Introduction to Machine Learning

M. Kundegorski

3rd March 2020

Centre for Ecological Sciences
Indian Institute of Science

Bengaluru, India

Is it just statistics?

- different aim: explanation vs prediction
- effective engineering to allow for a mathematical description of environment.
- If it works – <u>it works</u>.

What is machine “learning”?

“A computer program is said to ***learn*** *from* ***experience E*** *with respect to some* ***task T*** *and some* ***performance measure P****, if its performance on* ***T****, as measured by* ***P****, improves with experience* ***E****.”*

Tom Mitchel, 1998

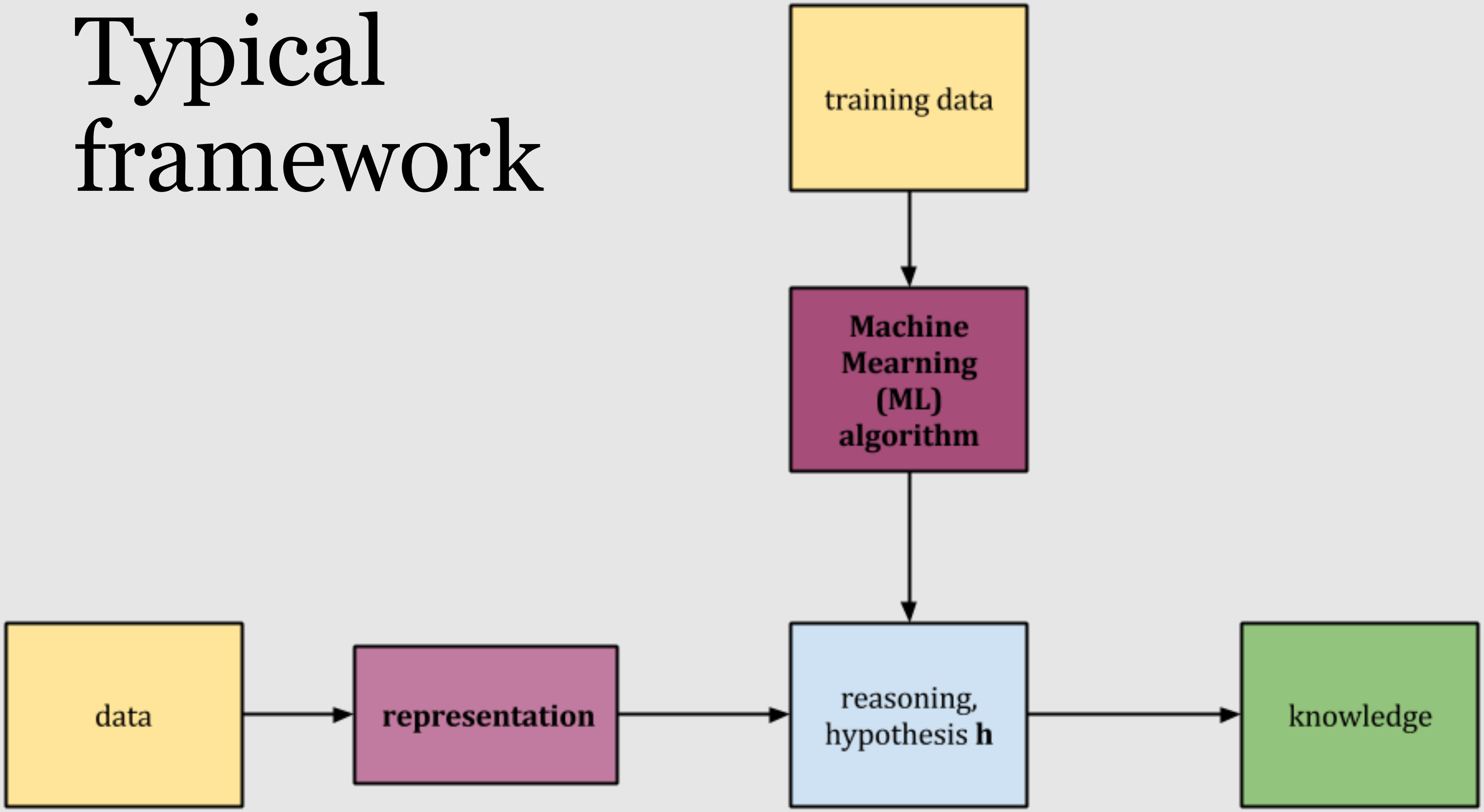
Types of ML

- Supervised – **task T** with ground-truth labels
  - goal: predict label from data

- Unsupervised – **task T** without labels
  - goal: find groupings and patterns in data

- Reinforcement Learning – **task T** is exploratory
  - goal: optimise processes
  - often RL elements are parts of Supervised and Unsupervised algorithms
  - often agent based (e.g. “game of life”)

Typical framework



Machine Learning areas

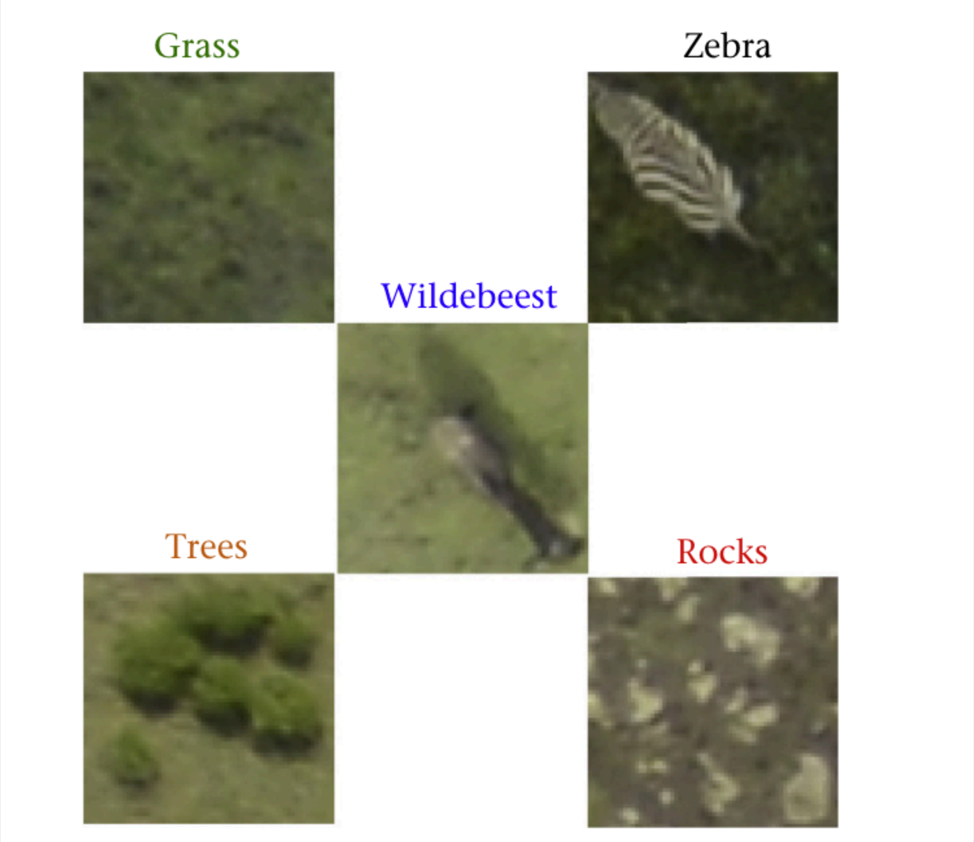
- **Big Data** – allowing us, humans to learn from data otherwise too sparse or huge to analyse with ordinary statistics .
- **Patter Recognition**: Exploring possible correlations in data which are elusive to ordinary analysis.
- **Computer Vision:** Understanding of visual data (video understanding, object detection, tracking, etc.)
- **Robotics:** physical presence of computers in the real world

Types of Computer Vision problems:

1. classification (binary and non-binary):
    - image recognition
2. regression
    - object localisation
    - object counting
3. association / clustering
    - texture clustering (object segmentation)
    - data mining
4. combination:
    - object detection (localisation + recognition)
    - object tracking (localisation + recognition + association)

Image recognition

We will focus on image recognition task – the most canonical.



Valletta 2017: Applications of machine learning in animal behaviour studies

Examples

Contents lists available at ScienceDirect

Animal Behaviour

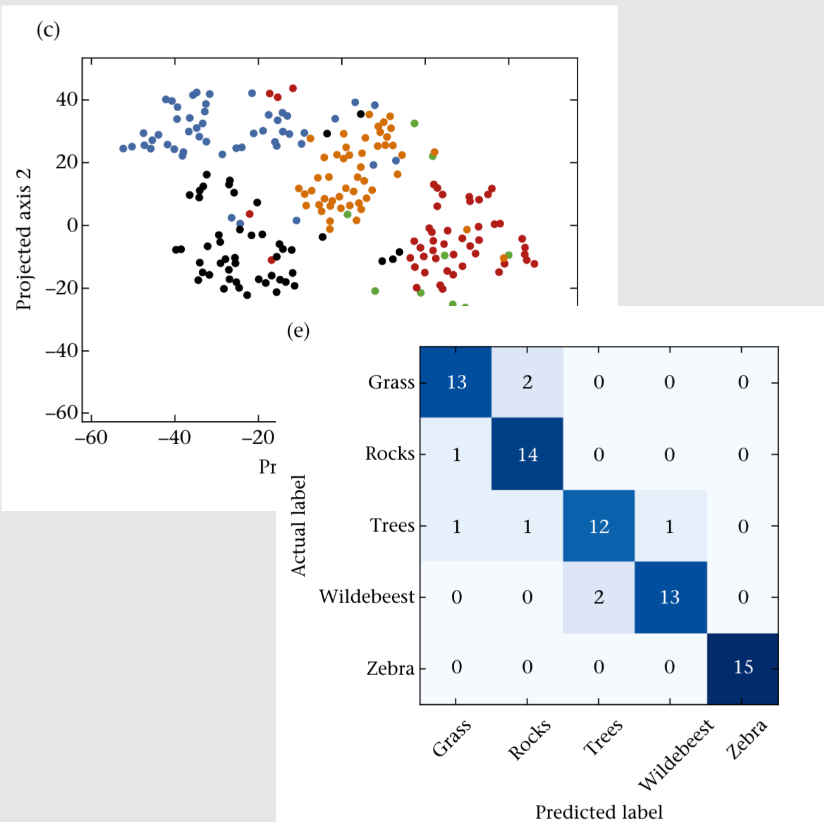
journal homepage: www.elsevier.com/locate/anbehav

Review

Applications of machine learning in animal behaviour studies

John Joseph Valletta [a, *], Colin Torney [a], Michael Kings [b], Alex Thornton [b], Joah Madden [c]

[a] Centre for Mathematics and the Environment, University of Exeter, Penryn Campus, Penryn, U.K.
[b] Centre for Ecology and Conservation, University of Exeter, Penryn Campus, Penryn, U.K.
[c] Centre for Research in Animal Behaviour, University of Exeter, Exeter, U.K.



github.com/mixmixmix/learnmachine

Counting via detection

The application of support vector machine classification to detect cell nuclei for automated microscopy

Ji Wan Han · Toby P. Breckon · David A. Randell · Gabriel Landini

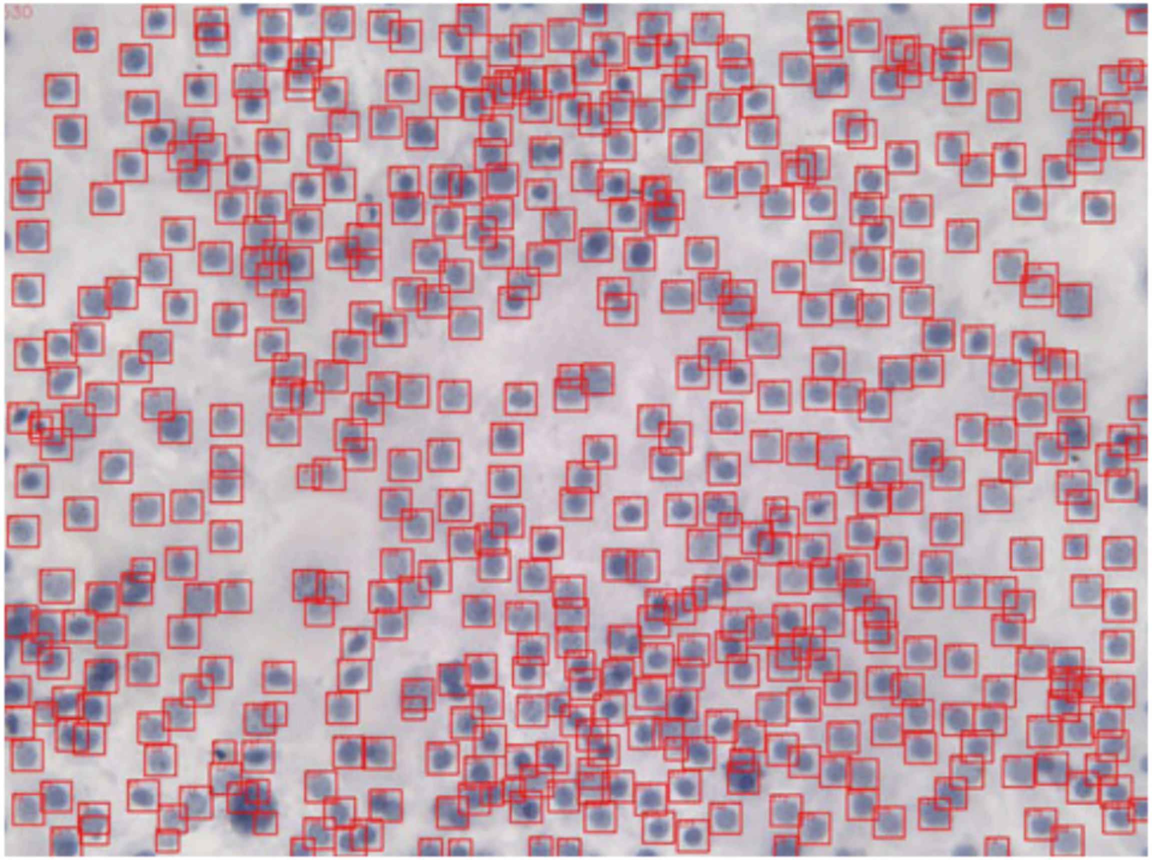
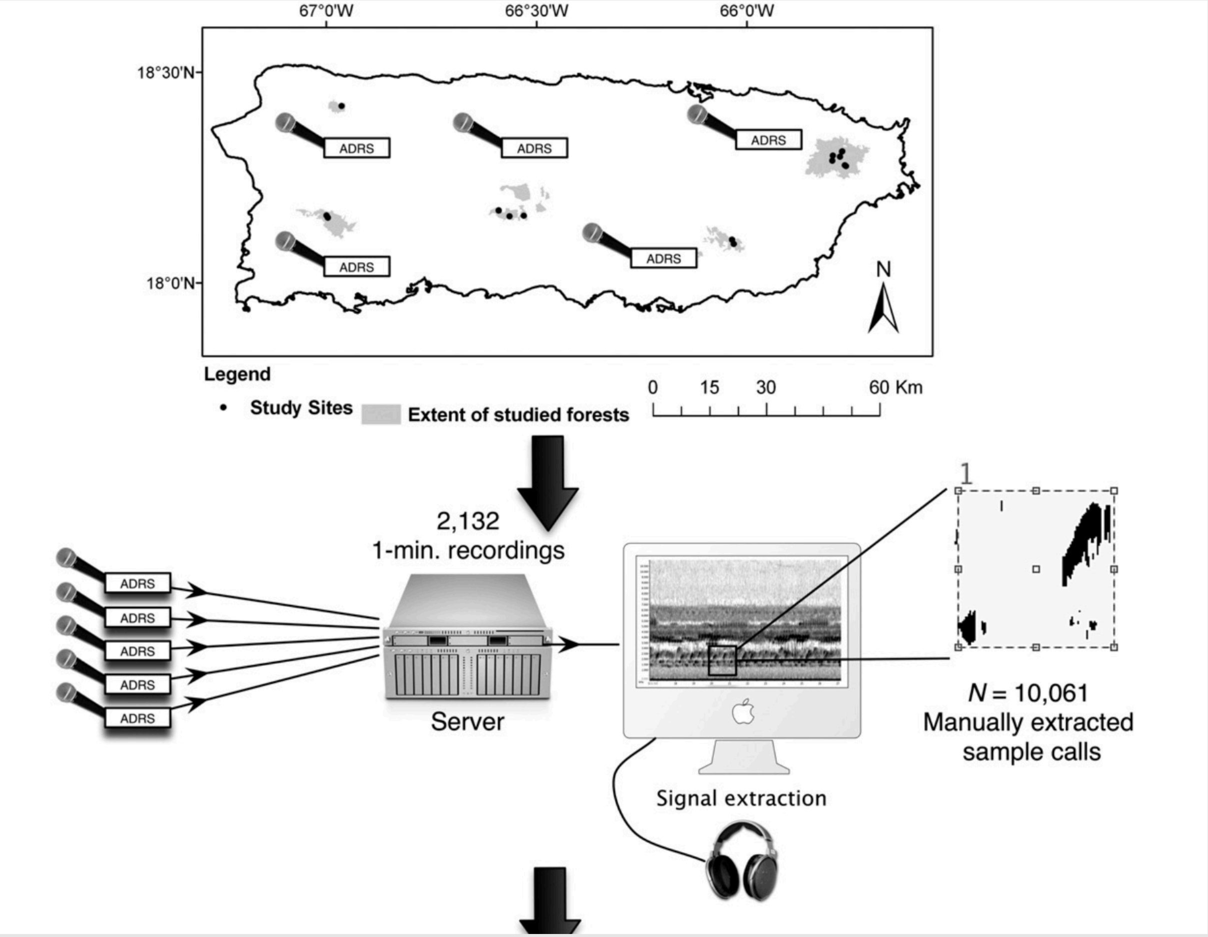


Fig. 9 Detection of 3t3-H-20× cell nuclei

Automated classification of bird and amphibian calls using machine learning: A comparison of methods

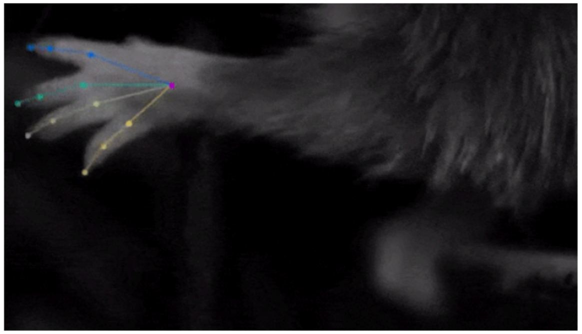
Miguel A. Acevedo [a,*], Carlos J. Corrada-Bravo [c], Héctor Corrada-Bravo [b], Luis J. Villanueva-Rivera [d], T. Mitchell Aide [a]

[a] *University of Puerto Rico, Department of Biology, Puerto Rico*
[b] *University of Wisconsin-Madison, Department of Computer Sciences, United States*
[c] *University of Puerto Rico, Department of Computer Science, Puerto Rico*
[d] *Purdue University, Department of Forestry and Natural Resources, United States*

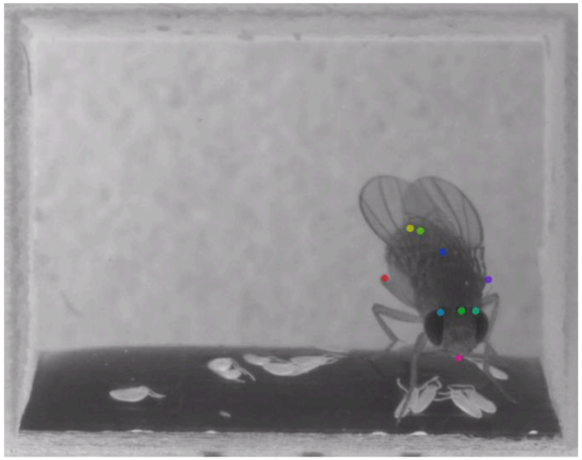


pypi package 2.1.6.2 | PyPi 3.1k/month | License LGPL v3 | forum 274 topics | chat on gitter | DeepLabCut 3.4k

DeepLabCut: a software package for animal pose estimation



DeepLabCut™ is an efficient method for **3D markerless pose estimation** based on transfer learning with **deep neural networks** that achieves excellent results (i.e. you can match human labeling accuracy) with minimal training data (typically 50-200 frames). We demonstrate the versatility of this framework by tracking various body parts in multiple species across a broad collection of behaviors.



The package is open source, fast, robust, and can be used to compute 3D pose estimates. Please see the original paper and the latest work below.

Where do you start?

PLoS One. 2012; 7(11): e49120.
Published online 2012 Nov 19. doi: 10.1371/journal.pone.0049120

PMCID: PMC3501513
PMID: 23185301

Movement Activity Based Classification of Animal Behaviour with an Application to Data from Cheetah *(Acinonyx jubatus)*

Steffen Grünewälder,[1,2,*] Femke Broekhuis,[3,4] David Whyte Macdonald,[4] Alan Martin Wilson,[5] John Weldon McNutt,[3] John Shawe-Taylor,[1,2] and Stephen Hailes[2,5]

Allan V. Kalueff, Editor

‣ Author information ‣ Article not

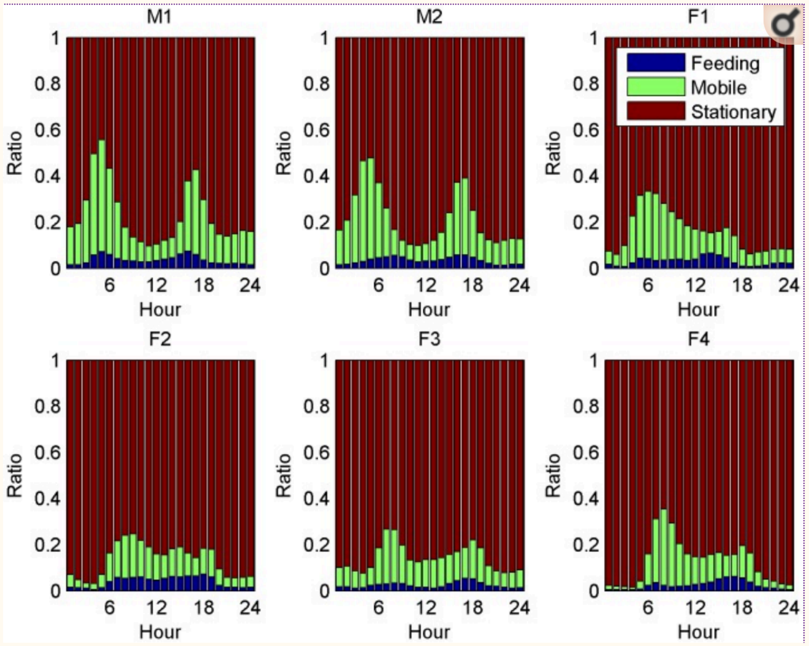


Figure 3

The classification of the daily activity of the six individuals is shown in the figure.

The activity is colour coded with feeding being at the bottom, mobile in the middle and stationary at the top. The activity is classified based on the hour of the day (local time) and normalized to one. Sunrise during wet season is between 5:19 and 6:25 and during the dry season between 5:30 and 7:01. Sunset during the wet season is between 18:16 and 19:11 and during the dry season between 17:35 and 18:31.

RESEARCH ARTICLE

Supervised accelerometry analysis can identify prey capture by penguins at sea

Gemma Carroll[1,*], David Slip[2], Ian Jonsen[2] and Rob Harcourt[2]

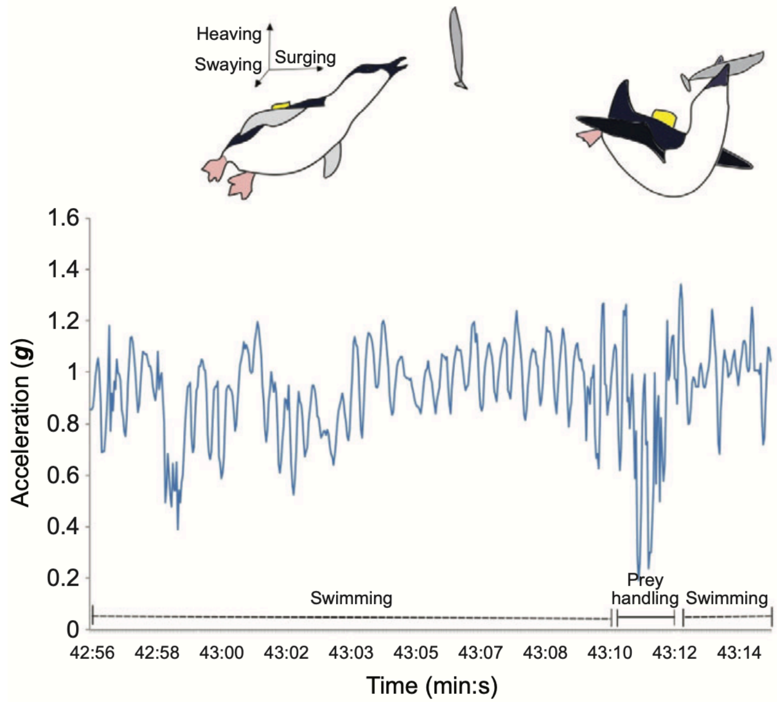


Fig. 2. Schematic diagram of a little penguin wearing an accelerometer (yellow) swimming towards a fish and handling prey. The accelerometer axes (heaving, surging and swaying; see Materials and methods) are shown on the left. Beneath is a sample raw accelerometry profile from the heaving axis (recorded at 30 Hz) of a penguin swimming and handling prey in captivity, labelled with the associated behaviours identified from HD video.

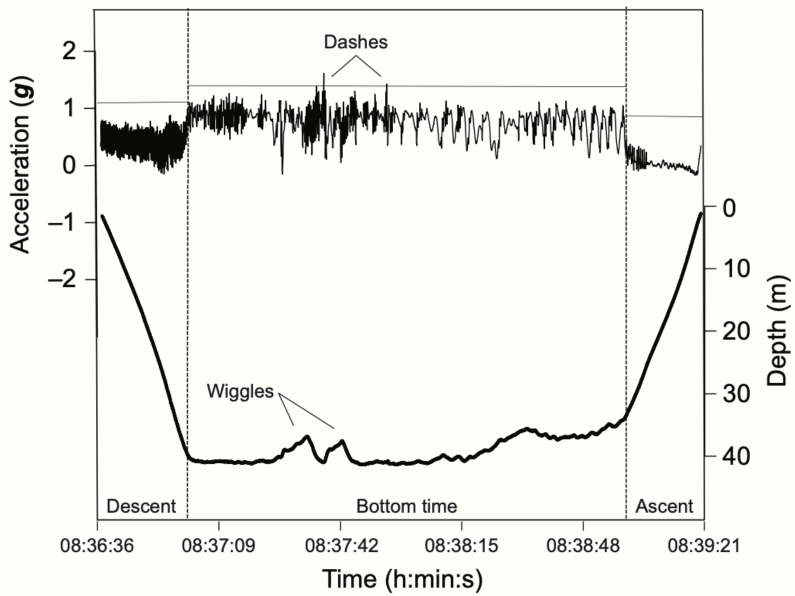


Fig. 4. Depth and acceleration in the vertical heaving axis during a sample little penguin dive that included both wiggles and dashes. Wiggles are undulations in the bottom phase of the dive occurring at >0.5 m s^{-1}; dashes are spikes above an acceleration threshold determined using a survival curve. Both wiggles and dashes have been used as proxies for prey encounter and are included in this paper to understand their relationship with predictions from the SVM estimates of prey capture.

Deep Learning Object Detection Methods for Ecological Camera Trap Data

Stefan Schneider*, Graham W. Taylor†, Stefan C. Kremer*
*School of Computer Science, University of Guelph
{sschne01, skremer}@uoguelph.ca
†School of Engineering, University of Guelph
gwtaylor@uoguelph.ca
†Vector Institute for Artificial Intelligence
†Canadian Institute for Advanced Research

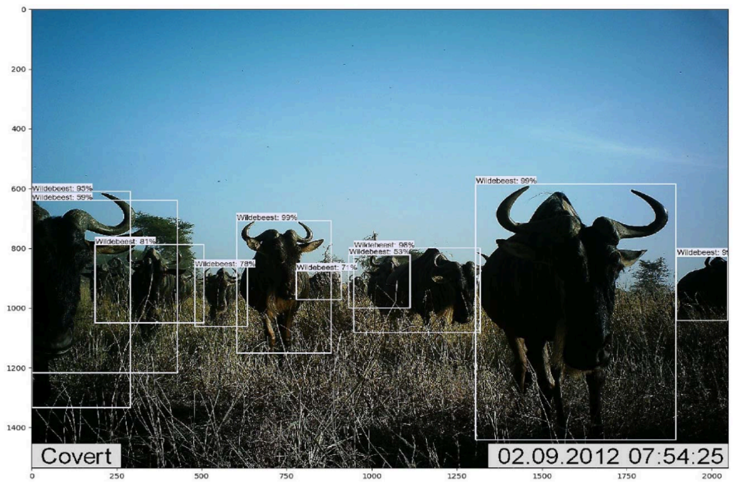


Fig. 4. Faster R-CNN output returning 10 Wildebeest from the GSSS data set, demonstrating one example of the high levels of obstruction within the data set.

ig. 6. Faster R-CNN output returning 10 Gazelle Thomsons from the ata set, demonstrating the difficulties of distances.

Fig. 1. Faster R-CNN output returning 1 White Nosed Agouti from the RCT data set in a highly camouflaged environment.

Segmentation

https://github.com/zhixuhao/unet

U-Net: Convolutional Networks for Biomedical Image Segmentation

Olaf Ronneberger, Philipp Fischer, and Thomas Brox

Computer Science Department and BIOSS Centre for Biological Signalling Studies,
University of Freiburg, Germany
ronneber@informatik.uni-freiburg.de,
WWW home page: http://lmb.informatik.uni-freiburg.de/

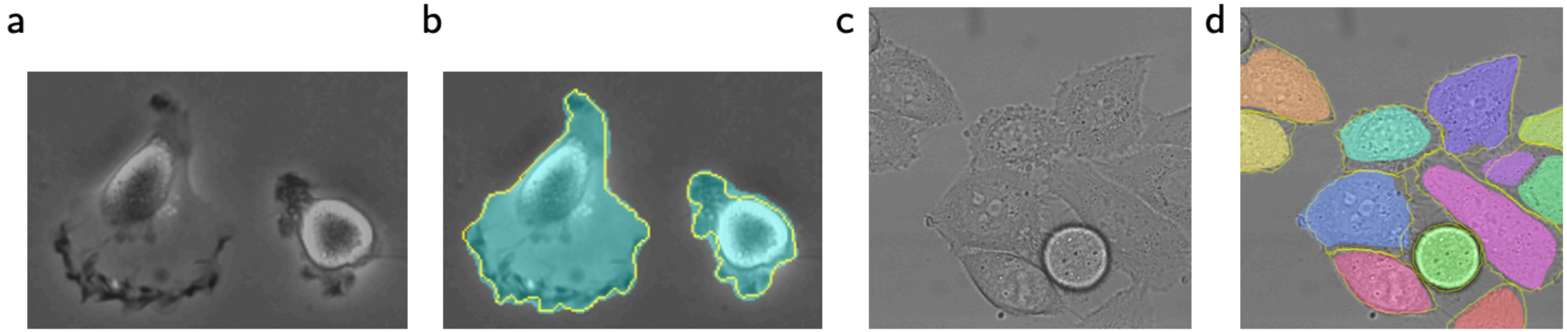


Fig. 4. Result on the ISBI cell tracking challenge. (**a**) part of an input image of the "PhC-U373" data set. (**b**) Segmentation result (cyan mask) with manual ground truth (yellow border) (**c**) input image of the "DIC-HeLa" data set. (**d**) Segmentation result (random colored masks) with manual ground truth (yellow border).

Object (phenotype) recognition

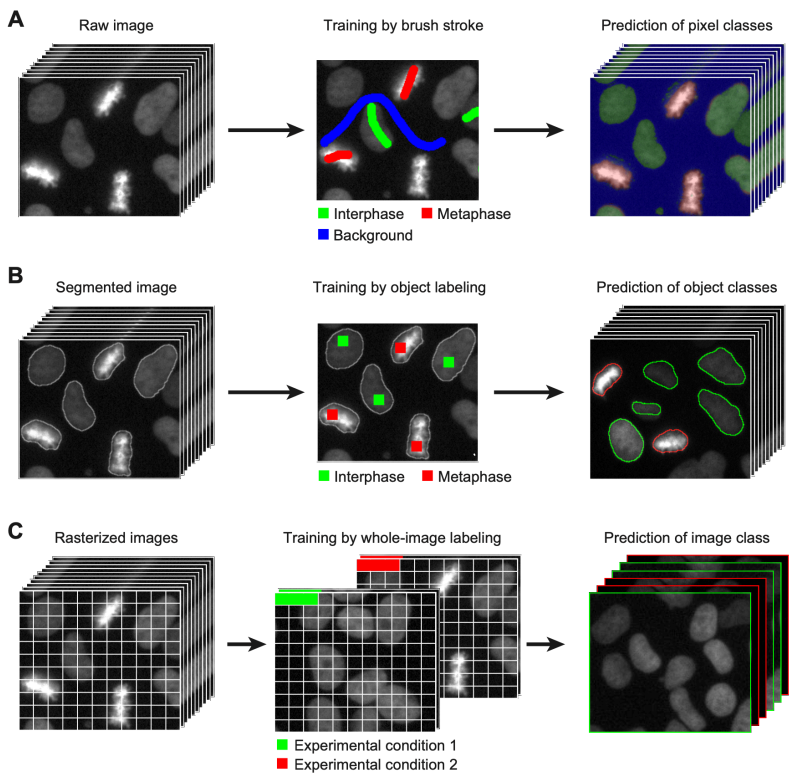
ARTICLE SERIES: Imaging Commentary 5529

Machine learning in cell biology – teaching computers to recognize phenotypes

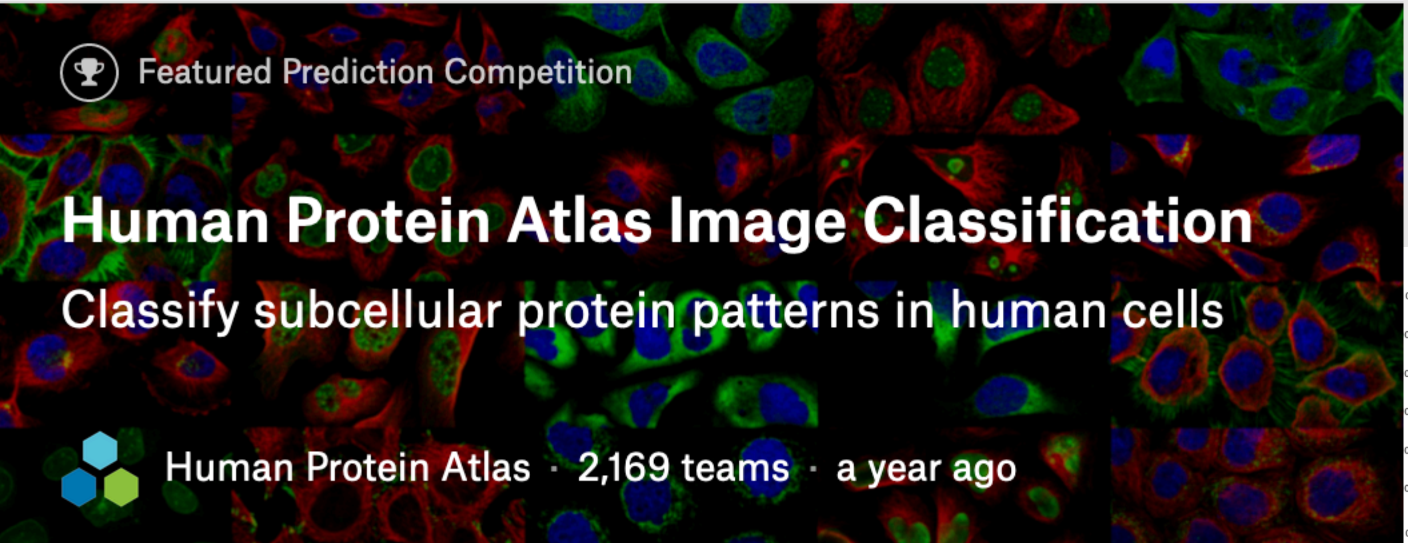
Christoph Sommer and Daniel W. Gerlich*
Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), 1030 Vienna, Austria
*Author for correspondence (daniel.gerlich@imba.oeaw.ac.at)

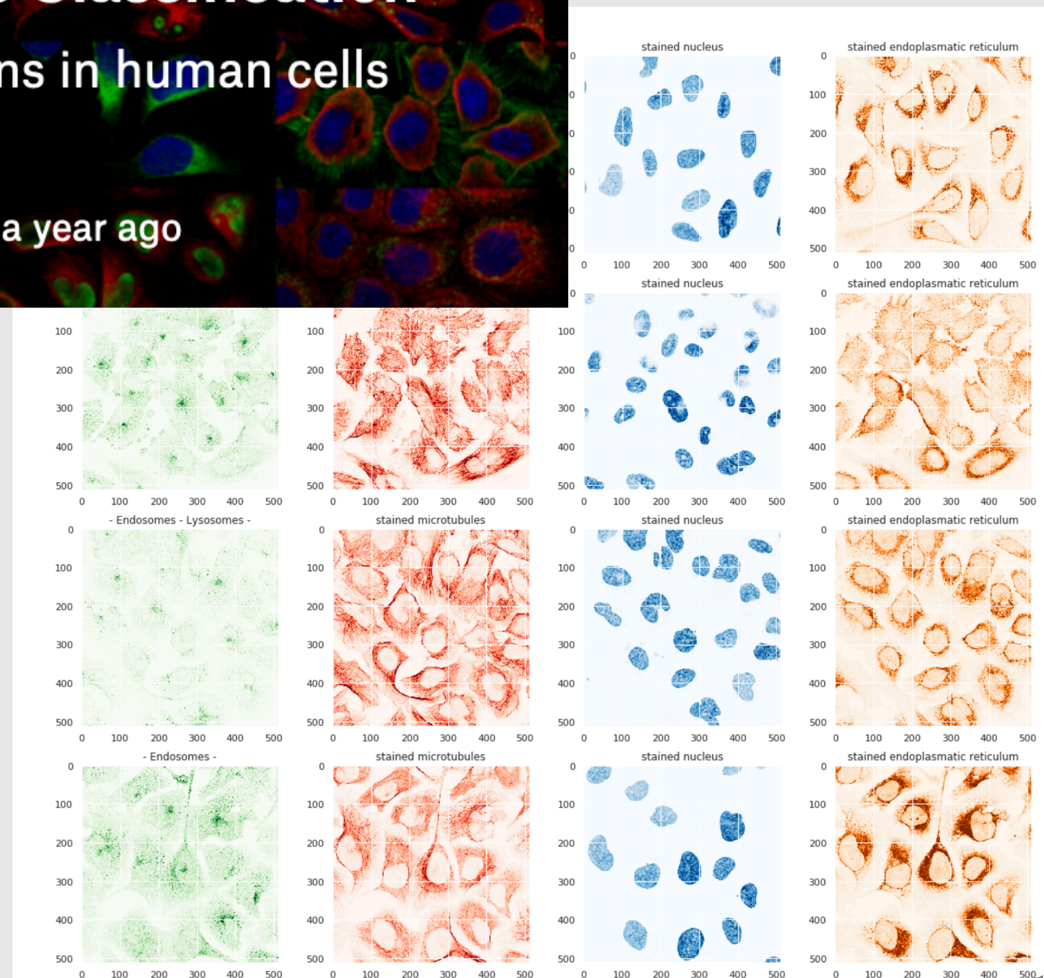
Journal of Cell Science 126, 5529–5539

doi: 10.1242/jcs.123604



Object (protein) detection





Predicting protein organelle localization labels for each sample.

DEEP LEARNING FOR LIFE SCIENCES

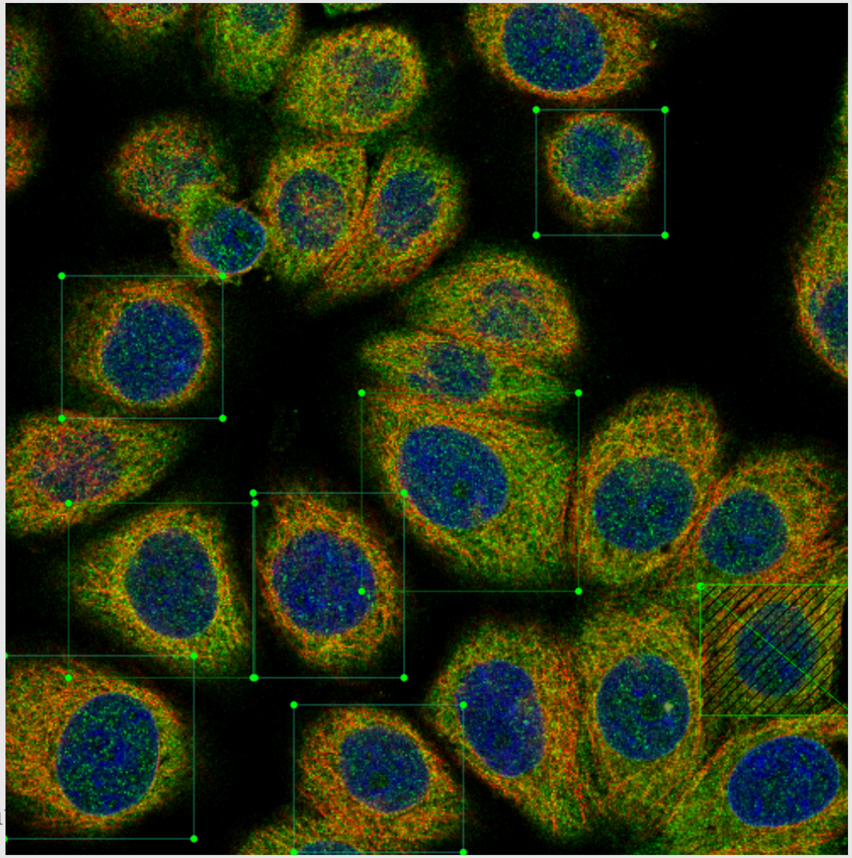
Deep Learning on Microscopy Imaging

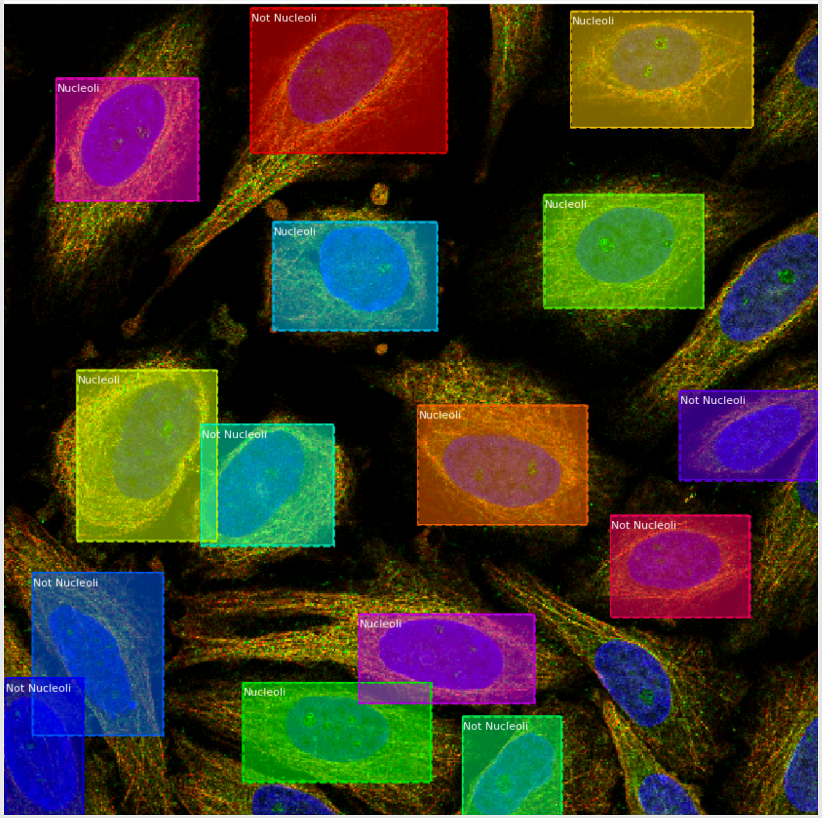
Detecting Good, Bad and Ugly Cells with Deep Learning

Nikolay Oskolkov Follow

Jul 7 · 9 min read ★

https://towardsdatascience.com/deep-learning-on-microscopy-imaging-865b521ec47c





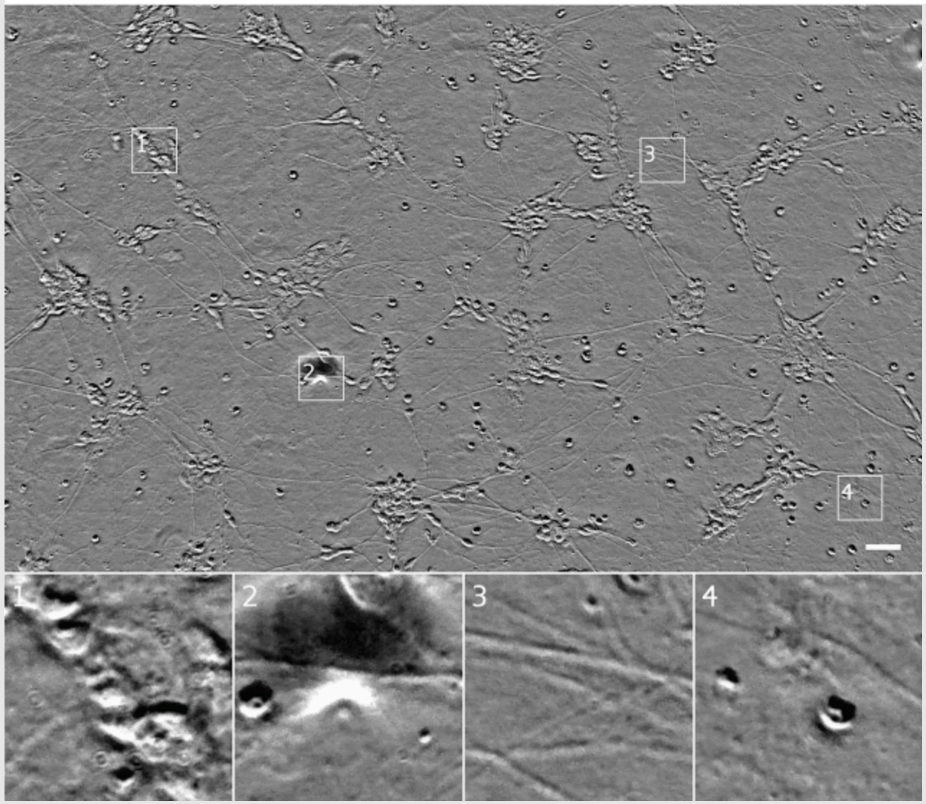
gith

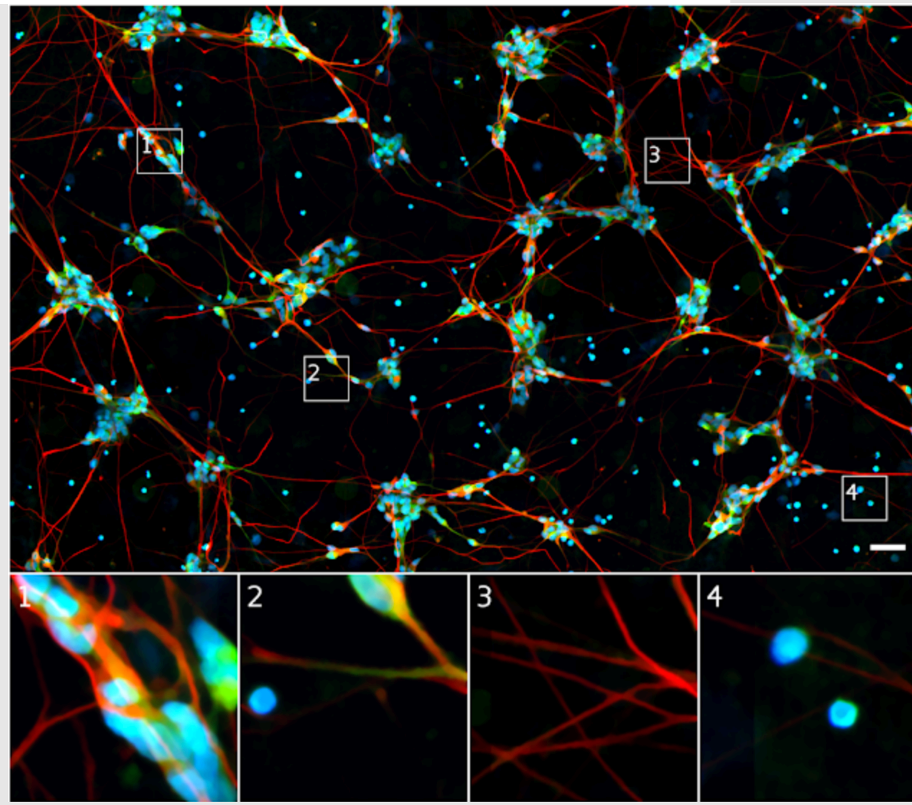
Deep Learning prediction

Seeing More with In Silico Labeling of Microscopy Images

Thursday, April 12, 2018

Eric Christiansen, Senior Software Engineer, Google Research





History of ML

- Early 20th century – Modern Statistics
- 1950's – Artificial Intelligence
- 1990's – Traditional Machine Learning
- 2000's – Deep Learning

Plan for today

- Session 1: **Basic Predictive Models** (statistics differently)
- ~~Session 2: **Features** (representing images in computer world)~~
- Session 3: **Unsupervised learning**
- Session 4: **Supervised learning**
- Session 5: **Deep Learning**

Acknowledgments

The content is heavily based on materials by Toby Breckon and Andrew Ng.

https://www.coursera.org/learn/machine-learning

http://breckon.eu/toby/teaching/mltutorial/

coursera

Stanford University