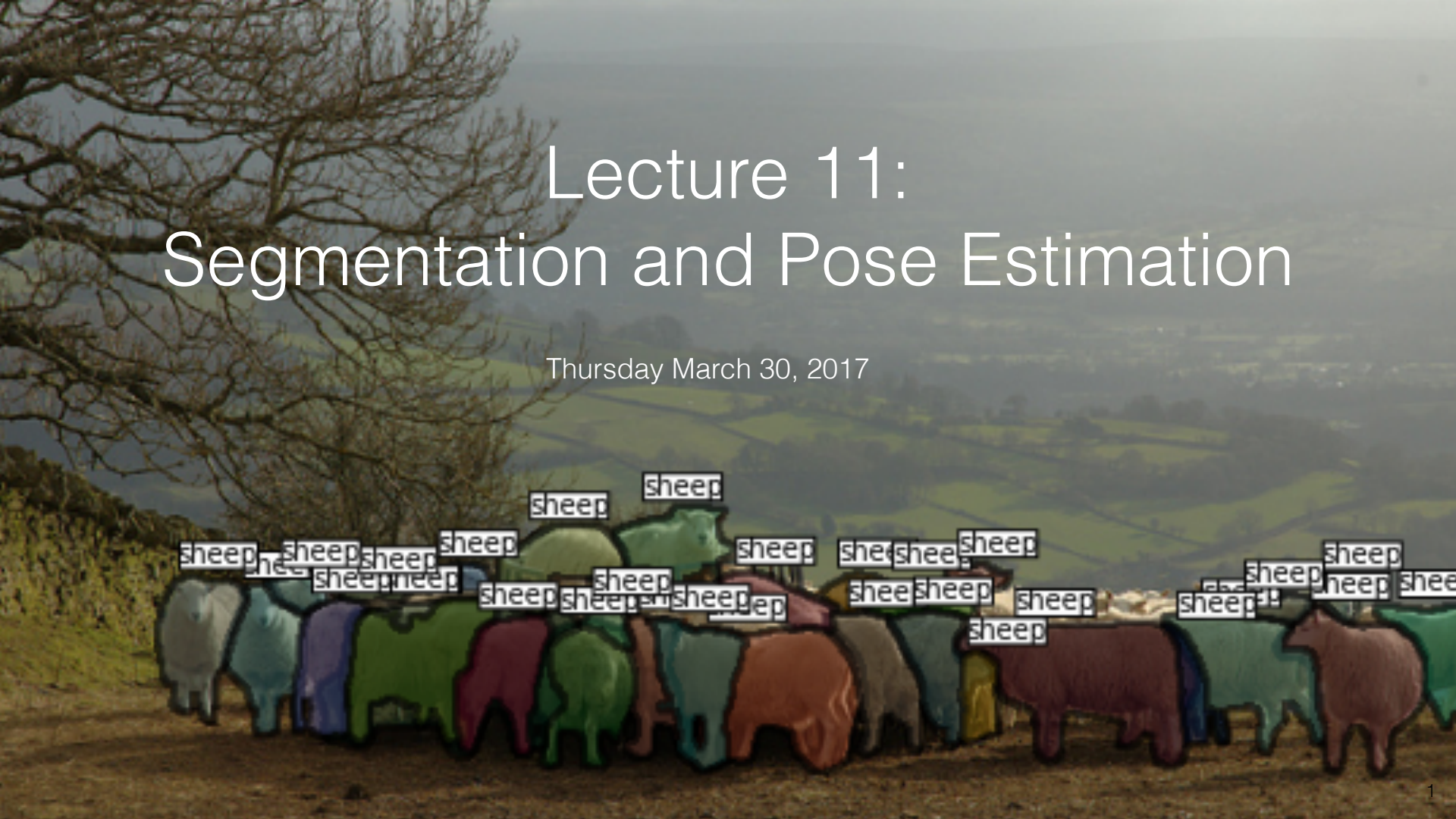
Lecture 11:
Segmentation and Pose Estimation

Thursday March 30, 2017



Announcements!

- HW #3 misunderstandings
    - Deadline for HW #3 re-try is next Thursday **April 6**
- Final Project milestones due next Tuesday **April 4**
- Vote for Final Day and Location

Python/Numpy of the Day

Decorators

How to write a decorator:

```
import time

def timeit(method):

    def timed(*args, **kw):
        ts = time.time()
        result = method(*args, **kw)
        te = time.time()

        print '%r (%r, %r) %2.2f sec' % \
              (method.__name__, args, kw, te-ts)
        return result

    return timed
```

How to use a decorator:

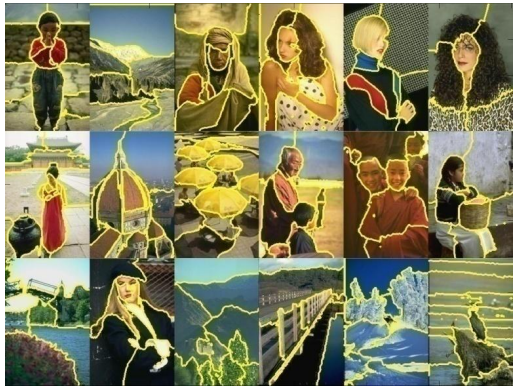
```
class Foo(object):

    @timeit
    def foo(self, a=2, b=3):
        time.sleep(0.2)

@timeit
def f1():
    time.sleep(1)
    print 'f1'
```

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 app.

Slide: Kristin Grauman

Examples of grouping in vision

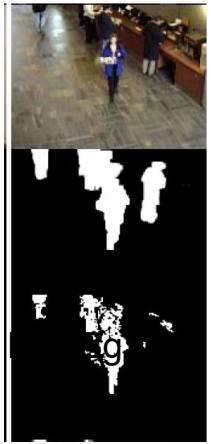


[Figure by J. Shi]

Determine image regions

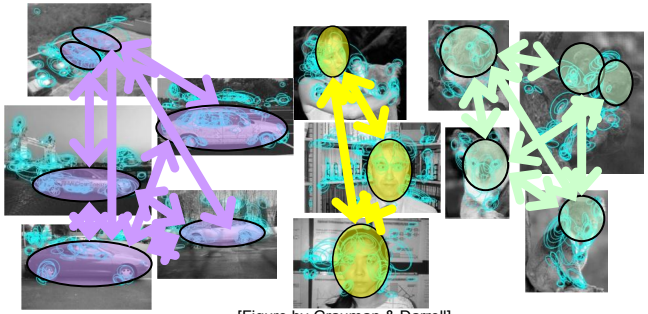
[http://poseidon.csd.auth.gr/LAB_RESEARCH/Latest/imgs/SpeakDepVidIndex_img2.jpg]

Group video frames into shots



[Figure by Wang & Suter]

Figure-ground



[Figure by Grauman & Darrell]

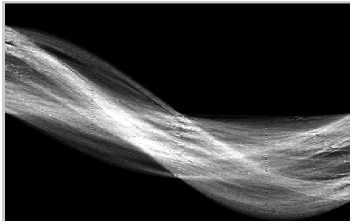
Object-level grouping

Slide: Kristin Grauman

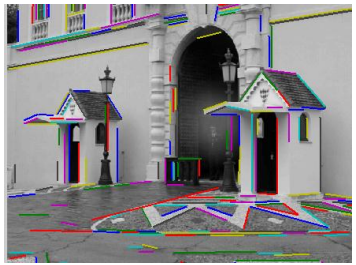
Edge and line detection

- Canny edge detector = smooth → derivative → thin → threshold → link

- Generalized Hough transform = points vote for shape parameters



- Straight line detector = canny + gradient orientations → orientation binning → linking → check for straightness



Slide: Derek Hoiem

Segmentation

Computer Vision Tasks

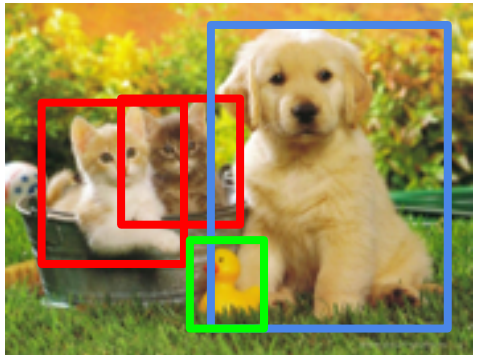
Classification

CAT

Classification + Localization

CAT

Object Detection



CAT, DOG, DUCK

Segmentation

CAT, DOG, DUCK

Single object

Multiple objects

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

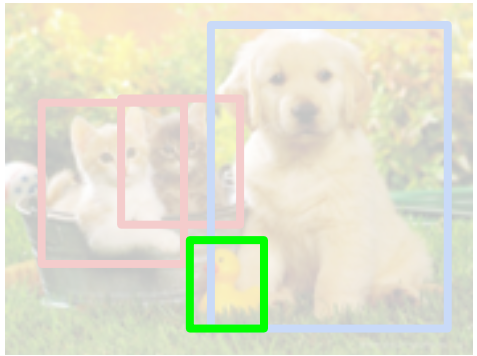
Computer Vision Tasks

Classification

Classification + Localization

Object Detection

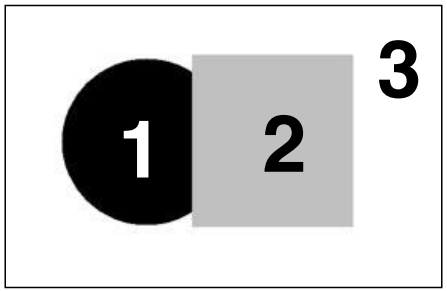
Segmentation



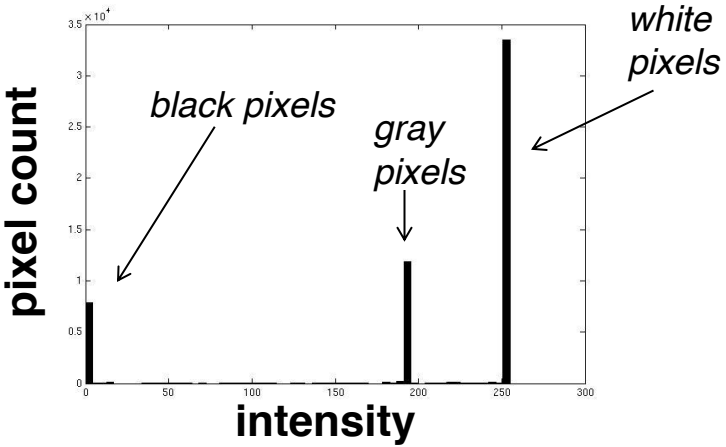
Today

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

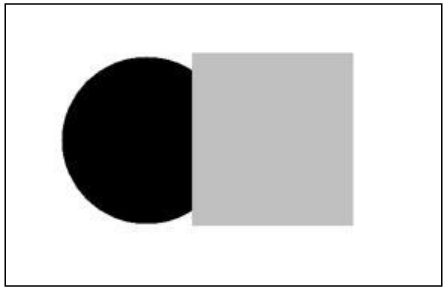
Image segmentation: toy example



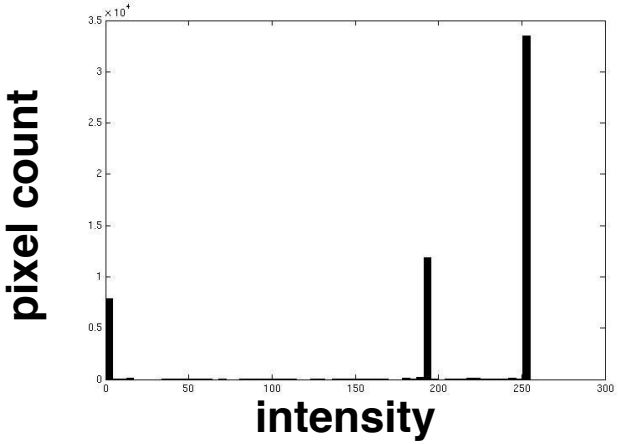
input image

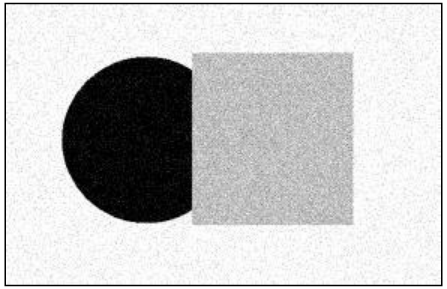


- These intensities define the three groups.
- We could label every pixel in the image according to which of these primary intensities it is.
  - i.e., *segment* the image based on the intensity feature.
- What if the image isn't quite so simple?

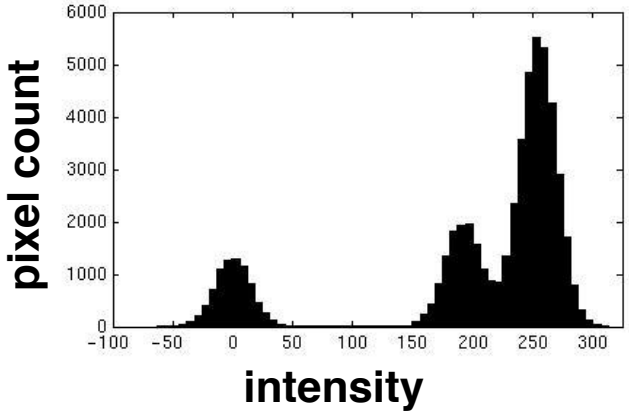


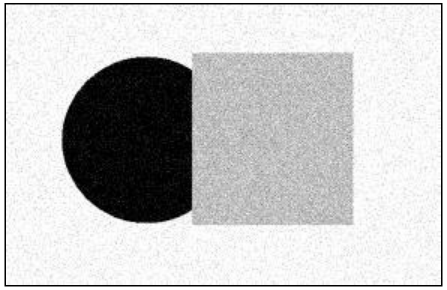
input image



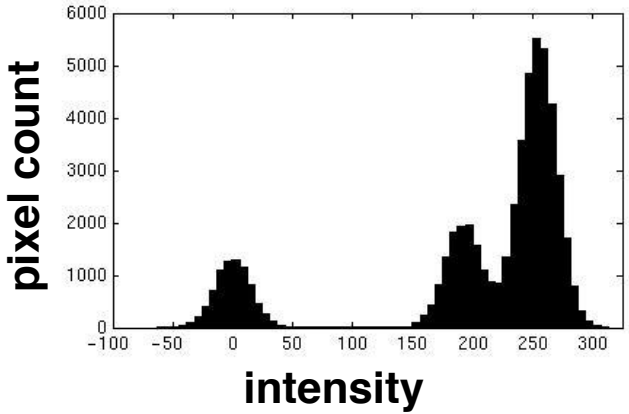


input image





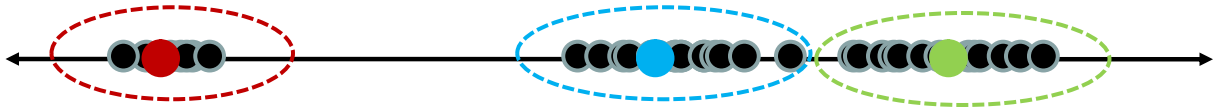
input image



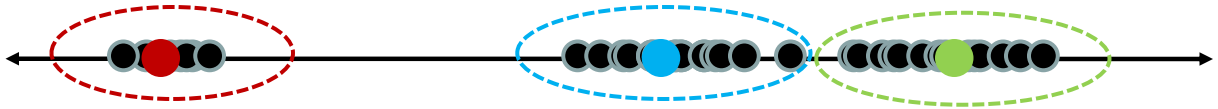
- Now how to determine the three main intensities that define our groups?
- We need to ***cluster.***

Clustering

- With this objective, it is a “chicken and egg” problem:
  - If we knew the **cluster centers**, we could allocate points to groups by assigning each to its closest center.

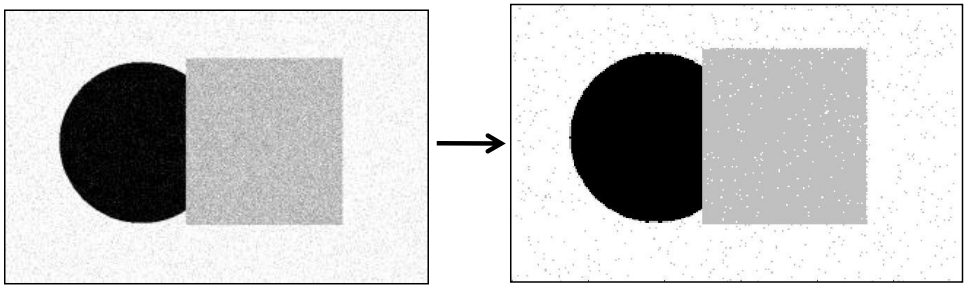


  - If we knew the **group memberships**, we could get the centers by computing the mean per group.



Smoothing out cluster assignments

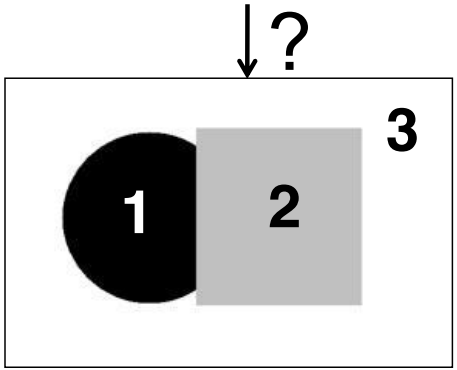
- Assigning a cluster label per pixel may yield outliers:



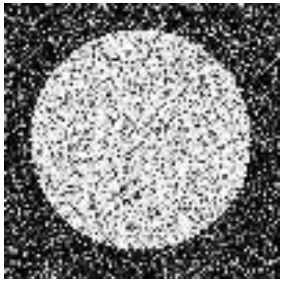
original

labeled by cluster center's intensity

- How to ensure they are spatially smooth?



Solution



P(foreground | image)

Encode dependencies between pixels

Normalizing constant

$$P(\mathbf{y};\theta,data) = \frac{1}{Z}\prod_{i=1..N} f_1(y_i;\theta,data) \prod_{i,j\in edges} f_2(y_i,y_j;\theta,data)$$

Labels to be predicted

Individual predictions

Pairwise predictions

Slide: Derek Hoiem

Writing Likelihood as an "Energy"

$$P(\mathbf{y};\theta,data)=\frac{1}{Z}\prod_{i=1..N}p_1(y_i;\theta,data)\prod_{i,j\in edges}p_2(y_i,y_j;\theta,data)$$

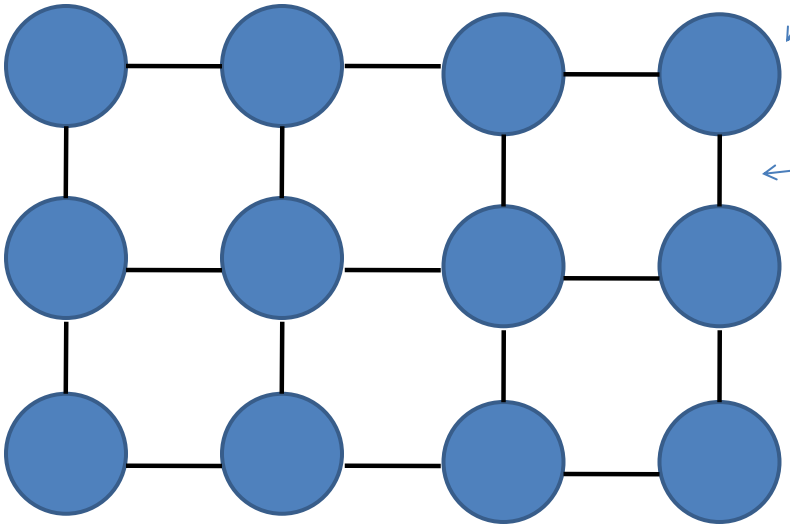
$$Energy(\mathbf{y};\theta,data)=\sum_i\psi_1(y_i;\theta,data)+\sum_{i,j\in edges}\psi_2(y_i,y_j;\theta,data)$$

"Cost" of assignment y_i

"Cost" of pairwise assignment y_i, y_j

Slide: Derek Hoiem

Markov Random Fields



Cost to assign a label to each pixel

Cost to assign a pair of labels to connected pixels

$$Energy(\mathbf{y};\theta,data) = \sum_i \psi_1(y_i;\theta,data) + \sum_{i,j\in edges} \psi_2(y_i,y_j;\theta,data)$$

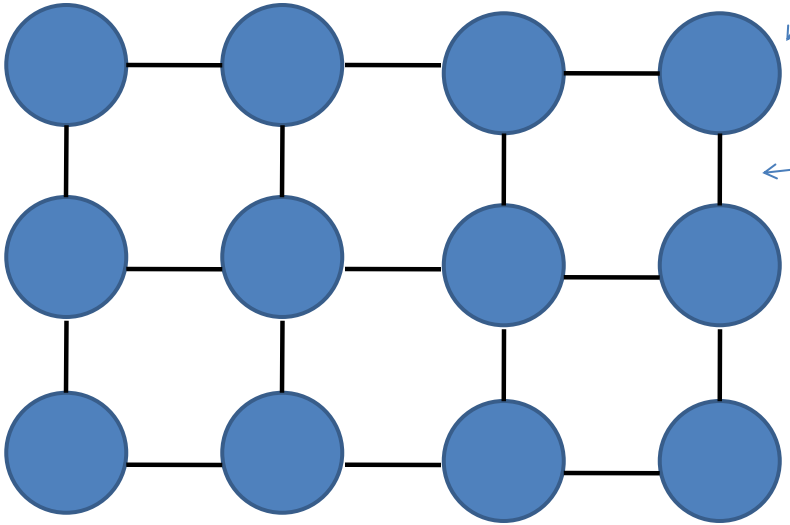
Slide: Derek Hoiem

Markov Random Fields

- Example: “label smoothing” grid

Unary potential

0: -logP(y_i = 0 ; data)
1: -logP(y_i = 1 ; data)



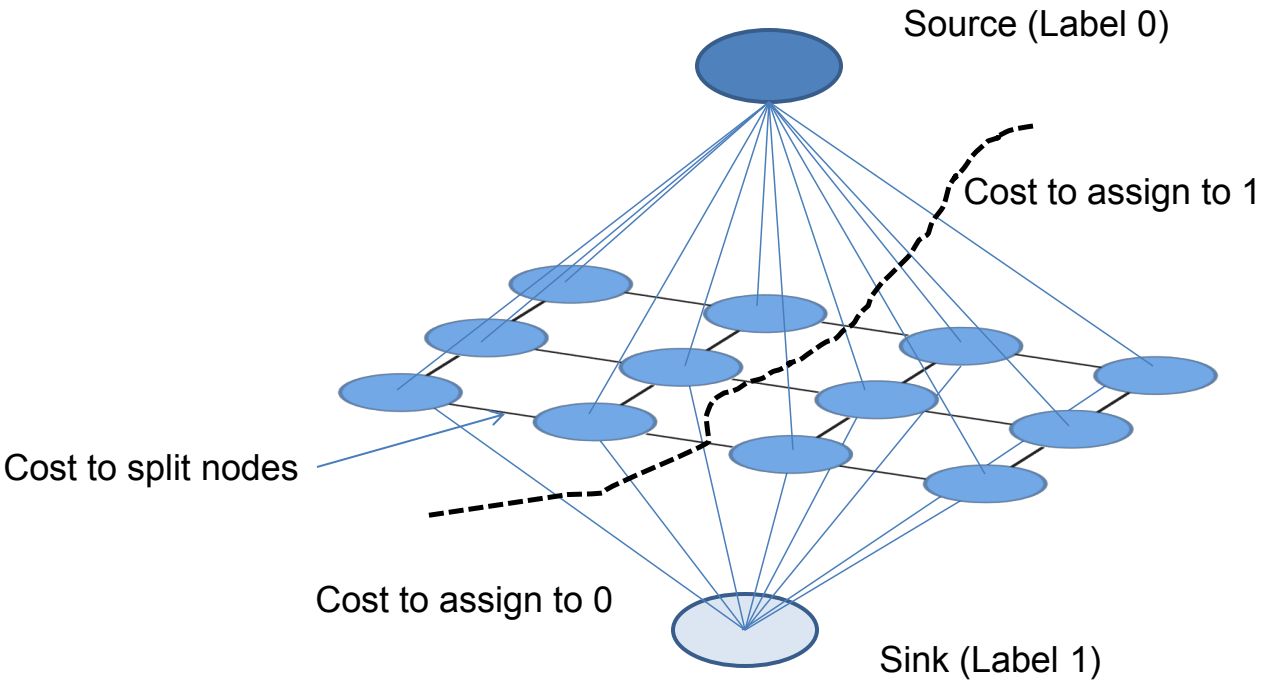
Pairwise Potential

	0	1
0	0	K
1	K	0

$$Energy(\mathbf{y};\theta,data) = \sum_i \psi_1(y_i;\theta,data) + \sum_{i,j \in edges} \psi_2(y_i,y_j;\theta,data)$$

Slide: Derek Hoiem

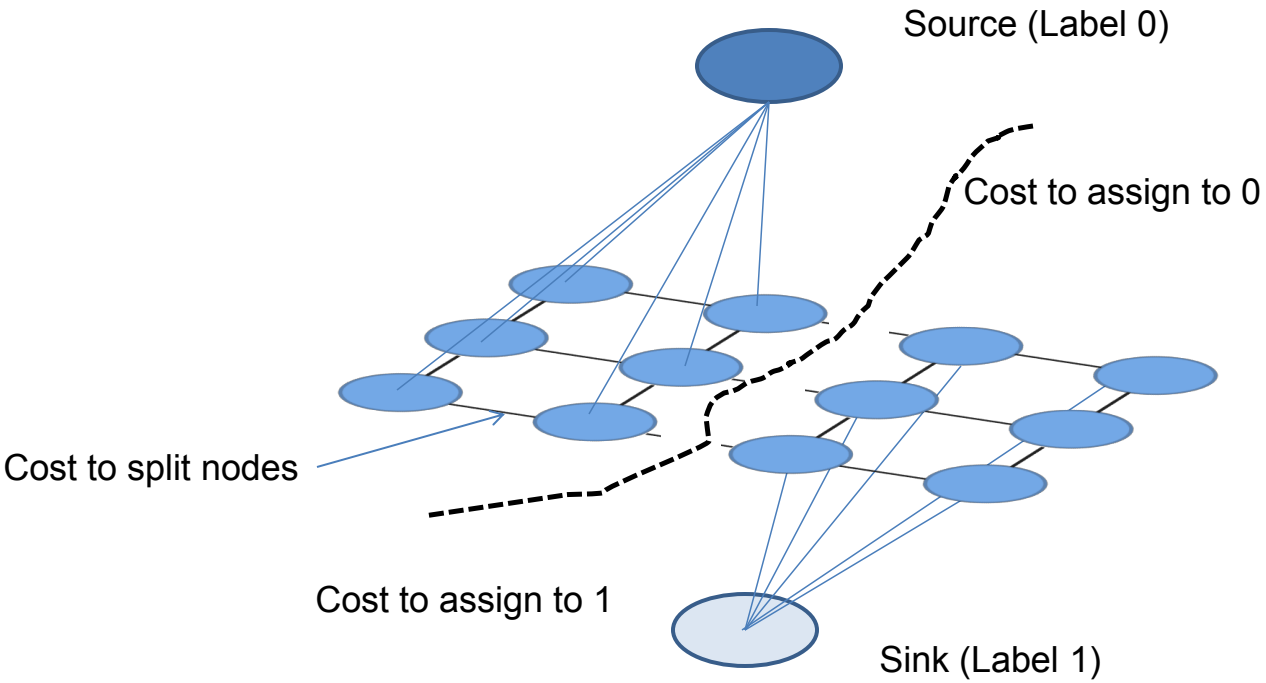
Solving MRFs with graph cuts



$$Energy(\mathbf{y};\theta,data) = \sum_i \psi_1(y_i;\theta,data) + \sum_{i,j\in edges} \psi_2(y_i,y_j;\theta,data)$$

Slide: Derek Hoiem

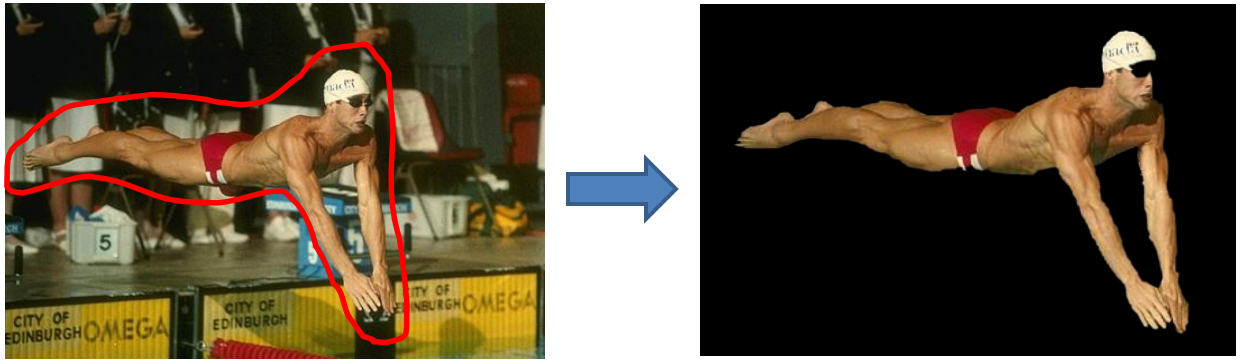
Solving MRFs with graph cuts



$$Energy(\mathbf{y};\theta,data) = \sum_i \psi_1(y_i;\theta,data) + \sum_{i,j\in edges} \psi_2(y_i,y_j;\theta,data)$$

Slide: Derek Hoiem

GrabCut segmentation

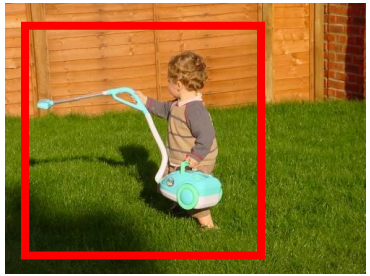


User provides rough indication of foreground region.

Goal: Automatically provide a pixel-level segmentation.

Slide: Derek Hoiem

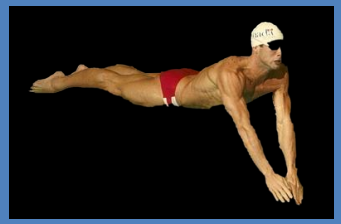
What is easy or hard about these cases for graphcut-based segmentation?



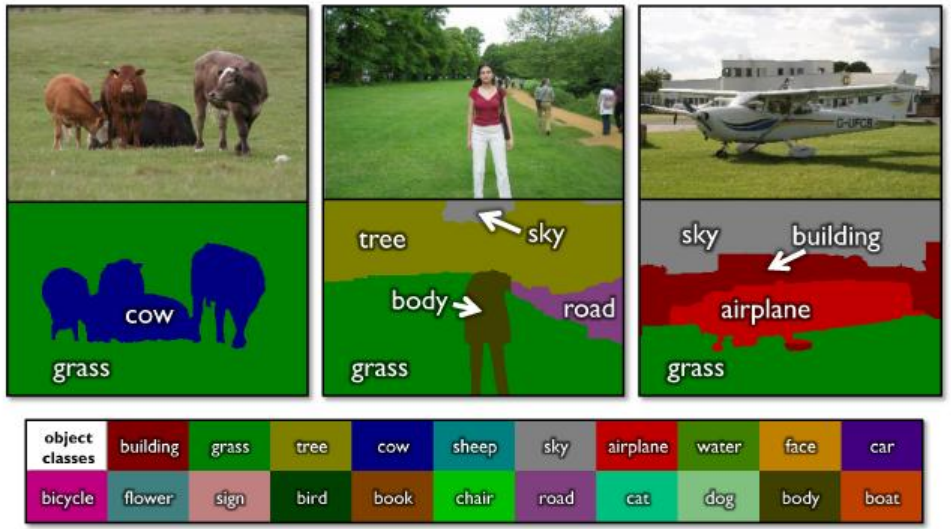
Slide: Derek Hoiem

Easier examples

More difficult Examples

	Camouflage & Low Contrast	Fine structure	Harder Case
Initial Rectangle			
Initial Result			

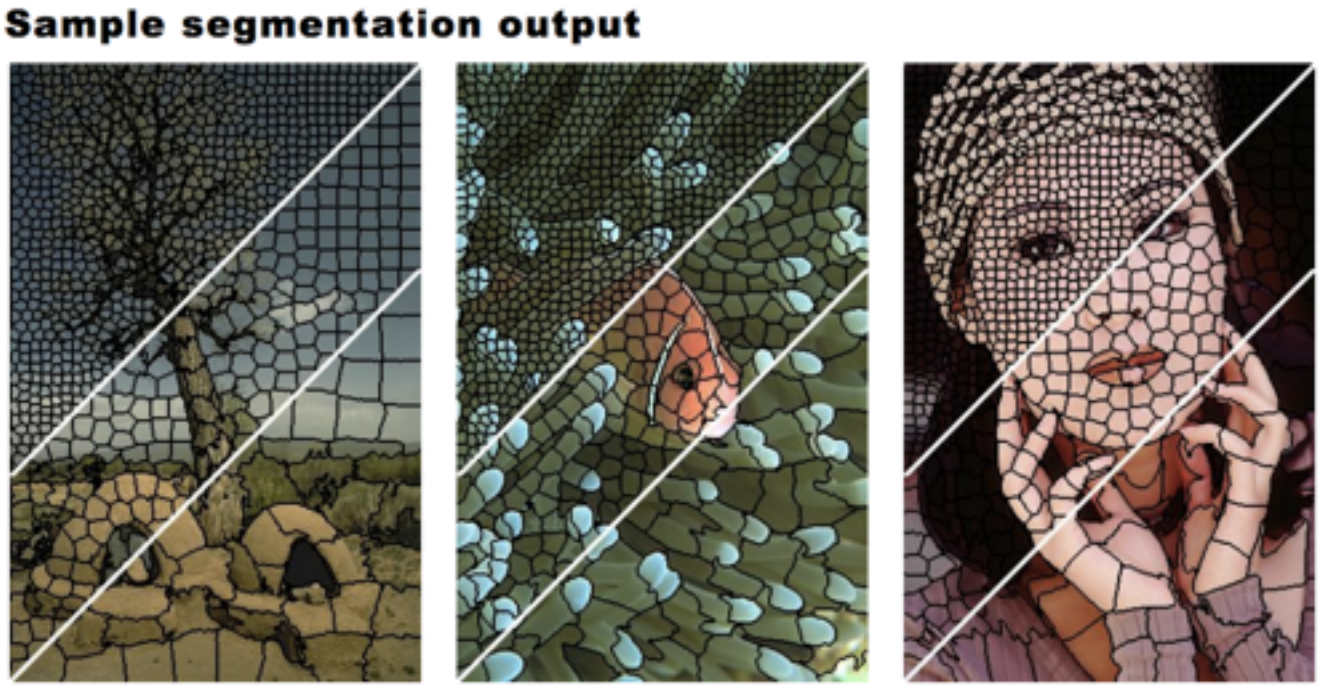
Using graph cuts for recognition



TextonBoost (Shotton et al. 2009 IJCV)

Unsupervised Segmentation

- Classic Example: **Superpixels**
- Cluster pixels with their neighbors
- Break clusters when large gradient occurs between neighboring pixels
  - Figure from Achanta et al. — “SLIC Superpixels Compared to State-of-the-art Superpixel Methods,” May 2012



Further reading and resources

- Graph cuts
  - http://www.cs.cornell.edu/~rdz/graphcuts.html
  - Classic paper: What Energy Functions can be Minimized via Graph Cuts? (Kolmogorov and Zabih, ECCV '02/PAMI '04)

- Belief propagation

  Yedidia, J.S.; Freeman, W.T.; Weiss, Y., "Understanding Belief Propagation and Its Generalizations", Technical Report, 2001: http://www.merl.com/publications/TR2001-022/

- Normalized cuts and image segmentation (Shi and Malik)

  *http://www.cs.berkeley.edu/~malik/papers/SM-n**cut**.pdf*

- N-cut implementation

  http://www.seas.upenn.edu/~timothee/software/ncut/ncut.html

Slide: Derek Hoiem


Touristbus
Amsterdam


Touristbus
Amsterdam

Semantic Segmentation

Label every pixel!

Don’t differentiate instances (cows)

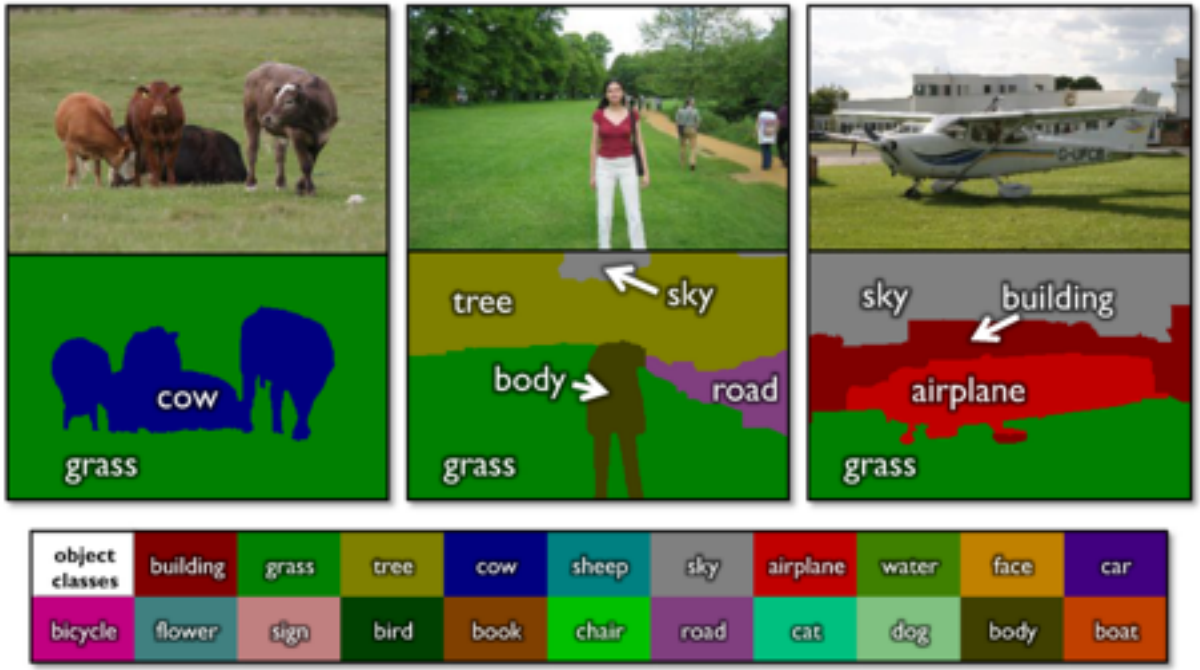


Figure credit: Shotton et al, “TextonBoost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context”, IJCV 2007

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Instance Segmentation

Detect instances, give category, label pixels

“simultaneous detection and segmentation” (SDS)

Lots of recent work (MS-COCO Challenges)

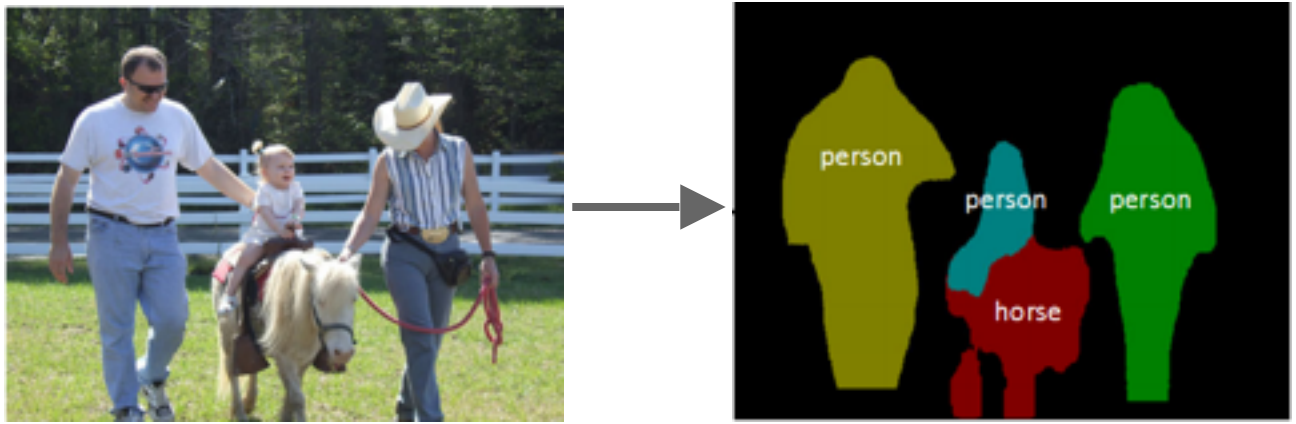


Figure credit: Dai et al, “Instance-aware Semantic Segmentation via Multi-task Network Cascades”, arXiv 2015

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Semantic Segmentation

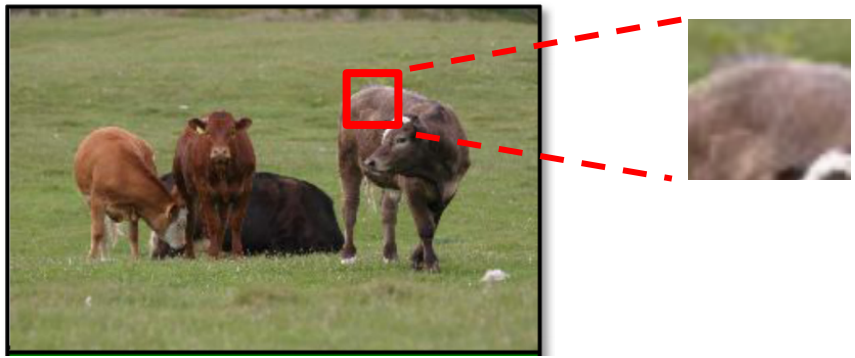
* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Semantic Segmentation

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

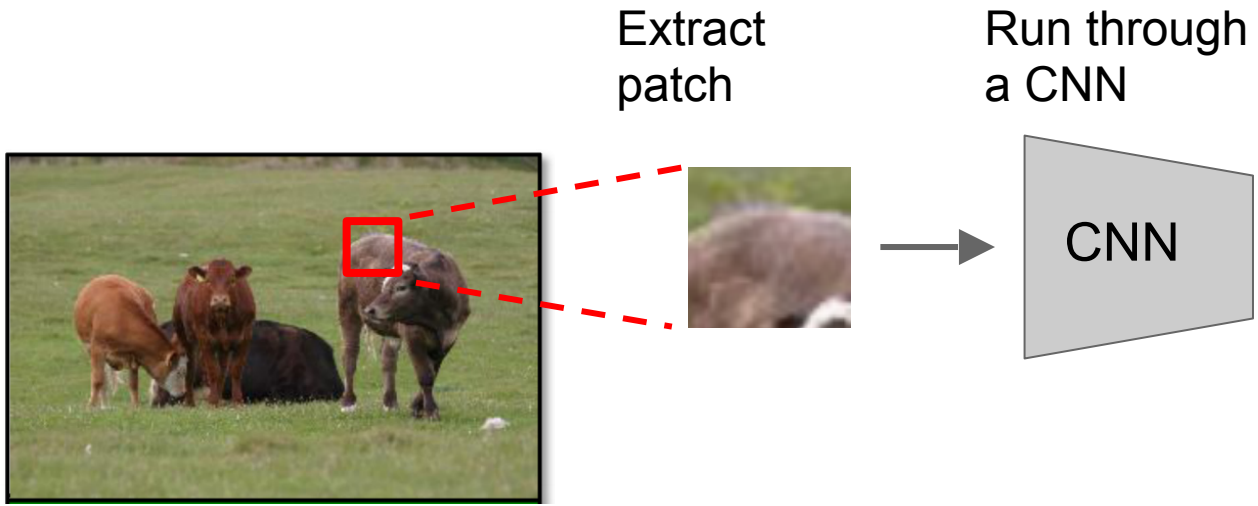
Semantic Segmentation

Extract patch



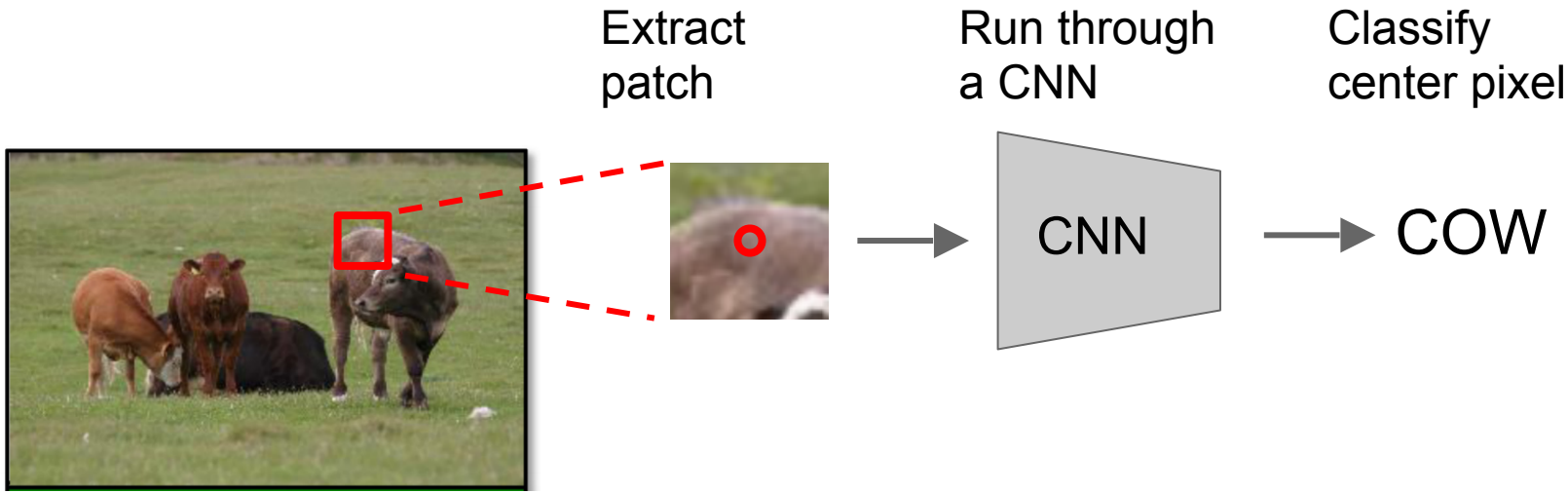
* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Semantic Segmentation



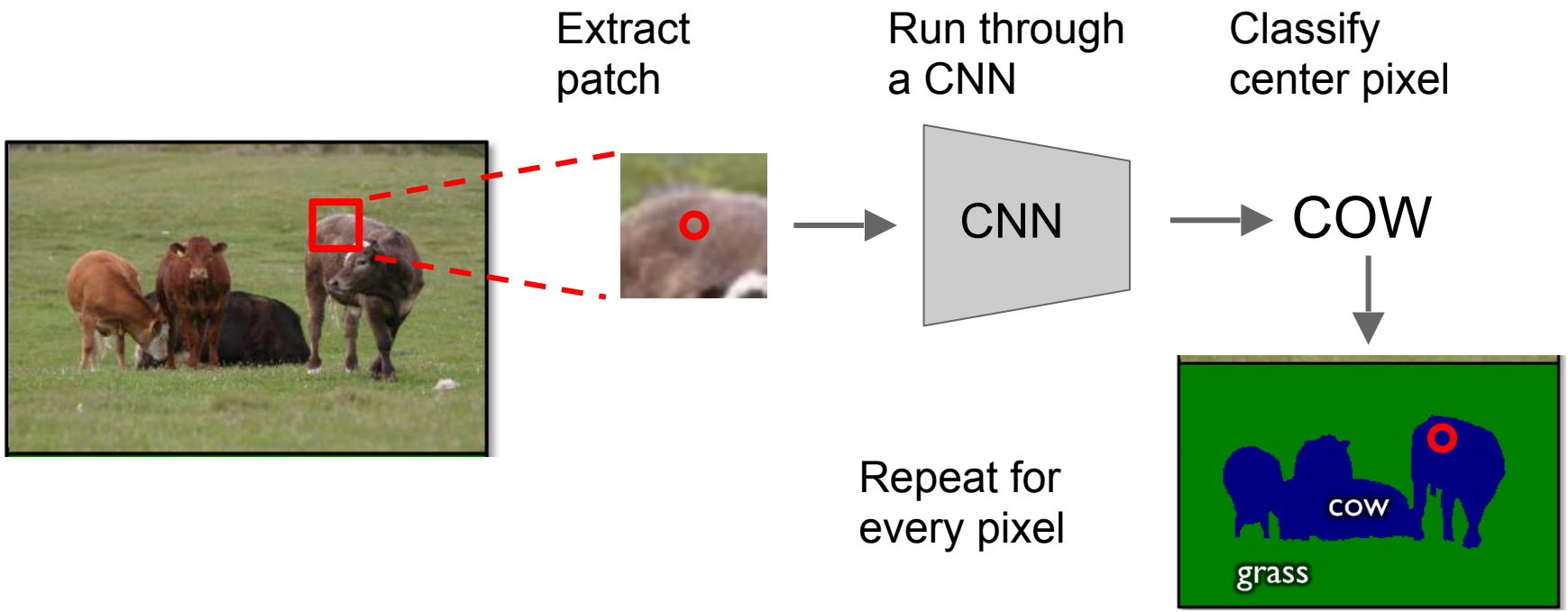
* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Semantic Segmentation



* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

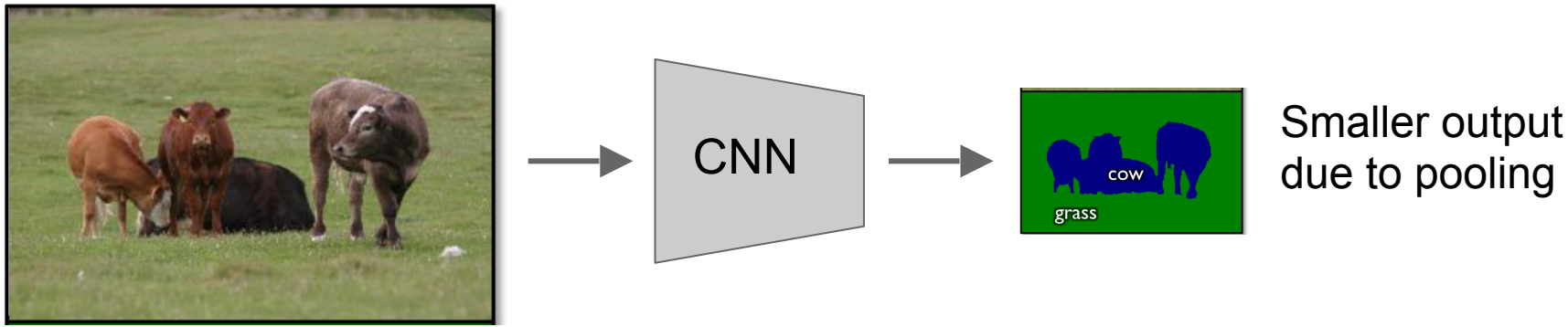
Semantic Segmentation



* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Semantic Segmentation

Run “fully convolutional” network
to get all pixels at once



Smaller output
due to pooling

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

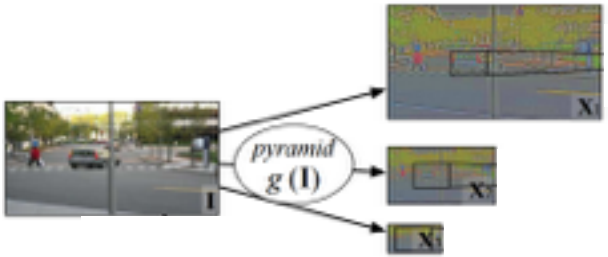
Semantic Segmentation: Multi-Scale

Farabet et al, “Learning Hierarchical Features for Scene Labeling,” TPAMI 2013

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Semantic Segmentation: Multi-Scale

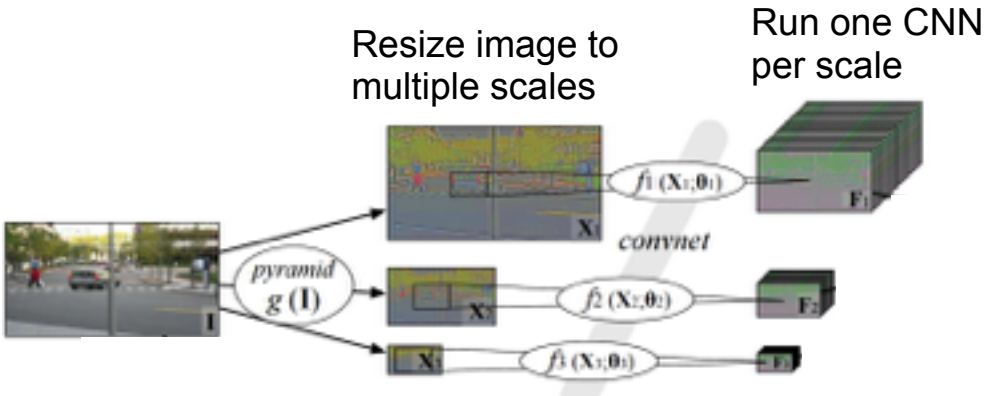
Resize image to multiple scales



Farabet et al, “Learning Hierarchical Features for Scene Labeling,” TPAMI 2013

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

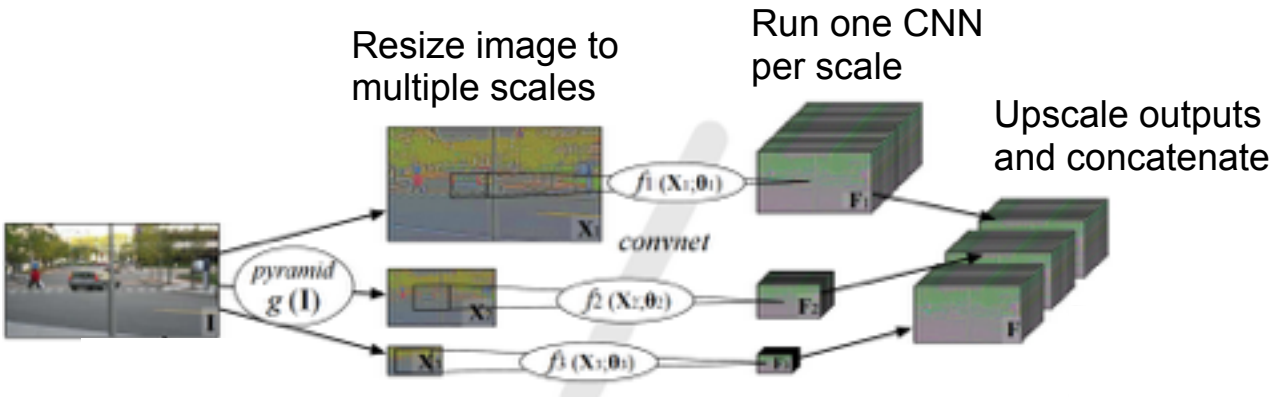
Semantic Segmentation: Multi-Scale



Farabet et al, “Learning Hierarchical Features for Scene Labeling,” TPAMI 2013

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

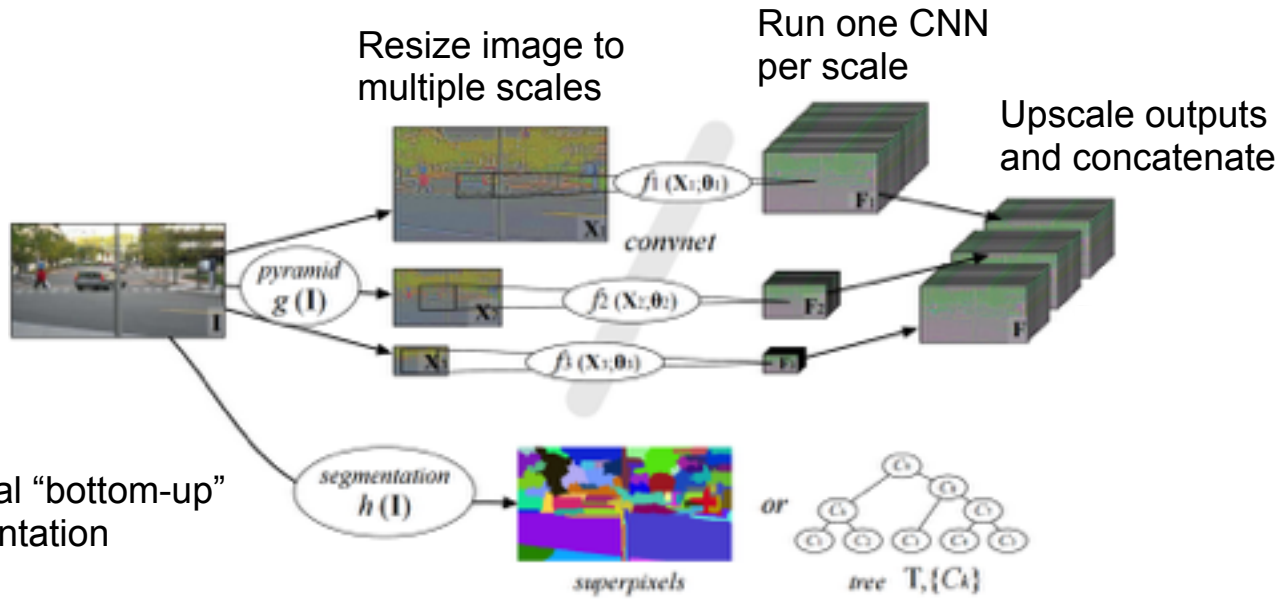
Semantic Segmentation: Multi-Scale



Farabet et al, “Learning Hierarchical Features for Scene Labeling,” TPAMI 2013

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

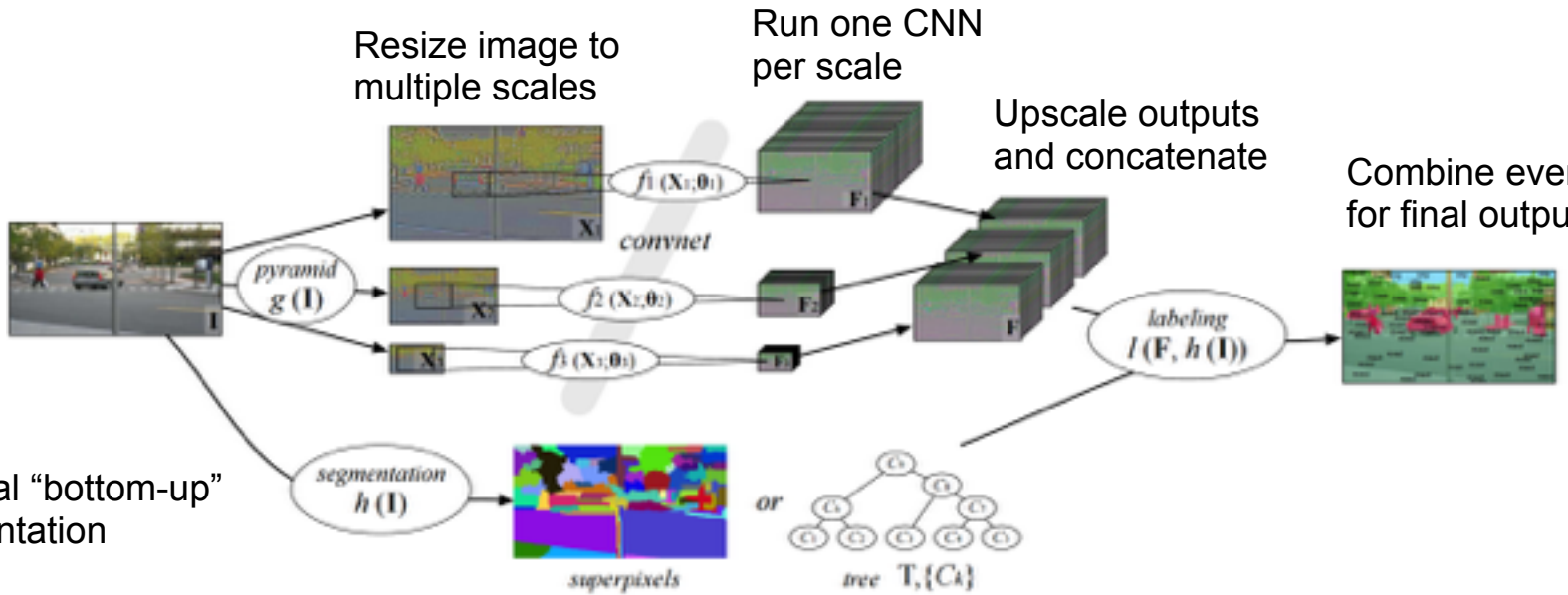
Semantic Segmentation: Multi-Scale



Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Semantic Segmentation: Multi-Scale

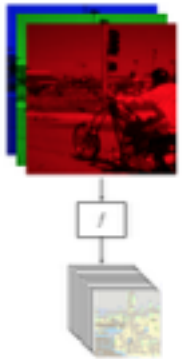


Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Semantic Segmentation: Refinement

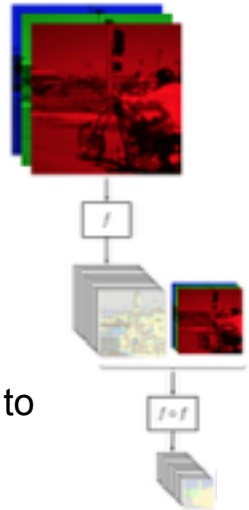
Apply CNN once
to get labels



Pinheiro and Collobert, “Recurrent Convolutional Neural Networks for Scene Labeling”, ICML 2014

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Semantic Segmentation: Refinement



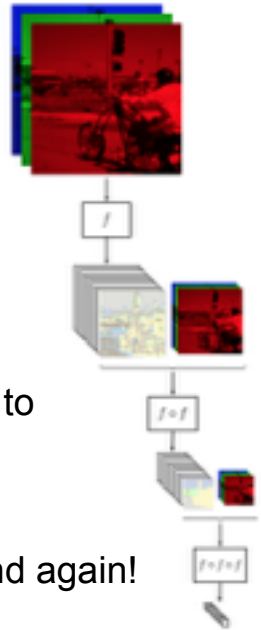
Apply CNN once to get labels

Apply AGAIN to refine labels

Pinheiro and Collobert, “Recurrent Convolutional Neural Networks for Scene Labeling”, ICML 2014

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Semantic Segmentation: Refinement



Apply CNN once to get labels

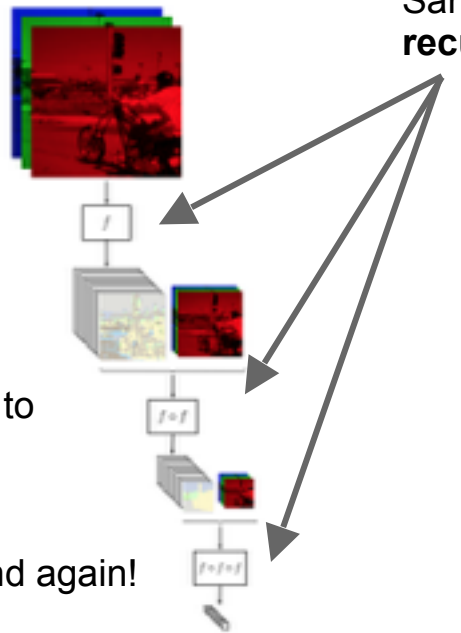
Apply AGAIN to refine labels

And again!

Pinheiro and Collobert, “Recurrent Convolutional Neural Networks for Scene Labeling”, ICML 2014

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Semantic Segmentation: Refinement



Same CNN weights:
recurrent convolutional network

Apply CNN once
to get labels

Apply AGAIN to
refine labels

And again!

Pinheiro and Collobert, “Recurrent Convolutional Neural Networks for Scene Labeling”, ICML 2014

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

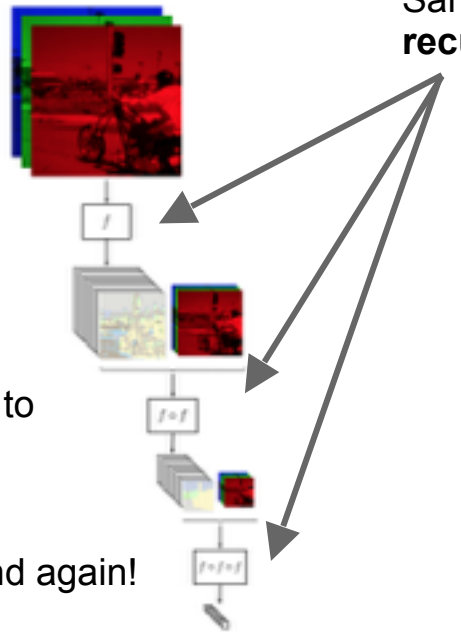
Semantic Segmentation: Refinement

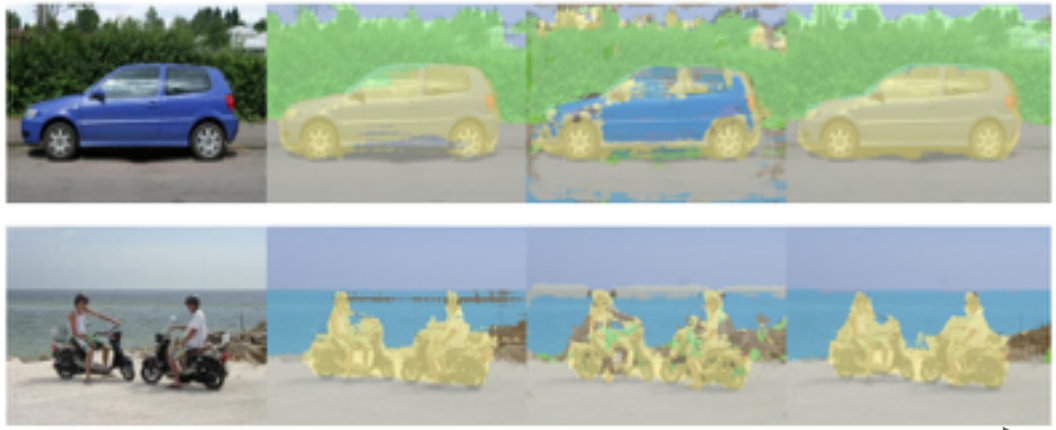
Same CNN weights:
recurrent convolutional network

Apply CNN once to get labels

Apply AGAIN to refine labels

And again!



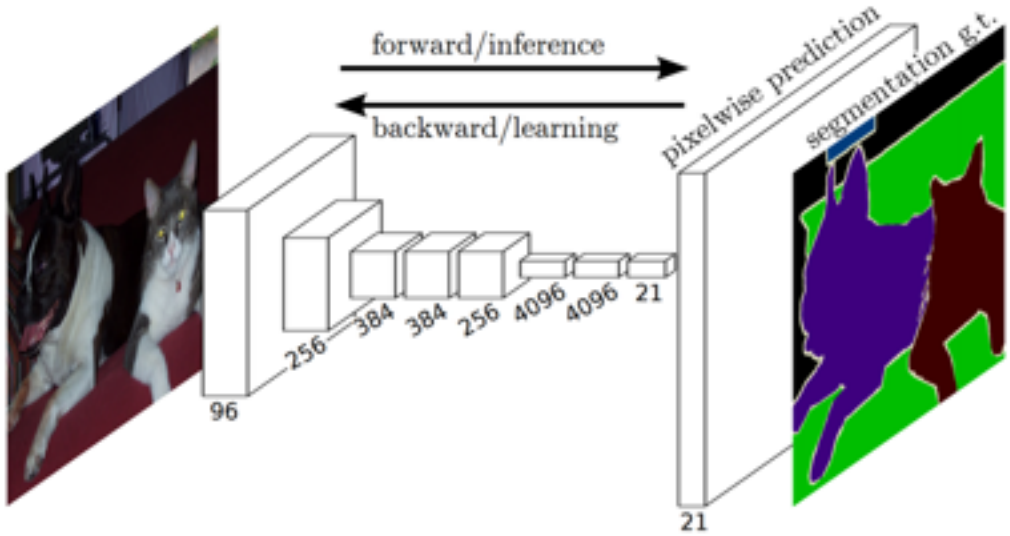


More iterations improve results

Pinheiro and Collobert, “Recurrent Convolutional Neural Networks for Scene Labeling”, ICML 2014

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

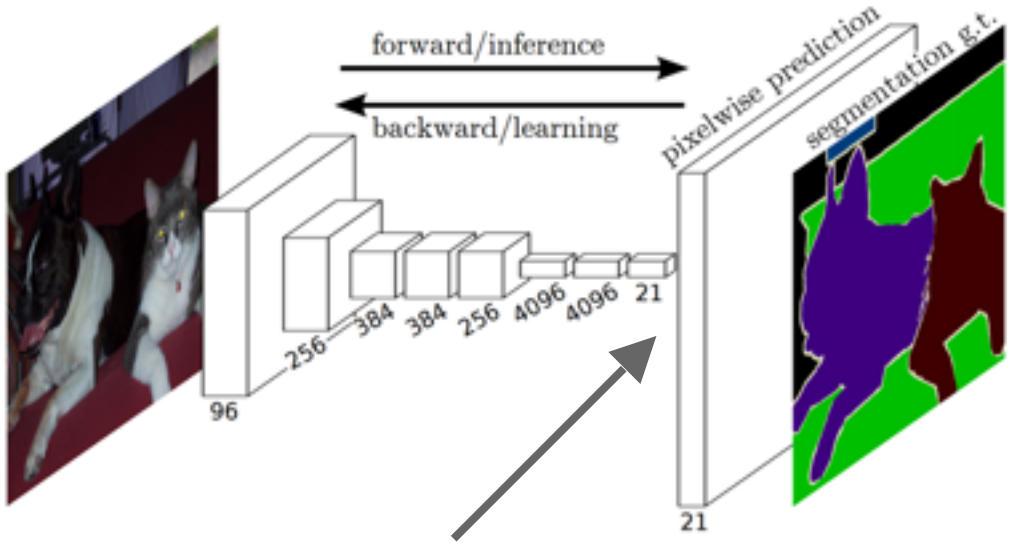
Semantic Segmentation: Upsampling



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

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Semantic Segmentation: Upsampling

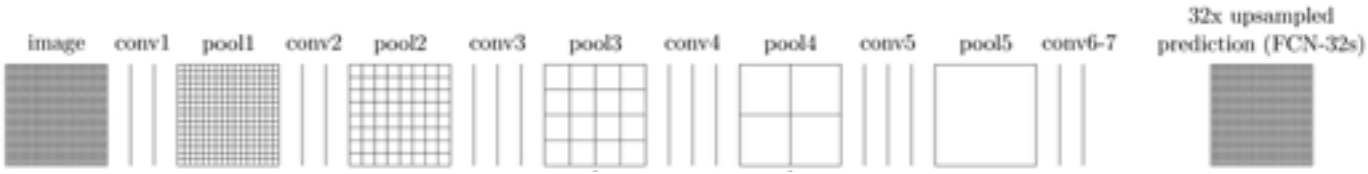


Learnable upsampling!

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

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

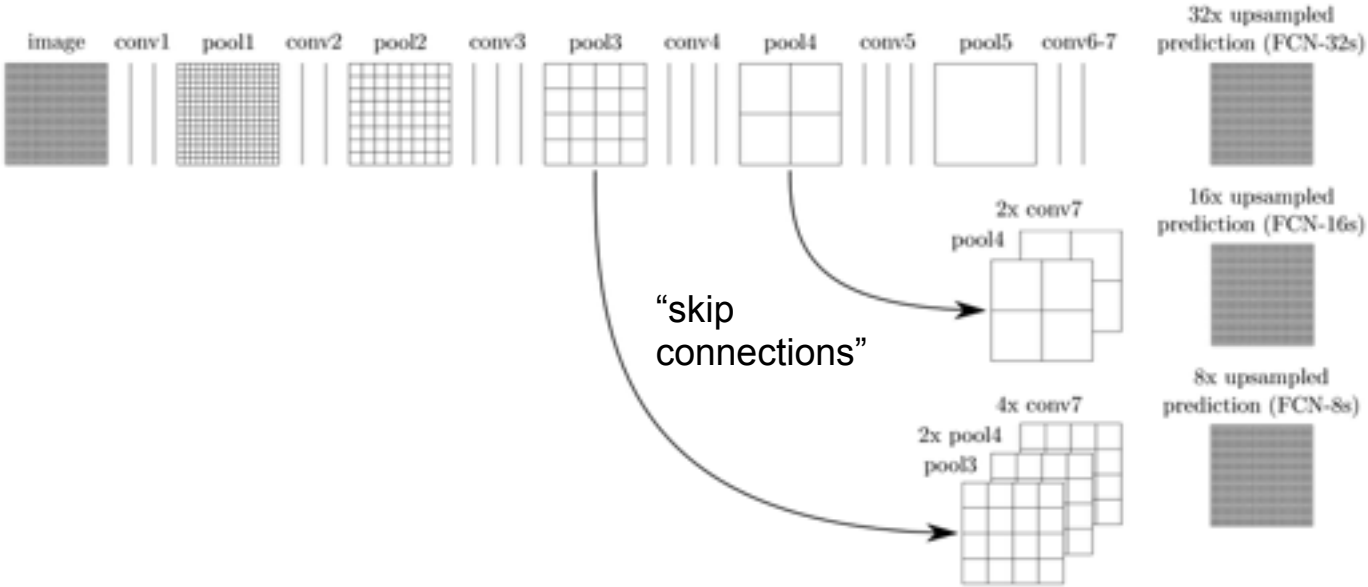
Semantic Segmentation: Upsampling



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

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

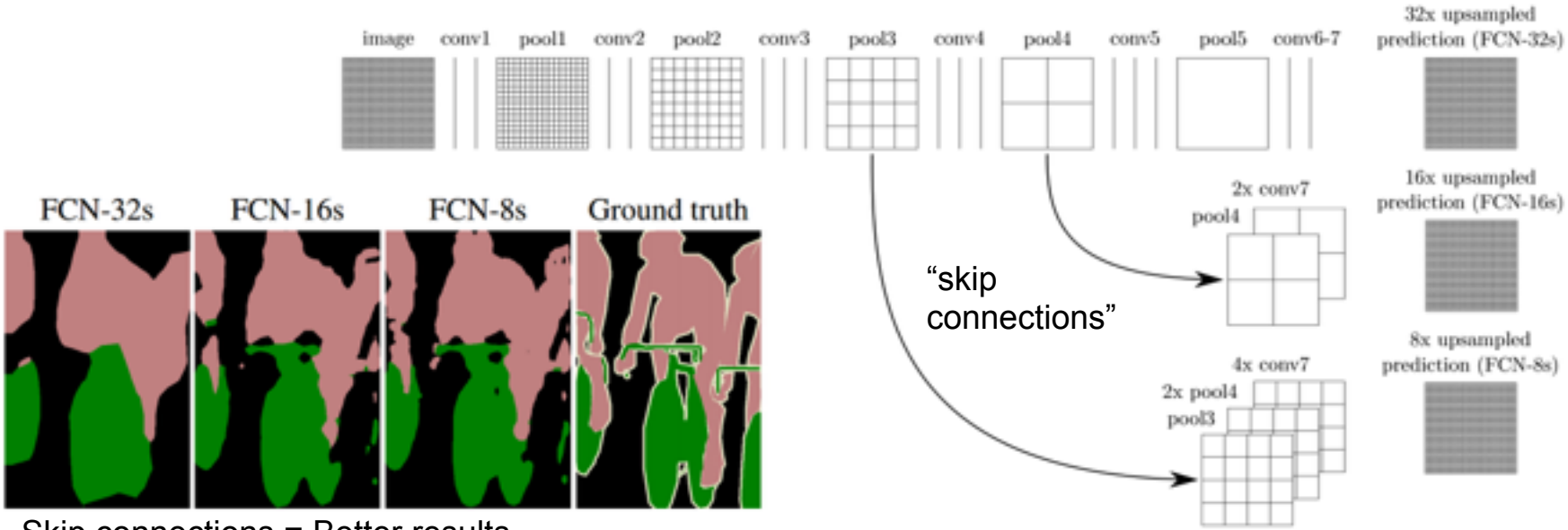
Semantic Segmentation: Upsampling



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

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Semantic Segmentation: Upsampling



Skip connections = Better results

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

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Learnable Upsampling: “Deconvolution”

Typical 3 x 3 convolution, stride 1 pad 1

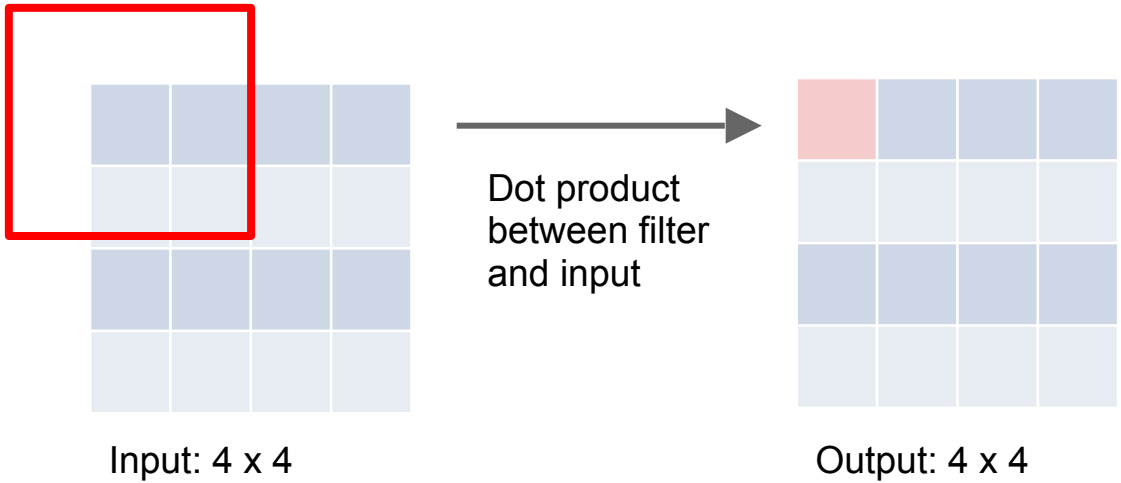
Input: 4 x 4

Output: 4 x 4

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Learnable Upsampling: “Deconvolution”

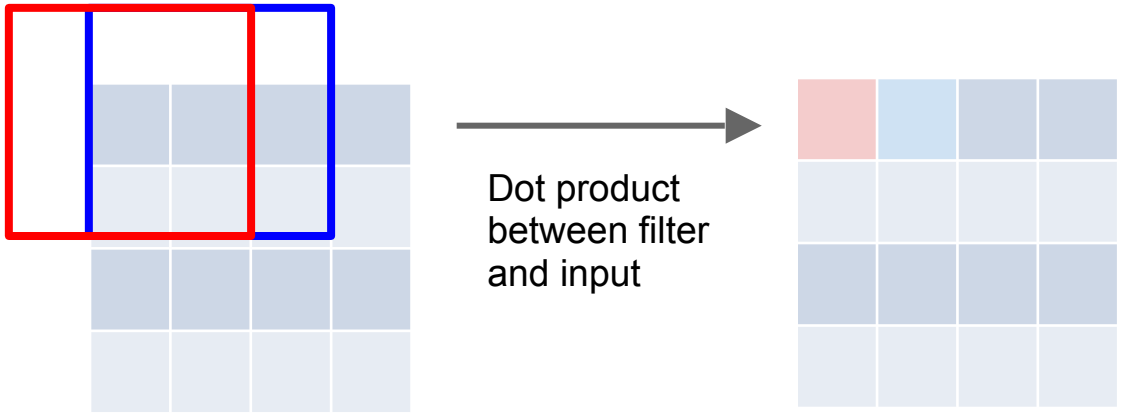
Typical 3 x 3 convolution, stride 1 pad 1



* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Learnable Upsampling: “Deconvolution”

Typical 3 x 3 convolution, stride 1 pad 1



* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Learnable Upsampling: “Deconvolution”

Typical 3 x 3 convolution, **stride 2** pad 1

Input: 4 x 4

Output: 2 x 2

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

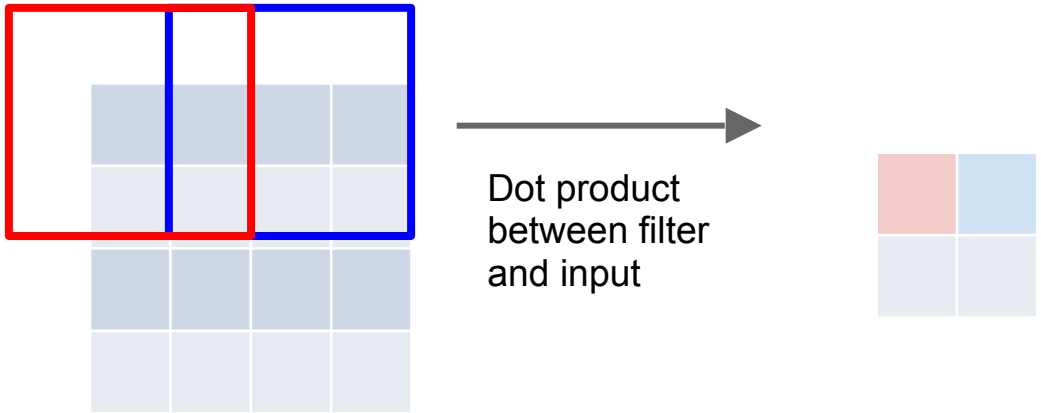
Learnable Upsampling: “Deconvolution”

Typical 3 x 3 convolution, stride 2 pad 1

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Learnable Upsampling: “Deconvolution”

Typical 3 x 3 convolution, stride 2 pad 1



Input: 4 x 4

Output: 2 x 2

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Learnable Upsampling: “Deconvolution”

3 x 3 “deconvolution”, stride 2 pad 1

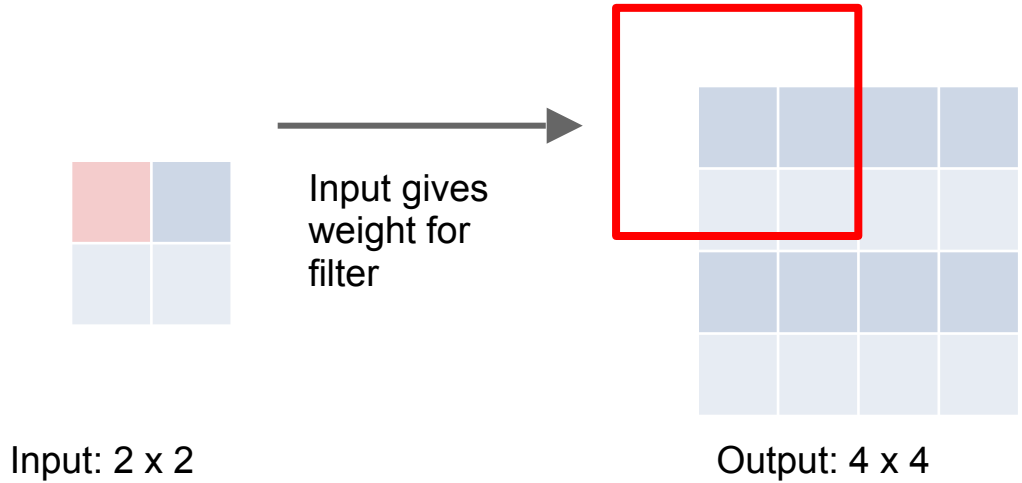
Input: 2 x 2

Output: 4 x 4

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

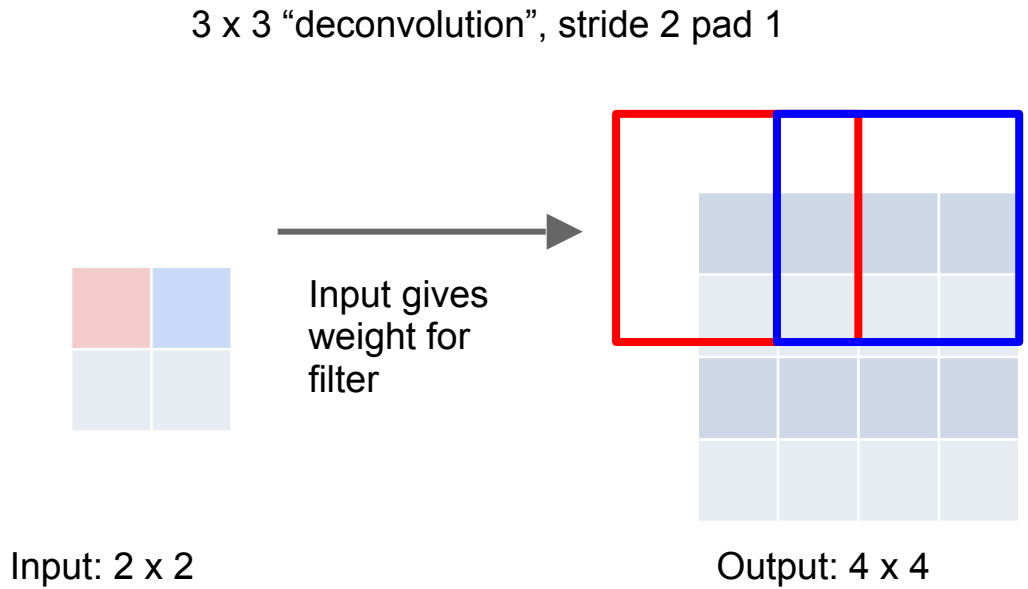
Learnable Upsampling: “Deconvolution”

3 x 3 “deconvolution”, stride 2 pad 1



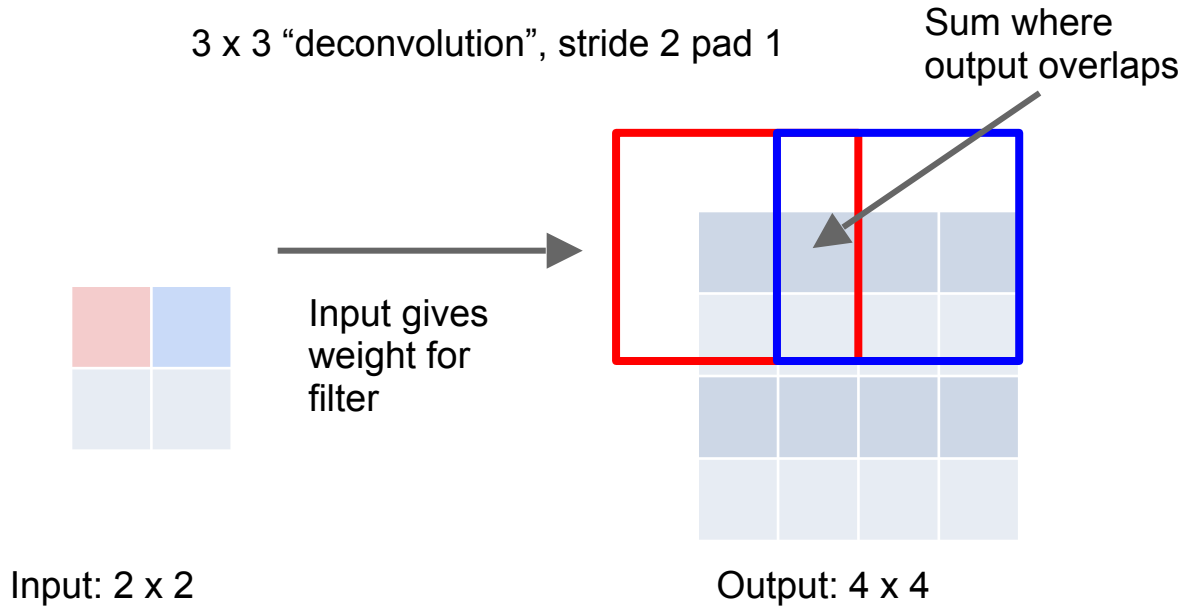
* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Learnable Upsampling: “Deconvolution”



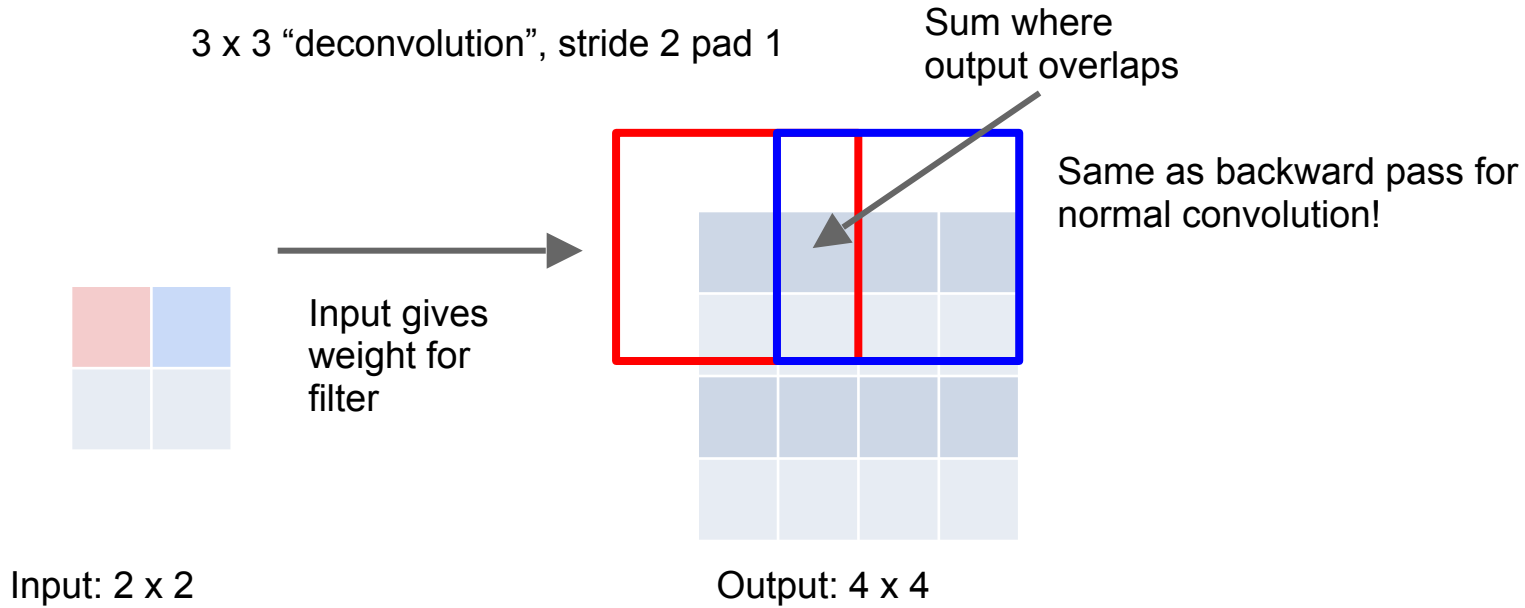
* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Learnable Upsampling: “Deconvolution”

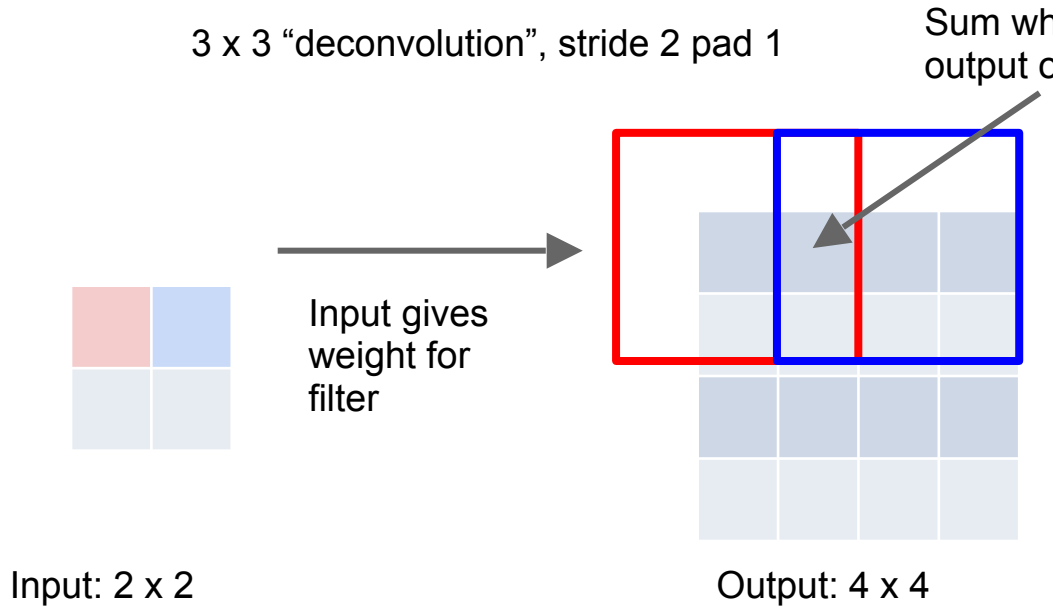


* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Learnable Upsampling: “Deconvolution”



Learnable Upsampling: “Deconvolution”



Same as backward pass for normal convolution!

“Deconvolution” is a bad name, already defined as “inverse of convolution”

Better names:
convolution transpose, backward strided convolution, 1/2 strided convolution, upconvolution

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Learnable Upsampling: “Deconvolution”

[1]It is more proper to say “convolutional transpose operation” rather than “deconvolutional” operation. Hence, we will be using the term “convolutional transpose” from now.

Im et al, “Generating images with recurrent adversarial networks”, arXiv 2016

A series of four fractionally-strided convolutions (in some recent papers, these are wrongly called deconvolutions)

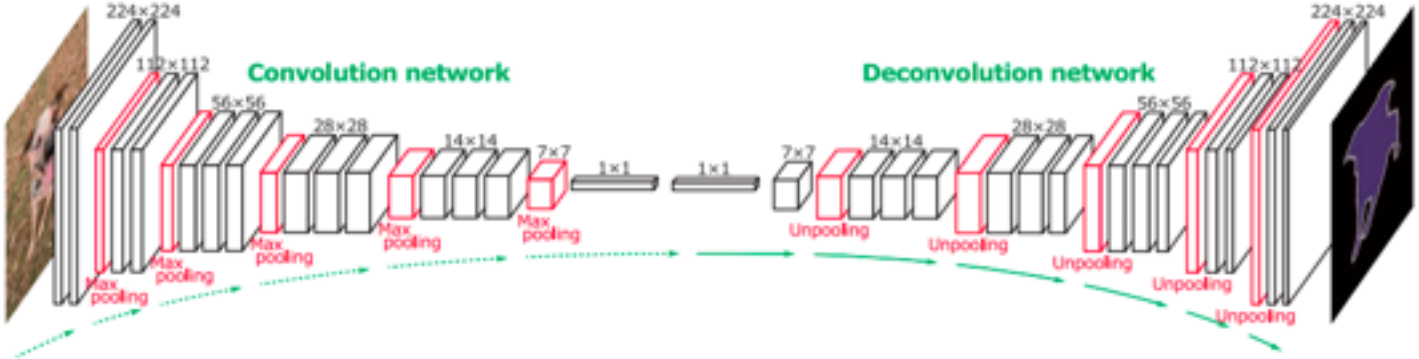
Radford et al, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”, ICLR 2016

“Deconvolution” is a bad name, already defined as “inverse of convolution”

Better names:
convolution transpose,
backward strided convolution,
1/2 strided convolution,
upconvolution

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

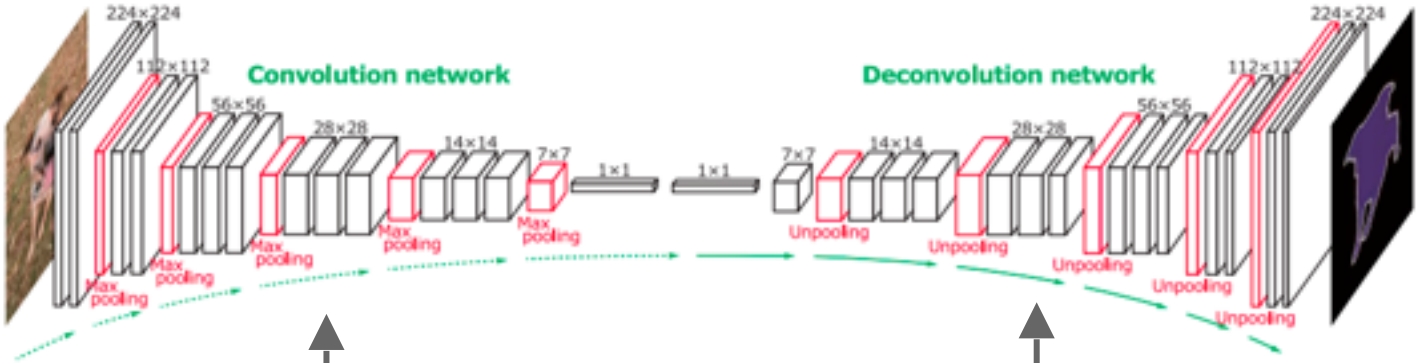
Semantic Segmentation: Upsampling



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

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Semantic Segmentation: Upsampling



Normal VGG

"Upside down" VGG

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

6 days of training on Titan X…

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Instance Segmentation

Instance Segmentation

Detect instances, give category, label pixels

“simultaneous detection and segmentation” (SDS)

Lots of recent work (MS-COCO)

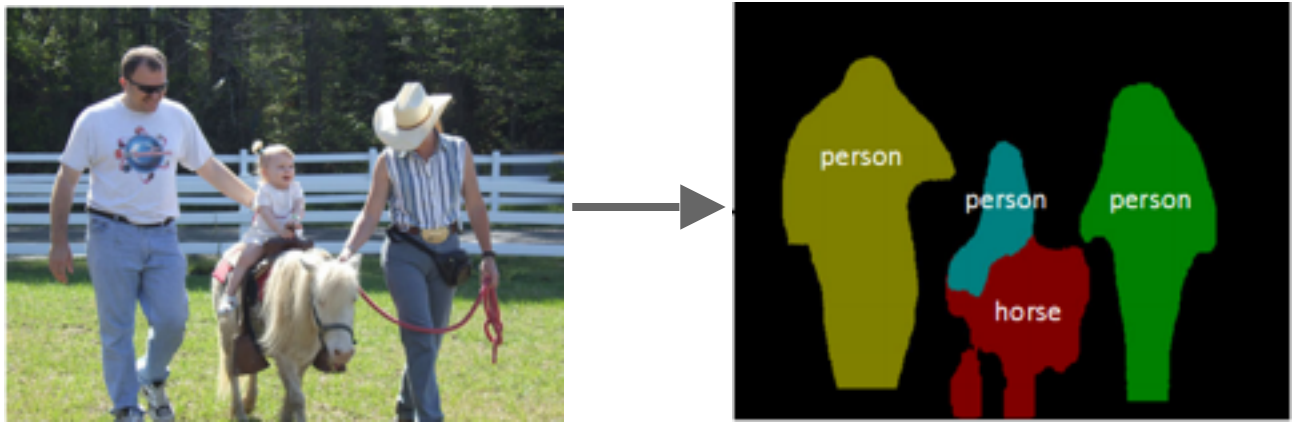


Figure credit: Dai et al, “Instance-aware Semantic Segmentation via Multi-task Network Cascades”, arXiv 2015

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

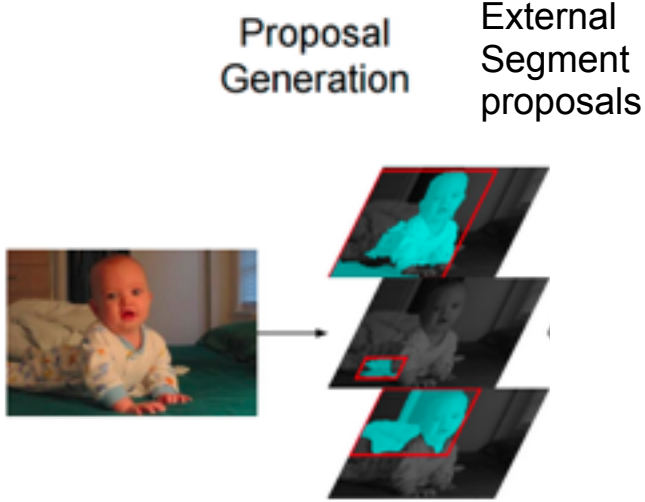
Instance Segmentation

Similar to R-CNN, but with segments

Hariharan et al, “Simultaneous Detection and Segmentation”, ECCV 2014

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Instance Segmentation

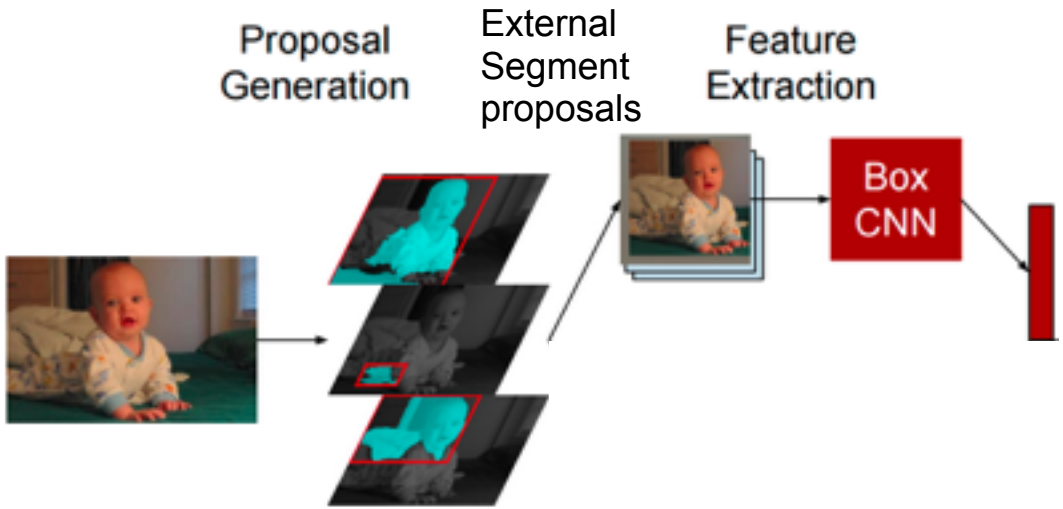


Similar to R-CNN, but with segments

Hariharan et al, “Simultaneous Detection and Segmentation”, ECCV 2014

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Instance Segmentation

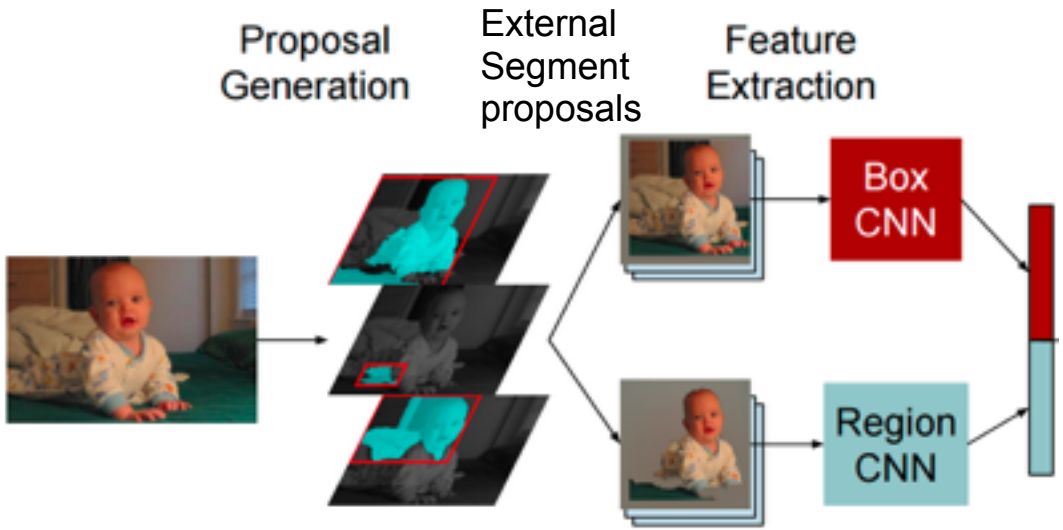


Similar to R-CNN, but with segments

Hariharan et al, “Simultaneous Detection and Segmentation”, ECCV 2014

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Instance Segmentation

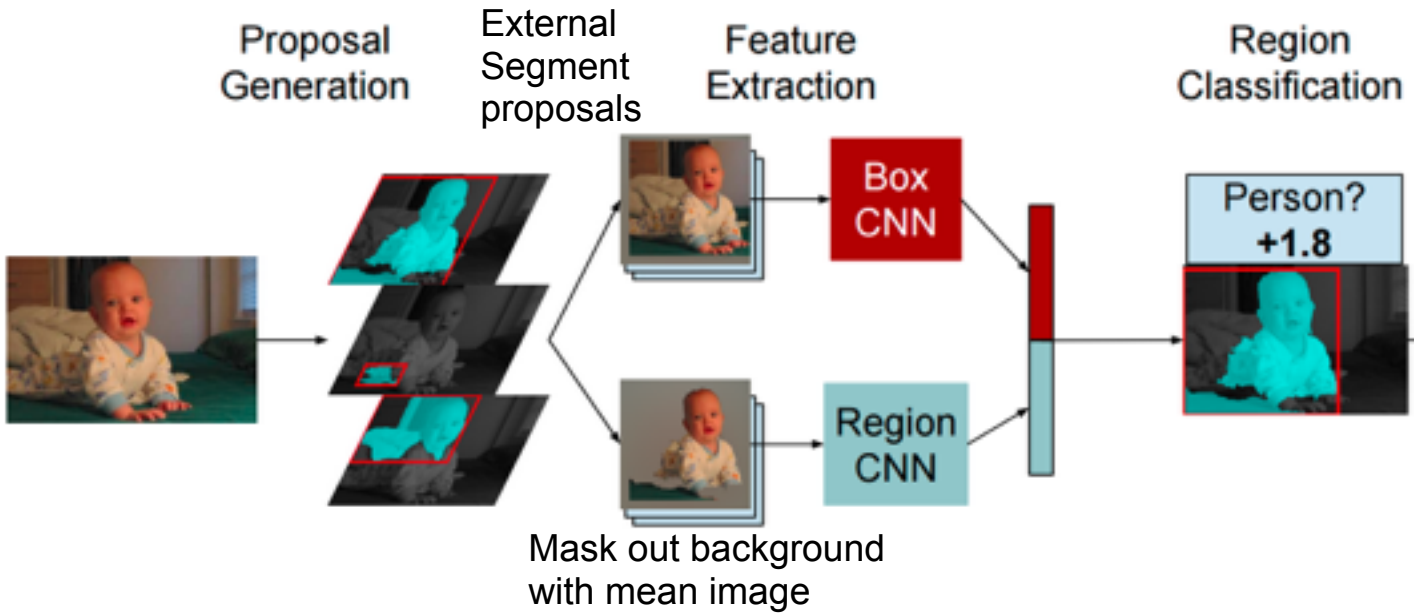


Similar to R-CNN, but with segments

Hariharan et al, “Simultaneous Detection and Segmentation”, ECCV 2014

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

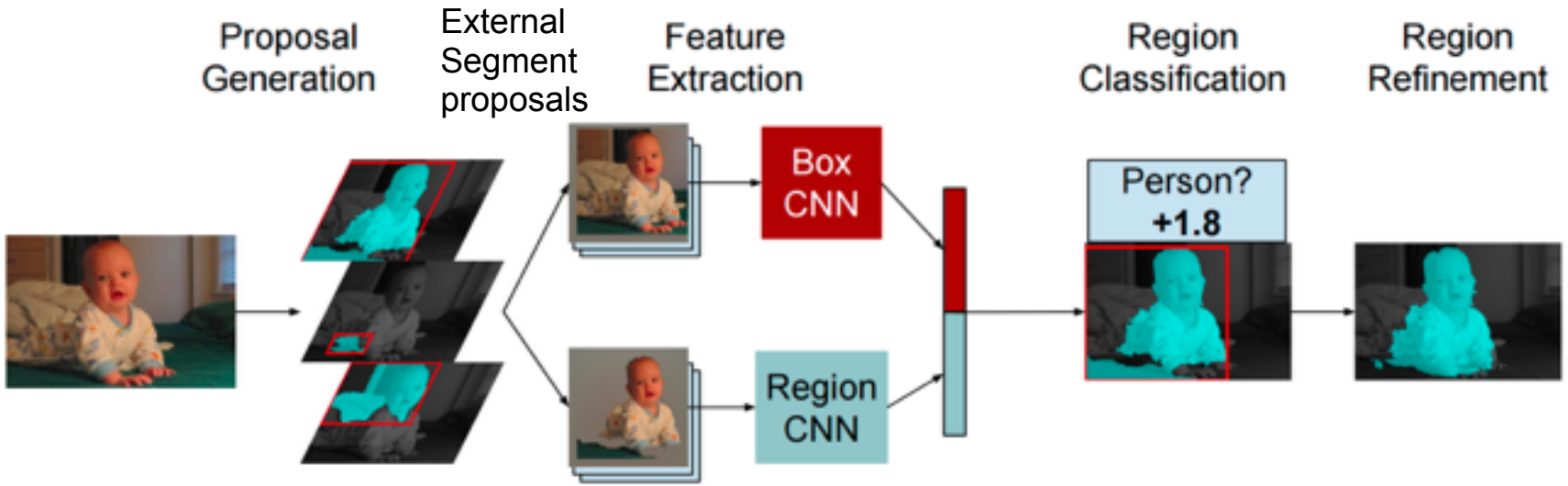
Instance Segmentation



Hariharan et al, “Simultaneous Detection and Segmentation”, ECCV 2014

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

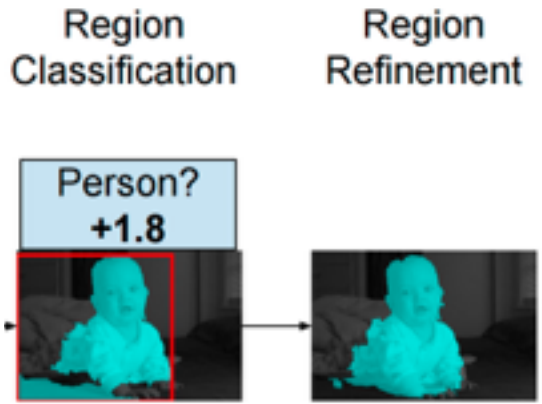
Instance Segmentation



Hariharan et al, “Simultaneous Detection and Segmentation”, ECCV 2014

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Instance Segmentation: Hypercolumns



Hariharan et al, “Hypercolumns for Object Segmentation and Fine-grained Localization”, CVPR 2015

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Instance Segmentation: Cascades

Similar to
Faster R-CNN

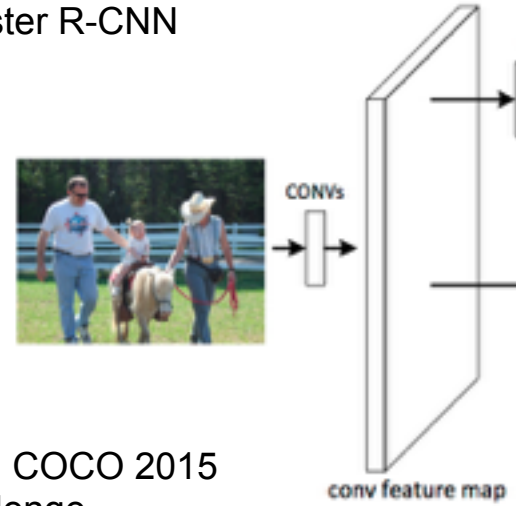
Won COCO 2015
challenge
(with ResNet)

Dai et al, “Instance-aware Semantic Segmentation via Multi-task Network Cascades”, arXiv 2015

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Instance Segmentation: Cascades

Similar to
Faster R-CNN



Won COCO 2015
challenge
(with ResNet)

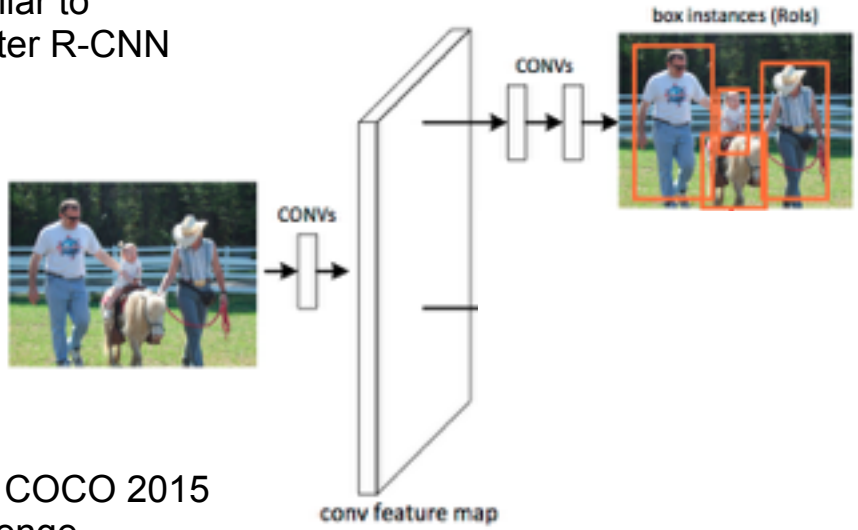
Dai et al, “Instance-aware Semantic Segmentation via Multi-task Network Cascades”, arXiv 2015

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Instance Segmentation: Cascades

Similar to
Faster R-CNN

Region proposal network (RPN)



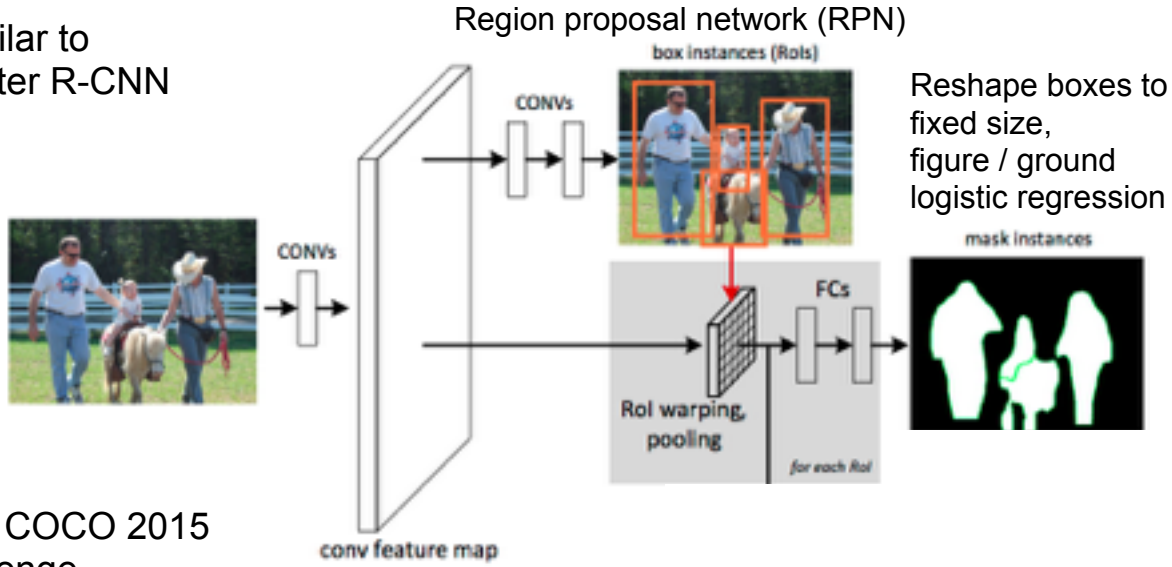
Won COCO 2015
challenge
(with ResNet)

Dai et al, “Instance-aware Semantic Segmentation via Multi-task Network Cascades”, arXiv 2015

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Instance Segmentation: Cascades

Similar to
Faster R-CNN



Won COCO 2015
challenge
(with ResNet)

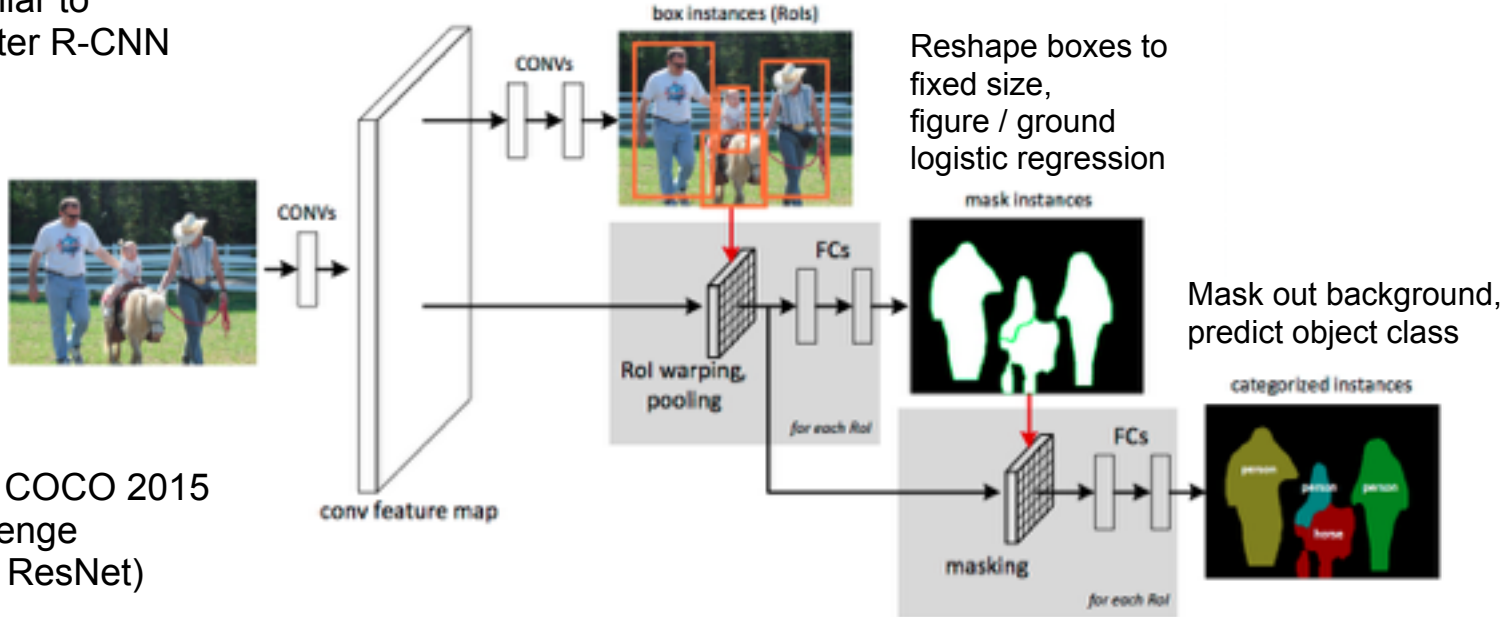
Dai et al, "Instance-aware Semantic Segmentation via Multi-task Network Cascades", arXiv 2015

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Instance Segmentation: Cascades

Similar to
Faster R-CNN

Region proposal network (RPN)



Won COCO 2015
challenge
(with ResNet)

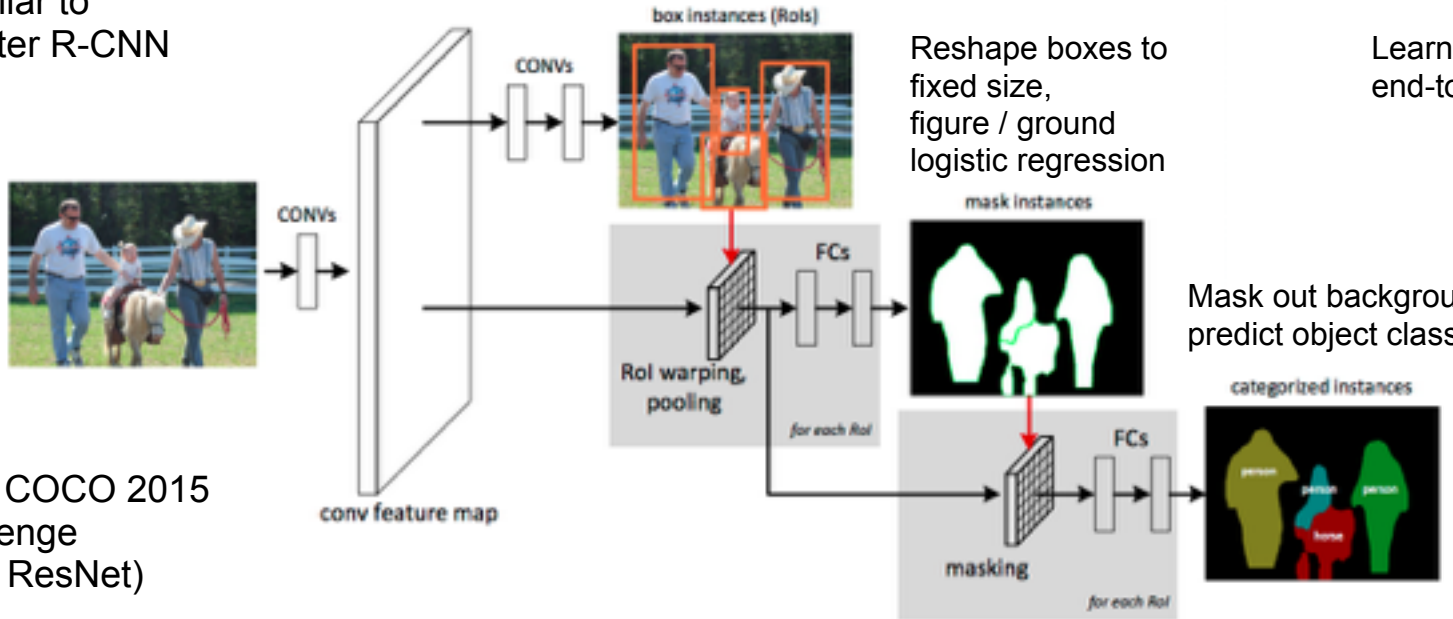
Dai et al, "Instance-aware Semantic Segmentation via Multi-task Network Cascades", arXiv 2015

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Instance Segmentation: Cascades

Similar to
Faster R-CNN

Region proposal network (RPN)



Reshape boxes to
fixed size,
figure / ground
logistic regression

Learn entire model
end-to-end!

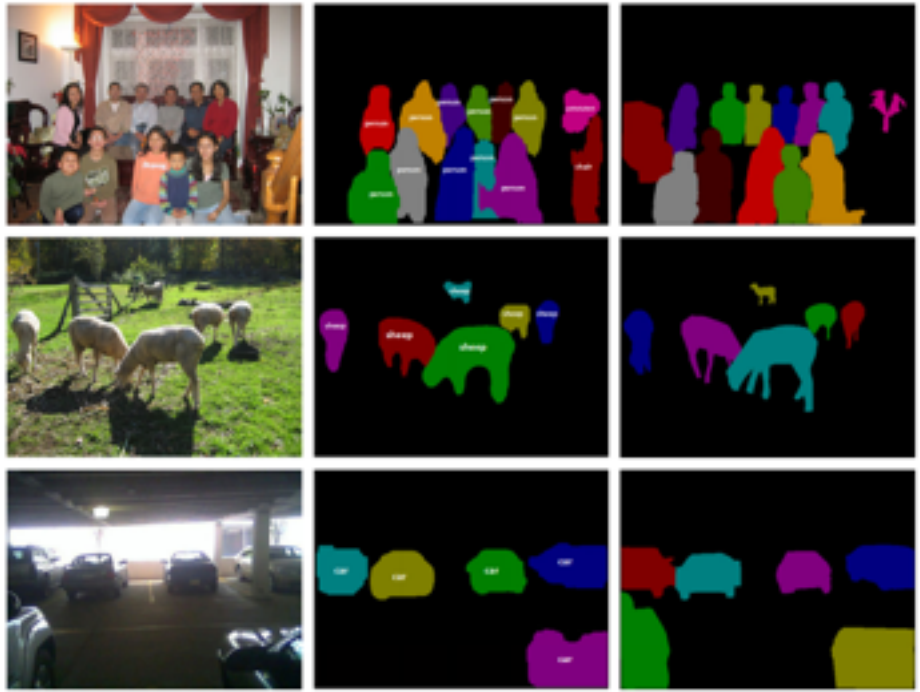
Mask out background,
predict object class

Won COCO 2015
challenge
(with ResNet)

Dai et al, "Instance-aware Semantic Segmentation via Multi-task Network Cascades", arXiv 2015

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Instance Segmentation: Cascades



Predictions **Ground truth**

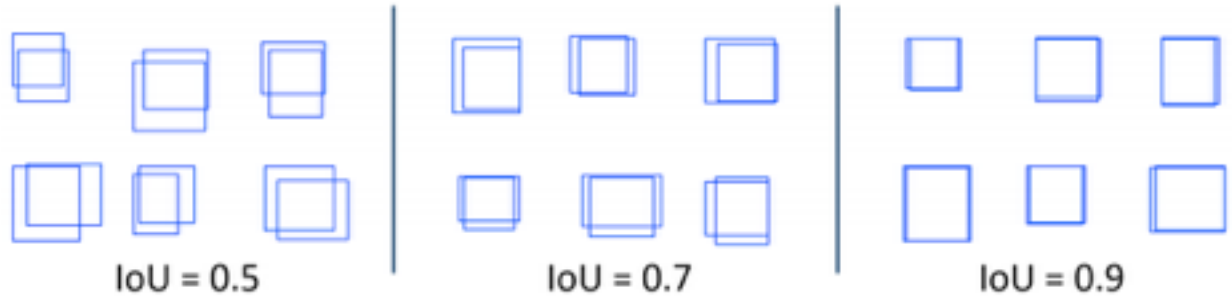
Dai et al, "Instance-aware Semantic Segmentation via Multi-task Network Cascades", arXiv 2015

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

Evaluation metrics

Detection Score: Average Precision (mAP is mean AP over all object categories)

- AP is averaged over multiple IoU values between 0.5 and 0.95 (and categories, size)
- More comprehensive comparison metric than the traditional AP at Intersection over Union (IoU) threshold of 0.5.

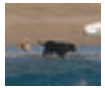


Evaluation Metrics

Detection Score: AP

- AP is averaged over multiple IoU values between 0.5 and 0.95
- AP is averaged over groups of objects or over object size in an image
- More comprehensive metric than the traditional AP at Intersection over Union (IoU) threshold of 0.5.

A < 32x32

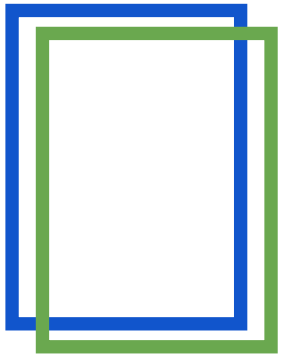


A > 96x96

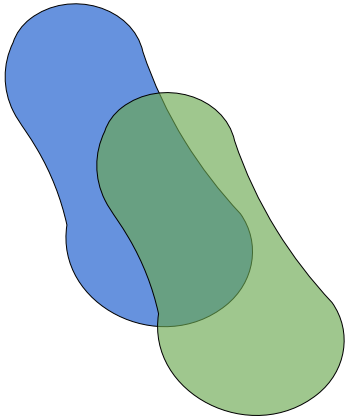
32x32 < A < 96x96

Evaluating Detection or Segmentation

To calculate AP we need:



Bounding Box IoU



Mask IoU

Segmentation Overview

- **Graph Cut**
  - Classic algorithm to quickly get foreground from background segmentation
  - Not trained to consider object shape prior
  - Can fail with complicated backgrounds
- **Unsupervised Segmentation**
  - Superpixels
  - Open area of research in ML
- **Semantic segmentation**
  - Classify all pixels
  - Fully convolutional models, downsample then upsample
  - Learnable upsampling: fractionally strided convolution
  - Skip connections can help
- **Instance Segmentation**
  - Detect instance, generate mask
  - Similar pipelines to object detection

* Original slides borrowed from Andrej Karpathy and Li Fei-Fei, Stanford cs231n

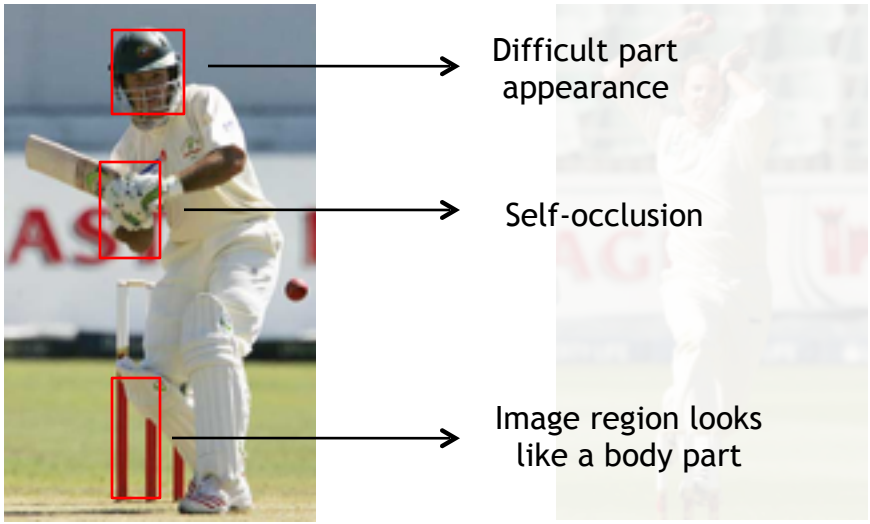
Pose Estimation
(aka Keypoints)

Classic Problem: Activity Recognition

What is this person doing?

Human pose estimation & Object detection

Human pose estimation is challenging.

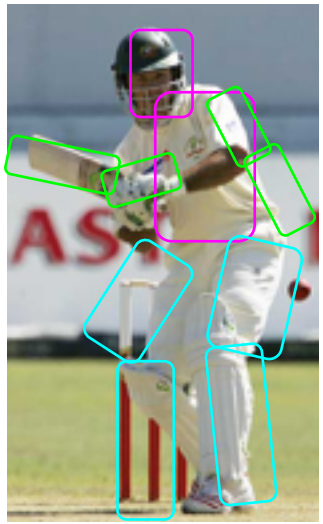


- Felzenszwalb & Huttenlocher, 2005
- Ren et al, 2005
- Ramanan, 2006
- Ferrari et al, 2008
- Yang & Mori, 2008
- Andriluka et al, 2009
- Eichner & Ferrari, 2009

Slide Credit: Yao/Fei-Fei

Human pose estimation & Object detection

Human pose estimation is challenging.



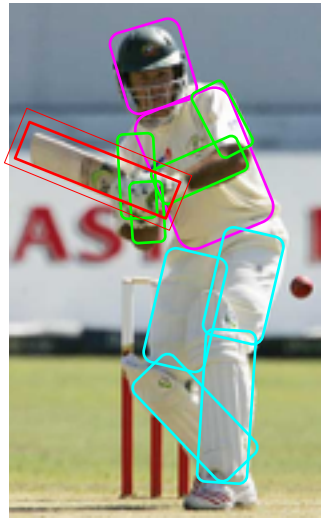
- Felzenszwalb & Huttenlocher, 2005
- Ren et al, 2005
- Ramanan, 2006
- Ferrari et al, 2008
- Yang & Mori, 2008
- Andriluka et al, 2009
- Eichner & Ferrari, 2009

Slide Credit: Yao/Fei-Fei

Human pose estimation & Object detection

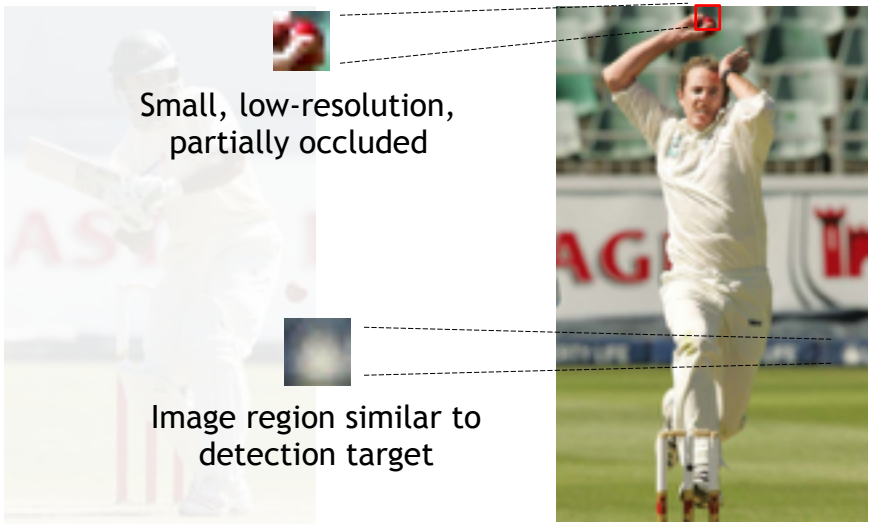
Facilitate

Given the object is detected.



Slide Credit: Yao/Fei-Fei

Human pose estimation & Object detection



Object detection is challenging

- Viola & Jones, 2001
- Lampert et al, 2008
- Divvala et al, 2009
- Vedaldi et al, 2009

Slide Credit: Yao/Fei-Fei

Human pose estimation & Object detection

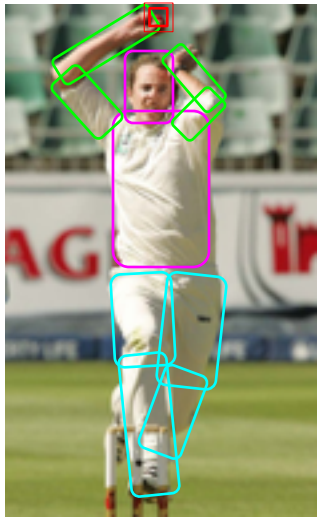
Object detection is challenging

- Viola & Jones, 2001
- Lampert et al, 2008
- Divvala et al, 2009
- Vedaldi et al, 2009

Slide Credit: Yao/Fei-Fei

Human pose estimation & Object detection

Facilitate

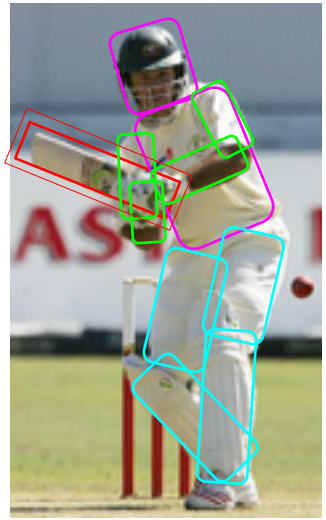


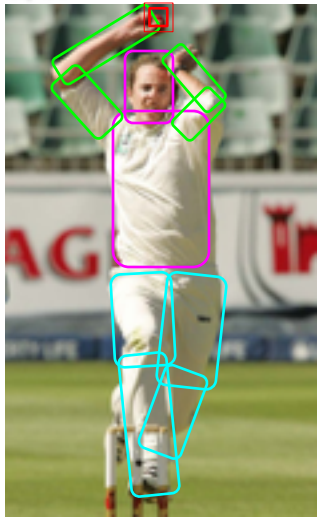
Given the pose is estimated.

Slide Credit: Yao/Fei-Fei

Human pose estimation & Object detection

Mutual Context





Slide Credit: Yao/Fei-Fei

Multi-Person Pose Estimation using Part Affinity Fields

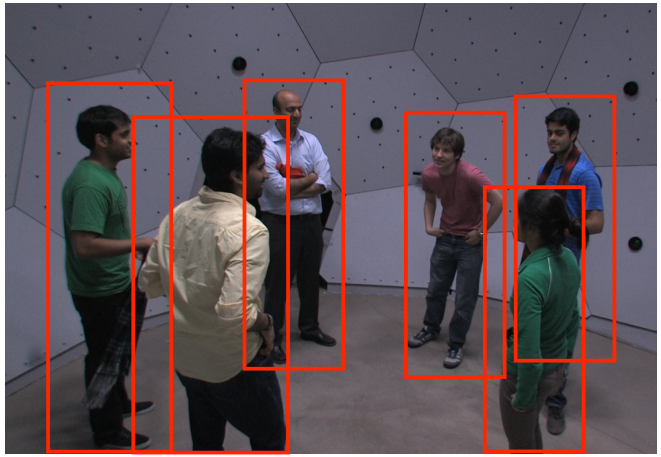
Zhe Cao, Shih-En Wei, Tomas Simon, Yaser Sheikh
Carnegie Mellon University

Top-down Approach: Person Detection + Pose Estimation

Top-down

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Top-down Approach: Person Detection + Pose Estimation



Top-down

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Top-down Approach: Person Detection + Pose Estimation

Top-down

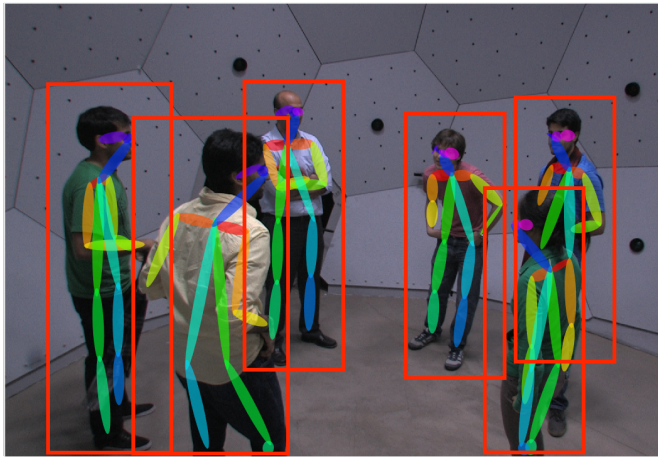
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Top-down Approach: Person Detection + Pose Estimation

Top-down

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

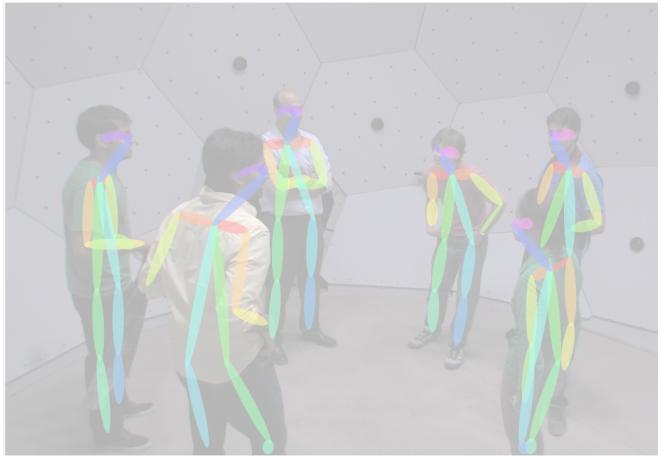
Top-down Approach: Person Detection + Pose Estimation



Top-down

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Our Method: Parts Detection + Parts Association

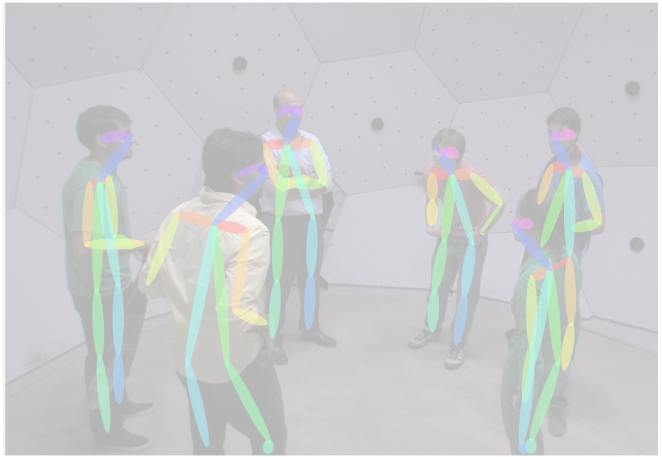


Top-down

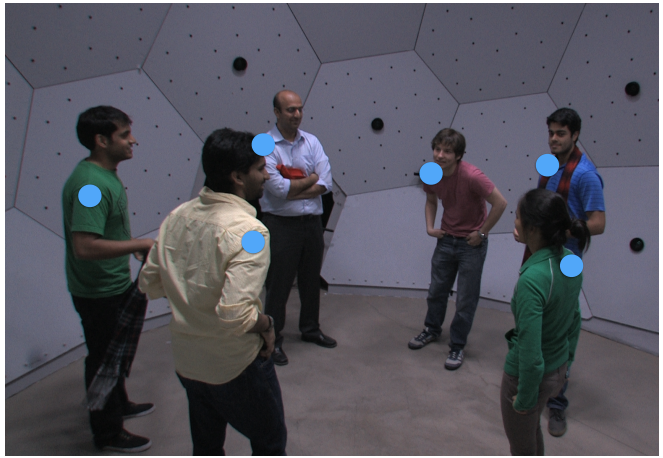
Ours

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Our Method: Parts Detection + Parts Association



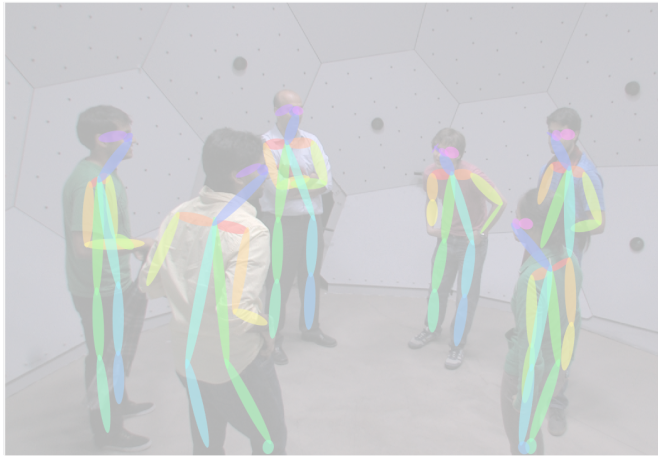
Top-down



Ours

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Our Method: Parts Detection + Parts Association

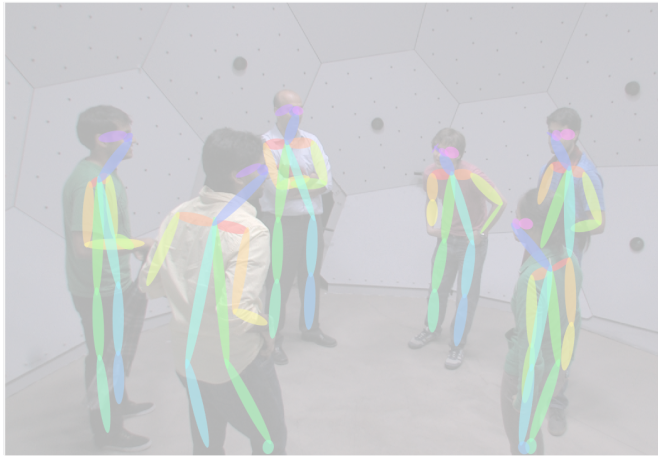


Top-down

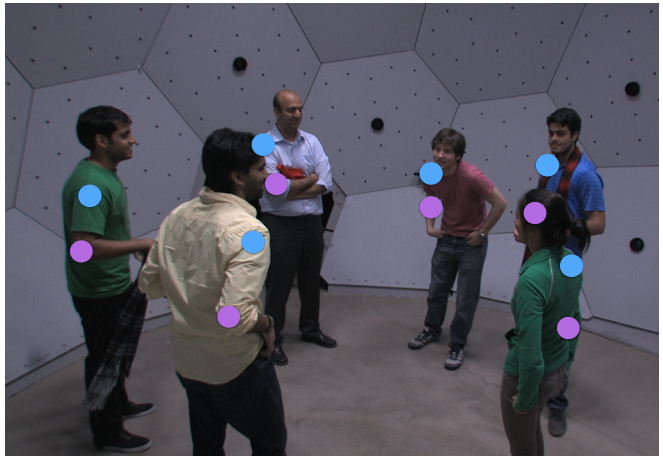
Ours

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Our Method: Parts Detection + Parts Association



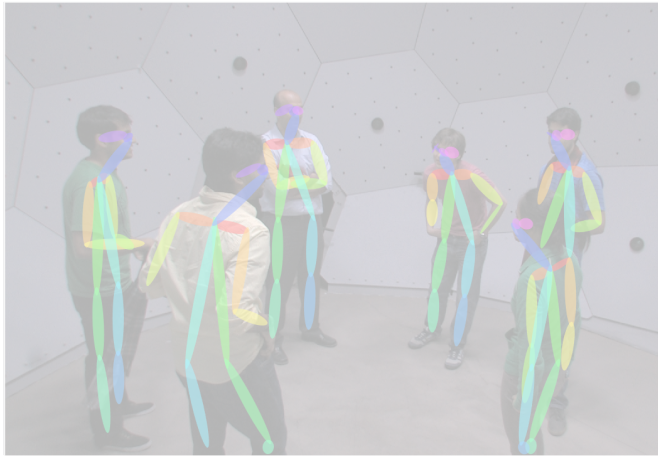
Top-down



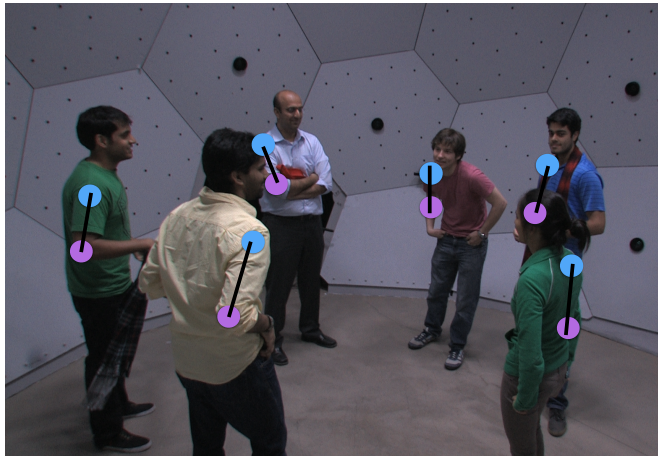
Ours

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Our Method: Parts Detection + Parts Association

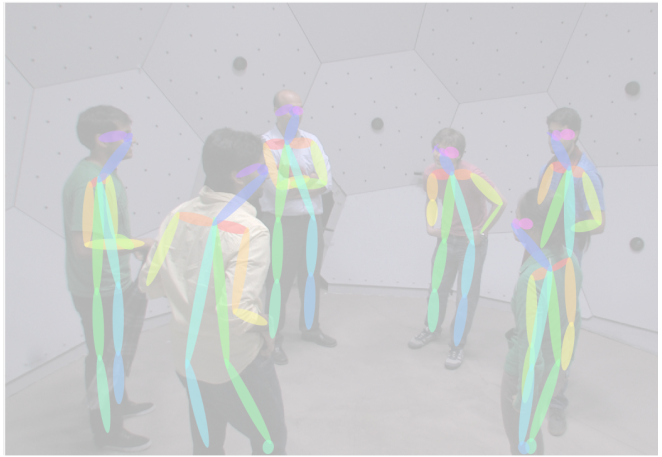


Top-down



* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Our Method: Parts Detection + Parts Association

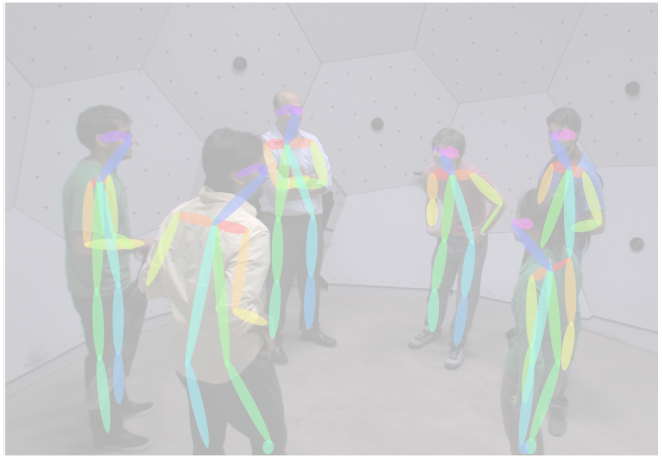


Top-down

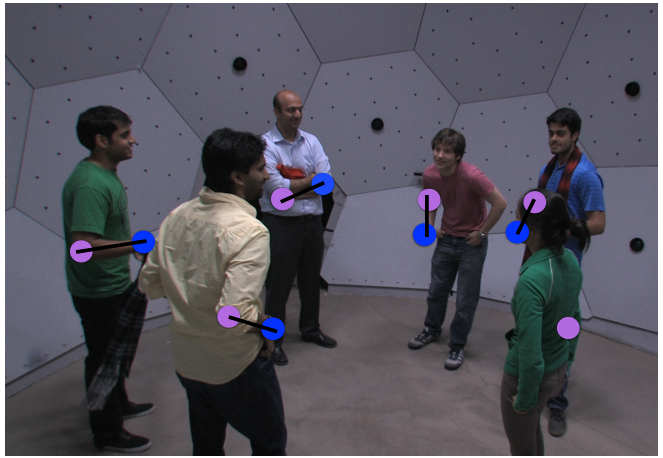
Part Affinity Fields

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Our Method: Parts Detection + Parts Association

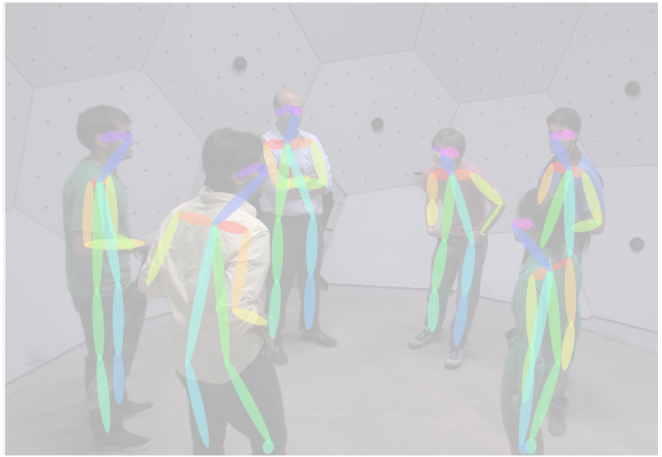


Top-down



* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Our Method: Parts Detection + Parts Association

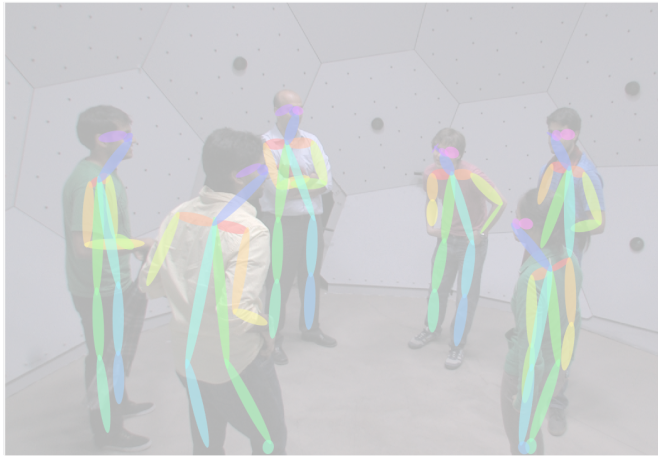


Top-down

Part Affinity Fields

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Our Method: Parts Detection + Parts Association

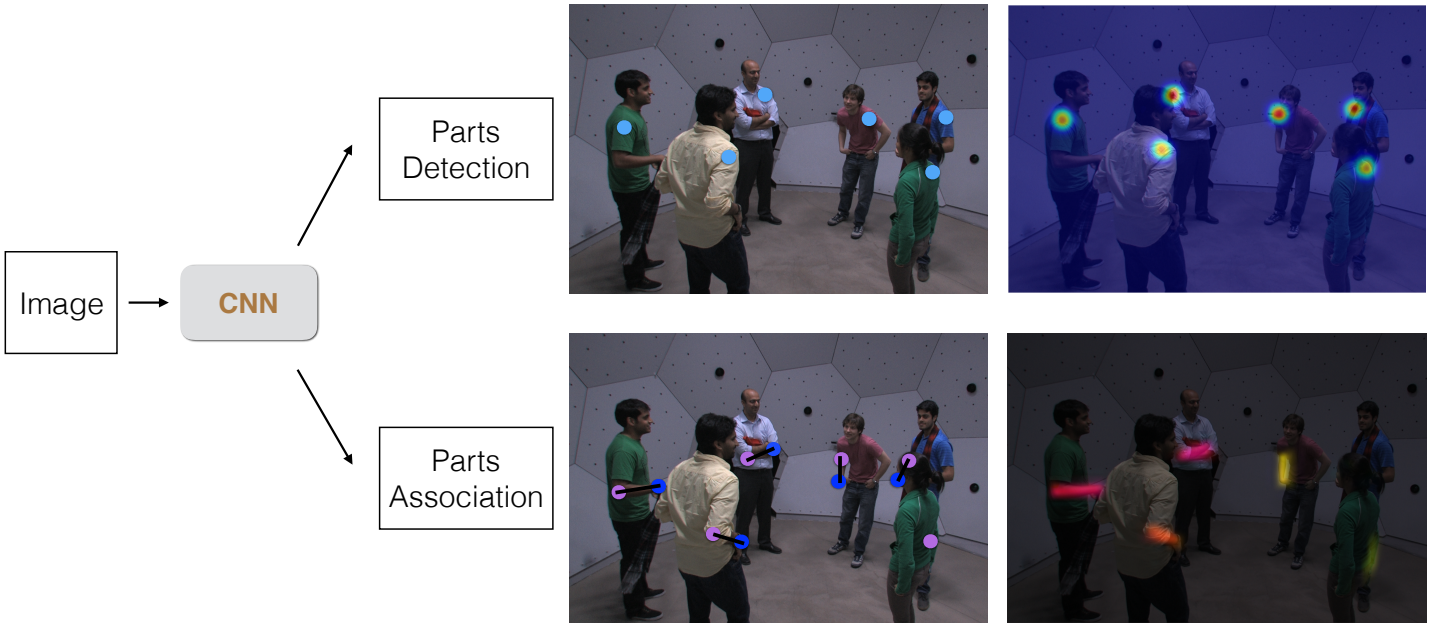


Top-down

Ours

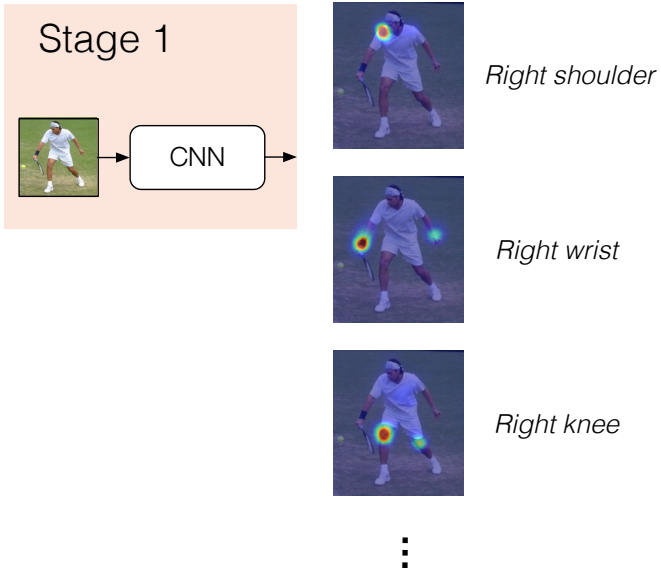
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Novelty: Jointly Learning Parts Detection and Parts Association



* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

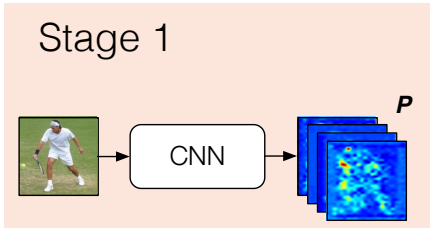
Sequential Prediction with Learned Spatial Context



Convolutional Pose Machines, Wei, Ramakrishna, Kanade, Sheikh, CVPR 2016

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

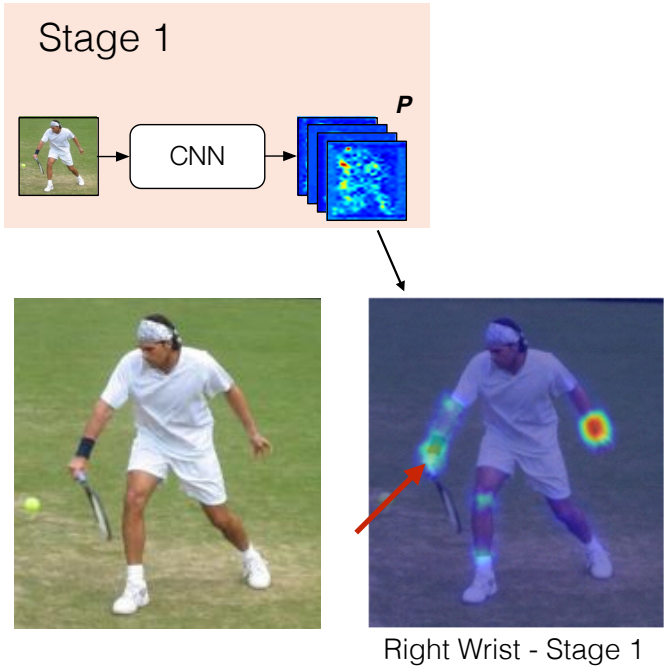
Sequential Prediction with Learned Spatial Context



Convolutional Pose Machines, Wei, Ramakrishna, Kanade, Sheikh, CVPR 2016

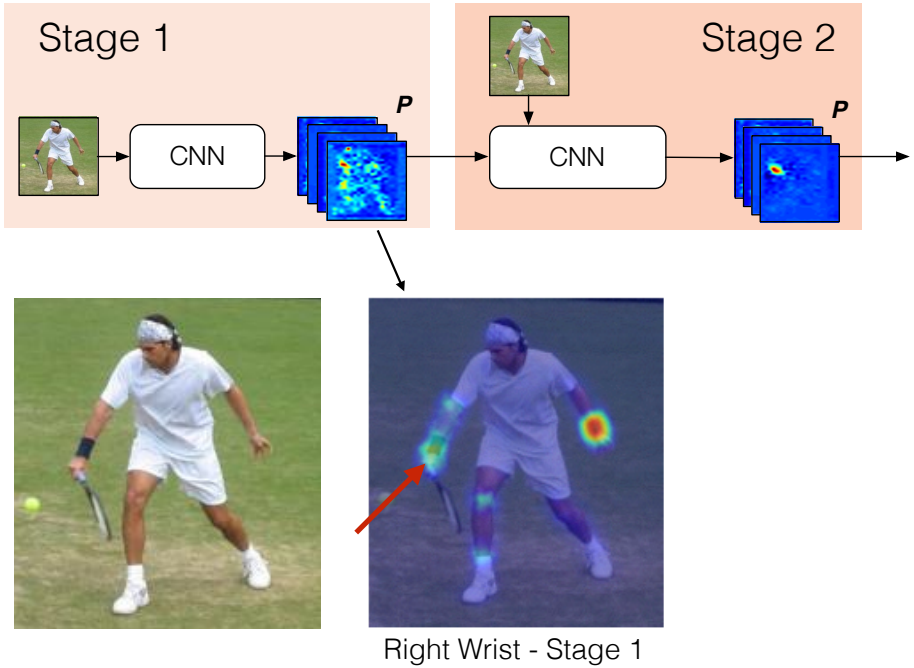
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Sequential Prediction with Learned Spatial Context



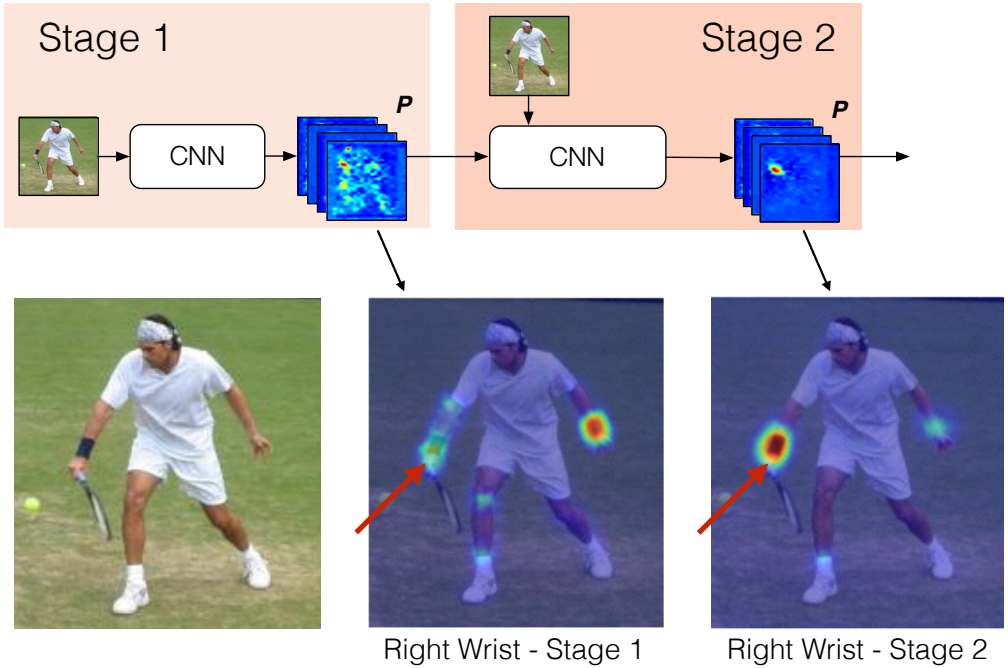
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Sequential Prediction with Learned Spatial Context



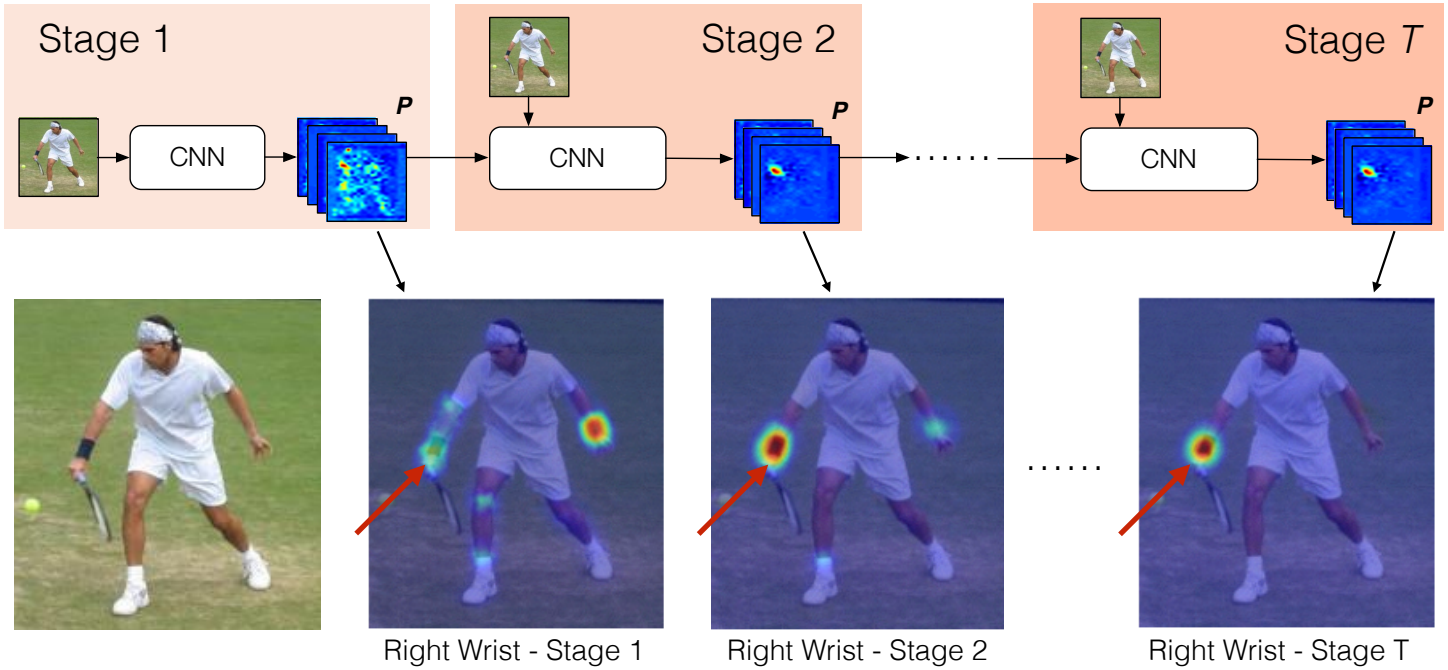
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Sequential Prediction with Learned Spatial Context



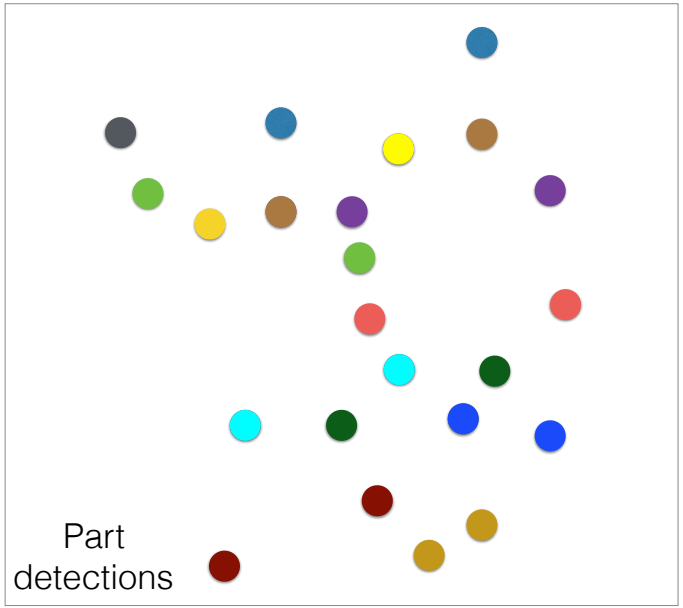
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Sequential Prediction with Learned Spatial Context



* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

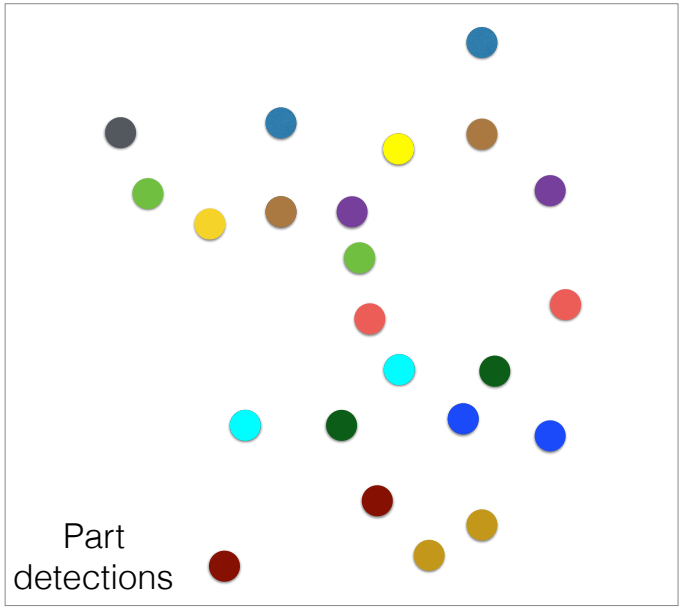
Part-Person Association for Multi-Person Pose Estimation

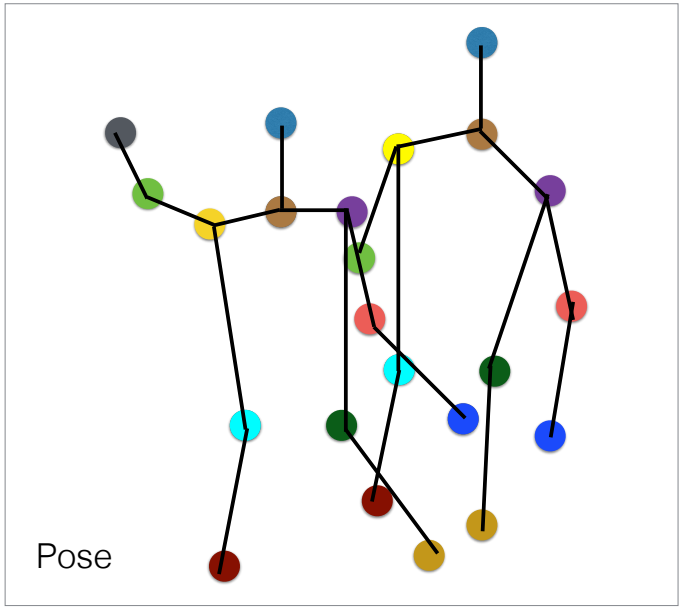


Elbow
Wrist
⋮

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

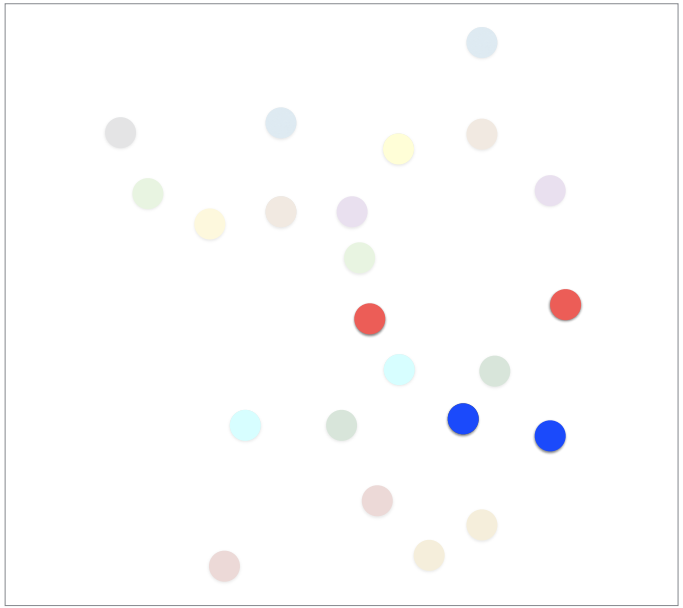
Part-Person Association for Multi-Person Pose Estimation





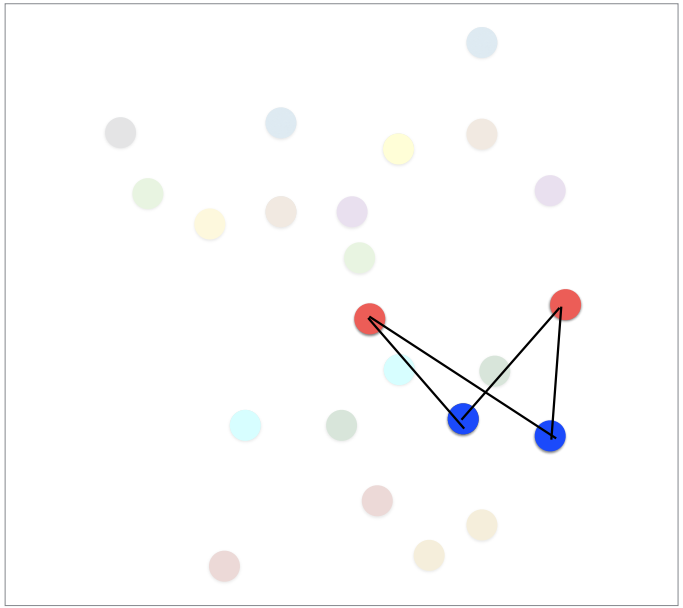
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Part-to-Part Association for Multi-Person Pose Estimation



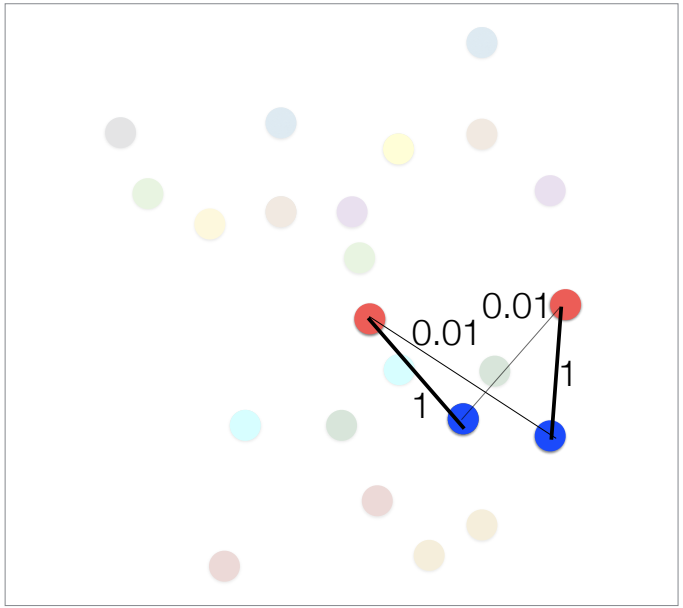
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Part Affinity Score Guides the Connection



* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Part Affinity Score Guides the Connection

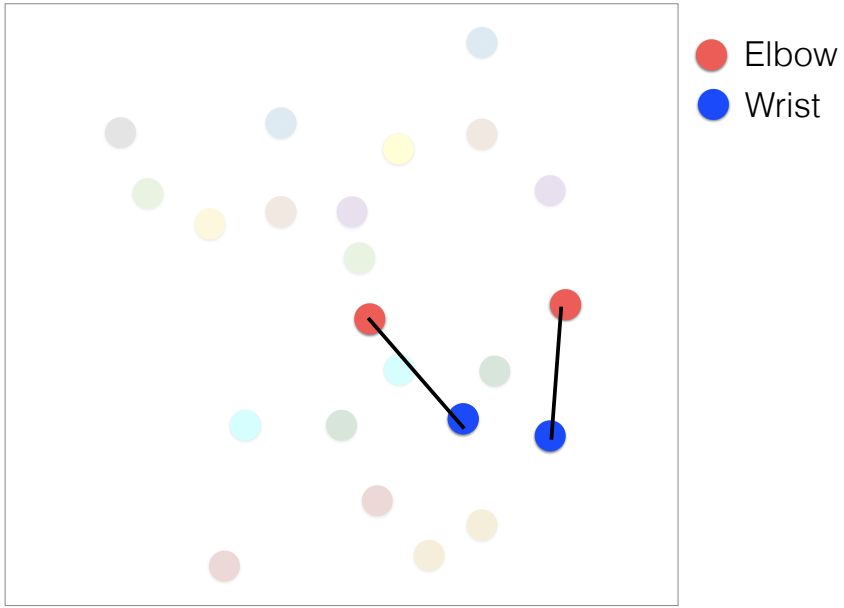


Elbow

Wrist

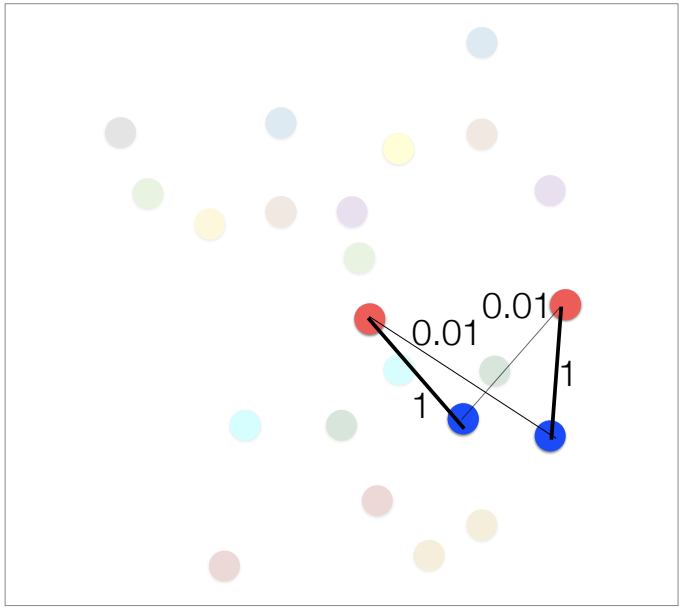
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Part Affinity Score Guides the Connection



* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

How to Obtain the Part Affinity Score

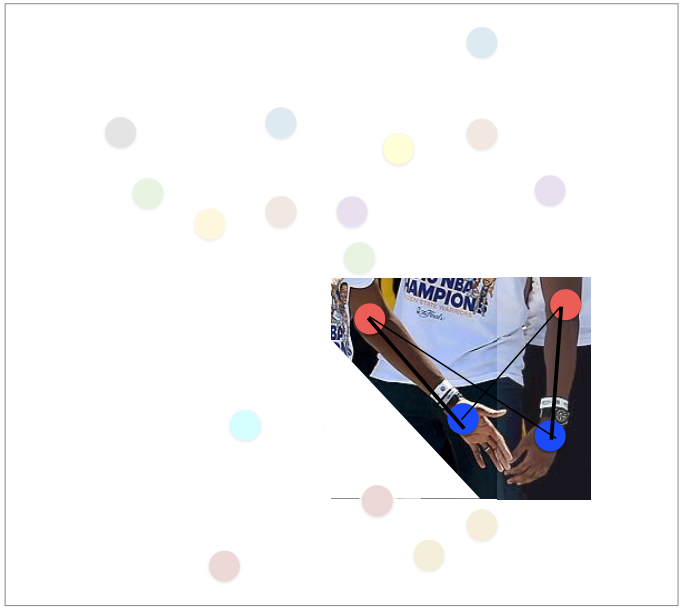


Elbow

Wrist

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Part Affinity Score is Dependent on Visual Appearance

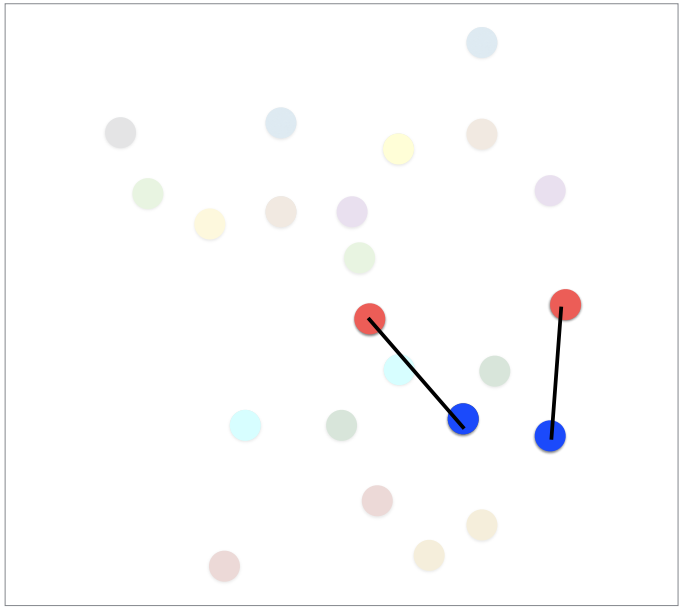


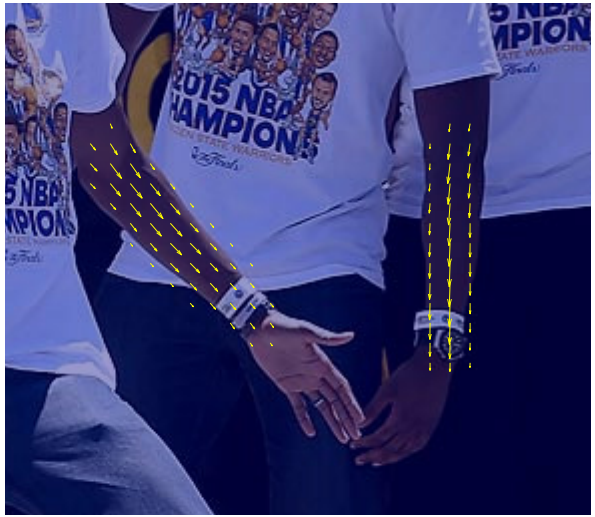
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Part Affinity Score is Dependent on Visual Appearance

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Key Idea: Encode the Part Affinity Score on the Image Plane

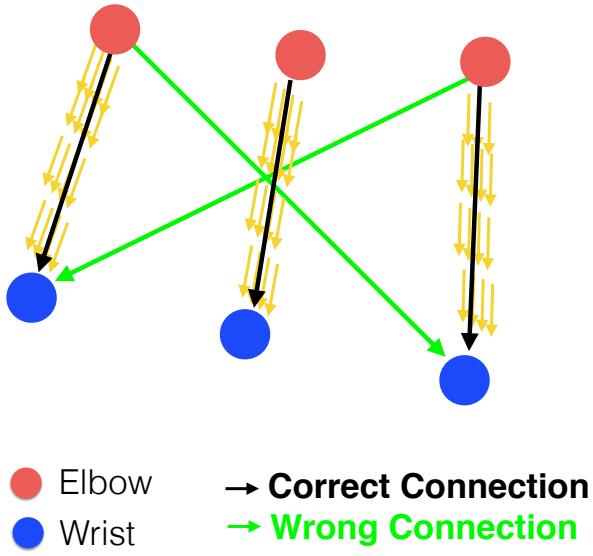




Part Affinity Fields
encode **direction** and **position**

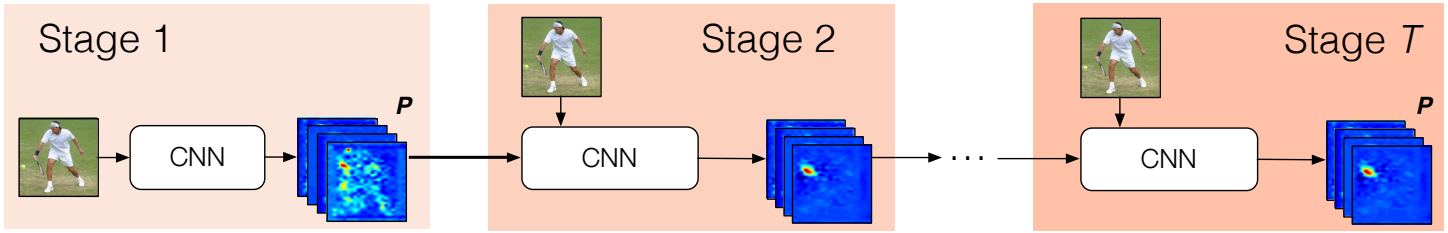
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Part Affinity Fields Avoid Spatial Ambiguity



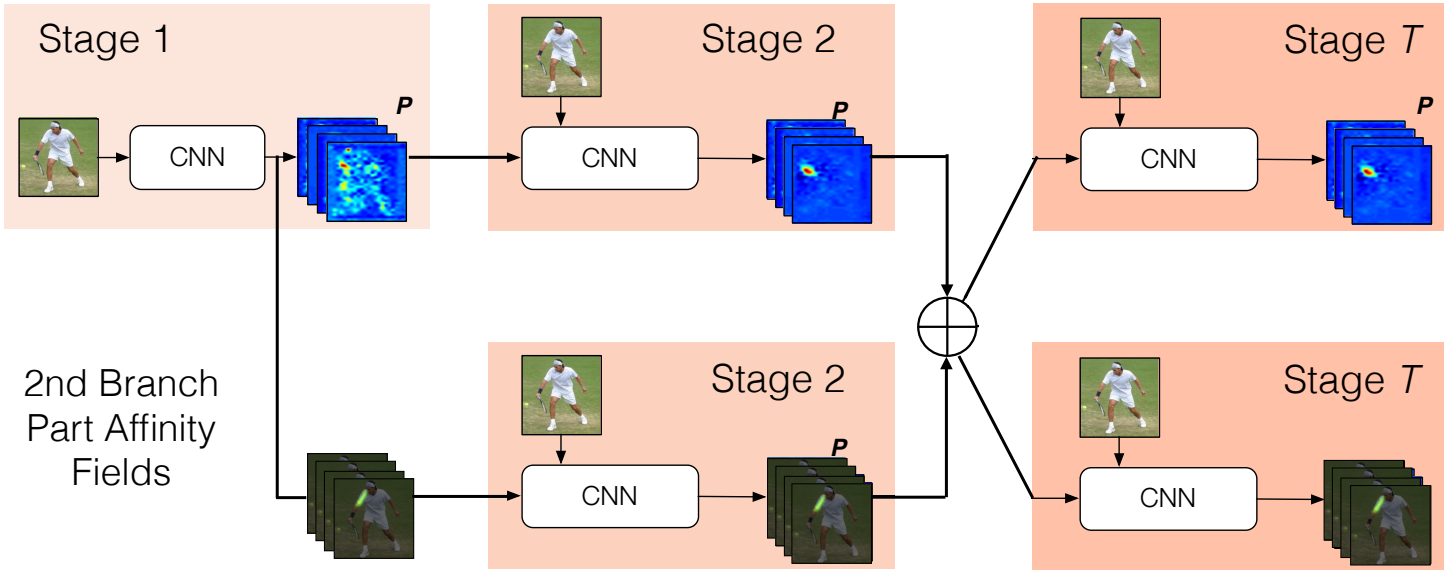
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Jointly Learning Parts Detection and Parts Association



* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Jointly Learning Parts Detection and Parts Association

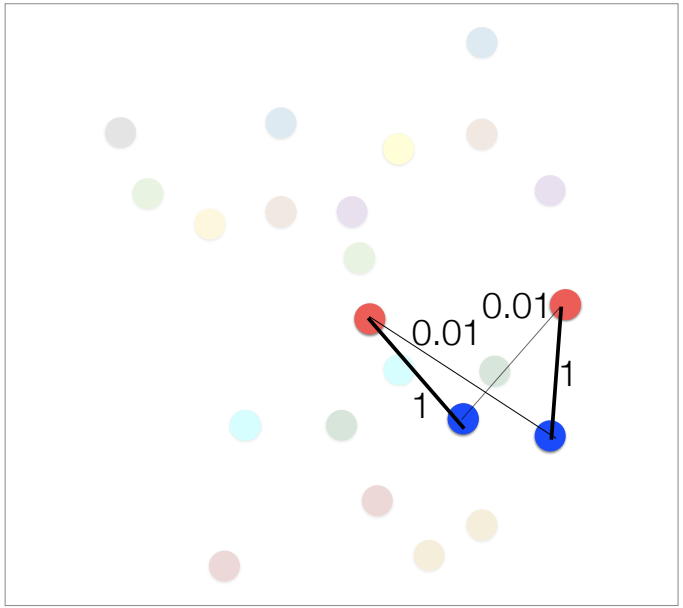


* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Greedy Algorithm for Body Parts Association

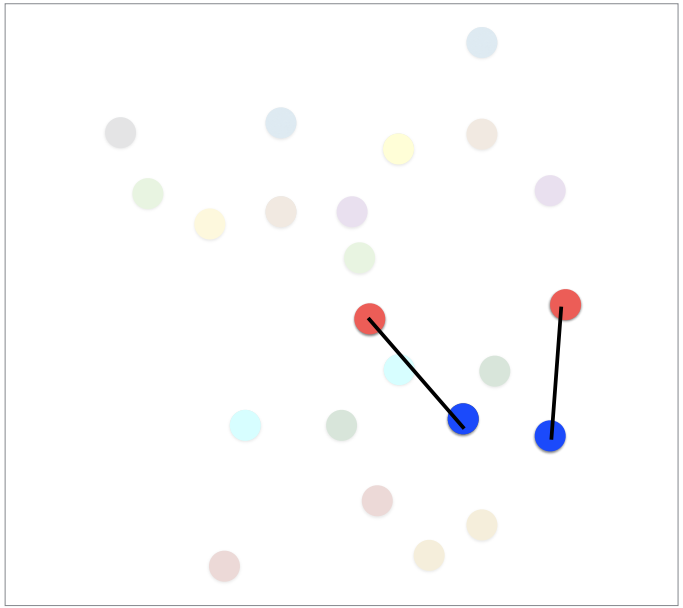


Elbow

Wrist

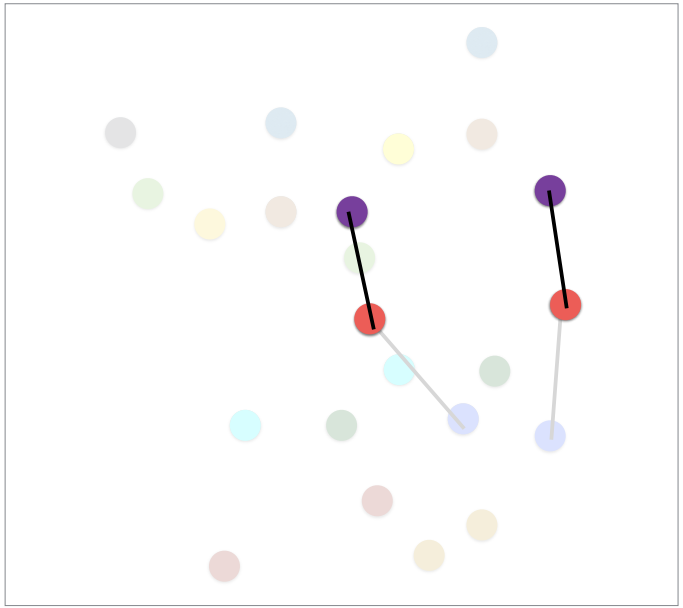
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Greedy Algorithm for Body Parts Association



* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Greedy Algorithm for Body Parts Association

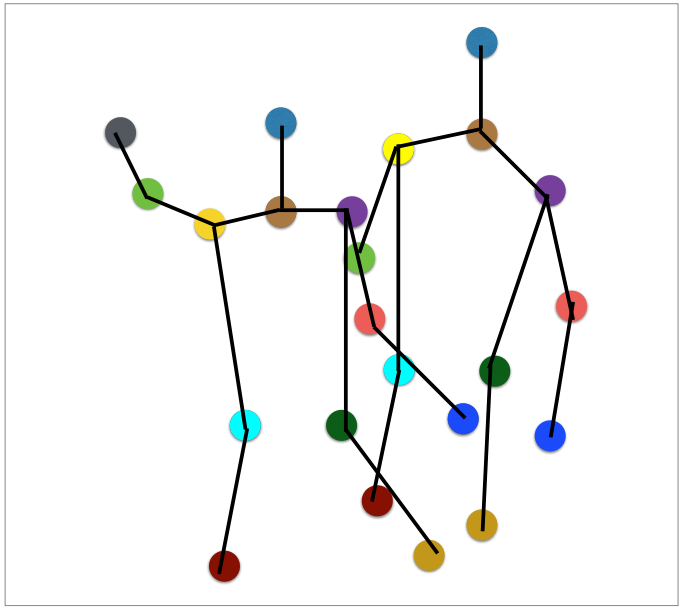


Elbow

Shoulder

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Greedy Algorithm for Body Parts Association



* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Results on COCO Challenge Validation Set

Top-down

Method	AP on val
GT bbox + CPM [1]	63
SSD [2] **+ CPM** [1]	53
Our Method	58.5
Ours + Refinement	61

[1] Convolutional Pose Machines [Wei et al. 2016]
[2] SSD: Single Shot MultiBox Detector [Liu et al. 2015]

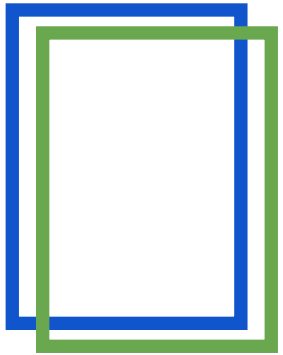
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

https://youtu.be/pW6nZXeWlGM

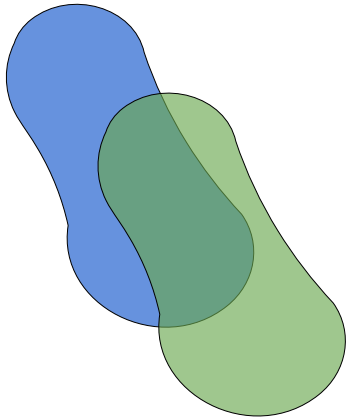
* Slide: Zhe Cao: “Multi-Person Pose Estimation Using Part Affinity Models

Evalutation: Keypoints

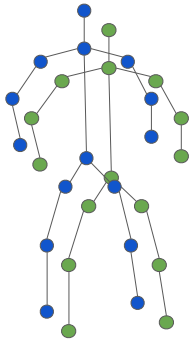
To calculate AP we need:



Bounding Box IoU



Mask IoU



Object Keypoint Similarity

Object Keypoint Similarity -OKS