Semantic Segmentation

TINGWU WANG

MACHINE LEARNING GROUP,
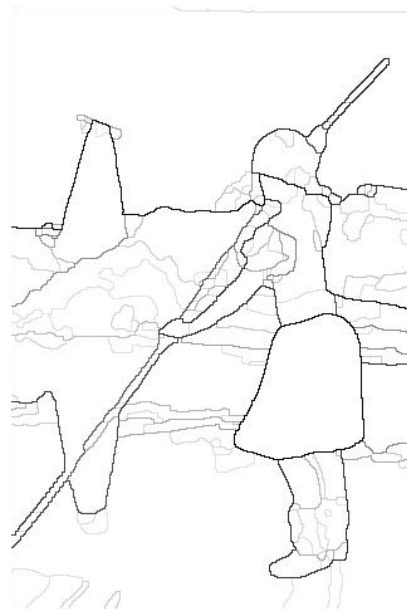
UNIVERSITY OF TORONTO

Contents

What is semantic segmentation

1. What is segmentation in the first place?
    1. Input: images
    2. Output: regions, structures
        1. line segments, curve segments, circles, etc.

What is semantic segmentation

1. What is segmentation in the first place?
   1. Input: images
   2. Output: regions, structures
   3. Most of the time, we need to "process the image"
      1. filters
      2. gradient information
      3. color information
      4. etc.

That's not quite so human.

What if we want to **understand** the image?

Arbelaez, Pablo, et al. [1]

What is semantic segmentation

1. What is semantic segmentation?
    1. Idea: recognizing, understanding what's in the image in pixel level.

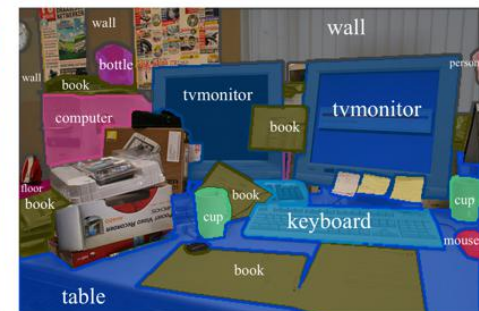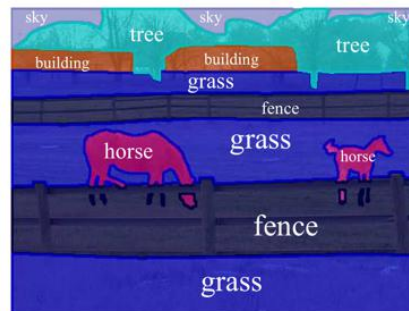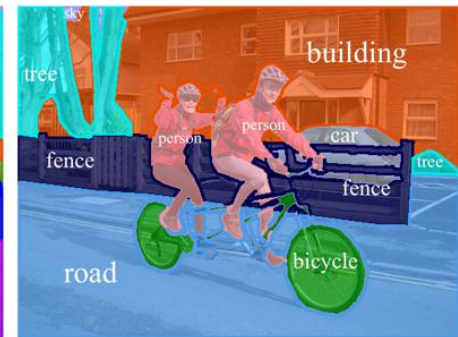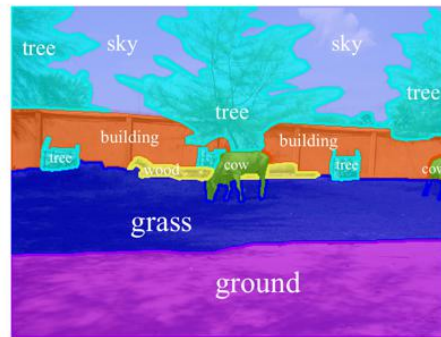   "Two men riding on a bike in front of a building on the road. And there is a car."

What is semantic segmentation

1. What is semantic segmentation?
   1. Idea: recognizing, understanding what's in the image in pixel level.
   2. A lot more difficult

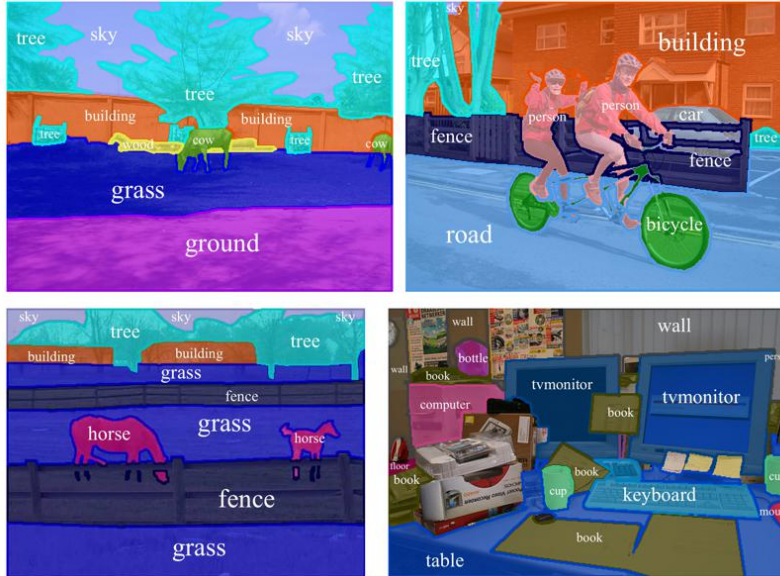   (Most of the traditional methods cannot tell different objects.)

What is semantic segmentation

1. What is semantic segmentation?
   1. Idea: recognizing, understanding what's in the image in pixel level.
   2. A lot more difficult

      (Most of the traditional methods cannot tell different objects.)

      No worries, even the best ML researchers find it very challenging.
   3. Output: regions with different (and limited number of) classes
      1. COCO detection challenge: 80 classes.
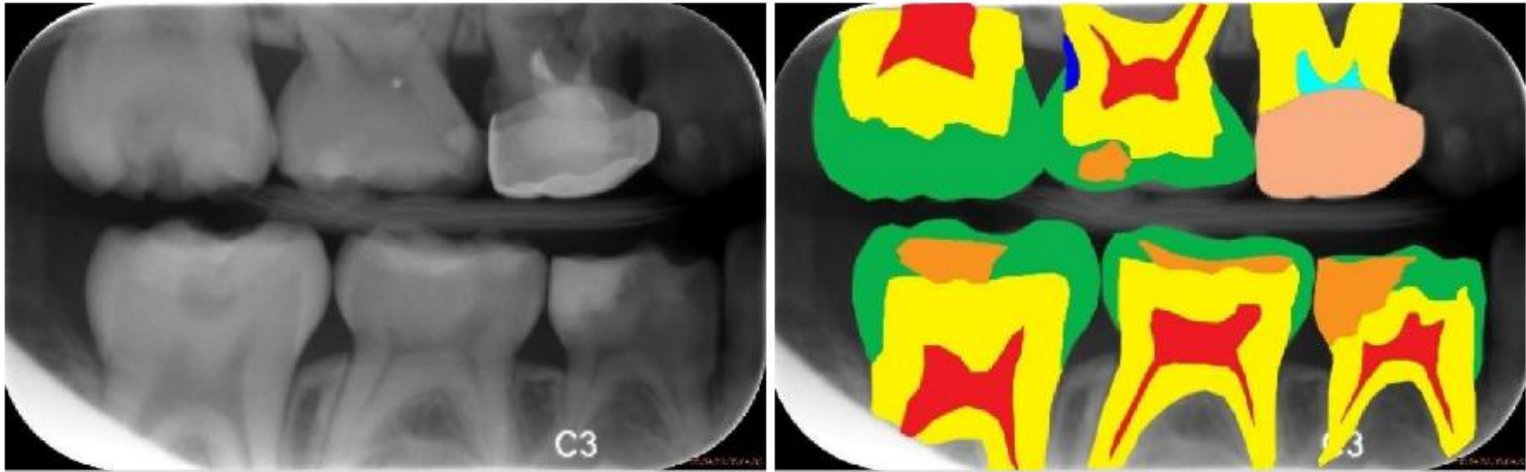   2. PASCAL VOC challenge: 21 classes

What is semantic segmentation

1. Why semantic segmentation?
    1. robot vision and understanding
    2. autonomous driving (remember your assignment?)

What is semantic segmentation

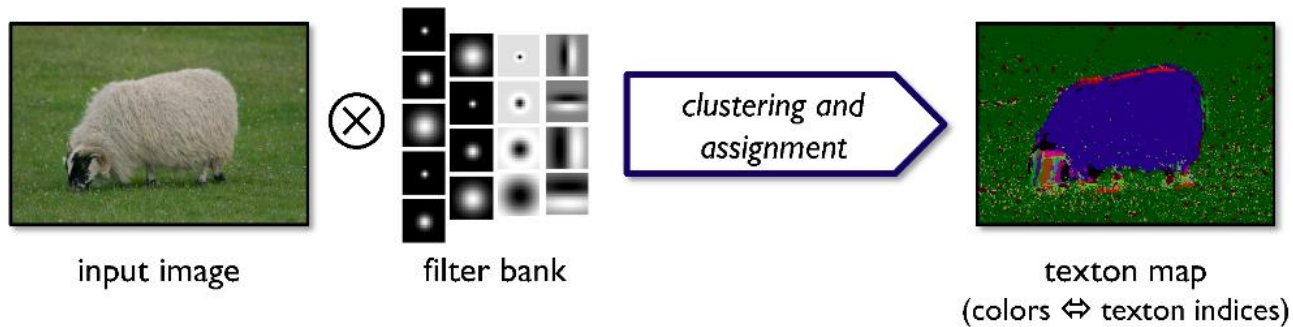1. Why semantic segmentation?
   3. medical purposes (ISBI Challenge)

Contents

Deep Learning in semantic Segmentation

1. Semantic segmentation before deep learning
   1. relying on conditional random field.
   2. operating on pixels or superpixels
   3. incorporate local evidence in unary potentials
   4. interactions between label assignments

input image ⊗ filter bank → *clustering and assignment* → texton map (colors ⇔ texton indices)

Deep Learning in semantic Segmentation

1. What is conditional random field?
   1. probabilistic framework for labeling and segmenting structured data
   2. no need to understand the math, just know the idea

   what it tries to model is the relationship between pixels, e.g.:

   1. nearby pixels more likely to have same label

   2. pixels with similar color more likely to have same label

   3. the pixels above the pixels "chair" more likely to be "person" instead of "plane"

   4. refine results by iterations

Deep Learning in Semantic Segmentation
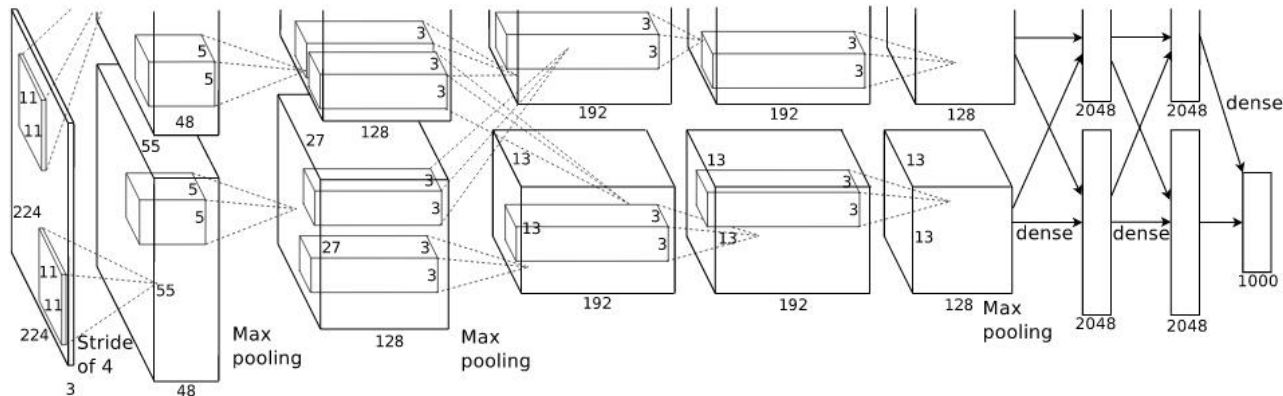
1. A Brief Review on Classification

    0. Again, it is totally fine if you don't understand the deep neural network.

        imagine it as a black magic box if you want :)
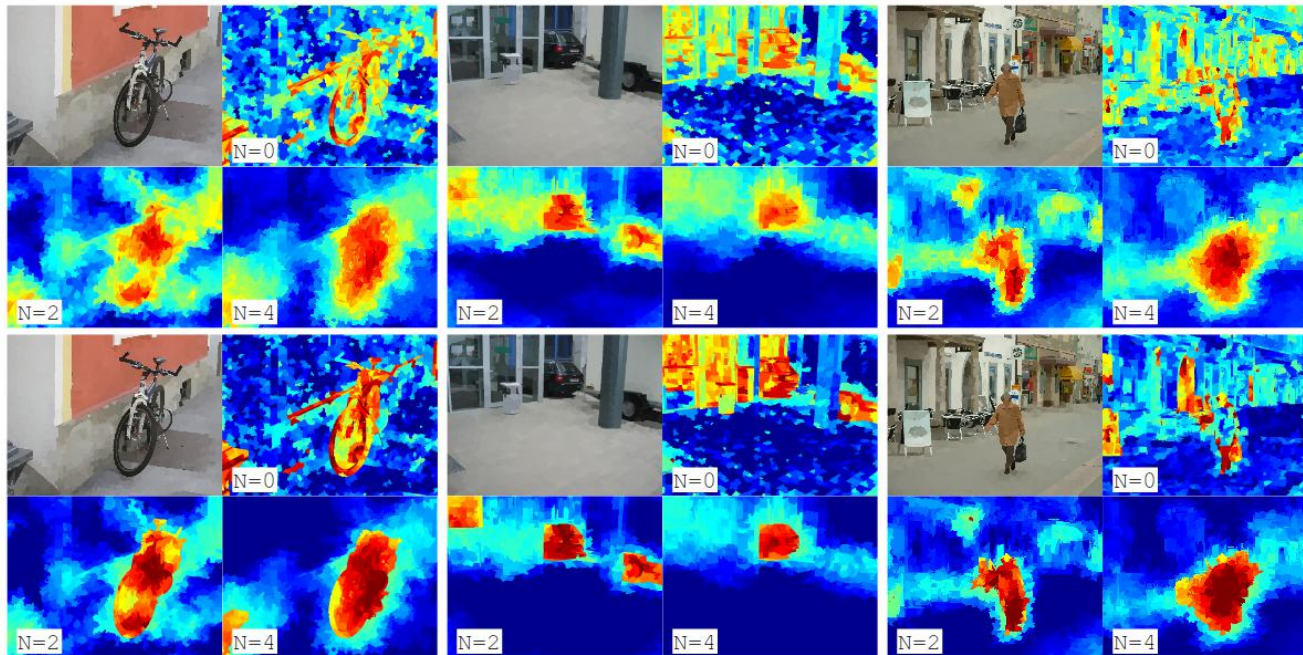
    1. Deep learning in classification.

        1. input: the whole image
        2. output: the probability of each class (person, dog, cat, ...)
        3. not appliable on semantic segmentation

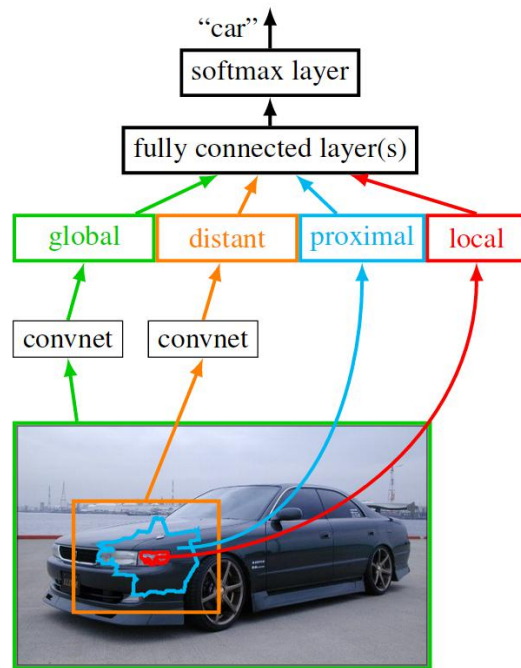A. Krizhevsky, et al. [4]

Deep Learning in Semantic Segmentation

1. How to move from classification to semantic segmentation?
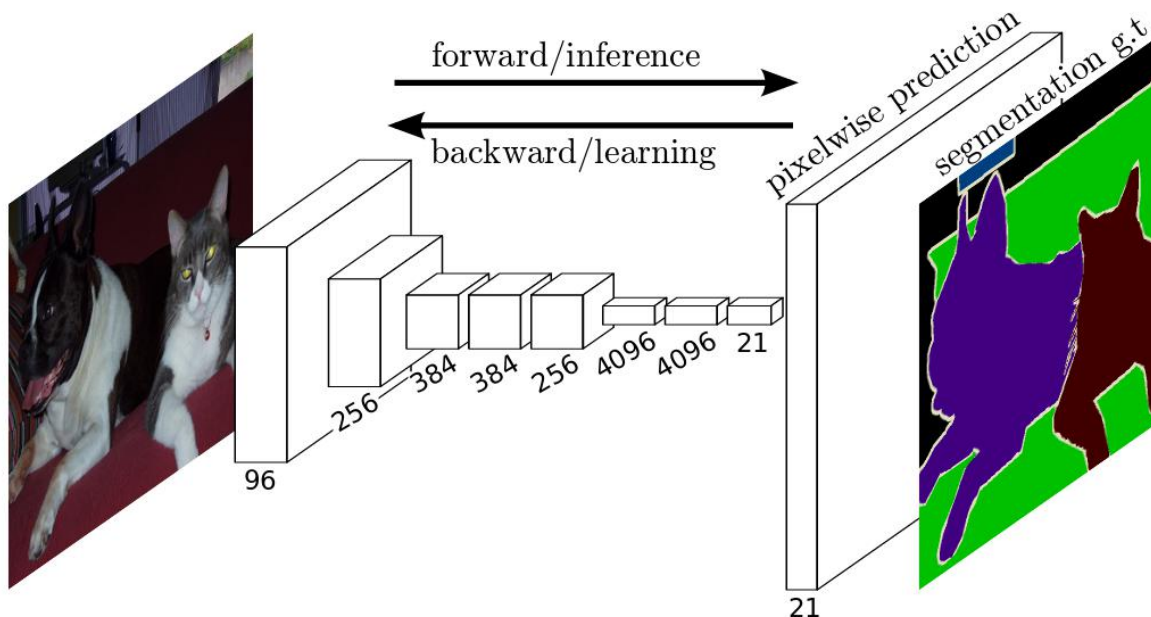   1. remember traditionally we use superpixels (Polygon)?

Deep Learning in Semantic Segmentation

1. Transition to segmentation; early ideas
   1. superpixel proposals
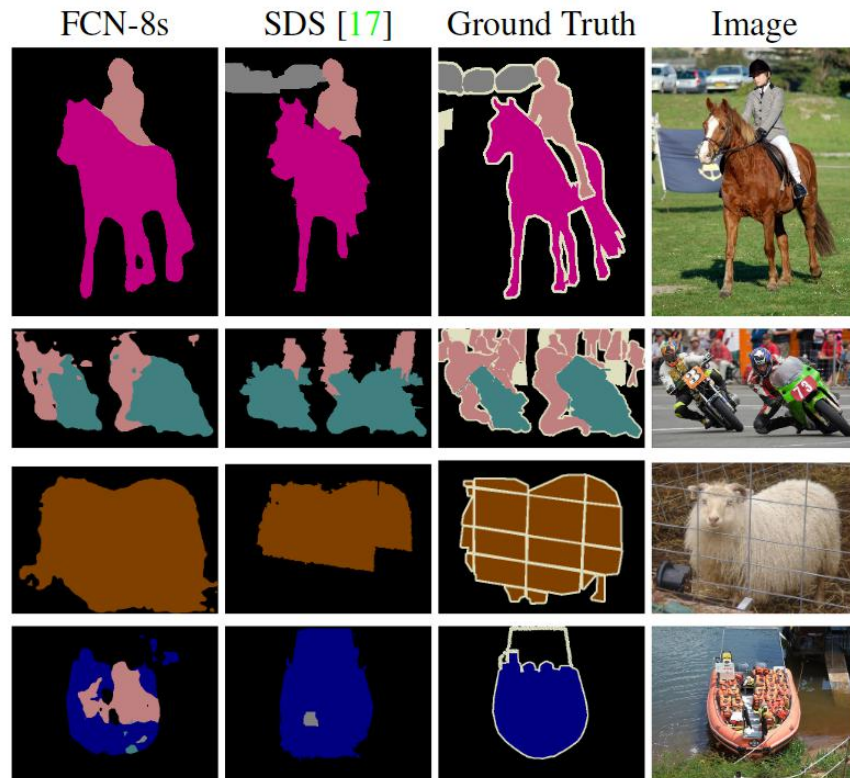   2. do classification on each superpixel.

Deep Learning in Semantic Segmentation

1. Fully Convolutional Networks for Semantic Segmentation
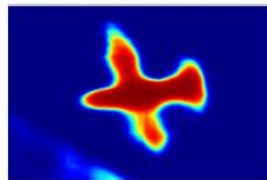   1. forget about pixels/superpixel input

Deep Learning in Semantic Segmentation

1. Fully Convolutional Networks for Semantic Segmentation

Deep Learning in Semantic Segmentation

1. Fully Convolutional Networks + CRF
   1. the output from DCNN is blurry and inaccurate
   2. rediscovery of CRF
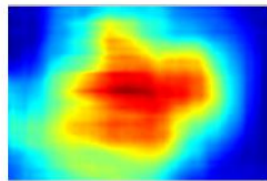
Image/G.T.   DCNN output   CRF Iteration 1   CRF Iteration 2   CRF Iteration 10
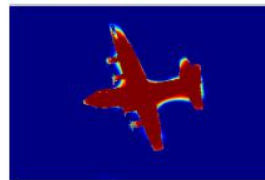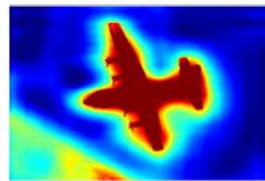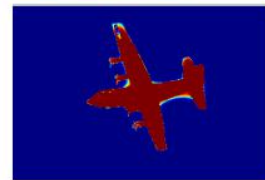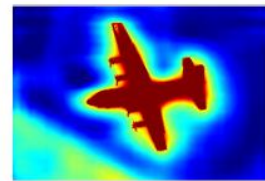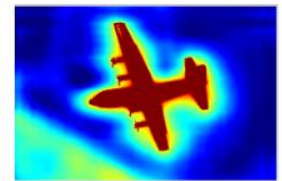
Deep Learning in Semantic Segmentation

1. Conditional Random Fields as Recurrent Neural Networks
    1. end-to-end training

        optimize(A) + optimize(B given A) < optimize(A, B together)
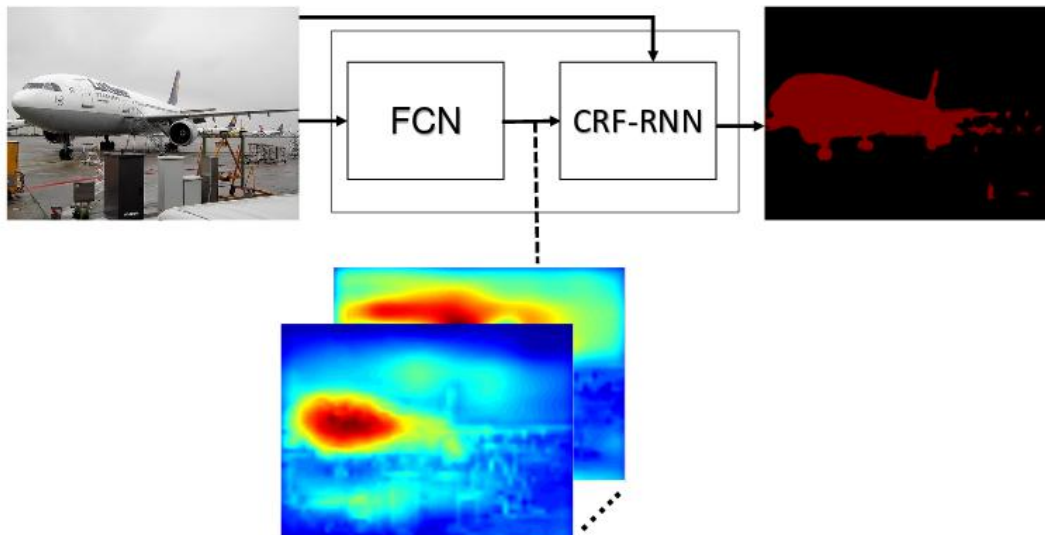
Contents

Discussions and Demos

1. Online Demos about CRF as RNN semantic segmentation

Zheng S., et al. [10]

Discussions and Demos

1. Segmentation from Natural Language Expression
   1. what does it mean?

   e.g., the phrase "two men sitting on the right bench" requires segmenting only the two people on the right bench and no one standing or sitting on another.
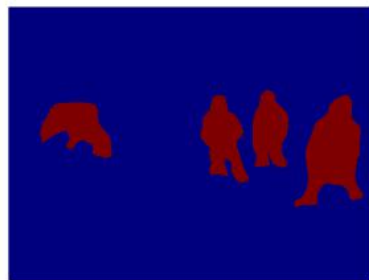
(a) input image | (b) object class segmentation of class *people* | (c) object instance segmentation of class *people* | (d) segmentation from expression *"people in blue coat"*
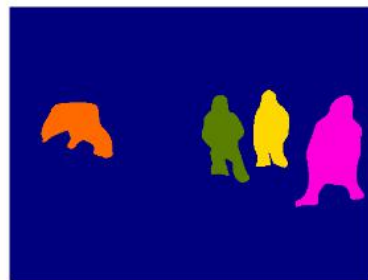
R. Hu, et al. [11]

Discussions and Demos

1. Segmentation from Natural Language Expression
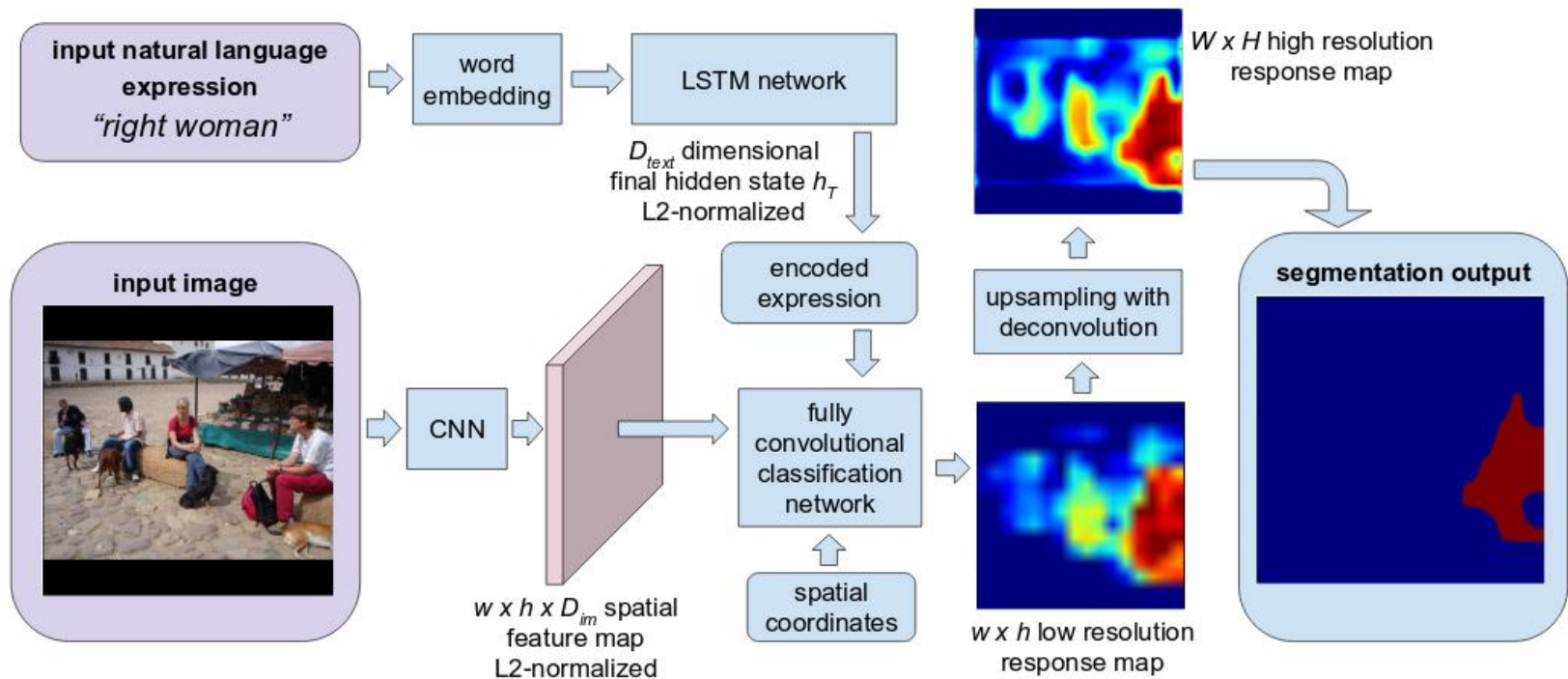
Discussions and Demos

1. Make Probabilistic Graphical Model Great Again?
   1. what happened to DPM [12]
      1. mixtures of multiscale deformable part models
      2. later people found DPM could be placed by a CNN layer [13]
      3. no one uses dpm now.
   2. what happened to object proposals in detection
      1. Human designed proposals (selective search, edge box, …) [14]
      2. later people found proposal generating could be replaced by a CNN layer [15, 16]
      3. no one (well, maybe still many people) uses human designed proposals now.

   3. what is happening to CRF in semantic segmentation
      1. pairwise relationship between pixels
      2. later people find CRF could be replaced by a CNN layer
      3. no one uses CRF? well, we don't know future

Discussions and Demos

1. The powerfulness of deep learning

Agent Smith: If you can't beat us…

Agent Smith Clone: Join us!

References

[1] Arbelaez, Pablo, et al. "Contour detection and hierarchical image segmentation." IEEE transactions on pattern analysis and machine intelligence, 2011.

[2] Roozbeh Mottaghi, Xianjie Chen, Xiaobai Liu, Nam-Gyu Cho, Seong-Whan Lee, Sanja Fidler, Raquel Urtasun, Alan Yuille. CVPR, 2014.

[3] Shotton, Jamie, et al. "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context." International Journal of Computer Vision, 2009.

[4] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[5] del Toro, Oscar Alfonso Jiménez, Orcun Goksel, Bjoern Menze, Henning Müller, Georg Langs, Marc-André Weber, Ivan Eggel et al. "VISCERAL—VISual Concept Extraction challenge in RAdioLogy: ISBI 2014 challenge organization." Proceedings of the VISCERAL Challenge at ISBI 1194 (2014): 6-15.

[6] Mostajabi, Mohammadreza, Payman Yadollahpour, and Gregory Shakhnarovich. "Feedforward semantic segmentation with zoom-out features." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[7] Fulkerson, Brian, Andrea Vedaldi, and Stefano Soatto. "Class segmentation and object localization with superpixel neighborhoods." In ICCV, 2009.

References

[8] Long, J., Shelhamer, E. and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[9] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L., 2014. Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062.

[10] Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C. and Torr, P.H., 2015. Conditional random fields as recurrent neural networks. In Proceedings of the IEEE International Conference on Computer Vision.

[11] Hu, Ronghang, Marcus Rohrbach, and Trevor Darrell. "Segmentation from Natural Language Expressions." arXiv preprint arXiv:1603.06180 (2016).

[12] Felzenszwalb, Pedro, David McAllester, and Deva Ramanan. "A discriminatively trained, multiscale, deformable part model." In Computer Vision and Pattern Recognition, 2008. CVPR.

[13] Girshick, Ross, et al. "Deformable part models are convolutional neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[14] Uijlings, Jasper RR, et al. "Selective search for object recognition." International journal of computer vision, 2013.

[15] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE International Conference on Computer Vision. 2015.

[16] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." arXiv preprint arXiv:1506.02640 (2015).

Q&A

For those who are interested in CRF and want to know the math, I recommend this tutorial:

[17] Nowozin, Sebastian, and Christoph H. Lampert. "Structured learning and prediction in computer vision." Foundations and Trends in Computer Graphics and Vision 6.3–4 (2011): 185-365.

(might take a long time to understand. good luck ;P)