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

Anonymous Author(s)
Affiliation
Address
email

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 analyzing facial communication 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 essential conduits for conveying information critical to social animals, with relevance to psychology, neuroscience, evolutionary biology, and conservation. Among the many signals the brain and body produce, facial movements are a particularly high-dimensional and information-rich stream of data. The primary challenge lies in reliably quantifying this signal by transforming raw video into a structured, continuous kinematic signal – a process fundamental to decoding the intricate dynamics of social communication.

While deep learning has driven progress in human facial analysis, powered by massive datasets like WIDERFace (Yang and others [2016] and COCO-WholeBody (Jin and others [2020], equivalent tools for non-human primates remain limited (Bala and others [2020] Carugati et al. [2025]). Existing approaches are typically trained on small, taxonomically narrow datasets, leading to specialized models that fail to generalize across the immense morphological heterogeneity of the primate order (Schofield and others [2023]). This diversity, coupled with the limitations of labor-intensive 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. A foundational resource – a large-scale, taxonomically diverse pretraining dataset – is required to learn truly generalizable representations.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

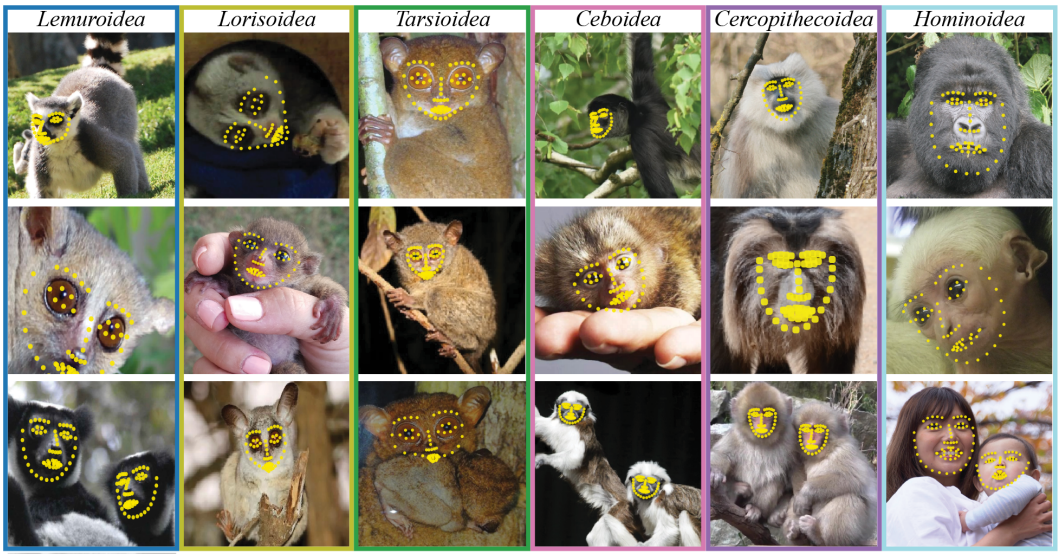

Figure 1: PrimateFace dataset provides a diverse foundation for cross-species analysis.

Here, we introduce PrimateFace, a foundational resource designed to address this challenge and accelerate research in behavioral biosignal analysis. Our contributions are threefold:

1. We constructed the largest and most taxonomically diverse dataset of annotated primate faces, designed for pretraining models that learn generalizable representations.
2. We developed a suite of pretrained models and demonstrate that their powerful generalization properties stem from pre-training on taxonomically diverse data.
3. We showcase the utility of PrimateFace as a front-end for diverse scientific applications, from cross-species behavioral analysis to automated individual recognition.

2 A Cross-Species Dataset for Pretraining

The foundation of our resource is a large-scale, taxonomically diverse dataset curated specifically for pretraining generalizable models, comprising over 260,000 images of more than 60 primate genera. We prioritized taxonomic breadth, spanning all six primate superfamilies, from Lemuroidea to Hominoidea (Figure 1). Exposing models to this diversity is critical for learning the invariant features that define a primate face, forcing the model to move beyond species-specific traits and develop a more fundamental understanding of primate facial structure, which is the key to generalization.

The annotations in PrimateFace are designed to transform unstructured pixel data into structured, analyzable biosignals Figure A.1. Every face is annotated with a bounding box and a standardized 68-point landmark configuration. When applied to video, models trained on this data produce a continuous time-series of landmark coordinates – a high-dimensional communicative signal that serves as the input for downstream analysis.

3 PrimateFace Enables Cross-Species Facial Analysis

A foundational resource is only as valuable as the models it can produce. We trained several computer vision models, including models from the OpenMMLab ecosytem, DeepLabCut (Mathis and others [2018],) SLEAP (Pereira and others [2022],) and Ultralytics (e.g., Zhang and others [2025]) and evaluated their generalization properties.

Our experiments confirm that pretraining on the taxonomically diverse PrimateFace dataset yields models with remarkable generalization. When evaluated on the challenging COCO-WholeBody-Face human benchmark in a zero-shot setting, our PrimateFace-trained model achieves a Normalized Mean Error (NME) of 0.061, performing competitively with specialist models trained exclusively on human data (0.053 NME) Figure A.2. This generalization is notably asymmetric; a model trained only on

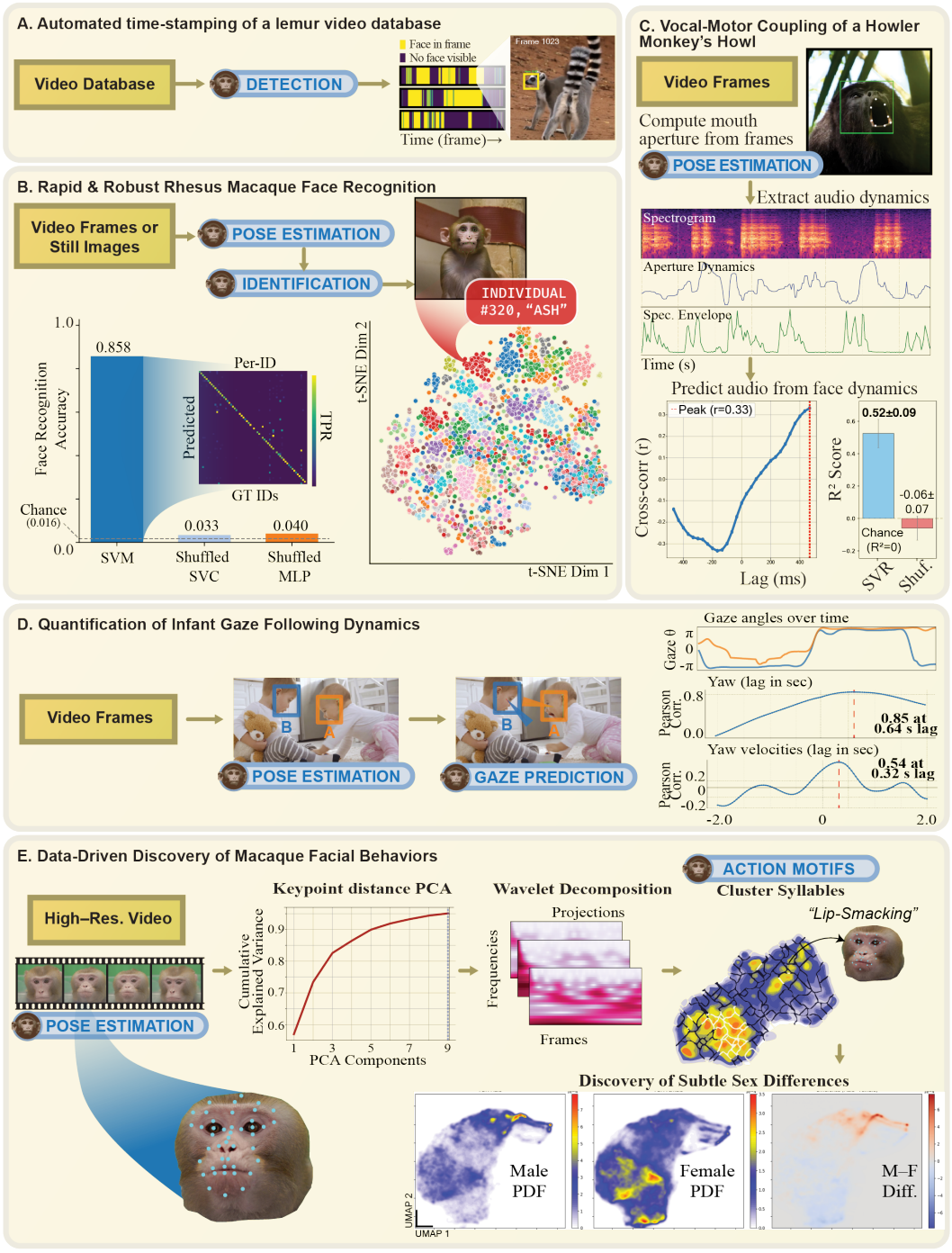

Figure 2: PrimateFace accelerates and enables diverse research applications.

human data exhibits a substantial performance degradation when evaluated on our diverse primate dataset (0.122 NME vs. 0.029 NME) (Figure A.3). This result powerfully illustrates the benefit of our approach: **pretraining on morphologically diverse data yields general representations that transfer broadly, whereas narrow pretraining results in specialized models that fail to generalize.**

4 Downstream Applications Enabled by PrimateFace

To demonstrate the utility of PrimateFace as a resource for studying facial communication, we present its application in five distinct domains Figure 2, ranging from lemur face tracking in the wild to howling analysis to facial motif discovery.

Scientific Automation: Automated Time-Stamping. Manually logging when a face is visible in video is a time-consuming bottleneck in observational research. PrimateFace's cross-species detectors automate this critical step. Our pipeline processes video frame-by-frame to effectively compress days of raw footage into concise visualizations of individual visibility over time, significantly enhancing the efficiency of longitudinal monitoring.

Scientific Automation: Rapid Individual Recognition. Building individual recognition systems is historically labor-intensive. PrimateFace automates the critical front-end steps of detection and alignment. Using our models, a pipeline to detect, align, and generate embeddings for a classifier was executed on a public dataset of 62 macaques in under an hour, achieving 0.858 top-1 accuracy. This demonstrates how PrimateFace enables researchers to rapidly create accurate ID systems for large cohorts without specialized development.

Cross-Signal Analysis: Vocal-Motor Coupling of Howler Monkey's Howl. Understanding vocal communication requires coordinating facial movements and sound. We applied our facial landmark estimation model to extract a continuous kinematic signal of mouth aperture from video of a howling howler monkey. Aligning this with the acoustic signal (the spectrogram's temporal envelope) allows for the quantification of precise coupling between mouth motion and vocal output, enabling a more mechanistic understanding of vocal production.

Cross-Species Generalization: Quantifying Social Gaze in Human Infants. Analyzing joint attention in developmental psychology is notoriously labor-intensive. Our resource's demonstrated cross-species generalization allows us to apply PrimateFace models "off-the-shelf" to human data. We use our robust face detector to track interacting infants, which then serve as inputs to a downstream gaze estimation model (Ryan and others [2024]). This pipeline enables automated, objective, and scalable quantification of fine-grained behavioral synchrony.

Cross-Subject, Data-Driven Discovery of Facial Action Motifs. Traditional ethological approaches rely on pre-defined behavioral categories that may miss subtle patterns. PrimateFace's precise landmark tracking enables 'behavioral syllable' discovery from facial kinematics. Using a pipeline inspired by traditional unsupervised approaches (Berman and others [2014]), we automatically identified over 80 recurrent movement patterns from high-resolution macaque video, discovering both stereotyped movements (e.g., lip-smacking) and subtle, previously unobserved sex-specific differences in communication repertoires.

5 Discussion

We introduce PrimateFace, a resource for the analysis of facial communication across the primate family. Training models on PrimateFace, a large-scale, taxonomically diverse dataset, is critical towards overcoming the generalization failures of previous species-specific approaches in studying primate facial communication. Our models' limitations outline our next steps: as a 2D system, it cannot fully disentangle expression from pose, motivating the development of 3D models. Furthermore, performance correlates with taxonomic density, highlighting the need for continued, targeted data collection for the most morphologically unique genera (e.g., Daubentonia). From an ethical standpoint, while powerful, face analysis technologies carry risks of misuse, such as in mass surveillance. In PrimateFace, we note all human data used in our demonstrations were from licensed stock footage. Ultimately, PrimateFace empowers a new generation of scalable, data-driven studies into the intricate links between brain, body, and behavior.

References

Praneet C. Bala and others. Automated markerless pose estimation in freely moving macaques with OpenMonkeyStudio. *Nature communications*, 11:1, 2020.

Gordon J. Berman and others. Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, 11:99, 2014.

Filippo Carugati, Olivier Friard, Elisa Protopapa, Camilla Mancassola, Emanuela Rabajoli, Chiara De Gregorio, Daria Valente, Valeria Ferrario, Walter Cristiano, Teresa Raimondi, Valeria Torti, Brice Lefaux, Longondraza Miaretsoa, Cristina Giacoma, and Marco Gamba. Discrimination between the facial gestures of vocalising and non-vocalising lemurs and small apes using deep learning. *Ecological Informatics*, 85:102847, March 2025. ISSN 1574-9541. doi: 10.1016/j.ecoinf.2024.102847. URL `https://www.sciencedirect.com/science/article/pii/S1574954124003893`.

Sheng Jin and others. Whole-body human pose estimation in the wild. In *European Conference on Computer Vision. Cham*. Springer International Publishing, 2020.

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

Talmo D. Pereira and others. *SLEAP: A deep learning system for multi-animal pose tracking*. 1-10, Nature methods, 2022.

Fiona Ryan and others. Gaze-LLE: Gaze Target Estimation via Large-Scale Learned Encoders. _eprint: 2412.09586, 2024.

Daniel P. Schofield and others. 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, June 2020. ISSN 01497634. doi: 10.1016/j.neubiorev.2020.02.031. URL `https://linkinghub.elsevier.com/retrieve/pii/S0149763419302404`.

Shuo Yang and others. Wider face: A face detection benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

Jia-Jin Zhang and others. A deep learning lightweight model for real-time captive macaque facial recognition based on an improved YOLOX model. *Zoological Research*, 46:2, 2025.

A Technical Appendices and Supplementary Material

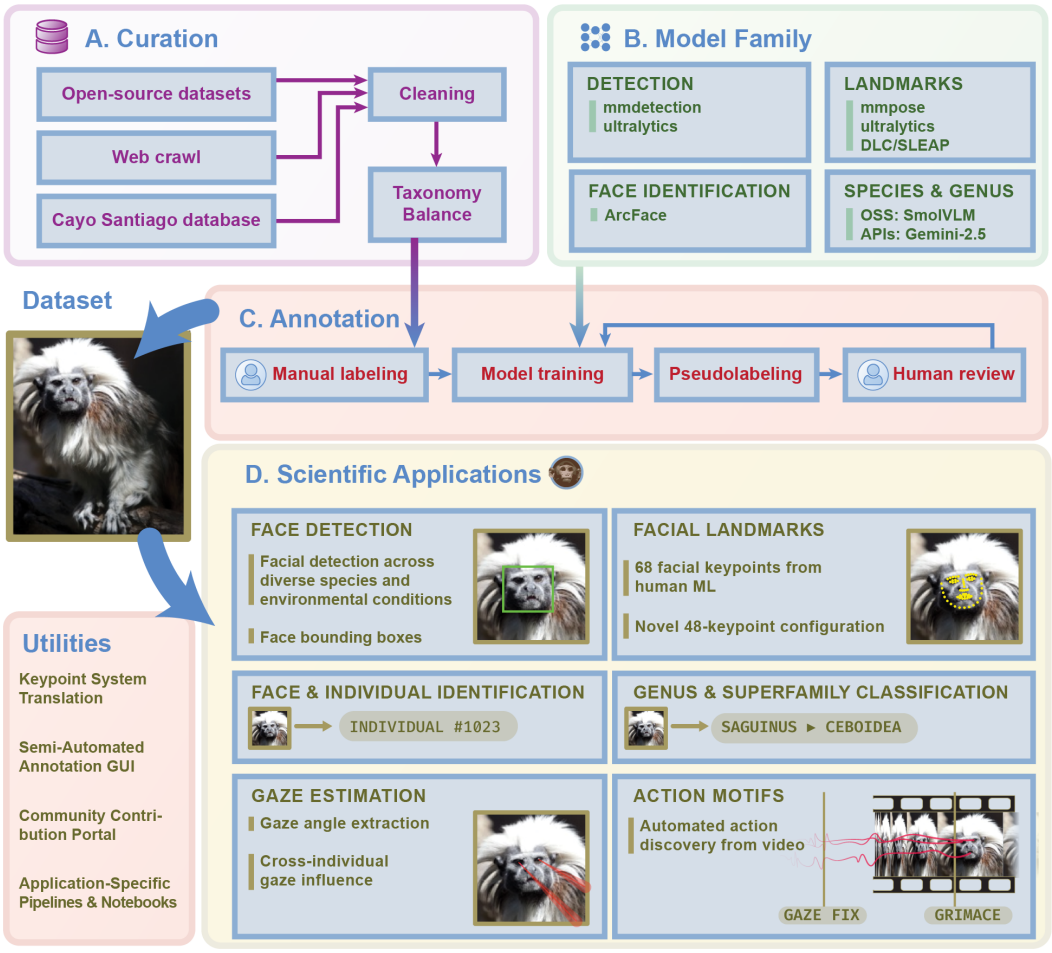

Figure A.1: **The PrimateFace Ecosystem.** An overview of our integrated workflow for building a foundational resource. The process unifies (A) large-scale data curation, (B) development of models using multiple open-source frameworks, and (C) a scalable, semi-automated annotation pipeline. This iterative loop enables the creation of (D) diverse downstream scientific applications.

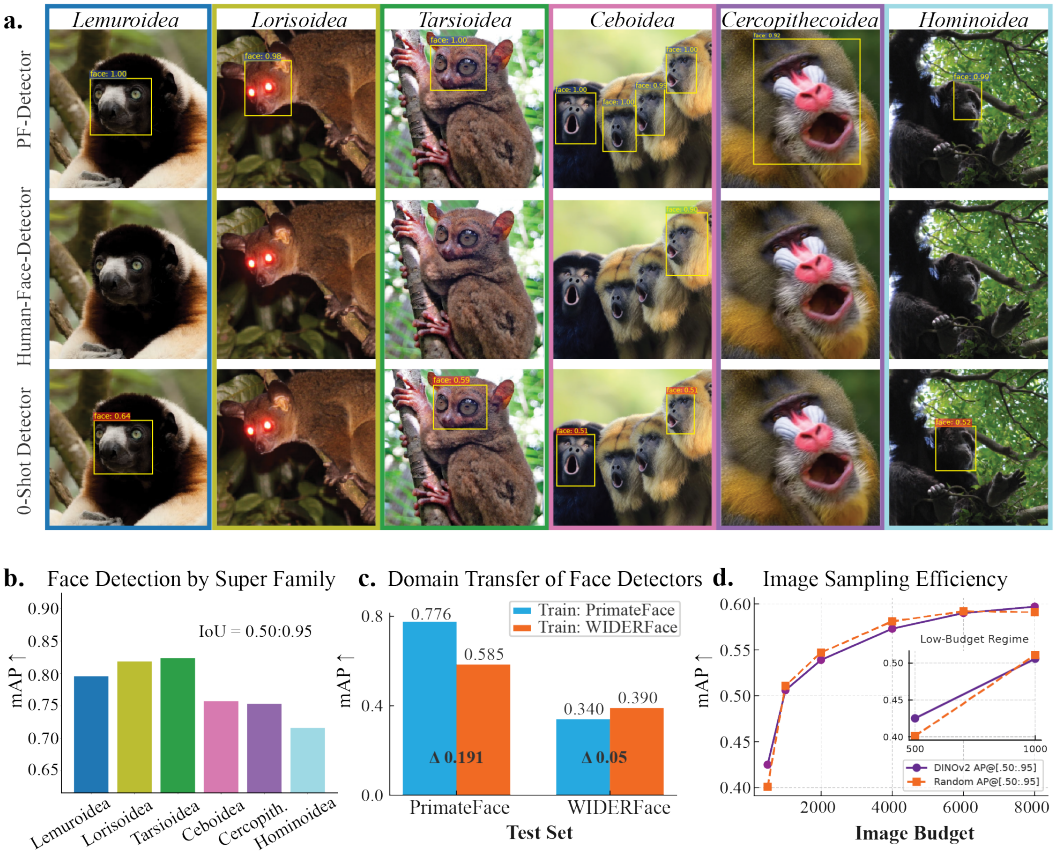

Figure A.2: **Evaluation of Face Detection Models.** (A) Qualitative comparison showing the superior performance of our PrimateFace-trained detector (top row) over a standard human-face detector (middle) and a zero-shot detector (bottom). (B) Detection performance (mAP) is robust across all six primate superfamilies. (C) Our PrimateFace-trained model shows strong zero-shot generalization to the human WIDERFace benchmark, while the human-trained model fails to generalize to our primate test set.

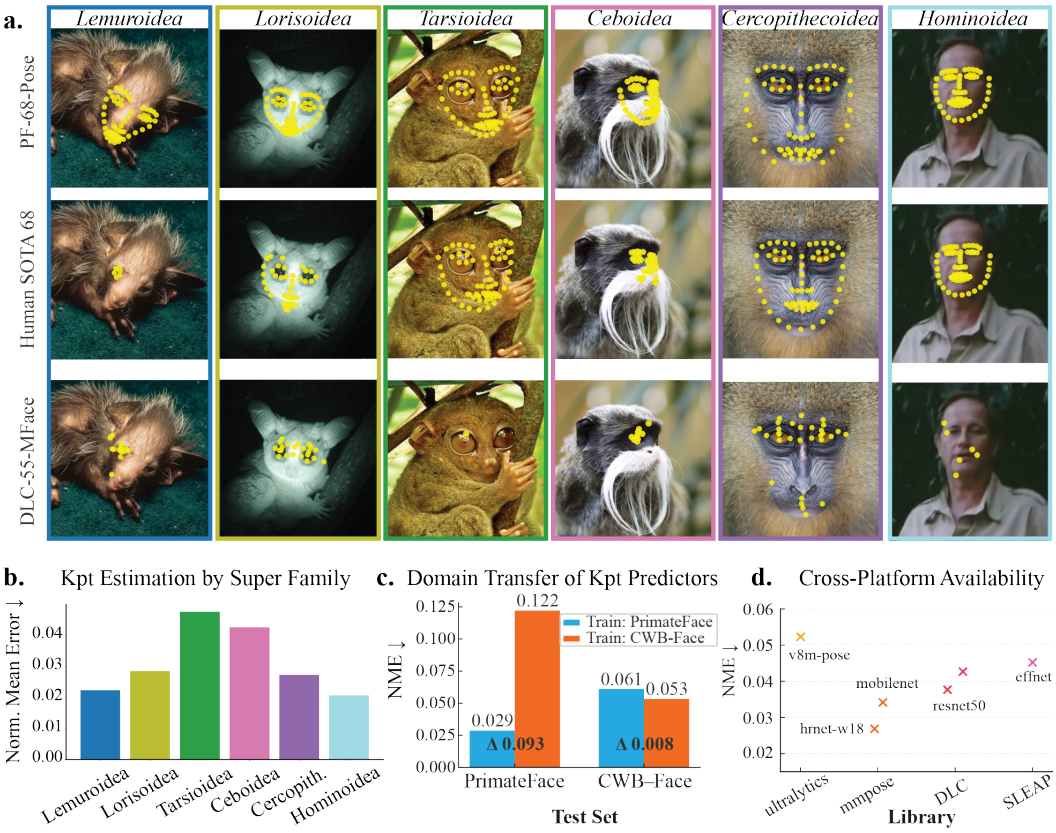

Figure A.3: **Evaluation of Facial Landmark Estimation (FLE) Models.** (A) Qualitative comparison showing our PrimateFace-trained model (top row) accurately localizes landmarks where human-specific (middle) and macaque-specific (bottom) models fail. (B) Normalized Mean Error (NME) is consistently low across superfamilies. (C) Our model generalizes to the human COCO-WholeBody-Face benchmark with performance competitive to a specialist model. This generalization is asymmetric, as the human-trained model performs poorly on our primate test set, highlighting the benefit of diverse pretraining.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.

- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [NA]

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.

- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).

- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [Yes]

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [Yes]

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.