CoSSL: Co-Learning of Representation and Classifier for Imbalanced Semi-Supervised Learning

Yue Fan Dengxin Dai Anna Kukleva Bernt Schiele
{yfan, ddai, akukleva, schiele}@mpi-inf.mpg.de
Max Planck Institute for Informatics, Saarbrücken, Germany
Saarland Informatics Campus

In the supplementary material, we first present results for two additional experimental setups in Section A and B. Then, we provide pseudo-code of our co-learning framework CoSSL in Section C. More ablation studies of our co-learning framework can be found in Section D. Finally, we conduct more evaluation at unknown and known shifted test distributions in Section E.

A. Evaluation under $\gamma_l \neq \gamma_u$

The imbalance ratio of labeled data is not always the same as that of unlabeled data in practice. In this section, we compare different methods under $\gamma_l \neq \gamma_u$. Table 1 shows the results on CIFAR-100 with $\gamma_l = 50$ and $\gamma_u = 100$ with two different SSL backbone. In both cases, CoSSL gives superior performance to other methods.

CIFAR-100	$\gamma_l = 50$ $\gamma_l = 100$
ReMixMatch	42.07
w/ DARP	43.19
w/ DARP + cRT	46.59
w/ CReST+	42.31
w/ CReST+ + LA	41.42
w/ CoSSL	**47.33**
FixMatch	40.47
w/ DARP	41.20
w/ DARP + cRT	43.01
w/ CReST+	41.20
w/ CReST+ + LA	44.14
w/ CoSSL	**45.92**

Table 1. Comparison on CIFAR-100 with $\gamma_l = 50$, $\gamma_u = 100$.

B. Evaluation with less number of labeled data

In this section, we provide more evaluation of our method with less number of labeled data than that of the main paper. We compare different methods on CIFAR-100 with $\gamma = 100$. We set the number of labeled data as 50 for the first class. As is shown in Table 2, CoSSL outperforms other methods and achieves the best performance.

CIFAR-100	N_1=50	CIFAR-100	N_1=50
ReMixMatch	27.76	FixMatch	24.00
w/ Re-sample	27.22	w/ Re-sample	25.06
w/ LDAM-DRW	30.22	w/ LDAM-DRW	23.30
w/ DARP	28.29	w/ DARP	25.02
w/ DARP + cRT	30.13	w/ DARP + cRT	24.55
w/ CReST+	28.76	w/ CReST+	25.22
w/ CReST+ + LA	28.32	w/ CReST+ + LA	26.08
w/ CoSSL	**31.31**	w/ CoSSL	**28.42**

Table 2. Efficacy of CoSSL with less labeled data on CIFAR-100 with $\gamma = 100$.

C. CoSSL pseudo-code

We present the complete algorithm of our co-learning framework processing one batch of labeled and unlabeled images in algorithm 1.

D. Ablation study

In this section, we provide more ablation results about different design choices of our method.

Benefits of the co-learning framework. As argued in the main paper, we attribute the success of CoSSL to four aspects: (1) Decoupling representation and classifier while coupling them closely (Table 3). (2) Classifier helps representation via pseudo-labeling rather using gradient directly (Table 5). (3) Using the balanced classifier h_{CL} for pseudo-label generation (Table 4). (4) Using TFE for classifier learning (Table 6).

Sampling the fusion factor from a uniform distribution with lower bound. Here we study the effect of different μ from TFE Algorithm. Since the fusion factor λ is sampled from a uniform distribution between μ and 1, μ controls the

Algorithm 1 Co-learning of representation and classifier

Input: Labeled set $\mathcal{X} = \{(\mathbf{x}_n, y_n) : n \in (1, \ldots, N)\}$, unlabeled set $\mathcal{U} = \{\mathbf{u}_m : m \in (1, \ldots, M)\}$, feature encoder g, classifier head in representation learning h_r, classifier head in classifier learning h_c, control parameter for fusion factor μ, momentum coefficient m, batch size B, total number of training iterations T

$\xi^0 = g$; $g^0 = g$; $h_r^0 = h_r$; $h_c^0 = h_c$

for $t = 0$ **to** $T - 1$ **do**

  // *Sample labeled and unlabeled data for SSL*

  $\{\mathbf{x}_i^t, y_i^t\}_{i=0}^{B-1} \sim$ Random sampler($\mathcal{X}$)

  $\{\mathbf{u}_i^t\}_{i=0}^{B-1} \sim$ Random sampler($\mathcal{U}$)

  // *Pseudo-labeling with EMA encoder and classifier*

  $\hat{y}_i^t$ = Pseudo-label(ξ^t, h_c^t, $\mathbf{u}_i^t$) $\forall i$

  // *Apply TFE*

  $\{\tilde{\mathbf{z}}_i, \tilde{y}_i\}_{i=0}^{B-1}$ = TFE($\mathcal{X}$, $\mathcal{U}$, ξ^t, μ)

  // *EMA update of the encoder*

  $\xi^{t+1} = m\xi^t + (1 - m)g^t$

  // *Compute losses and update the model*

  $\mathcal{L}_x = \frac{1}{B}\sum_{i=1}^{B} \ell_{CE}(y_i^t, h_r^t(g^t(\mathbf{x}_i^t)))$

  $\mathcal{L}_u = \frac{1}{B}\sum_{i=1}^{B} \ell_{CE}(\hat{y}_i^t, h_r^t(g^t(\mathbf{u}_i^t)))$

  $\mathcal{L}_c = \frac{1}{B}\sum_{i=1}^{B} \ell_{CE}(\tilde{y}_i^t, h_c^t(\tilde{\mathbf{z}}_i^t))$

  $\mathcal{L} = \mathcal{L}_c + \mathcal{L}_x + \mathcal{L}_u$

  g^{t+1}, h_r^{t+1}, h_c^{t+1} = Update(g^t, h_r^t, h_c^t)

end for

return ξ^T, h_c^T // *Model for evaluation*

Benefits of decoupling		CIFAR-10			CIFAR-100		
		γ=50	γ=100	γ=150	γ=20	γ=50	γ=100
Fix.	-	81.44	75.31	69.16	48.41	41.76	36.79
	two-stage	82.93	78.51	73.52	49.95	44.11	39.54
	CoSSL	**86.42**	**82.60**	**80.24**	**52.76**	**47.04**	**42.09**
ReMix.	-	82.57	76.94	73.30	50.76	43.51	38.48
	two-stage	86.43	82.27	80.30	54.07	47.25	41.87
	CoSSL	**87.55**	**83.40**	**81.95**	**55.01**	**48.26**	**43.14**

Table 3. Both decoupled approaches (two-stage, CoSSL) show better results over the joint training. Particularly, our co-learning achieves the best performance across settings.

Pseudo-label generation		CIFAR-10			CIFAR-100		
		γ=50	γ=100	γ=150	γ=20	γ=50	γ=100
Fix.	h_{SSL}	85.48	81.20	78.23	52.24	45.90	40.43
	h_{CL}	**86.42**	**82.60**	**80.24**	**52.76**	**47.04**	**42.09**
ReMix.	h_{SSL}	86.90	82.88	80.22	54.39	47.81	42.09
	h_{CL}	**87.55**	**83.40**	**81.95**	**55.01**	**48.26**	**43.14**

Table 4. Benefits of using classifier learning module to generate pseudo-labels. h_{SSL} denotes the classifier from the representation learning module, h_{CL} denotes the classifier from the classifier learning module.

Test Acc.		CIFAR-10			CIFAR-100		
		γ=50	γ=100	γ=150	γ=20	γ=50	γ=100
Fix.	allow grad	84.29	79.21	76.46	50.24	43.72	39.82
	CoSSL	**86.42**	**82.60**	**80.24**	**52.76**	**47.04**	**42.09**
ReMix.	allow grad	78.18	69.99	68.12	54.28	47.06	42.65
	CoSSL	**87.55**	**83.40**	**81.95**	**55.01**	**48.26**	**43.14**

Table 5. Benefits of not updating the encoder from the gradient of the classifier module.

Test Acc.	Enhancement	CIFAR-10			CIFAR-100		
		γ=50	γ=100	γ=150	γ=20	γ=50	γ=100
Fix.	-	84.24	80.27	77.22	51.40	45.39	41.33
	mixUp	85.07	80.42	77.36	52.01	45.85	41.24
	MFW	85.54	81.77	77.91	52.05	46.09	41.61
	TFE	**86.42**	**82.60**	**80.24**	**52.76**	**47.04**	**42.09**
ReMix.	-	87.06	82.24	79.53	54.58	47.84	42.60
	mixUp	86.80	83.10	81.75	**55.01**	47.89	42.27
	MFW	87.37	**83.56**	81.48	54.77	47.96	42.51
	TFE	**87.55**	83.40	**81.95**	**55.01**	**48.26**	**43.14**

Table 6. Test accuracy of using different classifier learning methods in CoSSL.

regularization effect of feature blending. A large μ indicates less regularization as the newly generated feature will be dominated by the labeled feature. In the extreme cases, when $\mu = 1$, TFE reduces to vanilla cRT as the unlabeled portion in the new feature is 0. On the other hand, a small μ implies strong regularization as the new feature can potentially contain a large portion of unlabeled data while still using the same label. As is shown by the blue curve in Fig. 1 left, a μ with proper amount of regularization needs to be selected to maximize the model performance. While $\mu = 0.6$ gives the best result (80.24%), our model is quite robust within a large range of μ. Note that $\mu = 0.6$ is used as the default for all the results across datasets (CIFAR, ImageNet, and Food-101) in the main paper, which also indicates the robustness of our method.

Furthermore, as is compared in Figure 1 left, sampling from the other half of the uniform distribution performs worse for all μ but the full range. Since the newly generated feature shares the class label with its labeled component, therefore, it is more beneficial to set λ closer to 1 by sampling from a uniform distribution between μ and 1. Moreover, the best uniform distribution with $\mu = 0.6$ outperforms commonly used beta distribution as shown in Figure 1 right.

Effect of the number of warm-up epochs. Here we study the effect of the number of warm-up epochs for representation learning. Using a warm-up for representation learning can make the model enjoy both high precision of pseudo-labels in early training, and stronger class-rebalancing in late training. Similar strategies are also widely used in many other works [1, 4, 6]. As shown in Fig. 2, warming up

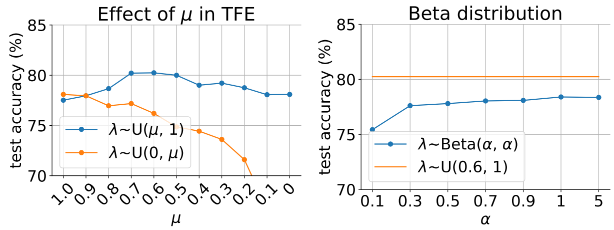

Figure 1. **Left:** It is important to make λ closer to 1 so that the blended feature is closer to the labeled feature, thus safer to share the label. Our method also shows good robustness within a wide range of μ. **Right:** Comparison with beta distribution.

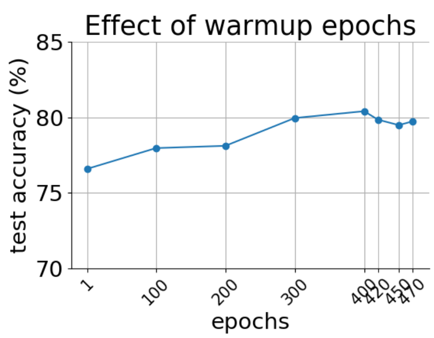

Figure 2. Effect of number of warm-up epochs for our method. Enabling our co-learning at a later epoch is more beneficial.

for longer than 300 epochs gives similar final results, and 400 epochs of warming-up achieves the best test accuracy, which corresponds to 80% of the training time. Our model shows good robustness in terms of the warming-up as we use a warm-up for the first 80% of the training epochs for all experiments in the main paper and achieve good performance.

Effect of the joint training. Here we extend the ablation study of the benefits of our co-learning framework compared with two-stage approaches. Specifically, we compare our CoSSL with three variants of two-stage methods: vanilla cRT [3], cRT with mixUp, and our TFE. For all three two-stage approaches, we first train a complete FixMatch for representation learning. Then, keeping the feature encoder fixed, the classification layer is reinitialized and trained for 20 epochs. Table 7 summarizes the results. In most cases, CoSSL outperforms two-stage methods, which demonstrates the benefits of the joint framework. Moreover, TFE also fits particularly well for imbalanced SSL as TFE is in the top-two performing methods across different settings among two-stage methods.

Class-imbalanced sampler during the classifier training. Here we study the effect of the class-imbalanced sampler in TFE under known shifted test distributions. When the shifted distribution is known prior to the training, we can

Ablation	CIFAR-10			CIFAR-100		
	γ=50	γ=100	γ=150	γ=20	γ=50	γ=100
FixMatch	81.44	75.31	69.16	48.41	41.76	36.79
+ cRT	82.93	78.51	73.52	49.95	44.11	39.54
+ cRT w/ mixUp	86.16	81.94	77.50	51.08	43.74	39.18
+ TFE	**86.83**	81.94	77.93	52.88	45.37	40.79
FixMatch + CoSSL	86.42	**82.60**	**80.24**	**52.76**	**47.04**	**42.09**

Table 7. Classification accuracy (%) of two-stage methods compared to our CoSSL. The better performance demonstrates the effectiveness of our co-learning framework.

leverage this information to improve the performance at important classes by replacing the class-balanced sampler in TFE with a sampler following the target distribution. Specifically, we train our models using class-imbalanced samplers with various imbalance ratios during the classifier training, and test them under three known shifted distributions. We report classwise accuracies on CIFAR-10-LT with an imbalance ratio of 150 and use FixMatch as the base SSL method.

Fig. 3 shows the classwise accuracies at known test distributions with imbalance ratio $\gamma = 32$, 1, and -32. While the class-balanced sampler (r=1) gives reasonable performance across classes, using class-imbalanced sampler during the classifier training can make the model in favor of head or tail classes. For example, when using a sampler with a large negative imbalance ratio -64, performance of tail classes can be improved further. The trend of the head classes is, however, the opposite, which shows a clear trade-off. Therefore, depending on the target distribution, an imbalanced sampler favoring the important classes should be deployed to improve the overall performance.

Table. 8 summarizes the average class accuracy of CoSSL trained with different class-imbalanced sampler under known shifted distributions. Replacing the class-balanced sampler in TFE with a sampler following the distribution of imbalance ratio 2 gives large improvement at positive test imbalance ratios and achieves the best numbers in most cases.

E. More evaluation at unknown and known shifted distributions

Here we extend the evaluation of different methods at shifted test distributions in Section 4.4. We report results at imbalance ratio $\gamma = 100$ on CIFAR-10-LT and $\gamma = 20$, 50 and 100 for CIFAR100-LT. All experiments are run with the same data split and the training protocol from Section 4.1. We take FixMatch as the base SSL method and test post-compensation (PC) [2], classifier retraining (cRT) [3], DARP [4], CReST+ [6], and our CoSSL over a family of shifted distributions. As PC takes in target distribution p_t to modify the logits at test time, we set p_t as the uniform

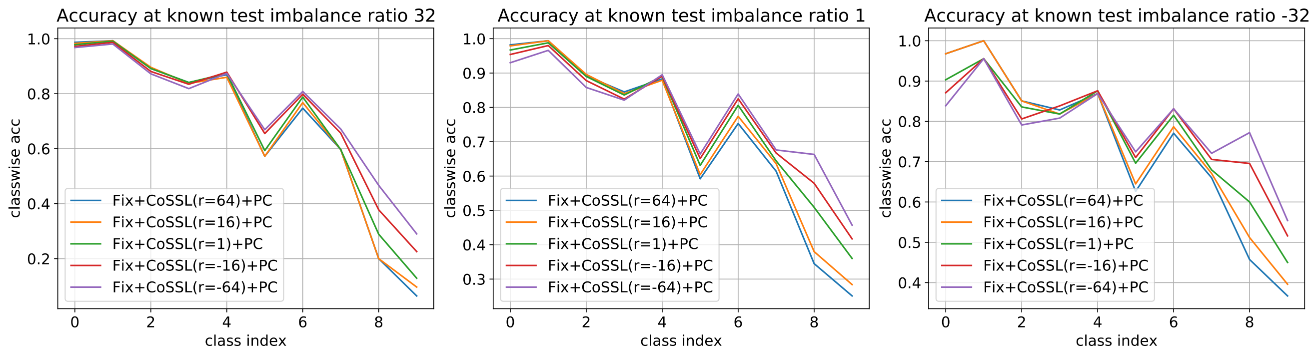

Figure 3. Rather than only looking at average class accuracy, this figure shows classwise accuracies of CoSSL. In particular, we train with different class-imbalanced samplers (r = 64, 16, 1, -16, -64) under three known shifted evaluation settings (left ratio 32, middle 1, right -32). The train data has an imbalance ratio of 150. In the case of test imbalance ratio of 32 (left figure), we can see that the class-imbalanced sampler has little effect on the head classes (classes 0, 1, etc) while having a strong influence on the tail classes (classes 9, 8, etc). This can be explained by the effect that a class-imbalanced sampler with a ratio of e.g. -64 will heavily oversample the tail classes and thus improve their performance overall. When the test imbalance ratio is further away (middle figure ratio 1, right figure ratio -32) from the train data imbalance of 150 we can see a similar trend for the tail classes, however, the trend for the head classes is the opposite. Thus there is a clear trade-off between head and tail classes depending on which class-imbalanced sampler is used for training the classifier. Depending on the application scenario it might be thus interesting to not only look at average accuracies but more closely at this trade-off.

Test imbalance ratio	512	256	**150**	128	64	32	16	8	4	2	1	-2	-4	-8	-16	-32	-64	-128	-256	-512	Mean
	Unknown test-time imbalance ratio																				
Fix	94.83	93.95	93.13	92.87	91.24	89.11	86.62	82.90	78.92	73.58	67.83	61.83	55.41	49.50	44.46	40.37	36.88	33.89	30.95	29.04	66.36
Fix + PC	94.63	93.95	93.30	92.95	91.54	89.89	87.87	84.89	82.05	77.97	73.49	68.86	63.88	59.45	55.70	52.76	50.24	47.90	45.77	44.23	72.57
Fix + vanilla cRT	94.78	93.90	93.17	92.83	91.24	89.24	86.87	83.75	80.29	75.54	70.40	65.10	59.47	54.36	49.86	46.35	43.39	40.81	38.34	36.61	69.31
Fix + DARP	**95.14**	**94.46**	**93.73**	**93.50**	**92.18**	**90.12**	87.70	84.39	81.03	76.26	71.15	66.12	60.99	56.10	52.28	48.84	45.75	43.25	40.79	39.17	70.65
Fix + CReST+	94.18	93.39	92.74	92.45	91.05	89.04	86.70	83.52	80.20	76.05	71.75	67.28	62.76	58.73	55.68	52.89	50.47	48.49	46.61	45.54	71.98
Fix + CoSSL	91.73	91.13	90.90	90.60	89.85	89.07	**87.95**	**86.24**	**84.60**	**82.61**	**80.40**	**78.39**	**76.03**	**74.19**	**73.21**	**72.49**	**71.43**	**70.64**	**70.02**	**69.71**	**81.06**
	Known test-time imbalance ratio																				
Fix + PC	94.98	94.00	93.13	92.83	91.16	89.24	87.03	84.00	81.03	77.31	73.49	70.10	66.79	64.21	62.69	61.89	62.41	63.26	64.80	66.50	77.04
Fix + vanilla cRT	95.14	94.32	93.39	93.25	91.35	89.24	86.73	83.45	79.85	75.04	70.40	65.76	60.65	56.67	53.81	52.04	51.07	51.09	49.98	51.60	72.24
Fix + DARP + PC	95.19	**94.46**	93.73	93.54	92.32	90.32	88.17	85.53	83.00	79.96	76.82	74.33	72.05	70.88	70.37	70.53	70.98	71.39	72.19	73.07	80.94
Fix + CReST+ + PC	94.48	93.44	92.74	92.49	91.09	89.17	87.20	84.75	82.60	79.86	77.74	76.09	74.41	74.03	74.40	75.40	76.38	77.22	78.66	80.29	82.62
Fix + CoSSL + PC	92.83	91.59	90.90	90.31	89.22	87.93	86.42	85.01	84.00	82.57	**82.00**	**81.70**	**81.72**	**81.66**	**82.94**	**84.66**	**85.77**	**86.83**	**87.58**	**88.31**	**86.20**
Fix + CoSSL(r=2) + PC	**95.24**	94.27	**93.95**	**93.67**	**92.51**	**91.07**	89.23	86.89	**85.05**	**82.64**	80.87	79.95	78.89	78.56	79.49	81.32	82.74	84.23	85.59	87.16	86.17
Fix + CoSSL(r=4) + PC	94.88	94.04	93.47	93.16	92.10	90.61	88.81	86.41	84.62	82.65	80.87	79.89	78.96	78.70	79.77	81.39	82.03	83.60	84.94	86.56	85.87
Fix + CoSSL(r=16) + PC	95.04	94.18	93.77	93.29	92.21	90.68	**89.29**	**86.98**	84.91	82.74	80.98	80.00	78.61	78.70	79.80	81.22	82.52	83.77	84.90	86.41	86.00

Table 8. Classification accuracy (%) on CIFAR-10-LT with imbalance ratio $\gamma = 150$. We test different methods on top of FixMatch [5] for known and unknown shifted distributions. Post-compensation (PC) [2] is deployed to utilize the information of the known test distribution.

distribution and the used test distribution for unknown and known distributions, respectively. For cRT, we reinitialize and train the classification layer for 20 epochs while keeping the feature encoder fixed after the representation learning.

Table 9, 10, 11, 12 show the evaluation results. For unknown distributions, while compromising at some positive ratios, CoSSL outperforms other methods by large margins at negative ratios, which leads to the overall higher mean accuracy across different settings. This indicates that our method addresses the imbalance better than other methods that only perform well at distributions closer to the ones used during the training. Similarly, we achieve a more balanced performance across various imbalance ratios for known distributions as well.

References

[1] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. In Advances in Neural Information Processing Systems, 2019. 2

[2] Youngkyu Hong, Seungju Han, Kwanghee Choi, Seokjun Seo, Beomsu Kim, and Buru Chang. Disentangling label distribution for long-tailed visual recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 3, 4, 5

[3] Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis. Decoupling representation and classifier for long-tailed recognition. In International conference on learning representations, 2020. 3

[4] Jaehyung Kim, Youngbum Hur, Sejun Park, Eunho Yang, SungJu Hwang, and Jinwoo Shin. Distribution aligning refinery of pseudo-label for imbalanced semi-supervised learning.

Test imbalance ratio	512	256	128	**100**	64	32	16	8	4	2	1	-2	-4	-8	-16	-32	-64	-128	-256	-512	Mean
											Unknown test-time imbalance ratio										
Fix	95.24	94.69	93.62	93.14	92.36	90.74	88.59	85.79	83.14	79.19	75.23	71.48	67.74	64.14	61.44	59.31	57.36	55.87	54.27	53.11	75.82
Fix + PC	94.83	94.46	**93.71**	93.38	92.70	91.59	**89.84**	**87.69**	**85.92**	83.07	80.43	78.07	75.81	73.75	72.45	71.64	70.95	70.30	69.56	69.36	81.98
Fix + vanilla cRT	95.64	94.92	93.79	93.22	92.40	90.87	88.76	86.02	83.76	79.69	76.27	73.08	69.67	66.90	64.69	63.04	61.74	60.65	59.45	58.93	77.67
Fix + DARP	95.29	94.60	93.58	92.94	92.25	90.74	89.29	86.36	83.96	80.18	76.60	73.28	69.79	66.69	64.39	62.71	61.25	60.07	58.85	58.12	77.55
Fix + CReST+	**95.44**	**95.10**	93.70	**94.25**	**93.00**	**91.63**	89.76	87.32	84.63	80.91	77.42	73.97	70.12	66.78	64.33	62.22	60.28	58.60	57.09	55.87	77.62
Fix + CoSSL	91.68	91.27	90.86	90.56	90.27	89.47	88.59	87.09	85.83	**84.04**	**82.52**	**81.09**	**79.91**	**78.63**	**77.88**	**77.53**	**77.12**	**76.85**	**76.40**	**76.38**	**83.70**
											Known test-time imbalance ratio										
Fix + PC	95.39	94.50	93.54	93.14	92.44	91.00	89.04	86.70	85.05	82.52	80.43	79.15	77.91	77.10	77.49	78.38	79.30	81.00	81.85	83.15	84.95
Fix + vanilla cRT	95.74	95.06	93.83	93.38	92.36	90.97	88.43	85.81	83.52	79.46	76.27	73.53	71.18	69.07	68.28	67.52	68.06	67.99	70.12	74.57	80.26
Fix + DARP + PC	95.29	94.50	93.54	92.94	92.21	90.94	89.59	87.37	85.89	83.43	81.57	80.53	79.58	79.12	79.97	81.35	82.22	83.56	84.80	85.96	86.22
Fix + CReST+ + PC	**95.69**	**95.15**	**93.70**	**94.30**	**92.89**	**91.59**	**89.84**	**87.94**	**86.02**	83.86	82.27	81.07	80.16	79.60	80.19	81.22	82.74	83.60	84.94	86.06	86.64
Fix + CoSSL + PC	92.83	91.87	91.07	90.56	90.08	88.98	87.48	85.99	84.82	**83.97**	**83.57**	**83.43**	**83.91**	**84.30**	**85.23**	**86.59**	**88.39**	**89.35**	**89.38**	**89.87**	**87.58**

Table 9. Classification accuracy (%) on CIFAR-10-LT with imbalance ratio $\gamma = 100$. We test different methods on top of FixMatch [5] for known and unknown test-time distributions. Post-compensation (PC) [2] is deployed to utilize the information of the known test distribution.

Test imbalance ratio	64	32	**20**	16	8	4	2	1	-2	-4	-8	-16	-32	-64	Mean
								Unknown test-time imbalance ratio							
Fix	69.07	67.18	65.25	64.56	61.02	57.15	53.01	48.30	43.89	39.53	35.27	31.23	28.12	25.60	49.23
Fix + PC	67.41	65.92	64.43	63.87	60.92	57.90	54.44	50.46	46.66	42.98	39.30	35.74	33.07	30.72	50.99
Fix + vanilla cRT	66.64	65.23	63.66	63.15	60.39	57.92	54.27	50.41	46.94	43.35	39.80	36.58	34.15	32.12	51.04
Fix + DARP	**69.41**	**67.91**	**65.96**	**65.30**	**62.22**	58.25	54.59	50.02	45.65	41.23	37.50	33.41	30.61	28.20	50.73
Fix + CReST+	65.96	64.66	63.41	63.18	60.70	58.29	55.16	51.85	48.45	45.49	42.36	39.68	37.62	35.49	52.31
Fix + CoSSL	65.87	65.27	63.95	63.51	61.40	**59.07**	**56.21**	**53.11**	**50.08**	**47.06**	**44.31**	**41.30**	**39.60**	**37.84**	**53.47**
								Known test-time imbalance ratio							
Fix + PC	69.28	67.08	65.25	64.62	60.97	57.81	54.21	50.46	47.14	44.30	41.60	39.68	39.49	39.42	52.95
Fix + vanilla cRT	67.92	66.14	64.39	63.54	60.44	57.84	53.78	50.41	47.29	44.39	42.15	40.37	39.03	38.61	52.59
Fix + DARP + PC	**69.84**	**68.01**	**65.96**	**65.42**	**62.38**	**58.74**	**55.96**	52.36	49.76	46.72	44.95	42.58	41.88	41.55	54.72
Fix + CReST+ + PC	66.51	64.80	63.41	63.24	60.82	58.55	55.61	**53.06**	50.91	49.18	47.83	46.76	46.97	46.67	55.31
Fix + CoSSL + PC	67.11	65.70	63.95	63.30	60.58	58.01	55.00	52.64	**51.00**	**49.25**	**48.00**	**46.79**	**47.08**	**47.31**	**55.41**

Table 10. Classification accuracy (%) on CIFAR-100-LT with imbalance ratio $\gamma = 20$.

In Advances in neural information processing systems, 2020. 2, 3

[5] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In Advances in Neural Information Processing Systems, 2020. 4, 5

[6] Chen Wei, Kihyuk Sohn, Clayton Mellina, Alan Yuille, and Fan Yang. Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 2, 3

Test imbalance ratio	64	**50**	32	16	8	4	2	1	-2	-4	-8	-16	-32	-64	Mean
								Unknown test-time imbalance ratio							
Fix	66.30	65.69	64.08	60.91	56.63	51.97	47.15	41.83	36.28	31.32	26.20	21.95	19.13	16.25	43.26
Fix + PC	65.61	65.16	63.83	60.97	57.35	53.41	49.09	44.30	39.51	34.87	30.38	26.46	23.86	21.46	45.45
Fix + vanilla cRT	64.80	64.28	63.00	60.08	56.75	53.06	48.79	44.52	39.63	34.84	30.59	26.84	24.40	22.10	45.26
Fix + DARP	**66.51**	**65.97**	**64.37**	**61.12**	57.16	52.40	47.32	42.14	36.65	31.33	26.54	22.31	19.21	16.51	43.54
Fix + CReST+	65.19	64.56	63.29	60.70	57.20	53.35	48.94	44.66	39.84	35.45	31.36	27.83	25.09	22.35	45.70
Fix + CoSSL	63.91	63.72	62.78	60.29	**57.54**	**54.30**	**50.99**	**47.12**	**42.69**	**38.19**	**34.16**	**30.64**	**27.98**	**25.81**	**47.15**
								Known test-time imbalance ratio							
Fix + PC	66.25	65.69	64.15	60.94	57.01	52.91	48.76	44.30	40.15	36.08	32.92	29.98	28.66	27.90	46.84
Fix + vanilla cRT	66.17	65.81	63.97	61.06	57.11	53.06	48.76	44.52	40.41	36.35	32.99	30.99	29.82	29.61	47.19
Fix + DARP + PC	**66.55**	**65.97**	**64.55**	**61.27**	**57.52**	53.19	48.77	44.76	40.71	36.72	33.66	30.90	29.57	28.50	47.33
Fix + CReST+ + PC	65.19	64.56	63.10	60.62	57.37	**53.86**	49.85	46.62	42.69	39.75	37.45	35.29	34.01	33.06	48.82
Fix + CoSSL + PC	64.29	63.72	62.35	59.48	56.32	53.46	**50.39**	**47.29**	**44.54**	**41.65**	**39.51**	**37.56**	**37.69**	**38.18**	**49.74**

Table 11. Classification accuracy (%) on CIFAR-100-LT with imbalance ratio $\gamma = 50$.

Test imbalance ratio	**100**	64	32	16	8	4	2	1	-2	-4	-8	-16	-32	-64	Mean
								Unknown test-time imbalance ratio							
Fix	**67.25**	**65.49**	**62.42**	58.61	53.70	48.42	42.87	37.09	31.17	25.78	20.67	16.45	13.18	10.24	39.52
Fix + PC	66.26	64.72	62.06	58.52	54.30	49.55	44.59	39.22	33.66	28.52	23.47	19.35	16.14	13.31	40.98
Fix + vanilla cRT	65.18	63.78	61.26	58.05	54.04	49.57	44.61	39.73	33.97	28.78	23.93	20.24	17.08	14.29	41.04
Fix + DARP	66.21	64.63	61.91	58.20	53.61	48.42	43.09	37.44	31.50	26.34	21.67	17.41	13.97	11.18	39.68
Fix + CReST+	65.65	64.38	61.88	**58.94**	54.62	49.93	44.81	39.60	34.31	29.58	24.38	20.01	16.68	13.48	41.30
Fix + CoSSL	64.15	62.93	61.23	58.47	**54.95**	**51.17**	**46.58**	**42.22**	**36.97**	**32.17**	**27.69**	**24.25**	**20.83**	**17.83**	**42.96**
								Known test-time imbalance ratio							
Fix + PC	**67.25**	**65.44**	**62.42**	58.76	54.06	49.29	44.38	39.22	33.97	29.53	25.13	22.16	19.89	18.05	42.11
Fix + vanilla cRT	66.73	64.76	62.02	58.14	53.92	49.31	44.32	39.73	34.37	29.68	25.34	22.51	20.32	19.03	42.16
Fix + DARP + PC	66.21	64.72	62.06	58.35	53.87	49.48	44.52	40.00	35.08	30.91	27.14	23.89	21.91	20.01	42.73
Fix + CReST+ + PC	65.65	64.29	61.81	**59.00**	**55.26**	**50.71**	**45.95**	41.56	37.09	33.25	29.68	27.08	25.49	24.32	44.37
Fix + CoSSL + PC	64.15	62.37	60.14	57.45	53.54	49.89	45.52	**42.27**	**38.25**	**34.99**	**32.63**	**30.99**	**29.89**	**28.37**	**45.03**

Table 12. Classification accuracy (%) on CIFAR-100-LT with imbalance ratio $\gamma = 100$.