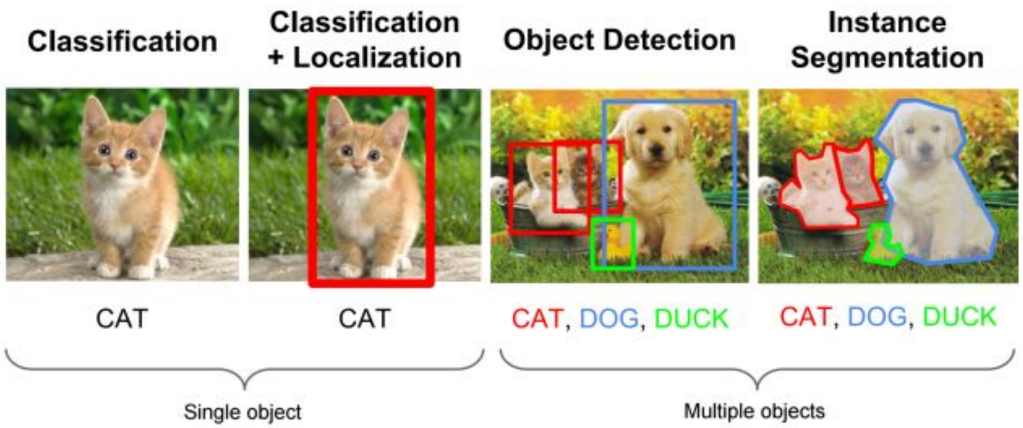
Instance Segmentation on the Wood Debris Problem

A study focused on the application of instance segmentation for the wood debris problem, as proposed by Dr. Nikolai Strigul.

Sarita Hedaya
EN-250 Fall 2019
Prof. Nick Strigul

A brief overview of Instance Segmentation…

What is Instance Segmentation?



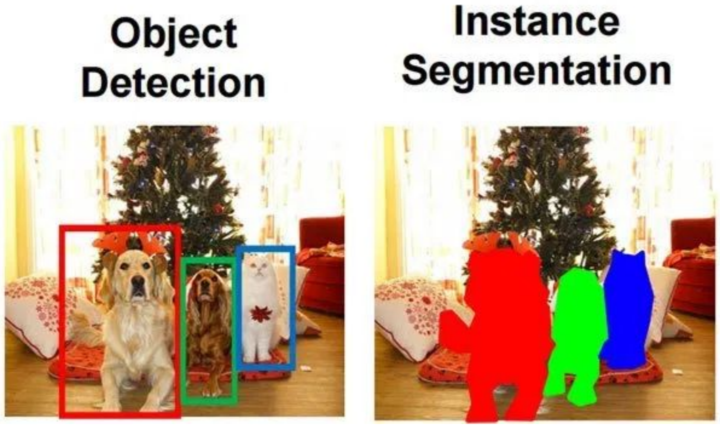
Classification: is there a cat in this picture?

Localization: roughly, where is the cat? (Bounding box)

Object detection: classification + localization with multiple classes

What's the shape of the objects? Where exactly are they? **Instance Segmentation** answers these questions.

How does Instance Segmentation work?



Instance segmentation creates a pixel-wise mask for each object in the image.

This technique gives us a far more granular understanding of the object(s) in the image.

History

Pre-deepLearning era

Used techniques that compared the pixel colors to some thresholds, and separated into classes just by pixel color.

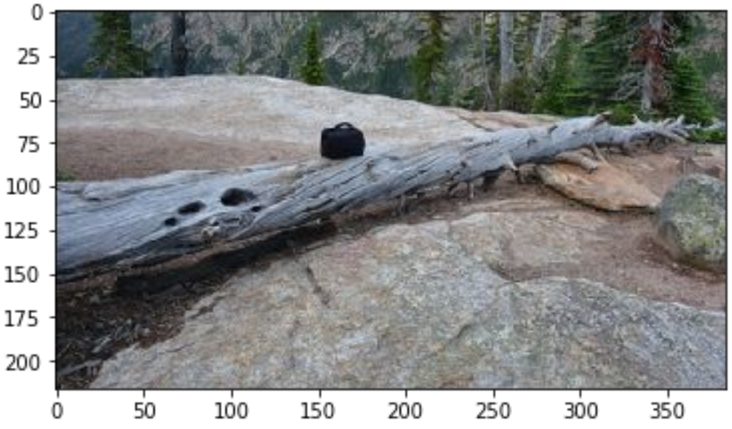
No consideration of shape, or continuity of an object.

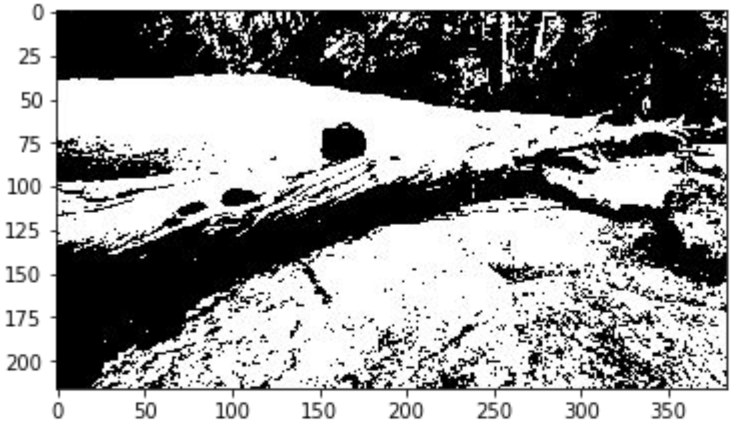
Model 1: Distance from mean color

1. Convert the image to grayscale
2. Flatten an image
3. Compare each pixel value to the mean
4. If above the mean, predict class 0.
5. If below the mean, predict class 1.

Results:

Simple approach doesn’t provide any meaningful segmentation.



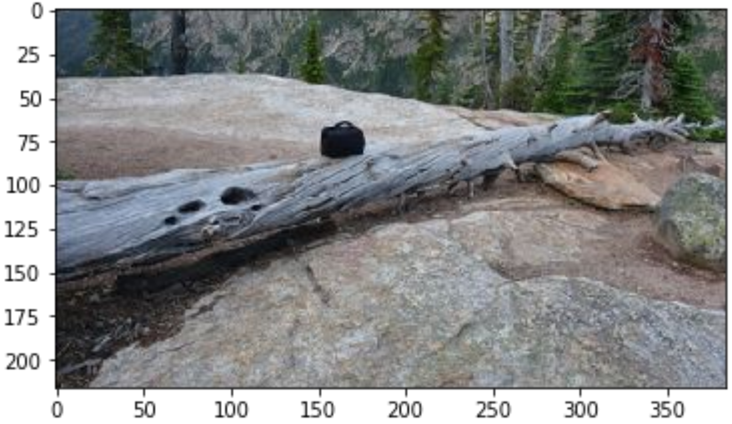


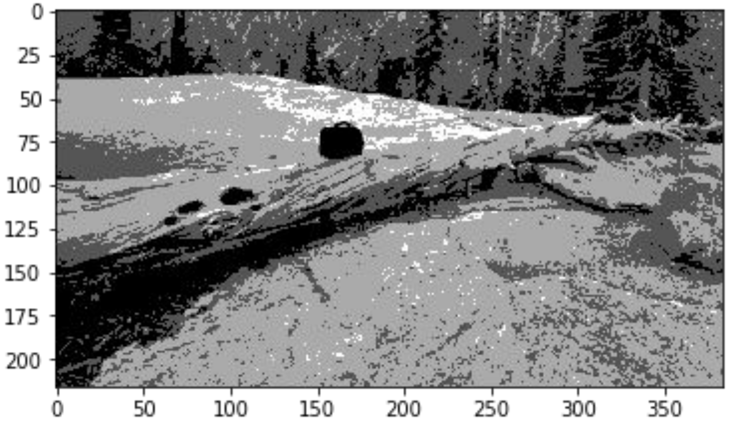
Model 2: Distance from mean color, multiclass

1. Convert to grayscale
2. Flatten an image
3. Compare each pixel value to the mean
4. If above A and below B, predict class 0.
5. If above B and below C, predict class 1.
6. If above C and below D, predict class 2.

Results:

Even by using 4 classes, this simple approach doesn't provide any meaningful segmentation.

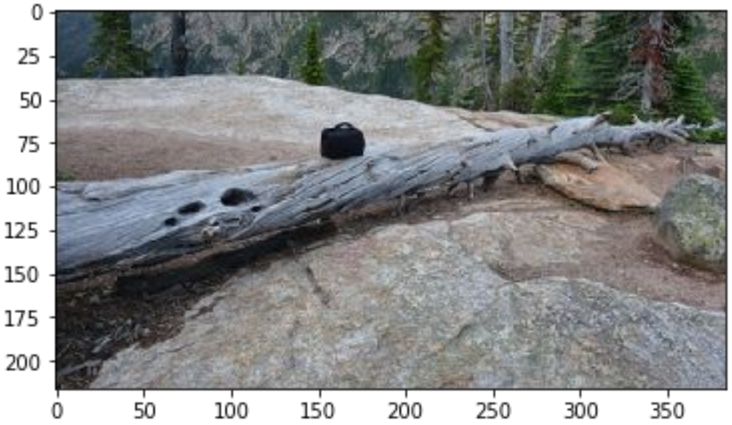


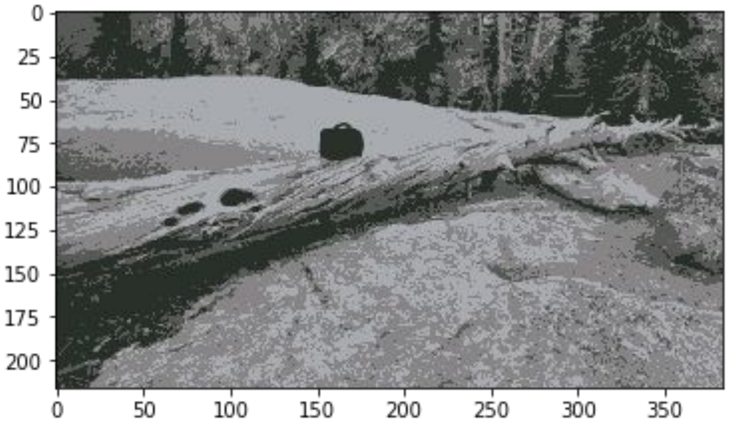


Model 3: Using k-means clustering

1. Flatten an image
2. Use sklearn Kmeans method
3. Select a number of clusters to split unto
4. Reshape the image back to its original shape

Results:
This method does not convert to grayscale, so we can use the colors as additional information.
Even so, it's too simple. This 4 clusters provide no meaningful segmentation



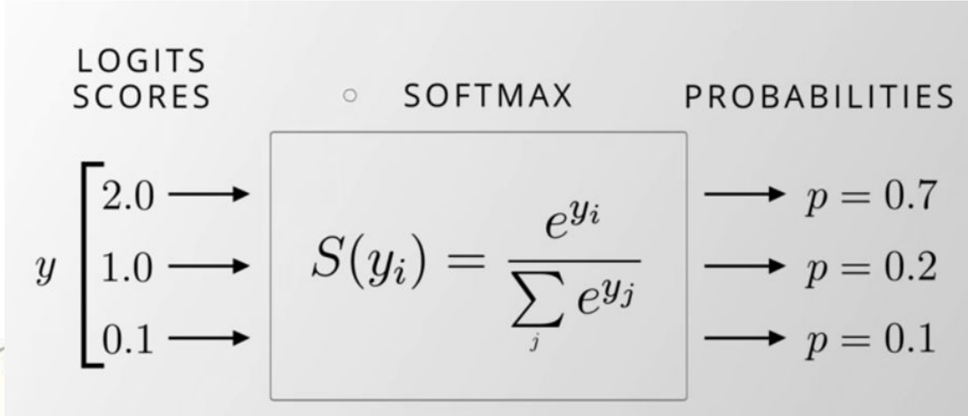


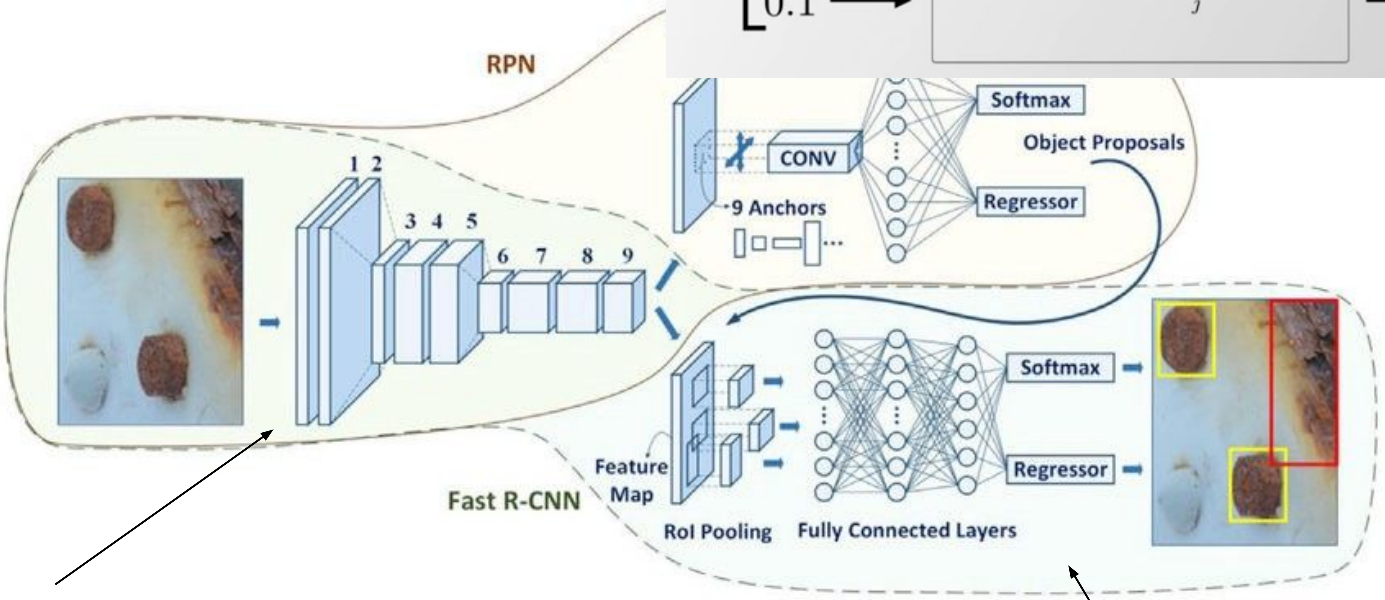
Mask R-CNN

State of the art instance segmentation technique

History

Built on top of **Faster R-CNN**



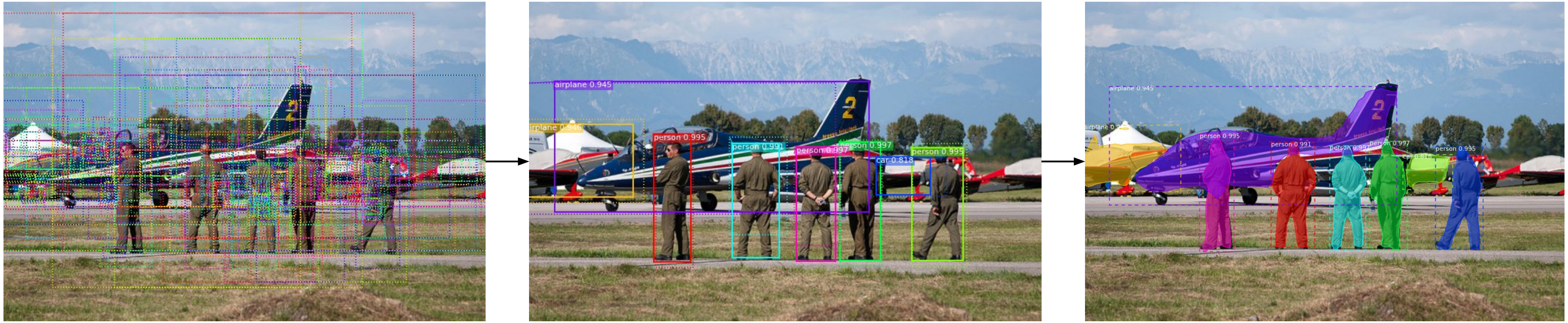


Part 1: Convolutional Layers
for feature extraction

Part 3: Detect objects

Model 4: Mask R-CNN

1. Anchor sorting and filtering
2. Bounding Box Refinement
3. **Generating masks**



1 2 3

Loss function

Standard Cross Entropy

Each pixel of the output of the network is compared with the corresponding pixel in the ground truth segmentation image. We will apply standard cross-entropy loss on each pixel.

Other loss functions: DeepMask's CE

$$\mathcal{L}(\theta) = \sum_k \left(\frac{1+y_k}{2w^o h^o} \sum_{ij} \log(1 + e^{-m_k^{ij} f_{segm}^{ij}(x_k)}) + \lambda \log(1 + e^{-y_k f_{score}(x_k)}) \right)$$

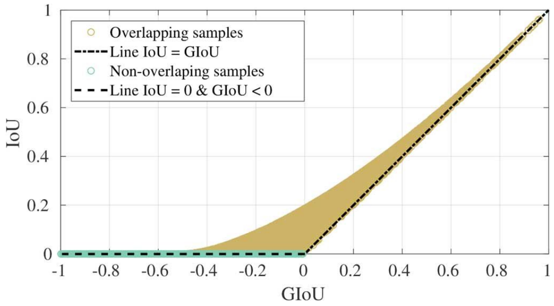
the loss function is a sum of binary logistic regression losses, one for each location of the segmentation network and one for the object score.

Metric: Intersection over Union

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{|I|}{|U|}$$

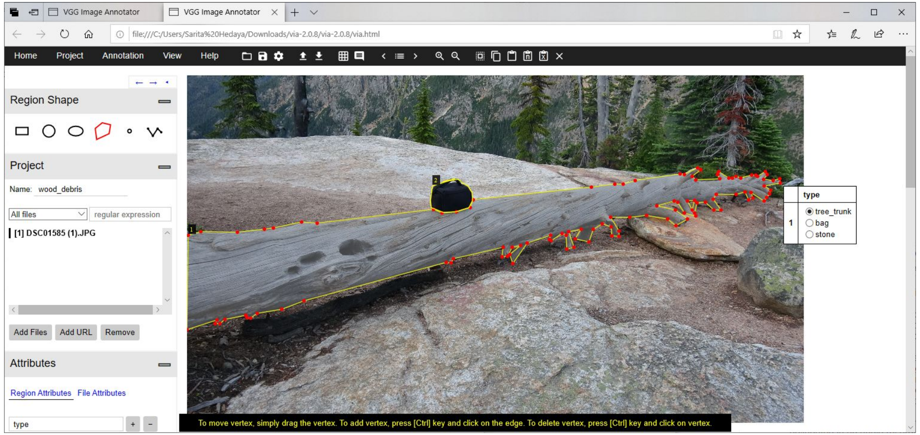
In practice we use GIoU because in a neural network, any given loss function must be differentiable to allow for backpropagation.

In cases where there is no intersection, IoU has no value and therefore no gradient. GIoU however, is always differentiable



How to train a Mask R-CNN model?

1. Select a dataset of images containing the objects to segment
2. Annotate the images by drawing boundaries around the objects (~3 minutes per image)
3. Use transfer learning from the COCO dataset. (~120,000 images)
4. Train using a modern GPU. Mask R-CNN uses ResNet101, a fairly large model.
5. Inspect your results

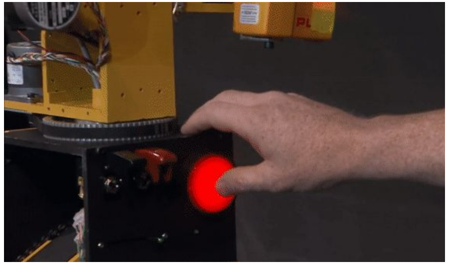


Using VIA (VGG Image Annotator) to draw boundaries

Using a GPU

ir

GPU	CPU
• hundreds of simpler cores	• few very complex cores
• thousand of concurrent hardware threads	• single-thread performance optimization
• maximize floating-point throughput	• transistor space dedicated to complex ILP
• most die surface for integer and fp units	• few die surface for integer and fp units



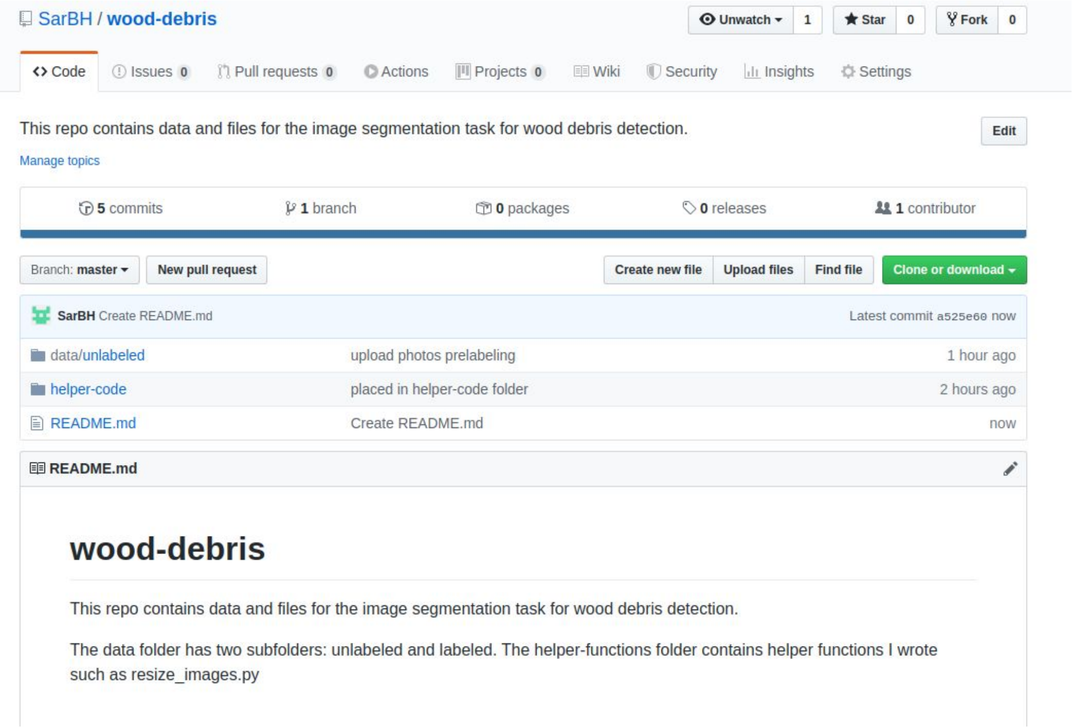
My progress training a Mask R-CNN

1. Create GitHub repository

A GitHub repository is used to store files, help with version control, manage changes, aid with teamwork, and backup data.

I created a GitHub repo that you can access with this link:

https://github.com/SarBH/wood-debris



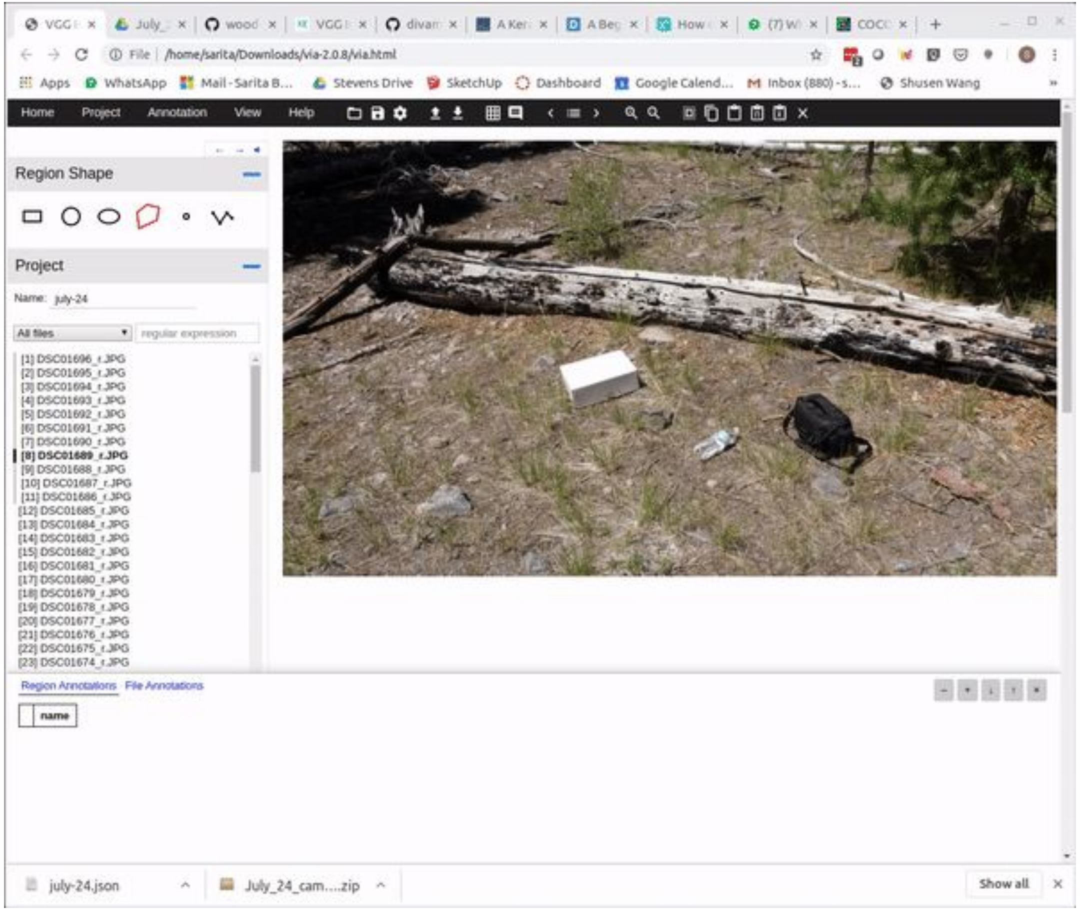
2. Prepare data

a. Select images to train on

b. Resize images to a more manageable size

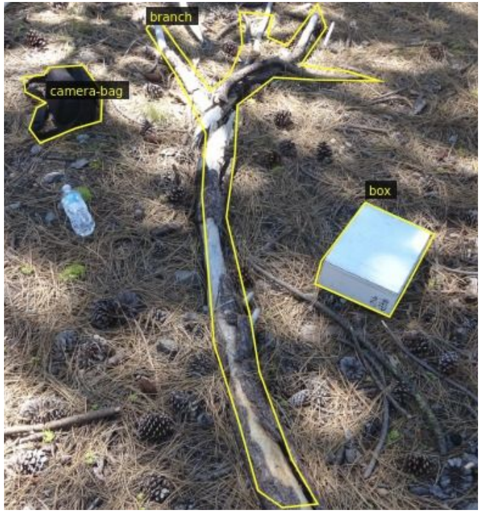
c. Manually annotate them to create masks

d. Create image mask from polygon points

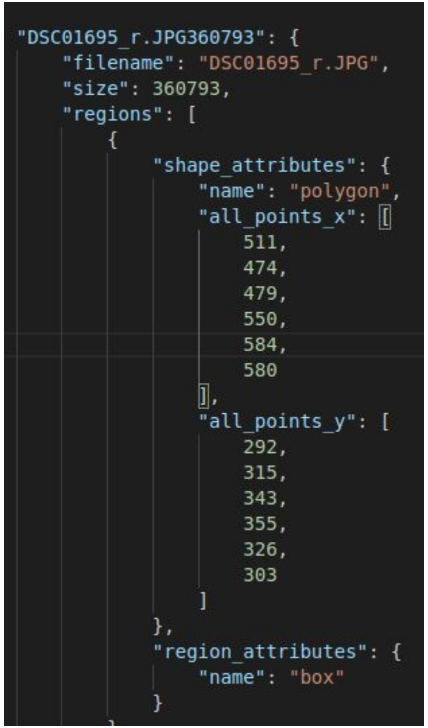


2. Prepare data

Original Image

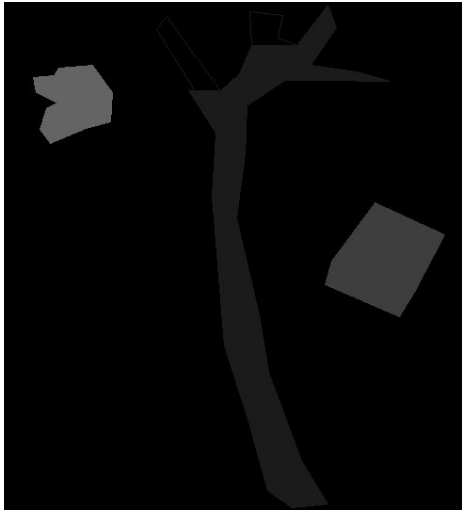


VIA annotation



```
"DSC01695_r.JPG360793": {
    "filename": "DSC01695_r.JPG",
    "size": 360793,
    "regions": [
        {
            "shape_attributes": {
                "name": "polygon",
                "all_points_x": [
                    511,
                    474,
                    479,
                    550,
                    584,
                    580
                ],
                "all_points_y": [
                    292,
                    315,
                    343,
                    355,
                    326,
                    303
                ]
            },
            "region_attributes": {
                "name": "box"
            }
```


VIA output as a .json file

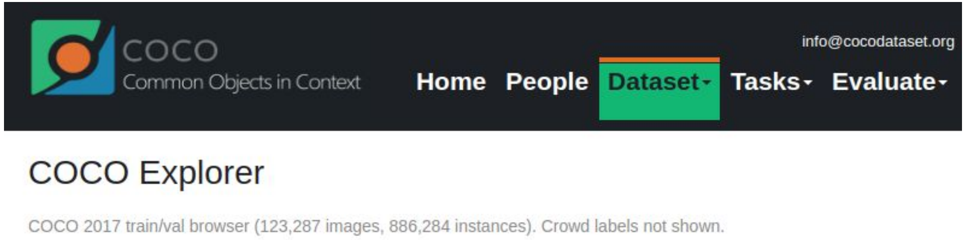


Mask

3. Baseline to measure progress

Machine learning is an **iterative process** that relies on **experimentation**. Because of this, it's important to **set baselines to measure improvement**.

Since my idea is to repurpose a pre-trained model, I baselined by **evaluating the original model on our data.**



Future improvements to our approach...

01	Even if we get segmentation to work reliably, camera bag can take multiple “sizes” because of the loose carrying belt, therefore should not be used to calibrate volume of trees.	Use reference objects that don’t change shape due to positioning, a non malleable solid object should work better.
02	Images are not taken from the “top view”, so depending on the angle and how far the reference objects are will change the mask size. Therefore cannot be reliably used to estimate debris size.	Take pictures from a birds eye view, possibly using a drone.
03	Current method of taking photos does not account for possible debris duplicates, i.e. multiple photos taken of the same branch will be added twice to the total debris calculation.	Replace current capturing system by an automated system that keeps a map of already captured areas.

Sources

Sharma, Pulkit. "Step-by-Step Implementation of Mask R-CNN for Image Segmentation." *Analytics Vidhya*, 22 July 2019, www.analyticsvidhya.com/blog/2019/07/computer-vision-implementing-mask-r-cnn-image-segmentation/?utm_source=blog&utm_medium=introduction-image-segmentation-techniques-python.

Gkioxari, Georgia, et al. "Mask R-CNN." ArXiv.org, Facebook AI Research (FAIR), 24 Jan. 2018, arxiv.org/abs/1703.06870

Khazri, Achraf. "Faster RCNN Object Detection." Medium, Towards Data Science, 9 Apr. 2019, towardsdatascience.com/faster-rcnn-object-detection-f865e5ed7fc4

Abdulla, Waleed. "Splash of Color: Instance Segmentation with Mask R-CNN and TensorFlow." Medium, Matterport Engineering Techblog, 10 Dec. 2018, engineering.matterport.com/splash-of-color-instance-segmentation-with-mask-r-cnn-and-tensorflow-7c761e238b46.

https://github.com/matterport/Mask_RCNN

Savarese, Silvio. "Generalized Intersection over Union." Stanford Artificial Intelligence Laboratory, giou.stanford.edu/.Gkioxari, Georgia, et al. "Mask R-CNN." ArXiv.org, Facebook AI Research (FAIR), 24 Jan. 2018, arxiv.org/abs/1703.06870

Gupta, Divam. *A Beginner's Guide to Deep Learning Based Semantic Segmentation Using Keras*. 6 June 2019, divamgupta.com/image-segmentation/2019/06/06/deep-learning-semantic-segmentation-keras.html.

"VGG Image Annotator (VIA)." *Visual Geometry Group*, University of Oxford, www.robots.ox.ac.uk/~vgg/software/via/.

Tsang, Sik-Ho. "Review: DeepMask (Instance Segmentation)." *Medium*, Towards Data Science, 20 Mar. 2019, towardsdatascience.com/review-deepmask-instance-segmentation-30327a072339.

Additional Slides

Go to https://github.com/SarBH/wood-debris

Transfer Learning

In practice, very few people train an entire Convolutional Network from scratch (with random initialization), because it is relatively rare to have a dataset of sufficient size.

Instead, it is common to pretrain a ConvNet on a very large dataset (e.g. ImageNet, which contains 1.2 million images with 1000 categories), and then use the ConvNet either as an initialization or a fixed feature extractor for the task of interest.

Implementation

Each pixel of the output of the network is compared with the corresponding pixel in the ground truth segmentation image. We will apply standard cross-entropy loss on each pixel.