Object detection in aerial imagery using deep learning

Melise Steyn

12 November 2020

CSIR

Touching lives through innovation

Introduction

Modern remote sensing technologies have developed rapidly in the past decade

∴ Optical imagery with very high spatial resolutions are now available

This imagery facilitates a wide range of applications, such as

- Disaster control
- Land planning
- Urban monitoring
- Traffic planning

Object detection from these high resolution images has gained increasing interest.

Thatched roofs

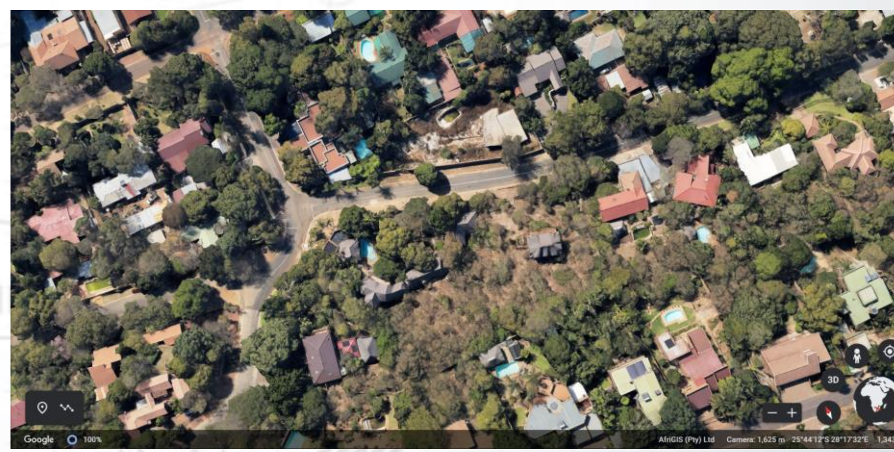
The thatched roof has formed an integral component of the South African architectural landscape due to its aesthetic appeal and insulation properties.

However, it can be a fire hazard.
→ Insurance companies are interested in knowing where these thatched roofs are located.

Potential solution: detecting thatched roofs using aerial photography

However, a thatch roof detection solution that depends on manual image interpretation only = time-consuming and challenging.

St Francis Bay, Eastern Cape



Silverton, Pretoria

Thatched roofs

The thatched roof has formed an i chitectural landscape due to its aesthetic app

However, it can be a fire hazard.
→ Insurance companies are inter ofs are located.

Potential solution: detecting thatch

However, a thatch roof detection s terpretation only = time-consuming and challenging

St Francis Bay, Eastern Cape

Silverton, Pretoria

Thatched roofs

The thatched roof has formed an integral component of the South African architectural landscape due to its aesthetic appeal and insulation properties.

However, it can be a fire hazard.
→ Insurance companies are interested in knowing where these thatched roofs are located.

Potential solution: detecting thatched roofs using aerial photography

However, a thatch roof detection solution that depends on manual image interpretation only = time-consuming and challenging.

St Francis Bay, Eastern Cape

Silverton, Pretoria

Thatched roofs

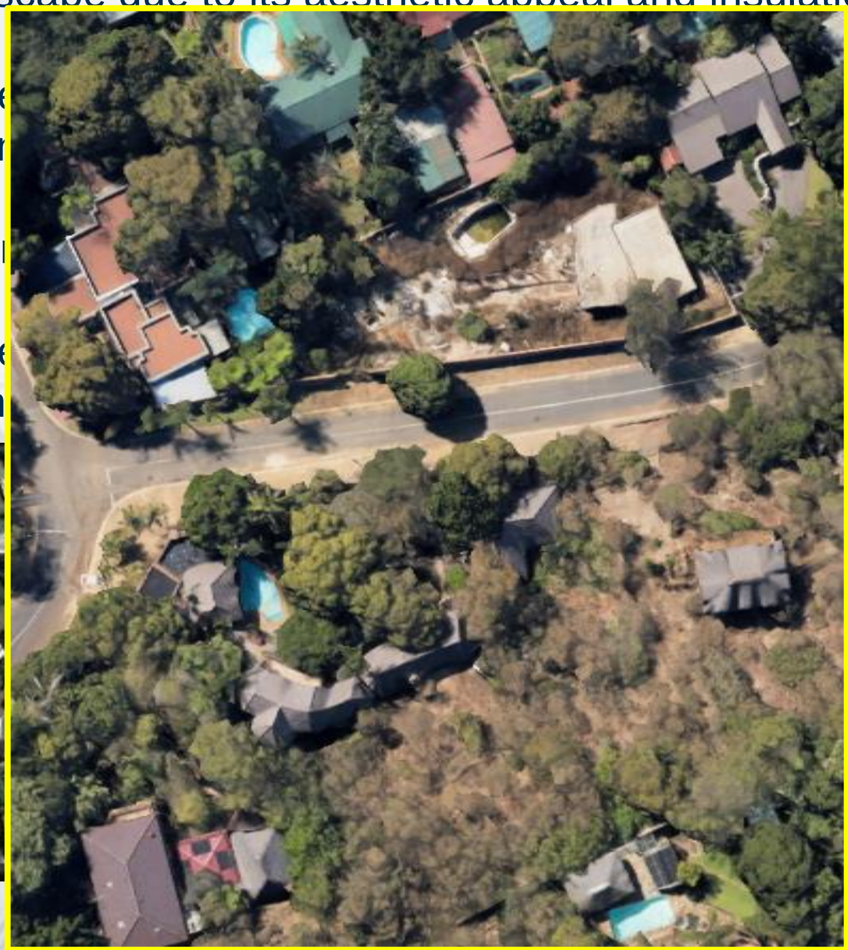
The thatched roof has formed an integral component of the South African architectural landscape due to its aesthetic appeal and insulation properties.

Howe

→ In … g where these thatched roofs are located.

Poten … rial photography

Howe … pends on manual image interpretation only = tim

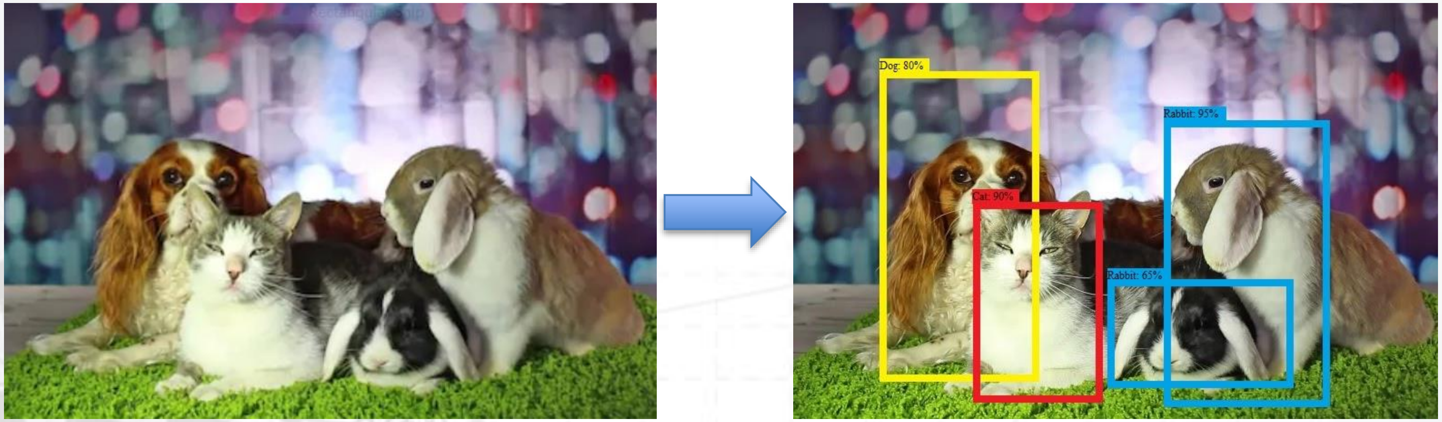


Silverton, Pretoria

Object detection

Object detection is a computer vision task that involves

1. identifying which of a known set of objects might be present in an image
2. giving information about their locations within the image
3. giving a confidence score for each detected object and its location



Using object detection model to locate objects in aerial images may enable user to

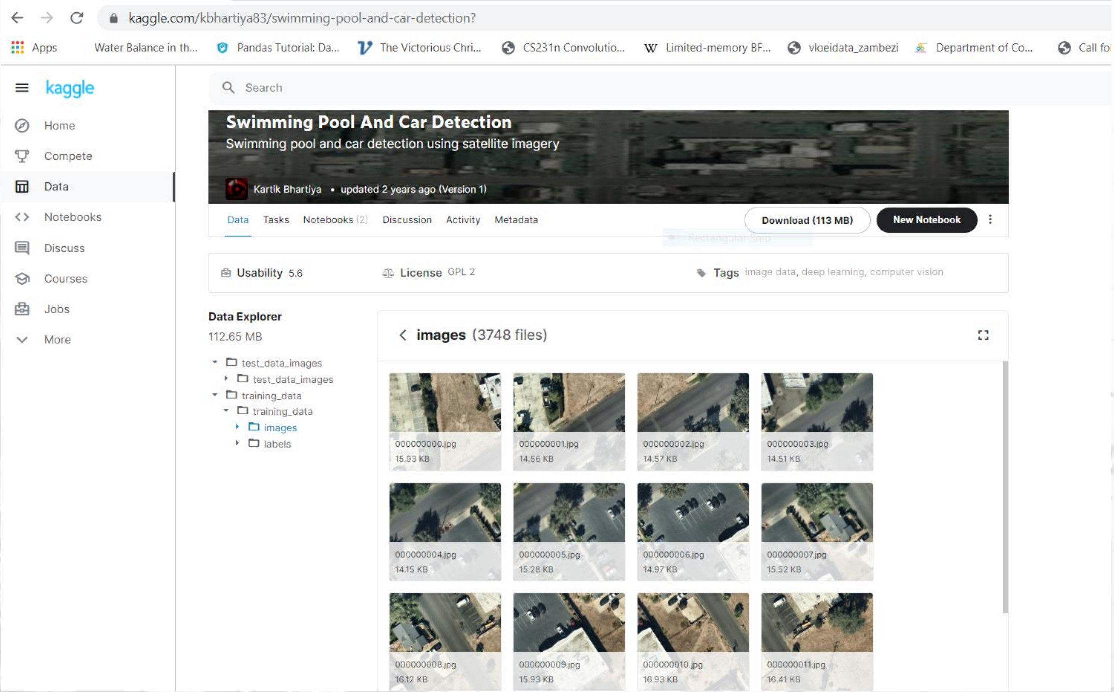
- monitor faster
- cover larger surfaces

Objective

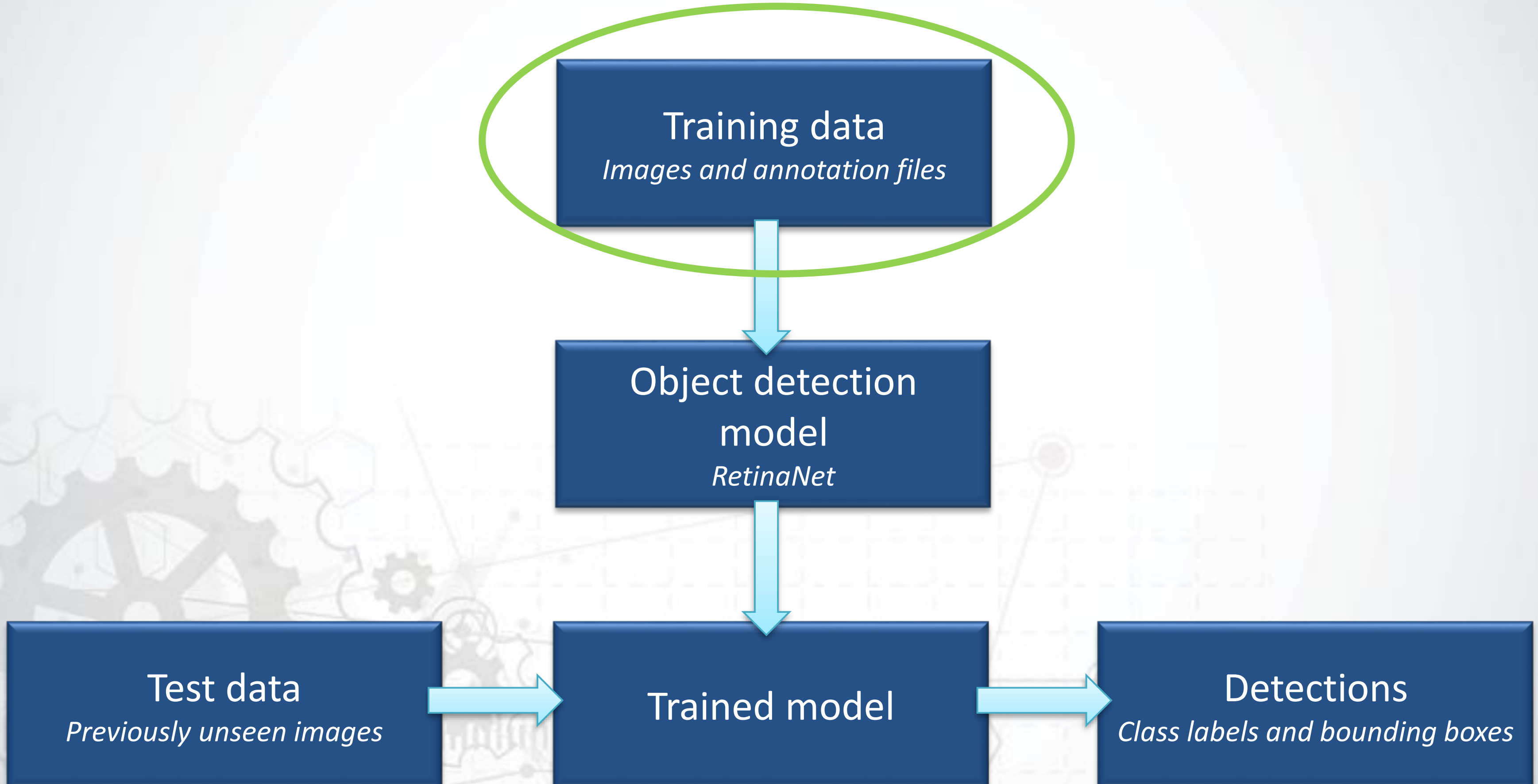
Unfortunately, no thatch roof aerial photography data is currently available.

However, an aerial imagery set of residential areas with annotation data for swimming pools and cars was available.

→ Could investigate the potential and feasibility of our project objective by developing a deep learning model to detect swimming pools and cars from aerial imagery.



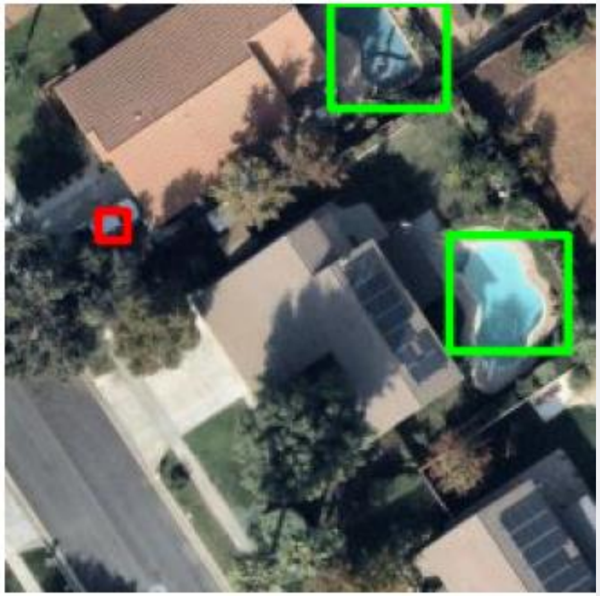
Model development



Methodology: Data preparation

- Aerial imagery set: 3748 RGB images of size 224x224
- For each image, ESRI provided corresponding annotation file containing **class labels** and **bounding box coordinates** of cars and swimming pools within the image


```
<annotation>
  <filename>000000017.jpg</filename>
  <source>
    <annotation>ArcGIS Pro 2.1</annotation>
  </source>
  <size>
    <width>224</width>
    <height>224</height>
    <depth>3</depth>
  </size>
  <object>
    <name>2</name>
    <bndbox>
      <xmin>167.73</xmin>
      <ymin>87.30</ymin>
      <xmax>212.18</xmax>
      <ymax>131.74</ymax>
    </bndbox>
  </object>
  <object>
    <name>2</name>
    <bndbox>
      <xmin>122.21</xmin>
      <ymin>0.00</ymin>
      <xmax>166.65</xmax>
      <ymax>39.58</ymax>
    </bndbox>
  </object>
  <object>
    <name>1</name>
    <bndbox>
      <xmin>34.63</xmin>
      <ymin>78.29</ymin>
      <xmax>45.74</xmax>
      <ymax>89.40</ymax>
    </bndbox>
  </object>
</annotation>
```

- Training = 80% of images; testing = 20% of images

	Number of images	Number of annotated cars	Number of annotated swimming pools	Total number of annotated objects
Training set	2998	10202	2524	12726
Test set	750	2712	598	3310

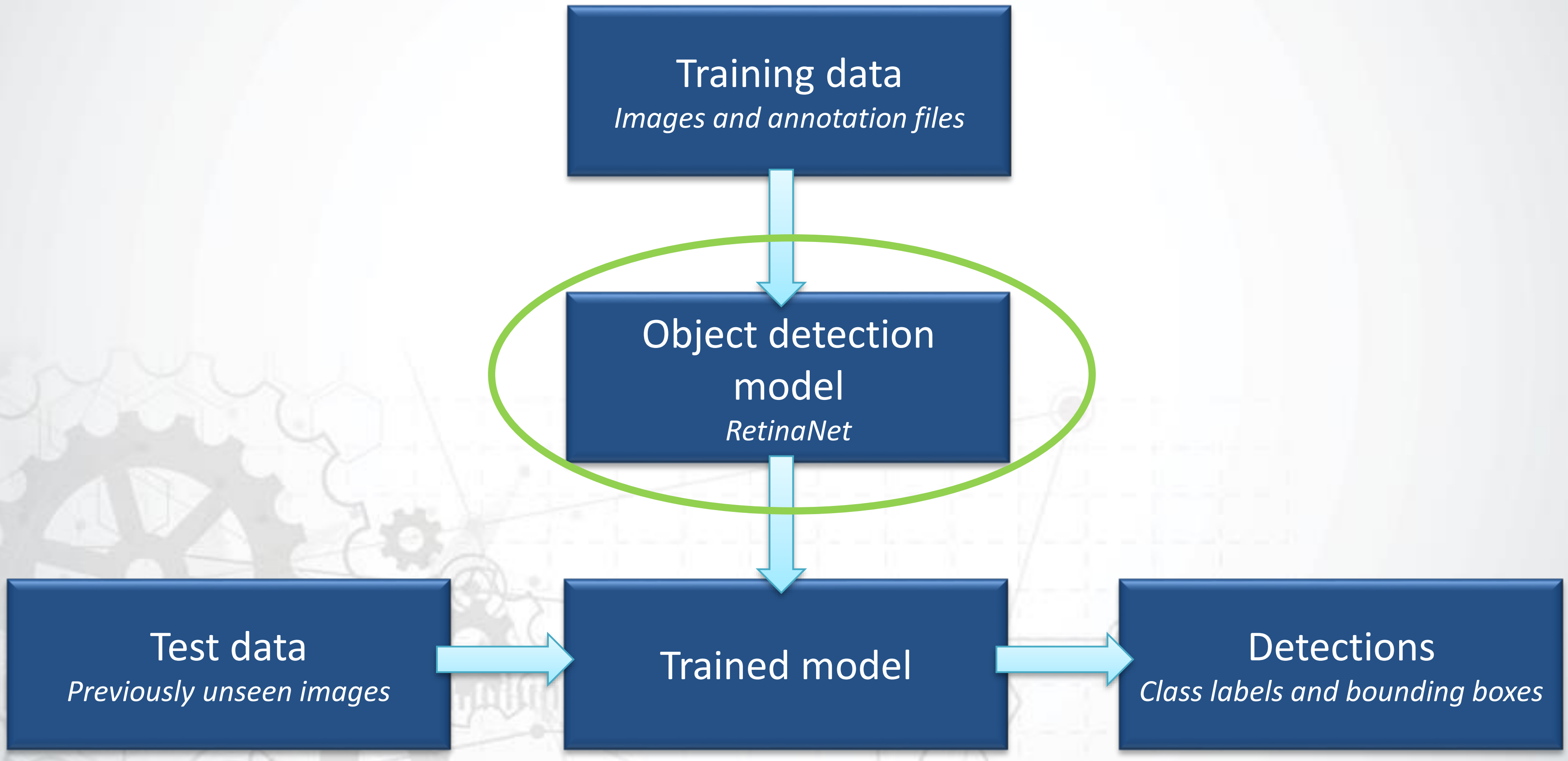
Methodology: Data preparation

- Aerial imagery set: 3748 RGB images of size 224x224
- For each image, ESRI provided corresponding annotation file containing **class labels** and **bounding box coordinates** of cars and swimming pools within the image

- Training = 80% of images; testing = 20% of images

	Number of images	Number of annotated cars	Number of annotated swimming pools	Total number of annotated objects
Training set	2998	10202	2524	12726
Test set	750	2712	598	3310

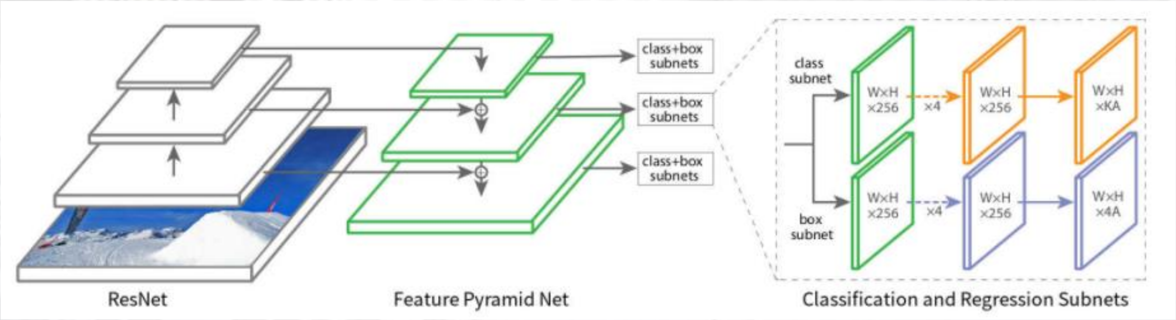
Model development



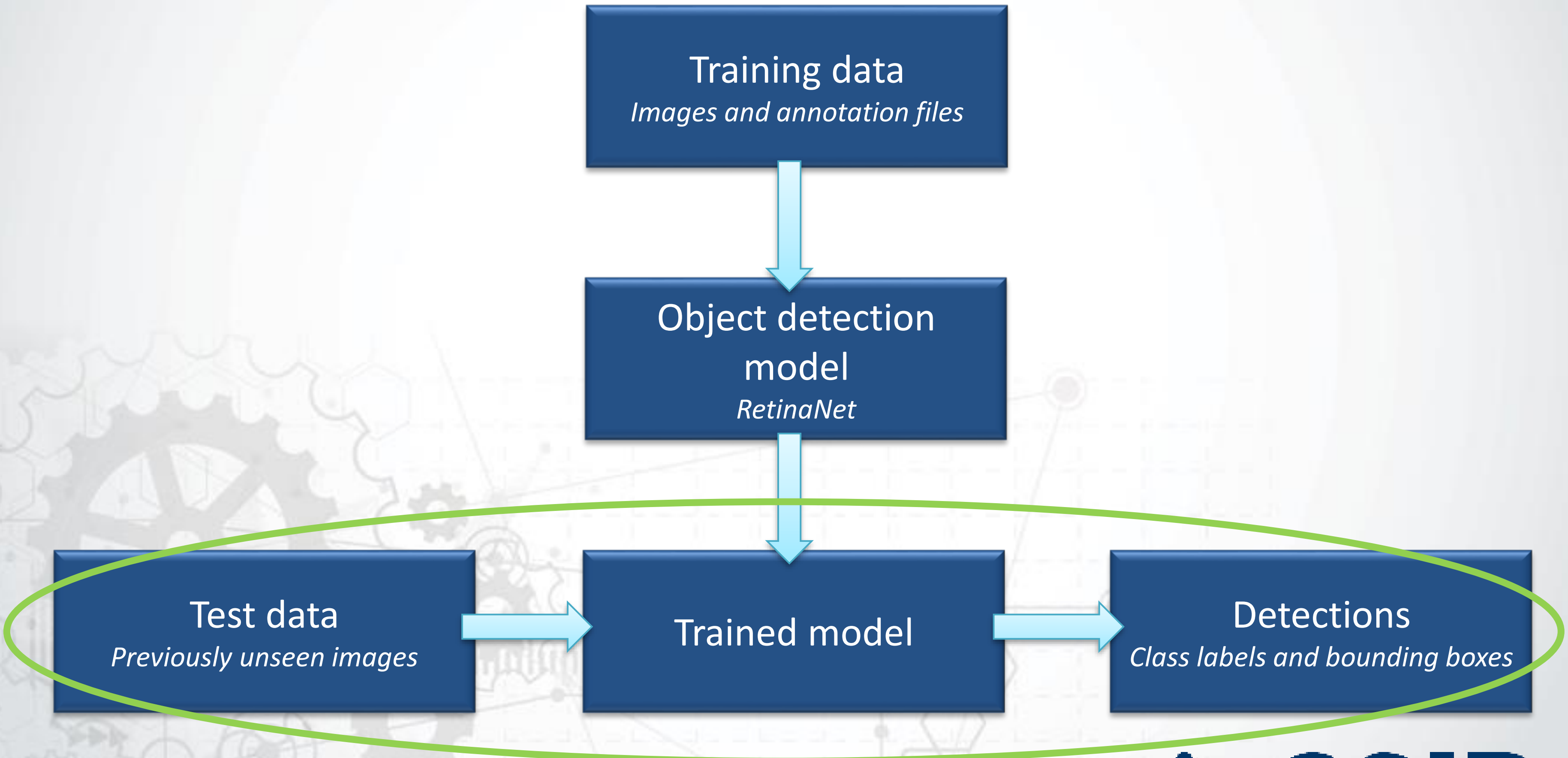
Object detection model: RetinaNet

RetinaNet:

- One-stage deep learning network
- Built on 2 crucial concepts:
  - Focal loss
  - Feature Pyramid Net
- A composite network composed of a
  - backbone network called Feature Pyramid Net, built on top of ResNet and responsible for computing convolutional feature maps of an entire image
  - subnetwork which performs object classification using the backbone's output
  - subnetwork which performs bounding box regression using the backbone's output
- Proven to be a useful and popular object detection model in the field of aerial photography and satellite imagery where objects of interest can be dense and small



Model development



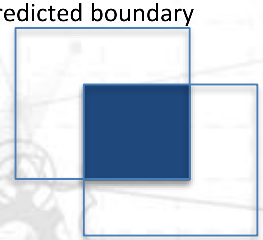
Performance criteria

Mean average precision (mAP)

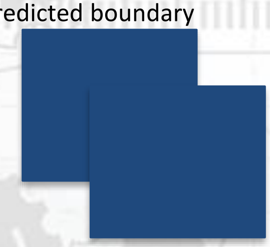
- is used to evaluate the RetinaNet model performance
- calculates the area under the precision-recall curve
- results in a value ranging from 0 to 1 (higher number = better performance)

Intersection over Union (IoU):

- Used to decide whether model's detection should be considered a true or false positive.



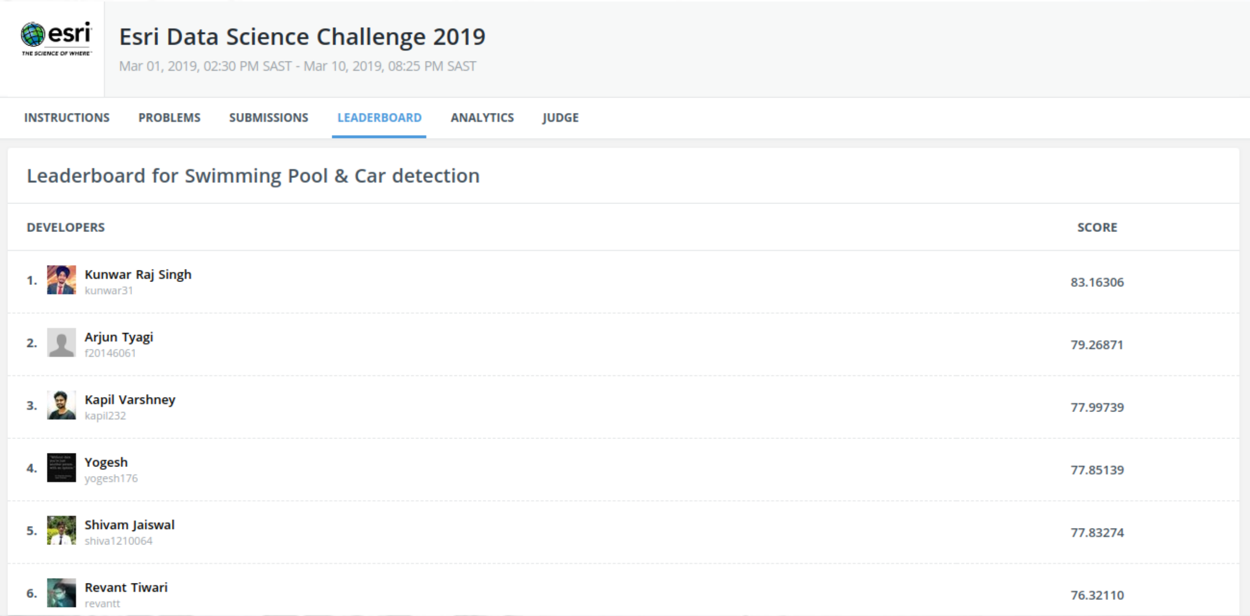
- $\mathrm{IoU} = \frac{\text{area of overlap}}{\text{area of union}} =$ (diagram above) / (diagram below) if $\mathrm{IoU} \geq 0.5$: True positive



Results

Performance evaluation results of object detection model on the training and test sets:

	Cars		Swimming pools		
	Number of annotations	Average precision	Number of annotations	Average precision	Mean average precision
Training set	10202	0.8025	2524	0.7697	0.7861
Test set	2712	0.7980	598	0.7646	0.7813


Esri Data Science Challenge 2019

Mar 01, 2019, 02:30 PM SAST - Mar 10, 2019, 08:25 PM SAST

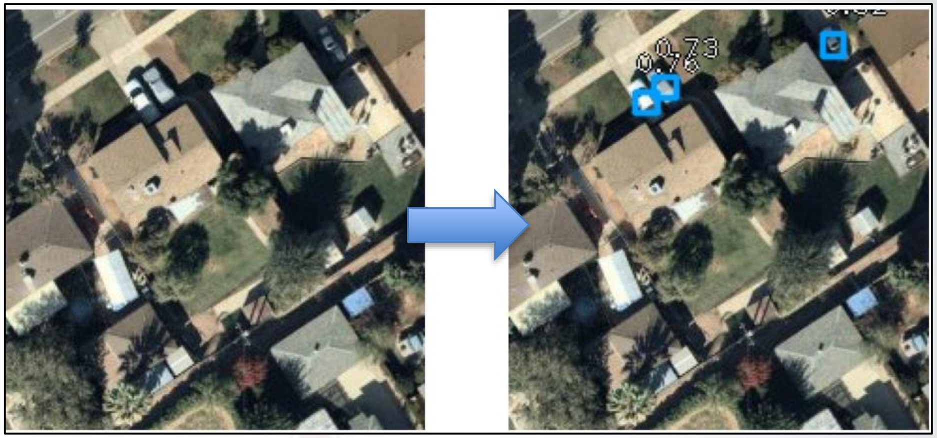
INSTRUCTIONS PROBLEMS SUBMISSIONS LEADERBOARD ANALYTICS JUDGE

Leaderboard for Swimming Pool & Car detection

	DEVELOPERS	SCORE
1.	Kunwar Raj Singh kunwar31	83.16306
2.	Arjun Tyagi f20146061	79.26871
3.	Kapil Varshney kapil232	77.99739
4.	Yogesh yogesh176	77.85139
5.	Shivam Jaiswal shiva1210064	77.83274
6.	Revant Tiwari revantt	76.32110

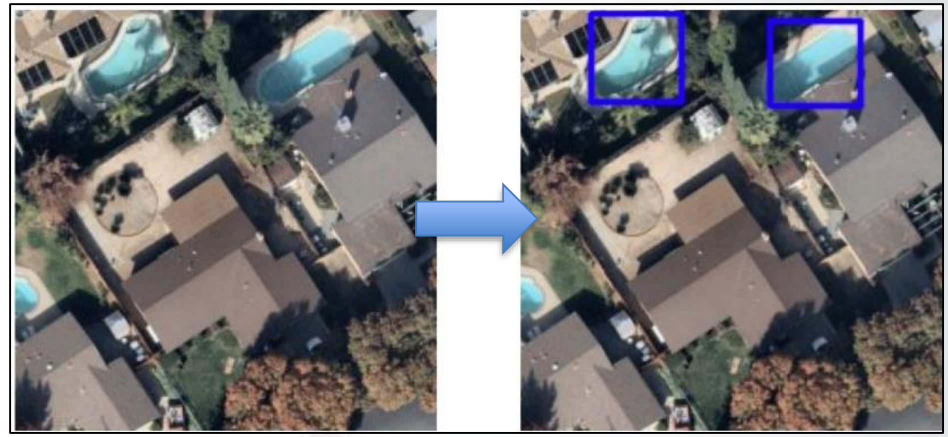
Results

Images from test set (IoU threshold = 0.6)



Results

Images from test set (IoU threshold = 0.6)



Results

Images from test set (IoU threshold = 0.6)

Conclusions and recommendations

- RetinaNet was capable of detecting a variety of cars positioned in various angles and located in different areas.

- Slightly less promising results for swimming pools. May be attributed to
  - less available annotations to train on
  - ambiguity of a swimming pool's features (shape, size and colour)

- Model performance could be improved by
  - collecting and annotating more images containing cars and swimming pools
  - developing more advanced data augmentation techniques
  - more carefully selecting images for training
  - incorporating information from near-infrared band (NIR)

END