Domain Adaptation for Fair and Robust Computer Vision.

Tarun Kalluri (sskallur@ucsd.edu)

Overview In recent years, the field of computer vision and deep learning has been propelled by the emergence of large-scale foundational models, which unlock remarkable capabilities in various tasks - from scene understanding and robot navigation to text-to-image synthesis and nuanced multimodal dialogue. However, their reliance on uncurated, web-sourced data presents new challenges, where the biases in training data can lead to unfair outcomes for under-represented subgroups and their lack of robustness outside their training domain limits their universal adoption. For instance self-driving technologies would significantly improve mobility and road-safety in high traffic-density geographies like Asia and Africa, but most benchmark datasets are instead collected from US or Europe with little to no representation from other countries, with notable domain gaps preventing robust transfer across geographies. Thus, my doctoral research focuses on improving the generalizability of vision models across under-represented domains in the real world. My past research successfully proposed scalable solutions for unsupervised domain adaptation Kalluri et al. (2022) and video frame interpolation Kalluri et al. (2023a), highlighted the severe limitation of current models in showcasing geographical robustness owing to biased training datasets Kalluri et al. (2023c) and devised novel algorithms to improve visual perception in an open-world Kalluri et al. (2023b). My future research is aimed at analyzing the fairness properties of emerging generative AI technology and leveraging large-scale multimodal foundational models to improve test-time robustness in computer vision tasks. My PhD research achieved prestigious recognitions, including a WACV'23 best paper finalist and the IPE PhD fellowship in 2021.

Research Highlights: Large-Scale Datasets and Scalable Solutions for Domain Adaptation
A major impediment to research progress in geographical fairness is the lack of suitable benchmarks informing the geographical sensitivity of existing methods. To address this limitation, I led the efforts in creating a large scale dataset and evaluation benchmark called GeoNet Kalluri et al. (2023c) dedicated to study geographical disparities on standard vision tasks. We analyze several salient properties of geographic adaptation, and highlight the limitations of several modern algorithms in bridging geographic domain gaps. We posit that GeoNet would not only provide researchers the opportunity to assess the suitability of state-of-the-art algorithms towards universal deployment, but also inspire design of robust AI models that can efficiently handle dynamic changes in geographies while maintaining superior performance. I also led the organization of a successful workshop and challenge based on the dataset at ICCV 2023.

As we move to real world adaptation scenarios, practical datasets invariably consist of plentiful categories introducing new challenges like smaller inter-class discriminability. In MemSAC Kalluri et al. (2022), we proposed and theoretically justified a novel variant of the contrastive loss, along with a memory augmented approach, to improve discriminative transfer across challenging domain shifts. Our work efficiently handles arbitrary number of classes with minimal negative alignment setting new state-of-the-art on challenging datasets like DomainNet and CUB-200, which still holds till date. Additionally, I also proposed a novel framework for universal semantic segmentation that allows joint deployment of single segmentation model across road scenes from diverse geographies while using very few labeled data from each, meeting the much desired dual needs of lower annotation and deployment costs Kalluri et al. (2019).

Alongside these fundamental innovations, my past research also proposed practical solutions for challenging applications like video frame interpolation Kalluri et al. (2023a), where our novel architecture inspired by 3D convolutions and training recipe based on large-scale, unlabeled video data yielded unprecedented generation quality while being upto 6x faster than all existing methods.

Future Vision: Trustworthy Foundational AI Models
I plan to focus my future research on two pivotal thrusts in the future. Firstly, I aim to explore the fairness properties of emerging generative AI applications, particularly Large Language Models (LLMs) and text-to-image (T2I) models. Notably, these models often face challenges in delivering optimal performance across low-resource domains like low and mid-income societies, presenting an open-challenge in extending their applicability to diverse populations. Secondly, I aspire to harness the recent advancements in foundational models towards enhancing the robustness capabilities in diverse applications. These models trained on multimodal, web-scale datasets showcase strong zero-shot and emergent intelligence capabilities which can drive progress in out-of-distribution generalization on several downstream tasks. This dual-pronged approach seeks to offer novel insights into the inclusivity challenges posed by the rapid progress in generative AI while providing practical solutions for the widespread adoption of these models in the future.

In summary, my research focuses on developing data-efficient algorithms to enhance test-time robustness and fairness of computer vision models, and my longer term goal is to make emerging AI technology universally deployable and more widely accessible.

References

Kalluri, T., Pathak, D., Chandraker, M., and Tran, D. (2023a). Flavr: Flow-agnostic video representations for fast frame interpolation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2071–2082.

Kalluri, T., Sharma, A., and Chandraker, M. (2022). Memsac: Memory augmented sample consistency for large scale domain adaptation. In *European Conference on Computer Vision*, pages 550–568. Springer.

Kalluri, T., Varma, G., Chandraker, M., and Jawahar, C. (2019). Universal semi-supervised semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5259–5270.

Kalluri, T., Wang, W., Wang, H., Chandraker, M., Torresani, L., and Tran, D. (2023b). Open-world instance segmentation: Top-down learning with bottom-up supervision. *arXiv preprint arXiv:2303.05503*.

Kalluri, T., Xu, W., and Chandraker, M. (2023c). Geonet: Benchmarking unsupervised adaptation across geographies. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15368–15379.