Tuesday, 31th of August 2021

Dealing with image analysis?
Let the computer do the heavy lifting

IMAS and AAPP Machine Learning Workshop

Presentation #5, Charley Gros

UNIVERSITY of TASMANIA

IMAS

Institute for Marine and Antarctic Studies

NeuroPoly

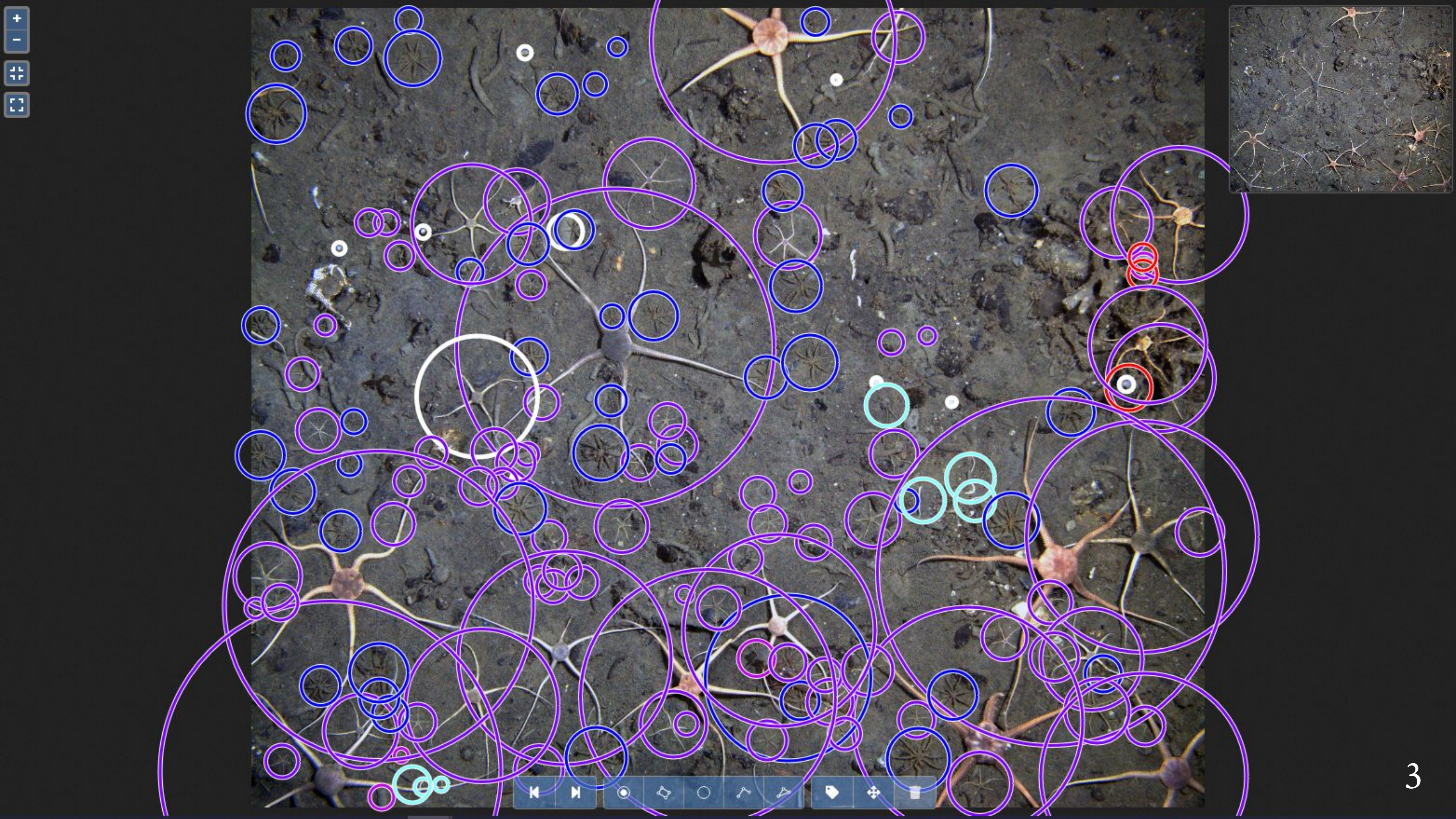
Mila

UNIVERSITY of TASMANIA

AAPP Australian Antarctic Program Partnership

I acknowledge the palawa people as the Traditional Owners and ongoing custodians of lutruwita (Tasmania)

and pay my respects to their Elders past, present and future.

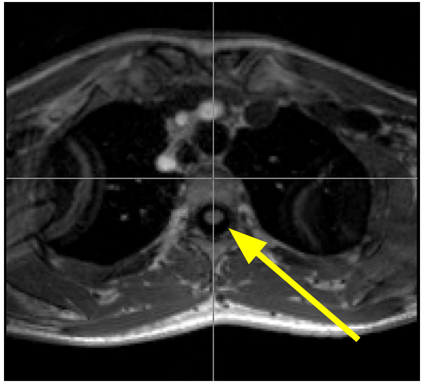


dazeinfo.com/2019/11/07/artificial-intelligence-modern-healthcare-practices/

When I first met CNNs...

Project with Dominique:

- 80 MRI scans of patients with Multiple Sclerosis, ~ 500 axial slices per patient
- Question: Is there a correlation between the spatial distribution of lesions in the spinal cord with patients' clinical score?



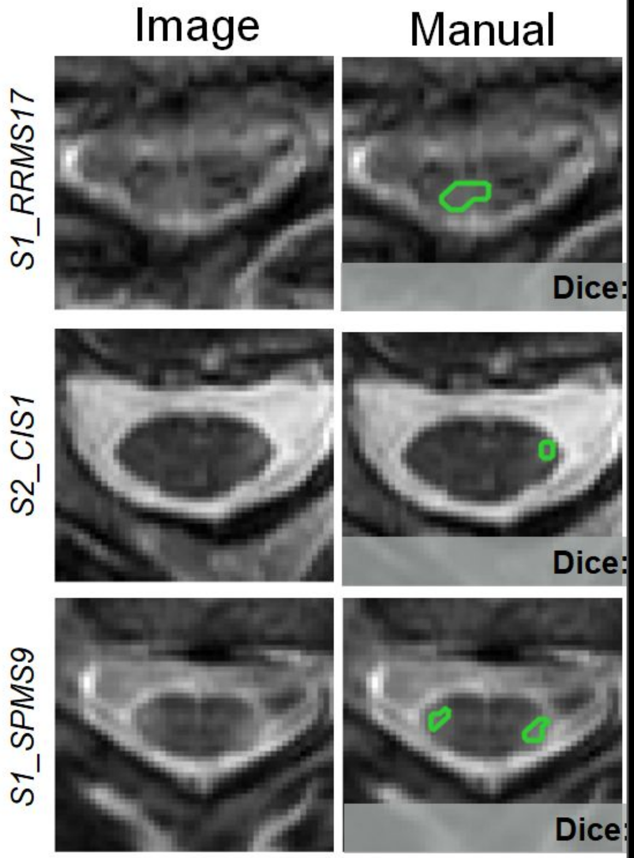
When I first met CNNs...

Manual delineation of lesions on 20 patients…

… used to train a Deep Learning algorithm to automatically segment lesions on the 60 remaining patients

… keys challenges include:

- Class imbalance
- Intensity / shape / size heterogeneity
- Intra- and inter-expert disagreement



Rectangle hiding a surprise

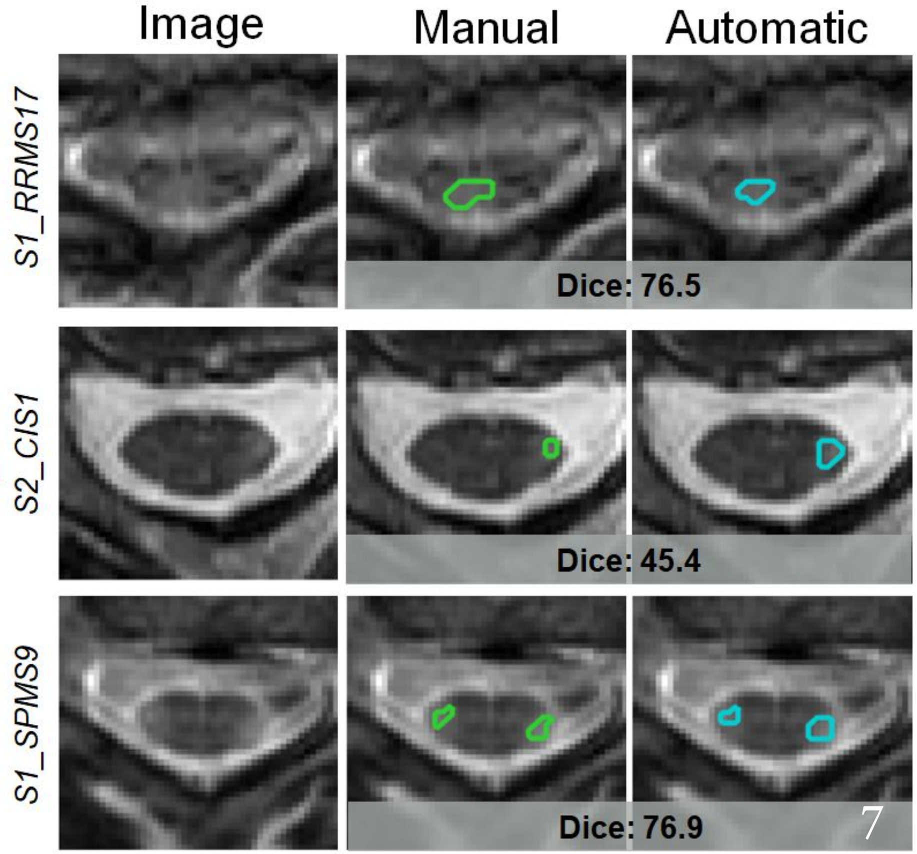
When I first met CNNs...

Manual delineation of lesions on 20 patients…

… used to train a Deep Learning algorithm to automatically segment lesions on the 60 remaining patients

… keys challenges include:

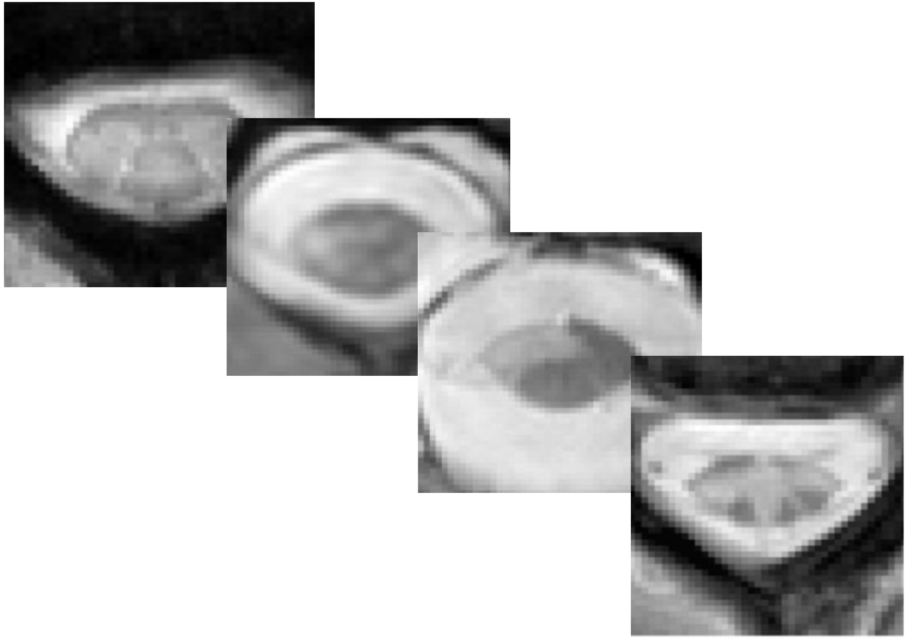
- Class imbalance
- Intensity / shape / size heterogeneity
- Intra- and inter-expert disagreement



When I first met CNNs...

- From 80 patients to… 672 patients, from 13 different clinical around the world.

- Re-train a segmentation model to account for the dataset heterogeneity

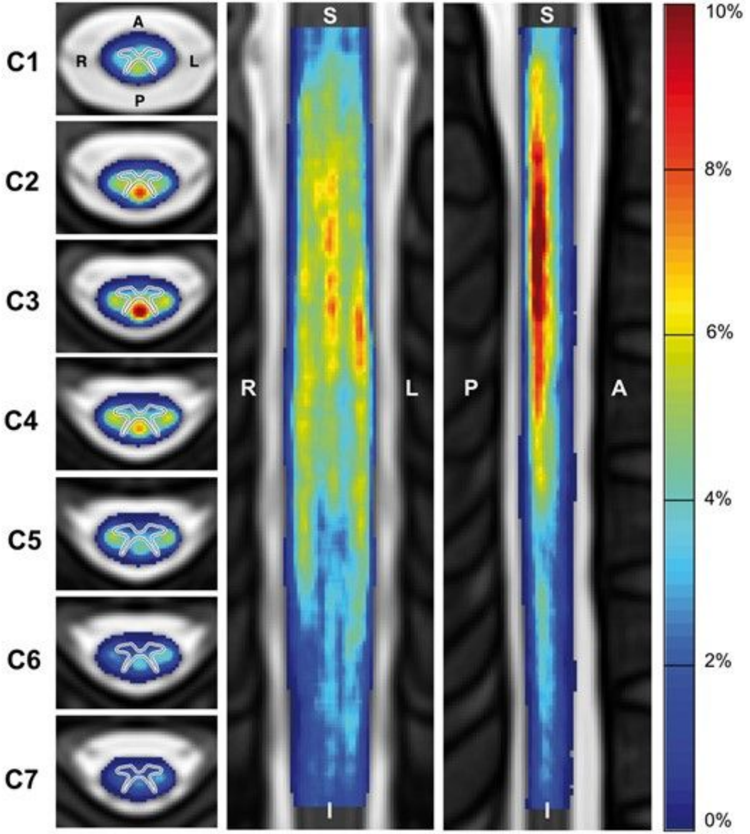


NeuroPoly ZUCKERBERG SAN FRANCISCO GENERAL Hospital and Trauma Center neuro XWHCMU Karolinska Institutet MGH 1811 CHU Rennes CHU

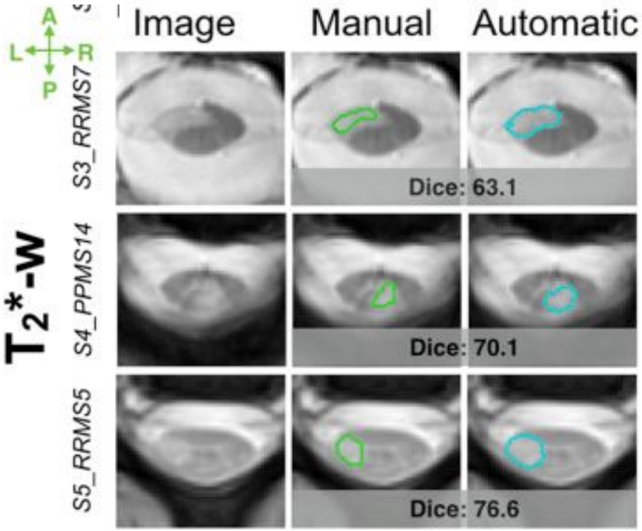
OFSEP OSPEDALE SAN RAFFAELE BWH UCL VANDERBILT UNIVERSITY NIH NYU Langone Health NIRS UNF

When I first met CNNs...

Frequency of multiple sclerosis lesions in the cervical spinal cord for all patients (n = 642)



Eden and Gros et al. 2019
doi: 10.1093/brain/awy352



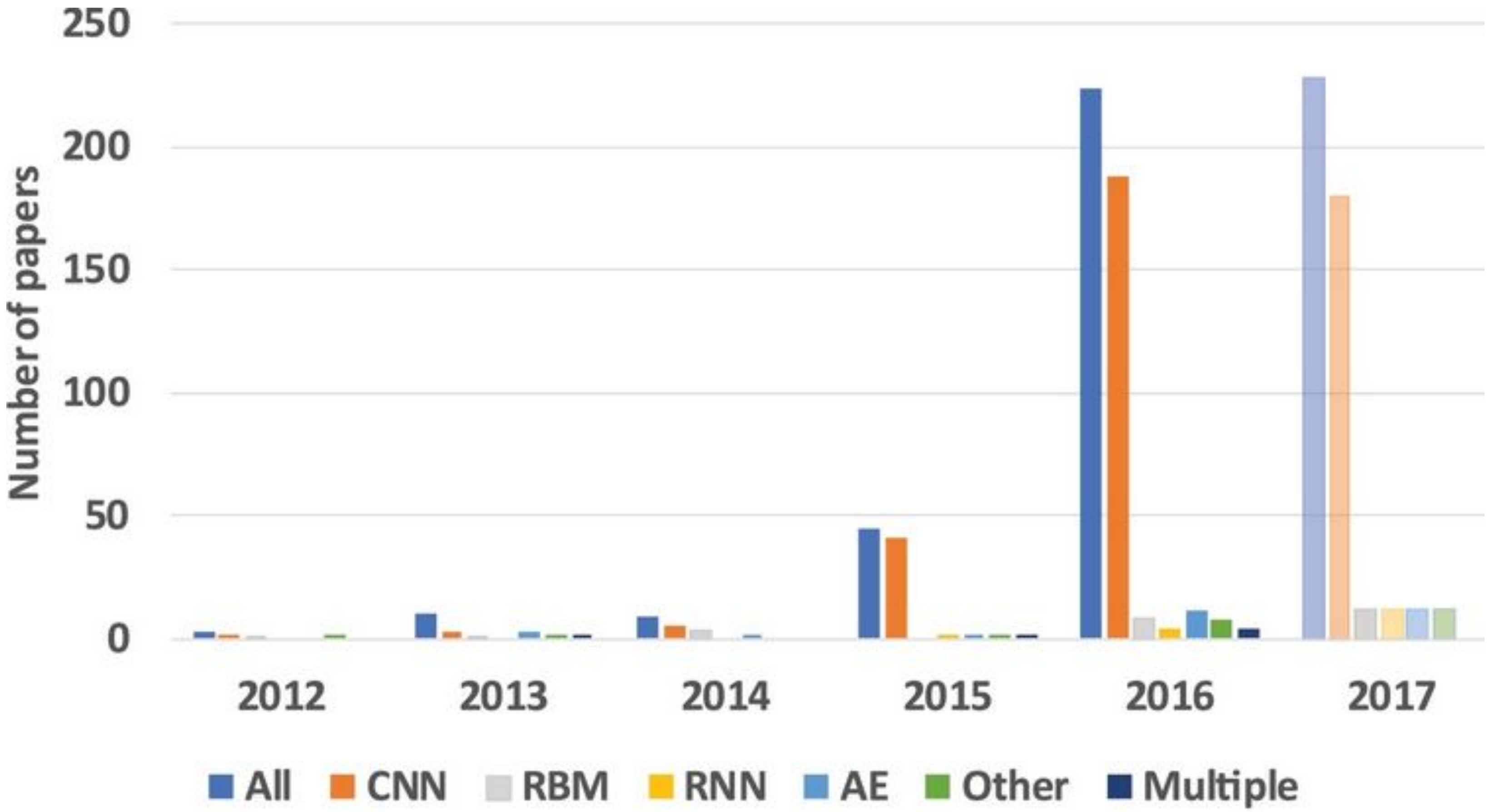
Gros et al. 2019
doi: 10.1016/j.neuroimage.2018.09.081

The automatic tool is now integrated into an open-source software, SCT (250+ citations).

SCT
Spinal Cord Toolbox

github.com/spinalcordtoolbox/spinalcordtoolbox

The 2015 sparkle

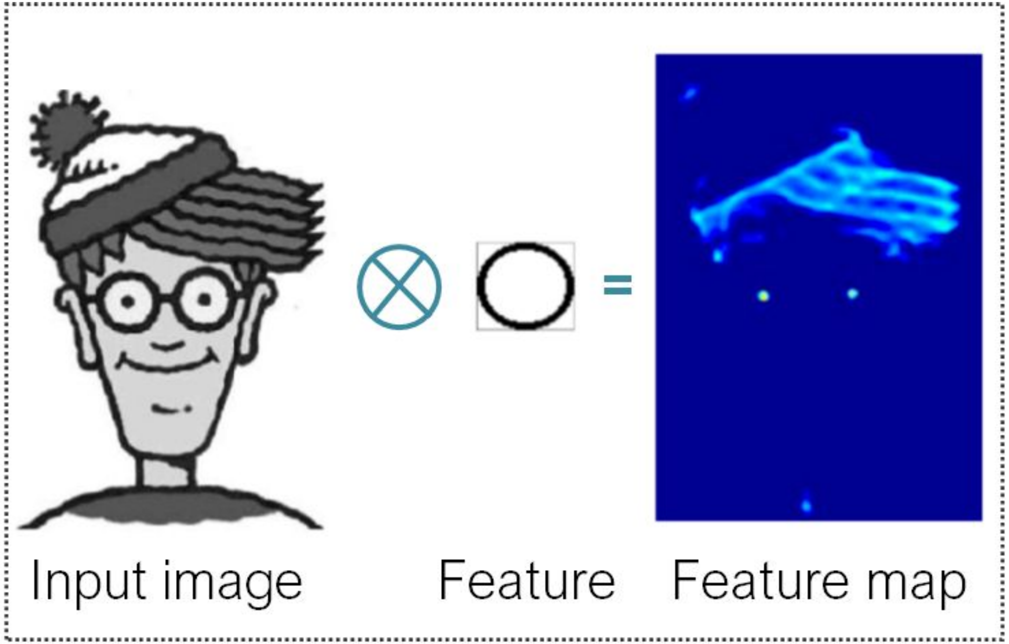


Litjens et al. 2017
doi: 10.1016/j.media.2017.07.005

Why do they work bloody so well?

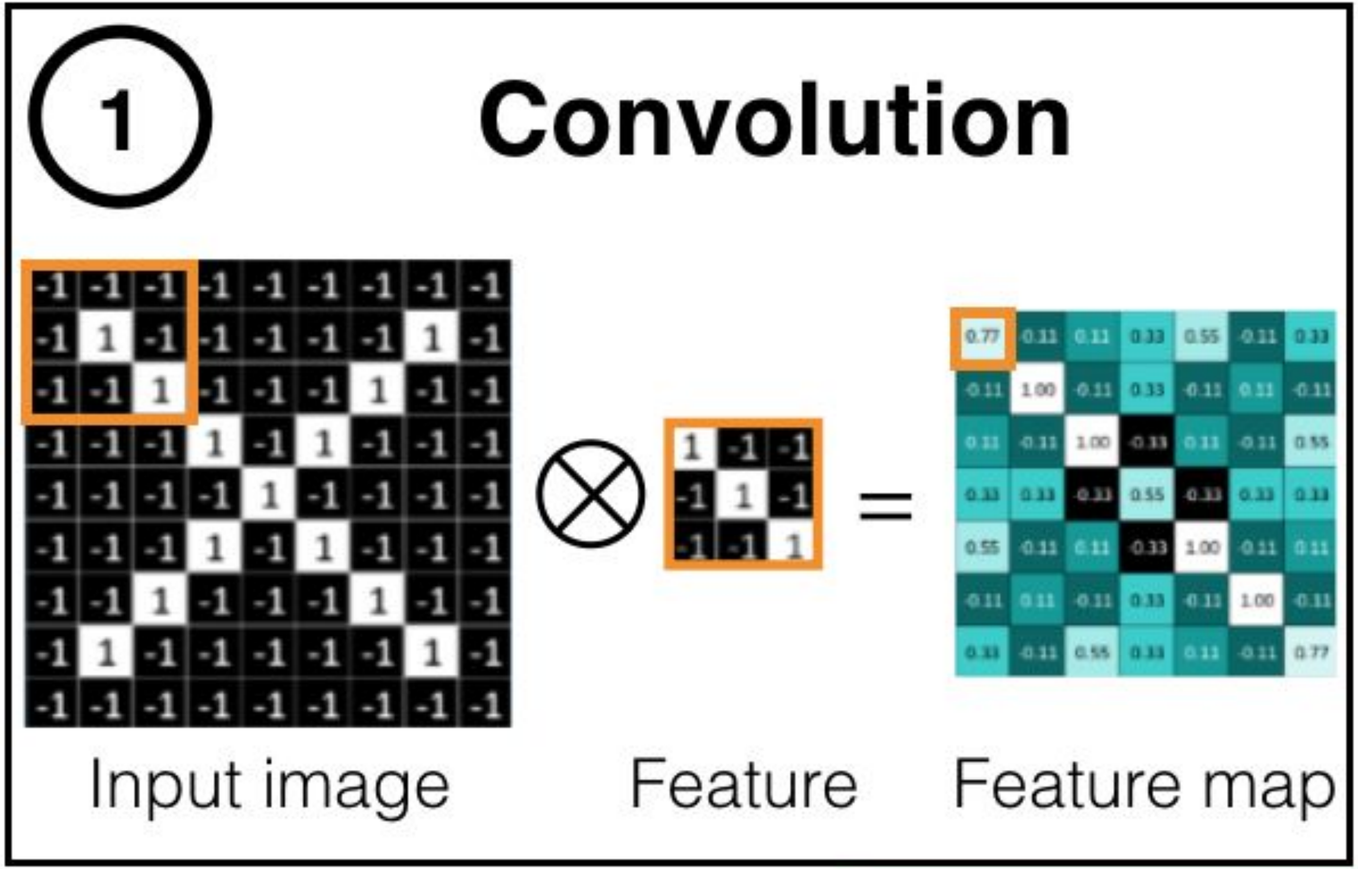
"Ideal" image

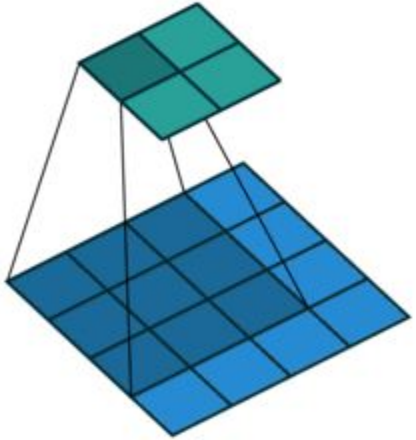
"Real life" image



From "Feature engineering" to "Automatic feature selection" ...

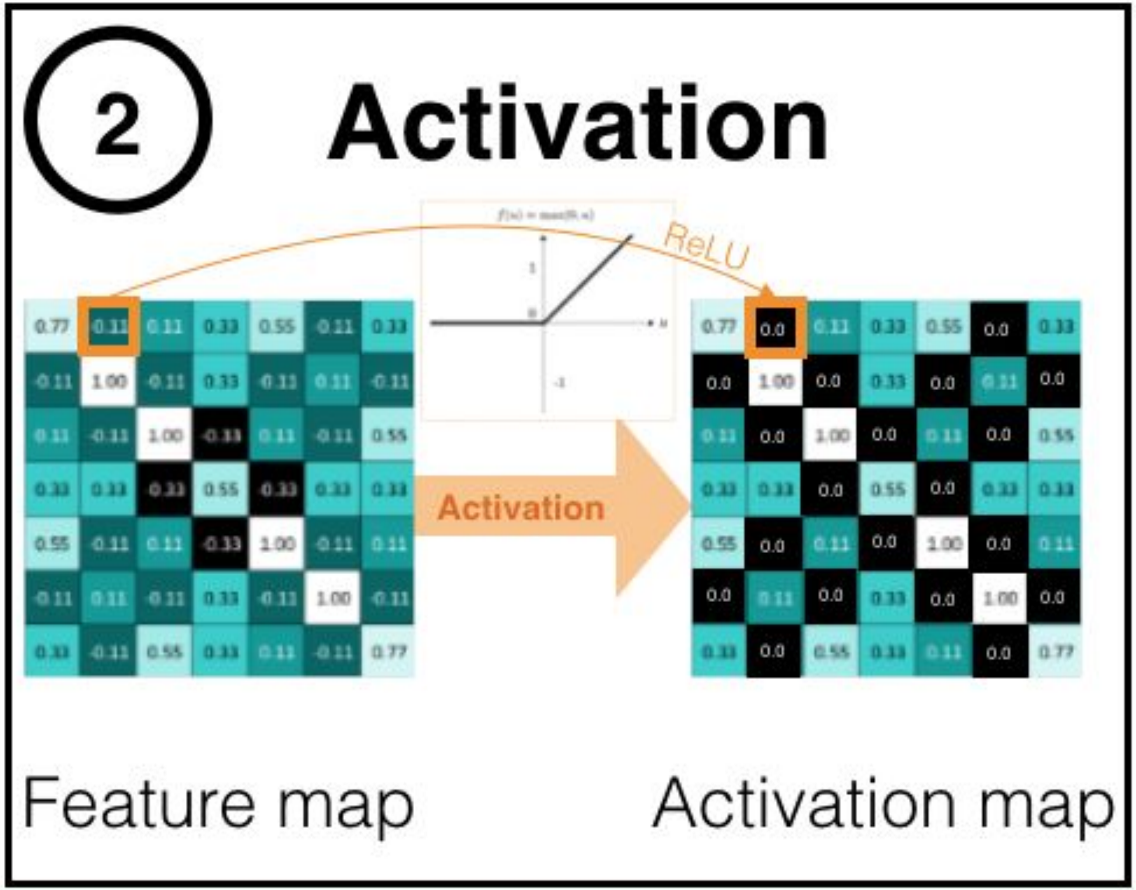
Why do they work bloody so well?



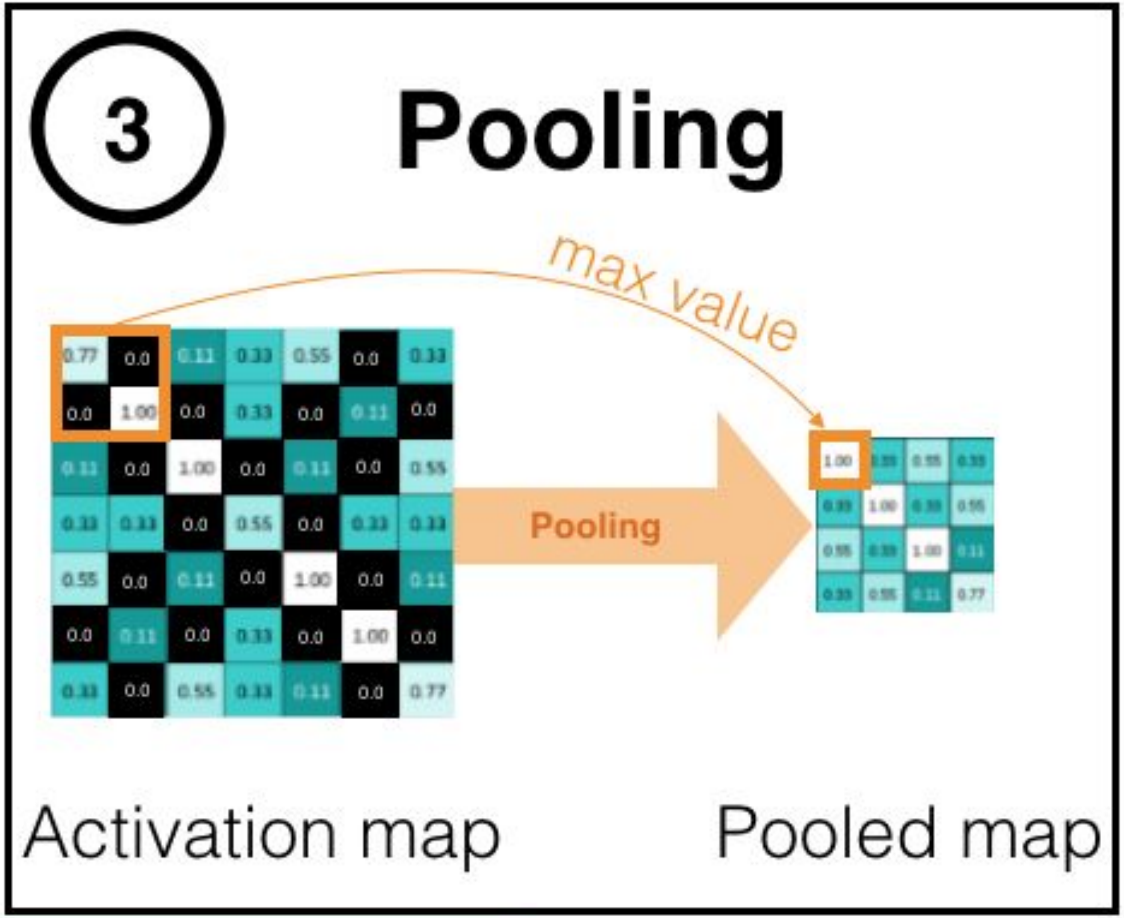


GIF from
github.com/vdumoulin/conv_arithmetic

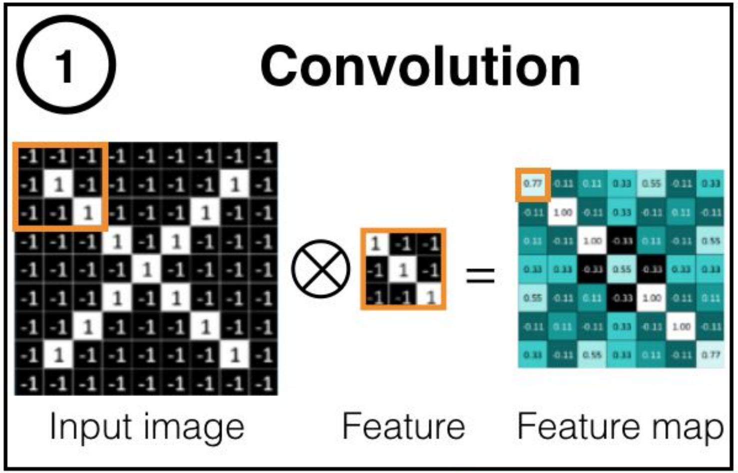
Why do they work bloody so well?

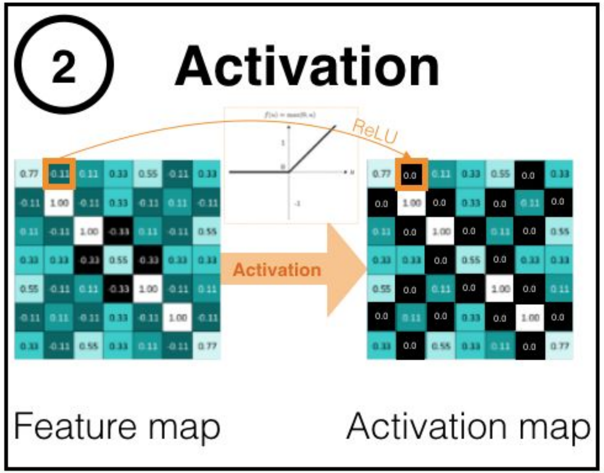


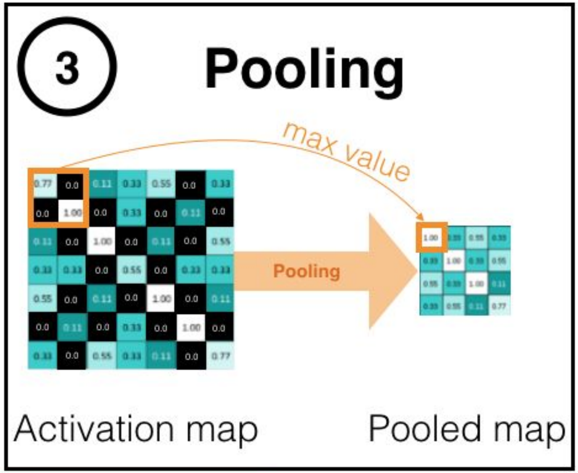
Why do they work bloody so well?

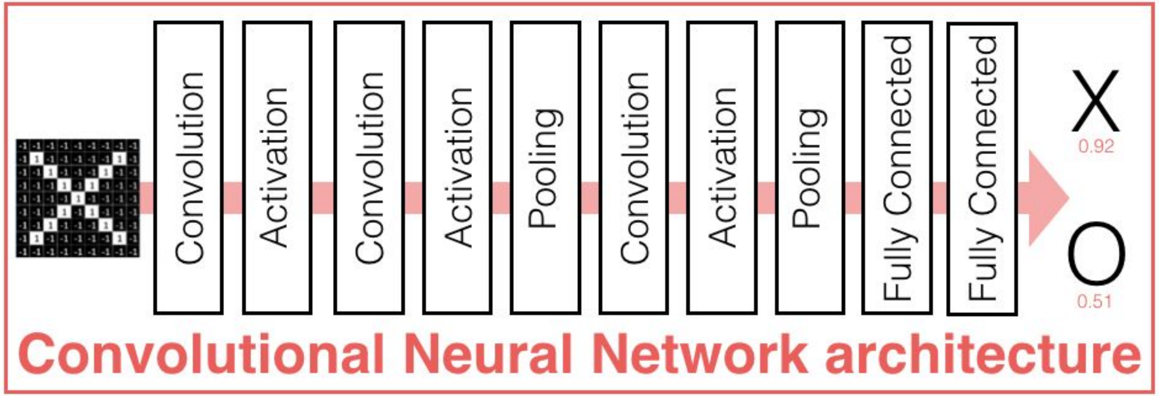


Why do they work bloody so well?

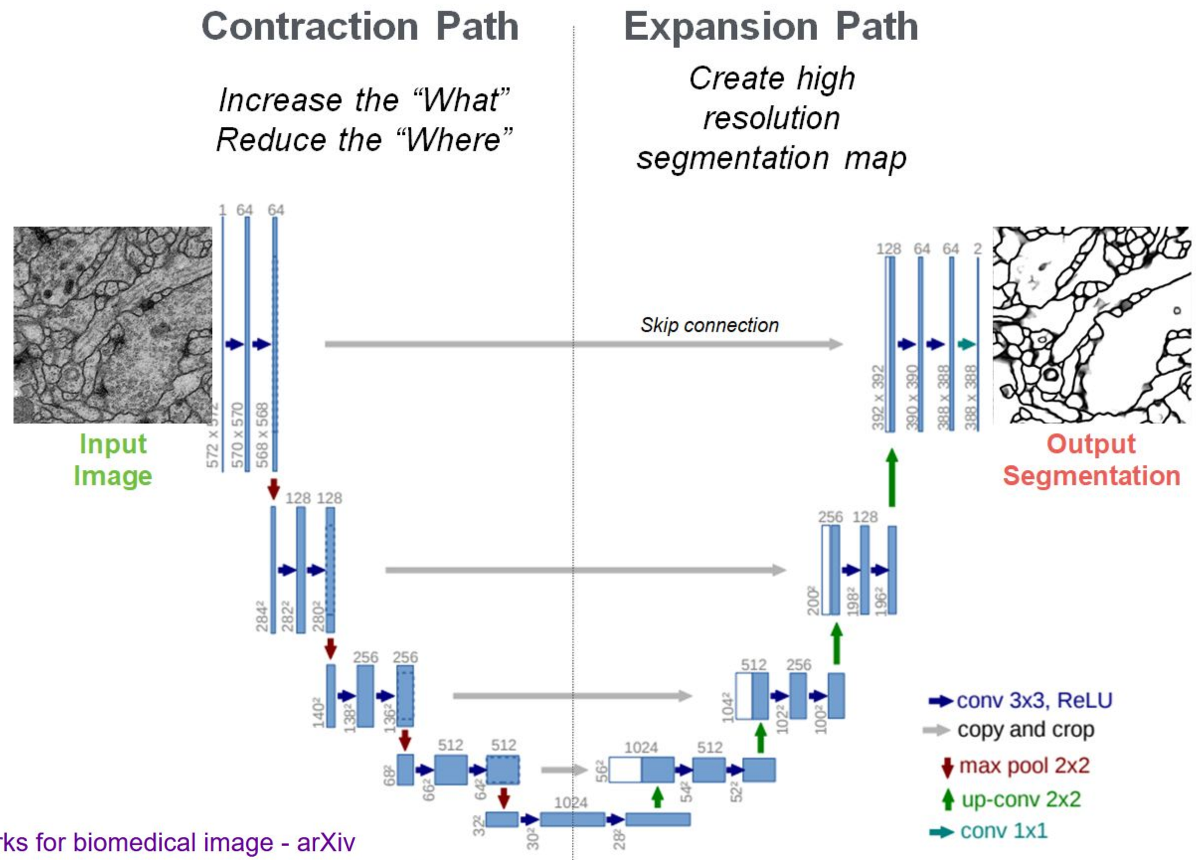








One famous critter: the U-Net



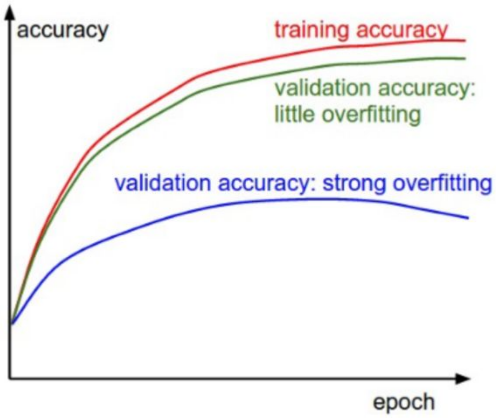
U-Net: Convolutional networks for biomedical image - arXiv

by O Ronneberger · 2015 · Cited by 30073 — In this **paper**, we present a network and training strategy that relies on the strong use of data augmentation to use the available annotated ...

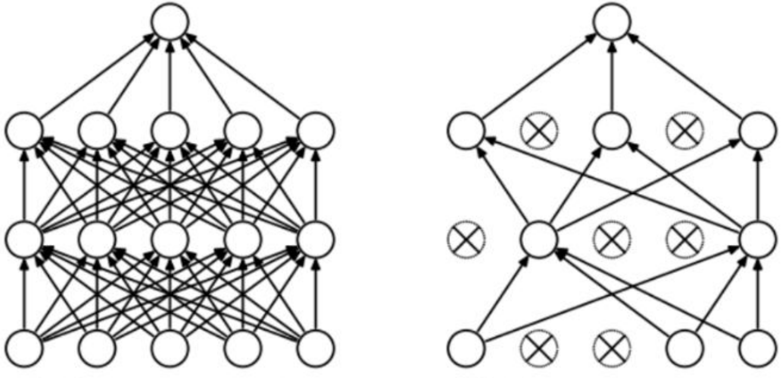
How to improve model's generalisation?

How to avoid overfitting:

- Reduce model complexity
- Data augmentation
- Dropout
- Early stopping
- Transfer learning



From cs231n.github.io/neural-networks-3/



(a) Standard Neural Net (b) After applying dropout.

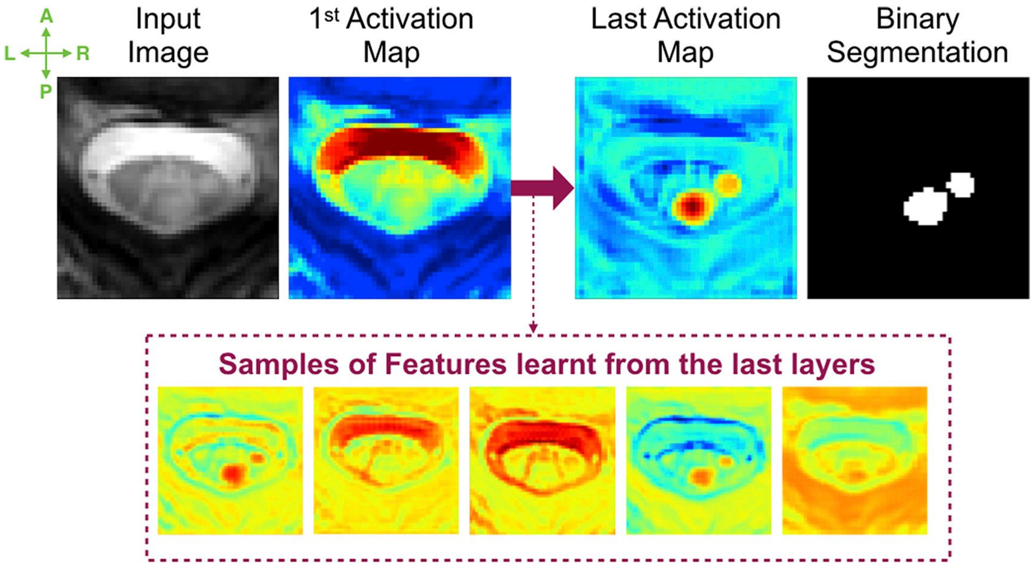
Srivastava et al. 2014

doi: 10.5555/2627435.2670313

github.com/aleju/imgaug

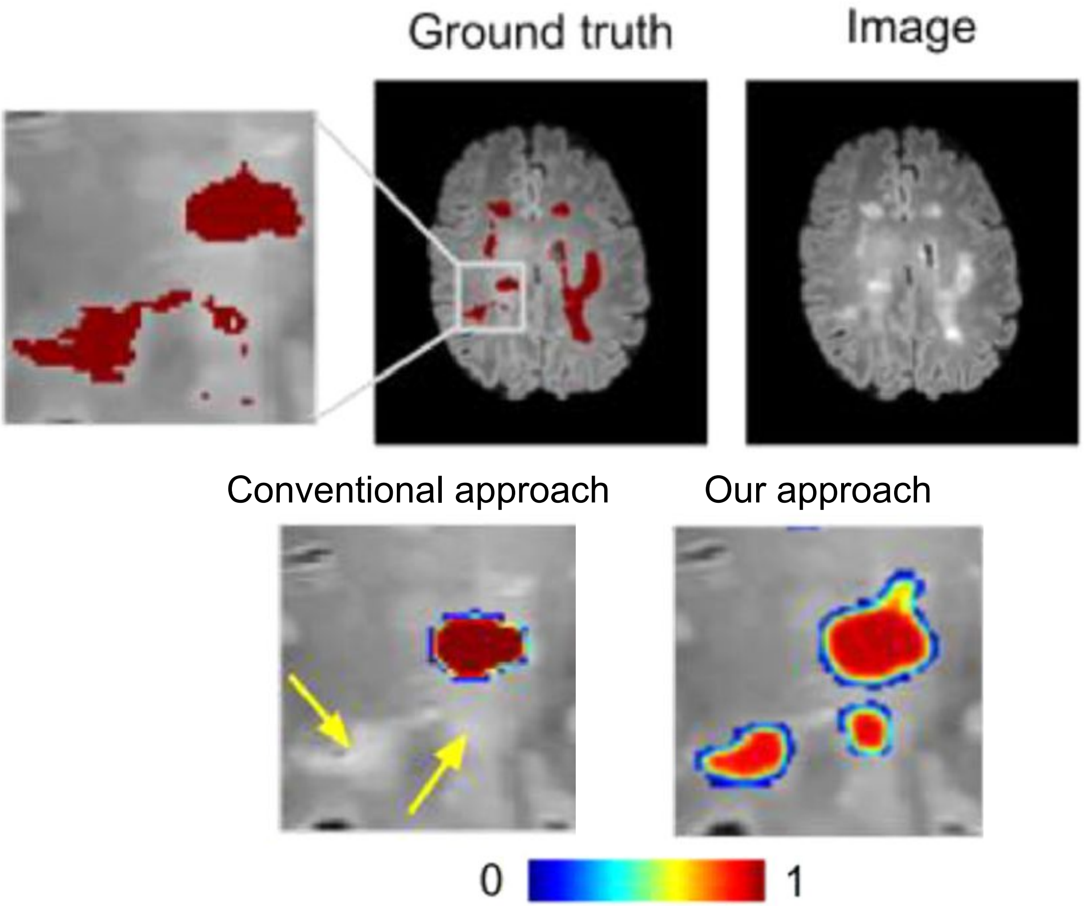
Under the hood

Zagoruyko and Komodakis 2016 ; arXiv: 1612.03928



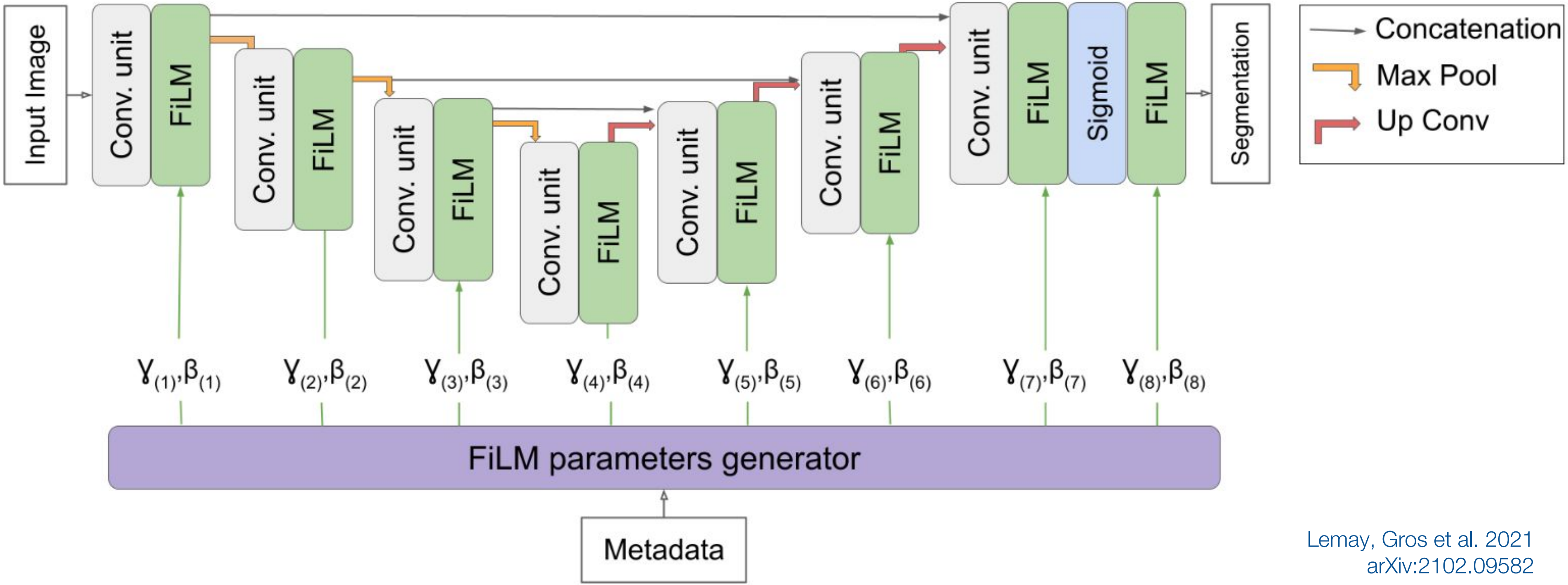
Gros et al. 2019 ; doi: 10.1016/j.neuroimage.2018.09.081

Uncertainty matters



Gros and Lemay et al. 2021
doi: 10.1016/j.media.2021.102038

Conditioning layers



Lemay, Gros et al. 2021
arXiv:2102.09582

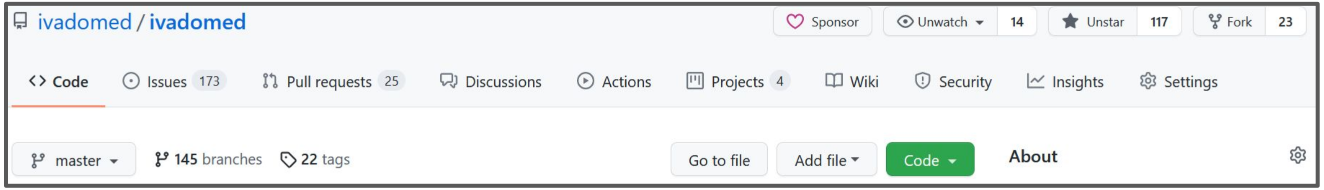
Why now is the time to jump in?

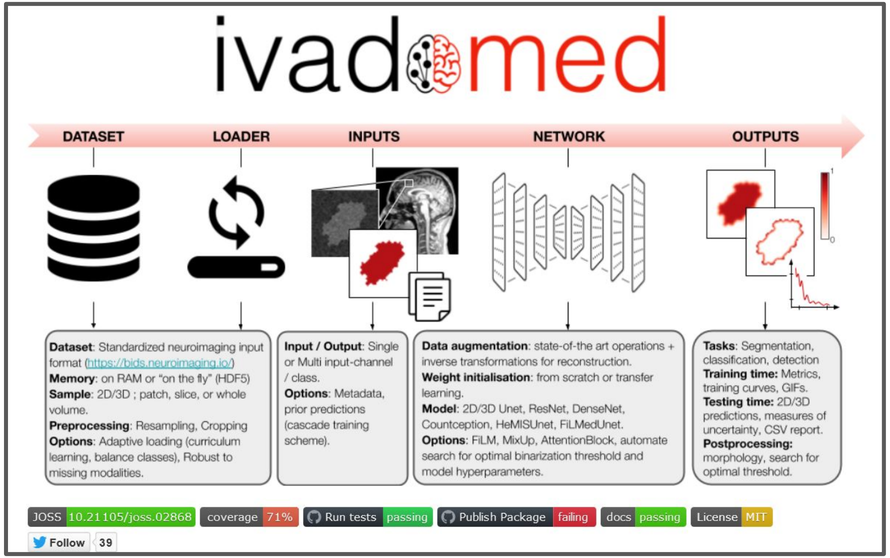
1. Quantity of data collected and shared

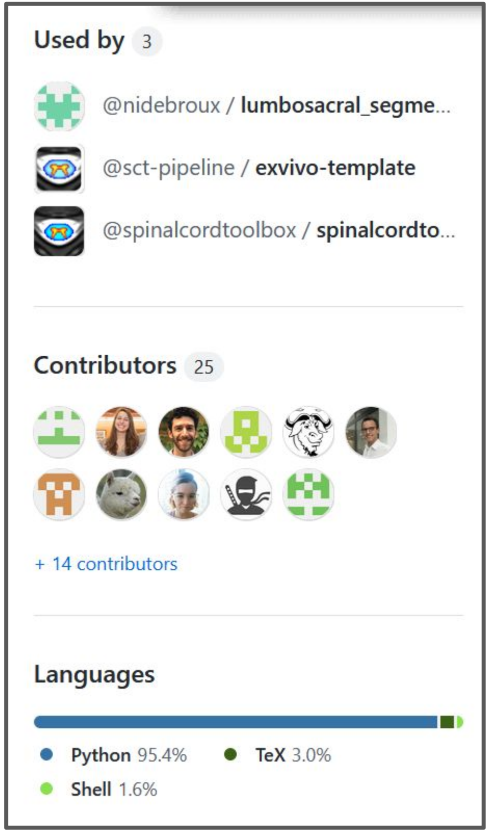
2. Computer resources available

3. Large and dynamic community, open-source culture

From your local code to an open-source library





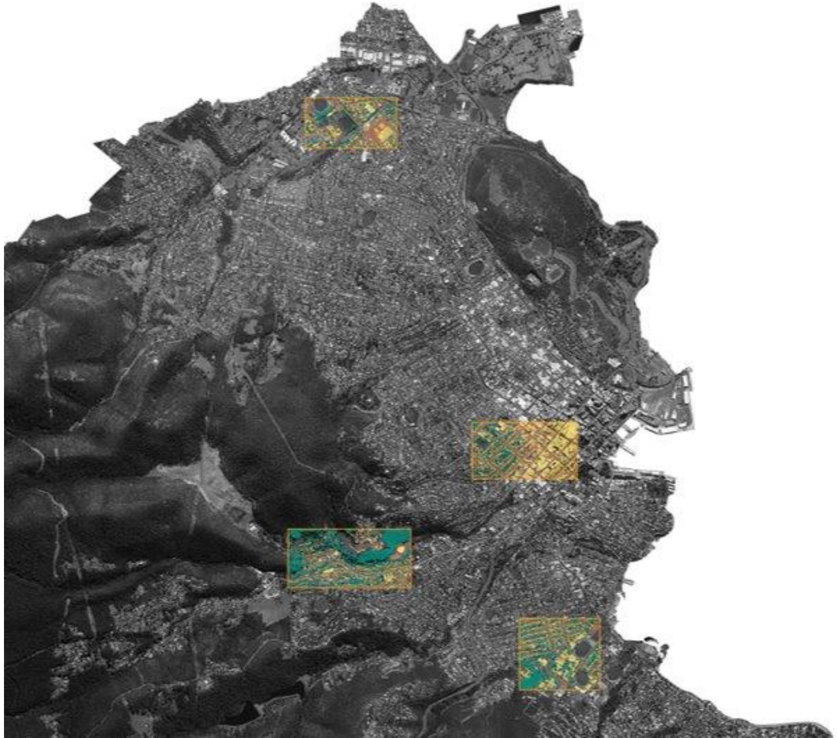


Documentation: ivadomed.org/en/latest/

Gros, Lemay et al. 2021
doi: 10.21105/joss.02868

Project with Bradley

Research question: *Differences in spatial distribution of human infrastructures vs trees, across Hobart's suburbs.*



Objective: fine-scale segmentation of Hobart area, ~80 km^2.

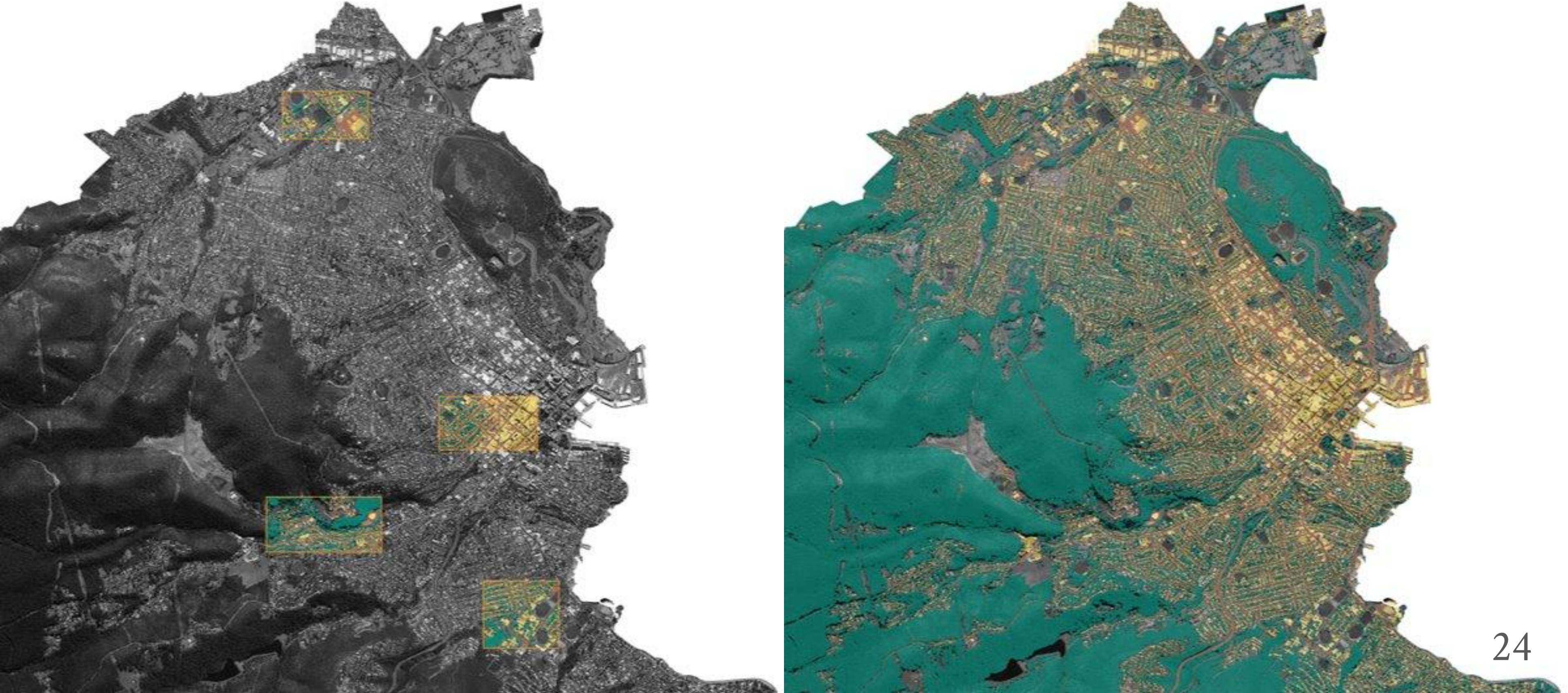
Target: trees vs. buildings vs. roads.

Input data: satellite imagery, 50 cm resolution.

Manual labelling: ~8% of the total area.

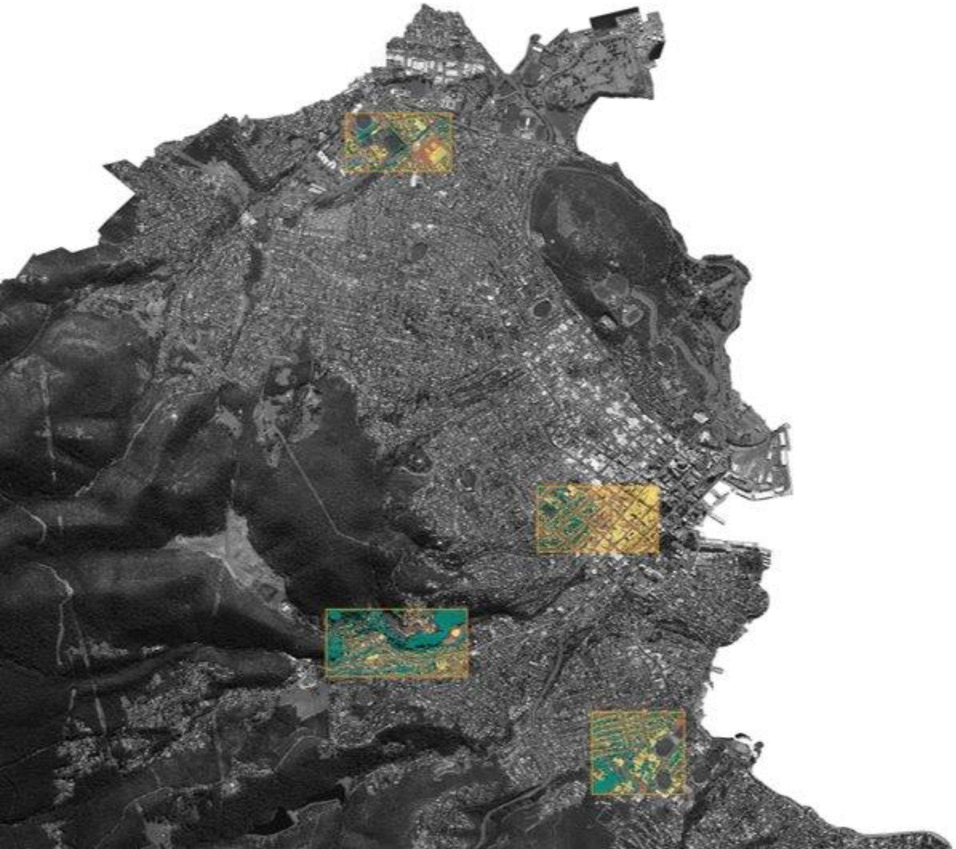
Method: U-Net, Dilated convolutions, DropOut, Batch normalisation, Dice Loss.

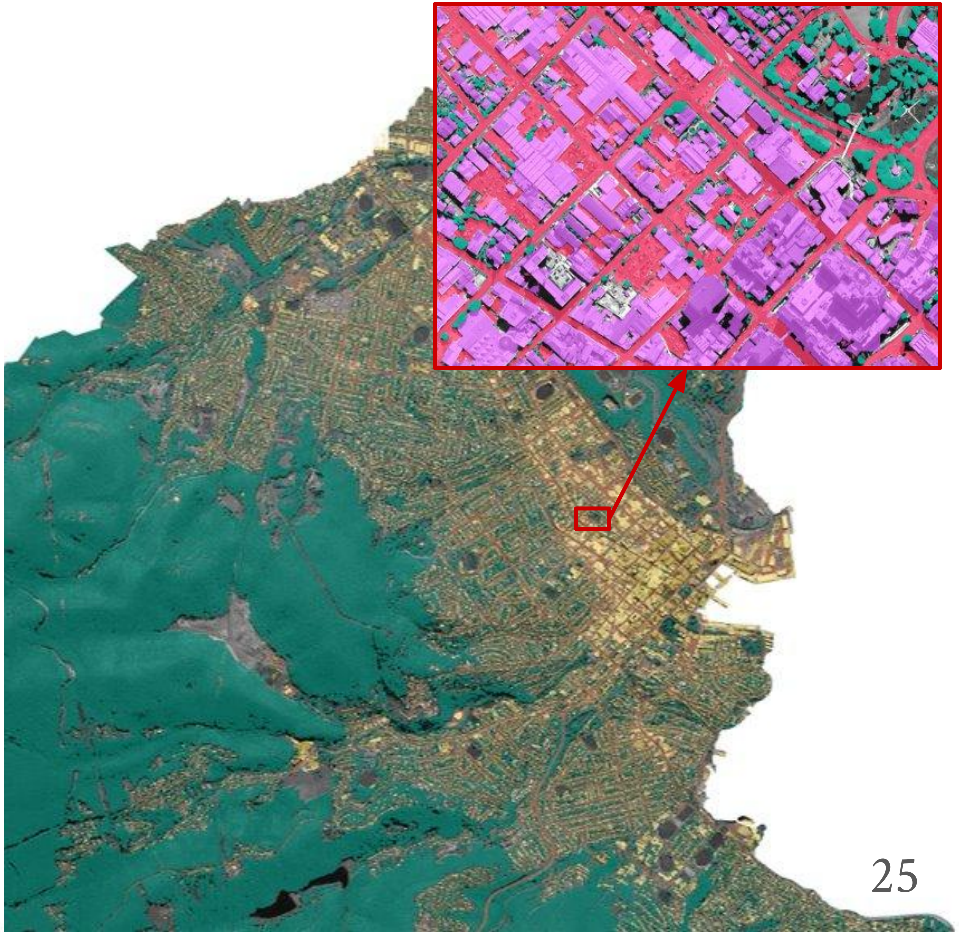
Project with Bradley



Project with Bradley

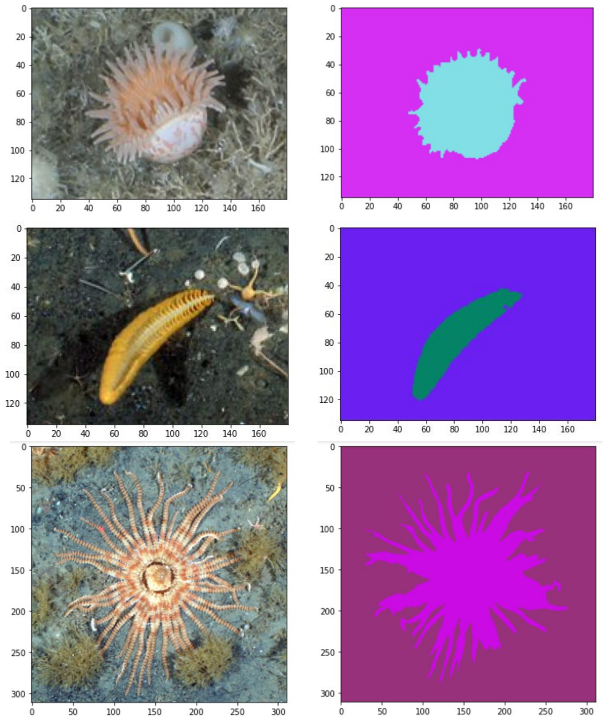
Geoneon





Project with Victor

Research question: *understand variability in organism size due to environmental factor or food availability.*



Objective: size estimation of organisms, from 90k+ annotated organisms.

Target: organism vs. background.

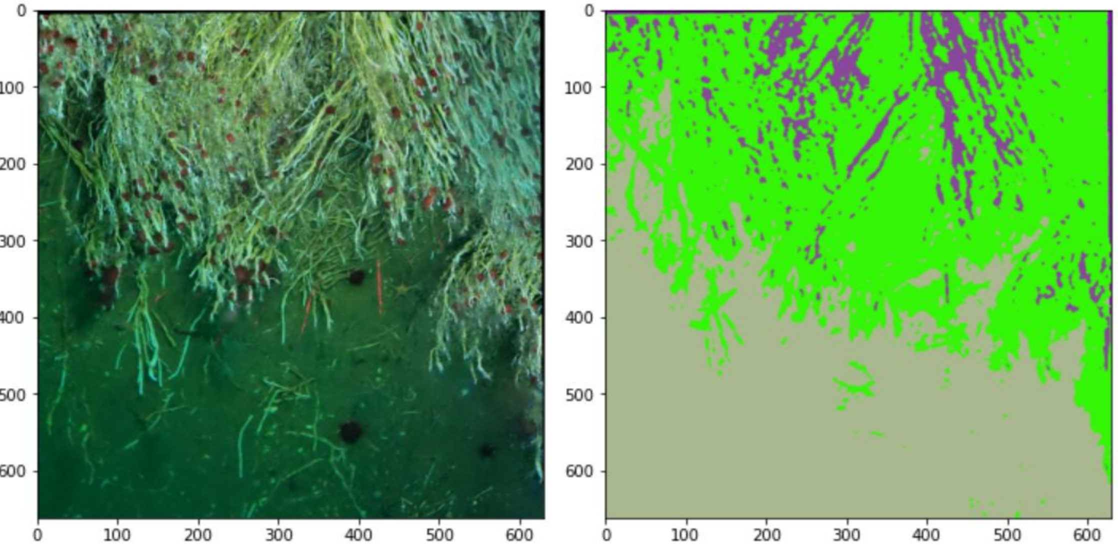
Input data: underwater imagery, heterogeneous acquisition method.

Method:

1. Unsupervised segmentation with spatial constraint
2. Review and cherry picking
3. Supervised training
4. Review and cherry picking + manual segmentation of some hard samples
5. Back to step 3 until satisfied

Project with Juan Carlos

Research question: *correlation between seafloor habitat and environmental variables.*



Originally proposed by Kim et al. 2020

doi: 10.1109/TIP.2020.3011269

Objective: clustering of the seafloor habitats without assumption on the number or type of habitat.

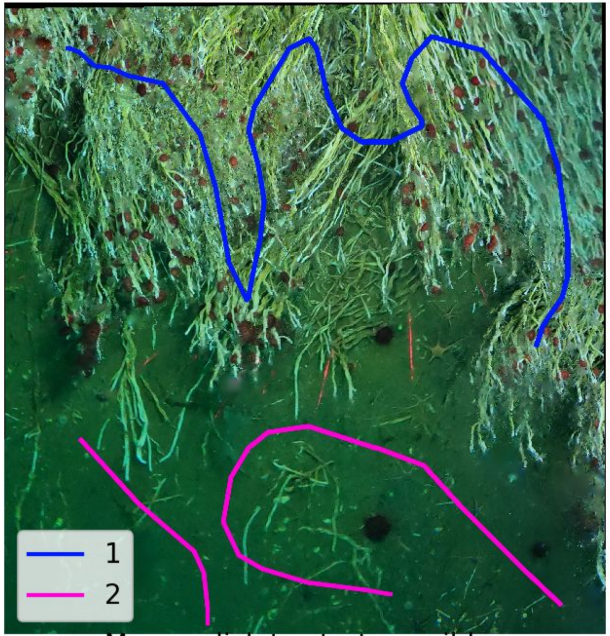
Target: spatial clustering of different habitats.

Input data: orthomosaics from photogrammetry.

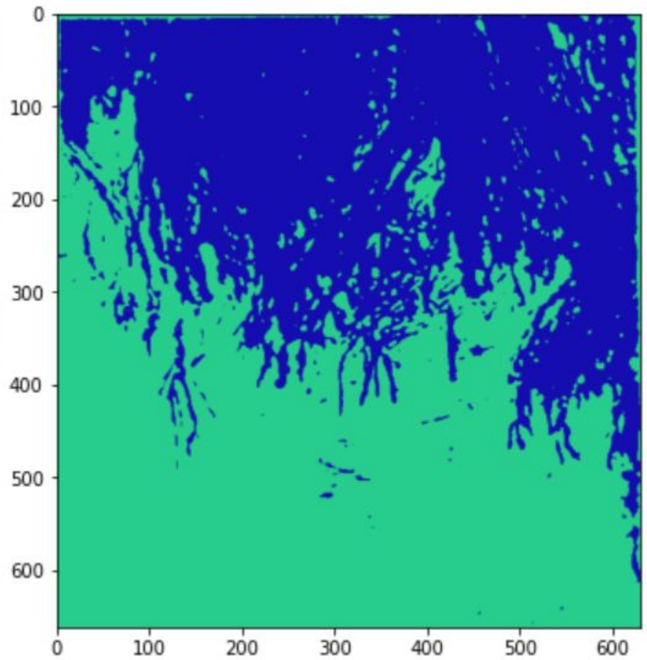
Method: Unsupervised image segmentation.

Project with Juan Carlos

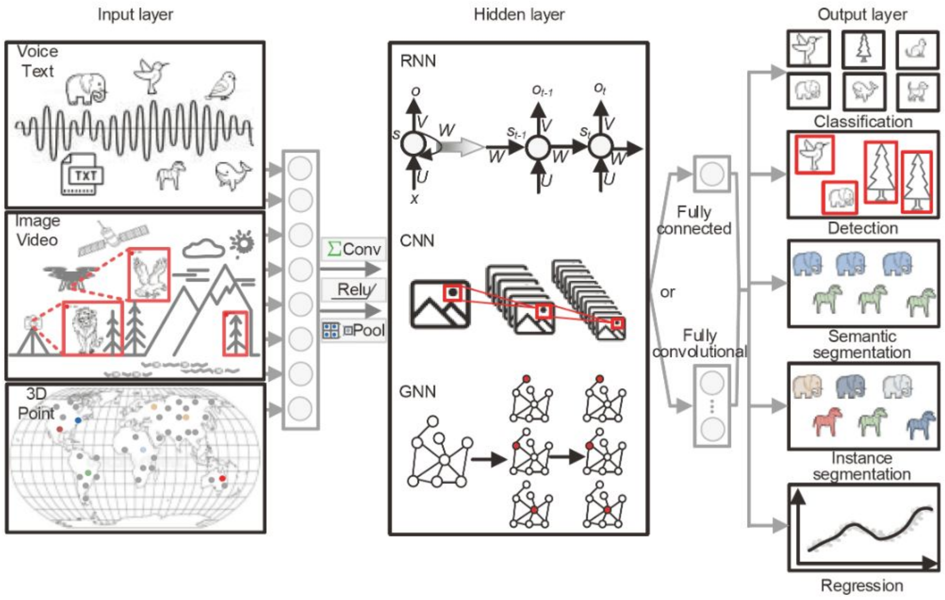
Alternative method: Weakly-supervised image segmentation.



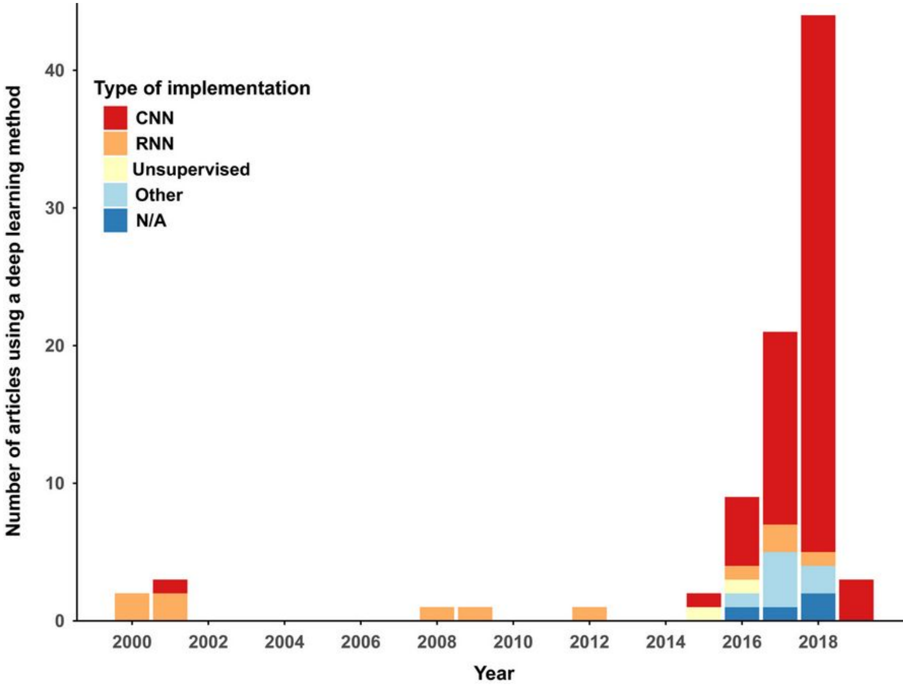
- Mouse click to start a scrible
- Change label with digits or <up>/<down> arrows
- <r> to reset current label, <t> to change random colors



What's up in Ecology?



Guo et al. 2020 ; doi: 10.1007/s11430-019-9584-9



Christin et al. 2019 ; doi: 10.1111/2041-210X.13256

Practice

Research question: *Where is bloody Wally?*

1. Dataset preparation and exploration
2. Data loading
3. Dataset splitting
4. Data augmentation
5. Model configuration
6. Model training and optimisation
7. Model evaluation on independent dataset

github.com/charleygros/IMAS_workshop_Where_is_Wally

PyTorch

colab

GitHub