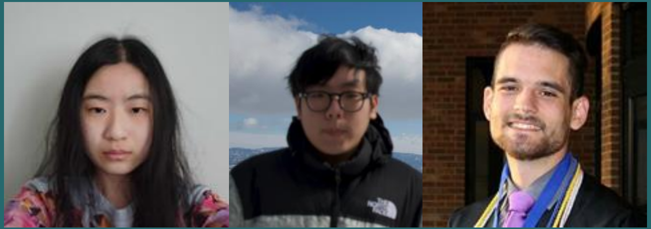
Anomaly Detection: Hybrid Butterflies

Elizabeth G. Campolongo

and

The Imageomics ML Challenge Team

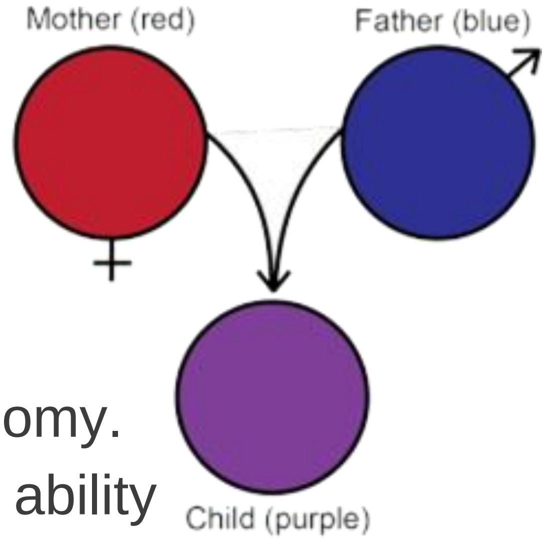


Special thanks to Lisa Wu, Ziheng Zhang, David Carlyn

Hybrid Detection

A brief history

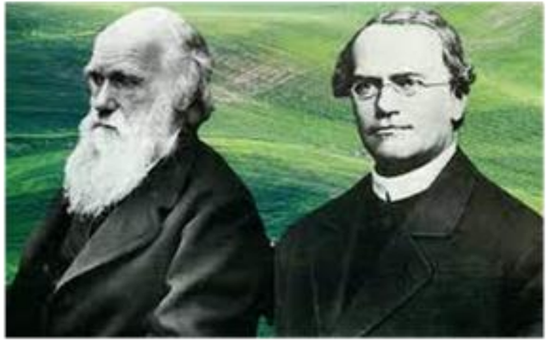
Hybrid Detection



- Researchers have sought a means to detect hybrids since the creation of the field of taxonomy.
- Detecting hybrids would give taxonomists the ability to determine what constitutes a true *species* or *subspecies.*
- The question is ***how***?
  - *How* do we recognize a hybrid?
  - What does a hybrid look like?

Hybrid Detection: History

- Darwin first posed this question of “What does a hybrid look like?”
- Mendel answered with his pea plant experiment.



Hybrid Detection: History

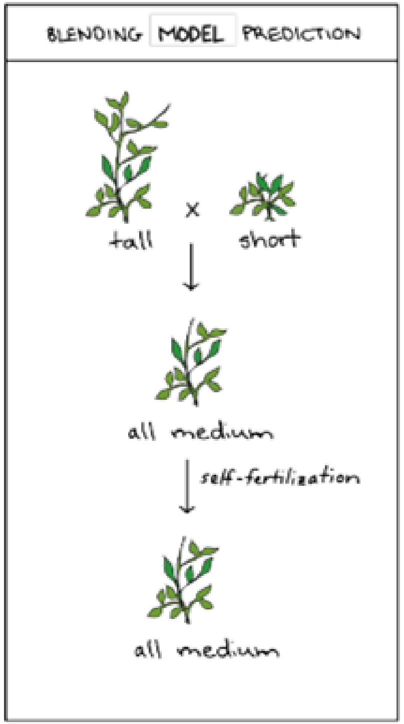
Mendel’s *Hypothesis*:

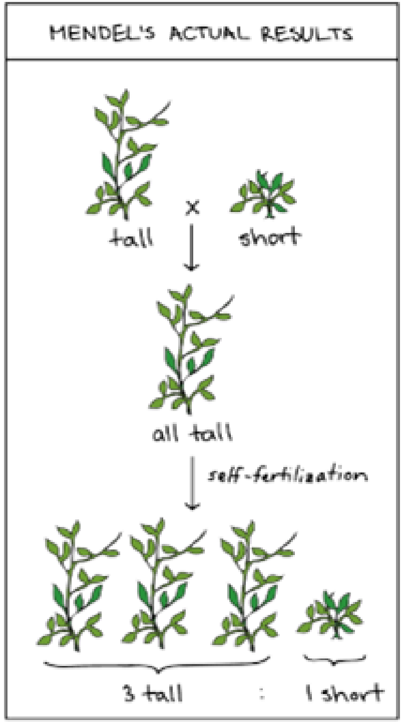
Blending Inheritance

- Inheritance of traits is ***continuous***.

Mendel’s *Results*:

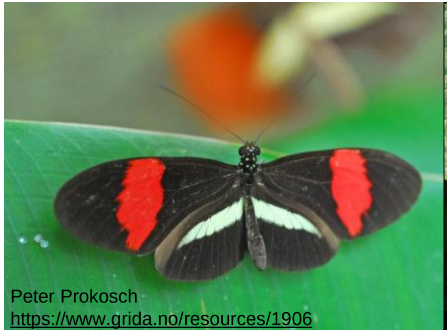
Inheritance is often ***discrete***.



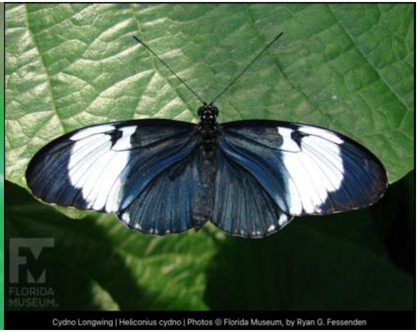


Hybrid Detection: Butterflies

- Consider these two species:
- Hybridization may lead to a variety of resulting patterns.
- There are several [dominant] genes that control color pattern on wings.
  - Ex: red on hindwings is a dominant trait.
- Dominance: hybrids may look like one parent.
- In practice, identifying hybrids requires knowledge of their parent species/subspecies.

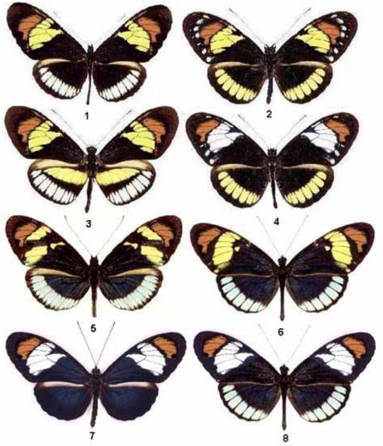


Peter Prokosch
https://www.grida.no/resources/1906



Cydno Longwing | Heliconius cydno | Photos © Florida Museum, by Ryan G. Fessenden

Hybrids between H. cydno and H. melpomene from Colombia
by Luis M. Constantino



Anomaly Detection

A brief history

Anomaly Detection: History

- Early topics include banking.
    - Detecting fraudulent or irregular spending or requests.
- In Machine Learning (ML), questions on classification:
    - Is the object a new one that the classifier has not seen?
- In Computer Vision (CV), questions for autonomous vehicles:
    - Is that a pedestrian or a deer that just ran into the road?

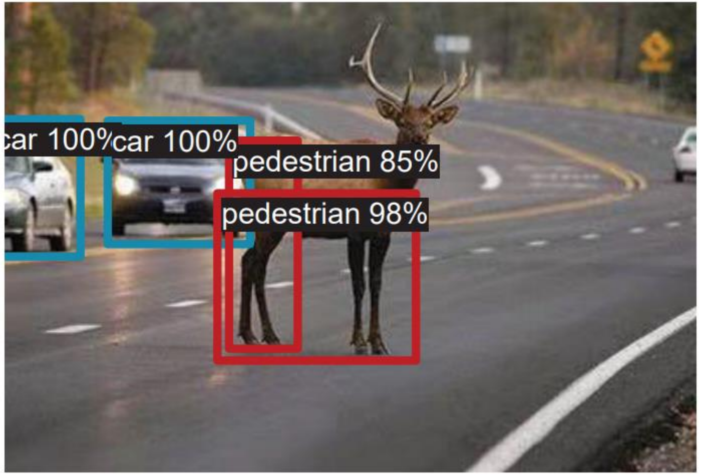
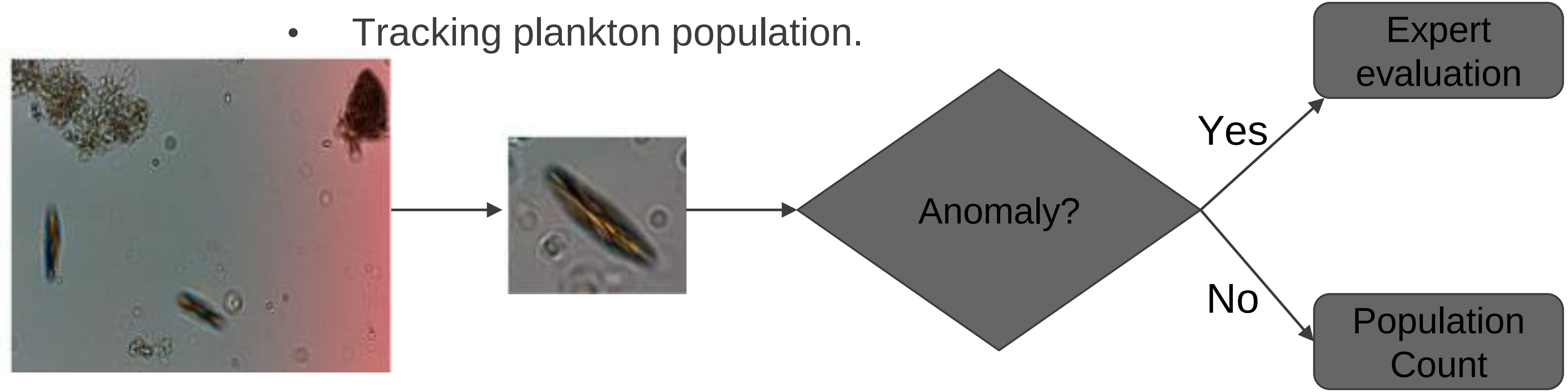


Figure: Xuefeng Du, Xin Wang, Gabriel Gozum, and Yixuan Li. "Unknown-aware object detection: Learning what you don't know from videos in the wild." *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13678-13688. 2022.

Anomaly Detection: History

- In Biology, questions on:
  - Gene function identification [1]
    - What phenotype anomaly resulted from a gene knockout?
  - Ecosystem health monitoring [2]
    - Tracking plankton population.



Anomaly?

Yes

No

Expert evaluation

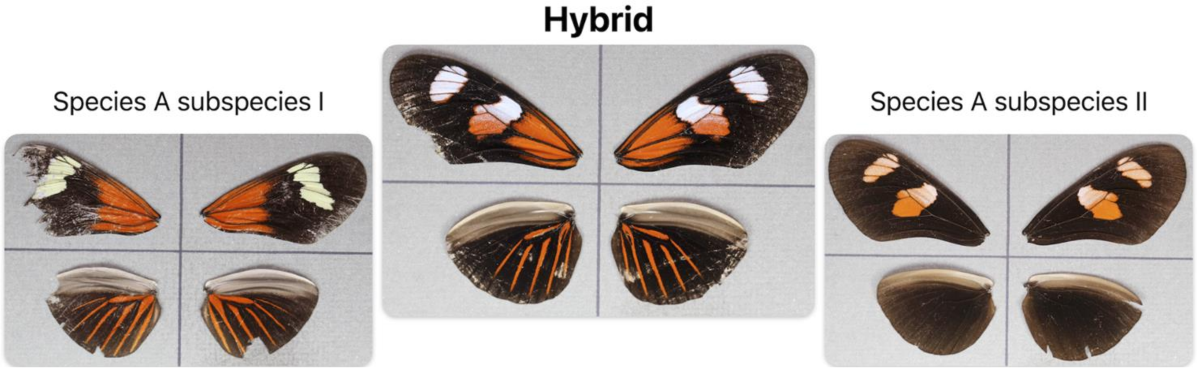
Population Count

[1] Ito, E. et al. (2022). Phenotype Anomaly Detection for Biological Dynamics Data Using a Deep Generative Model. In: Pimenidis, E., Angelov, P., Jayne, C., Papaleonidas, A., Aydin, M. (eds) Artificial Neural Networks and Machine Learning – ICANN 2022. ICANN 2022. Lecture Notes in Computer Science, vol 13530. Springer, Cham. https://doi.org/10.1007/978-3-031-15931-2_36

[2] Pastore, V.P., Zimmerman, T.G., Biswas, S.K. et al. Annotation-free learning of plankton for classification and anomaly detection. Sci Rep 10, 12142 (2020). https://doi.org/10.1038/s41598-020-68662-3

Our Challenge

How ***you*** can contribute to answering this important biological question



Images are from Zenodo records 2714333, 3082688, 2677821, 2686762, and 2549524, and are licensed under CC-BY 4.0.

Images are from Zenodo records 2714333, 3082688, 2677821, and 2686762, and are licensed under CC-BY 4.0.

Images are from Zenodo records 2714333 and 3082688, licensed under CC-BY 4.0.

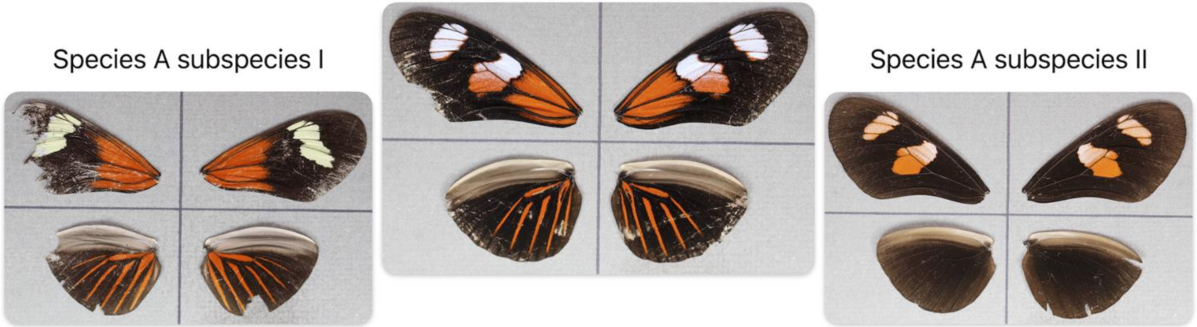
Hybrid graphic generated using Canva Magic Media [illegible] anually edited.

Our Challenge: Training Data

Training Data

- ~2200 images of Species A:
  - Multiple ***sub***species.
  - Selected signal hybrids of two ***sub***species.

Signal Hybrid



Images are from Zenodo records 2714333 and 3082688, licensed under CC-BY 4.0.

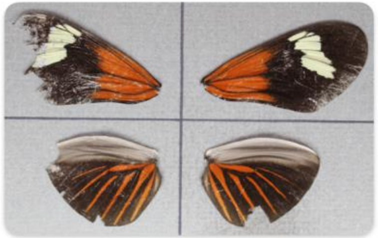
Our Challenge: Dev & Test Data

- Includes:
  - All Species A subspecies.
  - Signal hybrids from training data.
- Further introduces:
  - Other Species A hybrids (non-signal).
  - Species B: Mimics of Species A signal hybrid parents (& their hybrids).
- The numbers:
  - Validation Data (Dev): ~1100 images
  - Test Data: ~2200 images

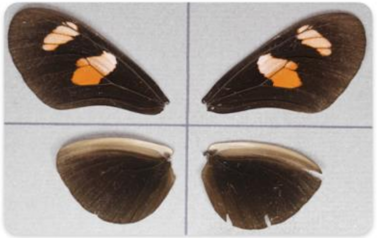
The Challenge: Find the Hybrids

- Among Species A & B, can your algorithm find…
  - Species A signal hybrids?
  - Species A non-signal hybrids?
  - Species B hybrids (mimics of Species A signal hybrids)?

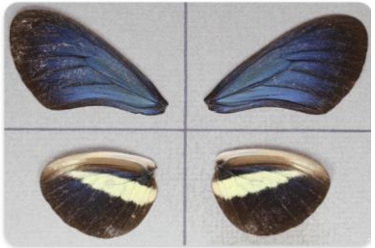
Species A subspecies I



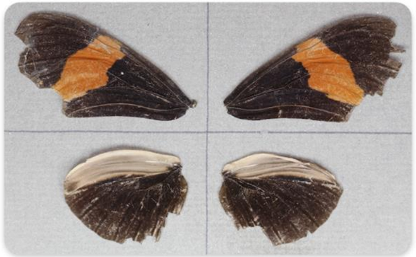
Species A subspecies II



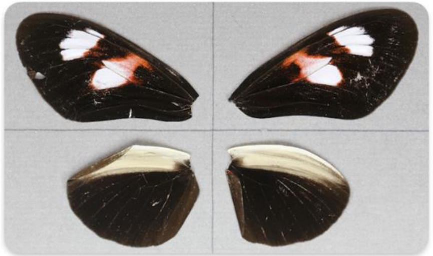
Species A subspecies III



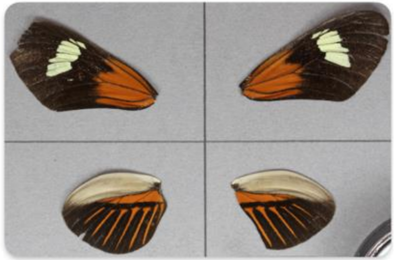
Species A subspecies IV



Species B subspecies II

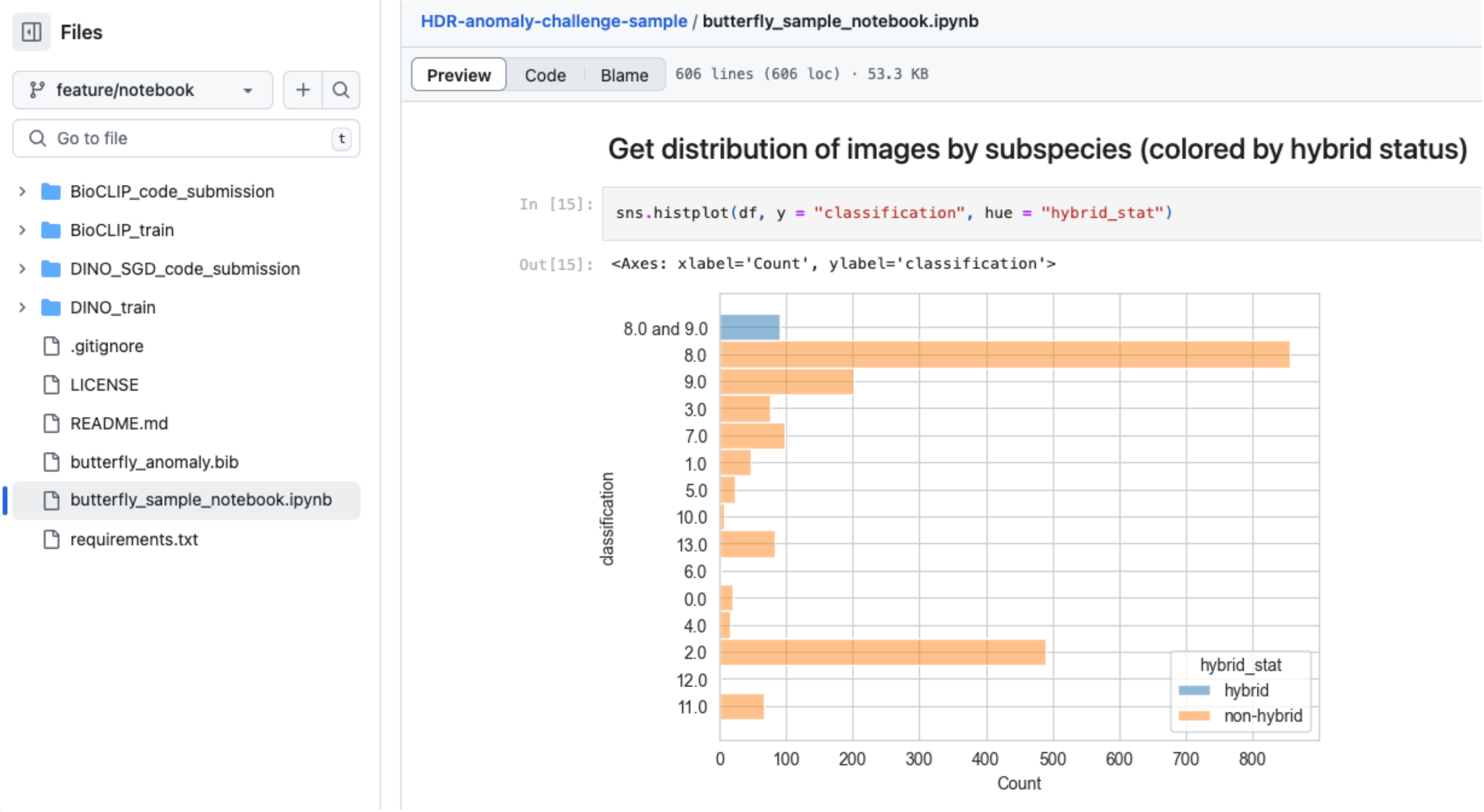


Species B subspecies I



Images are from Zenodo records 2714333, 3082688, 2677821, 2686762, and 2549524, and are licensed under CC-BY 4.0.

Sample Submissions Repository



Get distribution of images by subspecies (colored by hybrid status)

```
In [15]: sns.histplot(df, y = "classification", hue = "hybrid_stat")
```

```
Out[15]: <Axes: xlabel='Count', ylabel='classification'>
```

Join the Challenge!

Thank you!

Questions?