Deep learning to analyze avian displays: evaluating effectiveness across media types

Ashlynn Gonzalez and Emily H. DuVal
Florida State University, Department of Biological Sciences

FSU | FLORIDA STATE UNIVERSITY

FSU UNDERGRADUATE RESEARCH OPPORTUNITY PROGRAM
CENTER FOR UNDERGRADUATE RESEARCH & ACADEMIC ENGAGEMENT

INTRODUCTION

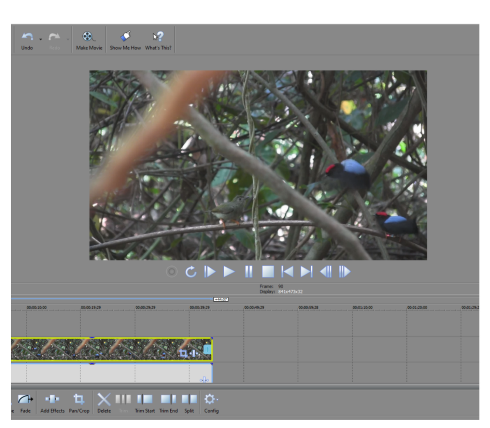
Video analysis helps quantify complex behaviors, such as the coordinated courtship displays of **lance-tailed manakins (*Chiroxiphia lanceolata*)**, which are central to sexual selection (DuVal, 2007). However, manually tracking movements across large datasets is impractical.

DeepLabCut (DLC), a machine learning tool for pose estimation, automates body part tracking, enabling rapid and objective video annotation (Mathis et al., 2018). **How does DLC perform across video types?**

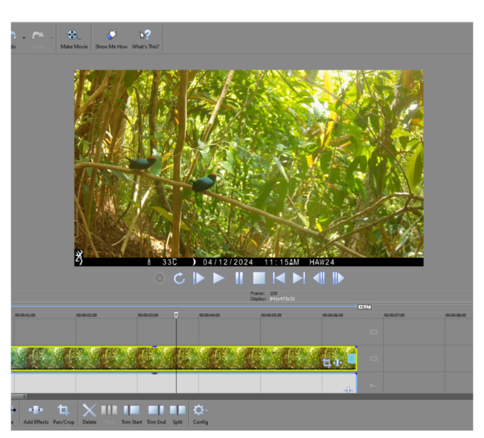
The DuVal Lab has recorded years of lance-tailed manakin displays using Webcam, Trailcam, and Handicam footage. This study aims to:

- Evaluate how camera type affects DLC tracking performance.
- Assess whether a model trained on one camera type can analyze footage from others or requires retraining.
- Provide recommendations for future video collection.

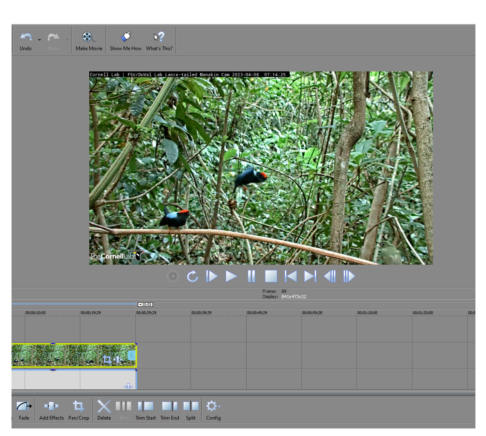
Understanding variation in analysis is essential for long-term data set reliability.



Handicam Video



Trailcam Video



Webcam Video

METHODS

Video Selection & Preprocessing:

- Extracted three clips per camera type (Webcam, Trailcam, Handicam) from longer recordings of lance-tailed manakins using Movie Studio Platinum.

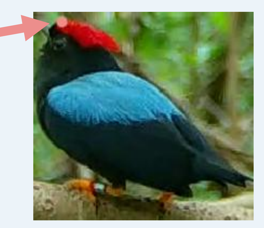
DeepLabCut Analysis:

- Processed videos in DeepLabCut (DLC) to track lance-tailed manakin body parts.
- Focused on, front crest, middleback, and tail base, landmarks that identify the key shapes of the bird.
- Analysis produces likelihood values that each body part for each bird was detected accurately by DLC.

Data Analysis:

- Likelihood values were analyzed using Kruskal-Wallis tests due to non-normal data distribution.
- Post-hoc Wilcoxon rank sum tests with Bonferroni correction were used for pairwise comparisons.

RESULTS



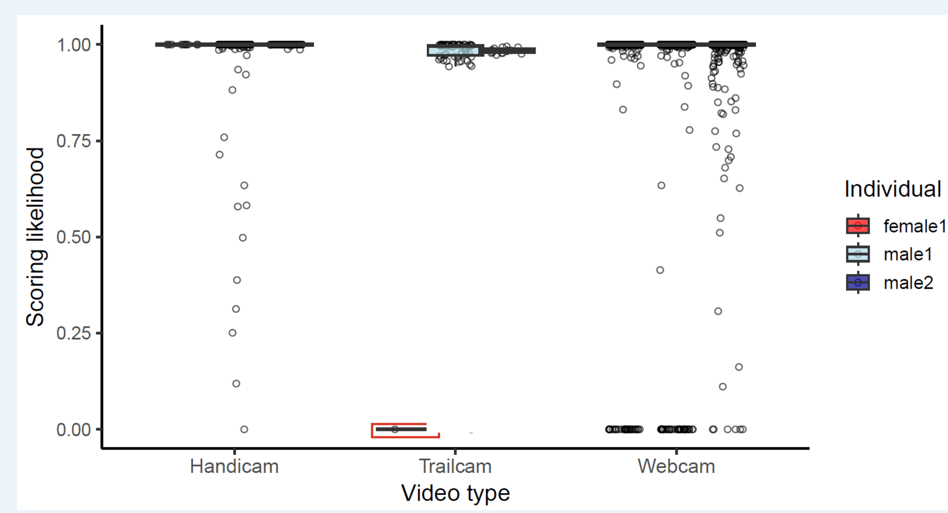


Figure 1: Front Crest Scoring Likelihood Across Video Types. Likelihood scores differed significantly across video types ($\chi^2 = 196.9$, $p < 2.2e\text{-}16$). Post-hoc Wilcoxon tests showed that Handicam and Webcam were not significantly different ($p = 0.544$), but both had significantly lower scores than Trailcam ($p < 2e\text{-}16$). N = 3945 Handicam, 263 Trailcam, and 2700 Webcam points identified by DLC.

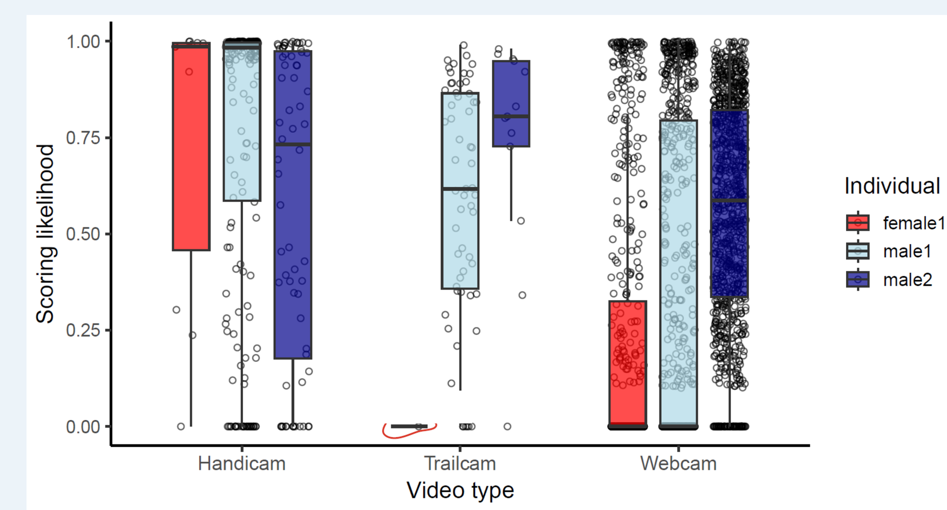


Figure 2: Tail Base Scoring Likelihood Across Video Types. Likelihood scores differed significantly across video types ($\chi^2 = 226.59$, $p < 2.2e\text{-}16$). Post-hoc Wilcoxon tests showed that Trailcam had significantly higher scores than both Handicam ($p = 2.3e\text{-}06$) and Webcam ($p < 2e\text{-}16$). Handicam and Webcam were also significantly different ($p = 1.4e\text{-}05$), with Webcam scoring the lowest. N = 3945 Handicam, 263 Trailcam, and 2700 Webcam points identified by DLC.

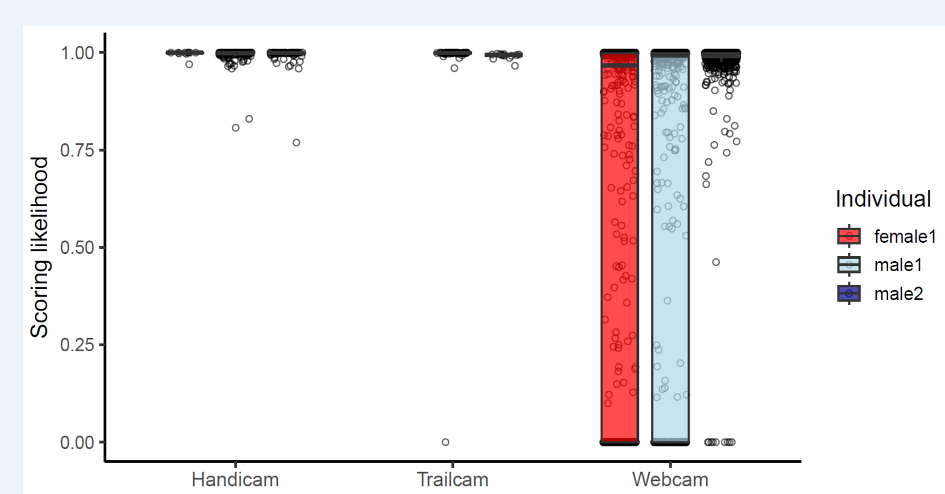


Figure 3: Middleback Scoring Likelihood Across Video Types. Likelihood scores differed significantly across video types ($\chi^2 = 196.9$, $p < 2.2e\text{-}16$), though there was high variation among individuals. Post-hoc Wilcoxon tests showed that Handicam and Webcam were significantly different ($p = 0.0027$), while Trailcam had significantly higher scores than Webcam ($p < 2e\text{-}16$). Increasing the number of videos analyzed may help reduce individual variation. N = 3945 Handicam, 263 Trailcam, and 2700 Webcam points identified by DLC.

Hypothesis 1: Likelihood scores will be highest in webcam footage as these were used for DLC training and lower in Handicam and Trailcam videos. → **Rejected.** Webcam had the lowest likelihood scores, while Trailcam performed best.

Hypothesis 2: Consistently visible body parts (front crest, middleback) will have higher likelihood scores than those whose visibility varies with movement (tail base). → **Mostly supported.** Consistently visible body parts had higher likelihood scores, while body parts affected by movement or angle had greater variability.

FUTURE DIRECTIONS

Expand Dataset & Improve Video Selection

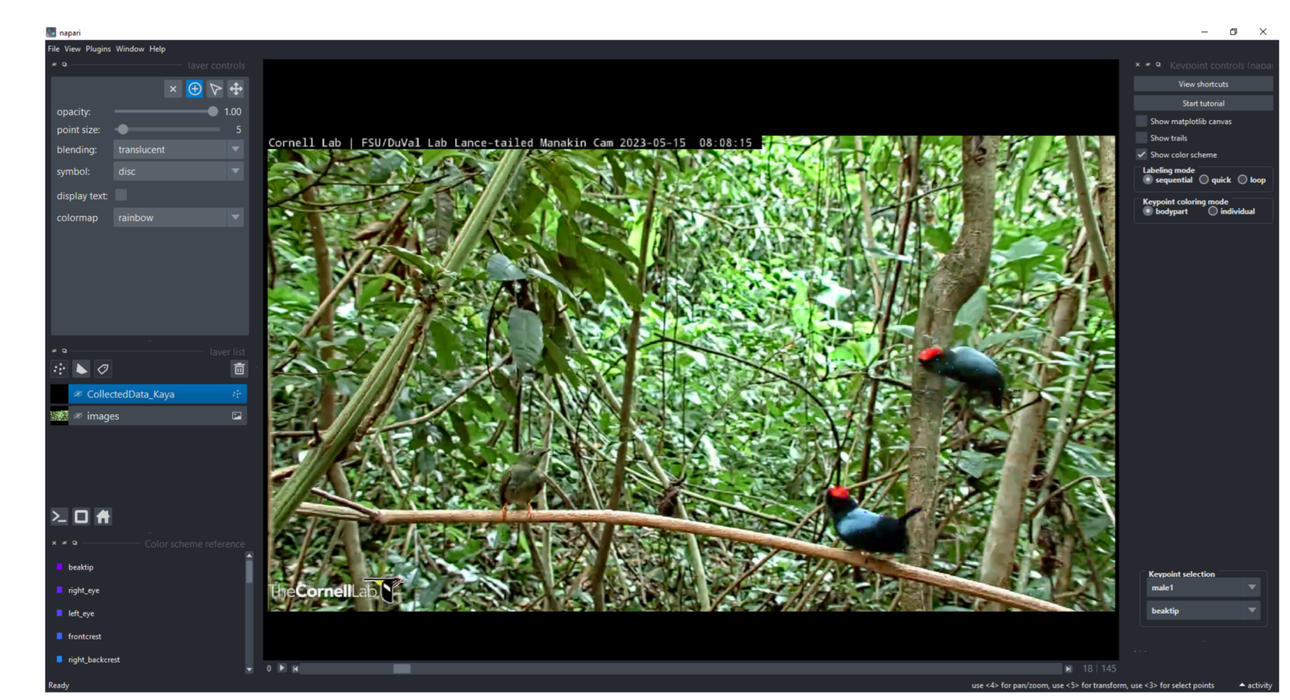
- Trailcam video included only two males, while Handicam and Webcam contained both two males and a female.
- Trailcam video had a smaller sample size of points identified by DLC compared to Handicam and Webcam,
- Comparing consistent videos with two males and a female will help us to better understand DLC's effectiveness across video types.

Assessing Camera Performance

- Our analysis indicates that Trailcam appears most reliable.
- This was unexpected since DLC was trained on Webcam videos.
- Next steps will investigate why Trailcams were the most successful in analysis.

Future Behavioral Analysis

- This work is an important step towards using DLC to analyze male courtship movements and display coordination.
- Analysis will quantify posture, wing movement, and head positioning in courtship.
- This will allow investigation of movement differences linked to copulation success.



DeepLabCut Program in Labeling Phase

ACKNOWLEDGEMENTS

I would like to thank everyone in the DuVal lab for their constant support throughout this project!

REFERENCES

DuVal E.H. (2007) Cooperative display and lekking behavior of the lance-tailed manakin (Chiroxiphia lanceolata). The Auk 124 (4): 1168-1185.

Mathis, A., Mamidanna, P., Cury, K.M. et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat Neurosci 21,1281–1289(2018).https://doi.org/10.1038/s41593-018-0209-y