Using Deep Learning to Quantify Phenotypic Similarities in Mimic Butterfly Species using Bird, Human, and Butterfly Acuities

Michelle Ramirez[1], Christopher Lawrence[2], David Carlyn[1], Matthew Thompson[1], Elizabeth Campolongo[1], Charles Stewart[3], Wei-Lun Chao[1], Daniel Rubenstein[2], and Tanya Berger-Wolf[1]

[1] The Ohio State University
[2] Princeton University
[3] Rensselaer Polytechnic Institute

Abstract. Quantifying variation in adaptive traits is essential for understanding evolutionary mechanisms such as mimicry in biology. Specifically, tropical Heliconius butterflies have bold wing patterns to warn off predators and have evolved by leveraging mimicry as a form of natural selection. In this work, we focus on taking a computational approach to modeling how butterflies perceive conspecifics, co-mimics, and other subspecies, and how they are perceived by external agents (birds, humans) using visual acuity measurements. We train a series of ResNet50 models with additive angular margin loss on images of Heliconius melpomene and Heliconius erato co-mimics processed under different visual acuities.

1 Introduction

Quantifying variation in adaptive traits is key to understanding evolutionary phenomena like mimicry. Heliconius butterflies (Lepidoptera; Nymphalidae) are aposematic, using defensive chemicals and warning coloration to deter predators. Multiple species converge on similar warning color patterns through Mullerian Mimicry, which leads to a learned aversion by predators. While predator-driven selection pushes for uniform wing patterns, sexual selection drives divergence to help butterflies distinguish between conspecifics (same subspecies) and heterospecifics (different species or groups) [3].

The mechanisms by which mimetic butterflies differentiate conspecifics from co-mimics remain unclear. It is believed to be multimodal, involving behavior, chemical signals, and visual cues [1,3,7]. For example, Heliconius butterflies use ultraviolet (UV) cues during mating to identify conspecifics, as the absence of UV leads to increased mimicry misidentification [3].

Behavioral and chemical studies seeking to understand identification mechanisms in co-mimic butterflies are informative but labor-intensive. Visual data, however, paired with modern computational methods such as deep learning, offer a more accessible starting point for understanding how mimetic butterflies perceive each other and how predators perceive them.

Deep learning, particularly in computer vision, is powerful for extracting distinguishing features from images. In this work, we apply deep learning to images processed under bird and butterfly visual acuities (R, AcuityView [2]) to investigate whether visual acuity alone can influence species classification.

Our study models how butterflies perceive conspecifics, co-mimics, and heterospecifics, and how they are perceived by birds and humans. We run species classification experiments using images of Heliconius melpomene and Heliconius erato subspecies, processed under different visual acuities. We build on [5] by incorporating visual acuity and an additive angular margin loss to improve learning across multiple subspecies.

Our deep learning experiments address three key questions: (1) Can computational methods detect similarities and differences in co-mimic images? (2) How does visual acuity (bird vs. butterfly) affect classification accuracy? (3) Does acuity influence the granularity of image traits used for classification?

We train models on all 20 subspecies (AllNet), only Heliconius erato subspecies (EratoNet), and only Heliconius melpomene subspecies (MelpomeneNet) to assess baseline performance and any classification bias arising from species-specific training. We demonstrate that deep learning can effectively differentiate traits among conspecifics, co-mimics, and heterospecifics, with visual acuity playing a stronger role in species classification for bird predators more so than butterflies.

2 Dataset

The dataset consists of a total of 3822 images sourced from the Heliconius Collection (Cambridge Butterfly) dataset [6]. Images consist of the dorsal view of 10 Heliconius erato subspecies and 10 Heliconius melpomene subspecies (Figure 1). The subspecies were chosen to be 1:1 H.erato and H.melpomene mimic pairings (Supplementary material, Table 3).

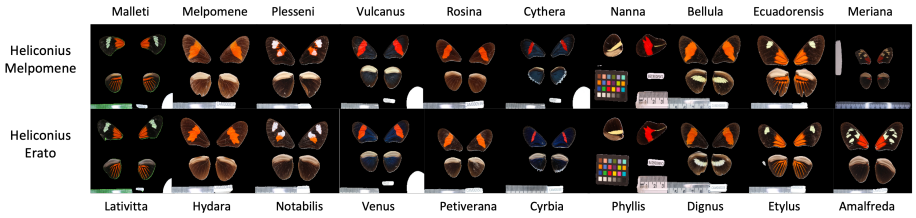


Fig. 1: Top row *Heliconius melpomene* subspecies left-to-right. **Bottom row** *Heliconius erato* subspecies left-to-right. Columns reflect comimic pairings between H.melpomene and H.erato.

We process all 3822 images by using a segmentation model to remove the background from all images and insert a black background (code available at:

`https://github.com/Imageomics/wing-segmentation`). We resize the images down from their original resolutions into dimensions of 256x256, as AcuityView requires images be square and dimensions be a factor of 2^n [2]. We use the Heliconius butterfly acuities (female, male, morphological, behavioral) measured in [8] to create four distinct Heliconius acuity datasets. Since there is little data on visual acuity of Jacamars (Heliconius' known bird predator) [3], we use the acuity belonging to Kingfishers (Supplementary material, Table 4) to approximate the visual acuity of a bird predator and create an additional dataset from our 3822 images. Finally, we keep a version of the original images unprocessed and denote this dataset version as "no acuity". We end up with a total of six different datasets of the same set of images processed under different visual acuities.

3 Methods

3.1 Model Architecture: ResNet50 and Additive Angular Margin Loss

We fine-tune ResNet50 (pretrained on ImageNet) with an ArcFace classification head individually on each acuity dataset. The ArcFace classification head incorporates additive angular margin loss to enhance discrimination between classes [4]. This is particularly beneficial for separating the visually similar classes of co-mimic butterflies. We implement several versions of the model, with the only difference being the data seen during training (the architecture remains the same). The differences are highlighted in Table 1. We ignore other light spectra (such as ultraviolet and violet), as well as behavior and motion components, leaving these aspects for future work.

Model	**Species seen at training time**
AllNet	*H. melpomene* and *H. erato*
EratoNet	*H. erato* only
MelpomeneNet	*H. melpomene*

Table 1: Species seen during training for different models.

3.2 Training AllNet

We begin by training our model for species classification on all 20 subspecies of *H. erato* and *H. melpomene* present in our image dataset. We refer to this model as AllNet and train a version of it for each acuity dataset (original, Heliconius male morphological, Heliconius female morphological, Heliconius male behavioral, Heliconius female behavioral, and kingfisher acuity). This setup aims to establish a baseline performance for the model's ability to discriminate between co-mimics for each acuity.

3.3 Training EratoNet and MelpomeneNet

We then set up an asymmetrical species classification experiment by training our model on one species (H. erato or H.melpomene) without access to their respective co-mimics (Table 1). We name these models EratoNet and MelpomeneNet, respectively. We aim to model the biological context in which model and mimic are presented to butterflies and birds. Furthermore, by training these asymmetrical models we can test whether there is any information loss when training on one species over the other.

4 Results

Given the dataset's class imbalance, we measure model performance through both micro and macro accuracy. For each trained model, we quantify the visual similarity of images processed under different acuities by computing pairwise Euclidean distances across image feature representations produced by the model. We also use GradCAM to analyze model predictions and gain insights into the specific image components that influence our models' ability to distinguish between co-mimics across different acuities.

4.1 AllNet

Table 2 shows the micro and macro accuracies of AllNet under various acuity conditions. AllNet achieved its best performance in mimic discrimination with Kingfisher acuity, reaching a micro accuracy of 0.972 and a macro accuracy of 0.875 (Table 2). Performance declined when AllNet was trained on datasets processed under the four butterfly acuities. Notably, models trained with Heliconius male visual acuities outperformed those with female acuities across both behavioral and morphological categories. Additionally, behavioral acuities outperformed morphological ones for both males and females.

4.2 EratoNet

Table 2 provides an evaluation of the EratoNet model, trained exclusively on Heliconius erato subspecies under six different visual acuity conditions. The performance is assessed on both Erato and Melpomene subspecies, highlighting how well the model is able to discriminate between co-mimics despite only seeing the Heliconius erato member of each erato:melpomene co-mimic pair at training.

EratoNet demonstrates consistent accuracy on Heliconius erato subspecies across various acuity model variants. Notably, the model trained with Heliconius male morphological acuity achieves the highest micro and macro accuracy (0.983 micro/ 0.873 macro Table 2). Accuracy (micro and macro) does not showcase a clear drop from high resolution images (Kingfisher and no acuity) to low resolution images (butterfly acuities). Comparison across models does not follow any clear pattern. Accuracies across acuity model variants are similar, and in some cases identical.

Table 2: Performance of AllNet, EratoNet, and MelpomeneNet Models Across Different Acuity Conditions

Acuity Condition	Model	Test On	Micro Acc	Macro Acc
No Acuity	AllNet	-	0.963	0.838
	EratoNet	Erato	0.978	0.844
	EratoNet	Melpomene	0.971	0.771
	MelpomeneNet	Melpomene	0.965	0.892
	MelpomeneNet	Erato	0.955	0.869
Heliconius Male Behavioral	AllNet	-	0.939	0.804
	EratoNet	Erato	0.969	0.784
	EratoNet	Melpomene	0.945	0.704
	MelpomeneNet	Melpomene	0.965	0.927
	MelpomeneNet	Erato	0.971	0.913
Heliconius Female Behavioral	AllNet	-	0.915	0.767
	EratoNet	Erato	0.978	0.844
	EratoNet	Melpomene	0.933	0.711
	MelpomeneNet	Melpomene	0.949	0.904
	MelpomeneNet	Erato	0.945	0.870
Heliconius Male Morphological	AllNet	-	0.916	0.760
	EratoNet	Erato	0.983	0.873
	EratoNet	Melpomene	0.938	0.746
	MelpomeneNet	Melpomene	0.953	0.863
	MelpomeneNet	Erato	0.941	0.805
Heliconius Female Morphological	AllNet	-	0.898	0.732
	EratoNet	Erato	0.971	0.809
	EratoNet	Melpomene	0.943	0.751
	MelpomeneNet	Melpomene	0.957	0.859
	MelpomeneNet	Erato	0.963	0.853
Kingfisher	AllNet	-	0.972	0.875
	EratoNet	Erato	0.978	0.844
	EratoNet	Melpomene	0.964	0.733
	MelpomeneNet	Melpomene	0.973	0.898
	MelpomeneNet	Erato	0.955	0.847

Testing EratoNet on melpomene subspecies led to a decrease in performance across all acuity model variants (Table 2). Again, no clear pattern was evident into deciphering how advantageous one form of acuity is over another, based on model accuracy.

4.3 MelpomeneNet

Table 2 provides an evaluation of the MelpomeneNet model, trained exclusively on Heliconius melpomene subspecies under six different visual acuity conditions. The Heliconius male behavioral acuity MelpomeneNet model had the highest

macro accuracy on the Melpomene test set of images (0.927, 2). When testing MelpomeneNet on erato subspecies (unseen at training time), the drop in performance (micro and macro accuracy) is not as large as those in EratoNet tested on melpomene (unseen at training time)(Tables 2, 2).

EratoNet and MelpomeneNet both had a butterfly acuity model variant outperform models trained on images processed at higher resolutions (no acuity applied and kingfisher acuity applied) (Tables 2, 2). However, in AllNet (all 20 subspecies seen at training time), models trained on images with higher resolutions (Kingfisher acuity and no acuity) outperformed models trained on images of lower resolution (butterfly acuities).

4.4 Embedding Distance

Training the model on all 20 subspecies (AllNet) produced results consistent with those observed in [5]. We find that the pairwise Euclidean distance of image embeddings for conspecifics is the smallest, while the distance between heterospecific image embeddings is the largest (Figure 2). The average distance between image embeddings of co-mimics falls between those of conspecifics and heterospecifics, but is much closer to the range of heterospecifics (Figure 2).

Similar trends are observed in the embedding distances for EratoNet and MelpomeneNet, though the range of co-mimic embedding distances shifts to resemble the range of conspecifics rather than heterospecifics (Figures 4, 5). Consequently, images of unseen subspecies are more likely to be misclassified as co-mimics across all acuities for both EratoNet and MelpomeneNet models.

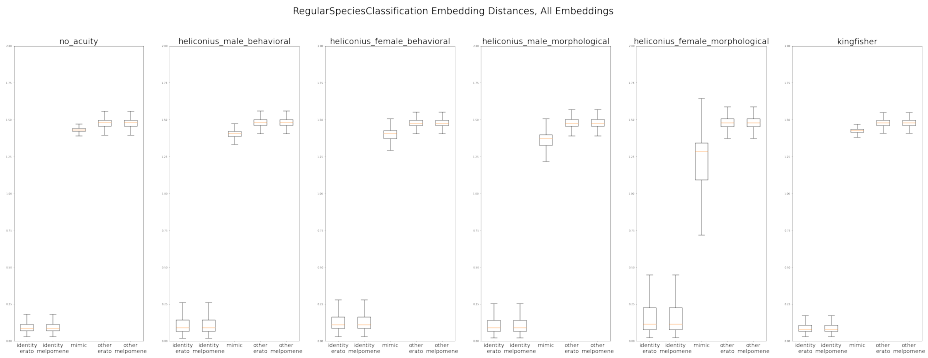


Fig. 2: Pairwise euclidean distances between embeddings for Resnet50 trained on all 20 H.erato and H.melpomene subspecies (All_Net).

5 Discussion

5.1 Effectiveness of ArcFace in Discriminating Co-Mimic Subspecies under Bird and Butterfly Acuity

The embedding distances derived from AllNet demonstrate the efficacy of integrating an ArcFace classification head to enhance class discrimination. Despite their visual similarity, the average distance between embeddings of co-mimic subspecies aligns more closely with heterospecific embeddings than conspecific embeddings (Figure 2). The model effectively discriminates between classes by maximizing distances even when confronted with nearly identical wing patterns among co-mimics. Notably, the model distinguishes co-mimics as distinct classes with minimal confusion, as evidenced by the larger average distance between co-mimic embeddings when compared to conspecifics (Figure 2). Additionally, the well-defined image embedding clusters in t-SNE visualizations (Figure 3) show a clear learned representation of our subspecies classes by AllNet.

Our findings align with prior observations made by [5], showing that deep learning can discriminate well enough between visually similar classes in the context of mimicry. We find that when training AllNet on all 20 subspecies, the distance between embeddings of co-mimics is more akin to that of heterospecifics (Figure 2) due to the inclusion of an ArcFace classification head in our model.

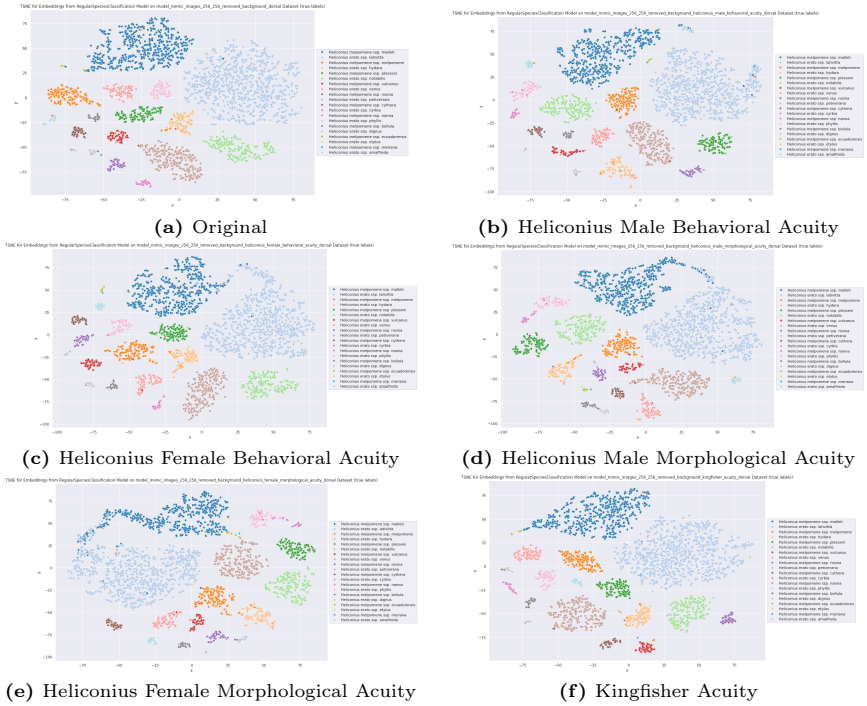
5.2 Acuity Impacts Trait Localization

Higher resolution images, such as those with kingfisher acuity and no acuity adjustments, led to improved species classification performance for AllNet (Table 2). Conversely, in the EratoNet and MelpomeneNet models, which do not simultaneously view both members of a co-mimic pair, models trained on images processed with butterfly visual acuity (lower resolution) outperform those using unprocessed images and images processed with kingfisher acuity. GradCAM results for EratoNet (Supplementary text, Figures 10, 11) reveal that under butterfly visual acuities, the model relies on larger, less detailed wing sections for classification. In contrast, for original images and those processed with kingfisher visual acuity, the model focuses on finer details and performs more precise trait localization. Similar patterns are observed in AllNet and MelpomeneNet (Supplementary text, Figures 9, 8, 12, 13).

T-SNE embedding projections for both EratoNet and MelpomeneNet (Supplementary text, Figures 6, 7) show that there is still a clear distinction made between model species (seen at train time) and mimic species (unseen at train time). Our GradCAM results reveal insight that although the same goal of species classification is achieved, the information that the model leverages to do so differs by acuity.

6 Conclusion

Our experiments show that deep learning can effectively learn how to differentiate visually similar co-mimic classes under bird and butterfly acuities when



(a) Original **(b)** Heliconius Male Behavioral Acuity

(c) Heliconius Female Behavioral Acuity **(d)** Heliconius Male Morphological Acuity

(e) Heliconius Female Morphological Acuity **(f)** Kingfisher Acuity

Fig. 3: T-SNE plots of image embeddings generated by AllNet under different acuities.

provided with all 20 subspecies classes at training time (AllNet). The results from EratoNet and MelpomeneNet show that which species you train on impacts the model's ability to recognize and differentiate the model species from the mimic species. The role of visual acuity was varied across different butterfly visual acuity estimates, but generally robust for Kingfisher acuity. Kingfisher acuity models consistently demonstrated a high macro accuracy and a large drop in performance when tested on the unseen mimic for the EratoNet and MelpomeneNet models, suggesting that the enhanced resolution of images via bird acuity is particularly beneficial for co-mimic differentiation. Results from the butterfly acuity model variants across experiments (AllNet, EratoNet, MelpomeneNet) demonstrate that species classification remains feasible, although the models leverage different wing sections for information to achieve this.

In this set of experiments, visual acuity showed to be advantageous for co-mimic differentiation to bird predators. As [3] have noted, species classification in Heliconius butterflies is believed to be multimodal. Thus, it is possible that the role of visual acuity for conspecific recognition in Heliconius butterflies is not as influential as visual cues in ultraviolet-sensitive vision systems (UVS), flight patterns, or chemosensory information.

Acknowledgements

The Imageomics Institute is funded by the US National Science Foundation's Harnessing the Data Revolution (HDR) program under Award #2118240 (Imageomics: A New Frontier of Biological Information Powered by Knowledge-Guided Machine Learning). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

1. Érika de Castro, Demirtas, R., Orteu, A.F., Olsen, C.E., Motawie, S., Cardoso, M., Zagrobelny, M., Bak, S.: The dynamics of cyanide defences in the life cycle of an aposematic butterfly: Biosynthesis versus sequestration. Insect Biochemistry and Molecular Biology **116**, 103259 (2019)
2. Caves, E.M., Johnsen, S.: AcuityView: An R package for portraying the effects of visual acuity on scenes observed by an animal. Methods Ecol Evol **9**, 793–797 (2018). https://doi.org/10.1111/2041-210X.12911, https://doi.org/10.1111/2041-210X.12911
3. Dell'Aglio, D.D., Troscianko, J., McMillan, W.O., Stevens, M., Jiggins, C.D.: The appearance of mimetic *Heliconius* butterflies to predators and conspecifics. Evolution; international journal of organic evolution **72**(10), 2156–2166 (2018). https://doi.org/10.1111/evo.13583, https://doi.org/10.1111/evo.13583
4. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4690–4699. The Computer Vision Foundation (2019)
5. Hoyal Cuthill, J.F., Guttenberg, N., Ledger, S., Crowther, R., Huertas, B.: Deep learning on butterfly phenotypes tests evolution's oldest mathematical model. Sci Adv **5**(8), eaaw4967 (Aug 2019). https://doi.org/10.1126/sciadv.aaw4967
6. Lawrence, C., Campolongo, E.G.: Heliconius collection (cambridge butterfly). Hugging Face (2024), https://huggingface.co/datasets/imageomics/Heliconius-Collection_Cambridge-Butterfly
7. Page, E., Queste, L.M., Rosser, N., Salazar, P.A., Nadeau, N.J., Mallet, J., Srygley, R.B., McMillan, W.O., Dasmahapatra, K.K.: Pervasive mimicry in flight behavior among aposematic butterflies. Proceedings of the National Academy of Sciences of the United States of America **121**(11), e2300886121 (2024). https://doi.org/10.1073/pnas.2300886121, https://doi.org/10.1073/pnas.2300886121
8. Wright, D.S., Manel, A.N., Guachamin-Rosero, M., Chamba-Vaca, P., Bacquet, C.N., Merrill, R.M.: Quantifying visual acuity in *Heliconius* butterflies. Biology Letters **19**(12), 20230476 (2023). https://doi.org/10.1098/rsbl.2023.0476, https://royalsocietypublishing.org/doi/abs/10.1098/rsbl.2023.0476