Detection and tracking of football players in digital video using Convolutional Neural Networks

Ilmurat Bazarov
Kazan national Research technical University
named after A.N. Tupolev-KAI
Kazan, Russian Federation
ibazarov@stud.etu.ru

***Abstract*— The industry of professional team sports is growing with an enormous speed. To be at the top, every aspect of the game, starting from training and the game itself should be perfected. Every move and decision of players during the game is crucial, it affects overall performance and results of the team. Tracking them in the game, will give the opportunity to know the heat-map, statistics, pass coverage, fails, attacking and defending capabilities of each player. Recent advances in technology makes it possible to apply them in real life, especially in sports industry to solve different tasks. In this paper, different methods are reviewed that solved object detection and tracking problem. Detection, classification and the tracking of football players in the digital video using convolutional neural networks (CNN) is analyzed. YOLOv3 is used to test the of deep neural network's capabilities in the detection and tracking of football players in football match. Outcomes of the experiment are quite good as player detection, tracking them is achieved with 90 percent accuracy prompting to further research and experiments.**

Keywords—computer vision, CNN, Object Detection and Tracking, YOLO.

I. Introduction

The ability to visually detect and track multiple persons across a scene has been a long-standing challenge within Computer Vision and ML communities. In terms of sports analytics, automatic player detection and tracking is critical for team tactics, player activity analysis, camera planning and even enjoyment in broadcast sports video.

Compared with single and multiple objects tracking in other scenes, multiple players tracking in sports video is much more difficult due to following reasons: sports players often interact with others in a complex ways and occlusions are much more frequent and severe. All of these issues together have posed quite a great challenge to the tracking system, which requires not only reliable observations but also a sophisticated tracking strategy to make the system robust

Player detection and tracking in broadcast soccer video plays an important role in multimedia analysis [], which contributes to players observation, player tracking can help coaches each player. Variety of applications are based on the algorithm such as players running diagram, ball possession, shots, pass combinations and so on. Object detection and tracking is hot topic in computer vision research field.

II. State of the Art

As a result, many algorithms have ben proposed to deal with multi-player detection and tracking problem in sport games. One of the most popular approach is particle filter and their enhancement [11], [9], [10]. Although they solve problem of object detection and tracking, particle filter-based methods lack analysis of particle numbers. Due to different target occlusion in tracking, it is inappropriate to track players with fixed number particles. In addition, likelihood functions depend only upon single feature, and the weight computed by likelihood functions presents error, which may result in loss in tracking []. Another limitation with those approaches is that in some way or another they require a selection of hand-crafted features, which most of the time, makes their scalability to even similar sports really challenging.

Recent years, have seen an extraordinary success of Deep neural Networks. Starting with AlexNet[], a deep Convolutional Neural Network (CNN) trained to classify more than a million images, we have seen more and more models that uses DNN to perform image classification and object recognition tasks.

CNN usage has also grown a lot. In sports especially, we Matija Buriç [] et al. proposed ball and player detection using different models based on YOLOv2. As their approach almost perfectly solves objectives that they set.

In this paper, we propose usage of new YOLOv3, a state-of-the-art real-time object detection system [7] model for object detection and tracking in broadcast video.

III. YOLOv3 object detector

YOLO stands for "You Look Only Once" which describes an approach used by a single-stage network architecture that predicts the class probabilities along with corresponding bounding boxes in a single stage, as in Figure 1.

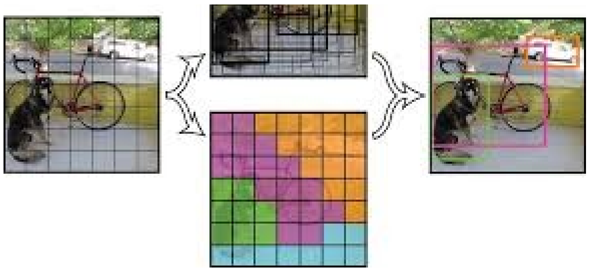

Fig1: YOLO detection pipeline: image is divided ins S x S grid where the bounding boxes are simultaneously predicted with confidence and class probability for a final decision

YOLOv3 uses a variant of Darknet [8], which has 53-layer network trained on ImageNet. Additionally, 53 layers are stacked onto it, giving us 106 layer fully convolutional underlying architecture YOLOv3. The newer architecture boasts of residual skip connections, and unsampling. The most salient feature of v3 is that it makes detections at three different scales. YOLO is fully convolutional network and its eventual output is generated by applying a 1 x 1 kernel on a feature map. In YOLO v3, the detection is done by applying 1 x 1detection kernels on feature maps of three different sizes at three different places in the network.

In addition, detections at different layers helps address the issue of detecting small objects. The unsampled layers concatenated with the previous layers help preserve the fine-grained feature which help in detecting small objects. This feature especially important in player detection as in broadcast video, view of camera constantly changes and moves with the technology of spider camera.

To sum up, YOLOv3 is fast, has at par accuracy and makes it a very powerful object detection model. Applications of object detection in variety of field, especially on sports need models to be fast and accurate. This makes YOLOv3 best model to choose on these kinds of applications either because the products real-time, where it can also be used, or where data is just big.

IV. Experimental Results

A. Dataset

In the experiment, we look at performance of YOLOv3 trained on an image dataset comprising publicly available sports images from Google Open Image Dataset. Dataset consists of only several classes: person, ball, football. For specifically training on YOLOv3, dataset extensions have been changed [5]. Dataset consists more than 1k images with variety of sizes. Most of the images are person playing football consisting ball as well. Darknet[7] developers offer pretrained weights in COCO dataset. It was decided that, pretrained weights will be used to conduct experiment faster, avoid computational power lack.

B. Object tracking

As Multi-Object tracking remains one of the problems of computer vision, in this work we used DeepSort[16] for multitracking in our video . DeepSort tracks not just the distance, velocity but also what that person looks like. DeepSort allows us to add this feature by computing deep features for every bounding box and using the similarity between deep features to also factor into the tracking logic. The whole pipeline of the work is presented in the Fig 3

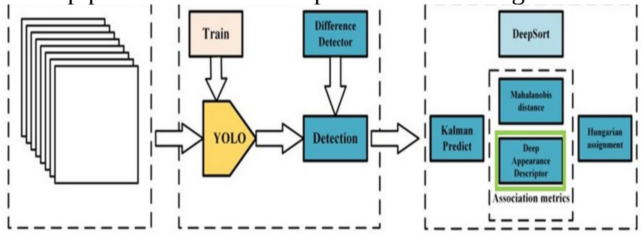

Fig 2. Object detection and tracking pipeline.

C. Evaluation

The evaluation is based on the mean average precision(mAP) criteria like one used in the PASCAL VOC 2012 competition. Evaluation results are presented on Fig 3-4. In the figure 7, we can see that camera view is directed to the player whilst ignoring the whole field. On the other hand, in Fig 8, camera is directed to the field. We can clearly see that, every player is clearly detected regardless the view of camera. In addition, in every frame players are successfully tracked. By analyzing the results, we can confidently say that Yolov3 with DeepSort successfully detects and tracks all the players in the pitch with the great accuracy.

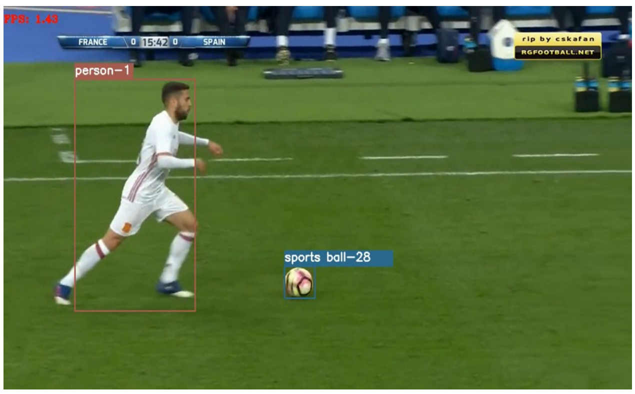

Fig 3: Detection of player and the ball in close view

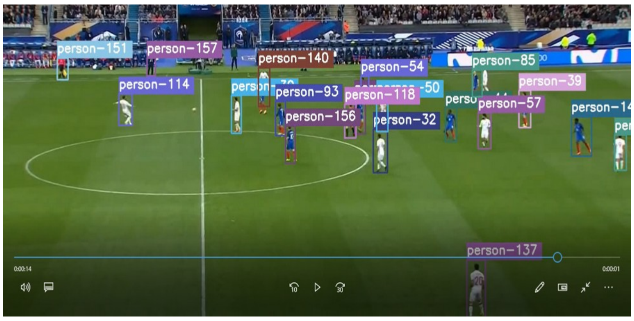

Fig 4: Detection of all the players in the field, tele broadcast view.

TABLE I. RESULTS

Model	Iterations	Mean value of loss function	Max mAP
YOLov3	4000	90	13%

Overall, YOLOv3 showed good results. It successfully solved the object detection and tracking problem. The approach we proposed showed great accuracy. As players are in constant motion and the effectiveness of YOLOv3 with DeepSort proved that, occlusion problem can be tackled and show good results in test video.

V. CONCLUSION

YOLOv3 is a good detector. It's fast, it's accurate. It solved the main task, the player detection and tracking. However, some aspects should be improved in the future. First of all, retrain model so that it detects and tracks only players in the pitch and differentiate each player by their teams.

In future, this approach could be used in real-time systems, as YOLOv3 makes it possible. Showing statistics of each player in real time, could be revolutionary in the sports industry.

ACKNOWLEDGMENT

Kazan national Research technical University named after A.N. Tupolev-KAI in collaboration with Technische Universität Ilmenau

REFERENCES

[1] Buric, M., Ivasic-Kos., Pobar, M., in Press. Object Detection in Sports Videos. 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO).

[2] Z. Zhao, P. Zheng, S. Xu and X. Wu, "Object Detection With Deep Learning: A Review," in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3212-3232, Nov. 2019, doi: 10.1109/TNNLS.2018.2876865.

[3] Buric, Matija & Pobar, Miran & Ivašić-Kos, Marina. (2019). Adapting YOLO Network for Ball and Player Detection. 845-851. 10.5220/0007582008450851.

[4] Yoon, Young & Hwang, Heesu & Choi, Yongjun & Joo, Minbeom & Oh, Hyeyoon & Park, Insun & Lee, Keon-Hee & Hwang, Jin-Ha. (2019). Analyzing Basketball Movements and Pass Relationships using Realtime Object Tracking Techniques based on Deep Learning. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2913953.

[5] X. Jiang, Z. Liu, and Y. Wang, Tracking Multiple Players in Beach Volleyball Videos, pp. 65–71. Singapore: Springer Singapore, 2016

[6] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2016

[7] Redmon, J., Farhadi A., 2018. Yolov3: An incremental improvement. arXiv preprint , VolumearXiv:1804.02767.

[8] Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Visionand Pattern Recognition (CVPR), Las Vegas, NV, pp.779-788.

[9] M. Heydari A.M.E. Moghdam, An MLP-based player detection and tracking in broadcast soccer video, in: International Conference on Robotics and Artificial Intelligence, 2012, pp. 195-199

[10] E.Y Huan, R. Li, Particle filter object tracking based on adaptive feature fusion, Comp,Sci (02) (2015) 3016-319

[11] N.Najafzadeh, M. Fotohoui, S.Kasaei, Multiple soccer players tracking in International Symposium on Artificial Intelligence and Signal processing, IEEE, 2015. Pp. 310-315

[12] P. Peng , Y. Tian , Y. Wang , J. Li , T. Huang , Robust multiple cameras pedestrian detection with multi-view bayesian network, Pattern Recognit. 48 (5) (2015) 1760–1772 .

[13] P. Baqué, F. Fleuret , P. Fua , Deep occlusion reasoning for multi-camera multi–target detection, in: Proceedings of the IEEE International Conference on Com- puter Vision, 2017, pp. 271–279 .

[14] T. Dorazio , M. Leo , A review of vision-based systems for soccer video analysis, Pattern Recognition. 43 (8) (2010) 2911–2926 .

[15] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in ICML, 2015.

[16] . Nicolai Wojke, Alex Bewley, & Dietrich Paulus. (2017). Simple Online and Realtime Tracking with a Deep Association Metric.