Proceedings of ICSP2020

Hierarchical Multi-Classification for Sensor-based Badminton Activity Recognition

Ya Wang[†], Jinwen Ma[†*], Xiangchen Li[‡], Albert Zhong[$]
[†]*School of Mathematical Sciences and LMAM, Peking University, Beijing, China*
[‡]*China Institute of Sport Science*
[$]*Hangzhou Zhidong Sports Co. LTD*
wangyachn@pku.edu.cn, jwma@math.pku.edu.cn, lixiangchen@ciss.cn, zhongl1978@outlook.com

2020 15th IEEE International Conference on Signal Processing (ICSP) | 978-1-7281-4480-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICSP48669.2020.9320935

***Abstract*—Fast development of sensor technology makes sensor equipments more and more smart and wearable. It further boost the need of sensor-based human activity recognition. Due to the lack of large-scale labeled datasets in practical AI applications, it is important to utilize prior information of the categories in sensor-based human activity recognition. In this paper, we propose a Hierarchical Multi-Classificaion (HMC) framework for sensor-based badminton activity recognition with the help of the prior information of badminton activity categories. Specifically, the multi-class sensor-based badminton activity recognition task is performed in two steps: (1). Any input data for a badminton activity are classified into one of the major classes which are based on their characteristic features; (2). They are further classified into one of the specific categories of badminton activity as required. It is demonstrated by the experimental results on BSS-V2 dataset that our proposed method can get up to 83.9% badminton activity recognition accuracy which is 1.7% better than previous state-of-the-art models.**

***Keywords*—Hierarchical classification, Badminton Activity Recognition, Multi-Classification.**

I. INTRODUCTION

Human activity recognition (HAR) is a fundamental and practical task, which has attracted increasing research efforts in recent years due to its extremely broad applications in many areas, such as healthcare [1], gesture recognition [2] and smart environment [3]. According to the forms of data, it can be classified into two typical categories: *video-based HAR* [4]–[6] and *sensor-based HAR* [7]. Literally, video-based HAR aims to recognize the activities in videos, while sensor-based HAR focuses on the data taken from smart sensors such as accelerometers, gyroscopes, bluetooth, sound sensors and so on. Compared with videos, sensor data has two advantages: Firstly, it is much cleaner for its freedom from noisy background [7]. Secondly, in general, the storage of sensor data is much less than video if both of them are taken with the same frequency and time duration. Benefited from it, the computation cost of sensor-based HAR is usually much less than that of video-based HAR. Due to the advantages above, sensor-based HAR is quite applicable and efficient to be inserted into various equipments, such as wearable devices [2], [8] and sports equipments [7]. However, though the task of sensor-based HAR is emergent for artificial intelligence applying to our daily life, it suffers from two challenges waiting to be overcomed by researchers: (1) Due to the cost of human labeled data, the dataset of most of realistic applications are not sufficient to train a deep CNN with satisfactory generalization. (2) The relationship of categories which is evident for human knowledge cannot be captured by classifiers effectively, especially in small scale datasets.

First of all, different categories are not independent of each other, in other words, the categories cannot be naively seen as individuals. In the field of badminton sports, as for the correlation of "similarity", the relationship of "Forehand Net Kill" and "Backhand Net Kill" is more significant than that of "Forehand Net Kill" and "Midfield Forehand Smash". Single Deep Neural Networks (DNN) learn the relation of categories with large-scale datasets implicitly, however when it comes to small-scale datasets, maybe they will fail to capture the complex relationship of various categories. Secondly, most of single DNNs can be formed as a two-stage mapping: $y = F_1 \circ F_0(X)$, where $F_0 : X \mapsto f$ denotes the feature extraction process and $F_1 : f \mapsto y$ is the mapping from feature space to label space. Naturally, a question is raised, is it reasonable to take all samples of different categories equally and cast them into the same feature space? Taking "Forehand Net Kill", "Backhand Net Kill", "Midfield Forehand Smash" and "Midfield Backhand Smash" as an example, "Forehand Net Kill' and "Backhand Net Kill" are separable with each other in accelerometer, while it is easier to distinguish "Midfield Forehand Smash" and "Midfield Backhand Smash" with spatial attitude angles. Deep DNN deals with this problem with redundant learnable features. However, compared with shallow DNN which maybe cannot generate a good enough feature space to cover all needed discriminative feature, deep DNN is usually easier to overfit especially small-scale datasets. It is inspired that casting samples into different feature spaces can help to settle this problem.

As far as we know, there has been no effective method which can utilize the relationship of categories in the area of sensor-based HAR. However, in the other fields, some methods have been proposed to mine the label correlations to improve the recognition accuracy. For example, KSSNet [4] utilizes a graph convolution network to boost the performance of multi-label video recognition, probabilistic graph model [9], [10] and RNN [11] are used to capture dependencies among labels. However, with more label-relation parameters

* Corresponding author.

introduced, the training of their models need sufficient labeled samples. However, it is usually not available in many tasks.

In this paper, we try to establish a new method to capture the relationship of categories as well as casting samples to their own characteristics-based feature spaces. Actually, we propose a novel Hierarchical Multi-Classification (*HMC*) framework for this task. Our strategy is to separate the classification task into two sub-tasks: casting the input data into a feature space and using a classifier affiliated with the feature space to predict the specific category of the input date. Specifically, we firstly construct some feature spaces and each of them has a specific classifier as the feature encoder. Then a lightweight coarse-grained classifier is utilized to cast the input date into the right feature spaces. After that, a specific classifier encodes the feature of input data and sends it to the fine-grained classifier to make a prediction of the category. As the mapping of the output of coarse-grained classifier and the choice of feature spaces are specified by human prior, our approach does not need extra dataset to train the relationship of the categories.

II. Hierarchical Multi-Classification

We now propose a Hierarchical Multi-Classification (HMC) framework for leveraging the badminton activity recognition performance with human priors. As Table I shows, it releases a complex HAR task to two-stage hierarchical subtasks. Given the input sensor-based data, Task 1 is conducted to recognize its *main-classes*, which is affiliated with a specific feature space. Then, Task 2 is implemented for recognizing their exact activities (called *subclasses*). In detail, it chooses a subsequencial classifier according to the result of Task 1 and utilizes it to extract the deep CNN feature of the raw input sensor-based data, finally predicts the subclass of the input sample. The mapping from Task 1 to the classifiers of Task 2 is determined by human priors, which is not experit and easy to be implemented. Using this strategy, our proposed HMC framework can superimpose knowledge priors into various state-of-the-art CNN-based classifiers, such as our work AFEB-MobileNet [7] which is also the previous state-of-the-art in our collected BSS-V2 dataset.

Usually, conventional CNN-based classifiers use a single CNN to output the prediction of input data, however as we mentioned in Section I, it suffers from two drawbacks: depending on large-scale labeled datasets and lacking the capacity to capture the relationship of categories. In order to overcome these problems, our proposed HMC is designed as a two-stage classifier. Formally, it can be formulated as a composite function:

$$Z = f(g(X), X) \tag{1}$$

where $X \in \mathbf{R}^{N \times C}$ is the input sensor-based data. N is the number of records and C is the dimension of each record. g is a main-class classifier (M-classifier) whose output is the main-class of input data. It acts as a jobs allocator choosing a specific downstream classifier from a set of classifiers. f is a binary function built with prior knowledge. It takes the first argument to choose a sub-classifier, the second one as the input to form a fine-grained classification task and outputs the predicted class of X. We first segment the total classes into three groups, the classes in the same groups have similar characteristics, while the classes lying in different groups are more distinctly different from each other. This structure is guided by expert knowledge. After that, g is trained on the training set where all samples are relabeled with main-classes to make a coarse-grained classification of the input sensor-based data. Next, a specific subclass classifier (S-classifier) acting as CNN encoder for each main-class is trained with the samples of this main-class. Each S-classifier only takes the subclasses belonging to its corresponding main-class as labels. Therefore, it is much easier to be trained compared with considering all classes simultaneously.

Usually, as for some tasks which have complex label relationship, the accuracies of g and f are much higher than that of the single CNN version of classifier, the reasons are following:

- As a hierarchical multi-classification framework, both g and f have significantly less classes (even orders of magnitude less) than single classifier variant. So the classification difficulty of our HMC is alleviated significantly.
- The labels of g are distinctly different from each other in the expert knowledge, which is not available in the task of single classifier variant. It results that the classification task of main-classifiers and sub-classifiers is much easier.
- The framework of our proposed HMC employs a M-classifier to map the input data into different feature spaces, after that the chosen sub-classifier only deals with the samples of the same main-class, therefore it can focus on learning to distinguish samples from different subclasses of the same main-class. Obviously, each S-classifier has much less labels than the single classifier which has to classify all classes. Though the similarity of the subclasses in a main-class is remarkable, in summary both the main-classifier and the sub-classifiers of HMC are easier to be learnt compared with the single classifier variant, because the later has much more extra different labels, which will impact the accuracy of classification task.

III. Experiment

In this section, we conduct experiments to show that our proposed framework can achieve pretty good performance in our self collected Badminton Single Sensor dataset (BSS-V2). Then, we carry out ablation studies to evaluate the effectiveness of the proposed HMC and to analyze the influence of the number of main-classes.

1) BSS-V2 Dataset: Recently, we have extended our previous self-collected dataset: Badminton Single Sensor (BSS) dataset [7] to a new version, called BSS-V2. It is also collected with a fixed specialized sensor on a badminton racket. The badminton rocket is composed of four parts: the LED, controller, sensor and rocket, The former three are fixed on the racket. The frequency of sensor is fixed to 200 HZ. The data of each record includes acceleration in a three-dimensional orthogonal system and spatial attitude angles (roll, pitch and yaw). The collected data include 37 major activities, which cover almost all badminton activities, such as “Backhand Hook Diagonal”, “Forehand Clear” and “Backhand High Clear”. On average, each activity contains about 215 samples. On average, each sample contains about 368.9 records, about 1.8s. However, samples from different activities, even the same activity, have significantly different numbers of records. For example, as for the samples of “Backhand Lift”, the shortest sample has 269 records, while the longest one has 719 records.

We segment BSS-V2 dataset into two parts, 80% of it as training set and 20% of it as test set randomly. Besides, we use this splitting strategy five times, as a result, we have five pairs of training and test set. Without otherwise stated, we evaluate all the methods five times with each pair of training and test set, then report the average results of them.

A. Implementation Details

1) Experiment on BSS-V2: We utilize the low-pass ButterWorth filter to eliminate the noise of sensor. In order to reduce the cost of computation, we choose a lightweight CNN, MobileNet [12], as the backbone of our HMC, which is pretrained on ImageNet. The connection of Task 1 and Task 2 is designed manually as Table I, where Task 1 has three main-classes, each main-class has 15, 10 and 12 subclasses, respectively.

TABLE I
THE MAPPING FROM TASK 1 TO TASK 2. “M1”, “M2” AND “M3” ARE THE MANUALLY SPECIFIED MAIN-CLASSES OF TASK 1.

Task 1	M1	M2	M3
Task 2	Forehand High Serve	Forehand Net Lift	Forehand High Clear
	Backhand High Clear	Backhand Net Lift	Overhead High Clear
	Backhand Clear	Forehand Intercept	Forehand Clear
	Forehand Serve	Backhand Intercept	Overhead Clear
	Backhand Serve	Forehand Intercept Drive	Forehand Smash
	Forehand Net Shot	Backhand Intercept Drive	Overhead Smash
	Backhand Net Shot	Forehand Intercept Straight	Midfield Forehand Smash
	Forehand Hook Diagonal	Backhand Intercept Straight	Midfield Backhand Smash
	Backhand Hook Diagonal	Forehand Intercept Diagonal	Forehand Drop Shot
	Forehand Net	Backhand Intercept Diagonal	Overhead Drop Shot
	Backhand Net	-	Forehand Full Strike
	Forehand Net Kill	-	Overhead Full Strike
	Backhand Net Kill	-	-
	Forehand Return	-	-
	Backhand Return	-	-

During training, The input sensor data are cropped with SWAB, following [7]. After that, we resize all samples into 300 records with linear interpolation and compression, as done in [7]. The models without AFEB need extra artificial feature extraction. We compute the mean values, standard deviation, skewness and kurtosis of each attributes as supplement of raw data. As for the models embedded with AFEB, we simply take the raw attribute: three-dimensional acceleration and spatial attitude angles (roll, pitch and yaw) as input. The sliding window of LSTM is 30 records with a overlap of 15 records. We also horizontally flipped the rocket with probability of 0.5 as data augmentation. For network optimization, Adam is used as the optimizer with a momentum of 0.9, weight decay of 10^{-4} and batch size of 64. The initial learning rate of Adam is 0.001. All models are trained for 500 epochs in total. Without otherwise stated, all the experiments are implemented on a linux server with two 1080 GPUs.

B. Comparison with Baselines

In this part, we present comparisons with some methods on BSS-V2 to show the effectiveness of our proposed methods.

As shown in Table II, compared with HMCs and their single classifier versions, the improvements of accuracy are 1.7%, 0.4% and 1.9% for the versions of MobileNetv2, ResNet50 and ResNet101 respectively. It indicates that human prior can leverage the performance of badminton human activity recognition. As the depth of backbone CNNs goes deeper, the performance of single classifiers and our HMCs all become worse due to the lack of training data.

C. Ablation Studies

In this section, we perform ablation studies to evaluate the influence of mapping from the M-classifiers to S-classifiers, illustrated in Table I, and the influence of the number of main-classes.

1) Mapping from the M-classifiers to S-classifiers: Considering the effective enhancement of our proposed HMC, there is a question left: does the human prior is really useful in this task? That is, whether the enhancement of HMC is gained by the human prior mapping or just the strategy of splitting one complicated classification task to a number of easier classification tasks? To answer this question, we construct random mappings from the M-classifier to S-classifiers as well as some other mappings with human prior. The backbones of all versions are set as AFEB-MobileNetv2.

Table VII summarizes the results of HMCs with different mappings from the M-classifiers to S-classifiers. The experiment shows that HMC relys on the mappings constructed manually. Detailly, the mappings built with human priors can boost the performance of backbone models. HMC-MobileNets with mappings from Table V, Table VI and Table I achieve 1.5%, 1.0% and 1.7% gains compared with their backbone AFEB-MobileNet. In addition, HMC-MobileNets with mapping from Table I outperforms other methods, which indicates that better human priors (badminton activities similarity) can facilitate better accuracy, though simple naive human priors (the names of badminton activities) can also be helpful to improve the classification results. The performance of HMCs with random mappings is even worse than their backbone. The reason is that: random mappings will result in the accuracy of M-classifier going down remarkably because the main-classes are meaningless and the separability of subclasses in the same main-classes are probably not more significant than those of different main-classes.

2) Influence of the Number of Main-classes: In this part, we conduct ablation to study the influence of the number of main-classes.

Experimental results on BSS-V2 dataset are shown in Table VIII. It is obvious, with humna prior, even with just two main-classes, our HMC still obtains 0.9% improvement

TABLE II

QUANTITATIVE RESULTS OF BASELINES AND HMC. WE REPORT THE AVERAGE ACCURACY OF FIVE TEST DATASETS WITH THE STRATETY OF CROSS VALIDATION.

Method	**Backbone**	**Pretrain**	**Accuracy**
LSTM	LSTM	-	68.4%
MobileNet	mobilenetv2	ImageNet	80.0%
ResNet50	ResNet50	ImageNet	79.5%
ResNet101	ResNet101	ImageNet	79.4%
AFEB-MobileNet	mobilenetv2	ImageNet	82.2%
AFEB-ResNet50	ResNet50	ImageNet	82.0%
AFEB-ResNet101	ResNet101	ImageNet	81.7%
HMC-MobileNet	AFEB-MobileNetv2, AFEB-MobileNetv2	ImageNet	**83.9%**
HMC-ResNet50	AFEB-ResNet50, AFEB-ResNet50	ImageNet	83.6%
HMC-ResNet101	AFEB-ResNet101, AFEB-ResNet101	ImageNet	82.4%

TABLE III

THE RANDOM MAPPING FROM MAIN-CLASSES TO SUB-CLASSIFIERS. EACH MAIN-CLASS COVERS 12, 12, 13 SUBCLASSES.

Task 1	M1	M2	M3
Task 2	Backhand Hook Diagonal	Overhead Clear	Backhand Return
	Backhand Intercept Diagonal	Forehand Net Kill	Forehand Serve
	Backhand Net Shot	Midfield Backhand Smash	Overhead Full Strike
	Forehand Drop Shot	Forehand Intercept	Forehand Intercept Straight
	Forehand Intercept Diagonal	Forehand Return	Backhand Net Kill
	Backhand Clear	Forehand Net Shot	Forehand High Clear
	Forehand Intercept Drive	Backhand Net	Overhead Drop Shot
	Forehand Net	Forehand Full Strike	Midfield Forehand Smash
	Overhead High Clear	Backhand Intercept	Forehand Hook Diagonal
	Backhand Intercept Straight	Backhand Serve	Backhand High Clear
	Forehand Clear	Backhand Intercept Drive	Forehand Smash
	Forehand High Serve	Backhand Net Lift	Forehand Net Lift
	-	-	Overhead Smash

TABLE IV

ANOTHER RANDOM MAPPING FROM MAIN-CLASSES TO SUB-CLASSIFIERS. EACH MAIN-CLASS COVERS 12, 12, 13 SUBCLASSES.

Task 1	M1	M2	M3
Task 2	Overhead Clear	Backhand Net Kill	Forehand Net Lift
	Backhand Intercept Diagonal	Forehand Net Kill	Forehand Hook Diagonal
	Forehand Net	Forehand Net Shot	Midfield Backhand Smash
	Forehand Clear	Forehand Intercept Drive	Backhand Intercept Straight
	Backhand Intercept Drive	Forehand High Serve	Overhead Drop Shot
	Overhead Smash	Backhand Net Shot	Forehand Intercept Diagonal
	Backhand Intercept	Forehand Drop Shot	Backhand Net
	Overhead Full Strike	Forehand Serve	Backhand Return
	Forehand Full Strike	Midfield Forehand Smash	Overhead high Clear
	Backhand High Clear	Forehand Intercept Straight	Forehand Smash
	Forehand Intercept	Backhand Net Lift	Backhand Hook Diagonal
	Backhand Serve	Backhand Clear	Forehand Return
	-	-	Forehand High Clear

TABLE V

THE MAPPING FROM TASK 1 TO TASK 2. "M1", "M2" AND "M3" ARE THE MANUALLY SPECIFIED MAIN-CLASSES OF TASK 1. IT IS BUILT ACCORDING TO THE NAMES OF BADMINTON ACTIVITIES REGARDLESS OF THEIR REAL SIMILARITY.

Task 1	M1	M2	M3
Task 2	Forehand High Serve	Forehand High Clear	Forehand Net
	Forehand Serve	Overhead High Clear	Backhand Net
	Backhand Serve	Backhand Clear	Forehand Hook Diagonal
	Forehand Intercept	Forehand Clear	Backhand Hook Diagonal
	Backhand Intercept	High Clear	Forehand Intercept Diagonal
	Forehand Intercept Drive	Overhead Clear	Backhand Intercept Diagonal
	Backhand Intercept Drive	Forehand Smash	Forehand Full Strike
	Forehand Intercept Straight	Overhead Smash	Overhead Full Strike
	Backhand Intercept Straight	Midfield Forehand Smash	Forehand Net Kill
	Forehand Net Lift	Midfield Backhand Smash	Backhand Net Kill
	Backhand Net Lift	Forehand Net Shot	Forehand Return
	-	Backhand Net Shot	Backhand Return
	-	Forehand Drop Shot	-
	-	Overhead Drop Shot	-

compared with its backbone. HMC-MobileNet with 2 main-

TABLE VI

ANOTHER MAPPING FROM TASK 1 TO TASK 2 BUT CONSIDERING THE NAMES OF BADMINTON ACTIVITIES. "M1", "M2" AND "M3" ARE THE MANUALLY SPECIFIED MAIN-CLASSES OF TASK 1.

Task 1	M1	M2	M3
Task 2	Forehand High Serve	Forehand Intercept	Forehand Net
	Forehand Serve	Backhand Intercept	Backhand Net
	Backhand Serve	Forehand Intercept Drive	Forehand Full Strike
	Forehand High Clear	Backhand Intercept Drive	Overhead Full Strike
	Overhead High Clear	Forehand Intercept Straight	Forehand Net Kill
	Backhand Clear	Backhand Intercept Straight	Backhand Net Kill
	Forehand Clear	Forehand Intercept Diagonal	Forehand Smash
	High Clear	Backhand Intercept Diagonal	Overhead Smash
	Overhead Clear	Forehand Net Lift	Midfield Forehand Smash
	Forehand Hook Diagonal	Backhand Net Lift	Midfield backhand Smash
	Backhand Hook Diagonal	-	Forehand Net Shot
	-	-	Backhand Net Shot
	-	-	Forehand Drop Shot
	-	-	Overhead Drop Shot
	-	-	Forehand Return
	-	-	Backhand Return

classes gets the best accuracy among three HMC versions in Table VIII. HMC-MobileNet with 3 main-classes also obtains a competitive result.

IV. CONCLUSION

We have established a novel hierarchical multi-classification (HMC) framework for sensor-based badminton activity recognition with the help of the prior information of badminton activity categories. In fact, it is constructed with a M-classifier and certain S-classifier. M-classificator is utilized to cast the input data into a specific feature map. Each feature map has its own S-classifier which acts as an encoder. The fine-grained prediction of input data is given by the chosen S-classificator. In this hierarchical framework, human prior information of the badminton activity categories is manipulated to design the mapping from the main-classes to the sub-classifiers. It is demonstrated by the experimental results on BSS-V2 dataset that our proposed HMC framework is effective and even outperforms the previous state-of-the-art model considerably.

ACKNOWLEDGMENT

This work was supported by the Joint Laboratory of Intelligent Sports of China Institute of Sport Science (CISS).

REFERENCES

[1] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga, "Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey," in *23th International conference on architecture of computing systems 2010*. VDE, 2010, pp. 1–10.

TABLE VII
RESULTS WITH DIFFERENT FORMS OF MAPPINGS FROM THE M-CLASSIFIERS TO S-CLASSIFIERS.

Method	Mapping	Backbone	Pretrain	Accuracy
AFEB-MobileNet	-	MobileNetv2	ImageNet	82.2%
HMC-MobileNet	Table III	AFEB-MobileNetv2, AFEB-MobileNetv2	ImageNet	81.4%
HMC-MobileNet	Table IV	AFEB-MobileNetv2, AFEB-MobileNetv2	ImageNet	81.9%
HMC-MobileNet	Table V	AFEB-MobileNetv2, AFEB-MobileNetv2	ImageNet	83.7%
HMC-MobileNet	Table VI	AFEB-MobileNetv2, AFEB-MobileNetv2	ImageNet	83.2%
HMC-MobileNet	Table I	AFEB-MobileNetv2, AFEB-MobileNetv2	ImageNet	**83.9%**

TABLE VIII
RESULTS WITH DIFFERENT NUMBER OF MAIN-CLASSES.

Method	Main-classes	Mapping	Backbone	Pretrain	Accuracy
HMC-MobileNet	2	Table IX	AFEB-MobileNetv2, AFEB-MobileNetv2	ImageNet	83.1%
HMC-MobileNet	3	Table I	AFEB-MobileNetv2, AFEB-MobileNetv2	ImageNet	**83.9%**
HMC-MobileNet	4	Table X	AFEB-MobileNetv2, AFEB-MobileNetv2	ImageNet	83.8%

TABLE IX
THE MAPPING FROM TASK 1 TO TASK 2. BOTH "M1" AND "M2" ARE THE MANUALLY SPECIFIED MAIN-CLASSES OF TASK 1. IT IS BUILT BY SPLITTING "M3" IN TABLE I INTO TWO PARTS, AND ADD THEM TO "M1" AND "M2" RESPECTIVELY.

Task 1	M1	M2
Task 2	Forehand High Serve	Forehand Net Lift
	Backhand High Clear	Backhand Net Lift
	Backhand Clear	Forehand Intercept
	Forehand Serve	Backhand Intercept
	Backhand Serve	Forehand Intercept Drive
	Forehand Net Shot	Backhand Intercept Drive
	Backhand Net Shot	Forehand Intercept Straight
	Forehand Hook Diagonal	Backhand Intercept Straight
	Backhand Hook Diagonal	Forehand Intercept Diagonal
	Forehand Net	Backhand Intercept Diagonal
	Backhand Net	Forehand Smash
	Forehand Net Kill	Overhead Smash
	Backhand Net Kill	Midfield Forehand Smash
	Forehand Return	Midfield Backhand Smash
	Backhand Return	Forehand Drop Shot
	Forehand High Clear	Overhead Drop Shot
	Overhead High Clear	Forehand Full Strike
	Forehand Clear	Overhead Full Strike
	Overhead Clear	-

TABLE X
THE MAPPING FROM TASK 1 TO TASK 2. "M1", "M2", "M3" AND "M4" ARE THE MANUALLY SPECIFIED MAIN-CLASSES OF TASK 1. THE "M1" IN TABLE I IS SEGMENTED INTO TWO MAIN-CLASSES TO CONSTRUCT THIS MAPPING.

Task 1	M1	M2	M3	M4
Task 2	Forehand High Serve	Forehand Net Lift	Forehand High Clear	Forehand Hook Diagonal
	Backhand High Clear	Backhand Net Lift	Overhead High Clear	Backhand Hook Diagonal
	Backhand Clear	Forehand Intercept	Forehand Clear	Forehand Net
	Forehand Serve	Backhand Intercept	Overhead Clear	Backhand Net
	Backhand Serve	Forehand Intercept Drive	Forehand Smash	Forehand Net Kill
	Forehand Net Shot	Backhand Intercept Drive	Overhead Smash	Backhand Net Kill
	Backhand Net Shot	Forehand Intercept Straight	Midfield Forehand Smash	Forehand Return
	-	Backhand Intercept Straight	Midfield Backhand Smash	Backhand Return
	-	Forehand Intercept Diagonal	Forehand Drop Shot	-
	-	Backhand Intercept Diagonal	Overhead Drop Shot	-
	-	-	Forehand Full Strike	-
	-	-	Overhead Full Strike	-

[2] F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.

[3] P. Rashidi and D. J. Cook, "Keeping the resident in the loop: Adapting the smart home to the user," *IEEE Transactions on systems, man, and cybernetics-part A: systems and humans*, vol. 39, no. 5, pp. 949–959, 2009.

[4] Y. Wang, D. He, F. Li, X. Long, Z. Zhou, J. Ma, and S. Wen, "Multi-label classification with label graph superimposing," *arXiv preprint arXiv:1911.09243*, 2019.

[5] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7794–7803.

[6] J. Lin, C. Gan, and S. Han, "Tsm: Temporal shift module for efficient video understanding," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 7083–7093.

[7] Y. Wang, W. Fang, J. Ma, X. Li, and A. Zhong, "Automatic badminton action recognition using cnn with adaptive feature extraction on sensor data," in *International Conference on Intelligent Computing*. Springer, 2019, pp. 131–143.

[8] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019.

[9] Q. Li, M. Qiao, W. Bian, and D. Tao, "Conditional graphical lasso for multi-label image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2977–2986.

[10] X. Li, F. Zhao, and Y. Guo, "Multi-label image classification with a probabilistic label enhancement model." in *UAI*, vol. 1, 2014, p. 3.

[11] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, "Cnn-rnn: A unified framework for multi-label image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2285–2294.

[12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.