Re-Evaluating the Impact of Unseen-Class Unlabeled Data on Semi-Supervised Learning Model

Rundong He[2] , Yicong Dong[2] , Lanzhe Guo[3] , Yilong Yin[2†] , Tailin Wu[1†]

[1] Department of Artificial Intelligence, School of Engineering, Westlake University
[2] School of Software, Shandong University
[3] School of Intelligence Science and Technology, Nanjing University

Problem Statement & Motivations

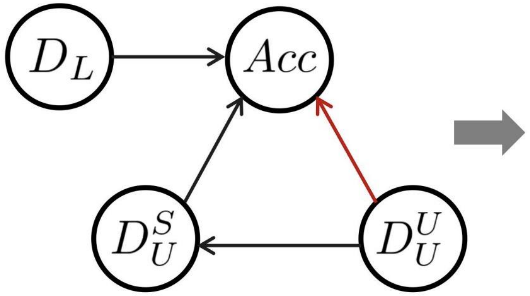
Current SSL Paradigm

• Dominant assumption: Unlabeled data shares *identical class space* with labeled data
• Widely adopted safe-SSL methods (DS3L, OpenMatch, Fix-A-Step) claim:
✖ Unseen-class unlabeled data *necessarily harms* model performance

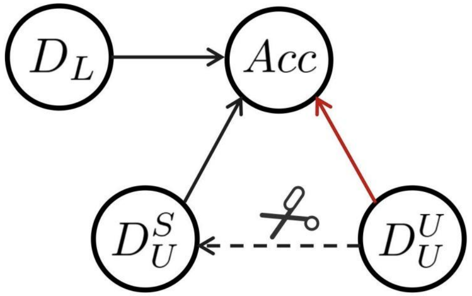
Critical Flaw in Prior Evaluations

• Violation of of controlling variables

Previous works fix total unlabeled data size while varying unseen-class ratio r_u
Creates confounding variables:
✓ $r_u \uparrow \rightarrow$ seen-class samples $D_U^S \downarrow$ (e.g., CIFAR10: r_u=0.6 reduces seen-class samples by 60%)
✓ Performance drop could stem from insufficient seen-class samples, not unseen-class presence



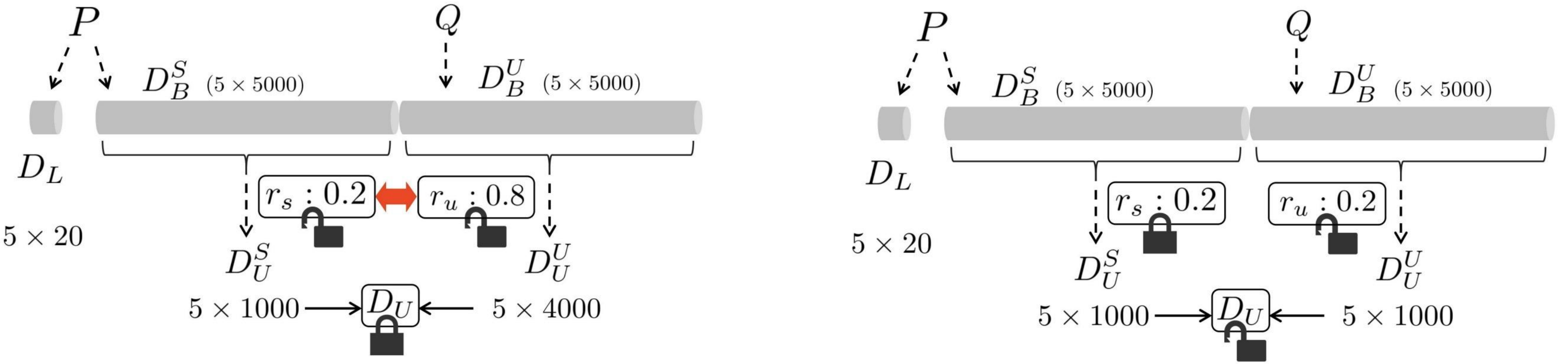
(1) old



(2) ours

Insight: Removing the spurious correlation between seen-class unlabeled data D_U^S and unseen-class unlabeled data D_U^U is critical.

RE-SSL Evaluation Framework



(a) Previous evaluation (b) RE-SSL evaluation

Key Components：
1. Controlled Variable Protocol: Fix seen-class ratio, vary unseen-class ratio
2. Five Impact Dimensions: Unseen classes’ sample-number, category-number, category-index, nearness, label distribution
3. Global and local robustness analysis: Slope of regression function, global magnitude, worst-case adjacent discrepancy, best-case adjacent discrepancy, probability

Key Experimental Findings

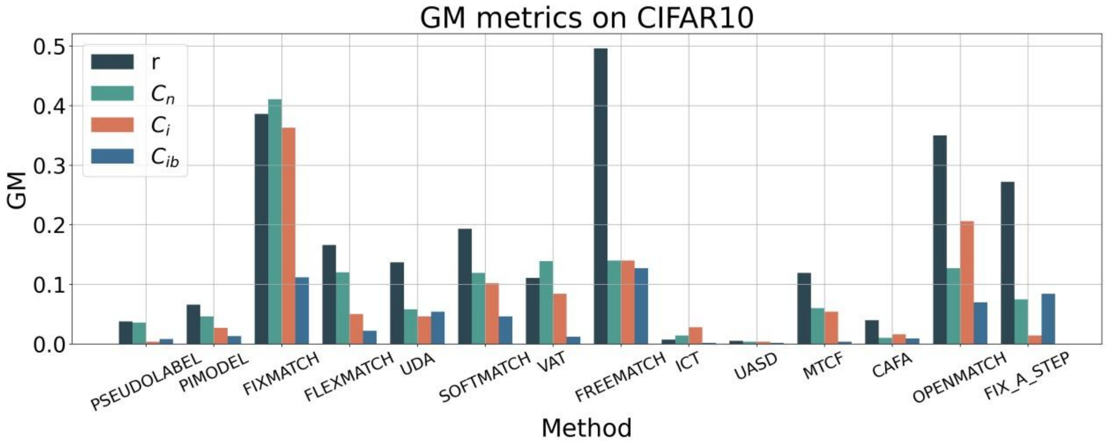
“Unseen Classes Do NOT Always Hurt SSL Performance!”

Table 1: Evaluation on CIFAR10 with 100 labels under inconsistent label spaces.

METHOD	r=0	r=0.2	r=0.4	r=0.5	r=0.6	r=0.8	r=1.0	R_{slope}	GM	BAD	WAD	$P_{AD\geq 0}$
SUPERVISED	0.617	0.617	0.617	0.617	0.617	0.617	0.617	0.000	0.000	0.000	0.000	1.000
PSEUDOLABEL	0.677	0.668	0.664	0.660	0.660	0.660	0.654	-0.020	0.038	**0.000**	-0.045	0.333
PIMODEL	0.667	0.651	0.644	0.637	0.636	0.636	0.630	-0.034	0.066	**0.000**	-0.080	0.167
FIXMATCH	0.612	0.547	0.515	0.514	0.505	0.442	0.365	-0.223	0.386	-0.010	-0.385	0.000
FLEXMATCH	0.770	0.717	0.702	0.704	0.704	0.686	0.638	-0.107	0.166	0.020	-0.265	0.333
MIXMATCH	0.731	0.708	0.708	0.706	0.701	0.697	0.676	-0.045	0.075	**0.000**	-0.115	0.167
UDA	0.645	0.621	0.604	0.601	0.591	0.590	0.552	-0.082	0.137	-0.005	-0.190	0.000
SOFTMATCH	0.764	0.710	0.689	0.697	0.692	0.682	0.607	-0.112	0.193	0.080	-0.375	0.167
VAT	0.710	0.677	0.663	0.663	0.657	0.651	0.639	-0.063	0.111	**0.000**	-0.165	0.167
FREEMATCH	0.760	0.723	0.665	0.635	0.608	0.605	0.440	-0.287	0.496	-0.015	-0.825	0.000
ICT	0.618	0.619	0.621	0.620	0.621	0.619	0.621	**0.002**	0.007	0.010	**-0.010**	**0.667**
UASD	0.618	0.619	0.617	0.616	0.617	0.618	0.618	0.000	**0.005**	0.010	**-0.010**	**0.667**
MTCF	0.772	0.743	0.731	0.723	0.725	0.716	0.692	-0.070	0.119	0.020	-0.145	0.167
CAFA	0.652	0.652	0.640	0.653	0.641	0.642	0.640	-0.013	0.040	0.130	-0.120	0.500
OPENMATCH	0.713	0.606	0.586	0.595	0.579	0.517	0.473	-0.211	0.350	0.090	-0.535	0.167
FIX_A_STEP	0.662	0.634	0.615	0.582	0.555	0.585	0.509	-0.138	0.272	0.150	-0.380	0.167

✓ From the perspective of global robustness: SSL algorithms PseudoLabel and ICT, as well as the robust SSL algorithms UASD and CAFA, are robust to unseen classes.

✓ From the perspective of local robustness: FlexMatch and ICT display relatively better performance stability, both under worst-case and best-case scenarios.



SSL learning models are most sensitive to the quantity of unseen class samples when dealing with unseen categories, while being relatively robust to changes in the label distribution of unseen classes.

Thank you!