Does Unconstrained Unlabeled Data Help Semi-Supervised Learning?

Shuvendu Roy, Ali Etemad
Dept. ECE and Ingenuity Labs Research Institute
Queen's University, Kingston, Canada
{shuvendu.roy, ali.etemad}@queensu.ca

Abstract

In this work, we study the role of unconstrained unlabeled data in semi-supervised learning and propose a semi-supervised learning framework which can learn effective representations from such unlabeled data. Most existing semi-supervised methods rely on the assumption that labelled and unlabeled samples are drawn from the same distribution, which limits the potential for improvement through the use of free-living unlabeled data. Consequently, the generalizability and scalability of semi-supervised learning are often hindered by this assumption. Our method aims to overcome these constraints and effectively utilize unconstrained unlabeled data in semi-supervised learning. UnMixMatch consists of three main components: a supervised learner with hard augmentations that provides strong regularization, a contrastive consistency regularizer to learn underlying representations from the unlabeled data and a self-supervised loss to enhance the representations that are learnt from the unlabeled data. We perform extensive experiments on 4 commonly used datasets and demonstrate superior performance over existing semi-supervised methods with a performance boost of 4.79%. Extensive ablation and sensitivity studies show the effectiveness of each of the proposed components of our method.

1 Introduction

Semi-supervised learning (SSL) [1, 2, 3, 4] uses large amounts of *unlabeled* data along with small amounts of *labelled* data to reduce the reliance on fully-labelled datasets. Most existing semi-supervised methods can be broadly divided into two main categories: pseudo-labelling [1, 2] and consistency regularization [3, 5, 6]. Existing SSL methods rely on the assumption that the labelled and unlabeled data belong to the same distributions, an assumption that is not necessarily true in real-world scenarios. Moreover, this assumption prohibits us from leveraging free-living unlabeled data. In fact, it has been shown in previous studies that incorporating out-of-distribution data with the unlabeled set for SSL impairs performance [7, 8].

To adopt a less constrained approach regarding unlabeled data, *open set* SSL has been proposed [9, 10, 11, 12], which allows the unlabeled set to contain samples from classes which are not necessarily present in the labelled set. Most open set SSL methods [13] learned to distinguish known classes from unknown ones, effectively avoiding samples from unknown classes in the learning process. Nevertheless, this setting still places certain restrictions on the unlabeled data, necessitating the inclusion of samples from **every** known class and ensuring that its data distribution is similar. These constraints create two important challenges. First, collecting an unlabeled dataset that necessarily includes samples from certain classes can be challenging in real-world settings. Second, this approach severely restricts the scalability of SSL to large, web-scale, unconstrained unlabeled data since such data do not hold the aforementioned constraints. Most existing semi-supervised methods are not suitable for learning from unconstrained data as they rely on pseudo-label predictions, which require the 'unlabeled' set to have the same classes as the 'labelled' set.

NeurIPS 2023 Workshop: Self-Supervised Learning - Theory and Practice.

In this paper, we propose a novel SSL approach called UnMixMatch, which can learn effective representations from unconstrained unlabeled data and effectively enable SSL to scale up using web-scale unlabeled data. UnMixMatch comprises three main components, which have some similarities to previous SSL methods but have been specifically modified and tailored toward the 'scalability' criteria: (1) A supervised learner with hard augmentations: We introduce a new hard augmentation module that combines RandAug with MixUp to prevent overfitting on the small labelled set. The convention of using soft augmentations of the existing literature does not perform well when learning from unconstrained unlabeled data. (2) A contrastive consistency regularizer: The primary unsupervised learning component of our method involves a contrastive regularizer, which learns the underlying data representations by enforcing the model to produce consistent predictions under strong perturbations. In contrast to existing SSL methods, we do not regularize the class predictions, as the unlabeled set contains unknown classes. Instead, we enforce consistency in the predicted embedding space. (3) A self-supervised pre-text learning module: To further enhance the learned representations, we include a pre-text task called rotation prediction on the unlabeled data, where the model learns by predicting the degree of rotation applied to each sample.

We conduct extensive research on four common datasets: CIFAR-10, CIFAR-100, SVHN, and STL-10. First, we re-implement and benchmark 13 recent semi-supervised methods with unconstrained unlabeled data, using ImageNet-100 as the unlabeled set. We find that existing methods experience performance degradation in unconstrained settings. In comparison, UnMixMatch outperformed existing methods by an average of 4.79%. Additionally, UnMixMatch exhibits robust scaling capabilities regarding the size of the unlabeled datasets, as we observe an additional 5.61% improvement when we increase the unlabeled data size by a factor of 10. Furthermore, we achieved state-of-the-art results in the open set SSL. Finally, we ablate each component of UnMixMatch and demonstrate the crucial role that each component plays in the performance.

In summary, we make the following contributions: (**1**) We propose a novel semi-supervised method that can learn effective representations from unconstrained unlabeled data for the first time. (**2**) We conduct an extensive study to benchmark the performance of existing semi-supervised methods when the unlabeled data are not constrained to match the distribution of the labelled data. (**3**) We demonstrate that our method outperforms previous methods by a large margin in unconstrained learning and sets a new state-of-the-art for open set SSL. We also show that the performance of our method scales up by increasing the amount of unconstrained unlabeled data.

2 Method

Supervised Module. SSL requires a labelled set X_L to be learned using a supervised component. The first contribution of our method is, therefore, to create a supervised learner suitable for our purpose of scalable SSL. Here, we hypothesize that given large amounts of unlabeled data in an unconstrained setting and relatively very small amounts of labelled data, the supervised module may overfit the small labelled set. Thus, unlike FixMatch [4], MixMatch [14], and ReMixMatch [15], which use weak augmentations for their supervised modules, we apply hard augmentations on the labelled samples in our supervised module. This acts as a regularizer for supervised learning and is better able to deal with overfitting compared to weaker augmentations. We utilize RandAug [16] as the hard augmentation followed by the MixUp operation [14, 15]. We denote RandAug plus MixUp as the RandMixUp operation. Finally, a supervised loss is applied to a batch of samples.

RandAug is a hard augmentation technique for generating diverse samples by employing a sequence of transformations [16]. More specifically, it applies $R_n \in 1..13$ transformations randomly chosen from a list of 13 augmentations, including rotation, translation, and colour distortion. The magnitude of each transformation is sampled randomly from a pre-defined range. We denote the augmentation operation as $\hat{x} = RandAug(x)$. On the other hand, MixUp operation interpolates between two data points (x_1, y_1), (x_2, y_2) to generate mixed samples and labels as: $\bar{x} = \lambda \cdot x_1 + (1 - \lambda) \cdot x_2$; $\bar{y} = \lambda \cdot y1 + (1 - \lambda) \cdot y2$, where λ is the mixing coefficient. Following MixMatch [14], we sample λ from a beta distribution ($\lambda \sim Beta(\alpha, \alpha)$) with hyper-parameter α.

For a batch of unlabeled samples $X_u = ((\hat{x}_i); i \in (1, .., b))$, with batch size b, we first generate the pseudo-label for each sample X_i as $p_i = P_\theta(x_i)$, where P_θ is the encoder with a classification head. Next, for a batch of labelled samples $X_l = ((\hat{x}_i, y_i); i \in (1, ..., b))$ and the unlabeled samples with pseudo-labels $X_p = ((\hat{x}_i, p_i); i \in (1, .., b))$, we augment all the samples using $\bar{X} = RandAugMix(X_l, X_p)$. Accordingly, we define the supervised loss of our method as: $\mathcal{L}_{sup} = \frac{1}{b} \sum_{\bar{x}, \bar{y} \in \bar{X}} \mathcal{H}(\bar{y}, P_\theta(y|\bar{x}))$.

Table 1: Comparison of our method against other SSL methods with *unconstrained unlabeled data* on 4 datasets.

	CIFAR-10			CIFAR-100			SVHN			STL-10	
Methods	40 labels	250 labels	4000 labels	400 labels	2500 labels	10000 labels	40 labels	250 labels	1000 labels	1000 labels	**Avg.**
Supervised	24.24±1.1	43.33±2.2	83.76±0.4	10.39±0.3	39.57±0.4	63.4±0.1	24.67±2.1	75.42±2.4	87.74±0.6	60.96±1.3	51.35
Pi-Model [5]	24.16±1.5	48.14±1.1	84.19±0.3	12.81±0.9	40.01±0.3	63.55±0.1	28.02±1.1	76.56±2.2	87.83±0.6	69.73±0.6	53.5
Mean Teacher [3]	26.22±1.0	49.35±1.7	83.7±0.4	13.72±0.9	41.57±0.3	63.83±0.1	28.57±1.3	76.78±2.3	87.68±0.6	70.12±0.1	54.15
VAT [6]	24.7±1.3	46.18±1.3	84.73±0.2	11.5±0.8	41.73±0.2	63.76±0.1	41.95±2.5	76.18±1.9	88.07±0.5	63.12±0.6	54.19
Pseudo-label [1]	24.88±1.6	50.29±1.3	84.11±0.1	12.12±0.1	39.72±0.8	63.57±0.0	36.04±2.7	78.04±1.2	88.91±0.3	65.6±0.9	54.33
UDA [22]	28.12±2.5	65.59±1.7	88.31±0.1	21.11±0.8	51.82±0.6	69.42±0.6	48.8±4.1	77.73±1.9	88.83±0.3	82.54±0.2	62.23
MixMatch [14]	32.58±0.8	58.24±0.3	84.59±0.3	20.26±0.6	45.94±0.4	65.89±0.2	57.46±1.7	77.65±1.2	89.95±0.2	71.85±0.6	60.44
ReMixMatch [15]	35.56±0.8	64.71±0.6	87.64±0.3	18.9±0.8	49.11±0.9	69.38±0.1	56.94±4.2	79.57±0.9	90.57±0.3	79.13±0.8	63.15
FixMatch [4]	27.91±3.9	64.98±1.0	88.18±0.1	21.2±0.6	51.4±1.3	67.8±0.3	43.71±5.5	74.52±1.0	88.06±0.6	81.85±0.4	60.96
FlexMatch [23]	32.8±0.8	63.22±1.0	87.82±0.1	20.87±0.7	51.28±0.8	69.52±0.5	60.5±2.5	79.72±0.6	88.85±0.4	82.67±0.4	63.72
CoMatch [24]	41.68±0.7	62.31±0.7	84.52±0.3	22.6±0.6	44.0±1.0	58.55±0.3	45.87±2.8	73.19±0.3	86.45±0.2	82.0±0.0	60.12
CCSSL [12]	30.89±5.9	67.2±1.5	88.77±0.1	24.53±1.5	**56.3±0.2**	71.13±0.3	50.02±6.6	80.39±0.6	88.6±0.3	82.0±0.0	63.98
SimMatch [25]	23.77±1.8	57.72±1.3	84.12±0.7	18.65±1.2	47.33±1.0	66.54±0.8	51.23±1.6	74.48±1.1	88.57±1.0	77.23±1.2	58.96
ScMatch [26]	27.81±1.1	56.78±0.6	83.09±0.2	18.14±1.1	46.21±0.7	64.24±0.2	56.59±0.9	75.08±0.8	89.23±0.3	79.44±0.7	59.66
UnMixMatch	**47.93±1.1**	**68.72±0.6**	**89.58±0.2**	**26.13±1.1**	54.18±0.7	**71.73±0.2**	**72.9±0.9**	**80.78±0.8**	**91.03±0.3**	**84.73±0.7**	**68.77**

Consistency Regularization Module. To deal with the unconstrained nature of existing unlabeled data and learn effective representations, we apply a consistency regularizer as our second contribution. A consistency regularizer learns from the unlabeled data by enforcing consistency on its predictions under different augmentations. Prior works that have used consistency regularization for SSL [5, 14, 15] enforce consistency on the **class predictions** under different perturbations. However, regularization over class predictions is not useful for learning in *unconstrained* settings where unlabeled data do not necessarily come from the same classes as the labelled data. To address this, we enforce consistency in the low-dimensional embedding space using a contrastive loss. Using contrastive loss on the embedding space enables the model to learn class-agnostic representations from unconstrained unlabeled data. In UnMixMatch, we adopt the Noise Contrastive Estimation loss, a.k.a InfoNCE [17].

Contrastive learning learns from positive (perturbations of the same sample) and negative samples (all other samples) by bringing embeddings of the positive samples together and pushing them away for the negatives. For each sample, $x_i \in X_U$, two augmentations are applied to generate two augmented images $\hat{x_i} = RandAug(x_i)$. These are first passed through the encoder and a projection head (shallow linear layers with non-linearity and batch normalization) to obtain embeddings $z_i = P_{\theta_p}(z|\hat{x_i})$. The contrastive loss is accordingly defined as: $\mathcal{L}_{con} = -\frac{1}{2b}\sum_{i=1}^{2b} log \frac{exp(z_i, z_{\kappa(i)}/\tau)}{\sum_{k=1}^{2b} \mathbf{1}_{[k \neq i]} exp(z_i, z_k/\tau)}$, where, $\kappa(i)$ is the index of the second augmented sample, $\mathbf{1}_{[k \neq i]}$ is an indicator function which returns 1 when k is not equal to i, and 0 otherwise. τ is a temperature parameter.

Self-supervised Module. Finally, we intend to enhance the quality of the representations extracted from the unconstrained unlabeled data using the consistency regularizer.It has been recently shown that self-supervised techniques can be employed to learn underlying domain-invariant representations for unlabeled data [18, 19]. Moreover, this idea has already been demonstrated to be useful in conjunction with SSL [20, 21, 15]. As a result, we integrate a straightforward yet effective self-supervised pre-text task called rotation prediction, which learns by predicting the degree of rotations applied to unlabeled images. In practice, a rotation module randomly samples one of the following rotations and applies it to an unlabeled image: $0°, 90°, 180°, 270°$. As a result, the rotation prediction task can be viewed as a four-way classification problem, represented as: $\mathcal{L}_{rot} = \frac{1}{b}\sum_{u \in U} \mathcal{H}(r, P_{\theta_r}(r|Rotate(u))$. Here, P_{θ_r} is the encoder with a rotation head that predicts the rotation, and $\mathcal{H}$ is the cross-entropy loss.

Total Loss. Finally, we incorporate the loss functions for the three modules above to create the total loss: $\mathcal{L}_{UnMixMatch} = \mathcal{L}_{sup} + \beta\mathcal{L}_{con} + \gamma\mathcal{L}_{rot}$. Here, β and γ are hyper-parameters that balance the significance of the contrastive and rotation losses.

3 Experiments and Results

For our main experiments, we follow the standard semi-supervised evaluation protocol from prior works [4], and present the results on four datasets: CIFAR-10 [27], CIFAR-100 [27], SVHN [28], and STL-10 [29]. We use ImageNet-1K [30] as unconstrained unlabeled data.

Performance in Unconstrained Settings. Table 1 presents the main results of our work on the four datasets. Here, we first re-implement 13 semi-supervised methods and report the results with *unconstrained* unlabeled data. We report the average accuracies and standard deviations across three individual runs for each setting. We also report the average accuracy across all settings for overall comparison. It should be noted that the performance of prior methods is considerably lower than what has been reported in the original papers, where the unlabeled and labelled samples came from the same datasets (unlabeled data were not unconstrained). Next, we observe that UnMixMatch demonstrates superior performance compared to other methods, with an average improvement of 4.79%. We obtain considerable improvement across all datasets and splits, except using CIFAR-100 with 2500 labels, where CCSSL achieves a better result. When considering the number of labelled samples, we notice that the differences between UnMixMatch and other methods are more pronounced when the labelled set size is small. For example, with only 40 labelled samples from CIFAR-10, UnMixMatch achieves a 17.04% performance gain over CCSSL (which has the second highest overall average performance after ours), and 6.25% higher than the next best result for this specific setting, which was obtained by CoMatch. A similar pattern is observed for SVHN, where UnMixMatch outperforms CCSSL and FlexMatch by 22.88% and 12.4%, respectively.

Scaling Up the Unlabeled Set. Our main motivation for using unconstrained unlabeled data is to take advantage of the abundance of free-living unlabeled data. In this experiment, we evaluate the performance of UnMixMatch as the size of the unlabeled set is increased. The results of this study are presented in detail in Table 2. Specifically, we increase the size of the unlabeled set from 130K images of ImageNet-100 (a subset of ImageNet-1K) to 1.28M images of ImageNet-1K, with two more subsets of 450K and 850K images from ImageNet-1K. We perform this experiment on CIFAR-10 with 40 labelled samples with the two best methods (CoMatch and ReMixMatch) on this setting and observe an increasing trend in the accuracy of UnMixMatch as the number of images in the unlabeled set increases. With ImageNet-1K used as the unlabeled set, which is approximately 10 times larger than ImageNet-100, the accuracy of UnMixMatch improves from 47.93% to 53.54%, a significant improvement of 5.61% by simply increasing the size of the unlabeled set. CoMatch and ReMixMatch, on the other hand, show very small improvements with the increase in the unlabeled data, but the performance difference with our method further increases.

Table 2: Impact of unlabeled set size.

Data	**IN-100**	**Subset 1**	**Subset 2**	**IN-1K**
No. of samples	130K	450K	850K	1.28M
ReMixMatch	35.56	35.72	36.15	36.24
CoMatch	41.68	42.52	42.31	43.38
UnMixMatch	**47.93**	**50.01**	**51.99**	**53.54**

Results on Open Set Settings. Next, we investigate the performance of UnMixMatch on open set SSL. Open set SSL is a relatively less challenging setting than *unconstrained* settings, where the unlabeled set may contain images of unknown classes but must contain images of all known classes [9, 10]. For learning in open set settings, we employ a variant of UnMixMatch which takes advantage of the fact that the unlabeled set contains samples of all known classes from the labelled set and learns from the predicted pseudo-labels on the unlabeled set. In this variant, we replace the contrastive loss in our method with the class-aware contrastive loss of CCSSL [12]. This method first predicts the pseudo-labels for the unlabeled samples and uses them with contrastive loss to learn clusters of known classes in the embedding space. For this experiment, we follow the experimental setup of OpenMatch [9], which reports the results for CIFAR-10 with a 6/4 split. This split means that the labelled set contains images of 6 classes from CIFAR-10, while the unlabeled set includes images of 6 known and 4 unknown classes. Like OpenMatch, we take 6 animal classes as the known classes and 4 object classes as the unknown classes. We perform this experiment using three different splits with 50, 100, and 400 labelled samples per class in the known set. The results of this experiment are presented in Table 3, where it can be observed that our method outperforms the existing methods and sets a new state-of-the-art for *open set SSL*. Once again, UnMixMatch better demonstrates its effectiveness when the amount of labelled data is limited. With 50 labelled samples per class, our approach provides a 6.1% improvement over the second-best method, OpenMatch. For 100 and 400 labelled samples per class, UnMixMatch shows 3.9% and 3.1% improvements, respectively.

Table 3: Performance on open set SSL for CIFAR-10 with 6/4 split.

	Labelled samples/class		
Methods	**50**	**100**	**400**
Supervised	64.3±1.1	69.5±0.7	80.0±0.3
FixMatch [4]	56.8±1.2	70.2±0.6	83.7±0.5
MTC [10]	79.7±0.9	86.3±0.9	91.0±0.5
OpenMatch [9]	89.6±0.9	92.9±0.5	94.1±0.5
UnMixMatch	**95.7±0.8**	**96.8±0.5**	**97.2±0.4**

4 Conclusion

Existing semi-supervised methods struggle to learn when the assumption that the unlabeled data comes from the same distribution as the labelled data, is violated. This work proposes a new semi-supervised method for learning from unconstrained unlabeled data. Our method shows large improvements over existing methods and even larger improvements under low-labelled data settings. Our approach also outperforms existing methods on *open set* settings. Most importantly, UnMixMatch scales up in performance when the size of unlabeled data increases. We hope this research will draw attention to this more challenging and realistic SSL setting with unconstrained unlabeled data.

References

[1] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on Challenges in Representation Learning, ICML*, volume 3, page 896, 2013.

[2] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V Le. Self-training with noisy student improves imagenet classification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10687–10698, 2020.

[3] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in Neural Information Processing Systems*, 30, 2017.

[4] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in Neural Information Processing Systems*, 33:596–608, 2020.

[5] Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. *Advances in Neural Information Processing Systems*, 29, 2016.

[6] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8):1979–1993, 2018.

[7] Avital Oliver, Augustus Odena, Colin A Raffel, Ekin Dogus Cubuk, and Ian Goodfellow. Realistic evaluation of deep semi-supervised learning algorithms. *Advances in Neural Information Processing Systems*, 31, 2018.

[8] Jong-Chyi Su, Zezhou Cheng, and Subhransu Maji. A realistic evaluation of semi-supervised learning for fine-grained classification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12966–12975, 2021.

[9] Fan Yang, Kai Wu, Shuyi Zhang, Guannan Jiang, Yong Liu, Feng Zheng, Wei Zhang, Chengjie Wang, and Long Zeng. Class-aware contrastive semi-supervised learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14421–14430, 2022.

[10] Qing Yu, Daiki Ikami, Go Irie, and Kiyoharu Aizawa. Multi-task curriculum framework for open-set semi-supervised learning. In *European Conference on Computer Vision*, pages 438–454, 2020.

[11] Lan-Zhe Guo, Zhen-Yu Zhang, Yuan Jiang, Yu-Feng Li, and Zhi-Hua Zhou. Safe deep semi-supervised learning for unseen-class unlabeled data. In *International Conference on Machine Learning*, pages 3897–3906, 2020.

[12] Fan Yang, Kai Wu, Shuyi Zhang, Guannan Jiang, Yong Liu, Feng Zheng, Wei Zhang, Chengjie Wang, and Long Zeng. Class-aware contrastive semi-supervised learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14421–14430, 2022.

[13] Ryota Yoshihashi, Wen Shao, Rei Kawakami, Shaodi You, Makoto Iida, and Takeshi Naemura. Classification-reconstruction learning for open-set recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4016–4025, 2019.

[14] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. *Advances in Neural Information Processing Systems*, 32, 2019.

[15] David Berthelot, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Kihyuk Sohn, Han Zhang, and Colin Raffel. Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. *arXiv preprint arXiv:1911.09785*, 2019.

[16] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 702–703, 2020.

[17] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*, pages 1597–1607, 2020.

[18] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*, 2018.

[19] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *European Conference on Computer Vision*, pages 649–666, 2016.

[20] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*, 2018.

[21] Xiaohua Zhai, Avital Oliver, Alexander Kolesnikov, and Lucas Beyer. S4l: Self-supervised semi-supervised learning. In *IEEE/CVF International Conference on Computer Vision*, pages 1476–1485, 2019.

[22] Qizhe Xie, Zihang Dai, Eduard Hovy, Thang Luong, and Quoc Le. Unsupervised data augmentation for consistency training. *Advances in Neural Information Processing Systems*, 33:6256–6268, 2020.

[23] Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in Neural Information Processing Systems*, 34:18408–18419, 2021.

[24] Junnan Li, Caiming Xiong, and Steven CH Hoi. Comatch: Semi-supervised learning with contrastive graph regularization. In *IEEE/CVF International Conference on Computer Vision*, pages 9475–9484, 2021.

[25] Mingkai Zheng, Shan You, Lang Huang, Fei Wang, Chen Qian, and Chang Xu. Simmatch: Semi-supervised learning with similarity matching. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14471–14481, 2022.

[26] Guan Gui, Zhen Zhao, Lei Qi, Luping Zhou, Lei Wang, and Yinghuan Shi. Improving barely supervised learning by discriminating unlabeled samples with super-class. In *Advances in Neural Information Processing Systems*, 2022.

[27] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[28] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. 2011.

[29] Adam Coates, Andrew Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 215–223. JMLR Workshop and Conference Proceedings, 2011.

[30] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE conference on computer vision and pattern recognition*, pages 248–255, 2009.