Semi-automatic Annotation for Semantic Segmentation: A Case Study for Underwater Image Segmentation

Sameer Chaudhary and Md Alimoor Reza

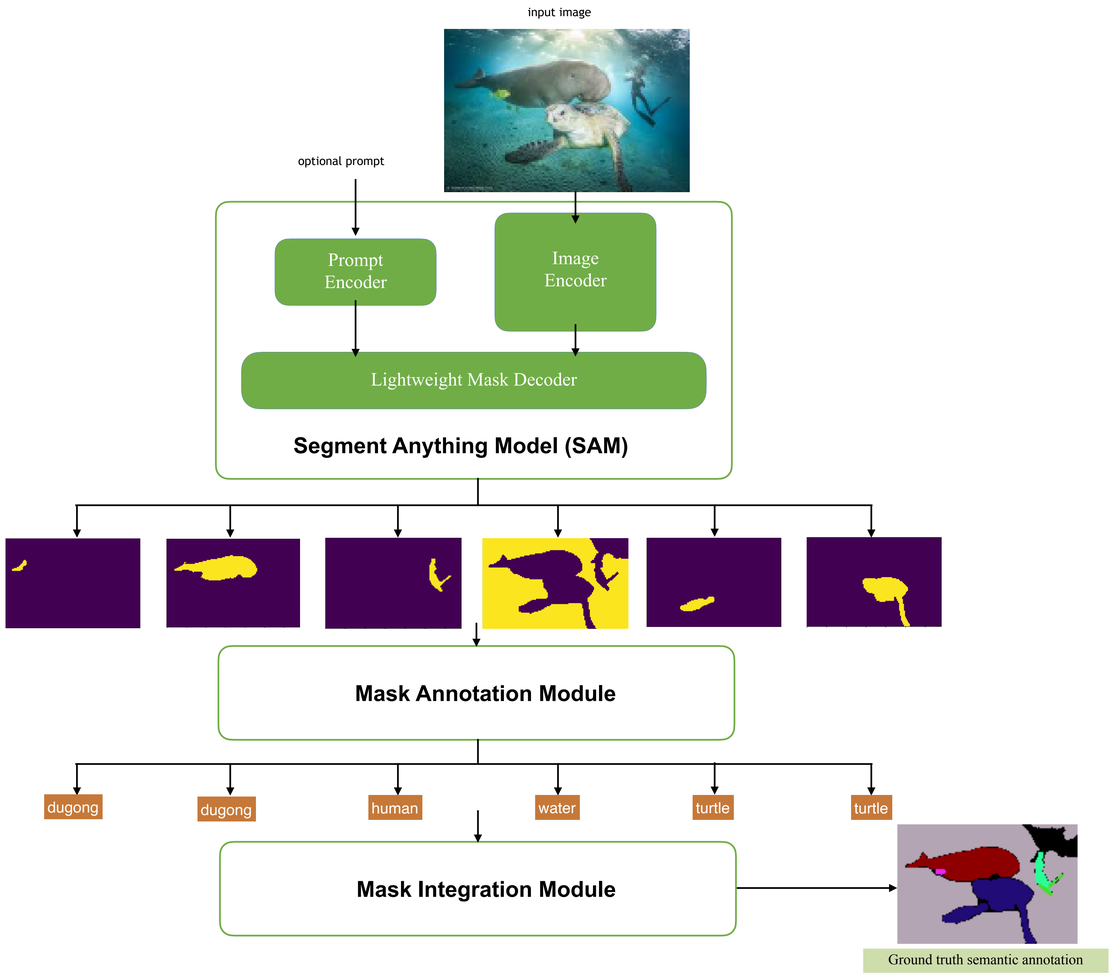
Drake University

Motivation

- Semantic segmentation is a dense image prediction task where the goal is to label different parts of an image from a set of predefined set of object categories e.g., *table, chair, floor* (indoor) or *sky, car* (outdoor)
- Semantic segmentation has applications in underwater environment exploration, autonomous driving, virtual reality, etc.
- Effective solution for semantic segmentation has two requirements:
  1. Deep Neural Network (DNN) models — CNNs or Transformer models
  2. Collecting a significant number of images along with dense pixel-wise annotations. These annotated images are used for training the DNNs
- Existing annotation [1] is i) time consuming, ii) laborious, and iii) costly

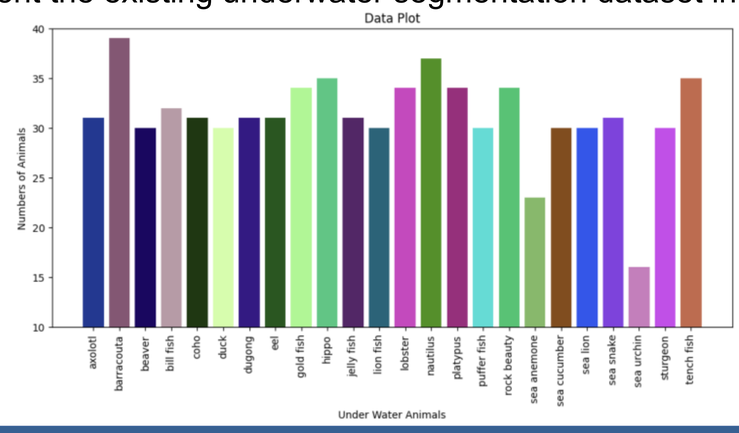
Semi-automatic Annotation Algorithm

- We proposed a semi-automatic algorithm that can annotate every pixel in an image. It only requires entering the name of a segment (vs. laborious manual marking of a region using existing tool [1])
- Our algorithm mitigates some of the limitations of manual annotations — less time-consuming, less labor-intensive, and less costly
- At the core of our algorithm, we have a powerful segment mask generation method called “Segment Anything Model (SAM)” [5]. It can be thought of as a general-purpose segmentation mask generation model — a large language model (LLaMa) equivalent such as ChatGPT. SAM is promotable and has a superior zero-shot performance on a new image distribution
- Our methodological pipeline is shown below:

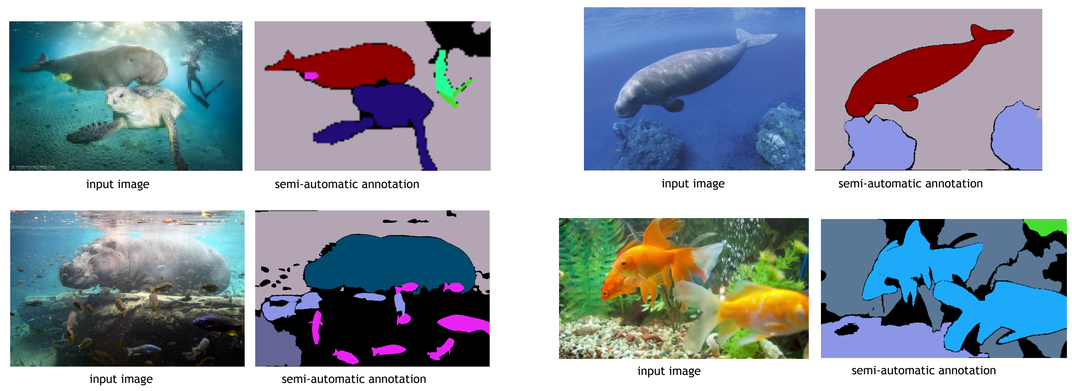


New Dataset for Semantic Segmentation

- We collected images for 24 new animal categories as shown below. It will complement the existing underwater segmentation dataset introduced in [4]



Semi-automatic Annotation Results



Baseline Segment Mask Generation Results

- U²Net [2] can generate segmentation masks but has several limitations:
  - fails to detect all foreground objects
  - merges multiple foregrounds

Future Work

- The semi-automatically annotated images will be used to train DNN semantic segmentation models [3][4]
- Apply our semi-automatic annotation method on other datasets

References

1. Label Studio: https://labelstud.io
2. X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. Zaiane, and M. Jagersand. “U2-Net: Going Deeper with Nested U-Structure for Salient Object Detection”. Pattern Recognition 2020
3. M. J. Islam, C. Edge, Y. Xiao, P. Luo, M. Mehtaz, C. Morse, S. Enan, and J. Sattar, “Semantic Segmentation of Underwater Imagery: Dataset and Benchmark,” in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020
4. I. Kabir, S. Shaurya, V. Maigur, N. Thakurdesai, M. Latnekar, M. Raunak, D. Crandall, and M. Reza, “Few-shot segmentation and Semantic Segmentation for Underwater Imagery” - International Conference on Intelligent Robots and Systems (IROS’23) (under review)
5. A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. Berg, W. Lo, P. Doll'ar, and R. Girshick, “Segment Anything” - arXiv’2023