Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 129 (2018) 59–65

Procedia Computer Science

www.elsevier.com/locate/procedia

2017 International Conference on Identification, Information and Knowledge in the Internet of Things

Multi-sensor Golf Swing Classification Using Deep CNN

Libin Jiao[a,b], Hao Wu[a], Rongfang Bie[a,*], Anton Umek[b], Anton Kos[b]

[a] College of Information Science and Technology, Beijing Normal University, Beijing, China, 100875

[b] Faculty of Electrical Engineering, University of Ljubljana, Tržaška cesta 25, 1000 Ljubljana, Slovenia

Abstract

In recent years smart sport equipments have achieved great success in professional and amateur sports, as well as body sensory systems; now discovering interesting knowledge in the surge of data from those embedded sensors used in sports is necessary and the focus of our research. In this paper, we investigate golf swing data classification method based on deep convolutional neural network (deep CNN) fed with multi-sensor golf swing signals. Our smart golf club integrates two orthogonally affixed strain gage sensors, 3-axis accelerometer and 3-axis gyroscope, and collects real-world golf swing data from professional and amateur golf players. Furthermore we explore the performance of our well-trained CNN-based classifier and evaluate it on the real-world test set in terms of common indicators including accuracy, precision-recall, and F1-score. Experiments and corresponding results show that our CNN-based model can satisfy the requirement of accuracy of golf swing classification, and outperforms support vector machine (SVM) method.

Selection and peer-review under responsibility of the scientific committee of the 2017 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI2017).

Keywords: Smart sport equipment; Sensor signals; Golf swing classification; Convolutional neural network

1. Introduction

Data science and technology matter more than ever, especially in professional sports, amateur sports, even recreational sports. After collecting the sport performance data with the integration of lightweight sensors, sensor networks and communication technologies discovering interesting knowledge in these surge of data is necessary. Consequently novel data mining technologies should be employed given the circumstances, and the data volume and time constraints are both demanding aspects in the data analysis.

From the sensors attached to bodies or embedded into smart equipments instructors and analysers are capable of overseeing the complete actions in a group sport match and even get into any particular details in any movements. In practice using data from customized biofeedback application, particular biomechanical feedback systems with terminal and/or concurrent feedback [1] is insightful, therefore, a system for proper action learning can be effectively

* Corresponding author. Tel.: +86-010-58804050.
E-mail address: 92xianshen@mail.bnu.edu.cn, wuhao@bnu.edu.cn, rfbie@bnu.edu.cn, anton.umek@fe.uni-lj.si, anton.kos@fe.uni-lj.si

Selection and peer-review under responsibility of the scientific committee of the 2017 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI2017).
10.1016/j.procs.2018.03.046

implemented by identifying and preventing those wrong actions [2]. We ultimately aim to devise a real-time procedure notifying users when their improper movements occur, during they are performing actions or immediately after a complete action.

Convolutional neural network (CNN) has been one of the most striking classification approaches in computer vision fields including image classification, object recognition, image retrieval, and image generation [3, 4, 5, 6, 7, 8]. Some incarnations including AlexNet [9], VGGNet [10], GoogLeNet [11], and ResNet [12] have achieved superb accuracy compared with traditional classification approaches. Inspired by features of CNN-based models such as its automatic feature extraction and high accuracy, we design a CNN-based classifier to classify our real-world golf swing signals. Collecting golf swing data by our smart golf club integrating two orthogonally affixed strain gage sensors, 3-axis accelerometer and 3-axis gyroscope [1], we assemble a multi-class real-word golf swing dataset from professional and amateur golf players. We also explore the performance of our well-trained CNN-based classifier and evaluate it on this real-world test set in terms of common indicators involving accuracy, precision-recall, and F1-scores. Experimental results show that our CNN-based model can satisfy the requirement of accuracy of golf swing classification, and outperform support vector machine (SVM) method in terms of quantitatively exceeding these indicators.

The paper is organised as follows: Section 2 summaries some related work regarding classification methods based on CNN. Section 3 introduces our network architecture. Section 4 presents the experiments and results to validate our model's effectiveness. Section 5 concludes the paper.

2. Related Work

Since AlexNet [9] achieved a striking performance with top 5 error of 15.4% and exceeded the runner-up that was with 26% error in ImageNet ILSVRC challenge [13] in 2012, convolutional neural network has been popularized in computer vision fields including image classification, object detection, image retrieval and image generation [3, 4, 5, 6, 7, 8, 14, 15, 16]. Derived from Yann LeCun's LeNet-5 [17], particular incarnations including AlexNet [9], VGGNet [10] with stacked small receptive fields, GoogLeNet [11] with Inception module, and ResNet [12] with stacked residual-blocks have reached the top 5 errors of 15.4%, 7.3%, 6.7% and 3.57%, respectively. Besides, CNN-based model quantitatively exceeded human-level performance, with 5.1% top-5 test error [18], which has shown a noticeable superiority in image recognition. Models with dramatical improvements are being gradually proposed with the development of intermediate architecture, such as ResNeXt [19] with parallel pathways and width-increasing shortcut connection, FractalNet [20] with fractal architecture, and Densely Connected Convolutional Networks [21] with dense blocks, all of which improved the classification quality of CNN-based classifiers as well.

In recent years smart sport equipment has been achieved a great success in professional and amateur sports, as well as body sensory systems and wireless networks for personal computing [22, 23, 24, 25, 26, 27, 28]; the combination of smart equipments and data analysis technologies improves the quality of experience in these sports. Inspired by promising aforementioned CNN-based classifiers, we intend to transfer the state-of-the-art classification quality to our golf swing classification by designing a CNN-based model to classify a set of 1-D golf swing signals. These advanced models have given us an inspiration to build our CNN-based classifier. To the best of our knowledge, our CNN-based classifier is an innovative trial in golf swing classification.

3. Architecture of Vanilla Convolutional Neural Network

Our CNN-based classifier for golf swing classification is composed of 3 conventional categories of layers as is the case in general convolutional neural network [9, 29, 30]: convolutional layers, pooling layers, and fully-connected layers, as shown in Figure 1. The network architecture hierarchically involves these three layers, which stacks the convolutional layer with learnable filters and the pooling layer with a fixed stride layer by layer. The network takes as input 8-channel golf swing signals, and produces the probability vector indicating to which predefined class the input signal possibly belongs to. The intermediate 1D convolutional layer convolves the 8-channel signals with trainable kernels to extract features automatically, and propagates forward the activation of the brewed feature maps presents. The max-pooling layers downsample the feature maps from the front convolutional layer, and are activated by the maximum within the predefined stride of the front feature maps. With performing the dimensionality reduction, max-pooling lay-

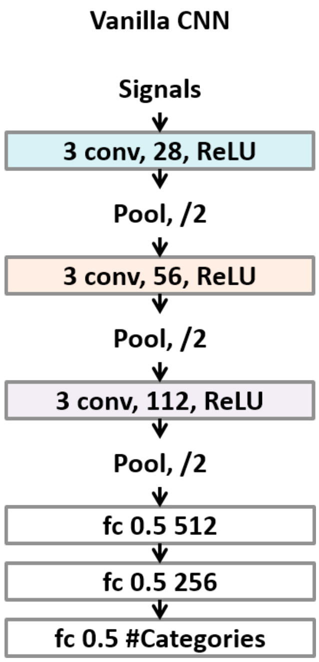

Fig. 1. Vanilla convolutional neural network for golf classification

ers highlight significant features in the forward-propagation to the bottom fully-connected layers nonlinearly classify the extracted features derived from input signals and produce probability vector to determine the inference.

4. Experiments & Results

In this section we present the experiment procedure including our real-world dataset, 10-fold cross-validation to validate the effectiveness of our classifier, indicators of accuracy and precision-recall to show the superiority of our classifier over SVM on behalf of traditional methods. First some meta-information of our real-world golf swing dataset is presented in Table 1 to show the overview of the dataset. Then 10-fold cross-validation is employed to validate our CNN-based classifier on the training set, showing our classifier is adequate to steadily classify golf swings and output coherent results with increasing epochs in terms of growing and steady accuracy. Furthermore our classifier is evaluated and quantitatively compared with SVM, and it can be concluded that our classifier quantitatively outperforms SVM in terms of indicators including overall accuracy and precision-recall.

4.1. Data Augmentation

The dataset is labelled by the hybrid combination of golf swing shapes and golf players. We have collected our data from 4 professional and amateur players denoted with aliases from 1 to 4. Each player performs several predefined golf swings with shapes shown in Table 1. Since the counts of various golf swings in our multi-class dataset are not well-balanced, the dataset is augmented to balance the swing counts in each class, namely, minorities are oversampled with random time translation and data rescaling. Finally the counts of golf swings in each class are balanced to be closer to that of each other, shown in Table 1, where GolferID, SwingShape, NumericalLabelID, Count, and AugCount denote the aliases of golf players, the golf swing shapes, the predefined numerical labels, the counts of various collected golf swings, and the counts of each golf swings after augmentation.

Table 1. Meta-information about real-world golf swing dataset

GolferID	SwingShape	NumericalLabelID	Count	AugCount
1	Straight	1	79	79
	Pull	3	9	54
	Pull-Hook	6	3	42
	Slice	7	9	54
	Push	8	9	54
	Hook	10	5	45
	Fade	14	7	49
	Push-Slice	15	1	32
	Draw	17	11	55
2	Straight	0	30	60
	Pull	5	1	32
	Push	18	4	44
3	Straight	12	15	60
4	Fade	2	2	38
	Push-Slice	4	1	32
	Draw	9	4	44
	Straight	11	11	55
	Slice	13	2	38
	Push	16	10	50

4.2. 10-fold Cross-Validation

10-fold cross-validation is employed to validate the effectiveness and accuracy coherence of our CNN-based classifier. We split the training set into 10 groups by randomly sampling 2/3 data from the whole dataset as the training set, each of which is preserved as a validation set and other 9 groups to train the classifier in repeated 10 turns. Apparently, the accuracy coherence can be used to convincingly demonstrate the effectiveness of our classifier. Since each swing is used for both training for 9 times and validation just only once, which reinforce classifier to sufficiently capture the distribution of the training set, so it will perform a reliably precise inference if the coherent accuracy is detected.

We found out that the coherent accuracy is presented in the 10-fold cross-validation; the accuracy is coherent over 96% and the standard deviations of accuracy are constrained after 40 epochs, so it can be concluded that our classifier can consistently perform well in classification in terms of the coherent accuracy and its convergent standard deviations, as shown in Figure 2. In addition, it is also inferred that our classifier should be converged after 40 epochs of iterations, according to the trend of accuracy in Figure 2.

4.3. Overall Accuracy

In Table 2 the overall accuracy is presented where SVM achieves 86.8% accuracy and our classifier achieves 95.0% accuracy. Apparently our classifier quantitatively outperform SVM on our real-world dataset in terms of accuracy. We attribute this achievement to the nature of our classifier - CNN-based model can be tolerant of data with time translation and data distortion where SVM can possibly fail in classifying. Besides, the mere discrepancy between this accuracy and the accuracy in 10-fold cross-validation also demonstrates our classifier has sufficiently captured the distribution of golf swing dataset, and performs well in generalization.

Table 2. Overall accuracy of vanilla CNN

	Model	
	SVM	CNN
all sensors	0.867986799	0.95049505

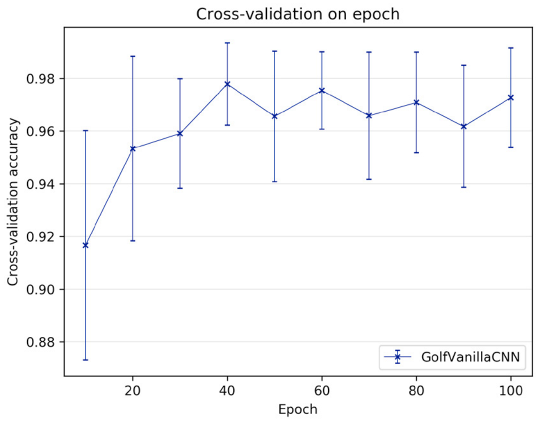

Fig. 2. 10-fold cross-validation for vanilla CNN

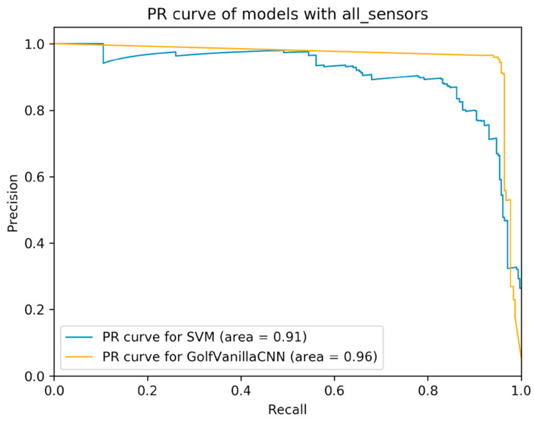

Fig. 3. Precision-Recall curve of classification from all sensors

4.4. Precision, Recall, and F1-score

Precision-recall is employed here to evaluate the quality of classification, more specifically, the quality of classification of each class and generalization of our classifier. In our case the precision is a measure reflecting golf swing relevancy that indicates the rate of correct golf swings the classifier can retrieve; recall is a measure reflecting how many truly relevant golf swings are retrieved that indicates the sensitivity of the classifier when it is confronting plausibly incorrect golf swings. Moreover the synthesized F1-score can evaluate the quality of classification by taking precision and recall into consideration. The precision-recall reports and illustration are shown in Table 3 and Figure 3.

Table 3. Classification report of SVM and Convolutional Neural Net (ConvNet) with all sensors

	SVM				ConvNet			
	precision	recall	f1-score	support	precision	recall	f1-score	support
0	1.00	0.62	0.77	24	1.00	0.83	0.91	24
1	0.93	0.90	0.91	29	0.88	0.79	0.84	29
2	1.00	1.00	1.00	11	0.92	1.00	0.96	11
3	1.00	1.00	1.00	17	0.85	1.00	0.92	17
4	1.00	1.00	1.00	9	1.00	1.00	1.00	9
5	1.00	1.00	1.00	11	1.00	1.00	1.00	11
6	0.61	1.00	0.76	14	0.93	1.00	0.97	14
7	0.94	0.89	0.92	19	1.00	1.00	1.00	19
8	0.93	0.81	0.87	16	1.00	1.00	1.00	16
9	0.85	0.73	0.79	15	1.00	1.00	1.00	15
10	1.00	1.00	1.00	16	0.94	1.00	0.97	16
11	0.56	0.88	0.68	17	1.00	0.94	0.97	17
12	1.00	1.00	1.00	19	1.00	1.00	1.00	19
13	1.00	0.33	0.50	12	1.00	1.00	1.00	12
14	0.88	0.88	0.88	16	1.00	1.00	1.00	16
15	1.00	1.00	1.00	10	1.00	1.00	1.00	10
16	0.87	0.87	0.87	15	1.00	0.93	0.97	15
17	1.00	0.78	0.88	23	0.91	0.87	0.89	23
18	0.53	1.00	0.69	10	0.71	1.00	0.83	10
avg / total	0.91	0.87	0.87	303	0.95	0.95	0.95	303

Taking the averages of precision and recall into consideration altogether, it can be found out that our classifier can quantitatively outperform SVM in terms of both precision and recall, which means our classifier performs better in capturing precise golf swings and retrieve as many relevant swings from the dataset as it is able to. F1-scores in Table 3 and the precision-recall curve in Figure 3 also support our above conclusion. Additionally the generalization of our classifier is basically guaranteed on account of the high individual precisions and recalls, which means our classifier is adequate to classify each golf swing.

5. Conclusion

In this paper we investigate golf swing data classification methods based on deep convolutional neural network (deep CNN) fed with multi-sensor golf swing signals. Training on our real-world golf swing data collected from professional and amateur golf players by our smart club integrating two orthogonally affixed strain gage sensors, 3-axis accelerometer and 3-axis gyroscope, we explore the performance of our well-trained CNN-based classifier and evaluate it on the real-world test set in terms of common indicators involving accuracy, precision-recall, and F1-scores. According to the experimental results it can be concluded that our CNN-based model satisfies the requirements of accuracy and precision of golf swing classification, and effectively outperform support vector machine (SVM) method.

In future we plan to investigate advanced CNN-based modules including stacked micro receptive fields, Inception modules and residual-block modules, and figure out whether these modules can bring improvements into our classification. How to choose a proper sequence length is also a valuable topic since the real-time feedback can be satisfied once CNN-based classifier can use short sequence to classify accurately. Furthermore the relevancy and redundancy of sensors is another proper topic to explore thoroughly since the time and resource consumption can be reduced once the classifiers can classify with low-dimensional signals as accurately as they with full-dimensional signals, which is also a prerequisite for real-time analysis.

Acknowledgements

(a) This research is partially sponsored by National Natural Science Foundation of China (No.61571049, 61401029, 11401028, 61472044, 61472403, 61601033) and the Fundamental Research Funds for the Central Universities (No.2014KJJCB32, 2013NT57), the BNU Graduate Students Platform for Innovation & Entrepreneurship Training Program (No.3122121F1), and the Fundamental Research Funds for the Central Universities(2016NT14), and by SRF for ROCS, SEM. (b) The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P2-0246).

References

[1] Anton Umek, Sašo Tomažič, and Anton Kos. Wearable training system with real-time biofeedback and gesture user interface. *Personal and Ubiquitous Computing*, 19(7):989–998, 2015.

[2] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review*, 20(1):21–53, 2013.

[3] Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S Lew. Deep learning for visual understanding: A review. *Neurocomputing*, 187:27–48, 2016.

[4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[5] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

[6] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[7] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2536–2544, 2016.

[8] Chao Yang, Xin Lu, Zhe Lin, Eli Shechtman, Oliver Wang, and Hao Li. High-resolution image inpainting using multi-scale neural patch synthesis. *arXiv preprint arXiv:1611.09969*, 2016.

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[10] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[11] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[13] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[14] Hao Wu, Yueli Li, Zhenjiang Miao, Yuqi Wang, Runsheng Zhu, Rongfang Bie, and Yi Wang. Creative and high-quality image composition based on a new criterion. *Journal of Visual Communication & Image Representation*, 38(C):100–114, 2016.

[15] Hao Wu, Zhenjiang Miao, Yi Wang, and Manna Lin. Optimized recognition with few instances based on semantic distance. *Mathematics & Computers in Simulation*, 31(4):1–9, 2015.

[16] Hao Wu, Zhenjiang Miao, Yi Wang, Jingyue Chen, Cong Ma, and Tianyu Zhou. Image completion with multi-image based on entropy reduction. *Neurocomputing*, 159(C):157–171, 2015.

[17] Yann LeCun, LD Jackel, Léon Bottou, Corinna Cortes, John S Denker, Harris Drucker, Isabelle Guyon, UA Muller, Eduard Sackinger, Patrice Simard, et al. Learning algorithms for classification: A comparison on handwritten digit recognition. *Neural networks: the statistical mechanics perspective*, 261:276, 1995.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[19] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. *arXiv preprint arXiv:1611.05431*, 2016.

[20] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Fractalnet: Ultra-deep neural networks without residuals. *arXiv preprint arXiv:1605.07648*, 2016.

[21] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten. Densely connected convolutional networks. *arXiv preprint arXiv:1608.06993*, 2016.

[22] Zaobo He, Zhipeng Cai, Siyao Cheng, and Xiaoming Wang. Approximate aggregation for tracking quantiles and range countings in wireless sensor networks. *Theoretical Computer Science*, 607:381–390, 2015.

[23] Siyao Cheng, Zhipeng Cai, Jianzhong Li, and Hong Gao. Extracting kernel dataset from big sensory data in wireless sensor networks. *IEEE transactions on knowledge and data engineering*, 29(4):813–827, 2017.

[24] Siyao Cheng, Zhipeng Cai, Jianzhong Li, and Xiaolin Fang. Drawing dominant dataset from big sensory data in wireless sensor networks. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*, pages 531–539. IEEE, 2015.

[25] Zhipeng Cai, Randy Goebel, Mohammad R Salavatipour, and Guohui Lin. Selecting dissimilar genes for multi-class classification, an application in cancer subtyping. *BMC bioinformatics*, 8(1):206, 2007.

[26] Jiguo Yu, Wenchao Li, and Li Feng. Connected dominating set construction in cognitive radio networks. In *International Conference on Identification, Information, and Knowledge in the Internet of Things*, pages 1–13, 2016.

[27] Dongxiao Yu, Li Ning, Yifei Zou, Jiguo Yu, Xiuzhen Cheng, and Francis C. M. Lau. Distributed spanner construction with physical interference: Constant stretch and linear sparseness. *IEEE/ACM Transactions on Networking*, PP(99):1–14, 2017.

[28] Jiguo Yu, Shengli Wan, Xiuzhen Cheng, and Dongxiao Yu. Coverage contribution area based *k*-coverage for wireless sensor networks. *IEEE Transactions on Vehicular Technology*, PP(99):1–1, 2017.

[29] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[30] Jake Bouvrie. Notes on convolutional neural networks. 2006.