leg movement with a pendulum in 1997 (Wan *et al.*, 2019) and the changes in the inclination of the legs were used for gait recognition. As compared to motion-based methods, model-based methods can be more robust to many variations, only if human bodies are correctly and accurately modeled (Liao *et al.*, 2020). Some traditional model-based methods use a simple stick model to simulate legs, and then the leg movement is simulated by an articulated pendulum during walking. Then, human identification frequency components are extracted as gait features at the end.

Motion-Based Approach

Motion-based approach is referred to as Image measuring techniques that use the samples' walking form to determine the gait features. Therefore, these methods do not require working on a model of the walking steps of a human being. Motion-based approaches can be broadly categorized as spatial and temporal. However, they can be further bifurcated into four subcategories (Rida *et al.*, 2019). These subcategories are contour, optical flow, silhouette, moments, and gait energy/entropy/motion history (Rida, 2019). The contours can be interrupted by intra-class variations, but have a low computational cost (Zhang *et al.*, 2010). An example of gait recognition-based contour features was introduced by (Hayfron-Acquah *et al.*, 2003). Silhouettes can be taken into consideration as a whole per subject. This can have more advantages because the errors of silhouette segmentation can be avoided (Boisvert *et al.*, 2013). The spatial and temporal features of gait are extracted by energy features using a single and robust signature (Roy *et al.*, 2012). The optical flow extracts the dynamic aspect of human motion and represents a robust feature representation against the various intra-class variations.

Gait biometrics means recognizing an individual based on his/her walking style. In General, gait recognition can be implemented on two types of data: a sequence of images (e.g., from a video), or an inertial gait time series generated by inertial sensors (Zou *et al.*, 2020). Various deep-learning algorithms (Kececi *et al.*, 2020) have been used in this area of research to recognize gait, abnormal gait, etc. in surveillance systems, biometrics, Rehabilitation centers, hospitals, and various other places. Gait phase recognition has traditionally been performed at broader levels, such as stance and swing phases. Furthermore, significant efforts have been directed towards classifying the sub-phases within these categories, including initial contact, loading response, mid-stance, and terminal stance.

The PHASEY technique is a new object detection and localization technique followed by the classification of walking patterns. PHASEY stands for PHASE YOLOv9, which operates on the principles of a Contrastive Multiphase GaitNet scheme. It applies a self-supervised learning technique with a comparison of positive and negative samples. It effectively and precisely segments and classifies all the different phases of a gait cycle by using some advanced object detection models. This paper addresses a new challenge, the simultaneous localization of the object and classification of the subject's walking phase. This approach is divided into two stages: the first stage includes object detection, and the second stage is phase classification. Previous machine learning algorithms have been mainly designed to extract features and classify walking phases, with many studies evaluating the efficiency and performance of various deep learning (DL) models in this domain. This work builds upon these advances to propose a better methodology. The PHASEY framework can extract task-related features that are necessary for differentiating various gait phases. A dual-layered approach is used to achieve this reorganization for gait phases at both the initial and final levels. PHASEY implements feature representation at multiple layers. It extracts the common differences between different gait phases by contrasting positive and negative samples within each layer independently. The PHASEY model can detect and analyze the detailed view of small changes that take place while a person moves through different phases of walking or running. The PHASEY model extracts features at various levels of the neural network as it processes input data, such as video frames or images of a person walking. This model calculates a contrastive loss for each layer, which indicates the level of accuracy in differentiating the features that represent various phases of the gait cycle. This model performs object detection and classification and focuses on minute, delicate differences within the gait cycle.

The major contribution of this paper lies in highlighting the growing need for automated recognition systems, which have practical applications across diverse domains such as security systems, healthcare, and other critical fields. To address this, we propose a dual-phase model, PHASEY, that integrates both object detection and classification. At the core of the model, CSPDarknet-53 functions as the primary feature extractor, ensuring robust performance. Furthermore, the effectiveness of PHASEY has been evaluated through its ability to accurately localize and classify gait phases, along with their bounding boxes, in an efficient manner.

Related Work

An interesting biometric technique that identifies people by their gait is called gait recognition. Since 2015, deep learning has changed the direction of this field of study by making it possible for it to learn discriminative representations autonomously. Deep learning-based approaches (Sepas-Moghaddam & Etemad, 2023) for recognizing gaits, now dominate the most recent developments in the field and have encouraged practical applications. The authors in reference (Niyogi & Adelson, 1994) proposed the earliest gait recognition system, which was based on a small gait database. Then, the HumanID initiative, funded by the Defense Advanced Research Projects Agency (DARPA) (Sarkar *et al.*, 2005) created the first publicly accessible database for gait identification.

Early Methods and Approaches

According to biomechanical and clinical research conducted in early research, each person has a different gait due to the combined activities of hundreds of limbs, joints, and muscles. In addition, gait can determine the presence of certain sicknesses or emotions. It was demonstrated that a person's gait variability was constant and could not be changed easily, and was difficult to alter. Although the majority of the studies in these early databases were only medically oriented (Lee *et al.*, 2014).

Developments and Milestones of Existing Techniques

In the domain of Gait recognition (Bari & Gavrilova, 2019), developments were made by introducing deep learning methodology (Alharthi *et al.*, 2019) which is a popular machine learning (ML) technique (Kolaghassi *et al.*, 2021), and opens new doors for advanced analysis (Yam & Nixon, 2021) of human motion (Han & Bhanu, 2006). The architecture and operation of biological neural networks serve as an inspiration for deep-structured learning. Deep learning is based on the idea of a multi-layer Artificial Neural Network (ANN) to learn the data representations automatically. Typically, "deep" refers to the number of layers in one of the following types of network structures: Boltzmann Machine (BM), Generative Adversarial Networks (GAN), Convolutional Neural Networks (CNN) with achieved accuracy of 97.1%, Recurrent Neural Networks (RNN), Deep Belief Networks (DBN), Feedforward Deep Networks (FDN), Long-Short Term Memory (LSTM), a specific type of RNN (dos Santos *et al.*, 2022). The LSTM model on its own had lower performance, with accuracies greater than 86.8% and F-scores greater than 86.4% (Narayan *et al.*, 2023).

Structural-Based Approach

Structural and motion models are required in a standard model-based approach to provide the basis for tracking and feature extraction. These models can be two or three-dimensional, although most existing methods are two-dimensional and have demonstrated the potential to provide encouraging recognition results on big databases. The topology or shape of human body components, such as the head, torso, hip, thigh, knee, and ankle, is represented by parameters like length, breadth, and position in a structural model. Stick figures, random forms describing the edges of various body parts, or primitive shapes (cylinders, cones, and blobs) can be used to create this model (Jun *et al.*, 2020).

Appearance-Based Approach

Generally, human Silhouettes are used as raw input data in appearance-based methods. Gait Energy Images (GEI) (Benouis *et al.*, 2016) has been widely used and is the most popular feature that can achieve a high recognition rate and has a low computational cost. GEI-based methods follow a common pipeline, which includes extracting the human silhouettes from videos and then finding out the average and aligning the silhouettes, computing the Gait Energy Image (GEI), and afterward calculating the similarities between two GEIs. Figure 1 gives a diagrammatic representation of gait recognition approaches.

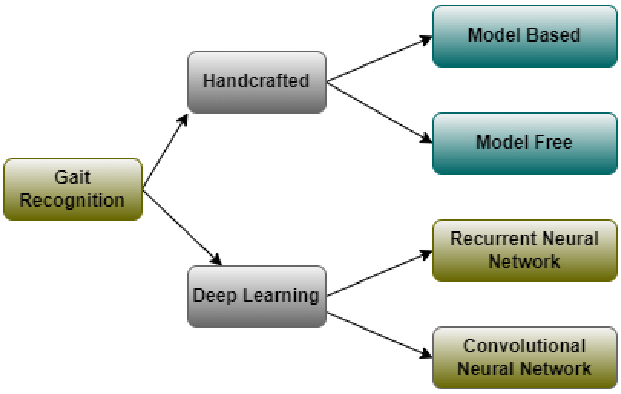

Fig. 1: Diagrammatic Representation of Previous Gait Recognition Approaches

Table 1: Deep learning techniques used for Gait recognition application

DL Technique/ Hybrid Techniques	Applications	Dataset Used	Recognition Rate
Convolutional Neural Network (CNN) (Nguyen *et al.*, 2023)	Feature detection, Biometric authentication	OU-ISIR (Nguyen *et al.*, 2023)	91.5%
Capsule Networks (Narayan *et al.*, 2023)	Tablet Identification	OU-ISIR (Narayan *et al.*, 2023)	74.4%
Auto Encoder (Mehmood *et al.*, 2024)	Functions by combining and separating input characteristics	OU-ISIR (Mehmood *et al.*, 2024)	96%
Recurrent Neural Network (RNN) (Jun *et al.*, 2020)	Speech Recognition and Prediction Problems	CASIA B (dos Santos *et al.*, 2022)	91%
Generative Adversarial Networks (Yu *et al.*, 2017)	Image to Image Translation, Video Prediction	CASIA A and CASIA B (dos Santos *et al.*, 2022)	82%
Combination of Convolutional Neural Networks (CNN) - Recurrent Neural Networks (RNN) (Zhen *et al.*, 2020)	Motion Feature Extraction of sequential temporal data from accelerometers and gyroscopes.	whuGAIT and OU-ISIR (dos Santos *et al.*, 2022)	99.75%
Convolutional Neural Networks (CNN) + Long Short Term Memory (LSTM) (Zhen *et al.*, 2020)	Recognition of severe gait abnormalities	whuGAIT and OU-ISIR (Zhen *et al.*, 2020)	99.75%
Hybrid CNN-Support Vector Machine (SVM) (Liu *et al.*, 2018)	Face Classification, Gender Recognition, and other recognition tasks	CASIA A and CASIA B (dos Santos *et al.*, 2022)	82%
SVM-Bayesian Network (Gupta *et al.*, 2015)	Gait Recognition	OU-ISIR	97.6%

Deep Learning and Hybrid Approaches

Researchers have used and put forth various hybrid deep learning approaches to classify and recognize gait movements. Before deep learning, machine learning methods were used to recognize human Gait, but those methods had certain limitations in using the features that were handcrafted. Table 1 represents some of the deep-learning techniques and algorithms used in Gait.

Materials and Methods

The phases for gait phases recognition are described here.

Feature Extraction and Representation

Three main phases are involved in gait recognition: first is the segmentation of the silhouette, feature extraction, and classification. Initially, from the gait sequence, human silhouettes are identified and separated. In a gait sequence, the background removal approach is frequently used to locate the moving human silhouette. Then, in the feature extraction stage, from the obtained human silhouettes, gait features are extracted by using the hand-crafted approach (Gupta *et al.*, 2015). Research on Gait has been done under controlled and uncontrolled environments, too. Controlled environments are the conditions under which the gait data captured are carefully monitored and standardized. This includes factors like lighting, background, camera angles, and the type of surface on which the subject is walking. The aim is to minimize external variables that could affect the gait patterns, thus allowing for more accurate and consistent data collection (Wan *et al.*, 2019). On the other hand, uncontrolled conditions consist of the real-world settings where the conditions are not regulated and various factors influence gait (Xia *et al.*, 2024). For example, Gait recognition in the wild (Zheng *et al.*, 2022).

Datasets in Gait Recognition

Various datasets have been used in gait recognition, which are described below.

Video-Based Datasets

Chalidabhongse *et al.* (2001) released the UMD dataset in 2001. Two walking-outside datasets make up the UMD dataset. In the larger setup, 55 people walk in T patterns in front of two orthogonally positioned cameras in a parking lot to record gait data. Furthermore, this dataset has four views: frontal, left, right, and rear. The National Institute of Standards and Technology (NIST) Dataset, often known as the HumanID dataset, was released in 2005 by Sarkar *et al.* (2005). The CASIA dataset was released in 2003 by Wang *et al.* (2003), with dataset B being the most commonly utilized. There were 124 people in all, representing 11 distinct perspectives and four different walking speeds. Likewise, in 2004, 2005, 2007, and 2014, CMU, USF, datasets published by Nixon and Carter, and TUM-GAID datasets were published (Hofmann *et al.*, 2014).

Accelerometer-Based Datasets

In 2005, 2007, 2012, 2014, 2014, 2016, Speed Dataset, Motion-Recording-Sensor-Based Dataset, Walking Pattern Dataset, Android phone Google G1 Dataset, large accelerometer-based gait dataset, Human Activities and Postural Transitions, were published respectively (Saha *et al.*, 2024).

Floor-Sensor-Based Datasets

The First Floor-sensor-based dataset was published by Orr and Abowd. Furthermore, from 2004 to 2007, further advancements were made in floor-sensor-based datasets (Dong & Noh, 2024).

Radar-Based Datasets

Otero released the first dataset based on wave radars in 2005. 49 people provided gait data for this dataset. People in this dataset moved closer to and further away from the radar. Wang and Fathy released a dataset in 2011 that contained gait information from a single person. Later, further developments were made by researchers (Collado Pérez, 2023).

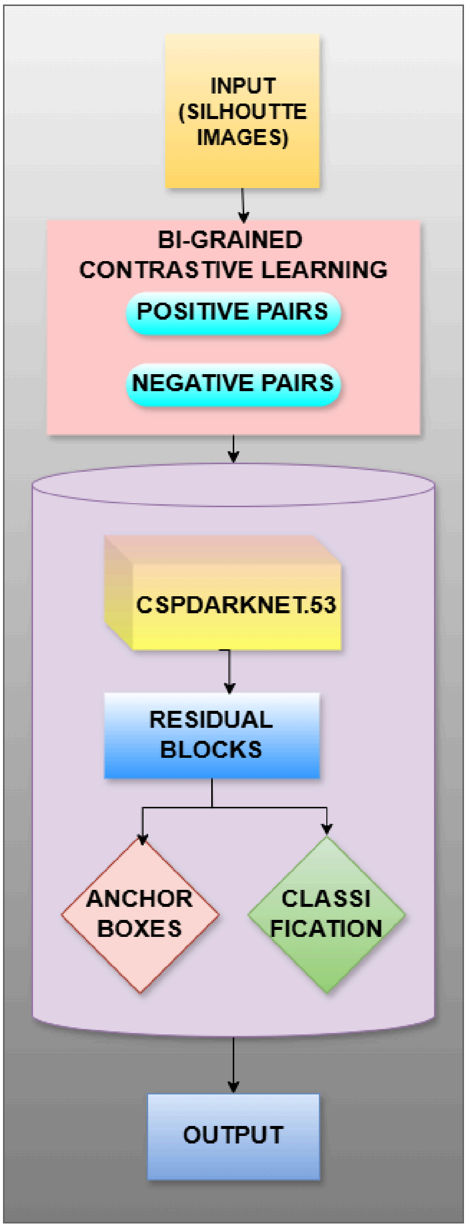

Fig. 2: Proposed methodology: Silhouette input, contrastive learning, CSPDarknet 53, and residual blocks

Contrastive Multiphase GaitNet

Contrastive learning is a technique that is used in vision tasks to enhance performance by using the principle of contrasting the samples against each other to learn attributes that are common between data classes and attributes that differentiate one data class from another (Gao *et al.*, 2023). It has an extraordinary ability to grab and learn different feature representations through a self-supervised methodology by comparing positive and negative samples (Hu *et al.*, 2024). This concept is widely used in many domains, such as natural language processing (Zhang *et al.*, 2022) and important visual tasks. These methods usually treat each instance and its augmented version as a positive pair, while other randomly selected instances are regarded as negative samples. The memory bank is usually used to store the features of the training data (Liu *et al.*, 2023). The overview of the proposed methodology for gait phase localization and classification has been presented in Figure 2.

In this paper, a new model called Contrastive MultiPhase GaitNet (PhaseY) has been proposed, designed to effectively extract and recognize multiple gait phases. The proposed model has expertise in categorizing both Stance and Swing phases. Unlike traditional gait phase recognition models, our approach simultaneously detects multiple gait phases by implementing both coarse-grained and fine-grained levels of gait phase data recognition across different walking patterns. Contrastive learning is essential for improving the model's capacity to differentiate between the various gait cycle phases for gait phase identification. The contrastive learning configuration is used in the PHASEY to achieve high accuracy in segmenting and identifying various gait phases.

Objective of Contrastive Learning

The main motive of contrastive learning in PHASEY is finding representations that maximize similarity between positive pairs (features indicating the same gait phase) and minimize similarity between negative pairs (features representing different gait phases). This method enables the model to get a more thorough understanding of the small differences among various gait phases, which is essential for precise identification.

Positive and Negative Pair Selection

In this model architecture, positive and negative pairs of samples are defined to configure contrastive learning. Positive pairs are feature representations taken from separate frames that are extracted from the same gait phase. For example, a positive pair is formed when features from consecutive frames that depict the "Initial Contact" phase come together. Negative pairs are feature representations that correspond to different gait phases (e.g., "initial contact" vs. "mid-stance"). These pairings have been chosen to illustrate the most challenging cases, in which there are few but significant changes between phases.

Contrastive Loss Function

A contrastive loss function is created to maximize the feature representations acquired by the YOLOv9 model (Ali & Zhang, 2024; Hussain, 2024). This function guides the contrastive learning process in PHASEY. Figure 3 shows the contrastive feature transformation through sampling and contrastive learning.

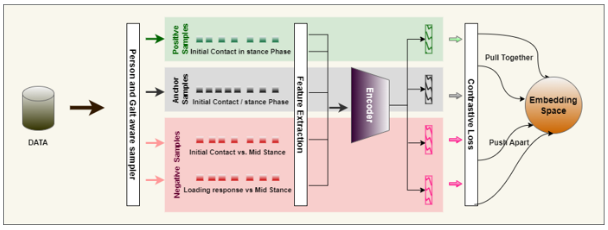
DATA
Person and Gait aware sampler
Positive Samples
Initial Contact in stance Phase
Anchor Samples
Initial Contact / stance Phase
Negative Samples
Initial Contact vs. Mid Stance
Loading response vs Mid Stance
Feature Extraction
Encoder
Contrastive Loss
Pull Together
Push Apart
Embedding Space

Fig. 3: Contrastive Feature Transformation Process through GAITNET Contrastive Learning

Overview of PhaseY model

You Only Look Once (YOLO) is the most common and widely used algorithm (Jiang *et al.*, 2022; Kang *et al.*, 2025). The phases of the PhaseY model have been described below.

Network Architecture of PHASEY Model

The proposed method works on the network architecture that is derived from the recent object detectors like Mask-RCNN (Jia *et al.*, 2024), YOLO (Jia *et al.*, 2024). Typically, the object detectors consist of an end-to-end architecture including a Feature extractor, a proposal generator (Tao *et al.*, 2024), and a mask predictor, respectively. They can locate the objects and classify their semantic categories simultaneously. Limiting the object categories to Stance and Swing phases naturally solves some of the drawbacks of gait recognition methods to a certain extent. Inspired by that, the authors adopt a robust and effective architecture and propose a new bi-grained contrastive multiphase GaitNet scheme customized for the feature extractor, proposal generator, and mask predictor (Zhang *et al.*, 2024).

Bi-Grained Contrastive Learning

To enhance the model's capacity to categorize between various Gait phases, a Bi-grained contrastive learning technique is followed, which includes learning representations at two different levels of granularity, that is coarse-grained and fine-grained (Zhang & Ran, 2024). This method is efficient and works well for tasks where it is important to capture the overall structure as well as the minor details (Huang *et al.*, 2024). The general contrastive GaitNet technique serves as the model for our study, as it allows us to identify the Gait stance and swing phases by examining how the walking phases are related to each other. Our method specifically looks for these relationships on the end-to-end architecture at the coarse-grained (proposal-wise) and fine-grained (pixel-wise) levels (Huang *et al.*, 2024). The items in the feature extractor's features represent different stance and swing leg proposals across various scales, and the items in the mask predictor's features correspond to different pixels in legs in the stance and swing phases. Figure 4 shows the architecture of bi-grained contrastive learning.

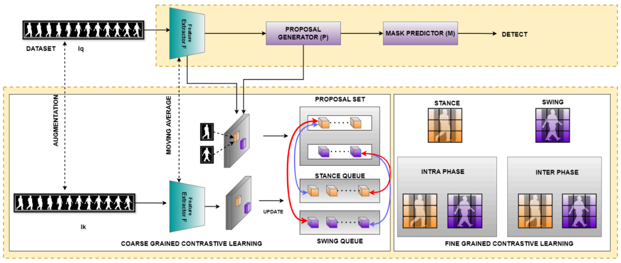

Fig. 4: Architecture of Bi-grained contrastive learning

Coarse-Grained Contrastive GaitNet Model

The coarse-grained contrastive GaitNet technique focuses on capturing the discriminative walking features among positive and negative proposals based on the feature extractors, i.e., putting together the features of proposals in the same category and segregating the ones of different categories (Liu *et al.*, 2024b). Two different samples from different views are generated, and for contrastive learning, a pair of feature extractors is generated. For an input image, where the two feature extractors share the same network architecture with different parameters. After that, the feature extractors get each of these two samples, in turn, to provide proposals at various levels (Yin *et al.*, 2025). Contrastive learning (Ju *et al.*, 2024) is conducted at each layer independently to capture multi-layer proposal-wise diverse walking patterns by matching the proposals with distinct walking steps, as proposals at various levels cannot be directly compared. Next, the contrastive learning configuration and the data view configuration are executed in order.

Dataset Description

A labeled or unlabeled dataset of the gait phases is the primary requirement for estimating and developing human pose estimation and gait recognition approaches (Bansal *et al.*, 2024). This study uses the publicly accessible OU-ISIR dataset that contains walking patterns of people for all age groups. Silhouette images from the OU-ISIR dataset, which is accessed online (Takemura *et al.*, 2018). are used, with 4,163 images taken. The dataset is applicable for many applications, including security and surveillance systems, sports performance optimization, smart wearable technologies, as well as robotics and exoskeleton development. In the PHASEY model, used in this paper for Gait recognition, data pre-processing techniques have been carried out in an effective approach to combat the data scarcity problems and improve the effectiveness of the deep learning model (Parashar *et al.*, 2023). In this paper, the dataset used is taken from OU-ISIR (Takemura *et al.*, 2018). Figure 5 shows the collection of OU-ISIR gait image samples in the form of silhouette images.

Fig. 5: Samples of silhouette images

Preliminary Conceptualization for the PHASEY Model

In this paper, the focus is on achieving accuracy and efficiency in recognizing and classifying the gait sequences (Chao *et al.*, 2021) falling under the stance (Perry & Burnfield, 2024) and swing (Liu *et al.*, 2024a) phases. Gait phase recognition is achieved by using silhouette images. These are binary images, where the subjects are represented specifically in white (Foreground) against black (background). The input images that are used & have dimensions of 224 x 224 pixels in the PHASEY model, although this can be changed depending on the hardware limitations and dataset. Figure 6 shows the layered diagram of the PHASEY model.

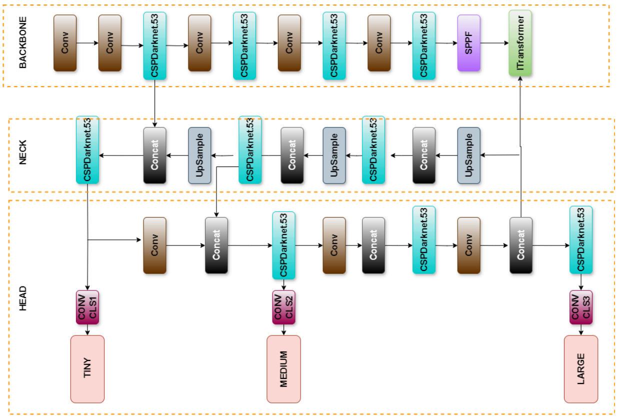

Fig. 6: Architecture Diagram of PHASEY model

Backbone of PHASEY

CSPDARKNET53 backbone has been used (Guo *et al.*, 2025) for extracting the features of input silhouette images. The process can be described as follows.

Feature Extraction and Bounding Box Regression

Feature Extraction (Ray *et al.*, 2024) and Bounding Box Regression (Ming *et al.*, 2024) are important tasks in the context of gait phase detection utilizing the PHASEY model with the CSPDarknet 53 backbone. These activities allow the identification of various gait phases, including the stance and swing phases. The steps in this process can be explained as under.

Feature Mapping

Convolutional Networks are multilevel architectures consisting of multiple stages (Y. Yu *et al.*, 2024). The inputs and outputs in each stage are the sets of arrays called feature maps. A series of convolutional layers is incorporated into our proposed model to process the input images given in the form of silhouettes. The feature extractor used i.e., CSPDARKNET.53 processes the input images and derives the extremely complex features like shapes, edges, and textures. These features proceed from basic elements such as edges and textures in the outer levels to more abstract elements in the inner layers that depict various aspects of the human gait. The layers included in this process make a series of operations like Convolution, Activation, and Normalization.

Semantic Mapping of Convolution:

In the PhaseY model, the first convolution layer consists of 65 filters with dimensions 3 x 3 x 3. This implies that the walking silhouette is processed using 64 different convolutional filters, commonly referred to as kernels. To extract certain characteristics from the silhouette, such as gradients or edges, each filter moves through the sample and applies the convolution process. Since the walking image used here is a picture with three channels (Red, Green, and Blue), so the input shape is 640×640×3. The convolution process involves sliding 3×3×3 filters through the walking silhouette image and applying a dot product between the filter and the local areas of the image. A feature map is generated by the filter, which extracts the specific features in the Image.

Residual Blocks

The model proposed uses a series of residual blocks to capture the high-level features. residual blocks are included in CSPDarknet 53, which have been organized in the following stages.

Step 1: A lot of convolutional and residual blocks at the initial stage.

Step 2: For the process of downsampling the feature maps obtained, deeper blocks with bigger strides and pooling layers are used.

Step 3: More residual blocks are used for even more downsampling.

Step 4: Final residual blocks that consist of aggregated features from previous stages.

Anchor boxes: To predict the bounding boxes our model uses anchor boxes around the detected and specific parts of the body that are involved in the gait. This model uses anchor boxes as reference boxes.

PHASEY Model Pseudocode

```
Step 1: Define Inputs and Outputs
Input: Gait Dataset OU-ISIR: A collection of gait phase
sequences with labeled frames, "Swing", "Stance".
Step 2: Preprocess Data for Pair Selection
Input: Gait Dataset
    for each sequence in D:
        for each frame_i, frame_j in sequence:
            if gait_phase(frame_i) == gait_phase(frame_j):
                add (frame_i, frame_j) to positive_pairs
            else:
                add (frame_i, frame_j) to negative_pairs
Output: Positive and Negative Pairs
Step 3: Coarse-Grained Contrastive Learning
Input: Positive and Negative Pairs
    for each pair in positive_pairs + negative_pairs:
        input1, input2 = pair
        features1 = FeatureExtractor(input1)
        features2 = FeatureExtractor(input2)
        if pair in positive_pairs:
            compute loss(features1, features2,
            label="positive")
        else:
            compute loss(features1, features2,
            label="negative")
Step 4: Fine-Grained Proposal-Based Learning
Input: Frames and Proposals
    for each frame in D:
        proposals =
        ProposalGenerator(FeatureExtractor(frame))
        for proposal in proposals:
            coarse_features =
            extract_features_at_layer(proposal,
            layer="coarse")
            fine_features =
            extract_features_at_layer(proposal,
            layer="fine")
            compute
            proposal_matching_loss(coarse_features,
            fine_features)
Output: Classification Loss, Contrastive Loss
Step 5: Precision Gait Phase Segmentation
Input: Frames
Classification
    for each frame in D:
        predictions =
        MaskPredictor(FeatureExtractor(frame))
        ground_truth = get_ground_truth(frame)
        compute segmentation_loss(predictions,
        ground_truth)
Output: Segmentation Loss for Accurate Gait Phase
Step 6: Optimize Model
Input: Total Loss (Contrastive Loss + Proposal Loss +
Segmentation Loss)
total_loss = contrastive_loss + proposal_matching_loss +
segmentation_loss
optimize_model(total_loss)
Output: PhaseY model capable of accurately classifying gait
phases and segmenting activities.
```

Performance Parameters

For the evaluation of the PHASEY model, several performance parameters have been used. The description of each performance parameter is provided hereunder.

Precision

Calculates the ability of the model to recognize positive pairs of images while minimizing the false positives (Aman *et al.*, 2024). It is computed as the ratio of true positives to the total predicted positives.

Recall

It is called the true positive rate in common. It can be defined as the ratio of actual positives (the total of true positives and false negatives) to true positives (properly anticipated positive samples (Lai *et al.*, 2020). Recall measures the scale up to which this model captures all true positives without missing any of them.

Intersection over Union (IoU)

It evaluates the accuracy of the model in detecting objects and in determining the accuracy with which a bounding box overlaps with the original bounding box. It assists in recognizing the phases of the gait cycle, stance, and swing. It is calculated as the ratio of the area of overlap to the area of union (Lanshammar, 1982). The values of IoU range from 0 to 1. Zero means no overlap between the predicted and actual bounding boxes, and it means that there is perfect overlap.

Results

All experiments were performed on a computer with an Intel(R) Core (TM) i5-10310U CPU running at 1.70 GHz and 2.21 GHz, 16GB of RAM, and a 64-bit operating system. Python version 3.7 was used with the Pytorch platform on a 64-bit Windows 11 Pro system.

Dataset Preparation and Preprocessing

In computer vision fields and Image processing, this process has been observed for a long time and includes foreground and background segmentation (Wang *et al.*, 2003). A binary silhouette is used to make the suggested solution insensitive to variations in the color and texture of clothing (Ji *et al.*, 2024). Figure 7 represents the training dataset consisting of walking silhouettes of people from both the gait stance and swing phases. The data-preprocessing technique followed is described below.

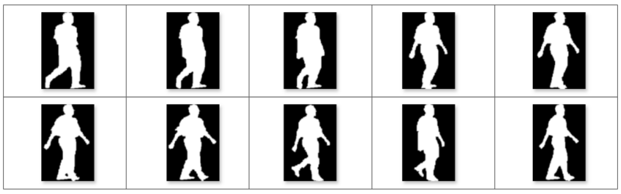

Fig. 7: Samples of Silhouette images

Silhouette Extraction

In this paper, the initial and unprocessed gait sequences from OU-ISIR are first processed to extract silhouette Images. Every image frame is transformed into a binary silhouette picture, where the human figure is represented in white pixels with the backdrop represented by black pixels. Concerning the dataset available online, 4016 silhouette images were taken from the website. The dataset consists of people walking on the ground recorded by two cameras at 30 frames per second at 640 by 480. The datasets are made available with a size normalization of 224 by 224 pixels. Figure 8 shows the silhouette images before and after implementing the extraction process.

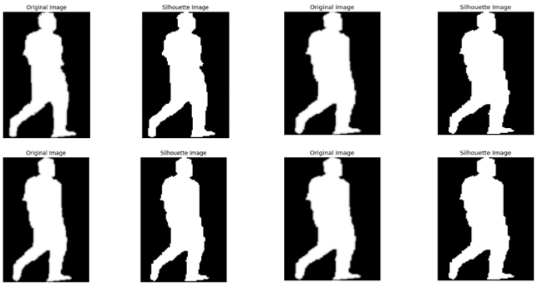

Fig. 8: Pre-processed Silhouette Images

Image Resizing

The Silhouette images extracted from real human walking sequences are processed first. The silhouette preprocessing process is then implemented on the extracted silhouette sequences. For our PHASEY model, image consistency was maintained by resizing the images to 224×224 pixels. It made sure that the silhouette images were fed properly and suited the dimensions of our proposed model without any loss of information about gait characteristics. After image resizing, the dimensions of images on a consistent scale have been changed through the normalization technique (Jlassi & Dixon, 2024).

Image Augmentation

Image augmentation is an efficient way to increase the boosting speed of the dataset. As the number of samples in the dataset is too small it can decrease the training accuracy of the PhaseY model. To increase the speed of the model's accuracy through data augmentation techniques. The data augmentation techniques (Chandrasekaran *et al.*, 2024) such as flipping, cropping, and rotation help to increase the dataset size. After data augmentation, a total of 17,500 images were obtained, which were divided into training sets and testing sets in the ratio of 7:3, thus preventing the detection model from overfitting. 12250 images were kept for training purposes, and 5250 images were used for testing. The augmented dataset in the form of image samples has been presented in Figure 9.

Dataset Diversity

The dataset derived from the OU-ISIR collection was initially imbalanced. To correct this problem, image augmentation techniques were applied to it, which also boosted the performance of the PhaseY model. After augmentation, the dataset was divided into the training and testing subsets. The process generated a total of 17,500 images, that are five times more than the original dataset. Moreover, 317 images were annotated to delineate the pictograms for each stance and swing phase class. Table 2 summarizes the distribution of training and testing images across classes after augmentation. Distribution of training and testing images across classes after augmentation, where the number of classes 0-3 represents the stance phase and 4-7 shows the swing phase classes.

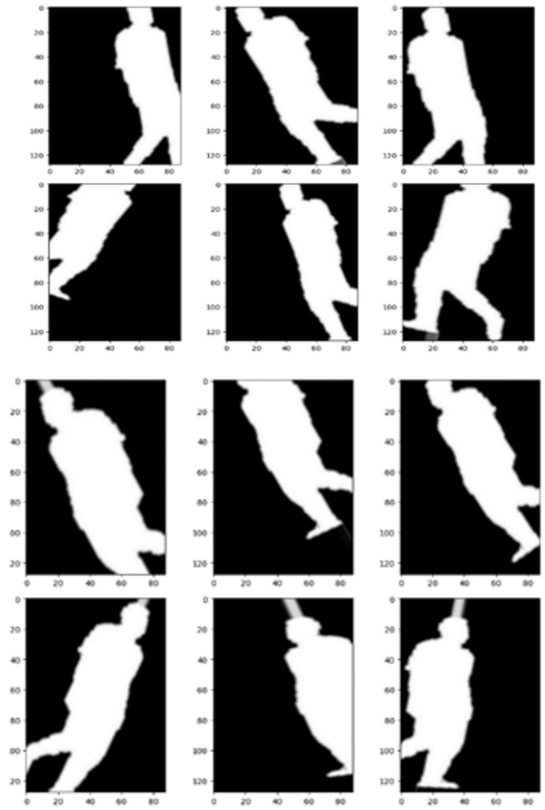

Fig. 9: Augmented Images

>**Table 2:** Images per class distribution

Total number of images for training								
Class number	0	1	2	3	4	5	6	7
Pictogram	IC	LR	MS	TS	PS	TO	MS	TS
No of images	1463	1517	1312	1711	960	1687	1690	1593
Total number of images for testing								
Class number	0	1	2	3	4	5	6	7
Pictogram	IC	LR	MS	TS	PS	TO	MS	TS
No of images	650	635	646	639	590	710	688	692

Prediction Results of PHASEY

The gait phase recognition is done by using the PHASEY model, which has three core components: Backbone, Neck, and Head. The contribution of each component has been discussed in this section.

Output as a Backbone

From the input silhouette images, the feature extraction is done using CSPDarknet 53. The architecture consists of five convolutional blocks and four feature extractors. Deep feature extraction is done by capturing the high-level features important for recognizing gait features (Marimon *et al.*, 2024). The initial stage convolutional layers process the input silhouettes, and the convolutional layers at the later stages identify the gait phases, like stance and swing. Using CSPNet (Cross-Stage Partial Networks) optimizes gradient flow, enhancing model learning while reducing computing costs (Han *et al.*, 2024). By minimizing repetition, this structure ensures that important features are maintained throughout the levels. The backbone consists of Spatial Pyramid Pooling SPPF and a Transformer block (Peng *et al.*, 2024). The transformer improves the important spatial features (Bilal *et al.*, 2024), which is important for differentiating between minute differences in gait phases. Table 3 shows the output of the backbone in terms of its parameters.

Table 3: Parameter details of PHASEY

Layer (type)	Output Shape	Param #
Input layer (InputLayer)	(None, 224, 224, 3)	0
zero_padding2d (ZeroPadding2D)	(None, 230, 230, 3)	0
conv1_conv (Conv2D)	(None, 112, 112, 64)	9,408
conv1_bn (BatchNormalization)	(None, 112, 112, 64)	256
conv1_relu (Activation)	(None, 112, 112, 64)	0
zero_padding2d_1 (ZeroPadding2D)	(None, 114, 114, 64)	0
pool1 (MaxPooling2D)	(None, 56, 56, 64)	0
conv2_block1_0_bn	(None, 56, 56, 64)	256
conv2_block1_0_relu (Activation)	(None, 56, 56, 64)	0
conv2_block1_1_conv (Conv2D)	(None, 56, 56, 128)	8,192
conv2_block1_1_bn (BatchNormalization)	(None, 56, 56, 128)	512
conv2_block1_1_relu (Activation)	(None, 56, 56, 128)	0

Total params: 8,062,504 (30.76 MB)
Trainable params: 7,978,856 (30.44 MB)
Non-trainable params: 83,648 (326.75 KB)

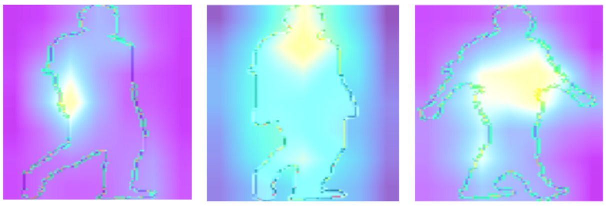

Fig. 10: Neck output of PHASEY Model

Output of the Neck

The multi-scale feature maps are combined at the neck of the architecture. This section contains up-sampling layers, concatenation blocks, and additional CSPDarknet53 modules. Figure 10 shows the output from this layer.

Detection of Gait Phases

Three types of detections are carried out in the three output layers that make up the Head. Figure 10 shows the detection accuracy of different YOLO models for gait phase recognition.

Figure 11 presents the detection performance achieved by YOLOv3, YOLOv4, YOLOv5, and the proposed YOLOv9-based PHASEY model. As shown in the figure, the PHASEY model outperforms earlier YOLO versions, achieving higher detection accuracy for both stance and swing phases of the gait cycle. This graphical comparison further validates the effectiveness of integrating contrastive learning and the CSPDarknet 53 backbone within the PHASEY framework.

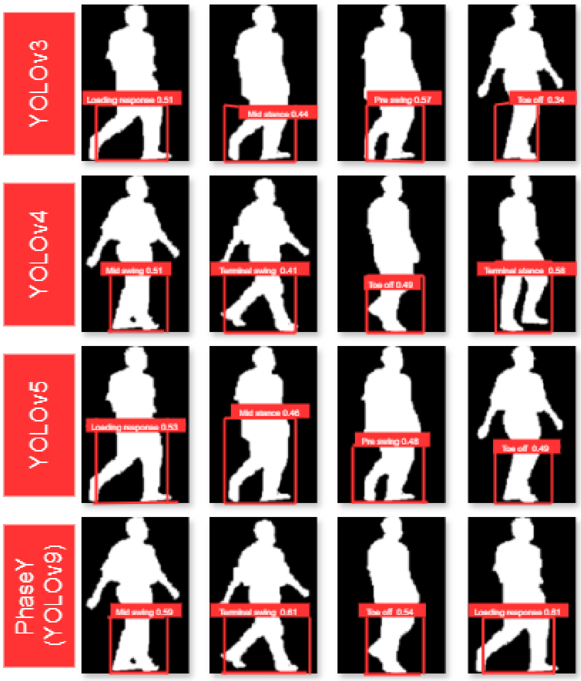

Fig. 11: Detection accuracy analysis of YOLOv3, YOLOv4, YOLOv5, and YOLOv9 for swing and stance gait phase recognition

Performance Evaluation

In this paper, the experimental platform used is PyTorch. Following data augmentation, a total of 17,500 walking images were obtained. These were split into training and testing sets at a 7:3 ratio to avoid the overfitting of the detection model. A total of 5250 images were used for testing, while 12250 images were retained for training. Even the total loss calculated during each phase of the PhaseY model with different iterations has been presented in Figure 12.

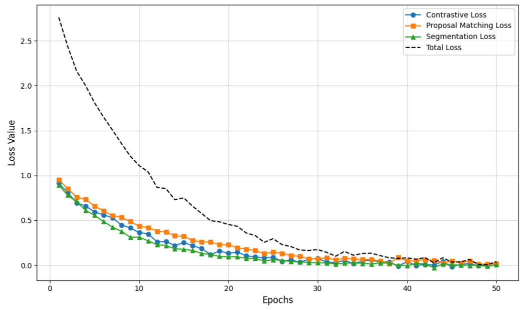

Fig. 12: Iteration-based total loss fluctuation for PhaseY model

Discussion

In this experiment, the ability to recognize the gait phases (stance and swing phases) in silhouette images is evaluated utilizing the CSPDarknet 53 backbone of several versions of the YOLO object recognition models (YOLOv3, YOLOv4, YOLOv5, YOLOv9). About 12,250 of the 17,500 images in the dataset are utilized for training, while the remaining 5,250 are used for testing. Results are based on important measures such as Intersection over Union (IoU), accuracy, precision, recall, and 40 model training epochs. Table 4 shows the Training and testing accuracy of different YOLO models.

Table 4: Training and testing accuracy of different YOLO models

Model	Tr. Acc	Testing Acc.	Prec.	Recall	IoU	IT
YOLOv3	87.5%	84.2 %	85.4%	83.9 %	78.5%	30
YOLOv4	91.3%	88.6%	89.1%	88.0%	82.9%	32
YOLOv5	93.6%	90.8%	91.5%	90.3%	85.5%	35
PHASEY	94.8%	92.5%	93.1%	91.9%	87.8%	38

Table 5: State-of-art comparison

References	Models	Dataset	Accuracy
A. S. M. H. Bari et.al	DLNN - Deep Learning Neural Network	UPVC gait dataset- images from 30 subjects, 55-120 frames per sequence Kinetic Gait Biometry Dataset- 164 subjects, 500-600 frames.	UPVC Dataset- 85.30% Kinetic Gait Biometry Dataset- 88.8%
J. Han et.al	Naïve Bayes	USF, HumanID Database, 122 Individuals, 122 Sequences	91%
M. Benouis et.al	DBN- Deep Belief Network	CASIA-B Gait Dataset Gait Sequences of 124 subjects, Total sequences- 13,640	90.8%
T. Zhen et.al	CNN+Autoencoder	Private Gait Dataset - 16 subjects, 144 sequences	91.2%
PHASE Y	Bi-contrastive Learning based YOLOv9 model	OU-ISIR	94.8%

State-of-the-Art Methods: A Comparative Evaluation

The comparison of the proposed PhaseY model with previous state-of-the-art methods for gait phase recognition with different datasets has been presented in Table 5.

In gait phase identification, particularly in the recognition of stance and swing phases, the IoU metric is often used to evaluate the accuracy of segmentation in time-series data or spatial data. Misclassification loss is a critical aspect of evaluating model performance, as it shows how often the model incorrectly classifies the stance and swing phases. Different IoU thresholds can influence how strictly the model's predictions are compared to ground truth, thus affecting misclassification loss. The misclassification error at different IoU values has been presented in Figure 13. Moreover, the confusion matrix obtained by the PHASEY model with the SGD optimizer during the validation step on OU-ISIR data is shown in Figure 14.

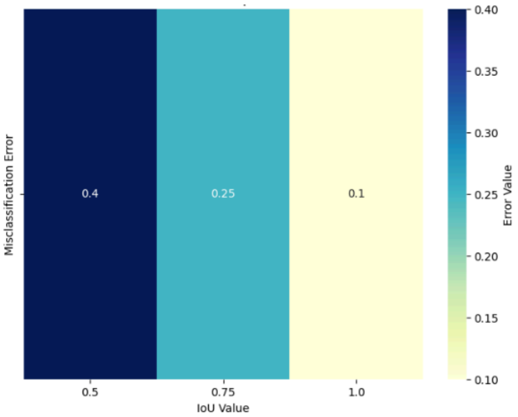

Fig. 13: Misclassification errors at different IoU values

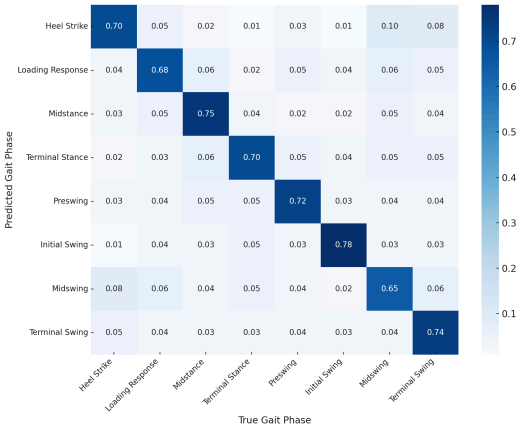

Fig. 14: Confusion matrix showing misclassification errors during validation for gait stance and swing phase classification using the PHASEY model

Practical Implications of the PHASEY Model

The PHASEY model aims to enhance the understanding and precision of gait phase recognition.

The practical implications of the PHASEY model are described below.

- Improving Clinical Diagnoses: By accurately segmenting and classifying gait phases (Xu *et al.*, 2024; Ranjan *et al.*, 2025). The PHASEY model helps to recognize gait abnormalities caused by conditions such as Parkinson's disease (Burtscher *et al.*, 2024), stroke, or musculoskeletal disorders that allow for more targeted treatment plans.
- Advancing Rehabilitation Techniques: The model provides real-time insights into gait dynamics, enabling tailored rehabilitation programs for individuals recovering from surgeries or injuries, thus accelerating recovery processes.
- Enhancing Athletic Performance: PHASEY supports the analysis of athletes' stance and swing phases, optimizing running or walking techniques, reducing the risk of injury, and improving overall performance through data-driven training adjustments.
- Wider Accessibility for Remote Monitoring: The implementation of PHASEY in wearable devices or remote monitoring systems allows clinicians to analyze gait activities without requiring patients to visit healthcare facilities, facilitating broader access to healthcare.
- Supporting Robotics and Prosthetics Development: Insights from PHASEY can improve robotic systems and prosthetic device design by providing precise data on human gait mechanics, leading to more natural and efficient movement solutions.

Conclusion

In this study, the OU-ISIR gait database has been used, which is a secondary source to work on the reorganization of gait swing and stance phases. Efficient pre-processing techniques have been applied, like silhouette extraction, normalization, and data augmentation, to deal with issues like insufficient data. For Gait stance and swing phase recognition, this study used the CSPDarknet 53 backbone to implement and test several YOLO models (YOLOv3, YOLOv4, YOLOv5, and YOLOv9) for gait phase detection. According to the findings, YOLOv9 with the CSPDarknet 53 backbone performed noticeably better than previous iterations in terms of recall, accuracy, precision, IoU, and inference speed. With 92.5% testing accuracy, 93.1% precision, 91.9% recall, and 87.8% IoU, YOLOv9 is the most effective and dependable model for identifying gait phases. Furthermore, YOLOv9 is appropriate for real-time applications because it has an enhanced feature extraction architecture that maintains high accuracy with speed, which is important for real-time applications. The outcomes of this study find their implementations in various domains like security and surveillance, rehabilitation and medical diagnosis, and in the sports field. In crowded places, where traditional biometrics, e.g., Facial recognition, is not feasible, this model can be integrated into non-interfering gait-based biometric systems to identify individuals. It can be an advantage for long-range identification in smart cities and airports. In the medical field, doctors can benefit by launching the model to monitor abnormalities in gait caused by any neurological disorder or stroke recovery. Moreover, this model, when implemented in IoT devices, can assist in tracking the rehabilitation progress of patients recovering from surgeries or major injuries. Last but not least, this model can assist in optimizing the performance of sportspersons by helping analyze the inefficiencies in their running and walking styles. Besides these advantages, the model has certain limitations too. If we talk about the gait patterns that are unfamiliar or not common, recognizing those walking patterns remains a challenge, particularly when the movements are unusual. Although the dataset used is quite diverse, but may lack in capturing global variations in gait patterns that environmental and physical factors might influence. This study can be implemented to capture the dynamic nature of gait phases using Long Short-Term Memory (LSTM) models, and Temporal Convolutional Networks (TCNs) to improve the model's understanding between stance and swing phases.

Data Availability Statement

The public dataset has been taken from the Institute of Scientific and Industrial Research, Osaka University, for providing access to the dataset. (OU-ISIR).

Acknowledgment

The authors would like to thank the Department of Intelligent Media, The Institute of Scientific and Industrial Research, Osaka University, for providing access to the dataset. (OU-ISIR).

Author’s Contributions

Urvashi: Drafted and wrote the manuscript.

Deepak Kumar: Developed and wrote the methodology; reviewed the manuscript.

Vinay Kukreja: Reviewed the manuscript; prepared figures and tables.

Ayush Dogra: Prepared figures and tables.

References

Alharthi, A. S., Yunas, S. U., & Ozanyan, K. B. (2019). Deep Learning for Monitoring of Human Gait: A Review. *IEEE Sensors Journal*, *19*(21), 9575-9591. https://doi.org/10.1109/jsen.2019.2928777

Ali, M. L., & Zhang, Z. (2024). The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in object detection. *Computers*, *13*(12), 336. https://doi.org/10.3390/computers13120336

Aman, N., Islam, M. R., Ahamed, M. F., & Ahsan, M. (2024). Performance Evaluation of Various Deep Learning Models in Gait Recognition Using the CASIA-B Dataset. *Technologies*, *12*(12), 264. https://doi.org/10.3390/technologies12120264

Bansal, A., Jain, A., & Bharadwaj, S. (2024). An exploration of gait datasets and their implications. *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, 1-6. https://doi.org/10.1109/SCEECS61402.2024.10482347

Bari, A. S. M. H., & Gavrilova, M. L. (2019). Artificial Neural Network Based Gait Recognition Using Kinect Sensor. *IEEE Access*, *7*, 162708-162722. https://doi.org/10.1109/access.2019.2952065

Benouis, M., Senouci, M., Tlemsani, R., & Mostefai, L. (2016). Gait recognition based on model-based methods and deep belief networks. *International Journal of Biometrics*, *8*(3/4), 237. https://doi.org/10.1504/ijbm.2016.082598

Bilal, M., Jianbiao, H., Mushtaq, H., Asim, M., Ali, G., & ElAffendi, M. (2024). Gaitstar: Spatial--temporal attention-based feature-reweighting architecture for human gait recognition. *Mathematics*, *12*(16), 2458. https://doi.org/10.3390/math12162458

Boisvert, J., Shu, C., Wuhrer, S., & Xi, P. (2013). Three-dimensional human shape inference from silhouettes: reconstruction and validation. *Machine Vision and Applications*, *24*(1), 145-157. https://doi.org/10.1007/s00138-011-0353-9

Burtscher, J., Moraud, E. M., Malatesta, D., Millet, G. P., Bally, J. F., & Patoz, A. (2024). Exercise and gait/movement analyses in treatment and diagnosis of Parkinson's Disease. *Ageing Research Reviews*, *93*, 102147. https://doi.org/10.1016/j.arr.2023.102147

Chalidabhongse, T., Kruger, V., & Chellappa, R. (2001). *The UMD database for human identification at a distance*. Technical Report, University of Maryland

Chandrasekaran, M., Francik, J., & Makris, D. (2024). *Enhancing gait recognition: data augmentation via physics-based biomechanical simulation*. 170-188. https://doi.org/10.1007/978-3-031-91575-8_11

Chao, H., Wang, K., He, Y., Zhang, J., & Feng, J. (2021). GaitSet: Cross-view gait recognition through utilizing gait as a deep set. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(7), 3467-3478. https://doi.org/10.1109/TPAMI.2021.3057879

Collado Pérez, M. (2023). *Evaluation and analysis of open-set radar-based human gait recognition performance with an adapted radar dataset*.

Dong, Y., & Noh, H. Y. (2024). Ubiquitous Gait Analysis through Footstep-Induced Floor Vibrations. *Sensors*, *24*(8), 2496. https://doi.org/10.3390/s24082496

dos Santos, C. F. G., Oliveira, D. de S., Passos, L. A., Pires, R. G., Santos, D. F. S., Valem, L. P., Moreira, T. P., Santana, M. C. S., Roder, M., Papa, J. P., & Colombo, D. (2022). Gait Recognition Based on Deep Learning: A Survey. *ACM Computing Surveys*, *55*(2), 1-34. https://doi.org/10.1145/3490235

Gao, B., Zhao, X., & Zhao, H. (2023). An Active and Contrastive Learning Framework for Fine-Grained Off-Road Semantic Segmentation. *IEEE Transactions on Intelligent Transportation Systems*, *24*(1), 564-579. https://doi.org/10.1109/tits.2022.3218403

Guo, X., Jiang, F., Chen, Q., Wang, Y., Sha, K., & Chen, J. (2025). Deep learning-enhanced environment perception for autonomous driving: MDNet with CSP-DarkNet53. *Pattern Recognition*, *160*, 111174. https://doi.org/10.1016/j.patcog.2024.111174

Gupta, A., Prasad, P. W. C., Alsadoon, A., & Bajaj, K. (2015). *Hybrid method for Gait recognition using SVM and Baysian Network*. 89-94. https://doi.org/10.1109/iwcia.2015.7449468

Han, J., & Bhanu, B. (2006). Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(2), 316-322. https://doi.org/10.1109/tpami.2006.38

Han, N., Ryu, S. J., & Nam, Y. (2024). Real-time moving object tracking on smartphone using cradle head servo motor. *Sensors*, *24*(4), 1265. https://doi.org/10.3390/s24041265

Hayfron-Acquah, J. B., Nixon, M. S., & Carter, J. N. (2003). Automatic gait recognition by symmetry analysis. *Pattern Recognition Letters*, *24*(12), 2175-2183. https://doi.org/10.1016/S0167-8655(03)00086-2

Hofmann, M., Geiger, J., Bachmann, S., Schuller, B., & Rigoll, G. (2014). The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits. *Journal of Visual Communication and Image Representation*, *25*(1), 195-206. https://doi.org/10.1016/j.jvcir.2013.02.006

Hu, H., Wang, X., Zhang, Y., Chen, Q., & Guan, Q. (2024). A comprehensive survey on contrastive learning. *Neurocomputing*, *610*, 128645. https://doi.org/10.1016/j.neucom.2024.128645

Huang, D., Deng, X., Chen, D.-H., Wen, Z., Sun, W., Wang, C.-D., & Lai, J.-H. (2024). Deep clustering with hybrid-grained contrastive and discriminative learning. *IEEE Transactions on Circuits and Systems for Video Technology*, *34*(10), 9472-9483. https://doi.org/10.1109/TCSVT.2024.3399596

Huang, P. S., Harris, C. J., & Nixon, M. S. (1999). Human Gait Recognition in Canonical Space Using Temporal Templates. *IEE Proceedings - Vision Image and Signal Processing*, *146*(2), 93-100. https://doi.org/10.1049/ip-vis:19990187

Hussain, M. (2024). YOLOv1 to v8: Unveiling Each Variant-A Comprehensive Review of YOLO. *IEEE Access*, *12*, 42816-42833. https://doi.org/10.1109/ACCESS.2024.3378568

Ji, B., Chen, X., Yang, W., & Zhu, F. (2024). *Boosting robustness of silhouette-based gait recognition against adversarial attacks*. 72-84. https://doi.org/10.1007/978-981-97-5594-3_7

Jia, Z., Zhang, Y., & Yang, H. (2024). *Research on High-Precision object detection and instance segmentation using Mask-RCNN*. 1050-1055. https://doi.org/10.1109/ICCASIT62299.2024.1082 7917

Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A Review of Yolo Algorithm Developments. *Procedia Computer Science*, *199*, 1066-1073. https://doi.org/10.1016/j.procs.2022.01.135

Jlassi, O., & Dixon, P. C. (2024). The effect of time normalization and biomechanical signal processing techniques of ground reaction force curves on deep-learning model performance. *Journal of Biomechanics*, *168*, 112116. https://doi.org/10.1016/j.jbiomech.2024.112116

Ju, W., Wang, Y., Qin, Y., Mao, Z., Xiao, Z., Luo, J., Yang, J., Gu, Y., Wang, D., Long, Q., Yi, S., Luo, X., & Zhang, M. (2024). Towards graph contrastive learning: A survey and beyond. *ArXiv*, arXiv:2405.11868.

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. https://doi.org/10.1109/access.2020.2967845

Kang, S., Hu, Z., Liu, L., Zhang, K., & Cao, Z. (2025). Object detection YOLO algorithms and their industrial applications: Overview and comparative analysis. *Electronics*, *14*(6), 1104. https://doi.org/10.3390/electronics14061104

Kececi, A., Yildirak, A., Ozyazici, K., Ayluctarhan, G., Agbulut, O., & Zincir, I. (2020). Implementation of machine learning algorithms for gait recognition. *Engineering Science and Technology, an International Journal*, *23*(4), 931-937. https://doi.org/10.1016/j.jestch.2020.01.005

Kolaghassi, R., Al-Hares, M. K., & Sirlantzis, K. (2021). Systematic Review of Intelligent Algorithms in Gait Analysis and Prediction for Lower Limb Robotic Systems. *IEEE Access*, *9*, 113788-113812. https://doi.org/10.1109/access.2021.3104464

Kusakunniran, W. (2020). Review of gait recognition approaches and their challenges on view changes. *IET Biometrics*, *9*(6), 238-250. https://doi.org/10.1049/iet-bmt.2020.0103

Lai, B., Sasaki, J. E., Jeng, B., Cederberg, K. L., Bamman, M. M., & Motl, R. W. (2020). Accuracy and precision of three consumer-grade motion sensors during overground and treadmill walking in people with Parkinson disease: cross-sectional comparative study. *JMIR Rehabilitation and Assistive Technologies*, *7*(1), e14059. https://doi.org/10.2196/14059

Lanshammar, H. (1982). On practical evaluation of differentiation techniques for human gait analysis. *Journal of Biomechanics*, *15*(2), 99-105. https://doi.org/10.1016/0021-9290(82)90041-0

Lee, T. K. M., Belkhatir, M., & Sanei, S. (2014). A comprehensive review of past and present vision-based techniques for gait recognition. *Multimedia Tools and Applications*, *72*(3), 2833-2869. https://doi.org/10.1007/s11042-013-1574-x

Liao, R., Yu, S., An, W., & Huang, Y. (2020). A model-based gait recognition method with body pose and human prior knowledge. *Pattern Recognition*, *98*, 107069. https://doi.org/10.1016/j.patcog.2019.107069

Liu, J., Tan, X., Jia, X., Li, T., & Li, W. (2024a). A gait phase recognition method for obstacle crossing based on multi-sensor fusion. *Sensors and Actuators A: Physical*, *376*, 115645. https://doi.org/10.1016/j.sna.2024.115645

Liu, J., Wang, W., Yi, B., Shen, X., & Zhang, H. (2024b). Contrastive multi-interest graph attention network for knowledge-aware recommendation. *Expert Systems with Applications*, *255*, 124748. https://doi.org/10.1016/j.eswa.2024.124748

Liu, T., Ye, X., & Sun, B. (2018). *Combining Convolutional Neural Network and Support Vector Machine for Gait-based Gender Recognition*. 3477-3481. https://doi.org/10.1109/cac.2018.8623118

Liu, Z., Alavi, A., Li, M., & Zhang, X. (2023). Self-supervised contrastive learning for medical time series: A systematic review. *Sensors*, *23*(9), 4221. https://doi.org/10.3390/s23094221

Marimon, X., Mengual, I., López-de-Celis, C., Portela, A., Rodriguez-Sanz, J., Herráez, I. A., & Pérez-Bellmunt, A. (2024). Kinematic analysis of human gait in healthy young adults using IMU sensors: exploring relevant machine learning features for clinical applications. *Bioengineering*, *11*(2), 105. https://doi.org/10.3390/bioengineering11020105

Mehmood, A., Amin, J., Sharif, M., Kadry, S., & Kim, J. (2024). Stacked-gait: A human gait recognition scheme based on stacked autoencoders. *Plos One*, *19*(10), 0310887. https://doi.org/10.1371/journal.pone.0310887

Ming, Q., Miao, L., Zhou, Z., Song, J., & Pizurica, A. (2024). Gradient calibration loss for fast and accurate oriented bounding box regression. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1-15. https://doi.org/10.1109/TGRS.2024.3367294

Narayan, V., Awasthi, S., Fatima, N., Faiz, M., & Srivastava, S. (2023). *Deep Learning Approaches for Human Gait Recognition: A Review*. 763-768. https://doi.org/10.1109/aisc56616.2023.10085665

Nguyen, K., Nguyen, V. V., Mai, N. T., Nguyen, A. H., & Nguyen, A. V. (2023). Human Gait Analysis Using Hybrid Convolutional Neural Networks. *Journal of Computer Science and Cybernetics*, *39*(2), 125-142. https://doi.org/10.15625/1813-9663/18067

Niyogi, & Adelson. (1994). *Analyzing and recognizing walking figures in XYT*. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA. https://doi.org/10.1109/cvpr.1994.323868

Parashar, A., Parashar, A., Ding, W., Shabaz, M., & Rida, I. (2023). Data preprocessing and feature selection techniques in gait recognition: A comparative study of machine learning and deep learning approaches. *Pattern Recognition Letters*, *172*, 65-73. https://doi.org/10.1016/j.patrec.2023.05.021

Peng, G., Li, R., Li, A., & Wang, Y. (2024). Synthesis Pyramid Pooling: A Strong Pooling Method for Gait Recognition in the Wild. *IEEE Signal Processing Letters*, *31*, 3159-3163. https://doi.org/10.1109/LSP.2024.3470749

Perry, J., & Burnfield, J. M. (2024). Phases of gait. *Gait Analysis*, 9-16.

Ranjan, R., Ahmedt-Aristizabal, D., Armin, M. A., & Kim, J. (2025). Computer Vision for Clinical Gait Analysis: A Gait Abnormality Video Dataset. *IEEE Access*, *13*, 45321-45339. https://doi.org/10.1109/ACCESS.2025.3545787

Ray, A., Uddin, M. Z., Hasan, K., Melody, Z. R., Sarker, P. K., & Ahad, M. A. R. (2024). Multi-Biometric Feature Extraction from Multiple Pose Estimation Algorithms for Cross-View Gait Recognition. *Sensors*, *24*(23), 7669. https://doi.org/10.3390/s24237669

Rida, I. (2019). Towards Human Body-Part Learning for Model-Free Gait Recognition. *ArXiv:1904.01620*.

Rida, I., Almaadeed, N., & Almaadeed, S. (2019). Robust gait recognition: a comprehensive survey. *IET Biometrics*, *8*(1), 14-28. https://doi.org/10.1049/iet-bmt.2018.5063

Roy, A., Sural, S., & Mukherjee, J. (2012). Gait recognition using Pose Kinematics and Pose Energy Image. *Signal Processing*, *92*(3), 780-792. https://doi.org/10.1016/j.sigpro.2011.09.022

Saha, U., Saha, S., Kabir, M. T., Fattah, S. A., & Saquib, M. (2024). Decoding human activities: Analyzing wearable accelerometer and gyroscope data for activity recognition. *IEEE Sensors Letters*, *8*(8), 1-4. https://doi.org/10.1109/LSENS.2024.3423340

Sarkar, S., Phillips, P. J., Liu, Z., Vega, I. R., Grother, P., & Bowyer, K. W. (2005). The humanID gait challenge problem: data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(2), 162-177. https://doi.org/10.1109/tpami.2005.39

Sepas-Moghaddam, A., & Etemad, A. (2023). Deep Gait Recognition: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 264-284. https://doi.org/10.1109/tpami.2022.3151865

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. on Computer Vision and Applications*, *10*(4), 1-14.

Tao, H., Zheng, Y., Wang, Y., Qiu, J., & Stojanovic, V. (2024). Enhanced feature extraction YOLO industrial small object detection algorithm based on receptive-field attention and multi-scale features. *Measurement Science and Technology*, *35*(10), 105023. https://doi.org/10.1088/1361-6501/ad633d

Umberger, B. R. (2010). Stance and swing phase costs in human walking. *Journal of The Royal Society Interface*, *7*(50), 1329-1340. https://doi.org/10.1098/rsif.2010.0084

Wan, C., Wang, L., & Phoha, V. V. (2019). A Survey on Gait Recognition. *ACM Computing Surveys*, *51*(5), 1-35. https://doi.org/10.1145/3230633

Wang, L., Tan, T., Ning, H., & Hu, W. (2003). Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(12), 1505-1518. https://doi.org/10.1109/tpami.2003.1251144

Xia, Y., Sun, H., Zhang, B., Xu, Y., & Ye, Q. (2024). Prediction of freezing of gait based on self-supervised pretraining via contrastive learning. *Biomedical Signal Processing and Control*, *89*, 105765. https://doi.org/10.1016/j.bspc.2023.105765

Xu, D., Zhou, H., Quan, W., Jiang, X., Liang, M., Li, S., Ugbolue, U. C., Baker, J. S., Gusztav, F., Ma, X., Chen, L., & Gu, Y. (2024). A new method proposed for realizing human gait pattern recognition: Inspirations for the application of sports and clinical gait analysis. *Gait & Posture*, *107*, 293-305. https://doi.org/10.1016/j.gaitpost.2023.10.019

Yam, C.-Y., & Nixon, M. S. (2021). Model-based Gait Recognition. *Encyclopedia of Biometrics*, 1082-1088.

Yin, T., Wang, J., Zhao, Y., Wang, H., Ma, Y., & Liu, M. (2025). Fine-grained adaptive contrastive learning for unsupervised feature extraction. *Neurocomputing*, *618*, 129014. https://doi.org/10.1016/j.neucom.2024.129014

Yu, S., Chen, H., Reyes, E. B. G., & Poh, N. (2017). *GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Networks*. 30-37. https://doi.org/10.1109/cvprw.2017.80

Yu, Y., He, Y., Karimi, H. R., Gelman, L., & Cetin, A. E. (2024). A two-stage importance-aware subgraph convolutional network based on multi-source sensors for cross-domain fault diagnosis. *Neural Networks*, *179*, 106518. https://doi.org/10.1016/j.neunet.2024.106518

Zhang, C., Qi, H., Wang, S., Li, Y., & Lyu, S. (2024). COMICS: End-to-End Bi-Grained Contrastive Learning for Multi-Face Forgery Detection. *IEEE Transactions on Circuits and Systems for Video Technology*, *34*(10), 10223-10236. https://doi.org/10.1109/tcsvt.2024.3405563

Zhang, F., Li, R., & Liu, S. (2010). Contour extraction of gait recognition. *Procedia Engineering*, *7*, 275-279. https://doi.org/10.1016/j.proeng.2010.11.044

Zhang, R., Ji, Y., Zhang, Y., & Passonneau, R. J. (2022). Contrastive data and learning for natural language processing. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Tutorial Abstracts*, 39-47. https://doi.org/10.18653/v1/2022.naacl-tutorials.6

Zhang, S., & Ran, N. (2024). Fine-grained and coarse-grained contrastive learning for text classification. *Neurocomputing*, *596*, 128084. https://doi.org/10.1016/j.neucom.2024.128084

Zhen, T., Kong, J., & Yan, L. (2020). Hybrid Deep-Learning Framework Based on Gaussian Fusion of Multiple Spatiotemporal Networks for Walking Gait Phase Recognition. *Complexity*, *2020*, 1-17. https://doi.org/10.1155/2020/8672431

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*. 20228-20237. https://doi.org/10.1109/cvpr52688.2022.01959

Zou, Q., Wang, Y., Wang, Q., Zhao, Y., & Li, Q. (2020). Deep Learning-Based Gait Recognition Using Smartphones in the Wild. *IEEE Transactions on Information Forensics and Security*, *15*, 3197-3212. https://doi.org/10.1109/tifs.2020.2985628