Betta Fish Evolutionary Morphology

Shijie He, CongCheng Yan, Chutian Chen, Chengchao Jin, Wancheng Chen
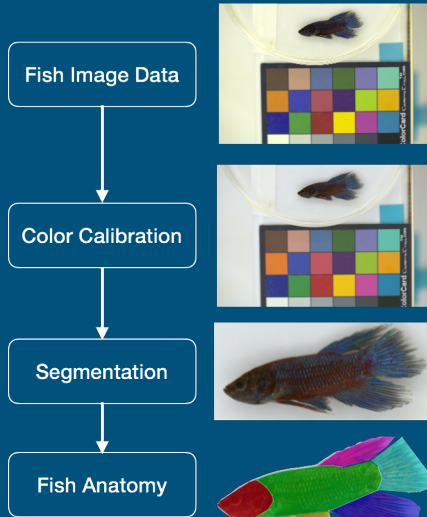
Content

1.Introduction

Introduction

Image Analysis

QTL Mapping

Fish Image Data

Color Calibration

Segmentation

Fish Anatomy

Genetic Data

Mapping

Color Percentage

Size Features

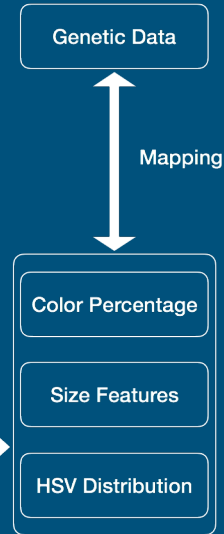HSV Distribution

Workflow of project

Goal: Figure out how evolution can affect the traits of Betta fish. The project is consisted of image analysis and genetic mapping.

Image Analysis:

1. **Read raw data (CR3 pictures)**
2. **Color calibration (self-designed algorithms)**
3. **Segmentation (self-designed algorithms)**
4. **Fish Anatomy and feature extraction (transfer learning)**

Genetic Mapping:

1. **Utilize the final output from image analysis part as input**
2. **Mapping genetic data with features of fish**

2. Data

Data

Data Description:
- The data of the project is Betta fish pictures. There are pure blue or red fishes and also many descendant fishes which mixed blue and red colors in different degree.
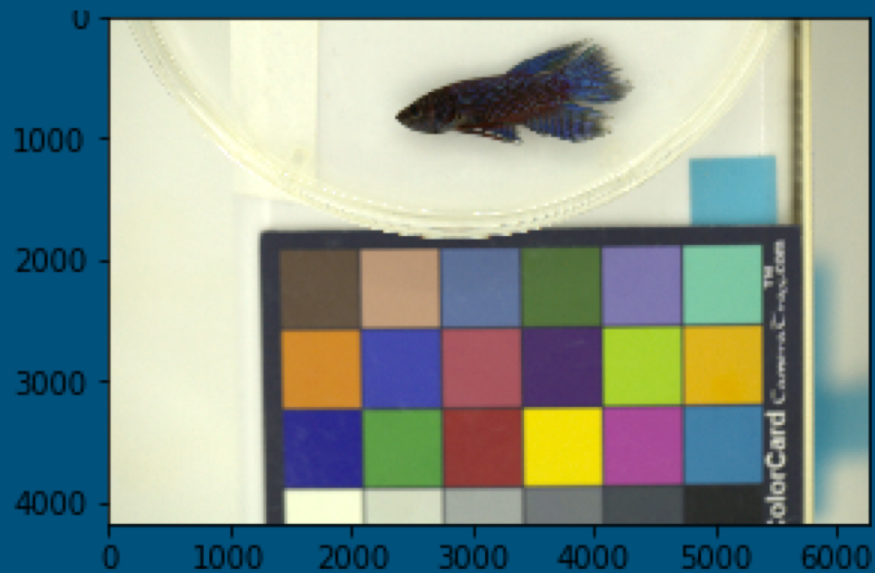- Obtained 331 fishes and 14,909 pictures in total

Data Input:
Remove the influence of the automatic adjustment to the photos, restore real color, make analysis more accurate
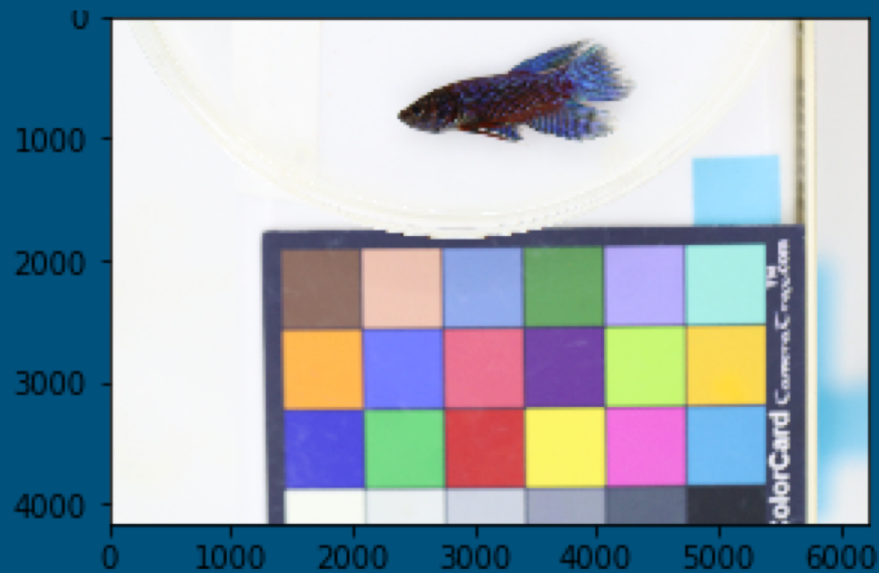
Methods:
- Use raw images (.CR3 images) instead of .JPG images
- Convert .CR3 to .tiff with package "Rawpy", transform the images into RGB data

Data

Raw Image

JPG Image

3. Image Analysis
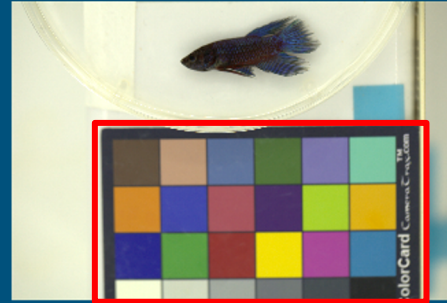
a. Color Calibration

Based on the color cards on the images, calibrate the color of the images according to the standard color card.
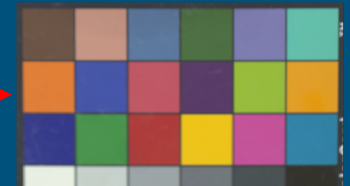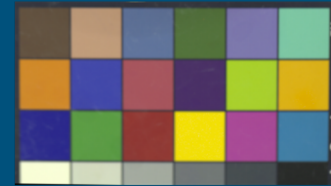
Steps:
1. Extract the color cards
2. Detect the color patches in the cards
3. Match the color patches to the standard color
4. Calibrate the color with machine learning methods (Partial least square regression)

* Use average PLS model to do calibration work for blurry images

**Before Calibration**

After Calibration

b. Fish Segmentation

Part 1: Whole Fish Cropping

Use OpenCV package to get all the contours and automatically select the fish contour based on contour length and cropped picture size.

Steps:
1. Read the images and convert to grey images
2. Get binary images according to different thresholds
3. Find the most complex contour for each threshold since the other objects in the image are rectangles
4. Set the output image size limit and get the best result, over 90% outputs are qualified
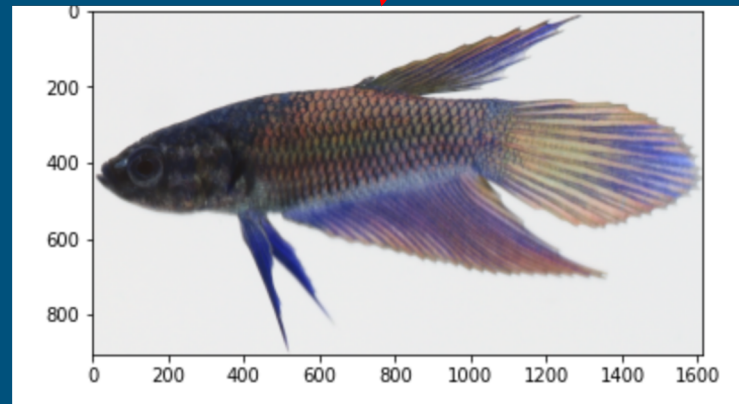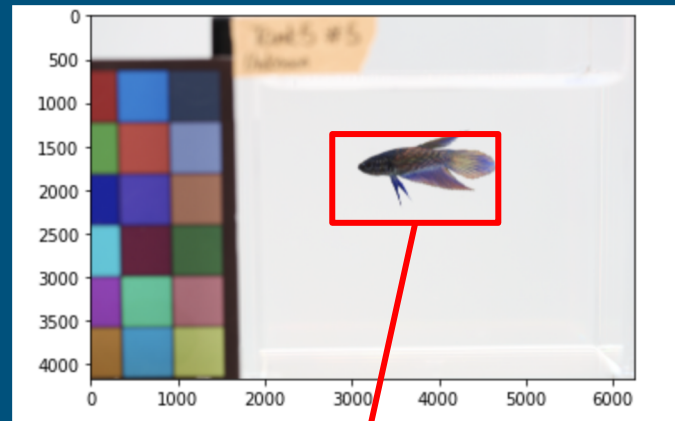
b. Fish Segmentation

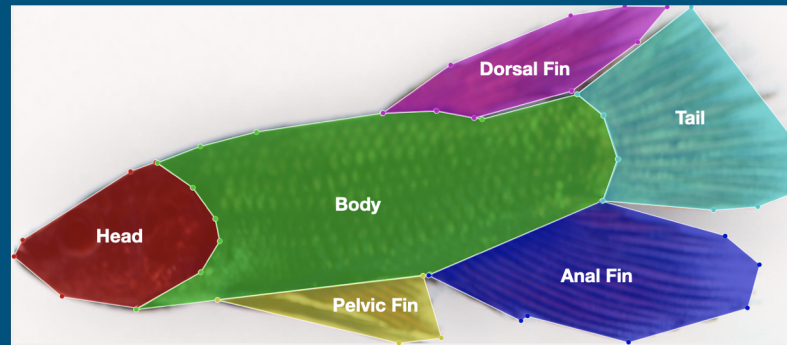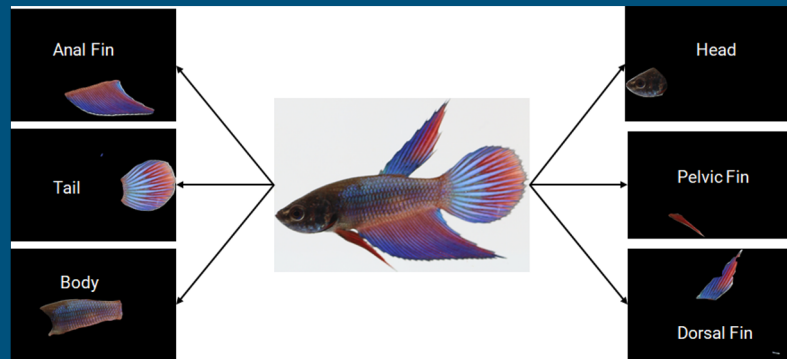## Part 2: Fish Anatomy

Label the images and train a model based on FCN Resnet. Do the segmentation automatically by the model and manually check the outputs.

Steps:
1. Label the images
2. Feed the images into the model and train it with data augmentation, f-score>0.9
3. Manually check the outputs of every fish and keep the best one for each

Labelled image

Model output

c. Feature Extraction

We extract 9 features from fish images.

Size:

1. Size percentage

Saturation Level:

2. Saturation level of whole fish
3. Red saturation level
4. Blue saturation level

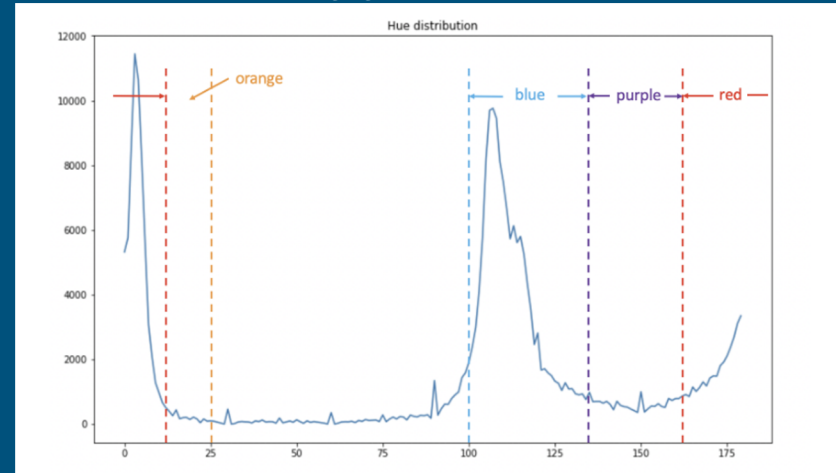**Color Percentage:**

5. Red percentage
6. Blue percentage
7. Purple percentage
8. Orange percentage
9. Dark percentage

Saturation Level

Hue Distribution

c. Feature Extraction

The final data frame look like:

	Index	Part	size_percent	whole_sat_level	red_percent	red_sat_level	blue_percent	blue_sat_level	purple_percent	orange_percent	dark_percent
0	149	body	0.247564	high	0.014360	low	0.532154	high	0.005681	0.038191	0.662371
1	149	dorsal_fin	0.223330	high	0.008278	low	0.798865	high	0.000645	0.004823	0.006377
2	149	anal_fin	0.008493	low	0.123951	low	0.447130	high	0.046788	0.090298	0.367084
3	149	pelvic_fin	0.034287	low	0.022757	low	0.022988	else	0.003511	0.131843	0.529483
4	149	head	0.038726	low	0.006487	very_low	0.018069	high	0.000890	0.026937	0.864057
...	...	...	...	...	...	...	...	...	...	...	...
505	150	head	0.092358	high	0.095792	high	0.015681	low	0.002121	0.202443	0.442001
506	150	body	0.346126	high	0.061871	low	0.421715	high	0.017837	0.194854	0.476735
507	150	dorsal_fin	0.069633	high	0.119692	high	0.761696	high	0.076944	0.005680	0.021441
508	150	pelvic_fin	0.004500	high	0.845182	high	0.021303	low	0.008174	0.050780	0.000000
509	150	tail	0.245970	low	0.204225	high	0.554919	high	0.098123	0.005086	0.000189

4. Genetic Mapping

Genetic Mapping

QTL mapping is to locate the gene that has the greatest impact on the specific phenotype.
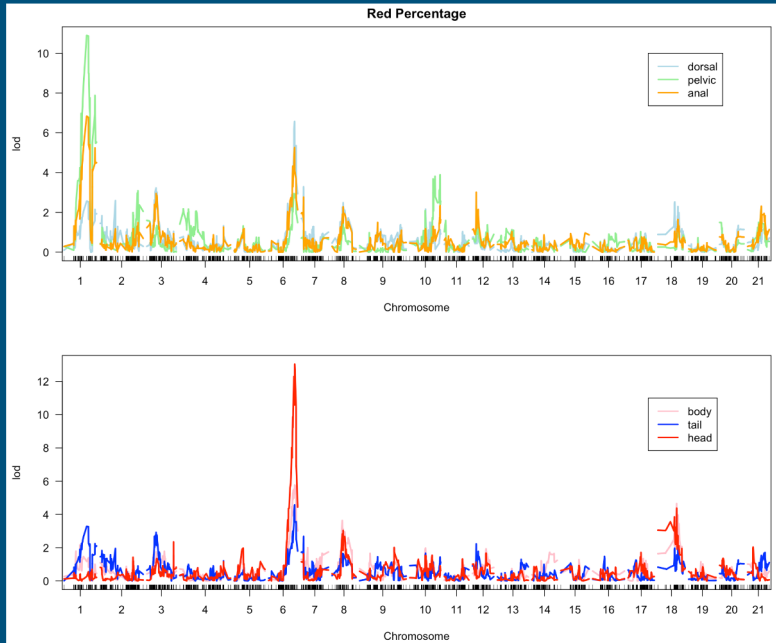
We use Haley-Knott regression to locate such gene. The marker with maximum LOD is the most dominant one for each individual phenotype.
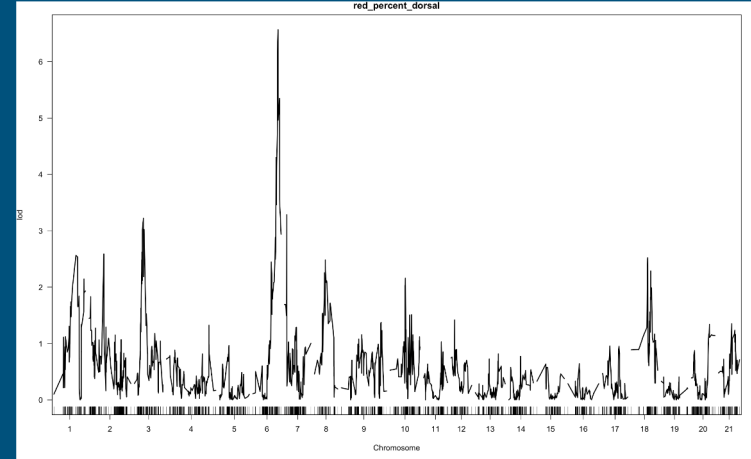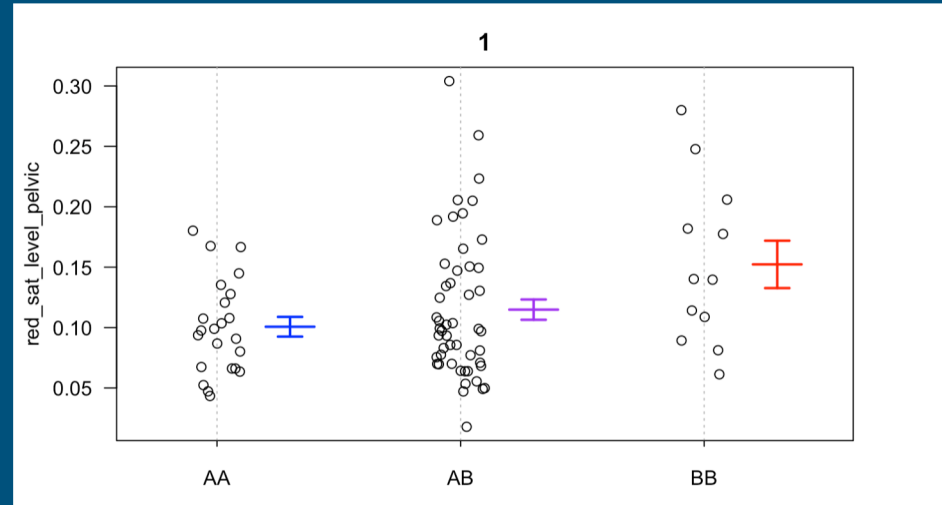

Feature: Red Percentage on Dorsal Fin

Genetic Mapping

Comparison of one phenotype on different body parts.

Function "plotPXG" to determine if gene is dominant, semi-dominant or recessive

5. Conclusion

Conclusion

What we have done

Develop a python-based model to read, calibrate, segment, analyze the fish images automatically

Based on the features extracted from images and the genetic data, run the QTL mapping between genotypes and phenotypes to locate genomic regions affecting the size and color of the fish
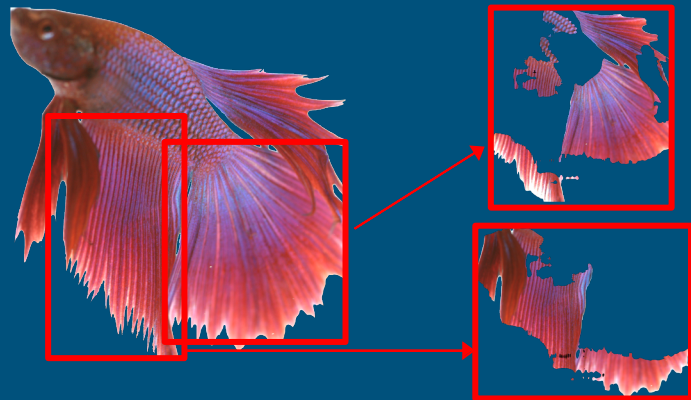
Problem

1. Poor performance of fish anatomy on fish with huge fins and tails

2. Hard to detect some parts of fish based on side view

3. Overlap and curling of fins and tails lead to inaccurate features

Problem and Future Work

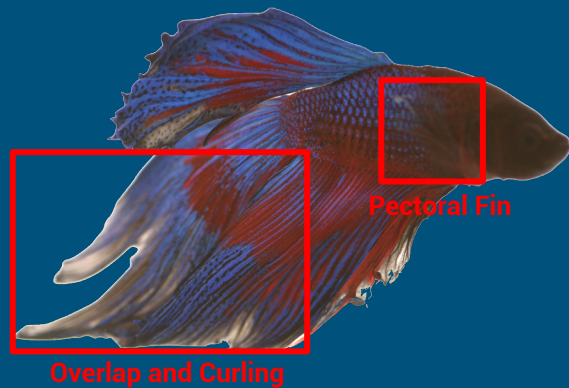**Problem 1: poor performance of fish anatomy on fish with huge fins and tails**

Future Work: fine tuning the neural network

Problem 2: hard to detect some parts of fish based on side view

Problem 3: overlap and curling of fins and tails lead to inaccurate features

Future Work: analyze the images from different angle (side view and top view) together

Pectoral Fin

Overlap and Curling

Thank you