Two-Line Summary. I am a final year PhD student at Indian Institute of Science (IISc), my work focuses on learning from limited and imperfect data. My works on these topics have been published at NeurIPS (x2), CVPR, ECCV, ICCV, ICML, UAI etc.

The availability of large sets of annotated datasets like ImageNet has fueled a lot of progress in Computer Vision and Machine Learning. These datasets have enabled the development of effective deep learning techniques for variety of downstream real-world applications. However, **the distribution of the well-curated datasets is significantly different, from the messy and imperfect data in the real-world**. Hence, a more practical goal for us is to *develop robust algorithms which utilize the limited and imperfect datasets*, to train efficient and effective deep learning models. Towards this goal, in my PhD thesis I have focused on developing algorithms which can lead to generalizable models, despite presence of long-tailed imbalances and distribution shifts between test and training data. Fig. 1 presents an overview of my research works. The first line of work focuses on training Generative Adversarial Networks (GANs) on long-tailed distributions, mitigating issues of mode-collapse and class confusion. The second line of work has focused on optimizing practical metrics like Recall H-mean, Worst-case Recall etc., for the long-tailed imbalanced datasets. These metrics focus on improving the model performance on **tail classes**, compared to accuracy, which treats all classes equally. Further, I have worked on improving generalization by inducing flatness in loss landscape, while learning from imperfect data. Below we describe the works in detail.

Generative Models for Long Tail Data. Generative models are ML models that aim to model the data x distribution $\mathcal{P}(x)$, either in an explicit way by learning the $\mathcal{P}(x)$ (e.g., VAEs, GMMs, etc.) or in an implicit way explicitly by providing ways to sample from $\mathcal{P}(x)$ (i.e., GANs, Diffusion Models, etc.). The implicit models have demonstrated superior photorealism and generality; hence, they are more widely used. In this line of work, we aim to analyze the limits of the generative models, specifically GANs, to learn from practical long-tailed datasets.

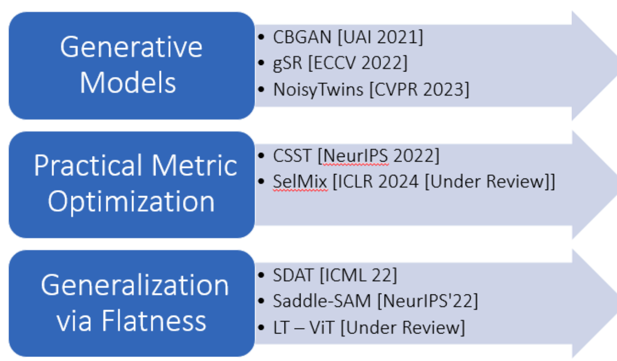

Figure 1: Summary of Works done in PhD

Our empirical analysis found that GANs don't model long-tailed categories well and suffer from mode-collapse on tail categories. As there had already been a plethora of work on training discriminative classifiers on the long-tailed data, we aimed to leverage that to guide the generation of generative models. While training, we pass the generated images to a classifier and keep an estimate of the class statistics. We use this estimate to regularize the generations from GAN, by proposing a theoretically principled weighted entropy regularizer. The regularized Class Balancing GAN (CBGAN) [1] leads to balanced generations for diverse datasets. This broad idea of using *classifier to guide generations of Generative Models* was also used for seminal Diffusion [2] model work. To further mitigate the issue of mode collapse fundamentally, we took a closer look at the reasons for it. We discovered that the spectral norm of the conditioning parameters were accumulating extra gradients, which when regularized through inexpensive group spectral norm regularizer (gSR) [3] made GAN training stable and secure on long-tailed datasets.

We found out that Spectral Norm Issue did solve the problem for cGANs, but due to differences in architecture our results didn't scale to SotA StyleGANs. To improve the conditioning further, in our CVPR'23 work NoisyTwins [4] we enforce explicit structure in conditioning latent space where we parameterized each class regions with a Gaussian. To enforce structure in latent space we introduced contrastive learning based technique to form distinct clusters for each class generations, which significantly improve the generation and mitigate the mode collapse by first principle construction.

Optimizing Practical Metrics via Feedback. Modern deep models have been commonly trained with common loss functions such as Cross-Entropy loss, which are surrogates for the average classifier accuracy. However, for the deep models to be deployed in practice, metrics like worst-case or H-mean of recall across classes are well suited as they holistically evaluate the model. Further, the trained models should also satisfy constraints to ensure *Fairness* in Decision making. Optimizing these practical metrics has been largely studied for linear models like SVM in supervised learning. However, modern deep networks currently operate in a semi-supervised setting where a large amount of unlabeled data (to learn representations) and a small amount of labeled data (to perform specific tasks) is used. In our work CSST [5], we introduce the first framework to optimize practical non-decomposable metrics (e.g., Worst-Case Recall, etc.) in a semi-supervised setup, providing rigorous theoretical guarantees for the consistency of the framework. However, this framework requires neural networks to be trained from scratch for each objective. As training from scratch is expensive, we introduce a framework SelMix [6] to fine-tune a pre-trained (a.k.a. foundational) model to optimize for a given non-decomposable objective. The SelMix framework combines the empirical effectiveness of Mixup [7] framework, with theoretical guarantees for linear classifiers. In SelMix, we selectively mixup samples of classes based on an optimal sampling distribution such that it optimizes the desired non-decomposable objective. The SelMix framework comes with theoretical guarantees and in practice, leads to SotA results on even ImageNet, demonstrating its principled nature and generic applicability.

Inducing Flatness for Generalization on Imperfect Data. Often fitting the model on a certain objective is often easy for overparameterized models, even on imperfect data. The real test is a generalization, for which I have closely analyzed modern deep networks trained on long-tailed datasets. In our NeurIPS'22 work [8], we discovered that re-weighting (cost-sensitive) loss techniques lead to convergence to saddle points in the loss landscape, which hinders generalization. We demonstrate that escaping saddles and converging to flat minima improves generalization and that Sharpness Aware Optimization (SAM) [9] is the most effective method for escaping saddle points in long-tailed cases. In our ICML'22 work, we analyze a similar issue of generalization on the target domain (in domain adaptation), using SAM. We show that SAM leads to sub-optimal performance in min-max (adversarial) optimization setups, and generalization improvement is only obtained by applying SAM (i.e., inducing flatness) for min (or max) optimization. This work has been impactful and is a backbone for many SotA follow-up works [10, 11]. Further, our recent submission demonstrates that

CNN models with SAM (i.e., flat models) can be used for effective knowledge distillation to large-scale ViT (Vision Transformers), with improved generalization on long-tailed data.

I have also worked on inducing flatness in Federated Learning (WACV'24) and using active learning to select samples for domain adaptation (ICCV'21).

1 Future Work

I want to develop principled methods that have a grounding in real-world applications. Towards this goal, I want to study theoretical ideas on learning and consider how those principled ideas can be used in the real world. I believe it's important to learn ML models aligned with humans to improve their real-world usability. Toward this goal, the following are some research directions that I want to study and explore, both in terms of theory and practice, in the following years :

Non-Decomposable (Human) Feedback for Generative Models. In this line of work, we aim first to develop metrics that capture the human preferences for the model. However, using RL and naive human feedback might not be very sample-efficient. Hence, we aim to develop metrics and ML models that are aligned with humans and can also provide feedback about the human alignment of the generative model's output. We aim to study the developed metrics (e.g., FID [12]) and properties of the human-aligned models, and then how to exploit their feedback to improve the generations. One concrete problem here is improving the distribution of images from a generative model in the context of fairness. I want to explore the non-decomposable and bandit optimization literature and use those methods to develop feedback mechanisms for generative models.

Compositional Generalization for Long-Tailed (Few-Shot) Data. One of the generative models' important features is their capability to generalize on tail categories and their emergent few-shot performance. This behavior results from the model's ability to compose the knowledge from other tasks to identify tail categories with only a few samples. For example, let's say we give the model a task to generate a rare species like a blue hummingbird; in such a case, if the model has learned a generic representation (e.g., parts) of a common hummingbird, it can just do some specific modifications to tailor it to a rare hummingbird. However, most theoretical works have only focused on analyzing the i.i.d. (independently and identically distributed) case. Hence, there is a requirement to re-think generalization by keeping the compositional properties in mind. Further, in empirical works, there is scope to study the compositional properties of the generative models, which I am to pursue diligently. In this process, I also aim to uncover ways to utilize the compositional properties of these models for creative and novel generations.

My experience with Generative Models and my work on Non-Decomposable Metrics have equipped me with the required skills to progress in the above areas. Further, my experience as a research intern at Adobe has prepared me to tackle generative models on a large scale. I firmly believe that these experiences will help me make a significant impact while working on the above-mentioned research areas.

References

[1] Harsh Rangwani, Konda Reddy Mopuri, and R Venkatesh Babu. Class balancing gan with a classifier in the loop. In *Uncertainty in Artificial Intelligence*, pages 1618–1627. PMLR, 2021.

[2] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[3] Harsh Rangwani, Naman Jaswani, Tejan Karmali, Varun Jampani, and R Venkatesh Babu. Improving gans for long-tailed data through group spectral regularization. In *European Conference on Computer Vision*, pages 426–442. Springer, 2022.

[4] Harsh Rangwani, Lavish Bansal, Kartik Sharma, Tejan Karmali, Varun Jampani, and R Venkatesh Babu. Noisytwins: Class-consistent and diverse image generation through stylegans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5987–5996, 2023.

[5] Harsh Rangwani, Sho Takemori, Kato Takashi, Yuhei Umeda, et al. Cost-sensitive self-training for optimizing non-decomposable metrics. *Advances in Neural Information Processing Systems*, 35:26994–27007, 2022.

[6] Shrinivas Ramasubramanian, Harsh Rangwani, Sho Takemori, Kunal Samanta, Yuhei Umeda, and Venkatesh Babu Radhakrishnan. Selmix: Selective mixup fine tuning for optimizing non-decomposable metrics. In *ICML 2023 Workshop on Differentiable Almost Everything: Differentiable Relaxations, Algorithms, Operators, and Simulators*, 2023.

[7] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.

[8] Harsh Rangwani, Sumukh K Aithal, Mayank Mishra, et al. Escaping saddle points for effective generalization on class-imbalanced data. *Advances in Neural Information Processing Systems*, 35:22791–22805, 2022.

[9] Pierre Foret, Ariel Kleiner, Hossein Mobahi, and Behnam Neyshabur. Sharpness-aware minimization for efficiently improving generalization. *arXiv preprint arXiv:2010.01412*, 2020.

[10] YiFan Zhang, Jian Liang, Zhang Zhang, Liang Wang, Rong Jin, Tieniu Tan, et al. Free lunch for domain adversarial training: Environment label smoothing. In *The Eleventh International Conference on Learning Representations*, 2022.

[11] Lukas Hoyer, Dengxin Dai, Haoran Wang, and Luc Van Gool. Mic: Masked image consistency for context-enhanced domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11721–11732, 2023.

[12] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.