Convolutional Neural Networks in Computer Vision and Beyond

Machine Perception

Siyu Tang

02 April 2020

[Some slides adopted from Karpathy, De Freitas, Serre]

Advanced Interactive Technologies

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Last week(s)

Backpropagation

Convolutional Neural Networks (CNNs)

Recurrent Neural Networks (RNNs)

This Week

CNNs – case studies from computer vision

- Image Classification
- Object detection

Fully convolutional networks

- Semantic Segmentation

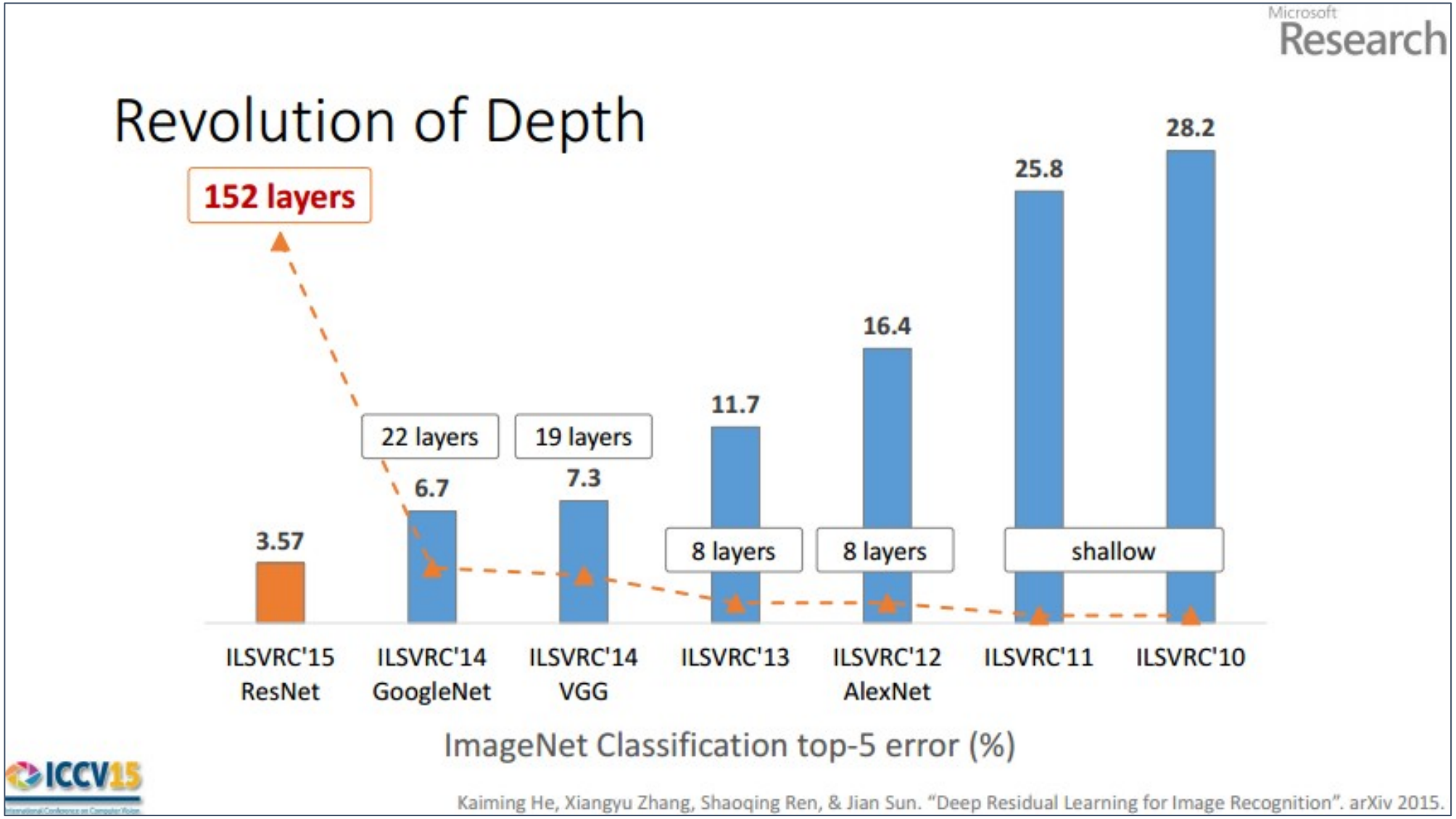
This Week

CNNs – case studies from computer vision

- Image Classification - ***different CNN architectures***
- Object detection - ***how to use CNNs as building blocks***

Fully convolutional networks

- Semantic Segmentation



(slide from Kaiming He's recent presentation)

Case Study: VGG – more depth, fewer parameters

Smaller filters

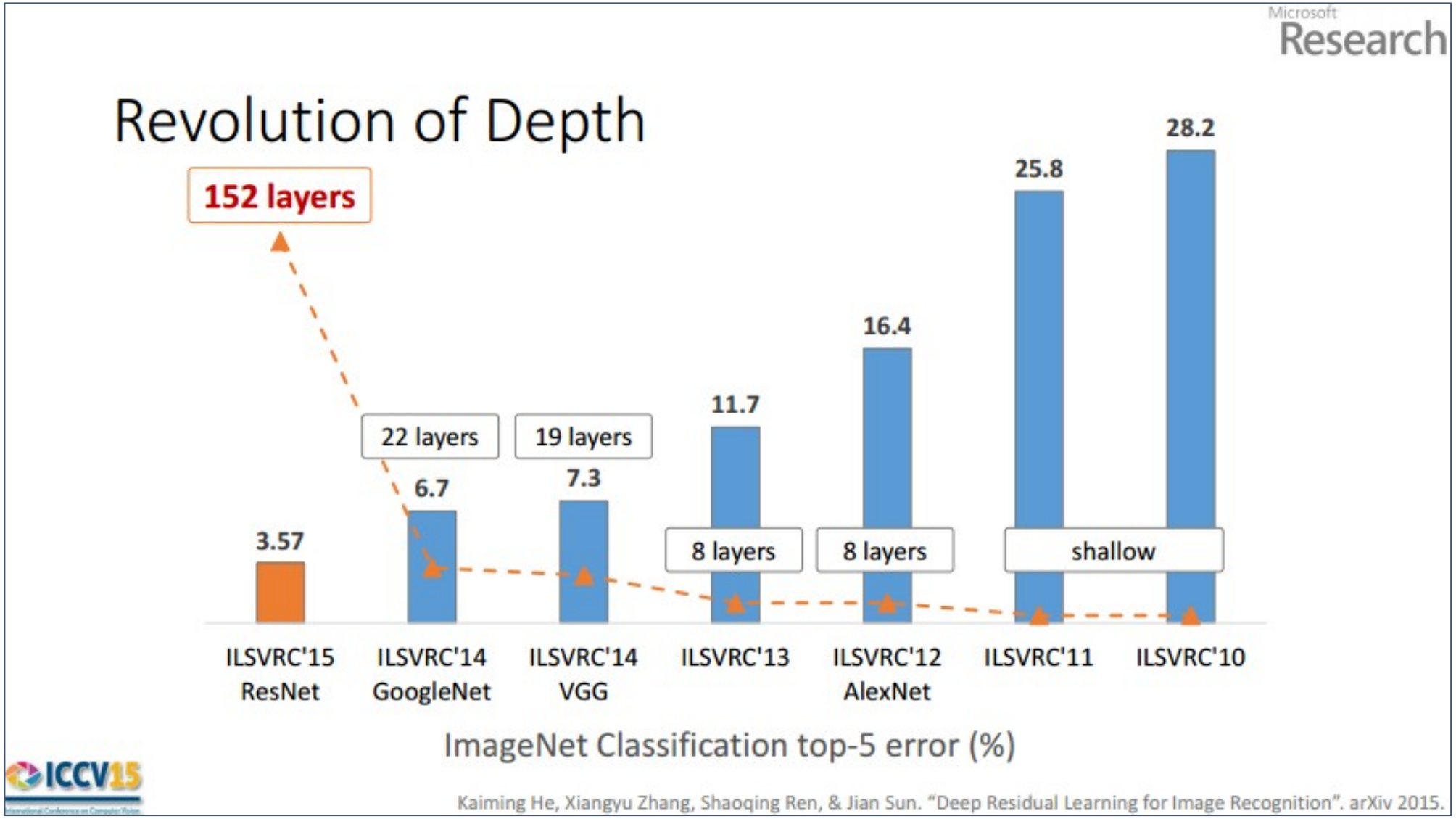
More layers

Due to depth:

- Large receptive field
- Despite smaller filters

Due to smaller filters:

- Fewer parameters



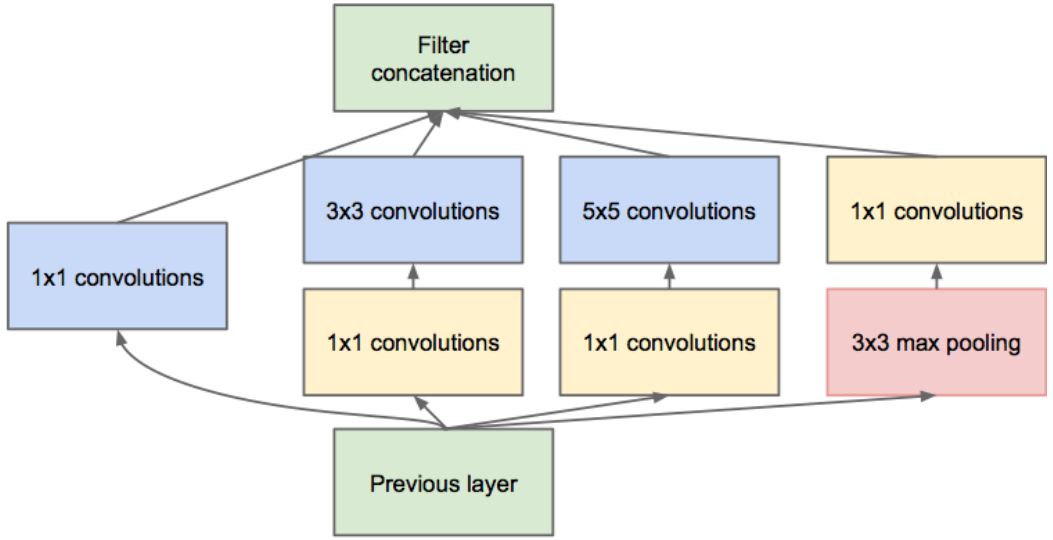
(slide from Kaiming He's recent presentation)

Case Study: GoogLeNet

[Szegedy et al., 2014]

Even deeper network with computational efficiency

- 22 layers
- Efficient “Inception” module
- Only 5 million parameters
  - 12x less than AlexNet
  - ILSVRC 2014 winner (6.7% top 5 error)



Inception module

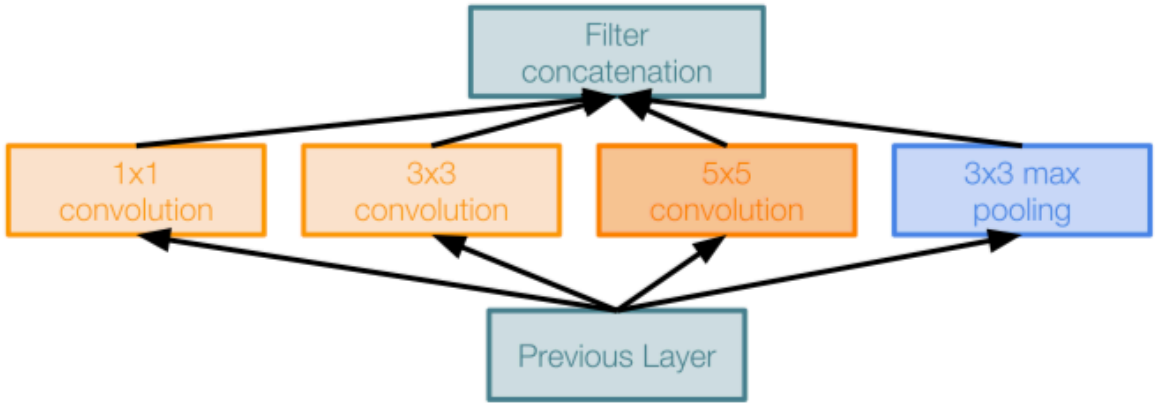
[Slides adopted from: Li, Karpathy]

Computational aspects

- 1x1 convolutions do not aggregate information over space but only over different channels

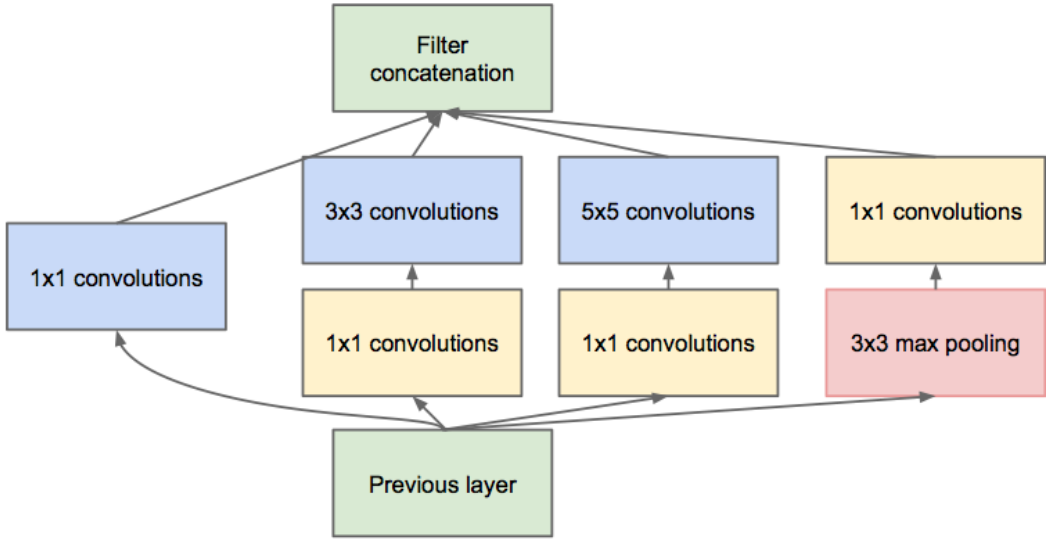
就是说1*1的感受野只有一个pixel，因此它无法聚合信息即看不到更多的视野，但是它可以看到不同的通道

Naïve Inception



Continuous increase in dimensionality

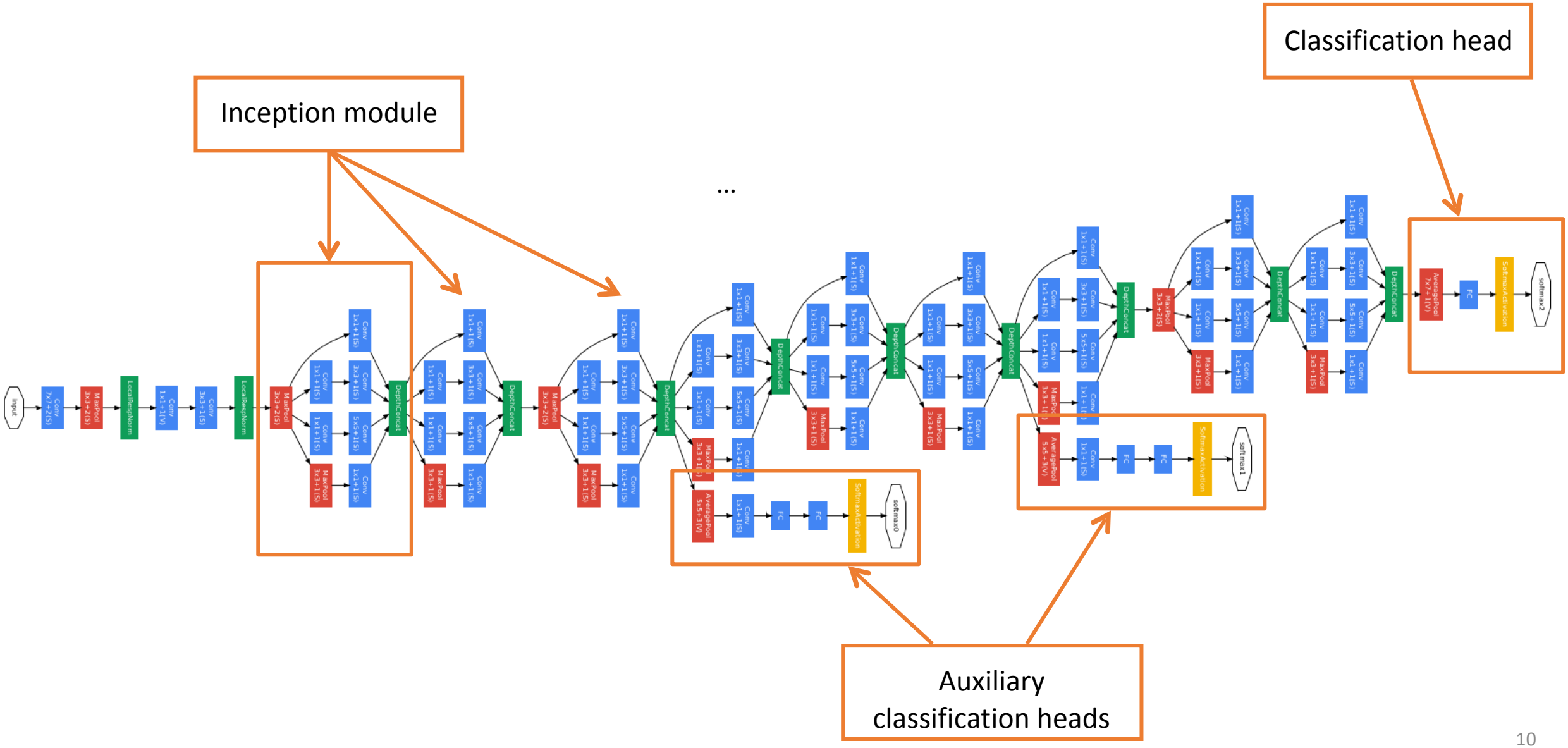
Final Inception module

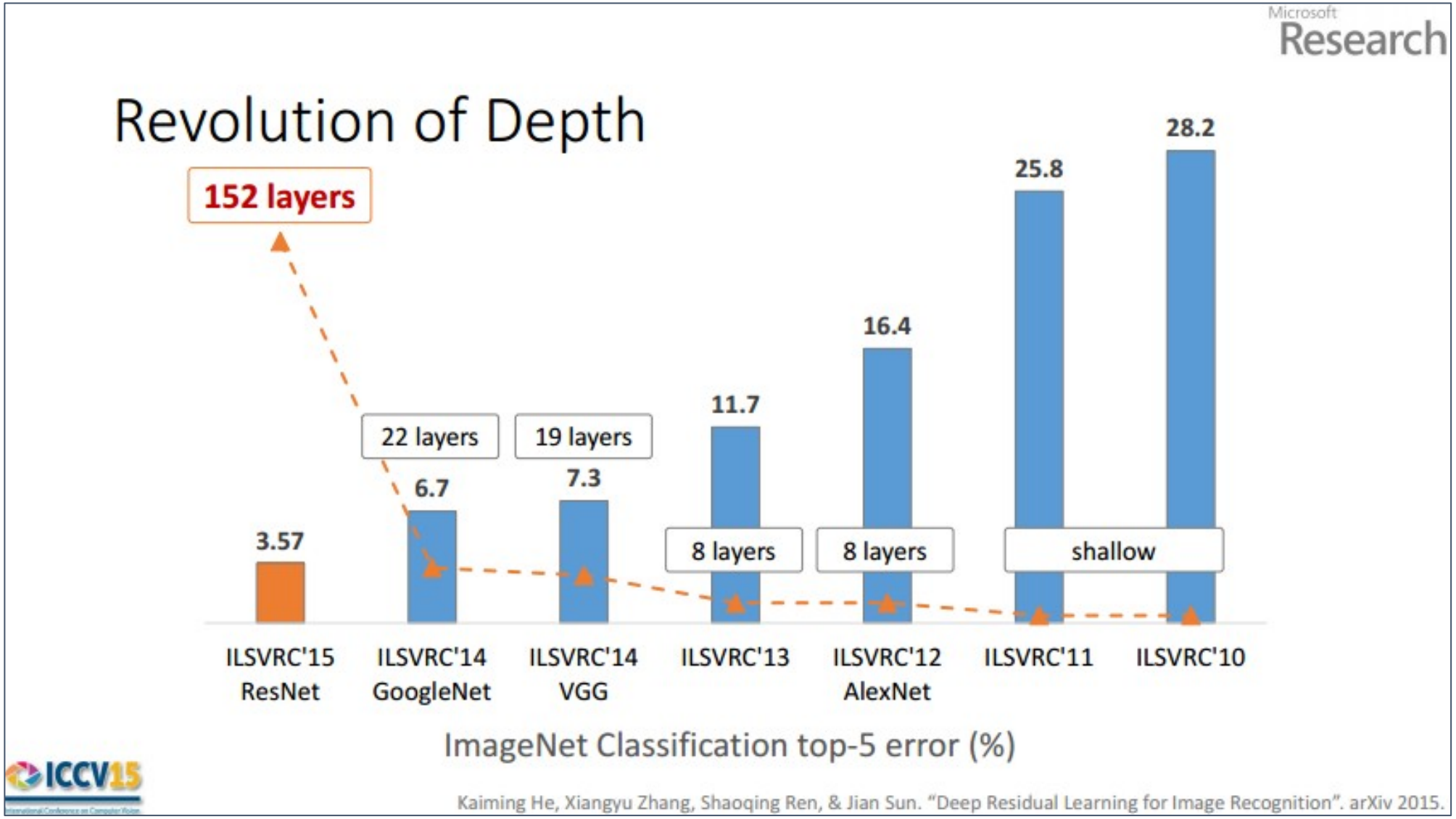


Dimensionality reduction via 1x1 convolutions

Case Study: GoogLeNet

[Szegedy et al., 2014]



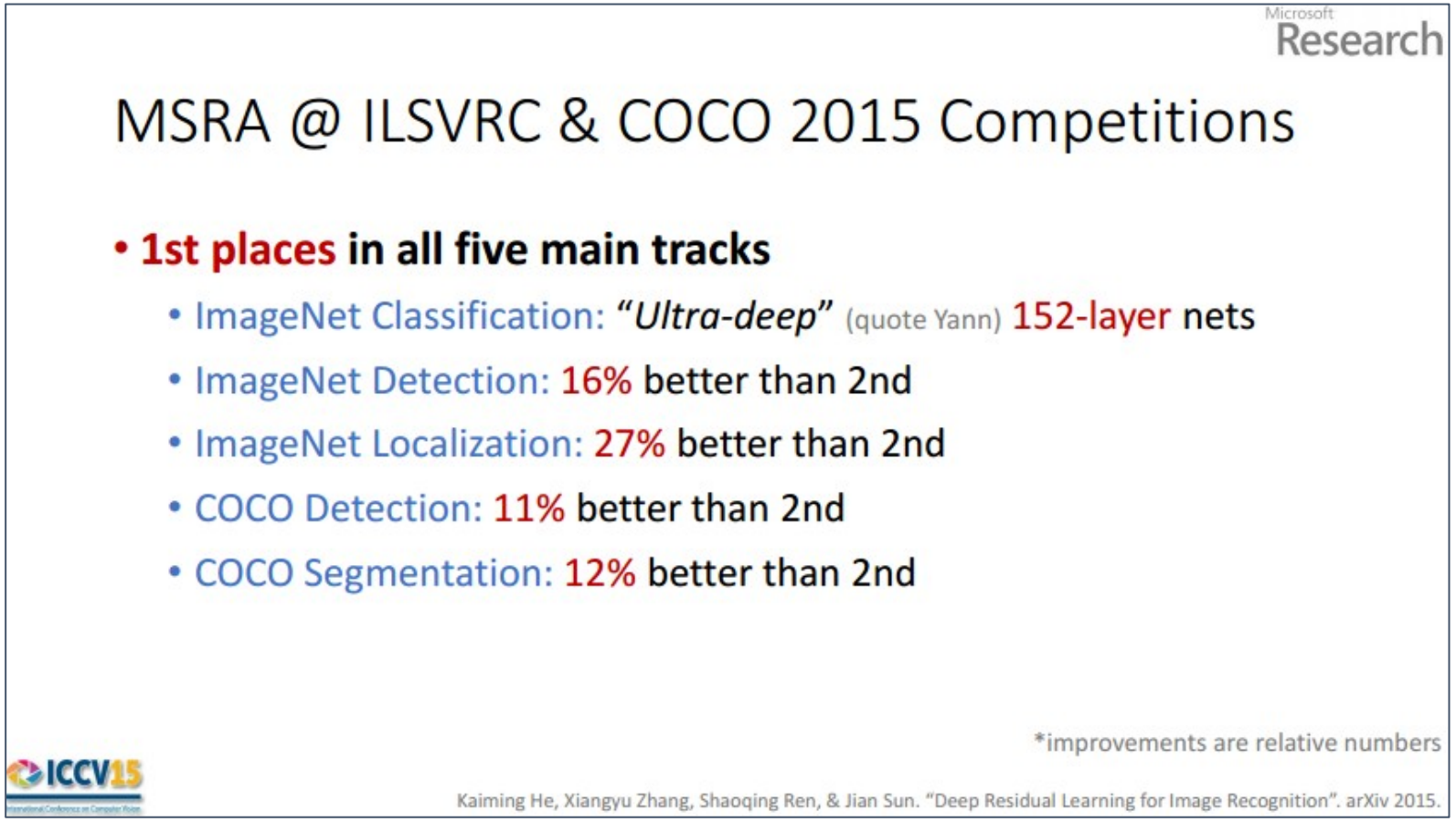


(slide from Kaiming He's recent presentation)

Case Study: ResNet

[He et al., 2015]

ILSVRC 2015 winner (3.6% top 5 error)

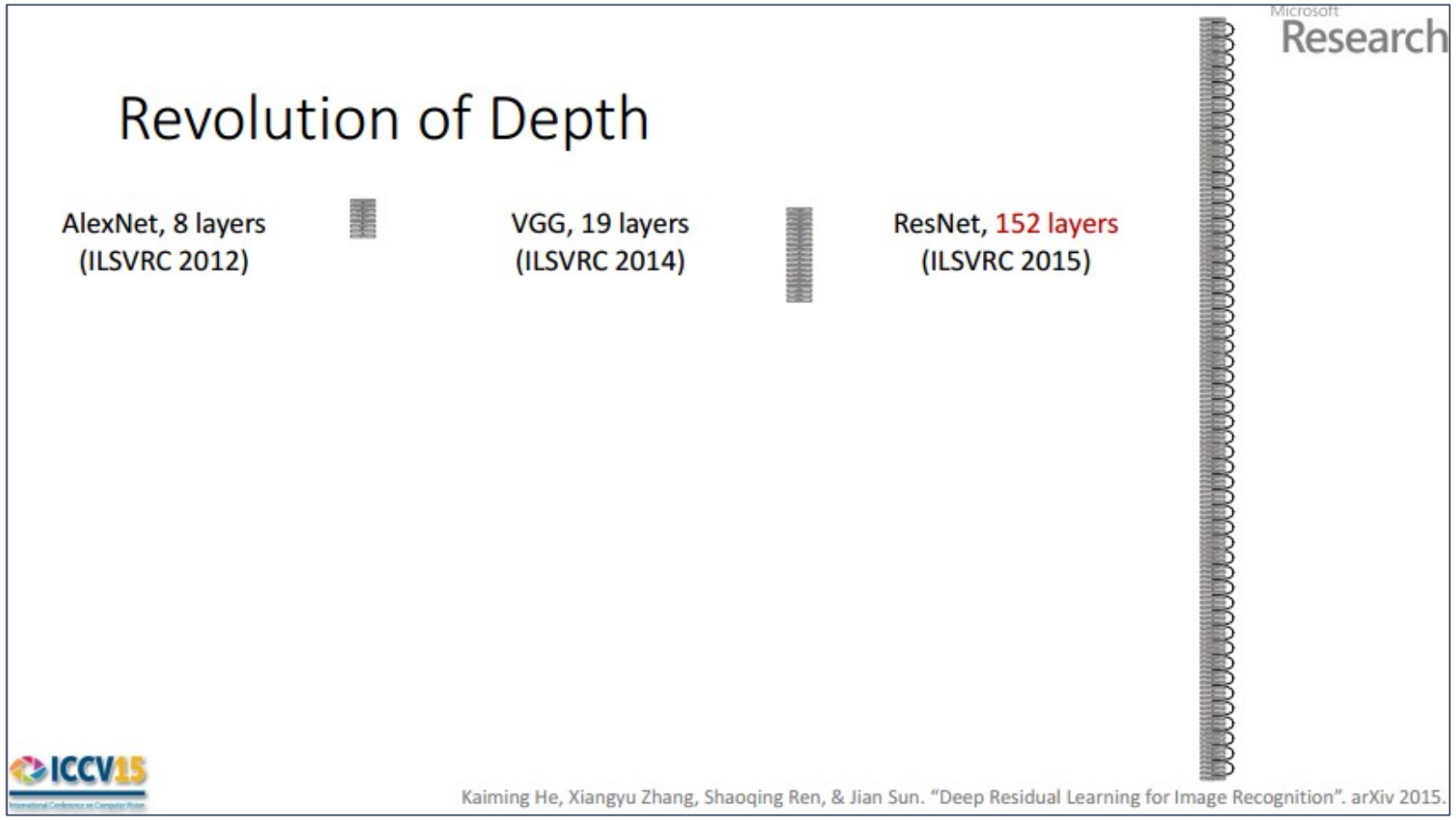


Slide from Kaiming He's recent presentation https://www.youtube.com/watch?v=1PGLj-uKT1w

Case Study: ResNet

[He et al., 2015]

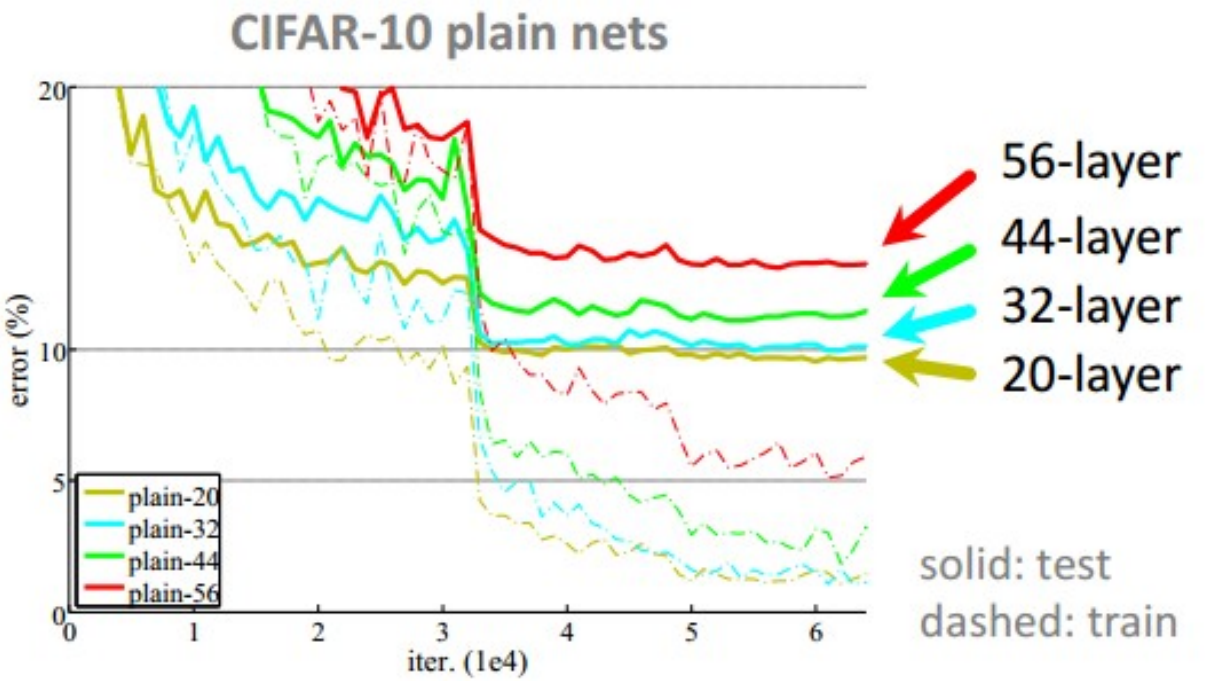
ILSVRC 2015 winner (3.6% top 5 error)



2-3 weeks of training on 8 GPU machine

at runtime: faster than a VGGNet! (even though it has 8x more layers)

(slide from Kaiming He's recent presentation)

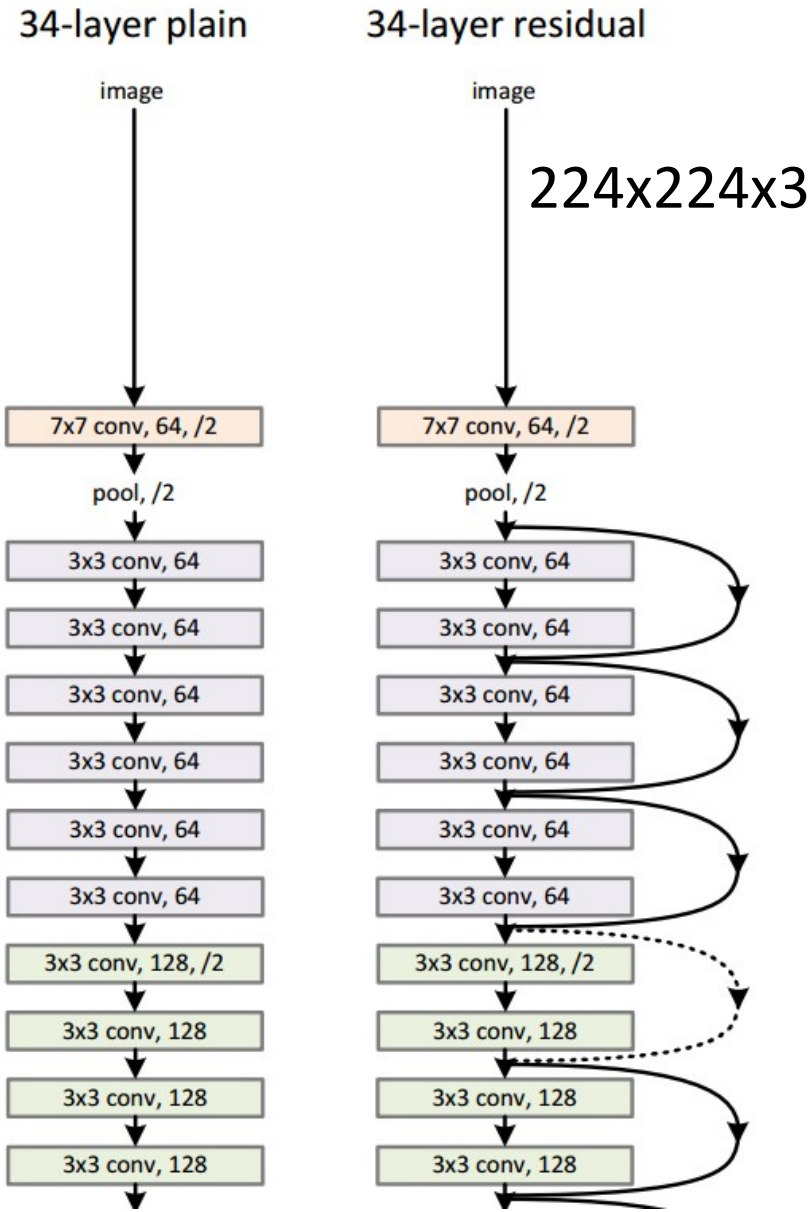


Failure to optimize

- A deeper network should perform at least as good as a shallow architecture
- “Proof”: take shallow network, copy layers and set the output to the identity map.

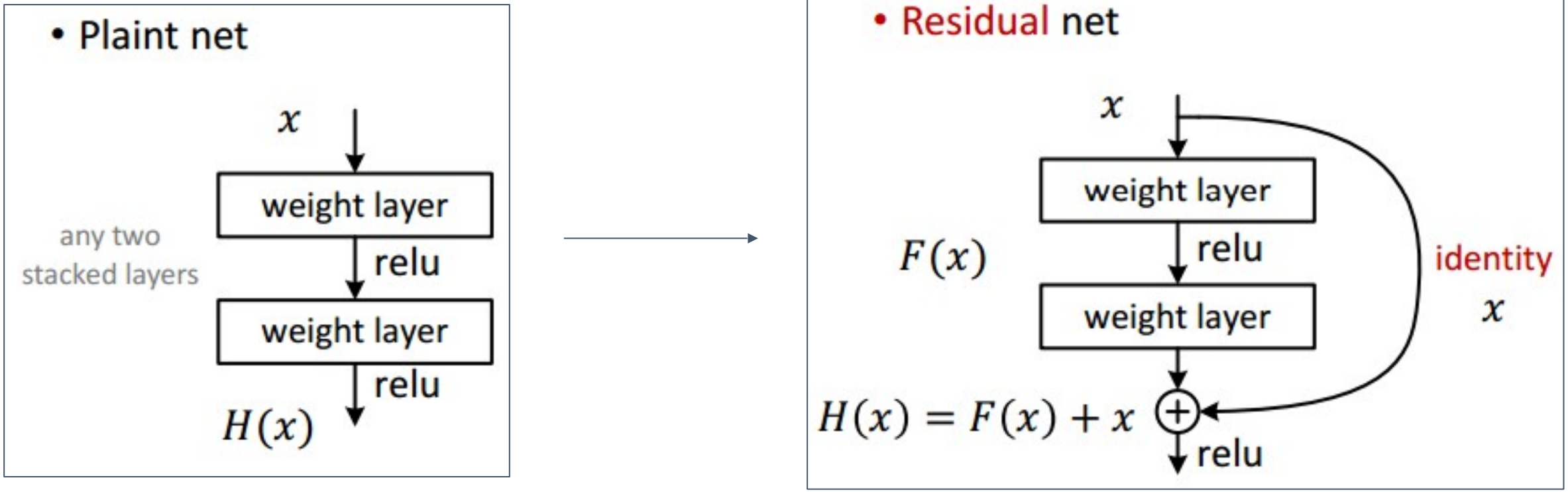
Case Study: ResNet

[He et al., 2015]

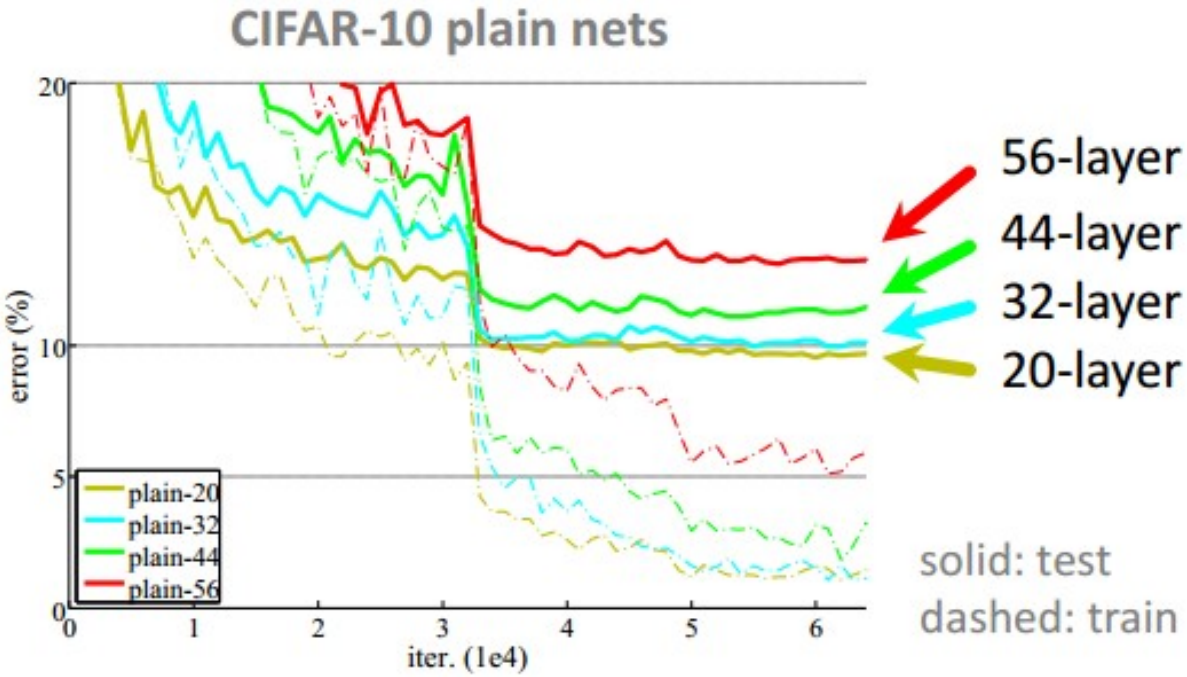


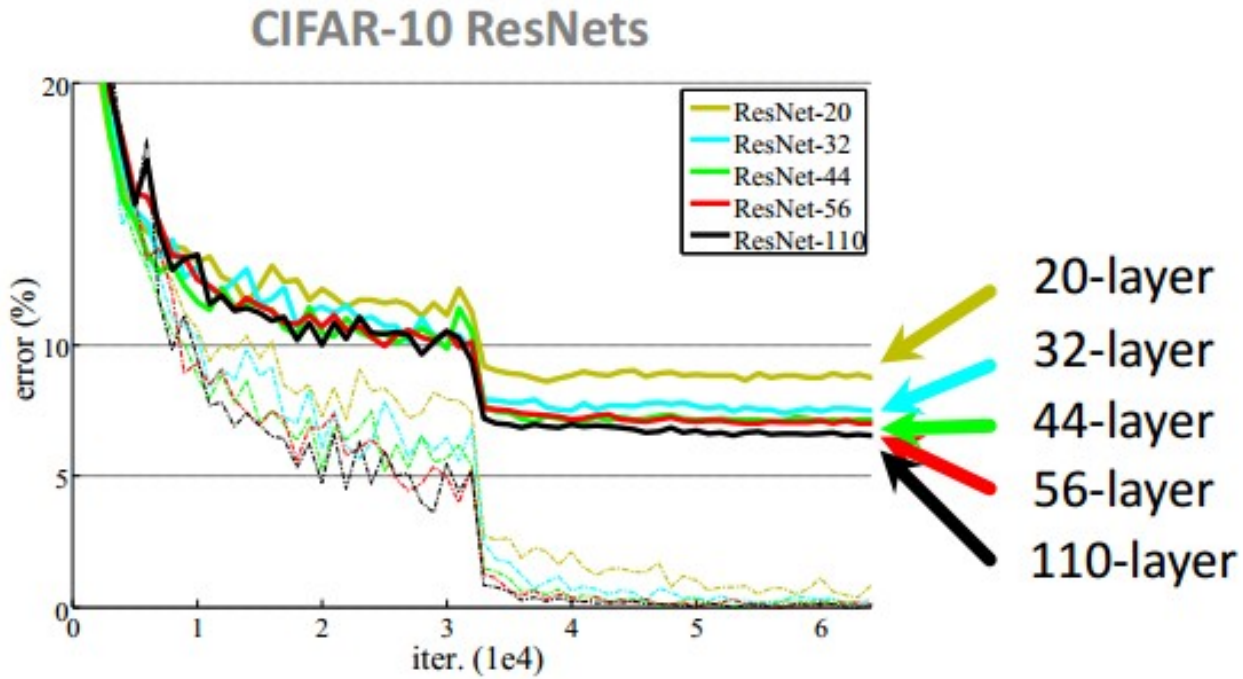
Case Study: ResNet

[He et al., 2015]



CIFAR-10 experiments





Other CNN architectures

Wide ResNet

ResNetX

DenseNet

…

And fully convolutional networks (see second half of this lecture)

HourGlass

U-Net

…

Next: object detection

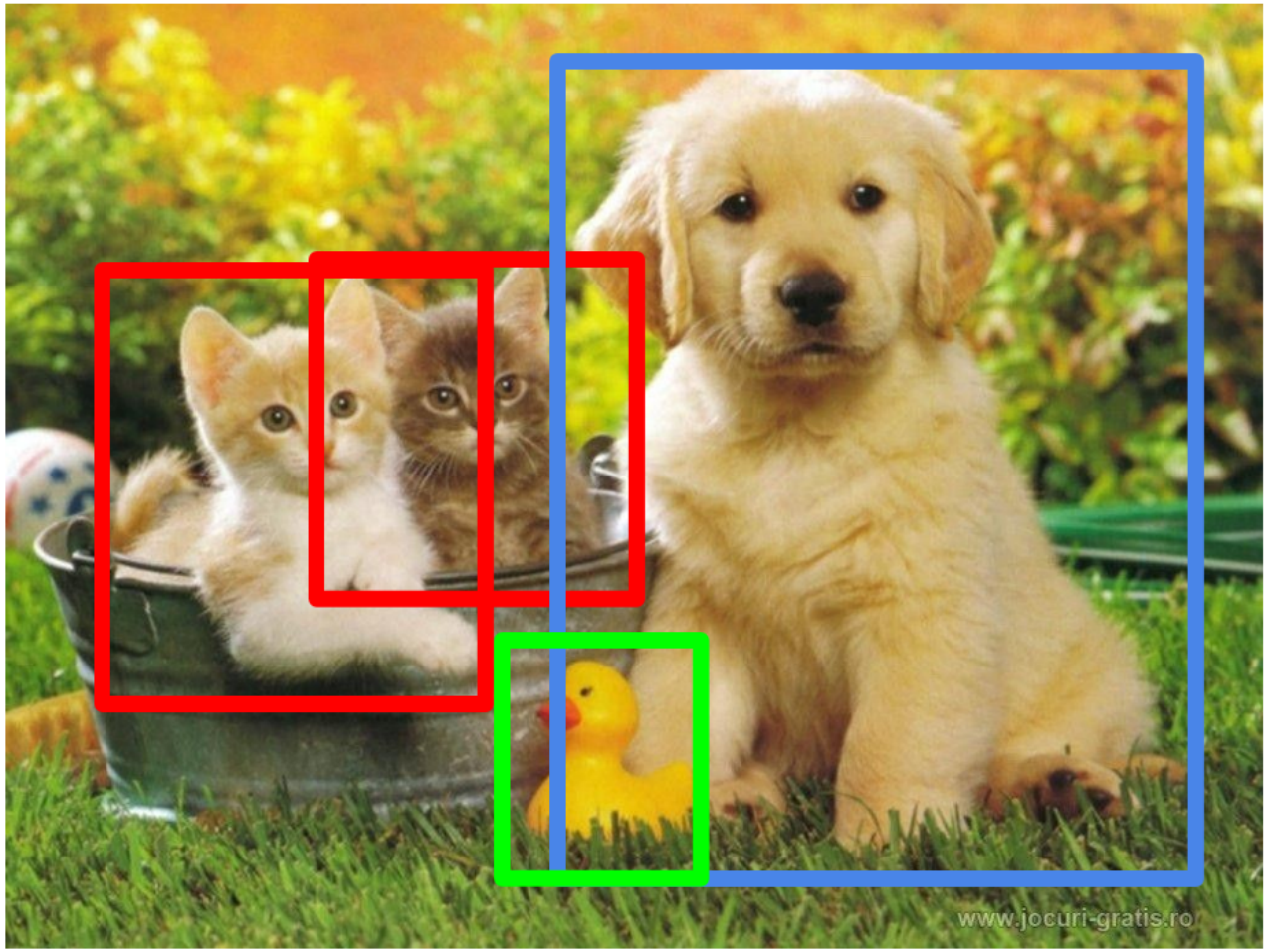
CNNs – case studies from computer vision

- Image Classification - *different CNN architectures*
- **Object detection - *how to use CNNs as building blocks***

Fully convolutional networks

- Semantic Segmentation

Object Detection



[Slides adopted from: Li, Karpathy, Johnson]

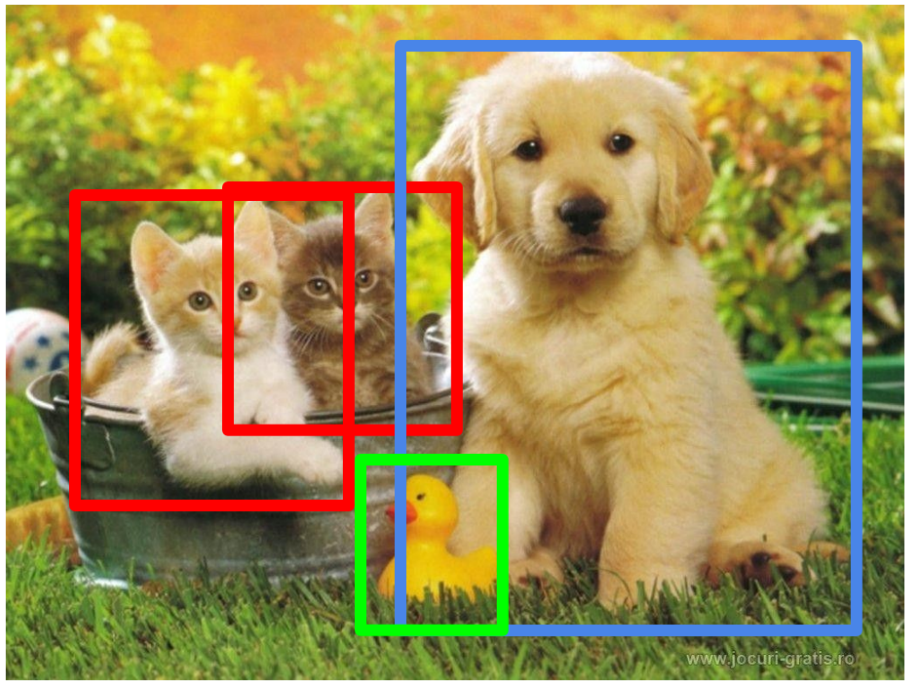
Object Detection

PASCAL Visual Object Classes Challenge

20 Classes, ~10K images, ~25K annotated objects

Training, validation, test dataset.

Object Detection as Regression



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

= 16 numbers

[Slides adopted from: Li, Karpathy, Johnson]

Object Detection as Regression

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

= 8 numbers

[Slides adopted from: Li, Karpathy, Johnson]

Object Detection as Regression

CAT, (x, y, w, h)
CAT, (x, y, w, h)
....
CAT (x, y, w, h)

= many numbers

[Slides adopted from: Li, Karpathy, Johnson]

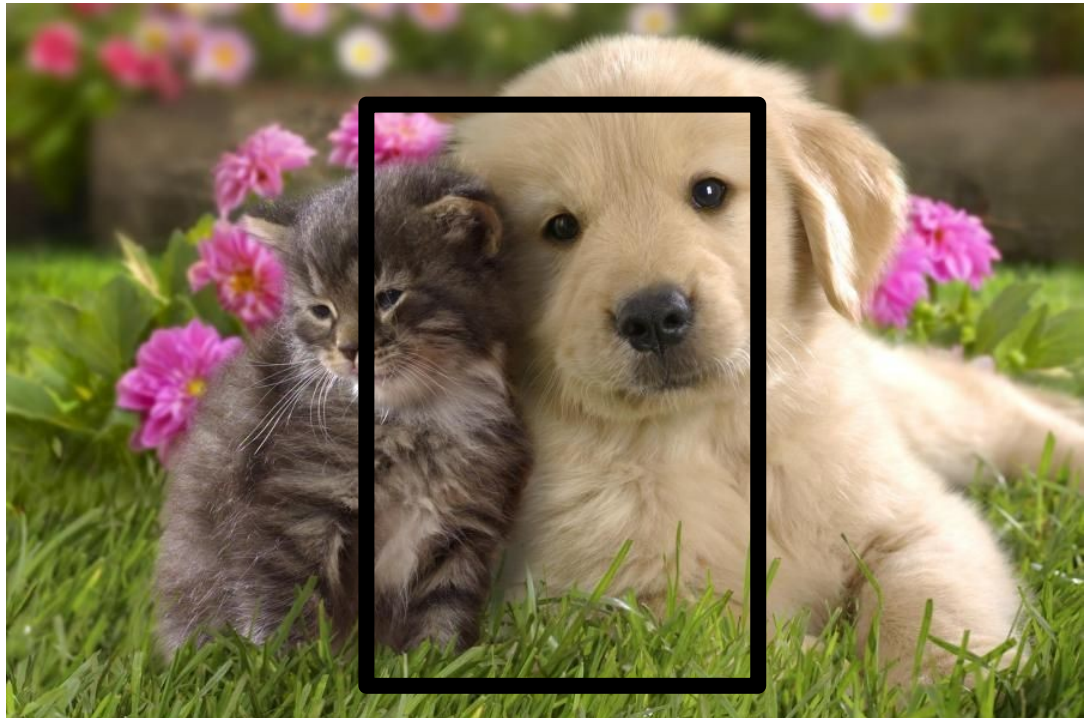
Object Detection as Classification

CAT? YES!

DOG? NO

[Slides adopted from: Li, Karpathy, Johnson]

Object Detection as Classification

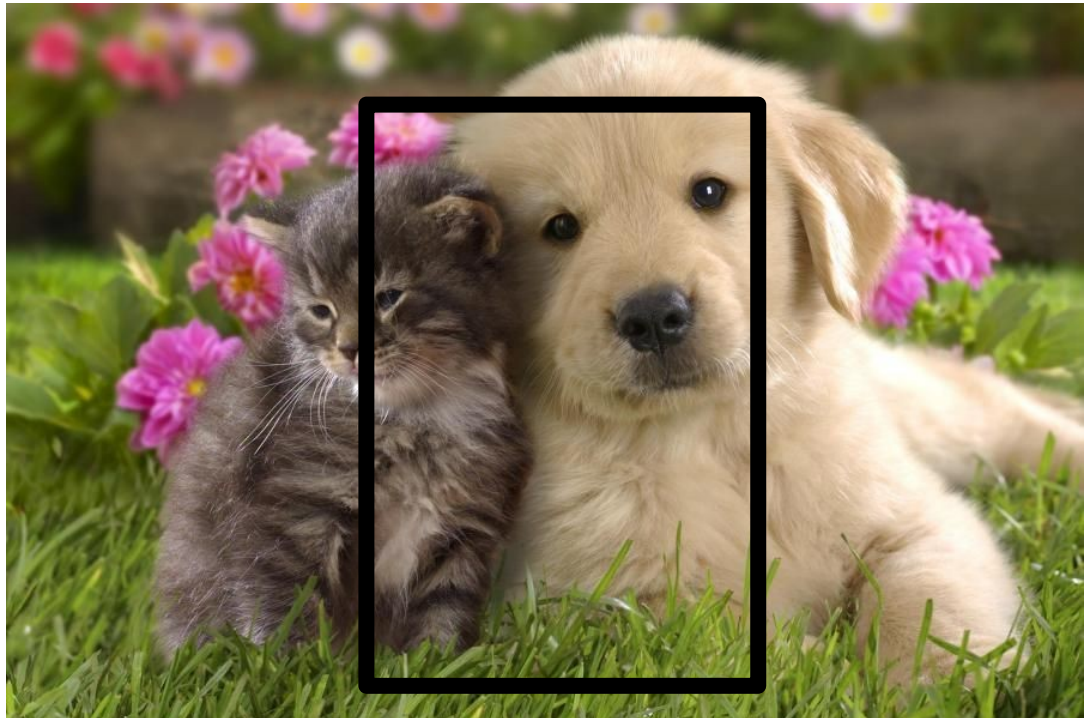


CAT? NO

DOG? NO

[Slides adopted from: Li, Karpathy, Johnson]

Object Detection as Classification



CAT? NO

DOG? NO

[Slides adopted from: Li, Karpathy, Johnson]

Object Detection as Classification

Problem: Need to test many positions and scales

Solution: If your classifier is fast enough, just do it

[Slides adopted from: Li, Karpathy, Johnson]

Object Detection as Classification

Problem: Need to test many positions and scales, and use a computationally demanding classifier (CNN)

Solution: Only look at a tiny subset of possible positions

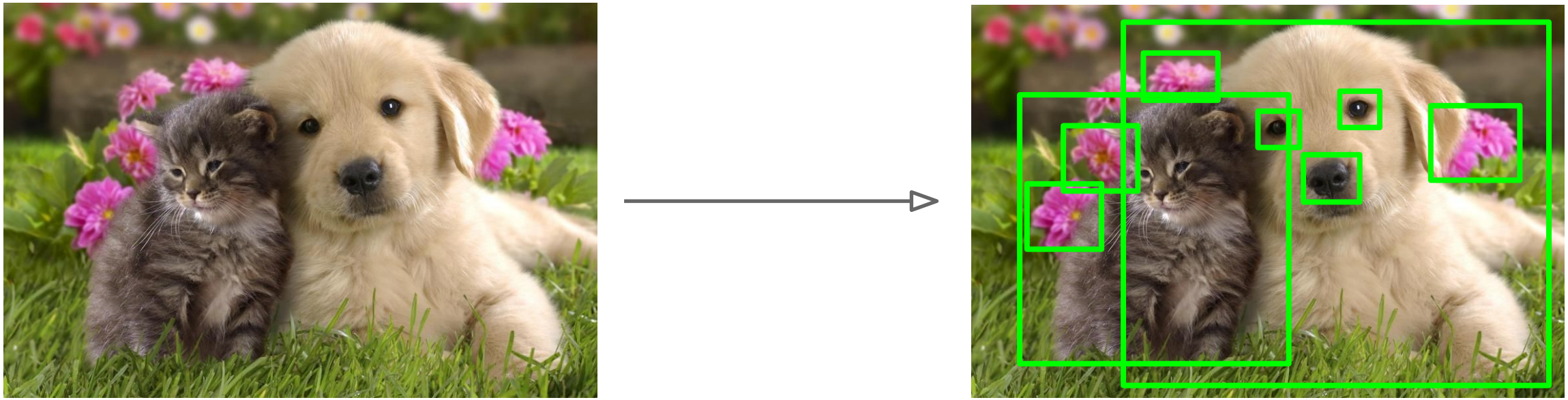
[Slides adopted from: Li, Karpathy, Johnson]

Region Proposals

Find “blobby” image regions that are likely to contain objects

“Class-agnostic” object detector

Look for “blob-like” regions

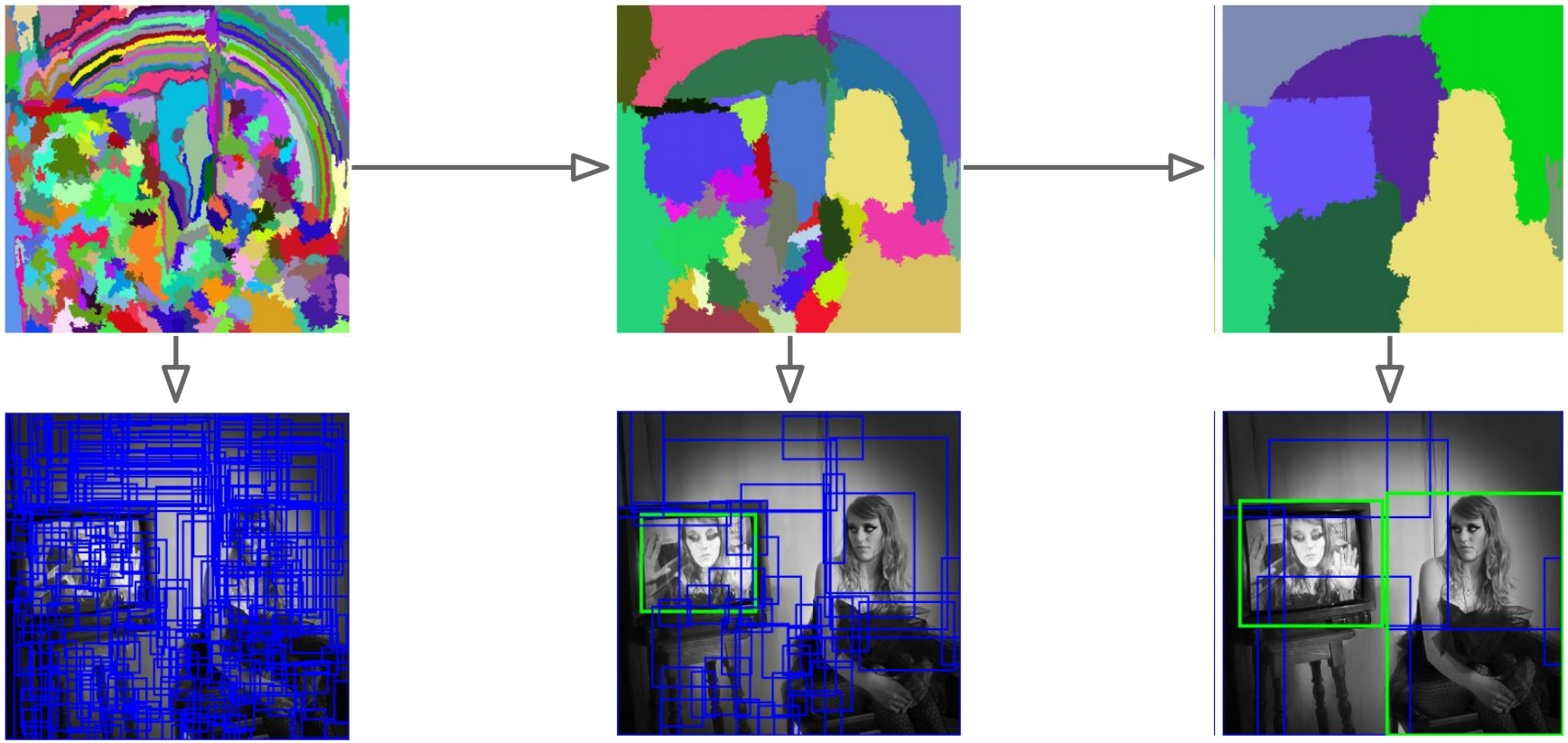


[Slides adopted from: Li, Karpathy, Johnson]

Region Proposals: Selective Search

Bottom-up segmentation, merging regions at multiple scales

Convert regions to boxes



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

[Slides adopted from: Li, Karpathy, Johnson]

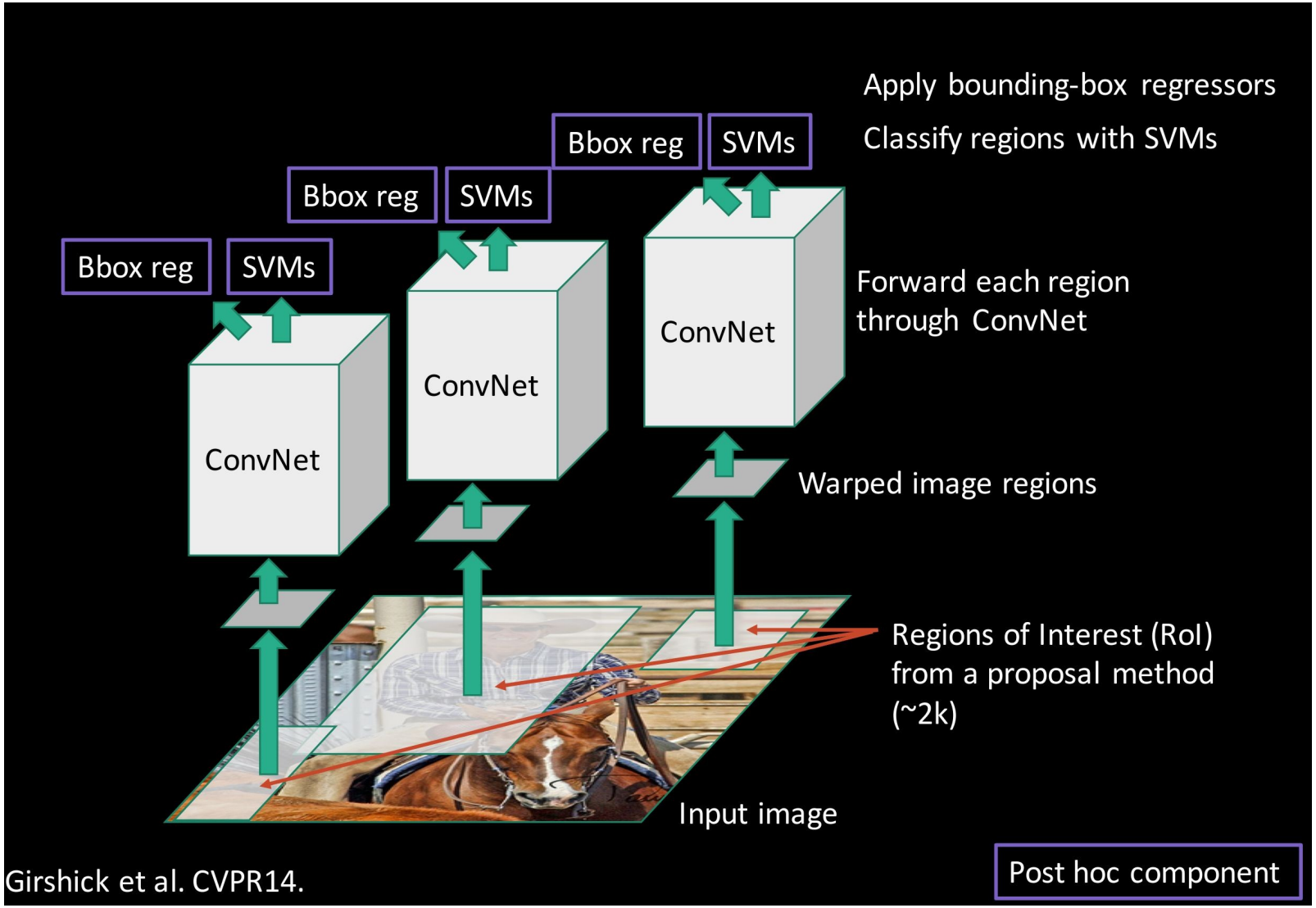
Region Proposals: Many other choices

Method	Approach	Outputs Segments	Outputs Score	Control #proposals	Time (sec.)	Repea-tability	Recall Results	Detection Results
Bing [18]	Window scoring		✓	✓	0.2	⋆⋆⋆	⋆	·
CPMC [19]	Grouping	✓	✓	✓	250	-	⋆⋆	⋆
EdgeBoxes [20]	Window scoring		✓	✓	0.3	⋆⋆	⋆⋆⋆	⋆⋆⋆
Endres [21]	Grouping	✓	✓	✓	100	-	⋆⋆⋆	⋆⋆
Geodesic [22]	Grouping	✓		✓	1	⋆	⋆⋆⋆	⋆⋆
MCG [23]	Grouping	✓	✓	✓	30	⋆	⋆⋆⋆	⋆⋆⋆
Objectness [24]	Window scoring		✓	✓	3	·	⋆	·
Rahtu [25]	Window scoring		✓	✓	3	·	·	⋆
RandomizedPrim's [26]	Grouping	✓		✓	1	⋆	⋆	⋆⋆
Rantalankila [27]	Grouping	✓		✓	10	⋆⋆	·	⋆⋆
Rigor [28]	Grouping	✓		✓	10	⋆	⋆⋆	⋆⋆
SelectiveSearch [29]	Grouping	✓	✓	✓	10	⋆⋆	⋆⋆⋆	⋆⋆⋆
Gaussian				✓	0	·	·	⋆
SlidingWindow				✓	0	⋆⋆⋆	·	·
Superpixels		✓			1	⋆	·	·
Uniform				✓	0	·	·	·

Hosang et al, "What makes for effective detection proposals?", PAMI 2015

[Slides adopted from: Li, Karpathy, Johnson]

Region Proposals + CNN: R-CNN



Girschick et al, “Rich feature hierarchies for accurate object detection and semantic segmentation”, CVPR 2014

Slide credit: Ross Girschick

This Week

CNNs – case studies from computer vision

- Image Classification
- Object detection

Fully convolutional networks

- Semantic Segmentation