Supplementary Material for HyperMatch: Noise-Tolerant Semi-Supervised Learning via Relaxed Contrastive Constraint

A. Comprehensive Experimental Results

A.1. Hyperparameters

We use almost identical hyperparameters of HyperMatch on CIFAR10, CIFAR100, STL10 and Semi-iNat and a complete list of hyperparameters is provided in Tab. 8. For hyperparameters related to consistency regularization, we follow the settings in [3, 4]. For those related to our proposed relaxed contrastive loss (τ_g, K, t), we strictly keep the same values. It is shown that relaxed contrastive loss is robust to different experiment scenarios.

	CIFAR10	CIFAR100	STL10	Semi-iNat
Arch	WRN28-2	WRN28-8	ResNet-18	ResNet-50
Optimizer	SGD			
Weak Aug	RandomCrop, RandomHorizontalFlip			
Strong Aug	RandAugment [1]			
B	64			
μ	7			
λ_1	1			
Weight Decay	1e-3	1e-3	5e-4	1e-3
λ_2	1	1	5	2
τ_u	0.95	0.95	0.95	0.6
τ_g	0.6			
K	2			
t	0.07			

Table 8. Experiment hyperparameters on CIFAR10, CIFAR100, STL10 and Semi-iNat datasets.

A.2. Fluctuation of Results over Different Runs

The random seeds chosen for averaged results are fixed for different settings. We show the highest and lowest accuracies achieved through repetitive experiments in Tab. 9 to better compare with other methods.

The training fluctuations do exist in our experiments with different random seeds and reporting the cherry-picked result of a single run would be biased. The averaged results give a much more convincing conclusion.

A.3. Performance Degradation

Confirmation bias in a well-known problem in SSL and the fast convergence brought by contrastive loss also accelerates the whole process. Acc_{test} and Acc_{label} refer to the test and pseudo label accuracy. We show the best accuracy and last epoch accuracy in Tab. 10 to verify the effectiveness of HyperMatch. Since FixMatch takes around 1000 epochs to finally converge and achieve the best performance, FixMatch is not included for comparisons. Evidently, the accuracy drop of HyperMatch is much smaller than CCSSL, proving its advantage when resisting the confirmation bias.

	CIFAR100		
	400	2500	10000
FixMatch [3]	51.15 ± 1.75	71.71 ± 0.11	77.40 ± 0.12
CCSSSL [4]	61.19 ± 1.65	75.7 ± 0.63	80.68 ± 0.16
HyperMatch	63.01 ± 0.57	76.45 ± 0.35	81.09 ± 0.28
HyperMatch (highest)	63.66	76.95	81.43
HyperMatch (lowest)	62.38	76.12	80.83

Table 9. Highest and lowest accuracies achieved in repetitive experiments on CIFAR100.

	Acc_{test}			Acc_{pseudo}		
	Best	Last	Drop	Best	Last	Drop
CCSSL [4]	60.74	43.22	17.52	68.07	53.12	14.95
HyperMatch	62.77	52.85	9.92	70.38	60.88	9.5

Table 10. The performance degradation of HyperMatch and CCSSL on CIFAR100@400.

A.4. Deployment to other SSL architectures

We further combined CoMatch with our proposed relaxed contrastive loss and conduct experiments on Semi-iNat in Tab. 11. The improved performance indicates that relaxed contrastive loss can regarded as a complementary technique and applied to other frameworks.

A.5. Selection of K

To better explore the selection of K, we add experiments on CIFAR10 in Tab. 12. Setting K=2 also improves FixMatch by a large margin (12.06%) on the more challenging Semi-iNat dataset. Hence fixing K to 2 is a good empirical value among all investigated tasks. To better decide K in unknown scenarios, a heuristic way is to only train a few epochs, and settings with faster convergence usually tend to behave better. In future, we're considering a more soft re-

Method	Semi-iNat
CoMatch	20.94
CoMatch + HyperMatch	**24.67**

Table 11. Deploy relaxed contrastive loss to CoMatch.

Top-K	1	2	3	5
CIFAR10@250	94.87	**95.18**	94.98	95.12
CIFAR10@4000	95.45	**95.96**	95.73	95.61
CIFAR100@400	60.51	**62.84**	61.24	60.67
CIFAR100@2500	75.87	76.62	**76.73**	75.86

Table 12. Results of different K on CIAR10 and CIFAR100.

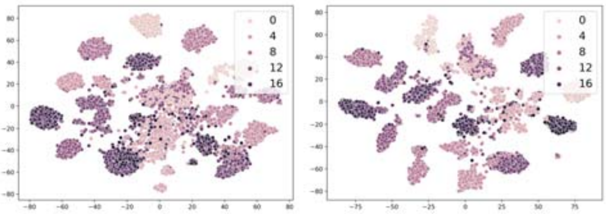
a) FixMatch b) HyperMatch

Figure 5. t-SNE visualization of intermediate features in FixMatch and HyperMatch on CIFAR100.

weight strategy by computing weights for *ALL* classes (instead of top-K) in Eq. 10, thus could skip the selection of K. As for fine-grained datasets, although larger K can mitigate the risk of the true class not being included in top-K, it also introduces more confusing classes in the hyper-class, thus may interfere with the feature learning. In Tab. 12, though CIFAR100 is more fine-grained than CIFAR10, they both achieve almost the best results when K=2.

B. Additional Qualitative Analysis

B.1. Feature Representations

The relaxed contrastive loss helps learn well-clustered feature representations and we visualize the intermediate features before the final classification layer in Fig. 5 to show the difference. For clarity, we randomly choose 20 classes and visualize the high-dimensional representations through t-SNE. HyperMatch shows a more separable and well-clustered feature representations compared with FixMatch. The improved feature quality obviously contributes to the outperformance of HyperMatch.

B.2. Top-K Accuracy

We show the top-1 and top-2 accuracy of clean and noisy labels in the CIFAR100@400 experiment. As the figure shows in Fig. 6, noisy labels benefit more all over the training process than clean labels and the accuracy gain is around 20%. This helps find the ground-truth class for unlabeled instances and relieve the inaccurate pseudo labels.

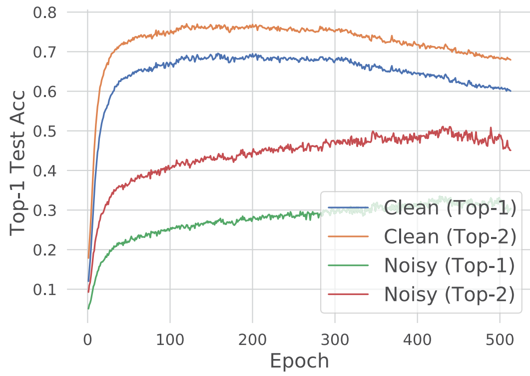

Figure 6. Top-1 and top-2 accuracy of clean and noisy labels in the training process.

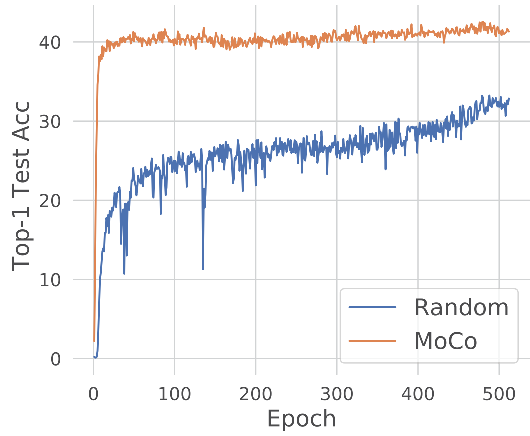

Figure 7. The test accuracy of HyperMatch trained with different pretrained weights on Semi-iNat.

B.3. Pretrained Weights

In Semi-iNat experiments, we tried different pretrained weights (training from randomly inited weights and MoCo [2] pretrained weights) and the training accuracy is plotted in Fig. 7. It is shown that the self-supervised pretrained weights significantly boost the SSL performance after only several iterations. This inspires us that the performance can be further improved by exploiting more unlabeled instances (even out-of-distribution samples) with contrastive loss in a unified SSL framework.

References

[1] Ekin Dogus Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. Randaugment: Practical automated data augmentation with a reduced search space. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3008–3017, 2020. 1

[2] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross B. Girshick. Momentum contrast for unsupervised visual representation learning. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9726–9735, 2020. 2

[3] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin Dogus Cubuk, Alexey Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *ArXiv*, abs/2001.07685, 2020. 1

[4] Fan Yang, Kaixing Wu, Shuyi Zhang, Guannan Jiang, Yong Liu, Feng Zheng, Wei Zhang, Chengjie Wang, and Long Zeng. Class-aware contrastive semi-supervised learning. *ArXiv*, abs/2203.02261, 2022. 1