
Computer vision

Computer Vision

Lecture 9: Image Segmentation

Dr. Dina Khattab

dina.khattab@cis.asu.edu.eg

Scientific Computing Department

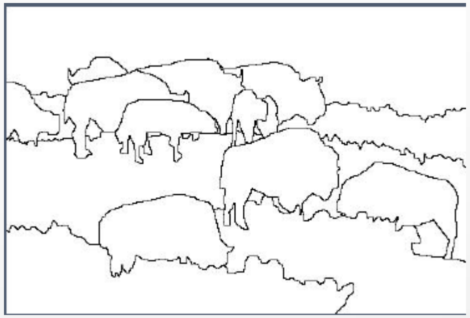
Instructor:	Dr. Dina Khattab
Email:	dina.khattab@cis.asu.edu.eg
Office:	Main Building – 4th floor – Room 302
Office Hours:	Monday 12:00 AM to 1:00 PM Thursday 11:00 AM to 12:00 PM

Agenda

- Segmentation By Clustering
- Semantic Segmentation
- Instance Segmentation

Image Segmentation

Partitioning of an image into the set of regions, which represent meaningful areas of the image.



- Separate the foreground regions (object) from the background regions which are ignored.

Image Segmentation

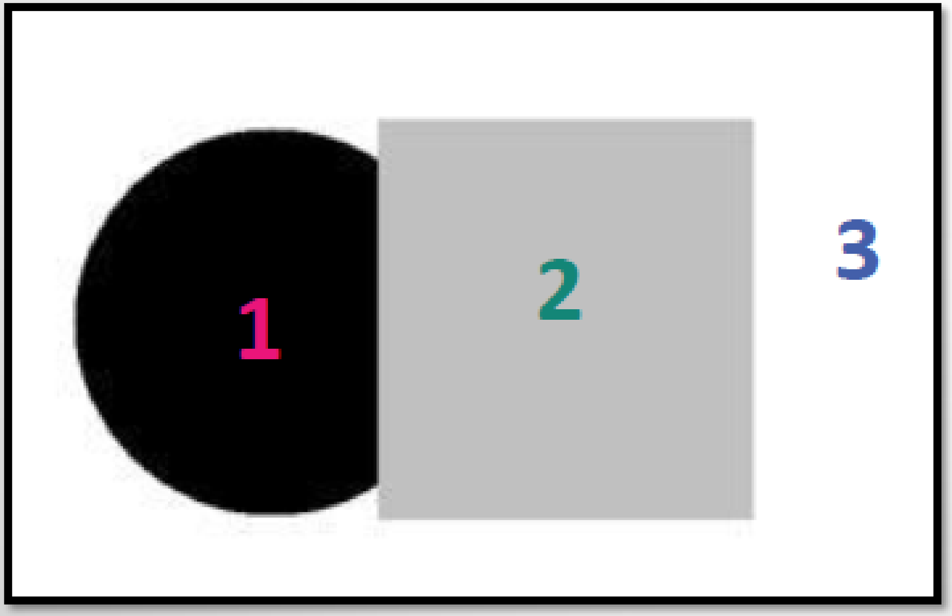
- Segmentation have two main objectives:
  - Decompose the image into parts for further analysis.
  - Perform change of representation.

- Regions of image segmentation should be uniform and homogenous with respect to some characteristics such as gray level, color or texture.

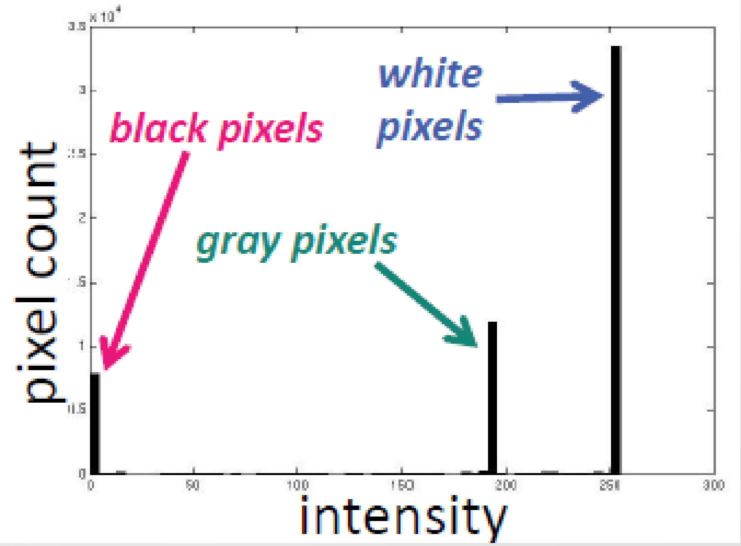
- The regions that humans see as homogenous may not be homogenous in terms of low-level features available to the segmentation system, so higher-level knowledge may have to be used.

Segmentation as clustering

Image segmentation: Toy example

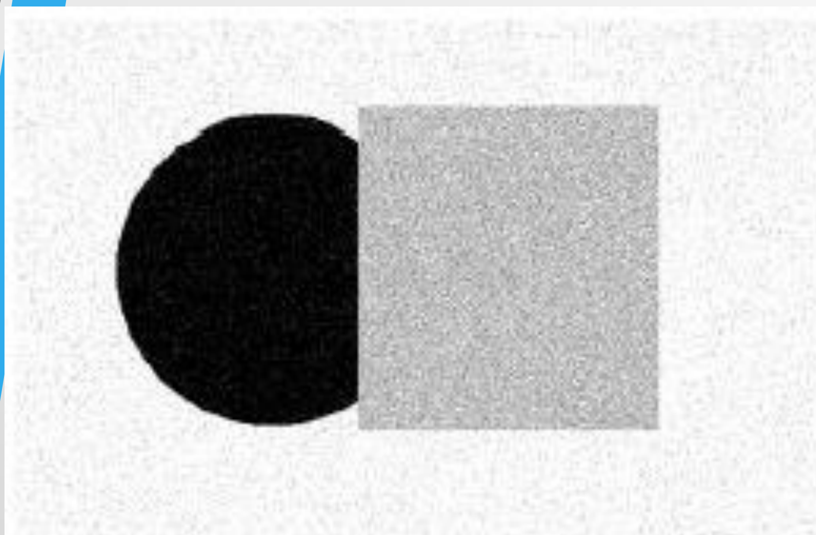


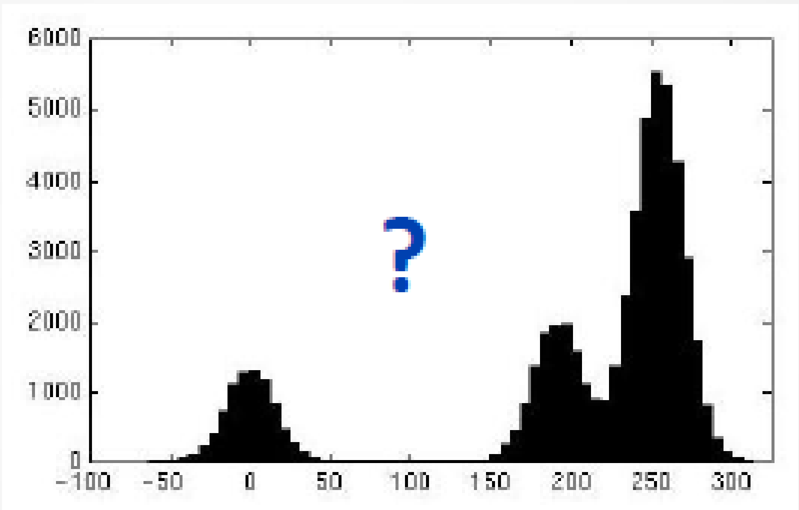
Input image



Intensity histogram

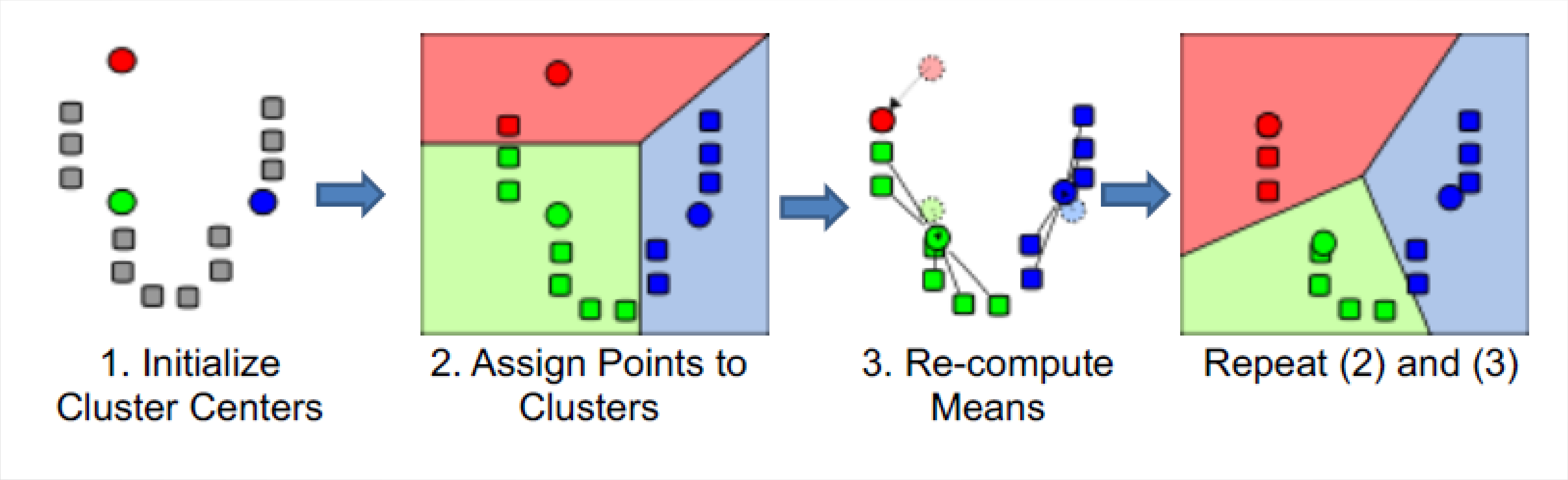
Noisy Images





- How to determine the three main intensities that define our groups?
- We need to *cluster*.

K-means clustering

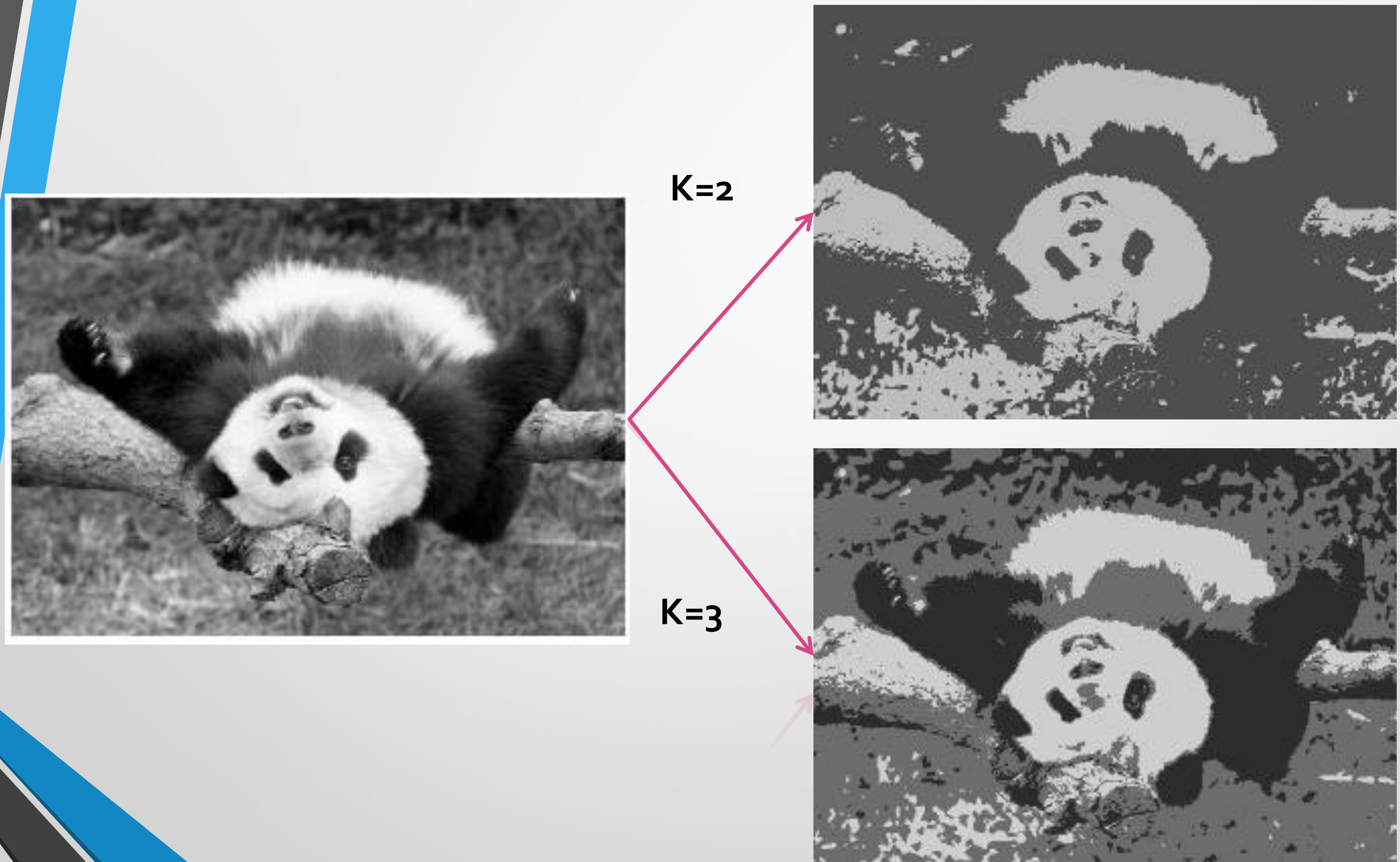


Segmentation as clustering

Depending on what we choose as the *feature space*, we can group pixels in different ways.

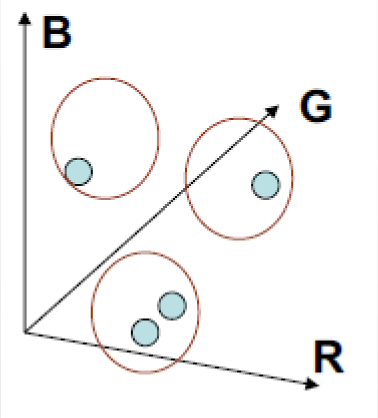
Feature space:
intensity value (1-d)

Number of Clusters

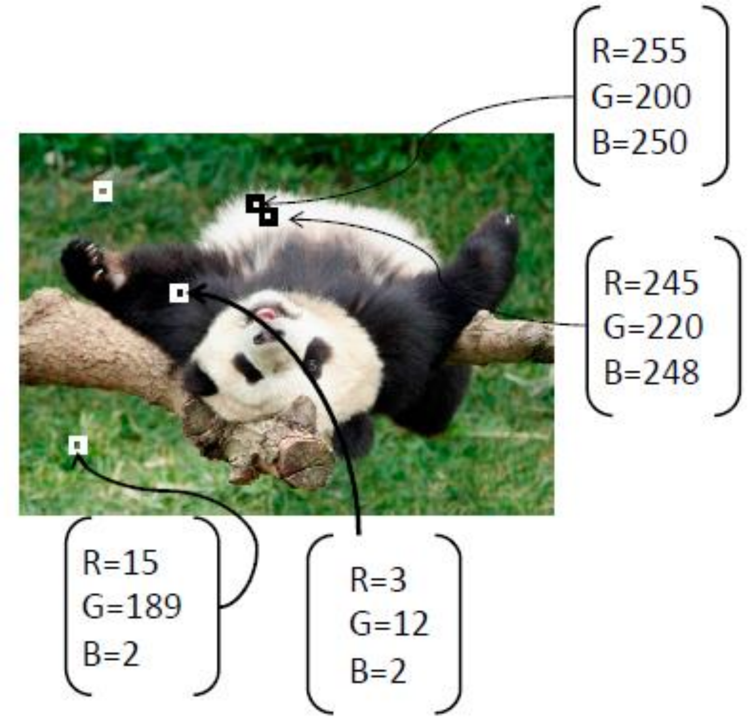


Segmentation as clustering

- Depending on what we choose as the *feature space*, we can group pixels in different ways.

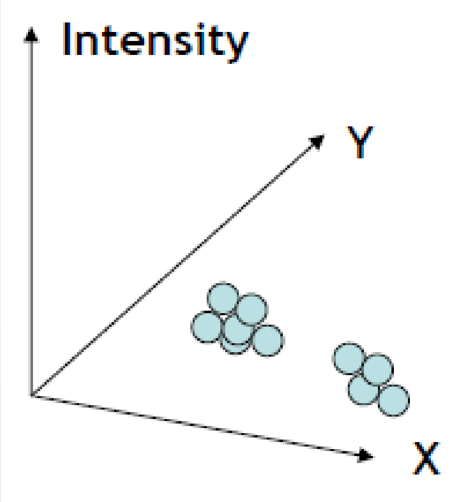


Feature space:
color value (3-d)

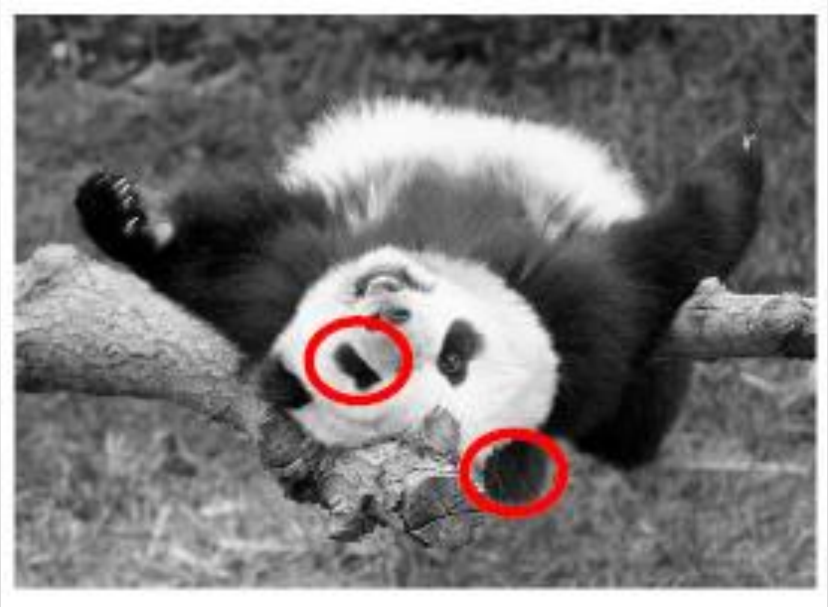


Segmentation as clustering

- Depending on what we choose as the *feature space*, we can group pixels in different ways.

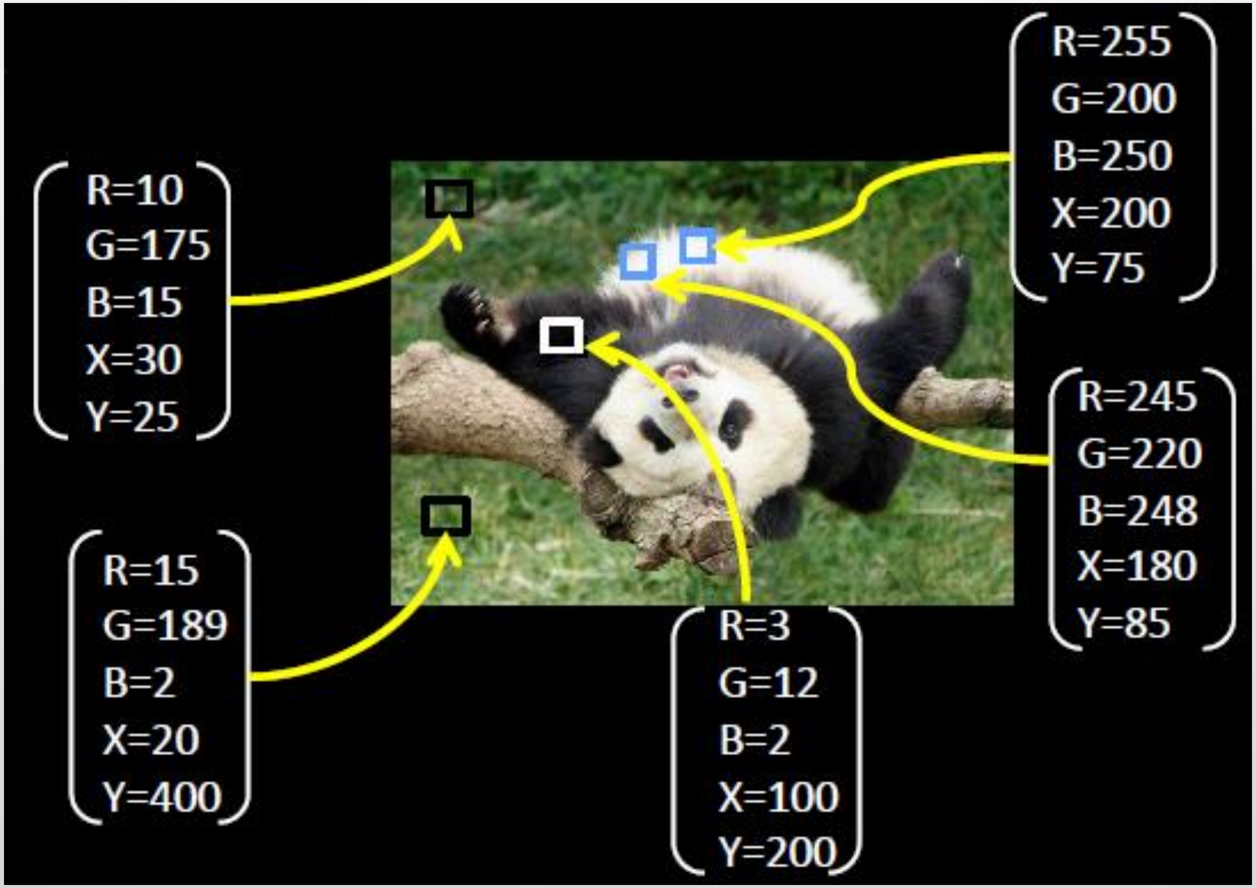


Feature space:
Intensity + position



Segmentation as clustering

- Can combine color and location…

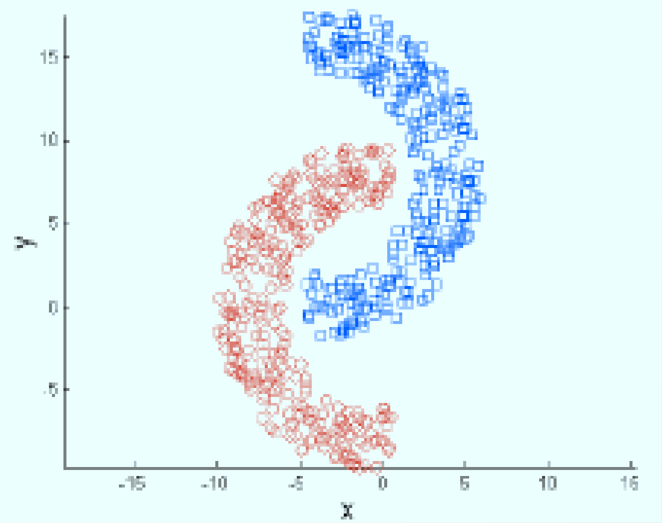


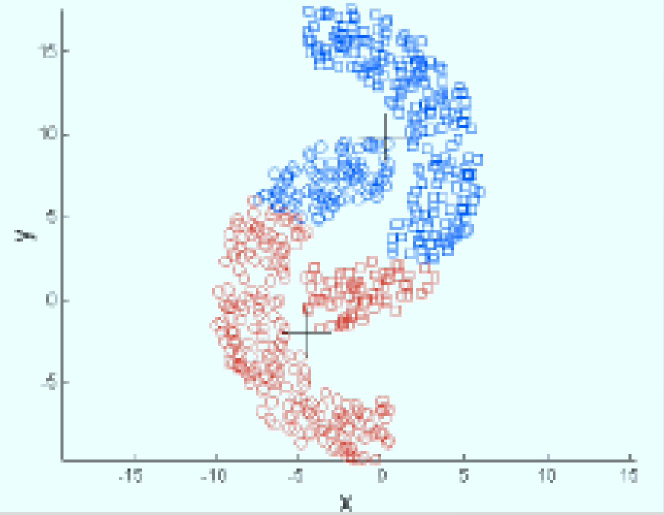
K-Means for segmentation

- Pros
  - ✓ Very simple method
  - ✓ Converges to a local minimum of the error function

K-Means for segmentation

- Cons
  - ✓ Memory-intensive
  - ✓ Need to pick K
  - ✓ Sensitive to initialization
  - ✓ Sensitive to outliers
  - ✓ Only finds “spherical” clusters

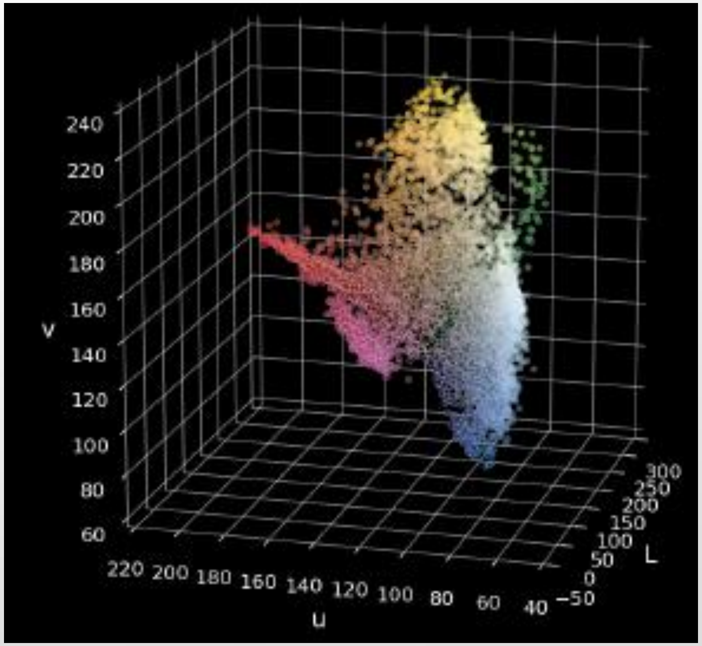




Mean shift algorithm

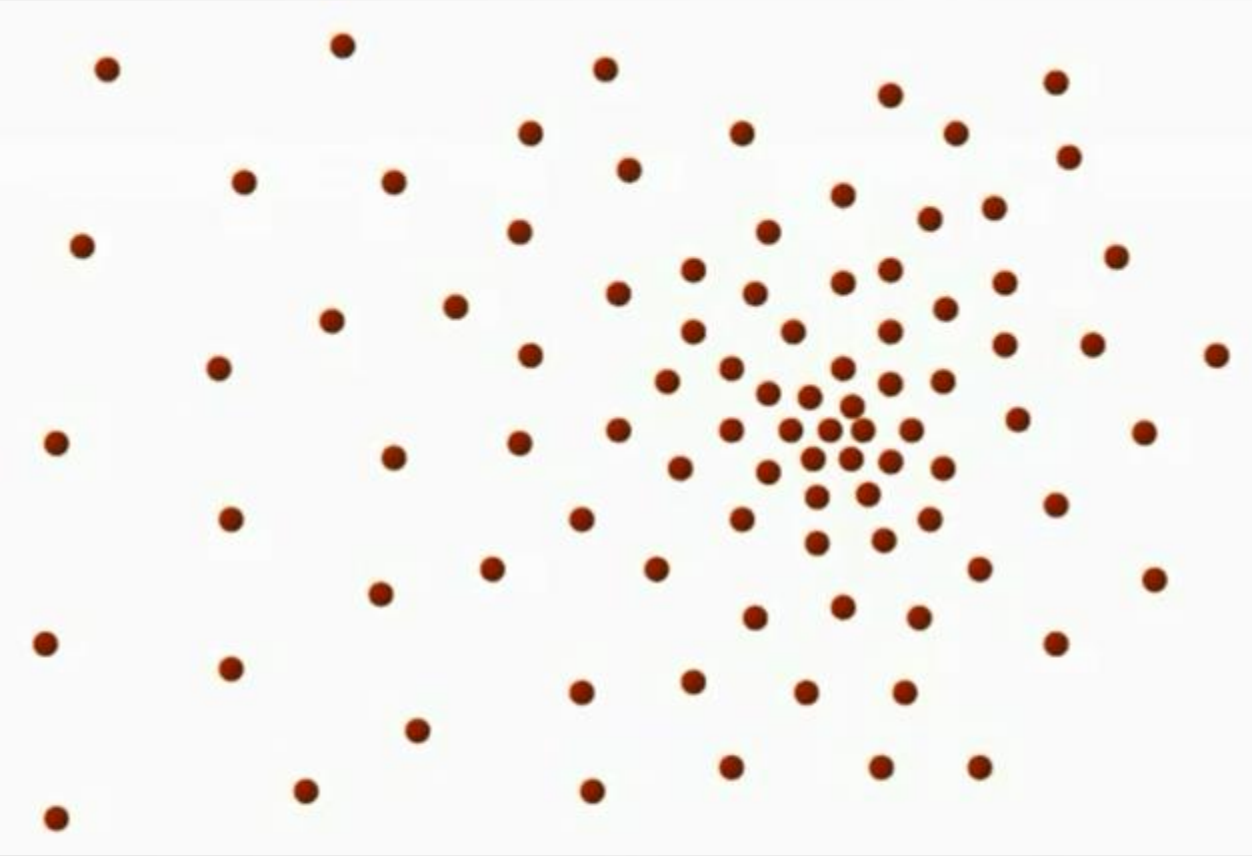
- The mean shift algorithm seeks *modes* or local maxima of density in the feature space

Input image

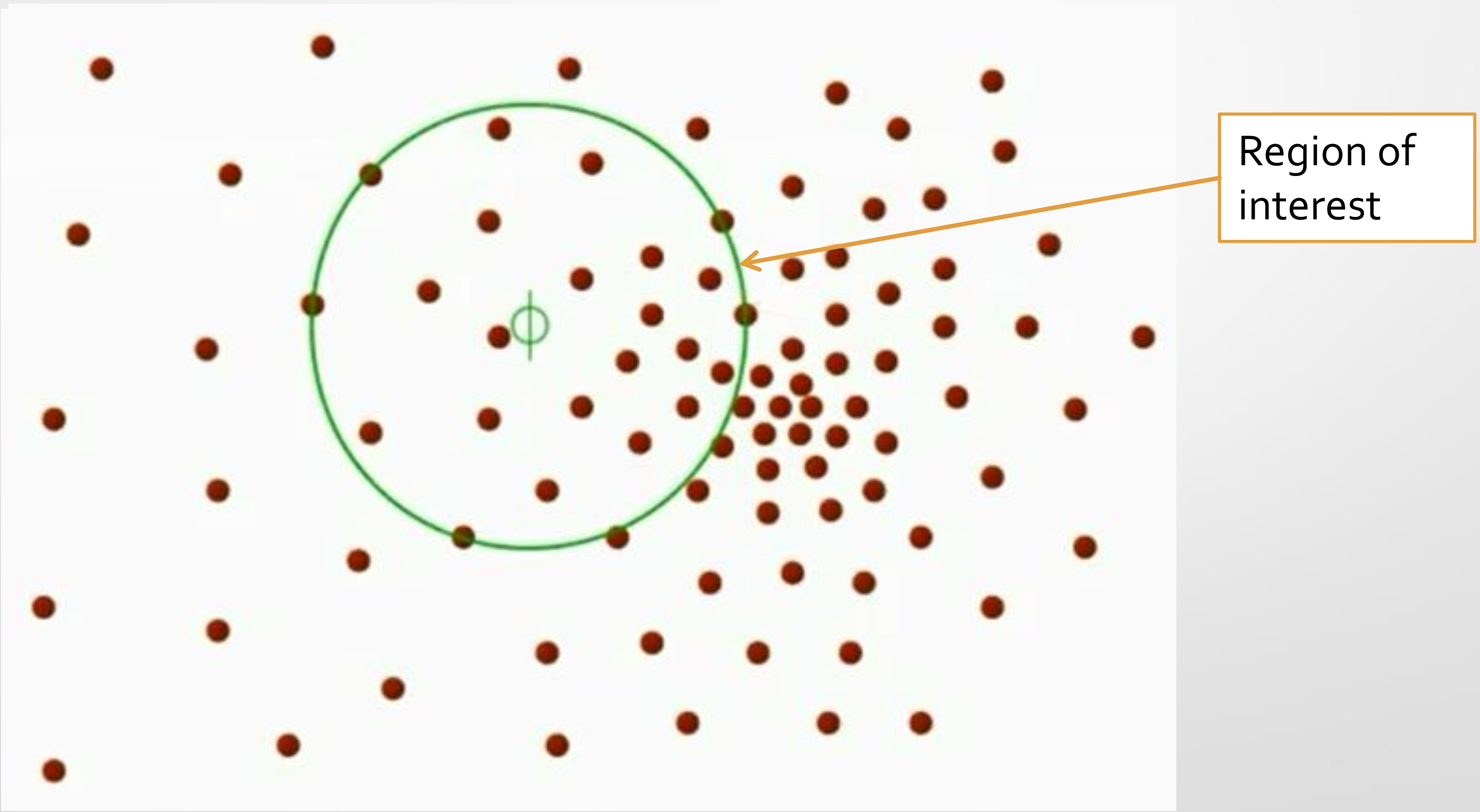


Feature space
(L*u*v* color values)

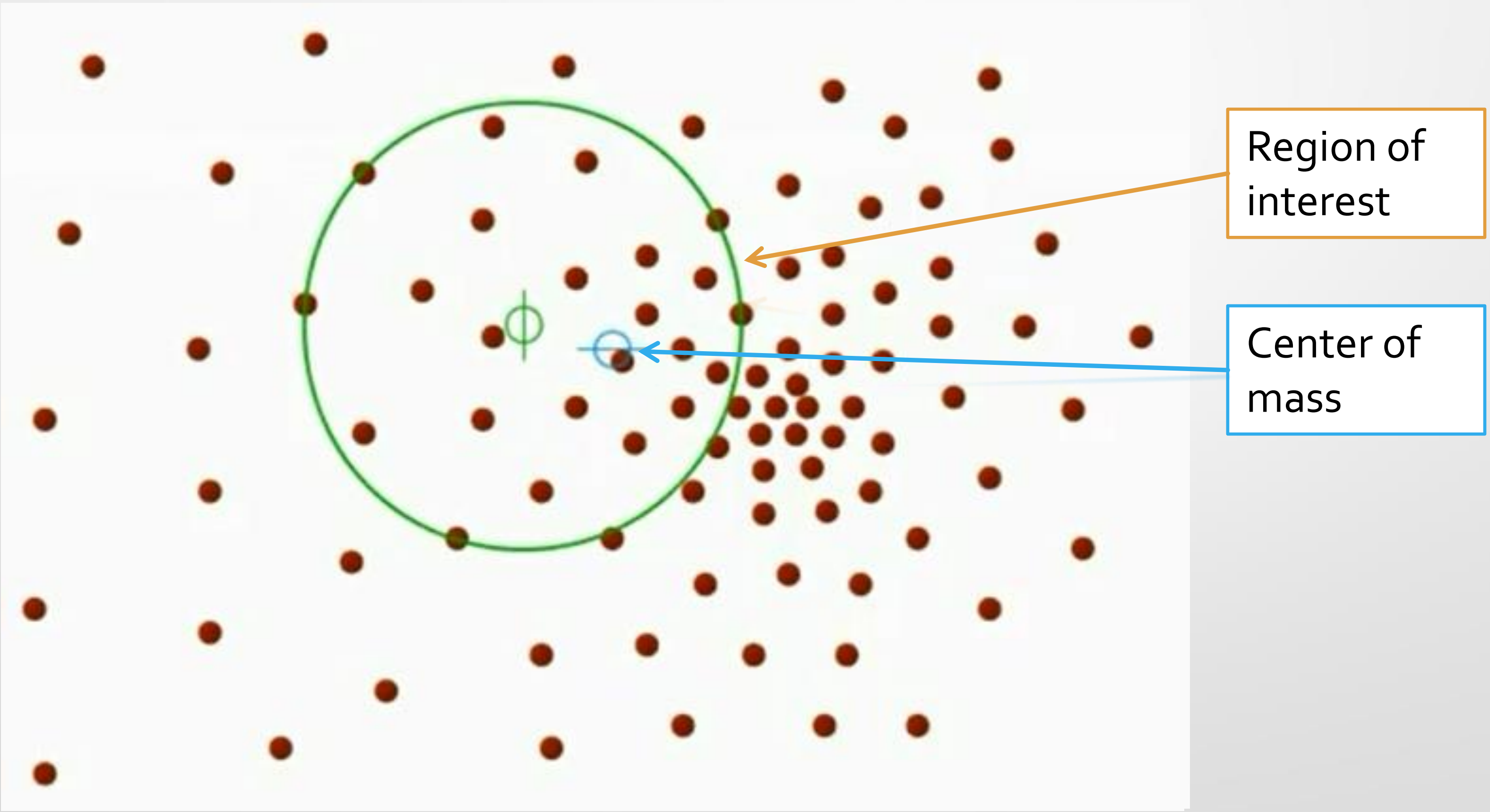
Mean Shift in space



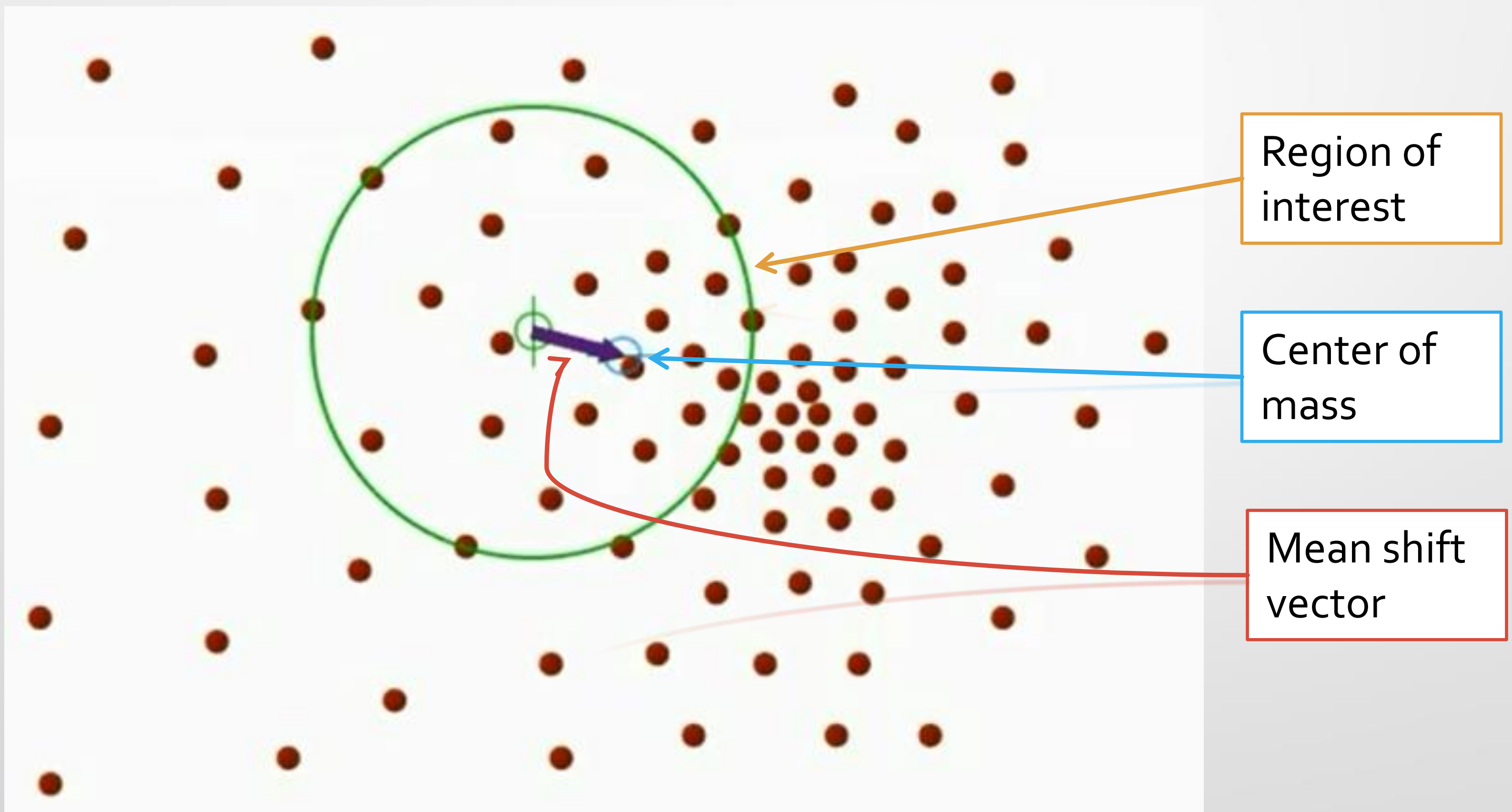
Mean Shift in space



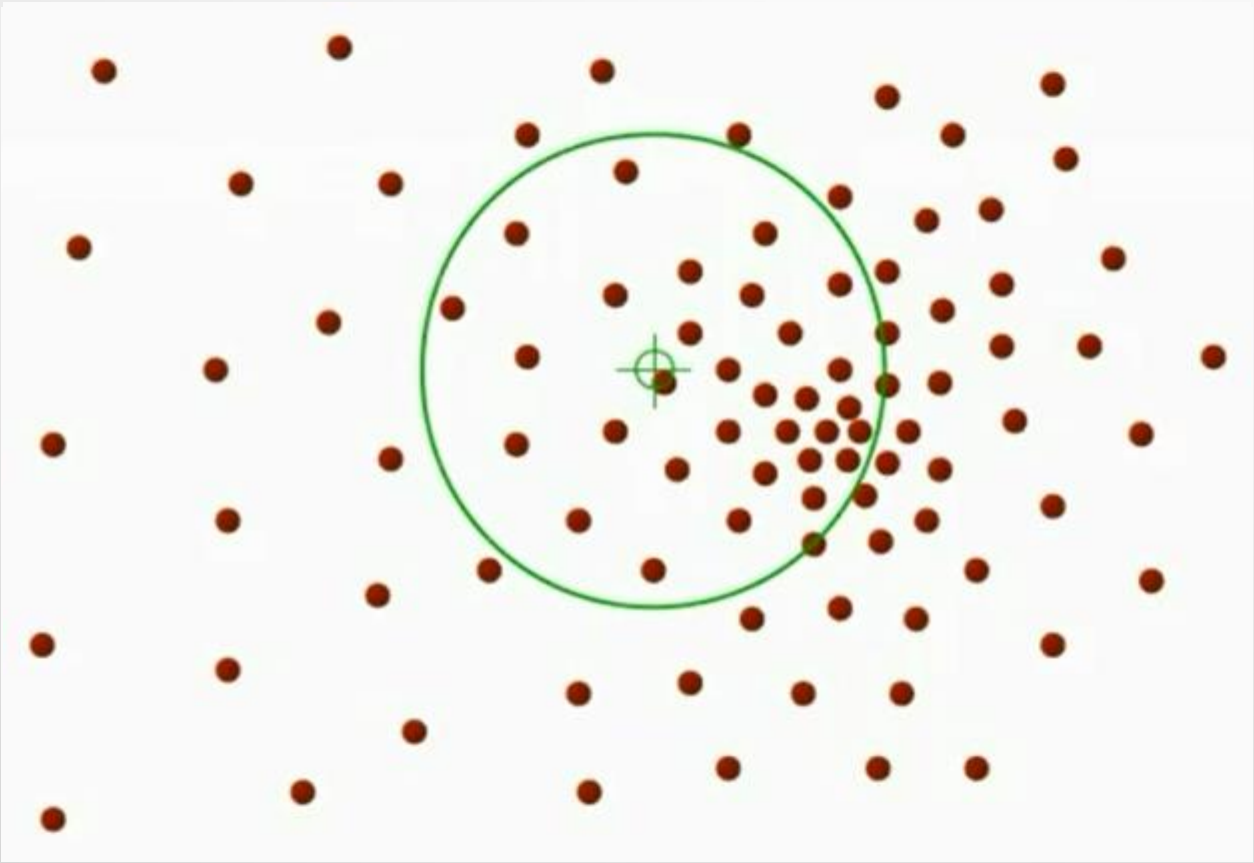
Mean Shift in space



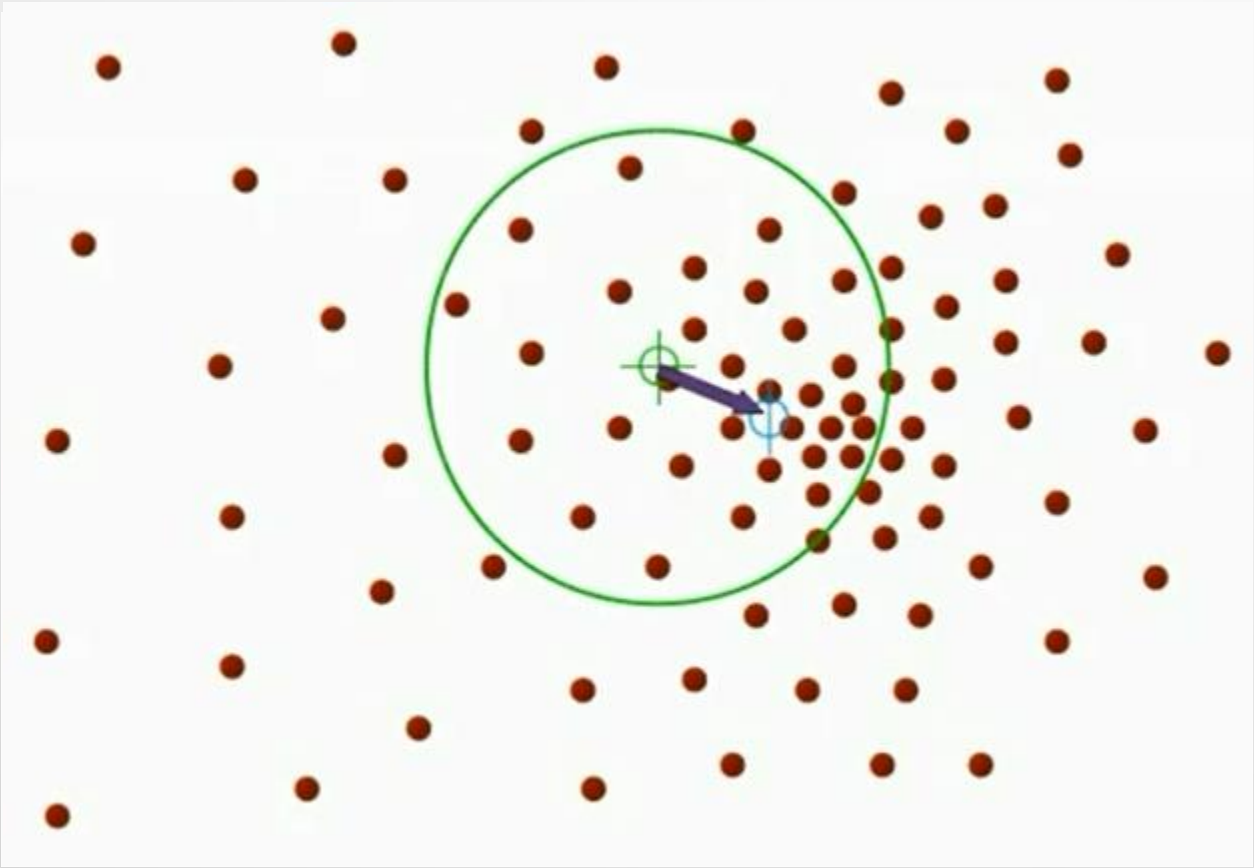
Mean Shift in space



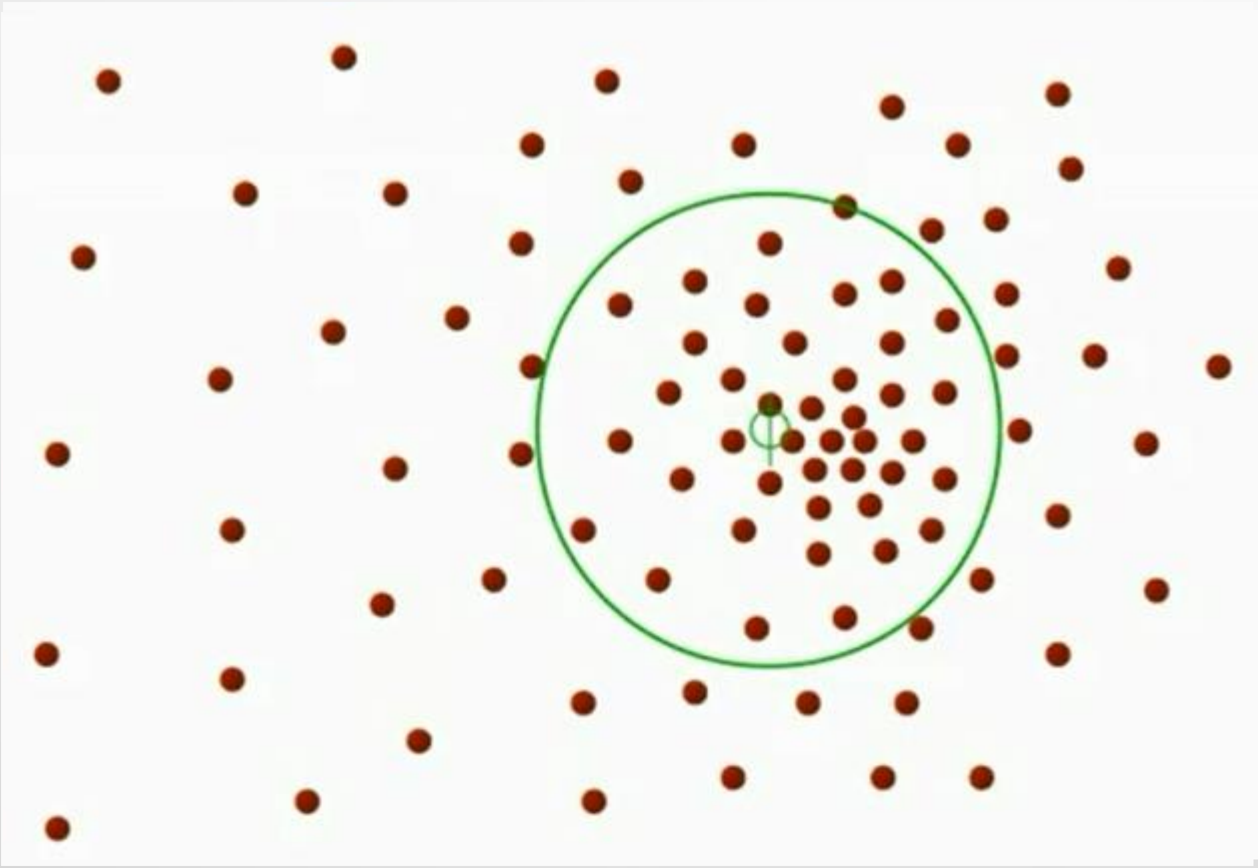
Mean Shift in space



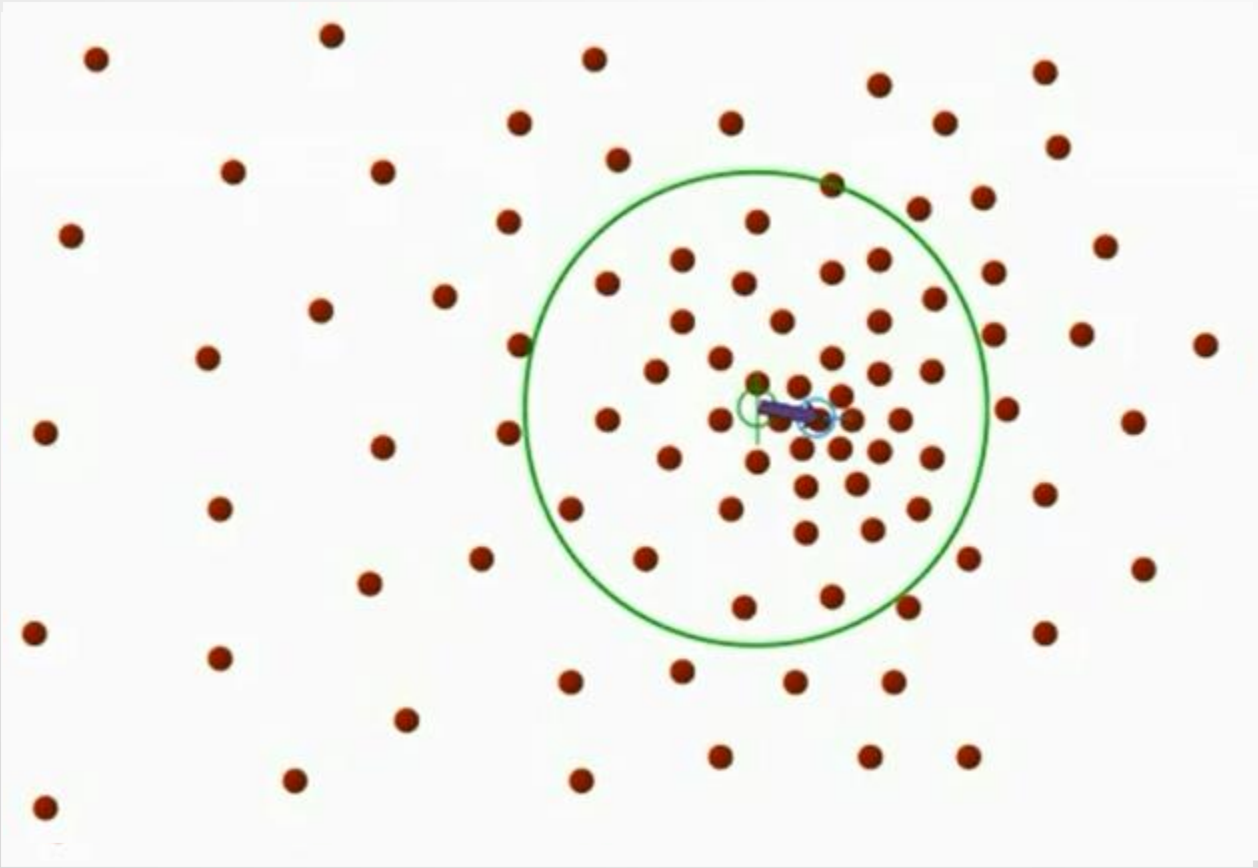
Mean Shift in space



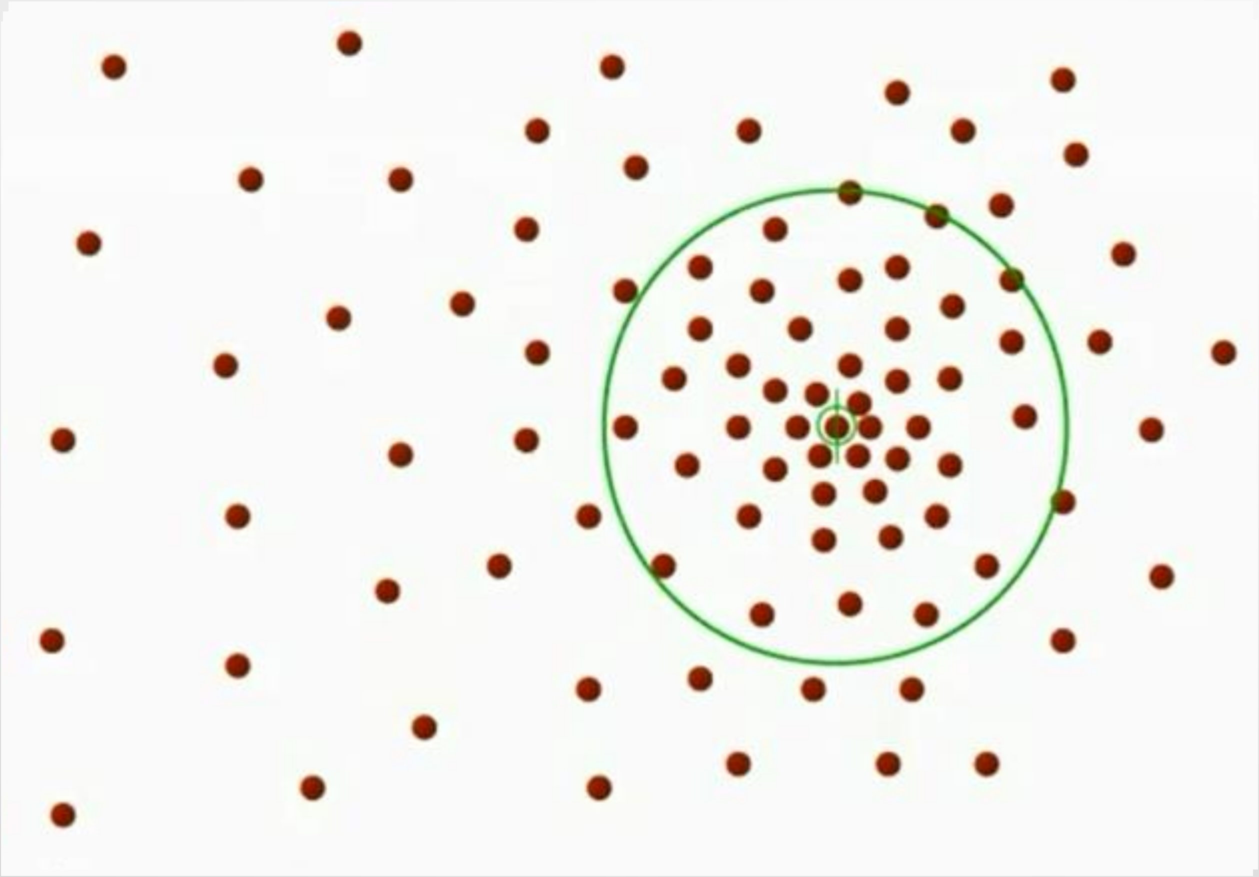
Mean Shift in space



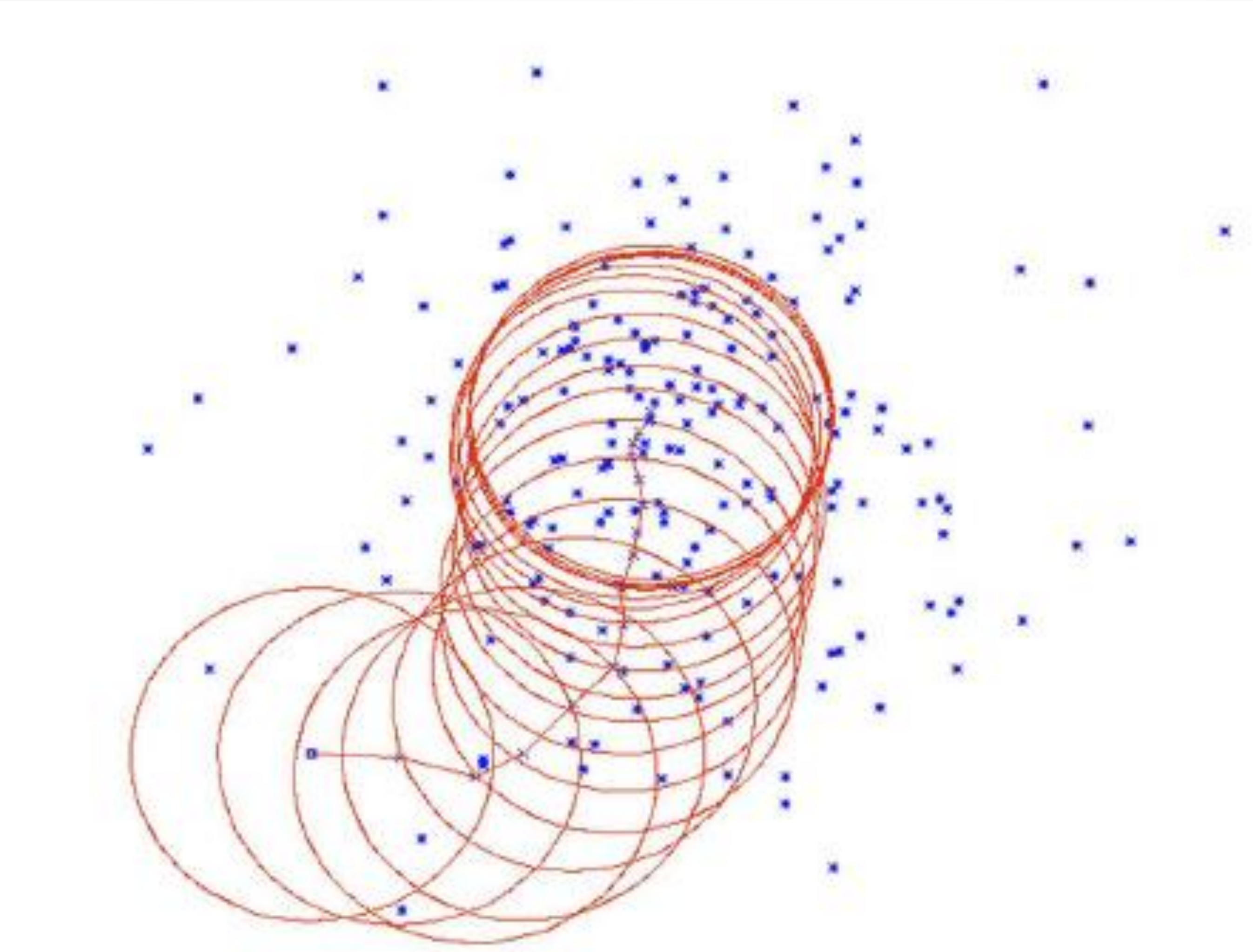
Mean Shift in space



Mean Shift in space

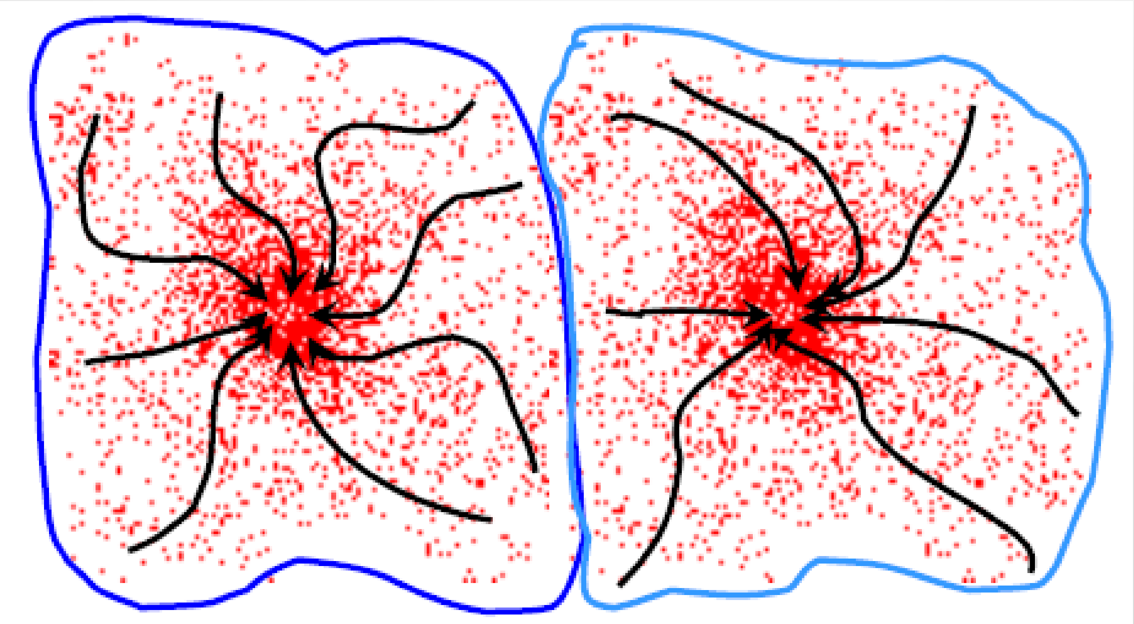


Mean Shift in space

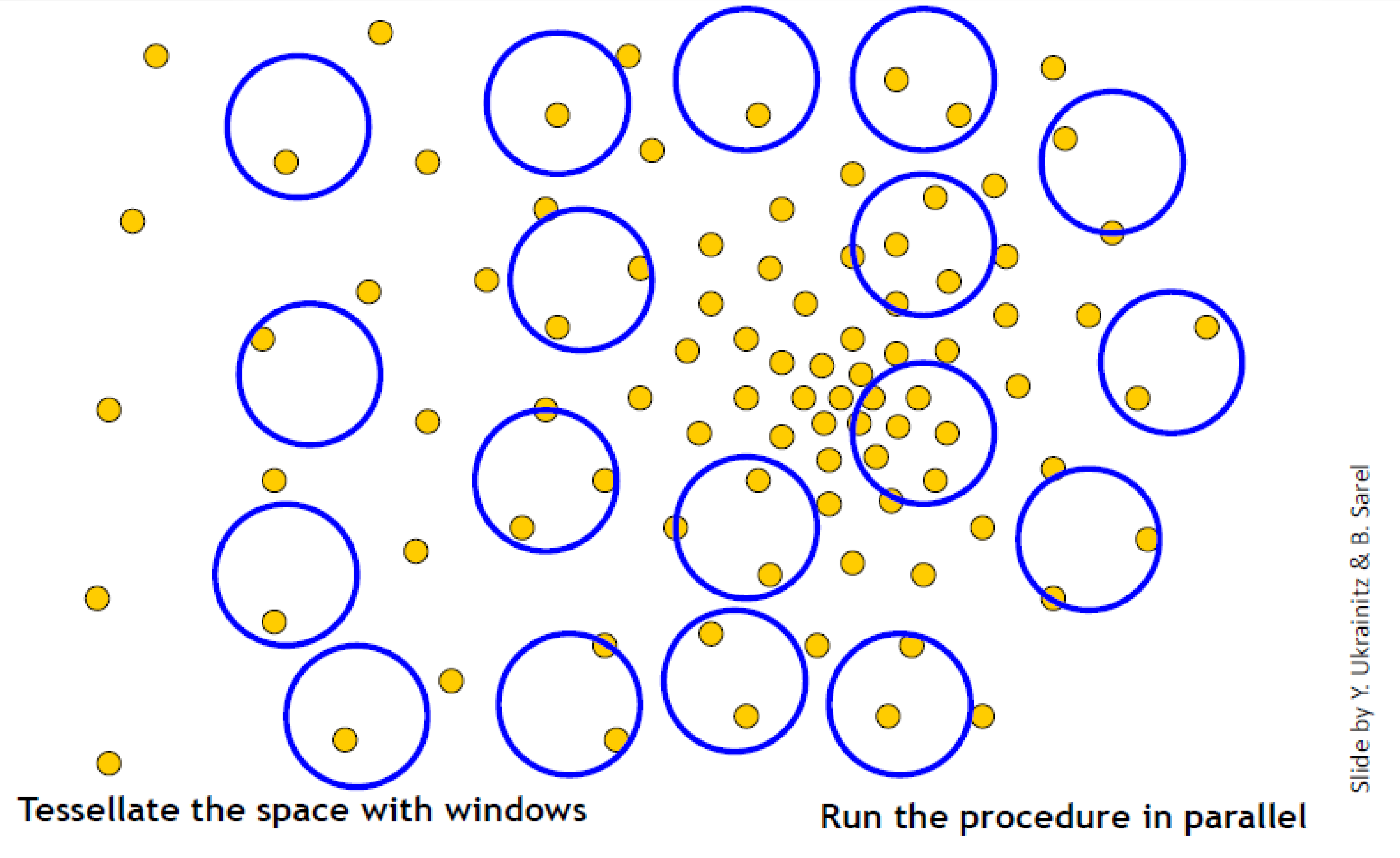


Mean shift clustering

- **Mean shift:** is a procedure for locating maxima of a density function given discrete data samples from that function.
- Cluster: all data points in the *attraction basin* of a mode
- Attraction basin: the region for which all trajectories lead to the same mode



Mean shift clustering



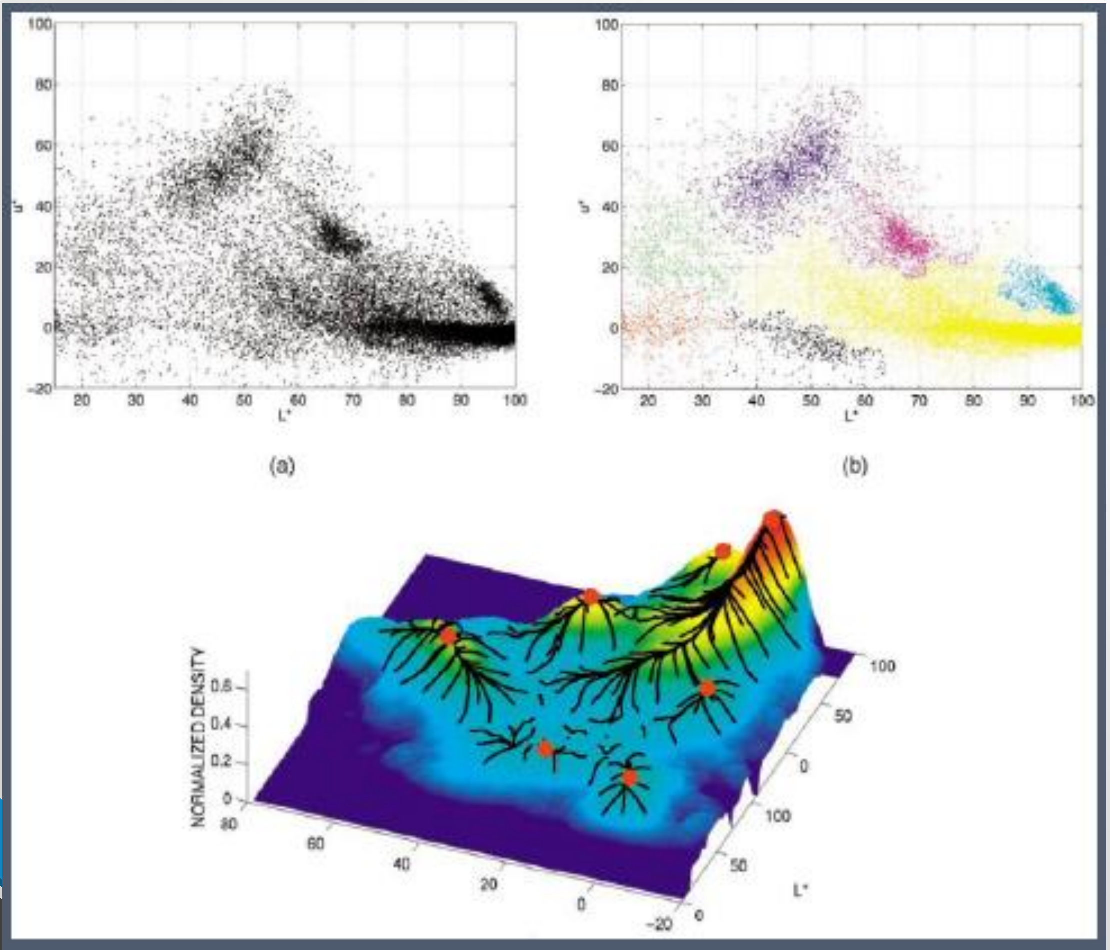
Mean Shift Algorithm

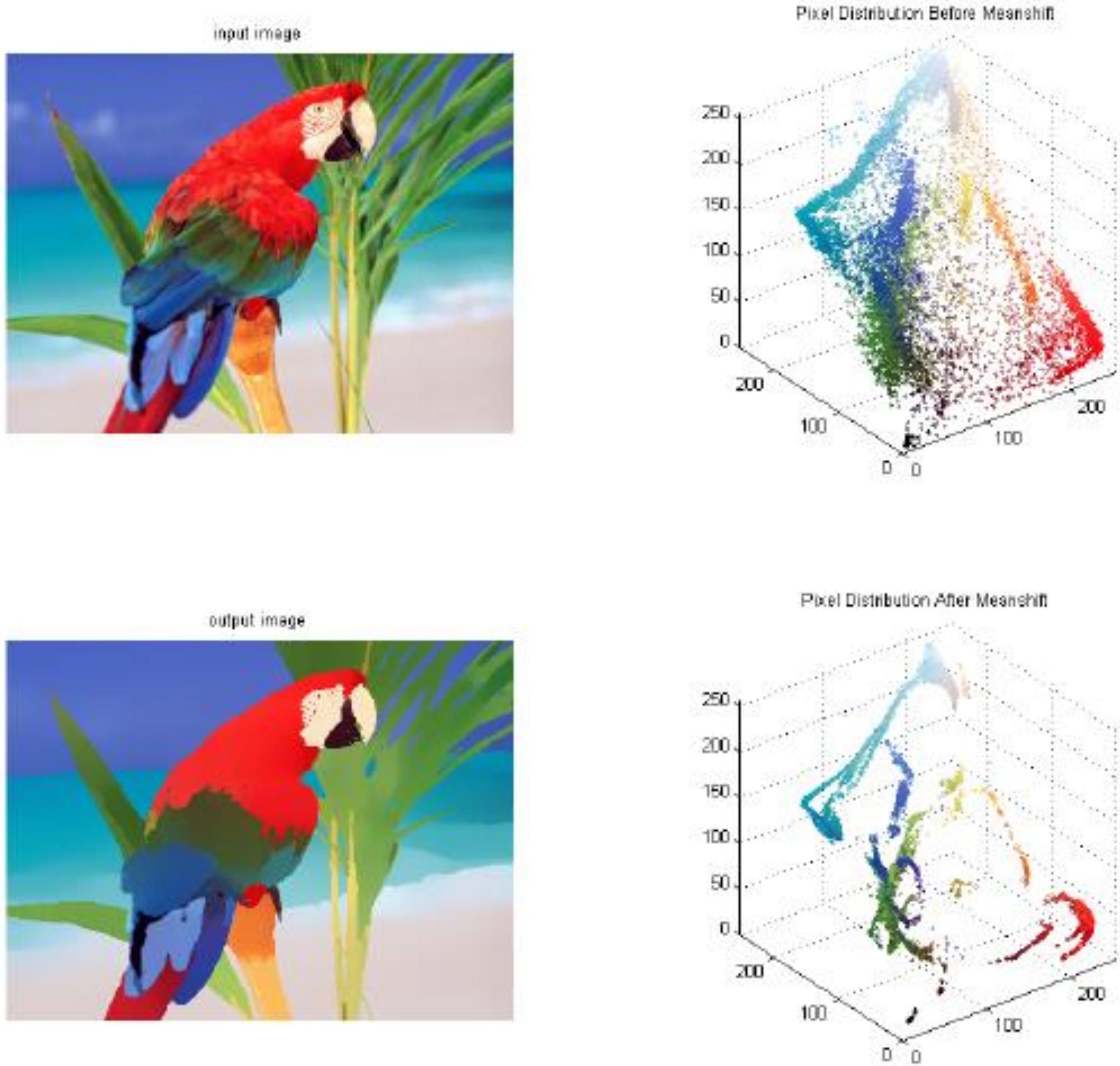
1. Choose a search window size.
2. Choose the initial location of the search window.
3. Compute the mean location (centroid of the data) in the search window.
4. Center the search window at the mean location computed in Step 3 (shift)
5. Repeat Steps 3 and 4 until convergence.

Mean Shift Clustering Algorithm

- Find features (color, gradients, texture, etc.).
- Initialize windows at individual feature points.
- Perform mean shift algorithm for each window until convergence.
- Merge windows that end up near the same "peak" or mode.

Mean shift finds 7 regions in the image



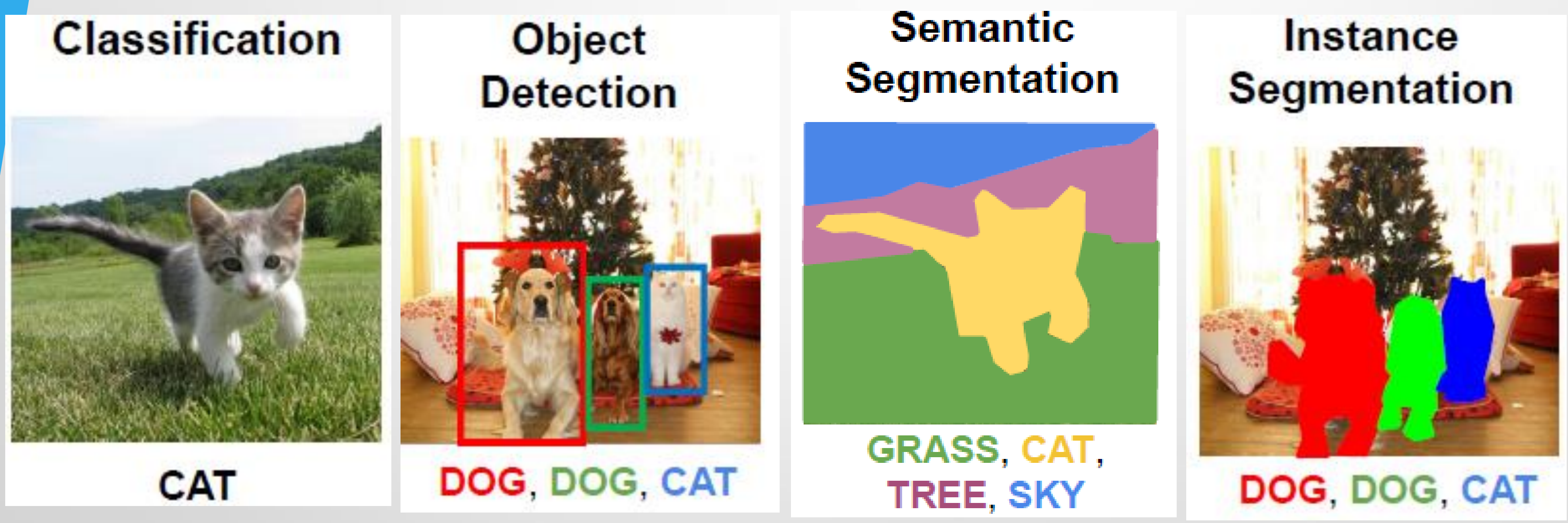

input image
Pixel Distribution Before Meanshift
250
200
150
100
50
0
200
100
0
0
100
200
output image
Pixel Distribution After Meanshift
250
200
150
100
50
0
200
100
0
0
100
200

Mean Shift

- Pros:
  - Automatically finds basins of attraction
  - One parameter choice (window size)
  - Robust to outliers
  - Generic technique
  - Find variable number of modes
- Cons:
  - Output depend on window size
  - Computationally expensive
  - Does not scale well with dimension of feature space

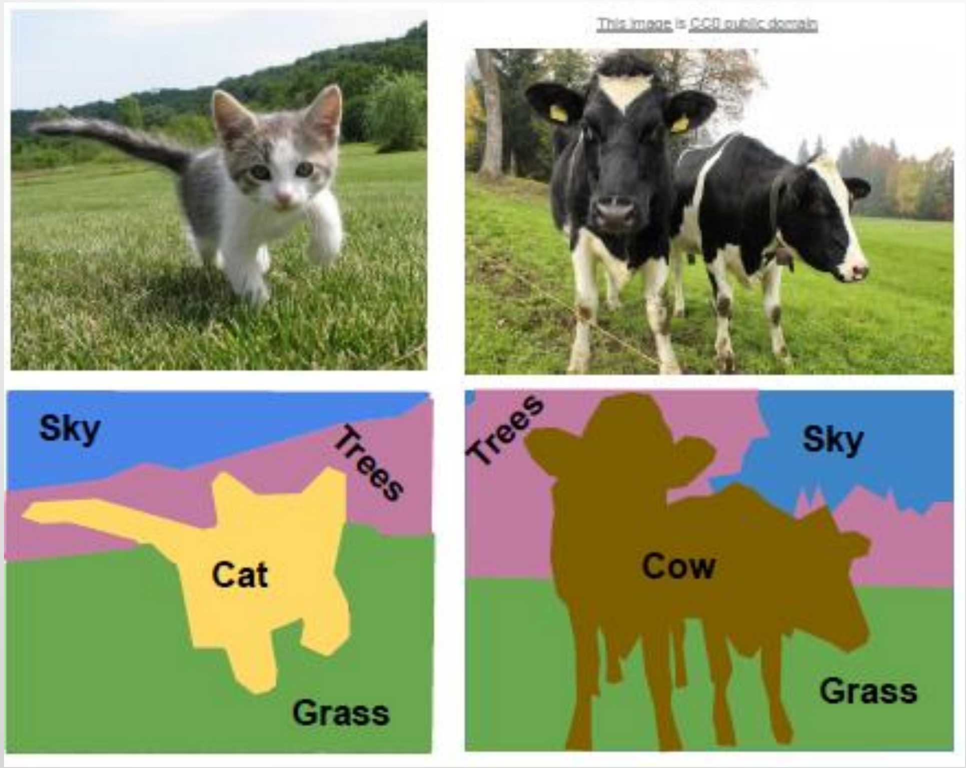
Semantic & Instance Segmentation

Semantic & Instance Segmentation



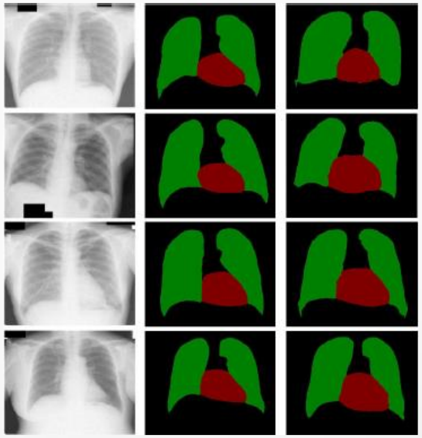
Semantic Segmentation

- Label each pixel in the image with a category label (pixel-level annotation)
- Don’t differentiate instances, only care about pixels



Semantic Segmentation Applications

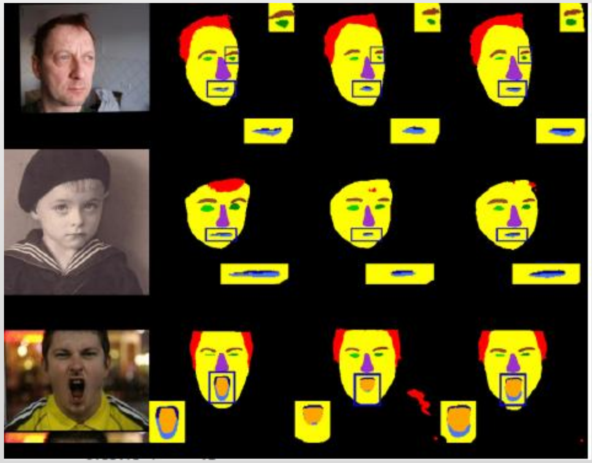
- A key part of Scene Understanding

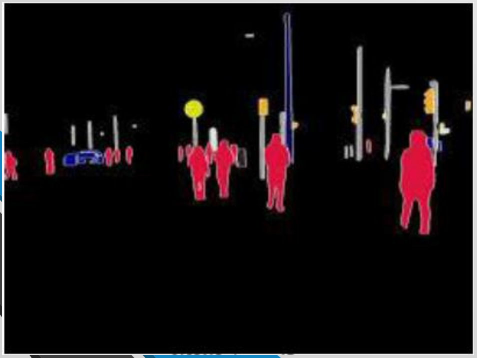


Medical diagnosis

Autonomous navigation

Image editing

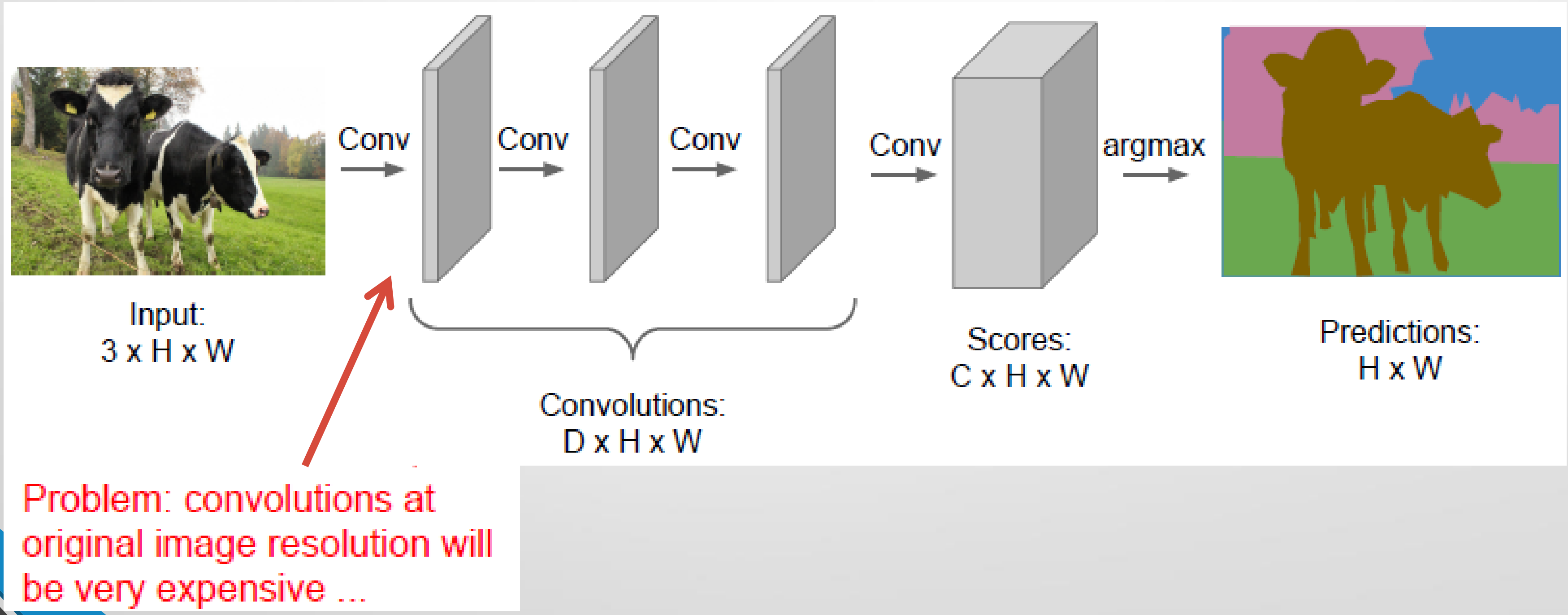




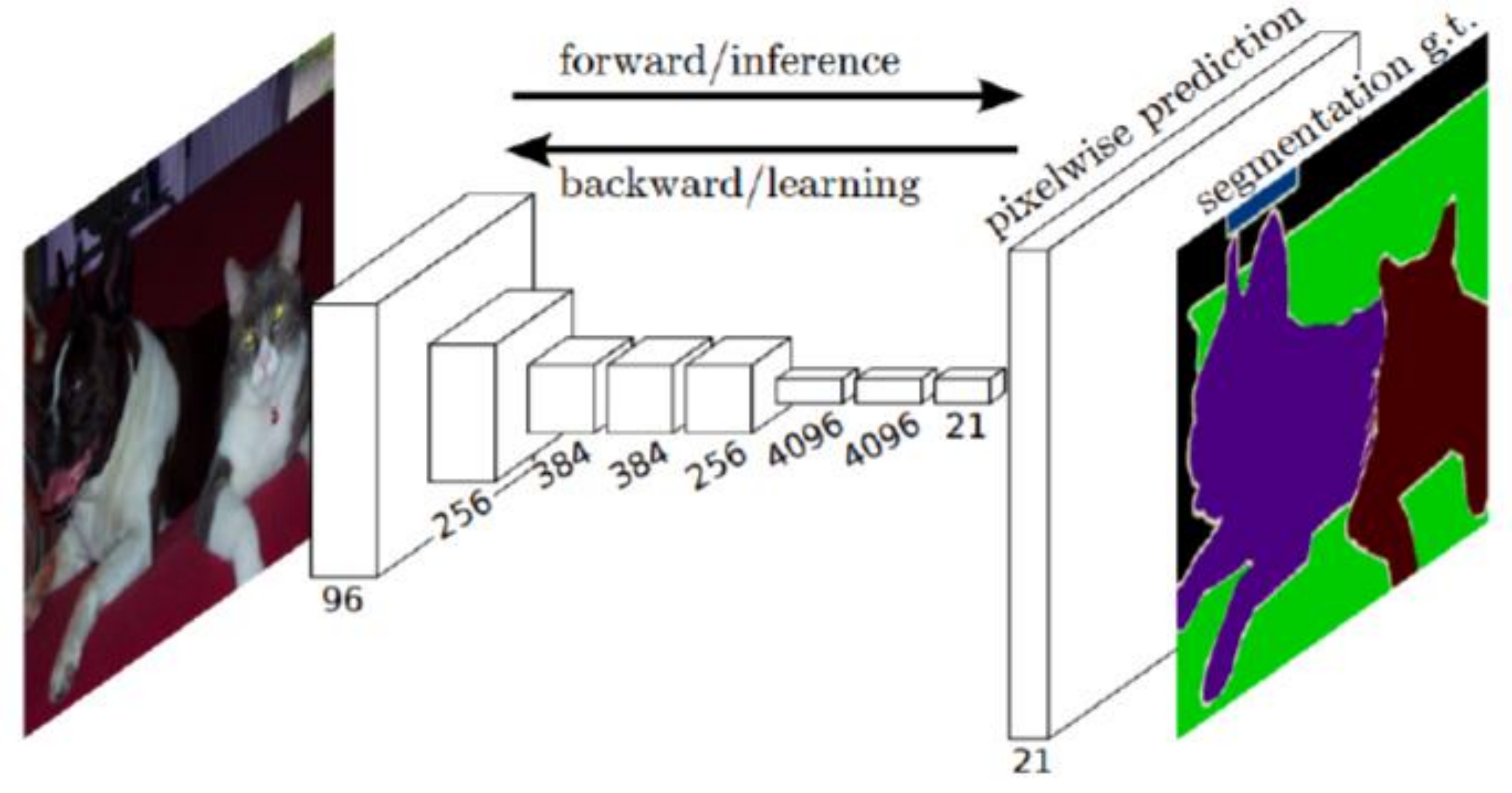
Assisting the partially sighted

Semantic Segmentation Idea

- Design a network as a bunch of convolutional layers to make predictions for pixels all at once!



Fully Convolutional Network (FCN)



Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015

In-Network upsampling: “Unpooling”

Nearest Neighbor

Input: 2 x 2

1	2
3	4

Output: 4 x 4

1	1	2	2
1	1	2	2
3	3	4	4
3	3	4	4

“Bed of Nails”

Input: 2 x 2

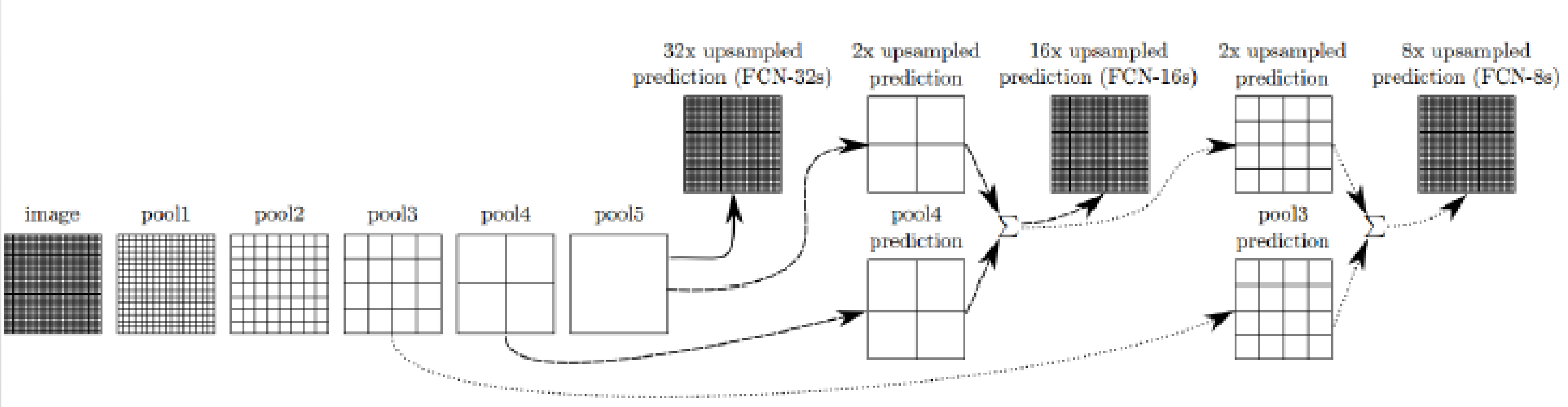
1	2
3	4

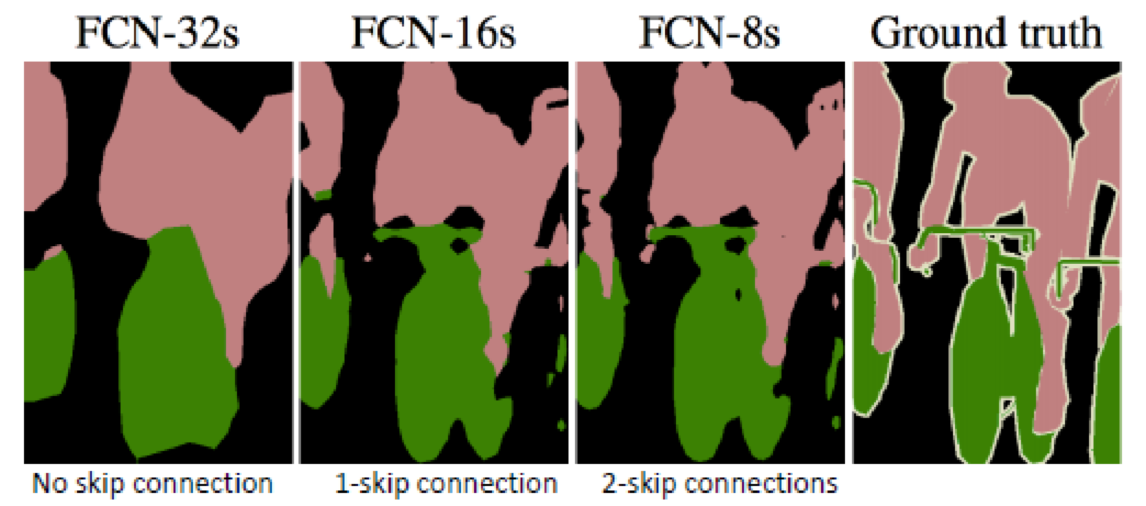
Output: 4 x 4

1	0	2	0
0	0	0	0
3	0	4	0
0	0	0	0

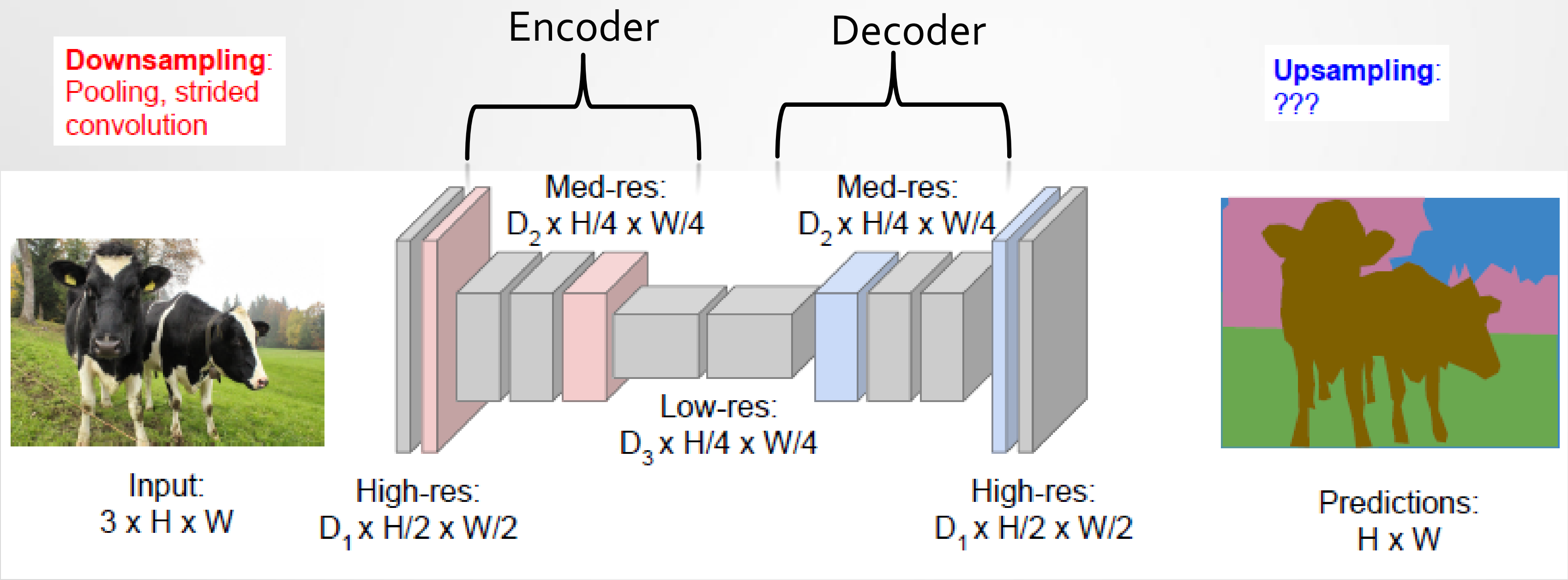
FCN (Skip Connections)

- Utilize skip-layer concept to improve the segmentation accuracy.





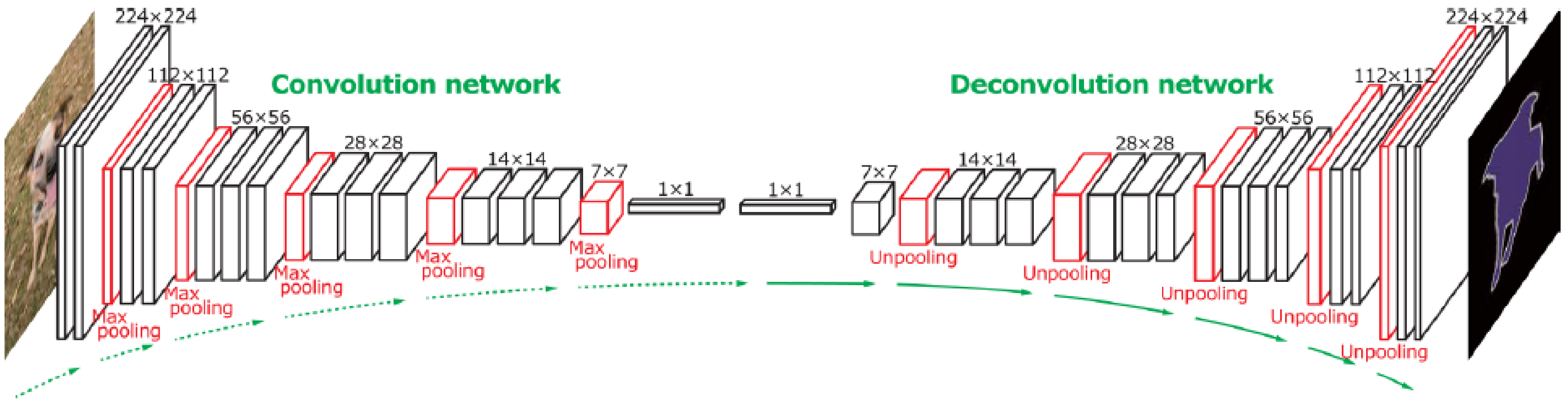
Deconvolution Network for Semantic Segmentation



Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

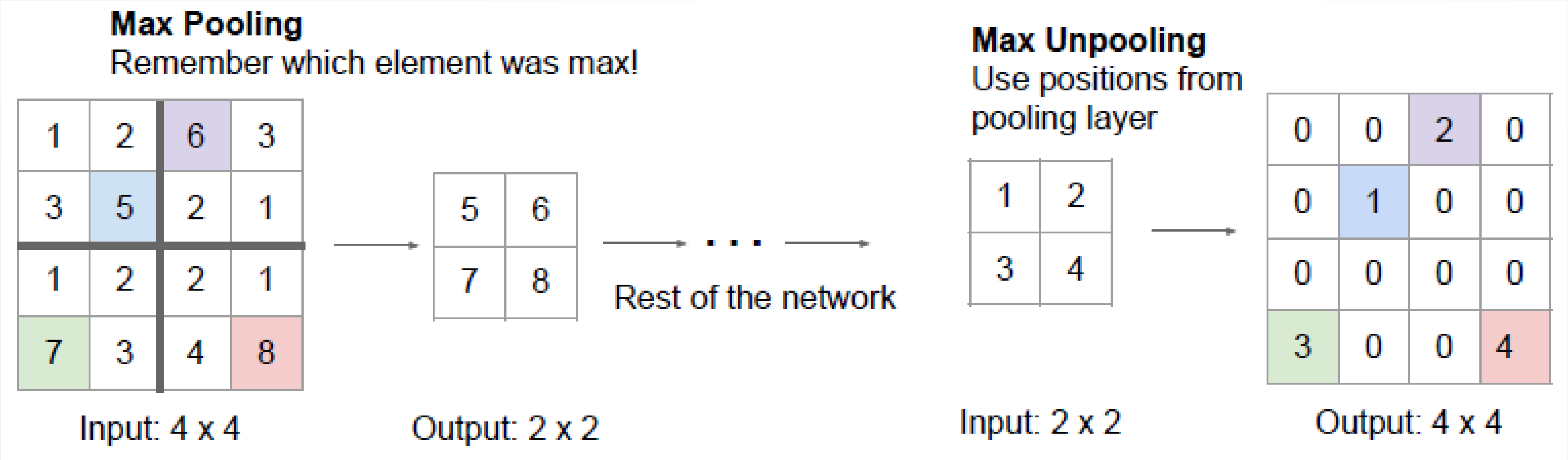
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

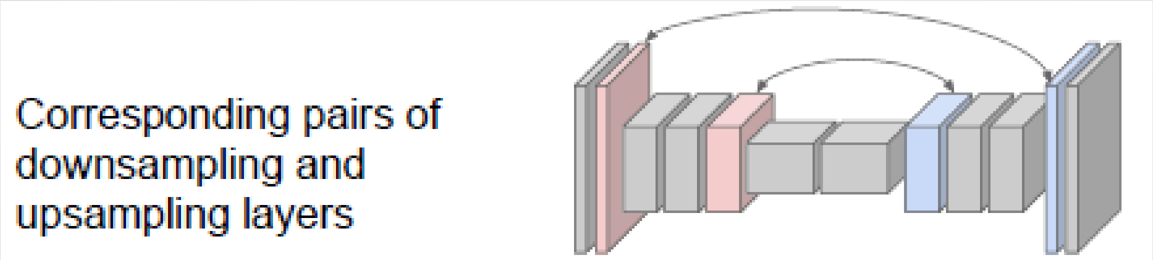
Deconvolution Network for Semantic Segmentation



Noh et al, “Learning Deconvolution Network for Semantic Segmentation”, ICCV 2015

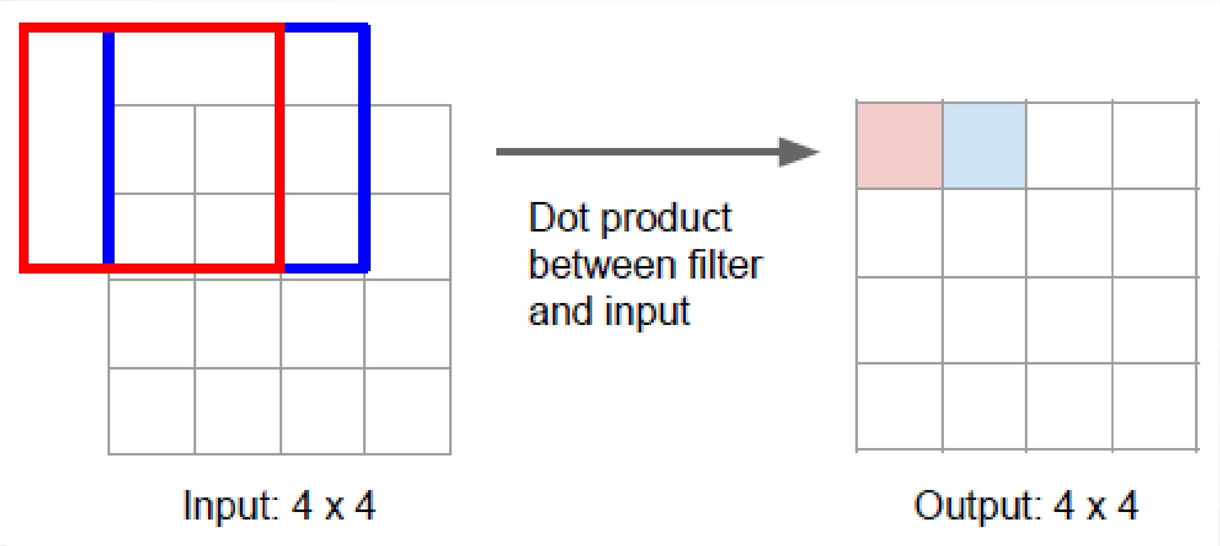
In-Network upsampling: "Max Unpooling"





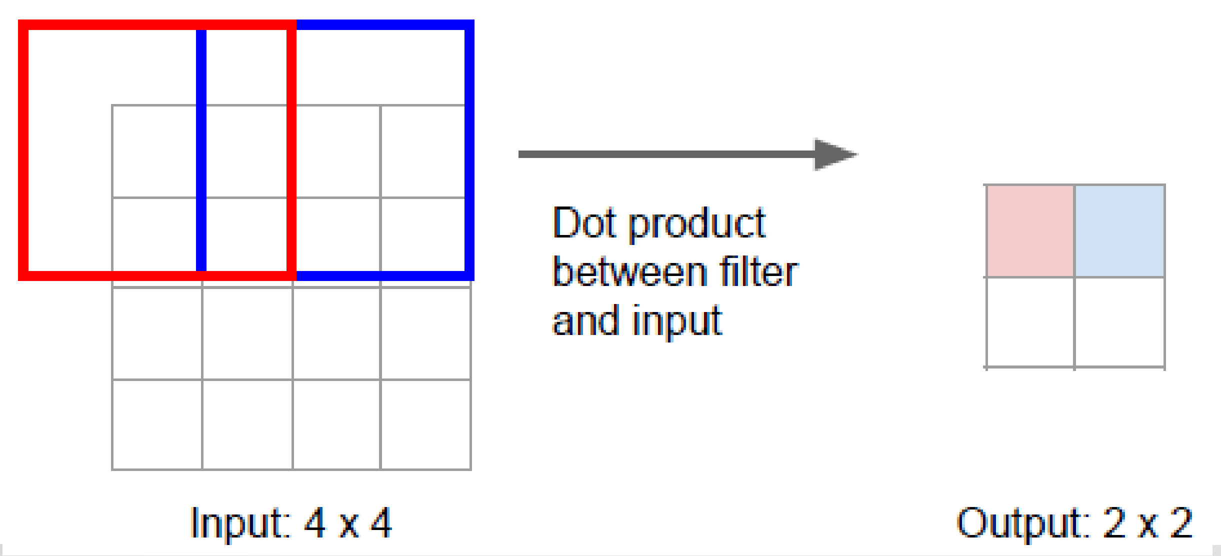
Learnable Upsampling: Transpose Convolution

- **Recall:** Normal 3 x 3 convolution, stride 1 pad 1



Learnable Upsampling: Transpose Convolution

- **Recall:** Normal 3 x 3 convolution, stride 2 pad 1

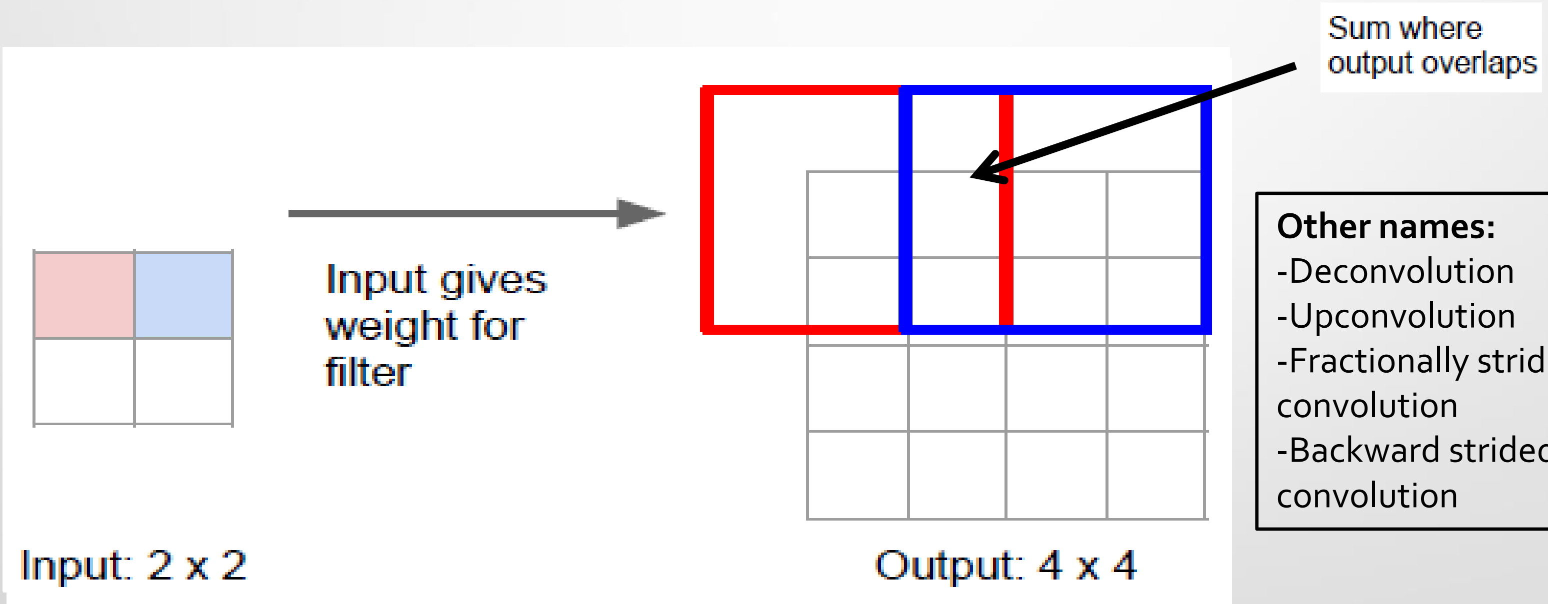


Filter moves 2 pixels in the input for every one pixel in the output

Stride gives ratio between movement in input and output

Learnable Upsampling: Transpose Convolution

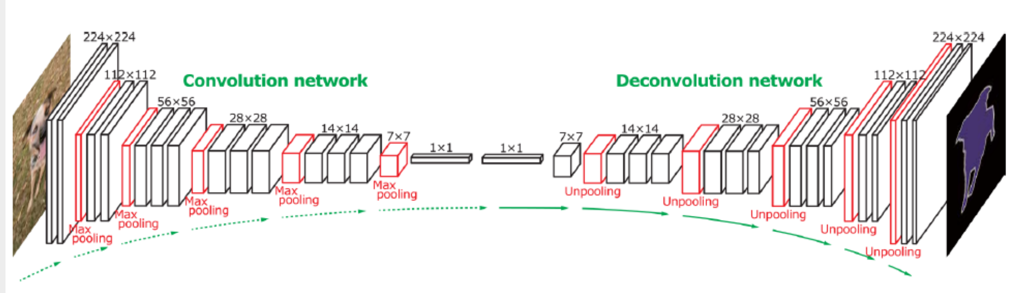
3 x 3 transpose convolution, stride 2 pad 1

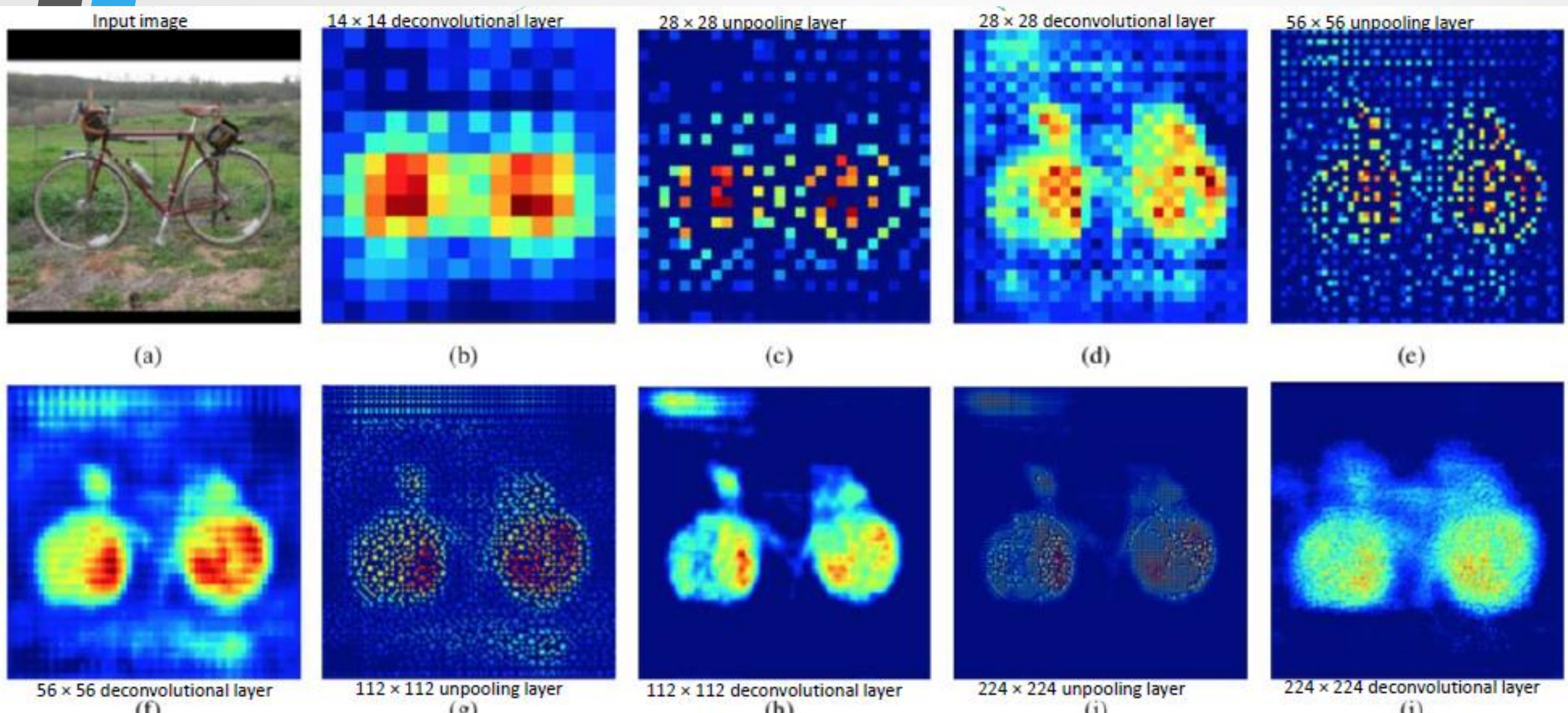


Other names:
-Deconvolution
-Upconvolution
-Fractionally strided convolution
-Backward strided convolution

Filter moves 2 pixels in the output for every one pixel in the input

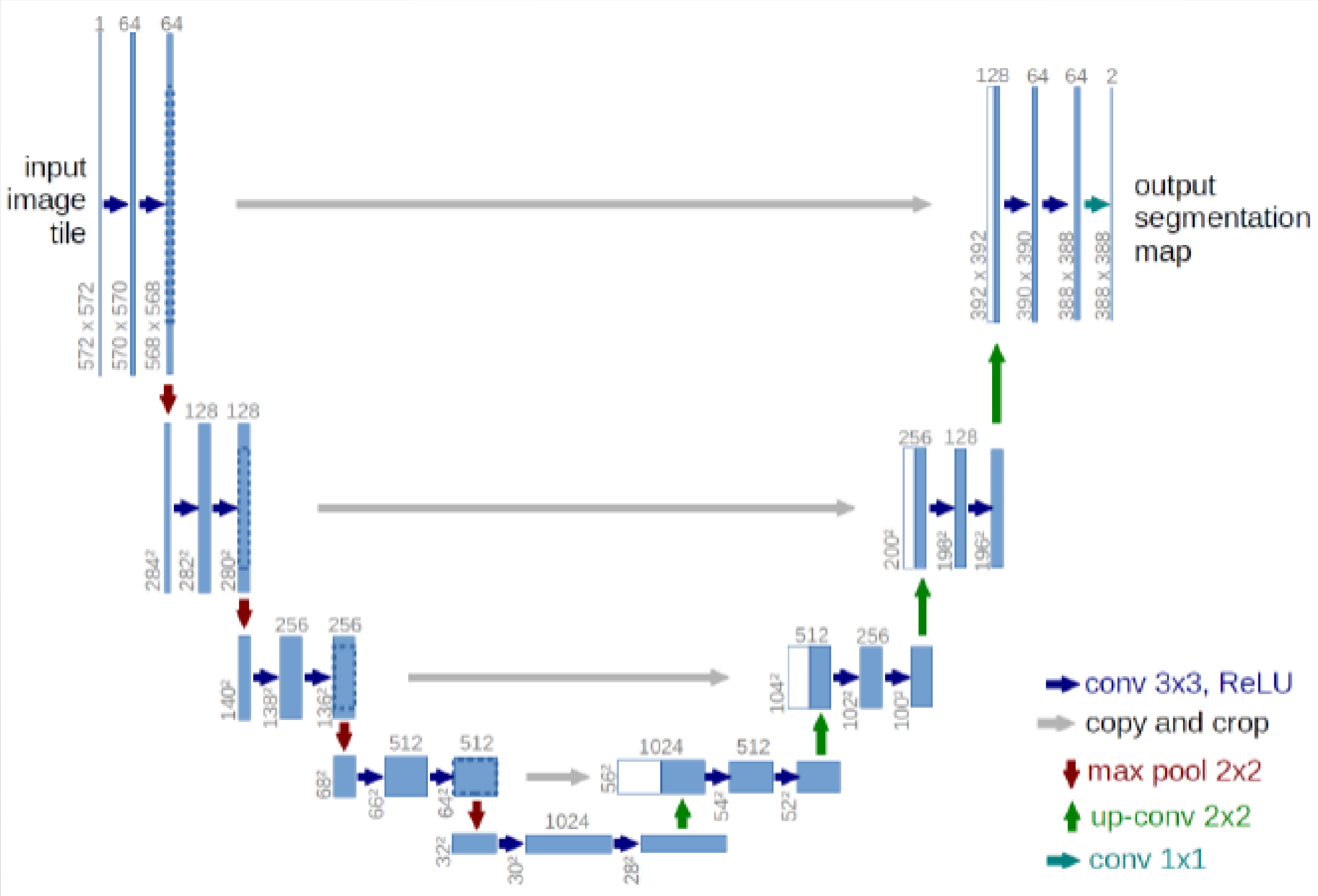
Stride gives ratio between movement in output and input





Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

U-Net



- Discard un-pooling and keep up-sampling (deconvolution), in addition to skip connections from same down-sampling layer to up-sampling layer

Source: Olaf Ronneberger, Philipp Fischer, Thomas Brox"U-Net: Convolutional Networks for Biomedical Image Segmentation", MICCAI, 2015

Instance Segmentation

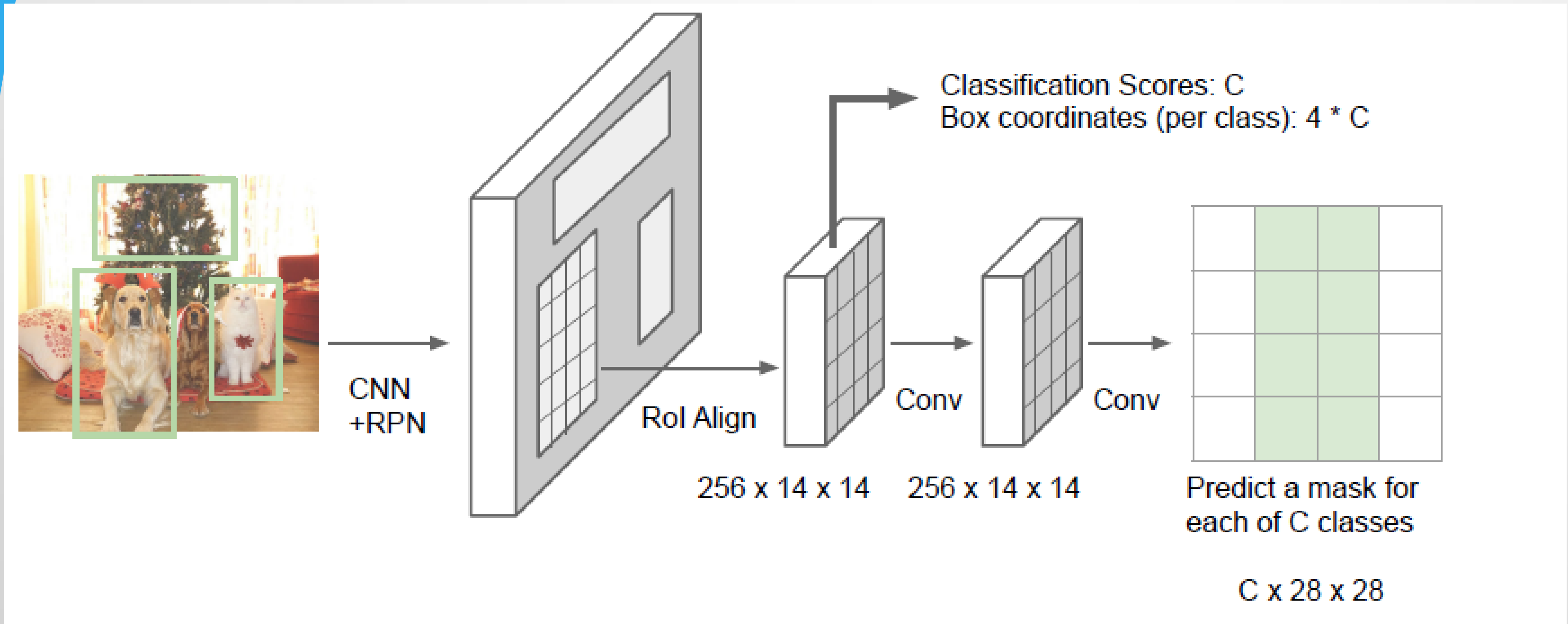
- Segment each instance of the same class separately.

Semantic & Instance Segmentation

- Instance Segmentation: (hybrid between semantic segmentation & object detection)

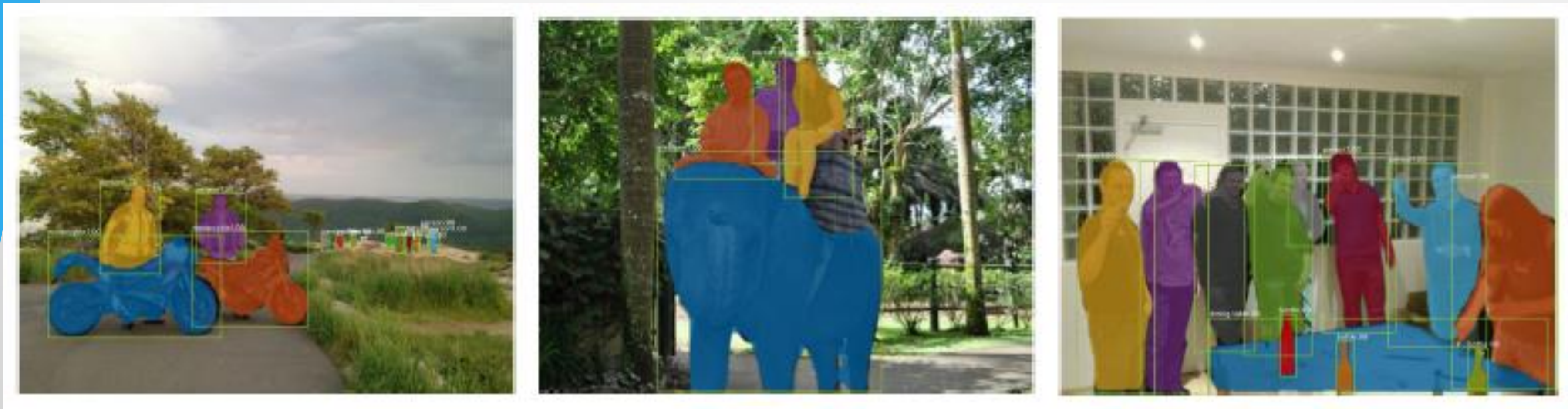
Mask R-CNN

Mask R-CNN = Faster R-CNN + FCN on RoIs



K. He, G. Gkioxari, P. Dollar, and R. Girshick, Mask R-CNN, ICCV 2017 (Best Paper Award)

Mask R-CNN: Very Good Results!



Credit for

CS 4495 Computer Vision (Spring 2015)

A. Bob - College of Computing, Georgia Tech.

CS231n "Convolutional Neural Networks for Visual Recognition"
by University of Stanford (Lecture 11)

CAP5415 " Computer Vision " University of Central Florida,
Center of Research in Computer Vision (UCF CRCV), Fall 2020