Computer vision in medicine and life science

Sonja Aits

Lund University, 2025-09-23

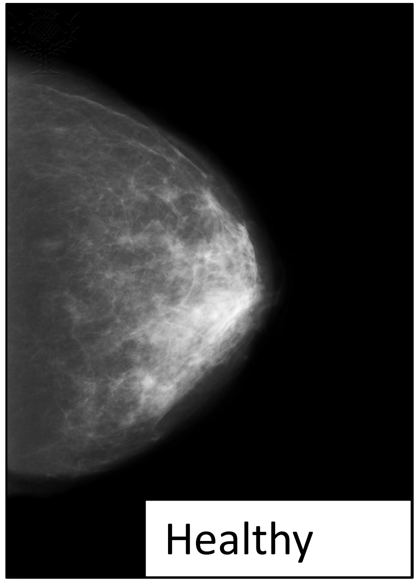
Computer Vision

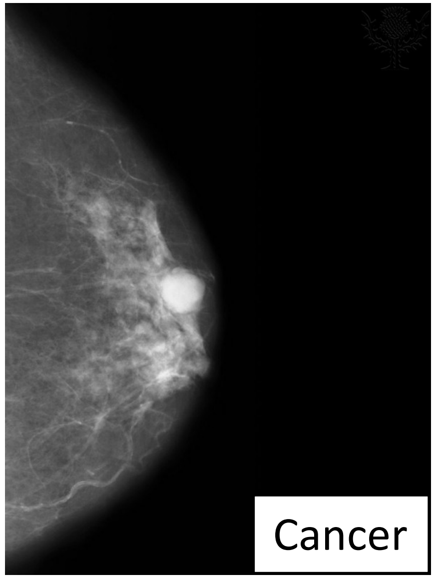
= AI for processing, analysis or generation of images or videos

Computer vision tasks

Image classification

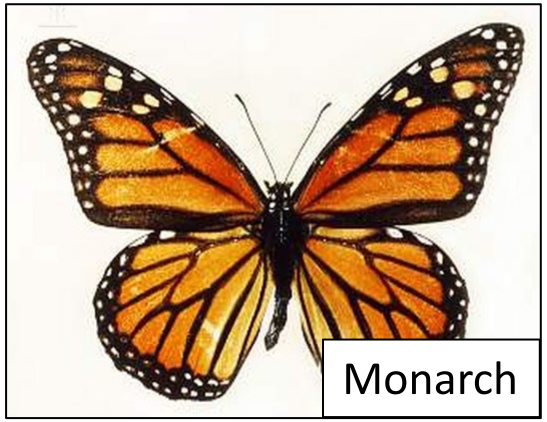
Binary

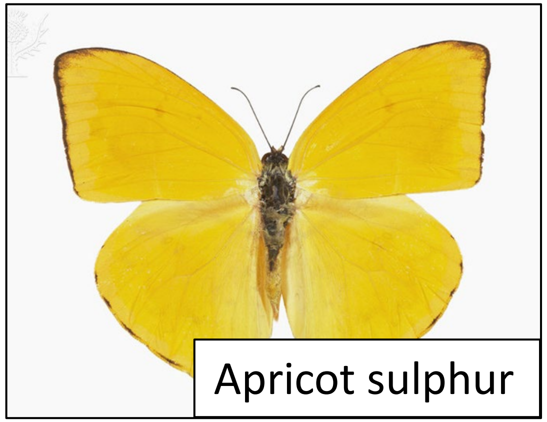




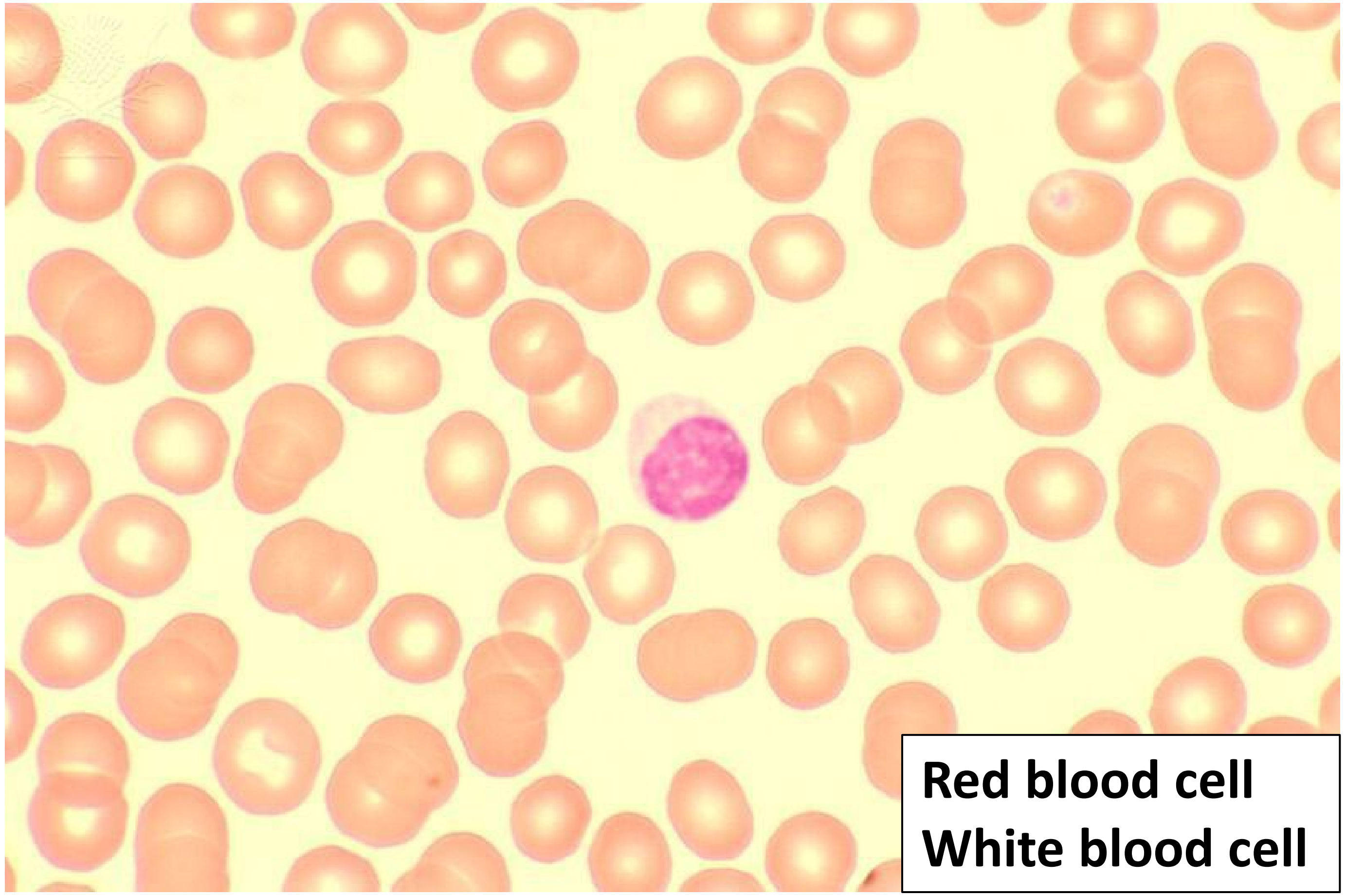
Output:
Image class

Multi-class



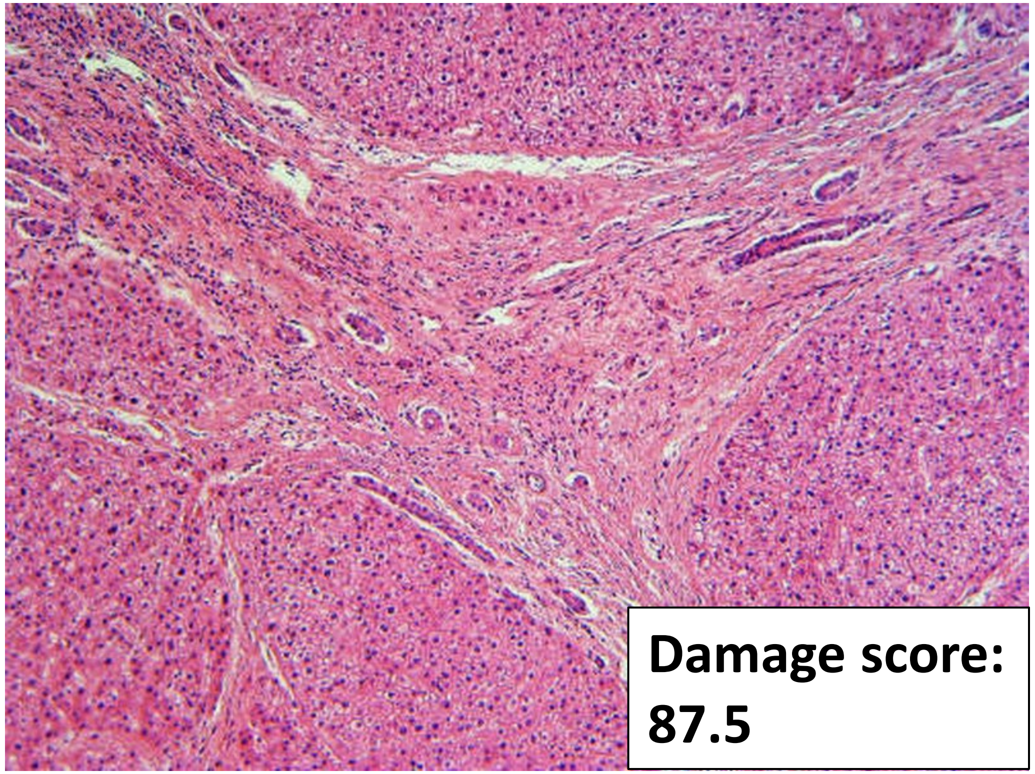


Multi-label image classification



Output:
Multiple
image classes

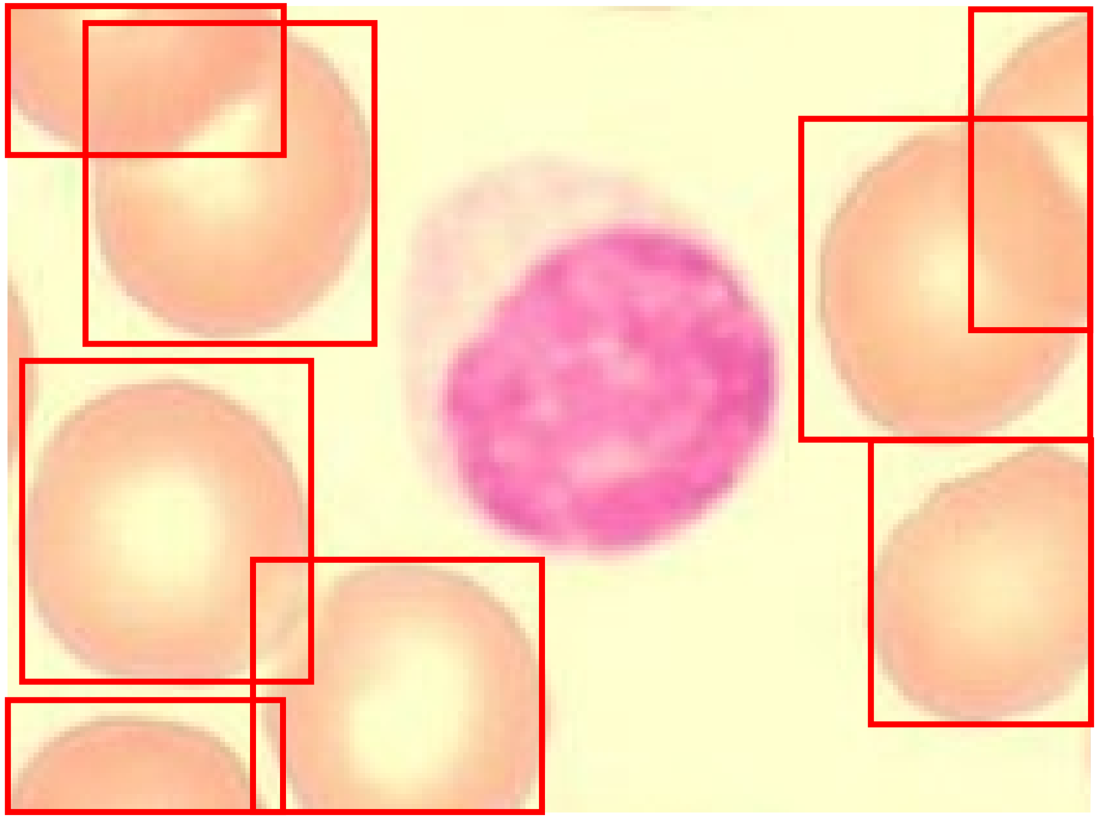
Image regression



Output:
Continuous value

Object detection

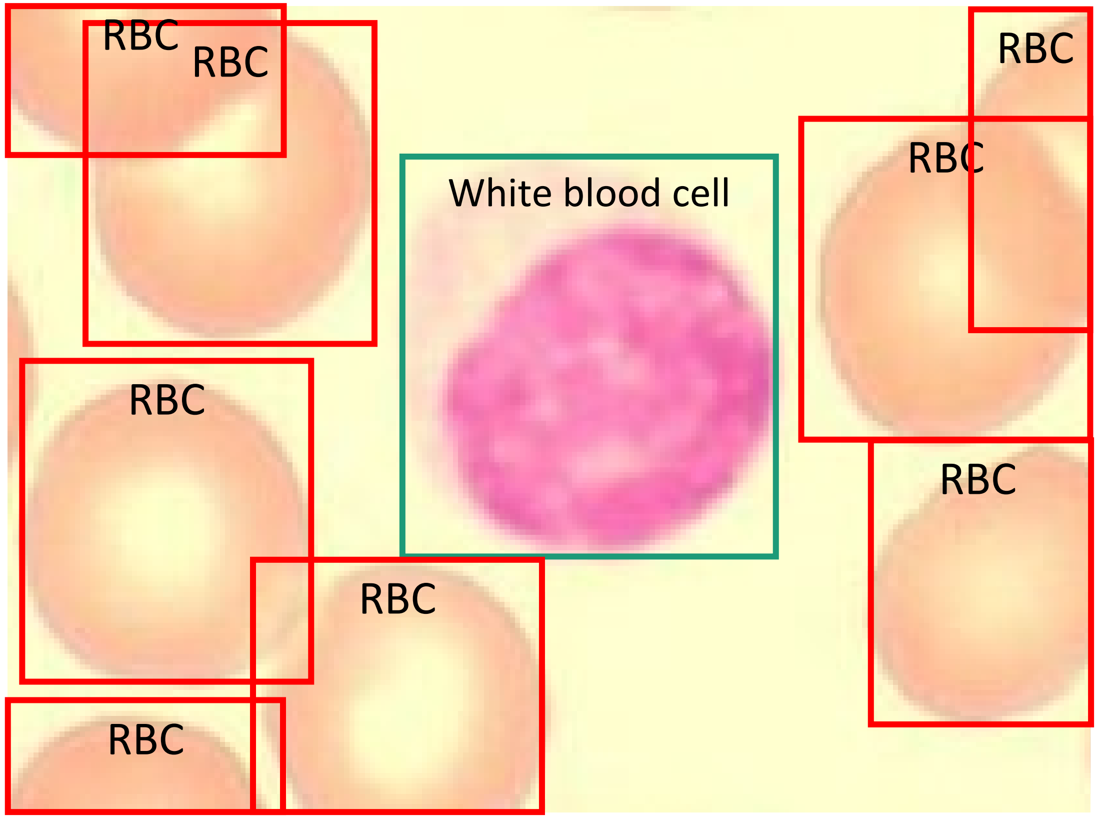
Single-class



Output:
Set of bounding boxes
(x, y, width, height)

Object detection

Multi-class

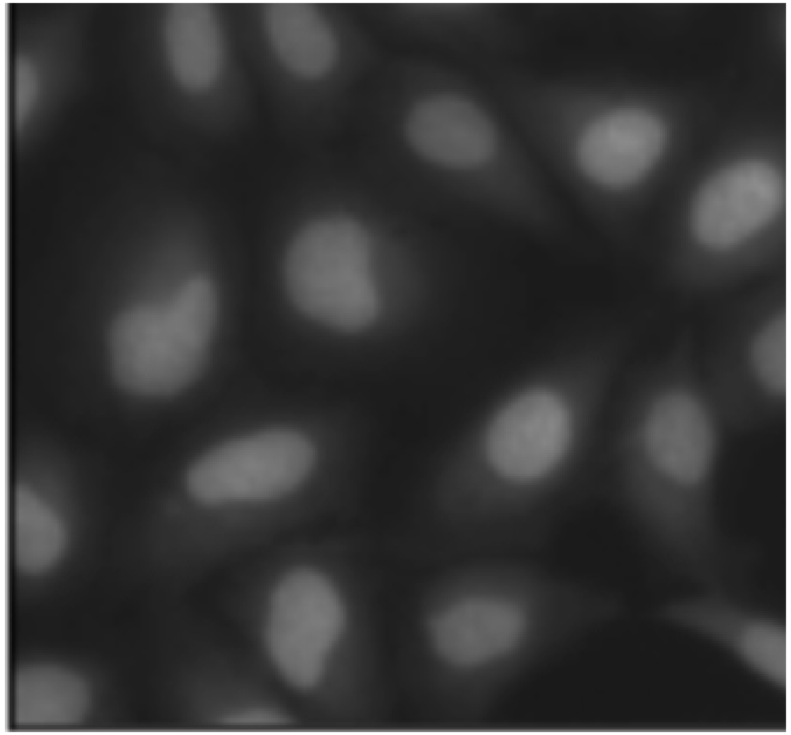


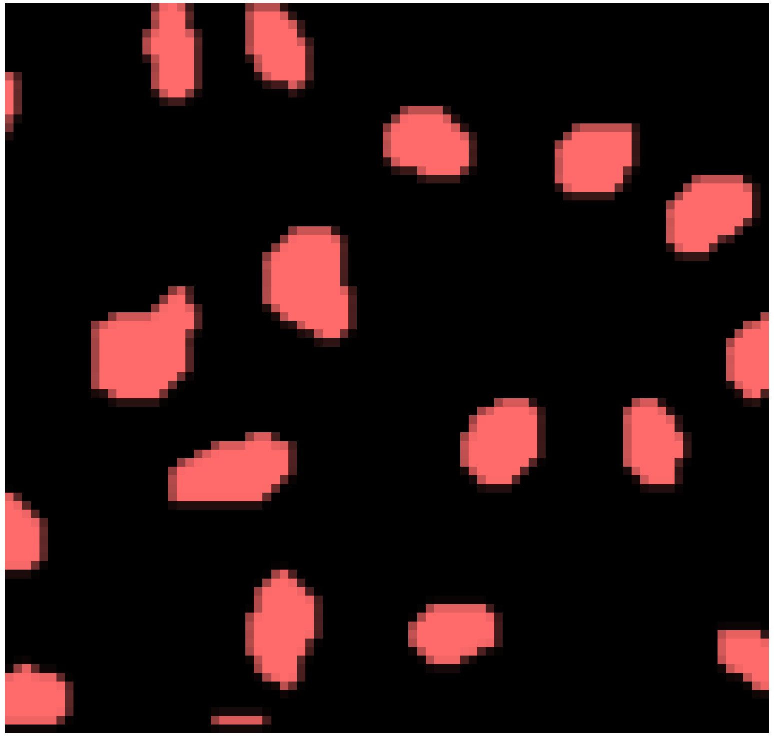
Output:
Set of bounding boxes with object class

Object tracking

https://www.youtube.com/watch?v=2Pq5upuFtrA

Semantic segmentation





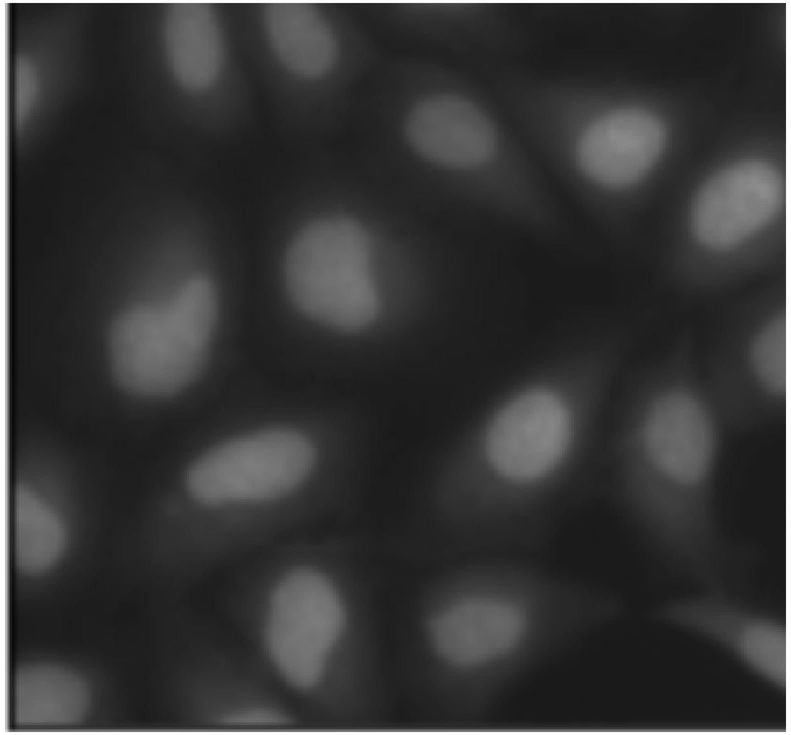
Background

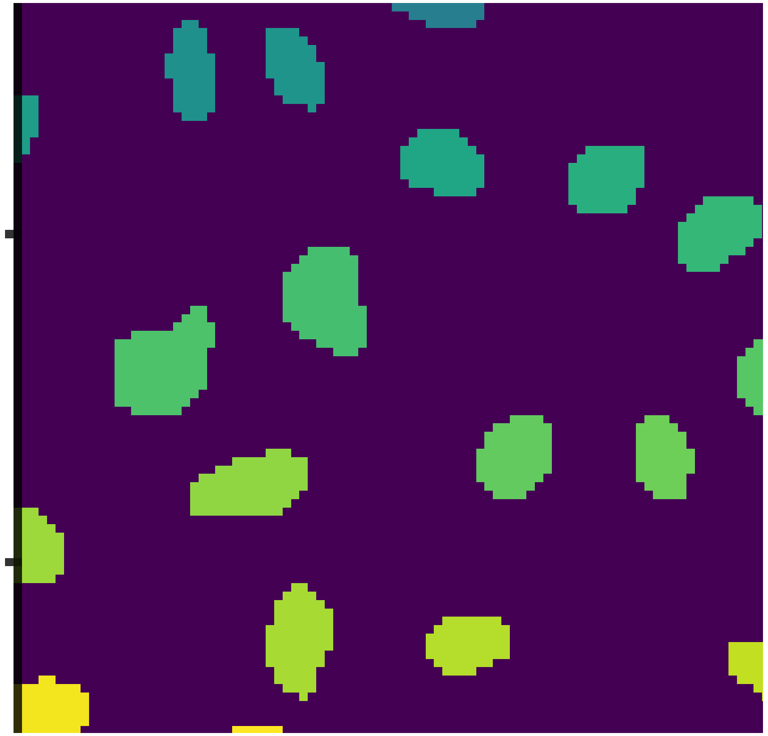
Cells

Output:
Class of each pixel

No information on individual objects!

Instance segmentation





Output:
Outlined objects

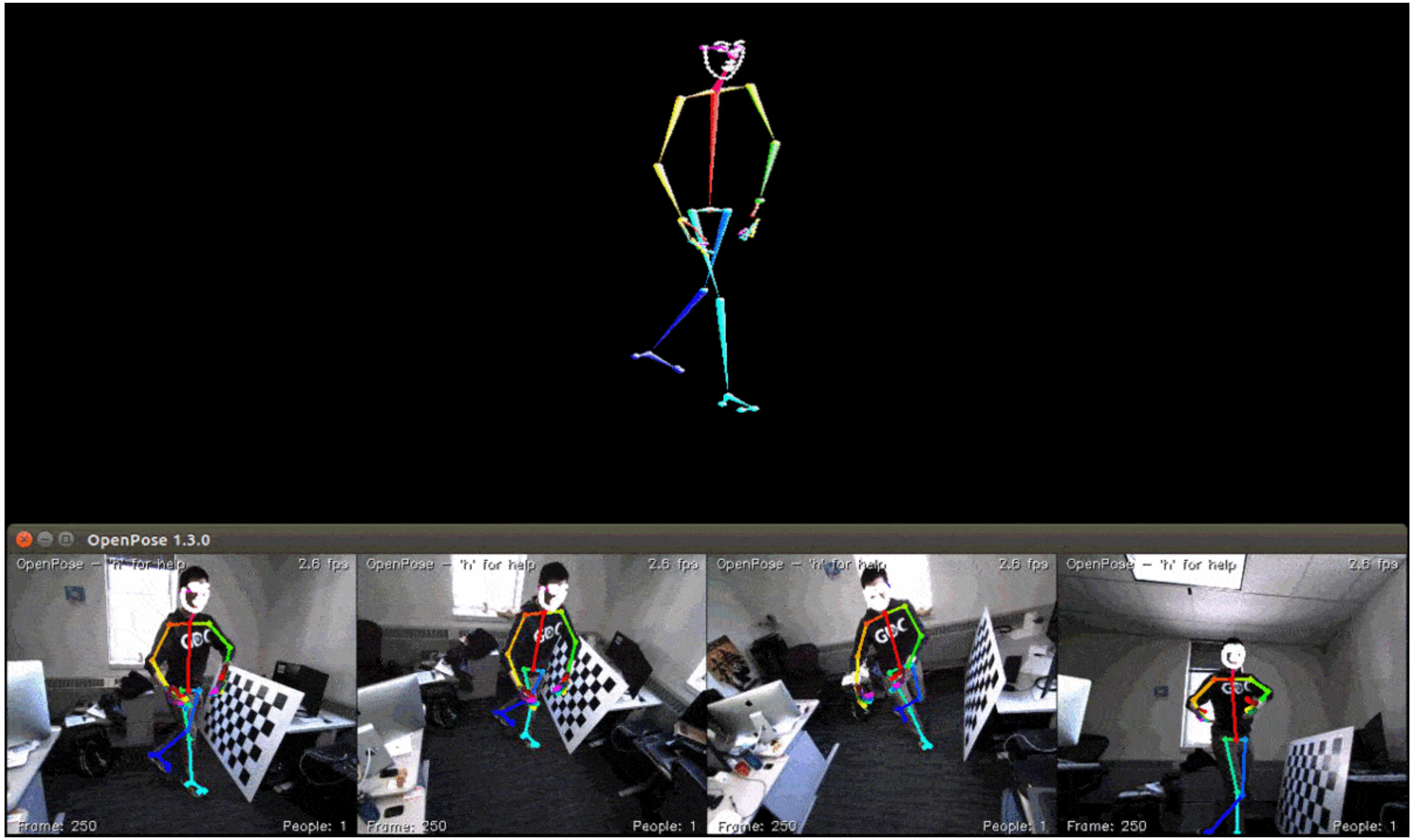
Mini quiz: Object detection, Instance segmentation, Semantic segmentation

- Counting bees in a photograph
- Calculating the area covered by forest in a satellite image
- Determining the average size of bacterial colonies in a photograph
- Determining the average level of green fluorescent protein in cells
- Calculating the circularity of cell nuclei
- Determining how many types of vegetation can be seen in drone images

Pose estimation

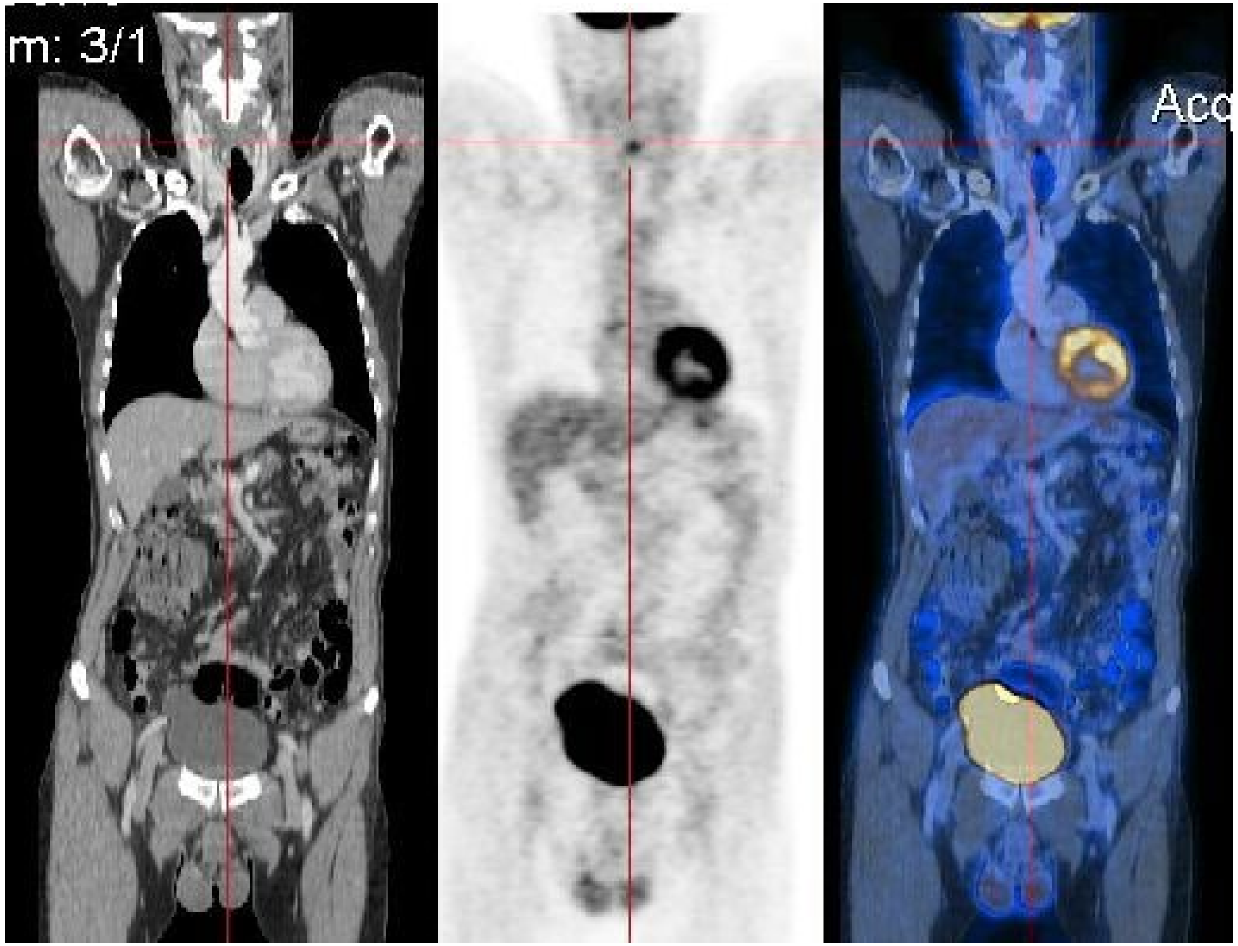
https://github.com/CMU-Perceptual-Computing-Lab/openpose/

Pose estimation: 3D



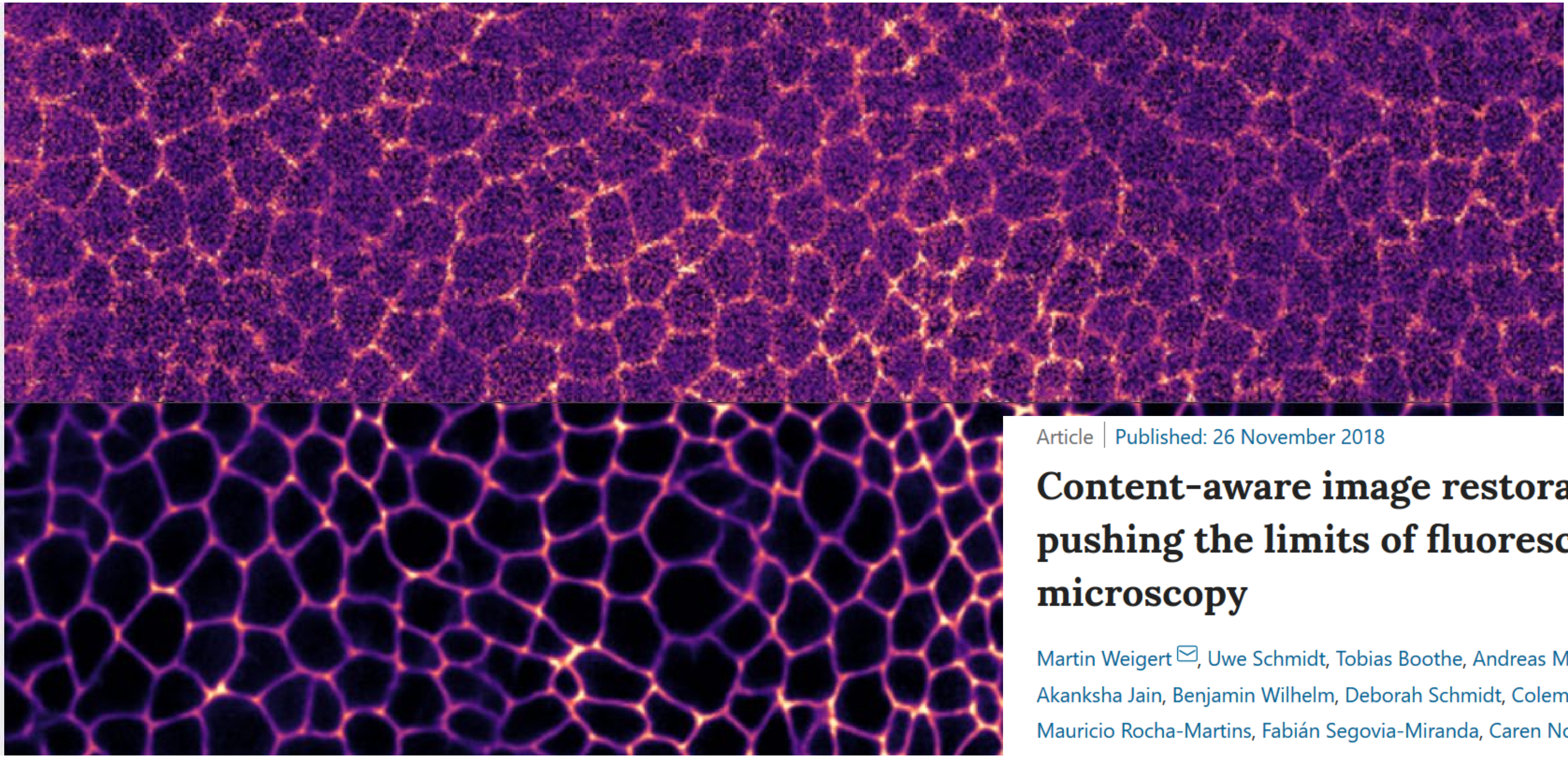
https://github.com/CMU-Perceptual-Computing-Lab/openpose/

Image registration



https://commons.wikimedia.org/w/index.php?curid=20682407

Image restoration



Article | Published: 26 November 2018

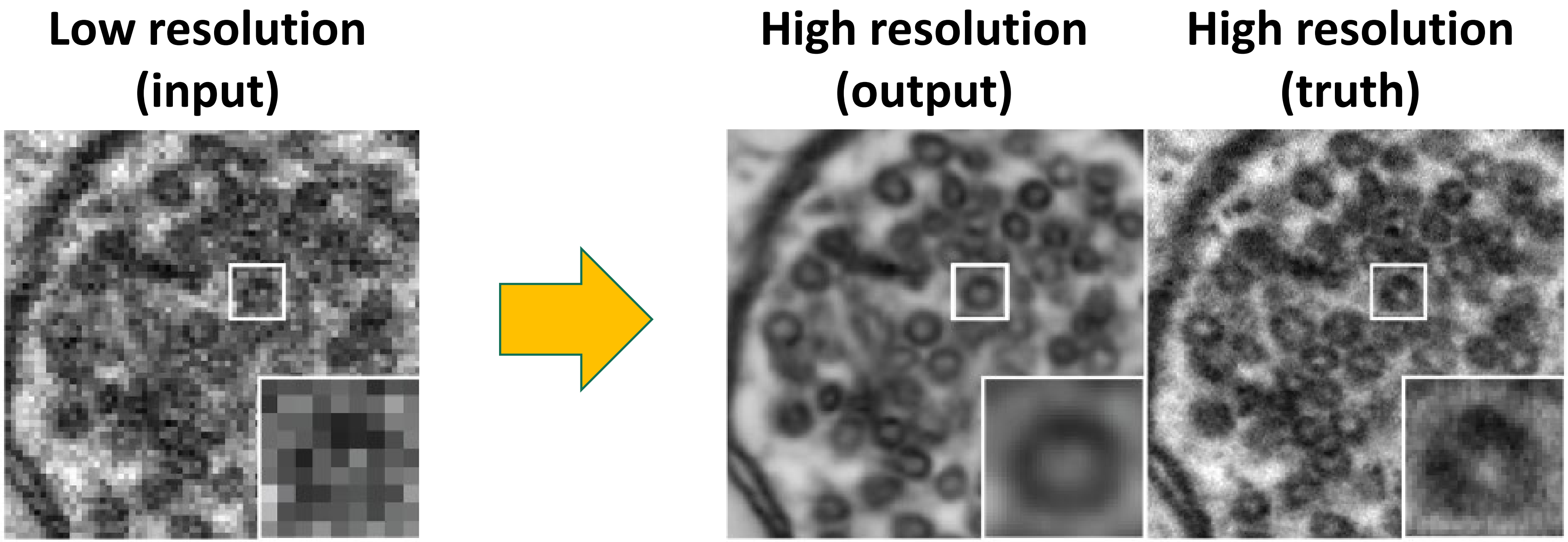
Content-aware image restoration: pushing the limits of fluorescence microscopy

Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, Mauricio Rocha-Martins, Fabián Segovia-Miranda, Caren Norden, Ricardo Henriques, Marino Zerial, Michele Solimena, Jochen Rink, Pavel Tomancak, Loic Royer, Florian Jug & Eugene W. Myers

Nature Methods **15**, 1090–1097(2018) | Cite this article

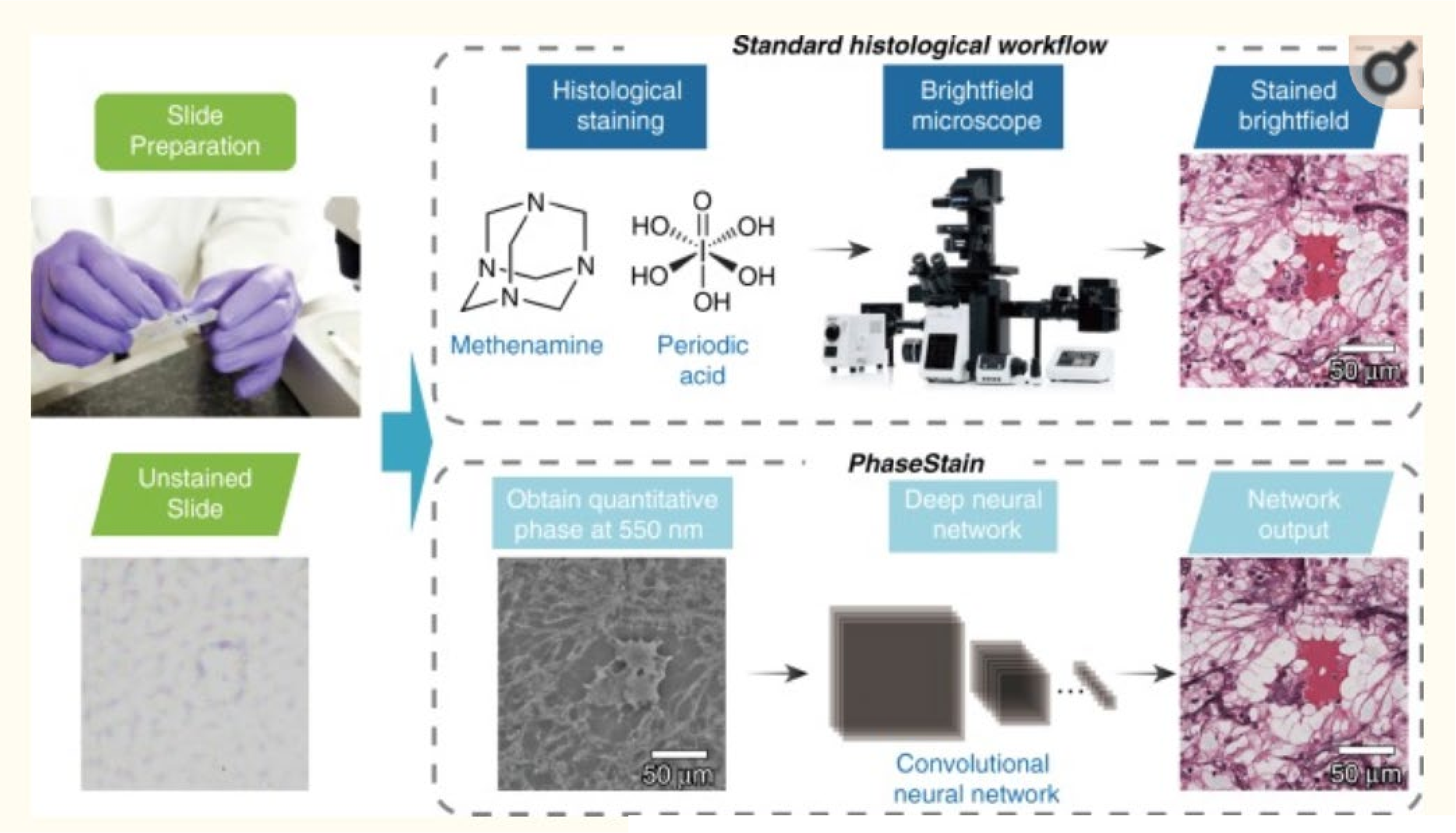
combined surface prediction + denoising

Super-resolution



https://www.biorxiv.org/content/10.1101/740548v7.full

Virtual staining

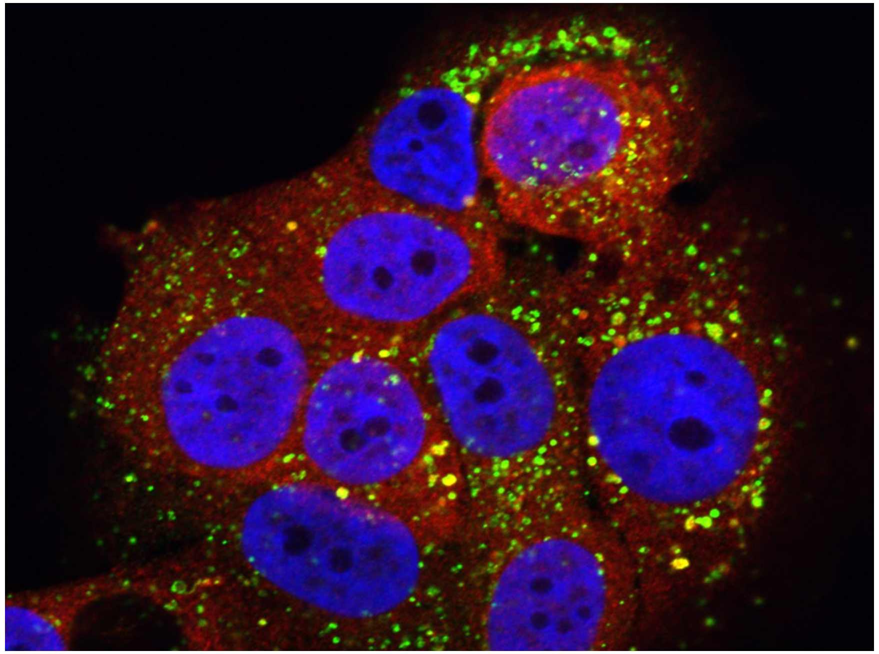


[PhaseStain: the digital **staining** of label-free quantitative **phase** microscopy images using deep learning.]

Rivenson Y, Liu T, Wei Z, Zhang Y, de Haan K, **Ozcan** A.

Light Sci Appl. 2019 Feb 6;8:23. doi: 10.1038/s41377-019-0129-y. eCollection 2019.

Image captioning



Output:
Descriptive word sequence

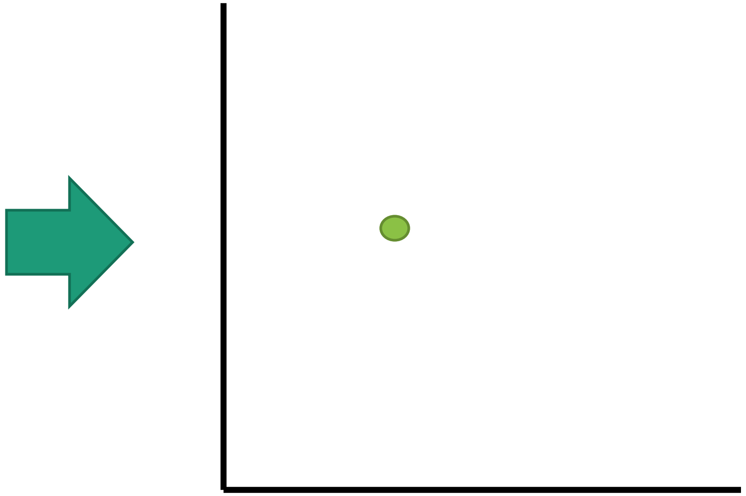
“a fluorescence microscopy image showing a group of cells with blue nuclei and bright green cytosolic spots”

Image clustering

Output:
Grouped data

Dimensionality reduction

High-dimensional data



Low-dimensional data

Output:
Low dimensional data

https://en.wikipedia.org/wiki/Hyperspectral_imaging#/media/File:HyperspectralCube.jpg

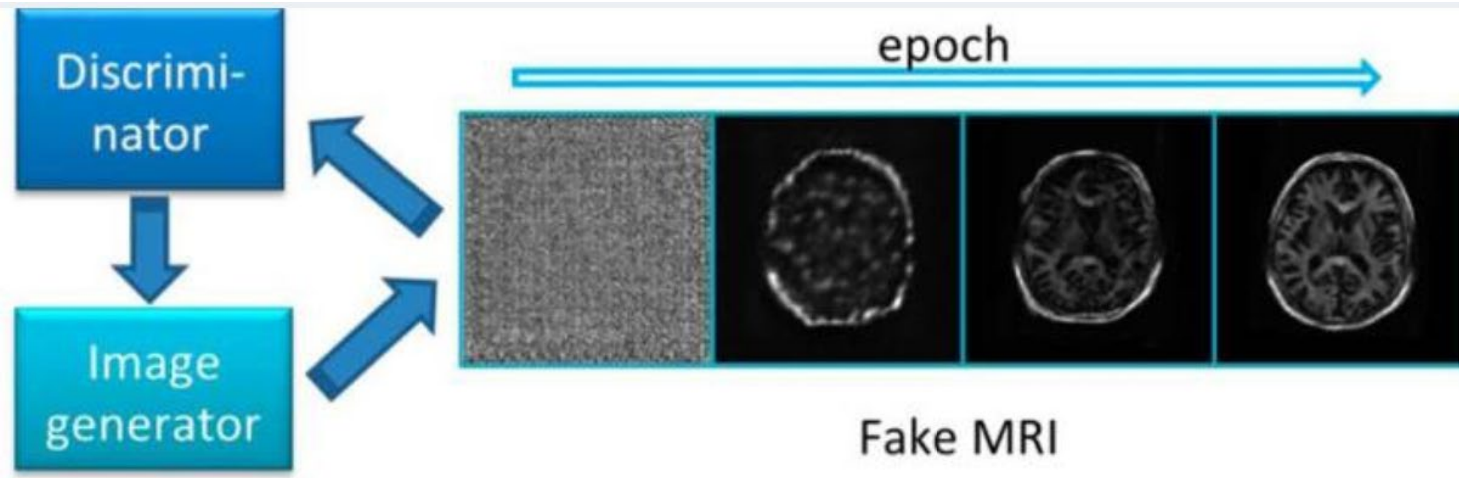
Image generation

https://thispersondoesnotexist.com/

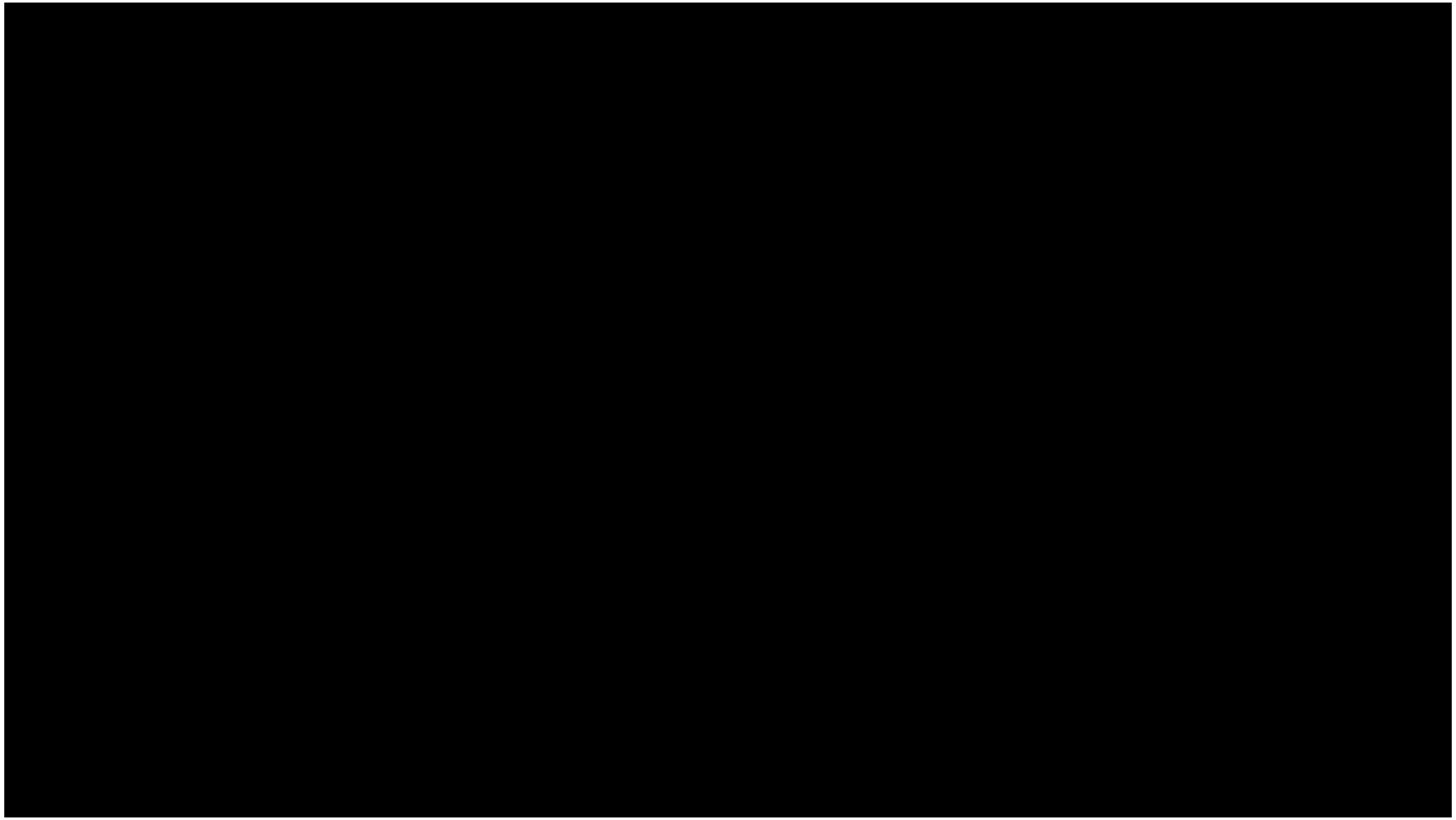
Tomography. 2018 Dec;4(4):159-163. doi: 10.18383/j.tom.2018.00042.

Generative Adversarial Networks for the Creation of Realistic Artificial Brain Magnetic Resonance Images.

Kazuhiro K[1], Werner RA[2,3,4], Toriumi F[5], Javadi MS[2], Pomper MG[2,6,7], Solnes LB[2], Verde F[7], Higuchi T[1,3,4], Rowe SP[2,6,7].



Autonomous robotic surgery

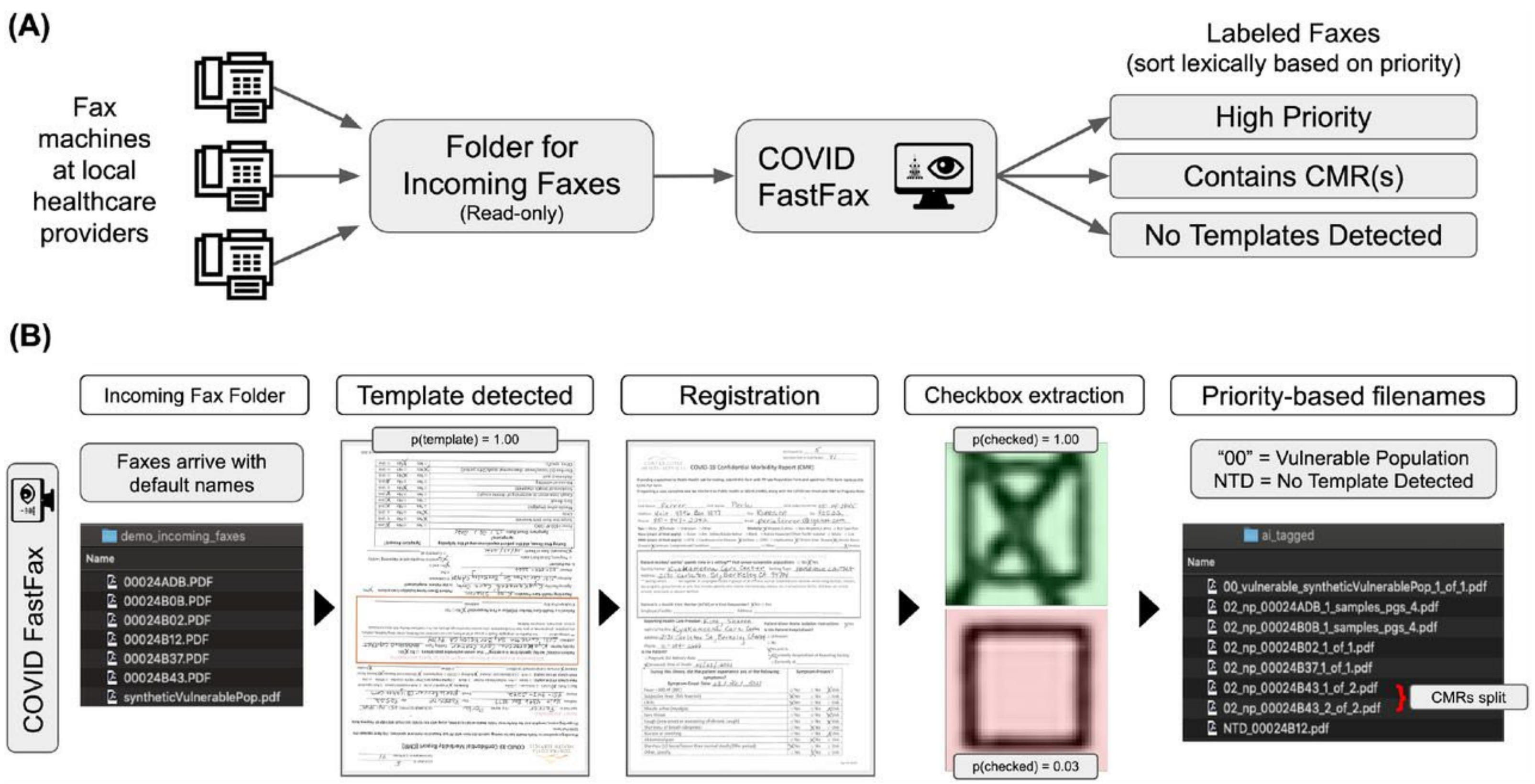


https://youtu.be/XeYeAzsQ1-U?feature=shared

Computer vision-driven assistant devices

https://youtu.be/vAGybwLb3kg?feature=shared

Assisting hospital logistics: COVID FastFax



Lavertu 2020, medRxiv https://www.medrxiv.org/content/10.1101/2020.12.15.20248256v1

Supervised learning

Labeled data (x,y)

Defined target

Aim: mapping function x -> y

Common uses: regression, classification

Unsupervised learning

Unlabeled data

No defined target

Aim: Learn structure of the data

Common uses: clustering, dimensionality reduction

Mini quiz: Supervised or unsupervised?

- Predicting whether patient has COVID-19 from a CT scan of the lung
- Determining the severity grade of a tumor from histology images
- Clustering cells of similar shape from microscopy images
- Tracking bird flightpaths in a video
- Outlining areas of deforestation on satellite images
- Flagging blurry MRI images
- Finding unknown types of image corruption in a large dataset

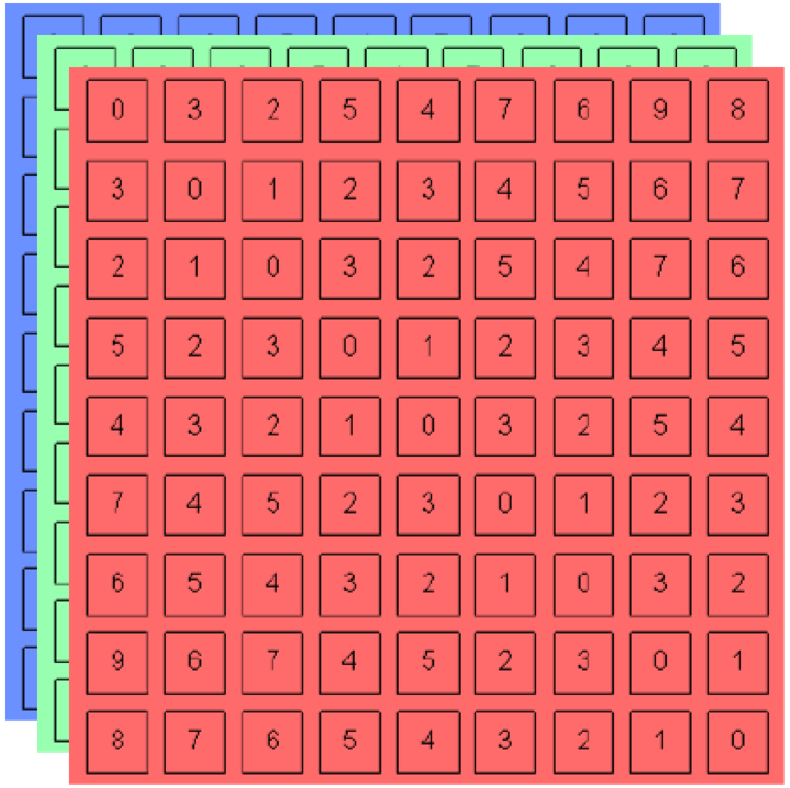
The technical side…

How do computers perceive images and videos?

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

Pixel intensities:
8-bit greyscale: 0 – 255 (2^8)
16-bit greyscale: 0 - 65 535 (2^{16})

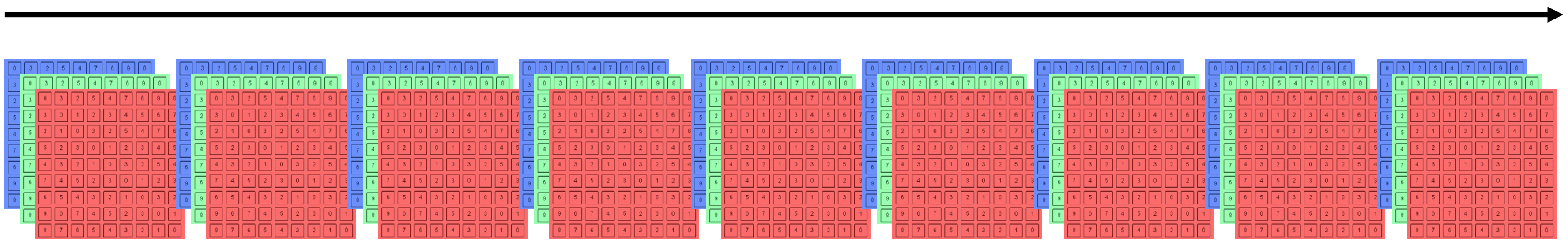
How do computers perceive images and videos?



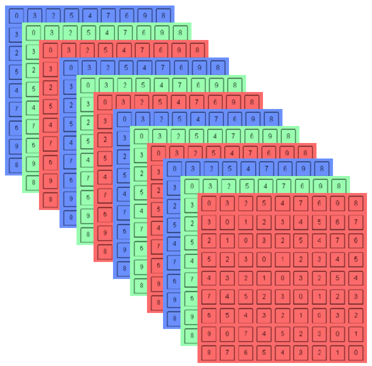
RGB

How do computers perceive images and videos?

Video

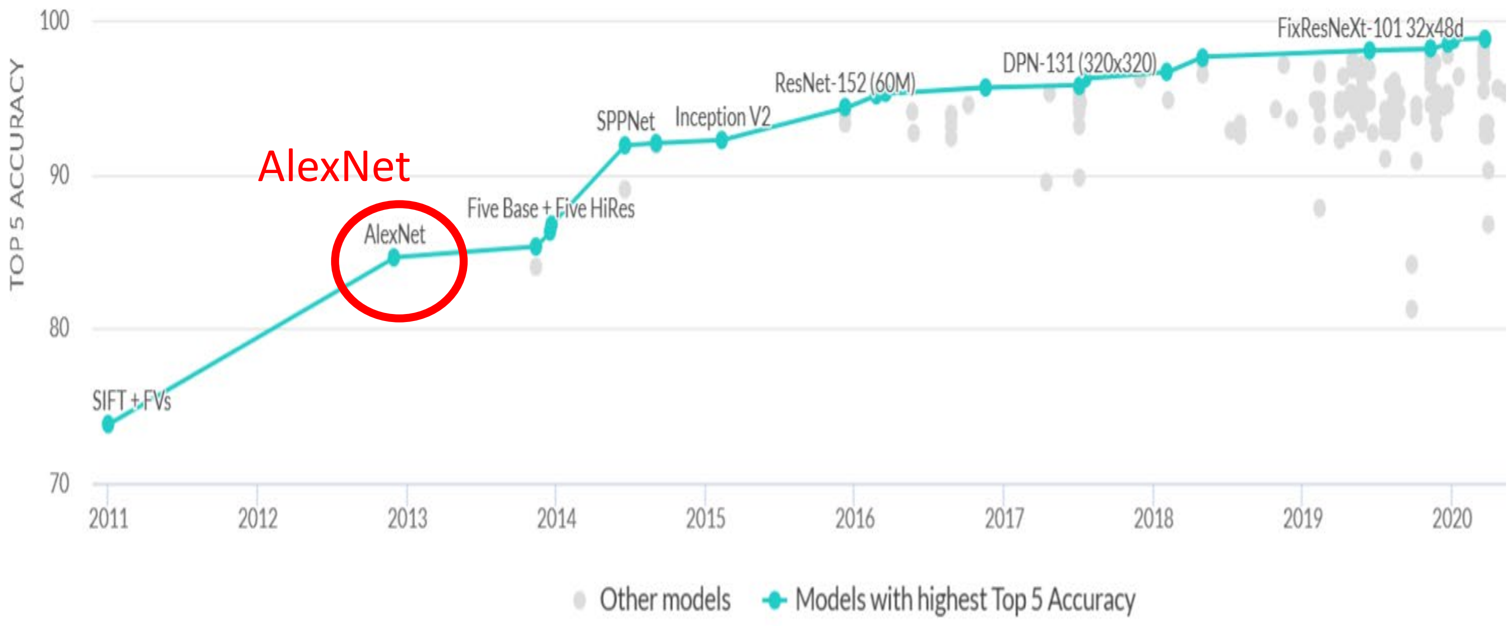


3D



AlexNet: the breakthrough of GPU-driven deep learning for computer vision

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)



https://paperswithcode.com/sota/image-classification-on-imagenet?metric=Top%205%20Accuracy
https://cs.stanford.edu/people/karpathy/cnnembed/

Convolutional neural networks (CNNs)

- Widely used for computer vision
- Ideal for recognizing small details
- Characterized by convolutional layers
- Neurons arranged in 3D: width, height, depth
- Neurons connected to small area of previous layer

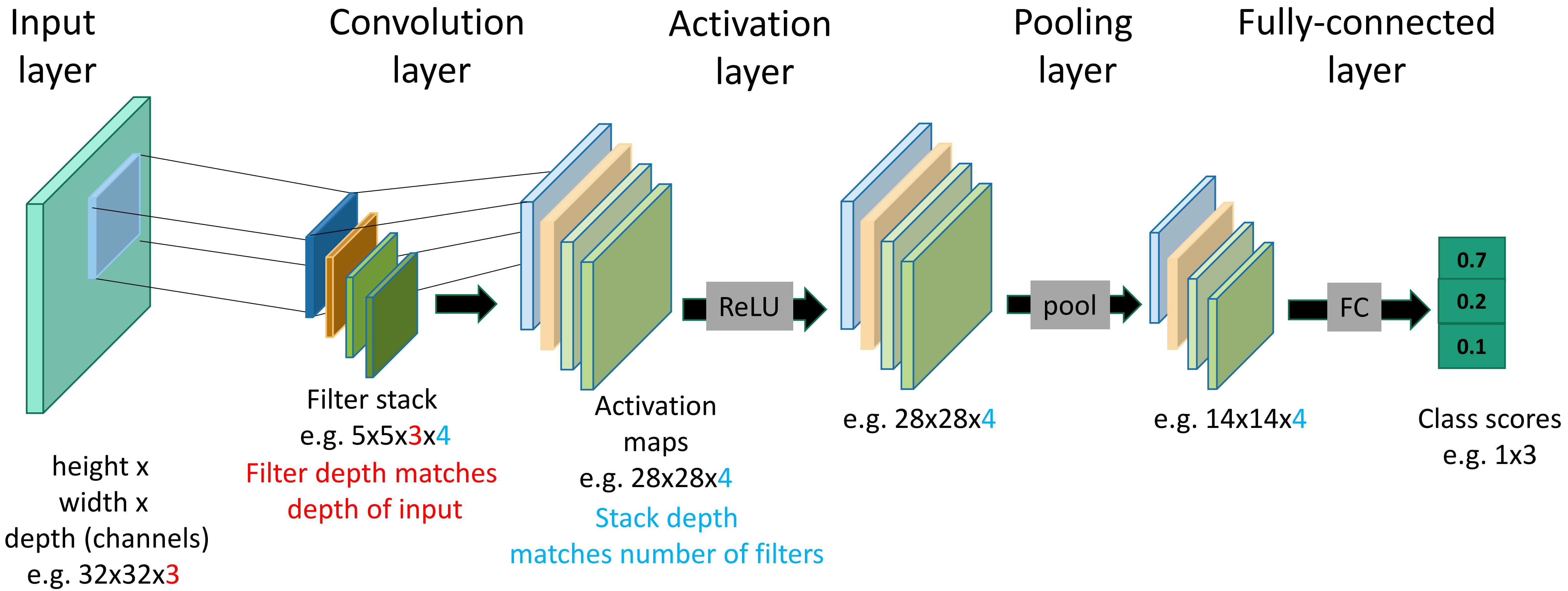
  ↔ Fully connected network

CNN components

- Input layer: holds pixel values of image/video
- Convolutional layer: slides filters over input computing dot product
- Pooling layer: reduces width and height by pooling (downsampling)
- Fully connected layer: every output neuron connected to all input neurons → can recombine all input

- Activation layer: applies activation function (element-wise)
- Normalization layer: normalizes its input
- Dropout layer: randomly sets input units to 0 → prevents overfitting

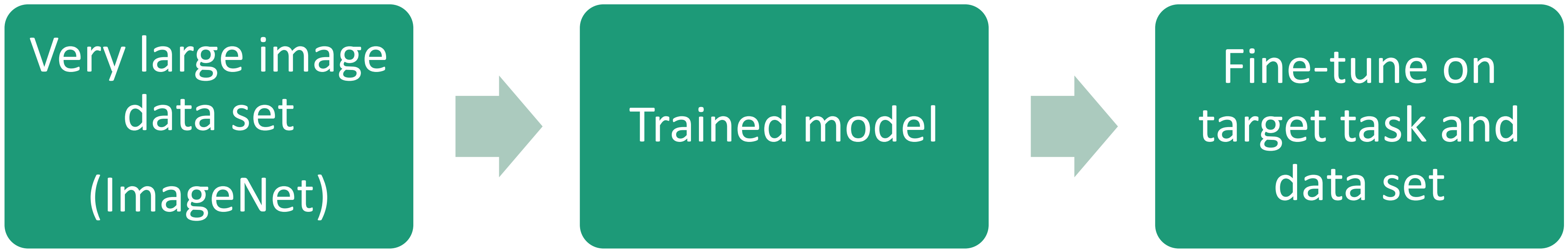
Simple CNN


Input
layer
Convolution
layer
Activation
layer
Pooling
layer
Fully-connected
layer
ReLU
pool
FC
0.7
0.2
0.1
Filter stack
e.g. 5x5x3x4
Filter depth matches
depth of input
Activation
maps
e.g. 28x28x4
Stack depth
matches number of filters
e.g. 28x28x4
e.g. 14x14x4
Class scores
e.g. 1x3
height x
width x
depth (channels)
e.g. 32x32x3

Vision transformers (ViTs)

- Divide images into patches which are encoded in a linear vector together with positional information
- Attention mechanism to weigh importance of each patch in relation to every other patch
- Can "look across the whole image" → relationships between distant parts
- Attention maps can increase explainability
- Can be combined with CNN features into hybrid architectures

Training computer vision models often uses transfer learning



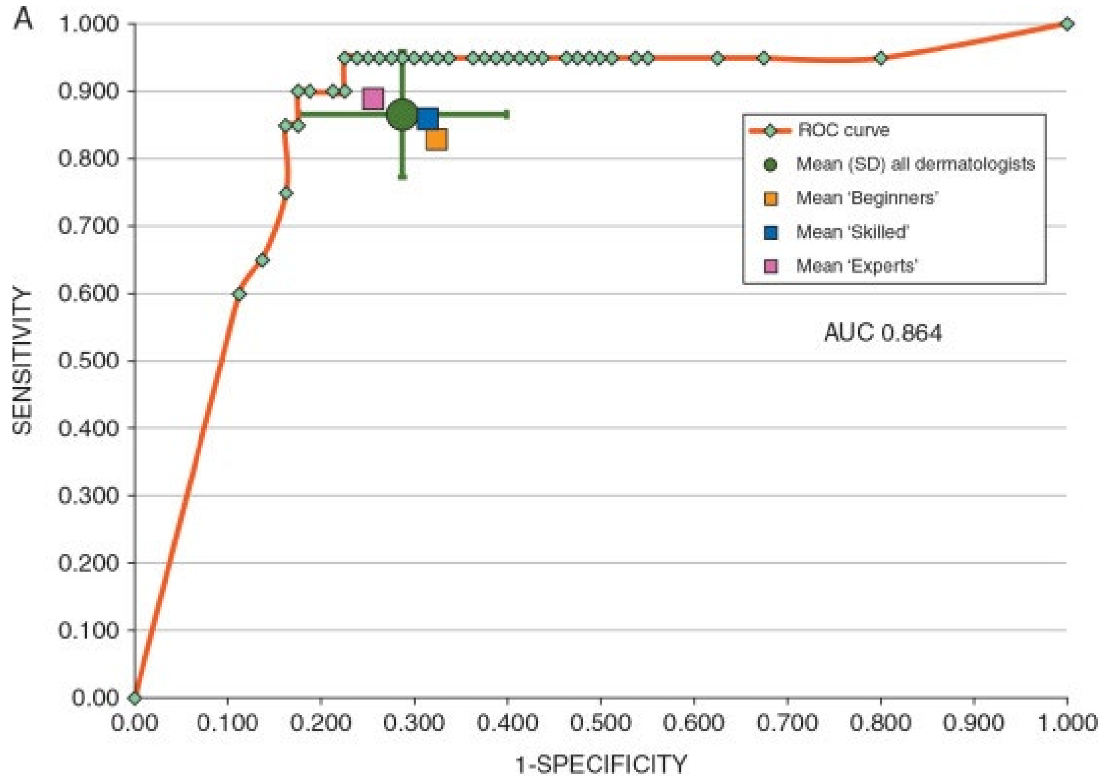
Data augmentation for images

- Flip
- Rotate
- Zoom in
- Add noise
- Split images into tiles
- Illumination variation
- Hue variation
- Blur
- Shear

AI surpasses human performance in many individual studies

Observational Study > Ann Oncol. 2018 Aug 1;29(8):1836-1842. doi: 10.1093/annonc/mdy166.

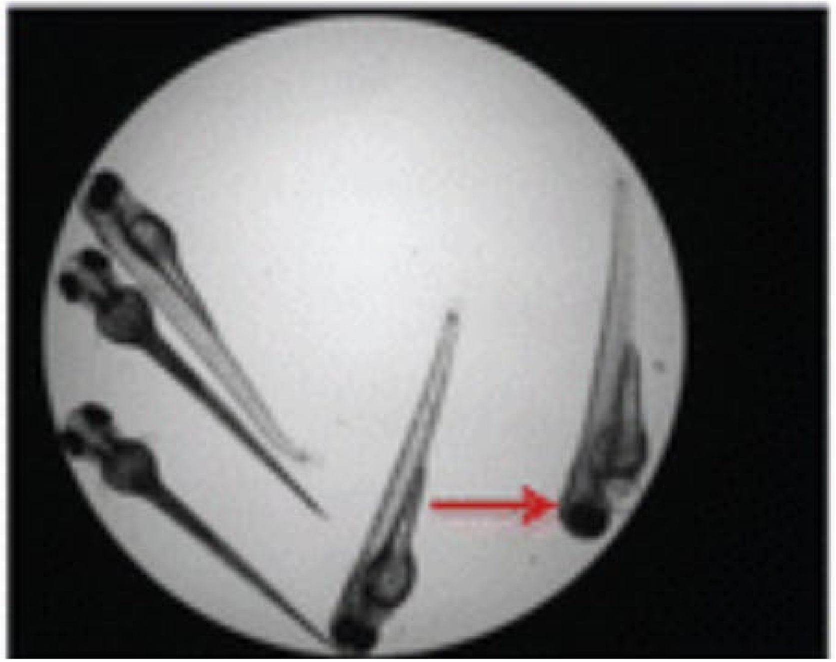
Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists



Neural networks “see” things humans cannot perceive…

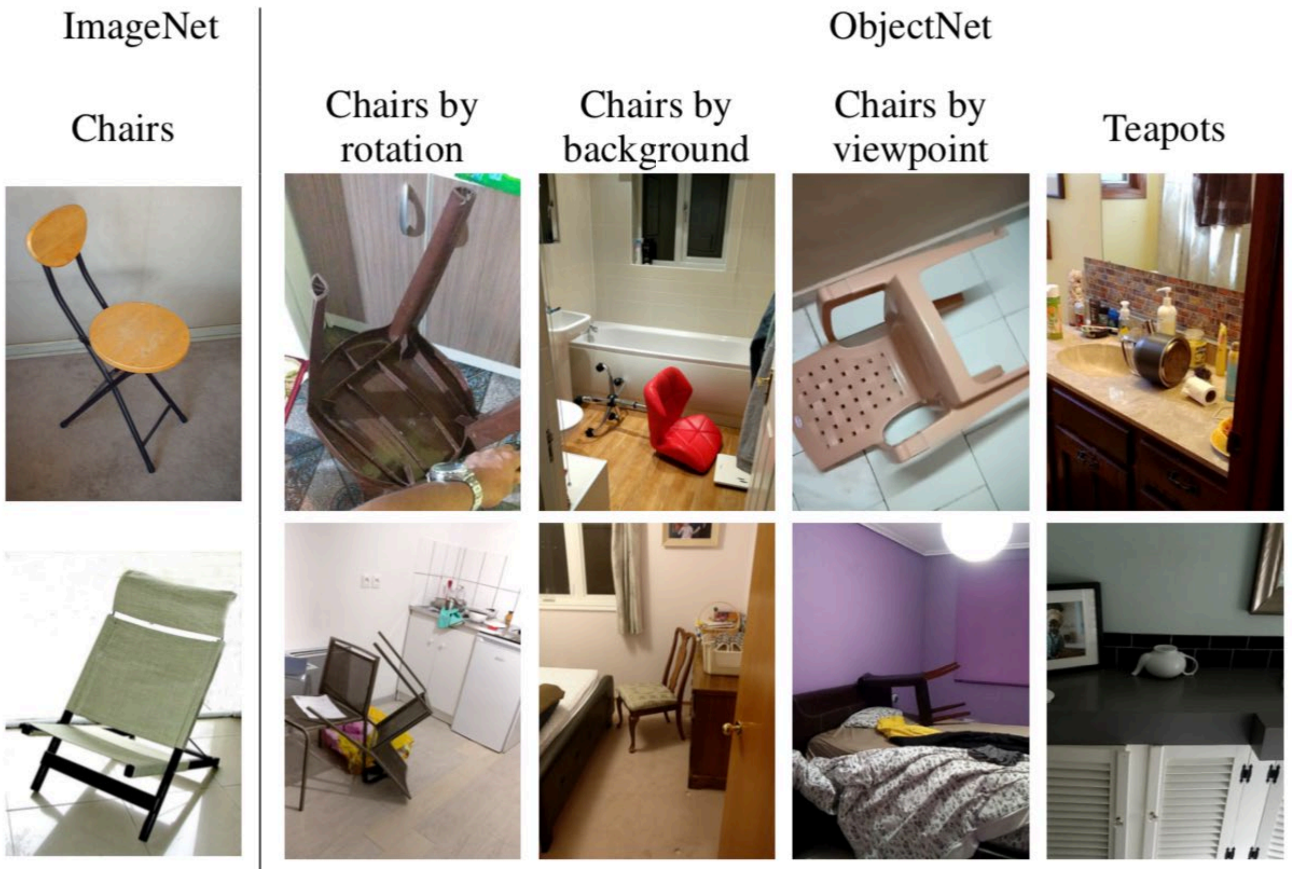
Unhealthy zebrafish

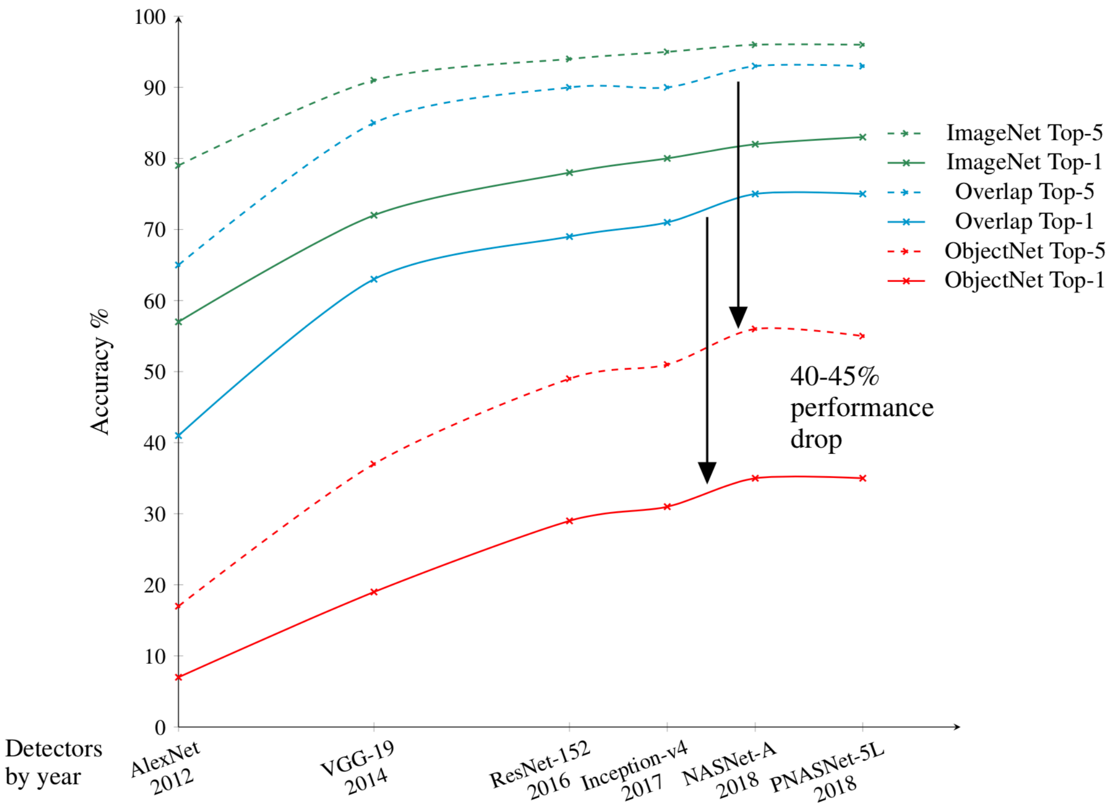
Healthy zebrafish



Ishaq, Sadanandan, Wählby. 2017 22(1):102-107. doi: 10.1177/1087057116667894

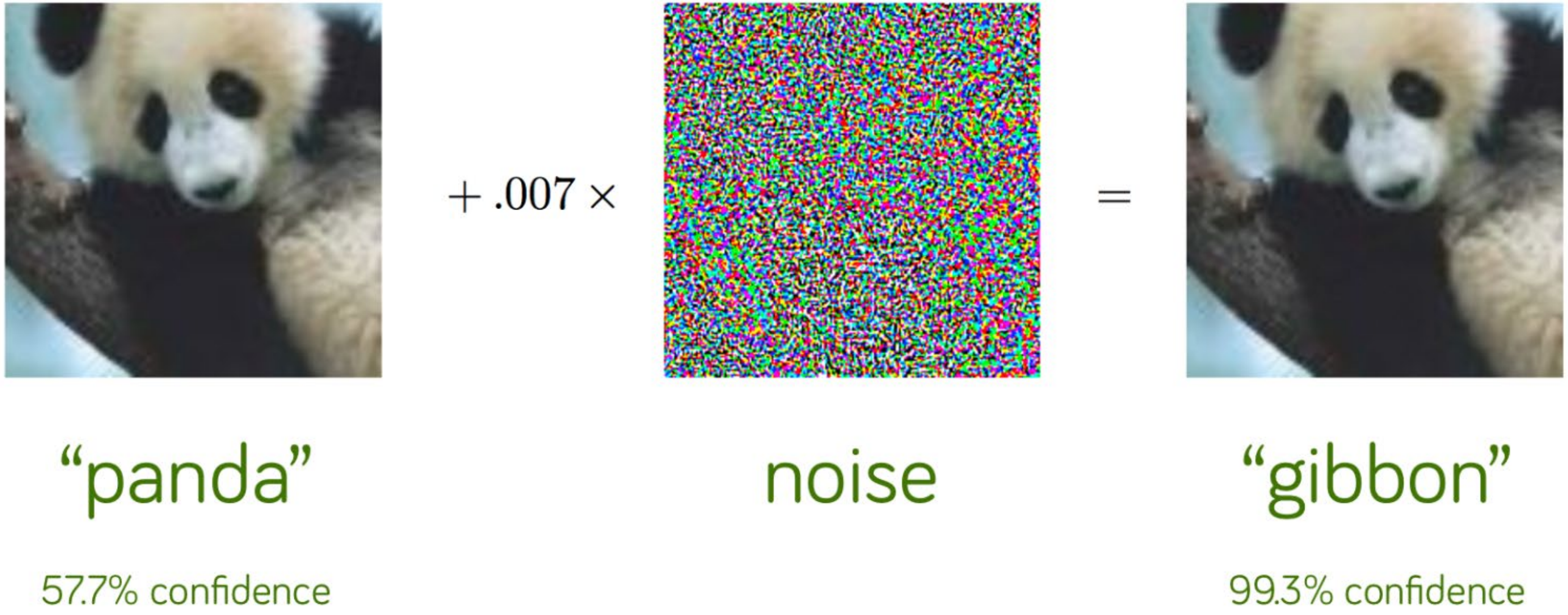
...but can struggle with problems that are easy for humans





Barbu et al. Advances in Neural Information Processing Systems 32, pages 9448–9458. 2019.

…but can struggle with problems that are easy for humans



Goodfellow et al, ICLR 2015

From Karen Zack
@teenybiscuit

Computer vision is challenging

- Variable viewpoints
- Noise
- “Accidental” similarities
- Variable scale
- Deformable and/or moving objects
- Occlusion
- Variable illumination
- Intra-class variation
- Variable background

AI-based software for bioimage analysis

Comm

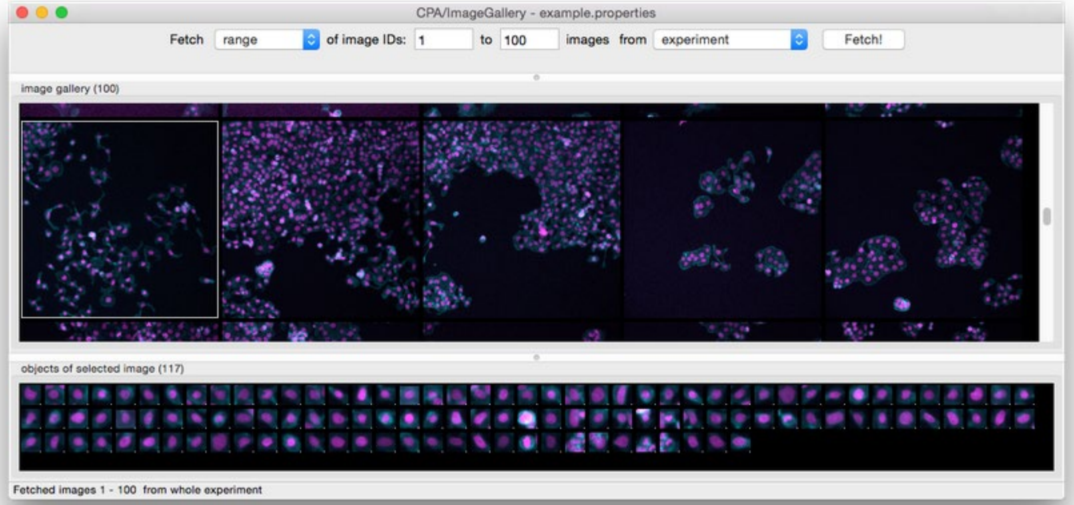
DeepImageJ: A user-friendly plugin to run deep learning models in ImageJ

Estibaliz Gómez-de-Mariscal, Carlos García-López-de-Haro, Laurène Donati, Michael Unser, Arrate Muñoz-Barrutia, Daniel Sage

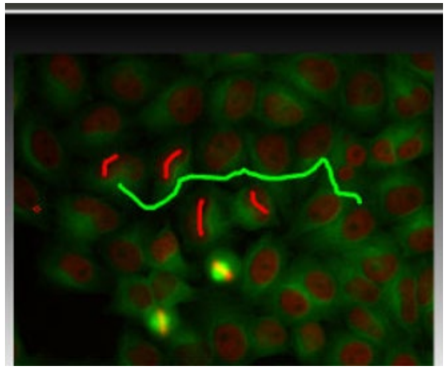
doi: https://doi.org/10.1101/799270

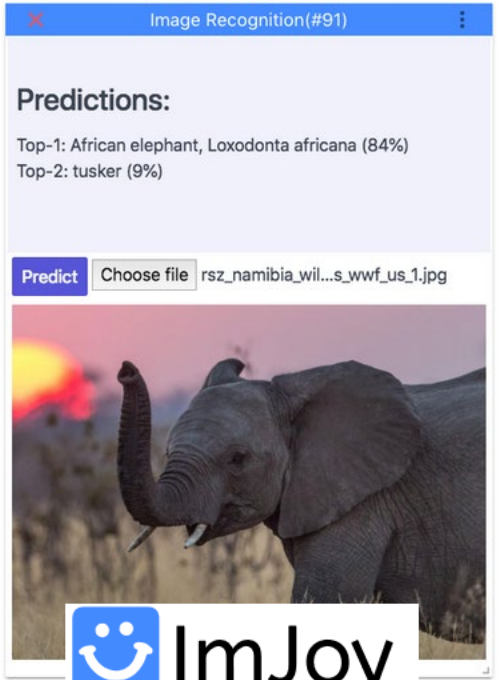
CellProfiler Analyst

Interactive data exploration, analysis, and classification of large biological image sets



ilastik





ImJoy

Summary: common computer vision tasks

- Classification
- Regression
- Object detection
- Object tracking
- Semantic segmentation
- Instance segmentation
- Pose estimation
- Image registration
- Image restoration
- Super-resolution
- Image captioning
- Image clustering
- Dimensionality reduction
- Image generation

Summary

- Convolutional layers are the defining characteristic of many computer vision models (“convolutional neural networks”)
- Vision transformers for tasks that require the whole image to be considered $\leftrightarrow$ CNNs for fine details (hybrid architectures exist)
- Transfer learning, e.g. with networks pre-trained on ImageNet, is default choice