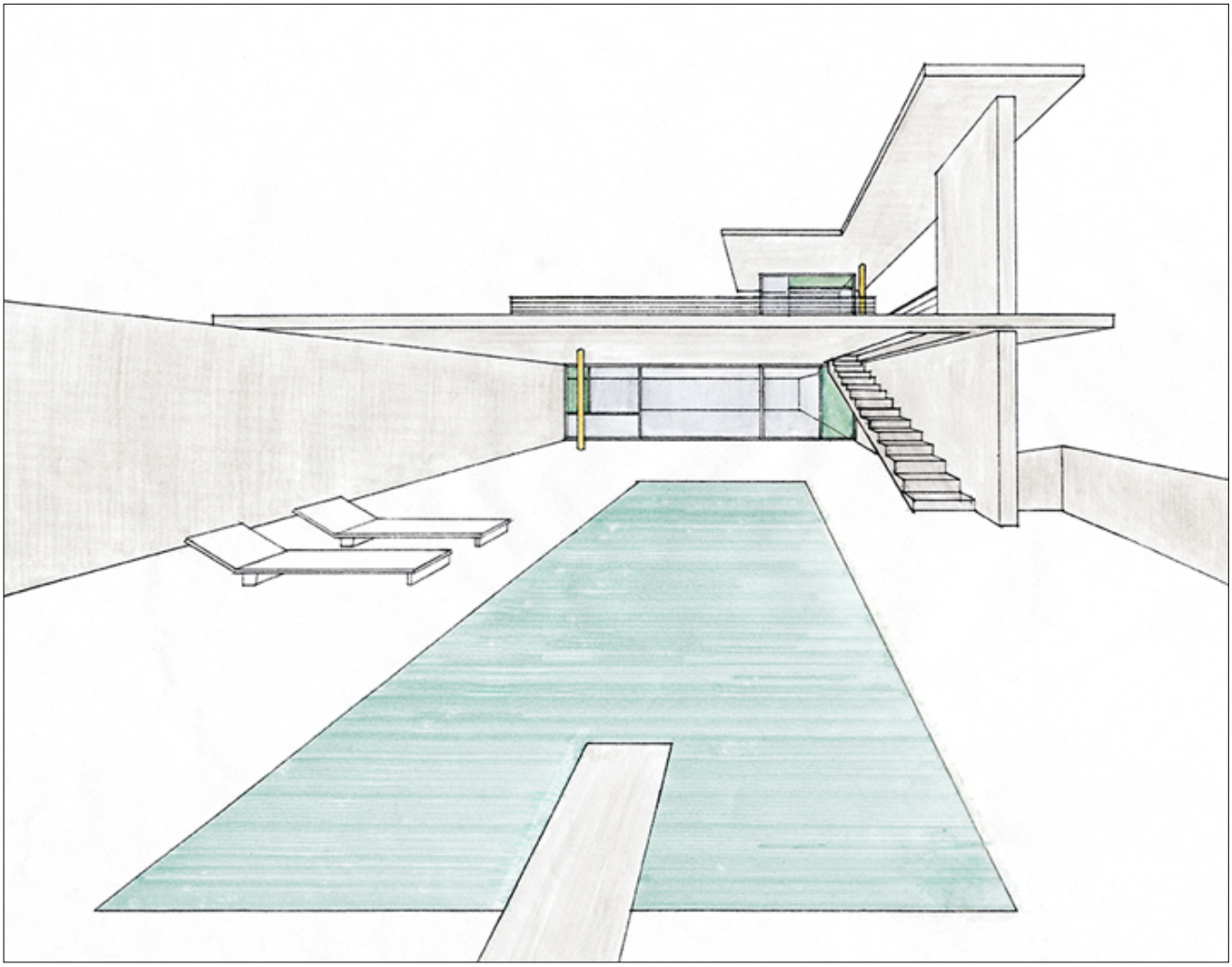
CAPSTONE 2



Project Proposals

Prepared for: Personal, Title

Prepared by: Ozkan Serttas, Job Title

February 13, 2019

Proposal number: 123-4567

CANDIDATE PROJECT 1:

Problem Description: Amazon Inventory Management

Items can be misplaced while being carried so that bin inventory does not match with the real inventory. Using image detection models could help identifying items and their counts in the bins so that potential mismatch can be detected.

The Data

Amazon uses a random storage scheme where items are placed into accessible bins with available space, so the contents of each bin are random, rather than organized by specific product types. Thus, each bin image may show only one type of product or a diverse range of products. Occasionally, items are misplaced while being handled, so the contents of some bin images may not match the recorded inventory of that bin.
The Amazon Bin Image Dataset contains over 500,000 images and metadata from bins of a pod in an operating Amazon Fulfillment Center. The bin images in this dataset are captured as robot units carry pods as part of normal Amazon Fulfillment Center operations. Each image/metadata tuple corresponds to a bin with products. The metadata shows the actual count of items in the bin that will be used as labels to training set.

Goals

- Find the best solution to the mismatch problem using bin images provided by Amazon so that pickers could find the item precisely.
- The output of the model should be a prediction of item counts in the bins.

Project Outline

Amazon Fulfillment Centers are bustling hubs of innovation that allow Amazon to deliver millions of products to over 100 countries worldwide with the help of robotic and computer vision technologies. The Amazon Bin Image Dataset contains images and metadata from bins of a pod in an operating Amazon Fulfillment Center. The bin images in this dataset are captured as robot units carry pods as part of normal Amazon Fulfillment Center operations.

CANDIDATE PROJECT 2:

Problem Description : Identifying Flower Species from Images

This project is to build an image classifier from scratch that will identify different species of flowers.

The Data

The data set contains images of flowers from 102 different species. We are provided a training set and a validation set total size of 297 mb.

Goals

- Train a model that will predict the flower species with high accuracy.

Project Outline

Image classification is one of the key research area in various fields as security, health care, transportation etc. We can think of many use cases such as identifying a criminal via street cams, unlocking personal devices, detecting cancer cells and detecting threats to a driver while driver and so on. In this project, we will use the same image classification techniques to identify different species of flowers by using the image data.

CANDIDATE PROJECT 3:

Problem Description : Maleria Detection test

Malaria is an infectious disease that causes over 400,000 deaths per year.
Malaria is a true endemic in some areas of the world, meaning that the disease is regularly found in the region. In other areas of the world, malaria is an epidemic — it's widespread in the area but not yet at endemic proportions. Yet in other areas of the world malaria is rarely, if ever, found at all.
So, what makes some areas of the world more susceptible to malaria while others are totally malaria free? There are many components that make an area susceptible to an infectious disease outbreak. The higher the poverty level, the higher the risk of infectious disease, although some researchers will say the opposite — that malaria causes poverty. Regions of the world that are below poverty levels most likely do not have access to proper healthcare. Without good healthcare, proper treatment, and if necessary, quarantine, infectious diseases can spread quickly. Testing a blood is tricky problem, there is high accuracy with a clinician but the speed of the process and the number of clinicians are not enough. Clinician basically counts the number of infected red blood cells. There is another practical way to do the testing by using a device called Rapid Diagnosis Testing but this is not accurate enough!

The Data

Malaria datasets is taken from NIH Communications Engineering Branch website which is a repository of segmented cells from the thin blood smear slide images from the Malaria Screener research activity. he dataset contains a total of 27,558 cell images with equal instances of parasitized and uninfected cells. An instance of how the patient-ID is encoded into the cell name is shown herewith: "P1" denotes the patient-ID for the cell labeled "C33P1thinF_IMG_20150619_114756a_cell_179.png". We have also included the CSV files containing the Patient-ID to cell mappings for the parasitized and uninfected classes. The CSV file for the parasitized class contains 151 patient-ID entries. The slide images for the parasitized patient-ID "C47P8thinOriginal" are read from two different microscope models (Olympus and Motif). The CSV file for the uninfected class contains 201 entries since the normal cells from the infected patients' slides also make it to the normal cell category (151+50 = 201).

Goals

- Count red blood cells that are infected.

Project Outline

Train a neural network that is computationally efficient (convolutional most used for images), to obtain the same level accuracy as NIH which is about 95 %. This model can be turned into a phone application as in the case of researches of NIH.