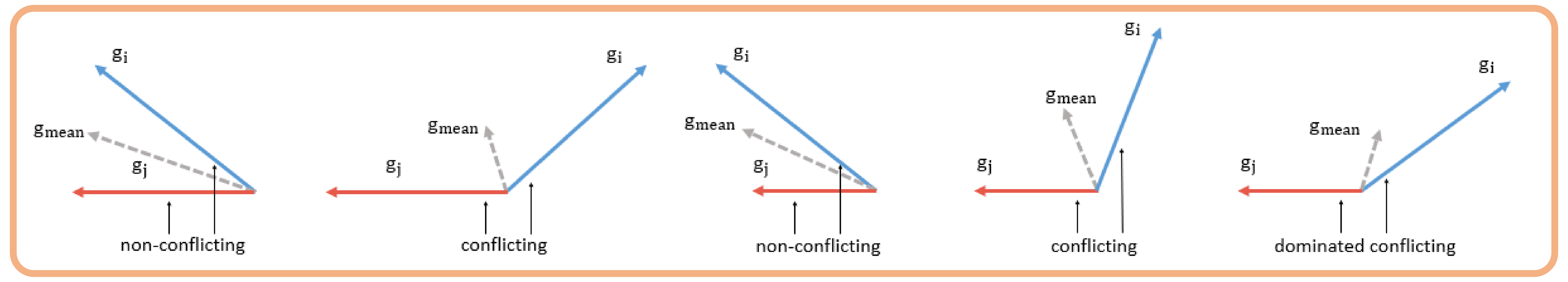
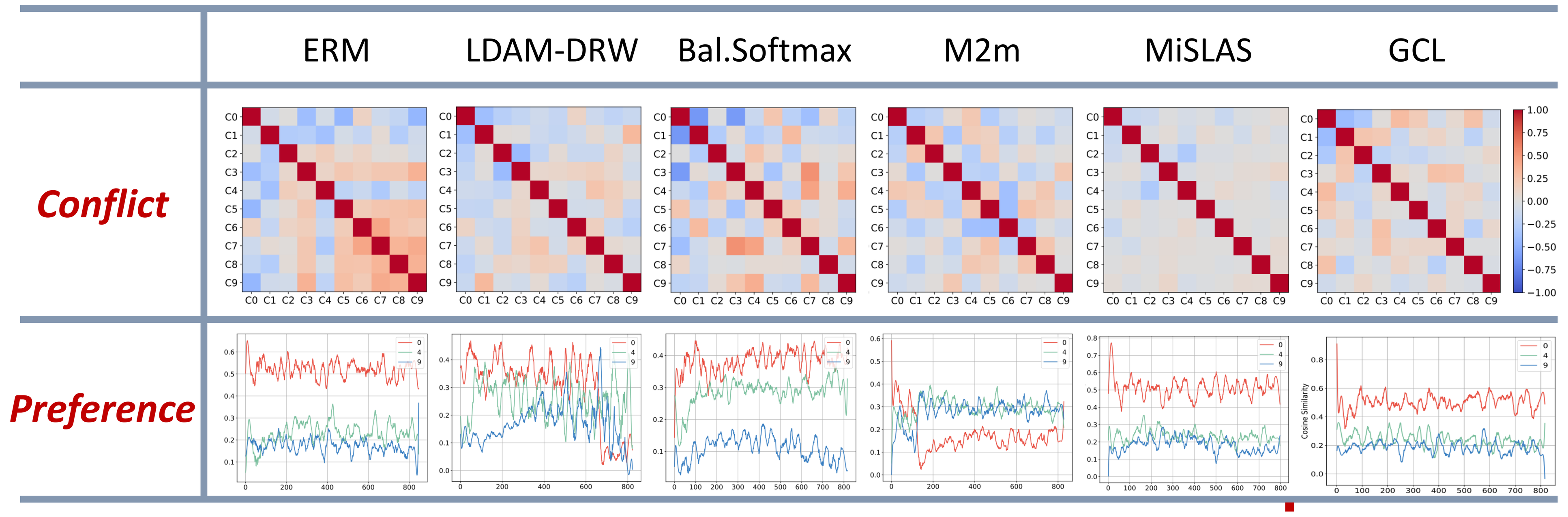
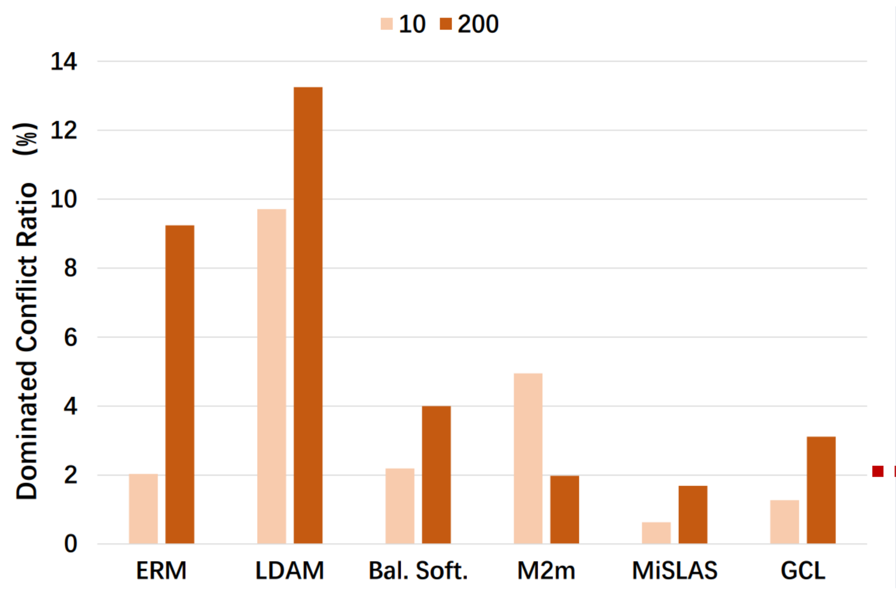
Pareto Deep Long-Tailed Recognition: A Conflict-Averse Solution

Zhipeng Zhou (USTC), Liu Liu (Tencent), Peilin Zhao (Tencent), Wei Gong (USTC)

Motivation

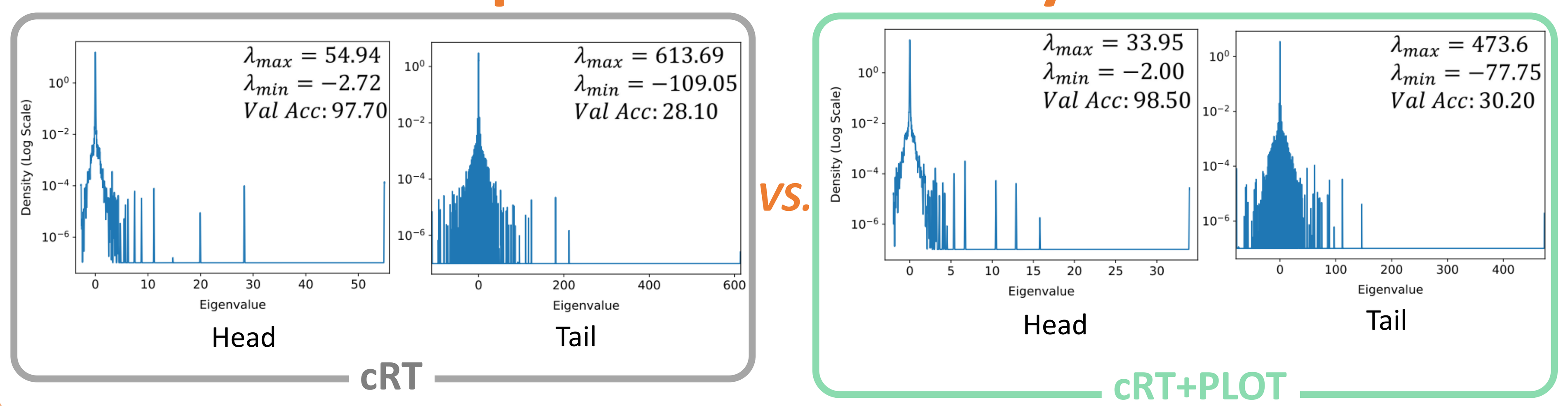




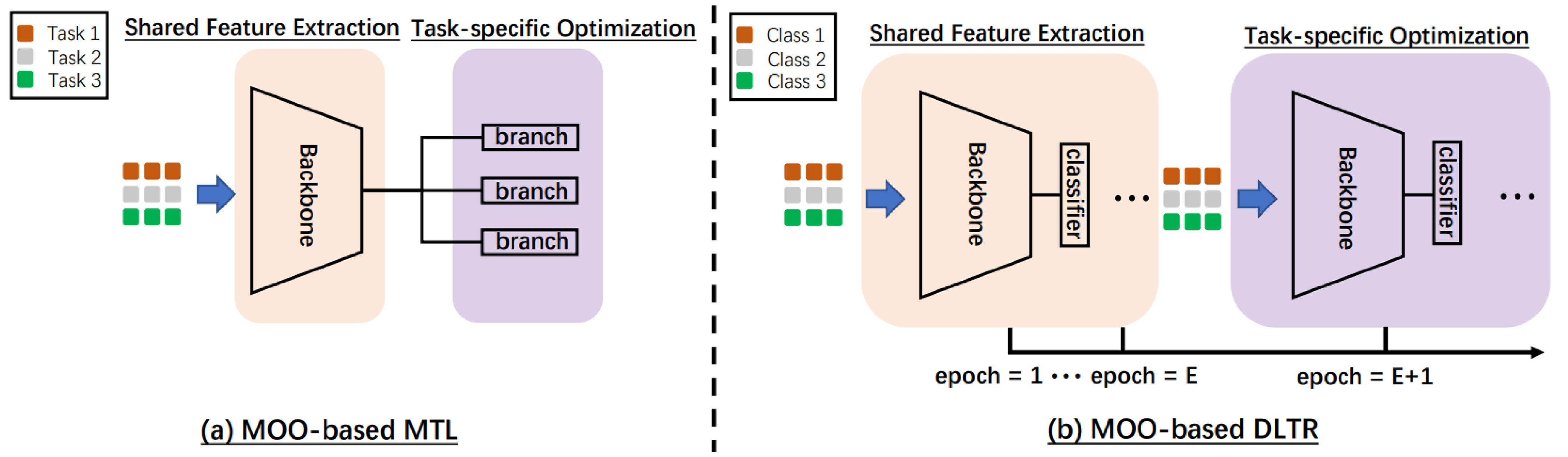


- There exist ***conflict*** and ***preference*** issues in LT models.
- The issues are ***exacerbated*** in the LT setting.

Representation Analysis



MOO: From MTL to DLTR



Mitigate the conflict issue during representation learning

Can MOO Benefit LT Models?

Imb.	cRT+Mixup							LDAM-DRW						
	Vanilla	w/ EPO		w/ MGDA		w/ CAGrad		Vanilla	w/ EPO		w/ MGDA		w/ CAGrad	
200	73.06	33.45	76.24▲	68.05	75.98▲	75.15	76.02▲	71.38	56.04	73.64▲	67.18	74.08▲	55.80	73.28▲
100	79.15	34.27	79.69▲	73.71	79.26▲	79.58	80.16▲	77.71	66.49	77.25▼	73.70	77.79▲	66.49	76.86▼
50	84.21	36.53	83.79▼	79.27	84.15▼	83.52	84.49▲	81.78	72.60	81.62▼	78.24	81.58▼	69.26	81.85▲
Imb.	**Balanced Softmax**							**M2m**						
	Vanilla	w/ EPO		w/ MGDA		w/ CAGrad		Vanilla	w/ EPO		w/ MGDA		w/ CAGrad	
200	81.33	45.37	81.40▲	74.13	80.90▼	79.20	80.93▼	73.43	51.90	73.07▼	57.14	72.63▼	70.95	73.84▲
100	84.90	44.33	85.30▲	79.06	85.10▲	83.77	85.40▲	77.55	57.89	76.57▼	52.37	76.48▼	76.24	77.95▲
50	89.17	41.43	88.97▼	79.43	88.90▼	88.00	89.27▲	80.94	42.07	81.19▲	46.38	80.66▼	78.19	81.11▲
Imb.	**MiSLAS**							**GCL**						
	Vanilla	w/ EPO		w/ MGDA		w/ CAGrad		Vanilla	w/ EPO		w/ MGDA		w/ CAGrad	
200	76.59	36.62	76.97▲	63.40	76.12▼	76.30	77.43▲	79.25	62.08	79.73▲	75.43	80.03▲	78.73	80.08▲
100	81.33	39.92	81.22▼	68.09	82.00▲	82.10	82.47▲	82.85	74.78	82.75▼	79.01	82.81▼	82.48	83.48▲
50	85.23	44.78	84.60▼	70.20	84.84▼	85.20	85.33▲	86.00	78.42	84.55▼	81.89	85.58▼	85.31	85.90▼

- Generally, applying MOO with our framework benefits LT models
- **CAGrad** shows a more **stable** performance compared to others

Generalization Guarantee

Theorem 1 *(MOO-based DLTR Generalization Bound)* *If the loss function l_k belonging to k_{th} category is M_k-Lipschitz, and $\forall(\boldsymbol{x},\boldsymbol{y}),(\boldsymbol{x}',\boldsymbol{y}') \in \mathcal{X}\times\mathcal{Y}$, $\forall \boldsymbol{h} \in \boldsymbol{\mathcal{H}}$: $\|[\boldsymbol{h}(\boldsymbol{x}),\boldsymbol{y}]-[\boldsymbol{h}(\boldsymbol{x}'),\boldsymbol{y}']\| \leq \mathcal{D}_{\mathcal{H}}$, assume $M_k\mathcal{D}_{\mathcal{H}}$ is bounded by M, then for any $\epsilon > 0$ and $\delta > 0$, with probability at least $1-\delta$, the following inequality holds for $\forall \boldsymbol{h} \in \boldsymbol{\mathcal{H}}$ and $\forall \boldsymbol{\omega} \in \boldsymbol{\mathcal{W}}$:*

$$\mathcal{L}_{\boldsymbol{\omega}}(\boldsymbol{h}) \leq \hat{\mathcal{L}}_{\boldsymbol{\omega}}(\boldsymbol{h}) + 2\mathfrak{R}_S(\boldsymbol{\mathcal{G}},\boldsymbol{\omega}) + \sum_{k=1}^{K}\boldsymbol{\omega_k} M\sqrt{\frac{\log\frac{1}{\delta}}{2}}$$

Derived Upper Bound

⇩

Variability Collapse Loss

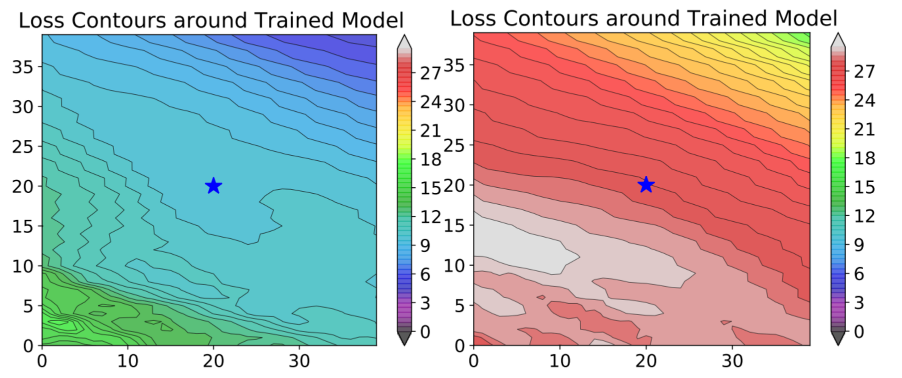
$$\mathcal{L}_{vc} = \frac{1}{K}\sum_{k=1}^{K}\mathrm{Std}(\tilde{l}(\boldsymbol{x}_*^{\boldsymbol{k}},\boldsymbol{y}_*^{\boldsymbol{k}}))$$

- Provide a theoretically derived design for generalization
- Happens to share the same formula with ***Neural Collapse***

Convergence Guarantee

Theorem 2 *(Convergence of CAGrad in DLTR) With a fix step size α and the assumption of H-Lipschitz on gradients, i.e., $\|\nabla\mathcal{L}_i(\boldsymbol{\theta}) - \nabla\mathcal{L}_i(\boldsymbol{\theta}')\| \leq H\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|$ for i = 1, 2, ..., K. Denote $\mathbf{d}^*(\boldsymbol{\theta_t})$ as the optimization direction of CAGrad at step t, then we have:*

$$\mathcal{L}(\boldsymbol{\theta_{t+1}}) - \mathcal{L}(\boldsymbol{\theta_t}) \leq -\frac{\alpha}{2}(1-c^2)\|\boldsymbol{g_0}(\boldsymbol{\theta_t})\|^2 + \frac{\alpha}{2}(H\alpha-1)\|\boldsymbol{d}^*(\boldsymbol{\theta_t})\|^2,$$



SAM: $\min_{\boldsymbol{\theta}}\max_{\boldsymbol{\epsilon}(\boldsymbol{\theta})}\mathcal{L}(\boldsymbol{\theta}+\boldsymbol{\epsilon}(\boldsymbol{\theta}))$, where $\|\boldsymbol{\epsilon}(\boldsymbol{\theta})\|_2 \leq \rho$

Provide convergence guarantee for **MOO in DLTR**

See experimental results in our paper

Contact Us!

Email: zzp1994@mail.ustc.edu.cn
Implementation: https://github.com/zzpustc/PLOT