PrimateFace: A Resource for Automated Face Analysis in Human and Non-human Primates

F. Parodi[1] **J.K. Matelsky**[1,2] **A.P. Lamacchia**[1] **M. Segado**[1]

Y. Jiang[1] **A. Regla-Vargas**[1] **L. Sofi**[1] **C. Kimock**[3]

B.M. Waller[3] **M.L. Platt**[1] **K.P. Kording**[1,4]

[1]University of Pennsylvania
[2]Johns Hopkins University Applied Physics Laboratory
[3]Nottingham Trent University
[4]CIFAR

Abstract

Machine learning has revolutionized human face analysis, but equivalent tools for non-human primates remain limited and species-specific, hindering progress in neuroscience, anthropology, and conservation. Here, we present PrimateFace, a comprehensive, cross-species platform for primate facial analysis comprising a systematically curated dataset of 260,000+ images spanning over 60 genera, including a genus-balanced subset of 60,000 images, annotated with bounding boxes and facial landmark configurations. Face detection and facial landmark estimation models trained on PrimateFace achieve high cross-species performance, from tarsiers to gorillas, achieving performance comparable to baseline models trained exclusively on human data (0.34 vs. 0.39 mAP for face detection; 0.061 vs. 0.053 normalized landmark error), demonstrating the generalization benefits of cross-species training. PrimateFace enables diverse downstream applications including individual recognition, gaze analysis, and automated extraction of stereotyped (e.g., lip-smacking) and subtle (e.g., soft left turn) facial movements. PrimateFace provides a standardized platform for facial phenotyping across the primate order, empowering data-driven studies that advance the health and well-being of human and non-human primates.

1 Introduction

Faces are a primary canvas for communication in social animals, conveying identity, intent, emotional state, and direction of attention. While research in animal communication has often focused on bioacoustics, facial movements represent a rich, high-dimensional, and comparatively understudied communication channel. The primary challenge lies in reliably quantifying these signals—transforming raw video into structured, kinematic data fundamental to decoding the intricate dynamics of social behavior.

While deep learning has revolutionized human facial analysis, powered by massive datasets like WIDERFace [Yang et al., 2016], equivalent tools for non-human primates remain limited and species-specific [Bala et al., 2020]. Existing approaches are typically trained on small, taxonomically narrow datasets, leading to specialized models that fail to generalize across the immense morphological

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI for Non-Human Animal Communication.

heterogeneity of the primate order [Schofield et al., 2023]. This limitation, coupled with the laborious nature of manual coding methods like the Facial Action Coding System (FACS) [Waller et al., 2020], has made large-scale, comparative studies of facial communication computationally intractable.

Here, we introduce PrimateFace, a foundational resource designed to address this challenge. Our contributions are threefold: (1) We constructed the largest and most taxonomically diverse dataset of annotated primate faces; (2) We developed a suite of pretrained models that generalize across species; and (3) We showcase the utility of PrimateFace for applications central to understanding animal communication.

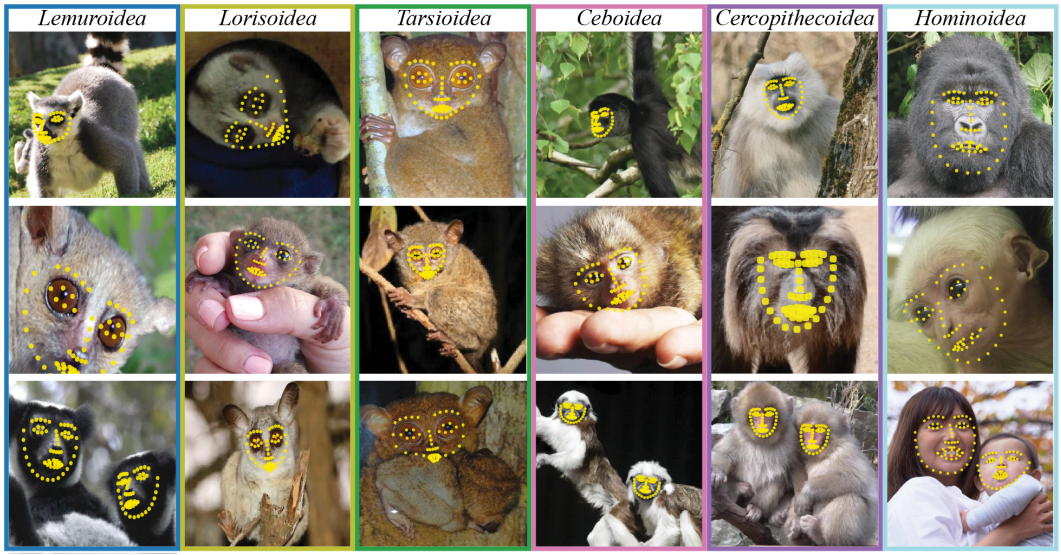

Figure 1: **PrimateFace dataset provides a diverse foundation for cross-species analysis.** The dataset spans all six primate superfamilies, capturing immense morphological diversity crucial for training generalizable models.

2 PrimateFace is an integrated ecosystem for primate face analysis

Dataset Construction. The foundation of our resource is a large-scale, taxonomically diverse dataset curated specifically for pretraining generalizable models. It comprises over 260,000 images of more than 60 primate genera, with prioritized taxonomic breadth spanning all six primate superfamilies, from Lemuroidea to Hominoidea (Figure 1). This diversity is critical for learning the invariant features that define a primate face, forcing models to move beyond species-specific traits. Every face is annotated with a bounding box and a standardized 68-point landmark configuration, transforming unstructured pixels into a structured kinematic biosignal ready for downstream analysis.

Model Development. Using this dataset, we trained several standard computer vision models (e.g., HRNet for landmark estimation, also known as Facial Landmark Estimation (FLE)) using frameworks like OpenMMLab, DeepLabCut [Mathis et al., 2018], and Ultralytics. Our goal was not to invent new architectures, but to create robust, off-the-shelf tools that empower researchers to apply AI to their own data.

3 Results: From Generalizable Models to Communication Insights

3.1 Taxonomic Scaling Improves Cross-Dataset Transfer

Our experiments confirm that pretraining on the taxonomically diverse PrimateFace dataset yields models with strong generalization. When evaluated on the COCO-WholeBody-Face human benchmark, our model achieves a Normalized Mean Error (NME) of 0.061, performing competitively with specialist models trained exclusively on human data (0.053 NME). In contrast, a model trained only on human data exhibits substantial degradation on our diverse primate dataset (0.122 NME vs.

0.029 NME). Because PrimateFace includes human faces, these results reflect the benefit of broader taxonomic coverage—pretraining on morphologically diverse data yields general representations that transfer broadly—rather than a claim of symmetry. A stricter symmetry test would hold out humans during pretraining. This insight is critical for developing foundation models for the broader animal kingdom.

3.2 Applications in Deciphering Primate Communication

PrimateFace is not just a dataset; it is an engine for discovery. We showcase its utility in domains central to animal communication research (Figure 2).

Multimodal Communication: Vocal-Motor Coupling in Howler Monkeys. Understanding vocal communication requires coordinating facial movements and sound. We applied our FLE model to video of a howling howler monkey to extract a continuous kinematic signal of mouth aperture (Figure 2A). Aligning this with the acoustic envelope of the vocalization allows for the precise quantification of coupling between mouth motion and vocal output, enabling a more mechanistic understanding of vocal production.

Data-Driven Discovery of a Facial "Vocabulary" in Macaques. Traditional ethology relies on predefined behavioral categories that may miss subtle patterns. PrimateFace's precise landmark tracking enables the data-driven discovery of 'behavioral syllables' from facial kinematics. Using a pipeline inspired by traditional unsupervised approaches [Berman et al., 2014], we automatically identified over 80 recurrent movement patterns from high-resolution macaque video (Figure 2B). This approach discovered both stereotyped movements (e.g., lip-smacking) and subtle, previously unobserved sex-specific differences in communication repertoires, revealing a rich facial "vocabulary."

Quantifying Social Building Blocks: Identity and Gaze. Effective communication depends on knowing who is signaling and where their attention is directed. PrimateFace automates the front-end steps for both. For identity, we built a pipeline that achieved 85.8

4 Conclusion

PrimateFace is a foundational resource that bridges the gap between AI and the study of non-human animal communication. By providing the largest, most taxonomically diverse primate face dataset and a suite of robust models, we enable a critical shift from manual, small-scale analysis to automated, large-scale investigation of facial communication. Our work provides the first empirical proof of a 'taxonomic scaling' principle for behavior, where model generalization scales with the morphological heterogeneity of the training data.

This resource empowers a new generation of scalable, data-driven studies into the intricate links between brain, body, and communication. It also provides a challenging new benchmark for the AI community to develop models for multimodal learning, few-shot recognition, and the data-driven discovery of behavioral structure. From an ethical standpoint, we note all human data were from licensed stock footage, and our lightweight models are accessible, promoting responsible use. Ultimately, PrimateFace is a step towards a more comprehensive understanding of the rich communicative lives of our closest relatives.

References

Praneet C Bala et al. Automated markerless pose estimation in freely moving macaques with openmonkeystudio. *Nature communications*, 11(1):1–12, 2020.

Gordon J Berman et al. Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, 11(99), 2014.

Alexander Mathis et al. Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience*, 21(9):1281–1289, 2018.

Daniel P Schofield et al. Automated face recognition using deep neural networks produces robust primate social networks and sociality measures. *Methods in Ecology and Evolution*, 14(8):1937–1951, 2023.

B. M. Waller, E. Julle-Daniere, and J. Micheletta. Measuring the evolution of facial 'expression' using multi-species facs. *Neuroscience Biobehavioral Reviews*, 113:1–11, 2020.

Shuo Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5525–5533, 2016.

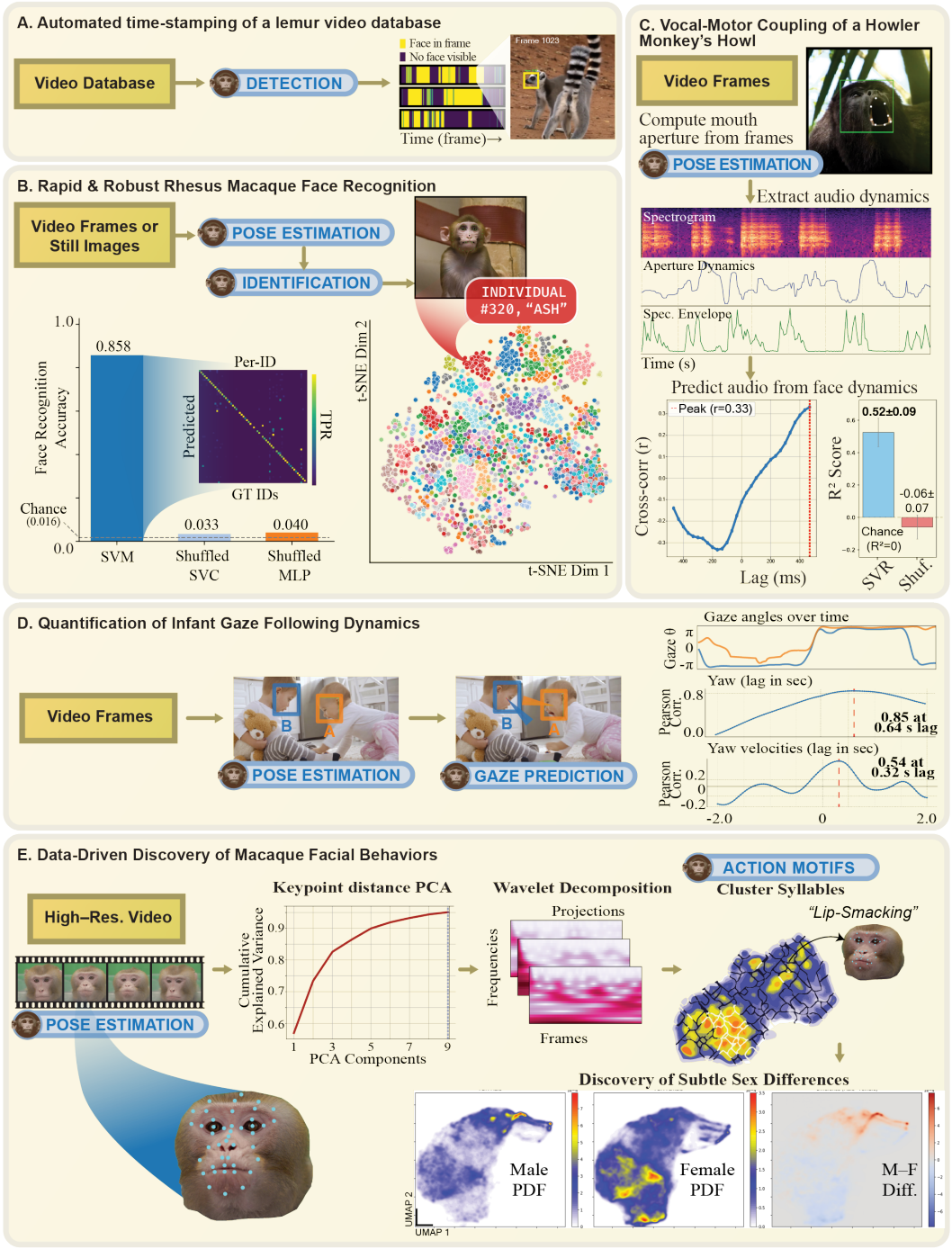

Figure 2: **PrimateFace accelerates and enables diverse research in animal communication.** (A) Vocal-motor coupling analysis in a Howler Monkey. (B) Data-driven discovery of a macaque facial "vocabulary". (C) Rapid, robust face recognition for identity signaling. (D) Quantifying social gaze dynamics in human infants.