NEURAL INFORMATION PROCESSING SYSTEMS

The 38 Annual Conference on Neural Information Processing Systems

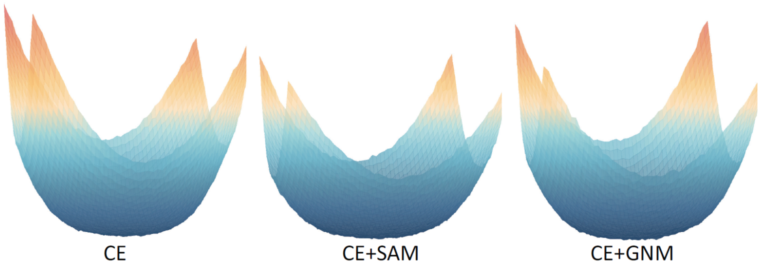
DECEMBER 10-15, 2024 | VANCOUVER, CANADA

Improving Visual Prompt Tuning by Gaussian Neighborhood Minimization for Long-Tailed Visual Recognition

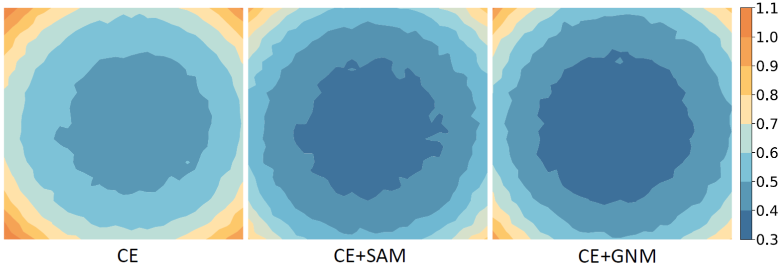
Mengke Li, Ye Liu, Yang Lu, Yiqun Zhang, Yiu-ming Cheung, Hui Huang*

Reported by Ye Liu

Introduction



(a) Side view



(b) Top view

SAM (Pierre et al. 2021) improves model generalization by flattening minima.

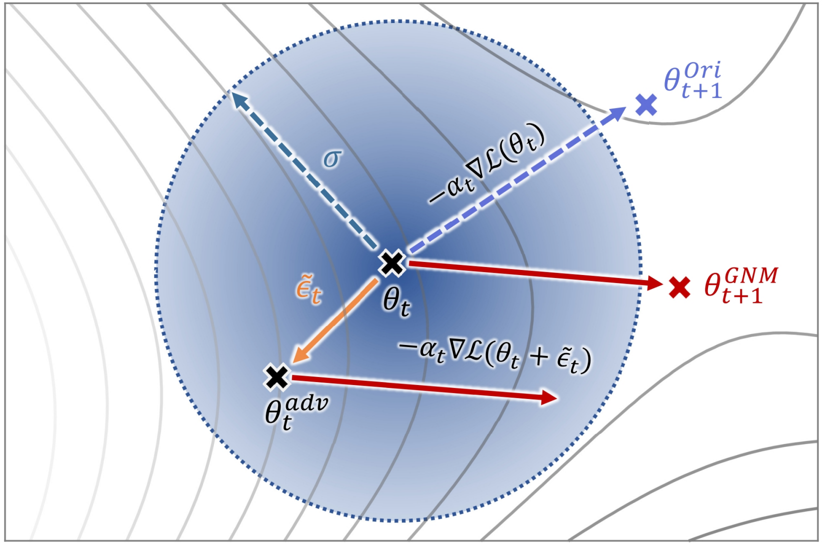
Its generalization performance is impacted by imbalanced data distributions

Motivation:

- **Flatten loss landscape** to enhance model generalization.
- Introduce **distribution-independent** perturbation.

Pierre F, Ariel K, Hossein M, and Behnam N. Sharpness-aware minimization for efficiently improving generalization. In ICLR, 2021.

Method



- Gaussian neighborhood loss:

$$L_{\mathcal{T}}^{GN}(\boldsymbol{\theta}) = \mathbb{E}_{\varepsilon_i \in \mathcal{N}(0,\sigma^2)}[L_{\mathcal{T}}(\boldsymbol{\theta} + \boldsymbol{\varepsilon})]$$

- Parameter update strategy:

$$\tilde{\boldsymbol{\varepsilon}}_t = \rho_{GNM} \cdot [\varepsilon_i]_{i=1}^{k},\ \varepsilon_i \sim \mathcal{N}(0,\sigma^2)$$

$$\boldsymbol{\theta}_{t+1}^{GNM} = \boldsymbol{\theta}_t - \alpha_t \left(\nabla_{\boldsymbol{\theta}_t} L_{\mathcal{T}}(\boldsymbol{\theta}_t)|_{\boldsymbol{\theta}_t + \tilde{\boldsymbol{\varepsilon}}_t} + \lambda \boldsymbol{\theta}_t \right)$$

Analysis

Comparison of parameter update strategies between GNM and SAM :

GNM

$$\tilde{\boldsymbol{\varepsilon}}_t = \rho_{GNM} \cdot [\varepsilon_i]_{i=1}^{k},\ \varepsilon_i \sim \mathcal{N}(0, \sigma^2),$$
$$\boldsymbol{\theta}_{t+1}^{GNM} = \boldsymbol{\theta}_t - \alpha_t \big(\nabla_{\boldsymbol{\theta}_t} L_{\mathcal{T}}(\boldsymbol{\theta}_t)|_{\boldsymbol{\theta}_t + \tilde{\boldsymbol{\varepsilon}}_t} + \lambda \boldsymbol{\theta}_t \big).$$

SAM

$$\hat{\boldsymbol{\varepsilon}}_t = \rho_{SAM} \frac{\nabla_{\boldsymbol{\theta}} L_{\mathcal{T}}(\boldsymbol{\theta}_t)}{\|\nabla_{\boldsymbol{\theta}} L_{\mathcal{T}}(\boldsymbol{\theta}_t)\|_2^2},$$
$$\boldsymbol{\theta}_{t+1}^{SAM} = \boldsymbol{\theta}_t - \alpha_t \big(\nabla_{\boldsymbol{\theta}_t} L_{\mathcal{T}}(\boldsymbol{\theta}_t)|_{\boldsymbol{\theta}_t + \hat{\boldsymbol{\varepsilon}}_t} + \lambda \boldsymbol{\theta}_t \big).$$

Our proposed GNM:

- Is well-suited for long-tailed data. GNM is in **sample-independent** manner .
- Saves computational overhead. The parameter update in GNM does **not** need additional forward and backward pass to calculate perturbations.

Experiments

Method	200	100	50	10
DNN-based model (Backbone: ResNet32)				
BBN [68]	37.2	42.6	47.0	59.1
RIDE [57]	45.8	50.4	55.0	-
MisLAS [66]	43.5	47.0	52.3	63.2
BCL [72]	-	51.9	56.6	64.9
GCL [32]	44.8	48.6	53.6	-
NCL [29]	-	54.2	58.2	-
GPaCo [7]	-	52.3	56.4	65.4
SHIKE [22]	-	56.3	59.8	-
DNN-based model with SAM				
CCSAM [71]	45.7	50.8	53.9	-
ImbSAM [70]	-	54.8	59.3	59.7
Self-attention-based model (Backbone: ViT-B/16)				
VPT [21]	72.8	81.0	84.8	89.6
LiVT [62]	-	58.2	-	69.2
LPT [10]	**87.9**	**89.1**	**90.0**	**91.0**
GNM-PT (ours)	**89.2**	**90.3**	**91.2**	**91.8**

Comparison results on CIFAR100-LT

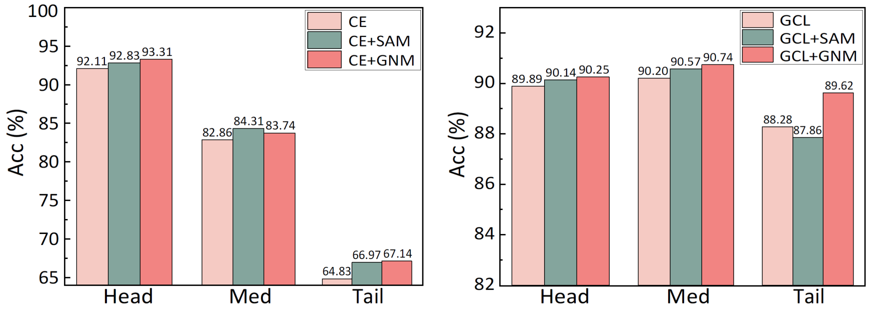
Experiments

Method	Head	Med	Tail	Overall
DNN-based model (Backbone: ResNet152)				
LWS [23]	40.6	39.1	28.6	37.6
RIDE [57]	44.4	40.6	33.0	40.4
MisLAS [66]	39.6	43.3	36.1	40.4
GCL [32]	38.6	42.6	38.4	40.3
NCL [29]	-	-	-	41.8
GPaCo [7]	39.5	47.2	33.0	41.7
SHIKE [22]	43.6	39.2	44.8	41.9
DNN-based model with SAM				
CCSAM [71]	41.2	42.1	36.4	40.6
MHSA-based model (Backbone: ViT-B/16)				
Supplementary with linguistic data				
VL-LTR [52]	54.2	48.5	42.0	**50.1**[3]
RAC [38]	48.7	48.3	41.8	47.2[3]
Visual-only				
Decoder [60]	-	-	-	46.8
LPT [10]	47.6	52.1	48.4	49.7[5]
LiVT [62]	48.1	40.6	27.5	40.8
GNM-PT (ours)	46.6	53.3	49.4	**50.1**
GNM-PT (ours)	48.6	52.1	47.9	**50.0**[4]

Method	Head	Med	Tail	Overall
DNN-based model (Backbone: ResNet50)				
LWS [23]	72.9	71.2	69.2	70.5
RIDE [57]	76.5	74.2	70.5	72.8
MisLAS [66]	73.2	72.4	70.4	71.6
GCL [32]	-	-	-	72.0
NCL [29]	72.7	75.6	74.5	74.9
GPaCo [7]	-	-	-	75.4
SHIKE [22]	-	-	-	75.4
DNN-based model with SAM				
LDAM+SAM [47]	64.1	70.5	71.2	70.1
CCSAM [71]	65.4	70.9	72.2	70.9
ImbSAM [70]	68.2	72.5	72.9	71.1
MHSA-based model (Backbone: ViT-B/16)				
Supplementary with linguistic data				
VL-LTR [52]	-	-	-	**76.8**[3]
RAC [38]	75.9	80.5	81.1	**80.2**[3]
Visual-only				
Decoder [60]	-	-	-	59.2
LPT [10]	-	-	79.3	76.1
LiVT [62]	78.9	76.5	74.8	76.1
GNM-PT (ours)	61.5	77.1	79.3	**76.5**
GNM-PT (ours)	76.3	77.6	75.0	**76.3**[4]

Comparison results on Places-LT and iNaturalist 2018.

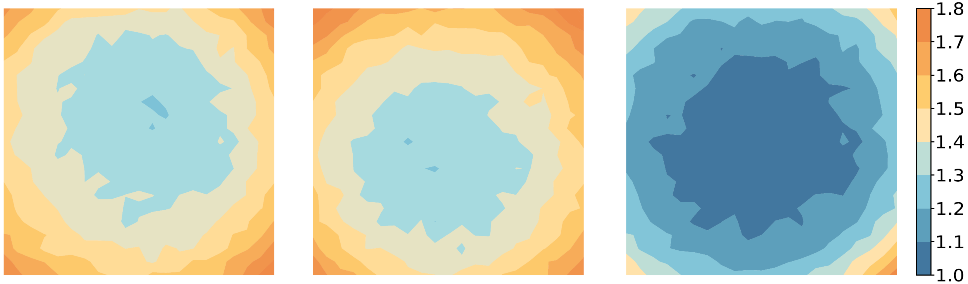
Experiments



(a) CE loss (b) GCL loss

Method	Acc. (%)	NET (s)
CE	81.02	39.78
CE+SAM	82.48	72.51
CE+GNM	**82.50**	40.16 (↓ 44.61%)
GCL+DRW	89.58	40.00
GCL+DRW+SAM	89.69	74.36
GCL+DRW+GNM	**90.28**	41.87 (↓ 43.69%)

- Consistently enhance the performance of GCL across all categories in every scenario.
- Save computational overhead



GCL GCL+SAM GCL+GNM

- Achieve a flat loss landscape.

Conclusion

Pros:

Simple and effective:

- Balance the generalization capabilities of both head and tail classes;
- Little additional computational cost.

Cons:

Need to further re-balancing the classifier:

- A rebalancing strategy is also needed to obtain a more balanced classifier.

Thanks

- More details: http://arxiv.org/abs/2410.21042
- Code: https://github.com/Keke921/GNM-PT
- Contact: limengke@gml.ac.cn; zbdly226@gmail.com