

Evaluating YOLOv8-pose variants for soccer keypoint detection and homography mapping

Durrotun Nashihin¹, Yohanes Yohanie Fridelin Panduman², Dwi Cahyo Kartiko³, Anna Noordia⁴,
Khoirul Islam¹

¹Department of Informatics Engineering, State University of Surabaya, Surabaya, Indonesia

²Department of Multimedia Engineering, The University of Osaka, Suita, Japan

³Department of Physical Education, State University of Surabaya, Surabaya, Indonesia

⁴Department of Sports Science, State University of Surabaya, Surabaya, Indonesia

Article Info

Article history:

Received Dec 1, 2025

Revised Aug 4, 2026

Accepted Aug 12, 2026

Keywords:

Broadcast data

Computational efficiency

Detection accuracy

Homography mapping

Homography reliability

Keypoint detection

YOLOv8-pose variants

ABSTRACT

Soccer analytics intensely relies on accurate pitch keypoint detection and homography mapping, yet soccer broadcast data poses challenges caused by camera angle variation and partial pitch coverage. This study systematically evaluates you only look once version 8 (YOLOv8)-pose variants (nano, small, medium, large, and x-large) for keypoint detection and homography mapping. These models were retrained on broadcast and scouting datasets, while homography mapping was computed using direct linear transform (DLT) and random sample consensus (RANSAC). Three metrics were used for evaluation: detection accuracy, computational efficiency, and homography reliability. Results showed that the medium variant achieved the best balance, with the lowest mean pixel error (35.67 px), competitive inference speed 49.6 frames per second (FPS), and moderate model size (52.2 MB). Furthermore, it also reached the highest homography success rate (96.2%), surpassing all other variants. These findings suggest that the medium variant is the most effective balance for real-time soccer analytics, providing robust performance in accuracy, efficiency, and reliability under diverse broadcast data. Future studies would focus on enlarging the dataset with additional keypoints, incorporating temporal smoothing, and integrating a player tracking system. For reproducibility, the dataset and code are openly available on GitHub.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Durrotun Nashihin

Department of Informatics Engineering, State University of Surabaya

Building A10, Ketintang Wiyata Street, Surabaya-60231, Indonesia

Email: durrotunnashihin@unesa.ac.id

1. INTRODUCTION

Despite the rapid development of soccer analytics, open-source platforms and accessible soccer scouting data remain limited. This situation can be addressed by utilizing broadcast data for low-cost soccer analytics. Major tournaments run by Fédération Internationale de Football Association (FIFA), like English Premier League and World Cup, even have been generating thousands of hours of high plausible data for scalable soccer analytics [1]. This condition confirms that the broadcast data potentially offers substantial resources for large-scale soccer analytics. However, in comparison to scouting data, broadcast data usually contains some challenges, such as varying camera angles, zoom levels, and occluded pitch regions [2], [3]. Therefore, such change in camera perspectives poses a much more challenging task to ensure accurate and stable soccer analytics. One of the pivotal tasks used to support soccer analytics through broadcast data is

accurate detection of field landmark/keypoint [4]. This process is crucial to standardize a top-down view of the pitch, a process known as homography mapping. The alignment between keypoint detection and homography mapping is commonly a prerequisite for downstream tasks, such as player tracking and tactical analysis [5]. In this way, both become two core components to enable transition from broadcast data into accurate and scalable soccer analytics.

Recent architectures provide useful information towards addressing these issues. Spotlight, the first winner of the 2023 SoccerNet camera calibration challenge, used high-resolution network version 2-width 48 (HRNetV2-w48) to detect pitch landmarks via heatmaps. They achieved accurate homography mapping from video frames to the top-down pitch view [6]. Despite its high-quality results, HRNet is computationally costly and might not be applicable to real-time analytics. However, the second winner, Spiideo, used you only look once X-pose (YOLOX-pose) as player pelvis keypoints, which leads to proving that even lightweight and speed-oriented pose detectors can be effective [7]. Based on such insights, the YOLO-pose family may become an excellent alternative to detecting soccer pitch landmarks. YOLO-pose can be retrained to detect field landmarks, although originally it was used to detect human keypoint. It provides five variants (nano, small, medium, large, and extra-large) that give flexibility in terms of speed and the complexity of the model [8]. As an example, YOLOv8n-pose (nano) took 131.8 ms to run on CPU-ONNX with the model size of 3.3 MB compared with 1,607 ms per frame and 69.4 MB in YOLOv8x-pose (extra-large) [9]. The benchmarks outline a noticeable speed accuracy trade-off, proposing the need to focus on the choice of a variant that will adopt the time-sensitive sports visual analytics.

Regardless of flexibility for choosing the YOLO-pose models, it is not specified yet which variant provides the most effective performance for detecting soccer pitch keypoints in diverse camera angles and partial occlusions [10]. This statement can be supported by the recent studies demonstrating that the actual effectiveness of each variant related to keypoint detection or soccer analytics is still unclear [9], [11]. Moreover, the ability of a reliable algorithm to map keypoint detected with YOLO-pose to homography mapping has not been widely explored. In this case, stable and reliable homography mapping is needed to overcome a situation where YOLO-pose model mispredicts keypoints, resulting in outliers. This is because inaccurate homography mapping will directly impact downstream processes, such as automated player tracking or tactical analysis tracking [12]. These issues highlight a clear gap in understanding the real performance of YOLO-pose variants for keypoint detection and what is the appropriate way to utilize these detected keypoints into reliable homography mapping. Therefore, a deeper investigation into these aspects is inevitably needed.

According to the mentioned issues, this paper aims to systematically evaluate the YOLO-pose variant for pitch keypoint detection and its homography mapping. To our knowledge, this paper introduces the first reproducible baseline of YOLO-pose model variants used in soccer homography mapping. Our contribution in detail can be summed up as follows:

- i) Robust keypoint detection accuracy: determine which YOLO-pose variant performs best under noisy broadcast data with occlusions and scouting data.
- ii) Computational efficiency: examine the latency, inference speed and model size to learn the trade-offs among real-time feasibility and detection accuracy of soccer analytics application.
- iii) Homography reliability: assess the robustness of homography mappings under varying video conditions to support downstream tactical analysis tasks.

In order to lead the readers through these contributions, the paper is organized in the following ways: section 2 provides related works, section 3 describes the methodology, section 4 includes results and discussion, and finally, section 5 gives the conclusion to the study.

2. RELATED WORKS

The recent innovations in pose estimation have demonstrated positive outcomes in football field localization [6], [7]. The performance of HRNet-based models in keypoint detection was observed through the SoccerNet camera calibration challenge. However, it is not ideal for low-end computers or real-time conditions due to its high computational cost [6]. Conversely, YOLO-pose presents lighter forms that offer tradeoffs between accuracy and speed, which is a good solution for limited computer resources [7], [8]. As the use of YOLO frameworks has been growing, other researchers have also begun to compare various YOLO generations (e.g., YOLOv5, YOLOv7, and YOLOv8) to learn more about trade-offs between inference speed versus its accuracy. Nonetheless, the majority of these researches are aimed at object detection and not at keypoint or landmark estimation [13], [14]. Consequently, comparisons in performance between different variants of a particular YOLO-pose are not well studied, especially when detecting soccer field keypoints in both broadcast and scouting scenarios.

Homography mapping is a crucial step to achieve soccer field localization, consisting of estimating a geometric transformation using detected pitch features into a standardized field template. Earlier approach

was based on classical computer vision algorithms like Hough-based line detection, ellipse fitting center circle, and template matching penalty box [15]. Although it worked well in a controlled environment, like scouting data, it is generally worse when used in a broadcast environment. Therefore, it is not a proper option for a large-scale automated analysis [16]. With the emergence of deep learning, the genre changed to convolutional neural networks (CNNs) with the ability to learn discriminative visual representations. Early studies were based on semantic segmentation networks, including fully convolutional networks (FCNs) and DeepLab to segment field lines or areas of interest [17], [18]. Segmentation-based pipelines, though more robust, were frequently costly in terms of annotations and had a difficult time competing with real-time pipelines. According to this current situation, keypoint-based methods is necessary to be explored as they provide a highly promising, lightweight, and flexible alternative for homography mapping.

3. METHOD

Figure 1 presents the system overview of this study with two input datasets (broadcast and scouting), five YOLOv8-pose model variants for keypoint detection, and three main evaluations. Although broadcast data is our main concern of this study, scouting data is still included since we want to emphasize that our approach can work well across different camera perspectives and field conditions. Then, YOLOv8-pose is chosen because it has been widely used and well-validated pose model among researchers in different area [19], [20]. Subsection 3.1 explains the dataset, preprocessing and training strategy, while subsections 3.2-3.4 describe the three main evaluations of this study: accuracy evaluation, computational efficiency, and homography mapping reliability. The details of the evaluation metrics can be seen in Table 1.

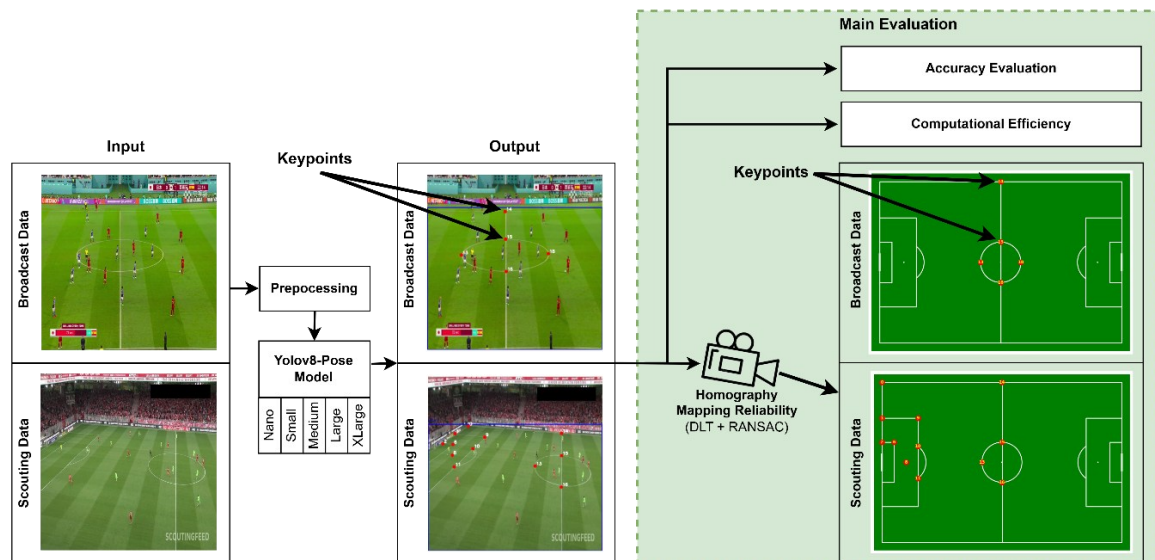


Figure 1. System overview of soccer field keypoint detection and homography mapping

Table 1. Summary of evaluation metrics

Evaluation	Metrics
Accuracy evaluation	Loss functions, mean average precision (mAP) and mean pixel error (MPE)
Computational efficiency	Latency, inference speed, and model size
Homography mapping reliability	Ratio of success rate (%)

3.1. Dataset, preprocessing, and training strategy

This study uses the Soccer field localization dataset from Roboflow [21], containing 514 manually annotated keypoint images from scouting and broadcast data. For further reproducibility, dataset and code can be accessed in GitHub at <https://github.com/durrotunnashihin/yolov8pose-soccer-keypoint-homography-mapping>. To make it clear, the meaning of keypoint in this research is a given point on the soccer field, like corners or markers of penalty areas, annotated with pixel coordinates. For the preprocessing step, the flip_idx parameter in the data.yaml was slightly changed to ensure that the left-right keypoints are properly matched in augmentation process. Furthermore, the size of all input images was standardized to a certain resolution (640×640 px) to standardize the datasets and optimize computing efficiency. In the training strategy, the hyperparameters of all

models were configured equally. The purpose of this setting is to ensure the performance's difference was based on model capacity, and not on optimization bias. Then, the stochastic gradient descent (SGD) optimizer with initial rate 0.01 was chosen due to its stability and efficiency, especially for large-scale training [22]. To address the stochastic effects, the data augmentation was as light as possible, turning mosaic augmentation off, and using simple geometric transformations.

3.2. Accuracy evaluation

The accuracy of the keypoint detection models was evaluated using three metrics: loss functions, mAP and MPE. During training, loss functions (training and validation losses) were used to monitor model performance. The training loss measures how well the models fit between predicted and actual keypoints, while the validation loss reflects the model's capability to generalize the previous unseen data [23]. In addition, mAP at two standard thresholds (mAP@0.5 and mAP@0.5:0.95) was implemented to measure the accuracy of keypoint localization under different levels of strictness. As shown in (1), the mAP is computed as the mean of average precision (AP) values across multiple thresholds τ [24]. To complement mAP, this study adopts MPE from prior work on object pose estimation to quantify the accuracy of keypoint detection in our soccer pitch keypoint detection task [25]. This metric as in (2) directly measures the average Euclidean distance in pixels between predicted $(\hat{x}_i, \hat{y}_i)$ and ground-truth (x_i, y_i) keypoints.

$$mAP = \frac{1}{|T|} \sum_{\tau \in T} AP(\tau) \quad (1)$$

$$MPE = \frac{1}{N} \sum_{i=1}^N \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2} \quad (2)$$

3.3. Computational efficiency

To measure the computational efficiency, this study used three metrics: latency, inference speed, and model size. Together, the combination of these metrics can support the practical implementation of real-time soccer analytics. First metric is latency defined as the average time to process a single frame. It is mathematically stated as in (3), where T_{total} is the sum of inference duration over N test frames. As seen in (3), the second metric is frames per second (FPS), describing how many total executed frames that models can process per second. This study considered these metrics because they are widely used in for AI applications, such as YOLO-fall, a lightweight edge model in which speed and accuracy are essential [26]. The model is the last metric, where larger model size usually requires greater computational and memory resources.

$$Latency = \frac{T_{total}}{N} \times 100, \quad FPS = \frac{N}{T_{total}} \times 100 \quad (3)$$

3.4. Homography mapping reliability

The homography mapping was calculated with direct linear transform (DLT) which matches identified field keypoints (source points) with their annotated reference points (destination points) on the soccer pitch template [27]. To make it robust against noise, we only used detected keypoints with a confidence level more than 0.8. Furthermore, the homography mapping is considered successful only if minimal detected keypoints are four with less geometric alignment. Besides, we also used the random sample consensus (RANSAC) algorithm, which repeatedly chooses random subsets of keypoint correspondences to hypothesize a homography matrix and then evaluates how many correspondences agree with this hypothesis [28]. Keypoints with reprojection error less than the threshold are considered inliers, and the rest are outliers. To measure the homography reliability, each YOLOv8-pose variant was evaluated on each image on the test set. For each image, a binary outcome was obtained, successful mapping as 1 and unsuccessful mapping as 0. As seen in (4), the formula is made to calculate the ratio of images with successful mapping over N all images. Different with mAP or MPE, which measure accuracy, this metric directly captures the usability of the system in a downstream task, i.e. tactical analysis and visualization of player trajectories.

$$R_{map} = \frac{1}{N} \sum_{i=1}^N is_mapping(i) \quad (4)$$

4. RESULTS AND DISCUSSION

4.1. Experimental setup

This section outlines the experimental setup used to evaluate the proposed approach. It covers three key aspects: dataset composition, hyperparameter settings, and hardware specifications.

4.1.1. Dataset composition

The dataset composition is summarized in Table 2, while the 32 annotated keypoints are illustrated in Figure 2. Besides, Figure 3 summarizes the distribution of keypoint visibility in the Soccer Field Localization dataset, with the overall counts for the three visibility categories (invisible, occluded, and visible) presented in Figure 3(a), and the distribution of visible keypoints shown in Figure 3(b). The highest numbers of annotated keypoints are observed at reference points 15, 14, and 13, indicating that the dataset majorly covers the middle area of the pitch. This distribution provides an important basis to measure the performance of homography mapping in various visibility conditions.

Table 2. Dataset composition

Subset	Number of images	Description
Training	402	254 scouting images, 148 broadcasting images
Validation	60	31 scouting images, 29 broadcasting images
Testing	52	32 scouting images, 20 broadcasting images
Total	514	-

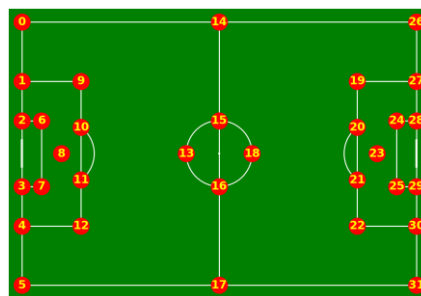


Figure 2. Keypoint annotation of a soccer pitch with 32 reference keypoints

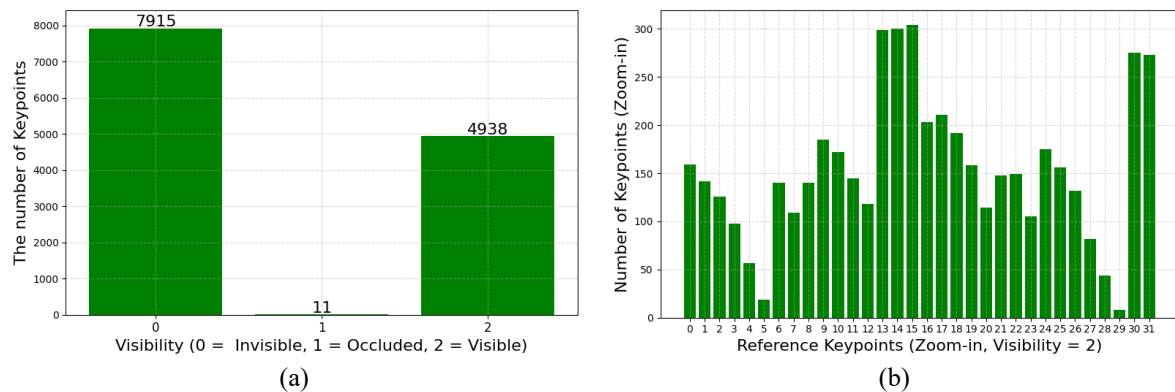


Figure 3. Distribution of keypoint visibility in the Soccer Field Localization dataset: (a) the overall counts for each visibility class and (b) the frequency of each reference point when labeled as visible

4.1.2. Hyperparameter settings

Table 3 displays the hyperparameter settings in this study. This configuration compromised both computational efficiency and reproducibility. So, lightweight and heavyweight YOLOv8-pose variants could be trained using the same experimental conditions.

Table 3. Hyperparameter settings for YOLOv8-pose model training

Parameter	Value	Notes
Epochs	500	Long schedule to ensure convergence
Image size	640 × 640	Fixed input resolution
Batch size	32	Balanced for GPU memory efficiency
Optimizer	SGD	Stable for keypoint detection tasks
Initial learning rate (lr0)	0.01	Default warm-up rate
Mosaic augmentation	0.0	Disabled to preserve pitch geometry

4.1.3. Hardware specification

The experimental environment is operationalized on Google Colab Pro. Google Colab Pro is a cloud-based platform that provides versatile computing resources for deep learning experiments. Details of the hardware specification can be seen in Table 4.

Table 4. Hardware specification of the experimental environment

Component	Specification
GPU	NVIDIA A100 (40 GB VRAM)
CPU	Intel Xeon (12 vCPUs, ~2.20 GHz)
RAM	83.5 GB
Disk storage	112.6 GB
Device	GPU (0)
Workers	12

4.2. Accuracy performance across models

Figure 4 depicts the loss function across the YOLOv8 variants over 500 epochs. Figure 4(a) shows the training loss, while Figure 4(b) shows the validation loss. Figure 5 depicts the accuracy metrics across the YOLOv8 variants over 500 epochs. Figure 5(a) shows the mAP@0.5, while Figure 5(b) shows the mAP@0.5:0.95. In general, the medium model should be considered as the best model due to its lowest value in validation loss and its highest value in mAP@0.5:0.95 after 500 epochs. In the loss function curves, the nano model always has a higher amount of residual. The x-large model, on the other hand, have much faster reductions in validation loss and stabilizes earlier like the medium one. However, after around 450 epochs, the medium model clearly performs the lowest than the others. In the mAP@0.5 and mAP@0.5:0.95, the nano is the only model with the slowest convergence, while the x-large can still be defeated by the medium model, especially after 450 epochs in mAP@0.5:0.95. To support that the medium should be chosen as the best model, Table 5 shows the MPE of all model variants. It is clear that the medium model has the lowest MPE (35.67 px) than the others. It also proves that this model can be used for more precise keypoint detection.

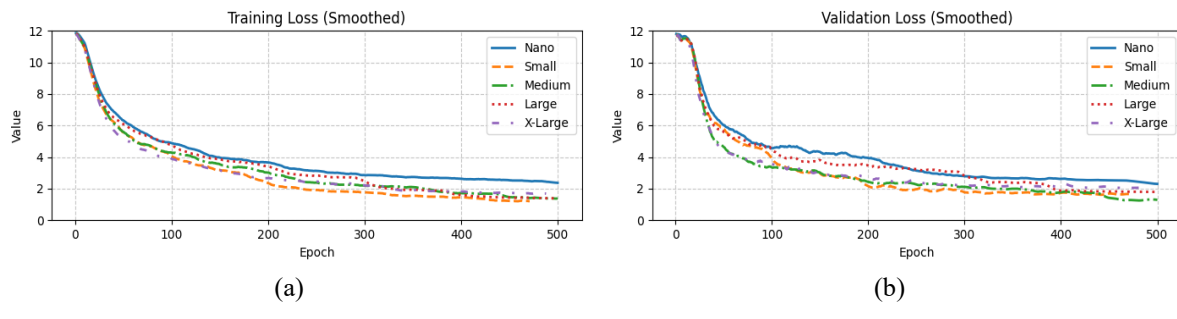


Figure 4. Training and validation loss across YOLOv8 variants (nano, small, medium, large, and x-large) over 500 epochs: (a) training loss and (b) validation loss

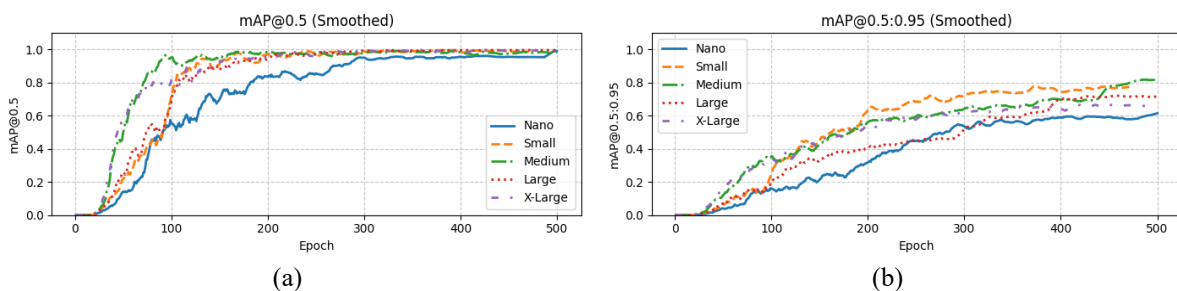


Figure 5. Accuracy metrics across YOLOv8 variants (nano, small, medium, and large, x-large) over 500 epochs: (a) mAP@0.5 and (b) mAP@0.5:0.95

Table 5. MPE across YOLOv8-pose model variants

Model variants	Nano	Small	Medium	Large	X-large
MPE (px)	76.31	60.00	35.67	55.59	67.35

4.3. Trade-offs in computational efficiency

Table 6 shows the computation efficiency values across the five YOLOv8-pose variants. The findings demonstrate that the nano model presents the lowest latency (16.1 ms), the highest inference speed (62.1 FPS) and the smallest model size (7.35 MB). However, this model does not necessarily the most practical solution for real-time soccer analytics as accuracy performance is also matters, not computational efficiency only. Combining accuracy performance in subsection 4.1 and computational efficiency in Table 6, the medium model achieves the most balanced solution with competitive latency (19.7 ms), inference speed (50.8 FPS) and model size (52.2 MB), while maintaining robust detection accuracy. The prior research has identified a similar trade-off between speed and accuracy in earlier works about object detection with deep learning models [29]. Nevertheless, the observed performance still has a limitation, where the limited dataset and fixed resolution (640×640) may favor models of medium capacity since larger models may not fully exploit their potential.

Table 6. Computational efficiency measured in latency, inference speed and model size

Model variants	Nano	Small	Medium	Large	X-large
Latency (ms)	16.1	16.7	19.7	22.8	23.8
Inference speed (FPS)	62.1	59.8	50.8	43.9	42.0
X-Model size (MB)	7.35	23.6	52.2	86.3	133.7

4.4. Homography mapping reliability findings

Figure 6 reports the homography success rate of different YOLOv8-pose model variants based on two dataset types (broadcast and scouting). The broadcast data (20 images) yields nano 80% (16/20), small 85% (17/20), medium 90% (18/20), large 80% (16/20), and x-large 80% (16/20). Then, the scouting data (32 images) has a more stable result, where nano 100% (32/32), small 100% (32/32), medium 100% (32/32), large 100% (32/32), and x-large 96.9% (31/32). Across the full dataset (52 images), the overall success rates are nano 92.3%, small 94.2%, medium 96.2%, large 92.3%, and x-large 90.4%. Two consistent trends emerge. First, scouting images are almost always solvable, due to the fixed vantage point and full-field coverage. Second, broadcast images represent the primary challenge: success rates majorly plateau near 80% even for the largest models, while the medium achieves 90% and thus provides the best choice for reliability.

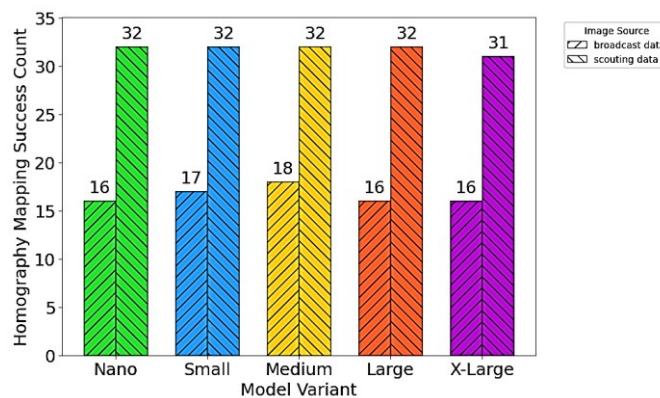


Figure 6. Success counts of homography mapping across YOLOv8-pose variants

Figure 7 illustrates how strong estimation can help to overcome the outlier correspondences using RANSAC algorithm. The keypoint labeled as #27 (highlighted in yellow) acts as a geometric outlier of the pitch template. Yet, RANSAC correctly excludes it during model fitting, allowing the homography to be accurately recovered. In this study, the RANSAC reprojection threshold was set to 3 px, which is the default value from OpenCV's library [30]. This is equivalent to about 0.47% at 640×640 resolution. RANSAC is the paradigm of choice in robust geometric fitting in the face of outlier contamination [31], and its effectiveness here highlights its necessity for stable homography mapping in the presence of noisy detections. Similar applications have been applied to sports analytics whereby strong outlier management is essential to registering the correct pitch [16].

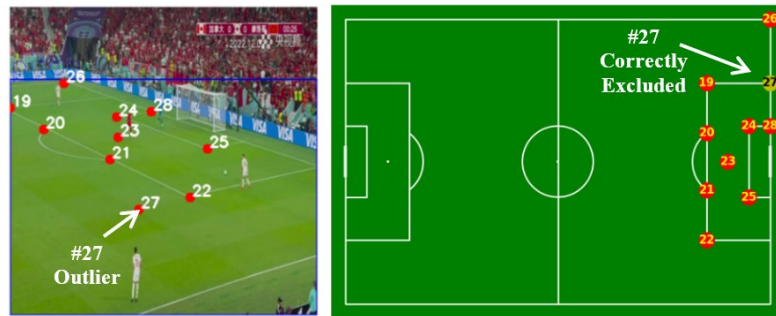


Figure 7. Homography mapping with RANSAC: keypoint #27 (yellow) acts as an outlier but is correctly excluded, enabling accurate homography recovery despite noisy detections

4.5. Discussion

According to the findings in subsections 4.2 and 4.3, the medium model offers the best compromise as it obtains the highest keypoint detection accuracy, while maintaining computational efficiency. This indicates that its model complexity is already sufficient to capture the characteristic of our dataset with fixed input resolution (640×640). For deeper understanding why this occurs, we can take a look at the characteristic of the nano as the smallest model and the x-large as the largest model. The smaller models, like the nano, has lower residuals in the loss functions due to its smaller representational ability. In the contrary, the larger models, like x-large, cannot be the best as it cannot maximize its ability to capture the limited dataset with fixed input resolution (640×640). Overfitting can tend to be common when the model capacity is larger than the data used in tasks. The issue is also supported by other studies describing that a model capacity and dataset size are important factor for keypoint detection [32], [33].

Taking into account the findings in subsection 4.4, this proposed system still has a limitation despite the robustness of the homography mapping. Figure 8 examines the two failure cases of the YOLOv8-pose medium model, which otherwise achieved the highest overall success rate (96.2%, 50/52 images). Both failures occur in broadcast images and share the same cause: fewer than four keypoints were detected. Since a minimum of four non-collinear correspondences is mathematically required to compute a projective transformation, RANSAC could not produce a valid homography. This highlights why the medium model could not achieve 100% reliability under broadcast data, despite outperforming all other models. In contrast, scouting images consistently provide enough well-distributed keypoints, resulting in perfect homography mapping. Similar constraints on keypoint density and spatial spread have been observed in other pitch registration pipelines, where insufficient features lead to mapping collapse [16].



Figure 8. Failure cases of YOLOv8-pose medium in broadcast data due to fewer than four detected keypoints

In summary, three main limitations can be identified based on our findings, along with several potential solutions. First, the medium performs as the best model in our limited dataset with fixed input size (640×640). In practice, the variety of camera angles can be more diverse, so enlarging this dataset size with diverse input resolution is necessary to get more robust model. Then, the applicability of this approach may also be further evaluated with multi-camera setups or in other sports domains, like basketball. Second, keypoint density in the visible area is another bottleneck in broadcast frames. Even such estimators, like DLT + RANSAC, cannot find a stable homography without sufficient detection. To solve this, the dataset

should be improved by adding more keypoints to enhance homography stability. Additionally, temporal smoothing between frames based on the adjacent frame information is also necessary to reconstruct lost keypoints [34]. Third, accurate keypoint detection with stable homography mapping cannot stand independently without player tracking. Thus, integrating it with player tracking is mandatory to enable comprehensive tactical analysis.

5. CONCLUSION

This paper presents a systematic evaluation of different YOLOv8-pose model variants for keypoint detection and homography mapping using broadcast and scouting datasets. The results demonstrate that the medium YOLOv8-pose variant provides the best trade-off balance between detection accuracy, computational efficiency, and homography mapping reliability. Compared to other variants, it achieves the best detection accuracy, while maintaining real-time inference speed and model complexity. Furthermore, the medium variant combined with DLT and RANSAC enables robust homography mapping, not only under scouting data but also broadcast data, generating more diverse camera angles. Thus, this model offers the most promising and practical solution for real-time soccer analytics. To the best of our knowledge, this work provides the first reproducible benchmark of different YOLOv8-pose variants for soccer homography mapping. Finally, future work can focus on utilizing larger datasets with additional keypoint labels, incorporating temporal smoothing, and integrating a tracking system to produce more robust and complete real-time player homography mapping.

FUNDING INFORMATION

This research was funded by a non-APBN grant from State University of Surabaya under Contract No. B/40006/UN38.III.1/LK.04.00/2025.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Durrotun Nashihin	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓
Yohanes Yohanie	✓	✓		✓						✓	✓			
Fridelin Panduman														
Dwi Cahyo Kartiko						✓				✓				✓
Anna Noordia	✓					✓	✓		✓					✓
Khoirul Islam	✓								✓				✓	✓

C : **C**onceptualization

M : **M**ethodology

So : **S**oftware

Va : **V**alidation

Fo : **F**ormal analysis

I : **I**nvestigation

R : **R**esources

D : **D**ata Curation

O : Writing - **O**riginal Draft

E : Writing - Review & **E**diting

Vi : **V**isualization

Su : **S**upervision

P : **P**roject administration

Fu : **F**unding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY

The dataset with codes is openly available and can be accessed through the Github repository at <https://github.com/durrotunnashihin/yolov8pose-soccer-keypoint-homography-mapping>.

REFERENCES

- [1] A. Deliege *et al.*, "SoccerNet-v2: a dataset and benchmarks for holistic understanding of broadcast soccer videos," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2021, pp. 4503–4514, doi: 10.1109/CVPRW53098.2021.00508.
- [2] S. Akan and S. Varlı, "Use of deep learning in soccer video analysis: survey," *Multimedia Systems*, vol. 29, no. 3, pp. 897–915, Jun. 2023, doi: 10.1007/s00530-022-01027-0.

- [3] F. Magera, T. Hoyoux, O. Barnich, and M. V. Droogenbroeck, "BroadTrack: broadcast camera tracking for soccer," in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Feb. 2025, pp. 6177–6187, doi: 10.1109/WACV61041.2025.00602.
- [4] H.-Y. Wang, P.-Y. Chou, Y.-Y. Kao, and C.-H. Lin, "GloFANet: global feature augmented networks for data-efficient soccer field keypoint detection," in *2025 IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2025, pp. 1–2, doi: 10.1109/ICCE63647.2025.10930162.
- [5] M. Manafifard, "A review on camera calibration in soccer videos," *Multimedia Tools and Applications*, vol. 83, no. 6, pp. 18427–18458, Jul. 2023, doi: 10.1007/s11042-023-16145-8.
- [6] N. S. Falaleev and R. Chen, "Enhancing soccer camera calibration through keypoint exploitation," in *Proceedings of the 7th ACM International Workshop on Multimedia Content Analysis in Sports*, Oct. 2024, pp. 65–73, doi: 10.1145/3689061.3689074.
- [7] H. Ardö *et al.*, "Spiideo SoccerNet SynLoc: single frame world coordinate athlete detection and localization with synthetic data," in *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2025, pp. 278–285, doi: 10.5220/0013108200003912.
- [8] Ultralytics, "Explore ultralytics YOLOv8," docs.ultralytics.com. Accessed: Aug. 15, 2025. [Online]. Available: <https://docs.ultralytics.com/models/yolov8/>
- [9] Ultralytics, "Pose estimation with ultralytics YOLO," docs.ultralytics.com. Accessed: Aug. 15, 2025. [Online]. Available: <https://docs.ultralytics.com/tasks/pose/>
- [10] P. Skalski, "Camera calibration in sports with keypoints," blog.roboflow.com, 2024. Accessed: Aug. 15, 2025. [Online]. Available: <https://blog.roboflow.com/camera-calibration-sports-computer-vision/>
- [11] P. S. S. Markappa, C. O'Leary, and C. Lynch, "A review of YOLO models for soccer-based object detection," in *2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS)*, Oct. 2024, pp. 1–7, doi: 10.1109/ICDS62089.2024.10756443.
- [12] G. Nousias, K. K. Delibasis, and I. G. Maglogiannis, "Intelligent sampling consensus for homography estimation in football videos using featureless unpaired points," *IEEE Access*, vol. 13, pp. 187843–187857, 2025, doi: 10.1109/ACCESS.2025.3627538.
- [13] K. Jannet, M. S. Sunar, M. M. Islam Molla, and M. A. As'Ari, "A deep learning approach to badminton player footwork detection based on YOLO models: a comparative study," in *2024 IEEE 8th International Conference on Signal and Image Processing Applications (ICSIPA)*, Sep. 2024, pp. 1–6, doi: 10.1109/ICSIPA62061.2024.10686537.
- [14] M. Atik, Z. Duran, and R. Özgünlük, "Comparison of YOLO versions for object detection from aerial images," *International Journal of Environment and Geoinformatics*, vol. 9, no. 2, pp. 87–93, Jun. 2022, doi: 10.30897/ijegeo.1010741.
- [15] D. Berjón, C. Cuevas, and N. García, "Soccer line mark segmentation and classification with stochastic watershed transform," *Signal Processing: Image Communication*, vol. 118, Oct. 2023, doi: 10.1016/j.image.2023.117014.
- [16] M. G.-Pérez and A. Agudo, "PnLCalib: sports field registration via points and lines optimization," *Computer Vision and Image Understanding*, vol. 267, Apr. 2026, doi: 10.1016/j.cviu.2026.104712.
- [17] S. A. Ghasemzadeh, G. V. Zandycke, M. Istasse, N. Sayez, A. Moshtaghpour, and C. D. Vleeschouwer, "DeepSportLab: a unified framework for ball detection, player instance segmentation and pose estimation in team sports scenes," in *Proceedings of the British Machine Vision Conference 2021*, 2021, doi: 10.5244/C.35.210.
- [18] B. M. A., S. Venkataraman, M. S. Khan, Nidhi, S. Padmanabhuni, and S. Natarajan, "Optimizing semantic segmentation for enhanced football analytics: a pixel-level approach," *Procedia Computer Science*, vol. 235, pp. 2662–2673, 2024, doi: 10.1016/j.procs.2024.04.251.
- [19] B. Ren, X. Zheng, T. Guan, and J. Wang, "Vibrator rack pose estimation for monitoring the vibration quality of concrete using improved YOLOv8-pose and vanishing points," *Buildings*, vol. 14, no. 10, Oct. 2024, doi: 10.3390/buildings14103174.
- [20] C. Dong and G. Du, "An enhanced real-time human pose estimation method based on modified YOLOv8 framework," *Scientific Reports*, vol. 14, no. 1, Apr. 2024, doi: 10.1038/s41598-024-58146-z.
- [21] Adzuki, "Soccer-field-localization computer vision model," universe.roboflow.com. Accessed: Aug. 15, 2025. [Online]. Available: <https://universe.roboflow.com/adzuki/soccer-field-localization>
- [22] Y. Tian, Y. Zhang, and H. Zhang, "Recent advances in stochastic gradient descent in deep learning," *Mathematics*, vol. 11, no. 3, Jan. 2023, doi: 10.3390/math11030682.
- [23] Y. Bai, *et al.*, "How important is the train-validation split in meta-learning?" in *Proceedings of the 38th International Conference on Machine Learning (ICML)-PMLR 139*, 2021, pp. 543–553.
- [24] Z. Jiang, H. Ji, C.-Y. Yang, and J.-N. Hwang, "2D human pose estimation calibration and keypoint visibility classification," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2024, pp. 6095–6099, doi: 10.1109/ICASSP48485.2024.10448474.
- [25] M. Jiono, H.-I. Lin, W.-H. Chen, and S. Nanda, "6-DoF pose estimation using an end-to-end pipeline of neural network with BP-n-P," *IEEE Sensors Journal*, vol. 25, no. 2, pp. 3837–3847, Jan. 2025, doi: 10.1109/JSEN.2024.3505832.
- [26] X. Zhang, J. Bai, G. Qiao, X. Xiao, L. Meng, and S. Hu, "YOLO-fall: a YOLO-based fall detection model with high precision, shrunk size, and low latency," *The Computer Journal*, vol. 68, no. 7, pp. 804–812, Jul. 2025, doi: 10.1093/comjnl/bxaf005.
- [27] OpenCV, "Basic concepts of the homography explained with code," docs.opencv.org. Accessed: Aug. 15, 2025. [Online]. Available: https://docs.opencv.org/4.13.0/d9/dab/tutorial_homography.html
- [28] W. He *et al.*, "A real-time and high precision hardware implementation of RANSAC algorithm for visual SLAM achieving mismatched feature point pair elimination," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 71, no. 11, pp. 5102–5114, Nov. 2024, doi: 10.1109/TCSI.2024.3422082.
- [29] M. Zhang and Z. Zhou, "Speed-accuracy trade-off in lightweight-based hand pose estimation," *The Journal of Supercomputing*, vol. 81, no. 12, Aug. 2025, doi: 10.1007/s11227-025-07699-5.
- [30] OpenCV, "Camera calibration and 3D reconstruction," docs.opencv.org. Accessed: Aug. 15, 2025. [Online]. Available: https://docs.opencv.org/4.x/d9/d0c/group_calib3d.html
- [31] F. Kluger and B. Rosenhahn, "PARSAC: accelerating robust multi-model fitting with parallel sample consensus," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 3, pp. 2804–2812, Mar. 2024, doi: 10.1609/aaai.v38i3.28060.
- [32] T. Zhang, Q. Li, J. Wen, and C. L. P. Chen, "Enhancement and optimisation of human pose estimation with multi-scale spatial attention and adversarial data augmentation," *Information Fusion*, vol. 111, Nov. 2024, doi: 10.1016/j.inffus.2024.102522.
- [33] R. B. Neupane, K. Li, and T. F. Boka, "A survey on deep 3D human pose estimation," *Artificial Intelligence Review*, vol. 58, no. 1, Nov. 2024, doi: 10.1007/s10462-024-11019-3.
- [34] Y. Zhao and P. Krähenbühl, "Real-time online video detection with temporal smoothing transformers," in *Computer Vision—ECCV 2022*, 2022, pp. 485–502, doi: 10.1007/978-3-031-19830-4_28.

BIOGRAPHIES OF AUTHORS



Durrotun Nashihin    is currently a lecturer in Department of Informatics Engineering at State University of Surabaya. He received his Bachelor of Applied Science degree in Computer Engineering from Politeknik Elektronika Negeri Surabaya, Indonesia, in 2018. He also obtained a master's degree in Biosystem Engineering with specialization in Information Technology at Wageningen University and Research (WUR), the Netherlands, in 2023. His research focuses include machine learning, deep learning and big data. He can be contacted at email: durrotunnashihin@unesa.ac.id.



Yohanes Yohanie Fridelin Panduman, Ph.D.,    works as a specially appointed assistant professor in The University of Osaka, Japan. In 2024, he graduated from his Ph.D. degree in Communication and Information Systems at Okayama University, Japan. He also finished his Master of Engineering in Electronic Engineering and his Bachelor of Applied Science in Computer Engineering at Politeknik Elektronika Negeri Surabaya, Indonesia, in 2014 and in 2018, respectively. His research interests consist of internet of things, big data, database systems, and machine learning. He can be contacted at email: yohanes@ist.osaka-u.ac.jp.



Prof. Dr. Dwi Cahyo Kartiko    is currently a professor in Department of Physical Education at State University of Surabaya. In 2015, he obtained his doctor in Sports Science at State University of Surabaya, Indonesia. He also completed his bachelor's degree in Sports Education and his master's degree in Sports Health Science at State University of Surabaya in 1998 and 2005, respectively. His research areas include sport science, sports biomechanics, and athletic performance. He can be contacted at email: dwicahyo@unesa.ac.id.



Dr. Anna Noordia    obtained her doctor of Sports Science from State University of Surabaya, Indonesia, in 2025. In 2024, she also finished her master of Public Health from Airlangga University, Indonesia. Her research areas of interest include sport science, food technology, and sport health. She can be contacted at email: annanoordia@unesa.ac.id.



Khoirul Islam    is currently a lecturer in Informatics Engineering at State University of Surabaya. He finished his Master's degree in Informatics Study Program at Universitas Amikom Yogyakarta, Indonesia, in 2019. In 2025, he also completed his Bachelor's degree in Information Systems Study Program at Universitas Pesantren Tinggi Darul 'Ulum, Indonesia. His research areas include business intelligence and human-AI interaction. He can be contacted at email: khoirulislam@unesa.ac.id.