Task 1 Deliverable — Dataset Generation

Overview

The purpose of Task 1 was to build a preprocessing pipeline for simulated multi-coil MRI phantom data. This pipeline included undersampling, augmentation, normalization, and tensor formatting in order to produce training-ready examples for later contrastive learning tasks.

Transformations Implemented

1. Normalize

Goal: Standardize the dynamic range across samples for stable training. Approach: Compute magnitude = $\sqrt{(real^2 + imag^2)}$, estimate global mean/std, and apply normalization to both real and imaginary channels. Optional clipping is supported. Design choice: Using magnitude statistics ensures consistent scaling between real and imaginary parts.

2. EquispacedUndersample

Goal: Simulate accelerated MRI acquisition by undersampling k-space lines. Approach: Apply FFT, retain every N-th line, preserve central low-frequency band, and inverse FFT. Parameters: acceleration=4, center_fraction=0.08. Design choice: Simple, reproducible approximation of accelerated MRI acquisition.

3. Augmentation

Goal: Introduce lightweight spatial variability to encourage model robustness. Approach: Small random rotation (±10°) and translation (±2 pixels) applied jointly to real & imaginary parts. Design choice: Mild augmentations mimic realistic variability without distorting anatomical structure.

4. ToCompatibleTensor

Goal: Convert complex multi-coil data into a format usable by Conv2d models. Approach: (C,H,W,2) → (C*2,H,W) stacking real and imaginary channels. Design choice: Preserves phase explicitly while remaining model-agnostic. Trade-off: doubled channels.

Transformation Order

The order applied is: 1. EquispacedUndersample → simulate accelerated acquisition. 2. Augmentation → add realistic spatial variability. 3. Normalize → stabilize intensity statistics. 4. ToCompatibleTensor → prepare data for PyTorch models. Rationale: Acquisition corruption is applied first, augmentation adds variability, normalization stabilizes intensity, and tensor conversion ensures model compatibility.

Example Outputs

The preview script (preview_task1_dataset.py) was executed with parameters: accel=4, center_frac=0.08, rot=10, shift=2, seed=123. This produced PNGs such as task1_sample_0.png, task1_sample_1.png, task1_sample_2.png. (Example figures can

be inserted here.)

Optional / Additional Notes

- A clipping option was added to Normalize to cap extreme outliers. - Augmentation uses a unified affine transform applied jointly to real and imaginary channels. - Preview script is lightweight and designed only for inspection/report figures.

Summary

The Task 1 pipeline successfully simulates accelerated MRI data, introduces realistic variability, normalizes across samples, and outputs a format suitable for standard neural networks. This provides a strong, production-ready foundation for subsequent tasks (model design and contrastive training).