ARIA

APPLIED RESEARCH IN ACTION

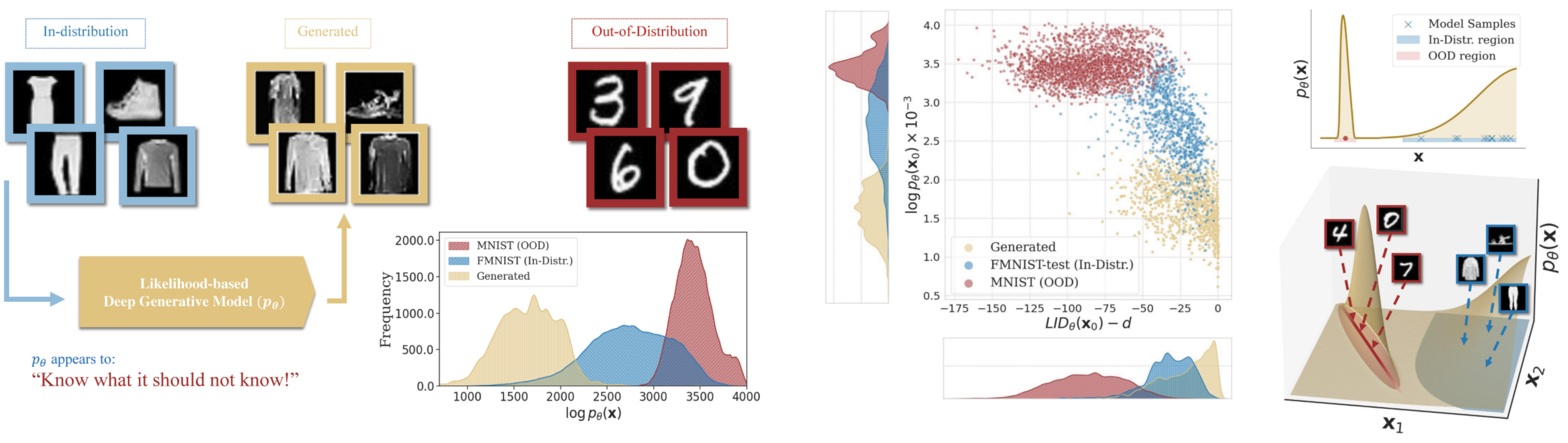
Explaining the Out-of-Distribution Detection Paradox through Likelihood Peaks

Explaining a popular paradox in deep generative models from a manifold learning perspective, leading to enhanced reliability of generative models in detecting out-of-distribution data.

Hamid Kamkari

Rahul G. Krishnan
ACADEMIC SUPERVISOR

Jesse C. Cresswell, Anthony L. Caterini, and Gabriel Loaiza-Ganem
INDUSTRY SUPERVISORS



PROJECT SUMMARY

Likelihood-based deep generative models (DGMs) commonly exhibit a puzzling behaviour: when trained on a relatively complex dataset, they assign higher likelihood values to out-of-distribution (OOD) data from simpler sources. Adding to the mystery, OOD samples are never generated by these DGMs despite having high likelihoods. This two-pronged paradox has yet to be conclusively explained, making likelihood-based OOD detection unreliable. Our primary observation is that high-likelihood regions will not be generated if they contain minimal probability mass, which can occur if the density is sharply peaked. We demonstrate how this seeming contradiction of large densities yet low probability mass can occur on data confined to low dimensional manifolds. We also show that this scenario can be identified through local intrinsic dimension (LID) estimation, and propose a method for OOD detection which pairs the likelihoods and LID estimates obtained from a pre-trained DGM. Moreover, we provide an efficient method for estimating LID from a normalizing flow model, improving upon existing estimators, and enabling state-of-the-art OOD detection performance with respect to comparable flow-based benchmarks.

REFERENCES

1. Nalisnick, Eric, et al. "Do Deep Generative Models Know What They Don't Know?" International Conference on Learning Representations. 2019.
2. Ren, Jie, et al. "Likelihood Ratios for Out-of-distribution Detection." Advances in Neural Information Processing Systems. Vol. 32. 2019.
3. Kirichenko, Polina, Izmailov, Pavel, and Wilson, Andrew G. "Why Normalizing Flows Fail to Detect Out-of-distribution Data." Advances in Neural Information Processing Systems. Vol. 33. 2020.
4. Serrà, Joan, et al. "Input Complexity and Out-of-distribution Detection with Likelihood-based Generative Models." International Conference on Learning Representations. 2020.
5. Tempczyk, Piotr, et al. "LIDL: Local Intrinsic Dimension Estimation Using Approximate Likelihood." International Conference on Machine Learning. PMLR, 2022. pp. 21205-21231.
6. Yoon, Sangwoong, Noh, Yung-Kyun, and Park, Frank. "Autoencoding Under Normalization Constraints." International Conference on Machine Learning. PMLR, 2021. pp. 12087-12097.

layer 6 AI at TD

Computer Science UNIVERSITY OF TORONTO | Master of Science in Applied Computing