nature methods

Article

https://doi.org/10.1038/s41592-024-02499-w

A foundation model for joint segmentation, detection and recognition of biomedical objects across nine modalities

Received: 21 May 2024

Accepted: 2 October 2024

Published online: 18 November 2024

Check for updates

Theodore Zhao[1,6], Yu Gu[1,6], Jianwei Yang[1], Naoto Usuyama[1], Ho Hin Lee[1], Sid Kiblawi[1], Tristan Naumann[1], Jianfeng Gao[1], Angela Crabtree[3], Jacob Abel[2], Christine Moung-Wen[2], Brian Piening[2,3], Carlo Bifulco[2,3], Mu Wei[1] ✉, Hoifung Poon[1] ✉ & Sheng Wang[4,5] ✉

Biomedical image analysis is fundamental for biomedical discovery. Holistic image analysis comprises interdependent subtasks such as segmentation, detection and recognition, which are tackled separately by traditional approaches. Here, we propose BiomedParse, a biomedical foundation model that can jointly conduct segmentation, detection and recognition across nine imaging modalities. This joint learning improves the accuracy for individual tasks and enables new applications such as segmenting all relevant objects in an image through a textual description. To train BiomedParse, we created a large dataset comprising over 6 million triples of image, segmentation mask and textual description by leveraging natural language labels or descriptions accompanying existing datasets. We showed that BiomedParse outperformed existing methods on image segmentation across nine imaging modalities, with larger improvement on objects with irregular shapes. We further showed that BiomedParse can simultaneously segment and label all objects in an image. In summary, BiomedParse is an all-in-one tool for biomedical image analysis on all major image modalities, paving the path for efficient and accurate image-based biomedical discovery.

Biomedical image analysis is critical to biomedical discovery because imaging is one of the most important tools for studying physiology, anatomy and function at multiple scales from the organelle level to the organ level[1–4]. Holistic image analysis comprises multiple subtasks, such as segmentation, detection and recognition of biomedical objects. Segmentation aims to divide an image into segments representing different objects, often requiring the aid of a user-provided bounding box for each object of interest[5,6]. Detection aims to identify the location of an object of interest in the image[7], whereas recognition aims to identify all objects within an image[8]. Standard image analysis methods typically approach these tasks separately, using specialized tools for individual tasks[9]. Despite their encouraging performance, such a disjointed approach misses opportunities for joint learning and reasoning across these interdependent tasks.

For example, a lot of previous image analysis works focus on segmentation alone, thus ignoring key semantic information from interdependent tasks, such as metadata and object type names. This results in suboptimal segmentation while imposing substantial burden on users, as many state-of-the-art segmentation tools require users to provide a tight bounding box indicating the location of an object of interest[10,11].

[1]Microsoft Research, Redmond, WA, USA. [2]Providence Genomics, Portland, OR, USA. [3]Earle A. Chiles Research Institute, Providence Cancer Institute, Portland, OR, USA. [4]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA. [5]Department of Surgery, University of Washington, Seattle, WA, USA. [6]These authors contributed equally: Theodore Zhao, Yu Gu. ✉e-mail: muhsin.wei@microsoft.com; hoifung@microsoft.com; swang@cs.washington.edu