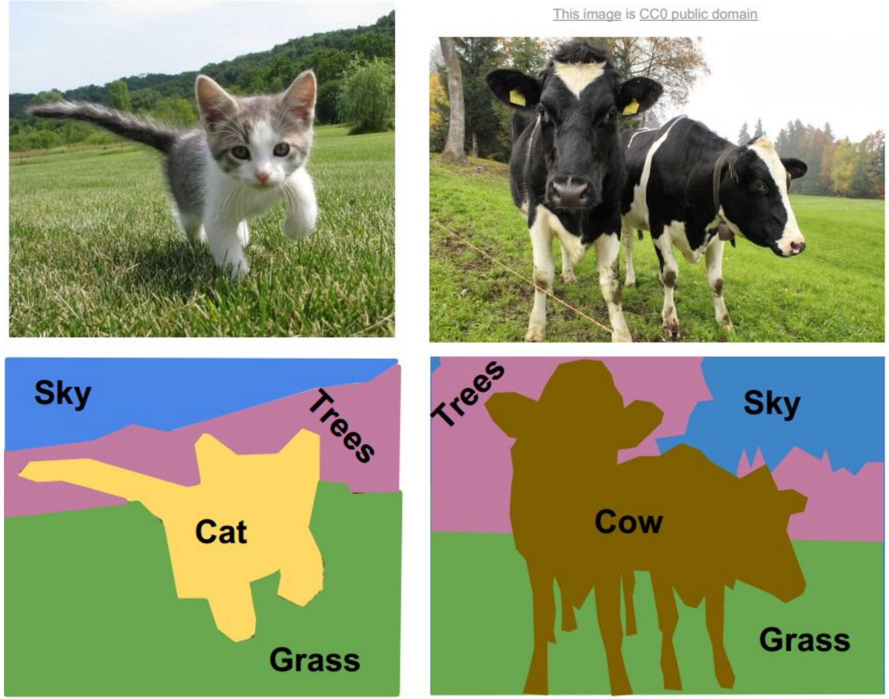
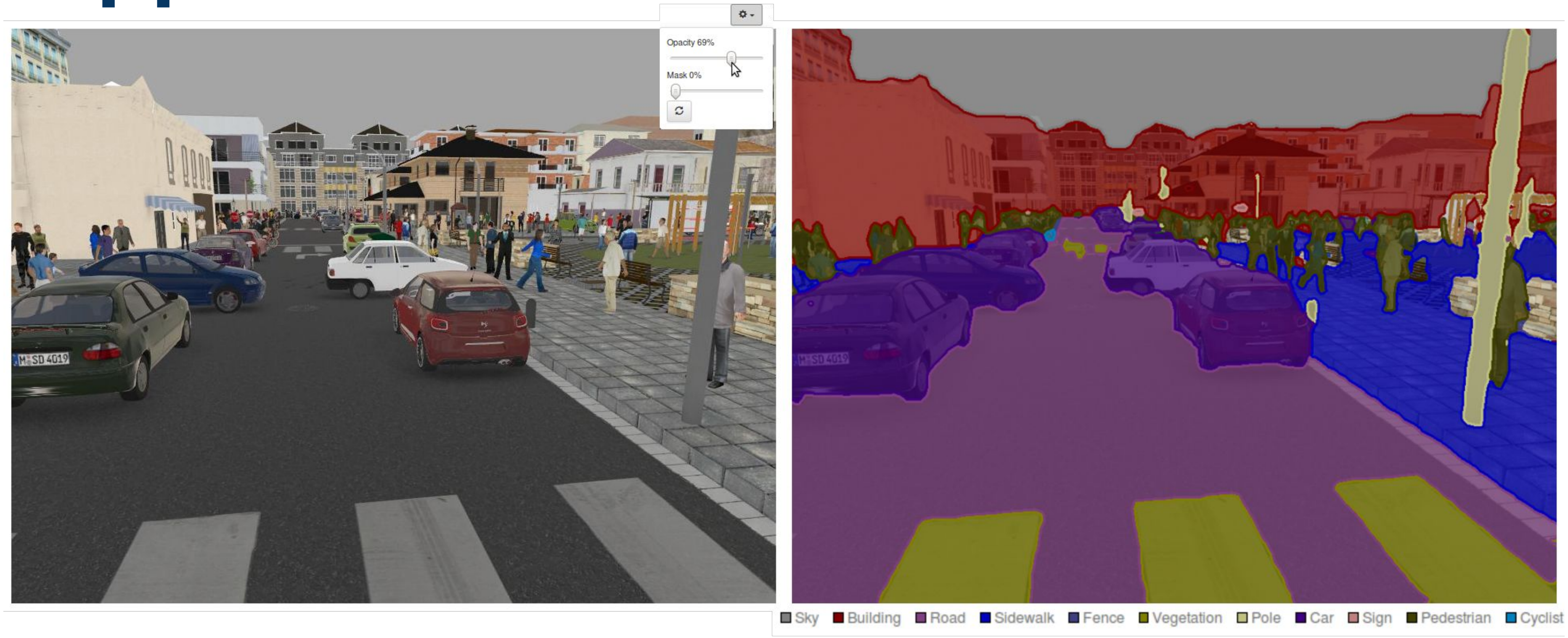
Image Segmentation

What is semantic segmentation?



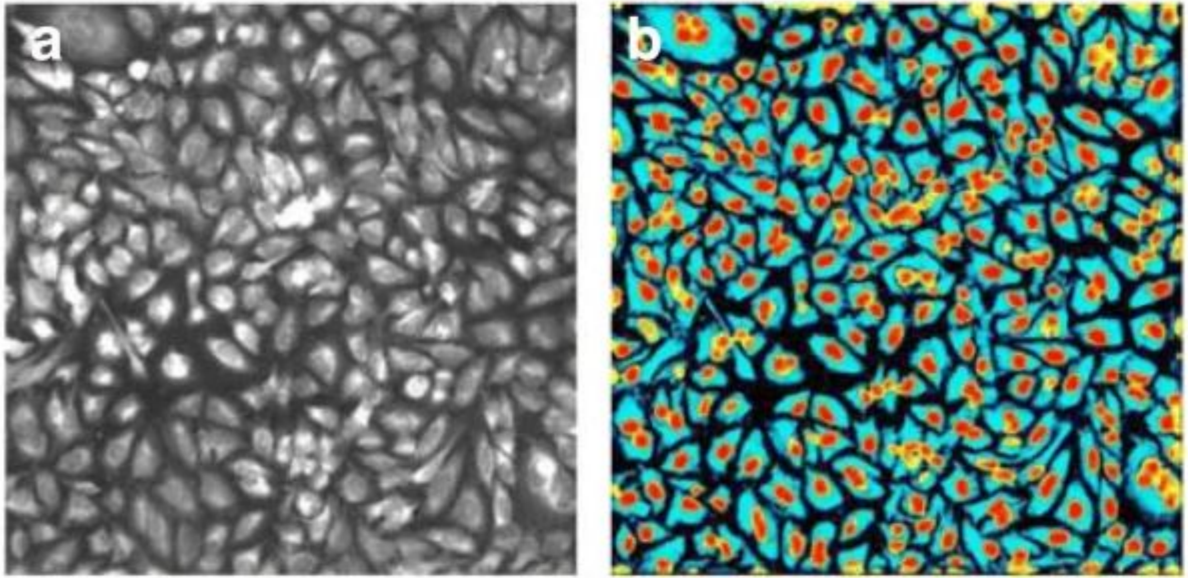
Source: https://tariq-hasan.github.io/concepts/computer-vision-semantic-segmentation/

Application: car vision



Source: https://developer.nvidia.com/blog/image-segmentation-using-digits-5/

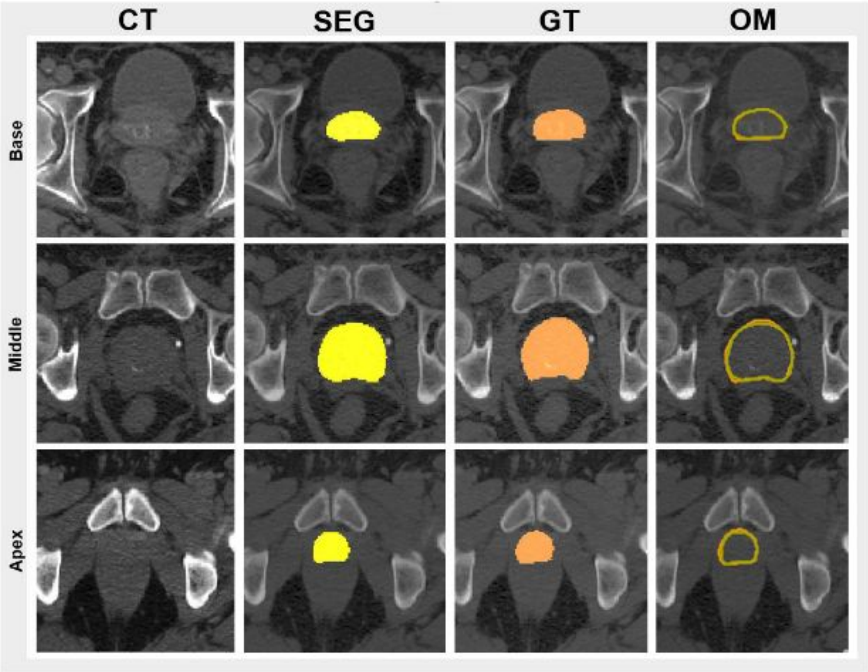
Application: cell images



a) Input image
b) Nuclei (Yellow-Red) and Cells (Blue-Cyan) prediction map.

Source: Yousef Al-Kofahi et al., A deep learning-based algorithm for 2-D cell segmentation in microscopy images. BMC Bioinformatics, 2018

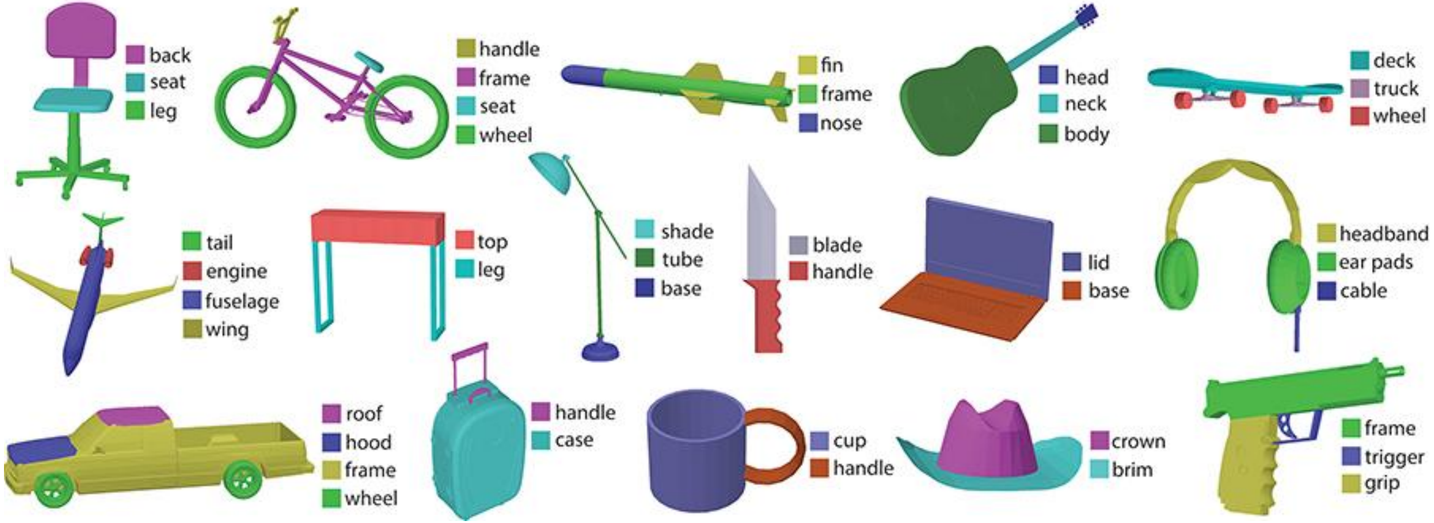
Application: organs contouring



CT: original CT scan image
SEG: U-Net segmentation
GT: ground truth
OM: overlay map of ground truth and segmented images.

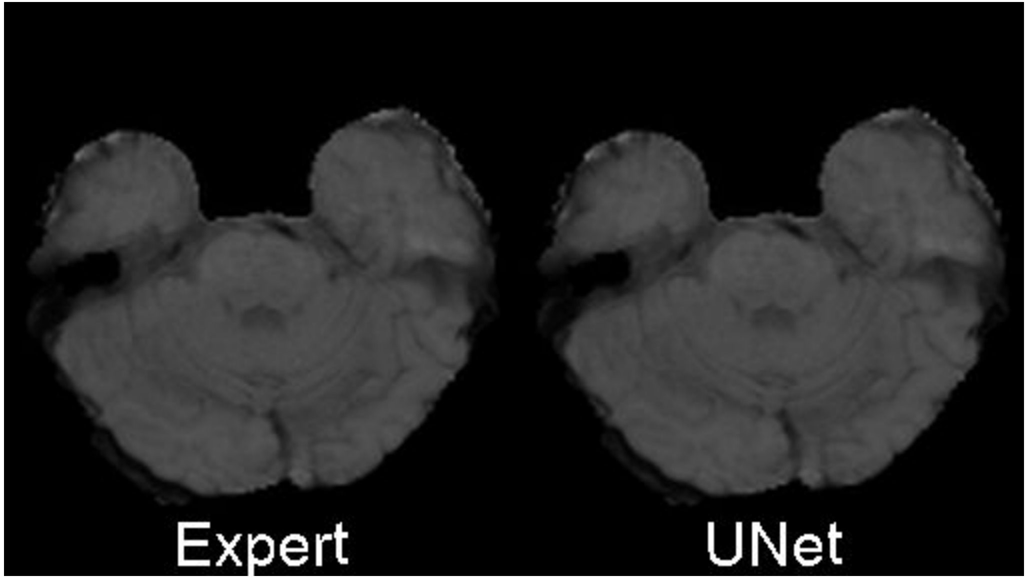
Source: Kazemifar et al., Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning. Biomedical Physics & Engineering Express, 2018

Beyond bidimensional images: 3D



Source: Kalogerakis et al., 3D Shape Segmentation with Projective Convolutional Networks. Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR) 2017

Application: tumor recognition

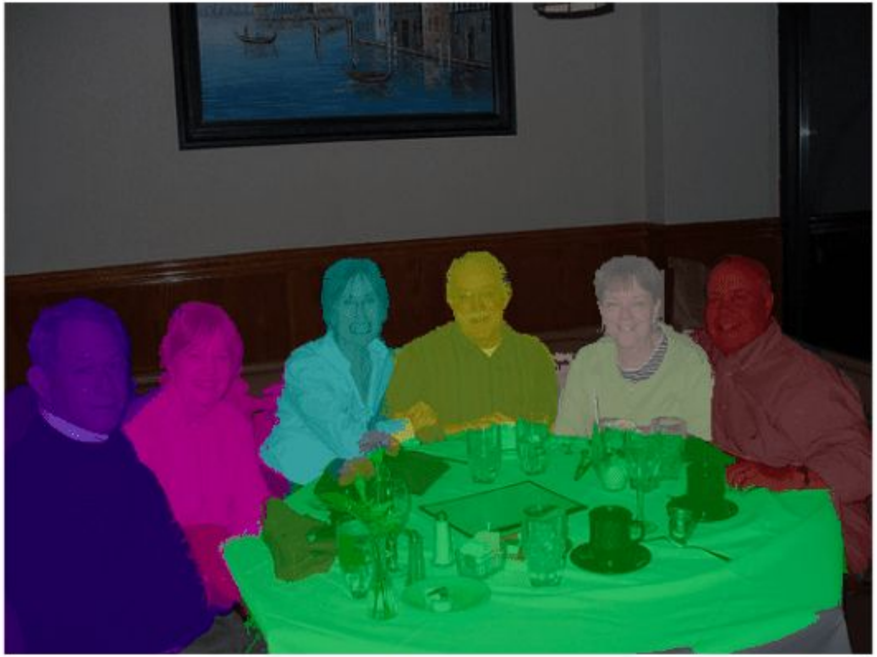


A CT scan is a set of (usually equally spaced) slices

Source: https://pythonawesome.com/keras-3d-u-net-convolution-neural-network-designed-for-medical-image-segmentation/

Semantic vs. Instance

Semantic Segmentation



Instance Segmentation

Source: Anurag Arnab, Shuai Zheng et. al 2018 “Conditional Random Fields Meet Deep Neural Networks for Semantic Segmentation”

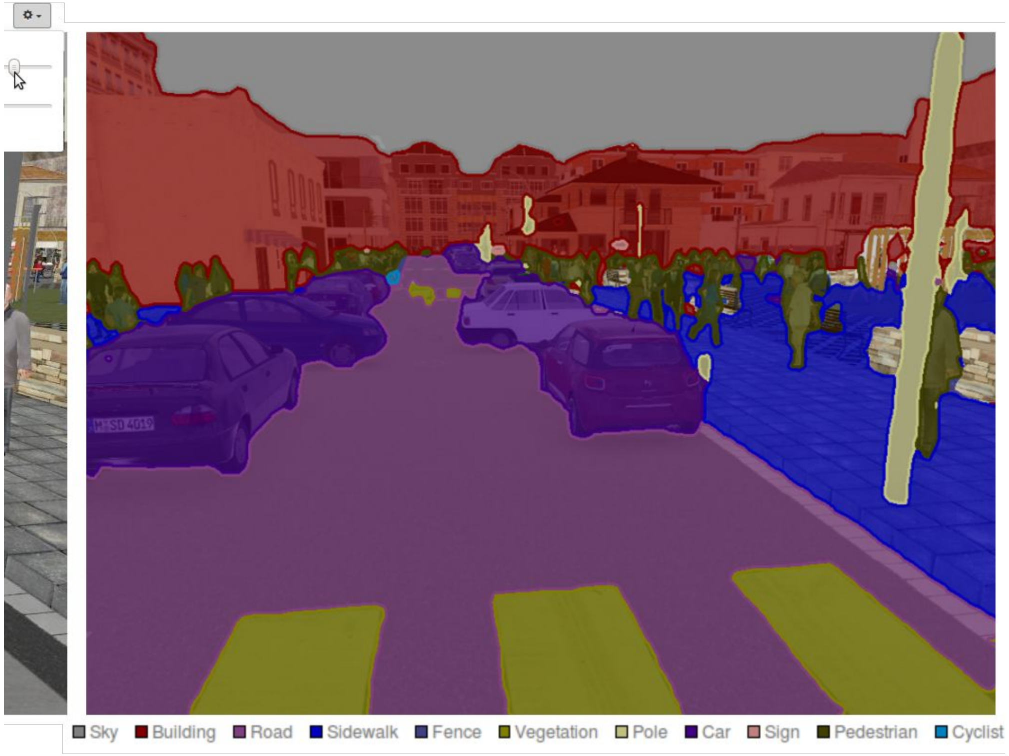
Classification vs. Segmentation

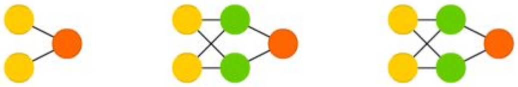
INPUT	
Image (2d, 3d…)	Image (2d, 3d…)
Ground Truth: CLASS	Ground Truth: MASK
OUTPUT	
Predicted class	Predicted mask
LOSS	
Accuracy, Binary cross entropy, …	???

Metric proposal: pixel accuracy

```
                 Number of pixel correctly classified
Pixel Accuracy = ------------------------------------
                 Total number of pixels in the image
```

Metric proposal: pixel accuracy

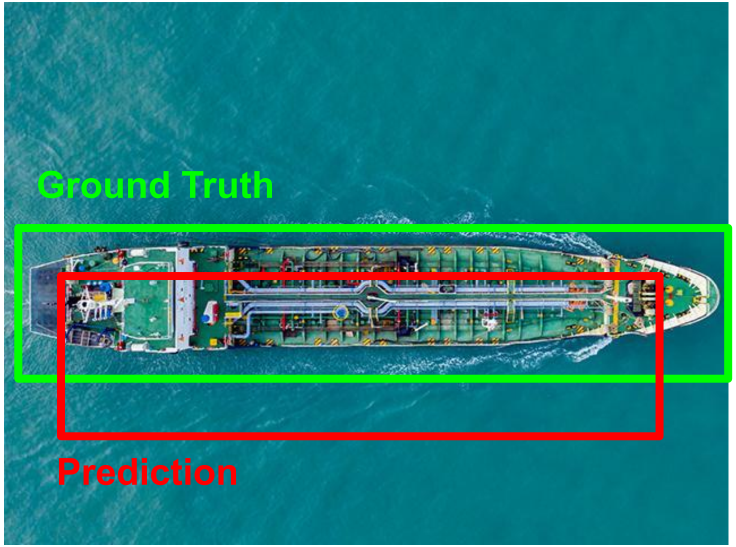




Other metrics

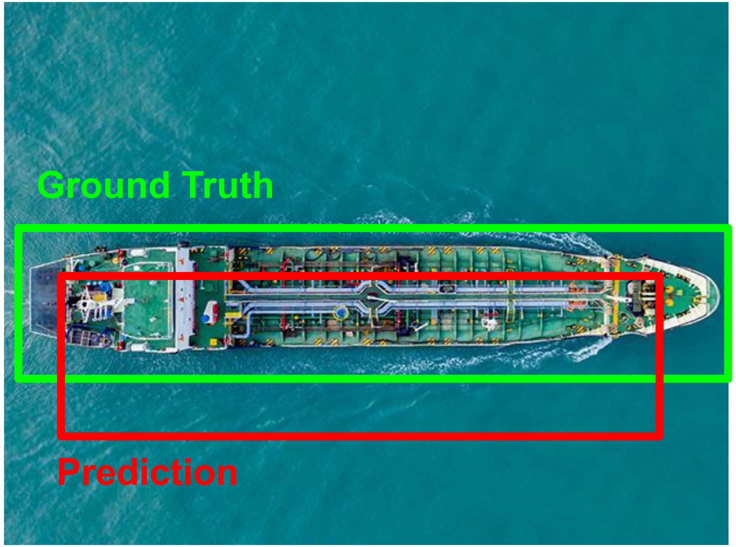
- PA - Pixel accuracy
    - MPA - Mean Pixel Accuracy
- Intersection over Union (IoU) - Jaccard Index
- Dice coefficient
- (Precision/Recall/F1 index)

Intersection Over Union




```
       Number of pixel correctly
classified
IoU =
-------------------------------------
       Union(Ground Truth, Prediction)
```

Dice coefficient



$$DC = \frac{\text{2 x Number of pixel correctly classified}}{\text{Ground Truth + Prediction}}$$

Metrics summary

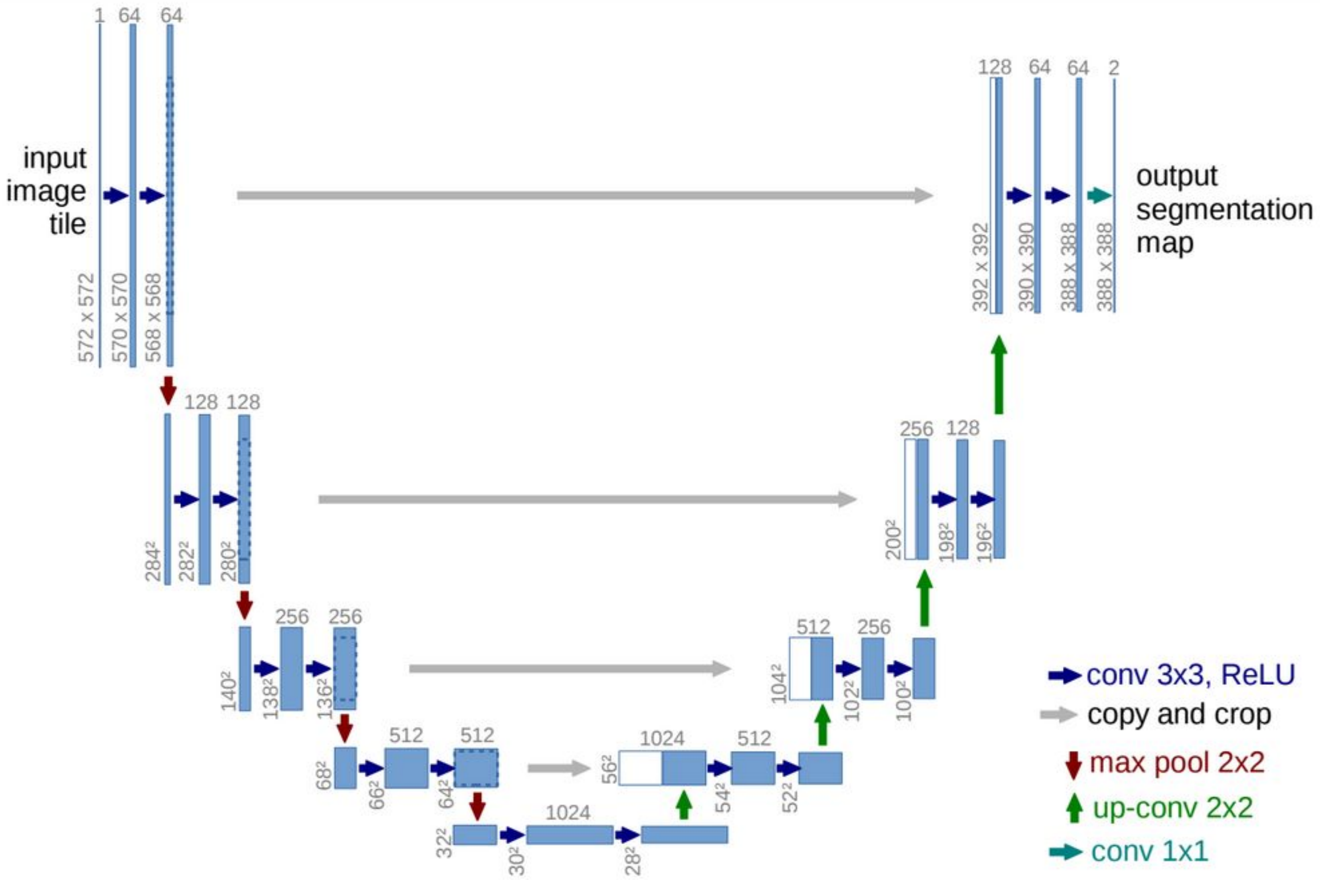
```
                     Number of pixel correctly classified
Pixel Accuracy     = ------------------------------------
                     Total number of pixels in the image


                     Number of pixel correctly classified
IoU                = ------------------------------------
                     Union(Ground Truth, Prediction)


                     2 x Number of pixel correctly classified
Dice Coefficient   = ----------------------------------------
                     Ground Truth + Prediction
```

Let’s do that

U-Net



Souce: Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.

Let’s also do that

Tips

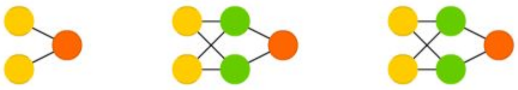
- Expensive to generate training set
- Less standard than classification
- Actively developed…
- ...but you need to follow the architecture authors
- Transfer learning is your friend

[REF] Architectures

- **U-Net**
  - Olaf Ronneberger et. al 2015 “U-net architecture image segmentation”
  - https://arxiv.org/abs/1505.04597
- **FastFCN**
  - Huikai Wu et.al 2019 “FastFCN: Rethinking Dilated Convolution in the Backbone for Semantic Segmentation”
  - https://arxiv.org/abs/1903.11816
- **Gated-SCNN**
  - Towaki Takikawa et. al 2019 “Gated-SCNN: Gated Shape CNNs for Semantic Segmentation”
  - https://arxiv.org/abs/1907.05740
- **DeepLab**
  - *Liang-Chieh Chen et. al 2016 “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs”*
  - *https://arxiv.org/abs/1606.00915*
- **Mask R-CNN**
  - Kaiming He et. al 2017 “Mask R-CNN”
  - https://arxiv.org/abs/1703.06870

[REF] Databases

- **[2D] Pascal Visual Object Classes (VOC)**
  - http://host.robots.ox.ac.uk/pascal/VOC/index.html
  - 21 object classes (vehicle, household, animal, airplane…)
- **[2D] Microsoft COCO: Common Objects in Context**
  - https://paperswithcode.com/dataset/coco
  - “91 objects types that would be easily recognizable by a 4 year old”
- **[2D] Cityscapes**
  - https://www.cityscapes-dataset.com/
  - 5 000 images with high quality annotations · 20 000 images with coarse annotations · 50 different cities
- **[2.5D] Sun RGB-D**
  - https://rgbd.cs.princeton.edu/
  - four different sensors and contains 10,000 RGB-D images
- **[3D] Stanford 2D-3D-Semantics Dataset**
  - http://buildingparser.stanford.edu/dataset.html
  - It covers over 6,000 m2 and contains over 70,000 RGB images, along with the corresponding depths, surface normals, semantic annotations, global XYZ images



[REF] Useful stuff

- **Image Segmentation Using Deep Learning: A Survey**
  - Minaee, Shervin, et al. "Image segmentation using deep learning: A survey." IEEE Transactions on Pattern Analysis and Machine Intelligence (2021).
- **Data preprocessing and augmentation using Torch.IO**
  - (take away the message, even if using pyTorch)
  - https://colab.research.google.com/github/fepegar/torchio-notebooks/blob/main/notebooks/Data_preprocessing_and_augmentation_using_TorchIO_a_tutorial.ipynb