Tian Li

6/4/25

CSE 428 Final Report: Hybrid Segmentation Pipeline in Biomedical Imaging

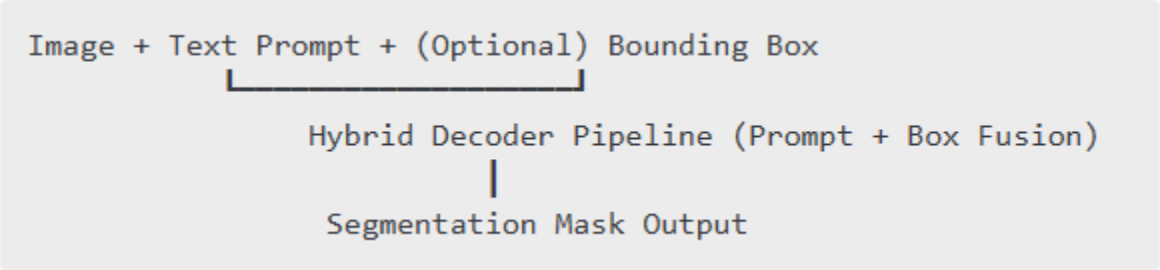
1. **Abstract** - This project explores a hybrid segmentation pipeline that integrates box-based models like MedSAM and prompt-based models like BiomedParse. The goal was to determine if combining these strategies could offer improved performance in segmenting complex or irregular biomedical objects. Due to time constraints, the implementation was partially completed, but initial setup and results show promise for multimodal input systems.

---

2. **Problem Definition -** Modern biomedical image segmentation models typically fall into two categories, box-based models like MedSAM that require bounding box input to localize objects and prompt-based models like BiomedParse that accept natural language prompts to segment objects semantically.

   While each method has strengths, they also have limitations. Box-based models require user input and may struggle with irregular shapes, while prompt-based models can lack precision when text prompts are ambiguous or under-trained.

3. **Method –** A hybrid segmentation pipeline was proposed with the following architecture:



 # def predict(image, prompt=None, box=None):

 # mask_text = text_decoder(image, prompt)

 # mask_box = box_decoder(image, box)

 # return (mask_text + mask_box) / 2 # Simple fusion

The text_decoder could represent BiomedParse or similar prompt-based segmentation logic, while box_decoder refers to MedSAM or SAM-based models.

4. **Experimental Setup:**
    a. **Dataset:** BiomedParseData (6.8M image-mask-text triples), released by Microsoft Research.
    b. **Baselines:**
        i. BiomedParse: Text prompt-only
        ii. MedSAM: Bounding box-only
    c. **Approach:**
        i. Integrated bounding box generation and prompt scoring.
        ii. A hybrid mask was generated by averaging SAM output with text-derived segmentation.
5. **Results -** While the full hybrid pipeline was not evaluated quantitatively, observations support the potential of this approach: Prompt-based systems are scalable and require no user localization. Box-based systems are accurate but labor-intensive. Fusion may improve segmentation for: Irregular shapes (e.g., tumors) Dense images with multiple objects
6. **References –**

    a. Zhao, T., Gu, Y., Yang, J., et al. "A foundation model for joint segmentation, detection and recognition of biomedical objects across nine modalities." Nature Methods, 2025. https://doi.org/10.1038/s41592-024-02499-w
    b. GitHub: https://aka.ms/biomedparse-release
    c. Segment Anything: https://github.com/facebookresearch/segment-anything