Using Machine Learning to Quantify Complex Behavior in a Tropical Bird

Kaya Simmons, Ashlynn Gonzalez, Emily DuVal
Florida State University, Department of Biological Science

Introduction

Animal behavioral repertoires can be complicated and difficult to characterize objectively. Researchers often quantify behavior manually through watching and annotating video files, which can be limited by user bias, requires expert knowledge, and can be very time-consuming (Luxem et al. 2023). Recent technological developments allow for more efficient analysis of complex behaviors while minimizing human error. Machine learning uses computer algorithms to identify patterns in data with minimal human guidance. Here we evaluate the use of DeepLabCut (DLC), an open-source deep learning tool for markerless pose estimation (Nath et al. 2019), to rapidly and effectively analyze fine-scale differences in individual movements during courtship display. We specifically assessed (1) its accuracy in identifying a range of male and female landmarks and (2) its performance in analyzing video generated by different media (Handicam, Webcam, and Trailcam).

Our System

Lance-Tailed Manakins (*Chiroxiphia lanceolata)* are small, tropical passerine birds that participate in complex cooperative courtship displays. These birds have a lek mating system, in which males perform displays at set locations while females explore options and freely choose their mates. These displays involve multiple males performing a fast-moving, coordinated dance. They can involve up to 11 unique display elements and can last up to 45 minutes (DuVal 2007). These qualities make observation and quantification of courtship displays difficult for humans to manage. Machine learning holds great potential to provide new insight into how and why these displays vary.

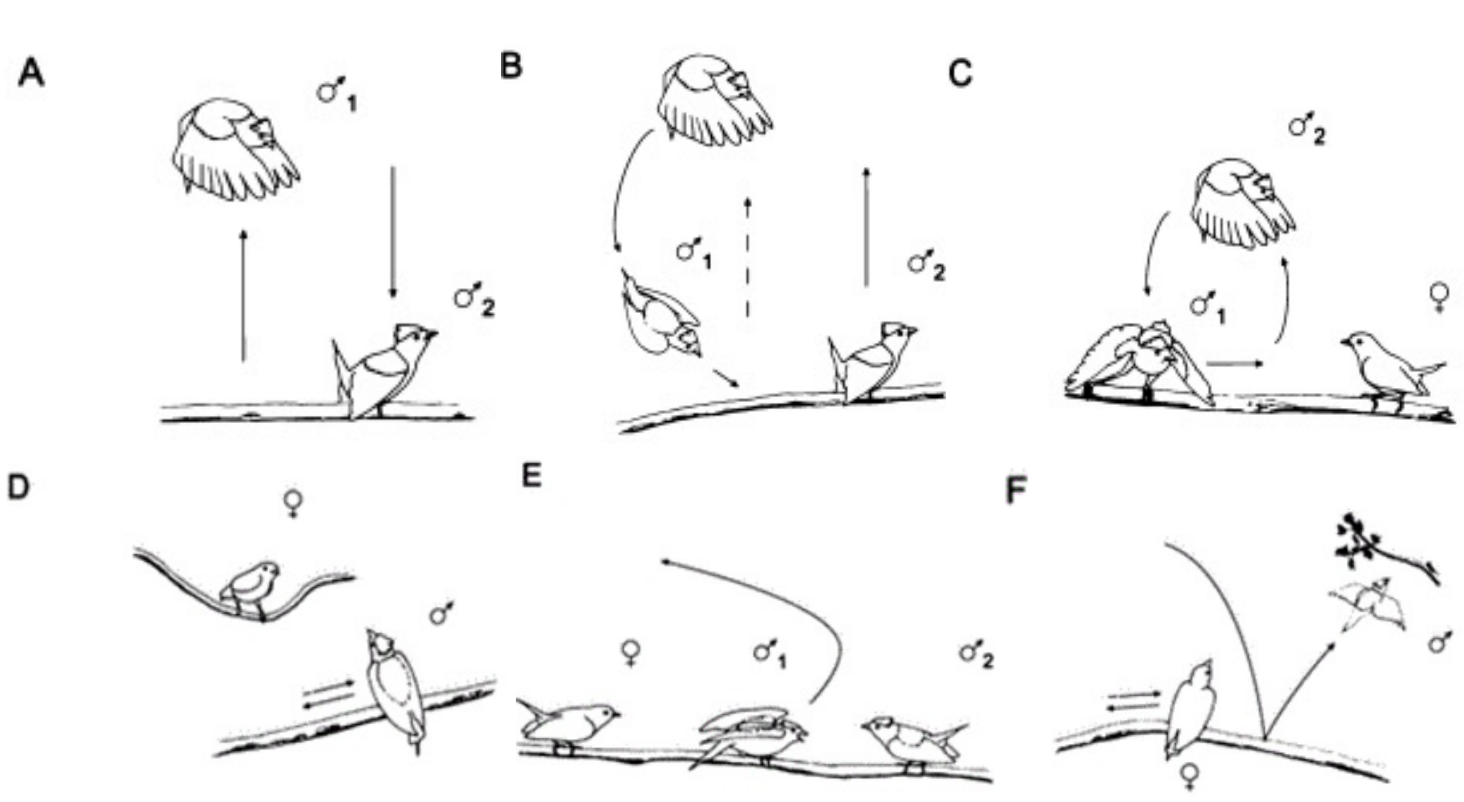

Figure 1. Example manakin display elements (DuVal, 2007)

Pose-Estimation and Tracking

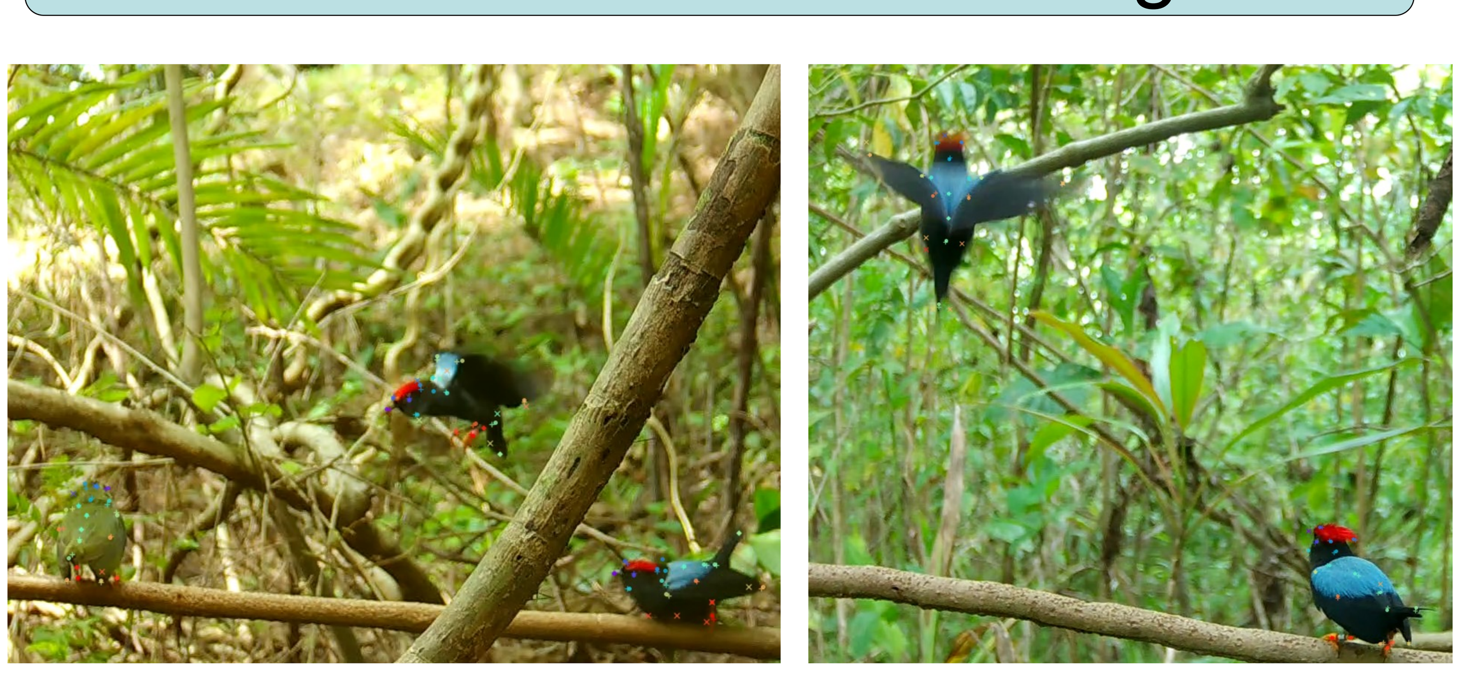

Figure 2. Examples of labeled frames of manakin displays post-training and evaluation depicting DeepLabCut's performance. Plus signs (+) are human-labeled points, dots are predictions with a likelihood >.6, and crosses (X) are predictions with a likelihood <.6.

References

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

Luxem, K., Sun, J. J., Bradley, S. P., Krishnan, K., Yttri, E., Zimmermann, J., ... & Laubach, M. (2023). Open-source tools for behavioral video analysis: Setup, methods, and best practices. *Elife*, 12, e79305.

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature protocols*, 14(7), 2152-2176.

Goal 1: Determine program accuracy in body-part identification

Methods:

- Using the AI software DeepLabCut, we extracted 150 frames from two videos of manakin displays involving a female and multiple males.
- We identified and labeled 24 body parts for each individual across frames and used these to create a training dataset that the algorithm used to predict points.
- After training, we allowed the program to evaluate its accuracy. This generated data on confidence for each predicted point (Figure 3), as well as frames to visualize performance (Figure 2).

Results:

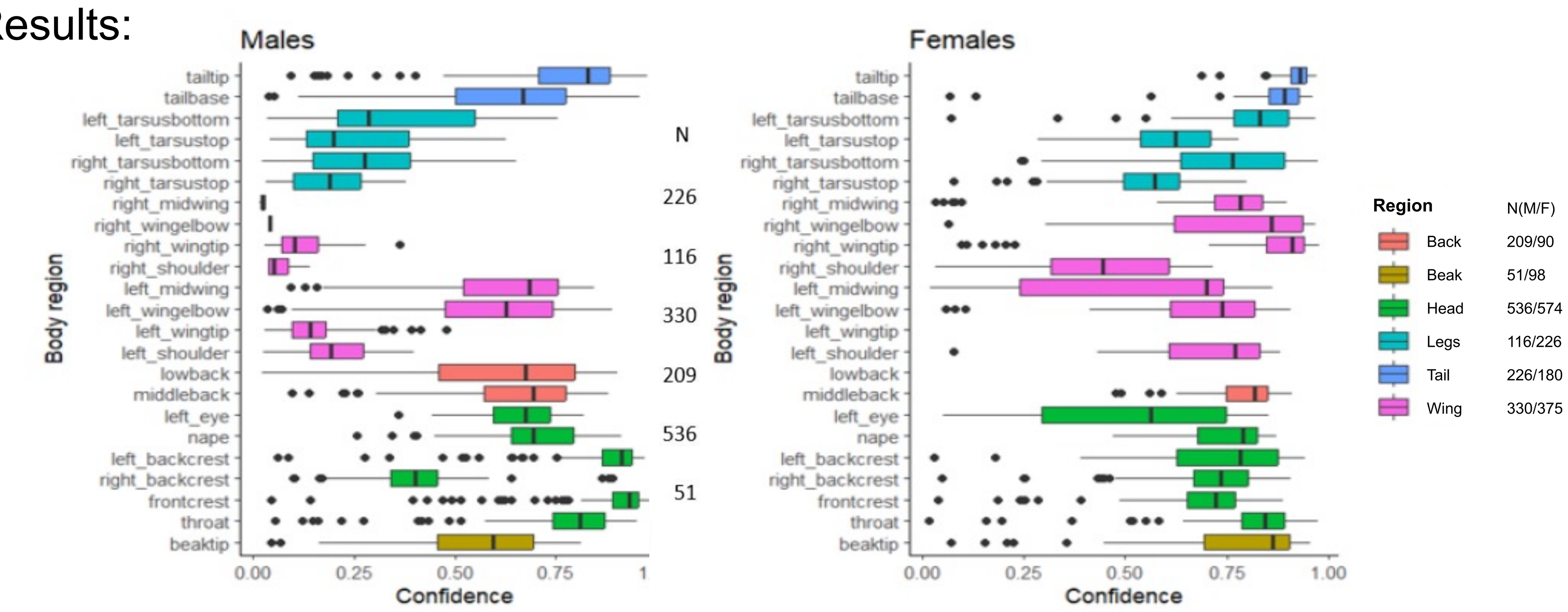

Figure 3. Box plots comparing program's confidence in identifying all labeled body parts by sex. N is the number of measures compared by color-coded body region.

- Body parts of females were on average more confidently assigned than those of males (Kruskal-Wallis X^2 (1)= 252.8, p<0.0001).
- Within sexes, body regions varied significantly in confidence: Males (Kruskal-Wallis $X^2(5)$= 518.67, p < 0.0001); Females (Kruskal-Wallis X2(5)= 212.96, p < 0.0001).

Goal 2: Determine program accuracy across media types

Methods:

- Extracted one clip per camera type (Webcam, Trailcam, Handicam) from longer recordings of lance-tailed manakins using Movie Studio Platinum.
- Processed videos in DeepLabCut to track lance-tailed manakin body parts. Focused on front crest and tail base, landmarks that identify key shapes of the bird.
- Analysis produced likelihood values that each body part for each bird was detected accurately. 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:

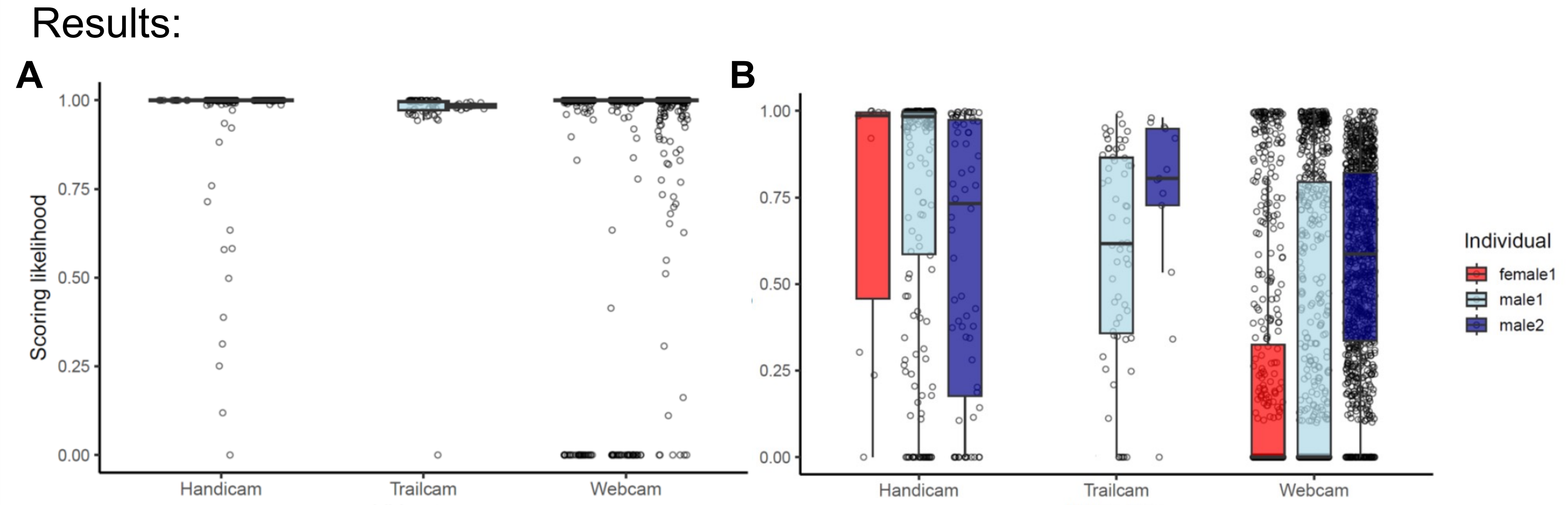

Figure 5: Scoring Likelihood Across Video Types. **A.** Front crest (*χ^2 = 196.9, p < 2.2e-16;* N = 3945 Handicam, 263 Trailcam, and 2700 Webcam frames). **B.** Tail base (*χ^2 = 226.59, p < 2.2e-16,* N = 3945 Handicam, 263 Trailcam, and 2700 Webcam frames).

- Likelihood scores differed significantly across video types for two different body points.
- Front crest: 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-16*).
- Tail base: post-hoc Wilcoxon tests showed that Trailcam had significantly higher scores than both Handicam (*p = 2.3e-06*) and Webcam (*p < 2e-16*). Handicam and Webcam were also significantly different (*p = 1.4e-05*), with Webcam scoring the lowest.

Summary and Next Steps

- Confidence of DLC's body part identification varied by region, and fast-moving parts (e.g., wings) had lower confidence than more stable parts (e.g., tail).
- Evaluation of female poses showed higher accuracy than males, likely due to less movement during recording.
- DLC performed well across media types. Recording format differences did not limit pose estimation.
- Some videos/image types may need retraining to improve recognition and performance.
- The ability to accurately track fine-scale movements in video of manakin displays enables development of further machine learning tools to evaluate:
  - How do successful and unsuccessful displays vary?
  - How does male display performance change over years of experience?
  - How do fine-scale male and female interactions during a display affect its progression?