

YUPU YAO

yupuyao.github.io

(+86) 18181918616 $\diamond$ yupu.yao@psu.edu

EDUCATION

Pennsylvania State University

Ph.D. in Computer Science and Engineering

Aug 2026 - Expected June 2031

University of Electronic Science and Technology of China

B.Eng. in Communication Engineering

Sept 2022 - June 2026

RESEARCH EXPERIENCE

Deconstructing Attention Mechanisms via Implicit Hilbert Ladder *Feb 2024 - June 2024*

- Proposed and implemented the Implicit Hilbert Ladder (IHL) framework, modeling Multi-Head Self-Attention (MSA) from a Reproducing Kernel Hilbert Space (RKHS) perspective to provide a new theoretical understanding of Transformers.
- Developed a novel mechanism, Hilbertian Attention Cascade (HAC), replacing self-attention with kernel mappings in RKHS. Achieved a $>5\%$ accuracy increase in early-stage training on CIFAR-10, with 16% fewer parameters and a 12.3% reduction in GFLOPs.
- Validated the theoretical advantages of IHL via Rademacher complexity generalization bounds, Neural Tangent Kernel (NTK) convergence proofs, and gradient flow analysis, establishing a theoretically-grounded attention alternative

Resource Allocation for Semantic-Aware Communication Networks *Aug 2022 - Feb 2024*

- Deployed and trained a Deep Q-Network (DQN)-based intelligent resource allocation algorithm, improving average user Quality of Experience (QoE) by 23.5% and system convergence speed by 18% in simulated wireless networks.
- Proposed a distributed two-tier control strategy integrating local DQNs with a global coordination module, reducing training communication overhead by 28.6% while maintaining near-optimal global performance.
- Effectively addressed the joint task of beamforming, power control, and semantic resource allocation, demonstrating a 19% improvement in average policy robustness under multi-task, high-dimensional state spaces.

VQ-GAN based Semantic Communication System

Dec 2022 - Mar 2023

- Proposed and implemented a novel end-to-end semantic communication architecture, introducing a VQ-GAN encoding module to replace traditional source-channel separation and achieve joint optimization of compression and representation.
- Improved semantic reconstruction accuracy by 21.3% and bandwidth efficiency by 35.7% compared to conventional systems.

PUBLICATIONS AND MANUSCRIPTS

[1] **Yupu Yao** . Generalize neural network through smooth hypothesis function. *ICLR 2024(Tiny)*

[2] **Yupu Yao**, Bowen Yang. GeoPE: A Unified Geometric Positional Embedding for Structured Tensors. *In Preprint*

[3] **Yupu Yao**, Shangqi Deng, Zihan Cao, Harry Zhang, Liang-jian Deng. APLA: Additional perturbation for latent noise with adversarial training enables consistency. *In Preprint*

SERVICE

Conference Reviewer

NeurIPS

2026

TECHNICAL SKILLS

- **ML/DL Frameworks:** Pytorch, Tensorflow, JAX, Triton
- **Training Frameworks:** DeepSpeed
- **Inference Frameworks:** SGLang, Vllm