Team: CoCaffeine
Project Title: Image Panoptic Segmentation Challenges using Coco Dataset

Project Summary:

Image panoptic segmentation [1] is a task which performs both instance segmentation (things) and semantic segmentation (stuff) simultaneously on images. Panoptic segmentation generates segmented output image that not only separates out each individual object, but also separates different groups of objects from each other. Today there are many areas in which panoptic segmentation is being heavily applied. In self-driving cars, for example, the software needs to use panoptic segmentation in order to tell from camera images which parts of the surroundings are objects and which parts are the road that the vehicle is driving on. Given to its wide range of applications, Panoptic segmentation has emerged as one of the most fast-advancing research areas in computer vision and deep learning. This project aims to explore some state-of-the-art works on panoptic segmentation, and in addition, explore some of our own modifications in hope of improving the performance.

Approach:

Based on our preliminary research, there are multiple approaches to perform image panoptic segmentation tasks. We want to start by collecting and analyzing varying approaches, for example: Mask-RCNN [2], DeepLab V2 [3], EfficientLPS [4], Axial-attention Model [5]. We expect to reproduce [2] and baseline other approaches.

As a stretch goal, we want to explore possible directions to improve these papers. One option is to use an ensemble of different models and combine the segmented results to ensure prediction consistency.

Another stretch goal is to come up with a new scoring on segmentation output in hope to improve benchmarks.

Ethical Implications:

There are no direct ethical implications for this project.

Resources/Related Work:

Panoptic segmentation is believed to be a breakthrough in the areas of computer vision [6]. Its recent development majorly lied in the variations of neutral network structures. Methods differ in the sequence by which instance and semantic networks interact with each other. In other words, Segmentation and semantic networks may be trained separately (but simultaneously) (e.g., UPSNet [7]), in sequence [8], or in combination [9] to result in final panoptic segmentation. The majority of previous studies use RGB images that are widely seen in such sources as video

cameras [10] and computer renderings [11]. In certain domains such as healthcare and remote sensing the methods are also applied using medical images [12] and LIDAR data (satellite images) [13]. These well-developed frameworks provide us a good start point to conduct different experiments and inspire us to modify existing techniques for a better performance.

References

[1] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, Piotr Dollár: “Panoptic Segmentation”, 2018; [http://arxiv.org/abs/1801.00868 arXiv:1801.00868].
[2] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick: “Mask R-CNN”, 2017; [http://arxiv.org/abs/1703.06870 arXiv:1703.06870].
[3] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille: “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs”, 2016; [http://arxiv.org/abs/1606.00915 arXiv:1606.00915].
[4] Kshitij Sirohi, Rohit Mohan, Daniel Büscher, Wolfram Burgard, Abhinav Valada: “EfficientLPS: Efficient LiDAR Panoptic Segmentation”, 2021, IEEE Transactions on Robotics (T-RO), 2021; [http://arxiv.org/abs/2102.08009 arXiv:2102.08009].
[5] Huiyu Wang, Yukun Zhu, Bradley Green, Hartwig Adam, Alan Yuille, Liang-Chieh Chen: “Axial-DeepLab: Stand-Alone Axial-Attention for Panoptic Segmentation”, 2020; [http://arxiv.org/abs/2003.07853 arXiv:2003.07853].
[6] Elharrouss, O., Subramanian, N., Ottakath, N., Almaadeed, N., & Himeur, Y. (2021). Panoptic Segmentation: A Review. *arXiv*. https://doi.org/10.48550/arXiv.2111.10250
[7] Y. Xiong, R. Liao, H. Zhao, R. Hu, M. Bai, E. Yumer, R. Urtasun, Upsnet: A unified panoptic segmentation network, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 8818–8826.
[8] Y. Wu, G. Zhang, H. Xu, X. Liang, L. Lin, Auto-panoptic: Cooperative multi-component architecture search for panoptic segmentation, in: 34th Annual Conference on Neural Information Processing Systems (NeurIPS), 2020, pp. 1–8.
[9] S. Caelles, K.-K. Maninis, J. Pont-Tuset, L. Leal-Taixé, D. Cremers, L. Van Gool, One-shot video object segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 221–230.
[10] M. Weber, J. Luiten, B. Leibe, Single-shot panoptic segmentation, arXiv preprint arXiv:1911.00764 (2019)
[11] F. Saeedan, S. Roth, Boosting monocular depth with panoptic segmentation maps, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 3853–3862.
[12] D. Liu, D. Zhang, Y. Song, H. Huang, W. Cai, Cell r-cnn v3: A novel panoptic paradigm for instance segmentation in biomedical images, arXiv preprint arXiv:2002.06345 (2020).
[13] J. Behley, A. Milioto, C. Stachniss, A benchmark for lidarbased panoptic segmentation based on kitti, arXiv preprint arXiv:2003.02371 (2020).

Datasets:

COCO dataset

Team Members:
Anurag Sen (asen48@gatech.edu)
Jie Song (jsong407@gatech.edu)
Zhiling Zhou (zhiling_zhou@gatech.edu)
Zixiang (David) Zhu (zzhu72@gatech.edu)