

A Review on Deep Learning Techniques Applied to Semantic Segmentation

Yuan An

Image semantic segmentation is more and more being of interest for computer vision and machine learning researchers. Many applications on the rise need accurate and efficient segmentation mechanisms: autonomous driving, indoor navigation, and even virtual or augmented reality systems to name a few. This demand coincides with the rise of deep learning approaches in almost every field or application target related to computer vision, including semantic segmentation or scene understanding. The paper of Garcia-Garcia *et al.* [3] provides a review on deep learning method for semantic segmentation applied to various application areas. First, the authors describe the terminology of this field as well mandatory background concepts. Next, the main dataset and challenges are exposed to help researchers decide which are the ones that best suit their needs and their targets. Then, existing methods are reviewed, highlighting their contributions and their significance in the field. Finally, quantitative results are given for the described methods and the datasets in which they were evaluated, following up with a discussion of the results. At last, they point out a set of promising future works and draw conclusions about the state of the art of semantic segmentation using deep learning techniques.

Introduction

Semantic segmentation applied to still 2D images, video, and even 3D or volumetric data is one of the key problems in the field of computer vision. In the macro sense, semantic segmentation is one of the high-level task that paves the way towards complete scene understanding. As an essential problem of computer vision, the importance of scene understanding is highlighted by the fact that an increasing number of applications nourish from inferring knowledge from imagery (i.e. abstract from concrete). Some of those applications include autonomous driving [2, 4, 1], human-machine interaction [5], computational photography [7], image search engines [6], and augmented reality to name a few. Such problem has been addressed in the past using various traditional computer vision and machine learning techniques. Despite the popularity of those kind of methods, the deep learning revolution has turned the tables so that many computer vision problems including semantic segmentation are being tracked using deep architectures, usually convolu-

tional neural networks (CNNs), which are surpassing other approaches by a large margin in terms of accuracy and sometimes even efficiency. However, deep learning is far from the maturity achieved by other old-established branches of computer vision and machine learning.

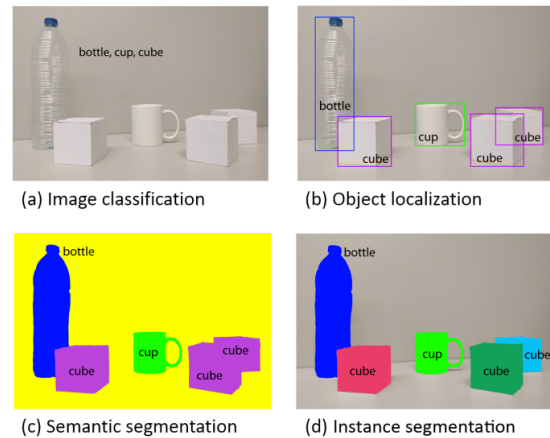


Figure 1. Evolution of object recognition or scene understanding from coarse-grained to fine-grained inference: classification, detection or localization, semantic segmentation, and instance segmentation.

Figure 1 shows the difference of image classification, object localization, semantic segmentation and instance segmentation.

This paper is the first review to focus explicitly on deep learning for semantic segmentation.

The key contributions of their work are as follows:

- Provide a broad survey of existing datasets that might be useful for segmentation projects with deep learning techniques.
- An in-depth and organized review of the most significant methods that use deep learning for semantic segmentation, their origins, and their contributions.
- A thorough performance evaluation which gathers quantitative metrics such as accuracy, execution time, and memory footprint.
- A discussion about the aforementioned results, as well as a list of possible future works that might set the

course of upcoming advances, and a conclusion summarizing the state of the art of the field.

References

- [1] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *CVPR*, pages 3213–3223, 2016. [1](#)
- [2] A. Ess, T. Mueller, H. Grabner, and L. v. Gool. Segmentation-based urban traffic scene understanding. In *BMVC*, pages 1–11, 2009. [1](#)
- [3] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. G. Rodríguez. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*. [1](#)
- [4] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *CVPR*, pages 3354–3361, 2012. [1](#)
- [5] M. Oberweger, P. Wohlhart, and V. Lepetit. Hands deep in deep learning for hand pose estimation. *arXiv preprint arXiv:1502.06807*, 2015. [1](#)
- [6] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li. Deep learning for content-based image retrieval: A comprehensive study. In *ACMMM*, pages 157–166, 2014. [1](#)
- [7] Y. Yoon, H. Jeon, D. Yoo, J. Lee, and I. S. Kweon. Learning a deep convolutional network for light-field image super-resolution. In *ICCV Workshops*, pages 24–32, 2015. [1](#)