All Labels Are Not Created Equal: Enhancing Semi-supervision via Label Grouping and Co-training

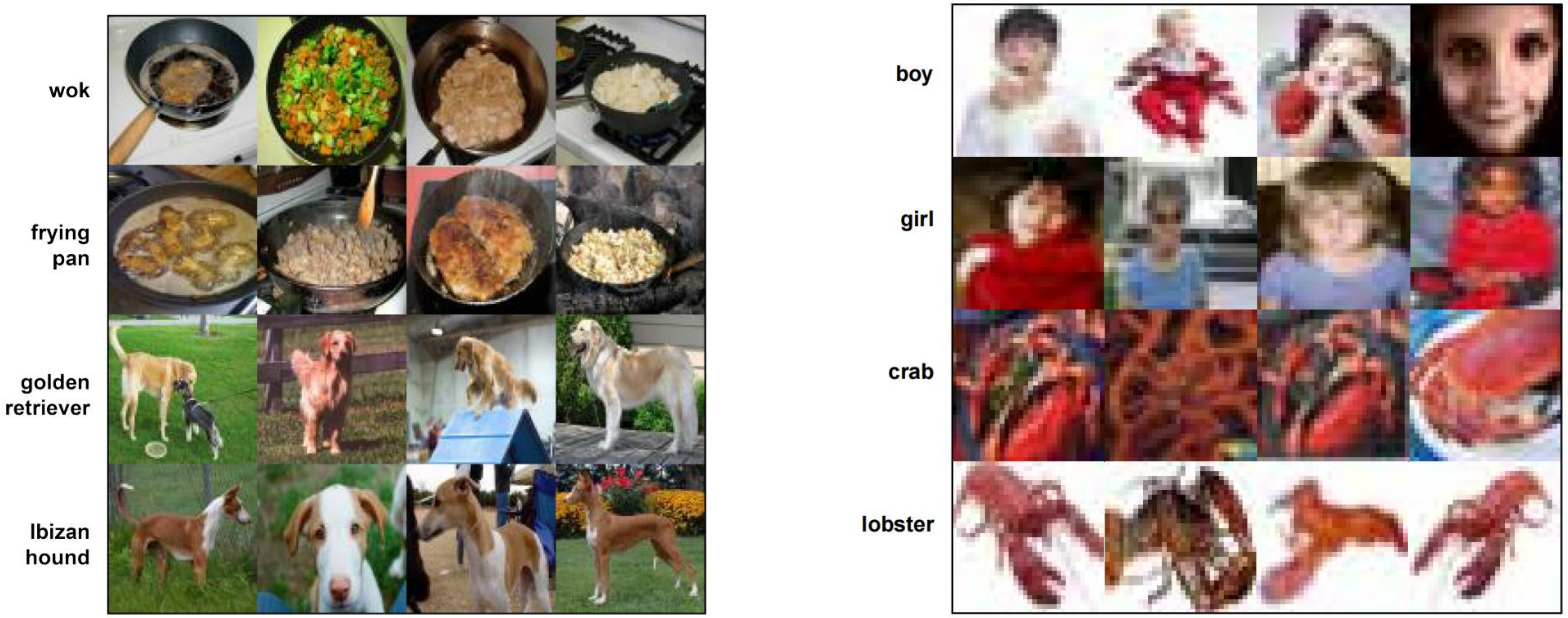
Islam Nassar[1,*], Samitha Herath[1], Ehsan Abbasnejad[2], Wray Buntine[1], Gholamreza Haffari[1]
[1] Dept of Data Science and AI, Faculty of IT, Monash University, Australia
[2] Australian Institute for Machine Learning, University of Adelaide, Australia

CVPR 2021

https://arxiv.org/abs/2104.05248

Problem Statement

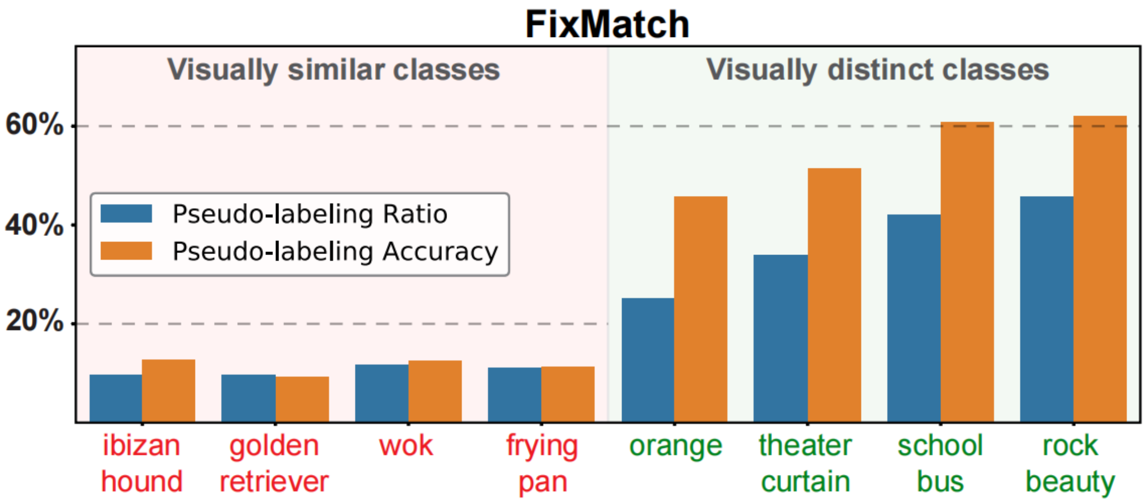
- Visually similar classes often produce low-confidence predictions
- FixMatch (Confidence Filtering) -> Class Imbalanced



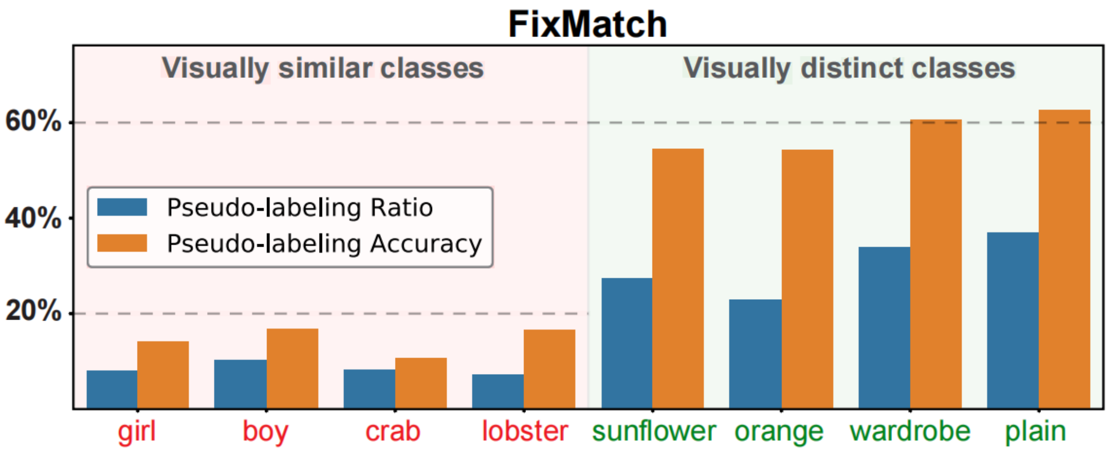
(a) Mini-ImageNet (b) CIFAR-100

Figure 7: The most confused images for the 4 most visually similar classes of Mini-ImageNet (left) and CIFAR-100 (right). The caption next to each image group denotes the true class to which the image group belongs.

Problem Statement



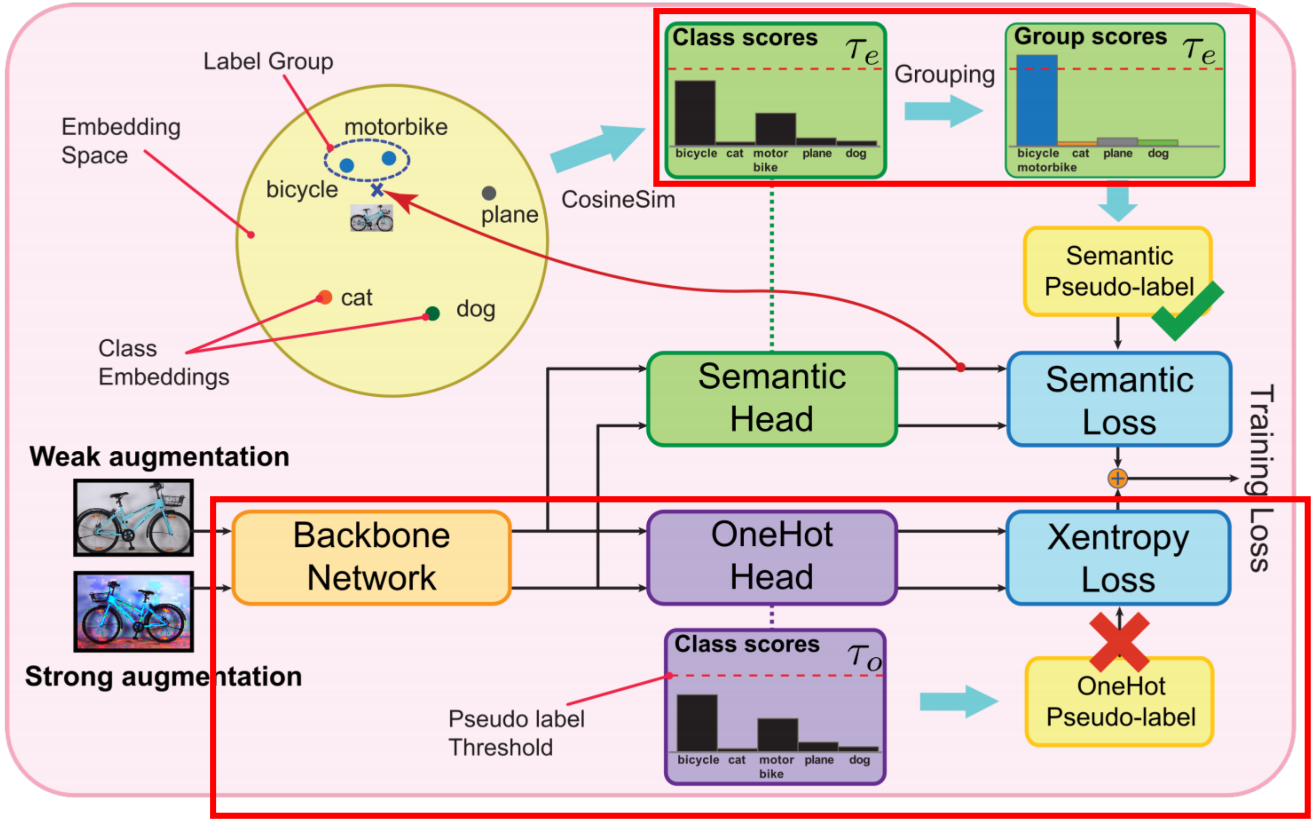
Mini-ImageNet



CIFAR100

- Visually similar classes often produce low-confidence predictions
- This leads to class imbalance among the pseudo-labeled instances which potentially misguides SSL training.

Method



FixMatch

Figure 1: A conceptual diagram of our co-training solution

Method

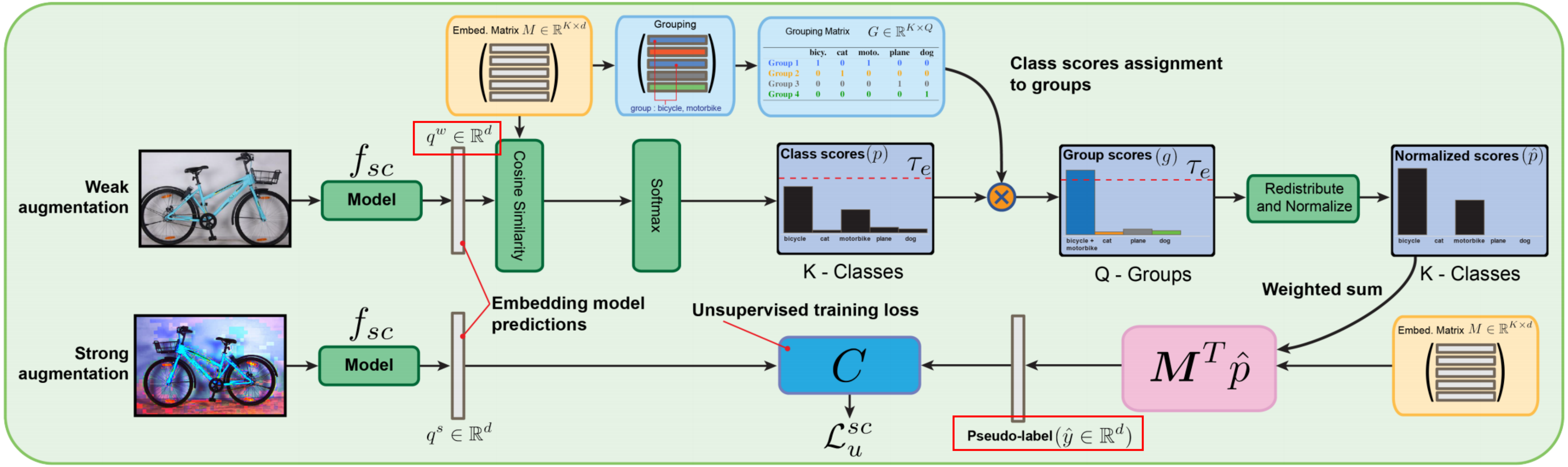


Figure 3: Unsupervised loss for the *Semantic Classifier* - A weakly augmented image is used (upper path) to obtain a predicted embedding, which is then used to obtain class scores. The class scores are summed for each label group (as identified by our grouping method) to obtain group scores. If one of the group scores exceeds the threshold, it is retained for pseduo-labeling. The pseudo-label is then calculated as an average of the group members embeddings weighted by their class scores. The loss is then enforced against the predicted embedding for a strongly augmented image (lower path).

Where M From

- **Knowledge graph embeddings**
- **Using Class Attributes Annotations**

Loss

- **Semantic Classifier**

$$\mathcal{L}_s^{sc} = \frac{1}{n} \sum_{i=1}^{n} C(M^T \boldsymbol{y}_i, f_{sc}(\boldsymbol{x}_i)),$$

with $C(\boldsymbol{z}, \boldsymbol{z}') = 1 - \text{CosineSim}(\boldsymbol{z}, \boldsymbol{z}')$.

$$\mathcal{L}_u^{sc} = \frac{1}{\mu \cdot n} \sum_{j=1}^{\mu \cdot n} C(\boxed{\hat{\boldsymbol{y}}_j}, f_{sc}(\mathcal{A}_s(\boldsymbol{u}_j))) \cdot \eta_j^{sc}.$$

$$\eta_j^{sc} = \mathbb{1}(\max(\boldsymbol{g}_j) \geq \tau_e).$$

- **One-Hot Classifier**

$$\mathcal{L}_s^{oh} = \frac{1}{n} \sum_{i=1}^{n} H(\boldsymbol{y}_i, f_{oh}((\boldsymbol{x}_i))).$$

$$\mathcal{L}_u^{oh} = \frac{1}{\mu \cdot n} \sum_{j=1}^{\mu \cdot n} H(\hat{\boldsymbol{y}}_j, f_{oh}(\mathcal{A}_s(\boldsymbol{u}_j))) \cdot \eta_j^{oh},$$

Loss

- **Co-training Loss**

$$\mathcal{L}_{co} = \frac{1}{\mu \cdot n} \sum_{j=1}^{\mu \cdot n} C(M^T \hat{\boldsymbol{y}}_j, f_{sc}(\mathcal{A}_s(\boldsymbol{u}_j))) \boxed{\cdot\, \eta_j^{oh}} + H(\arg\max(\boldsymbol{p}_j), f_{oh}(\mathcal{A}_s(\boldsymbol{u}_j))) \boxed{\cdot\, \eta_j^{sc}}$$

- **Total Loss**

$$\mathcal{L}_{total} = \mathcal{L}_s^{sc} + \mathcal{L}_s^{oh} + \lambda_u(\mathcal{L}_u^{sc} + \mathcal{L}_u^{oh}) + \lambda_{co}\mathcal{L}_{co}.$$

Results

Table 1: Error rates for CIFAR-10, CIFAR-100 and Mini-ImageNet. We report results for two different values of μ - i.e. ratio between unlabeled and labeled data in a mini-batch, for our method and FixMatch. † denotes that the results reported are using the same codebase. $*$ denotes that the result is based on using CNN-13 model. We report the mean and standard deviation across 3 different splits of labeled data for each experiment.

	CIFAR-10		**CIFAR-100**			**Mini-ImageNet**		
Total Labelled Samples	250	4000	2500	4000	10000	1000	4000	10000
Pseudo-labeling [19]	49.78±0.43	16.09±0.28	-	-	-	-	-	-
Mean teacher [34]	32.32±2.30	9.19±0.19	-	-	-	-	72.51±0.22	57.55±1.11
UDA [42]	8.82±1.08	4.88±0.18	33.13±0.22	-	24.50±0.25	-	-	-
Label Propagation [14]	-	12.69±0.29*	-	-	-	-	70.29±0.81	57.58±1.47
PLCB [1]	24.81±5.35	6.28±0.30	-	37.55±1.09*	32.15±0.50*	-	56.49±0.51	46.08±0.11
MixMatch† [3]	11.29±0.75	6.24±0.07	39.70±0.27	-	28.59±0.31	60.97±0.31	49.79±0.11	44.27±0.23
FixMatch†($\mu = 3$) [31]	5.78±0.23	4.52±0.01	38.45±0.51	32.22±0.21	28.42±0.09	66.23±1.13	59.73±5.45	44.66±0.12
FixMatch† ($\mu = 7$)	**4.55±0.12**	4.49±0.05	33.64±0.07	31.27±1.30	26.13±0.18	60.97±0.31	49.79±0.11	44.27±0.23
Ours (SemCo)† ($\mu = 3$)	5.87±0.31	4.43±0.01	33.80±0.57	29.40±0.18	25.07±0.04	**55.35±0.71**	**46.01±0.93**	**41.25±0.76**
Ours (SemCo)† ($\mu = 7$)	5.12±0.27	**3.80±0.08**	**31.93±0.01**	**28.61±0.23**	**24.45±0.12**	59.35±0.23	49.46±2.20	42.78±0.35

Results

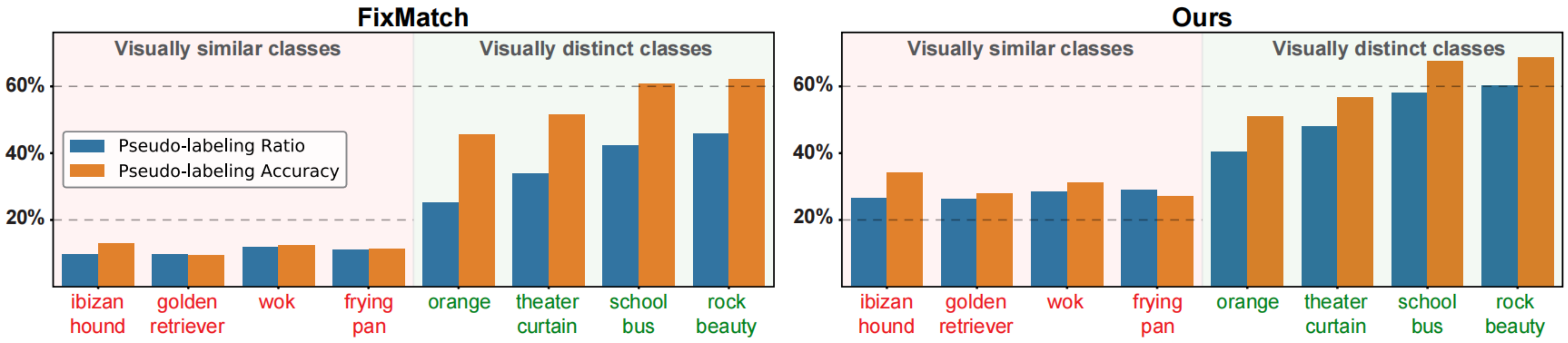


Figure 2: Confidence-based pseudo-labeling comparison between the baseline (left) and our method (right). *Accuracy* values show how much, on average, pseudo-labels for a given class match the true label, while *Ratio* values show the percentage of samples of a given class which are retained for pseudo-labeling (i.e. with confidence score above the threshold). The two metrics are calculated for the 4 most (red) and least (green) visually similar classes over the first 10 epochs of training.

Results

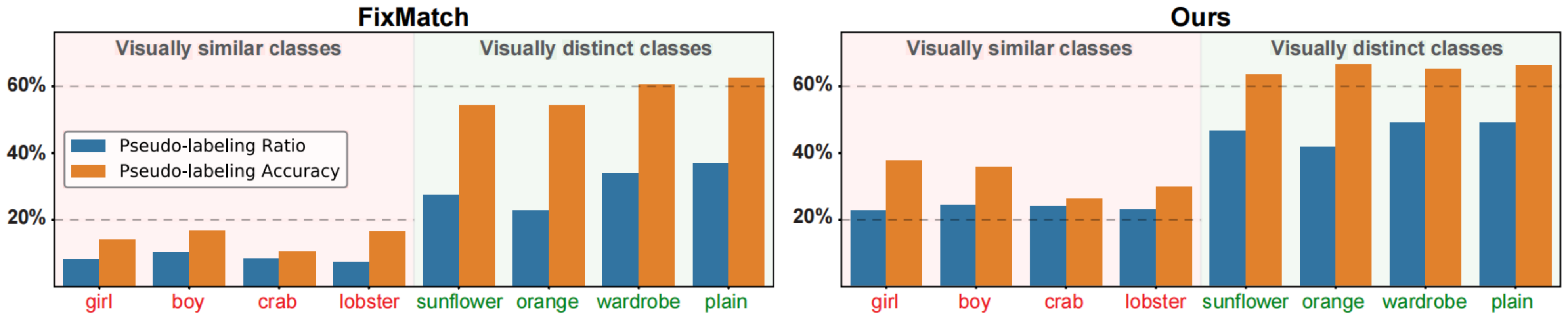


Figure 6: CIFAR-100 confidence-based pseudo-labeling comparison between the baseline (left) and our method (right). *Accuracy* values show how much, on average, pseudo-labels for a given class match the true label, while *Ratio* values show the percentage of samples of a given class which are retained for pseudo-labeling (i.e. with confidence score above the threshold). The two metrics are calculated for the 4 most (red) and least (green) visually similar classes over the first 10 epochs of training.

Results

Table 2: Error rates on CUB-200 dataset and DomainNet Real. Errors are reported based on 1 split for each of the amounts of labeled data. Poor baseline results are omitted.

CUB-200	Total Labeled Samples	
Method	1000	2000
Supervised baseline	-	70.11
FixMatch	84.35	72.15
Ours (SemCo)	**79.44**	**66.76**

DomainNet Real	Total Labeled Samples	
Method	6900	10350
Supervised baseline	47.9	45.2
FixMatch	41.34	39.04
Ours (SemCo)	**35.32**	**32.89**

Results

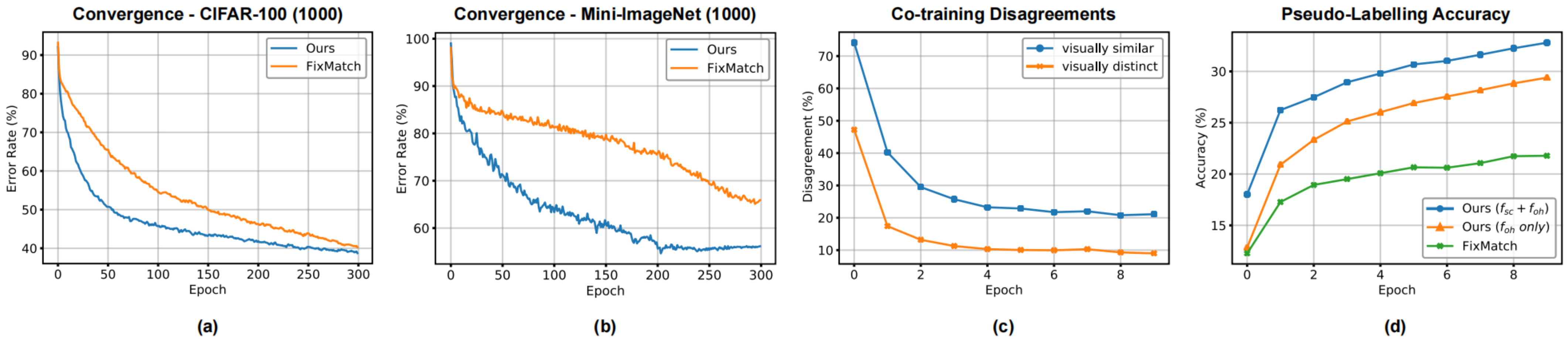


Figure 4: Experimental analysis plots showing: (a,b): Convergence trends of our method and the baseline for CIFAR-100 (a) and Mini-ImageNet (b) with 1000 labeled examples. (c,d): Co-training analysis plots showing the disagreements between our two classifiers for visually similar and distinct classes (c) and the associated pseudo-labeling accuracies (d). The co-training plots are spanning only the first 10 epochs of training.

Results

Table 3: Error Rates for different settings of Co-training and Label Grouping

		Mini-ImageNet 1000	**CIFAR-100 2500**
Label Grouping	Co-training	Error Rate	
✓	✓	55.35	31.93
-	✓	59.60	33.09
✓	-	60.39	33.19
-	-	62.16	34.25

Table 4: Error Rates when using Embedding Targets versus One-Hot Targets for our *Semantic Classifier*, reported on CIFAR-100 and Mini-ImageNet

	Embeddings Target	One-Hot Target
CIFAR-100 (2500)	31.93	33.33
Mini-ImageNet (1000)	55.35	60.33