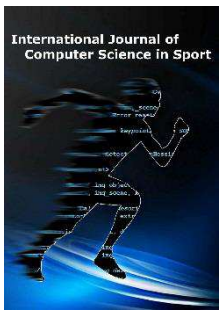
International Journal of Computer Science in Sport

Volume 24, Issue 1, 2025

Journal homepage: http://iacss.org/index.php?id=30

sciendo

DOI: 10.2478/ijcss-2025-0002



Deep Learning with 3D ResNets for Comprehensive Dual-Lane Speed Climbing Video Analysis

Xie, Y.[1,2]*, Mariano, V.*[1]

[1]*College Of Computing and Information Technologies, National University, Manila, Philippines*

[2]*School of Social Management, Jiangxi College of Applied Technology, Ganzhou, China*

Abstract

Analyzing dual-lane speed climbing videos provides critical insights into data-driven performance evaluation in sports climbing. This study introduces an enhanced deep learning approach based on 3D ResNets to classify and analyze speed climbing states. Leveraging an annotated dataset of 872 high-resolution videos covering 15 state combinations, the model integrates optimized 3D convolutions and residual connections, achieving significant improvements in classification accuracy and computational efficiency. With a test accuracy of 92.78%, the model significantly outperforms 2D CNNs and C3D models. Additionally, its lightweight architecture and reduced computational complexity equip it with the potential for real-time deployment in controlled environments. While challenges such as data imbalance and limited generalization remain, this research provides a robust technical framework for speed climbing video analysis and lays the groundwork for broader applications in spatiotemporal modeling and intelligent sports analytics.

KEYWORDS: 3D RESNETS, SPEED CLIMBING, VIDEO ANALYSIS, DEEP LEARNING, SPORTS ANALYTICS.

Introduction

Background and Motivation

Speed climbing has emerged as a highly regarded competitive climbing discipline in recent years, focusing on athletes completing a standard 15.5-meter climbing wall in the shortest possible time (Pandurevic, Draga, Sutor, & Hochradel, 2022), as shown in Figure 1. Unlike disciplines such as lead climbing and bouldering, speed climbing emphasizes high-intensity physical coordination, instantaneous power, and precision (Askari Hosseini & Wolf, 2023). With its inclusion in the 2020 Tokyo Olympics, the sport has garnered increased international attention (International Olympic Committee, 2021; Reveret, Chapelle, Quaine, & Legreneur, 2020).

Figure 1. The International Federation of Sport Climbing (IFSC) World Cup Climbing Men's Speed Climbing Competition site, Chamonix, France, 2018. Photograph by Jan Kriz, via Wikimedia Commons. (https://en.wikipedia.org/wiki/Speed_climbing_wall).

Despite the growing global popularity of speed climbing, traditional performance analysis methods still rely on manual video review and subjective evaluation, which have several limitations: (1) time-consuming processes, (2) accuracy dependent on the evaluator's expertise, and (3) low efficiency in scenarios involving multiple athletes climbing simultaneously on different lanes (Legreneur, Rogowski, & Durif, 2019), such as during training sessions where two or more climbers practice in parallel. While less common, similar situations may also occur in competitions where multiple pairs of climbers compete simultaneously.

The rapid advancement of computer vision has revolutionized sports analysis, with deep learning models excelling in video analysis due to their ability to handle complex spatiotemporal data (Hara, Kataoka, & Satoh, 2018). Traditional CNNs have proven effective in tasks such as pedestrian recognition, enabling efficient object detection in dynamic environments (Liqin, Blancaflor, & Abisado, 2023), and facial expression recognition, where optimized network structures enhance the identification of complex dynamic features (Ding & Mariano, 2023). Expanding upon 2D CNNs, 3D CNNs incorporate temporal dimensions, allowing the extraction of spatial and temporal dynamics (Ji, Xu, Yang, & Yu, 2013). Notably, 3D ResNets employ residual connections to address gradient issues, achieving exceptional performance in video classification and action recognition (Dong, Fang, Li, Bi, & Chen, 2021; Hara et al., 2018; Qiu, Yao, & Mei, 2017).

In speed climbing video analysis, several advancements have been made. Pieprzycki et al. (2023) developed a method to analyze climbing phases, but it remained sensitive to lighting and video quality (Pieprzycki et al., 2023). Pandurevic et al. (2022) utilized human keypoint detection for feature extraction but struggled with low classification accuracy in multi-climber scenarios

(Pandurevic et al., 2022). Reveret et al. (2020) introduced markerless video tracking for 3D motion visualization, achieving high accuracy but facing cost-related limitations (Reveret et al., 2020).

Despite progress, challenges remain. Models show limited adaptability to environmental conditions like low lighting or complex backgrounds (Pandurevic et al., 2022). Furthermore, most research in speed climbing predominantly focuses on technical movement analysis, while practical outcome classifications such as "flash," "slip," and "fall" are often overlooked, limiting their applicability in real-world training contexts (Askari Hosseini & Wolf, 2023; Diez-Fernández, Ruibal-Lista, Rico-Díaz, Rodríguez-Fernández, & López-García, 2023; Pandurevic, Sutor, & Hochradel, 2023; Saul, Steinmetz, Lehmann, & Schilling, 2019). Moreover, the lack of high-quality annotated datasets hinders model training and further advancements (Richter, Beltrán, Köstermeyer, & Heinkel, 2023).

Objectives and Questions

This study aims to address the lack of dedicated algorithms for climbing state classification in speed climbing, particularly under dual-lane scenarios. To this end, a deep learning model based on 3D ResNet is developed for the automated classification of 15 climbing state combinations, including "flash," "slip," "fall," and "empty" (excluding cases where both lanes are "empty"). By optimizing the model architecture, the study seeks to achieve high classification accuracy and computational efficiency, providing reliable support for performance analysis in speed climbing.

The proposed 3D ResNet model, specifically tailored for dual-lane speed climbing scenarios, is hypothesized to effectively capture the spatiotemporal dynamics of speed climbing videos through lightweight design and structural optimization. This approach is expected to deliver precise classification of 15 state combinations while achieving significant improvements in computational efficiency and classification accuracy compared to existing general-purpose video analysis models. By doing so, the study aims to fill a critical gap in the field of speed climbing performance analysis.

Innovations and Contributions

Dataset Contribution. This study developed the first annotated dataset of speed climbing videos comprising 872 high-quality samples. The dataset includes 15 combinations of dual-lane states, such as "flash," "slip," "fall," and "empty," providing valuable resources for future related research.

Model Innovation. The 3D ResNet architecture was improved to adapt to the processing requirements of short video sequences in speed climbing, reducing computational complexity while maintaining high classification performance.

Practical Significance. The study proposed an automated analysis method capable of quickly providing feedback on performance during training. This method offers efficient training assistance for climbing enthusiasts, athletes, and coaches, reducing the subjectivity of manual analysis as well as labor and time costs. It enables the rapid analysis and recording of climbing performances in scenarios with multiple athletes, few coaches, or even no professional coaching.

Methods

Dataset Description

The dataset used in this study comprises 872 dual-lane speed climbing training videos, collected from the routine training sessions of the climbing team at Jiangxi Applied Technology

Vocational College. Data collection was approved by the college's ethics committee (Approval No.: 2024-PYZX-001). All participants were students aged 17 to 22, completed flash climbs in 5.00–7.00 seconds (males) or 7.00–9.00 seconds (females), with longer times for slips or falls. Before participating in the study, all climbers signed informed consent forms, ensuring they understood the research objectives, potential risks, and benefits.

Data Collection

Equipment and Setup. High-definition stationary cameras were used for recording, with a resolution of 1920x1080 pixels and a frame rate of 30 frames per second. This ensured sufficient video clarity to support motion analysis and climbing outcome classification.

Collection Period. Data was collected between March 4, 2024, and May 13, 2024, covering the spring training season.

Recording Environment. All videos were recorded on speed climbing walls that met the standards set by the International Federation of Sport Climbing (IFSC). The dual-lane design adhered to uniform width, slope, and standardized climbing wall layouts (see Figure 2).

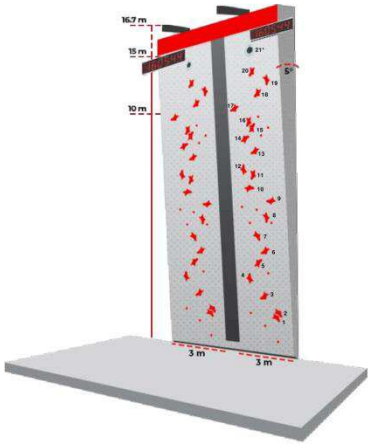


Figure 2. Diagram of the official speed climb wall including numbered hand holds. Adapted from (Walltopia 2020). *Final button or 'hold 21' (Lau, 2021)

Data Content

Climbing Scenarios. The dataset includes the following three dual-lane scenarios (see Figure 3):

- Dual-lane, dual-climber scenario: Two athletes climb simultaneously, one on the left lane and the other on the right lane.
- Dual-lane, left-lane single climber: An athlete climbs on the left lane, while the right lane remains empty.
- Dual-lane, right-lane single climber: An athlete climbs on the right lane, while the left lane remains empty.

Figure 3. Images of the three training situations for speed climbing, (1)Dual-Lane double climbing, both lanes are occupied, (2)Dual-Lane single climbing on the left lane, the left lane is occupied and the right lane is empty, (3)Dual-Lane single climbing on the right lane, the right lane is occupied and the left lane is empty.

Climbing States. As shown in Figure 4, the climbing state for each lane is categorized into the following three types:

- Flash: The athlete successfully taps the top timing panel to complete the climb. The timer stops, and the top light turns green.
- Slip: The athlete fails to accurately grip or step on a hold, causing a brief pause before resuming and completing the climb.
- Fall: The athlete loses contact with the climbing wall and falls. The top light remains red.

Figure 4. Images of the Three States in Speed Climbing: Flash, Slip, and Fall, (1)Flash, the climber has tapped the timer to stop and the display light is green, (2)Slip, the climber's left foot steps out of the air and slips (3)Fall, he climber's feet are dangling in the air, falling downward slowly, and the timer continues to keep time and the display light is red.

Data Annotation

Annotation Rules. Two professional coaches annotated the dataset, labeling the climbing results of both lanes for each video to create independent classification data. The label categories are as follows:

- 1: Flash
- 2: Slip
- 3: Fall
- 4: Empty

Excluding invalid scenarios where both lanes are empty, the dual-lane state combinations result in 15 unique labels (see Table 1). The labels are encoded as integers (0–14) to align with the model training requirements. Multiple rounds of review were conducted to ensure annotation accuracy and consistency.

Table 1. Comparison table for classification, labelling and coding of video status for dual lane speed climbing.

Dual-lane climb state	Annotation	Encode	Videos Quantities
left flash, right flash	1-1	0	237
left flash, right slip	1-2	1	86
left flash, right fall	1-3	2	57
left flash, right empty	1-4	3	72
left slip, right flash	2-1	4	93
left slip, right slip	2-2	5	45
left slip, right fall	2-3	6	20
left slip, right empty	2-4	7	31
left fall, right flash	3-1	8	44
left fall, right slip	3-2	9	10
left fall, right fall	3-3	10	29
left fall, right empty	3-4	11	15
left empty, right flash	4-1	12	90
left empty, right slip	4-2	13	20
left empty, right fall	4-3	14	23

Data Statistics

The dataset in this study consists of 872 videos, covering dual-lane scenarios in speed climbing training. These videos were preprocessed and annotated to form a comprehensive training dataset, providing a robust foundation for training and testing the 3D ResNet model.

Duration Distribution. The average video duration is 11.38 seconds, capturing the full sequence of the athletes' climbing process.

File Size. The average file size is 12,163,891.33 bytes (approximately 11.6 MB), ensuring sufficient resolution for feature extraction.

State Distribution. The dataset includes 15 dual-lane state combinations, with detailed distributions shown in Table 1.

Data Preprocessing

To ensure data quality and efficiency in model training, this study designed a preprocessing workflow:

Frame Extraction. Frames were extracted at 30 frames per second, using multi-threaded extraction via OpenCV and Python's ThreadPoolExecutor (Joshi, Escriva, & Godoy, 2016). This process ensured the capture of dynamic climbing details, enabling the model to learn temporal features effectively.

Frame Cropping. Frames were cropped from 1920 to 1600 pixels in height using OpenCV to focus on athletes' movements and the climbing wall, standardizing the format, reducing irrelevant background, lowering computational load, and improving training efficiency.

Dataset Creation. Frames were resized to 112x112 pixels and organized into temporal sequences with a shape of (30, 112, 112, 3), representing 30 time steps, spatial resolution, and RGB channels. These sequences were paired with labels and stored as NumPy arrays for efficient loading.

After preprocessing, the dataset included the processed results of 872 videos, totaling 10,937 frame sequences. The labels covered 15 states with codes ranging from 0 to 14, fully aligned with the model's classification requirements. The total dataset size was approximately 11.4 GB.

Model Design and Optimization

This study employs an optimized 3D ResNet architecture to classify dual-lane climbing states in speed climbing videos. 3D ResNets, known for their robust spatiotemporal feature extraction, integrate temporal convolutions to capture both spatial details and dynamic temporal patterns, addressing complex motion video challenges (Hara et al., 2018; Qiu et al., 2017). Compared to earlier 3D networks like C3D, 3D ResNets use residual learning to improve training stability and performance (Tran, Bourdev, Fergus, Torresani, & Paluri, 2015). Task-specific optimizations, including reduced computational complexity and enhanced sensitivity to motion, further tailor the model for speed climbing analysis (Du et al., 2021).

Model Architecture

Input Layer. The input to the model is temporal sequence data with the shape (30,112,112,3), where each sequence consists of 30 frames, each being a 112×112 pixel RGB image. The input layer is designed to capture dynamic features of the athlete's climbing process, with a temporal resolution of 30 frames per second. This input dimension balances spatiotemporal resolution with computational resource requirements, optimized for the efficient handling of short temporal sequences by the 3D ResNet model.

Convolutional Layer. The initial convolutional layer employs a 3D convolutional kernel of size 7×7×7 with a stride of (2,2,2), containing 64 filters. The 3D convolution simultaneously processes spatial and temporal dimensions, extracting low-level spatiotemporal features from the input video. The convolutional layer is followed by batch normalization (Batch Normalization) and a ReLU activation function to stabilize the training process and introduce nonlinearity, enhancing the model's expressiveness. A MaxPooling3D layer is then added with a pooling kernel size of 3×3×3 and a stride of (2,2,2), which reduces dimensionality to decrease computational complexity while retaining critical features.

Residual Modules. The core of the model consists of four groups of residual modules, each containing two to three residual units. The structure of each unit includes the following components:

- Convolutional Layers: Each unit uses a 3×3×3 convolutional kernel. The first group of modules contains 64 filters, with subsequent groups increasing the filter count to 128, 256, and 512, respectively. The stride increases progressively from (1,1,1) to (2,2,2).
- Residual Connections: Skip connections are used to directly add the input to the output, forming an identity mapping. This structure effectively mitigates the vanishing gradient problem in deep networks and enhances feature extraction efficiency.
- Customized Optimization: The convolutional kernel sizes and module layers are adjusted to suit the characteristics of short video sequences in speed climbing, improving the model's adaptability to rapid movements and dynamic changes.

Fully Connected Layers. After passing through all residual modules, the feature maps are flattened into a one-dimensional vector and fed into fully connected layers designed as follows:

- Global Feature Extraction: A Dense layer with 512 neurons and ReLU activation is used to extract global features.
- Dropout Layer: A dropout rate of 0.5 is applied to randomly drop part of the neuron outputs, reducing the risk of overfitting.
- Output Layer: The final layer is a Softmax classifier that outputs probabilities for 15 categories, corresponding to the dual-lane climbing state combinations.

The model has a total of 37,425,231 parameters, with 37,415,503 trainable parameters and 9,728 non-trainable parameters. As illustrated in Figure 5, the model leverages multi-level feature extraction and classification processes, efficiently capturing the spatiotemporal features of dual-lane speed climbing videos to classify complex actions accurately.

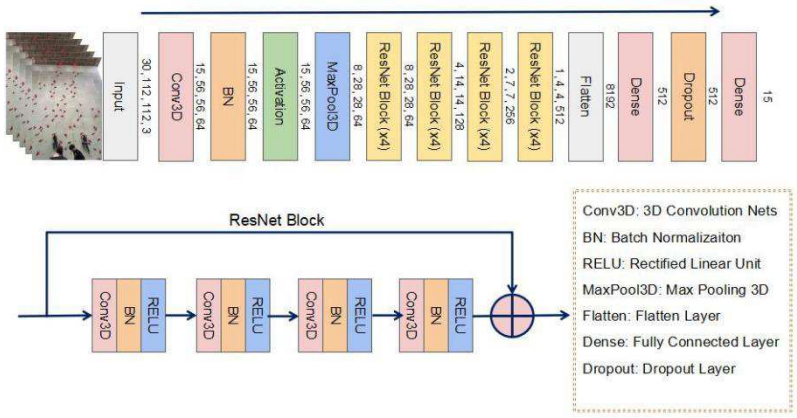


Figure 5. Simplified diagram of the 3D ResNet model architecture designed for this study.

Model Optimization

To accommodate the characteristics of short video sequences in speed climbing while ensuring efficiency and accuracy, the 3D ResNet model was optimized across several dimensions.

Lightweight Design:

- Parameter Optimization: Convolutional kernel sizes and residual module layers were reduced, capping total parameters at 37 million for input size (30, 112, 112, 3), significantly lowering computational demands compared to typical deep 3D networks.
- Module Pruning: Residual modules were adjusted to 2–3 layers, with filter counts incrementing from 64 to 512, balancing computational efficiency and feature

representation.

- Adaptation for Short Video Sequences: Designed for 30-frame inputs, the model efficiently captures dynamic changes without performance loss from redundant data.

Regularization Techniques:

- L2 Regularization: Applied with a coefficient of 0.01 to convolutional and fully connected layers, penalizing large weights to reduce complexity and overfitting.
- Dropout: A 0.5 dropout rate in fully connected layers randomly drops neuron outputs, preventing over-reliance on specific neurons and mitigating overfitting risks.

Loss Function

The model employs the Categorical Cross-Entropy Loss, which is suitable for multi-class classification tasks (Ho & Wookey, 2020). The categorical cross-entropy loss quantifies the difference between the true labels and the predicted probability distribution. By minimizing this loss, the model gradually learns accurate classification rules.

Optimization Algorithm

To efficiently optimize the model parameters, the Adam optimizer was selected with the following configurations (Bock & Weiß, 2019).

Learning Rate. The initial learning rate was set to 0.0001. A ReduceLROnPlateau callback was applied to dynamically reduce the learning rate by a factor of 0.2 when the validation loss showed no significant improvement for two consecutive epochs. The minimum learning rate was set to 0.00001.

Momentum Parameters. Adam incorporates a momentum mechanism for gradient descent, using default values of $\beta_1 = 0.9$ and $\beta_2 = 0.999$. These parameters smooth gradient updates and reduce the impact of local extrema on weight updates.

Robustness. Compared to traditional optimization algorithms such as stochastic gradient descent (SGD), Adam demonstrates superior robustness in handling high-dimensional data and sparse gradients. This enables the model to converge more quickly to a global optimum.

Training Process

Data Splitting

Data splitting was performed using the train_test_split method from the scikit-learn library to ensure that training and testing data were independent, enabling an accurate evaluation of the model's generalization performance:

Splitting Strategy. The dataset was split into 80% training set and 20% testing set randomly.

Fixed Random Seed. A random seed of 42 was used to ensure reproducibility of the split.

Splitting Results:

- Training Set: Consists of 8,749 samples with a shape of (8749,30,112,112,3) for the data and (8749,15) for the labels.
- Testing Set: Consists of 2,188 samples with a shape of (2188,30,112,112,3) for the data and (2188,15) for the labels.

Given that the dataset comprises a total of 10,937 samples, its size was sufficient to directly divide it into training and testing sets for model evaluation.

The validation set was not initially included for the following reasons:

Time and Resource Constraints. Due to the long training cycles of the model, preliminary experiments focused on assessing generalization performance using the test set.

Dataset Coverage. The 80% training and 20% testing split ensured diversity within both sets, and the size of the test set was sufficient for evaluating the model's performance.

Training Configuration

Training Hardware Environment. Due to the large scale of the dataset, the training process required significant computational resources. All training tasks were conducted on a high-performance computer with the following specifications:

- Processor (CPU): 20-core Intel processor
- Graphics Processor (GPU): NVIDIA RTX 3090
- Storage and Memory: 4TB mechanical hard drive and 64GB RAM

Training Hyperparameters. To achieve efficient training and optimal classification performance, the following hyperparameters were tailored to the 3D ResNet architecture and dataset size:

- Batch Size: Set to 8. A smaller batch size alleviates GPU memory pressure and helps capture more sample details in smaller datasets.
- Epochs: The maximum number of training epochs was set to 40. The actual number of epochs was dynamically determined using an early stopping mechanism to prevent overfitting.
- Learning Rate: The initial learning rate was set to 0.0001 to stabilize the optimization process.
- Optimizer: The Adam optimizer was employed for its adaptive learning rate mechanism, which adjusts the step size for each parameter, improving convergence efficiency.

Callbacks

Early Stopping:

- Purpose: Monitors changes in validation loss (val_loss) and halts training when there is no improvement in validation loss over 3 consecutive epochs.
- Parameter Settings:
  (1) monitor='val_loss': Tracks validation loss as the monitoring metric.
  (2) patience=3: Allows 3 epochs without improvement before stopping.
  (3) restore_best_weights=True: Restores the model weights from the epoch with the best validation loss at the end of training.

Learning Rate Scheduler (ReduceLROnPlateau):

- Purpose: Dynamically adjusts the learning rate when validation loss fails to improve over 2 consecutive epochs, reducing it by a factor of 0.2 to avoid local minima.
- Parameter Settings:
  (1) monitor='val_loss': Tracks validation loss as the monitoring metric.
  (2) factor=0.2: Multiplies the learning rate by 0.2 upon triggering.

(3) patience=2: Reduces the learning rate after 2 epochs without improvement.

(4) min_lr=1e-5: Sets a minimum learning rate to prevent excessively slow convergence.

These configurations and mechanisms ensured efficient training, improved model performance, and minimized the risk of overfitting, laying a solid foundation for further experiments.

Evaluation Metrics

To comprehensively evaluate the model's performance, this study employed a range of metrics: accuracy, measuring overall correctness; precision, assessing the avoidance of false positives; recall, evaluating the identification of relevant instances; and F1 score, balancing precision and recall for imbalanced datasets (Vujovic, 2021). Additionally, a confusion matrix was used to provide detailed insights into classification performance, highlighting true positives, false positives, false negatives, and true negatives. Together, these metrics offer a robust framework for assessing the model's effectiveness in classifying speed climbing states and addressing various classification challenges.

Baseline Comparison

In order to verify the effectiveness of the improved 3D ResNets model, a comparison experiment with 2D CNN (2D Convolutional Neural Network) and C3D (3D Convolutional Network) is designed in this study. In this experiment, 2D CNN (Ge, Cao, Li, & Fu, 2020) and C3D (Tran et al., 2015) are used as baseline models to classify 15 climbing states, and their performance will be fully compared with the improved 3D ResNet.

Results

Overall Performance Evaluation

Classification Accuracy

The improved 3D ResNet model achieved a classification accuracy of 92.78% on the test set, demonstrating its capability to effectively distinguish between 15 dual-lane speed climbing state combinations. As shown in Figure 6, from the curve trend, it can be seen that as the number of training rounds increases, the training accuracy gradually rises and the fluctuation decreases, and the testing accuracy also shows a rising trend and stabilizes in the late stage, indicating that the model is continuously optimized in the learning process and does not show obvious overfitting phenomenon.

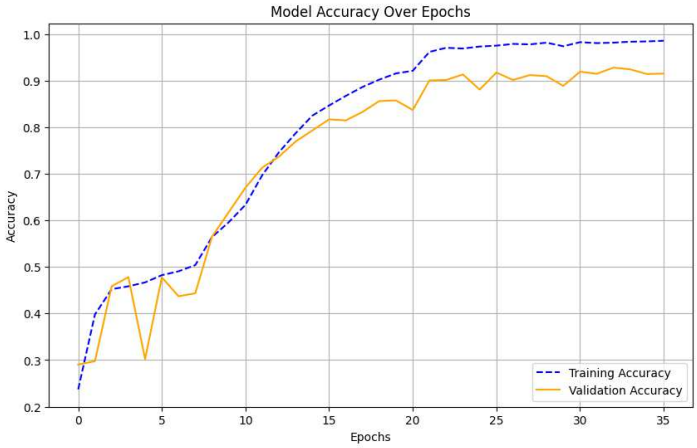


Figure 6. Accuracy Curve of 3D ResNet Model. The horizontal "Epoch" indicates the number of training rounds, starting from 0 and incrementing, demonstrating the model's training iteration. The vertical "Accuracy" represents the model's accuracy in classifying 15 climbing results on the Training and Testing Accuracy sets, ranging from 0 to 1.

Loss Curves

The training and testing loss curves exhibit good convergence (as shown in Figure 7). The continuous decrease of the training loss curve indicates that the model continuously adjusts the parameters to reduce the error during the training process, and the testing loss curve first decreases and then tends to flatten, which further validates the training effect and stability of the model.

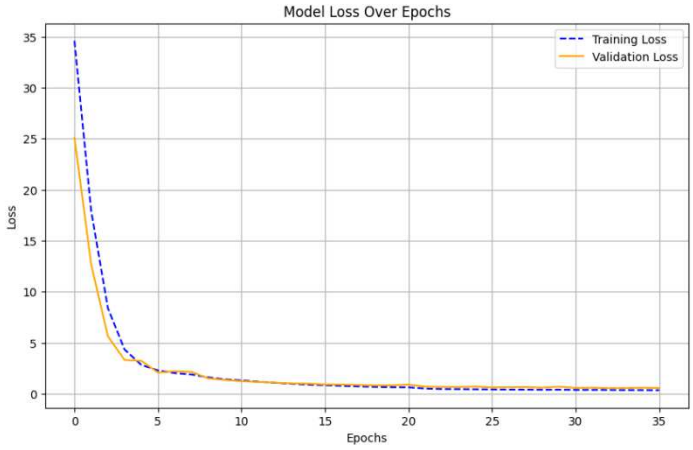


Figure 7. Loss Curve of 3D ResNet Model. The horizontal coordinate is also "Epoch" and the vertical coordinate "Loss" indicates the loss value of the model on the training set and the Testing set , the magnitude of which reflects the model's predicted the degree of difference between the results and the true labels.

Overall Performance Metrics

Using the **classification_report** function, the following overall performance metrics for the model were obtained:

- Precision: 0.93
- Recall: 0.93
- F1 Score: 0.93

These results indicate that the model exhibits stable performance and high accuracy in the classification task.

Detailed Classification Performance

Class-Level Performance Evaluation

On the test set, the model's classification performance is detailed in Table 2, covering each class's Precision, Recall, F1 Score, and the corresponding Support (number of samples).

Table 2. Table of 3D ResNet model classification report

Class	Precision	Recall	F1-Score	Support
0 (L-Flash; R-Flash)	0.91	0.95	0.93	594
1 (L-Flash; R-Slip)	0.85	0.91	0.88	190
2 (L-Flash; R-Fall)	0.95	0.82	0.88	131
3 (L-Flash; R-Empty)	0.95	0.96	0.95	164
4 (L-Slip; R-Flash)	0.94	0.82	0.88	244
5 (L-Slip; R-Slip)	0.93	0.92	0.93	125
6 (L-Slip; R-Fall)	0.89	0.98	0.93	57
7 (L-Slip; R-Empty)	0.95	0.94	0.95	83
8 (L-Fall; R-Flash)	0.93	0.93	0.93	124
9 (L-Fall; R-Slip)	0.95	0.95	0.95	20
10 (L-Fall; R-Fall)	0.84	0.91	0.88	58
11 (L-Fall; R-Empty)	1.00	0.79	0.88	24
12 (L-Empty; R-Flash)	0.99	0.98	0.99	250
13 (L-Empty; R-Slip)	0.99	0.99	0.99	72
14 (L-Empty; R-Fall)	0.98	1.00	0.99	52

Confusion Matrix

The confusion matrix (as shown in Figure 8) provides a visual representation of the 3D ResNet model's classification performance across 15 categories. It highlights the number of true positives (correct classifications), false positives, false negatives, and true negatives for each category.

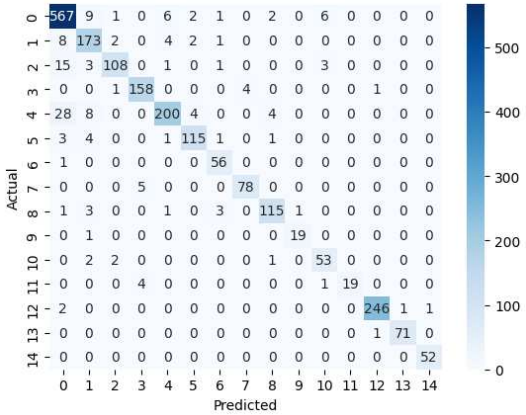


Figure 8. Confusion Matrix for 3D ResNet Model Performance on Speed Climbing Video Analysis. Rows: Represent the actual classes. Columns: Represent the predicted classes. Diagonal Values: Indicate correct classifications, with higher values reflecting better performance. Off-Diagonal Values: Represent misclassifications, identifying areas where the model struggled.

Comparison Experiment Results

To verify the advantages of the improved 3D ResNet model, comparative experiments were conducted against 2D CNN and C3D models, with results shown in Table 3.

Table 3. Performance Comparison of 3D ResNet, 2D CNN and C3D in Terms of Accuracy and Loss.

Model	Accuracy	Loss
3D ResNet	92.78%	0.57
2D CNN	25.62%	2.42
C3D	27.15%	2.51

The comparison demonstrates that the 3D ResNet model significantly outperforms C3D and 2D CNN models in terms of accuracy. This indicates that traditional 2D CNN and C3D models are inadequate for classifying the numerous categories in speed climbing video analysis.

Discussion

Model Strengths and Challenges

Strengths

The improved 3D ResNet demonstrated several notable strengths in its ability to classify 15 dual-lane speed climbing states:

Exceptional Class-Level Performance:

The model achieved outstanding F1 scores close to 0.99 for specific classes such as Class 12 ("L-Empty; R-Flash"), Class 13 ("L-Empty; R-Slip"), and Class 14 ("L-Empty; R-Fall"). These results highlight the model's robustness in recognizing climbing states involving empty lanes. The distinctiveness of these categories, with one lane being empty and less dynamic, likely contributed to the model's high accuracy.

General Robustness Across Categories:

Overall, the model maintained precision, recall, and F1 scores above 0.88 for most classes. This consistent performance reflects the model’s strong adaptability to diverse climbing scenarios, making it well-suited for real-world applications in speed climbing analysis.

Computational Efficiency:

The optimized architecture, with lightweight residual modules and reconfigured kernels, enabled efficient processing. With a sample processing time of approximately 100 milliseconds on an NVIDIA RTX 3090 GPU, the model shows potential for near real-time video analysis, which is critical for both training feedback and competition evaluation.

Challenges

Despite its strengths, the model encountered specific challenges that highlight areas for further improvement:

Class 0 ("L-Flash; R-Flash") Misclassification Issues. Class 0 ("L-Flash; R-Flash") represents the largest category in the dataset, with 594 samples, making it a dominant class in the feature space. While the model performed well in recognizing Class 0, this dominance introduced misclassification challenges. Specifically, Class 1 ("L-Flash; R-Slip"), Class 2 ("L-Flash; R-Fall"), and Class 4 ("L-Slip; R-Flash") had 8, 15, and 28 samples, respectively, misclassified as

Class 0. These errors suggest overlapping features in shared "Flash" states and the model's tendency to prioritize these dominant signals. Furthermore, reverse misclassifications occurred where 9 samples from Class 0 were classified as Class 1, 6 as Class 4, and 6 as Class 10 ("L-Fall; R-Fall"). This bidirectional misclassification highlights the difficulty in distinguishing subtle feature variations between "Flash" and other dynamic states like "Slip" and "Fall," especially in categories with less distinctive features or lower sample representation.

Class 11 ("L-Fall; R-Empty") Recall Issues. The model struggled with Class 11, achieving a recall of only 0.79. This indicates difficulty in distinguishing this category from others, particularly Class 3 ("L-Flash; R-Empty"), with 4 misclassified samples. The shared "R-Empty" state likely caused feature overlap, complicating the classifier's decision-making process. Additionally, the limited representation of Class 11 in the dataset (24 samples) constrained the model's ability to generalize effectively.

Model Optimization and Innovations

This study incorporates task-specific improvements to the traditional 3D ResNet structure to meet the requirements of speed climbing video classification:

Lightweight and Efficient Design. The model's parameters were reduced to 37 million, significantly lower than the original 3D ResNet (63 million) (Tran et al., 2015) and C3D (79 million) models (Hara et al., 2018). Simplified residual modules, reconfigured kernels, and constrained fully connected layers collectively enhanced computational efficiency. This lightweight architecture enables deployment on a range of hardware, including high-performance laptops and mobile GPUs (Foo, Gong, Fan, & Liu, 2023).

Advanced Spatiotemporal Feature Extraction. Hierarchical kernels (7×7×7 in initial layers, 3×3×3 in residual modules) were employed to effectively capture both global and local climbing features, addressing the complexities of speed climbing video analysis. This multi-scale feature extraction, supported by prior studies (Hara et al., 2018; Kong, Satarboroujeni, & Fu, 2016; Tran et al., 2015), improves the model's ability to distinguish nuanced action categories and temporal dependencies.

Enhanced Classification Accuracy and Real-Time Potential. With Table 3, the optimized 3D ResNet achieved a test accuracy of 92.78%, significantly outperforming 2D CNN (25.62%) and C3D (27.15%). Its loss convergence (0.57) further demonstrated stability and efficiency during training. The model processes each sample in approximately 100 milliseconds on an NVIDIA RTX 3090 GPU, highlighting its potential for near real-time applications in both training and competition settings.

Limitations and Future Improvements

Data Imbalance. The dataset's skewed sample distribution posed challenges, particularly for underrepresented classes like Class 9 ("L-Fall; R-Slip") and Class 11 ("L-Fall; R-Empty"). These smaller sample sizes hindered the model's training and reduced classification accuracy for these categories. Augmentation or GANs could improve sample diversity (Johnson & Khoshgoftaar, 2019; Masko & Hensman, 2015).

Lack of Validation Set. The absence of a validation set limited intermediate evaluations during training. Future work should incorporate one to refine hyperparameter tuning (Ghosh et al., 2024).

Generalization Capability. The dataset, collected from a single venue and athlete group, limited adaptability. Broader data sources would enhance generalization to diverse environments (Krawczyk, 2016).

Conclusion

This study introduced an efficient analysis method for speed climbing videos using an optimized 3D ResNet architecture, achieving a classification accuracy of 92.78% and outperforming baseline models like 2D CNN and C3D. The model demonstrated strong performance in capturing the dynamic features of speed climbing, validating its effectiveness in analyzing complex spatiotemporal patterns and providing reliable technical support for this domain.

Future research can focus on the following directions:

Cross-Scenario Adaptability. Enhancing the model's generalization by diversifying data sources across environments, age groups, and skill levels to expand its application in real-time, multi-scenario analysis (Mandour, 2024).

Deployment on Mobile Devices. Developing lightweight models optimized for laptops and smartphones to facilitate real-time analysis in resource-constrained settings (Ahmed & Nyarko, 2024; Kim, Yu, & Xiong, 2024).

Expanded Application Domains. Extending this approach to other sports like swimming, skiing, and running to validate adaptability and support intelligent sports research (Tao & Long, 2023).

In summary, this research not only advances speed climbing video analysis but also broadens technical perspectives in spatiotemporal feature analysis, offering a robust foundation for future studies in intelligent sports.

References

Ahmed, A. A., & Nyarko, B. (2024). Smart-Watcher: An AI-Powered IoT Monitoring System for Small-Medium Scale Premises. *2024 International Conference on Computing, Networking and Communications (ICNC)*, 139–143.

Askari Hosseini, S., & Wolf, P. (2023). Performance indicators in speed climbing: Insights from the literature supplemented by a video analysis and expert interviews. *Frontiers in Sports and Active Living*, *5*. https://doi.org/10.3389/fspor.2023.1304403

Bock, S., & Weiß, M. (2019). A Proof of Local Convergence for the Adam Optimizer. *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Diez-Fernández, P., Ruibal-Lista, B., Rico-Díaz, J., Rodríguez-Fernández, J. E., & López-García, S. (2023). Performance Factors in Sport Climbing: A Systematic Review. *Sustainability*, *15*, 16687.

Ding, X., & Mariano, V. Y. (2023). Research on expression recognition algorithm based on improved convolutional neural network. *International Conference on Computer, Artificial Intelligence, and Control Engineering (CAICE 2023)*, *12645*, 787–792. SPIE.

Dong, M., Fang, Z., Li, Y., Bi, S., & Chen, J. (2021). AR3D: Attention Residual 3D Network for Human Action Recognition. *Sensors*, *21*, 1656.

Du, X., Li, Y., Cui, Y., Qian, R., Li, J., & Bello, I. (2021, September 3). *Revisiting 3D ResNets for Video Recognition*. arXiv.

Foo, L. G., Gong, J., Fan, Z., & Liu, J. (2023). *System-Status-Aware Adaptive Network for Online Streaming Video Understanding*. 10514–10523.

Ge, Z., Cao, G., Li, X., & Fu, P. (2020). Hyperspectral Image Classification Method Based on 2D–3D CNN and Multibranch Feature Fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 5776–5788.

Ghosh, K., Bellinger, C., Corizzo, R., Branco, P., Krawczyk, B., & Japkowicz, N. (2024). The class imbalance problem in deep learning. *Machine Learning*, *113*, 4845–4901.

Hara, K., Kataoka, H., & Satoh, Y. (2018). *Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?* 6546–6555.

Ho, Y., & Wookey, S. (2020). The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling. *IEEE Access*, *8*, 4806–4813.

International Olympic Committee. (2021, August 6). Women's Combined Finals—Climbing | Tokyo 2020 Replays. Retrieved March 3, 2024, from International Olympic Committee website: https://olympics.com/en/video/women-s-combined-finals-climbing-tokyo-2020-replays

Ji, S., Xu, W., Yang, M., & Yu, K. (2013). 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*, 221–231.

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, *6*, 27.

Joshi, P., Escriva, D. M., & Godoy, V. (2016). *OpenCV By Example*. Packt Publishing Ltd.

Kim, D., Yu, X., & Xiong, S. (2024, August 5). *A Practical and Robust Skeleton-Based Artificial Intelligence Algorithm for Multi-Person Fall Detection on Construction Sites Considering Occlusions* [SSRN Scholarly Paper]. Rochester, NY: Social Science Research Network.

Kong, Y., Satarboroujeni, B., & Fu, Y. (2016). Learning hierarchical 3D kernel descriptors for RGB-D action recognition. *Computer Vision and Image Understanding*, *144*, 14–23.

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*, *5*, 221–232.

Lau, E. (2021). *Identifying physiological demands of Speed Climbing within a sample of recreational climbers*. https://doi.org/10.13140/RG.2.2.18266.06089

Legreneur, P., Rogowski, I., & Durif, T. (2019). Kinematic analysis of the speed climbing event at the 2018 Youth Olympic Games. *Computer Methods in Biomechanics and Biomedical Engineering*, *22*, S264–S266.

Liqin, L., Blancaflor, E., & Abisado, M. (2023). Research on Pedestrian Recognition Based on Deep Learning. *2023 5th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, 18–22.

Mandour, H. Y. (2024). Artificial Intelligence in Measurement and Evaluation for Athletes. *International Sports Science Alexandria Journal*, *6*, 7–9.

Masko, D., & Hensman, P. (2015). *The Impact of Imbalanced Training Data for Convolutional Neural Networks*. Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-166451

Pandurevic, D., Draga, P., Sutor, A., & Hochradel, K. (2022). Analysis of Competition and Training Videos of Speed Climbing Athletes Using Feature and Human Body Keypoint Detection Algorithms. *Sensors*, *22*, 2251.

Pandurevic, D., Sutor, A., & Hochradel, K. (2023). Towards statistical analysis of predictive parameters in competitive speed climbing. *Sports Engineering*, *26*, 1–12.

Pieprzycki, A., Mazur, T., Krawczyk, M., Krol, D., Witek, M., & Rokowski, R. (2023). *Computer-Aided Methods for Analysing Run of Speed Climbers*. https://doi.org/10.20944/preprints202302.0166.v1

Qiu, Z., Yao, T., & Mei, T. (2017). *Learning Spatio-Temporal Representation With Pseudo-3D Residual Networks*. 5533–5541.

Reveret, L., Chapelle, S., Quaine, F., & Legreneur, P. (2020). 3D Visualization of Body Motion in Speed Climbing. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.02188

Richter, J., Beltrán, R., Köstermeyer, G., & Heinkel, U. (2023). Climbing with Virtual Mentor by Means of Video-Based Motion Analysis: *Proceedings of the 3rd International Conference on Image Processing and Vision Engineering*, 126–133. Prague, Czech Republic: SCITEPRESS - Science and Technology Publications.

Saul, D., Steinmetz, G., Lehmann, W., & Schilling, A. F. (2019). Determinants for success in climbing: A systematic review. *Journal of Exercise Science & Fitness*, *17*, 91–100.

Tao, T., & Long, J. (2023). RETRACTED ARTICLE: Simulation of video image detection in leisure sports tourism industry based on convolutional neural network. *Soft Computing*. https://doi.org/10.1007/s00500-023-08426-z

Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). *Learning Spatiotemporal Features With 3D Convolutional Networks*. 4489–4497.

Vujovic, Z. (2021). Classification Model Evaluation Metrics. *International Journal of Advanced Computer Science and Applications*, *Volume 12*, 599–606.