

Object detection on robosoccer environment using convolution neural network

Diana Steffi¹, Shilpa Mehta¹, K. A. Venkatesh²

¹Department of Electronics and Communication Engineering, Presidency University, Bangalore, India

²School of Mathematics and Natural Science, Chanakya University, Bangalore, India

Article Info

Article history:

Received Jul 23, 2022

Revised Sep 12, 2022

Accepted Sep 28, 2022

Keywords:

Computer vision

Convolutional neural network tracking

Nao

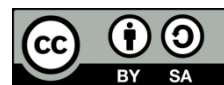
Object detection

Robosoccer

ABSTRACT

Robots with autonomous capabilities depend on vision capabilities to detect and interact with objects and their environment. In the field of robotic research, one of the focus areas is the robosoccer platform that is being used to implement and test new ideas and findings on computer vision and decision making. In this article, an efficient real-time object detection algorithm is employed in a robosoccer simulation environment by deploying a convolution neural network and Kalman filter based tracking algorithms. This study's objective is to classify nao, ball, and the goalpost as well as to validate nao and ball tracking without human intervention from initial frame to last frame. In comparison with the existing methods, the proposed method is robust and fast in identifying three classes namely nao, ball, and goalpost with a speed of 1.67 FPS and a mAP of 95.18%. By implementing this approach, soccer playing robots can make appropriate decisions during game play.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Diana Steffi

Department of Electronics and Communication Engineering, Presidency University

Bangalore-560064, India

Email: steffiseelan@gmail.com

1. INTRODUCTION

In Stanford research institute, the research team under Charles Rosen developed a robot called "Shakey". Shakey wheeled around the room and even in unfamiliar surroundings, observe the scene with his television eyes, responded to the environment to a certain extent. In 1960, Joseph Eagleburger who is known as the father of robotics developed an industrial robot by modifying the Devol robot. Hence then, the field of robotics have grown and developed rapidly. In the 21st century, the automation in various industries are facilitated by industrial robots, which are getting better day by day due to the development of artificial intelligence technology.

There are many artificial intelligence (AI) events that occur across the globe, and one such event is robocup. In the International Joint Conference on Artificial Intelligence (IJCAI-95) held at Montreal, Canada, an official announcement was made regarding organizing the first robot world soccer game and conferences in association with IJCAI-97 in Nagoya, by giving a two-year preparation time. The first robocup which was held in 1997 was a huge success in which more than 40 teams participated and it was identified as a great research platform for many researchers. The objective of the robocup is to focus on the research and development in all major AI tasks: perception, control, navigation, strategy and planning. Perfect identification and detection of the targets is very crucial for the robot players to perform all the five above mentioned AI tasks on the robot soccer environment. The computer vision of the robot players forms the basis of the entire AI task that the robot has to perform on the soccer environment. Object detection and recognition is a key factor that should be perfect and at the best for improving the game playing. Feature extraction from an image is the basic technology

of machine vision that involves extraction of certain key information with subsequent recognition and classification, which are now being used in real sports also for improving the playing skills. Along with the well-developed machine vision the robot player needs time coordination and collision avoidance strategies as discussed by Tabasso *et al.* [1]. The development and the on-going research in the field of intelligent soccer playing robots is aimed at industrial robot to perform the next level of intelligence and also in development of autonomous cars [2] which involve object detection, obstacle avoidance, tracking just like the soccer playing robots.

However, an enormous amount of research is being carried out on AI and deep learning for automation which is tested on the robocup platform, the traditional existing, object detection and tracking algorithm has its own merits and limitations like, slow recognition and tracking, low accuracy and it needs human intervention for operations. This article proposes an improved approach to object detection and tracking with no human intervention based on AI coupled with deep learning and this method improves the accuracy and the speed of playing robot. The model proposed in this article is known as ROBOSOCCER, which is used to convolve the image layers, and perform classification of the three class: Nao, Ball and the goal-post and for tracking stage Kalman Filter is employed. The robustness, fast and accuracy of the proposed approach is evident from the simulation results.

2. ROBOCUP AND OBJECT DETECTION

In this session a brief study of the trends in the development of object detection and tracking algorithm is done. The mobile robot needs to identify various objects based on the application in which field it is deployed. In the robosoccer platform, there are two categories of moving object: nao, ball and one static object, the goal-post.

2.1. The development trend in robocup on object detection and tracking

Menashe *et al.* [3] and Cruz *et al.* [4], proposed ball detection based on machine learning. Speck *et al.* [5] proposed an approach that uses neural networks for real-time classification and detection on a humanoid nao robot. Apart from deep learning methods, there are approaches that depends on colour classification, that works with the idea that the entire soccer environment is green and the objects can be recognized by finding the gaps as proposed by Hoffmann *et al.* [6], Lenser and Veloso [7], Leiva *et al.* [8], Dijk and Scheunemann [9] proposed a neural network based semantic segmentation that combines the robot vision techniques on grayscale images and the convolutional neural network (CNN) classification. The object detection and tracking on the robosoccer environment is a complex problem as it depends on various factors like shadows, reflections, obstructions, vanishing points, high activity points and there are main challenges like occlusions, speed, multiple scales and limited data. The vision sensors of the robot players is used to capture the environment, but due their mobility, the captured data is imprecise. Due to the uncertainty in the captured data, the identification of the object is harder and hence there is a necessity of improvement.

The robot soccer league had moved from orange ball to black and white ball from 2016 league. The work done by Menashe *et al.* [3] detects a ball, without knowing the ball's position for which series of heuristic region of interest identification techniques and super-vised machine learning methods was used. Kukleva *et al.* [10] proposes a CNN approach for detecting and tracking the robotsoccer ball, in their approach they use the history of frames instead of using the current frame to detect the ball. Loncomilla and Solar [11] proposed YoloSPoc which uses maximal activation convolutions descriptors in which good quality object proposals are done by YoloV3. Poppinga and Laue [12] proposed just enough time (JET-NET) convolutional neural network frame which is able to perform at its best in the real time robot detection on the robosoccer environment. JET-NET does only player detection using transfer learning. Szemenyei and Castro [13], proposed a new architecture for object detection on robot soccer environment, which again uses transfer leaning, and the proposed architecture is said to have outperformed tiny-yolo in terms of speed and accuracy.

Teimouri *et al.* [14] proposed a new method for detecting soccer ball for low-cost humanoid robot, which achieves high accuracy of up to 97.17% in which the entire ball is extracted by recursive algorithm using key image-based feature after which a light weight CNN is used. Houliston and Chalup [15] proposed a geometric input transformation called visual mesh to generate a plot on visual space, which reduces the complexity in computation, by standardizing the pixel and the object's feature density. This proposed work was an enhancement of CNN for object detection in resource-constraint robots. Szemenyei and Castro [16] proposed an end-to-end neural network for nao vision on robot soccer environment in which there are two key neural networks, one for segmentation and the other for propagation. Szemenyei and team had trained the models on dummy datasets and later tuned them on real images from Nao robot. Leiva *et al.* [8] proposed a framework that detects the ball and other nao players, their orientation and key features of the field without using any colour information, as all processing is done on grayscale images and cascade methodology is used which combines classical approaches and modern CNN based classifiers. Felbinger *et al.* [17] proposed the

genetic algorithm approach, which optimized the CNN hyper parameters, with minimum size dataset that resulted in cost effective inference on nao robot.

2.2. The development trend in CNN on object detection

Lecun *et al.* [18], established the basic of image classification using the convolution neural network. Even though LeNet had proved that CNN could do magic in image classification, there was no big development due to the fact of limited computing power and availability of data and it was believed that CNN could be used for digit identification only. In the complex recognitions like face detection or object detection, a HarrCascade or SIFT feature extractor with an SVM classifier was used. Krizhevsky *et al.* [19], proposed an approach that used the multi-layer CNN concept from LeNet and reached an accuracy of 84.7%. In their work, the size of the CNN was increased and rectified linear unit (ReLU) non-linearity was used that accounted for computation cost reduction and the ImageNet size was also drastically increased.

Rethage *et al.* [20], representing the visual geometric group, proposed the VGG network which is now being used as the back-bone network for most of the computer vision models like fuzzy cognitive networks (FCN) that are used for semantic segmentation and for Faster R-CNN for object detection. The VGG reached an excellent result of 93.2%. Szegedy *et al.* [21], focused on reducing the computation cost and gradient diminishing problems rather than focusing on performance improvement like VGG and GoogLeNet. Lofte and Szegedy [22] introduced the concept of batch normalization and used mini-batches to approximate the entire dataset and increased the training time. They also introduced the learnable parameters: scale and shift, which makes the network, normalize each layers on its own.

He *et al.* [23] proposed the ResNet model, in which irrespective of number of layers used, the gradient flow in the network is found to be better. The residual model learns the variations between the input and the output without each layer fitting to the feature mapping, which results in better results with less information. Chollet [24] introduced the Xception network that outperforms both ResNet and InceptionV3 and became source of open tool for deep learning approaches. They suggested that in convolution neural networks, decoupling of cross-channel correlations and spatial correlations is possible. Harjoseputro *et al.* [25] introduced MobileNet, which is same to same like Xception, but MobileNet uses depth-wise convolution with minimum parameters and high efficiency.

Zoph *et al.* [26] proposed to search for an architectural building block on a small dataset and then transfer the block to a larger dataset. New search space called the neural architecture search network (NASNet) search space was introduced in the work which enables transferability and the authors have proposed a new regularization technique called scheduled drop path that significantly improves generalization in NASNet models. Tan and Le [27] proposed a new scaling method that uniformly scales all dimensions of depth, width, resolution using an effective compound coefficient. Neural architecture search is used to design a new baseline network call it up to obtain a family of models called Efficient Nets, which achieves higher accuracy and efficiency. Sergey *et al.* [28] proposed a deep learning-based object detection and position estimation and the potential applicability of the developed work framework was then demonstrated on an experimental robot-manipulation setup realizing a simplified object pick and place scenario. Kanthi *et al.* [29] proposed a multi-scale 3D-convolutional neural network for hyperspectral image classification. There are many other works in the world of CNN towards image recognition that are famous in their own area, like CNN proposed by Murugan *et al.* [30] and Cui *et al.* [31] used in underwater object detection.

3. RESEARCH METHOD

The proposed approach includes the dataset generation, deploying the CNN to classify the predefined classes, tracking the ball and other robot players in the environment. This session explains clearly about the CNN architecture and the algorithm that were used for the research. The tracking algorithm which is very crucial for the game is also explained.

3.1. Dataset

For the work, there was no proper dataset available. SPQR team had released open dataset for research purpose, but it did not meet the needs. So, the datasets were created by generating screenshots of the robosoccer videos as per the frame rate. Figure 1 shows some sample images from the generated dataset. Many open online converters were available for the same. Randomly images were generated and the same was annotated with the three class that the work is concerned about: nao, ball and the goal-post. Annotation is the process of labelling all the images in the dataset and for this work, it is done using a graphical tool called labeling. Some images with environments in natural and artificial illumination at the same time which makes some areas subjected to high contrast and different brightness were included in the dataset, for the work, blurred and occluded images for better training and efficient identification and over-fitting problems.



Figure 1. Random images from the dataset generated

3.2. CNN architecture and training

The CNN architecture proposed in the work is called as ROBOSOCER model. We have modified CSPDarknet-53 network, which is used as the backbone of the proposed object detector model. Feature pyramid network (FPN) is used as the neck network and the fully connected layer as the head network, in addition spatial pyramid pooling (SPP) layers are used to remove the fixed size constraint of the network. Since for the object detection, we are concerned in detecting only the three class: nao, ball and the goal-post, small network with less convolution layers is only needed. The speed of the object detection is increased because less number of layers means, reduced learnable parameters and reduced computation. The reduced number of layers may result in errors in detection, but since the number of class in the work is very less, the experimental results has proved the ROBOSOCER is an efficient model for detecting nao, ball and the goal-post. The proposed ROBOSOCER model has 50 layers, several changes to the original model was introduced so that it has an optimized performance in the robot soccer environment. The original CSPDarknet-53 Yolo has the best optimal performance when trained on complex data set such as COCO which detects 80 class of object. Since the work is comparatively simple in terms of detecting only three class, in the ROBOSOCER strided convolution is used for downscaling and maxpooling layers are replaced with strided convolution, because Sabour *et al.* [32] proposed that the maxpooling layers reduces the spatial information. The aggressive downscaling and increase in stride should decrease the accuracy of the model as per the theory, but the replacement of the max pooling has balanced the loss. Due to the limited capability of softmax function [33] rectified linear unit (ReLU) activation function is used in the proposed ROBOSOCER model.

To minimize the loss function SGDM optimization algorithm is used on the proposed ROBOSOCER Model. A mini-batch of 600 sample images from the training dataset is used for network parameters update. A learning rate of 0.001 was experimentally selected and the CNN was trained using our ball-nao-goal-post dataset divided in a ratio of 90:10 for train and test set. The training can be done until the network converges using loss function threshold or for a predefined number of iterations. In the work, the process is repeated for 6000 iteration which is 193 epochs. The test mAP@50 of 95.18% and training mAP@50 of 96.33% was achieved and average error of 4.065%, which will further reduce on increasing the iterations.

3.3. Feature extraction

It is known fact that feature extraction is a very important part of the computer vision. The feature extraction is usually done by various techniques such as principle component analysis (PCA), independent component analysis (ICA), linear discriminant analysis (LDA), locally linearly embedding (LLE), t-distributed stochastic neighbour embedding (t-SNE) and auto encoder (AE). Now days in computer vision, deep CNN itself is used for feature extraction and classification. The image enhancement through the features extracted by the CNN is better than the traditional methods and for the proposed work, Darknet-53 is used for feature extraction. In CNN, the convolution of images with the filters is used to extract various invariant features from the image that is given to the next layer. In the next layer the input features are convolved with a new filter and further more invariant features are extracted as shown in Figure 2. The process continues till we get the final feature or the output of the feature extraction CNN.

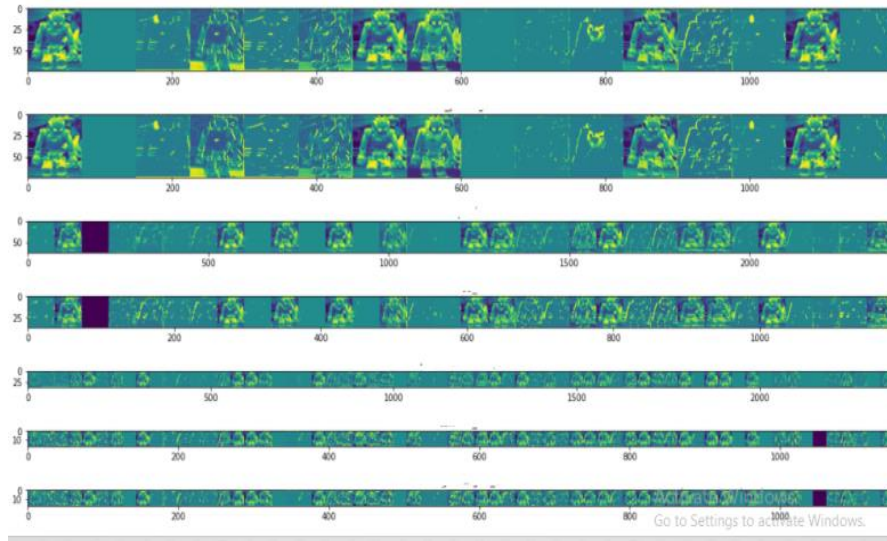


Figure 2. Feature extraction through CNN

3.4. Tracking algorithm

In the proposed work, the image capture or live input of the nao robot is not always frontal, because based on the position of the robot the front camera might not capture the image as the detector needs and many time the detector fails. But the three class: the nao, the ball and the goal-post will be blobby in the captures of the nao front camera. Yang [34] proposed the moving video image target tracking and recognition based on CNN feature selection in which feature centers are generated as per the distance matrix between feature images and feature dimensions are compressed. In the literature, there are many studies on face detection and tracking algorithms, Mliki and Hammami [35], had introduced a full automatic approach to achieve face detection and tracking with estimation in video sequences. In their work, they propose a combination between detection and tracking to overcome the various challenging problems that might occur while detecting or tracking faces.

The ROBOSOCER model and the Kalman filter are combined for tracking the ball and the nao on the robot soccer environment is depicted in Figure 3. First object detection is employed to ROBOSOCER model for detecting the ball/nao and its position in the first frame. Then the next frame is given to the model and the new position is got [36]. Now with the old measurement, the new measurement is updated. After measurement update, posteriori state estimate and posteriori estimate error covariance will be employed to do the time update to predict before state estimate and a before estimate error covariance in the next frame. The motion state variable of the Kalman filter is defined as $s = \{x, y, a, b\}$, here the x, y are the coordinates of the object and a, b is the velocity of the object in two directions. The state at each frame k and the measurement are governed by (1) and (2) respectively [37], [38]. The prediction noise and the measurement noise are represented as $n(k)$ and $m(k)$ with covariance N and M . X is the $n \times n$ matrix that denotes the priori estimate at present state from the previous state and Y is the $n \times l$ matrix that denotes the control with the state s . H is the observation matrix.

$$s_k = Xs_{k-1} + Ya_{k-1} + n_{k-1} \quad (1)$$

$$z_k = Hs_k + b_k \quad (2)$$

There are two updates,

- The Time update: A priori state estimate given in (3) and the priori estimate error covariance is given in (4),

$$\hat{s}_k^- = X\hat{s}_{k-1} + Ya_{k-1} \quad (3)$$

$$P_k^- = XP_{k-1}X^T + N \quad (4)$$

- The Measurement update: Kalman Gain,

$$K_k = P_k^- H^T (HP_k^- H^T + M)^{-1} \quad (5)$$

Posteriori state estimate from the measurement z_k ,

$$\hat{s}_k = \hat{s}_k^- + K_k(z_k + H\hat{s}_k^-)$$

posteriori estimate error coefficient:

$$P_k = (I - K_kH)P_k^-$$

3.5. Novelities in approach

In this work a novel ROBOSOCER CNN model is proposed by modifying CSP Darknet CNN. The model is trained to classify the nao, ball and the goal-post classes using transfer learning. There are no suitable open datasets are available for these kinds of studies, so creation of new dataset with annotation has been taken care. Figure 3 demonstrates the tracking algorithm.

In this proposed work, Kalman filter with time update and measurement update to track the ball and the nao is used. The proposed work explores the performance measure, intersection over union (IoU) for the three class’s detection and tracking. The proposed approach for tracking doesn’t need any object with bounding boxes in the first frame, instead it detects itself and do the tracking. The proposed model achieves an overall speed of 1.67 FPS, which will support the robosoccer application.

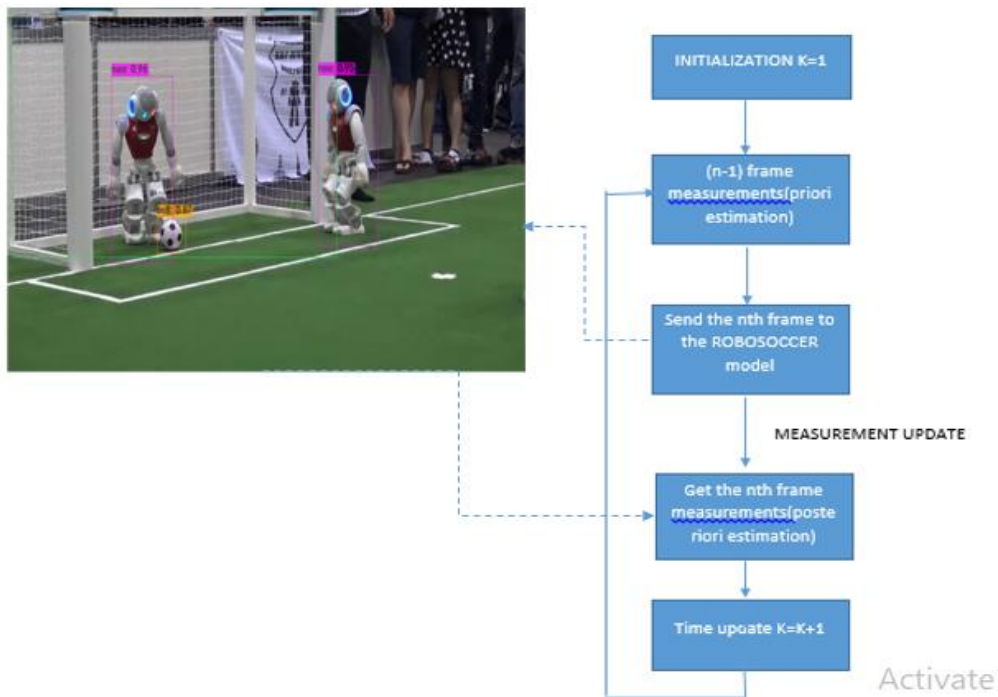


Figure 3. Tracking algorithm

4. RESULTS AND DISCUSSIONS

In the work, Nvidia Tesla P100-PCIE GPU and 25GB RAM was used for the training and testing the different models to arrive at the proposed ROBOSOCER model. The proposed model was evaluated at different IoU threshold. The minimum permissible difference between the ground truth and predicted anchor boxes to decide whether the detection is correct is defined as IoU value. When the threshold is set at small values, it means the model is evaluated leniently in terms of localization. Even very small errors in localization will make the IoU to fall drastically is the problem with this evaluation method. This issue will lead this method of evaluation to ignore the classification or the confidence error. To balance this, mean average precision (mAP) values were computed for the validation dataset at different error measure, like the Euclidean distance between the anchor box centers. The results as in Table 1, shows that the pro-posed ROBOSOCER model performs well and achieves higher detection. In the work, many tests were run, by increasing the retraining the initial layers, made the model to learn the dataset better, which results in faster convergence. In this process of retraining smaller learning rate was used. Precision-recall curve as shown in Figure 4. The PR curve for the prosed model shows that the model performance is good for use as in Figure 4(a). The PR curve was also drawn for each of the three class as in Figures 4(b)-(d). The proposed model is compared with the existing methods as seen in Table 2. Figure 5 shows the mAP and the IOU changes during the training process. This mAP metrics for evaluating the proposed model

focuses on the rigidity of the object detected rather than the anchor boxes shape errors.

Table 1. Evaluation metrics

IoU Threshold	Precision	AvgIoU (in %)	F1-Score	mAP (in %)
0.5	0.94	79.76	0.95	95.18
0.55	0.92	79.41	0.94	92.80
0.6	0.88	78.14	0.92	88.81
0.65	0.82	75.35	0.88	80.06
0.7	0.76	71.67	0.82	72.45
0.75	0.70	67.50	0.76	67.56

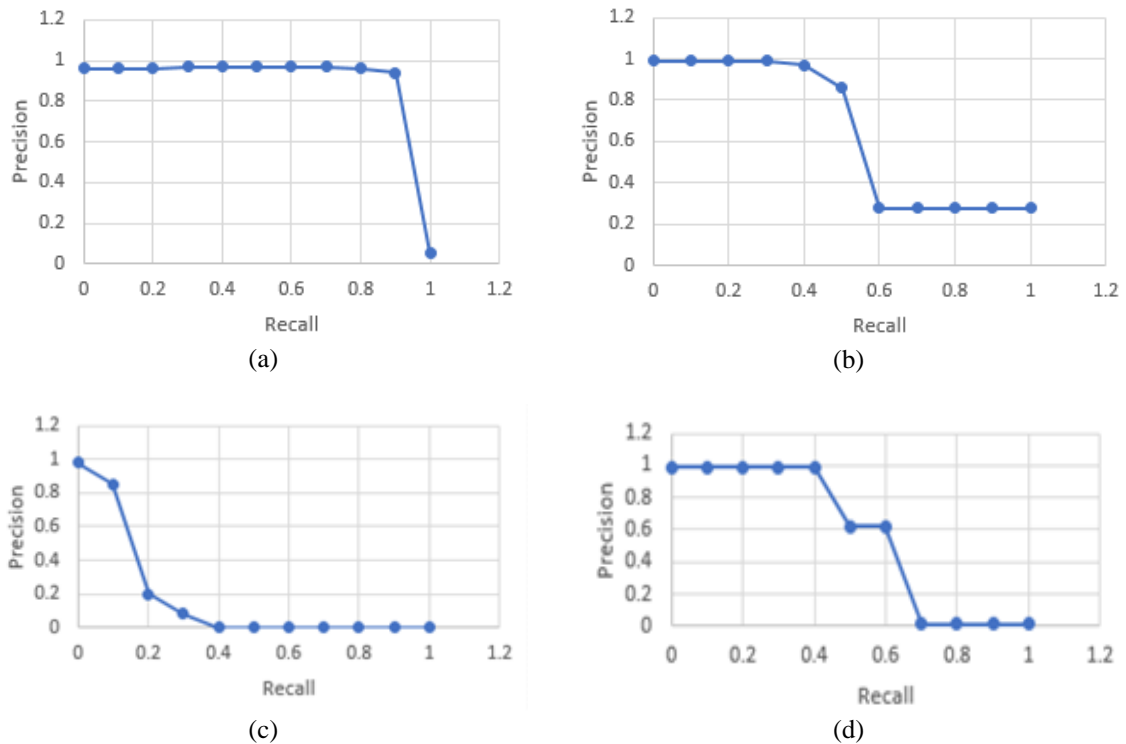


Figure 4. Precision-recall curve (a) PR curve of the ROBOSOCCEER model, (b) PR curve of the class2: Goal-post, (c) PR curve of the class1: Ball, and (d) PR curve of the Class0: Nao

Table 2. Comparison with existing object detector in robosoccer soccer environment

Model	Reference	mAP	Focused Class of Identification
MobileNet SSD v1	Yang [34]	68.81%	Small Sized League
Neural approach using CNN	Cruz <i>et al.</i> [4]	83%	Only Ball
JET-Net	Loncomilla and Solar [11]	85%	Only Nao
ROBO	Poppinga and Laue [12]	81.38%	Nao, Ball and Goal-post
BallNet	Szemenyei and Castro [13]	89.72%	Only Ball
ROBOSOCCEER (Proposed)		96.18%	Nao, Ball and Goal-Post

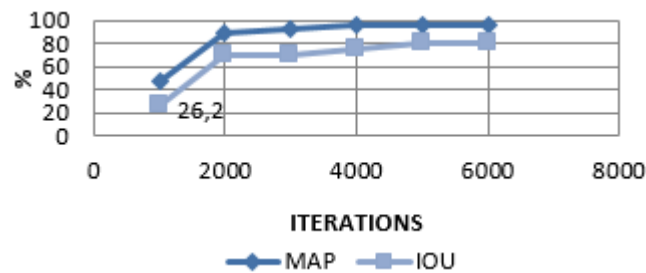


Figure 5. mAP and IOU during the training

5. CONCLUSION

In this work, a novel ROBOSOCER CNN model is proposed by modifying CSP Darknet CNN. The model is trained to classify the nao, ball and the goal-post classes using transfer learning. There are no suitable open datasets available for these kind of studies, so creation of new dataset with annotation has been taken care. In this proposed work, Kalman filter with time update and measurement update to track the ball and the nao is used. The proposed work explores the performance measure, intersection over union (IoU) for the three class's detection and tracking. The proposed approach for tracking don't need any object with bounding boxes in the first frame, instead it detects itself and do the tracking. The proposed model achieves an overall speed of 1.67 FPS, which will support the robosoccer application.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the management of Presidency University, Bangalore and Dr. C. S. Ramesh (R&D Dean, Presidency University, Bangalore) for the support extended in carrying out the research. The authors extend the heartfelt thanks to the management of MIIT, Myanmar for providing access to the robotics laboratory. <https://doi.org/10.48550/arXiv.1706.06702>





REFERENCES

- [1] C. Tabasso, V. Cichella, S. B. Mehdi, T. Marinho, and N. Hovakimyan, "Time coordination and collision avoidance using leader-follower strategies in multi-vehicle missions," *Robotics*, vol. 10, no. 1, pp. 34, 2021, doi: 10.3390/robotics10010034.
- [2] C. Liu, C. Zhou, W. Cao, F. Li, and P. Jia, "A novel design and implementation of autonomous robotic car based on ROS in indoor scenario," *Robotics*, vol. 9, no. 1, p. 19, 2020, doi: 10.3390/robotics9010019.
- [3] J. Menashe *et al.*, "Fast and precise black and white ball detection for robocup soccer," *Robot world cup2017*, pp. 45-58, Springer, Cham, 2017, doi: 10.1007/978-3-030-00308-1_4.
- [4] N. Cruz, K. Lobos-Tsunekawa, and J. Ruiz-del-Solar, "Using convolutional neural networks in robots with limited computational resources: detecting NAO robots while playing soccer," *Robot World Cup 2017*, pp. 19-30, Springer, Cham, 2017, doi: 10.48550/arXiv.1706.06702.
- [5] D. Speck, P. Barros, C. Weber, and S. Wermter, "Ball localization for robocup soccer using convolutional neural networks," *Robot World Cup 2016*, pp. 19-30, Springer, Cham, 2016, doi: 10.1007/978-3-319-68792-6_2.
- [6] J. Hoffmann, M. Jünger, and M. Löttsch, "A vision based system for goal-directed obstacle avoidance," *Robot Soccer World Cup 2004*, pp. 418-425, Springer, Berlin, Heidelberg, 2004, doi: 10.1007/978-3-540-32256-6_35.
- [7] S. Lenser and M. Veloso, "Visual sonar: Fast obstacle avoidance using monocular vision," *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)*, (Cat. No.03CH37453), vol. 1, pp. 886-891, 2003, doi: 10.1109/IROS.2003.1250741.
- [8] F. Leiva, N. Cruz, I. Bugueño, and J. Ruiz-del-Solar, "Playing soccer without colors in the spl: A convolutional neural network approach," *Robot World Cup*, pp. 122-134, Springer, Cham, 2018, doi: 10.1007/978-3-030-27544-0.
- [9] S. G. Dijk and M. M. Scheunemann, "Deep learning for semantic segmentation on minimal hardware," *Robot World Cup*, pp. 349-361, Springer, Cham, 2018, doi: 10.1007/978-3-030-27544-0-29.
- [10] A. Kukleva, M. A. Khan, H. Farazi, and S. Behnke, "Utilizing temporal information in deep convolutional network for efficient soccer ball detection and tracking," *Robot World Cup*, pp. 112-125, Springer, Cham, 2019, doi: 10.1007/978-3-030-35699-6_9.
- [11] P. Loncomilla, J. Ruiz-del-Solar, "YoloSPoC: Recognition of multiple object instances by using yolo-based proposals and deep SPoC-based descriptors," *Robot World Cup*, pp. 154-165, Springer, Cham, 2019, doi: 10.1007/978-3-030-35699-6_12.
- [12] B. Poppinga, T. Laue, "JET-Net: real-time object detection for mobile robots," *Robot World Cup*, pp. 227-240, Springer, Cham, 2019, doi: 10.1007/978-3-030-35699-6_18.
- [13] M. Szemenyei and V. E. Castro, "Fully neural object detection solutions for robot soccer," *Neural Computing and Applications*, pp. 1-14, 2021, doi: 10.1007/s00521-021-05972-1.
- [14] M. Teimouri, M. H. Delavaran, and M. Rezaei, "A real-time ball detection approach using convolutional neural networks," *Robot World Cup*, pp. 323-336, Springer, Cham, 2019, doi: 10.1007/978-3-030-35699-6_25.
- [15] T. Houliston and S. K. Chalup, "Visual mesh: Real-time object detection using constant sample density," *Robot World Cup*, pp. 45-56, Springer, Cham, 2018, doi: 10.1007/978-3-030-27544-0_4.
- [16] M. Szemenyei and E. V. Castro, "Real-time scene understanding using deep neural networks for RoboCup SPL," *Robot World Cup*, pp. 96-108, Springer, Cham, 2018, doi: 10.1007/978-3-030-27544-0.
- [17] G. C. Felbinger, P. Göttisch, P. Loth, L. Peters, and F. Wege, "Designing convolutional neural networks using a genetic approach for ball detection," *Robot World Cup*, pp. 150-161, Springer, Cham, 2018, doi: 10.1007/978-3-030-27544-0_12.
- [18] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998, doi: 10.1109/5.726791.
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017, doi: 10.1145/3065386.
- [20] D. Rethage, J. Wald, J. Sturm, N. Navab, and T. Tombari, "Fully-convolutional point networks for large-scale point clouds," *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 596-611, doi: 10.1007/978-3-030-01225-0_37.
- [21] C. Szegedy *et al.*, "Going deeper with convolutions," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.
- [22] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *International Conference on Machine Learning*, 2015, pp. 448-456.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
- [24] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251-1258, doi: 10.1109/CVPR.2017.195.





- [25] Y. Harjoseputro, I. Yuda, and K. P. Danukusumo, "MobileNets: Efficient convolutional neural network for identification of protected birds," *International Journal on Advanced Science, Engineering and Information Technology (IJASEIT)*, vol. 10, no. 6, pp. 2290-6, 2020, doi: 10.18517/ijaseit.10.6.10948.
- [26] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8697-8710, 2018, doi: 10.1109/CVPR.2018.00907.
- [27] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *International Conference on Machine Learning*, 2019, pp. 6105-6114, doi: 10.48550/arXiv.1905.11946.
- [28] S. Soltan, A. Oleinikov, M. F. Demirci, and A. Shintemirov, "Deep learning-based object classification and position estimation pipeline for potential use in robotized pick-and-place operations," *Robotics*, vol. 9, no. 3, p. 63, 2020, doi: 10.3390/robotics9030063.
- [29] M. Kanthi, T. HitendraSarma, and C. ShobaBindu, "Multi-scale 3D-convolutional neural network for hyperspectral image classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 25, no. 1, pp. 307-316, 2022, doi: 10.11591/ijeecs.v25.i1.pp307-316.
- [30] S. Murugan, T. R. G. Babu, and C. Srinivasan, "Underwater object recognition using KNN classifier," *International Journal of MC Square Scientific Research*, vol. 9, no. 3, pp. 48-52, 2017, doi: 10.20894/IJMSR.117.009.003.007.
- [31] S. Cui, Y. Zhou, Y. Wang, and L. Zhai, "Fish detection using deep learning," *Applied Computational Intelligence and Soft Computing*, 2020, doi: 10.1155/2020/3738108.
- [32] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," *Advances in Neural Information Processing Systems*, 2017.
- [33] L. Zhou, Q. Li, G. Huo, and Y. Zhou, "Image classification using biomimetic pattern recognition with convolutional neural networks features," *Computational Intelligence and Neuroscience*, 2017, doi: 10.1155/2017/3792805.
- [34] J. W. Yang, "Target tracking and recognition of a moving video image based on convolution feature selection," *International Journal of Biometrics*, vol. 13, no. 2-3, pp. 180-94, 2021, doi: 10.1504/IJBM.2021.114646.
- [35] H. Mliki and M. Hammami, "Face analysis in video: face detection and tracking with pose estimation," *International Journal of Biometrics*, vol. 10, no. 2, pp. 121-141, 2018, doi: 10.1504/IJBM.2018.10012738.
- [36] M. H. Ismail, S. A. Dawwd, and F. H. Ali, "Static hand gesture recognition of Arabic sign language by using deep CNNs," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 1, pp. 178-188, 2021, doi: 10.11591/ijeecs.v24.i1.pp178-188.
- [37] G. Welch and G. Bishop, "An introduction to the Kalman filter," pp. 127-132, 1995.
- [38] A. Yahya and A. G. E. Abdalla, "Enhancement of voip performance in MANET using fuzzy logic," *International Journal of Advances in Signal and Image Sciences*, vol. 6, no. 2, pp. 40-49, 2020, doi: 10.1016/j.jestch.2022.101099.

BIOGRAPHIES OF AUTHORS







Diana Steffi     is a research scholar in Presidency University, Bangalore. Her research interests include image/signal processing, object detection, robotics and path planning. She can be contacted at email: steffiseelan@gmail.com.



Shilpa Mehta     Completed her B.E. in Electronics Engineering in 1991 from Government Engineering College Ujjain affiliated to VVV Ujjain, where she bagged the university Gold Medal as the University Merit List Topper. She went on to do her M.E. in Digital Techniques and Instrumentation Engineering from SGSITS under DAVV Indore. She worked as a Lecturer in Electronics and Telecommunication Engineering Department in SGSITS for 6 years and then moved to Hospet, Karnataka where she worked as a Lecturer for 7 years and later as Associate Professor for 8 years. She pursued her PhD research work under JNTU during this tenure. In 2013 she relocated to Bangalore where she worked as a Professor in RITM, and SVCE. She is now working as Professor and Dean-Acadamic in the SOE of Presidency University. She can be contacted at email: shilpamehta@presidencyuniversity.in.



Dr. K. A. Venkatesh     has 3 decades of teaching experience and 4 years of Industry experience, served as CTO. He functioned as a Registrar, Dean-Research, Head of Department, Principal in couple of universities in Bangalore. He has 75 publications in journal, conference proceedings, study materials and coauthored a book on Discrete Mathematics. He has completed his Ph.D. from Alagappa University. Also authored 11 book chapters. He is a member of Systems Society of India, RMS, SIAM, USA, Academy of Discrete Mathematics. Served as Adjunct Professor at IIIT-B and He has handled international assignments as a Professor at MIIT, Mandalay at Myanmar while on deputation from IIIT Bangalore. Currently he serving as a Dean, School of Mathematics and Natural Sciences, Chanakya University, Bengaluru. He can be contacted at email: kavenkatesh@chanakyauniversity.in or ka.venkatesh@chanakyauniversity.edu.in.