An End-to-End Pipeline for Virtual Banner Replacement in Football Broadcasts

Victor Gaspar[1*], Anthony Cioppa[1*], Jan Held[1,2], Silvio Giancola[2], Marc Braham[1], Adrien Deliège[1], Bernard Ghanem[2], Marc Van Droogenbroeck[1]
[1] University of Liège, Belgium [2] KAUST

Abstract

Augmented reality has been used in sports broadcasting since the 1990s to enhance viewer engagement through virtual overlays. A key application is virtual advertising, which replaces physical advertisement banners with dynamic digital content, enabling targeted and region-specific advertisements. This technology optimizes advertising space and increases monetization opportunities for broadcasters. However, traditional augmented reality solutions require specialized hardware, such as instrumented cameras and virtual-ready LED panels, along with manual calibration and prior environmental knowledge. These constraints make its implementation costly and less adaptative. In this work, we propose a first fully automated end-to-end pipeline that seamlessly integrates augmented reality advertising into sports broadcasts using only the main camera feed. Our approach leverages state-of-the-art deep neural networks to identify the advertisement banner, estimate camera motion, and dynamically composite virtual content without additional hardware or manual intervention. We validate our pipeline on football broadcasts using our novel SoccerNet-banner dataset, the first dataset for training and evaluating banner segmentation models, and demonstrate high-quality virtual banner replacement on SoccerNet videos. Therefore, our pipeline unlocks new possibilities for personalized content and advances AI-powered sports broadcasting by eliminating hardware dependencies and manual calibration. Our code and dataset are available at `https://github.com/SoccerNet/sn-banner`.

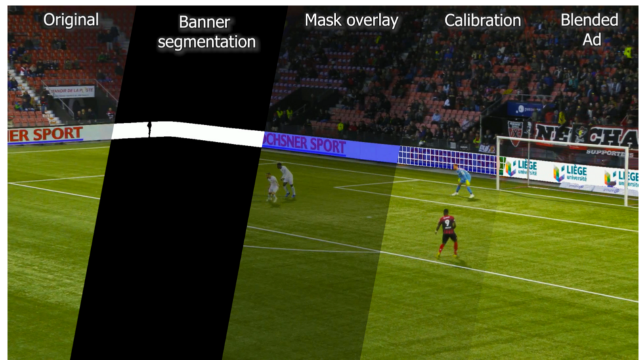

Figure 1. **Steps of banner replacement.** We propose an end-to-end pipeline for banner replacement which includes the steps of banner segmentation, camera calibration for re-projection, and composition of a virtual blended content on the banner.

1. Introduction

Advertising in sports broadcasts represents a significant revenue stream [4], particularly in popular sports such as football or basketball. Traditional advertising methods, like physical banners and billboards, have long been employed to capitalize on viewer engagement. However, these approaches have inherent limitations, including static placement, potential geographical irrelevance, and logistical constraints. Augmented reality-based virtual ads have emerged as a solution, enabling dynamic and region-specific advertisements that adapt to different markets. These techniques allow broadcasters to overlay existing physical advertising panels on the field with virtual advertisements targeted to specific broadcast regions. For example, advertisements can be translated into any language, leading to greater relevance and engagement. Today's virtual advertising solutions leverage advancements in computer vision, real-time camera calibration, and chroma-keying technology to achieve integration [57, 66, 67]. Chroma keying allows a specified color range in the footage to become transparent, enabling the superimposition of virtual content over the affected area.

However, these systems come with notable limitations. Chroma keying, for instance, is highly sensitive to lighting conditions, requiring careful planning and preparation before each game to account for weather conditions, camera angles, the stadium's 3D model, and even players' jersey colors [70, 72, 73]. Another challenge is that they cannot rely solely on the raw camera feed. Even software-driven solutions, which minimize the need for additional hardware,

[*] These authors contributed equally.

still require precise camera calibration to ensure proper alignment and occlusion handling [67]. This calibration process must be repeated for each broadcast setup, making the deployment of virtual advertising solutions much more demanding than traditional static banners.

This paper introduces the first automatic end-to-end pipeline for seamlessly superimposing new content onto dynamically changing advertisement banners in football game videos, relying solely on the main camera feed; the steps of our new pipeline are illustrated in Fig. 1. We begin by segmenting the advertisement banner using semantic segmentation models, trained and evaluated on our new *SoccerBanner* dataset, part of which will be publicly released. We also estimate and track the camera's position and orientation using the known dimensions of the soccer field as a reference for calibration. To refine the camera parameters, we propose a three-filter process: the first estimates missing camera parameters, the second removes outliers, and the third smoothens the parameters to enhance temporal consistency and attenuate high frequencies. Then, we propose a new 3D model to determine the billboards' positions and orientations, allowing the system to adapt to varying stadium layouts. Finally, the extracted data is integrated with compositing tools to overlay new content naturally onto the segmented banners, with an adaptive positioning method and a blending algorithm. Compared to existing methods, our pipeline only requires three inputs: the broadcast video, the image for the new advertisement, and the speed at which the new content moves along the billboards. Our experiments illustrate the effectiveness of our pipeline, paving the way for customizable advertising solutions.

Contributions. We summarize our contributions as follows: **(i)** We introduce the first automatic end-to-end pipeline for virtual banner replacement in football broadcast videos, integrating banner segmentation, camera calibration, and compositing. Our approach achieves high-quality virtual replacements, as demonstrated by compelling qualitative results on test videos. **(ii)** We publicly release the first dedicated banner segmentation dataset, enabling the training and benchmarking of banner segmentation models. **(iii)** We enhance the current state-of-the-art camera calibration method by incorporating temporal smoothing and introducing a novel 3D banner model, leading to improved quantitative performance and visually superior results.

2. Related work

Sports video understanding. Sports have always captivated audiences worldwide, and similarly, the automatic understanding of sports videos has garnered increasing attention in the research community [26]. This growing interest has been driven by the availability of football-centric large-scale annotated datasets [8,13,15,19,20,28,29,45,58] and the recent advances in deep learning for general video understanding. Over the past years, a wide range of tasks were tackled, including player segmentation, detection, and tracking [11, 12, 51, 52, 60, 65, 69], summarizing games [27, 53], player re-identification [52, 64], action spotting in untrimmed videos [5,9,10,30,31,40,61–63,74], pass prediction and feasibility [3, 39], camera calibration [48–50], foul recognition [34–37], or dense video captioning for football broadcasts commentaries [2, 55]. In this work, we leverage sports understanding tool for content generation, by introducing the first pipeline for virtual banner replacement, enabling personalized advertisement placement.

Augmented reality (AR) in sports. Augmented reality has been enhancing sports viewing experiences since the 1990s [70]. An early example is the 1st & Ten system [71], which virtually inserts a yellow first-down line at fixed coordinates onto the field during live American football broadcasts. Recent advancements in deep learning have significantly improved AR's capabilities and realism in sports broadcasting [42, 47]. For example, Rematas *et al*. [59] leverage monocular depth estimation to watch the game on a tabletop with AR glasses. Nowadays, novel-view synthesis methods [38, 43, 54] enable the generation of new, previously inaccessible viewpoints. Lewin *et al*. [46] demonstrated the effectiveness of these techniques in synthesizing dynamic, immersive football scenes. Beyond football, AR has also been utilized in basketball for real-time player analytics [25] or in Formula 1 for enhanced broadcasting through immersive replays, driver performance analytics, and interactive AR storytelling [16, 41]. Complementing these advancements, our work extends AR into the domain of virtual advertising, providing broadcasters the capability to dynamically replace physical advertising panels on the field with virtual advertisements.

Localizing elements in sports broadcasts. Accurately localizing elements in sports broadcasts, such as players [11, 12], ball position [68], and camera positions [49], is fundamental for applications like game analysis and augmented reality overlays. In the context of camera calibration, sports field markings serve as essential geometric references. Methods such as field line detection using Hough transforms provide a reliable approach to extracting these features [1, 21, 24]. To further enhance camera tracking and localization, state-of-the-art methods [22, 23, 32, 33, 48] integrate high-level semantic analysis of sports field markings with broadcast camera modeling, developing a tracking system that achieves impressive performances on the SoccerNet-GSR dataset [65]. In this work, we leverage recent advancements in camera calibration and improve their temporal consistency with low-level signal processing techniques, and train the transformer-based segmentation model Mask2Former [7] on our new SoccerNet-Banner dataset to accurately determine the position and movement of physical banners.

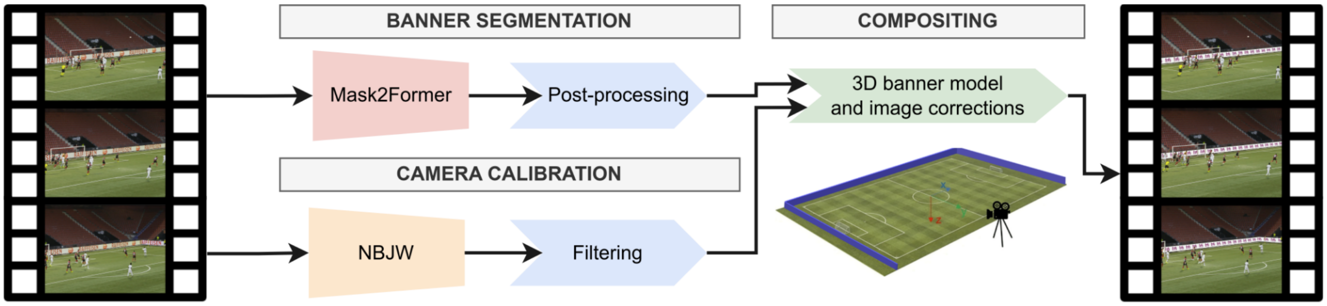

Figure 2. **Overview of our end-to-end pipeline for virtual banner replacement in football broadcast videos.** Our pipeline consists of three main components: (1) Banner segmentation, where Mask2Former segments the banner region, followed by a post-processing step to refine the segmentation map; (2) Camera calibration where *No-Bells-Just-Whistles* (NBJW) estimates the camera parameters, further refined through multi-stage filtering; and (3) Compositing where the new virtual banner is seamlessly integrated into the scene based on the extracted segmentation mask and calibrated camera parameters leveraging a 3D banner model and image correction techniques.

3. Methodology

This section presents our end-to-end pipeline for virtual banner replacement in football broadcasts. Our pipeline is designed to function solely from the camera feed, eliminating the need for additional hardware or manual intervention. It consists of three main modules: (1) segmentation of perimeter banners, (2) camera calibration for spatial alignment, and (3) virtual banner compositing for seamless integration. An overview of our pipeline is illustrated in Fig. 2.

3.1. Problem formulation

Given a football broadcast video sequence $V = \{I_t\}_{t=1}^{T}$, where each frame I_t has dimensions $(H, W, 3)$, the objective is to replace the perimeter banners dynamically while maintaining visual and temporal consistency. The output is a modified sequence $\hat{V}$ where perimeter banners are substituted with virtual advertisements seamlessly blended into the broadcast. This task presents multiple challenges, including non-uniform banner dimensions, occlusions from players, dynamic camera motion, and variations in lighting conditions. Additionally, LED-based banners introduce artifacts that further complicate segmentation and alignment.

3.2. Banner segmentation

Semantic segmentation assigns a class label to each pixel in the image, allowing to extract the precise regions occupied by the banners. To accurately localize the perimeter banners in each frame, we employ state-of-the-art semantic segmentation models. Particularly, we employ the Mask2Former [7] model, a transformer-based segmentation model trained on our SoccerNet-Banner dataset (see Section 4), to classify each pixel as either in one of four classes. This model processes individual frames and generates a 4 binary mask $M_t \in \{0, 1\}^{H \times W \times 4}$, delineating the banners.

Despite the accuracy of deep learning models, incorrect classifications occur in some frames, particularly when non-banner regions such as stadium LED panels or crowd areas share similar colors or textures with banners. To mitigate this, we apply a filtering step that enforces geometric constraints based on expected banner placement. The segmentation map is analyzed using an 8-connectivity grid to extract connected components. Only the largest detected region near the field boundary is retained, leveraging the assumption that the perimeter banner extends across the entire camera view. We further segment objects that occlude the banner to prevent erroneous banner segmentation. This post-processing step significantly improves segmentation robustness, as depicted in Fig. 3.

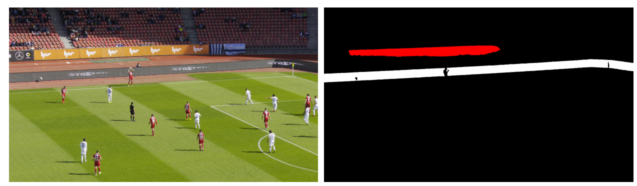

Figure 3. **Illustration of our banner segmentation.** Non-banner regions (in red) might be incorrectly segmented, but our post-processing step filters them out.

3.3. Camera calibration and registration

On top of segmenting the perimeter banners, we estimate the camera parameters required to align virtual advertisements with real-world positions. Since perimeter banner dimensions are not standardized across stadiums, we use the football field as a calibration reference, leveraging its known geometry. Following the *No-Bells-Just-Whistles* (NBJW) method [32], keypoints such as penalty box corners and midfield lines are detected and matched with their corresponding locations in a canonical field model. Using these correspondences, we estimate the homography matrix $H_t \in \mathbb{R}^{3\times3}$ via RANSAC to robustly align field features.

However, estimating an homography frame-by-frame is

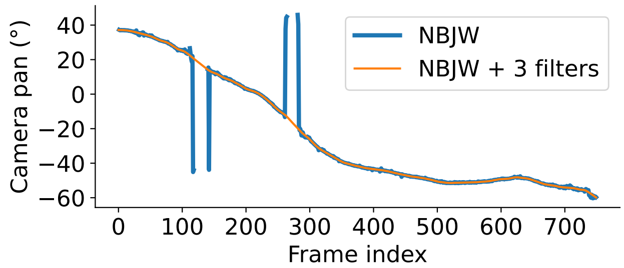

Figure 4. **Camera parameter refinement.** Example evolution of the camera horizontal Euler angle (pan) about the upright Z-axis in video 129 of the SoccerNet-GSR dataset before (blue) and after (orange) applying our three filters.

sensitive to noise and prone to temporal inconsistencies. To improve stability, we refine the estimated camera parameters using a low-level signal processing filtering framework that enforces temporal consistency. First, missing camera parameters in frames where registration fails are interpolated using adjacent frames. Next, a median filtering step detects and removes outlier estimates by evaluating deviations across consecutive frames. Finally, a Savitzky-Golay filter is applied to smooth parameter variations while preserving sharp transitions due to intentional camera movements. This filtering process enhances the temporal stability of the estimated camera parameters, mitigating flickering artifacts in virtual banner placement, as shown in Fig. 4.

3.4. Virtual banner compositing

With the segmentation masks and camera parameters estimated, virtual advertisements are projected onto the detected perimeter banners. A general banner model is constructed by defining three connected rectangular regions, aligned with the field sidelines and constrained to be level with the ground at $Z = 0$. The position and orientation of these rectangles are determined based on the detected banner boundaries and calibrated camera parameters. The banner height is estimated using reference world points and a scale factor computed as:

$$Z_t = Z_r \frac{y_b - y_t}{y_b - y_r}, \quad (1)$$

where Z_r is a reference height, and y_b, y_t correspond to the bottom and top image coordinates of the detected banner.

To compensate for imperfections in segmentation and calibration, polynomial regression is applied per frame to approximate the physical curvature of the banner, reducing distortions introduced by lens effects. The virtual advertisement is then warped using the estimated homography and blended into the scene. To ensure seamless integration, the composited advertisement undergoes a multistep refinement process that includes HSV-based color adaptation to match real-world lighting conditions, antialiasing through upscaling and downscaling, Gaussian blurring to smooth transitions, and masking to restrict overlay to detected banner regions. These steps ensure that the final virtual advertisement is indistinguishable from real-world perimeter banners while maintaining correct perspective alignment within the broadcast scene, as shown in Fig. 5.

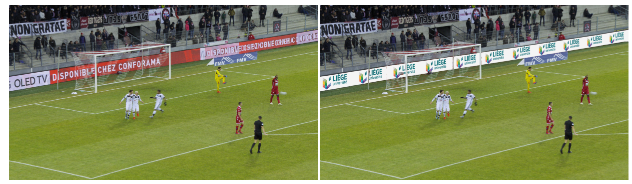

Figure 5. **Compositing results.** Original image (left) and final composited result after Gaussian blurring (right).

4. SoccerNet-banner dataset

We introduce *SoccerNet-Banner*, a novel image dataset of banner segmentation in football broadcasts. It consists of 2,960 Full HD (1920×1080 pixels) RGB images with manually annotated semantic segmentation maps. The images are extracted from the SoccerNet Game State Reconstruction dataset [65] and follow the same split. Particularly, the train, validation, and test splits may include images from the same stadium but never from the same match, reflecting a real-world scenario where a deployed system may have been trained on past matches within a stadium but not on future ones. Each annotation is a single-channel Full HD image where each pixel is assigned to one of four classes: (1) the area outside the advertisement banner, (2) visible advertisement banner regions, (3) goal nets, and (4) occlusions in front of the banner, including for instance field players and referees. Although the goal net allows partial visibility of the billboard behind it, we treat it as a non-occlusion class and preserve this distinction for future improvements. The annotation campaign lasted 2 months and required pixel-level accuracy. Hence, no semi-automated tools were used.

We propose SoccerNet-Banner in two versions: a public one, openly accessible to any researcher, and a private one, with segregated annotation for a future challenge. The public version, is a subset of the private dataset, containing 100 images per split (train, validation, and test), making it a lightweight reproducible benchmark for segmentation models. The private dataset comprises the full 2,960-image collection and is used for training and evaluation in this paper. Let us note that we will publicly release the checkpoints of all models, including those trained on private dataset.

5. Experiments

In this section, we conduct a comprehensive evaluation of our pipeline. First, we present the experimental settings.

Second, we quantitatively assess each module individually and provide some ablation studies. Third, we showcase qualitative results of the full pipeline. Finally, we discuss the limitations of our approach and outline potential directions for future improvements. Videos are provided in supplementary material, alongside the code for reproducibility.

5.1. Experimental settings

Banner segmentation settings. All banner segmentation models are sourced from the open-source project MMSegmentation[1] [17] from OpenMMLab[2] as it provides convenient high-level APIs and enough modularity to configure, train, test, and deploy models easily. On top of Mask2Former [7], we also experiment with alterative segmentation models, including DDRNet [56], PointRend [44], and SegFormer [75]. Those models were selected based on their state-of-the-art mean Intersection over Union (mIoU) score on the Cityscapes dataset [18]. All models are fine-tuned on either our public or private SoccerNet-Banner dataset from a pre-trained ImageNet backbone on a single Tesla V100 GPU for $40{,}000$ iterations using the SGD optimizer with a momentum of 0.9, a weight decay of 5×10^{-4}, an initial learning rate of 10^{-2}, a polynomial decay scheduler with a power of 0.9, an end learning rate of 10^{-4}, and a warmup of 500 iterations. The batch size is set to 1 for all models with accumulated gradients for two iterations to account for GPU memory limitations. Data augmentation is applied during training by random cropping and horizontally flipping. During training, models are evaluated on the validation set every 135 iterations to control overfitting. During inference, Test-Time Augmentation (TTA) from OpenMMLab[3] applies transformations such as flipping and scaling to the input image, generating multiple augmented versions, and merges the predictions. Finally, the results are presented on our public and private test sets, and evaluated using the Boundary IoU [6], mBIoU_d, at different distances d. For small d, the score provides information about the performance near the mask boundaries, *i.e.* its effectiveness in precisely retrieving object boundaries, and for greater distances d, about its global performance to segment all "objects".

Camera calibration settings. We use the publicly available state-of-the-art method *No-Bells-Just-Whistles* (NBJW) [32], with a condition added to use homography estimation instead of 3D camera calibration if the overall root-mean-square re-projection error is higher than 10.0 and if the homography estimation error is lower than the 3D camera calibration error. The model is trained model on the SoccerNet-GSR dataset [65], with the same hyperparameters [14]. Our three temporal filters have only three parameters: the median, Savitzky-Golay filter window lengths, and the polynomial degree. To optimize the filter parameters, we perform a grid search on SoccerNet-GSR [65] and find that values of window lengths of 13 and 23 frames for the median and Savitzky-Golay filters, with a polynomial degree of 2. To evaluate the performance of camera calibration, we use the Jaccard index, JaC_τ introduced by Magera *et al*. [48], with $\tau = 5$.

Compositing settings. For blending, we reduce the saturation and value of the new content image by 4% and 10%, respectively. Then, the warped new content background color is set to the new content's top left pixel color. The new content is then upscaled and downscaled by a factor of 2 to remove aliasing. Finally, to refine the compositing, we apply a smoothing step by blending the original image with its Gaussian-blurred version, using a weight of $\alpha = 0.2$. This further enhances the seamless integration of the virtual content while preserving important visual details of the new advertisement. All these values were determined empirically through visual inspection, but can be adjusted based on specific broadcast setups.

5.2. Quantitative results

For banner segmentation, we compare the results of four segmentation methods. Similarly, we compare the performances of NBJW with and without our camera parameter filtering. Currently, no benchmark exists for evaluating compositing in sports. While certain metrics can assess blending without a dedicated dataset, we only conduct a qualitative evaluation of the composition module.

Banner segmentation results. Table 1 presents the mean Boundary IoU (mBIoU_d) of banner segmentation models trained and tested on both the public and private SoccerNet-Banner datasets. First, we observe that Mask2Former almost consistently outperforms all other models across all distances d and testing settings. As a result, Mask2Former is confirmed as the best choice for banner segmentation, provided real-time performance is not a constraint. Next, we note that the private test set is slightly more challenging than the public test set, as all models exhibit slightly lower performance. This is due to its increased diversity. However, since the performance gap remains small and the ranking of methods remains mostly unchanged, the public test set still serves as a reliable benchmark for model evaluation. Regarding the impact of training data, we observe that training on the private dataset improves performance by approximately 3 points across all methods. This confirms that additional training data enhances model performance, although the public dataset alone remains a viable option. Finally, we find that test-time augmentation (TTA) consistently improves the performance. However, it significantly increases inference time, making it impractical for real-time

[1] https://github.com/open-mmlab/mmsegmentation
[2] https://openmmlab.com/
[3] https://mmengine.readthedocs.io/en/latest/advanced_tutorials/test_time_augmentation.html

			Public test set mBIoU$_d$ [%]					Private test set mBIoU$_d$ [%]				
Method	Training Set	TTA	1	3	5	10	∞	1	3	5	10	∞
Mask2Former [7]	Public	×	15.67	40.96	53.11	65.66	81.59	14.95	38.77	50.26	62.40	78.49
DDRNet [56]	Public	×	13.47	34.59	45.53	57.51	75.94	13.13	33.17	43.47	55.17	73.12
PointRend [44]	Public	×	13.31	35.52	47.05	59.69	77.17	12.93	33.89	44.60	56.58	73.90
SegFormer [75]	Public	×	9.77	28.22	39.56	53.88	76.20	9.39	27.16	38.07	51.97	74.24
Mask2Former [7]	Public	✓	19.18	45.44	56.92	68.33	82.67	18.08	42.69	53.52	64.67	79.39
PointRend [44]	Public	✓	14.97	36.62	47.54	59.74	77.29	14.87	35.65	45.81	57.37	74.59
SegFormer [75]	Public	✓	14.54	36.10	47.23	59.95	78.44	14.22	34.99	45.80	58.56	77.18
Mask2Former [7]	Private	×	19.39	46.05	57.51	68.79	82.63	18.55	43.09	53.60	64.07	78.02
DDRNet [56]	Private	×	15.88	39.96	50.99	62.42	78.22	15.24	38.23	51.17	63.24	77.09
PointRend [44]	Private	×	15.49	41.50	54.09	66.91	81.14	15.17	39.55	51.17	63.24	77.96
SegFormer [75]	Private	×	11.15	33.15	45.87	60.62	79.63	10.62	31.56	43.61	57.71	77.10
Mask2Former [7]	Private	✓	22.17	49.88	61.07	71.70	83.75	20.61	45.69	55.84	65.82	79.09
PointRend [44]	Private	✓	20.00	47.11	58.85	70.42	83.09	19.14	44.01	54.95	66.12	79.46
SegFormer [75]	Private	✓	16.14	40.57	52.50	65.29	81.48	15.55	38.56	49.83	62.15	78.73

Table 1. **Quantitative results and ablation study of our banner segmentation module.** Comparison of banner segmentation performance across different methods, training sets (public vs. private), and the impact of test-time augmentation (TTA). Results are reported as mean Boundary IoU (mBIoU$_d$) at different distances ($d = 1, 3, 5, 10, \infty$) for both our public and private test sets. The inclusion of TTA (checkmark) as well as training on the larger private dataset generally improves performance across all methods and test sets.

Model	Completeness [%]	JaC$_5$ [%]	Final Score [%]
NBJW	93.65	38.17	35.74
NBJW + 1 filter	100.00	35.88	35.88
NBJW + 2 filters	100.00	36.46	36.46
NBJW + 3 filters	100.00	38.46	38.46

Table 2. **Comparison of camera calibration performance using different levels of filtering.** The completeness metric reaches 100% after applying the first filter (linear filter), while additional filtering improves the final score. The best performance is obtained using all three filters in sequence: Filter 1 is a linear filter for interpolates missing values, Filter 2 is a median filter for removing outliers, and Filter 3 is a Savitzky-Golay filter for preserving local trends while refining the signal.

applications. Therefore, TTA should be reserved for offline scenarios where inference speed is not critical.

Camera calibration results. As shown in Tab. 2, while the JaC score slightly decreases with the first *linear interpolation filter*, it effectively recovers all missing camera parameters, leading to a small improvement in the final score. As expected, the completeness rate reaches 100% after applying the linear interpolation, whose objective is to join abrupt variations by averaging neighboring values, reducing noise while preserving the general trend of the camera parameters. Adding a second *median filter* replaces each value with the median of its surrounding values, making it effective at removing outliers without blurring sharp transitions. Although this step provides only a marginal improvement over using the linear filter alone, it ensures that extreme deviations do not interfere with calibration. The most significant performance boost comes when introducing the third *Savitzky-Golay filter* which fits a polynomial to local windows of data to perform smoothing while maintaining local trends and fine details. This three-stage filtering approach demonstrates that properly handling missing values and outliers before applying a final smoothing step significantly improves the camera calibration final score. The best results are obtained when all three filters are applied sequentially, leading to more stable and precise camera parameters.

5.3. Qualitative results

We qualitatively evaluate our pipeline on videos from the challenge set, emphasizing on the edge cases to show its limitations. To produce the following qualitative results, we use the Mask2Former banner segmentation model trained on the private SoccerNet-Banner dataset with test-time augmentation (TTA) and the three filters for camera calibration. Video samples are provided in supplementary material.

Banner segmentation qualitative results. Figure 6 showcases the performance of Mask2Former in segmenting the visible perimeter banner. In well-functioning cases, the model accurately delineates the banner while preserving fine details near occlusions, even handling blurry obstructions caused by fast-moving players. However, failure cases include misclassification of painted ground ads as banners and occasional confusion between a player's body parts and the banner itself, particularly for goalkeepers. Additionally, segmentation masks tend to be slightly wider than the actual

Figure 6. **Qualitative banner segmentation results.** Examples of successful (top) and erroneous (bottom) segmentation of the banner. Failures include segmentation of printed ads on the field (left) and incorrect occlusion handling of the goalkeeper's head (right).

banner, creating a subtle but perceptible halo effect where the physical banner content bleeds into occluded areas.

Camera calibration qualitative results. The camera calibration pipeline effectively aligns the new content with the physical banner in most scenarios, ensuring temporal stability and accurate placement during camera motion. The three-stage filtering approach (linear, median, and Savitzky-Golay filters) reduces flickering and corrects missing or outlier camera parameters. Figure 7 provides a visual comparison of the initial and filtered camera parameters. The red lines represent the projected soccer pitch lines before filtering, while the blue lines depict the same lines after filtering. The green points indicate the projected soccer corners based on the filtered camera parameters. This visualization highlights how the filtering process stabilizes parameter estimation, leading to more consistent virtual banner placement. However, when prolonged periods of missing or erroneous camera parameters occur, particularly at the start or end of videos, filtering alone is insufficient for correction, resulting in misaligned advertisements as shown in Fig. 8. Additionally, rapid changes in camera movement direction may cause momentary desynchronization between the overlay and the physical banner, particularly when the Savitzky-Golay filter smooths sharp motion transitions excessively.

Compositing qualitative results. The compositing step generally produces seamless virtual banner replacement, as demonstrated in Fig. 5. The polynomial-based fitting method successfully adapts to slight distortions and variations in banner shape, ensuring consistent overlay alignment. However, compositing failures occur when the banner intersects with the image border, causing the overlay to appear compressed as shown in Fig. 9 (top). Additionally, the blending technique does not dynamically adjust to weather conditions, sometimes resulting in unrealistic integration under varying lighting conditions as displayed in Fig. 9 (bottom). Other artifacts include aliasing effects when the banner is small in the frame and sharp, stair-like transitions at segmentation mask boundaries.

Discussions on potential improvements. Overall, the pipeline achieves seamless banner replacement in most conditions, maintaining alignment and temporal consistency. When assumptions hold, the results are visually convincing, even under camera motion. However, failure cases arise when segmentation mislabels banner regions, camera calibration lacks sufficient data for correction, or compositing artifacts emerge due to environmental variations. Despite these challenges, the method generalizes well across different stadiums and conditions, making it a practical solution for virtual advertising applications. Yet, several aspects can be enhanced. First, segmentation errors, such as misclassifications of ground ads and minor occlusions, could be reduced by incorporating temporal consistency mechanisms or refining the segmentation model with additional occlusion-specific training data. Second, camera calibration

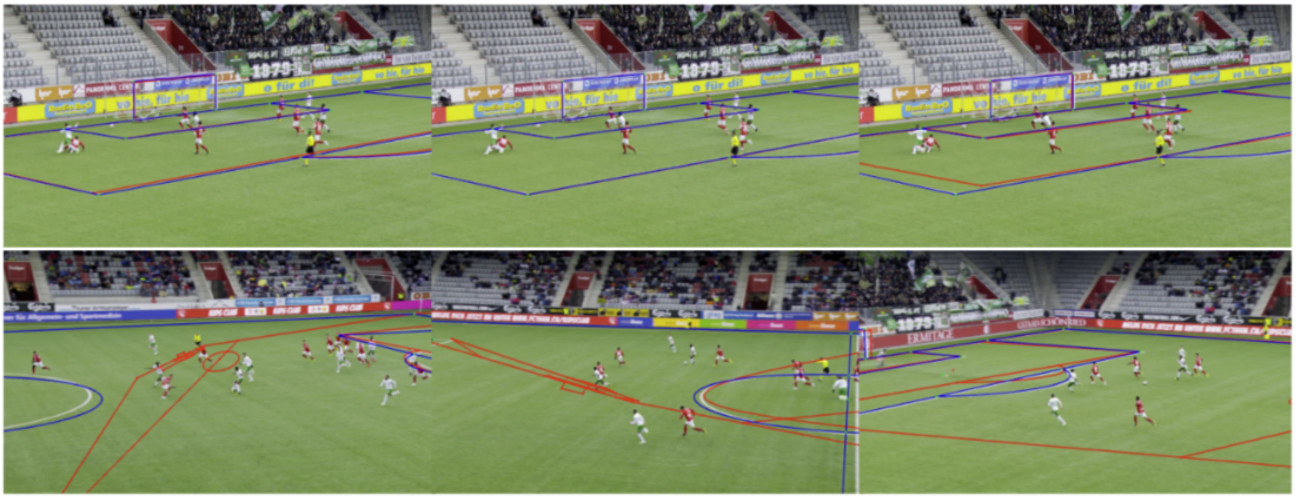

Figure 7. **Visualization of initial (red) versus filtered camera parameters (blue) for two sequences (each row).** The red lines correspond to the projected soccer pitch lines using the initial NBJW camera parameters. Blue lines show the filtered projection based on the filtered parameters. The bottom row depicts failing initial camera parameters (red), yet properly refined (blue) by our filters.

Figure 8. **Qualitative camera calibration error** leading to wrongly aligned advertisement in the banner.

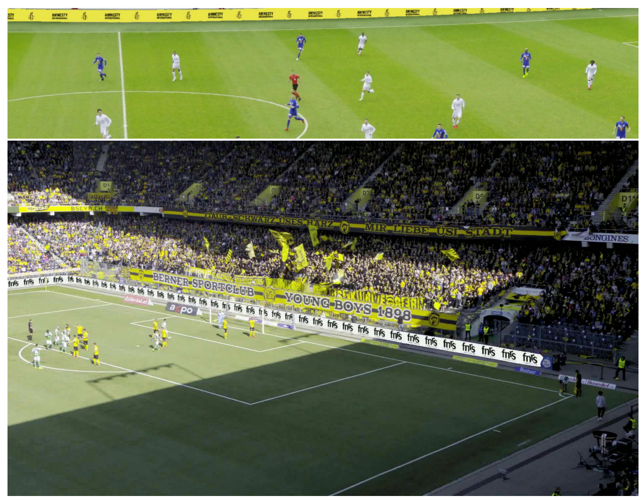

Figure 9. **Qualitative compositing limitation.** Top: Overlay distortion due to banner intersection with the image border. Bottom: Compositing failure due to a sharp shadow.

stability could be improved with more robust filtering techniques such as BroadTrack [49], a recent camera parameter tracking method that handles missing and outlier values more effectively. Third, compositing could benefit from adaptive blending techniques that adjust to varying lighting conditions and incorporate edge refinement to smooth stair-like segmentation boundaries. Additionally, addressing aliasing artifacts when the banner is small in the frame would further enhance visual quality. Future work could explore end-to-end learning approaches that jointly optimize segmentation, calibration, and compositing for a fully automated, real-time system suitable for live broadcasts.

6. Conclusion

In this work, we introduced the first end-to-end automatic pipeline for virtual banner replacement in football broadcast videos, integrating banner segmentation, camera calibration, and compositing. We proposed a novel segmentation approach and released a dedicated SoccerNet-Banner dataset to facilitate future research in this domain. Our improved camera calibration method, leveraging multi-stage filtering, demonstrated enhanced temporal consistency. Extensive quantitative and qualitative evaluations validated the effectiveness of our approach, showing that Mask2Former achieves the best segmentation performance, additional training data improves results, and multi-filtering significantly refines camera calibration. We hope this work paves the way for robust and scalable solutions in virtual advertising for sports broadcasts.

Acknowledgments. J. Held and A. Deliège are funded by the F.R.S.-FNRS. This work was partly funded, initially, by the European VIBBRE Project from the Eurostars programme (grant No. 12347). This work is supported by the KAUST Center of Excellence for Generative AI under award number 5940.

References

[1] Miguel. Alemán-Flores, Luis Alvarez, Luis Gómez, Pedro Henriquez, and Luis Mazorra. Camera calibration in sport event scenarios. *Pattern Recognit.*, 47(1):89–95, Jan. 2014.

[2] Peter Andrews, Oda Elise Nordberg, Stephanie Zubicueta Portales, Njål Borch, Frode Guribye, Kazuyuki Fujita, and Morten Fjeld. AiCommentator: A multimodal conversational agent for embedded visualization in football viewing. In *Int. Conf. Intell. User Interfaces*, page 14–34, Greenville, SC, USA, Mar. 2024.

[3] Adrià Arbués Sangüesa, Adriàn Martín, Javier Fernández, Coloma Ballester, and Gloria Haro. Using player's body-orientation to model pass feasibility in soccer. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 3875–3884, Seattle, WA, USA, Jun. 2020.

[4] Associated Press. Why advertisers are paying record prices to get your attention during this year's Super Bowl. https://apnews.com/article/super-bowl-ads-commercials-pepsi-bud-fox-04ec2eddd72bb7569abad6afc1475655, Jan. 2025.

[5] Bruno Cabado, Anthony Cioppa, Silvio Giancola, Andrés Villa, Bertha Guijarro-Berdiñas, Emilio J. Padrón, Bernard Ghanem, and Marc Van Droogenbroeck. Beyond the Premier: Assessing action spotting transfer capability across diverse domains. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, volume 27, pages 3386–3398, Seattle, WA, USA, Jun. 2024.

[6] Bowen Cheng, Ross Girshick, Piotr Dollar, Alexander C. Berg, and Alexander Kirillov. Boundary IoU: Improving object-centric image segmentation evaluation. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 15329–15337, Nashville, TN, USA, Jun. 2021.

[7] Bowen Cheng, Ishan Misra, Alexander G. Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 1280–1289, New Orleans, LA, USA, Jun. 2022.

[8] Anthony Cioppa, Adrien Deliège, Silvio Giancola, Bernard Ghanem, and Marc Van Droogenbroeck. Scaling up SoccerNet with multi-view spatial localization and re-identification. *Sci. Data*, 9(1):1–9, Jun. 2022.

[9] Anthony Cioppa, Adrien Deliège, Silvio Giancola, Bernard Ghanem, Marc Van Droogenbroeck, Rikke Gade, and Thomas B. Moeslund. A context-aware loss function for action spotting in soccer videos. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 13123–13133, Seattle, WA, USA, Jun. 2020.

[10] Anthony Cioppa, Adrien Deliège, Silvio Giancola, Floriane Magera, Olivier Barnich, Bernard Ghanem, and Marc Van Droogenbroeck. Camera calibration and player localization in SoccerNet-v2 and investigation of their representations for action spotting. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW), CVsports*, pages 4532–4541, Nashville, TN, USA, Jun. 2021.

[11] Anthony Cioppa, Adrien Deliege, Maxime Istasse, Christophe De Vleeschouwer, and Marc Van Droogenbroeck. ARTHuS: Adaptive real-time human segmentation in sports through online distillation. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW), CVsports*, pages 2505–2514, Long Beach, CA, USA, Jun. 2019.

[12] Anthony Cioppa, Adrien Deliège, and Marc Van Droogenbroeck. A bottom-up approach based on semantics for the interpretation of the main camera stream in soccer games. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW), CVsports*, pages 1846–1855, Salt Lake City, UT, USA, Jun. 2018.

[13] Anthony Cioppa, Silvio Giancola, Adrien Deliege, Le Kang, Xin Zhou, Zhiyu Cheng, Bernard Ghanem, and Marc Van Droogenbroeck. SoccerNet-tracking: Multiple object tracking dataset and benchmark in soccer videos. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW), CVsports*, pages 3490–3501, New Orleans, LA, USA, Jun. 2022.

[14] Anthony Cioppa, Silvio Giancola, Vladimir Somers, Victor Joos, Floriane Magera, Jan Held, Seyed Abolfazl Ghasemzadeh, Xin Zhou, Karolina Seweryn, Mateusz Kowalczyk, Zuzanna Mróz, Szymon Łukasik, Michał Hałoń, Hassan Mkhallati, Adrien Deliège, Carlos Hinojosa, Karen Sanchez, Amir M. Mansourian, Pierre Miralles, Olivier Barnich, Christophe De Vleeschouwer, Alexandre Alahi, Bernard Ghanem, Marc Van Droogenbroeck, Adam Gorski, Albert Clapés, Andrei Boiarov, Anton Afanasiev, Artur Xarles, Atom Scott, ByoungKwon Lim, Calvin Yeung, Cristian Gonzalez, Dominic Rüfenacht, Enzo Pacilio, Fabian Deuser, Faisal Sami Altawijri, Francisco Cachón, HanKyul Kim, Haobo Wang, Hyeonmin Choe, Hyunwoo J. Kim, Il-Min Kim, Jae-Mo Kang, Jamshid Tursunboev, Jian Yang, Jihwan Hong, Jimin Lee, Jing Zhang, Junseok Lee, Kexin Zhang, Konrad Habel, Licheng Jiao, Linyi Li, Marc Gutiérrez-Pérez, Marcelo Ortega, Menglong Li, Milosz Lopatto, Nikita Kasatkin, Nikolay Nemtsev, Norbert Oswald, Oleg Udin, Pavel Kononov, Pei Geng, Saad Ghazai Alotaibi, Sehyung Kim, Sergei Ulasen, Sergio Escalera, Shanshan Zhang, Shuyuan Yang, Sunghwan Moon, Thomas B. Moeslund, Vasyl Shandyba, Vladimir Golovkin, Wei Dai, WonTaek Chung, Xinyu Liu, Yongqiang Zhu, Youngseo Kim, Yuan Li, Yuting Yang, Yuxuan Xiao, Zehua Cheng, and Zhihao Li. SoccerNet 2024 challenges results. *arXiv*, abs/2409.10587, 2024.

[15] Anthony Cioppa, Silvio Giancola, Vladimir Somers, Floriane Magera, Xin Zhou, Hassan Mkhallati, Adrien Deliège, Jan Held, Carlos Hinojosa, Amir M. Mansourian, Pierre Miralles, Olivier Barnich, Christophe De Vleeschouwer, Alexandre Alahi, Bernard Ghanem, Marc Van Droogenbroeck, Abdullah Kamal, Adrien Maglo, Albert Clapés, Amr Abdelaziz, Artur Xarles, Astrid Orcesi, Atom Scott, Bin Liu, Byoungkwon Lim, Chen Chen, Fabian Deuser,

Feng Yan, Fufu Yu, Gal Shitrit, Guanshuo Wang, Gyusik Choi, Hankyul Kim, Hao Guo, Hasby Fahrudin, Hidenari Koguchi, Håkan Ardö, Ibrahim Salah, Ido Yerushalmy, Iftikar Muhammad, Ikuma Uchida, Ishay Be'ery, Jaonary Rabarisoa, Jeongae Lee, Jiajun Fu, Jianqin Yin, Jinghang Xu, Jongho Nang, Julien Denize, Junjie Li, Junpei Zhang, Juntae Kim, Kamil Synowiec, Kenji Kobayashi, Kexin Zhang, Konrad Habel, Kota Nakajima, Licheng Jiao, Lin Ma, Lizhi Wang, Luping Wang, Menglong Li, Mengying Zhou, Mohamed Nasr, Mohamed Abdelwahed, Mykola Liashuha, Nikolay Falaleev, Norbert Oswald, Qiong Jia, Quoc-Cuong Pham, Ran Song, Romain Hérault, Rui Peng, Ruilong Chen, Ruixuan Liu, Ruslan Baikulov, Ryuto Fukushima, Sergio Escalera, Seungcheon Lee, Shimin Chen, Shouhong Ding, Taiga Someya, Thomas B. Moeslund, Tianjiao Li, Wei Shen, Wei Zhang, Wei Li, Wei Dai, Weixin Luo, Wending Zhao, Wenjie Zhang, Xinquan Yang, Yanbiao Ma, Yeeun Joo, Yingsen Zeng, Yiyang Gan, Yongqiang Zhu, Yujie Zhong, Zheng Ruan, Zhiheng Li, Zhijian Huang, and Ziyu Meng. SoccerNet 2023 challenges results. *Sports Eng.*, 27(2):1–18, July 2024.

[16] Filip Cleeren. AI replays and more augmented reality: What's new for F1's TV coverage in 2023. https://www.motorsport.com/f1/news/ai-replays-and-more-augmented-reality-whats-new-for-f1s-tv-coverage-in-2023/10439326/, Mar. 2023.

[17] MMSegmentation Contributors. MMSegmentation: OpenMMLab semantic segmentation toolbox and benchmark. https://github.com/open-mmlab/mmsegmentation, 2020.

[18] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 3213–3223, Las Vegas, NV, USA, Jun. 2016.

[19] Mohamad Dalal, Artur Xarles, Sivlio Giancola, Anthony Cioppa, Sergio Escalera, Thomas B. Moeslund, Marc Van Droogenbroeck, Bernard Ghanem, and Clapés Albert. Action anticipation from soccernet football video broadcasts. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, Nashville, TN, USA, Jun. 2025.

[20] Adrien Deliège, Anthony Cioppa, Silvio Giancola, Meisam J. Seikavandi, Jacob V. Dueholm, Kamal Nasrollahi, Bernard Ghanem, Thomas B. Moeslund, and Marc Van Droogenbroeck. SoccerNet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 4503–4514, Nashville, TN, USA, Jun. 2021.

[21] Elan Dubrofsky and Robert J. Woodham. Combining line and point correspondences for homography estimation. In *Adv. Vis. Comput.*, volume 5339 of *Lect. Notes Comput. Sci.*, pages 202–213, 2008.

[22] Nikolay Falaleev. Top-1 solution of SoccerNet camera calibration challenge 2023. https://github.com/NikolasEnt/soccernet-calibration-sportlight, Jun. 2023.

[23] Nikolay S. Falaleev and Ruilong Chen. Enhancing soccer camera calibration through keypoint exploitation. In *Int. ACM Work. Multimedia Content Anal. Sports (MMSports)*, volume 6, pages 65–73, Melbourne, Victoria, Aust., Oct. 2024.

[24] Dirk Farin, Susanne Krabbe, Peter H. N. de With, and Wolfgang Effelsberg. Robust camera calibration for sport videos using court models. In *Storage and Retrieval Methods and Applications for Multimedia*, volume 5307 of *Proc. SPIE*, pages 1–12, San Jose, CA, USA, 2004.

[25] Eric Fisher. The NBA is wisely opening up its SportVU data to fans and media outlets. https://www.sportsbusinessjournal.com/Daily/Issues/2016/01/20/Technology/the-nba-is-wisely-opening-up-its-sportvu-data-to-fans-and-media-outlets/, Jan. 2016.

[26] Keisuke Fujii. *Machine Learning in Sports: Open Approach for Next Play Analytics*. Springerbriefs Comput. Sci. Springer Nat. Singap., 2025.

[27] Sushant Gautam, Cise Midoglu, Saeed Shafiee Sabet, Dinesh Baniya Kshatri, and Pål Halvorsen. Soccer game summarization using audio commentary, metadata, and captions. In *Work. User-centric Narrat. Summ. Long Videos (narsum)*, pages 13–22, Lisboa Portugal, Oct. 2022.

[28] Silvio Giancola, Mohieddine Amine, Tarek Dghaily, and Bernard Ghanem. SoccerNet: A scalable dataset for action spotting in soccer videos. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 1792–179210, Salt Lake City, UT, USA, Jun. 2018.

[29] Silvio Giancola, Anthony Cioppa, Adrien Deliège, Floriane Magera, Vladimir Somers, Le Kang, Xin Zhou, Olivier Barnich, Christophe De Vleeschouwer, Alexandre Alahi, Bernard Ghanem, Marc Van Droogenbroeck, Abdulrahman Darwish, Adrien Maglo, Albert Clapés, Andreas Luyts, Andrei Boiarov, Artur Xarles, Astrid Orcesi, Avijit Shah, Baoyu Fan, Bharath Comandur, Chen Chen, Chen Zhang, Chen Zhao, Chengzhi Lin, Cheuk-Yiu Chan, Chun Chuen Hui, Dengjie Li, Fan Yang, Fan Liang, Fang Da, Feng Yan, Fufu Yu, Guanshuo Wang, H. Anthony Chan, He Zhu, Hongwei Kan, Jiaming Chu, Jianming Hu, Jianyang Gu, Jin Chen, João V. B. Soares, Jonas Theiner, Jorge De Corte, José Henrique Brito, Jun Zhang, Junjie Li, Junwei Liang, Leqi Shen, Lin Ma, Lingchi Chen, Miguel Santos Marques, Mike Azatov, Nikita Kasatkin, Ning Wang, Qiong Jia, Quoc Cuong Pham, Ralph Ewerth, Ran Song, Rengang Li, Rikke Gade, Ruben Debien, Runze Zhang, Sangrok Lee, Sergio Escalera, Shan Jiang, Shigeyuki Odashima, Shimin Chen, Shoichi Masui, Shouhong Ding, Sin-wai Chan, Siyu Chen, Tallal El-Shabrawy, Tao He, Thomas B. Moeslund, Wan-Chi Siu, Wei Zhang, Wei Li, Xiangwei Wang, Xiao Tan, Xiaochuan Li, Xiaolin Wei, Xiaoqing Ye, Xing Liu, Xinying Wang, Yandong Guo, Yaqian Zhao, Yi Yu, Yingying Li, Yue He, Yujie Zhong, Zhenhua Guo, and Zhiheng Li. SoccerNet 2022 challenges results. In *Int. ACM Work. Multimedia Content Anal. Sports (MMSports)*, pages 75–86, Lisbon, Port., Oct. 2022.

[30] Silvio Giancola, Anthony Cioppa, Julia Georgieva, Johsan Billingham, Andreas Serner, Kerry Peek, Bernard Ghanem, and Marc Van Droogenbroeck. Towards active learning for action spotting in association football videos. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 5098–5108, Vancouver, Can., Jun. 2023.

[31] Silvio Giancola, Anthony Cioppa, Bernard Ghanem, and Marc Van Droogenbroeck. Deep learning for action spotting in association football videos. *arXiv*, abs/2410.01304, 2024.

[32] Marc Gutiérrez-Pérez and Antonio Agudo. No bells, just whistles: Sports field registration by leveraging geometric properties. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 3325–3334, Seattle, WA, USA, Jun. 2024.

[33] Marc Gutiérrez-Pérez and Antonio Agudo. PnLCalib: Sports field registration via points and lines optimization. *arXiv*, abs/2404.08401, 2024.

[34] Jan Held, Anthony Cioppa, Silvio Giancola, Elaf Almahmoud, Katherine M. Collins, Umang Bhatt, Bernard Ghanem, and Marc Van Droogenbroeck. Enhancing football refereeing with AI: VARS and X-VARS for assisted decision-making. In *MathSport Conference*, Luxembourg, Jun. 2025.

[35] Jan Held, Anthony Cioppa, Silvio Giancola, Abdullah Hamdi, Christel Devue, Bernard Ghanem, and Marc Van Droogenbroeck. Towards AI-powered video assistant referee system (VARS) for association football. *arXiv*, abs/2407.12483, 2024.

[36] Jan Held, Anthony Cioppa, Silvio Giancola, Abdullah Hamdi, Bernard Ghanem, and Marc Van Droogenbroeck. VARS: Video assistant referee system for automated soccer decision making from multiple views. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 5086–5097, Vancouver, Can., Jun. 2023.

[37] Jan Held, Hani Itani, Anthony Cioppa, Silvio Giancola, Bernard Ghanem, and Marc Van Droogenbroeck. X-VARS: Introducing explainability in football refereeing with multi-modal large language models. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 3267–3279, Seattle, WA, USA, Jun. 2024.

[38] Jan Held, Renaud Vandeghen, Abdullah Hamdi, Adrien Deliège, Anthony Cioppa, Silvio Giancola, Andrea Vedaldi, Bernard Ghanem, and Marc Van Droogenbroeck. 3D convex splatting: Radiance field rendering with 3D smooth convexes. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2025.

[39] Yutaro Honda, Rei Kawakami, Ryota Yoshihashi, Kenta Kato, and Takeshi Naemura. Pass receiver prediction in soccer using video and players' trajectories. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 3502–3511, New Orleans, LA, USA, Jun. 2022.

[40] James Hong, Haotian Zhang, Michaël Gharbi, Matthew Fisher, and Kayvon Fatahalian. Spotting temporally precise, fine-grained events in video. In *Eur. Conf. Comput. Vis. (ECCV)*, volume 13695 of *Lect. Notes Comput. Sci.*, pages 33–51, Tel Aviv, Israël, 2022.

[41] immersiv.io. Augmented reality live experience with the F1 & TF1. https://www.immersiv.io/portfolio/formula-1-tf1-ar/.

[42] Intel. Intel True View technology. https://www.intel.com/content/www/us/en/sports/technology/true-view.html, 2025.

[43] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):1–14, Jul. 2023.

[44] Alexander Kirillov, Yuxin Wu, Kaiming He, and Ross Girshick. PointRend: Image segmentation as rendering. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 9796–9805, Seattle, WA, USA, Jun. 2020.

[45] Arnaud Leduc, Anthony Cioppa, Silvio Giancola, Bernard Ghanem, and Marc Van Droogenbroeck. SoccerNet-Depth: a scalable dataset for monocular depth estimation in sports videos. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, volume 12, pages 3280–3282, Seattle, WA, USA, Jun. 2024.

[46] Sacha Lewin, Maxime Vandegar, Thomas Hoyoux, Olivier Barnich, and Gilles Louppe. Dynamic NeRFs for soccer scenes. In *Int. ACM Work. Multimedia Content Anal. Sports (MMSports)*, pages 113–121, Ottawa, Ontario, Can., Oct. 2023.

[47] Tica Lin, Zhutian Chen, Yalong Yang, Daniele Chiappalupi, Johanna Beyer, and Hanspeter Pfister. The quest for omnioculars: Embedded visualization for augmenting basketball game viewing experiences. *arXiv*, abs/2209.00202, 2022.

[48] Floriane Magera, Thomas Hoyoux, Olivier Barnich, and Marc Van Droogenbroeck. A universal protocol to benchmark camera calibration for sports. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 3335–3346, Seattle, WA, USA, Jun. 2024.

[49] Floriane Magera, Thomas Hoyoux, Olivier Barnich, and Marc Van Droogenbroeck. BroadTrack: Broadcast camera tracking for soccer. In *IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, pages 6177–6187, Tucson, AZ, USA, Feb. 2025.

[50] Floriane Magera, Thomas Hoyoux, Martin Castin, Olivier Barnich, Anthony Cioppa, and Marc Van Droogenbroeck. Can geometry save central views for sports field registration? In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, Nashville, TN, USA, Jun. 2025.

[51] Adrien Maglo, Astrid Orcesi, and Quoc-Cuong Pham. Efficient tracking of team sport players with few game-specific annotations. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 3460–3470, New Orleans, LA, USA, Jun. 2022.

[52] Amir M. Mansourian, Vladimir Somers, Christophe De Vleeschouwer, and Shohreh Kasaei. Multi-task learning for joint re-identification, team affiliation, and role classification for sports visual tracking. In *Int. ACM Work. Multimedia Content Anal. Sports (MMSports)*, page 103–112, Ottawa, Ontario, Can., Oct. 2023.

[53] Cise Midoglu, Saeed Shafiee Sabet, Mehdi Houshmand Sarkhoosh, Mohammad Majidi, Sushant Gautam,

Håkon Maric Solberg, Tomas Kupka, and Pål Halvorsen. AI-based sports highlight generation for social media. In *Proceedings of the Mile-High Video Conference*, page 7–13, Denver, Colorado, USA, Feb. 2024.

[54] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF. *Commun. ACM*, 65(1):99–106, Dec. 2021.

[55] Hassan Mkhallati, Anthony Cioppa, Silvio Giancola, Bernard Ghanem, and Marc Van Droogenbroeck. SoccerNet-caption: Dense video captioning for soccer broadcasts commentaries. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 5074–5085, Vancouver, Can., Jun. 2023.

[56] Huihui Pan, Yuanduo Hong, Weichao Sun, and Yisong Jia. Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes. *IEEE Trans. Intell. Transp. Syst.*, 24(3):3448–3460, Mar. 2023.

[57] PTF Lab. What is virtual advertising & how it's changing the game? https://ptf-lab.com/tpost/bhhal5ggy1-what-is-virtual-advertising-how-its-chan, Nov. 2024.

[58] Jiayuan Rao, Haoning Wu, Hao Jiang, Ya Zhang, Yanfeng Wang, and Weidi Xie. Towards universal soccer video understanding. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2025.

[59] Konstantinos Rematas, Ira Kemelmacher-Shlizerman, Brian Curless, and Steve Seitz. Soccer on your tabletop. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 4738–4747, Salt Lake City, UT, USA, Jun. 2018.

[60] Karolina Seweryn, Gabriel Cheć, Szymon Łukasik, and Anna Wróblewska. Improving object detection quality in football through super-resolution techniques. *arXiv*, abs/2402.00163, 2024.

[61] Karolina Seweryn, Anna Wróblewska, and Szymon Łukasik. Survey of action recognition, spotting and spatio-temporal localization in soccer – current trends and research perspectives. *arXiv*, abs/2309.12067, 2023.

[62] João V. B. Soares and Avijit Shah. Action spotting using dense detection anchors revisited: Submission to the SoccerNet challenge 2022. *arXiv*, abs/2206.07846, 2022.

[63] João V. B. Soares, Avijit Shah, and Topojoy Biswas. Temporally precise action spotting in soccer videos using dense detection anchors. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 2796–2800, Bordeaux, France, Oct. 2022.

[64] Vladimir Somers, Christophe De Vleeschouwer, and Alexandre Alahi. Body part-based representation learning for occluded person Re-Identification. In *IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, pages 1613–1623, Waikoloa, HI, USA, Jan. 2023.

[65] Vladimir Somers, Victor Joos, Anthony Cioppa, Silvio Giancola, Seyed Abolfazl Ghasemzadeh, Floriane Magera, Baptiste Standaert, Amir M. Mansourian, Xin Zhou, Shohreh Kasaei, Bernard Ghanem, Alexandre Alahi, Marc Van Droogenbroeck, and Christophe De Vleeschouwer. SoccerNet game state reconstruction: End-to-end athlete tracking and identification on a minimap. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 3293–3305, Seattle, WA, USA, Jun. 2024.

[66] Supponor. Making virtual advertising a reality. https://supponor.com/air/, 2025.

[67] uniqFEED. Virtual advertising solutions for football. https://www.uniqfeed.com/our-solutions/football, 2025.

[68] Gabriel Van Zandycke and Christophe De Vleeschouwer. 3D ball localization from a single calibrated image. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 3471–3479, New Orleans, LA, USA, Jun. 2022.

[69] Renaud Vandeghen, Anthony Cioppa, and Marc Van Droogenbroeck. Semi-supervised training to improve player and ball detection in soccer. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW), CVsports*, pages 3480–3489, New Orleans, LA, USA, Jun. 2022.

[70] Vox. How the NFL's magic yellow line works. YouTube video: https://www.youtube.com/watch?v=1Oqm6eO6deU, Feb. 2016.

[71] Wikipedia. 1st & Ten (graphics system). https://en.wikipedia.org/w/index.php?title=1st_%26_Ten_(graphics_system), 2024.

[72] Wikipedia. Chroma key. https://en.wikipedia.org/w/index.php?title=Chroma_key, 2024.

[73] Wikipedia. FoxTrax. https://en.wikipedia.org/w/index.php?title=FoxTrax, 2024.

[74] Artur Xarles, Sergio Escalera, Thomas B. Moeslund, and Albert Clapés. T-DEED: Temporal-discriminability enhancer encoder-decoder for precise event spotting in sports videos. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 3410–3419, Seattle, WA, USA, Jun. 2024.

[75] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. SegFormer: Simple and efficient design for semantic segmentation with transformers. In *Adv. Neural Inf. Process. Syst. (NeurIPS)*, volume 34, pages 12077–12090, 2021.