ICCV #12805

Semi-ViM: Bidirectional State Space Model for Mitigating Label Imbalance in Semi-Supervised Learning

A. Theoretical analysis

SSMixup (@CBgn and 1hky) - We thank the reviewers for suggesting improving the theoretical analysis of SSMixup. In the original submission, we rely on results, (e.g., Table 4), to emphasize the advantages of SSMixup. Here, we provide complementary theoretical analysis to justify the stabilizing and regularizing effects of SSMixup within bidirectional state-space models (SSMs).

Stability via Contractive Bidirectional SSM Dynamics. Let $h_t^f(x_i), h_t^b(x_j)$ denote the forward and backward hidden states of two inputs x_i, x_j, respectively. SSMixup interpolates these via:

$$h_t^{\text{mix}} = \psi h_t^f(x_i) + (1-\psi) h_t^b(x_j), \quad \psi \in [0,1].$$

Assuming standard SSM dynamics $h_{t+1} = Ah_t + Bx_t$ with contractive transition matrix $\|A\|_2 < 1$, the interpolation remains within a bounded attractor set:

$$\|h_t^{\text{mix}}\| \leq \max(\|h_t^f\|, \|h_t^b\|),$$

ensuring dynamical stability and preventing latent explosion from noisy pseudo-labels. Bidirectional modeling further preserves both causal and anti-causal temporal structures in the interpolated dynamics.

Regularization via Lipschitz Continuity and Residual Injection. Let f_{SSM} denote an L-Lipschitz operator. Then the output deviation from SSMixup satisfies:

$$\|y_t^{(i,j)} - y_t^{(j,i)}\| \leq \nu L \|x_i - x_j\|,$$

where ν is the residual injection coefficient (see Eq. 16 in the paper). This Lipschitz-bound implies that SSMixup enforces smoothness in the model's input-output mapping, similar to manifold mixup but extended temporally. The residual path $\nu C h_t^{\text{mix}}$ acts as Tikhonov-style regularization on the latent trajectory, stabilizing gradients and enhancing robustness to pseudo-label noise—particularly when used in conjunction with LyapEMA.

We will revise the submission to include these explanations, which we believe address the reviewers' concern.

B. Additional experiments

LTSSL (@CBgn and xVLc) - We focus on evaluating Semi-ViM on the full version of ImageNet-LT, which represents a more difficult task. Several LTSSL (Long-Tailed Semi-supervised Learning) SOTA methods perform evaluations on its subset, i.e., ImageNet-127. We believe that semi-supervised training on larger datasets will be more beneficial and significant for the future development of models and applications. Table 1 provides further results.

Table 1. Top-1 Accuracy (%) comparison on ImageNet-LT under 1% and 10% label settings. Results for other methods are computed using the same experimental setting as ours.

Method	Backbone	1% Labels	10% Labels
FixMatch w/ ACR [2]	ResNet-50	56.4	61.8
FixMatch w/ SimPro [1]	ResNet-50	57.2	65.5
FixMatch w/ BEM [3]	ResNet-50	58.9	66.2
Semi-ViM-Small (Ours)	ViM-Small	**63.2**	**74.6**
Semi-ViM-Base (Ours)	ViM-Base	**66.3**	**77.4**

LyapEMA (@xVLc and 1hky) - While LyapEMA was introduced in the SSL context, its design—grounded in Lyapunov stability—makes it applicable beyond SSL. LyapEMA dynamically adjusts momentum by stability, supporting robust training under noise, as validated on DeiT (Table 2), LyapEMA adaptively stabilizes updates under noise, improving convergence and accuracy in both supervised and semi-supervised learning.

Regarding tuning complexity, LyapEMA introduces two interpretable hyperparameters: the stability coefficient, λ, and the sensitivity, γ. In practice, we find these parameters to be robust across datasets once roughly calibrated, and their effects are intuitive (e.g., larger λ encourages smoother updates). Moreover, our formulation enables potential future work on heuristic or adaptive scheduling strategies (e.g., EMA warm-up or variance-based λ scaling), which we plan to explore.

Table 2. Top-1 Accuracy (%) of DeiT-Small on ImageNet-1K under full supervision using different update strategies.

Update Strategy	Top-1 Accuracy
EMA	79.8
LyapEMA (Ours)	**81.1**

Overall (@xVLc , CBgn and 1hky) - Thank you to the reviewers for your valuable feedback. We will include a more comprehensive set of LTSSL experiments in the final version. Supplementary material includes experiments under extreme imbalance and out-of-distribution (OOD) settings. We plan to explore the potential of applying LyapEMA to other **domains** and **tasks** as part of our future work.

References

[1] Chaoqun et al. Du. Simpro: A simple probabilistic framework towards realistic long-tailed semi-supervised learning. *arXiv preprint arXiv:2402.13505*, 2024. 1

[2] Tong Wei and Kai Gan. Towards realistic long-tailed semi-supervised learning: Consistency is all you need. In *CVPR*, 2023. 1

[3] Hongwei et al. Zheng. Bem: Balanced and entropy-based mix for long-tailed semi-supervised learning. In *CVPR*, 2024. 1