The impact of data augmentation on YOLOv8 performance for object detection in basketball games

Kyana Marckx

January 25, 2026

Abstract

Object detection plays an increasingly important role in sports analytics, enabling the automatic identification and analysis of players and game-related elements in video footage. In dynamic sports such as basketball, detection models must operate accurately under challenging conditions, including rapid motion, occlusions and varying lighting. Although YOLO-based detectors are widely used for real-time applications, the influence of different data augmentation strategies on their performance in sports contexts remains insufficiently explored.

This study investigates the impact of several image data augmentation techniques on the performance of a YOLOv8-based object detection model trained on basketball game footage. Five training configurations are evaluated: a non-augmented baseline, geometric augmentation, color-based augmentation, kernel-based augmentation and a combined augmentation strategy. All models are trained under identical conditions to ensure fair comparison. Performance is assessed using standard detection metrics, per-class analysis and inference speed.

The results show that all augmentation strategies improve detection performance compared to the baseline model. The combined augmentation approach achieves the highest overall accuracy and recall while maintaining real-time inference speed. In particular, improvements are observed for small and fast-moving objects such as the basketball and shot clock. These findings demonstrate that carefully selected and combined data augmentation techniques can significantly enhance YOLOv8 performance in basketball detection tasks without compromising efficiency.

Keywords – data augmentation, YOLOv8, object detection, basketball

1 Introduction

Object detection has become a fundamental component of modern computer vision systems, enabling the automatic localization and classification of objects in images and videos. In recent years, advances in deep learning have significantly improved detection accuracy while maintaining high processing speeds, making real-time applications increasingly feasible. One domain where these developments are particularly relevant is sports analytics, where object detection can be used to identify and track players, referees and game-related elements in live video streams.

Among the various object detection architectures, the You Only Look Once (YOLO) family has emerged as a popular choice due to its favorable balance between accuracy and computational efficiency. YOLO models perform detection in a single forward pass, allowing them to operate at real-time speeds. YOLOv8, developed by Ultralytics[1], further improves upon earlier versions by introducing architectural and optimization changes that enhance performance, especially for small and fast-moving objects. As a result, YOLOv8 is well suited for analyzing dynamic sports scenes such as basketball games.

Despite these advances, training robust object detection models remains challenging, particularly when working with limited or domain-specific datasets. Basketball broadcasts present several difficulties, including rapid player movements, frequent occlusions, variable lighting conditions and small objects such as the ball or shot clock. When training data does not sufficiently capture this variability, models are prone to overfitting and poor generalization to unseen footage.

Data augmentation is a commonly used technique to address this problem by artificially increasing the diversity of training data through transformations such as geometric distortions, color modifications and kernel filters. While data augmentation is widely adopted in object detection pipelines, its effectiveness

[1] https://www.ultralytics.com

strongly depends on the type of augmentation applied and the target application domain. In the context of sports analytics, and basketball in particular, there is limited empirical analysis on how different augmentation strategies influence the performance of modern YOLO-based detectors.

This study claims to fill this gap by systematically evaluating the impact of several image data augmentation categories on the performance of a YOLOv8-based object detection model trained on basketball game footage. Specifically, geometric, color-based, kernel-based and combined augmentation strategies are compared agains a non-augmented baseline. All experiments are conducted under identical training conditions to ensure fair comparisons, and performance is evaluated using both global and per-class metrics as well as inference speed.

The contributions of this work are threefold. First, it provides a controlled comparison of commonly used image augmentation strategies in the context of basketball object detection. Second, it offers a detailed per-class analysis that highlights how augmentation affects small and challenging objects. Finally, it demonstrates that carefully selected augmentation strategies can improve detection accuracy without compromising real-time performance, making the findings directly applicable to practical sports analytics systems.

2 Literature

2.1 Object detection and the YOLO family

2.1.1 Object detection

Object detection is a fundamental task within computer vision that aims to both localize and classify objects in images or video frames [4]. Unlike image classification, object detection requires models to predict bounding boxes in addition to class labels, which makes it a more complex problem. Recent advances in deep learning have led to significant improvements in accuracy and robustness, enabling object detection models to be deployed in real-time applications.

2.1.2 The YOLO family

The *You Only Look Once (YOLO)* family of architectures is among the most widely used approaches for real-time object detection. Since the introduction of the original YOLO model by Redmon et al. [9], successive versions have focused on improving detection accuracy, inference speed and architectural stability. A key advantage of YOLO-based models is their single-stage detection pipeline, which allows them to process images efficiently while maintaining competitive performance.

YOLOv8, introduced by Ultralytics in 2023 [2], represents a substantial update compared to earlier versions. It features an improved backbone, anchor-free detection heads and updated loss functions, leading to improved performance, particularly for small and fast-moving objects. Due to its high inference speed, YOLOv8 is well suited for real-time applications such as surveillance systems, autonomous systems and sports analytics.

Recent comparative studies further support the effectiveness of YOLOv8. Research comparing YOLOv5, YOLOv8 and YOLOv9 in combat sport scenarios showed that YOLOv8 offers a strong balance between detection accuracy and computational efficiency across object detection, tracking and human action recognition tasks [8]. These findings reinforce the suitability of YOLOv8 for dynamic sports environments.

At the time of this research (January 2026), YOLOv12 represents the most recent officially released version, with further versions already announced. Nevertheless, YOLOv8 remains a stable and well-supported model, making it a suitable choice for systematic experimental evaluation.

2.2 Data augmentation

A recurring challenge in deep learning-based computer vision systems is the limited availability of diverse and representative training data. Data augmentation addresses this limitation by artificially increasing dataset diversity through transformations applied to existing images. These techniques have been shown to reduce overfitting and improve the generalization capabilities of deep learning models [3].

2.2.1 Augmented data versus synthetic data

Although both augmented data and synthetic data aim to increase dataset size and diversity, they differ fundamentally in their approach. Augmented data is created by applying relatively simple transformations to real images, such as rotation, cropping or brightness adjustments. In contrast, synthetic data is fully generated using techniques such as generative models or 3D simulations. Synthetic data enables the creation of controlled and rare scenarios that may be difficult to capture in real-world settings and is particularly valuable in domains with privacy or data availability constraints, such as medical imaging [5].

2.2.2 Types of data augmentation

Data augmentation techniques can be categorized based on the data modality to which they are applied; most commonly audio, text and image data [1]. Since this research focuses on visual data from basketball broadcasts, image-based augmentation techniques are of primary relevance.

Image augmentation methods can be grouped into several categories:

- **Geometric transformations** such as *crop, flip, scale, rotate, translate* and *stretch* improve robustness to camera viewpoint changes and object orientation.
- **Color space transformations** including adjustments to *brightness, contrast* and *saturation* simulate varying lighting conditions commonly encountered in broadcast footage.
- **Kernel-based filters** such as *blur* or *sharpen* help models become more robust to *motion blur* and focus variations.
- **Random erasing** removes random regions of an image to simulate partial occlusions.
- **Image mixing techniques** combine multiple images to increase contextual diversity.

These forms of augmentation are particularly relevant for sports analytics, where fast movement and changing lighting conditions frequently occur.

2.3 Object detection in sports and basketball analytics

Object detection has become an essential component of modern sports analytics, enabling automated player detection, referee identification, ball tracking and the extraction of game statistics. Due to their real-time processing capabilities, YOLO-based models are frequently adopted in sports-related computer vision systems.

Several studies have demonstrated the effectiveness of YOLO architectures in sports contexts. In football analytics, systems combining YOLOv5 with DeepSORT have achieved strong performance in player detection and tracking tasks [10]. In basketball-specific research, YOLOv8 has been successfully applied to detect players, the ball and court elements, demonstrating its suitability for complex basketball scenes [6].

Despite these advances, object detection in sports remains a challenging task. Sports footage is characterized by rapid movement, frequent camera angle changes, varying illumination and significant occlusions between players. Ball detection poses an additional challenge due to the small size and high velocity of the ball, as highlighted in recent reviews of ball detection techniques in sports [7]. These factors can negatively affect detection performance, particularly when models are trained on limited or insufficiently diverse datasets.

2.4 Research gap and motivation

The existing literature clearly shows that YOLO-based object detection models achieve strong performance in real-time and sports-related applications. Additionally, data augmentation is widely recognized as an effective strategy for improving model generalization and robustness. However, the specific impact of different image augmentation strategies on YOLOv8 performance in basketball contexts has not been studied in depth.

Most existing research focuses on model architecture comparisons or tracking frameworks, while data augmentation is often treated as a secondary component. This research addresses this gap by systematically evaluating how different categories of image augmentation influence YOLOv8 detection performance for basketball-specific objects, including players, referees, the ball, the hoop and scoreboard elements. The goal is to provide practical insights into building more robust and reliable real-time basketball detection systems.

3 Research questions

Based on the preceding literature review and problem definition, this study focuses on analyzing the impact of different image data augmentation techniques on the performance of the YOLOv8 model for object detection in basketball game footage.

3.1 Main research question

What is the effect of different image data augmentation techniques on the detection performance of YOLOv8 when identifying objects such players, the basketball, the shot clock and other relevant elements in basketball games?

3.2 Sub-research questions

- To what extent do different image augmentation techniques (such as geometric transformations, color space adjustments and kernel-based filters) improve the detection performance of YOLOv8 compared to a model trained on a non-augmented dataset?
- Which augmentation techniques have the greatest impact on the detection of small and fast-moving objects such as the basketball?
- Are certain object classes (e.g. players, ball, shot clock) more sensitive to specific augmentation techniques than others?

3.3 Hypotheses

- A YOLOv8 model trained using data augmentation achieves a significantly higher mean Average Precision (mAP) compared to a model trained without data augmentation.
- Models trained on a combination of multiple augmentation techniques outperform models trained on individual augmentation categories. Among individual augmentation types, kernel-based filters are expected to have the strongest positive impact on detection performance.
- The effectiveness of data augmentation varies across object classes, with smaller and faster-moving objects such as the basketball benefiting more from augmentations such as blur, rotation and brightness adjustments.

4 Methodology

4.1 Overview

This study investigates the impact of different image data augmentation strategies on the performance of a YOLOv8-based object detection system for basketball game footage. The methodology is structured as a sequence of experimental steps, each implemented in a separate notebook. The workflow starts with dataset preparation and augmentation, followed by model training, evaluation and quantitative comparison of results across multiple experimental settings. The complete experimental pipeline is publicly available in a GitHub repository[2].

All experiments are conducted on an NVIDIA GeForce RTX 4060 Laptop GPU using the same base model architecture and training configuration to ensure fair and reproducible comparisons.

4.2 Dataset and preprocessing

The experiments are based on a publicly available basketball object detection dataset[3] containing annotated images of players, referees, the ball, the hoop and scoreboard elements (period, shot clock, team name, team points and time remaining). An example of an original image from this dataset is shown in Figure 1, and an annotated image is shown in Figure 2. Before training, a preprocessing pipeline was applied uniformly across the original dataset since it serves as the baseline for all augmentations. This pipeline includes automatic image orientation correction and resizing to a fixed resolution of 640x640 pixels with white padding to preserve the aspect ratio of the original image. No augmentations were applied at this stage to the original dataset.

The dataset is split into training (120 images, 70%), validation (25 images, 15%) and test (25 images, 15%) sets using a consistent split across all experiments to ensure comparability of evaluation results.

[2] https://github.com/kyanamarckx/data-augmentation-detection-basketball

[3] https://app.roboflow.com/kiejaana/basketball-players-fy4c2-dqb1n/1

4.3 Creation of augmented datasets

Three augmented datasets were created by applying specific categories of image augmentation techniques to the original dataset. Each augmented dataset focuses on a single augmentation category to allow for isolated evaluation of its effect. The considered augmentation groups include geometric transformations, color space adjustments and kernel-based filters.

Care is taken to avoid duplication of original images across datasets. Each augmented dataset contains the original images alongside their augmented counterparts, ensuring that the total dataset size increases while preserving the original data distribution.

4.4 YOLOv8 training on baseline dataset

To establish a reference point, a baseline YOLOv8[4] model was trained using the original, non-augmented dataset. The lightweight YOLOv8n variant was selected to prioritize real-time performance while maintaining sufficient detection accuracy.

The baseline model is trained using fixed hyperparameters, including image size, batch size, number of epochs and random seed. During baseline training, built-in data augmentation within YOLOv8 is disabled to isolate the effect of external augmentation strategies introduced later in the workflow.

The trained baseline model is evaluated on the held-out test set, and standard object detection metrics such as mean Average Precision (mAP50 and mAP50-95), precision and recall are recorded.

4.5 YOLOv8 training on augmented datasets

Separate YOLOv8 models are trained on each augmented dataset using the same training configuration as the baseline experiment. This ensures that any observed performance differences can be attributed to the applied augmentation strategy rather than changes in model architecture or hyperparameters.

For each augmented dataset, the trained model is evaluated on the same test set used for the baseline model. Evaluation metrics are collected consistently across all experiments.

4.6 Creation of combined dataset

In addition to training on individual augmentation types, a combined dataset is constructed by merging the different augmented datasets into a single training set. Duplicate images are removed to prevent overrepresentation of original samples. This combined dataset aims to capture the complementary benefits of multiple augmentation strategies within a single training process.

4.7 YOLOv8 training on combined dataset

Another YOLOv8 model is trained on this combined dataset using the same training configuration and evaluated on the test set.

4.8 Evaluation metrics

Model performance is assessed using standard object detection metrics, including mean Average Precision at an Intersection over Union (IoU) threshold of 0.5 (mAP50), mean Average Precision averaged over thresholds from 0.5 to 0.95 (mAP50-95), precision and recall. In addition to accuracy-based metrics, inference-related timing metrics are recorded to evaluate the suitability of each model for real-time deployment.

All metrics are computed using the YOLOv8 validation framework to ensure consistency across experiments.

4.9 Results aggregation and comparison

To facilitate quantitative comparison, evaluation results from all experiments are exported to CSV files and subsequently concatenated into a single CSV file. This aggregated results file enables direct comparison across baseline, individual augmentation strategies and the combined dataset.

Visualizations are generated to compare performance metrics and inference times across experiments, allowing trends and trade-offs to be clearly identified.

[4] https://docs.ultralytics.com/models/yolov8/

5 Results

5.1 Overall detection performance

The overall detection performance of all trained models is summarized in Table 1, which reports the global metrics mAP50, mAP50-95, precision, recall and inference time. The base model achieves an mAP50 of 0.797 and serves as the reference configuration for all comparisons.

All augmentation strategies result in improved detection performance compared to the base dataset. Geometric, color and kernel augmentations each increase mAP50 by approximately 2-4 percentage points. Among these, kernel-based augmentation achieves the highest mAP50-95 value (0.522), indicating improved localization accuracy across stricter IoU thresholds.

The combined dataset, which integrates all augmentation categories, yields the best overall performance. It reaches the highest mAP50 (0.832) and recall (0.824), demonstrating that combining complementary augmentations leads to a more robust and generalizable model. These improvements suggest that exposure to diverse visual variations during training allows the model to better handle the complexity of real broadcast basketball footage.

5.2 Impact of augmentation strategies

The relative impact of the different augmentation strategies is illustrated in Figure 3, which compares mAP50 values across datasets. The figure clearly shows that each augmentation type contributes positively to detection performance, with no augmentation strategy causing a degradation relative to the base model.

Color-based augmentation produces the highest precision score (0.893), as can be seen in Table 1, indicating fewer false positives. However, this improvement comes with only a modest gain in recall. In contrast, the combined dataset slightly sacrifices precision but achieves the highest recall, reflecting a better ability to detect all relevant objects present in an image. In the context of sports analytics, where missed detections are often more detrimental than occasional false positives, this trade-off is generally acceptable.

5.3 Per-class performance analysis

A detailed per-class analysis is visualized in Figure 4 and Figure 5, which show class-wise mAP50 for all datasets. The results indicate that data augmentation has the strongest effect on small or visually challenging objects.

The basketball itself is the most difficult class to detect. The base model achieves an mAP50 of only 0.328 for this class, while the combined dataset improves this score to 0.383. Similar improvements are observed for the shot clock, where the combined model reaches an mAP50 of 0.882 compared to 0.703 for the base dataset.

For larger and more stable objects such as players, hoops and scoreboard elements, performance is already high in the base model. Nevertheless, augmentation consistently maintains or slightly improves detection quality for these classes, confirming that the applied transformations do not negatively affect the recognition of dominant objects.

5.4 Precision-recall trade-offs

Although the combined model does not achieve the highest precision score overall, it provides the most balanced performance across all metrics (see Table 1). The precision-recall trade-off is especially visible when comparing the color-augmented and combined datasets. Color augmentation leads to fewer false positives, whereas the combined dataset improves recall and reduces missed detections.

This balance is particularly important in basketball analysis, where missing fast-moving or partially occluded objects, such as the ball or referees, can significantly impact downstream tasks like player tracking or event detection.

5.5 Inference speed and real-time feasibility

Inference speed results are also summarized in Table 1 and visualized in Figure 6. The base model processes an image in approximately 14.8ms, while augmented models (except for color-based) exhibit slightly higher inference times due to increased model complexity.

The combined model records an inference time of approximately 19.5ms per image. Despite this increase, the model remains well within real-time constraints, corresponding to over 50 frames per second ($1000/19.4 \approx 51.5$ FPS). This confirms that the performance gains obtained through data augmentation do not compromise real-time applicability, which is essential for live sports analysis and video-based decision support systems.

5.6 Summary of findings

Across all experiments, data augmentation consistently improves detection performance. While individual augmentation strategies already yield measurable gains, the combined dataset produces the most robust and well-balanced model. It achieves the highest overall detection accuracy, strong per-class performance and maintains real-time inference speed, making it the most suitable configuration for practical basketball object detection applications.

6 Conclusion

This study investigated the effect of different image data augmentation strategies on the performance of a YOLOv8-based object detection model applied to a basketball game footage. By systematically comparing a baseline dataset with datasets augmented using geometric, color-based, kernel-based and combined augmentation techniques, the impact of each strategy on detection accuracy and efficiency was evaluated in a controlled and reproducible manner.

The results demonstrate that data augmentation consistently improves object detection performance compared to training on the original dataset alone. All individual augmentation strategies led to higher mAP scores, confirming their effectiveness in increasing model robustness. Among them, kernel-based augmentations yielded strong improvements in localization accuracy, while color-based augmentations achieved the highest precision, reducing the number of false positives.

The combined augmentation approach produced the best overall results, achieving the highest mAP50 and recall while maintaining real-time inference performance. Although this configuration did not yield the highest precision score, the improved recall indicates a lower rate of missed detections, which is particularly important in sports analysis scenarios where fast-moving and partially occluded objects are common. The observed trade-off between precision and recall suggests that combined augmentation provides a more balanced and practically useful model.

Importantly, all trained models remained well within real-time constraints, with inference times below 20 milliseconds per image (except for geometric augmentation). This confirms that the performance gains achieved through data augmentation do not compromise the applicability of YOLOv8 for live basketball analysis and similar real-world use cases.

Overall, this work highlights the importance of carefully selecting and combining data augmentation techniques when training object detection models on limited or domain-specific datasets. The findings contribute empirical insights into how different augmentation categories affect YOLOv8 performance in a sports context and provide practical guidance for developing robust, real-time detection systems for basketball analytics.

7 Discussion

The goal of this study was to analyze how different image data augmentation strategies influence the performance of a YOLOv8-based object detection model in the context of basketball game footage. The results clearly show that data augmentation plays a crucial role in improving both detection accuracy and robustness, especially when training data is limited and visually complex, as is typical for sports broadcasts.

7.1 Effectiveness of individual augmentation strategies

Each augmentation category contributed positively to overall model performance, but their effects differed depending on the evaluated metric and object class.

Geometric augmentations improved the model's ability to handle variation in camera angle, scale and player positioning. This is reflected in higher mAP scores for dynamic objects such as players and referees, which frequently appear under different orientations and spatial configurations in basketball footage.

Color-based augmentations resulted in the highest precision among all configurations. This suggests that variations in brightness, contrast and color distribution help the model better distinguish foreground objects from the background, thereby reducing false positives. However, these gains in precision were not accompanied by equally strong improvements in recall, indicating that color augmentation alone may not sufficiently address missed detections under challenging conditions such as occlusion or motion blur.

Kernel-based augmentations, such as blurring and sharpening, showed strong improvements in mAP50-95. This indicates enhanced localization performance, particularly under stricter IoU thresholds. These augmentations appear to increase robustness to motion blur and broadcast artifacts, which are common in fast-paced basketball scenes.

7.2 Benefits of combined augmentation

The combined augmentation strategy consistently achieved the best overall performance across most metrics. It delivered the highest mAP50 and recall, as well as strong per-class improvements, particularly for small and difficult-to-detect objects such as the basketball and the shot clock. These results support the hypothesis that combining complementary augmentation techniques exposes the model to a wider range of visual variations, enabling better generalization to unseen data.

While the combined model did not achieve the highest precision, the observed decrease was relatively small and accompanied by a substantial increase in recall. In the context of basketball analytics, this trade-off is often acceptable as missing critical objects (e.g. the ball or referees) can have a greater negative impact on downstream tasks than occasional false detections. Therefore, from a practical perspective, the combined model offers the most balances and useful performance.

7.3 Per-class observations and dataset characteristics

The per-class analysis revealed that augmentation has the greatest impact on small, fast-moving, or visually ambiguous objects. The basketball consistently remained the most challenging object to detect across all datasets, although notable improvements were achieved with combined augmentation. Larger and more stable objects, such as players, hoops and scoreboard elements, already achieved high performance with the base dataset, but still benefited modestly from augmentation.

These findings highlight the importance of class-specific evaluation. Aggregated metrics alone may ask significant improvements for critical object categories, reinforcing the need for detailed per-class analysis when evaluating object detection models in sports contexts.

7.4 Real-time performance considerations

Despite the increased dataset complexity and improved detection accuracy, all models maintained inference times well below the threshold required for real-time processing. Even the combined model, which exhibited almost the highest inference time, processed images at over 50 FPS. This confirms that the application of data augmentation does not compromise the real-time suitability of YOLOv8, making it viable for live basketball analysis and other time-sensitive applications.

7.5 Limitations

This study has several limitations.

First, the experiments were conducted using a single model variant (YOLOv8n). While this ensured fair comparisons, results may differ for larger model variants or newer YOLO versions.

Second, the dataset size remained relatively limited, and although augmentation increased visual diversity, it cannot fully replace additional real-world data.

Finally, the evaluation focused on image-based detection rather than full video-based tracking, which may introduce additional challenges such as temporal consistency.

7.6 Implications and future work

The findings of this study suggest that careful selection and combination of augmentation strategies can significantly improve object detection performance in sports analytics without sacrificing efficiency. Future work could explore adaptive or class-specific augmentation strategies, evaluate the impact of augmentation on object tracking performance, or extend the analysis to other sports domains. Additionally, comparisons with newer YOLO variants or transformer-based detection models could provide further insights into the generalizability of these results.

8 Tables and figures

8.1 Tables

Dataset	mAP50	mAP50-95	Precision	Recall	Inference (ms)
Base	0.79652	0.49286	0.84642	0.76642	14.81764
Geometric	0.81750	0.50987	0.88305	0.75614	21.20283
Color	0.82542	0.51404	0.89344	0.76163	14.17165
Kernel	0.83023	0.52229	0.84457	0.79597	16.02768
Combined	0.83244	0.52101	0.84962	0.82409	19.46548

Table 1: Overall detection performance across datasets

8.2 Figures

Figure 1: Example of original image from base dataset (without preprocessing)

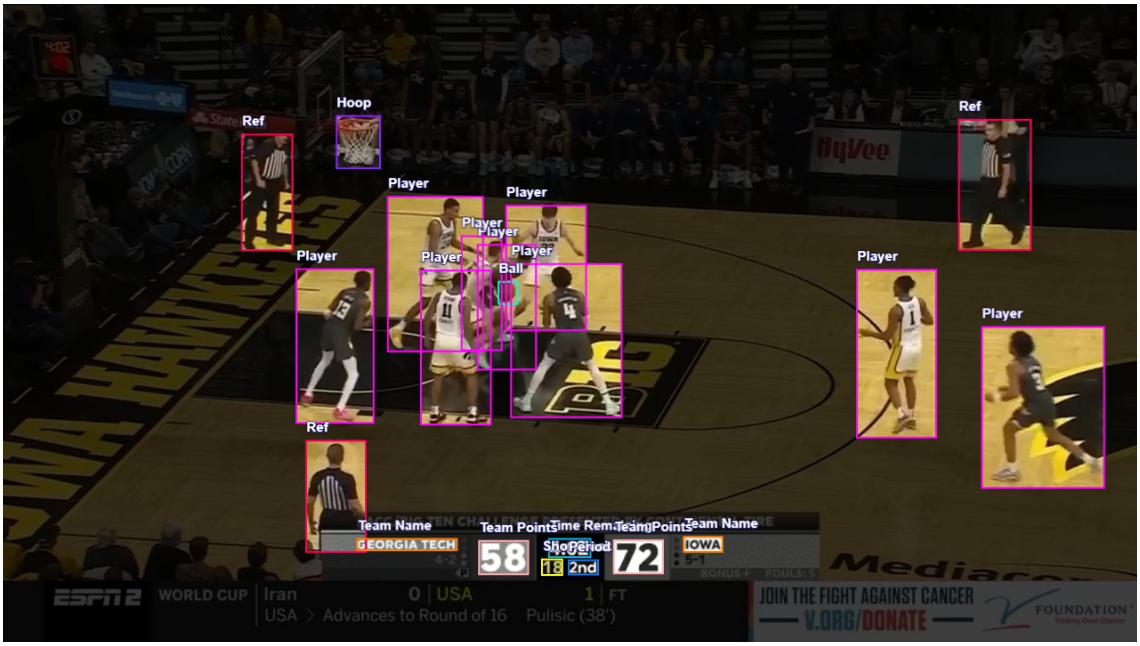

Figure 2: Example of annotated image from base dataset (without preprocessing)

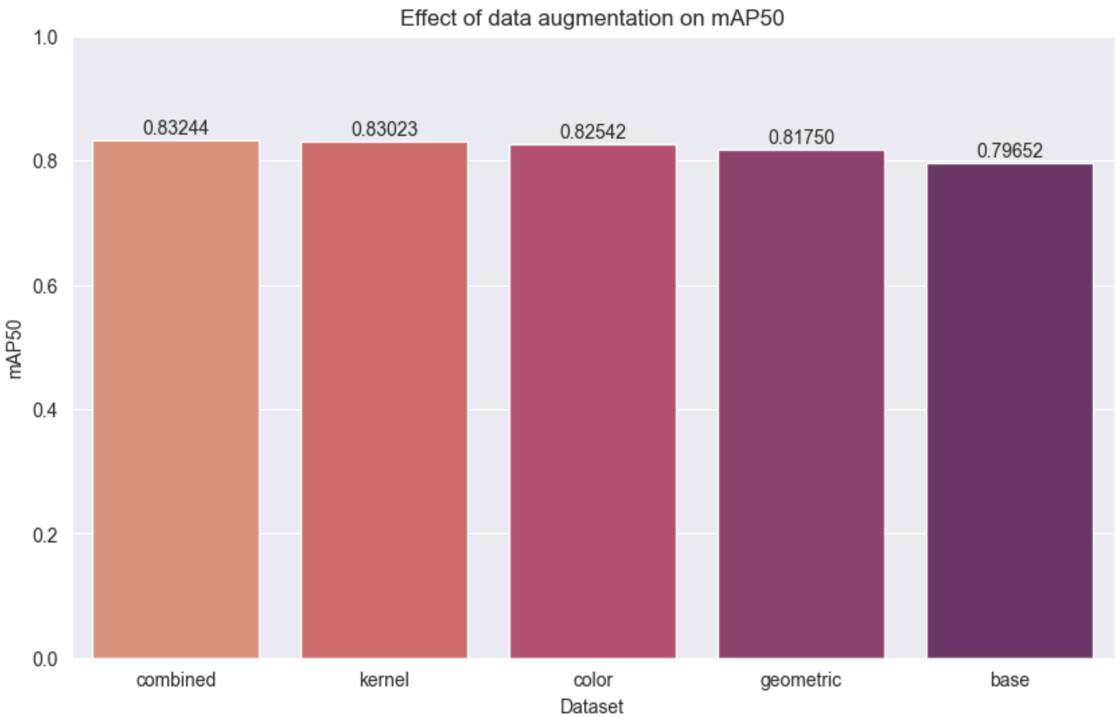

Figure 3: Effect of data augmentation across datasets on mAP50

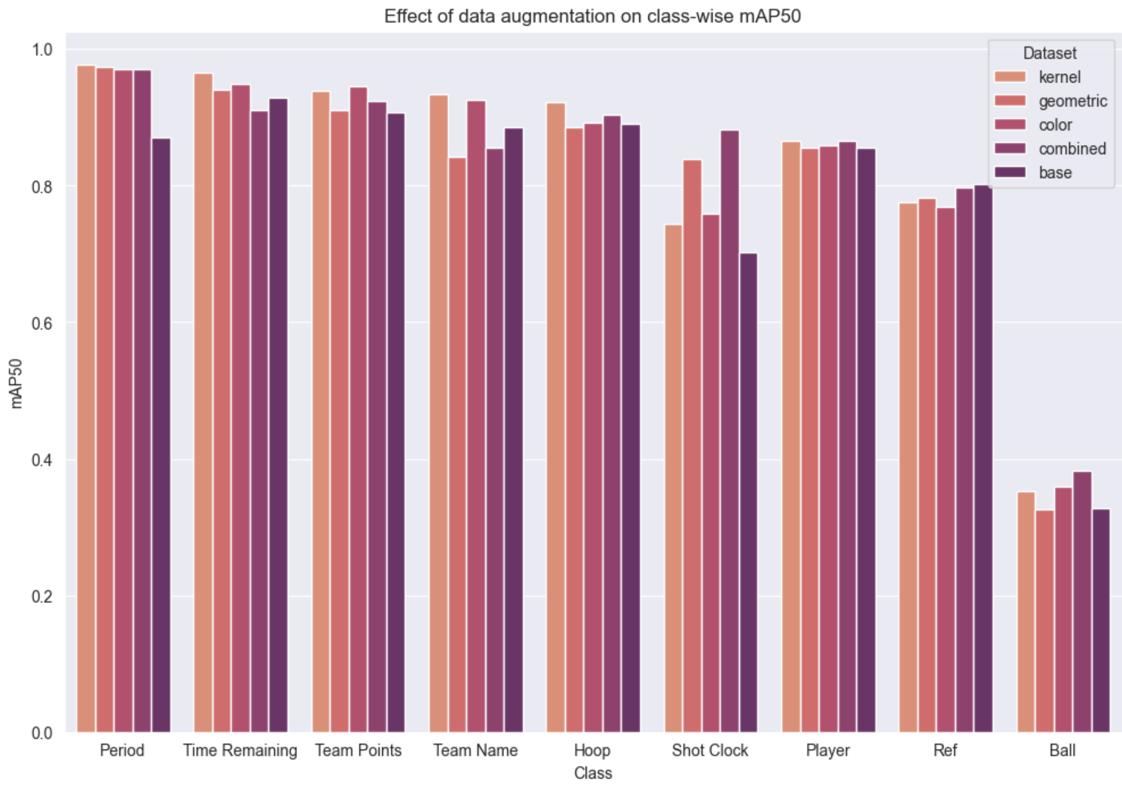

Figure 4: Effect of data augmentation per class on mAP50

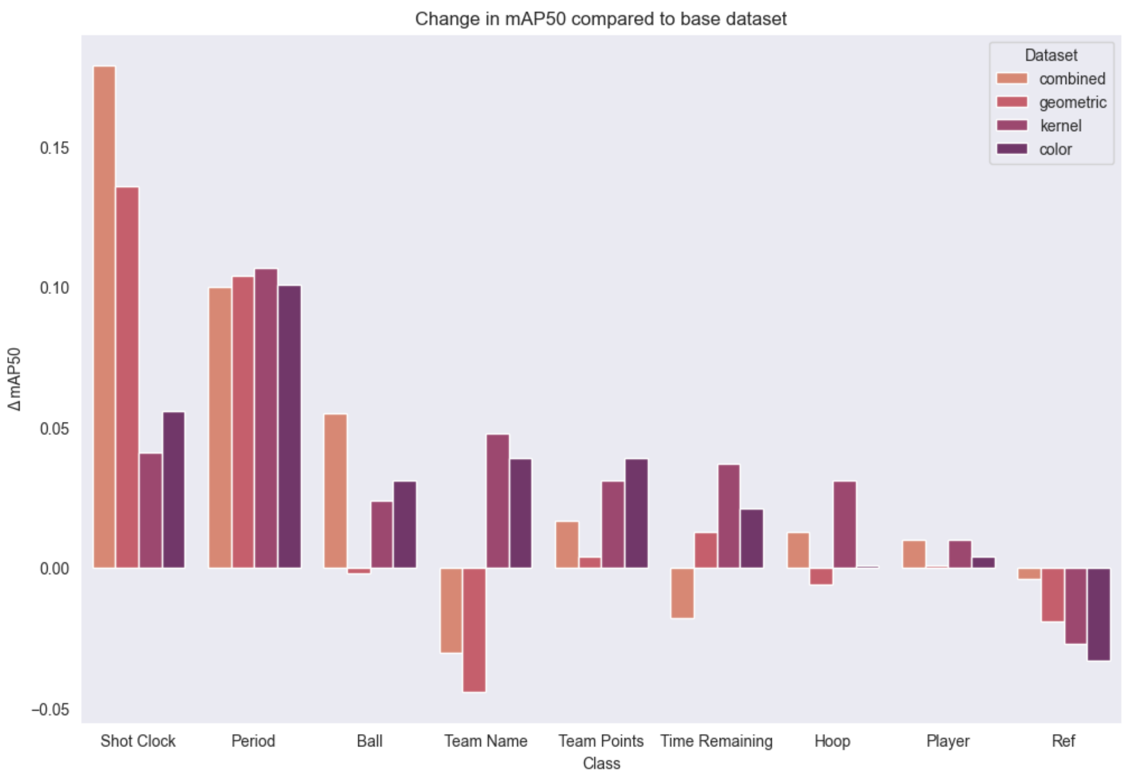

Figure 5: Effect of data augmentation per class on mAP50 compared to base dataset

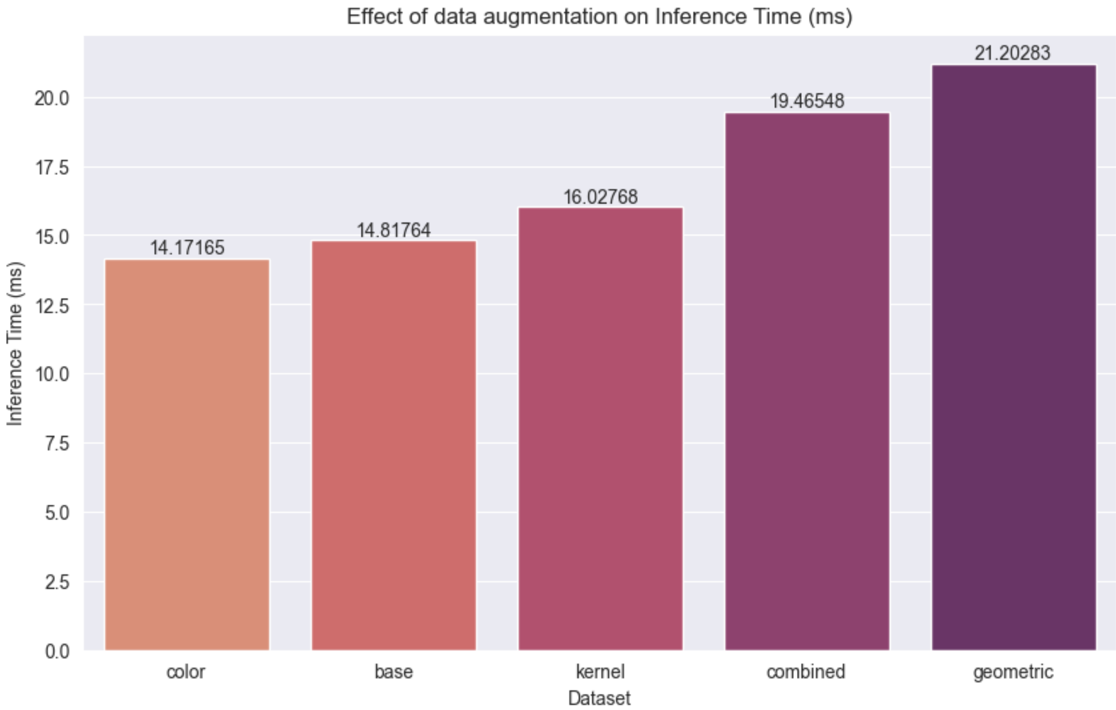

Figure 6: Effect of data augmentation across datasets on Inference Time (ms)

References

[1] A. A. Awan. A complete guide to data augmentation, Dec. 2024.

[2] G. Jocher, A. Chaurasia, and J. Qiu. Ultralytics yolov8, 2023.

[3] E. Kavlakoglu and J. Murel. What is data augmentation? — ibm, May 2024.

[4] E. Kavlakoglu and J. Murel. What is Object Detection? | IBM, Jan. 2024.

[5] D. Keskin. Synthetic data and data augmentation, Nov. 2023.

[6] Z. Liang, J. Wang, T. Huang, Z. Sang, and J. Zhang. Basketball detection based on yolov8. *PLOS ONE*, 20(8):e0326964, Aug. 2025.

[7] C. Moreira, L. Ferreira, and P. Coelho. A comprehensive review of ball detection techniques in sports. *PeerJ Computer Science*, 11:e3079, 08 2025.

[8] E. Quinn and N. Corcoran. Comparative analysis of yolov5/v8/v9 for object detection, tracking, and human action recognition in combat sports. *International Conference on AI Research*, 4:364–373, 12 2024.

[9] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015.

[10] B. Wang. Football sports video tracking and detection technology based on yolov5 and deepsort. *Discover Applied Sciences*, 7(6):563, May 2025.