VolleyVision

BENAROCHE Eyal, MANUS Raphael

I. Introduction

During our 3rd year at Ecole polytechnique, we took the class *INF573 - Image Analysis and Computer Vision*. The goal of computer vision is to compute properties of the world from digital images. Challenges in this field encompass 3D shape recovery, motion estimation, and object recognition, all through the analysis of images and videos. This course served as an introduction to image analysis and computer vision, covering topics such as feature detection, motion estimation, image mosaicing, and 3D vision. We were tasked with a free of choice project related to computer vision, and here is our Volley Vision Project.

A. Motivation

Integrating computer vision into volleyball presents a compelling opportunity to enhance player performance analysis and automate statistical tracking, even in a sport where financial investment may be limited compared to more mainstream sports like football. As one of us is a dedicated volleyball player and enthusiast seeking to improve personal skills, leveraging computer vision technology can provide valuable insights into his game.

In the pursuit of enhancement, teams frequently capture footage of their games from the rear of the court for subsequent review. Although this practice is crucial for progress, the game review process can be arduous and time-consuming. Volleyball matches, in particular, are characterized by substantial downtime between points, requiring patience as one waits for significant moments to analyze on screen—from the conclusion of one point to the commencement of the next serve. With game recordings extending beyond two hours, identifying key moments or compiling statistics becomes a demanding and time-intensive task.

Assuming footage from the back of the court, a computer vision model can be trained to recognize and track various actions in a volleyball match, such as serves, spikes, blocks, and digs. This automated approach allows for a comprehensive understanding of individual and team performance over time.

Automated statistics generated through computer vision can offer real-time feedback on your strengths and areas for improvement. This data-driven approach not only supports your personal development but also contributes to a more sophisticated understanding of the sport itself. As the volleyball community embraces technology, the application of computer vision in the sport could lead to more advanced coaching techniques, enhanced player development, and a richer overall experience for players and fans alike.

II. Volleyball Tracking

Our exploration into volleyball tracking started with a comprehensive review of existing literature and projects in the domain. This preliminary research led us to identify a GitHub repository named VolleyVision [shukkkur, 2023], which aligned closely with our objectives of volleyball tracking, field tracking, and player action classification. Recognizing the potential of VolleyVision as a foundational base, we decided to not only implement but also enhance its methodologies.

The initial phase focused on the volleyball tracking component as outlined in VolleyVision. This involved the utilization of YOLOv7, known for its versatility in performing tasks like classification, posing, and tracking, specifically for instance segmentation of the volleyball. Following the further development suggestions from VolleyVision, we transitioned to training YOLOv8 on the same dataset provided by the repository's author to harness YOLOv8's improved performance for ball

detection.

In order to visualizing the volleyball's trajectory, we overlaid the last seven detected positions of the ball onto the current frame. This was achieved by rendering these positions as overlapping yellow circles, thereby creating a dynamic and informative visual representation of the ball's movement.

Subsequent to the successful training of YOLOv8, we turned our attention to optimizing the inference process. A key modification involved the direct utilization of OpenCV's VideoWriter output as an in-memory image, as opposed to the less efficient method of saving each frame as a JPEG. This change significantly expedited the video analysis process. Additionally, we addressed a technical challenge concerning color format conversion. The initial frame processing in BGR format was hindering accurate inference, which we overcame by transitioning to RGB format.

An important aspect of our optimization efforts was the streamlining of the original script. We achieved a considerable reduction in complexity, bringing down the script from 600 lines of code with over 10 imports to a bit more than 200 lines with only 7 imports. This refinement not only enhanced the efficiency of the script but also improved its readability and maintainability.

Our efforts resulted in more accurate ball tracking, streamlined processing, and a more efficient and accessible codebase.

III. Field Tracking

Our initial approach to field tracking in volleyball match analysis involved the implementation of the VolleyVision solution available on GitHub. The field tracking in this solution utilized a semantic segmentation model provided by Roboflow, a platform offering tools for dataset creation and model training. The model, whose specific architecture was not disclosed by Roboflow, was trained on a custom dataset comprising 36 images of volleyball courts. These images were extracted from video footage of both professional and amateur volleyball matches.

The semantic segmentation model generated a semantic mask of the volleyball court, which was then processed using contour detection techniques from Opencv. The visible field was subsequently approximated as a polygon, which was overlayed on the output image. While this method showed satisfactory performance on images and videos from the original dataset, we encountered limitations when applied to new images, particularly those captured from uncommon angles or featuring volleyball courts with varying color schemes. This issue was particularly pronounced in amateur settings, where the courts are often demarcated by colored lines on a uniformly colored floor, complicating the distinction between the court and its surroundings in lower-quality videos.

To address these challenges, we shifted our approach to training and running our own model locally, using YOLOv8 for instance segmentation. YOLOv8 was chosen for its capability in segmentation. We replicated the original dataset in Roboflow and annotated it for YOLOv8 training. Despite achieving comparable performance to the original online model, our initial efforts with YOLOv8 did not significantly improve field detection from diverse viewpoints.

To enhance the model's robustness, we expanded our dataset in Roboflow, doubling the number of images and ensuring a wide variety of field types and angles were included. Roboflow's data augmentation capabilities were then employed, leveraging techniques such as gamma, angle, and skew variations to enrich the dataset effectively. This improved dataset yielded significantly better performance on new inputs and allowed for faster processing since we could run the model directly on our local computers.

Integrating our trained YOLOv8 model into the original script required several adaptations. Firstly, YOLOv8 necessitates input images to be resized to 640x640 pixels. Initial attempts to deal with this requirement by resizing only during inference resulted in skewed outputs, which was resolved

by resizing to the required input dimensions. Secondly, to enhance video processing efficiency, we modified the script exactly as with volleyball tracking to improve on performance and lisibility.

There are possibilities for further enhancing field tracking. Temporal filtering, which involves using information from previous frames to improve the stability and accuracy of detection in subsequent frames, could be a valuable addition. This approach would be particularly effective in volleyball match analysis, where camera movements are typically minimal or inexistent. Additionally, inspired by the methodology outlined in "Sports Camera Calibration via Synthetic Data" [Chen and Little, 2018], there is potential for 3D camera positioning. This could be achieved by generating a dataset correlating camera poses with corresponding field edge images, enabling the calculation of a homography matrix to produce a bird's eye view of the volleyball court. Such a perspective, combined with player bounding boxes, would facilitate an estimation of the players' 3D positions, offering valuable insights for match analysis.

In conclusion, our work in volleyball match field tracking demonstrates the importance of dataset diversity, the effectiveness of local model training and optimization, and the potential for advanced techniques like temporal filtering and 3D positioning in sports analytics.

IV. Player Action Classification

In our volleyball match analysis project, a significant component was the player action classification, which involved the implementation of two distinct models focusing on player and action segmentation.

The player segmentation model was tasked with identifying players on the court. However, this model faced challenges in differentiating players from non-players like coaches and referees, particularly due to its sensitivity to occlusions. To address these issues, we could consider limiting the model to recognize no more than 12 players at any given time and using the court's layout or the position of the ball to enhance accuracy. Despite these strategies, the problem of misidentification remained complex. Another limitation was the inability to track individual players, which restricted the generation of player-specific statistics. We theorized that analyzing overlaps in bounding boxes across consecutive frames could enable consistent player tracking. The dataset for this model, comprising 206 images, was not retrained for our specific requirements as we used the provided YOLOv8 trained models from the repository, particularly lacking in rear-view court perspectives, which are crucial for our analysis.

Our second focus was on action segmentation, using the YOLOv8 model from the repository which uses 13 different classes. This model, despite being trained on an extensive dataset of 13,920 images, faced significant challenges. Determining the exact start and end of an action was difficult, especially when actions spanned only a few frames. The VolleyVision github considered validating an action if it appeared consistently over three frames, but we did not implemented it due to the model's inadequate performance on diverse video inputs and its failure to detect rapid actions. A notable issue was the model's inability to distinguish between similar actions, such as a 'set' and a 'pass', and the lack of recognition for critical actions like 'feints'. Additionally, the model was hindered by occlusions, echoing the challenges faced by the player segmentation model.

Among all the components of our project, the action segmentation model necessitates the most extensive further development. Initially conceptualized as a three-person project, resource constraints led us to focus on an overview rather than detailed refinement. Future work could involve extensive retraining of the models, particularly focusing on rear-view perspectives for player segmentation and enhancing the action model to distinguish between nuanced and rapid actions.

In conclusion, while the player action classification segment of our project presented significant challenges, it also highlighted areas for potential

improvements and future research directions, underscoring the complexity and dynamism inherent in volleyball match analysis.

V. 3D Scene Reconstruction

In our project, we aimed to implement 3D scene reconstruction using the SLAMHR Python implementation, based on the principles outlined in "Decoupling Human and Camera Motion from Videos in the Wild" [Ye et al., 2023]. This method offers a novel approach to reconstruct global human trajectories from videos by effectively separating camera motion from human movements, allowing for accurate placement of players within a unified world coordinate frame.

We believed this technology would be highly beneficial in analyzing volleyball matches, enabling the visualization of players in 3D meshes and their relative spatial positioning. Such a system could potentially identify specific volleyball actions through body mesh analysis and provide various 3D perspectives for in-depth action review.

However, the complexity of implementing SLAMHR proved to be a significant challenge, and we were unable to successfully integrate this method into our analysis framework within our project timeline. This implementation difficulty limited our ability to achieve a full 3D recreation of volleyball matches, which would have included both player movements and ball tracking in a cohesive 3D environment.

Despite this setback, the potential of this approach for advanced sports analysis remains clear, suggesting promising directions for future research in volleyball match analysis and beyond.

VI. Conclusion

In concluding our Volley Vision Project, we reflect on the achievements and future directions of our volleyball match analysis using computer vision.

The combined model we developed operates satisfactorily, producing visually appealing results. However, from a scientific standpoint, the current state of the project yields limited insights. The model, while effective in its basic functionality, requires further development to achieve a deeper, more analytical utility.

Looking ahead, the next crucial step involves the creation of a statistical interface. This interface would serve as a tool for extracting and presenting comprehensive data derived from the match footage. By transforming raw analytical outputs into accessible statistics, this interface would enable both players and coaches to gain meaningful insights into performance metrics, team dynamics, and individual player contributions.

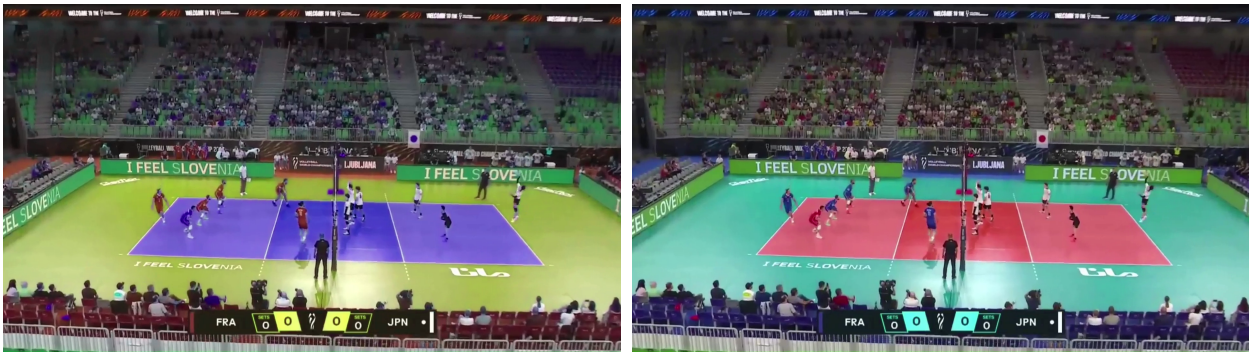
Another significant advancement would be the implementation of the discussed 3D reconstruction. This development would elevate the project from two-dimensional analysis to a more holistic three-dimensional perspective. With 3D reconstruction, we can offer a more nuanced understanding of player movements, interactions, and spatial dynamics on the court. Such a feature would not only enhance the visual aspect of the analysis but also provide a deeper level of strategic understanding, potentially influencing training methods and game strategies.

In essence, while our project has laid a solid foundation in applying computer vision to volleyball match analysis, its true potential lies in these future developments. By integrating a statistical interface and 3D reconstruction, we can transform our project from a primarily visual tool into a comprehensive analytical platform.

Our project is now available here https://github.com/Shaamallow/VolleyVision

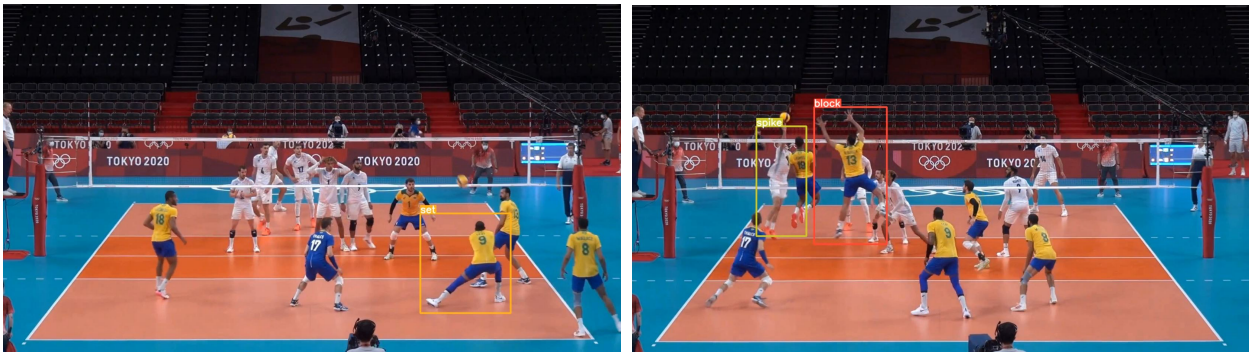
VII. Annexes

Fig. 1. Combination of all models



Video frame with wrong color formatting Video frame with correct color formatting

Fig. 2. Illustration of the problem encountered with video processing



mis-classification Correct classification

Fig. 3. Limitation of the action classification YOLOv8 model

Fig. 4. Overidentification of the player YOLOv8 model

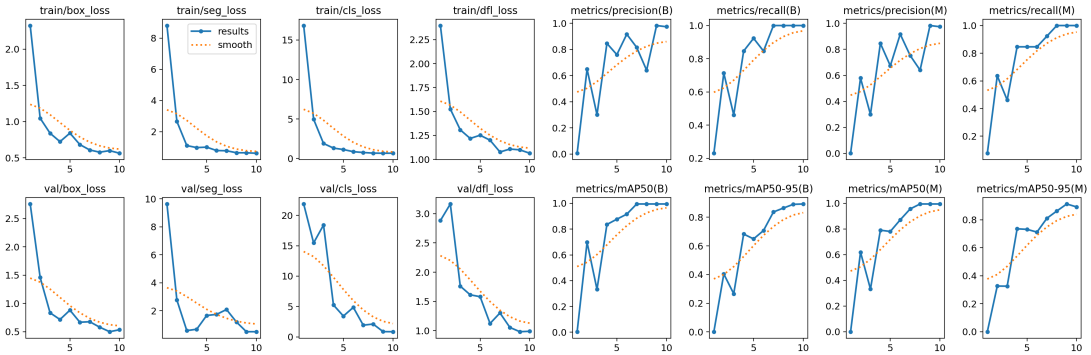


Fig. 5. Training results of the field dataset

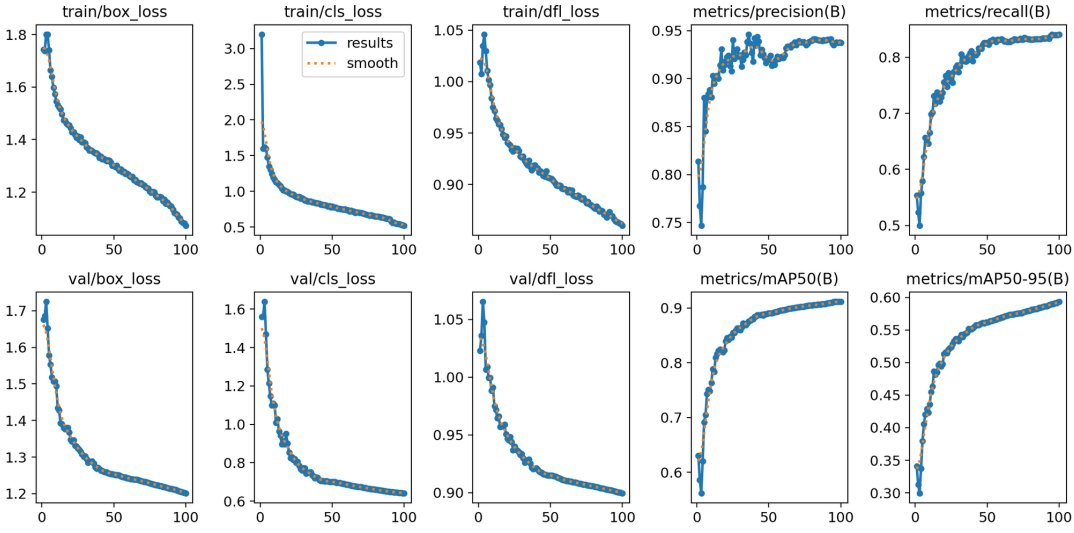


Fig. 6. Training results of the volleyball dataset

REFERENCES

[Chen and Little, 2018] Chen, J. and Little, J. J. (2018). Sports camera calibration via synthetic data.
[shukkkur, 2023] shukkkur (2023). Volleyvision. https://github.com/shukkkur/VolleyVision.
[Ye et al., 2023] Ye, V., Pavlakos, G., Malik, J., and Kanazawa, A. (2023). Decoupling human and camera motion from videos in the wild.