

## **Project**

# **BoxingX: A Multi-Modal Benchmark for Boxing Action Recognition and Localization**

## **Abstract**

Accurate analysis of combat sports using computer vision has gained traction in recent years, yet the development of robust datasets remains a major bottleneck due to the dynamic, unstructured nature of actions and variations in recording environments. In this work, we present a comprehensive, well-annotated video dataset tailored for punch detection and classification in boxing. The dataset comprises 6,915 high-quality punch clips categorized into six distinct punch types, extracted from 20 publicly available YouTube sparring sessions and involving 18 different athletes. Each clip is manually segmented and labeled to ensure precise temporal boundaries and class consistency, capturing a wide range of motion styles, camera angles, and athlete physiques. This dataset is specifically curated to support research in real-time vision-based action recognition, especially in low-resource and unconstrained environments. By providing a rich benchmark with diverse punch examples, this contribution aims to accelerate progress in movement analysis, automated coaching, and performance assessment within boxing and related domains.