

Neural Networks and Deep Learning: Localization & Segmentation

Nicolas Thome

Conservatoire National des Arts et Métiers (Cnam)
Département Informatique

Deep Features: Domain Adaptation for Localized Tasks



From [Noh et al., 2017]



From [Cao et al., 2017]

- ▶ Local information needed: various applications, e.g. localization, segmentation, retrieval, pose estimation, *etc*

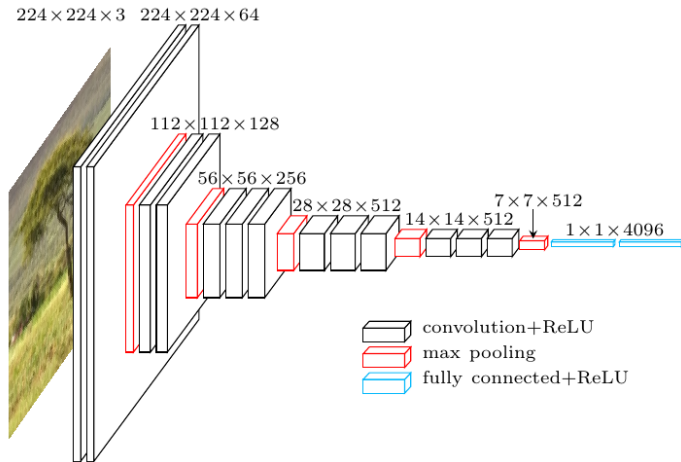
Deep Features for Localized Tasks



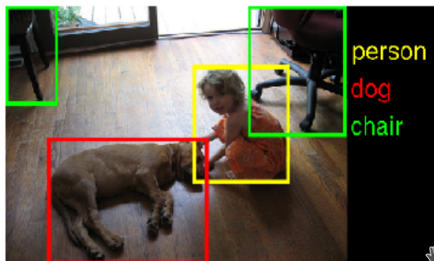
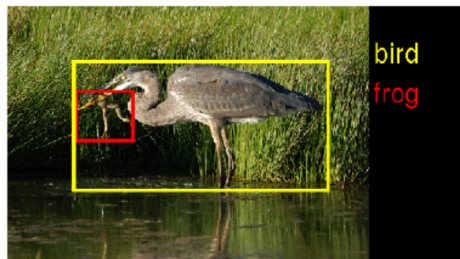
- ▶ Core (simple) idea: deep features for local information in image regions
 - ▶ Crop given image sub-area
 - ▶ Rescale \rightarrow ImageNet input size, e.g. 224×224

Deep Features for Localized Tasks

- Core idea: deep features for local information in image regions
 - Extract Deep Features with ConvNet pre-trained on ImageNet



Example: Object Localization



- ▶ **Object Localization:** rectangular Bounding Box (BB) around each object in the image
- ▶ **Localization as classification:** classify each region into $K+1$ (background) classes

Localization with Region-CNN (R-CNN) [Girshick et al., 2014]

1. R-CNN, 1st step: extract a set of region proposal candidates
 - Goal: pre-select candidates based on their "objectness"
 - Low-level, unsupervised
 - Many approaches, e.g. selective search [Uijlings et al., 2013]



Localization with Region-CNN (R-CNN) [Girshick et al., 2014]

2. R-CNN, 2nd step: classify each regions proposal

- Rescale proposal & extract deep feature
- Add transfer layer with $K + 1$ classes
 - +BB regression, *i.e.* remap proposal (red) → GT BB (green)



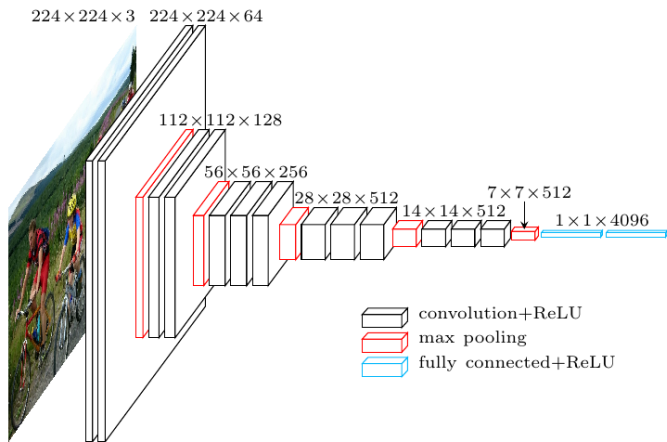
Semantic Image Segmentation

- ▶ **Label each image pixel into $K + 1$ (background) classes**
- ▶ Extract deep features on regions centered at each pixel (cf localization)?
 - ▶ Naive solution very inefficient , does not scale!
 - ▶ Ex: 500×500 image \Rightarrow 25000 regions with a single scale!



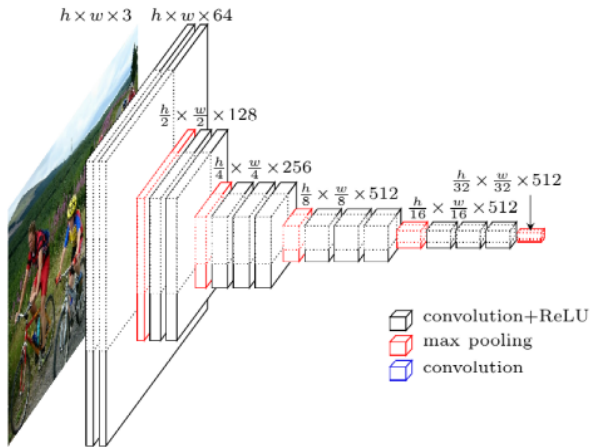
Semantic Segmentation with Fully Convolutional Networks

- ▶ 224×224 input image: apply [Conv-FC], e.g. VGG



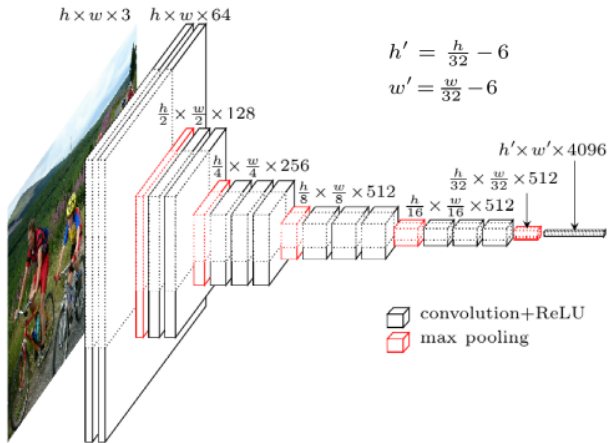
Semantic Segmentation with Fully Convolutional Networks

- ▶ Conv layer directly applicable to bigger image, size $w \times h$
- ▶ How to transfer FC layers? (direct with base FCN, e.g. ResNet)



Semantic Segmentation with Fully Convolutional Networks

- ▶ FC \Leftrightarrow conv with $7 \times 7 \times 512$ filters
- ▶ Ex: input image = 512^2 , $w' = 10$, $h' = 10$



Semantic Segmentation with Fully Convolutional Networks

- ▶ Ex: input image = 512×512 , $w' = 10$, $h' = 10$



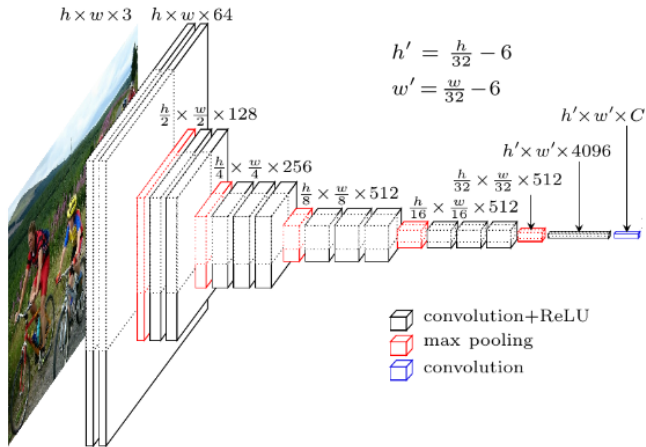
Semantic Segmentation with Fully Convolutional Networks

- ▶ Ex: input image = 512×512 , $w' = 10$, $h' = 10$
- ▶ Receptive field, features extracted \approx rescaled region and apply ConvNet



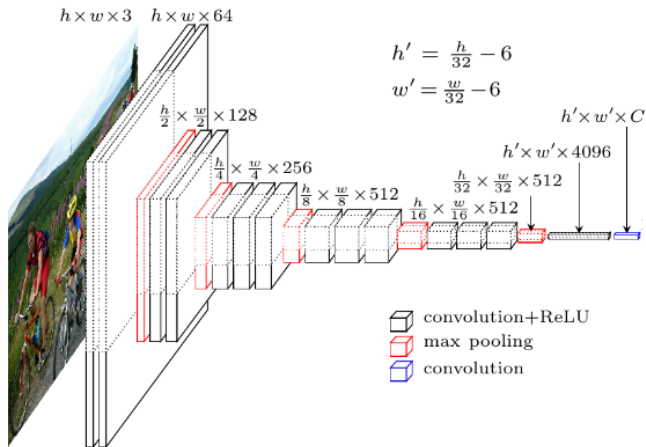
Semantic Segmentation with Fully Convolutional Networks

- ▶ Add transfer layer ($C = K + 1$ classes) to classify each of the $w' \times h'$ regions
- ▶ Fully connected layer on each region: 1×1 convolution + softmax



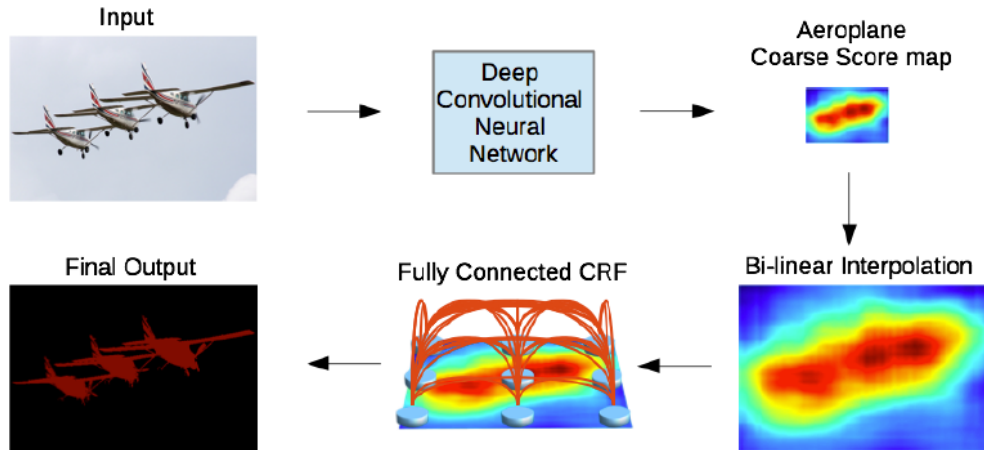
Semantic Segmentation: DeepLab [Chen et al., 2015]

- ▶ Fully Convolutional Network outputs $w' \times h' \times C$ tensor
- ▶ How to train it from $w \times h \times C$ annotations?



Semantic Segmentation: DeepLab [Chen et al., 2015]

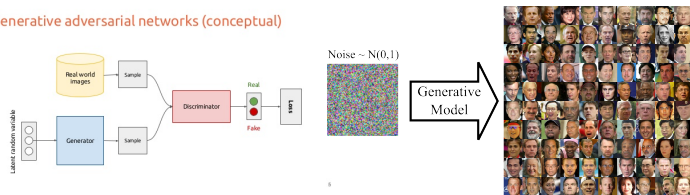
- ▶ DeepLab: simply interpolate maps $\rightarrow w \times h \times C$
- ▶ Cross-entropy loss for each pixel, CRF as post-processing



Application of ConvNets for Localized Tasks: Conclusion

- ▶ Core idea: computing deep features on regions
- ▶ Adapting architecture depending on the task
- ▶ Fully convolutional architecture key to scalability
- ▶ **Ongoing work & perspectives on unsupervised learning? \Rightarrow following!**

Generative adversarial networks (conceptual)



References I



Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y. (2017).

Realtime multi-person 2d pose estimation using part affinity fields.

In *CVPR*.



Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2015).

Semantic image segmentation with deep convolutional nets and fully connected crfs.

In *ICLR*.



Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014).

Rich feature hierarchies for accurate object detection and semantic segmentation.

In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.



Noh, H., Araujo, A., Sim, J., Weyand, T., and Han, B. (2017).

Large-scale image retrieval with attentive deep local features.

In *The IEEE International Conference on Computer Vision (ICCV)*.



Uijlings, J. R. R., van de Sande, K. E. A., Gevers, T., and Smeulders, A. W. M. (2013).

Selective search for object recognition.

International Journal of Computer Vision, 104(2):154–171.