

# Machine Learning for Geospatial Raster Data

Workshop #2    Please visit [bit.ly/geoml-2](https://bit.ly/geoml-2)

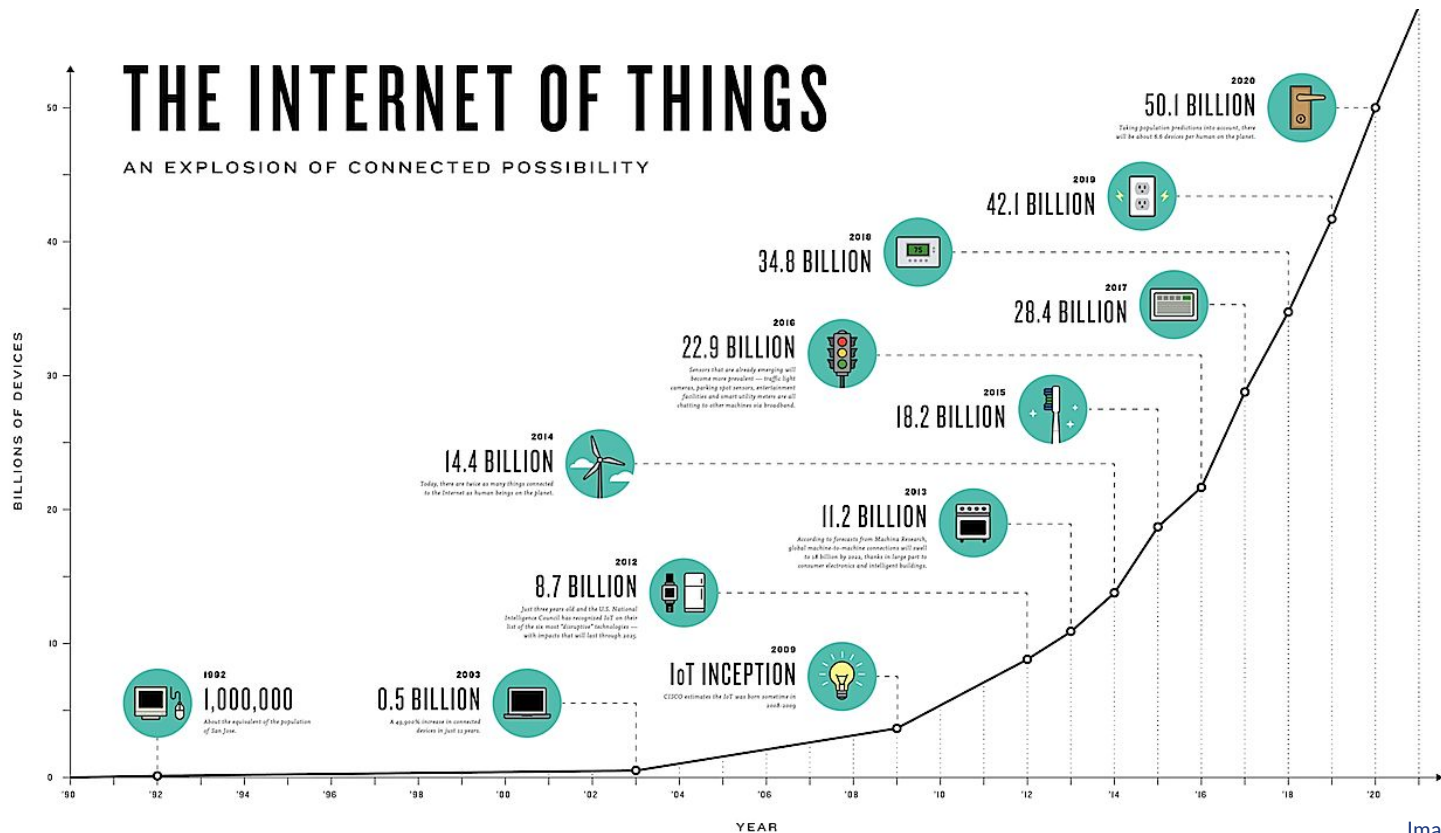
Presented by Juan Carrillo & Jaydeep Mistry

On behalf of the Geospatial Club - Winter 2019  
In partnership with the University of Waterloo Geospatial Centre



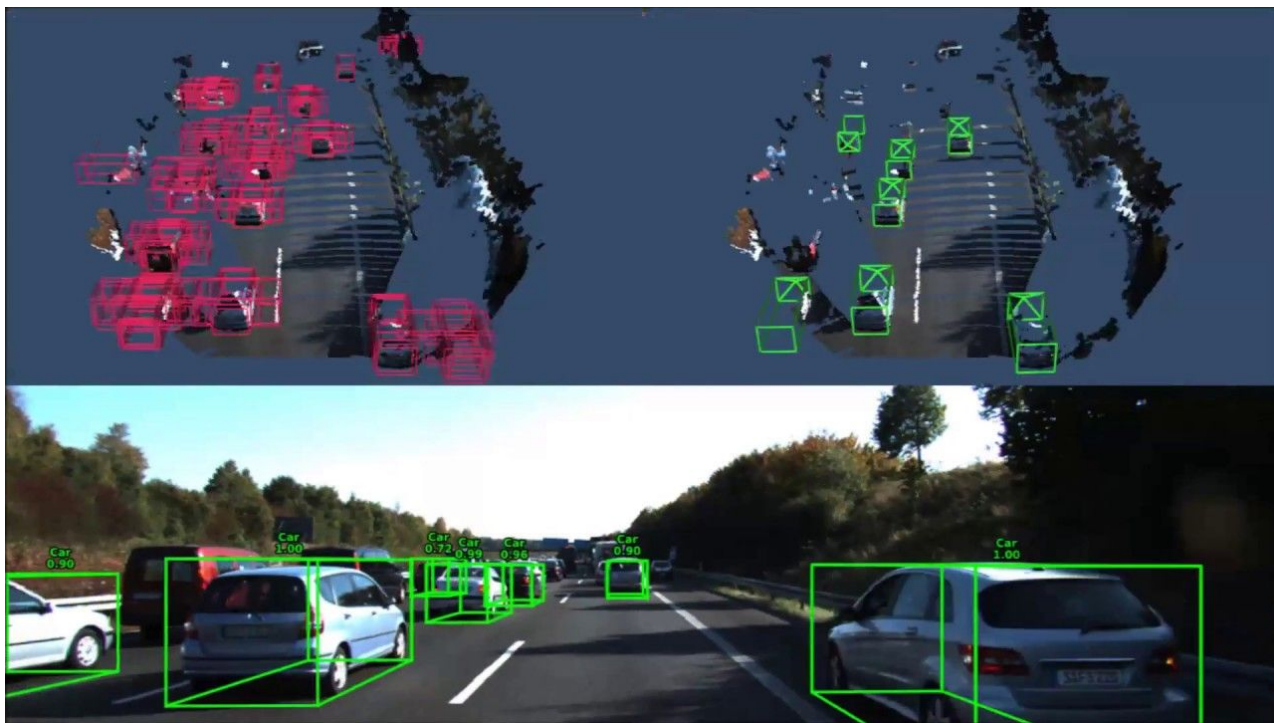
# Agenda

1. Why should we care about Machine Learning for Raster data?
2. Problem context: Understanding the Amazon from Space
3. Setup of the cloud-computing platform
4. Data exploration
5. Convolutional Neural Networks
6. Wrap-up



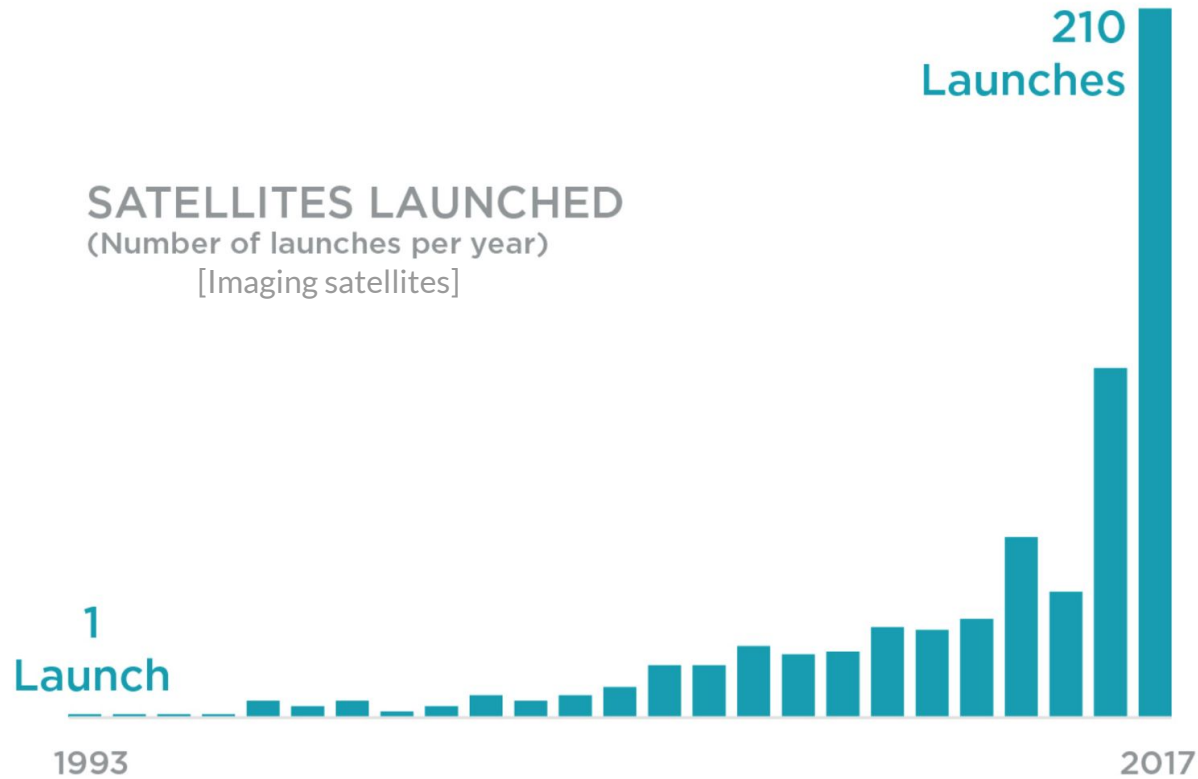
[Image Source](#)

## Why does Machine Learning matter?



[Image Source](#)

# Common Computer Vision tasks



[Image Source](#)

# Why does Raster data matter?

**Classification**



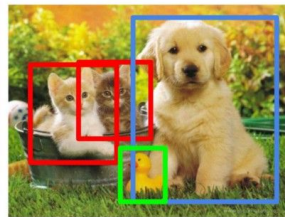
CAT

**Classification  
+ Localization**



CAT

**Object Detection**



CAT, DOG, DUCK

**Instance  
Segmentation**

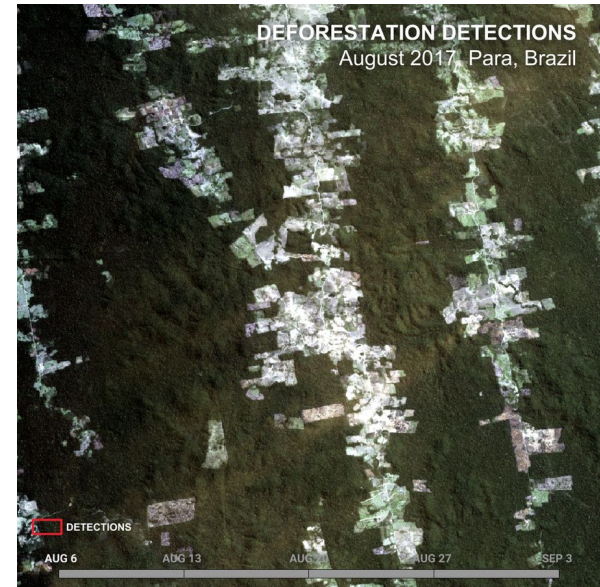


CAT, DOG, DUCK

Single object

Multiple objects

[Image Source](#)



[Image Source](#)

# Common Computer Vision tasks

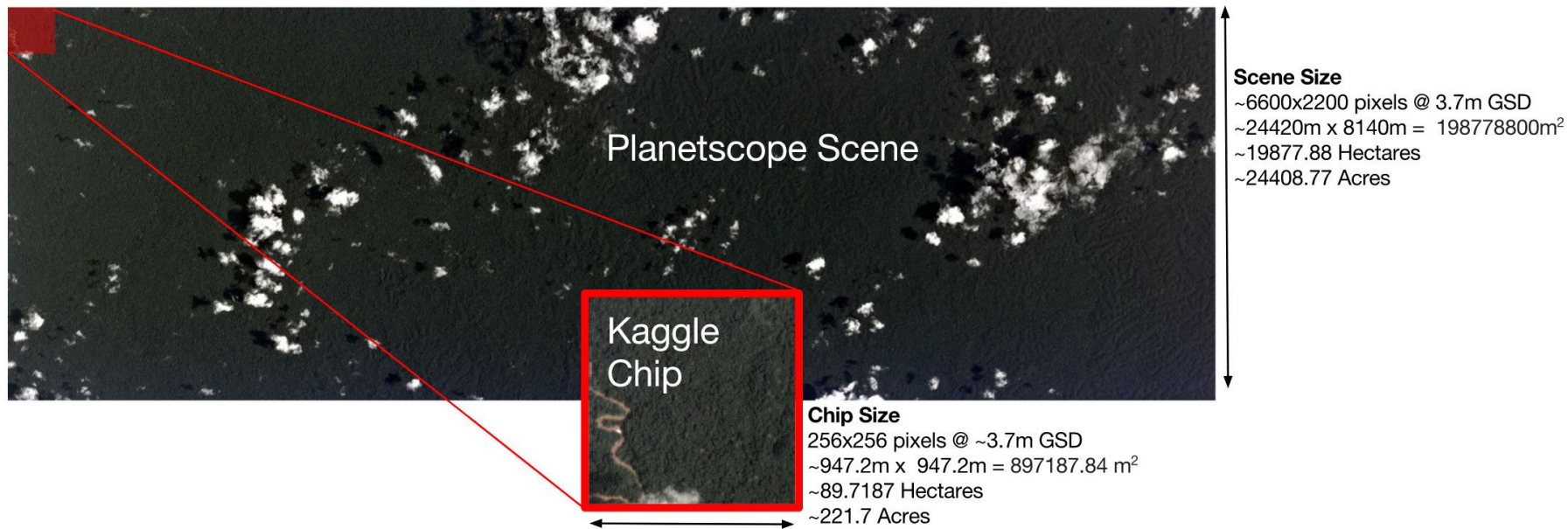


# The Problem:

## Identify Human Activity in the Amazon Rainforest

1. Use satellite data
2. Track human footprint in the Amazon rainforest
3. [Popular Kaggle Competition](#) for \$60,000





[Image Source](#)

# Dataset: 40,000+ Satellite Images





## How to perform Machine Learning online?

kaggle

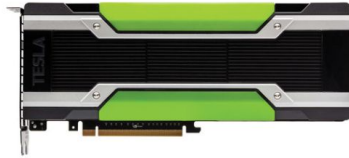
- Free-to-use platform
- Designed for Data Science Competitions
- Full integration with **Python** and **R** for Cloud Computing

Optional: [Google Colab](#), [Earth Engine](#)

# Specs of a Kaggle Kernel virtual machine

## NVIDIA Tesla K80 GPU Accelerator for Servers

B&H # NVTK80 • MFR # 900-22080-0000-000



### Special Order ⓘ

Expected availability: 7-14 business days

[Calculate Shipping](#)

### Product Highlights

- 560 MHz Core - Boostable to 876 MHz
- 4992 CUDA Cores
- 24GB GDDR5 vRAM
- 10 GHz Effective Memory Clock

[Show more](#)

Reviews 0 [Write a review](#) | 1 Questions, 1 Answers

1

QTY

CAD \$6,617.68  
USD \$4,999.00

Add to Cart

Add to Wish List

Attn: International Customers

Used from \$3,999.00



**True Know-How**  
Ask Our Experts  
Live Chat: 800.336.9338  
[Email](#)

16 Gigabytes of disk storage

4 CPU cores

17 Gigabytes of RAM

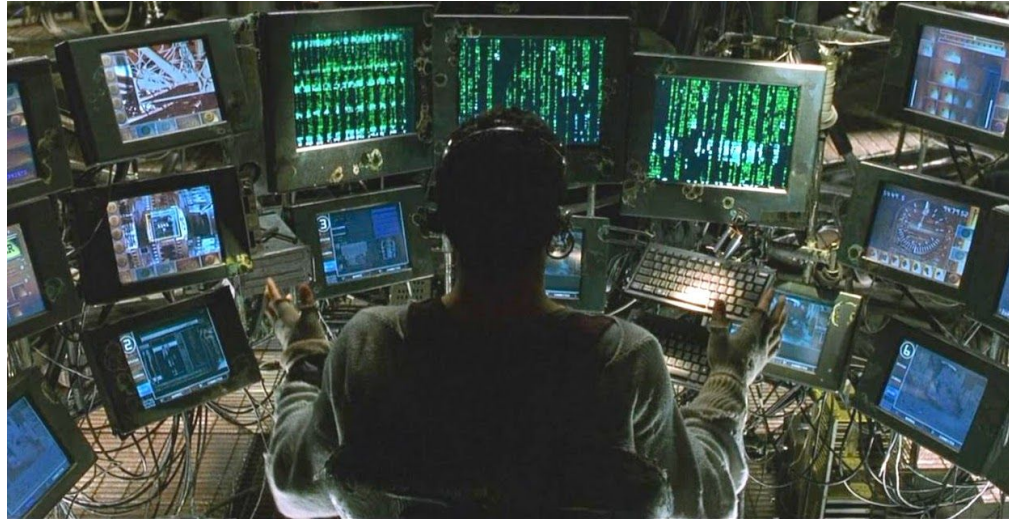
NVidia K80 GPU (12.5X speedup)

Source: [Kaggle](#) and [BH seller](#)

---

# Let's get started!

- Dataset
- Notebook



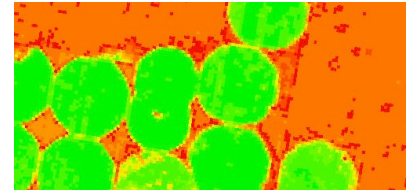
[Image source](#)

# Normalized Differential Vegetation Index (NDVI)

- Quantifies vegetation by using difference between Near-Infrared and Red light
  - Near-Infrared reflected by vegetation
  - Red light absorbed by vegetation
  - Ranges between -1 to +1



$$\text{NDVI} = \frac{(\text{NIR} - \text{Red})}{(\text{NIR} + \text{Red})}$$





# t-Distributed Stochastic Neighbor Embedding (t-SNE)

- For data dimensionality reduction
- Visualization of high dimensional datasets
- Uses t-Distribution to avoid having clusters group together

Why use this?

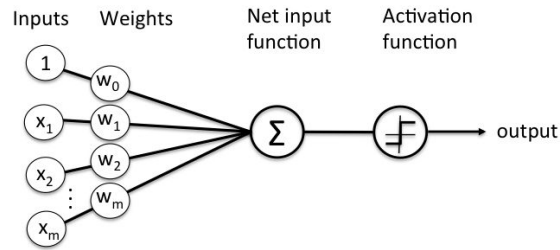
- Our Amazon rainforest dataset has
  - 40,000+ images
  - 256 x 256 pixels per image
  - Each image has 4 bands of data: Blue, Green, Red, Near-Infrared

Creator: [Visual Examples](#)

# Convolutional Neural Networks (CNN)

What are Neural Networks?

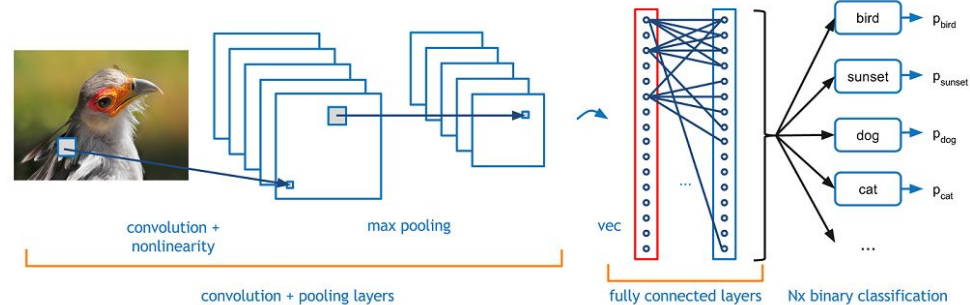
- System of connected 'Neurons' inspired by biological neural networks



Schematic of Rosenblatt's perceptron.

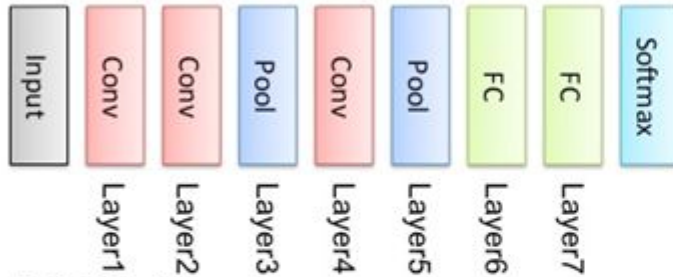
What is CNN?

- Multi-layered system of Neurons, where the data gets Pooled to remove noise

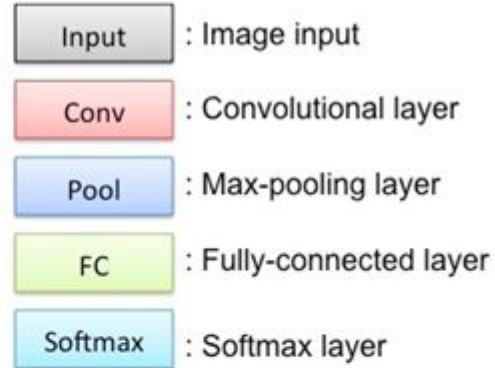
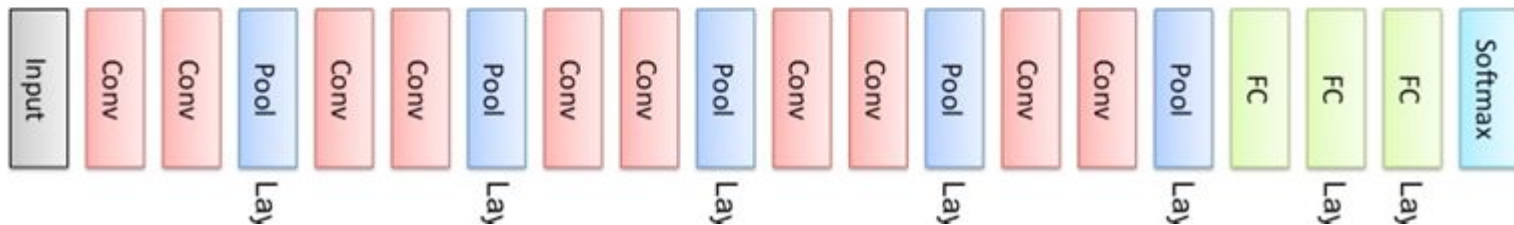


# Convolutional Neural Networks (CNN)

AlexNet

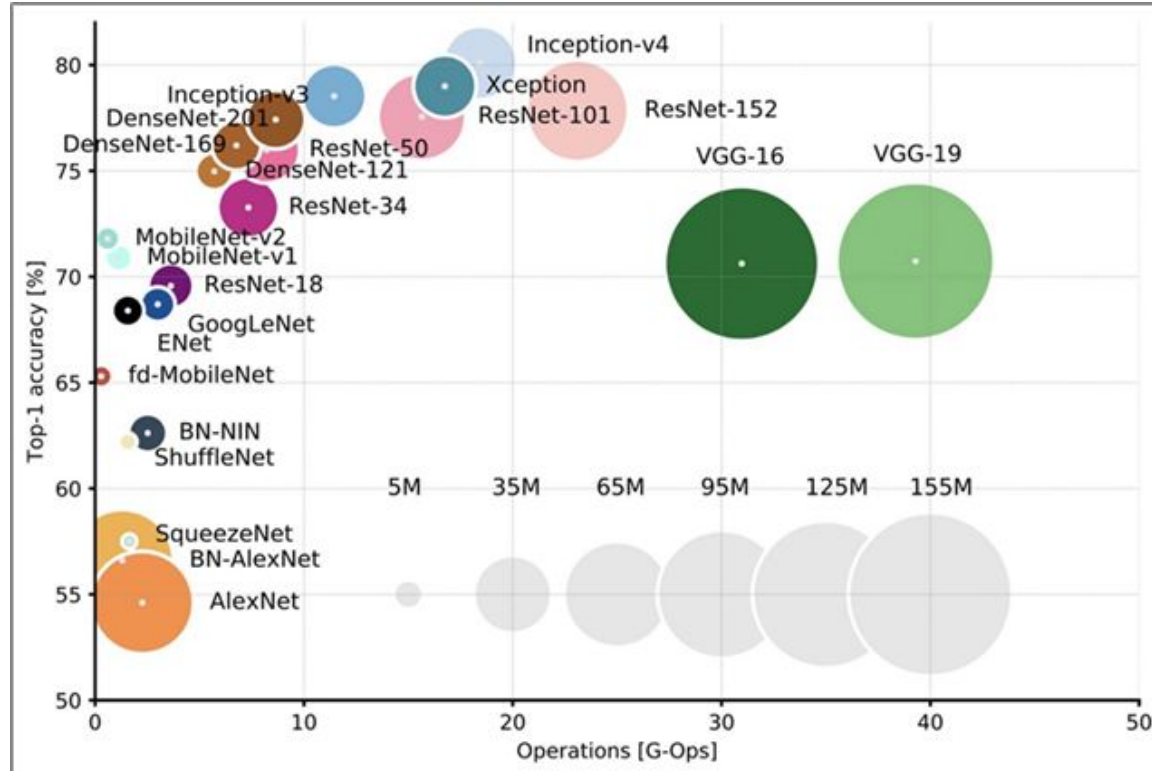


VGGNet

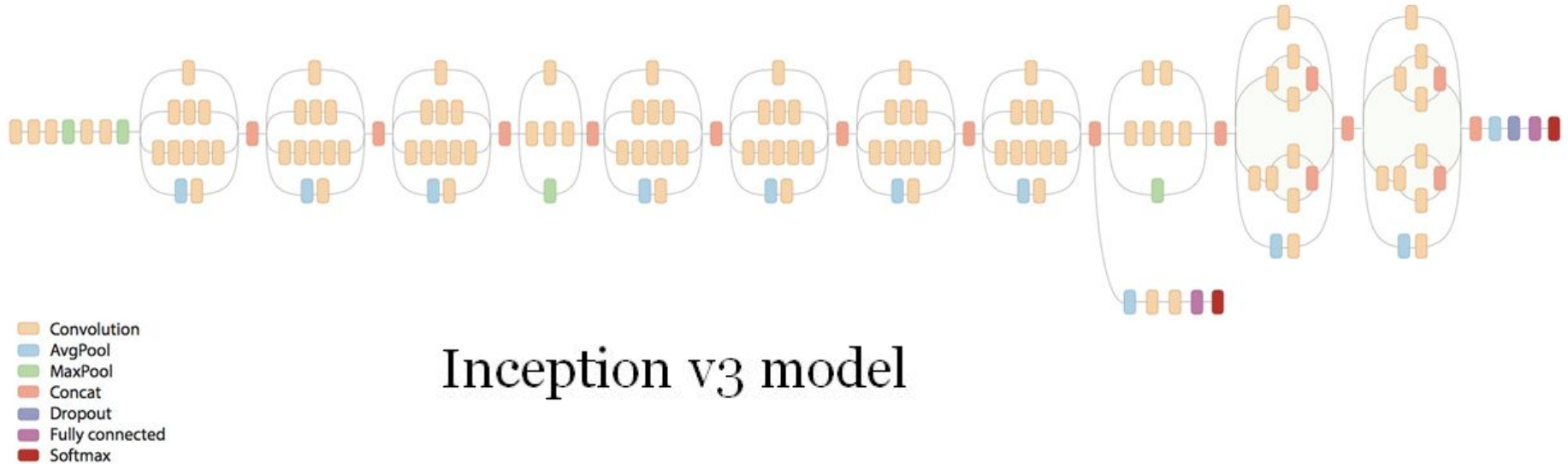




# Deep Convolutional Neural Networks (CNN)



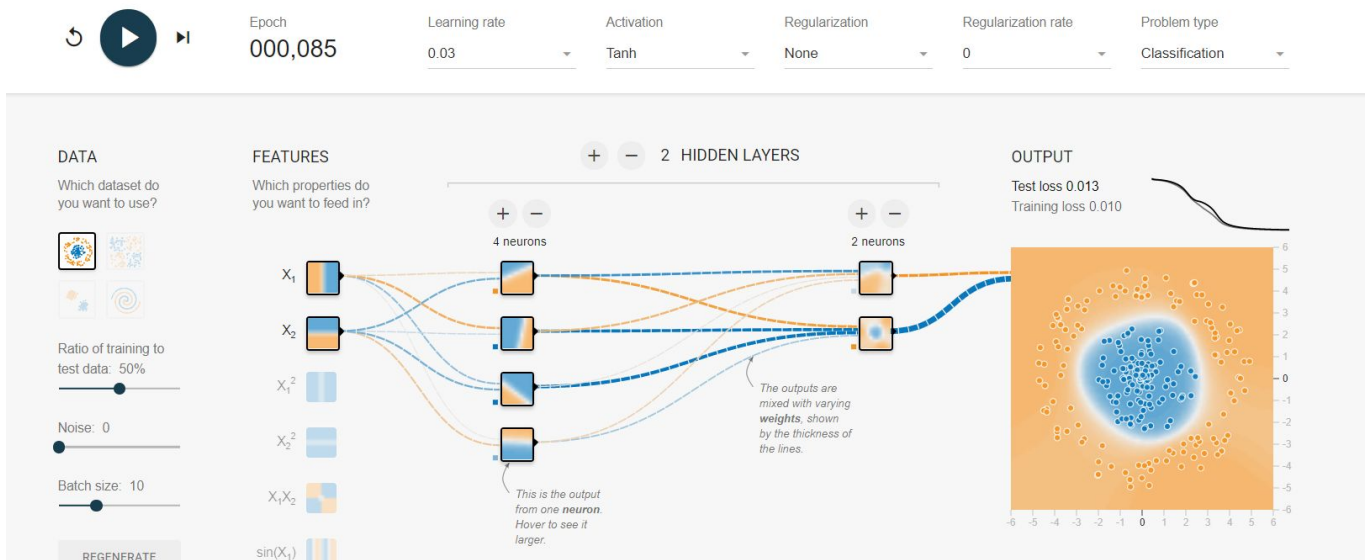
# Deep Convolutional Neural Networks (CNN)



Inception v3 model

# More on Neural Networks - online demo

Tinker With a **Neural Network** Right Here in Your Browser.  
Don't Worry, You Can't Break It. We Promise.



[Try it here!](#)



# Time to Train a basic CNN

Computing on CPU - 5-6 mins Avg

```
Train on 35000 samples, validate on 5479 samples
Epoch 1/4
35000/35000 [=====] - 91s 3ms/step
Epoch 2/4
35000/35000 [=====] - 90s 3ms/step
Epoch 3/4
35000/35000 [=====] - 90s 3ms/step
Epoch 4/4
35000/35000 [=====] - 90s 3ms/step
```



Computing on GPU - 30s Avg (**10x Faster**)

```
Train on 35000 samples, validate on 5479 samples
Epoch 1/4
35000/35000 [=====] - 11s 309us/step
Epoch 2/4
35000/35000 [=====] - 6s 172us/step -
Epoch 3/4
35000/35000 [=====] - 6s 172us/step -
Epoch 4/4
35000/35000 [=====] - 6s 173us/step -
```





# Wrap-Up

Thank You to the Geospatial Centre  
for this collaboration!

Slide of this workshop: [bit.ly/geoml-2](https://bit.ly/geoml-2)

Please upvote Kernel and dataset :)

Contact us at:

<https://www.facebook.com/uwgeospatial/>

Jaydeep Mistry  
[jaydeep.mistry@uwaterloo.ca](mailto:jaydeep.mistry@uwaterloo.ca)

Juan Carrillo  
[jmcarril@uwaterloo.ca](mailto:jmcarril@uwaterloo.ca)