Environmental change detection WITH DINOv3
(Using OSCD Dataset)

Karimov Temurbek
Kadirov Bekzod

Project Proposition Value

- Our project proposes an automated change detection system that compares satellite images from different years to identify how landscapes, infrastructure, and urban areas have evolved over time.
- This solution enables analysts, researchers, and local authorities to efficiently monitor long-term changes without manual inspection.

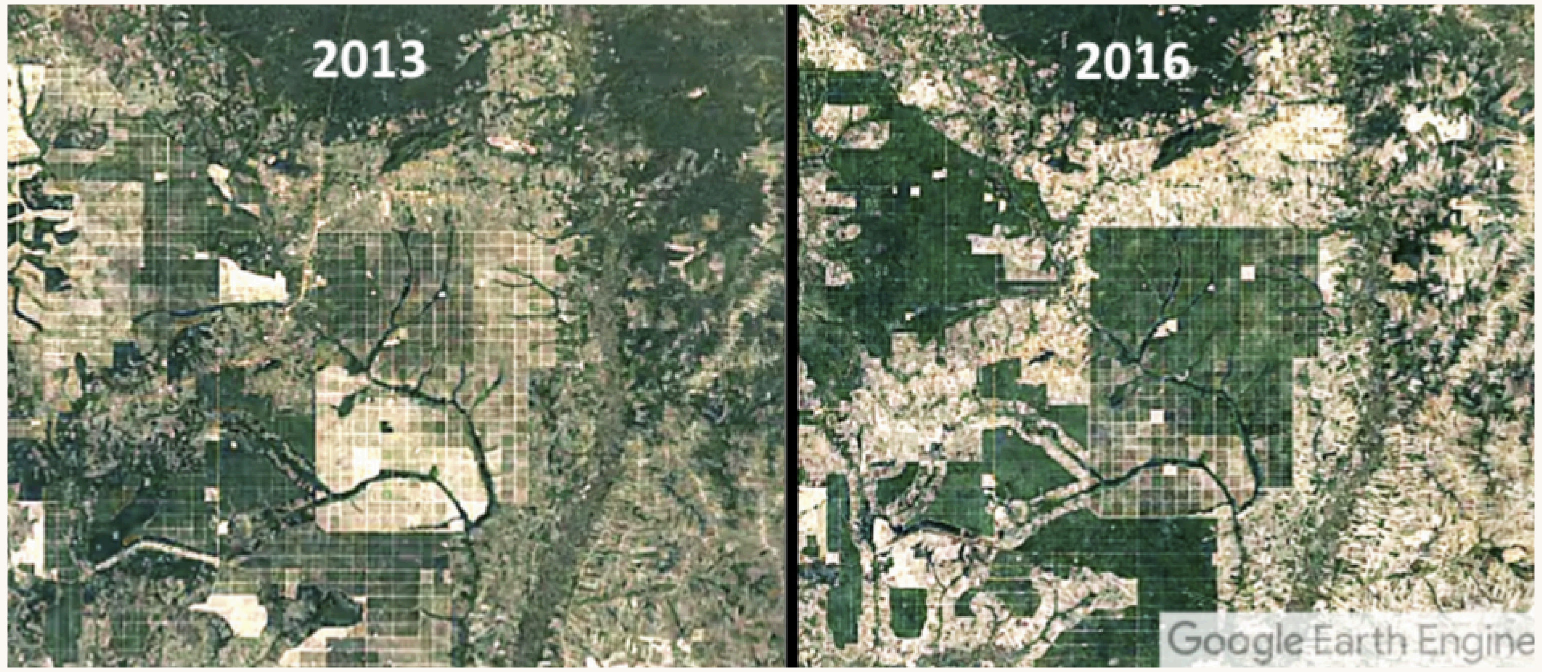
Value Provided

- Automated detection of year-to-year changes in buildings, vegetation, land use, and infrastructure
- Improved accuracy using transformer-based features, including DINOv3
- Faster analysis compared to manual comparison of satellite images
- Scalable monitoring for large geographic regions
- Supports data-driven decision-making for urban planning, environmental studies, and disaster prevention

Why This Problem Matters ?

Urban planning: Cities expand, infrastructure evolves, new buildings appear.

Environmental monitoring: Forest loss, agricultural expansion, water-level changes.

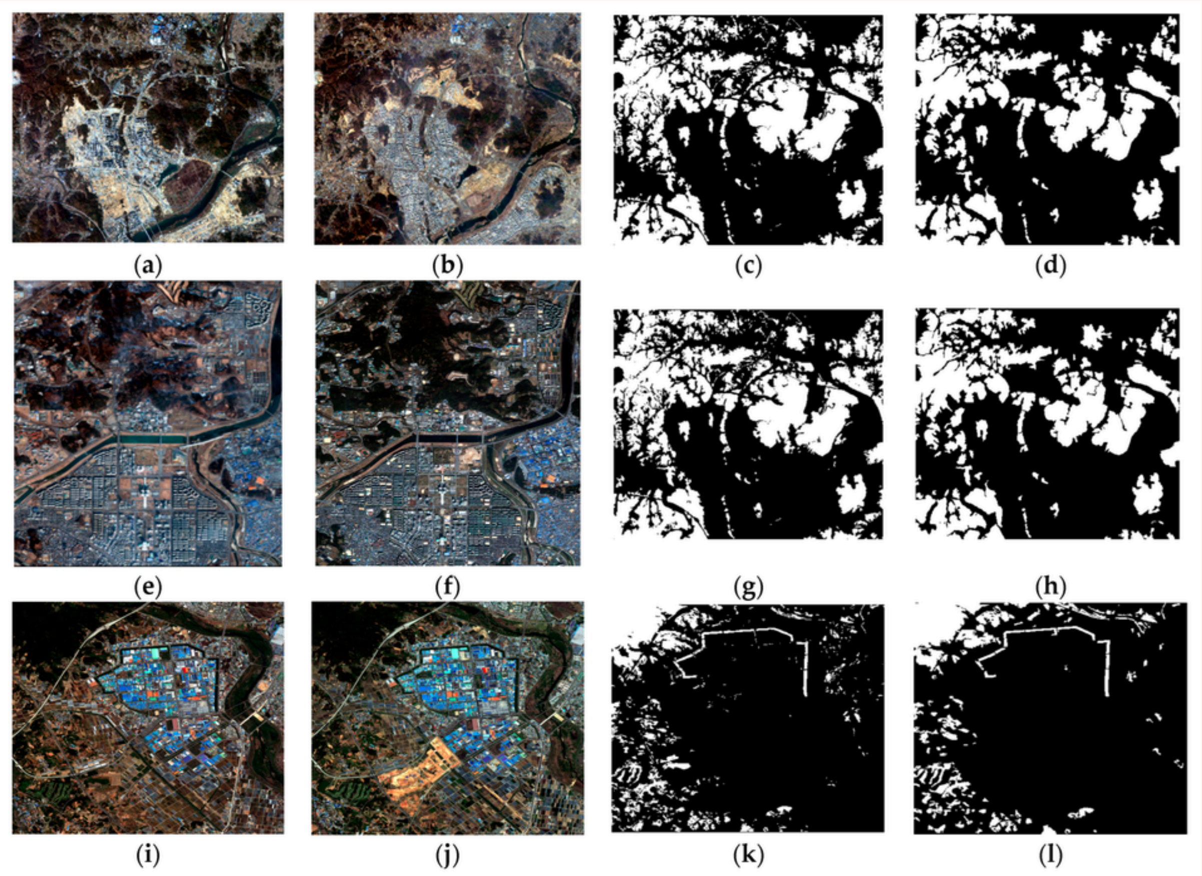


Infrastructure or landscape transformation over years

Research Questions & Hypotheses

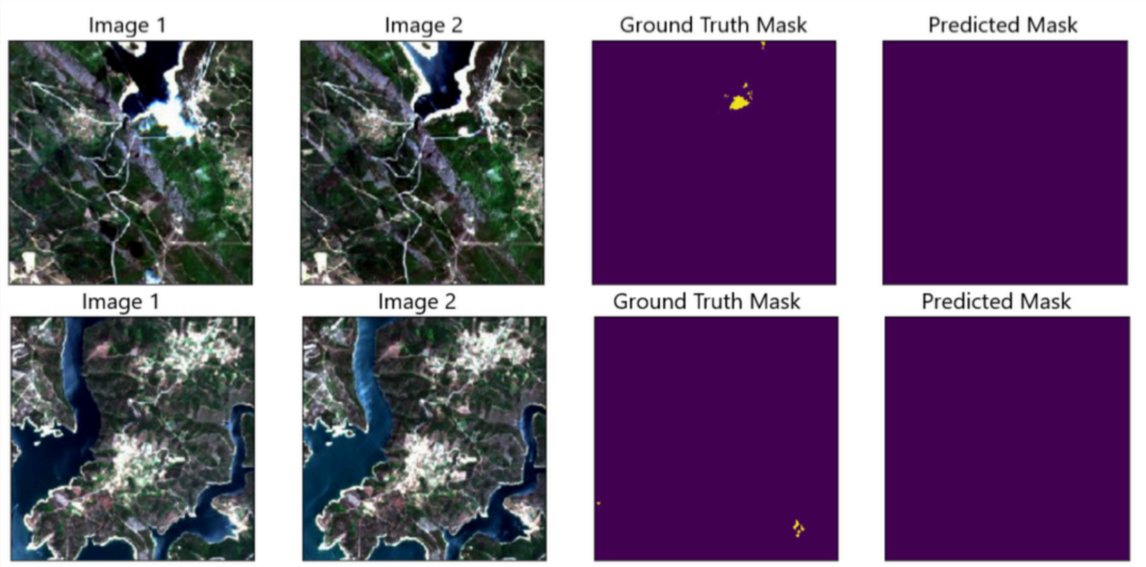
Research Questions

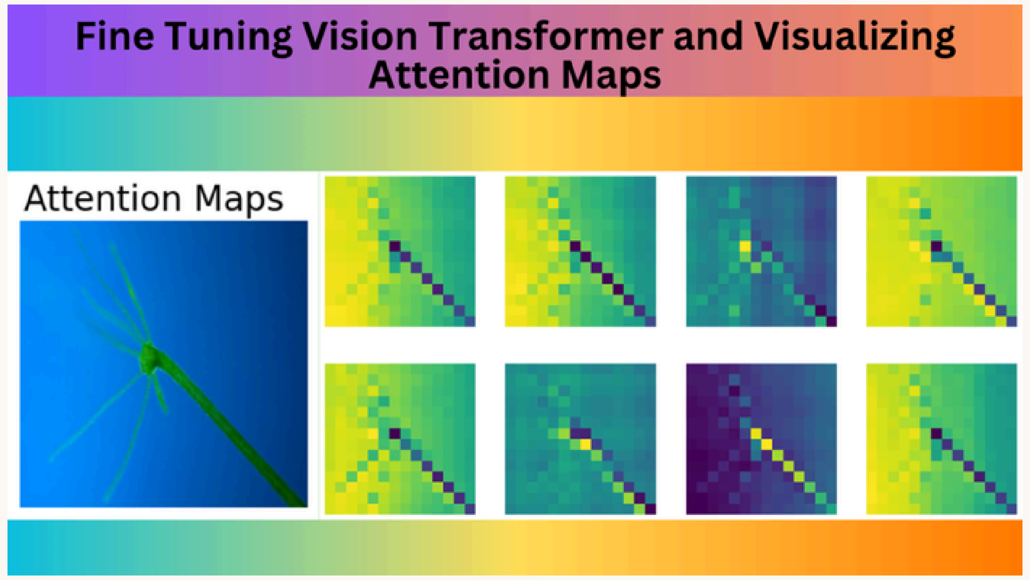
- RQ1 — Can DINOv3 detect subtle changes?
- RQ2 — Can we detect different types of changes?
- RQ3 — Does multi-year comparison improve consistency?
- RQ4 — Which fusion method works best?
- RQ5 — Can the model generalize to different regions?

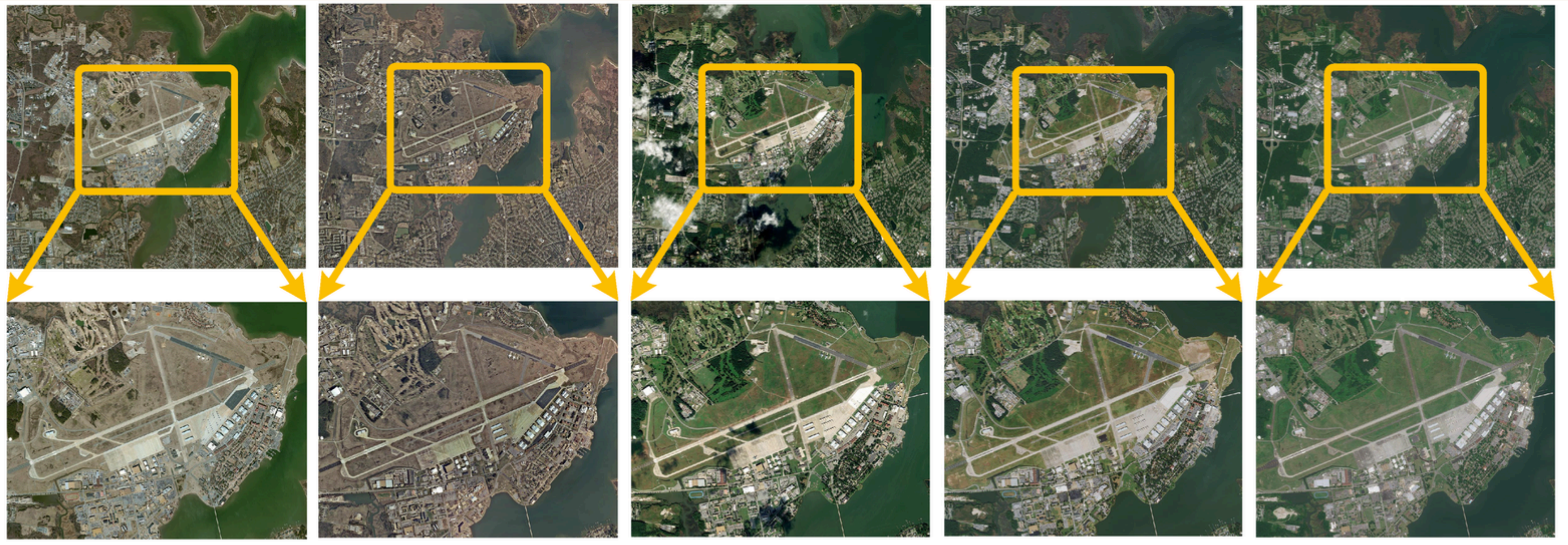


Hypotheses

- H1: DINOv3 features improve sensitivity to fine-grained structural and land-cover changes.
- H2: A dual-branch or multi-temporal architecture outperforms single-image models for identifying long-term transformations.
- H3: Combining multiple years of imagery results in smoother and more reliable change maps.
- H4: Spatial augmentations and patch-based processing increase robustness to noise, lighting differences, and seasonal effects.





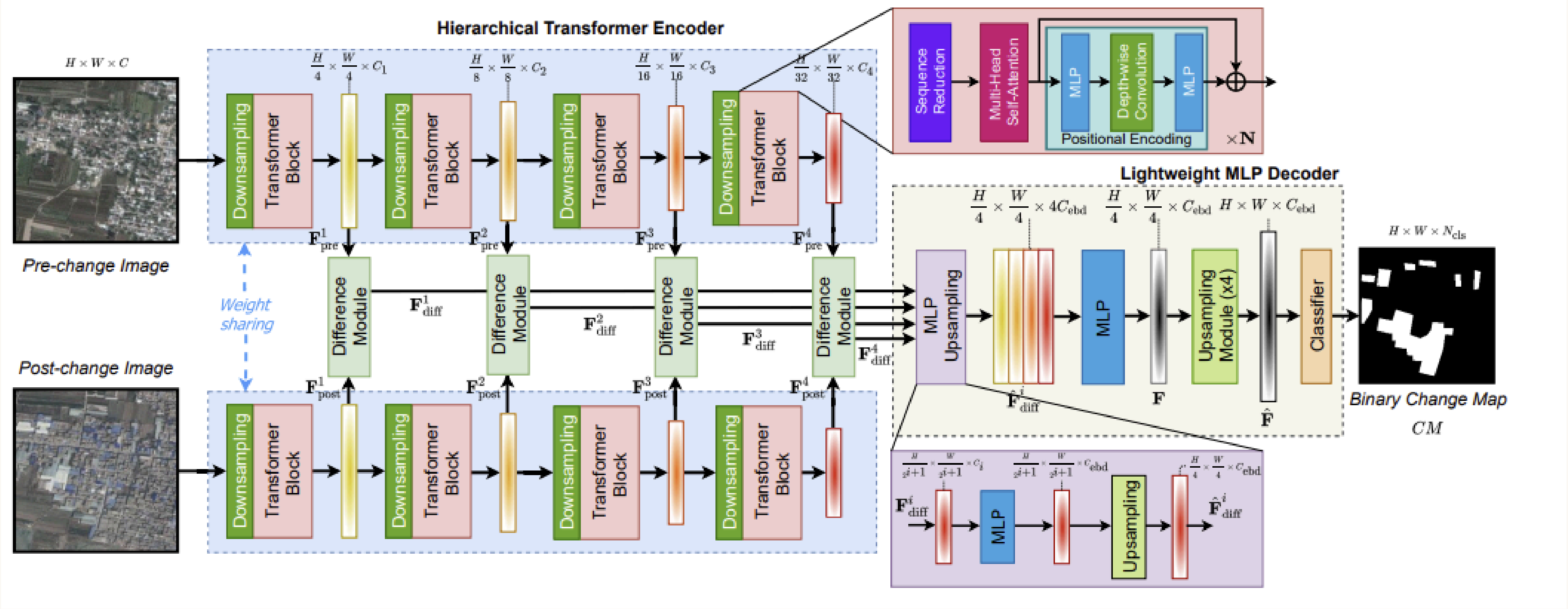


(a) 2015 (b) 2018 (c) 2019 (d) 2020 (e) 2022

Development Choices & Techniques

Model Architecture Choices

- Use DINOv3 Vision Transformer as the core feature extractor due to its strong capability to capture spatial patterns and subtle differences.
- Adopt a two-branch or multi-temporal fusion architecture, where each year’s satellite image is processed by the same encoder.
- Combine extracted features through feature fusion (concatenation, subtraction, or transformer cross-attention) to highlight differences.
- Use patch-based processing to handle high-resolution satellite imagery efficiently.

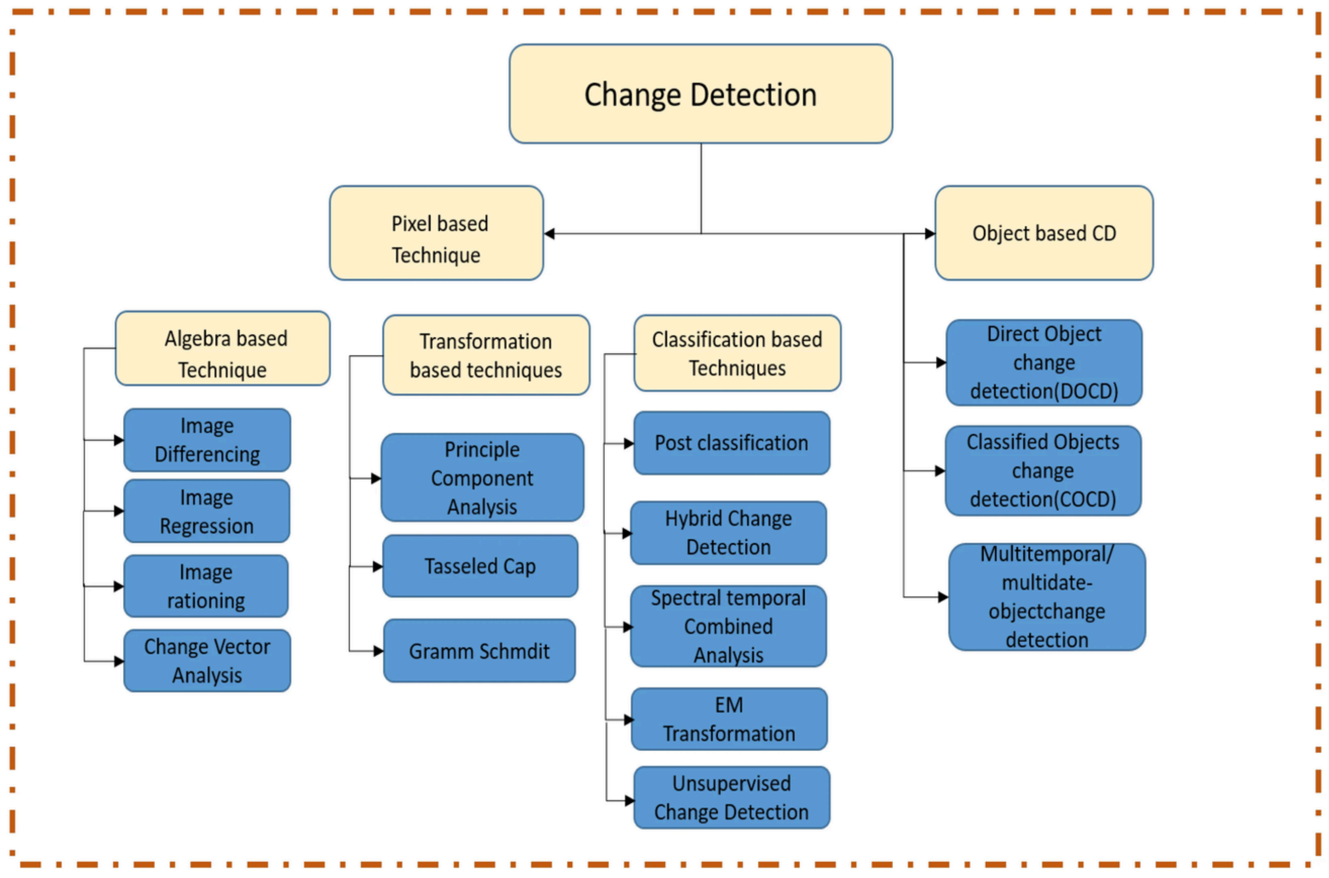


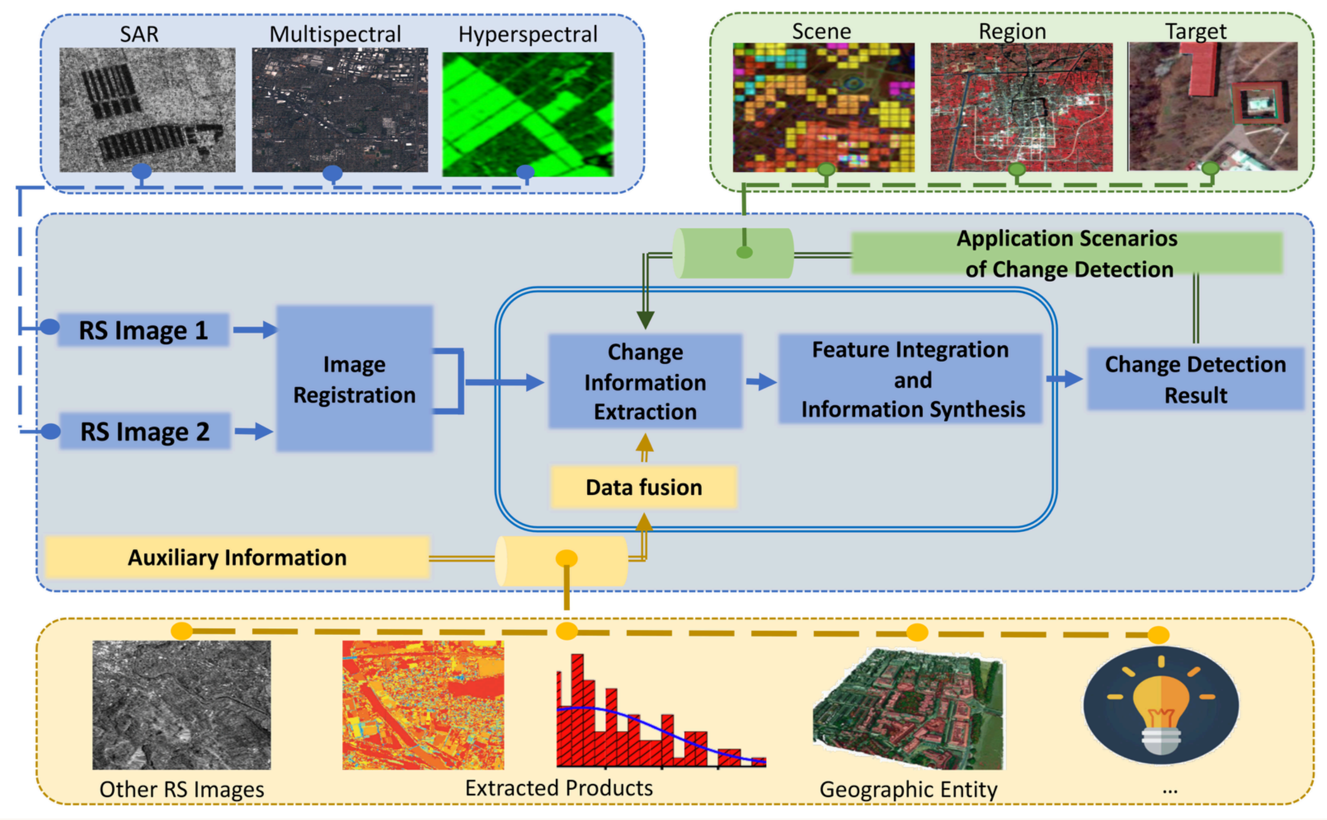
Training & Evaluation Choices

- Use change-detection labels or difference masks (if available) or evaluate model outputs manually.
- Loss functions:
  - Cross-entropy for change/no-change classification
  - Dice or Focal loss for imbalanced cases
- Use metrics:
  - Pixel-level accuracy
  - IoU for change regions
  - F1-score for rare changes
- Compare results using:
  - Single-year vs. multi-year input
  - CNN baseline vs. DINOv3 transformer

Data Processing Choices

- Satellite preprocessing:
  - Normalize lighting and contrast differences across years.
  - Align images spatially to ensure correct pixel-level comparison.
- Split images into overlapping patches (e.g., 256×256) to increase pixel diversity.
- Apply augmentations: random flips, rotations, color jitter — to reduce seasonal and illumination bias.





Evaluation Strategy & Expected Outcomes

Evaluation Strategy

1. Quantitative Metrics
   - Pixel-level Accuracy: Measures how correctly the model identifies change vs. no-change.
   - IoU for Change Regions: Evaluates overlap between predicted change areas and reference maps.
   - F1-score for Rare Changes: Useful when change areas are small or imbalanced.
   - Temporal Consistency Check: Ensures predictions are stable across sequences (2010→2013→2016).
2. Qualitative Evaluation
   - Visual inspection of multi-year change maps
   - Comparison with manually annotated regions
   - Overlay predictions on satellite imagery to assess structural correctness
3. Baseline Comparisons
   - CNN baseline vs. DINOv3 transformer
   - Single-year difference vs. multi-year fusion
   - Raw subtraction maps vs. learned features

Expected Outcomes

- More accurate identification of long-term changes (urban expansion, deforestation, infrastructure growth)
- Cleaner and more interpretable change maps compared to classical methods
- Higher sensitivity to subtle transformations thanks to DINOv3 features
- Robust performance across years with varying lighting and seasonal artifacts
- Practical usefulness for analysts, planners, and environmental teams

Questions?