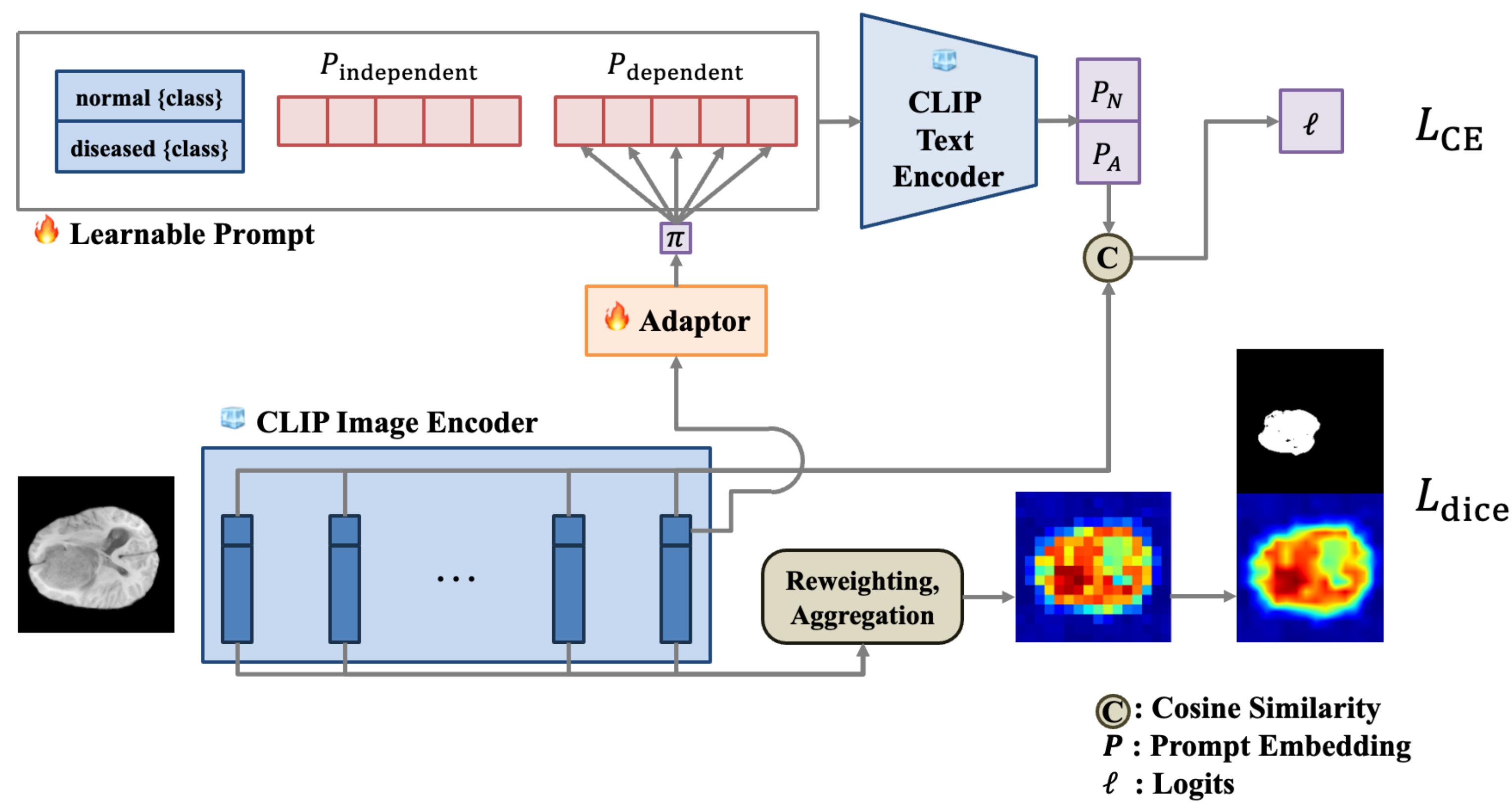


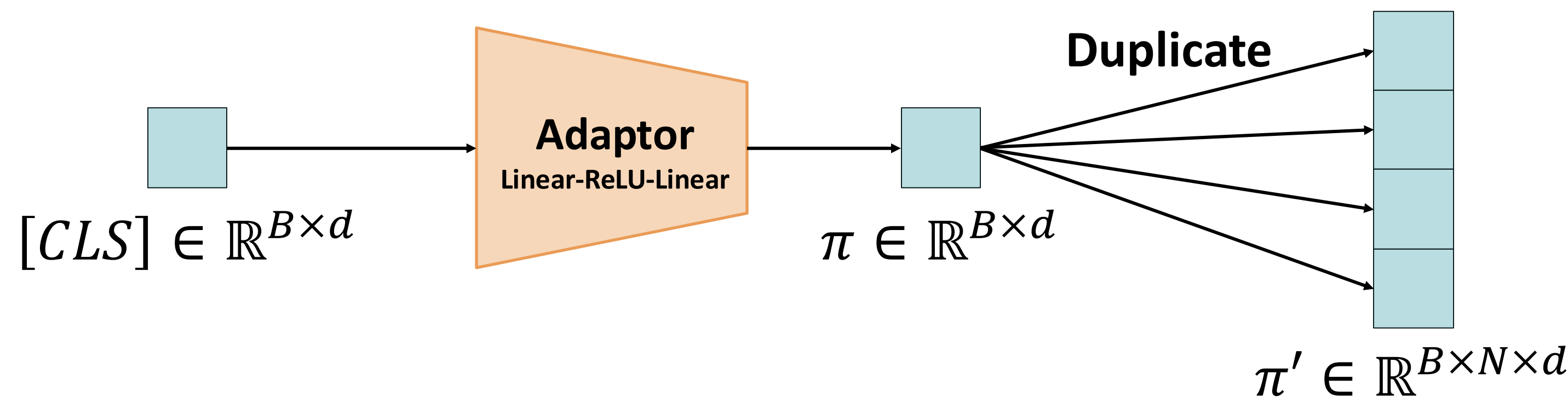
Introduction

- ◆ **Robust Algorithm applicable to diverse domains**
 - Diverse MRI sequences (T1, T2, T1CE, FLAIR)
 - Diverse Modalities (X-ray, OCT...)
- ◆ **Giving Condition to Prompt for Adaptation**
 - Existing works (AnomalyCLIP, AdaCLIP, MediCLIP) use CoOp method solely, limiting model's adaptability
- ◆ **A Prompting module to guide CLIP:**
 - **CLIP** prompt gUidance for mEdical image
 - Guiding CLIP patch embedding to understand medical anomalies in diverse slices

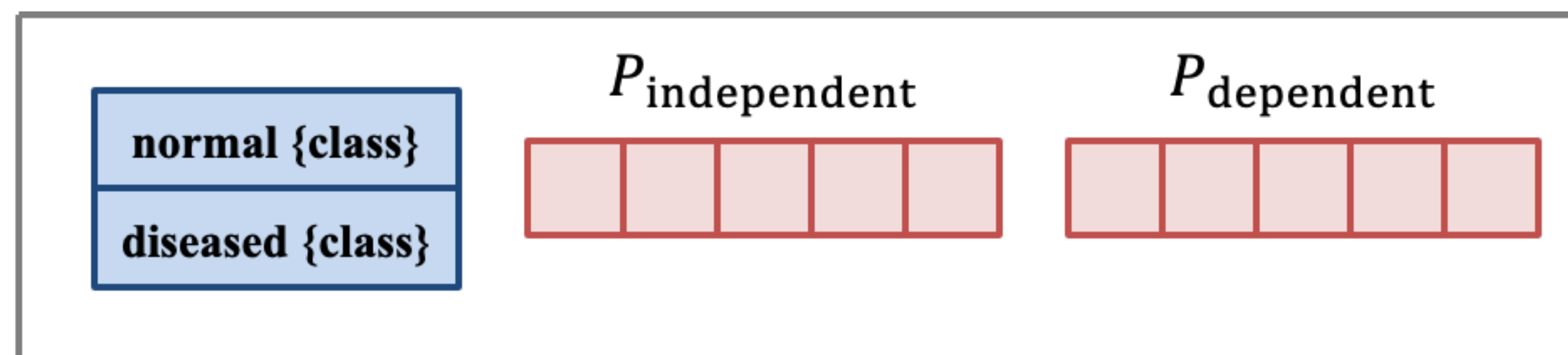
Methods



- ◆ **Image Adaptor**
 - Input: CLS token from final layer of the image encoder
 - Output: Prompt bias conditioned on input image



- ◆ **Prompt Design**
 - Learnable prompt is composed as below;



- $P_{\text{dependent}} \in \mathbb{R}^{B \times N \times d}$
 - π' is added to incorporate the information of the image.
- $P_{\text{independent}} \in \mathbb{R}^{B \times N \times d}$
- $P_{\text{class}} \in \mathbb{R}^{B \times 2 \times d}$
- Finally, the prompt fed into CLIP is as follows.
 - $P = \text{concat}(P_{\text{class}}, P_{\text{independent}}, P_{\text{dependent}})$
- In this work, we used ViT-L/14, and N=32
- $v \in \mathbb{R}^{B \times (16 \times 16 + 1) \times d}$ is obtained from last layer

- ◆ **Reweighting & Upscaling**
 - This makes patches into 224x224 anomaly maps.

Algorithm 1 Anomaly Map Computation

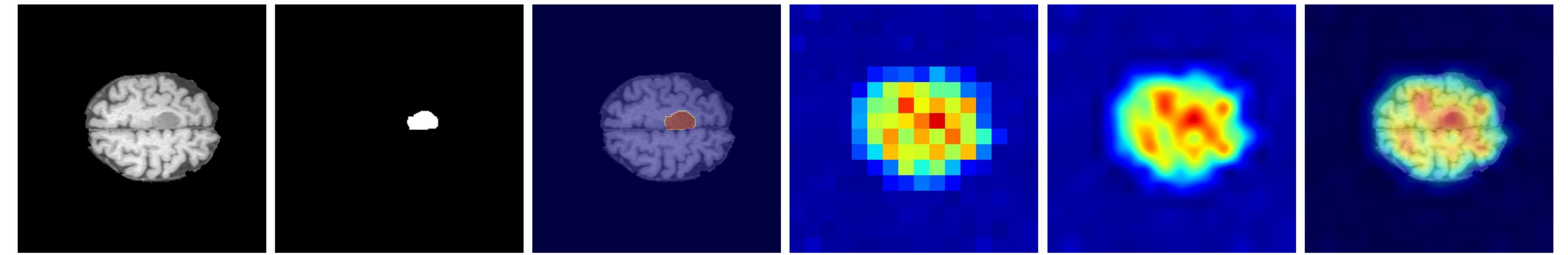
Require: Image Feature I , Prompt Feature P
Initialize M as an empty list
for $l = 1$ to L **do**
 Extract patch features F_l from $I[l]$
 Compute anomaly score for patch: $M_l = \cos(F_l, P_A) - \cos(F_l, P_N)$
end for
Compute weighted anomaly scores: $w = \text{softmax}(\frac{1}{HW} \sum_l M_l)$
Compute final anomaly map: $M_{\text{final}} = \sum w_l M_l$
Upscale M_{final} to size (224, 224) and normalize
Return M_{final}

Results

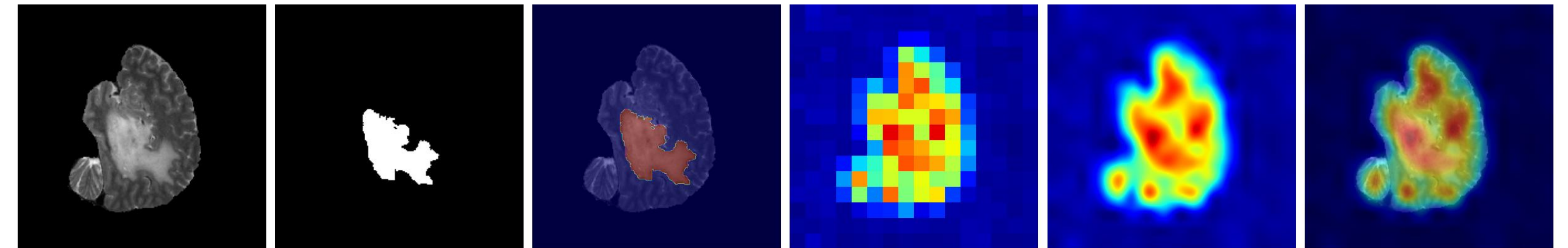
- ◆ **Experiment on diverse views and sequences**
 - Train on T1w MRI with axial Slices
 - Test on the others

MRI Sequences	Viewpoint			(Pixel-AUROC, %)
	Axial	Sagittal	Coronal	
T1	93.19	87.43	87.16	
T1CE	85.97	83.52	84.74	
T2	90.26	85.92	87.42	
FLAIR	88.43	84.78	85.20	

- ◆ **In-Domain (T1w MRI, axial slice)**



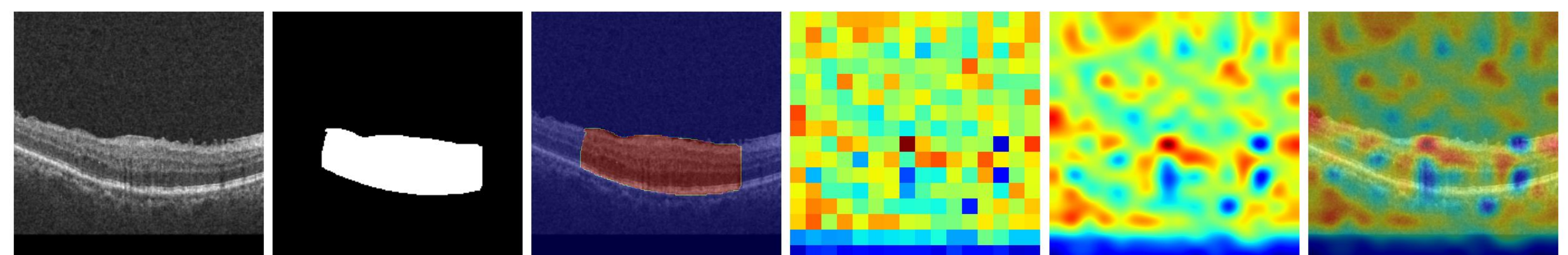
- ◆ **Domain Shift (T2w MRI, sagittal slice)**



- ◆ **Experiment on diverse organs and modalities**
 - Zero-Shot Transfer
 - Chest X-ray, Retina OCT

	Image-AUROC	Pixel-AUROC
Chest	55.47	N/A
Retina	55.83	53.15

- ◆ **Actually NOT a successful performance. Rather, FAILED.**
 - Still Notable, since better than random prediction



Discussion

- ◆ Map generation algorithm might not be optimal
 - Unstable due to non-parametric method, Interpolation
- ◆ Diverse class or domain during training can improve
 - Unstable due to non-parametric method, Interpolation
 - Fine tuning or using auxiliary data might be solution

Conclusion

- ◆ Without fine-tuning CLIP itself, prompt learning can guide CLIP to understand medical anomalies
- ◆ Without additional decoder, patch embeddings can represent the anomalous location at large