SemiReward: A General Reward Model for Semi-supervised Learning

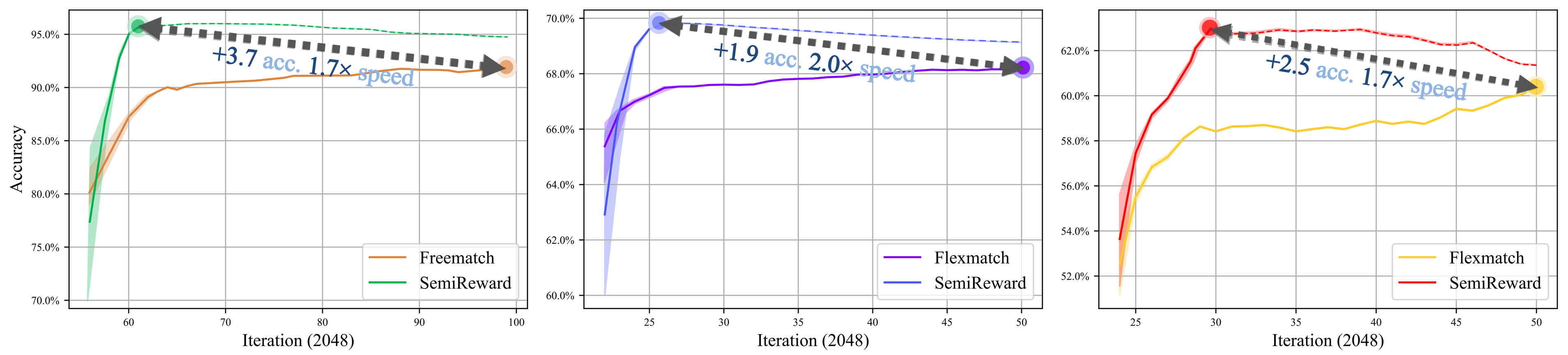
The 12th Internaltional Conference on Learning Representations

Siyuan Li[1,2,*], **Weiyang Jin**[2,*], Zedong Wang[1,2], Zicheng Liu[1,2], Cheng Tan[1,2], and Stan Z. Li[2,#]

[1] Zhejiang University, [2] Westlake University, [*] Equal contribution [#] Corresponding author

Summary of Contributions

- We first introduce the reward model to Semi-supervised Learning (SSL) with a well-defined reward score as the pseudo-label quality indicator (for Pseudo-Labeling).
- A plugin-and-play semi-supervised reward framework is designed to filter out high-quality pseudo labels for classification and regression SSL tasks.
- SemiReward shows significant performance gains and faster convergence speeds on 13 standard SSL datasets across three modalities applying up SSL algorithms.



(a) CV: Euro-SAT (20) (b) NLP: Yahoo! Answer (2000) (c) Audio: ESC-50 (250)

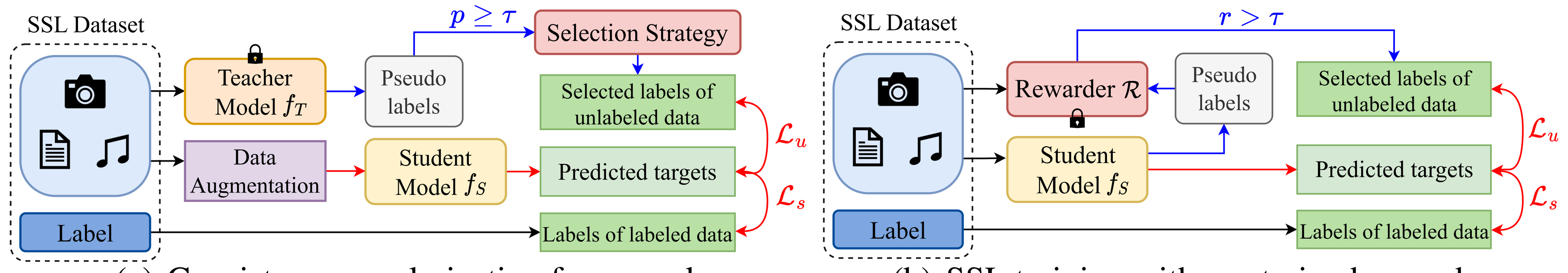
Figure 1. Top-1 Acc v.s. training iterations (×2048) on SSL datasets of three modalities.

Quality Indicator: Reward Score

Given a labeled dataset $D_L = \{x_i^l, y_i^l\}_{i=1}^{N_L}$ and an unlabeled dataset $D_U = \{x_i^u\}_{i=1}^{N_U}$ with the sample number $N_L \ll N_U$. SSL is to train student model $f_S(x) = y \in \mathbb{R}^C$ with D_L and the pseudo-label set $\widehat{D}_U = \{x_i^y, \hat{y}_i^u\}$, selected by $\hat{y}^u = \mathbb{I}(y^u, \tau)$. We parameterize $\mathbb{I}(.,.)$ by a new reward score:

$$r(y^u, y^l) = \mathcal{S}(y^u, y^l) \simeq \mathcal{R}(x, y^u) \in [0,1].$$

$$\mathcal{S}(y_i, y_j) = \frac{y_i \cdot y_j}{2\,\|y_i\|\,\|y_j\|} + 0.5 \in [0,1].$$

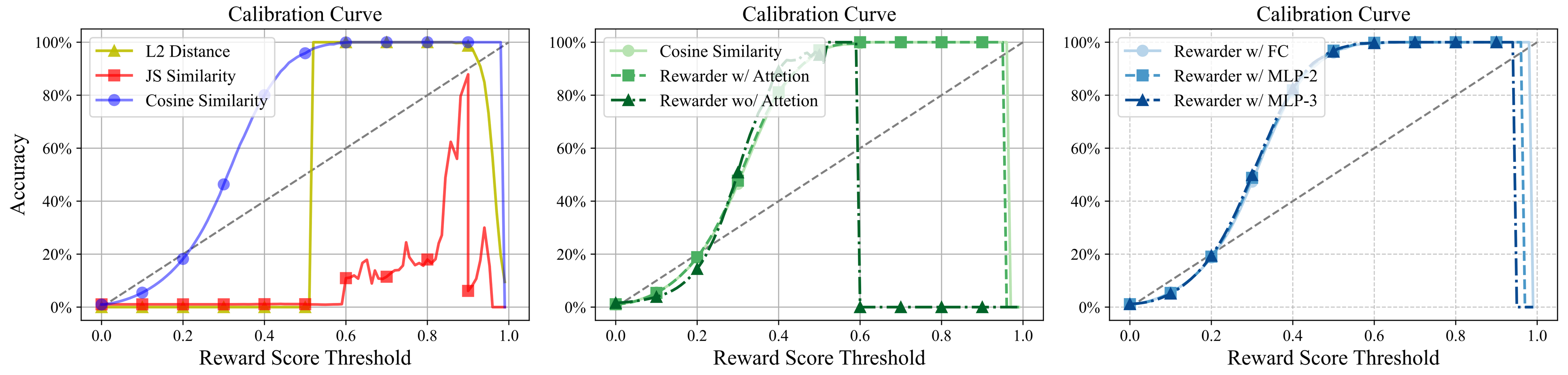


(a) Consistency regularization framework (b) SSL training with pre-trained rewarder

Figure 2. Illustration of SSL training paradigm, where blue lines denote pseudo-labeling pipeline and red lines denote gradient propagation, and (b) is SemiReward.

Formulating the rewarder $\mathcal{R}(.,.)$ to approximate $r(y^u, y^l)$:

$$\mathcal{R}(x^u, y^u) = \mathrm{Sigmoid}\Big(\mathrm{MLP}\Big(\mathrm{CA}\big(\mathrm{Emb}(f(x^u)), \mathrm{Emb}(y^u)\big)\Big)\Big)$$



(a) Various reward similarities (b) Attention module in Rewarder (c) MLP module in Rewarder

Figure 3. How $\mathcal{R}$ works illustrated by reward scores v.s. top-1 Acc on CIFAR-100 (400).

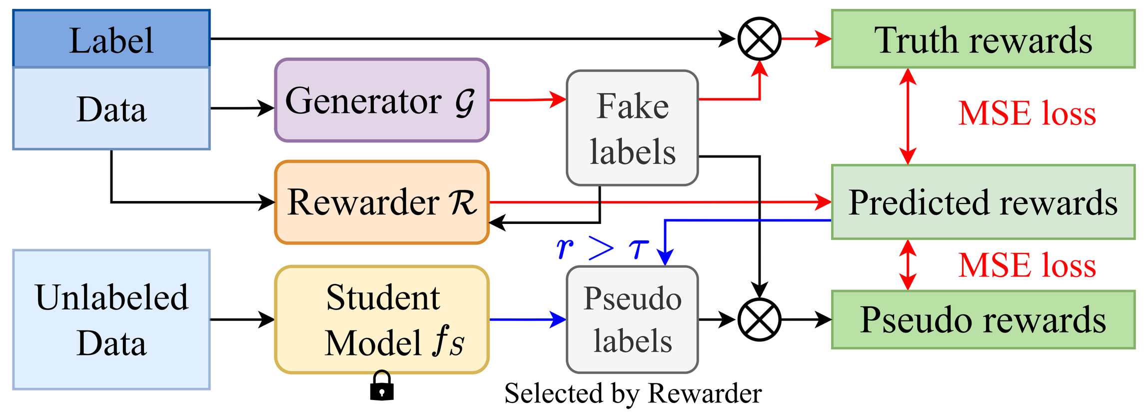
SemiReward Learning Framework

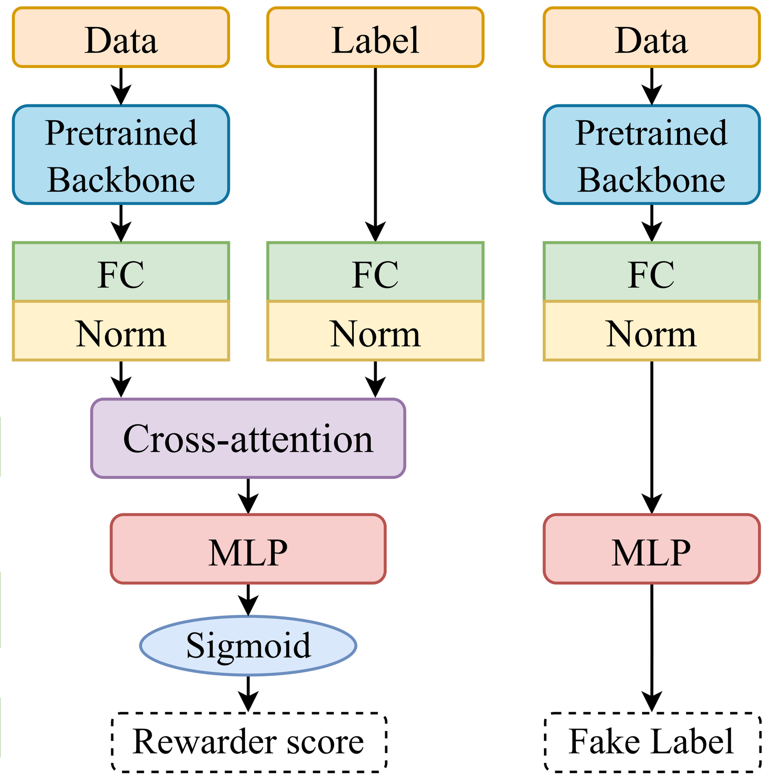
Two-stage SSL training with a schedule T and a Generator $\mathcal{G}(x) = y^f$ alternatively optimizing $\mathcal{L}_{aux} = \mathcal{L}_{\mathcal{R}} + \mathcal{L}_{\mathcal{G}}$.

$$\mathcal{L}_{\mathcal{R}} = \frac{1}{B_R}\sum_{i=1}^{B_R} \ell_2\Big(\mathcal{R}\big(x_i^r, \overline{\mathcal{G}}(x_i^r)\big), \mathcal{S}\big(y_i^r, \overline{\mathcal{G}}(x_i^r)\big)\Big)$$

$$\mathcal{L}_{\mathcal{G}} = \frac{1}{B_R}\sum_{i=1}^{B_R} \ell_2\Big(\overline{\mathcal{R}}\big(x_i^r, \mathcal{G}(x_i^r)\big), 1\Big)$$

Figure 4. Training pipeline and network architecture.





(a) Rewarder (b) Generator

Experiment Results

- Comparison experiments upon existing SSL methods with 13 classification and regression datasets of audio, natural language, vision modalities based on USB codebase, reporting top-1 error rate and speed-up times.

Domain	Dataset (Setting)	Pseudo Label Base	Pseudo Label +SR	FlexMatch Base	FlexMatch +SR	SoftMatch/FreeMatch Base	SoftMatch/FreeMatch +SR	Average Gain	Average Speed.
Audio	ESC-50 (250)	38.42±0.85	**33.33**±0.97	36.83±0.51	**32.58**±0.51	32.71±0.82	**29.71**±0.64	+4.11	×1.73
	ESC-50 (500)	28.92±0.24	**27.65**±0.32	27.75±0.41	**25.92**±0.31	29.07±1.27	**25.98**±0.49	+2.06	×2.07
	FSDnoisy18k (1773)	34.60±0.55	**33.24**±0.82	26.29±0.17	**25.63**±0.28	29.39±1.83	**26.10**±0.83	+1.77	×1.30
	UrbanSound8k (100)	37.74±0.96	**36.47**±0.65	37.88±0.46	**36.06**±0.93	37.68±1.82	**34.97**±1.02	+1.93	×1.70
	UrbanSound8k (400)	27.45±0.96	**25.27**±0.65	23.78±0.46	**23.45**±0.93	23.78±0.13	**19.39**±0.33	+2.30	×1.08
NLP	AG News (40)	15.19±3.07	**13.90**±0.21	13.08±3.94	**12.60**±0.69	11.69±0.12	**10.67**±0.90	+0.93	×2.77
	AG News (200)	14.69±1.88	**12.10**±0.58	12.08±0.73	**11.05**±0.14	11.75±0.17	**10.02**±0.82	+1.78	×2.30
	Yahoo! Answer (500)	34.87±0.50	**35.08**±0.40	34.73±0.09	**33.64**±0.73	33.02±0.02	**30.92**±0.90	+0.99	×1.80
	Yahoo! Answer (2000)	33.14±0.70	**32.50**±0.42	31.06±0.32	**29.97**±0.10	30.34±0.18	**29.11**±0.15	+0.99	×3.53
	Yelp Review (250)	46.09±0.15	**42.99**±0.14	46.09±0.15	**42.76**±0.33	43.91±0.19	**42.68**±0.12	+2.55	×1.40
	Yelp Review (1000)	44.06±0.14	**42.08**±0.15	40.38±0.33	**37.58**±0.19	40.43±0.12	**38.43**±0.14	+2.26	×1.01
CV	CIFAR-100 (200)	32.78±0.20	**31.94**±0.57	25.72±0.35	**23.74**±1.39	21.07±0.72	**20.06**±0.41	+1.28	×1.04
	CIFAR-100 (400)	25.16±0.67	**23.84**±0.20	17.80±0.57	**17.59**±0.35	15.97±0.24	**15.62**±0.71	+0.63	×1.57
	STL-10 (40)	20.53±0.12	**17.37**±0.47	11.82±0.51	**10.20**±1.11	17.51±0.61	**9.72**±0.62	+4.19	×1.07
	STL-10 (100)	11.25±0.81	**10.88**±1.48	7.13±0.20	**7.59**±0.57	8.10±0.35	**7.10**±1.39	+0.30	×1.11
	Euro-SAT (20)	25.25±0.72	**23.65**±0.41	5.54±0.16	**4.86**±1.00	5.51±0.54	**4.22**±0.34	+1.19	×1.03
	Euro-SAT (40)	12.82±0.81	**8.33**±0.33	4.51±0.24	**3.88**±0.69	5.46±0.34	**3.94**±0.71	+2.21	×1.13

Table 2: RMSE and MAE, performance gain, and training speedup times on three SSL regression datasets with 1% labels.

Method	RCF-MNIST RMSE	RCF-MNIST MAE	IMDB-WIKI RMSE	IMDB-WIKI MAE	AgeDB RMSE	AgeDB MAE
Supervised	62.02±0.34	22.81±0.07	14.92±0.14	11.52±0.09	14.51±0.13	11.77±0.27
Pseudo Label	62.72±0.11	23.07±0.05	14.90±0.22	11.44±0.53	14.76±0.12	11.71±0.53
Π-Model	63.24±0.63	23.54±0.63	14.80±0.12	11.35±0.12	14.76±0.14	11.92±0.09
MeanTeacher	63.44±0.32	23.25±0.13	15.01±0.64	11.66±0.32	14.99±0.99	12.07±0.48
CRMatch	101.66±0.84	85.45±0.72	22.42±0.23	18.77±0.43	20.42±0.10	17.11±0.49
PseudoLabel+SR	**61.71**±0.34	**22.45**±0.05	**14.80**±0.53	**10.91**±0.12	**14.01**±0.12	**10.77**±0.22
Gain	-0.90	-0.99	-0.10	-0.53	-0.75	-0.94

Table 3: Top-1 error rate (%), performance gain, and training speedup times on ImageNet with 100 labels per class.

Method	Top-1	Gain	Speedup
FixMatch	43.66	+0.00	×1.00
FixMatch+SR	**41.72**	+1.94	×1.98
FlexMatch	41.85	+0.00	×0.00
FreeMatch	40.57	+1.28	×1.50
SoftMatch	40.52	+1.33	×1.46
FlexMatch+SR	**40.36**	+1.49	×2.35

- Ablation of losses and training scheduler / possesses of the rewarder model on CIFAR-100 (400 labels).

Scheduler T	Loss MSE	Loss BCE	Error (%)
0%	✓		19.65
5%	✓		17.89
10%	✓		**16.65**
10%		✓	17.66
15%	✓		16.82

MSE	BCE	Weighted	Accuracy(%)
✓		–	83.35
✓		0.1	80.99
✓		0.5	81.25
✓		0.9	79.85
	✓	–	82.34
	✓	0.1	80.02
	✓	0.5	81.11
	✓	0.9	81.01