Rowan University
COLLEGE OF SCIENCE & MATHEMATICS

Computation Offloading for Precision Agriculture using Cooperative Inference

Suraj Bitla, Gopi Krishna Patapanchala, Stephen Piccolo, Shen-Shyang Ho, Ph.D. (hos@rowan.edu), Department of Computer Science
Nicholas Bovee, College of Engineering
Paolo Rommel Sanchez, College of Engineering and Agro-Industrial Technology, University of the Philippines Los Baños

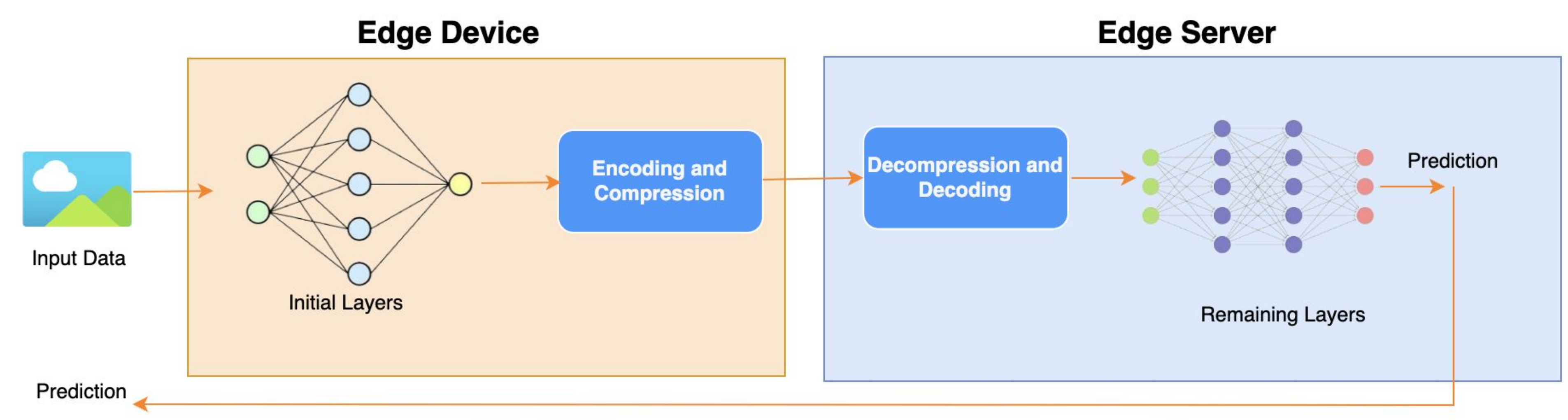
Project Goal: **Use of a cooperative inference framework to support edge devices with limited power and computational resources on agricultural tasks.**

Background

iStock. (2023). Image of robotics in agriculture https://www.foodmanufacturing.com/supply-chain/news/22882565/robotics-power-the-future-of-farming-forestry

- Precision Agriculture (PA) employs advanced technology and data analysis to maximize farm resource efficiency, guiding farmers to make decisions based on real-time crop, soil, and environmental data.
- It ensures efficient use of water, fertilizers, and pesticides, contributing to sustainable and effective farming practices.
- Deep learning, such as convolutional neural networks in machine vision systems, is central to PA, enhancing tasks like targeted herbicide application to minimize waste and boost crop yields, signaling a shift towards a new era in agriculture.

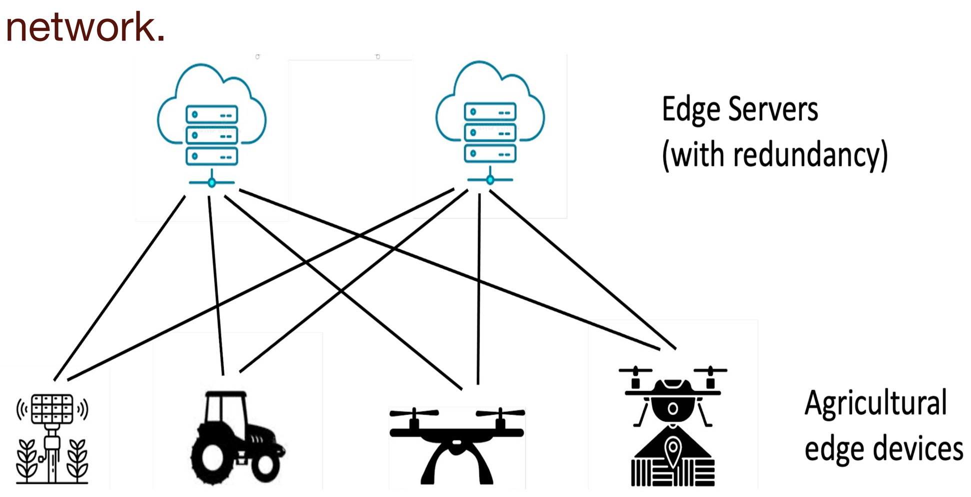
Execution Flow



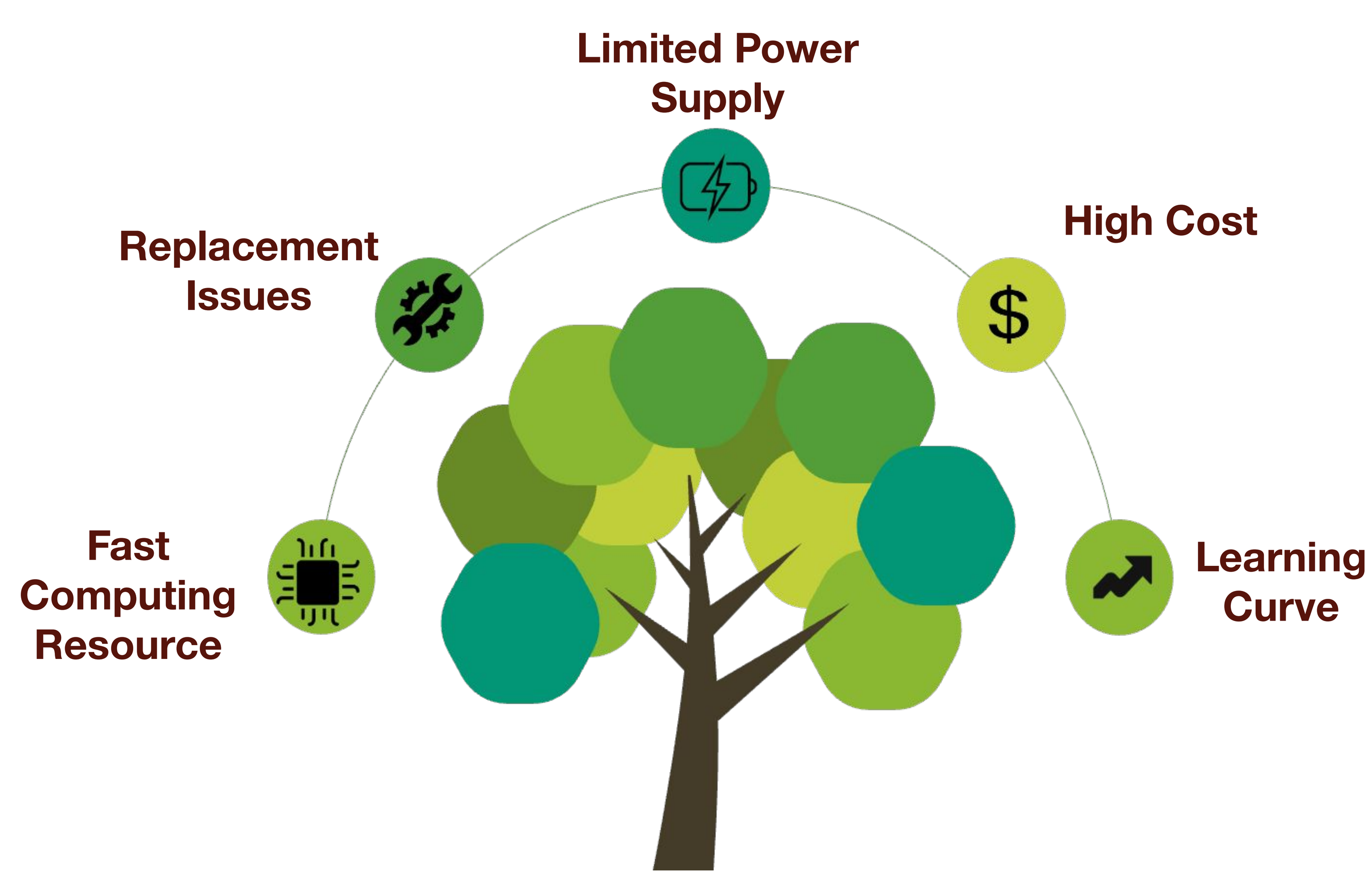
- Use of advanced deep learning architectures such as **AlexNet** and **ResNet** for predictive model.
- Splitting the architecture into two parts: One for Edge device and the other part for Edge Server.
- Compression and wireless transmission of the tensor output from device to server.
- Decompression and final layer processing on the Edge Server, with predictions relayed back to the local device.

Proposed System Design

We utilize split computing for **cooperative inference,** it is a modern and innovative approach within the field of deep learning that has gained significant attention in recent years. At the core of this approach lies the strategic partitioning and distribution of computational tasks across multiple devices, servers, or nodes within a network.



Challenges of PA



Benefits of System Design

Efficiency Gains:
- Reduces energy usage by offloading tasks to edge devices.
- Improves real-time processing capabilities and reduces latency.

Cost Reduction:
- Lowers the need for expensive computing equipment for farmers.
- Makes advanced PA technologies more accessible and affordable.

Performance Enhancement:
- Increases computational capacity for AI tasks in agriculture.
- Enhances responsiveness and adaptability of services in agricultural IoT.

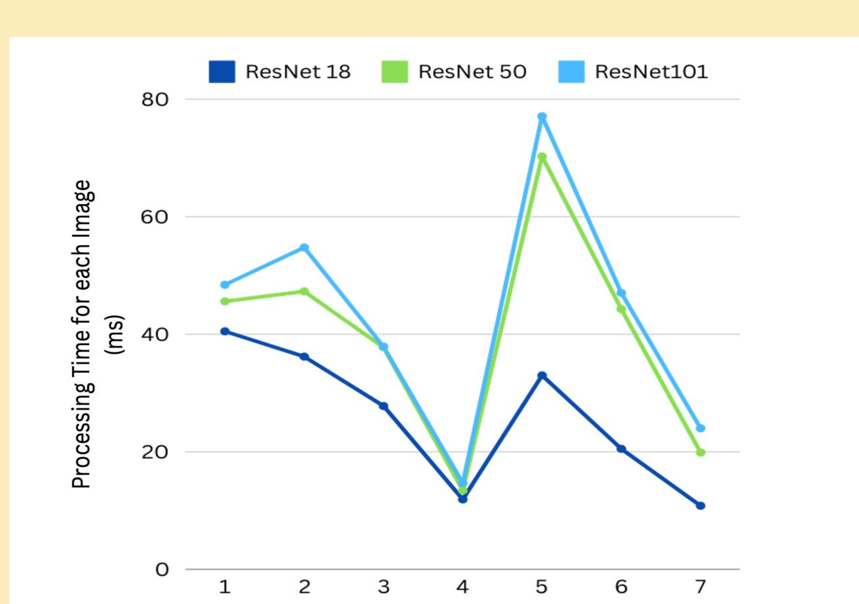
Sustainability and Energy Efficiency:
- Aligns with the energy constraints of agricultural environments for sustainable farming.

Splitting Points

- We perform experiments on both **classification models** (AlexNet, variants of ResNet) and **object detection models** (R-CNN and YOLO) and determine optimal layer split points in terms of computational cost and network latency.
- The provided diagrams illustrate the strategic split points within ResNet models 50, 101, and 152. These splits are designed to **optimize the computational workload and data transmission** between edge devices and edge servers. It also facilitates **real-time processing**, contributes to **energy-efficient computing**, and enhances the system's maintenance and scalability.

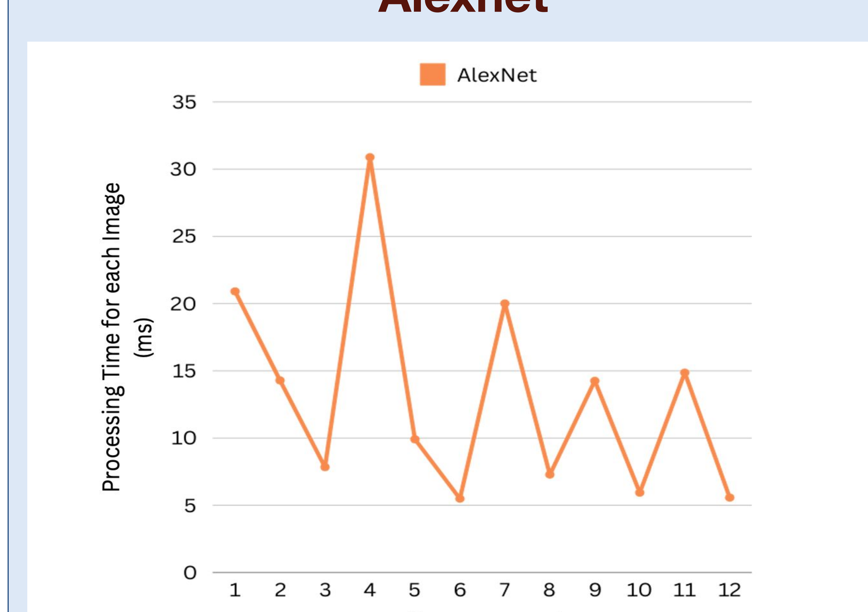
Experimental Results

Resnet 18/101/152



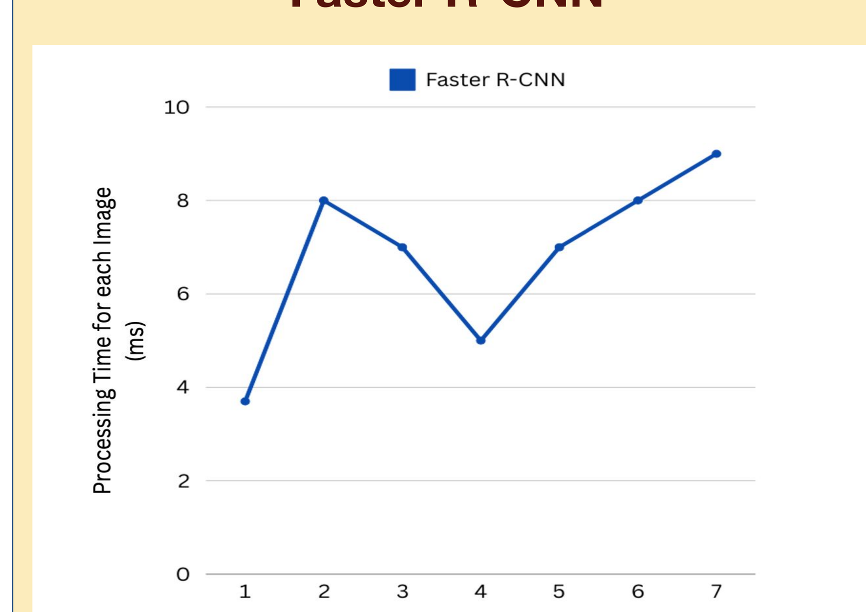
- Dataset: MNIST
- Best Split : 4
- Used Blosc2 for compression

Alexnet

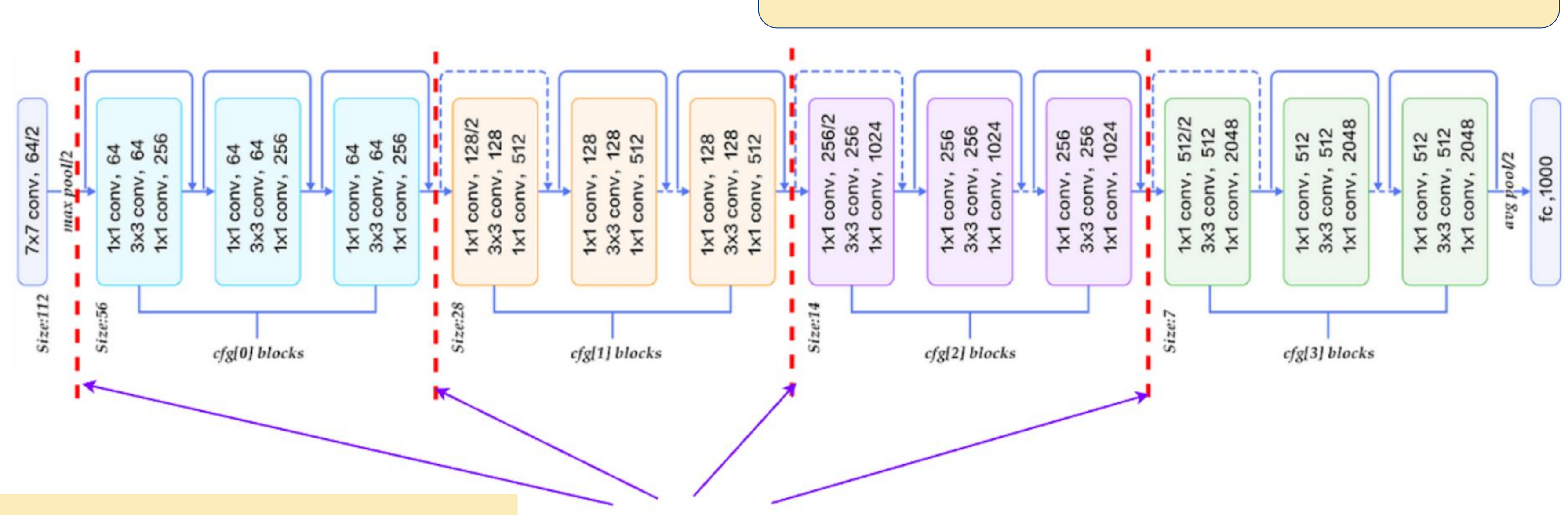


- Dataset: MNIST
- Best Split : 6
- Used Blosc2 for compression

Faster R-CNN



- Dataset: Individual_Weed
- Best Split : 1
- Used Blosc2 for compression



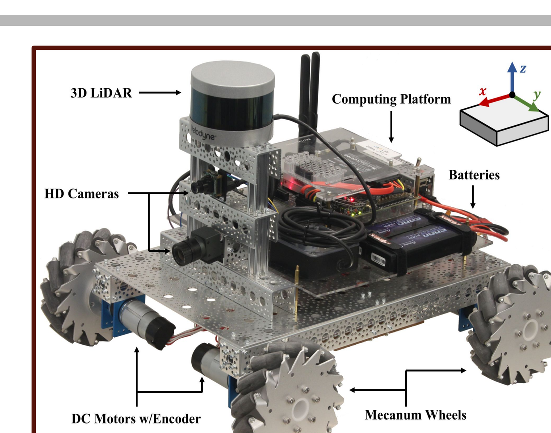
Implementation Challenges

- Finding places to split some models was tricky, especially when the models have a complex structure.
- Moving data from the edge device to the server was slow because of limited network speed.
- Determining the ideal split point required thorough examination, which included assessing energy efficiency, total data transmission time, and overall processing time to ensure the most effective division of tasks between the edge device and the server at each possible split position.

Future Works:

- Conduct extensive testing within the developed testbed to refine the cooperative inference framework.
- Implement the proposed research within a vehicular edge computing platform to validate its effectiveness in real-world scenarios.
- Customize and utilize object detection models such as RCNN and YOLO to enhance the adaptability and accuracy of the system.

(This project is supported by NSF Award #2128341)



Reference:
[1] N. Bovee, S. Piccolo, S. S. Ho, and N. Wang, "Experimental test-bed for computation offloading for cooperative inference on edge devices," in EdgeComm: The Fourth Work-shop on Edge Computing and Communications (at ACM/IEEE Symposium on Edge Computing). IEEE, 2023
[2] P. R. Sanchez, H. Zhang, S.-S. Ho, and E. De Padua, "Comparison of one-stage object detection models for weed detection in mulched onions," in 2021 IEEE International Conference on Imaging Systems and Techniques(IST). IEEE, 2021, pp. 1–6
[3] P. R. Sanchez and H. Zhang, "Evaluation of a cnn-based modular precision sprayer in broadcast-seeded field," Sensors, vol. 22, no.24, p. 9723, 2022
[4] CAR Lab. "Hydra Robot Diagrams." Hydra One, http://weisongshi.org/hydraone/index.html. Accessed 27 Mar. 2023.