Detection and Re-Identification in the case of Horse Racing

Will Binning
https://www.linkedin.com/in/will-binning-66a9b7221/
Sadegh Rahmani
https://www.linkedin.com/in/sadegh-rahmani/
Xu Dong
www.linkedin.com/in/xudong-442302166
Andrew Gilbert
https://andrewjohngilbert.github.io/

C-CATS,
Centre of Creative Arts and Technology
University of Surrey
UK

Despite its popularity and the substantial betting it attracts, horse racing has seen limited research in machine learning. Some studies have tackled related challenges, such as adapting multi-object tracking to the unique geometry of horse tracks [3] and tracking jockey caps during complex manoeuvres [2]. Our research aims to create a helmet detector framework as a preliminary step for re-identification using a limited dataset. Specifically, we detected jockeys' helmets throughout a 205-second race with six disjointed outdoor cameras, addressing challenges like occlusion and varying illumination. Occlusion is a significant challenge in horse racing, often more pronounced than in other sports. Jockeys race in close groups, causing substantial overlap between jockeys and horses in the camera's view, complicating detection and segmentation. Additionally, motion blur, especially in the race's final stretch, and the multi-camera broadcast capturing various angles—front, back, and sides—further complicate detection and consecutively re-identification (Re-ID). To address these issues, we focus on helmet identification rather than detecting all horses or jockeys. We believe helmets, with their simple shapes and consistent appearance even when rotated, offer a more reliable target for detection to make the Re-ID downstream task more achievable.

The Architecture: A summary of the architecture is presented in Figure 1. The system focuses on helmet detection and classification. A multi-class Convolutional Neural Network (ConvNet) based on ResNet-18 is trained on labelled helmet classes using a cross-entropy loss function. For helmet detection and segmentation, we employ Grounded-SAM [4] with the prompt "helmet" to accurately detect and segment jockeys' helmets, even capturing partially visible helmets in cases of mild occlusion. These segmented helmets are then cropped from the images to create our dataset. Based on the ResNet-18 architecture, the classification model processes these cropped helmet images and is trained using cross-entropy loss and the Adam optimizer. After training, we apply a confidence threshold to filter out false positives. Finally, we use the detected helmets to annotate the original images with colour-coded bounding boxes corresponding to each class.

Dataset: The racing broadcast company Racetech [1] provided the data and industry context that enabled the research presented in this paper. The dataset comprises a single outdoor competitive horse race, captured by six moving cameras, with a total duration of 205 seconds and 1026 frames sampled at 5 FPS. The race features 12 jockeys, and we developed a proof-of-concept labelled dataset focusing on 5 jockeys (or classes) across the six cameras. For semi-automated ground truth labelling, the cropped helmet images were sorted by primary colour, followed by a manual review to remove misclassifications. The model was trained on 80% of the samples from camera 1 and tested on the remaining samples from camera 1 and unseen data from cameras 2-6.

Results: The accuracy of the helmet detector across the 6 cameras is illustrated in Figure 2, highlighting the model's overall performance. Helmets with simple, solid-coloured designs, such as classes 1 and 2, achieved higher accuracy. However, helmets with intricate designs, like those in classes 4 and 5, faced challenges, particularly in wide shots where lower resolution impacted detection. Additionally, cameras positioned around turns (such as cameras 4 and 5) were less successful due to issues with occlusion and blurring, which adversely affected both ResNet feature extraction and the initial helmet detection. However, there is a good performance in the classification of helmets despite only using examples from a single camera view to train the model. The camera view 1 helmet examples used to train this model ranged from 70 to 160 images per class, which, in conjunction with the single-angle training set and relatively simple model, shows that there is a prospect of success in this technique. For instance, the most successful class (class 1) had a mean of 79.1% successful detection and classification during the race. Overall, we see this technology as helpful in augmenting the broadcast footage and possibly retrieving specific jockey footage of past races. The near 80% is an excellent initial figure, but we'd expect this to be needed to have far few false negatives and positives for broadcast.

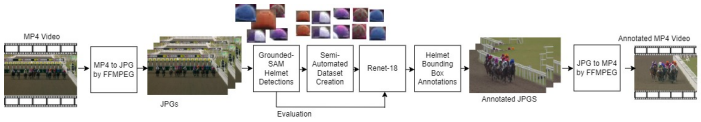

Figure 1: The model segments helmets from RGB images and utilizes a multi-class classifier to train the helmet detector. The final output is an annotated MP4 video, where helmet bounding boxes are colour-coded to represent the five distinct classes.

Figure 2: Confusion matrix for individual classes on each angle. Values are percentages of total frames that the class has been successfully annotated. After running through the model, these values were manually verified to remove false positives but do not account for frames where the class is not present.

Conclusions: We initially selected five jockeys with distinct helmets to simplify dataset creation. However, expanding the dataset revealed challenges in manually classifying similar helmets. Future research will combine distinct helmet detections with jockey or horse tracking to improve classification, acknowledging that not all helmets in a race will be unique. Further exploration of the data requirements for effective helmet detection is also recommended.

[1] Racetech. `https://www.racetech.co.uk/`. Accessed: 2024-08-27.

[2] Mohammad Hedayati, Michael J Cree, and Jonathan B Scott. Tracking jockeys in a cluttered environment with group dynamics. In *Proceedings Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports*, pages 67–73, 2019.

[3] Wing W.Y. Ng, Xuyu Liu, Xuli Yan, Xing Tian, Cankun Zhong, and Sam Kwong. Multi-object tracking for horse racing. *Information Sciences*, 638:118967, 2023.

[4] Tianhe Ren, Shilong Liu, Ailing Zeng, Jing Lin, Kunchang Li, He Cao, Jiayu Chen, Xinyu Huang, Yukang Chen, Feng Yan, et al. Grounded sam: Assembling open-world models for diverse visual tasks. *arXiv preprint arXiv:2401.14159*, 2024.