Acknowledging the Unknown for Multi-label Learning with Single Positive Labels

Donghao Zhou[1,2], Pengfei Chen[3], Qiong Wang[1], Guangyong Chen[4*], Pheng-Ann Heng[1,5]

[1] SIAT, CAS [2] UCAS [3] Tencent [4] Zhejiang Lab [5] CUHK

Single Positive Multi-label Learning

	person	dog	bus	bicycle	apple	boat
(a)	✓	×	✓	✓	×	×
(b)	✓	×	?	✓	?	×
(c)	✓	?	?	?	?	?

✓: positive labels ×: negative labels ?: unknown Labels

(a): Multi-label Learning (b): Multi-label Learning with Missing Labels (MLML) (c): Single Positive Multi-label Learning (SPML)

In SPML, each multi-label training image **has only one positive label** and **other labels remain unannotated**.

Traditional Solution

Assuming-Negative (AN) Loss: assumes all unannotated labels are negative and follows BCE loss.

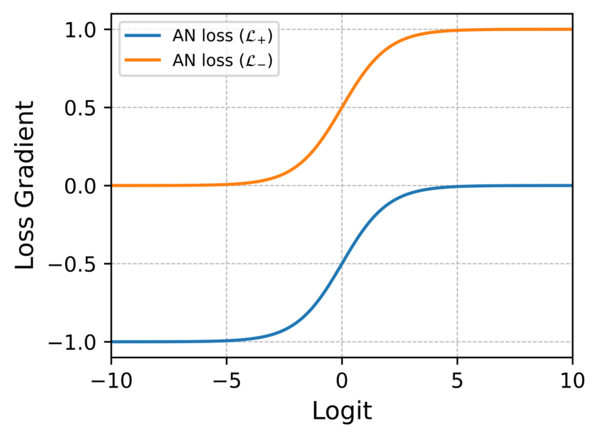
$$\mathcal{L}_{\text{AN}}(\mathbf{f}^{(n)},\mathbf{y}^{(n)}) = -\frac{1}{C}\sum_{c=1}^{C}[\mathbb{1}_{[y_c^{(n)}=1]}\log(f_c^{(n)}) + \mathbb{1}_{[y_c^{(n)}=0]}\log(1-f_c^{(n)})]$$

The gradient regime of AN loss:

$$\mathcal{L}_+ = -\log(p) \qquad \mathcal{L}_- = -\log(1-p) \quad \Rightarrow \quad \begin{cases} \frac{\partial \mathcal{L}_+}{\partial g} = \frac{\partial \mathcal{L}_+}{\partial p}\frac{\partial p}{\partial g} = \frac{-e^{-g}}{1+e^{-g}}, & y_c^{(n)}=1 \\ \frac{\partial \mathcal{L}_-}{\partial g} = \frac{\partial \mathcal{L}_-}{\partial p}\frac{\partial p}{\partial g} = \frac{1}{1+e^{-g}}, & y_c^{(n)}=0 \end{cases}$$

It results in three issues:

1. Dominance of Assumed Negative Labels
2. Introduced Label Noise
3. Over-Suppression for Confident Positive Predictions



Motivation

Una. labels need to be treated with **a better gradient regime**. Instead of making an unrealistic assumption, we choose to **acknowledge the fact that they are unknown**.

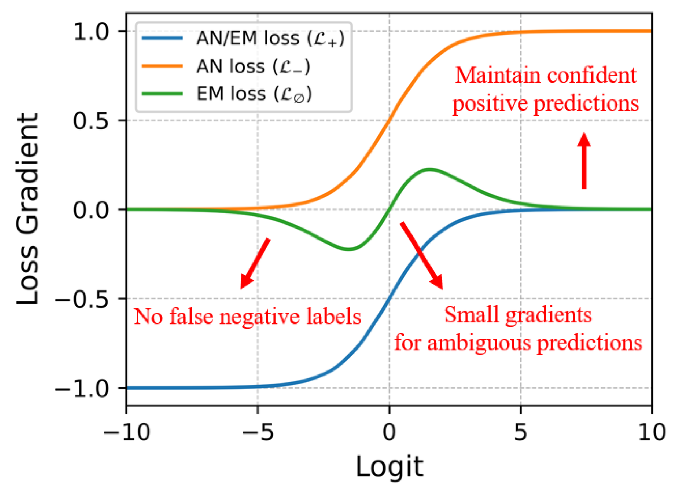
Acknowledging the Unknown

1. **Entropy-Maximization (EM) Loss**: maximizes the entropy of predicted probabilities for una. labels.
2. **Asymmetric Pseudo-Labeling (APL)**: adopts asymmetric-tolerance PL strategies.

$$\mathcal{L}_{\text{EM}}(\mathbf{f}^{(n)},\mathbf{y}^{(n)}) = -\frac{1}{C}\sum_{c=1}^{C}[\mathbb{1}_{[y_c^{(n)}=1]}\log(f_c^{(n)}) + \mathbb{1}_{[y_c^{(n)}=0]}\alpha H(f_c^{(n)})]$$

$$H(f_c^{(n)}) = -[f_c^{(n)}\log(f_c^{(n)}) + (1-f_c^{(n)})\log(1-f_c^{(n)})]$$

EM loss



APL

Algorithm 1 Asymmetric Pseudo-Labeling

Input: Training set $\mathcal{D}$ and model f_{T_w} trained with Eq. 3 for T_w epochs
Parameter: Total training epoch T_t, sample proportion $\theta\%$ and loss weight β
Output: Well-trained model f_i

```
i ← T_w, θ'% ← θ%/(T_t − T_w)
repeat
    Generate pseudo-labels using f_i by following Eq. 6
    Train f_{i+1} from f_i with Eq. 8
    i ← i + 1
until early stopping or i = T_t
return f_i
```

Benchmark Results

Experimental results with mAP on four large-scale multi-label datasets

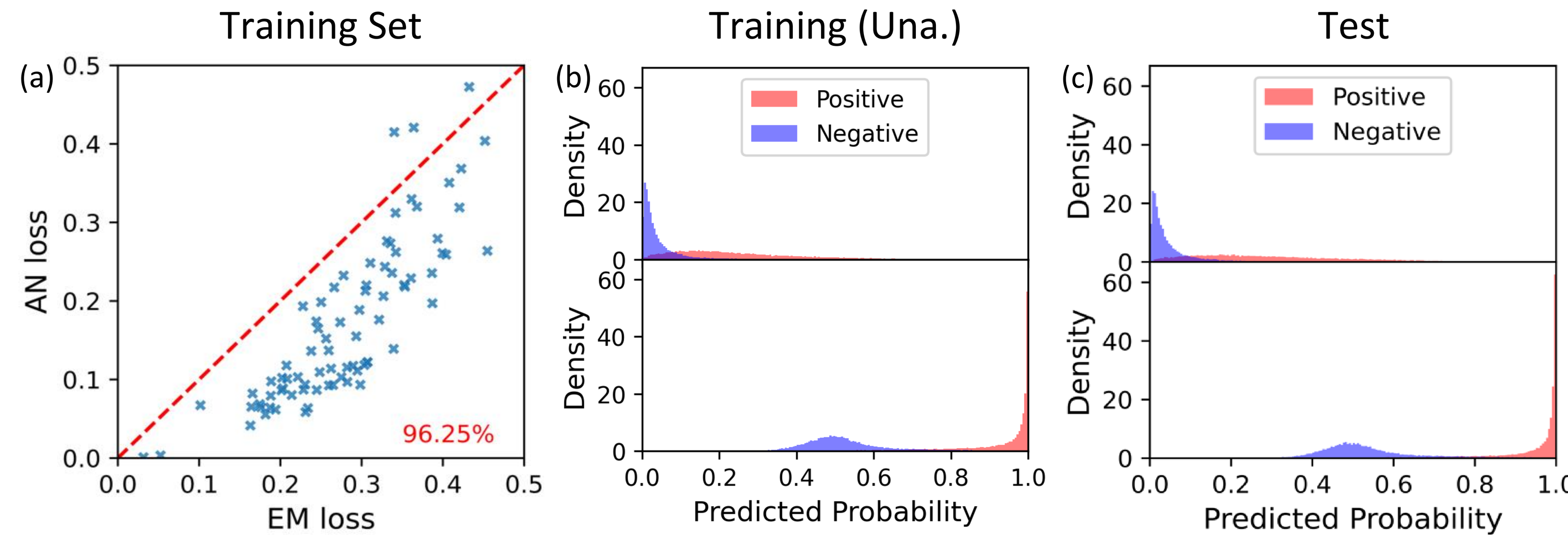
Ann. Labels	Methods	VOC	COCO	NUS	CUB
All P. & All N.	BCE loss	89.42±0.27	76.78±0.13	52.08±0.20	30.90±0.64
1 P. & All N.	BCE loss	87.60±0.31	71.39±0.19	46.45±0.27	20.65±1.11
1 P. & 0 N.	AN loss	85.89±0.38	64.92±0.19	42.27±0.56	18.31±0.47
	DW	86.98±0.36	67.59±0.11	45.71±0.23	19.15±0.56
	L1R	85.97±0.31	64.44±0.20	42.15±0.46	17.59±1.82
	L2R	85.96±0.36	64.41±0.24	42.72±0.12	17.71±1.79
	LS	87.90±0.21	67.15±0.13	43.77±0.29	16.26±0.45
	N-LS	88.12±0.32	67.15±0.10	43.86±0.54	16.82±0.42
	EntMin	53.16±2.81	32.52±5.55	19.38±3.64	13.08±0.15
	Focal loss	87.59±0.58	68.79±0.14	47.00±0.14	19.80±0.30
	ASL	87.76±0.51	68.78±0.32	46.93±0.30	18.81±0.48
	ROLE	87.77±0.22	67.04±0.19	41.63±0.35	13.66±0.24
	ROLE+LI	88.26±0.21	69.12±0.13	45.98±0.26	14.86±0.72
1 P. & 0 N.	EM loss	*89.09±0.17*	*70.70±0.31*	*47.15±0.11*	*20.85±0.42*
	EM loss+APL	**89.19±0.31**	**70.87±0.23**	**47.59±0.22**	**21.84±0.34**

➡ Oracles (BCE loss rows)
➡ AN Loss and Improved AN Loss (AN loss – N-LS)
➡ Other Comprising Methods (EntMin – ROLE+LI)
➡ Ours (EM loss, EM loss+APL)

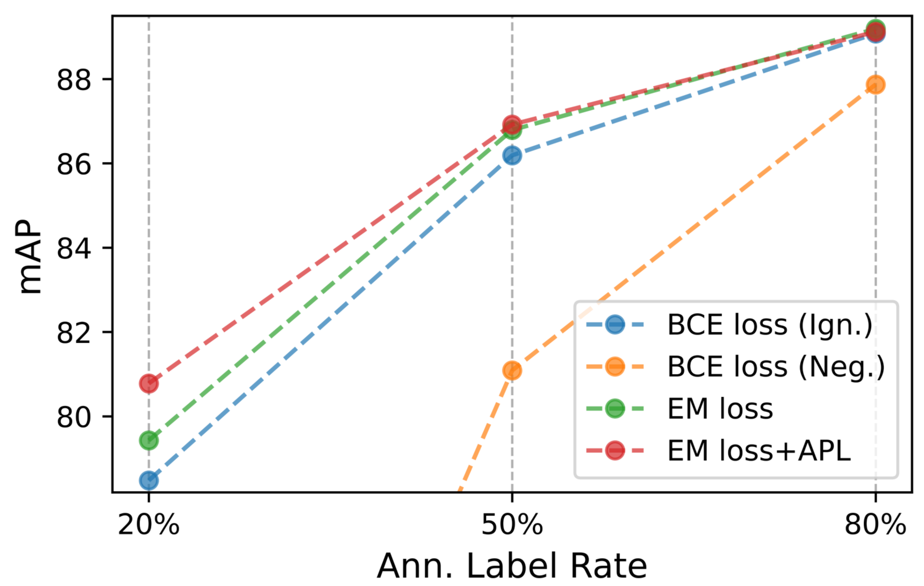
The proposed method achieves **SOTA results** on all four benchmarks, and even **approaches to the results of training with full annotations** in some cases.

Further Analysis

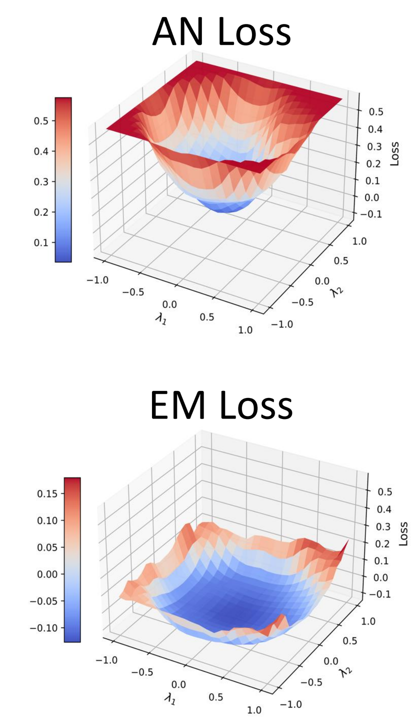
Distinguishability of the predictions for pos. and neg. labels
(a): Wass. distances (b)&(c): Densities of an example class



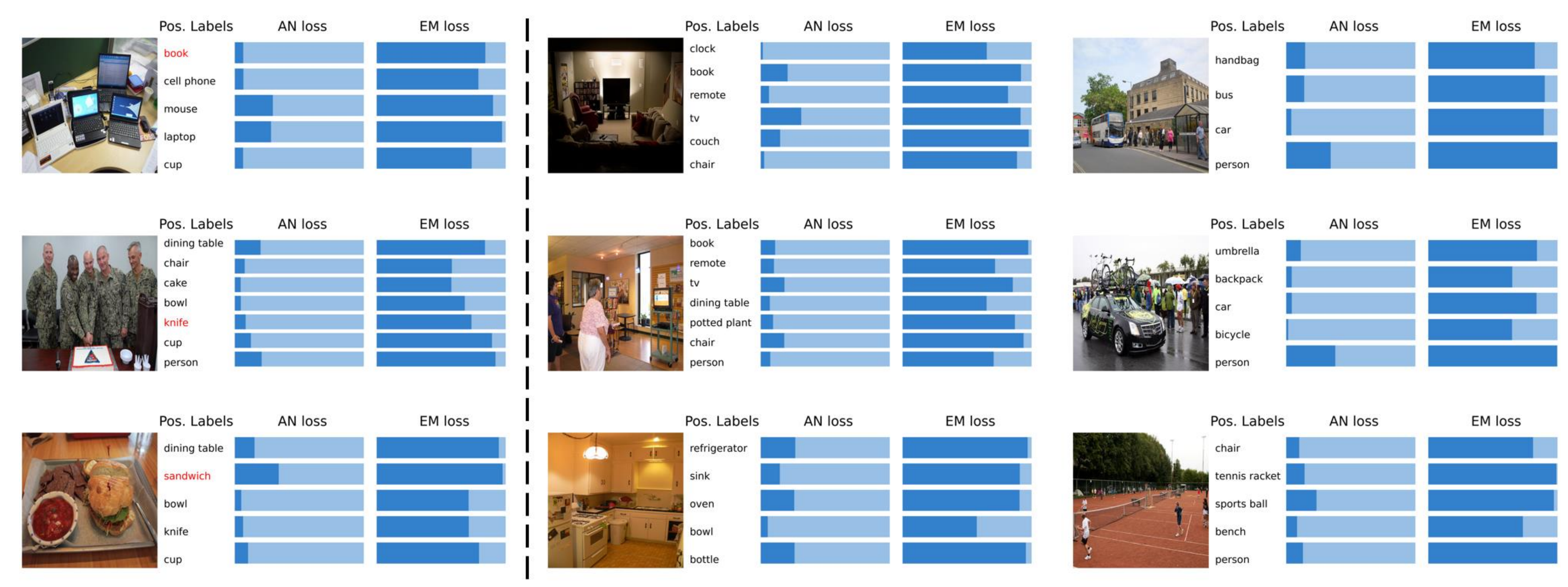
Performance in a more general scenario (MLML)



Generalization Evaluation by Loss Landscapes



Qualitative Results



Training Image Test Image

Paper and Code are publicly available: https://github.com/Correr-Zhou/SPML-AckTheUnknown