Task-driven Prompt Evolution for Foundation Models

Rachana Sathish, Rahul Venkataramani, K S Shriram, Prasad Sudhakar
rahul.venkataramani.ge.com
GE HealthCare

GE HealthCare

MOTIVATION

- Promptable foundation models, particularly Segment Anything Model (SAM) [1], have emerged as a promising alternative to the traditional task-specific supervised learning for image segmentation.
- Choice of location of click-based prompt inside the region of interest affects the quality of segmentation as demonstrated in Fig.1.
- **Challenges**:
  - Since X-ray is a summative modality, the intensity values under the lung mask are a result of superimposition of soft tissue, ribs, cardiac region, and occasional extraneous objects such as PICC lines.
  - Though visually the lung region may appear equally dark in X-ray images to the user, it is not homogeneous, and its heterogeneity is further amplified by the presence of pathology.

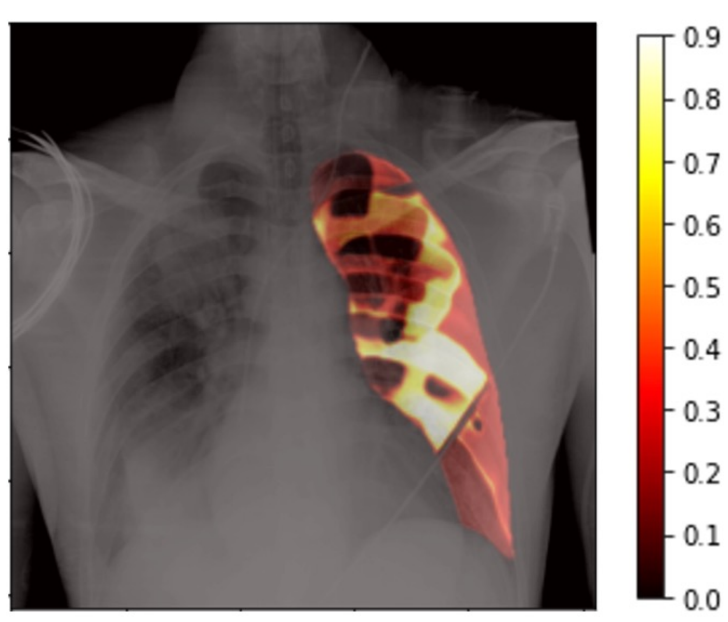


Fig.1. Heat-map of Dice values obtained by placing the prompt at various locations in the lung.

METHOD

- Proposed solution – **SAM P**rompt **O**ptimization **T**echnique (SAMPOT)
- We design an unsupervised segmentation performance scorer that generates a proxy for the supervised performance metric like the Dice value.
- At inference, given a test image and prompt, we iteratively maximize this task-based score to evolve the location of the prompt to produce superior results compared to utilizing initial prompt location provided by user.

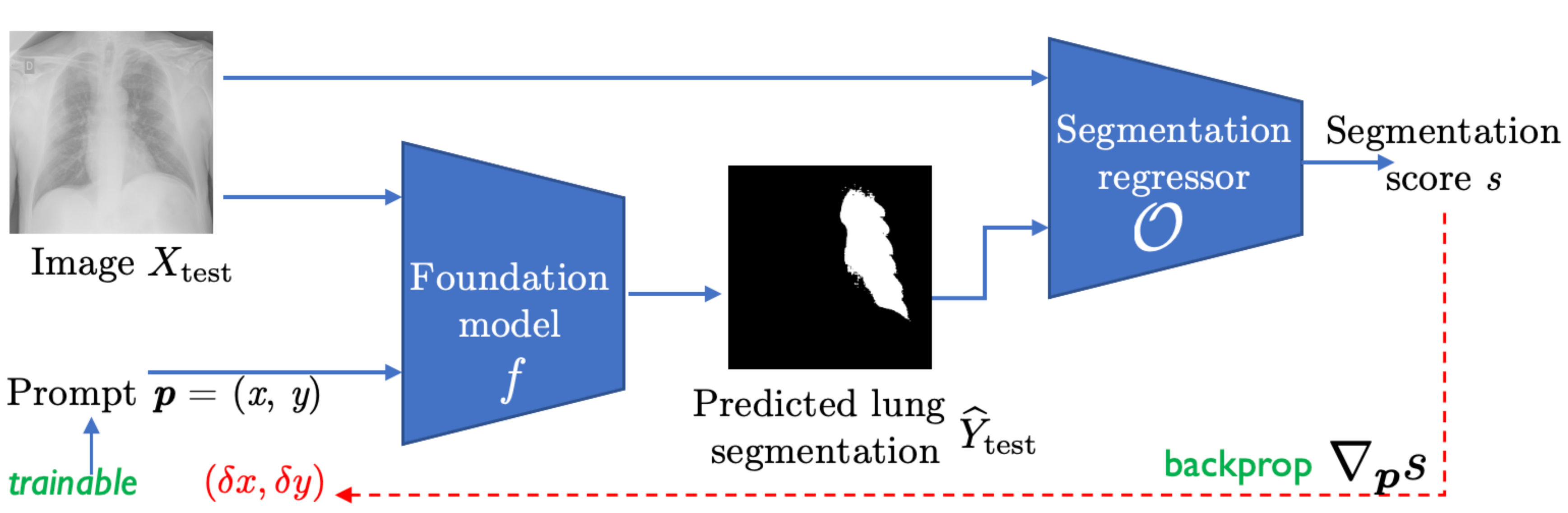
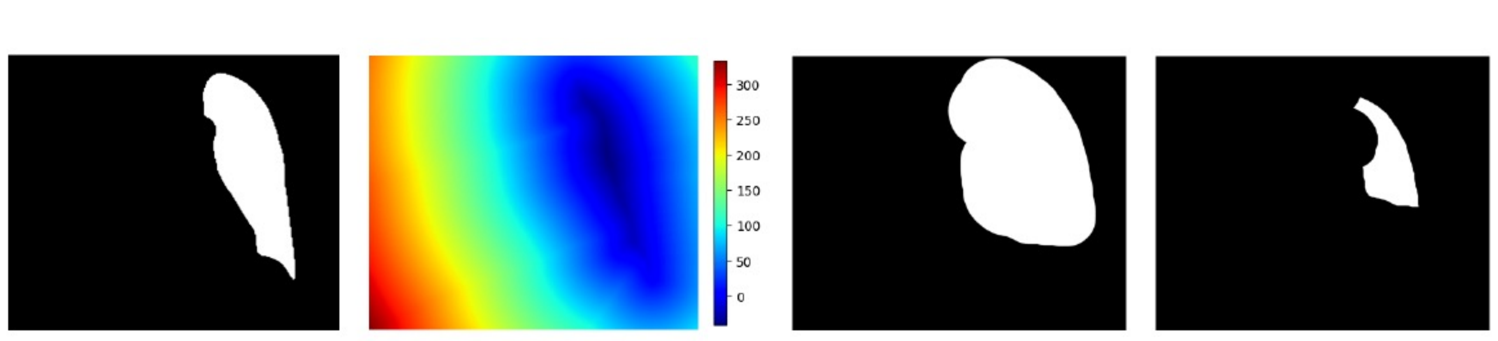


Fig.2. Overview of proposed method - SAMPOT

- Prompt optimization:

$$\boldsymbol{p}^* := \arg\max_{\boldsymbol{p}} \mathcal{O}(X_{\text{test}}, \widehat{Y}_{\text{test}}), \text{ where } \widehat{Y}_{\text{test}} := f_{\texttt{SAM}}(X_{\text{test}}, \boldsymbol{p}).$$

- Learning to score: The oracle $\mathcal{O}$ is expected to score the quality of segmentation blindly in the absence of ground truth. To this end, we train a segmentation regressor which learns to predict the Dice directly from the input image and the corresponding predicted mask.



(a) Sample Mask (b) Distance map (c) Over-seg. (d) Under-seg.

Fig.3. Data preparation: We used the level-sets of ground truth annotation to generate under- and over-segmented instances of the lung field

RESULTS

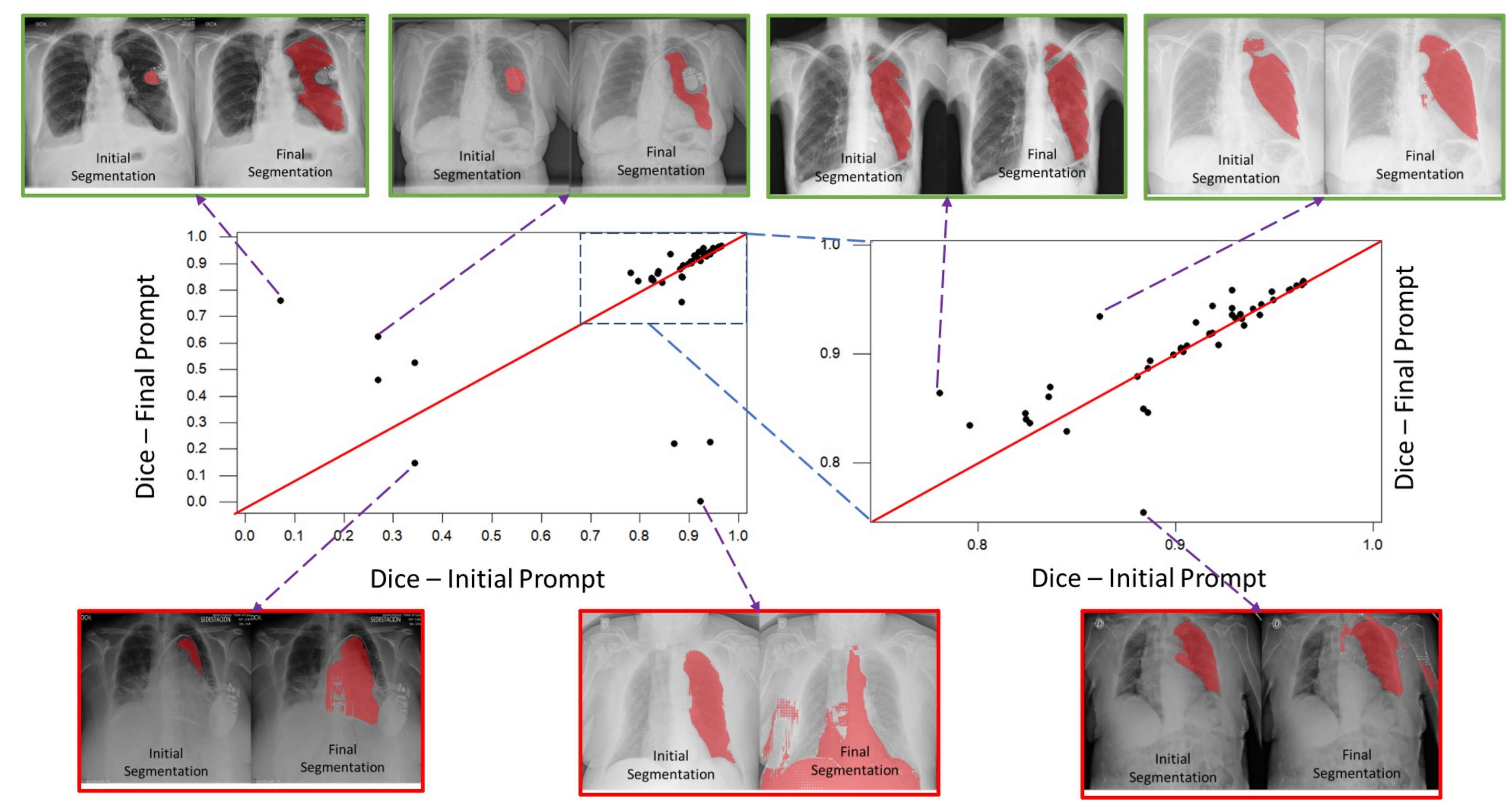


Fig.4. Scatter plot of Dice coefficients resulting from initial prompts and the final evolved prompts on the test set.

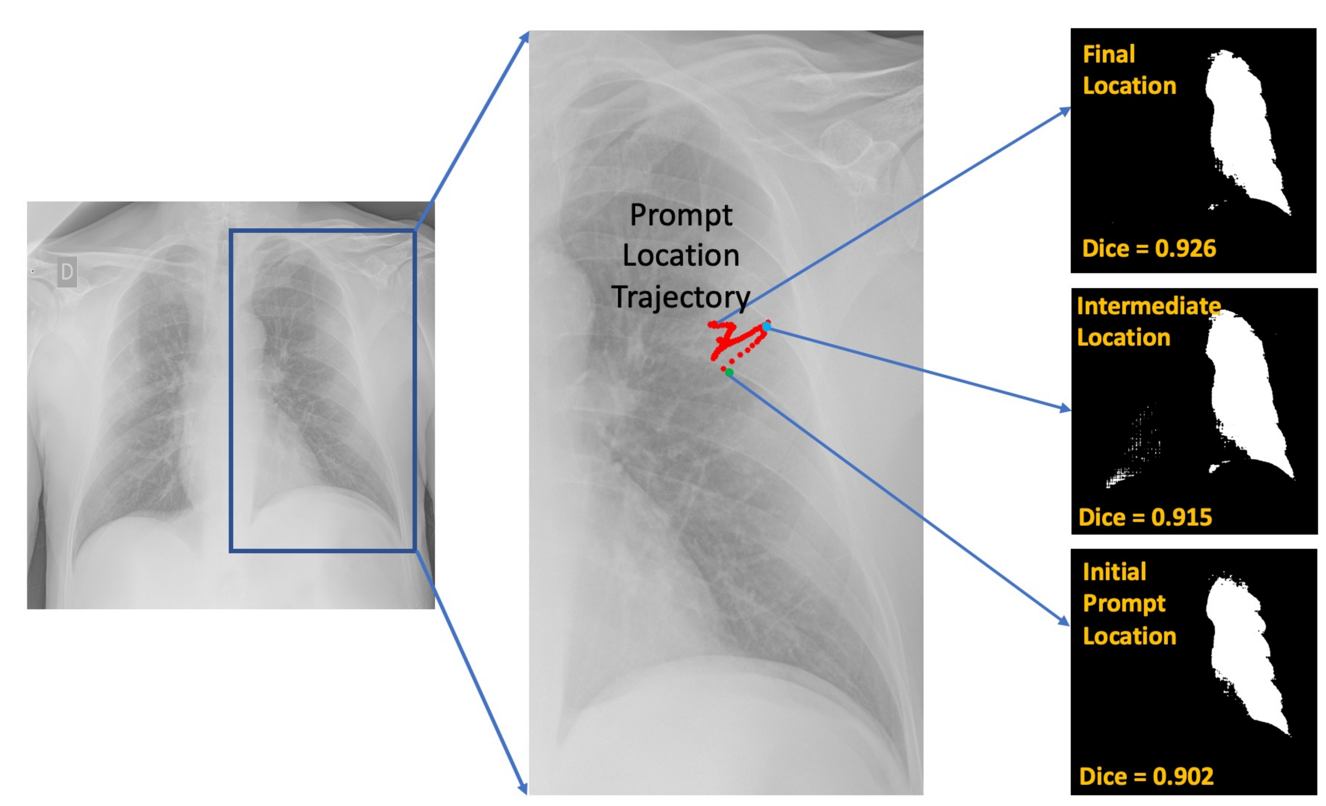
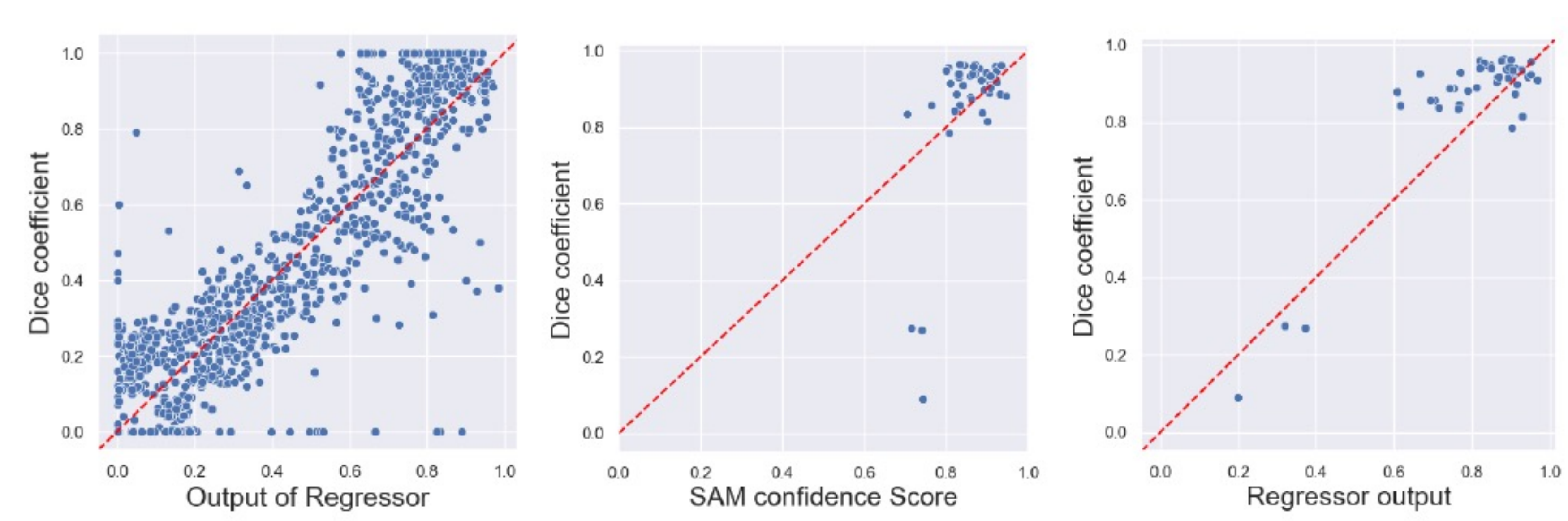


Fig.5. Trajectory of the prompt during the optimization process. The initial prompt is set at the centroid of the ground truth lung field annotation.

Ablation Study:



(a) (b) (c)

Fig.6. Comparison of (a) Dice against regressor output for unseen synthetically generated masks (1205 samples); on the test set (53 samples) (b) Dice against SAM confidence score and (c) Dice against regressor output when prompts are placed at the centroid of the lung mask.

FUTURE WORK

- Optimization of multiple prompts: click-based prompts coupled with box prompts.
- Optimization of positive and negative prompts.

REFERENCES

[1] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T.,Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. arXiv preprint arXiv:2304.02643 (2023)