Research Statement - Jiaru Zhang

My doctoral research mainly focused on the **robustness, generality and interpretation** of **Bayesian Neural Networks (BNNs), which** are particular neural networks where the parameters are modeled as probabilistic distributions instead of constants.

1. In the **CVPR2021** paper titled *Robust Bayesian Neural Networks by Spectral Expectation Bound Regularization*, I introduced a generalized Lipschitz constraint to enhance the **adversarial robustness** of BNNs.

2. In the **AAAI2022** paper titled *Improving Bayesian Neural Networks by Adversarial Sampling*, I proposed a method called Adversarial Sampling to mitigate the negative impact of sampling randomness during training and inference in Bayesian neural networks.

3. In the **ECAI2023** paper titled *Information Bound and its Applications in Bayesian Neural Networks*, I introduced the concept of Information Bound as a metric to estimate the amount of information in BNNs and showcased its various applications.

Recently, I started to conduct **research about Diffusion Models (DMs)**. Concretely,

1. I participated in two projects exploring the **trustworthy problems in DMs**. In the **ICML2023 oral** paper titled *Adversarial Example Does Good: Preventing Painting Imitation from Diffusion Models via Adversarial Examples*, we proposed to apply adversarial attacks on DMs to protect images from being imitated and copied. In the **CVPR 2024** paper titled *CGI-DM: Digital Copyright Authentication for Diffusion Models via Contrasting Gradient Inversion*, we proposed a new framework to authenticate digital copyrights and thus validate infringements in DMs.

2. In my recent work titled *Exploring Diffusion Models' Corruption Stage in Few-Shot Fine-tuning and Mitigating with Bayesian Neural Networks,* I proposed to apply BNNs to mitigate a critical *corruption* problem in DM fine-tuning.

Furthermore, during my **internship at Microsoft Research Asia (MSRA)**, I had the opportunity to explore the utilization of deep neural networks for solving **causal discovery** problems.

- During the internship, I finished a paper titled *Supervised Causal Learning of Identifiable Causal Structures*. It introduces a novel DNN-based supervised casual learning approach by modeling pairwise features and learning identifiable casual structures.

Moreover, I have a rich research experience in other fields, including approximate nearest neighbor search, tensor-based feature fusion in neural networks, and energy cost in deep learning. Currently I am also learning the knowledge about LLMs, and I would like to research on LLMs in the future.