Deep learning for

Segmentation

Where are we now

TensorMask: A Foundation for Dense Object Segmentation

Xinlei Chen Ross Girshick Kaiming He Piotr Dollár
Facebook AI Research (FAIR)

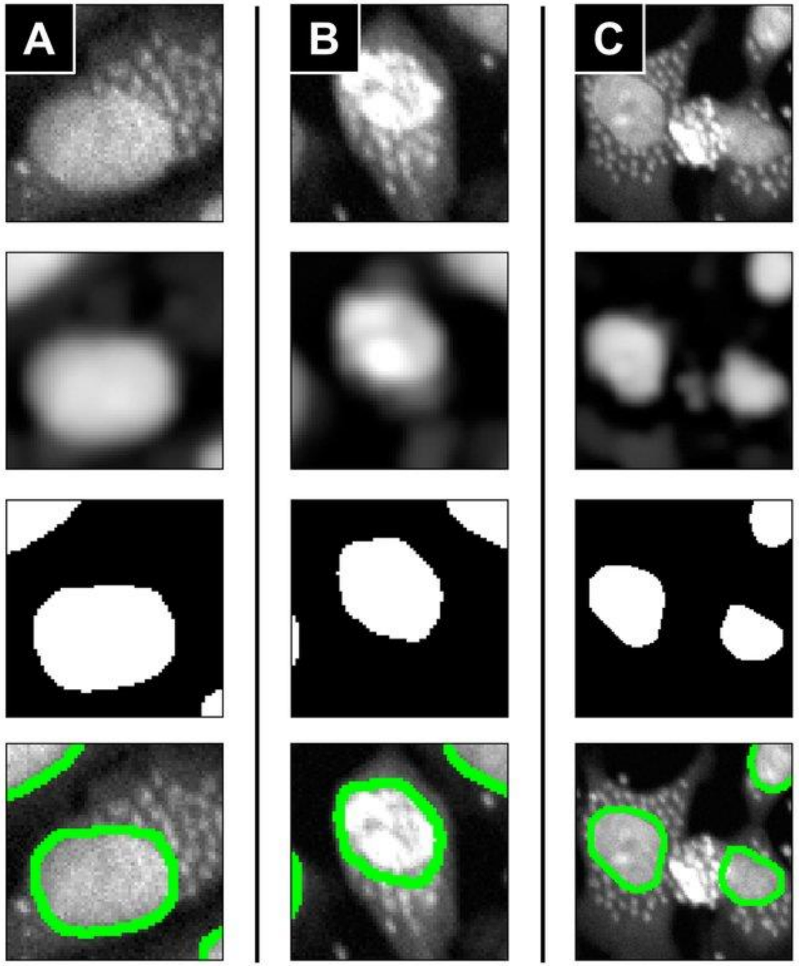
Abstract

Sliding-window object detectors that generate bounding-box object predictions over a dense, regular grid have advanced rapidly and proven popular. In contrast, modern instance segmentation approaches are dominated by methods that first detect object bounding boxes, and then crop and segment these regions, as popularized by Mask R-CNN. In this work, we investigate the paradigm of dense sliding-window instance segmentation, which is surprisingly under-explored. Our core observation is that this task is fundamentally different than other dense prediction tasks such as semantic segmentation or bounding-box object detection, as the output at every spatial location is itself a geometric structure with its own spatial dimensions. To formalize this,

Figure 1. Selected output of *TensorMask*, our proposed framework for performing *dense sliding-window instance segmentation*. We treat dense instance segmentation as a prediction task over *structured* 4D tensors. In addition to obtaining competitive quantitative results, TensorMask achieves results that are *qualitatively* reason-

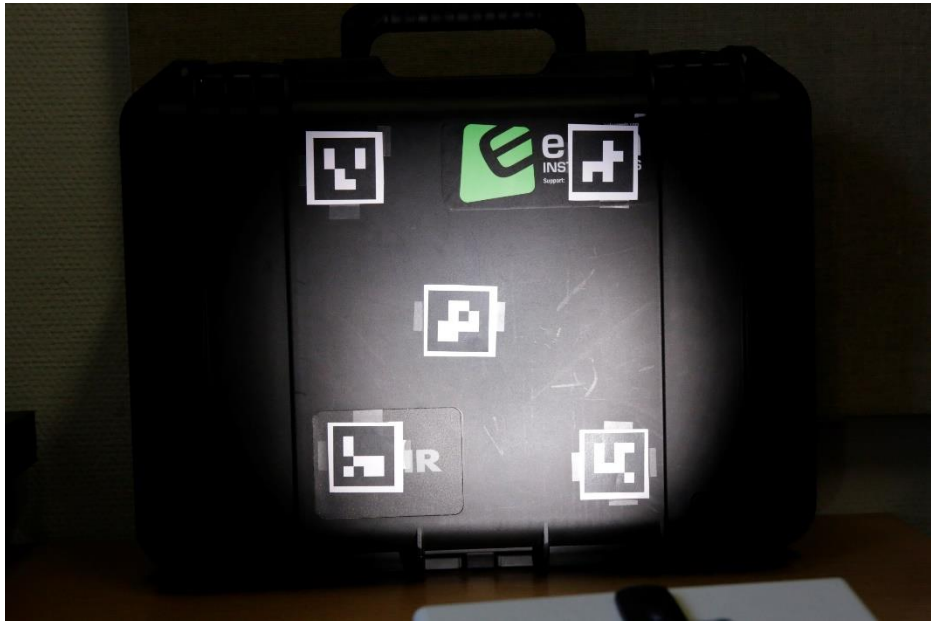
Why is segmentation difficult

- Non uniform illumination
- Partial occlusion
- Overlapping objects
- Shape inconsistency
- Class imbalance
- Noise
- Variance in ground truth

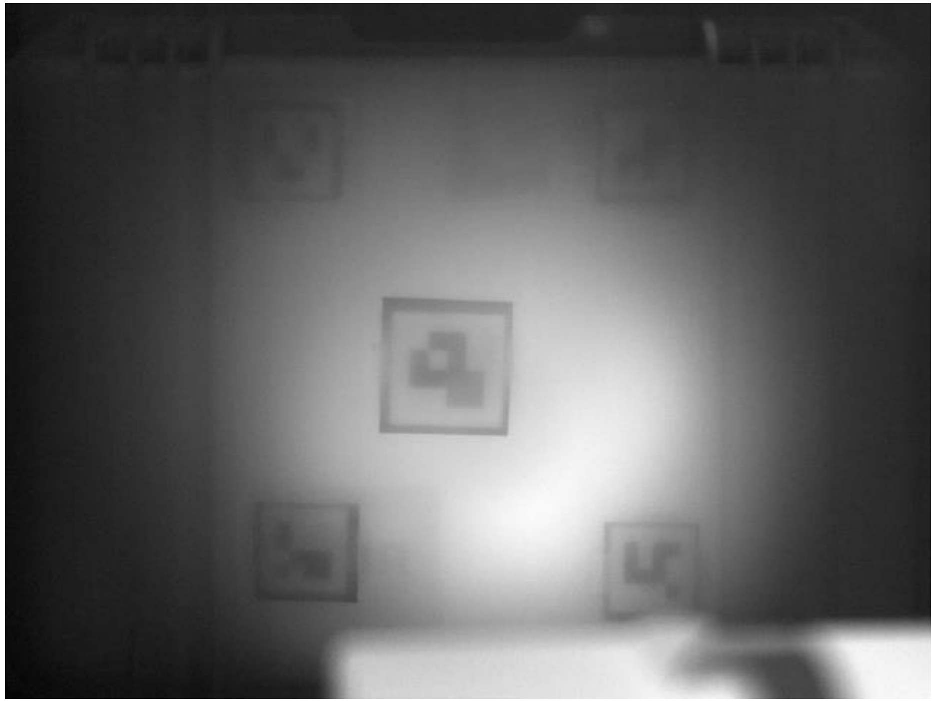


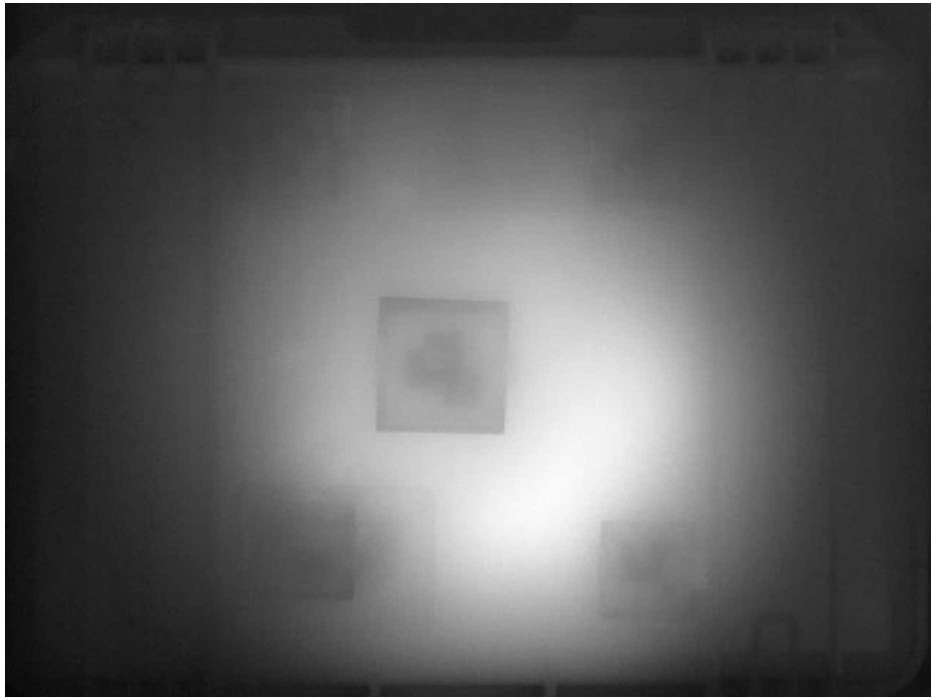
Classical approaches

- Watershed
- Snake
- Active contour
- Superpixel
- Graphcut
- Morphological filtering
- etc.

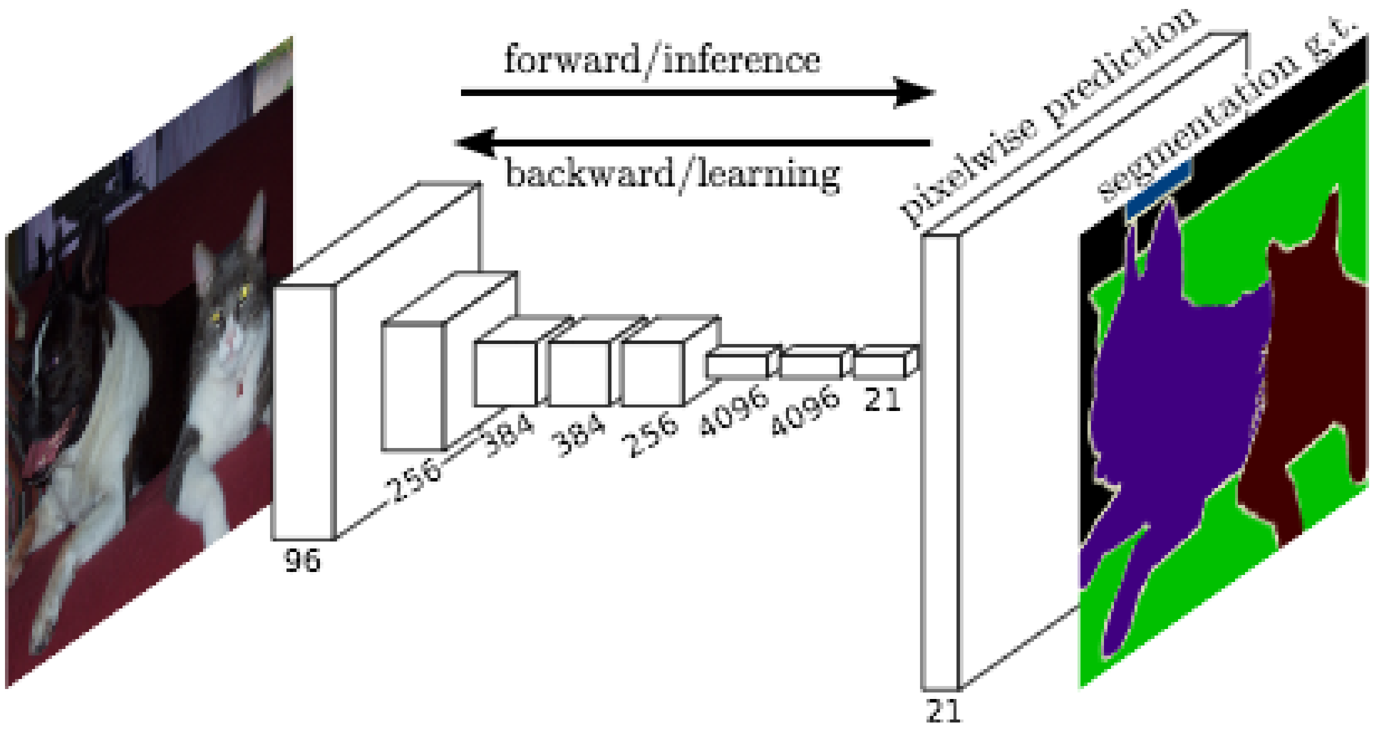


What happens to images like this



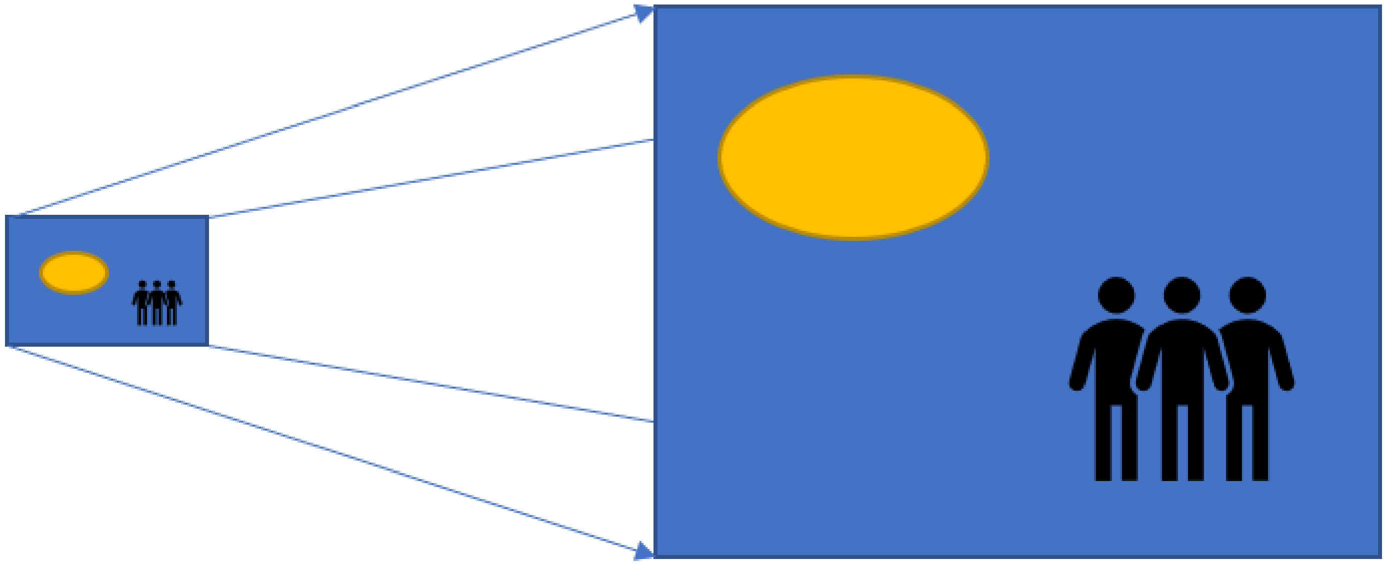


Deep learning for segmentation (2015)



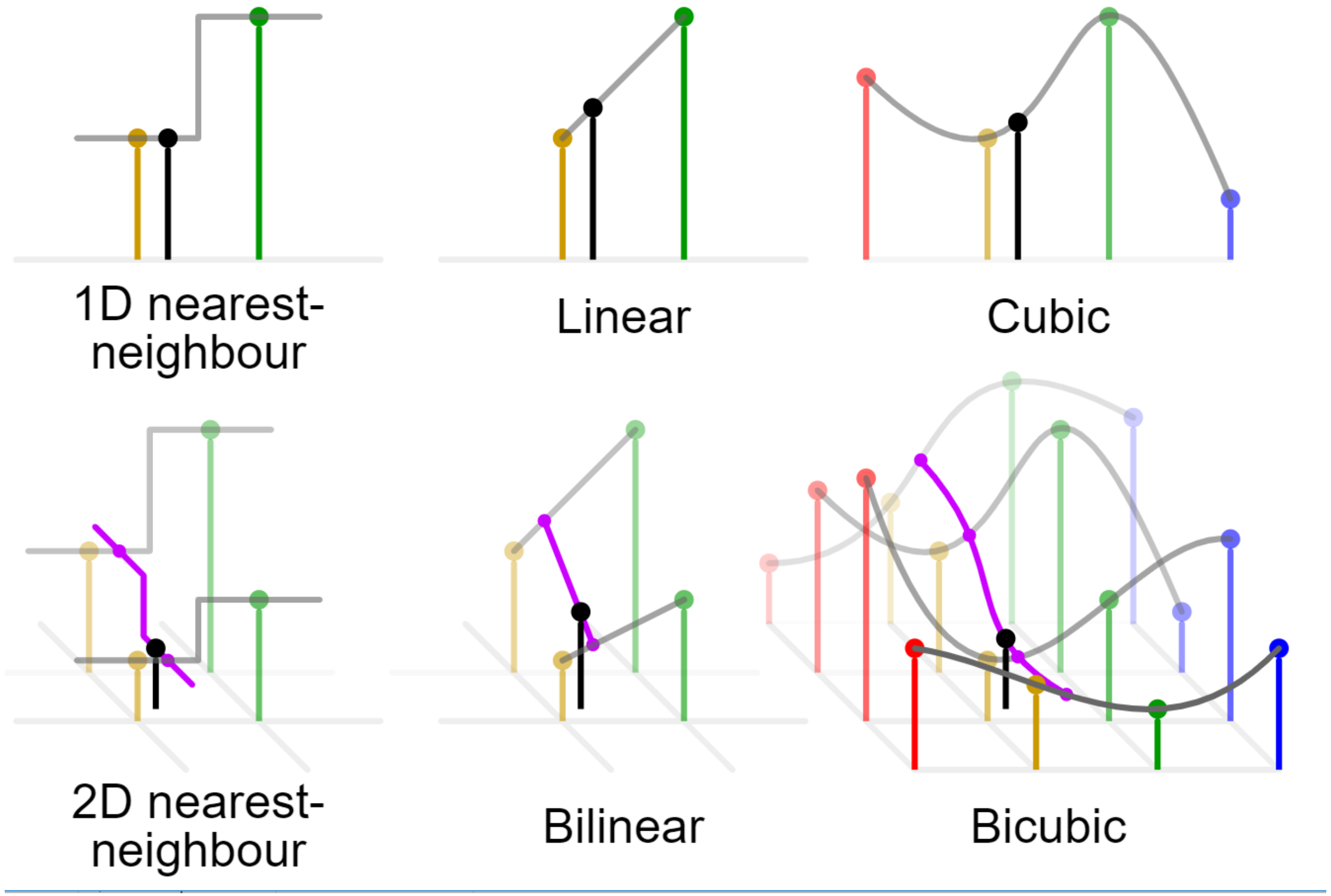
https://people.eecs.berkeley.edu/~jonlong/long_shelhamer_fcn.pdf

Up-sampling with Transposed Convolution



- Fractionally-strided convolution
- Deconvolution

Interpolation



General convolution

4	5	8	7
1	8	8	8
3	6	6	4
6	5	7	8

conv

1	4	1
1	4	3
3	3	1

$\rightarrow$

122	148
126	134

Convolution by Matrix multiplication

$$
\begin{bmatrix}
1 & 4 & 1 & 0 & 1 & 4 & 3 & 0 & 3 & 3 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 4 & 1 & 0 & 1 & 4 & 3 & 0 & 3 & 3 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 4 & 1 & 0 & 1 & 4 & 3 & 0 & 3 & 3 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 4 & 1 & 0 & 1 & 4 & 3 & 0 & 3 & 3 & 1
\end{bmatrix}
\otimes
\begin{bmatrix} 4 \\ 5 \\ 8 \\ 7 \\ 1 \\ 8 \\ 8 \\ 8 \\ 3 \\ 6 \\ 6 \\ 4 \\ 6 \\ 5 \\ 7 \\ 8 \end{bmatrix}
\rightarrow
\begin{bmatrix} 122 \\ 148 \\ 126 \\ 134 \end{bmatrix}
$$

Transposed multiplication

1	0	0	0
4	1	0	0
1	4	0	0
0	1	0	0
1	0	1	0
4	1	4	1
3	4	1	1
0	3	0	1
3	0	1	0
3	3	4	1
1	3	3	4
0	1	0	3
0	0	3	0
0	0	3	3
0	0	1	3
0	0	0	1

$\otimes$

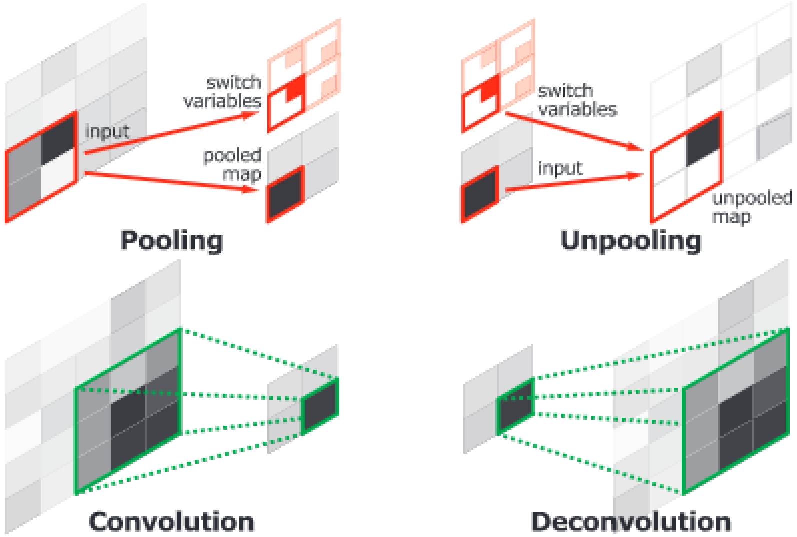
2
1
4
4

$\rightarrow$

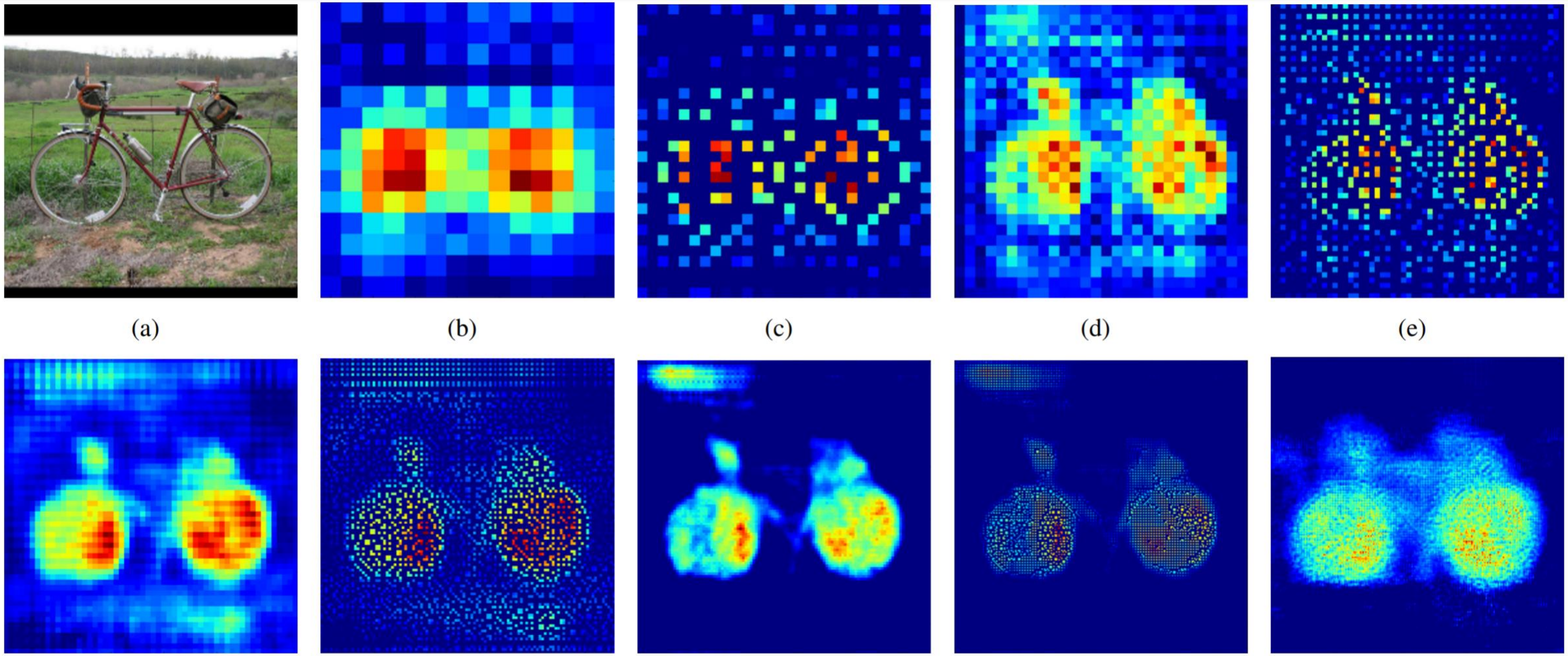
2
9
6
1
6
29
30
7
10
29
33
13
12
24
16
4

Convolutional and Deconvolutional Networks (2015)

- https://arxiv.org/pdf/1505.04366.pdf

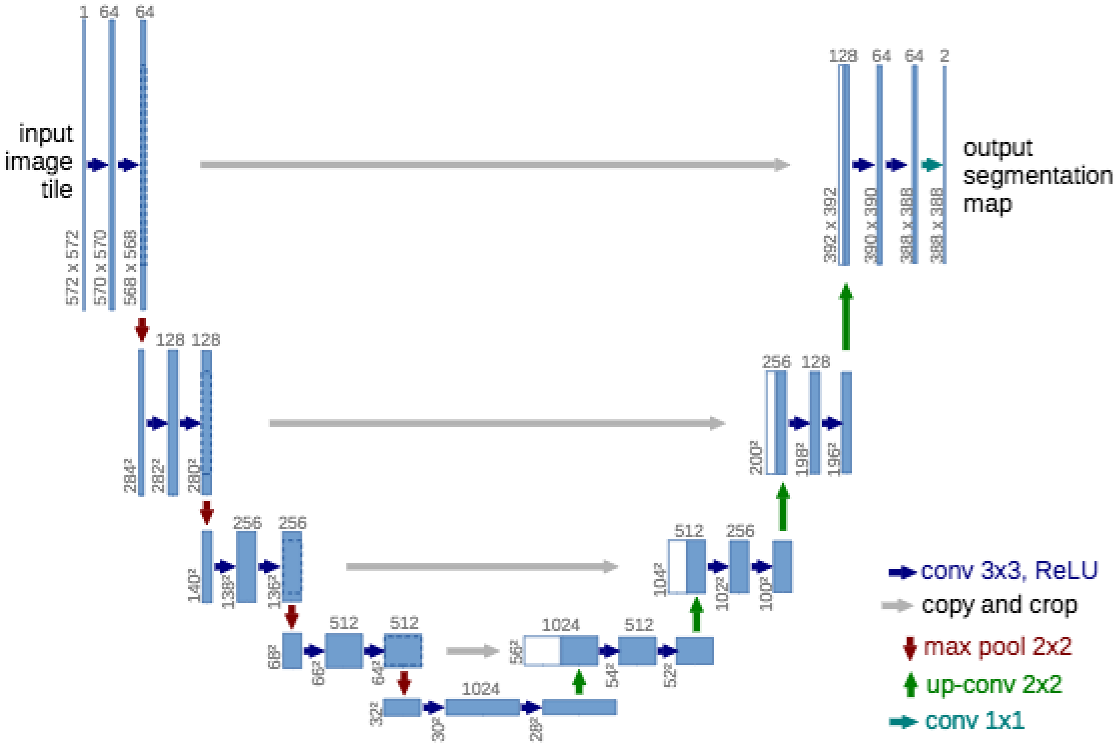


Impact of unpooling and deconvolution

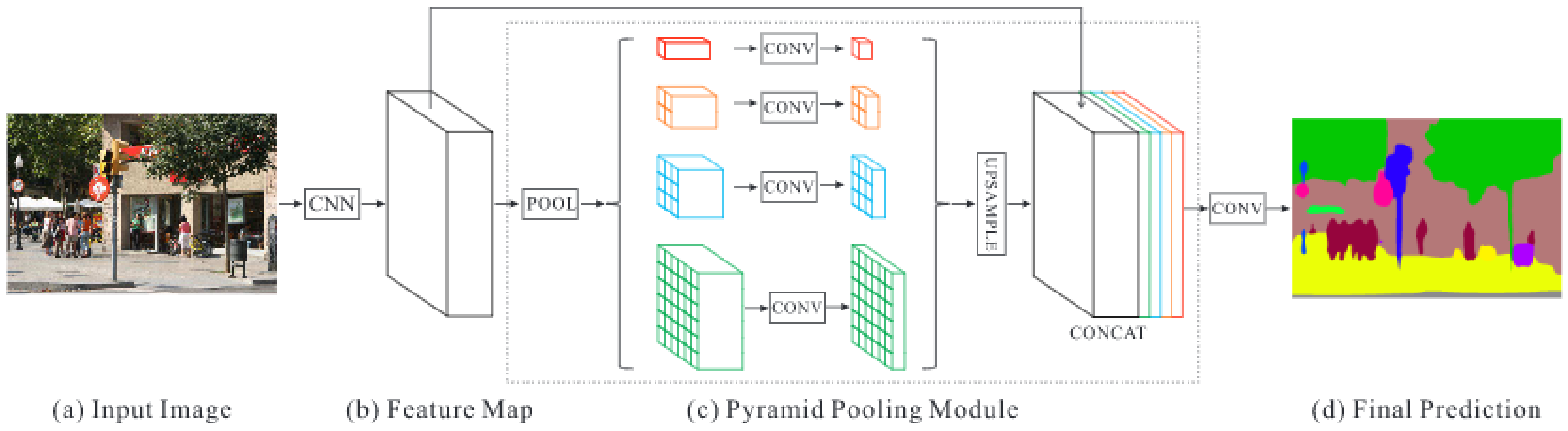


(a) (b) (c) (d) (e)

U-net (2015)

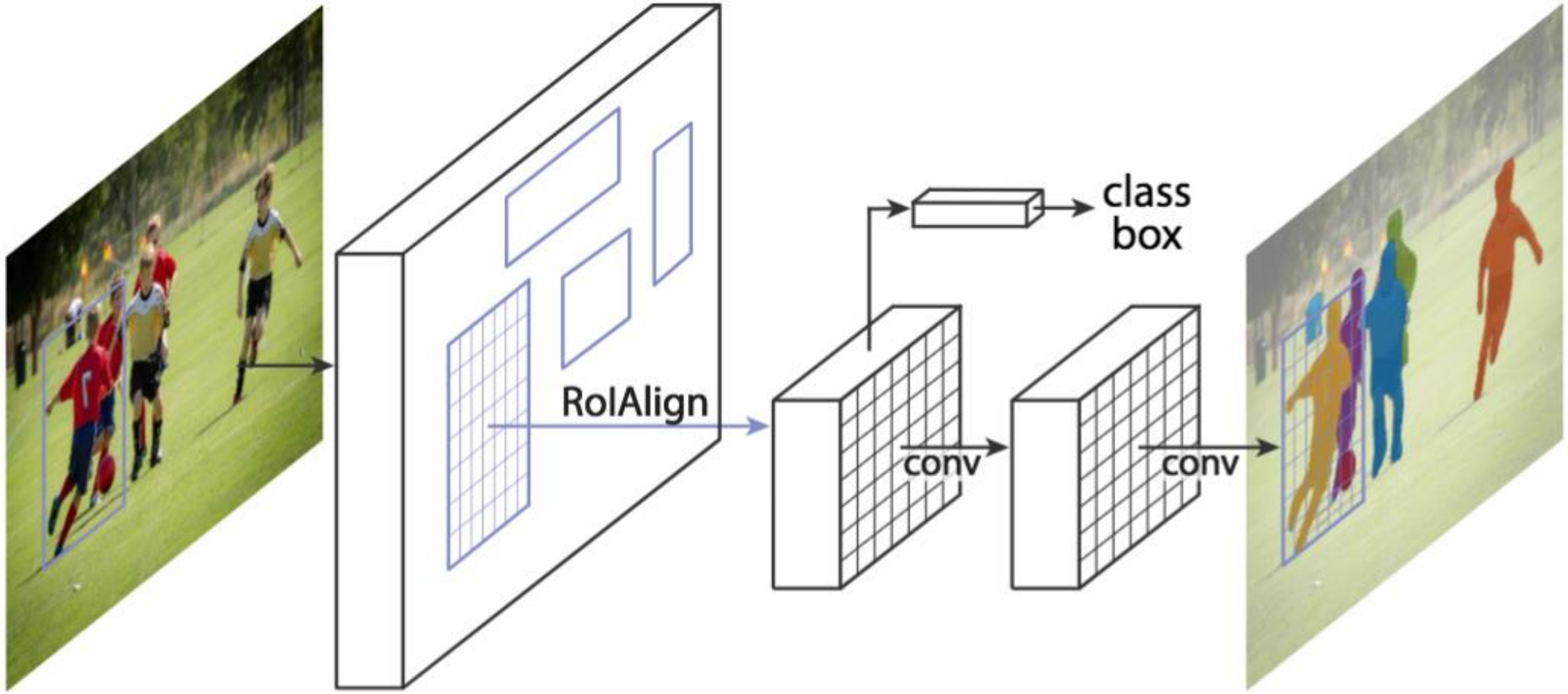

input image tile
output segmentation map
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1
572 x 572
570 x 570
568 x 568
392 x 392
390 x 390
388 x 388
388 x 388

Pyramid Scene Parsing Network (2017)

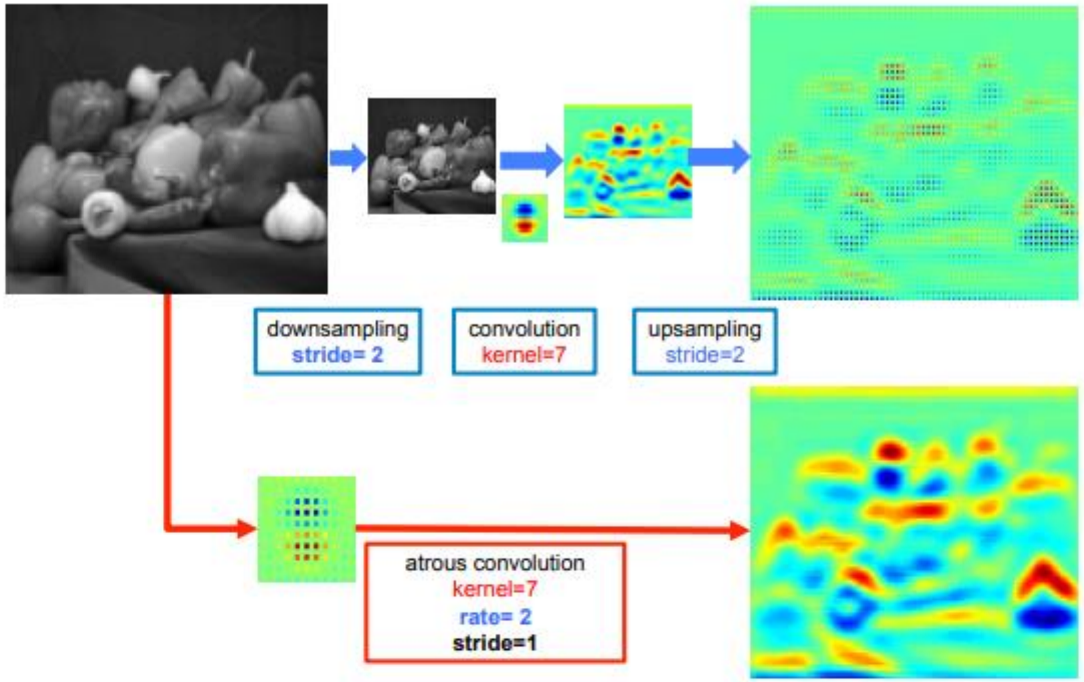


https://arxiv.org/pdf/1612.01105.pdf

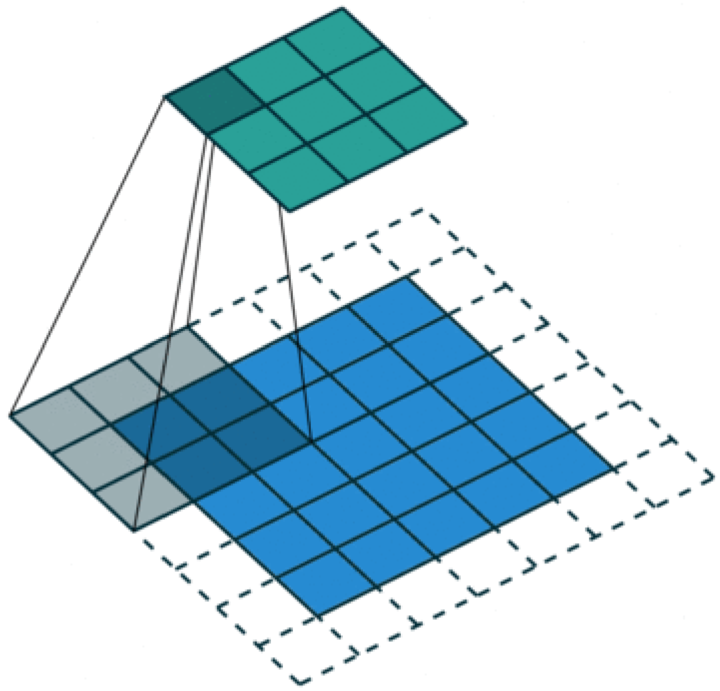
Mask RCNN



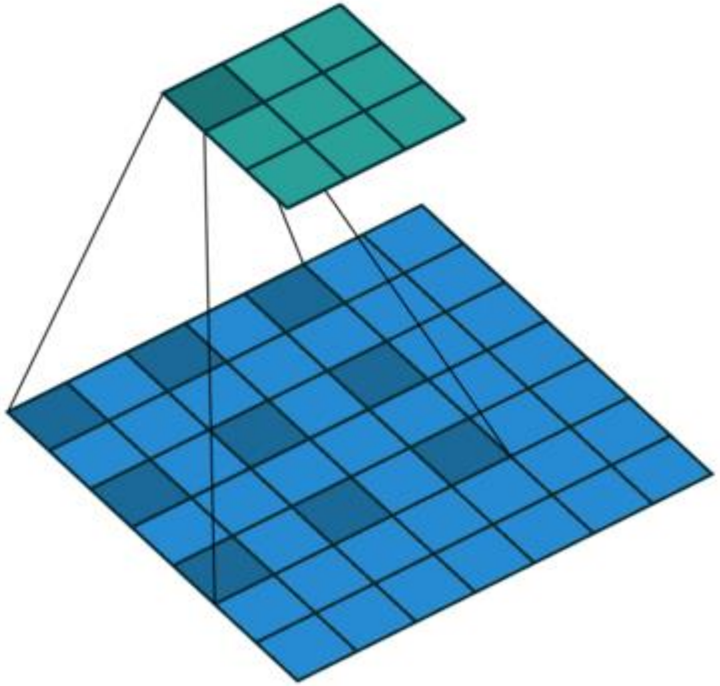
Deeplab (2017)



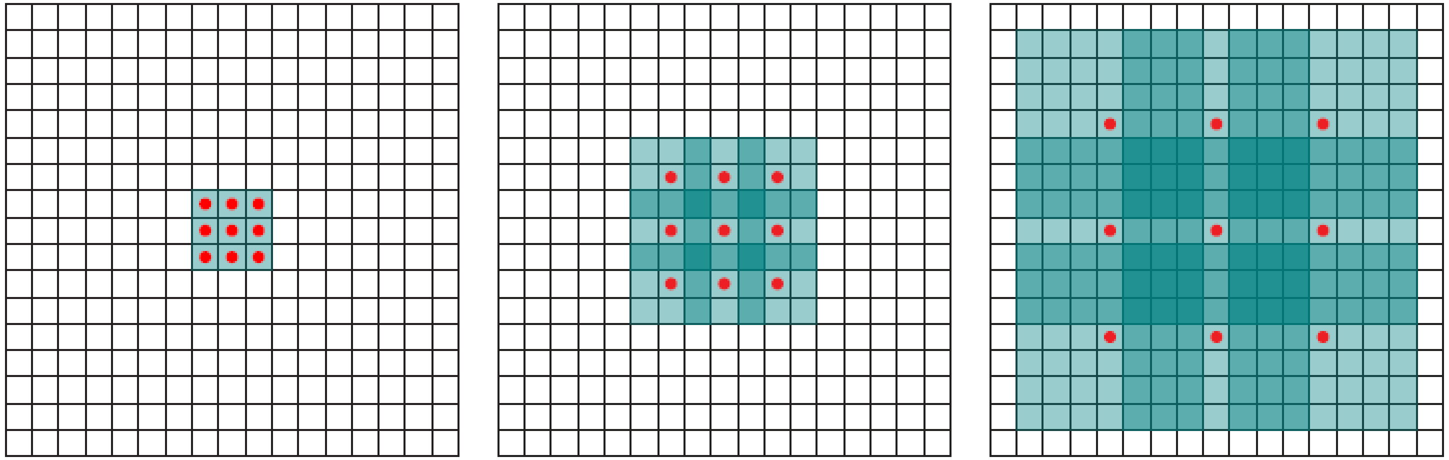
Convolution



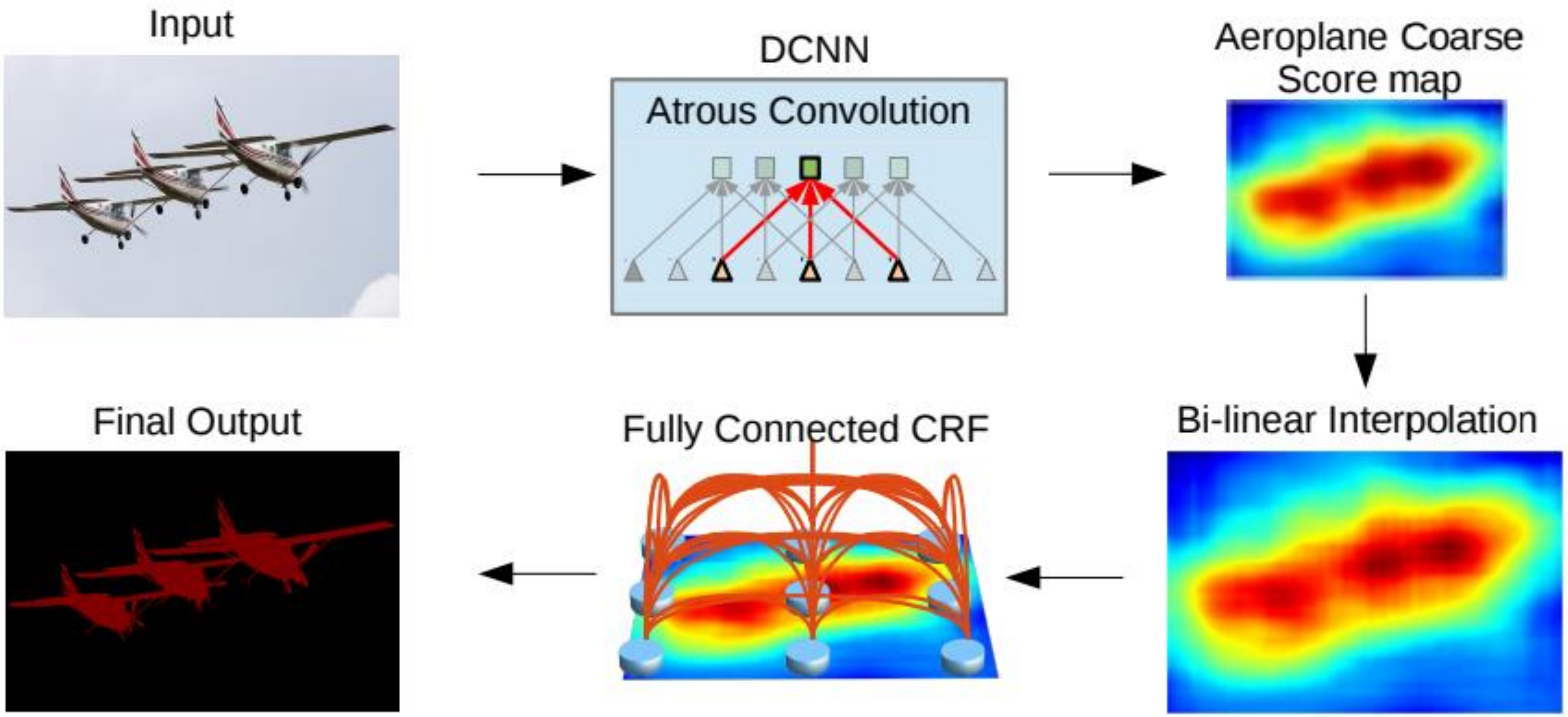
Dilated convolution



Scaling up the dilation

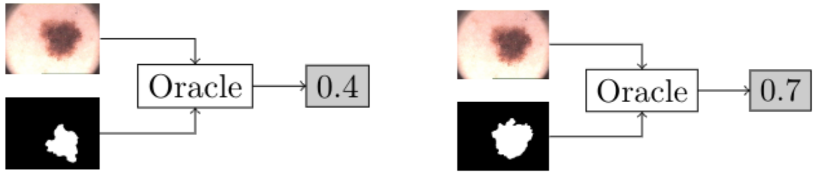


Deeplab Progression



Oracle segmentation

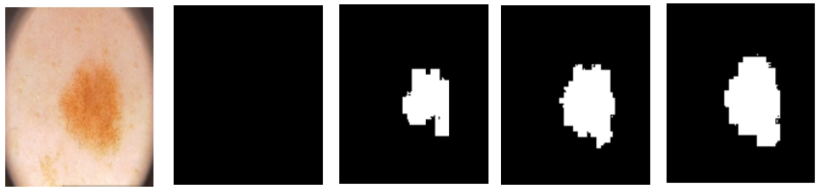
1 Training the oracle



The oracle is trained to produce a score based on segmentations of differing qualities, which are synthetically produced:

Original | Elastic | Dilation | Erosion | Random switches | Rotate | Flip & offset

2 Using the oracle for segmentation



(a) Source image (c) Initial mask (d) 5 steps (e) 10 steps (f) 100 steps

Once the oracle is trained, a mask can be given which can be improved by gradient ascent:

$$\hat{m}_{ij} \leftarrow \hat{m}_{ij} + \alpha \frac{\partial \hat{q}}{\partial \hat{m}_{ij}}$$

Advantage • segmentation is performed iteratively; it can be used to improve existing segmentations.

Fernandes, Kelwin, Ricardo Cruz, and Jaime S. Cardoso. "Deep image segmentation by quality inference." 2018 International Joint Conference on Neural Networks (IJCNN). IEEE, 2018.

Other networks

- Path Aggregation Network (PANet)
- TensorMask
- Context Encoding Network (EncNet)
- Pixelnet
- Enet
- Icnet
- Renet
- Reseg
- Parsenet
- etc

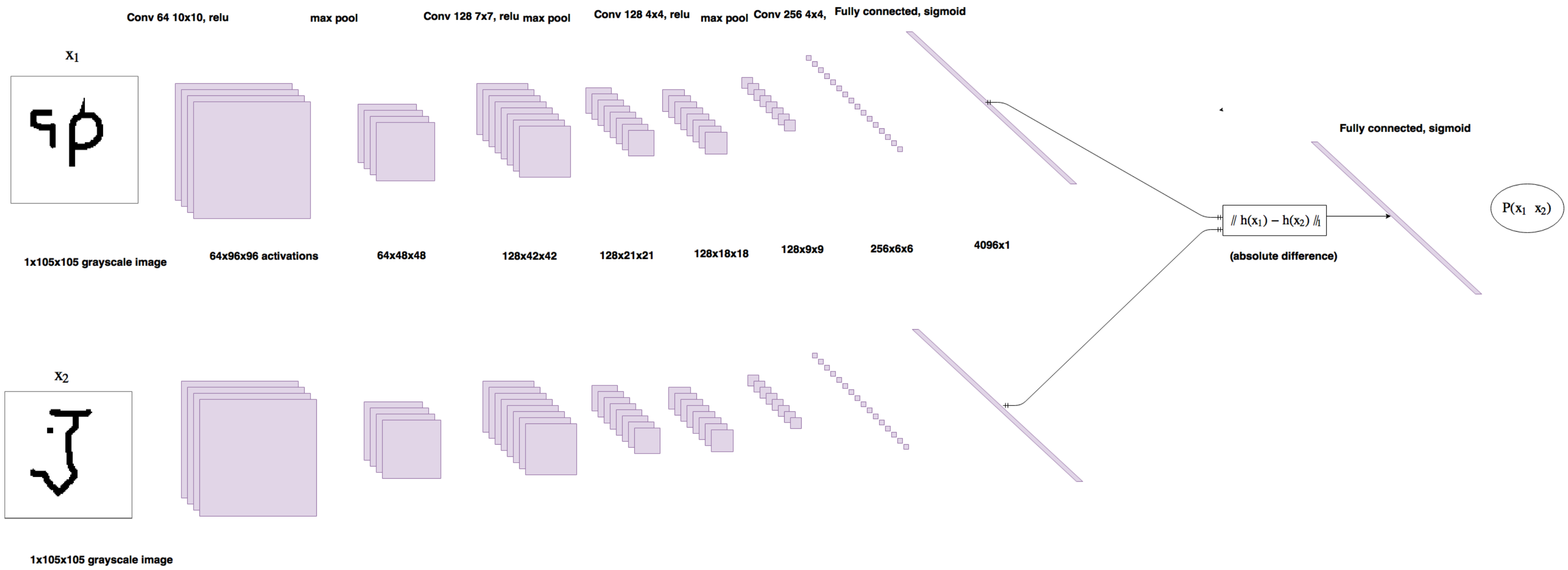
Open topic

- Useful datasets for training
- Oneshot learning / Triplet loss
- Hard negative mining
- Augmentation
- Class imbalance / condition imbalance
- Transfer learning
- Deep learning for 3D
- Style transfer with deep learning
- Overfitting
- Types of validation
- Smart applications of deep learning

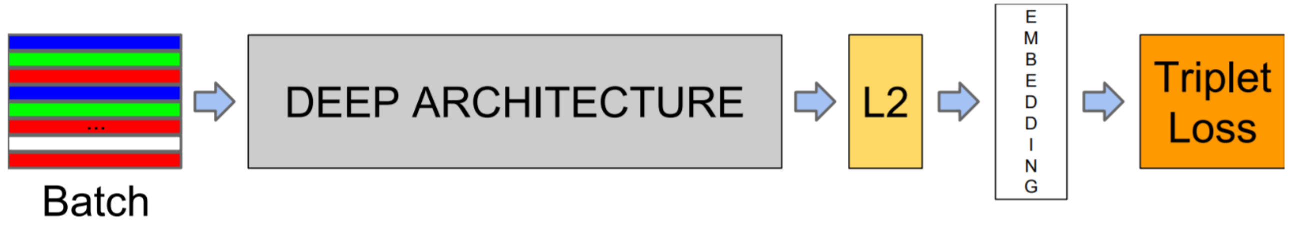
Useful Dataset

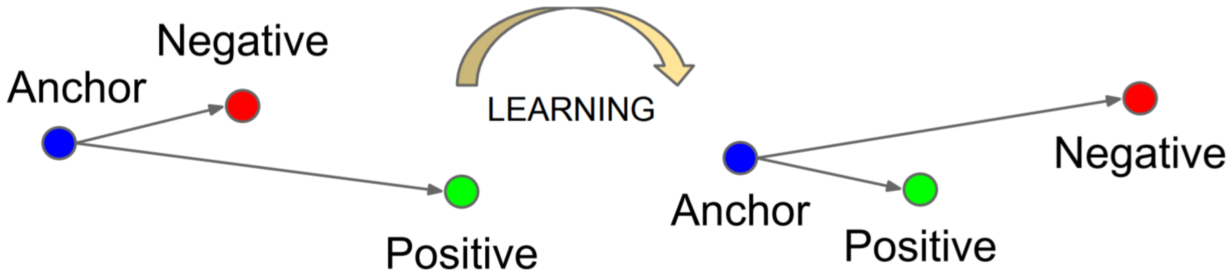
- IMAGENET
- PASCAL VOC
- MS COCO
- SPORTS-1M
- YOUTUBE-8M
- CIFAR-10
- CALTECH datasets
- MNIST
- LABELME
- COIL100
- Labelled Faces in the Wild
- Google’s Open Images
- Food101
- etc

One shot learning – Siamese network

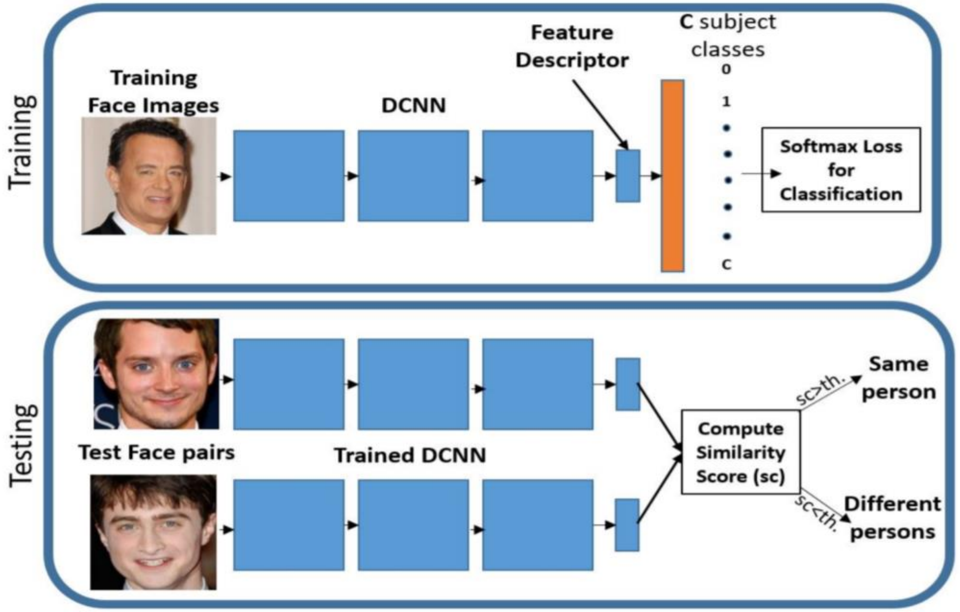


Triplet loss





Class imbalance



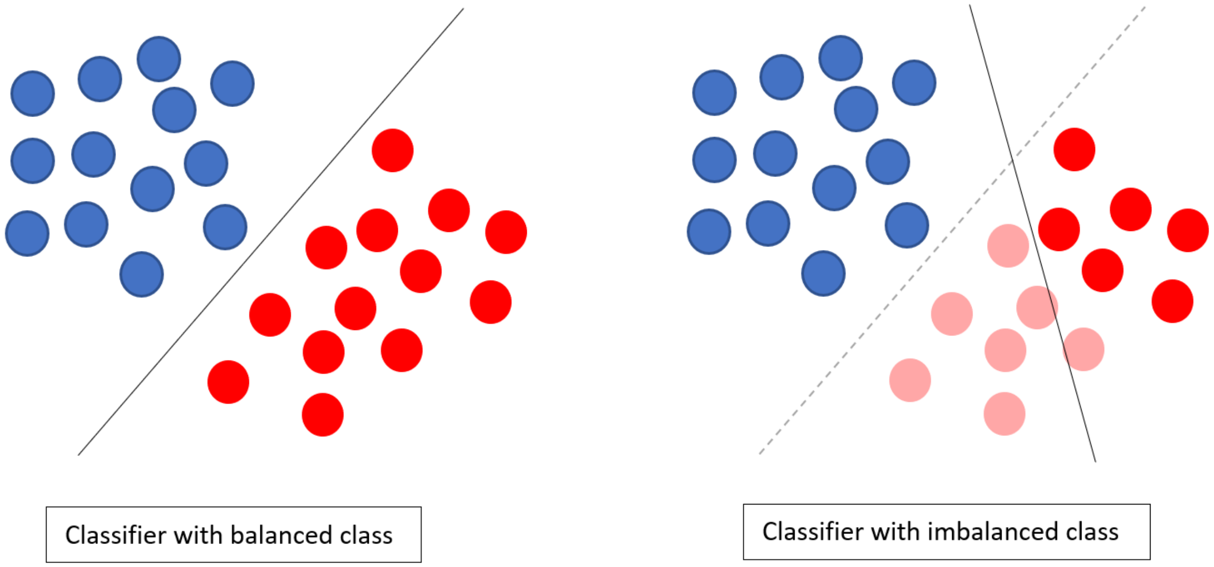
Deep learning – class imbalance tackling

Data level methods

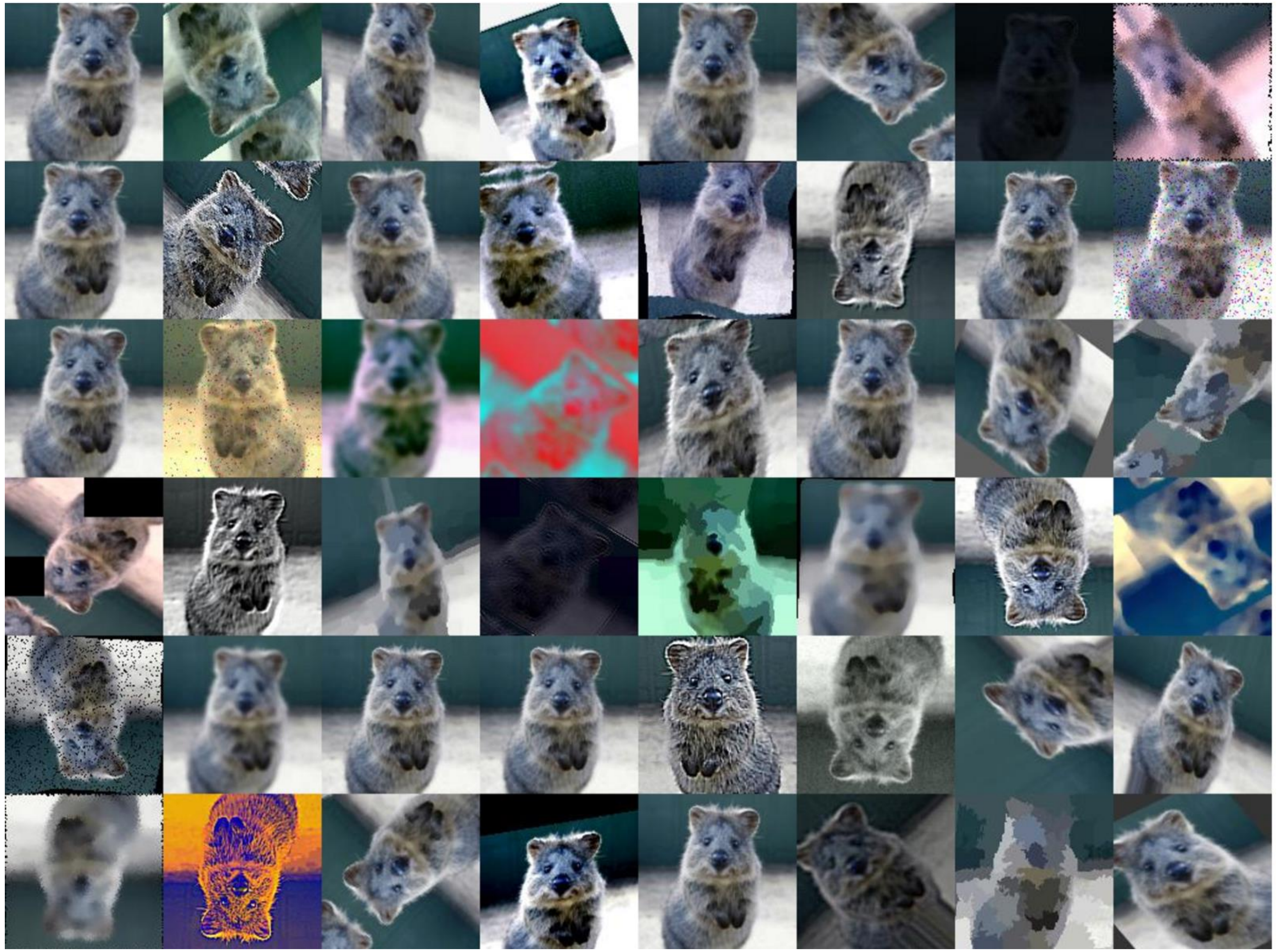
Oversampling
Undersampling

Classifier level methods

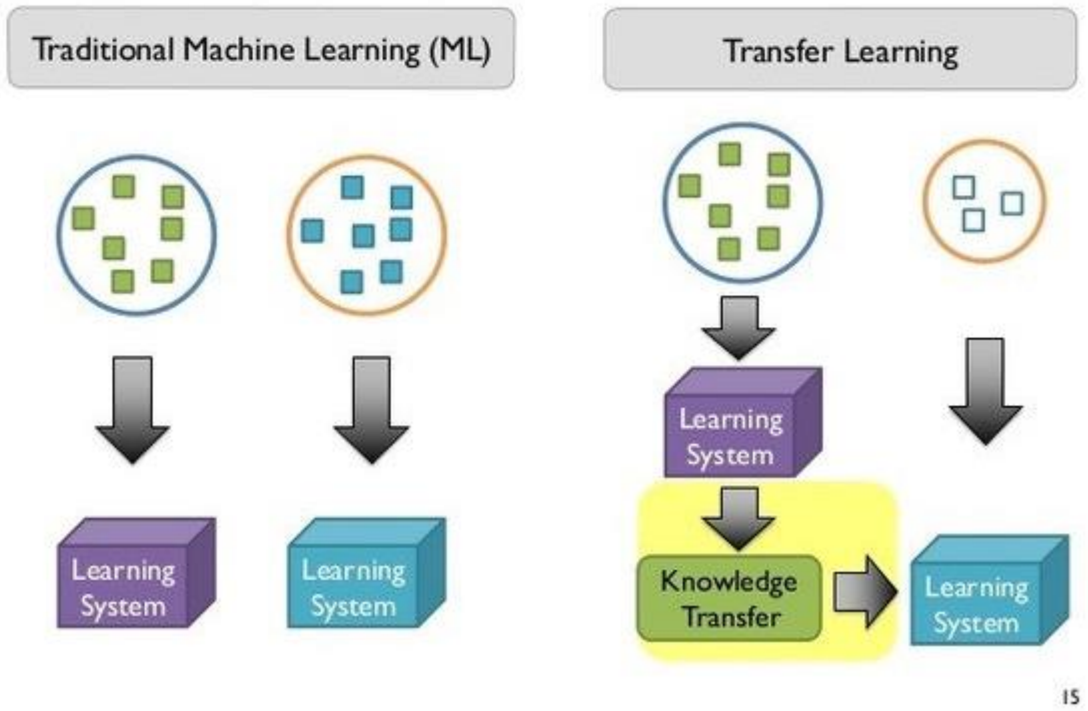
Thresholding
Cost sensitive learning
One-class classification
Hybrid of methods



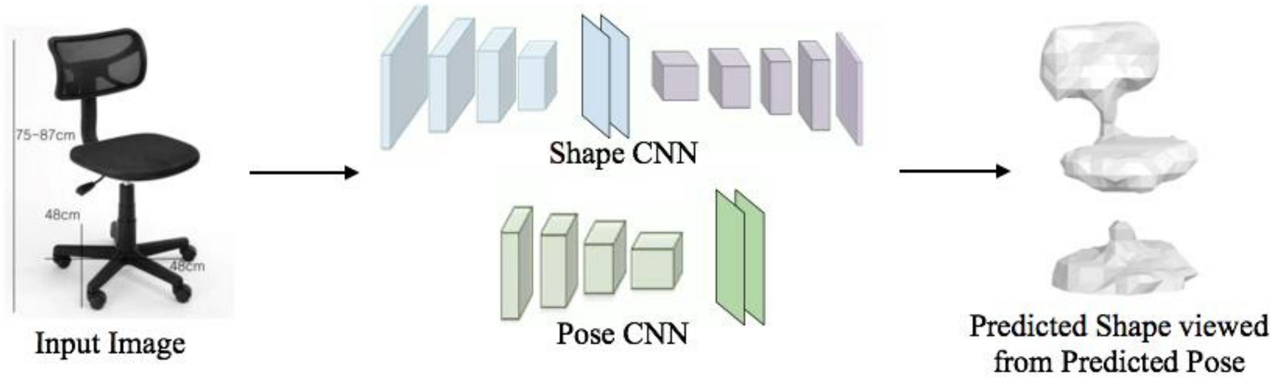
Augmentation



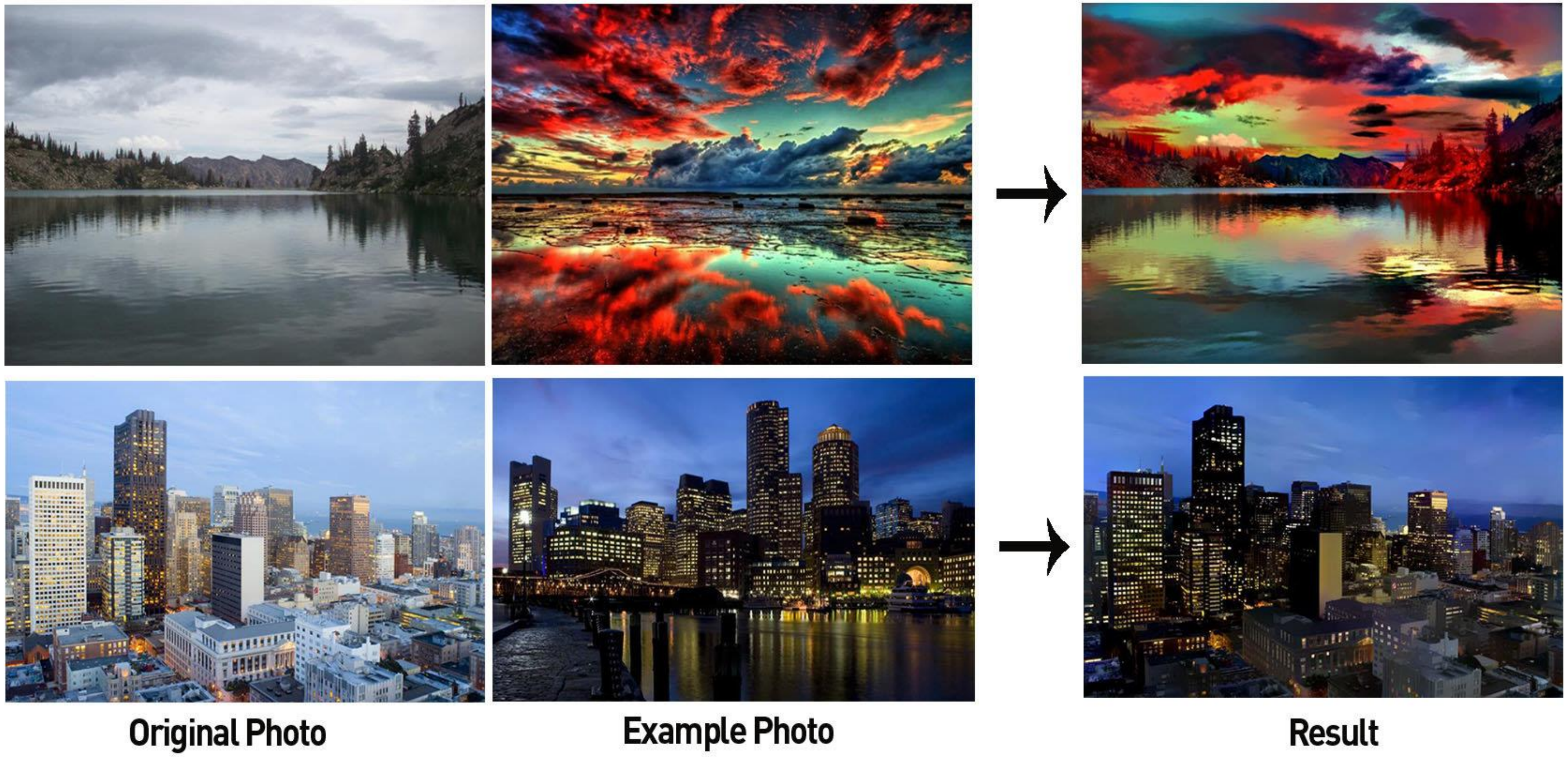
Transfer learning



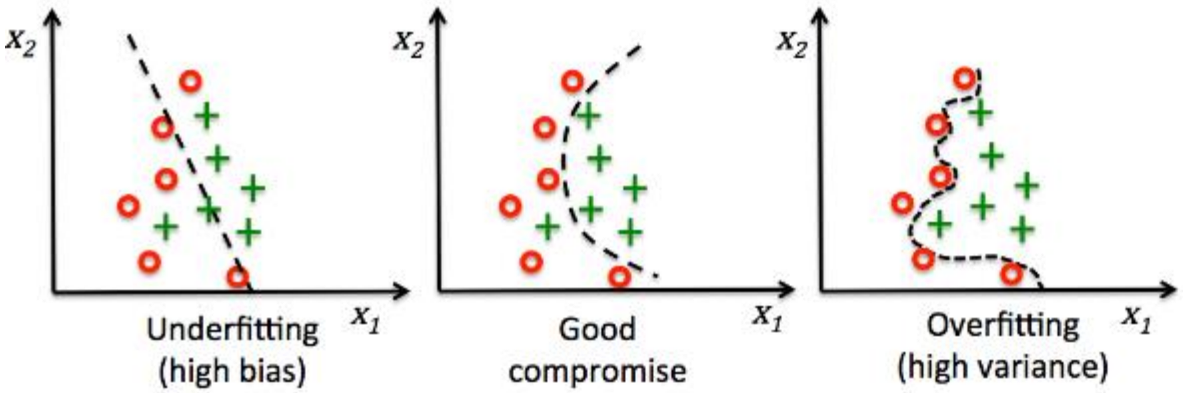
ShapeNet for 3D object

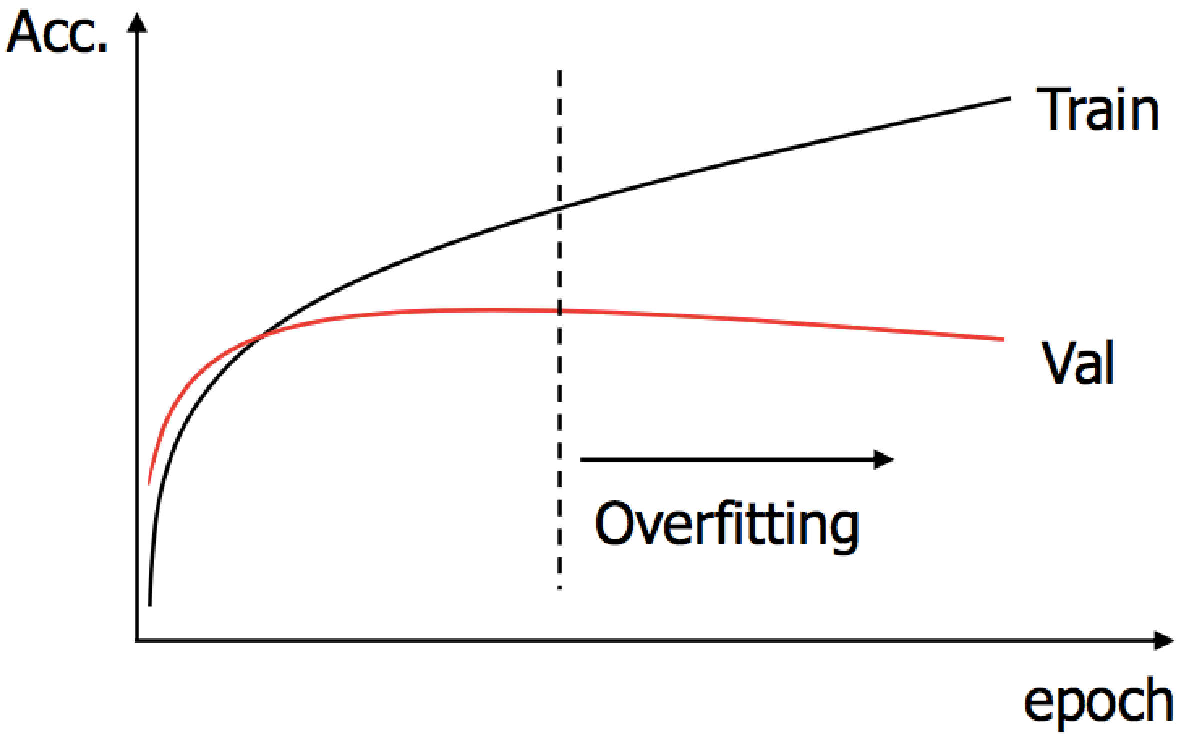


Style transfer – deep learning



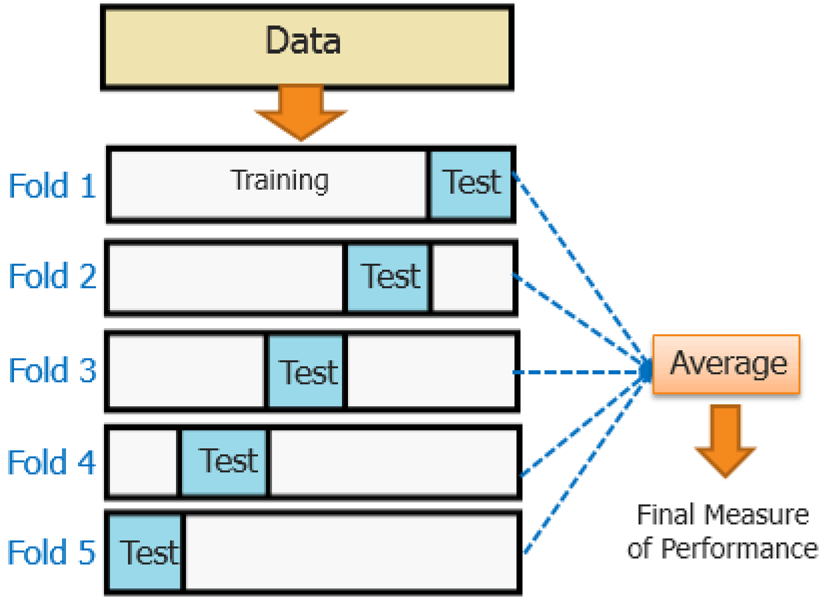
Overfitting





Validation

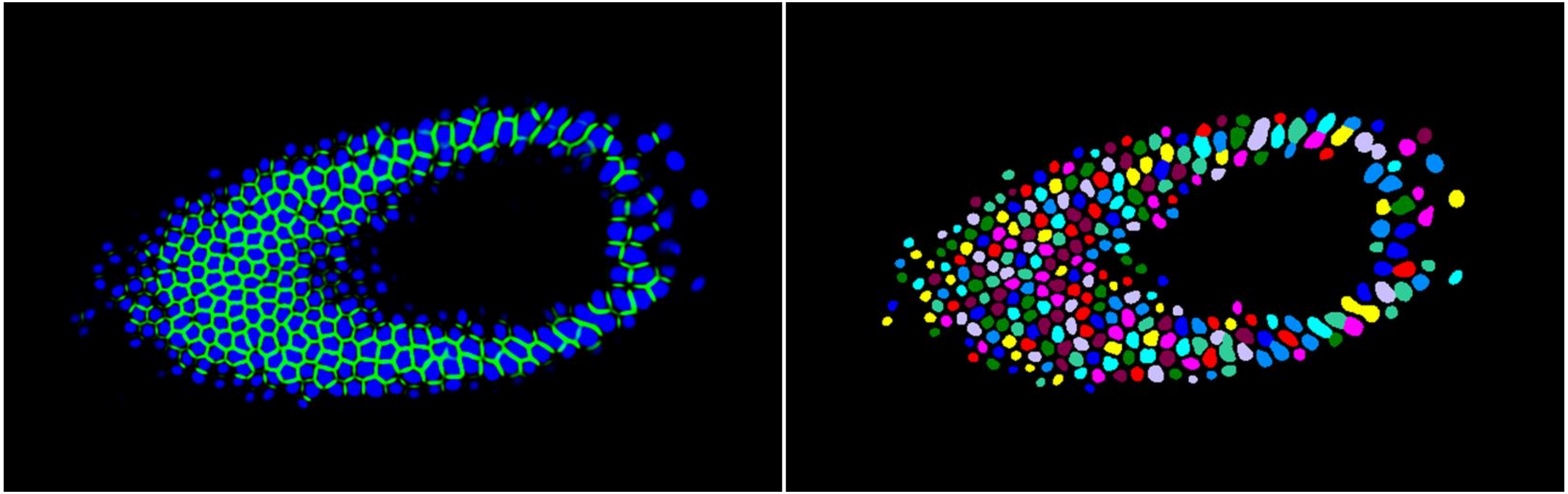
- K-crossfold validation
- Leave P out cross validation
- Leave one out cross validation



Smart applications

- https://deepart.io
- https://quickdraw.withgoogle.com/
- https://distill.pub/2016/handwriting/
- https://paperswithcode.com/task/semantic-segmentation

Assignment



Practicalities

- This week's plan
- Group topics
- Assignment expectations
- 5 image discussion
- Dateline