

# Neural Networks and Deep Learning: Deep Features

**Nicolas Thome**

Conservatoire National des Arts et Métiers (Cnam)  
Département Informatique

# Training Deep ConvNets on Small Datasets

- ▶ Ex: PASCAL VOC'07: 20 categories, 5000 training samples



Aeroplanes



Bicycles



Birds



Boats



Bottles



Buses



Cars



Cats



Chairs



Cows



Dining tables



Dogs



Horses



Motorbikes



People



Potted plants



Sheep



Sofas



Trains

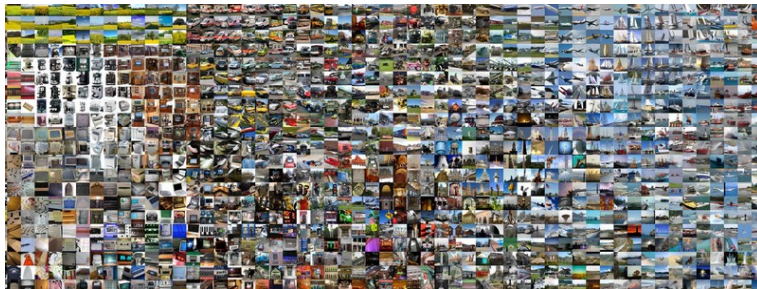


TV/Monitors

- ▶ Training Deep model from scratch (VGG) vs Handcrafted BoW (FV) [Perronnin et al., 2010]
- ▶ Deep « Handcrafted

| Model | Test mAP (%) |
|-------|--------------|
| VGG   | ≈ 40         |
| FV    | ≈ 70         |

# Training Deep ConvNets on Small Datasets



- ▶ ImageNet:  
deep » handcrafted
- ▶ VOC'07  
deep « handcrafted
  - ▶ Not enough training samples
  - ▶ Complex images



Dining tables



Dogs



Horses



Motorbikes



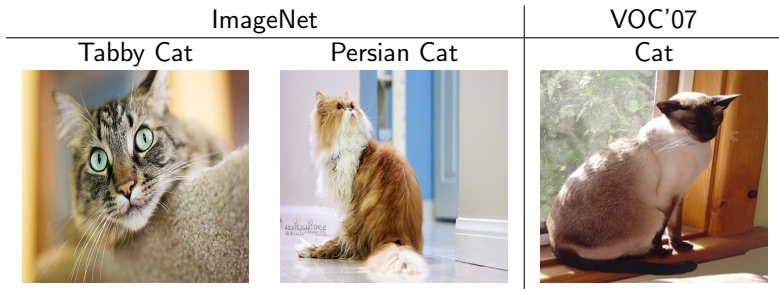
People



Deep Features

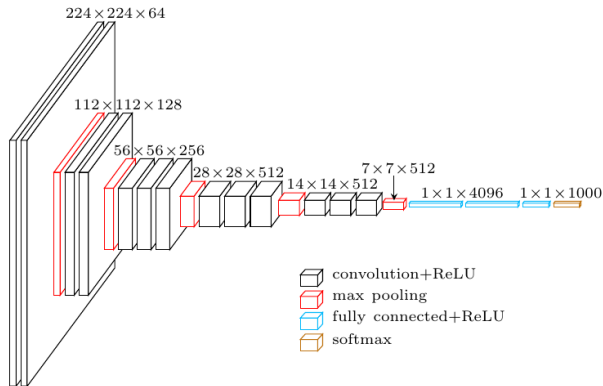
# Transfer Learning

- ▶ **Idea:** export knowledge from source domain to target domain
  - ▶ Source: good performances, e.g. many samples
  - ▶ Target: more challenging, e.g. few samples
    - ⇒ Deep ConvNet good in imagenet, but not as good in VOC'07
- ▶ **Assumption:** source and target classes different but related
  - ▶ Learned representations in ImageNet (source) relevant for VOC'07 (target)
  - ▶ Ex: Various breeds of cat (tabby, persian) vs cat



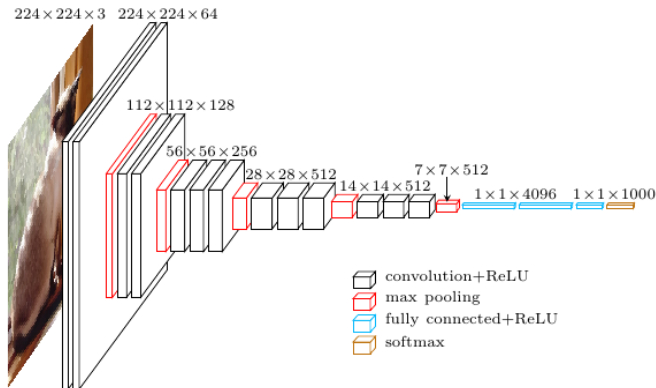
# Transferring Representations learned from ImageNet

- ▶ Most naive transfer learning approach:
  - ▶ Load ConvNet model pre-trained on ImageNet, e.g. VGG

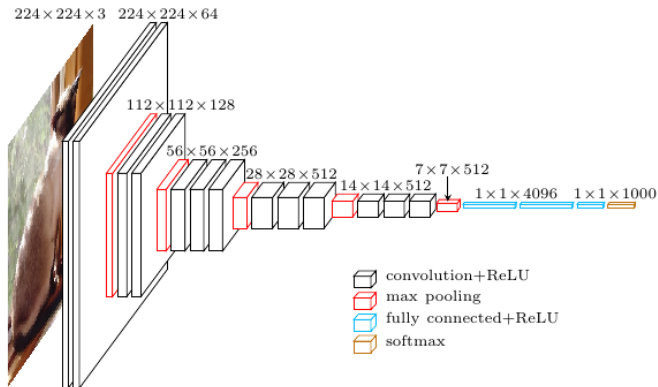


# Transferring Representations learned from ImageNet

- ▶ Most naive transfer learning approach:
  - ▶ Load ConvNet model pre-trained on ImageNet
  - ▶ Apply ConvNet on each target dataset image, e.g. VOC
  - ▶ Extract a given layer activation: **"Deep Features" (DF)**

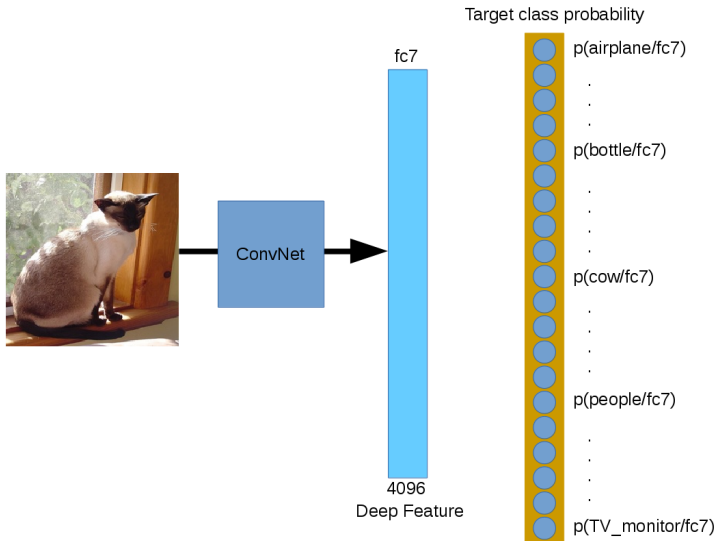


# Deep Features (DF) for Classification



- ▶ **Deep Feature (DF), e.g. fc7:** use it as any visual descriptor
- ▶ Fc7 (4096 before classif): relevant features for discriminating classes related to target class, e.g. tabby/persian cat for cat

# Deep Features (DF) for Classification



► **Deep Feature (DF), e.g. fc7: use it as any visual descriptor**

- DF: very non-linear feature extractor
- Can use any Machine Learning flat model for target class prediction

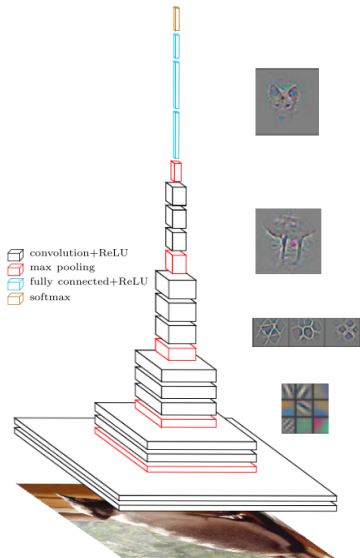


# Deep Features (DF) for Classification

## Which layer to use for classification ?

- Layers close to classification: specific to ImageNet
- Layers less close to classification: more generic features

⇒ Dependent on semantic similarity between target task & ImageNet



## Deep Features (DF) for localized visual tasks



- Localized visual tasks: localization, segmentation, retrieval
- Use DF to describe image region content

# Deep Features (DF) for Visual Recognition

- ▶ **Deep Features: ConvNet success beyond ImageNet**  
⇒ No need huge dataset for using / training deep models
- ▶ **Off-the-shelf features for any visual recognition task**

Increasing distance from ImageNet  
↓

## Image Classification

PASCAL VOC Object [9]  
MIT 67 Indoor Scenes [33]  
SUN 397 Scene [45]

## Attribute Detection

H3D human attributes [6]  
Object attributes [10]  
SUN scene attributes [30]

## Fine-grained Recognition

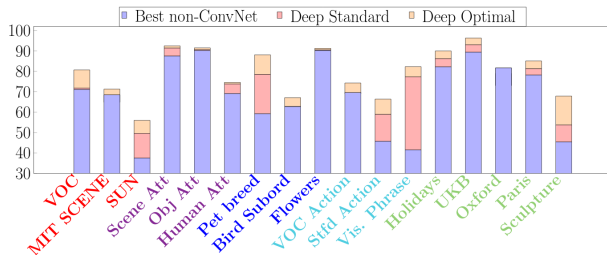
Cat&Dog breeds [29]  
Bird subordinate [43]  
102 Flowers [27]

## Compositional

VOC Human Action [9]  
Stanford 40 Actions [46]  
Visual Phrases [34]

## Instance Retrieval

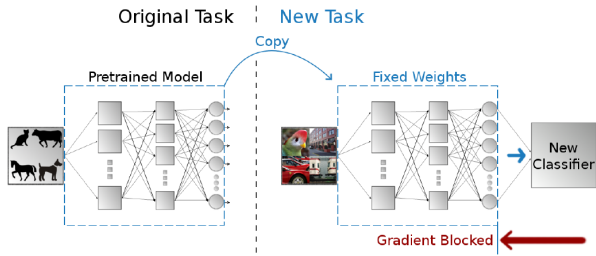
Holiday scenes [17]  
Paris buildings [31]  
Sculptures [4]



Credit: Razavian et. al. [Azizpour et al., 2016]

# Deep Features (DF) for Visual Recognition: Conclusion

- ▶ **Off-the-shelf features for any visual recognition task**
- ▶ ImageNet: 1000 classes, large set of visual concepts
  - ▶ Transfer very well even to task with large domain shift, e.g. medical, images
- ▶ **How implement transfer for classification & localized tasks?**  
⇒ following!



# References I



**Azizpour, H., Razavian, A. S., Sullivan, J., Maki, A., and Carlsson, S. (2016).**

**Factors of transferability for a generic convnet representation.**  
*IEEE Trans. Pattern Anal. Mach. Intell.*, 38(9):1790–1802.



**Perronnin, F., Sánchez, J., and Mensink, T. (2010).**

**Improving the fisher kernel for large-scale image classification.**  
In *Computer Vision—ECCV 2010*, pages 143–156. Springer.