

# AI-Based Sports Highlight Generation for Social Media

Cise Midoglu<sup>1,3</sup>, Saeed Shafiee Sabet<sup>1,3</sup>, Mehdi Houshmand Sarkhoosh<sup>2,3</sup>,  
Sayed Mohammad Majidi Dorcheh<sup>2,3</sup>, Sushant Gautam<sup>1,2</sup>, Tomas Kupka<sup>3</sup>, Pål Halvorsen<sup>1,2,3</sup>

<sup>1</sup> SimulaMet, Norway

<sup>2</sup> Oslo Metropolitan University, Norway

<sup>3</sup> Forzasys, Norway

Social media plays a significant role for sports organizations with millions of active fans [1-8], but social media publishing is often a tedious manual operation. With the development of AI, new tools are available for content generation and personalization to engage audiences [9-22]. We present an AI-based multimedia production pipeline for automatic soccer and ice hockey highlight sharing on social media, which ingests a broadcast (VoD / stream) and performs event detection, clipping, smart cropping (aspect ratio retargeting), thumbnail generation, summarization, and social media sharing.

**Event detection:** Automatic event detection from soccer broadcasts suffers traditionally from high latency and low accuracy, making it challenging to operate at the live edge. We present an algorithm to detect events in soccer videos such as goals, cards, substitutions in real-time, using 3D convolutional neural networks [23,24]. Tested on three datasets including SoccerNet [25], our algorithm can detect events with high recall, low latency, and accurate time estimation. Sports broadcasts contain both video and audio streams, but detection approaches often focus solely on the visual modality. We evaluate the possibility of increasing detection performance by utilizing both modalities.

**Clipping:** The most time-consuming and expensive operation is the extraction of highlight clips, since manual annotations and trimming (where human operators define the start and end of an event and trim away the unwanted scenes) are traditionally used. We automate this process using scene boundary detection, logo detection, and optional cheering removal. We experiment with different neural network architectures and present two models that have been evaluated both quantitatively and qualitatively with a user study [26]. Our AI-based solution can identify appropriate event intervals, e.g., scene changes, game turnovers, logo transitions, cheering, and replays.

**Cropping:** SmartCrop is a tool for cropping videos to fit desired aspect ratios which are supported by different social media platforms, based on tracking Points of Interest (POI), with the soccer ball / hockey puck serving as the primary POI. If the ball / puck isn't visible, outlier detection and interpolation are used. It segments videos using the TransNetV2 model and employs object detection through a custom YOLOv8-medium model [27], which has been retrained on a novel soccer and hockey dataset. For

instances where the ball or puck isn't visible, the system utilizes outlier detection methods (average, IQR, Z-Score). Interpolation techniques (linear, polynomial, ease-in-out, heuristic) are applied to enhance the video quality and accuracy. Frames are then cropped either around the POI or the frame's center, based on user preferences and POI visibility. Before sharing, we format the content to meet the unique requirements of social media platforms, ensuring optimal viewing experiences across all channels.

**Thumbnail generation:** Thumbnails must capture the essence of highlight clips and engage viewers, but are time-consuming to generate. We extend the HOST-ATS framework [28,29], which performs logo detection, close-up shot detection, face detection, blur and image quality analysis on frames, to rank them as thumbnail candidates according to relevance and quality, with a postprocessing module to introduce overlays supporting multimodality.

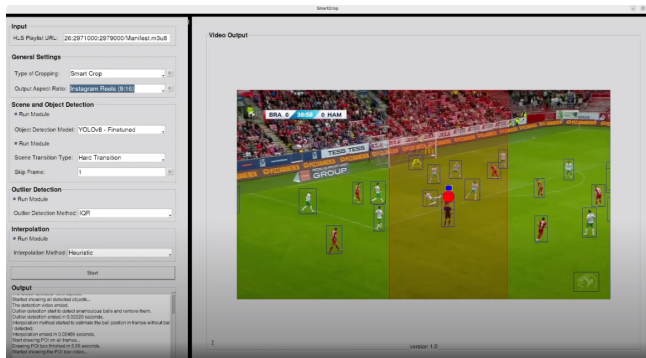
**Summarization:** We craft a summary (highlight description augmented with metadata) tailored to specific timeframes and priorities [17-19]. To improve the textual output, we rely on multimodal scene understanding based on architecture from [30-33], use modality specific encoders for each multimedia type in the highlight (video, audio, image, text), and adapter layers trained on an instruction dataset (QA), followed by a conversational LLM. We also recommend pertinent hashtags and captions derived from the video content, match specifics, and current trending topics. This strategy ensures that our content is discoverable by a broader audience, thereby increasing fan engagement [34].

**Sharing:** We can directly publish the generated highlights to various social media platforms [35] through integration with their APIs (TikTok, Instagram). The use of AI not only enhances the production capabilities but also provides valuable insights: analyzing user engagement metrics from shared content allows us to continually refine algorithms, ensuring that the highlights resonate with audience preferences.

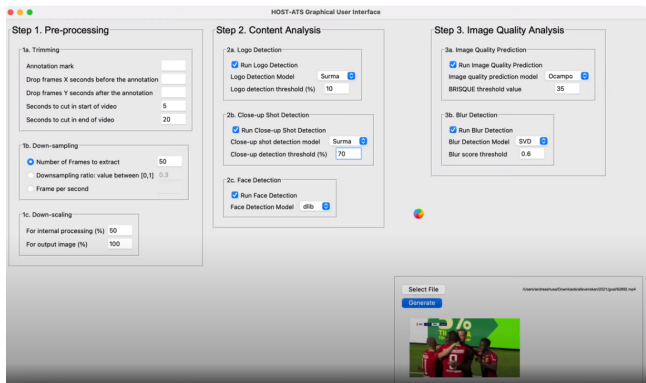
**Next steps:** We have tested our proposed pipeline on content from Scandinavian soccer and ice hockey leagues, and ran subjective user studies for evaluation, which show that an AI-based automated approach relying on advanced multimodal scene understanding facilitates social media engagement and increases QoE.

APPENDIX - Demonstration

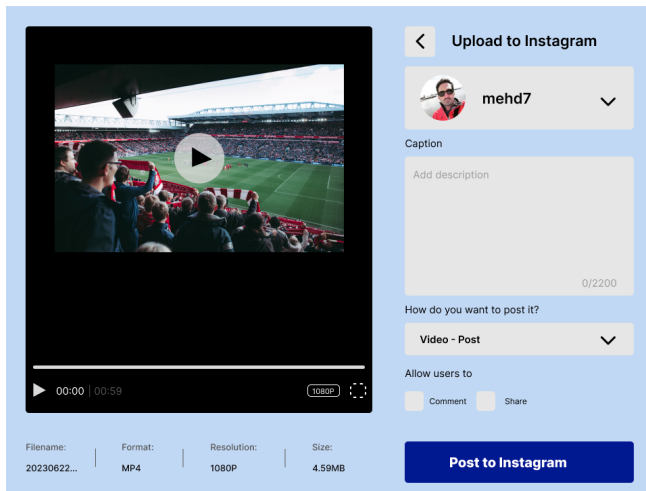
Cropping:



Thumbnail selection:



Social media sharing:



Demonstration videos for individual components are also available (thumbnail selection: <http://youtu.be/HHMCdMucorI>, <http://youtu.be/VZQaEy2VauQ>, cropping: <http://youtu.be/aqqPWfrPmsE>).

## REFERENCES

- [1] McCarthy et al., Social media marketing strategy in English football clubs. *Soccer & Society*, pages 513–528, April 2022.
- [2] Obradovic et al., Use of Social Media in Communication Strategies of Premier league Football Clubs. *Proceedings of the International Scientific Conference - Sinteza 2019*, 2019.
- [3] Yun et al., Drivers of soccer fan loyalty: Australian evidence on the influence of team brand image, fan engagement, satisfaction and enduring involvement. *Asia Pacific Journal of Marketing and Logistics*, 33(3):755–782, July 2020.
- [4] Trail et al., A longitudinal study of team-fan role identity on self-reported attendance behavior and future intentions. *JAS*, 3(1), March 2017.
- [5] Alduayji, Online co-creation behavior in a sports context. the UWA Profiles and Research Repository, 2019.
- [6] Winand et al., International Sport Federations' Social Media Communication: A Content Analysis of FIFA's Twitter Ac-count. *International Journal of Sport Communication*, 12(2):209–233, June 2019.
- [7] Eddy, et al., Examining Engagement With Sport Sponsor Activations on Twitter. *International Journal of Sport Communication*, 14(1):79–108, January 2021.
- [8] Michael L et al., User engagement from within the Twitter community of professional sport organizations. *Managing Sport and Leisure*, pages 275–293, June 2019.
- [9] Darbinyan et al., How AI Transforms Social Media. *Forbes*, March 2023.
- [10] Belsky et al., How Generative AI Is Changing Creative Work. *Forbes*, May 2023.
- [11] Kapoor et al., Advances in Social Media Research: Past, Present and Future. *Inf. Syst. Front.*, 20(3):531–558, June 2018.
- [12] How Different Generations Use Social Media—and What this Means for Your Business, July 2021. [Online; accessed 28. Sep. 2023].
- [13] Wang et al., PAC-Net: Highlight Your Video via History Preference Modeling, 2022. [Online; accessed 28. Sep. 2023].
- [14] Hou et al., Deep Learning for Logo Detection: A Survey. *ACM Trans. Multimedia Comput. Com-mun. Appl.*, 2022.
- [15] Sosnovik et al., Learning to Summarize Videos by Contrasting Clips. *ArXiv*, 2023.
- [16] Xiong et al., Dual-Stream Multimodal Learning for Topic-Adaptive Video Highlight Detection. *Proceedings of the 2023 ACM International Conference on Multimedia Retrieval*, 2023.
- [17] Gautam et al., Assisting soccer game summarization via audio intensity analysis of game high-lights. In *Proceedings of 12th IOE Graduate Conference*, volume 12, pages 25 – 32. Institute ofEngineering, Tribhuvan University, Nepal, October 2022.
- [18] Gautam et al., AI-based soccer game summarization: From video highlights to dynamic text summaries. Master's Thesis, Tribhuvan University, 2022.
- [19] Gautam et al., Soccer game summarization using audio commentary, metadata, and captions. In *Proceedings of the 1st Workshop on User-centric Narrative Summarization of Long Videos*, pages 13–22, 2022.
- [20] Apostolidis et al., A fast smart-cropping method and dataset for video retargeting. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 2618–2622. IEEE, 2021.
- [21] Jin et al., Object-ratio-preserving video retargeting framework based on segmentation and inpainting. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, pages 497–503, January 2023.
- [22] Mallmann et al., Ppcensor: Architecture for real-time pornography detection in video streaming. *Future Generation Computer Systems*, 112:945–955, 2020.
- [23] Rongved et al., "Real-Time Detection of Events in Soccer Videos using 3D Convolutional Neural Networks," *IEEE International Symposium on Multimedia (ISM)*, 2020, pp. 135-144, doi: 10.1109/ISM.2020.00030.
- [24] Rongved et al., "Using 3D Convolutional Neural Networks for Real-time Detection of Soccer Events," *International Journal of Semantic Computing* 15.02 (2021): 161-187.
- [25] Giancola et al., "SoccerNet: A Scalable Dataset for Action Spotting in Soccer Videos," *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 1792-179210, doi: 10.1109/CVPRW.2018.00223.
- [26] Valand et al., "Automated Clipping of Soccer Events using Machine Learning," *IEEE International Symposium on Multimedia (ISM)*, 2021.
- [27] Jiang et al., A Review of Yolo algorithm developments. *Procedia Computer Science*, 199, 1066-1073.
- [28] Husa, et al. Automatic Thumbnail Selection for Soccer Videos using Machine Learning. *MMSys 2022*. DOI: 10.1145/3524273.3528182.
- [29] Husa et al., HOST-ATS: Automatic Thumbnail Selection with Dashboard-Controlled ML Pipeline and Dynamic User Survey. *MMSys 2022*. DOI: 10.1145/3524273.3532908.
- [30] Liu et al., "Visual instruction tuning", *arXiv preprint: 2304.08485* (2023)
- [31] Maaz et al., "Video-Chat GPT: Towards Detailed Video Understanding via Large Vision and Language Models", *arXiv preprint: 2306.05424* (2023)
- [32] Zhang et al., "Video-LLaMa: An instruction- tuned audio-visual language model for video understanding", *arXiv preprint: 2306.02858* (2023)
- [33] Wu et al., "NExT-GPT: Any-to-Any Multimodal LLM.", *arXiv preprint: 2309.05519* (2023)
- [34] Yu et al., Generating Hashtags for Short-form Videos with Guided Signals. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 9482-9495). July 2023.
- [35] Sundaram et al. (2014). Building Applications with Social Networking API's. *International Journal of High Performance Computing and Networking*. 5. 2070-2075.