Sumo Action Classification Using Mawashi Keypoints

Yuki Utsuro[1][a], Hidehiko Shishido[2][b] and Yoshinari Kameda[2][c]
[1]*Master's Program in Intelligent and Mechanical Interaction Systems, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki, Japan*
[2]*Center for Computational Sciences, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki, Japan*

Keywords: Action Classification, Action Recognition, Extended Skeleton Model, Pose Estimation, Pose Sequences, Computer Vision in Sports, Sumo, Japanese Wrestling.

Abstract: We propose a new method of classification for kimarites in sumo videos based on kinematic pose estimation. Japanese wrestling sumo is a combat sport. Sumo is played by two wrestlers wearing a mawashi, a loincloth fastened around the waist. In a sumo match, two wrestlers grapple with each other. Sumo wrestlers perform actions by grabbing their opponents' mawashi. A kimarite is a sumo winning action that decides the outcome of a sumo match. All the kimarites are defined based on their motions. In an official sumo match, the kimarite of the match is classified by the referee, who oversees the classification just after the match. Classifying kimarites from videos by computer vision is a challenging task. There are two reasons. The first reason is that the definition of kimarites requires us to examine the relationship between the mawashi and the pose. The second reason is the heavy occlusion caused by the close contact between wrestlers. For the precise examination of pose estimation, we introduce a wrestler-specific skeleton model with mawashi keypoints. The relationship between mawashi and body parts is uniformly represented in the pose sequence with this extended skeleton model. As for heavy occlusion, we represent sumo actions as pose sequences to classify the sumo actions. Our method achieves an accuracy of 0.77 in action classification by LSTM. We confirmed that the skeleton model extension by mawashi keypoints improves the accuracy of action classification in sumo through the experiment results.

1 INTRODUCTION

Japanese wrestling sumo is a two-person combat sport. Sumo wrestlers wear a mawashi, a loincloth fastened around their body. Two wrestlers grapple and fight against each other in a circular sumo ring called dohyo. They are allowed to hold the opponent's mawashi to control the opponent's body. Therefore, during a sumo match, the wrestlers are in close contact with each other (Figure 1). This contact causes heavy occlusion. One of the unique features of sumo is the mawashi. The mawashi is tightened so that it adheres closely to the wrestler's body.

A kimarite is a sumo winning action that decides the outcome of a sumo match. All the kimarites are defined based on their motions. The definition of sumo actions involves the positional relationship between the hands of one wrestler and the mawashi of the other. Therefore, examination of this relationship is an indispensable property for discriminating actions. In an official sumo match, the kimarite of the match is classified by the referee, who oversees the classification just after the match.

Classifying kimarites from videos by computer vision is a challenging task. There are two reasons for this. The first reason is that the definition of kimarites requires us to examine the relationship between the mawashi and the pose. The second reason is the heavy occlusion caused by the close contact between wrestlers.

We propose a new method of classification for kimarites in sumo videos based on kinematic pose estimation. For the precise examination of pose estimation, we newly introduce a wrestler-specific

[a] https://orcid.org/0009-0008-7166-2235
[b] https://orcid.org/0000-0001-8575-0617 (Currently at Soka University)
[c] https://orcid.org/0000-0001-6776-1267

Utsuro, Y., Shishido, H. and Kameda, Y.
Sumo Action Classification Using Mawashi Keypoints.
DOI: 10.5220/0012337000003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 2: VISAPP*, pages 401-408
ISBN: 978-989-758-679-8; ISSN: 2184-4321

skeleton model with mawashi keypoints. The relationship between mawashi and body parts is uniformly represented as the pose of this extended skeleton model. As for handling heavy occlusion, we represent sumo actions as pose sequences to classify the sumo actions so that the later classification procedure can achieve better classification performance.

The first contribution of our research is the new skeleton model with the mawashi keypoints. Since the mawashi is tightened to the wrestler's body, it can be regarded as a part of the body. We extended the existing human skeleton model in OpenPose (Cao et al., 2019) by adding two new keypoints on the mawashi worn by the wrestler (Figure 2). This extension makes it possible to uniformly represent the relationship between the mawashis and the body parts within a single pose.

The second contribution is the utilization of pose sequences for the representation of sumo actions so that the kimarite classification procedure can achieve better classification performance. The kimarite, the last action of a match that decides the outcome of the match, lasts approximately one-second movement of the wrestlers. Sumo referees can classify the action by observing the one-second movement. Even in situations with heavy occlusions during matches, action classification should be successfully made by referring to the pose sequence that might miss some parts in some poses. The pose sequences are classified by LSTM (Hochreiter and Schmidhuber, 1997).

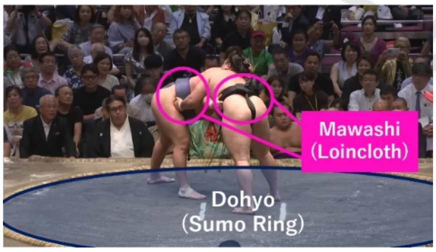

Figure 1: An overview of sumo. The loincloths that wrestlers wear are called "mawashi." The ring is called "dohyo."

2 RELATED WORKS

2.1 Action Classification in Sports

Methods for action classification from specific sports videos often differ from those for daily-action classification due to the need to consider the unique characteristics of the sport. Regarding action classification in sports videos, studies have been reported focusing either on a single player or multiple players. Studies that focus on the action of just one player in sports are reported across various sports such as table tennis (Kulkarni and Sheony, 2021), basketball (Zakharchenko, 2020; Liu et al., 2021), ice hockey (Fani et al., 2017; Vats et al., 2019), karate (Guo et al., 2021), taekwondo (Liang and Zuo, 2022; Luo et al., 2023), figure skating (Liu et al., 2021; Liu et al., 2020), and tennis (Vainstein et al., 2014). In these sports, little occlusion occurs between players. These studies show that if the action is made by a single player and little occlusion is found in the sports video, it is feasible to classify actions.

There are also studies on multi-player action classification in some sports, such as volleyball (Gavrilyuk et al., 2020; Azar et al., 2019), hockey (Rangasamy et al., 2020), soccer (Gerats, 2020), and badminton (Ibh et al., 2023). These studies use information from multiple players as input, but no heavy occlusions occurred for the players under observation. This is because close contact between players does not frequently happen in these sports.

In sports where two athletes are in close contact, leading to heavy occlusion, a study has been reported on action classification in freestyle wrestling (Mottaghi et al., 2020). This study classifies actions using features extracted from the silhouette of grappling wrestlers. This method employs SVM or k-NN as a classifier, which might not consider the temporal changes adequately. Since it does not perform pose estimation based on the human skeleton, this approach may not be suitable for tasks that require more detailed observation of body movements.

Regarding sumo research using machine learning, there is a study (Li and Sezan, 2001) that estimates the start timing of the match using SVM. However, this is not a study of action classification.

2.2 Extension of Human Skeleton Model for Pose Estimation

In the field of pose estimation in sports, there have been reports of research using skeleton models that have been extended by adding sport-specific keypoints to the existing human skeleton model.

Neher et al. have proposed an extended skeleton model that has two new keypoints of the upper and lower parts of an ice hockey stick (Neher et al., 2018). They show that the introduction of this extended skeleton model improves the accuracy of pose estimation for ice hockey players.

Ludwig et al. have proposed a pose estimation method using an extended skeleton model that incorporates two endpoints of the ski board in ski

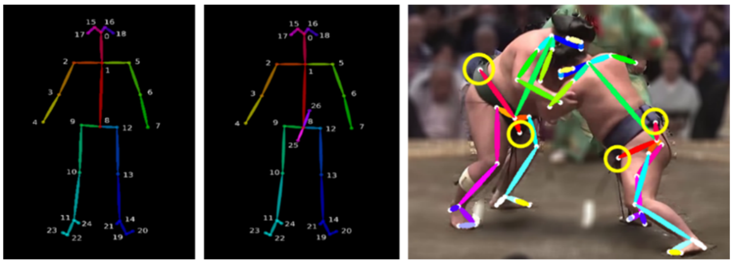

Figure 2: Proposed extended skeleton model. (Left) Original skeleton model of OpenPose (Cao et al., 2019). This picture is cited from the OpenPose document webpage (CMU Perceptual Computing Lab, 2021). (Middle) Our extended skeleton model. Keypoint no. 25 and no. 26 are the new keypoints on mawashi. (Right) Pose estimation with the extended skeleton model. Keypoints highlighted by yellow circles are mawashi keypoints.

jumping as the keypoints (Ludwig et al., 2020). They have reported that this model improves the accuracy of estimating flight parameters in ski jumping videos.

Einfalt et al. have proposed an extended skeleton model based on a skeleton model without foot keypoints. They added four points of the left and right toes and the heels for track and field athletes (Einfalt et al., 2019). They have confirmed that this model improves the accuracy of frame-level estimation of the timing of steps in long jump and triple jump videos.

These studies show that adding unique keypoints to the existing skeleton model improves the accuracy of pose-based action recognition tasks in sports.

3 SUMO ACTIONS “KIMARITE”

The Japan Sumo Association officially recognizes 82 kimarites. In this study, we focus on classifying the most popular four kimarites based on the frequency of kimarite appearance over the past five years (Japan Sumo Association, 2023). There are *Oshidashi* (Frontal Push Out), *Yorikiri* (Frontal Force Out), *Hatakikomi* (Slap Down), and *Tsukiotoshi* (Thrust Down). These top four kimarite actions account for approximately 65% of the total matches. These kimarites are shown in Figure 3. Usually, kimarites last one second or so.

4 SUMO ACTION CLASSIFICATION

4.1 Procedure Overview

We assume sumo videos used for classification are taken from a single viewpoint, where the camera captures both the ring and the wrestlers. The video is taken from a perspective where the two wrestlers are positioned on the left and right in the frame at the start of the match. Additionally, the video is taken from a sufficiently distant place from the ring, such that it can be approximated by a weak perspective projection (as shown in Figure 4). Note that the camera changes the pan/tilt/zoom parameters slightly to let the wrestlers in the video.

Figure 3: Four kimarite actions we classify. Upper-left is Oshidashi (Frontal Push Out), upper-right is Yorikiri (Frontal Force Out), lower-left is Hatakikomi (Slap Down), and lower-right is Tsukiotoshi (Thrust Down). These pictures are cited from the official website with the courtesy of the Japan Sumo Association (Japan Sumo Association, 2023).

Figure 4: A snapshot of sumo videos used for classification. All videos are captured like this figure.

When given an input video, we first perform the frame-by-frame pose estimation and detect the keypoints of the sumo wrestlers' skeleton. Our approach is based on OpenPose (Cao et al., 2017; Cao et al., 2019). We employ the extended skeleton model (Figure 2) tuned for sumo wrestlers, which consists of

27 keypoints, including two new keypoints on mawashi. Using this model for pose estimation, we obtain the sets of skeleton positional vectors for the two sumo wrestlers.

These two skeleton positional vectors are concatenated into one integrated feature vector as a pose sequence.

Since sumo takes place within the ring, the coordinates should be normalized to the ring. Since the ring is a circle shape, it is expressed by the planner cartesian coordinate system with the range of -1.0 to 1.0 for both axes.

We train the LSTM network with a normalized pose sequence (Hochreiter and Schmidhuber, 1997) by using the class name of the kimarite as the correct label.

4.2 Extended Skeleton Model

We introduce a new extended 27-point skeleton model (Figure 2-middle) based on the existing 25-point model of OpenPose (Figure 2-left) by adding two new keypoints of the mawashi. The added keypoints represent the knot point between the front mawashi and the front vertical flap (Figure 5-A) and the knot point between the back mawashi and the back vertical flap (Figure 5-B).

We created a custom dataset of 723 images from sumo matches with keypoint annotations of the sumo wrestlers. Using this dataset, we fine-tuned our extended skeleton model with the pretrained parameters of the 25-point model as the initiation.

By using the extended skeleton model, it becomes possible to represent the relationship between the mawashi and body parts such as hands, which is essential for classifying sumo actions in the uniformed form of skeleton keypoints.

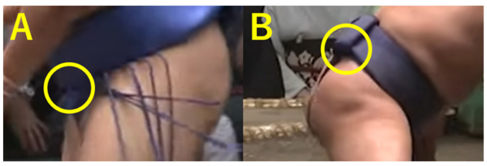

Figure 5: Two additional keypoints of mawashi.

4.3 Normalization to the Sumo Ring

We obtain the planner coordinates of the skeleton keypoints of the two wrestlers by applying the extended skeleton model to a video frame. Since the camera moves slightly, the coordinates should be normalized to the circular sumo ring. Since the ring is observed as an ellipse in the frame, an ellipse approximation is used in image processing to extract the circle information of the sumo ring. We normalize the coordinates of the keypoints to the planner cartesian coordinate system with the range of -1.0 to 1.0 for both axes. The two axes are set by the definition of the sumo ring as it is placed to align north, south, east, and west.

4.4 Pose Sequence

In sumo, the class of the kimarite is determined based on the movements made by the two wrestlers from the start of the match-finishing action to its end. When classifying sumo actions, it is essential to consider the movements of both wrestlers. We represent the kimarite action as time series data related to the movements of the two wrestlers. We first concatenate the two sets of the planner coordinates of the skeleton keypoints of the two wrestlers into one frame vector. Then, by arranging the frame vectors in time-series order, we obtain a pose sequence that corresponds to a certain time length of the sumo match.

4.5 Training LSTM

As vision-based action classifications are one of the popular research domains, research such as graph convolution (Shi et al., 2019), Vision Transformer (Dosovitskiy et al., 2020), ViViT (Arnab et al., 2021), and TemPose (Ibh et al., 2023) have been reported. As the pose sequence of kimarite is not long, we would rather adopt LSTM, one of the RNN-based methods.

We make a classifier model composed of an LSTM layer, Dropout layers, and Dense layers. The structure of the model is shown in Figure 6. The model has two Dense layers and three Dropout layers between the first LSTM layer and the last Dense layer. These layers are set to prevent overfitting and stabilize the learning process.

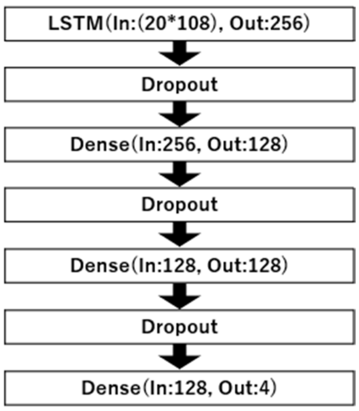

Figure 6: The classifier model.

5 EXPERIMENT

5.1 Dataset

We take up four classes of kimarite action; *Oshidashi* (Frontal Push Out), *Yorikiri* (Frontal Force Out), *Hatakikomi* (Slap Down), and *Tsukiotoshi* (Thrust Down). We made a video dataset collected from the NHK Sports webpage (NHK, 2023) for our experiments. All videos were taken by a camera placed at a certain distance from the sumo ring. The resolution of the videos is 960 pixels in width and 540 pixels in height at 30 fps.

Kimarite is the last action of the match. Therefore, as an input video clip for classification, we segment the last 20 frames of the match video. The end frame of the match is marked visually. The numbers of the collected video clips for the four kimarites are shown in Table 1. The constructed dataset is randomly split so that the ratio of training data to test data is 7:3.

Table 1: The numbers of kimarites in the dataset.

Action	# of video clips
Hatakikomi (Slap Down)	223
Oshidashi (Frontal Push Out)	235
Tsukiotoshi (Thrust Down)	209
Yorikiri (Frontal Force Out)	251

5.2 Data Augmentation

Since the number of video clips is not large, we apply data augmentation to the training video clips. The data augmentation should be tailored to fit the sumo property.

5.2.1 Random Change

As for the general data augmentation, we randomly add a small positional noise to the original training dataset. The noise is generated based on the standard normal distribution. An example of the random change is shown in Figure 7. The red and grey skeletons are the original, and the pink and yellow skeletons are the slightly changed ones in the figure.

5.2.2 Horizontal Flip

For all sumo matches, the name of the kimarite action remains the same even if the two sumo wrestlers are horizontally flipped around the centre of the sumo ring. In Figure 8, the pink and yellow skeletons on the right are the flipped skeletons of the red and grey original skeletons on the left. We doubled the number of pose sequence data for training by the horizontal flip.

5.2.3 Relocation

The kimarite actions of *Hatakikomi* (Slap Down) and *Tsukiotoshi* (Thrust Down) should end within the sumo ring. Therefore, these two actions are still valid even when the actions are relocated within the sumo ring.

We generated new pose sequences with relocation by adding the same values to all the coordinates in a pose sequence. The random values are generated based on the uniform distribution. The values are set so as not to go over the sumo ring.

This process allows us to create multiple new skeletons while maintaining the shape of the original skeleton, as shown in Figure 9. In the figure, the red and grey skeletons represent the original skeletons, while the pink and yellow skeletons, as well as the light blue and lavender skeletons, are the relocations of the original skeletons.

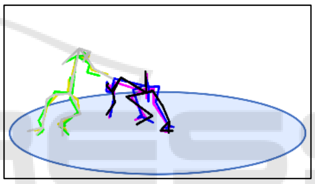

Figure 7: Random change of skeleton poses. The blue circle indicates the sumo ring.

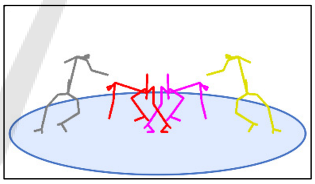

Figure 8: Horizontal flipping of skeleton poses. The blue circle indicates the sumo ring.

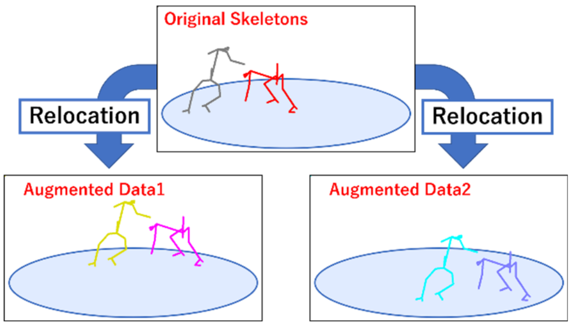

Figure 9: Relocation of skeleton poses. The blue circle means the sumo ring.

5.2.4 Enhanced Training Dataset

On kimarite classification, we do not need to distinguish the winner and the loser of a match. Therefore, we created pose sequences by swapping the two sumo wrestlers in the pose sequences of all the video clips.

We applied all the augmentation approaches and prepared the enhanced training dataset. The final numbers of the pose sequences in the enhanced training dataset are shown in Table 2.

Table 2: Pose sequences of the enhanced training dataset.

Action	# of pose seq.
Hatakikomi (Slap Down)	3099
Oshidashi (Frontal Push Out)	3263
Tsukiotoshi (Thrust Down)	2903
Yorikiri (Frontal Force Out)	2632

5.3 Model Training

We trained the model shown in Figure 6. The loss function for the model is cross-entropy. Layer normalization is performed after the LSTM layer. All the activation functions in the Dense layer are ReLU, and the dropout rate in the Dropout layer is set at 0.2. 30% of the video clips were randomly selected for validation. Training was carried out for 50 epochs, and the weights from the epoch with the lowest validation loss were adopted.

The training loss and training accuracy are shown in Figure 10, and the validation loss and validation accuracy are shown in Figure 11.

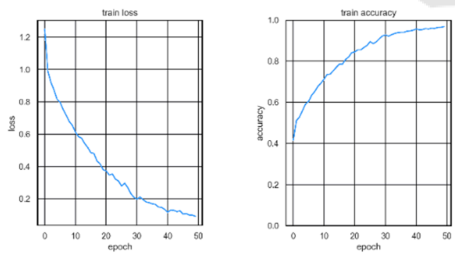

Figure 10: Training loss and training accuracy.

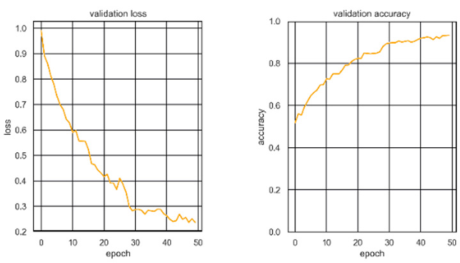

Figure 11: Validation loss and validation accuracy.

5.4 Results and Evaluations

5.4.1 Evaluation of Our Method

As the evaluation of the trained model, we performed classification with the test dataset. As mentioned in section 5.1, this test dataset is randomly picked 30% of the overall dataset shown in Table 1. We calculate the precision, recall, and F1-score of the results of classification. The results are shown in Table 3. Additionally, the confusion matrix is shown in Figure 12. Figure 12 shows that the correctly predicted class is the most common for any of the classes. The accuracy for classifying the four kimarite actions was 0.77. Based on these results, our method is effective in classifying kimarite actions from sumo videos.

From Figure 12, we can confirm that the actions tend to be misclassified into certain specific classes. Specifically, *Hatakikomi* are mostly misclassified as *Tsukiotoshi*, and vice versa. A similar case can be found for the pair *Oshidashi* and *Yorikiri*. The pair of *Hatakikomi* and *Tsukiotoshi* look similar from the definition (lower row of Figure 3). The pair of *Oshidashi* and *Yorikiri* (upper row of Figure 3) also look similar. This implies that most of the misclassifications occur between actions that have similar forms. We expect that further training of the LSTM model with a larger number of the dataset could result in better recognition performance.

5.4.2 Improvement by Mawashi Keypoints

To verify the effectiveness of adding the mawashi keypoints, we compared the results of the 25-point model (Figure 2), which did not include the two extended mawashi keypoints. The result of the 25-point model is presented in Table 4, and the confusion matrix is shown in Figure 13. Furthermore, the result of the comparison of accuracy is shown in Table 5.

As shown in Table 5, using the extended skeleton model with the mawashi keypoints is more effective for kimarite action classification. From this result, we can confirm the advantage of the extended skeleton model tuned for sumo wrestlers.

Table 3: Results of kimarite action classification using the proposed extended skeleton model with mawashi keypoints. The total accuracy of the 4-class classification is 0.77.

Action	Precision	Recall	F1-score
Hatakikomi	0.72	0.84	0.77
Oshidashi	0.78	0.76	0.77
Tsukiotoshi	0.78	0.63	0.70
Yorikiri	0.78	0.82	0.80

Table 4: Results of kimarite action classification using the 25-point skeleton model without mawashi keypoints. Total accuracy is 0.74, as shown in Table 5.

Action	Precision	Recall	F1-score
Hatakikomi	0.70	0.79	0.74
Oshidashi	0.86	0.69	0.77
Tsukiotoshi	0.65	0.62	0.63
Yorikiri	0.77	0.86	0.81

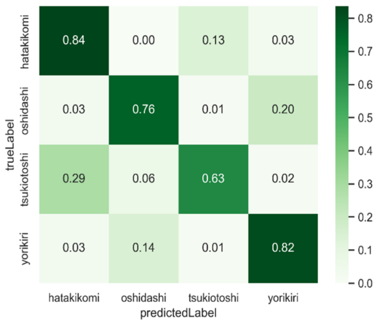

Figure 12: Confusion matrix of the action classification result using the proposed extended skeleton model with mawashi keypoints.

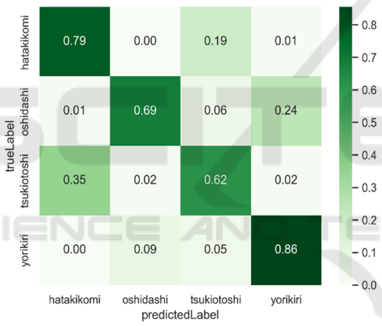

Figure 13: Confusion matrix of the action classification result using the 25-point skeleton model without mawashi keypoints.

Table 5: Comparison of classification accuracy between two skeleton models (with mawashi keypoints or not).

Skeleton model	Accuracy
Proposed: With mawashi keypoints (27 points)	**0.77**
Without mawashi keypoints (25 points)	0.74

6 CONCLUSIONS

We propose a new method of classification for kimarites in sumo video based on our proposed extended skeleton model that has the two mawashi keypoints to fit sumo wrestlers.

The unique feature of the extended skeleton model is the new keypoints that correspond to mawashi. The relationship between mawashi and body parts is uniformly represented in the extended skeleton model. As to deal with heavy occlusion, we represent sumo actions as pose sequences so that the classification procedure based on LSTM can achieve better classification performance.

The extended skeleton model is trained by our own video dataset, which was enhanced with the data augmentation. The data augmentation was designed so as not to break the definition of kimarites.

With this approach, we achieved an accuracy of 0.77 in the task of classifying the four popular kimarites. We also confirmed that the introduction of the extended skeleton model with mawashi keypoints improved classification accuracy by the ablation study with the 25-point model.

ACKNOWLEDGEMENTS

We thank the Japan Sumo Association for the permission to use their images in this paper. We had a written permission from them. Part of this work was supported by JSPS KAKENHI Grant Number 22K19803.

REFERENCES

Arnab, A., Dehghani, M., Heigold, G., Sun, C., Lučić, M., and Schmid, C. (2021). Vivit: A video vision transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp.6836-6846.

Azar, S. M., Atigh, M. G., Nickabadi, A., and Alahi, A. (2019). Convolutional relational machine for group activity recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.7892-7901.

Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., and Sheikh, Y. (2019). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. In *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol.43, no.1, pp.172-186.

Cao, Z., Simon, T., Wei, S. E., and Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pp.7291-7299.

CMU Perceptual Computing Lab. (2021). OpenPose Doc – Output, OpenPose Doc, Available at: https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/02_output.md (Accessed: 10 October 2023).

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Einfalt, M., Dampeyrou, C., Zecha, D., and Lienhart, R. (2019). Frame-level event detection in athletics videos with pose-based convolutional sequence networks. In *Proceedings Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports*, pp.42-50.

Fani, M., Neher, H., Clausi, D. A., Wong, A., and Zelek, J. (2017). Hockey action recognition via integrated stacked hourglass network. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp.29-37.

Gavrilyuk, K., Sanford, R., Javan, M., and Snoek, C. G. (2020). Actor-transformers for group activity recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.839-848.

Gerats, B. G. A. (2020). Individual action and group activity recognition in soccer videos. *Master's thesis, University of Twente*.

Guo, J., Liu, H., Li, X., Xu, D., and Zhang, Y. (2021). An attention enhanced spatial–temporal graph convolutional LSTM network for action recognition in karate. *Applied Sciences*, vol.11, no.18, article no.8641.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, vol.9, no.8, pp.1735-1780.

Ibh, M., Grasshof, S., Witzner, D., and Madeleine, P. (2023). TemPose: A New Skeleton-Based Transformer Model Designed for Fine-Grained Motion Recognition in Badminton. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.5198-5207.

Japan Sumo Association. (2023). 82 kimarite. *Nihon Sumo Kyokai Official Grand Sumo Home Page*, Available at: https://www.sumo.or.jp/Kimarite (Accessed: 10 October 2023).

Kulkarni, K. M., and Shenoy, S. (2021). Table tennis stroke recognition using two-dimensional human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp.4576-4584.

Li, B., and Sezan, M. I. (2001). Event detection and summarization in sports video. In *Proceedings IEEE workshop on content-based access of image and video libraries (CBAIVL 2001)*, pp.132-138. IEEE.

Liang, J., and Zuo, G. (2022). Taekwondo Action Recognition Method Based on Partial Perception Structure Graph Convolution Framework. *Scientific Programming*, vol.2022, Article ID 1838468.

Liu, R., Liu, Z., and Liu, S. (2021). Recognition of basketball player's shooting action based on the convolutional neural network. *Scientific Programming*, vol.2021, Article ID 3045418.

Liu, S., Liu, X., Huang, G., Feng, L., Hu, L., Jiang, D., Zhang, A., Liu, Y., and Qiao, H. (2020). FSD-10: a dataset for competitive sports content analysis. *arXiv preprint arXiv:2002.03312*.

Liu, S., Zhang, A., Li, Y., Zhou, J., Xu, L., Dong, Z., and Zhang, R. (2021). Temporal segmentation of fine-gained semantic action: A motion-centered figure skating dataset. In *Proceedings of the AAAI conference on artificial intelligence*, vol.35, no.3, pp.2163-2171.

Ludwig, K., Einfalt, M., and Lienhart, R. (2020). Robust estimation of flight parameters for ski jumpers. In *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp.1-6. IEEE.

Luo, C., Kim, S. W., Park, H. Y., Lim, K., and Jung, H. (2023). Agnostic Taekwondo Action Recognition Using Synthesized Two-Dimensional Skeletal Datasets. *Sensors*, vol.23, no.19, p.8049.

Mottaghi, A., Soryani, M., and Seifi, H. (2020). Action recognition in freestyle wrestling using silhouette-skeleton features. *Engineering Science and Technology, an International Journal*, vol.23, no.4, pp.921-930.

Neher, H., Vats, K., Wong, A., and Clausi, D. A. (2018). Hyperstacknet: A hyper stacked hourglass deep convolutional neural network architecture for joint player and stick pose estimation in hockey. In *2018 15th Conference on Computer and Robot Vision (CRV)*, pp. 313-320. IEEE.

NHK (Japan Broadcasting Corporation). (2023). Movies of Grand Sumo, NHK Sports, Available at: https://www3.nhk.or.jp/sports/special/sumomovies (Accessed: 10 October 10, 2023)

Rangasamy, K., As'ari, M. A., Rahmad, N. A., and Ghazali, N. F. (2020). Hockey activity recognition using pre-trained deep learning model. *ICT Express*, vol.6, no.3, pp.170-174.

Shi, L., Zhang, Y., Cheng, J., and Lu, H. (2019). Two-stream adaptive graph convolutional networks for skeleton-based action recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp.12026-12035.

Vainstein, J., Manera, J. F., Negri, P., Delrieux, C., and Maguitman, A. (2014). Modeling video activity with dynamic phrases and its application to action recognition in tennis videos. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 19th Iberoamerican Congress, CIARP 2014*, pp.909-916, Springer.

Vats, K., Neher, H., Clausi, D. A., and Zelek, J. (2019). Two-stream action recognition in ice hockey using player pose sequences and optical flows. In *2019 16th Conference on computer and robot vision (CRV)*, pp.181-188. IEEE.

Zakharchenko, I. (2020). Basketball Pose-Based Action Recognition. *Ukrainian Catholic University, Faculty of Applied Sciences: L'viv.*