英語論文#1

2023/05/08

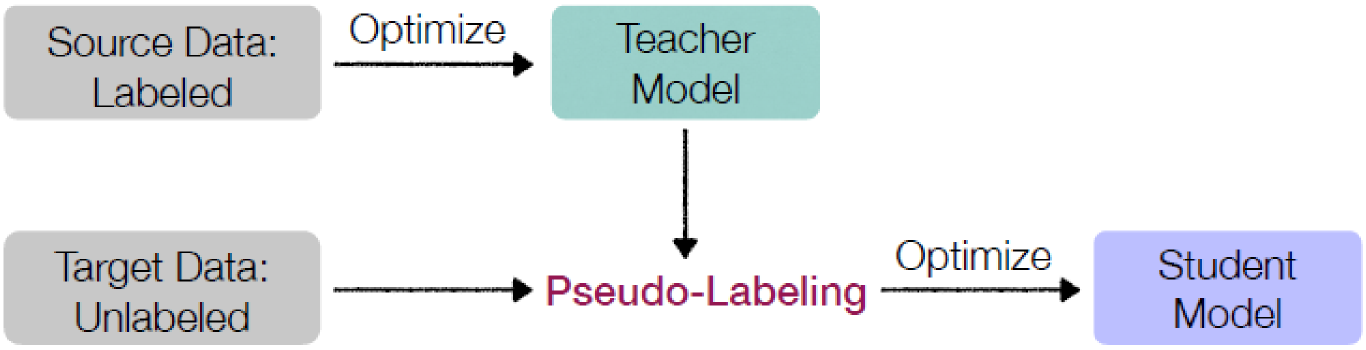
M1
建元　了

論文の概要

- タイトル
  - 「**Debiased Learning from Naturally Imbalanced Pseudo-Labels**」
- 執筆者
  - Xudong Wang, Zhirong Wu, Long Lian, Stella X, Yu
- 掲載
  - CVPR 2022
- 選択理由
  - SSL(・ZSL)において予測のバイアスを改善する研究であるため

背景

- **疑似ラベル**は、データセットで事前学習したモデルがラベル付かない画像に対して予測値をラベルとして扱い、データセットと組み合わせる半教師あり学習の手法として広く利用

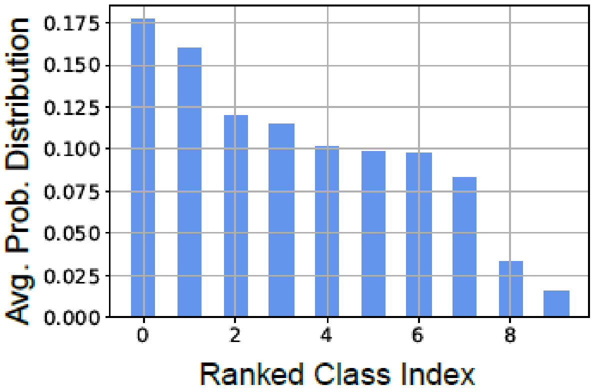


(a) Framework of pseudo-labeling

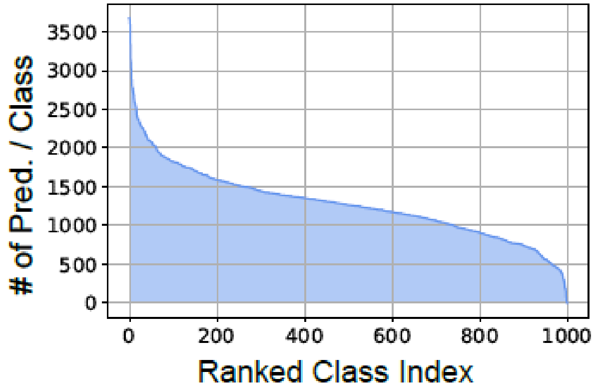
背景

- 既存の厳選されクラスのバランスのとれたデータセットで事前学習したモデルであっても**疑似ラベルの付与に偏りが生じる**

- 不均衡なラベルによってモデルの精度にも影響でてしまうため、バイアスを除去する手法が必要となっている



(b) FixMatch on CIFAR10 SSL



(c) CLIP on ImageNet ZSL

導入

- 擬似ラベルは、バランスのとれたデータセットで学習しバランスのとれた疑似ラベルで評価した場合でも、内在するデータの類似性により自然にバイアスがかかる
- 疑似ラベルに起因する不均衡な分類問題において、動的にバイアスを除去する脱バイアス学習を提案
- ImageNet-1Kにおいてアノテーション0.2%のSSＬでは26%、ZSLでは9%

Fixmatch(Semi-Supervised Learnig)

- consistency regularization(一致性正規化)とPseudo-Labeling(疑似ラベル)を組み合わせた手法

- Labeled: dataset $X_L = \{(x_i, y_i)\}_{i=1}^{L}$
- Unlabeled: dataset $X_U = \{(x_i)\}_{i=L+1}^{L+U}$
  - 入力画像 x_i
  - C個のクラス $y_i = [{y_i}^1, \dots, {y_i}^C] \in \{0,1\}^C$

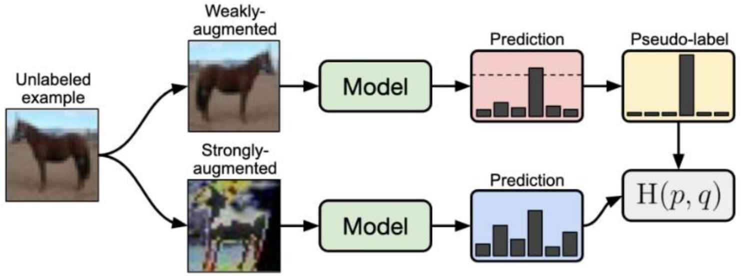


Figure 1: Diagram of FixMatch. A weakly-augmented image (top) is fed into the model to obtain predictions (red box). When the model assigns a probability to any class which is above a threshold (dotted line), the prediction is converted to a one-hot pseudo-label. Then, we compute the model's prediction for a strong augmentation of the same image (bottom). The model is trained to make its prediction on the strongly-augmented version match the pseudo-label via a cross-entropy loss.

Fixmatch(SSL)

- 最適化

$$\mathcal{L} = \mathcal{L}_s + \lambda_U \mathcal{L}_U$$

  - モデル予測と正解との交差エントロピー

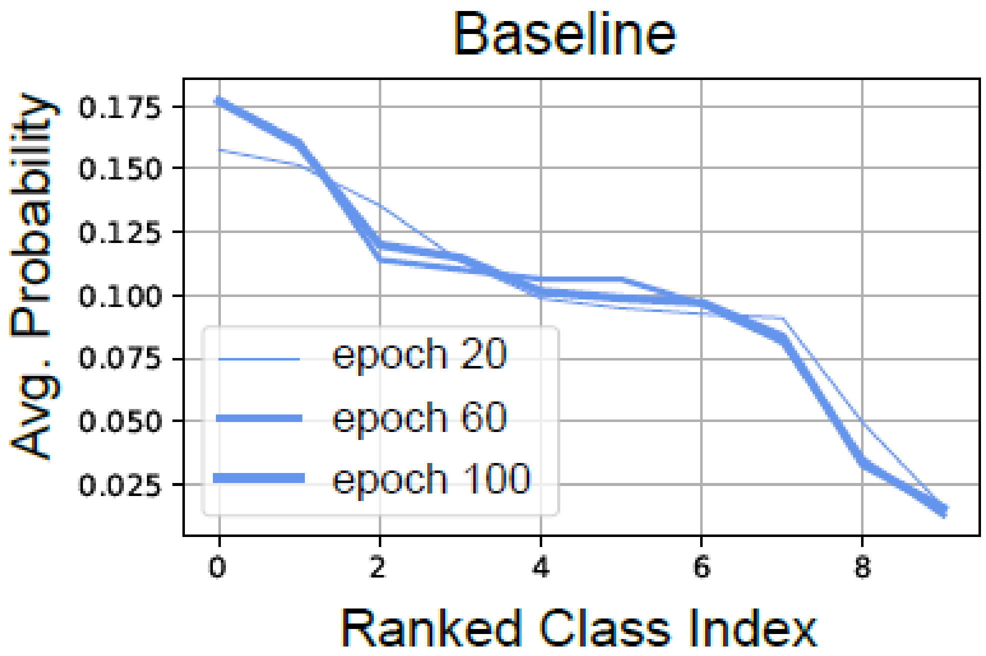
$$\mathcal{L}_s = \frac{1}{B}\sum_{i=1}^{B} H(y_i, p(\alpha(x_i)))$$

  - 教師なし損失

$$\mathcal{L}_U = \frac{1}{\mu B}\sum_{i=1}^{\mu B} \mathbb{I}[\max\left(p\left(\alpha(x_i)\right)\right) \geq \tau] \cdot H\left(y_i, p\left(\beta(x_i)\right)\right)$$

Fixmatch(SSL)

- ラベル付き、ラベルなしともに厳選されクラスバランスがとれたデータを学習した場合でも疑似ラベルのクラスバランスは悪い



CLIP(Zero-Shot Learnig)

- image – textの対になったデータをそれぞれエンコードした後、コサイン類似度を使って高くなるよう学習

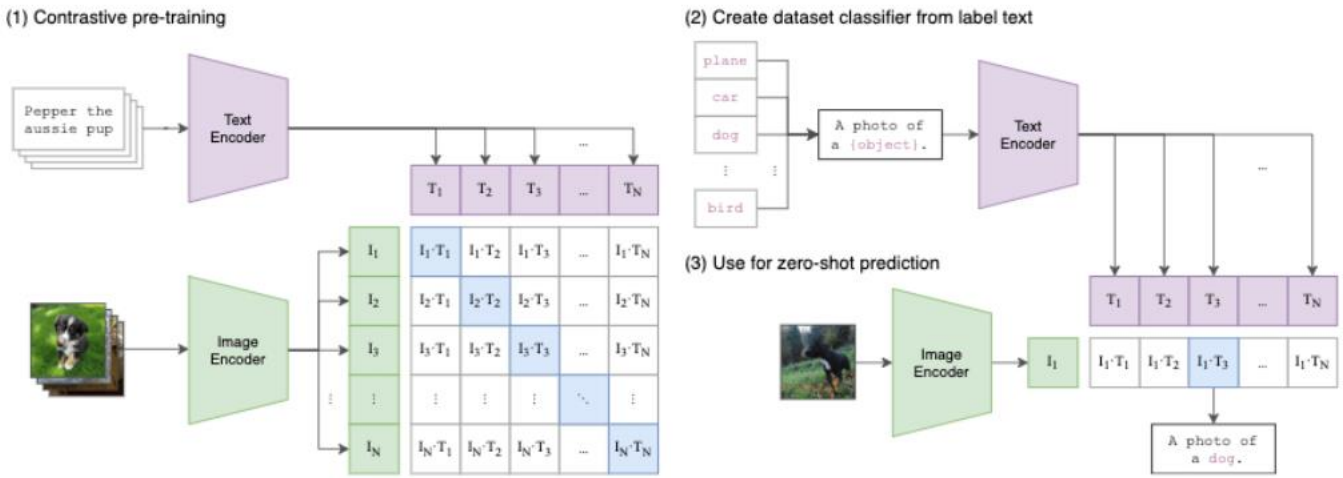


Figure 1. Summary of our approach. While standard image models jointly train an image feature extractor and a linear classifier to predict some label, CLIP jointly trains an image encoder and a text encoder to predict the correct pairings of a batch of (image, text) training examples. At test time the learned text encoder synthesizes a zero-shot linear classifier by embedding the names or descriptions of the target dataset's classes.

CLIP(ZSL)

- 1.3M ImageNetで事前学習したCLIPを用いた疑似ラベル付与のクラスごとのPrecisionとRecall
- Recallが高い多数のクラスではPrecisionの低い疑似ラベルを持つことが多い

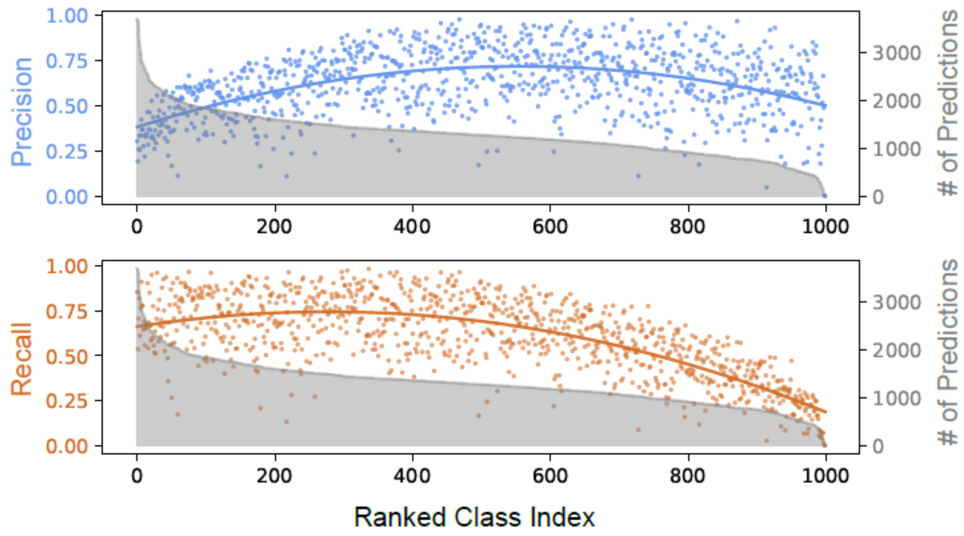


Figure 3. Per-class precision and recall of pseudo-label predictions on 1.3M ImageNet instances with a pre-trained CLIP. The majority classes with high recall often have less precise pseudo-labels.

CLIP(ZSL)

- ベンチマーク結果

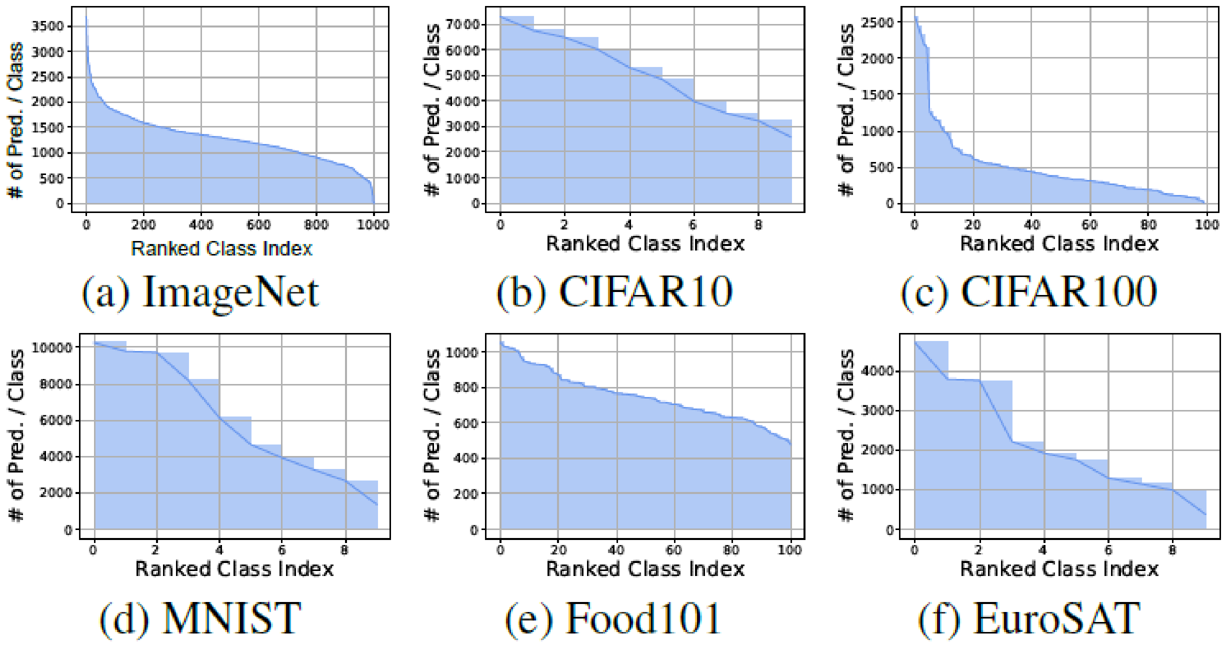
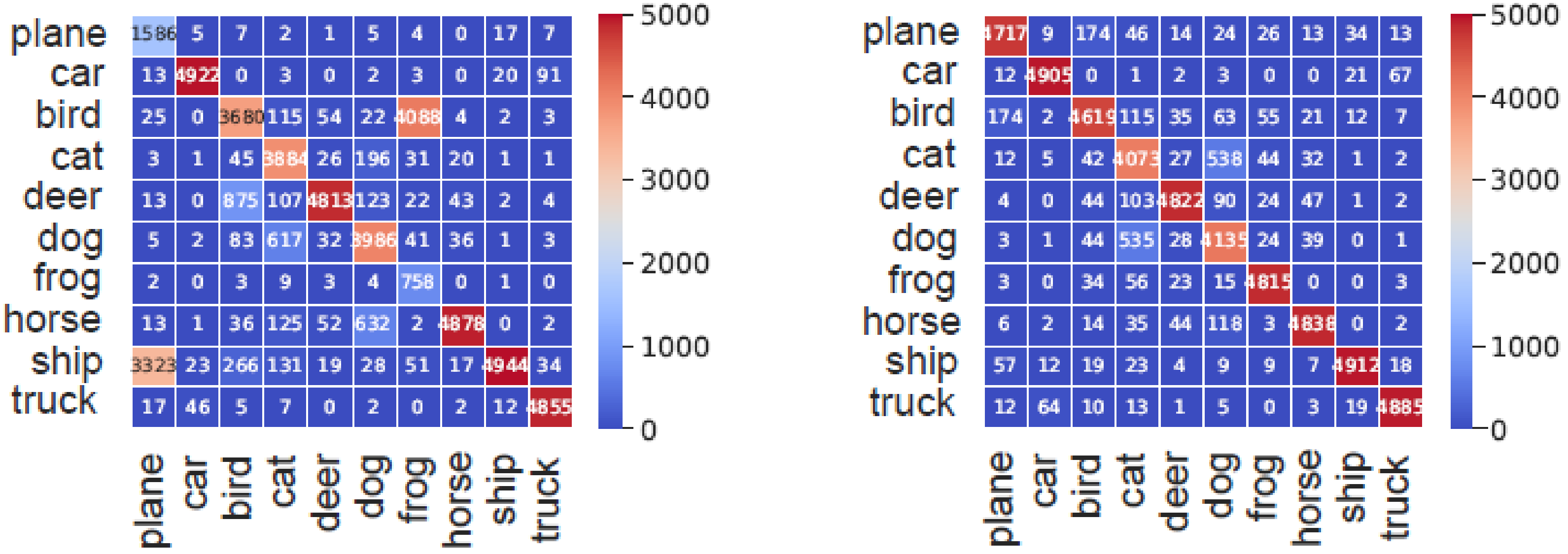


Figure 4. CLIP's zero-shot predictions are highly biased for various datasets and benchmarks.

バイアスの原因

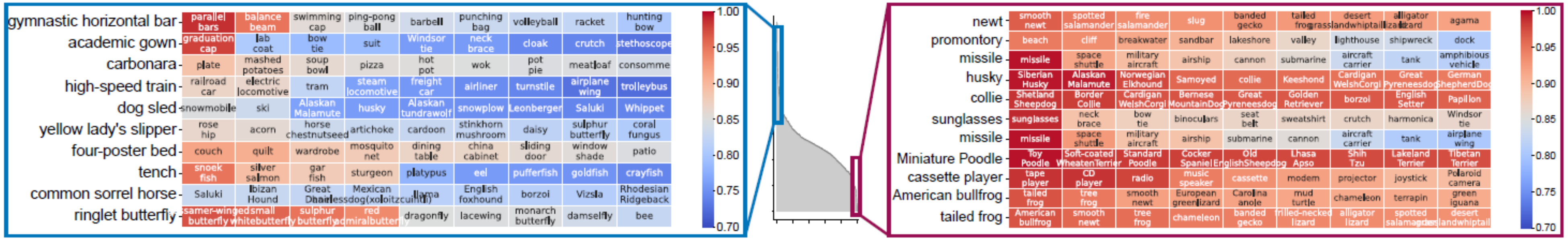
- Fixmatchの疑似ラベル混同行列
- クラス間の交絡に大きく起因する



(a) FixMatch　　(b) DebiasPL

バイアスの原因

- クラス間の相関を分析するため、事前学習CLIPから画像エンコーダによって抽出された正規化画像特徴の平均を取得
- クラスごとに１つの画像重心(image centroid)を計算
- least-10 classesはクラス間の混同が強い



a) top-10 classes　　　b) least-10 classes

Figure 5. The low-frequency classes of ImageNet, with the least-10 number of CLIP predictions per class, usually have strong inter-class correlations, while the high-frequency classes are the opposite. We compare the cosine similarity between each class's image embedding centroid and embedding centroids of its nine closest "negative" classes. (better view zoomed in)

Debiased Pseudo-Labeling(DebiasPL)

- 異なる分布に従うデータセットとラベルなし画像であってもクラス分布の事前知識を必要とせず、生徒モデルに対する偏った疑似ラベルの影響を動的に緩和する

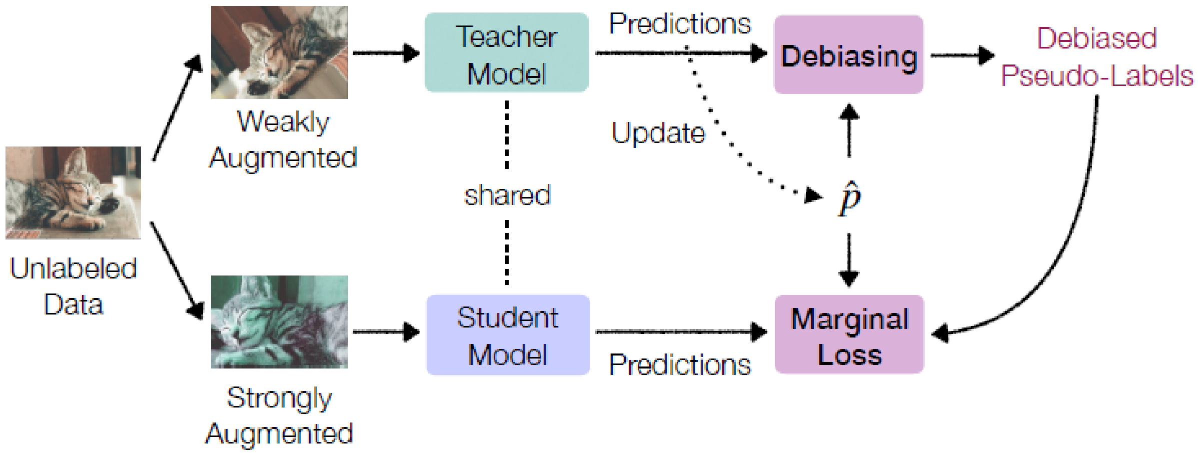


Figure 7. Diagram of the proposed Adaptive Debiasing module and Adaptive Marginal Loss, added to the top of FixMatch.

Controlled Direct Effect(CDE)

- バイアスの影響を緩和するために、A_iからYを直接的な因果効果を追求

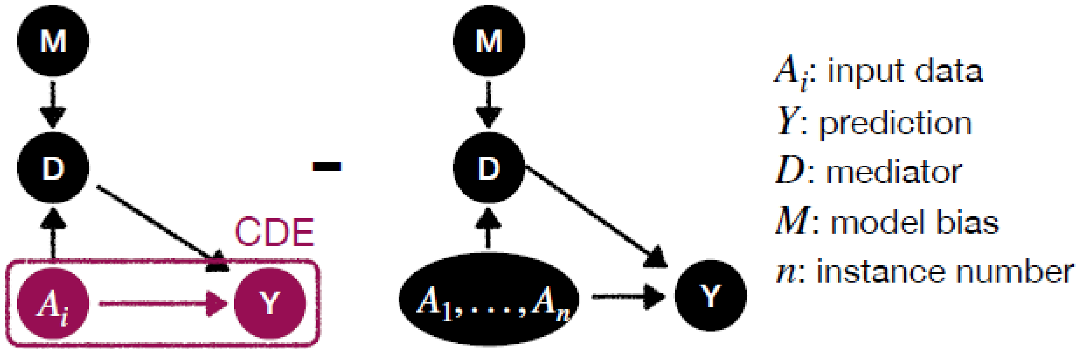


Figure 8. Causal graph of debiasing with counterfactual reasoning.

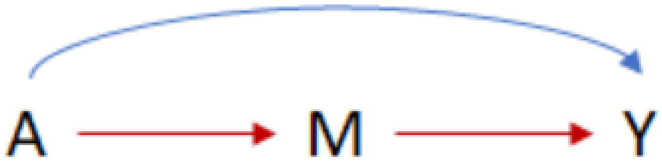
$$\mathrm{CDE}(Y_i) = [Y_i | do(A_i), do(D)] - [Y_i | do(\hat{A}), do(D)]$$

Controlled Direct Effect(CDE)

- **平均因果効果（Average Causal Effect）**
  - ある集団に対して潜在的結果変数の差の期待値

$$E[Y_{a=1}] - E[Y_{a=0}]$$

- **CDE**
  - 条件Mを固定したAからYへの影響



$$CDE(m) = Y_{1m} - Y_{0m}$$

Aximated Controlled Direct Effect

- すべての訓練データ１つずつで**反実仮想結果(counterfactual outcome)**を測定するのは計算がかかる
- ACDEはバイアスが大きく変化しない前提なので近似として用いる
- 反実推論を用いた脱バイアスロジット関数

$$\tilde{f}_i = f(\alpha(x_i)) - \lambda \log \hat{p}$$

$$\hat{p} \leftarrow m\hat{p} + (1-m)\frac{1}{\mu B}\sum_{k=1}^{\mu B} p_k$$

Adaptive Marginal Loss

- 偏りの少ないクラスと大きいクラスのマージンを大きくする
- 大きいクラスが他のクラスを圧倒しないように調整
- AMLの定式化

$$\mathcal{L}_{\mathrm{AML}} = -\log \frac{e^{(z_{\hat{y_i}} - \Delta_{\hat{y_i}})}}{e^{(z_{\hat{y_i}} - \Delta_{\hat{y_i}})} + \sum_{k \neq \hat{y_i}}^{C} e^{(z_k - \Delta_k)}}$$

where $\Delta_j = \lambda \log(\frac{1}{\hat{p}_i})$ for $j \in \{1, ..., C\}$, $z = f(\beta(x_i))$.

- AMLを用いて、 $\mathcal{L}_U$ の $H\left(y_i, p(\beta(x_i))\right)$ を置換

従来手法との比較

- 従来手法ではマージンの調整は静的だった
- 提案手法では動的な変化の過程であるべきと主張

Desired Properties	LA or LDAM	Causal Norm	DA	Ours
Improve representation learning at training time	✓	✗	✓	✓
No prior knowledge on true marginal class distribution	✗	✓	✗	✓
Adaptive as the training progresses	✗	✗	✓	✓
Applicable to both imbalanced and balanced data	✗	✗	✓	✓
Source and target data can come from varying distributions	✗	✗	✗	✓

Table 1. Our method is the only one with all these desired properties. Comparisons with previous works concentrating on resolving training data distribution issues, including LA [35], LDAM [10], DA [4], Causal Norm [54] and our DebiasPL, in key properties. Desired (undesired) properties are in green (red).

実験(SSL)

- CIFAR10, CIFAR10-LT, ImageNet-1Kなどで検証
- クラス分布に関する予備知識ない場合

Method	CIFAR10-LT: # of labels (percentage)				CIFAR10: # of labels (percentage)		
	γ=100		γ=200				
	1244 (10%)	3726 (30%)	1125 (10%)	3365 (30%)	40 (0.08%)	80 (0.16%)	250 (2%)
UDA [63] §	-	-	-	-	71.0 ±6.0	-	91.2 ±1.1
MixMatch [5] §	60.4 ±2.2	-	54.5 ±1.9	-	51.9 ±11.8	80.8 ±1.3	89.0 ±0.9
CReST w/ DA [62]	75.9 ±0.6	77.6 ±0.9	64.1 ±0.22	67.7 ±0.8	-	-	-
CReST+ w/ DA [62]	78.1 ±0.8	79.2 ±0.2	67.7 ±1.4	70.5 ±0.6	-	-	-
CoMatch w/ SimCLR [12, 31]	-	-	-	-	92.6 ±1.0	94.0 ±0.3	95.1 ±0.3
FixMatch [53] §	67.3 ±1.2	73.1 ±0.6	59.7 ±0.6	67.7 ±0.8	86.1 ±3.5	92.1 ±0.9	94.9 ±0.7
FixMatch w/ DA w/ LA [4, 35, 53, 62] §	70.4 ±2.9	-	62.4 ±1.2	-	-	-	-
FixMatch w/ DA w/ SimCLR [4, 12, 53] §	-	-	-	-	89.7 ±4.6	93.3 ±0.5	94.9 ±0.7
DebiasPL (w/ FixMatch)	**79.2 ±1.0**	**80.6 ±0.5**	**71.4 ±2.0**	**74.1 ±0.6**	**94.6 ±1.3**	**95.2 ±0.1**	**95.4 ±0.1**
gains over the best FixMatch variant	**+8.8**	**+7.5**	**+9.0**	**+6.4**	**+4.9**	**+1.9**	**+0.5**

Table 2. Without any prior knowledge of the marginal class distribution of unlabeled/labeled data, the performance of DebiasPL on *both* **CIFAR** ***and*** **CIFAR-LT SSL benchmarks** surpasses previous SOTAs, which are *either designed for balanced data or meticulously tuned for long-tailed data*. DibasMatch is experimented with the same set of hyper-parameters across all benchmarks. § states the best-reported results of counterpart methods, copied from [31], [53] or [62]. γ: imbalance ratio. We report results averaged on 5 different folds.

不均衡率

- （疑似ラベルの）**クラス間でのバランスを測る**
  - データセット:N クラス:C
  - $N_1 \geq N_2 \geq \cdots \geq N_C$:クラスごとのデータ数

$$\gamma = \frac{N_1}{N_C}$$

  - $\gamma \in [100, 200]$とはクラスの最大と最小で100~200倍

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, pages 10857–10866, 2021. 2, 7

実験(SSL)

- ラベル付きデータの割合を変えて検証

Method	B.S.	#epochs	Pre-train	1%		0.2%	
				top-1	top-5	top-1	top-5
FixMatch w/ DA [4, 53]	4096	400	✗	53.4	74.4	-	-
FixMatch w/ DA [4, 53]	4096	400	✓	59.9	79.8	-	-
FixMatch w/ EMAN [9, 53]	384	50	✓	60.9	82.5	43.6*	64.6*
DebiasPL w/ FixMatch	384	50	✓	**63.1** (+2.2)	**83.6** (+1.1)	**47.9** (+3.7)	**69.6** (+5.0)
DebiasPL (multi-views)	768	50	✓	**65.3** (+4.4)	**85.2** (+2.7)	**51.6** (+8.0)	**73.3** (+8.7)
DebiasPL (multi-views)	768	200	✓	**66.5** (+5.6)	**85.6** (+3.1)	**52.3** (+8.7)	**73.5** (+8.9)
DebiasPL (multi-views)	1536	300	✓	**67.1** (+6.2)	**85.8** (+3.3)	-	-
DebiasPL w/ CLIP [45]	384	50	✓	**69.1** (+8.2)	**89.1** (+6.6)	**68.2** (+24.6)	**88.2** (+23.6)
DebiasPL w/ CLIP (multi-views) [45]	768	50	✓	**70.9** (+10.0)	**89.3** (+6.8)	**69.6** (+26.0)	**88.4** (+23.8)
CLIP (few-shot) [45, 68]	256	50	✓	53.4	-	40.0	-
SwAV [11]	4096	50	✓	53.9	78.5	-	-
SimCLRv2 (+ Self-distillation) [13]	4096	400	✓	60.0	79.8	-	-
PAWS (multi-crops) † [2]	4096	50	✓	66.5	-	-	-
CoMatch (multi-views) [31]	1440	400	✓	67.1	87.1	-	-

Table 3. DebiasPL delivers state-of-the-arts results on **ImageNet-1K semi-supervised learning** with various fractions of labeling samples, especially for extremely low-shot settings. All results are produced with a backbone of ResNet-50. †: unsupervised pre-trained for 800 epochs, except for PAWS [2], which is pre-trained for 300 epochs with pseudo-labels generated non-parametrically. *: reproduced.

実験(SSL)

- ラベル付きとなしの分布で異なる分布で従う場合
- CIFAR10におけるSSL手法の比較で5 foldの平均でTop-1

Method	Labeled: **LT**; 10% labeled, $\gamma = 200$	
	Unlabeled: **LT**	Unlabeled: **Balanced**
FixMatch [53]	62.3 ±1.6	72.1 ±2.3
DebiasPL	**71.4 ±2.0** (+9.1)	**83.5 ±2.4** (+11.4)

Table 4. DebiasPL consistently improves the performance of SSL when the unlabeled data is either the sames as labeled data, i.e., long-tailed distributed, or different with labeled data, i.e., balanced distributed across semantics. We report results averaged on 5 folds.

	FixMatch	MixMatch	UDA
Baseline	89.7 ± 4.6	47.5 ± 11.5	29.1 ± 5.9
+ DebiasPL	**94.6 ± 1.3**	**61.7 ± 6.1**	**43.2 ± 5.2**

Table 5. **DebiasPL is a universal add-on.** Top-1 accuracies of various SSL methods on CIFAR10, averaged on 5 folds, are compared. 4 instances per class are labeled.

実験(ZSL)

- ImageNet-1KでのZSLの結果

Method	#param	Accuracy (%) top-1	Accuracy (%) top-5
ConSE [39]	-	1.3	3.8
DGP [23]	-	3.0	9.3
ZSL-KG [38]	-	3.0	9.9
Visual N-Grams [30]	-	11.5	-
CLIP (prompt ensemble) [45]	26M	59.6	-
(ours) CLIP + FixMatch	26M	55.7	80.6
(ours) CLIP + DebiasPL	26M	**68.3** (+8.7)	**88.9** (+8.3)
CLIP (few-shot) [45, 68] †	26M	53.4	-
CLIP + CoOp (few-shot) [68] †	26M	60.9	-
CLIP (ViT-B/32) [45]	398M	63.2	-
CLIP (ResNet50x4) [45]	375M	65.8	-

Table 6. DebiasPL delivers state-of-the-art results of **zero-shot learning on ImageNet-1K**, outperforming CLIP with bigger models or fine-tuned with labels. †: CoOp and CLIP (few-shot) are fine-tuned with about 1.5% annotated data.

実験(ZSL)

- 様々なデータセットでZSLを行った場合でのDebiasPLはドメインシフトに対してより強い頑健性を示す

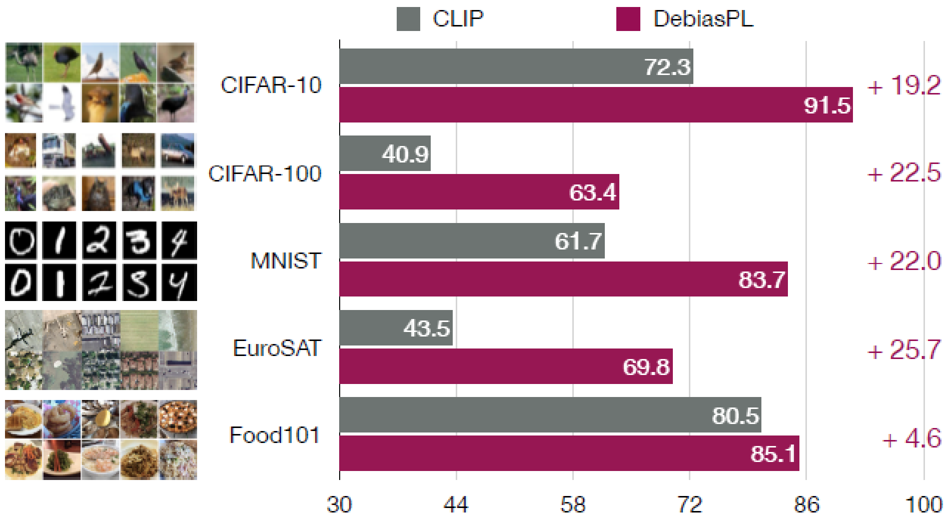


Figure 9. DebiasPL exhibits stronger robustness to domain shift when conducting **zero-shot learning on various datasets**. We experiment with ResNet-50 as a backbone network. CLIP results are reproduced with official codes.

結論

- 偏った疑似ラベルの原因がクラス間の交絡に起因
- 提案手法はSSL・ZSLともに従来手法を超える精度
- 様々なデータセット(CIFAR10, ImageNet-1Kなど)であっても精度がよい