Shot-Net: A Convolutional Neural Network for Classifying Different Cricket Shots

Md. Ferdouse Ahmed Foysal[1(✉)], Mohammad Shakirul Islam[1], Asif Karim[2], and Nafis Neehal[1]

[1] Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh
{ferdouse15-5274,shakirul15-311,nafis.cse}@diu.edu.bd

[2] College of Engineering and IT, Charles Darwin University, Darwin, Australia
asif.karim@cdu.edu.au

Abstract. Artificial Intelligence has become the new powerhouse of data analytics in this technological era. With advent of different Machine Learning and Computer Vision algorithms, applying them in data analytics has become a common trend. However, applying Deep Neural Networks in different sport data analyzing tasks and study the performance of these models is yet to be explored. Hence, in this paper, we have proposed a 13 layered Convolutional Neural Network referred as "Shot-Net" in order to classifying six categories of cricket shots, namely Cut Shot, Cover Drive, Straight Drive, Pull Shot, Scoop Shot and Leg Glance Shot. Our proposed model has achieved fairly high accuracy with low cross-entropy rate.

Keywords: Cricket shot classification · Convolution neural network · Deep learning

1 Introduction

Cricket is one of the most exciting games in the world, batting is the ability of hitting the cricket ball with a cricket bat, and there are different kinds of cricket shots. Batsmen have to accommodate to various conditions when playing on different cricket pitches, especially in different countries therefore, as well as having distinguished physical batting skills, top-level batsmen will have thunder reflex action, excellent decision making and be good strategists [6]. Application of computer vision and machine learning techniques in cricket for different analysis is an emerging domain now. In cricket plethora of technologies used for visualization and coaching [1–3]. From recent researches till now satisfactory results for detecting shots are not achieved.

We thought that we can do a deep convolution neural network (CNN) [2] based action detection. Therefore, we are proposing a novel approach to classify different types of cricket shots using Convolutional Neural Network and

K. C. Santosh and R. S. Hegadi (Eds.): RTIP2R 2018, CCIS 1035, pp. 111–120, 2019.
https://doi.org/10.1007/978-981-13-9181-1_10

Deep Learning. An intelligent device can recognize the human corpus parts by extracting features from real data using algorithms [4]. Then all the parts of the body will be classified [5] by applying various techniques of action for the representation of model. In machine learning image processing and pattern recognition, extraction of features started from a basic set of data being measured and constructs evolved values intended to be copious and non-surplus, the following generalization moves and generalization should be facilitating. Extraction of features is related to dimensional reduction. We proposed a CNN based model where we input images in three convolution layer, three max pooling layer, four dropout layer and two dense layer.

Our goal was to classify six types of cricket shot activities by our own developed model. On our own generated and modified data-set, our CNN model recognize the activity of Cricket Shots, and moreover the similarities and difference between different shots. This task was formulated as a deep learning based real life problem, of which the difference is thought by a feature extraction approach. All other section of this paper is described as few sections, Sect. 2 elaborates the background study and related work. In Sects. 3 and 4 our model representation, train and test procedure described, result discussion is given in Sect. 5.

2 Literature Review

Several studies published on Cricket since last decade. The Hawk-Eye [1] by Collins and Evans is studied about advanced system of coaching for cricket. In 2015 a research paper was published on cricket shot classification based on motion vector by a group of Bangladeshi researcher. For action recognition, they use 3D MACH to classify the shots and to detect cricket shots they define 8 classes of angle ranges [7]. In 2016 Dixit and Balakrishnan from Stanford University published a report on Deep Learning using CNN's for Ball-by-Ball Outcome Classification in Sports [8]. They compare the performance of three different Convolutional Neural Network architectures, inspired by literature on activity recognition in videos.

In 2010 Yao and Fei-Fei published a paper on Object Interaction Activities of Human pose by using mutual context modeling [9]. In their research paper they described a new model of random field to encode poses of human in activities of human-objects. They mold the learning structure problem as a learning task of model, the summary of human activity pose, and the parts of the body are calculated using a search of structure method and new max-margin algorithm used for estimating the parameters of the model.

In another research paper, Batra, Gupta, Yadav, Gupta and Yadav proposed a multi-valued automated decision whether a ball is no-ball or wide ball [6]. Presenting game specific concept selection and event selection criteria. Another cricket shot classification using computer vision proposed by Chowdhury and Jihan divided the approach into four phases of identifying batsman's hand stroke direction, tracking, detection of a collision of bat and ball and detection of human pose and skeleton joints [10]. Another work of a semantic video analysis based on image features of high and low level knowledge for cricket video sequences [12]. Angadi and Vilas Naik worked on detection of shot boundary technique using regional color moments. They said that it can become a measurement of discovering difference among sequent video frames [11]. In [13–19], different approaches of object classification, and sports analysis works has been shown.

3 Proposed Methodology

In this canto we described our model implementation, the process of cricket shot classification. The process of training the model, testing and validation, data-set collection, data preparation, data augmentation, data resize, proposed model description and finally training procedure of the model is described in this part.

3.1 Background of CNN

Deep learning is a technique for implementing Machine Learning. It is made of artificial neural networks. Neural networks work as similar as our brain. CNN that means Convolutional Neural Network is one of the strongest networks in deep learning. It is an artificial neural network, which is also known as feed-forward ANN. In a "feed-forward" network information flows right through the networks.

Yann LeCun was the inventor of CNN. Inspired from human processes he made it. Actually CNN works like biological visual cortex. CNN is one of the most successful models in image classification. CNN's classification accuracy is better than any other traditional image classification algorithms. In CNN we don't have to do feature selection, but in other image classification algorithms, we have to do it. There are different types of layers that are used in CNN.

Convolution layer has a moving filter or kernel which passes through the image. Generally it passes through a 2D matrix (representation of image) and take a certain portion and applies dot multiplication and stores it in another matrix (Fig. 1).

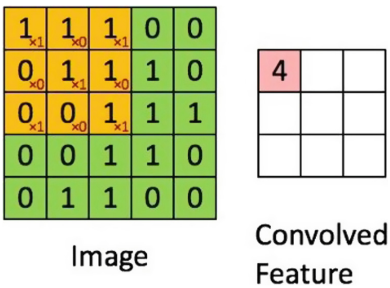

Fig. 1. Convolution of a filter over 2D image

Dimension of the output matrix can be calculated by an equation. We can see an equation bellow where:

n_{out} – Output dimension
n_{in} – Input dimension
f – Window size
S – Stride

$$n_{out} = floor(\frac{n_{in} - f}{S}) + 1$$

$$Equation\ to\ find\ Output\ dimension$$

Pooling layer generally sits next to convolution layer. It mainly used to reduce memory and for fast computation. It reduces the volume. Max pooling is one of the most used layers in CNN. It sets a kernel and finds the max number from the matrix (Fig. 2).

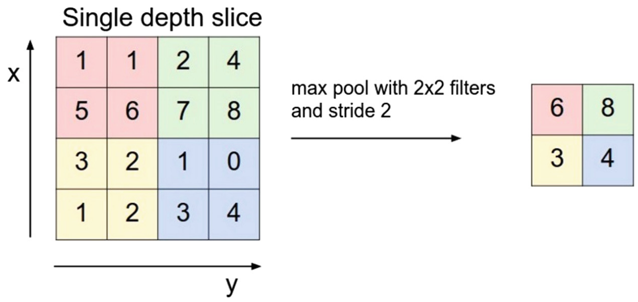

Fig. 2. Max pooling.

Fully connected layer gets 2D or 3D array as input from previous layer and converts the 2D or 3D array into 1D array.

The output layer of a convolutional neural network shows the probability of the classes. It is calculated by "Softmax" function. The equation of calculating the probability is given bellow.

$$\sigma(X_j) = \frac{e^{x_j}}{\sum_{i^{e^{x_i}}}}$$

3.2 Data-Set Collection

We have made a dataset of 3600 images. The dataset have 6 classes of cricket shots, each class contains 600 images. The classes are Cut shot, Cover drive, Straight drive, Pull shot, Leg glance shot and Scoop shot. We took 80% image of the dataset that means 2880 images to train the model and 20% image of the dataset that means 720 images for testing. In the train dataset, each class contains 480 images and in the test dataset, each class contains 120 images. All the data were processed and collected by the authors (Fig. 3).

Fig. 3. The example of our dataset.

3.3 Data Augmentation

We artificially expanded the dataset to avoid overfitting. It helps to increase the amount of relevant data in the dataset. We augmented the real main data-set in 5 different way.

- Rotate $-30°$
- Rotate $+30°$
- Shear by a certain amount
- Adding Salt and Pepper noise
- Shading

3.4 Data Preperation

All the images of our dataset have different dimension such as height, width and size. Since our model requires a similar pixel size data for train and test purpose, we resized the data-set into 100×100 pixels. We have also converted the images into grayscale. Because of lower GPU in our computer, we used grayscale images to train the model.

3.5 Proposed Model

We proposed our own CNN model, which have 13 layers. There are three convolutional layers:

- First layer have 32-3 × 3 filters and 'linear' as activation function.
- Second layer have 64-3 × 3 filters and 'linear' as activation function.
- Third layer have 128-3 × 3 filters and 'linear' as activation function (Fig. 4).

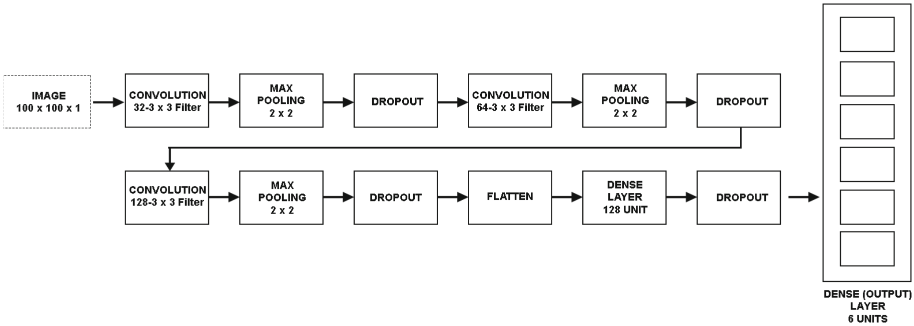

Fig. 4. Architecture of our model.

In addition, there are three max-pooling layers each of size 2 × 2. There are four dropout layer with parameter 0.20. We have a flatten layer, in the model. Lastly there are two dense layer, where in one we used 'linear' as activation function and in the other we used 'softmax' as activation function. We used softmax activation function to get the probability of each class.

3.6 Training the Model

We used Adam optimizer to compile our model. We used 80% of training dataset to train and 20% is used for validation. Training dataset has 2880 images so that we used 2304 images to train and 576 images to validate. We used a batch size of 64. We have trained the network for 40 epochs.

4 Performance Evaluation

Training accuracy is usually the accuracy when the model is applied on the training data. When the model is applied on a few selected data from random class, is known as validation accuracy. Figure 5 shows a graph which contains training and validation accuracy of our model.

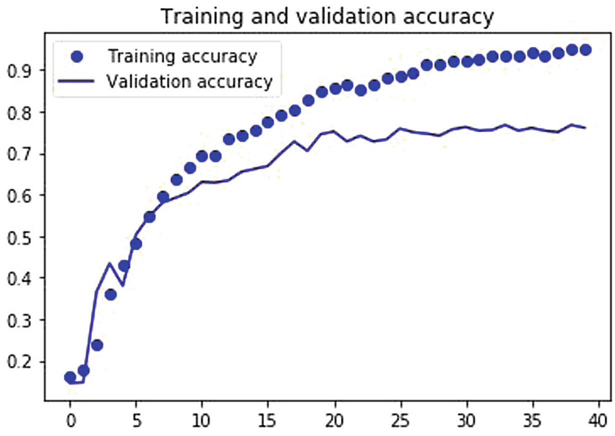

Fig. 5. Training and validation accuracy.

The error on the training data-set is called training loss. The error occurs after running the validation data-set through the trained network is known as validation loss. Figure 6 shows a graph which contains training and validation loss of our model.

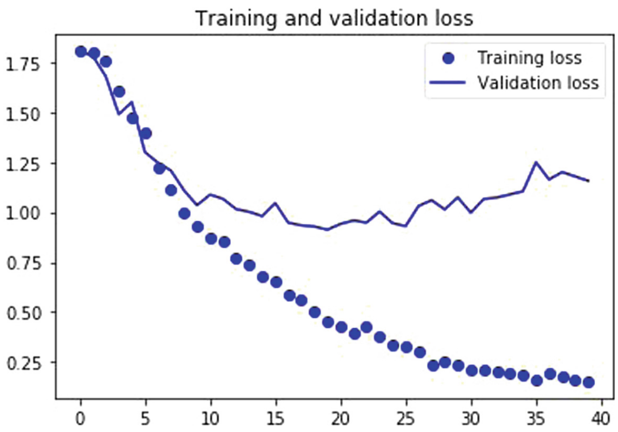

Fig. 6. Training and validation loss.

5 Result Discussion

We calculated Precision, Recall and F1-score from test dataset containing 840 images from the classification report we can see Precession average is 0.80, Recall average is 0.79 and average F1-score is 0.79. So it can be said that the performance of our classifier is pretty good, classification report given in Table 1.

From table of Classification report we can see that the classifier achieved a decent accuracy, which is 80%.

Table 1. Classification report.

Class	Precision	Recall	F-score
Cut Shot	0.69	0.76	0.72
Cover Drive	0.74	0.78	0.76
Straight Drive	0.78	0.83	0.81
Pull Shot	0.89	0.77	0.83
Leg glance Shot	0.79	0.88	0.83
Scoop Shot	0.88	0.72	0.79
Avg.	0.80	0.79	0.79

We describe the performance of our model by few figures, Fig. 7(a) Shows the confusion matrix without normalization and Fig. 7(b) Show the normalized confusion matrix.

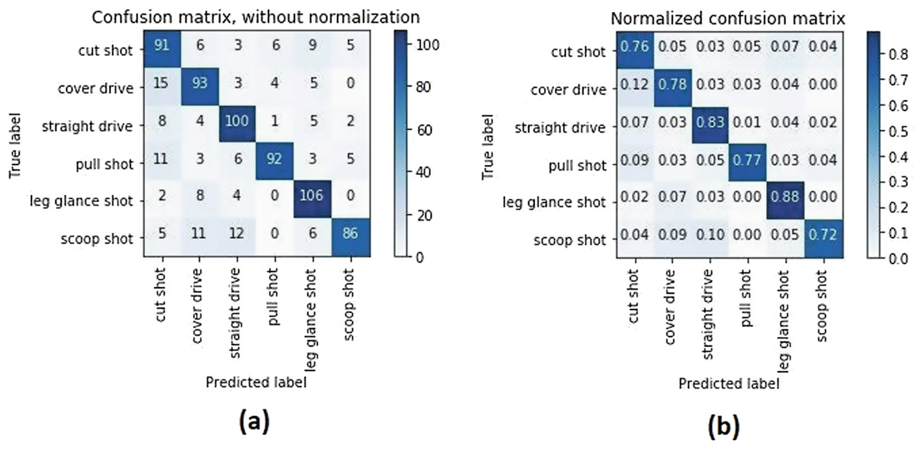

Fig. 7. (a) Confusion matrix, (b) Normalized confusion matrix.

6 Future Work

In our proposed method we can classify different kinds of cricket shots, we have used convolution neural networks to build a model for our Shot-Net data. Our future goal is to make a better neural network to get a better accuracy. We have a plan to do a 3D depth image based classification by deep learning. Where we will use MS Kinect or Intel RealSense, We will use different types of algorithm and that will select the efficient one.

7 Conclusion

In this paper, we provide an approach of cricket shot classification approach by our CNN model. We used three convolution layer, three max polling layer, four dropout layer, one flatten and two dense layer. We use dropout layers to reduce overfitting. The final achieved result we've found is so promising. We hope that this method will be developed as a real application in future for the welfare of cricket game. It will be effective for coaching system also, to improve bowling skill as-well as batting skill too.

References

1. Collins, H., Evans, R.: Hawkeye second edition public understanding revised 07 clean (2008)
2. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet Classification with Deep Convolutional Neural Networks (2012)
3. Collins, H., Evans, R.: Hawkeye-final-submission (2012)
4. Shotton, J., Sharp, T., Kipman, A.: Real-time human pose recognition in parts from single depth images. Commun. ACM (2013). http://dl.acm.org/citation.cfm?id=2398381
5. Andriluka, M., Roth, S., Schiele, B.: Pictorial structures revisited: people detection and articulated pose estimation. In: Computer Vision and Pattern, pp. 1014–1021 (2009)
6. Batra, N., Gupta, H., Yadav, N., Gupta, A., Yadav, A.: Implementation of augmented reality in cricket for ball tracking and automated decision making for no ball, pp. 316–321 (2014). https://doi.org/10.1109/ICACCI.2014.6968378
7. Wikipedia, Different Cricket Shots. https://en.wikipedia.org/wiki/Batting (cricket)
8. Dixit, K., Balakrishnan, A.: Deep learning using CNN's for ball-by-ball outcome classification in sports, report submission on the course of Convolutional Neural Networks for Visual Recognition, Stanford University (2016)
9. Yao, B., Fei-Fei, L.: Modeling Mutual Context of Object and Human Pose in Human-Object Interaction Activities. IEEE (2010). 978-1-4244-6985-7/10/26.00
10. Chowdhury, A.Z.M.E., Jihan, A.U.: Classification of Cricket Shots Using Computer Vision (2014)
11. Angadi, S.A., Naik, V.: A shot boundary detection technique based on local color moments in YCbCr color space. In: CS and IT-CSCP 2012 (2012)
12. Kolekar, M.H., Palaniappan, K., Sengupta, S.: Semantic event detection and classification in cricket video sequence. In: 2008 Sixth Indian Conference on Computer Vision, Graphics and Image Processing, Bhubaneswar, pp. 382–389 (2008)
13. Islam, M.S., Foysal, F.A., Neehal, N., Karim, E., Hossain, S.A.: InceptB: a CNN based classification approach for recognizing traditional bengali games. In: ICACC-2018 (2018)
14. Patel, H.A., Thakore, D.G.: Moving object tracking using Kalman filter. IJCSMC **2**(4), 326–332 (2013)
15. Zhu, G., Huang, Q., Xu, C., Xing, L., Gao, W., Yao, H.: Human behavior analysis for highlight ranking in broadcast racket sports video. IEEE Trans. Multimedia **9**(6), 1167–1182 (2007)

16. Simonyan, A., Zisserman, K.: Very deep convolutional networks for large-scale image recognition (2014)
17. Rock, R.A., Gibbs, A., Carlos, P.H.: The 5 the Umpire: Cricket"s Edge Detection System (2012)
18. Forsyth, D.A., Brien, V.O.: Computer Vision: A Modern Approach, 2 edn, pp. 88–101 (2003)
19. Lowe, D.G.: Distinctive image features from scale invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)