

CattleAct: Cattle Interaction Detection with Joint Learning of Action-Interaction Latent Space

Supplementary Material

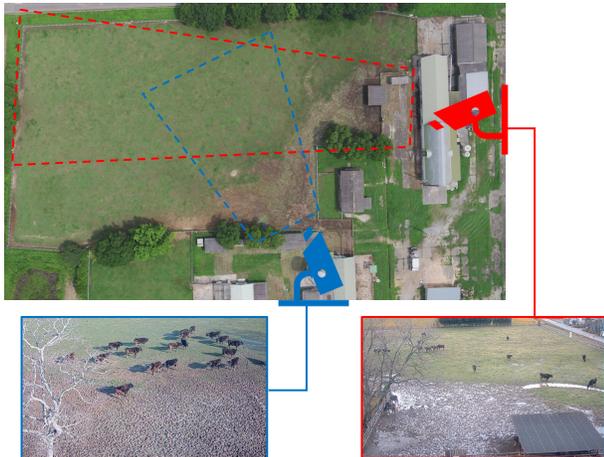


Figure S1. Data collection environment and camera configurations. Top: Aerial view of the pasture showing the overlapping fields of view of the two surveillance cameras. The red and blue dashed lines indicate the coverage areas of each camera. Bottom: Sample frames from each viewpoint, illustrating diversity in object scale and occlusion.

A. Dataset Details

This section describes the data collection environment and the definition of action labels.

A.1. Environmental Settings and Data Split

Our dataset was collected in a real-world pasture. As shown in Fig. S1, a multi-camera setup was employed to enable the learning of behavioral features independent of specific camera angles.

Camera Angles and GPS Synchronization A key challenge is aligning GPS location data with visual data. The training set utilizes video sequences from multiple angles to facilitate view-invariant feature learning. The test set, however, is restricted to a single view (the camera indicated in red in the figure). This view enables the most precise matching between GPS coordinates and bounding boxes, ensuring the performance required for practical application.

A.2. Action Label Definitions

As shown in Fig. S2, action labels are classified into two groups.

- Individual Action Labels: Physical state of the cattle. *Lying* (resting), *Grazing* (eating), *Standing* (idle or moving),

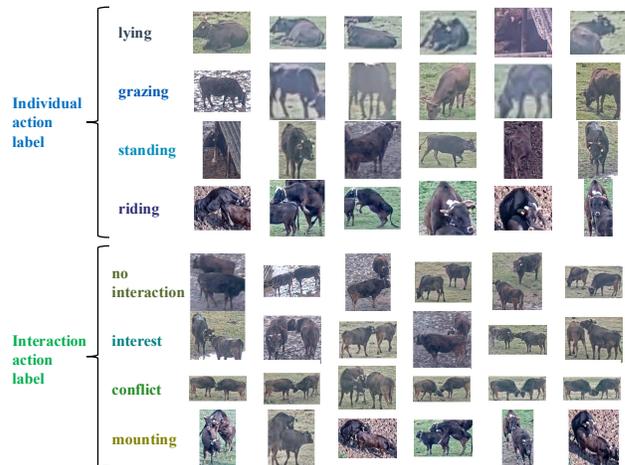


Figure S2. sample of cattle behavior labels. Top: Individual action labels. States of single subjects (e.g., *lying*, *grazing*, *standing*, *riding*). Bottom: Interaction action labels. Relationships between multiple subjects (e.g., *no interaction*, *interest*, *conflict*, *mounting*). High-level interactions often co-occur with specific individual actions.

and *Riding* (individual motion associated with mounting).

- Interaction Action Labels: Social relationships between cattle. *No Interaction*, *Interest*, *Conflict*, and *Mounting*.

These annotations aim to enable the understanding of herd dynamics beyond simple object detection.