

Article

Expert–Novice Level Classification Using Graph Convolutional Network Introducing Confidence-Aware Node-Level Attention Mechanism

Tatsuki Seino ¹, Naoki Saito ², Takahiro Ogawa ³, Satoshi Asamizu ⁴ and Miki Haseyama ^{3,*}

- ¹ Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan; seino@lmd.ist.hokudai.ac.jp
- ² Office of Institutional Research, Hokkaido University, Sapporo 060-0808, Japan; saito@lmd.ist.hokudai.ac.jp
- ³ Faculty of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan; ogawa@lmd.ist.hokudai.ac.jp
- ⁴ National Institute of Technology, Kushiro College, Kushiro 084-0916, Japan; asamizu@kushiro-ct.ac.jp
- * Correspondence: mhaseyama@lmd.ist.hokudai.ac.jp

Abstract: In this study, we propose a classification method of expert–novice levels using a graph convolutional network (GCN) with a confidence-aware node-level attention mechanism. In classification using an attention mechanism, highlighted features may not be significant for accurate classification, thereby degrading classification performance. To address this issue, the proposed method introduces a confidence-aware node-level attention mechanism into a spatiotemporal attention GCN (STA-GCN) for the classification of expert–novice levels. Consequently, our method can contrast the attention value of each node on the basis of the confidence measure of the classification, which solves the problem of classification approaches using attention mechanisms and realizes accurate classification. Furthermore, because the expert–novice levels have ordinalities, using a classification model that considers ordinalities improves the classification performance. The proposed method involves a model that minimizes a loss function that considers the ordinalities of classes to be classified. By implementing the above approaches, the expert–novice level classification performance is improved.



Citation: Seino, T.; Saito, N.; Ogawa, T.; Asamizu, S.; Haseyama, M. Expert–Novice Level Classification Using Graph Convolutional Network Introducing Confidence-Aware Node-Level Attention Mechanism. *Sensors* **2024**, *24*, 3033. <https://doi.org/10.3390/s24103033>

Academic Editor: Carlos M. Travieso-González

Received: 10 April 2024
Revised: 6 May 2024
Accepted: 8 May 2024
Published: 10 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: expert–novice level classification; motion data; graph convolutional network; attention mechanism

1. Introduction

In the context of sports, the transfer of “expert techniques” from outstanding athletes and coaches to the next generation of players is essential for development. However, most expert techniques are tacit knowledge, and the transfer of such techniques requires prolonged guidance from experienced athletes or coaches. Thus, the construction of support technologies to facilitate the efficient transfer of these expert techniques is expected. To effectively implement support technologies, it is essential to delineate the differences between expert and novice athletes. Therefore, the classification of athletes into “expert” and “novice” is a fundamental methodology [1]. In recent years, the popularization of wearable devices, such as smartwatches and motion capture devices, has facilitated the acquisition of biometric data, and various methods for expert–novice level classification using biometric data have been proposed [2–20]. For example, Kuo et al. proposed a classification method for laparoscopic surgical skills based on multiple machine learning methods such as a multilayer perceptron using gaze information [19]. Furthermore, Guo et al. proposed a skill-level classification method based on convolutional neural networks using a Single Inertial Sensor attached to the arm [20]. In particular, motion data are closely related to tacit knowledge. Several expert–novice level classification methods using motion data have been proposed [21–27]. For example, Ross et al. proposed a method for classifying athletes’ skill levels using machine learning techniques such as support vector machines

and logistics regression based on motion data collected during specific movements [25]. Furthermore, Vincenzo et al. proposed a method for classifying violin performance levels by the random forest method using motion data [26]. In addition, Xuan et al. proposed a method for classifying surgical skill levels using a convolutional neural network (CNN) and long short-term memory (LSTM) based on motion data collected during surgical simulation [27]. From the above methods, expert–novice classification is realized using a machine-learning-based approach with biometric information. In particular, motion data have attracted attention as information that can accurately classify expert–novice levels and are used in many previous methods.

In classification tasks using motion data, many methods handle motion as a graph structure, which is used to construct a graph convolutional network (GCN) [28] based on motion data [29–38]. GCNs allow the relationships between joints in the human body to be captured in a graph structure, facilitating the classification of complex movements. However, conventional GCNs, typically used in classification tasks, only output classification results without providing explanations. Therefore, there is a need for methods that elucidate the reasoning behind the classification results. In this regard, a classification method using motion data via a spatiotemporal attention GCN (STA-GCN), which introduces the attention mechanism into the GCN, has been proposed [39]. The STA-GCN improves classification performance and provides explanations for the classification results. In the STA-GCN, the feature extractor is placed close to the input, whereas the attention and perception branches are placed closer to the output. The attention branch performs classification using feature maps obtained using the feature extractor and generates attention nodes and edges. These generated attention nodes and edges are used to highlight the parts that are critical for accurate classification. Conversely, the perception branch performs the final classification using feature maps, attention nodes, and attention edges derived from both the feature extractor and the attention branch. From the above procedures, the attention mechanism in the STA-GCN enables the highlighting of important parts for classification. However, if the parts emphasized by the attention mechanism differ from the actual focal parts, there is a potential for reduced classification performance [40]. Therefore, in the attention mechanism, the influence of attention that fails to highlight important parts needs to be diminished.

To address this issue, we previously proposed an expert–novice level classification method (confidence-aware STA-GCN: ConfSTA-GCN) that introduces an attention mechanism that considers the confidence measure of critical parts [41]. Because the confidence measure is treated as the probability of class assignment obtained through expert–novice level classification in the attention branch, the same confidence measure is applied as a weight to all attention nodes in the previous method. However, it is anticipated that there will be variations in the confidence measure at each attention node for accurate classification. Therefore, by calculating different confidence measures for each attention node, classification performance can be improved. In addition, previous methods construct classification models under the assumption that there is no ordinality between the classes of the expert and novice levels; thus, they do not consider relationships between these classes. Given the ordinariness of the expert–novice levels, this can lead to limitations in classification performance.

In this study, we propose a method for expert–novice level classification using a GCN with a confidence-aware node-level attention mechanism. The proposed method calculates the probability of belonging to an actual expert–novice level when specific attention nodes are excluded. This process is repeated for the number of attention nodes, and the computed probabilities are regarded as confidence measures. A novel attention mechanism that considers the confidence measure of each attention node is one of the main contributions of this study. The perception branch outputs the final classification results using feature maps computed from these attention nodes, and these features are adjusted according to the confidence measure. Furthermore, the proposed method considers the ordinality of the expert–novice level, an aspect not considered in previous methods using attention

mechanisms. Because this allows for consideration of the relationships between classes, it is expected to further improve of expert–novice level classification performance. The main contributions of this study can be summarized as follows.

- Proposal of a method for improving existing GCN-based classification approaches by individually calculating and applying the confidence measure to attention nodes.
- Construction of a classification model that allows for consideration of the order of expert–novice levels among classes.

Note that this is an extended version of the ConfSTA-GCN for skeleton-based expert–novice level classification [41]. Specifically, the proposed method can calculate the confidence measure for each joint, separately, resulting in a node-level attention mechanism.

This paper is organized as follows. In Section 2, the classification of the expert–novice levels using the STA-GCN with a confidence-aware node-level attention mechanism is explained. The experimental results are described in Section 3 to evaluate the classification performance of our method. Finally, Section 4 concludes this study and describes future work.

2. Classification of Expert–Novice Levels Using STA-GCN with Confidence-Aware Node-Level Attention Mechanism

This section describes the proposed method to improve the existing approach and this study’s novelty. It also structures a classification model considering the expert–novice level ordering relationship between classes. An overview of the proposed method is shown in Figure 1. The proposed method comprises a feature extractor, an attention branch, and a perception branch. First, the proposed method uses a spatiotemporal graph (ST-graph) to represent spatial and temporal motion data as a graph structure. Using the feature extractor process, feature maps are calculated. Next, the attention branch obtains attention nodes, which represent the significance of each joint, and attention edges, which indicate the important relationships between joints. Furthermore, the attention branch uses the confidence-aware node-level attention mechanism to generate a new feature map in which important joints and their connections are emphasized for classification. Finally, by inputting the attention nodes and edges along with the feature map into the perception branch, the classification results of the expert–novice levels are obtained. During this process, the attention nodes computed using the attention branch are output as a visualization of important joints for expert–novice level classification.

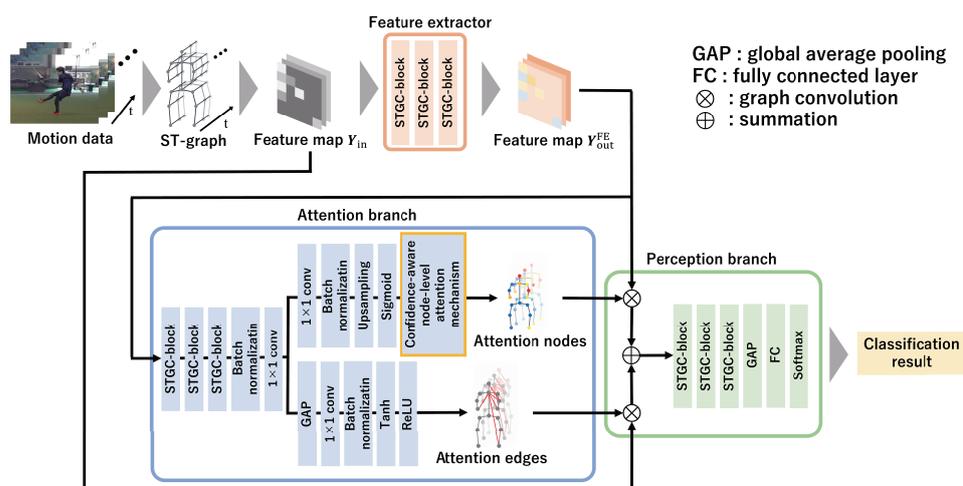


Figure 1. Overview of the proposed method. In the proposed method, feature maps extracted from graphed motion data are used to calculate attention nodes and edges through an attention mechanism that considers the confidence measure in emphasizing elements crucial for classification. Subsequently, the classification results of the expert–novice levels are obtained using the perception branch. Furthermore, the attention nodes used in the classification are visualized.

2.1. ST-Graph Construction and Feature Extractor

This subsection describes the calculation of features that take into account spatiotemporal information from motion data. Specifically, we describe the construction approach of an ST-graph and the feature extractor method separately. First, the proposed method constructs an ST-graph from motion data in the same manner as [29]. Specifically, the ST-graph represents human joints as nodes $v_{f,n}$ ($f = 1, 2, \dots, F$; F denoting the number of frames, $n = 1, 2, \dots, N$; N representing the number of nodes), as shown in Figure 2. The ST-graph connects them with inter-frame and intra-body edges. The inter-frame edges connect the same joint across consecutive frames, i.e., the f -th and $(f + 1)$ -th frames in the motion data. Conversely, the intra-body edges connect nodes in the ST-graph according to the adjacency relationships of each joint in the human body.

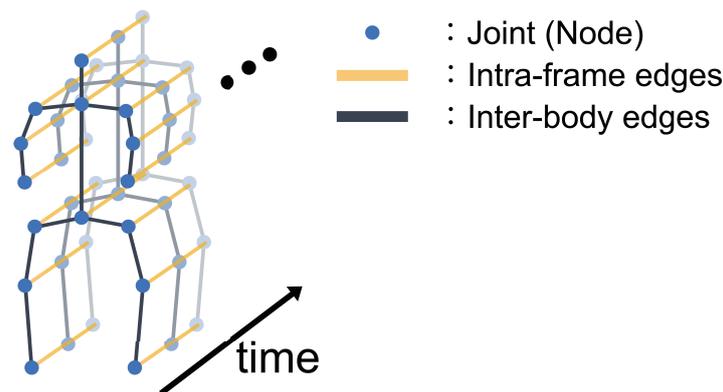


Figure 2. Overview of ST-graph constructed using the proposed method. The ST-graph is constructed by connecting nodes representing joints (blue points) with inter-frame edges (yellow lines) and intra-body edges (black lines).

The proposed method computes the feature map using a spatiotemporal graph convolutional (STGC)-block. The network configuration of the STGC-block is depicted in Figure 3. This block performs spatial graph convolution (S-GC) and temporal graph convolution (T-GC). Let $\mathbf{y}(v_{f,n}) \in \mathbb{R}^D$ (D denoting the dimension of node features) be the feature vector for the n -th node in the f -th frame. Our method defines the feature map obtained from the ST-graph as $\mathbf{Y}_{\text{in}} = [\mathbf{y}(v_{f,1}), \mathbf{y}(v_{f,2}), \dots, \mathbf{y}(v_{f,N})]^T \in \mathbb{R}^{N \times D}$. First, the output $\mathbf{Y}_{\text{out}}^{\text{space}} \in \mathbb{R}^{N \times N}$, which is obtained by applying S-GC to the feature map \mathbf{Y}_{in} , is computed as follows:

$$\mathbf{Y}_{\text{out}}^{\text{space}} = \sum_{h=1}^H \mathbf{W}_h^{\text{Edge}} \circ (\mathbf{\Lambda}_h^{-1/2} (\mathbf{A}_h^{\text{space}} + \mathbf{I}) \mathbf{\Lambda}_h^{-1/2}) \mathbf{Y}_{\text{in}} \mathbf{W}_h^{\text{Node}}, \quad (1)$$

where $\mathbf{W}_h^{\text{Node}}$ and $\mathbf{W}_h^{\text{Edge}}$ ($h = 1, 2, \dots, H$; H denoting the number of adjacent nodes connected by intra-body edges) denote the weight matrices of the nodes and edges, respectively, and $\mathbf{A}_h^{\text{space}}$ denotes the adjacency matrix in the spatial direction. The symbol “ \circ ” denotes the Hadamard product, and $\mathbf{I} \in \mathbb{R}^{N \times N}$ denotes the identity matrix. Furthermore, $\mathbf{\Lambda}_h \in \mathbb{R}^{N \times N}$ denotes a diagonal matrix whose diagonal elements are $\Lambda_{nn} = \sum_i (A_{ni} + I_{ni})$.

The proposed method calculates the output $\mathbf{Y}_{\text{out}}^{\text{time}} \in \mathbb{R}^{F \times N \times D}$ using T-GC as follows:

$$\mathbf{Y}_{\text{out}}^{\text{time}} = \begin{bmatrix} \mathbf{y}^{\text{time}}(v_{1,1}) & \mathbf{y}^{\text{time}}(v_{1,2}) & \cdots & \mathbf{y}^{\text{time}}(v_{1,N}) \\ \mathbf{y}^{\text{time}}(v_{2,1}) & \mathbf{y}^{\text{time}}(v_{2,2}) & \cdots & \mathbf{y}^{\text{time}}(v_{2,N}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}^{\text{time}}(v_{F,1}) & \mathbf{y}^{\text{time}}(v_{F,2}) & \cdots & \mathbf{y}^{\text{time}}(v_{F,N}) \end{bmatrix} \quad (2)$$

$$\mathbf{y}^{\text{time}}(v_{f,n}) = \sum_{\tau=-\lfloor \kappa/2 \rfloor}^{\lfloor \kappa/2 \rfloor} \alpha_{\tau} \circ \mathbf{x}(v_{f-\tau,n}) \in \mathbb{R}^D, \quad (3)$$

where κ denotes the size of the T-GC kernel and $\alpha_{\tau} \in \mathbb{R}^D$ denotes the weight vector of T-GC. In the STGC-block, $\mathbf{Y}_{\text{out}}^{\text{FE}} \in \mathbb{R}^{F \times N \times D}$ is calculated using the network architecture shown in Figure 3. The STGC-block consists of S-GC, batch normalization [42], the ReLU activation function [43], T-GC, and Dropout [44], with a skip connection [45].

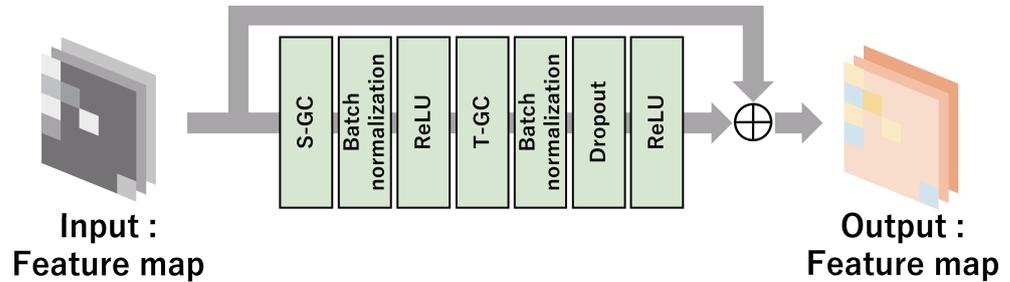


Figure 3. Network configuration of STGC-block in the proposed method.

2.2. Attention Branch

This work aims to improve the attention branch in the existing GCN-based classification. The attention edges $E(\mathbf{Y}_{\text{out}}^{\text{FE}}) \in \mathbb{R}^{1 \times N \times N}$ are derived by applying several 1×1 convolution layers [46] and global average pooling (GAP) [46] to the feature map $\mathbf{Y}_{\text{out}}^{\text{FE}}$. This is followed by batch normalization and the application of the Tanh and ReLU functions to convert the values of non-important parts to zero. The attention edges $E(\mathbf{Y}_{\text{out}}^{\text{FE}})$, which contain only the connections important for classification, are computed by employing several 1×1 convolution layers and GAP to the feature map $\mathbf{Y}_{\text{out}}^{\text{FE}}$. Subsequently, batch normalization is performed, and the Tanh and ReLU activation functions are used to convert the values of non-essential parts to zero, thereby computing the attention edges.

Furthermore, the attention branch in the proposed method employs a process to obtain the attention nodes and edges using the feature map $\mathbf{Y}_{\text{out}}^{\text{FE}}$ calculated in the previous subsection. The attention nodes $\mathbf{V}(\mathbf{Y}_{\text{out}}^{\text{FE}}) \in \mathbb{R}^{1 \times N \times N}$ are obtained by applying several 1×1 convolution layers, batch normalization, upsampling, and the sigmoid function to $\mathbf{Y}_{\text{out}}^{\text{FE}}$. In the upsampling process, linear interpolation is performed so that the number of frames in the feature map after the 1×1 convolution processes and batch normalization in T-GC and the input feature map $\mathbf{Y}_{\text{out}}^{\text{FE}}$ become the same. Using the computed attention nodes $\mathbf{V}(\mathbf{Y}_{\text{out}}^{\text{FE}})$, a new feature map $\mathbf{Y}_{\text{out}}^{\text{AN}} \in \mathbb{R}^{1 \times F \times N}$, which emphasizes information about important parts for expert–novice level classification, is computed as follows:

$$\mathbf{Y}_{\text{out}}^{\text{AN}} = \mathbf{V}(\mathbf{Y}_{\text{out}}^{\text{FE}}) \mathbf{Y}_{\text{out}}^{\text{FE}}. \quad (4)$$

The proposed method uses these attention nodes and edges for expert–novice level classification.

Our method applies the confidence-aware node-level attention mechanism to the feature map $\mathbf{Y}_{\text{out}}^{\text{AN}}$ to emphasize important nodes. We obtain a novel feature map $\mathbf{Y}_{\text{out}}^{\text{CAN}}$ from the attention nodes. In the confidence-aware node-level attention mechanism, we first calculate the confidence measure of each attention node. The calculation approach for the confidence measure is depicted in Figure 4. The proposed method computes the probability of belonging to each expert–novice level via a network in the attention branch when one of the attention nodes is masked, i.e., the attention value of the target node is set to zero. Let $c_{n,f}$ be the probability value calculated when the n -th attention node in the f -th frame is masked. The proposed method derives the confidence measure $\bar{c}_{n,f}$ of the n -th attention node in the f -th frame as follows:

$$\bar{c}_{n,f} = 1 - c_{n,f}. \quad (5)$$

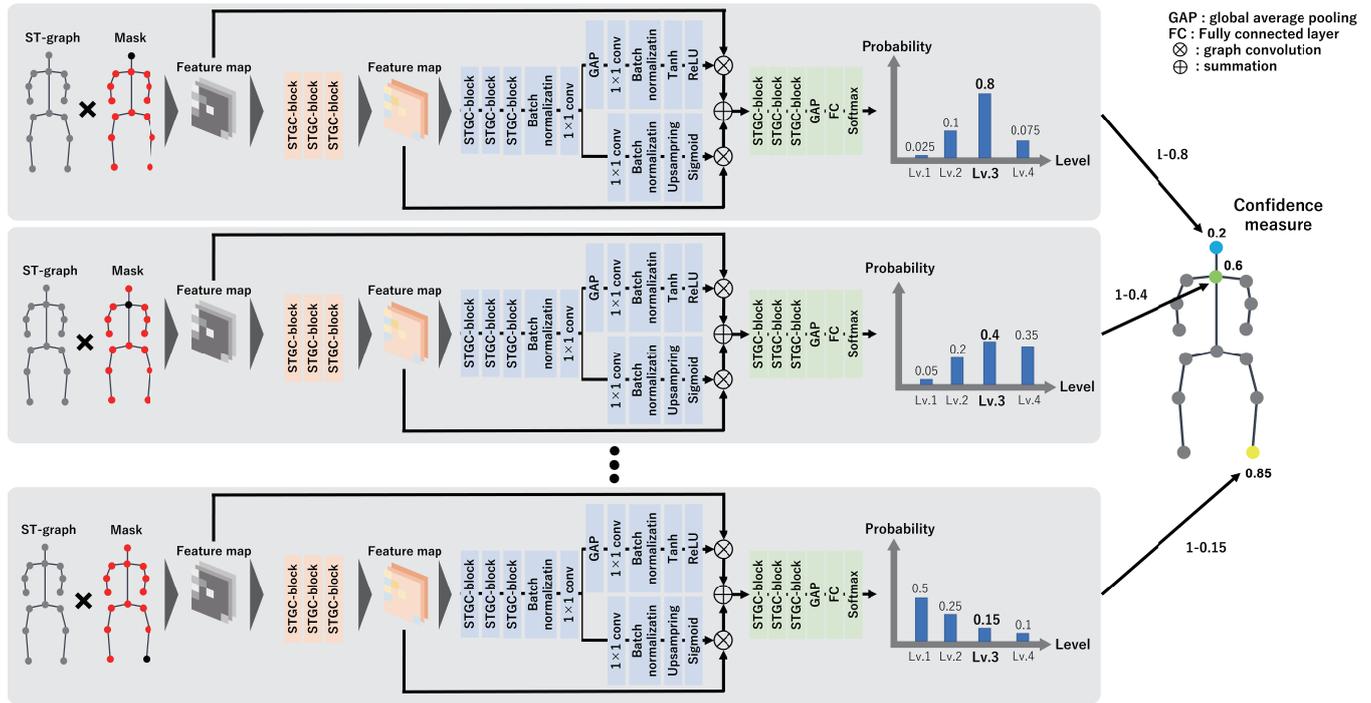


Figure 4. Overview of the calculation approach for the confidence measure. The proposed method calculates the probability of belonging to each expert–novice level by masking one attention node (setting its attention node to zero) and derives the confidence measure on the basis of the probability value. In this attention mechanism, the product of the calculated confidence measure and the attention node is taken, allowing the calculation of controlled attention nodes.

In the confidence-aware node-level attention mechanism, we calculate the confidence measure for all attention nodes, and the feature maps Y_{out}^{CAN} are calculated using the feature maps Y_{out}^{AN} and the confidence measure, as shown in the following equations:

$$Y_{out}^{CAN} = C \circ Y_{out}^{AN}, \quad (6)$$

$$C = \begin{bmatrix} \bar{c}_{11} & \bar{c}_{12} & \cdots & \bar{c}_{1N} \\ \bar{c}_{21} & \bar{c}_{22} & \cdots & \bar{c}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{c}_{F1} & \bar{c}_{F2} & \cdots & \bar{c}_{FN} \end{bmatrix} \in \mathbb{R}^{1 \times F \times N}. \quad (7)$$

Equation (6) enables the controlled influence of attention nodes calculated using the confidence-aware node-level attention mechanism, moderated by the confidence measure of each attention node. From the above, the proposed method obtains the attention edges $E(Y_{out}^{FE})$ and the feature maps Y_{out}^{CAN} calculated from the attention nodes $V(Y_{out}^{FE})$ in the attention branch to realize accurate expert–novice level classification.

2.3. Perception Branch

This work aims to compute classification results from features acquired in the feature extraction and attention branch. The perception branch obtains the final expert–novice level classification results by using a new feature map Y_{out} . First, the proposed method computes $Y_{out}^{per} \in \mathbb{R}^{N \times N}$ using the graph convolution of the attention edge and the feature map Y_{in} obtained from the ST-graph as follows:

$$Y_{out}^{per} = \sum_{\varphi=1}^{\phi} (\Lambda_{\varphi}^{-1/2} (A_{\varphi}^{per} + I) \Lambda_{\varphi}^{-1/2}) Y_{in} W_{\varphi}^{per}, \quad (8)$$

where $A_\phi^{\text{per}} \in \mathbb{R}^{N \times N}$ ($\phi = 1, 2, \dots, \phi$; ϕ denoting the number of attention edges) denotes a normalized adjacency matrix of the attention edges, and $W_\phi^{\text{per}} \in \mathbb{R}^{D \times N}$ denotes the weight matrix. Furthermore, $\Lambda_\phi \in \mathbb{R}^{N \times N}$ denotes a diagonal matrix. The proposed method obtains the final feature map $Y_{\text{out}} \in \mathbb{R}^{N \times N}$ using $Y_{\text{out}}^{\text{space}}$ calculated using Equation (1) and $Y_{\text{out}}^{\text{per}}$ as follows:

$$Y_{\text{out}} = Y_{\text{out}}^{\text{space}} + Y_{\text{out}}^{\text{per}}. \quad (9)$$

Using multiple STGC-blocks, GAP, a fully connected layer, and the softmax function, we calculate the probability of belonging to each expert–novice level class. The class with the highest probability is considered the final classification result in the proposed method.

2.4. Training Approach

The purpose of this work is to construct a GCN learning approach that considers ordinality. The proposed method learns the STA-GCN by minimizing the loss function $\mathcal{L}_{\text{total}}$, which is calculated on the basis of the probability of belonging to each expert–novice level class. Specifically, $\mathcal{L}_{\text{total}}$ is defined as follows:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{att}} + \mathcal{L}_{\text{per}}, \quad (10)$$

$$\mathcal{L}_{\text{att}} = - \sum_{m=1}^M q^{(m)} \log p_{\text{att}}^{(m)} - \sum_{m=1}^M \left(|\text{label} - m|^2 (1 - \delta^{(m)}) \log \left(1 - p_{\text{att}}^{(m)} \right) \right), \quad (11)$$

$$\mathcal{L}_{\text{per}} = - \sum_{m=1}^M q^{(m)} \log p_{\text{per}}^{(m)} - \sum_{m=1}^M \left(|\text{label} - m|^2 (1 - \delta^{(m)}) \log \left(1 - p_{\text{per}}^{(m)} \right) \right), \quad (12)$$

where M represents the number of classes corresponding to the expert–novice levels, and $\text{label} \in \{1, 2, \dots, M\}$ denotes the ground truth of the expert–novice levels. $p_{\text{att}}^{(m)}$ and $p_{\text{per}}^{(m)}$ denote the probabilities of belonging to the m -th class ($m = 1, 2, \dots, M$), as determined by the attention and perception branches, respectively. $\delta^{(m)}$ is defined as follows:

$$\delta^{(m)} = \begin{cases} 1 & \text{if } m = \text{label}, \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

In the proposed method, the squared difference between the ground truth and the classification result is used as a weight in the second term of Equations (11) and (12). Consequently, the loss function outputs larger values when there is a more significant discrepancy between the ground truth and the classification result. In addition, to ensure a certain level of accuracy in the attention edges and nodes, the sum of the loss functions \mathcal{L}_{per} and \mathcal{L}_{att} at the attention branch point is minimized. By minimizing the defined loss function $\mathcal{L}_{\text{total}}$, the parameters in the STA-GCN are determined.

3. Experimental Results

In this section, we present the experimental results to evaluate the classification performance of the proposed method. This experiment classifies expert–novice levels using motion data during sports activities, i.e., soccer and diving. In addition, we quantitatively evaluated the classification performance and discussed the effectiveness of explaining the classification results by visualizing the attention nodes.

3.1. Experimental Settings

In this subsection, we explain the experimental setting. To evaluate the classification performance of our GCN-based method, we used the expert–novice soccer dataset [47] and the action quality assessment (AQA) dataset [48]. These datasets contain motion data on sports and their expert–novice levels. Specifically, the expert–novice soccer dataset contains motion data of eight participants for nine types of soccer plays (penalty kick (PK), free kick (FK), direct shot (DS), cross shot (CS), volley, long dribble, straight dribble, short

dribble, and juggling), four times each, for 288 samples. These motion data were obtained using the PERCEPTION NEURON PRO (<https://neuronmocap.com>), which was used to capture whole-body motion [49,50]. The nine types of soccer plays in this dataset are illustrated in Figure 5. In this dataset, the number of motion data frames differs based on the participant and the specific play. Note that the proposed method requires the number of frames in the input motion data to be identical. Therefore, to unify the temporal duration of all motion data, downsampling was performed by sampling the data at regular intervals to match the shortest motion data. In addition, each soccer play was classified according to a four-tiered expert–novice level, predetermined by individuals with more than five years of soccer experience. The AQA dataset consists of videos of athletes from seven sports (e.g., diving and 10 m platform) taken during the summer and winter Olympics. With this dataset, the experiment used motion data extracted from videos via MediaPipe 2 (<https://google.github.io/mediapipe/>, accessed on 23 February 2024). Because of the challenges of capturing motion data from the complex movements and changing camera angles present in many of the videos in the AQA dataset, only the “10 m platform single dive” data were used. The 10 m platform single dive in the AQA dataset is illustrated in Figure 6. The AQA dataset shows variability in motion data acquisition time across athletes and actions, which is similar to the expert–novice dataset. Therefore, to unify the temporal duration of all motion data, downsampling was performed in the same manner as in the expert–novice dataset experiment. The obtained motion data comprised 367 samples, of which 321 samples were used as training data and the remaining samples as test data. Each sample was given a score between 21.60 and 102.60 points. In the experiment, samples were categorized into four expertise levels, ranging from novice to expert, on the basis of the quartiles derived from their scores. To evaluate classification performance, we used the mean absolute error (MAE) and accuracy, which are defined as follows:

$$\text{MAE} = \frac{1}{K} \sum_{k=1}^K |g_k - r_k|, \quad (14)$$

$$\text{Accuracy} = \frac{\text{Number of correctly classified samples}}{\text{Number of all samples}}. \quad (15)$$

In Equation (14), g_k and r_k ($k = 1, 2, \dots, K$; K representing the number of test samples) denote the ground truth and classification results for the k -th test sample, respectively. In the MAE equation, $|g_k - r_k|$ denotes the difference between the actual expert–novice levels and the classification results. Therefore, the MAE calculated from Equation (14) indicates the extent to which the classification results deviate from the actual expert–novice level. A lower MAE indicates smaller classification errors, whereas a higher accuracy indicates a larger number of samples in which the classification results match the actual expert–novice level.

To evaluate the classification performance of the proposed method (PM), it was compared with the following eight comparative methods: the ST-GCN [29], ST-GCN with the proposed loss function $\mathcal{L}_{\text{total}}$ (ST-GCN w/ $\mathcal{L}_{\text{total}}$), STA-GCN [33], STA-GCN with $\mathcal{L}_{\text{total}}$ (STA-GCN w/ $\mathcal{L}_{\text{total}}$), ConfSTA-GCN [41], ConfSTA-GCN with $\mathcal{L}_{\text{total}}$ (ConfSTA-GCN w/ $\mathcal{L}_{\text{total}}$), spatiotemporal graph ConvNeXt (TSGCNeXt) [51], and proposed method without $\mathcal{L}_{\text{total}}$ (PM w/o $\mathcal{L}_{\text{total}}$). The ST-GCN, which is capable of incorporating spatial and temporal information, is GCN-based. As a basic method in GCN-based classification that considers spatiotemporal information, the ST-GCN was used as a comparative method. The STA-GCN is a GCN-based classification method that introduces the conventional attention mechanism. The ConfSTA-GCN verifies the effectiveness of the computation of confidence measures in the PM. The ST-GCN w/ $\mathcal{L}_{\text{total}}$, STA-GCN w/ $\mathcal{L}_{\text{total}}$, ConfSTA-GCN w/ $\mathcal{L}_{\text{total}}$, and PM w/o $\mathcal{L}_{\text{total}}$ were used to verify the effectiveness of our loss function $\mathcal{L}_{\text{total}}$. TSGCNeXt is a state-of-the-art method for GCN-based classification using motion data.

In this experiment, the number of joints in the expert–novice soccer and AQA datasets is 22 and 33, respectively. For the proposed and comparative methods, the learning rate

and the batch size were set to 0.01 and 64, respectively, and this experiment used stochastic gradient descent [29] as the optimization approach. In addition, the kernel size of T-GC was set to nine, consistent with the conditions of [29,39].

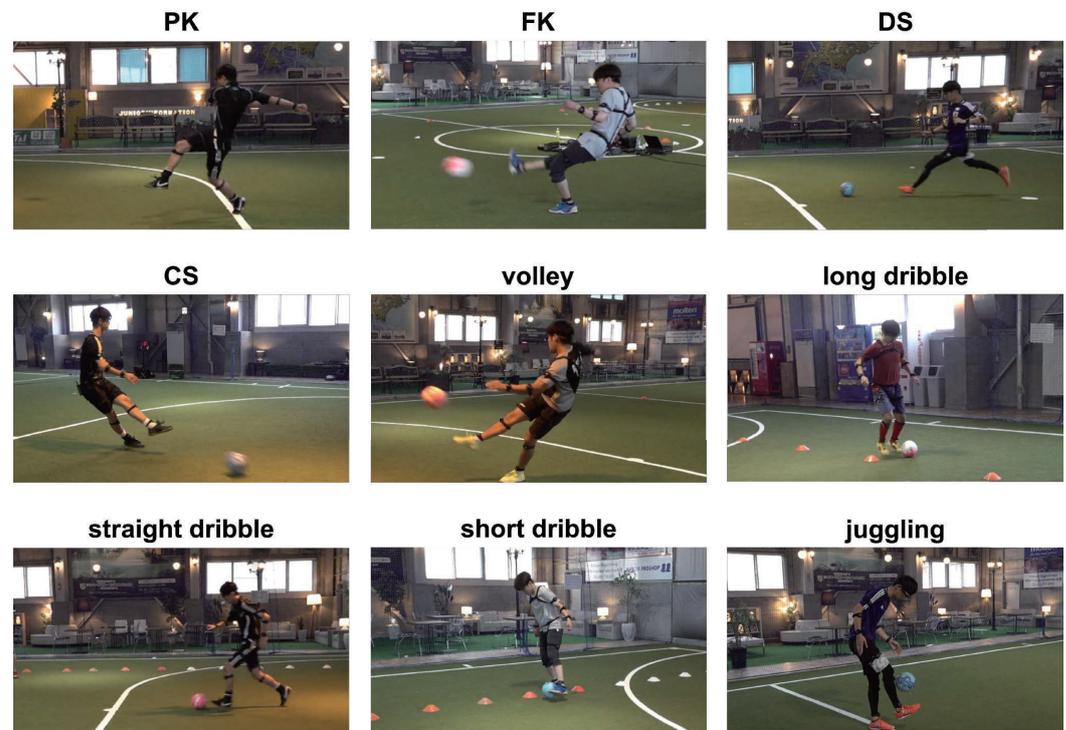


Figure 5. Nine types of soccer plays included in the expert–novice soccer dataset.



Figure 6. 10 m platform single dive included in the AQA dataset.

3.2. Evaluation of Expert–Novice Level Classification Performance

This subsection shows the performance of expert–novice level classification via the proposed and comparative methods. Tables 1 and 2 show the MAE and accuracy of the expert–novice level classification results obtained using the proposed and comparative methods for the expert–novice soccer dataset. Furthermore, the MAE and accuracy of the classification results for the AQA dataset are presented in Table 3. From these performance indices, the PM outperforms all comparative methods, demonstrating its effectiveness. Because the PM outperforms the ST-GCN and ST-GCN w/ \mathcal{L}_{total} , we can conclude that it can classify with higher accuracy than the ST-GCN when an attention mechanism is introduced. Furthermore, because our method outperforms the STA-GCN and STA-GCN w/ \mathcal{L}_{total} , accurate classification becomes feasible using the confidence-aware attention mechanism. By comparing the classification results of the PM, ConfSTA-GCN, and ConfSTA-GCN w/ \mathcal{L}_{total} , we can verify the effectiveness of the node-level attention mechanism that can control the impact of each attention node. Because the proposed method outperforms the PM w/o \mathcal{L}_{total} , we can confirm the effectiveness of the expert–novice level classification using the loss function \mathcal{L}_{total} . Finally, by comparing the classification results of the PM and TGCNext, we confirm that the proposed method outperforms the state-of-the-art method for GCN-based classification using motion data. These results confirm that the PM allows

for accurate classification of expert–novice levels by employing the confidence-aware node-level attention mechanism and the loss function \mathcal{L}_{total} .

Table 1. MAE (\downarrow) of expert–novice level classification results obtained using the proposed and comparative methods for the expert–novice soccer dataset.

	ST-GCN [29]	ST-GCN w/ \mathcal{L}_{total}	STA-GCN [39]	STA-GCN w/ \mathcal{L}_{total}	ConfSTA-GCN [41]	ConfSTA-GCN w/ \mathcal{L}_{total}	TGC NeXt [51]	PM w/o \mathcal{L}_{total}	PM
PK	0.844	0.375	0.594	0.313	0.281	0.344	0.563	0.250	0.188
FK	0.656	0.594	0.156	0.281	0.313	0.188	0.500	0.125	0.0625
DS	0.500	0.656	0.563	0.250	0.125	0.0938	1.16	0.0938	0.0938
CS	0.594	0.594	0.688	0.438	0.594	0.188	0.375	0.344	0.313
volley	0.594	0.688	0.250	0.219	0.250	0.125	0.656	0.188	0.156
long dribble	0.406	0.656	0.500	0.375	0.344	0.438	0.688	0.0625	0.0625
straight dribble	0.688	0.531	0.281	0.156	0.313	0.313	0.594	0.125	0.0938
short dribble	0.719	0.750	0.219	0.188	0.281	0.281	1.22	0.125	0.0313
juggling	0.531	0.438	0.563	0.375	0.531	0.406	0.906	0.344	0.188
Average	0.615	0.587	0.424	0.288	0.337	0.263	0.740	0.278	0.132

Table 2. Accuracy (\uparrow) of expert–novice level classification results obtained using the proposed and comparative methods on expert–novice soccer dataset.

	ST-GCN [29]	ST-GCN w/ \mathcal{L}_{total}	STA-GCN [39]	STA-GCN w/ \mathcal{L}_{total}	ConfSTA-GCN [41]	ConfSTA-GCN w/ \mathcal{L}_{total}	TGC NeXt [51]	PM w/o \mathcal{L}_{total}	PM
PK	0.469	0.719	0.656	0.813	0.750	0.750	0.563	0.750	0.813
FK	0.500	0.406	0.844	0.719	0.688	0.813	0.563	0.875	0.938
DS	0.594	0.500	0.781	0.813	0.875	0.906	0.344	0.938	0.906
CS	0.594	0.563	0.656	0.563	0.625	0.813	0.625	0.719	0.719
volley	0.688	0.688	0.750	0.781	0.813	0.938	0.500	0.813	0.844
long dribble	0.656	0.500	0.718	0.750	0.781	0.750	0.469	0.938	0.938
straight dribble	0.594	0.656	0.813	0.875	0.781	0.781	0.500	0.875	0.906
short dribble	0.656	0.563	0.875	0.875	0.844	0.813	0.375	0.875	0.969
juggling	0.531	0.563	0.688	0.656	0.594	0.688	0.406	0.719	0.813
Average	0.587	0.573	0.753	0.760	0.750	0.778	0.483	0.833	0.872

Table 3. MAE and accuracy of expert–novice level classification results obtained using the proposed and comparative methods for the AQA dataset.

	MAE (\downarrow)	Accuracy (\uparrow)
ST-GCN [29]	1.17	0.348
ST-GCN w/ \mathcal{L}_{total}	1.13	0.348
STA-GCN [39]	1.00	0.348
STA-GCN w/ \mathcal{L}_{total}	0.935	0.391
ConfSTA-GCN [41]	0.913	0.370
ConfSTA-GCN w/ \mathcal{L}_{total}	0.891	0.413
TGCNeXt [51]	1.09	0.391
PM w/o \mathcal{L}_{total}	0.870	0.478
PM	0.696	0.587

In addition, the confusion matrices for the PM, PM w/o \mathcal{L}_{total} , and ConfSTA-GCN w/ \mathcal{L}_{total} are shown in Figures 7 and 8. These results demonstrate that the proposed method is capable of accurately classifying the expert–novice levels compared with the PM w/o \mathcal{L}_{total} and ConfSTA-GCN w/ \mathcal{L}_{total} and that can classify the expert–novice levels closely to the ground truth.

Examples of the classification results of the PM and ConfSTA-GCN for the expert–novice soccer and AQA datasets are shown in Figures 9 and 10, respectively. These results confirm that by considering the ordinality between the expert and novice levels, we can accurately classify the expert–novice levels and enable classifications that are close to the ground truth. Consequently, in GCN-based classification, we verify the effectiveness of the confidence-aware node-level attention mechanism and the importance of considering the ordinality of the expert–novice levels.

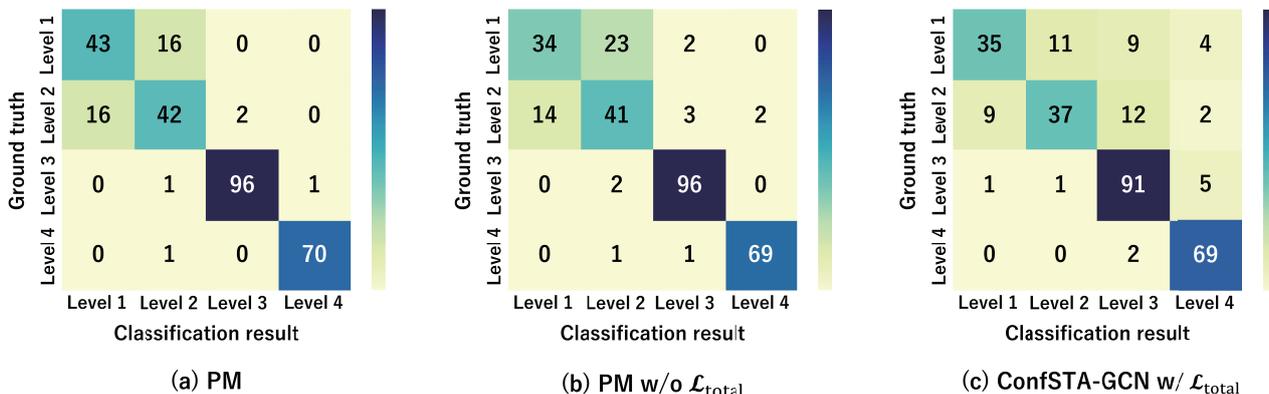


Figure 7. Confusion matrices of expert–novice level classification results obtained using the PM, PM w/o \mathcal{L}_{total} , and ConfSTA-GCN w/ \mathcal{L}_{total} for the expert–novice soccer dataset.

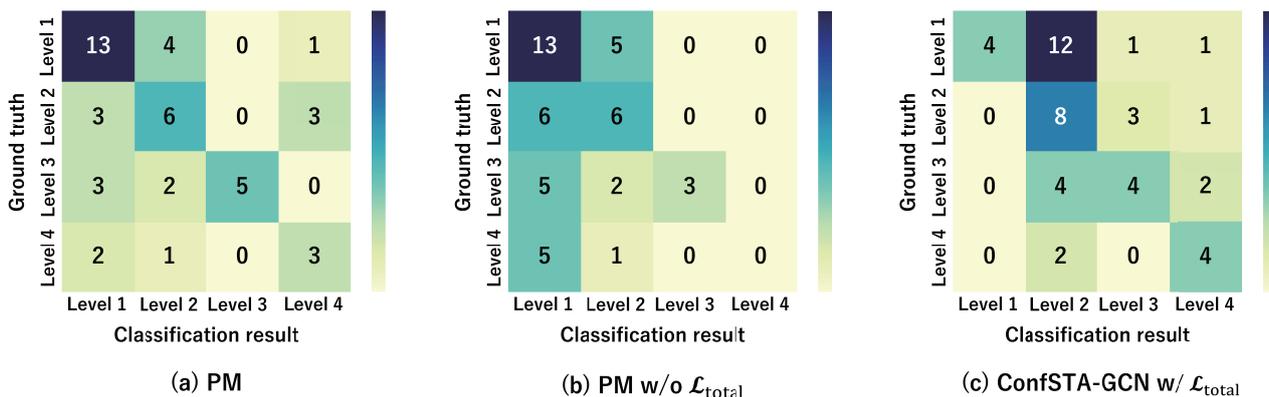


Figure 8. Confusion matrices of expert–novice level classification results obtained using the PM, PM w/o \mathcal{L}_{total} , and ConfSTA-GCN w/ \mathcal{L}_{total} for the AQA dataset.

This evaluation of expert–novice level classification performance demonstrates an improvement in classification performance, attributable to the contributions of this study, which include enhancements to the existing approach (ConfSTA-GCN) and the construction of a classification model that considers the ordinality between classes. Specifically, the effectiveness of improvements to the existing approach was verified by comparing the PM and ConfSTA-GCN w/ \mathcal{L}_{total} . Furthermore, the efficacy of the classification model that considers the ordinality between classes was confirmed through the comparison of the PM and PM w/o \mathcal{L}_{total} . Consequently, this study achieves the research objectives of enhancing the existing approach and improving the classification performance of expert–novice level classification through a model that accounts for the ordinal relationships between classes.



Figure 9. Examples of expert–novice level classification results obtained using the PM and previous method for the expert–novice soccer dataset.

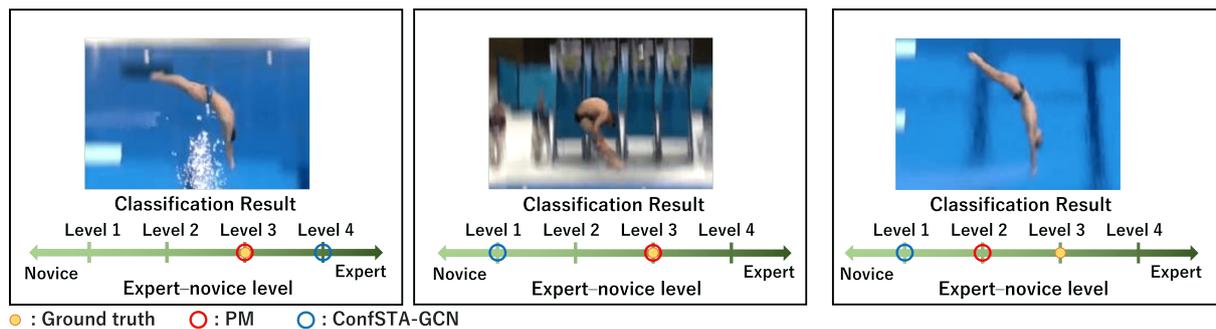


Figure 10. Examples of expert–novice level classification results obtained using the PM and previous method for the AQA dataset.

3.3. Visualization Results of Attention Nodes

This subsection shows the visualization results of the attention nodes and discusses the effectiveness of the PM. Figures 11–13 show examples of the visualization of the attention nodes for the PK, FK, and DS categories in the expert–novice soccer dataset. The frames visualized were selected as the frames with the largest standard deviation between attention nodes. Furthermore, Figure 14 shows an example of the visualization of the attention nodes in the AQA dataset regarding the frames when the standard deviation of the attention nodes is maximum. Figures 11–13 show that participants with lower expert–novice levels are confirmed to shoot using only their legs.

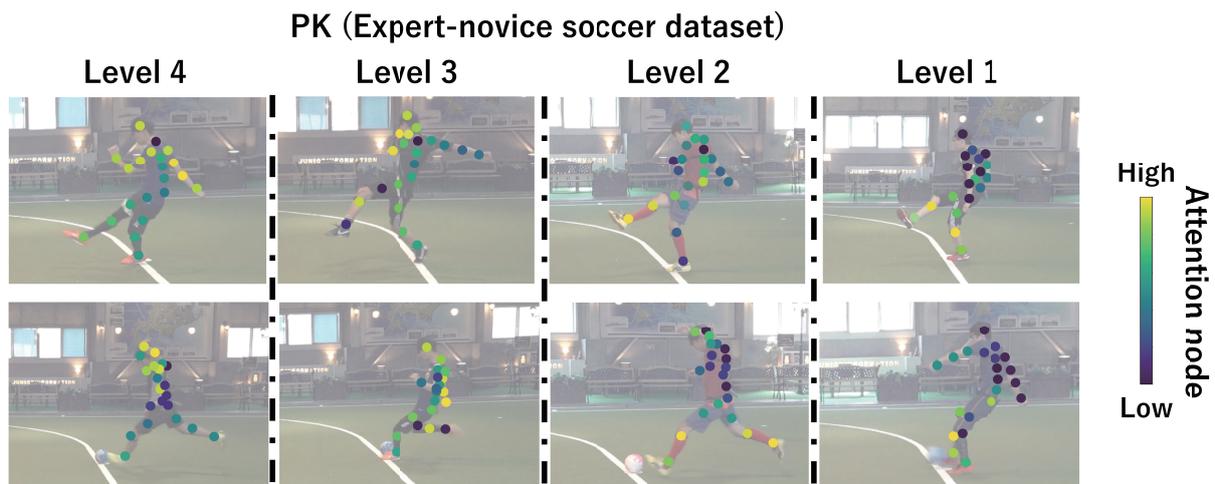


Figure 11. Examples of the visualization of attention nodes for penalty kick in the expert–novice soccer dataset using the PM.

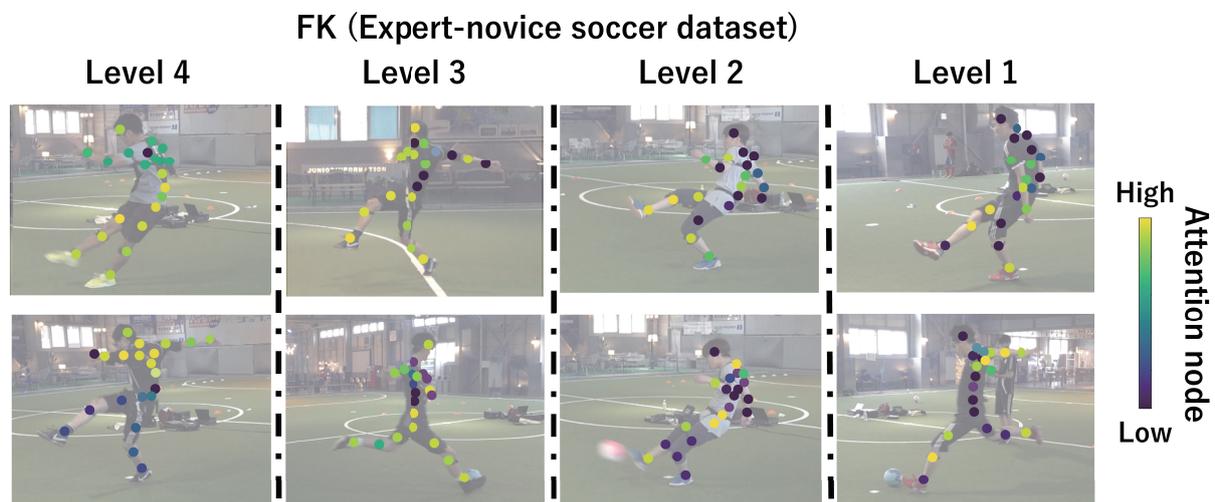


Figure 12. Examples of the visualization of attention nodes for free kick in the expert–novice soccer dataset using the PM.

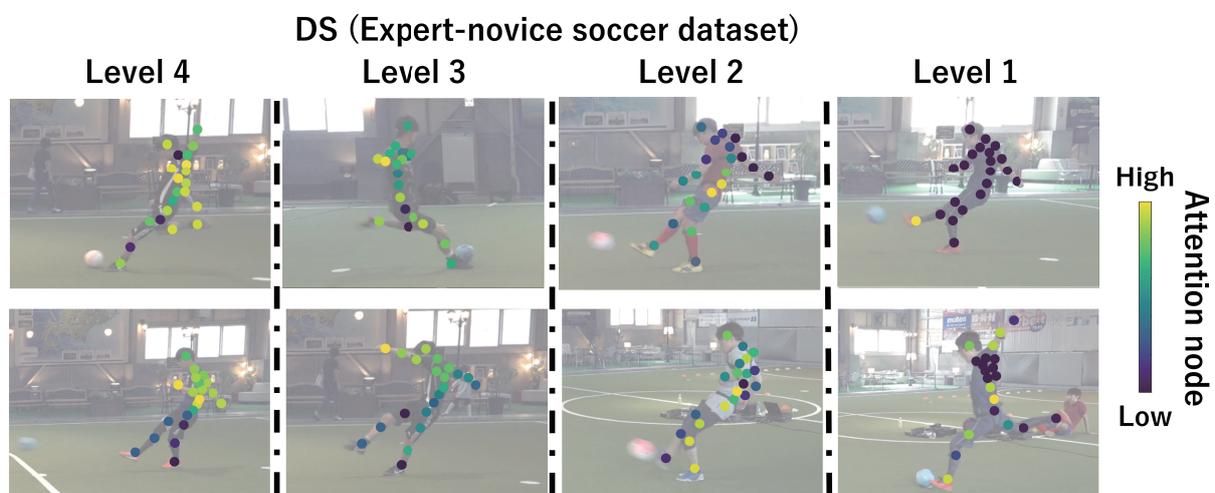


Figure 13. Examples of the visualization of attention nodes for direct shot in the expert–novice soccer dataset using the PM.

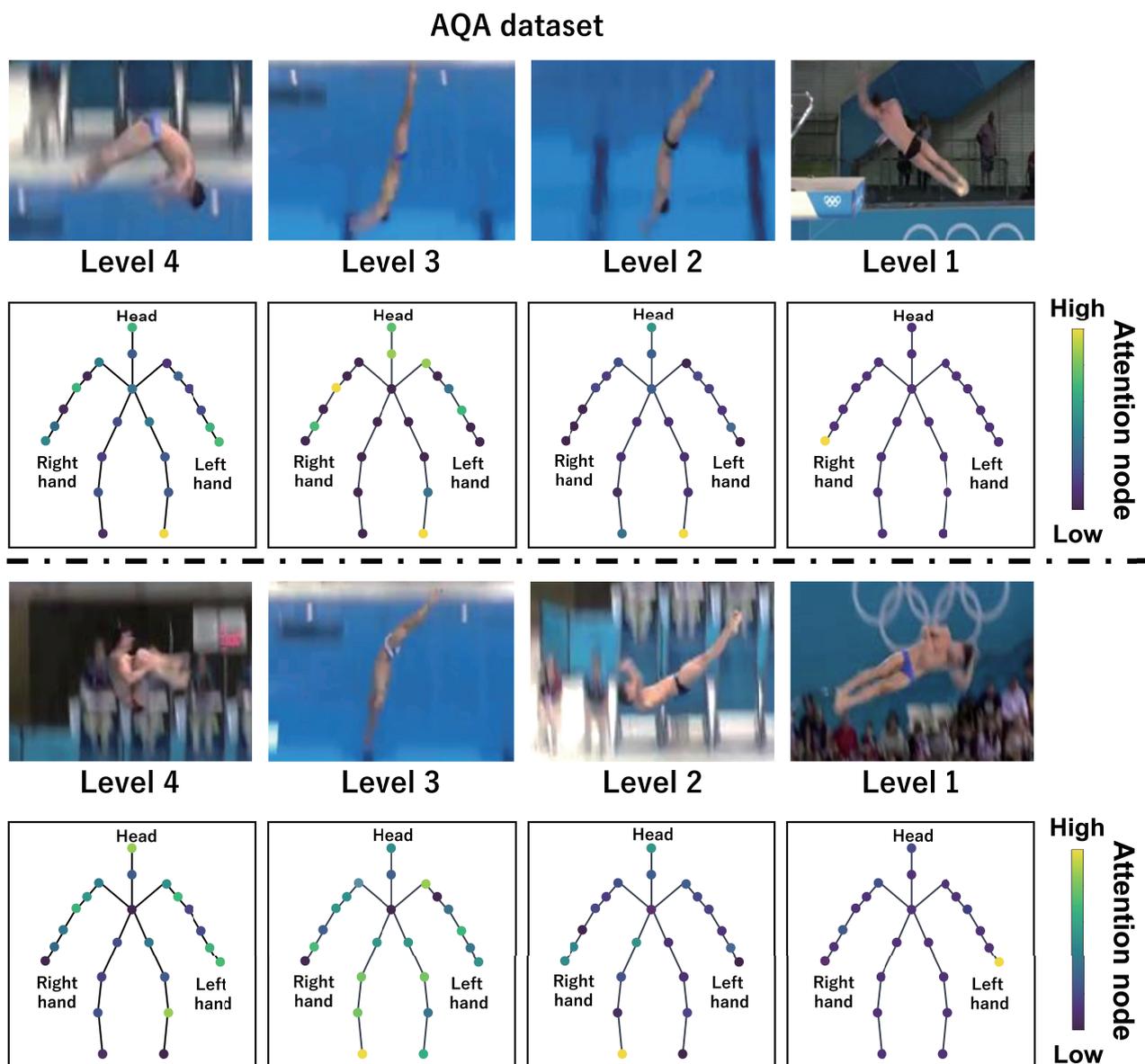


Figure 14. Examples of the visualization of attention nodes for the AQA dataset using the PM.

Specifically, the visualization results indicate that participants with lower expert–novice levels have higher values in the nodes associated with the lower body. Conversely, participants with higher expert–novice levels are observed to effectively use their upper body when shooting. Figure 14 demonstrates that expert levels of expertise have higher values of attention nodes in the head and shoulder.

Figures 15–18 focus on depicting the average values of attention nodes across all frames for each sample, confirming an overview of the trends in the whole sample. A comparison between attention nodes in Figures 11–14 and averaged attention nodes reveal that in PK, FK, and DS categories, participants with lower expert–novice levels confirm higher values in nodes associated with the lower body. Conversely, it can be confirmed that participants with higher expert–novice levels exhibit higher values in nodes related to the upper body. This result is consistent with the results confirmed by the visualization results in frames with the highest standard deviation across both datasets. These results confirm that the visualization outcomes of attention nodes in this experiment are independent of the frame.

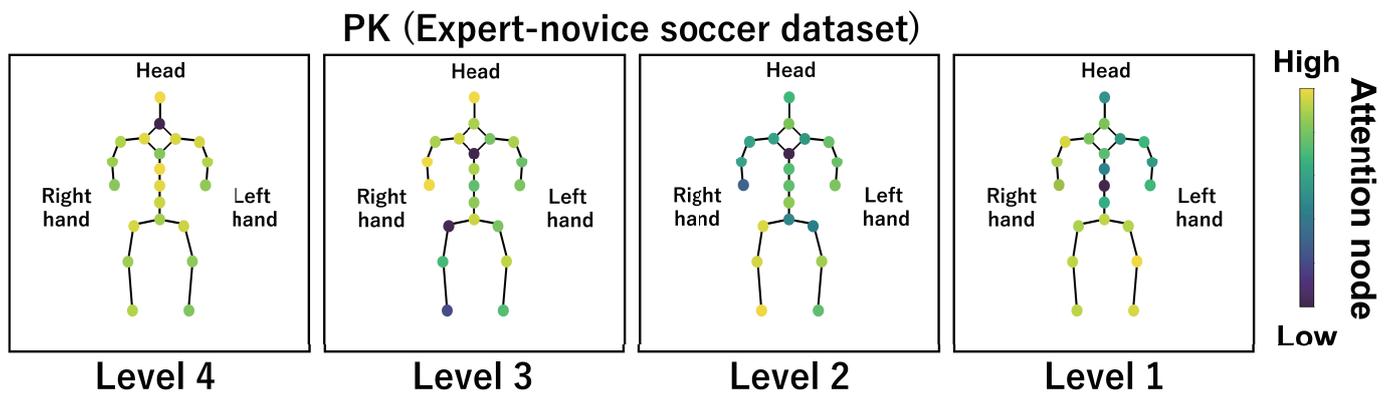


Figure 15. Examples of the visualization of attention nodes for the PK (expert–novice soccer dataset) using the PM. The attention nodes are averaged across all frames for each sample.

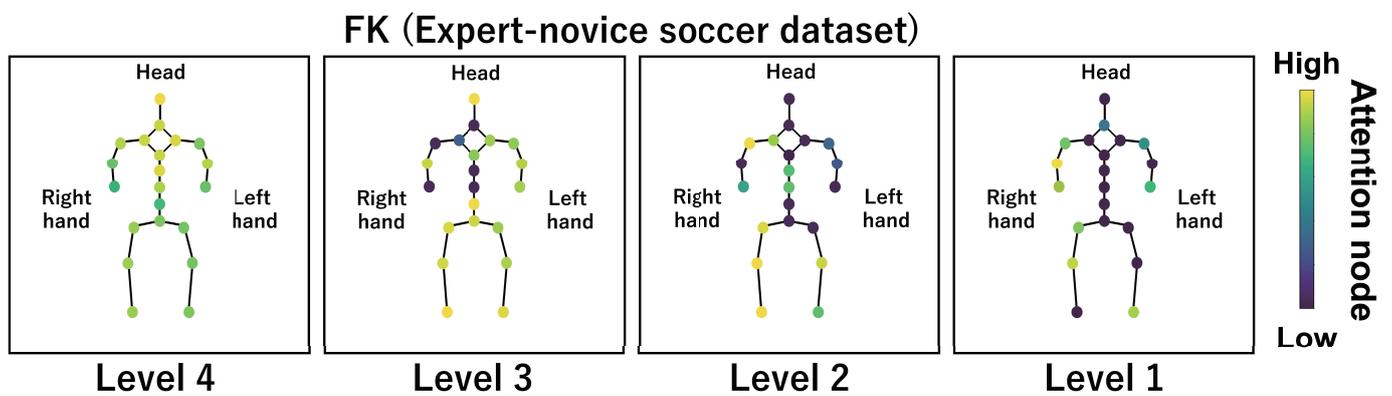


Figure 16. Examples of the visualization of attention nodes for the FK (expert–novice soccer dataset) using the PM. The attention nodes are averaged across all frames for each sample.

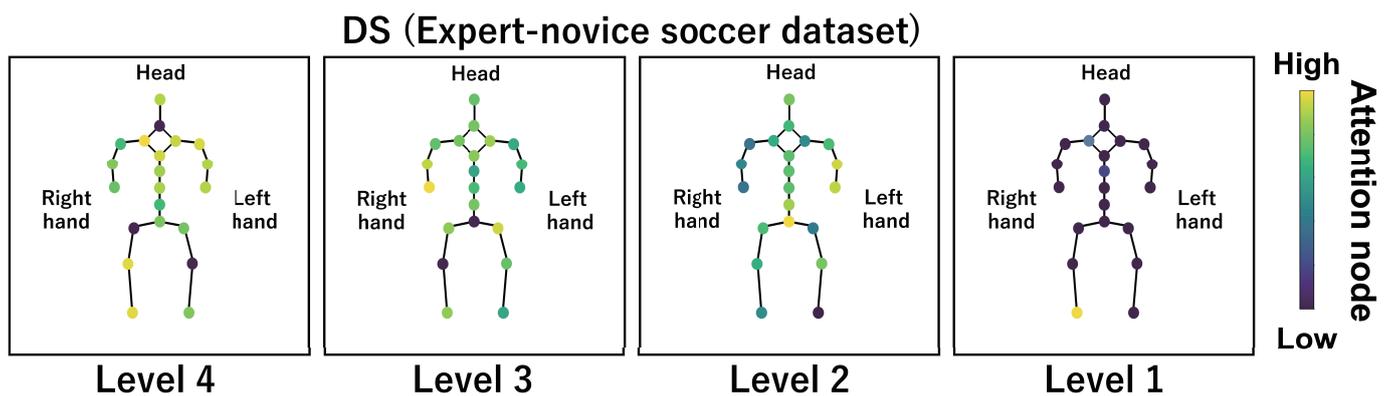


Figure 17. Examples of the visualization of attention nodes for the DS (expert–novice soccer dataset) using the PM. The attention nodes are averaged across all frames for each sample.

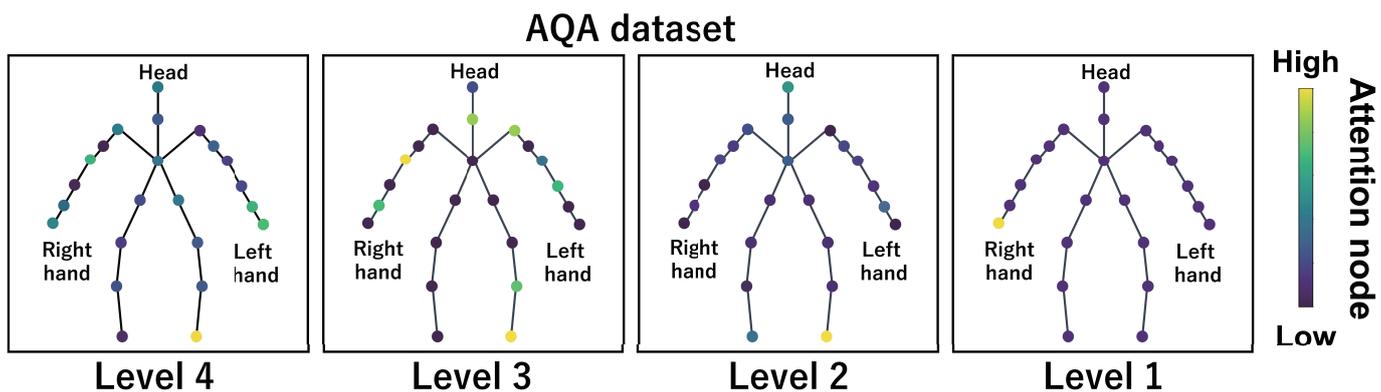


Figure 18. Examples of the visualization of attention nodes in the AQA dataset using the PM. The attention nodes are averaged across all frames for each sample.

Figures 19–22 depict the visualization of averaged attention nodes across all frames for each action at expert–novice levels, allowing for a comparison of how attention nodes between expert–novice levels. Despite the varying number of participants in each level, the values of attention nodes for the PK, FK, DS, and the AQA dataset exhibit similar trends to those confirmed in Figures 11–14 and Figures 15–18. In the expert–novice level soccer dataset, participants with higher expert–novice levels show elevated values in nodes associated with the upper body. In the AQA dataset, samples from expert participants show high values in nodes related to the upper body and foot. These results are corroborated by quantitative results from each dataset, confirming that the visualized attention nodes contribute to the classification process.

These results suggest that the visualization approach proposed consistently captures the importance of soccer-specific movements across different frame counts and samples, highlighting their relevance to varying expert–novice levels. Moreover, it successfully identifies the significance of movements specific to diving and swimming, demonstrating their relevance to expert–novice levels.

PK (Expert-novice soccer dataset)

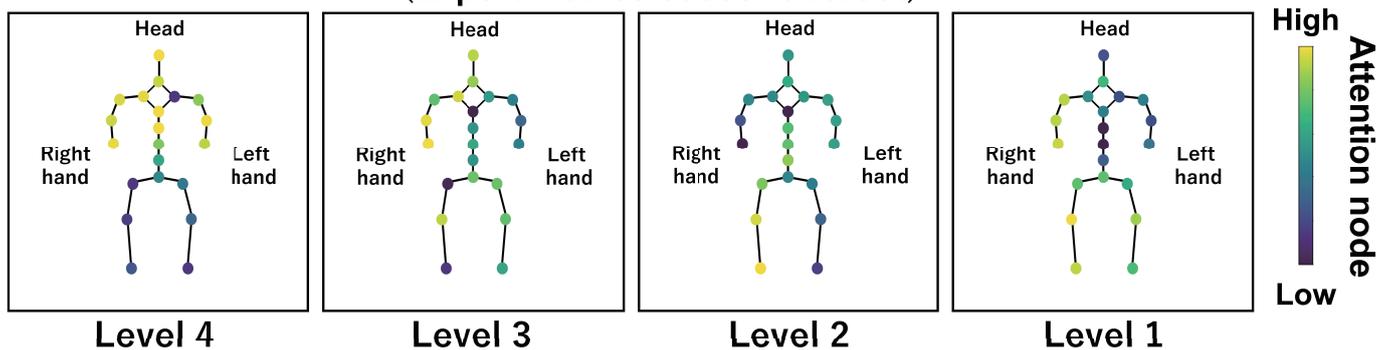


Figure 19. Examples of the visualization of attention nodes for the PK (expert–novice soccer dataset) using the PM. The attention nodes are averaged across all frames for each action by expert–novice levels.

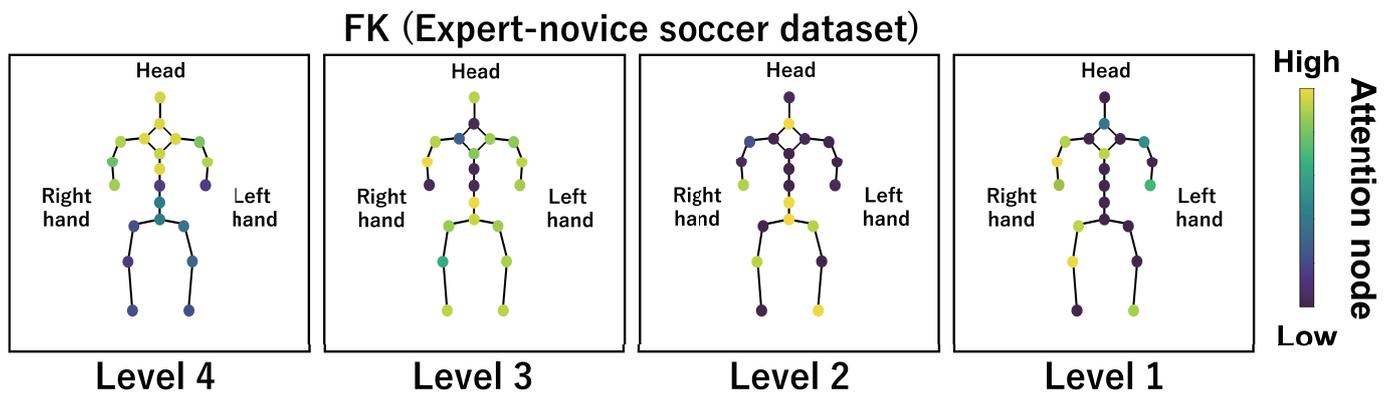


Figure 20. Examples of the visualization of attention nodes for the FK (expert–novice soccer dataset) using the PM. The attention nodes are averaged across all frames for each action by expert–novice levels.

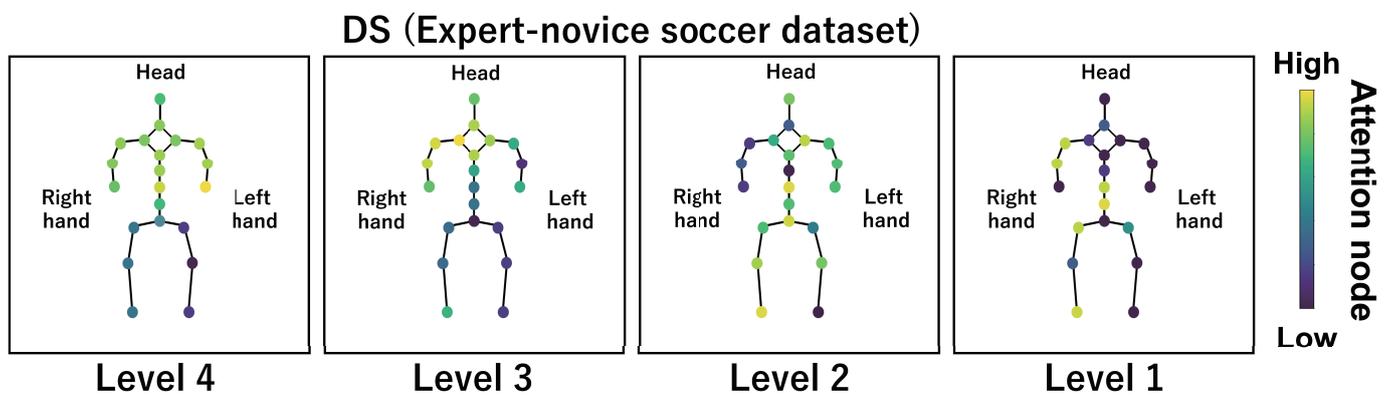


Figure 21. Examples of the visualization of attention nodes for the DS (expert–novice soccer dataset) using the PM. The attention nodes are averaged across all frames for each action by expert–novice levels.

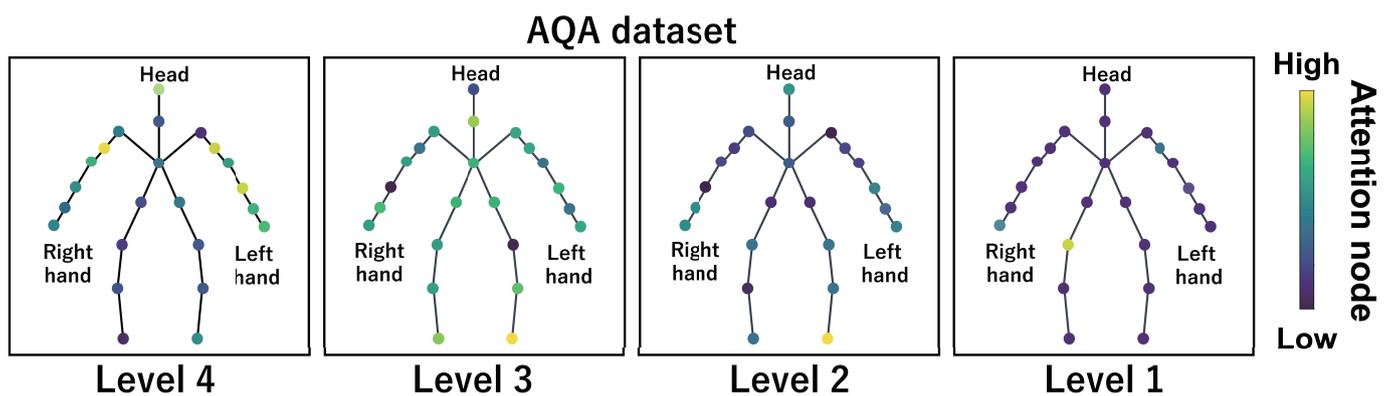


Figure 22. Examples of the visualization of attention nodes for AQA dataset using the PM. The attention nodes are averaged across all frames for each action by expert–novice levels.

4. Conclusions

In this study, we proposed a method for classifying expert–novice levels using motion data via a GCN that introduces a confidence-aware node-level attention mechanism. The PM effectively solves the problem of using unimportant features in existing methods. In particular, the PM calculates the probability of belonging to an actual expert–novice level when specific attention nodes are excluded, and the calculated probabilities are regarded as

a confidence measure. Consequently, our method can compare the attention value of each node based on the confidence of the classification. This solves the attention mechanism problem and enables accurate classification. Furthermore, because the expert–novice levels have ordinalities, we construct a classification model that considers ordinalities, thereby improving classification performance.

Because of the constraint in the PM, the number of frames in the input motion data must be uniform, and downsampling is performed. However, there is a problem with downsampling, which can result in the lack of important frames for accurate classification. Therefore, constructing a model capable of handling motion data with different time lengths remains a challenge for future work.

Author Contributions: Conceptualization, T.S., N.S., T.O., S.A. and M.H.; methodology, T.S., N.S. and T.O.; software, validation, and data curation, T.S.; writing—original draft preparation, T.S.; writing—review and editing, N.S., T.O. and M.H.; visualization, T.S.; funding acquisition, T.O. and M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partly supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant Number JP21H03456).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. The public datasets used in our experiment are available at <https://github.com/LMD-datasets/Expert-NoviceSoccerDataset> (accessed on 5 March 2024) and <http://rtis.oit.unlv.edu/datasets.html> (accessed on 5 March 2024).

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

References

1. Daley, B.J. Novice to expert: An exploration of how professionals learn. *Adult Educ. Q.* **1999**, *49*, 133–147. [[CrossRef](#)]
2. Meteier, Q.; Capallera, M. Classification of drivers' workload using physiological signals in conditional automation. *Front. Psychol.* **2021**, *12*, 596038. [[CrossRef](#)]
3. Toy, S.; Ozsoy, S.; Shafiei, S.; Antonenko, P.; Schwengel, D. Using electroencephalography to explore neurocognitive correlates of procedural proficiency: A pilot study to compare experts and novices during simulated endotracheal intubation. *Brain Cogn.* **2023**, *165*, 105938. [[CrossRef](#)] [[PubMed](#)]
4. Capogna, E.; Salvi, F.; Delvino, L.; Di Giacinto, A.; Velardo, M. Novice and expert anesthesiologists' eye-tracking metrics during simulated epidural block: A preliminary, brief observational report. *Local Reg. Anesth.* **2020**, *13*, 105–109. [[CrossRef](#)] [[PubMed](#)]
5. Hafeez, T.; Umar Saeed, S.M.; Arsalan, A.; Anwar, S.M.; Ashraf, M.U.; Alsubhi, K. EEG in game user analysis: A framework for expertise classification during gameplay. *PLoS ONE* **2021**, *16*, e0246913. [[CrossRef](#)] [[PubMed](#)]
6. Ihara, A.S.; Matsumoto, A.; Ojima, S.; Katayama, J.; Nakamura, K.; Yokota, Y.; Watanabe, H.; Naruse, Y. Prediction of second language proficiency based on electroencephalographic signals measured while listening to natural speech. *Front. Hum. Neurosci.* **2021**, *15*, 665809. [[CrossRef](#)]
7. Villagrán Gutiérrez, I.A.; Moënné-Loccoz, C.; Aguilera Sviragol, V.I.; Garcia, V.; Reyes, J.T.; Rodriguez, S.; Miranda Mendoza, C.; Altermatt, F.; Fuentes López, E.; Delgado Bravo, M.A.; et al. Biomechanical analysis of expert anesthesiologists and novice residents performing a simulated central venous access procedure. *PLoS ONE* **2021**, *16*, e0250941. [[CrossRef](#)] [[PubMed](#)]
8. Laverde, R.; Rueda, C.; Amado, L.; Rojas, D.; Altuve, M. Artificial neural network for laparoscopic skills classification using motion signals from apple watch. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Honolulu, HI, USA, 18–21 July 2018; pp. 5434–5437.
9. Pan, J.H.; Gao, J.; Zheng, W.S. Action assessment by joint relation graphs. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27–28 October 2019; pp. 6331–6340.
10. Xue, H.; Batalden, B.M.; Sharma, P.; Johansen, J.A.; Prasad, D.K. Biosignal-based driving skill classification using machine learning: A case study of maritime navigation. *Appl. Sci.* **2021**, *11*, 9765. [[CrossRef](#)]
11. Baig, M.Z.; Kavakli, M. Classification of user competency levels using EEG and convolutional neural network in 3D modelling application. *Expert Syst. Appl.* **2020**, *146*, 113202. [[CrossRef](#)]
12. Hosp, B.; Yin, M.S.; Haddawy, P.; Watcharopas, R.; Sa-Ngasoongsong, P.; Kasneci, E. Differentiating surgeons' expertise solely by eye movement features. In Proceedings of the International Conference on Multimodal Interaction, Montreal, QC, Canada, 18–22 October 2021; pp. 371–375.

13. Ahmidi, N.; Ishii, M.; Fichtinger, G.; Gallia, G.L.; Hager, G.D. An objective and automated method for assessing surgical skill in endoscopic sinus surgery using eye-tracking and tool-motion data. *Int. Forum Allergy Rhinol.* **2012**, *2*, 507–515. [[CrossRef](#)]
14. Berges, A.J.; Vedula, S.S.; Chara, A.; Hager, G.D.; Ishii, M.; Malpani, A. Eye tracking and motion data predict endoscopic sinus surgery skill. *Laryngoscope* **2023**, *133*, 500–505. [[CrossRef](#)]
15. Seong, M.; Kim, G.; Yeo, D.; Kang, Y.; Yang, H.; DelPreto, J.; Matusik, W.; Rus, D.; Kim, S. MultiSenseBadminton: Wearable sensor-based biomechanical dataset for evaluation of badminton performance. *Sci. Data* **2024**, *11*, 343. [[CrossRef](#)]
16. Soangra, R.; Sivakumar, R.; Anirudh, E.; Reddy, Y.S.V.; John, E.B. Evaluation of surgical skill using machine learning with optimal wearable sensor locations. *PLoS ONE* **2022**, *17*, e0267936. [[CrossRef](#)]
17. Shafiei, S.B.; Shadpour, S.; Mohler, J.L.; Sasangohar, F.; Gutierrez, C.; Seilanian Toussi, M.; Shafqat, A. Surgical skill level classification model development using EEG and eye-gaze data and machine learning algorithms. *J. Robot. Surg.* **2023**, *17*, 2963–2971. [[CrossRef](#)]
18. Dials, J.; Demirel, D.; Sanchez-Arias, R.; Halic, T.; Kruger, U.; De, S.; Gromski, M.A. Skill-level classification and performance evaluation for endoscopic sleeve gastropasty. *Surg. Endosc.* **2023**, *37*, 4754–4765. [[CrossRef](#)]
19. Kuo, R.; Chen, H.J.; Kuo, Y.H. The development of an eye movement-based deep learning system for laparoscopic surgical skills assessment. *Sci. Rep.* **2022**, *12*, 11036. [[CrossRef](#)]
20. Guo, X.; Brown, E.; Chan, P.P.; Chan, R.H.; Cheung, R.T. Skill level classification in basketball free-throws using a single inertial sensor. *Appl. Sci.* **2023**, *13*, 5401. [[CrossRef](#)]
21. Weinstein, J.L.; El-Gabalawy, F.; Sarwar, A.; DeBacker, S.S.; Faintuch, S.; Berkowitz, S.J.; Bulman, J.C.; Palmer, M.R.; Matyal, R.; Mahmood, F.; et al. Analysis of Kinematic differences in hand motion between novice and experienced operators in IR: A pilot study. *J. Vasc. Interv. Radiol.* **2021**, *32*, 226–234. [[CrossRef](#)]
22. Laube, M.; Sopidis, G.; Anzengruber-Tanase, B.; Ferscha, A.; Haslgrübler, M. Analyzing arc welding techniques improves skill level assessment in industrial manufacturing processes. In Proceedings of the 16th International Conference on Pervasive Technologies Related to Assistive Environments, Corfu, Greece, 5–7 July 2023; pp. 177–186.
23. Tao, L.; Elhamifar, E.; Khudanpur, S.; Hager, G.D.; Vidal, R. Sparse hidden markov models for surgical gesture classification and skill evaluation. In Proceedings of the Information Processing in Computer-Assisted Interventions: Third International Conference, IPCAI 2012, Pisa, Italy, 27 June 2012; pp. 167–177.
24. Uemura, M.; Tomikawa, M.; Miao, T.; Souzaki, R.; Ieiri, S.; Akahoshi, T.; Lefor, A.K.; Hashizume, M. Feasibility of an AI-based measure of the hand motions of expert and novice surgeons. *Comput. Math. Methods Med.* **2018**, *2018*, 9873273. [[CrossRef](#)]
25. Ross, G.B.; Dowling, B.; Troje, N.F.; Fischer, S.L.; Graham, R.B. Classifying elite from novice athletes using simulated wearable sensor data. *Front. Bioeng. Biotechnol.* **2020**, *8*, 814. [[CrossRef](#)]
26. D’Amato, V.; Volta, E.; Oneto, L.; Volpe, G.; Camurri, A.; Anguita, D. Understanding violin players’ skill level based on motion capture: A data-driven perspective. *Cogn. Comput.* **2020**, *12*, 1356–1369. [[CrossRef](#)]
27. Nguyen, X.A.; Ljuhar, D.; Pacilli, M.; Nataraja, R.M.; Chauhan, S. Surgical skill levels: Classification and analysis using deep neural network model and motion signals. *Comput. Methods Programs Biomed.* **2019**, *177*, 1–8. [[CrossRef](#)]
28. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.
29. Yan, S.; Xiong, Y.; Lin, D. Spatial temporal graph convolutional networks for skeleton-based action recognition. In Proceedings of the Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32, pp. 3482–3489.
30. Si, C.; Jing, Y.; Wang, W.; Wang, L.; Tan, T. Skeleton-based action recognition with spatial reasoning and temporal stack learning. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 103–118.
31. Li, M.; Chen, S.; Chen, X.; Zhang, Y.; Wang, Y.; Tian, Q. Actional-structural graph convolutional networks for skeleton-based action recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3595–3603.
32. Si, C.; Chen, W.; Wang, W.; Wang, L.; Tan, T. An attention enhanced graph convolutional lstm network for skeleton-based action recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1227–1236.
33. Shi, L.; Zhang, Y.; Cheng, J.; Lu, H. Skeleton-based action recognition with directed graph neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7912–7921.
34. Shi, L.; Zhang, Y.; Cheng, J.; Lu, H. Two-stream adaptive graph convolutional networks for skeleton-based action recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 12026–12035.
35. Zhao, L.; Peng, X.; Tian, Y.; Kapadia, M.; Metaxas, D.N. Semantic graph convolutional networks for 3D human pose regression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3425–3435.
36. Cheng, K.; Zhang, Y.; Cao, C.; Shi, L.; Cheng, J.; Lu, H. Decoupling GCN with dropgraph module for skeleton-based action recognition. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 536–553.
37. Zhang, J.; Ye, G.; Tu, Z.; Qin, Y.; Qin, Q.; Zhang, J.; Liu, J. A spatial attentive and temporal dilated (SATD) GCN for skeleton-based action recognition. *Chin. Assoc. Artif. Intell. Trans. Intell. Technol.* **2022**, *7*, 46–55. [[CrossRef](#)]

38. Thakkar, K.; Narayanan, P. *Part-Based Graph Convolutional Network for Action Recognition*; British Machine Vision Association: Durham, UK, 2018.
39. Shiraki, K.; Hirakawa, T.; Yamashita, T.; Fujiyoshi, H. Spatial temporal attention graph convolutional networks with mechanics-stream for skeleton-based action recognition. In Proceedings of the Asian Conference on Computer Vision, Kyoto, Japan, 30 November–4 December 2020.
40. Mitsuhashi, M.; Fukui, H.; Sakashita, Y.; Ogata, T.; Hirakawa, T.; Yamashita, T.; Fujiyoshi, H. Embedding human knowledge into deep neural network via attention map. *arXiv* **2019**, arXiv:1905.03540.
41. Seino, T.; Saito, N.; Ogawa, T.; Asamizu, S.; Haseyama, M. Confidence-aware spatial temporal graph convolutional network for skeleton-based expert-novice level classification. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Seoul, Republic of Korea, 14–19 April 2024.
42. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
43. Glorot, X.; Bordes, A.; Bengio, Y. Deep sparse rectifier neural networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 11–13 April 2011; pp. 315–323.
44. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
45. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
46. Lin, M.; Chen, Q.; Yan, S. Network in network. *arXiv* **2013**, arXiv:1312.4400.
47. Akamatsu, Y.; Maeda, K.; Ogawa, T.; Haseyama, M. Classification of expert-novice level using eye tracking and motion data via conditional multimodal variational autoencoder. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Toronto, ON, Canada, 6–11 June 2021; pp. 1360–1364.
48. Parmar, P.; Morris, B. Action Quality Assessment Across Multiple Actions. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 7–11 January 2019; pp. 1468–1476.
49. Roth, E.; Möncks, M.; Bohné, T.; Pumplun, L. Context-aware cyber-physical assistance systems in industrial systems: A human activity recognition approach. In Proceedings of the IEEE International Conference on Human–Machine Systems, Rome, Italy, 7–9 September 2020; pp. 1–6.
50. Demircan, E. A pilot study on locomotion training via biomechanical models and a wearable haptic feedback system. *Robomech J.* **2020**, *7*, 19. [[CrossRef](#)]
51. Liu, D.; Chen, P.; Yao, M.; Lu, Y.; Cai, Z.; Tian, Y. TSGCNeXt: Dynamic-static multi-graph convolution for efficient skeleton-based action recognition with long-term learning potential. *arXiv* **2023**, arXiv:2304.11631.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.