Big Bee: Hair Recognition & Quantification

Harper Klauke, Jonathan Wong-Lau, Harleen Kaur, and Nicholas Alexander Project Sponsor: Dr. Katja Seltmann , Mentor: Joshua Bang

UC SANTA BARBARA
Cheadle Center for Biodiversity and Ecological Restoration

Background

The **Cheadle Center for Biodiversity and Ecological Restoration (CCBER)** is working with our capstone team to continue the research and understanding of bee ecology through image and trait digitalization.

Bee species are **decreasing** in both number and diversity. As important pollinators, bees are responsible for creating and maintaining the ecosystems that many animals and humans rely on. Examining their physical traits and the differences between species will allow us to study the resiliency of different bees using computer vision and machine learning.

Develop Python scripts that can extract EXIF data from an image file and rename the images by reading the QR/Datamatrix

Create different models that can quantify the number of hair on a given bee image

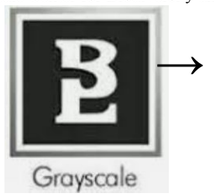
Extend our findings and methods to be accessible to partnering institutions and researchers

Data & Preprocessing

We worked with **550 high resolution lateral bee images** that were gathered from various institutions for the Bee Library, an online repository of bee image, trait, and specimen data.

Contributors: Museum of Comparative Zoology, Harvard University, University of Kansas Natural History Museum Entomology Division, UC Santa Barbara Cheadle Center for Biodiversity and Ecological Restoration

QR Code Scanner & MetaData Extractor
- Smaller sample of bee images that included barcode and specific metadata
- Using binarization to convert grayscale into black and white



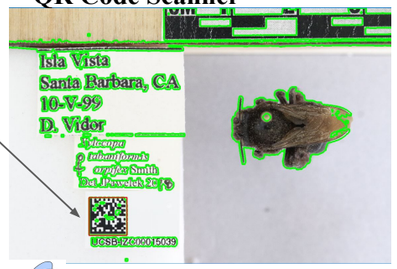
U-Net + Transfer Learning
- Manually created a bee binary mask dataset using Photoshop.
- Used data augmentation techniques to create a 300x300 cropped bee hair mask image dataset of the original bee images

Entropy Analysis
- Uses lateral images of bees
- Preprocessing
  - Removing pollen grains and other artifacts

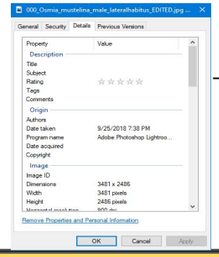
Results

QR Code Scanner

QR Boxed



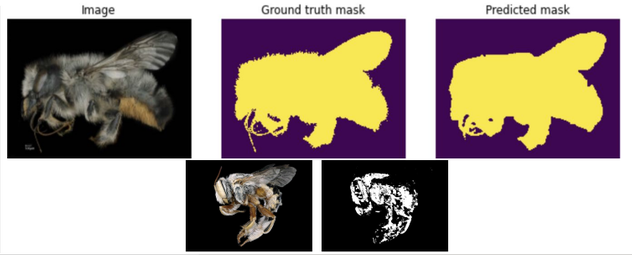
Metadata Extractor



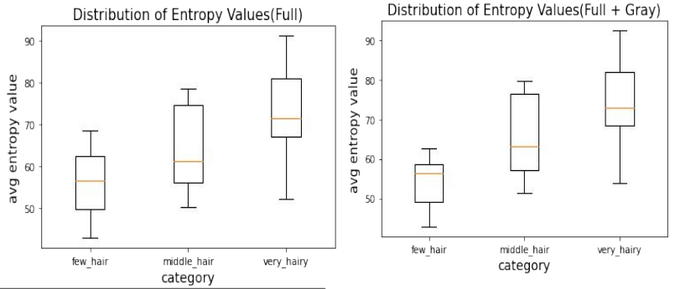
	B	C	D	E
1	FNumber	MaxApertu	DateTimeOriginal	DateTimeDigitized
2	f/2.4	NA	6/2/2021 11:08	6/2/2021 11:08
3	f/5.6	NA	10/28/2021 14:53	10/28/2021 14:53
4	f/5.6	NA	10/25/2021 12:41	10/25/2021 12:41
5	f/5.6	NA	10/25/2021 12:26	10/25/2021 12:26
6	f/5.6	NA	10/25/2021 12:59	10/25/2021 12:59

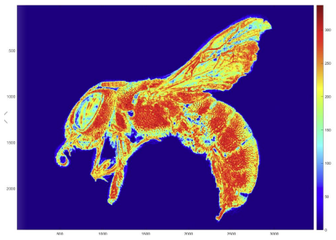
U-Net + Transfer Learning

Model	F1/ Accuracy
Bee Masking	97.68%
Hair Masking	87.2%



Entropy





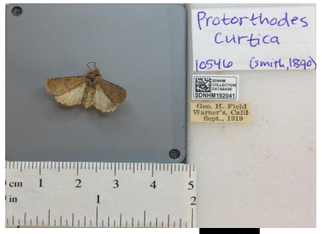
Methodologies

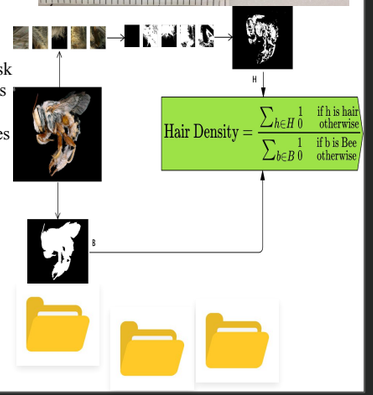
QR Code Scanner - Python Script
- Uses binarization, adaptive thresholding, and contours
- Input: Folder of images with QR/Datamatrix codes
- Output: Folder of images renamed by reading of decoded code
- Uses CV2, Pyzbar, Pylibdmtx

MetaData Extractor - Python Script
- Input: Folder of images containing metadata
- Output: CSV containing relevant exif data for each image

U-Net + Transfer Learning
- Used pre-trained U-Net for semantic segmentation task
- Hand crafted ground truth image set of hair/bee masks
- Takes a set of pre-trained weights from a similar task done on a different dataset and applies learned features our own dataset for further training
- Two separate models to quantify hair density

Entropy
- Per-pixel measure of local information variance
- Average entropy of three color values (RGB)
- Works best when images have little to no shine
- Tested accuracy by binning 50 images
  - Script ran for full and cropped images
  - Split into few hair, middle hair, and high hair
  - Comparing entropy values within the bins





Conclusions

The QR Code Scanner and MetaData Extractors are published on GitHub and are currently used by two academic institutions
- Scanner works best for images with a contrasting background to the QR code

Difficult to obtain a ground truth for hairiness
- The manual process to create a hair mask dataset is extremely time consuming and it would be more efficient to automate this process

Entropy algorithm is most accurate with specimens that have low reflectivity and minimal skin texture (e.g. pores)
- One possible route: Without binning, run the entropy script, then analyze the images with similar entropy values
- Binning images based on hairiness & used as ground-truth may lead to misclassification

References & Acknowledgements

This research was supported by the National Science Foundation: Extending Anthophila research through image and trait digitization (Big-Bee) project (DBI2102006).

Stavert JR, Liñán-Cembrano G, Beggs JR, Howlett BG, Pattemore DE, Bartomeus I. 2016. Hairiness: the missing link between pollinators and pollination. *PeerJ* 4:e2779 https://doi.org/10.7717/peerj.2779

Vladimir Iglovikov, Alexey Shvets. 2018. TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation. *arXiv:1801.05746* https://doi.org/10.48550/arXiv.1801.05746

Special thanks to Professor Yon Visell and Dr. B.S Manjunath for providing recommendations for methodologies.