IC🫵 : Multiple Object Detection in the Domain of Basketball

Akayla Hackson
akayla@stanford.edu
Project Repository: https://github.com/Akayla-Hackson/ICU/tree/main

Abstract

Basketball coaches dedicate countless hours to film analysis, thoroughly scouting their opponents to gain a competitive edge. This valuable time could alternatively be invested in refining their game strategies. However, some coaching teams are constrained by limited resources and personnel, for which they struggle to conduct thorough scouting. Enter "I See You" (ICU), a tool of deep computer vision architecture designed to enhance the scouting process for coaches across all levels of basketball. By leveraging existing object detection models, which have been fine-tuned for the specific nuances of the sport of basketball, ICU transforms how coaches prepare for games. This project dives into the evaluation of several object detection models, ultimately identifying that the Yolov8 stood out, achieving an 96% accuracy rate. The integration of this model with planned future enhancements positions ICU to transform the landscape of game preparation.

I. Introduction

Basketball has significantly evolved from its humble beginnings into a multifaceted sport that transcends age, culture, and ethnicity. Its popularity has not only expanded the game's reach but also transformed it into a lucrative industry, offering numerous opportunities for recreation, education, and professional advancement. The financial stakes involved have sparked the sport's evolution, enabling a wide range of careers: from athletes perfecting their skills and physique to coaches devising intricate strategies and scouting opponents.

Gone are the days when casual practices were enough. The competitive landscape now demands great preparation and strategy. Players dedicate endless hours to honing their abilities, while coaches analyze game footage to uncover their opponents' strengths and weaknesses. This intense preparation is critical, especially when teams face the challenge of playing games with as little as 24 hours' notice, leaving little time for proper preparation.

This schedule highlights the invaluable role of "I See You" (ICU), a pioneering solution designed to revolutionize game preparation and strategy. ICU alleviates the burden on coaching staff by automating the scouting process, freeing up precious time for strategic planning and player development. Although currently focused on object detection, the full implementation of ICU will not only advance game preparation but also reshape the sport itself. Within the realm of object detection, the initial phase of the ICU project has achieved remarkable success, consistently delivering accuracies exceeding 95% within the specialized domain of basketball. Thorough testing of various models, identified Yolov8 as the standout performer, distinguishing itself as the most effective and promising model for future developments.

II. Related Work

The field of object detection has been transformed by the advancement of deep learning, allowing more sophisticated models like SSD, Faster R-CNN, YOLOv8, and Detectron2 to exist; each model with distinctive advantages in speed, accuracy, and adaptability (LeCun, Y. et al. 2015; Liu, W. et al. 2016; Ren, Shaoqing et al. 2015; Reis, Dillon, et al. 2023; Wu, Yuxin, et al. 2019). Their application extends beyond traditional boundaries to fields like sports analytics, where they enable advanced game footage analysis, player tracking, and the extraction of previously inaccessible insights, marking a great leap in our capacity to decode complex visual data.

In sports analytics, and basketball in particular, models like SSD and Faster R-CNN offer a blend of rapid processing and high-precision object detection, facilitating intricate play pattern recognition to support tactical decision-making (Liu, W. et al. 2016; Ren, Shaoqing et al. 2015). Meanwhile, advancements in models like YOLOv8 and Detectron2 further elevate the potential for analytics, with their capability to efficiently handle different scenarios, predict play outcomes, and provide evaluations of player performances and strategies (Reis, Dillon, et al. 2023; Wu, Yuxin, et al. 2019). The integration of these models into basketball analytics exemplifies the impact of deep learning in enhancing our understanding and strategic approach to the game, setting the stage for future developments.

III. Dataset

For this project, the SportsMOT dataset, curated by Cui, Yutao, et al. in 2023, served as the primary data source. This diverse dataset includes 240 high-definition videos, spanning three major sports: basketball, soccer, and volleyball. Each video, recorded in 720p resolution at 25 frames per second, has an average length of 485 frames. This project focused exclusively on basketball, for which contained 30 videos, averaging 630 frames each. The basketball videos offer a rich variety, featuring 9 women's FIBA games, 4 men's FIBA games, 6 men's NCAA games, and 13 NBA games. This diversity ensures a broad representation of stadium environments, court designs, jersey patterns, and player types.

To facilitate object detection, the dataset was formatted according to the specifications of both the Common Objects in Context (COCO) framework (Lin, Tsung-Yi, et al. 2014) and YOLO standards. The dataset was split into training, validation, and test sets, containing 20, 5, and 5 videos respectively. It is important to note that the SportsMOT dataset was originally assembled for competition purposes, meaning it lacked ground truth labels for the test split. Therefore, the division into training, validation, and test splits was executed manually, ensuring that only videos with available ground truths were selected for experimentation and analysis.

IV. Methods

Following a series of experiments with SSD, Faster R-CNN, YOLOv8, and Detectron2 models, it became evident that YOLOv8 and Detectron2 significantly surpassed their counterparts in performance. Despite their superiority, there remained a margin for enhancement; to optimize their efficacy, fine-tuning these models was used. This process, as depicted in Figure 1, refined the models to achieve peak accuracy and reliability in object detection within the SportsMOT dataset.

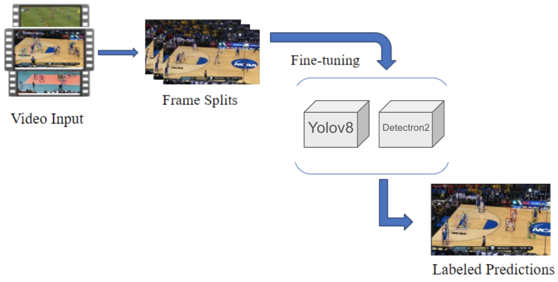

Figure 1: *The process flow from video input to predictions. Initially, individual frames are extracted from the video sequences. These frames then undergo a fine-tuning process using advanced object detection models, YOLOv8 and Detectron2. The refined models are subsequently applied to the extracted frames to generate labeled predictions during the evaluation phase on the test dataset.*

V. Metrics

In this project, various evaluation metrics were employed to thoroughly assess model performance, these metrics included Precision, Recall, Accuracy, and F1-Score. The foundation for these metrics, as shown in Equation 1, stems from the computation of the Intersection Over Union (IoU) for each predicted bounding box. This calculation facilitates the generation of a confusion matrix, providing a framework for model evaluation across the entirety of the hold out dataset. The specific methodology for calculating the IOU is shown in Equation 2. Among these metrics, Accuracy was designated as the primary metric used for this project.

$$Accuracy = \frac{TP + TN}{TP+TN+FP+FN} \quad Precision = \frac{TP}{TP+FP} \quad Recall = \frac{TP}{TP+FN} \quad F1\,Score = \frac{2*(precision*recall)}{precision + recall}$$

Where:

TP = True Positives; FP = False Positives; TN = True Negatives; FN = False Negatives

Equation 1: *Formulas for Accuracy, Precision, Recall, and F1 Score.*

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$ Where: J = Jaccard Distance; A = set 1; B = set 2

Equation 2: *Calculation for IoU*

VI. Experiments

To establish a baseline on the SportsMOT dataset, straightforward models such as SSD and Faster R-CNN were initially deployed, which attained accuracies of 3% and 7% respectively. Both models were pretrained on the COCO dataset, with the SSD model specifically being the torchvision's ssd300_vgg16 variant, enhanced with four trainable layers in its backbone. The chosen Faster R-CNN framework featured a ResNet-50-FPN backbone, incorporating three trainable layers.

The next evaluations involved the deployment of the Yolov8 and Detectron2 models, both of which had also been pretrained on the COCO dataset. Initial trials used the identification of all classes; however, given that the dataset only contained annotations for players on the court, later experiments were refined to focus solely on detecting the "person" class. Although the results from these advanced models were promising, showing significant enhancements over the simpler models, there remained a substantial margin for optimization. Detections not only comprised players on the court, but also included coaches, referees, and audience members, underscoring the necessity for targeted fine-tuning.

For the fine-tuning process, the Yolov8 model was refined using the SportsMOT training and validation splits, over 100 epochs, with a batch size of 16. Optimization was conducted via Stochastic Gradient Descent (SGD), at an optimal learning rate of 0.01 and momentum of 0.9, complemented by a weight decay of 0.0005. In contrast, the Detectron2 model was trained for 73 epochs, with adjustments made for its learning rate (0.001) and batch size (16), employing an SGD optimizer with identical momentum, tailored to accommodate the extensive computational demands relative to Yolov8.

Post fine-tuning, when evaluated on the hold out set, both models exhibited great performance, with accuracies surpassing 95%. A comprehensive comparison of all models tested is detailed in Table 1, highlighting Yolov8's superior performance over Detectron2.

See appendix for visualizations of predictions.

Model	Precision	Recall	Accuracy	F1 Score
SSD	0.11	0.04	0.03	0.06
Faster R-CNN	0.08	0.39	0.07	0.13
Yolov8	0.52	0.77	0.45	0.62
Dectron2	0.10	0.97	0.10	0.18
FT Yolov8	0.97	**0.98**	**0.96**	**0.98**
FT Detectron2	**0.99**	0.96	0.95	**0.98**

Table 1: *Comparison of all the models, excluding the models permitting all classes to be identified. Note FT is short for fine-tuned.*

VII. Analysis

Fine-tuning significantly enhanced the performance of both the Yolov8 and Detectron2 models on the SportsMOT dataset due to its ability to adapt the pretrained models, which were originally generalized for a broad range of classes in the COCO dataset, to the specific nuances and requirements of detecting players on a basketball court. This process effectively leveraged the rich feature representations learned from the large and diverse COCO dataset, enabling the models to refine these features for the specialized task at hand with far less data than would be required to train a model from scratch. Despite Detectron2 being fine-tuned for approximately 30 fewer epochs compared to Yolov8, leading to a marginally lower performance, Yolov8 emerged as the preferred model. One could argue that if the Detectron2 model was trained on the same amount of epochs as the Yolov8 model then its accuracy could be higher, therefore, changing the outcome of the preferred model. However, the preference was not solely based on Yolov8’s slight edge in accuracy but also its efficiency in training time. Where the Detectron2 model took, substantially, much longer to train. Yolov8's ability to achieve high performance in a significantly shorter duration made it especially attractive for the project, showing the importance of not just model accuracy but also operational efficiency.

VIII. Conclusion

The ICU project marks the beginning of an innovative application of deep computer vision in basketball coaching, with its current focus on player detection through the fine-tuning of models like Yolov8, which has already achieved a 96% accuracy. This foundational step lays the groundwork for a much broader vision. As ICU stands now, it's a promising prototype that hints at the future of game preparation but falls short of being a comprehensive tool for coaches. Future work includes exploring data augmentation, extending training durations for improved accuracy, and advancing to player tracking. These future steps are to transform ICU from a concept into a cornerstone of basketball strategy development, offering coaches a data-driven lens in analyzing and strategizing against opponents. The journey of ICU, while only at its beginning stage, promises to revolutionize basketball coaching by offering deep insights that were previously inaccessible.

Contributions:

This project was solely worked on by myself, Akayla Hackson. However, I did receive feedback and suggestions for which I implemented from Nick Harber, Merve Cerit, and the students in the workshop groups.

Citations

LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444 (2015). https://doi.org/10.1038/nature14539

Liu, W. et al. (2016). SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9905. Springer, Cham. https://doi.org/10.1007/978-3-319-46448-0_2

Ren, Shaoqing et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2015): 1137-1149.

Reis, Dillon, et al. "Real-Time Flying Object Detection with YOLOv8." arXiv preprint arXiv:2305.09972 (2023).

Wu, Yuxin, et al. "Detectron2." 2019, github.com/facebookresearch/detectron2.

Cui, Yutao, et al. "SportsMOT: A Large Multi-Object Tracking Dataset in Multiple Sports Scenes." arXiv preprint arXiv:2304.05170 (2023).

Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014.

APPENDIX

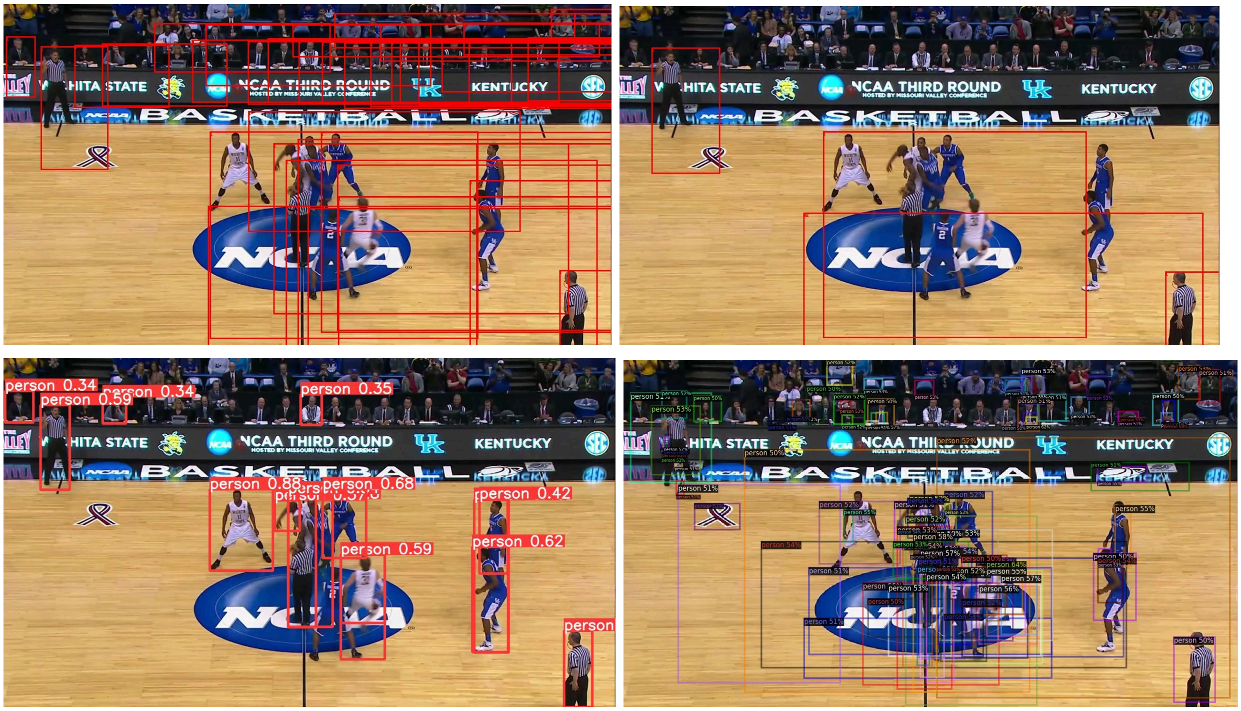

Figure 2: The model's prediction outputs without fine-tuning.
Top Left - Faster RCNN; Top Right - SSD; Bottom Left - Yolov8; Bottom Right - Detectron2
Note: The predictions are only using the "person" class

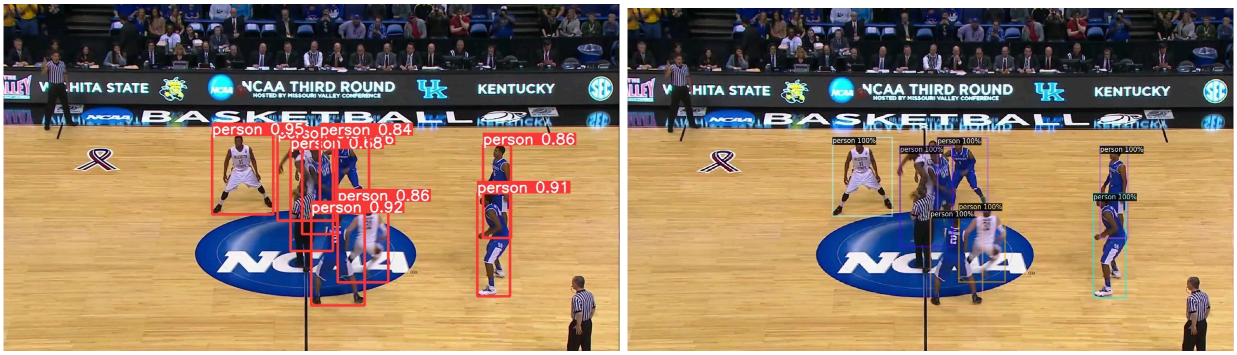

Figure 3: The model's prediction outputs WITH fine-tuning.
Left - Yolov8; Right - Detectron2
Note: The predictions are only using the "person" class