Toward Untethered Extended Reality Through Wireless Sensing and Communications Co-design

Mingzhe Chen (University of Miami) and Shiwen Mao (Auburn University).

ECCS-2434054 (UM), ECCS-2434053 (AU).

Project Overview

Project Goal: To develop a holistic wireless XR framework that utilizes mmWave for joint XR user movement detection and XR data transmission while satisfying the joint communication, computing, sensing, and XR service requirements.

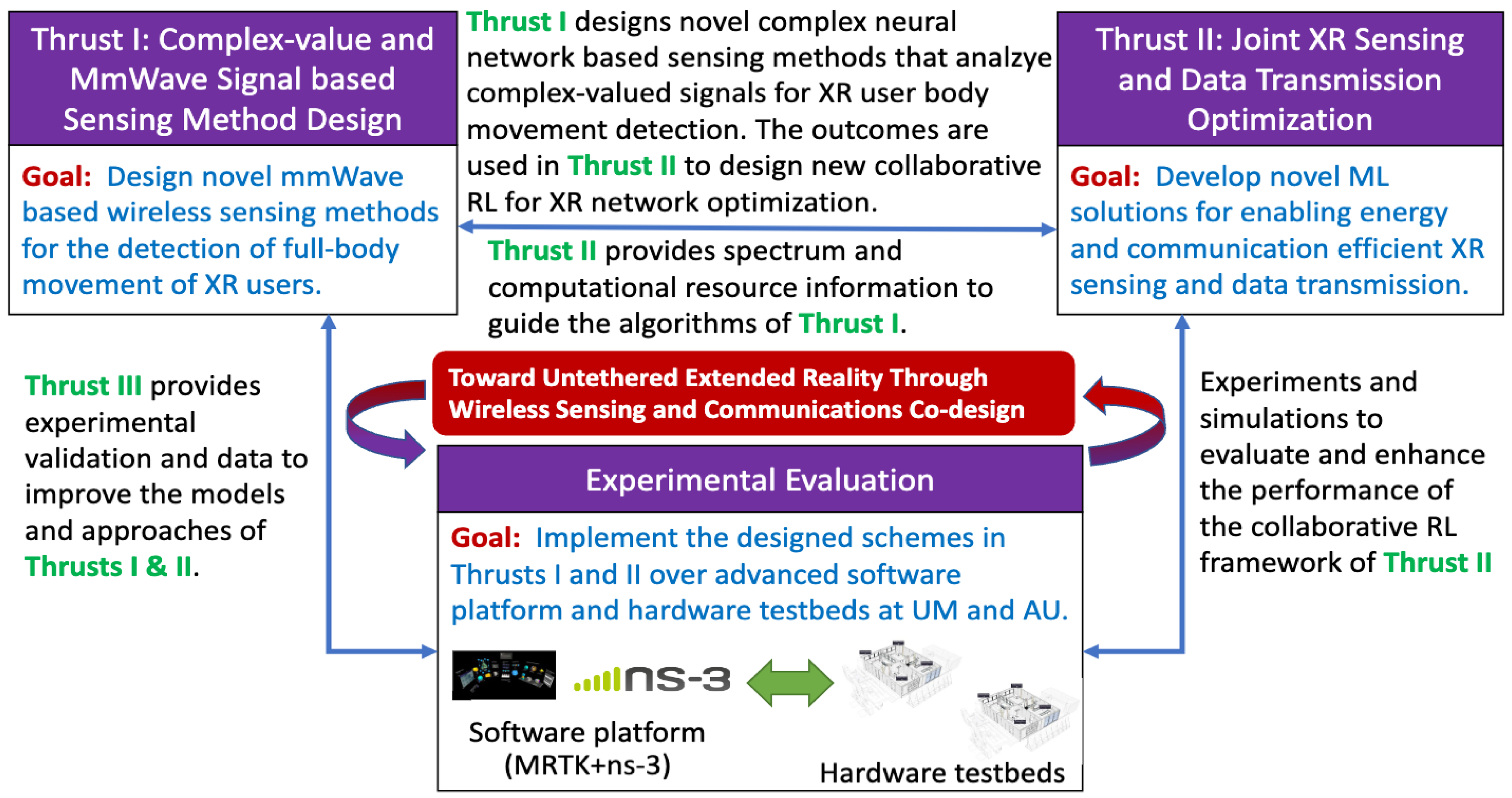

Fig. 1: Overview of the proposed framework

Complex-valued Neural Networks for User Positioning

Research Goal: We design novel complex-valued neural networks (CVNNs) such as complex number activation functions, and neural network layers.

CVNN based Positioning Method: Each complex valued neuron can output two dependent variables.

- Case I: The output is x-axis and y-axis of the user position.
- Case II: The output is two dependent channel state information (CSI) features (i.e., time of signal arrival and LoS/NLoS signal classification)

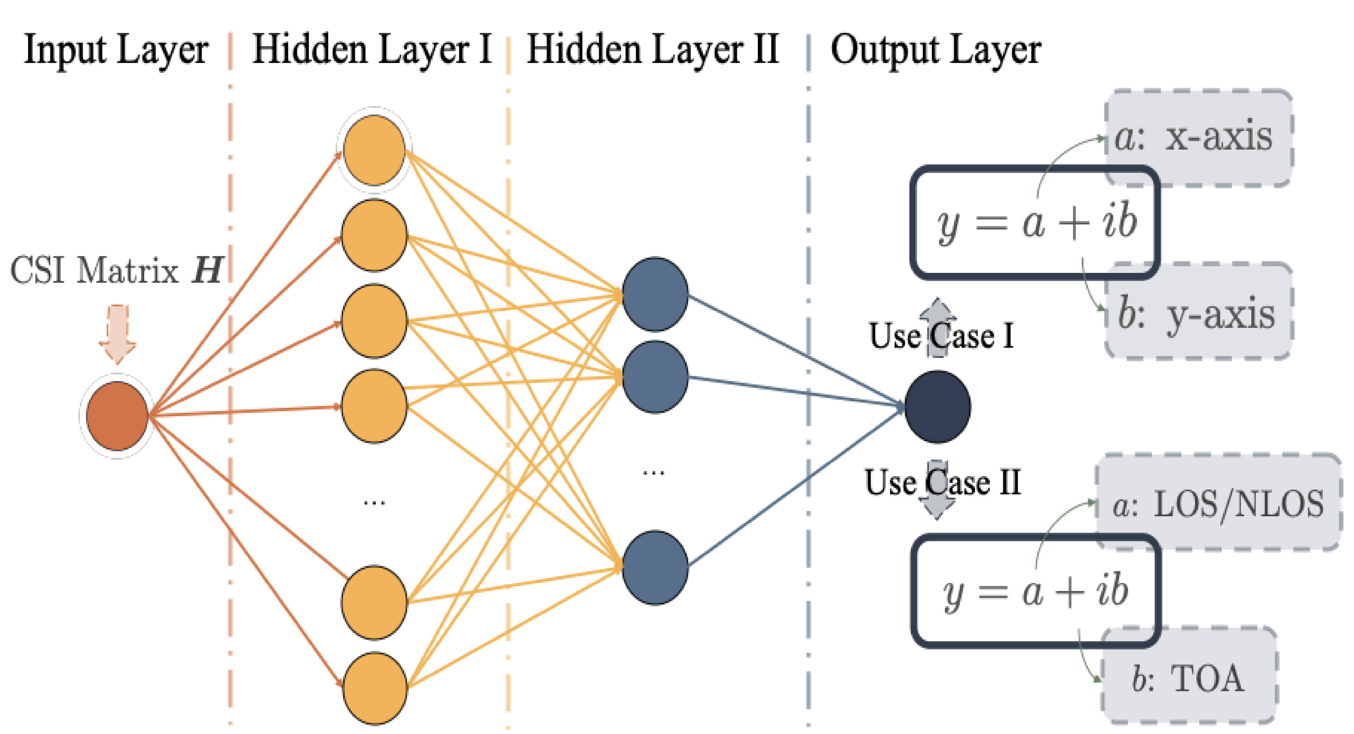

Fig. 2: The CVNN architecture.

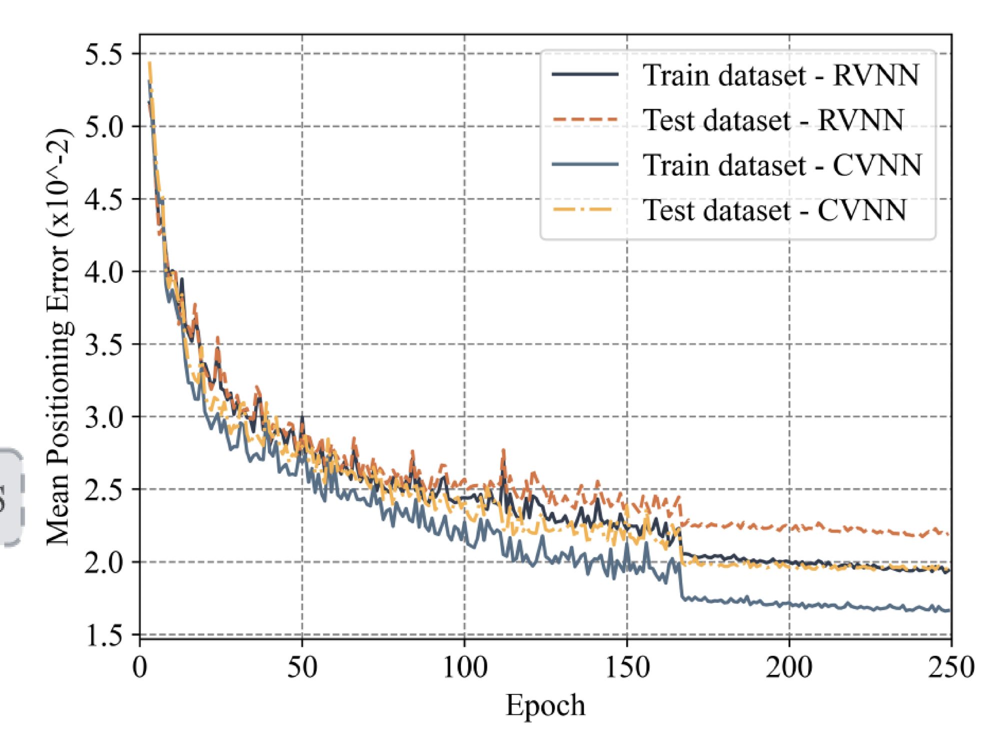

Fig. 3: Convergence.

CVNN based Federated Learning for Multi-User Positioning

Project Goal: To develop a multi-user positioning framework that enables a server to train a positioning model without the transmissions of users' CSI data.

Convergence analysis of CVNN based Federated Learning

Theorem 1. Given the transmission indicators $\boldsymbol{r}^t$ and $\boldsymbol{m}^t$, the optimal global FL model $\overline{\boldsymbol{g}}$, and the learning $h(\eta, t) = \frac{1}{Z}$, the upper bound of $\mathbb{E}\left(J(\boldsymbol{g}_{t+1}) - J(\overline{\boldsymbol{g}})\right)$ is

$$\mathbb{E}\left(J(\boldsymbol{g}_{t+1}) - J(\overline{\boldsymbol{g}})\right) \leq A^t \mathbb{E}\left(J(\boldsymbol{g}_1) - J(\overline{\boldsymbol{g}})\right) + \left(\frac{1 - A^{t-1}}{1 - A}\right) \frac{2\zeta_1 E}{ZN},$$

$$E = 2N - \mathbb{E}\left(\sum_{u=1}^{U} |\mathfrak{B}_u^t| r_u^t + \sum_{u=1}^{U} |\mathfrak{B}_u^t| m_u^t\right),$$

where $\mathbb{E}(\sum_{u=1}^{U} |\mathfrak{B}_u^t| r_u^t)$ ($\mathbb{E}(\sum_{u=1}^{U} |\mathfrak{B}_u^t| m_u^t)$) is the expected total training data samples of the users that send real (Imaginary) component of the local model to the server.

Simulation Results

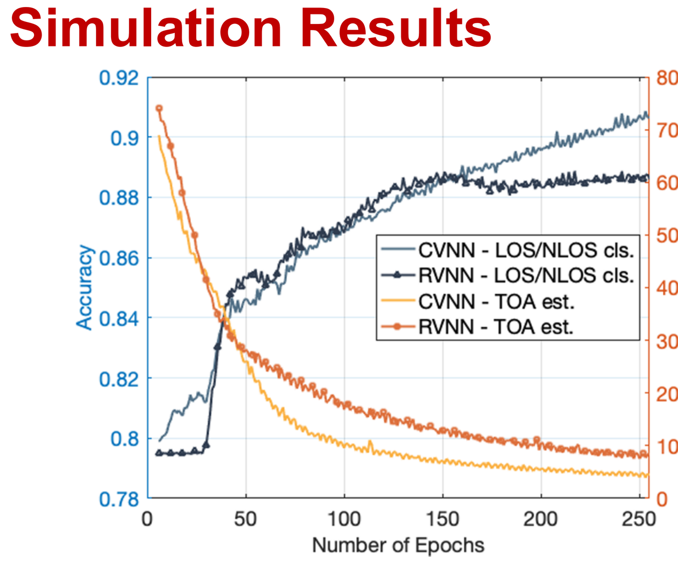

Fig. 4: CVNN vs. RVNN

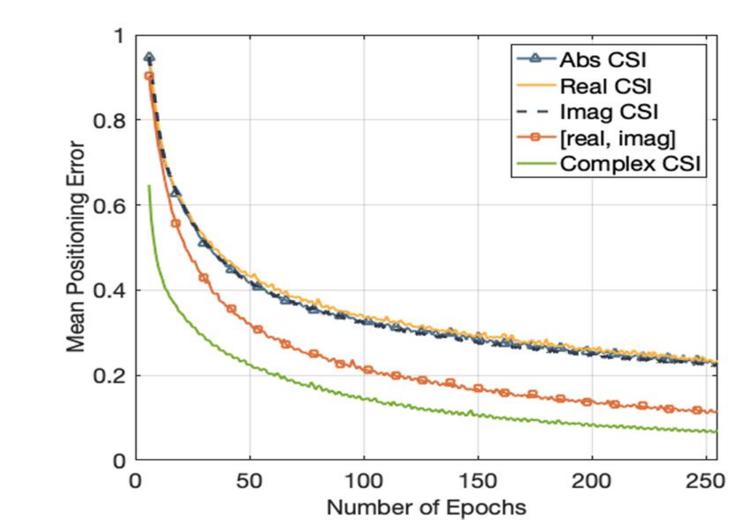

Fig. 5: Various CSI processing methods

Fig. 6: Testbed

Semantic Communications for XR Image Transmission

Project Goal: To design a novel semantic communication framework that enables a transmitter to 1) extract data meanings of the original data without neural network model training, and 2) select an appropriate semantic encoder model to meet both data transmission requirements (e.g., delay) and follow-up task requirements (e.g., classification accuracy).

Semantic Encoder and Decoder Design: Contrastive language–image pre-training model

Problem Formulation and Simulation Results: Our goal is to optimize the follow-up task performance (e.g., image regeneration or data classification) while meeting the delay and energy consumption requirements. This problem involves the optimization of resource block allocation $\boldsymbol{\alpha}_i$, the selection k_i of the CLIP model-based encoder

$$\min_{\boldsymbol{\alpha}_i, k_i} \sum_{i=1}^{U} f\left(\boldsymbol{\alpha}_i, k_i\right),$$

$$\text{s.t.} \quad \alpha_{i,q} \in \{0,1\}, \quad k_i \in \{0,1,2\}, \quad \forall i \in \mathcal{U}, \forall q \in \mathcal{Q},$$

$$\sum_{q=1}^{Q} \alpha_{i,q} \leq 1, \forall i \in \mathcal{U},$$

$$\sum_{i=1}^{U} \alpha_{i,q} \leq 1, \forall q \in \mathcal{Q},$$

$$l_{\text{total}}\left(k_i, \boldsymbol{\alpha}_i\right) \leq D,$$

$$e_{\text{total}}\left(k_i\right) \leq E,$$

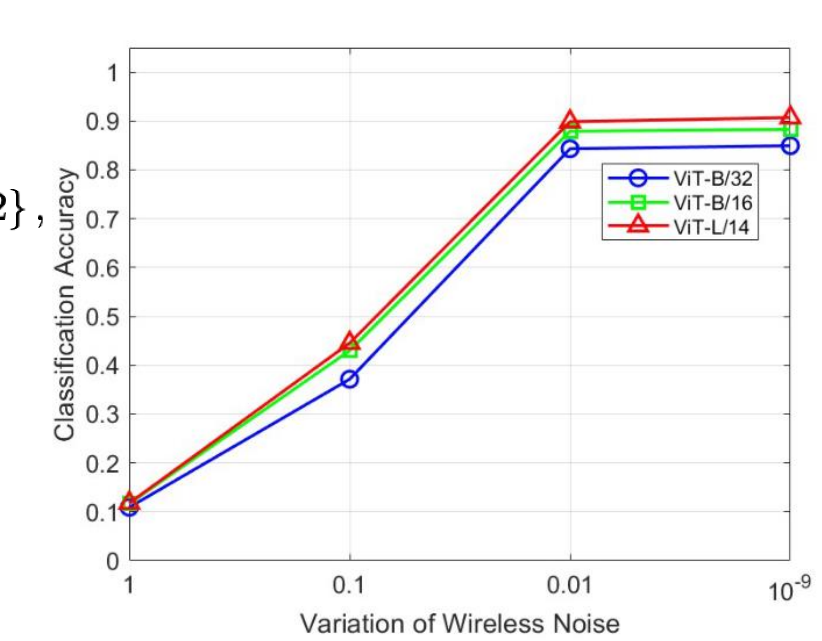

Fig. 7: Impacts of wireless noise on CLIP

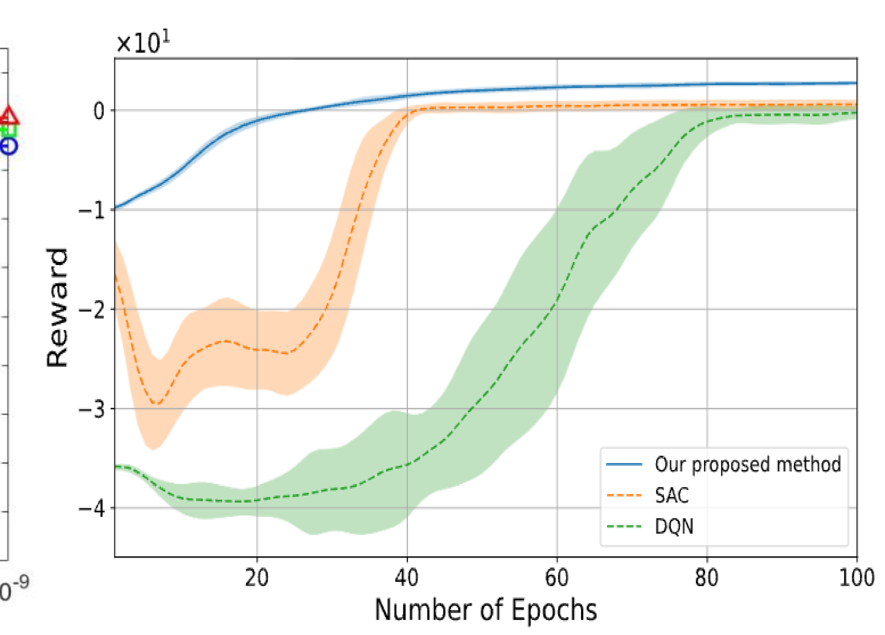

Fig. 8: Convergence