THESIS APPLIED INFORMATION TECHNOLOGY (2024-2025)

AI judge for recognition of jump rope skills in videos.

Mike De Decker
mike.dedecker@student.hogent.be

Promotor: Ms. L. De Mol & Mr. T. Parmentier
Co-promotor: Mr. D. Plummer (Case Western Reserve University / NextJump)

University of Applied Sciences and Arts Ghent, Valentin Vaerwyckweg 1, 9000 Ghent, Belgium

Abstract

Judging the difficulty of jump rope freestyles at high competitive levels, is prone to human errors. It is hard to calculate skill levels, watching multiple athletes at the same time, seeing all actions, skill modifiers or rope manipulations. Even though a routine consists of forty to sixty skills, wrongly assigning a single level may impact the ranking, deciding national or international selections. In order to correctly assign levels, difficulty judges, at higher level competitions, are allowed to review the routine at slower speeds in order to increase the accuracy of assigned scores. This is why this research looked for a way to recognize skills in a video. The scope has been limited to recognizing double dutch single freestyles. The solution provided in this research includes a sequential execution of three steps. The first step involves localizing the athletes in order to crop them out of the video. This cropped video can be used in the second step, namely segmenting the video into individual skill sections. Finally, each individual skill section can then be fed into the recognition model, predicting all aspects of the skills performed by the athletes. Training on a skewed limited dataset of less than an hour, containing about 2500 skills, show that more occurring skills reach an accuracy between 80-99%, while the lesser occurring skills reach a limited accuracy or even none at all. Mapping the skill predictions to scores, averages to minus 20.94% point difference, compared to the score assigned by judges. Populating the (train and test) dataset with more examples, should greatly increase the accuracy. Not only is this interesting for jump rope, but for other sports or movement analysis in general.

Specialisation: AI- & Data Engnineer

Keywords: Computer vision, Machine Learning, Neural networks, Human Activity Recognition, video classification, YOLO, YOLOv11, MViT, Multiscale Vision Transformer, Swin Transformer, Jump Rope, Rope Skipping, judging, sport

Source code: `https://github.com/mikeddecker/judge`

1. Introduction

Jump rope is an evolving sport. Year after year, an increasing amount of high-level competitors are pushing the limits of jump rope. This results in new skills, new combinations, better physiques, better rope material, and faster movements. For the judges to keep up with the jumpers and to correctly assess scores to a routine, Double Dutch freestyles, one of the jump rope disciplines, are reviewed at half speed in International competitions or even at nationals in Belgium. Head judges around the world question the best way to judge athletes correctly so as to give an accurate and objective ranking in competitions.

Thus, this research explored the possibility to recognize skills in double dutch single freestyles. The proposed solution in this research includes a sequential execution of three steps, namely localizing the athletes, segmenting the video and recognizing the action. As acquired recordings of videos lose a lot of information on surroundings, it is advised to zoom in on the athletes by cropping them out. These cropped videos can than be used in the segmentation step, which splits the video in various individual skill sections. This way, analysis on full videos can be performed. Thirdly, by labeling the different aspects present in the section, a recognition model can be trained, to predict skills in segmented sections.

Finally, predicted skills can be mapped to their corresponding levels and scores, allowing for judges to review scores on competitions. Not only is this interesting for jump rope, but for other sports like figure skating, synchronized swimming, gymnastics or movement analysis in general.

2. Jumper Localization

Ultralytics provides an easy-to-use pre-trained implementation of the YOLOv11 model for predicting people and objects in images which can be fine-tuned for specific use cases. Fine-tuning was needed as spectators, also humans, were also included in the predictions.

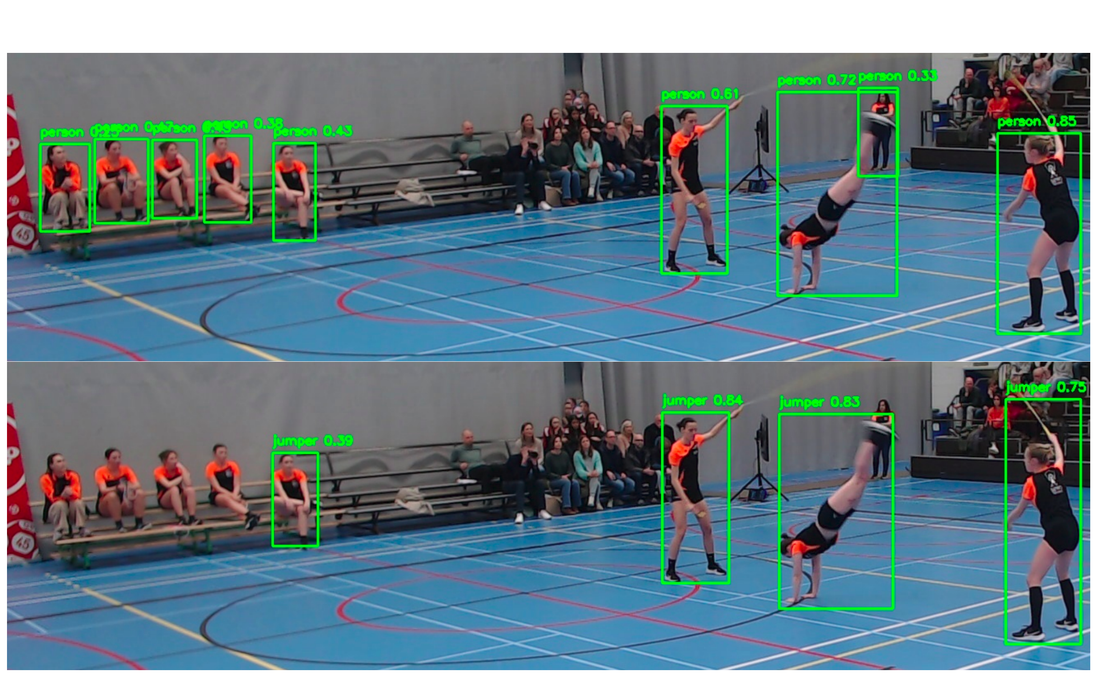

Figure 1: Raw YOLOv11 nano predictions compared to a prediction using fine tuned parameters. The prediction on this video still includes a spectator.

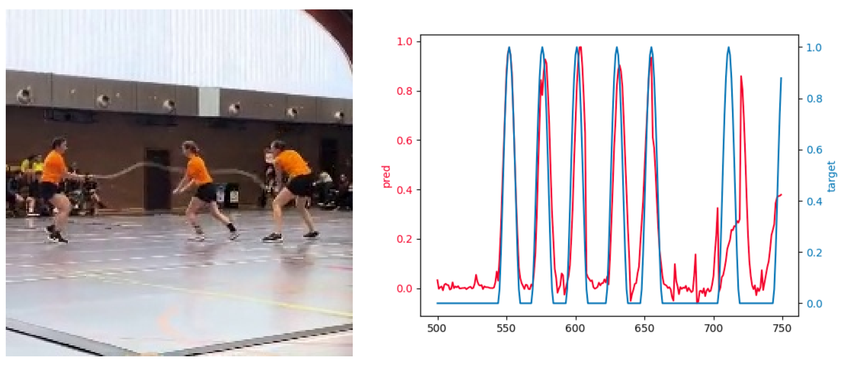

Figure 2: Image on the left: Crop example on a video. (224x224 pixels)
Plot on the right compares the predicted split values points, compared to the label.

3. Action segmentation

Using the cropped videos, actions can be isolated from the video. PyTorch has a pre-trained implementation of the Multiscale Vision Transformer (MViT). The pre-trained weights require an input of (1, 3, 16, 224, 224) or (batch size, channels, timesteps, height, width). Thus, a video is split into sequential sections of 16 frames, with a splitvalue between 0 and 1 for each frame. It then predicts 1 output value for each frame. This is illustrated in the segment prediction plot.

4. Skill recognition

After segmenting the video, each skills section can be predicted using the Multiscale Vision Transformer or the Swin Transformer. However, skills have varying length, which requires frames to be duplicated or skipped in order to create equally sized inputs of (1, 3, 16, 224, 224). Now, for each of the 13 jump rope labels, a prediction can be made. This includes the skill, the type of turning, the number of rotations, first turner, second turner etc. Training on a skewed and limited dataset of about one hour, correctly predicts the most frequent skills between 80-99%, while the lesser occurring skills reach a limited accuracy or none at all.

Click or use the link in the Readme of the repo for a video example.

After predicting the skills in a video, levels and scores can be assigned to the routine, which currently differ on average by minus 20.94%.

5. Conclusions

On competitions which use the AI assistant, the actual setting can be adapted in order to have no or less spectators influencing the localization predictions. This then increases further predictions, relying on the localization. Seeing high accuracies on frequent skills, or even starting accuracy on low to mid frequent skills, indicate the possibility of AI recognizable skills. However, a 20% gap with the scores assigned by judges is to large to consider usable on competitions.

6. Future research

This research can be enhanced by other experiments using more models, better metrics for action segmentation, more skill labels of the less occurring skills and skill aspects. Additionally, a benchmark can be created by comparing the score assigned by judges with the ground truth and the AI-models, deciding possible usage at competitions.

To conclude, this research showed the possibilities of the proposed three step architecture for recognizing skills in a full routine on a limited dataset. While results aren't perfect, great effort has been put in the proof of concept, enabling future research, additional labels and additional model experiments.

HO
GENT