

QuickNotes

Science Summaries from fRI Research

Can Deep Learning Find Moose and Deer in Thermal Drone Footage?

Hiroto Goto, Nicholas C. Coops, Gordon Stenhouse, Kwang Moo Yi, Leonard Hambrecht, Kirsten Solmundson

Uncrewed aerial vehicles (UAVs), or drones, equipped with thermal sensors offer a promising approach for surveying ungulates like moose in boreal environments. Thermal sensors detect the heat emitted by animals, revealing camouflaged animals and those partially hidden beneath tree canopy. But the vast amount of footage that drone surveys generate is a significant bottleneck, as manual review can be time-consuming, laborious, and error-prone. Deep learning, a form of artificial intelligence, can automate the detection of animals in this footage. However, detecting and distinguishing visually similar species, such as moose and deer, in low-resolution thermal imagery remains challenging and largely unexplored.



Example visible (RGB) and thermal imagery of a moose. While the RGB image (left) shows limited visibility, thermal image (right) reveals the moose beneath tree canopy.

We tested how well two modern deep learning models could detect and classify moose and deer in thermal drone footage, across multiple independent survey dates.

Key Findings

- **Overall performance:** The best performing model correctly identified 75% of its detections and found 62% of all moose and deer.
- **Moose vs. deer:** Moose were detected more reliably than deer. Deer were frequently missed and undercounted, and were confused with visually similar background objects like logs and fences.
- **Challenges:** Performance varied between surveys, reflecting the difficulty of applying models to new survey conditions and working with coarse, low-detail thermal imagery from high-altitude drone flights.
- **Operational significance:** Human verification is still necessary, but the model can be used as an initial filtering tool to assist manual review of drone footage.



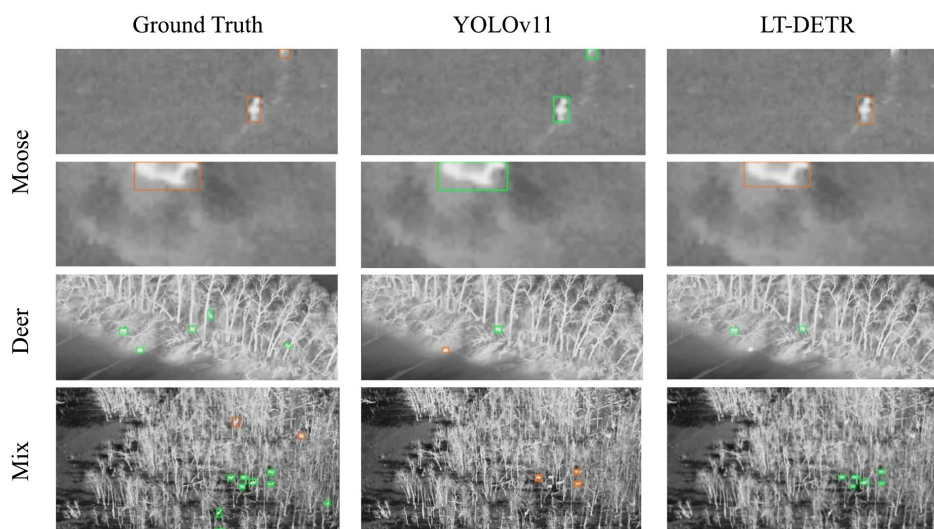
fRI Research

For more information about this research, contact:
hiroto.goto@ubc.ca

Methods

Thermal footage was collected by Superwake using a fixed-wing drone during aerial surveys for Manitoba's Chronic Wasting Disease monitoring program in February and March 2024. We extracted still frames containing moose or deer and drew a box around each animal, and wildlife specialists confirmed the species using the matching colour video. The final dataset included 4,787 images with 2,377 moose (14 individuals) and 6,866 deer (white-tailed and mule deer combined; 126 individuals). The dataset was characterized by large scale variation caused by camera zoom.

We tested two families of object detection models, which both locate and identify animals in an image by drawing a box around each one: the convolution-based YOLOv11 and the transformer-based LT-DETR. Since models often perform worse on unseen surveys, we tested both models on survey dates that were not seen during training.



Examples of moose (orange boxes) and deer (green boxes) in thermal drone imagery. Left: the true locations (ground truth); middle: YOLOv11 predictions; right: LT-DETR predictions. LT-DETR detected animals that YOLOv11 missed or misclassified, but both models struggled with small animals in cluttered backgrounds.

Implications

With a precision of 0.75, only roughly one in four detections would be a false alarm requiring manual correction, so the model can serve as a useful filtering tool for flagging potential moose and deer. However, with a recall of 0.62, a substantial proportion of individuals are missed (particularly deer), resulting in systematic undercounting. Combined with common confusion between moose and deer, the model is not suitable as a fully autonomous tool. Human verification is still required, especially when accurate species-level data are needed.

Therefore, these results mainly reflect the model's value as an assisted-review tool: screening flagged video, discarding empty frames, and generating preliminary annotations for new datasets. This can reduce manual processing effort, allowing resources to be reallocated toward expanding survey coverage and frequency. Future work should focus on expanding the dataset, adding species such as elk and bison, and applying tracking algorithms on video footage.

More Information

Read the full paper here: <https://doi.org/10.1016/j.ecoinf.2026.104068>

A subset of the dataset is available here: <https://doi.org/10.5281/zenodo.20018687>

This research was supported by the Forest Resource Improvement Association of Alberta (FRIAA).

