University of Economics in Katowice, Poland
Faculty of Informatics and Communication
Department of Machine Learning

ML in PL CONFERENCE 2024

Improved Scene Classification in Dynamic Combat Sports by Video Frame Segmentation

Piotr Stefański
Department of Machine Learning,
University of Economics in Katowice, 1 Maja 50, 40-287 Katowice, Poland

7-10 November 2024

Abstract

This study addresses the challenge of classifying video frames in sports analytics, specifically focusing on identifying boxing punches with minimal preprocessing. The core issue investigated is the low region of interest (ROI) to image ratio prevalent in dynamic sports footage, where critical actions like punches occupy less than 1.5% of the frame area. We conducted a comprehensive review of existing literature and introduced a novel video frame segmentation technique tailored for high-speed sports analytics. Our proposed method significantly improves classification performance while maintaining high processing speeds, crucial for real-time applications. Experimental results demonstrate the superiority of our approach over state-of-the-art methods in terms of computational efficiency. This study contributes to the fields of image processing and sports analytics by providing a scalable solution that improves the real-time classification of fast sports actions, potentially benefiting training and competitive strategies through immediate feedback.

Method

To address the problem of a low ROI to image ratio, which can adversely affect classification performance, three segmentation approaches from the literature were implemented to preprocess video frames before classification. These approaches are aimed at extracting the most informative parts of a frame, thereby improving the classifier's ability to accurately differentiate between "punch" and "not punch" frames. In addition to methods from the literature, a novel approach was also proposed. Therefore the following approaches were evaluated during the experiments:

- **original** - original frames without any transformations before classification (baseline);
- **back_n_frames** - frames with subtracted background by the proposed Algorithm 1 where n is the second parameter;
- **extract_by_knn** - frames with subtracted background by the algorithm which is based on K-nearest neighbours;
- **extract_by_mog2** - frames with subtracted background by the algorithm which is based on the Gaussian mixture.
- **BSUV-Net** - frames with subtracted background by the BSUV-Net algorithm which is based on the convolution neural networks.

Input: $frame$ – one frame from the video
Input: $compare\_with\_n\_back\_frame$ – a number specifying with how many earlier frames the algorithm should subtract the input frame
Output: $segmented\_image$ – video frame after segmentation

$gray = convert\_image\_rgb\_to\_grayscale(frame)$
$previous\_n\_frames.insert\_at\_first\_position(gray)$
if $length(previous\_n\_frames) <= compare\_with\_n\_back\_frame$ **then**
  $segmented\_image = frame$
else
  $diff = gray - previous\_n\_frames[compare\_with\_n\_back\_frame]$;
  $mask = normalize\_diff(diff)$
  $segmented\_image = apply\_mask\_on\_image(frame, mask)$
  $previous\_n\_frames.remove\_last\_element()$
end
result $segmented\_image$;

Algorithm 1: Proposed algorithm for video frame segmentation

Results

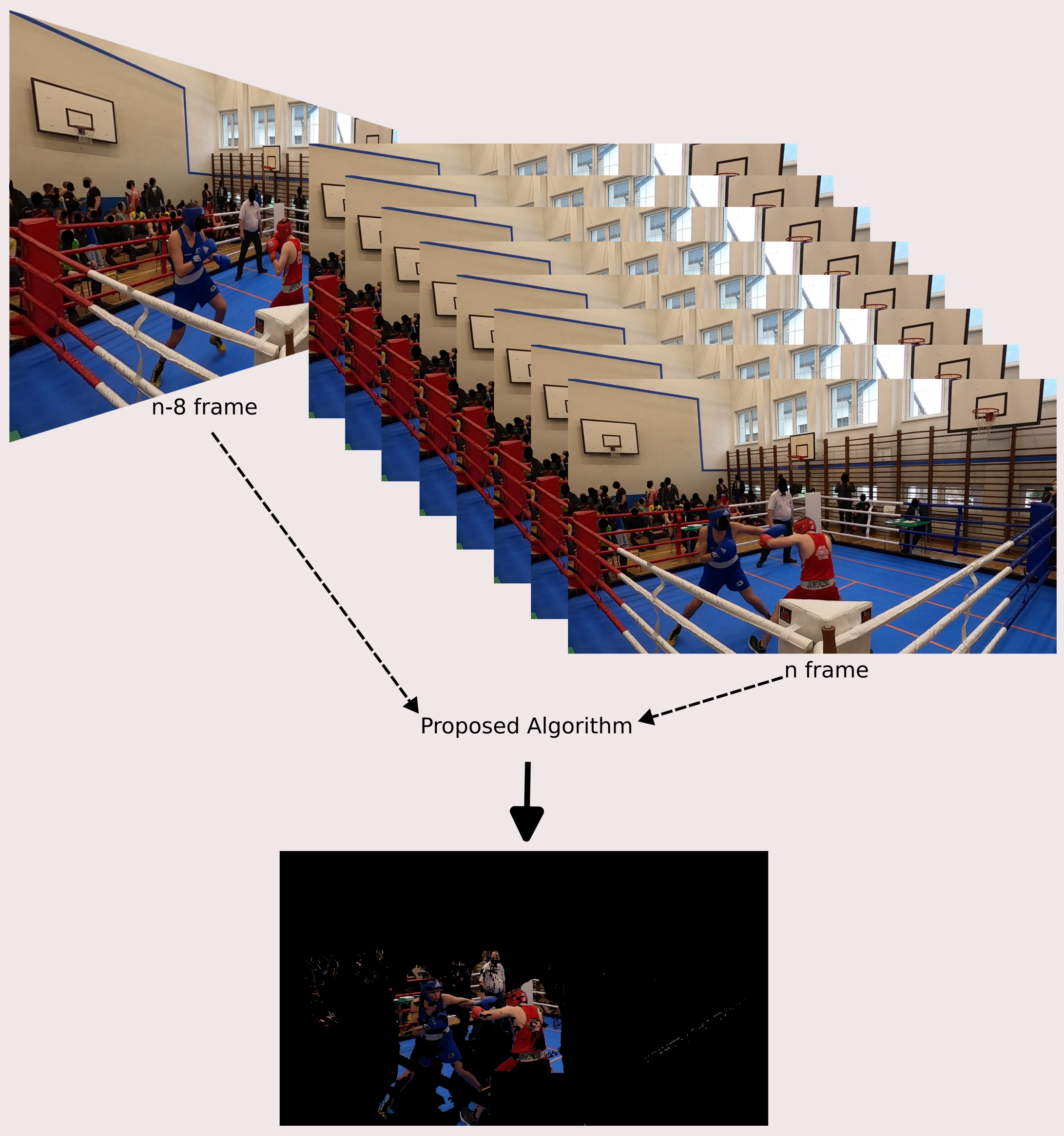

Figure: Visualization of the proposed algorithm for n = 8

Table: Medians of performance metrics for tested approaches

Approach name	Accuracy	Balanced accuracy	F1 punch	F1 not_punch
original	0.846	0.500	0.511	0.924
back_13_frames	0.893	0.839	0.583	0.939
extract_by_knn	0.895	**0.882**	**0.625**	0.939
extract_by_mog2	**0.900**	0.873	0.622	**0.942**
BSUV-Net	0.899	0.864	0.611	**0.942**

Table: Average processing time and resource utilization for the tested approaches

Approach name	Time (s)	CPU avg (%)	Mem avg (%)
back_13_frames	**88.629**	**28.147**	69.708
extract_by_knn	184.132	71.190	73.904
extract_by_mog2	273.155	82.387	63.760
BSUV-Net	40731.288	48.336	**19.089**

Conclusions

The study conclusively demonstrates that the proposed video frame segmentation techniques offer substantial advantages for high-speed sports analytics, particularly in boxing. The methods stand out for their enhanced accuracy and processing speed, crucial for real-time applications. Unlike existing approaches, the novel segmentation techniques effectively handle the low region of interest (ROI) to image ratios typical in dynamic sports scenes, ensuring that critical moments such as punches are captured and analyzed with high precision. These techniques also optimize computational efficiency, making them well-suited for scenarios with limited processing resources.