Image segmentation with deep learning: state of the art

Etienne Decencière

MINES ParisTech
PSL Research University
Center for mathematical morphology
Fontainebleau, France

Contents

Contents

A few words about myself

- Researcher on mathematical morphology and image processing - not an astronomer
- Main current application fields:
  - Ophthalmology
  - Dermatology, cosmetology
  - Astronomy
- I have been using deep learning methods since 2015

Acronyms

- ANN = artificial neural network
- CNN = convolutional neural network

Image Segmentation with ANNs

- Computer vision has been one of the main application domains of ANNs
- Image segmentation often is an important step in an image processing work flow
- Image segmentation has been a very active deep learning research field

Image segmentation example

Historical overview of image segmentation with ANNs

- 1979: Fukushima's neocognitron [Fukushima, 1979]
- 1989: Convolutional network + backpropagation [LeCun et al., 1989]
- 2012: Segmentation competition (neuronal membranes in electron microscopy images [Ciresan et al., 2012])

Image definition

Definition: image

An image I of size $p \times q$ ($p, q \in \mathbb{N}^*$) is a function from $D = [0, \ldots p-1] \times [0, \ldots q-1]$ into $\mathbb{R}^d$ ($d \in \mathbb{N}^*$).
The set of these images is $\mathcal{I}^d$.

Examples

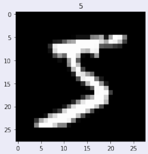


Figure: 28×28 grey level image ($d = 1$) from the MNIST dataset, and 481×321 colour image ($d = 3$) from the Berkeley segmentation dataset.

Image-to-image artificial neural networks

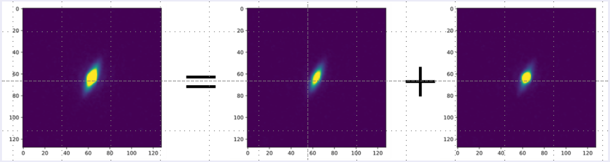
Definition: image-to-image ANNs

An image-to-image ANNs N is a ANN that transforms an image into an image of same size:

$$N : \mathcal{I}^{d_1} \longrightarrow \mathcal{I}^{d_2}$$
$$I \longmapsto N(I)$$

Note that the dimensions d_1 and d_2 of the value spaces can be different.

Example: bulge / disk decomposition



(Credits: Tuccillo, Huertas-Company, Velasco-Forero, Decencière)

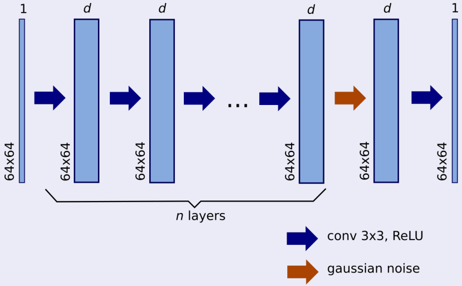
Image-to-image ANNs architecture

- Image-to-image ANNs are based on convolutional layers
- If downsampling is used, the corresponding upsampling is needed
- The receptive field of the network is an essential property

Image-to-image ANNs architecture

- Image-to-image ANNs are based on convolutional layers
- If downsampling is used, the corresponding upsampling is needed
- The receptive field of the network is an essential property

Example: Pang network [Pang et al., 2010]



Receptive field

Definition: links between neurons

In an ANN, we say that neuron a is linked to neuron b if there is an oriented path in the corresponding graph going from a to b.

Definition

The receptive field of a neuron in an ANN is the set of input neurons that are linked to that neuron.
The size of the receptive field is an essential property when designing an ANN architecture.

The specific case of image segmentation

Definition: image segmentation

Let I be an image defined on D. A segmentation of I is a partition of D. In practice the regions of the segmentation should correspond to the objects in I, which is application dependant.

- A partition is often represented as a labelled image
- In order to make the segments symmetric, each one is represented by a different channel

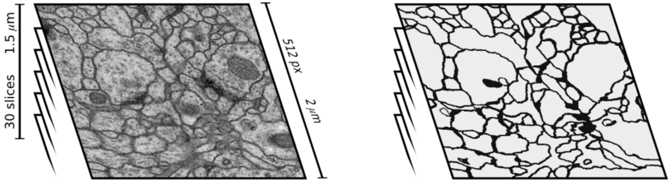
Some vocabulary on segmentation

- Object detection / localization: bounding box around the object(s).
- Binary segmentation: segmentation in 2 classes, background and object.
- Semantic segmentation: a label is given to each pixel, according to the object it belongs to.
- Instance segmentation: identify each separate object, even if they belong to the same class.

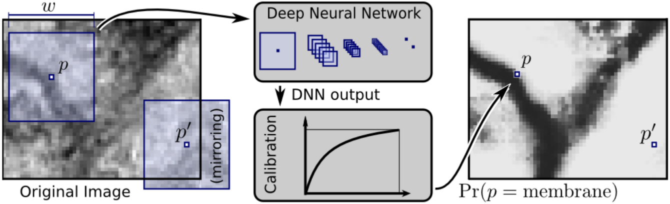
Contents

Neuron membrane segmentation challenge (ISBI 2012)

- Train: single stack of size $30 \times 512 \times 512$.
- Test: a second stack of same size.



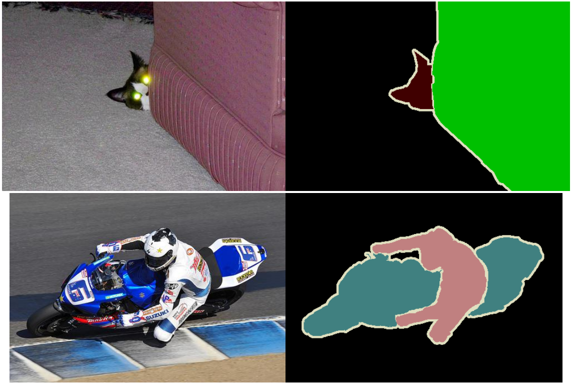
Neuron membrane segmentation challenge winner [Ciresan et al., 2012]



Contents

Pascal visual object classes segmentation challenge 2012 [Everingham et al., 2014]

- 1464 training and 1449 validation images
- automatic online test, with unknown images
- 20 image categories (cat, sofa, motorbike, person, etc.)

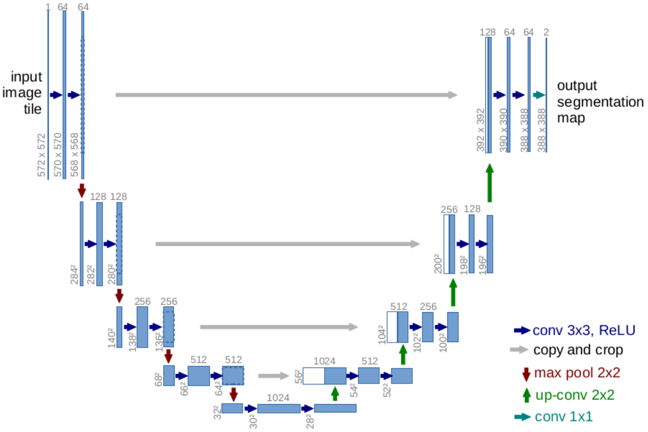


Convolutional nets for semantic image segmentation

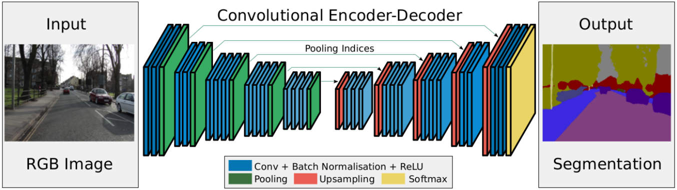
Three papers in 2015:

- Fully convolutional networks for semantic segmentation [Long et al., 2015]
- U-Net: convolutional networks for biomedical image segmentation [Ronneberger et al., 2015]
- SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [Badrinarayanan et al., 2015]

Example: U-Net architecture[Ronneberger et al., 2015]



Example: SegNet architecture[Badrinarayanan et al., 2015]



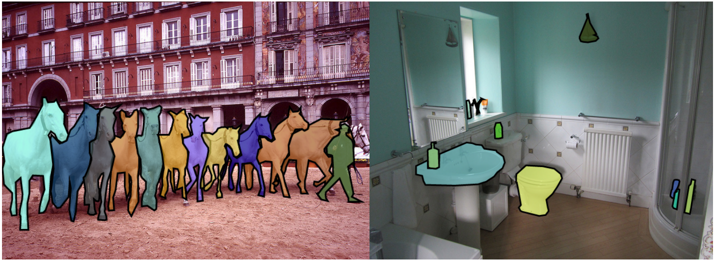
Remarks

- These architectures easily contain a number of parameters of the order of 10^7 (28 million for U-Net)
- Their optimization might be difficult
- For many segmentation applications, they are overkill

Contents

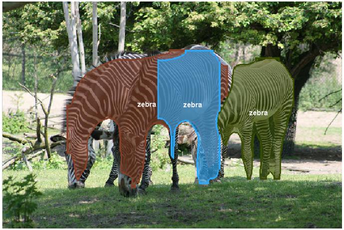
COCO: common objects in context [Lin et al., 2014]

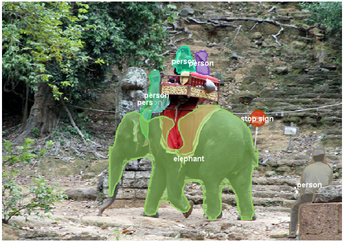
- 2 million objects, from 80 categories, in 300 000 images

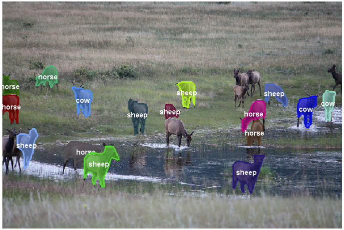


Winner 2016: Fully Convolutional Instance-aware Semantic Segmentation (Microsoft) [Li et al., 2016]

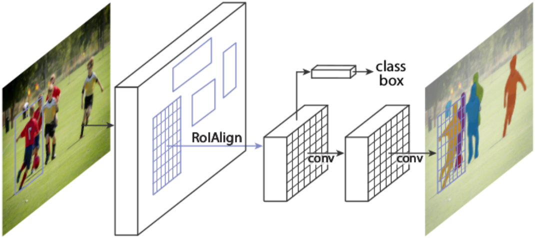
COCO instance segmentation challenge: examples of 2016 winner results



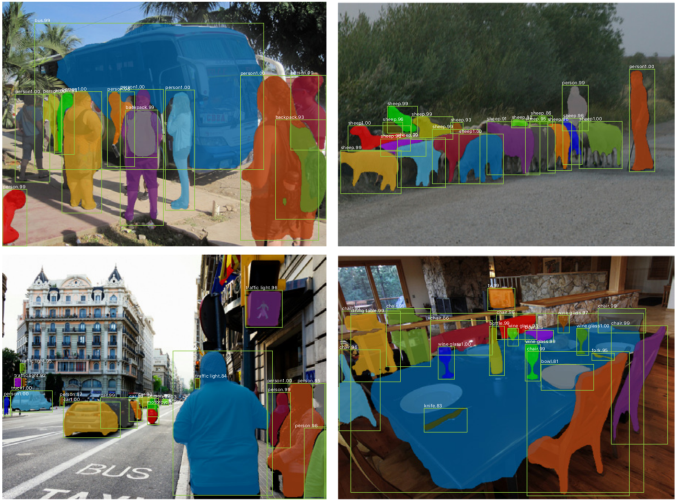




State of the art on the COCO database: Mask R-CNN [He et al., 2017]

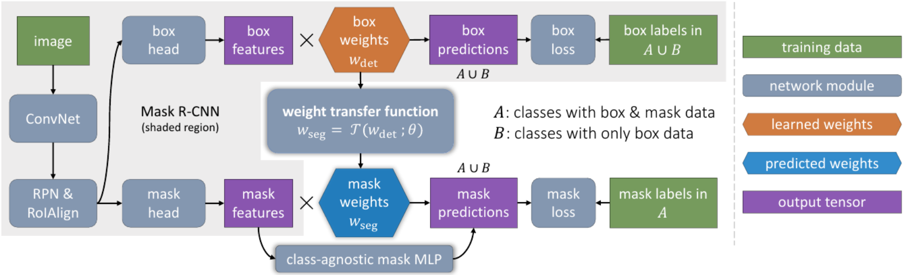


Mask R-CNN on the COCO database

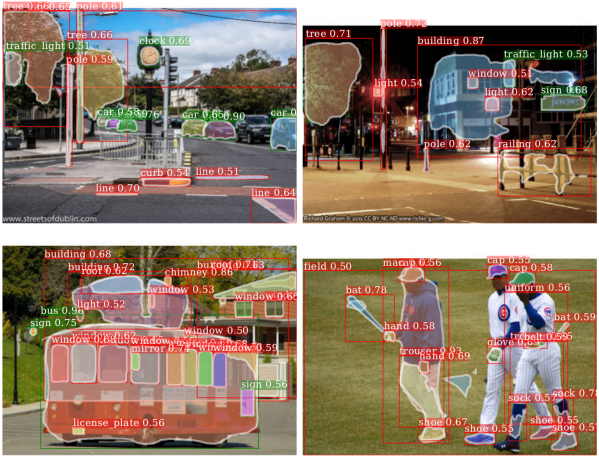


Partially supervised segmentation - [Hu et al., 2017]

- 80 segmented categories from COCO database
- 3000 visual concepts using box annotations from the Visual Genome dataset (100k images)



Partially supervised segmentation - learning to segment every thing



[Hu et al., 2017]

Current (?) trends for instance segmentation

- Region proposal +
- Fully convolutional (very deep) network +
- (Post-processing)

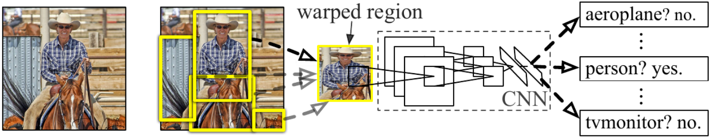


Figure: Regions with CNN features (R-CNN) (from [Girshick et al., 2014])

Current (?) trends for instance segmentation

- Region proposal +
- Fully convolutional (very deep) network +
- (Post-processing)

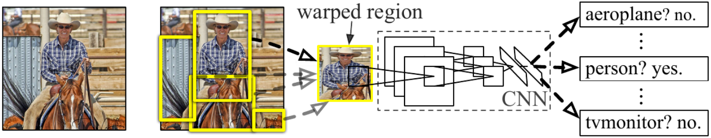


Figure: Regions with CNN features (R-CNN) (from [Girshick et al., 2014])

Meanwhile, on the object detection field...

- YOLO: you look only once [Redmon and Farhadi, 2016]
- SSD: single shot detector [Liu et al., 2016]

Contents

Modeling your problem

Casting your problem into the right representation

- Familiarize yourself with the training data (input and output images)
- Choose the right representation for your images
- Choose an architecture and train it
- Analyze the results on the validation data (look at the images!)
- Do you need preprocessing? Data augmentation? Post-processing?
- Iterate ...

Preprocessing

- Standard statistical preprocessing: not always useful, and sometimes problematic, when applied to images
- Morphological operators

Data augmentation

- Geometrical transformations: similarities
- Elastic transformations
- Specific methods: articulated objects, . . .
- Simulated data

Postprocessing for segmentation

- Superpixels (e.g. [Farabet et al., 2013])
- Conditional random fields (e.g. [Krähenbühl and Koltun, 2011])
- Mathematical morphology

Practical example

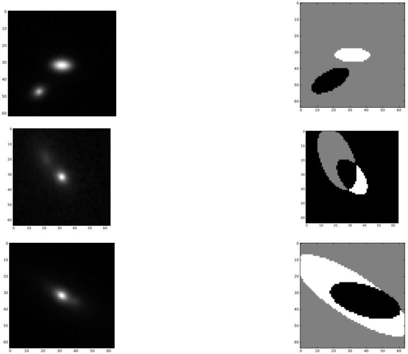
(Credits: ESA/Hubble, CC BY 4.0,

https://commons.wikimedia.org/w/index.php?curid=34205833)

How would you:

- segment the background?
- segment the sources?
- separate the sources?

Another example



(Credits: M. Huertas)

How are you representing this problem?

- semantic segmentation?
- instance segmentation?

Contents

A solved problem?

- Progress in image segmentation during the 5 last years has been enourmous
- Several complex problems have now satisfactory solutions
- Training can be a problem (large annotated databases, difficult optimization)
- Some remaining challenges:
  - Making the training database as small as possible
  - Taking *a priori* structural information into account

References I

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2015).
SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation.
arXiv:1511.00561 [cs].
arXiv: 1511.00561.

Ciresan, D., Giusti, A., Gambardella, L. M., and Schmidhuber, J. (2012).
Deep neural networks segment neuronal membranes in electron microscopy images.
In *Advances in neural information processing systems*, pages 2843–2851.

Everingham, M., Eslami, S. M. A., Gool, L. V., Williams, C. K. I., Winn, J., and Zisserman, A. (2014).
The Pascal Visual Object Classes Challenge: A Retrospective.
International Journal of Computer Vision, 111(1):98–136.

Farabet, C., Couprie, C., Najman, L., and LeCun, Y. (2013).
Learning Hierarchical Features for Scene Labeling.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8):1915–1929.

Fukushima, K. (1979).
Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position- Neocognitron.
ELECTRON. & COMMUN. JAPAN, 62(10):11–18.

References II

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014).
Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation.
pages 580–587.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017).
Mask R-CNN.
arXiv:1703.06870 [cs].
arXiv: 1703.06870.

Hu, R., Dollár, P., He, K., Darrell, T., and Girshick, R. (2017).
Learning to Segment Every Thing.
arXiv:1711.10370 [cs].
arXiv: 1711.10370.

Krähenbühl, P. and Koltun, V. (2011).
Efficient inference in fully connected crfs with gaussian edge potentials.
In *Advances in neural information processing systems*, pages 109–117.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989).
Backpropagation Applied to Handwritten Zip Code Recognition.
Neural Computation, 1(4):541–551.

References III

Li, Y., Qi, H., Dai, J., Ji, X., and Wei, Y. (2016).
Fully Convolutional Instance-aware Semantic Segmentation.
arXiv:1611.07709 [cs].
arXiv: 1611.07709.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P. (2014).
Microsoft COCO: Common Objects in Context.
arXiv:1405.0312 [cs].
arXiv: 1405.0312.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016).
SSD: Single Shot MultiBox Detector.
arXiv:1512.02325 [cs], 9905:21–37.
arXiv: 1512.02325.

Long, J., Shelhamer, E., and Darrell, T. (2015).
Fully Convolutional Networks for Semantic Segmentation.
In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440.

References IV

Pang, B., Zhang, Y., Chen, Q., Gao, Z., Peng, Q., and You, X. (2010).
Cell Nucleus Segmentation in Color Histopathological Imagery Using Convolutional Networks.
In *2010 Chinese Conference on Pattern Recognition (CCPR)*, pages 1–5.

Redmon, J. and Farhadi, A. (2016).
YOLO9000: Better, Faster, Stronger.
arXiv:1612.08242 [cs].
arXiv: 1612.08242.

Ronneberger, O., Fischer, P., and Brox, T. (2015).
U-Net: Convolutional Networks for Biomedical Image Segmentation.
In Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, number 9351 in Lecture Notes in Computer Science, pages 234–241. Springer International Publishing.