InstaGuard: An Intelligent Skating Protection System

Xiangyi Li
Shenzhen Longcheng High School, Shenzhen, China
yi88.aria@gmail.com

Keywords: Skating, Coccyx Safety, Intelligent Protection System, Computer Vision, Gravity Sensor, Multimodal Fusion

Abstract: In skating, the safety of the coccyx has always been one of the main risks faced by athletes. To address this issue, this paper proposes an intelligent protective skating system based on multimodal fusion of gravity and vision. First, this paper introduces the background of coccyx injuries in skating and the shortcomings of existing protective measures. Then, a novel intelligent protection system is proposed, which monitors and analyzes the athlete's posture and motion trajectory in real time by fusing gravity sensors and visual data to detect potential coccyx injury risks. Experimental results show that the system is superior to traditional methods in detection accuracy and reaction speed, and significantly improves the safety of the athlete's coccyx. The system provides an effective solution for coccyx protection in skating and has broad application prospects.

1. Introduction

Skating, a sport that emphasizes both speed and technique, is inherently risky [1–3]. At high speeds, even minor operational errors can cause falls, with common risk factors including poor rink conditions, ill-fitting equipment, incorrect posture, and fatigue. Injuries are typically categorized as contact (e.g., collisions, fractures) or non-contact (e.g., strains, sprains), with the lower limbs—especially knees and ankles—being most vulnerable. Head and torso injuries, though less frequent, can be severe.

Tailbone (coccyx) injuries are particularly common due to direct falls on the buttocks [4,5]. The tailbone, located at the spine's base, supports body weight while sitting. Falls can lead to fractures, intense pain, and difficulty with basic activities such as sitting, standing, or walking. Severe cases may require surgery and can result in complications like sciatica and urinary dysfunction, significantly affecting quality of life [4].

While humans instinctively use their hands to break falls—often resulting in wrist injuries—learning correct falling techniques, such as rolling to the side [5], is essential for reducing tailbone impact. However, developing these techniques demands extensive training and overcoming psychological barriers like fear of falling.

To address this issue, this study proposes an intelligent tailbone protection system powered by deep learning and sensor technologies [6–10]. Equipped with accelerometers and gyroscopes, the system monitors athletes' real-time motion, detects risky movements, and issues alerts (via vibration or sound) when dangerous posture is identified [7]. It encourages safer fall responses—like rolling sideways—and helps prevent injuries. The system also supports wireless data synchronization with smart devices for personalized feedback and training optimization [8].

Such a system offers both individual and societal benefits: it reduces injury rates, lightens healthcare burdens, and promotes safer participation in high-risk sports. It is especially valuable for both professional and amateur athletes seeking to improve performance and longevity while minimizing injury risk [10].

This paper's contributions include:

- ➢ Chapter 1: Introduction and background
- ➢ Chapter 2: Review of deep learning-based sports injury prevention
- ➢ Chapter 3: Design of the intelligent tailbone protection system

  DOI: 10.25236/icceme.2025.022

➢ Chapter 4: System evaluation
➢ Chapter 5: Conclusion and future directions

2. Intelligent Protection System for Enhancing Coccyx Safety Based on Gravity-Vision Multimodal Integration

The use of multimodal sensor technology to enhance coccyx safety has become a research hotspot, driven by advances in technology and increasing public concern for health [11–14]. In activities such as sports, work, and daily life—especially among individuals involved in heavy labor or high-risk sports—coccyx injuries are a common issue. By integrating visual sensors (e.g., cameras) and gravitational sensors (e.g., accelerometers, gyroscopes), it is possible to monitor human posture and motion status in real time, effectively preventing coccyx injuries [12].

2.1. Visual Human Motion Recognition Module

We designed a multi-stage deep learning architecture specifically for fall detection during skating. The system takes as input a sequence of human motion images with the primary goal of accurately detecting falls during sliding and outputs the corresponding action category. Figure 1 illustrates the overall architecture of the visual motion capture module for skating fall detection, showing how input images are processed through multiple neural network stages. Below is a detailed explanation of each component and its role in the overall framework.

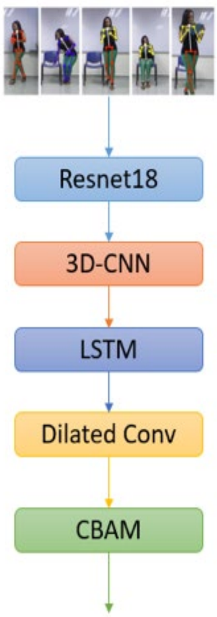

Figure 1 Visual Motion Capture Module.

The input to the network is a sequence of human motion images during skating. In preprocessing, OpenPose extracts human joint coordinates, allowing the model to focus on motion-relevant regions and reduce background noise. The data is first processed by ResNet18, a deep CNN with residual connections that effectively extracts high-level spatial features while mitigating vanishing gradients [15]. These features are then passed to a 3D-CNN, which captures both spatial and temporal information—essential for recognizing dynamic human motion [16]. The resulting spatiotemporal features are input into an LSTM network, which models long-term dependencies in action sequences, aiding in accurate fall detection [17]. Finally, the LSTM outputs are refined using Dilated Convolution, which enlarges the receptive field without increasing computation, enabling the network to capture broader context while preserving detail [18]. As shown in Figure 2, the dilated convolution expands the receptive field without increasing the number of parameters, enabling broader contextual awareness.

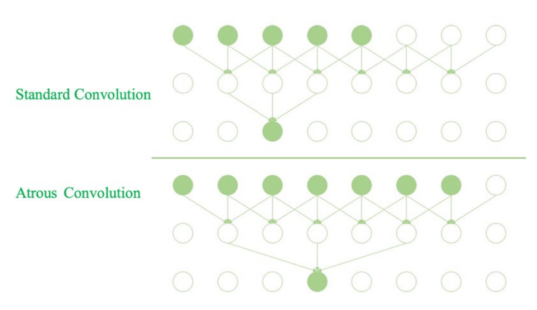

Figure 2 Dilated Convolution.

Finally, the network employs a Channel and Spatial Attention Module (CBAM). CBAM adaptively adjusts channel and spatial weights to emphasize features that are highly relevant to fall recognition while suppressing irrelevant information [19]. The CBAM structure used in our system is illustrated in Figure 3, showing both channel and spatial attention branches. This mechanism significantly improves fall detection performance under complex backgrounds and diverse motions.

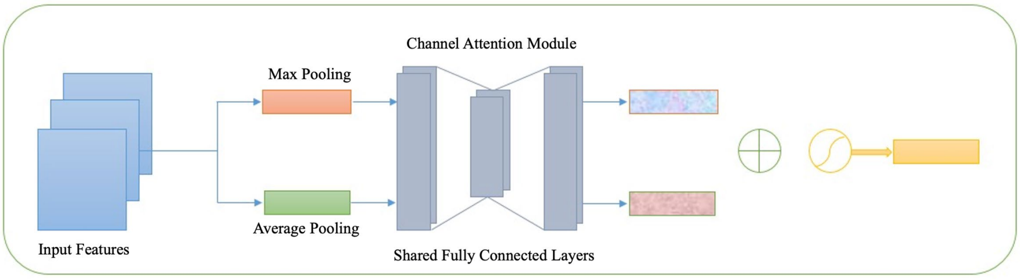

Figure 3 CBAM Structure

This multi-level deep learning architecture—combining ResNet18 for powerful feature extraction, 3D-CNN for spatiotemporal feature modeling, LSTM for temporal dynamics, dilated convolution for receptive field expansion, and CBAM for attention optimization—forms an efficient and comprehensive visual feature extraction module for fall detection in skating, offering real-time and accurate safety assurance.

2.2. Gravity Sensor-Based Motion Recognition Module

We designed a module based on the Temporal Convolutional Network (TCN) architecture specifically to process sequential data collected by gravity sensors during skating, aiming to detect falls. The input consists of sensor data, which is passed through several convolutional and feature extraction layers to output high-dimensional features for classifying motion types.

TCN (Temporal Convolutional Network) is a deep learning architecture tailored for sequence data [20]. The general structure of the TCN model used in this study is shown in Figure 4. Using 1D convolutions and dilated convolutions, TCN captures long-term temporal dependencies while preserving causality (i.e., only using current and past information), making it well-suited for real-time applications like speech recognition, natural language processing, and time series forecasting.

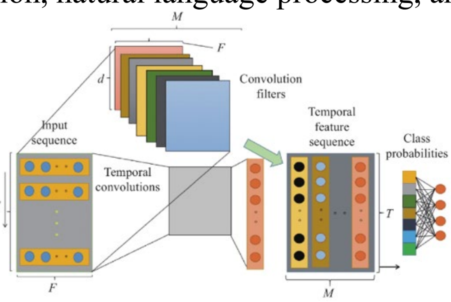

Figure 4 Temporal Convolutional Network (TCN)

We describe our model structure in detail below.

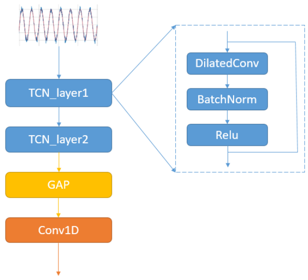

Figure 5 Gravity Sensor Recognition Module

The input consists of time-series data collected by gravity sensors during skating, recording changes in gravitational acceleration at different time points. To improve signal quality, preprocessing such as noise filtering and normalization is applied. The detailed pipeline of the gravity sensor-based motion recognition module is presented in Figure 5.

The preprocessed time-series data is passed through the first TCN layer, which employs dilated convolutions to extract preliminary temporal features across longer time spans without increasing computational cost.

Features extracted from the first layer are fed into a second TCN layer for deeper modeling, capturing more complex temporal patterns in gravitational changes.

The feature map is then processed through a GAP layer to reduce dimensionality and generate a compact, high-level feature vector. This step helps prevent overfitting and provides stable, representative features for downstream classification.

2.3. Multimodal Motion Recognition

The overall network for multimodal motion recognition consists of three main components: Vision Block, Sensor Block, and Adaptive Fusion. Below is an introduction to each component. Figure 6 shows the multimodal motion recognition architecture, including the Vision Block, Sensor Block, and Adaptive Fusion layer.

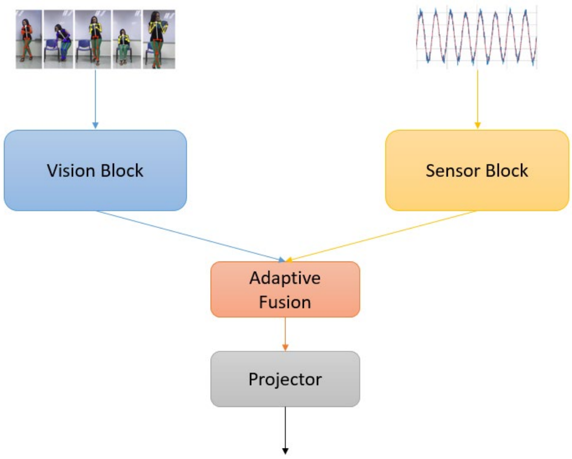

Figure 6 Multimodal Module

This module processes both visual data (e.g., images) and sensor data (e.g., accelerometers). The Adaptive Fusion layer integrates vision and sensor information by learning how to combine them effectively, based on their characteristics. It aligns the two modalities into a shared space and uses attention mechanisms to assign dynamic weights, enabling optimal fusion of features from both sources. This enhances model performance by focusing on the most informative aspects of each input. Below this layer, the Projection Module transforms the fused features into a format suitable for decision-making or further tasks.

3. Model Experimental Results Analysis

The deep learning training environment used in this study includes Ubuntu 20.04, Python 3.8, Keras 2.5.1, TensorFlow 2.2.0, PyTorch 1.13.0, OpenCV 4.2.0, CUDA 11.1, and cuDNN 7.8.5. The hardware environment includes CPU: Xeon E5-2678 v3, GPU: Tesla V100, and 32GB RAM.

3.1. Model Training

The model was trained for 100 epochs with an initial learning rate of 0.001, decaying by 1% every 10 epochs. This strategy helped reduce loss early and fine-tune weights later. The ADAM optimizer was used for its effectiveness with sparse gradients and non-smooth objectives. Key metrics included loss and accuracy, which showed steady improvement throughout training. Loss dropped significantly in early epochs and stabilized at a low level, while accuracy reached 0.95, indicating strong performance and feature extraction. Figure 7 shows the variation of training loss and accuracy over 100 epochs, illustrating the model's convergence behavior. To ensure generalization, validation performance was continuously monitored, and cross-validation was applied to prevent overfitting, confirming the model's robustness on unseen data.

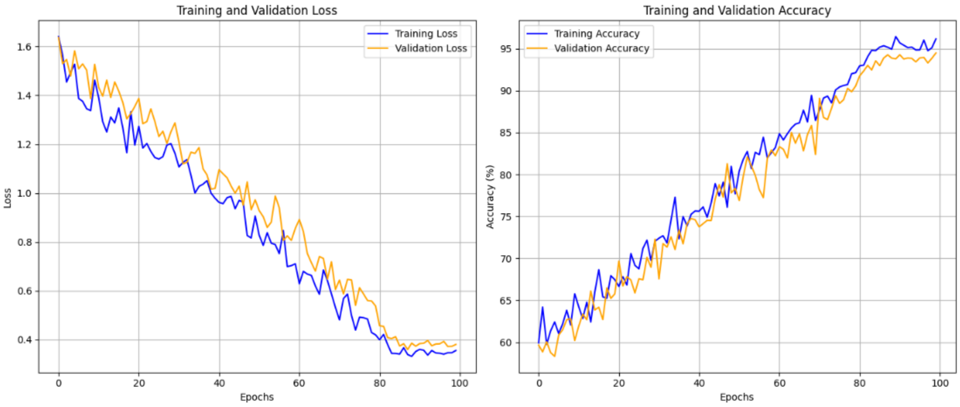

Figure 7 Model Training Process

3.2. Model Testing

After training, the model was evaluated using a confusion matrix across five action categories. The confusion matrix of the five-category classification task is shown in Figure 8, where the diagonal elements represent correct predictions. Normal skating had the highest accuracy with 192 correct predictions, though some were misclassified as standing or falling. Falling actions were also well identified (195 correct), with minor

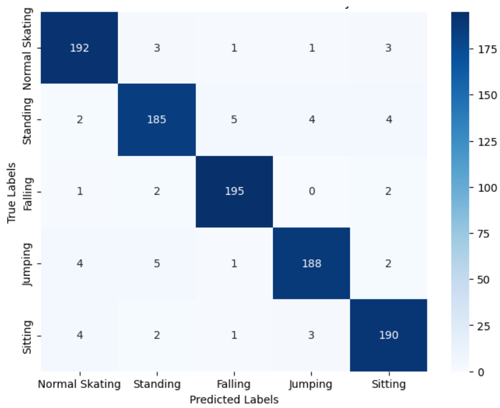

Figure 8 Confusion Matrix

Precision, recall, and F1-score are key metrics for evaluating classification models. To assess performance, we compared the multimodal model with the visual and gravity sensor models. While the visual model outperformed the sensor model, the multimodal model achieved the best results across all metrics, demonstrating superior overall performance.

4. Conclusion and Future Outlook

This paper presents an intelligent ice-skating protection system aimed at preventing coccyx injuries. Chapter 1 outlined the risks and limitations of current protective methods. Chapter 2 reviewed related sensor- and vision-based technologies. Chapter 3 introduced our multimodal system combining gravity and vision data to monitor posture and detect risks in real time. Chapter 4 validated the system's effectiveness, showing superior detection accuracy and response speed compared to traditional methods. Looking forward, the system can be optimized with better sensors and deep learning techniques, and extended to other high-risk sports. In summary, our multimodal system provides a practical and impactful solution for enhancing.

References

[1] Adams, Mary Louise. Artistic impressions: Figure skating, masculinity, and the limits of sport. University of Toronto Press, 2011.

[2] Jiang, Xinyan, et al. "Sports related injury mechanism on ice hockey skills: A system review." Journal of Medical Imaging and Health Informatics 10.5 (2020): 1149-1158.

[3] Šeniglová, Lenka. Injury and its association with training in female youth figure skaters. Diss. University of Saskatchewan, 2011.

[4] White, William D., et al. "The interdisciplinary management of coccydynia: A narrative review." PM&R 14.9 (2022): 1143-1154.

[5] Boyle, Michael. New functional training for sports. Human Kinetics, 2016.

[6] Brown, Christopher M. "Computer vision and natural constraints." Science 224.4655 (1984): 1299-1305.

[7] Li, Guoming, et al. "Practices and applications of convolutional neural network-based computer vision systems in animal farming: A review." Sensors 21.4 (2021): 1492.

[8] Learning, Deep. "Convolutional neural networks." Transfer 12.6 (2015).

[9] Shukla, Vidit, and Shilpa Choudhary. "Deep Learning in Neural Networks: An Overview." Deep Learning in Visual Computing and Signal Processing (2022): 29-53.

[10] Wu Y , Deng Y , Yang J ,et al.AniFaceGAN: Animatable 3D-Aware Face Image Generation for Video Avatars[J].arXiv e-prints, 2022.DOI:10.48550/arXiv.2210.06465.

[11] Zhang, Chi, et al. "Spatio-temporal feature extraction network based multi-performance indicators synergetic monitoring method for complex industrial processes." Expert Systems with Applications (2024): 125052.

[12] Oleotti, Marianna. "Spatio-Temporal Graph Convolutional Neural Networks (ST-GCN) for Semantic Understanding of Environment in Ambient Assisted Living (AAL)." (2023).

[13] Velliangiri, S., and S. J. P. C. S. Alagumuthukrishnan. "A review of dimensionality reduction techniques for efficient computation." Procedia Computer Science 165 (2019): 104-111.

[14] Diraco, Giovanni, et al. "Review on human action recognition in smart living: Sensing technology, multimodality, real-time processing, interoperability, and resource-constrained processing." Sensors 23.11 (2023): 5281.

[15] Liu, Yan, Guo-rong She, and Shu-xaing Chen. "Magnetic resonance image diagnosis of femoral head necrosis based on ResNet18 network." Computer methods and programs in biomedicine 208 (2021): 106254.

[16] Rao, Chengping, and Yang Liu. "Three-dimensional convolutional neural network (3D-CNN) for heterogeneous material homogenization." Computational Materials Science 184 (2020): 109850.

[17] Sherstinsky, Alex. "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network." Physica D: Nonlinear Phenomena 404 (2020): 132306.

[18] Yu, Fisher, and Vladlen Koltun. "Multi-scale context aggregation by dilated convolutions." arXiv preprint arXiv:1511.07122 (2015).

[19] Woo, Sanghyun, et al. "Cbam: Convolutional block attention module." Proceedings of the European conference on computer vision (ECCV). 2018.

[20] Lea, Colin, et al. "Temporal convolutional networks for action segmentation and detection." proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.