Wild Hogs and Big Data: How to deal with 7 million game camera images to answer questions on hog biology and management

University of Florida
Range Cattle Research and Education Center
Webinar 11 December, 2018
Presented by: Dr Raoul Boughton
rboughton@ufl.edu

Wild Hogs and Big Data
Introduction

5 year study (USDA): 2014-2019
1. Density and demography, reproduction and survival.
   • Uniquely marked individuals (captures)
   • Capture-Resight Spatial Models (game cameras)
2. Spatial ecology, movement, resource selection models and interactions with livestock and at point sources
   • GPS collars
   • UHF proximity collar studies among cattle, hogs and point sources
3. Population control RACI design
   • Does this change interactions, space use by hogs?
   • What effort is needed to remove hogs?
   • What is the time to population recovery?
4. Disease sampling and identification
   • Blood, feces, nasal, buccal, genital
   • Focus on ARMs and viruses (torque teno and circovirus to use in epidemiological models)
5. Aerial assessment of rooting damage
   • Drone flights and analyses of rooting damage
Tejon Ranch Conservancy
Buck Island Ranch

Study Location

Game Camera Deployment
- 44 Game Cameras, 1/km²
- Deployed August 2015
- Continuously running since ~3.5 years
- Maintained monthly
- ~100,000 images processed a month
Trailer build for hog study

Needed for reduced use of anesthetics, ease of handling, and access for taken samples from awake hogs.

Lots of help and hard work
Data Summary Information

- 772 Feral Hog Captures
  - 302 unique marks deployed (~100 per year)
  - 260 non-marked hogs removed during removals
  - 121 marked hogs removed during removals
  - 89 recaptures, resampled and released
- > 4 million images taken (in Florida)
  - ~ 276,975 of wild hogs
- 111 GPS units deployed (30 minute fixes)
  - 546,060 fixes obtained
  - Many early failures, short retention, water damage
Wild Hogs and Big Data Image Management

CPW Photo Warehouse

- Microsoft Access Application
- Archiving, summarizing, analyzing photos
- Double blind entry
- Third party verification
- Module creation and management for out of database identification (many uses at once)
- Tested and tested with many research projects
- Automated query functions to allow for data management
- Automated query functions to allow for data manipulation and format for analyses
  - Occupancy tables
  - Capture-recapture tables
- Excellent data management tool

CPW Photo Warehouse - Problems

- Too many photos for MS Access
  - Overburdened within 1-3 months of data collection
- Moved to SQL server platform requiring some changes in code. Issue solved
  - Issued on USDA server access by secure login only
- Unable to keep up with image classification
  - 100,000 images/month
  - 200,000 IDs needed/month (double blind)
  - Plus verification
    - ~ 80hrs/wk fulltime
    - ~ 1.5 million images identified and verified over 3.5 years. Not enough

Tried and tested with many research projects
- Automated query functions to allow for data manipulation and format for analyses
  - Occupancy tables
  - Capture-recapture tables
- Excellent data management tool
Stalled by ID process

- Currently 6.9 million images in database, continual growth at 100,000/month
- Only, 1.5 million identified 3 years in!!!

Machine Learning Process

- Supervised machine learning algorithms use training examples to "learn" how to complete a task.
- In our setting, we provided a set of animal images already identified (1.5 million) from camera traps of different species and their labels (species identifiers) to a deep neural network.
- We trained the model to identify species in training images.
- Once a model is trained, it learns how to classify new images that were not used for training.

Image classification neuron

- I1 to I3 are inputs in our case Red, Blue, Green
- In our species classification setting, the inputs to the network are normalized Red, Green, Blue (RGB) values of raw pixels of image
- We then interpret the output of the final layer as the probability of the presence of species in the image.
The math you should know but don’t have to

• $\theta = \text{ReLU}(w_1 I_1 + w_2 I_2 + w_3 I_3 + b)$ (eqn 1),
  Output of neuron calculated based on red, green, blue

• $L(P, Y) = - \sum_i P_i \log(Y_i)$ (eqn2)
  Loss calculated = predicted result compared to actual identification

• $w_i = w_{\text{initial}} - \frac{\partial L}{\partial w_i}$ (eqn 3).
  Weights adjusted to improve loss ($L$, eqn2) to best case scenario

---

Images used from across the USA

• We used 3,741,656 classified images to train and assess neural model
  California, Colorado, Florida, South Carolina, Texas and Canada
• 10% of each species retained for testing model (374,273 images)
• 27 Species or groups able to trained
Recall – ability of model to recognize species in training dataset for that species
Model Confidence Assignment

- Very high confidence achieved for most species
- But not every image is high
- Use confidence to assess what you will accept or need further verification

You don’t need to understand neural network mathematics and architecture to run the machine learning module. Go to github https://rdrr.io/github/mikeyEcology/MLWIC/README.md
Special Thanks

Collaborators:
- Dr Ryan Miller, USDA VS
- Dr Jesse Lewis, ASU
- Dr Samantha Wisely, UF WEC
- Dr JKC Leong, UF EFI
- Dr Michael Tabak, LSQA VS
- Dr Kim Pepin, USDA NWRC
- Dr Antoinette Plageo, USDA NWRC

Graduate Students:
- Wesley Anderson
- Tyler Buckley
- Connor Crank (MS)
- Ke Zhang (MS)
- Samantha Baraoidan
- Mary Merrill (PhD)
- Brittany Bankovich (MS)
- Philippe Hernandez (PhD)

Technicians:
- Bethany Wight
- Brandon Parker
- 30+ UF and CSU undergraduate student species identification experts

Wild Hogs and Big Data:
How to deal with 7 million game camera images to answer questions on hog biology and management

University of Florida
Range Cattle Research and Education Center
Webinar 11 December, 2018

Presented by: Dr Raoul Boughton
rboughton@ufl.edu