Technical Take-Home Exercise: Multi-Sensor Deforestation Detection

The Challenge (6 hours)

Build a machine learning system that combines multiple satellite sensors to detect deforestation in the Brazilian Amazon. This isn't about following a recipe - we want to see how you approach the messy reality of working with multi-modal satellite data for environmental monitoring.

Improved deforestation detection could help protect millions of hectares of Amazon rainforest - this work has real environmental impact.

The Problem

Current deforestation monitoring systems face a fundamental challenge: optical satellites like Sentinel-2 provide rich spectral information but are blocked by clouds (which cover the Amazon 70%+ of the time), while radar satellites like Sentinel-1 can see through clouds but provide different, less intuitive information about forest cover.

Organizations like INPE (Brazil's space agency), MapBiomas, and Global Forest Watch have built operational systems, but there's still room for improvement in accuracy, timeliness, and handling edge cases. The question is: how do you effectively combine these complementary data sources?

Your Mission

Design and implement a multi-sensor approach to deforestation detection that combines SAR (radar) and optical data. We're providing a basic Sentinel-2-only dataset loader as a starting point, but it has significant limitations you'll need to address.

Core Requirements:

- **Must use both SAR and optical sensors** (this is non-negotiable)
- Address the temporal nature of change detection (before/after comparison)
- Handle real-world data messiness (clouds, missing data, different providers)
- Show measurable improvement over single-sensor approaches

Minimum Viable Solution:

- Loads both SAR and optical data successfully

- Demonstrates some form of sensor fusion (even if simple)
- Shows basic change detection capability
- Provides quantitative comparison vs single-sensor baseline

Available Satellite Data

SAR (Radar): Sentinel-1 C-band SAR with VV and VH polarizations (~10m resolution)

Optical: Sentinel-2 (10m, multiple bands), Landsat 8/9 (30m), CBERS-4A (2m PAN, 8m multispectral)

Dataset: 100 deforestation alerts from the Brazilian Amazon with multi-sensor imagery from ~30 days before to ~14 days after each alert date, plus ground truth polygons.

Note: While 100 alerts is small for deep learning standards, it's sufficient to demonstrate multi-sensor fusion concepts. Focus on methodology over perfect accuracy.

Download this file: https://storage.googleapis.com/capacity_shared/amazon_deforestation_v0.1/download_dataset.sh and put it in a folder. Then:

```
./download_dataset.sh all --extract                    # Download and extract everything

./download_dataset.sh sentinel2 --extract              # Get only Sentinel-2 data

./download_dataset.sh essential                        # Minimum dataset for ML research

./download_dataset.sh metadata                         # Just get documentation first
```

Key Technical Challenges

- **Multi-Modal Fusion**: How do you combine SAR backscatter and optical reflectance data with fundamentally different characteristics?
- **Temporal Change Detection**: Design architectures for sparse, irregular time series that focus on change rather than static classification
- **Data Harmonization**: Handle different sensors, providers, resolutions, and quality levels in a unified system

Data Challenges You'll Encounter

- SAR vs optical data have fundamentally different characteristics (backscatter vs reflectance)
- Multiple data providers with different band naming and statistics
- Extensive cloud coverage and irregular temporal sampling

- Mixed spatial resolutions (2-30m) and missing/corrupted files

Key Questions to Consider

- How do you handle missing data (SAR available but clouds block optical, or vice versa)?
- Can you reliably detect deforestation using SAR alone? Optical alone?
- How do you distinguish deforestation from seasonal vegetation changes?

Deliverables

- **Code**: Working multi-modal implementation with both SAR and optical data processing
- **Analysis**: Document your methodology, key findings, and trade-offs. How did SAR + optical compare to each sensor alone?
- **Results**: Quantitative comparisons showing multi-sensor vs single-sensor performance
- **Demo**: Clear example showing your system processing an alert and producing deforestation predictions
- **Submission**: GitHub repository with code, documentation, and result visualizations

Evaluation

We're evaluating:

- **Technical execution**: Multi-sensor fusion and change detection implementation
- **Problem understanding**: Grasp of real-world forest monitoring challenges
- **Communication**: Clear explanation of approach and results

Time Management & Tips

Key Tips:

- Focus on working end-to-end system rather than perfecting components
- Don't get stuck on perfect sensor fusion - simple approaches can be effective
- Process alerts individually to avoid memory issues
- Document assumptions and shortcuts
- If you run out of time, tell us what you would have done next

Evaluation Baseline: Aim to beat simple baselines like single-sensor CNN, NDVI differencing, or basic SAR/optical concatenation.

Resources

See the attached notebook with dataset structure and baseline Sentinel-2 loader (you need to add SAR)

Libraries: Use whatever you're comfortable with - PyTorch required for deep learning (no TensorFlow)

A Note on Scope: This reflects real environmental monitoring challenges. We want to see how you navigate complexity and make reasonable trade-offs under time constraints.

Questions? Feel free to reach out for clarification on multi-modal requirements.

Next Steps: Technical discussion exploring your solution and different fusion approaches.