Does Mixed Label Imbalance Matter to Minority Collapse in Imbalanced Learning?

Hoyong Kim
[1] DGIST InnoCORE Research Group for Bio-Embodied Physical AI
[2] Seoul National University

Kangil Kim
Department of Artificial Intelligence,
Gwangju Institute of Science and Technology

Paper

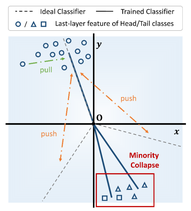
About Me

TL;DR

- We ask whether **mixed-label balance** matters for minority collapse under Mixup, and analyze it in the unconstrained features model with Mixup (UFM_λ)
- Class-balanced samplers balance mixed labels **only in expectation**: epoch-wise label imbalance remains and lowers accuracy
- **BMLS** balances mixed labels in every epoch; the parameter-free **MS classifier** treats each mixed label as its own target

Problem: Minority Collapse

- Classifiers become **biased toward head classes** with abundant samples
- **Class vectors of minority classes converge** and become nearly indistinguishable
- This **minority collapse** degrades accuracy on the minority classes
- Common remedy: Mixup with class-balanced sampling

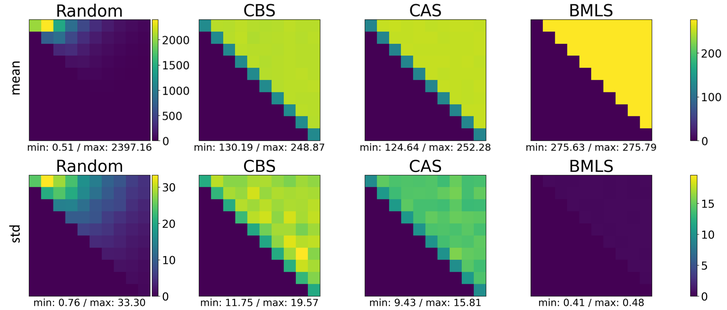


Trained tail-class vectors collapse (red box)

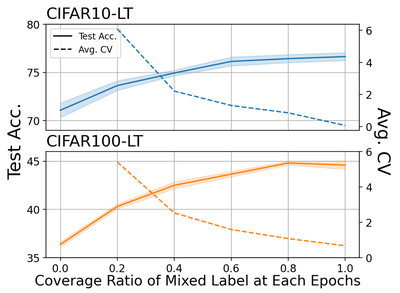
YES!

The imbalance of mixed labels invokes minority collapse, and existing class-balanced samplers do not resolve this problem

The Randomness of Mixup Breaks Epoch-wise Label Balance!



Class-balanced samplers (CBS, CAS) make the mean label frequency nearly uniform, but its epoch-wise standard deviation remains large; only BMLS makes both uniform



Accuracy rises as more mixed labels are covered each epoch

Take-Home Message

- **Balance your mixed-label samples!**
- Use **BMLS** when your data is long-tailed and you train with Mixup
- Add the **MS classifier** when mixed labels are balanced and each mixed label gets more than one sample per epoch (per-pair coverage > 1)
- Future work: extend the theory to deeper networks and other architectures (RNN, Transformer, diffusion models)
- How about CutMix or other Mixup variants?

Experiments

Method	imb	CIFAR10-LT 200	100	50	10	CIFAR100-LT 200	100	50	10
Mixup		66.77	72.94	78.64	88.05	39.06	42.88	48.31	63.03
+LOM		70.17	76.63	81.15	89.24	39.61	44.24	49.99	63.90
+CAS		69.90	76.43	81.42	89.24	40.28	44.65	50.07	63.57
+$BMLS_{MS}$		74.70	79.67	83.46	88.51	41.71	47.62	52.74	64.47
	diff.	+7.93	+6.73	+4.82	+0.46	+2.65	+4.74	+4.43	+1.44
ETF+DR		71.58	76.82	81.25	87.59	41.20	45.07	50.71	63.08
BMLS+$WETF_{MS}$+CE		77.73	80.31	84.22	88.26	42.73	47.10	52.44	64.10
	diff.	+6.15	+3.49	+2.97	+0.67	+1.53	+2.03	+1.73	+1.02
Remix		69.58	75.15	80.41	88.61	41.03	44.95	50.19	63.45
+BMLS		73.95	80.10	83.92	88.62	39.95	46.34	51.53	64.42
+$BMLS_{MS}$		73.18	78.00	83.70	88.20	40.25	46.82	49.78	63.54
	diff.	+3.60	+2.85	+3.29	-0.41	-0.78	+1.87	-0.41	+0.09
DBN-mix		77.40	82.40	86.05	**91.01**	40.71	45.52	50.47	62.68
+$BMLS_{MS}$		**79.73**	**84.30**	**87.28**	90.93	**44.42**	**49.08**	**55.41**	**65.42**
	diff.	+2.33	+1.90	+1.23	-0.08	+3.71	+3.56	+4.94	+2.74

Top-1 accuracy (%) on CIFAR10/100-LT by imbalance factor (imb)

	Method	Clf.	CIFAR10-LT many	med	few	all	CIFAR100-LT many	med	few	all
Smpl.	random	FC	**93.39**	74.05	50.99	72.94	**72.09**	41.10	8.77	42.88
	BMLS	FC	88.53	**77.84**	**70.53**	**78.85**	68.38	**46.89**	**14.37**	**45.20**
		diff.	-4.86	+3.79	+19.54	+5.91	-3.71	+5.79	+5.60	+2.32
Clf.	random	MS	**89.47**	62.24	41.15	64.08	**67.29**	33.84	2.78	36.87
	BMLS	K^2	58.35	34.31	25.93	39.01	14.15	9.36	1.21	8.72
	BMLS	MS	89.14	**76.34**	**74.63**	**79.67**	66.31	**49.80**	**21.80**	**47.62**
		diff.	-0.33	+14.10	+33.48	+15.59	-0.98	+15.96	+19.02	+10.75
MS	diff. ($BMLS_{MS}$ - BMLS)		+0.61	-1.50	+4.10	+0.82	-2.07	+2.91	+7.43	+2.42

Ablation of sampler and classifier (imbalance factor 100)

$BMLS_{MS}$ beats Mixup, LOM and CAS in every setting except CIFAR10-LT with imb 10 (up to +7.93), with the largest gains on few-shot classes

Theoretical Analysis

How about Mixup? Does the balance of mixed labels also matter to minority collapse?

Lemma 4.1. *Assume $p \geq 2K$ and the loss function $\mathcal{L}$ is convex in its first argument. Let $\boldsymbol{X}^\star$ be a minimizer of the convex program. Define $(\boldsymbol{W}^\star, \boldsymbol{H}^\star)$ as*

$$\left[\boldsymbol{h}^\star_{(1,1)}, \boldsymbol{h}^\star_{(1,2)}, \ldots, \boldsymbol{h}^\star_{(K,K)}, (\boldsymbol{W}^\star)^\top\right] = \boldsymbol{P}(\boldsymbol{X}^\star)^{1/2}$$
$$\boldsymbol{h}^\star_{k,i} = \boldsymbol{h}^\star_k, \qquad \forall\, i \in \mathcal{I}^\lambda_k, k \in \mathbb{K}^2,$$

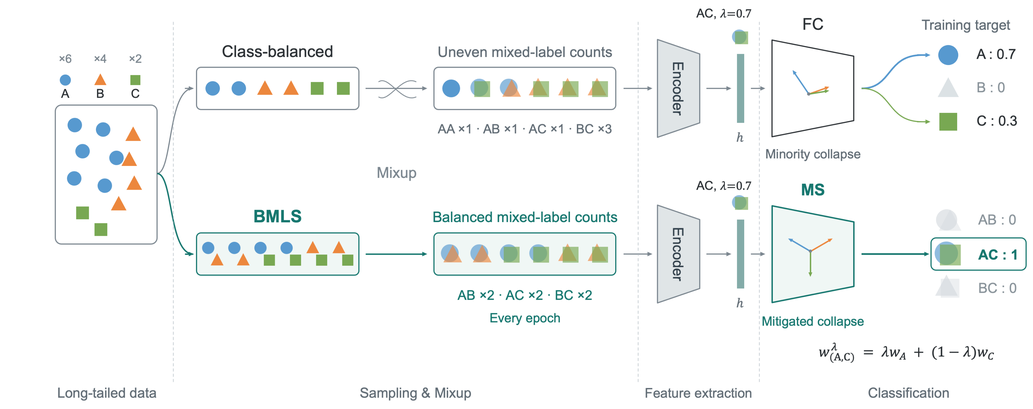
where $(\boldsymbol{X}^\star)^{1/2}$ denotes the positive square root of $\boldsymbol{X}^\star$and $\boldsymbol{P} \in \mathbb{R}^{p\times(K^2+K)}$ is any partial orthogonal matrix such that $\boldsymbol{P}^\top\boldsymbol{P} = \boldsymbol{I}_{K^2+K}$. Then, $(\boldsymbol{W}^\star, \boldsymbol{H}^\star)$ is a minimizer of UFM_λ. Moreover, if all $\boldsymbol{X}^\star$'s satisfy $\frac{1}{K^2}\sum_{k=1}^{K^2} \boldsymbol{X}^\star(k,k) = E_H$, then all the solutions of UFM_λ are in the form of $\boldsymbol{P}(\boldsymbol{X}^\star)^{1/2}$

Theorem 4.2. *Assume $p \geq K$ and $n_A/n_B \to \infty$, and fix K_A and K_B. Let $(\boldsymbol{W}^\star, \boldsymbol{H}^\star)$ be any global minimizer of the UFM_λ. As the imbalance factor $R \equiv n_A/n_B \to \infty$, we have*

$$\lim \boldsymbol{w}^\star_k - \boldsymbol{w}^\star_{k'} = \boldsymbol{0}_p, \forall\, K_A < k < k' \leq K.$$

From *Lemma 4.1* and *Theorem 4.2*, **UFM_λ also exhibits minority collapse** when mixed labels are imbalanced. Balancing mixed labels acts as oversampling the minority mixed labels and mitigates it (Proposition 4.3).

Proposed Method: BMLS and Mixed-Singleton (MS) Classifier



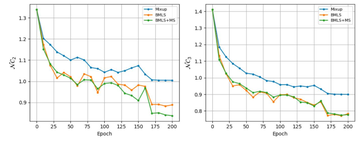
Toy long-tailed data (6, 4, 2 samples): class-balanced sampling leaves mixed-label counts uneven (1:1:1:3), whereas BMLS equalizes them (2:2:2) every epoch. FC scores a mixed sample against singleton classes; MS scores it against its own mixed label.

- **BMLS** pre-defines the index pairs $\tilde{\mathcal{J}}^\lambda$ so that every mixed label appears equally often in every epoch:

$$P((i,\pi(i))) = \frac{1}{N}, \qquad \forall (i,\pi(i)) \in \tilde{\mathcal{J}}^\lambda, \qquad n_{(a,b)} = n, \qquad \forall (a,b) \in \mathbb{K}^2$$

- **MS classifier** scores each mixed label as its own one-hot target against the interpolated class vector, adding no parameters:

$$\boldsymbol{W}^\lambda = [\lambda \boldsymbol{w}_a + (1-\lambda)\boldsymbol{w}_b]_{(a,b)\in\mathbb{K}^2}, \qquad \tilde{\boldsymbol{y}}^\lambda_i = \boldsymbol{e}^{\mathcal{I}^2(c_i, c_{\pi(i)})}$$



NC_2 (simplex ETF error) and NC_3 (self-duality error) on CIFAR100-LT, imb. 100; lower is better

Contribution

- **Theory:** minority collapse persists in UFM_λ under imbalanced mixed labels
- **Finding:** class-balanced samplers leave **epoch-wise label imbalance**, which lowers accuracy and worsens collapse geometry
- **BMLS** equalizes mixed-label frequencies every epoch and works with a standard FC classifier
- **MS classifier** interprets mixed labels as singletons and recovers the balanced simplex-ETF geometry with no extra parameters