

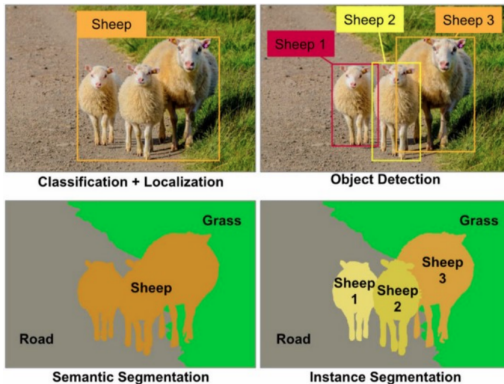
Image segmentation

Victor Kitov

v.v.kitov@yandex.ru

Types of tasks

Different image tasks



- Classification+localization could be done by extending classifier to output (x,y,h,w) of the bbox.

Image segmentation¹



- Segmentation - classification of every pixel of the image.
- Quality measures: pixel accuracy, intersection over union.

¹Picture source.

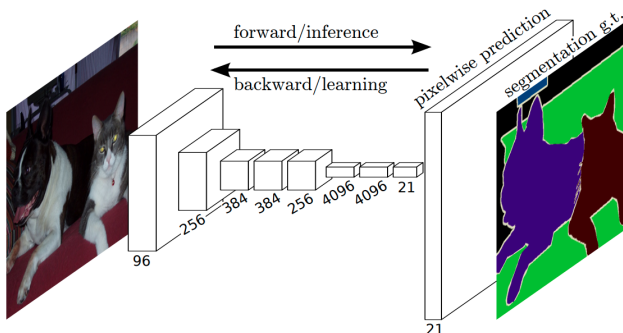
Applications

- satellite photos:
 - segment urban and agricultural areas
 - segment fields with different types of plants, their growth
- autonomous driving
 - segment people, other vehicles, signs, road obstacles
- medical applications:
 - tissues on the skin

Fully convolutional neural networks for semantic segmentation

May use VGG, ResNet as encoder, add upscaling at the end.

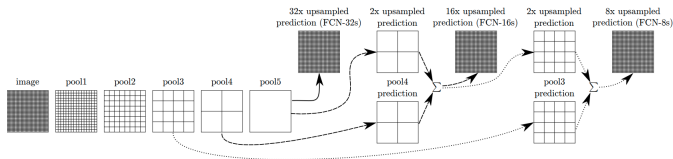
- gives rough imprecise object boundaries on rigid grid.



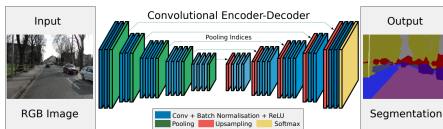
Improvement idea

- Upsampling and addition of prev. layers helps to recover both high and low level info.

Info from previous layers is used

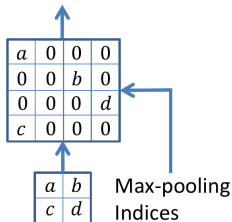


SegNet



- Encoder from VGG, ResNet, can finetune.
- Decoder - reversed encoder with pooling replaced with "unpooling" layer.

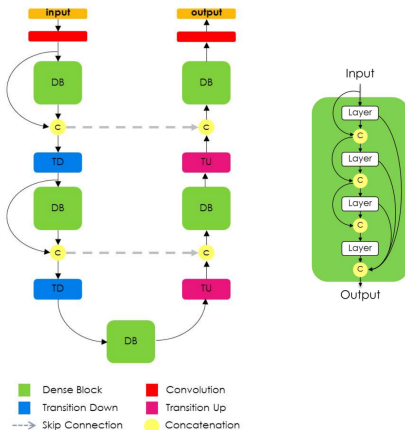
Unpooling upscased by placing max values at corresponding positions



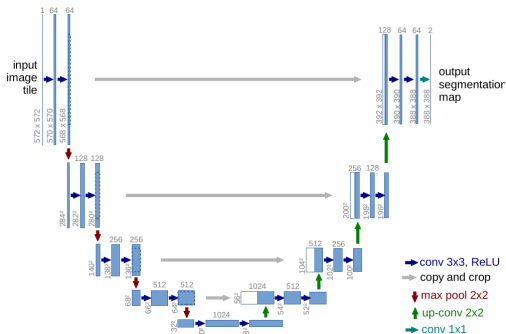
One Hundred Layers Tiramisu

Utilizes dense block - extended ResNet block with more identity connections.

Architecture and dense block



U-net architecture²



Horizontal numbers = #[channels]; vertical numbers = spatial size.
 White blocks - copied output of earlier layers; up-conv - rescaling & convolution.

²Ronneberger et al [2015].

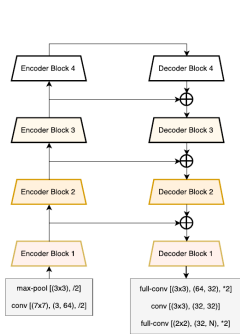
Discussion

Key ideas of U-net:

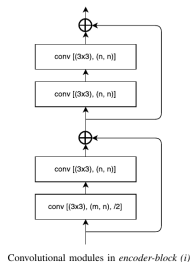
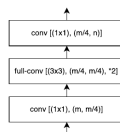
- preserve spatial info at each layer
 - use only convolution, pooling, scaling.
 - don't use vectorization & fully connected layers
- 1st half - encoder; 2nd half - decoder.
- Encoder aggregates wider and wider local information
 - creating more abstract features
- Decoder reconstructs local information from
 - more abstract features (green input on figure)
 - lower level features (gray input on figure)

LinkNet

Similar to U-net but uses summation to combine information.
ResNet blocks used in encoder. 1x1 conv used in decoder to be lightweight.

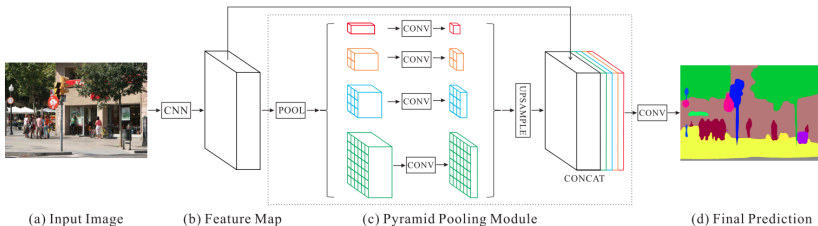


LinkNet Architecture

Convolutional modules in encoder-block (i) Convolutional modules in decoder-block (i)

PSPNet

Pretrained ResNet as CNN encoder. Pyramidal pooling & upsampling used to combine features with different resolution.



Object detection

- Need to find bboxes for arbitrary number of objects.
- Could apply CNN classifier with rolling window - too slow.
- R-CNN: apply CNN on region proposals (rescaled to std. resolution).
- Fast R-CNN: apply CNN to whole image
- Faster R-CNN: extract region proposals with CNN
- YOLO, SSD