Label-Assemble: Leveraging Multiple Datasets with Partial Labels

Mintong Kang[1], Bowen Li[2], Zengle Zhu[3], Yongyi Lu[2], Elliot K. Fishman[4], Alan Yuille[2], Zongwei Zhou[2]

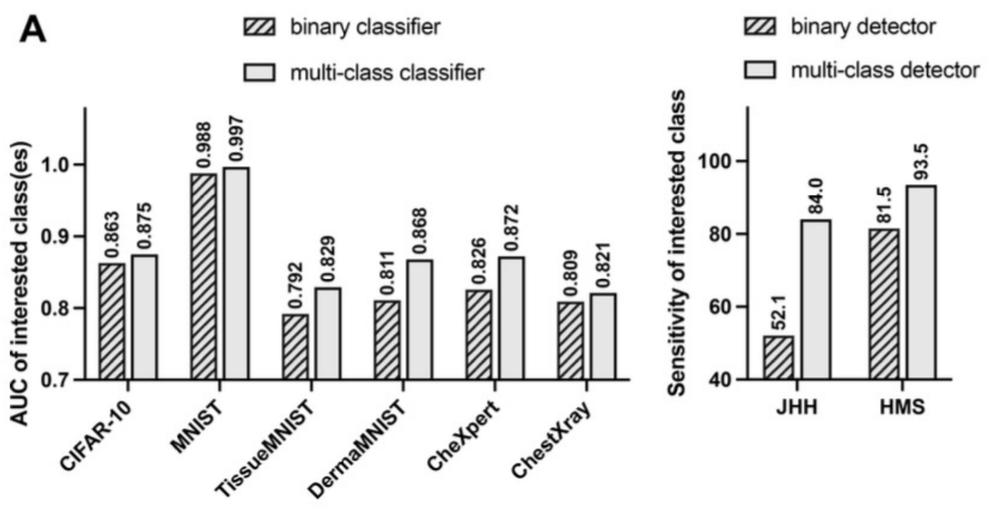
[1]University of Illinois Urbana-Champaign [2]Johns Hopkins University [3]Tongji University [4]Johns Hopkins University School of Medicine

Introduction

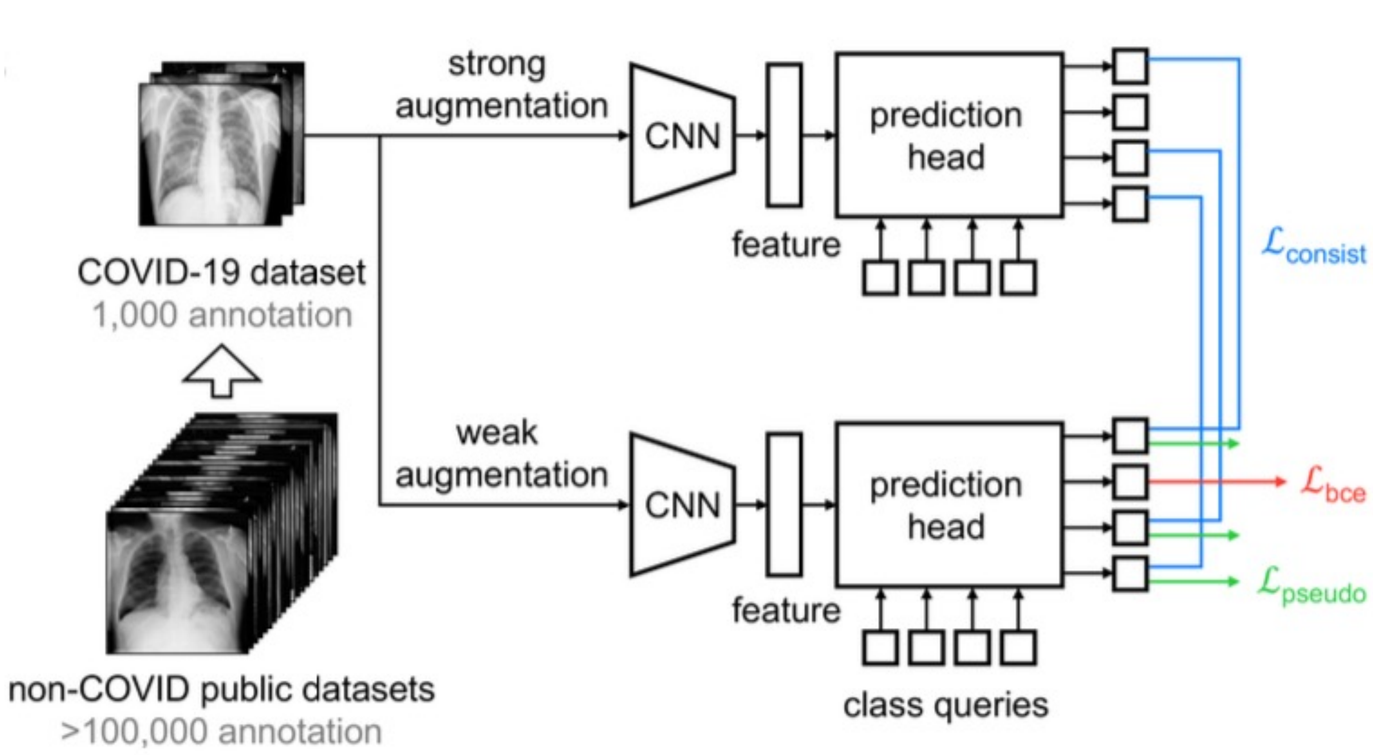
- **Challenge**: labels in public medical datasets are often incomparable, heterogeneous, or even conflicting
- **Question**: Can we integrate and exploit such a great number of publicly available datasets with *partial labels* to achieve an improved computer-aided diagnosis and detection of specific diseases?
- **Principal hypothesis:** a dataset that is labeled with various classes can foster more powerful models than one that is only labeled with the class of interest.
- **Method (label-assemble)**: 1) a new *class query* to encode different visual tasks, which can dynamically integrate partial labels across different datasets; 2) pseudo labels and consistency constraints are introduced for the missing part of labels and for mitigating the domain gap across different datasets.
- **Validation:** effectiveness of Label-Assemble in both computer-aided disease diagnosis and detection, supported by two clinical applications.

Motivation

- With the same amount of data, learning from classes of "negative examples" benefits the learning of the interested class.
- Verified by six classification tasks and two detection tasks.



Label-Assemble



Dynamic adapter with learnable class queries:

The output (a) is dynamically generated by class queries (q) and input image (x) as:

$$a = w(q;\theta_w) * f(x)$$

where $w(\cdot)$ parameterized with θ_w transforms class queries into classification weights and $f(\cdot)$ is the pre-trained feature extractor.

Pseudo labels & consistency constraints:

To unleash the full potential of unannotated labels, we introduce a sharpening operator to generate pseudo-labels:

$$\tilde{a} = \begin{cases} a + (1-a)/t & a \geq \tau \\ a - a/t & a \leq \tau \end{cases}$$

where t is the sharpen temperature and τ is the threshold (τ=0.5 in our experiments).

Loss to guide self-training on unlabeled data:

$$\mathcal{L}_{\text{pseudo}} = ||a_w - \tilde{a}_w||_2$$

where a_w denotes the answer of weakly augmented images.

Consistency loss can be formulated as:

$$\mathcal{L}_{\text{consist}} = ||a_s - \tilde{a}_w||_2$$

where a_s denotes the answer of strongly augmented images.

Overall loss function:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{bce}} + \mathcal{L}_{\text{pseudo}} + \mathcal{L}_{\text{consist}}$$

where $\mathcal{L}_{\text{bce}}$ is the binary cross-entropy loss.

Experimental Results

➢ Label-Assemble outperforms other methods developed for partial labels, and performs on par with the method of learning from *full* labels

Method	# labels	CheXpert (val) Card†	Pneu1†	Atel†	Edema	Effusion	Average	ChestX-ray14 (val) Cons†	Pneu2†	Atel†	Edema	Effusion	Average
DenseNet [19]	37,655	0.646	0.461	0.431	0.791	0.800	0.626	0.693	0.640	0.688	0.737	0.783	0.708
Med3D [20]	75,310	0.751	0.629	0.663	0.839	0.836	0.744	0.700	0.758	0.718	0.732	0.788	0.739
DoDNet [10]	75,310	0.778	0.598	0.646	0.859	0.845	0.745	0.706	0.756	0.721	0.745	0.769	0.740
Ours	75,310	0.832	0.675	**0.702**	**0.867**	**0.886**	0.792	**0.744**	0.805	**0.813**	0.710	0.778	0.770
DenseNet [19]	105,434	**0.835**	**0.683**	0.699	0.864	0.885	**0.793**	0.719	**0.810**	0.740	**0.811**	**0.812**	**0.778**

†Card, Pneu1, Atel, Cons, Pneu2 denote Cardiomegaly, Pneumonia, Atelectasis, Consolidation Pneumothorax, respectively.

➢ Label-Assemble achieves the best mean performance over all 14 thorax diseases on ChestXray-14 (official split).

	Ref. & Year	Architecture	mAUC
Ma et al. [23]	MICCAI 2019	DenseNet (×2)	0.817
Hermoza et al. [22]	MICCAI 2020	DenseNet121	0.821
Kim et al. [21]	CVPR 2021	DenseNet121	0.822
Taslimi et al. [24]	arXiv 2022	SwinT	0.810
Xiao et al. [25]	WACV 2022	ViT-S	0.823
Ours		DenseNet121	**0.832**

➢ Assembling labels of other chest diseases improves lung nodule classification.

➢ The performance gain is positively correlated to inter-class similarity between nodule and the assembled disease. The Pearson Coefficient is r = 0.83; p = 4.93e-4.