Reading: Assignment Overview: Data Loading and Augmentation Using PyTorch

Estimated time: 2 minutes

Welcome to the Data Loading and Augmentation using PyTorch lab. In this hands-on session, you will build data pipelines for deep learning models using PyTorch's in-built data handling utilities. You'll work with an agricultural satellite imagery dataset to understand how to load, augment, and prepare image data for machine learning.

By completing this lab, you will:

1. Create custom PyTorch Dataset classes for specialized data loading
2. Utilize PyTorch's built-in "ImageFolder" utility for standard directory structures
3. Implement data augmentation pipelines using torchvision "transforms" methods
4. Build efficient DataLoader objects for batched data loading
5. Visualize and analyze augmented datasets

There are five tasks to be completed by you in this lab:

1. **Build a custom transformation pipeline:** You will create a data augmentation pipeline using PyTorch's transform for data augmentation. This transform pipeline demonstrates how data augmentation helps prevent overfitting by showing your model varied versions of the same images during training.
2. **Implement ImageFolder Dataset creation:** Your next task will be to use PyTorch's "datasets.ImageFolder" class to load, in this case, the agricultural imagery dataset.
3. **Explore dataset properties:** After creating the dataset, you will investigate the class information from the automatically generated information from the "ImageFolder" dataset. This will help you understand the inner working of the dataset method.
4. **Visualize custom dataset results:** Finally, you will display a batch of images from your custom dataset implementation alongside their labels. This will give you a visual confirmation of your entire data pipeline, that it produces correctly augmented samples with correct labels.

By the end of this lab, you will have hands-on experience with production-level data loading patterns for real-life datasets used in deep learning projects, and you will be able to understand the modular design of data handling in PyTorch.

You will need to download and save the finished lab on your computer for final evaluation at the end of this course. Good luck!