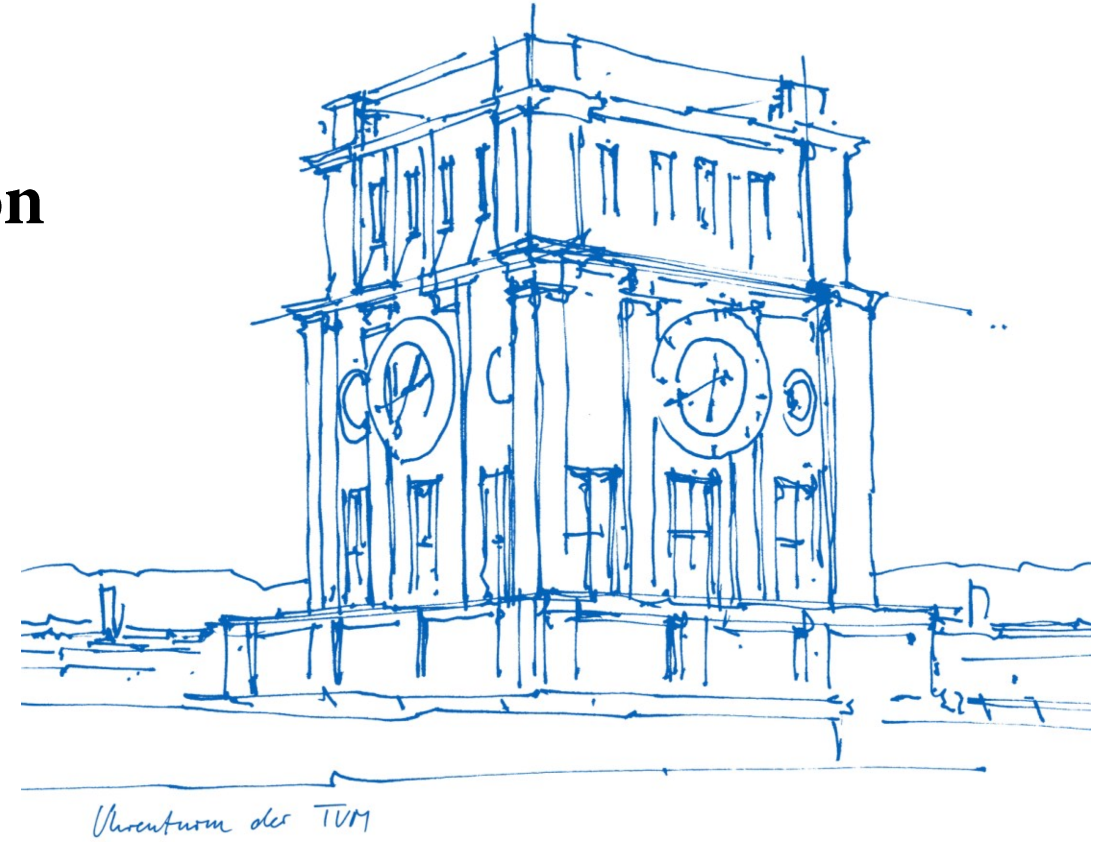
Technical University of Munich
Chair of Data Science in Earth Observation

Munich Center for Machine Learning

Easy and Efficient Fine-tuning of Foundation Models for Earth Observation

Sining Chen, Shan Zhao
{sining.chen, shan.zhao}@tum.de

Living Planet Symposium 2025
June 22, 2025



Who are we?

- Chair of Data Science in Earth Observation at the Technical University of Munich
- Led by Prof. Dr. –Ing Xiaoxiang Zhu

- Research Focuses
  - AI and data science in Earth Observation
  - Innovative signal processing and machine learning methods.
  - Big data analytics solutions to extract highly accurate large-scale geo-information.
  - Tackling societal grand challenges, e.g., global urbanization, UN's SDGs, and climate change.

Lab Website

Lab GitHub

Technical University of Munich
Chair of Data Science in Earth Observation

Agenda

- Introduction to Foundation Models
- Foundation Models for Earth Observation
  - Data
  - Training
  - Adaptation and Deployment
- Applications Cases and Hands-on Session
  - Wildfire Mapping
  - LCZ Classification
- Outlooks

Technical University of Munich
Chair of Data Science in Earth Observation

What are foundation models?

What are foundation models?

- Let's ask a foundation model.
- Key points
  - Large-scale model.
  - Vast amounts of data.
  - Fine-tuned for specific tasks and more.

Foundation models are a type of large-scale language model that serves as a starting point or foundation for various natural language processing (NLP) tasks. These models are typically pre-trained on vast amounts of text data and can be fine-tuned for specific tasks like text classification, language generation, question answering, and more.

The term "foundation model" emphasizes their role as foundational tools in NLP research and applications. They are designed to capture general language understanding and can be adapted or extended to handle a wide range of specific tasks with additional training on task-specific data. Examples of foundation models include OpenAI's GPT (Generative Pre-trained Transformer) models and Google's BERT (Bidirectional Encoder Representations from Transformers).

These models are crucial because they reduce the need for training large-scale models from scratch for every new task, thereby saving computational resources and time. They provide a robust starting point that can be customized and optimized for specific applications in various domains.

What are foundation models?

- Only for NLP?

Are foundation models only used for NLP tasks?

No, **foundation models are not limited to NLP tasks** — they are a broader category of models that serve as a general-purpose base for many types of AI tasks across different modalities.

Key Points:

1. Multimodal Scope:

Foundation models can be trained on various types of data, not just text. They are used in:

- **Vision** (e.g., image classification, object detection)
  - Examples: CLIP, DINO, SAM (Segment Anything Model)
- **Audio** (e.g., speech recognition, sound classification)
  - Examples: Whisper (speech recognition by OpenAI)
- **Multimodal tasks** (combining vision, language, and audio)
  - Examples: GPT-4 with vision, Gemini, Flamingo, or VILA

Technical University of Munich
Chair of Data Science in Earth Observation

What are foundation models?

- For EO?

What about using foundation models for Earth observation?

Yes, **foundation models are increasingly being applied to Earth Observation (EO)** — and they hold great promise for transforming how we analyze and interpret satellite and geospatial data at scale.

Technical University of Munich
Chair of Data Science in Earth Observation

Origin

- In March 2021, an informal community at Standard University was created.
- The community united researchers interested in some aspects of foundation models, from methods to applications.

- In August 2021, the report came out with the efforts of over 100 people from different backgrounds.

On the Opportunities and Risks of Foundation Models

Rishi Bommasani* Drew A. Hudson Ehsan Adeli Russ Altman Simran Arora
Sydney von Arx Michael S. Bernstein Jeannette Bohg Antoine Bosselut Emma Brunskill
Erik Brynjolfsson Shyamal Buch Dallas Card Rodrigo Castellon Niladri Chatterji
Annie Chen Kathleen Creel Jared Quincy Davis Dorottya Demszky Chris Donahue
Moussa Doumbouya Esin Durmus Stefano Ermon John Etchemendy Kawin Ethayarajh
Li Fei-Fei Chelsea Finn Trevor Gale Lauren Gillespie Karan Goel Noah Goodman
Shelby Grossman Neel Guha Tatsunori Hashimoto Peter Henderson John Hewitt
Daniel E. Ho Jenny Hong Kyle Hsu Jing Huang Thomas Icard Saahil Jain
Dan Jurafsky Pratyusha Kalluri Siddharth Karamcheti Geoff Keeling Fereshte Khani
Omar Khattab Pang Wei Koh Mark Krass Ranjay Krishna Rohith Kuditipudi
Ananya Kumar Faisal Ladhak Mina Lee Tony Lee Jure Leskovec Isabelle Levent
Xiang Lisa Li Xuechen Li Tengyu Ma Ali Malik Christopher D. Manning
Suvir Mirchandani Eric Mitchell Zanele Munyikwa Suraj Nair Avanika Narayan
Deepak Narayanan Ben Newman Allen Nie Juan Carlos Niebles Hamed Nilforoshan
Julian Nyarko Giray Ogut Laurel Orr Isabel Papadimitriou Joon Sung Park Chris Piech
Eva Portelance Christopher Potts Aditi Raghunathan Rob Reich Hongyu Ren
Frieda Rong Yusuf Roohani Camilo Ruiz Jack Ryan Christopher Ré Dorsa Sadigh
Shiori Sagawa Keshav Santhanam Andy Shih Krishnan Srinivasan Alex Tamkin
Rohan Taori Armin W. Thomas Florian Tramèr Rose E. Wang William Wang Bohan Wu
Jiajun Wu Yuhuai Wu Sang Michael Xie Michihiro Yasunaga Jiaxuan You Matei Zaharia
Michael Zhang Tianyi Zhang Xikun Zhang Yuhui Zhang Lucia Zheng Kaitlyn Zhou
Percy Liang*[1]

Center for Research on Foundation Models (CRFM)
Stanford Institute for Human-Centered Artificial Intelligence (HAI)
Stanford University

Bommasani et al., 2022

Technical University of Munich
Chair of Data Science in Earth Observation

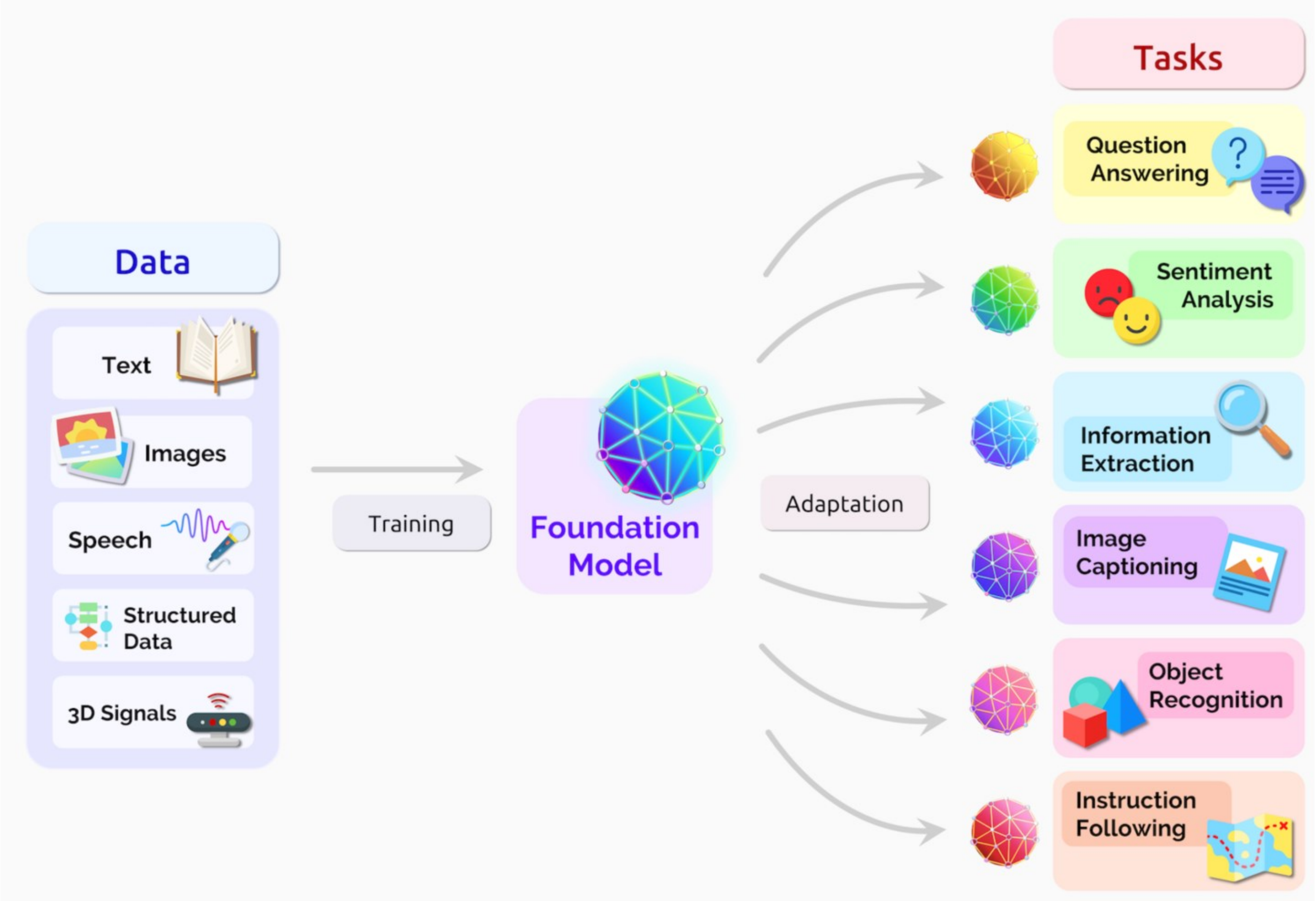
Definition

- Models trained on broad data that can be adapted to a wide range of downstream tasks.
- Examples
  - Large Language Models (LLMs)
    - e.g., GPT(OpenAI), Llama (Meta), DeepSeek, Gemini (Google), Grok (xAI)
  - Vision Foundation Models
    - Models trained with ImageNet
    - DINO, SAM, …
  - Vision-Language Models
    - Many recent LLMs are extended to VLMs, e.g., GPT-4V and Gemini

Technical University of Munich
Chair of Data Science in Earth Observation

From Deep Learning to Foundation Models

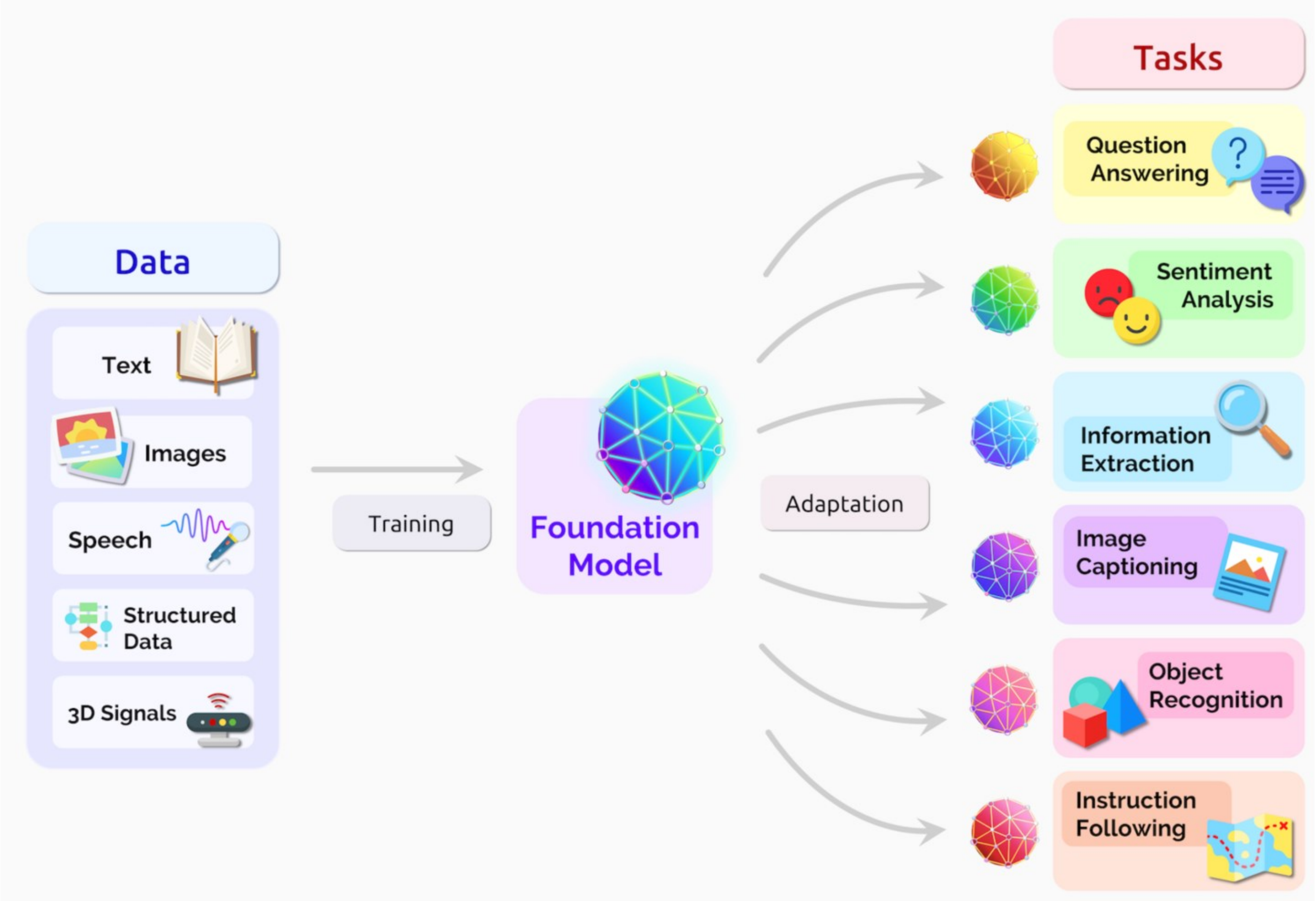
- Transfer Learning
  - Take the "knowledge" learned from one task to another.
- Scale
  - Hardware
    - Increasing GPU throughput and memory
  - Transformer architecture
    - Parallelism of hardware
    - More expressive models
  - Much more data



Technical University of Munich
Chair of Data Science in Earth Observation

Foundation Models for Earth Observation

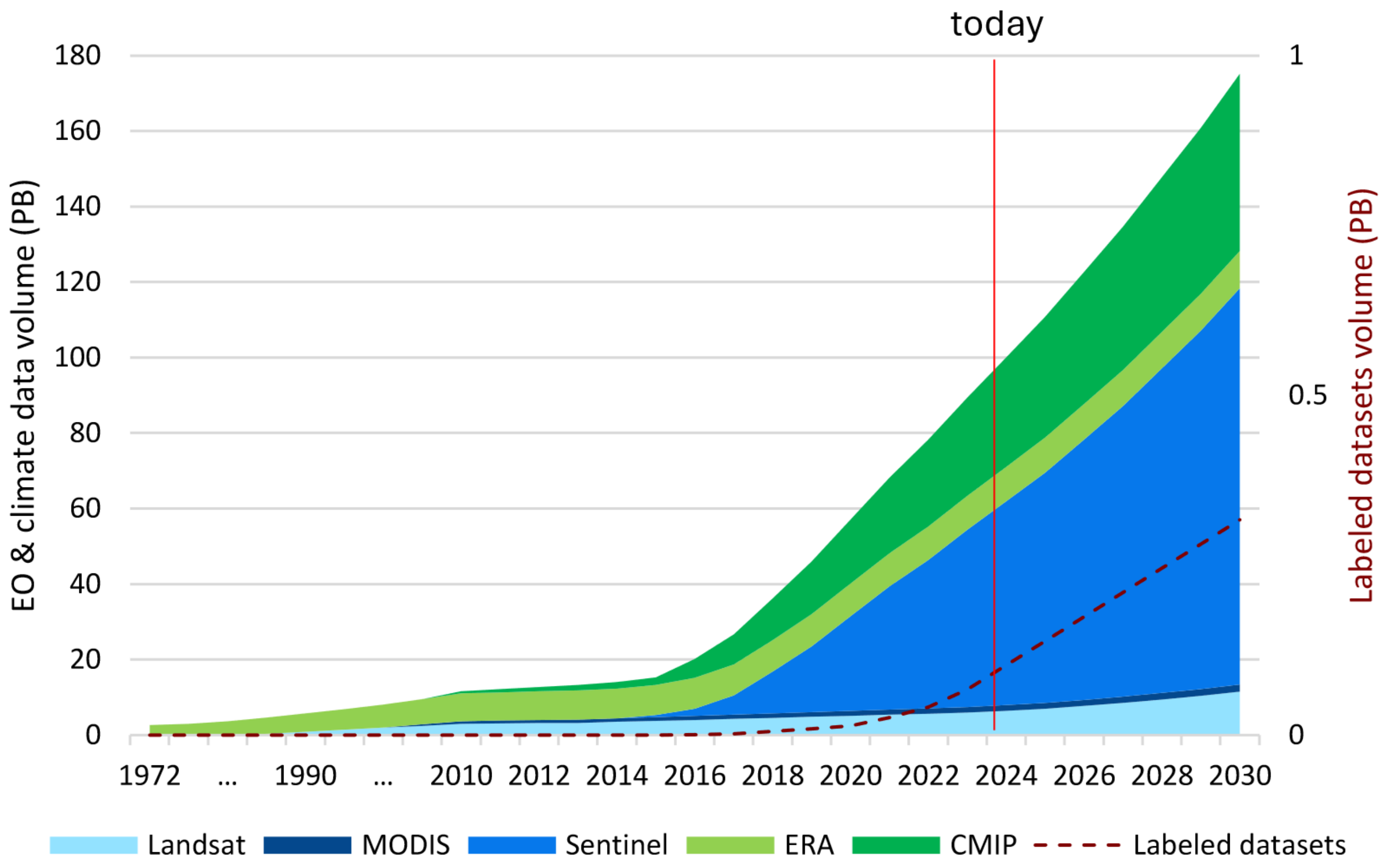
✓ • Transfer Learning
  - Take the "knowledge" learned from one task to another.
- Scale
  ✓ • Hardware
    - Increasing GPU throughput and memory
  ✓ • Transformer architecture
    - Parallelism of hardware
    - More expressive models
  ❓ • Much more data



Technical University of Munich
Chair of Data Science in Earth Observation

Foundation Models for Earth Observation – Big Data

- More than 1000 active remote sensing satellite in space.
- Continuous, high-quality, and open stream of satellite imagery.
- A wide range of Earth's features at various spatial, spectral, and temporal resolutions.



Zhu et al., 2024

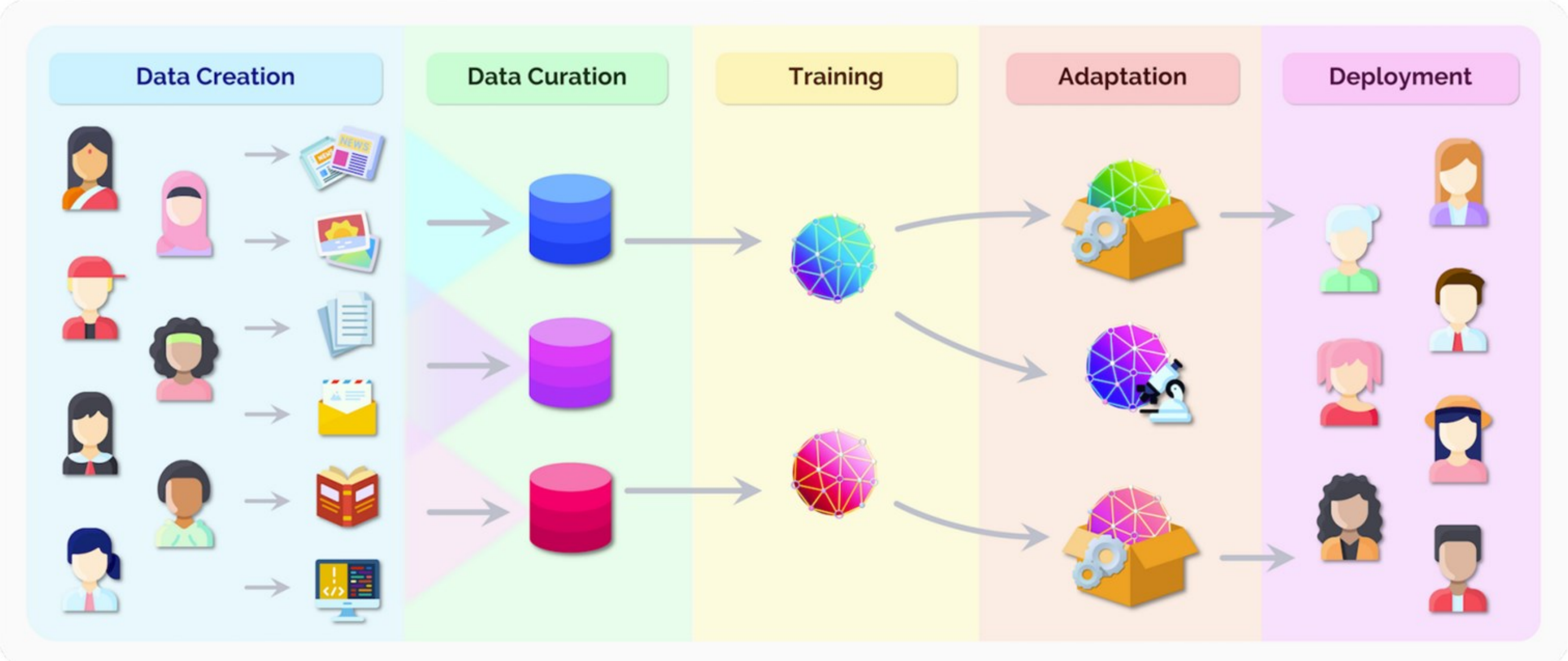
Foundation Models for Earth Observation – Challenges

- Domain discrepancies between natural and EO data
- More modalities in EO compared to natural data
- Geo-information
- Time series
- Physical constraints

Technical University of Munich
Chair of Data Science in Earth Observation

Foundation Models for Earth Observation

- Data
- Training
  - Training Strategy
  - Models
- Adaptation and Deployment
  - Fine-tuning
  - Implementation



Technical University of Munich
Chair of Data Science in Earth Observation

Data

- Big data are needed to
  - Train foundation models
  - Adapt foundation models for specific tasks

- Overview of Existing Large-scale Datasets for Foundation Models

- Works at our Chair
  - SSL4EO-S12: a large-scale pretraining dataset
  - EarthNets: empowering artificial intelligence for Earth observation

Technical University of Munich
Chair of Data Science in Earth Observation

Overview of Existing Large-scale Datasets for Foundation Models

- From ImageNet to EO data
  - In-domain Representation Learning for Remote Sensing (Neumann et al., 2019)

- Towards curated general-purpose pretraining datasets for EO
  - SEN12MS (Schmitt et al., 2019)
  - SeCo (Mañas et al., 2021)
  - SSL4EO-S12 (Wang et al., 2023)
  - SatlasPretrain (Bastani et al., 2023)
  - SSL4EO-L (Stewart et al., 2023)
  - SpectralEarth (Ait Ali Braham et al., 2024)
  - MMEarth (Nedungadi et al., 2024)

Technical University of Munich
Chair of Data Science in Earth Observation

Overview of Existing Large-scale Datasets for Foundation Models

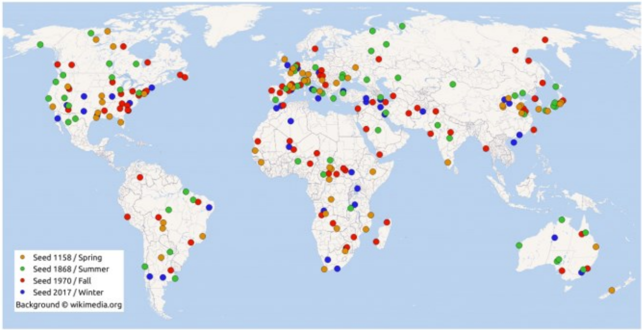
- On the way of scaling up dataset size, modality, resolution and time series.

	sensor	modality	GSD	# bands	# images	area (km^2)
SEN12MS (Schmitt et al., 2019)	S1, S2	SAR, MS	10-60m	2-13	180K*2	590K
SeCo (Mañas et al., 2021)	S2	MS	10-60m	12	200K*5	<1.4M
SSL4EO-S12 (Wang et al., 2023)	S1, S2	SAR, MS	10-60m	2-13	251K*4*3	1.8M
SSL4EO-L (Stewart et al., 2023)	LandSAT	MS	30-120m	7-11	250K*4*5	15.6M
SatlasPretrain (Bastami et al., 2023)	S2, RGB	MS, RGB	1m,10-60m	3-10	856K	21.3M
SpectralEarth (Ait Ali Braham et al., 2024)	EnMAP	HS	30m	202	253K	3.7M
MMEarth (Nedungadi et al., 2024)	S1, S2, DEM, …	12 (SAR, MS,…)	10m	46	1.2M	1.96M

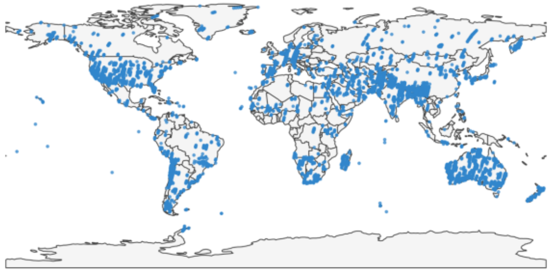
Technical University of Munich
Chair of Data Science in Earth Observation

Overview of Existing Large-scale Datasets for Foundation Models

SEN12MS (Sentinel)

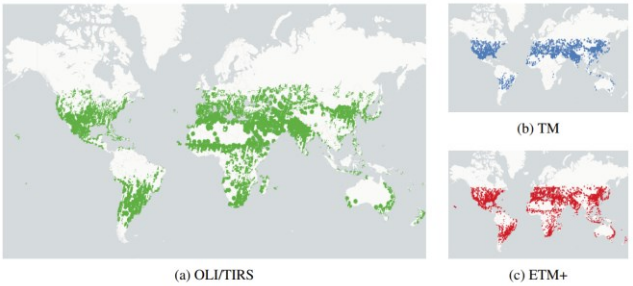


SpectralEarth (EnMAP)

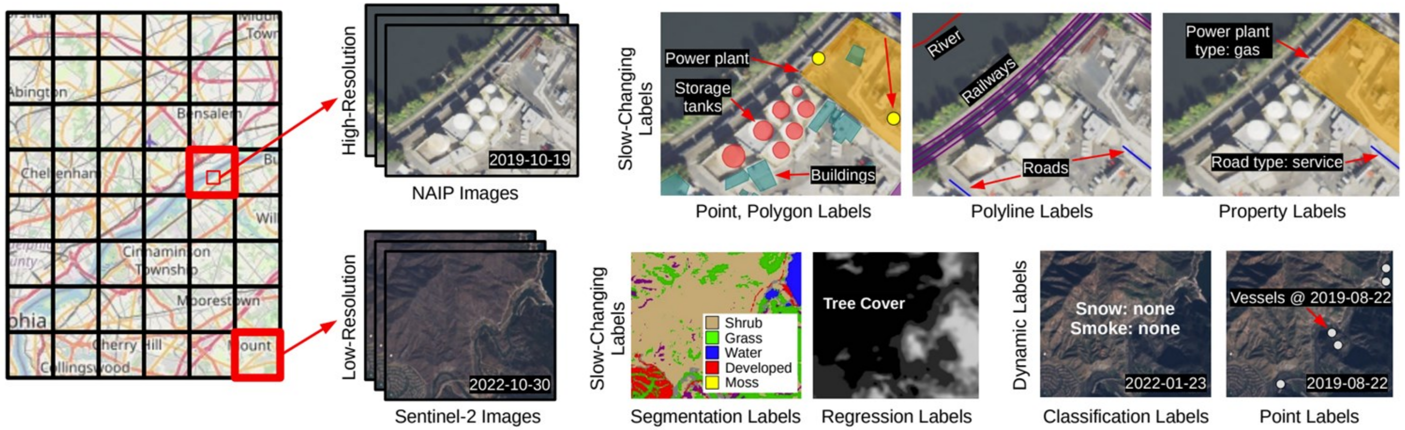


SSL4EO-S12 (Sentinel)

SSL4EO-L (Landsat)



Satlas Pretrain (Sentinel + NAIP)

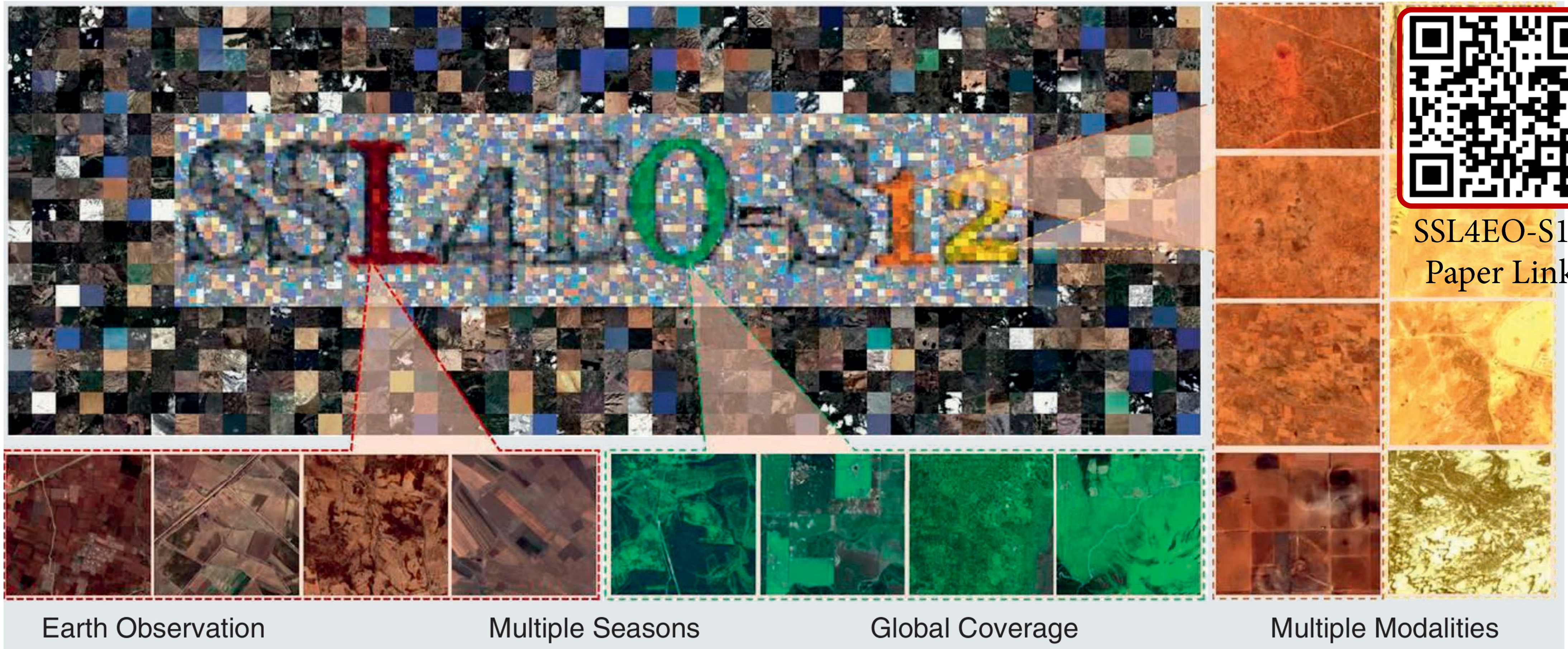


Schmitt et al., 2019; Wang et al., 2023; Stewart et al., 2021; Ait Ali Braham et al., 2024; Bastani et al., 2023

Technical University of Munich
Chair of Data Science in Earth Observation

dataset	EuroSAT	BE-10%	BE-100%
ImageNet (RGB) [4]	*86.4*	*70.5*	*71.8*
SeCo (RGB) [4]	*89.5*	*74.5*	*76.3*
SEN12MS (RGB)	94.9	76.6	79.6
SSL4EO-S12 (RGB)	**96.6**	**80.1**	**82.3**
SeCo^2 (all bands)	89.2	73.7	76.6
SEN12MS (all bands)	95.5	79.6	82.1
BigEarthNet (all bands)	94.4	80.6	83.9
SSL4EO-S12 (all bands)	**98.0**	**82.1**	**84.2**

SSL4EO-S12



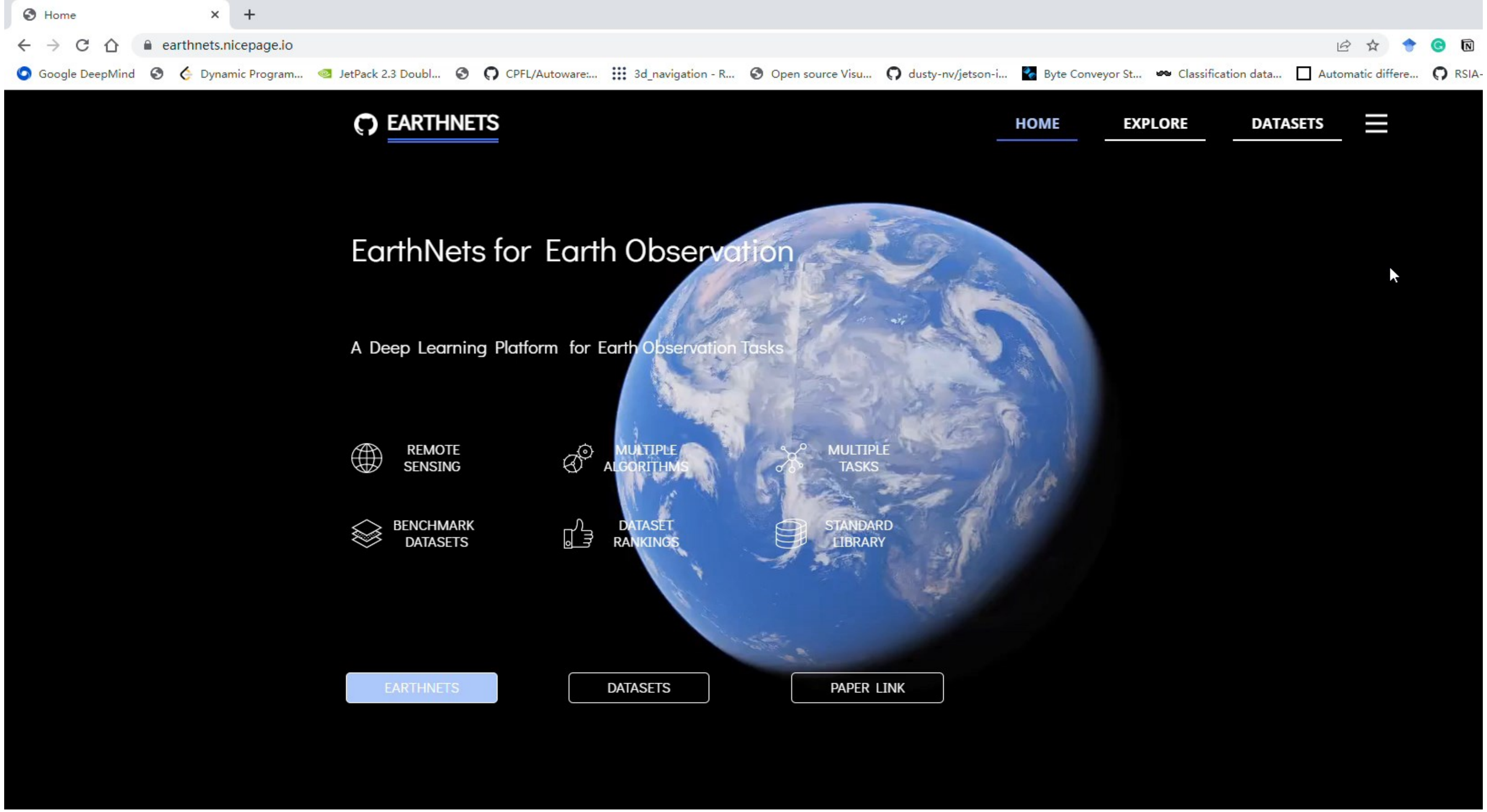
SSL4EO-S12
Paper Link

Wang et al., 2023

Technical University of Munich
Chair of Data Science in Earth Observation

Paper Homepage Catalogue

EarthNets: Empowering AI in Earth Observation



Xiong et al., 2024

Technical University of Munich
Chair of Data Science in Earth Observation

EarthNets: Empowering AI in Earth Observation

Catalogue Paper

- Review more than 500 remote sensing datasets (updating …)
  - Detailed attributes including ten different aspects
- Systematic dataset analyses
  - Volume
  - Bibliometric
  - Research Domains
  - Modalities
  - Resolution
  - Correlation

ID	Name	Publicatio...	# samples	Size of sa...	Task	# classes	Domain	Modality	Coverage	Resolution	Volume (G...	Annotation	# Citations	Links
1	MSTAR-8cl...	1,996	9,466	368	Class	8	Military	SAR	/	0.3m	0.444	image-level	–	https://ww...
2	Things And...	2,008	30	792	OD	2	Vehicles	RGB	/	0.5m	0.01	object-level	550	http://ai.sta...
3	OIRDS	2,009	900	256~640	OD	5	Vehicles	RGB	/	0.15m	0.153	object-level	30	https://sour...
4	Oakland 3-...	2,009	–	1.5 km	3DOD	5	Urban 3D P...	PointCloud	Single city	/	0.033	point-level	365	http://www...
5	UC Merced	2,010	2,100	256	Class	21	General Sc...	RGB	multiple citi...	0.3m	0.3	image-level	1,808	http://weeg...
6	Pavia Center	2,011	1	1096	SemSeg	9	Land Cover	Hyperspect...	single city	1.3m	0.121	pixel-level	–	http://www...
7	Pavia Unive...	2,011	1	610	SemSeg	9	Land Cover	Hyperspect...	single city	1.3m	0.032	pixel-level	–	http://www...
8	ISPRS 2D - ...	2,011	38	6000	SemSeg	6	General Sc...	RGB,nDSM	single city	0.05m	15.625	pixel-level	–	https://ww...
9	ISPRS 2D - ...	2,011	33	2200	SemSeg	6	General Sc...	RGB,nDSM	single city	0.09m	16.6	pixel-level	–	https://ww...
10	SZTAKI Air...	2,008	13	800	CD	2	Land Change	RGB	/	1.5m	0.04	pixel-level	193	http://web....

Technical University of Munich
Chair of Data Science in Earth Observation

Training

- How to train
  - Supervised learning
  - Self-supervised learning

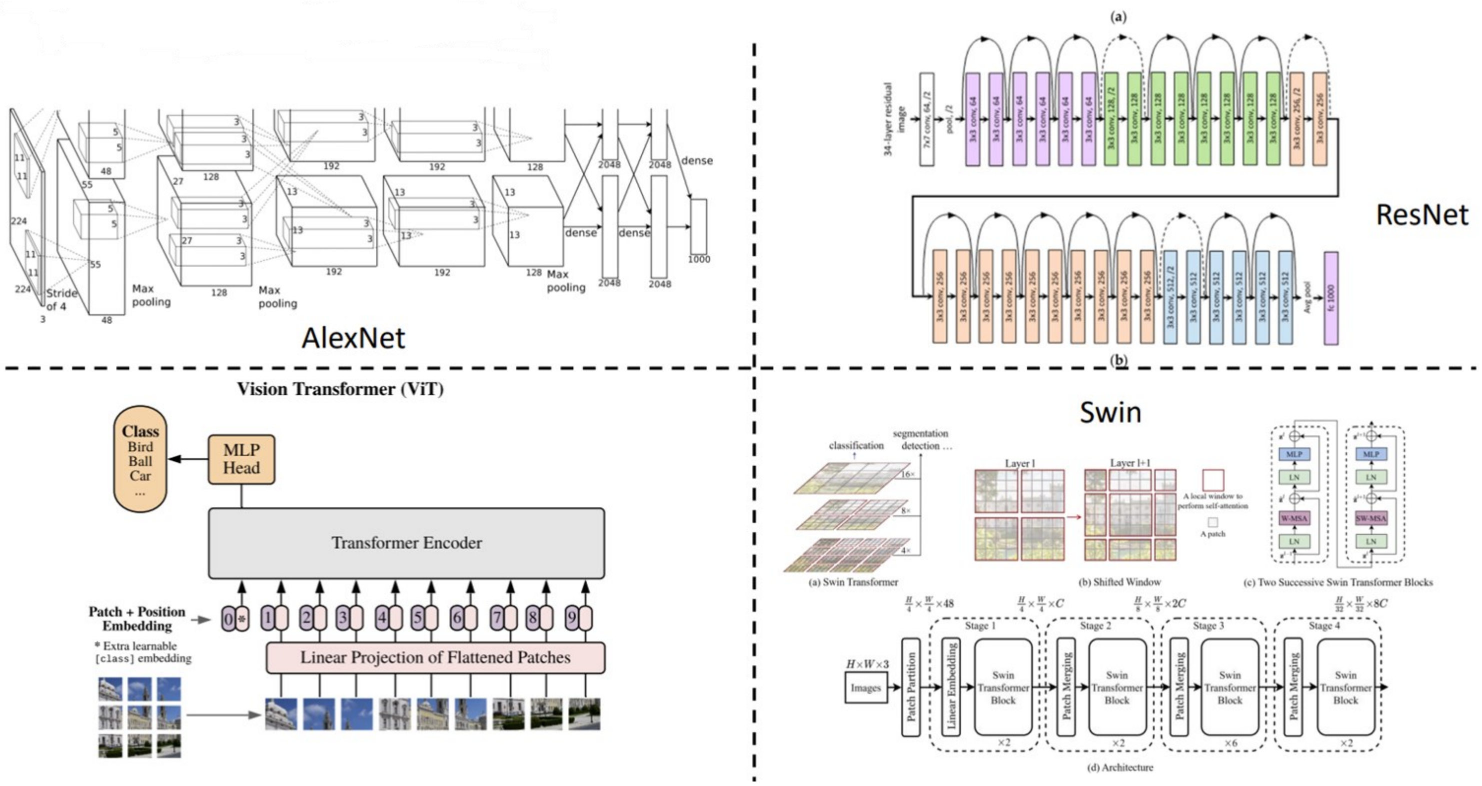
- What to train
  - Existing foundation models

- Works at our Chair
  - DeCUR: Decoupling Common and Unique Representations for Multimodal Self-supervised Learning
  - DOFA: Neural Plasticity-Inspired Multimodal Foundation Model for Earth Observation

Technical University of Munich
Chair of Data Science in Earth Observation

Supervised Learning

- Mapping an image to a discrete label which is associated to a visual concept.
  - Early/powerful foundation models when trained on large datasets.
  - Standard way to develop model backbones.
  - Annotation is expensive and limited!



Krizhevsky et al., 2012; He et al., 2015; Dosovitskiy et al., 2021; Liu et al., 2021

Technical University of Munich
Chair of Data Science in Earth Observation

Self-supervised Learning (SSL)

- Learning generic representations from large-scale unlabeled data through self-supervision.
- 3 types of SSL
    - Predictive
    - Contrastive
    - Generative

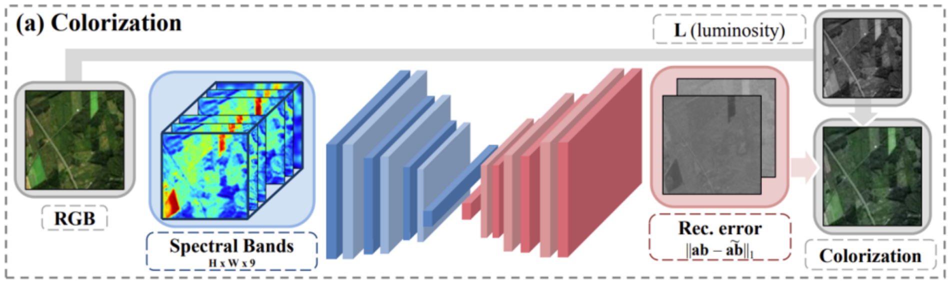
Technical University of Munich
Chair of Data Science in Earth Observation

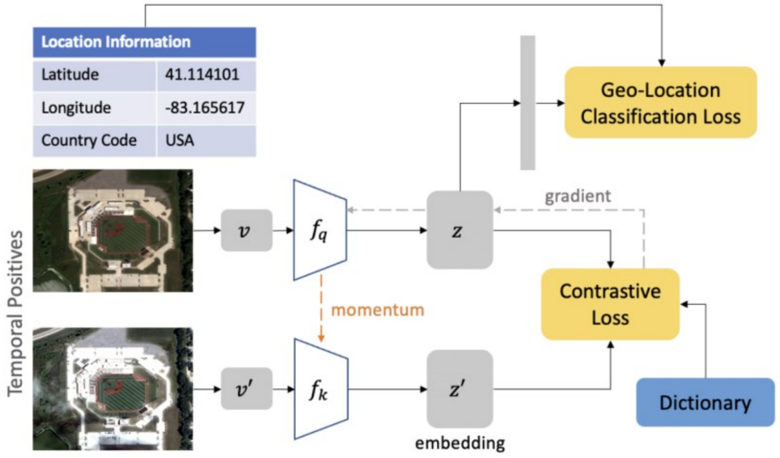
Predictive SSL (Outdated)

- Hand-designed pretext tasks
  - Solving Jigsaw puzzles
  - colorization
  - …
- In EO
  - Colorization (Vincenzi et al., 2020)
  - Geography-aware SSL (Ayush et al., 2020)

X =

Y = (9,5,8,3,2,4,7,1,6)



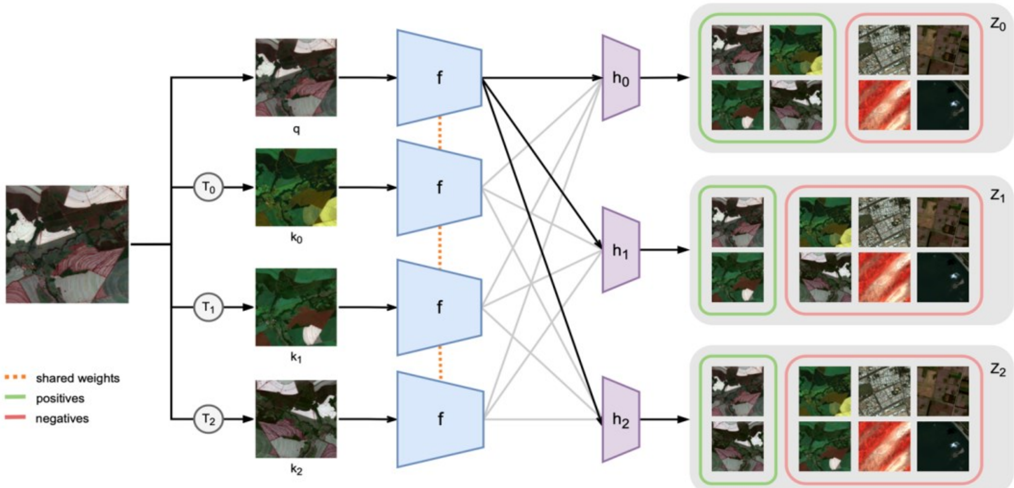


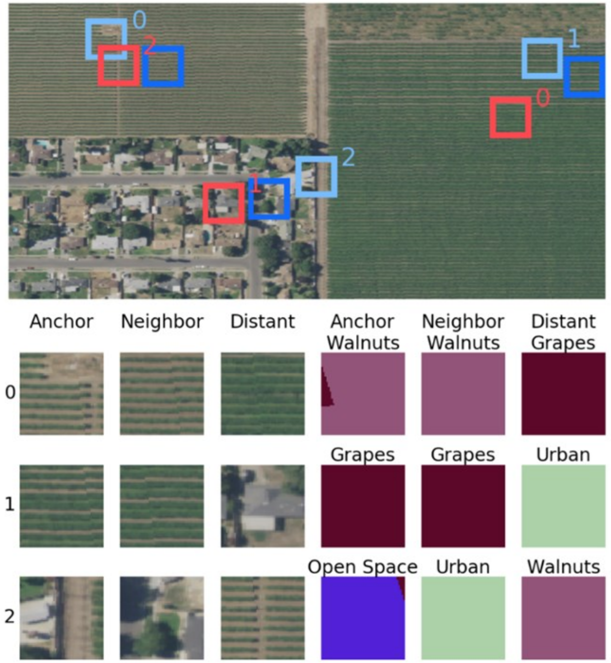
Technical University of Munich
Chair of Data Science in Earth Observation

Contrastive SSL

- Joint-embedding architectures with instance discrimination
  - e.g., SimCLR (Chen et al., 2020)

- In EO
  - Seasonal contrast (SeCo) (Mañas et al., 2021)
  - Geolocation contrast (Tile2Vec) (Jean et al., 2019)



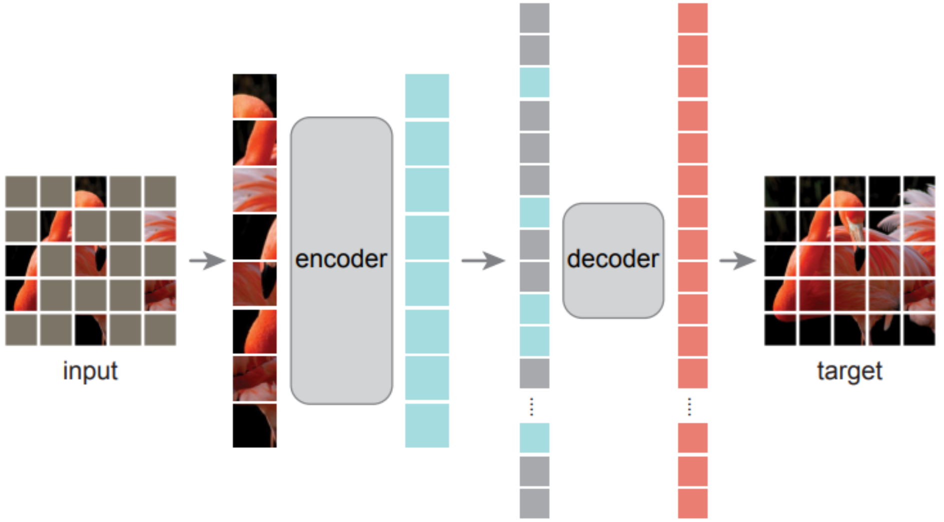


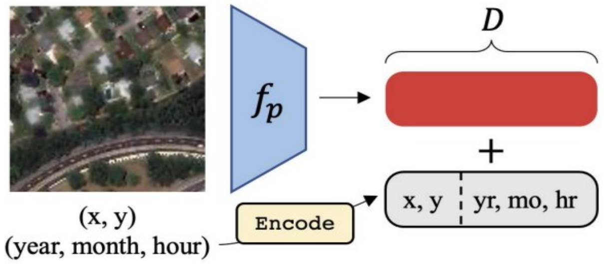
Vincenzi et al., 2020; Ayush et al., 2020

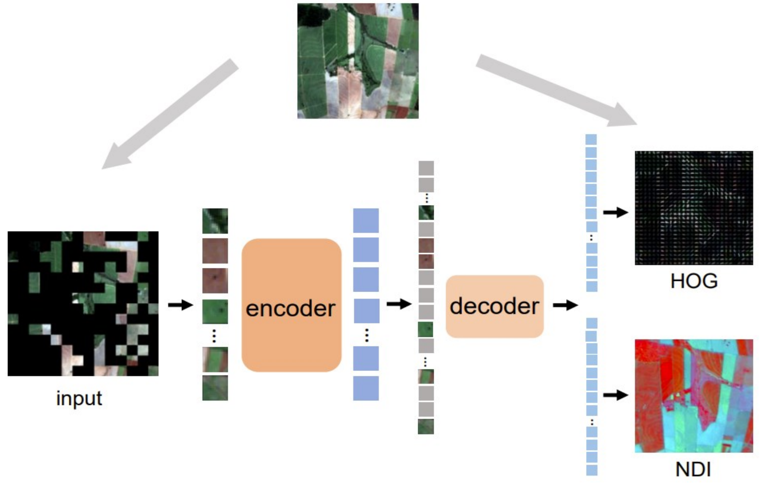
Technical University of Munich
Chair of Data Science in Earth Observation

Generative SSL

- Masked Image Modeling (MIM)
  - Masked Autoencoder (MAE) (He et al., 2021)
- In EO
  - Spectral/temporal positional encoding (SatMAE) (Cong et al., 2022)
  - Remote sensing feature reconstruction (FG-MAE) (Wang et al., 2023)



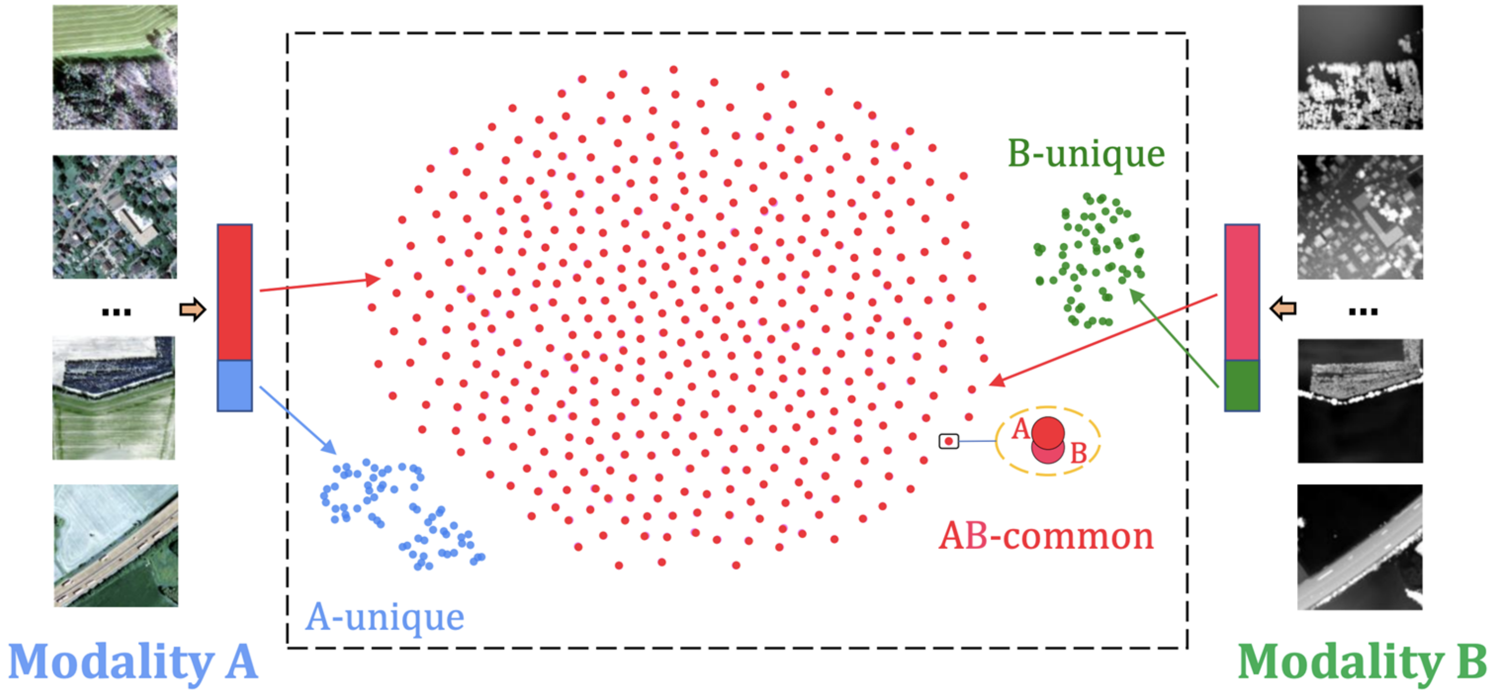




Technical University of Munich
Chair of Data Science in Earth Observation

SSL – DeCUR

- Motivation
  - Modality-unique information is neglected in existing multimodal works.



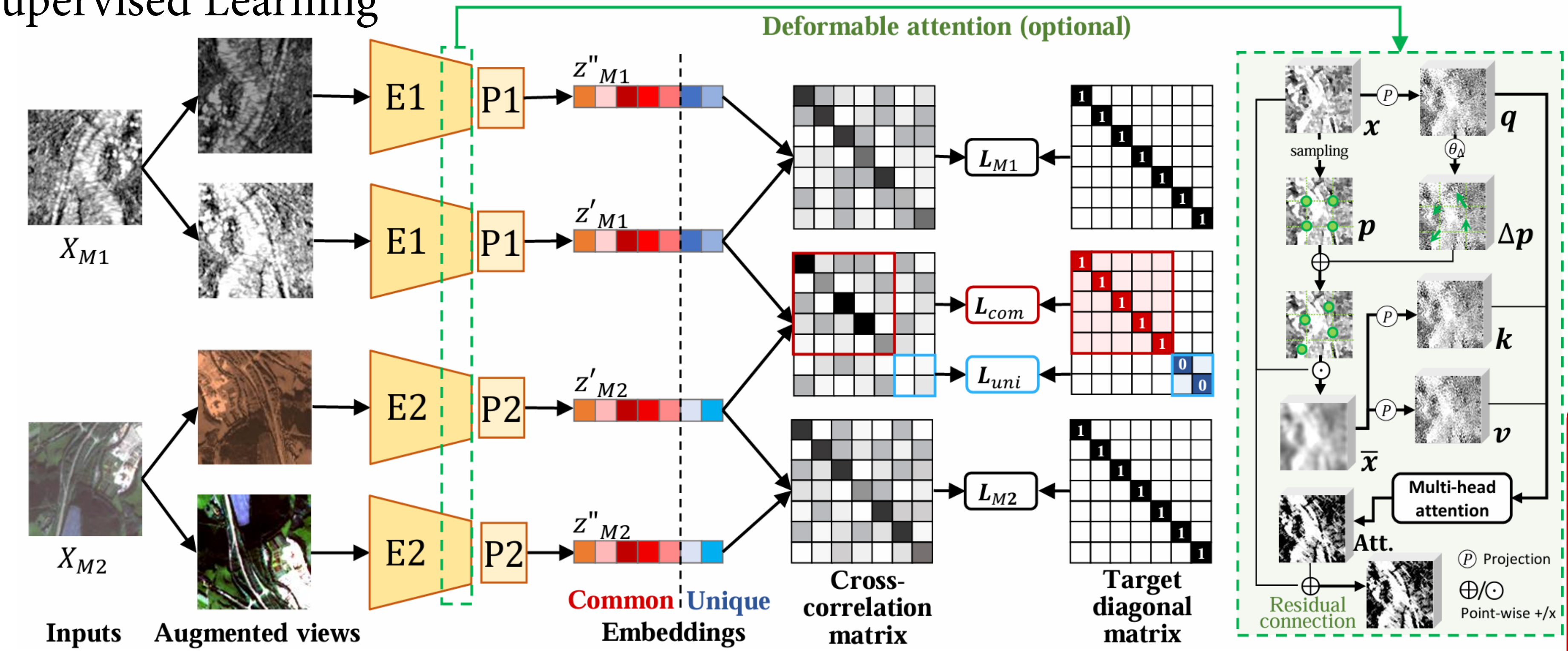
Wang et al., 2024

Technical University of Munich
Chair of Data Science in Earth Observation

SSL - DeCUR

SAR-optical	1% labels	100% labels
Rand. Init.	58.7	70.1
Supervised	77.0	88.9
SimCLR-cross	77.4/78.7	82.8/89.6
CLIP	77.4/78.7	82.8/89.6
Barlow Twins	78.7/80.3	83.2/89.5
VICReg	74.5/79.0	81.9/89.5
DeCUR (ours)	**79.8/81.5**	**86.2/89.8**

- Decoupling Common and Unique Representations for Multimodal Self-supervised Learning



DeCUR Paper Link

Technical University of Munich
Chair of Data Science in Earth Observation

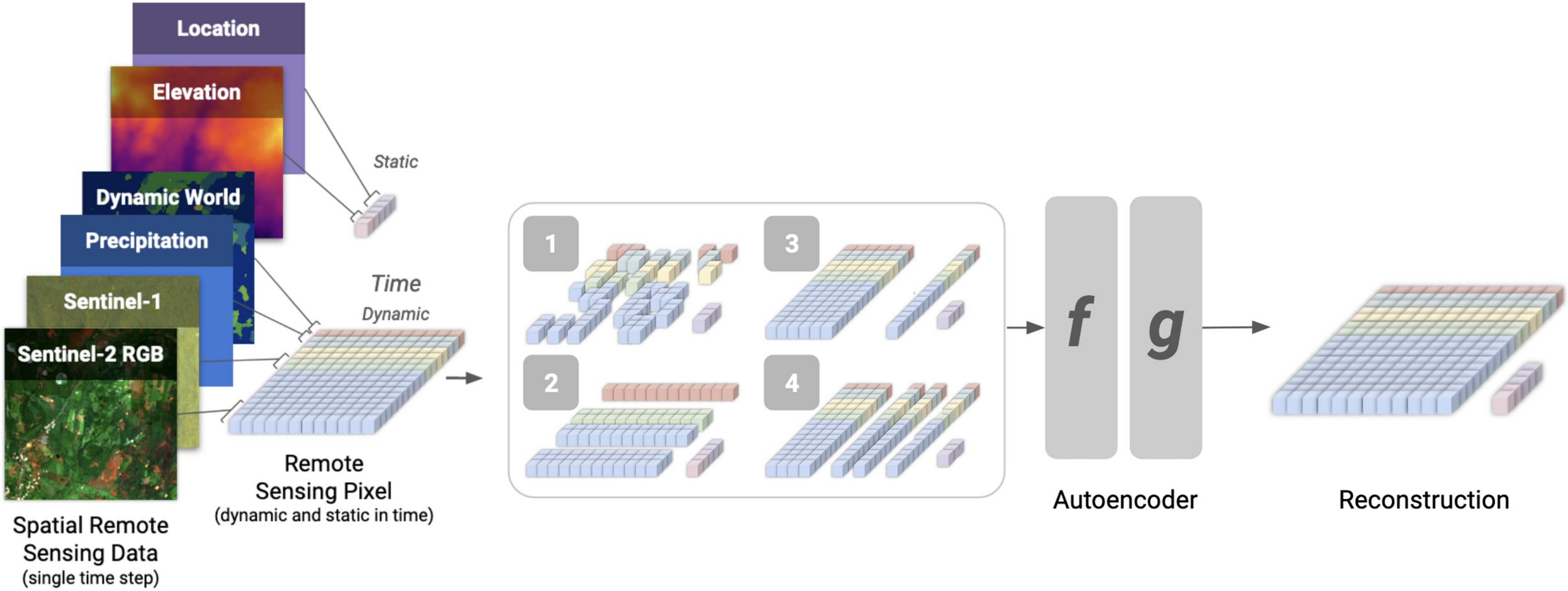
Models

- Mainstream EO foundation models are self-supervised by masked image modeling.
- Considering the challenges of EO
  - Time series and geo-encoded
    - Presto
    - Prithvi
  - Multi-modality
    - TerraMind
    - DOFA

Technical University of Munich
Chair of Data Science in Earth Observation

Models – Presto

- 4 types of masking strategies: random, channels groups, contiguous timesteps, timesteps.

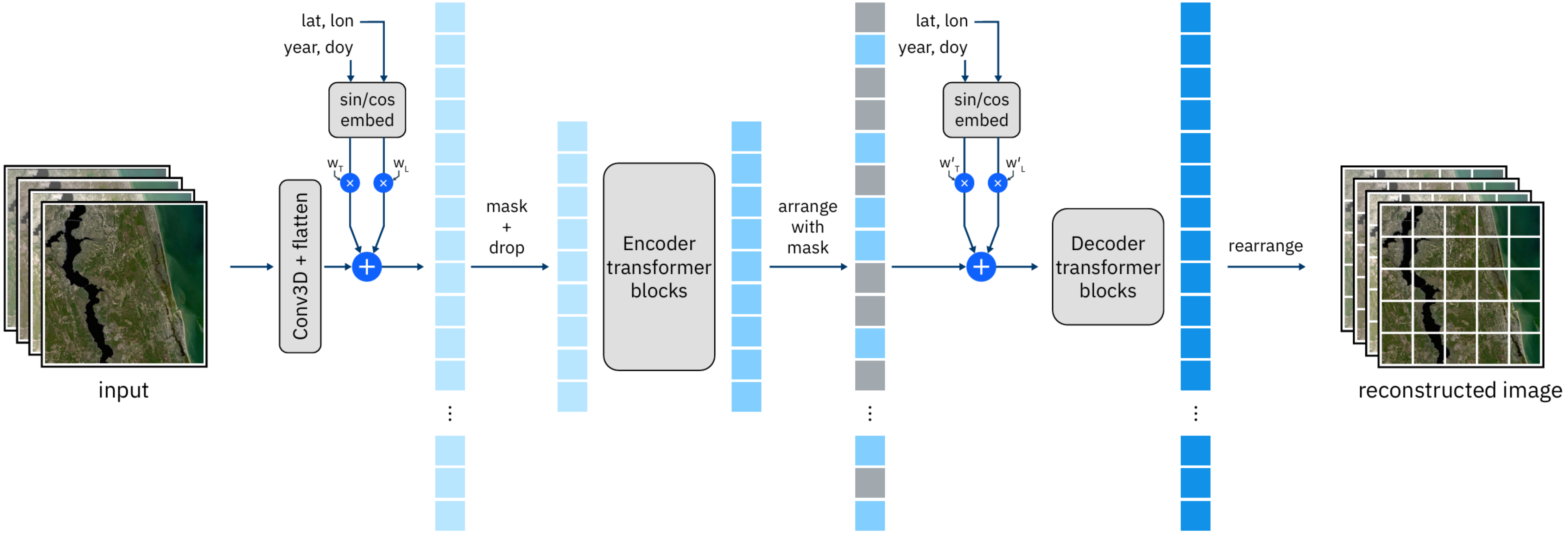


Tseng et al., 2023

Technical University of Munich
Chair of Data Science in Earth Observation

Models – Prithvi

- Temporal and spatial embeddings.

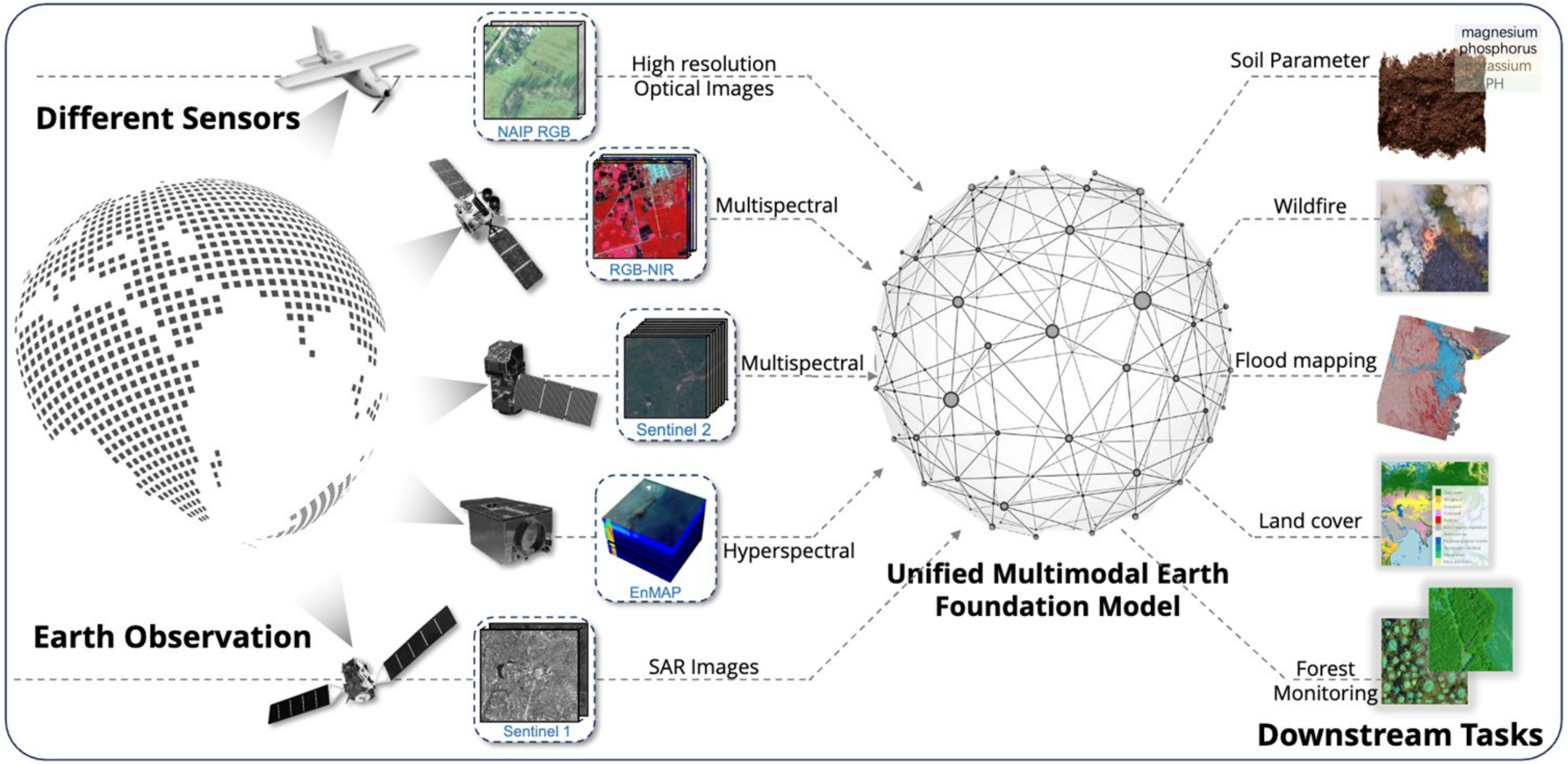


Szwarcman et al., 2024

Technical University of Munich
Chair of Data Science in Earth Observation

Models – DOFA

- Dynamic One-For-All foundation model for remote sensing and Earth observation
- Challenges
  - How to unify different data modalities?

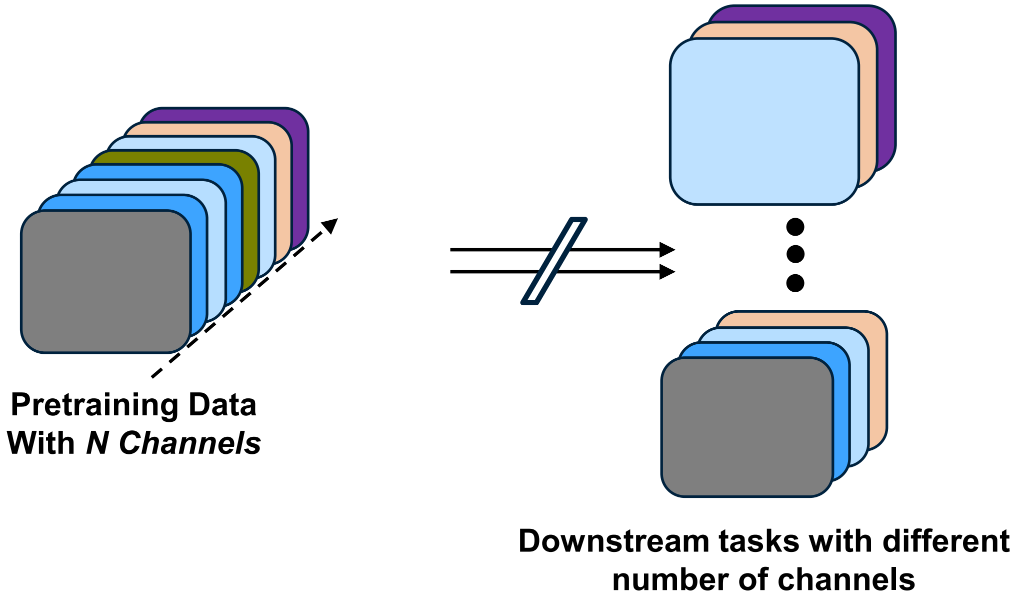


Xiong et al., 2024

Technical University of Munich
Chair of Data Science in Earth Observation

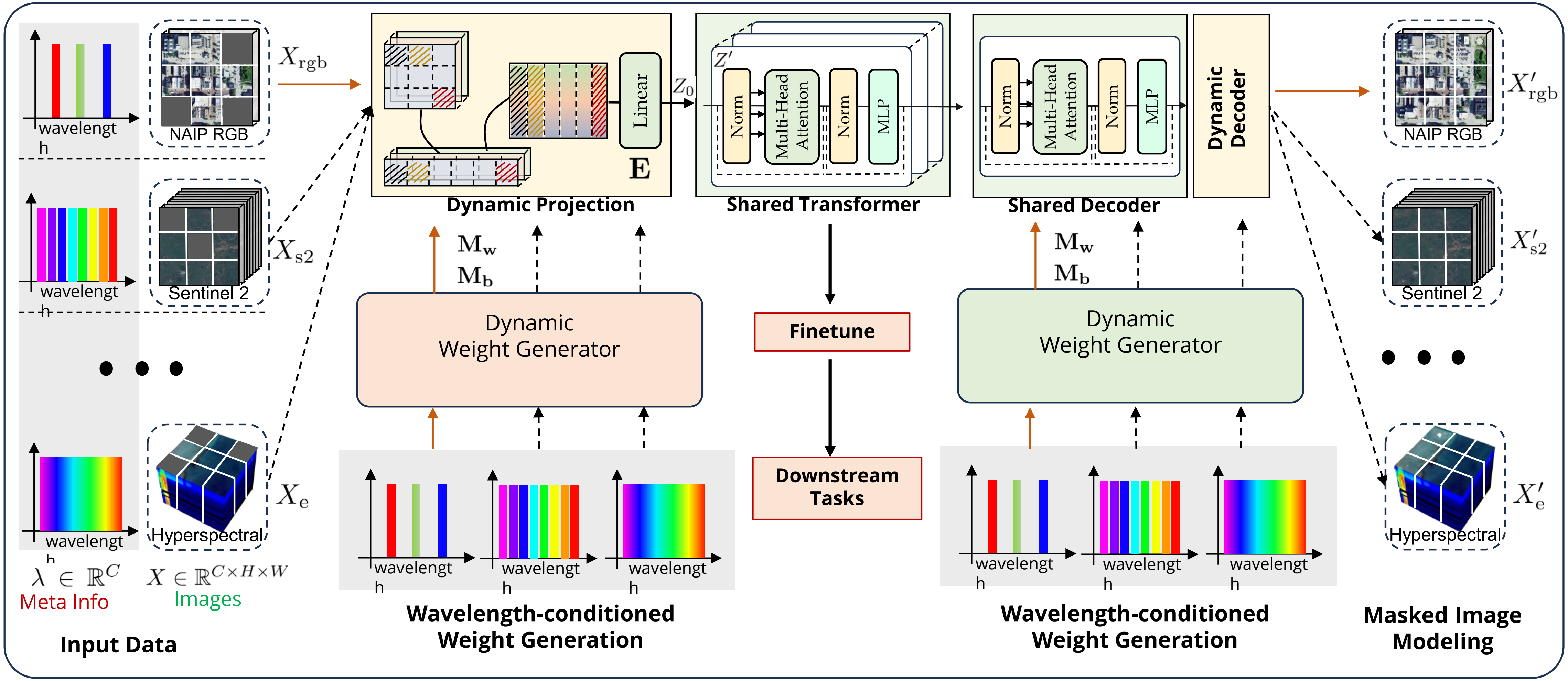
Models – DOFA

- Dynamic One-For-All foundation model for remote sensing and Earth observation
- Challenges
  - How to unify different data modalities?
  - How to unify different spectral bands?
    - Different combination in downstream tasks
    - New sensors never seen during training



Technical University of Munich
Chair of Data Science in Earth Observation

Models – DOFA – Architecture



Technical University of Munich
Chair of Data Science in Earth Observation

Models - DOFA

- DOFA is a unified multimodal foundation model for different data modalities in remote sensing and Earth observation.
- DOFA is inspired by neural plasticity.
  - Dynamic patch embedding
  - Dynamic weight generator

- More Findings
  - DOFA generates diverse weights dynamically.
  - DOFA optimizes separability in latent space.
  - DOFA masters arbitrary classification tasks.
  - DOFA outperforms single-source models for segmentation tasks.

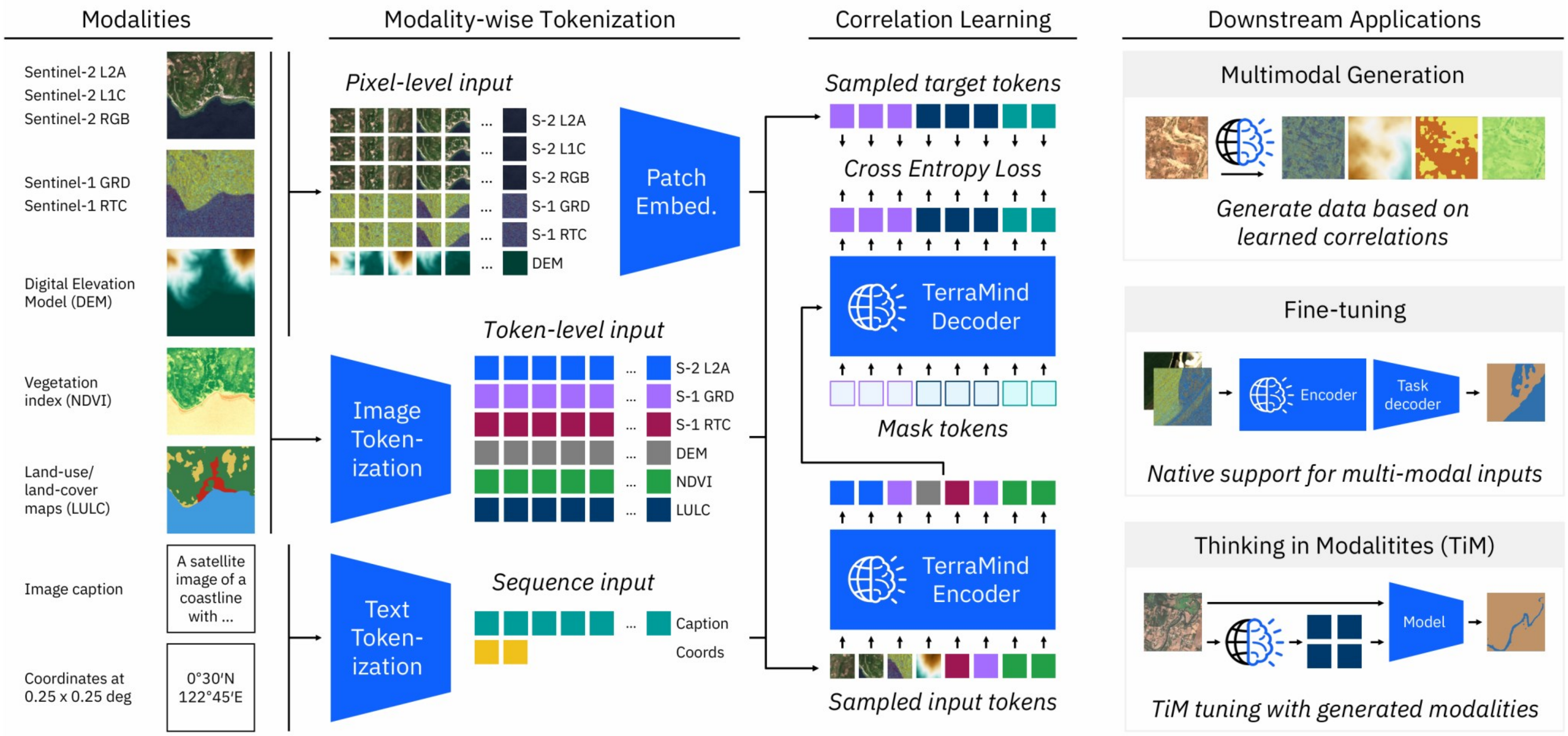
DOFA Paper Link

DOFA Code Repo

Technical University of Munich
Chair of Data Science in Earth Observation

Models – TerraMind

- Any-to-any generative, multimodal foundation model for Earth observation.



Jakubik et al., 2025

Technical University of Munich
Chair of Data Science in Earth Observation

Adaptation and Deployment

- Fine-tuning of foundation models for specific tasks.

- Conventional Fine-tuning
- Parameter-Efficient Fine-tuning (PEFT)

- Works at our Chair
  - Dataset4EO: data streaming for efficient deployment of foundation models
  - ExEBench: Benchmark data for extreme events

Technical University of Munich
Chair of Data Science in Earth Observation

Conventional Fine-tuning

- Mostly used fine-tuning approaches in EO.

- Linear Probing
  - Freeze all weights, train a single final layer at the end
  - Very efficient, but not very flexible

- Full fine-tuning
  - Train all model weights on your dataset
  - Flexible but inefficient, can lead to catastrophic forgetting

- Partly fine-tuning
  - Train model weights at your choice, e.g., freeze the encoder
  - More efficient than full fine-tuning, but less flexible

Technical University of Munich
Chair of Data Science in Earth Observation

Parameter Efficient Fine-tuning (PEFT)

- Motivation, for full fine-tuning, concerns arise like
  - Computational Overhead
  - Memory Constraints
  - Catastrophic Forgetting
  - Environmental Concerns

- PEFTs are proposed for
  - Efficiency
  - Scalability
  - Performance Retention
  - Generalizability

Chang et al., 2025

Technical University of Munich
Chair of Data Science in Earth Observation

Parameter Efficient Fine-tuning (PEFT)

- Examples
  - Adapter-based tuning
  - Low-rank Adaptation
  - Prefix tuning
  - Prompt Tuning
  - Selective parameter tuning
  - Hybrid and modular approaches
  - …

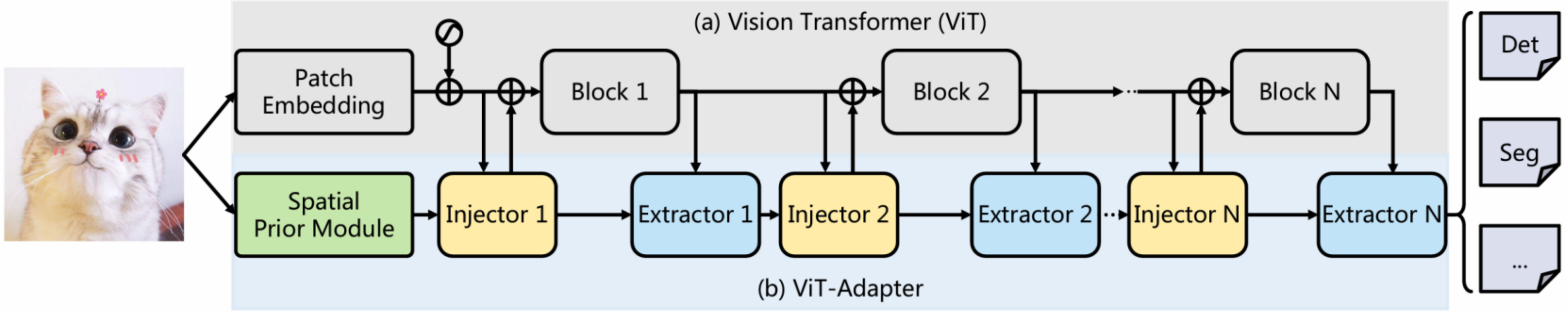
- Less used for fine-tuning foundation models for EO.

Technical University of Munich
Chair of Data Science in Earth Observation

PEFT – Adapter-based tuning

- Small, trainable modules into a frozen pre-trained model
- Usually in the bottleneck design in transformer layers, project input features into a lower-dimensional space before up-projecting them back to the original feature size.

- Advantages
  - Low computational cost
  - Modular transferability
  - Better knowledge retention
- Limitations
  - Inference overhead
  - Task-specific design
  - Limited adaptation capacity.



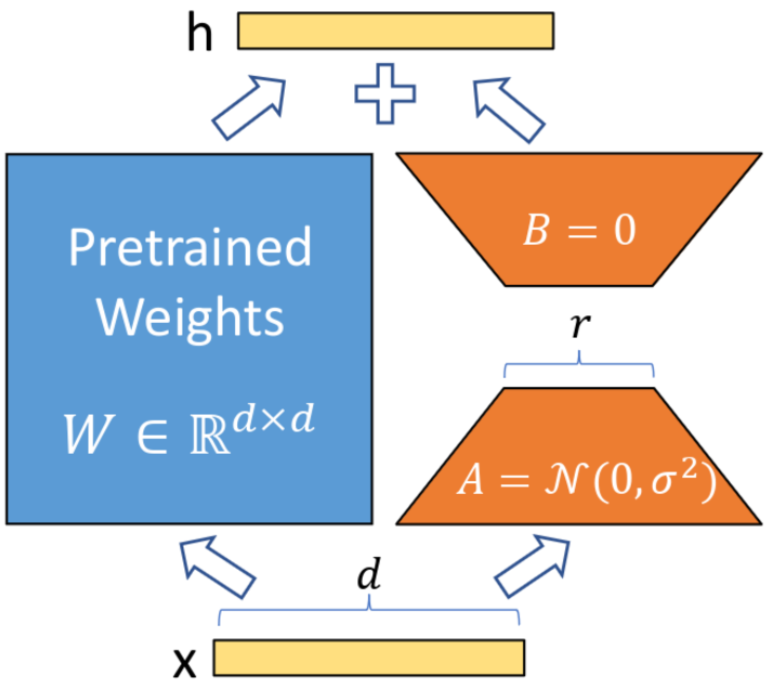
Chen et al., 2022

Technical University of Munich
Chair of Data Science in Earth Observation

PEFT – Low-Rank Adaptation (LoRA)

- Reduces the number of trainable parameters by decomposing weight updates into low-rank matrices.
- Usually modifies the query and value projection matrices in transformers.

- Advantages
  - Significant reduction in Trainable Parameters
  - No Additional Inference Latency
- Limitation
  - Rank Selection Sensitivity
  - Task Dependency



Hu et al., 2021

Technical University of Munich
Chair of Data Science in Earth Observation

PEFT – Prompt and Prefix Tuning

- Modify the input representations.
- Prompt tuning
  - Augment the input sequence with learnable prompt
- Prefix tuning
  - Augment the key and value matrices with prefixes
- Advantages
  - Minimal fine-tuning cost
  - Strong generalizability
  - Few-shot adaptation
- Limitations
  - Performance sensitivity
  - Limited representation capacity

Lester et al., 2021; Li et al., 2021

Technical University of Munich
Chair of Data Science in Earth Observation

Deployment - Easier Implementation

- Based on TerraTorch by IBM
  - A PyTorch domain library for geospatial data
  - Provide a flexible fine-tuning framework for geospatial foundation models
    - Flexible trainers for image segmentation, classification and pixel-wise regression
    - Model factories that allow easily combine backbones and decoders for different tasks
    - Launching of fine-tuning tasks through CLI and flexible configuration files, or via jupyter notebooks.
  - Supported foundation models
    - Prithvi, TerraMind, SatMAE, ScaleMAE, Satlas, **DOFA**, SSL4EO-L and **SSL4EO-S12** models…

  - Will be demonstrated in the following hands-on session.

TerraTorch

Gomes et al., 2025

Technical University of Munich
Chair of Data Science in Earth Observation

Deployment - Dataset4EO - Motivation

- For efficient data streaming while training/evaluating/fine-tuning foundation models.

- Two types of PyTorch datasets

	Map-Style Dataset `torch.utils.data.Dataset`	**Iterable-Style Dataset** `torch.utils.data.IterableDataset`
Access Pattern	Random access via index `__getitem__`	Sequential access via iteration `__iter__`
Ideal For	Static datasets on disk	Streaming, generated, or massive datasets
Memory Usage (RAM)	May require large memory if dataset is loaded fully	Typically **lower**, as data is streamed or generated on the fly

- RAM is usually a constraint when utilizing foundation models.

Technical University of Munich
Chair of Data Science in Earth Observation

Deployment - Dataset4EO - Motivation

- Define a map-style dataset
  - `__init__`
    - Images, annotations, and transforms (pre-processing)
  - `__getitem__`
    - Return the data batches

  - They should adapt to different datasets

- Multiple datasets are needed to train/evaluate foundation models.

```
class CustomImageDataset(Dataset):
    def __init__(self, annotations_file, img_dir,
transform=None, target_transform=None):
        self.img_labels = pd.read_csv(annotations_file)
        self.img_dir = img_dir
        self.transform = transform
        self.target_transform = target_transform

    def __len__(self):
        return len(self.img_labels)

    def __getitem__(self, idx):
        img_path = os.path.join(self.img_dir,
self.img_labels.iloc[idx, 0])
        image = decode_image(img_path)
        label = self.img_labels.iloc[idx, 1]
        if self.transform:
            image = self.transform(image)
        if self.target_transform:
            label = self.target_transform(label)
        return image, label
```

Technical University of Munich
Chair of Data Science in Earth Observation

Deployment - Dataset4EO - Motivation

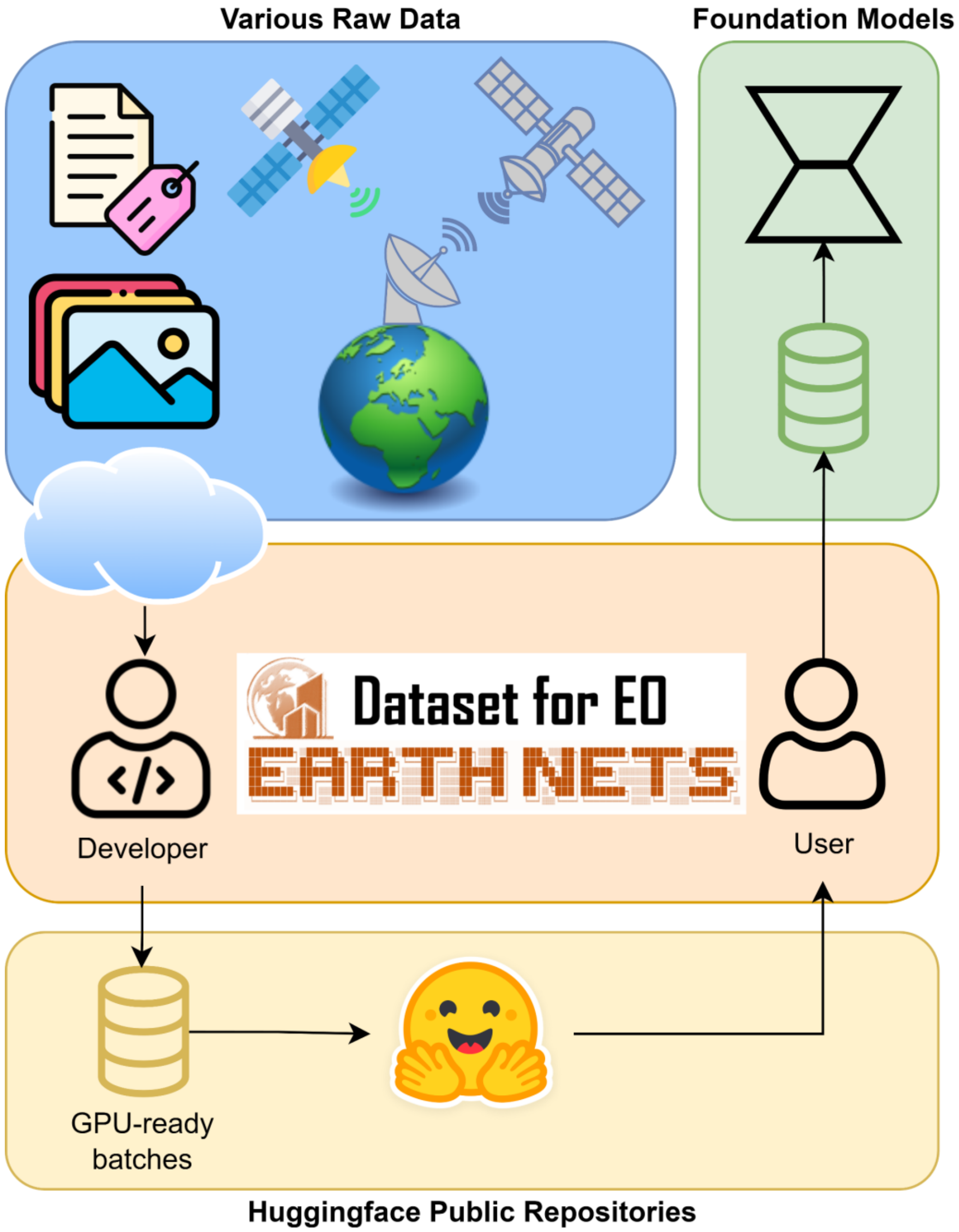
- To save RAM and enable more efficiency, provide a protocol of using iterable-style dataset for foundation models.
- To lift the burden of writing customized `torch.utils.data.Dataset` class for each dataset, provide processing-ready batches instead and unify the data loading procedure.

- To promote easy and efficient training/fine-tuning/evaluation of foundation models.

Technical University of Munich
Chair of Data Science in Earth Observation

Deployment - Dataset4EO – Pipeline

- Based on TorchGeo (Stewart et al., 2022) TorchGeo
- Developer
  - Process various EO raw data from the Internet into GPU-ready batches.
  - Upload them into huggingface public repositories.
- User
  - Fetch GPU-ready batches from the huggingface repositories.
  - Use them for foundation models.



Deployment - Dataset4EO – Key Features

- Channel-wise Dataset Support
  - Efficient storage and selective decoding of individual image channels.
- Full-Image Dataset Support
  - Traditional storage and decoding of entire multi-channel images.
- Integration with LitData[1]
  - Fully leverages LitData's streaming capabilities for handling large datasets.

- TODO: Performance Benchmarking
  - Tools to compare storage efficiency, memory usage, and decoding speed between channel-wise and full-image approaches.

1 LitData: A software that scales data processing tasks on local or cloud machines and enables optimizing datasets to accelerate AI model training and work with large remote datasets without local loading.

Technical University of Munich
Chair of Data Science in Earth Observation

Deployment - Dataset4EO – User Example

```
import dataset4eo as eodata
import litdata as ld
import time
from huggingface_hub import snapshot_download
repo_id = eodata.builtin_datasets['so2sat']            # get the repo_id of a built-in dataset, So2Sat as an example.

local_path = snapshot_download(                        # fetch the huggingface repo.
    repo_id=repo_id,
    repo_type = "dataset",
    cache_dir="./data_so2sat",
    revision="main"
)
split = "train"

train_dataset = eodata.StreamingDataset(               # load the downloaded chunk files.
    input_dir=f"{local_path}/{split}",
    num_channels=18,
    channels_to_select=[0, 3, 5, 7, 9],                # you can also choose to load only some channels.
    shuffle=True,
    drop_last=True
)
train_dataloader = ld.StreamingDataLoader(train_dataset) # here you go, you get the GPU-ready batches!
```

Technical University of Munich
Chair of Data Science in Earth Observation

Deployment - Dataset4EO

- More information on our GitHub repository.
- Join us to contribute as a developer!

Dataset for EO
EARTH NETS

Technical University of Munich
Chair of Data Science in Earth Observation

ExEBench

- Benchmarking Foundation Models on Extreme Earth Events

- Challenges
  - Heterogenous data sources
  - Far more complex scenarios
  - Limited finetuning data
  - Training biases

Paper Link

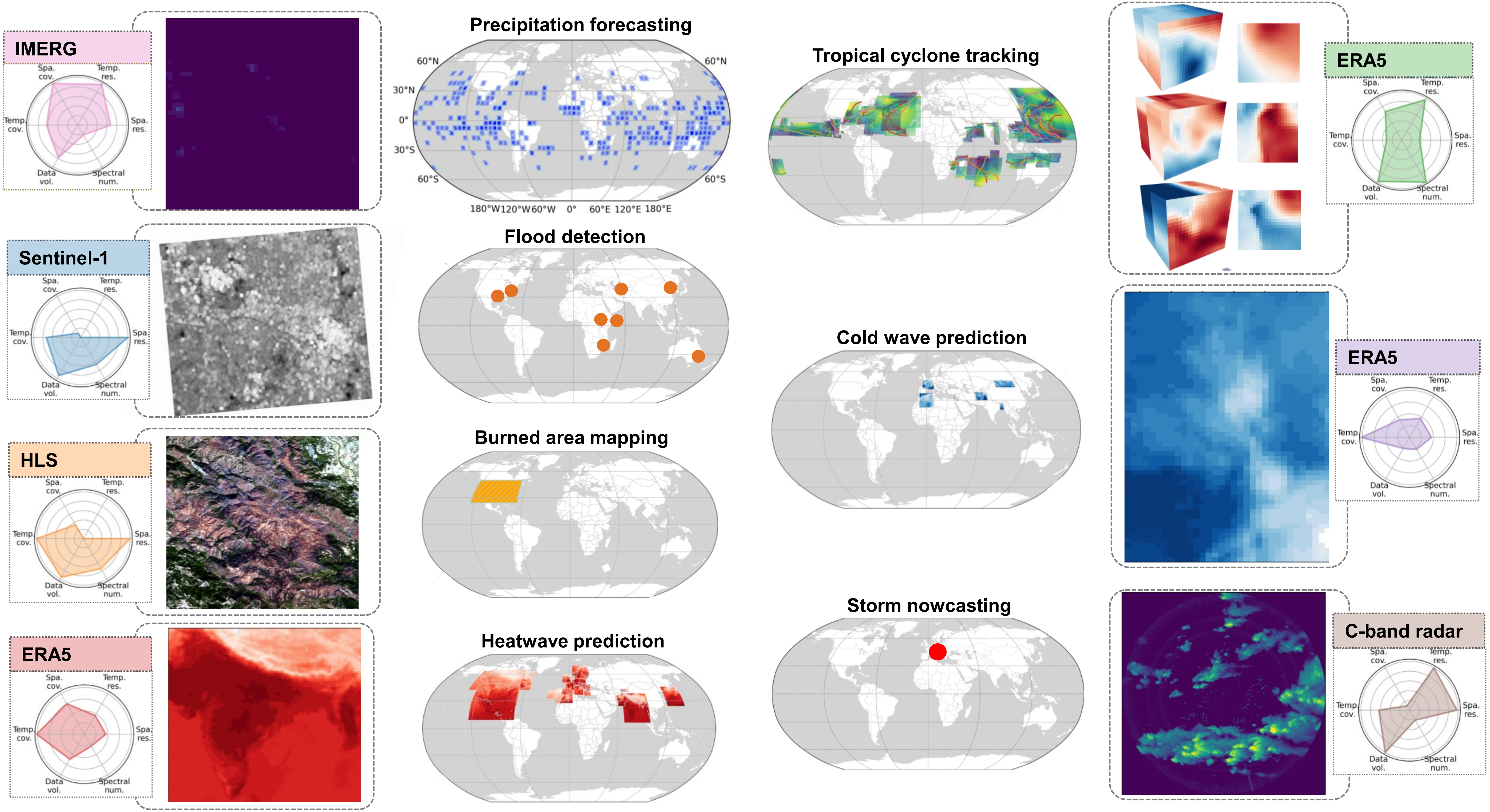
gaps

Foundation Models ⟷ Extreme Earth Events

Unified backbone

Diverse data and tasks

Zhao et al., 2025


IMERG
Sentinel-1
HLS
ERA5
ERA5
ERA5
C-band radar
Spa. cov.
Temp. res.
Temp. cov.
Spa. res.
Data vol.
Spectral num.
Precipitation forecasting
60°N
30°N
0°
30°S
60°S
180°W120°W60°W 0° 60°E 120°E 180°E
Flood detection
Burned area mapping
Heatwave prediction
Tropical cyclone tracking
Cold wave prediction
Storm nowcasting

Technical University of Munich
Chair of Data Science in Earth Observation

ExEBench

- Assess FM generalizability across diverse, high-impact tasks and domains.
- Promote novel ML methods that benefit disaster management.
- Analyze the interactions and cascading effects of extreme event

➢ To get the ML task on one extreme event
  - `ee_task = EETask(disaster="coldwave")`

➢ To get the dataset
  - `ee_task.get_loader()`

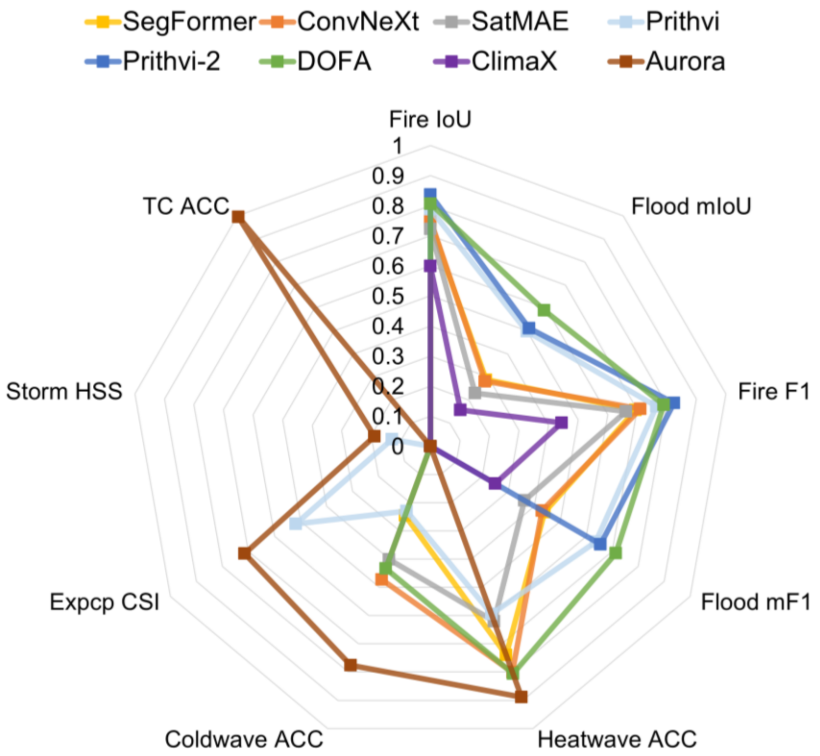
➢ To train and test the model
  - `ee_task.train_and_evaluate(mode="fully_finetune")`

Datasets
- 2 EO events
- 5 weather events
- Dataloader

Baseline models
- FM for WxC
- FM for RS
- FM for CV

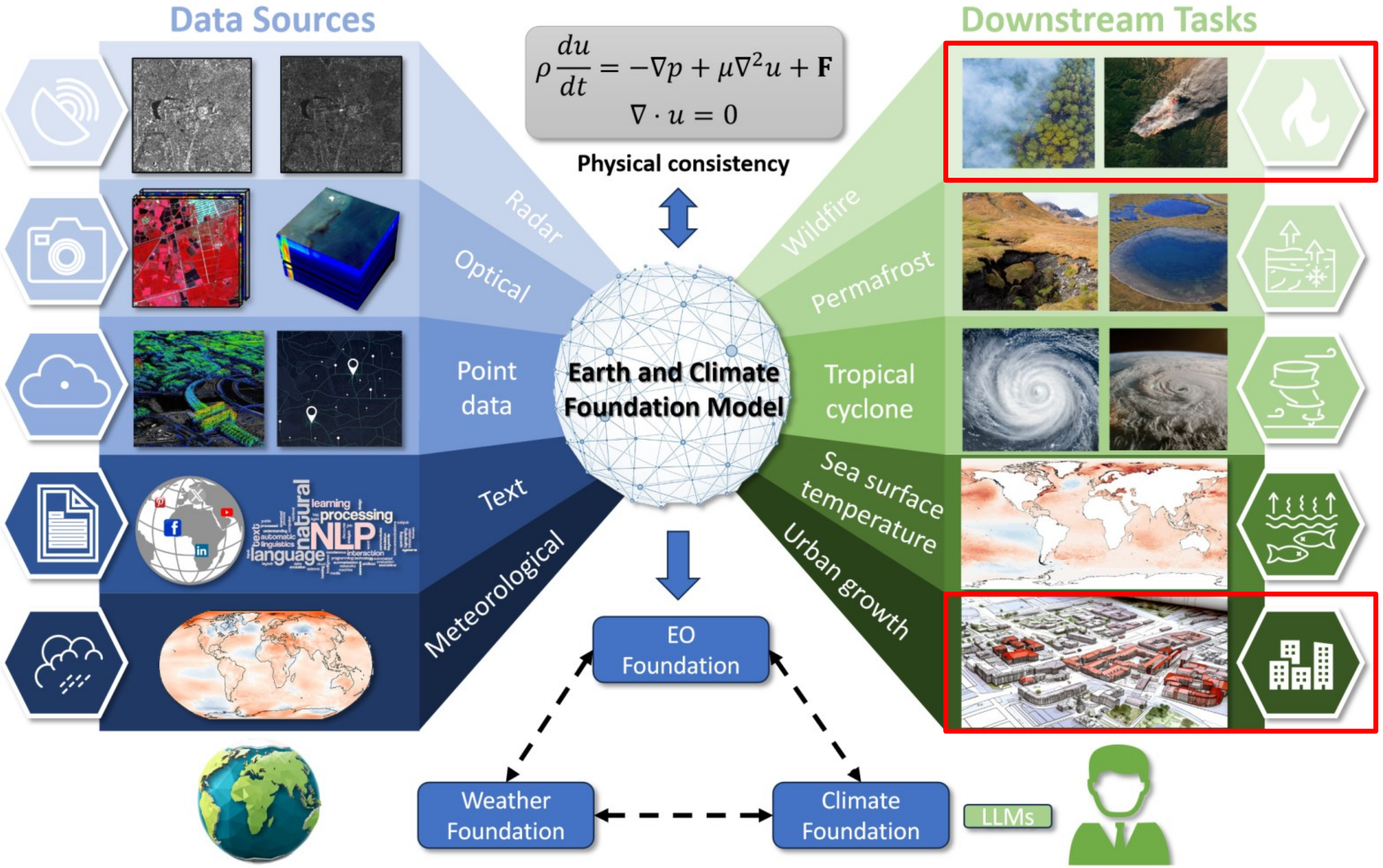
Evaluation
- Logging files
- Evaluation

Training
- Finetuning
- Training pipeline in PyTorch



Technical University of Munich
Chair of Data Science in Earth Observation

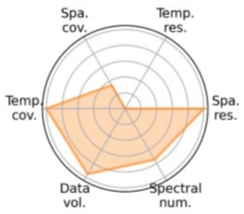
Application Cases and Hands-on Session



Zhu et al., 2024

Technical University of Munich
Chair of Data Science in Earth Observation

Burned Area Mapping with ExEBench



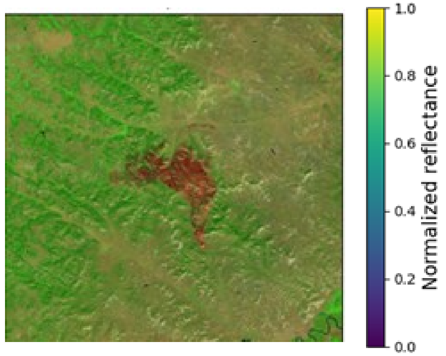
- Motivation
  - Visual appealing features
  - Moderate size

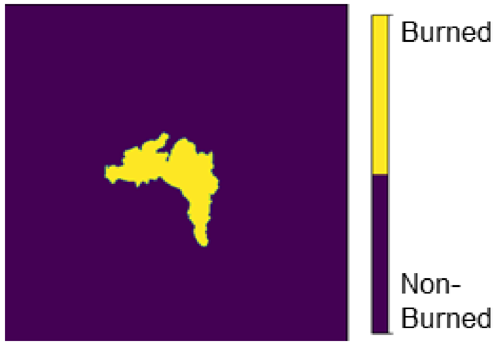
- Data
  - Image

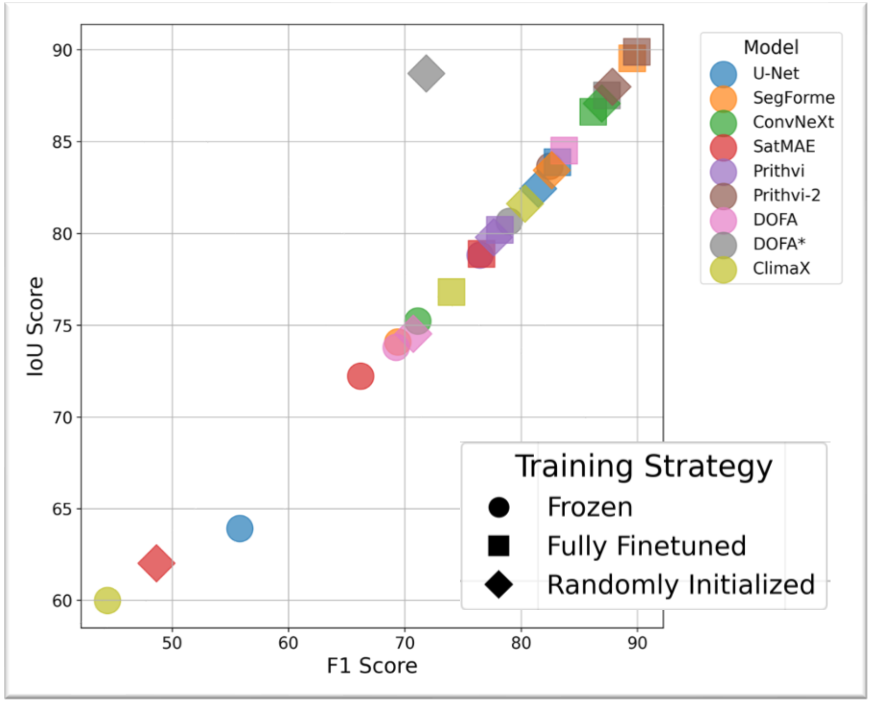
    1, Blue, B02;2, Green, B03; **3, Red, B04; 4, NIR, B8A; 5, SW 1, B11; 6, SW 2, B12**

  - Label

    0, Non burned; 1, Burned



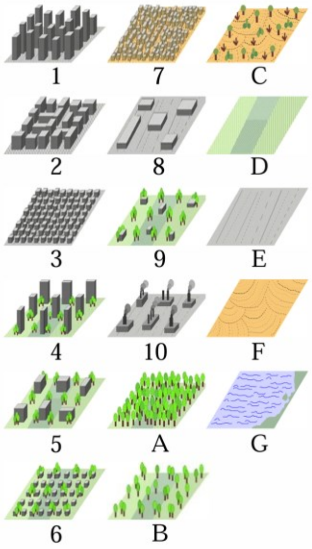


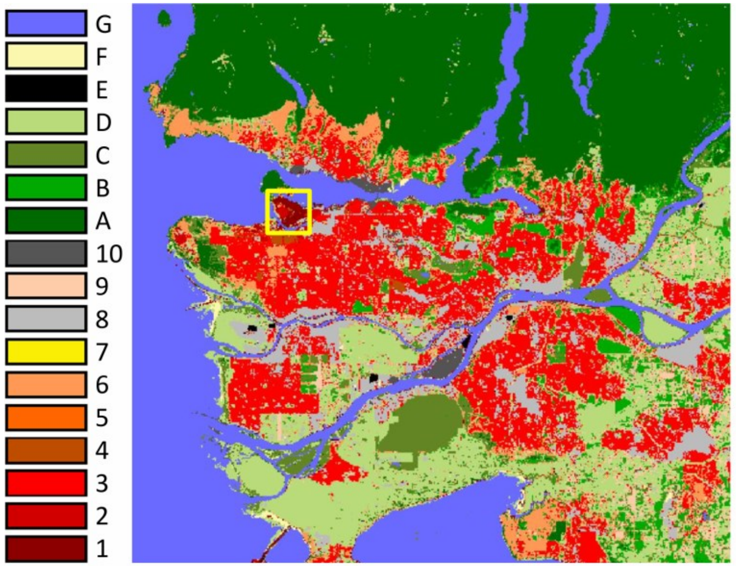


Technical University of Munich
Chair of Data Science in Earth Observation

Local Climate Zone (LCZ) Classification with So2Sat-LCZ42

- LCZ
  - Based on climate-relevant surface properties on local scale, mainly related to 3D surface structure, surface cover, and human factors.
  - Shows great potential in urban morphology mapping.



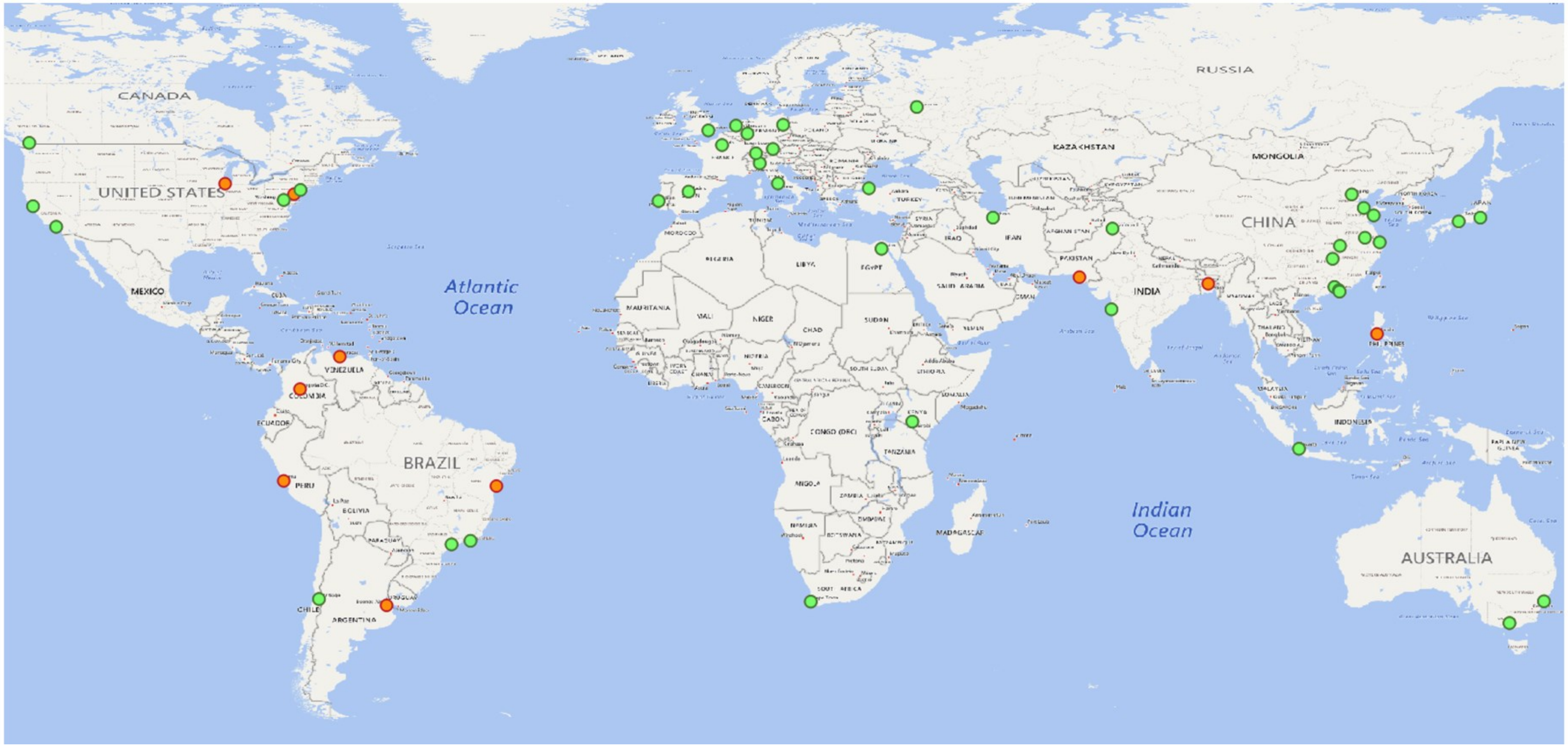


Zhu et al., 2019

Technical University of Munich
Chair of Data Science in Earth Observation

Local Climate Zone (LCZ) Classification with So2Sat-LCZ42

- So2Sat-LCZ42
  - A benchmark dataset for global LCZ classification.
  - 42 urban clusters around the globe.
  - 400,673 pairs of Sentinel-1 SAR and Sentinel-2 multi-spectral images.



Technical University of Munich
Chair of Data Science in Earth Observation

Notebook

https://shorturl.at/m9Ccf

Technical University of Munich
Chair of Data Science in Earth Observation

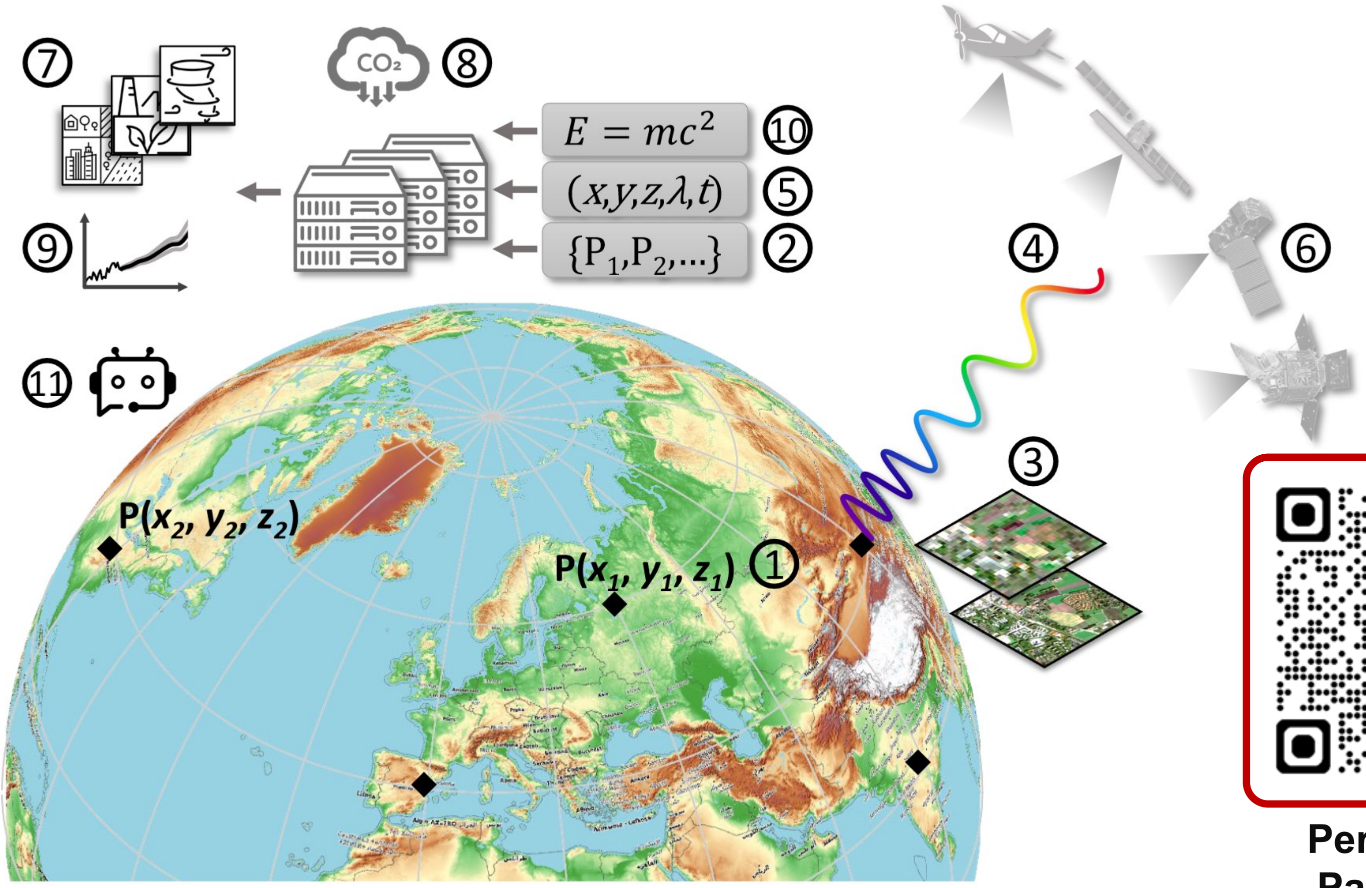
Outlooks – the Ideal Earth and Climate Foundation Models

Must-have

1. Geolocation embedding
2. Balanced geographical representations
3. Scale awareness
4. Wavelength embedding
5. The time variable
6. Multisensory
7. Task-agnostic
8. Carbon minimized

Highly desirable

9. Uncertainty quantification
10. Physical consistency
11. AI assistants



Perspective Paper Link

Zhu et al., 2024

Technical University of Munich
Chair of Data Science in Earth Observation

Outlooks – How to build new foundation models?

Monday - 23.06.2025

09:00

D.02.19 TUTORIAL - Foundation Models for Remote Sensing Applications

Chair(s)

N/A

Room

Room 0.11/0.12

Duration

80 Minutes

Details

Foundation models are Artificial Intelligence (AI) models trained on vast datasets that can be effectively adapted to a wide range of general tasks. Their recent emergence has brought a paradigm shift in AI, enabling solutions to complex problems with minimal task-specific data, often surpassing traditional machine learning models trained from scratch.

References (1)

- Introduction
  - Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, et al. "On the Opportunities and Risks of Foundation Models." arXiv, July 12, 2022. https://doi.org/10.48550/arXiv.2108.07258.
  - **Zhu, Xiao Xiang, Zhitong Xiong, Yi Wang, Adam J. Stewart, Konrad Heidler, Yuanyuan Wang, Zhenghang Yuan, Thomas Dujardin, Qingsong Xu, and Yilei Shi. "On the Foundations of Earth and Climate Foundation Models." arXiv, May 7, 2024. https://doi.org/10.48550/arXiv.2405.04285.**
- Data (2)
  - Neumann, Maxim, Andre Susano Pinto, Xiaohua Zhai, and Neil Houlsby. "In-Domain Representation Learning for Remote Sensing." arXiv, November 15, 2019. https://doi.org/10.48550/arXiv.1911.06721.
  - **Schmitt, Michael, Lloyd Haydn Hughes, Chunping Qiu, and Xiao Xiang Zhu. "SEN12MS -- A Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion." arXiv, June 18, 2019. https://doi.org/10.48550/arXiv.1906.07789.**
  - Mañas, Oscar, Alexandre Lacoste, Xavier Giró-i-Nieto, David Vazquez, and Pau Rodríguez. "Seasonal Contrast: Unsupervised Pre-Training From Uncurated Remote Sensing Data," 9414–23, 2021. https://openaccess.thecvf.com/content/ICCV2021/html/Manas_Seasonal_Contrast_Unsupervised_Pre-Training_From_Uncurated_Remote_Sensing_Data_ICCV_2021_paper.html.

References (2)

- Data (2)
  - **Wang, Yi, Nassim Ait Ali Braham, Zhitong Xiong, Chenying Liu, Conrad M. Albrecht, and Xiao Xiang Zhu. "SSL4EO-S12: A Large-Scale Multimodal, Multitemporal Dataset for Self-Supervised Learning in Earth Observation [Software and Data Sets]." *IEEE Geoscience and Remote Sensing Magazine* 11, no. 3 (September 2023): 98–106. https://doi.org/10.1109/MGRS.2023.3281651.**
  - Bastani, Favyen, Piper Wolters, Ritwik Gupta, Joe Ferdinando, and Aniruddha Kembhavi. "SatlasPretrain: A Large-Scale Dataset for Remote Sensing Image Understanding." In 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 16726–36. Paris, France: IEEE, 2023. https://doi.org/10.1109/ICCV51070.2023.01538.
  - **Stewart, Adam J, Nils Lehmann, Isaac A Corley, Yi Wang, Yi-Chia Chang, Nassim Ait Ali Braham, Shradha Sehgal, Caleb Robinson, and Arindam Banerjee. "SSL4EO-L: Datasets and Foundation Models for Landsat Imagery," n.d.**
  - Braham, Nassim Ait Ali, Conrad M. Albrecht, Julien Mairal, Jocelyn Chanussot, Yi Wang, and Xiao Xiang Zhu. "SpectralEarth: Training Hyperspectral Foundation Models at Scale." arXiv, August 15, 2024. https://doi.org/10.48550/arXiv.2408.08447.
  - Nedungadi, Vishal, Ankit Kariryaa, Stefan Oehmcke, Serge Belongie, Christian Igel, and Nico Lang. "MMEarth: Exploring Multi-Modal Pretext Tasks For Geospatial Representation Learning." arXiv, July 29, 2024. https://doi.org/10.48550/arXiv.2405.02771.
  - **Xiong, Zhitong, Fahong Zhang, Yi Wang, Yilei Shi, and Xiao Xiang Zhu. "EarthNets: Empowering Artificial Intelligence for Earth Observation." IEEE Geoscience and Remote Sensing Magazine, 2024, 2–36. https://doi.org/10.1109/MGRS.2024.3466998.**

References (3)

- Training – How to Train (1)
  - Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks." In Advances in Neural Information Processing Systems, Vol. 25. Curran Associates, Inc., 2012. https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.
  - He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." arXiv, December 10, 2015. https://doi.org/10.48550/arXiv.1512.03385.
  - Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, et al. "An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv, June 3, 2021. https://doi.org/10.48550/arXiv.2010.11929.
  - Liu, Ze, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows." arXiv, August 17, 2021. https://doi.org/10.48550/arXiv.2103.14030.
  - Vincenzi, Stefano, Angelo Porrello, Pietro Buzzega, Marco Cipriano, Pietro Fronte, Roberto Cuccu, Carla Ippoliti, Annamaria Conte, and Simone Calderara. "The Color out of Space: Learning Self-Supervised Representations for Earth Observation Imagery." arXiv, June 22, 2020. https://doi.org/10.48550/arXiv.2006.12119.
  - Ayush, Kumar, Burak Uzkent, Chenlin Meng, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. "Geography-Aware Self-Supervised Learning." arXiv, March 8, 2022. https://doi.org/10.48550/arXiv.2011.09980.

References (4)

- Training – How to Train (2)
  - Chen, Ting, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. "A Simple Framework for Contrastive Learning of Visual Representations." arXiv, July 1, 2020. https://doi.org/10.48550/arXiv.2002.05709.
  - Mañas, Oscar, Alexandre Lacoste, Xavier Giró-i-Nieto, David Vazquez, and Pau Rodríguez. "Seasonal Contrast: Unsupervised Pre-Training From Uncurated Remote Sensing Data," 9414–23, 2021. https://openaccess.thecvf.com/content/ICCV2021/html/Manas_Seasonal_Contrast_Unsupervised_Pre-Training_From_Uncurated_Remote_Sensing_Data_ICCV_2021_paper.html.
  - Jean, Neal, Sherrie Wang, Anshul Samar, George Azzari, David Lobell, and Stefano Ermon. "Tile2Vec: Unsupervised Representation Learning for Spatially Distributed Data." Proceedings of the AAAI Conference on Artificial Intelligence 33, no. 01 (July 17, 2019): 3967–74. https://doi.org/10.1609/aaai.v33i01.33013967.
  - He, Kaiming, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. "Masked Autoencoders Are Scalable Vision Learners." arXiv, December 19, 2021. https://doi.org/10.48550/arXiv.2111.06377.
  - "SatMAE: Pre-Training Transformers for Temporal and Multi-Spectral Satellite Imagery." Accessed June 18, 2025. https://sustainlab-group.github.io/SatMAE/.
  - **Wang, Yi, Hugo Hernández Hernández, Conrad M. Albrecht, and Xiao Xiang Zhu. "Feature Guided Masked Autoencoder for Self-Supervised Learning in Remote Sensing." arXiv, October 28, 2023. https://doi.org/10.48550/arXiv.2310.18653.**

References (5)

- Training – How to Train (3)
  - **Wang, Yi, Conrad M. Albrecht, Nassim Ait Ali Braham, Chenying Liu, Zhitong Xiong, and Xiao Xiang Zhu. "Decoupling Common and Unique Representations for Multimodal Self-Supervised Learning." In Computer Vision – ECCV 2024, edited by Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol, 15087:286–303. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2025. https://doi.org/10.1007/978-3-031-73397-0_17.**
- Training – Models
  - Tseng, Gabriel, Ruben Cartuyvels, Ivan Zvonkov, Mirali Purohit, David Rolnick, and Hannah Kerner. "Lightweight, Pre-Trained Transformers for Remote Sensing Timeseries." arXiv, February 5, 2024. https://doi.org/10.48550/arXiv.2304.14065.
  - Szwarcman, Daniela, Sujit Roy, Paolo Fraccaro, Þorsteinn Elí Gíslason, Benedikt Blumenstiel, Rinki Ghosal, Pedro Henrique de Oliveira, et al. "Prithvi-EO-2.0: A Versatile Multi-Temporal Foundation Model for Earth Observation Applications." arXiv, February 3, 2025. https://doi.org/10.48550/arXiv.2412.02732.
  - **Xiong, Zhitong, Yi Wang, Fahong Zhang, Adam J. Stewart, Joëlle Hanna, Damian Borth, Ioannis Papoutsis, Bertrand Le Saux, Gustau Camps-Valls, and Xiao Xiang Zhu. "Neural Plasticity-Inspired Multimodal Foundation Model for Earth Observation." arXiv, June 7, 2024. https://doi.org/10.48550/arXiv.2403.15356.**
  - Jakubik, Johannes, Felix Yang, Benedikt Blumenstiel, Erik Scheurer, Rocco Sedona, Stefano Maurogiovanni, Jente Bosmans, et al. "TerraMind: Large-Scale Generative Multimodality for Earth Observation." arXiv, June 11, 2025. https://doi.org/10.48550/arXiv.2504.11171.

References (6)

- Adaptation and Deployment (1)
  - Tseng, Gabriel, Ruben Cartuyvels, Ivan Zvonkov, Mirali Purohit, David Rolnick, and Hannah Kerner. "Lightweight, Pre-Trained Transformers for Remote Sensing Timeseries." arXiv, February 5, 2024. https://doi.org/10.48550/arXiv.2304.14065.
  - Szwarcman, Daniela, Sujit Roy, Paolo Fraccaro, Þorsteinn Elí Gíslason, Benedikt Blumenstiel, Rinki Ghosal, Pedro Henrique de Oliveira, et al. "Prithvi-EO-2.0: A Versatile Multi-Temporal Foundation Model for Earth Observation Applications." arXiv, February 3, 2025. https://doi.org/10.48550/arXiv.2412.02732.
  - Jakubik, Johannes, Felix Yang, Benedikt Blumenstiel, Erik Scheurer, Rocco Sedona, Stefano Maurogiovanni, Jente Bosmans, et al. "TerraMind: Large-Scale Generative Multimodality for Earth Observation." arXiv, June 11, 2025. https://doi.org/10.48550/arXiv.2504.11171.
  - Chang, Li, Jianhong Qiu, Wei Yang, Yahui Jia, Yan Shufen, Jingyi Yin, and Shufen Zhihao. "Making Foundation Models Adaptable: A Review of Parameter-Efficient Fine-Tuning Approaches," 2025. https://www.authorea.com/users/906274/articles/1280646-making-foundation-models-adaptable-a-review-of-parameter-efficient-fine-tuning-approaches.
  - Chen, Zhe, Yuchen Duan, Wenhai Wang, Junjun He, Tong Lu, Jifeng Dai, and Yu Qiao. "Vision Transformer Adapter for Dense Predictions." arXiv, February 13, 2023. https://doi.org/10.48550/arXiv.2205.08534.
  - Hu, Edward J., Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. "LoRA: Low-Rank Adaptation of Large Language Models." arXiv, October 16, 2021. https://doi.org/10.48550/arXiv.2106.09685.

References (7)

- Adaptation and Deployment (2)
  - Lester, Brian, Rami Al-Rfou, and Noah Constant. "The Power of Scale for Parameter-Efficient Prompt Tuning." arXiv, September 2, 2021. https://doi.org/10.48550/arXiv.2104.08691.
  - Li, Xiang Lisa, and Percy Liang. "Prefix-Tuning: Optimizing Continuous Prompts for Generation." arXiv, January 1, 2021. https://doi.org/10.48550/arXiv.2101.00190.
  - Gomes, Carlos, Benedikt Blumenstiel, Joao Lucas de Sousa Almeida, Pedro Henrique de Oliveira, Paolo Fraccaro, Francesc Marti Escofet, Daniela Szwarcman, Naomi Simumba, Romeo Kienzler, and Bianca Zadrozny. "TerraTorch: The Geospatial Foundation Models Toolkit." arXiv, March 26, 2025. https://doi.org/10.48550/arXiv.2503.20563.
  - Phillips, Christopher, et al. HLS Foundation Burnscars Dataset. Aug. 2023, https://huggingface.co/ibm-nasa-geospatial/hls_burn_scars. DOI: 10.57967/hf/0956.
  - **Stewart, Adam J., Caleb Robinson, Isaac A. Corley, Anthony Ortiz, Juan M. Lavista Ferres, and Arindam Banerjee. "TorchGeo: Deep Learning With Geospatial Data." arXiv, September 17, 2022. https://doi.org/10.48550/arXiv.2111.08872.**
  - **Zhao, Shan, Zhitong Xiong, Jie Zhao, and Xiao Xiang Zhu. "ExEBench: Benchmarking Foundation Models on Extreme Earth Events." arXiv, May 13, 2025. https://doi.org/10.48550/arXiv.2505.08529.**
  - **Zhu, Xiao Xiang, Jingliang Hu, Chunping Qiu, Yilei Shi, Jian Kang, Lichao Mou, Hossein Bagheri, et al. "So2Sat LCZ42: A Benchmark Dataset for Global Local Climate Zones Classification." arXiv, December 19, 2019. https://doi.org/10.48550/arXiv.1912.12171.**

References (8)

- Outlooks

  - **Zhu, Xiao Xiang, Zhitong Xiong, Yi Wang, Adam J. Stewart, Konrad Heidler, Yuanyuan Wang, Zhenghang Yuan, Thomas Dujardin, Qingsong Xu, and Yilei Shi. "On the Foundations of Earth and Climate Foundation Models." arXiv, May 7, 2024. https://doi.org/10.48550/arXiv.2405.04285.**

Thanks for your attention!

Sining Chen, Shan Zhao
{sining.chen, shan.zhao}@tum.de

Living Planet Symposium 2025
June 22, 2025

Check our GitHub repository for the materials