Apprentissage, réseaux de neurones et modèles graphiques (RCP209)
Neural Networks and Deep Learning

Nicolas Thome
Prenom.Nom@cnam.fr
`http://cedric.cnam.fr/vertigo/Cours/ml2/`

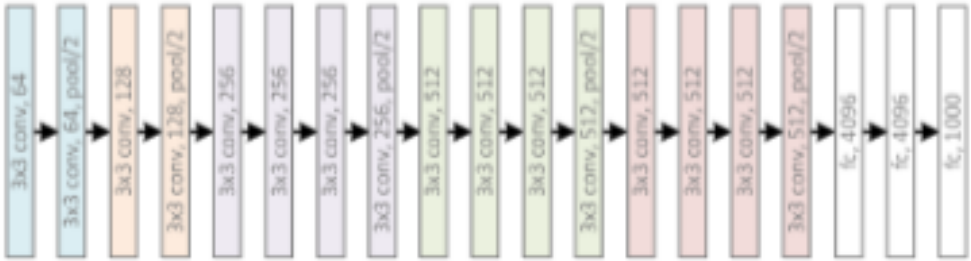
Département Informatique
Conservatoire Nationnal des Arts et Métiers (Cnam)

Outline

Deep Learning since 2012

More & more data (Facebook 10^9 images / day), larger & larger networks

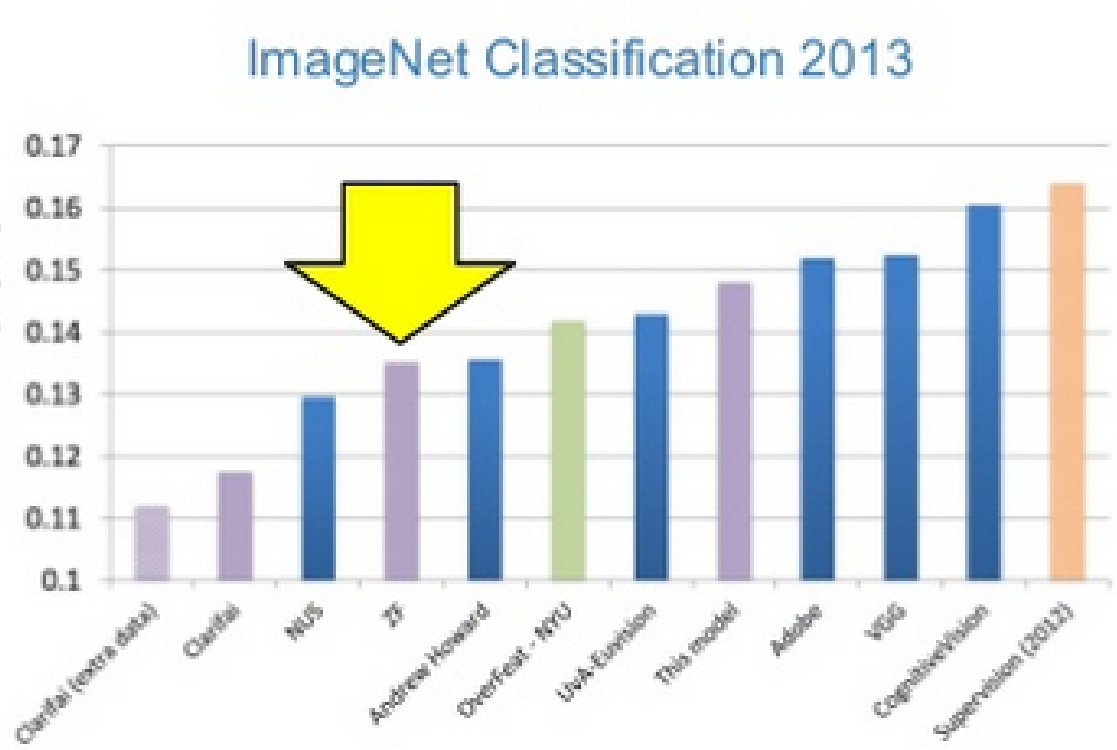
VGG, 16/19 layers, 2014



GoogleNet, 22 layers, 2014

ResNet, 152 layers, 2015

Deep Learning since 2012: ImageNet'13



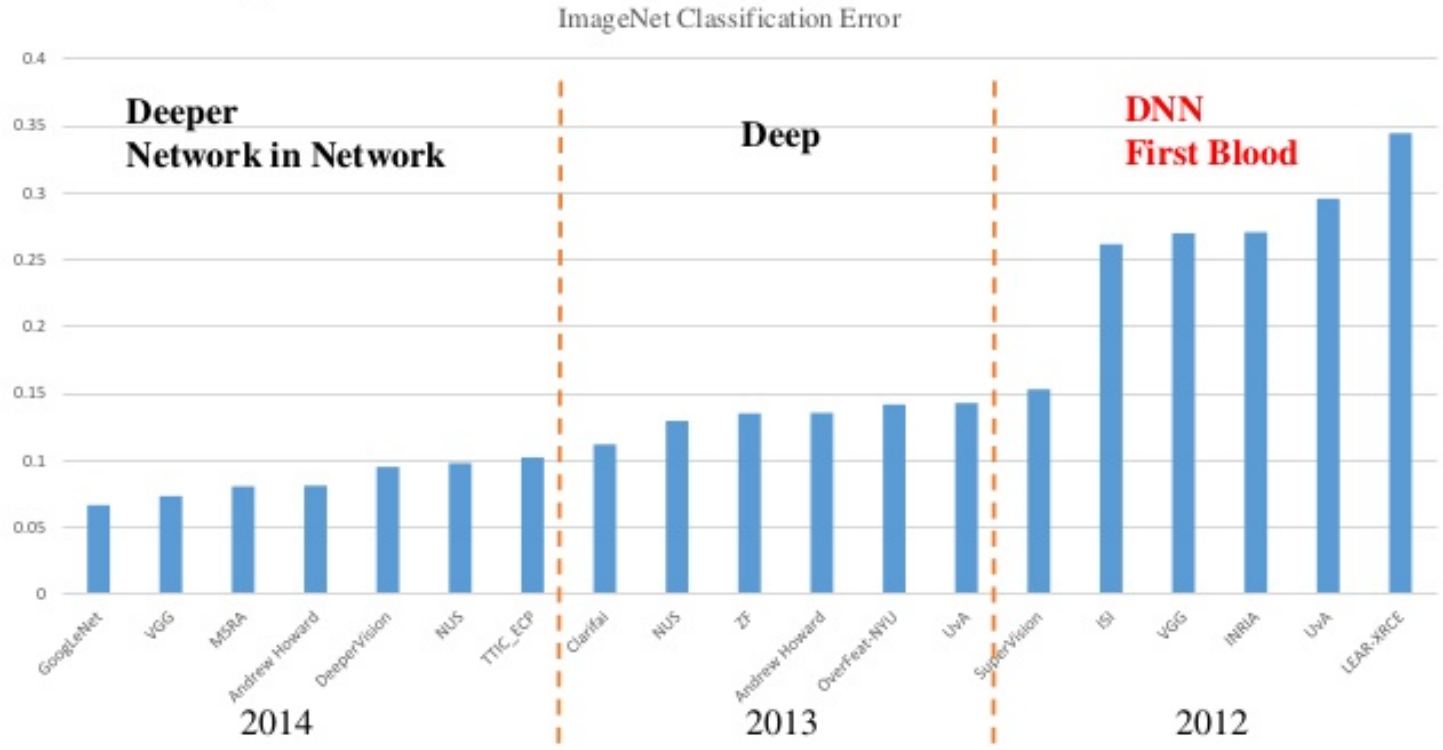
layer	size-in	size-out	kernel	param	FLPS
conv1	220×220×3	110×110×64	7×7×3, 2	9K	115M
pool1	110×110×64	55×55×64	3×3×64, 2	0	
norm1	55×55×64	55×55×64		0	
conv2a	55×55×64	55×55×64	1×1×64, 1	4K	13M
conv2	55×55×64	55×55×192	3×3×64, 1	111K	335M
norm2	55×55×192	55×55×192		0	
pool2	55×55×192	28×28×192	3×3×192, 2	0	
conv3a	28×28×192	28×28×192	1×1×192, 1	37K	29M
conv3	28×28×192	28×28×384	3×3×192, 1	664K	521M
pool3	28×28×384	14×14×384	3×3×384, 2	0	
conv4a	14×14×384	14×14×384	1×1×384, 1	148K	29M
conv4	14×14×384	14×14×256	3×3×384, 1	885K	173M
conv5a	14×14×256	14×14×256	1×1×256, 1	66K	13M
conv5	14×14×256	14×14×256	3×3×256, 1	590K	116M
conv6a	14×14×256	14×14×256	1×1×256, 1	66K	13M
conv6	14×14×256	14×14×256	3×3×256, 1	590K	116M
pool4	14×14×256	7×7×256	3×3×256, 2	0	
concat	7×7×256	7×7×256		0	
fc1	7×7×256	1×32×128	maxout p=2	103M	103M
fc2	1×32×128	1×32×128	maxout p=2	34M	34M
fc7128	1×32×128	1×1×128		524K	0.5M
L2	1×1×128	1×1×128		0	
total				140M	1.6B

- Zeiler / Fergus (ZF) network: archi ~ AlexNet (2012), *i.e.* conv + FC

Deep Learning since 2012: ImageNet'14

ImageNet Classification

- **1000** categories and **1.2** million training images

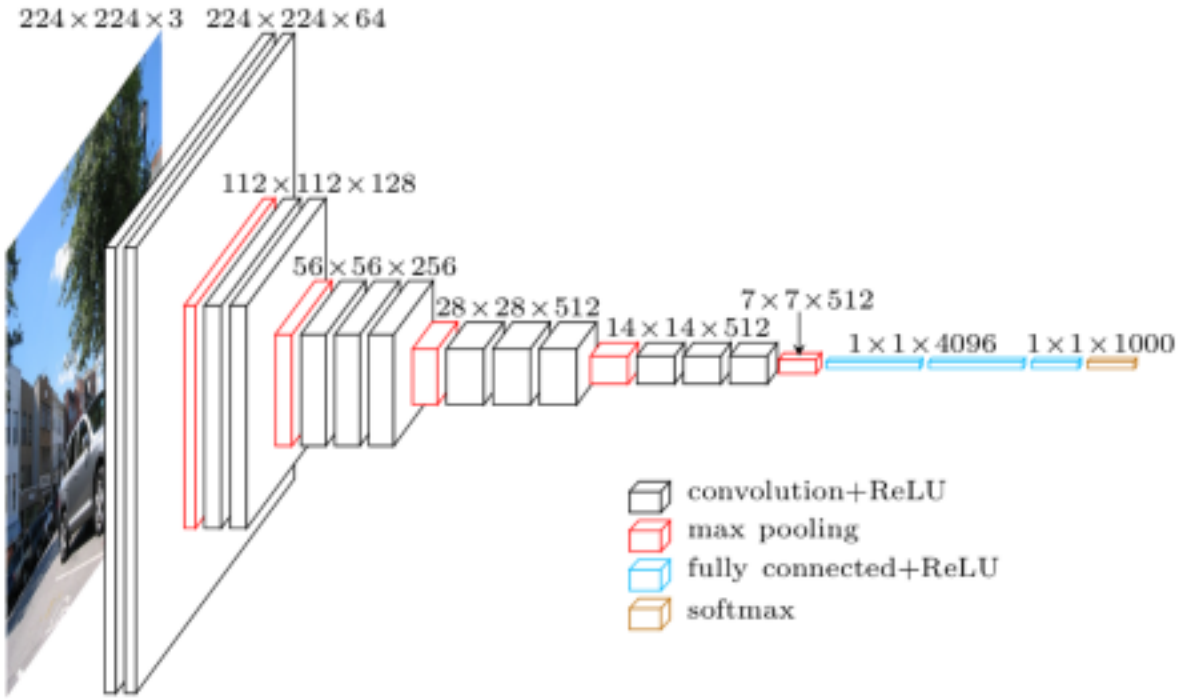


Li Fei-Fei: ImageNet Large Scale Visual Recognition Challenge, 2014 http://image-net.org/

9/3/2014 lipiji.pz@gmail.com 7

- Two leading nets: VGG & GoogLeNet

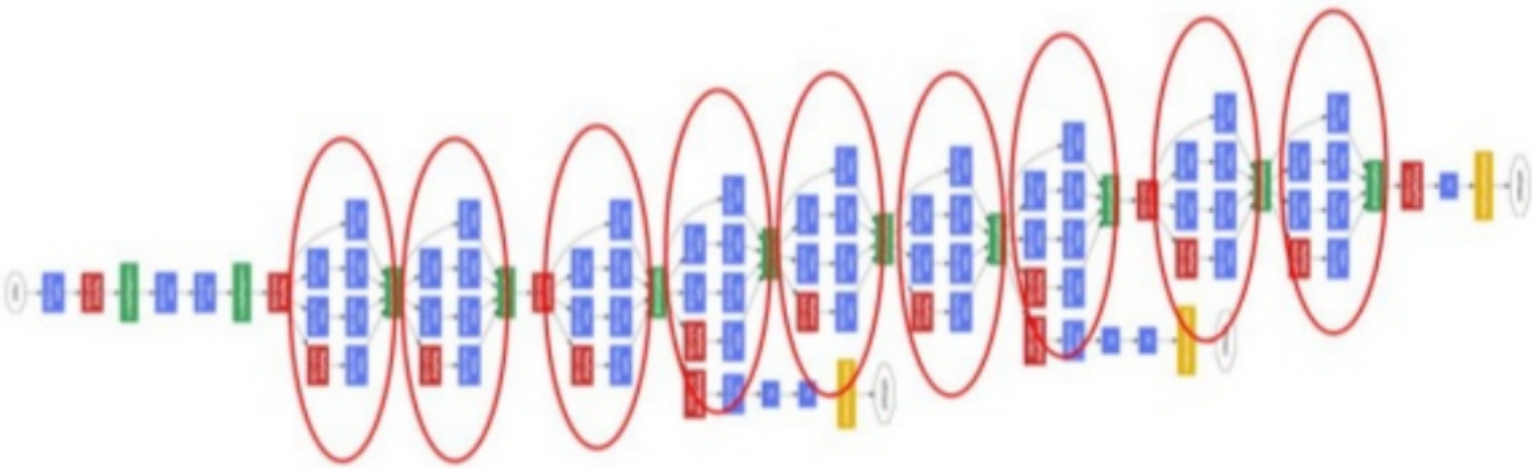
ImageNet'14: VGG



Still Conv + FC, BUT :

- No pooling between some convolutionnal mayers
- Convolution stride 1
- 3x3 convolulutions: two 3x3 conv ~ one 5x5 conv

ImageNet'14: GoogLeNet

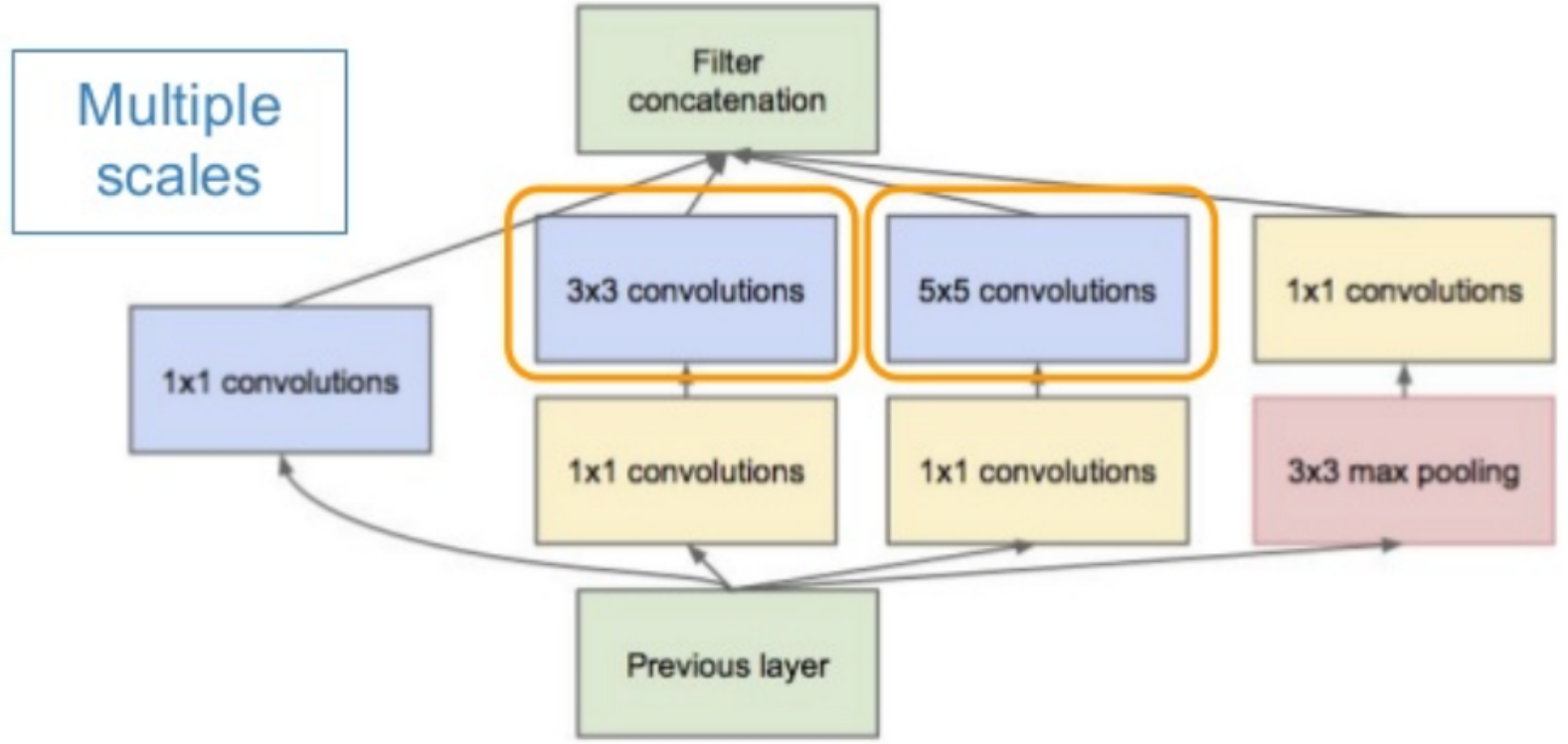


9 **Inception** modules

Network in a network in a network...

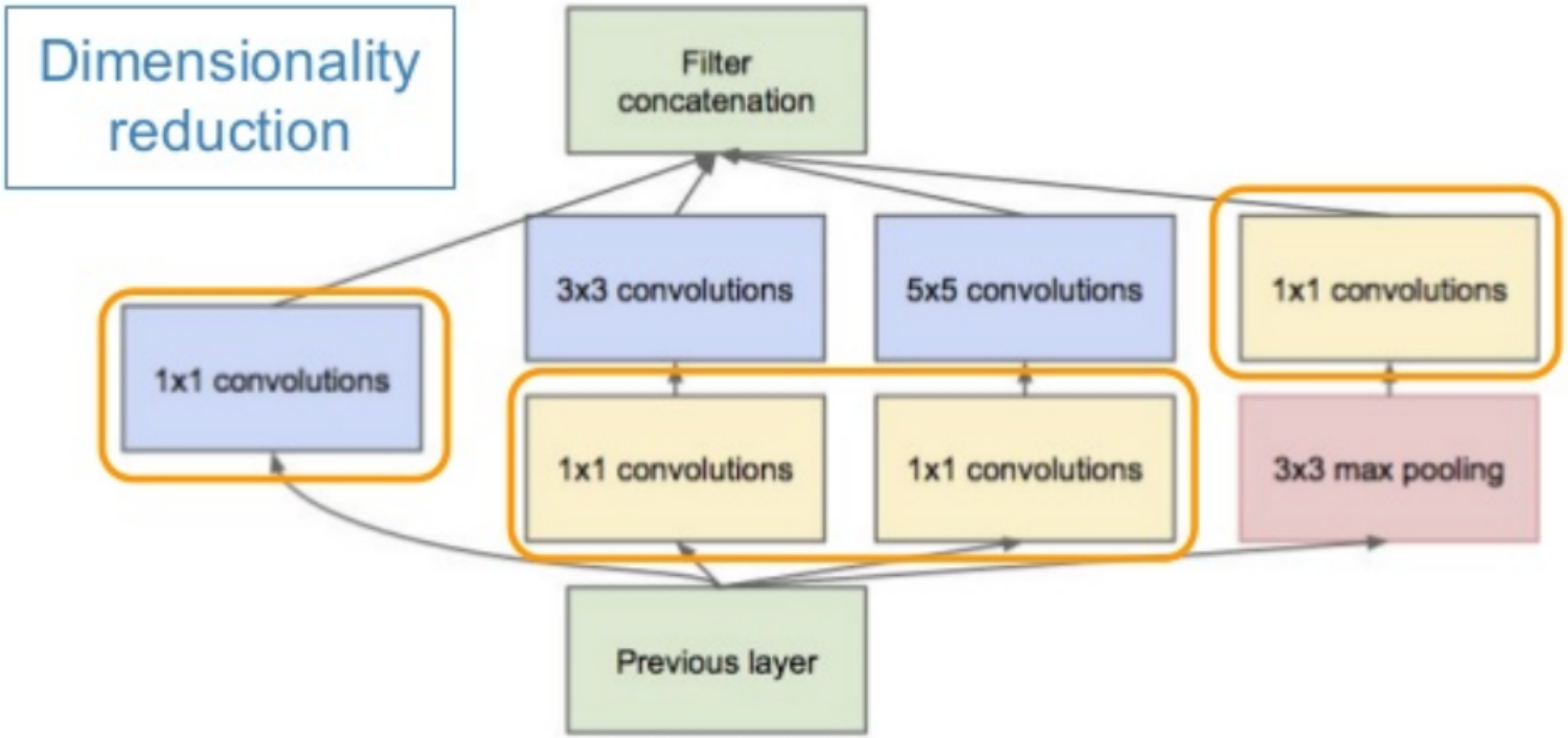
Convolution
Pooling
Softmax
Other

GoogLeNet: Inception Module



- Inspired from Network in Network idea / architecture [LCY13]

GoogLeNet: Inception Module



- Inspired from Network in Network idea / architecture [LCY13]

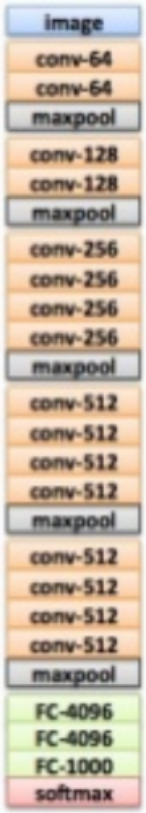
ImageNet'14: GoogLeNet

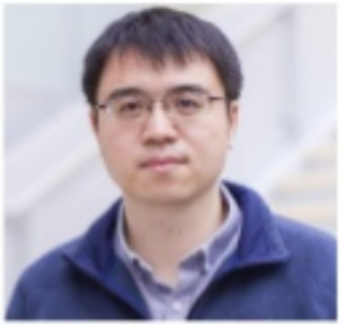
Two Softmax Classifiers at intermediate layers combat the vanishing gradient while providing regularization at training time.

...and no fully connected layers needed !

Deep Learning since 2012: ImageNet'15

E2E: Classification



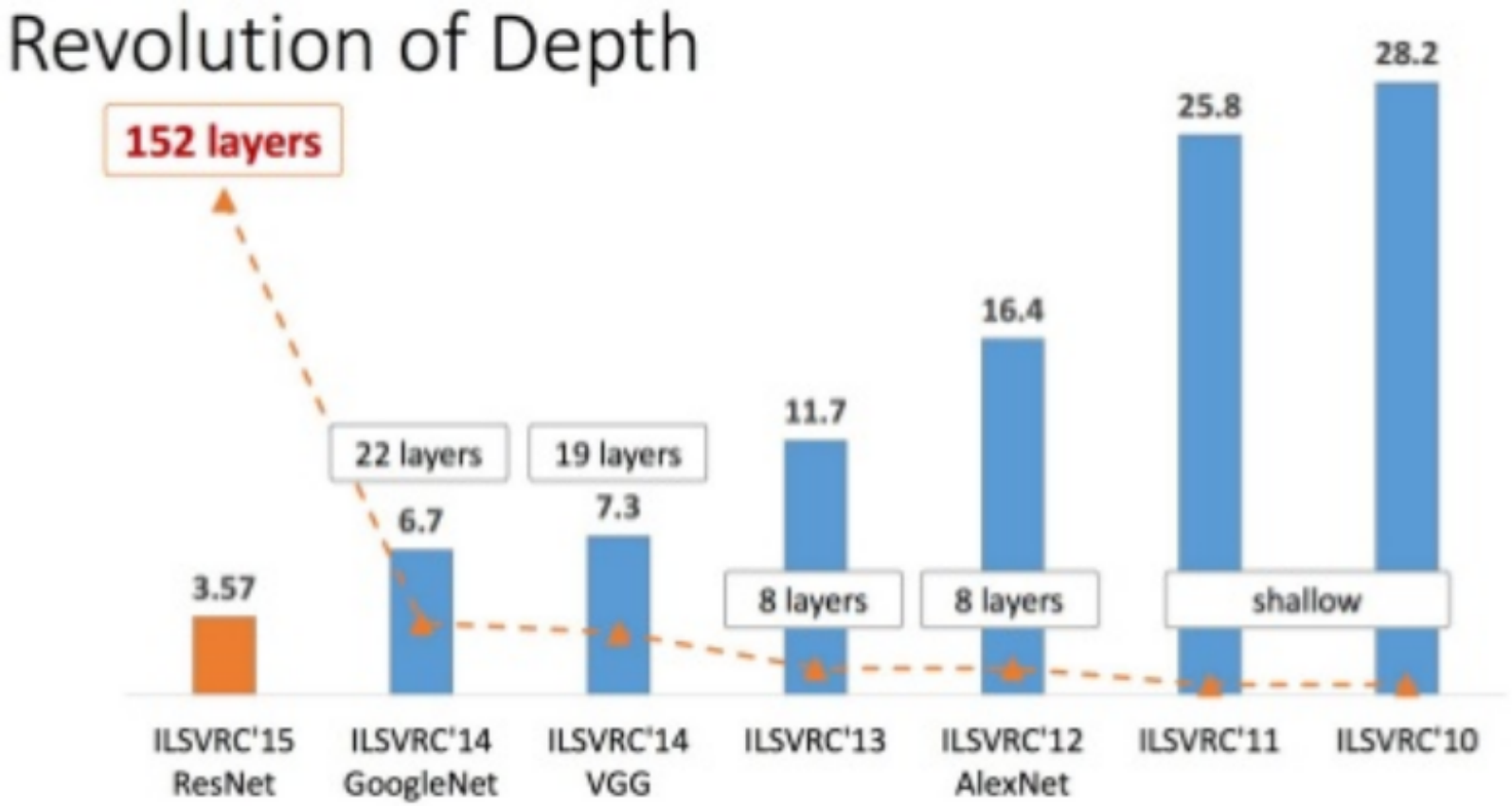


Microsoft Research

3.6% top 5 error…
with 152 layers !!

47

Deep Learning since 2012: ImageNet'15



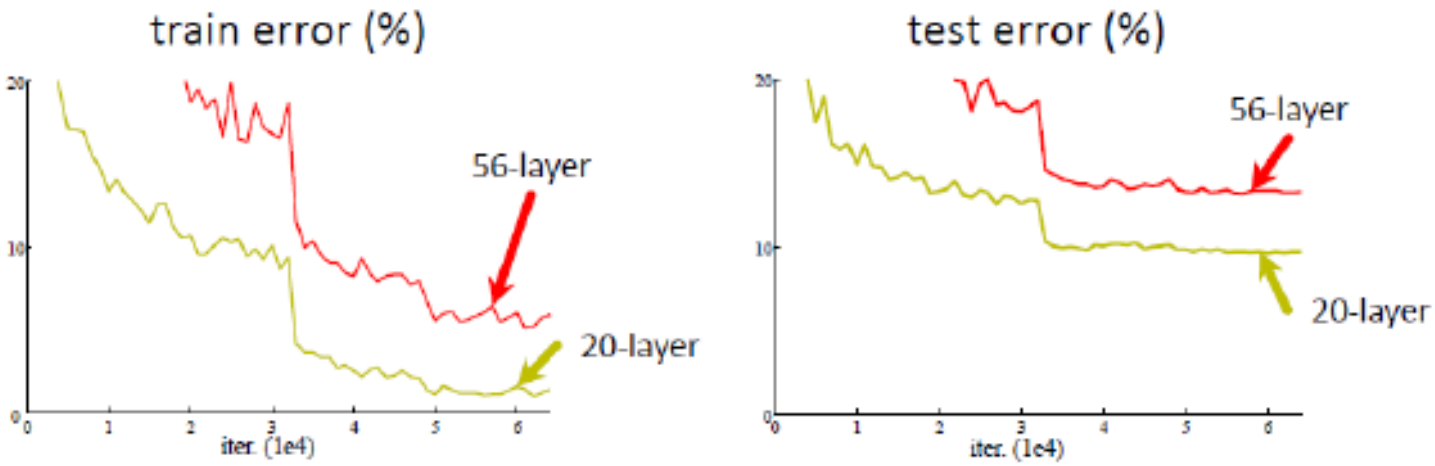
ImageNet'15: ResNet

Deeper VGG: 56 Plain Network

Plain nets: stacking 3x3 conv layers

- 56-layer net has higher training error and test error than 20-layers net

CIFAR-10

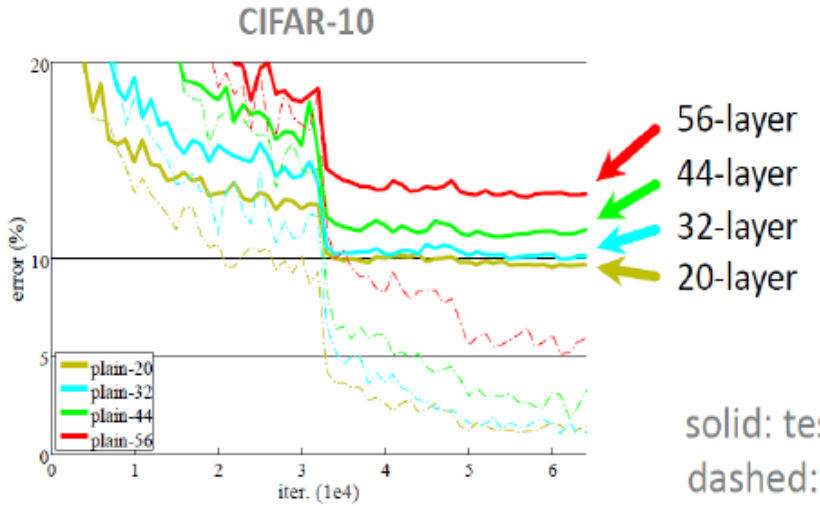


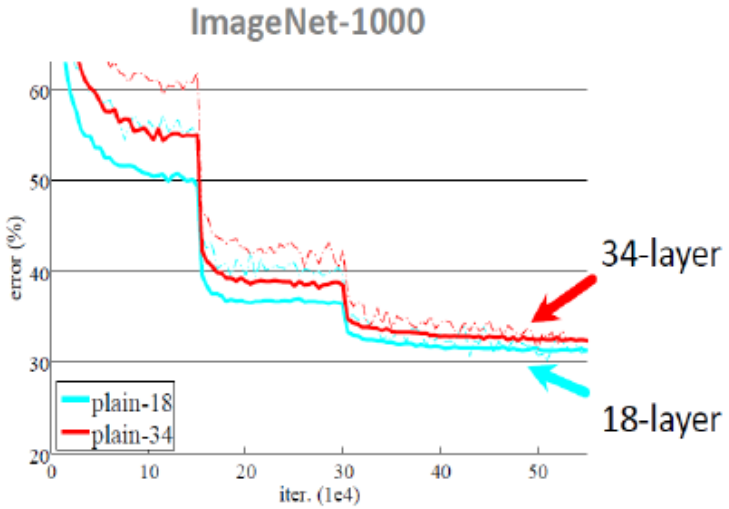
ImageNet'15: ResNet

Deeper VGG:

"Overly deep" plain nets have higher training error

A general phenomenon, observed in many datasets



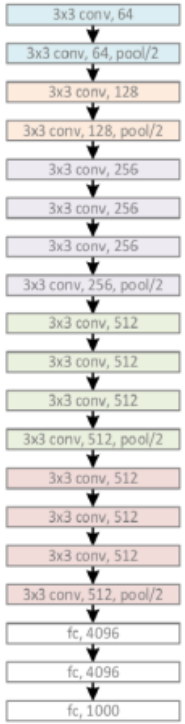


ImageNet'15: ResNet

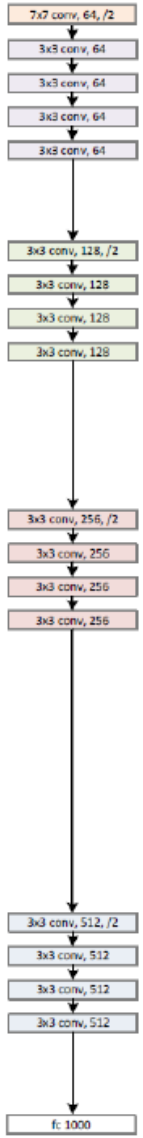
Residual Network

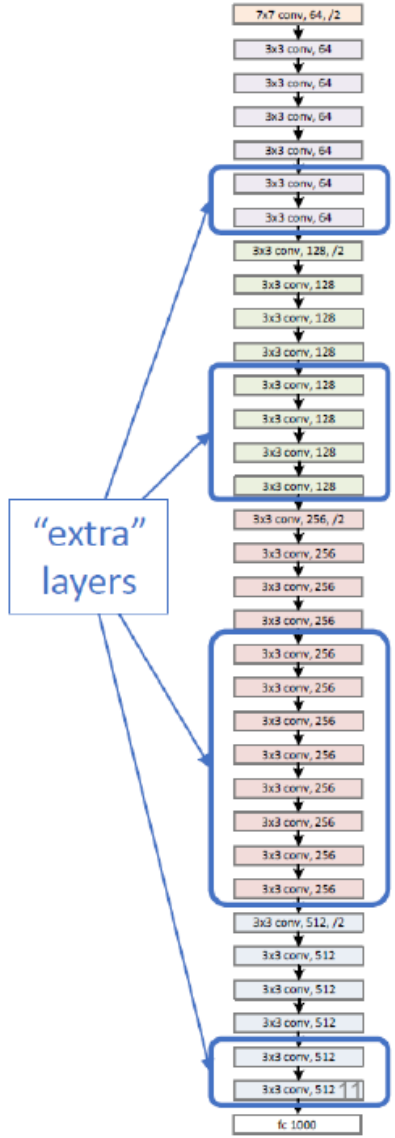
Naïve solution

If extra layers identity mapping, training error not increase



VGG, 16/19 layers, 2014

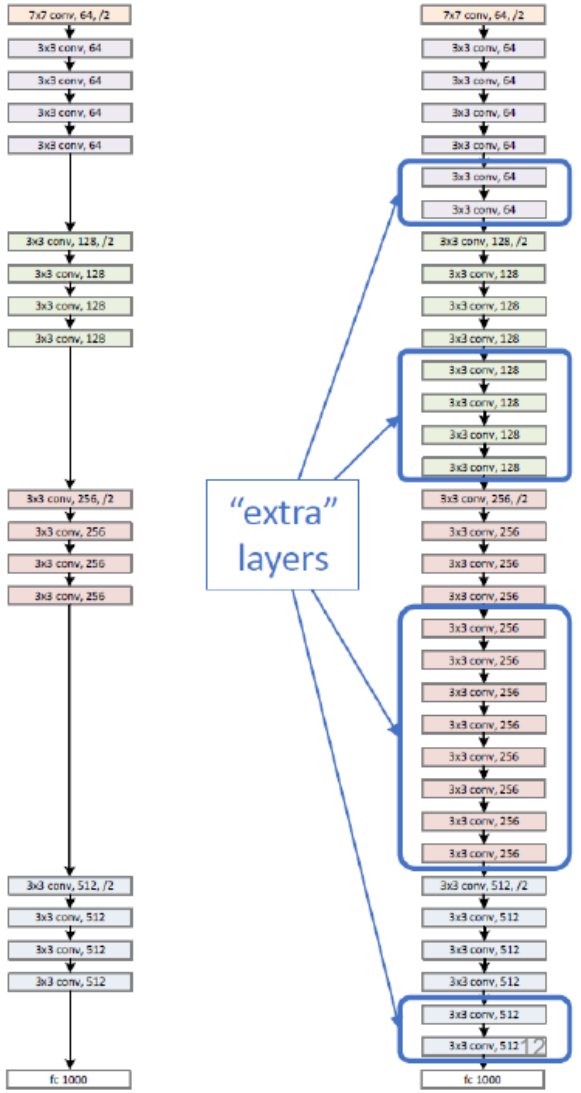




ImageNet'15: ResNet

Residual Network

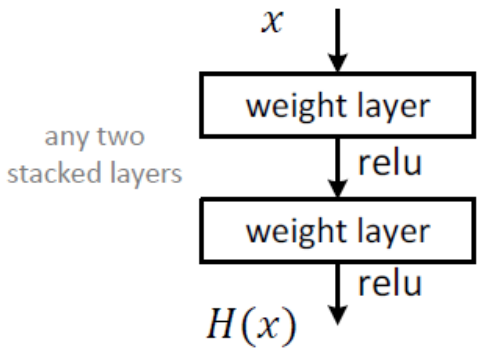
- Deeper networks maintain the tendency of results
  - Features in same level will be almost same
  - An amount of changes is fixed
  - Adding layers make smaller differences
  - Optimal mappings closer to an identity



ImageNet'15: ResNet

Plain block

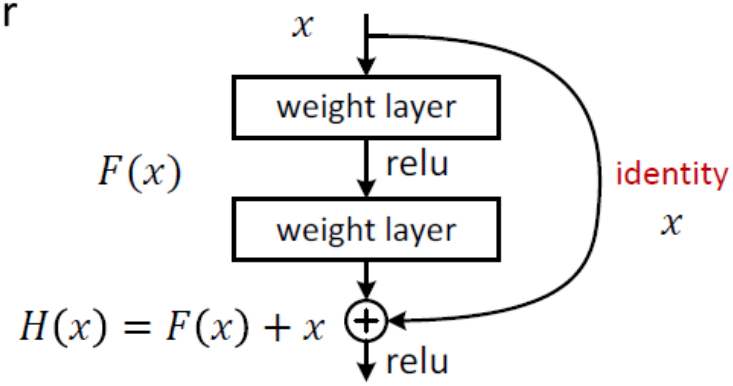
Difficult to make identity mapping because of multiple non-linear layers



Residual block

If identity were optimal, easy to set weights as 0
If optimal mapping is closer to identity, easier to find small fluctuations

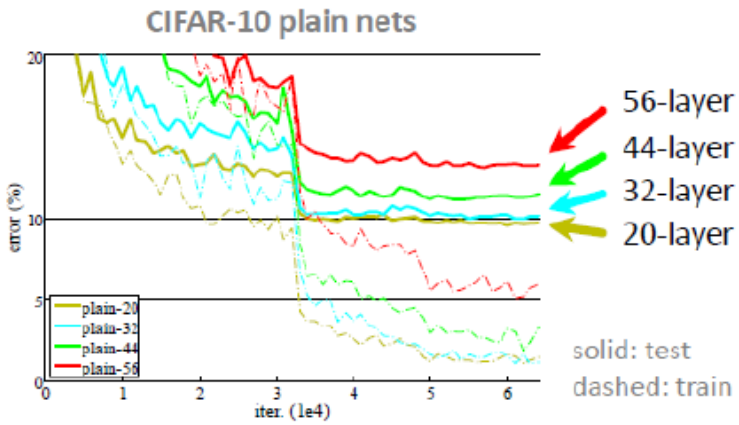
-> Appropriate for treating perturbation as keeping a base information

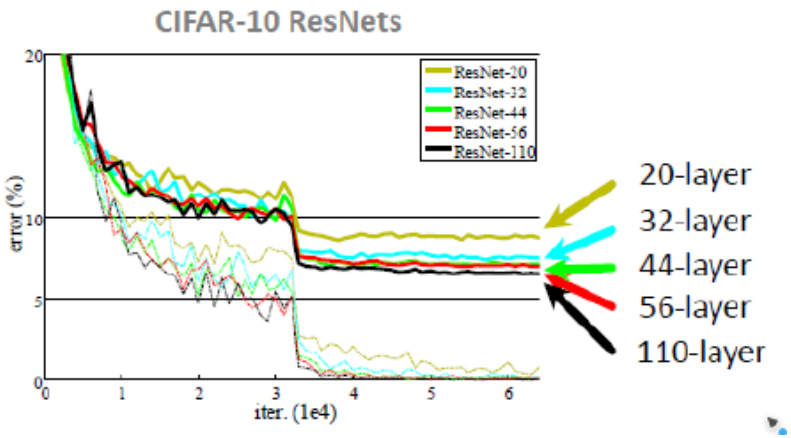


ImageNet'15: ResNet

Results

- Deep Resnets can be trained without difficulties
- Deeper ResNets have lower training error, and also lower test error



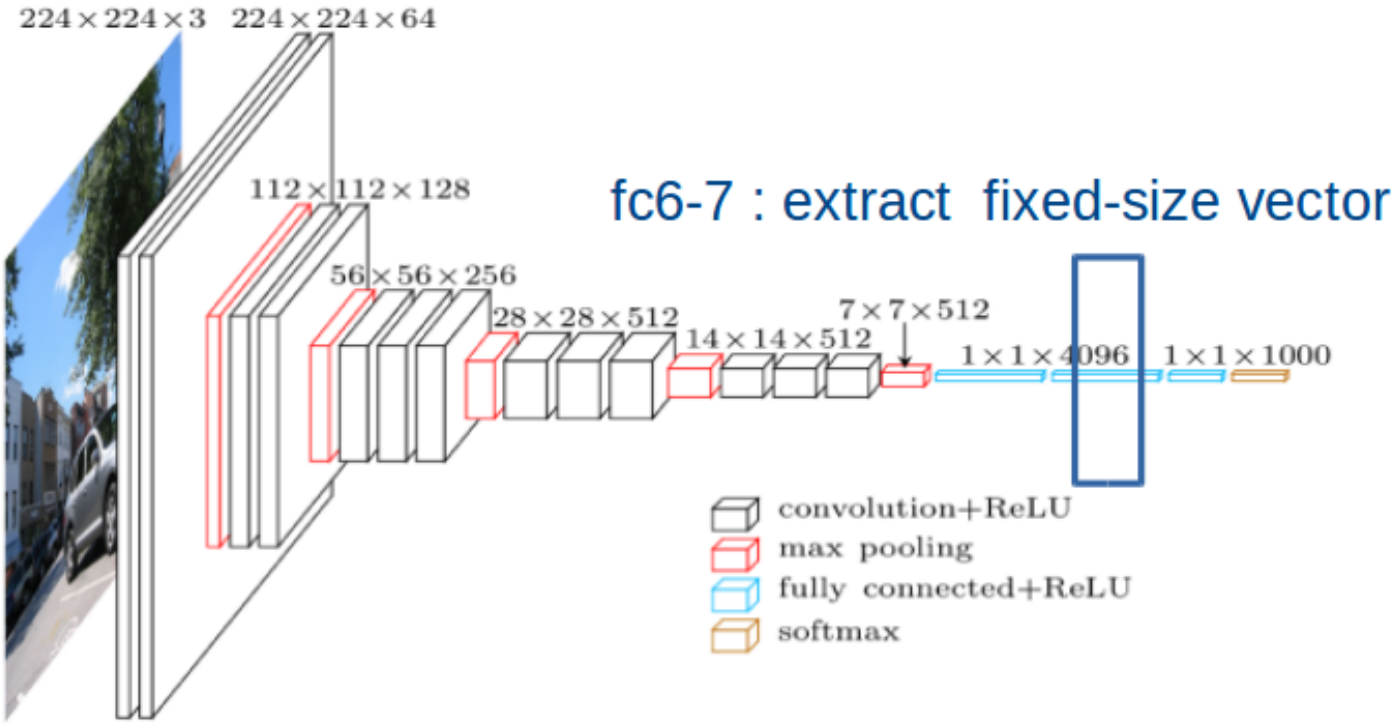


Outline

Deep Learning since 2012

Transferring Representations learned from ImageNet

- Deep ConvNets require large-scale annotated datasets
  - Huge # params ⇒ difficult to train from scratch on "small datasets"
- **BUT:** Extract layer ⇒ fixed-size vector: **"Deep Features" (DF)**



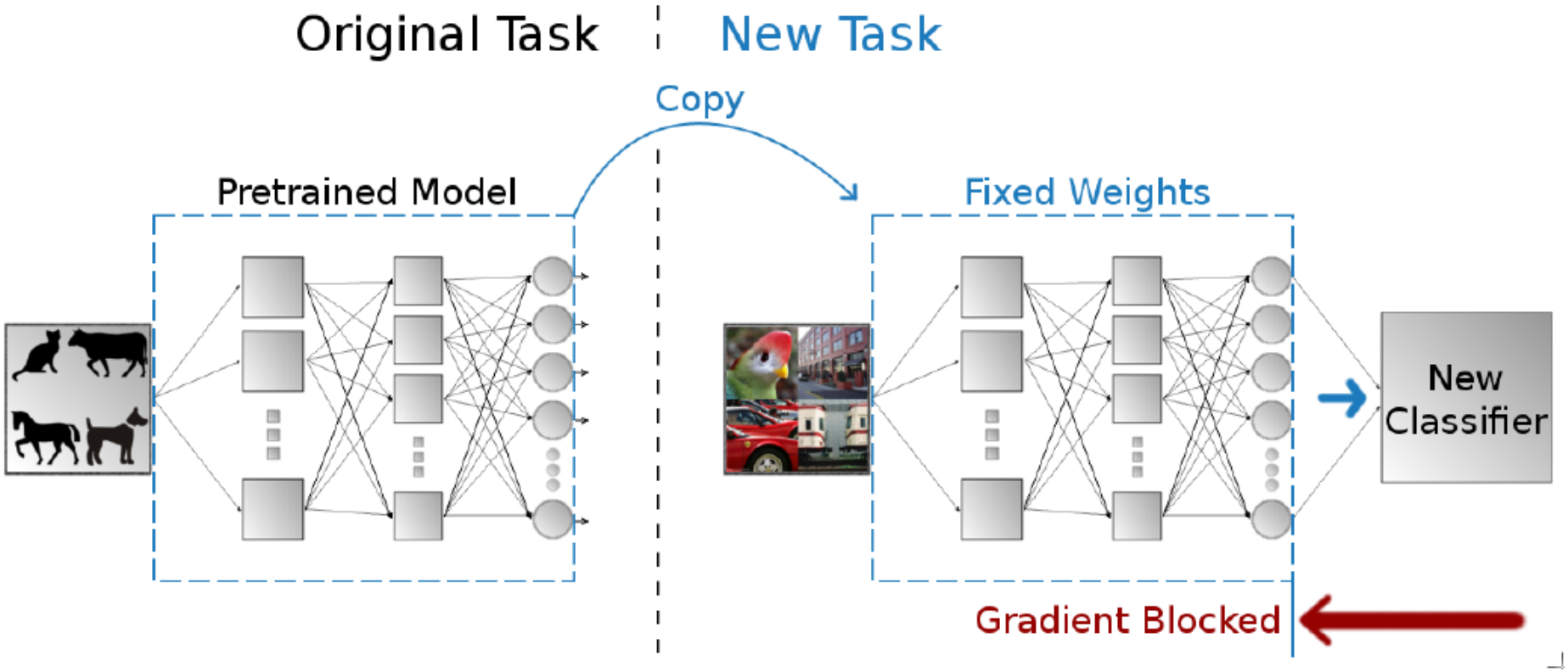
- **Now state-of-the-art for any visual recognition task**

Deep Features (DF) and Domain Adaptation

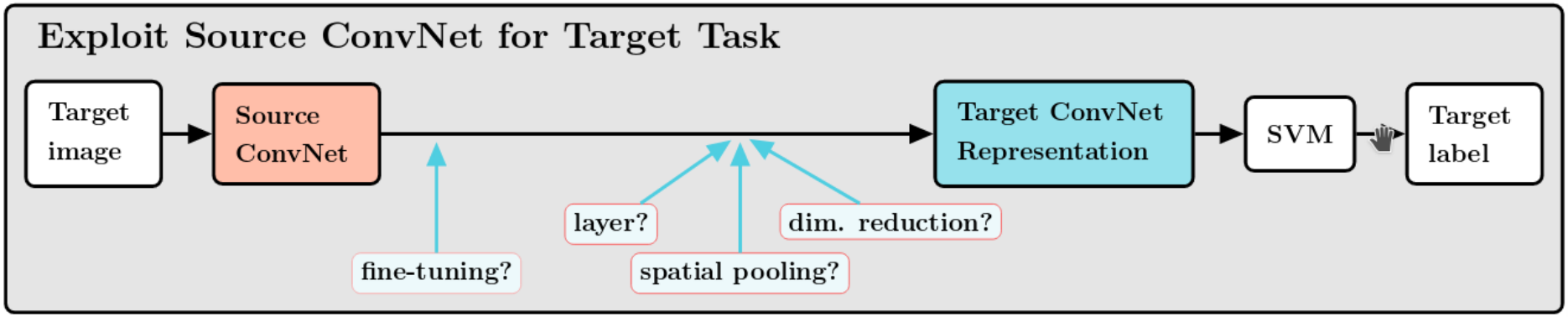
- DF ⇐ Deep ConvNet trained on a large-scale dataset (ImageNet, Places, *etc*)
- DF: off-the-shelf features for any visual recognition task
- DF: Generic features, very robust to :
  - Dataset change ⇒ apply DF for classification with different images, different classes
  - Domain change ⇒ apply DF for other visual tasks, *e.g.* localization, segmentation, pose estimation, retrieval *etc*
- Since 2012: all performance re-benchmarked with DF
- Need to adapt the final layer to match the target domain goal

Deep Features (DF) and Domain Adaptation

DF: off-the-shelf descriptors



Deep Features (DF) and Domain Adaptation

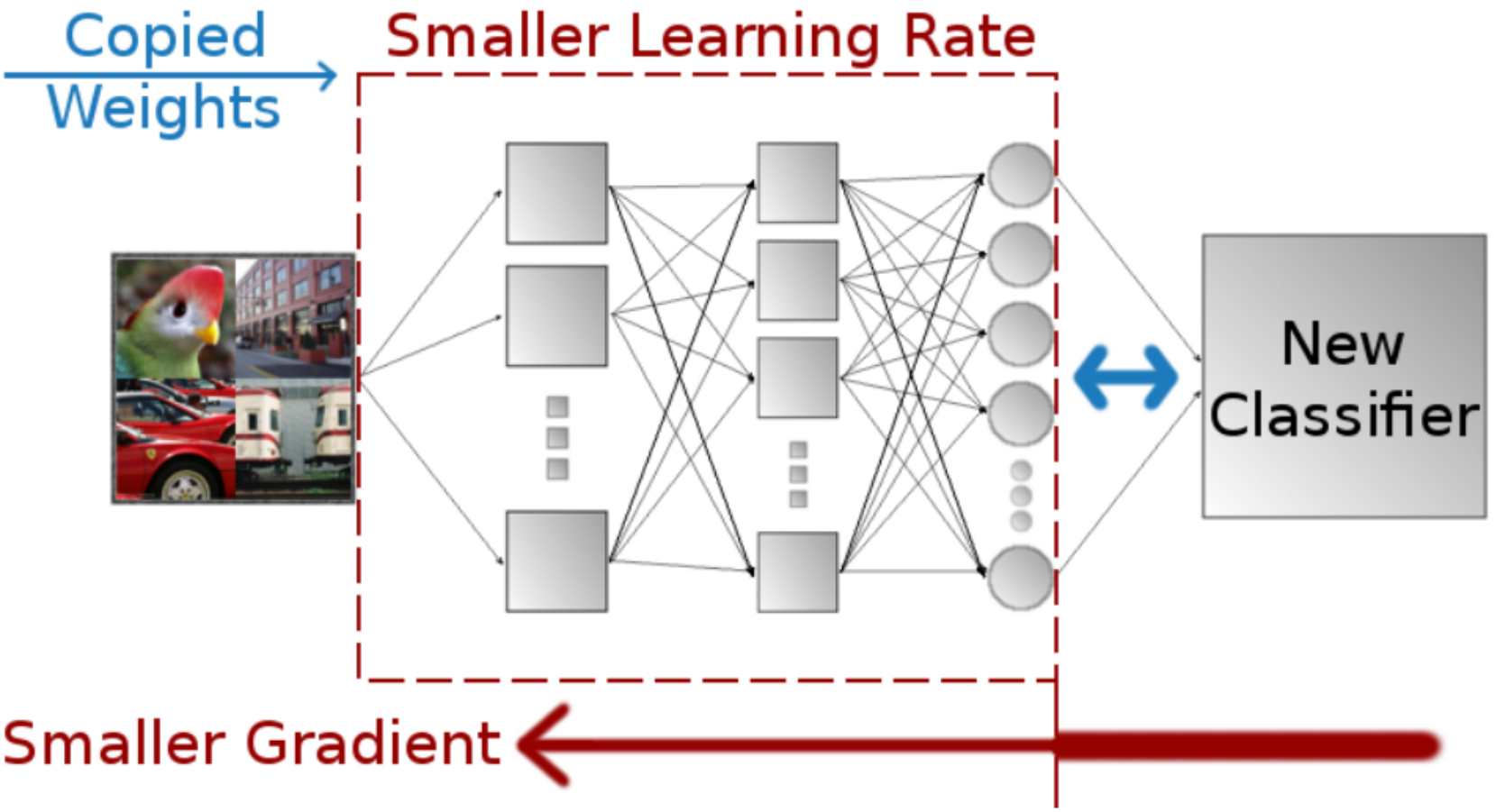


Credit: Razavian *et. al.* [ARS+16]

- Dataset change: fine-tuning ConvNet parameters on the target domain
  - Different learning for fined-tuned & from scratch parameters
- Domain (task) change: adapt archi to the target task (*e.g.* pooling)

Deep Features (DF) and Domain Adaptation

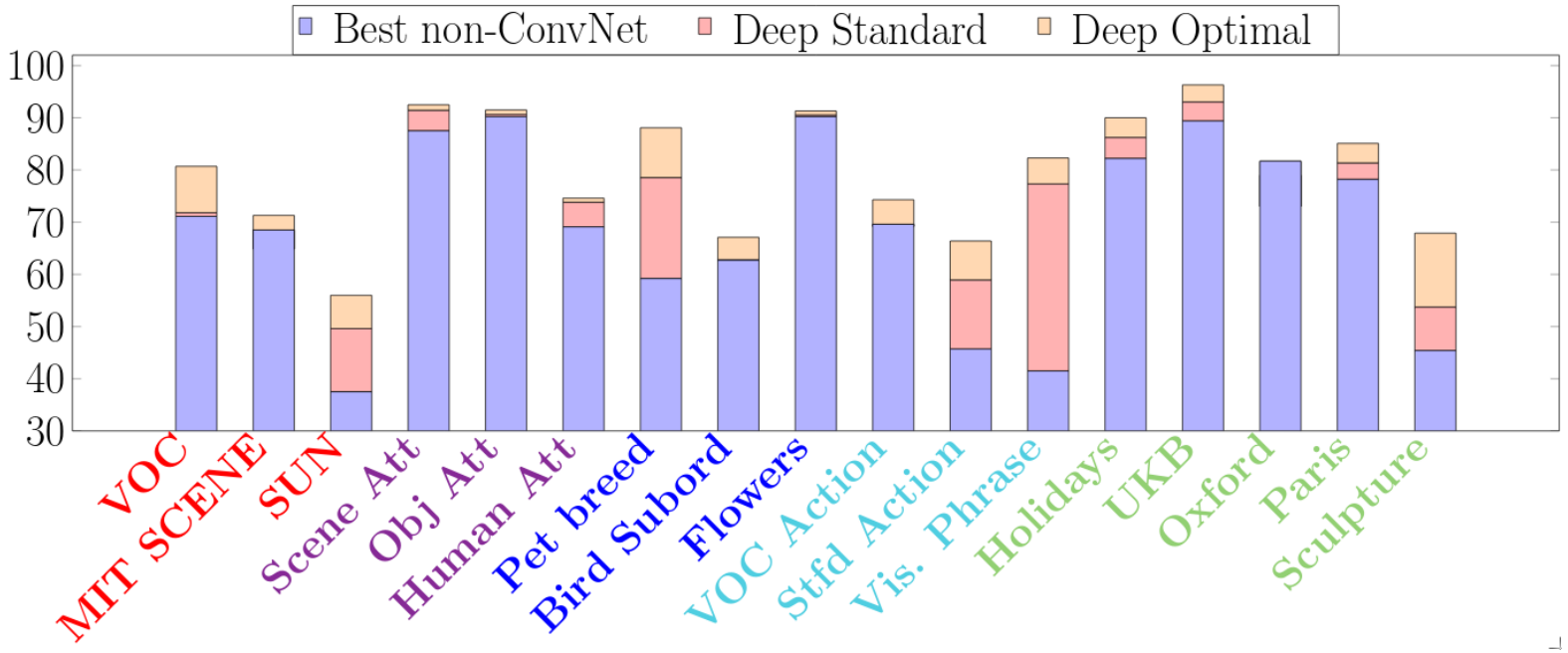
DF: fine-tuning



Deep Features (DF) and Domain Adaptation

Increasing distance from ImageNet →

Image Classification	Attribute Detection	Fine-grained Recognition	Compositional	Instance Retrieval
PASCAL VOC Object [9]	H3D human attributes [6]	Cat&Dog breeds [29]	VOC Human Action [9]	Holiday scenes [17]
MIT 67 Indoor Scenes [33]	Object attributes [10]	Bird subordinate [43]	Stanford 40 Actions [46]	Paris buildings [31]
SUN 397 Scene [45]	SUN scene attributes [30]	102 Flowers [27]	Visual Phrases [34]	Sculptures [4]

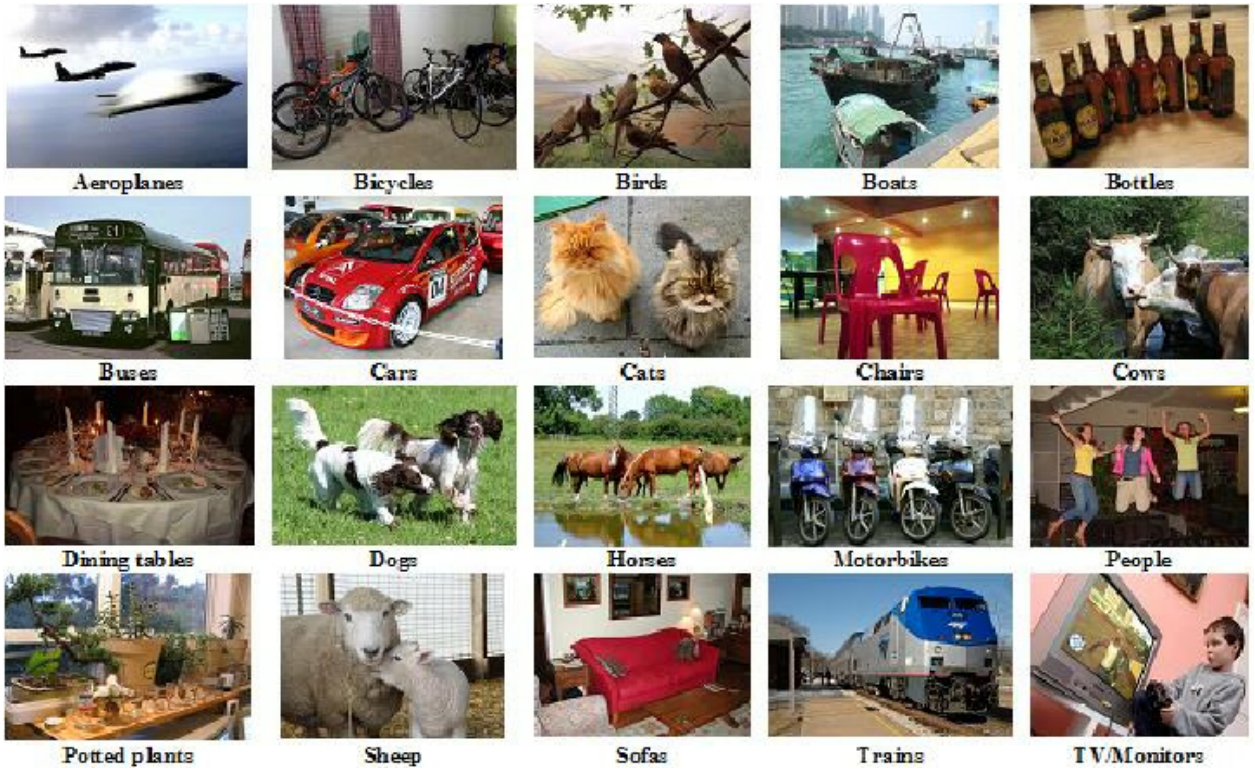


Credit: Razavian *et. al.* [ARS+16]

Deep Features (DF) and Domain Adaptation

DF for Image classification on other datasets

- **Small size datasets**: $\sim 10^3 - 10^4$ ex, *e.g.* VOC'07 (20 classes, 5000 ex)



Model type	Test mAP
From Scratch	39.79
BoW [ATC+13]	61.6
Transfer	83.22
Fine Tuning	**85.70**

Table : **Mean Average Precision.**

- Fine Tuning > Transfer >> Handcrafted (BoW)
- From scratch does not work (not enough training data)

Deep Features (DF) and Domain Adaptation

DF for Image classification on other datasets

- **Medium size datasets**, *e.g.* Food (LIP6) : 101 classes, 10^5 ex

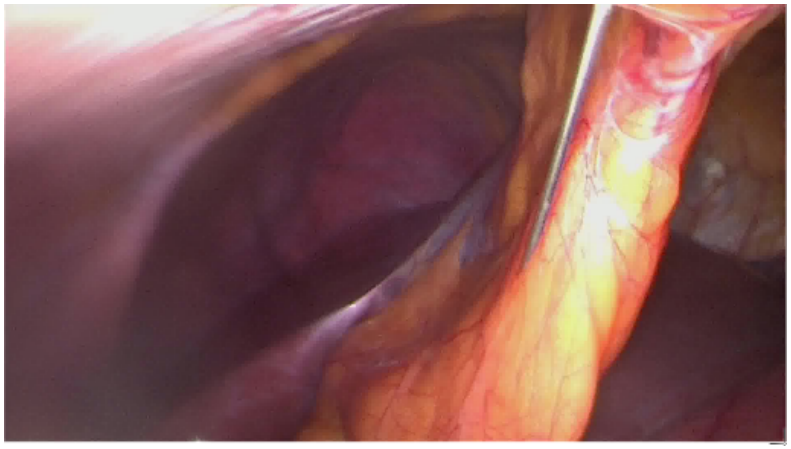
Modèle	Test top 1 (%)
(a) Bag of visual Words	23.96
Overfeat & Extraction	33.91
Overfeat & From Scratch	47.46
Overfeat & Fine Tuning	57.98
(b) Vgg16 & Extraction	40.21
Vgg16 & From Scratch	53.62
Vgg16 & Fine Tuning	**65.71**

- From scratch DOES work (well !)
- Fine Tuning >> From scratch >> Transfer >> Handcrafted (BoW)

Deep Features (DF) and Domain Adaptation

DF for Image classification on other datasets

- Another ex: M2CAI'16 challenge - large domain shift (medical images)
  - Medium-size: 22 videos, $\sim 60 \cdot 10^4$ images, 8 classes

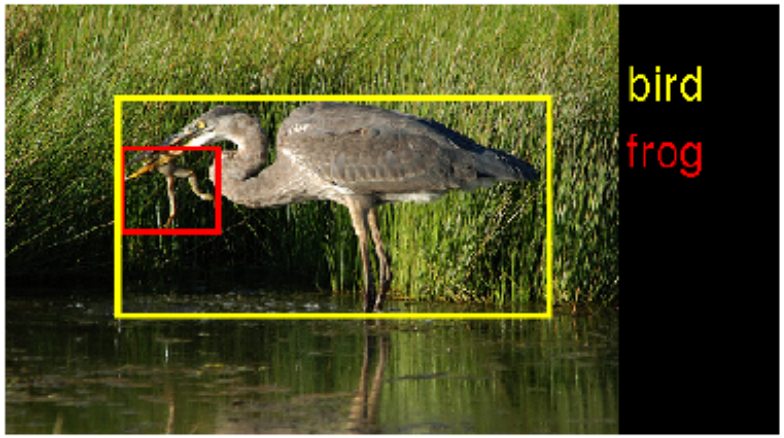


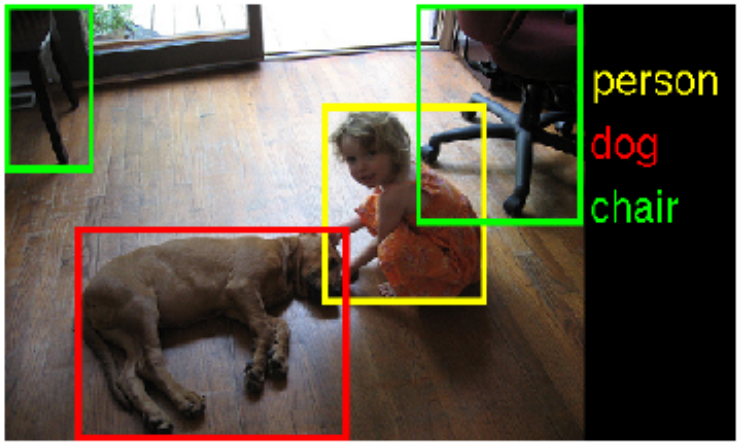
Model	Acc Top1(%)
Transfer	59.27
From Scratch	69.13
Fine Tuning	79.06

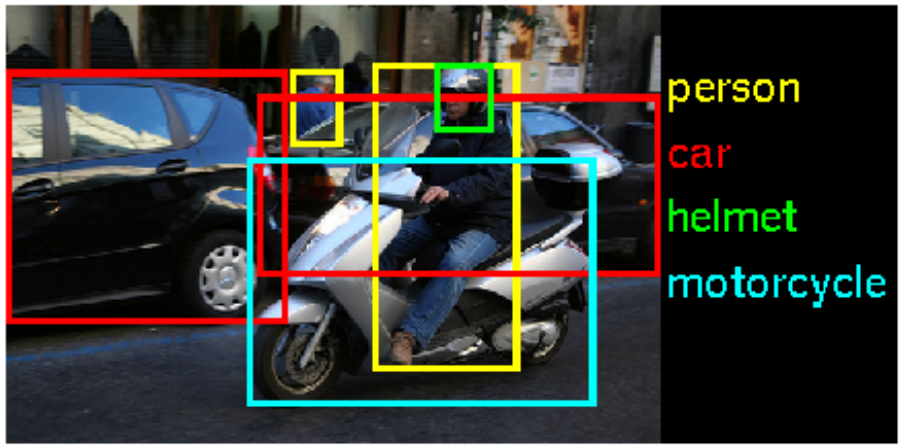
- Fine Tuning >> From scratch >> Transfer
- Transfer already good baseline despite big visual content shift

Deep Features (DF) and Domain Adaptation

Task Adaptation: Localization



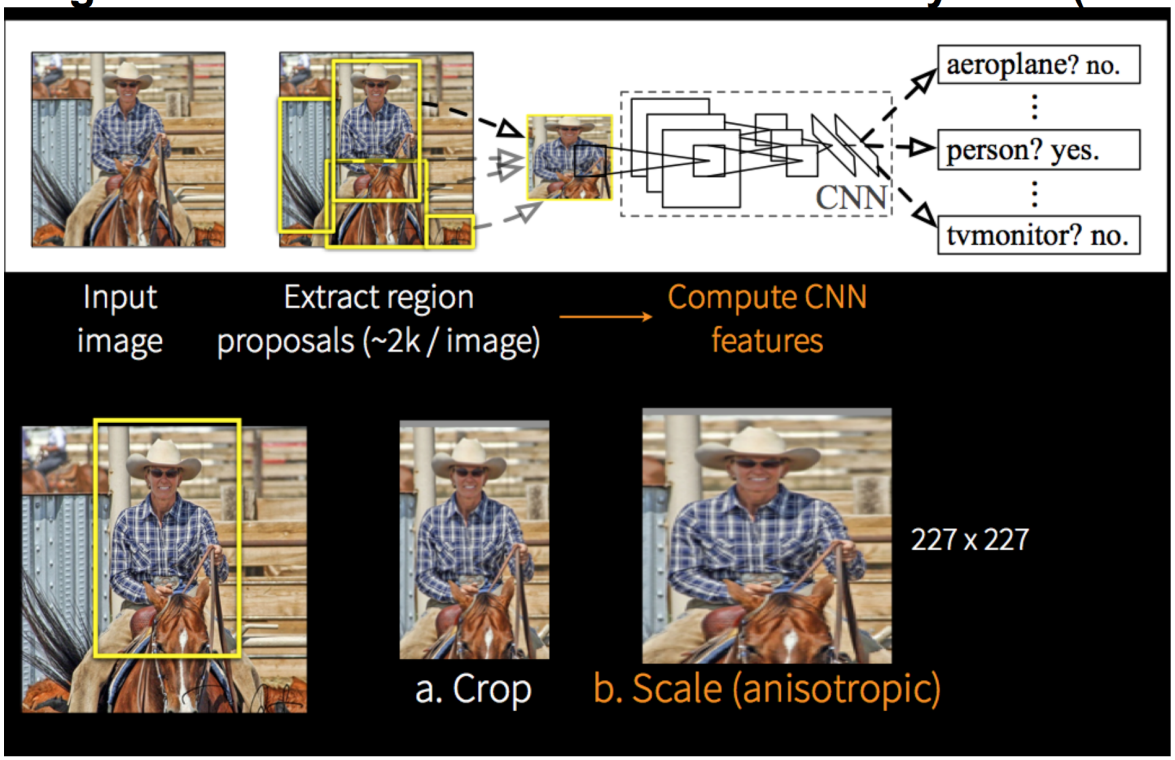




Deep Features (DF) and Domain Adaptation

Task Adaptation: Localization

Regions with Convolutional Neural Net.s system (RCNN)



R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *CVPR* 14

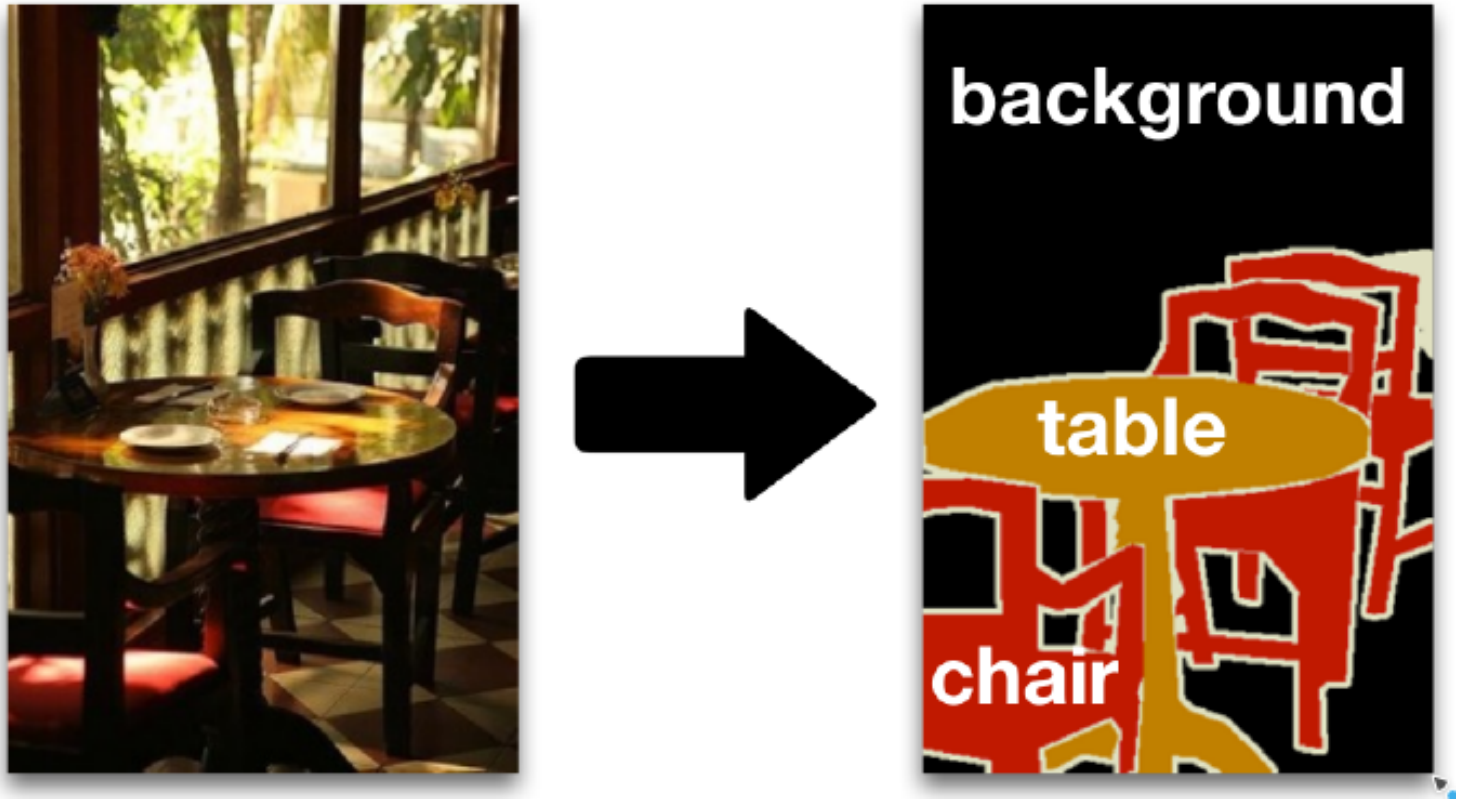
Eval on VOC'07:

R-CNN	**58.5**
DPM HoG	34.3

- R-CNN ⇒ region proposals ⇒ Deep Features ⇒ classify
- Significantly outperformed previous models (DPM on HoG features)

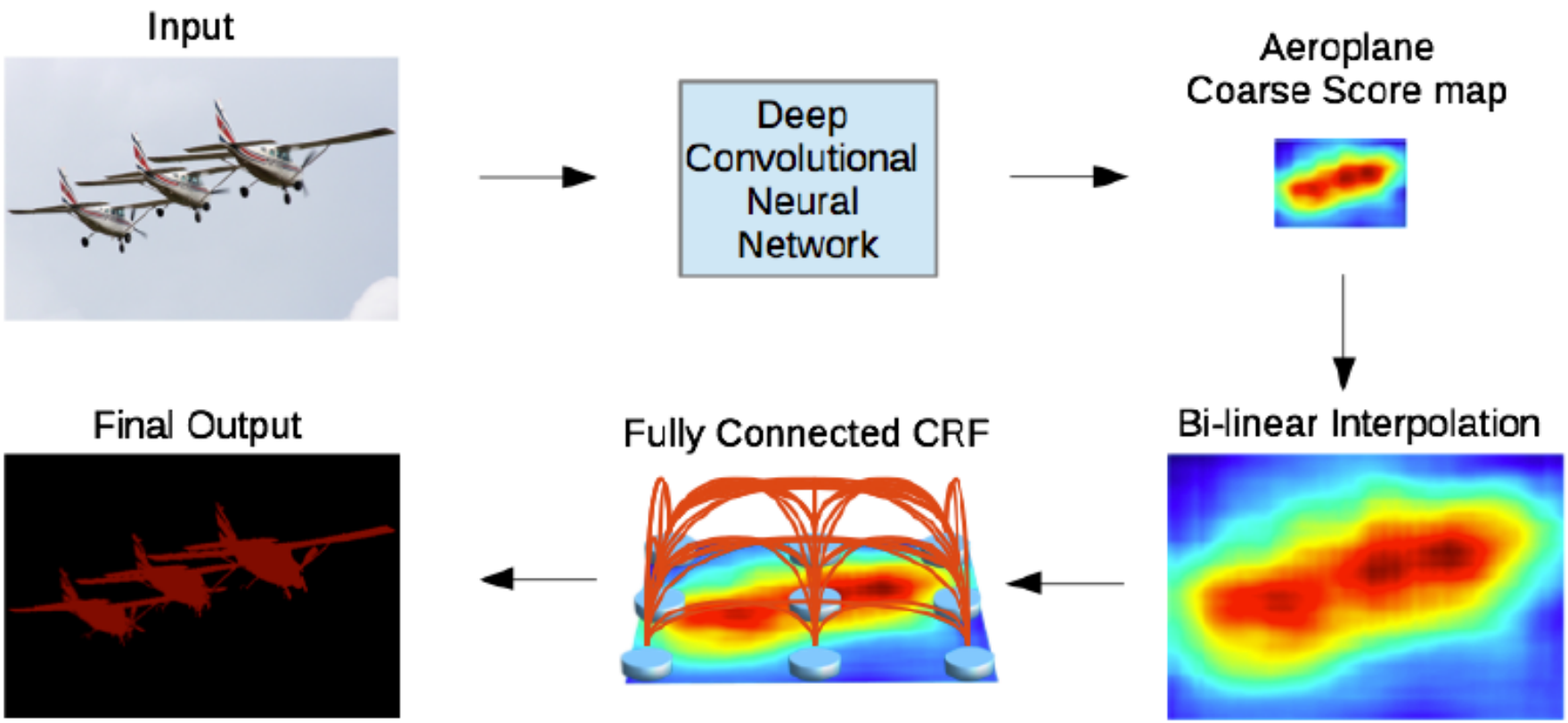
Deep Features (DF) and Domain Adaptation

Task Adaptation: Semantic Segmentation



Deep Features (DF) and Domain Adaptation

Task Adaptation: Semantic Segmentation

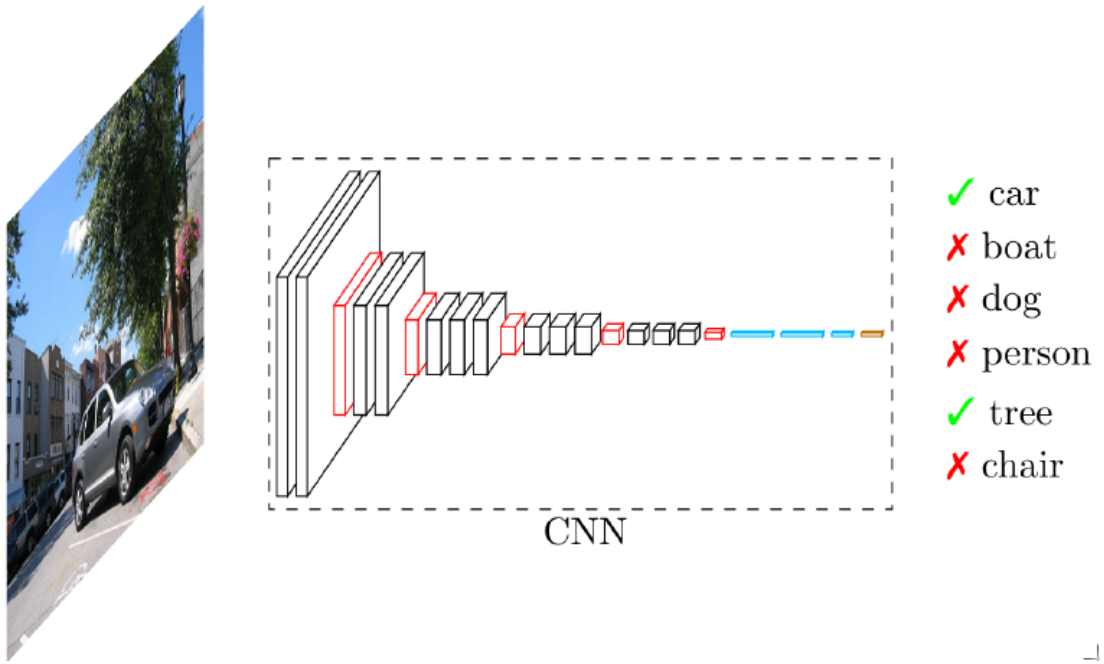


Chen *et.al.* ICLR'15

Deep Features (DF) and Domain Adaptation

Conclusion

- Deep Feature in transfer mode
  - Very good baseline
  - Ofter > descriptors based on expert knowledge
  - The solution for small databases
- From medium-size: from scratch possible and very competitive
- Fine-tuning always improves performances (small → large datasets)



References I


Hossein Azizpour, Ali Sharif Razavian, Josephine Sullivan, Atsuto Maki, and Stefan Carlsson, *Factors of transferability for a generic convnet representation*, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2016), no. 9, 1790–1802.


Min Lin, Qiang Chen, and Shuicheng Yan, *Network in network*, CoRR abs/1312.4400 (2013).