

# DeepRob Discussion 1

1/21/2025

# Session Agenda

- Recap
  - P0 Debrief
  - Last Session
- Datasets
  - PROPS Dataset
  - Deepfish Dataset
  - ERA5/Graphcast Dataset
- P1 Kickoff

# P0 Debrief

- Seemed to go well
  - Solid performance
  - Fewer issues than last semester
  - Primary Issue - Lack of saving of notebook before submission
- Student perspective
  - What went well?
  - What could have been better?



# P0 Debrief

- Learning Objectives
  - Introduction to PyTorch
    - Tensors
  - Introduction to Class Operations
    - Autograder, etc.

# Datasets

Learning objectives for today

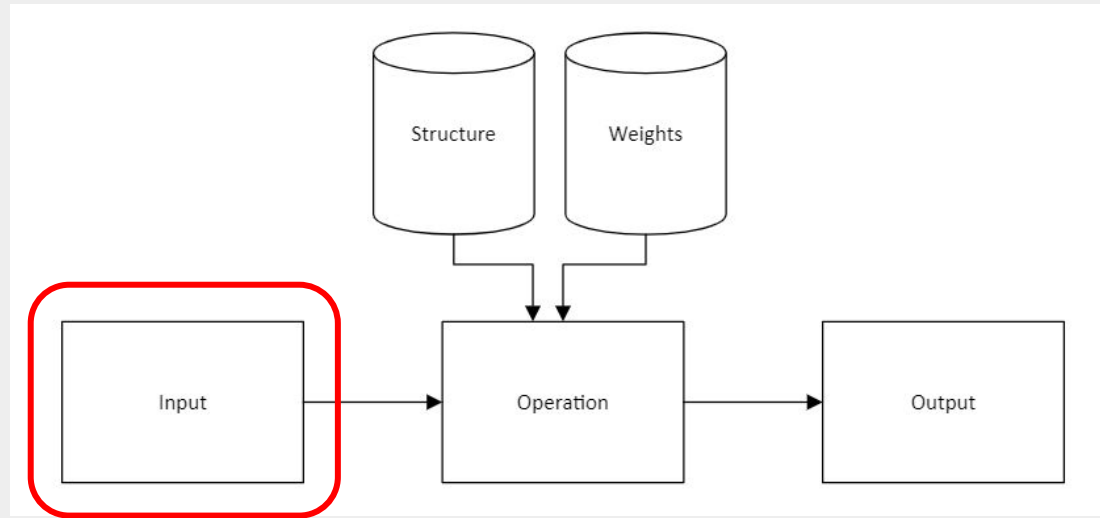
- Gather rough understanding of datasets
- Inspect three datasets
  - PROPS
  - DeepFish
  - ERA5
- Work through different aspects of dataset with regards to different machine learning models

# Learning Objective

- Describe usage of datasets with regard to machine learning tasks
- Describe structure and components of datasets
- Identify different mechanisms by which we can gather information about datasets

# AI/ML Background

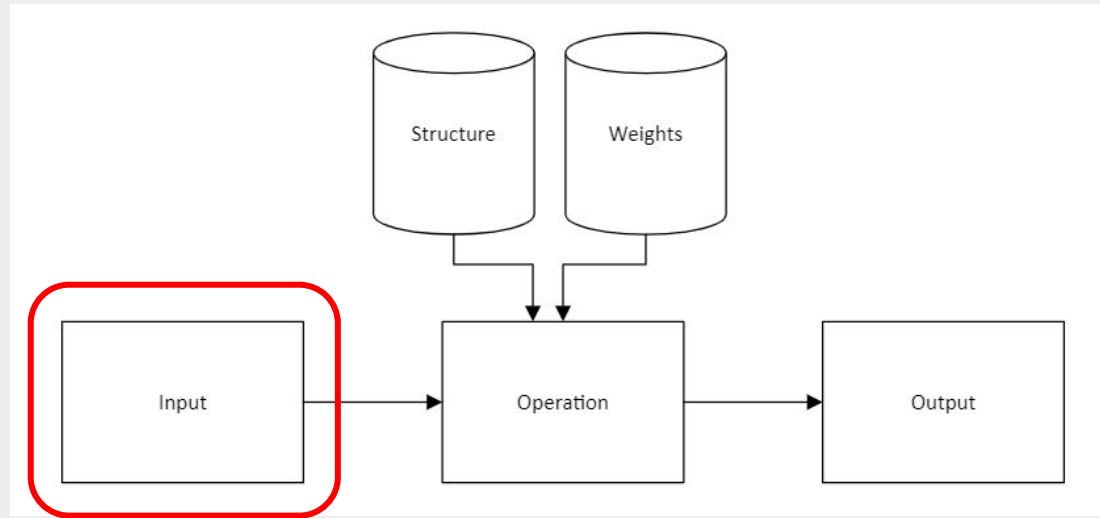
What were some of the formats we talked about last week?





# AI/ML Background

What are the different ways that datasets are used in this paradigm?



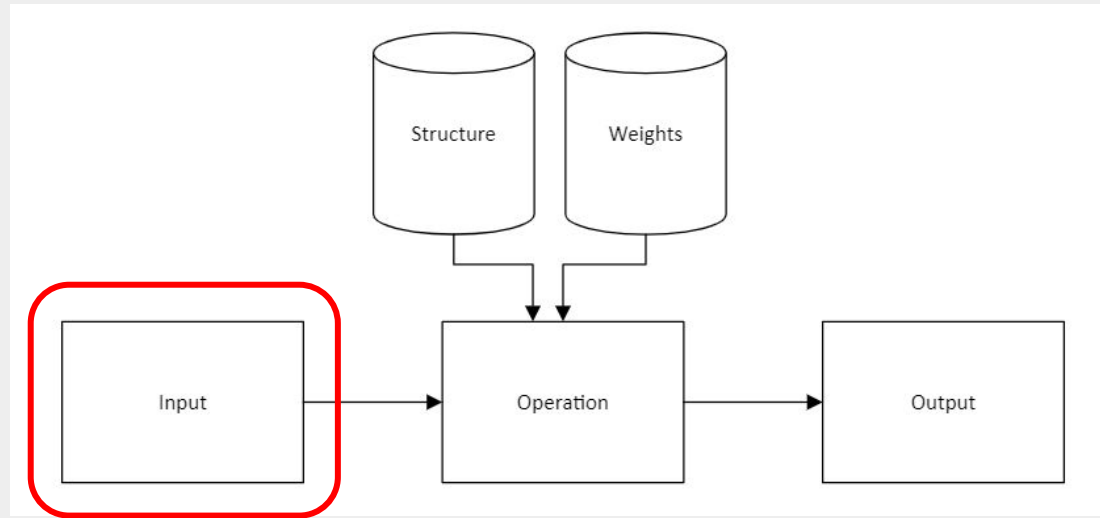




# AI/ML Background

What are the different ways that datasets are used in this paradigm?

- Inference
- Training
- Validation
- Testing





# What are the components of a dataset?

- Data
  - Multiple Formats
- Metadata
  - Data divisions
- Annotations
  - Provides ground truth
  - Can be generated in multiple ways
- Other Considerations
  - How was data collected?
  - Is the data complete?
  - Were there redactions?
  - Is the data synthetic?
  - How does the data relate to itself?
- These questions are most often answered in the paper introducing the dataset



# Demo - PROPS Dataset Slides



# Demo - DeepFish Dataset Inspection



# Demo - ERA5/Graphcast Dataset Interaction



**Class Activity - Find a dataset that  
matches your interests!  
Post into Piazza**



# Back to Our Learning Objective

- Describe usage of datasets with regard to machine learning tasks
  - Training
  - Validation
  - Testing
  - Inference

# Back to Our Learning Objective

- Describe structure and components of datasets
  - Data
  - Metadata
  - Annotations
  - Other considerations



# Back to Our Learning Objective

- Identify different mechanisms by which we can gather information about datasets
  - Inspect files themselves (DeepFish)
  - Inspect code used to load and process data (PROPS)
  - Research; read papers on dataset creation and usage (DeepFish/GraphCast)

Quick break!

# Class Objective - P1 Primer

Same as last time - Everyone get some points on the board!