SmartCrop-H: AI-Based Cropping of Ice Hockey Videos

Mohammad Majidi*
OsloMet, Forzasys
Norway

Mehdi Houshmand Sarkhoosh*
OsloMet & Forzasys
Norway

Cise Midoglu
SimulaMet & Forzasys
Norway

Saeed S. Sabet
Forzasys
Norway

Tomas Kupka
Forzasys
Norway

Dag Johansen
UiT The Arctic University of Norway
Norway

Pål Halvorsen
SimulaMet, OsloMet, & Forzasys
Norway

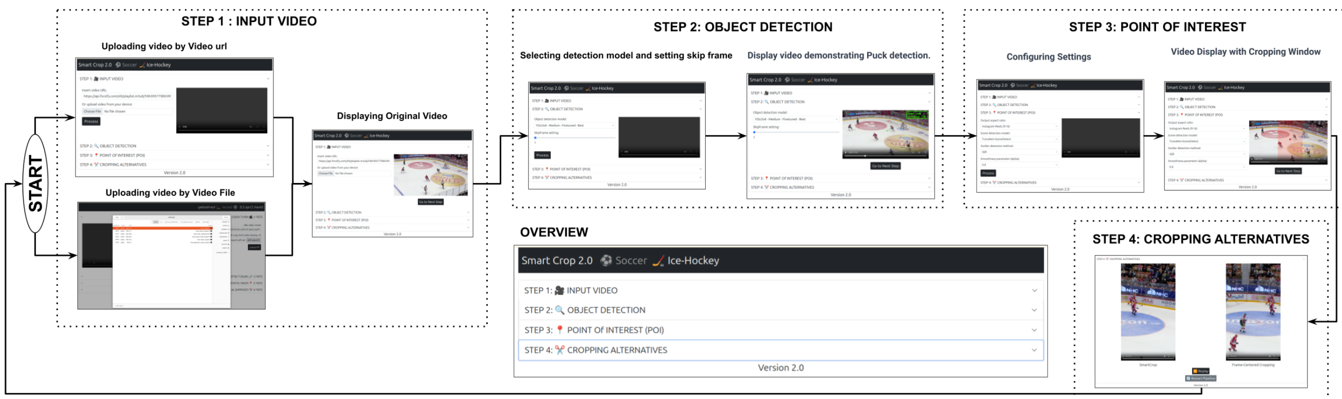


Figure 1: SmartCrop GUI application flow overview.

ABSTRACT

Sports multimedia plays a central role in captivating audiences on social media platforms. However, fast-paced sports such as ice hockey pose unique challenges due to their swift gameplay and the small puck size, making object tracking-based video adaptation for social media a complex task. In this context, we introduce SmartCrop-H, an innovative ice hockey video cropping tool powered by advanced AI models. It excels at tracking the puck and ensuring that crucial gameplay remains the center of attention, regardless of the desired target aspect ratio. The tool combines various techniques including object detection, scene detection, outlier detection, and smoothing, to deliver high-quality ratio-adapted videos. In this demonstration, we showcase SmartCrop-H in real-world scenarios through an intuitive step-by-step Graphical User Interface (GUI) that vividly illustrates how the tool works. The demonstration emphasizes the vital role of AI in enhancing the sports viewing experience, and its importance in the dynamic realm of social media content distribution. A video of the demo can be found here: https://youtu.be/rMmYOCM-k7A.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision tasks**; **Artificial intelligence**; **Machine learning**.

KEYWORDS

AI, video, cropping, aspect ratio, social media, ice hockey

ACM Reference Format:
Mohammad Majidi, Mehdi Houshmand Sarkhoosh, Cise Midoglu, Saeed S. Sabet, Tomas Kupka, Dag Johansen, and Pål Halvorsen. 2024. SmartCrop-H: AI-Based Cropping of Ice Hockey Videos. In *ACM Multimedia Systems Conference 2024 (MMSys '24), April 15–18, 2024, Bari, Italy*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3625468.3652195

1 INTRODUCTION AND BACKGROUND

In today's digital landscape, the demand for adaptive media content, particularly for globally popular sports is at an all-time high, and video aspect ratio retargeting is required for viewing content on different social media platforms and devices [23, 24]. Here,

*Both authors contributed equally to this research.

MMSys '24, April 15–18, 2024, Bari, Italy

ACM ISBN 979-8-4007-0412-3/24/04
https://doi.org/10.1145/3625468.3652195

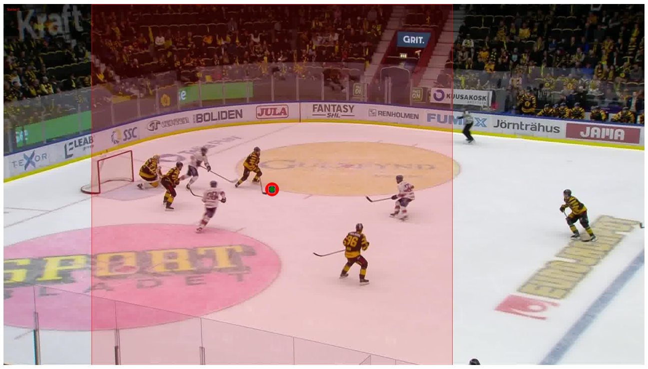


Figure 2: Cropping from 16:9 to 1:1 aspect ratio using object detection. Red dot = cropping center, red square = cropping area.

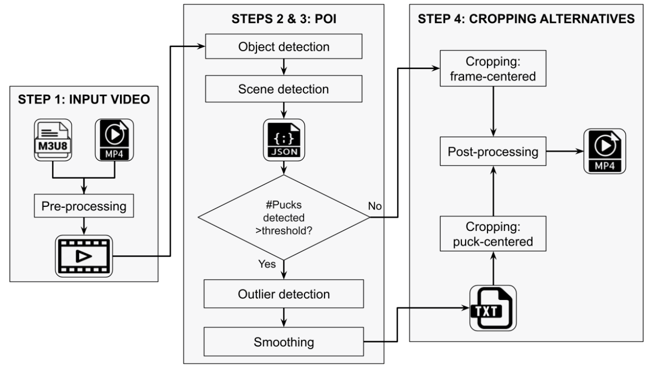


Figure 3: SmartCrop-H pipeline overview.

ice hockey, with its rapid gameplay and small puck size, presents unique challenges as cropping should be based on tracking of the salient objects. We increasingly find that traditional manual video cropping tools, such as Final Cut Pro, are inadequate for this task due to the need for high speed publishing based on real-time broadcasts. This shift towards automated solutions has occurred because they offer greater efficiency in general, but they still struggle with maintaining the visibility of the puck in dynamic and unpredictable ice hockey broadcasts [6].

Our work introduces SmartCrop-H, a novel pipeline for automated cropping of ice hockey videos, which optimizes the output for different social media platform representations, while ensuring the visibility of crucial gameplay elements. This solution builds upon existing aspect ratio adjustment methods and puck tracking technologies [1, 5, 8, 10, 12–14, 17, 20, 21, 29, 30, 32–34]. Advanced computer vision techniques have greatly improved puck tracking, which is critical given the puck's small size and the game's fast pace. However, professional broadcast environments have designed existing methodologies, often requiring high-definition video, which limits their generalizability. [4, 7, 18, 20].

SmartCrop-H addresses these limitations by integrating a modified YOLO model for puck detection, and adapting various techniques for scene detection, outlier detection, and smoothing [9, 31]. Our approach takes into account the fast pace of ice hockey, the characteristics of broadcast video properties, and the unique challenges of detecting the puck against a white background. We also highlight the commercial potential of our pipeline, recognizing the market gap for technologies tailored towards the dynamics of ice hockey, as seen in companies such as Magnifi, Pendular, Backlight, and WSC Sports, which focus on personalized highlight generation and video aspect ratio adjustments [2, 15, 19, 28]. Here, we demonstrate SmartCrop-H through a step-by-step tool as illustrated in Figure 1.

2 SMARTCROP-H PIPELINE

The core principle of the SmartCrop-H pipeline, inspired by the SmartCrop concept originally tailored for soccer [6, 16], is based on using a Point of Interest (POI) as the center of the cropping area, as illustrated in Figure 2. In our case, for ice hockey, the puck is designated as the POI. When in view, the puck is the primary focus for cropping, when it's not, our enhanced object detection and tracking techniques, in conjunction with outlier detection, identify an alternative focal point.

Figure 3 highlights our pipeline's functionality which is marked by several key steps: (Step 1) input video pre-processing, (Steps 2 and 3) combined object detection and scene detection, outlier detection, and smoothing, and finally (Step 4) cropping and post-processing. These modules are thoughtfully integrated, with intermediate logic to ensure the video is optimally processed for various aspect ratios, maintaining the focus on pivotal gameplay elements. The pipeline accepts an HTTP Live Streaming (HLS) playlist or a VoD file (URL or local) as input, and produces an MP4 file as output.

2.1 Pre-processing

Our pipeline handles various video formats. When processing HLS playlists, it selects the lowest quality stream to expedite object detection and reduce computational load. While this may reduce resolution, it minimally affects output quality, relying more on subsequent detection models. The approach significantly accelerates processing, irrespective of input quality spectrum. The chosen stream is then converted to H.264 format for optimal processing. Pre-processing also initializes necessary detection models for seamless pipeline transition.

2.2 Object Detection

In the development of this module, we trained various object detection models, including the versatile YOLOv8 Medium and our specialized Y8_Sc_m model. Our primary objective was to accurately identify pucks in ice hockey videos. Notably, the Y8_Sc_m model[1], trained on a dataset of 800 images from the Swedish Hockey League (SHL) [11], demonstrated exceptional performance with a 77% true positive rate for puck detection. This highlights its effectiveness in this specific task.

Additionally, we measured the Mean Average Precision (MAP) at an Intersection over Union (IoU) threshold of 0.5, which yielded a value of 0.576, indicating moderate accuracy across the model's predictions. During its operation, the model strikes a balance between

[1] Y8_Sc_m model weights: https://github.com/forzasys-students/SportsVision-YOLO

precision and recall. Initially, it exhibits high precision, signifying strong specificity. However, as the model aims to capture more true positives, there is a discernible decline in precision, showcasing the inherent challenge of balancing these aspects in object detection tasks.

2.3 Scene Detection

SmartCrop-H employs a dual-model approach for effective scene detection, incorporating both the TransNetV2 machine learning model [27] and the SceneDetect Python library [3, 25]. TransNetV2 is utilized for its proficiency in segmenting individual scenes, an essential capability for our pipeline's context-aware video cropping. In tandem, SceneDetect provides versatility in scene detection across varied video environments, proving particularly useful in hockey video analysis. The incorporation of SceneDetect within SmartCrop-H is vital for addressing the distinctive challenges posed by hockey videos, which often feature unique patterns of brightness and pixel composition. We have undertaken specific enhancements to tailor SceneDetect to these hockey-specific attributes, ensuring optimal performance. The SceneDetect enhancements for hockey videos:

- `luma_only`: `false` (Considers both luminance and color).
- `adaptive_threshold`: `1.5` (Optimizes sensitivity).
- `min_scene_len`: `140` (Ensures valid scene lengths).
- `min_content_val`: `20` (Threshold for significant content changes).
- `window_width`: `[15, 20, 25]` (Adapts to diverse pacing).

2.4 Outlier Detection

This module implements various techniques to eliminate anomalous data points from the predicted puck positions, enhancing the accuracy of our system. We support 3 alternative methods, from among which a selection can be made via configuration parameters:

- **Z-Score:** Outliers are identified by calculating their Z-scores and comparing them to a standard threshold [22].
- **Modified Z-Score:** This method uses the median and Median Absolute Deviation (MAD) to pinpoint outliers. A data point is considered an outlier if it significantly deviates from the median of the dataset of puck positions based on a predetermined threshold [22]. Choose for non-normal distributions or robustness against outliers.
- **Interquartile Range (IQR):** Outliers are determined using the IQR, defining any values outside a certain range (based on the IQR) as outliers. This range is typically set using a scaling factor [22]. For our pipeline, IQR provides the best performance. Outlier detection assessed with MAE: **IQR = 95.79**, outperforms Modified Z-Score (MAE = 125.131) and Z-Score (MAE = 125.230).

2.5 Smoothing

To facilitate a seamless transition between frames with respect to the POI, we have incorporated the Exponential Moving Average (EMA) technique [26]. EMA is a method for smoothing data series, prioritizing recent observations while diminishing the influence of older ones. It employs weighted averages with exponentially decreasing weights, giving greater importance to the most recent data and less to historical observations. EMA is calculated using the formula:

$$EMA_t = \alpha \times P_t + (1 - \alpha) \times EMA_{t-1} \tag{1}$$

where EMA_t represents the EMA value at time t, P_t is the current data point at time t, EMA_{t-1} is the EMA value at the previous time step, and α is the smoothing factor, typically within the range of (0, 1). The parameter α governs the degree of weighting reduction, with higher values giving greater emphasis to recent data points, making the EMA more responsive to new information. Conversely, smaller α values extend the influence of older data points, resulting in a more stable EMA output over time. In essence, EMA strikes a balance between the immediacy of the latest data point and the continuity of historical data, yielding a refined output.

2.6 Cropping

In this module, each frame is cropped using ffmpeg, based on the target aspect ratio specified using the relevant configuration parameter. To determine the center of the cropping area, we explore two primary options:

- **Frame-Centered Cropping:** This approach involves statically selecting the cropping center as the middle of each frame.
- **Puck-Centered Cropping:** Here, the cropping center is determined based on the coordinates of the puck as detected and calculated by the detection and smoothing modules, resulting in the output video from the SmartCrop-H pipeline, as shown in Figure 3.

2.7 Post-Processing

This module is responsible for generating an MP4 video file from the cropped frames, serving as the output of the pipeline. It also offers supplementary capabilities for organizing processed data, facilitating visualization, summarization, and subsequent analysis.

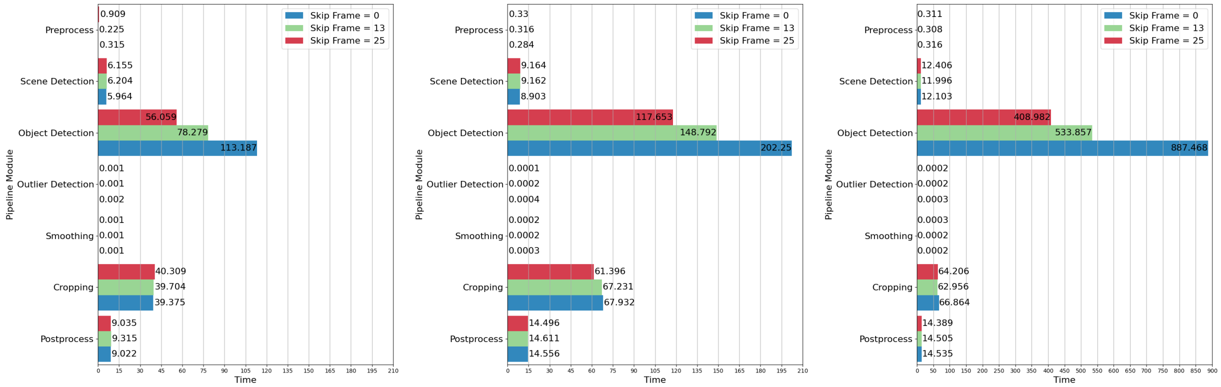
3 EVALUATION

3.1 System Performance

We tested the SmartCrop-H pipeline in a local deployment using a system with an NVIDIA Tesla T4 GPU. The test environment included a high-performance CPU and was assessed using a 20-second video clip, providing insights into the pipeline's performance under different configurations.

Execution time: We measured the execution time for each pipeline module, noting consistent performance across various configurations, particularly in modules like scene detection and post-processing. The object detection module showed variability in execution time influenced by the Skip Frame parameter.

CPU vs. GPU: Monitoring of CPU and GPU loads revealed dynamic resource allocation based on task requirements, with the object detection module significantly leveraging GPU resources. The comparison of CPU and GPU performance also underscored the superior performance of GPU-accelerated configurations, particularly in reduced execution times (S). The CPU-only scenario



(a) 20 second video on GPU. (b) 30 second video on GPU. (c) 30 second video on CPU.

Figure 4: Runtime per module on GPU and CPU, with different SkipFrame settings, for 20 and 30 second videos.

	Centering	Description	O	S
1	frame-centered	static no padding	✗	✗
2	frame-centered	static w/black padding to 16:9	✗	✗
3	puck-centered	use last detected puck position	✗	✗
4	puck-centered	w/smoothing	✗	✔
5	puck-centered	w/outlier detection	✔	✗
6	puck-centered	w/outlier detection & smoothing	✔	✔

Table 1: Alternative cropping types used in the subjective evaluation (O: outlier detection, S: smoothing).

highlighted the limits of CPU processing, particularly with high-definition video inputs and complex detection algorithms, emphasizing the importance of GPU acceleration.

Video duration: Testing the impact of video duration on system performance with an extended 30-second video indicated a linear increase in processing time, demonstrating the scalability of our pipeline.

3.2 Subjective User Experience

We conducted a comprehensive user study to evaluate the perceptual impact of different video cropping techniques implemented in SmartCrop-H. This study was pivotal in understanding how various cropping methods influence viewer satisfaction and content appreciation.

Experimental setup: We used an online survey where participants were instructed to view a series of videos on a mobile device, replicating a typical real-world setting for engaging with content. Each participant was exposed to 6 alternative cropping methods (as listed in Table 1, with alternative 6 corresponding to SmartCrop-H with full functionality) applied to 4 different videos. The videos, both originally in a 16:9 aspect ratio, were cropped to either 9:16 or 1:1 aspect ratio, creating diverse viewing experiences.

Evaluation criteria: Participants rated their viewing experience based on three key aspects: overall Quality of Experience (QoE), smoothness of window transitions, and the effectiveness of each cropping method in preserving the essence of the original content.

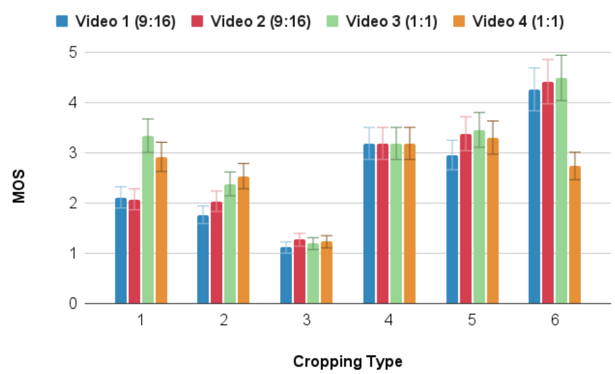


Figure 5: QoE ratings for different target aspect ratios.

Participants: 26 individuals participated in the study, with diverse backgrounds in terms of age, gender, social media usage, and video editing experience. This diversity ensured that our findings were representative of a wide range of potential users.

Results and insights: Figure 5 illustrates the Mean Opinion Score (MOS) for QoE, revealing that as a group, smart cropping alternatives 4, 5, and 6 enhance QoE compared to the static cropping alternatives 1, 2, and 3, for both target aspect ratios, which demonstrates the importance of smart cropping in general, and more specifically outlier detection and smoothing. Notably, there was no significant difference between smart cropping types 4 and 5, indicating that outlier detection and smoothing were of relatively similar importance. Overall, we see that the SmartCrop-H pipeline with its full functionality (alternative 6) outperforms all other alternatives for both videos with the 9:16 target aspect ratio, and one video with the 1:1 target aspect ratio, where Video 4 is the notable exception. We observe that videos cropped to the 1:1 aspect ratio generally achieve higher QoE scores across the board compared to the 9:16 videos, again with the exception of Video 4 with cropping alternative 6. Static cropping alternatives 1 and 2, which could maintain a reasonable level of QoE for the 1:1 target

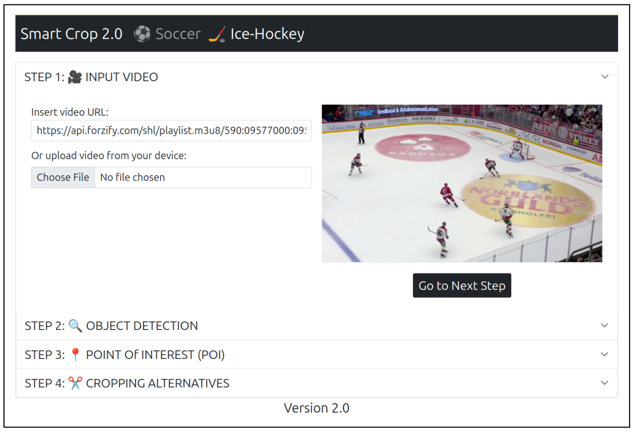


Figure 6: Uploading the input video (Step 1).

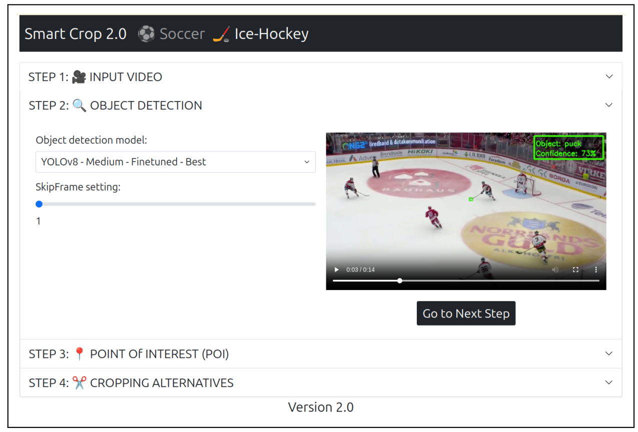


Figure 7: Puck detection (Step 2).

aspect ratio fell short in the 9:16 target aspect ratio, potentially due to the limited cropping window. The MOS ratings for video smoothness and user ratings for content closely align with the QoE results.

4 DEMONSTRATION

The SmartCrop-H GUI presents the workflow of the SmartCrop-H in 4 configurable steps, as illustrated by Figure 1, demonstrated in the video https://youtu.be/rMmYOCM-k7A, and explained below.

4.1 Step 1: Input Video

As depicted in Figure 6, the process begins with the user selecting an input video. There are two methods catering to user convenience based on the source of their video content:

- **Playlist/video URL:** Users can submit a direct URL for online-hosted videos (HLS playlist or VoD). The system support multiple video formats, such as `.mp4`, `.avi`, `.mov`, and `.ts`, and supports HLS streaming via `.m3u8` links.
- **File upload:** The platform allows direct video file uploads, optimally supporting the MP4 format. Leveraging FFmpeg, it also facilitates conversion and processing of various other video formats, enhancing usability and format compatibility.

Once the desired video is selected or the URL is entered, the user can initiate the upload process by clicking the 'Process' button.

4.2 Step 2: Object Detection

In this step, as depicted in Figure 7, the interface provides users with the option to select from a range of fine-tuned YOLOv8 models for object detection. These models include nano, small, medium, large, and xlarge. The YOLOv8 medium is particularly recommended for its balanced trade-off between speed and accuracy, making it an optimal choice for most scenarios.

An essential feature provided in this step is the 'Skip Frame' setting, which ranges from 1 to 50. This parameter allows users to adjust the frequency at which frames are analyzed for object detection. By increasing the 'Skip Frame' value, the system analyzes fewer frames per second. This reduction can significantly speed up the processing time, especially beneficial when dealing with lengthy videos or when quick results are more crucial than detailed frame-by-frame analysis. However, it is important to note that a higher 'Skip Frame' value may result in decreased detection accuracy, as fewer frames are examined, potentially missing crucial moments or details. Thus, users must find a balance based on their specific needs and the nature of the video content. Fine-tuning this parameter provides flexibility and control, enabling users to tailor the object detection process to their specific requirements and constraints.

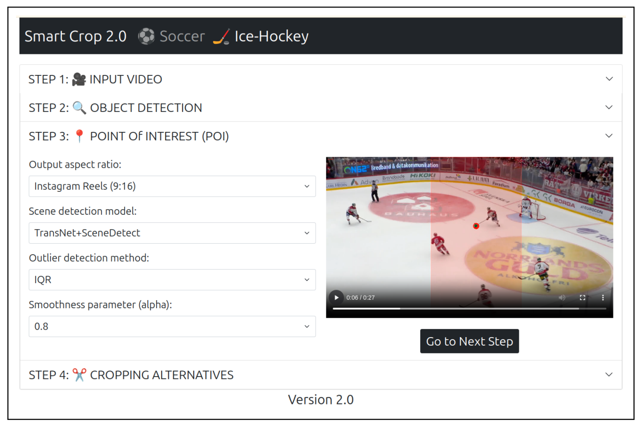


Figure 8: Dynamic POI identification and cropping (Step 3).

Figure 7 illustrates the user interface for this step, showcasing the model selection dropdown and the 'SkipFrame' slider, which users can adjust to optimize the object detection process according to their needs.

4.3 Step 3: Points of Interest (POI)

In this crucial phase(8), users are provided with several configuration options to fine-tune the identification of POI in the video. The settings include:

Aspect Ratio Configuration: This dropdown allows users to select the desired aspect ratio for the output video. These options enable users to tailor the video output for specific platforms or viewing experiences.The available options are:

- *9:16 (Instagram Reels, TikTok)*: Ideal for mobile-oriented vertical videos.
- *1:1 (Instagram Post, Facebook Post)*: Suitable for square format, widely used on social media platforms.
- *4:5*: A slightly taller variation than the square format, commonly used for Instagram posts.
- *1:2.4 (LinkedIn)*: A wide aspect ratio, typically used for professional or business-oriented content.

Scene Detection Configuration: Users can choose between different scene detection models to identify changes in scenes or shots within the video. The available models are:

- *TransNet V2*: An advanced neural network model specialized in detecting hard cuts and gradual transitions in videos.
- *Scene Detect Model*: A general-purpose scene detection tool that analyzes the video for sudden changes.
- *Both*: Utilizes both TransNet V2 and Scene Detect Model for comprehensive scene detection coverage.

Outlier Detection Configuration: This option allows users to select a method for detecting and handling outliers in POI data. The methods, as described in section 2.4, include *Z-Score*, *Modified Z-Score*, and *IQR*.

Smoothness Configuration: Set alpha values to adjust POI path smoothness in exponential smoothing (Eq. (1)). Options:

- *Alpha=0.8*: Higher weight to recent data, more sensitive to recent changes.
- *Alpha=0.6*: Balanced weighting between recent and past data.
- *Alpha=0.4*: Increased weight to historical data, less influenced by recent changes.
- *Alpha=0.2*: Predominantly weights past observations, minimizing recent data impact.

Upon configuring these settings, clicking the 'Process' button processes the video. The output is displayed with a red box highlighting the cropped area, indicating the dynamically identified points of interest based on the selected configurations. This visualization, as illustrated in Figure 8, assists users in understanding how the POI is determined and cropped in different scenes of the video.

Best Configuration: After thorough subjective and objective evaluations, default settings were determined for the best POI identification accuracy and efficiency. These configurations, derived from extensive testing, ensure effective performance across diverse video content, providing users with a well-balanced mix of accuracy, efficiency, and visual fluidity.

- **Object Detection Model:** *Finetuned YOLOv8 medium* - Balances accuracy and speed effectively.
- **Skip Frame Setting:** *SkipFrame=1* - Ensures comprehensive frame-by-frame detection.
- **Scene Detection:** *TransNet V2 + SceneDetect* - Combines models for robust scene detection.
- **Outlier Detection:** *Interquartile Range (IQR)* - Reliable for managing outliers.
- **Smoothness (Alpha):** *Alpha=0.8* - Ideal for smooth POI paths in videos with steady movements.

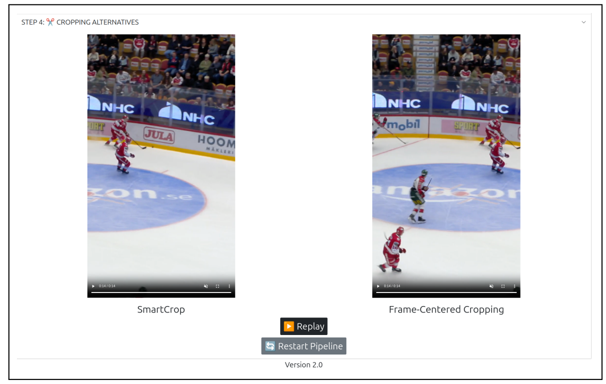


Figure 9: Visual comparison of SmartCrop and frame-centered cropping (Step 4).

4.4 Step 4: Cropping Alternatives

In the final stage, SmartCrop-H offers an intuitive side-by-side comparative analysis for evaluating its smart cropping algorithm. This vital comparison highlights the advantages of SmartCrop-H's dynamic, content-aware cropping over standard static cropping.

- **The SmartCrop-H Output:** This video showcases the results of the smart cropping algorithm applied by SmartCrop-H. It reflects all the configurations and settings applied in the previous steps, offering a custom-tailored cropping based on the identified points of interest and scene transitions.
- **A Frame-Centered Crop:** This video serves as a baseline comparison, utilizing a standard frame-centered cropping approach. It provides a uniform crop across the entire video, centered around the middle of the frame, without any dynamic adjustments based on the content of the video.

The 'Re Play' button lets users replay both videos simultaneously for detailed comparison of scene handling and points of interest. The 'Restart Pipeline' feature enables users to return to the start of the process for adjustments or to try different settings, promoting a flexible and user-centric approach to video analysis.

Figure 9 illustrates the user interface for this step, showcasing the comparative functionality between the two cropping methods.

5 CONCLUSION

SmartCrop-H stands as a groundbreaking tool in the realm of sports video processing, here demonstrated particularly for ice hockey. It adeptly addresses the unique challenges posed by the sport's fast pace and the small size of the puck. The tool integrates advanced AI models for object detection with a user-friendly graphical interface, enabling precise tracking of the puck and dynamic video cropping tailored for various social media platforms.

ACKNOWLEDGMENTS

This research was partly funded by the Research Council of Norway, project number 346671 (AI-Storyteller).

REFERENCES

[1] K. Apostolidis and V. Mezaris. 2021. A fast smart-cropping method and dataset for video retargeting. In *Proc. of IEEE ICIP*. 2618–2622.

- [2] Backlight. 2023. Cloud-Based video CMS, playout, distribution and app builder. https://www.backlight.co/streaming.
- [3] Brandon Castellano. 2023. SceneDetect. https://github.com/Breakthrough/PySceneDetect/tree/main.
- [4] Rick Cavallaro. 1997. The foxtrax hockey puck tracking system. *IEEE computer graphics and applications* 17, 2 (1997), 6–12.
- [5] T. Deselaers, P. Dreuw, and H. Ney. 2008. Pan zoom scan – time-coherent trained automatic video cropping. In *Proc. of IEEE CVPR*. 1–8.
- [6] Sayed Mohammad Majidi Dorcheh, Mehdi Houshmand Sarkhoosh, Cise Midoglu, Saeed Shafiee Sabet, Tomas Kupka, Michael A. Riegler, Dag Johansen, and Pal Halvorsen. 2023. SmartCrop: AI-Based cropping of soccer videos. In *Proc. of IEEE ISM*.
- [7] Xin Duan. 2011. *Automatic determination of puck possession and location in broadcast hockey video*. Ph. D. Dissertation. University of british columbia.
- [8] E. Jain, Y. Sheikh, A. Shamir, and J. Hodgins. 2015. Gaze-driven video re-editing. *ACM TOG* 34, 2 (2015), 1–12.
- [9] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. 2023. *Ultralytics YOLOv8*. https://github.com/ultralytics/ultralytics
- [10] H. Kaur, S. Kour, and D. Sen. 2019. Video retargeting through spatio-temporal seam carving using kalman filter. *IET Image Processing* 13, 11 (2019), 1862–1871.
- [11] Swedish Hockey League. 2023. SHL. https://www.shl.se/shl-in-english.
- [12] H.-S. Lee, G. Bae, S.-I. Cho, Y.-H. Kim, and S. Kang. 2019. Smartgrid: video retargeting with spatiotemporal grid optimization. *IEEE Access* 7 (2019), 127564–127579.
- [13] Muyu Li, Henan Hu, and Hong Yan. 2023. Ice hockey puck tracking through broadcast video. *Neurocomputing* (2023), 126484.
- [14] F. Liu and M. Gleicher. 2006. Video retargeting: automating pan and scan. In *Proc. of ACM MM*. 241–250.
- [15] Magnifi. 2023. Automated sports highlights, AI video highlights. https://magnifi.ai/.
- [16] Cise Midoglu, Saeed Shafiee Sabet, Mehdi Houshmand Sarkhoosh, Mohammad Majidi, Sushant Gautam, Håkon Maric Solberg, Tomas Kupka, and Pål Halvorsen. 2024. AI-Based Sports Highlight Generation for Social Media. In *Proceedings of the 3rd Mile-High Video Conference on Zzz (MHV '24)*. Association for Computing Machinery, New York, NY, USA, 7–13. https://doi.org/10.1145/3638036.3640799
- [17] H. Nam, D. Park, and K. Jeon. 2020. Jitter-robust video re-targeting with kalman filter and attention saliency fusion network. In *Proc. of IEEE ICIP*. 858–862.
- [18] Juha-Pekka Parto. 2021. *Ice hockey checking detection from indoor localization data*. Master's thesis. Tampere Univeristy, Finland.
- [19] Pendular. 2023. Content generation technology for broadcasters and publishers. https://pendular.io/.
- [20] Hemanth Pidaparthy and James Elder. 2019. Keep your eye on the puck: Automatic hockey videography. In *Proc. of IEEE WACV*. 1636–1644.
- [21] K.-K. Rachavarapu, M. Kumar, V. Gandhi, and R. Subramanian. 2018. Watch to edit: video retargeting using gaze. *Computer Graphics Forum* 37 (2018), 205–215.
- [22] Sehar Saleem, Maria Aslam, and Mah Rukh Shaukat. 2021. A review and empirical comparison of univariate outlier detection methods. *Pakistan Journal of Statistics* 37, 4 (2021).
- [23] Mehdi Houshmand Sarkhoosh, Sayed Mohammad Majidi Dorcheh, Sushant Gautam, Cise Midoglu, Saeed Shafiee Sabet, and Pål Halvorsen. 2023. Soccer on social media. *arXiv preprint arXiv:2310.12328* (2023).
- [24] Mehdi Houshmand Sarkhoosh, Dorceh Sayed Mohammad Majidi, Cise Midoglu, Saeed Shafiee Sabet, Tomas Kupka, Michael A. Riegler, Dag Johansen, and Pal Halvorsen. 2024. AI-Based cropping of soccer videos for different social media representations. In *Proc. of MMM*.
- [25] Zheng Shou, Junting Pan, Jonathan Chan, Kazuyuki Miyazawa, Hassan Mansour, Anthony Vetro, Xavier Giro-i Nieto, and Shih-Fu Chang. 2018. Online detection of action start in untrimmed streaming videos. In *Proc. of ECCV*. 534–551.
- [26] Robert H Shumway, David S Stoffer, and David S Stoffer. 2000. *Time series analysis and its applications*. Vol. 3. Springer.
- [27] Tomás Soucek and Jakub Lokoc. 2020. TransNet V2: an effective deep network architecture for fast shot transition detection. *CoRR* (2020). arXiv:2008.04838
- [28] Wsc sport. 2023. Wsc sport. https://wsc-sports.com/.
- [29] Kanav Vats, Mehrnaz Fani, David A Clausi, and John Zelek. 2021. Puck localization and multi-task event recognition in broadcast hockey videos. In *Proc. of the IEEE/CVF CVPR*. 4567–4575.
- [30] Kanav Vats, William McNally, Chris Dulhanty, Zhong Qiu Lin, David A. Clausi, and John Zelek. 2021. PuckNet: Estimating hockey puck location from broadcast video. arXiv:1912.05107 [cs.CV]
- [31] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. 2022. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696* (2022).
- [32] S. Wang, Z. Tang, W. Dong, and J. Yao. 2020. Multi-operator video retargeting method based on improved seam carving. In *Proc. of IEEE ITOEC*. 1609–1614.
- [33] Mehmet Yakut and Nasser Kehtarnavaz. 2016. Ice-hockey puck detection and tracking for video highlighting. *Signal, Image and Video Processing* 10 (2016), 527–533.
- [34] Xionghao Yang. 2021. Where is the puck? Tiny and fast-moving object detection in videos.