Assignment 3

COL-780

December 2020

This assignment is based on Object Detection and Semantic Segmentation. For this assignment, you are required to register for the ICVGIP Visual Data Challenge. The Challenge includes two contests:

1. **Semantic Segmentation for Autonomous Driving:** This challenge involves obtaining pixel-level semantic labeling of images in the context of autonomous driving. It utilizes the large-scale, public, India-centric dataset, India Driving Dataset (IDD) to evaluate the solutions provided by the competing teams.
2. **Object Detection for Anti-Poaching Aerial Patrolling of Protected Areas:** This challenge addresses the problem of mitigating poaching attacks on endangered animals in the wild. It utilizes the BIRDSAI dataset for automatic detection of animals and humans in thermal IR aerial images collected from UAVs for augmenting anti-poaching patrols of protected areas.

You may register for either of these contests. Refer to the challenge website for more details.

Guidelines

1. The dataset for both the challenges can be downloaded from the challenge website.
2. You should start with a baseline model. Obtain a pre-trained object detection or semantic segmentation model and train it on the challenge dataset. You can start with one of these models [1, 2, 3, 4, 5, 6] or look up other recent papers.
3. Analyse the weaknesses of your model, and suggest a series of improvements on top of it.

Submission Instructions

1. **Register for the contest by 12th December, 2020.**
2. You may join the challenge individually or form groups of two.
3. Note that the contest ends on 17th December, 2020. The assignment deadline will, however, be at a later date (to be announced in class).
4. You are allowed to use publicly available code and libraries with appropriate citation.
5. You are not allowed to discuss or borrow code from other groups.

Evaluation

Your evaluation will be based on

1. Contest leader-board: A leader-board rank of top 5 will fetch you 100% marks in the assignment (and a bonus for top 3). There is also a cash prize for the top 3 teams in the contest.
2. Report and Viva: In addition to your submission on the contest website, you must also submit your code and report on Moodle. Report must include details of your method and results.
    - Baseline model - 3 marks
    - Improvements on baseline model - 5 marks
    - Report and viva - 2 marks

References

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[3] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[4] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[5] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[6] E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo, "Erfnet: Efficient residual factorized convnet for real-time semantic segmentation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 263–272, 2017.