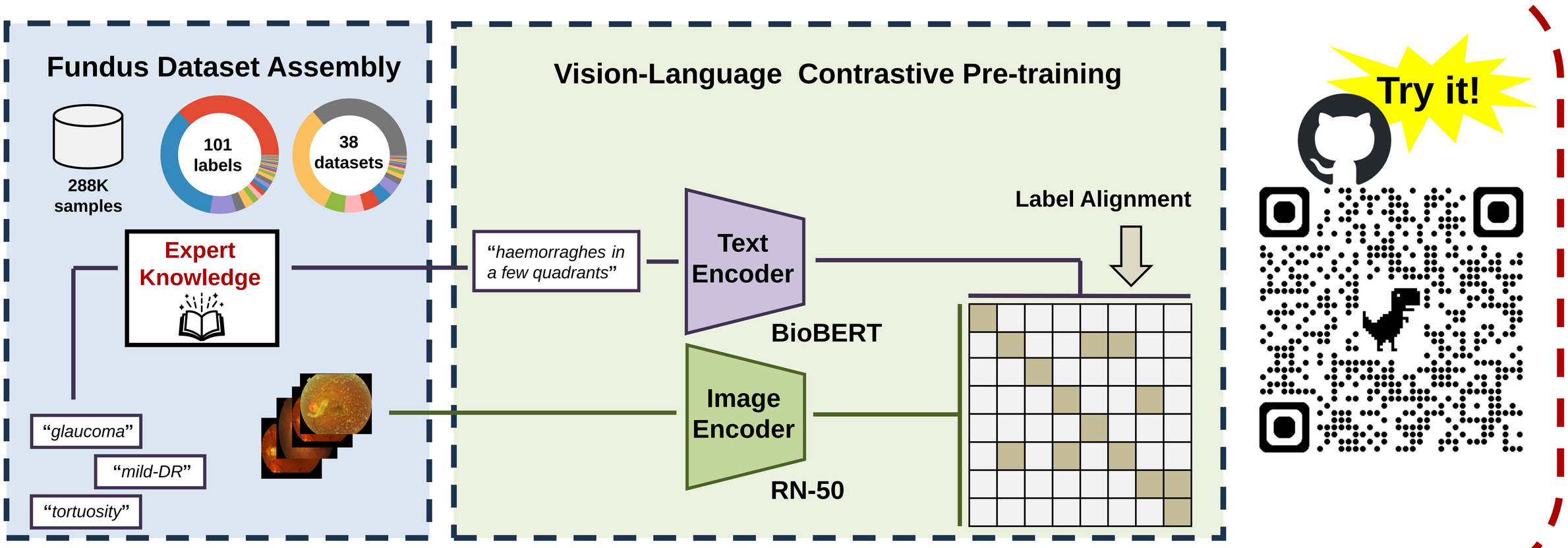
On the Importance of Expert Knowledge to Improve Foundation Models for Retinal Fundus Images

Julio Silva-Rodríguez[1] · Hadi Chakor[2] · Riadh Kobbi[2] ·
Jose Dolz[1] · Ismail Ben Ayed[1]
ÉTS Montreal[1]· DIAGNOS Inc.[2]

MIDL Paris 2024

A Foundation Language-Image Model of the Retina · FLAIR



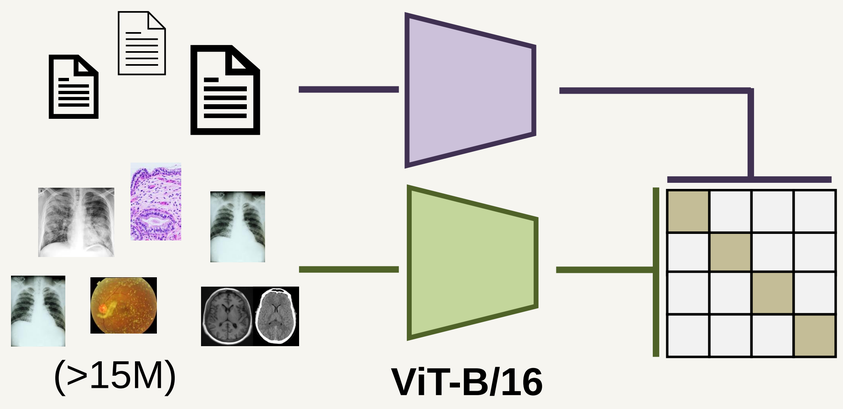
TAKE-HOME

In Nutshell

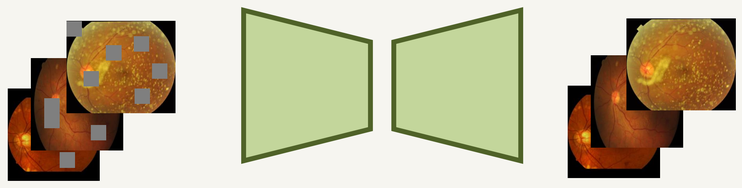
- **Expert descriptions encode hierarchical dependencies of lesions/diaseases.**
- **Specialized foundation models are prefered over larger-scale generalist or self-supervised models.**

COMPETITORS

BiomedCLIP



RETFound



METHODS

- **Vision-language-label pre-training.** Vision ,$\boldsymbol{\theta}$, and text , $\boldsymbol{\phi}$, encoders map input modalities into a common L^2 – normalized embedding space, with vision and language features $\boldsymbol{u}$ and $\boldsymbol{v}$. Pre-training consist of aligning paired projections in the label space:

$$L_{i2t}(\theta,\phi,\tau|B) = -\sum_{i\epsilon X_B}\frac{1}{|P_{\tau_B}(i)|}\sum_{i'\epsilon P_{\tau_B}(i)|}\log\frac{\exp(u_i^T\, v_{i'}/\tau)}{\sum_{j\epsilon\tau_B}\exp(u_i^T\, v_j/\tau)}$$

$$L_{t2i}(\theta,\phi,\tau|B) = -\sum_{j\epsilon\tau_B}\frac{1}{|P_{X_B}(j)|}\sum_{j'\epsilon P_{X_B}(j)|}\log\frac{\exp(u_{j'}^T\, v_j/\tau)}{\sum_{i\epsilon X_B}\exp(u_i^T\, v_j/\tau)}$$

- **Encoding expert knowledge.** We introduce *domain expert knowledge* by mapping categorical labels to ophtalmologist's descriptions.

Given a category, $y^* \rightarrow \{T^*\}_1^P = \pi_{EK}(y^*)$

"exudates" → "small white or yellow deposits with sharp margins"

RESULTS

(a) *Zero-shot*		MESSIDOR	FIVES	REFUGE	20x3	$\text{ODIR}_{200\text{x}3}$	MMAC	**Avg.**
CLIP	ViT-B/32	0.200	0.256	0.433	0.333	0.480	0.183	0.314
BiomedCLIP	ViT-B/16	0.207	0.415	0.624	0.617	0.583	0.274	0.453
FLAIR	RN50	**0.604**	**0.735**	**0.883**	**0.983**	**0.667**	**0.400**	**0.712**
(b) *Linear Probing*								
ImageNet	RN50	0.424	0.741	0.733	0.983	0.887	0.631	0.733
CLIP	ViT-B/32	0.491	0.800	0.720	0.950	0.917	0.642	0.753
BiomedCLIP	ViT-B/16	0.433	0.654	0.776	0.866	0.883	0.678	0.715
RETFound	ViT-B/16	0.457	0.765	0.747	0.950	0.887	0.547	0.725
FLAIR	RN50	**0.719**	**0.879**	**0.843**	**1.000**	**0.935**	**0.740**	**0.852**