Pet Tracking and Activity Visualization Using

AI-Powered Heatmaps

CPSC 4420 - Final Project Fall 24

Trisha Andres, trishaa@g.clemson.edu

Kate Fullero, ktfulle@g.clemson.edu

Problem Statement

The Challenge:

- Understanding animal behavior for welfare assessment, veterinary diagnostics, and environmental monitoring
- Traditional manual observation methods are labor-intensive and error-prone

Specific Issues Addressed:

- Visualizing pet activity patterns in real time
- Challenges in detecting and tracking multiple animals in dynamic or edited video content (YouTube)

CLEMSON
College of ENGINEERING, COMPUTING AND APPLIED SCIENCES

Motivation

Why This Matters:

- Provides insights into pet behavior for owners, veterinarians, and researchers
- Offers an intuitive, automates alternative to manual tracking

Potential Impact:

- Enables smart home surveillance
- Facilitates behavior monitoring in veterinary clinics
- Assists pet owners in detecting abnormal patterns or identifying favorite spots

Proposed Solution

System Components:

- Object Detection: YOLOv5 model for detecting and tracking pets in videos
- Activity Visualization: Heatmaps generated using Gaussian-based methods

Key Features:

- processes videos frame-by-frame, overlays bounding boxes, and creates heatmaps
- Optimized for real-time performance at 30 FPS

Methodology

Data Collection:

- Five videos in total with dogs, cats, and cattle
  - 3 Videos downloaded from Youtube
    - Compilation of a singular dog walking
    - Compilation of numerous cat videos with different amounts of cats
    - Video of numerous cattle roaming
  - 2 Personal Videos of Trisha’s family dog Izzy
- Varied Scenarios: uninterrupted personal recordings vs. edited YouTube videos

Detection and Tracking:

- YOLOv5 with a confidence threshold of 0.1
- Challenges: Multi-animal tracking and occlusions

$$H(x,y) = \sum_{i=1}^{N} exp\left(-\frac{(x-x_i)^2+(y-y_i)^2}{2\sigma^2}\right)$$

Heatmap Generation:

- Gaussian contributions aggregated for activity visualization
- Warm colors in heatmaps represent areas of high activity density

CLEMSON College of ENGINEERING, COMPUTING AND APPLIED SCIENCES

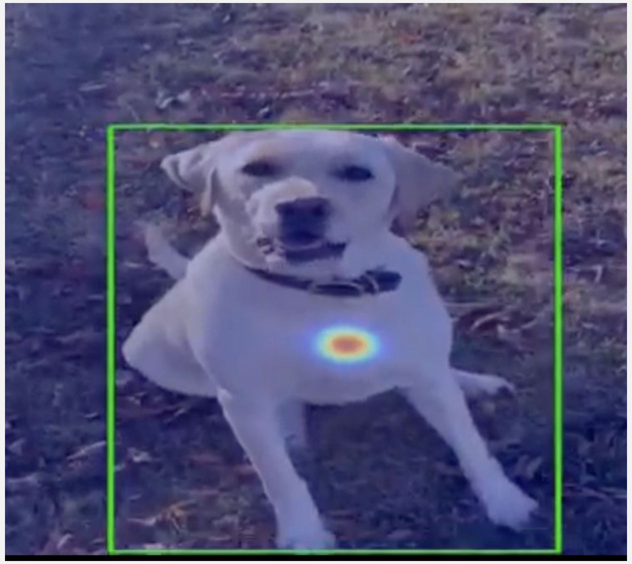
Quantitative Analysis

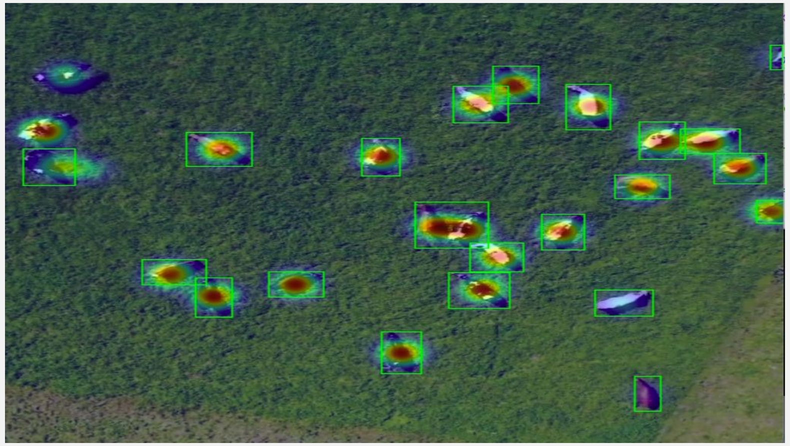
Performance Metrics:

- Speed:
  - 30 FPS on a mid-range GPU
- Accuracy:
  - Personal videos: ~90%
  - YouTube videos: ~75% (due to edits & scene transitions)
- Outcome:
  - Real-time tracking and intuitive heatmap visualization

Qualitative Analysis

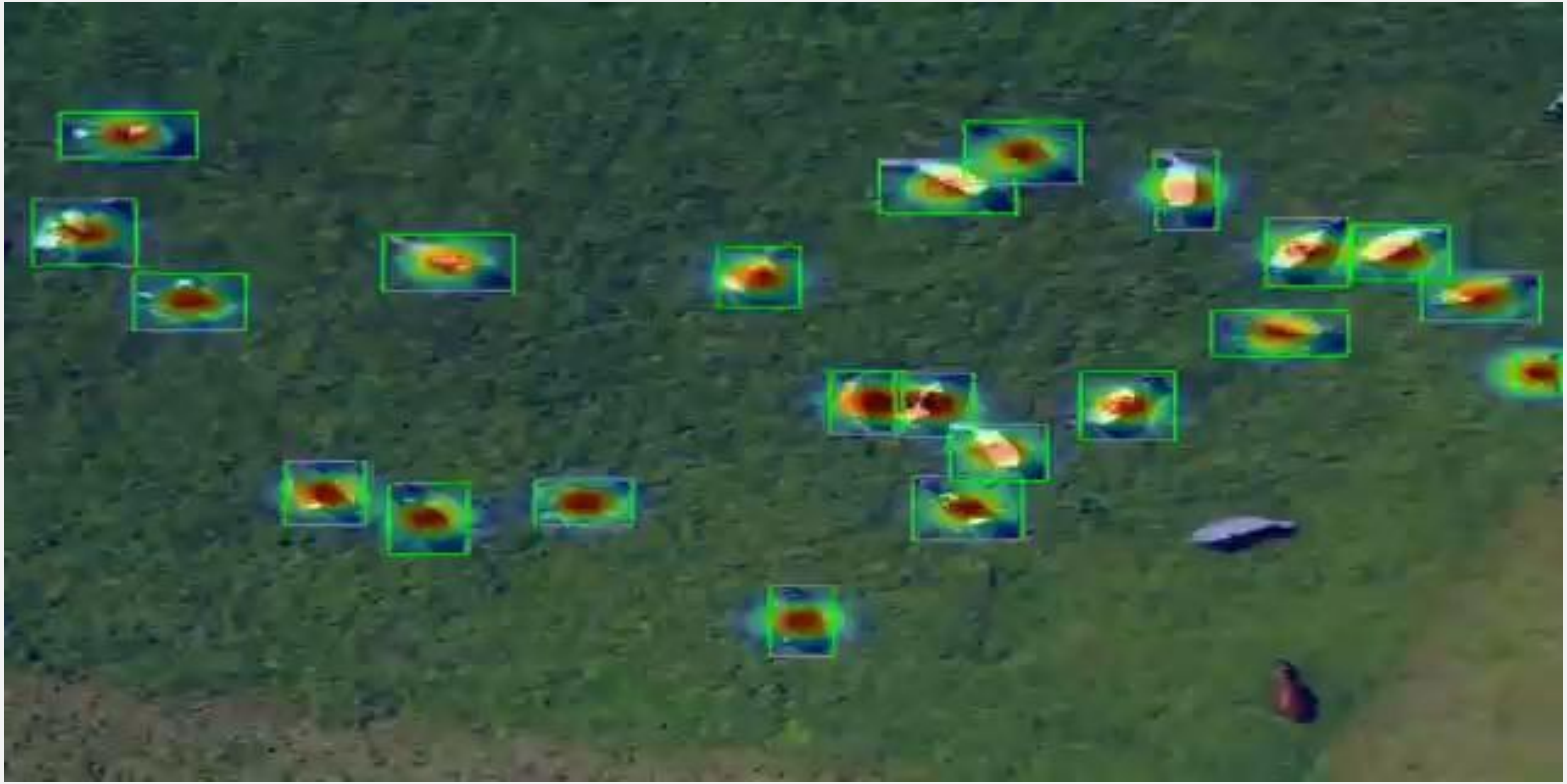
Screenshots from four of the videos tested showing the heatmap and bounded box around the animal(s).



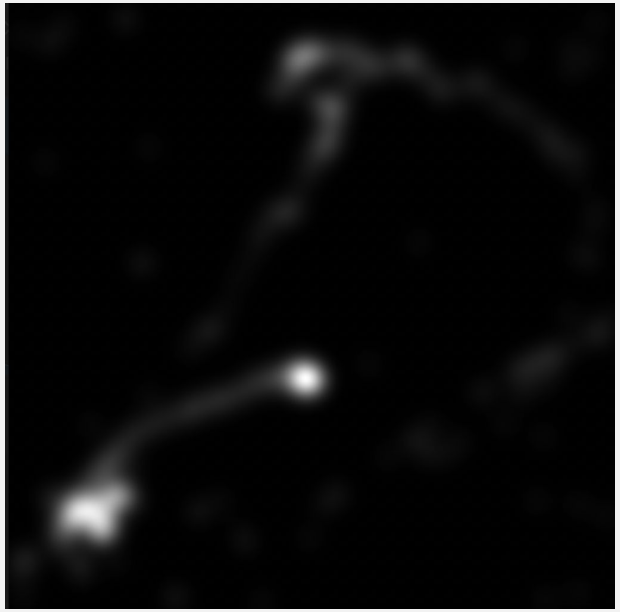


CLEMSON
College of ENGINEERING, COMPUTING AND APPLIED SCIENCES

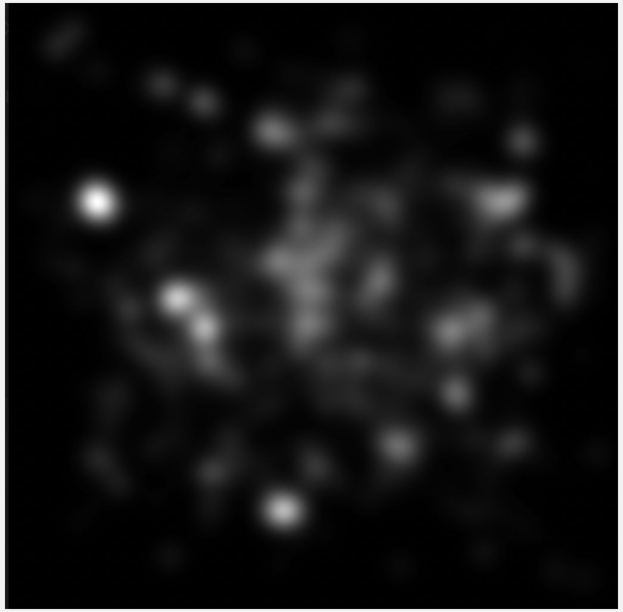
CLEMSON
College of ENGINEERING,
COMPUTING AND APPLIED SCIENCES



Qualitative Analysis



Final Heatmap from personal video of a singular dog.



Final Heatmap from YouTube video of numerous clips of cats.

Final Heatmap from Youtube video of cattle in a field.

CLEMSON
College of ENGINEERING, COMPUTING AND APPLIED SCIENCES

Discussion

Comparison to Related Work:

- Novelty: Real-time performance using consumer-grade video inputs
- Unique Contribution: Activity visualization through heatmaps

Limitations:

- Multi-animal tracking difficulties in compilation videos with abrupt edits and clipping
- Edited video content disrupts tracking continuity

Potential Improvements:

- Incorporate advanced algorithms (e.g., DeepSORT or ByteTrack)
- Expand dataset diversity

Results

Key Takeaways:

- Hypothesis validates: Accurate activity visualization (>85% accuracy in controlled settings)
- System demonstrates potential for real-time pet behavior analysis

Future Directions:

- Improve tracking robustness for multi-animal scenarios in edited videos
- Optimize for edge device deployment
- Enhance dataset variety for better generalization

References

[1] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.

[2] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934.

[3] Girshick, R. (2015). Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 1440-1448.

[4] Papageorgiou, C., & Poggio, T. (2000). A Trainable System for Object Detection. International Journal of Computer Vision, 38(1), 15-33.

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.

[6] Zhang, Z., et al. (2019). DeepSORT: Deep Learning to Track Multiple Targets in Videos. International Conference on Image Processing (ICIP).

[7] Seitz, S. M., & Dyer, C. R. (1999). Photorealistic Scene Reconstruction by Voxel Coloring. International Journal of Computer Vision, 35(2), 151-173.