

# Lecture 23

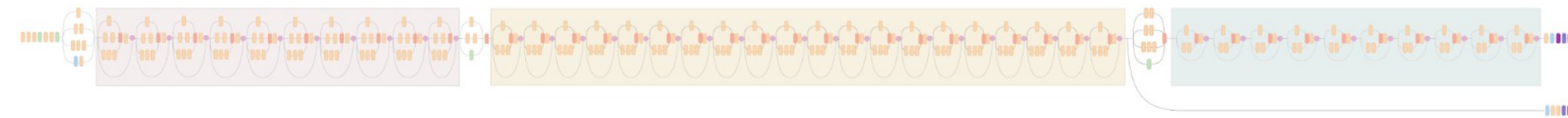
## Deep Learning: Segmentation

COS 429: Computer Vision

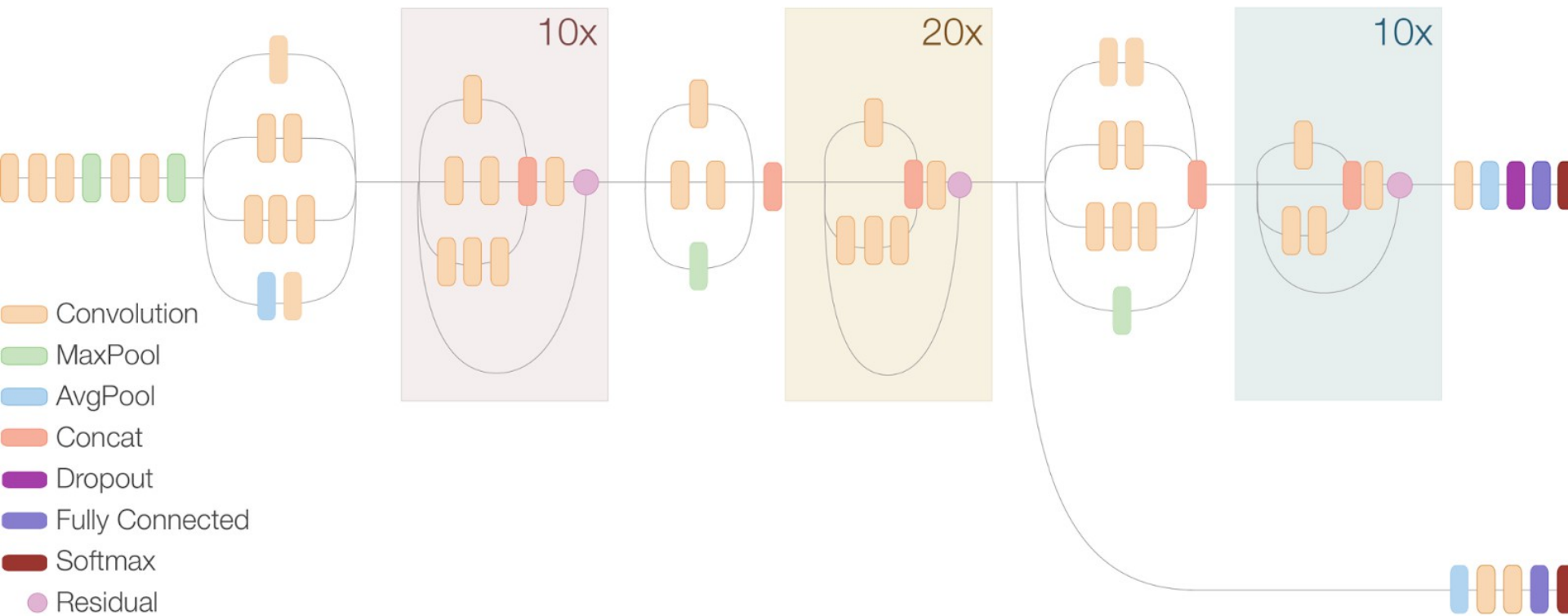


Thanks: most of these slides shamelessly adapted from  
Stanford CS231n: Convolutional Neural Networks for Visual Recognition  
Fei-Fei Li, Andrej Karpathy, Justin Johnson  
<http://cs231n.stanford.edu/>

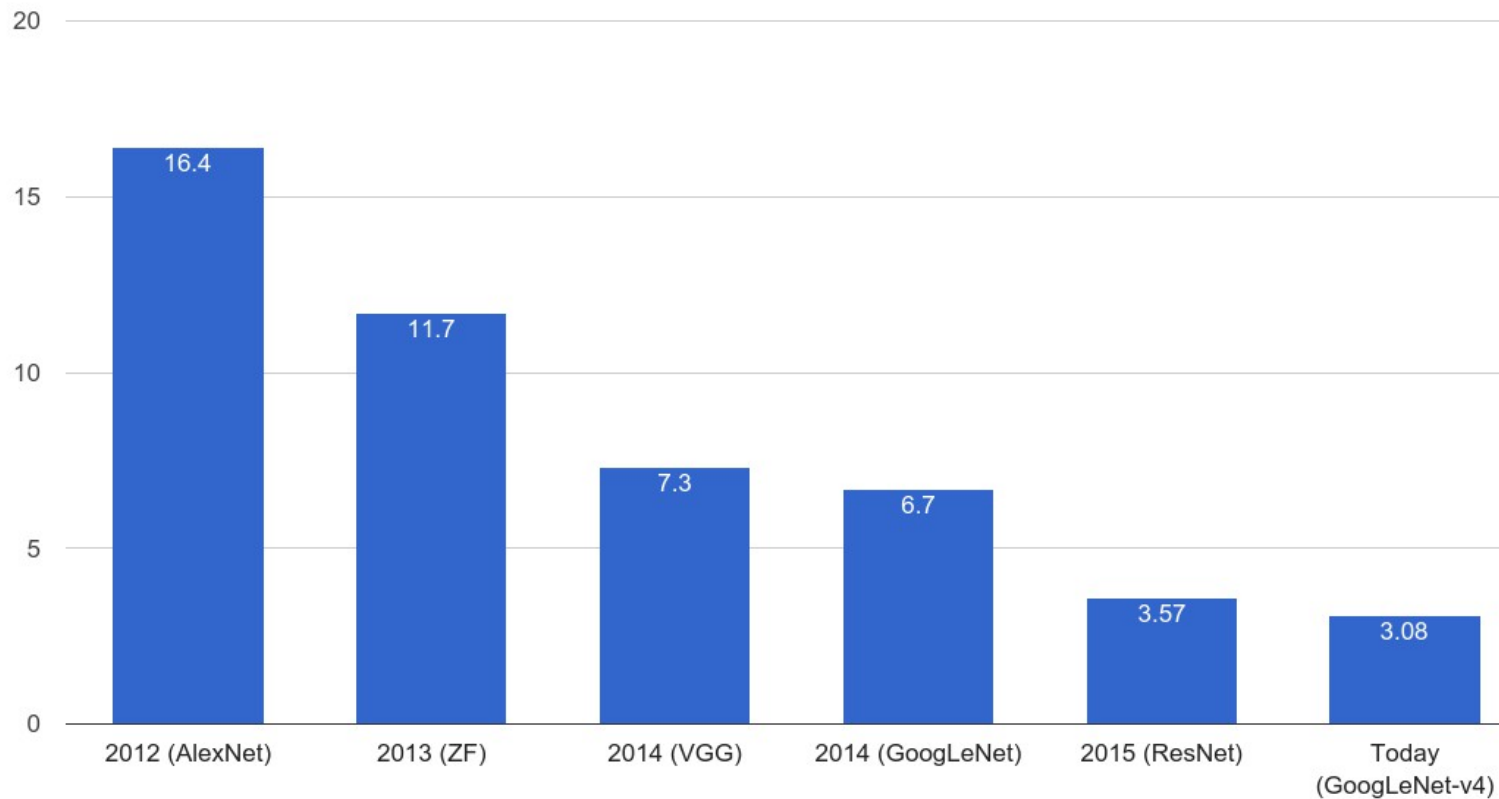
# Inception Resnet V2 Network



## Compressed View



## ImageNet Classification Error (Top 5)



Szegedy et al, Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, arXiv 2016

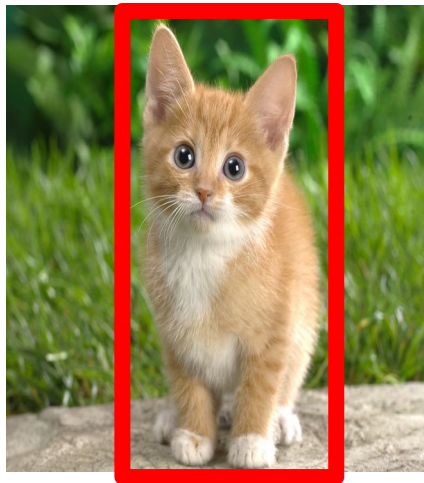
# Computer Vision Tasks

**Classification**



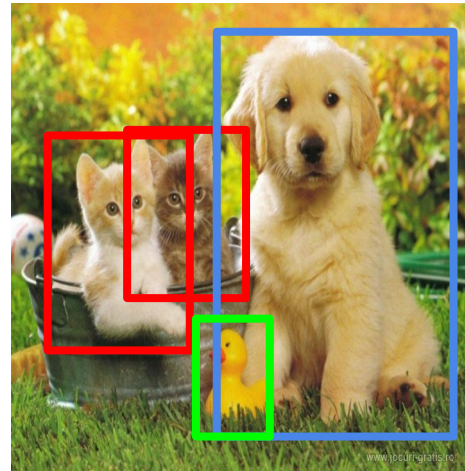
CAT

**Classification  
+ Localization**



CAT

**Object Detection**



CAT, DOG, DUCK

**Instance  
Segmentation**



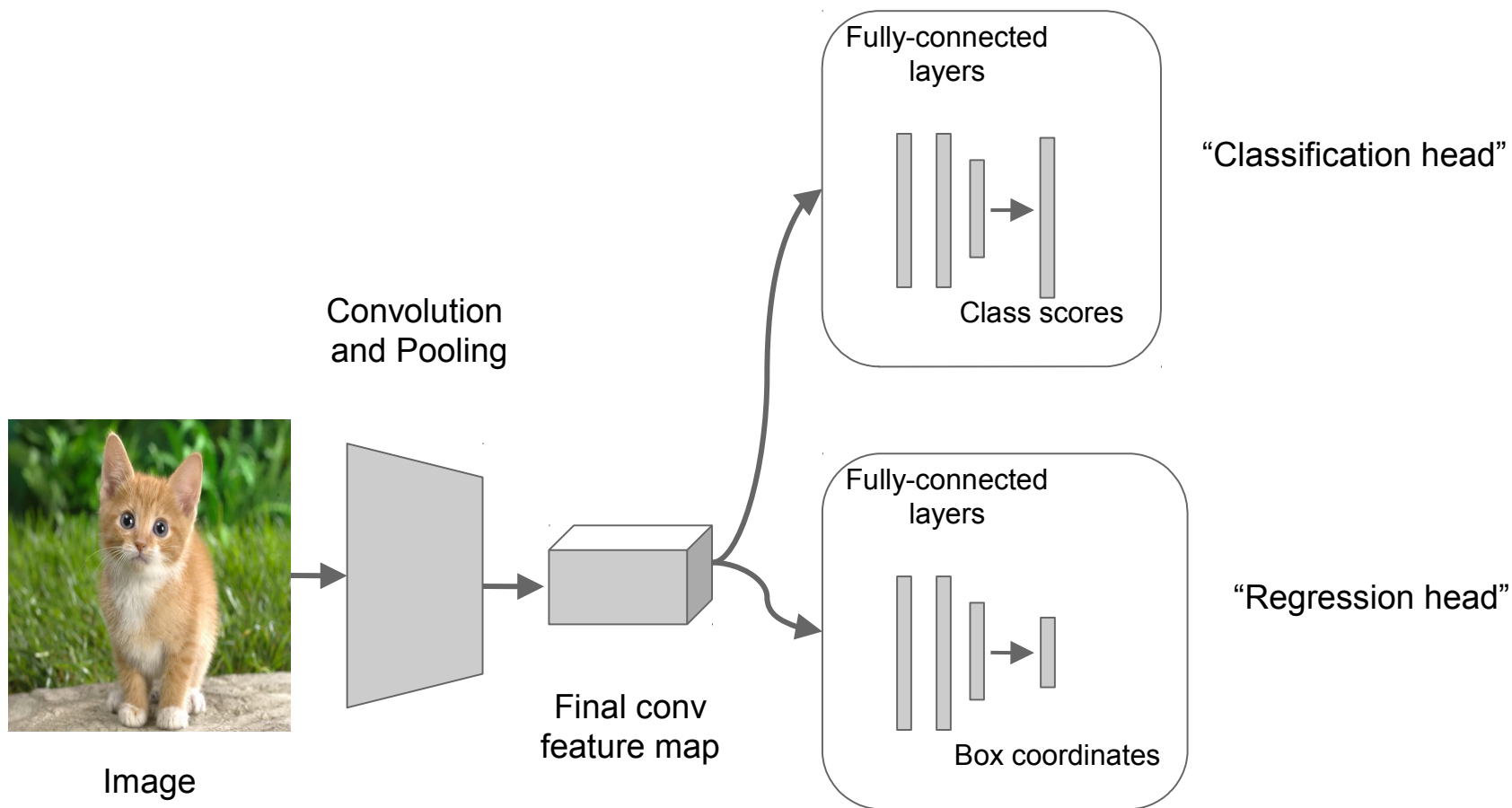
CAT, DOG, DUCK

Single object

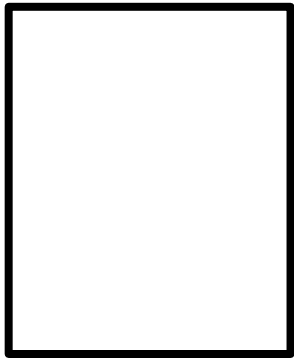
Multiple objects

# Simple Recipe for Classification + Localization

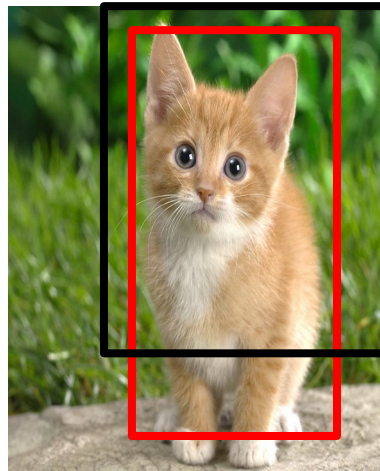
**Step 2:** Attach new fully-connected “regression head” to the network



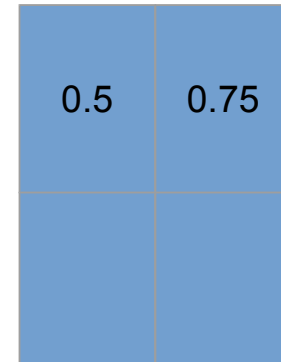
# Sliding Window: Overfeat



Network input:  
3 x 221 x 221

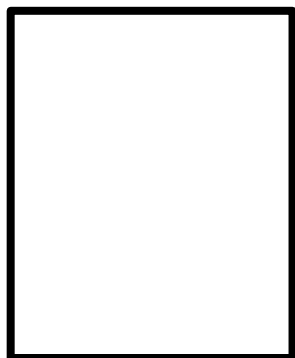


Larger image:  
3 x 257 x 257

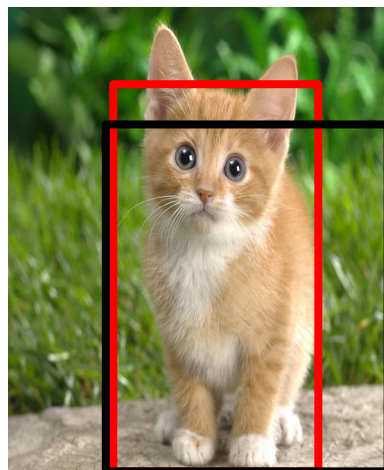


Classification scores:  
P(cat)

# Sliding Window: Overfeat



Network input:  
3 x 221 x 221

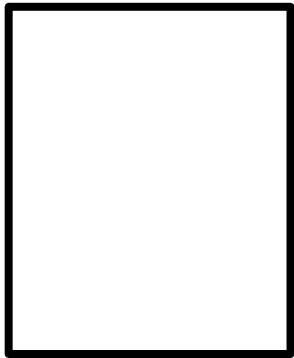


Larger image:  
3 x 257 x 257

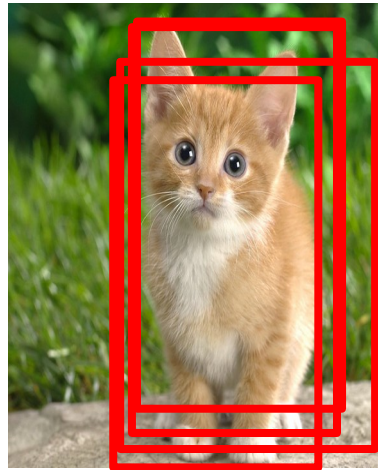
0.5	0.75
0.6	0.8

Classification scores:  
P(cat)

# Sliding Window: Overfeat



Network input:  
3 x 221 x 221



Larger image:  
3 x 257 x 257

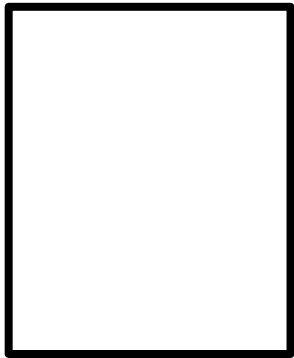
0.5	0.75
0.6	0.8

Classification scores:  
P(cat)

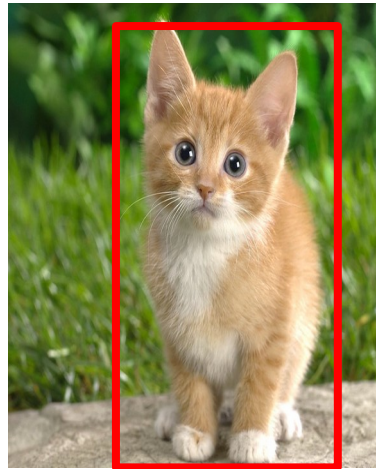


# Sliding Window: Overfeat

Greedily merge boxes and scores (details in paper)



Network input:  
3 x 221 x 221



Larger image:  
3 x 257 x 257

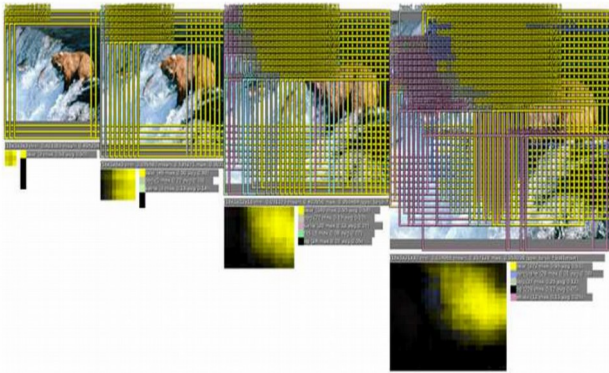
0.8

Classification score:  
P(cat)

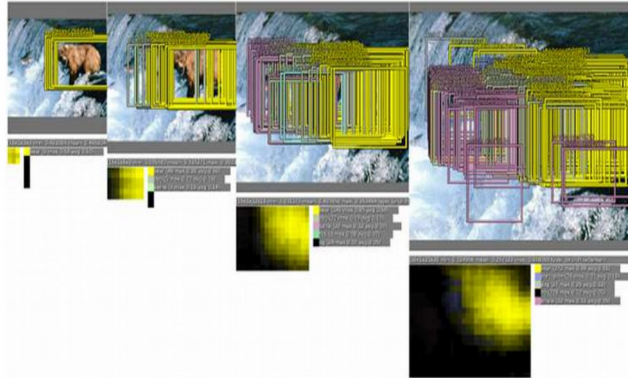
# Sliding Window: Overfeat

In practice use many sliding window locations and multiple scales

Window positions + score maps



Box regression outputs



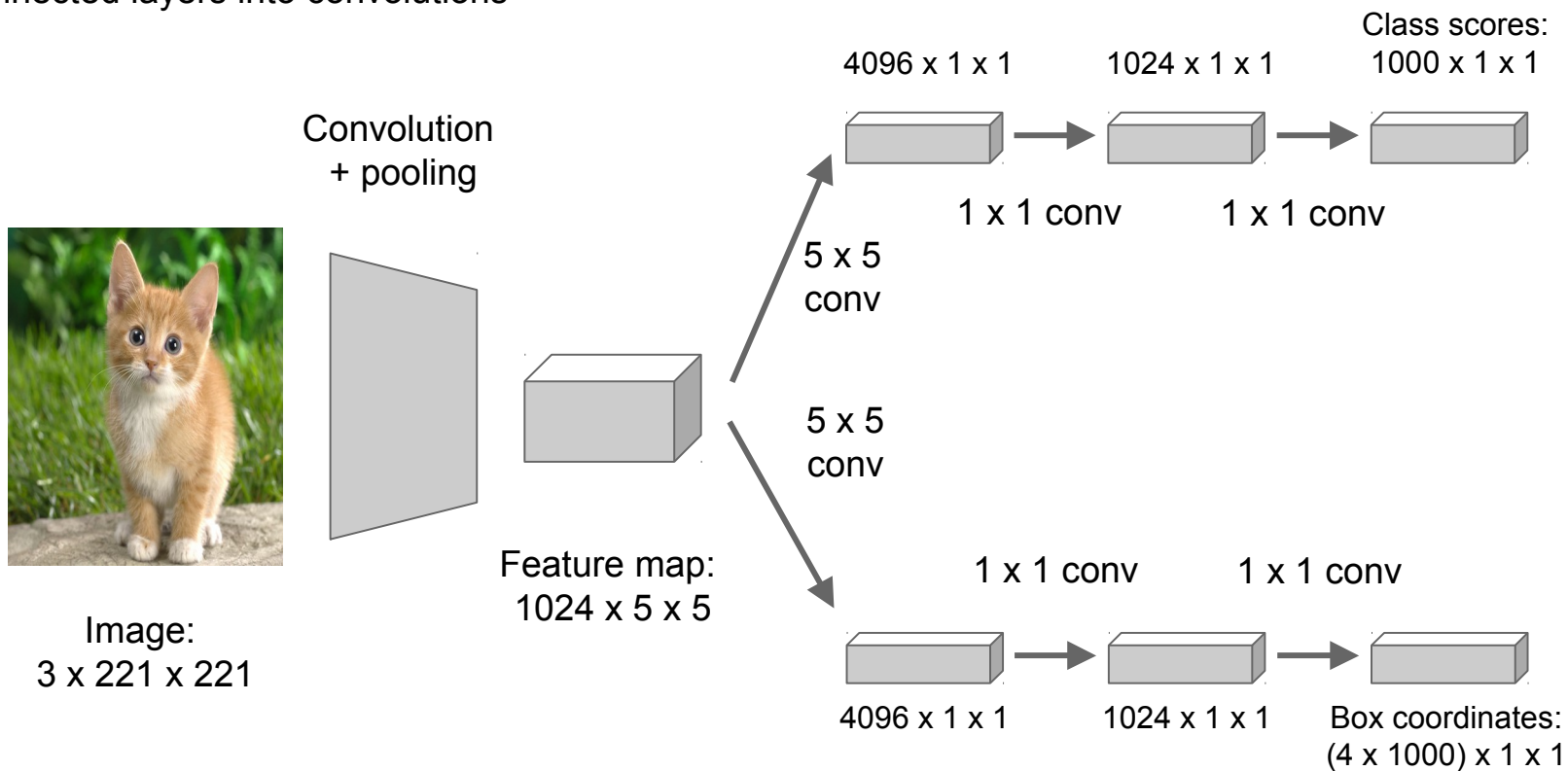
Final Predictions



Sermanet et al, "Integrated Recognition, Localization and Detection using Convolutional Networks", ICLR 2014

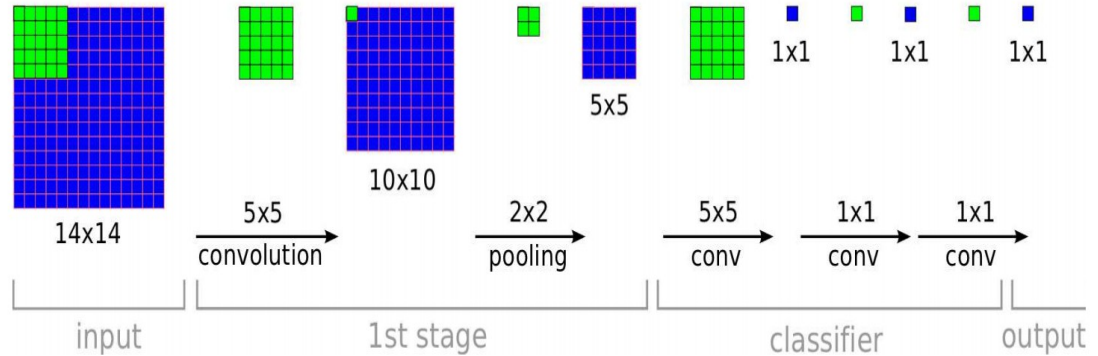
# Efficient Sliding Window: Overfeat

Efficient sliding window by converting fully-connected layers into convolutions

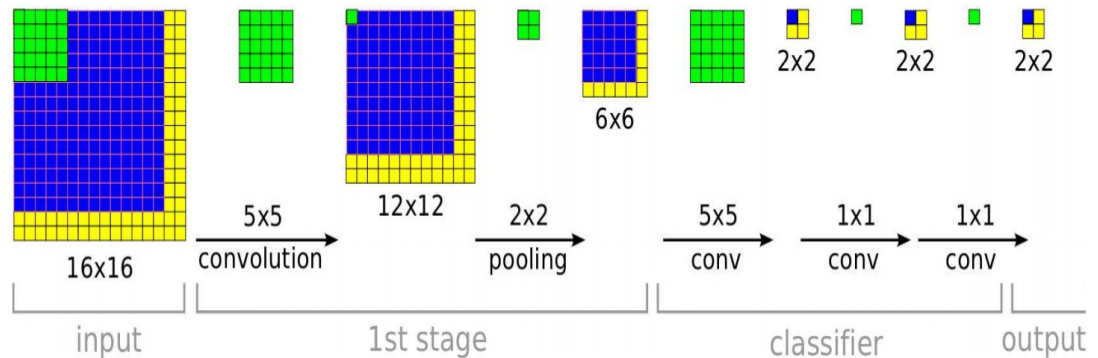


# Efficient Sliding Window: Overfeat

**Training time:** Small image, 1 x 1 classifier output



**Test time:** Larger image, 2 x 2 classifier output, only extra compute at yellow regions



Sermanet et al, "Integrated Recognition, Localization and Detection using Convolutional Networks", ICLR 2014

# Computer Vision Tasks

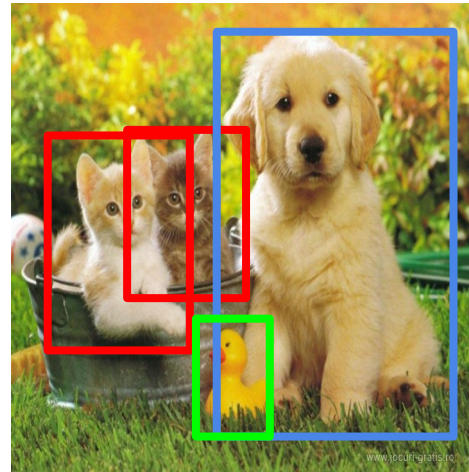
Classification



Classification  
+ Localization



Object Detection

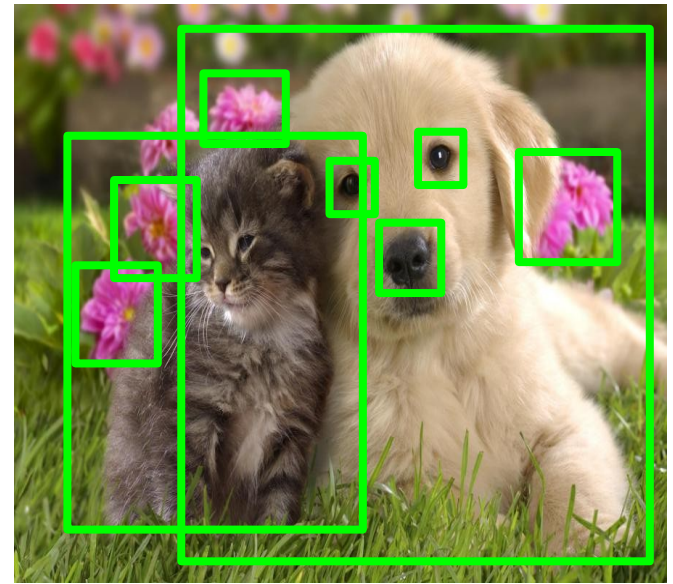


Instance  
Segmentation



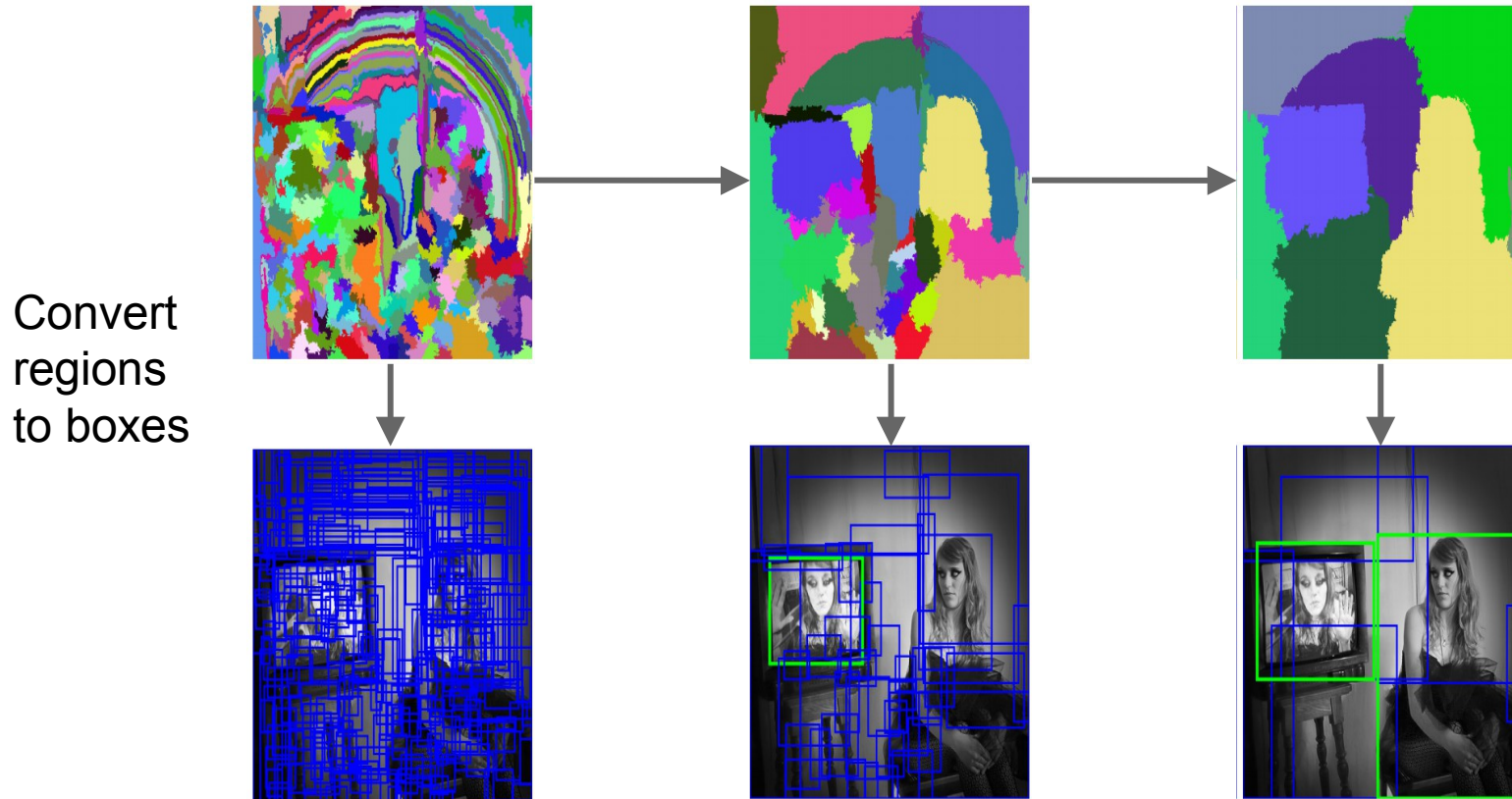
# Region Proposals

- Find “blobby” image regions that are likely to contain objects
- “Class-agnostic” object detector
- Look for “blob-like” regions



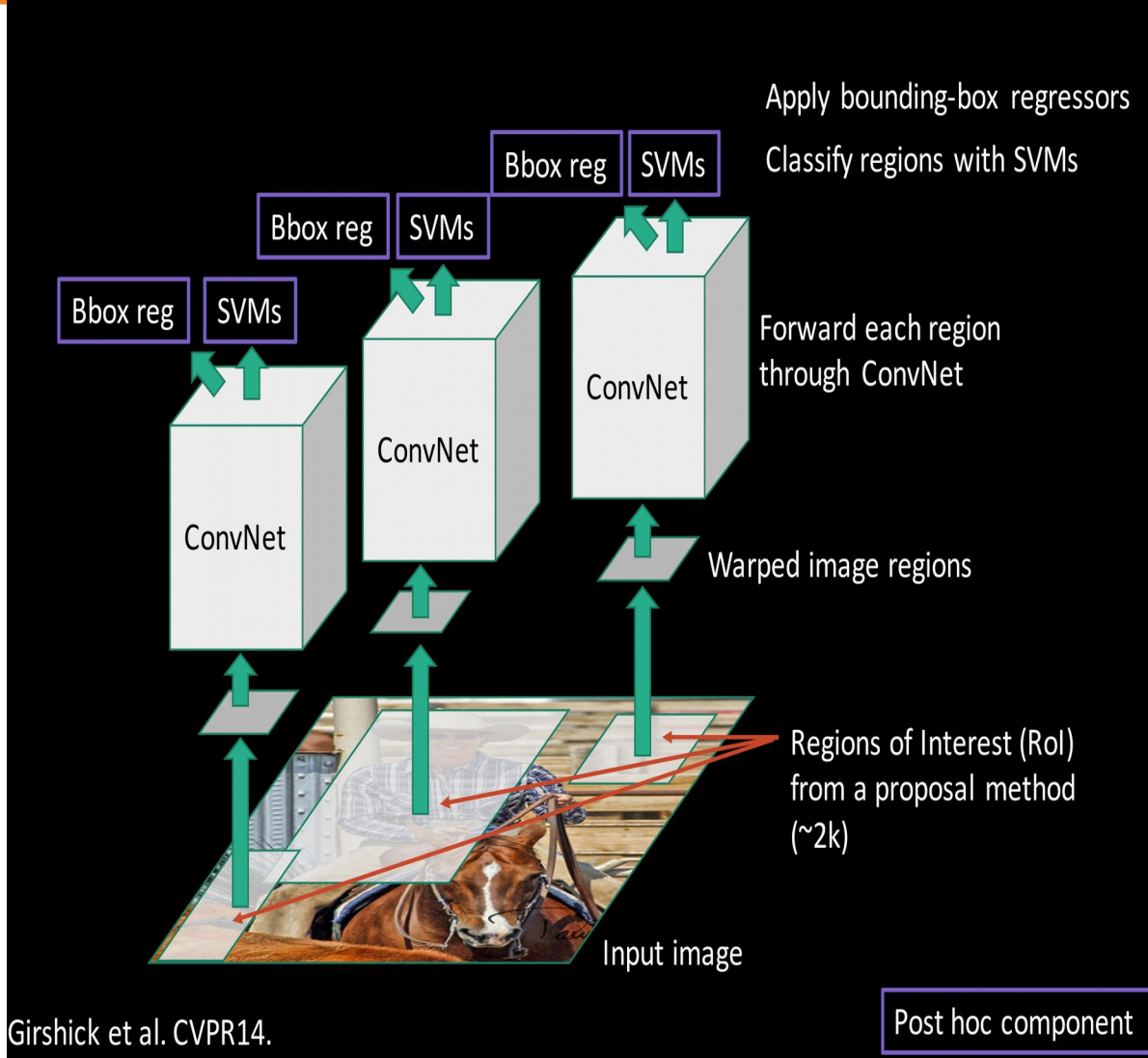
# Region Proposals: Selective Search

Bottom-up segmentation, merging regions at multiple scales



Uijlings et al, "Selective Search for Object Recognition", IJCV 2013

# R-CNN

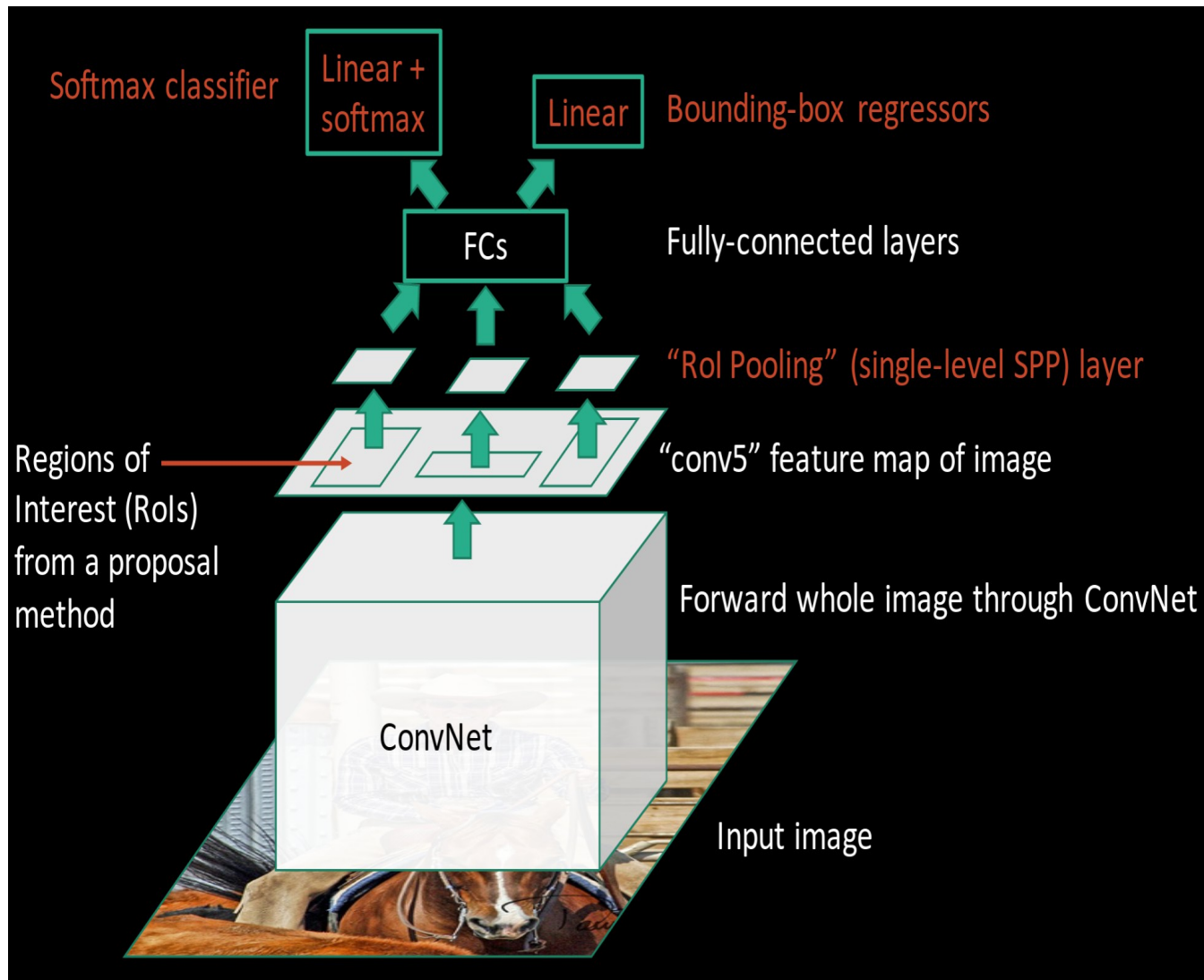


Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR 2014

Slide credit: Ross Girshick



# Fast R-CNN



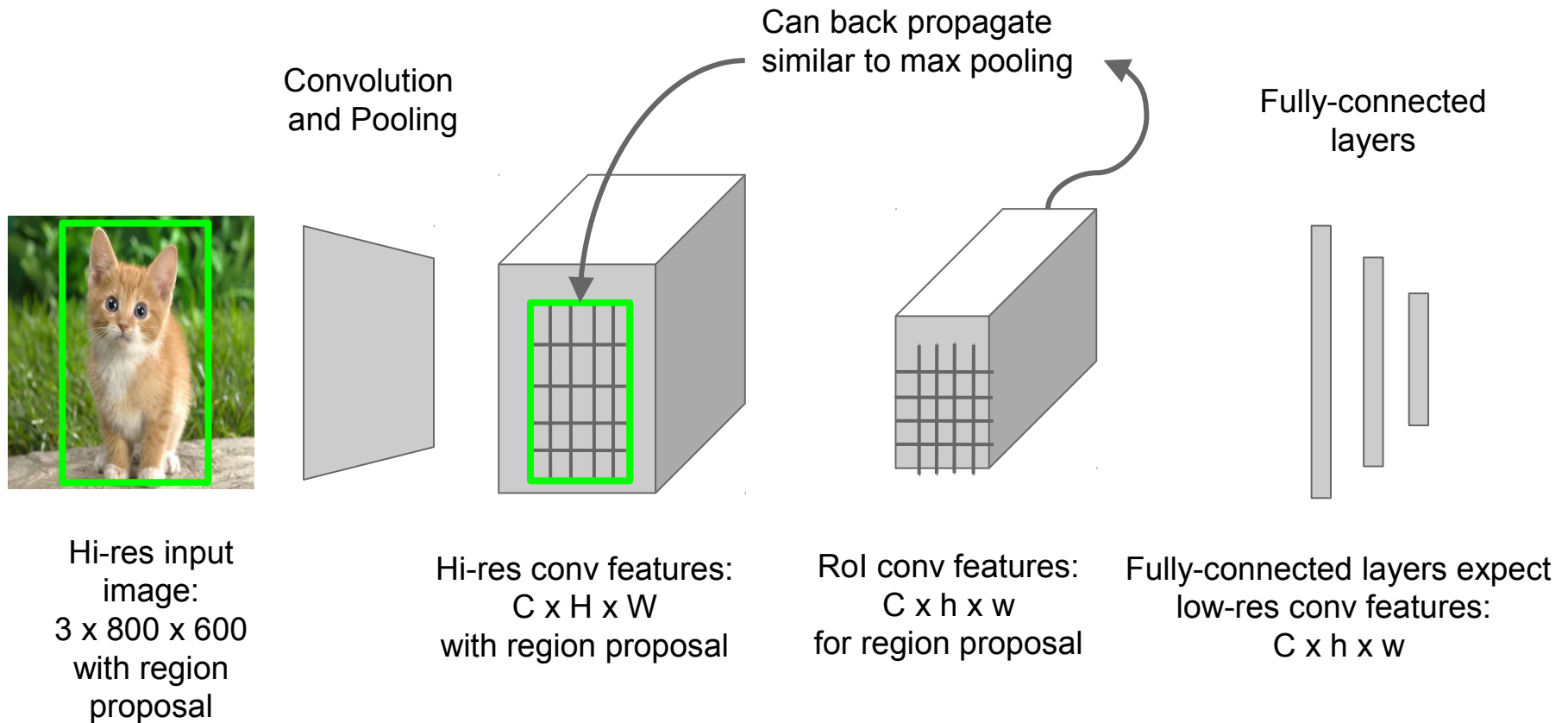
**R-CNN Problems:**  
Slow at test-time due to independent forward passes of the CNN

**Solution:**  
Share computation of convolutional layers between proposals for an image

**R-CNN Problems:**  
- Post-hoc training: CNN not updated in response to final classifiers and regressors  
- Complex training pipeline

**Solution:**  
Just train the whole system end-to-end all at once!

# Fast R-CNN: Region of Interest Pooling



# Faster R-CNN: Training

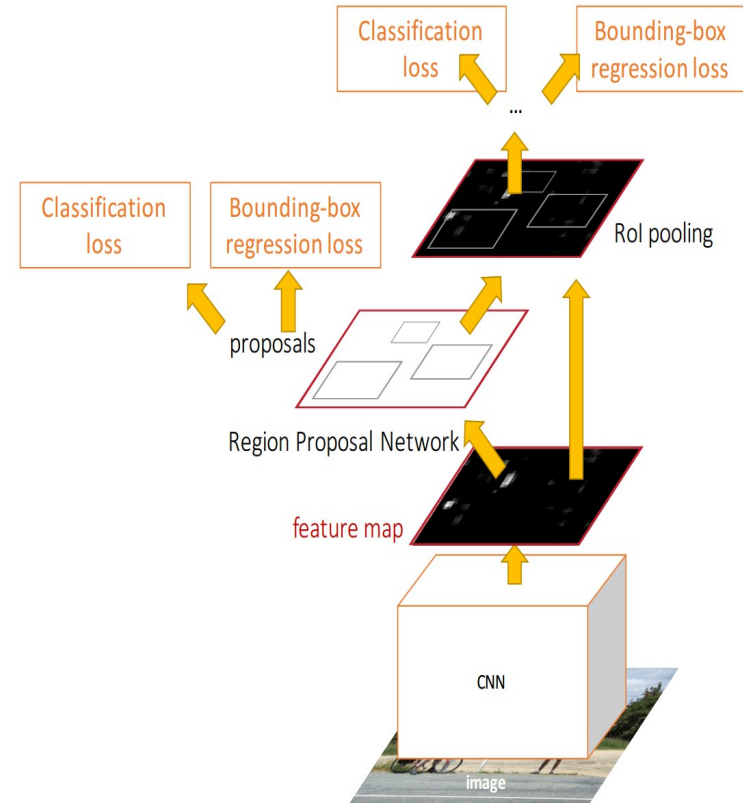
In the paper: Ugly pipeline

- Use alternating optimization to train RPN, then Fast R-CNN with RPN proposals, etc.
- More complex than it has to be

Since publication: Joint training!

One network, four losses

- RPN classification (anchor good / bad)
- RPN regression (anchor  $\rightarrow$  proposal)
- Fast R-CNN classification (over classes)
- Fast R-CNN regression (proposal  $\rightarrow$  box)



Slide credit: Ross Girschick

# Faster R-CNN: Results

	<b>R-CNN</b>	<b>Fast R-CNN</b>	<b>Faster R-CNN</b>
Test time per image (with proposals)	50 seconds	2 seconds	<b>0.2 seconds</b>
(Speedup)	1x	25x	<b>250x</b>
mAP (VOC 2007)	66.0	<b>66.9</b>	<b>66.9</b>

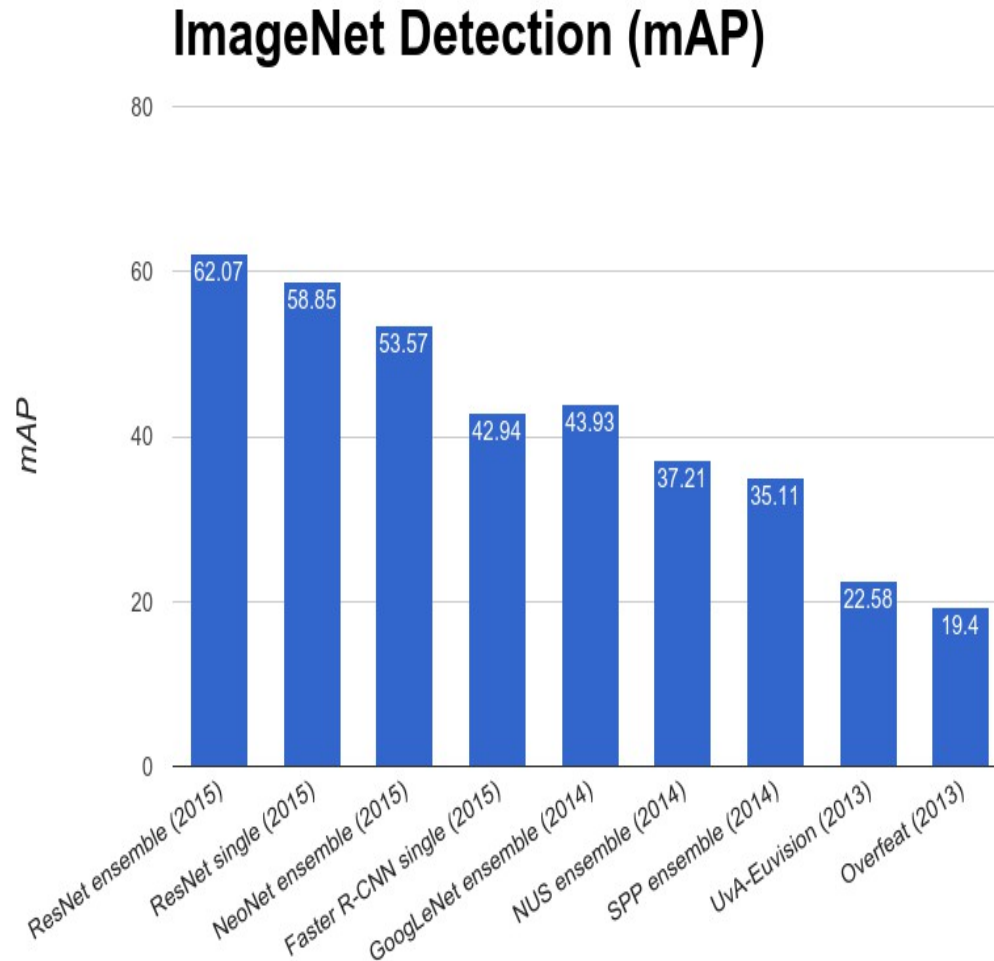
# Object Detection State-of-the-art:

## ResNet 101 + Faster R-CNN + some extras

training data	COCO train		COCO trainval	
test data	COCO val		COCO test-dev	
mAP	@.5	@[.5, .95]	@.5	@[.5, .95]
baseline Faster R-CNN (VGG-16)	41.5	21.2		
baseline Faster R-CNN (ResNet-101)	48.4	27.2		
+box refinement	49.9	29.9		
+context	51.1	30.0	53.3	32.2
+multi-scale testing	53.8	32.5	<b>55.7</b>	<b>34.9</b>
ensemble			<b>59.0</b>	<b>37.4</b>

He et. al, "Deep Residual Learning for Image Recognition", arXiv 2015

# ImageNet Detection 2013 - 2015



# YOLO: You Only Look Once

## Detection as Regression

Divide image into  $S \times S$  grid

Within each grid cell predict:

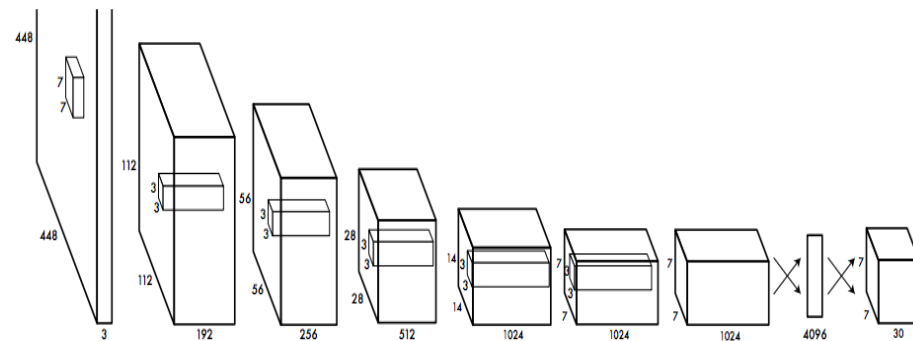
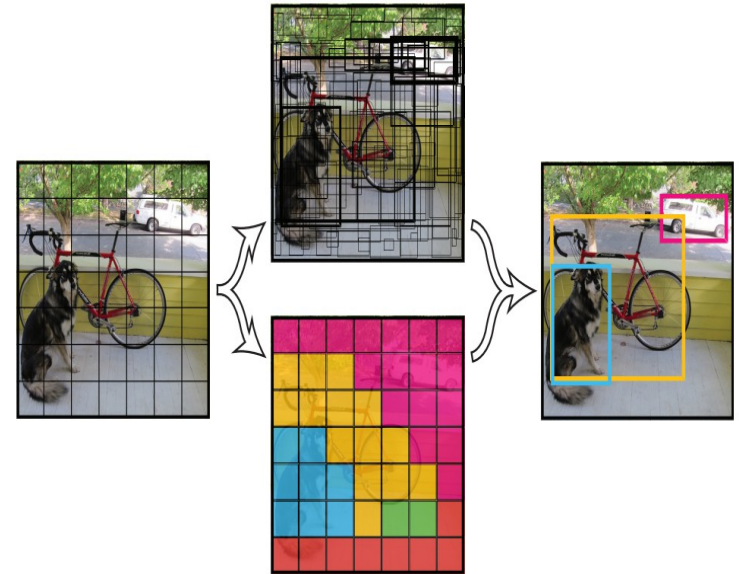
B Boxes: 4 coordinates + confidence

Class scores: C numbers

Regression from image to  
 $7 \times 7 \times (5 * B + C)$  tensor

Direct prediction using a CNN

Redmon et al, "You Only Look Once:  
Unified, Real-Time Object Detection", arXiv 2015



# YOLO: You Only Look Once

## Detection as Regression

Faster than Faster R-CNN, but not as good

Redmon et al, "You Only Look Once: Unified, Real-Time Object Detection", arXiv 2015

Real-Time Detectors	Train	mAP	FPS
100Hz DPM [30]	2007	16.0	100
30Hz DPM [30]	2007	26.1	30
Fast YOLO	2007+2012	52.7	<b>155</b>
YOLO	2007+2012	<b>63.4</b>	45
<hr/>			
Less Than Real-Time			
Fastest DPM [37]	2007	30.4	15
R-CNN Minus R [20]	2007	53.5	6
Fast R-CNN [14]	2007+2012	70.0	0.5
Faster R-CNN VGG-16[27]	2007+2012	73.2	7
Faster R-CNN ZF [27]	2007+2012	62.1	18



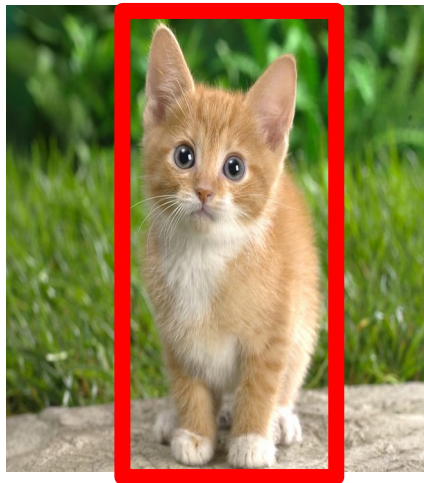
# Computer Vision Tasks

**Classification**

**Classification  
+ Localization**

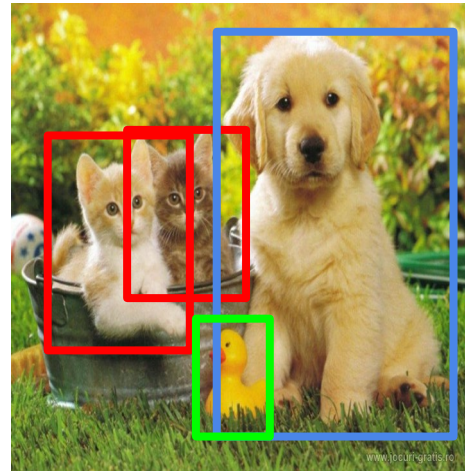
**Object Detection**

**Segmentation**

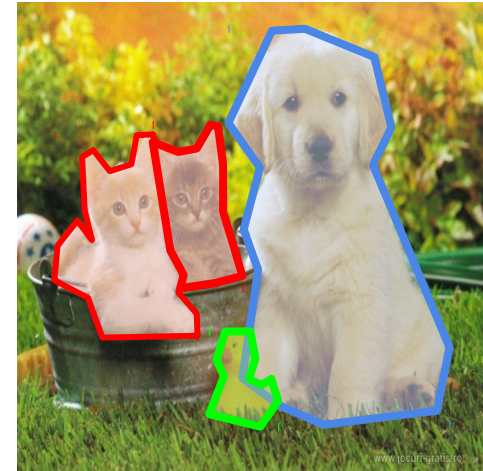


CAT

CAT



CAT, DOG, DUCK



CAT, DOG, DUCK

Single object

Multiple objects

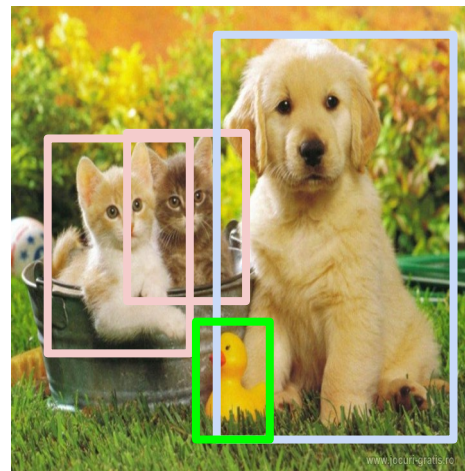
# Today

Classification

Classification  
+ Localization

Object Detection

Segmentation



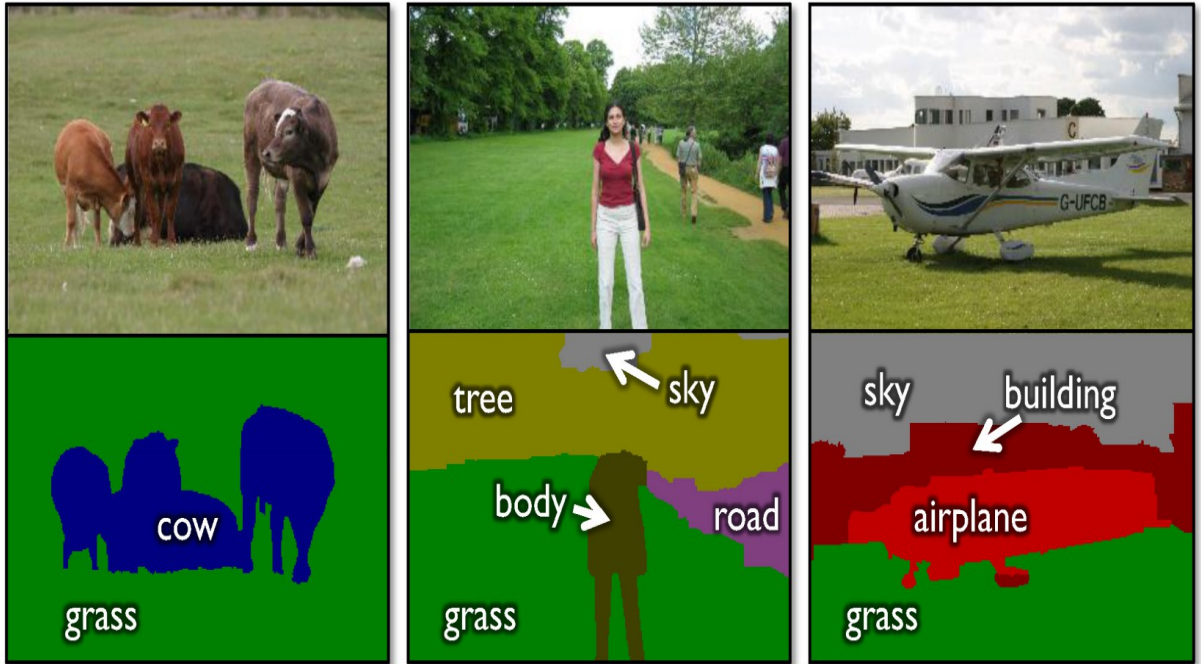
Today

# Semantic Segmentation

Label every pixel!

Don't differentiate instances (cows)

Classic computer vision problem



object classes	building	grass	tree	cow	sheep	sky	airplane	water	face	car	
	bicycle	flower	sign	bird	book	chair	road	cat	dog	body	boat

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

# Instance Segmentation

Detect instances,  
give category, label  
pixels

“simultaneous  
detection and  
segmentation” (SDS)

Lots of recent work  
(MS-COCO)

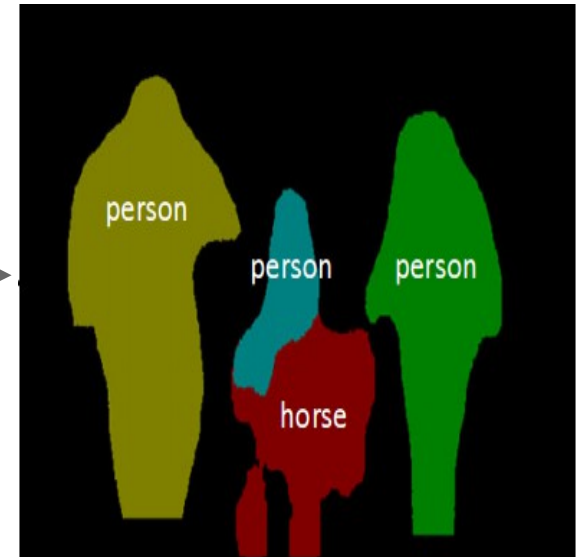
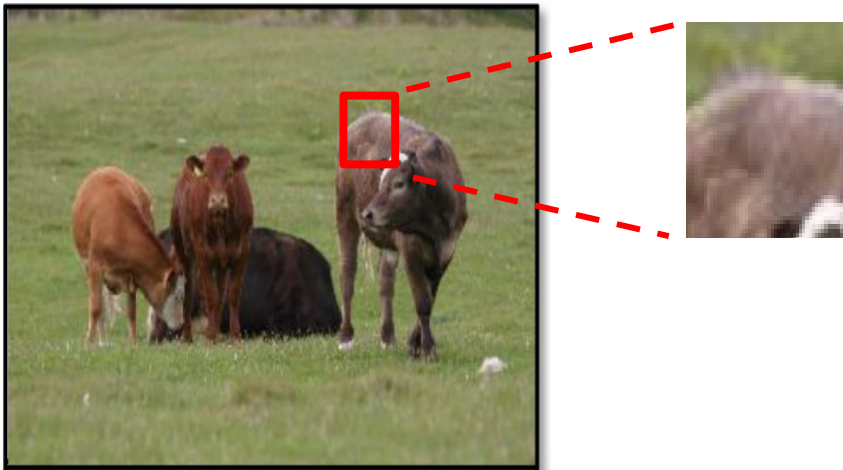


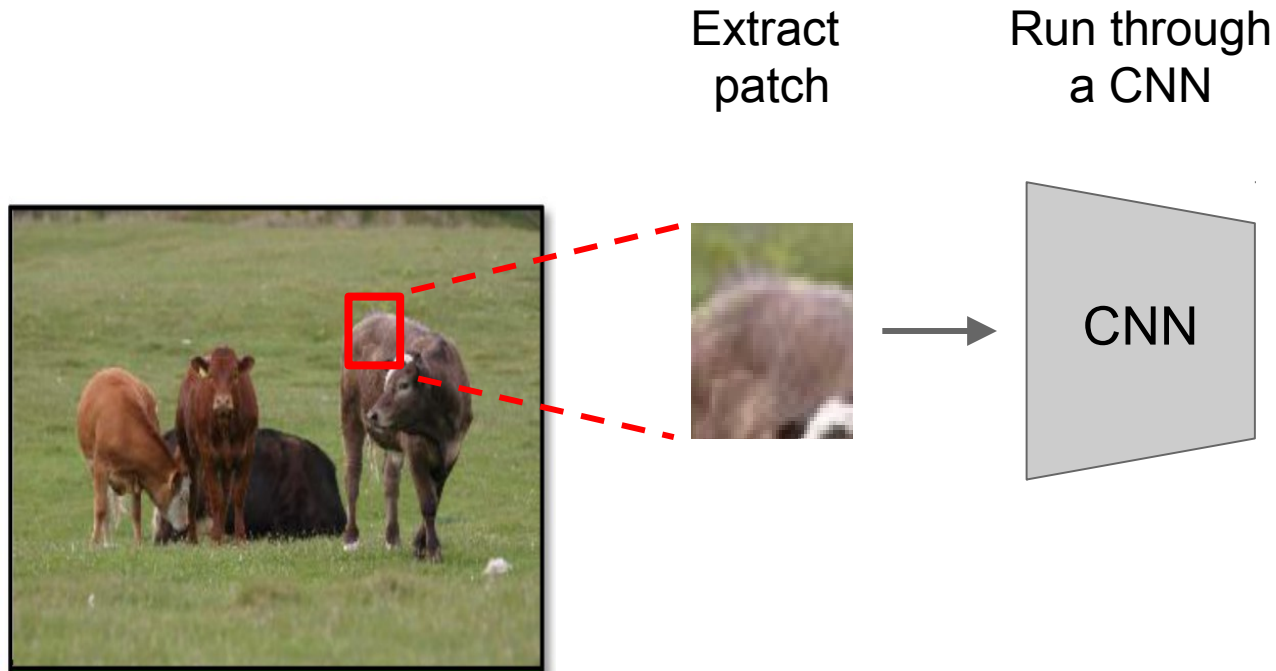
Figure credit: Dai et al, “Instance-aware Semantic Segmentation via Multi-task Network Cascades”, arXiv 2015

# Semantic Segmentation

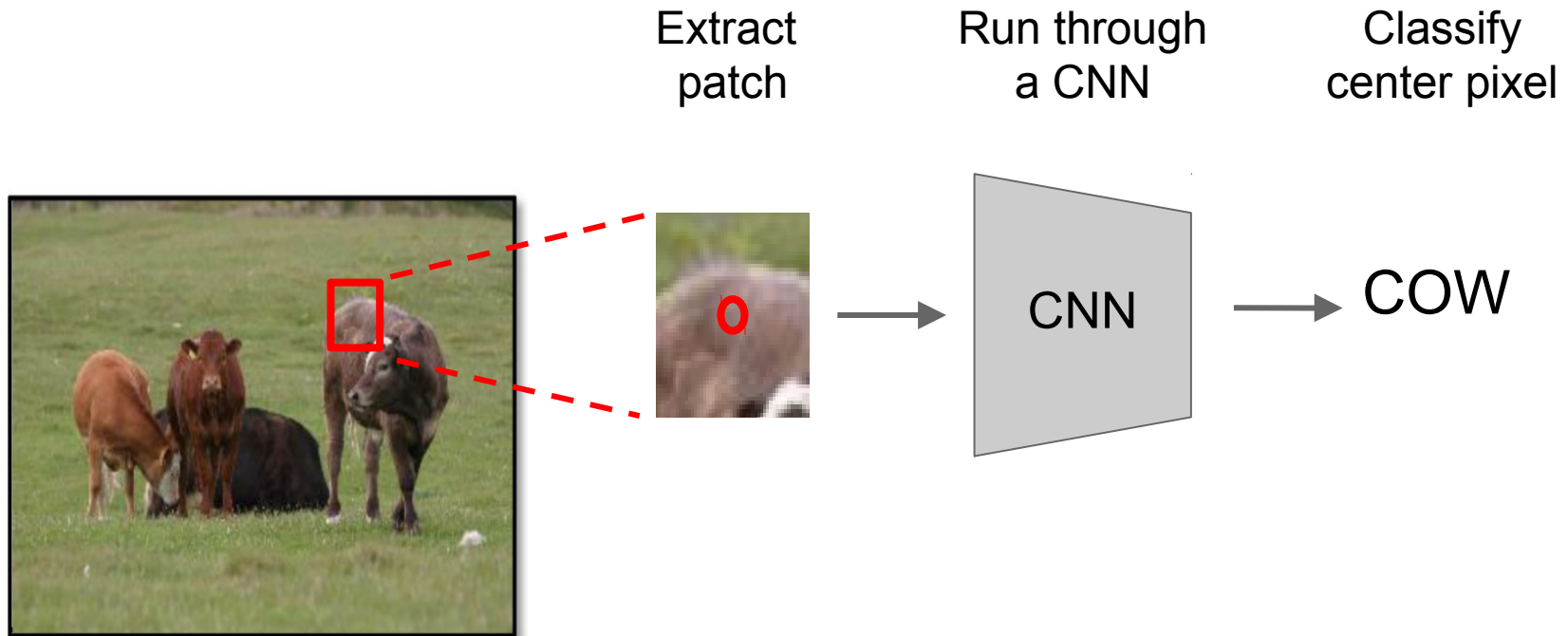
Extract  
patch



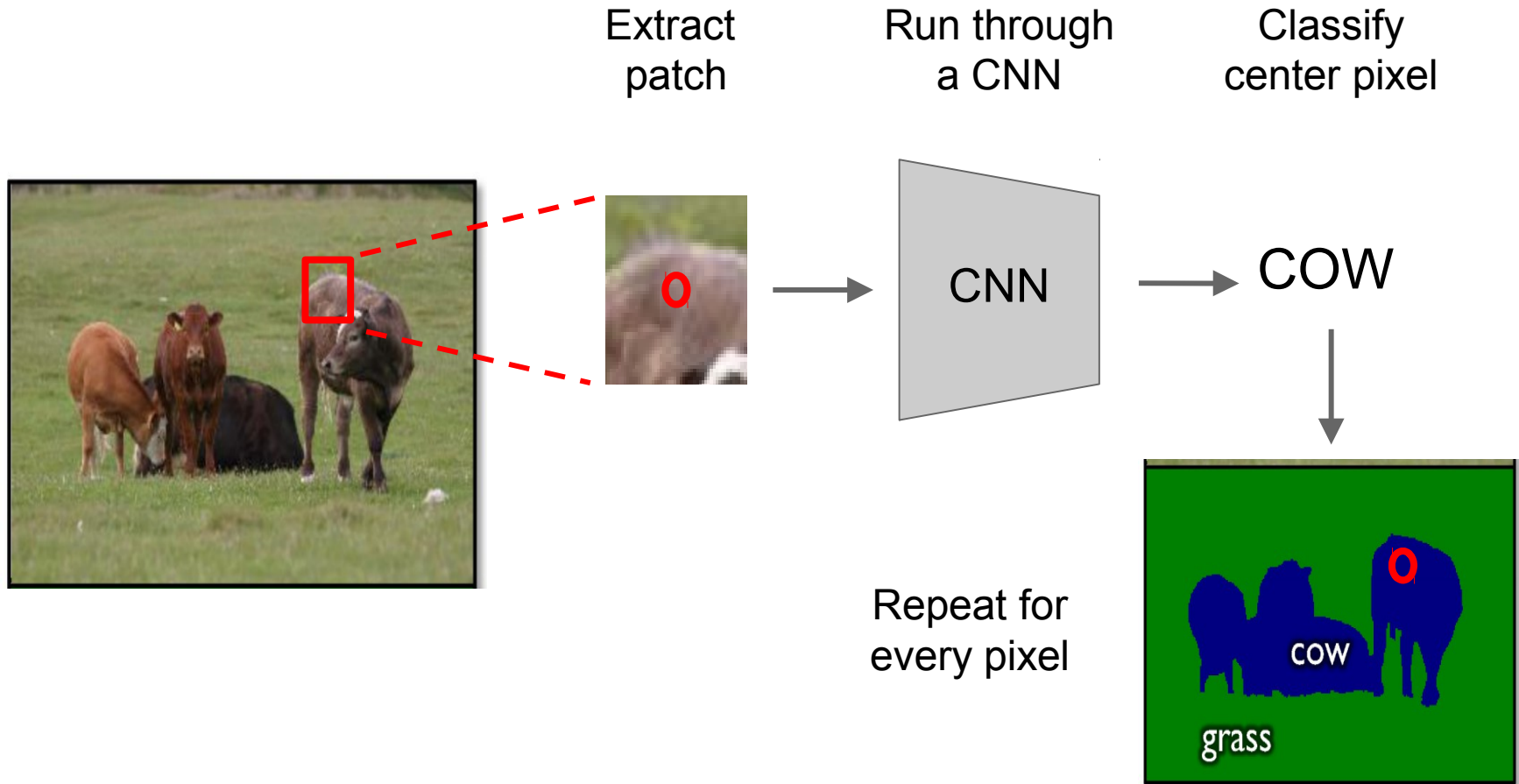
# Semantic Segmentation



# Semantic Segmentation



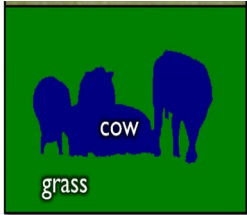
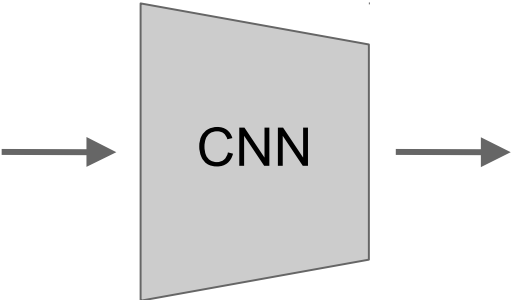
# Semantic Segmentation





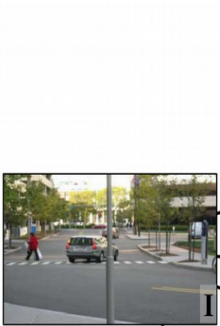
# Semantic Segmentation

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



Smaller output  
due to pooling

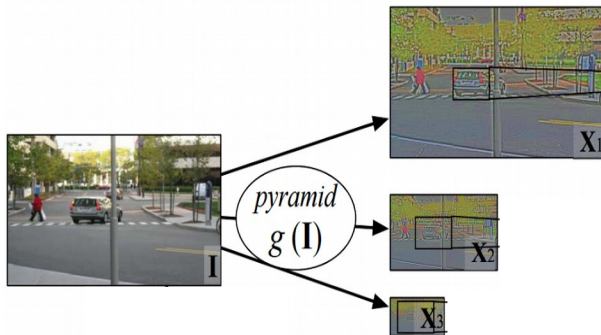
# Semantic Segmentation: Multi-Scale



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

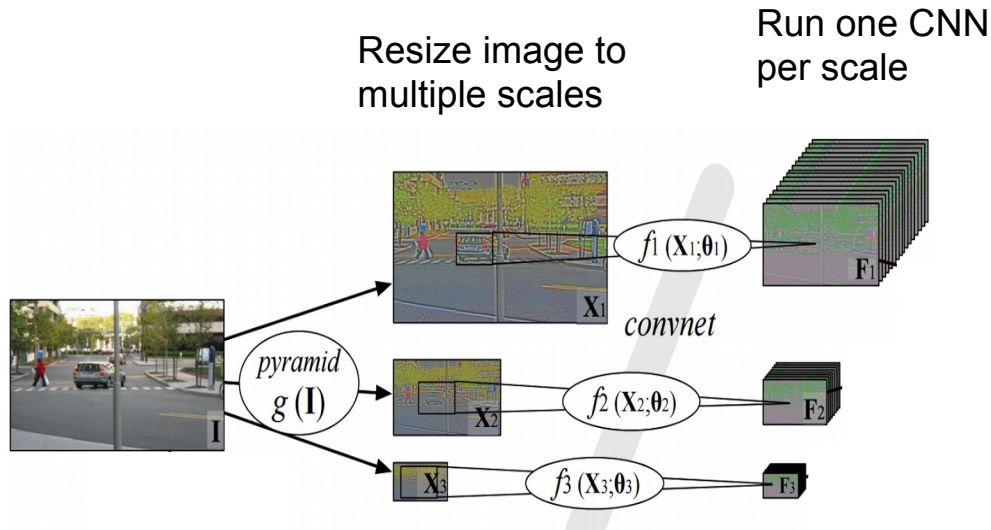
# Semantic Segmentation: Multi-Scale

Resize image to  
multiple scales



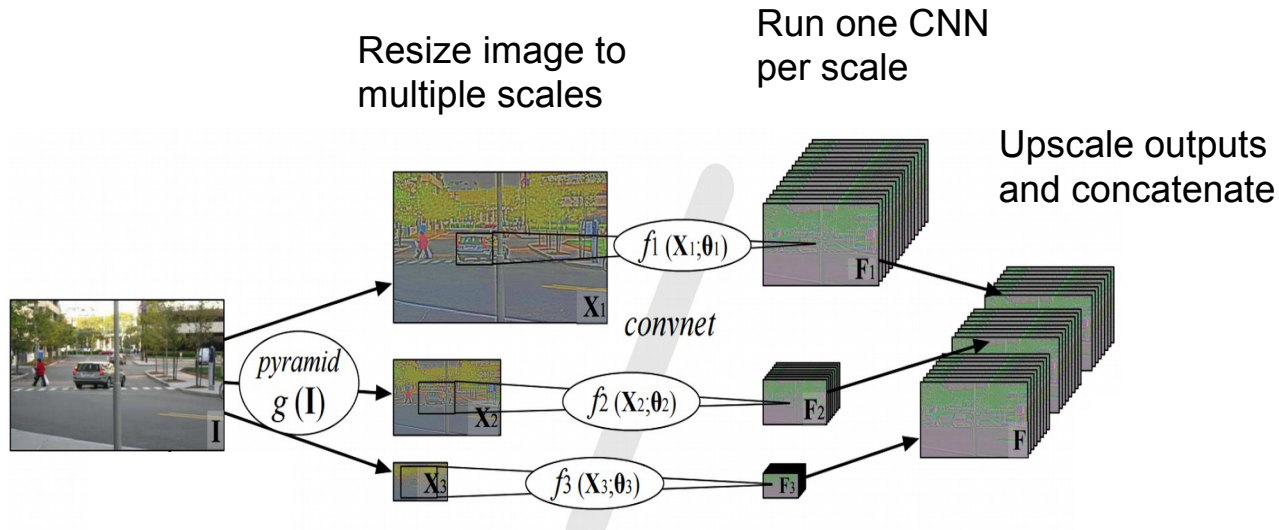
Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013

# Semantic Segmentation: Multi-Scale



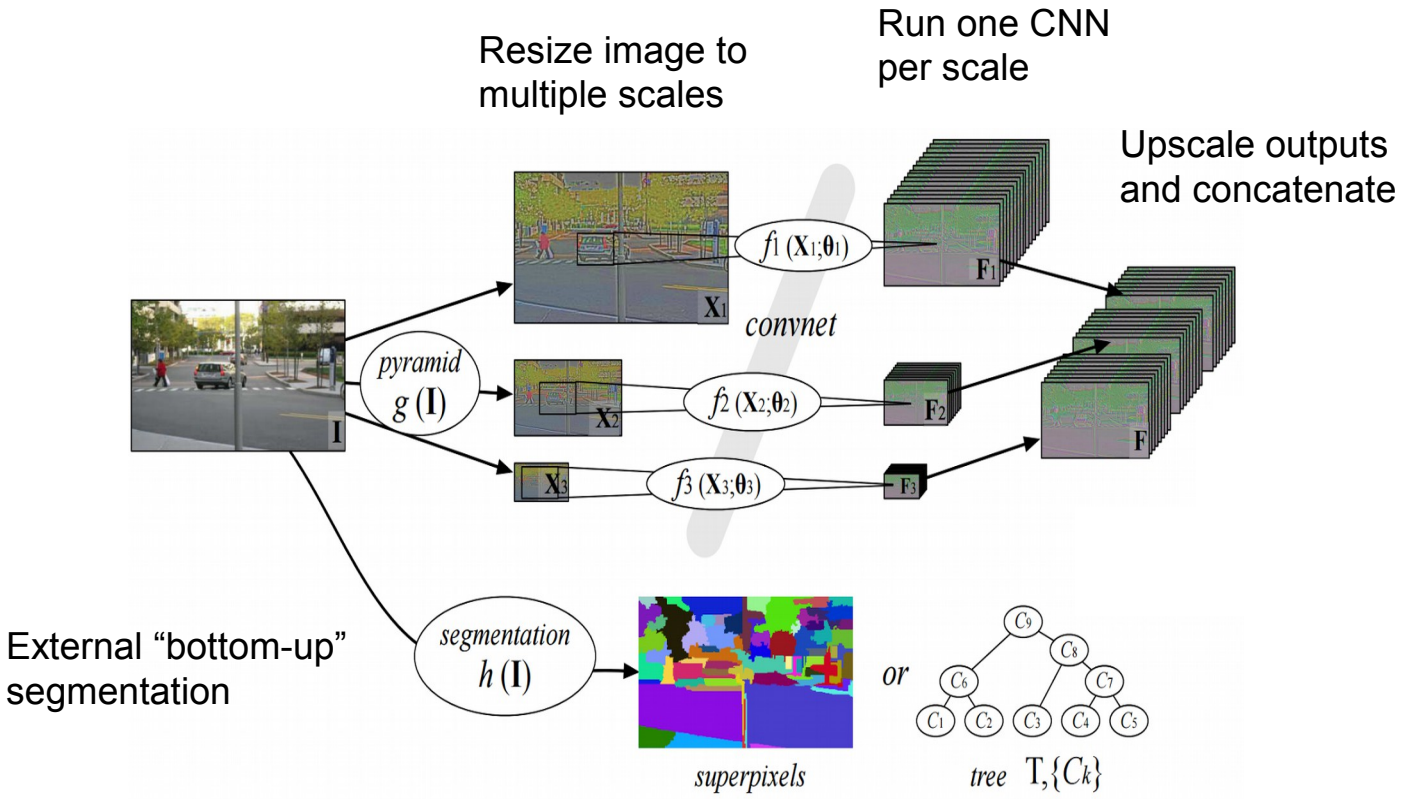
Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013

# Semantic Segmentation: Multi-Scale



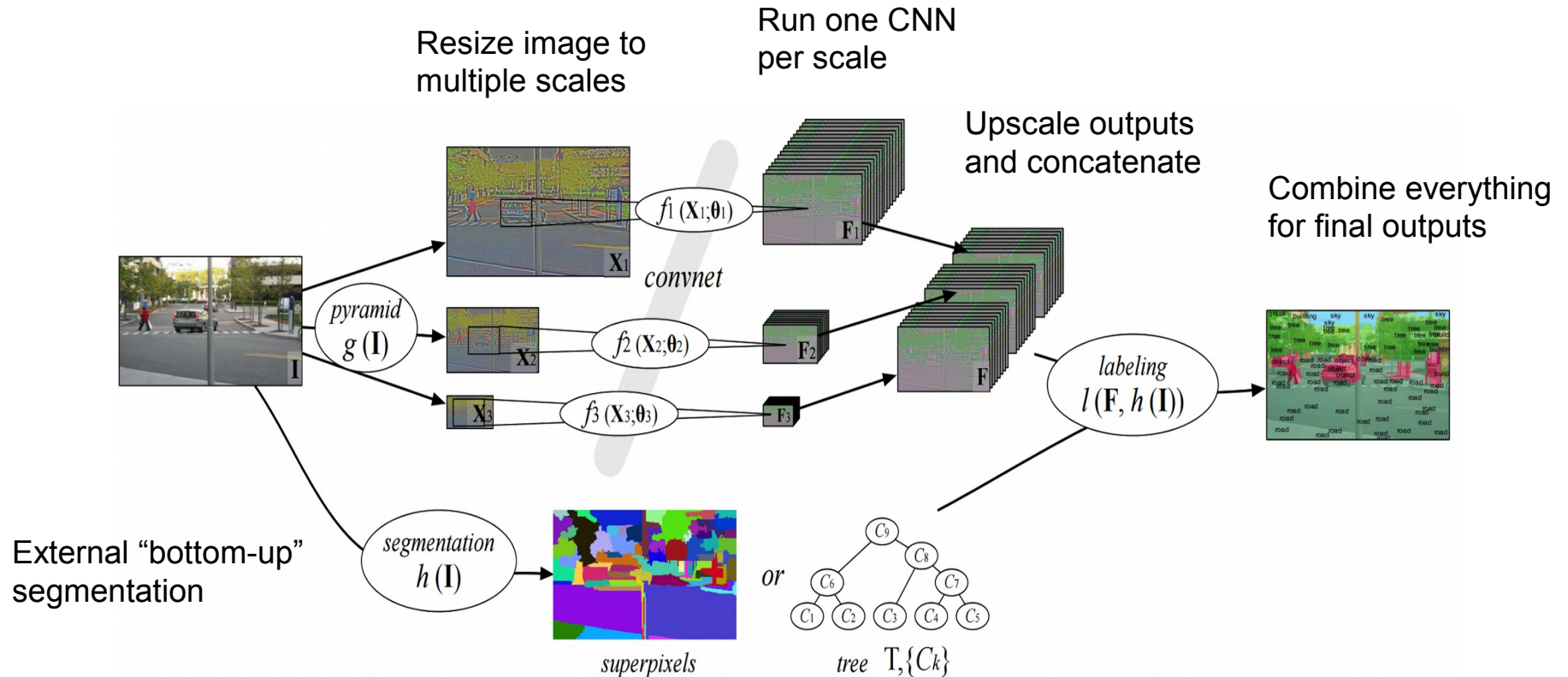
Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013

# Semantic Segmentation: Multi-Scale



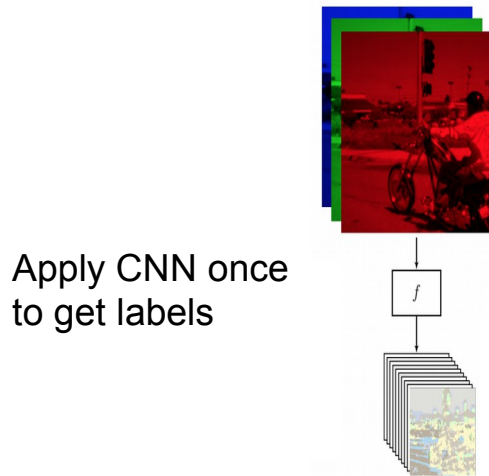
Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013

# Semantic Segmentation: Multi-Scale



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

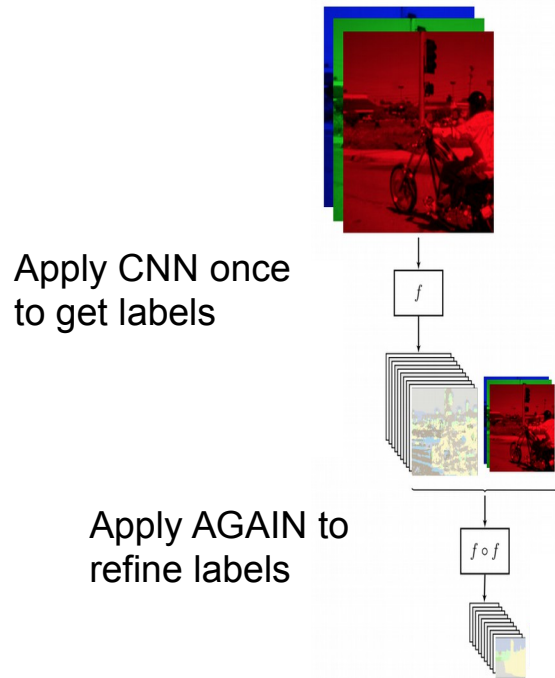
# Semantic Segmentation: Refinement



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

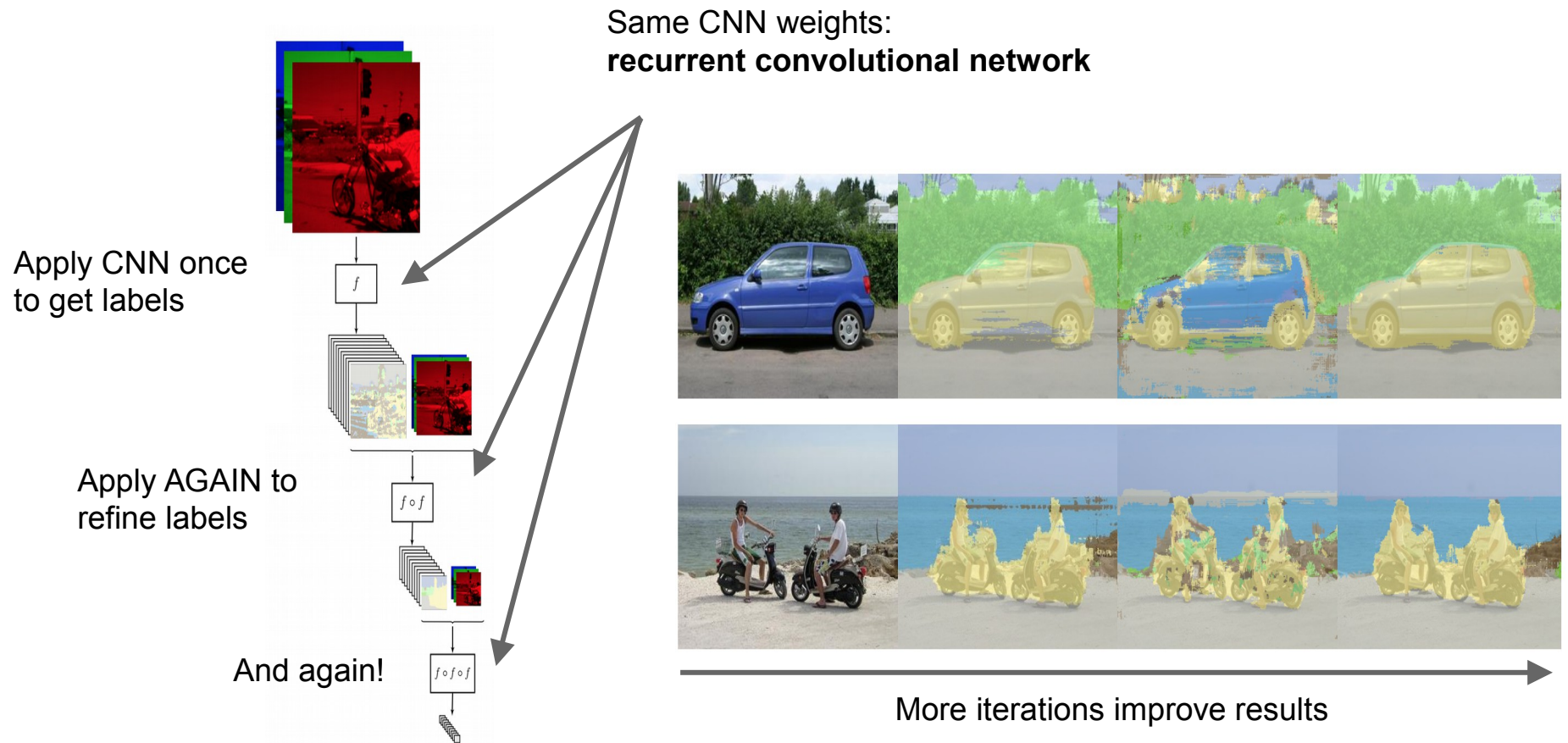


# Semantic Segmentation: Refinement



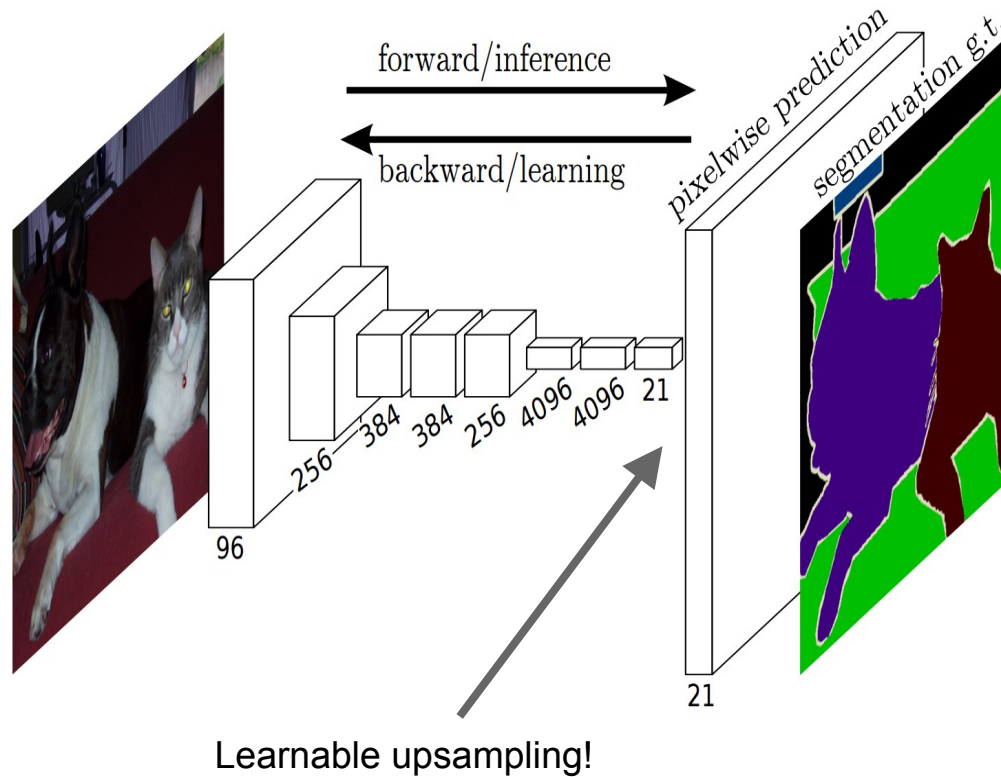
Pinheiro and Collobert, "Recurrent Convolutional Neural Networks for Scene Labeling", ICML 2014

# Semantic Segmentation: Refinement



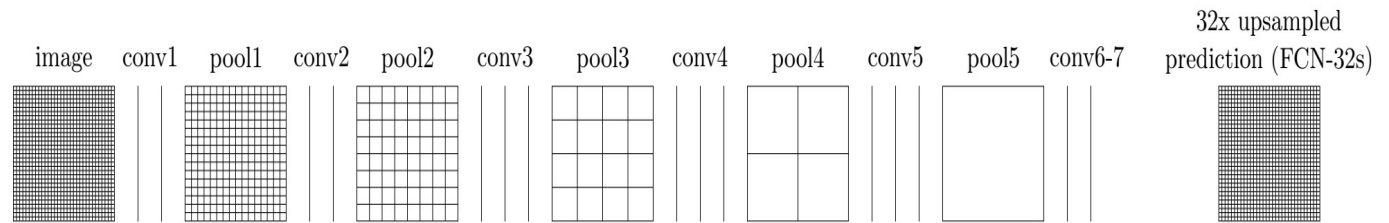
Pinheiro and Collobert, "Recurrent Convolutional Neural Networks for Scene Labeling", ICML 2014

# Semantic Segmentation: Upsampling



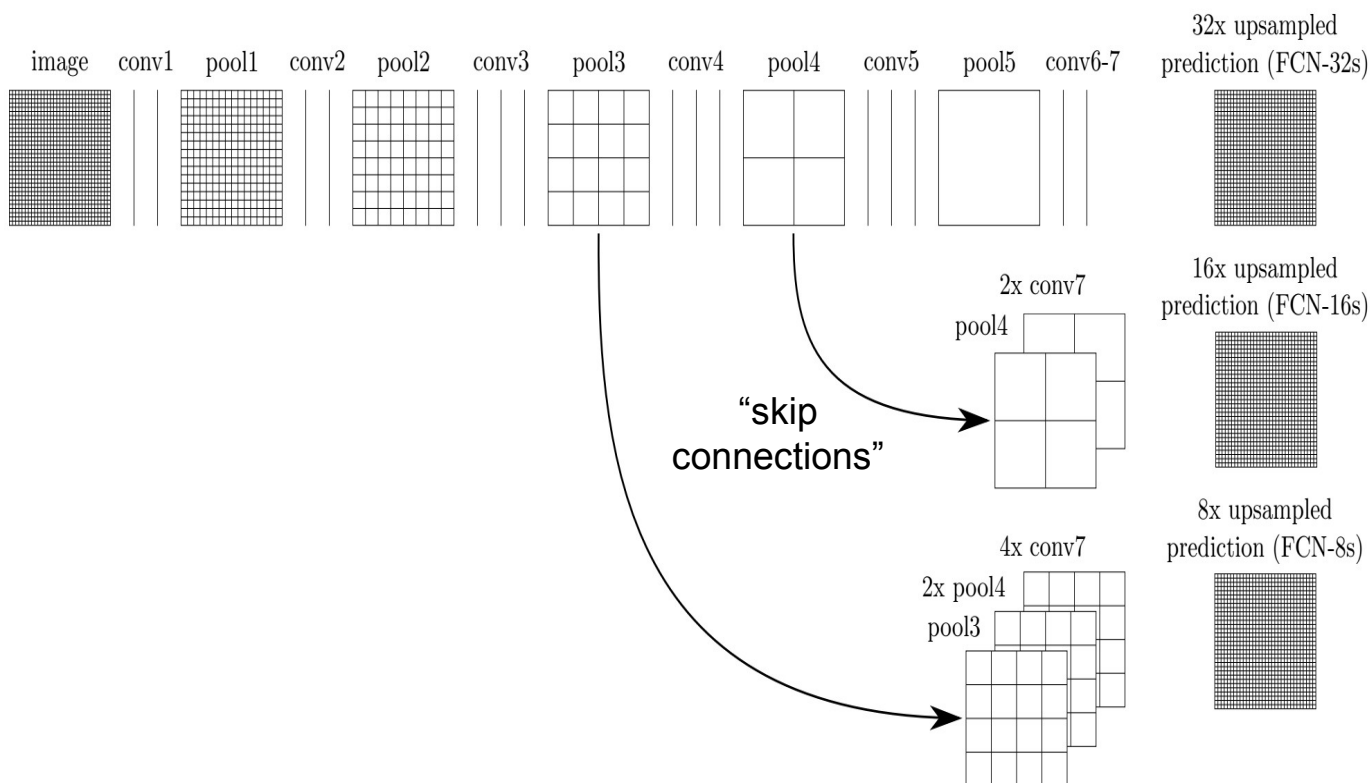
Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015

# Semantic Segmentation: Upsampling



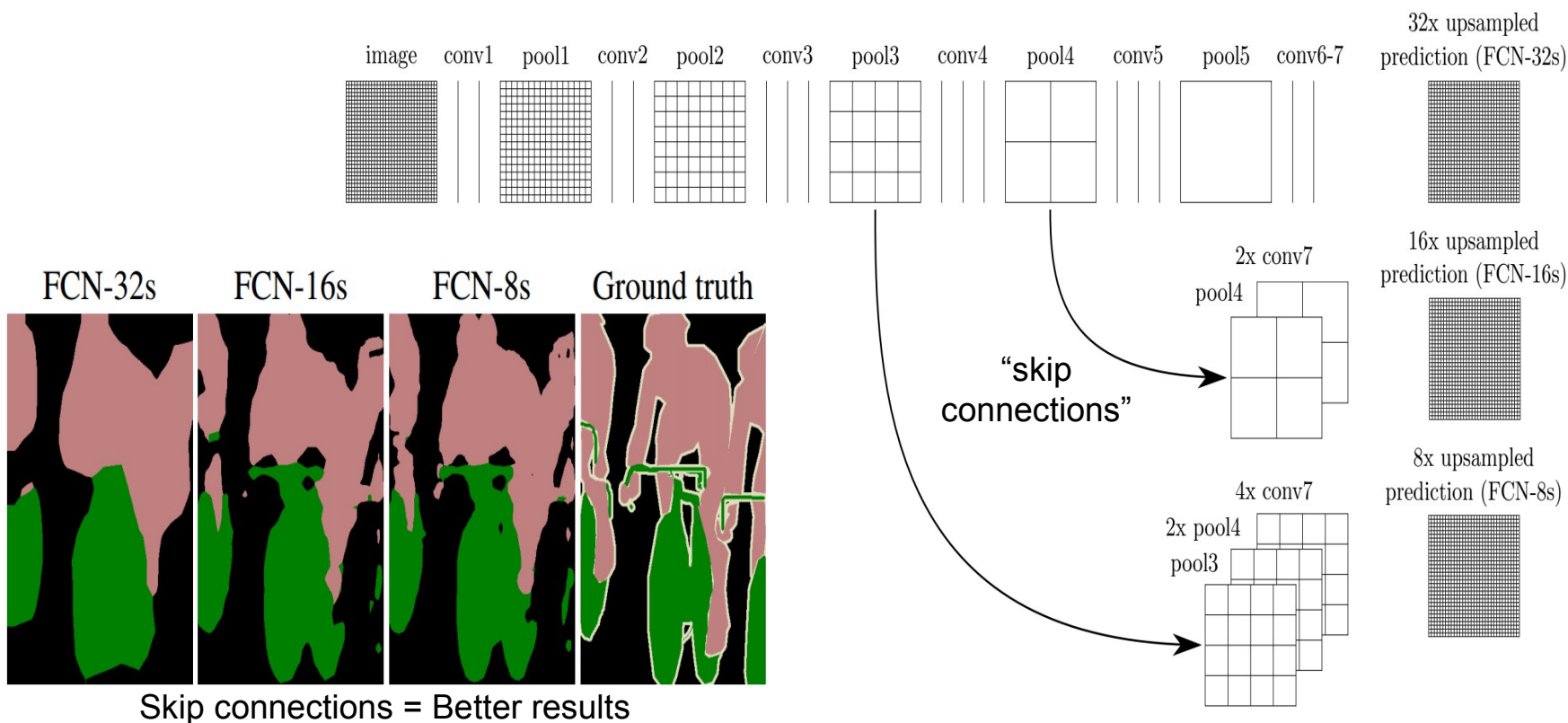
Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015

# Semantic Segmentation: Upsampling



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

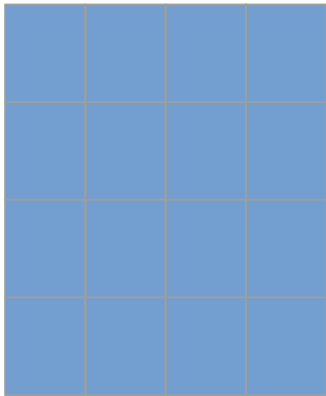
# Semantic Segmentation: Upsampling



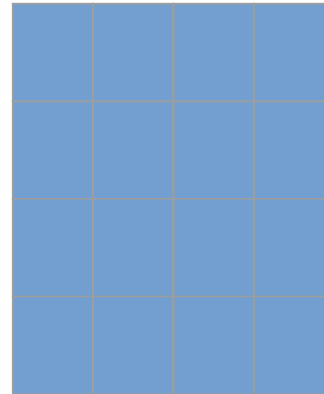
Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015

# Learnable Upsampling: “Deconvolution”

Typical 3 x 3 convolution, stride 1 pad 1



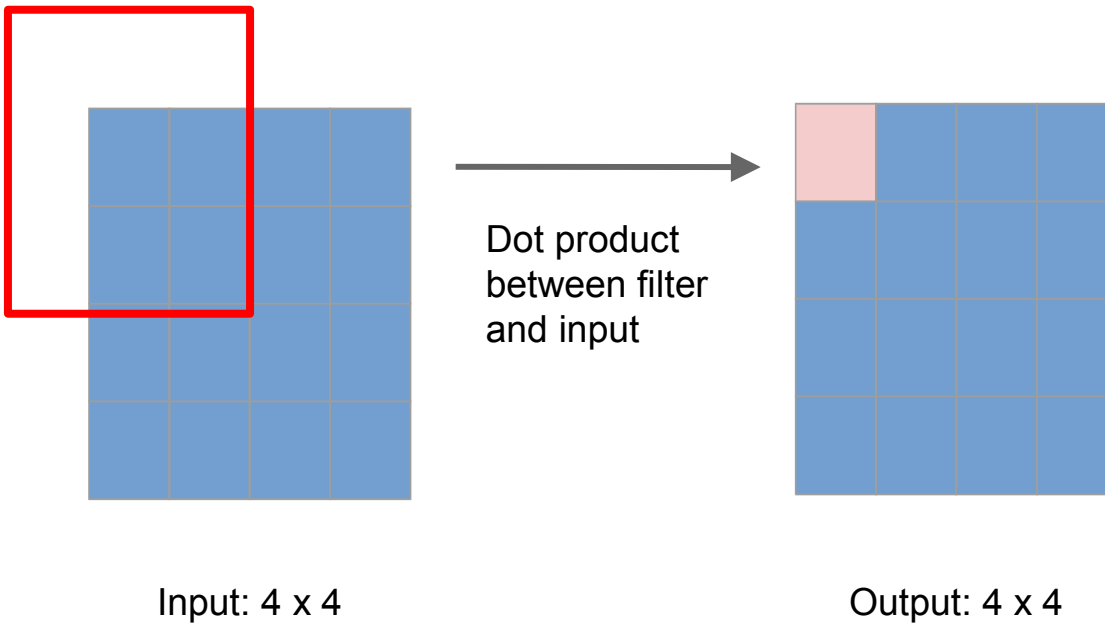
Input: 4 x 4



Output: 4 x 4

# Learnable Upsampling: “Deconvolution”

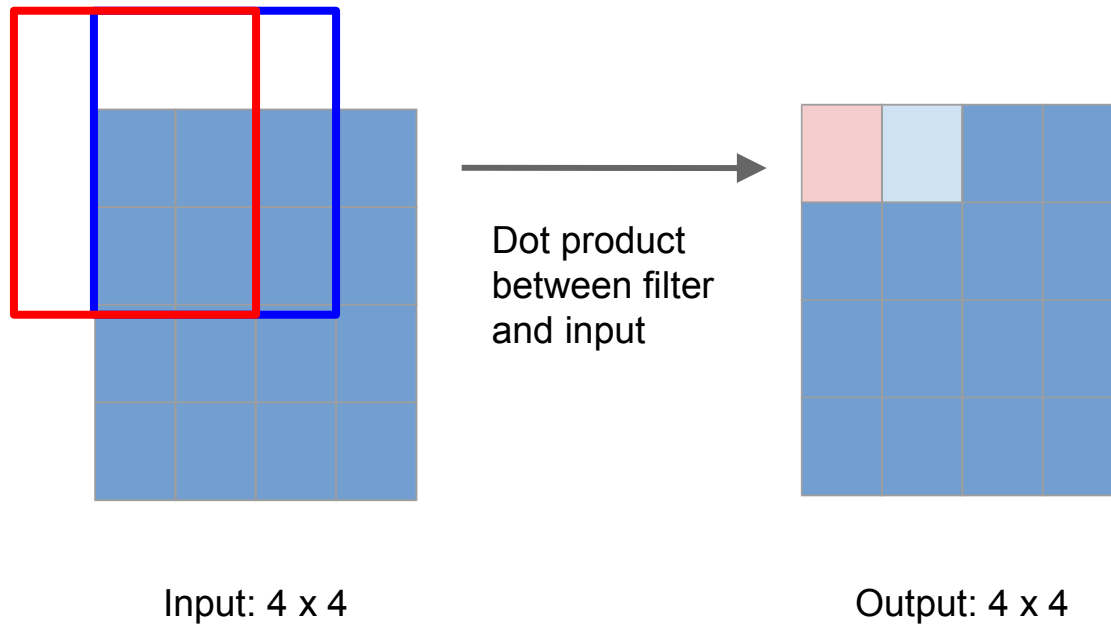
Typical 3 x 3 convolution, stride 1 pad 1





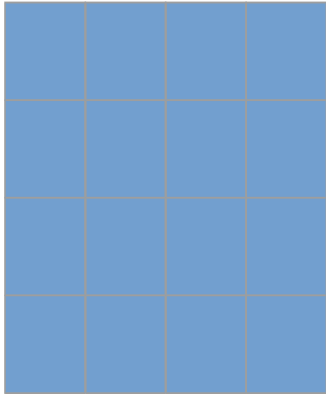
# Learnable Upsampling: “Deconvolution”

Typical 3 x 3 convolution, stride 1 pad 1

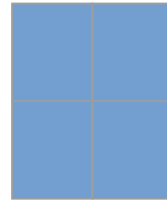


# Learnable Upsampling: “Deconvolution”

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



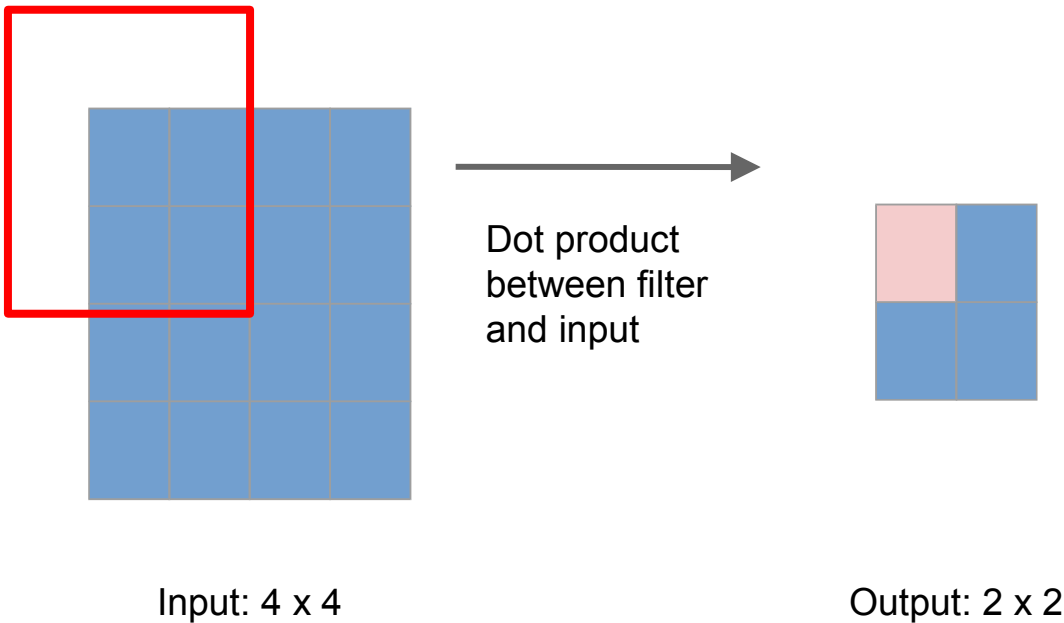
Input: 4 x 4



Output: 2 x 2

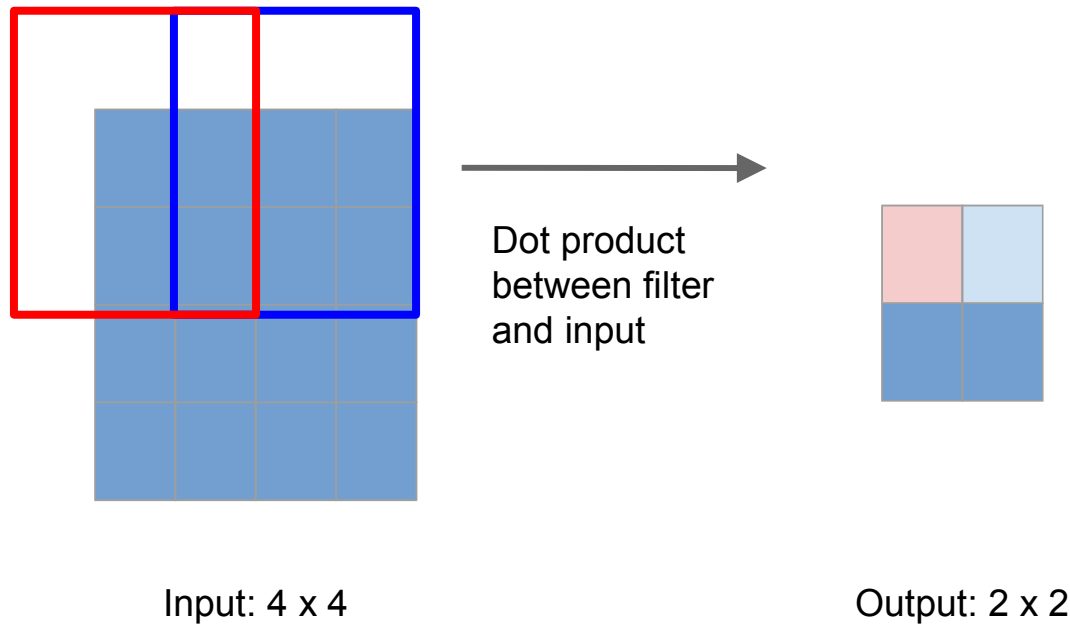
# Learnable Upsampling: “Deconvolution”

Typical 3 x 3 convolution, stride 2 pad 1



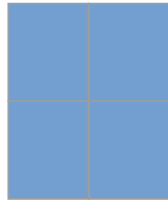
# Learnable Upsampling: “Deconvolution”

Typical 3 x 3 convolution, stride 2 pad 1

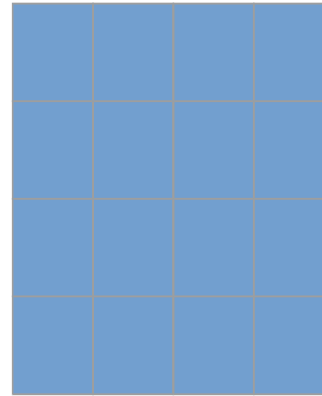


# Learnable Upsampling: “Deconvolution”

3 x 3 “deconvolution”, stride 2 pad 1



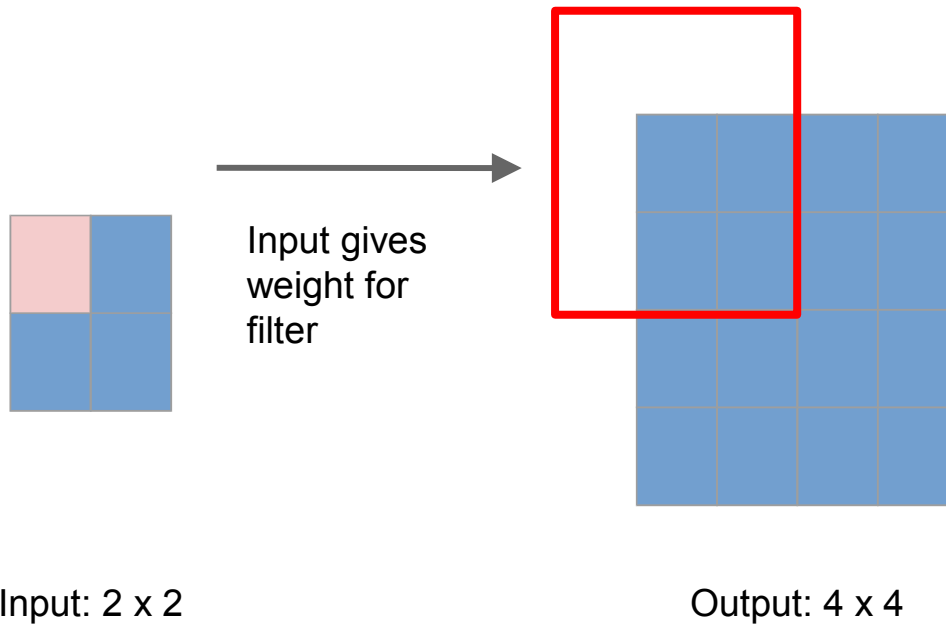
Input: 2 x 2



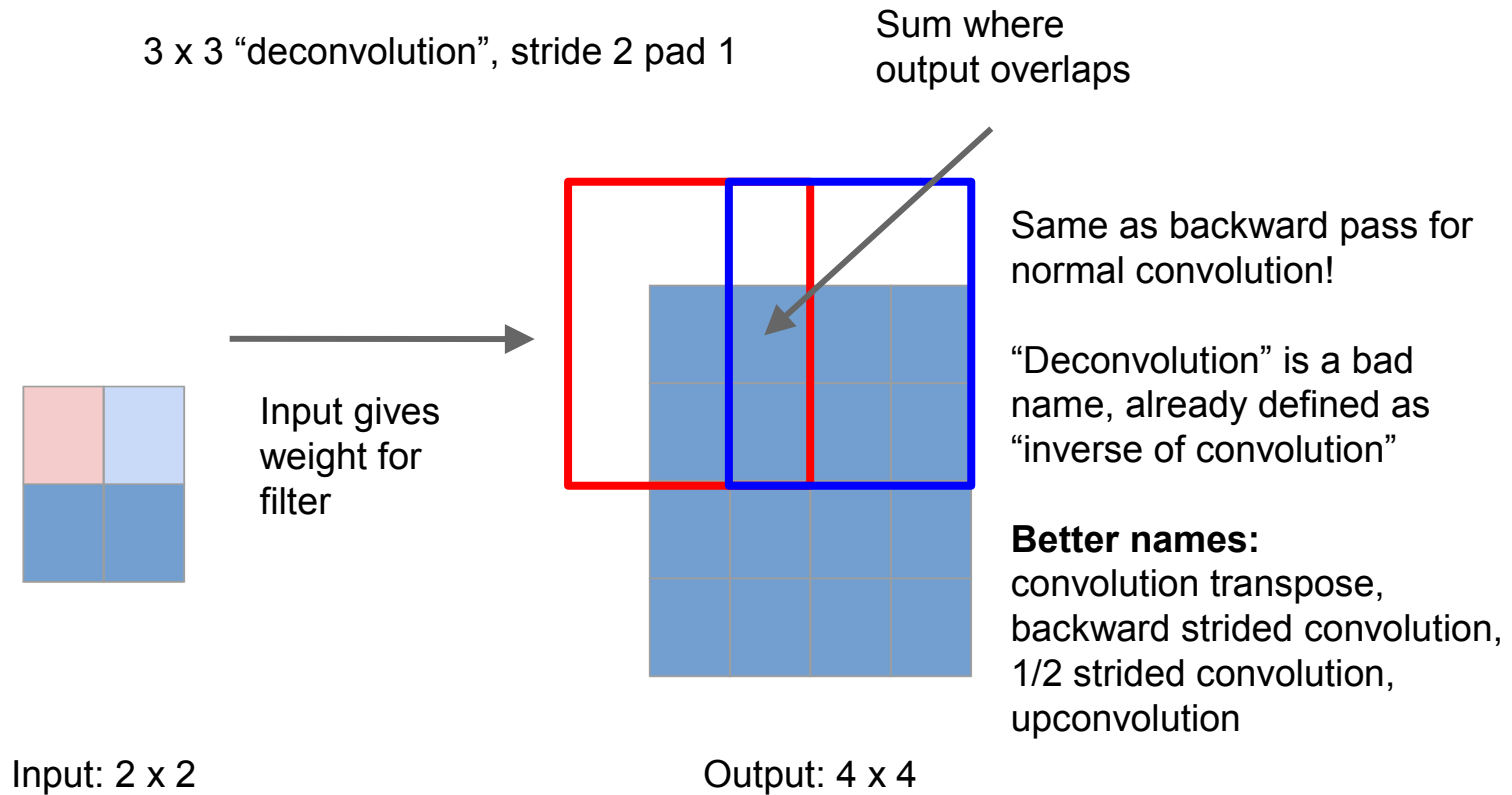
Output: 4 x 4

# Learnable Upsampling: “Deconvolution”

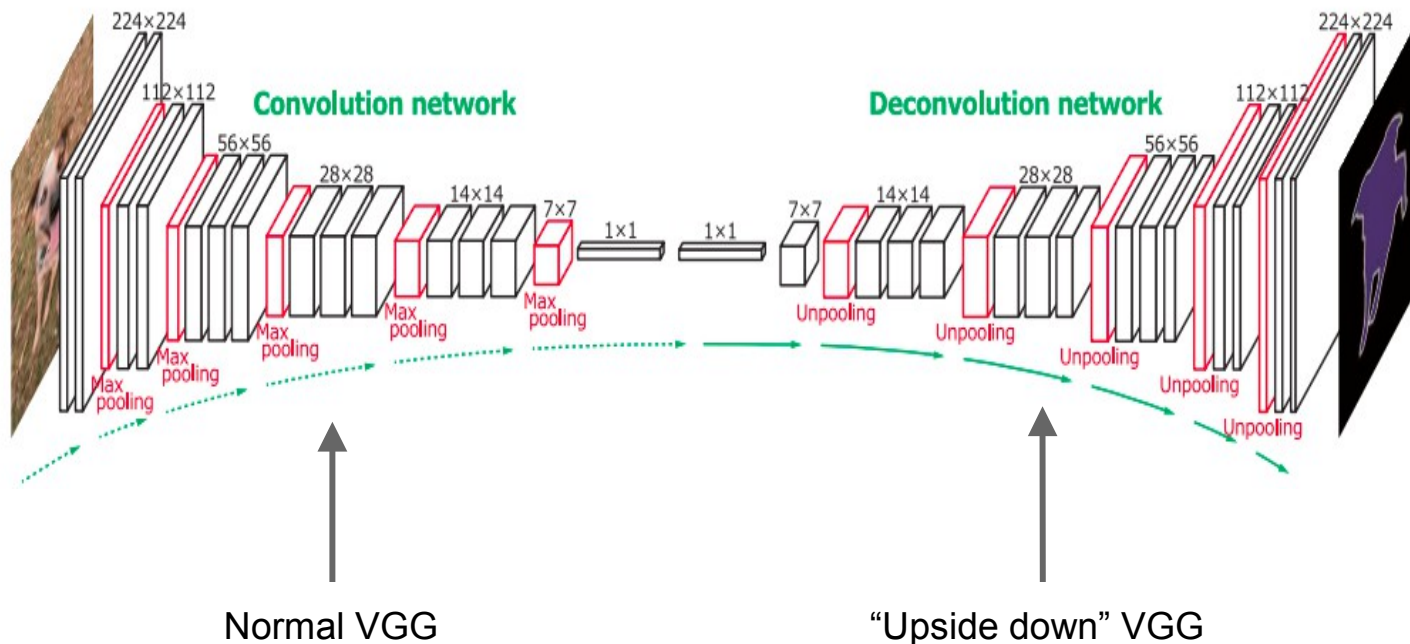
3 x 3 “deconvolution”, stride 2 pad 1



# Learnable Upsampling: “Deconvolution”



# Semantic Segmentation: Upsampling



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

6 days of training on Titan X...



# Instance Segmentation

Detect instances,  
give category, label  
pixels

“simultaneous  
detection and  
segmentation” (SDS)

Lots of recent work  
(MS-COCO)

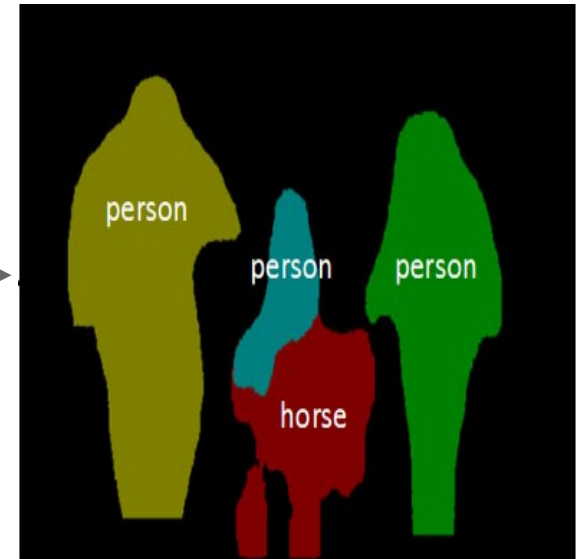


Figure credit: Dai et al, “Instance-aware Semantic Segmentation via Multi-task Network Cascades”, arXiv 2015

# Instance Segmentation

Similar to R-CNN, but  
with segments

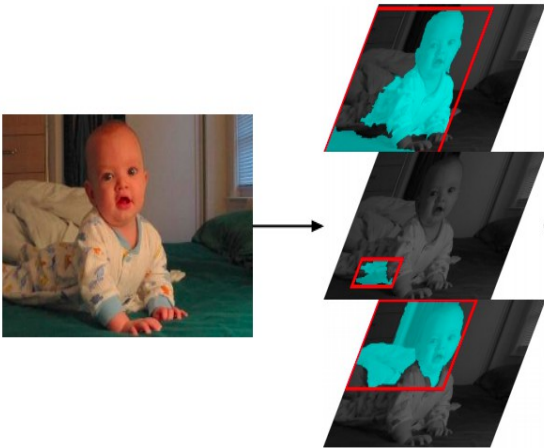


Hariharan et al, "Simultaneous Detection and Segmentation", ECCV 2014

# Instance Segmentation

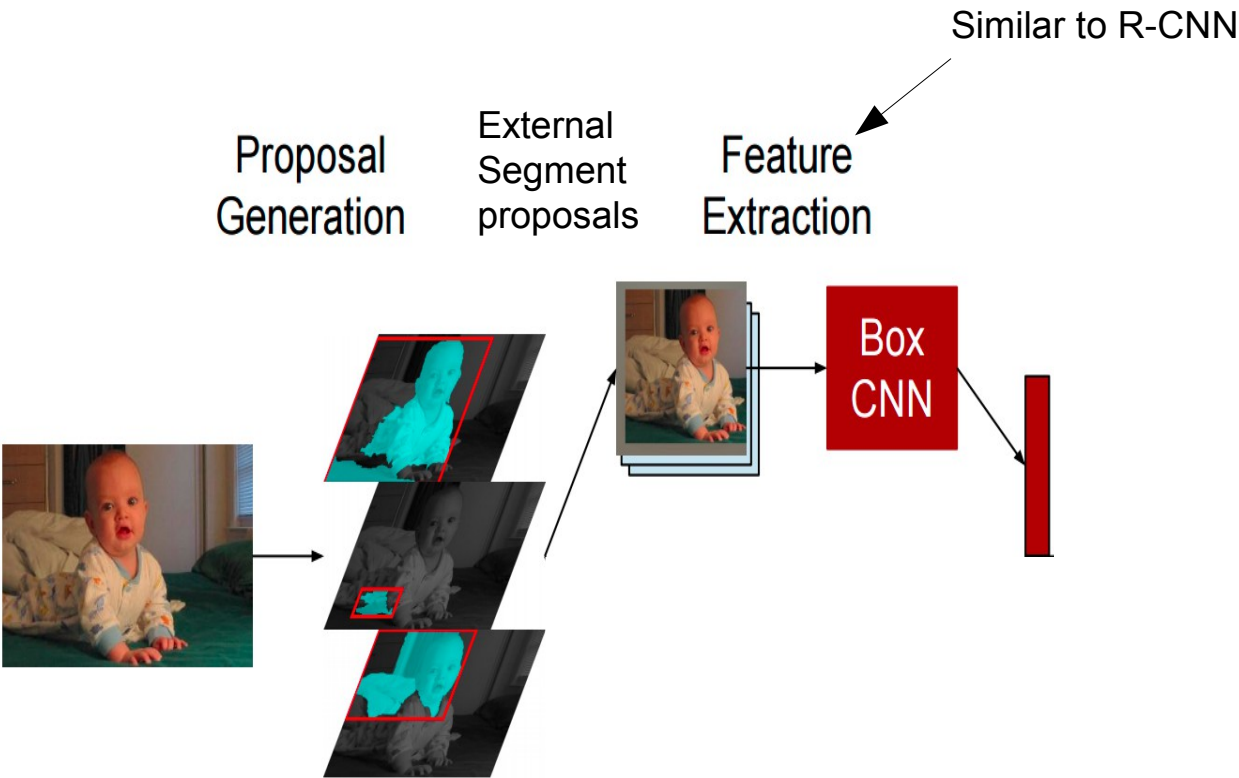
Similar to R-CNN, but  
with segments

Proposal  
Generation      External  
Segment  
proposals



Hariharan et al, "Simultaneous Detection and Segmentation", ECCV 2014

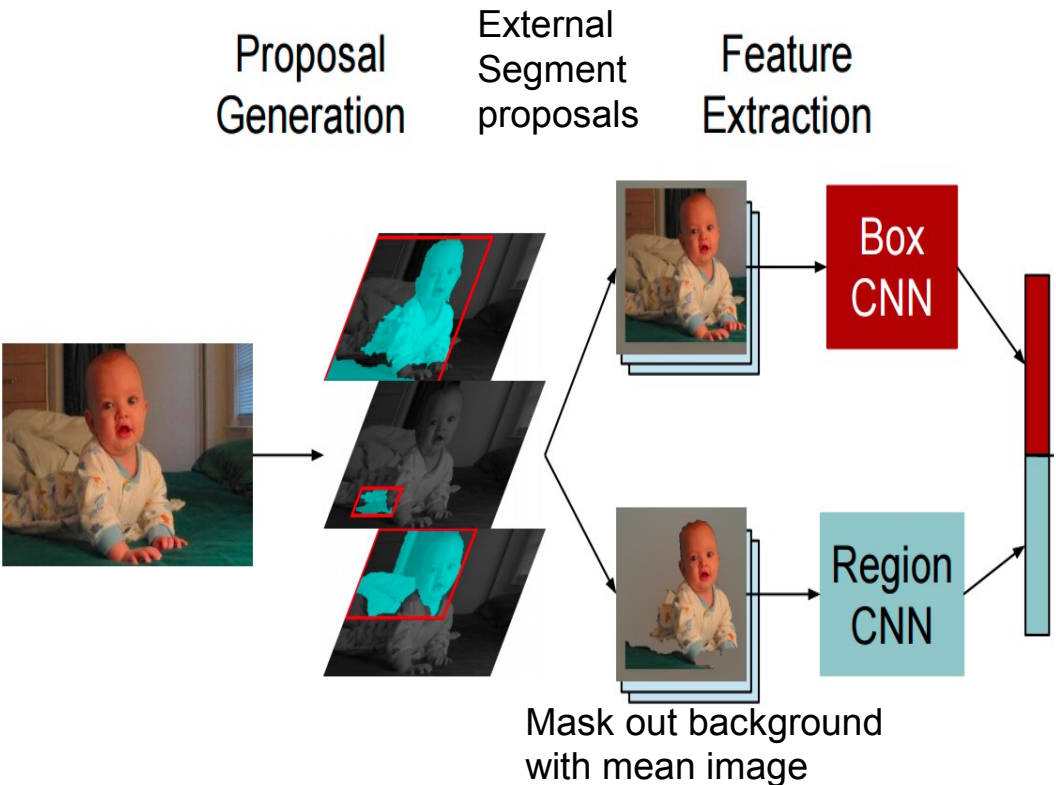
# Instance Segmentation



Hariharan et al, "Simultaneous Detection and Segmentation", ECCV 2014

# Instance Segmentation

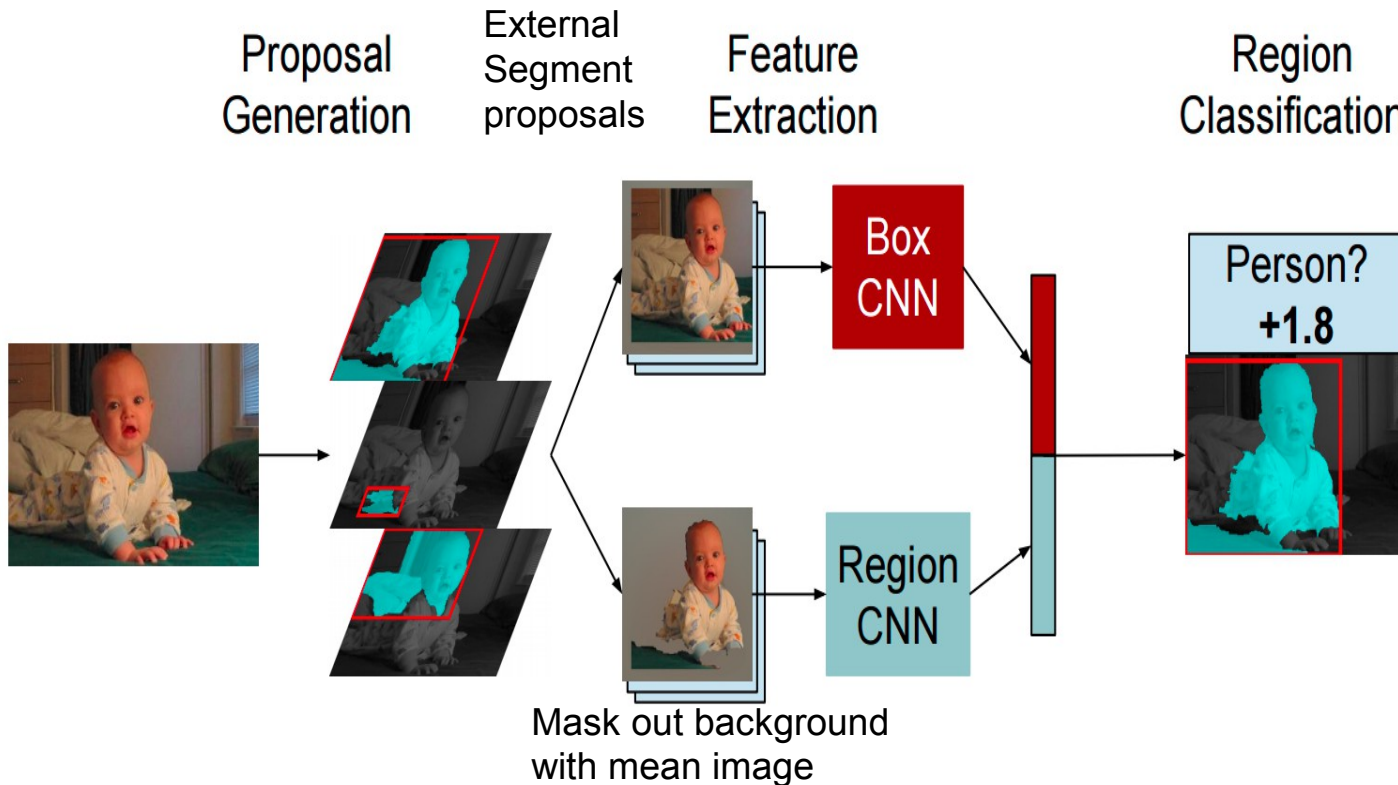
Similar to R-CNN, but with segments



Hariharan et al, "Simultaneous Detection and Segmentation", ECCV 2014

# Instance Segmentation

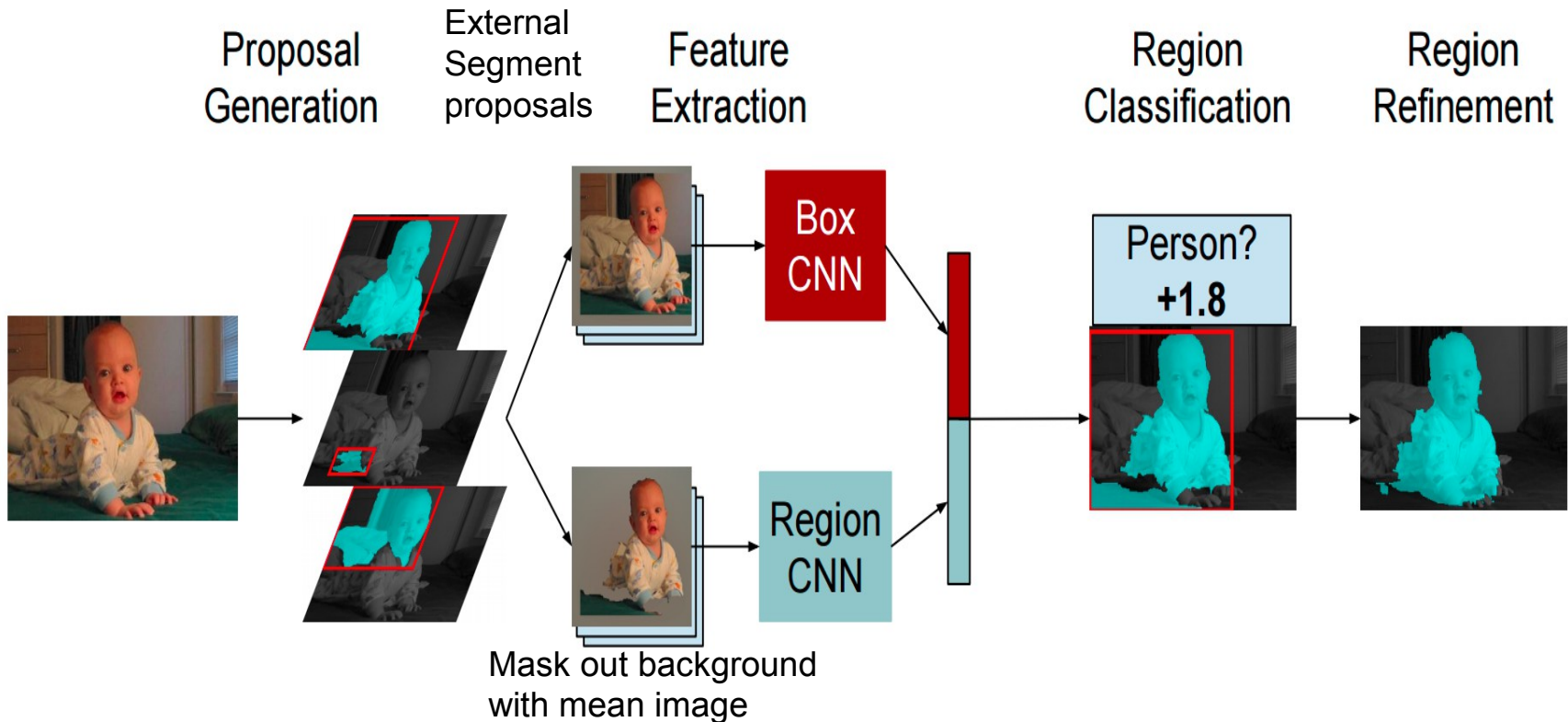
Similar to R-CNN, but  
with segments



Hariharan et al, "Simultaneous Detection and Segmentation", ECCV 2014

# Instance Segmentation

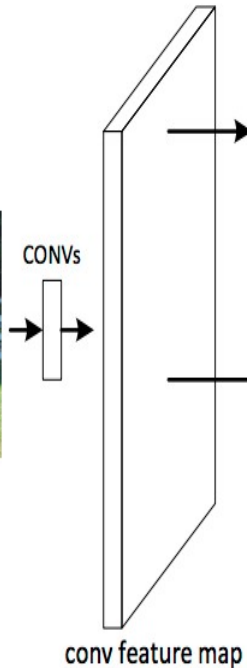
Similar to R-CNN, but  
with segments



Hariharan et al, "Simultaneous Detection and Segmentation", ECCV 2014

# Instance Segmentation: Cascades

Similar to  
Faster R-CNN



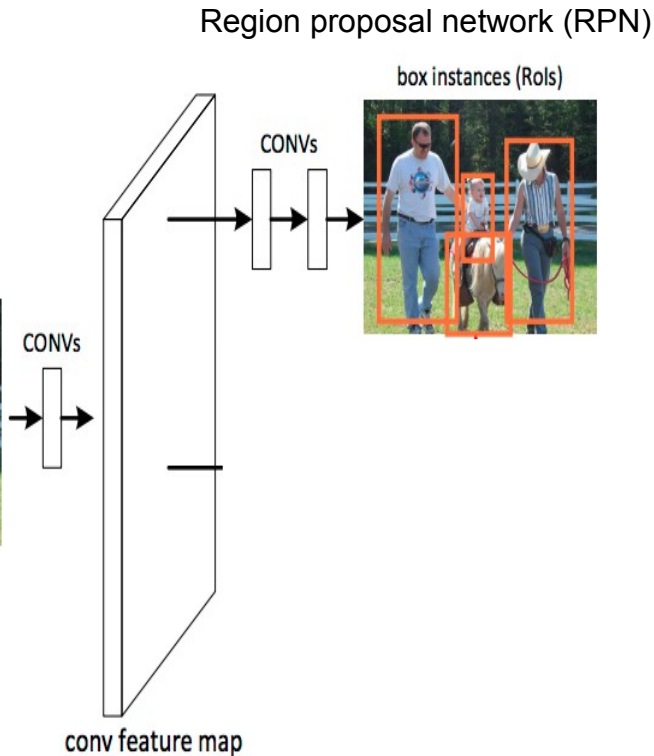
Won COCO 2015  
challenge  
(with ResNet)

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



# Instance Segmentation: Cascades

Similar to  
Faster R-CNN

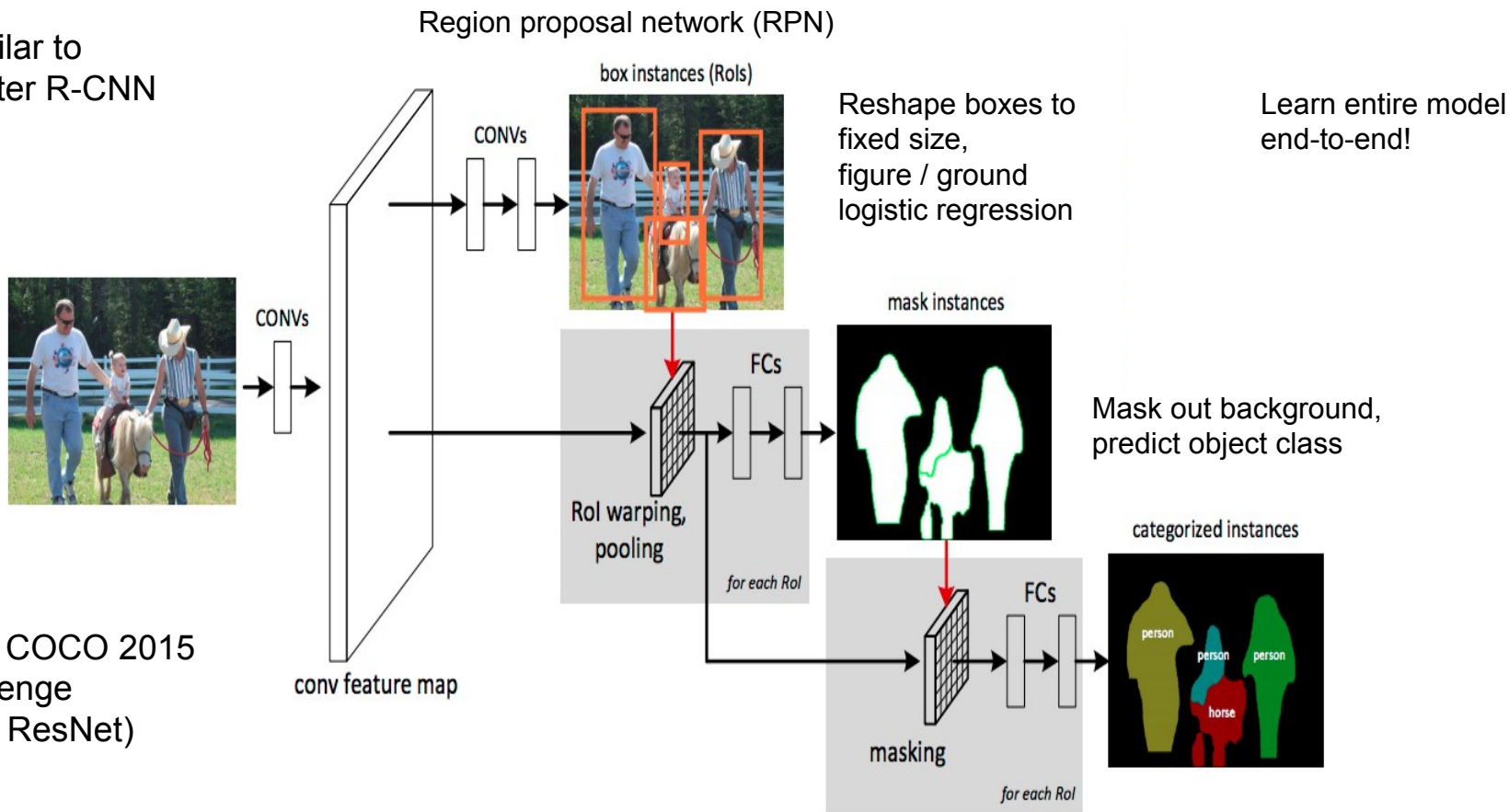


Won COCO 2015  
challenge  
(with ResNet)

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

# Instance Segmentation: Cascades

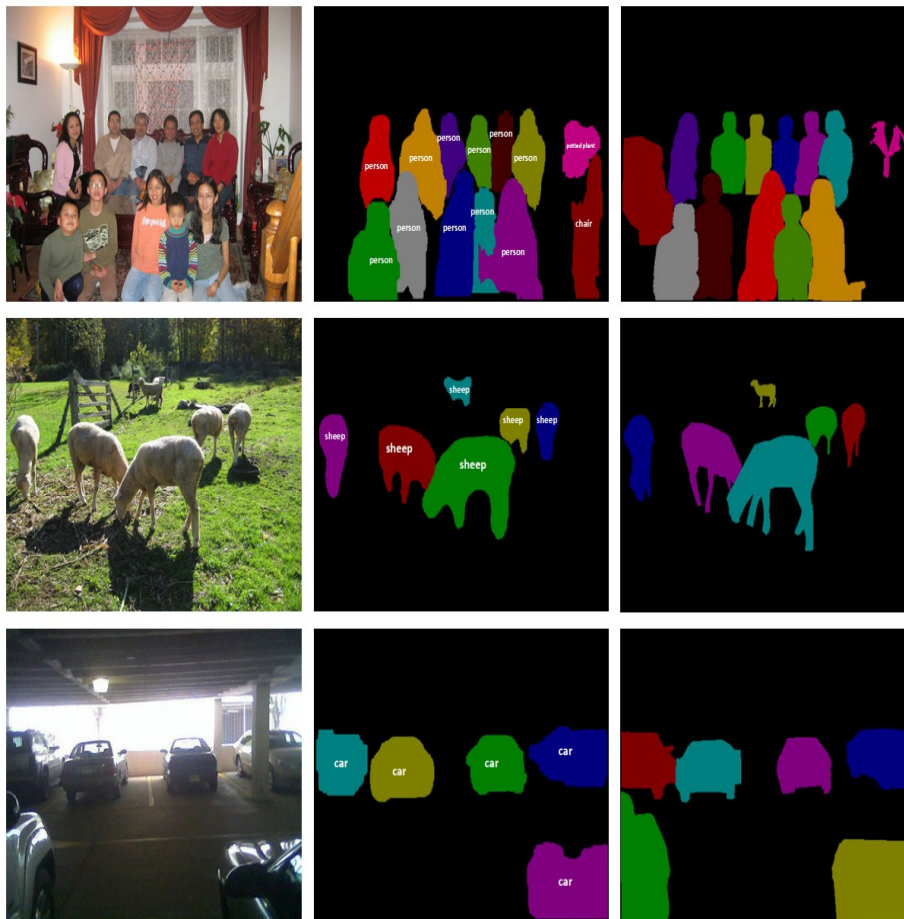
Similar to  
Faster R-CNN



Won COCO 2015  
challenge  
(with ResNet)

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

# Instance Segmentation: Cascades



Predictions

Ground truth

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

# Segmentation Overview

---

## 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

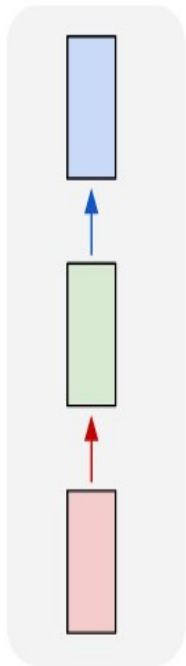
- Similar pipelines to object detection

---

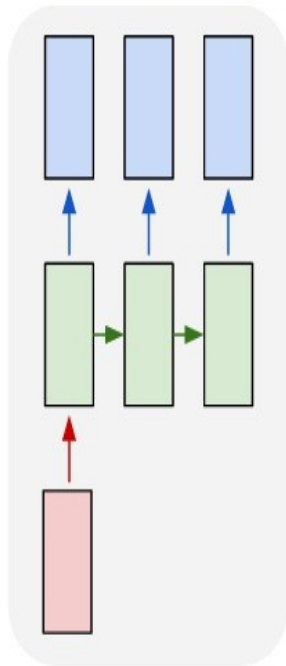
# Quick overview of Other Topics

# Recurrent Neural Networks (RNN)

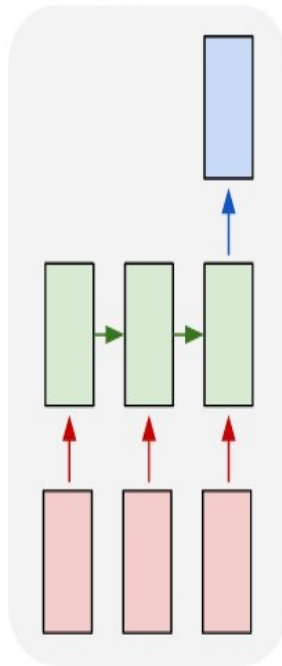
one to one



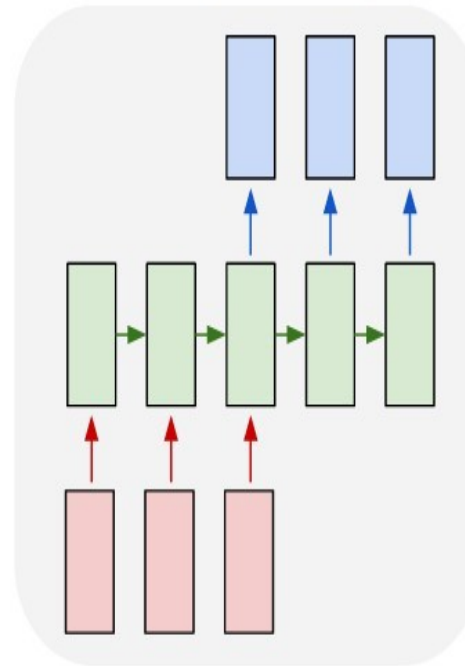
one to many



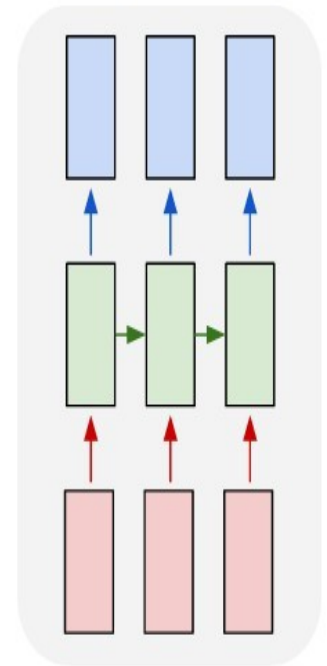
many to one



many to many



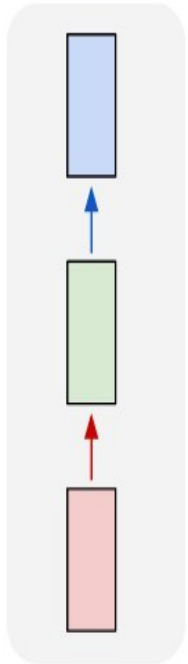
many to many



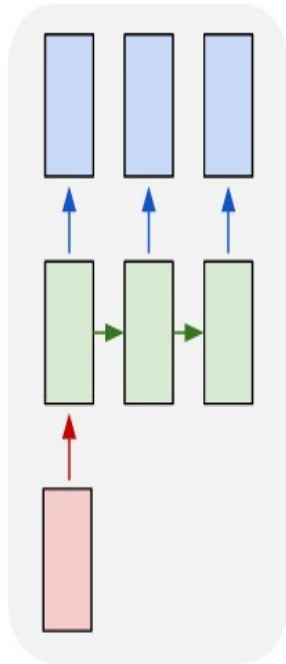
↖ **Vanilla Neural Networks**

# Recurrent Neural Networks (RNN)

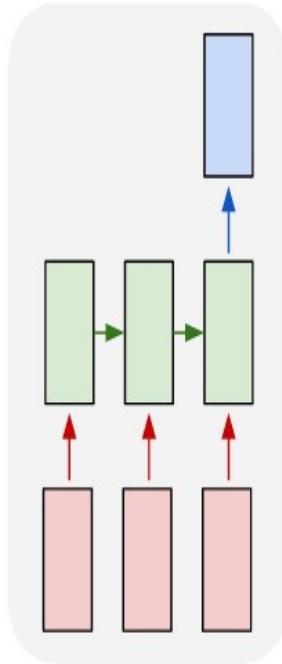
one to one



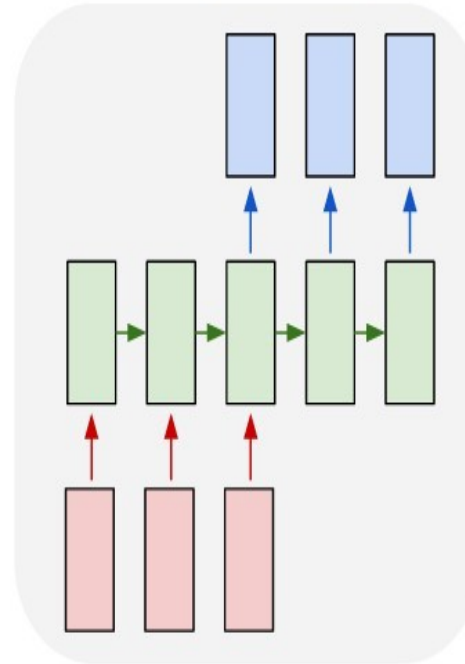
one to many



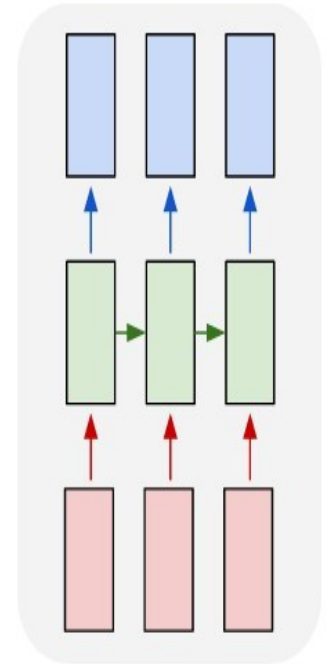
many to one



many to many



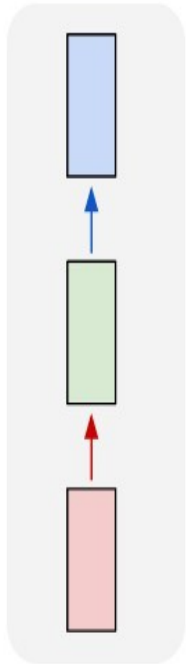
many to many



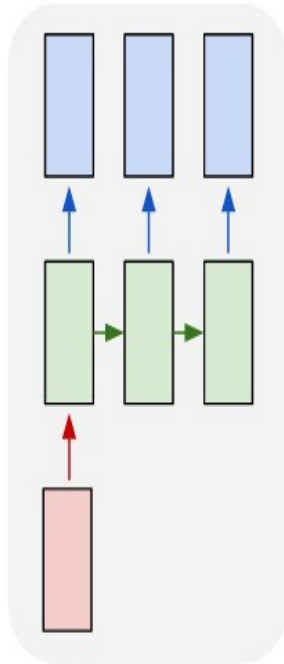
e.g. **Image Captioning**  
image -> sequence of words

# Recurrent Neural Networks (RNN)

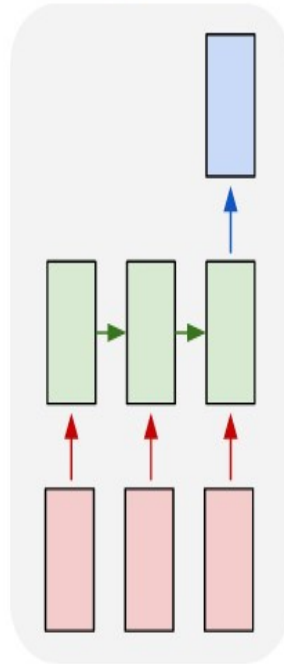
one to one



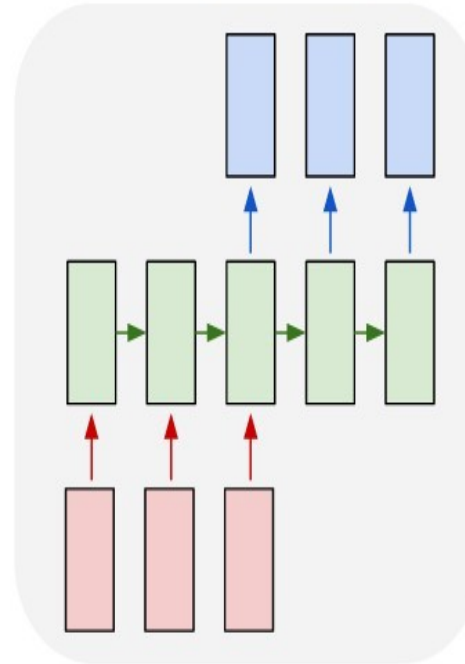
one to many



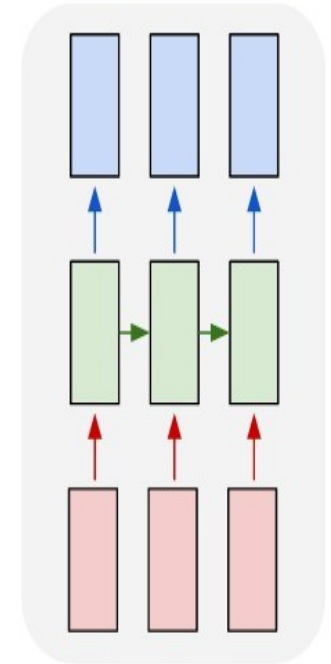
many to one



many to many



many to many

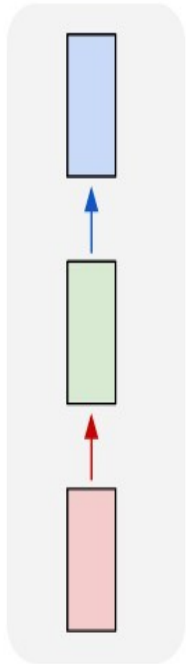


e.g. **Sentiment Classification**  
sequence of words -> sentiment

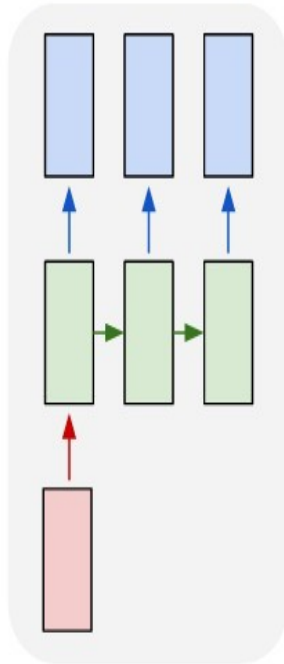


# Recurrent Neural Networks (RNN)

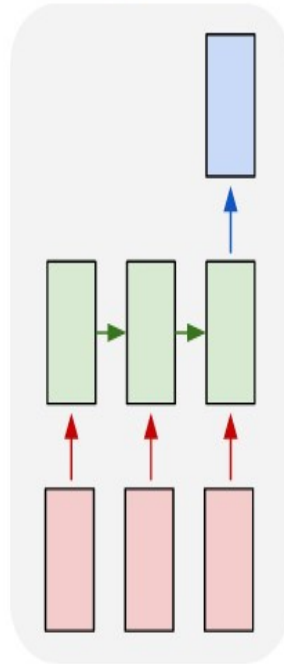
one to one



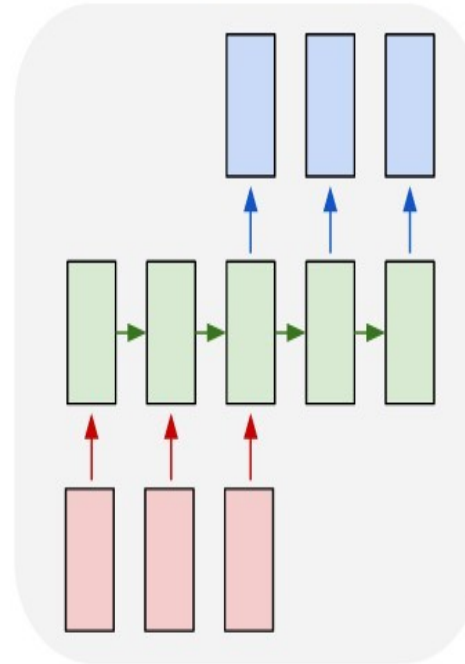
one to many



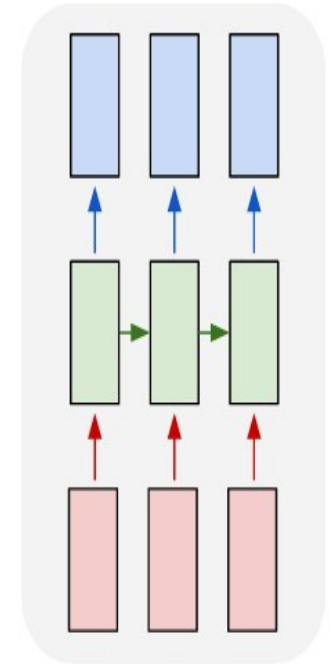
many to one



many to many



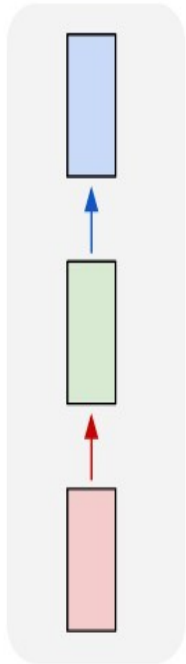
many to many



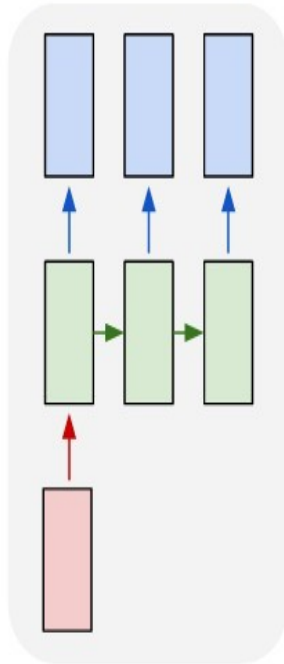
↑ e.g. **Machine Translation**  
seq of words → seq of words

# Recurrent Neural Networks (RNN)

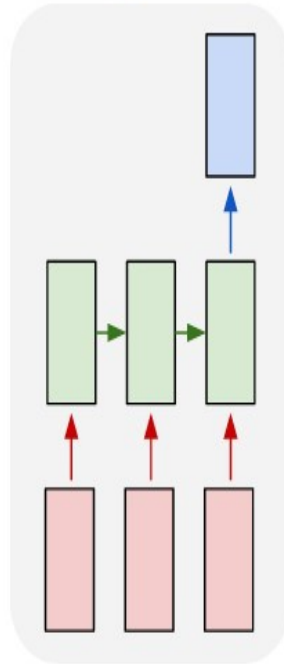
one to one



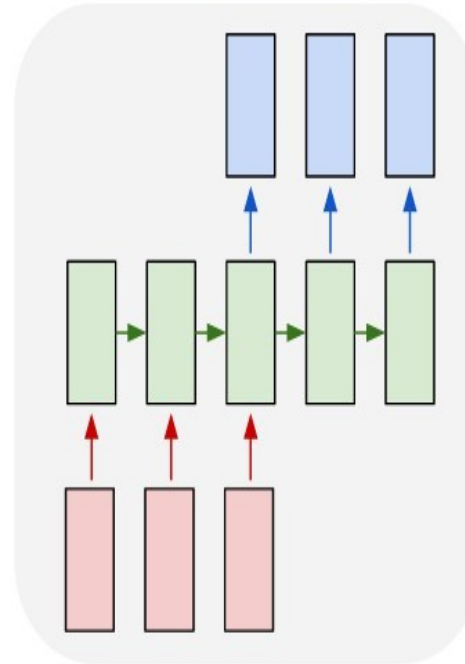
one to many



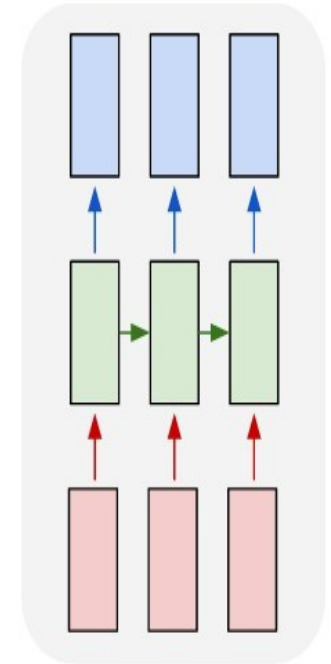
many to one



many to many



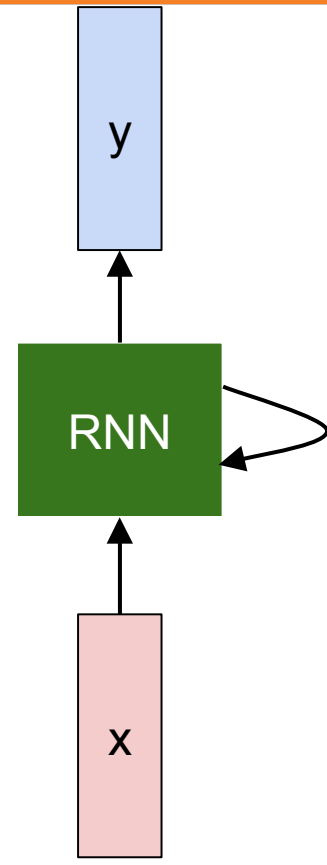
many to many



e.g. **Video classification on frame level**



<p class="clear">Products: Laser-Printers The fundamental everyday requirement for mono and colour laser printing throughout today's offices is perfectly met with the extensive Epson laser printer range. The latest AcuLaser printer range offers users exceptionally and affordable colour laser printing for far too long. The traditionally high costs and poor speeds of colour lasers has left many offices looking a bit, well, grey. But not any more with the Epson AcuLaser C1900. Epson brings both colour and monochrome laser printing together at a black and white price. more Where to Buy Support Epson AcuLaser C3000 The fastest colour laser printer in its class. The perfect printer for small businesses and work groups, the Epson AcuLaser C3000 prints high volumes in black and white and vibrant colour, at high speed and with low running costs. more Where to Buy High quality resolution: 2400dpi R77 Large paper capacity: 600 sheets, expandable up to 1,600 sheets Compatible Windows and Mac High speed USB and EpsonNet (10/100 Base Tx Ethernet interfaces as standard) Epson AcuLaser Resolution Improvement Technology EpsonNet 10/100 Base Tx Ethernet standard with Epson AcuLaser C3000N model only AcuLaser C3000 64MB Memory, 100 sheet MP Tray, 500 sheet cassette, Duplex printing as standard AcuLaser C3000N 64MB Memory, 100 sheet MP Tray, 500 sheet cassette, Duplex printing, 10/100BaseTX Ethernet Interface Networked compact colour laser printer for professional enterprises. Businesses have been denied simple and affordable colour laser printing for far too long. The traditionally high costs and poor speeds of colour lasers has left many offices looking a bit, well, grey. But not any more with the Epson AcuLaser C1900. Epson brings both colour and monochrome laser printing together at a black and white price. Key Features cost effective mono printing for day to day business needs and vivid versatile colour when required. search Search Epson UK Epson AcuLaser C500 Outstanding professional colour printing for business Add colour to your business with the Epson AcuLaser C500 from Epson. Its perfect for the smaller workgroup, being a compact and cost effective laser printing workhorse that offers amazing colour output as well as high performance black and white production. more Where to Buy Support As cost efficient to run as a mono-only laser printer Paper capacity of 700 sheets from two media sources Easy to operate with advanced printer driver Memory expandable from 320 to 1024Kb Pre-configured models available with Wireless G02 11b, Adobe® PostScript® Level 3™ and two-sided printing The AcuLaser C1900 is available in 5 configurations - AcuLaser C1900S with 320MB, 200 Sheet MP Tray, 10/100BaseTX Networking - AcuLaser C1900 with 32MB, 200 Sheet MP Tray, 500 Sheet Cassette, 10/100BaseTX Networking Support Epson AcuLaser C4100 High performance colour lasers for all your business printing needs The Epson AcuLaser C4100 provides businesses with a high performance colour and monochrome printing solution. It adds crucial colour to your business, while producing high quality monochrome output at lower costs than many monochrome-only printers, and is just as easy to operate. So now there's no reason to buy two printers, because perfect monochrome and colour solutions are available in one. more Where to Buy Support Epson AcuLaser C8600 Professional high performance A3W colour laser printer Epson AcuLaser C8600 is the perfect professional printing solution for users who require exceptional quality colour and mono output on a range of media formats from C5 up to A3W in size. The Epson AcuLaser C8600 is able to achieve superb print quality by utilising a combination of Epson's exclusive AcuLaser Color Laser Technologies. more Where to Buy Support - AcuLaser C1600PS with A4x2x2 PostScript® 3™, 96MB, 200 Sheet MP Tray, 500 Sheet Cassette, 10/100BaseTX Networking - AcuLaser C1900N with Duplex unit (two sided printing) 96MB, 200 Sheet MP Tray, 500 Sheet Cassette, 10/100BaseTX Networking - AcuLaser C1900 WiFi with 32MB, 200 Sheet MP Tray, 500 Sheet Cassette, Wireless Networking facility Add colour to your business with the Epson AcuLaser C500 from Epson. Its perfect for the smaller workgroup, being a compact and cost effective laser printing workhorse that offers amazing colour output as well as high performance monochrome printing. more Where to Buy Support Epson AcuLaser C4000 High performance colour laser The Epson AcuLaser C4000 provides businesses with high performance colour and monochrome printing solutions. more Where to Buy Support Epson AcuLaser C9100 High speed A3 colour laser printer Why have separate black and white and colour printers when you can have the Epson AcuLaser C9100? Epson has taken the lead in laser technology to deliver a complete high-performance solution for all your colour and mono printing needs. Support EPL-6200L High performance A4 mono laser professional printers The Epson EPL-6200 and EPL-6200L are the ideal printing solutions for small to medium workgroups and personal users. They deliver professional performance quickly, easily, reliably and cost-effectively, and are perfect for users who need high levels of laser quality and productivity at a low investment. more performance black and white production. For the first time, you can now bring the power of high quality colour to your documents without suffering the high costs or low speeds traditionally associated with colour



## Character RNN during training

tyntd-iafhatawiaoihrdemot lytdws e ,tfti, astai f ogoh eoase rrranbyne 'nhtnee e  
plia tklrqd t o idoe ns,smtt h ne etie h,hregtrs nigtike,aoaenns lng

train more

"Tmont thithey" fomesscerliund  
Keushey. Thom here  
sheulke, anmerenith ol sivh I lalterthend Bleipile shuwv fil on aseterlome  
coaniogennc Phe lism thond hon at. MeiDimorotion in ther thize."

train more

Aftair fall unsuch that the hall for Prince Velzonski's that me of  
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort  
how, and Gogition is so overelical and offer.

train more

"Why do what that day," replied Natasha, and wishing to himself the fact the  
princess, Princess Mary was easier, fed in had oftened him.  
Pierre aking his soul came to the packs and drove up his father-in-law women.

PANDARUS:

Alas, I think he shall be come approached and the day  
When little strain would be attain'd into being never fed,  
And who is but a chain and subjects of his death,  
I should not sleep.

Second Senator:

They are away this miseries, produced upon my soul,  
Breaking and strongly should be buried, when I perish  
The earth and thoughts of many states.

DUKE VINCENTIO:

Well, your wit is in the care of side and that.

Second Lord:

They would be ruled after this chamber, and  
my fair news begun out of the fact, to be conveyed,  
Whose noble souls I'll have the heart of the wars.

Clown:

Come, sir, I will make did behold your worship.

VIOLA:

I'll drink it.

VIOLA:

Why, Salisbury must find his flesh and thought  
That which I am not apt, not a man and in fire,  
To show the reining of the raven and the wars  
To grace my hand reproach within, and not a fair are hand,  
That Caesar and my goodly father's world;  
When I was heaven of presence and our fleets,  
We spare with hours, but cut thy council I am great,  
Murdered and by thy master's ready there  
My power to give thee but so much as hell:  
Some service in the noble bondman here,  
Would show him to her wine.

KING LEAR:

O, if you were a feeble sight, the courtesy of your law,  
Your sight and several breath, will wear the gods  
With his heads, and my hands are wonder'd at the deeds,  
So drop upon your lordship's head, and your opinion  
Shall be against your honour.

```

static void do_command(struct seq_file *m, void *v)
{
    int column = 32 << (cmd[2] & 0x80);
    if (state)
        cmd = (int)(int_state ^ (in_8(&ch->ch_flags) & Cmd) ? 2 : 1);
    else
        seq = 1;
    for (i = 0; i < 16; i++) {
        if (k & (1 << i))
            pipe = (in_use & UMXTHREAD_UNCCA) +
                ((count & 0x00000000ffffffff) & 0x0000000f) << 8;
        if (count == 0)
            sub(pid, ppc_md.kexec_handle, 0x20000000);
        pipe_set_bytes(i, 0);
    }
    /* Free our user pages pointer to place camera if all dash */
    subsystem_info = &of_changes[PAGE_SIZE];
    rek_controls(offset, idx, &soffset);
    /* Now we want to deliberately put it to device */
    control_check_polarity(&context, val, 0);
    for (i = 0; i < COUNTER; i++)
        seq_puts(s, "policy ");
}

```

# Generated C code

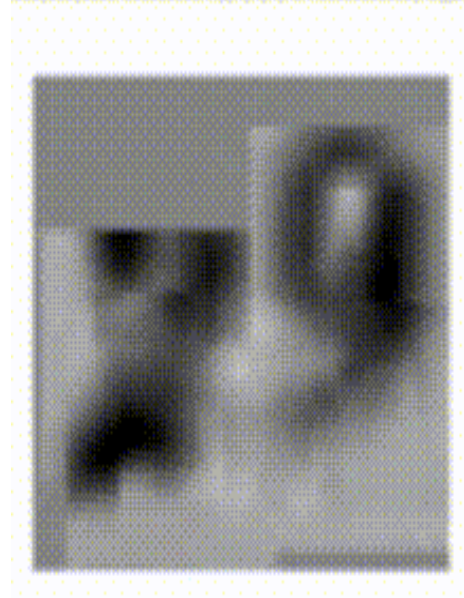
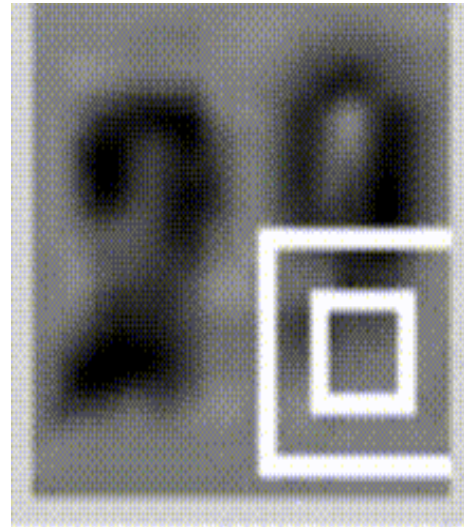
# Searching for interpretable cells

"You mean to imply that I have nothing to eat out of.... On the contrary, I can supply you with everything even if you want to give dinner parties," warmly replied Chichagov, who tried by every word he spoke to prove his own rectitude and therefore imagined Kutuzov to be animated by the same desire.

Kutuzov, shrugging his shoulders, replied with his subtle penetrating smile: "I meant merely to say what I said."

quote detection cell

# Sequential Processing of fixed inputs

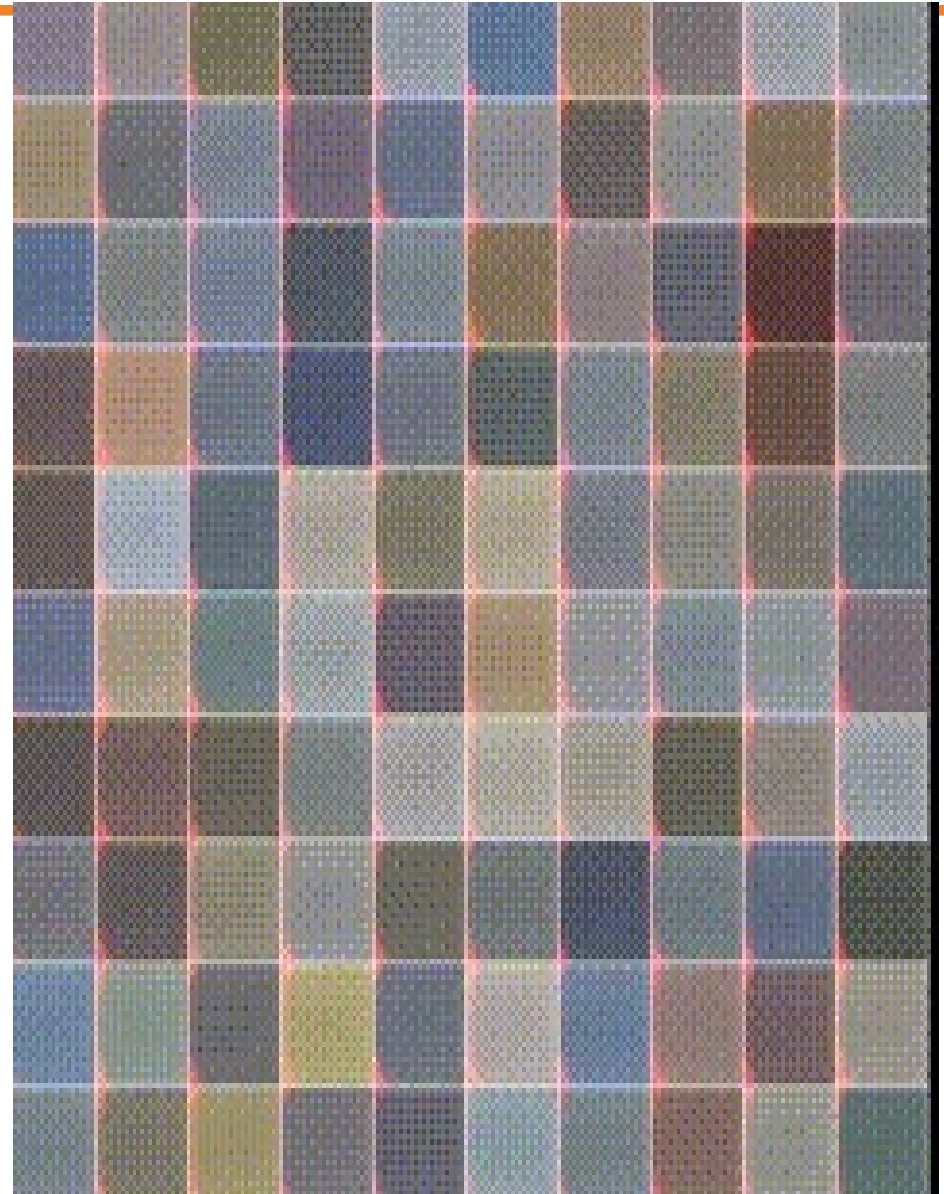


Multiple Object Recognition with  
Visual Attention, Ba et al.

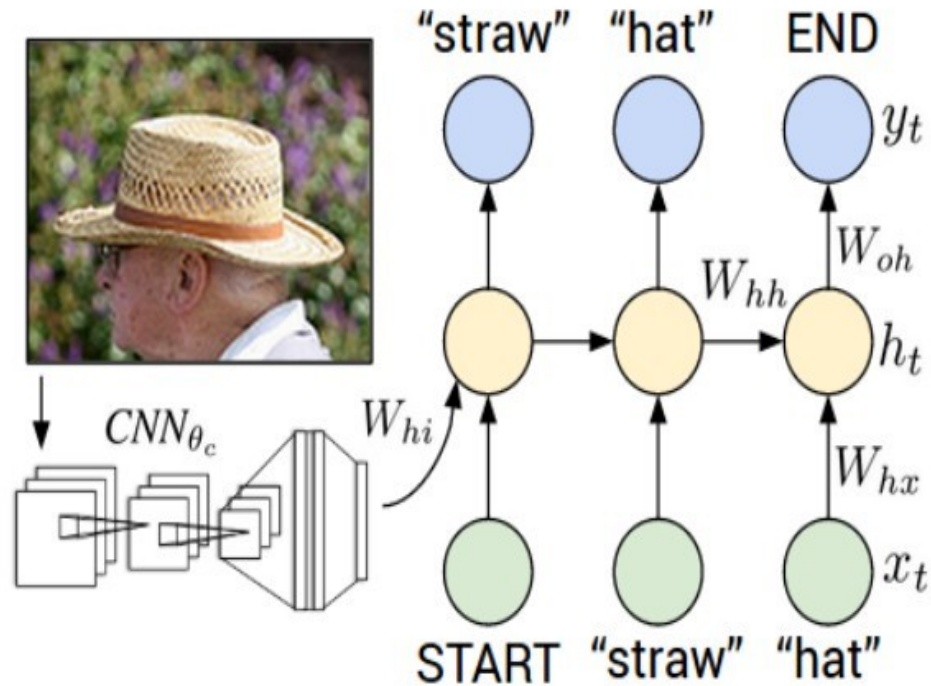


# Sequential Processing of fixed outputs

DRAW: A Recurrent  
Neural Network For  
Image Generation,  
Gregor et al.



# Image Captioning



Explain Images with Multimodal Recurrent Neural Networks, Mao et al.

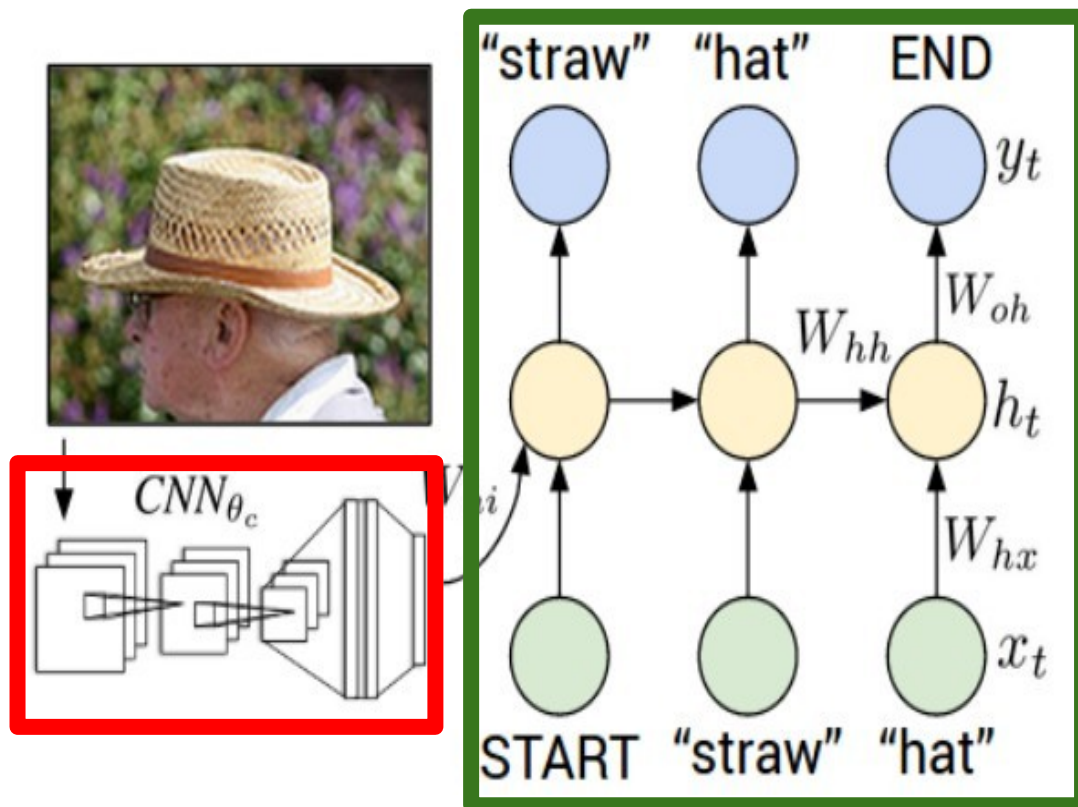
Deep Visual-Semantic Alignments for Generating Image Descriptions, Karpathy and Fei-Fei

Show and Tell: A Neural Image Caption Generator, Vinyals et al.

Long-term Recurrent Convolutional Networks for Visual Recognition and Description, Donahue et al.

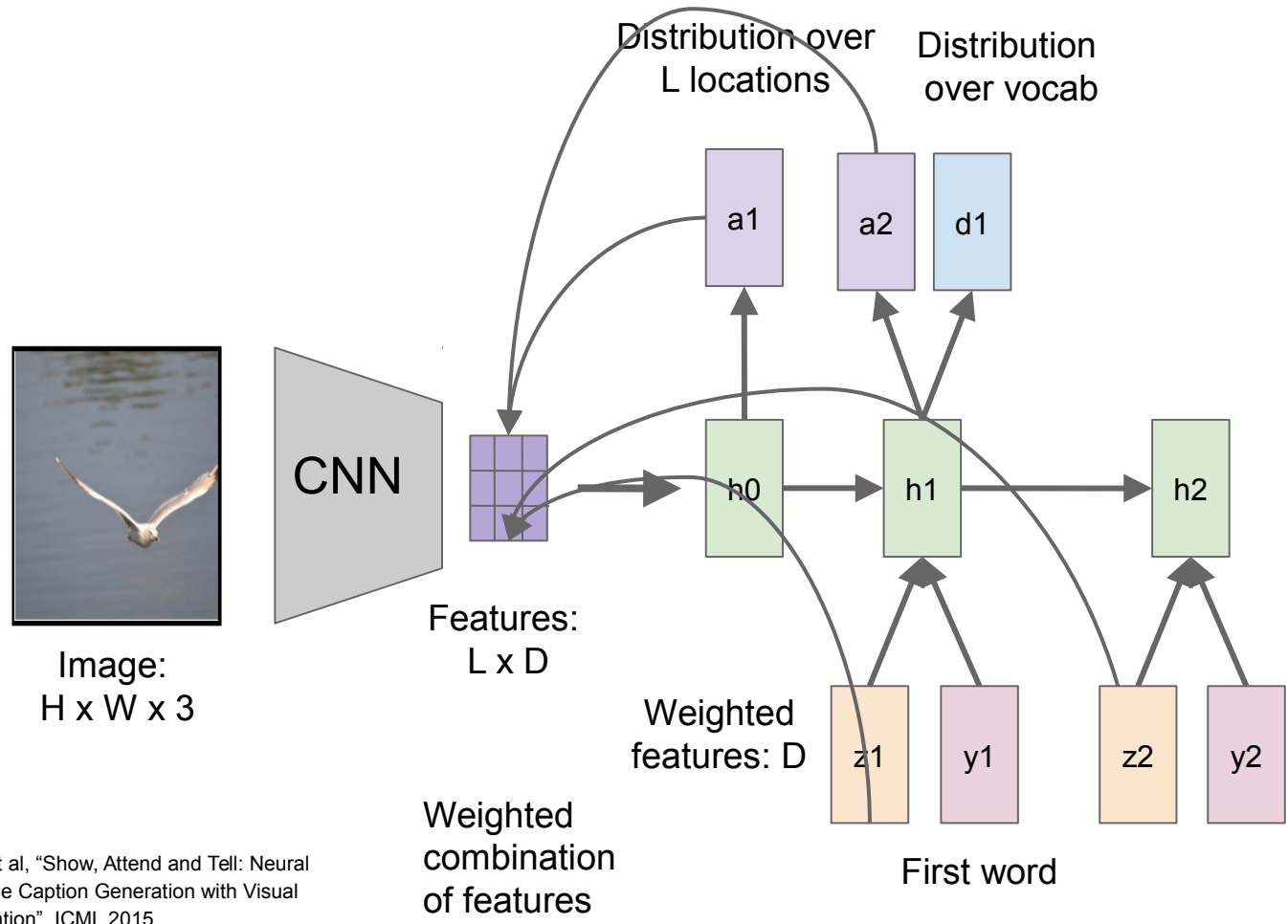
Learning a Recurrent Visual Representation for Image Caption Generation, Chen and Zitnick

# Recurrent Neural Network



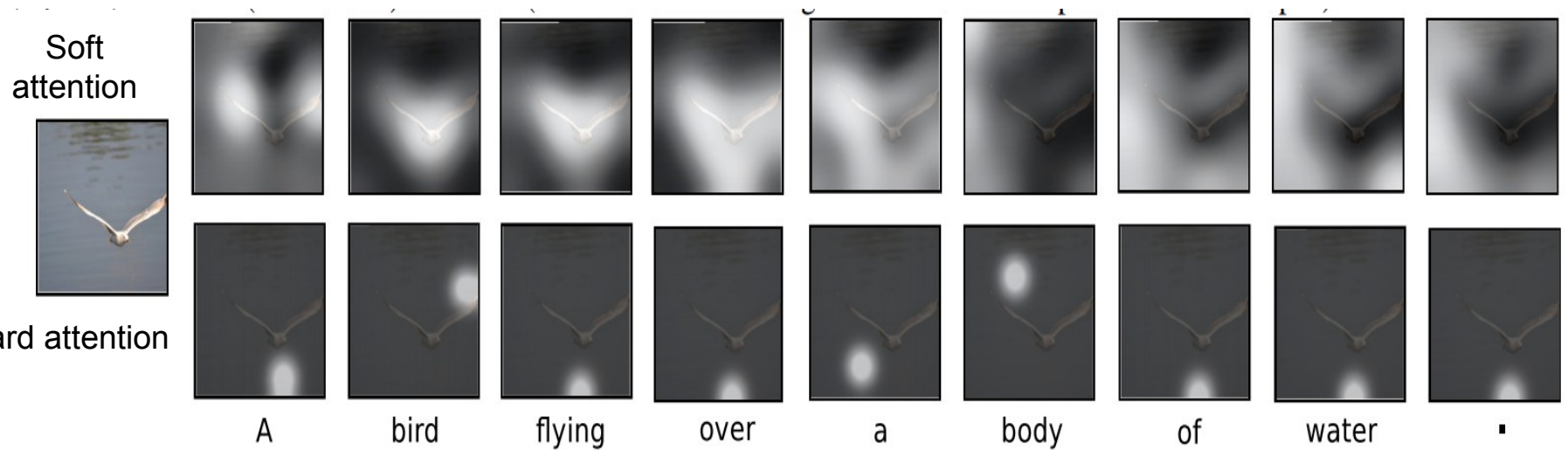
## Convolutional Neural Network

# Soft Attention for Captioning



Xu et al, "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention", ICML 2015

# Soft Attention for Captioning



Xu et al, "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention", ICML 2015

# Soft Attention for Captioning



A woman is throwing a frisbee in a park.



A dog is standing on a hardwood floor.



A stop sign is on a road with a mountain in the background.



A little girl sitting on a bed with a teddy bear.



A group of people sitting on a boat in the water.



A giraffe standing in a forest with trees in the background.

Xu et al, "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention", ICML 2015

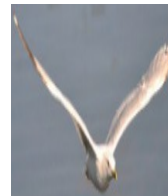
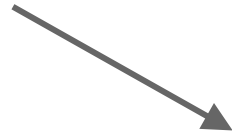
# Spatial Transformer Networks



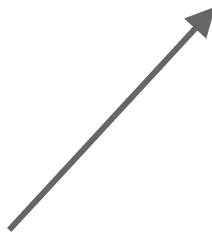
Input image:  
 $H \times W \times 3$

Box Coordinates:  
 $(x_c, y_c, w, h)$

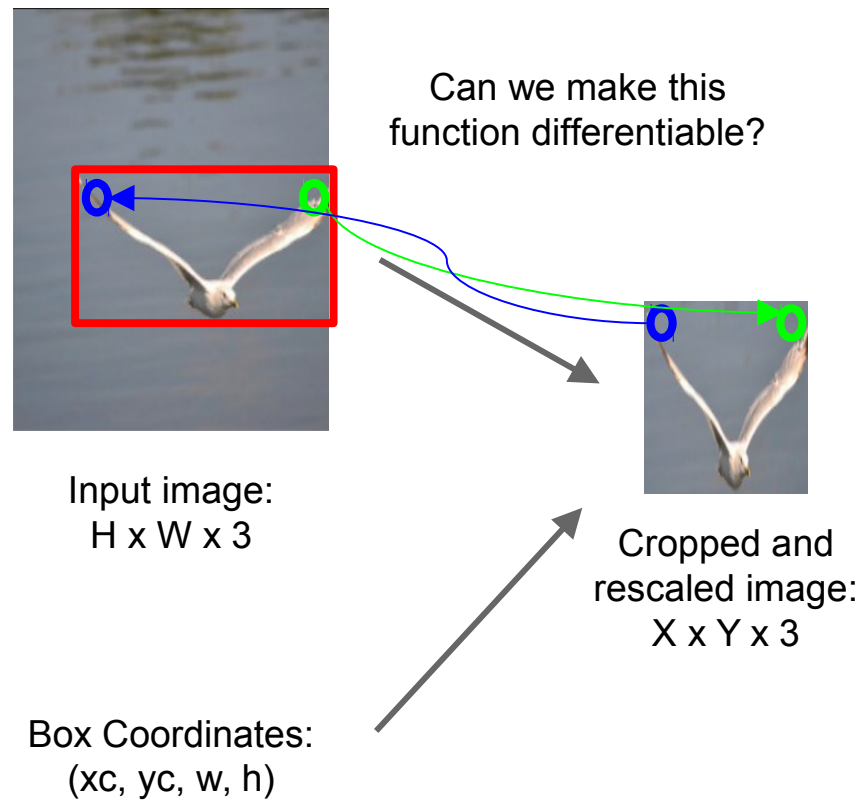
Can we make this  
function differentiable?



Cropped and  
rescaled image:  
 $X \times Y \times 3$



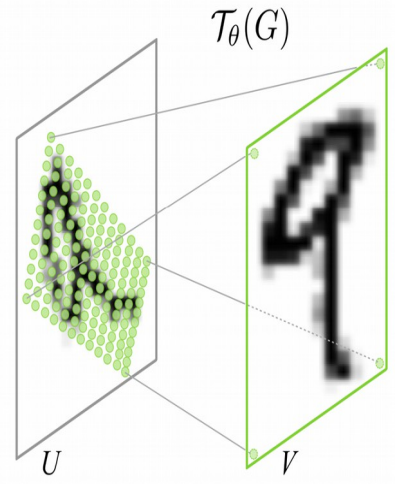
# Spatial Transformer Networks



**Idea:** Function mapping *pixel coordinates* (xt, yt) of output to *pixel coordinates* (xs, ys) of input

Network attends to input by predicting  $\theta$

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix}$$



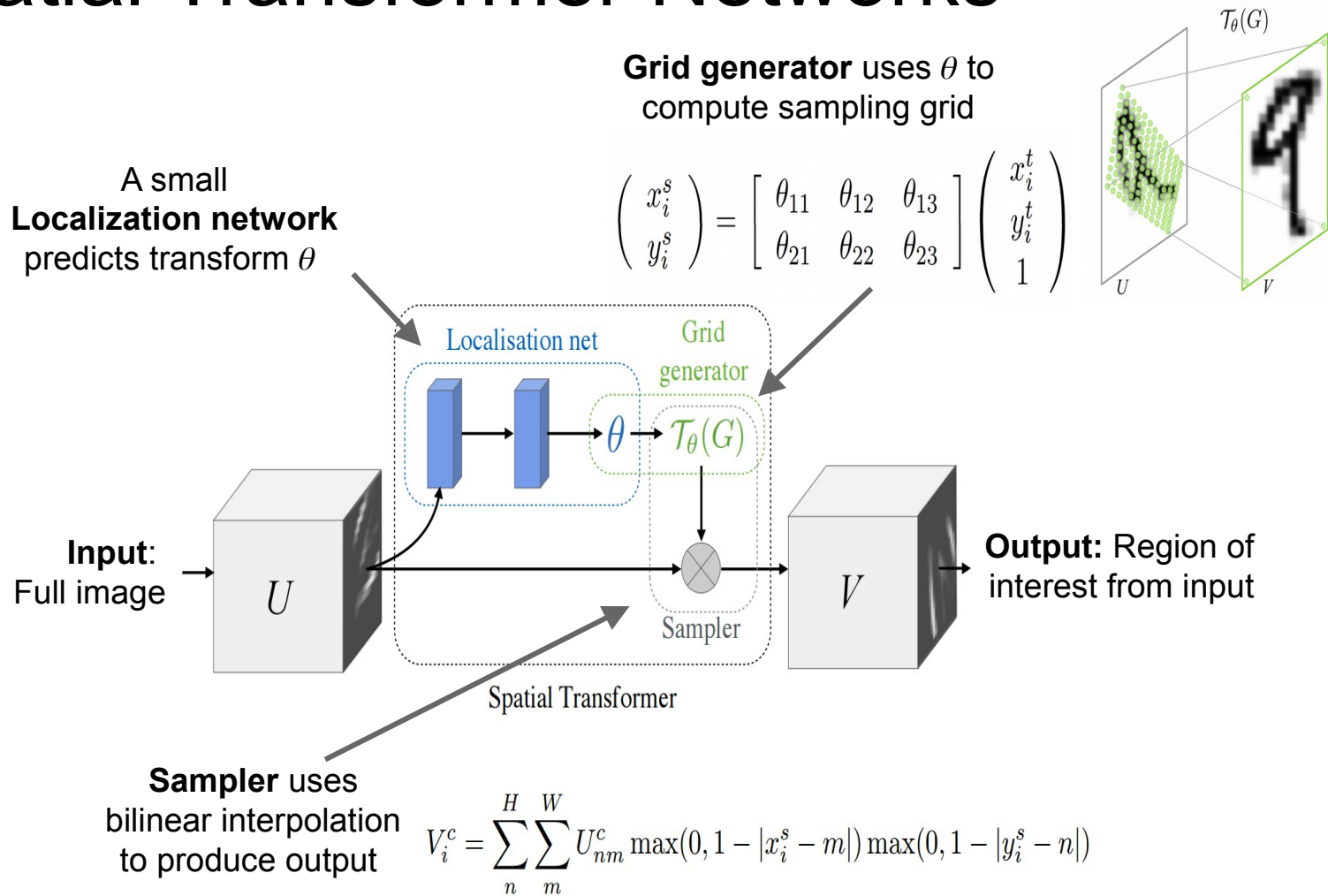
Repeat for all pixels in *output* to get a **sampling grid**

Then use **bilinear interpolation** to compute output

Jaderberg et al, "Spatial Transformer Networks", NIPS 2015



# Spatial Transformer Networks



# Spatial Transformer Networks

Insert spatial transformers into a classification network and it learns to attend and transform the input

Differentiable “attention / transformation” module

