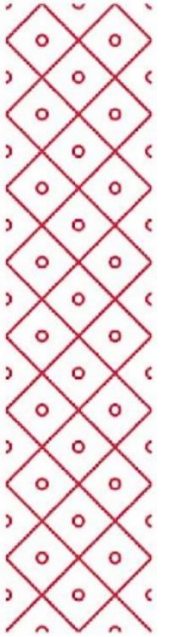


Advancing Remote Sensing with Deep Learning

Challenges, Innovations, and Future Directions

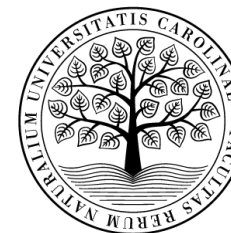


LUKÁŠ BRODSKÝ

DEPARTMENT OF APPLIED GEOINFORMATICS AND
CARTOGRAPHY, CHARLES UNIVERSITY

KEYNOTE SPEECH

44TH EARSEL SYMPOSIUM, PRAGUE 28/MAY/2025



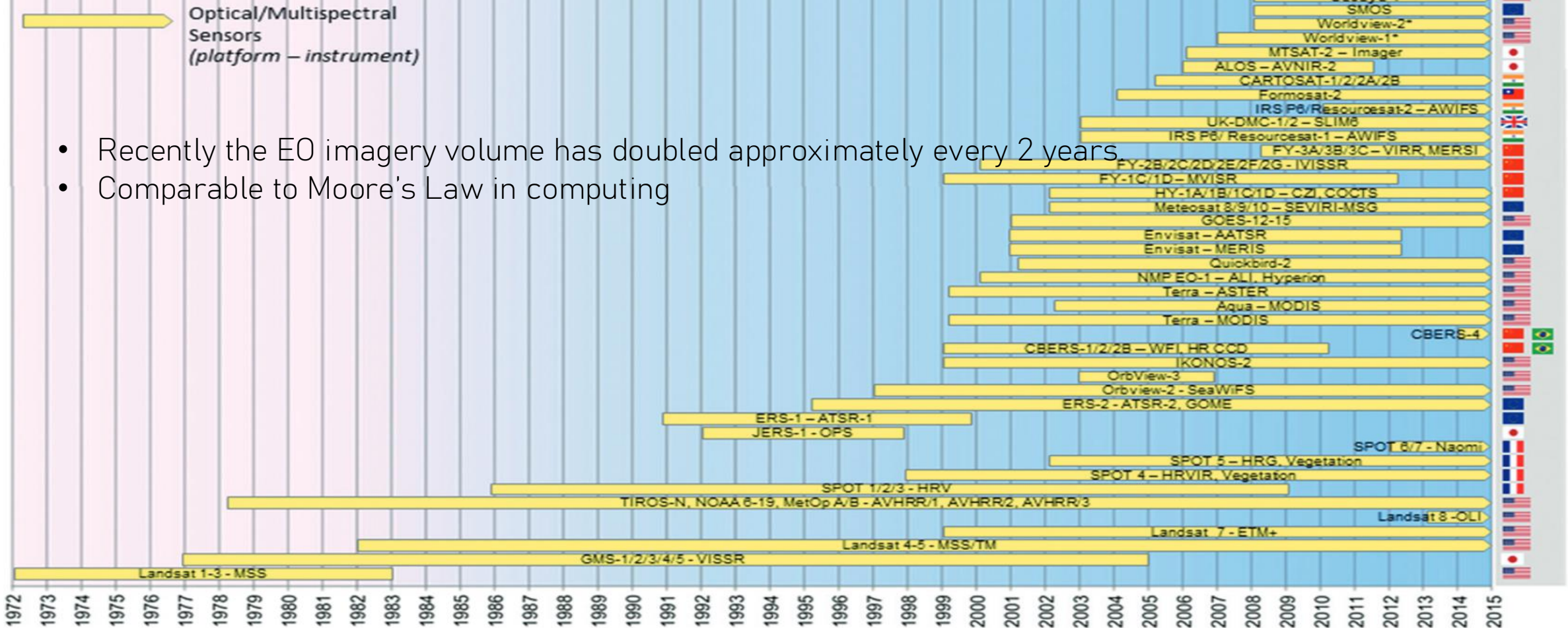
Faculty of Science
Charles University


Motivation

- Earth Observation (EO) data is growing at an unprecedented pace.
- Deep Learning (DL) has transformed AI—can it do the same for Remote Sensing?

Explosion of EO Data

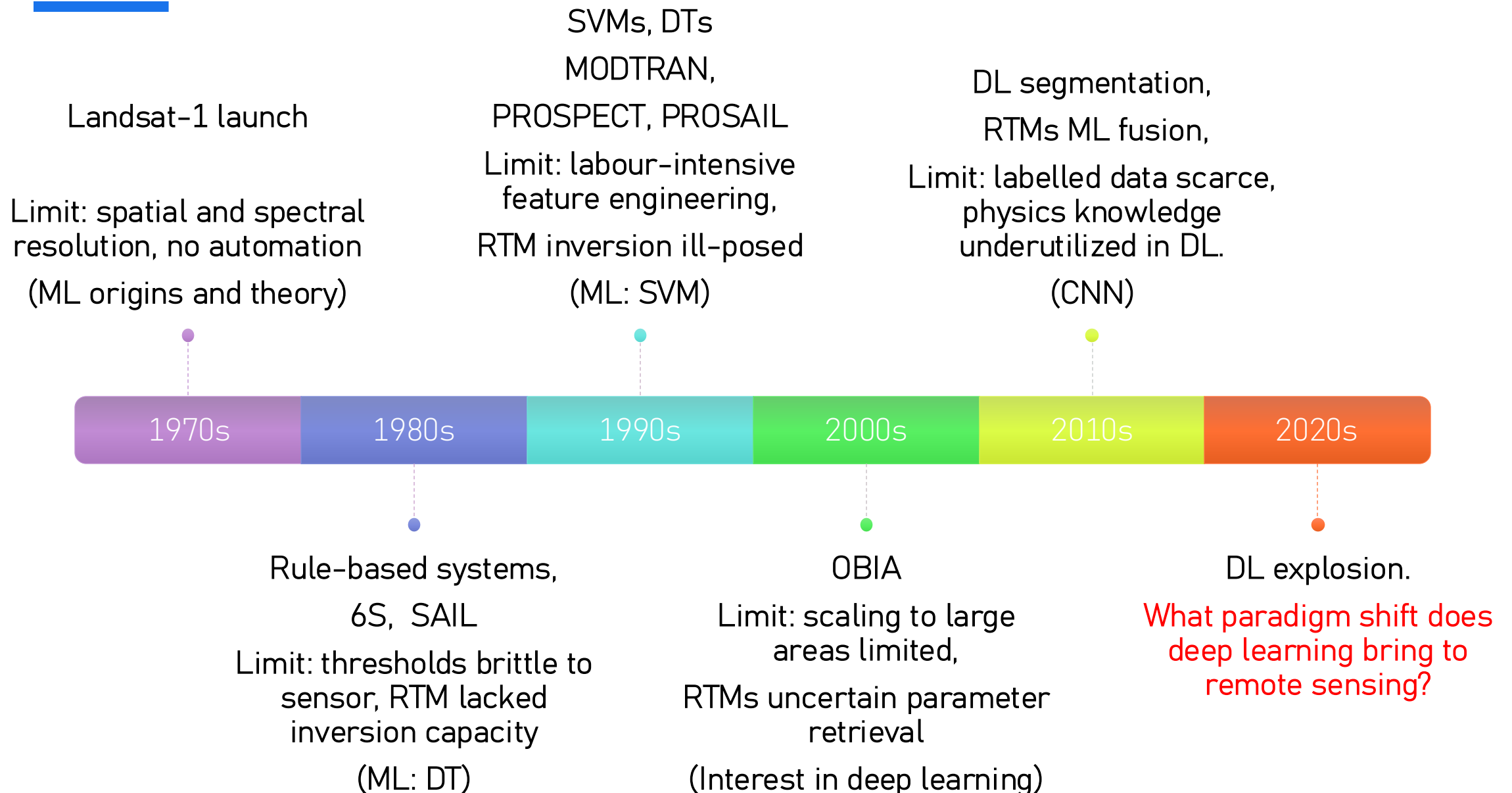
- Today, there are over hundreds active sensors, and the number keeps rising.
- Satellites capture vast amount of images daily—far beyond what traditional methods can handle.





The challenge is not collecting data.
It is making sense of the vast amount of
data we capture daily!

Traditional techniques face limits: scale, noise, and data variability.





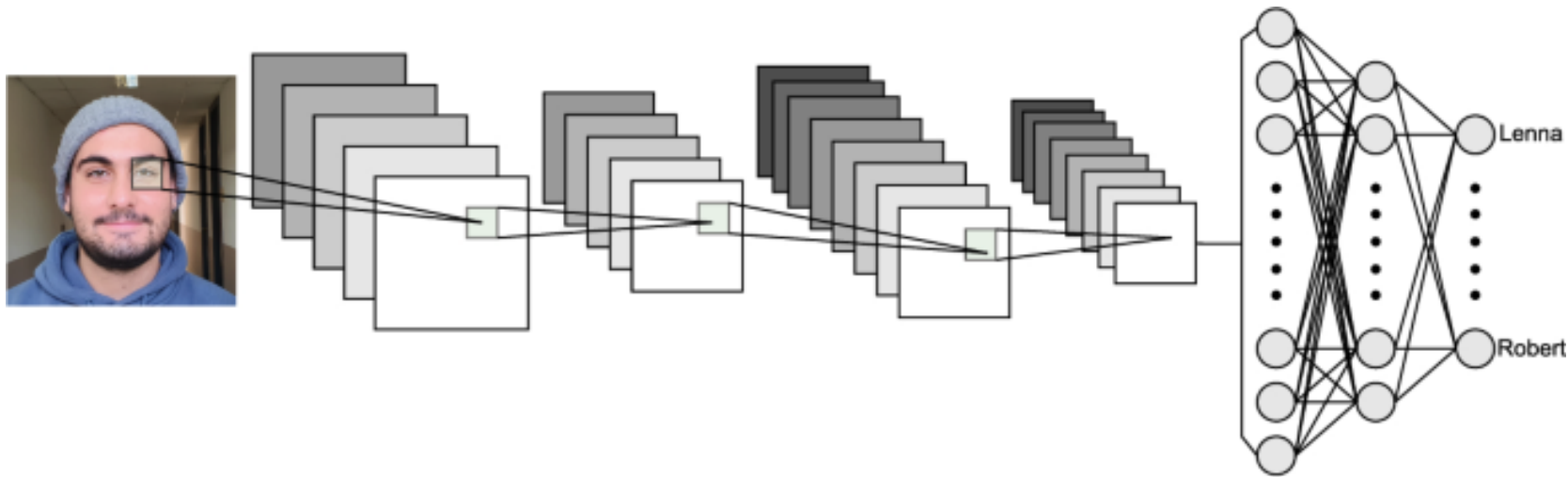
What paradigm shift does deep learning
bring to remote sensing?



Three main architectures.

Deep Learning Architecture Spotlight

- CNNs specialize in spatial feature extraction.

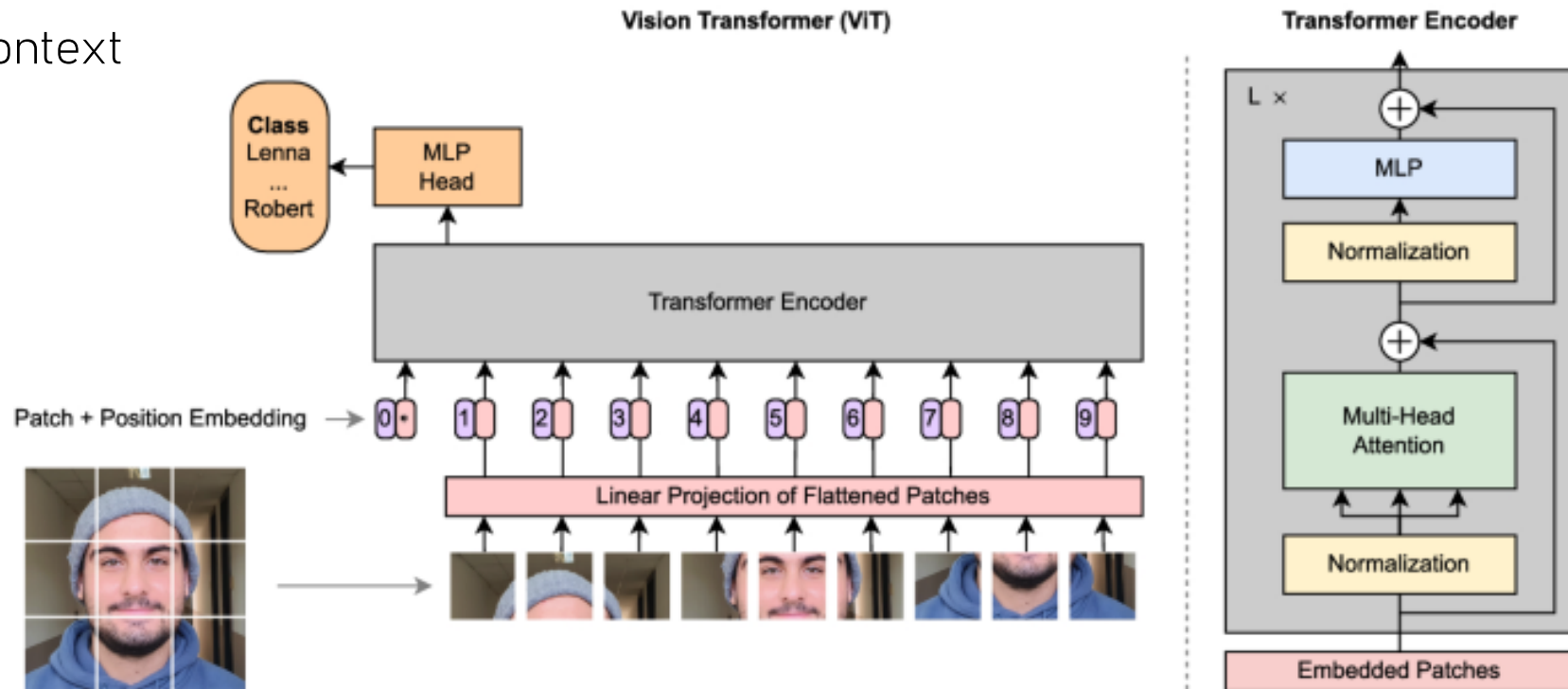


(a) Common CNN architecture

Rodrigo, M., Cuevas, C. & García, N. Comprehensive comparison between vision transformers and convolutional neural networks for face recognition tasks. *Sci Rep* 14, 21392 (2024). <https://doi.org/10.1038/s41598-024-72254-w>

Deep Learning Architecture Spotlight

- CNNs: Spatial feature extraction.
- Transformers (ViTs): Global context

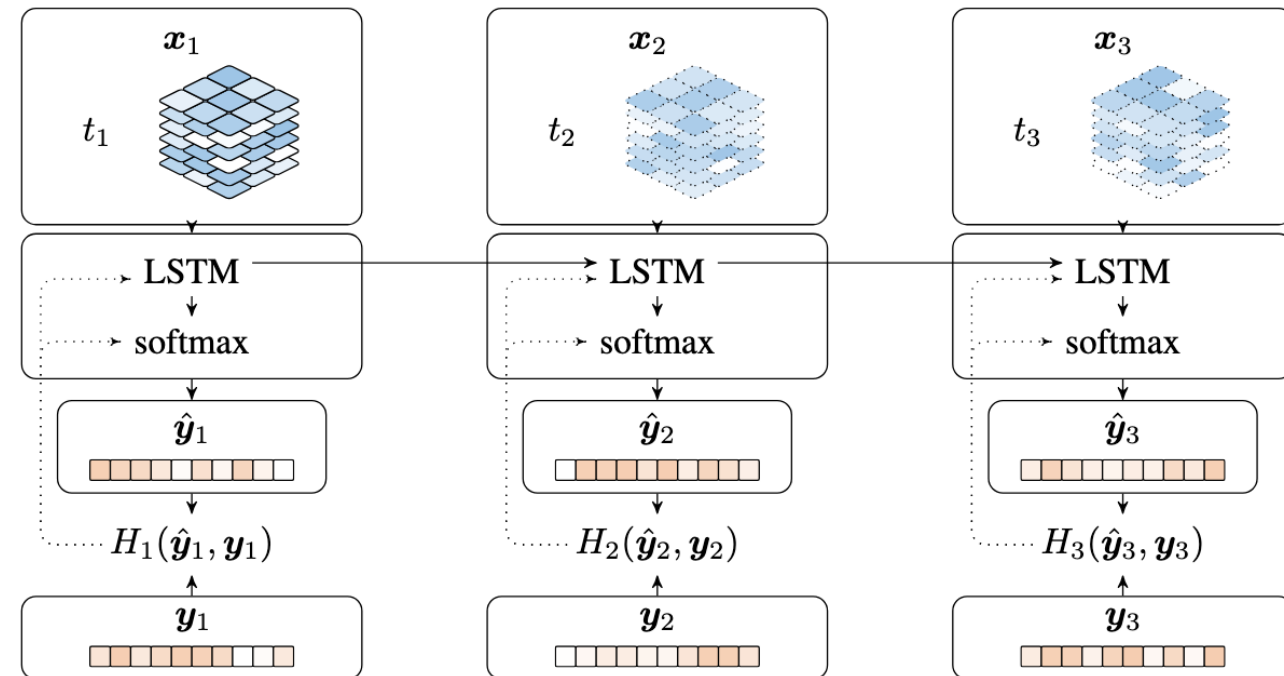


Rodrigo, M., Cuevas, C. & García, N. Comprehensive comparison between vision transformers and convolutional neural networks for face recognition tasks. *Sci Rep* 14, 21392 (2024). <https://doi.org/10.1038/s41598-024-72254-w>

Key Deep Learning Architectures in RS

- CNNs: Spatial feature extraction.
- Transformers (ViTs): Global context
- RNNs/LSTMs: Temporal pattern modelling.

Rußwurm et al. (2017)



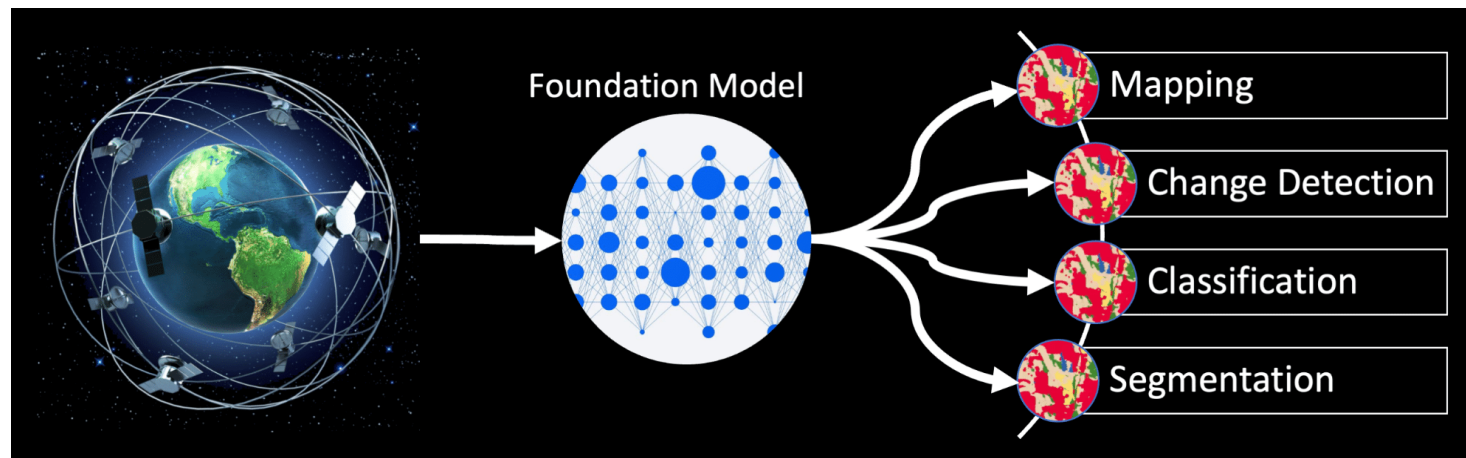


How do these architectures help to cope
with Big EO Data?

{volume, variety, value, veracity, and
velocity}

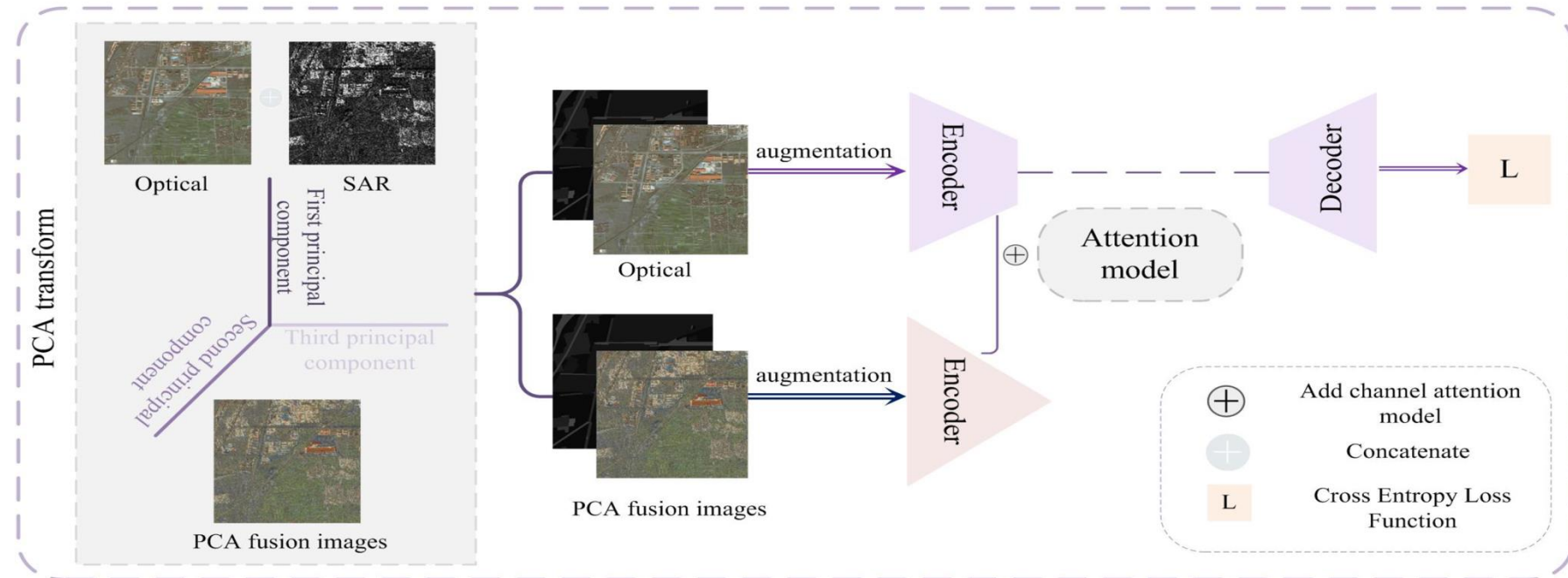
Problem: Data Volume & Model Transferability

- Can deep learning models trained on one region or sensor generalize to others?
- Solution: **Foundation Models**. Illustration: Szwarcman *et al.* 2024 (Prithvi-EO-2.0 model)
<https://www.clarku.edu/centers/geospatial-analytics/projects/prithvi-foundation-model>
- Prithvi-EO-2.0: A Versatile Multi-Temporal Foundation Model for Earth Observation Applications
- Improved performance on a range of environmental mapping tasks



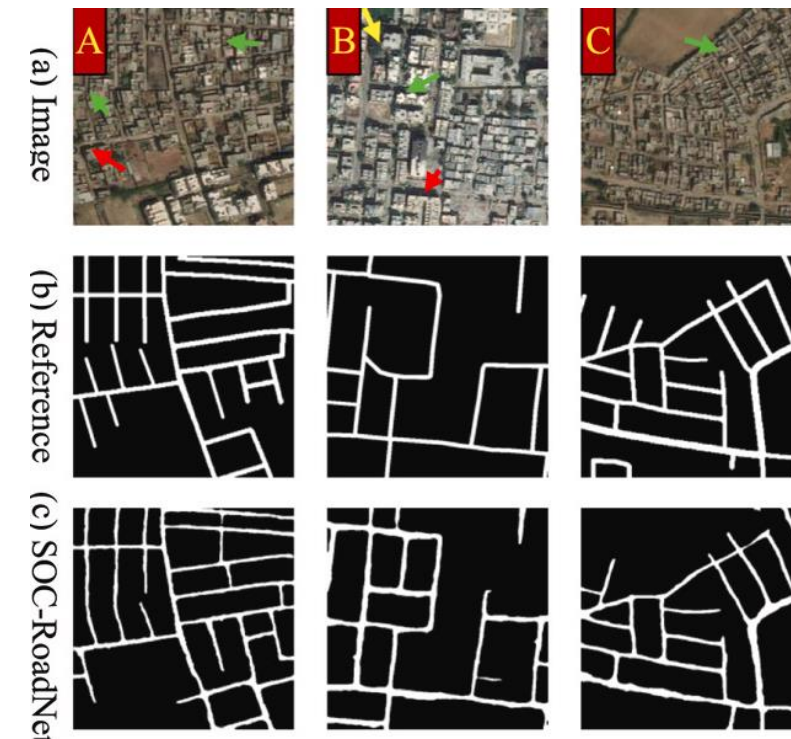
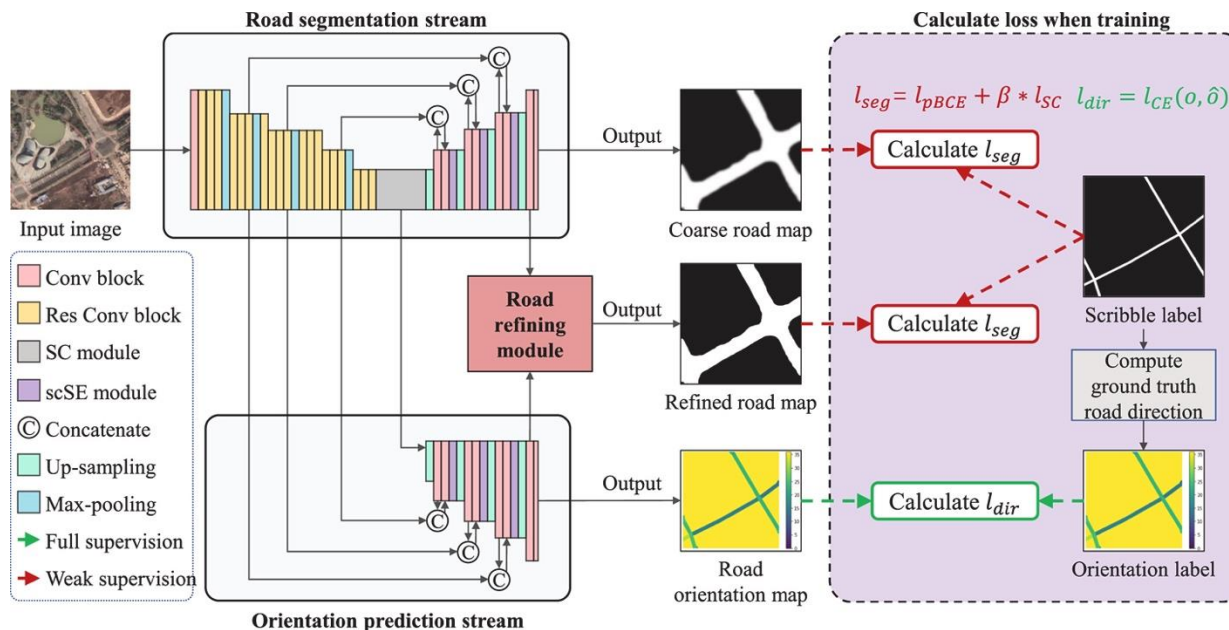
Problem: Sensor Variety (Optical, SAR, ...)

- Solution: **Multimodal fusion** models – combining sensors to compensate for weaknesses.
- Example: Learning SAR-Optical Cross-Modal Features for Land Cover Classification (Quan et al. 2024)



Problem: Value – Mapping Linear Features

- Solution: Convolutional model with **problem-specific loss function**
- RoadNet, Loss = $L_{BCE} + L_{SC}$ (Cross Ent. + Orientation prediction stream)
- Zhou et al. (2022) doi.org/10.1016/j.isprsjprs.2022.09.005

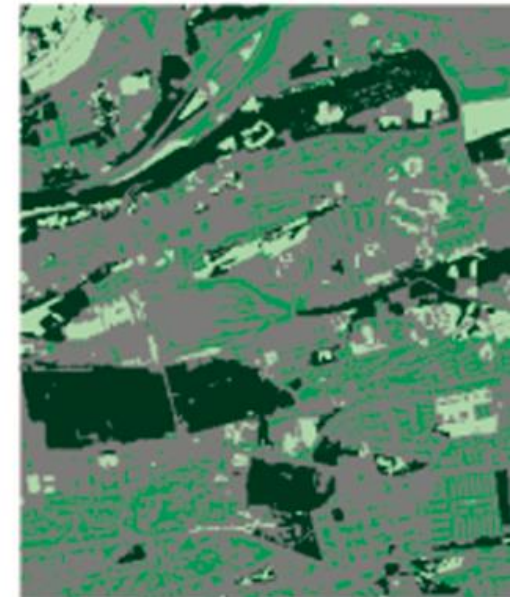


Problem: Value – Land Use mapping

- Solution: **local context modelling** (CNN)
- Pešek et al. (2024)
<https://doi.org/10.1016/j.rsase.2024.101238>
- Level-3 semantic segmentation of Urban Green land use

Legend

- non-vegetated
- non-recreational vegetation
- low recreational vegetation
- high recreational vegetation

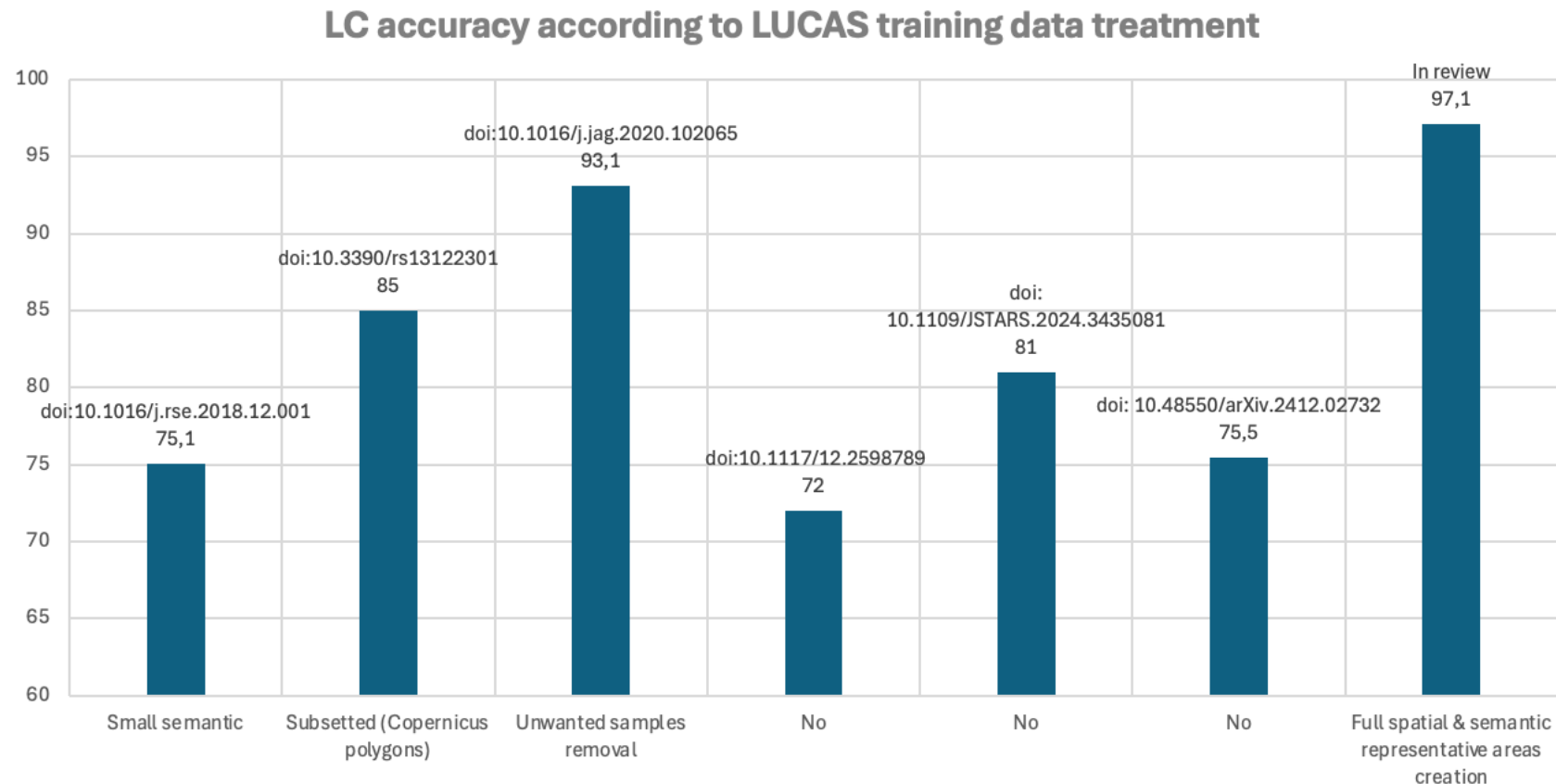


Problem: Veracity – Label Quality

- Solution: **Curated datasets** (e.g. European LUCAS for Land Cover Classification)

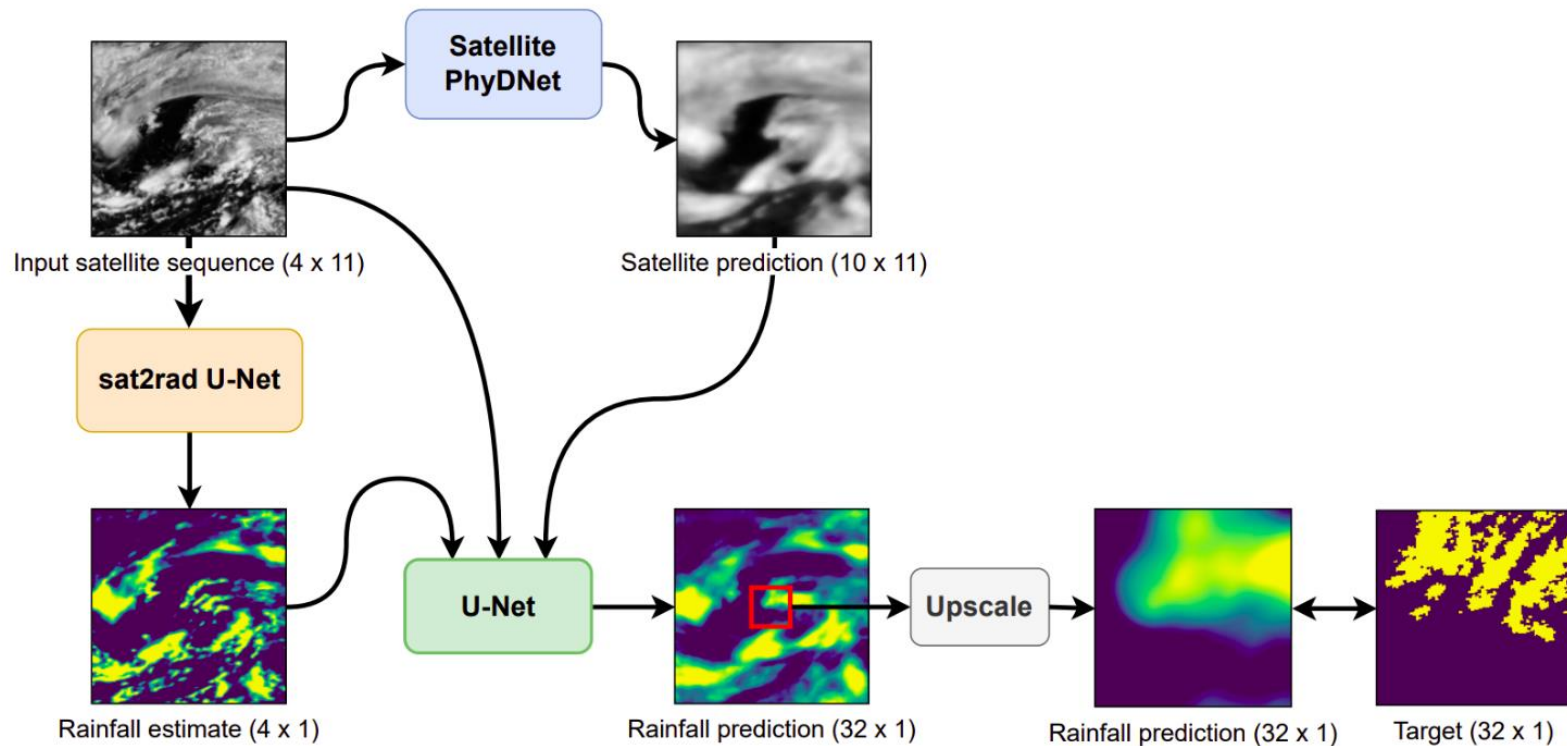
Prithovi-2.0 uses

Sen4Map (No quality
curation!)



Problem: Velocity – Dynamic Changes

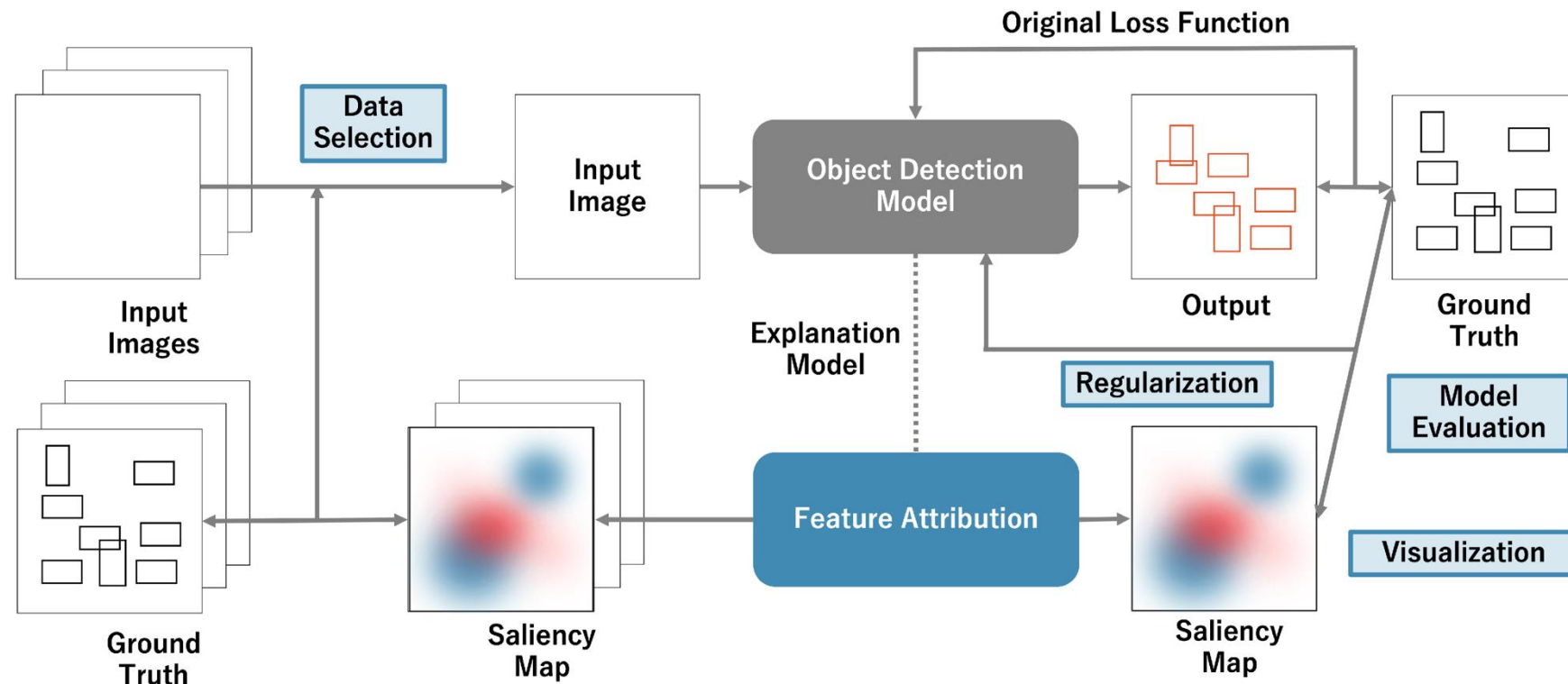
- *Challenge: Prediction of Precipitation from Satellite Data* (NeurIPS Weather4Cast Challenge)
- **Solution: Spatiotemporal DL and PhyDNet fusion:** Predicting Precipitation from Satellite Data.
Petr Šimánek (2022)



Problem: Explainability

- Solution: xAI techniques (e.g. LIME, SHAP, Saliency Map – visual explanations that rely on backpropagation of gradients)

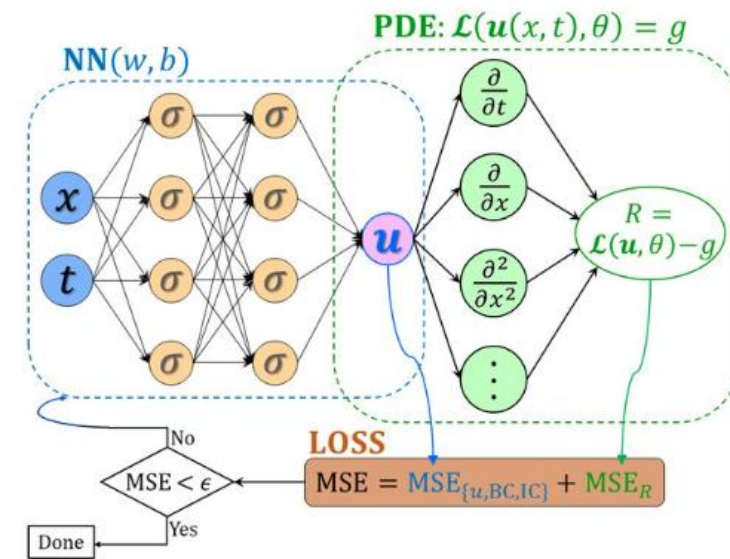
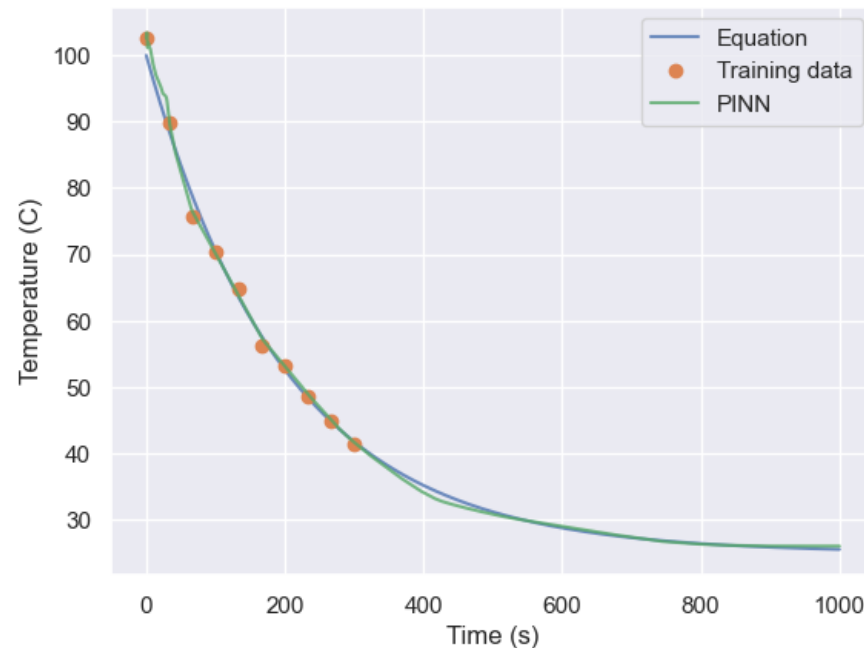
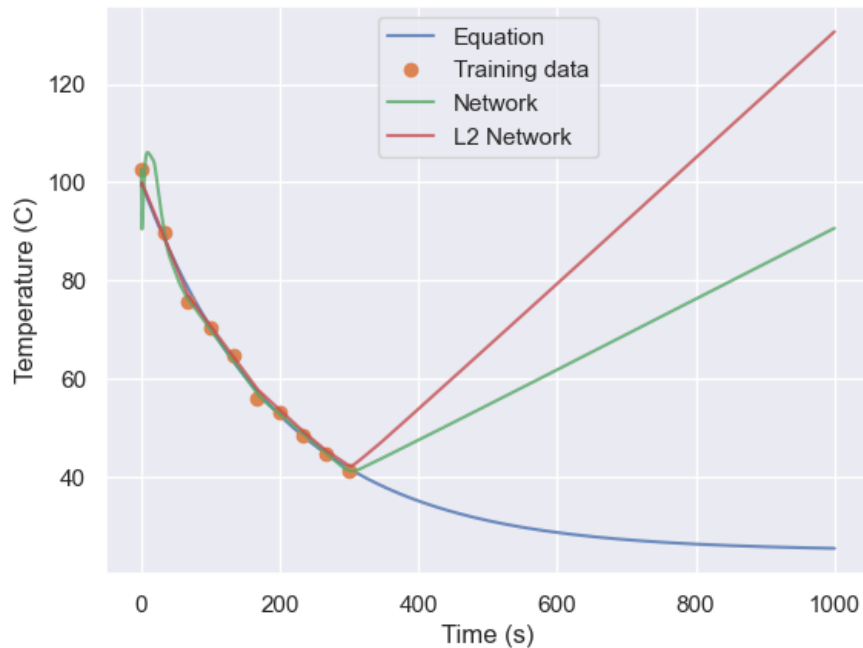
Kawauchi and Fuse (2022)



Problem: Physical plausibility

- Solution: Physics informed Neural Networks (PINN)

$$L = L_{data} + \alpha L_{physics}$$





Building Blocks

Key Deep Learning Blocks

Convolutional Block: extract spatial features from image data.

Residual Block: allows deep networks to learn identity mappings, improving gradient flow and convergence.

Generative Adversarial Network (GAN) Block: generate realistic data via adversarial training

Transformer Block: long-range dependencies and global context via self-attention

Contrastive Module: learn discriminative representations by comparing similar/dissimilar samples.

Recurrent Module: capture sequential dependencies in time series

Graph Neural Network (GNN) Block: relationships between non-grid-structured data

Physics-Informed Layers (PINNs, Physical Constraints)



Can we innovate by recombining
these blocks?



Innovation through Recombination?

- "There is a huge field of opportunity for everyone. As the number of small innovations—the building blocks of more innovations—grows, it becomes harder to predict which combinations will be transformative."
— *Brynjolfsson & McAfee, The Second Machine Age (2014)*
- Innovation in science and technology often emerges not from entirely new theories, but from **novel re-combinations of existing components**.
A striking example: **Kary Mullis' invention of PCR (Polymerase Chain Reaction)**, which revolutionized molecular biology, wasn't based on a new theory—but on an unexpected combination of known biochemical techniques.



Recombining the building blocks enables us to go beyond incremental gains.



Research Directions

- **1. Scalable Multimodal Models**—fuse data into shared representations

Direction: build multimodal EO foundation models trained on fused sensor data using contrastive learning and shared embedding spaces.

- **2. Incremental Learning for New Sensors**

Direction: develop continual learning frameworks that leverage transformer backbones.

- **3. Spatiotemporal Event Detection**—targeting floods, landslides, and glacier changes from EO streams.

Direction: combine RNNs or ConvLSTMs with spatial transformers to capture both local dynamics and global context in near-real-time EO pipelines.



Research Directions

- 4. Semi-Supervised Learning

Direction: use self-supervised and contrastive learning to pre-train models on global EO data and fine-tune with few-shot labelled samples.

- 5. Explainability

Direction: design explainable pipelines that combine physical constraints, saliency-based attribution, and uncertainty estimation.

Deep Learning

Stepping Stone or Destination?





Thank You for Your Attention!

- Github link or QR code to slides