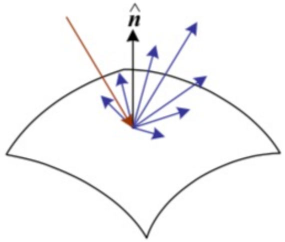
Segmentation & Custering

Disclaimer: Many slides have been borrowed from Devi Parikh and Kristen Grauman, who may have borrowed some of them from others. Any time a slide did not already have a credit on it, I have credited it to Kristen. So there is a chance some of these credits are inaccurate.



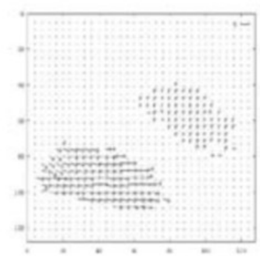
2. Image Formation

3. Image Processing

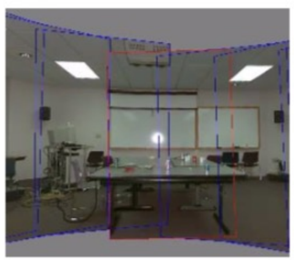
4. Features

5. Segmentation

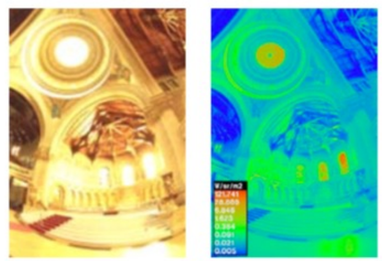
6-7. Structure from Motion



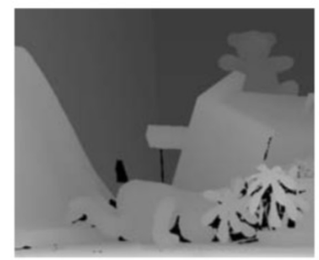
8. Motion



9. Stitching

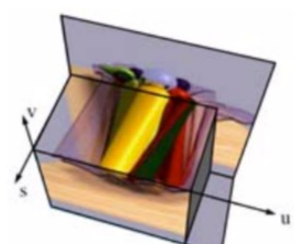


10. Computational Photography



11. Stereo

12. 3D Shape



13. Image-based Rendering

14. Recognition

Grouping in Vision
Segmentation as Clustering
Feature Representations
Data-driven Features

$\boxed{\text{Grouping in Vision}}$

Segmentation as Clustering

Feature Representations

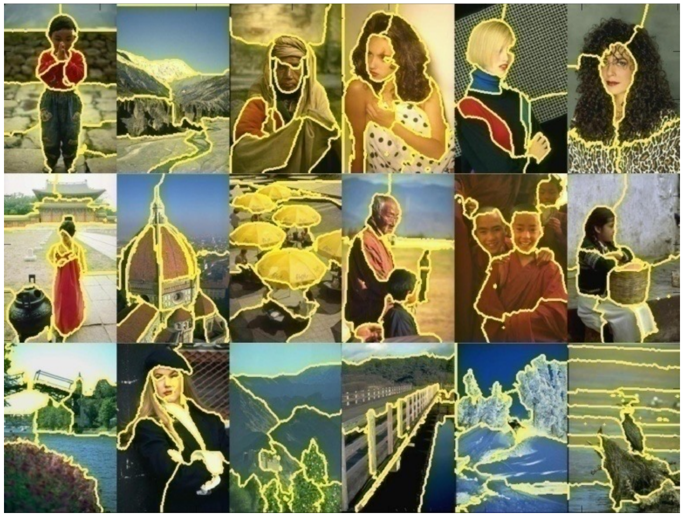
Data-driven Features

Grouping in vision

- Goals:
  - Gather features that belong together
  - Obtain an intermediate representation that compactly describes key image or video parts

Slide credit: Kristen Grauman

Examples of grouping in vision



[Figure by J. Shi]

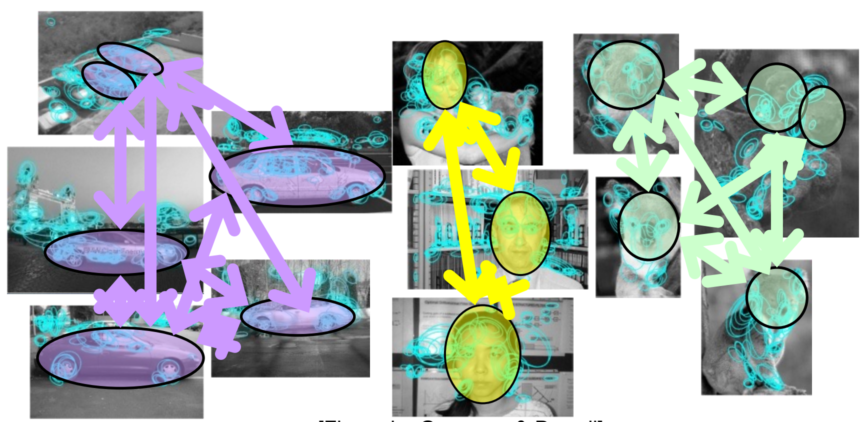
Determine image regions

[http://poseidon.csd.auth.gr/LAB_RESEARCH/Latest/imgs/SpeakDepVidIndex_img2.jpg]

Group video frames into shots

[Figure by Wang & Suter]

Figure-ground



[Figure by Grauman & Darrell]

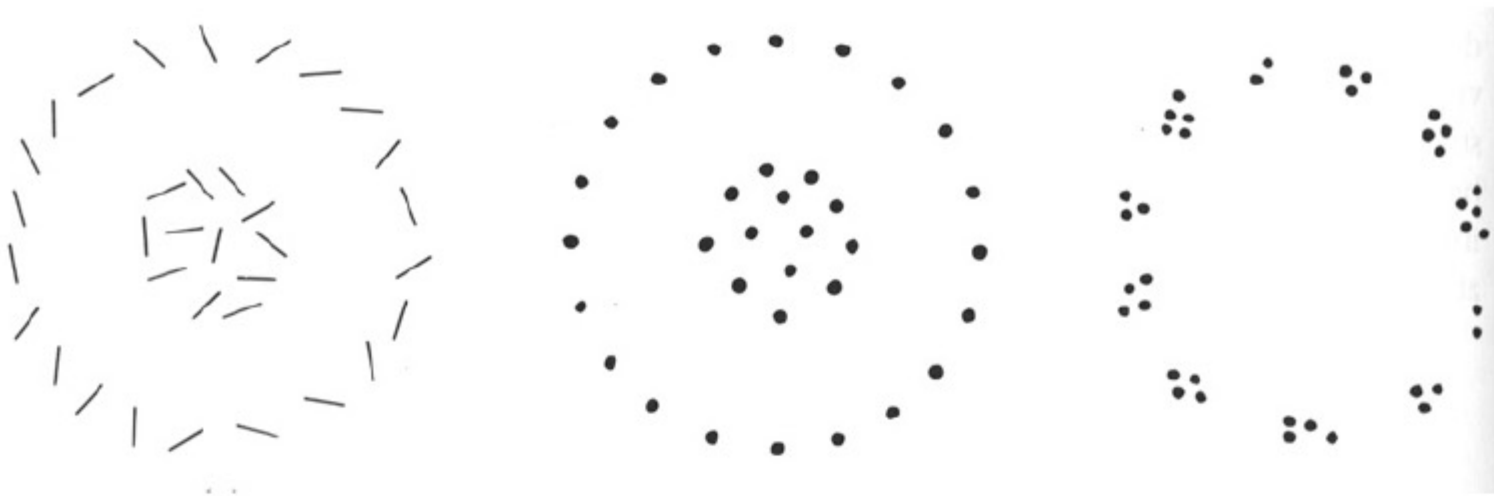
Object-level grouping

Slide credit: Kristen Grauman

Grouping in vision

- Goals:
  - Gather features that belong together
  - Obtain an intermediate representation that compactly describes key image (video) parts
- Top down vs. bottom up segmentation
  - Top down: pixels belong together because they are from the same object
  - Bottom up: pixels belong together because they look similar
- Hard to measure success
  - What is interesting depends on the application.

Slide credit: Kristen Grauman



Slide credit: Kristen Grauman

Gestalt

- Gestalt: whole or group
  - Whole is something other than sum of its parts
  - Relationships among parts can yield new properties/features

- Psychologists identified series of factors that predispose set of elements to be grouped (by human visual system)

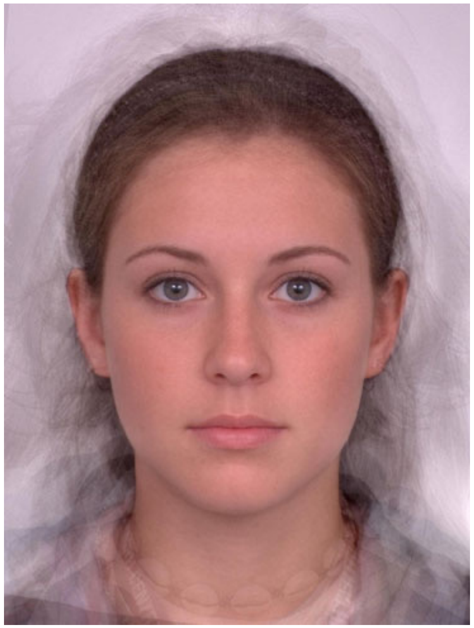
Slide credit: Kristen Grauman

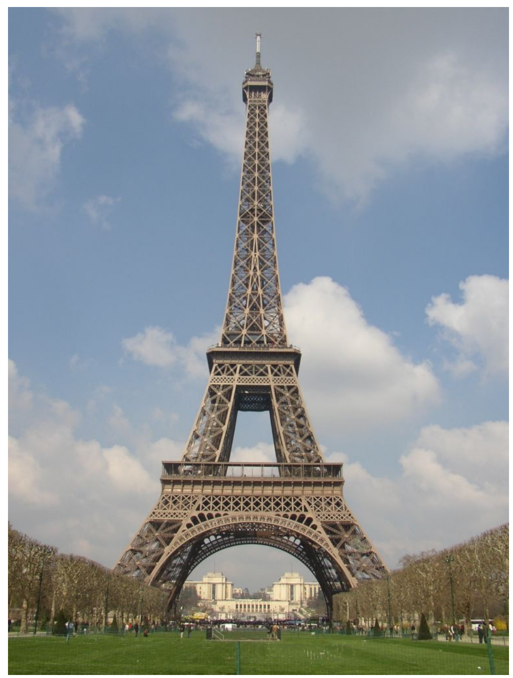
Similarity

Kristen Grauman

http://chicagoist.com/attachments/chicagoist_alicia/GEESE.jpg, http://wwwdelivery.superstock.com/WI/223/1532/PreviewComp/SuperStock_1532R-0831.jpg

Symmetry





Slide credit: Kristen Grauman

http://seedmagazine.com/news/2006/10/beauty_is_in_the_processingtim.php

Common fate

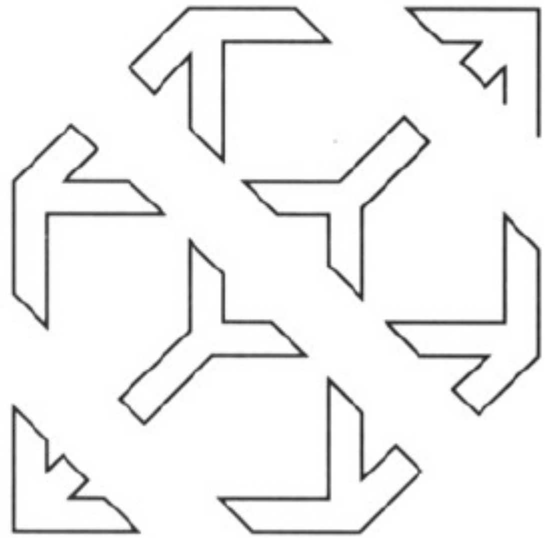
Image credit: Arthus-Bertrand (via F. Durand)

(coherent motion)

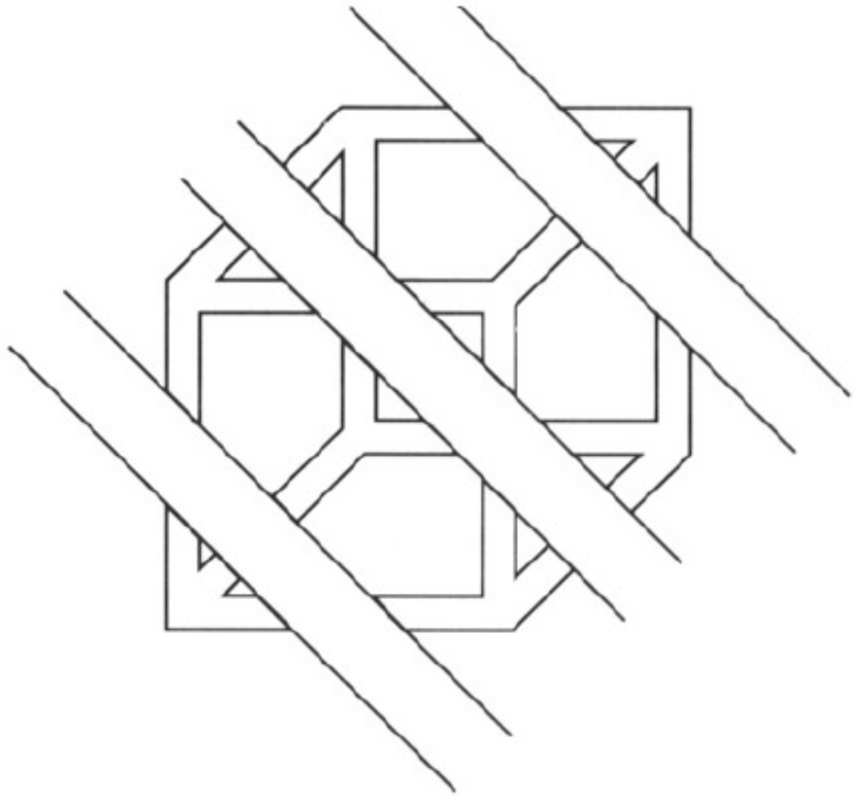
Proximity

Slide credit: Kristen Grauman

http://www.capital.edu/Resources/Images/outside6_035.jpg

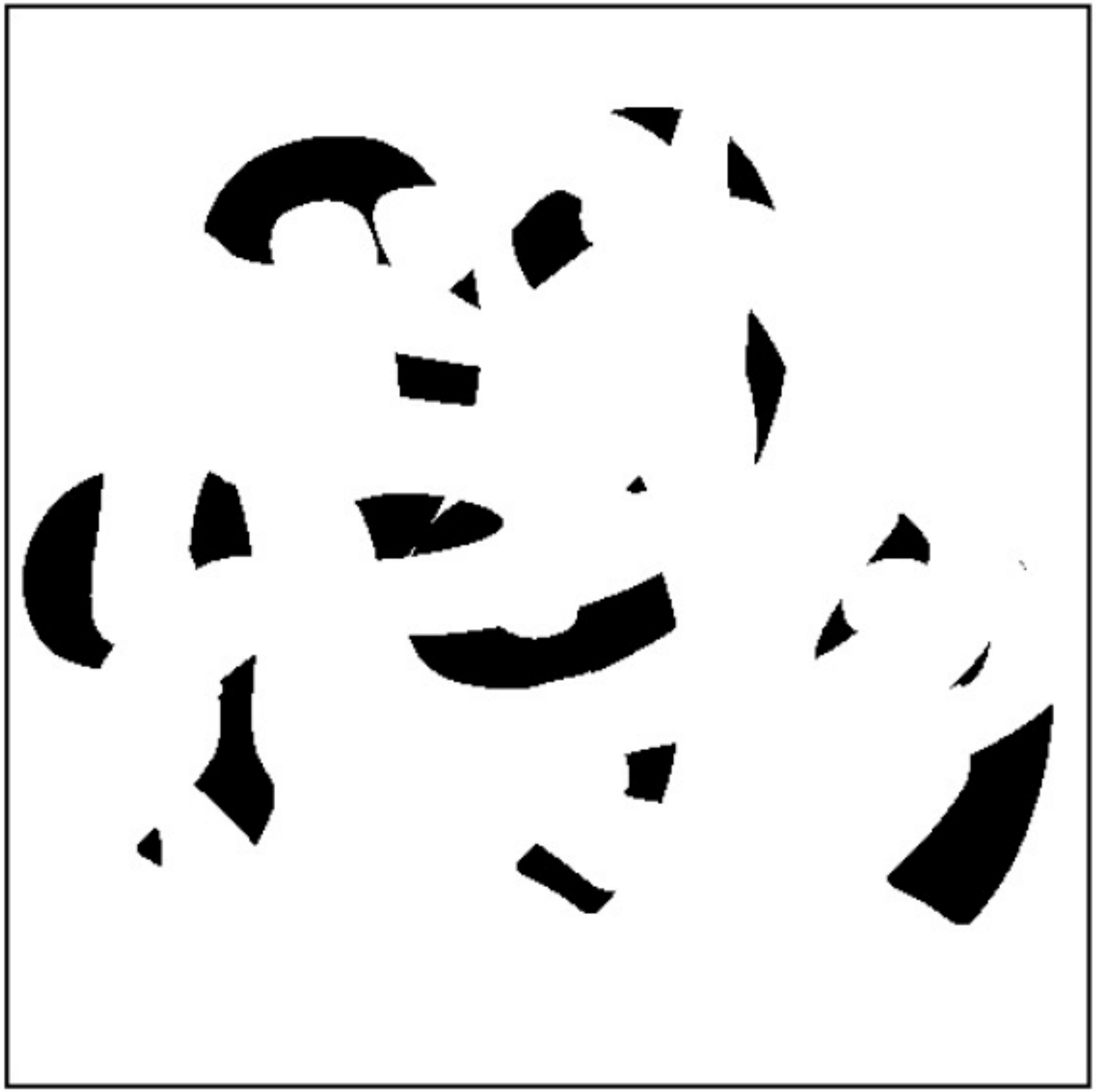


Slide credit: Kristen Grauman



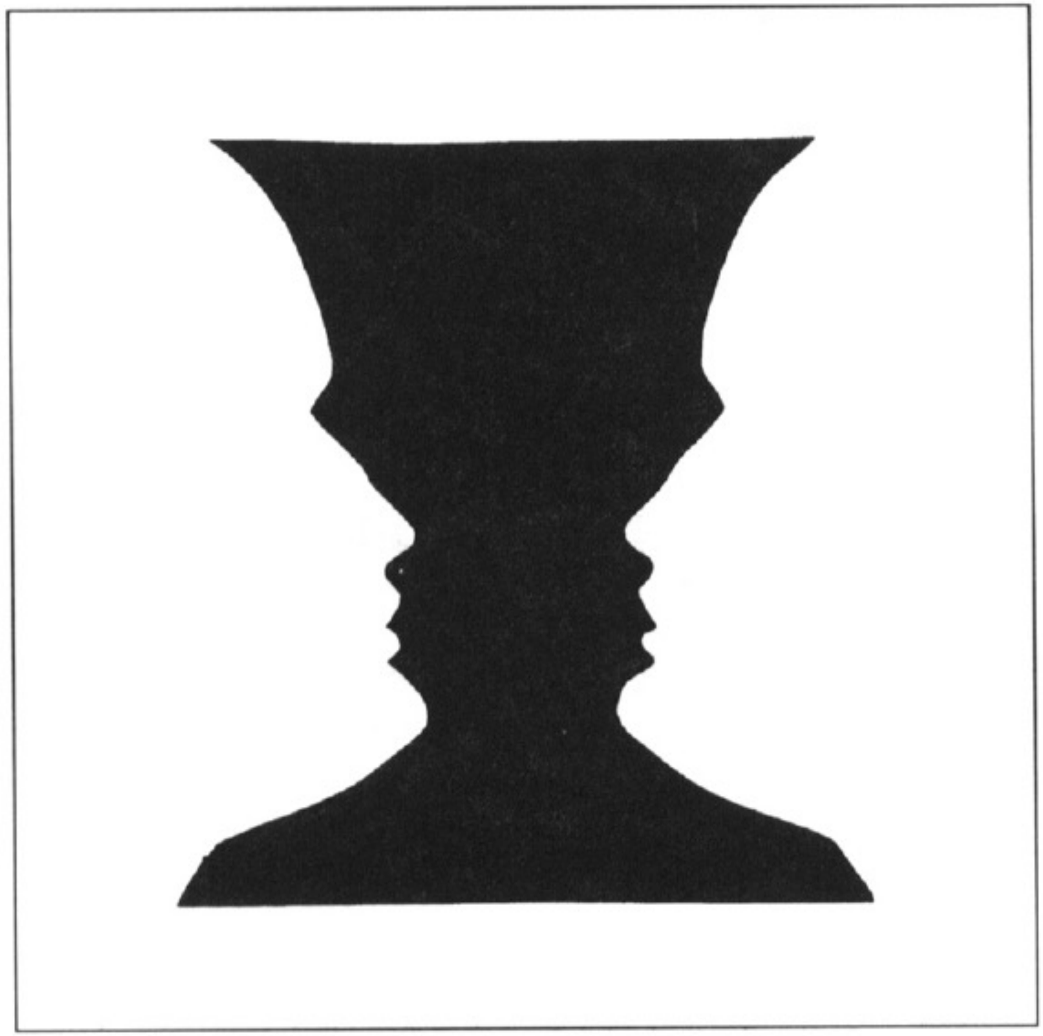
Continuity, explanation by occlusion

Slide credit: Kristen Grauman



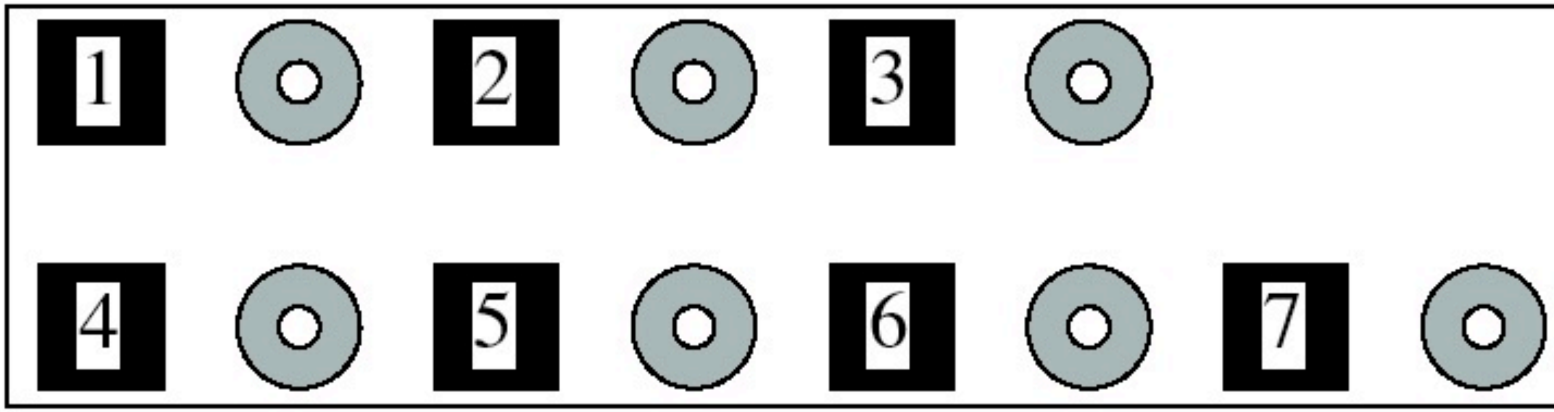
Slide credit: Kristen Grauman

Figure-ground



Slide credit: Kristen Grauman

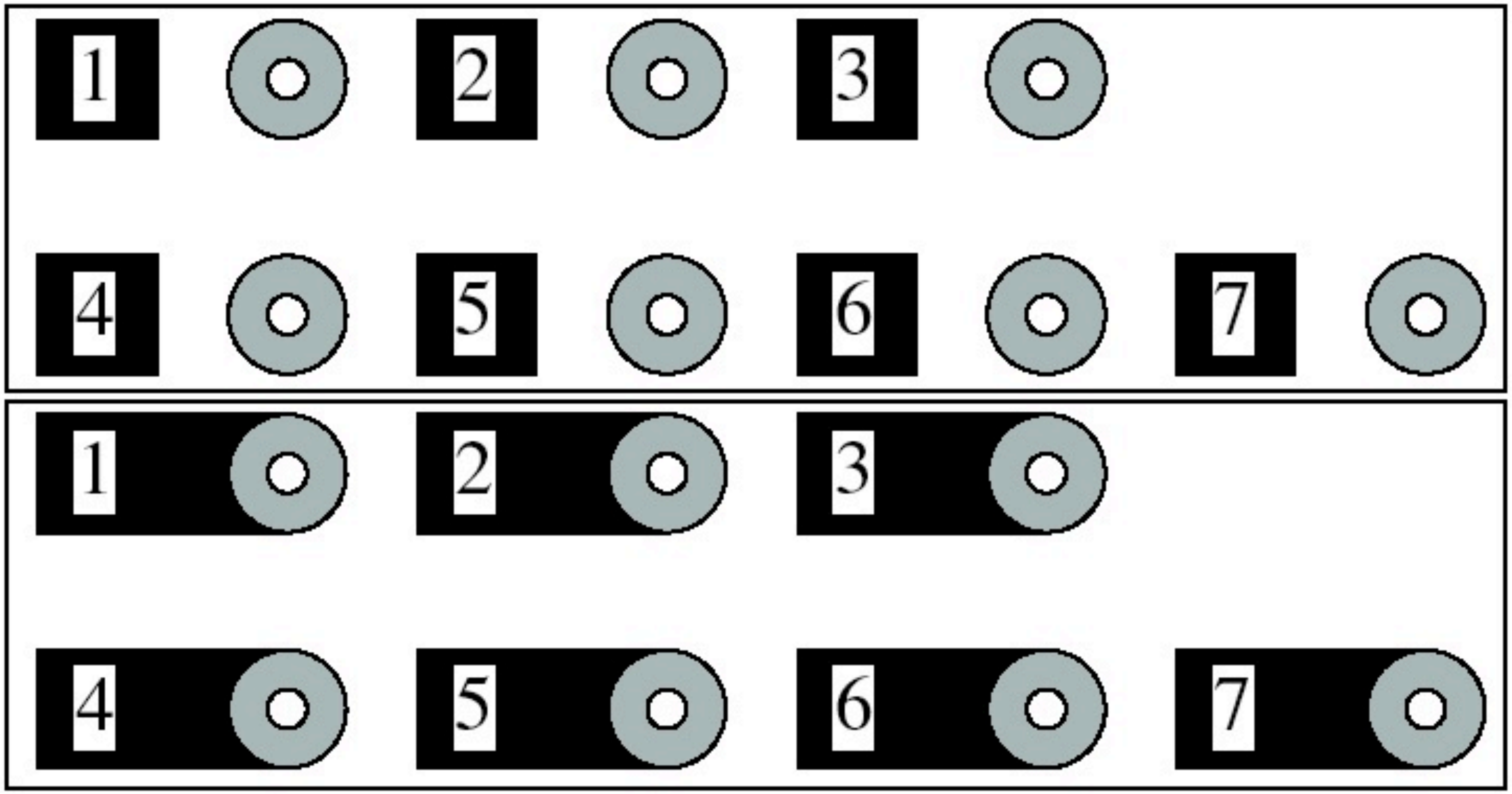
Grouping phenomena in real life



Forsyth & Ponce, Figure 14.7

Slide credit: Kristen Grauman

Grouping phenomena in real life



Forsyth & Ponce, Figure 14.7

Slide credit: Kristen Grauman

Gestalt

- Gestalt: whole or group
  - Whole is other than sum of its parts
  - Relationships among parts can yield new properties/features
- Psychologists identified series of factors that predispose set of elements to be grouped (by human visual system)
- Inspiring observations/explanations; challenge remains how to best map to algorithms.

Slide credit: Kristen Grauman

Grouping in Vision

Segmentation as Clustering

Feature Representations

Data-driven Features

The goals of segmentation

- Separate image into coherent “objects”

image   human segmentation

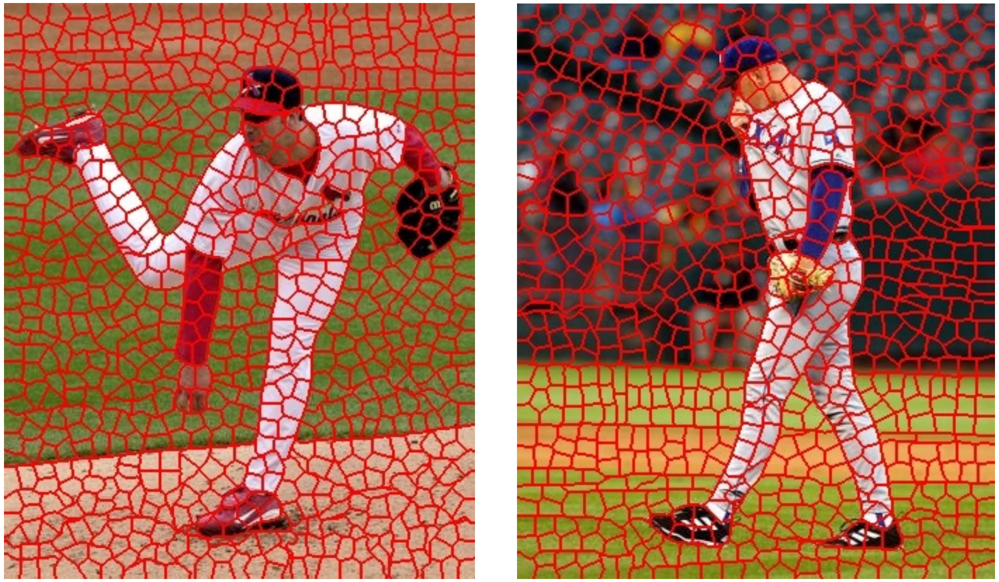
Source: Lana Lazebnik

The goals of segmentation

- Separate image into coherent "objects"

- Group together similar-looking pixels for efficiency of further processing

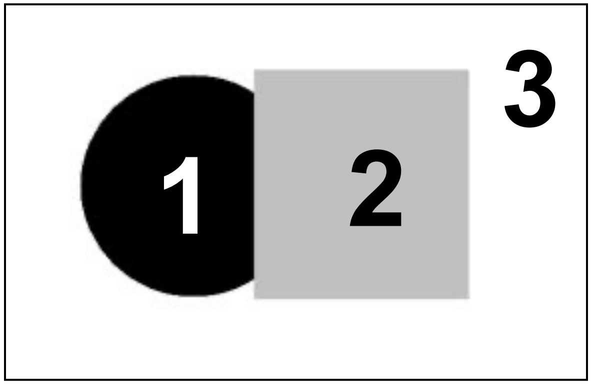
"superpixels"



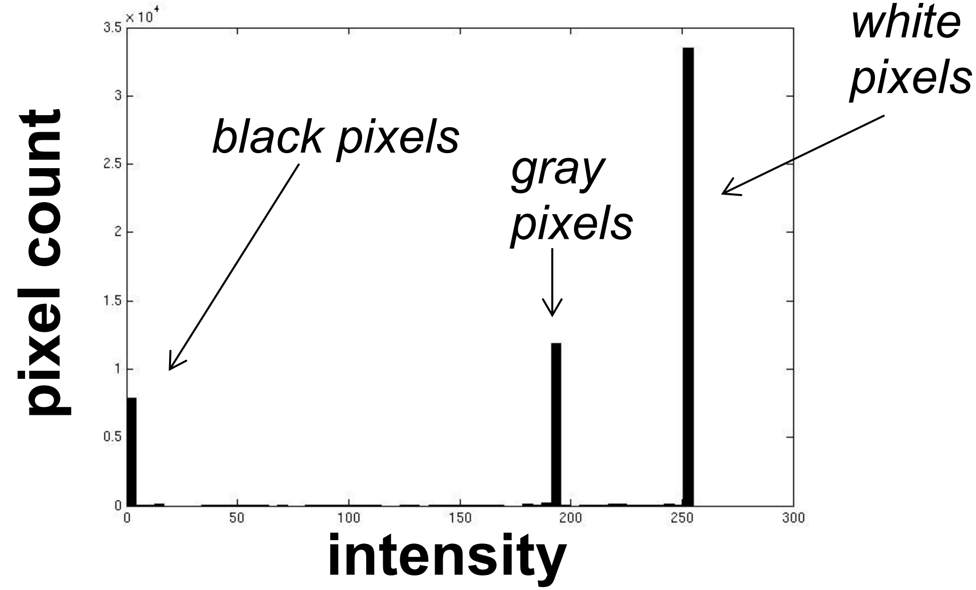
X. Ren and J. Malik. [Learning a classification model for segmentation.]() ICCV 2003.

Source: Lana Lazebnik

Image segmentation: toy example

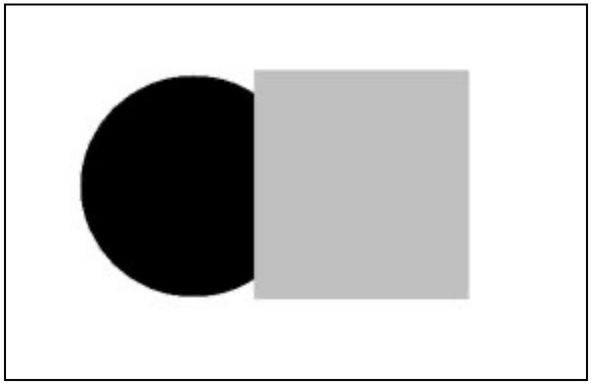


input image

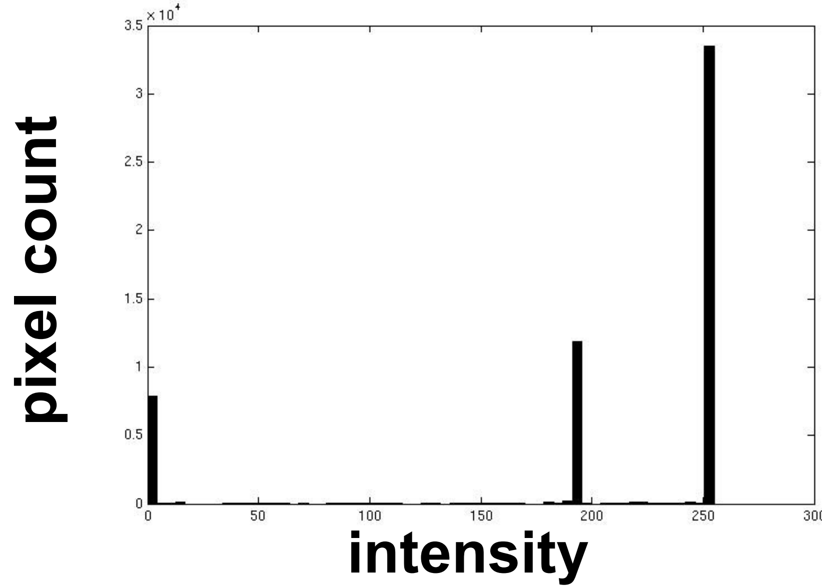


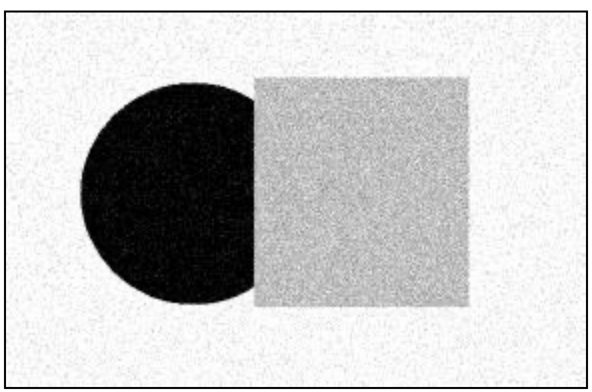
- These intensities define the three groups.
- We could label every pixel in the image according to which of these primary intensities it is.
  - i.e., *segment* the image based on the intensity feature.
- What if the image isn't quite so simple?

Kristen Grauman

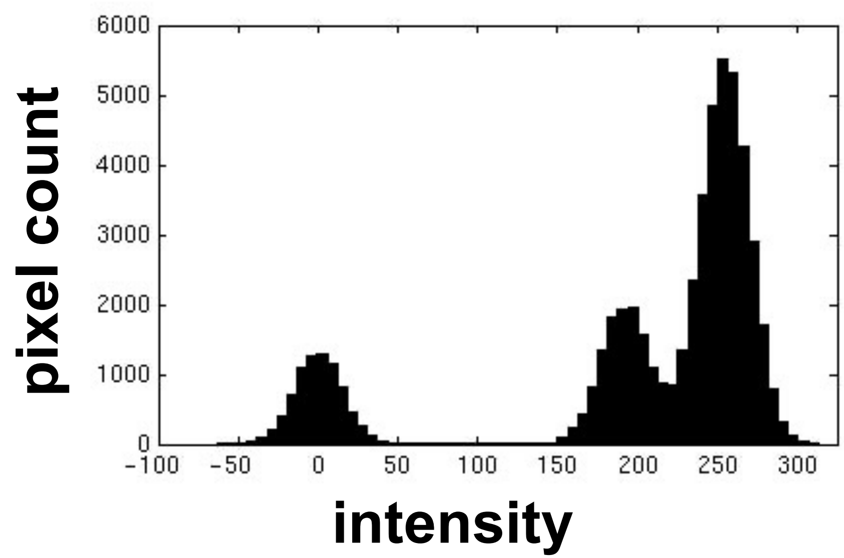


input image

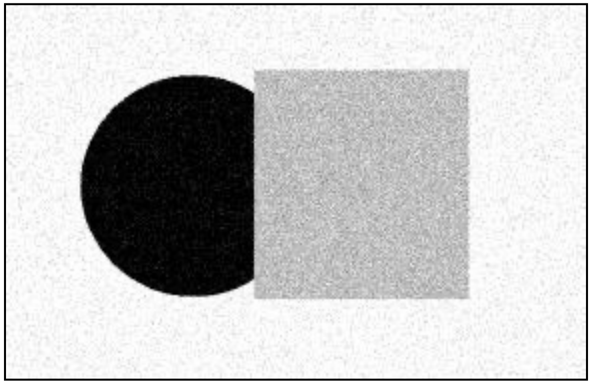




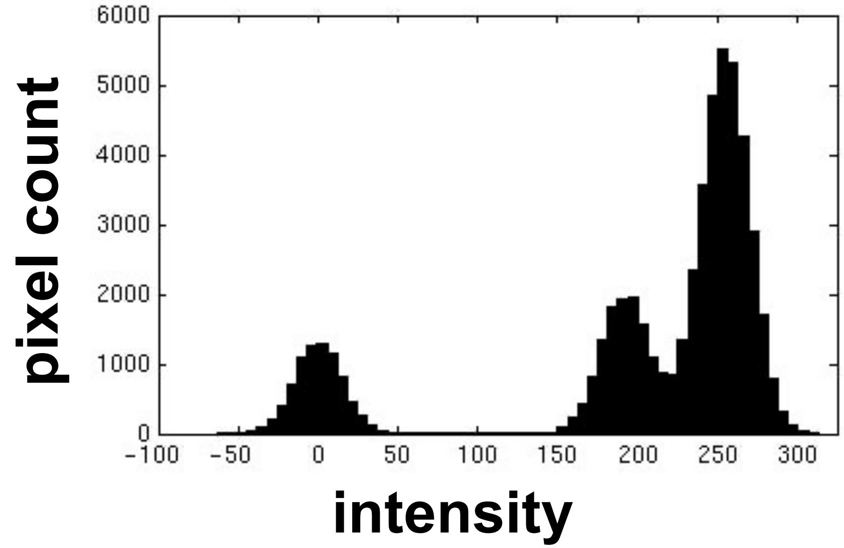
input image



Kristen Grauman



input image



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

Kristen Grauman

K-means clustering

- Basic idea: randomly initialize the *k* cluster centers, and iterate between the two steps we just saw.

  1. Randomly initialize the cluster centers, $c_1, \ldots, c_K$
  2. Given cluster centers, determine points in each cluster
     - For each point p, find the closest c_i. Put p into cluster i
  3. Given points in each cluster, solve for c_i
     - Set c_i to be the mean of points in cluster i
  4. If c_i have changed, repeat Step 2

Properties

- Will always converge to *some* solution
- Can be a "local minimum"
  - does not always find the global minimum of objective function:

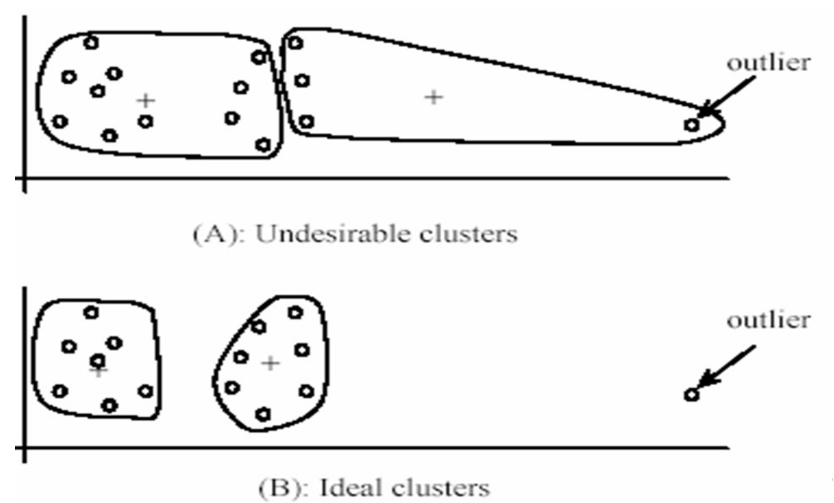
$$\sum_{\text{clusters } i} \sum_{\text{points p in cluster } i} ||p - c_i||^2$$

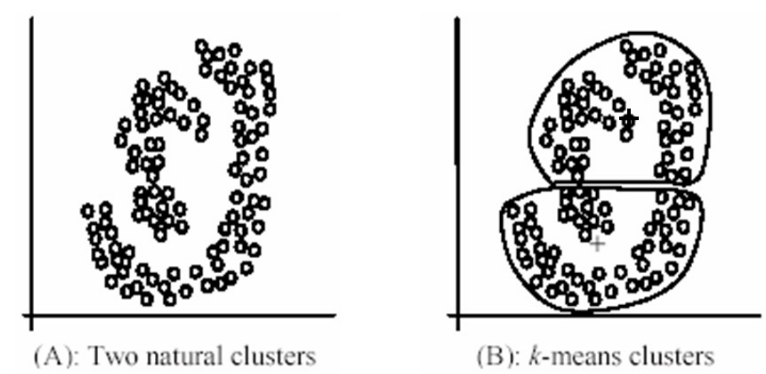
Source: Steve Seitz

K-means: pros and cons

<u>Pros</u>
- Simple, fast to compute
- Converges to local minimum of within-cluster squared error

<u>Cons/issues</u>
- Setting k?
- Sensitive to initial centers
- Sensitive to outliers
- Detects spherical clusters
- Assuming means can be computed

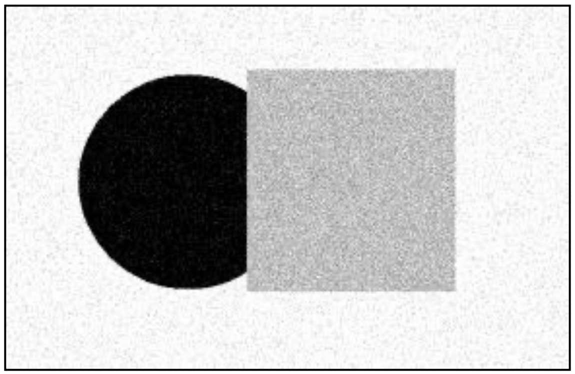




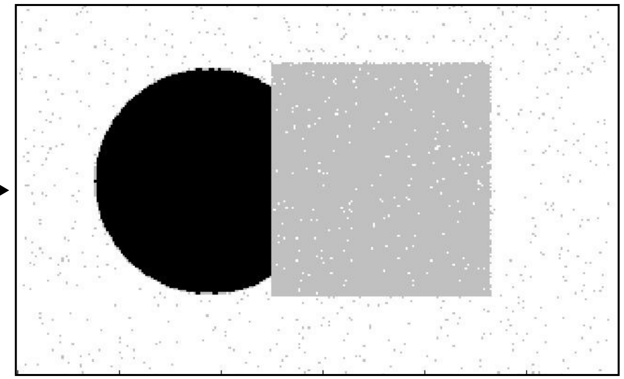
Slide credit: Kristen Grauman

An aside: Smoothing out cluster assignments

- Assigning a cluster label per pixel may yield outliers:



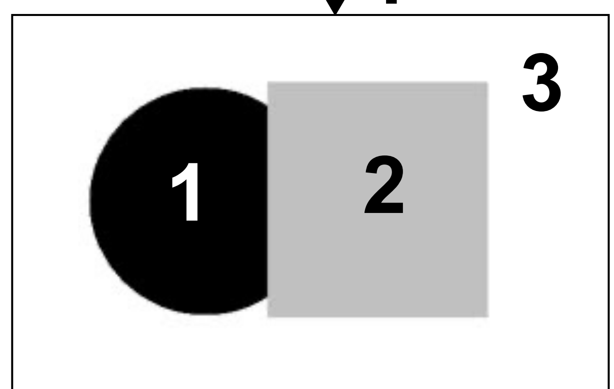
original



labeled by cluster center's intensity

?

- How to ensure they are spatially smooth?



Kristen Grauman

Grouping in Vision

Segmentation as Clustering

$\boxed{\text{Feature Representations}}$

Data-driven Features

Segmentation as clustering

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

Grouping pixels based on **intensity** similarity

Feature space: intensity value (1-d)

Slide credit: Kristen Grauman

K=2

K=3

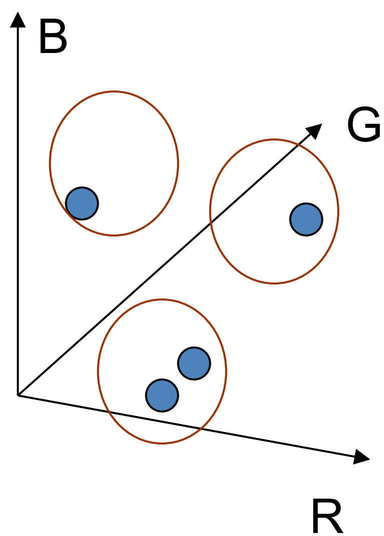
quantization of the feature space;
segmentation label map

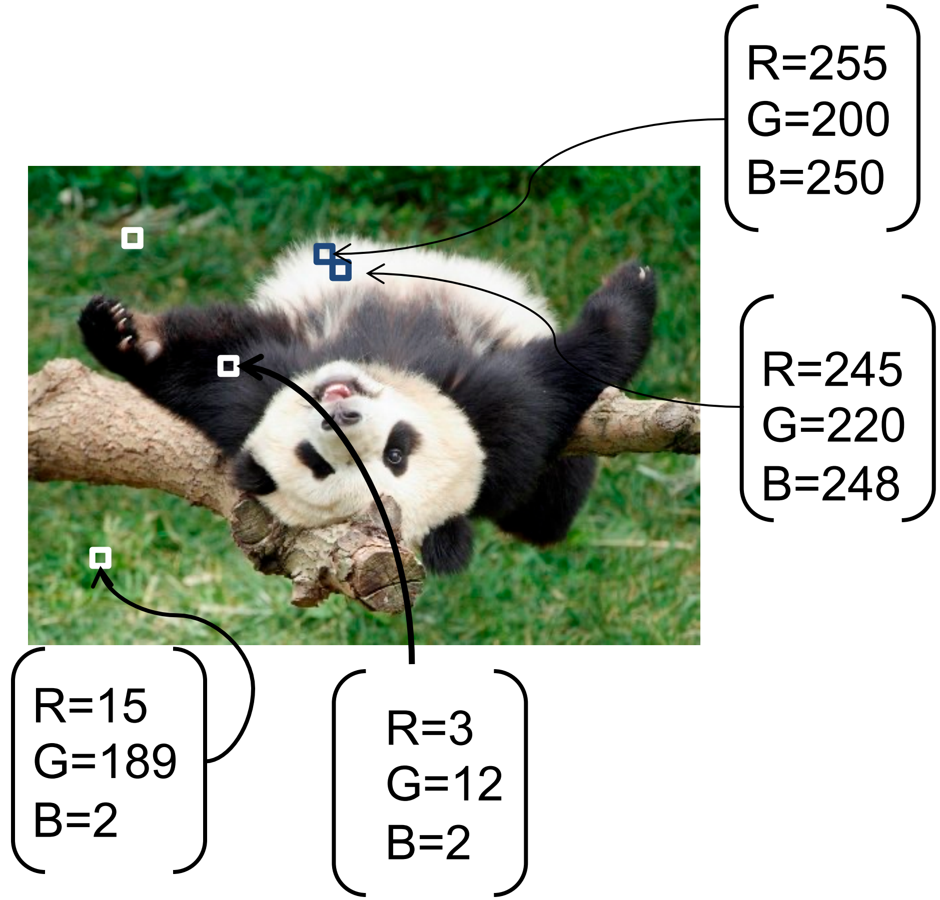
Slide credit: Kristen Grauman

Segmentation as clustering

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

Grouping pixels based on **color** similarity





Feature space: color value (3-d)

Kristen Grauman

Segmentation as clustering

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

Grouping pixels based on **intensity** similarity

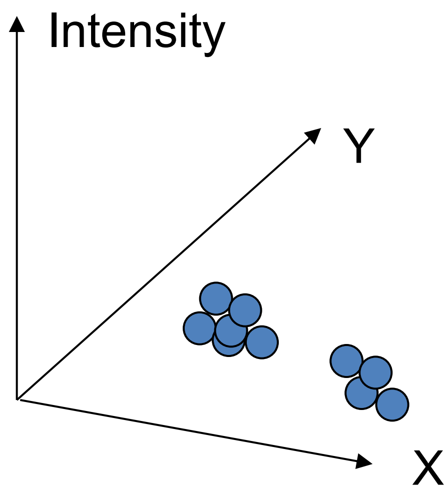
Clusters based on intensity similarity don’t have to be spatially coherent.

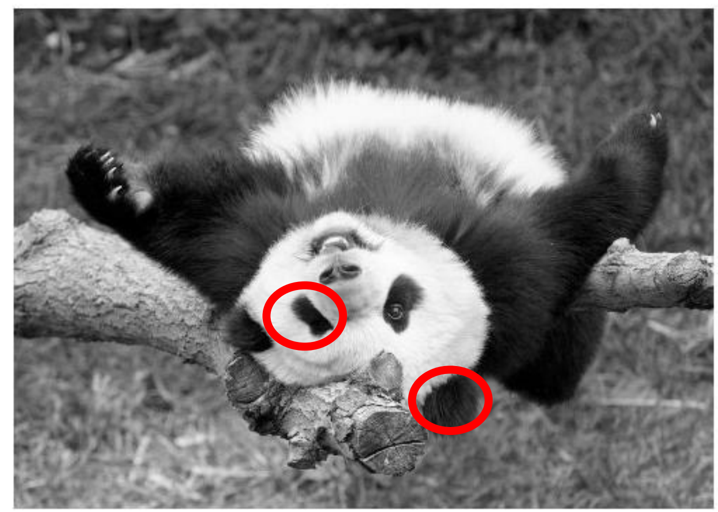
Kristen Grauman

Segmentation as clustering

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

Grouping pixels based on **intensity+position** similarity

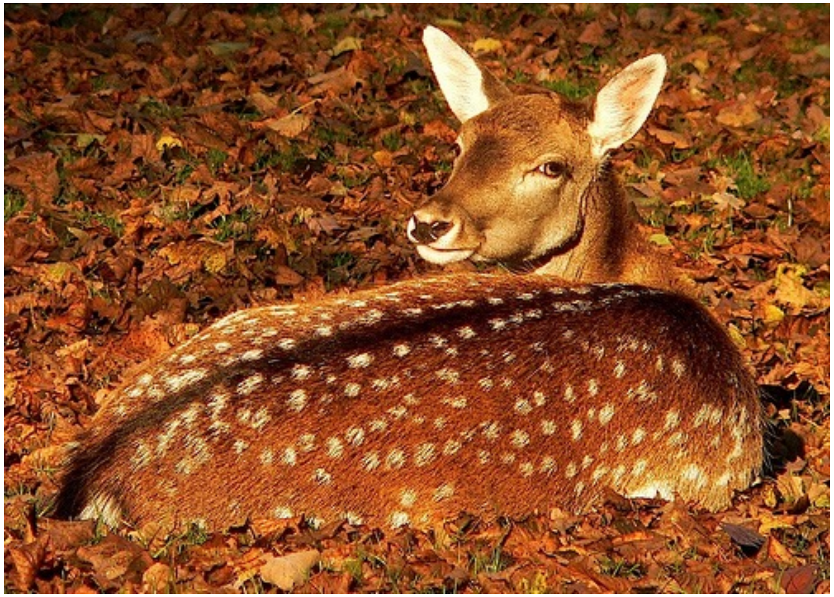




Both regions are black, but if we also include **position (x,y),** then we could group the two into distinct segments; way to encode both similarity & proximity.

Segmentation as clustering

- Color, brightness, position alone are not enough to distinguish all regions…



Slide credit: Kristen Grauman

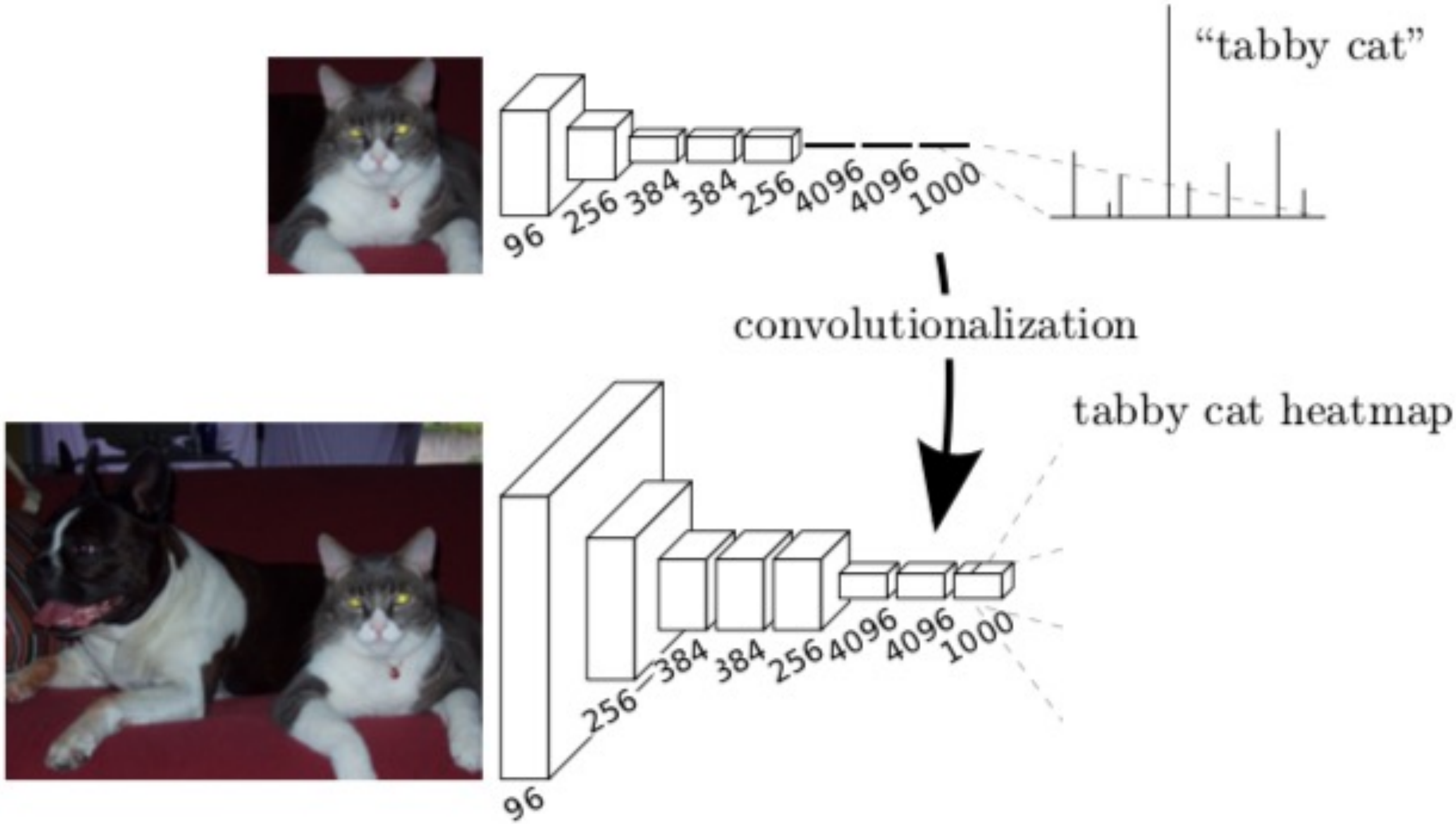
Grouping in Vision

Segmentation as Clustering

Feature Representations

$\boxed{\text{Data-driven Features}}$

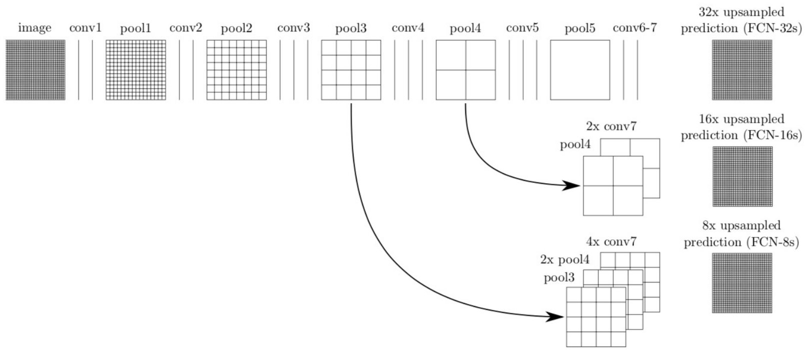
Fully convolutional nets…

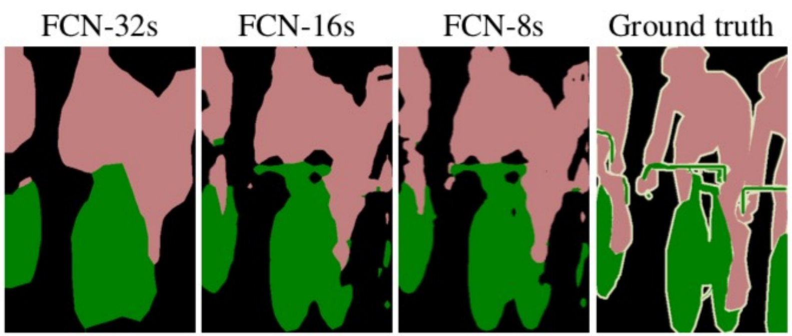


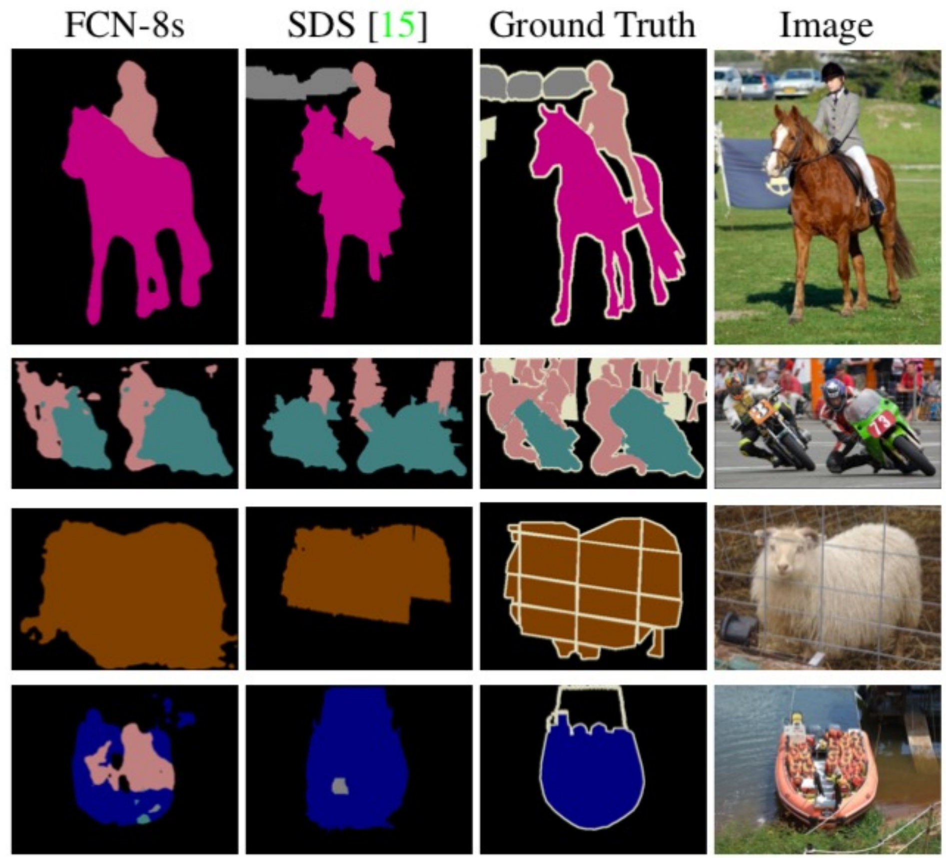
- "Expand" trained network to any size

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In CVPR 2014

…for segmentation



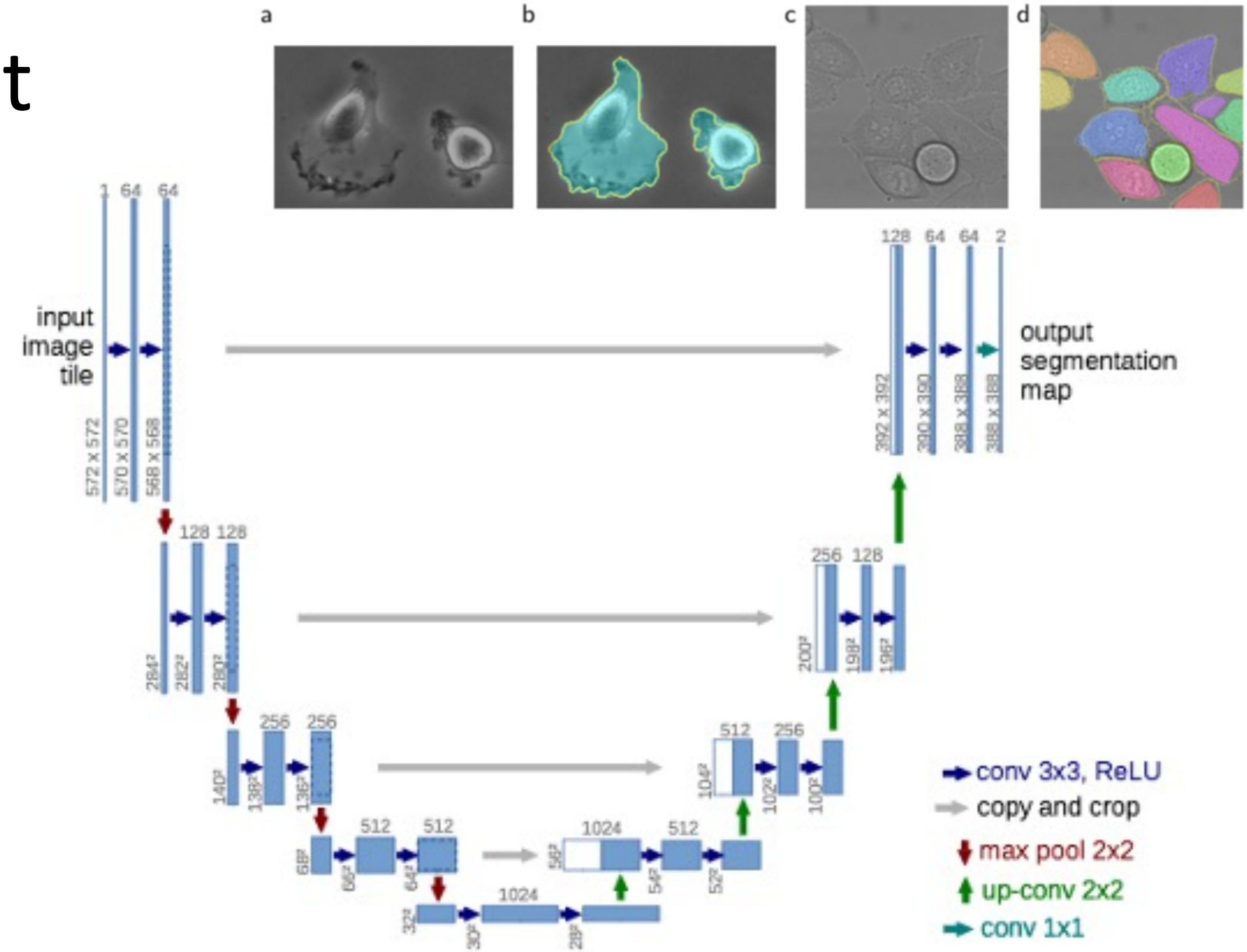




- Complicated upsampling strategies…
- Results not yet great

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In CVPR 2014

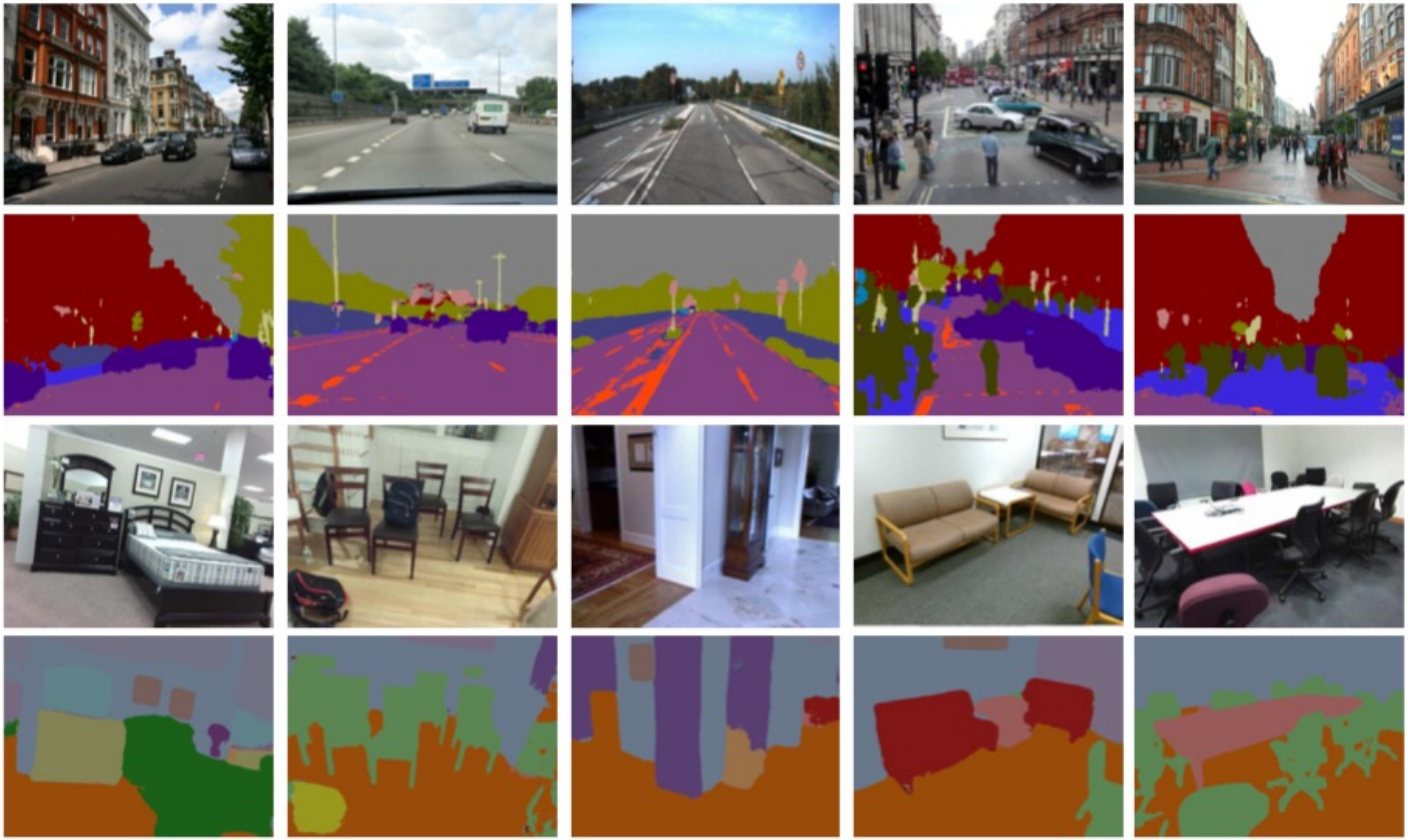
U-Net



- Builds on FCN, Contract-expand with skip…
- Almost symmetric, many channels at bottom!

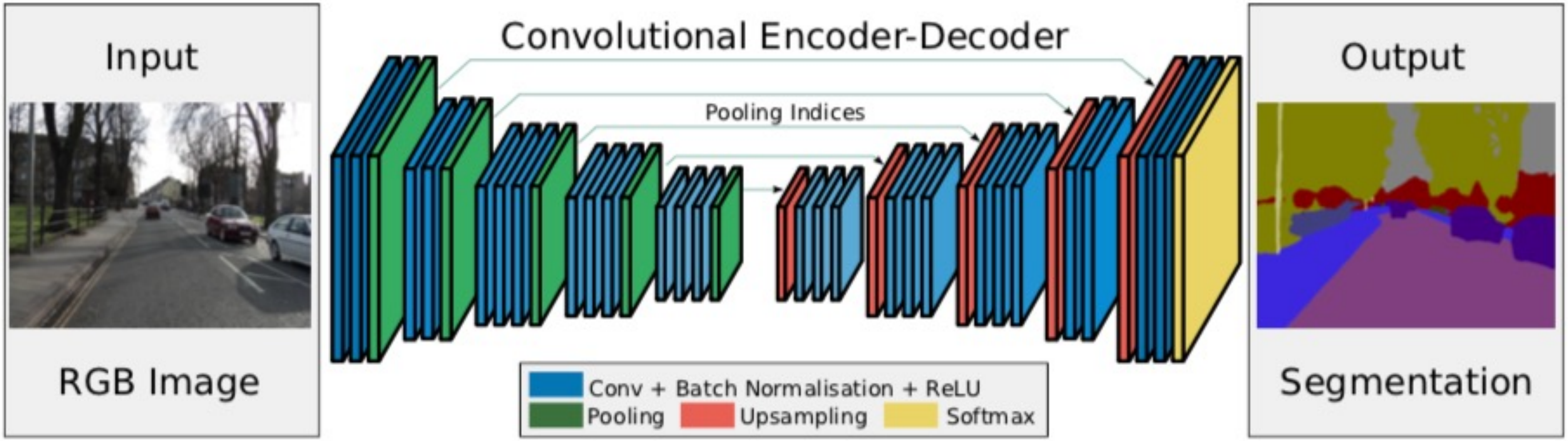
O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *MICCAI*, pp. 234–241, Springer, 2015.

Segnet



Segnet: A deep convolutional encoder-decoder architecture for image segmentation
V Badrinarayanan, A Kendall, R Cipolla - PAMI 2017

Segnet

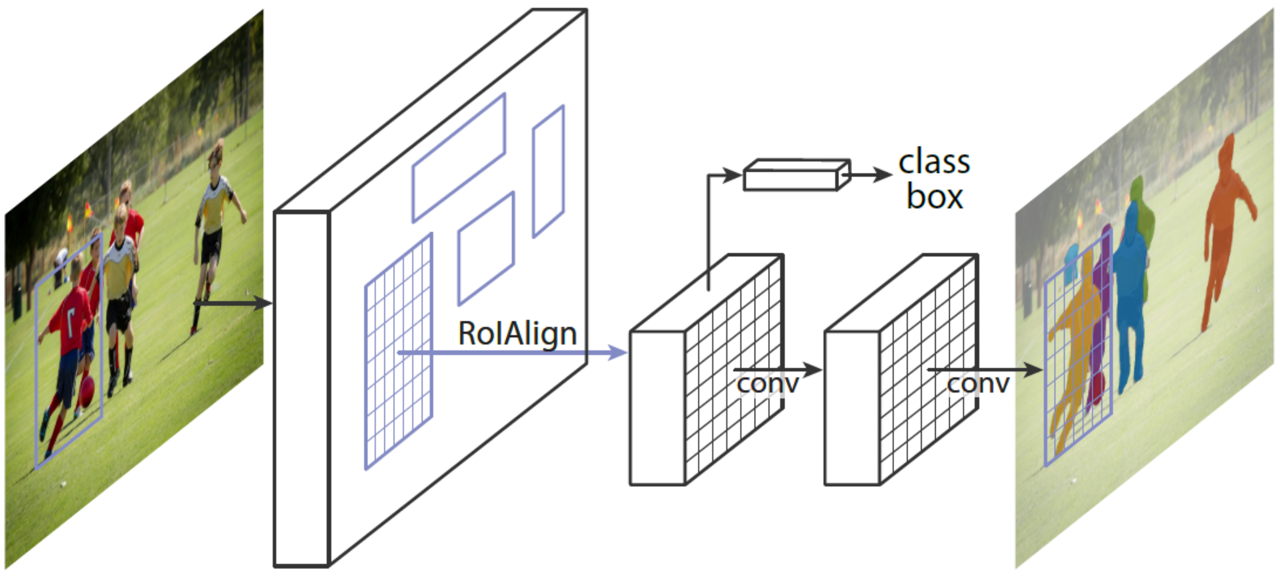


- Eliminates need to learn the upsamplng

Segnet: A deep convolutional encoder-decoder architecture for image segmentation
V Badrinarayanan, A Kendall, R Cipolla - PAMI 2017

Mask-RCNN…

- Neural networks to learn both local feature affinities and top-down context



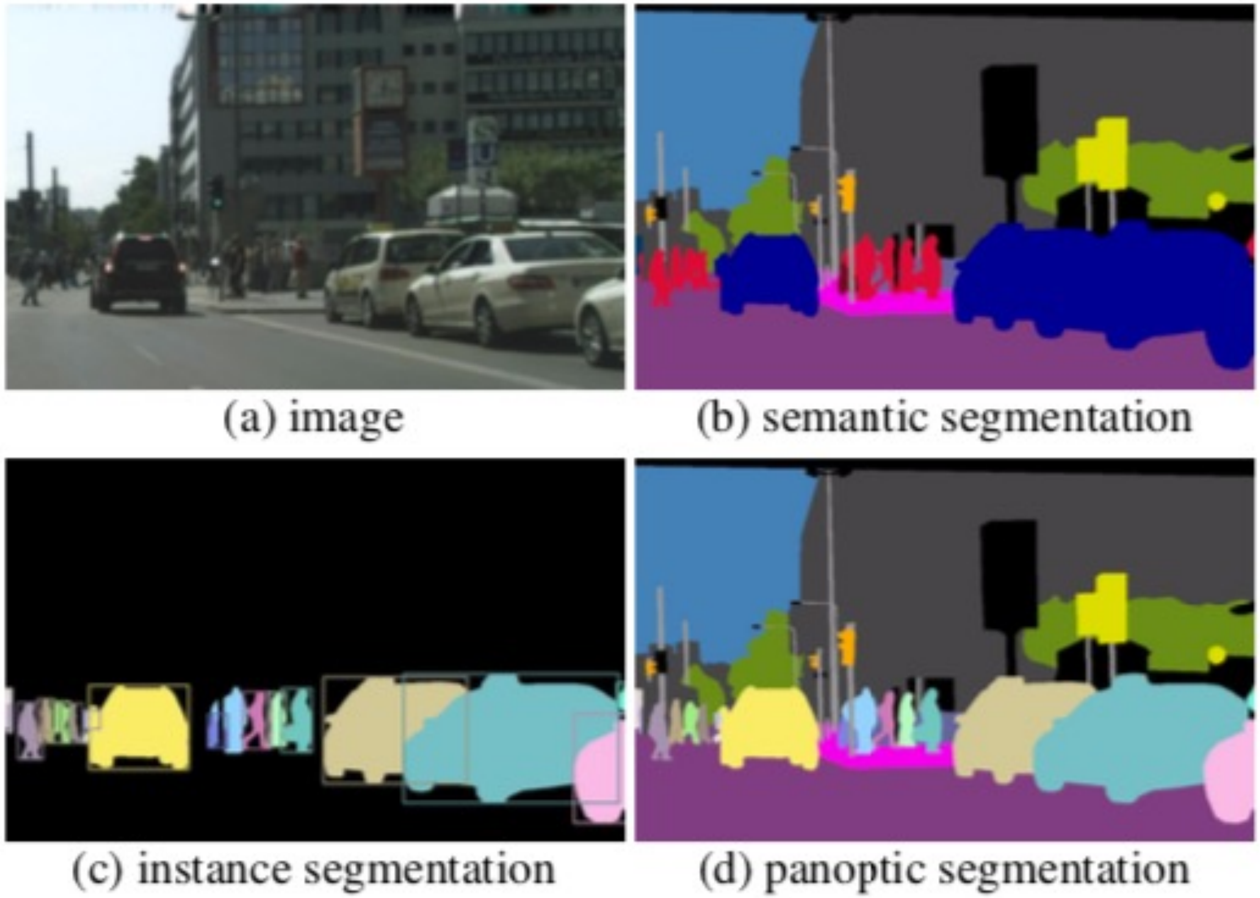
- He et al., [“Mask R-CNN,”] ICCV 2017 (Best paper)

Mask-RCNN…

- Results

- He et al., [“Mask R-CNN,”]() ICCV 2017 (Best paper)

“Panoptic” Segmentation



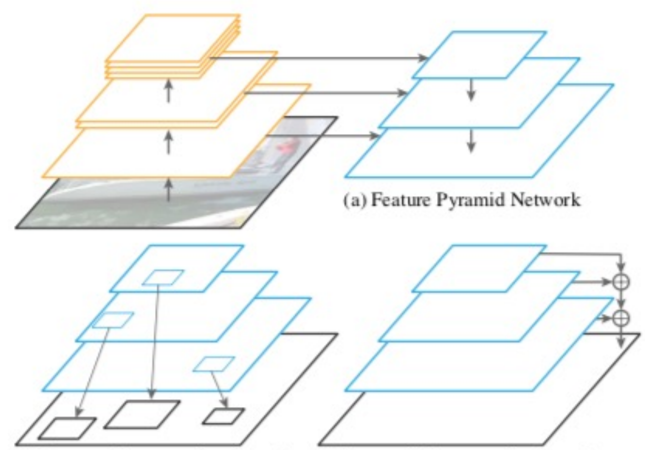
(a) image (b) semantic segmentation (c) instance segmentation (d) panoptic segmentation

- Segnet = semantic segmentation (every pixel)
- Mask-RCNN = instance segmentation (ojects)
- Panoptic = combined

Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic segmentation. In *CVPR*, 2019

Panoptic Feature Pyramid Networks

- Uses FPN architecture
- 2 heads



Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic Feature Pyramid Networks. In *CVPR*, 2019