DERMATOLOGY TEAM 14

Dr. Ricardo Henao, Adrian Lopez, Brian Kim, Kinjal Sanghvi, Meng Xiao, Matt Engelhard

Agenda

- Problem Statement
- Dataset
- Solution Approach
- Active Learning
- Results
- Accomplishments & Going Forward

PROBLEM STATEMENT

The dermatology department at the Duke Hospital has **25,509** images of skin lesions. All these images are unlabelled and scale from “a very clear centered lesion” to an image of a full body with tiny lesions.

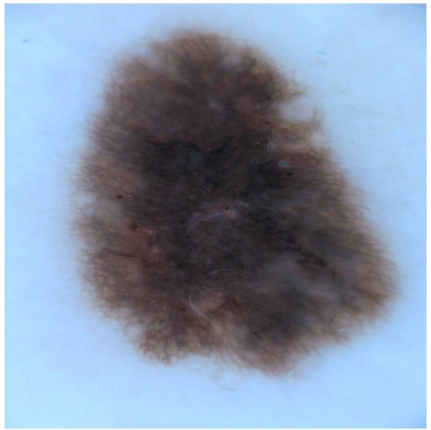
Our task is to train a model to create bounding boxes around each of these lesions and classify them.

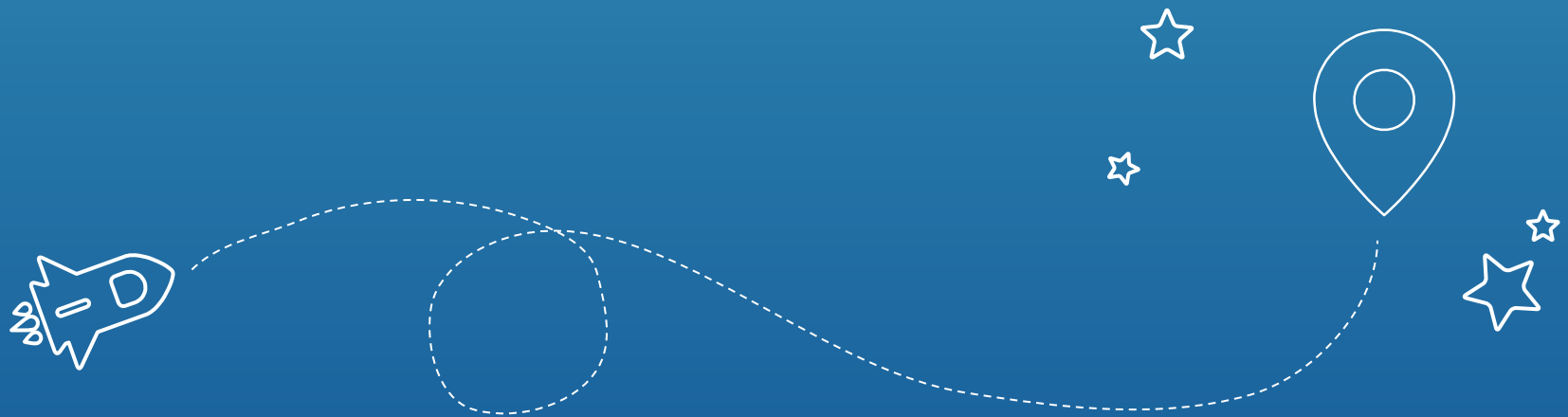
DATASETS

1. Our Dataset contains Duke Images as well as images from an online dataset (ISIC 2018).

2. To train our models, we use a combination of these images.

3. We have minimal number of ground truths to the 25,509 images provided by Duke Hospital. Therefore, we utilize Active Learning in training the model to maximize performance gain for each label.



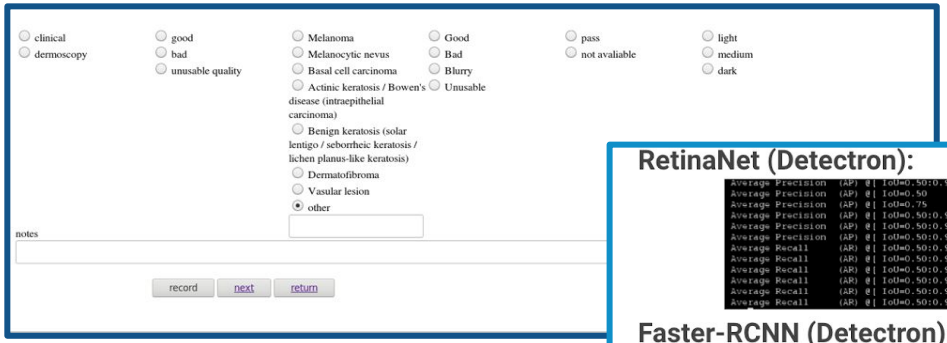


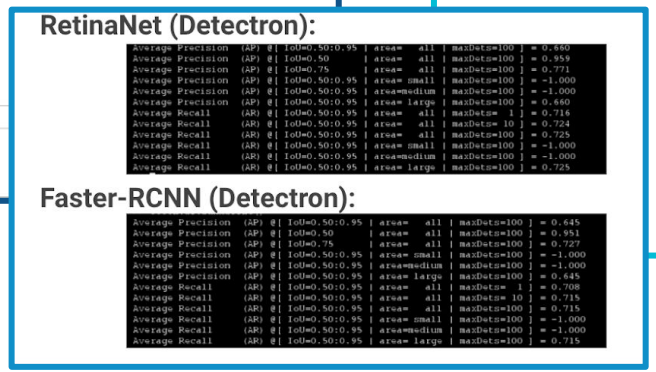
SOLUTION APPROACH

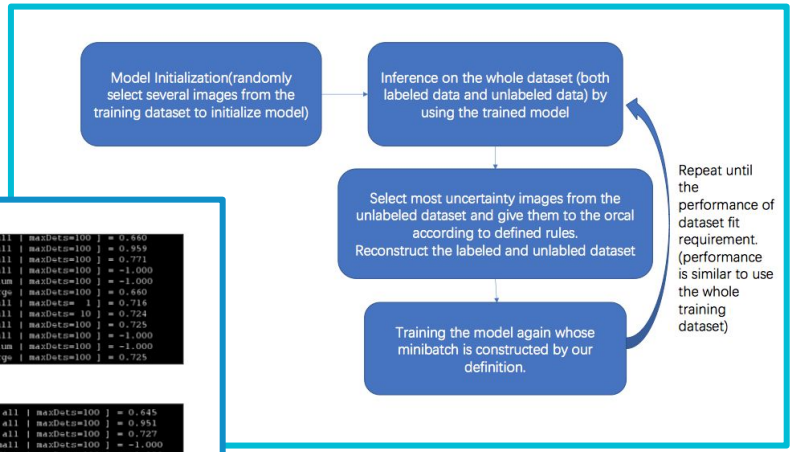
CREATE APPLICATION

TEST OBJECT DETECTION MODELS

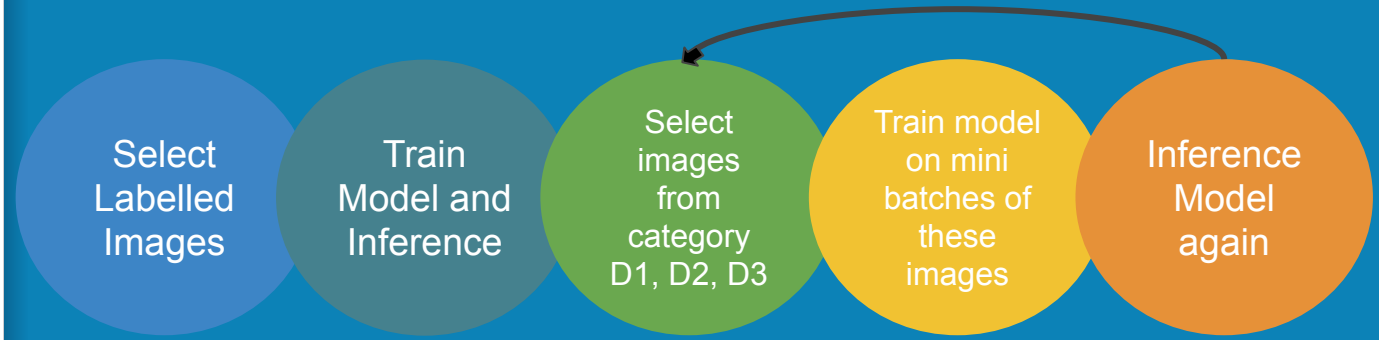
ACTIVE LEARNING SOLUTION







Active Learning



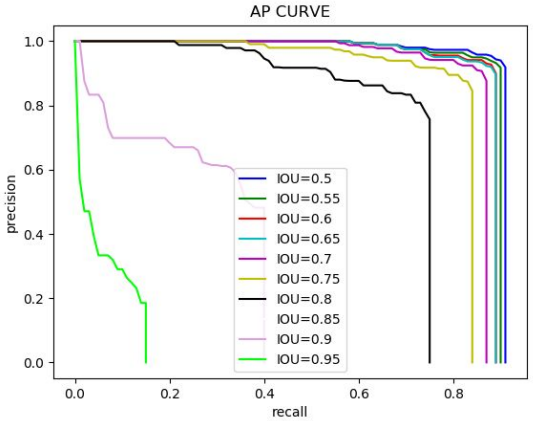
D1 = 1 Random Labelled Image
D2 = 1 Labelled Image with Score ~ 1/(N+1)
D3 = 2 Unlabelled Images with Score ~ 1/(N+1)

N = Number of Classes, 1 = Background Class

RESULTS

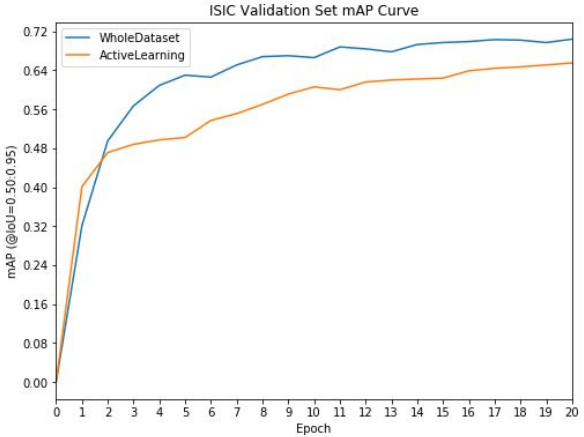
ISIC Dataset Results (RetinaNet):

Whole Dataset Training:

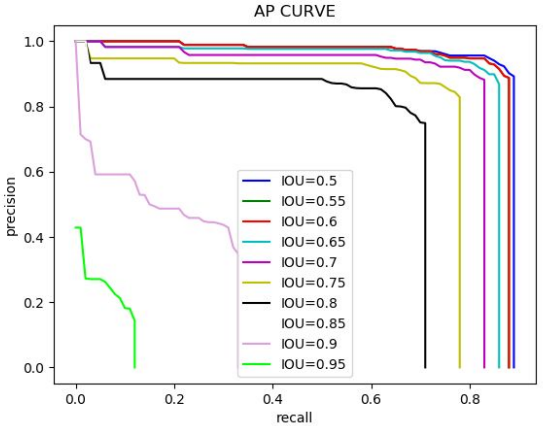


Training Process:

- Train on the whole dataset (**2334** images) at every epoch



Active Learning:



Training Process:

- Train on random subset of 40 labelled images for 5 epochs
- At each following epoch, train on 20 mini-batches of 4 images selected based on D1, D2, D3 criteria
- Total # images used per epoch: **80**

ACCOMPLISHMENTS & GOING FORWARD

Accomplishments:

1. Successfully set up an AL pipeline using RetinaNet

2. Tested and validated its effectiveness and efficiency using the ISIC dataset

3. Recognized some key problems with performance on Duke dataset: varying image quality, glare, size of lesions, etc.

Going Forward:

1. Incorporate the Duke clinical data into the pipeline

2. Further automate the process by integrating the lesion app within the AL pipeline to serve as an oracle

3. Continue to improve the performance on the highly varying Duke dataset

Thank You