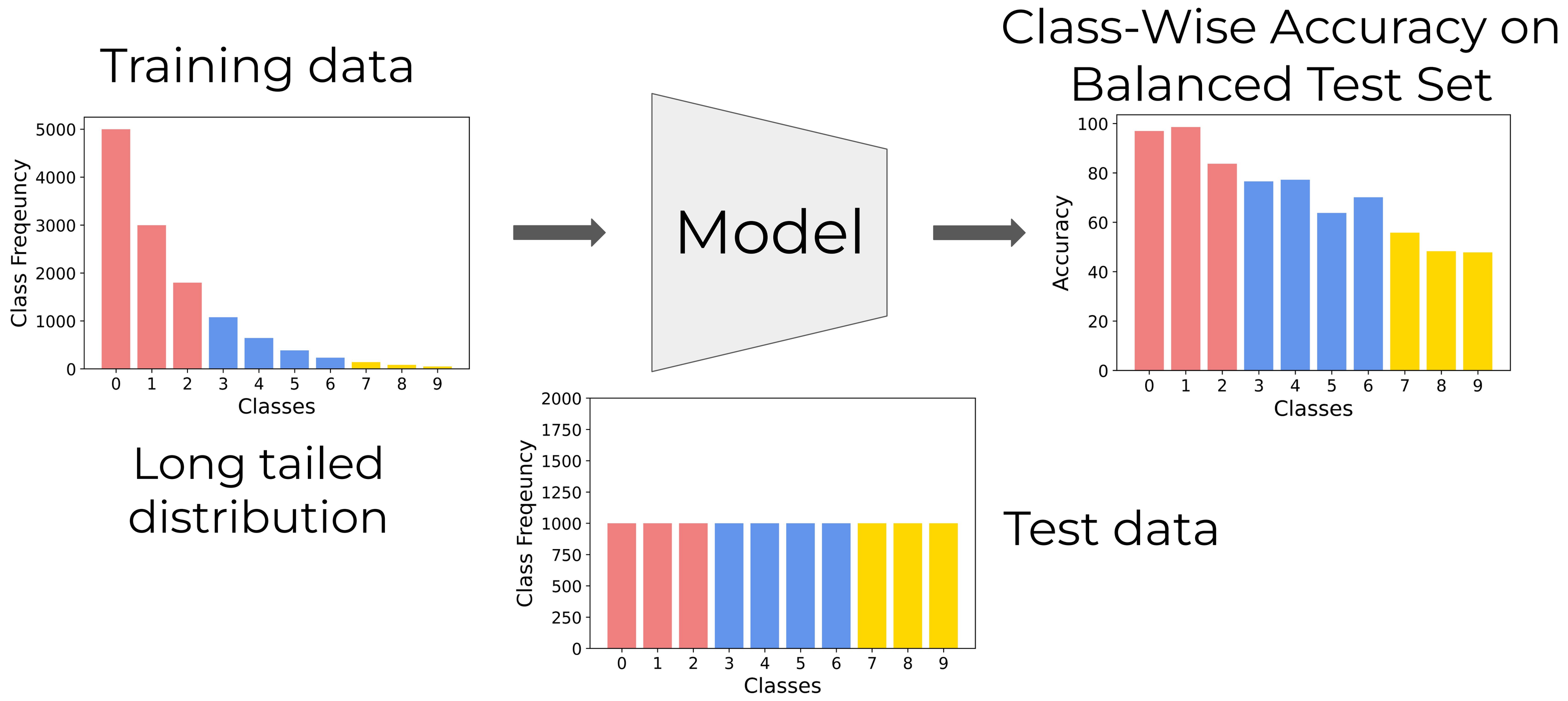
# Escaping Saddle Points for Effective Generalization on Class-Imbalanced Data

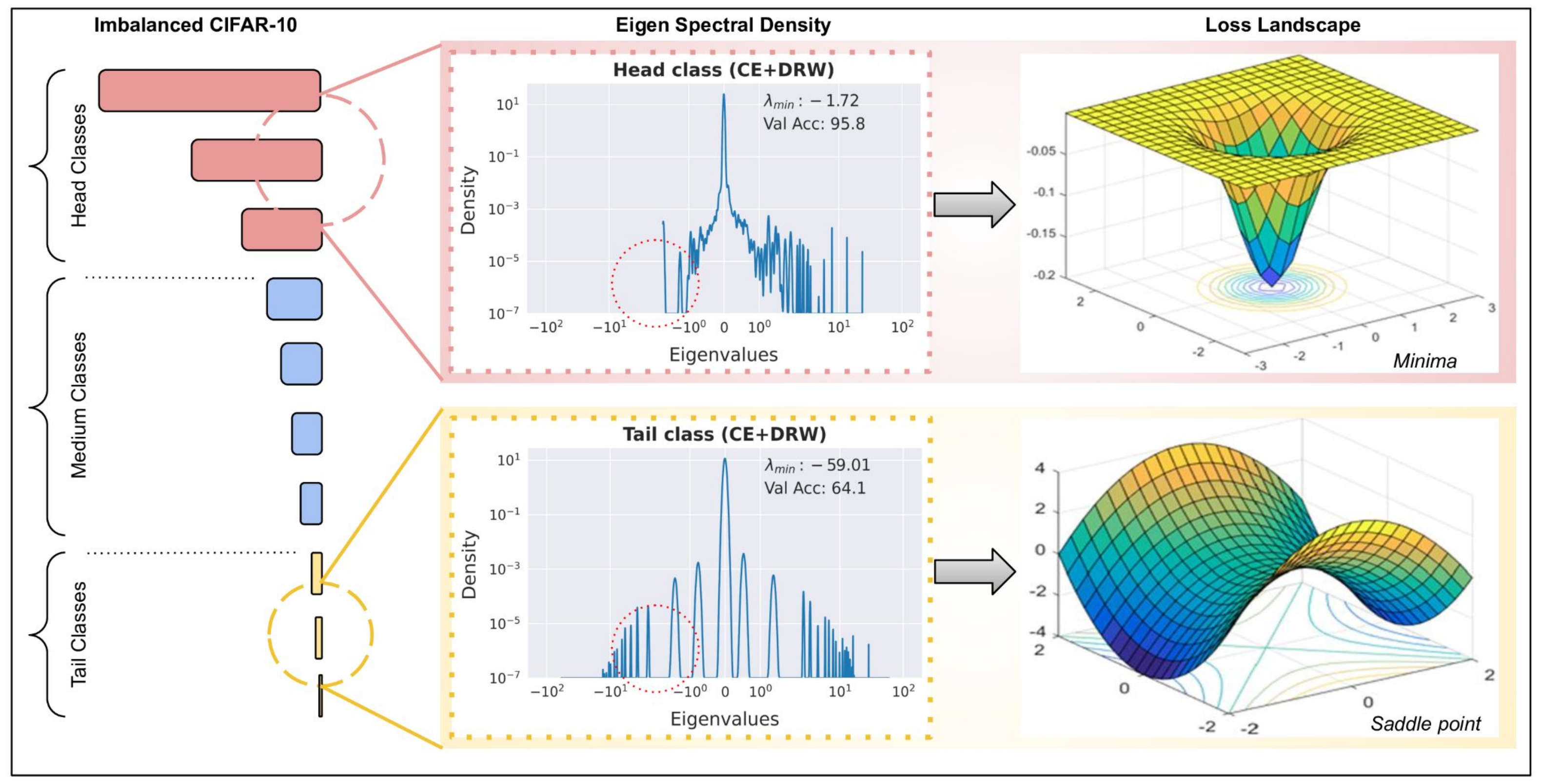
Harsh Rangwani* Sumukh K Aithal* Mayank Mishra R. Venkatesh Babu
Video Analytics Lab, Indian Institute of Science, Bangalore, India

## Motivation




- Real-world datasets are often imbalanced and deep neural networks show poor performance on the samples with few classes (tail class).
- In this work, we focus on analyzing the nature of loss manipulation methods for imbalance datasets.
  - Cross-Entropy+Deferred-Reweighting (CE+DRW)
  - LDAM (Margin Based Loss) [Cao et al. 2019]
  - Vector Scaling Loss (VS) [Kini et al. 2021]

## Convergence to Saddle Points in Tail Class Loss Landscape




- We propose Hessian analysis of per class loss in contrast to overall loss analysis in prior art. The properties of the per class loss landscape (saddle points or minima) can be observed by analyzing eigen spectral density of Hessian (centre).
  - Solution for tail classes reach a region of large negative curvature (high $\lambda_{min}$) indicating convergence to saddle point (bottom),
  - Head classes converge to minima (low $\lambda_{min}$) (top).

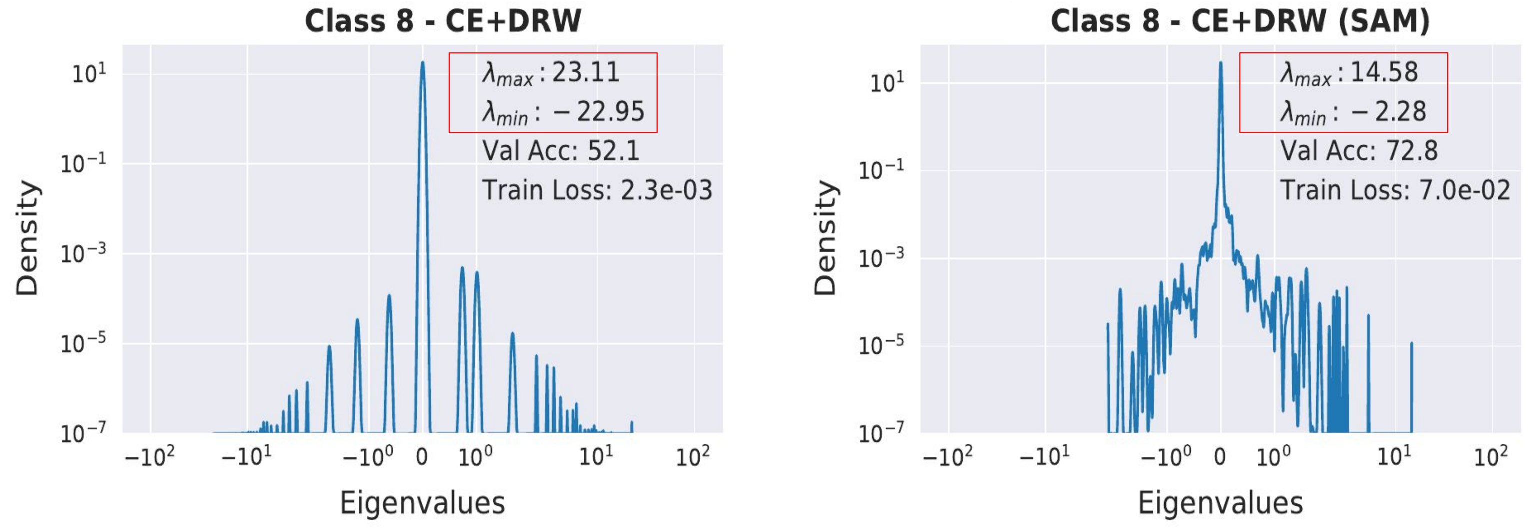
## Escaping Saddle Points with SAM Improves Generalization

**Sharpness Aware Minimization (SAM) [Foret et al. 2021]:** Optimization algorithm which incentivizes convergence to flat minima, given as:

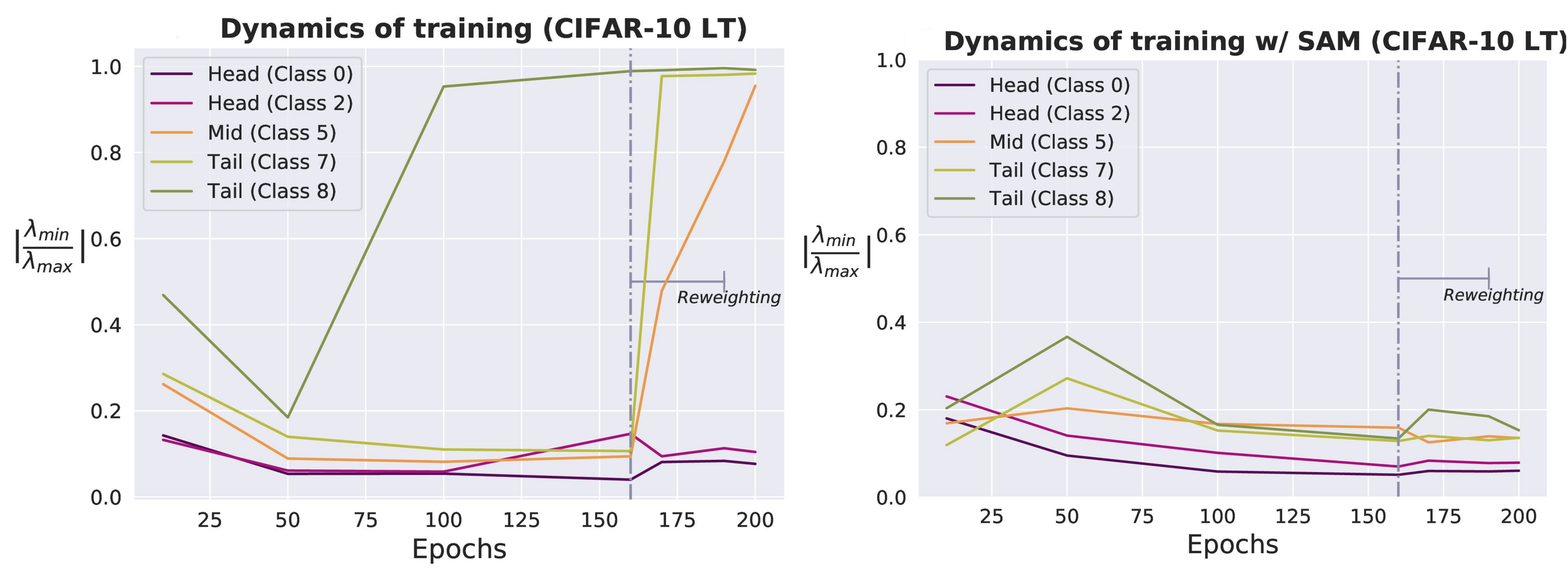
$$\min_{w} \max_{||\epsilon|| \le \rho} f(w+\epsilon) \Rightarrow w' = w - \eta \nabla f\left(w + \rho \frac{f(w)}{|f(w)|}\right)$$

**Theorem 1 (Informal Statement):** We show that using SAM amplifies the component of gradient by a factor of $(1 + \rho\lambda_{min})^2$ in the direction of negative curvature, hence SAM with high $\rho$ can effectively escape saddle points and improve generalization.

**Empirical Validation:** For CIFAR-10 LT tail class, we find that SAM converges to minima with low $\lambda_{min}$ .




## Dynamics of Training on Long-Tailed Datasets




Plot of $(|\lambda_{min}/\lambda_{max}|)$ value for Hessian of the class-wise loss for CIFAR-10 LT. It is observed that:

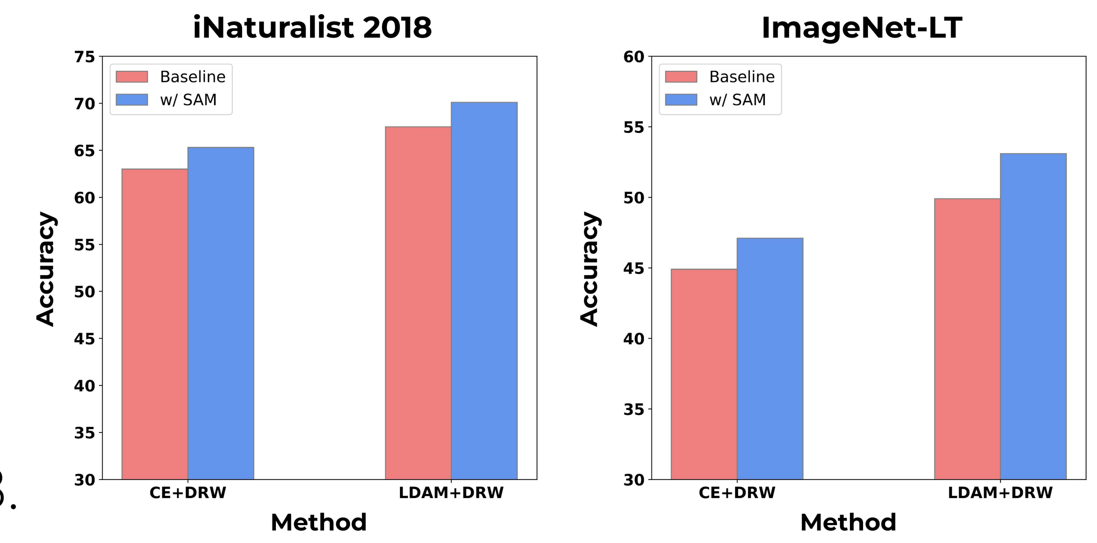
- With re-weighting for tail classes the networks moves to non-convex region with large negative $\lambda_{min}$ and converges to a saddle point.
- With SAM we find that it prevents entry to non-convex region and leads towards a minima.

## Experiments

| | CIFAR-10 LT | | CIFAR-100 LT | |
|---|---|---|---|---|
| | Acc | Tail | Acc | Tail |
| CE | $71.7_{\pm0.1}$ | $52.3_{\pm3.7}$ | $38.5_{\pm0.5}$ | $8.2_{\pm1.0}$ |
| CE + SAM | $73.1_{\pm0.3}$ | $51.7_{\pm1.0}$ | $39.6_{\pm0.6}$ | $8.0_{\pm0.6}$ |
| CE + DRW | $75.5_{\pm0.2}$ | $61.4_{\pm0.9}$ | $41.0_{\pm0.6}$ | $14.7_{\pm0.9}$ |
| CE + DRW + SAM | $80.6_{\pm0.4}$ | $73.1_{\pm0.9}$ | $44.6_{\pm0.4}$ | $20.7_{\pm0.6}$ |
| LDAM + DRW | $77.5_{\pm0.5}$ | $66.4_{\pm0.2}$ | $42.7_{\pm0.3}$ | $19.4_{\pm0.9}$ |
| LDAM + DRW + SAM | $81.9_{\pm0.4}$ | $76.4_{\pm1.1}$ | $45.4_{\pm0.1}$ | $20.8_{\pm0.3}$ |
| VS | $78.6_{\pm0.3}$ | $70.3_{\pm0.5}$ | $41.7_{\pm0.5}$ | $26.8_{\pm1.0}$ |
| VS + SAM | $82.4_{\pm0.4}$ | $78.0_{\pm01.2}$ | $46.6_{\pm0.4}$ | $31.7_{\pm0.1}$ |

Results on long tailed CIFAR 10 & 100 datasets. SAM improves the tail class accuracy for various methods.

SAM improves accuracy on large-scale Imbalanced datasets too: ImageNet-LT & iNaturalist 2018.

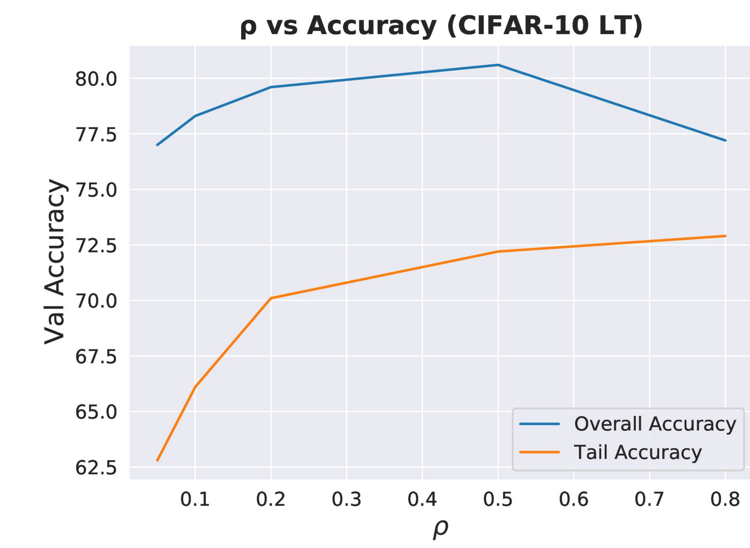



## Analysis

How do other optimization methods for escaping saddle points compare with SAM?

| | CIFAR-10 LT | | CIFAR-100 LT | |
|---|---|---|---|---|
| | Acc | Tail | Acc | Tail |
| CE + DRW | 75.5 | 61.4 | 41.0 | 14.7 |
| + PGD | 77.2 | 65.0 | 42.2 | 17.0 |
| + LPF-SGD | 78.5 | 67.2 | 42.9 | 15.8 |
| + SAM | 80.6 | 73.1 | 44.6 | 20.7 |

PGD: Perturbed Gradient Descent [Jin et al. 2017]
LPF-SGD: Low-Pass Filter SGD [Bisla et al. 2022]




How does neighborhood size of SAM ($\rho$) impact tail accuracy?

*As $\rho$ increases, the component of gradient in the direction of negative curvature increases. This leads to an increase in tail accuracy. (Theorem 1)*

## Acknowledgement

This work was supported in part by SERB-STAR Project (Project: STR/2020/000128), Govt. of India. Harsh Rangwani is supported by PMRF fellowship.

Paper: https://openreview.net/pdf?id=9DYKrsFSU2
Code: https://github.com/val-iisc/Saddle-LongTail