Article

Not peer-reviewed version

Action Recognition of Taekwondo Unit-Actions Using Action Images Constructed by Time-Warped Motion Profile

Junghwan Lim , Chenglong Luo , Seunghun Lee , Young Eun Song * , Hoeryong Jung *

Posted Date: 7 March 2024

doi: 10.20944/preprints202403.0436.v1

Keywords: Action recognition; convolution neural network; human action dataset; taekwondo

Article

Action Recognition of Taekwondo Unit-Actions Using Action Images Constructed by Time-Warped Motion Profile

Junghwan Lim [1], Chenglong Luo [2], Seunghun Lee [3], Young Eun Song [4,*] and Hoeryong Jung [2,*]

[1] Department of Motion, Torook Co., Ltd., 42, Toegye-ro 80-gil, Jung-gu, Seoul 04585, Korea; gpopl123@naver.com (J.L.)
[2] Department of Mechanical Engineering, Konkuk University, 120 Neungdong-ro, Jayang-dong, Gwangjin-gu, Seoul 05029, Korea; luo0611@konkuk.ac.kr (C.L.); junghl80@konkuk.ac.kr (H.J.)
[3] School of Mechanical and Aerospace Engineering, Seoul National University, Gwanak-ro 1, Gwanak-gu, Seoul 08826, Korea; erioer95@snu.ac.kr (S.L.)
[4] Department of Autonomous Mobility, Korea University, 2511 Sejong-ro, Sejong, 30019, Korea; tdsong@korea.ac.kr (Y.S.)
* Correspondence: tdsong@korea.ac.kr (Y.E.S.); junghl80@konkuk.ac.kr (H.J.); Tel.: +82-2-450-3903

Abstract: Taekwondo has evolved from a traditional martial art into an official Olympic sport. This study introduces a novel action recognition model tailored for Taekwondo unit actions, utilizing joint motion data acquired via wearable inertial measurement unit (IMU) sensors. The utilization of IMU sensor-measured motion data facilitates the capture of the intricate and rapid movements characteristic of Taekwondo techniques. The model, underpinned by a conventional convolutional neural network (CNN)-based image classification framework, synthesizes action images to represent individual Taekwondo unit actions. These action images are generated by mapping joint motion profiles onto the RGB color space, thus encapsulating the motion dynamics of a single unit action within a solitary image. To further refine the representation of rapid movements within these images, a time-warping technique was applied, adjusting motion profiles in relation to the velocity of the action. The effectiveness of the proposed model was assessed using a dataset compiled from 40 Taekwondo experts, yielding remarkable outcomes: an accuracy of 0.998, precision of 0.983, recall of 0.982, and an F1-score of 0.982. These results underscore the time-warping technique's contribution to enhancing feature representation, as well as the proposed method's scalability and effectiveness in recognizing Taekwondo unit actions.

Keywords: action recognition; convolution neural network; human action dataset; taekwondo

1. Introduction

Taekwondo, originating from Korea, has evolved from a traditional martial art into an official Olympic sport, becoming one of the world's most practiced martial arts. This discipline is divided into two main categories: Gyeorugi, a full-contact sparring between two competitors utilizing electronic scoring equipment for objective, quantitative assessments, and Poomsae, where individual competitors perform a series of predetermined movements, including basic attack and defense techniques, in front of judges. The evaluation of Poomsae is inherently subjective and qualitative, lacking objective tools for assessment, which complicates the recognition of competitors' movements and introduces fairness concerns due to variability in judges' assessments.

The application of action recognition technology, which enables precise motion measurement and quantitative evaluation, presents a promising solution to the challenges faced in current Poomsae evaluations [1–11]. A variety of vision-based action recognition methods have been developed specifically for Taekwondo. De Goma et al. [12] utilized a hidden Markov model (HMM) with

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 March 2024 doi:10.20944/preprints202403.0436.v1

skeletons extracted from RGB-D camera images for action recognition. Choi et al. [13] introduced a remote evaluation module for Poomsae using a multivision sensor action recognition approach. Seo et al. [14] developed a recognition algorithm based on Poisson distribution, leveraging one-dimensional spatial information from image sequences. Kong et al. [15] proposed an automatic analysis framework for Taekwondo videos broadcasted, integrating a structure-preserving object tracker with a principal component analysis (PCA) network. Liang et al. [16] explored a novel evaluation method for Taekwondo competitions combining long short-term memory (LSTM) with a spatial temporal graph convolutional network (ST-GCN). Recently, Lee et al. [17] reported over 80% recognition accuracy using a convolutional neural network (CNN)-based model that processes a sequence of key-frame images to identify basic Taekwondo unit actions. However, these vision-based strategies, relying on raw image sequences, face significant accuracy limitations due to environmental conditions and the appearance of competitors' attire.

Skeleton-based action recognition models represent a promising avenue for mitigating the limitations encountered in traditional vision-based action recognition, particularly by excluding extraneous background elements that are not pertinent to recognizing actions [18–21]. This methodology hinges on extracting the subject's skeleton from the image, which is then utilized as the input for the action classification model. By eliminating superfluous details from the raw images, this approach ensures that accuracy is not compromised by environmental variables. Graph convolutional network (GCN)-based strategies [22], which employ action classifiers trained on input graphs derived from skeletal configurations, along with 3D convolutional networks that utilize 3D heat maps of the skeletons [23], have been introduced to enhance the robustness of action recognition. These techniques enable the action classifier to discern the essential characteristics of an action through the geometric data of a joint and its adjacent points, independent of background elements. However, a significant challenge arises owing to the sensitivity to minor coordinate alterations, which can result in markedly divergent predictions [24–26]. This issue is particularly pronounced in martial arts like Taekwondo, where rapid and intricate movements may lead to inaccuracies in joint positioning and topological errors, consequently affecting the stability and consistency of action predictions [27,28].

Despite the predominance of vision-based methodologies in action recognition research, they exhibit several drawbacks when applied to Taekwondo action recognition. An alternative approach that circumvents these limitations involves the use of wearable inertial measurement unit (IMU) sensors. These sensors, affixed to the body's joints, capture three-dimensional linear accelerations and angular velocities, facilitating the computation of full-body motion in a rapid (> 200 Hz) and reliable manner. Prior research has explored the application of various machine learning techniques to process the raw IMU data for action classification, including recurrent neural networks (RNNs) [29], LSTM networks [30], CNNs [31], and hybrid CNN-LSTM models [32], showcasing the potential of IMU sensors in overcoming the challenges posed by vision-based action recognition methods.

Recent investigations have explored the conversion of raw IMU data into the RGB color space to generate action images, subsequently utilized as inputs for action classification [33]. This advanced technique captures both the spatial attributes of human motion—such as joint positions, velocities, and accelerations—and the temporal dynamics by integrating sequential action data within a defined timeframe into a singular image, thereby offering a comprehensive input for the action classifier. Despite the potential of this approach, the specific inclusion and representation of action characteristics within the action image have been inadequately addressed.

In this study, we introduce an action recognition model specifically designed for Taekwondo unit actions, utilizing motion data obtained from wearable IMU sensors. The proposed method refines the existing strategy of transforming joint motion data into RGB action images by applying a modulation of the time-domain motion profiles. This modification aims to amplify the depiction of rapid movements characteristic of Taekwondo unit actions. The main concept of our approach is the implementation of time-warping techniques to the motion profiles based on their velocity, thereby extending the portrayal of swift movements within the action image. A CNN-based classification model was adopted to evaluate the effectiveness of our proposed method, employing a dataset of

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 March 2024 doi:10.20944/preprints202403.0436.v1

Taekwondo unit actions from 40 skilled practitioners. The primary contributions of our study are outlined as follows:

1. We present a pioneering technique for creating an action image from joint motion data captured via IMU sensors. This method incorporates time-warping techniques into the motion profiles, significantly enhancing the representation of rapid movements within the action image.
2. The effectiveness of the proposed action recognition model was validated with a dataset comprising Taekwondo unit actions from 40 experts. The evaluation results not only affirm the model's accuracy but also its scalability, underscoring the viability of our approach in recognizing Taekwondo unit actions.

2. Materials and Methods

The action recognition framework proposed in this study is structured around three principal phases: data collection, action image generation, and action classification, as depicted in Figure 1.

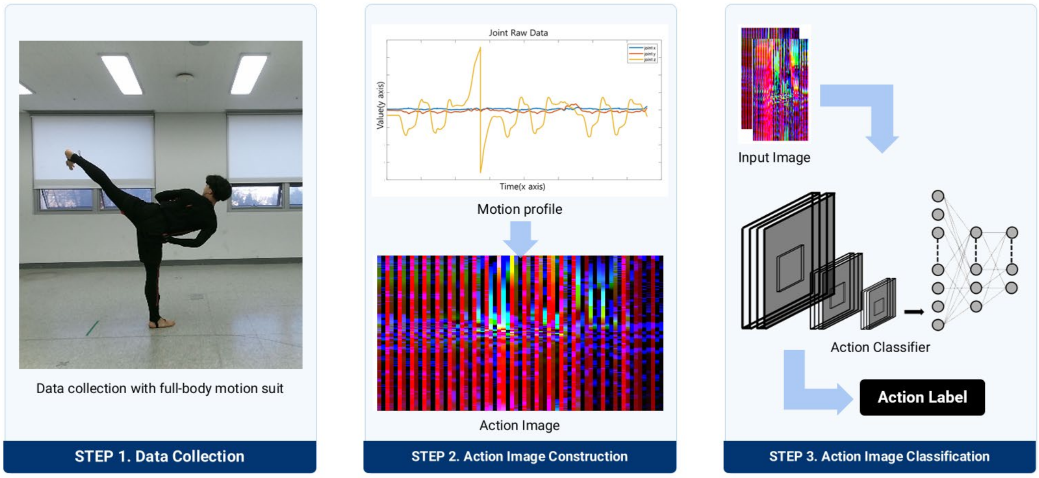


Figure 1. Overview of the proposed action recognition process.

Data collection: This phase involves the acquisition of IMU data from sensors affixed to participants as they execute specific Taekwondo unit actions. The gathered data are segmented by individual actions and annotated with the corresponding action names to facilitate subsequent analysis.

Action image generation: In this step, the segmented IMU data for each unit action are transformed into a singular action image. This transformation is achieved by mapping the IMU data values to the RGB color spectrum of each pixel, with the image's columns representing the sampling times and the rows reflecting the indices of the IMU sensors.

Action classification: The final phase employs a CNN classification model to discern the Taekwondo unit actions from the generated action images. This model is specifically designed to process action images derived from IMU data, outputting the designated label for each action.

2.1. Data Collection

During this study, a dataset comprising IMU data on Poomsae unit actions was compiled, drawing from the expertise of professional Taekwondo practitioners. The data collection protocol received approval from the Konkuk University Institutional Review Board (IRB) under protocol number 7001355-202004-HR-372, ensuring the process adhered to ethical guidelines. The dataset was devoid of any personally identifiable information, with informed consent secured from each participant prior to data collection. The participants were assured that the data would be exclusively utilized for scholarly research purposes.

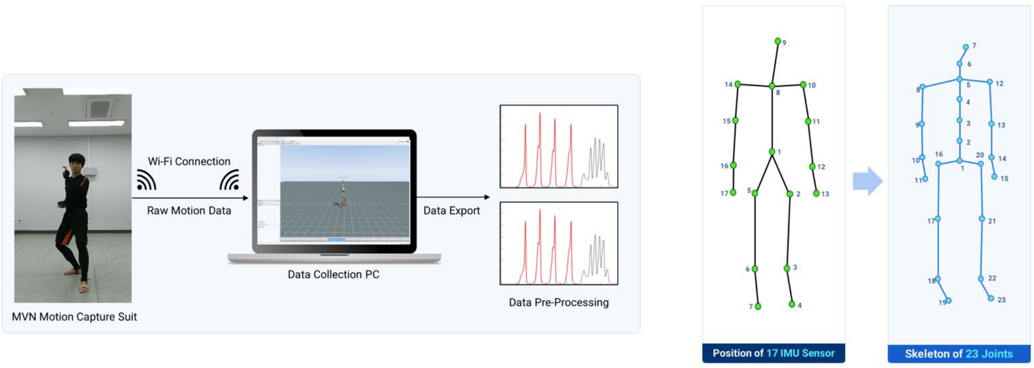
2.1.1. Participants

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 March 2024 doi:10.20944/preprints202403.0436.v1

For the data collection endeavor, forty adult professional Taekwondo Poomsae demonstrators were enlisted. These participants were assigned identifiers ranging from T1 to T40, enabling the systematic categorization of the data derived from different individuals.

2.1.2. Data Collection Protocol

Figure 2a illustrates the configuration utilized for data collection. Participants were equipped with a comprehensive motion capture ensemble, specifically the Xsens MVN system (Xsens Corp., Netherlands). This ensemble comprised a snugly fitting Lycra suit integrated with 17 IMU sensors designated for motion tracking (MTx, MVN, Xsens Corp., Netherlands), accompanied by a waist pack housing the batteries, a data acquisition (DAQ) unit, and a wireless transmission module facilitating data exchange with the central computing unit. The system afforded the capability of real-time, full-body motion tracking, capturing data across 23 joints. This included nine sets of translational data (encompassing position, velocity, and acceleration) and nine sets of angular data (covering orientation, angular velocity, and angular acceleration), all sampled at a frequency of 240 Hz. These measurements were derived from the 17 IMU sensors through sophisticated, embedded motion-tracking algorithms. For an in-depth explanation, reference [34] provides further details. The placement of the sensors and corresponding joints is depicted in Figure 2b. During the data collection phase, participants were instructed to sequentially perform 16 distinct Poomsae unit actions while donning the motion capture suit. These actions were: Act 1: Single Knife Hand Block, Act 2: Knife Hand Inward Outside, Act 3: Body Inside Block, Act 4: Low Block, Act 5: High Block, Act 6: Body Punch, Act 7: Double Punch, Act 8: Face Punch, Act 9: Spear Hand Thrust, Act 10: Inside Knife Hand Strike, Act 11: Side Knife Hand Strike, Act 12: Swallow Form Inside Strike, Act 13: Back Fist Front Strike, Act 14: Front Kick, Act 15: Side Kick, and Act 16: Turning Kick. Each unit action was performed four times in succession before the participant reverted to their starting position. This routine was replicated three times for each unit action, culminating in 12 iterations per action. Through this meticulous procedure, a total of 7,680 datasets of Taekwondo unit actions were amassed (calculated from 40 participants, each performing 16 unit actions, repeated 12 times), as depicted in Figure 3.



(a) (b)

Figure 2. Data collection setup (a) Configuration of data collection setup, (b) Placement of IMU sensors and the skeleton structure illustrating 23 joints.

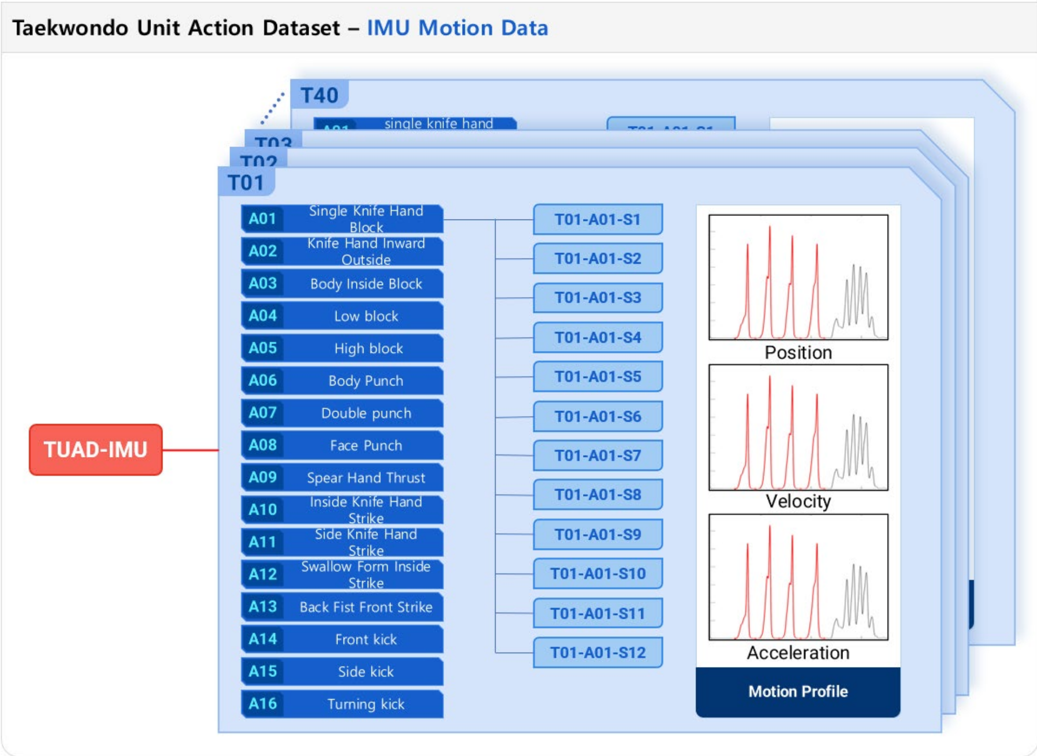


Figure 3. Structure of the Taekwondo unit action dataset (TUAD-IMU) derived from IMU motion data.

2.2. *Action Image Generation*

2.2.1. Action Segmentation

During the data acquisition phase, subjects were required to execute a specific unit action 12 times in succession. Given that each action image corresponds to a singular unit action, the IMU data, encompassing 12 repetitions of the unit action, needed to be segmented into 12 distinct datasets. The process employed the average velocity value—calculated by averaging the magnitudes of velocities recorded at the hands and feet—as a criterion to determine the initial and final samples of each action, as depicted in Figure 4. This velocity metric facilitated the straightforward segmentation of the motion profile by individual unit actions, as illustrated in the figure.

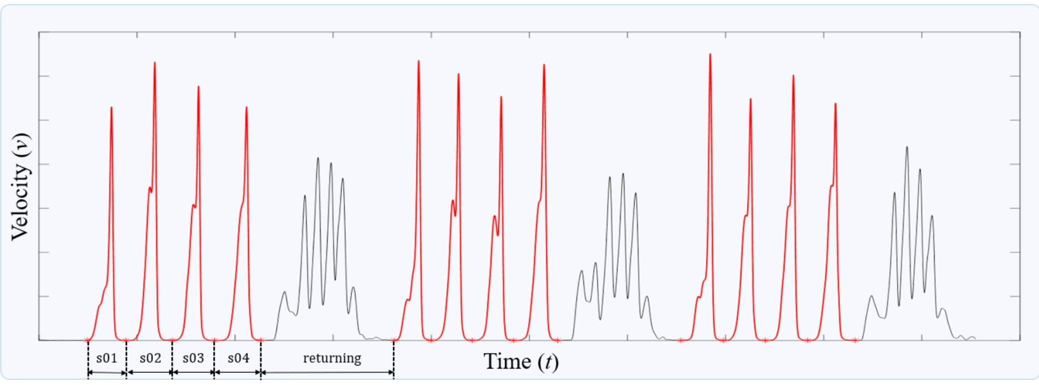


Figure 4. Process of segmenting a single action profile.

2.2.2. Projecting IMU Data onto RGB Color Space

The methodology for generating an action image involved mapping the IMU data values onto the RGB color space of image pixels. This process transformed the time series data of both linear and angular motions across 23 joints, associated with a single unit action, into one comprehensive action image. The linear and angular motion data included 3-dimensional aspects of position, velocity, and acceleration, culminating in 18 scalar values for each joint per time frame. Figure 5 elucidates the structure of the action images, where the column in the action image signifies the joint index and the type of motion data. The six motion data points were represented by six columns within the action image, with the x, y, and z dimensions of each motion data point corresponding to the r, g, and b color channels of the image pixels, respectively. For instance, the motion data for joint 1 (J01) would occupy the first six columns of the image, with the 3-dimensional space of the motion data being mapped onto the three channels of the image pixel.

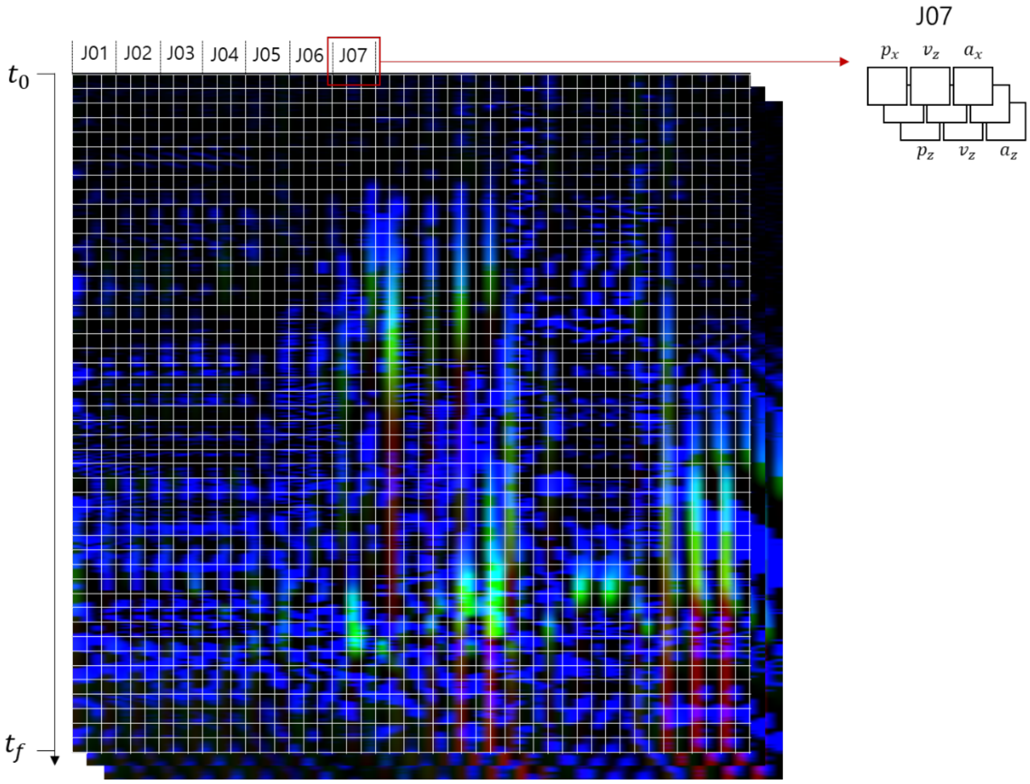


Figure 5. Generation of action images through mapping of IMU motion data onto RGB color space.

In the process of mapping the sampling time to pixel row coordinates and the joint index to pixel column coordinates, the RGB values of the image pixels were derived from the IMU data as per the following methodology:

$$C_{i,j}^{red} = \frac{1}{P_x(j)^{max} - P_x(j)^{min}} P_x(j)_i, \quad (1)$$

$$C_{i,j}^{green} = \frac{1}{P_x(j)^{max} - P_x(j)^{min}} P_y(j)_i, \quad (2)$$

$$C_{i,j}^{blue} = \frac{1}{P_x(j)^{max} - P_x(j)^{min}} P_z(j)_i, \quad (3)$$

where $C_{i,j}^{red}$ represents the pixel value of the red channel at the i^{th} row and j^{th} column, and $P(j)$ represents the function that returns the motion data corresponding to the j^{th} column of the image at the i^{th} time frame. $P(j)^{max}$ and $P(j)^{min}$ denote the maximum and minimum value of $P(j)$, respectively.

2.2.3. Time-Warping

The rapid movements inherent in Taekwondo unit actions present a significant challenge for action recognition. As these dynamic movements are pivotal in distinguishing Taekwondo actions, their representation in action images necessitates emphasis. In the described action image construction, each row correlates to a distinct timeframe of an action, distributing the temporal

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 March 2024 doi:10.20944/preprints202403.0436.v1

sequence of the action uniformly across the rows, irrespective of the action's velocity. This uniform distribution results in a diminished representation of rapid movements compared to slower ones. Despite the critical role of rapid movements in the recognition of Taekwondo actions, their depiction in action images has been insufficiently accentuated.

To address this issue, this study introduces a time-warping technique designed to enhance the depiction of rapid movements in action images. The concept of time warping mirrors the slow-motion effects utilized in cinematography, which decelerates the footage to capture and emphasize essential or fleeting details without missing them. By expanding the temporal segments corresponding to rapid movements within an action image, the technique aims to facilitate a more effective capture of the fundamental characteristics of Taekwondo actions. The application of this time-warping method employs a velocity indicator—calculated by averaging the velocity magnitudes of both hands and feet—to delineate the speed of motion, thereby adjusting the representation of rapid movements within the action image to ensure they are adequately emphasized.

$$V^{rep}(t) = \frac{1}{4}\left(\|\boldsymbol{V}^{hl}(t)\| + \|\boldsymbol{V}^{hr}(t)\| + \left\|\boldsymbol{V}^{fl}(t)\right\| + \left\|\boldsymbol{V}^{fr}(t)\right\|\right), \quad (4)$$

where $V^{rep}(t)$ represents the representative speed of motion at time t, and $\boldsymbol{V}^{hl}(t)$, $\boldsymbol{V}^{hr}(t)$, $\boldsymbol{V}^{fl}(t)$, $\boldsymbol{V}^{fr}(t)$ denote the velocities of the hands and feet at the corresponding time frame. Assume that the motion data of one unit of action is presented as a function of time t as follows:

$$p = p(t), where\ 0 < t < t_f \quad (5)$$

where t_f denotes the final time of the action. The time-warping algorithm modifies the time domain to accentuate segments corresponding to rapid movements, thereby transforming the original time domain into a warped time domain using the value of V^{rep} as follows:

$$t^{warp} = \frac{\int_0^t V^{rep}(t)dt}{\int_0^{t_f} V^{rep}(t)dt} \times t_f. \quad (5)$$

Upon substituting the time t into the proposed equation, the corresponding warped time t^{warp} is computed, allowing for the determination of position data within this altered time frame using the following expression:

$$p^{warp} = p(t^{warp}). \quad (5)$$

These position data, once resampled, serve as the basis for generating the action image utilized in the classification process.

2.4. CNN Architecture

The CNN-based classification model, which employs the generated action image as input, is tasked with identifying the appropriate action label. The model's architecture, as depicted in Figure 6, encompasses four convolutional layers followed by fully connected layers. The convolutional layers are equipped with filters of dimensions 5×5×16, 3×3×32, 3×3×64, and 3×3×128, respectively. The rectified linear unit (ReLU) functions as the activation mechanism, while max pooling is implemented with 2×2 windows and a stride of 2 to reduce spatial dimensions. Table 1 presents the output shapes and parameter counts for each layer within the model. For the training and validation of the model, 70% of the action images were allocated for training, with the remaining 30% dedicated to validation. To ensure thorough training and validation, a 5-fold cross-validation technique was employed, enhancing the robustness and reliability of the classification outcomes.

Table 1. Specifications of the classification model including output shapes and parameter counts.

Layer Index	Conv1	Conv2	Conv3	Conv4	FC1	FC2	Total
Output shape	196×116×16	96×58×16	46×26×64	23×13×64	6400×1	128×1	-
# of parameters	2416	4640	18496	73856	819328	4128	921854

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 March 2024 doi:10.20944/preprints202403.0436.v1

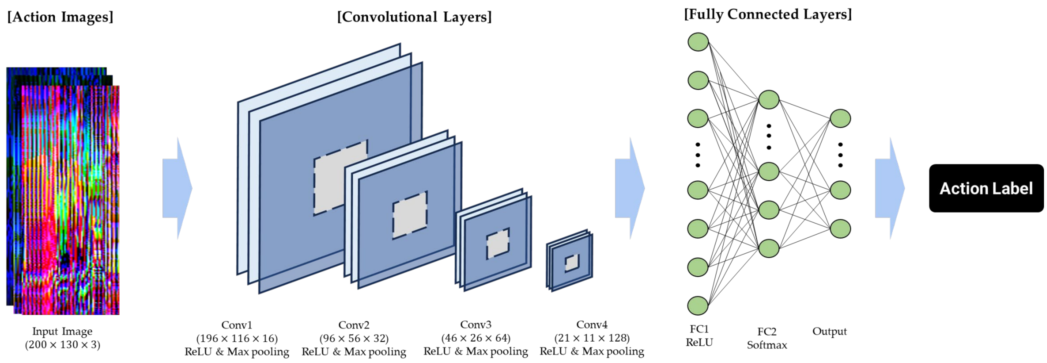


Figure 6. Architectural diagram of the CNN used for action classification.

2.5. Evaluation Metrics

The effectiveness of the action recognition model presented in this study was assessed utilizing four key metrics: accuracy, precision, recall, and F1 score. These metrics were calculated as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN},\ Precision = \frac{TP}{TP+FP},$$
$$Recall = \frac{TP}{TP+FN},\ F1\ score = 2 \cdot \frac{Precision \cdot recall}{Precision+recall}, \quad (4)$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

3. Results

The validation of the proposed action-recognition model was conducted through a dual approach. Initially, a performance evaluation was carried out to compare the models that utilized action images generated from time-warped motion data against those created from standard motion data. The objective was to ascertain the effectiveness of the time-warping algorithm. Subsequently, the model's performance was examined in relation to the number of joints considered for generating the action image, aiming to assess the algorithm's scalability. For this comparison, action images were generated using motion data from varying numbers of joints: 23, 8, and 4.

3.1. Performance Evaluation of Time-Warped Action Image

Figure 7 displays the action images for four distinct unit actions, created using both time-warped and normal motion data. The color intensity within these images, with brighter colors indicating higher data magnitudes, showcased that the application of time-warping techniques resulted in an expanded bright region. This expansion signifies an enhancement in the representation of rapid motion features within the action images. Table 2 delineates the results of the performance comparison across the 16 unit actions using the four evaluation metrics, with approximately 140 action images per unit action employed for model performance evaluation.

Table 2. Comparative analysis of performance metrics between time-warped and normal action images.

	Accuracy		Precision		Recall		F1 score	
	normal	warping	normal	warping	normal	warping	normal	warping
A1	0.997	0.999	0.966	0.980	0.993	1.000	0.979	0.990
A2	0.999	1.000	0.993	1.000	0.993	1.000	0.993	1.000
A3	0.991	0.993	0.977	0.921	0.882	0.965	0.927	0.942
A4	0.999	0.999	0.993	1.000	0.986	0.986	0.990	0.993
A5	0.998	0.999	0.993	1.000	0.972	0.979	0.982	0.989
A6	0.991	0.996	0.942	0.993	0.910	0.938	0.926	0.964
A7	0.995	1.000	0.952	0.993	0.972	1.000	0.962	0.997

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 March 2024 doi:10.20944/preprints202403.0436.v1

A8	0.993	0.996	0.939	0.941	0.958	0.993	0.948	0.966
A9	0.990	0.997	0.895	0.979	0.944	0.979	0.919	0.979
A10	0.990	0.993	0.890	0.971	0.958	0.917	0.923	0.943
A11	0.994	0.999	0.965	0.993	0.944	0.986	0.954	0.990
A12	0.998	0.999	0.993	0.980	0.972	1.000	0.982	0.990
A13	0.996	0.999	0.966	0.993	0.972	0.993	0.969	0.993
A14	0.999	1.000	0.986	1.000	1.000	1.000	0.993	1.000
A15	0.994	0.999	0.992	0.986	0.910	0.993	0.949	0.990
A16	0.994	0.998	0.928	0.993	0.986	0.979	0.956	0.986
Ave.	0.995	0.998	0.961	0.983	0.960	0.982	0.960	0.982

The comparative analysis revealed that both methodologies exhibited high levels of accuracy and precision across all evaluated actions. For example, in the case of Action1, the “time-warping” technique achieved an accuracy of 0.999 and a precision of 1.000, marginally surpassing the “normal” method, which recorded an accuracy of 0.997 and a precision of 0.993. This pattern of performance enhancement through the “time-warping” technique was consistent across the majority of actions. Actions such as Action2 and Action14 were particularly noteworthy, where the “time-warping” approach attained perfect scores (1.000) in both metrics, underscoring its effectiveness. The exploration of recall and F1 scores unveils subtle distinctions between the “normal” and “time-warping” methods. For instance, in Action3, the “normal” method yielded a recall of 0.977 and an F1 score of 0.927. In contrast, the “time-warping” method exhibited a slightly lower recall of 0.921 yet achieved a superior F1 score of 0.942. Such discrepancies suggest that while the “time-warping” method may occasionally compromise recall for precision, it broadly sustains a high level of performance, as reflected by its consistently elevated F1 scores.

Collectively, the time-warping method demonstrated marginally superior efficacy across the board. The average performance metrics for the “time-warping” method—accuracy: 0.998, precision: 0.982, recall: 0.982, F1 score: 0.982—surpassed those of the “normal” method—accuracy: 0.995, precision: 0.960, recall: 0.960, F1 score: 0.960. This denotes a slight yet consistent edge of the “time-warping” approach in action recognition endeavors, with detailed classification outcomes presented in Figures 8 and 9.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 March 2024 doi:10.20944/preprints202403.0436.v1

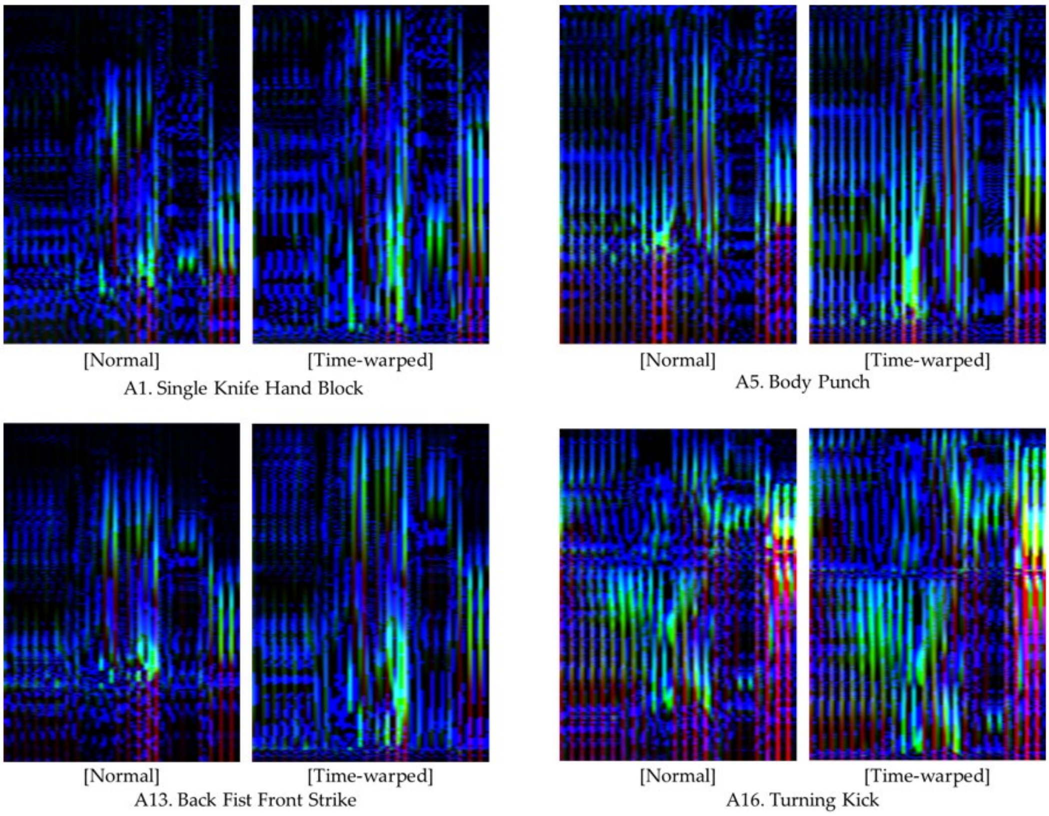


Figure 7. Comparative illustration of action images generated from normal and time-warped motion data.

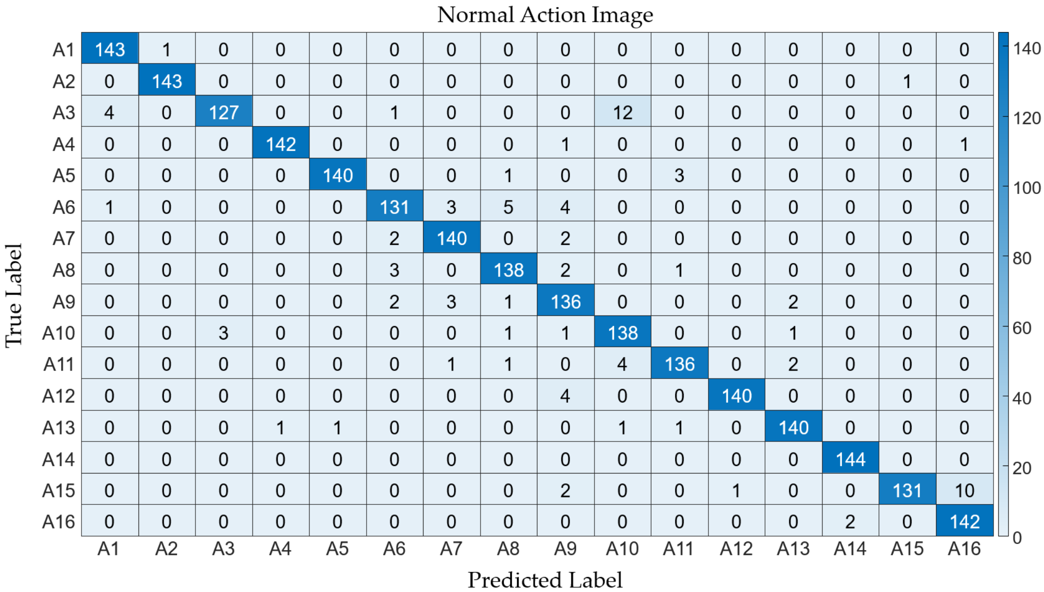


Figure 8. Confusion matrix illustrating the classification results of normal action image.

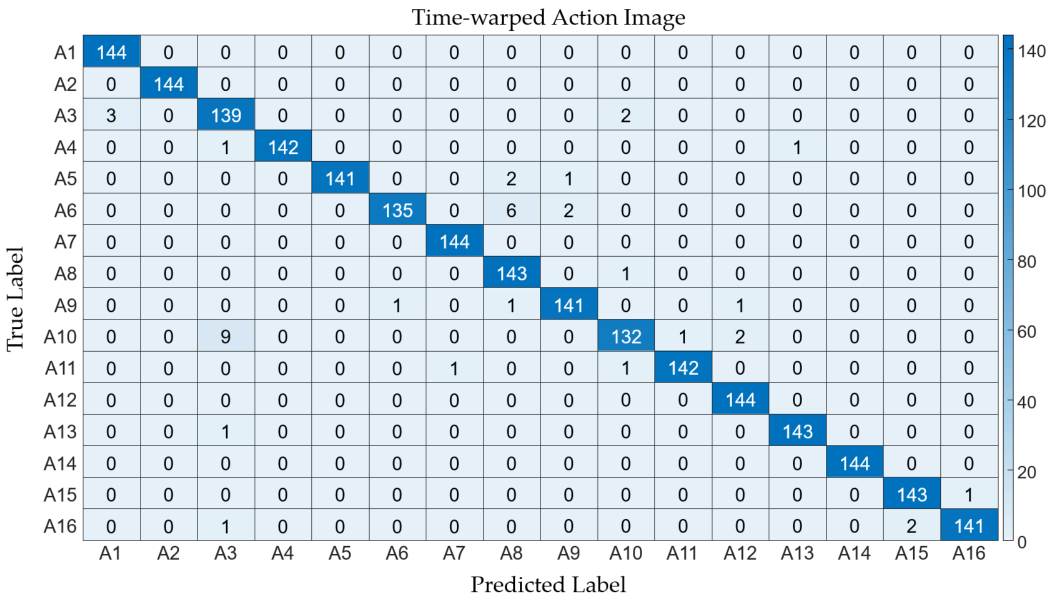


Figure 9. Confusion matrix illustrating the classification results of time-warped action images.

3.2. Performance Evaluation According to the Number of Joints

In the context of employing wearable IMU sensors for action recognition, reducing the number of sensors is imperative to curtail both cost and system complexity. This segment of the study delves into the scalability of the proposed method relative to the number of joints considered for generating an action image. Performance evaluations employed three variants of action images, derived from the time-warped motion data of 23, 8, and 4 joints, respectively. Figure 10 delineates these action image types. The intrinsic features of the action images are effectively conveyed in both the comprehensive (23 joints) and the condensed (eight and four joints) formats, as depicted in the figure. Table 3 enumerates the performance metrics across the three action image scales. It was observed that the average values of the performance metrics exhibit a slight decrement with the reduction in the number of joints involved. Specifically, the average accuracy for the comprehensive action images stood at 0.998, while the reduced-scale images registered average accuracies of 0.998 and 0.996 for eight and four joints, respectively, with analogous patterns observed in other metrics. Detailed classification outcomes are showcased in Figures 11 and 12.

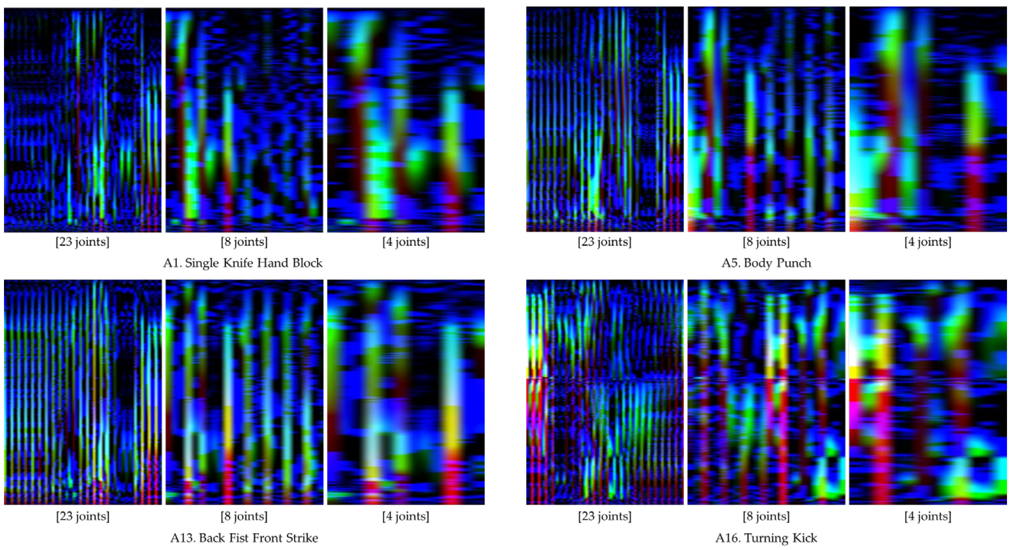


Figure 10. Comparative display of action images generated using motion data from 23, 8, and 4 joints.

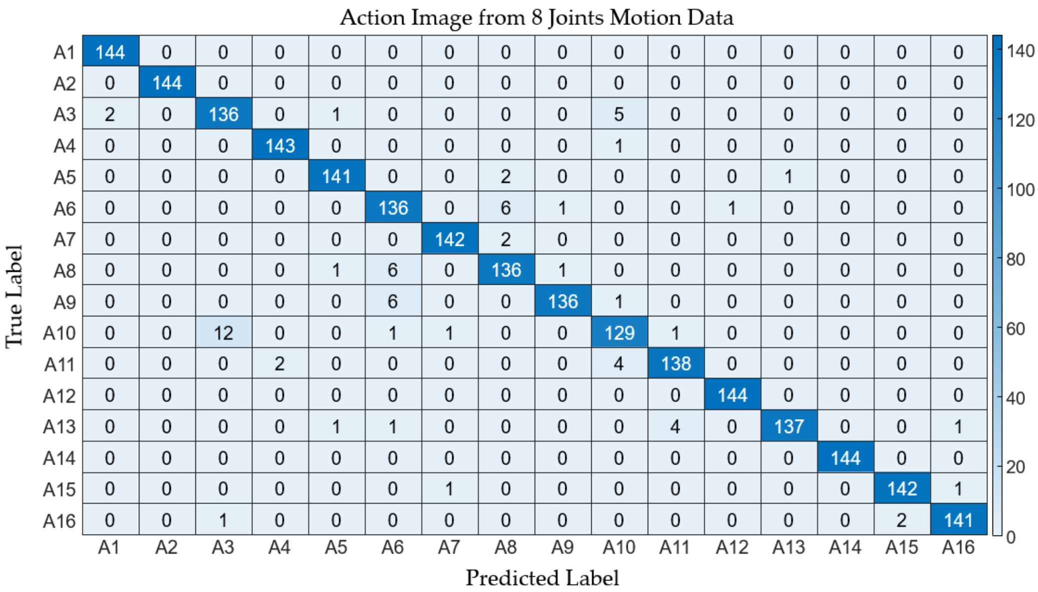


Figure 11. Confusion matrix depicting classification results for motion data involving 8 joints.

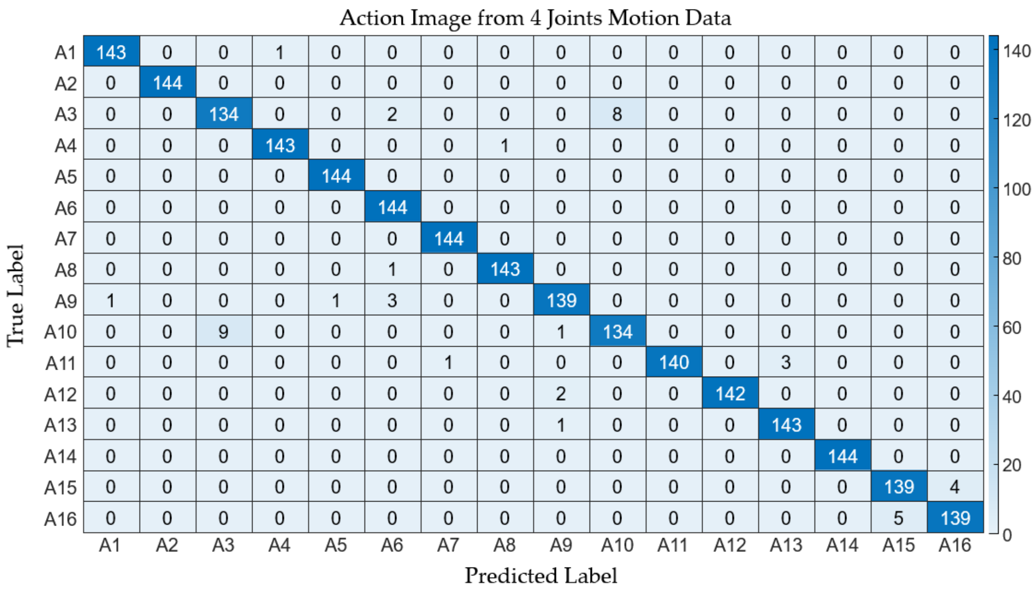


Figure 12. Confusion matrix depicting classification results for motion data involving 4 joints.

Table 3. Comparative analysis of performance metrics for action images generated from different scales of motion data.

	Accuracy			Precision			Recall			F1 score		
	23 joints	8 joints	4 joints	23 joints	8 joints	4 joints	23 joints	8 joints	4 joints	23 joints	8 joints	4 joints
A1	0.999	0.999	0.999	0.980	0.993	0.986	1.000	0.993	1.000	0.990	0.993	0.993
A2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
A3	0.993	0.992	0.991	0.921	0.937	0.913	0.965	0.931	0.944	0.942	0.934	0.928
A4	0.999	0.999	0.999	1.000	0.993	0.986	0.986	0.993	0.993	0.993	0.993	0.990
A5	0.999	1.000	0.997	1.000	0.993	0.979	0.979	1.000	0.979	0.989	0.997	0.979
A6	0.996	0.997	0.990	0.993	0.960	0.907	0.938	1.000	0.944	0.964	0.980	0.925
A7	1.000	1.000	0.998	0.993	0.993	0.986	1.000	1.000	0.986	0.997	0.997	0.986
A8	0.996	0.999	0.992	0.941	0.993	0.932	0.993	0.993	0.944	0.966	0.993	0.938

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 March 2024 doi:10.20944/preprints202403.0436.v1

A9	0.997	0.996	0.996	0.979	0.972	0.986	0.979	0.965	0.944	0.979	0.969	0.965
A10	0.993	0.992	0.989	0.971	0.944	0.921	0.917	0.931	0.896	0.943	0.937	0.908
A11	0.999	0.998	0.995	0.993	1.000	0.965	0.986	0.972	0.958	0.990	0.986	0.962
A12	0.999	0.999	1.000	0.980	1.000	0.993	1.000	0.986	1.000	0.990	0.993	0.997
A13	0.999	0.998	0.997	0.993	0.979	0.993	0.993	0.993	0.951	0.993	0.986	0.972
A14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
A15	0.999	0.996	0.998	0.986	0.965	0.986	0.993	0.965	0.986	0.990	0.965	0.986
A16	0.998	0.996	0.998	0.993	0.972	0.986	0.979	0.965	0.979	0.986	0.969	0.983
Ave.	0.998	0.998	0.996	0.983	0.981	0.970	0.982	0.980	0.969	0.982	0.981	0.969

4. Discussion

This study presents an innovative action recognition model designed for the quantitative evaluation of Poomsae using action images derived from motion data collected via wearable IMU sensors. This approach successfully addresses a critical shortcoming of previous vision-based methodologies, namely the difficulty in accurately capturing features indicative of rapid movements. By integrating rapidly updated IMU data with a time-warping modulation, the model effectively adjusts motion profiles within the time domain based on velocity, thereby significantly enhancing the depiction of rapid movements. Employing a CNN-based classification model to implement this method has showcased substantial efficacy across various metrics, underlining the model's robustness and potential applicability in the precise assessment of Poomsae performance.

The findings from the performance comparison substantiate that the time-warping method significantly augments action recognition capabilities. Through the application of time warping, there was a marked improvement in the average metrics of accuracy, precision, recall, and F1-score by 0.30%, 2.28%, 2.29%, and 2.29%, respectively. This method is particularly effective in enhancing the differentiation of actions that exhibit closely related motion characteristics. For example, the unit actions A6 (body punch) and A8 (face punch) demonstrate nearly indistinguishable motions, with the primary variance being the hand's direction during the punch. The rapid motion profile of the "punch" is more accurately represented through the application of time warping, leading to an improvement in performance. Specifically, the accuracy for A6 (body punch) increased from 99.1% to 99.6%, and for A8 (face punch), it rose from 99.3% to 99.6%. These enhancements underscore the efficacy of time warping in classifying motions with subtle differences, validating its effectiveness in the complex domain of martial arts technique recognition.

The utility of time warping extends to unit actions involving kicks, such as A14 (front kick), A15 (side kick), and A16 (turning kick), where it facilitated notable improvements across all evaluated performance metrics. For A14 (front kick), the metrics of accuracy, precision, recall, and F1 score improved from 99.9% to a perfect 100%. A15 (side kick) saw increases in accuracy (from 99.4% to 99.9%), precision (from 98.5% to 99.6%), recall (from 88.7% to 99.8%), and F1 score (from 93.8% to 99.6%). Similarly, A16 (turning kick) experienced enhancements in accuracy (from 99.4% to 99.8%), precision (from 93.4% to 99.6%), recall (from 92.8% to 99.2%), and F1 score (from 98.6% to 99.6%). These advancements in the recognition of Taekwondo kicks, especially where action profiles may share similar final poses but differ in movement trajectories and speeds, highlight the profound impact of time warping. It proves particularly beneficial in distinguishing between similar kick actions, thereby confirming its substantial value in the nuanced recognition of Taekwondo techniques.

In assessing the practical implications of the proposed action recognition model, this study delved into the system's scalability relative to the quantity of joint motion data utilized. A key advantage of reducing the number of sensors is the consequent decrease in both the cost and complexity of the equipment required, thereby rendering the technology more accessible and user-friendly. Crucially, the findings from this investigation reveal that the performance of the action recognition system remains robust, even when the amount of joint motion data is significantly reduced. This outcome is of paramount importance, demonstrating that a streamlined sensor setup, when integrated with the time-warping technique, is capable of maintaining high levels of accuracy and efficiency. Such insights are invaluable for practical implementations where the objectives include minimizing costs and simplifying operational complexity without sacrificing the accuracy of the system.

The focus of current research efforts has been on the classification of the 16 individual unit actions characteristic of Taekwondo. However, it is essential for future research endeavors to extend beyond this scope to encompass the evaluation of Poomsae. Poomsae represents sequences that combine various unit actions, wherein the precision and fluidity of each constituent movement are of critical importance. Drawing upon the knowledge acquired from the successful recognition and analysis of individual unit actions, advancing the development of algorithms capable of evaluating and analyzing the comprehensive execution of Poomsae emerges as a vital subsequent step. This progression will not only enhance the understanding and assessment of Poomsae performances but also contribute significantly to the broader field of martial arts technique analysis and evaluation.

5. Conclusions

This paper presented an action recognition model tailored for Taekwondo unit actions, employing action images generated from full-body joint motion data captured via IMU sensors. The proposed model augmented the representation of rapid motion by modulating motion profiles through the application of time-warping techniques, thereby facilitating the identification of subtle differences in motion characteristics. The comparative analysis of performance underscored not only the method's efficacy but also its scalability, affirming its utility across varying scales of joint motion data. In conclusion, our research has contributed a novel approach to the recognition and analysis of Taekwondo actions, underscoring the potential of wearable IMU sensors in capturing the nuances of fast and complex movements. This advancement marks a significant step forward in the integration of technology with traditional sports training and assessment practices, heralding a future where technological innovations enhance athletic performance and training methodologies.

6. Patents

Supplementary Materials:

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript." Please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: This work was supported in part by Human Resources Development Program of the Korea Institute of Energy Technology Evaluation and Planning(KETEP) grant funded by the Ministry of Trade, Industry and Energy, Republic of Korea (No.RS-2023-00237035), and was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2018R1A5A7023490). This paper was written as part of Konkuk University's research support program for its faculty on sabbatical leave in 2022.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Konkuk University (7001355-202004-HR-372).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data cannot be provided due to the security reason.

Acknowledgments:

Conflicts of Interest: The authors declare no conflict of interest.

References

1. J. Ryu, J. Seo, H. Jebelli, and S. Lee, "Automated Action Recognition Using an Accelerometer-Embedded Wristband-Type Activity Tracker," J. Constr. Eng. Manag., vol. 145, no. 1, p. 04018114, Jan. 2019, doi: 10.1061/(ASCE)CO.1943-7862.0001579.
2. K. M. Rashid and J. Louis, "Times-series data augmentation and deep learning for construction equipment activity recognition," Adv. Eng. Inform., vol. 42, p. 100944, Oct. 2019, doi: 10.1016/j.aei.2019.100944.
3. J. Zhang, L. Zi, Y. Hou, M. Wang, W. Jiang, and D. Deng, "A Deep Learning-Based Approach to Enable Action Recognition for Construction Equipment," Adv. Civ. Eng., vol. 2020, pp. 1–14, Nov. 2020, doi: 10.1155/2020/8812928.
4. T. T. Zin et al., "Real-Time Action Recognition System for Elderly People Using Stereo Depth Camera," Sensors, vol. 21, no. 17, p. 5895, Sep. 2021, doi: 10.3390/s21175895.
5. H. Yan, B. Hu, G. Chen, and Z. E, "Real-Time Continuous Human Rehabilitation Action Recognition using OpenPose and FCN," in 2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), Shenzhen, China, Apr. 2020, pp. 239–242. doi: 10.1109/AEMCSE50948.2020.00058.
6. G. Maeda, M. Ewerton, G. Neumann, R. Lioutikov, and J. Peters, "Phase estimation for fast action recognition and trajectory generation in human–robot collaboration," Int. J. Robot. Res., vol. 36, no. 13–14, pp. 1579–1594, Dec. 2017, doi: 10.1177/0278364917693927.
7. M.-A. Fiedler, P. Werner, A. Khalifa, and A. Al-Hamadi, "SFPD: Simultaneous Face and Person Detection in Real-Time for Human–Robot Interaction," Sensors, vol. 21, no. 17, p. 5918, Sep. 2021, doi: 10.3390/s21175918.
8. R. Cuperman, K. M. B. Jansen, and M. G. Ciszewski, "An End-to-End Deep Learning Pipeline for Football Activity Recognition Based on Wearable Acceleration Sensors," Sensors, vol. 22, no. 4, p. 1347, Feb. 2022, doi: 10.3390/s22041347.
9. P. Yanan, Y. Jilong, and Z. Heng, "Using Artificial Intelligence to Achieve Auxiliary Training of Table Tennis Based on Inertial Perception Data," Sensors, vol. 21, no. 19, p. 6685, Oct. 2021, doi: 10.3390/s21196685.
10. A. Tejero-de-Pablos, Y. Nakashima, T. Sato, N. Yokoya, M. Linna, and E. Rahtu, "Summarization of User-Generated Sports Video by Using Deep Action Recognition Features," IEEE Trans. Multimed., vol. 20, no. 8, pp. 2000–2011, Aug. 2018, doi: 10.1109/TMM.2018.2794265.
11. S. Sharma, R. Kiros, and R. Salakhutdinov, "Action Recognition using Visual Attention." arXiv, Feb. 14, 2016. Accessed: Jun. 18, 2022. [Online]. Available: http://arxiv.org/abs/1511.04119
12. J. C. de Goma, M. S. Bustos, J. A. Sebastian, and J. J. E. Macrohon, "Detection of Taekwondo Kicks Using RGB-D Sensors," in Proceedings of the 2019 3rd International Conference on Software and e-Business, Tokyo Japan, Dec. 2019, pp. 129–133. doi: 10.1145/3374549.3374576.
13. Choi, C.-H.; Joo, H.-J. Motion recognition technology based remote Taekwondo Poomsae evaluation system. *Multimed Tools Appl* **2016**, *75*, 13135–13148, doi:10.1007/s11042-015-2901-1.
14. Jeong Min Seo, Il Ki Jang, Jeong Hoon Choi, Sang Moon Lee A STUDY OF THE TAEKWONDO POOMSAE RECOGNITION SYSTEM USED BY MOTION RECOGNITION TECHNIQUES.
15. Kong, Y.; Wei, Z.; Huang, S. Automatic analysis of complex athlete techniques in broadcast taekwondo video. *Multimed Tools Appl* **2018**, *77*, 13643–13660, doi:10.1007/s11042-017-4979-0.
16. J. Liang and G. Zuo, "Taekwondo Action Recognition Method Based on Partial Perception Structure Graph Convolution Framework," Sci. Program., vol. 2022, pp. 1–10, Feb. 2022, doi: 10.1155/2022/1838468.
17. J. Lee and H. Jung, "TUHAD: Taekwondo Unit Technique Human Action Dataset with Key Frame-Based CNN Action Recognition," Sensors, vol. 20, no. 17, p. 4871, Aug. 2020, doi: 10.3390/s20174871.
18. Sohaib Laraba et al., "3D skeleton-based action recognition by representing motion capture sequences as 2D-RGB images," *Computer Animation and Virtual Worlds*, vol. 28, no.3-4. P. e1782, 2017
19. H. Duan, Y. Zhao, K. Chen, D. Lin, and B. Dai, "Revisiting Skeleton-Based Action Recognition," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 2969–2978.
20. M. Feng and J. Meunier, "Skeleton Graph-Neural-Network-Based Human Action Recognition: A Survey," *Sensors*, 2022, 22, 2091.
21. Wang, X.; Dai, Y.; Gao, L.; Song, J. Skeleton-Based Action Recognition via Adaptive Cross-Form Learning. In Proceedings of the 30th ACM International Conference on Multimedia, 2022; pp. 1670–1678.
22. Chan, W.; Tian, Z.; Wu, Y. GAS-GCN: Gated Action-Specific Graph Convolutional Networks for Skeleton-Based Action Recognition. Sensors 2020, 20, 3499; DOI:10.3390/s20123499.
23. Wang, K.; Deng, H. TFC-GCN: Lightweight Temporal Feature Cross-Extraction Graph Convolutional Network for Skeleton-Based Action Recognition. Sensors 2023, 23, 5593; DOI:10.3390/s23125593.
24. Liu, D.; Xu, H.; Wang, J.; Lu, Y.; Kong, J.; Qi, M. Adaptive Attention Memory Graph Convolutional Networks for Skeleton-Based Action Recognition. Sensors 2021, 21, 6761.

25. Dong, J.Q.; Gao, Y.B.; Lee, H.J.; Zhou, H.; Yao, Y.F.; Fang, Z.J.; Huang, B. Action Recognition Based on the Fusion of Graph Convolutional Networks with High Order Features. Applied Sci. 2020, 10, 1482
26. Jiang, Y.; Sun, Z.; Yu, S.; Wang, S.; Song, Y. A Graph Skeleton Transformer Network for Action Recognition. Symmetry 2022, 14, 1547.
27. Tu, Z.; Zhang, J.; Li, H.; Chen, Y.; Yuan, J. Joint-bone Fusion Graph Convolutional Network for Semi-supervised Skeleton Action Recognition.
28. Sanchez, J.; Neff, C.; Tabkhi, H. Real-World Graph Convolution Networks (RW-GCNs) for Action Recognition in Smart Video Surveillance. CoRR 2022, abs/2201.05739. Available online: https://arxiv.org/abs/2201.05739
29. Jaramillo, I.E.; Jeong, J.G.; Lopez, P.R.; Lee, C.H.; Kang, D.Y.; Ha, T.J.; Oh, J.H.; Jung, H.; Lee, J.H.; Lee, W.H.; et al. Real-Time Human Activity Recognition with IMU and Encoder Sensors in Wearable Exoskeleton Robot via Deep Learning Networks. Sensors 2022, 22, 9690
30. Brownlee, J. LSTMs for Human Activity Recognition Time Series Classification. Machine Learning Mastery. 2018. Available online: https://machinelearningmastery.com/how-to-develop-rnn-models-for-human-activity-recognition-time-series-classification/
31. Kwon, H.; Abowd, G.D.; Plötz, T. Complex Deep Neural Networks from Large Scale Virtual IMU Data for Effective Human Activity Recognition Using Wearables. Sensors 2021, 21, 8337. Available online: https://www.mdpi.com/1424-8220/21/24/8337
32. Yin, X.; Liu, Z.; Liu, D.; Ren, X. A Novel CNN-based Bi-LSTM Parallel Model with Attention Mechanism for Human Activity Recognition with Noisy Data. Scientific Reports 2022, 12, 1–11.
33. Rey, V.; Hevesi, P.; Kovalenko, O.; Lukowicz, P. Let there be IMU data: generating training data for wearable, motion sensor based activity recognition from monocular RGB videos. In Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the ACM International Symposium on Wearable Computers, 2019; pp. 699–708.
34. Roetenberg, D.; Luinge, H.; Slycke, P. Xsens MVN: Full 6DOF human motion tracking using miniature inertial sensors. Xsens Motion Technol. BV Tech. Rep. 2009, 1, 1–7.