DEEP LEARNING

Computer Vision Applications

Ing. Zese Riccardo
Ing. Bizzarri Alice
alice.bizzarri@unife.it

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Outline

Practical applications and research directions

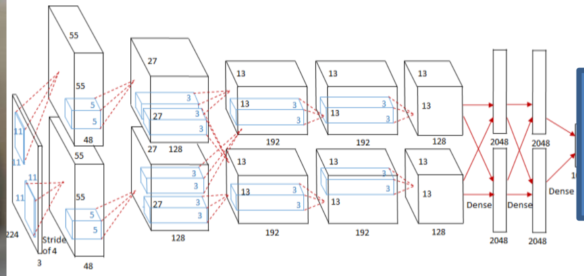
- Computer Vision
- Autoencoders
- GAN
- Transfer Learning

Computer Vision

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Computer Vision, Object Detection

So far: Image Classification



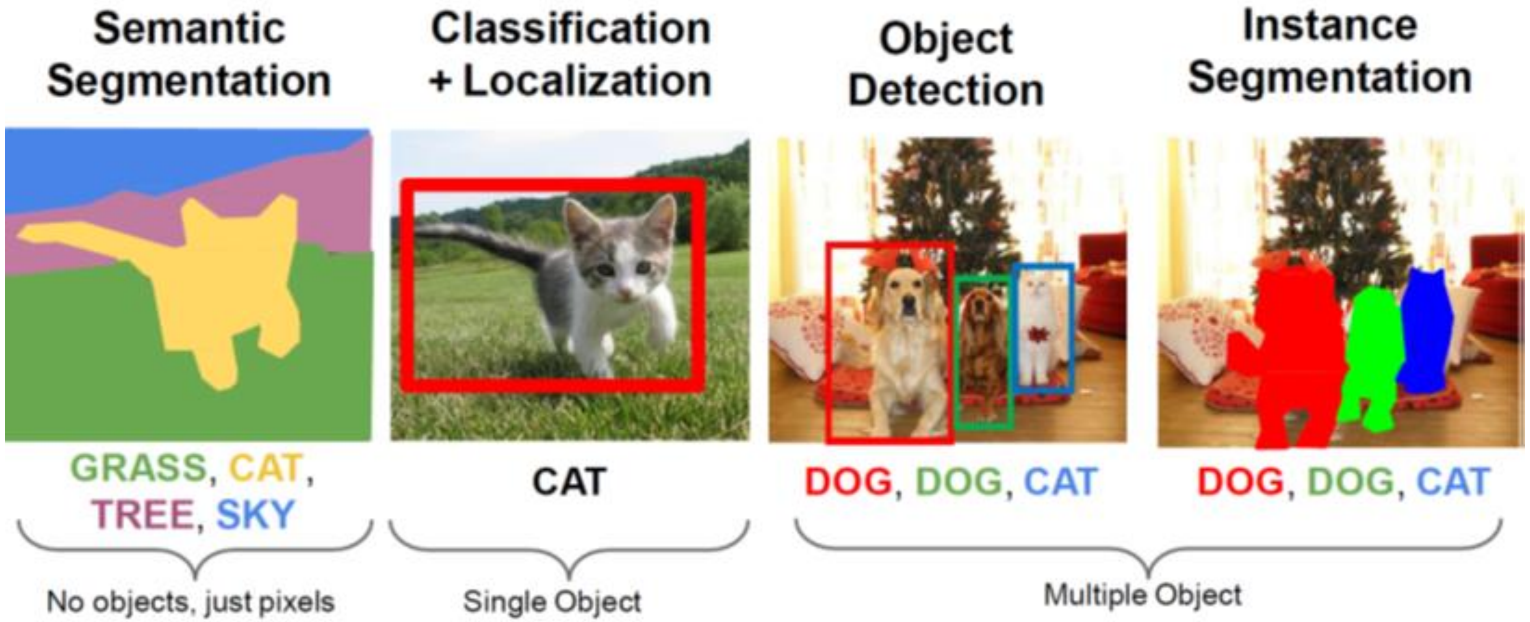
FC

Fully-connected
From 4096 to 100

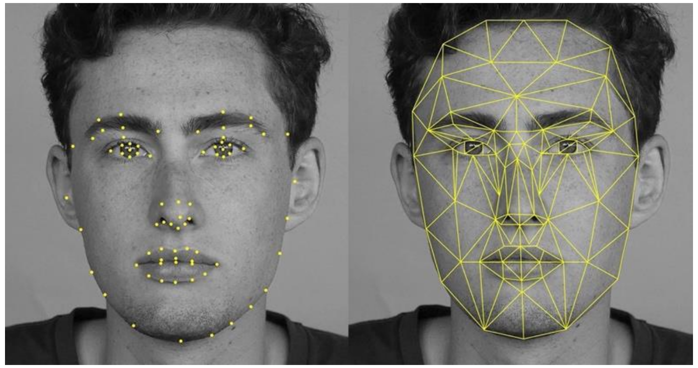
Class Scores
Quokka: 0.84
Cat: 0.1
Dog: 0.05
Car: 0.01

http://7wallpapers.net/quokka/

Other Computer Vision Tasks



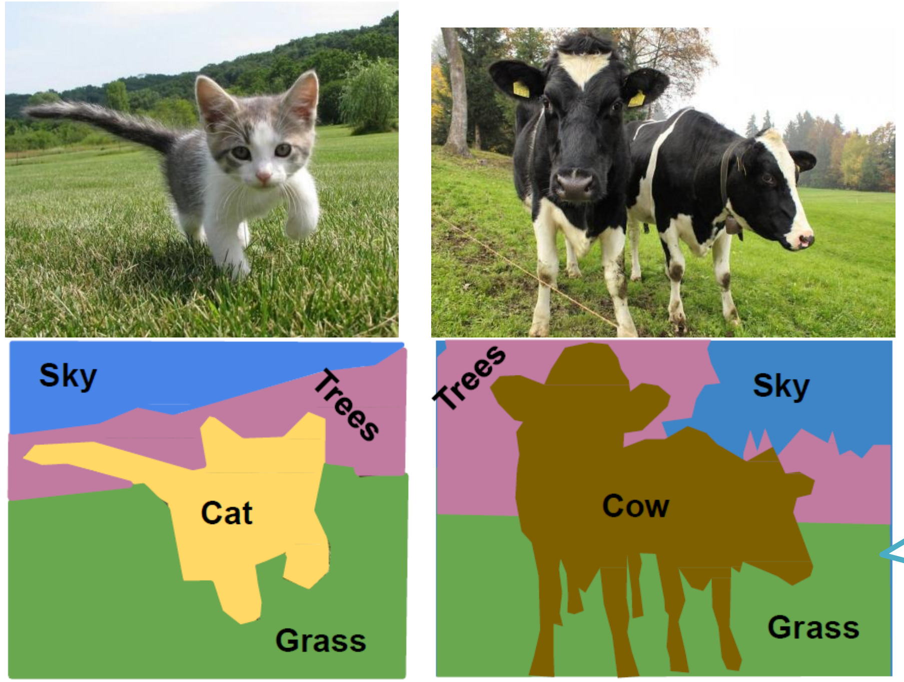
Keypoint detection



https://pixabay.com/photos/pets-christmas-dogs-cat-962215/ - https://pixabay.com/p-1246693/ - CC0 public domain (https://creativecommons.org/publicdomain/zero/1.0/deed.en)

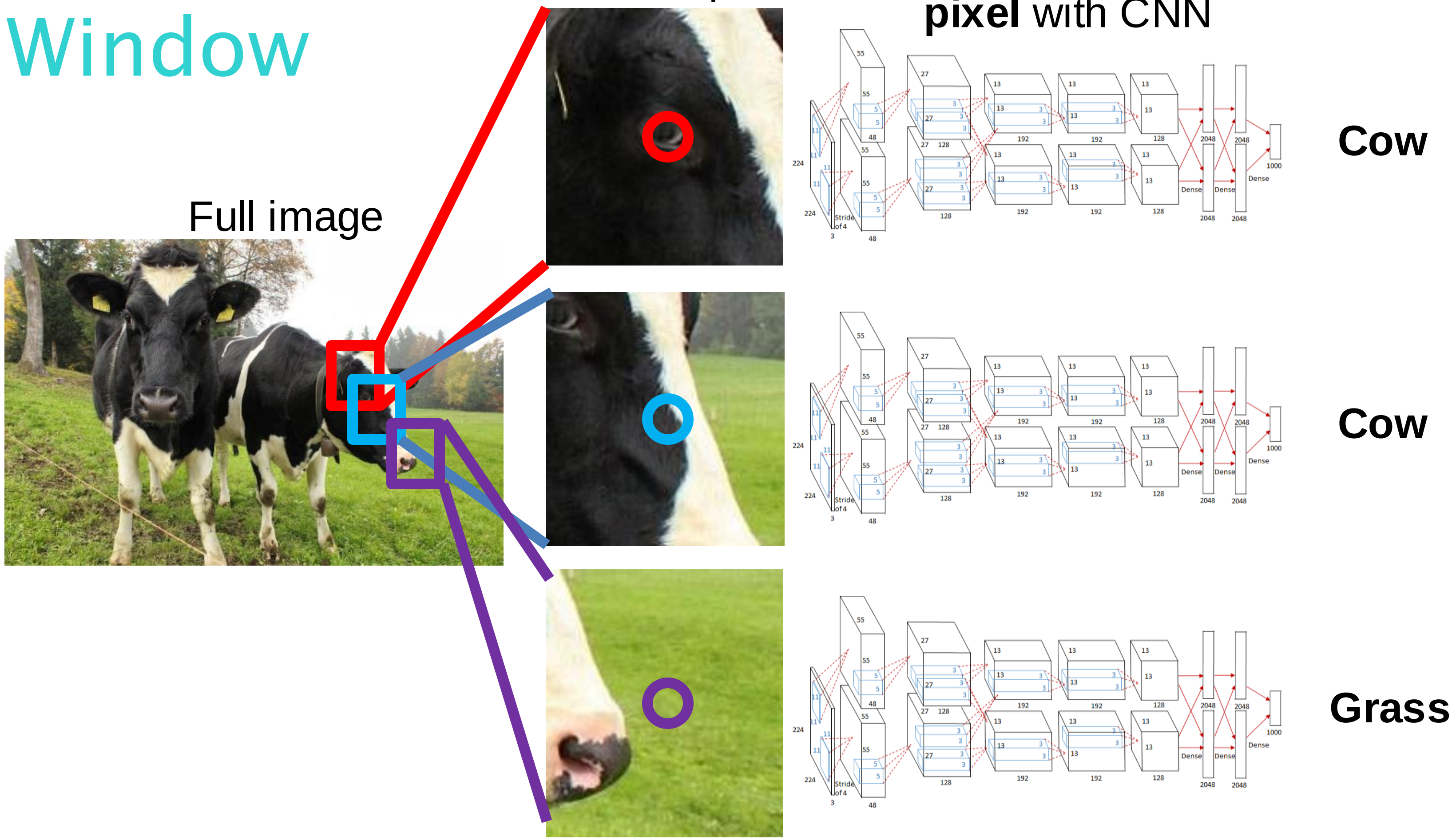
Semantic Segmentation

- Label each pixel in the image with a category label
- Don't differentiate instances, only care about pixels
- We assume we know which classes to search in images (cat, cow, tree, dog, …)



NOTE: it's only about pixel. Here we do not distinguish the two cows. Every pixel is classified independently → we have a big mass of pixels classified as cow instead of two different cows.

Semantic Segmentation Idea: Sliding Window

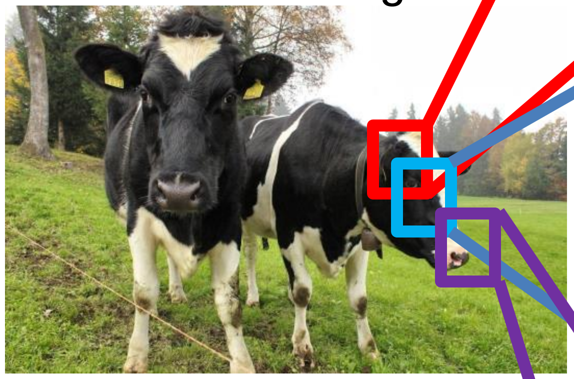


Farabet et al, “Learning Hierarchical Features for Scene Labeling,” TPAMI2013
Pinheiro and Collobert, “Recurrent Convolutional Neural Networks for Scene Labeling”, ICML2014

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

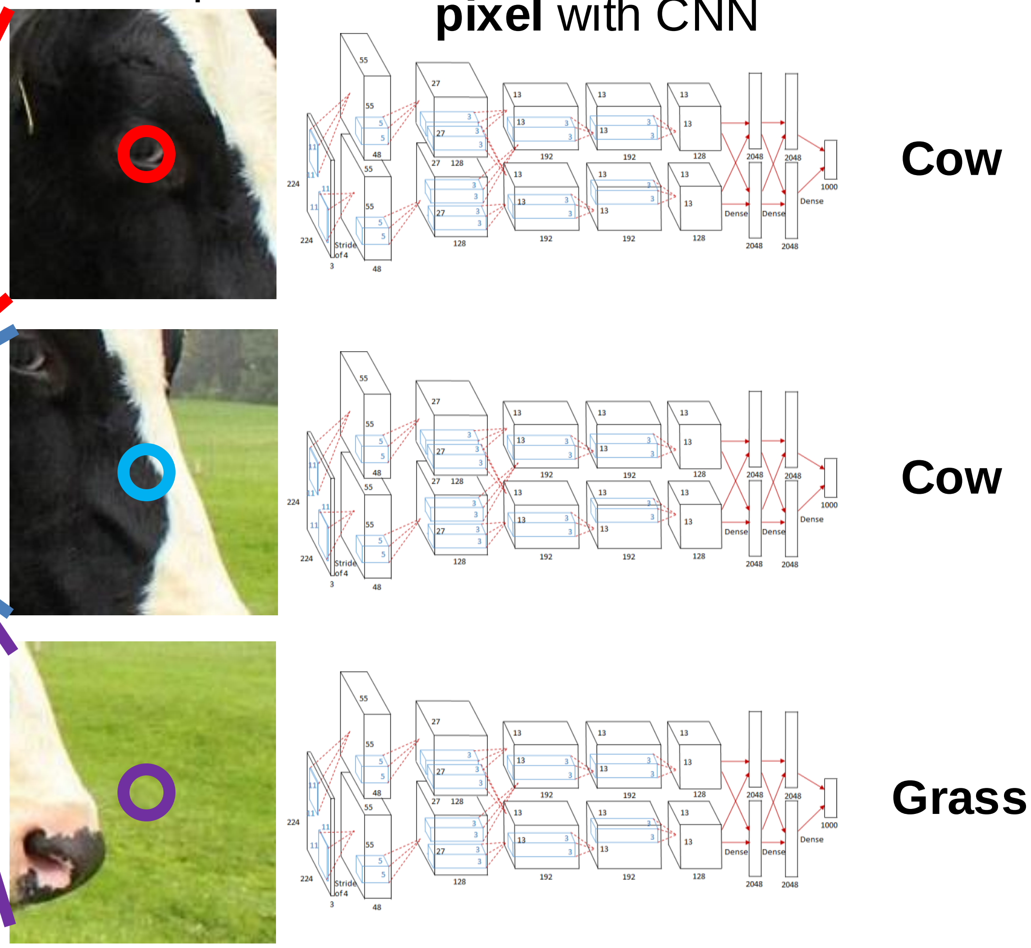
Semantic Segmentation Idea: Sliding Window

Full image



Extract patch

Classify **center pixel** with CNN



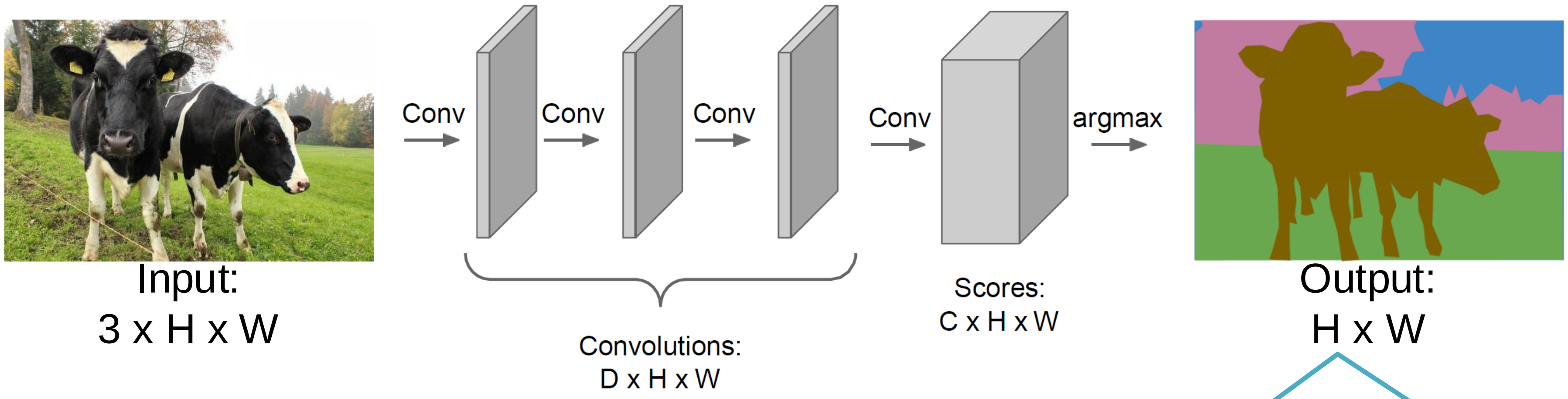
Problem: Very inefficient! Not reusing shared features between overlapping patches

Farabet et al, “Learning Hierarchical Features for Scene Labeling,” TPAMI2013
Pinheiro and Collobert, “Recurrent Convolutional Neural Networks for Scene Labeling”, ICML2014

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Semantic Segmentation
Idea: Fully Convolutional

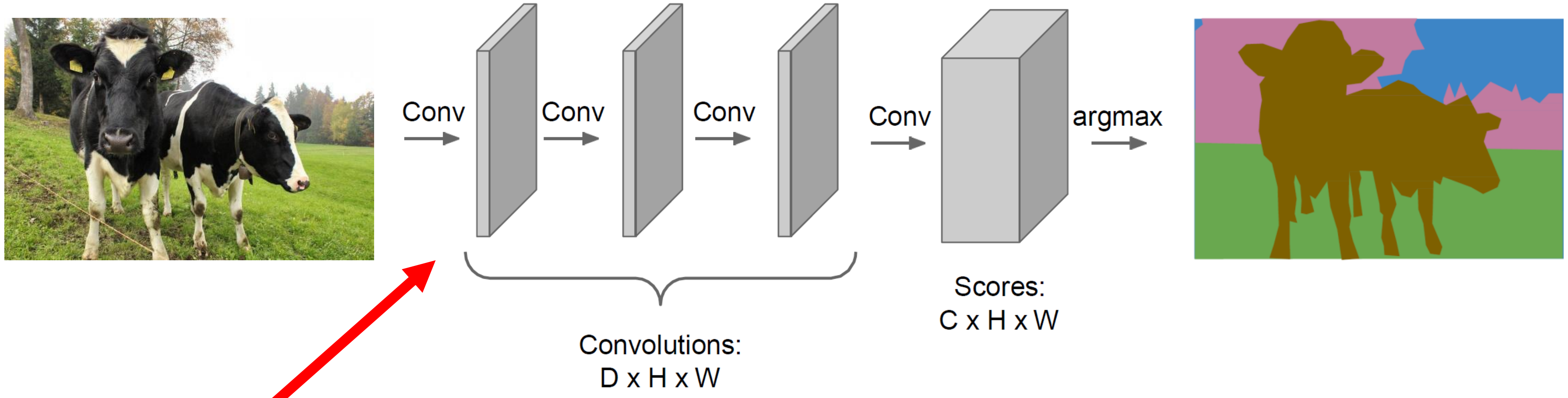
Design a network as a bunch of convolutional layers
to make predictions for pixels all at once!



Loss function: Cross entropy
loss on every pixel.

Semantic Segmentation Idea: Fully Convolutional

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

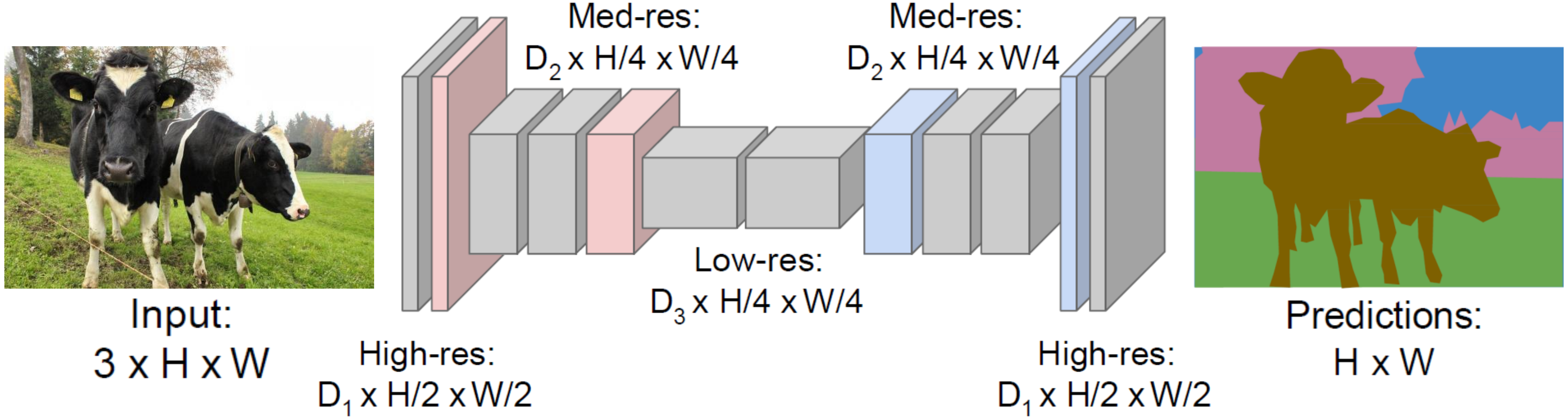


Problem: convolutions at original image resolution will be very expensive

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Semantic Segmentation
Idea: Fully Convolutional

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



Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

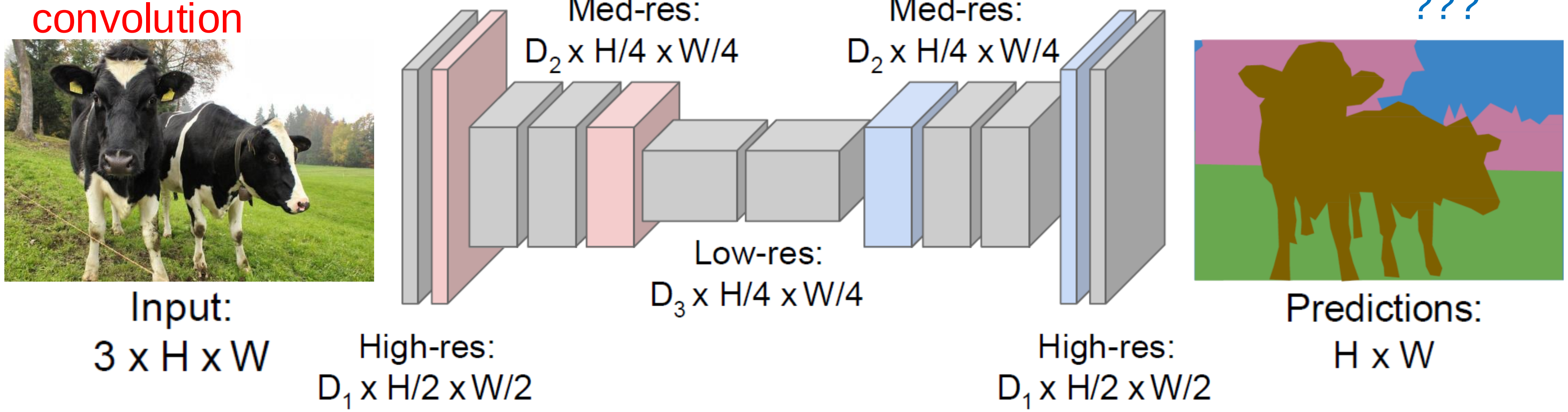
AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Semantic Segmentation
Idea: Fully Convolutional

Downsampling
Pooling, strided convolution

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

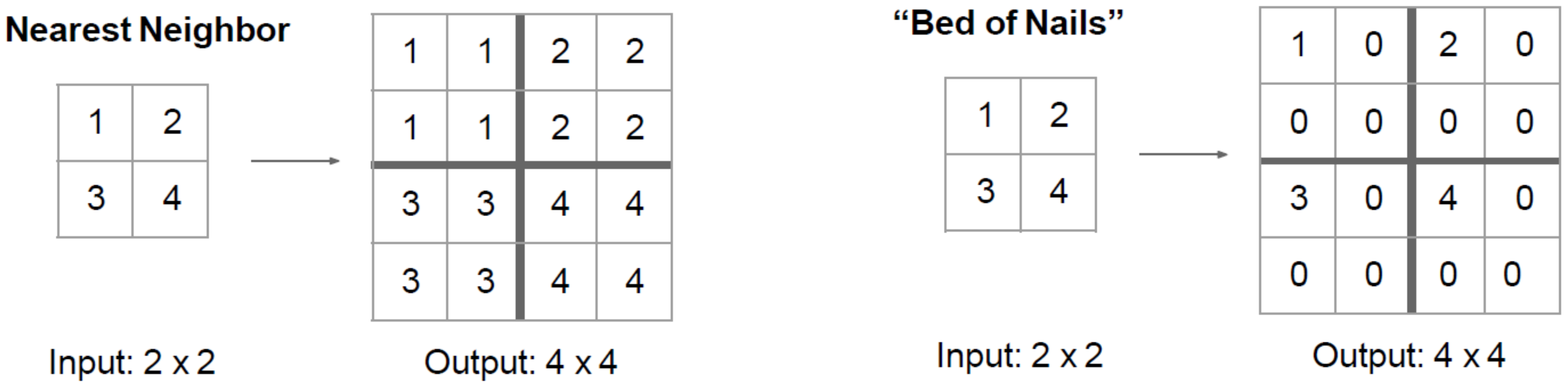
Upsampling
???



Long, Shelhamer, and Darrell, “Fully Convolutional Networks for Semantic Segmentation”, CVPR 2015
Noh et al, “Learning Deconvolution Network for Semantic Segmentation”, ICCV 2015

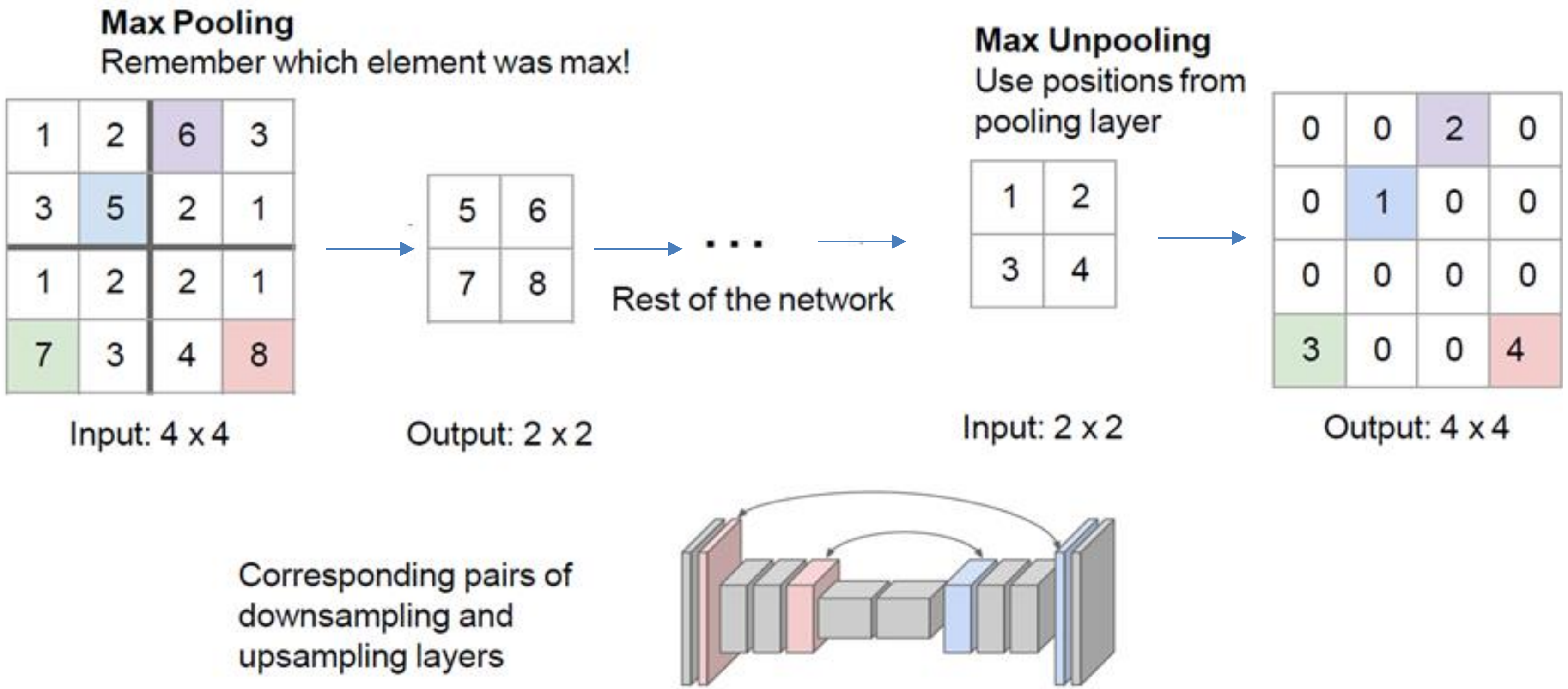
AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

In-network Upsampling
Unpooling



AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

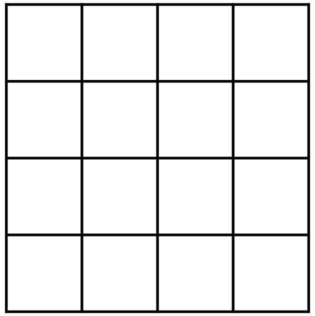
In-network Upsampling
Max Unpooling



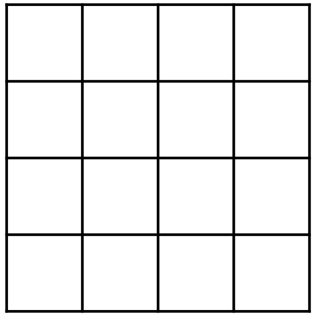
AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Learnable Upsampling
Transpose Convolution

Recall: Typical 3 x 3 convolution, stride 1 pad 1



Input: 4x4



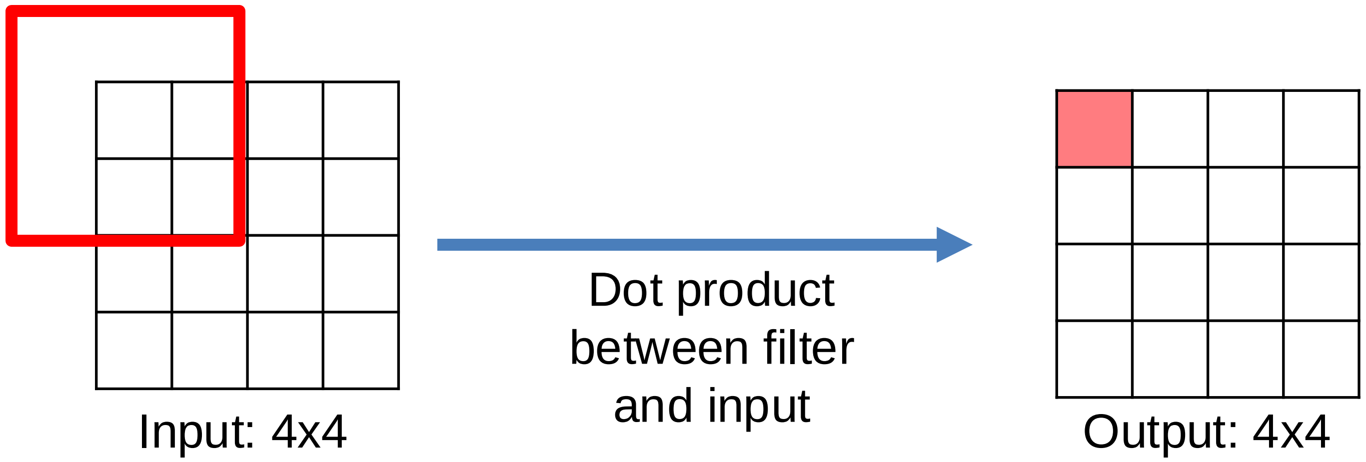
Output: 4x4

Unpooling are fixed functions, here we learn some weigths guiding the upsample

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Learnable Upsampling
Transpose Convolution

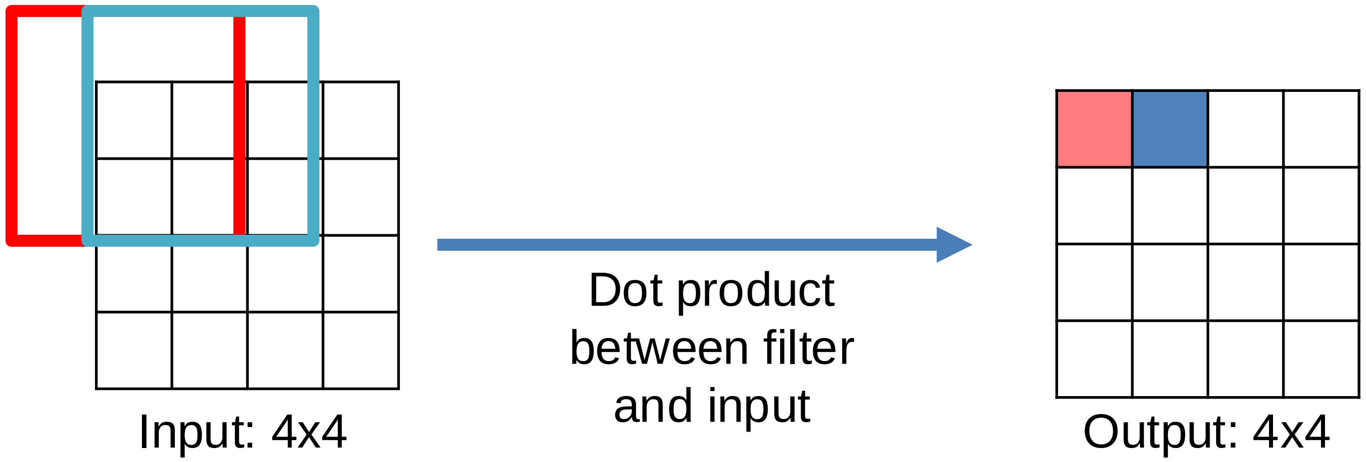
Recall: Typical 3 x 3 convolution, stride 1 pad 1



AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Learnable Upsampling
Transpose Convolution

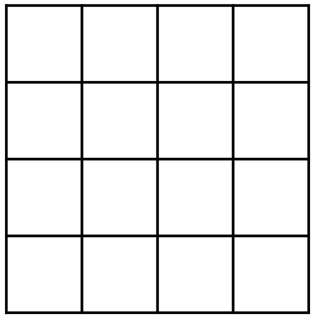
Recall: Typical 3 x 3 convolution, stride 1 pad 1



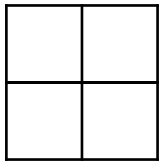
AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Learnable Upsampling
Transpose Convolution

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



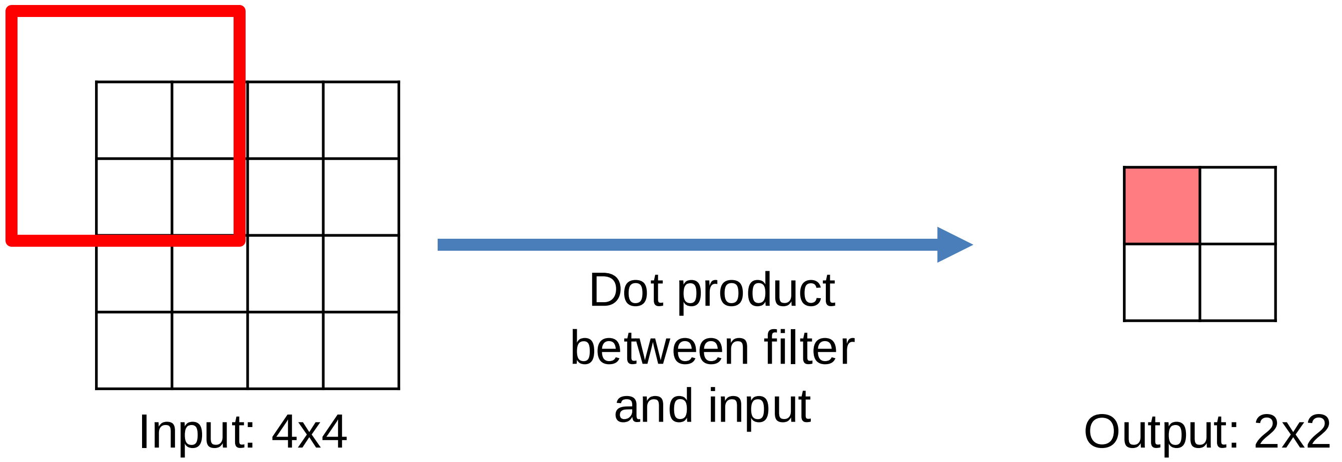
Input: 4x4



Output: 2x2

Learnable Upsampling
Transpose Convolution

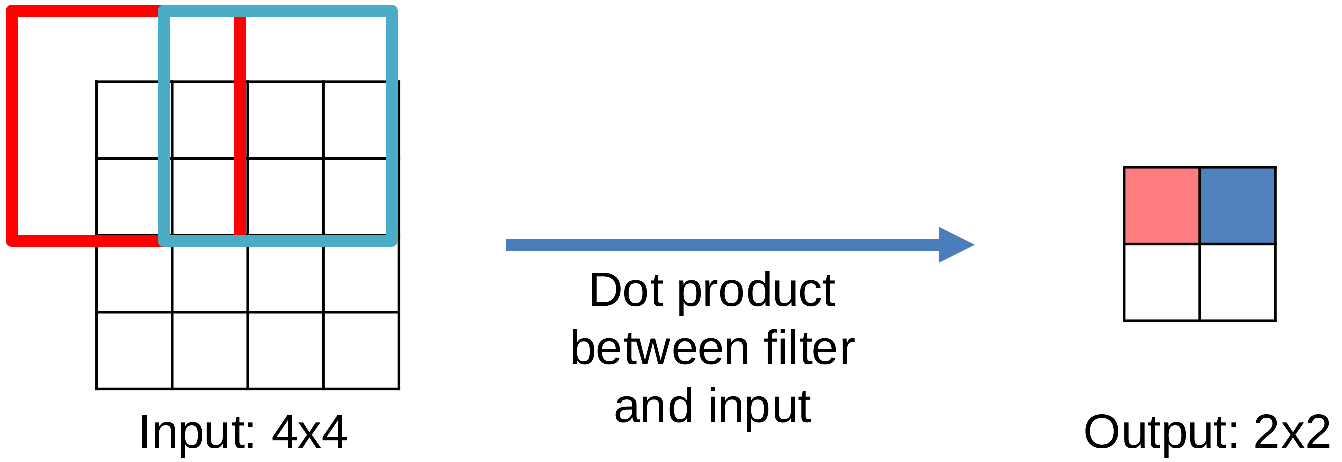
Recall: Typical 3 x 3 convolution, **stride 2** pad 1



AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Learnable Upsampling
Transpose Convolution

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



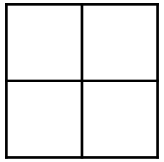
Filter moves 2 pixels in the input for every one pixel in the output

Stride gives ratio between movement in input and output

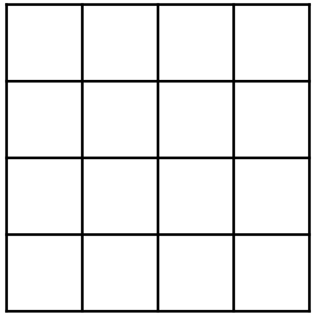
AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Learnable Upsampling
Transpose Convolution

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



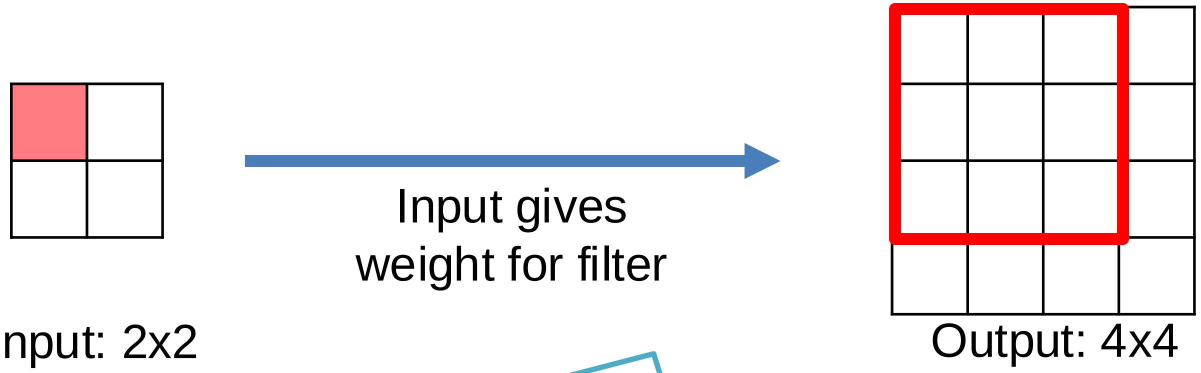
Input: 2x2



Output: 4x4

Learnable Upsampling
Transpose Convolution

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

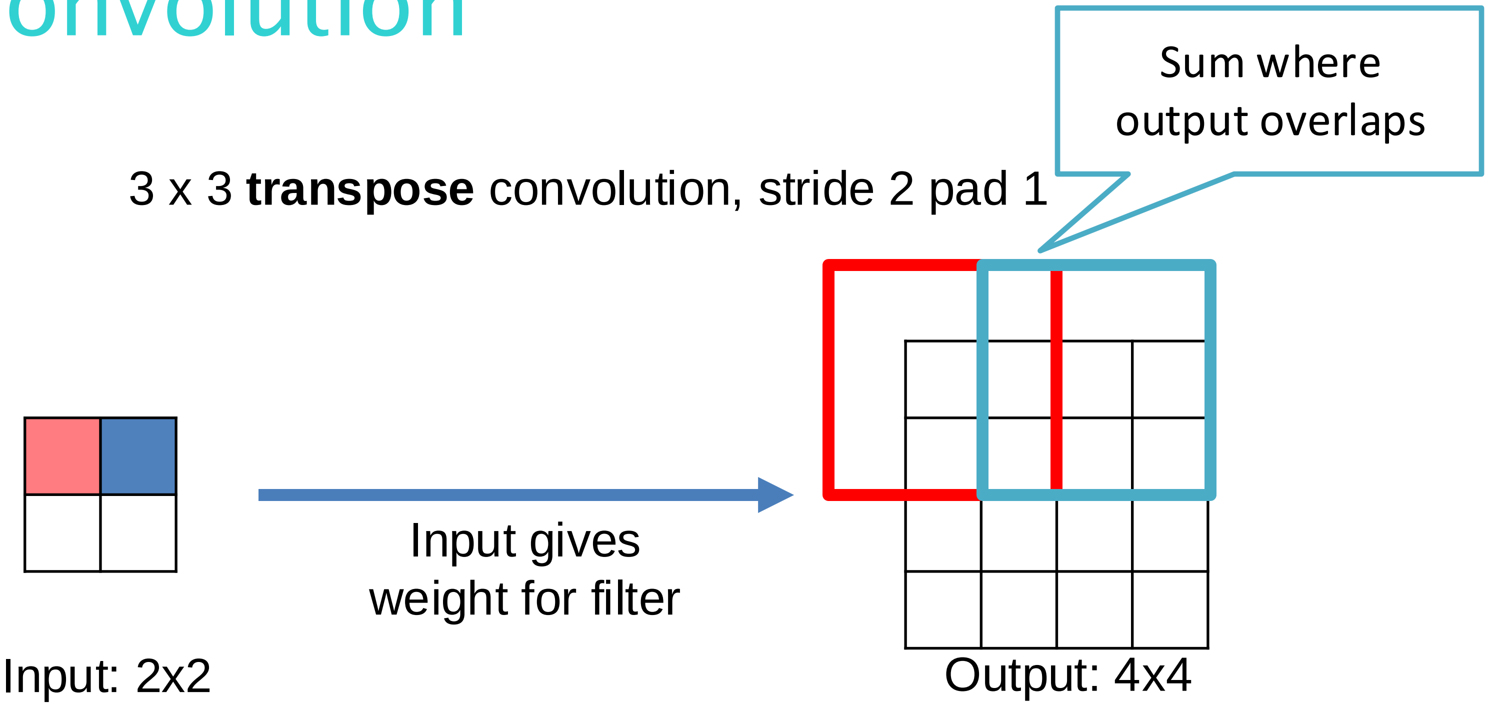


We take the value of the input, multiply the filter with the value and copy the result in the 3x3 zone in output.

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Learnable Upsampling
Transpose Convolution

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



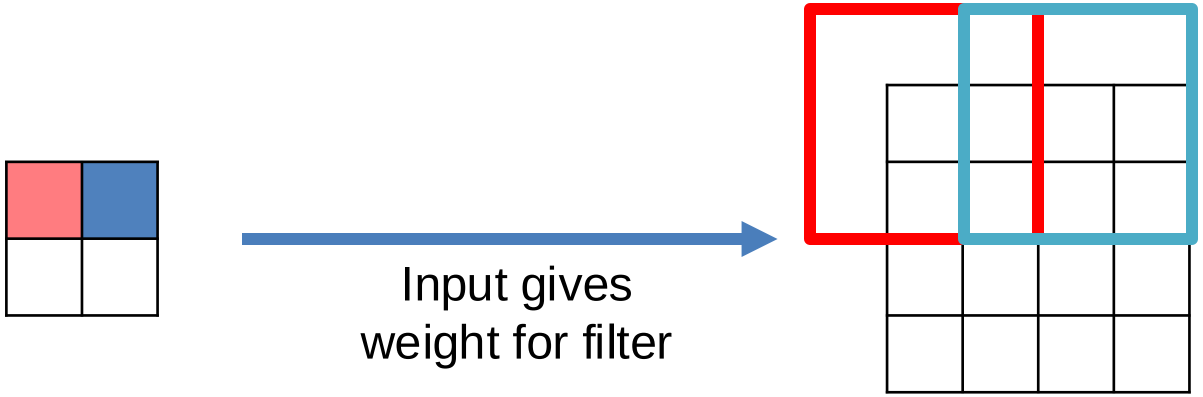
Filter moves 2 pixels in the **output** for every one pixel in the **input**

Stride gives ratio between movement in output and input

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Learnable Upsampling
Transpose Convolution

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



Input: 2x2

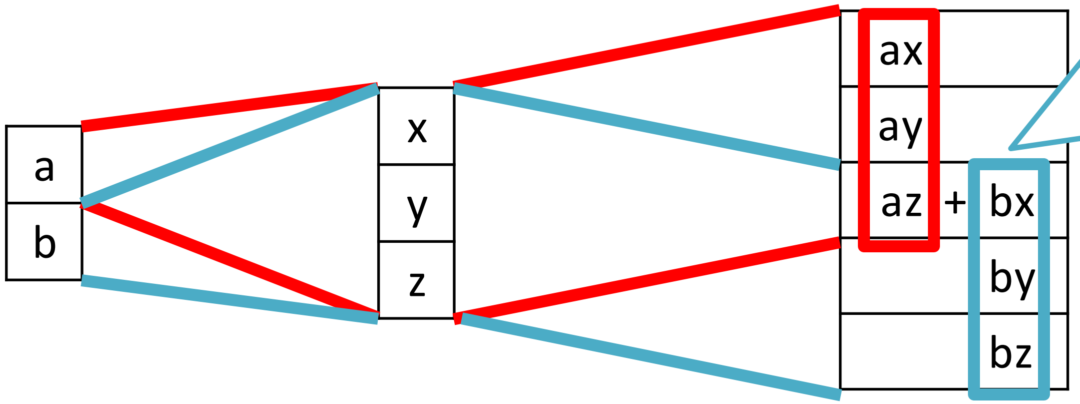
Output: 4x4

Other names:

- Deconvolution
- Upconvolution
- Fractionally strided convolution
- Backward strided convolution

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Transpose Convolution
1D Example



The stride is important here to avoid summing to much values for each shared cell.

Output contains copies of the filter weighted by the input, summing at where it overlaps in the output

Need to crop one pixel from output to make output exactly 2x input

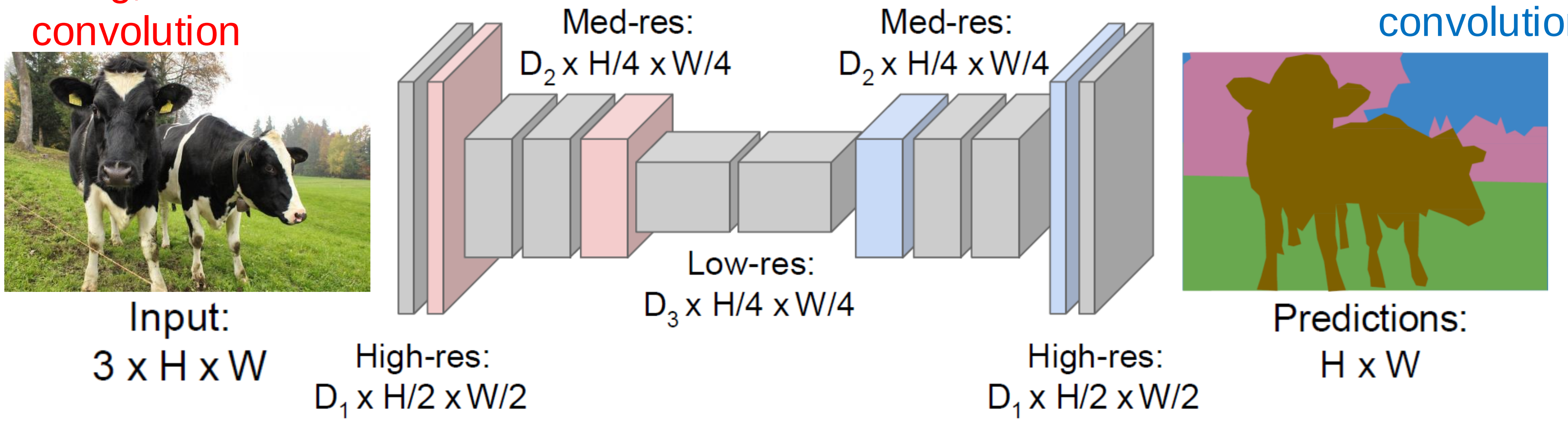
AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Semantic Segmentation
Idea: Fully Convolutional

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

Downsampling
Pooling, strided convolution

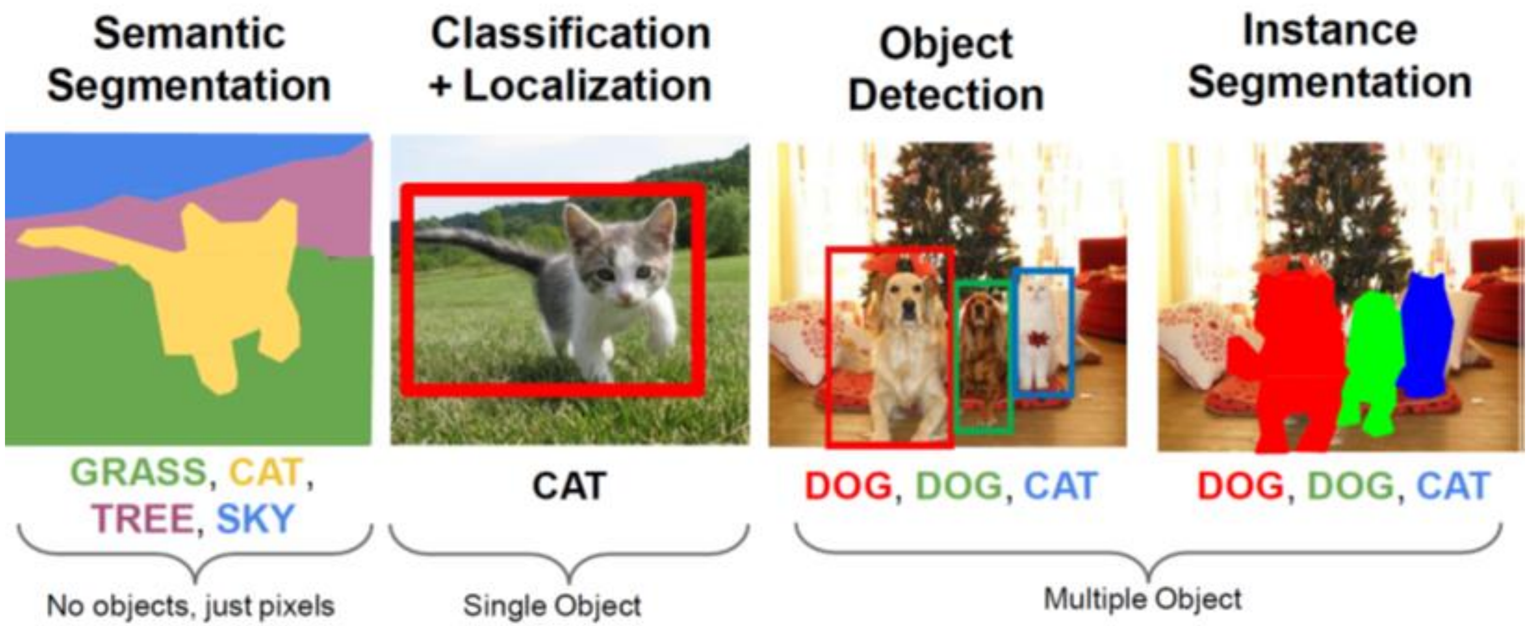
Upsampling
Unpooling or strided transpose convolution



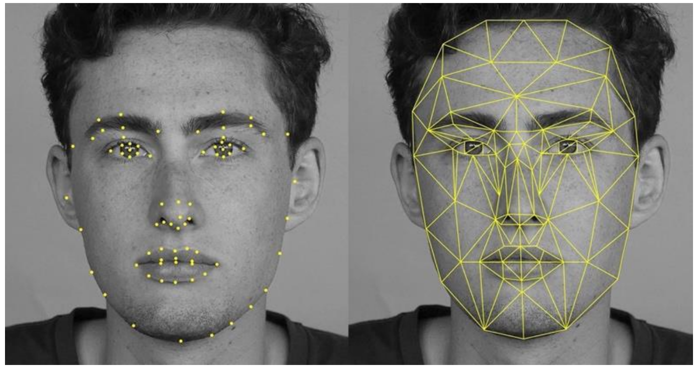
Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

Other Computer Vision Tasks

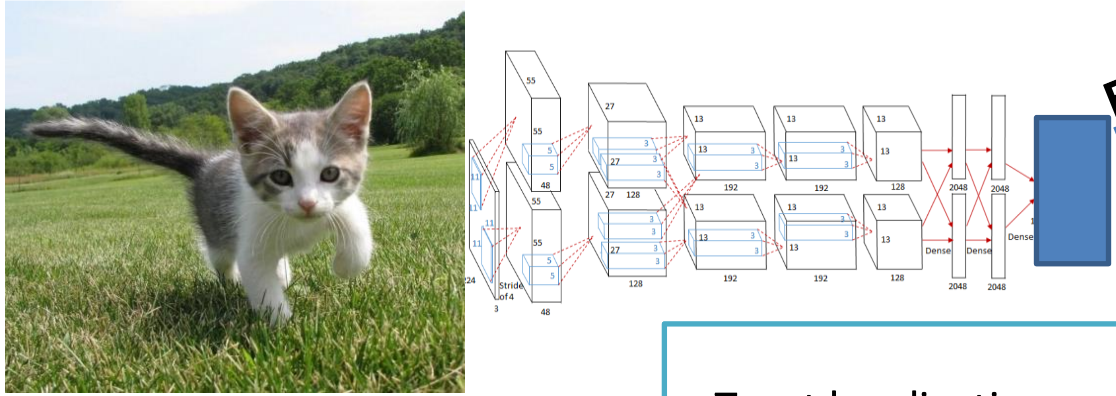


Keypoint detection



Classification + Localization

Supposing 100 classes



Fully-connected
From 4096 to 100

Fully-connected
From 4096 to 4

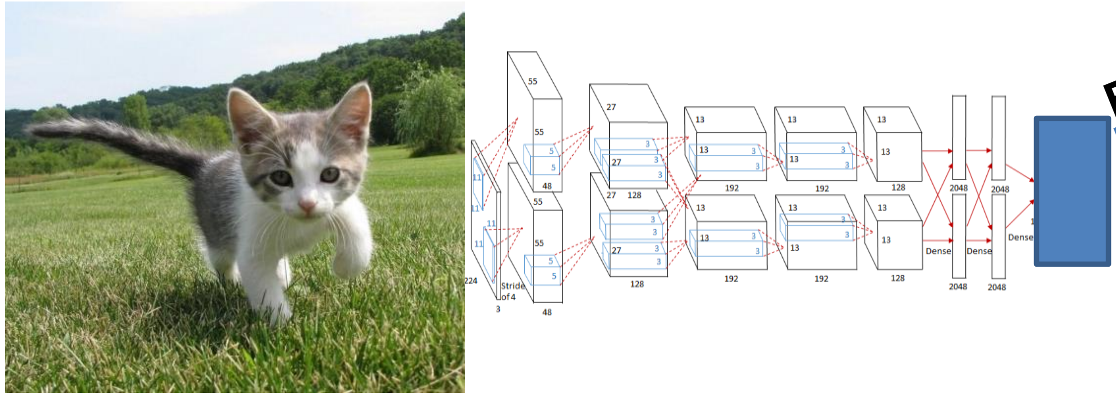
Class Scores
Cat: 0.84
Quokka: 0.1
Dog: 0.05
Car: 0.01

Box Coords
x: 1
y: 5
w: 20
h: 18

Treat localization as a regression problem!

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Classification + Localization



Softmax Loss
(with correct label)

Class Scores
Cat: 0.9
Quokka: 0.1
Dog: 0.05
Car: 0.01

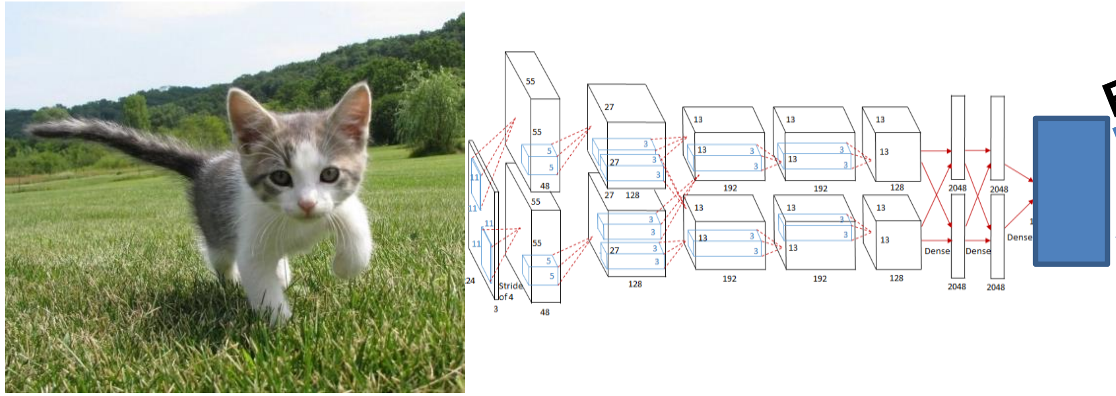
Fully-connected
From 4096 to 100

Fully-connected
From 4096 to 4

Box Coords
x: 1
y: 5
w: 20
h: 18

L2 Loss
(with correct coords)

Classification + Localization



Softmax Loss

Class Scores

Cat: 0.9
Quokka: 0.1
Dog: 0.05
Car: 0.01

Fully-connected
From 4096 to 100

Fully-connected
From 4096 to 4

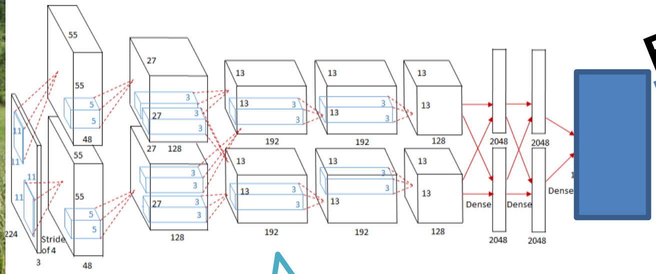
Box Coords

x: 1
y: 5
w: 20
h: 18

L2 Loss

Multitask Loss
Sum of the different loss
Softmax + L2

Classification + Localization



Softmax Loss

Class Scores

Cat: 0.9
Quokka: 0.1
Dog: 0.05
Car: 0.01

Fully-connected
From 4096 to 100

Fully-connected
From 4096 to 4

Box Coords

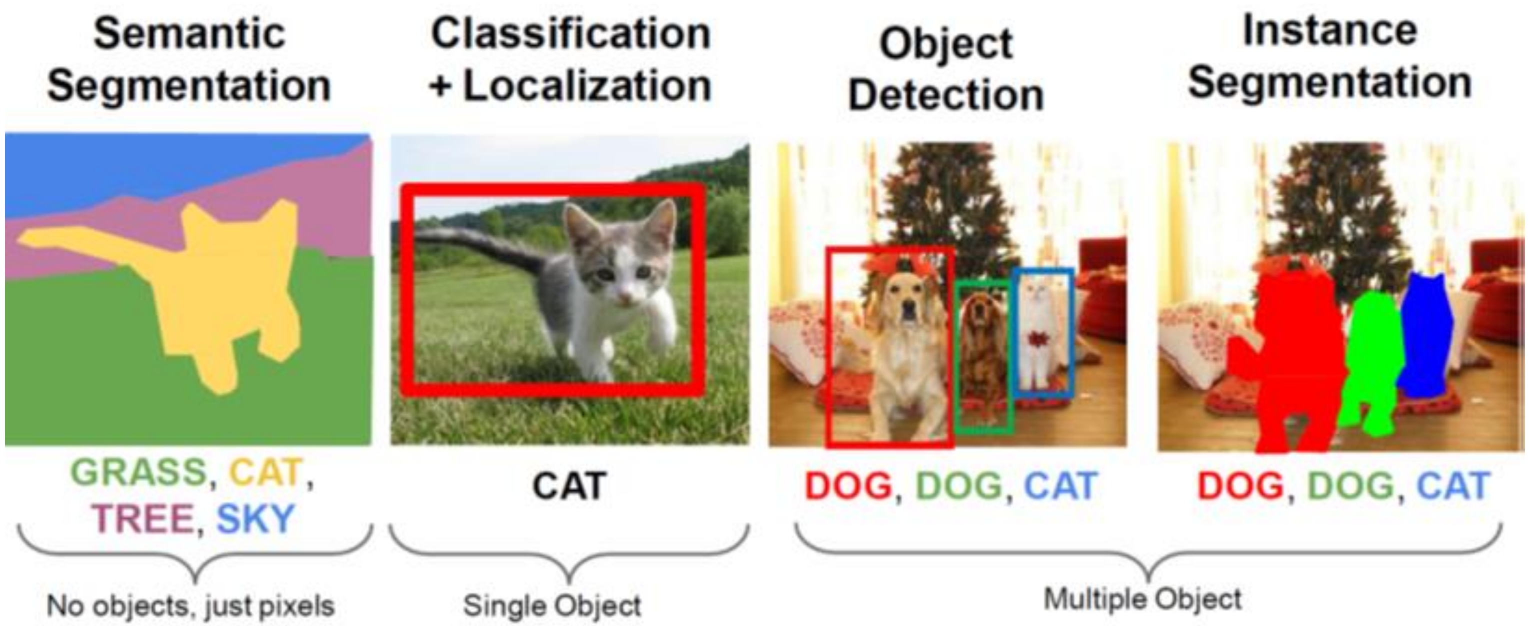
x: 1
y: 5
w: 20
h: 18

L2 Loss

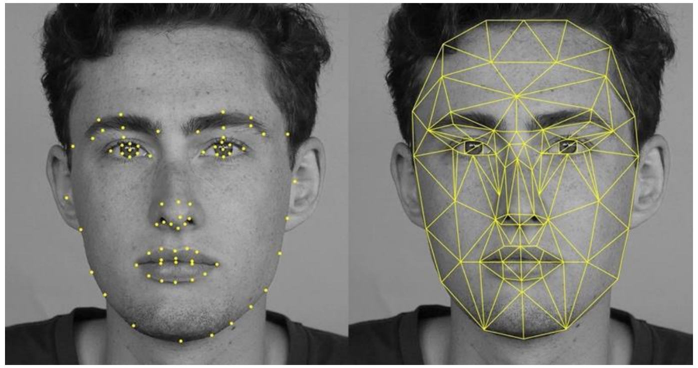
Often pre-trained on ImageNet
→ **Transfer Learning**

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Other Computer Vision Tasks



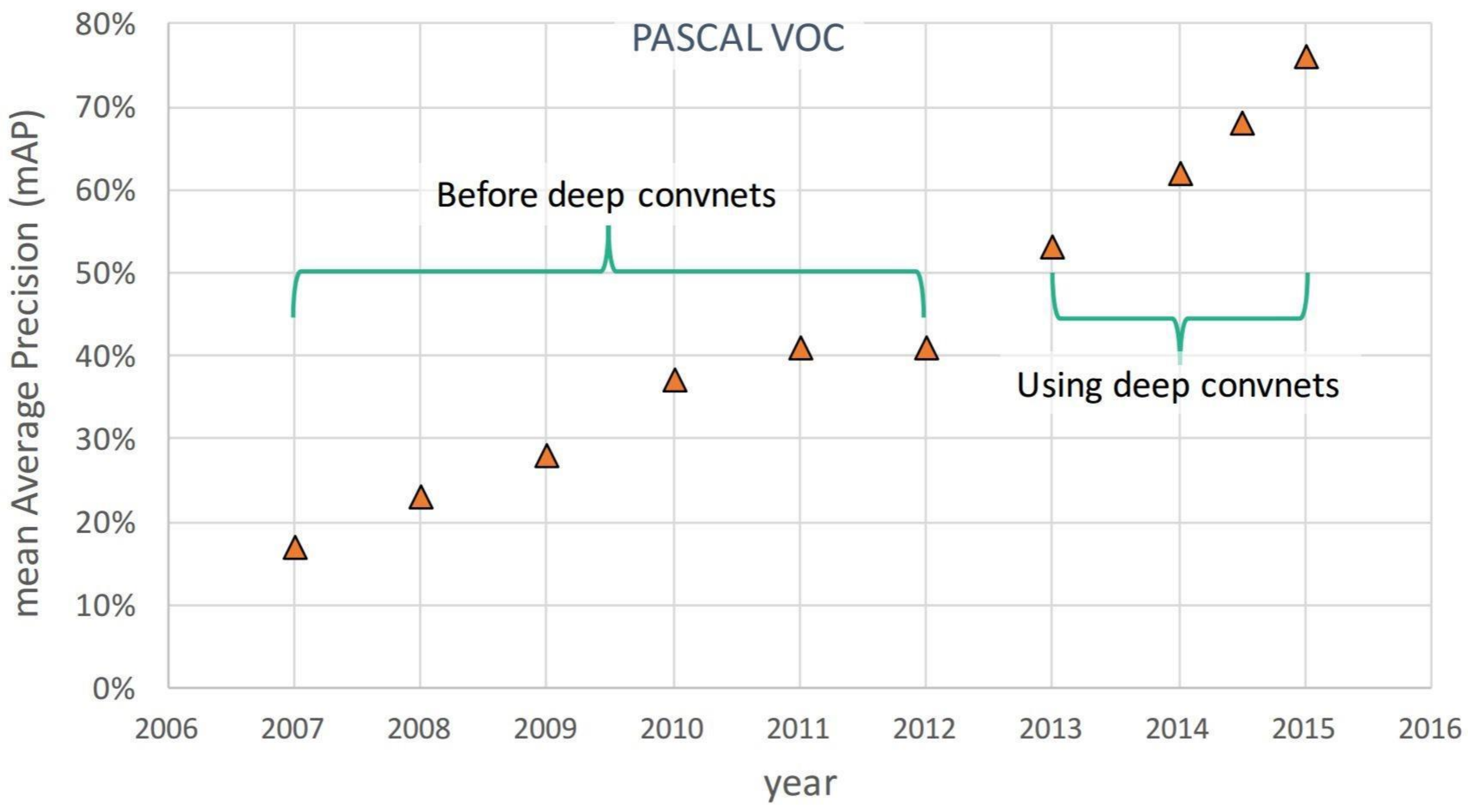
Keypoint detection



https://pixabay.com/photos/pets-christmas-dogs-cat-962215/ - https://pixabay.com/p-1246693/ - CC0 public domain (https://creativecommons.org/publicdomain/zero/1.0/deed.en)

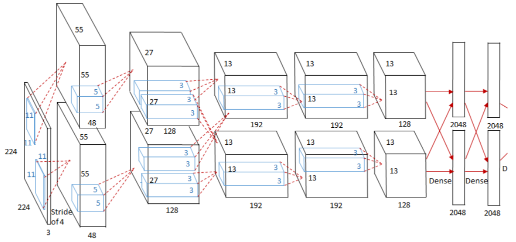
AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

Object Detection

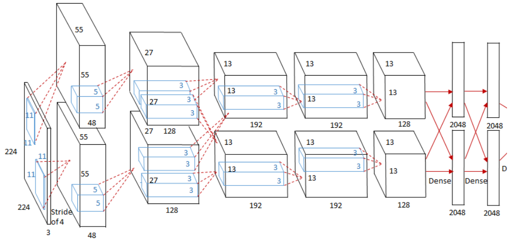


AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

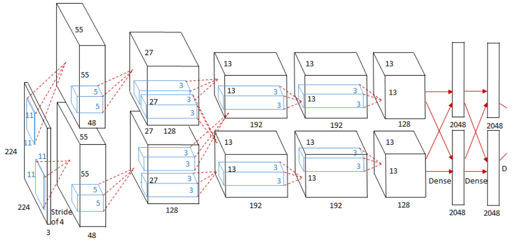
Object Detection as Regression?



CAT: (x, y, w, h)



DOG: (x, y, w, h)
DOG: (x, y, w, h)
CAT: (x, y, w, h)



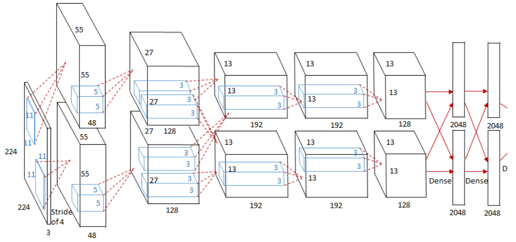
DUCK: (x, y, w, h)
DUCK: (x, y, w, h)
…

We have a set of classes to find, in each image every time we recognize an object among given classes we have to draw a box around it

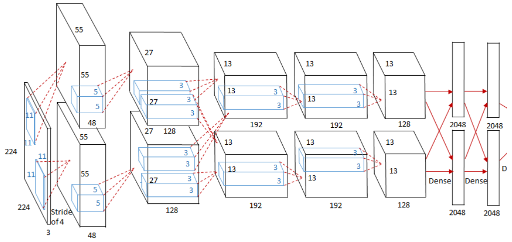
AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Object Detection as Regression?

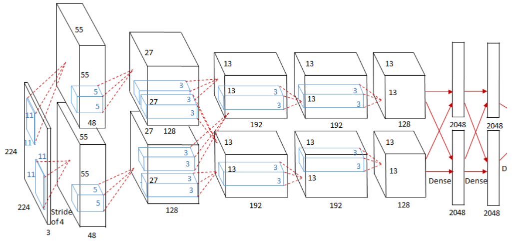
Each image needs a different number of outputs!



CAT: (x, y, w, h) 4 numbers



DOG: (x, y, w, h)
DOG: (x, y, w, h) 12 numbers
CAT: (x, y, w, h)

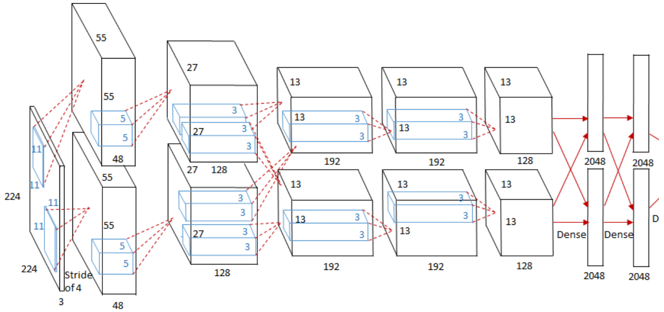


DUCK: (x, y, w, h)
DUCK: (x, y, w, h) Many numbers
…

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Object Detection as Classification: Sliding Window

Apply a CNN to many different crops of the image,
CNN classifies each crop as object or background

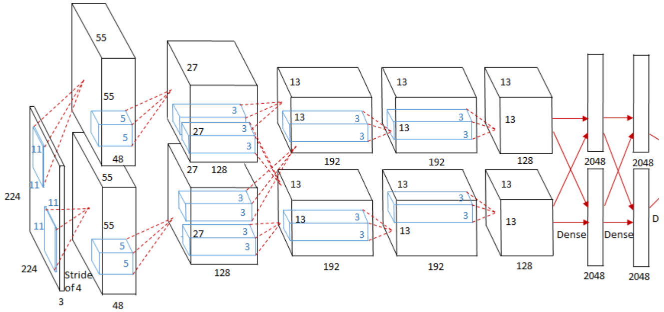


Dog? NO
Cat? NO
Background? YES

In addition to out categories (classes) we consider another general category → **background**

Object Detection as Classification: Sliding Window

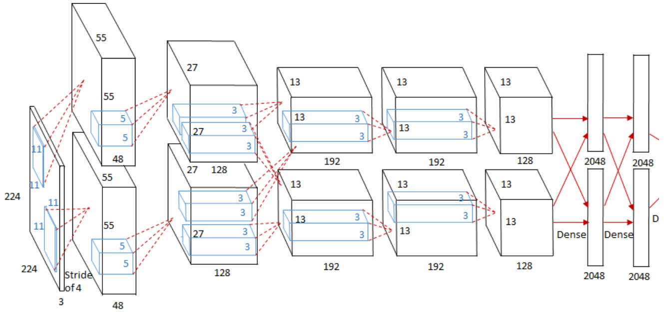
Apply a CNN to many different crops of the image, CNN classifies each crop as object or background



Dog?	YES
Cat?	NO
Background?	NO

Object Detection as Classification: Sliding Window

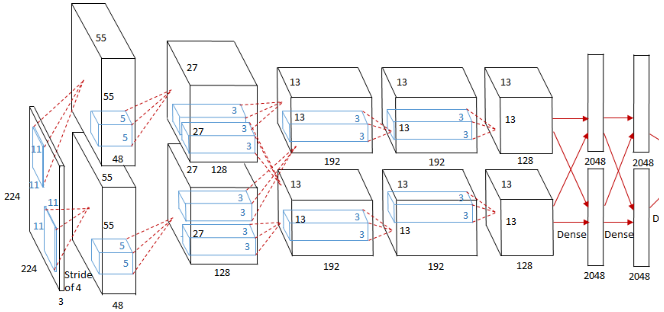
Apply a CNN to many different crops of the image, CNN classifies each crop as object or background



Dog?	YES
Cat?	NO
Background?	NO

Object Detection as Classification: Sliding Window

Apply a CNN to many different crops of the image, CNN classifies each crop as object or background

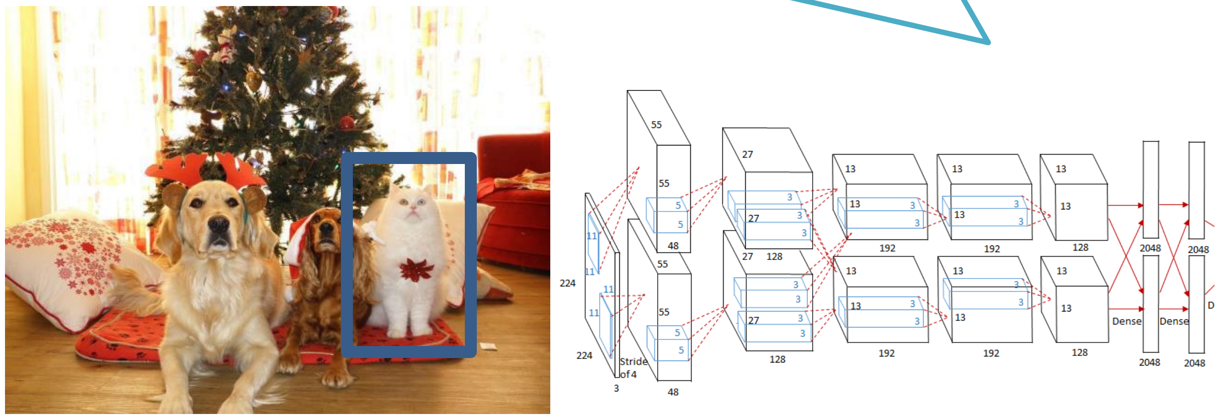


Dog?	NO
Cat?	YES
Background?	NO

AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

Object Detection as Classification: Sliding Window

Problem
Need to apply CNN to huge number of locations and scales, very computationally expensive!



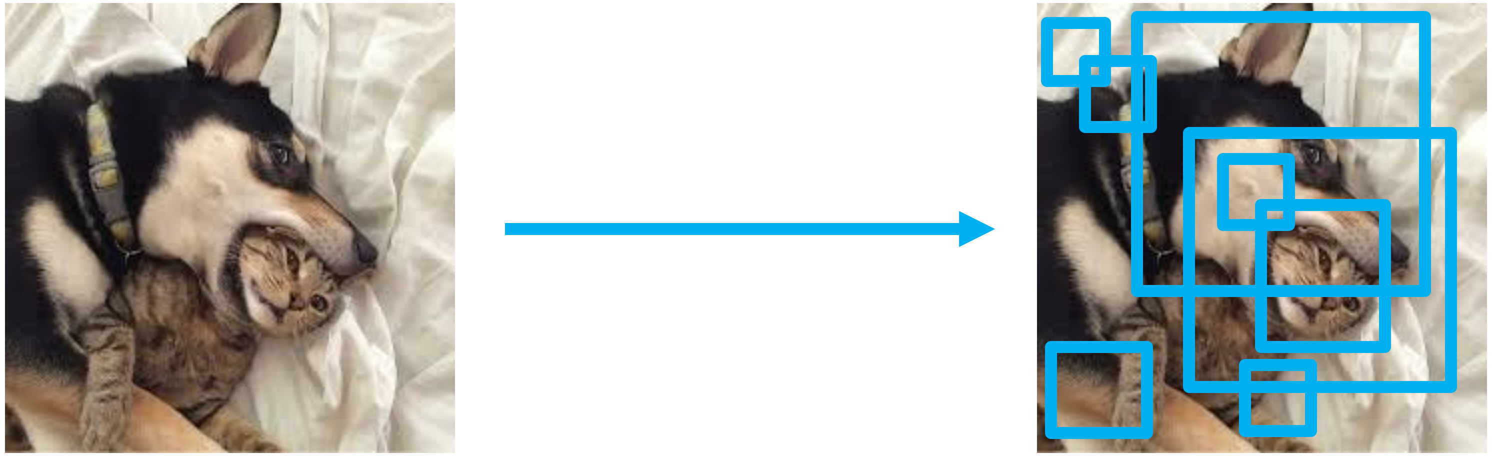
Dog? NO
Cat? YES
Background? NO

This approach tries many different windows that vary in size and position. It is unfeasible and this approach is not used in practice.

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Region Proposals

Used in classical computer vision techniques before DL
Find “blobby” image regions that are likely to contain objects
Relatively fast to run; e.g. Selective Search gives 1000 region proposals where the object may be in a few seconds on CPU



Alexe et al, “Measuring the objectness of image windows”, TPAMI 2012
Uijlings et al, “Selective Search for Object Recognition”, IJCV 2013
Cheng et al, “BING: Binarized normed gradients for objectness estimation at 300fps”, CVPR 2014
Zitnick and Dollar, “Edge boxes: Locating object proposals from edges”, ECCV 2014

R-CNN: CNN with Regions

Input image

Girshick et al, “Rich feature hierarchies for accurate object detection and semantic segmentation”, CVPR 2014. Figure copyright Ross Girshick, 2015. Reproduced with permission.
https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

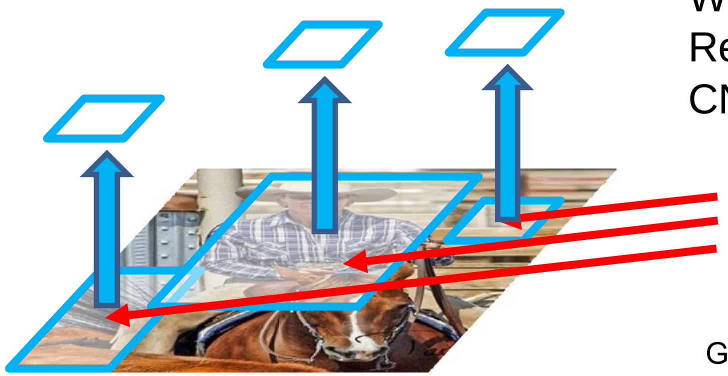
R-CNN: CNN with Regions

Input image

Girshick et al, “Rich feature hierarchies for accurate object detection and semantic segmentation”, CVPR 2014. Figure copyright Ross Girshick, 2015. Reproduced with permission. https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

R-CNN: CNN with Regions



Warped image regions:
Resize the region to the input size of the CNN

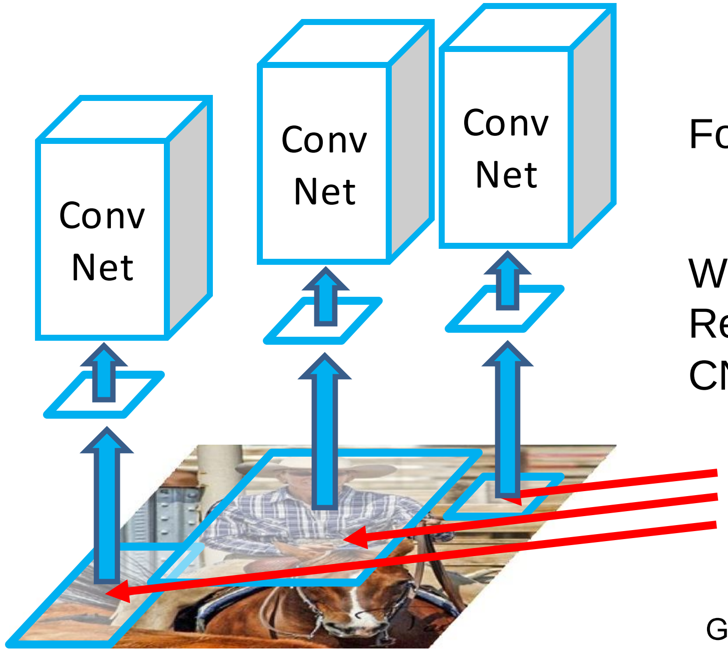
Regions of Interest
(RoI) (c.ca 2k)

Input image

Girshick et al, “Rich feature hierarchies for accurate object detection and semantic segmentation”, CVPR 2014. Figure copyright Ross Girshick, 2015. Reproduced with permission.
https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

R-CNN: CNN with Regions



Forward each region through ConvNet

Warped image regions:
Resize the region to the input size of the CNN

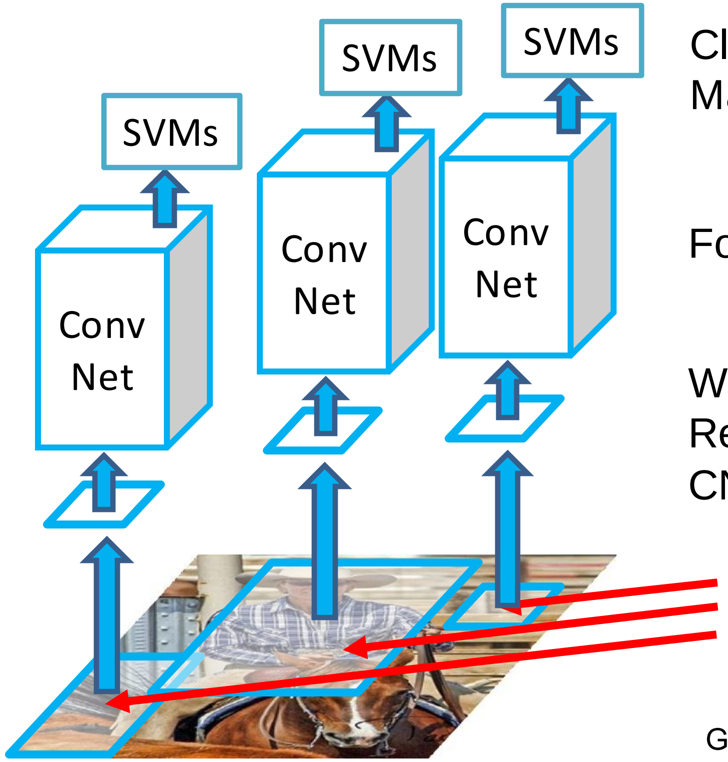
Regions of Interest
(RoI) (c.ca 2k)

Input image

Girshick et al, “Rich feature hierarchies for accurate object detection and semantic segmentation”, CVPR 2014. Figure copyright Ross Girshick, 2015. Reproduced with permission.
https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

R-CNN: CNN with Regions



Input image

Classify regions with Support Vector Machines

Forward each region through ConvNet

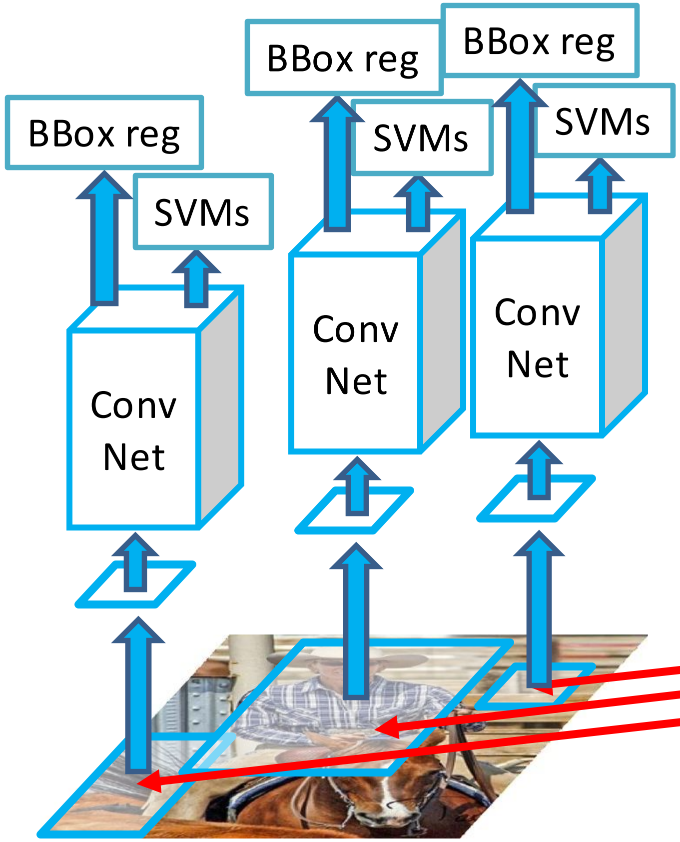
Warped image regions:
Resize the region to the input size of the CNN

Regions of Interest (RoI) (c.ca 2k)

Girshick et al, “Rich feature hierarchies for accurate object detection and semantic segmentation”, CVPR 2014. Figure copyright Ross Girshick, 2015. Reproduced with permission.
https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

R-CNN: CNN with Regions



Linear Regression for bounding box offset to correct the bounding boxes to contain only the object (RoIs can contain background or not the entire obj)

Classify regions with Support Vector Machines

Forward each region through ConvNet

Warped image regions:
Resize the region to the input size of the CNN

Regions of Interest (RoI) (c.ca 2k)

Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR 2014. Figure copyright Ross Girshick, 2015. Reproduced with permission.
https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

R-CNN: Problems

Ad hoc training objectives

- Fine tune network with softmax classifier (log loss)
- Train post hoc linear SVMs (hinge loss)
- Train post hoc bounding box regressions (least squares)

Training is slow (84h), takes a lot of disk space

Inference (detection) is slow

- 47s / image with VGG16 [Simonyan & Zisserman. ICLR15]
- Fixed by SPP-net [He et al. ECCV14]

~2000 ConvNet passes per image

AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

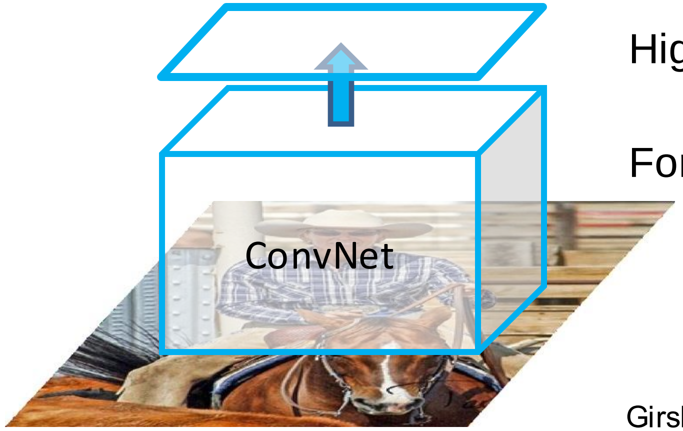
SPP-Net: Fast R-CNN

Input image

Girshick, “Fast R-CNN”, ICCV 2015. Figure copyright Ross Girshick, 2015. Reproduced with permission. https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

SPP-Net: Fast R-CNN



High resolution convolution feature map of image

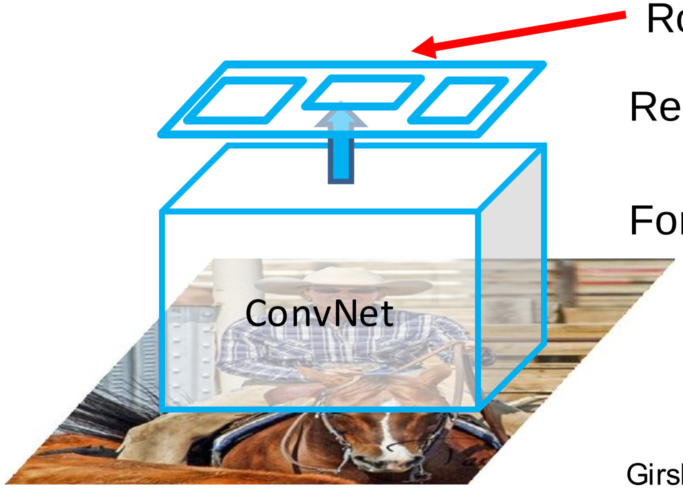
Forward whole image through ConvNet

Input image

Girshick, “Fast R-CNN”, ICCV 2015. Figure copyright Ross Girshick, 2015. Reproduced with permission. https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

SPP-Net: Fast R-CNN

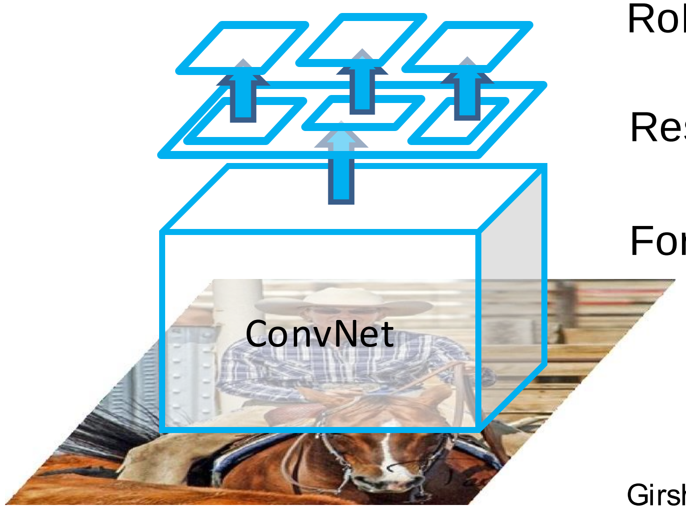


Input image

Girshick, “Fast R-CNN”, ICCV 2015. Figure copyright Ross Girshick, 2015. Reproduced with permission. https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

SPP-Net: Fast R-CNN



RoI Pooling or Spatial Pyramid Pooling

Resulting feature map of image

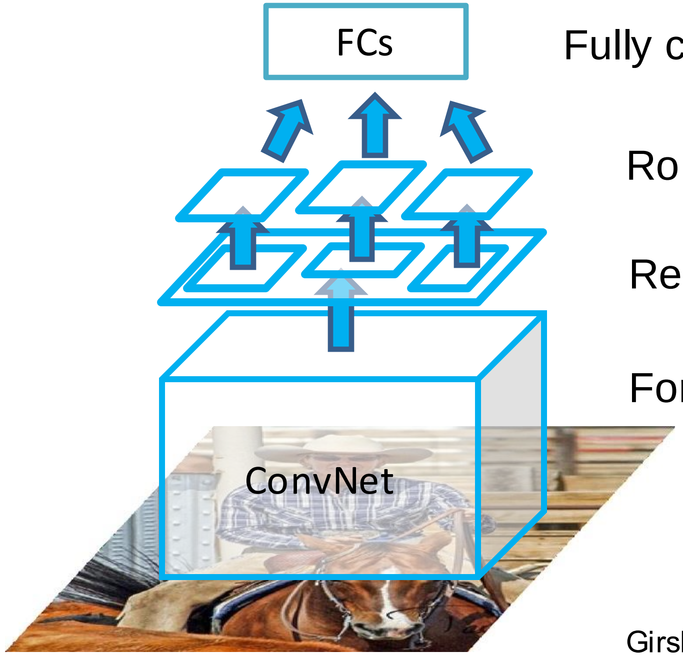
Forward whole image through ConvNet

Input image

Girshick, “Fast R-CNN”, ICCV 2015. Figure copyright Ross Girshick, 2015. Reproduced with permission. https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

SPP-Net: Fast R-CNN



Fully connected layers

RoI Pooling or Spatial Pyramid Pooling

Resulting feature map of image

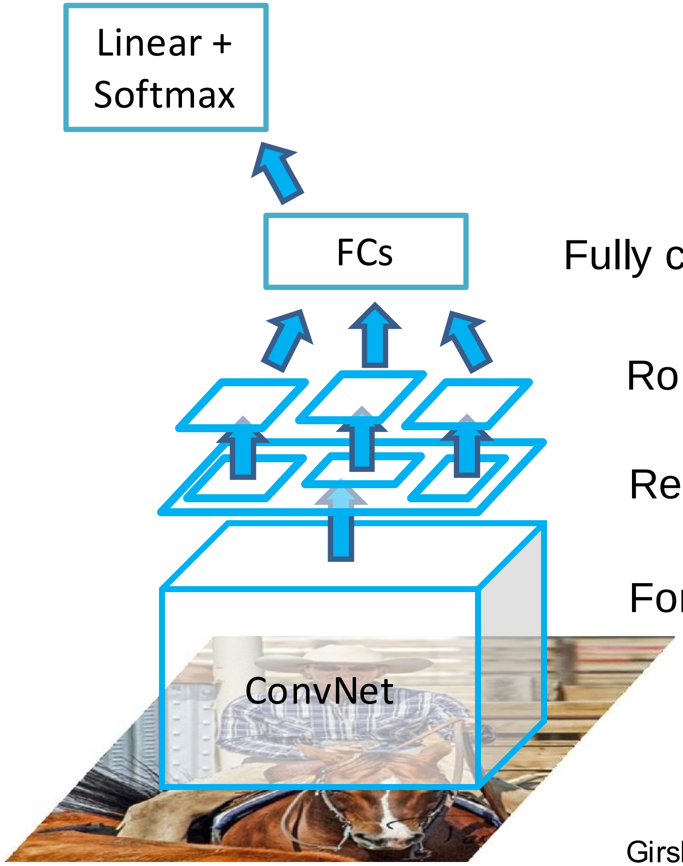
Forward whole image through ConvNet

Input image

Girshick, "Fast R-CNN", ICCV 2015. Figure copyright Ross Girshick, 2015. Reproduced with permission. https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

SPP-Net: Fast R-CNN



Input image

Softmax classifier for the objects

Fully connected layers

RoI Pooling or Spatial Pyramid Pooling

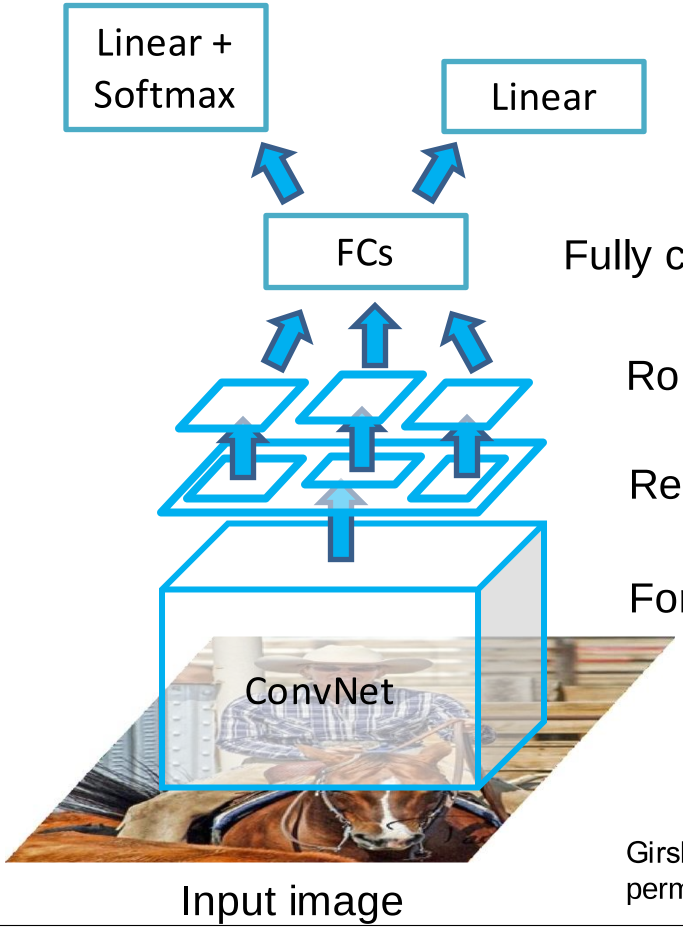
Resulting feature map of image

Forward whole image through ConvNet

Girshick, “Fast R-CNN”, ICCV 2015. Figure copyright Ross Girshick, 2015. Reproduced with permission. https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

SPP-Net: Fast R-CNN



Softmax classifier for the objects +
Linear regression for adjusting bounding boxes

Fully connected layers

RoI Pooling or Spatial Pyramid Pooling

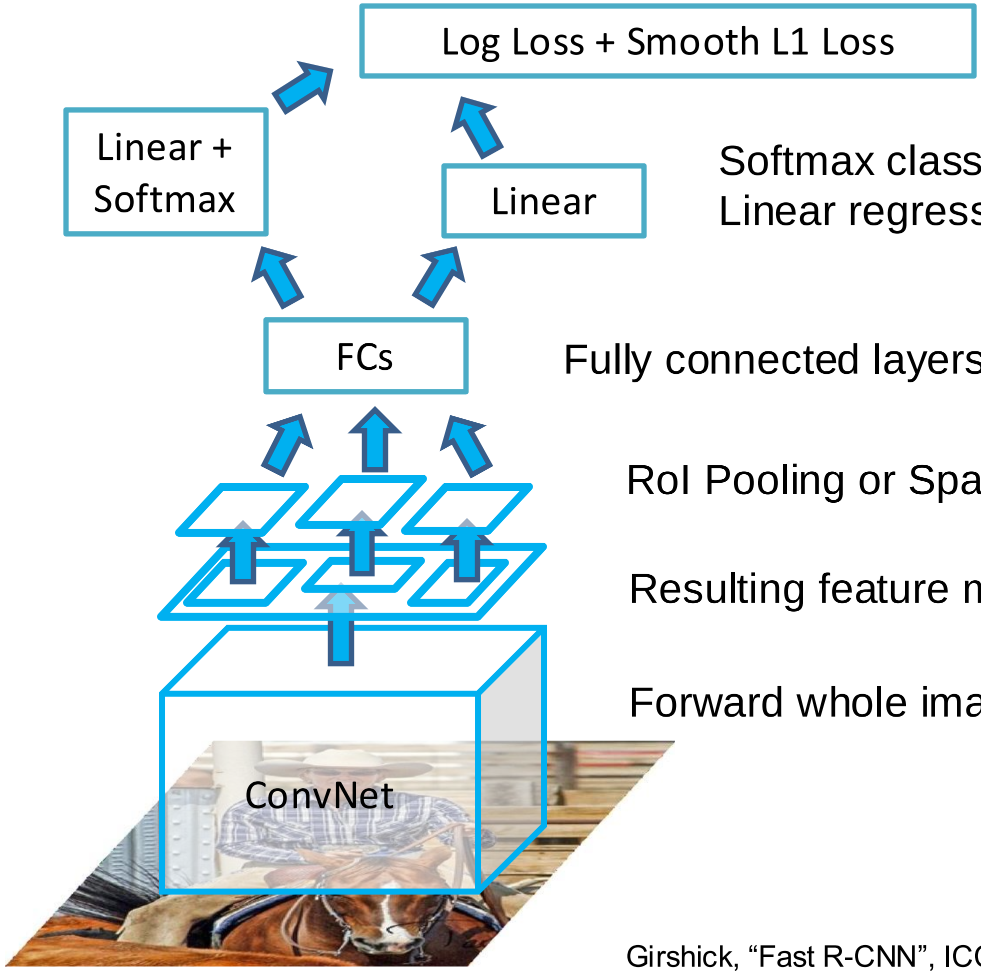
Resulting feature map of image

Forward whole image through ConvNet

Girshick, “Fast R-CNN”, ICCV 2015. Figure copyright Ross Girshick, 2015. Reproduced with permission. https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

SPP-Net: Fast R-CNN (Training)



Multi-task loss + Backprop

Softmax classifier for the objects +
Linear regression for adjusting bounding boxes

Fully connected layers

RoI Pooling or Spatial Pyramid Pooling

Resulting feature map of image

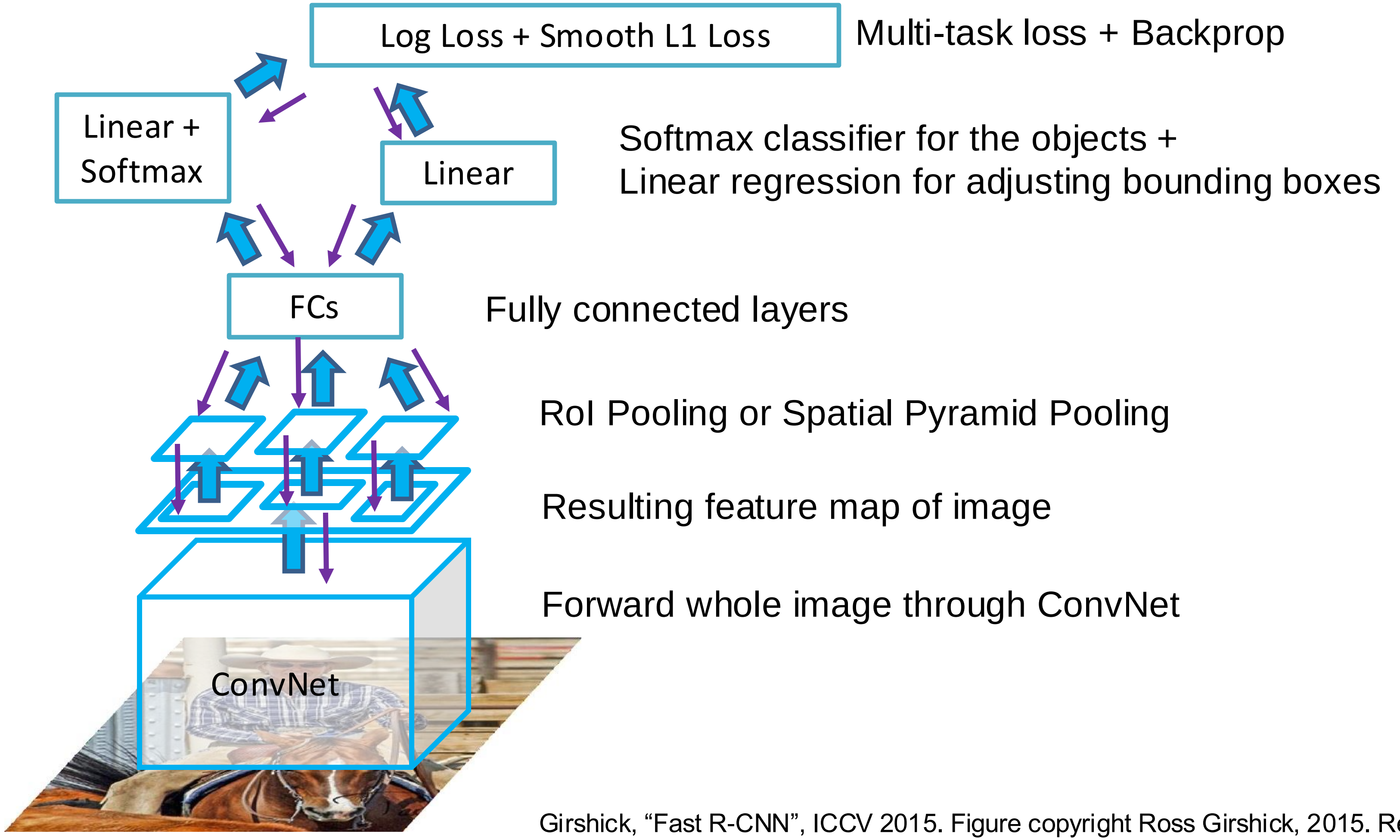
Forward whole image through ConvNet

Input image

Girshick, “Fast R-CNN”, ICCV 2015. Figure copyright Ross Girshick, 2015. Reproduced with permission. https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

SPP-Net: Fast R-CNN (Training)

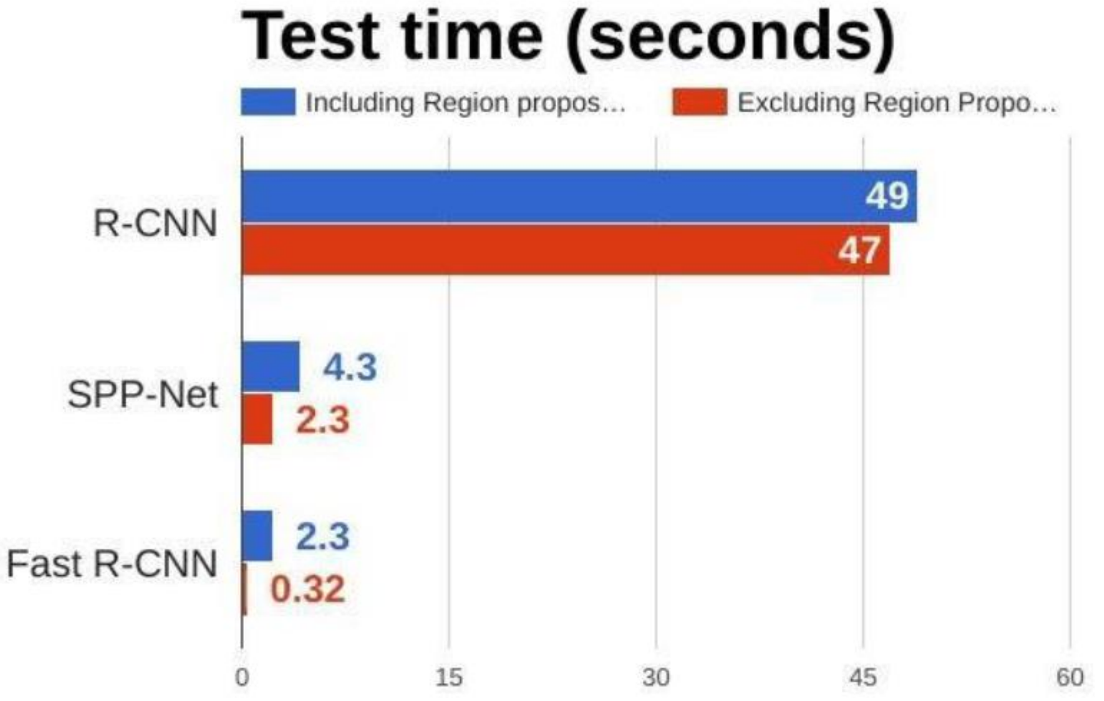


Input image

Girshick, “Fast R-CNN”, ICCV 2015. Figure copyright Ross Girshick, 2015. Reproduced with permission. https://dl.dropboxusercontent.com/s/vlyrkgd8nz8gy5l/fast-rcnn.pdf?dl=0

AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

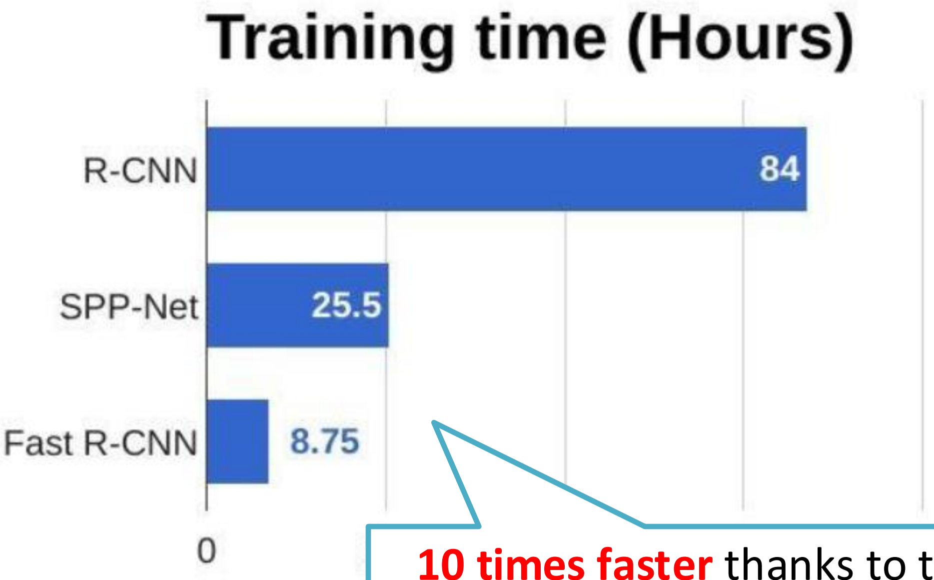
R-CNN vs SPP vs Fast R-CNN



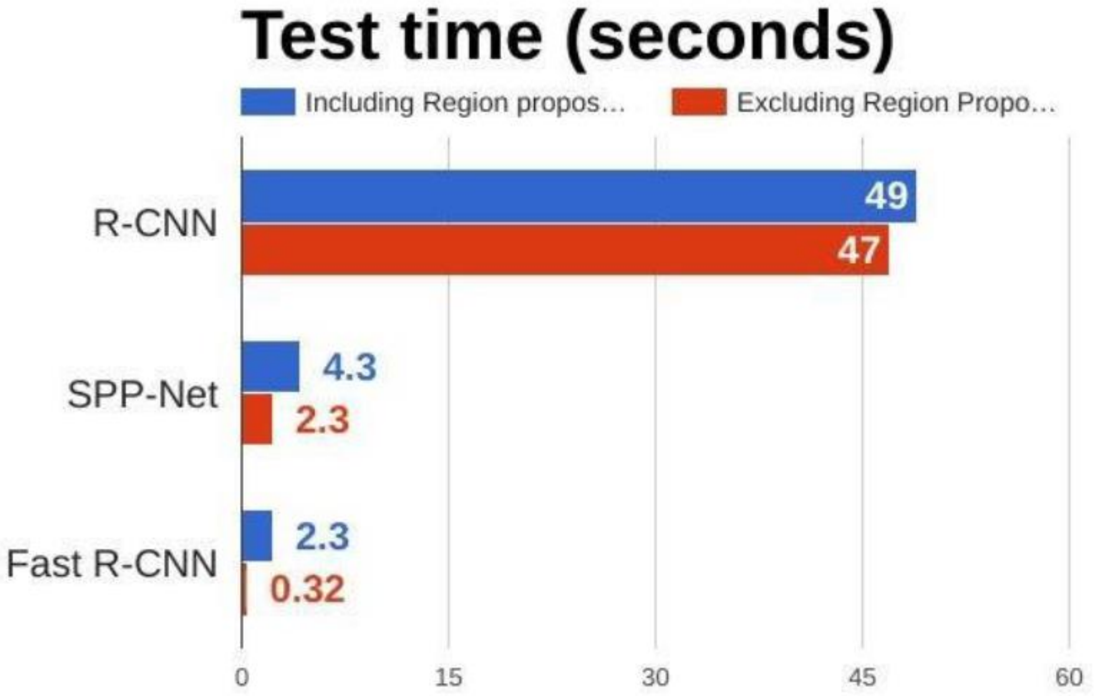
Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR 2014.
He et al, "Spatial pyramid pooling in deep convolutional networks for visual recognition", ECCV 2014
Girshick, "Fast R-CNN", ICCV 2015

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

R-CNN vs SPP vs Fast R-CNN



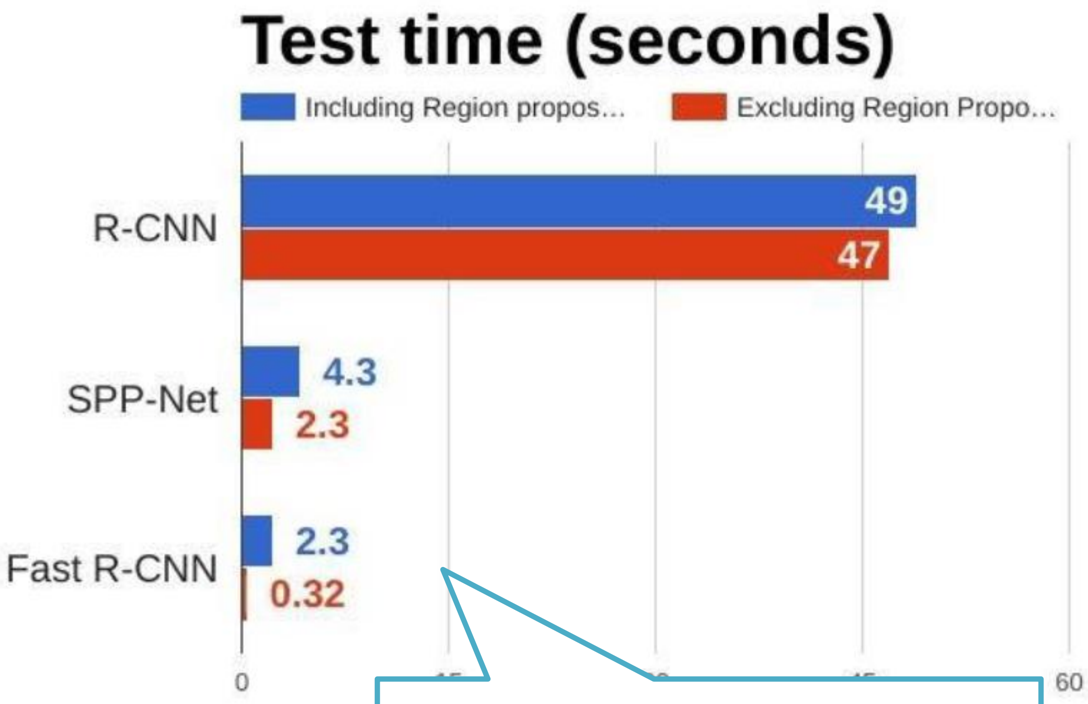
10 times faster thanks to the sharing of computations between feature masks



Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR 2014.
He et al, "Spatial pyramid pooling in deep convolutional networks for visual recognition", ECCV 2014
Girshick, "Fast R-CNN", ICCV 2015

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

R-CNN vs SPP vs Fast R-CNN



Problem
Runtime dominated
by computing region proposals!
Computing RoIs takes about 2s

Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR 2014.
He et al, "Spatial pyramid pooling in deep convolutional networks for visual recognition", ECCV 2014
Girshick, "Fast R-CNN", ICCV 2015

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Faster R-CNN

Developed to solve the bottleneck given by computing region proposals → make CNN do proposals
Insert **Region Proposal Network** (**RPN**) to predict proposals from features

Jointly train with 4 losses:
1. RPN classify object / not object
2. RPN regress box coordinates
3. Final classification score (object classes)
4. Final box coordinates

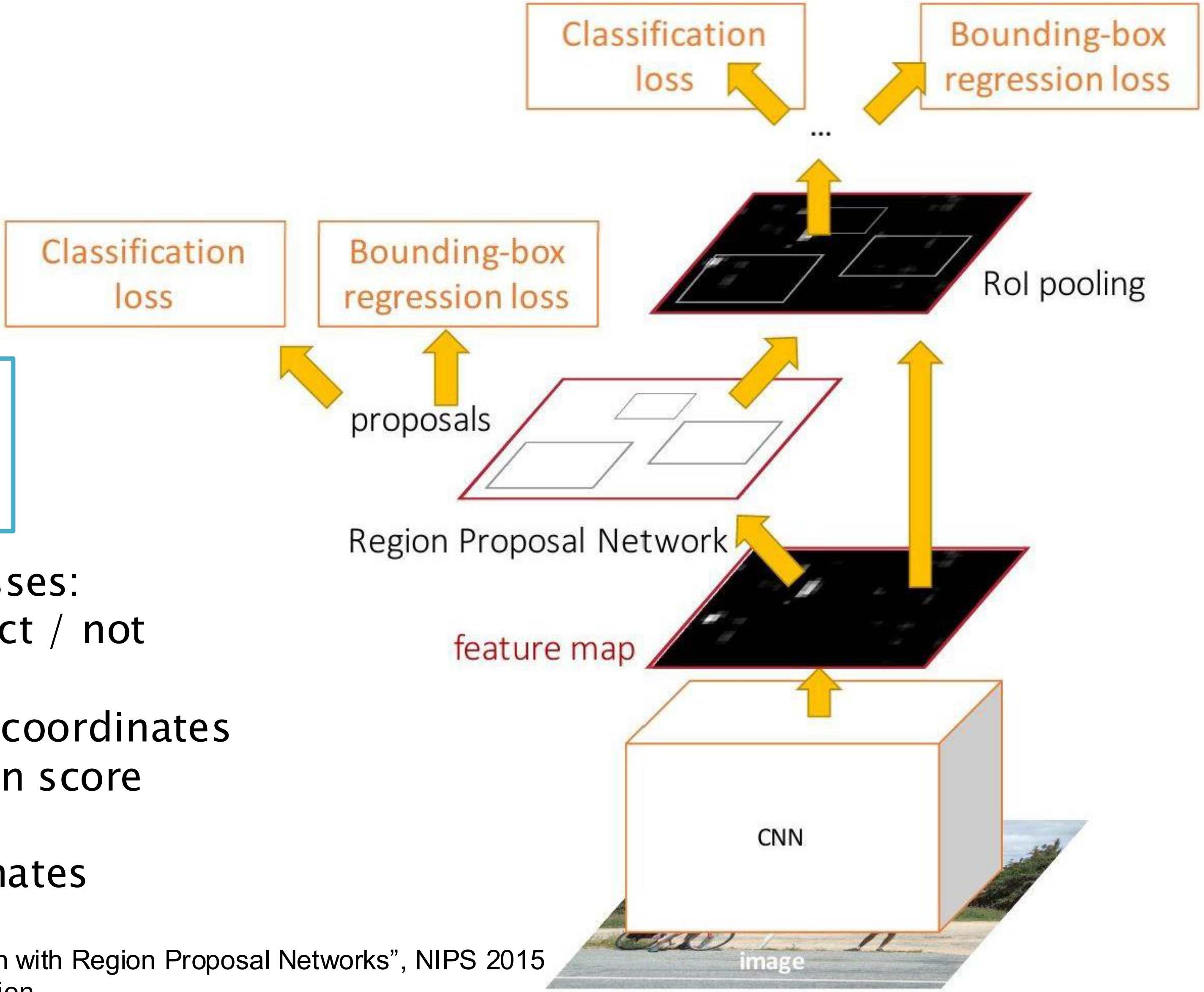
AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

Faster R-CNN

Make the net itself predict its own region proposals

Jointly train with 4 losses:

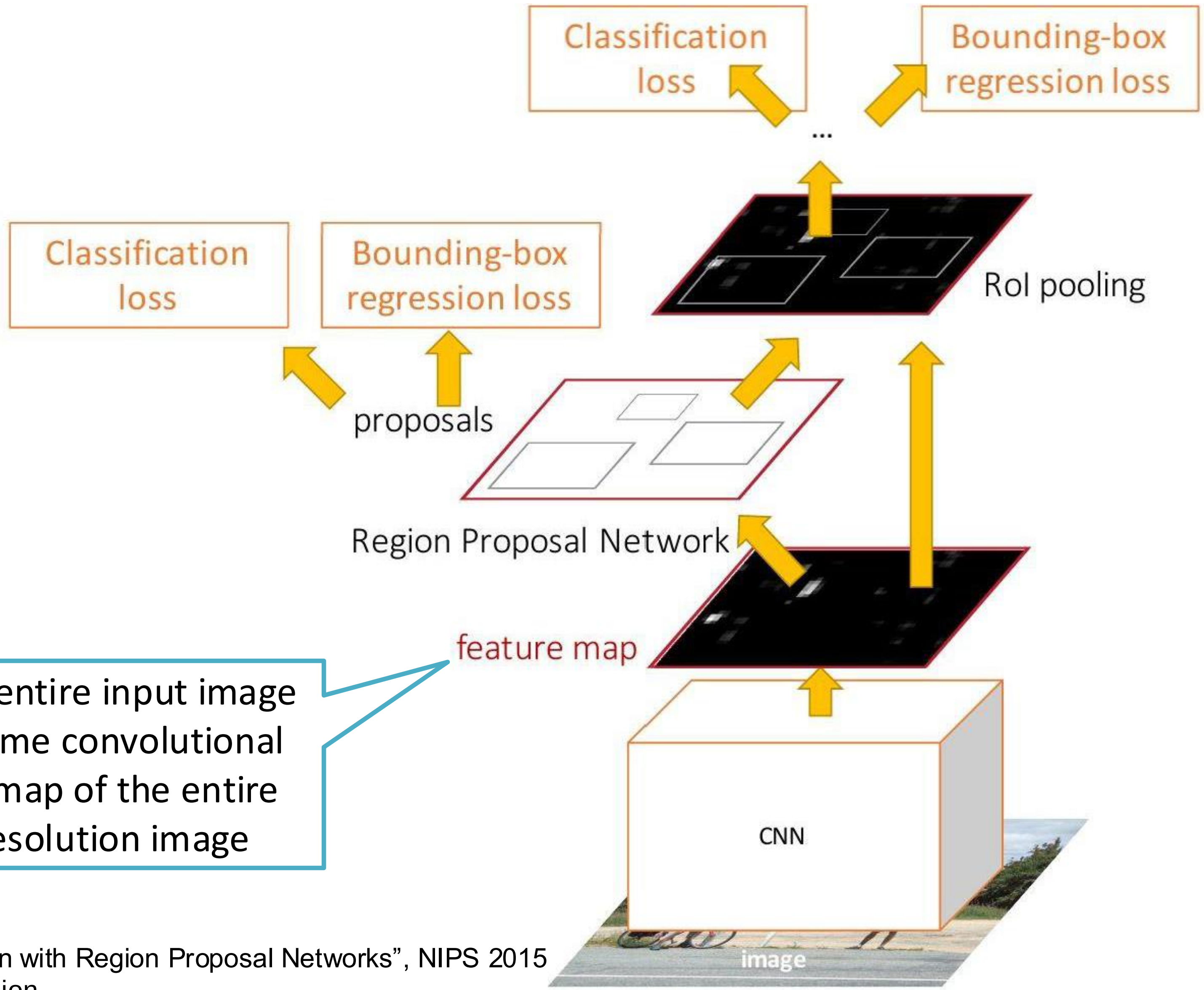
1. RPN classify object / not object
2. RPN regress box coordinates
3. Final classification score (object classes)
4. Final box coordinates



Ren et al, “Faster R-CNN: Towards Real Time Object Detection with Region Proposal Networks”, NIPS 2015

AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

Faster R-CNN

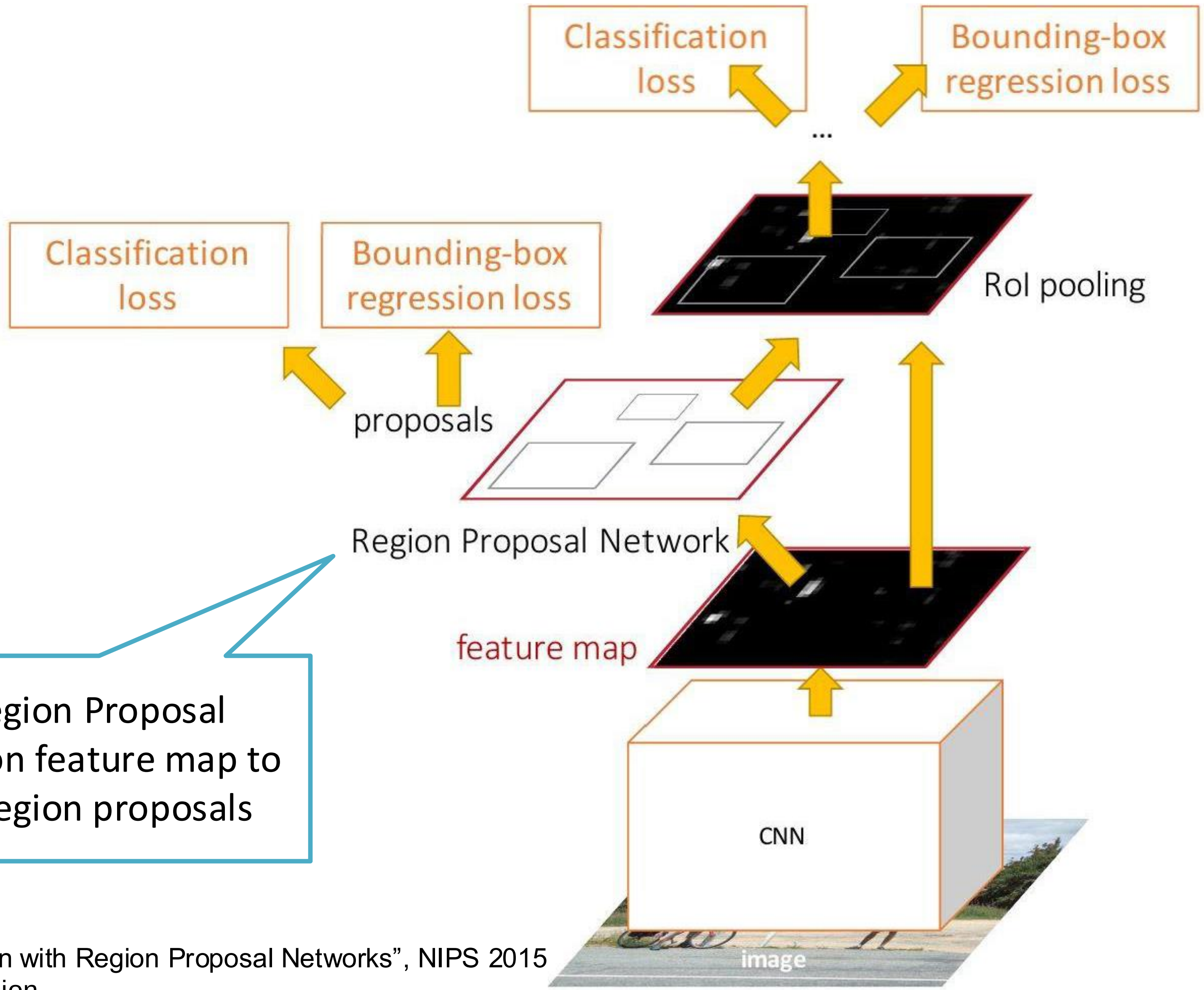


Take the entire input image to get some convolutional feature map of the entire high-resolution image

Ren et al, “Faster R-CNN: Towards Real Time Object Detection with Region Proposal Networks”, NIPS 2015

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Faster R-CNN

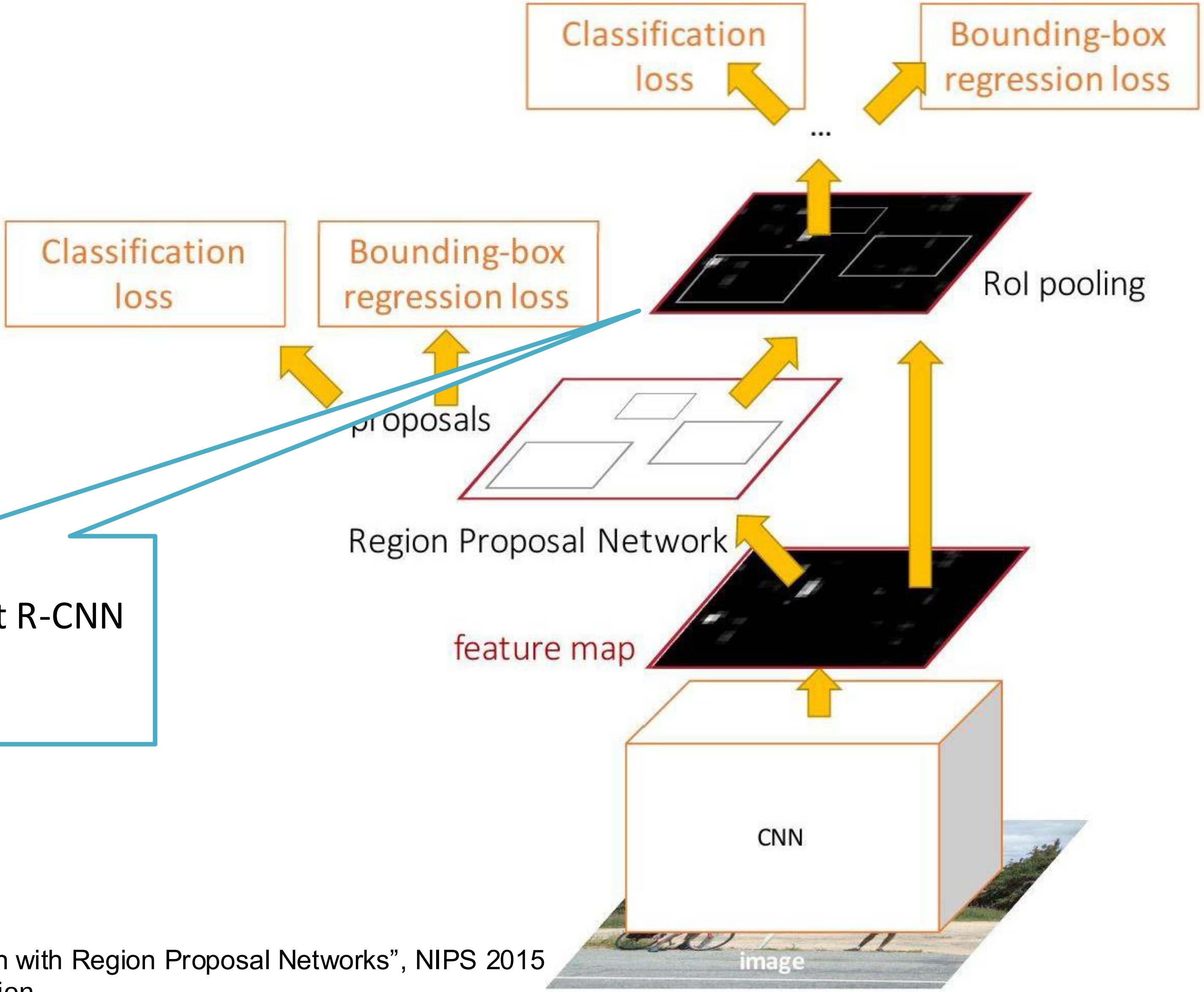


Run Region Proposal Network on feature map to gather region proposals

Ren et al, "Faster R-CNN: Towards Real Time Object Detection with Region Proposal Networks", NIPS 2015

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Faster R-CNN

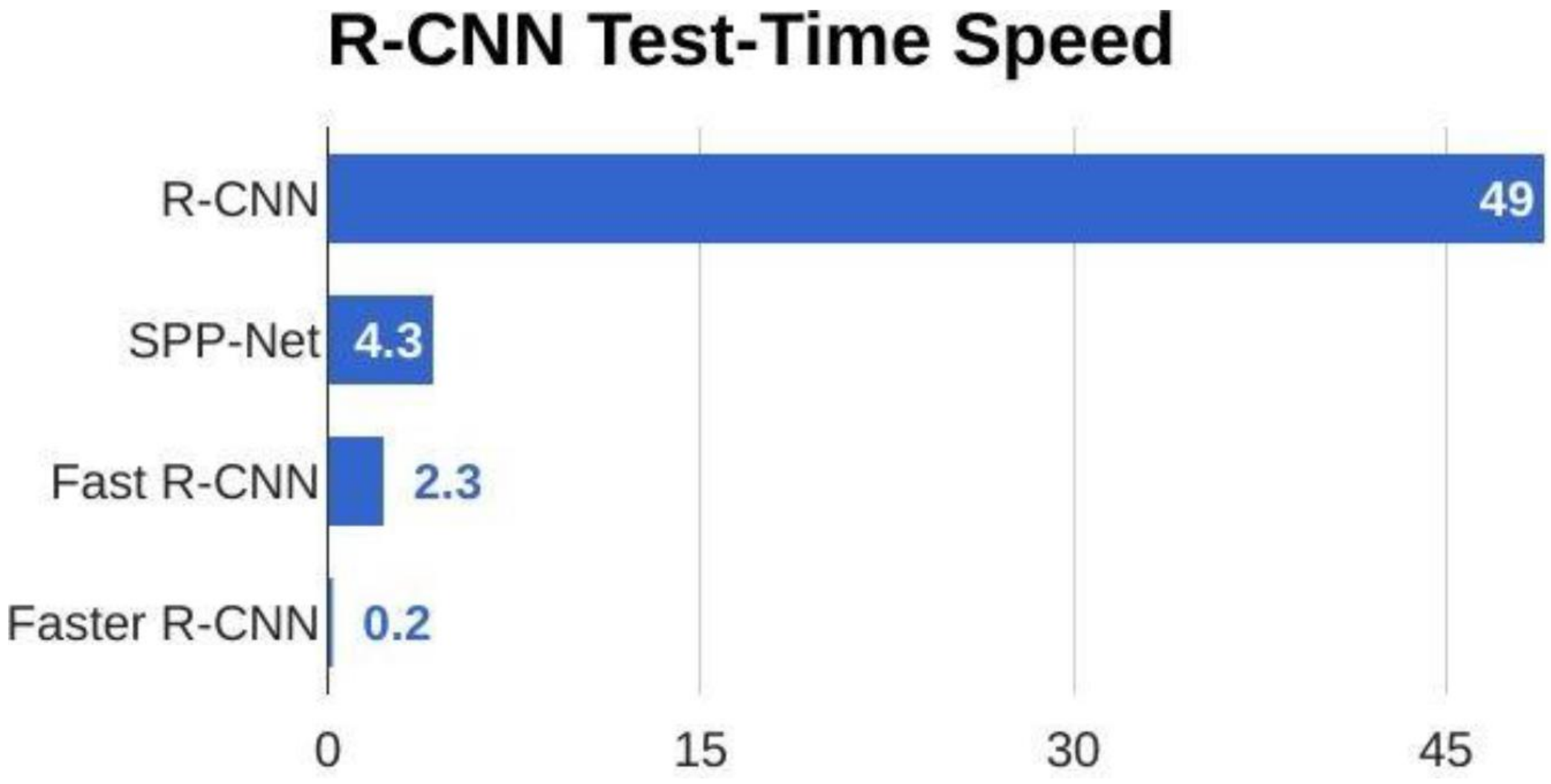


Finally, classical Fast R-CNN is used.

Ren et al, "Faster R-CNN: Towards Real Time Object Detection with Region Proposal Networks", NIPS 2015
Figure copyright 2015, Ross Girshick; reproduced with permission

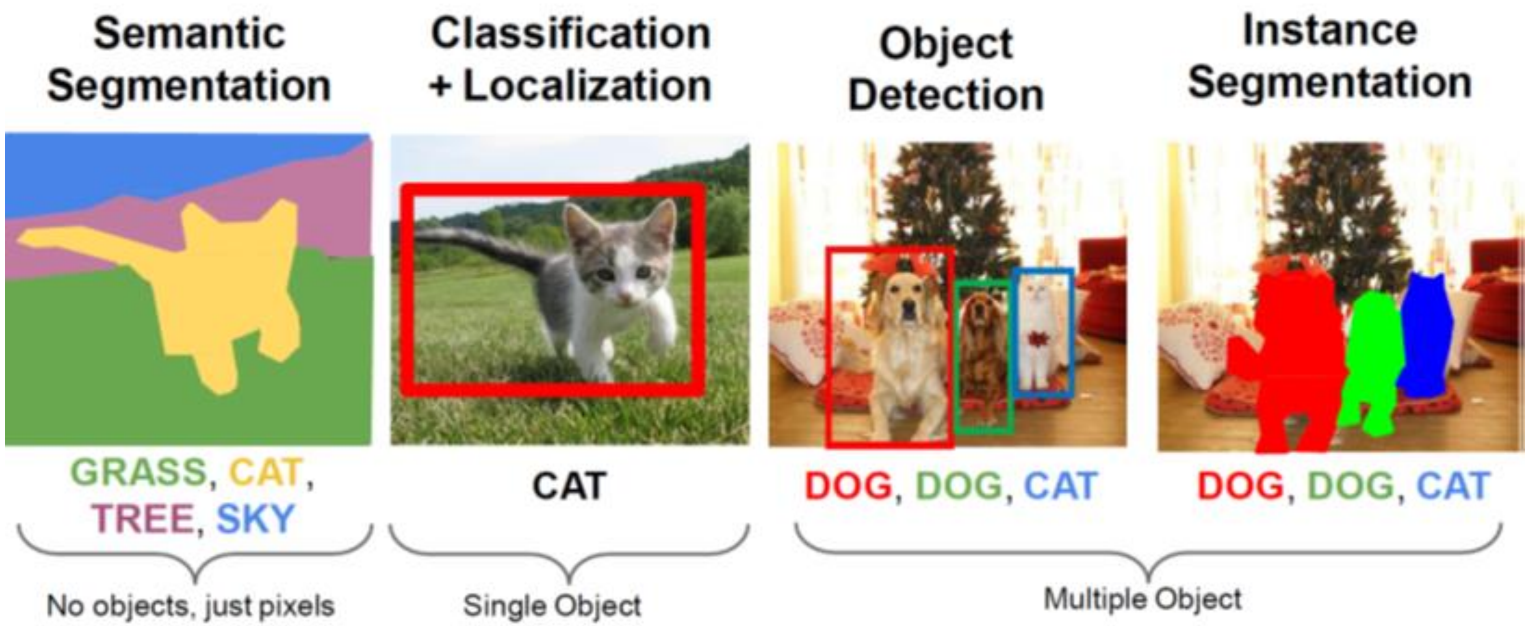
AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

R-CNN Test-Time Speed

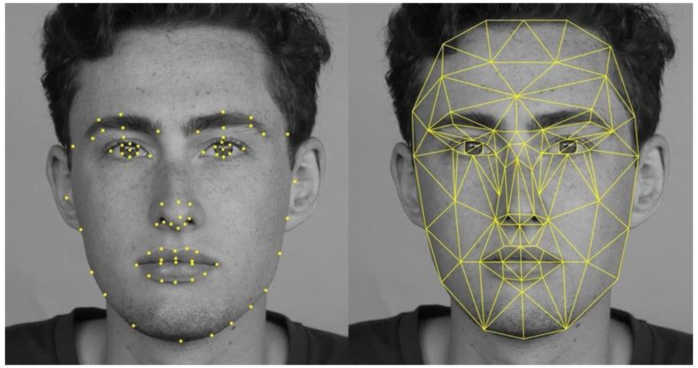


AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Other Computer Vision Tasks



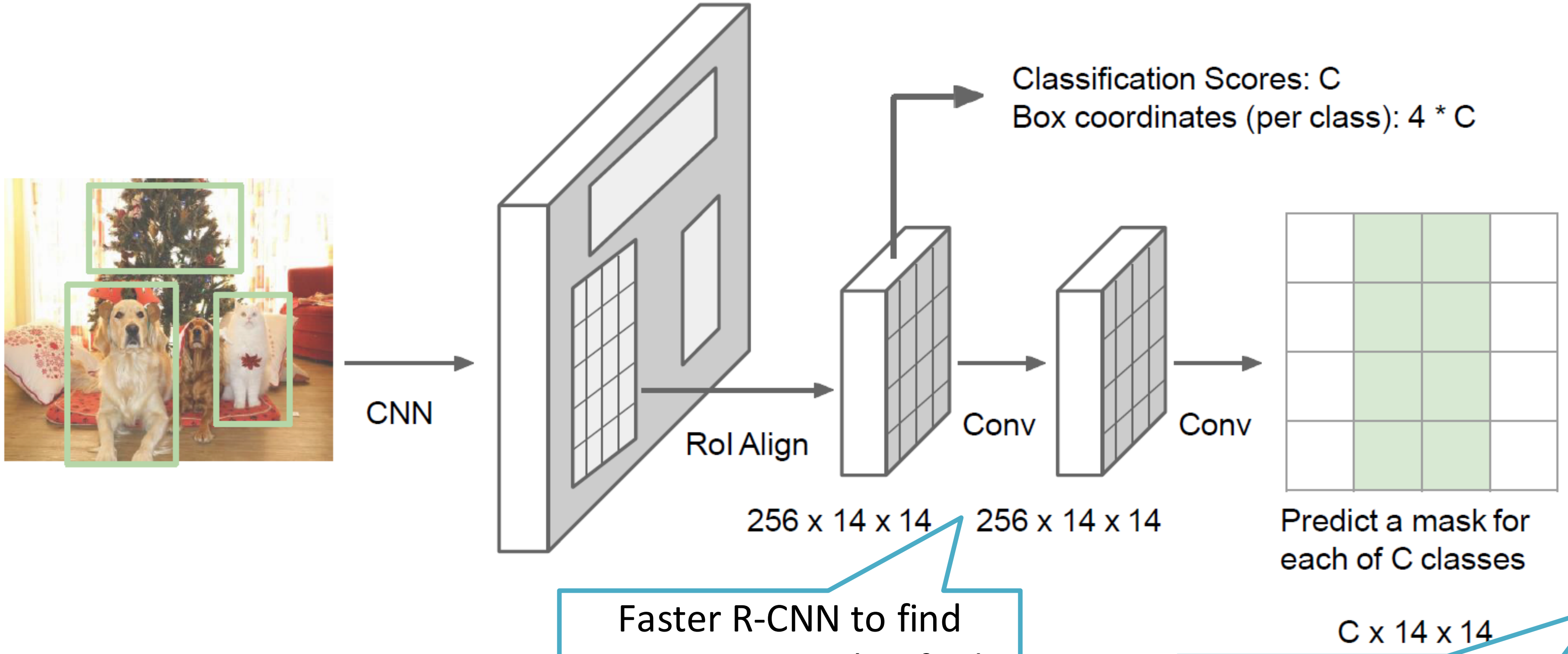
Keypoint detection



https://pixabay.com/photos/pets-christmas-dogs-cat-962215/ - https://pixabay.com/p-1246693/ - CC0 public domain (https://creativecommons.org/publicdomain/zero/1.0/deed.en)

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Mask R-CNN



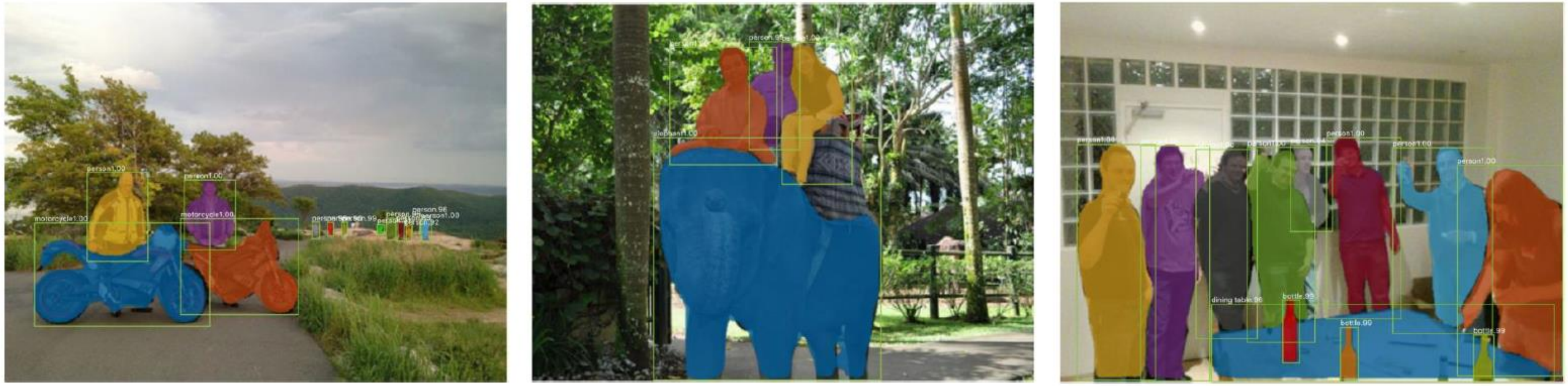
Faster R-CNN to find region proposal to find classes and bounding boxes

Check each pixel if it is part of an object or not.

He et al, “Mask R-CNN”, arXiv 2017

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Mask R-CNN: Very Good Results!

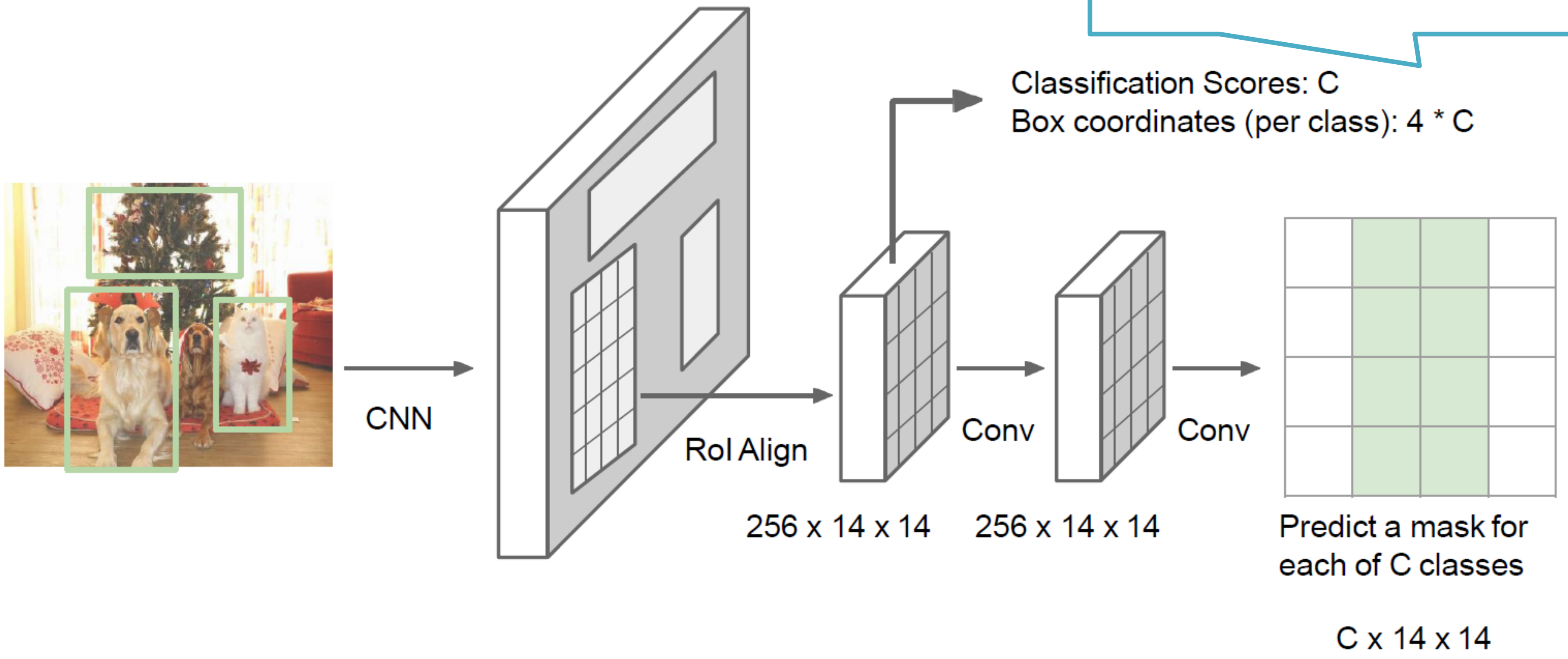


He et al, “Mask R-CNN”, arXiv 2017
Figures copyright Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, 2017.

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Mask R-CNN

Mask R-CNN can also do pose estimation by returning **joint coordinates**



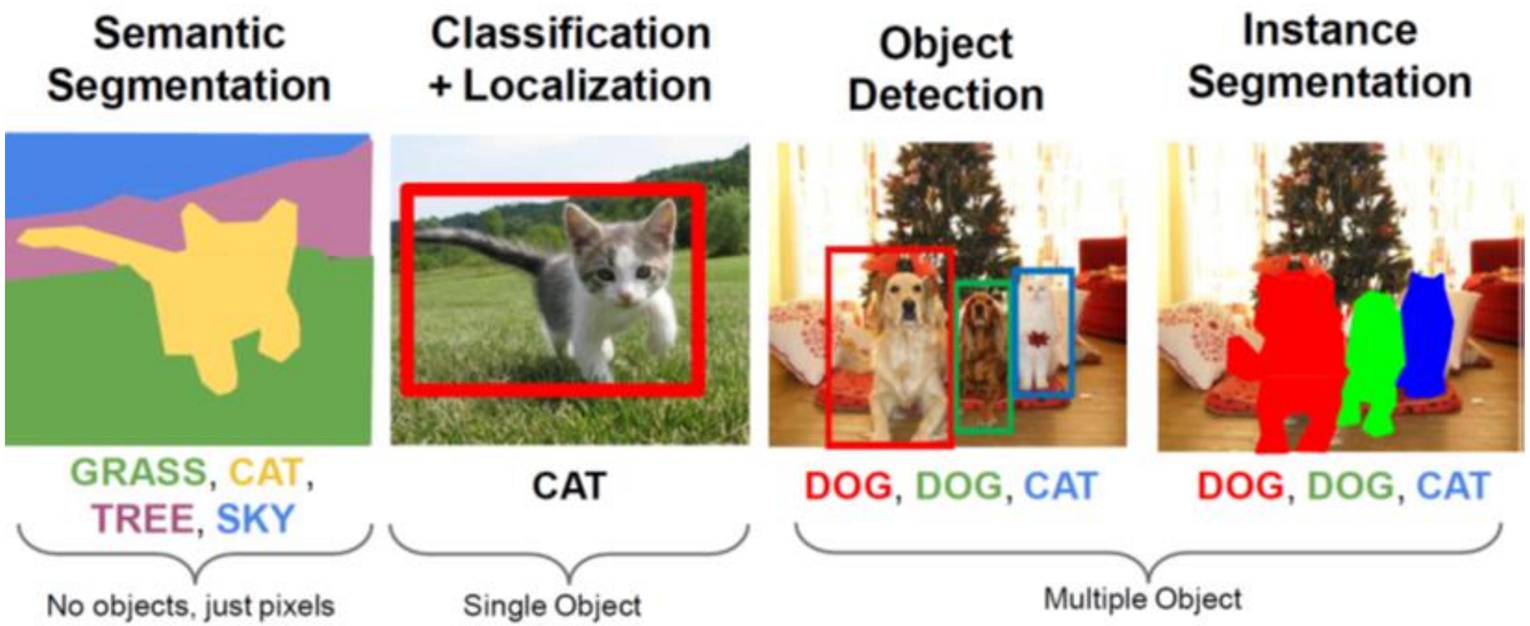
He et al, “Mask R-CNN”, arXiv 2017

AIDA4Edge UK Research and Innovation Funded by the European Union University of Ferrara

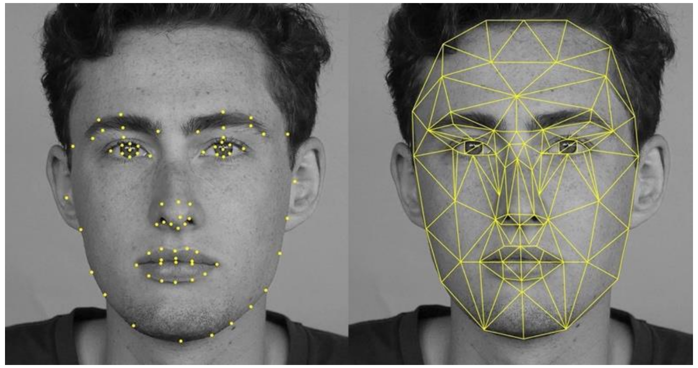
Mask R-CNN and pose estimation

He et al, "Mask R-CNN", arXiv 2017

Other Computer Vision Tasks



Keypoint detection



https://pixabay.com/photos/pets-christmas-dogs-cat-962215/ - https://pixabay.com/p-1246693/ - CC0 public domain (https://creativecommons.org/publicdomain/zero/1.0/deed.en)

Keypoint Detection

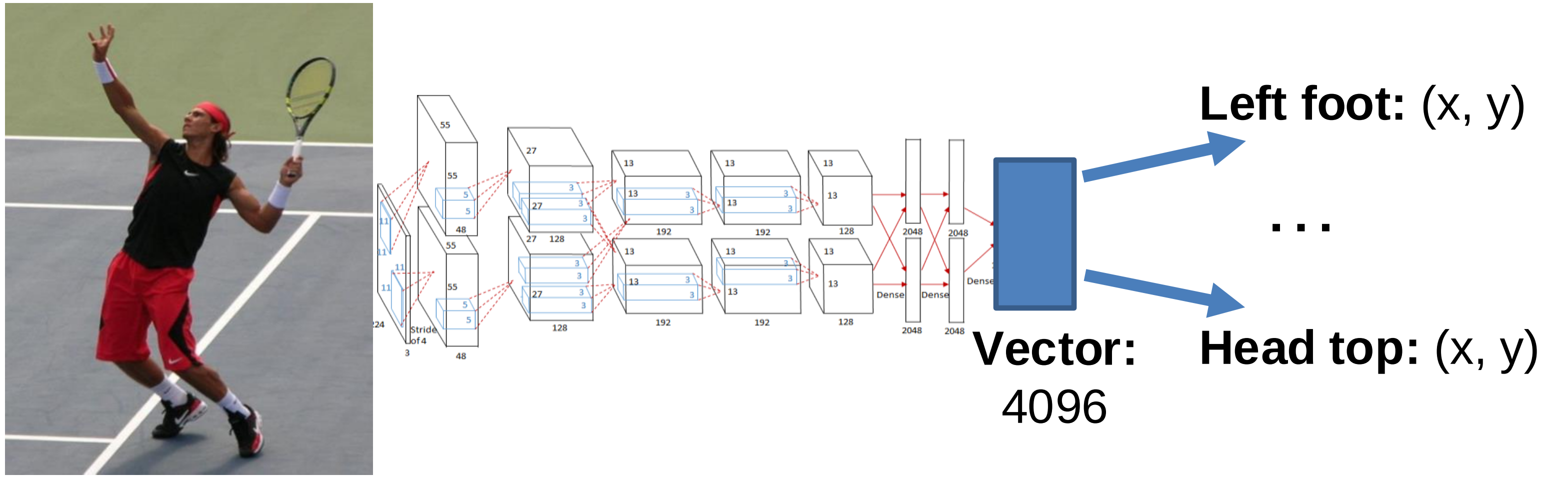
Represent pose as a set of 14 joint positions:

Left / right foot
Left / right knee
Left / right hip
Left / right shoulder
Left / right elbow
Left / right hand
Neck
Head top

Johnson and Everingham, “Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation”BMVC2010

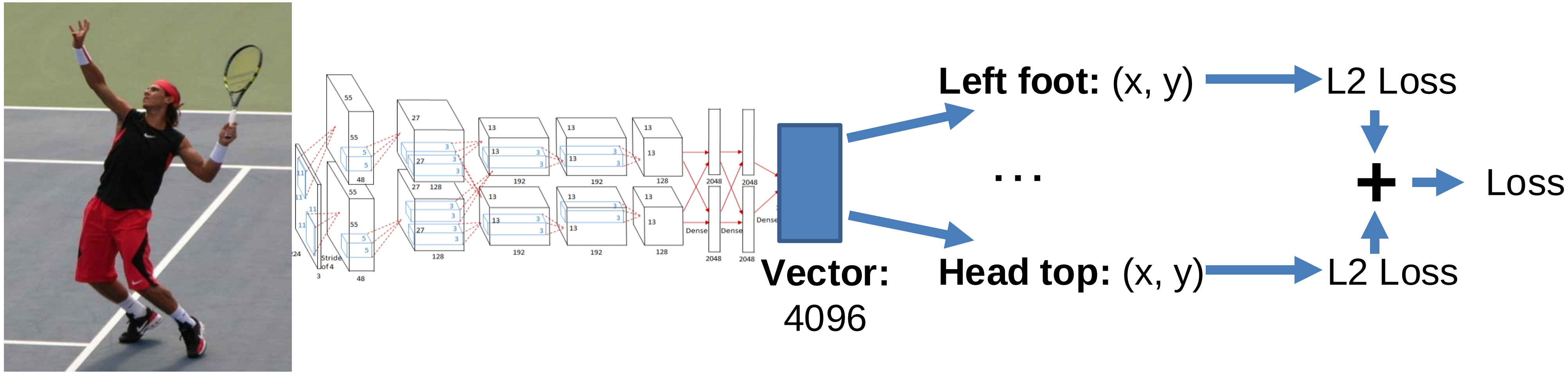
AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara

Classification + Localization



Toshev and Szegedy, “DeepPose: Human Pose Estimation via Deep Neural Networks”
CVPR 2014

Classification + Localization



Toshev and Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks"
CVPR 2014

AIDA4Edge
UK Research and Innovation
Funded by the European Union
University of Ferrara