Learning More from Less: Resource-Constrained Generative AI for Classification, Generation, and Personalization

Aniket Roy
Johns Hopkins University

Abstract

The rapid advancement of generative models has created new opportunities for addressing core challenges in computer vision, including data scarcity, image quality, and efficient personalization. My research develops principled, resource-aware methods that enable models to generalize effectively from limited supervision, adapt efficiently to new concepts, and generate high-fidelity visual content. I first address few-shot learning through augmentation-driven uncertainty-guided mixup, improving robustness in data-constrained regimes. Building on this, I propose caption-guided multi-modal augmentation techniques that enrich visual diversity while mitigating real-to-synthetic domain gaps. To enhance the quality and realism of generated images, I introduce diffusion models grounded in natural image statistics, yielding perceptually aligned outputs suitable for downstream tasks. To advance personalization, I develop parameter-efficient mechanisms for combining low-rank adapters, enabling fine-grained control over content and style without retraining. I further extend personalization to a zero-shot setting through a training-free textual-inversion-based method that customizes arbitrary objects directly within the diffusion process. Finally, I present a frequency-guided multi-LoRA fusion framework that leverages wavelet-domain cues and timestep-aware weighting for accurate, training-free concept composition. Collectively, these contributions move toward a unified vision of generative models that are efficient, adaptive, and capable of high-quality, customizable image synthesis.

Introduction. The rapid advancement of generative models (Roy et al. 2023; Pal et al. 2024; Roy et al. 2024) has opened new avenues for addressing critical challenges in computer vision (Dhar et al. 2021; Fazlyab et al. 2023), such as data scarcity, image quality enhancement, and personalization. Recent progress has concentrated on improving the adaptability, efficiency, and quality of these models to meet the growing demand for parameter-efficient fine-tuning and adaptation of large vision-language and generative models (Roy et al. 2025b; Pramanick, Roy, and Patel 2022). In this work, we begin by tackling the challenges of resource-constrained learning (Roy et al. 2022). We then leverage powerful vision-language models to address these issues in a parameter-efficient manner. Additionally, we aim to enhance state-of-the-art generative models—specifically diffusion models—by incorporating natural image priors (Roy et al. 2023). We also explore joint concept merging through the lens of low-rank adapter merging, applying it to content-style personalization. Finally, we address the challenge of zero-shot personalization of any object without requiring additional training. We conclude by devising a frequency-guided method for training-free multi-LoRA composition, which is more appropriate for deployment on edge devices.

Few-shot learning. Few-shot learning poses a fundamental challenge: effectively learning from limited data while maintaining generalization. FeLMi (Roy et al. 2022) introduces an augmentation-guided solution through uncertainty-guided hard mixup. By generating synthetic data points that mix novel and base class samples using entropy-based uncertainty metrics, FeLMi effectively augments the training distribution. This strategy enhances model robustness, particularly in low-data scenarios, and provides a foundation for developing more reliable few-shot learning frameworks. The method exemplifies the potential of data augmentation to bridge the gap between data scarcity and model performance. FeLMi is published in **NeurIPS'22**.

Generative augmentation. Extending the idea of augmentation (Shah et al. 2023), Cap2Aug (Roy et al. 2025b) integrates textual captions into the generative process to enable more semantically diverse and contextually rich synthetic data. By leveraging caption-guided diffusion models, Cap2Aug not only augments visual data but also aligns it with linguistic information, creating a multimodal approach to data enrichment. This strategy proves particularly effective for imbalanced and long-tail datasets, where the combination of real and synthetic data enhances feature diversity and model discrimination capabilities. It also mitigate the real-to-synthetic domain gap using a Maximum Mean Discrepancy (MMD) loss. Cap2Aug illustrates the transformative potential of merging textual and visual modalities to address data limitations. Cap2Aug is published in **WACV'25**.

Image quality improvement of generative models. While FeLMi and Cap2Aug focus on augmenting and enriching data, DiffNat (Roy et al. 2023) shifts the focus to enhancing the inherent quality of generated images. By grounding diffusion models in natural image statistics, DiffNat introduces statistical priors that guide the generative process toward producing visually realistic outputs. This alignment with natural image distributions not only improves perceptual quality but also ensures that the generated content meets the standards of high-quality applications, e.g., personalization, unconditional and super-resolution image generation. DiffNat addresses the critical demand for reliability and aesthetics in generative models, paving the way for their adoption in sensitive and quality-focused domains. DiffNat is published in **TMLR'25**.

Content-style personalization. As generative models grow in complexity, the ability to efficiently adapt them to new concepts without extensive retraining becomes a critical requirement. DuoLoRA (Roy et al. 2025a) introduces a parameter-efficient mechanism by combining multiple low-rank adaptation (LoRA) modules. This dual adaptation strategy allows for precise customization of pretrained diffusion models, balancing computational efficiency with fine-grained control over style and content. DuoLoRA's approach exemplifies how adaptable and lightweight methodologies can democratize the use of generative models, particularly in creative and personalized applications. DuoLoRA is published in **ICCV'25**.

Zero-shot personalization. Recent text-to-image diffusion models excel at generating high-quality images, but adapting them quickly for custom content—especially beyond human subjects—remains challenging. Existing identity-based methods work well for people but falter on arbitrary objects. We propose a novel, training-free approach that uses textual inversion to learn object-specific embeddings and injects them directly into the diffusion UNet timesteps (Roy, Suin, and Chellappa 2025b). This enables rapid, text-conditional customization of a wide range of objects with minimal overhead. Our extensive evaluations demonstrate that this method is both fast and flexible, effectively filling a critical gap in image generation by supporting inclusive, high-fidelity customization across diverse object categories.

Training-free multi-LoRA fusion. Low-Rank Adaptation (LoRA) has become a compute-efficient way to fine-tune generative models for precise control over visual attributes, but existing methods struggle to combine multiple pretrained LoRAs into a single composite image without additional training. Prior approaches—like averaging or alternating LoRA scores—fail to enforce spatial locality (e.g., placing a dress on a person) or adaptively weight each concept, resulting in visual artifacts and misplaced elements. Noting that high-frequency details emerge earlier in diffusion timesteps and low-frequency details later, we introduce MultLFG (Roy, Suin, and Chellappa 2025a): a training-free framework that fuses LoRAs in the wavelet frequency domain with pixel-wise attention for spatial coherence and adaptive score weighting per timestep. This frequency-aware, localization-driven composition consistently outperforms existing methods on the ComposLoRA benchmark.

Conclusion. In conclusion, the rapid evolution of generative models has provided fertile ground for tackling key challenges in computer vision—from addressing data scarcity and enhancing image quality to enabling seamless personalization. This work advanced the landscape by first resolving the constraints of resource-limited learning, then leveraging vision-language models for more parameter-efficient adaptation. Additionally, we improved state-of-the-art diffusion models by integrating natural image priors and explored low-rank adapter merging for joint concept personalization. Finally, we demonstrated zero-shot personalization of arbitrary objects without additional training and training-free multiLoRA composition. Collectively, these contributions move us closer to truly versatile, efficient, and adaptive generative models in computer vision.

References

Dhar, P.; Gleason, J.; Roy, A.; Castillo, C. D.; and Chellappa, R. 2021. Pass: protected attribute suppression system for mitigating bias in face recognition. In *ICCV*.

Fazlyab, M.; Entesari, T.; Roy, A.; and Chellappa, R. 2023. Certified robustness via dynamic margin maximization and improved lipschitz regularization. *NeurIPS*.

Pal, B.; Roy, A.; Kathirvel, R.; O'Toole, A.; and Chellappa, R. 2024. DiversiNet: Mitigating Bias in Deep Classification Networks across Sensitive Attributes through Diffusion-Generated Data. In *IJCB*. IEEE.

Pramanick, S.; Roy, A.; and Patel, V. M. 2022. Multimodal learning using optimal transport for sarcasm and humor detection. In *WACV*.

Roy, A.; Borse, S.; Kadambi, S.; and Das, D. 2025a. DuoLoRA : Cycle-consistent and Rank-disentangled Content-Style Personalization. *arXiv*.

Roy, A.; Roy, A.; Mitra, S.; and Ghosh, K. 2024. Bri3l: A brightness illusion image dataset for identification and localization of regions of illusory perception. In *ICIP*. IEEE.

Roy, A.; Shah, A.; Shah, K.; Dhar, P.; Cherian, A.; and Chellappa, R. 2022. FeLMi : Few shot Learning with hard Mixup. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Roy, A.; Shah, A.; Shah, K.; Roy, A.; and Chellappa, R. 2025b. Cap2aug: Caption guided image to image data augmentation. *WACV*.

Roy, A.; Suin, M.; and Chellappa, R. 2025a. MultLFG: Multi-LoRA fusion using frequency-domain guidance. *arXiv*.

Roy, A.; Suin, M.; and Chellappa, R. 2025b. Zero-shot customizing of objects via textual inversion. *arXiv*.

Roy, A.; Suin, M.; Shah, A.; Shah, K.; Liu, J.; and Chellappa, R. 2023. Diffnat: Improving diffusion image quality using natural image statistics. *arXiv preprint arXiv:2311.09753*.

Shah, A.; Roy, A.; Shah, K.; Mishra, S.; Jacobs, D.; Cherian, A.; and Chellappa, R. 2023. Halp: Hallucinating latent positives for skeleton-based self-supervised learning of actions. In *CVPR*.