Using Spatiotemporal Features for Butterfly Classification

MARTA SKRETA, SASHA LUCCIONI, DAVID ROLNICK

Climate Change and Butterflies

BUTTERFLIES

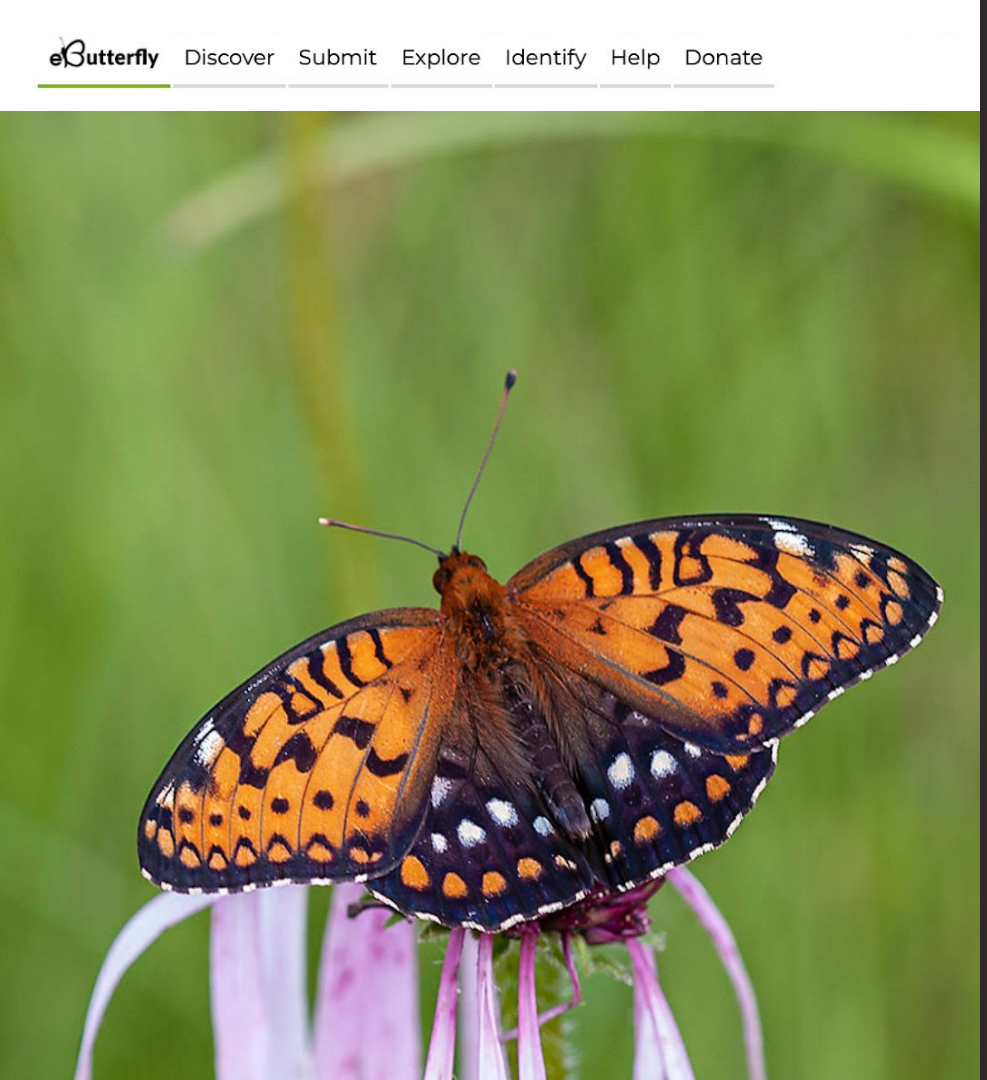
Temperature/weather impact

Indirect via habitat loss

ECOSYSTEM

Predators of butterflies/caterpillars

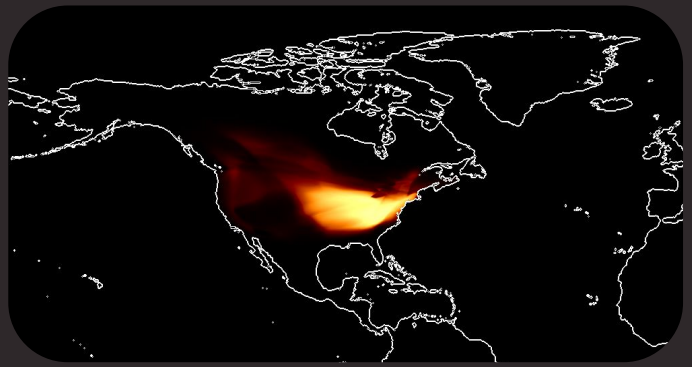
Plants that butterflies pollinate



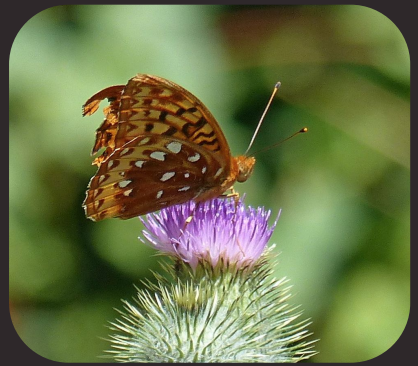
eButterfly project

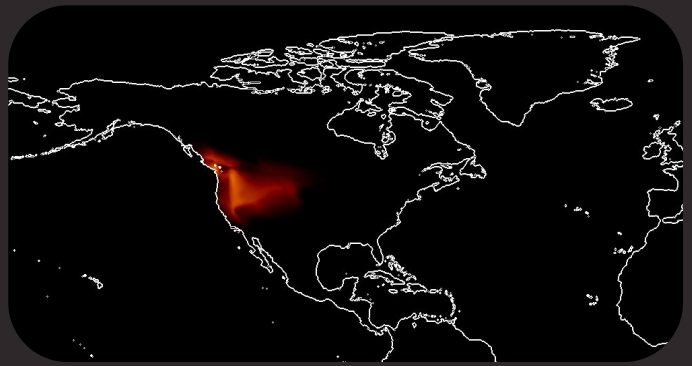
- **> 400,000 observations** in North America by **citizen scientists**
- **> 600 species**
- **Difficult to label** images by hand
- **Machine learning** can be useful

S. cybele



S. zerene





Can we use **WHERE** and **WHEN** the image was taken to improve classification?

Related work

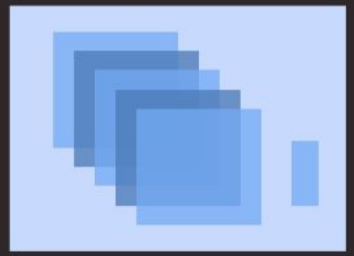
- Networks **trained** on images and geocoordinates **together**[1]
  - Assumption that test sample has location
  - Can't learn from spatiotemporal information that doesn't have image
- **Bayesian approach:**

  Train image and spatiotemporal **models separately**, **combine them at test time**[2]
  - Successfully used to classify birds & other animals
- Image-only classifiers have been built for butterfly identification[3]

[1] Chu et al. Geo-aware networks for fine-grained recognition. ICCV 2019
[2] Aodha et al. Presence-only geographical priors for fine-grained image classification. ICCV 2019
[3] Kantor et al. Guided attention for fine-grained and hierarchical classification. 2020

TRAIN TIME

Image, **I**



ResNet-50 CNN

$P(y|\mathbf{I})$

`(lat, lon, date)`

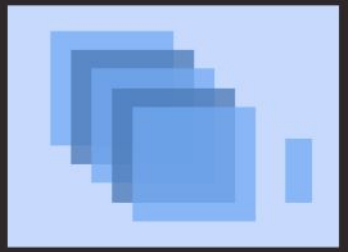
Spatiotemporal features, **x**

[sin(**x**), cos(**x**)]

Spatiotemporal encoder

$P(y|\mathbf{x})$

TEST TIME



$P(y|I)P(y|x) \propto P(y|I,x)$

(lat, lon, date)

[sin(x), cos(x)]

Accuracy	Image only	**Image + (Lat, Lon, Date)**
Top 1, Micro	84.56	**86.53**
Top 1, Macro	59.87	**65.65**
Top 3, Micro	93.84	**95.38**
Top 3, Macro	77.53	**83.74**

Micro accuracy: total correct/total number samples

Macro accuracy: average of species accuracies

Data augmentation

- Dataset is **imbalanced**:
  - > 400 species have < 100 observations
  - < 200 species have up to 2700 observations

- We use **iNaturalist** to **increase rare species** representation
  - Sample from iNaturalist until **each species has 100** observations

Accuracy	eButterfly	eButterfly + iNat	**eButterfly + iNat + (Lat, Lon, Date)**
Top 1, Micro	84.56	84.94	**87.90**
Top 1, Macro	59.87	69.51	**75.73**
Top 3, Micro	93.84	93.94	**95.86**
Top 3, Macro	77.53	83.59	**89.38**

Micro accuracy: total correct/total number samples

Macro accuracy: average of species accuracies

Conclusion & Future Work

- Using **spatiotemporal features improves** classification
- **Augmenting rare species** increases macro accuracy
- Working on improving geo model & testing on other species
- **Model is being deployed** on eButterfly website

martaskreta@cs.toronto.edu