

Gait Patterns as Biomarkers: A Video-Based Approach for Classifying Scoliosis

Zirui Zhou^{1†}, Junhao Liang^{1†}, Zizhao Peng^{1,2}, Chao Fan¹, Fengwei An¹, and Shiqi Yu^{1*}

¹ Southern University of Science and Technology, Shenzhen, China

² The Hong Kong Polytechnic University, Hong Kong, China

Abstract. Scoliosis presents significant diagnostic challenges, particularly in adolescents, where early detection is crucial for effective treatment. Traditional diagnostic and follow-up methods, which rely on physical examinations and radiography, face limitations due to the need for clinical expertise and the risk of radiation exposure, thus restricting their use for widespread early screening. In response, we introduce a novel video-based, non-invasive method for scoliosis classification using gait analysis, effectively circumventing these limitations. This study presents Scoliosis1K, the first large-scale dataset specifically designed for video-based scoliosis classification, encompassing over one thousand adolescents. Leveraging this dataset, we developed ScoNet, an initial model that faced challenges in handling the complexities of real-world data. This led to the development of ScoNet-MT, an enhanced model incorporating multi-task learning, which demonstrates promising diagnostic accuracy for practical applications. Our findings demonstrate that gait can serve as a non-invasive biomarker for scoliosis, revolutionizing screening practices through deep learning and setting a precedent for non-invasive diagnostic methodologies. The dataset and code are publicly available at <https://zhouzi180.github.io/Scoliosis1K/>.

Keywords: Scoliosis · Gait analysis · Non-invasive screening · Deep learning · Computer vision.

1 Introduction

Scoliosis, a complex spinal deformity characterized by three-dimensional curvature, significantly impacts adolescents' physical well-being and quality of life globally. Clinically, scoliosis is assessed by measuring the Cobb angle, defined as the angle between the upper and lower end vertebrae on standing X-rays (Figure 1 (a)), with a threshold exceeding 10° indicating scoliosis [19,16]. This

[†] Equal contribution.

* Corresponding author: yusq@sustech.edu.cn.

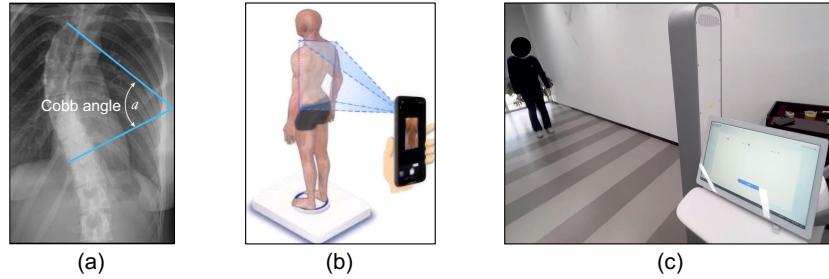


Fig. 1: Comparative overview of scoliosis diagnosis methods: (a) Traditional X-ray examination, the clinical gold standard [18]; (b) Non-invasive analysis of bareback photos [24]; (c) Our proposed gait analysis approach, enabling efficient, large-scale early adolescent screening to identify cases requiring further radiographic investigation, highlighting its non-invasive and privacy-preserving characteristics.

early-stage condition is often asymptomatic, potentially leading to severe health issues if undiagnosed [14]. In China, scoliosis affects approximately 5.14% of school-aged children [6], highlighting the critical need for effective early screening methods.

Traditionally, scoliosis diagnosis and monitoring have relied on physical exams and radiography, which require significant clinical expertise and expose patients to radiation, limiting early, widespread screening. Innovations in deep learning have prompted the exploration of non-invasive scoliosis assessment methods, such as bareback photo analysis [21,24]. However, these alternatives often raise concerns about privacy and efficiency. To address these challenges, we introduce a novel video-based method using gait as a biomarker for scoliosis, eliminating the need for direct bodily exposure (Figure 1 (b) v.s. (c)). Despite its potential, there is a lack of related work benchmarking this promising direction, primarily due to the absence of public datasets and baseline models.

In response, we developed Scoliosis1K—a groundbreaking dataset featuring over 1k adolescents and 400k frames, setting new standards for video-based adolescent scoliosis screening. To respect privacy and enhance usability, we opted for silhouette data and utilized existing gait recognition techniques [1,3,4,9,2] to create ScoNet, demonstrating the viability of gait as a scoliosis biomarker. Furthermore, ScoNet’s initial vulnerability to sample imbalance led to the development of ScoNet-MT, an enhanced multi-task learning model.

This work significantly contributes to the field by (1) creating Scoliosis1K, the first large-scale dataset for video-based scoliosis classification, establishing a new benchmark for the research community; (2) introducing ScoNet, the first baseline model for scoliosis classification through gait analysis, and evolving it into ScoNet-MT to better handle real-world data complexities; and (3) demonstrating that ScoNet-MT exhibits promising diagnostic accuracy for practical applications, underscoring the potential of gait as a non-invasive biomarker for scoliosis and showcasing the transformative impact of deep learning in healthcare diagnostics.

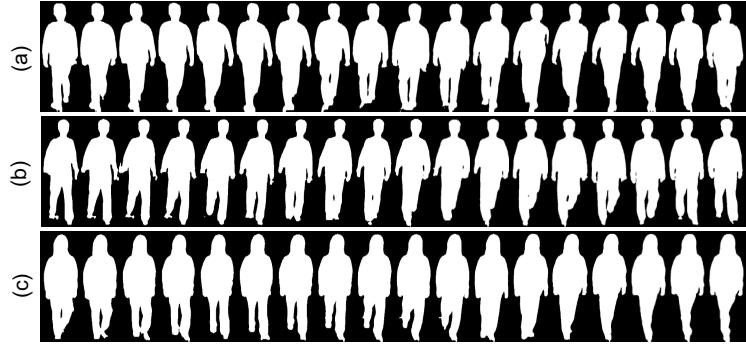


Fig. 2: Silhouettes from the Scoliosis1K dataset: (a) positive, (b) neutral, and (c) negative samples.

2 Dataset

2.1 Overview of Scoliosis1K

To our knowledge, Scoliosis1K is the first large-scale dataset specifically designed for video-based scoliosis classification. The samples are categorized into three diagnostic groups: positive (Cobb angle $> 10^\circ$), neutral (Cobb angle $\approx 10^\circ$), and negative (Cobb angle $< 10^\circ$). The dataset comprises 1,050 adolescent participants from a middle school in China. It includes 447,900 silhouette images from 1,493 video sequences. Figure 2 shows three sequences representing the different categories. The study received approval from the relevant ethics committee, with informed consent obtained from all participants and their guardians.

2.2 Data Collection and Preprocessing

The videos were captured at 720p resolution using a camera. Participants were instructed to walk along a corridor. The camera was positioned 1.4 to 4.2 meters from the participants. Data collection was designed to simulate a controlled yet natural walking environment conducive to accurate biomechanical analysis. Each sequence, containing approximately 300 frames at 15 frames per second, was annotated by scoliosis experts. Experts used well-established scoliosis screening methods, including visual assessments and the Adams forward bend test, to classify participants. Experts did not classify participants by watching videos; instead, they assessed the participants' bodies directly. Subsequently, all videos from a participant were labeled according to their assessed category.

Due to the relative scarcity of positive scoliosis cases, diagnosed individuals were encouraged to contribute multiple sequences. This approach helped mitigate potential class imbalance, enhancing the dataset's analytical reliability and robustness.

During preprocessing, raw video footage was rigorously converted into binary silhouette sequences. Choosing silhouettes serves two primary purposes:

Table 1: Statistics of the Scoliosis1K dataset, highlighting diversity in diagnostic categories and participant demographics.

	Attributes	All	Positive	Neutral	Negative
Scoliosis1K	Number of Participants	1050	176	82	792
	Number of Sequences	1493	493	200	800
	Sex(F/M)	641/409	113/63	49/33	479/313
	Age(years, mean ± std)	15.2 ± 1.5	14.3 ± 1.0	14.0 ± 0.6	15.5 ± 1.5
	Height(cm, mean ± std)	163.2 ± 8.8	161.6 ± 7.1	161.4 ± 6.7	163.7 ± 9.3
	Weight(kg, mean ± std)	51.9 ± 10.7	48.3 ± 8.4	46.7 ± 7.8	53.3 ± 11.1

anonymizing participant data to safeguard privacy, and focusing the deep learning model on the body region rather than the background. These silhouettes, a fundamental component of our dataset, retain critical gait features while discarding irrelevant background information. A detailed description of silhouette extraction is provided in Section 3.1.

2.3 Demographics and Dataset Characteristics

Table 1 provides a demographic and clinical overview of Scoliosis1K, illustrating its comprehensiveness for adolescent scoliosis screening. The dataset’s design emphasizes scale and diversity, encompassing a wide array of participant attributes and gait patterns. This variety enhances the dataset’s potential for training deep models in scoliosis classification based on gait.

2.4 Implications for Scoliosis Research

Scoliosis1K can advance scoliosis diagnosis research in the following aspects:

- **Scale and Scope:** To our knowledge, it is the first large-scale dataset for automatic scoliosis diagnosis, making computer vision-based scoliosis diagnosis feasible.
- **Innovation in Non-Invasive Screening:** Scoliosis1K provides high-quality, annotated silhouette data, addressing the critical need for non-invasive diagnostic tools in scoliosis screening. This fosters innovation by enabling the exploration of methods that prioritize patient safety and privacy.

Furthermore, Scoliosis1K bridges a critical gap in the availability of high-quality, annotated images for non-invasive scoliosis screening. This contribution not only catalyzes innovation in healthcare technology but also lays the groundwork for expansive future research in automated scoliosis diagnosis. The dataset paves the way for improving public health, particularly in regions with limited medical services.

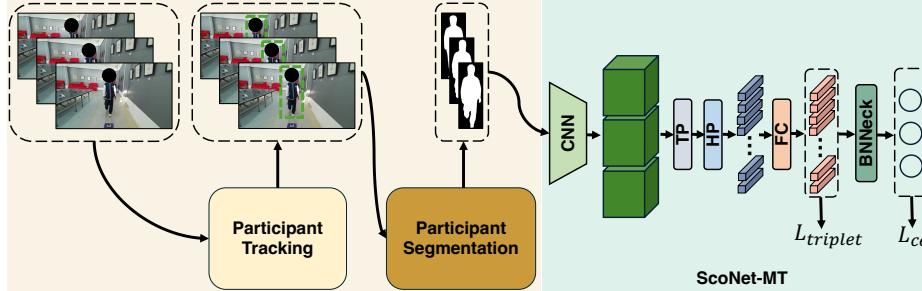


Fig. 3: The Proposed Pipeline: The participant is tracked throughout the video, excluding non-participant entities like clinicians. The participant’s silhouette is then segmented, followed by scoliosis classification using ScoNet-MT based on gait analysis.

3 Methodology

This section presents our novel approach to scoliosis classification through gait pattern analysis in videos. The method involves three key stages: participant tracking, segmentation, and gait-based scoliosis classification, as shown in Figure 3.

3.1 Participants Tracking and Segmentation

We used BYTETracker [25] for precise person tracking in videos. BYTETracker enhances tracking accuracy by evaluating each detection box, recovering occluded objects, and eliminating irrelevant false detections. It employs tracklet similarity to effectively differentiate between the participant and background, ensuring consistent tracking.

After tracking, participant segmentation is performed by feeding the cropped image region into PP-HumanSeg [22,10], a robust deep model for precise human segmentation. PP-HumanSeg employs a deep neural network architecture combining an encoder, Spatial Pyramid Pooling Module (SPPM), and Flexible Lightweight Decoder (FLD) to generate binary masks, or silhouettes. We then normalized the silhouettes to a fixed size following gait recognition methods [7].

3.2 Scoliosis Classification based on Gait

Recognizing gait as a non-invasive biomarker for scoliosis, we introduce ScoNet and its enhanced model, ScoNet-MT, designed to automate the classification process with high efficiency.

ScoNet employs a ResNet-inspired \mathcal{E} architecture to convert participant silhouettes \mathbf{s} into 3D feature maps \mathbf{f} :

$$\mathbf{f} = \mathcal{E}(\mathbf{s}) \in R^{n \times c \times h \times w},$$

where n , c , h , and w represent gait frames, channels, height, and width, respectively. Temporal Pooling (TP) [1] condenses these into essential features z :

$$z = TP(\mathbf{f}) \in R^{c \times h \times w}.$$

Horizontal Pooling (HP) [5] further segments these maps, pooling them into vectors z_s , with global pooling across 16 horizontal segments for comprehensive feature extraction:

$$\mathbf{f}' = \text{maxpool}(z_s) + \text{avgpool}(z_s).$$

A fully connected layer then maps these vectors into the metric space, with BNNeck [11] refining the feature space before final classification using cross-entropy loss:

$$L_{ce} = - \sum_{i=1}^n y_i \log(\hat{y}_i).$$

ScoNet-MT builds on ScoNet by incorporating multi-task learning with a Gait Recognition (GR) task that highlights distinct human motion patterns, reducing bias from non-gait factors. The model uses triplet loss to distinguish subtle gait variations critical for scoliosis classification. In each training batch, N triplets are formed, each comprising an anchor sequence \mathbf{s}_i^a , a positive sequence \mathbf{s}_i^p of the same identity, and a negative sequence \mathbf{s}_i^n of a different identity:

$$L_{triplet} = \sum_{i=1}^N \max(0, \|f(\mathbf{s}_i^a) - f(\mathbf{s}_i^p)\|_2^2 - \|f(\mathbf{s}_i^a) - f(\mathbf{s}_i^n)\|_2^2 + \alpha),$$

where $f(\cdot)$ is the embedding function transforming each sequence into an embedding space, $\|\cdot\|_2$ is the euclidean norm, and α is a margin enforcing separation between positive and negative matches. This setup enhances the model's capacity to discriminate between classes. The total loss function, combining cross-entropy and triplet losses, optimizes ScoNet-MT for accurate scoliosis classification:

$$L_{total} = L_{ce} + L_{triplet}.$$

4 Experiments

4.1 Setup

Evaluation Protocol. The dataset was divided into a training set with 745 sequences and a test set with 748 sequences, maintaining a realistic ratio of positive:neutral:negative samples at 1:1:8 in the training set. Specifically, the sequence counts for the three classes are 74, 74, and 596. Model performance was evaluated using three metrics: accuracy, sensitivity, and specificity, defined as follows:

- **Accuracy:** The proportion of correctly classified samples out of the total number of samples.

Table 2: Comparison of our method with conventional scoliosis screening techniques. * indicates results directly cited from [8]. The best results are highlighted in bold.

Method	Accuracy	Sensitivity	Specificity
Adams Test* [8]	-	84.4%	95.2%
Scoliometer* [8]	-	90.6%	79.8%
ScoNet (Ours)	51.3%	100.0%	33.2%
ScoNet-MT (Ours)	82.0%	99.0%	76.5%

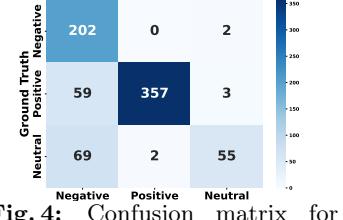


Fig. 4: Confusion matrix for ScoNet-MT.

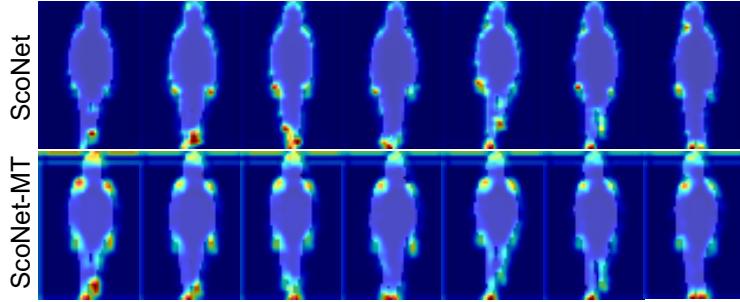
- **Sensitivity:** The proportion of true positives (actual scoliosis cases correctly classified) out of all positive cases.
- **Specificity:** The proportion of true negatives (actual normal cases correctly classified) out of all negative cases.

Implementation Details. Our models, ScoNet and ScoNet-MT, were implemented using PyTorch [13] and OpenGait [3], with input image sizes of 64×44 . The training utilized triplet loss with a margin of 0.2. The positive sequence in each triplet was selected from the same gait sequence as the anchor sequence but consisted of different frames. For experiments, 30 frames were selected from each gait sequence as input. All models were trained using an SGD optimizer [17] with an initial learning rate of 0.1 and a weight decay of 0.0005. The learning rate was reduced by a factor of 10 at 10,000, 14,000, and 18,000 iterations, with training continuing for 20,000 iterations.

4.2 Results

ScoNet-MT shows significant improvements in accuracy and specificity compared to the initial ScoNet model, as shown in Table 2. The model’s accuracy increased by 30.7%, and specificity by 43.3%, highlighting the effectiveness of multi-task learning. Although ScoNet-MT’s sensitivity surpasses traditional methods like the *Adams forward bend test* and the *Scoliometer*, its specificity, while slightly lower, suggests room for further enhancement. These findings highlight the potential of our approach to surpass human expertise in adolescent scoliosis screening. The confusion matrix (Figure 4) offers detailed insights into the model’s precision, particularly in distinguishing between negative cases and others, despite some overlap between positive and neutral classifications.

The heatmaps in Figure 5 show the regions of interest highlighted by our model using the technique from [23]. While ScoNet mainly focuses on the extremities, ScoNet-MT, through multi-task learning, extends its analysis to critical areas like the head and shoulders, aligning with key motion pattern indicators related to scoliosis identified in the literature [12,15,26,20].

**Fig. 5:** Visualization results of our method.**Table 3:** Performance comparison of ScoNet-MT against baseline models and ScoNet. The best results are highlighted in bold.

Method	Accuracy	Sensitivity	Specificity
1) Baseline CNN	45.5%	99.0%	27.0%
2) Baseline CNN-Triplet	56.3%	77.0%	85.0%
3) Baseline CNN-MT	57.8%	96.1%	44.4%
4) ScoNet	51.3%	100.0%	33.2%
5) ScoNet-Triplet	61.2%	100.0%	46.6%
ScoNet-MT (Ours)	82.0%	99.0%	76.5%

Table 4: Test set accuracy of ScoNet and ScoNet-MT, as influenced by various class distributions (positive, neutral, negative) during training.

Pos:Neu:Neg	ScoNet	ScoNet-MT
1:1:2	91.4%	95.2%
1:1:4	88.6%	90.5%
1:1:8	51.3%	82.0%
1:1:16	23.7%	49.5%

4.3 Ablation Studies

Baseline Comparisons. As this is the first video-based solution using a large-scale image dataset and a deep model, we could not find similar methods in the literature for comparison. To demonstrate feasibility and effectiveness, we compared ScoNet-MT with various baseline models in the ablation experiments. The baseline models include: 1) *Baseline CNN*: A basic model trained with cross-entropy loss, excluding horizontal pooling and BNNeck. 2) *Baseline CNN-Triplet*: An improved version of the Baseline CNN, adding triplet loss to enhance category differentiation. 3) *Baseline CNN-MT*: An extension of the Baseline CNN, incorporating multi-task learning with identity information for better generalization. 4) *ScoNet*: Our initial model. 5) *ScoNet-Triplet*: A variation of ScoNet that includes triplet loss to improve feature discrimination. Table 3 presents the performance comparison, showing that ScoNet-MT significantly outperforms all baseline models, particularly highlighting the effectiveness of multi-task learning in enhancing diagnostic precision.

Class-Imbalanced Distribution. Given the inherent imbalance in scoliosis case distribution, we evaluated the performance of ScoNet and ScoNet-MT across different class ratio settings in the training set. The sequence numbers for these settings were 186:186:373 for approximately 1:1:2, 124:124:497 for 1:1:4, and 41:41:663 for 1:1:16, reflecting real-world scenarios where negative cases predominate. Table 4 presents our findings, indicating that ScoNet-MT consistently out-

performs ScoNet across all ratios. This robust performance, even in highly imbalanced conditions, underscores ScoNet-MT’s adaptability and ability to mitigate overfitting, highlighting its superior diagnostic accuracy in realistic settings.

5 Conclusion

Our study demonstrates the effectiveness of using gait as a non-invasive biomarker for scoliosis. The introduction of the Scoliosis1K dataset, along with the development of the ScoNet and ScoNet-MT models, marks a significant advancement in this field, enabling early and accurate scoliosis classification.

The implications of our work are far-reaching, offering a scalable and privacy-preserving diagnostic tool with the potential to revolutionize adolescent scoliosis screening, especially in resource-limited regions. Future work will focus on expanding the diversity of our dataset, identifying additional biomarkers, and exploring more effective methods. Developing a mature large-scale scoliosis screening solution using vision-based gait analysis could benefit a vast number of children, particularly in developing countries.

Acknowledgments. This work was supported in part by the Shenzhen International Research Cooperation Project (Grant No. GJHZ20220913142611021) and ACCESS (AI Chip Center for Emerging Smart Systems), which is sponsored by InnoHK funding, Hong Kong.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

- Chao, H., Wang, K., He, Y., Zhang, J., Feng, J.: Gaitset: Cross-view gait recognition through utilizing gait as a deep set. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(7), 3467–3478 (2021)
- Fan, C., Hou, S., Wang, J., Huang, Y., Yu, S.: Learning gait representation from massive unlabelled walking videos: A benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023)
- Fan, C., Liang, J., Shen, C., Hou, S., Huang, Y., Yu, S.: Opengait: Revisiting gait recognition towards better practicality. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9707–9716 (2023)
- Fan, C., Peng, Y., Cao, C., Liu, X., Hou, S., Chi, J., Huang, Y., Li, Q., He, Z.: Gaitpart: Temporal part-based model for gait recognition. In: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition. pp. 14225–14233 (2020)
- Fu, Y., Wei, Y., Zhou, Y., Shi, H., Huang, G., Wang, X., Yao, Z., Huang, T.: Horizontal pyramid matching for person re-identification. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 8295–8302 (2019)
- Hengwei, F., Zifang, H., Qifei, W., Weiqing, T., Nali, D., Ping, Y., Junlin, Y.: Prevalence of idiopathic scoliosis in chinese schoolchildren: a large, population-based study. *Spine* **41**(3), 259–264 (2016)

7. Iwama, H., Okumura, M., Makihara, Y., Yagi, Y.: The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Transactions on Information Forensics and Security* **7**(5), 1511–1521 (2012)
8. 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)
9. Liang, J., Fan, C., Hou, S., Shen, C., Huang, Y., Yu, S.: Gaitedge: Beyond plain end-to-end gait recognition for better practicality. In: European Conference on Computer Vision. pp. 375–390. Springer (2022)
10. Liu, Y., Chu, L., Chen, G., Wu, Z., Chen, Z., Lai, B., Hao, Y.: Paddle-seg: A high-efficient development toolkit for image segmentation. arXiv preprint arXiv:2101.06175 (2021)
11. Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W.: Bag of tricks and a strong baseline for deep person re-identification. In: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition workshops. pp. 0–0 (2019)
12. Mahaudens, P., Detrembleur, C., Mousny, M., Banse, X.: Gait in adolescent idiopathic scoliosis: energy cost analysis. *European Spine Journal* **18**, 1160–1168 (2009)
13. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* **32** (2019)
14. Payne III, W.K., Ogilvie, J.W., Resnick, M.D., Kane, R.L., Transfeldt, E.E., Blum, R.W.: Does scoliosis have a psychological impact and does gender make a difference? *Spine* **22**(12), 1380–1384 (1997)
15. Kramers-de Quervain, I.A., Müller, R., Stacoff, A., Grob, D., Stüssi, E.: Gait analysis in patients with idiopathic scoliosis. *European Spine Journal* **13**, 449–456 (2004)
16. Reamy, B.V., Slakey, J.B.: Adolescent idiopathic scoliosis: review and current concepts. *American family physician* **64**(1), 111–117 (2001)
17. Ruder, S.: An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747 (2016)
18. Thuaimer, A., Knipe, H., M, E.: Cobb angle: Radiology reference article (2024), <https://radiopaedia.org/articles/cobb-angle?lang=us>, accessed: 27 June 2024
19. Weinstein, S.L., Dolan, L.A., Cheng, J.C., Danielsson, A., Morcuende, J.A.: Adolescent idiopathic scoliosis. *The lancet* **371**(9623), 1527–1537 (2008)
20. Wen, J.X., Yang, H.H., Han, S.M., Cao, L., Wu, H.Z., Yang, C., Li, H., Chen, L.L., Li, N.N., Yu, B.H., et al.: Trunk balance, head posture and plantar pressure in adolescent idiopathic scoliosis. *Frontiers in Pediatrics* **10**, 979816 (2022)
21. 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. *Communications biology* **2**(1), 390 (2019)
22. Yuan, Y., Chen, X., Wang, J.: Object-contextual representations for semantic segmentation. In: European Conference on Computer Vision. pp. 173–190. Springer (2020)
23. Zagoruyko, S., Komodakis, N.: Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. arXiv preprint arXiv:1612.03928 (2016)

24. Zhang, T., Zhu, C., Zhao, Y., Zhao, M., Wang, Z., Song, R., Meng, N., Sial, A., Diwan, A., Liu, J., et al.: Deep learning model to classify and monitor idiopathic scoliosis in adolescents using a single smartphone photograph. *JAMA Network Open* **6**(8), e2330617–e2330617 (2023)
25. Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., Wang, X.: Bytetrack: Multi-object tracking by associating every detection box. In: European Conference on Computer Vision. pp. 1–21. Springer (2022)
26. Zhu, F., Hong, Q., Guo, X., Wang, D., Chen, J., Zhu, Q., Zhang, C., Chen, W., Zhang, M.: A comparison of foot posture and walking performance in patients with mild, moderate, and severe adolescent idiopathic scoliosis. *PLoS One* **16**(5), e0251592 (2021)