Multimedia Tools and Applications (2026) 85:611
https://doi.org/10.1007/s11042-026-21756-y

Occlusion-aware advertisement placement in soccer penalty area

Sukriti Dhang[1,4] · Fucheng Zheng[2] · Peter Han Joo Chong[2] · Mimi Zhang[3] · Soumyabrata Dev[1,3,4]

Received: 5 December 2025 / Revised: 15 May 2026 / Accepted: 22 June 2026

Abstract

Advert placement in sports broadcasts is a growing strategy to boost sponsor visibility without disrupting live gameplay. Achieving realism, however, requires careful handling of scene geometry and dynamic occlusions from players and the sports ball. In this work, we propose an occlusion-aware and perspective-consistent framework specifically for virtual advert placement in the soccer penalty area. We introduce an automatic procedure to select a geometrically consistent quadrilateral region inside the penalty area from predicted field coordinates, which is then used for homography-based warping. We integrate instance-level occlusion masks with Laplacian Alpha Blending for dynamic occlusion-aware blending so that the virtual advert is correctly placed behind players and the ball. Quantitative evaluations demonstrate that our occlusion-aware advert placement method preserves high visual fidelity with an average SSIM of 0.97 and PSNR of 31.6 dB, while maintaining temporal consistency with flicker index increase of <7%. Furthermore, we analyze occlusion preservation by comparing advert insertion with and without occlusion handling. A decrease in this metric indicates that objects overlapping the advert region become incorrectly hidden after advert insertion. The results highlight the importance of occlusion-aware blending for maintaining scene integrity and visual realism. By effectively managing occlusions, the proposed framework reduces visual artifacts and improves perceptual quality, producing augmented sports footage that is realistic and visually coherent.

Keywords Virtual advertising · Candidate space · Penalty area · Advert placement · Sport videos

1 Introduction

Advertising plays a crucial role in media, offering brands the ability to reach large audiences. With the increasing digitization of content, there has been a growing interest in integrating advert, particularly virtual billboards, into live video streams. Despite advancements in this domain, advert insertion in video remains largely manual, requiring significant effort from video editors. This challenge is especially significant in sports broadcasting, where the

Extended author information available on the last page of the article fast-paced nature of the content demands both automation and a high level of visual realism. Therefore, it is important to automatically identify candidate spaces in a video frame, wherein new adverts can be implanted, such that the candidate space should match the scene perspective, and also have a high quality of realism [6]. Advances in deep learning such as one-stage object detectors and multi-object tracking frameworks have also driven progress in automated highlight detection in sports videos [14, 29]. However, most existing methods emphasize global event classification or action recognition [5, 9], with limited focus on spatially constrained tasks such as localized advert placement.

Recent work in soccer video analysis has increasingly incorporated spatial reasoning, including player–ball interaction modeling, field registration, and homography estimation, which enable more precise scene understanding. These techniques provide a foundation for downstream applications that depend on stable geometric alignment, such as spatially grounded content (advert) insertion. Broadcast technology enables dynamic insertion of virtual advert around the perimeter boards, which are static and have pre-existing boards [30]. Consequently, the penalty area is not currently used for commercial advertisement placement in official matches. Nevertheless, the penalty area remains one of the most visually salient regions in broadcast coverage, receiving concentrated camera attention during penalties, corners kicks. From a theoretical media-visibility perspective, such high-focus zones represent areas of potential commercial interest should regulatory frameworks evolve. In modern broadcasts, dynamic advert insertion has become standard, allowing networks to monetize live feeds while maintaining immersion. Unlike static overlays, realistic virtual adverts must adhere to the physical geometry of the field and interact naturally with dynamic scene elements like players and the ball. Effective advertisement placement in sports video requires addressing two key challenges. First, adverts must be accurately aligned with the perspective of the field, ensuring that they appear naturally integrated within the scene geometry. This requires precise estimation of the camera viewpoint and application of appropriate geometric transformations. Second, occlusion handling is essential so that adverts appear behind players or objects when appropriate, preserving scene realism and avoiding unnatural overlaps. To address this problem, we propose a framework for advertisement placement in soccer videos that is both occlusion- and perspective-aware. Our method does not rely on manually annotated advert regions. We predict pitch extremities using a pretrained SoccerNet model for video dataset. Based on these predicted pitch extremities, we then introduce a fully automatic geometric procedure that derives a quadrilateral region inside the penalty area. No manual region annotation is required. Thus, the advert region is geometrically inferred from field structure. Our contribution shifts the problem from static advert blending to geometry-driven, video-level, occlusion-aware dynamic advert placement in broadcast soccer footage. We summarize our contributions as follows:

i) A geometry-driven framework that automatically selects geometrically consistent quadrilateral region within the penalty area, eliminating the need for manual region annotation.

ii) An occlusion-aware dynamic advert placement pipeline based on homography transformation and instance segmentation enabling realistic advert integration in broadcast soccer videos.

iii) An extensive experimental evaluation demonstrating improved visual realism.

The rest of the paper is organized as follows: Section 2 provides background and discusses related work. Section 3 details our proposed methodology and advert placement framework.

Experimental evaluations are presented in Section 4. Finally, Section 5 concludes the paper by summarizing the findings with discussions and potential directions for future research.

2 Related works

The proposed framework integrates three key components: occlusion handling, homography estimation, and advert placement within sports broadcasts. Occlusion handling ensures realistic compositing of virtual elements with dynamic foreground objects such as players and the ball. Homography estimation enables accurate geometric alignment between the broadcast image and the field plane. Finally, advert placement techniques focus on the seamless integration of advert into broadcasts. Accordingly, this section reviews related works in these three areas.

2.1 Occlusion handling

Occlusion handling is a critical component for the integration of virtual elements in live sports scenes. This enables virtual content to appear naturally behind the objects, thereby enhancing the realism and immersion of augmented sports broadcasts. Such pixel-level accuracy is especially important in soccer, where players frequently occlude critical regions of the field and the visual dynamics change rapidly. Recent research has explored various approaches to handle occlusion in sports video analysis, particularly within the context of soccer. An Occlusion detection and resolving is a forward–backward tracking algorithm designed to improve player detection and tracking under frequent occlusions [25]. By leveraging temporal consistency, the method effectively recovers trajectories lost due to overlapping players, a common issue in dense scenes such as the penalty area. Similarly, Wang [28] incorporates an optimized YOLOv5 architecture with attention mechanisms, combined with a modified Kalman filter in DeepSORT to maintain robust tracking across frames. While both methods primarily rely on bounding box–based tracking, they contribute valuable strategies for managing occlusions in dynamic, real-time environments. Additional relevant works, such as SoccerNet-Tracking [3], provide large-scale benchmarks for multi-object tracking in soccer and highlight the difficulties posed by frequent occlusions and similar appearances among players. Building upon these insights, more fine-grained occlusion handling can be achieved through instance segmentation techniques. Models such as Mask R-CNN [10] have been widely adopted for their ability to generate accurate object masks of players and other scene elements [12, 13]. In the context of ball detection and marking [22], Mask R-CNN provides greater precision in object localization by incorporating pixel-level segmentation information. Additionally, Buric et al. [2] observed that YOLO-based approaches tend to struggle more with occlusion, particularly in scenarios involving overlapping objects such as the ball, whereas Mask R-CNN demonstrates improved robustness under such conditions. For instance, Kiruthika et al. [15] demonstrate that Mask R-CNN is well-suited for sports footage, as it can effectively distinguish individual players. Their study also highlights that YOLO-based approaches may struggle to differentiate objects with visual characteristics similar to the background, such as players wearing jerseys that blend with the pitch.

However, we selected Mask R-CNN for our work due to its well-established performance and its ability to capture fine-grained object boundaries, making it particularly suitable for scenes with overlapping objects commonly found in sports footage. Our method leverages instance segmentation to detect key occluders, such as players and the ball, and applies a visibility-aware compositing strategy to ensure that virtual logos and text are seamlessly integrated into the scene. This approach preserves spatial realism without requiring explicit depth information, while remaining effective across diverse game situations and varying camera angles.

2.2 Homography estimation

Recent advancements in homography estimation have significantly enhanced the understanding and augmentation of sports scenes, particularly in soccer broadcasts where dynamic camera motion and frequent occlusions pose challenges for accurate field registration [4]. Liu et al. [18] explores learning geometric correspondences through a coarse-to-fine estimation framework, enabling robust performance under varying scales of transformation. This approach is especially effective in scenarios involving large viewpoint changes caused by camera zoom, pan, or tilt—common in live sports broadcasts. Complementing this, the Enhanced Homography-Based Sports Image Components Analysis System [1] integrates feedback-based homography estimation with additional tasks such as player localization and reference-plane mapping. By incorporating contextual cues like team jersey colors, the system improves alignment accuracy and demonstrates the utility of combining homography estimation with scene understanding for more precise visual overlays. Zhang and Izquierdo [31] reported the Four-Point Camera Calibration Method for Sport Videos, which estimates homography using four control points derived after foreground segmentation via a conditional GAN. This preprocessing step effectively removes occluding objects such as players, allowing for cleaner homography computation through Direct Linear Transform and subsequent refinement using Enhanced Correlation Coefficient alignment. These methods collectively demonstrate the evolving sophistication of homography estimation techniques and their crucial role in achieving spatially accurate, visually coherent augmented reality experiences in live sports environments.

Methods such as SFLNet [26] have demonstrated high accuracy in detecting and aligning field lines, even under difficult conditions involving occlusions or lighting changes. This mapping is essential for localizing semantically important zones, including the center circle, goal area, and penalty box.

2.3 Advert placement

In recent years, deep convolutional neural networks (CNNs) have demonstrated remarkable performance in image and video processing tasks, with applications in advert classification and advert detection [7]. These techniques have been particularly effective in outdoor environments. Augmented reality (AR) has also gained significant traction in sports broadcasting, enabling the seamless overlay of virtual content such as logos, scores, and branding elements onto live video feeds. While object detection in sports settings has been widely studied [3, 11], relatively few works address the challenge of integrating visual elements like logos or text directly onto the playing field, particularly in dynamic and occlusion-prone regions such as the penalty area. Prior approaches such as ST-GAN [17] and context-aware scene editing methods [21] have explored geometric and contextual consistency for virtual content embedding. However, these methods often overlook real-time occlusion modeling, which is essential in fast-paced environments like soccer matches. Other works in augmented reality broadcasting have addressed the integration of virtual billboards and field markings, typically relying on homography estimation [20, 27]. Nevertheless, occlusion handling remains a significant limitation in these systems. Existing approaches often fall short in one or both of these areas, some rely solely on homography-based alignment without considering occlusions leading to visual artifacts [24]. While others depend on depth-estimation [16, 23], which remains unreliable in outdoor environments like soccer fields due to issues like scale ambiguity and uniform textures.

Dhang et al. [8] introduced LaBINet, a method for seamless scene integration under varying lighting and perspective conditions within a pre-existing advert region and on static images. While effective for static images, LaBINet does not address automatic region derivation from field geometry or dynamic occlusion reasoning with moving players and the ball. To address these limitations jointly, we propose an occlusion-aware framework with homography-based alignment for the realistic and robust advert placement in dynamic soccer scenes. This work differs from prior methods by an automatic field-region selection with visibility-aware placement, ensuring that virtual adverts remain geometrically consistent and visually coherent in broadcasts, instead of relying on manual region annotation.

3 Proposed methodology

We propose a framework for the insertion of virtual advertisements into soccer broadcast videos, specifically within the penalty area. The method ensures that adverts are naturally integrated into the scene by addressing key challenges such as occlusion from dynamic elements (e.g., players and the ball) and maintaining geometric consistency under varying camera perspectives.

3.1 Dataset

For this study, we focus on the field layouts calibration set, which includes line segment essential for camera parameter estimation. The image training set in the SoccerNet dataset [19] provides robust annotations including field layouts as shown in Fig. 1. The purpose of Fig. 1 is to visually illustrate all the soccer pitch extremities, which refer to the field coordinates. The different colors correspond to these field extremities (e.g., penalty areas, goal areas, center circle, sidelines, and goal structures). Each colored segment represents a specific class in the annotation scheme, allowing unambiguous identification of pitch boundaries and key reference lines. Figure 2 depicts a sample of automatically selected penalty area regions from the training images, used to better understand and visualize the different polygon areas. These calibrated data are essential for accurate spatial understanding and perspective mapping, forming the foundation for realistic virtual advert insertion.

For evaluation purposes, we utilize videos dataset from SoccerNet. The description of the video dataset is summarized in Table 1. The discrepancy between the total frames and the frames actually processed (e.g., 284 vs. 268 frames for Video 1) arises from decoding

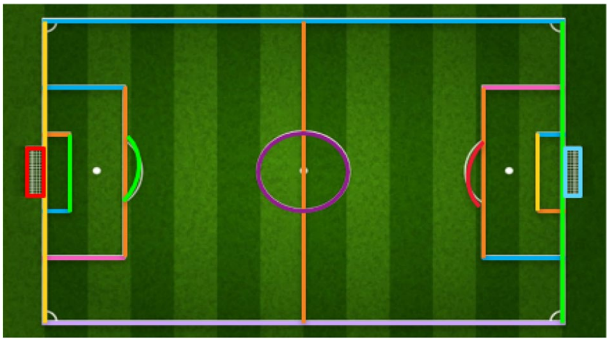


Fig. 1 All pitch extremities: Big rect. left bottom, Big rect. left main, Big rect. left top, Big rect. right bottom, Big rect. right main, Big rect. right top, Circle central, Circle left, Circle right, Goal left crossbar, Goal left post left, Goal left post right, Goal right crossbar, Goal right post left, Goal right post right, Middle line, Side line bottom, Side line left, Side line right, Side line top, Small rect. left bottom, Small rect. left main, Small rect. left top, Small rect. right bottom, Small rect. right main, Small rect. right top. The different colors correspond to distinct semantic field elements (e.g., penalty areas, goal areas, center circle, sidelines, and goal structures)

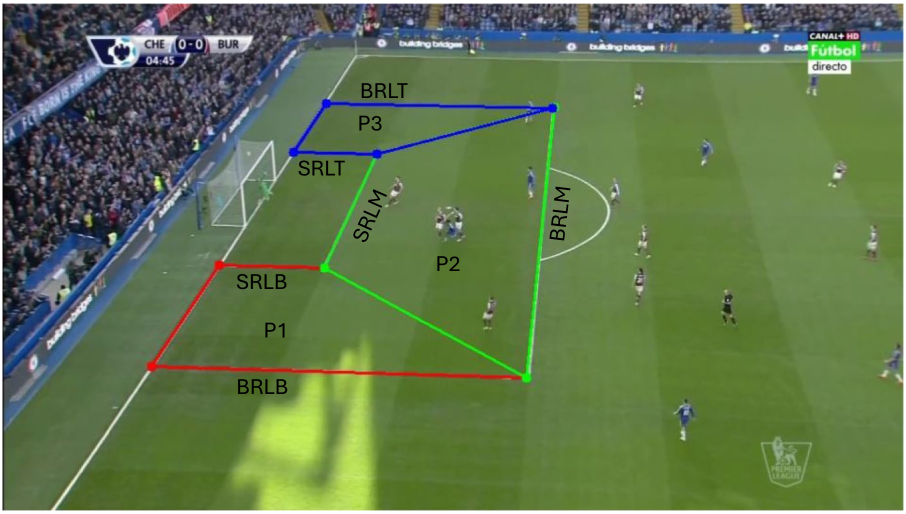


Fig. 2 Sample of automatically selected penalty area coordinates and marked polygon area from training set: Red polygon (P1): Big rect. left bottom (BRLB), Small rect. left bottom (SRLB). Green polygon (P2): Small rect. left main (SRLM), Big rect. left main (BRLM). Blue polygon (P3): Small rect. left top (SRLT), Big rect. left top (BRLT). Note that the penalty area is located on the left side of the field; therefore, all annotations correspond to the left side

limitations in OpenCV. All subsequent analyses were performed on the 268 frames that were successfully read for Video 1. Therefore, the effective FPS for the analyzed portion remains consistent with the original encoding.

The selected video clips were chosen to represent diverse real-world conditions, with the penalty area occupying varying positioning and lighting conditions. Specifically, Video 1 (total 268 frames) contains mixed frames with and without the penalty area under

Table 1 Dataset description

Name	Video 1	Video 2	Video 3	Video 4	Video 5	Video 6
Duration (mm:ss)	(00:10)	(00:30)	(00:26)	(00:33)	(00:02)	(00:04)
Dimension	1280 x 720	1920 x 1080	1920 x 1080	1280 x 720	1280 x 720	1280 x 720
Color grading	RGB	RGB	RGB	RGB	RGB	RGB
FPS	25	25	30	25	25	25
Total number of frames	284	750	807	827	66	105
Total number of frames actually readable	268	750	807	827	66	105
No. of frames with penalty area	159	448	807	134	66	105
No. of frames with no-penalty area	109	302	-	693	-	-

(a) Advert1. (b) Advert2. (c) Advert3.

Fig. 3 Represents the adverts in different sizes and colors. The top row displays non-transparent images, while the bottom row displays transparent images

bright lighting conditions. Video 2 (total 750 frames) includes footage with varied camera angles and noticeable lighting variations. Video 3 (total 807 frames) consists entirely of penalty-area frames, enabling evaluation under continuous target presence. Video 4 (total 827 frames) consists of mixed frames with and without the penalty area. Video 5 (total 66 frames) and Video 6 (total 105 frames) consist entirely of penalty-area, enabling evaluation under continuous target presence. Across these sequences, the dataset covers different camera perspectives, zoom levels, varying proportions of visible penalty-area regions, and considering occluding objects which includes players, the ball, and pitch lines. To understand the robustness of the proposed method, we chose three different types of images as advert with different sizes, colors as shown in Fig. 3. The transparent advert better reflect real-world virtual advertising scenarios where seamless integration with the field texture is required. Therefore we considered, images with transparent backgrounds to allow the advert to integrate naturally with the field in our video dataset.

Figure 4 illustrates the overall framework for inserting virtual advert into soccer broadcast videos by automatically selecting a geometrically consistent quadrilateral region within the penalty area.

3.2 Polygon coordinates selection for homography

As video dataset has no label pitch extremities, we used a pre-trained SoccerNet model to predict pitch extremities for each frame. Using these predicted extremities, we automatically extract geometrically consistent quadrilateral region inside the penalty area. These extremities are used to compute a homography matrix that maps the advert image onto the field surface with geometric fidelity.

Given an image frame f_i and a set of detected polygons $\{P_{ij}\}_{j=1}^{M}$, we compute the following geometric properties for each polygon P_{ij}:

- Area (A_{ij}): The area of polygon P_{ij}, represented as a sequence of 4 normalized points $\{(p_k^x, p_k^y)\}_{k=1}^{4}$, is computed using the Shoelace formula. The normalized coordinates are first scaled by the image width w and height h, where $(x_k, y_k) = (p_k^x \cdot w,\ p_k^y \cdot h)$

$$A_{ij} = \frac{1}{2}\left|\sum_{k=1}^{4}(x_k y_{k+1} - x_{k+1} y_k)\right|$$

where $(x_5, y_5) = (x_1, y_1)$ to complete the loop.

- Aspect ratio (R_{ij}): The aspect ratio of the bounding rectangle enclosing the polygon is: $R_{ij} = \frac{w_{ij}}{h_{ij}+\epsilon}$ where w_{ij} and h_{ij} are the width and height of the bounding box, and $\epsilon = 1e{-}5$ prevents division by zero.

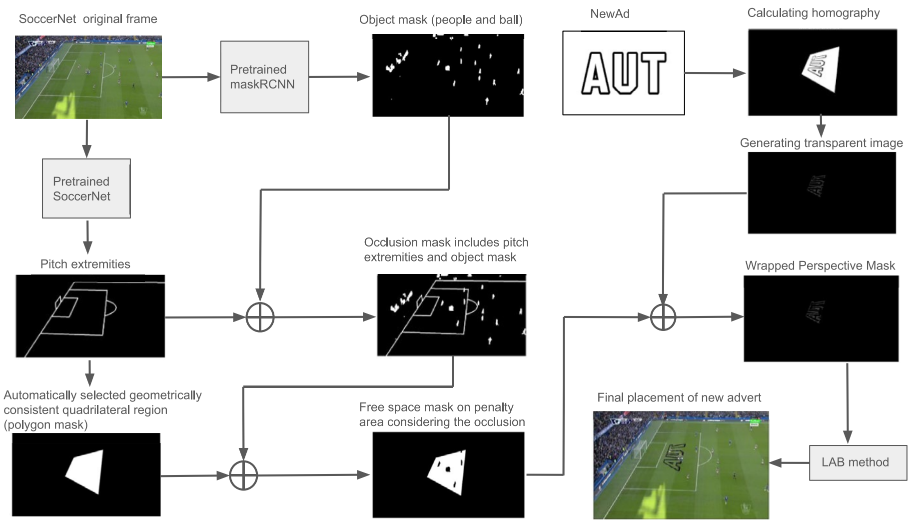


Fig. 4 Framework for advert placement in penalty area considering occlusion - person, sports ball, coordinates

- Similarity score (S_{ij}): Defined as the polygon's area, $S_{ij} = A_{ij}$, favoring larger, more field-representative regions.

We consider a polygon P_{ij} to be valid if it satisfies the following criteria: *i)* has exactly four unique points: $|P_{ij}| = 4$ *ii)* It meets a minimum area requirement: $A_{ij} \geq A_{\min}$, with $A_{\min} = 30000$ *iii)* the polygon is convex *iv)* each internal angles θ_k satisfies the constraint: $|\theta_k - 90°| \leq \alpha$ where α is the angle deviation threshold. We prioritize polygons with approximately right internal angles and enforce a minimum area constraint. These criteria serve as thresholds to ensure that the detected region is sufficiently large and visually appropriate for placing the advert. The minimum area constraint helps avoid selecting very small regions where the inserted advertisement would not be perceptually clear, while the near-orthogonal angle constraint favors quadrilateral shapes that better resemble planar surfaces used for advert placement. Finally, among all valid polygons $\{P_{ij}\}$ in a frame f_i, the best polygon P_i^* is selected as:

$$P_i^* = \arg\max_j \left\{ S_{ij} \;\middle|\; P_{ij} \text{ is valid} \right\}$$

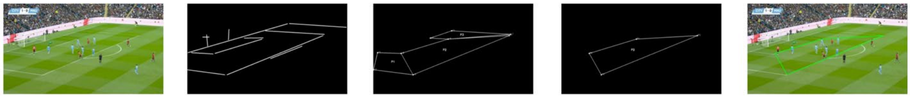
This polygon provides the basis for homography estimation and accurate advert warping. Our framework effectively preserves the spatial integrity of adverts, even under occlusion and varying camera motion, by coupling geometric transformations with semantic segmentation. The Algorithm 1 describes automatic extraction of penalty area and polygon selection. Pitch extremities in Algorithm 1 explicitly refer to the field coordinates predicted by a pre-trained SoccerNet model on the SoccerNet video dataset. Figure 5 illustrates the output for frames in Video 1, Video 2, Video 3, Video 4, Video 5 and Video 6 using Algorithm 1.

Algorithm 1 Penalty area extraction and polygon selection.

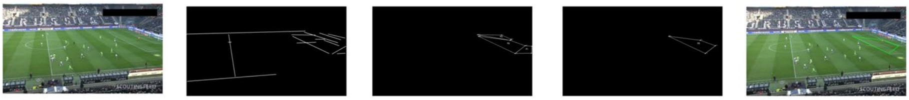
```
Input: Input video V = {V1, V2, V3, V4, V5, V6}
Output: Best-fit polygon coordinates for advert placement
Let F = {f1, f2, ..., fn} be the set of frames extracted from V
Load pre-trained SoccerNet model M_soccer
for each frame f_i ∈ F do
    Predict pitch extremities: E_i ← M_soccer(f_i)        ▷ Determines all pitch coordinates
    Generate binary field mask M_i of all the extremities
    Extract candidate penalty area coordinates C_i from E_i
    for each coordinate set c_ij ∈ C_i do
        Construct polygon P_ij from c_ij                  ▷ P1, P2, P3
        Compute properties of P_ij      ▷ Area A_ij, Orientation θ_ij, similarity score S_ij
    end for
end for
Select best polygon P* based on properties of P_ij
    return Coordinates of selected polygon P*
```

3.3 Advert warping and occlusion-aware blending

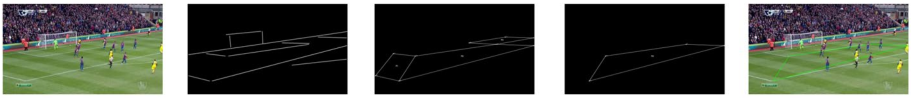
Once the optimal polygon region is selected, the next step is to project the advert onto the field surface while handling occlusions caused by dynamic objects such as players and the ball. To robustly handle occlusions in dynamic soccer scenes, we integrate instance-level



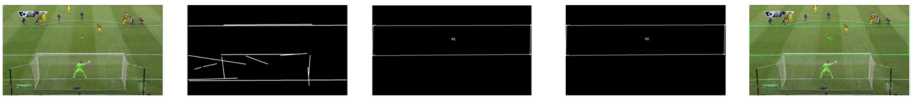
(a) Represents output for frame 262 in Video 1



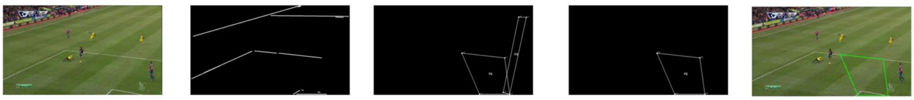
(b) Represents output for frame 729 in Video 2



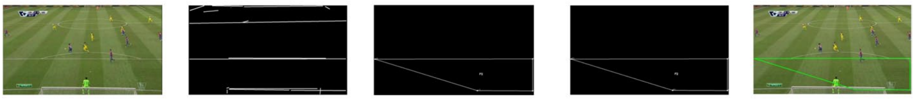
(c) Represents output for frame 795 in Video 3



(d) Represents output for frame 754 in Video 4



(e) Represents output for frame 20 in Video 5



(f) Represents output for frame 18 in Video 6

Fig. 5 Represents output for automatically selecting polygon using Algorithm 1. From left-to-right: Original frame, Predicted output of all pitch extremities obtained from the pretrained SoccerNet model, Identified polygon areas, Automatically extracted best polygon (binary mask) and the original frame with the selected polygon

pretrained occlusion masks from Mask R-CNN model [10] to segment players and the ball in each video frame. This generates a binary occlusion mask that highlights regions occupied by dynamic objects, enabling frame-by-frame occlusion quantification and broadcast scene analysis. The Mask R-CNN with a ResNet-50-FPN backbone, contains approximately 44.4 million trainable parameters.

For each video frame $f_i \in V$, the occlusion mask O_i is obtained using the pretrained Mask R-CNN model, $O_i \leftarrow \mathcal{M}_{rcnn}(f_i)$. Next, the new advert image A, along with its alpha channel α, is warped into the target frame using a homography transformation H_i derived from the selected polygon coordinates P^*, such that for each corresponding point pair $\mathrm{A}_i \leftrightarrow \mathrm{P}_i^*$:

$$\begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} \sim \mathrm{H} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

where, $(x_i, y_i) \in$ coordinates of new advert, $(x'_i, y'_i) \in$ polygon coordinates in target frame, $\sim$ indicates equality up to a non-zero scale (projective transformation). The resulting warped advert A_i is thus correctly aligned with the field geometry $A_i \leftarrow \mathrm{warpPerspective}(A, H_i)$. warpPerspective is defined as the homography-based perspective transformation used to project the advertisement into the estimated penalty area. To warp the image frame $f_s(x, y)$ using H, we compute the output image $f_o(x', y')$ of size $\mathrm{target\_w} \times \mathrm{target\_h}$ as:

$$(x, y)^T = \left(\frac{h_{11}x' + h_{12}y' + h_{13}}{h_{31}x' + h_{32}y' + h_{33}}, \frac{h_{21}x' + h_{22}y' + h_{23}}{h_{31}x' + h_{32}y' + h_{33}} \right)$$

$$f_o(x', y') = f_s(x, y)$$

This inverse mapping ensures that each pixel in the output frame f_o is computed by sampling the corresponding location in the source image f_s using the homography H.

To account for occlusions, we construct a pixel-wise alpha mask $\alpha_i(x, y)$ that incorporates occlusion information:

$$\alpha_i(x, y) = \begin{cases} 0, & \text{if } (x, y) \in O_i \\ \mathcal{M}_{\mathrm{lab}}(x, y), & \text{otherwise} \end{cases}$$

Here, O_i denotes the occluded region, and $\mathcal{M}_{\mathrm{lab}}$ is a LAB-based blending [8] model designed to ensure smooth transitions and consistent color blending between the advert and the field background. While our previous work blended advert within a pre-existing advert region and on static images, in this work, we dynamically select an optimal polygonal area based on scene geometry and perform blending within that region in video frames. The equation $\nabla u = \nabla g$ used in the context of blending to ensure that the composited image u preserves the gradient field of the advertisement image g within the selected insertion region. Here, u represent the pixel values in the selected polygonal region of the target frame, and g those from the corresponding source region. The blending process enforces $\nabla u = \nabla g$ inside the selected region, while the boundary values are fixed according to the target (background) image. This formulation preserves the internal gradient structure of the advert while maintaining seamless continuity with the surrounding scene at the region boundaries. This results in a sparse linear system $Au = b$, where ∇^2 is the Laplacian operator discretized via finite differences on a uniform grid. For full derivation and implementation details, refer to [8].

The final blended frame $\hat{f}_i$ is produced by compositing the warped advert A_i with the original frame f_i guided by the occlusion-aware mask α_i. Foreground occluders (players, ball) are reinserted on top of the blended image to preserve proper scene depth. This process is repeated across all frames to yield a modified video $\hat{V} = \{\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_n\}$, where each frame contains a geometrically accurate and context-aware virtual advert placement. The complete advert insertion and blending process across video frames is described in Algorithm 2. Building upon occlusion-aware segmentation and homography-based geometric alignment, we therefore insert a virtual advert into the field.

Algorithm 2 Advert insertion with occlusion-aware.

Input: Best polygon coordinates P^*, binary mask M, input video V, New advert A
Output: Video frames with seamlessly blended virtual advert
Let $F = \{f_1, f_2, \ldots, f_n\}$ be the set of frames extracted from V
Load Mask R-CNN model $\mathcal{M}_{rcnn}$ for occlusion segmentation
Load LAB blending model $\mathcal{M}_{lab}$ for advert placement
for each frame $f_i \in F$ **do**
  Detect occluders: $O_i \leftarrow \mathcal{M}_{rcnn}(f_i)$
  Define advert placement region $\mathcal{R}_i$ using polygon coordinates P^*
  Warp and resize advert $A_i \leftarrow \text{warp}(A, \mathcal{R}_i)$
  Generate alpha mask α_i using O_i and $\mathcal{M}_{lab}$:
  Blend advert: $\hat{f}_i \leftarrow \mathcal{M}_{lab}(f_i, A_i, \alpha_i)$ ▷ advert layer: A_i blended using α_i
  Compose final frame $\hat{f}_i \leftarrow (O_i, f_i, \hat{f}_i)$ ▷ target frame: f_i, occlusions: O_i
end for
  return Modified video $\hat{V} = \{\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_n\}$

3.4 Limitations

While the proposed framework demonstrates strong performance in realistic advert placement, several limitations persist:

- Field Line Detection Errors: Field lines near the penalty area are often partially occluded by players or shadows. Variability in lighting, grass texture, and camera angle further complicates accurate detection, sometimes resulting in erroneous or missed lines as depicted in Fig. 6.
- Interference from Goal Net: The goal net frequently obstructs parts of the field but is not classified as an occluding object by Mask R-CNN as depicted in Fig. 7. Consequently, the model may attempt to place adverts in visually blocked areas, degrading placement realism. As a potential mitigation, fine-tuning the segmentation model with annotated goal-net examples could help; although such annotations are not publicly available, creating a dedicated dataset for goal nets would be a feasible approach for future work.
- Polygon out-of shape: Due to inaccurate corner detection or perspective calibration, the generated field polygon may become irregular or warped. This misalignment impacts both the homography estimation and the perceptual integration of adverts as depicted in the Fig. 8.

Fig. 6 Examples of field line detection errors due to occlusion and challenging lighting conditions

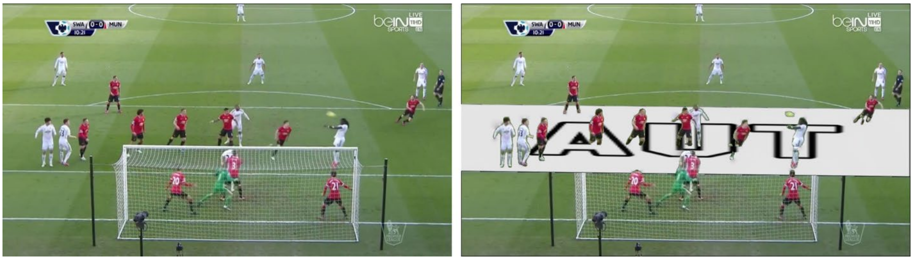


Fig. 7 Original frame (left) and advert placement (right) showing incorrect insertion over the goal net area due to lack of occlusion detection

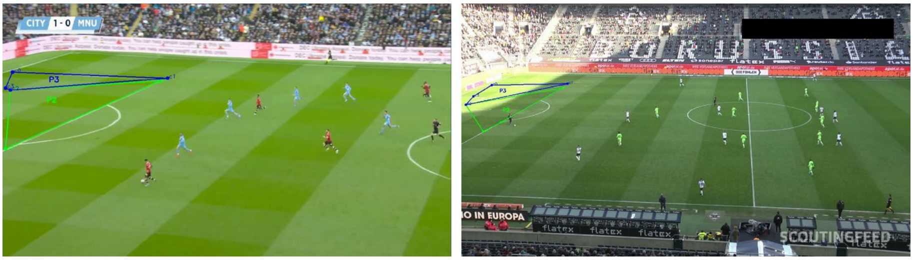


Fig. 8 Examples of irregular shape-of polygons resulting from filed line detection. Represents polygon out-of-shape

4 Experimental results

In this section, we evaluate our proposed framework through quantitative metrics and qualitative visualizations on the video dataset. Our quantitative results demonstrate measurable improvements, while qualitative visuals highlight the visual realism of occlusion-aware advertisement placement. We considered adverts[1] with transparent background to allow them to integrate naturally with the field in our video dataset.

4.1 Quantitative analysis

To assess the visual fidelity of our advert insertion framework, we computed the Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR) between the original and modified video frames. These metrics help quantify how visually similar the modified video is to the original. To evaluate temporal stability and motion preservation, we analyze temporal consistency using optical flow consistency and flicker index metrics. These metrics help determine whether player movements and overall video dynamics remain unaffected by the inserted advertisement.

1. **Temporal consistency motion-preservation metric:** The optical flow consistency metric measures deviations in motion fields before and after advert integration. Lower values indicating better preservation of temporal motion cues. Let f_i denote the i-th frame of a video. For each consecutive pair of frames we compute the dense flow field for the original

[1] Note: Advert1, Advert2, Advert3 correspond to the bottom row in Fig. 3, which illustrates the new advert labeling used throughout the experimental section.

and the modified video with the advert inserted using `calcOpticalFlowFarneback` method to obtain $F_i^{\text{orig}} = \text{Flow}(f_i^{\text{orig}}, f_{i+1}^{\text{orig}})$ and $F_i^{\text{mod}} = \text{Flow}(f_i^{\text{mod}}, f_{i+1}^{\text{mod}})$, for $i = 1, \ldots, N-1$, where N is the number of frames. We measure the mean squared deviation between the two flow sequences (per-pixel, per-frame) using the below equation:

$$\text{Flow\_error} = \frac{1}{N-1} \sum_{i=1}^{N-1} \left\| F_i^{\text{orig}} - F_i^{\text{mod}} \right\|_2^2, \tag{1}$$

A lower flow error Flow_Error indicates that the advert insertion preserves the natural motion of the original video more closely.

2. **Frame intensity flicker metric:** The metric measures mean absolute difference between consecutive frames before and after insertion. A lower flicker index suggests better temporal consistency. We quantify how much the frame intensity changes over consecutive frames. We compute the mean absolute frame-to-frame intensity difference of the original video:

$$FI^{\text{orig}} = \frac{1}{N-1} \sum_{i=1}^{N-1} \text{mean}\left(\left| f_i^{\text{orig}} - f_{i+1}^{\text{orig}} \right| \right) \tag{2}$$

Similarly, we computed the mean absolute frame-to-frame intensity difference of the modified video FI^{mod}. A lower flicker index FI^{mod} compared to FI^{orig} suggests that the modified video maintains better temporal coherence, supporting the effectiveness of occlusion-aware blending.

We calculated SSIM and PSNR, focusing only on the penalty region of the frame. To obtain more informative metrics to quantify the similarity between the images before and after adding the advert on the penalty area, we calculated SSIM for an image frame from Video 1 as shown in Fig. 9a, before placing the advert on the penalty area (without players), SSIM = 1.00 and after placing advert on the same frame with SSIM = 0.987 using the proposed

(a) Frame 121 from Video 1. (b) Frame 504 from Video 2. (c) Frame 18 from Video 3.

Fig. 9 Calculated SSIM for an image frame before and after placing the advert on the penalty area (without players) method. Similarly, we calculated SSIM for video 2 after placing advert in frame 504, SSIM = 0.993 as shown in Fig. 9b, and for video 3, SSIM = 0.991, shown in Fig. 9c. From these findings, we demonstrate that the values remain very close to 1, indicating that the proposed method preserves the visual quality of the original frames while seamlessly integrating the advert into the penalty area, causing minimal perceptual distortion. Similarly, we quantify the similarity between the images before and after adding the advert in the penalty area for Video4, Video5, Video6.

Table 2 presents a quantitative comparison of video quality and temporal consistency metrics between the original broadcast video and its modified video containing inserted adverts. From a temporal perspective, the mean flow deviation for Video 1 was measured at 0.1505, signifying that motion dynamics remain well-preserved in the modified videos. The flicker index, a measure of frame-to-frame intensity variation, showed only a marginal increase from 8.15 in the original to 8.72 in the advert-inserted video, corresponds to an increase of approximately 0.569 in absolute value, which translates to about 7% relative increase. Tables 3, 4, 5, 6, 7, and 8 provides a detailed numerical analysis of metrics across five consecutive frames for each video, comparing the original video and advert placement (with Advert1), with and without occlusion awareness. For Video 1, the original video exhibits a flicker index ranging from 2.825 to 3.64 and an occlusion ratio between 0.240 and 0.259. With advert placement, the flicker index increases to values between 3.441 and 5.549. The occlusion-aware occlusion ratio ranges from 0.255 to 0.268, while the occlusion ratio without occlusion handling varies between 0.205 and 0.221. Optical-flow consistency for the advert placement ranges from 0.08 to 0.198. In Video 2, the flicker index remains largely unchanged after advert placement (9.8 to 10.9), suggesting minimal temporal dis-

Table 2 Comparison of video quality and temporal consistency metrics between original and modified videos. The reported values are the average across all frames for each video. (Note: Modified videos refer to videos with advert placement that is occlusion-aware of players and balls. Refer to Fig. 3 for New advert labeling)

Videos	New advert	SSIM (↑)	PSNR (dB) (↑)	Flow Consistency (↓)	Flicker Index (↓)	
					Original Video	Modified Video
Video 1	Advert1	0.974	31.60	0.151	8.146	8.715
	Advert2	0.970	32.09	0.166	8.146	8.627
	Advert3	0.980	32.65	0.108	8.146	8.382
Video 2	Advert1	0.978	36.30	0.102	7.952	7.936
	Advert2	0.978	36.28	0.103	7.952	7.953
	Advert3	0.978	36.32	0.097	7.952	7.924
Video 3	Advert1	0.964	29.07	0.300	5.734	6.409
	Advert2	0.961	29.52	0.298	5.734	6.213
	Advert3	0.971	30.56	0.213	5.734	5.993
Video 4	Advert1	0.970	35.90	0.546	5.330	8.263
	Advert2	0.966	35.73	0.812	5.330	8.266
	Advert3	0.973	35.74	0.430	5.330	7.627
Video 5	Advert1	0.940	28.60	0.249	3.483	4.589
	Advert2	0.934	26.93	0.334	3.483	4.633
	Advert3	0.951	27.26	0.192	3.483	4.315
Video 6	Advert1	0.943	30.90	0.499	5.210	7.150
	Advert2	0.931	30.41	0.303	5.210	6.526
	Advert3	0.955	30.74	0.564	5.210	6.989

Table 3 Metrics across five consecutive frames of Video 1, comparing the original video and advert placement

	Metric					
		Frame 262	Frame 263	Frame 264	Frame 265	Frame 266
Org Video	Flicker index	3.040	3.640	2.867	2.825	2.848
	Occlusion ratio	0.252	0.240	0.248	0.259	0.256
Advert placement	Optical-flow consistency	0.133	0.198	0.122	0.157	0.08
	Flicker index	4.05	5.549	3.840	3.867	3.441
	Occlusion ratio (occlusion-aware)	0.265	0.268	0.262	0.255	0.263
	Occlusion ratio (without occlusion)	0.215	0.207	0.209	0.205	0.221

Table 4 Metrics across five consecutive frames of Video 2, comparing the original video and advert placement

	Metric					
		Frame 729	Frame 730	Frame 731	Frame 732	Frame 733
Org Video	Flicker index (original)	10.490	10.219	10.857	10.274	9.88
	Occlusion ratio (original video)	0.051	0.039	0.043	0.044	0.041
Advert placement	Flicker index	10.492	10.136	10.956	10.257	9.818
	Optical-flow consistency	0.170	0.133	0.176	0.125	0.142
	Occlusion ratio (occlusion-aware)	0.051	0.039	0.045	0.041	0.041
	Occlusion ratio (without occlusion)	0.049	0.035	0.038	0.038	0.033

Table 5 Metrics across five consecutive frames of Video 3, comparing the original video and advert placement

	Metric					
		Frame 795	Frame 796	Frame 797	Frame 798	Frame 799
Org Video	Flicker index	5.187	5.707	6.104	6.507	7.334
	Occlusion ratio	0.041	0.040	0.041	0.043	0.043
Advert placement	Optical-flow consistency	0.297	0.134	0.375	0.247	0.167
	Flicker index	6.109	6.135	7.256	7.437	8.14
	Occlusion ratio (occlusion-aware)	0.045	0.038	0.059	0.039	0.043
	Occlusion ratio (without occlusion)	0.034	0.030	0.028	0.031	0.039

Table 6 Metrics across five consecutive frames of Video 4, comparing the original video and advert placement

	Metric					
		Frame 754	Frame 755	Frame 756	Frame 757	Frame 758
Org Video	Flicker index	0.147	6.581	5.811	6.576	5.364
	Occlusion ratio	0.073	0.077	0.086	0.084	0.089
Advert placement	Optical-flow consistency	0.086	0.235	0.856	0.309	0.135
	Flicker index	1.273	8.950	9.555	9.441	7.114
	Occlusion ratio (occlusion-aware)	0.060	0.066	0.077	0.073	0.083
	Occlusion ratio (without occlusion)	0.059	0.063	0.069	0.066	0.070

Table 7 Metrics across five consecutive frames of Video 5, comparing the original video and advert placement.

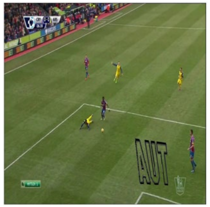
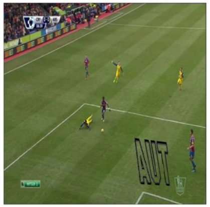
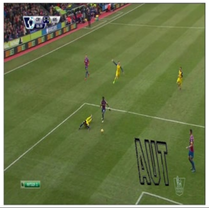
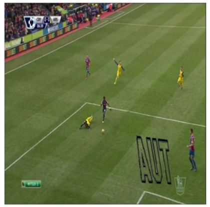
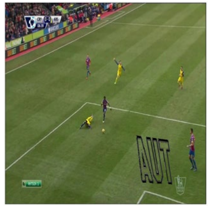
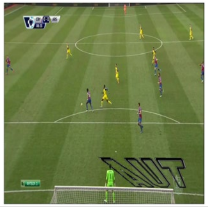
	Metric					
						
		Frame 20	Frame 21	Frame 22	Frame 23	Frame 24
Org Video	Flicker index (original)	4.096	5.156	4.152	1.561	5.405
	Occlusion ratio (original video)	0	0	0	0	0
Advert placement	Flicker index	5.087	6.134	5.828	2.394	6.920
	Optical-flow consistency	0.110	0.0924	0.248	0.0707	0.428
	Occlusion ratio (occlusion-aware)	0	0	0	0	0
	Occlusion ratio (without occlusion)	0	0	0	0	0

Table 8 Metrics across five consecutive frames of Video 6, comparing the original video and advert placement

	Metric					
						
		Frame 18	Frame 19	Frame 20	Frame 21	Frame 22
Org Video	Flicker index	6.133	5.743	0.319	4.016	5.344
	Occlusion ratio	0.070	0.056	0.068	0.061	0.066
Advert placement	Optical-flow consistency	1.045	0.763	0.217	0.235	0.310
	Flicker index	8.450	7.185	1.613	5.0271	6.417
	Occlusion ratio (occlusion-aware)	0.071	0.058	0.069	0.060	0.064
	Occlusion ratio (without occlusion)	0.063	0.054	0.067	0.052	0.058

turbance. Similarly, optical-flow consistency remains stable (0.125 to 0.176), indicating that motion coherence is preserved in this case. For Video 3, the flicker index of the original video varies between 5.187 and 7.334, with occlusion ratio from 0.040 to 0.043. After advert placement, the moderate increases flicker index to 6.109–8.14, along with optical-flow consistency fluctuates between 0.134 and 0.375. The occlusion-aware method again maintains more consistent occlusion estimates. Similar trends are observed across the other videos, where advert placement (with Advert2, Advert3) generally moderates flicker deviation while occlusion-aware processing provides more stable and reliable occlusion estimates across frames.

Notably, in Video 5 (Table 7), frames 20–24 contain no objects within the penalty area, resulting in an occlusion ratio of zero. Overall, these results highlights how advert placement affects flicker and occlusion ratio across consecutive frames in videos.

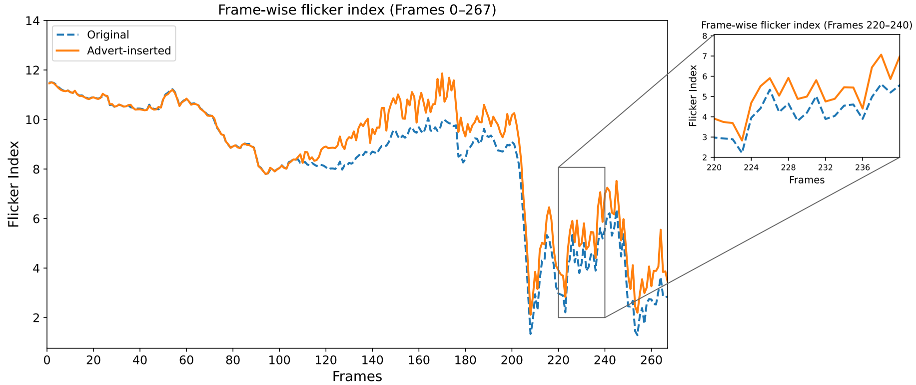
Frame-wise analysis as illustrated in Figs. 10 and 11 shows that the flicker index for both original and modified videos closely align, indicating that temporal brightness continuity is preserved. In Fig. 10a, the leftmost figure depicts the flicker values for all 268 frames for Video 1. However, the magnified view, depicted in the rightmost includes a zoomed-in section of the figure for closer examination from Frames 220 to 240, containing scenes with advert placement with occlusion handling. Small deviations in flicker and optical flow around certain frames correspond to transient scene dynamics but remain within perceptually acceptable limits, confirming robust temporal coherence. Similarly, Fig. 10b and c depicts flicker values for Video 2 and 3 and Fig. 11a, b and c depicts flicker values for Video 4, 5 and 6 respectively. The near-identical patterns reinforce that the advert insertion process induces negligible temporal disturbance, highlighting the robustness of the proposed pipeline across diverse video content.

4.2 Occlusion preservation analysis

Correct occlusion ordering between objects (e.g., players and the ball) and the inserted advertisement is essential for maintaining visual realism. If occlusion is not handled properly, the advert may incorrectly appear on top of objects, causing them to disappear and resulting in unrealistic rendering. To verify that the proposed method preserves these occlusion relationships, we conduct an ablation study comparing advert insertion with and without occlusion handling. To quantitatively assess this behavior, we compute the occluded area ratio. Let M_{obj} denote the binary object mask and M_{poly} denote the binary mask corresponding to the selected polygon, where the advert is placed. The occlusion ratio is defined as:

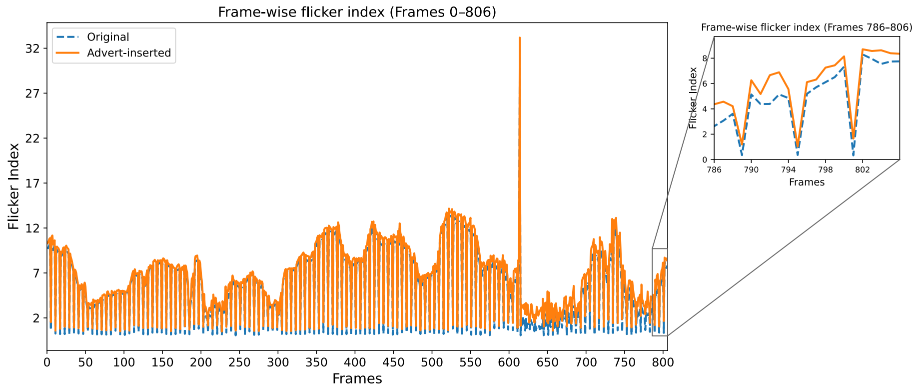
$$O = \frac{|M_{\mathrm{obj}} \cap M_{\mathrm{poly}}|}{|M_{\mathrm{obj}}|} \tag{3}$$

This metric measures the proportion of object pixels that overlap with the advert region. A decrease in this value indicates that objects overlapping the advert region become hidden after advert insertion. Table 9 reports the average occlusion ratio across all frames for each video. The results show that the proposed occlusion handling method produces values that remain close to those of the original videos, indicating that the natural occlusion relationships between players and the advertisement region are preserved. In contrast, inserting



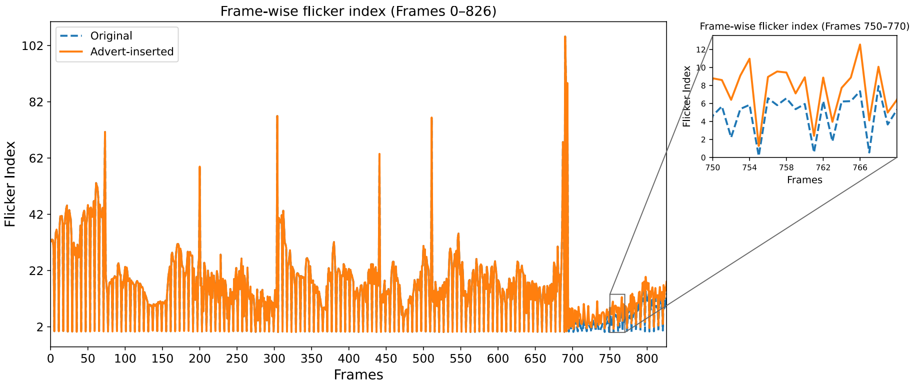
(a) Video 1

(b) Video 2

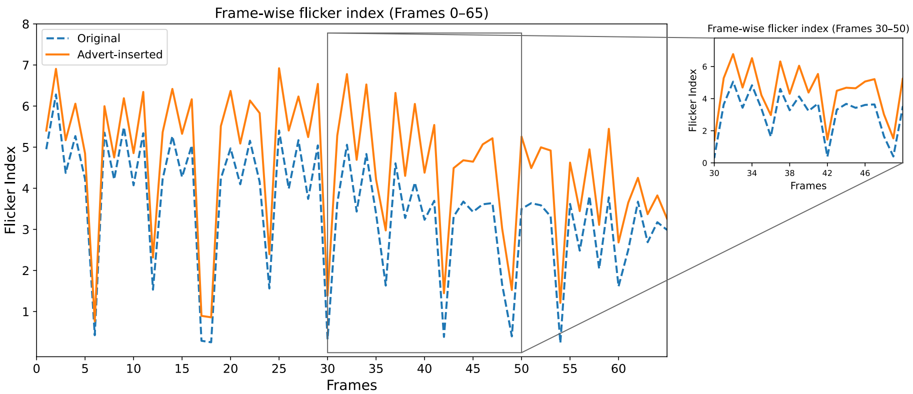


(c) Video 3

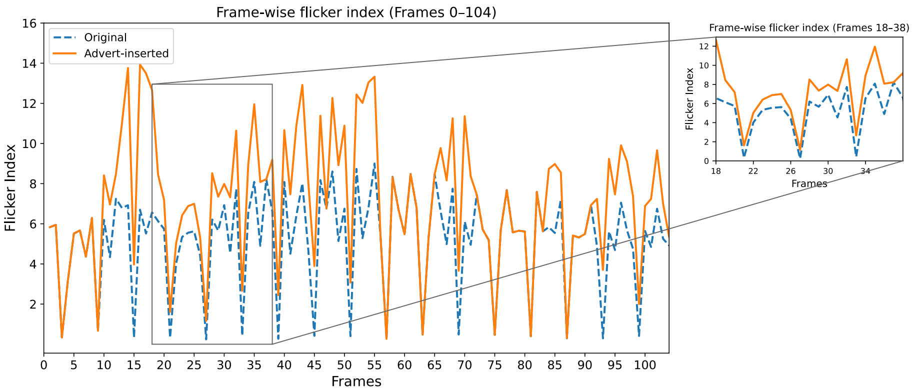
Fig. 10 Frame-wise flicker index for both original and modified videos. The leftmost figure depicts the flicker values for all frames. However the magnified view, depicted in the rightmost includes a zoomed-in section of the figure for closer examination from selected frames



(a) Video 4



(b) Video 5



(c) Video 6

Fig. 11 Frame-wise flicker index for both original and modified videos. The leftmost figure depicts the flicker values for all frames. However the magnified view, depicted in the rightmost includes a zoomed-in section of the figure for closer examination from selected frames

Table 9 Occlusion ratio comparison between original videos (no advert) and advert-inserted videos with and without occlusion handling. The metric measures the proportion of the object mask that becomes occluded after advert placement. Lower values indicate less occlusion. Results are averaged over all frames of each video. (Refer to Fig. 3 for New advert labeling)

Videos	Original video	NewAdvert	With occlusion handling	Without occlusion handling
Video 1	0.1911	Advert1	0.1973	0.1734
		Advert2	0.1984	0.1643
		Advert3	0.1965	0.1888
Video 2	0.0398	Advert1	0.0404	0.0384
		Advert2	0.0407	0.0384
		Advert3	0.0418	0.0410
Video 3	0.1265	Advert1	0.1295	0.1238
		Advert2	0.1268	0.1147
		Advert3	0.1267	0.1243
Video 4	0.2471	Advert1	0.2460	0.2414
		Advert2	0.2455	0.2334
		Advert3	0.2470	0.2430
Video 5	0.0790	Advert1	0.0780	0.0741
		Advert2	0.0770	0.0736
		Advert3	0.0785	0.0770
Video 6	0.2271	Advert1	0.2259	0.2194
		Advert2	0.2264	0.2166
		Advert3	0.2361	0.2253

advert without occlusion handling consistently reduces the occlusion ratio, suggesting that objects are incorrectly rendered behind the advertisement and therefore become partially or completely hidden.

The visualization for the occlusion handling for Video 1, Video 3 and Video 4 includes the target image, object mask, polygon mask, and the resulting occluded area, as shown in Figs. 12, 13 and 14 which further illustrate this effect qualitatively. When object occlusion

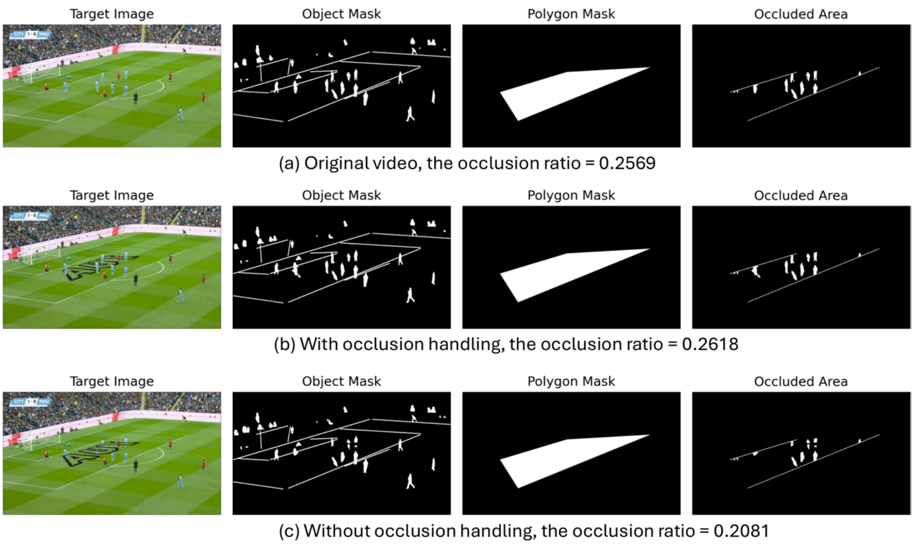


Fig. 12 Visualization of the occluded area for Video 1

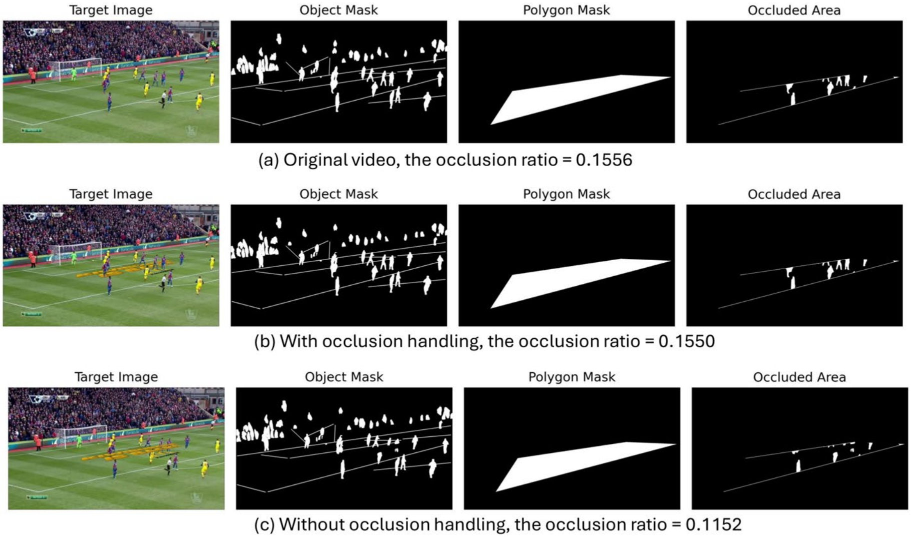


Fig. 13 Visualization of the occluded area for Video 3

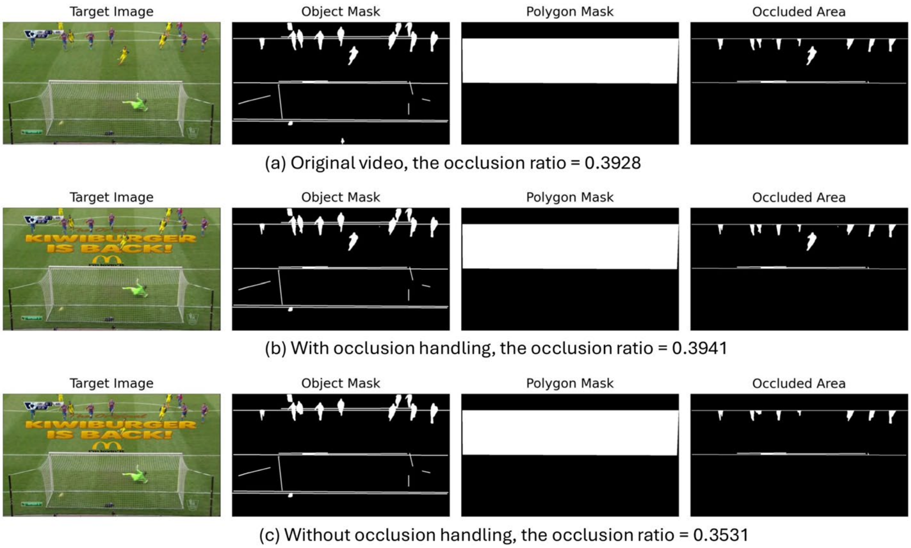


Fig. 14 Visualization of the occluded area for Video 4

is considered, the occluded regions closely match those observed in the original frames, ensuring that players correctly appear in front of the advertisement surface (see Figs. 12b, 13b and 14b). However, without considering occlusion, the advert is rendered on top of the scene, causing objects to disappear within the polygon region (see Figs. 12c, 13c and 14c). This qualitative observation is consistent with the quantitative results demonstrating that the proposed method effectively preserves realistic occlusion relationships during advert insertion. Similarly, we perform occlusion analysis for other videos.

The qualitative results for comparing advert placement approaches for Video1, Video2, Video3, Video4, Video5, and Video6 are shown in Fig. 15. For all the videos, showing results from left to right, the original frame is followed by advert placement without occlusion handling and with occlusion handling, highlighting the effect of occlusion-aware placement. Each figure compares advert placement with and without occlusion handling. Figure 15(a) presents advert placement results with and without occlusion handling, corresponding to representative example of Advert2 placement on Video1. Note that the advert labeling follows the convention defined in Fig. 3. Similarly, Fig. 15(b)-(f) illustrate the same comparison for Advert1, Advert2, and Advert3 across other videos.

The figures depict that without occlusion handling, the advert is unrealistically overlaid on objects such as players and the ball. In contrast, with occlusion handling, these objects are correctly rendered in front of the advert, resulting in more realistic integration.

4.3 Qualitative visuals

To assess robustness, we conducted visual experiments using advert images with different colors, sizes, and both transparent and non-transparent images for training images. Figure 16 illustrates advert placement using non-transparent images (Advert1, Advert2, Advert3 as shown in Fig. 3 top row), with different colors and background. Additionally, Fig. 17 present experiments using transparent adverts (Advert1, Advert2, Advert3 as shown in Fig. 3 bottom row). The visual results demonstrate that the proposed method remains robust across variations in color, sizes, background type (transparent vs. non-transparent), and visual appearance of the advertisement. The transparent advert better reflect real-world virtual advertising scenarios where seamless integration with the field texture is required. Therefore we considered, images with transparent backgrounds to allow the advert to integrate naturally with the field in our video dataset.

We evaluated our framework on six different video sequences using three distinct transparent advertisements. All quantitative results are reported on this video dataset. A detailed illustration of the complete pipeline from automatic polygon selection to final advertisement placement, including occlusion handling is presented for all video sequences in Fig. 18. The figure specifically demonstrates the placement of Advert1, Advert2 and Advert3. These examples visually demonstrate the improved realism achieved by incorporating occlusion-aware into the advert placement. The visualization results present representative examples selected to clearly demonstrate the observed effects across the dataset, as showing all frames would be redundant. The output videos are available online at vimeo.

4.4 Implementation details

All experiments were conducted on a high-performance computing (HPC) cluster equipped with NVIDIA A100-SXM4-80GB, used driver version 560.35.03, and CUDA version 12.6. NVLink. The experimental pipeline was implemented using Python 3.10.18 with PyTorch 2.7.1.

(a) Video 1

(b) Video 2

(c) Video 3

(d) Video 4

(e) Video 5

(f) Video 6

Fig. 15 We present the results for Video1, Video2, Video3, Video4, Video5, Video6. For each video, the sequence from left to right shows the original frame (first column), advert placement without occlusion handling (second column), advert placement with occlusion handling (third column)

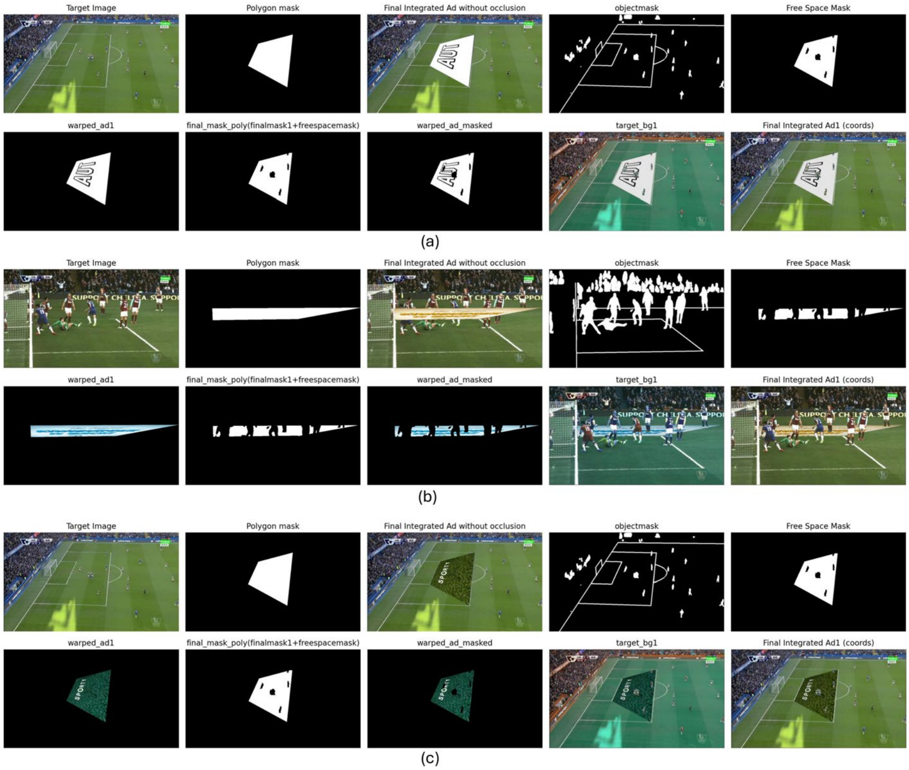


Fig. 16 Represents Advert placement. (**a**), (**b**), (**c**) Top Row (left to right): Target image, polygon mask for advert placement, advert placement without considering occlusion, object mask (detected occlusion entities), free space mask (visible region within the polygon). Bottom Row (left to right): Homography transformed of newad, final mask combining polygon and free space masks, warped advert with occlusion-aware masking, target background with occluding elements, final integrated advert with occlusion handling respecting occlusions and preserving scene realism

The computational complexity is approximately 207.95 GFLOPs. The average runtime per-frame to perform all operations is 13.69s.

5 Conclusion

Advert insertion in sports broadcasts is an increasingly popular way to boost sponsor visibility without interrupting live gameplay. Achieving realistic placement, however, requires careful handling of scene geometry and dynamic occlusions caused by players and the ball. We presented an occlusion-aware and perspective-consistent framework for realistic virtual advert placement in the soccer penalty area. The method automatically selects a geometrically valid quadrilateral within the penalty area with the predicted pitch extremities, which is then used as the target region for homography-based advert warping. To ensure visual realism, instance-level occlusion masks derived from Mask R-CNN are combined with Laplacian Alpha Blending, allowing the inserted advert to appear correctly

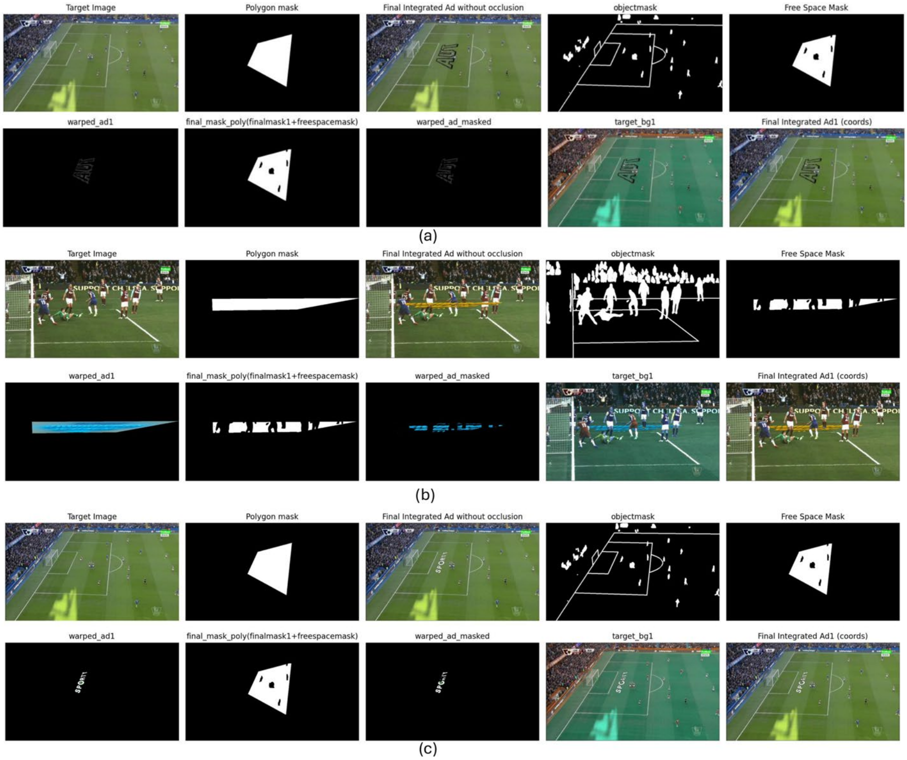


Fig. 17 Represents Advert placement. (**a**), (**b**), (**c**) Top Row (left to right): Target image, polygon mask for advert placement, advert placement without considering occlusion, object mask (detected occluding entities), free space mask (visible region within the polygon). Bottom Row (left to right): Homography transformed of newad, final mask combining polygon and free space masks, warped advert with occlusion-aware masking, target background with occluding elements, final integrated advert with occlusion handling respecting occlusions and preserving scene realism

behind players and the ball. This approach provides a robust, fully automated pipeline for integrating virtual content into soccer broadcasts while maintaining spatial accuracy and immersion. Quantitative results demonstrate minimal visual distortion, and only a marginal increase in flicker index (<7%) confirms strong temporal consistency. Quantitative evaluations demonstrate that our occlusion-aware advert placement method preserves high visual fidelity with an average SSIM of 0.97 and PSNR of 31.6 dB. The results show that the proposed occlusion handling method produces values that remain close to those of the original videos, indicating that the natural occlusion relationships between players and the advertisement region are preserved. These findings emphasize the importance of occlusion-aware blending to preserve realism and scene integrity. Overall, the framework produces visually coherent and realistic advert insertions, making it suitable for sports broadcast augmentation.

Future work will focus on incorporating a wider range of other soccer production footage, while also extending the approach to other sports such as basketball and cricket to assess its generalization and robustness across diverse broadcast scenarios.

(a) Video 1

(b) Video 2

(c) Video 3

(d) Video 4

(e) Video 5

(f) Video 6

Fig. 18 We present the results for Video1, Video2, Video3, Video4, Video5, Video6. For each video, the sequence from left to right shows: the original frame, extracted polygons in the penalty area, selected best polygon, and final advertisement placement with occlusion handling

Author Contributions S. Dhang conducted the experiments and wrote the manuscript text. F. Zheng, P. H. J. Chong, M. Zhang and S. Dev also wrote the manuscript text. All authors reviewed the manuscript.

Funding Open Access funding provided by the IReL Consortium. This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No. 18/CRT/6224. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. This research was conducted with the financial support of Science Foundation Ireland under Grant Agreement No. 13/RC/2106_P2 at the ADAPT SFI Research Centre at University College Dublin. ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology, is funded by Science Foundation Ireland through the SFI Research Centres Programme.

Availability of data and materials The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethical Approval Not applicable.

Competing interests The authors declare that they have no conflict of interest.

References

1. Atrish A, Singh N, Kumar V (2018) Enhanced homography-based sports image components analysis system. In: First International Conference on Artificial Intelligence and Cognitive Computing: AICC 2018, Springer, pp 495–505, https://doi.org/10.1007/978-981-13-1580-0_48
2. Buric M, Pobar M, Ivasic-Kos M (2018) Ball detection using yolo and mask r-cnn. In: 2018 International Conference on Computational Science and Computational Intelligence (CSCI), pp 319–323, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8947818
3. Cioppa A, Giancola S, Deliege A, et al (2022) Soccernet-tracking: Multiple object tracking dataset and benchmark in soccer videos. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3491–3502, https://openaccess.thecvf.com/content/CVPR2022W/CVSports/papers/Cioppa_SoccerNet-Tracking_Multiple_Object_Tracking_Dataset_and_Benchmark_in_Soccer_Videos_CVPRW_2022_paper.pdf
4. Citraro L, Márquez-Neila P, Savare S et al (2020) Real-time camera pose estimation for sports fields. Mach Vis Appl 31(3):16. https://doi.org/10.1007/s00138-020-01064-7
5. Deliege A, Cioppa A, Giancola S, et al (2021) Soccernet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 4508–4519, https://openaccess.thecvf.com/content/CVPR2021W/CVSports/papers/Deliege_SoccerNet-v2_A_Dataset_and_Benchmarks_for_Holistic_Understanding_of_Broadcast_CVPRW_2021_paper.pdf
6. Dev S, Javidnia H, Hossari M, et al (2019) Identifying candidate spaces for advert implantation. In: 2019 Ieee 7th International Conference on Computer Science and Network Technology (ICCSNT), IEEE, pp 503–507,https://doi.org/10.1109/ICCSNT47585.2019.8962510
7. Dhang S, Zhang M, Dev S (2024) Adsegnet: a deep network to localize billboard in outdoor scenes. SIViP 18(10):7221–7235. https://doi.org/10.1007/s11760-024-03388-7
8. Dhang S, Zhang M, Dev S (2025) LaBINet-an approach for seamlessly integrating new advertisement into an existing scene. IEEE Transactions on Artificial Intelligence 6(8):2281–229. https://doi.org/10.1109/TAI.2025.3544595
9. Giancola S, Amine M, Dghaily T, et al (2018) Soccernet: A scalable dataset for action spotting in soccer videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 1711–1721, https://openaccess.thecvf.com/content_cvpr_2018_workshops/papers/w34/Giancola_SoccerNet_A_Scalable_CVPR_2018_paper.pdf
10. He K, Gkioxari G, Dollár P, et al (2017) Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp 2961–2969, https://openaccess.thecvf.com/content_ICCV_2017/papers/He_Mask_R-CNN_ICCV_2017_paper.pdf
11. Homayounfar N, Fidler S, Urtasun R (2017) Sports field localization via deep structured models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 5212–5220, https://openaccess.thecvf.com/content_cvpr_2017/papers/Homayounfar_Sports_Field_Localization_CVPR_2017_paper.pdf

12. Husein AM, Ciptady KC, Suryadi R, et al (2023) Detecting and tracking player in football videos using two-stage mask r-cnn approach. In: IAIC International Conference Series, pp 132–138, https://doi.org/10.34306/conferenceseries.v4i1.643
13. Kanimozhi S, Mala T, Kaviya A, et al (2021) Key object classification for action recognition in tennis using cognitive mask rcnn. In: Proceedings of International Conference on Data Science and Applications: ICDSA 2021, Volume 2, Springer, pp 121–128, https://doi.org/10.1007/978-981-16-5348-3_9
14. Khan A, Lazzerini B, Calabrese G, et al (2018) Soccer event detection. Computer Science & Information Technology pp 119–129. https://arpi.unipi.it/bitstream/11568/931066.9/2/csit88509.pdf
15. Kiruthika SU, K. K, Raja SKS (2022) Object recognition in soccer sports videos. In: 2022 Smart Technologies, Communication and Robotics (STCR), pp 1–5, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10009543
16. Li J, Gong X, Li B (2020) End-to-end texture-aware and depth-aware embedded advertising for videos. In: Proceedings of the 2020 6th International Conference on Computer and Technology Applications, pp 57–61,https://doi.org/10.1145/3397125.3397133
17. Lin CH, Yumer E, Wang O, et al (2018) St-gan: Spatial transformer generative adversarial networks for image compositing. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 9455–9464, https://openaccess.thecvf.com/content_cvpr_2018/papers/Lin_ST-GAN_Spatial_Transformer_CVPR_2018_paper.pdf
18. Liu Y, Huang Q, Hui S (2024) Semantic-aware representation learning for homography estimation. Proceedings of the 32nd ACM International Conference on Multimedia. pp 2506–2514. https://doi.org/10.1145/3664647.3681021
19. Magera F, Hoyoux T, Barnich O, et al (2024) A universal protocol to benchmark camera calibration for sports. In: IEEE International Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), CVsports, Seattle, Washington, USA, https://openaccess.thecvf.com/content/CVPR2024W/CVsports/papers/Magera_A_Universal_Protocol_to_Benchmark_Camera_Calibration_for_Sports_CVPRW_2024_paper.pdf
20. Medioni G, Guy G, Rom H, et al (1999) Real-time billboard substitution in a video stream. In: Multimedia Communications. Springer, p 71–84,https://doi.org/10.1007/978-1-4471-0859-7_6
21. Niklaus S, Liu F (2018) Context-aware synthesis for video frame interpolation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1701–1710, https://openaccess.thecvf.com/content_cvpr_2018/papers/Niklaus_Context-Aware_Synthesis_for_CVPR_2018_paper.pdf
22. Pobar M, Ivasic-Kos M (2018) Mask R-CNN and optical flow based method for detection and marking of handball actions. In: 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), pp 1–6, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8633201
23. Ranftl R, Bochkovskiy A, Koltun V (2020) Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 44(3):1621–16. https://doi.org/10.1109/TPAMI.2020.3019967
24. Remez T, Huang J, Brown M (2018) Learning to segment via cut-and-paste. In: Proceedings of the European conference on computer vision (ECCV), pp 37–52, https://openaccess.thecvf.com/content_ECCV_2018/papers/Tal_Remez_Learning_to_Segment_ECCV_2018_paper.pdf
25. Sani MR (2023) Multi object tracking in soccer video focusing on occlusion detection and resolving. Multimedia Tools and Applications 82(23):35913–3594. https://doi.org/10.1007/s11042-023-14798-z
26. Tarashima S (2019) SFLNet: direct sports field localization via cnn-based regression. Asian Conference on Pattern Recognition. Springer, pp 677–690. https://doi.org/10.1007/978-3-030-41404-7_48
27. Wan K, Yan X (2007) Advertising insertion in sports webcasts. IEEE MultiMedia 14(2):78–https://doi.org/10.1109/MMUL.2007.26
28. Wang B (2025) Football sports video tracking and detection technology based on yolov5 and deepsort. Discover Applied Sciences 7(6):563. https://doi.org/10.1007/s42452-025-07116-9
29. Watve A, Sural S (2008) Soccer video processing for the detection of advertisement billboards. Pattern Recogn Lett 29(7):994–1006. https://doi.org/10.1016/j.patrec.2008.01.022
30. Yang C, Zheng F, Al-Hamid DZ, et al (2025) Deep learning based advertisement replacement on dynamic background videos. Journal of Image and Graphics 13(1). https://www.joig.net/2025/JOIG-V13N1-34.pdf
31. Zhang N, Izquierdo E (2023) A four-point camera calibration method for sport videos. IEEE Trans Circuits Syst Video Technol 33(8):3811–382. https://doi.org/10.1109/TCSVT.2023.3243126

Authors and Affiliations

Sukriti Dhang[1,4] · Fucheng Zheng[2] · Peter Han Joo Chong[2] · Mimi Zhang[3] · Soumyabrata Dev[1,3,4]

✉ Soumyabrata Dev
soumyabrata.dev@ucd.ie; devs@tcd.ie

Sukriti Dhang
sukriti.dhang@ucdconnect.ie

Fucheng Zheng
fucheng.zheng@aut.ac.nz

Peter Han Joo Chong
peter.chong@aut.ac.nz

Mimi Zhang
Mimi.Zhang@tcd.ie

1 School of Computer Science, University College Dublin, Dublin, Ireland

2 Auckland University of Technology, Auckland, New Zealand

3 School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

4 ADAPT SFI Research Centre, Dublin, Ireland