‘Class-Aware Contrastive Semi-Supervised Learning’ Supplementary for FAQs

Best Pseudo Accuracy	CIFAR100			CIFAR10		
	400	2500	10000	40	250	4000
FixMatch	62.45	83.82	92.01	94.81	96.09	**97.79**
+CCSSL	**69.79**	**85.46**	**92.08**	**96.01**	**96.51**	97.59

Table 7. Best pseudo-label accuracy for various data setttings.

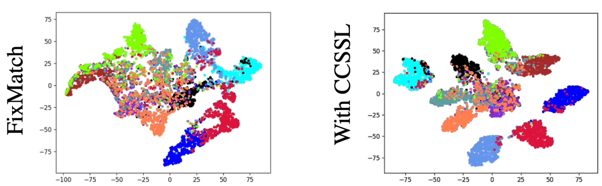

Figure 4. TSNE of high dimensional features with/without CC-SSL.

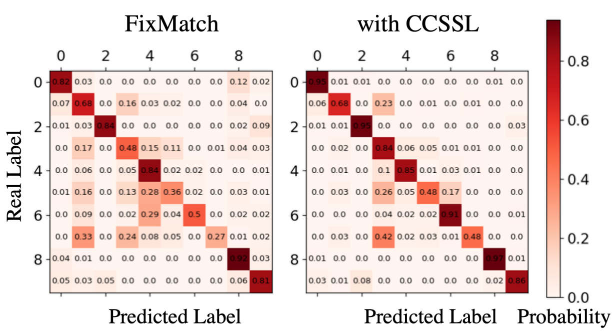

Figure 5. Confusion Matrix with/without CCSSL. STL10 is used for both because CIFAR10 is easy and CIFAR100 has too many categories to be visualized.

A. What is the intuition of using contrastive in SSL? Are there some proofs other than the final performance?

The intuition of CCSSL is to end-to-end reduce noise on the feature level by contrastive learning while maintaining model’s clustering ability.

(1) Noise reduction by a projection head. The noise reduction is on the feature level and not directly on pseudo labels. **Qualitatively**, [35] verified that the uniformity of embeddings from the projection head is essential to learn general separable features. As in Fig. 4, the features are more separable with CCSSL. **Quantitatively**, the model’s confusion on similar categories is halved with CCSSL, as in Fig. 5. With CCSSL, the model can be more resilient to misclassification and thus has less noise on pseudo-labels. So if the model starts with a good pretrained model with good features, like MoCo, CCSSL can only benefit the training process and less helpful.

(2) Quantitative analysis for pseudo-labels. As in Tab. 7, the model’s pseudo label quality is better with CCSSL on many data settings. The more noise (less labels), the more salient of improvements. We cannot evaluate pseudo-label accuracy on Semi-iNat 2021 and STL10 because labels are not provided. We will provide more evaluation results on our public github page https://github.com/TencentYoutuResearch/Classification-SemiCLS.

B. Details of method formulation

(1) **Relationship with infoNCE.** Our method improves upon the **infoNCE**, so we provide S, W_{con} and $\mathcal{L}_{\text{infoNCE}}$ as references and has not directly used them in the final equation for simplicity.

(2) **Cross Entropy of $\mathcal{L}_c$.** The **cross-entropy** between S and W_{target} for $\mathcal{L}_c$ in Eq. (9) means that we take columns as classes and rows as samples to calculate the soft cross-entropy between two matrices. Soft means the targets have been re-weighted.

(3) **Meaning of $\mathbf{P_i}$. $\mathbf{z_p}$** is the high dimensional feature of sample p from $\mathbf{P_i}$, which represents the sample set with the same pseudo-label of sample i.

C. Details of training

The training strategy is generally the same as FixMatch for a fair comparison.

(1) **Warm-up Peroid**. There is no warm-up period, and CCSSL can help a SSL model from a noisy starting. Because the confidence is low in the beginning, the contrastive part of CCSSL is triggered and helps the model learn a general representation instead of over-fitting on the training data. After the model learns a few epochs and can differen-

tiate in-distribution and out-of-distribution data, the clustering part of CCSSL will then be triggered with less noise.

(2) **Augmentations**. The augmentations of supervised and semi-supervised branches are the same as FixMatch, and the strong augmentation of CCSSL branch is the same as MoCo for a fair comparison.

(3) **Backbone**. The model size is the same as FixMatch during inference and the backbone follows the convention with wid-resnet 28-2, 28-8 for CIFAR10 and CIFAR100, resnet18 for STL10, resnet50 for Semi-iNat.

References

[35] Feng Wang and Huaping Liu. Understanding the behaviour of contrastive loss, 2021. 1