Supercharging Computer Vision with Foundation Model Fine-Tuning

Shubing Wang
Merck Research Laboratories

Abstract

Hailed as the GPT-4 of computer vision, the Segment Anything Model (SAM) is a foundational model that exhibits robust zero-shot performance across a variety of computer vision tasks, including segmentation and object detection, applicable to diverse imaging platforms. Despite SAM's superior out-of-the-box adaptation compared to previous methods, there remains significant potential for improvement, particularly in challenging scenarios such as cryo-electron microscopy (cryoEM) image analysis. By leveraging smaller, noisier annotated datasets from earlier approaches, fine-tuning SAM in different contexts—such as region of interest (ROI) detection, particle segmentation, and classification in cryoEM image analysis—has significantly enhanced performance in these tasks. However, the large size of fine-tuned SAM models often results in slow training and inference, requiring substantial computational resources. To address these challenges, we demonstrate that fine-tuned SAM can significantly assist in fine-tuning smaller foundational models, such as YOLO, to improve efficiency through knowledge distillation. This presentation will demonstrate how fine-tuning foundation models can transform cryo-electron microscopy (cryoEM) from a specialized technique into a fundamental platform for the development of various vaccines and adjuvants.