Teaching Scalable Wildlife Image Processing with NAIRR Jetstream2 GPUs to Undergraduates

Katherine Nunn[1], Bernie Boscoe[1], Shawn Johnson[2] , Karen Mager[2] , Chandler Campbell[1]
Southern Oregon University Computer Science[1], Southern Oregon University Environmental Science[2]

Motivation & Goals

Undergraduate data-science courses rarely give students access to **research-grade, GPU-accelerated image-processing pipelines**

This module shows how NAIRR Jetstream 2 GPUs can be turned into a **hands-on teaching platform** for wildlife-conservation data, lowering the barrier between classroom learning and real-world ecological research

We use camera trap videos and still images of wildlife and 1) run them through a YOLO-based computer vision tool, MegaDetector, 2) put bounding boxes on them, and 3) transfer them to the Environmental Scientists for further processing and analysis

Target Audience & Prerequisites

For basic course module, **undergraduate students** who have completed an introductory Python course.

No prior experience with high-performance computing, cloud platforms, or deep-learning models is required—everything is provisioned and demonstrated in the notebook.

Knowledge useful to have: Jupyter notebook, JupyterHub, file paths to connect data, CUDA to check GPUs, ways to upload large datasets to server

GitHub Repo:

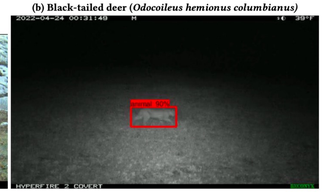
Dataset- Camera Trap Images

30,000 raw images from a multi-year camera-trap network within the **Cascade-Siskiyou mountains**
Contains a variety of mammals and birds with challenging detection conditions (weather, forest)
Dataset is stored on a **persistent Jetstream 2 volume** so students can explore the full collection without worrying about local storage limits

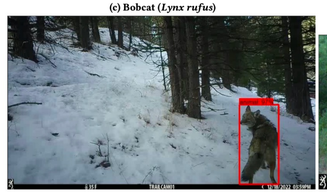
(a) Black bear (*Ursus americanus*)

(b) Black-tailed deer (*Odocoileus hemionus columbianus*)

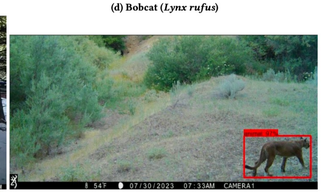
(c) Bobcat (*Lynx rufus*)



(d) Bobcat (*Lynx rufus*)



(e) Coyote (*Canis latrans*)



(f) Cougar (*Puma concolor*)

NAIRR + Jetstream2 Infrastructure

NAIRR classroom-pilot allocation | **JupyterHub** launch from browser| **Jetstream 2 VM with GPU** | **m3.medium** (2 vCPU, 8 GB RAM, 1 × NVIDIA A100) **Jupyter Notebook**| Provides **GPU-accelerated inference** for MegaDetector. | **Verification scripts** Simple Python checks (**torch.cuda.is_available()**, benchmark timers)

References

B. Boscoe, S. Johnson, A. Osbon, C. Campbell, and K. Mager, "GreenCrossingAI: A Camera Trap/Computer Vision Pipeline for Environmental Science Research Groups," in Practice and Experience in Advanced Research Computing (PEARC) 2025: The Power of Collaboration, Columbus, OH, July 2025. doi: 10.1145/3708035.

Sara Beery, Dan Morris, and Siyu Yang. 2019. Efficient Pipeline for Camera Trap Image Review. arXiv:1907.06772

https://github.com/agentmorris/MegaDetector

Workflow for Student

-Login to JupyterHub, open notebook, obtain correct path to dataset
-Ensure mounted volume is visible
-Check code, install dependencies if needed
-Run either Video processing or Image processing notebook
-Note benchmarks and time
-Check outputs
-Download results for further analysis

Pedagogical Highlights

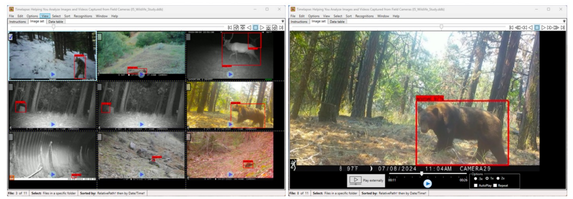
Scalable computing concepts – data transfer, persistent storage, virtual-machine lifecycle
Performance engineering – measuring latency, throughput, and GPU utilization
Reproducibility – notebook version-controlled on GitHub
Real-world relevance – students work with authentic conservation data rather than toy examples
Collaborations with Domain Scientists (in Environmental Science) CS students contribute to the project, with the larger aim to build a land bridge across the I-5 in Southern Oregon. The pre-animal bridge data will help to understand what animals are in the area, and how they currently cross the freeway

Expected Learning Outcomes

- Students access the National Artificial Intelligence Research Resource (NAIRR) computing ecosystem and learn to provision Jetstream2 virtual machines with GPU capability
- Students verify GPU availability and performance within Python, practice timing and benchmarking GPU execution, and manage software dependencies
- Students work with mounting persistent storage volumes to Jetstream instances and uploading large datasets to connected volumes
- Students gain hands-on experience with research-grade cyberinfrastructure, including JupyterHub notebooks, GPU computing platforms, and cloud-based data processing
- Students develop troubleshooting skills for environment setup, and dependency management while working with real conservation data
- The module lowers barriers to working with NAIRR resources while giving students authentic exposure to conservation-oriented data science workflows

Future Work & Extensions

- **Refactor notebook into reusable functions/modules** – extract repeated steps (data mounting, inference, benchmarking) into well-named functions
- **Add Python classes to the notebook** e.g., a **DataManager** for dataset handling, a **Detector** wrapper for MegaDetector, and a **Benchmark** utility for performance logging
- **Model fine-tuning option** – let students fine-tune the last layers of MegaDetector on a small curated subset of the camera-trap data
- **Introduce a basic CUDA tutorial** – simple kernel & timing demo



Media about GreenCrossing Project

https://sou.edu/academics/sou-students-support-wildlife-project-on-i-5/

https://www.siskiyoudaily.com/story/news/2022/08/31/animal-crossings-northern-california-highways/7943113001/

https://www.vcstar.com/story/news/local/california/2024/10/08/freeway-crossings-for-wildlife-built-in-three-california-counties/75577735007/

https://www.oregonlive.com/environment/2024/12/oregon-gets-33m-to-construct-wildlife-crossing-over-i-5-in-southwest-oregon.html

Special Thanks

NAIRR Pilot — National Artificial Intelligence Research Resource Pilot

NAIRR Classroom Award #240359:
Machine Learning for CS Undergraduates