

# Project Proposal

---

## Client:

The client is a Wildlife Conservation Organization.

---

## Question/Need:

The organization works to preserve lands home to biodiversity and sensitive wildlife populations. A large component of this process is understanding and tracking wildlife populations and behaviors.

The goal of this project is to develop a system for the organization to identify wildlife present.

---

## Data Description:

We will be using the [Open Images Dataset](#) by Google. The dataset is quite large so we will be limiting it to relevant labels for our wildlife detection. We will then using images as our feature set. We will use labels and bounding boxes as our targets.

---

## Solution Path:

We see our solution to this problem as two stage:

1. Create a general purpose model as a proof of concept.
    - We will use a [YOLO](#) like model for predicting our classes and boxes.
    - We will not be concerned with model speed or size here.
  2. Create a smaller specialize model.\*
    - Utilze First model as a teacher. Perform knowledge distilation for training the smaller model. See [Keras implementation](#) as a resource.
- **Feature Engineering**
    - Images will need to be resized to a smaller resolution.
    - Images may be augmented to help generalization to images in the wild.
  - **Model**
    - YOLO\*
    - Faster rCNN
    - rFCN

\*See Constraints

\*\*Leading option

---

## Criteria for Success:

- Our model is able to detect the presence of wildlife in naturalistic images.
- Our model is lightweight enough to run on a small device.\*

\*See Constraints

---

## **Assumptions, Risks and Constraints:**

The organization will need to gather data from remote regions where access to the internet may be unavailable. The will need to gather data from large areas requiring many monitoring points. This scale will make it infeasible for the organization to deploy stations with network capabilities. Without the ability to upload images regularly storage will become an issue. The smaller model we can develop the easier it will be to deploy in specialize devices with limited compute. Rather than storing images, we will only need to store records of the wildlife scene.

Our data augmentation strategy will generalize to real life incoming documentation.

---

## **Tools:**

1. Keras
  2. Numpy
  3. OpenCV
  4. Gradio
  5. Albumentations
  6. Google Colab
- 

## **Communication:**

A slide presentation will be made explaining the model building process and showcasing the model utility.

A demo web app will be presented allowing for live demonstration of the tool.

---

## **MVP Goal:**

An MVP for this project will consist of a model generated to detect the presence of a small number of animals.