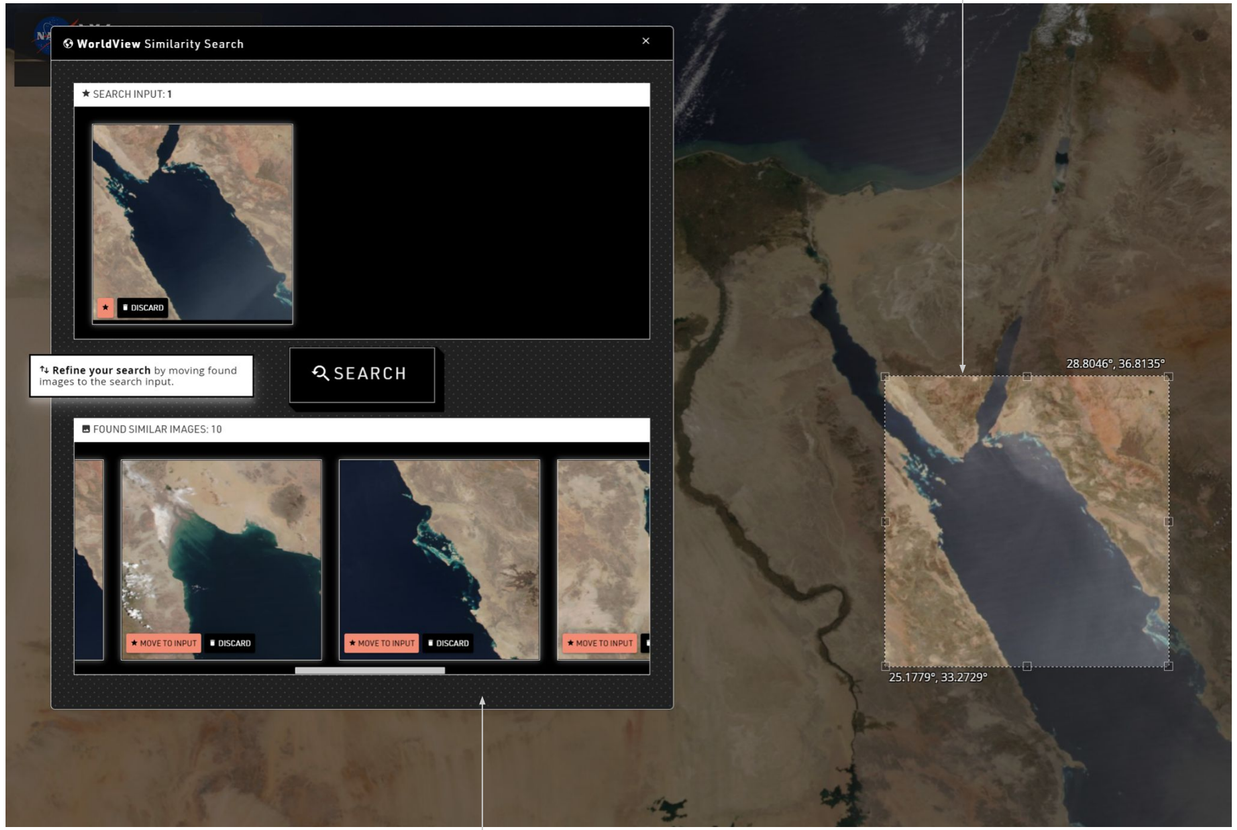
SPACEML

WorldView Search

Searching unlabeled petabyte-scale imagery with self-supervised Machine Learning.



FDL FRONTIER DEVELOPMENT LAB

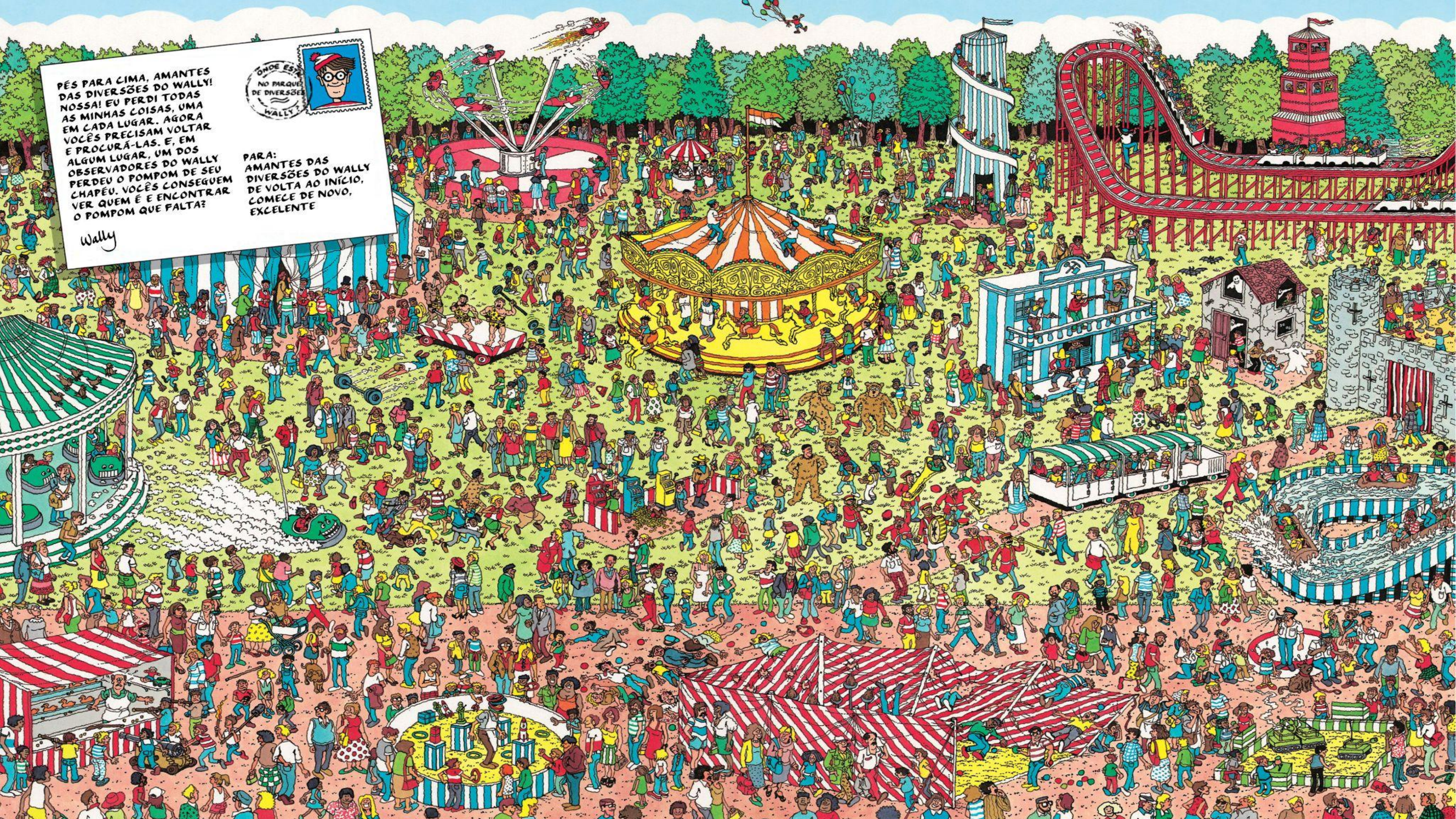
NASA PARTNER

Frontier Development Lab (FDL) is a NASA co-operative agreement with the SETI Institute.

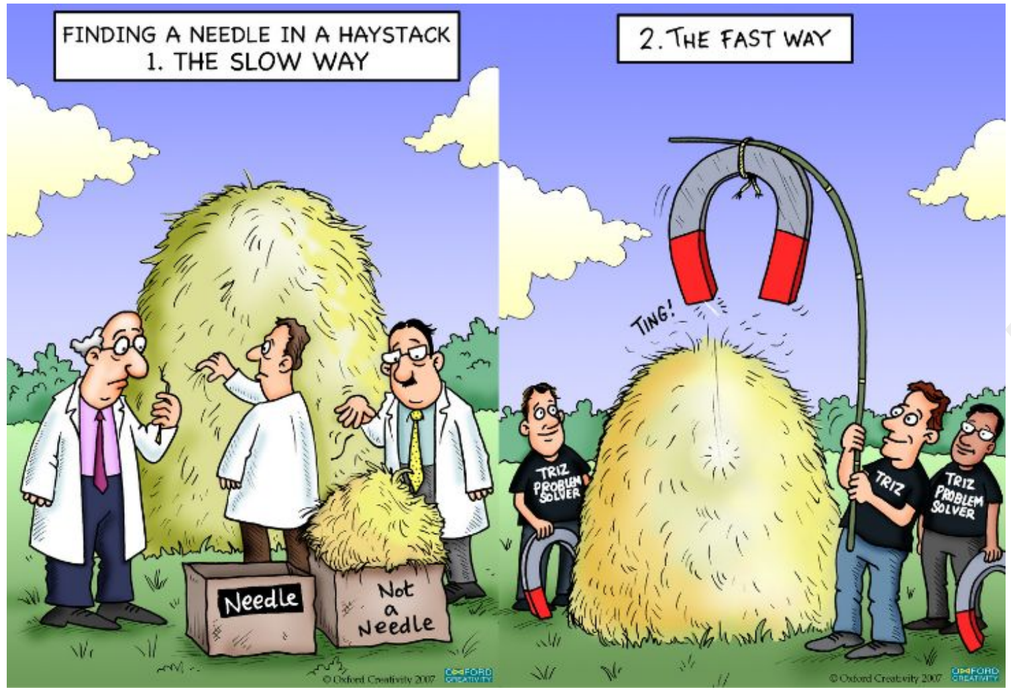
ARTIFICIAL INTELLIGENCE RESEARCH FOR SPACE SCIENCE, EXPLORATION & ALL HUMANKIND

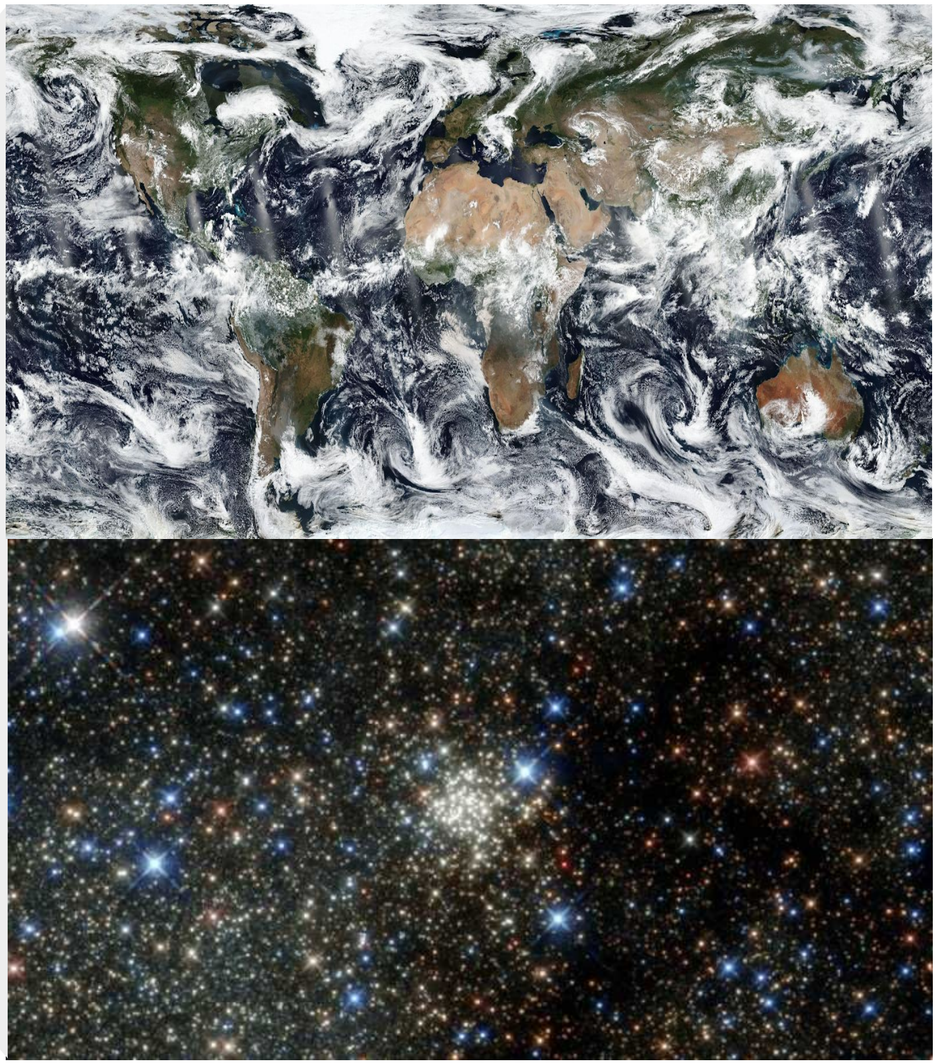
SPACE ML

The Search Problem

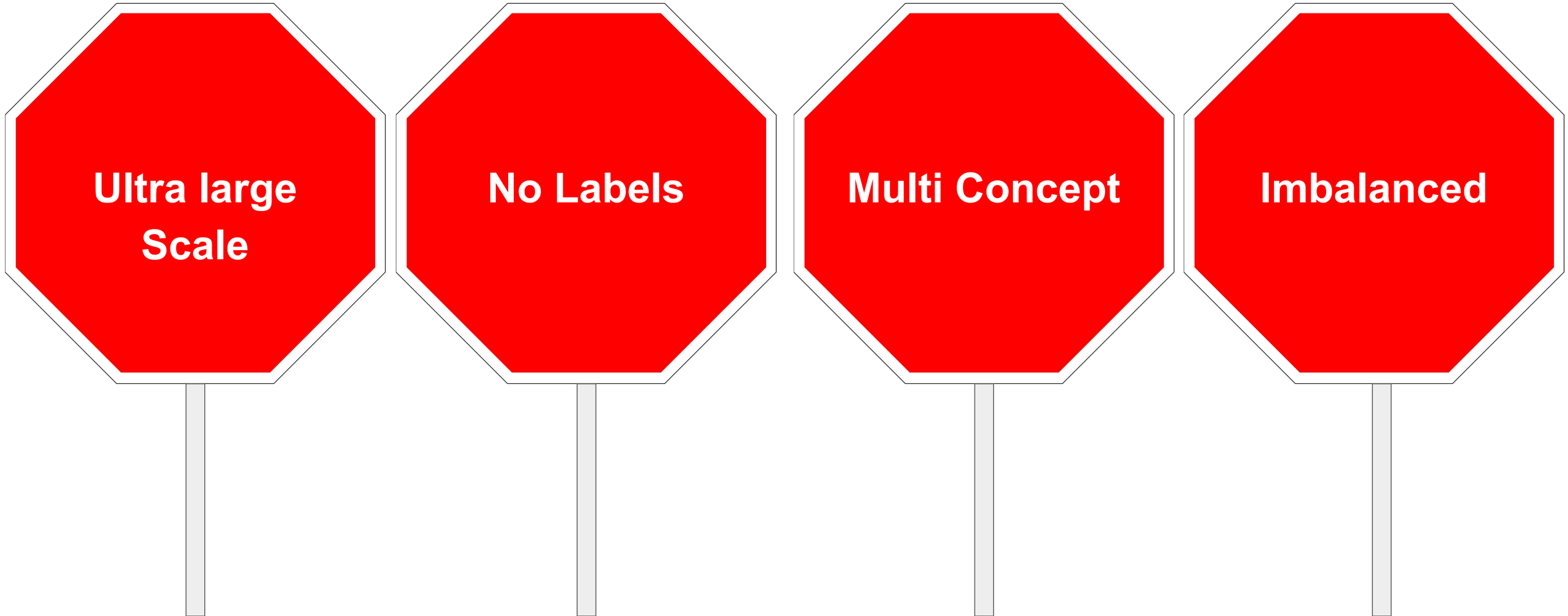

PÉS PARA CIMA, AMANTES DAS DIVERSÕES DO WALLY! NOSSA! EU PERDI TODAS AS MINHAS COISAS, UMA EM CADA LUGAR. AGORA VOCÊS PRECISAM VOLTAR E PROCURÁ-LAS. E, EM ALGUM LUGAR, UM DOS OBSERVADORES DO WALLY PERDEU O POMPOM DE SEU CHAPÉU. VOCÊS CONSEGUEM VER QUEM É E ENCONTRAR O POMPOM QUE FALTA?
Wally
ONDE ESTÁ WALLY?
NO PARQUE DE DIVERSÕES
PARA:
AMANTES DAS DIVERSÕES DO WALLY
DE VOLTA AO INÍCIO, COMECE DE NOVO, EXCELENTE

Finding a needle in a haystack





Why is this a hard problem?

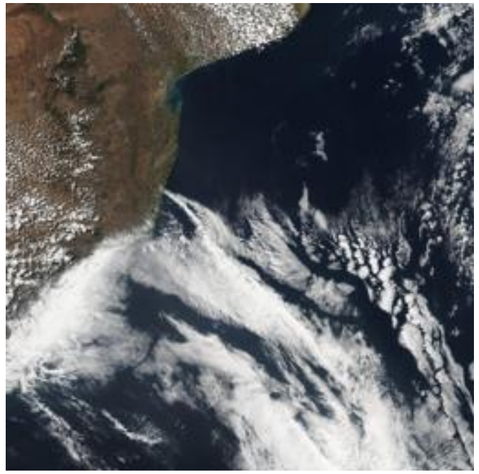


SPACE ML

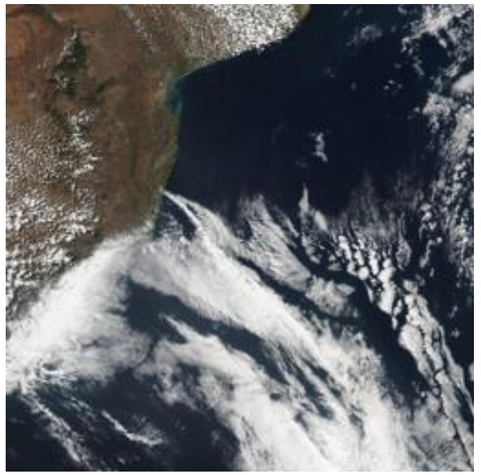
Demo

SPACE ML

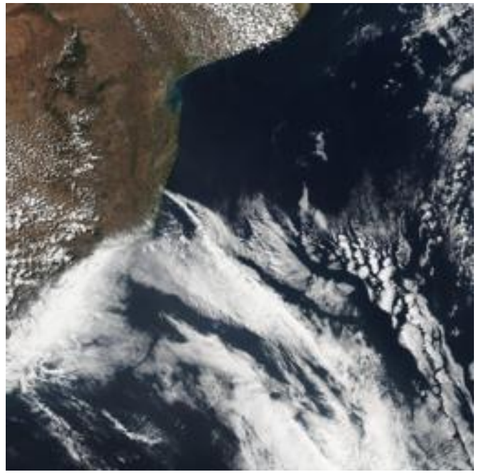
Self Supervised Learning

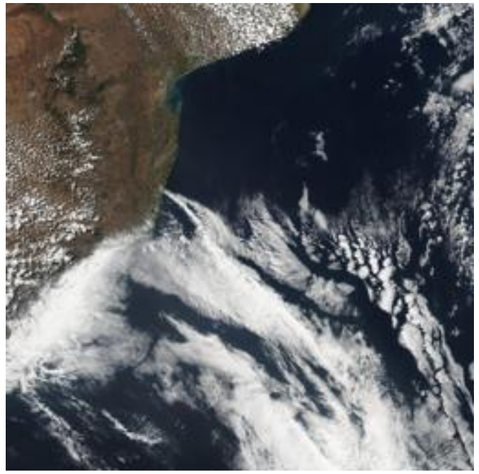


Label:
Ocean, Coastal Fog, Forest

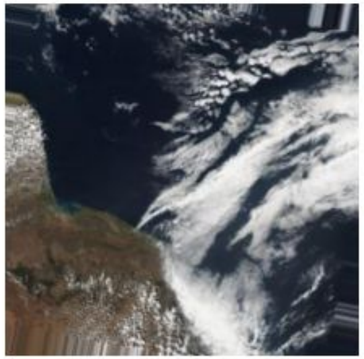


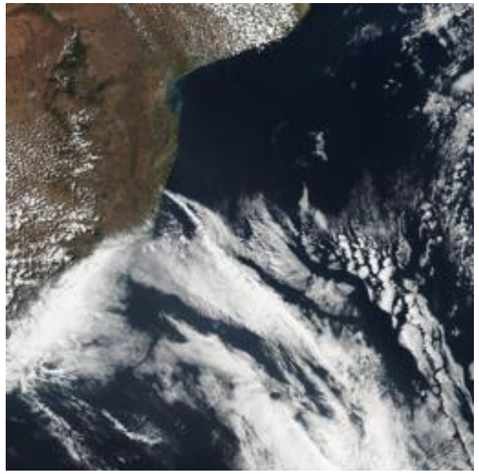
Label:
Ocean, Coastal Fog, Forest



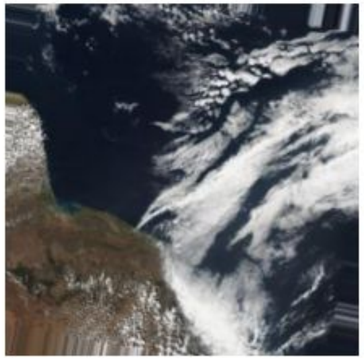


Rotate (0-360°)

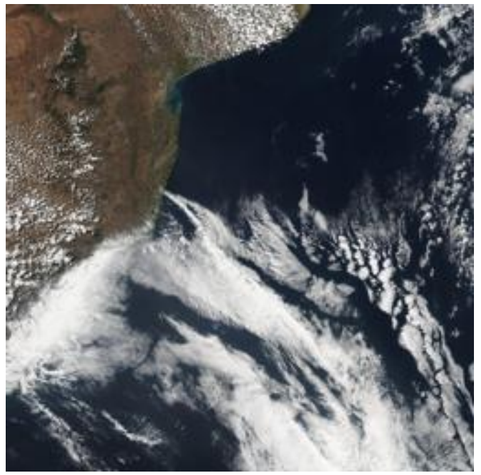




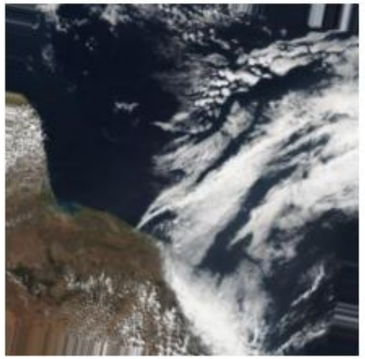
Rotate (0-360°)



Crop

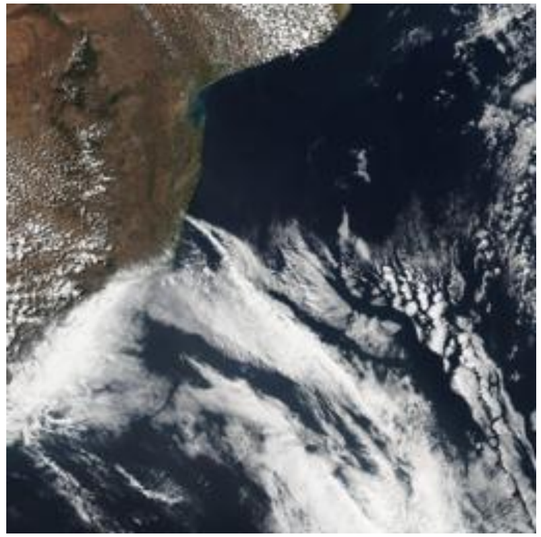


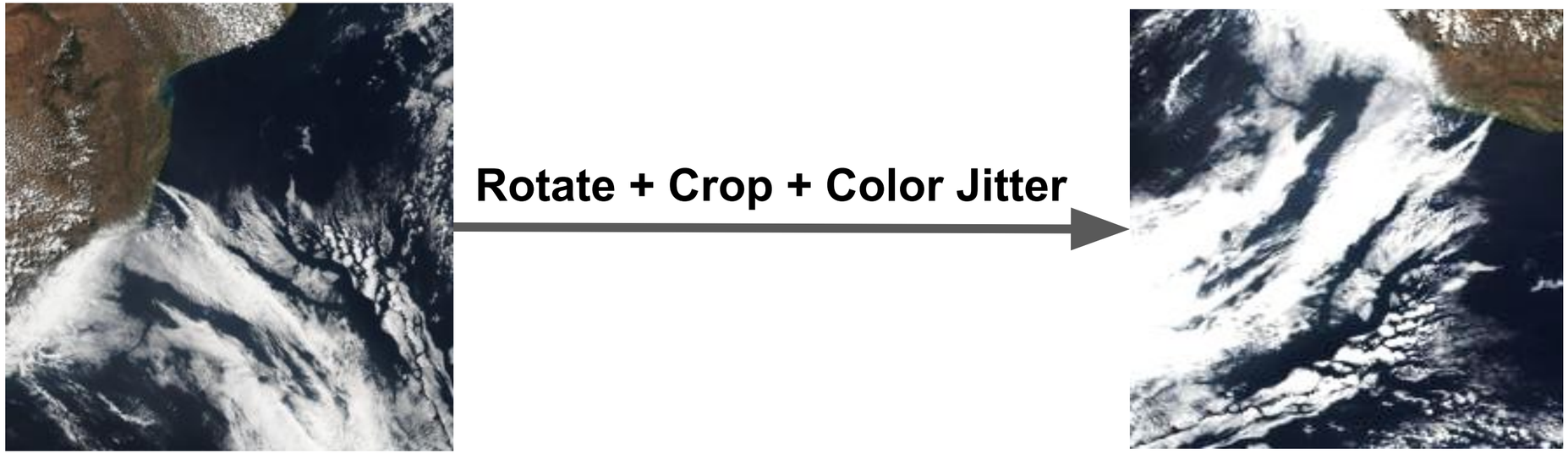
Rotate (0-360°)

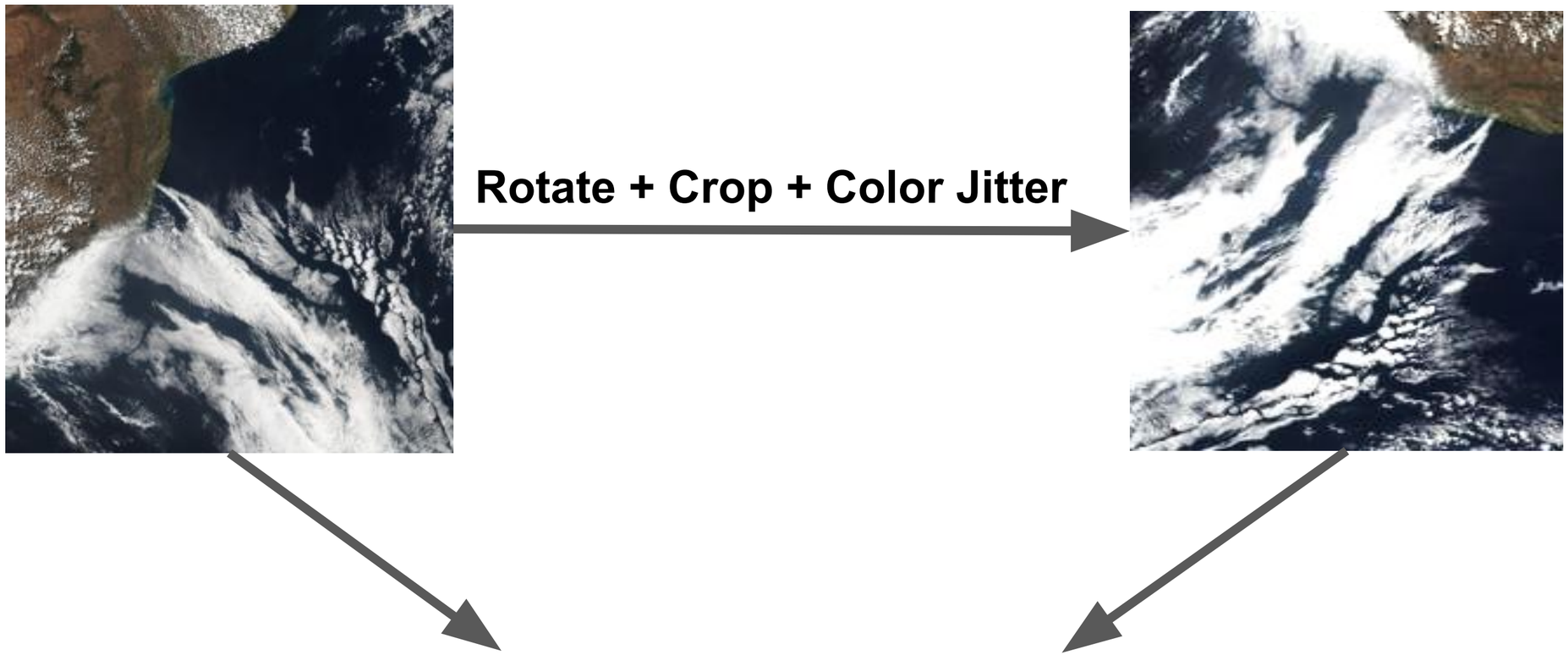


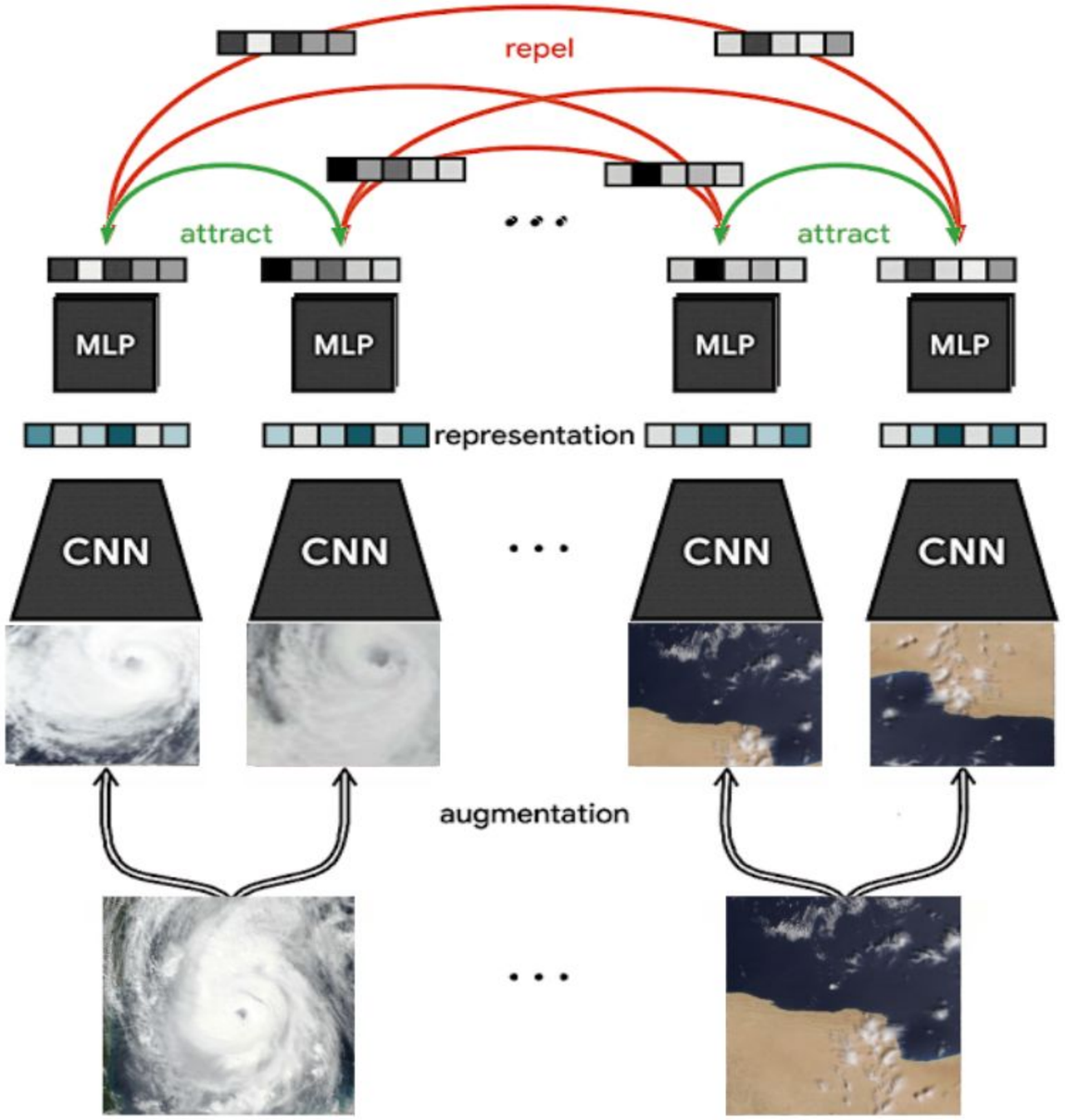
Crop

Color Jitter



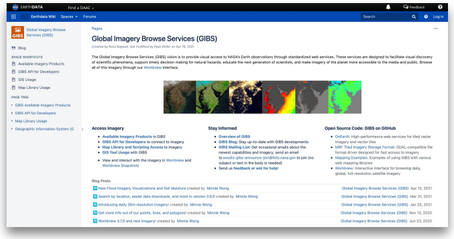

Rotate + Crop + Color Jitter

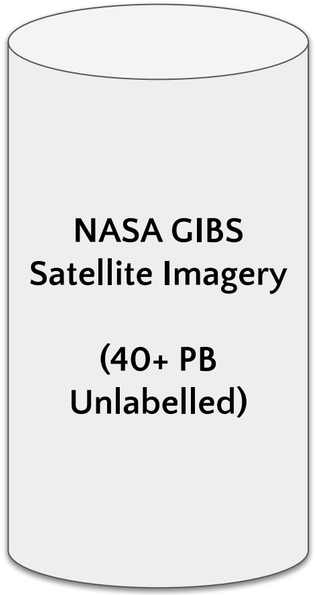

Rotate + Crop + Color Jitter
Teach network to generate
similar embedding

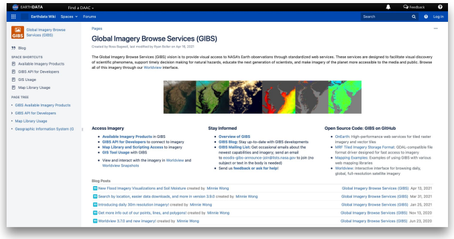

repel
attract
attract
MLP
MLP
MLP
MLP
representation
CNN
CNN
CNN
CNN
augmentation

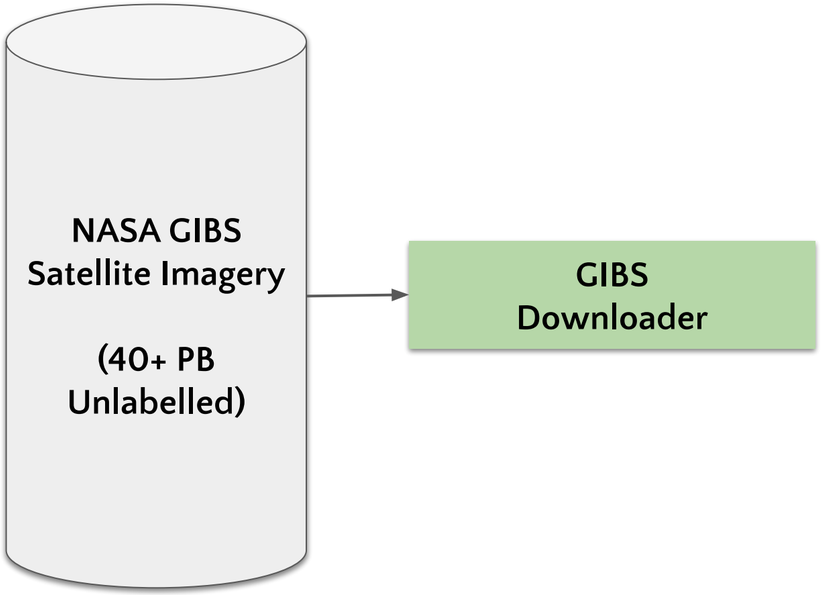
SPACE ML

Worldview Search Pipeline


EARTHDATA
Find a DAAC
Feedback
Earthdata Wiki
Spaces
Forums
Search
Log in
Global Imagery Browse Services (GIBS)
Blog
SPACE SHORTCUTS
Available Imagery Products
GIBS API for Developers
GIS Usage
Map Library Usage
PAGE TREE
GIBS Available Imagery Products
GIBS API for Developers
Map Library Usage
Geographic Information System (I
Pages
Global Imagery Browse Services (GIBS)
Access Imagery
Stay Informed
Open Source Code: GIBS on GitHub
Blog Posts

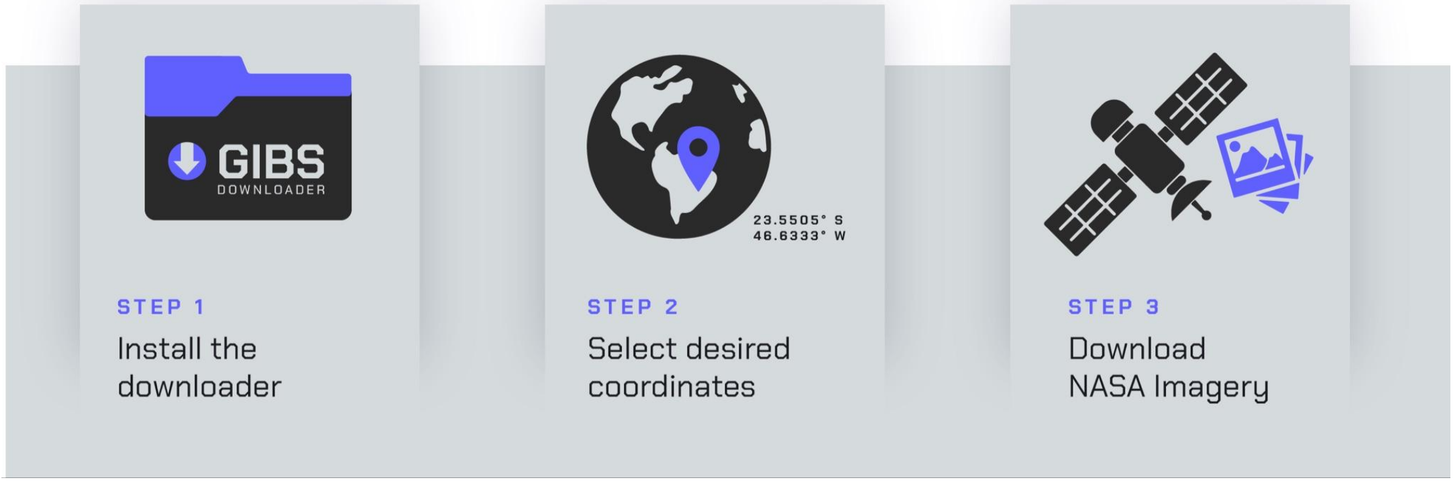

NASA GIBS
Satellite Imagery
(40+ PB
Unlabelled)





GIBS
Downloader

Introducing...

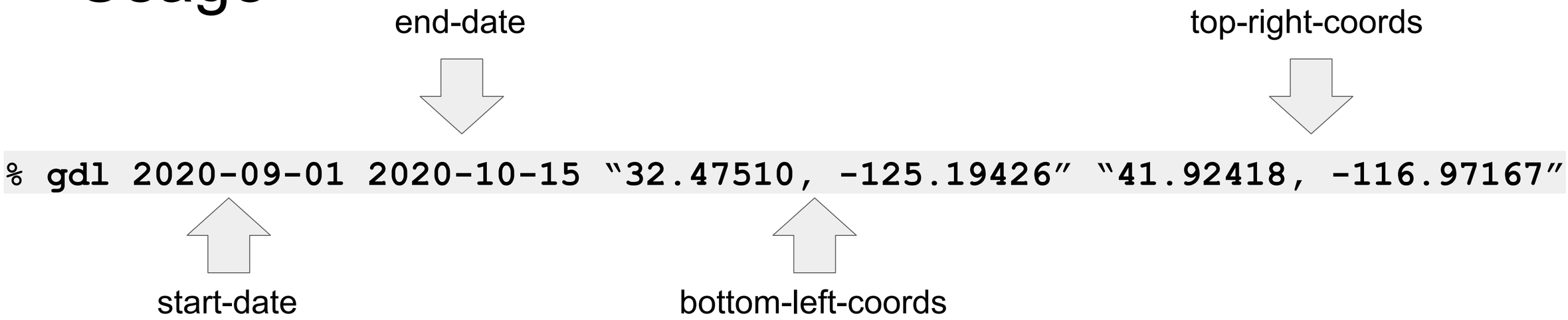


Fast and easy access to decades of data

https://github.com/spaceml-org/NASA-GIBS-Downloader

Researchers: Fernando Lisboa, Navya Sandadi, Shivam Verma

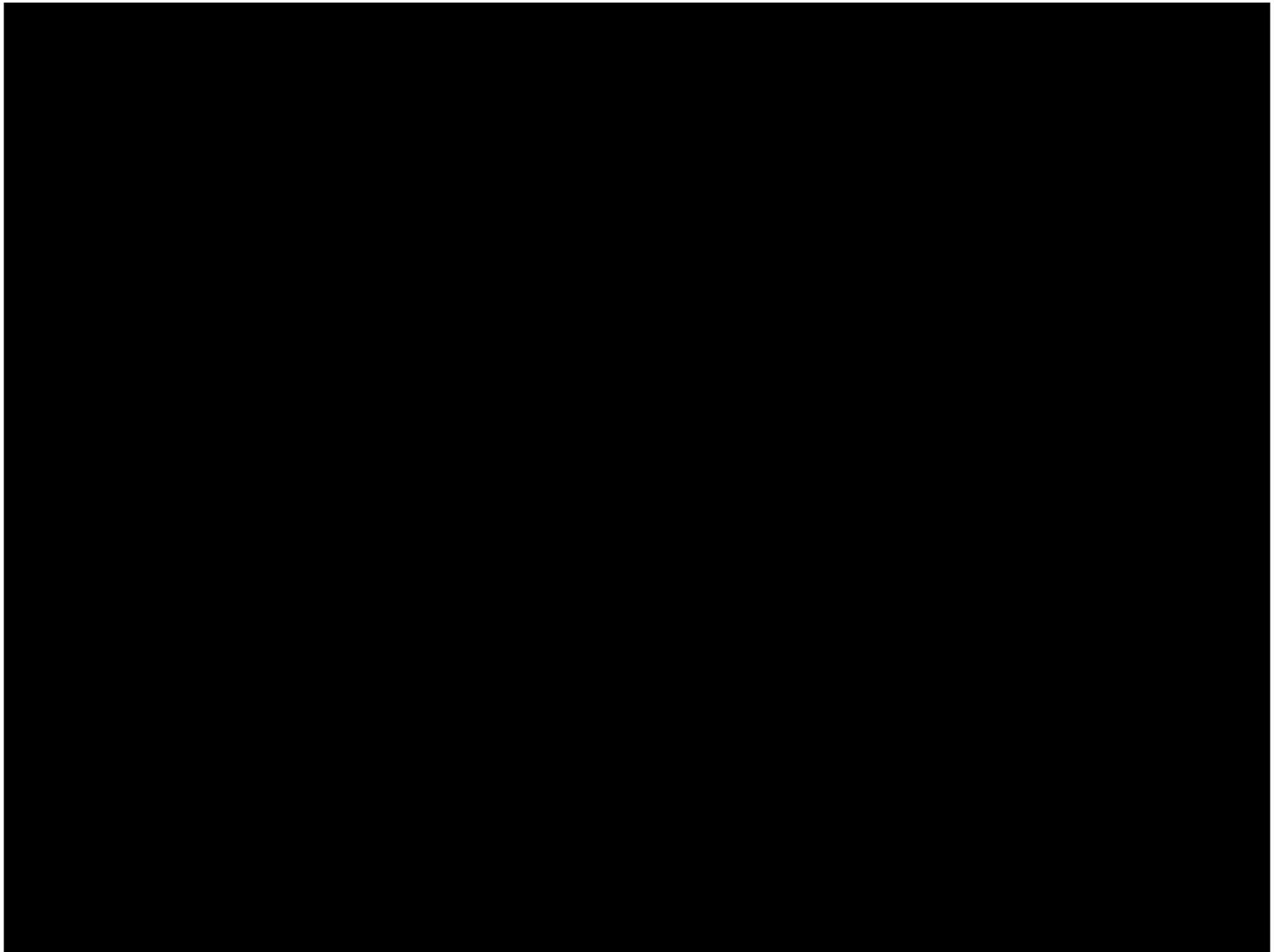
Usage

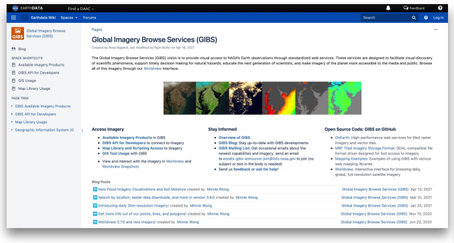


Researchers: Fernando Lisboa, Navya Sandadi, Shivam Verma

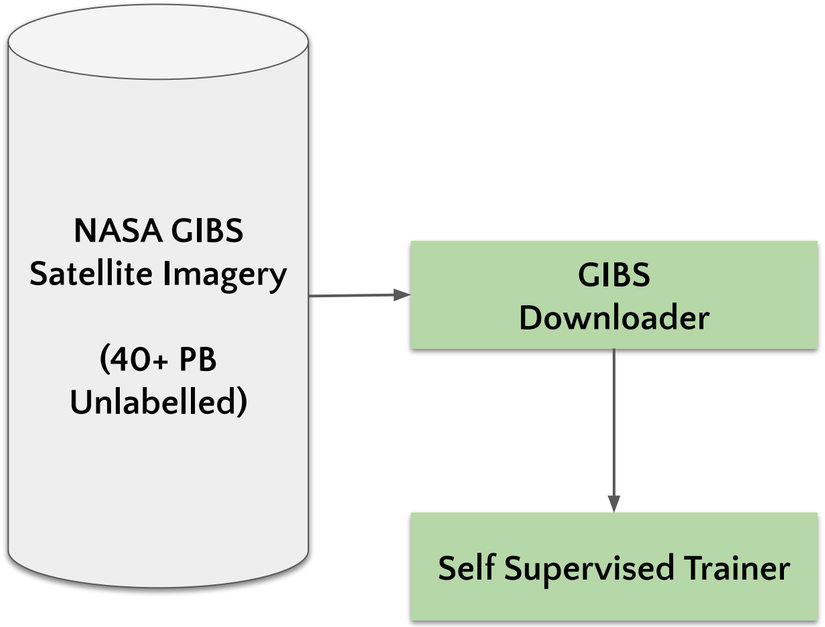
GIBS Downloader

--animate




Global Imagery Browse Services (GIBS)
Access Imagery
Stay Informed
Open Source Code: GIBS on GitHub

GIBS
DOWNLOADER


NASA GIBS
Satellite Imagery
(40+ PB
Unlabelled)
GIBS
Downloader
Self Supervised Trainer

Self Supervised Trainer

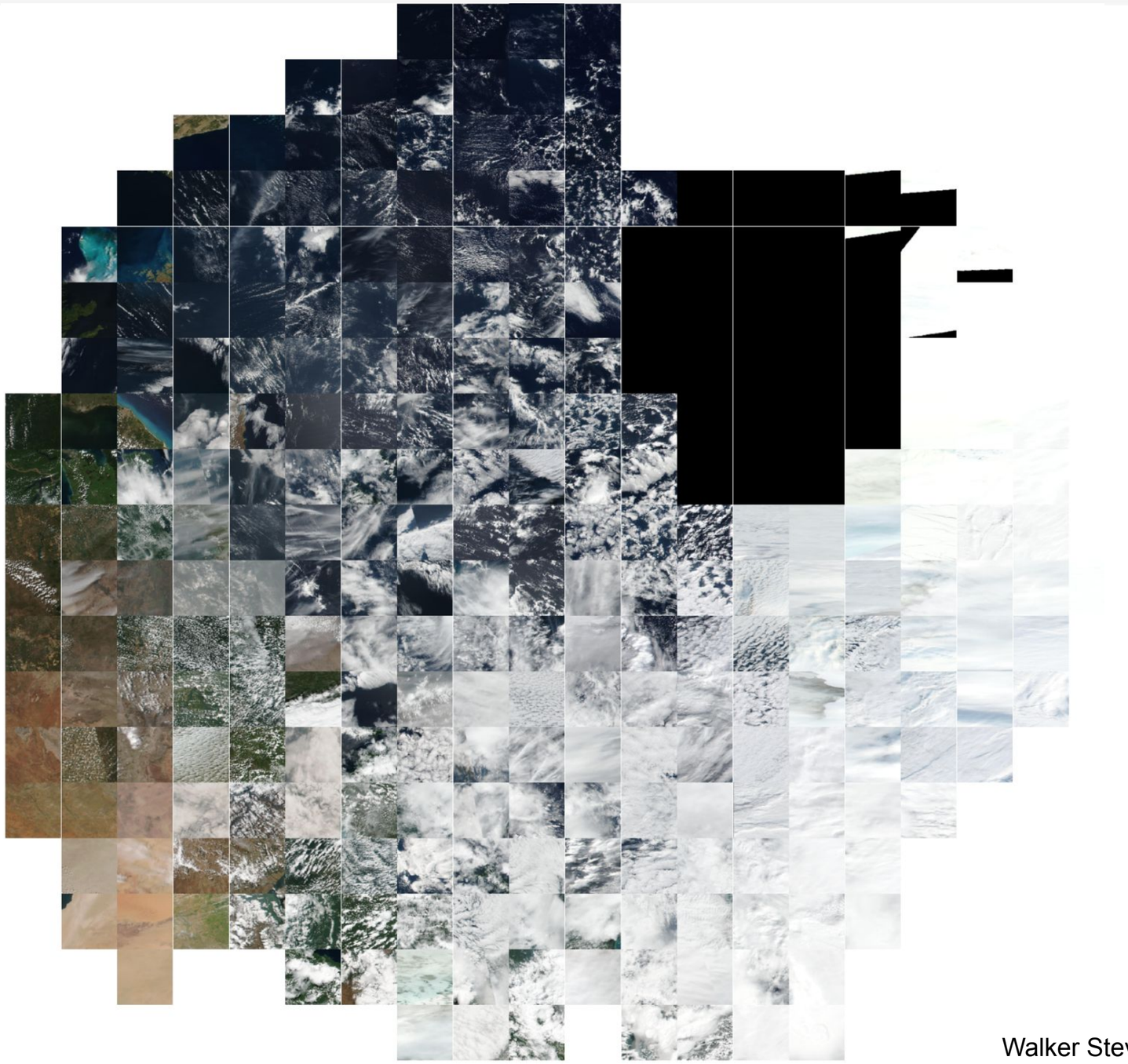
```
python train.py --DATA_PATH myDataFolder/AllImages
```

```
python train.py --technique CLASSIFIER --model ./models/SIMCLR_ssl.ckpt
--DATA_PATH myDataFolder/LabeledImages
```

- Command line, no programming needed
- Modular & upgradable, new SSL algorithms addable in 3 days
- High performance, substituting GPU operations for CPU
- Scales to multiple GPU

https://github.com/spaceml-org/Self-Supervised-Learner

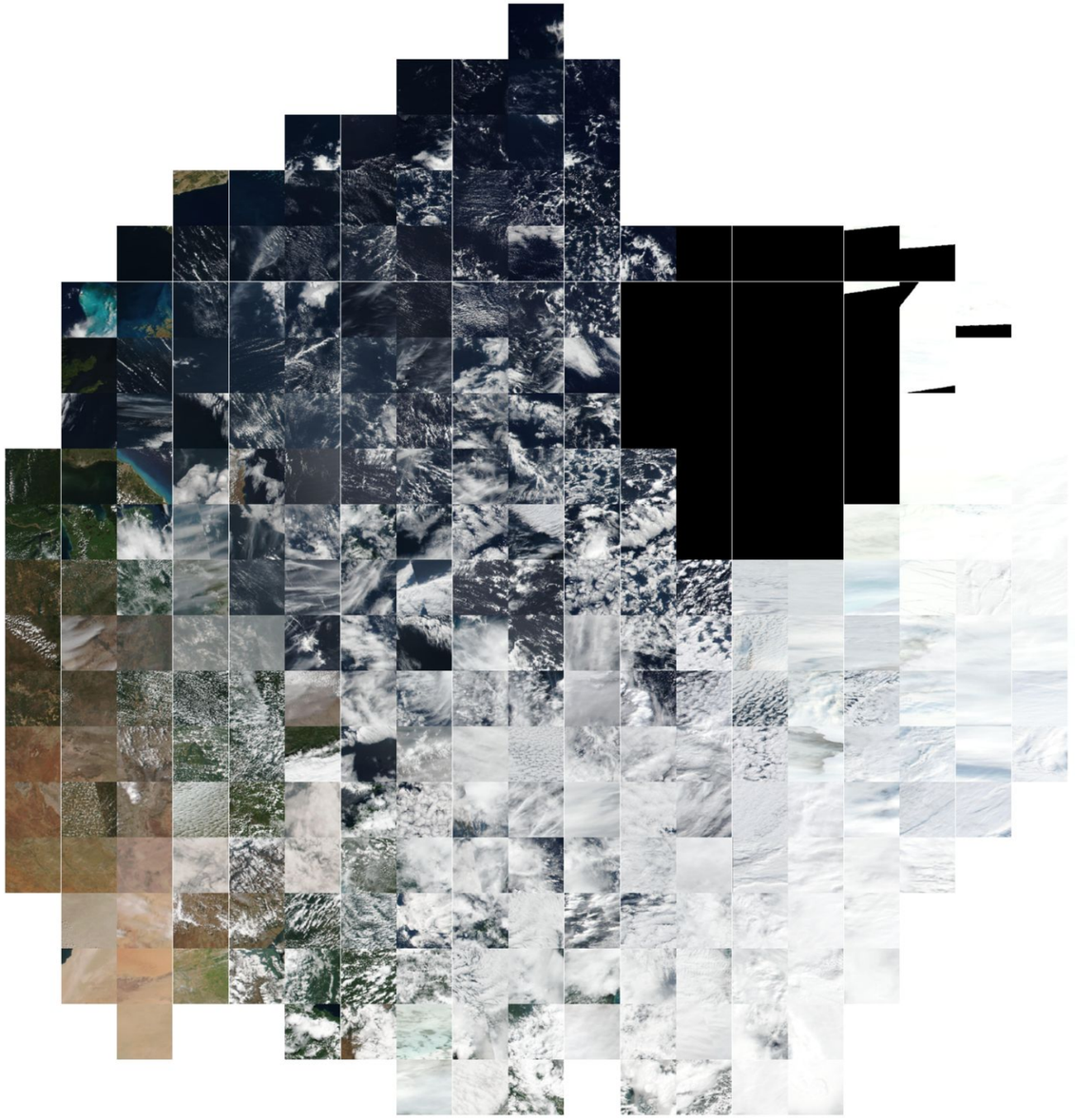
Researcher: Rudy Venguswamy



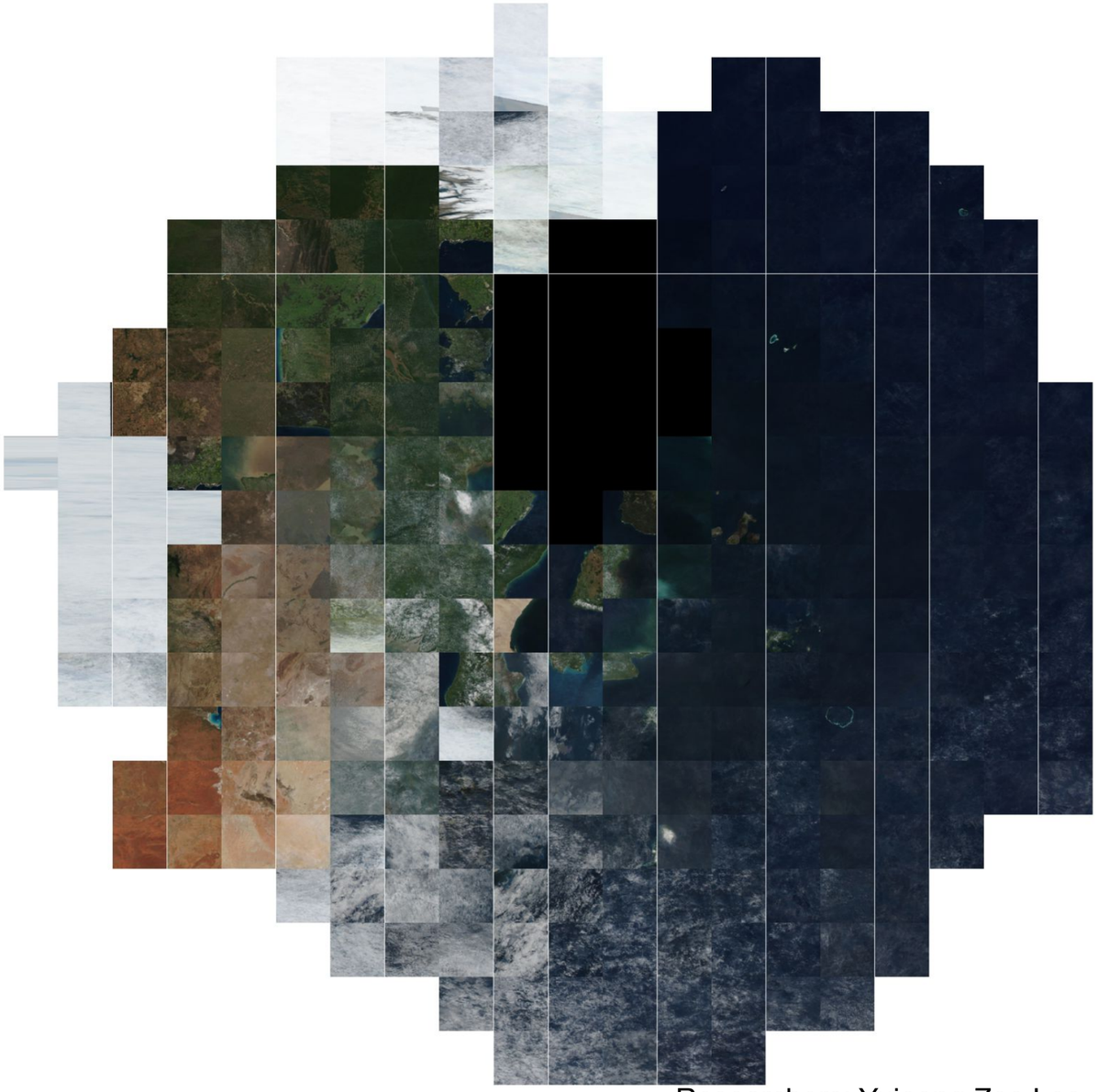
[128 km x 128 km]
VIIRS

Researchers: Yujeong Zoe Lee,
Walker Stevens, Ajay Krishnan, Tarun Narayanan

[128 km x 128 km - VIIRS]

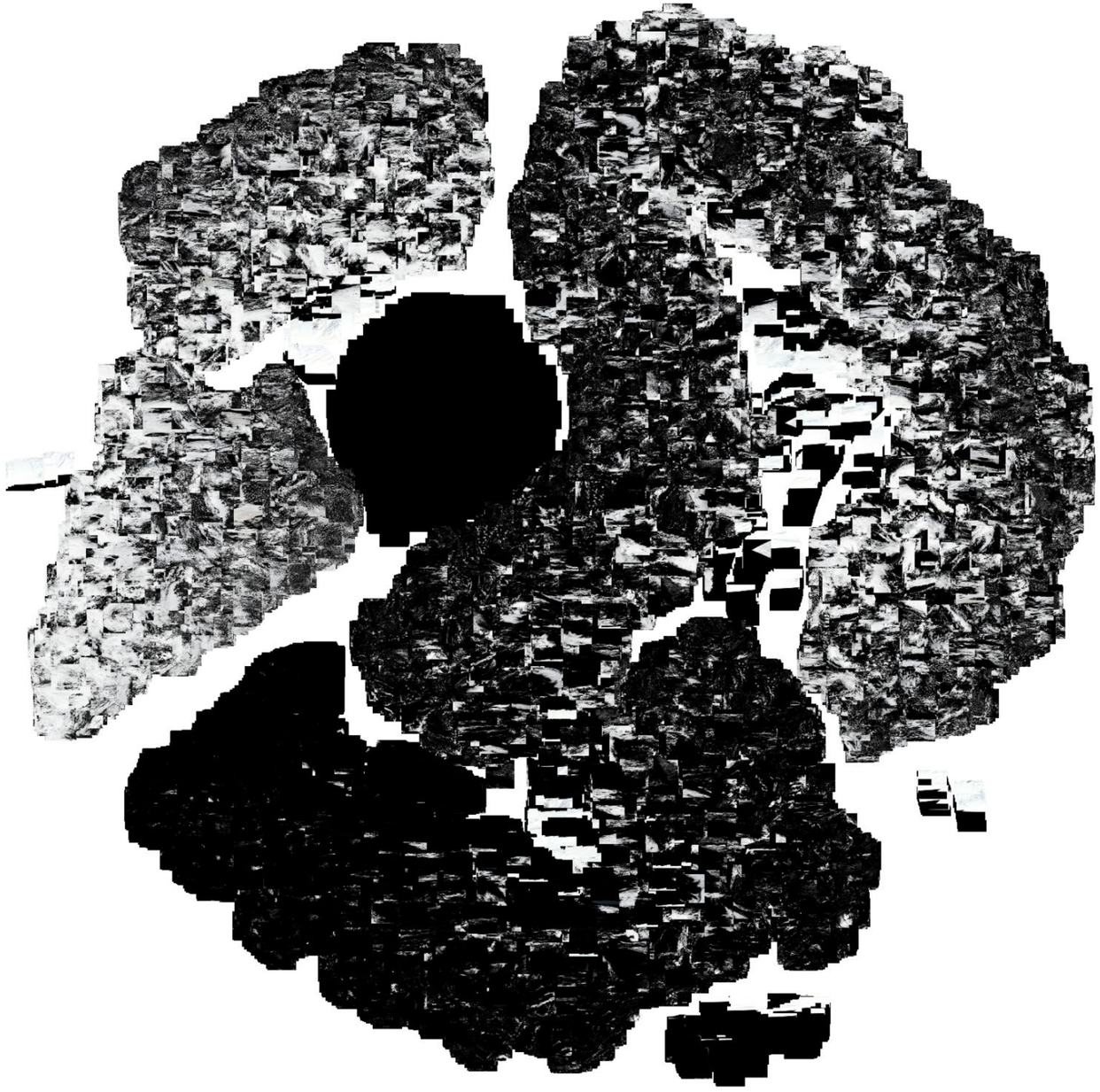


[128 km x 128 km - Cloudless]

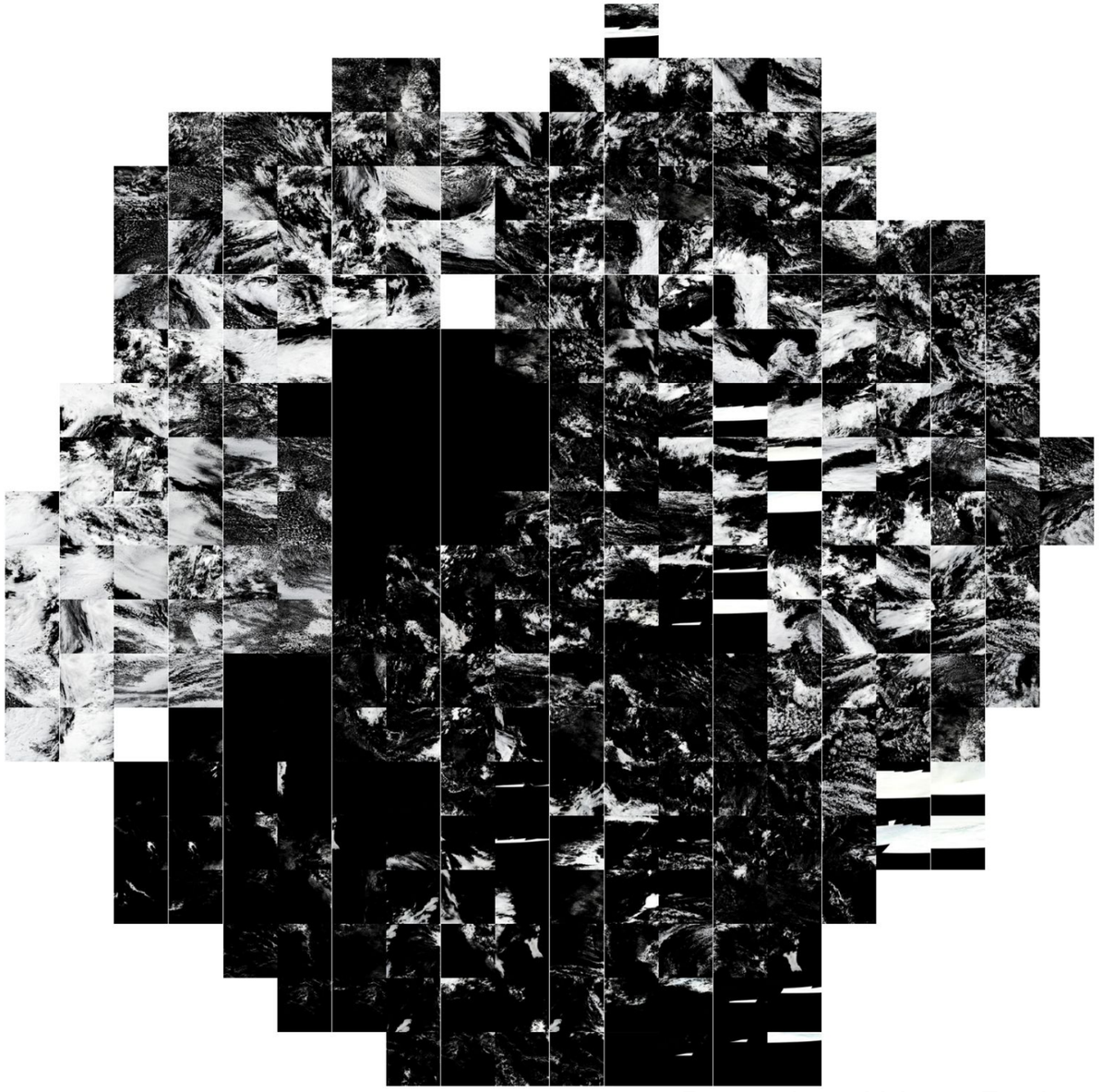


Researchers: Yujeong Zoe Lee,
Walker Stevens, Ajay Krishnan, Tarun Narayanan

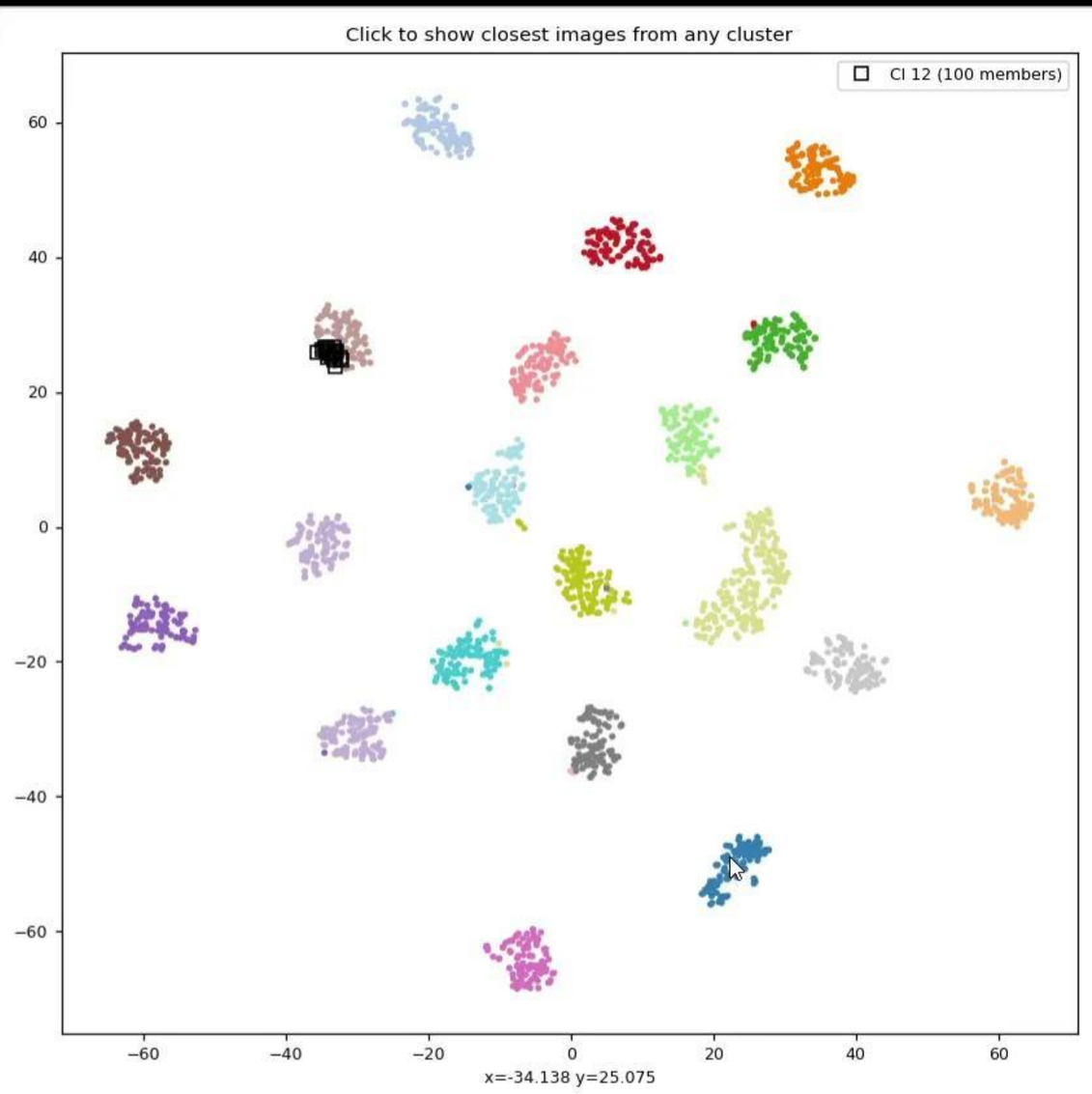
[TSNE plot - Raw]



[TSNE plot - Grid]



Researchers: Yujeong Zoe Lee,
Walker Stevens, Ajay Krishnan, Tarun Narayanan, Surya Ambardar, Aaron Banze


Click to show closest images from any cluster
Cl 12 (100 members)
60
40
20
0
−20
−40
−60
−60
−40
−20
0
20
40
60
x=-34.138 y=25.075

parkinglot (-34.1,25.2) 0.000
parkinglot (-33.5,25.4) 0.608
parkinglot (-33.9,25.8) 0.612
parkinglot (-34.3,26.0) 0.877
parkinglot (-33.2,24.7) 0.941
parkinglot (-33.1,25.7) 1.132
parkinglot (-33.1,26.1) 1.307
parkinglot (-34.8,26.4) 1.448
parkinglot (-34.1,26.7) 1.545
parkinglot (-32.9,26.2) 1.576
parkinglot (-34.4,26.7) 1.581
parkinglot (-33.1,23.9) 1.601
parkinglot (-35.6,25.9) 1.707
parkinglot (-33.2,26.7) 1.794
parkinglot (-32.3,25.2) 1.809
parkinglot (-32.2,24.7) 1.924

Are these the same image category?

Dataset: Resisc-45

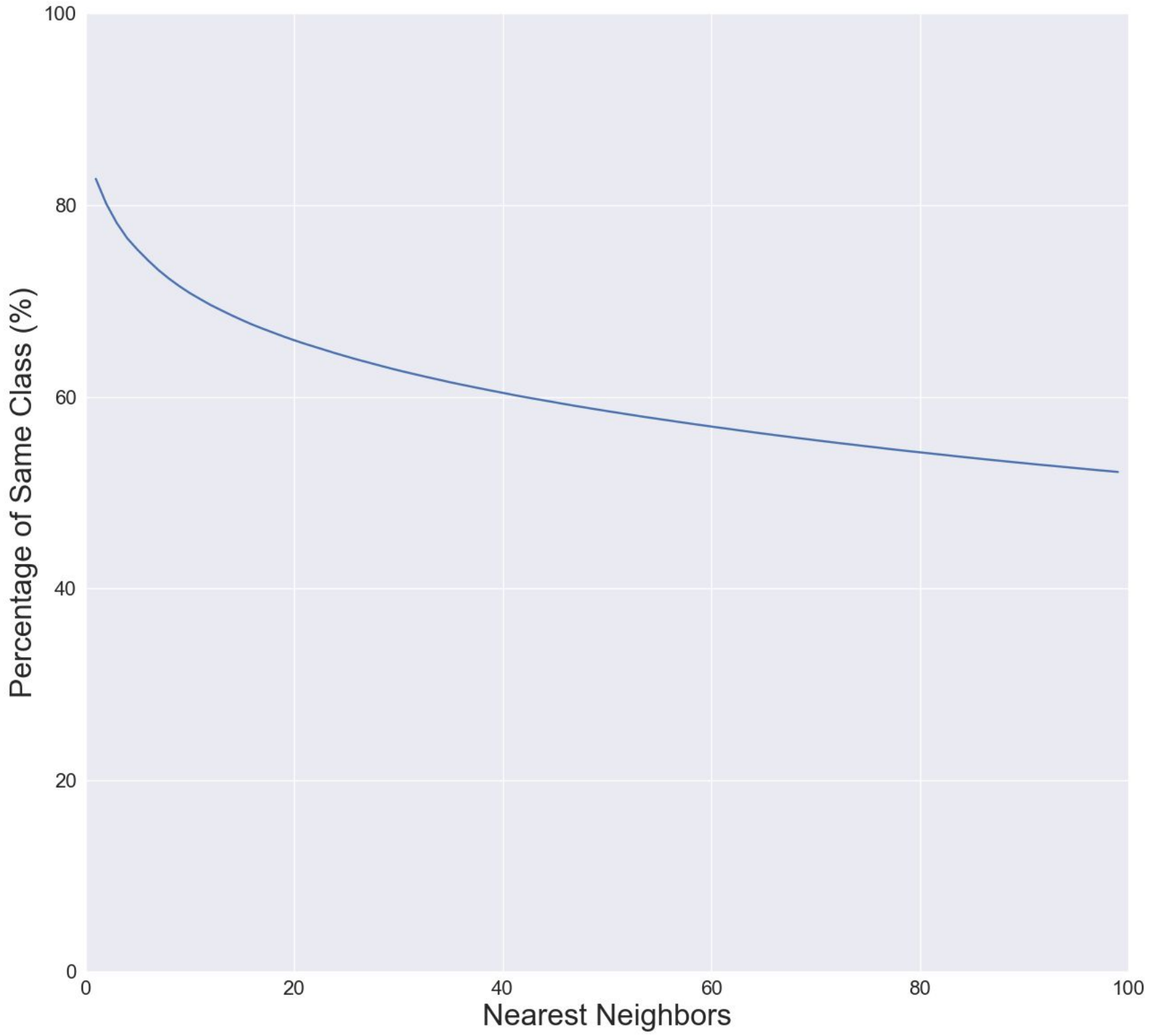
Are these the same image category?

Dataset: Resisc-45

Are these the same image category?

Dataset: Resisc-45

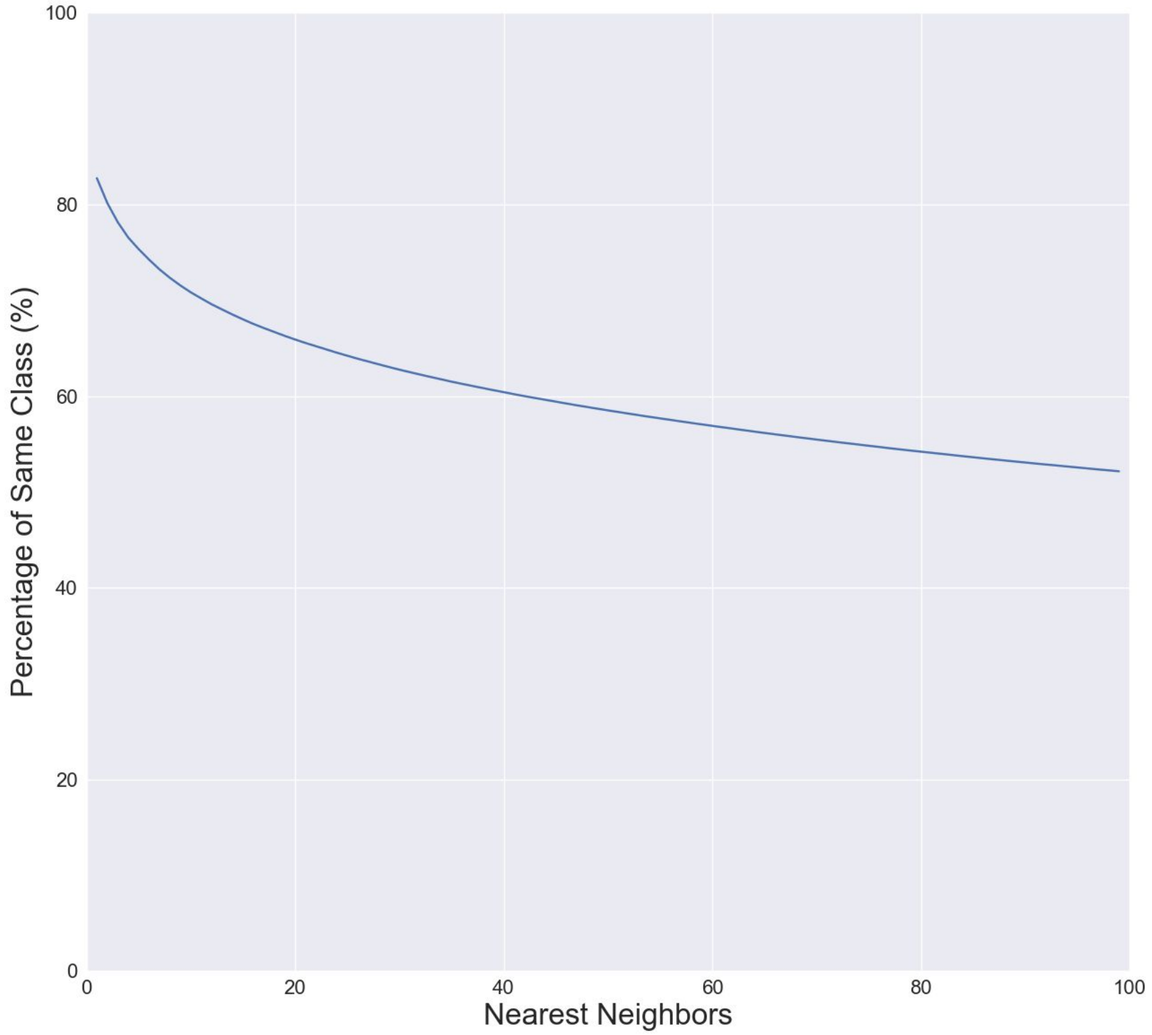
Similarity Search Benchmark



- Dataset: RESISC45 [45 Classes, 31500 images]
- Given an image, for the nearest 100 images, do they belong to the same class as the query image.

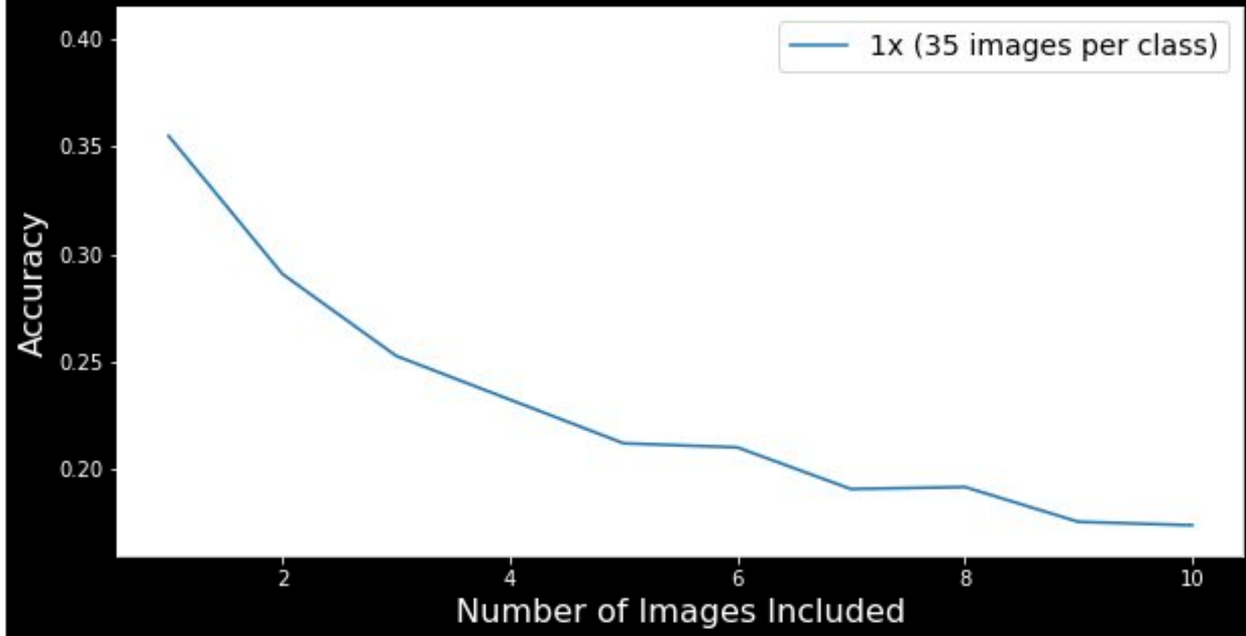
Researcher: Walker Stevens

Circa Aug 2021



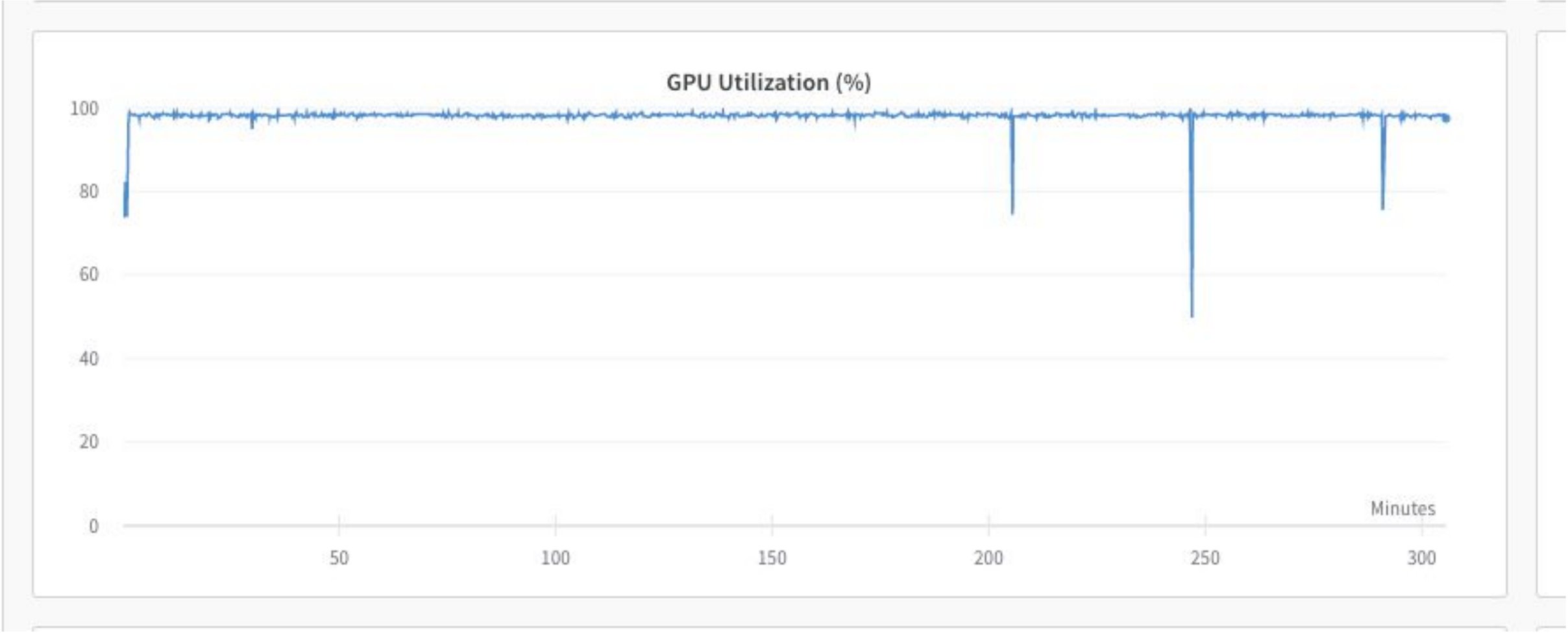
Researcher: Walker Stevens

Circa Oct 2020

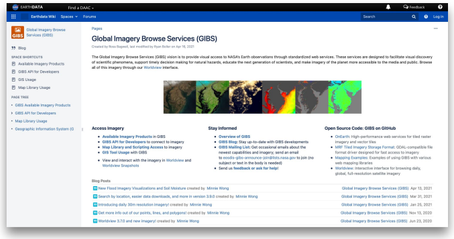


Researcher: Suhas Kotha

Performance



Researcher: Rudy Venguswamy



NASA GIBS Satellite Imagery

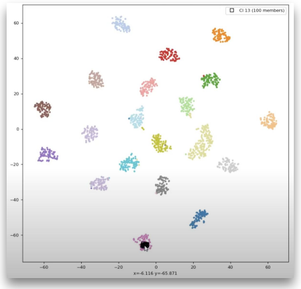
(40+ PB Unlabelled)

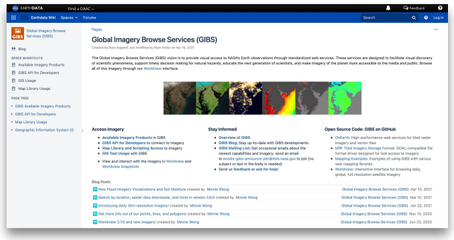
GIBS DOWNLOADER

GIBS Downloader

Self Supervised Trainer

Model Store





NASA GIBS Satellite Imagery

(40+ PB Unlabelled)

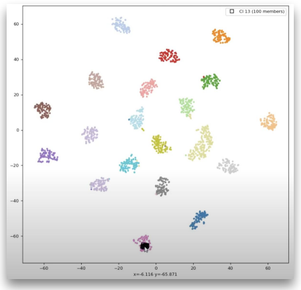
GIBS Downloader

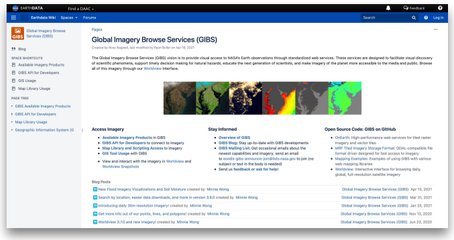
Embedding Store

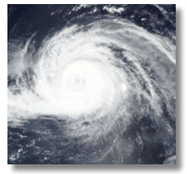
Image Embedding Indexer

Self Supervised Trainer

Model Store







Example Image

NASA GIBS Satellite Imagery

(40+ PB Unlabelled)

GIBS DOWNLOADER

GIBS Downloader

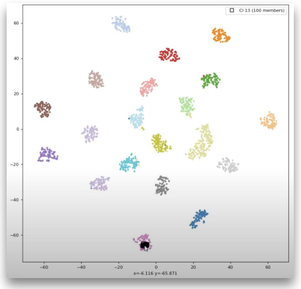
Embedding Store

Nearest Neighbor Search API

Image Embedding Indexer

Self Supervised Trainer

Model Store

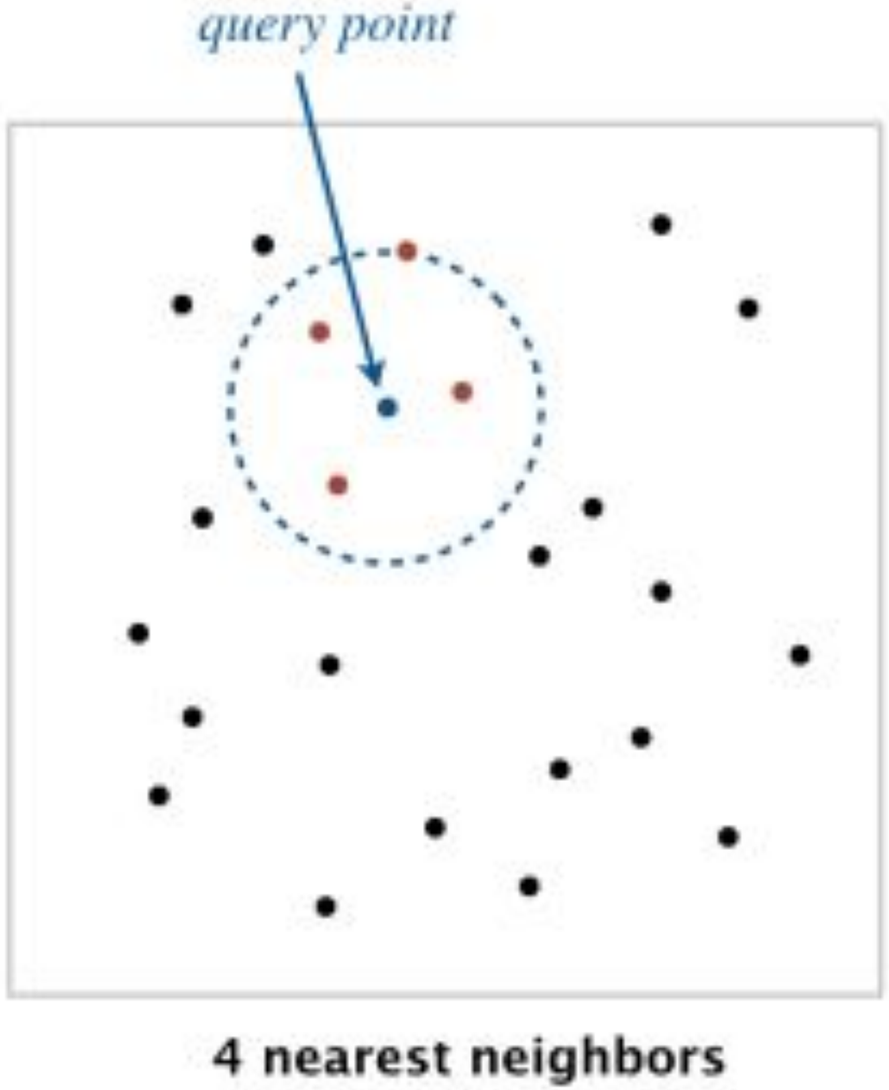


Scalable Indexing and Search API

- Parallized downloading and embedding generation
- Approximate nearest neighbour searching for sub-second million scale search for each index
- Partitioning - Individual Indexes per product, day, resolution
- API accept images (or embeddings), return nearest images
- Decoupled components: Indexes addition without taking API down
- And 30+ cloud tricks for fast but cost effective operations

Researchers: Kai Priester, Stefan Pessolano, Daniela Fragoso, Nathan Hilton
Mentors: Sherin Thomas, Dharini Chandrasekaran

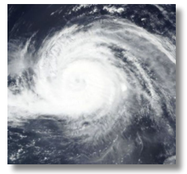
K Nearest Neighbor Search



Approximate nearest neighbor (ANN)

Searching indices intelligently

Researchers: Kai Priester, Stefan Pessolano, Daniela Fragoso, Nathan Hilton
Mentors: Sherin Thomas, Dharini Chandrasekaran



Example Image

Embedding Store

Nearest Neighbor Search API

NASA GIBS Satellite Imagery

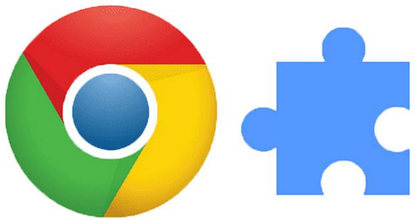
(40+ PB Unlabelled)

GIBS Downloader

Image Embedding Indexer

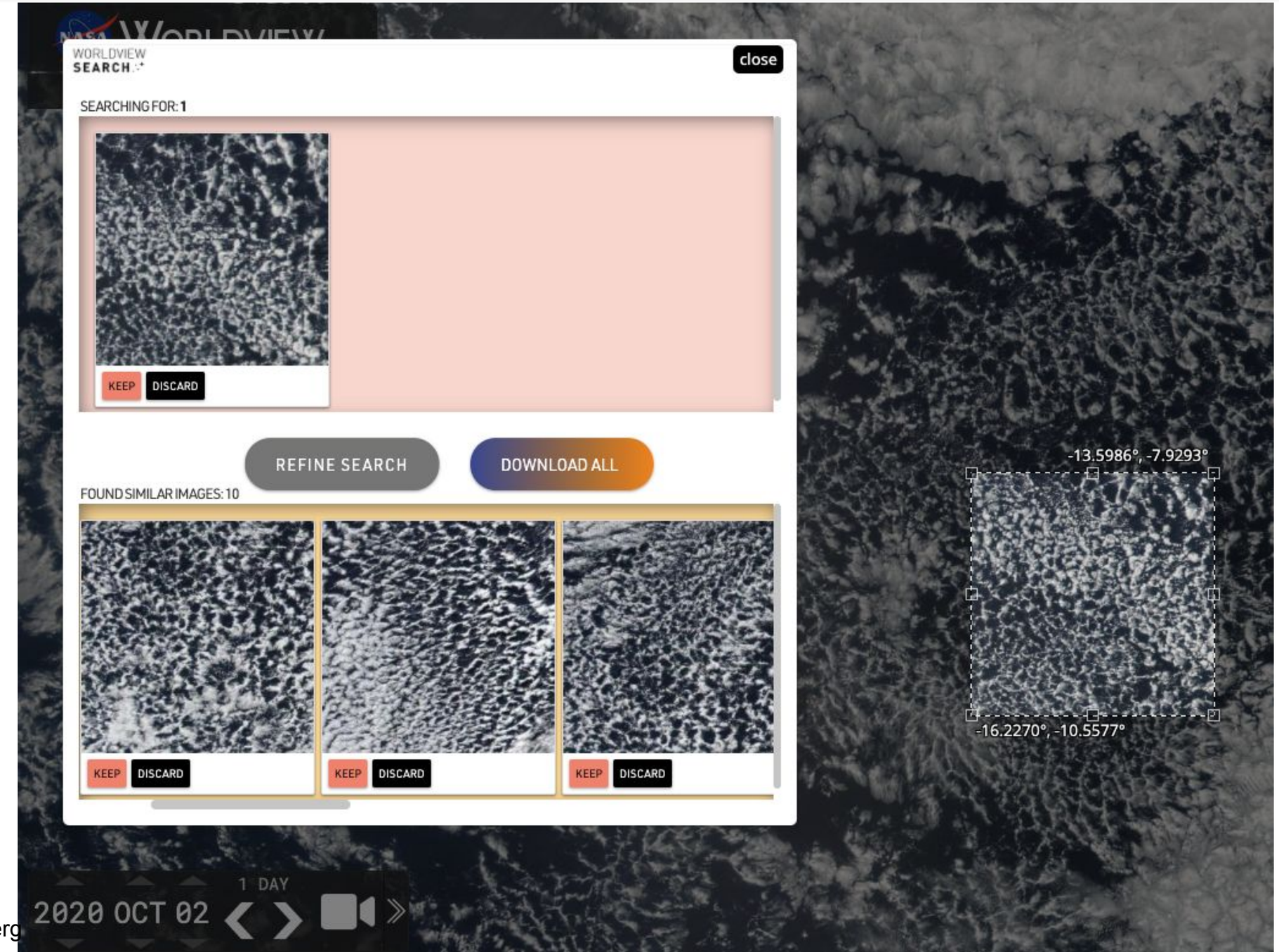
Self Supervised Trainer

Model Store

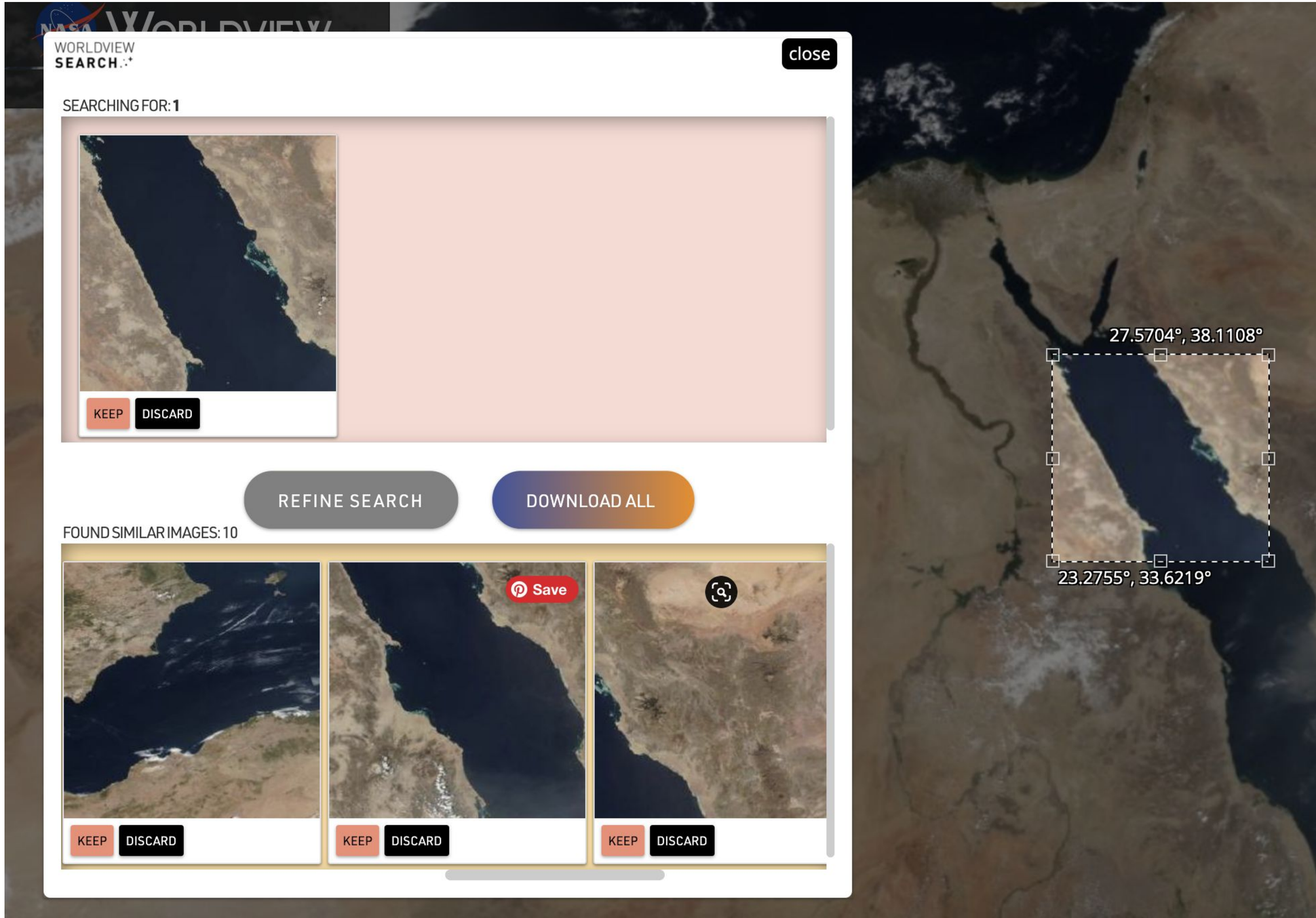


Worldview Search

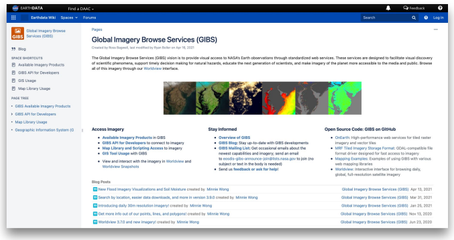
Chrome Extension

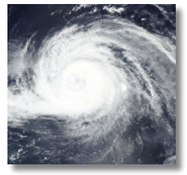


Researchers:
Sumanth Ramesh, Ricardo Lisboa, Leonard Silverberg

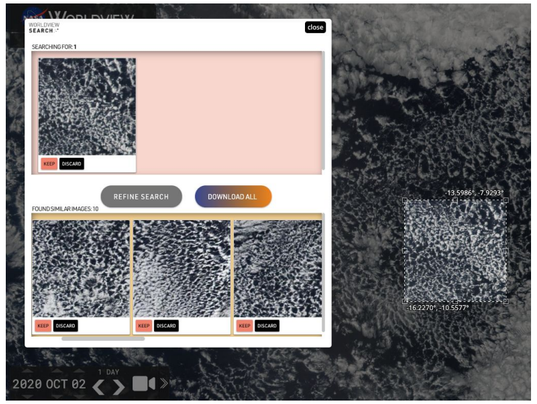


Researchers:
Sumanth Ramesh, Leonard Silverberg





Example Image



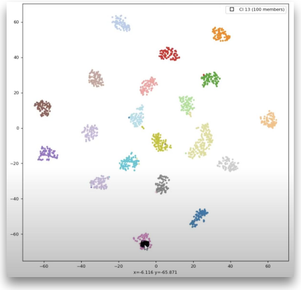
NASA GIBS Satellite Imagery

(40+ PB Unlabelled)

GIBS DOWNLOADER

GIBS Downloader

Self Supervised Trainer



Embedding Store

Image Embedding Indexer

Model Store

Nearest Neighbor Search API

Identify Relevant Examples

Expand Search on Positive Examples

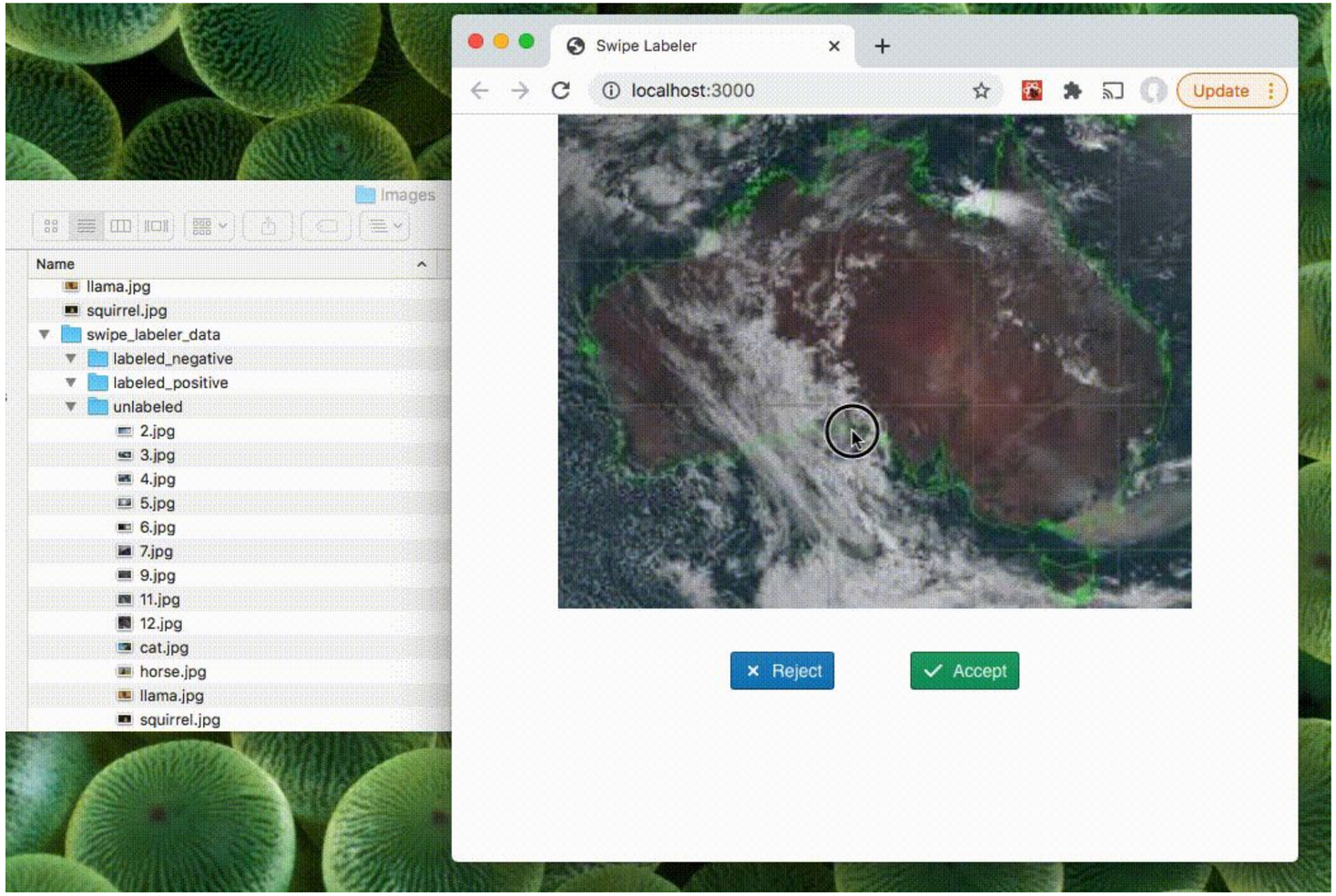
Swipe Data Labeler

SPACE ML

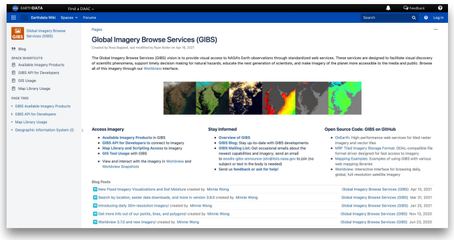
Producing labeled datasets

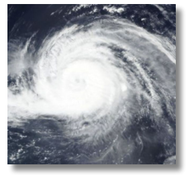
Introducing : “Swipe Labeler”

- Label with Ease
- On Mobile & Website
- Single line command
- Collaborative & Sharable



Researchers:
Sumanth Ramesh, Jenessa Peterson





Example Image

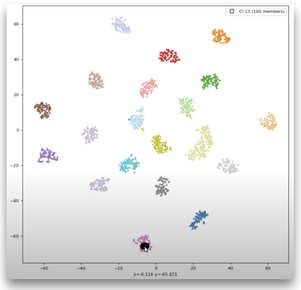
NASA GIBS Satellite Imagery

(40+ PB Unlabelled)

GIBS DOWNLOADER

GIBS Downloader

Self Supervised Trainer



Embedding Store

Image Embedding Indexer

Model Store

Nearest Neighbor Search API

Identify Relevant Examples

Expand Search on Positive Examples

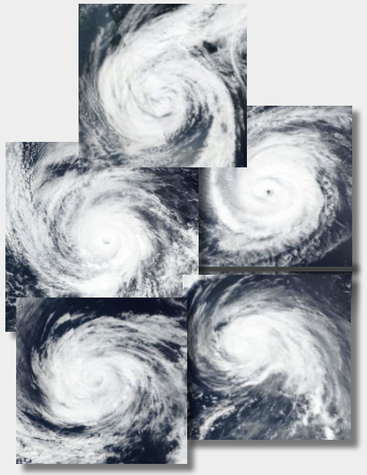
Swipe Data Labeler

Train Classifier

Label Most Confusing Examples

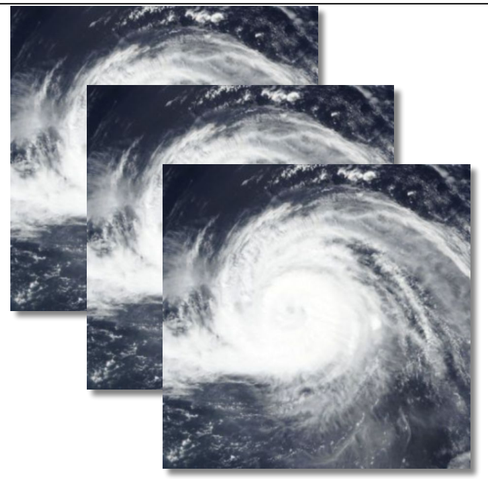
Active Learner

Catalogue Quality Datasets



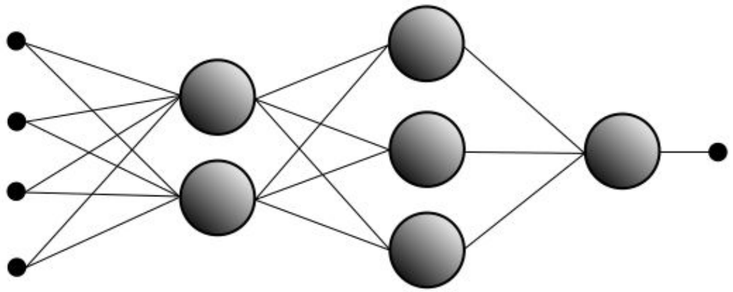
SPACE ML

How do we improve performance?



Similarity Search

Accuracy
30%

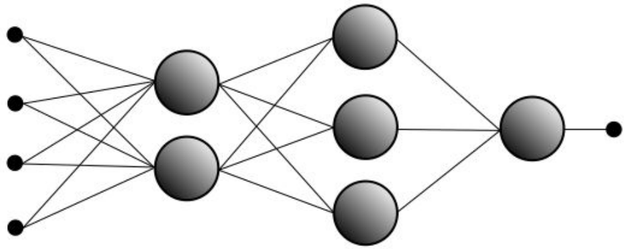


Classifier

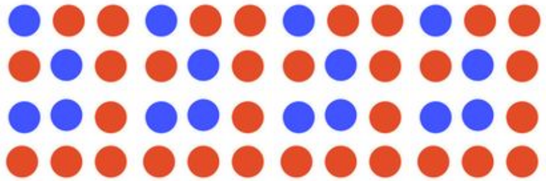
Accuracy
80 - 90%

Researchers: Subhiksha Muthukrishnan, Mandeep Khokhar, Ajay Krishnan, Tarun Narayanan, Rudy Venguswamy

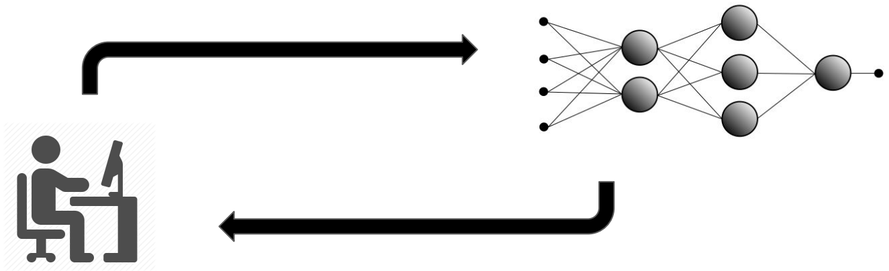
Classifier



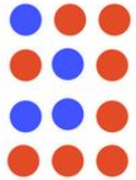
Labeled Data Needed



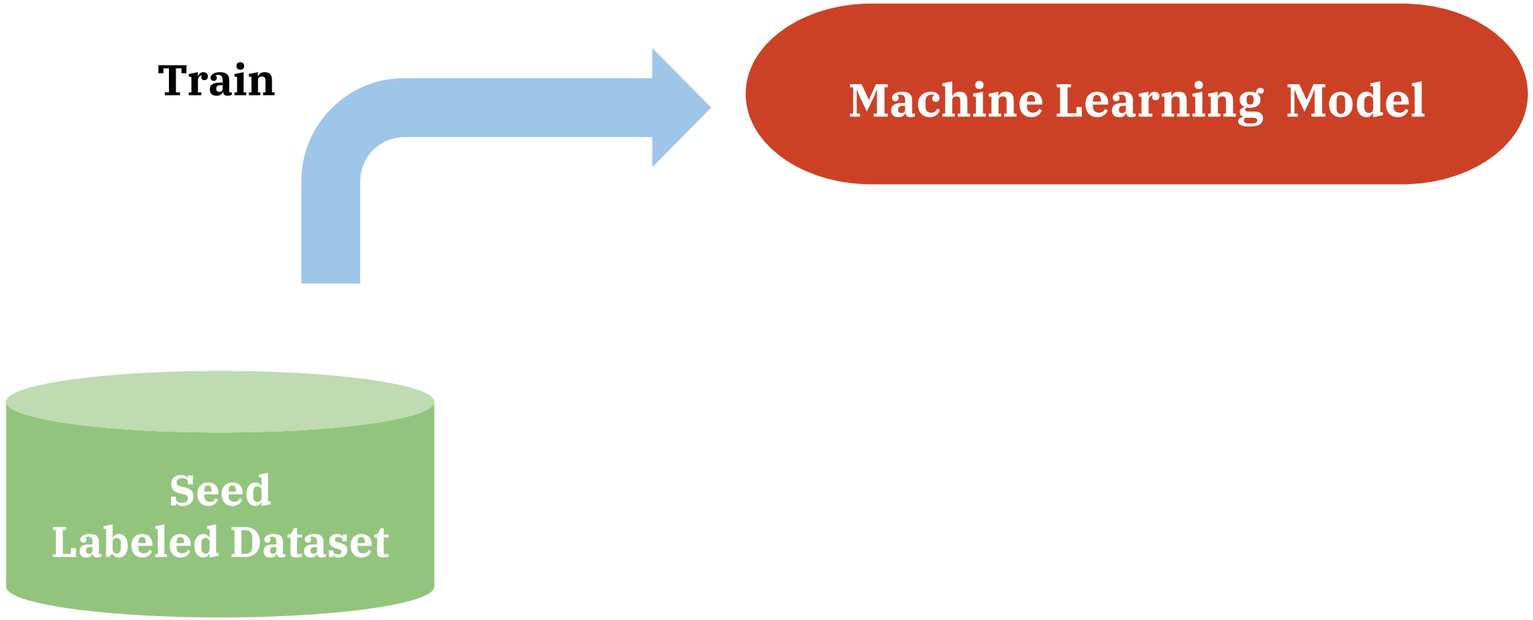
Active Learning



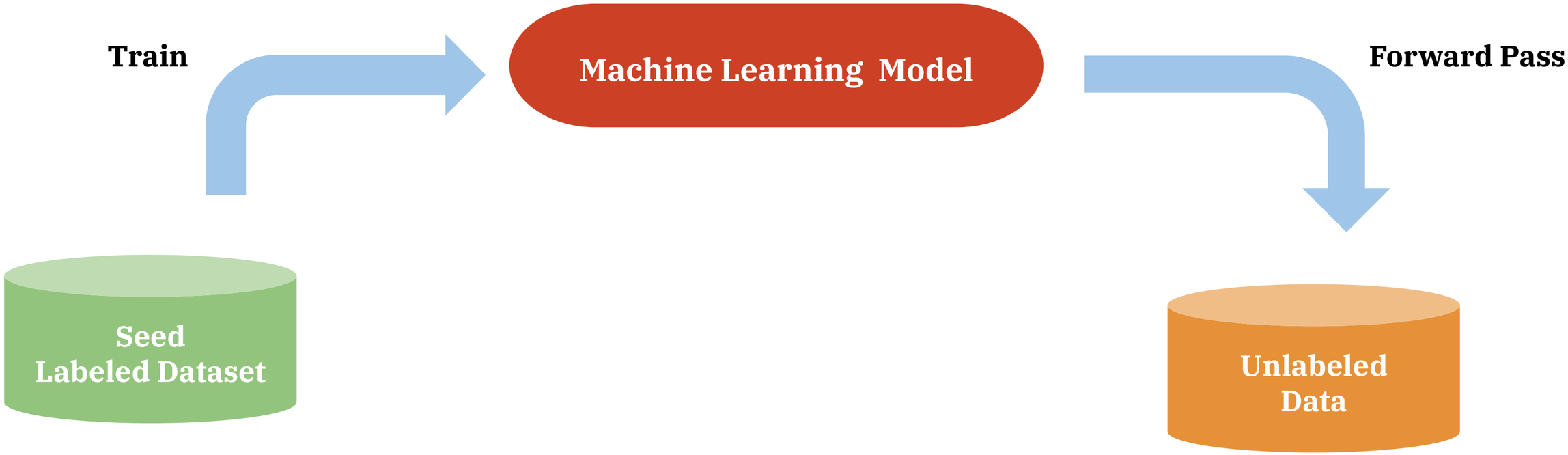
Labeled Data Needed



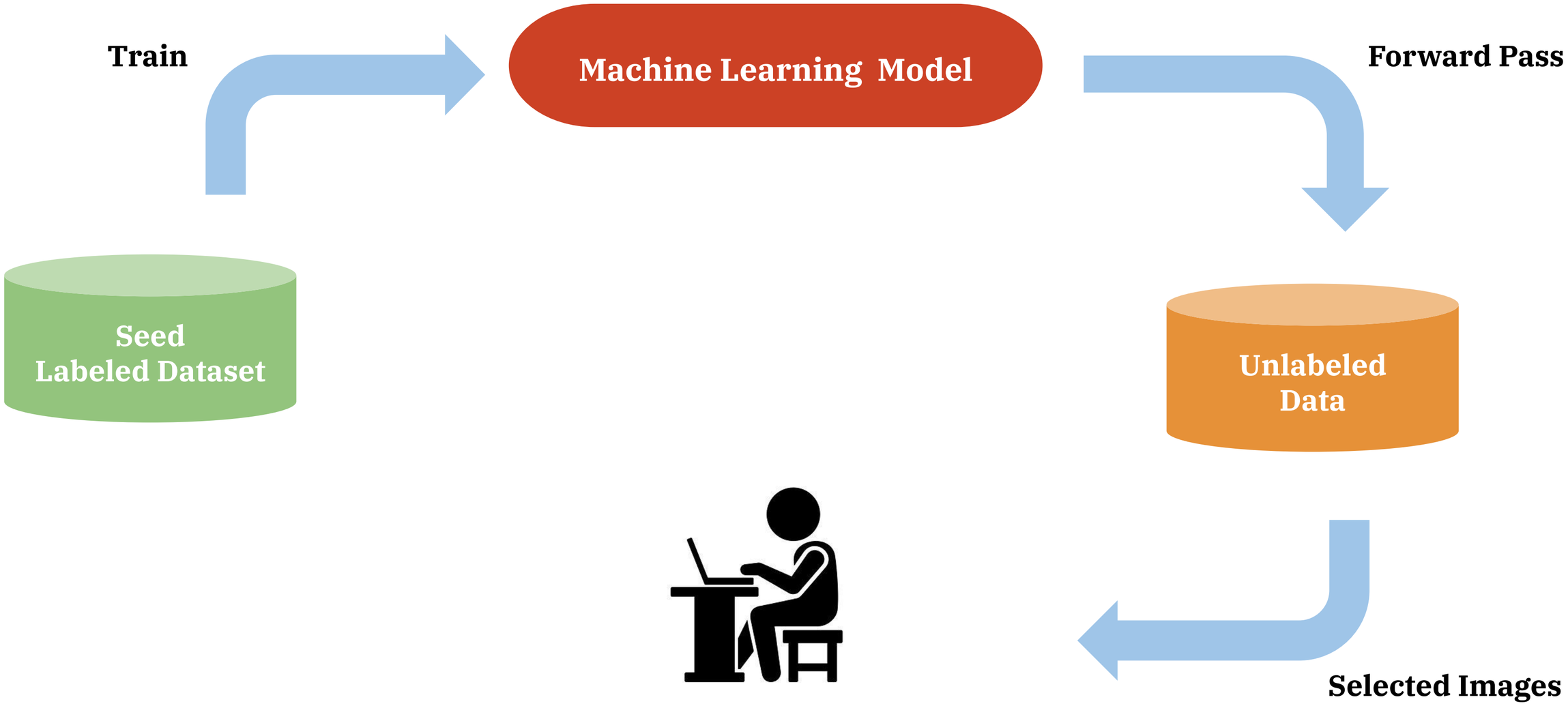
Researchers: Subhiksha Muthukrishnan, Mandeep Khokhar, Ajay Krishnan, Tarun Narayanan, Rudy Venguswamy



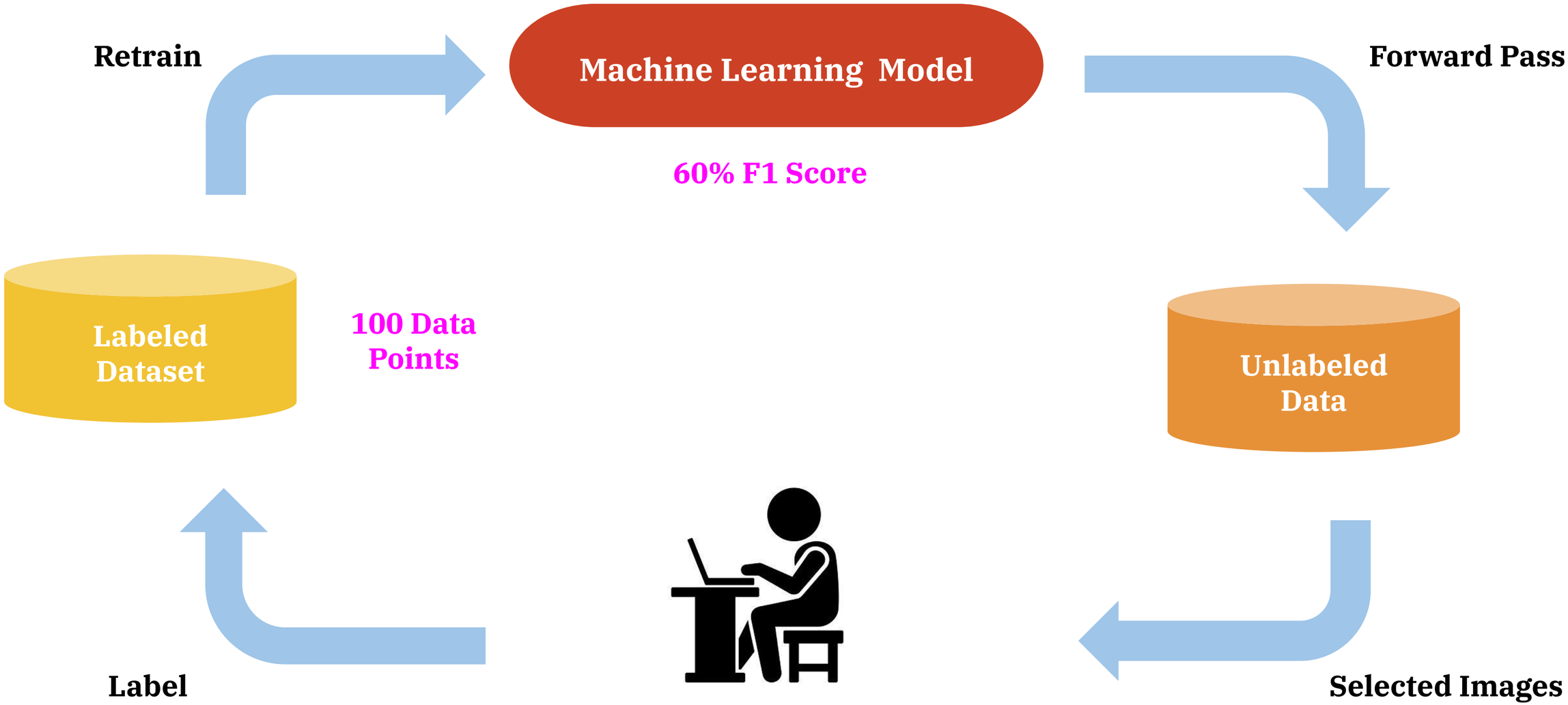
Researchers: Subhiksha Muthukrishnan, Mandeep Khokhar, Ajay Krishnan, Tarun Narayanan, Rudy Venguswamy



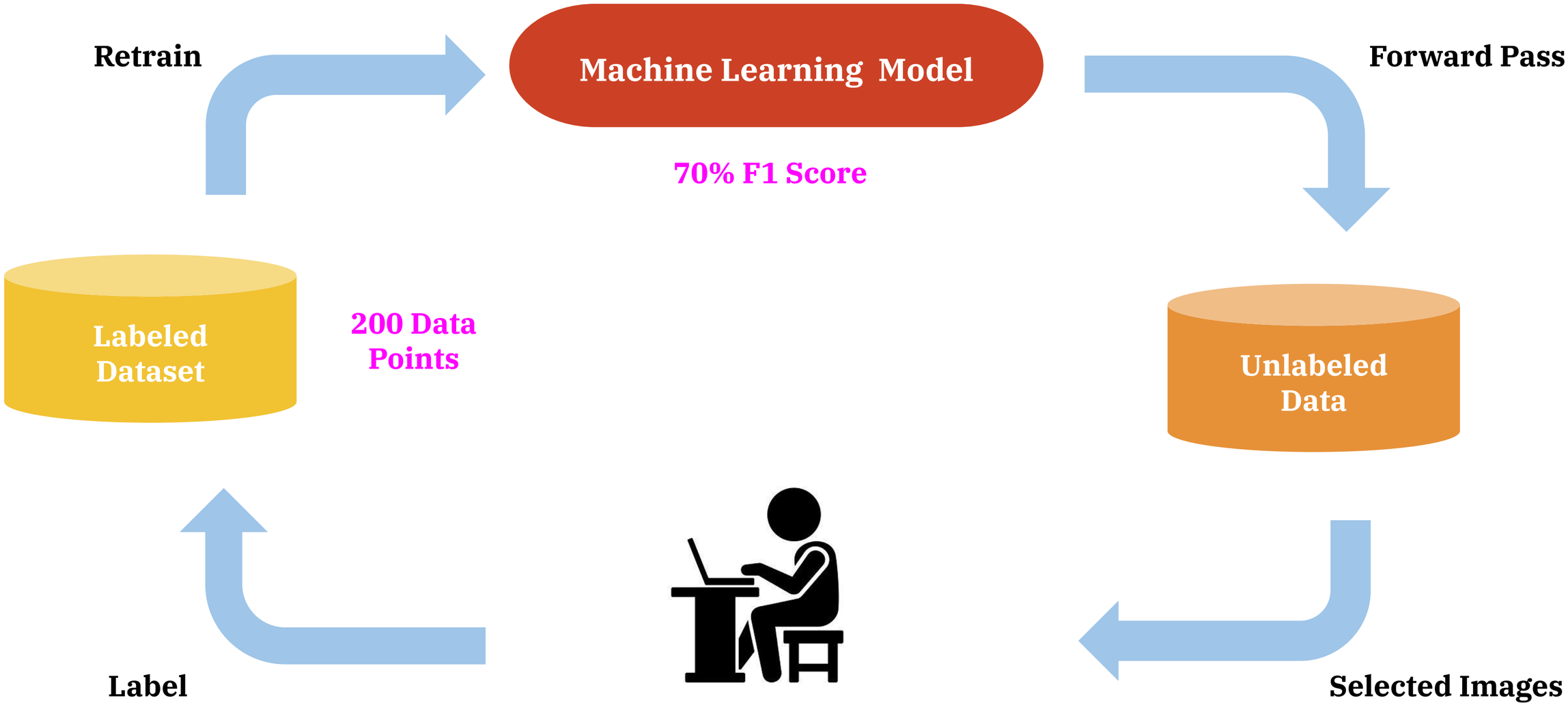
Researchers: Subhiksha Muthukrishnan, Mandeep Khokhar, Ajay Krishnan, Tarun Narayanan, Rudy Venguswamy



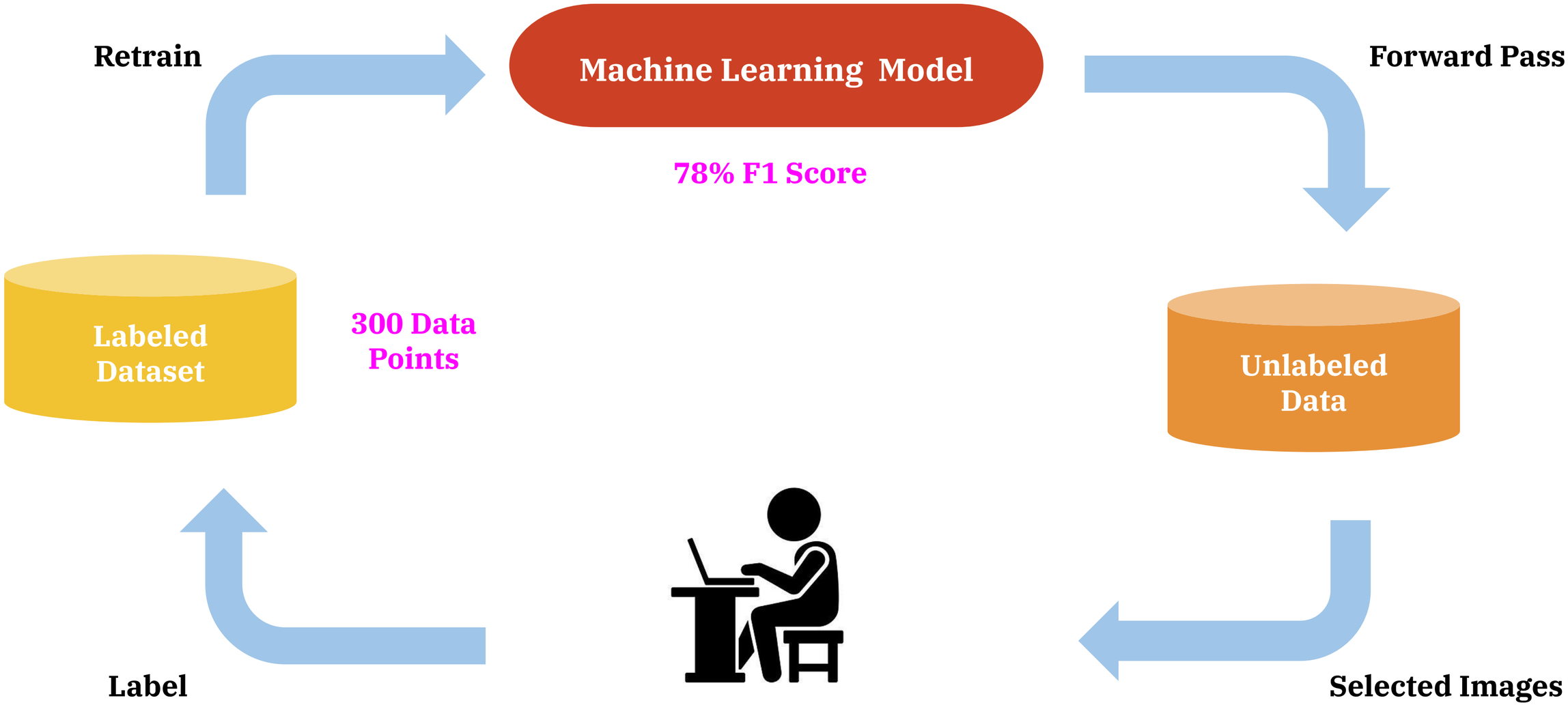
Researchers: Subhiksha Muthukrishnan, Mandeep Khokhar, Ajay Krishnan, Tarun Narayanan, Rudy Venguswamy



Researchers: Subhiksha Muthukrishnan, Mandeep Khokhar, Ajay Krishnan, Tarun Narayanan, Rudy Venguswamy



Researchers: Subhiksha Muthukrishnan, Mandeep Khokhar, Ajay Krishnan, Tarun Narayanan, Rudy Venguswamy



Researchers: Subhiksha Muthukrishnan, Mandeep Khokhar, Ajay Krishnan, Tarun Narayanan, Rudy Venguswamy

Retrain

Machine Learning Model

Forward Pass

85% F1 Score

Labeled Dataset

400 Data Points

Unlabeled Data

Label

Selected Images

Researchers: Subhiksha Muthukrishnan, Mandeep Khokhar, Ajay Krishnan, Tarun Narayanan, Rudy Venguswamy

Active Labeling - Solving for Scale

- Issue: Need to go over all images to generate probabilities
  - For 10B images, each forward pass takes $30K per forward pass and ~21K hours
  - Each experiment takes at least 10 forward passes, $300K for just forward passes
- Solution:
  - Generate a small representative sample of all embeddings, e.g. 100K
  - Forward pass on representative sample, ~13 minutes
  - For identified ‘most uncertain images’, use similarity search to access neighbourhood

Access to billions, Cost of pennies

Researchers: Subhiksha Muthukrishnan, Mandeep Khokhar, Ajay Krishnan, Tarun Narayanan, Rudy Venguswamy

Benchmark

- Dataset: RESISC45 Remote Sensing
- Aim: Given 1 image, curate images of same class

# Images Shown to Human for Labeling	% Same class curated	Time Taken
27,000	100%	38 Hours

Researchers: Subhiksha Muthukrishnan, Mandeep Khokhar, Ajay Krishnan, Tarun Narayanan, Rudy Venguswamy

Benchmark

- Dataset: RESISC45 Remote Sensing
- Aim: Given 1 image, curate images of same class

# Images Shown to Human for Labeling	% Same class curated	Time Taken
27,000	100%	38 Hours
~2100	**~82%**	**2.9 Hours**

Researchers: Subhiksha Muthukrishnan, Mandeep Khokhar, Ajay Krishnan, Tarun Narayanan, Rudy Venguswamy

Active Labeling - Solving for Scale

- Issue: Need to go over all images to generate probabilities
  - For 10B images, each forward pass takes $30K per forward pass and ~21K hours
  - Each experiment takes at least 10 forward passes, $300K for just forward passes

Researchers: Subhiksha Muthukrishnan, Mandeep Khokhar, Ajay Krishnan, Tarun Narayanan, Rudy Venguswamy

Active Labeling - Solving for Scale

- Issue: Need to go over all images to generate probabilities
  - For 10B images, each forward pass takes $30K per forward pass and ~21K hours
  - Each experiment takes at least 10 forward passes, $300K for just forward passes
- Solution:
  - Generate a small representative sample of all embeddings, e.g. 100K
  - Forward pass on representative sample, ~13 minutes
  - For identified ‘most uncertain images’, use similarity search to access neighbourhood

Access to billions, Cost of pennies

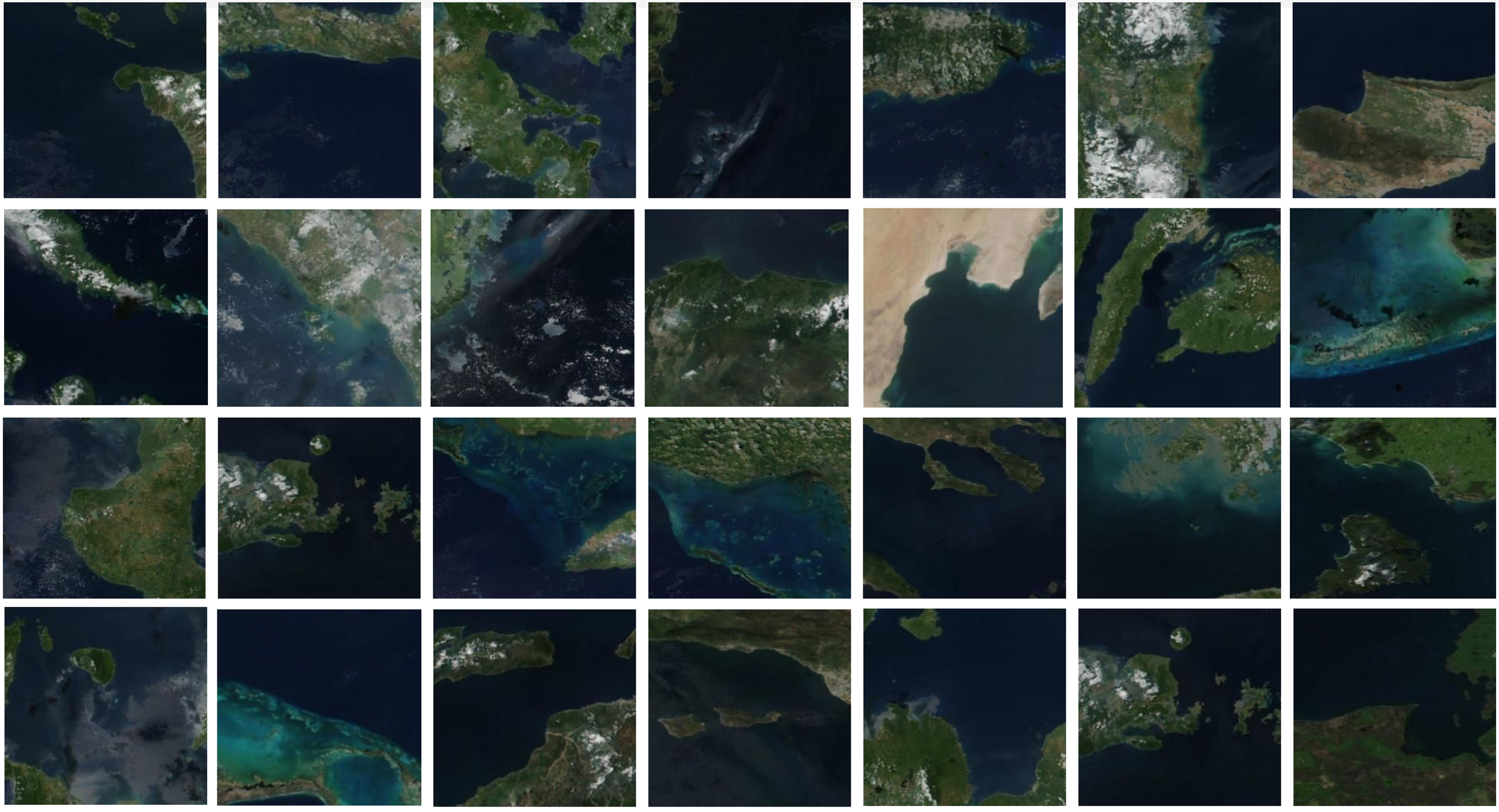
Researchers: Subhiksha Muthukrishnan, Mandeep Khokhar, Ajay Krishnan, Tarun Narayanan, Rudy Venguswamy

Using this →

out of 5M potential tiled images.

We Identified
~1000 islands out of
With probability > 0.75

In just 52 minutes

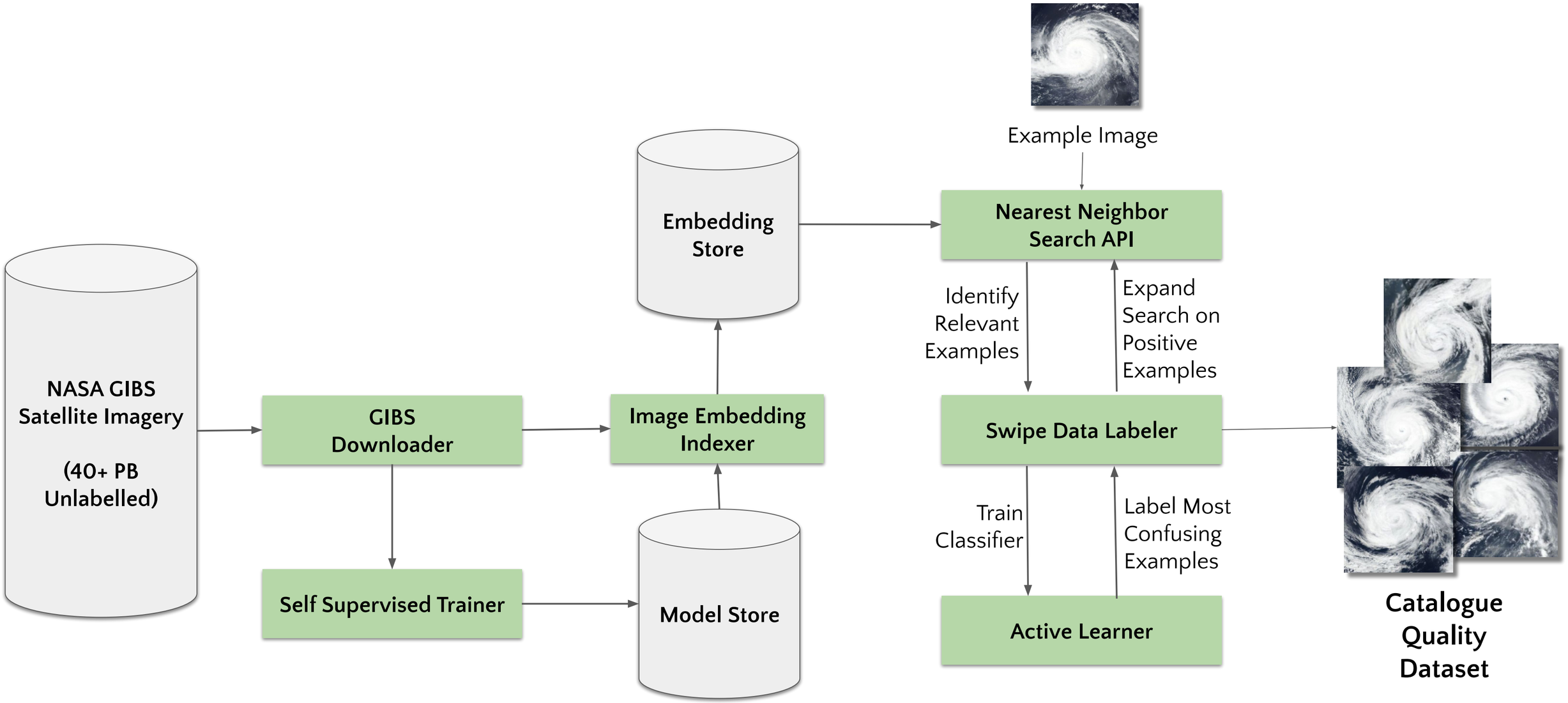


Hand Labeling = expensive

Assumptions	Details
5 Million Tiles Daily	Tiles at different resolutions
5 Seconds an Image (7K Hours)	Our Experience as Unskilled Labeler
$15/Hr to Hire Annotator	$105,000 Annotation Cost

~~Hand Labeling = expensive~~

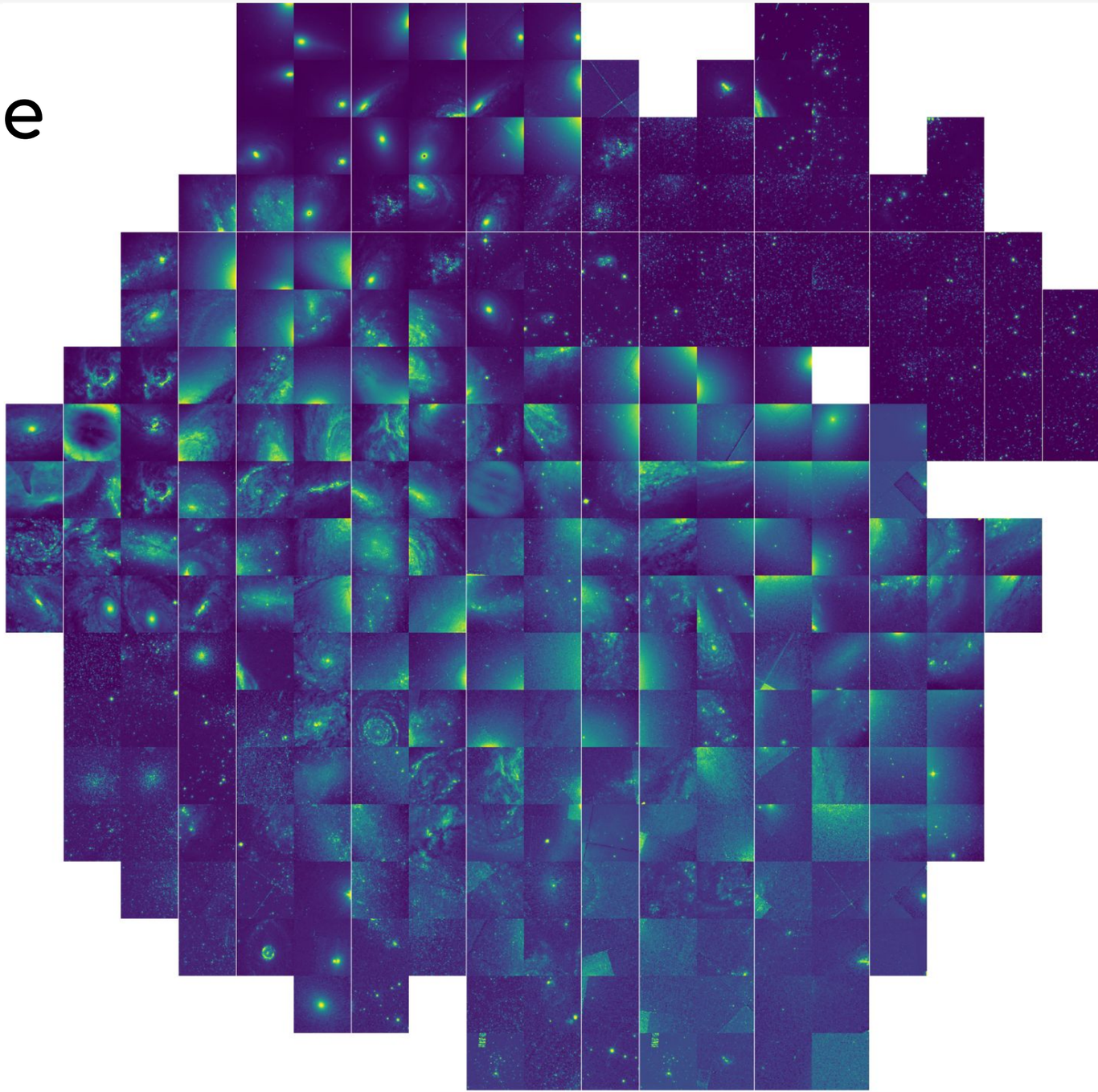
Assumptions	Details
5 Million Tiles Daily	Tiles at different resolutions
5 Seconds an Image (7K Hours)	Our Experience as Unskilled Labeler
~~$15/Hr to Hire Annotator~~ Citizen Science!	~~$105,000 Annotation Cost~~ Priceless


NASA GIBS
Satellite Imagery
(40+ PB
Unlabelled)
GIBS
Downloader
Self Supervised Trainer
Image Embedding
Indexer
Embedding
Store
Model Store
Example Image
Nearest Neighbor
Search API
Identify
Relevant
Examples
Expand
Search on
Positive
Examples
Swipe Data Labeler
Train
Classifier
Label Most
Confusing
Examples
Active Learner
Catalogue
Quality
Dataset

Cross Domain Applications

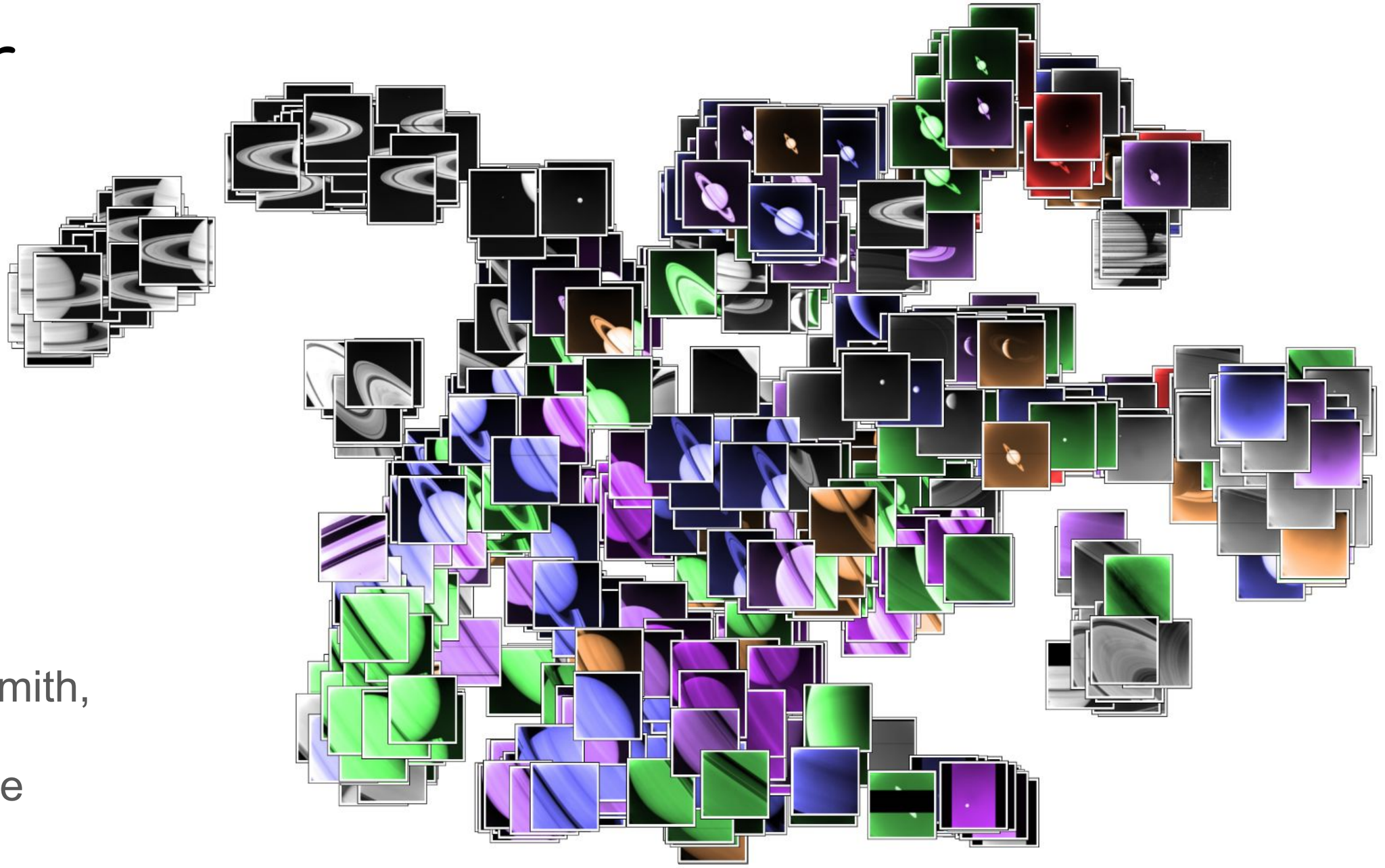
Hubble Space Telescope

- In collaboration with Rick White and Joshua Peek from Space Telescope Science Institute
- Dataset: Hubble Image Similarity Project



Researchers: Walker Steven

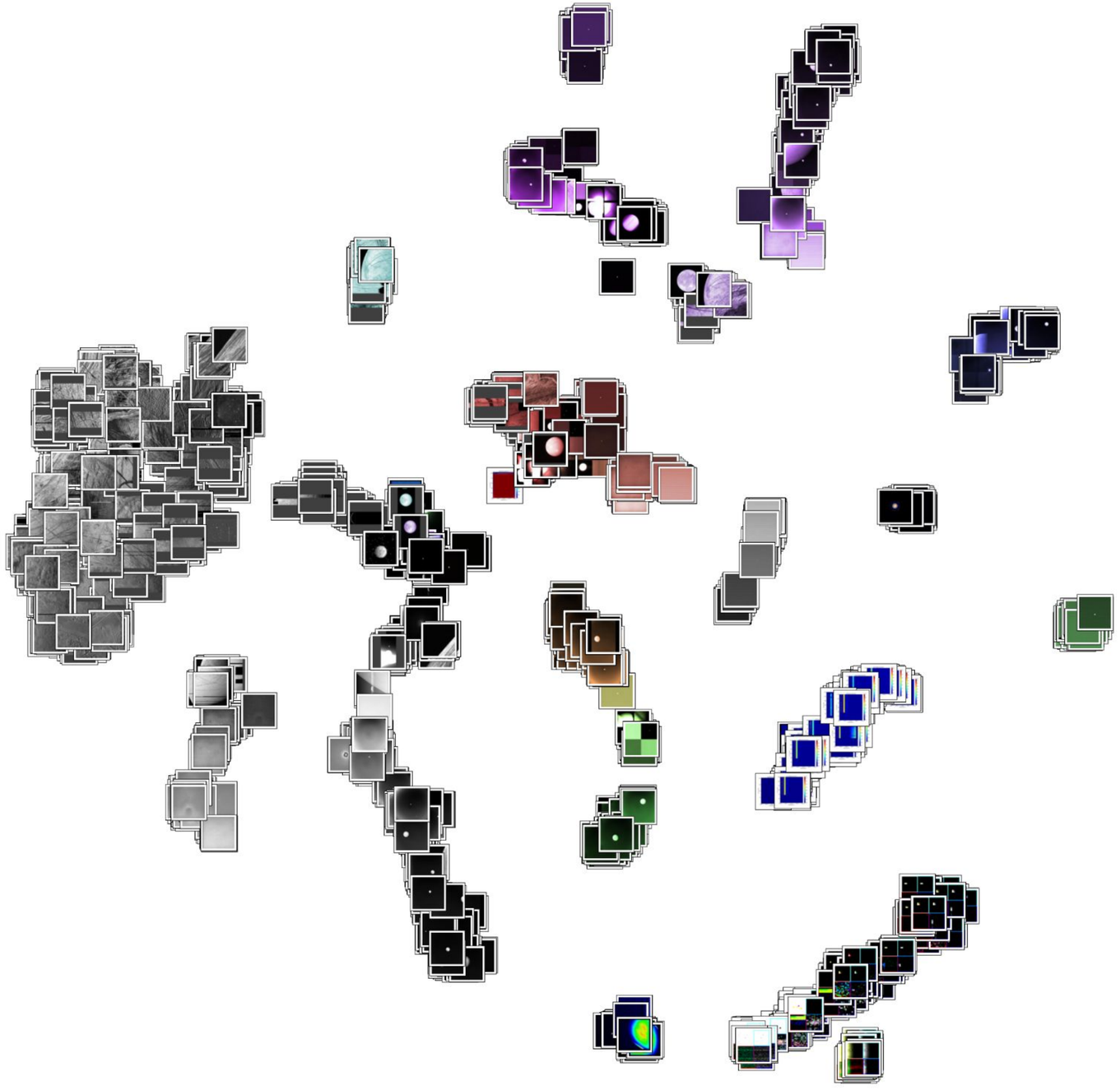
Voyager



- By Jeffrey Smith,

  SETI Institute

Europa



- By Jeffrey Smith,

  SETI Institute

Other Wins

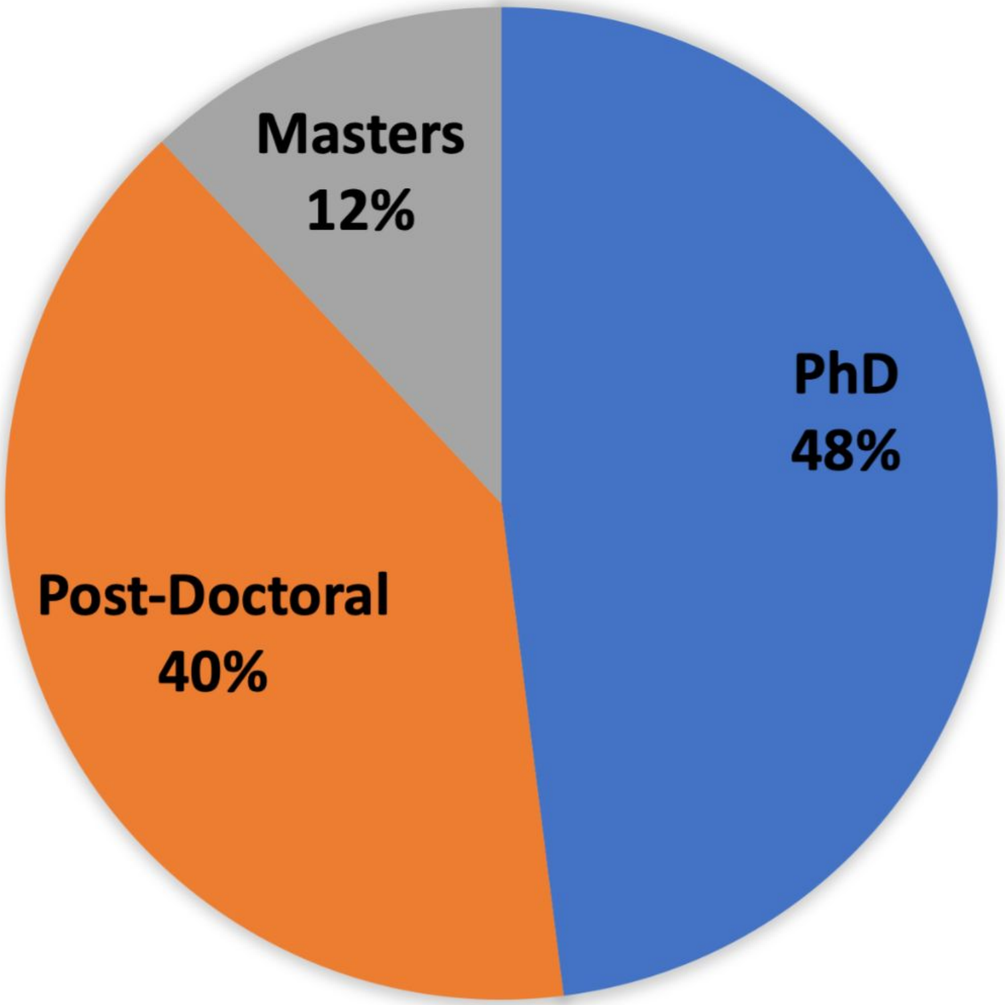
- NASA/MSFC IMPACT Team - Integration in Phenomena portal
- NASA/MSFC IMPACT Team - Automate creation of new labels
- NASA/GSFC ESDIS (Ryan Boller) - Investigating infusion into WorldView

SPACE ML

Meet our Scientists

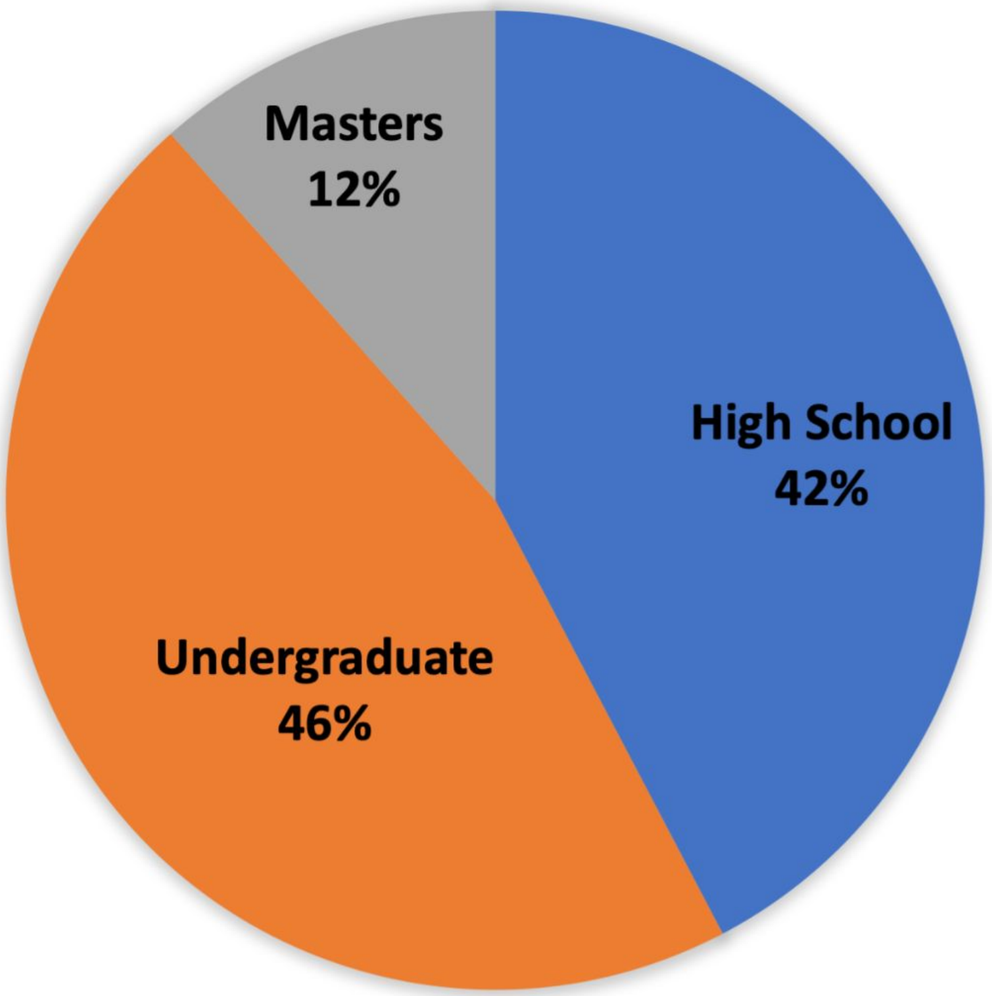
FDL PARTICIPANT BACKGROUND

FDL FRONTIER DEVELOPMENT LAB



SPACEML PARTICIPANT BACKGROUND

SPACEML



SPACEML PARTICIPANT BACKGROUND

Mexico 🇲🇽

South Korea

India 🇮🇳

USA 🇺🇸

Singapore 🇸🇬

Germany

Thank You

www.spaceML.org

Available now

FDL FRONTIER DEVELOPMENT LAB

PARTNER

Frontier Development Lab (FDL) is a NASA co-operative agreement with the SETI Institute.

ARTIFICIAL INTELLIGENCE RESEARCH FOR SPACE SCIENCE, EXPLORATION & ALL HUMANKIND

SPACEML