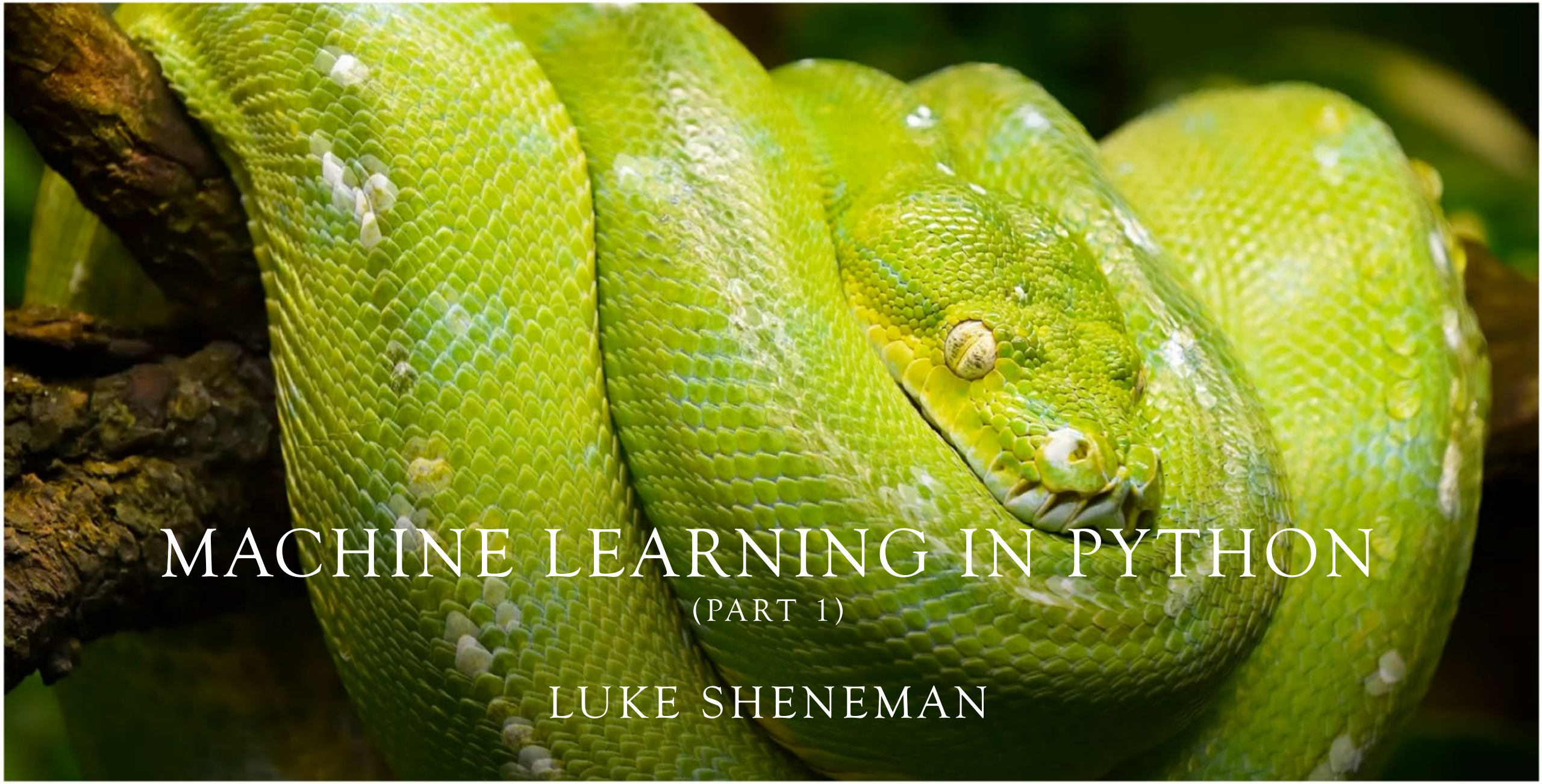


MACHINE LEARNING IN PYTHON

(PART 1)

LUKE SHENEMAN

OVERVIEW

- Background
  - Why Machine Learning?
  - Why Python?
  - My Current Machine Learning Projects

- **Part 1**: Introduction to Machine Learning in Python: ***scikit-learn***
- **Part 2**: Introduction to Deep Learning in Python: ***TensorFlow & Keras***
- **Part 3**: Introduction to Deep Learning in Python: ***PyTorch***

WHY MACHINE LEARNING?

- Easily build exceptionally flexible and powerful models to:
  - Predict values
  - Optimize systems
  - Classify, categorize, and cluster data
  - Parse and generate natural language
  - Generate or denoise sounds, images, video, etc.

- Use *data* rather than explicit programming to create models
  - ML can perform tasks not explicitly programmed
  - Generalize / robust

- Applications
  - Science of all kinds
  - Agriculture
  - Medicine
  - Robotics / Manufacturing
  - Banking / Business / Economics
  - Entertainment

AI/ML is in its infancy and will come to define an era

WHY PYTHON?

- Easy to learn
- Consistent, concise, readable code
- Collaborative
- Extensive libraries
- Platform independent
- High performance (cython)
- Massive head start (TensorFlow, PyTorch)

Sheneman L, Vasdekis AE, et al.

Deep learning classification of lipid droplets in quantitative phase images.

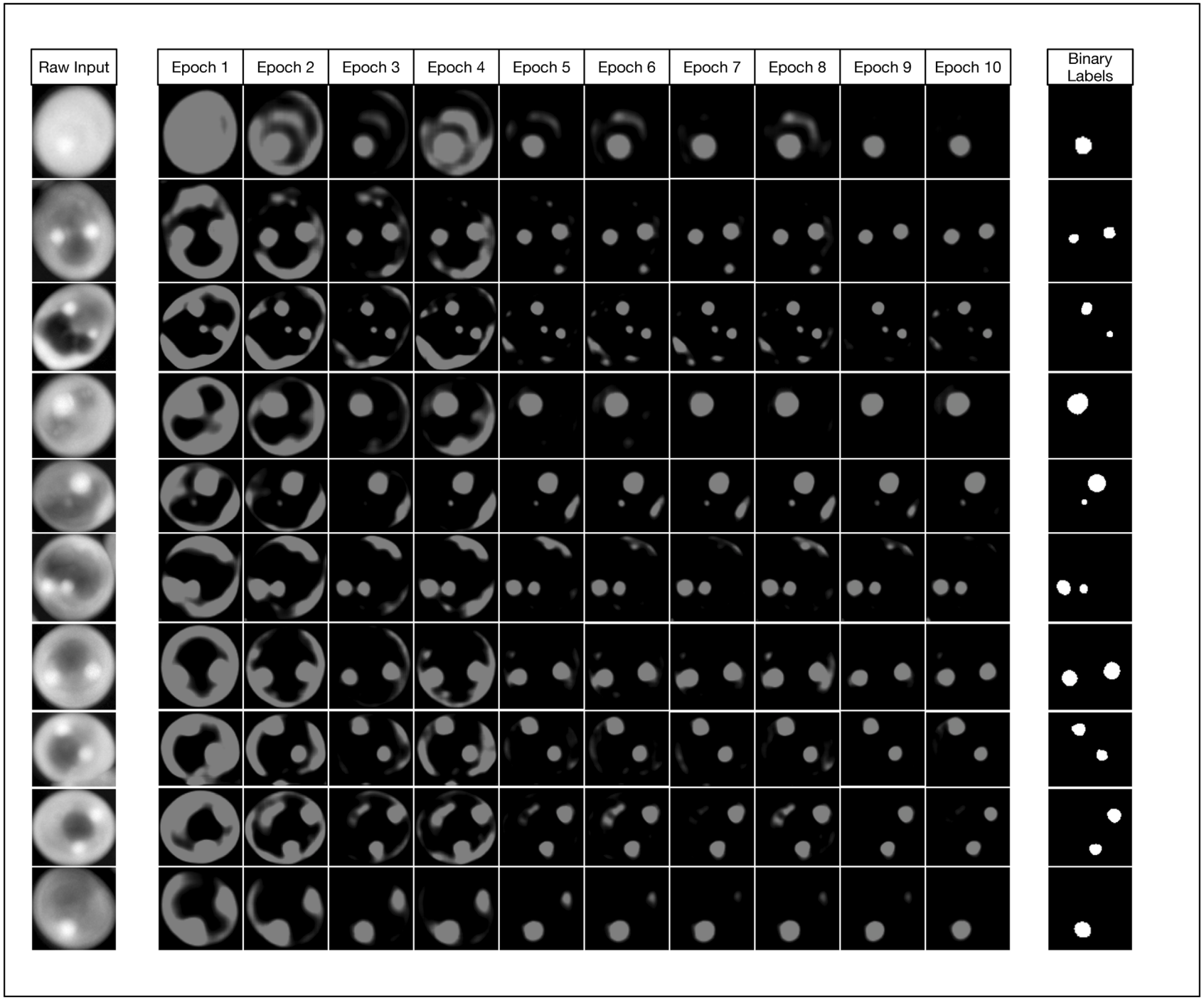
PLoS ONE 2021 16(4)

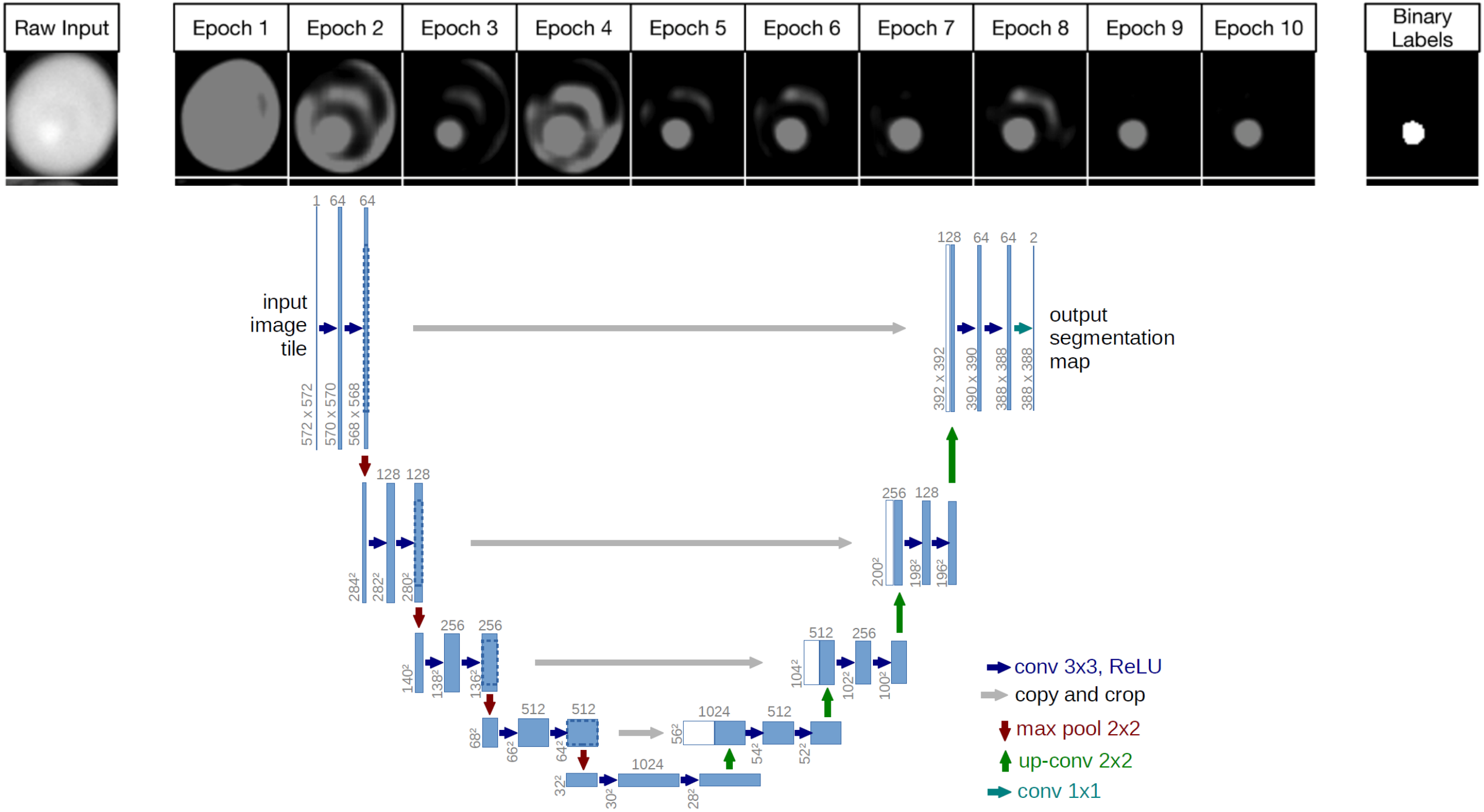
github.com/sheneman/deep_lipid

Tristin Sanchez, Luke Sheneman, Andreas E. Vasdekis et al.

Photon-Sparse, Poisson Light-Sheet Microscopy

ACS Photonics 2021 8 (10)




Raw Input
Epoch 1
Epoch 2
Epoch 3
Epoch 4
Epoch 5
Epoch 6
Epoch 7
Epoch 8
Epoch 9
Epoch 10
Binary Labels
input image tile
output segmentation map
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1

NanoDetector Workflow

On timer

MegaDetector AI Model

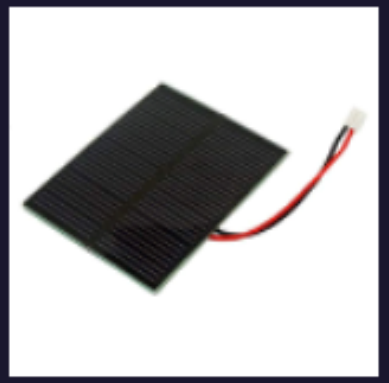
NanoDetector AI Model

encoded inferences

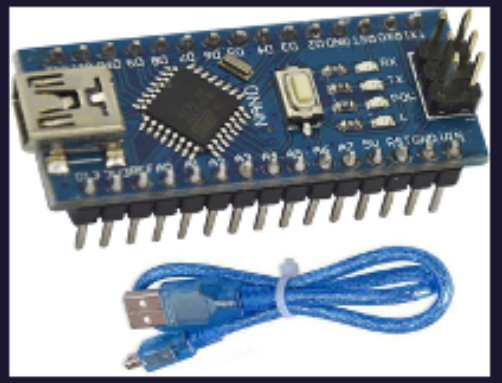
University of Idaho

Developing the NanoDetector Edge AI Device

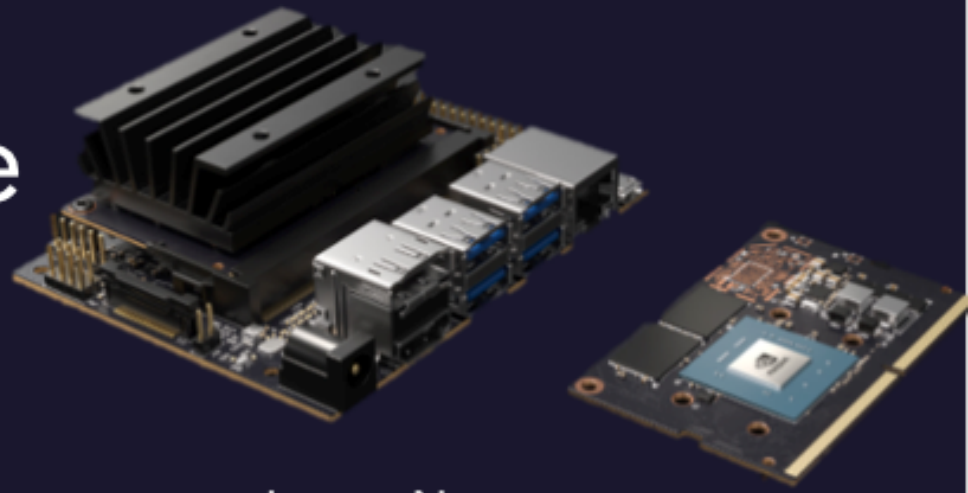
Moultrie Camera



9W Solar Panel



Arduino Microcontroller



Jetson Nano

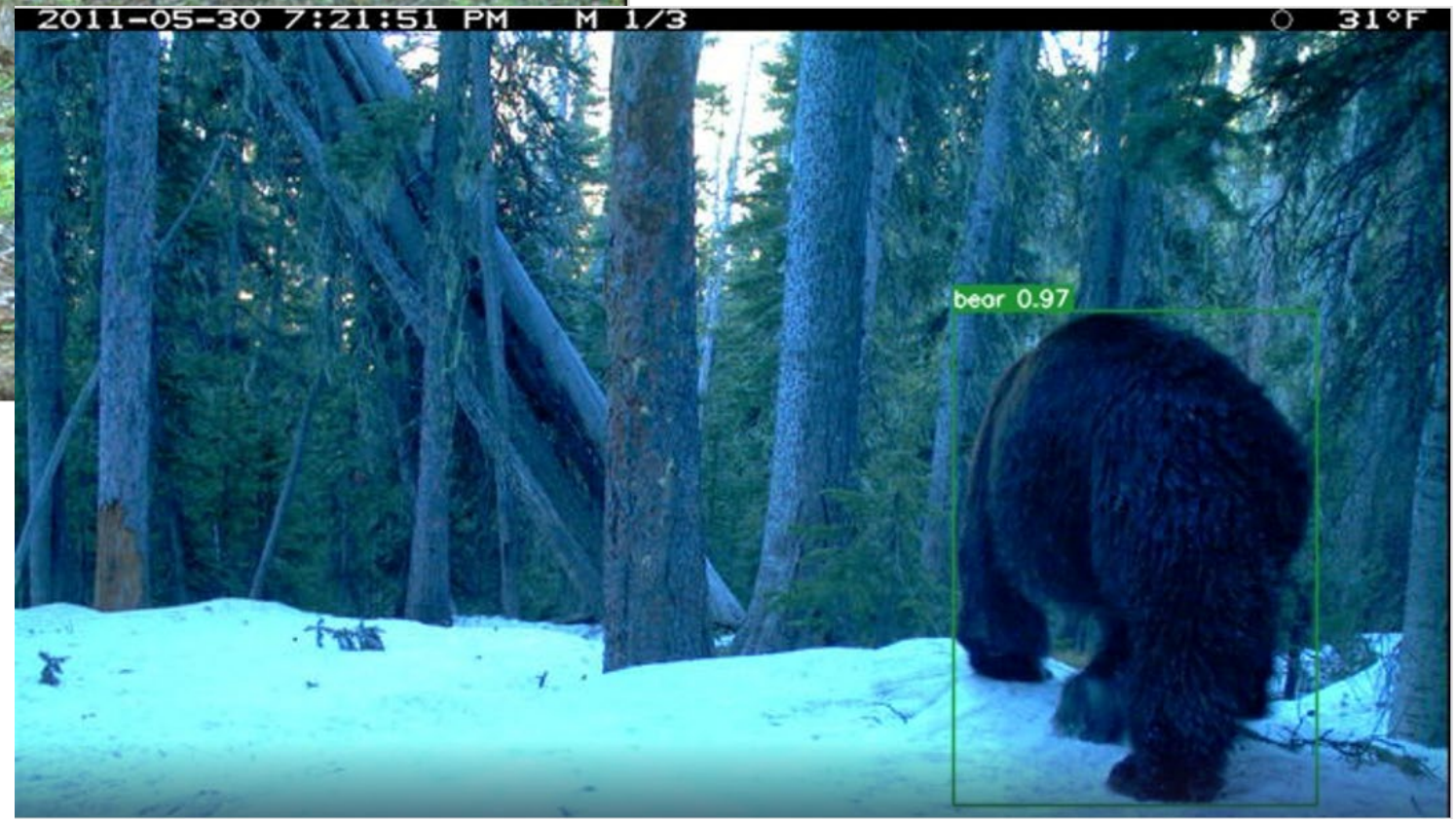
RockBlock Satellite Modem

19,200mAh, 75 Watt Hour
Always-On Li-IonBattery

External Antenna

Clock and
Temperature


moose 0.74
moose 0.96

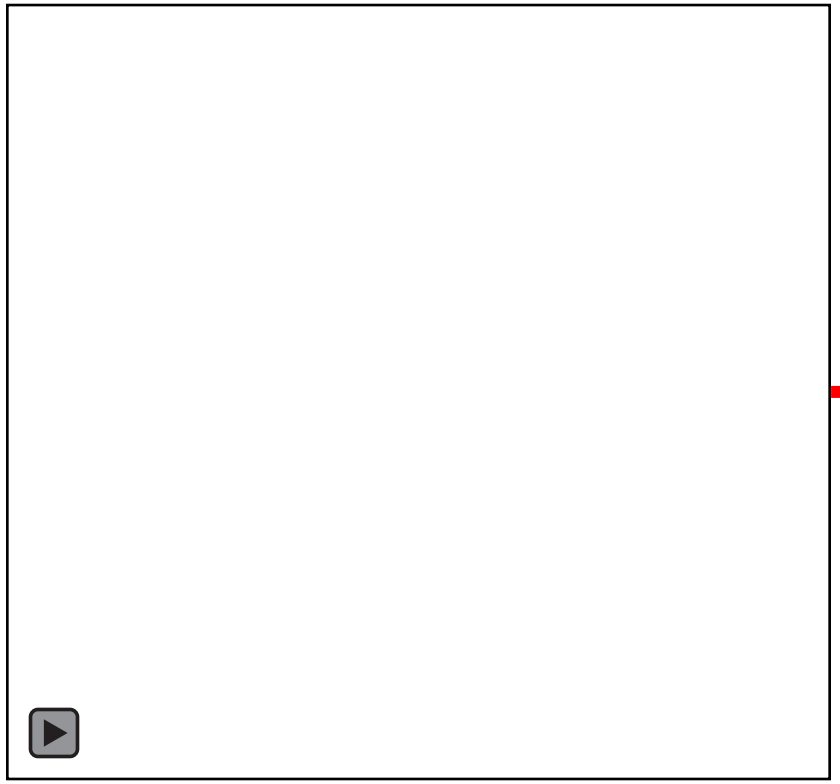

2011-05-30 7:21:51 PM M 1/3
31°F
bear 0.97

Integrative imaging of Plant Roots during symbiosis with Mycorrohizal Fungi

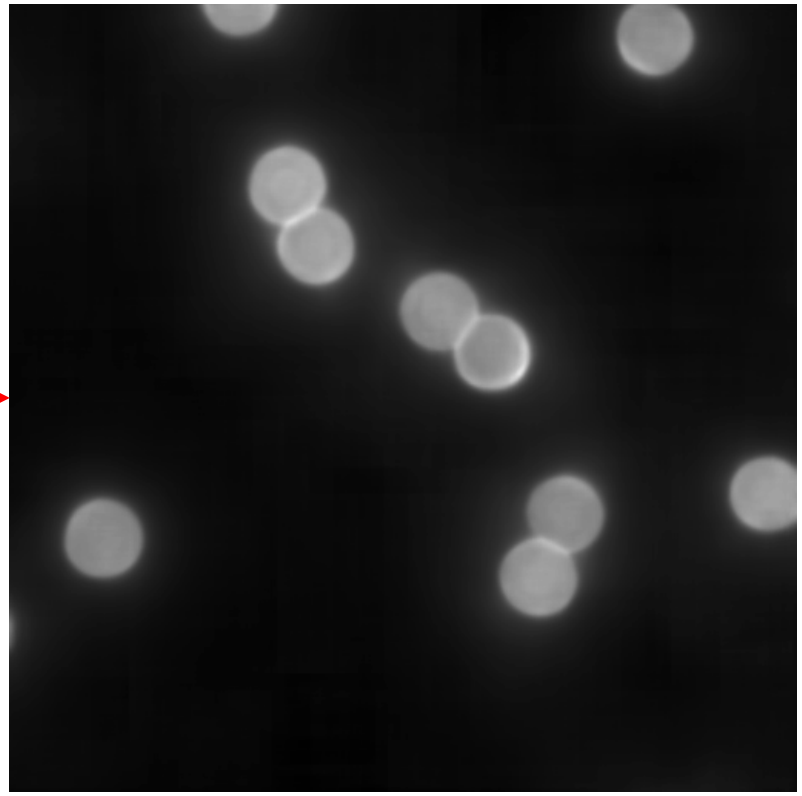
DOE Award #: DE-SC0022282

Sparse Photon Microscopy

"How Low Can You Go?"



Photon Clouds

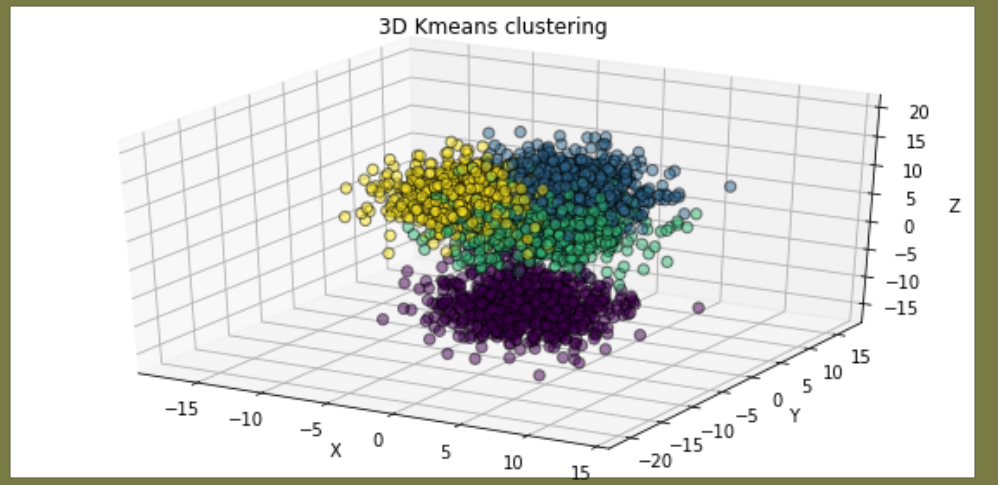


Generative AI

NEPALESE TIGER

ACCELEROMETER DATA

Tri-Axial Accelerometer
4 months of 16 Hertz Data

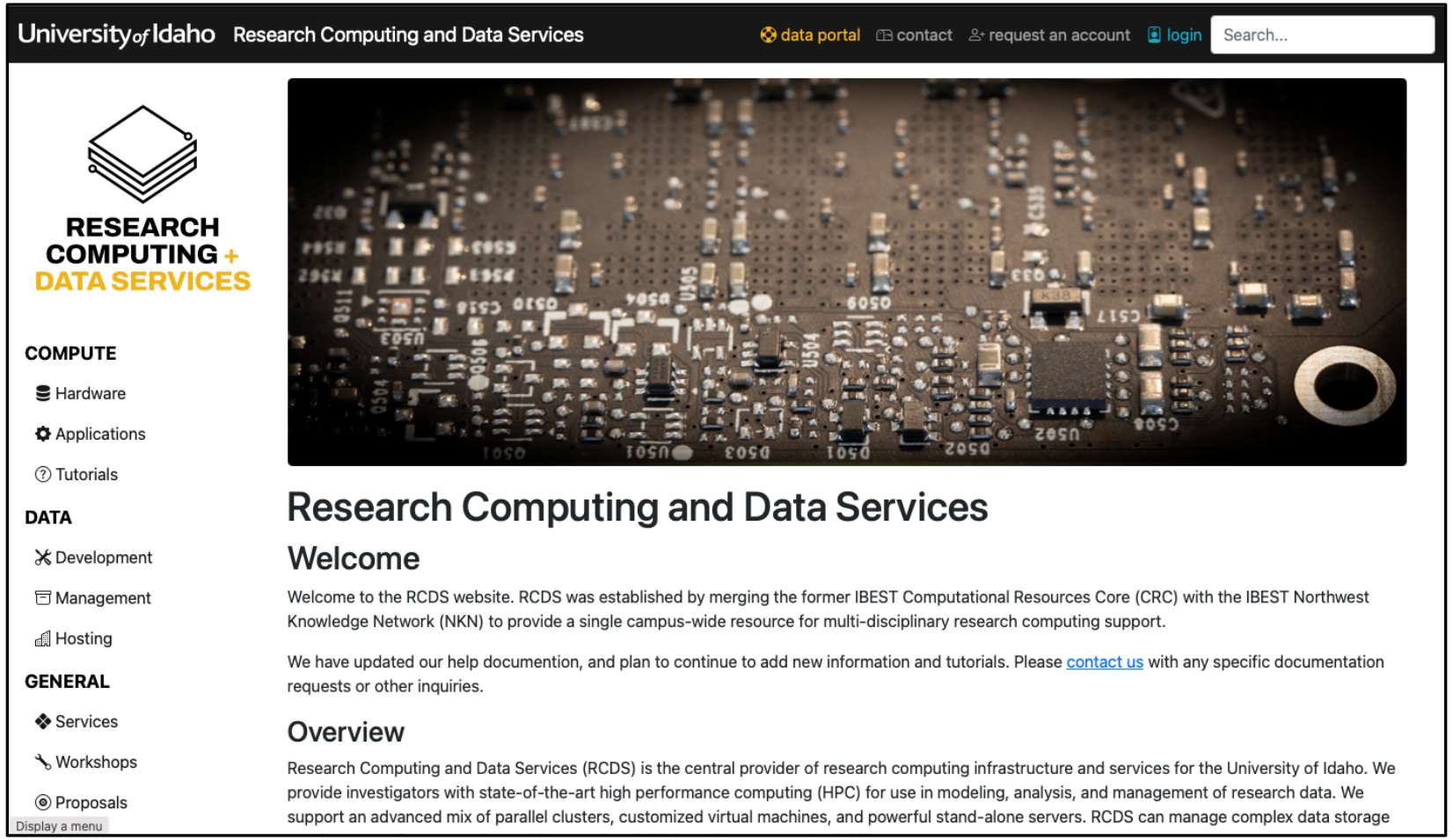


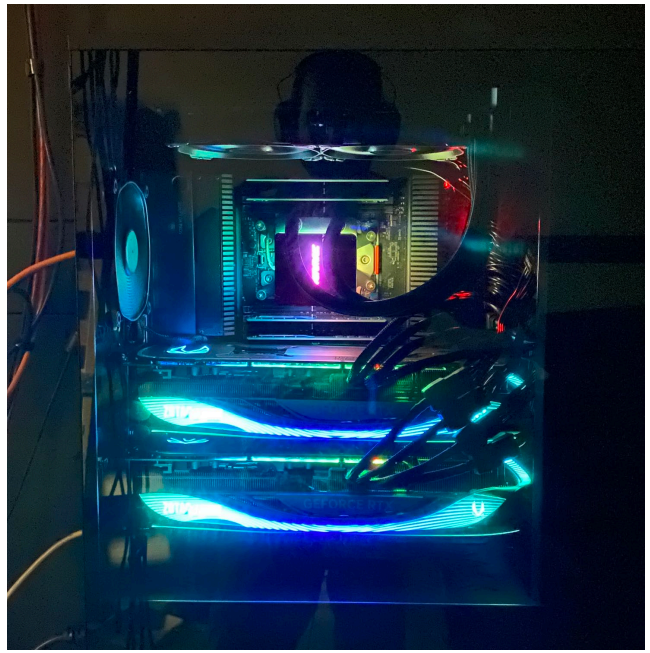
https://theconversation.com/gps-tracking-could-help-tigers-and-traffic-coexist-in-asia-158751

Associate Professor
Ecosystem Science and Management
Geospatial Data Sciences
Conservation + Restoration

University of Michigan

MACHINE LEARNING INFRASTRUCTURE





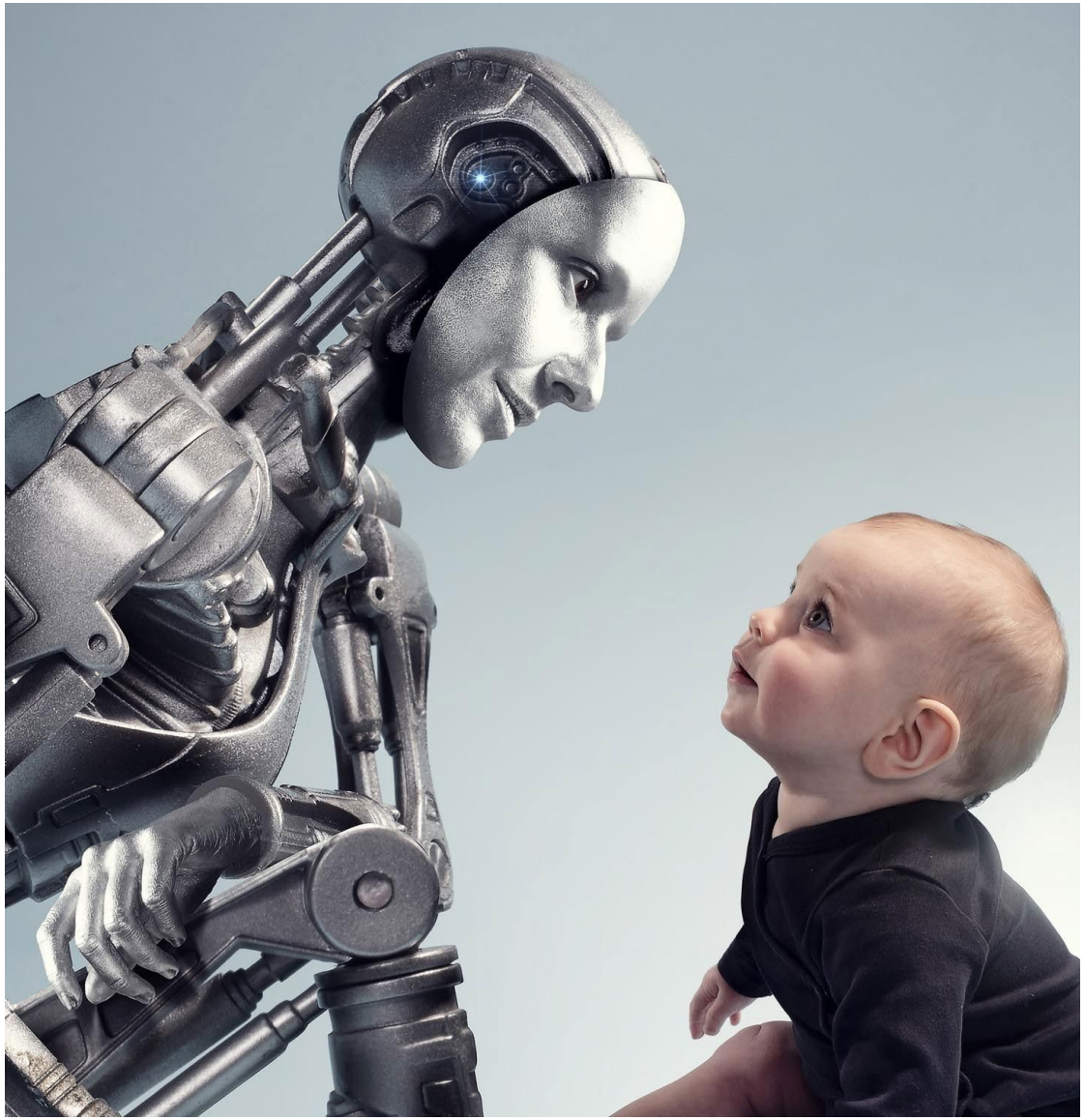
RCDS GPUs

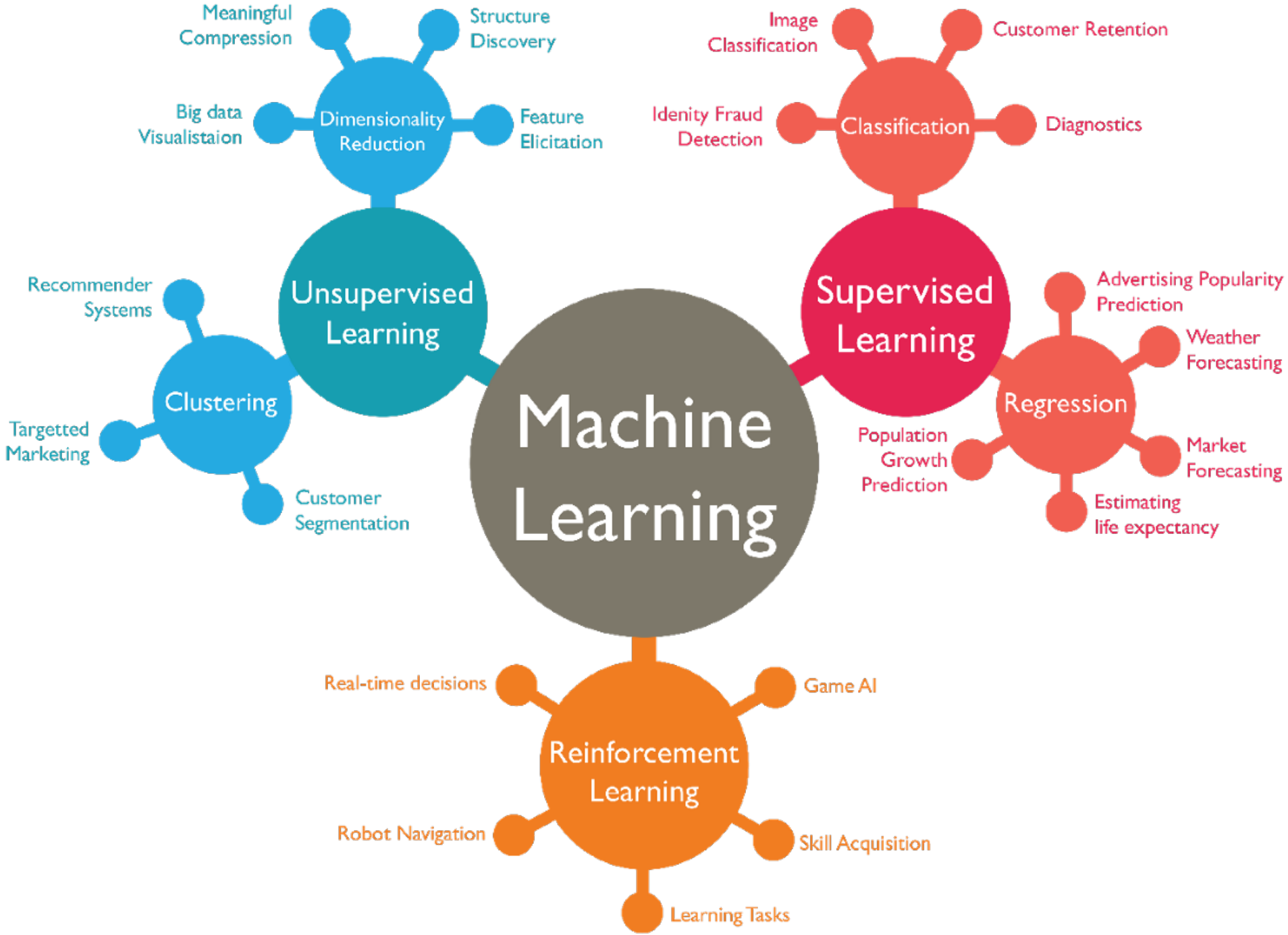
Google Colaboratory

hpc.uidaho.edu

ACCESS | Advancing Innovation

MACHINE LEARNING BASICS




Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

SCIKIT-LEARN

https://scikit-learn.org/

- Datasets
- Data Splitting
- Linear/Logistical Regression
- Decision Trees
- Bagging / Boosting
- Random Forest
- Gradient Boost
- Confusion Matrices / Scoring
- K-Means Clustering
- K-Fold Cross Validation
- Principal Component Analysis
- Feature Extraction
- Neural Networks

TRAIN A HANDWRITING RECOGNITION MODEL

Supervised Machine Learning

Use scikit-learn *Datasets*

Perform *Data Splitting*

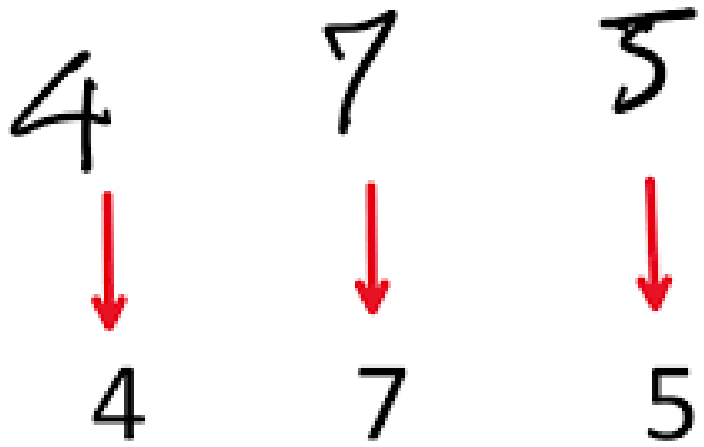
Perform Feature Extraction

Use Multiple Machine Learning Methods

1. *Random Forest*
2. *Gradient Boost*
3. *Neural Network (MLP)*

Confusion Matrices and Scoring

Compare Methods



MNIST

THE "HELLO WORLD" OF IMAGE CLASSIFICATION DATASETS

70,000 28x28 grayscale images
(with labels)