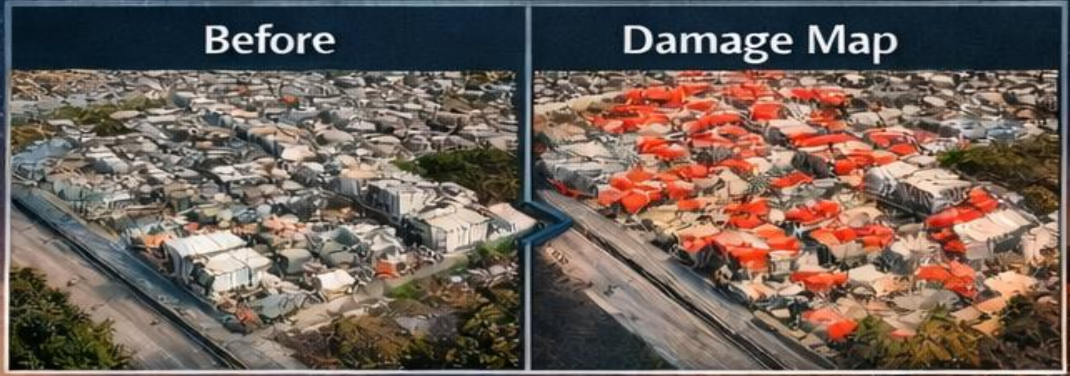
AI for Resilient Infrastructure:

Vision Transformers in Earthquake Damage Mapping



Amulya Sri Pulijala

Indian Space Research Organisation (ISRO), India

Presentation Road Map….

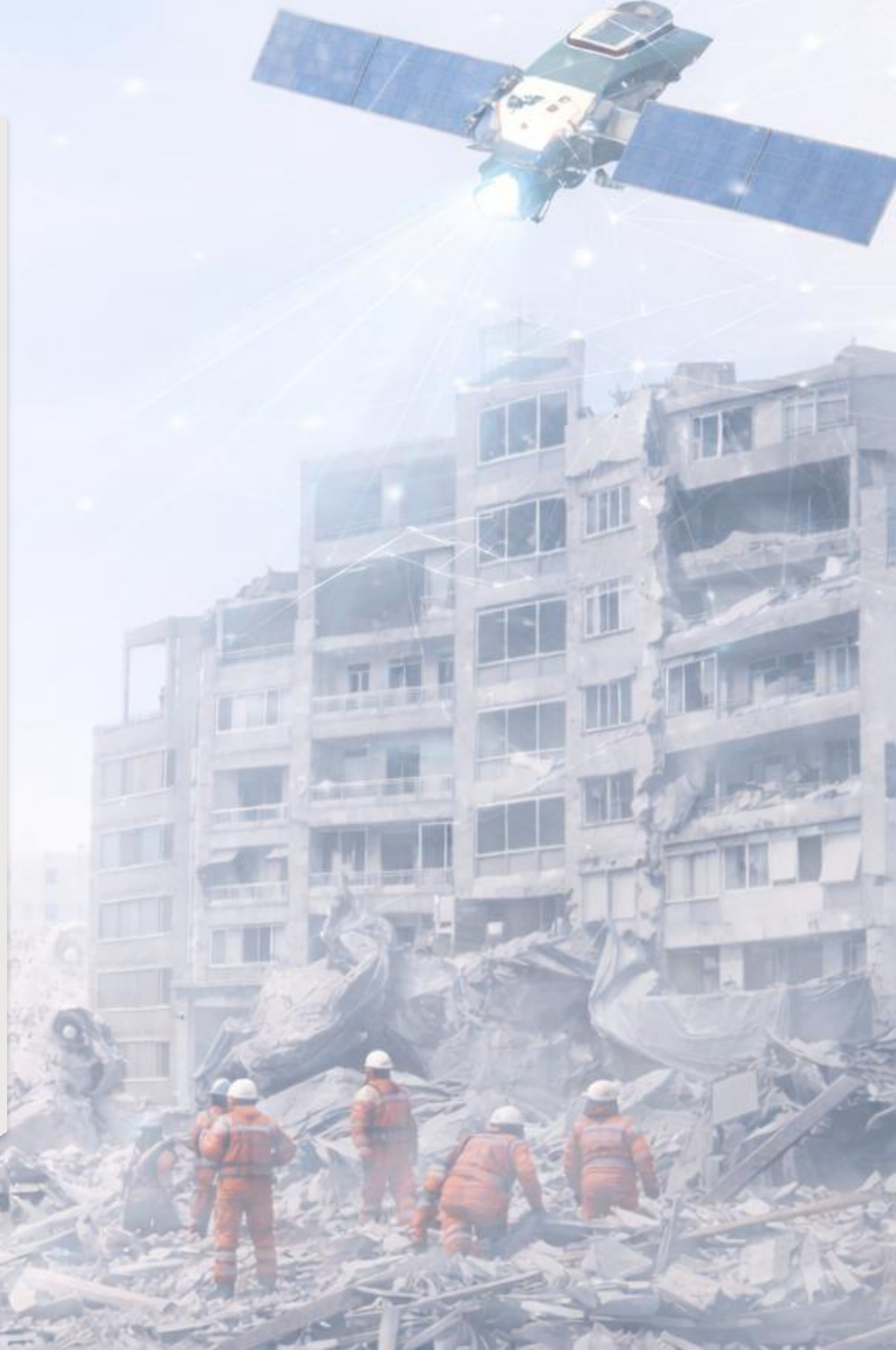
- Setting the Objectives
- Why Earthquake Damage Mapping Matters
- Seeing Disasters from Space : RS + AI for Rapid Damage Assessment
- Manual Assessment Vs AI at Scale
- Beyond CNNs
- The Rise of Vision Transformers
- Anatomy of ViT Model
- Self Supervised ViT : DinoV3
- From Theory to Hands-on Implementation

Setting the Objectives

- Understanding key challenges in rapid post-earthquake damage assessment
- Recognizing limitations of local-context learning in CNN-based models
- Appreciating the role of global attention in damage mapping
- Gaining insight into transformer-based learning for EO imagery
- Interpreting AI model decisions through explainability techniques
- Identifying characteristics of reliable and deployable damage assessment models

Why Earthquake Damage Assessment Matters

- Rapid identification of affected infrastructure
- Prioritization of rescue and emergency response
- Minimizing loss of life and economic impact
- Supporting data-driven decision making
- Enabling resilient recovery and reconstruction



Seeing Disasters from Space

- Wide-area monitoring across diverse disaster types
- Rapid assessment over large and inaccessible regions
- Consistent observations independent of ground conditions
- Multi-temporal views for change detection

Satellite imagery enables timely, large-scale situational awareness across disasters.

Seeing Disasters from Space:

Remote Sensing for Rapid Damage Assessment

Remote Sensing Capabilities

- Rapid, wide-area coverage
- Repeat observations (pre- and post-event)
- Consistent, sensor-based data acquisition

AI-Driven Enablement

- Automated analysis at scale
- Objective and repeatable damage assessment
- Near-real-time damage mapping

Operational Impact

- Faster situational awareness
- Improved prioritization of rescue and relief
- Reduced risk to human responders
- Scalable, data-driven decision support

From observation to actionable insight at scale……

Manual Assessment vs AI at Scale

Scale of Major Earthquakes

- Tens of thousands of buildings
- Hundreds of square kilometers
- Decisions within hours to days

Delayed Assessment Leads To

- Slower emergency response
- Higher casualty risks
- Extended recovery timelines

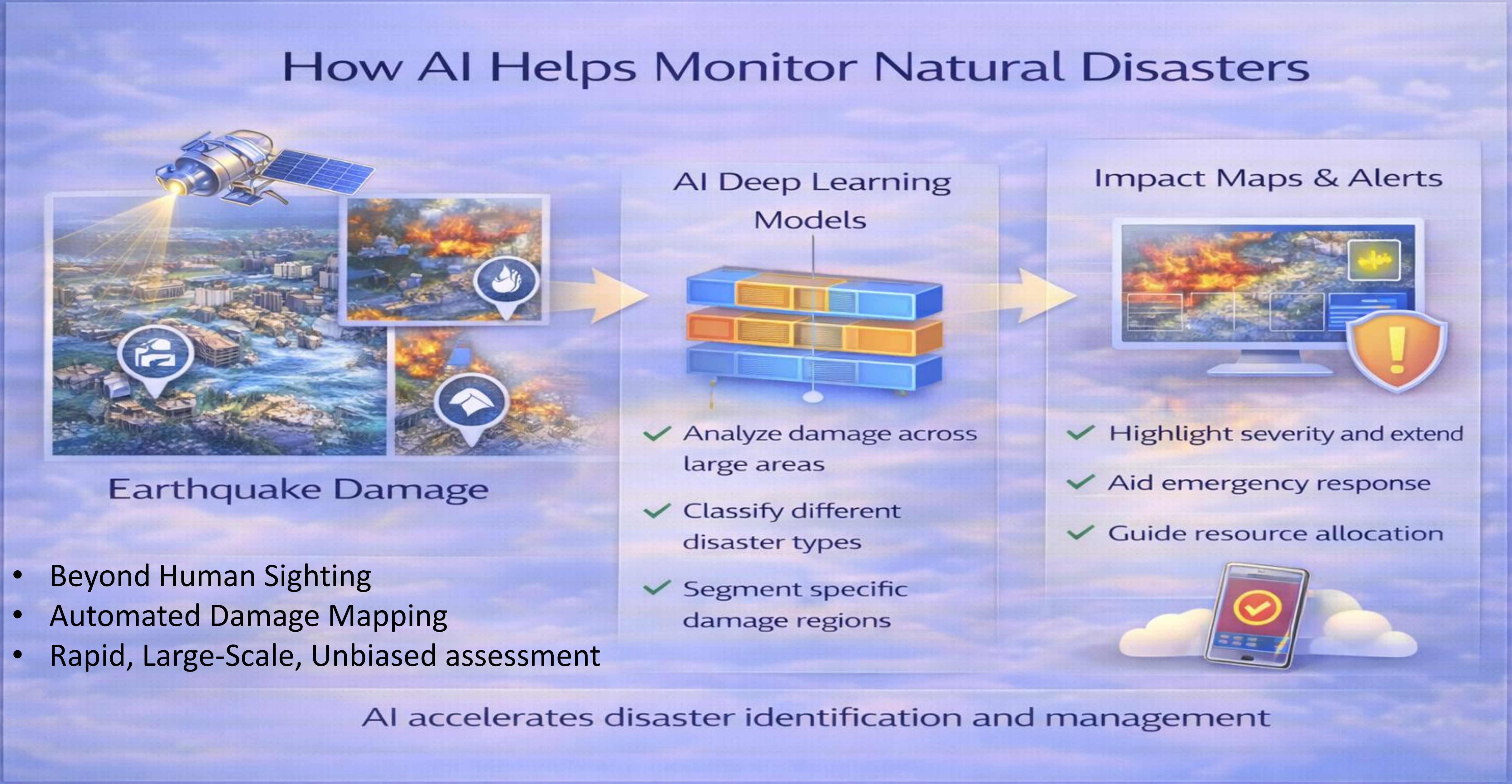
Constraints of Manual Assessment

- Time- and labor-intensive surveys
- Dependence on expert availability
- Limited scalability and subjectivity
- Accessibility and safety barriers

AI-Enabled Damage Assessment

- Rapid, large-area damage screening
- Consistent and objective analysis
- Near-real-time situational awareness
- Decision support for emergency response

Motivation for scalable, AI-driven damage mapping

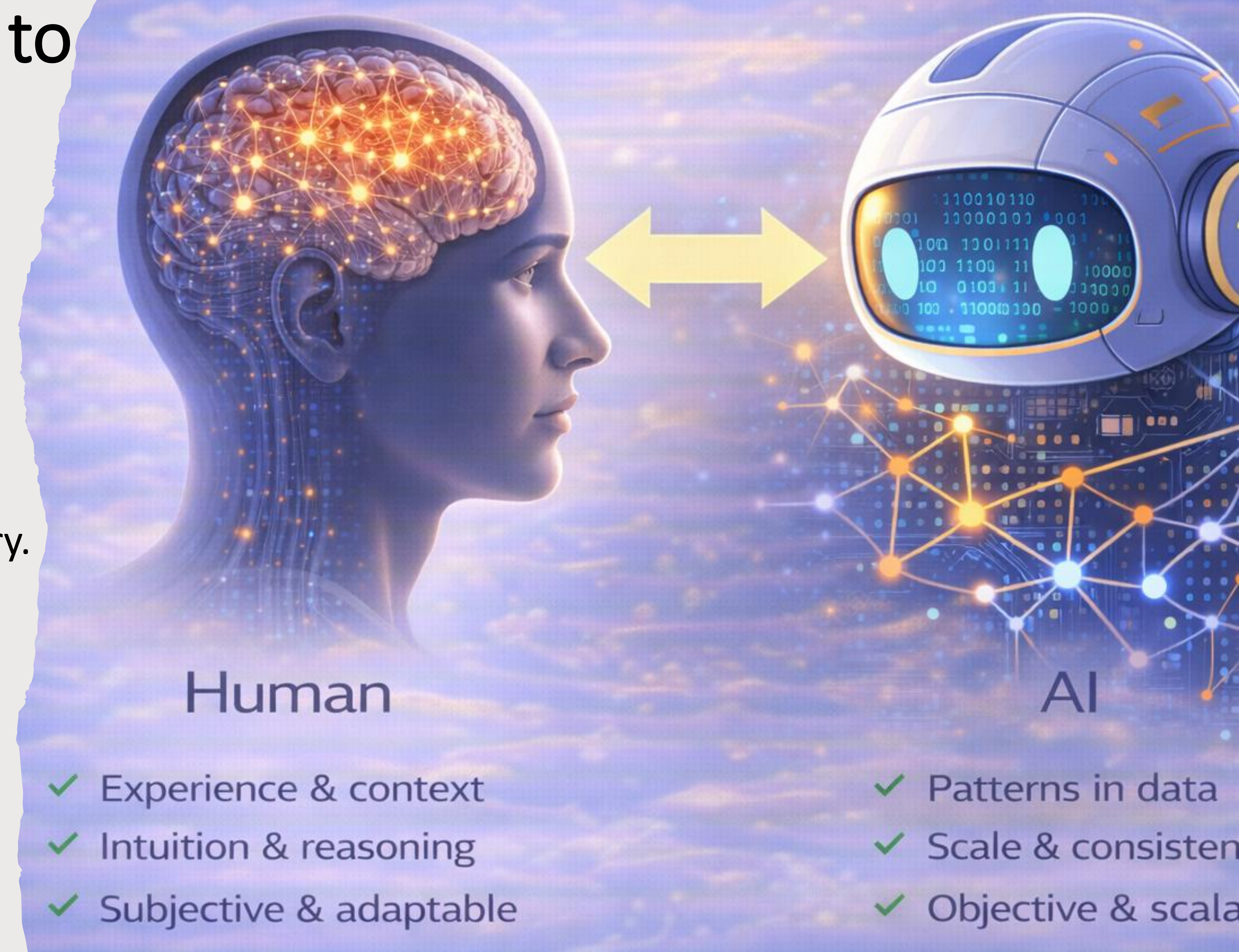


- Beyond Human Sighting
- Automated Damage Mapping
- Rapid, Large-Scale, Unbiased assessment

The goal of machine vision is not to replace human reasoning, but to scale it consistently.

From Human Perception to Artificial Intelligence

How machines learn to interpret visual Information…..



- AI decisions are data-driven, not arbitrary.
- Machine predictions reflect learned structure in data.

This shift from human perception to machine learning led to convolution-based models for visual understanding.

Convolutional Neural Networks (CNNs) in Damage Mapping

Why CNNs Became Popular

- ✓ Strong local feature extraction
- ✓ Translation invariance
- ✓ Effective for texture and edge detection
- ✓ Well-suited for pixel-level tasks

Common CNN Architectures

- ✓ ResNet – deep feature learning
- ✓ U-Net – dense prediction and segmentation
- ✓ DenseNet – feature reuse and efficiency

Common CNN Architectures

How CNNs Help in Disaster Assessment

CNNs provide a strong baseline for image-based damage analysis

Beyond CNNs

CNNs in Remote Sensing

- Dominant models for vision-based damage mapping
- Widely used architectures: ResNet, U-Net, DenseNet
- Effective for local feature extraction

Why This Matters for Damage Mapping

- Damage cues are spatially distributed
- Structural damage is context-dependent
- Local textures may be misleading

Inherent Limitations

- Local receptive fields by design
- Global context emerges only in deeper layers
- Limited modeling of long-range spatial dependencies

Resulting Challenges

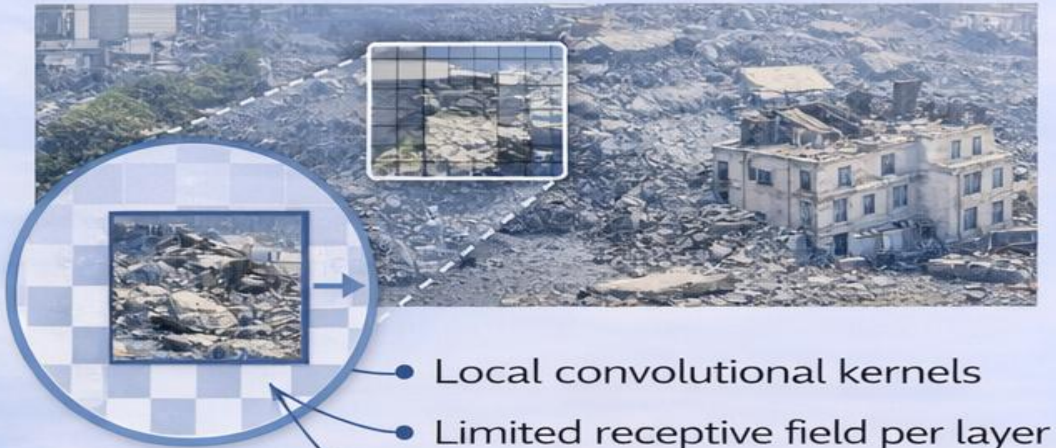
- Sensitivity to background clutter
- Misclassification of non-damage patterns
- Reduced robustness in complex urban scenes

Damage interpretation requires global structural context, not isolated local patterns.

The Rise of Vision Transformers

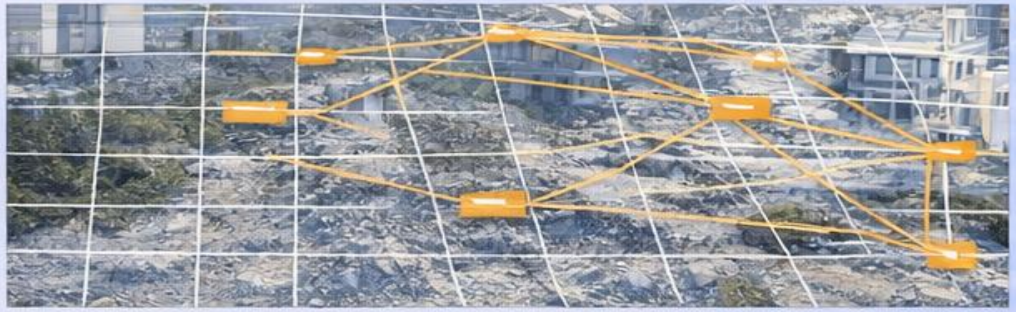
Why Global Context Matters for Damage Mapping

CNNs



- Local convolutional kernels
- Limited receptive field per layer
- Global context captured indirectly
- Sensitive to local textures and clutter

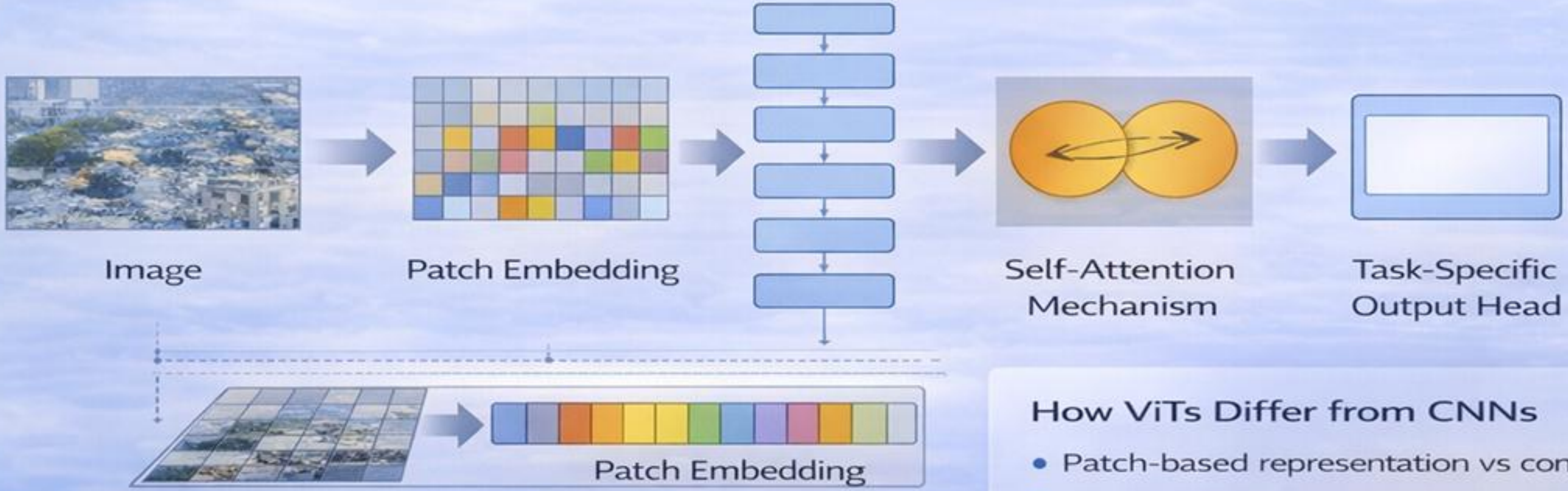
Vision Transformers



- Image represented as patches
- Self-attention across all patches
- Explicit modeling of long-range dependencies
- Structure-aware interpretation

From local texture recognition to global structural understanding

Anatomy of a Vision Transformer



How ViTs Differ from CNNs

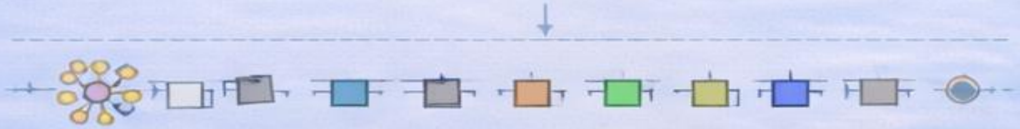
- Patch-based representation vs convolutional kernels
- Global self-attention vs local receptive fields
- Explicit long-range dependency modeling
- Context-aware feature learning from early layers

From Image to Tokens

- Image divided into fixed-size patches
- Flattening and linear projection
- Patch embeddings as input tokens
- Positional encoding for spatial awareness

Patch Embeddings

Patch Embeddings

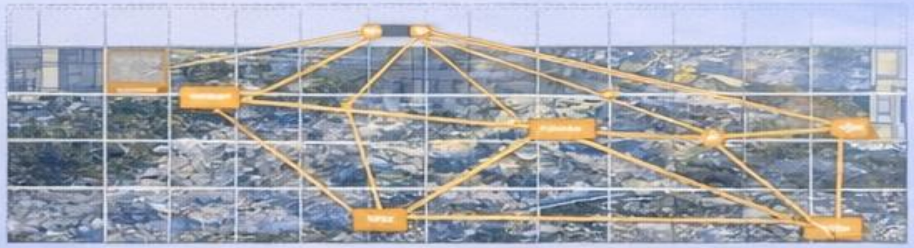


Positional Encoding

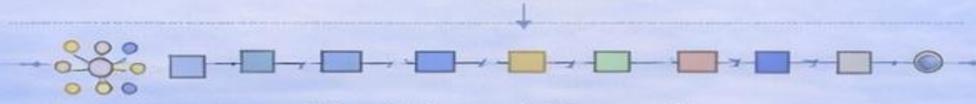
Self-Attention & Global Context

- Self-attention across all image patches
- Modeling long-range spatial dependencies
- Context-aware feature representation
- Multi-head attention for diverse relationships

Patch Embeddings



Self-Attention Mechanism



Positional Encoding

"Vision Transformers are powerful — but how do we train them when labels are scarce?"

From Vision Transformers to **Self-Supervised Learning**

Vision Transformers Enable

- Global, context-aware representation learning
- Long-range spatial reasoning

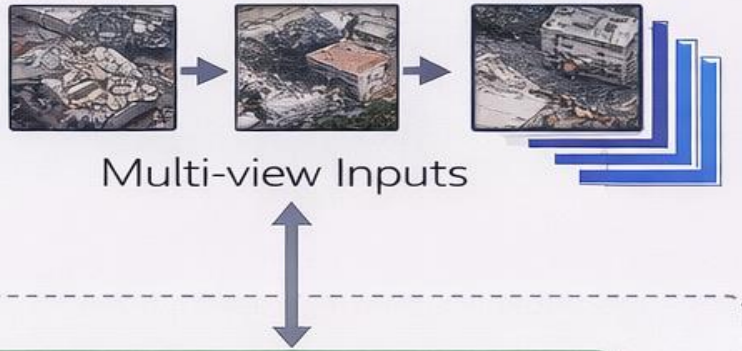
The Practical Challenge

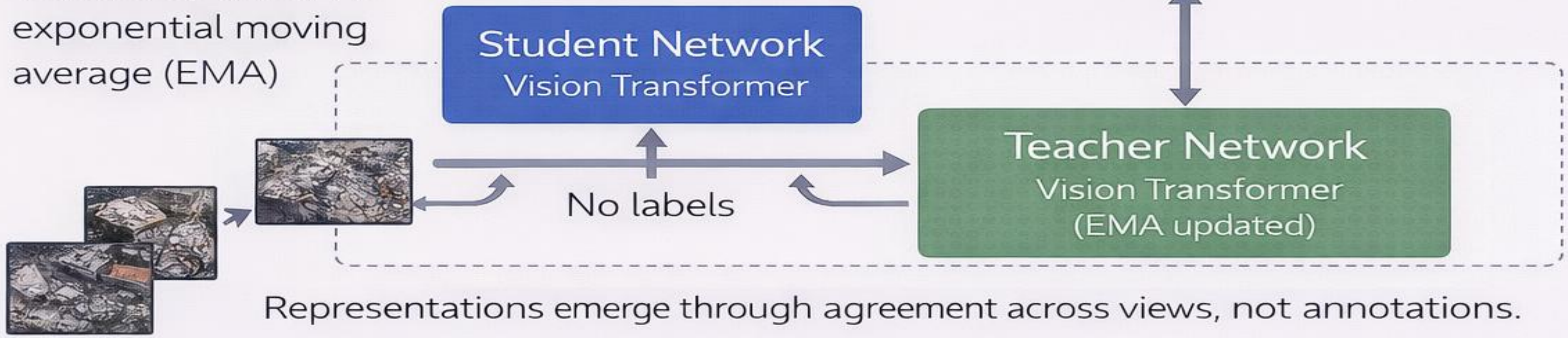
- Large-scale labeled data is **limited**
- Manual annotation is **costly and subjective**
- Disaster scenarios vary across regions and events

How can Vision Transformers learn robust representations without relying on extensive labels?

DINOv3: Self-Supervised Vision Transformer Architecture

- Teacher—student Vision Transformer framework
- Self-distillation without labels
- Multi-view consistency learning
- Teacher updated via exponential moving average (EMA)





Representations emerge through agreement across views, not annotations.

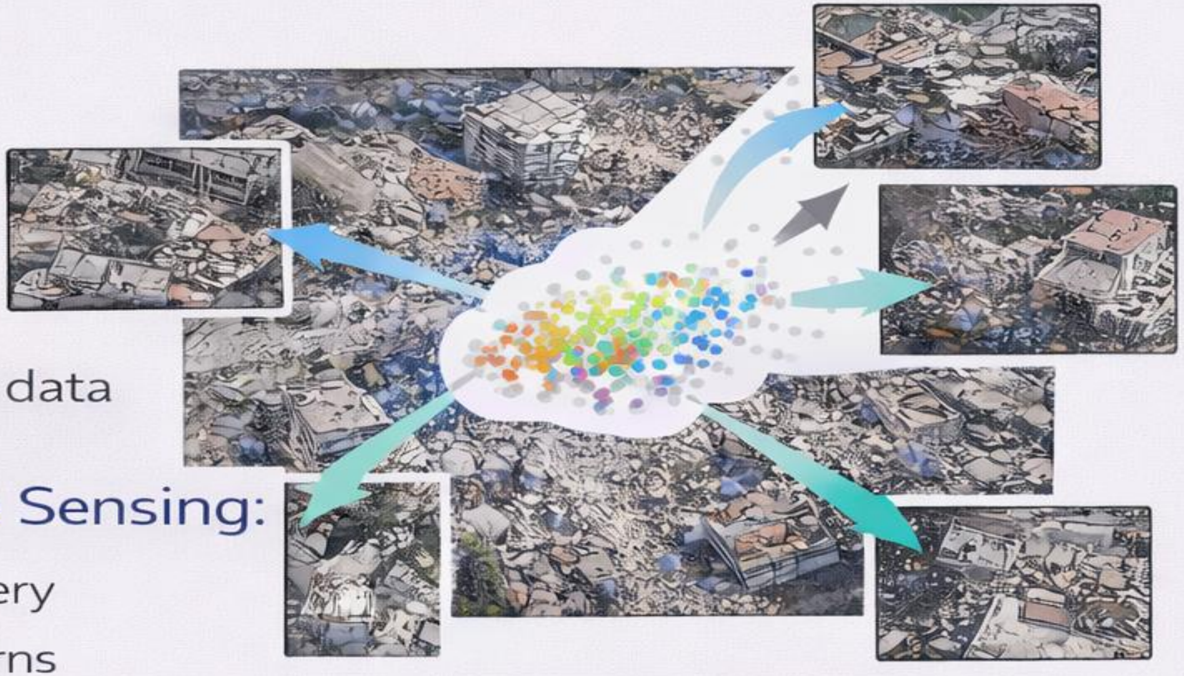
Representations emerge through agreement across **views, not annotations.**

Caron et al., "DINOv3: Self-Supervised Vision Transformers with Global Representations," Meta AI, 2024.

How DINOv3 Works Without Labels

Key Principles:

- Different views of the same image share semantics
- Consistency across views replaces supervision
- Structure emerges naturally from data



Why This Works for Remote Sensing:

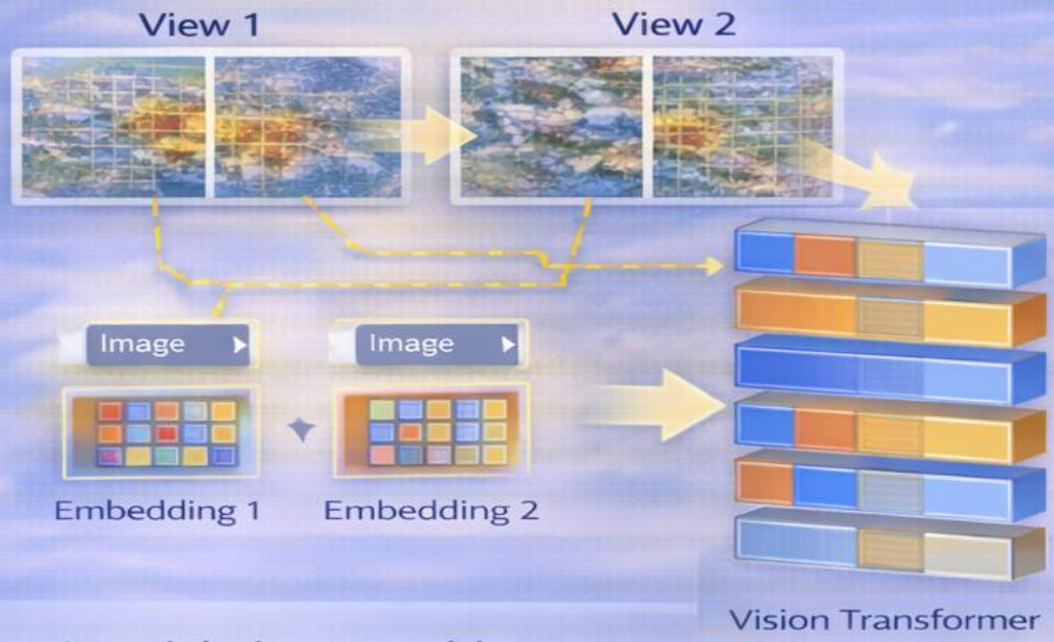
- Abundant unlabeled satellite imagery
- Strong spatial and structural patterns
- Reduced annotation bias
- Better cross-region generalization

DINOv3 learns structure first – labels come later.

Caron et al., "DINOv3: Self-Supervised Vision Transformers with Global Representations," Meta AI, 2024.

Understanding DINOv3

Self-Supervised Transformers for Deep Image Representations

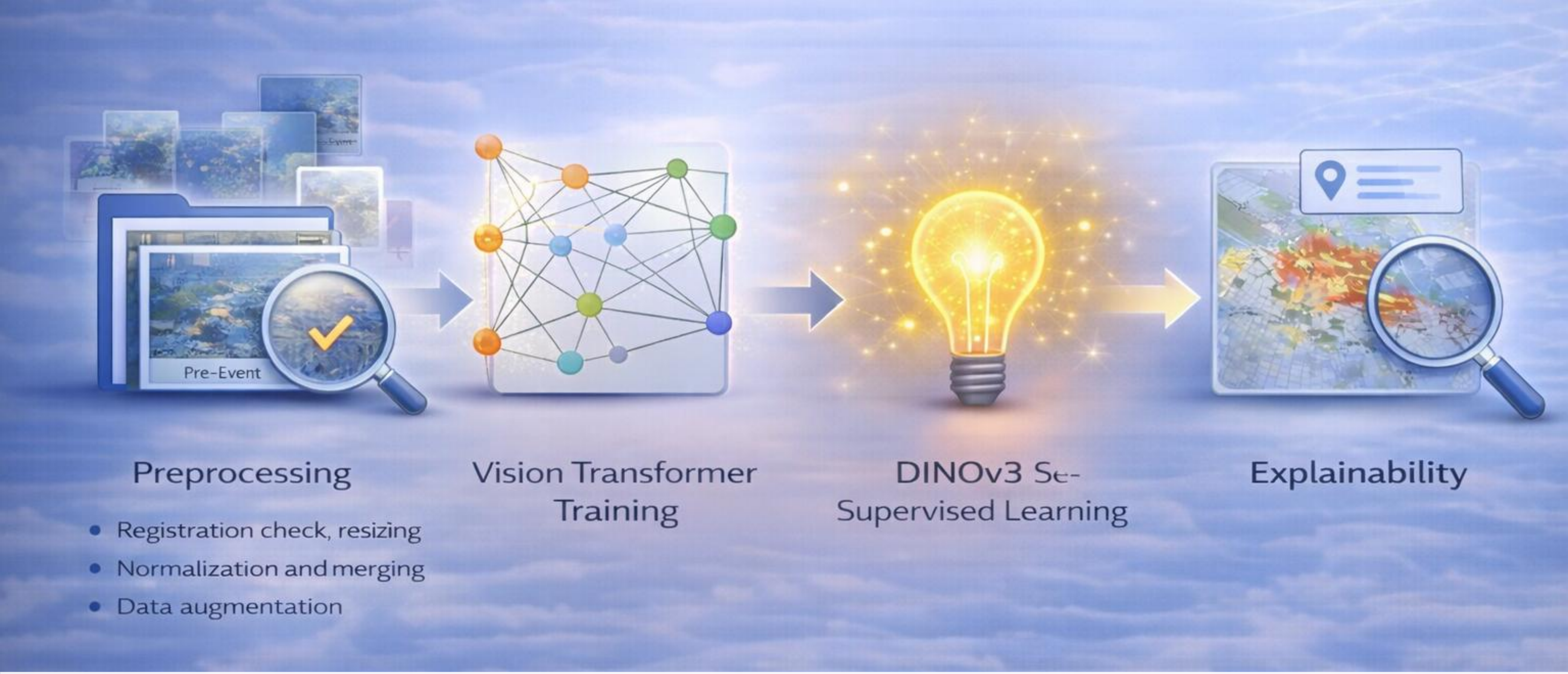


- ✓ No labels, no problem
- ✓ Learn powerful features with DINOv3
- ✓ Foundational model for ViTs

Unlocking the Power of DINOv3

Zero-shot and low-shot learning in satellite imagery

Technology Stack



From Theory to Hands-On Implementation...

Design Choices in the Hands-On Workflow

Key Design Decisions

- Patch-based representation over pixel-level modeling
- Pre-post image pairing for change awareness
- Self-supervised learning to address label scarcity
- Attention-based explainability for trust

What the Hands-On Will (and Will Not) Cover

Will Cover

- Dataset structure and loading
- Preprocessing steps
- ViT and DINOv3 workflow
- Explainability outputs

Will Not Cover

- Full model training from scratch
- Hyperparameter tuning
- Infrastructure optimization
- Benchmark comparisons

Lightweight, reproducible, and research-oriented stack

Live Demonstration Using a Reproducible Google Colab Notebook

Hands-On Tutorial Overview

- Google Colab–based execution (no local installation required)
- Publicly available earthquake damage datasets
- End-to-end workflow: preprocessing → ViT → DINOv3 → explainability
- Fully reproducible and reusable notebook

Open data ; Open methods ; Reproducible practice

Hands-On Tutorial Environment

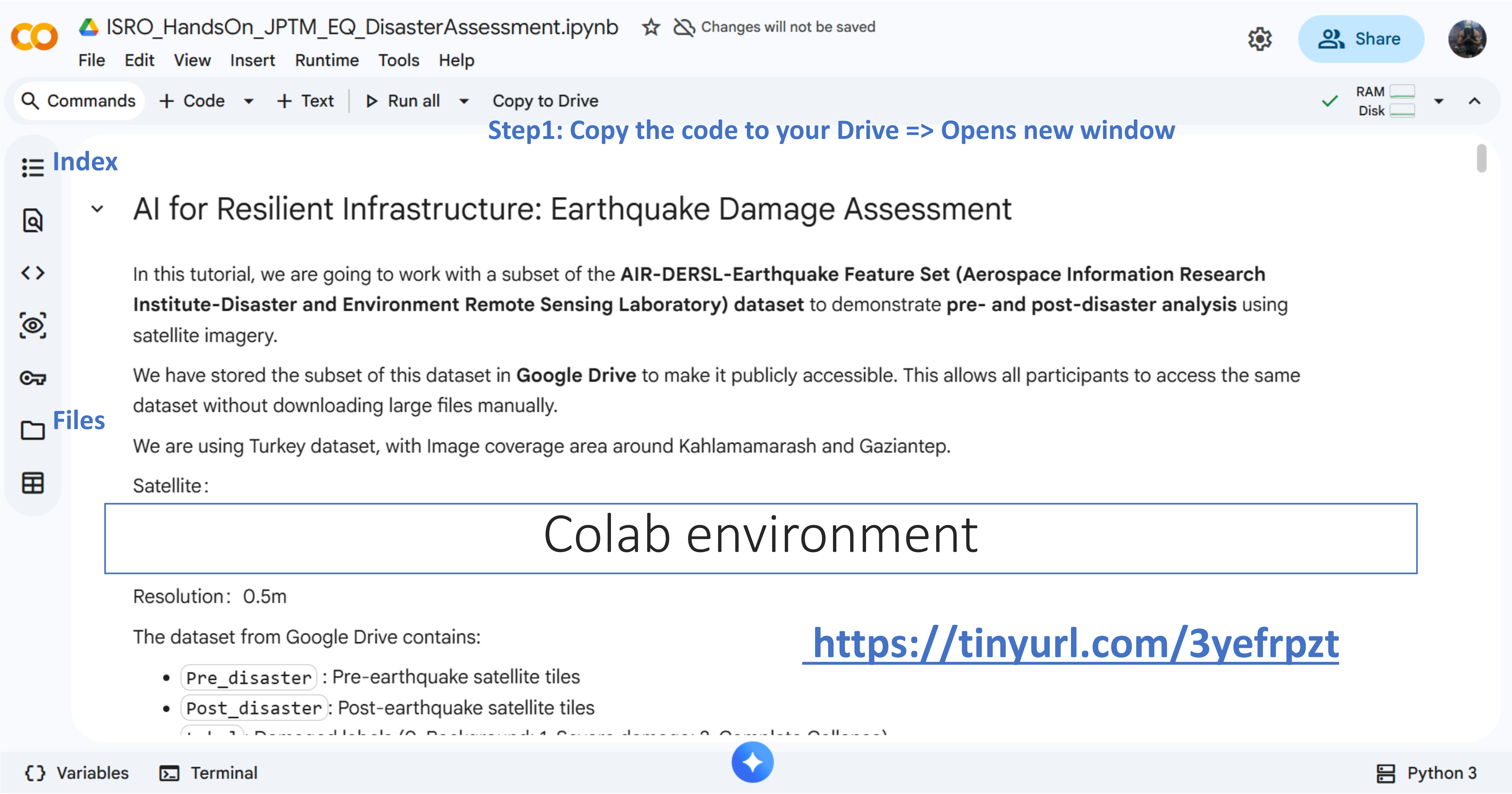
Why Google Colab

- No local installation required
- Runs entirely in browser
- Preconfigured Python environment
- Accessible from anywhere

What You Will See

- Publicly available imagery
- End to End workflow execution
- Vision Transformer Configuration, Training
- Explainable AI

https://tinyurl.com/3yefrpzt


ISRO_HandsOn_JPTM_EQ_DisasterAssessment.ipynb
Changes will not be saved
File Edit View Insert Runtime Tools Help
Share
Commands
+ Code
+ Text
Run all
Copy to Drive
RAM
Disk
Step1: Copy the code to your Drive => Opens new window
Index
Files
AI for Resilient Infrastructure: Earthquake Damage Assessment
In this tutorial, we are going to work with a subset of the AIR-DERSL-Earthquake Feature Set (Aerospace Information Research Institute-Disaster and Environment Remote Sensing Laboratory) dataset to demonstrate pre- and post-disaster analysis using satellite imagery.
We have stored the subset of this dataset in Google Drive to make it publicly accessible. This allows all participants to access the same dataset without downloading large files manually.
We are using Turkey dataset, with Image coverage area around Kahlamamarash and Gaziantep.
Satellite:
Colab environment
Resolution: 0.5m
The dataset from Google Drive contains:
Pre_disaster : Pre-earthquake satellite tiles
Post_disaster: Post-earthquake satellite tiles
https://tinyurl.com/3yefrpzt
Variables
Terminal
Python 3

Colab Notebook

https://tinyurl.com/3yefrpzt

From Experiments to Principles

What the Hands-On Demonstrated

- Model behavior is shaped by training data
- Architecture influences how damage is interpreted
- Failures reveal limitations, not just errors
- Explainability is essential for trust

Hands-on experiments reveal not just results, but underlying principles.

How Training Data Shapes Machine Outcomes

Data Matters More Than Models

- Diverse training data improves generalization
- Rare damage classes require explicit exposure
- Augmentation simulates unseen conditions
- Bias in labels propagates to predictions

Practical Implication:

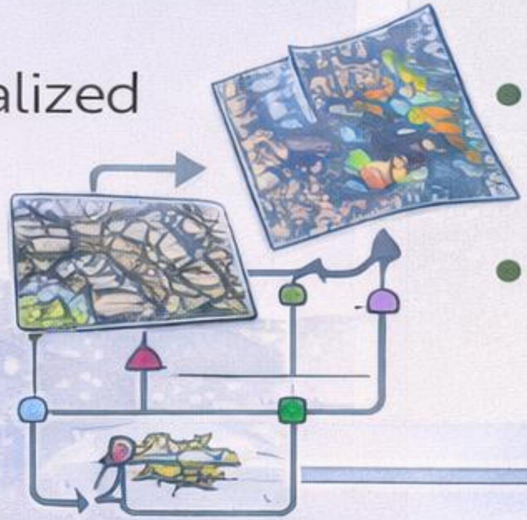
Models learn what they see — and fail where they haven't.

Caron et al., Meta AI (2024).

Siamese Models vs Vision Transformers

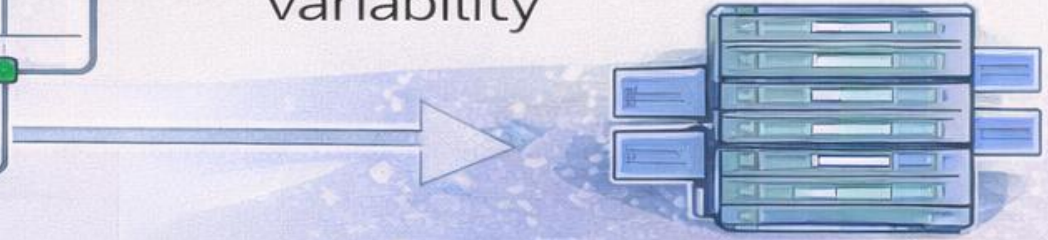
Siamese Architectures

- Explicit pre–post change modeling
- Effective for localized differences
- Sensitive to registration and noise.



Vision Transformers

- Global context and structural reasoning
- Learn implicit change representations
- More robust to spatial variability



Siamese models learn change explicitily — ViTs learn structure implicitly.

What Makes a Model Win in Practice

Winning Is Not Just Accuracy

- Balanced data exposure (oversampling rare damage)
- Robust training via augmentation
- Stable predictions through ensembles
- Consistent preprocessing and normalization

Learning the Right Signal

- Pansharpening enhances structural detail
- Siamese + ViT models learn change, not pixels
- Synthetic data helps when labels are scarce

Strong data, *smart training, and interpretable* models win in the real world.

Challenges and Future Directions

Current Challenges

- Label scarcity and inconsistency
- Generalization across regions and sensors
- Ambiguous or partial damage patterns
- Timeliness vs accuracy trade-offs

Why Explainability Matters

- Builds trust with decision-makers
- Identifies failure modes
- Supports responsible deployment

Looking Ahead

- Foundation models for Earth Observation
- More self-supervised and multimodal learning
- Human–AI collaboration in disaster response.

The future is not better models alone, but better understanding of models.

Thank you!

Any questions?

amulyasri.pulijala@gmail.com
amulya_sri@nrsc.gov.in