Mammo-CLIP: A Vision Language Foundation Model to Enhance Data Efficiency and Robustness in Mammography

Shantanu Ghosh[1], Clare B. Poynton[2], Shyam Visweswaran[3], Kayhan Batmanghelich[1]

[1]Dept. Of Electrical and Computer Engineering, Boston University

[2]Boston University Chobanian & Avedisian School of Medicine

[3]Intelligent Systems Program (ISP), University of Pittsburgh

MICCAI 2024 Marrakesh MOROCCO

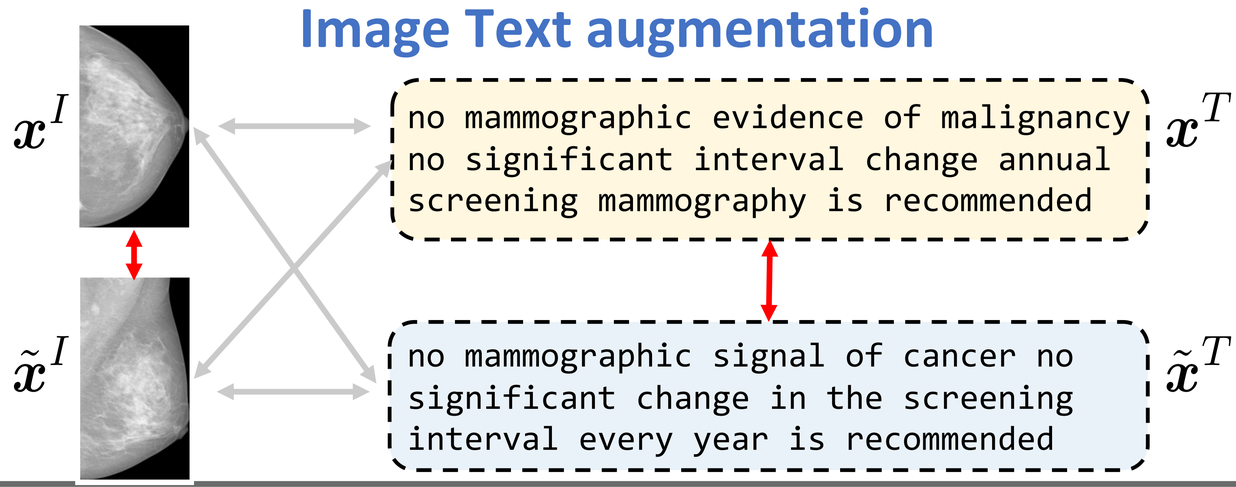
BOSTON UNIVERSITY | BATMAN LAB

TLDR: A vision language model trained on both mammogram-report pairs and mammogram-attribute datasets, enhancing data efficiency, robustness, and interpretability

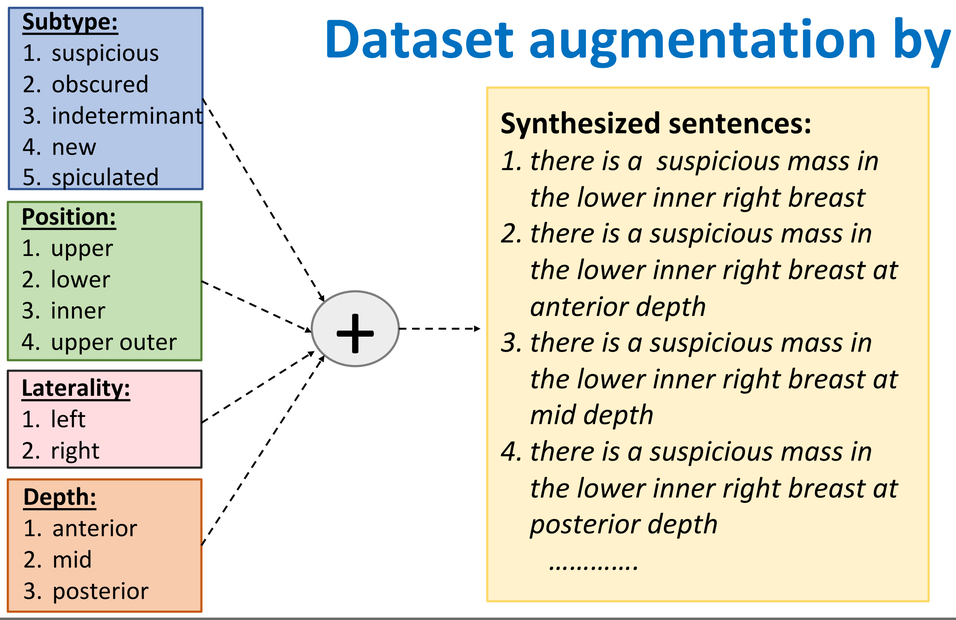
Motivation

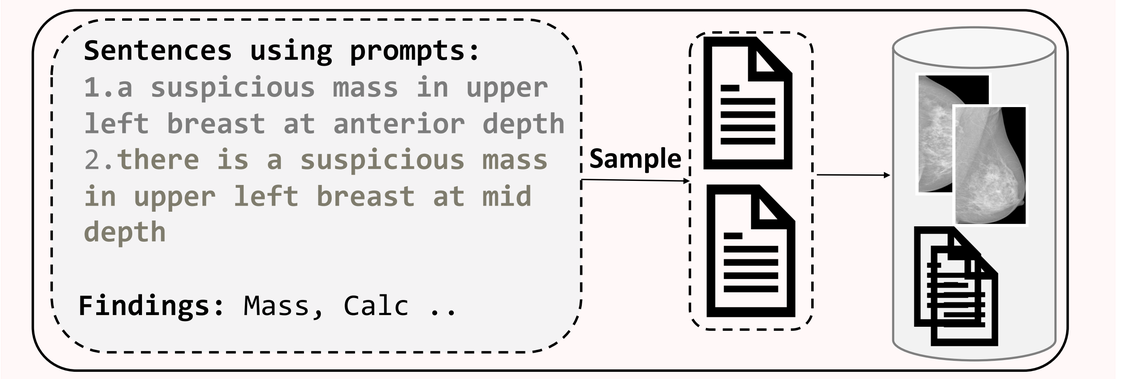
- Scarcity of diverse, annotated mammogram datasets for effective CAD training.
- Vision-Language Models enhance robustness and data efficiency for medical imaging..
- Existing models lose critical diagnostic details due to reduced image resolution.
- Improving AI transparency with feature alignment between images and reports.

Image Text augmentation

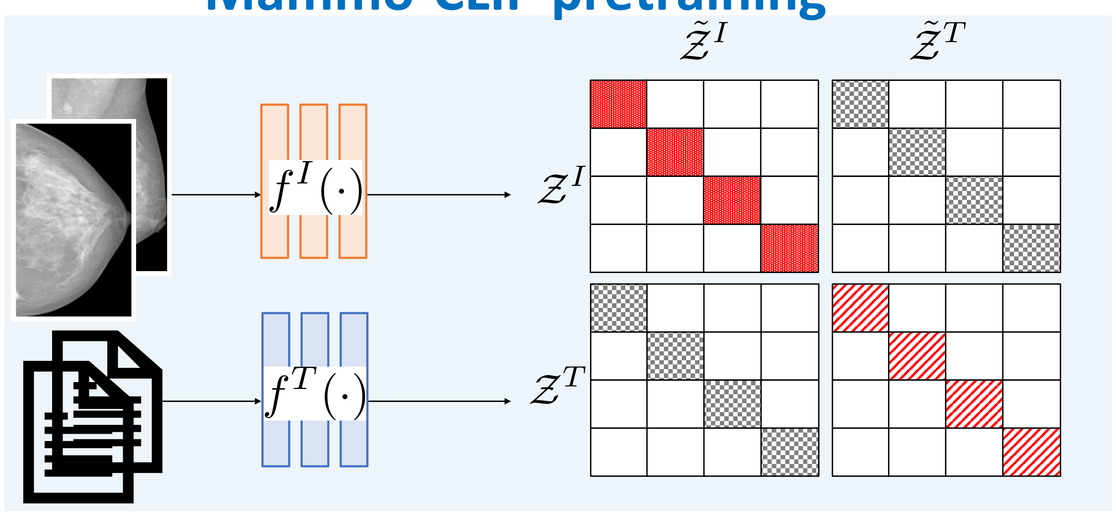


Dataset augmentation by synthesizing reports using image-label datasets

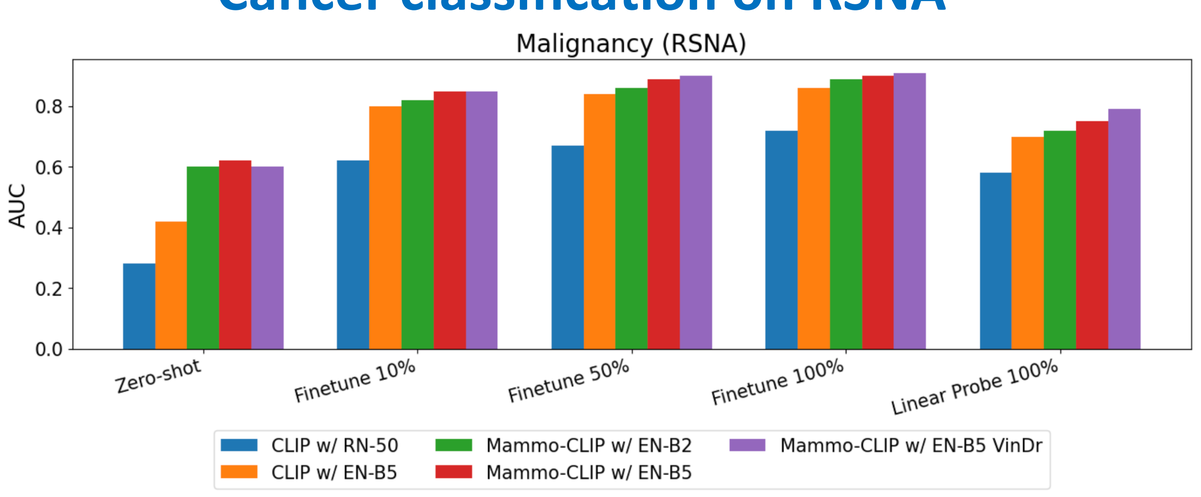




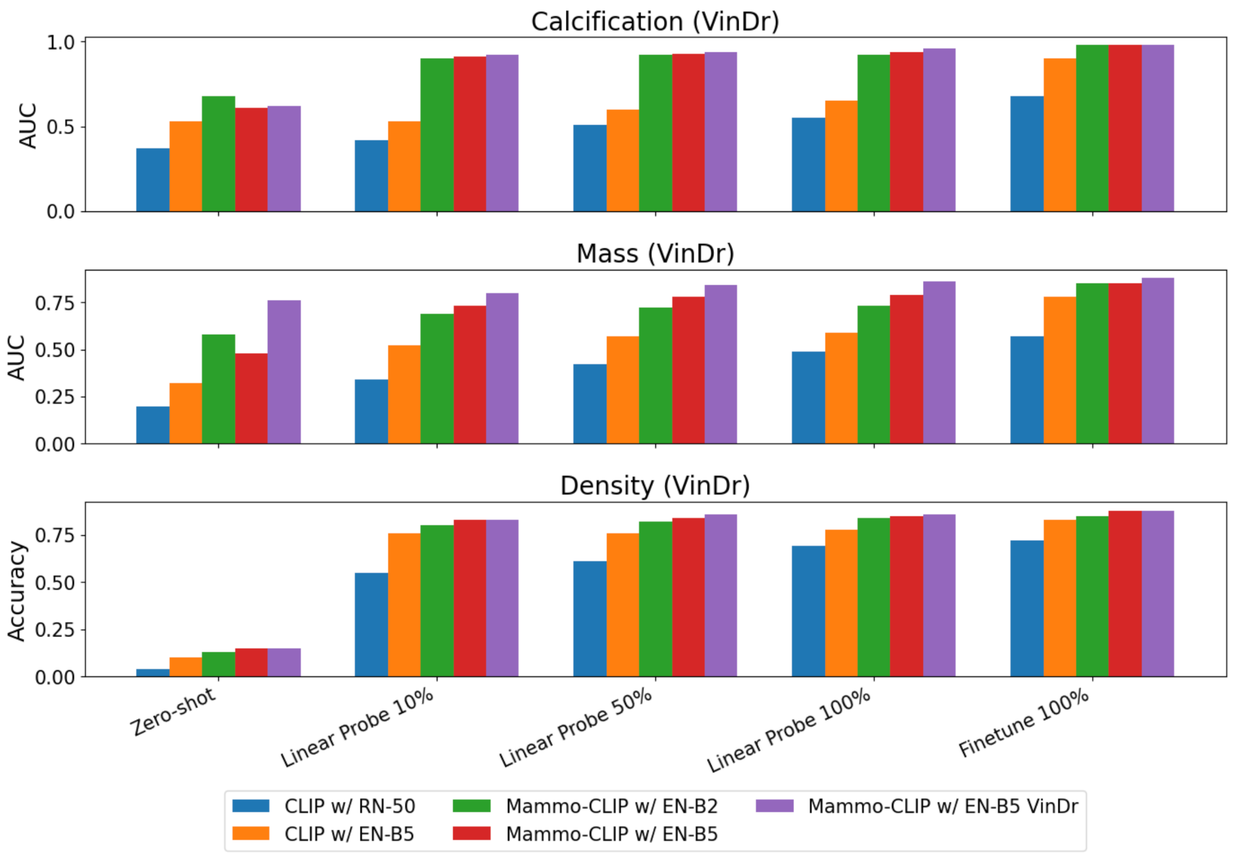
Mammo-CLIP pretraining



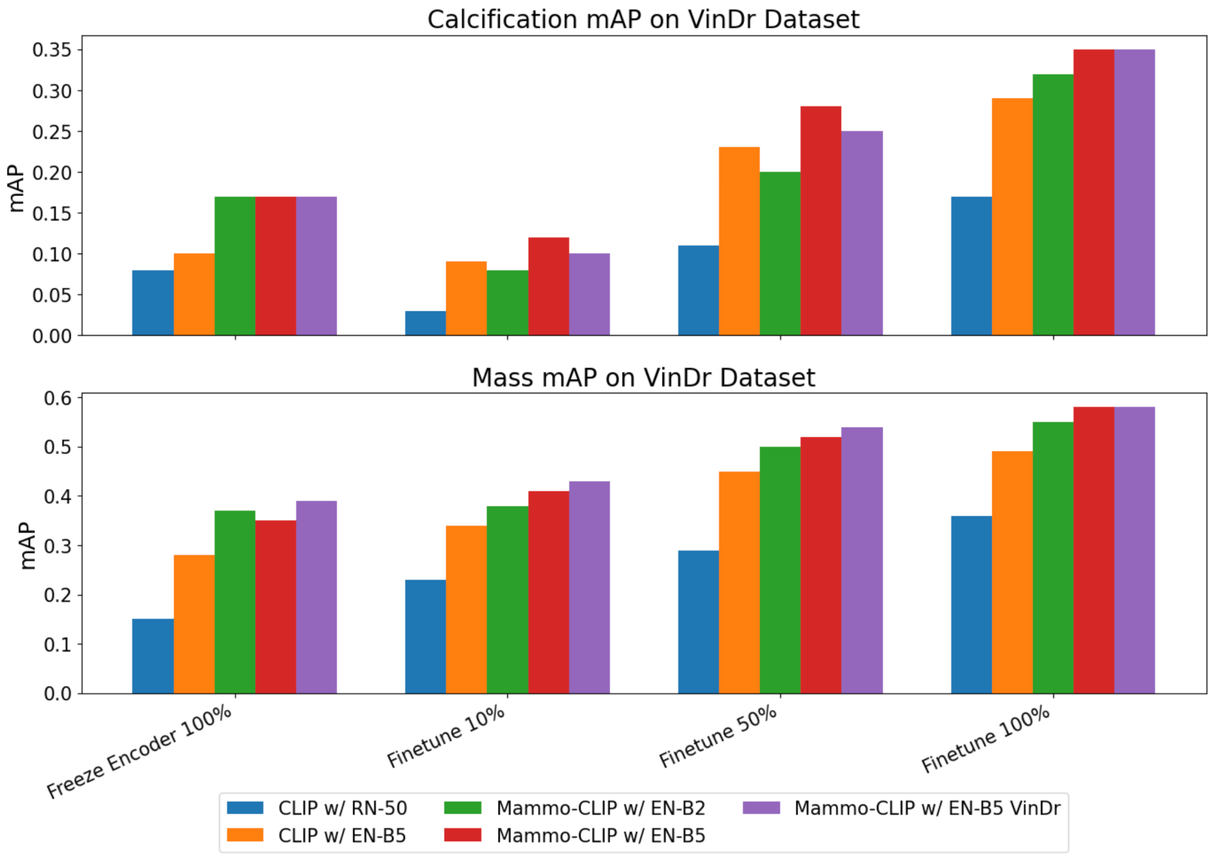
Cancer classification on RSNA



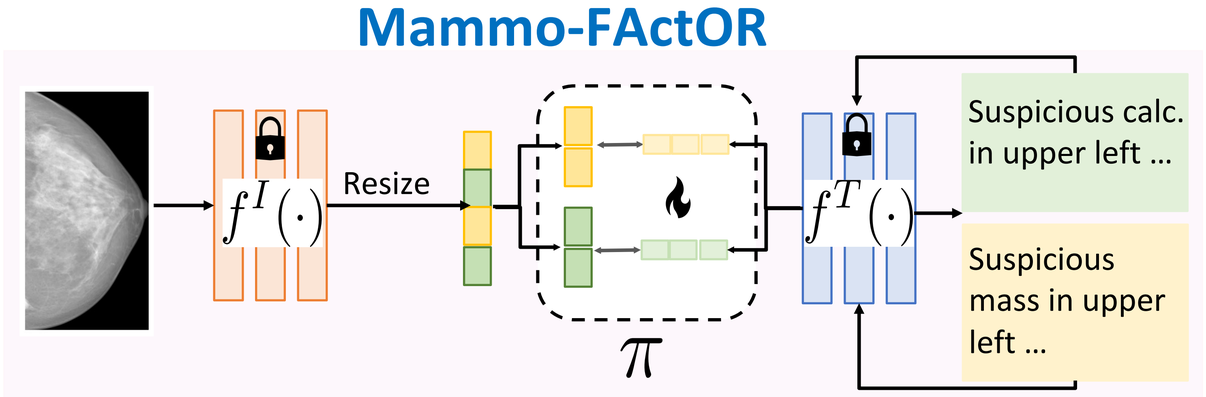
Findings classification on VinDr



Findings localization on VinDr



Mammo-FActOR



Mammo-FActOR localization