The Visual Computer (2025) 41:6823–6835
https://doi.org/10.1007/s00371-025-03983-w

REVIEW

Graph convolutional networks for 3D skeleton-based scoliosis screening using gait sequences

Zizhao Peng[1] · Zihan Wang[1] · Mengying Sun[1] · Zheng Lv[2] · Yan Wang[3] · Ping Li[1] · Fengwei An[4]

Accepted: 30 April 2025 / Published online: 27 May 2025

Abstract

Adolescent idiopathic scoliosis is a significant health concern, ranked as the third most prevalent issue among adolescents after obesity and myopia. Traditional screening methods rely on the use of complex and expensive measuring instruments and expert physicians to interpret X-ray images. These methods can be both time-consuming and inaccessible for widespread screening efforts. To address these challenges, we propose a standardized protocol for the collection of scoliosis gait dataset. This protocol enables the systematic capture of relevant gait characteristics associated with scoliosis, leading to the creation of a comprehensive, annotated dataset tailored for research and diagnostic purposes. Leveraging this dataset, we developed an effective deep learning algorithm based on graph convolutional networks, which outperforms traditional CNN by effectively modeling the complex spatial and temporal dynamics of human gait and posture, leveraging skeletal structure as a graph for more accurate and robust scoliosis screening. We also explored various optimization strategies to enhance the model's accuracy and efficiency, ensuring robust performance across diverse scenarios. Our innovative approach allows for the rapid and non-invasive recognition of scoliosis. This method is not only scalable but also eliminates the need for specialized equipment or extensive medical expertise, making it ideal for large-scale screening initiatives. By improving the accessibility and efficiency of scoliosis detection, our approach has the potential to facilitate early intervention.

Keywords Adolescent idiopathic scoliosis · Graph convolutional networks · Medical imaging · Gait dataset

✉ Ping Li
p.li@polyu.edu.hk

✉ Fengwei An
anfw@sustech.edu.cn

Zizhao Peng
21051291r@connect.polyu.hk

Zihan Wang
20031683r@connect.polyu.hk

Mengying Sun
mengying.sun@connect.polyu.hk

Zheng Lv
lzlgzxyy@163.com

Yan Wang
13562898274@163.com

[1] Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong

[2] Department of Rehabilitation, Longgang District Central Hospital of Shenzhen, Shenzhen, China

[3] Rehabilitation Medicine Center, The Affiliated Taian City Central Hospital of Qingdao University, Taian, China

[4] School of Microelectronics, Southern University of Science and Technology, Shenzhen, China

Abbreviations

CTZX	ChuangTang middle school in Heyuan, CN
CTZXXX	ChuangTang Center primary school in Heyuan, CN
DHXX	DongHua school in Heyuan, CN
DJZX	Dongjiang school in Heyuan, CN
DJZXGS	Dongjiang school third-year senior, in Heyuan, CN
DTDEXX	Dengta Second primary school, in Heyuan, CN
DTDYXX	Dengta First primary school, in Heyuan, CN
DTZX	Dengta middle school, in Heyuan, CN
DTZXXX	Dengta center primary school, in Heyuan, CN
DYCWXX	Dongyuan ChongWen school, in Heyuan, CN

DYDEXX	Dongyuan second primary school, in Heyuan, CN
DYDSXX	Dongyuan fourth primary school, in Heyuan, CN
DYDSXX3	Dongyuan third primary school, in Heyuan, CN
DYDWXX	Dongyuan fifth primary school, in Heyuan, CN
DYDYXX	Dongyuan first primary school, in Heyuan, CN
DYGJZX	Dongyuan senior high school, in Heyuan, CN
DYSYZX	Dongyuan experimental middle school, in Heyuan, CN
DYZX	Dongyuan middle school, in Heyuan, CN
HCZX	Huangcun middle school, in Heyuan, CN
HCZXXX	Huangcun center primary school, in Heyuan, CN
HTZX	Huangtian middle school, in Heyuan, CN
HTZXXX	Huangtian center primary school, in Heyuan, CN
KHZXXX	Kanghe center primary school, in Heyuan, CN
LHZX	Luohu middle school, in Heyuan, CN
LKZX	Lanko middle school, in Heyuan, CN
MDCXX	Meidi City primary school, in Heyuan, CN
SDTAYY	Shandong Taian hospital, in Taian, CN
SJZX	Shuangjiang middle school, in Heyuan, CN
STZX	Shuntian middle school, in Heyuan, CN
SUSTECH11	Southern University of Science and Technology, in Shenzhen, CN
SZDXZYY	Shenzhen University General Hospital, in Shenzhen, CN
SZETYY	Shenzhen Children's Hospital, in Shenzhen, CN
SZLGYY	Shenzhen Longgang center Hospital, in Shenzhen, CN
SZRMYY	Shenzhen People's Hospital, in Shenzhen, CN
XGZXXX	Xingang center primary school, in Heyuan, CN
XTXX	Xiangtang primary school, in Heyuan, CN
XTZX	Xiangtang middle school, in Heyuan, CN
XXW	New Hope company, in Heyuan, CN
YHZX	Yihe middle school, in Heyuan, CN
YTZX	Yetan middle school, in Heyuan, CN

1 Introduction

Adolescent idiopathic scoliosis (AIS) is the most common form of scoliosis, emerging in children and adolescents during growth. This spinal deformity is characterized by a lateral curvature and vertebral rotation, potentially causing a visible hump and asymmetrical waist. The idiopathic nature of AIS indicates that its exact cause remains unknown, despite its significant incidence and potential impact on health and quality of life.

Early detection and intervention are crucial for effectively managing AIS and reducing the risk of progression. Groot [9] examined the effectiveness of educating parents to recognize scoliosis. Participants assessed two sets of 14 cases, including 8 with scoliosis and 6 without, each with photographs of a child's back both standing and in forward flexion. Parents determined the need for physician referral. Despite education, only 74% accurately identified scoliosis.

Although radiographic measurement of the Cobb angle currently remains the clinical gold standard for the diagnosis of scoliosis, its inherent limitations are becoming increasingly apparent. Studies have shown that a single full-length radiographic examination of the spine in the standing position can expose adolescent patients to an effective radiation dose of approximately 0.14−0.20 mSv [5], while the International Commission on Radiological Protection (ICRP) has stated that each Sv of radiation dose will result in an incremental lifetime risk of cancer death of 5.5%, which poses a potential health threat to the adolescent population who require regular review [19]. In clinical practice, Cobb angle measurements are subject to measurement errors of up to 5°–7°, which mainly stem from subjective differences in vertebral end-plate identification [17]. Additionally, radiography relies on expensive digital imaging equipment and the operation of specialized radiologic technologists, resulting in high screening costs, which severely limits the accessibility of large-scale screening in areas where healthcare resources are poor.

Recently, neural networks have gained interest for non-invasive scoliosis detection. Convolutional neural networks (CNNs) excel in image recognition and medical imaging analysis, identifying patterns in spinal images that may escape human detection, enabling earlier and more accurate detection. Yang et al. [25] developed a CNN for automatic scoliosis screening using bare back images, outperforming human experts in detecting curves of 20° or more and classifying severity. However, CNN-based methods often rely on localized features or static image representations, which may not fully capture the complex spatial and temporal dynamics of human gait and posture. In contrast, the proposed use of graph convolutional network (GCN) offers significant advantages by inherently modeling the skeletal structure as a graph, allowing for direct and efficient representation of the spatial relationships between joints. Furthermore, GCN excels in dynamic spatiotemporal feature modeling, effectively capturing the intricate interactions and motion patterns over time that are critical for accurate scoliosis detection. These capabilities enable GCN-based approaches to provide

a more comprehensive analysis of gait abnormalities, leading to improved screening performance compared to CNN.

GCN effectively learns relationships among nodes by encoding graph structures and node features, showing impressive classification results. Integrating gait analysis with GCN models introduces a novel scoliosis screening approach. This method harnesses artificial intelligence to process complex biomechanical data and identify scoliosis correlations. Benefits include reduced radiography reliance, earlier detection, and the ability to monitor scoliosis progression.

This work uses deep learning algorithms for gait analysis to achieve scoliosis recognition. The main contributions are as follows:

- We applied gait analysis to scoliosis detection and, in conjunction with deep registration, developed a scoliosis detection algorithm centered on three-dimensional skeletal data that utilizes graph convolutional networks.
- We establish a 3D scoliosis dataset and develop standards for scoliosis sample collection.
- We further investigated the impact of binary classification and the symmetry-based skeleton feature grouping strategy.

2 Related works

Conventional diagnostic methods, such as radiography, are the clinical standard for scoliosis assessment but have limitations, including radiation exposure and challenges in maintaining a standard measurement protocol [20]. Computer-aided detection (CAD) systems have been developed to enhance the accuracy and efficiency of scoliosis assessment in radiographs [27]. However, these systems require high-quality images and do not eliminate radiation use.

Recent research highlights the potential of neural networks in image-based scoliosis detection. For gait data with depth information, modern CNN-based deep learning methods outperform traditional geometric approaches in pose estimation on account of their ability to handle different types of input data: depth maps [8] (2D images with depth values), 3D volumetric data [2] (3D grid representations), and point clouds [10] (collections of 3D points). Researchers use specialized neural network designs like 2D CNN [21] (processing depth maps as images), 3D CNN [2] (analyzing volumetric data), MLPs (multilayer perceptrons) [7] (for point cloud processing), and transformer models [10] (capturing long-range dependencies). Comparative analyses [2] demonstrate that 3D CNN maintain superior spatial coherence preservation notwithstanding their substantial computational overhead, while alternative approaches like the anchor-to-joint paradigm [21] employ two-stage processing with 2D heatmap generation followed by depth offset estimation from local depth distributions, yet still underperform relative to volumetric probability estimation via 3D CNN hourglass architectures [2].

Recent advances in graph convolutional network (GCN) have revolutionized gait analysis through their inherent capability to model skeletal topology. Some predominant GCN-based approaches have emerged: Skeleton-sequence GCN [24] directly operates on raw joint coordinates, constructing spatiotemporal graphs where nodes represent body joints and edges encode natural bone connections. While computationally efficient [18], these methods often neglect global inter-joint relationships beyond predefined edges. Multi-scale GCN [3] addresses this limitation by learning hierarchical graph representations through adaptive edge weighting and multi-hop neighborhood aggregation, achieving superior performance on cross-view gait recognition at the cost of increased computational complexity (+38% FLOPs).

Decoupled spatial-temporal GCN [23] proposes independent spatial and temporal convolution modules, reducing parameters by 27% compared to coupled architectures while maintaining competitive accuracy on CASIA-B dataset (98.1% vs. 98.4%). While decoupled spatial-temporal graph convolutional networks achieve improved parameter efficiency through modular separation, their architectural design fundamentally suffers from inherent limitations in cross-modal synergistic modeling. The rigid decoupling of spatial and temporal features artificially dissociates the dynamically coupled relationships between limb spatial configurations and temporal evolution patterns essential for action recognition, thereby constraining representational capacity for complex multi-joint coordinated movements—particularly prone to spatiotemporal feature misalignment when handling non-uniform temporal sampling or abrupt motion transitions.

Integrating gait analysis with advanced neural network techniques is an emerging field. Initial studies using deep learning algorithms on gait data have laid the groundwork for developing non-invasive, low-cost, and accurate methods for early scoliosis detection [1]. These approaches promise to revolutionize scoliosis screening and monitoring, especially in settings with limited access to advanced medical imaging. The AIS diagnostic landscape is undergoing a paradigm shift with machine learning and neural networks. The convergence of image processing and gait analysis through these advanced models holds significant promise for developing novel diagnostic and monitoring tools. This shift has the potential to improve patient outcomes and provide deeper insights into the underlying mechanisms of AIS.

Table 1 Statistical results of our datasets

Label	Male	Female	Age (years, mean±std)	Height (cm, mean±std)	Weight (kg, mean±std)
Negative	4388	4025	13.79 ± 1.85	157.03 ± 11.32	46.11 ± 12.30
Critical	1751	2191	13.96 ± 1.49	158.28 ± 10.29	45.50 ± 10.36
Positive	555	1136	15.14 ± 4.70	158.87 ± 9.90	45.31 ± 9.87

Due to changes in recording protocols and the extensive time span of the dataset, some data lack complete personal information. The above statistics are based on 14,043 records that contain full individual details

3 Dataset

3.1 Overview of the dataset

To address scoliosis, we have curated a novel dataset comprising 31,463 cases from 36 schools and 5 hospitals, as shown in Table 5. Data collection was approved by the hospital ethics committee, with informed consent from participants. We used Kinect cameras to capture parallel streams of color and depth video at a 720p resolution. Each sample consists of 300 frames recorded at 15 frames per second.

Professional doctors specializing in scoliosis meticulously labeled these samples, ensuring precise dataset annotations. Scoliosis severity was measured using the Cobb angle, a standardized metric reflecting spinal axial rotation and lateral curvature. Following medical guidelines, the dataset was categorized into three classifications: 'positive' for Cobb angles greater than 10 degrees, 'negative' for angles less than 10 degrees, and 'critical' for angles around the 10-degree threshold. The dataset includes 3909 positive samples, 18,396 negative samples, and 9178 critical samples. To address the lower prevalence of positive samples, participants with positive Cobb angles were asked to contribute additional samples, ensuring a more balanced dataset for analysis.

The detailed sample distribution is presented in Table 1. During the data collection process, we ensured a diverse and representative sample by covering a wide range of demographic groups. The gait samples in the dataset encompass various age groups, genders, as well as differences in height and weight, reflecting the richness and diversity of the data. This heterogeneous composition of samples is instrumental in enhancing the model's generalization ability in real-world applications, thereby ensuring more robust and reliable performance across different populations.

3.2 Setup

The site layout is illustrated in Fig. 1. To replicate a natural gait, participants were instructed to walk within a range of 1.4 to 4.2 ms from the camera. This methodological approach was designed to simulate a controlled yet natural walking environment conducive to accurate biomechanical analysis. The Kinect camera is positioned 1.3 ms above the ground, angled downward at 13 degrees from the horizontal plane. The RGB camera has a field of view of 75° × 65°, while the lidar camera has a field of view of 90° × 59°, both outputs fixed at 1080p, 15 fps, and 300 frames.

Due to the Kinect camera's lidar sensor sensitivity to light, testing environments are typically required to be indoors without direct sunlight on the subject. It's also advisable to avoid backlighting scenarios where the camera faces a window directly. During detection, subjects should wear tight-fitting or properly fitting clothing, as long skirts may prevent the device from accurately recording gait information. It is recommended that subjects avoid wearing black clothing, as some black materials absorb light, potentially leading to inaccurate detection data. Subjects should remove hats to expose their ears and tie up their hair to reveal their shoulders.

3.3 Preprocess

In preparation for further investigation, the dataset underwent extensive preprocessing. Due to the Kinect's lidar unit's sensitivity to sunlight and color, and to prevent inaccuracies in gait recognition caused by clothing or environmental interference, we conducted a manual screening to exclude videos with quality issues. Identified problems included direct sunlight exposure, dark clothing that absorbs light, long skirts or garments obscuring the legs, subjects being too close or too far from the camera, and background disturbances from other people. We present some typical examples of excluded samples in Fig. 2. Given the lower quality of depth information from the lidar, we optimized the depth values of key points to address missing data issues that could compromise quality.

To further enhance the dataset, we utilized GroundingDINO [15] to accurately identify and extract participant presence within video frames. Subsequently, SAMHQ [12] was applied to segment the high-resolution color and depth images into masked images without size constraints. This process was crucial in isolating relevant anatomical features from the surrounding environment, thereby enhancing data quality and focus for analysis.

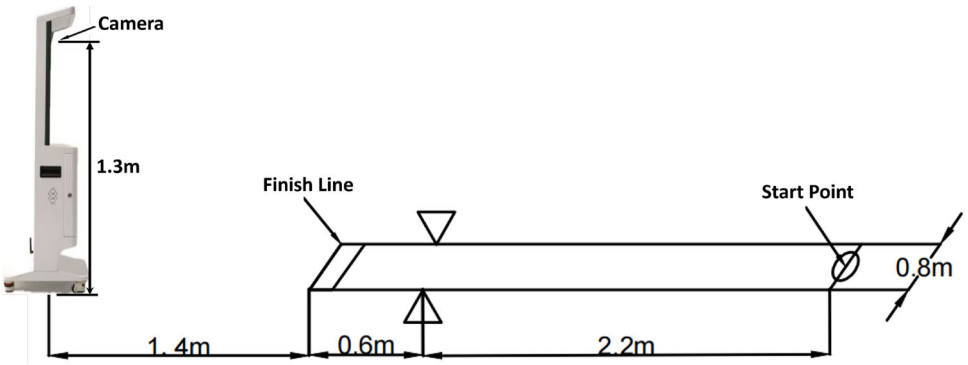


Fig. 1 Data collection setup

Table 2 Experiment dataset

Dataset	Negative	Critical	Positive
Train	5395	3219	2470
Test	1349	805	618

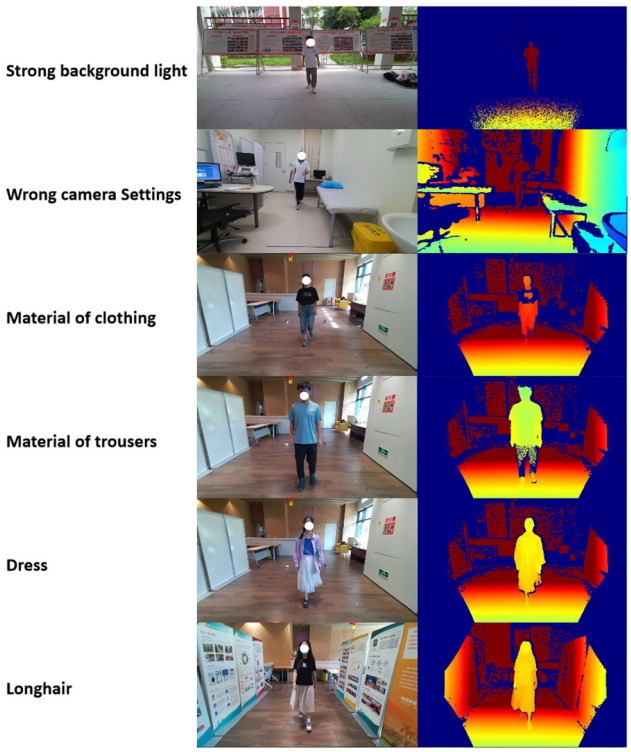


Fig. 2 Typical examples of excluded samples

4 Methodology

Our system architecture, illustrated in Fig. 3, presents a comprehensive scoliosis detection system. This system converts video streams into a series of 3D skeleton coordinates to represent gait features, and then employs Graph Convolutional Networks to thoroughly analyze gait patterns based on these skeleton.

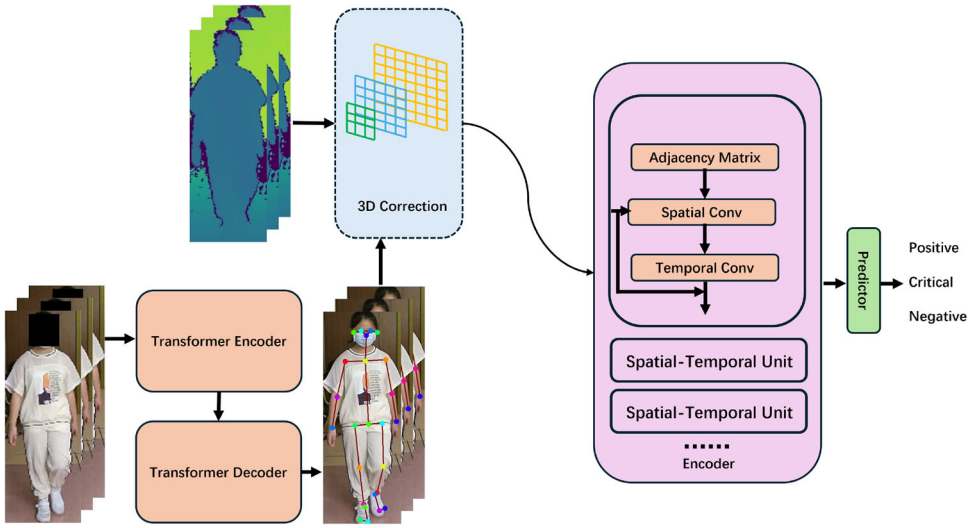
The Spatial Convolution module, implemented through hierarchical dynamic graph convolution, captures the structural relationships between joints in the human skeleton. It adaptively learns the correlations between different body parts by utilizing a learnable adjacent matrix, which evolves during training to reflect the optimal spatial dependencies for action recognition. This module effectively processes the anatomical constraints and dynamic connections inherent in human movements.

The Temporal Convolution module employs a multi-scale architecture to capture motion patterns across different time spans. Through parallel branches with varying dilated convolutions and pooling operations, it can simultaneously model both short-term subtle movements and long-term action dependencies. This design enables the module to be robust against variations in action speed and temporal deformation.

The Spatial-Temporal Unit combines these two components in a sequential manner, where spatial features are first extracted from joint configurations, followed by temporal feature extraction from the motion sequence. This unit is augmented with residual connections and attention mechanisms, allowing for efficient gradient flow and selective feature enhancement. The integration of spatial and temporal convolutions enables comprehensive modeling of both structural and dynamic aspects of human actions.

The encoder architecture comprises 10 cascaded spatial-temporal units, organized in a progressive manner. Starting with 64 base channels, the encoder gradually expands its capacity through three stages: The first stage (units 1–4) maintains spatial and temporal dimensions while building basic motion patterns; the second stage (units 5–7) doubles the channel dimension to 128 with stride-2 temporal downsampling; the final stage (units 8–10) further increases channels to 256, capturing high-level action semantics. This hierarchical design, processing from 3 input channels to 256 output channels, enables the encoder to construct a comprehensive feature pyramid, where early layers capture local joint movements and deeper layers abstract complex action patterns. The architecture is complemented by skip connec-

Fig. 3 Overall architecture



tions and attention mechanisms, facilitating both fine-grained motion details and global action understanding.

4.1 2D skeleton estimation

For the acquired gait dataset, it is essential to preprocess the data into a format that is easily comprehensible for the model. This involves minimizing noise and reducing factors that might lead to model overfitting. In the context of scoliosis detection, we propose transforming RGB and depth images, which often contain redundant information, into a simplified skeletal point format. This approach facilitates the identification of asymmetries and differences in stride frequency and amplitude between the legs of scoliosis patients. With the advent of the Vision Transformer (ViT) [4], various ViT-based visual models have achieved remarkable success across numerous application domains. Currently, ViTPose [22] has become the state of the art in the field of keypoint detection. We utilized a model fine-tuned with ViTPose on the COCO-25 dataset as our 2D pose recognition model. This model efficiently and accurately extracts human keypoints, providing a reliable foundation for subsequent 3D skeleton construction.

4.2 3D skeleton construction

To obtain 3D pose information in a single-view scenario, we registered the detected 2D skeleton with the depth information from the Kinect's lidar camera. However, the lidar camera is susceptible to interference from external light sources, causing significant noise in the original depth data. Direct usage can lead to convergence issues in the loss function during training.

We designed a correction module to correct contaminated depth data and generate high-quality 3D skeleton data to address this issue. The input skeleton points form a sequentially ordered two-dimensional series of 25 points.

$$\mathbf{P} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_{25}, y_{25})\}, \tag{1}$$

Each point corresponds to a z values on the lidar image:

$$\mathbf{Z} = \{z_1, z_2, \ldots, z_{25}\}, \tag{2}$$

This module consists of multiple reference points and a series of progressively larger filters. We performed a statistical analysis to evaluate the accuracy of depth information acquisition across 25 points. Among these, the seven points with the highest accuracy—specifically the nose, chest, left shoulder, right shoulder, left hip, right hip, and pelvis—were chosen as reference points set $\mathbf{R} \subseteq \mathbf{P}$. These reference points are utilized to establish the baseline range for the skeleton's z_{ref} values.

$$z_{\text{ref}} = \frac{1}{|\mathbf{R}|} \sum_{(x_i, y_i) \in \mathbf{R}} z_i, \tag{3}$$

Filter sizes are applied sequentially from smallest to largest on the lidar image at each point.

$$\text{filter range} \in \{1, 2, 4, 9, 16, 25\}, \tag{4}$$

For each filter, compute the set of all points (x_j, y_j) within the neighborhood $\mathbf{N}_i$ such that:

$$\|(x_i, y_i) - (x_j, y_j)\| \leq \text{filter range}, \tag{5}$$

Then calculate the average value of points in the neighborhood that have a difference from the reference value within

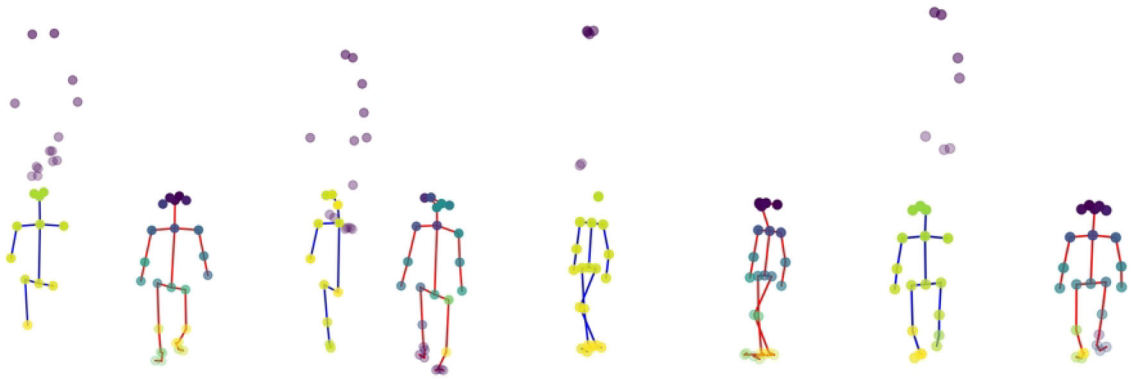


Fig. 4 Correction effect on depth noise. The blue skeleton represents the 3D joint positions before correction, while the red skeleton represents the 3D joint positions after correction

the threshold ϵ:

$$\bar{z}_i = \frac{\sum_{(x_j,y_j)\in \mathbf{N}_i, |z_j - z_{\text{ref}}| < \epsilon} z_j}{|\{(x_j, y_j) \in \mathbf{N}_i, |z_j - z_{\text{ref}}| < \epsilon\}|}, \tag{6}$$

Prioritize matching with smaller filter sizes to find the smallest filter size that meets the conditions. Combine the corrected z values with the original coordinates:

$$\mathbf{P}_3 = \{(x_i, y_i, \bar{z}_i) \mid i = 1, 2, \ldots, 25\}, \tag{7}$$

The correction effect on depth noise is illustrated in Fig. 4.

4.3 Scoliosis recognition

We employed a graph convolutional network as the backbone for our classification model. This model uses skeleton sequences as input to enhance the detection and classification of scoliosis-related gait abnormalities. In recent advancements in the field of graph neural networks, we have developed a new version of our model that outperforms the Hierarchical Dynamic Graph Convolutional Network [14].

Firstly, we have revised the aggregation function within the GNN framework. The aggregation function plays a crucial role in determining how information is collected from neighboring nodes to update the feature representations of the nodes in the graph. By optimizing this aggregation mechanism, our model can more effectively capture complex dependencies and relationships between nodes. This optimization allows for a richer representation of the graph structure, leading to improved prediction accuracy and robustness in various applications.

Secondly, in our model, we have improved an innovative approach that automatically learns an adjacency matrix based on the data itself. This automatic learning mechanism enables the model to dynamically ascertain the optimal relationships among nodes, eliminating the biases and inefficiencies associated with predefined groupings.

Our GCN module implements biomechanical modeling of spinal motion asymmetry through a dynamically learned adjacency matrix. The proposed hierarchical framework employs a three-level decomposition architecture to capture local vertebral motion, spinal segment dynamics, and global spine–pelvis coordination patterns. Within this structure, our dynamic adjacency matrix learning mechanism quantifies asymmetric intervertebral coupling strength through trainable feature transformation layers, where the sigmoid-activated weights explicitly encode lateral biomechanical response disparities during specific movements. To address asymmetric pathological patterns like scoliosis, we design directional-specific convolutional operators that extract coronal-sagittal plane motion discrepancies, integrated with an attention-based multi-level adjacency fusion scheme. This configuration enables our model to autonomously detect vertebral rotation offsets and lateral bending deformations through learned kinematic representations, establishing an interpretable computational framework for spinal motion analysis.

Mathematically, the relationship between the node feature matrix $\mathbf{X}$, the learned adjacency matrix $\mathbf{A}$, and the updated node features can be articulated through the following formula:

$$\mathbf{H}^{(l+1)} = \sigma\left(\mathbf{A} \cdot \mathbf{H}^{(l)} \cdot \mathbf{W}^{(l)}\right), \tag{8}$$

In this equation, $\mathbf{H}^{(l)}$ denotes the feature representation of the nodes at layer l, $\mathbf{W}^{(l)}$ represents the learnable weight matrix at that layer, and σ is a nonlinear activation function such as ReLU. This formulation illustrates how our model effectively aggregates features from neighboring nodes while utilizing an adjacency matrix that is adaptively learned, thereby enhancing the model's capacity to represent intricate node relationships within the graph.

Assume the skeleton data contain N joints; each joint's position at time t is given by $\mathbf{x}_i^t \in \mathbf{R}^d$, where i is the joint

index, and d is the feature dimension (commonly 3, such as (x, y, z) coordinates).

The skeleton data are decomposed into K local substructures; each substructure contains N_k joints, where k is the substructure index. Each local substructure k is represented as a graph $\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k)$, where $\mathcal{V}_k$ is the set of nodes and $\mathcal{E}_k$ is the set of edges. GCN is applied to each local substructure. For the l-th layer of the GCN, the feature update formula for local substructure k is:

$$\mathbf{H}_k^{(l+1)} = \sigma(\mathbf{D}_k^{-1/2}\mathbf{A}_k\mathbf{D}_k^{-1/2}\mathbf{H}_k^{(l)}\mathbf{W}_k^{(l)}), \tag{9}$$

where $\mathbf{H}_k^{(l)}$ is the node feature matrix at the l-th layer. $\mathbf{A}_k$ is the adjacency matrix of local substructure k. $\mathbf{D}_k$ is the diagonal matrix for normalization, defined as $\mathbf{D}_k = \text{diag}(\sum_j \mathbf{A}_{k_{ij}})$. $\mathbf{W}_k^{(l)}$ is the weight matrix for the l-th layer. σ is the activation function, such as ReLU.

All local substructures' graphs are integrated to form a global skeleton graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of all joints and $\mathcal{E}$ is the set of edges in the global skeleton. GCN is applied to the global skeleton graph to extract overall spatial features. The global feature update formula is:

$$\mathbf{H}^{(l+1)} = \sigma(\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\mathbf{H}^{(l)}\mathbf{W}^{(l)}), \tag{10}$$

where $\mathbf{A}$ is the adjacency matrix of the global skeleton graph.

Then temporal convolution is introduced to capture dynamic changes in actions. Let $\mathbf{H}_t$ be the feature matrix at time t, then the temporal convolution operation is defined as:

$$\mathbf{H}_t^{'} = \sum_{\tau=-T}^{T} \mathbf{H}_{t+\tau}\mathbf{W}_\tau, \tag{11}$$

where T is the size of the temporal convolution window. $\mathbf{W}_\tau$ is the weight matrix of the temporal convolution kernel.

Temporal features are combined with spatial features to form the final spatiotemporal feature representation:

$$\mathbf{H}^{\text{final}} = \text{Concat}(\mathbf{H}^{(L)}, \mathbf{H}_t^{'}), \tag{12}$$

where $\mathbf{H}^{(L)}$ is the output feature of the last GCN layer, and $\mathbf{H}_t^{'}$ is the feature after temporal convolution.

5 Experiment

5.1 Setup

Due to the sensitivity of lidar sensors to lighting conditions, clothing color, and material, as well as frequent deviations from the data collection standards in the scenarios and subjects, we conducted a comprehensive data cleaning process prior to the experiments. We retained 13,856 data points that exhibited minimal interference, accurate collection ranges, suitable lighting conditions, and adherence to the standard collection protocols. Among these, there were 6744 negative samples, 4024 critical samples, and 3088 positive samples. Subsequently, we divided the dataset into training and testing sets at a ratio of 4:1 (see Table 2).

To minimize interference from frequent turning and moving out of the effective recording range during the subjects' back-and-forth walking, we retained only the parts where the subjects walked toward and away from the camera. Sections involving turning near or far from the camera were removed.

The training hyperparameters are set as follows: the batch size is 64, the optimizer is Lamb, the initial learning rate is 0.001 with a minimum of 0.0001, and 200 iterations were conducted on the dataset.

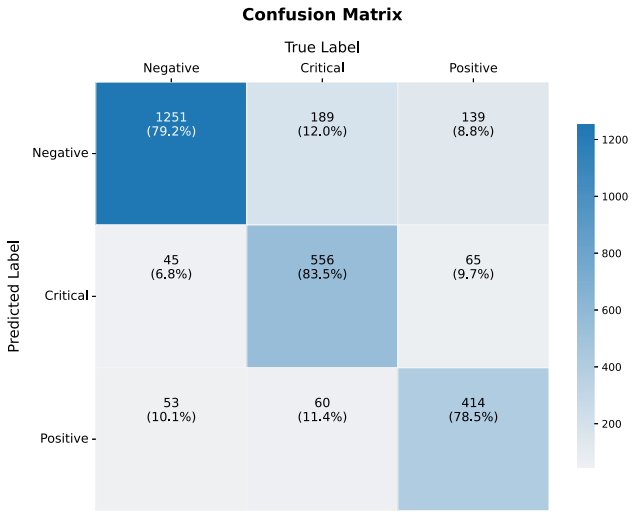


Fig. 5 Confusion matrix of result

5.2 Result

The experimental results are presented in Fig. 5, where we compare our results with traditional scoliosis screening methods and STGCN [24]. The comparative data are provided in Table 3.

Traditional screening methods for scoliosis generally include the Forward Bending Test (FBT), Scoliometer (S), and Moiré topography screening (M). The combination of multiple screening methods can enhance the accuracy of scoliosis detection. The team from Rochester [26] employed a combined approach of FBT and Scoliometer. In contrast, the Hong Kong team [6, 13, 16] adopted a more complex method that integrates FBT, Scoliometer, and Moiré topography screening, achieving commendable results. The Greek

Table 3 Comparison results

Method	Screening test	Accuracy	Sensitivity	Specificity
Rochester [26]	FBT+S	–	71.1% (54.1−84.6)	97.1% (96.3−97.7)
Hong Kong [6, 13, 16]	FBT+S+M	–	93.8% (93.3−94.3)	99.2%(99.2−99.2)
Greece [11]	FBT	–	84.4% (67.2−94.7)	95.2% (94.3−95.9)
	S	–	90.6% (75.0−98.0)	80.7% (79.1−82.1)
	M	–	100.0% (84.2-100)	85.4% (84.0−86.7)
STGCN [24]	–	53.8%	51.1%	75.9%
This work	–	80.1%	74.9%	95.9%
This work (Binary)	–	85.8%	82.9%	88.7%

team [11], on the other hand, conducted individual assessments to evaluate the efficacy of each method.

We also presented the testing results of the STGCN model on the scoliosis dataset. The training was conducted over 200 epochs with a batch size of 64 and a learning rate of 0.001. It is important to note that the STGCN exhibited a severe overfitting phenomenon on this dataset. The model achieved a maximum accuracy of 97.8% on the training set, but its performance on the test set was significantly lower, with an accuracy of only 53.8%. In fact, overfitting appears to be a widespread issue in our dataset, particularly when the training data does not include location-specific information. We also discussed this problem in the discussion section.

The results demonstrate that our purely visual and non-invasive rapid screening model for scoliosis has achieved a level of effectiveness comparable to that of traditional measurement instruments used by experts. In certain evaluation criteria, particularly when experts rely on a single instrument for detection, our model has even surpassed the performance of conventional screening methods.

It is worth noting that Hong Kong's screening program [6] has achieved a high level of specificity. This is mainly due to its design as a long-term screening initiative conducted every two years until students reach the age of 19. The program incorporates a two-tier screening system along with radiographic diagnosis. In the first stage, students undergo screening at community clinics using FBT, and their trunk rotation angle (ATR) is measured. Students with an ATR between 5° and 15° are further evaluated using moiré topography. If there are two or more moiré lines that differ between the left and right sides of the back, or if clinical signs of obvious spinal deformity are present, these students are referred for radiographic diagnosis. Students with an ATR of 15° or higher are directly referred to a specialist hospital for radiographic assessment.

In Fig. 6, we visualized the features to illustrate the focus of our model. We applied regularization to the feature layer outputs and represented the model's attention to specific points using the size of the halo surrounding each point. Our observations indicate that the chest and foot regions are

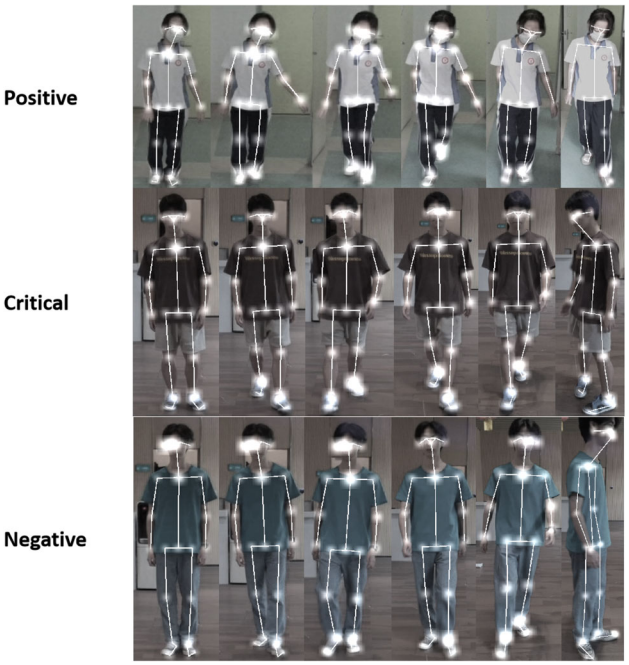


Fig. 6 Feature visualization of positive, critical and negative samples

the areas of greatest interest for the model. The chest region serves as a marker for the body's symmetry center, especially in cases of positive scoliosis, where a significant lateral tilt of the chest occurs, affecting the skeletal symmetry of the body. During walking, we observed that the model's attention alternates between the two supporting feet, reflecting its focus on the force-bearing foot at each phase of gait.

Furthermore, we investigated the results of binary classification. We combined the positive data with critical data to create a new positive dataset. This new dataset is more balanced, comprising 6744 negative samples and 7112 positive samples. We utilized this new dataset for binary classification training, with the results presented in Fig. 7 and compared in Table 3. Our findings indicate that the more balanced binary classification training led to a noticeable increase in the overall accuracy of the model, while the disparity between sensitivity and specificity was significantly reduced.

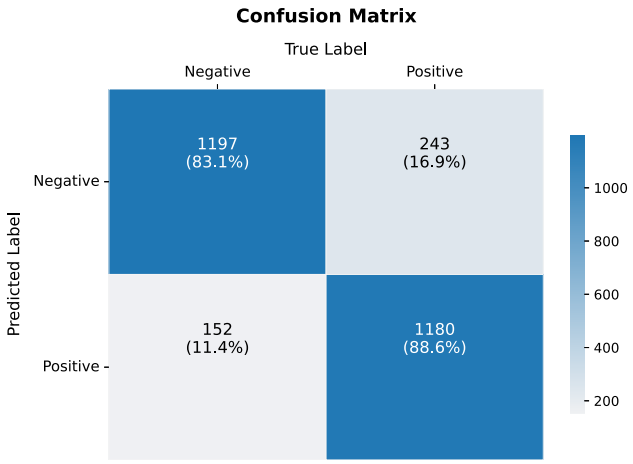


Fig. 7 Binary classification result

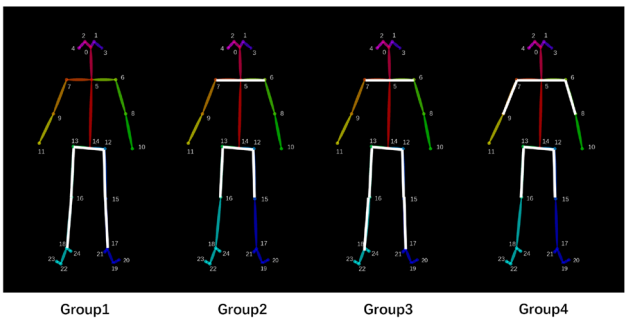


Fig. 8 Grouping strategies. Group 1 includes the left and right hips, middle hip, left and right knees, and left and right ankles. Group 2 includes the left and right hips, middle hip, left and right knees, left and right shoulders, and the chest. Group 3 combines all points from Group 1 and Group 2. Group 4 includes all points from Group 1 and left and right elbows

6 Discussion

6.1 Grouping strategy

Gait asymmetry in patients with scoliosis is a key characteristic identified through gait analysis for recognizing scoliosis. We also explored manually designed grouping features to identify skeletal points that significantly influence classification accuracy in the task of recognizing scoliosis. To this end, we further investigated the effects of different grouping strategies on the model, as depicted in Fig. 8, with the experimental results presented in Table 4.

The experimental results are consistent with those observed in feature visualization. The key factors influencing the model's classification are the points at the chest location and the footpoints. We believe that the model captures gait symmetry and upper-body tilt by analyzing the temporal changes in the central chest point and the points on both legs. This aligns with our prior understanding that gait asymmetry is a typical characteristic in scoliosis patients during walking.

Table 4 Grouping strategy experiment result

Group	Accuracy (%)	Sensitivity (%)	Specificity (%)
Group1	65.4	50.0	90.9
Group2	69.0	64.3	77.0
Group3	71.8	79.8	58.6
Group4	58.1	35.5	95.8

An interesting phenomenon, as shown in Group 4, is that the points on the arms in the feature visualization are not the primary factors affecting the model's decision. However, when arm points are included in the grouping, there is a significant negative impact on the model's sensitivity. This may be because the irregular swinging of the arms introduces noise or distracting information, interfering with the model's ability to accurately discern relevant gait features.

6.2 Overfitting

Enhancing the model's generalization capability is a challenging aspect of this work. Although iterating multiple times on small samples can easily yield high accuracy, it often leads to significant overfitting. According to our experiments, the model typically starts to overfit before reaching 200 epochs. To address this, we employed several strategies, including increasing data volume, refining model optimization, and adjusting hyperparameters. Additionally, we utilized data augmentation techniques such as Rotation, Shear, Frame Reversal, Gaussian Noise, and Subsampling. We also incorporated dropout and normalized the data to a 0–1 range to help reduce overfitting and improve the model's generalization performance. Despite these efforts, further refinement is still needed to enhance the model's ability to generalize to new data scenarios. We believe there is significant potential for exploration in this field.

7 Conclusion

In conclusion, our study demonstrates the potential of gait analysis as a reliable and scalable method for scoliosis screening, particularly suitable for large-scale settings such as school screening programs. The proposed deep learning approach enables rapid, non-invasive detection of scoliosis, making it a practical tool for early intervention. Additionally, by establishing a large, meticulously annotated gait dataset, we provide a valuable resource to support future research and development in this field. Looking ahead, integrating this methodology with common household devices such as home cameras can facilitate real-time, remote monitoring of spinal health, allowing for early detection of scoliosis risk outside

Table 5 The composition of locations in the complete dataset, with location names represented by their pinyin abbreviations as explained above

Location	Positive	Negative	Critical
CTZX	5	81	113
CTZXXX	0	61	60
DHXX	25	0	90
DJZX	58	3272	1046
DJZXGS	0	7	15
DTDEXX	0	0	60
DTDYXX	3	0	74
DTZX	18	0	264
DTZXXX	4	0	86
DYCWXX	64	2579	1235
DYDEXX	0	749	142
DYDSXX	6	336	36
DYDSXX3	17	1360	264
DYDWXX	214	590	0
DYDYXX	296	1573	550
DYGJZX	78	1760	334
DYSYZX	727	1987	1540
DYZX	1078	1797	1774
HCZX	5	164	110
HCZXXX	0	84	46
HTZX	5	0	75
HTZXXX	0	0	69
KHZXXX	0	0	171
LCZX	0	96	45
LHZX	0	116	57
LKZX	19	37	39
MDCXX	0	326	0
SDTAYY	55	0	0
SJZX	0	114	21
STZX	0	103	39
SUSTECH11	66	133	178
SZDXZYY	626	0	0
SZETYY	157	810	41
SZLGYY	289	0	161
SZRMYY	61	30	8
XGZXXX	5	13	51
XTXX	5	0	114
XTZX	15	0	144
XXW	5	0	0
YHZX	0	125	48
YTZX	0	88	78

clinical settings. This approach could significantly expand the accessibility and convenience of scoliosis screening, promoting timely medical attention and improving outcomes across diverse populations.

A Composition of the full dataset

Acknowledgements The authors would like to thank the editors and anonymous reviewers for their insightful comments and suggestions. This work was supported by The Hong Kong Polytechnic University under Grants P0044520, P0048387, P0050657, and P0049586, Southern University of Science and Technology, The Affiliated Taian City Central Hospital of Qingdao University and Longgang District Central Hospital of Shenzhen.

Author Contributions Zizhao Peng, Zihan Wang, and Mengying Sun participated in the experiment, Zizhao Peng and Zihan Wang contributed to the writing of the paper, Ping Li and Fengwei An supervised the project, Zheng Lv and Yan Wang provided medical data and professional guidance on medicine.

Funding Open access funding provided by The Hong Kong Polytechnic University

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest This study was conducted in compliance with ethical standards through a multi-institutional collaboration between Southern University of Science and Technology and The Hong Kong Polytechnic University, having obtained approval from the Institutional Review Board of several hospitals. All data collection procedures involving human participants across multiple medical institutions and academic facilities were performed under physician supervision with written informed consent obtained prior to enrollment. Participant confidentiality and gait data security were strictly maintained through encrypted storage in our institutional database compliant with international data protection regulations (GDPR/ISO 27001 standards), ensuring no unauthorized access or data leakage throughout the research process. The investigation adheres to the ethical principles outlined in the Declaration of Helsinki and its later amendments, with no conflict of interest declared by any participating researchers or institutions involved in this study.

References

1. Alharthi, A.S., Yunas, S.U., Ozanyan, K.B.: Deep learning for monitoring of human gait: a review. IEEE Sens. J. **19**(21), 9575–9591 (2019)
2. Chang, J.Y., Moon, G., Lee, K.M.: V2v-posenet: voxel-to-voxel prediction network for accurate 3d hand and human pose estimation from a single depth map. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 5079–5088 (2018)

3. Chen, Z., Li, S., Yang, B., Li, Q., Liu, H.: Multi-scale spatial temporal graph convolutional network for skeleton-based action recognition. In: AAAI Conference on Artificial Intelligence, vol. **35**, pp. 1113–1122 (2021)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: The International Conference on Learning Representations (2021)
5. Duke, A., Marchese, R., Komatsu, D.E., Barsi, J.: Radiation in adolescent idiopathic scoliosis management: estimated cumulative pre-operative, intra-operative, and post-operative exposure. Orthopedic Research and Reviews pp. 487–493 (2022)
6. Fong, D.Y., Cheung, K.M., Wong, Y.W., Wan, Y.Y., Lee, C.F., Lam, T.P., Cheng, J.C., Ng, B.K., Luk, K.D.: A population-based cohort study of 394,401 children followed for 10 years exhibits sustained effectiveness of scoliosis screening. Spine **15**(5), 825–833 (2015)
7. Ge, L., Cai, Y., Weng, J., Yuan, J.: Hand pointnet: 3d hand pose estimation using point sets. In: IEEE Conference on Computer Vision and Pattern Recognitionn, pp. 8417–8426 (2018)
8. Ge, L., Liang, H., Yuan, J., Thalmann, D.: 3d convolutional neural networks for efficient and robust hand pose estimation from single depth images. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1991–2000 (2017)
9. de Groot, C., Heemskerk, J.L., Willigenburg, N.W., Altena, M.C., Kempen, D.H.R.: Educating parents improves their ability to recognize adolescent idiopathic scoliosis: a diagnostic accuracy study. Children **9**(4), 563 (2022)
10. Huang, L., Tan, J., Liu, J., Yuan, J.: Hand-transformer: Non-autoregressive structured modeling for 3d hand pose estimation. In: European Conference on Computer Vision, pp. 17–33 (2020)
11. Karachalios, T., Sofianos, J., Roidis, N., Sapkas, G., Korres, D., Nikolopoulos, K.: Ten-year follow-up evaluation of a school screening program for scoliosis: Is the forward-bending test an accurate diagnostic criterion for the screening of scoliosis? Spine **24**(22), 2318 (1999)
12. Ke, L., Ye, M., Danelljan, M., Liu, Y., Tai, Y.W., Tang, C.K., Yu, F.: Segment anything in high quality. Adv. Neural Inf. Process. Syst. **36**, 29914–29934 (2023)
13. Lee, C., Fong, D.Y., Cheung, K.M., Cheng, J.C., Ng, B.K., Lam, T., Mak, K., Yip, P.S., Luk, K.D.: Referral criteria for school scoliosis screening: assessment and recommendations based on a large longitudinally followed cohort. Spine **35**(25), E1492–E1498 (2010)
14. Lee, J., Lee, M., Lee, D., Lee, S.: Hierarchically decomposed graph convolutional networks for skeleton-based action recognition. In: IEEE International Conference on Computer Vision, pp. 10444–10453 (2023)
15. Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., et al.: Grounding dino: marrying dino with grounded pre-training for open-set object detection. In: European Conference on Computer Vision, pp. 38–55 (2024)
16. Luk, K.D., Lee, C.F., Cheung, K.M., Cheng, J.C., Ng, B.K., Lam, T.P., Mak, K.H., Yip, P.S., Fong, D.Y.: Clinical effectiveness of school screening for adolescent idiopathic scoliosis: a large population-based retrospective cohort study. Spine **35**(17), 1607–1614 (2010)
17. Morrissy, R.T., Goldsmith, G., Hall, E., Kehl, D., Cowie, G.: Measurement of the cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error. J. Bone Joint Surg. Amer. Vol. **72**(3), 320–327 (1990)
18. Shi, L., Zhang, Y., Cheng, J., Lu, H.: Skeleton-based action recognition with multi-stream adaptive graph convolutional networks. IEEE Trans. Image Process. **29**, 9532–9545 (2020)
19. Valentin, J., et al.: The 2007 Recommendations of the International Commission on Radiological Protection, vol. 37. Elsevier, Amsterdam (2007)
20. Weinstein, S.L., Dolan, L.A., Cheng, J.C., Danielsson, A., Morcuende, J.A.: Adolescent idiopathic scoliosis. The Lancet **371**(9623), 1527–1537 (2008)
21. Xiong, F., Zhang, B., Xiao, Y., Cao, Z., Yu, T., Zhou, J.T., Yuan, J.: A2j: anchor-to-joint regression network for 3d articulated pose estimation from a single depth image. In: IEEE International Conference on Computer Vision, pp. 793–802 (2019)
22. Xu, Y., Zhang, J., Zhang, Q., Tao, D.: ViTPose: simple vision transformer baselines for human pose estimation. Adv. Neural Inf. Process. Syst. **35**, 38571–38584 (2022)
23. Yan, S., Xiong, Y., Lin, D.: Separated spatial-temporal graph convolutional networks for skeleton-based action recognition. In: AAAI Conference on Artificial Intelligence, vol. **32** (2018)
24. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: AAAI Conference on Artificial Intelligence, vol. **32**, pp. 7444–7452 (2018)
25. Yang, J., Zhang, K., Fan, H., Huang, Z., Xiang, Y., Yang, J., He, L., Zhang, L., Yang, Y., Li, R., et al.: Development and validation of deep learning algorithms for scoliosis screening using back images. Commun. Biol. **2**(1), 390 (2019)
26. Yawn, B.P., Yawn, R.A., Hodge, D., Kurland, M., Shaughnessy, W.J., Ilstrup, D., Jacobsen, S.J.: A population-based study of school scoliosis screening. J. Am. Med. Assoc. **282**(15), 1427–1432 (1999)
27. Zhang, J., Lou, E., Hill, D.L., Raso, J.V., Wang, Y., Le, L.H., Shi, X.: Computer-aided assessment of scoliosis on posteroanterior radiographs. Med. Biol. Eng. Comput. **48**, 185–195 (2010)

Zizhao Peng received the B.Eng. degree in Microelectronic Science and Engineering from Southern University of Science and Technology, Shenzhen, China, in 2021. He is currently pursuing the PhD degree in computing with The Hong Kong Polytechnic University, Hong Kong. His current research interests include posture recognition, graph convolutional neural networks, and deep learning.

Zihan Wang received the B.Eng. degree in Microelectronic Science and Engineering from Southern University of Science and Technology, Shenzhen, China, in 2020. He is currently pursuing the PhD degree in computing with The Hong Kong Polytechnic University, Hong Kong. His current research interests include medical image analysis, deep learning, posture recognition, and 3D medical image generation.

Mengying Sun received the B.Eng. degree in communications engineering from Shandong University, Qingdao, China, in 2021, and the M.Eng. degree in materials and chemical engineering from Southern University of Science and Technology, Shenzhen, China, in 2023. She is currently pursuing the PhD degree in computing with The Hong Kong Polytechnic University, Hong Kong. Her current research interests include image colorization, image synthesis, diffusion models, and deep learning.

Zheng Lv is a Chief Physician graduated with a specialization in Rehabilitation Medicine and Physiotherapy from Jilin University in 2004. From 2004 to 2019, he served in the Department of Rehabilitation Medicine at the First Affiliated Hospital of Harbin Medical University. Currently holding dual leadership roles as Director of Rehabilitation Medicine (encompassing Hyperbaric Oxygen Therapy and Psychosomatic Medicine) at Shenzhen Longgang Central Hospital and Director of the Second Rehabilitation Ward under the hospital group at Longgang Sixth People's Hospital, he is recognized as a Class III High-Level Talent in Shenzhen "Shenlong Elite" program. His academic contributions include leading 4 national/provincial-ministerial research projects and 2 municipal/district-level initiatives, authoring over 10 SCI-indexed publications as corresponding/first author alongside 20+ core journal papers, securing 7 patents (invention, utility model, and design), editing one professional monograph, and receiving two provincial-level new technology awards. His research focuses on central neural plasticity mechanisms and AI-driven rehabilitation technologies.

Yan Wang received the Master's Degree from Southern Medical University, Guangzhou, China, in 2009. She is now associate chief physician in the Affiliated Taian City Central Hospital of Qingdao University, Taian, China. She mainly focuses on the overall rehabilitation assessment and treatment of pregnancy and childbirth, diagnosis and treatment of neurogenic diseases as well as examination of electromyography and sensory threshold determination. She specializes in the treatment of postpartum rectus abdominis separation, pubic symphysis separation, sacroiliac joint disorder, pelvic floor muscle dysfunction, and stress urinary incontinence.

Ping Li received the PhD degree in computer science and engineering from The Chinese University of Hong Kong, Hong Kong, in 2013. He is currently an Assistant Professor with the Department of Computing and an Assistant Professor with the School of Design, The Hong Kong Polytechnic University, Hong Kong. He has published over 260 top-tier scholarly research articles, pioneered several new research directions and made a series of landmark contributions in his areas. He has an excellent research project reported by the *ACM TechNews*, which only reports the top breakthrough news in computer science worldwide. More importantly, however, many of his research outcomes have strong impacts to research fields, addressing societal needs and contributed tremendously to the people concerned. His current research interests include image/video stylization, colorization, artistic rendering and synthesis, computational art, and creative media.

Fengwei An received the PhD degree in Hiroshima University, Japan, in 2013. He was an associate professor at Hiroshima University, Japan, and also worked as a supervising engineer at Panasonic Semiconductor in Japan before joining the School of Microelectronics, Southern University of Science and Technology in March 2019. He is mainly engaged in the fields of high-performance video image processing chips and high-performance AI/digital signal processing chips. He has published several academic papers in top journals and conferences in the field of integrated circuit design (including TCAS-I, TCAS-II, TCSVT, TVLSI, CICC, ESSCIRC, A-SSCC, etc.), 9 Chinese invention patents and 3 Japanese patents.