Amy Tang & Ron Polonsky

Mentor: Yibo Zhang

Project Proposal: Swim Stroke Classification

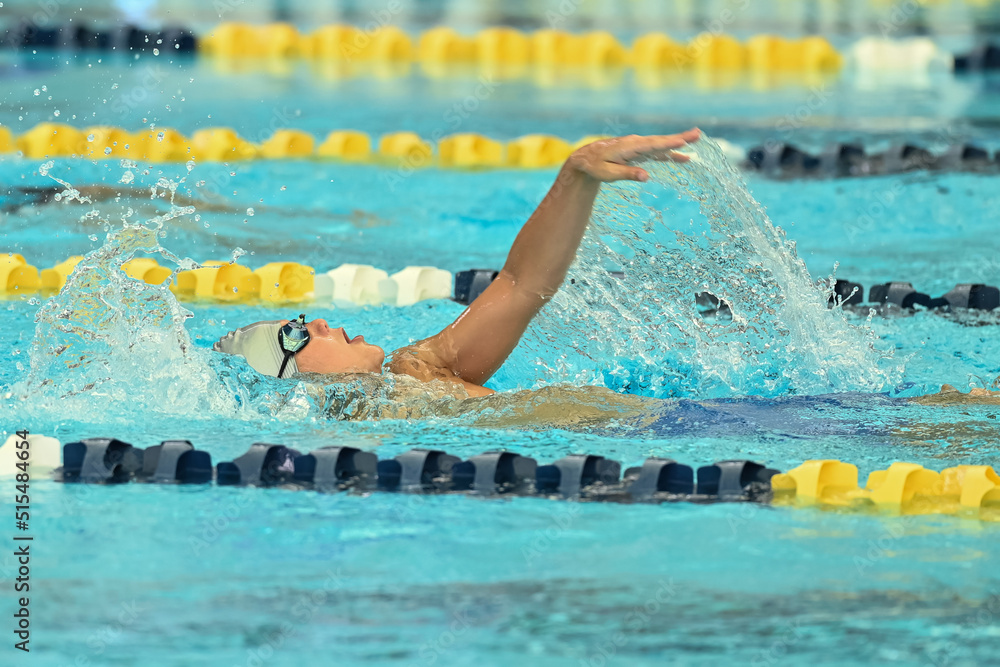
**Problem Statement and Task Definition:** Swimming is a very popular sport that many hundreds of thousands of young people participate in. However, the elite level of the sport is hard to reach and is often limited by financial means. Paying for private lessons costs hundreds of dollars per hour and professional stroke video analysis is even more expensive. Additionally, if only a few professionals are analyzing videos by hand, they can only help a small number of athletes.

The goal of our project is to input swimming images and output what swimming stroke is present in the image. This could be the foundation for a future project that could provide accessible stroke analysis software to everyone, reducing the financial barrier to swimming at the elite level.

**Input/Output Behavior:** The input variables are swimming images in either jpg or png format of one or multiple people in a race or practice performing a stroke. The output of my system will be a predicted label of which stroke is present in the image. The possible labels are butterfly, backstroke, breaststroke, and freestyle. In the case where there are multiple people doing different strokes, the system will predict the stroke of the person who takes up the largest space in the image.

**Concrete Examples**:

Input:



Output: Backstroke

Input:



Output: Freestyle

**Evaluation Metric:** We will evaluate the success of our system based on accuracy. We will measure accuracy as *number of correct predictions / number of total predictions*. We will also measure accuracy for each stroke to make sure our model is equally successful at predicting all strokes. Accuracy is the best metric for this problem because there is a simple right or wrong answer to each input, and each stroke class is distinct.

**Dataset:** We wrote a python script that uses bing\_image\_downloader to collect a dataset of 1000 swimming images with different strokes. We will use 80% of the dataset for training, 10% for validation, and 10% for testing.

**Related works:** Most of the previous related [projects](https://www.researchgate.net/publication/366691863_An_effective_swimming_stroke_recognition_system_utilizing_deep_learning_based_on_inertial_measurement_units) for analyzing swimming do not use images as their input. They have swimmers wear sensors on different parts of their bodies and use the numbers from the sensor to predict strokes. We only found one [other project](https://universe.roboflow.com/gecko-vision/swimming-strokes-detection) that analyzed swimming strokes using images. However, it was only of one person swimming in a pool the size of a hot tub, and it had a dataset of less than 200 images.

**Baseline:** Our baseline for our project will be a simple convolutional neural network trained from scratch on only our collected dataset without any pre-training on external datasets. This baseline provides a lower bound on performance because the model does not have access to millions of images of all types. It is trained on a relatively small dataset that is specific to swimming.

**Oracle:** Our oracle for this project will be human labelling. My partner and I will label images manually as an upper bound for accuracy. We should achieve virtually 100% accuracy.

**Methodology:** To solve our task, we will fine-tune a pre-trained CNN loaded from Pytorch. We will slightly adjust the model by changing the final layer to predict one of four strokes. Then we will train the model on our dataset and evaluate its success. We are considering using ResNet18 because it is small, fast, and easy to train while still being very accurate at image recognition.

**Challenges:** Some possible challenges include making sure our dataset is clean, pictures from different angles (front, back, side, overhead, underwater), and our model having higher accuracy for certain strokes than others. To address these challenges, we will manually delete images where it is hard for a human to tell what stroke is present or images that don’t belong in our dataset (non-swimming or cartoon-swimming images) and/or write a script to delete images and collect about an equal number of images from different strokes and camera angles.