

Scene Segmentation



Figure 1: Example images and labeled images in ADE20K dataset [1].

Scene segmentation, or pixel-wise dense labeling, is one of the key problems in computer vision. Specifically, one aims to give every pixel a label for images, which is more **fine-grained** than image-level recognition or object detection tasks. Existing dataset like ADE20K [1] has a large collection of carefully labeled natural scene images, with over 3000 classes. And there are also existing works on image segmentation tasks, one of the first is Fully Convolutional Neural Networks (FCN) [2]. In this project, we plan to **take part of the annotations**, say the **most common 150 objects** in the ADE20K dataset, to train our segmentation model. A good reference for state-of-the-art performance can be found on the Scene Parsing Challenge website, <http://sceneparsing.csail.mit.edu/>. We do not aim for the best performance, but we want the project to be exploratory where an original model can be proposed (e.g. **instance segmentation**, **improving semantic edges**, **compress large models**).

References

- [1] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, Semantic understanding of scenes through the ade20k dataset, *arXiv preprint arXiv:1608.05442*, 2016.
- [2] J. Long, E. Shelhamer, and T. Darrell, Fully convolutional networks for semantic segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.