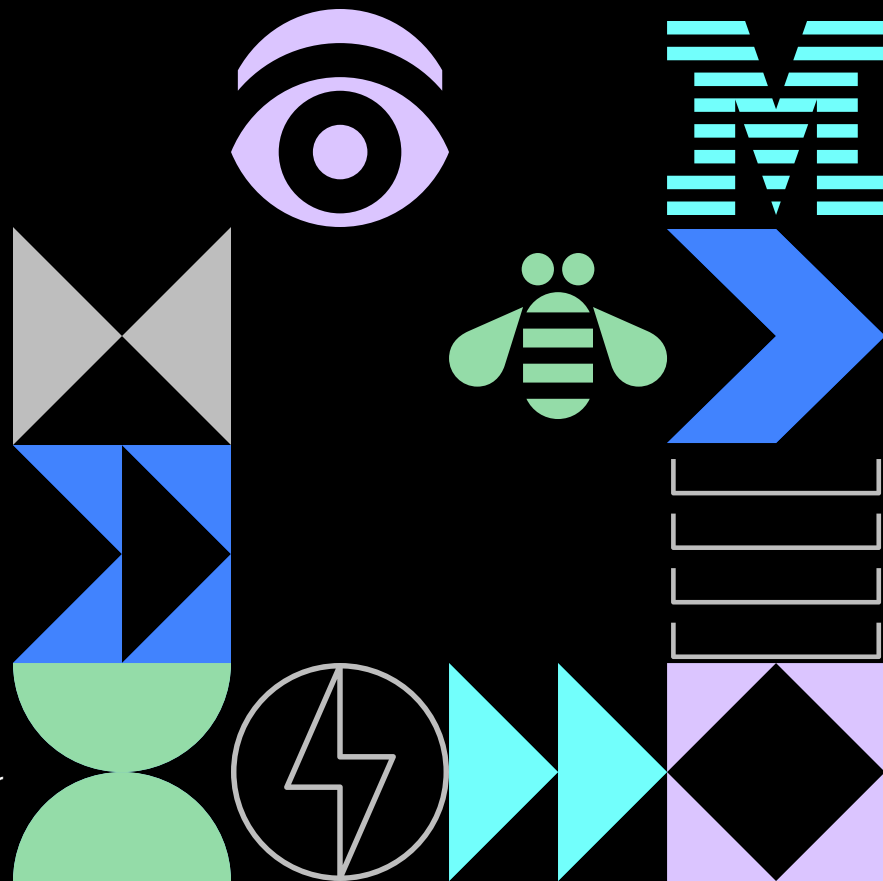


地理空間基盤モデルPrithviと その将来

木村大毅

日本IBM株式会社 Senior Account Technical Leader

電気通信大学 連携准教授



自己紹介 (<http://ibm.biz/daiki-kimura>)

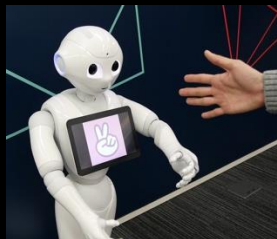


2015年 東京工業大学 博士課程修了、中学・高校教員免許取得（数学）

日本学術振興会 特別研究員 等を兼任，Microsoft Research Asia，スイスの大学で研究インターン

2015年 楽天技術研究所に入所 画像処理の研究に従事

2016年～2025年 IBM東京基礎研究所に入所（一時期、所長技術秘書に就任）



ゲームAI



病気の判定補助AI



ボトル検出AI

```
You are hungry! Let's cook a meal. Check the cookbook in the
kitchen for the recipe. Once done, enjoy your meal!

-- Kitchen --
You arrive in a typical kitchen. You begin to look around.

You see a closed fridge in the corner. You can see a stove.
But, it might come in handy, but there isn't anything on
it now. You see a glass over in a corner, where you can see a
counter. Make a note of this, you might have to put stuff on
it later. You do see a knife and a cookbook on it now.

There is a closed door leading south.

> examine inventory
You are carrying: a carrot.
> take knife from counter
Taken.
> alice carrot
What do you want to do with the carrot with?
> the knife
You slice the carrot.
[Your score increased by one point.]
```

説明性の高いAI



衛星画像AI

2025年～現在 政府機関/病院担当 Senior Account Technical Leader

2025年～現在 IBM Quantum Ambassador（兼務）

2025年～現在 電気通信大学 連携准教授 画像処理・AI担当（兼務）

2023年～現在 人工知能学会 理事 広報担当（兼務）



今日の流れ

- IBM Research（IBM基礎研究所）の簡単な紹介
- 基盤モデルの利点（少数データで高精度モデル）の紹介
- NASAとESA、IBMの衛星画像基盤モデル Prithviシリーズ
 - 概要
 - Prithvi 1.0: 基盤モデルとしての学習と、下流タスク、デモ
 - Prithvi 2.0: 1.0との違い
 - Prithvi WxC: 気象・気候
 - TerraMind: マルチモーダル生成
 - TerraTorch: プラットフォーム
- デジタルツインと、Prithviの将来

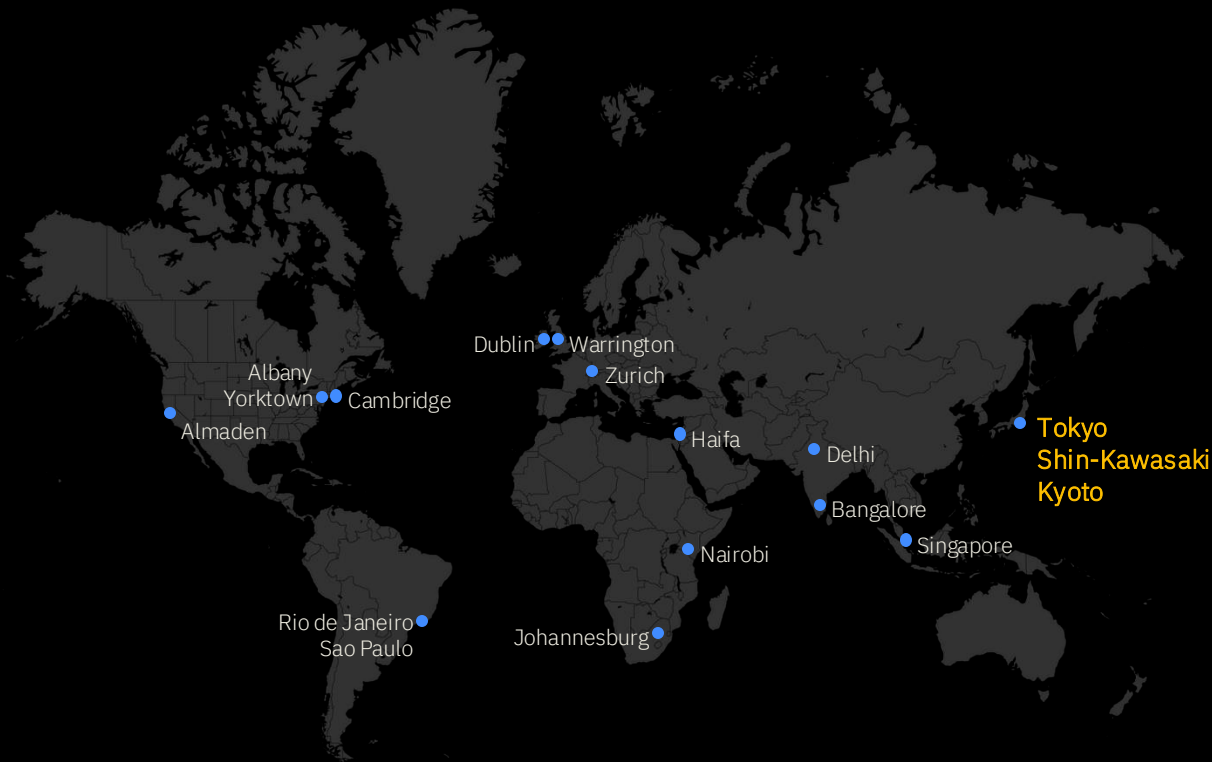
IBM Research

3000

Researchers

79

Year (since 1945)



6 Nobel Laureates



10 Medals of Technology

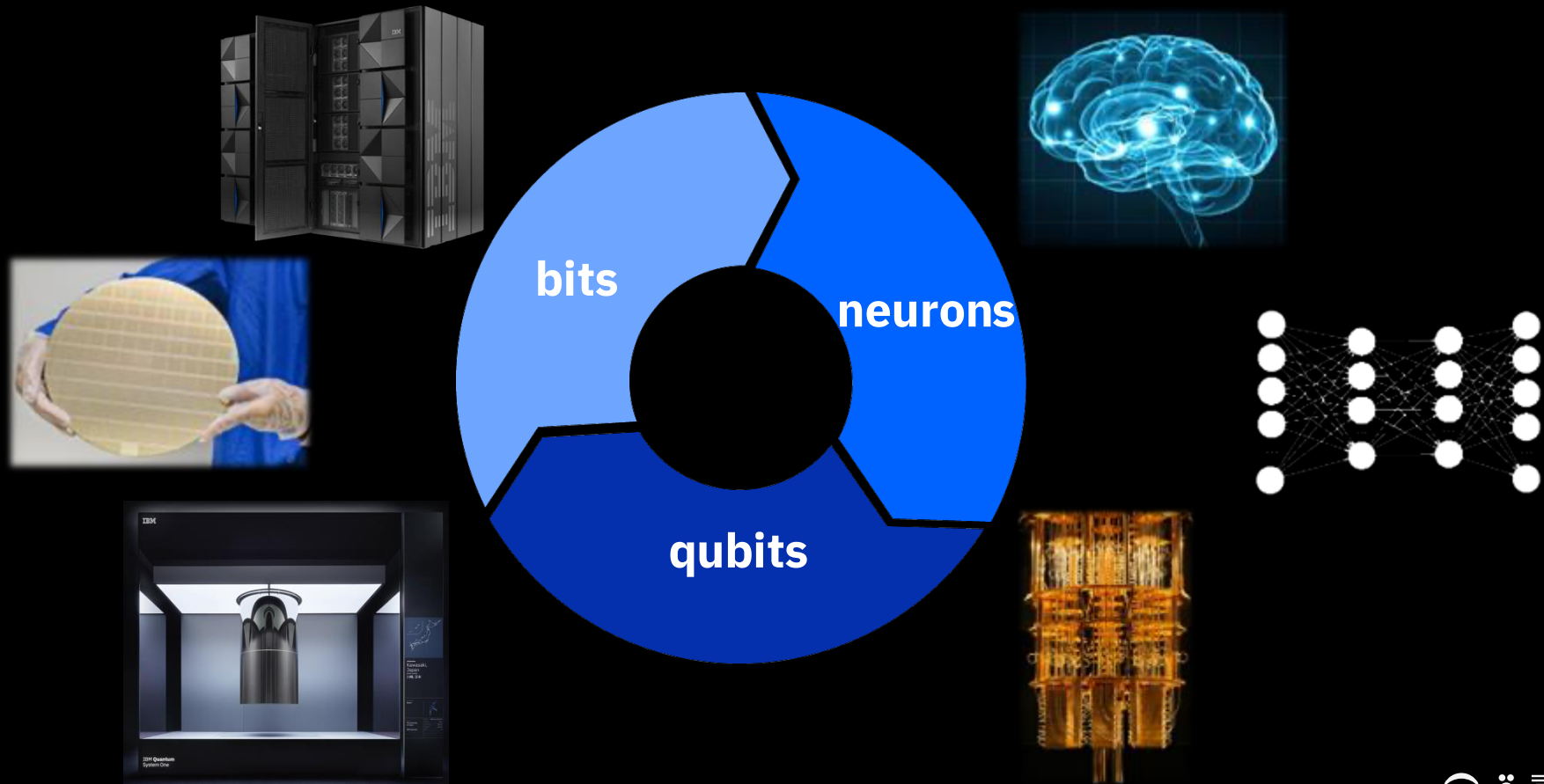


5 National Medals of Science

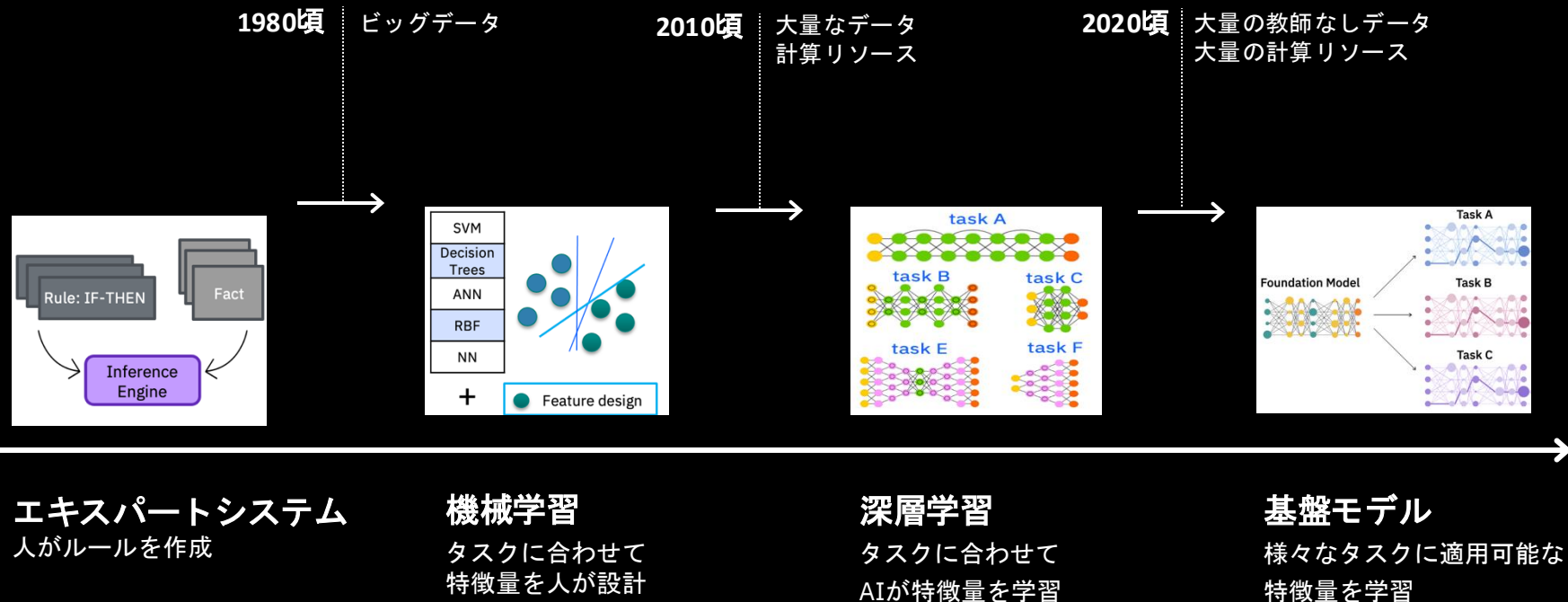


6 Turing Awards

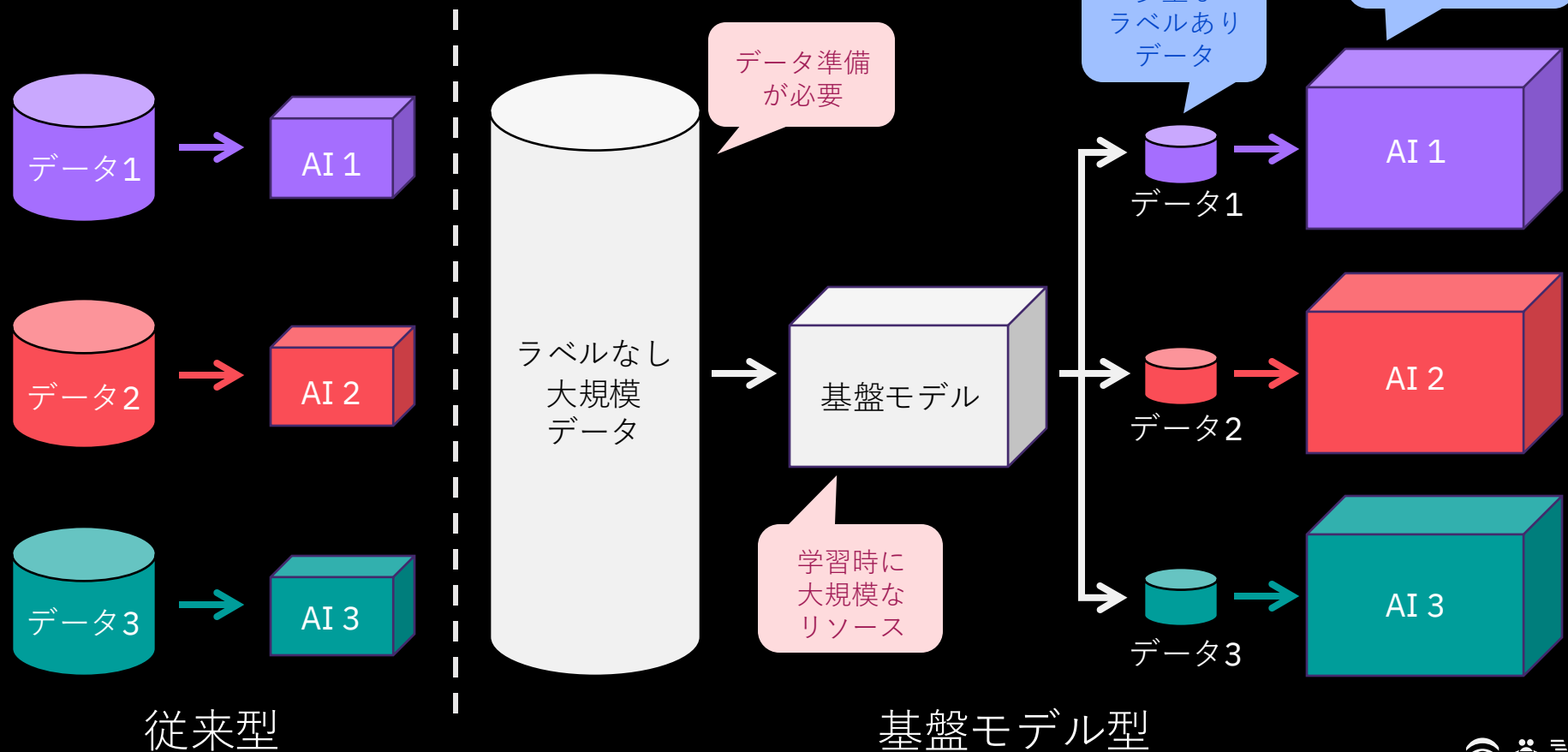
Researchの題目：コンピューティングの未来



AIの歴史



基盤モデルのメリットとデメリット



IBMがリリースしているオープンなモデル

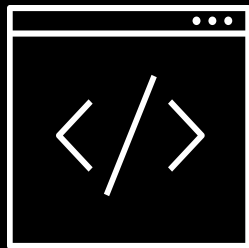


Granite

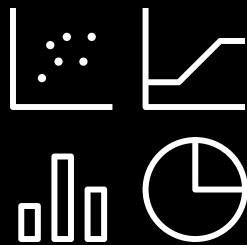
Hugging Face上で公開中 (<http://huggingface.co/ibm-granite>)



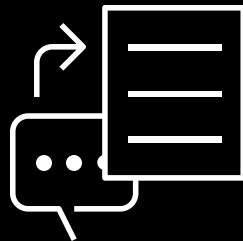
言語



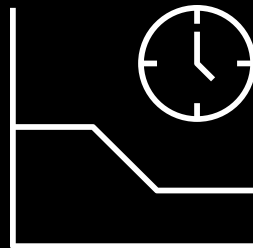
コード



画像



スピーチ

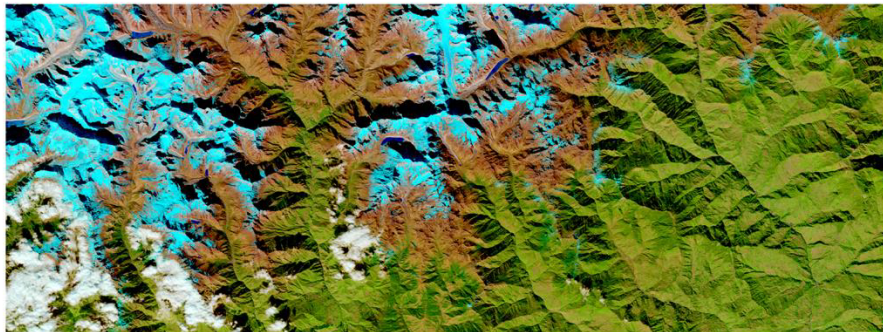


時系列



衛星画像

Feb 1, 2023



人

IBM and NASA plan to develop several new technologies to extract insights from Earth observations. One project will train an IBM geospatial intelligence foundation model on NASA's

[Subscribe to email](#)

2023年2月2日 12:11







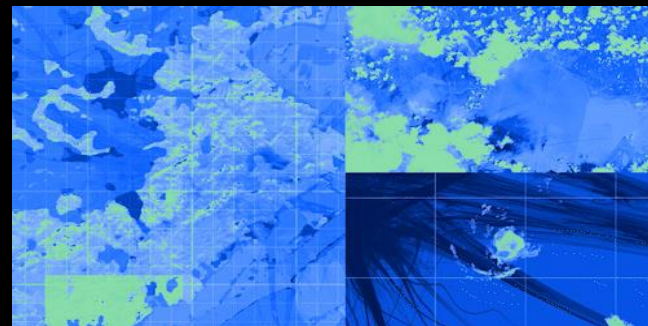

[米国ニューヨーク州アーモンク-2023年2月1日（現地時間）]

リリース本文中の「関連資料」は、こちらのURLからご覧ください。

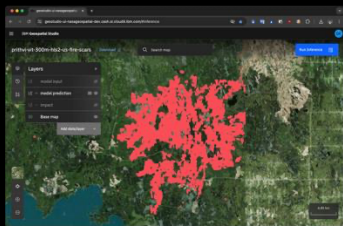
地理空間基盤モデル (Prithvi: Geospatial Foundation Model)

科学的理解と気候変動問題の迅速な解決を目的とした
NASAが保有する膨大な「衛星データ」を学習した基盤モデル

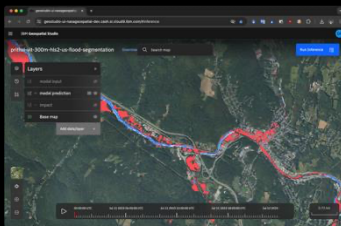
- 2023年2月
気候変動の影響に関する研究の協働開始
- 2023年8月
HuggingFace上にPrithvi 1.0の公開
- 2024年12月
HuggingFace上にPrithvi 2.0を公開



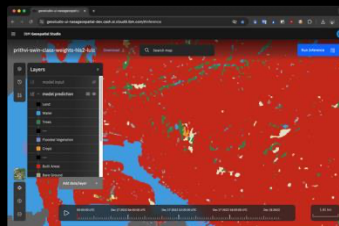
山火事



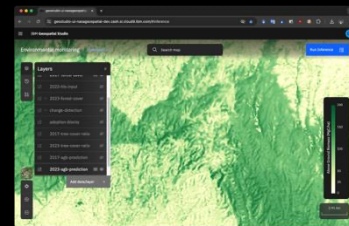
洪水



土地利用



地上部バイオマス

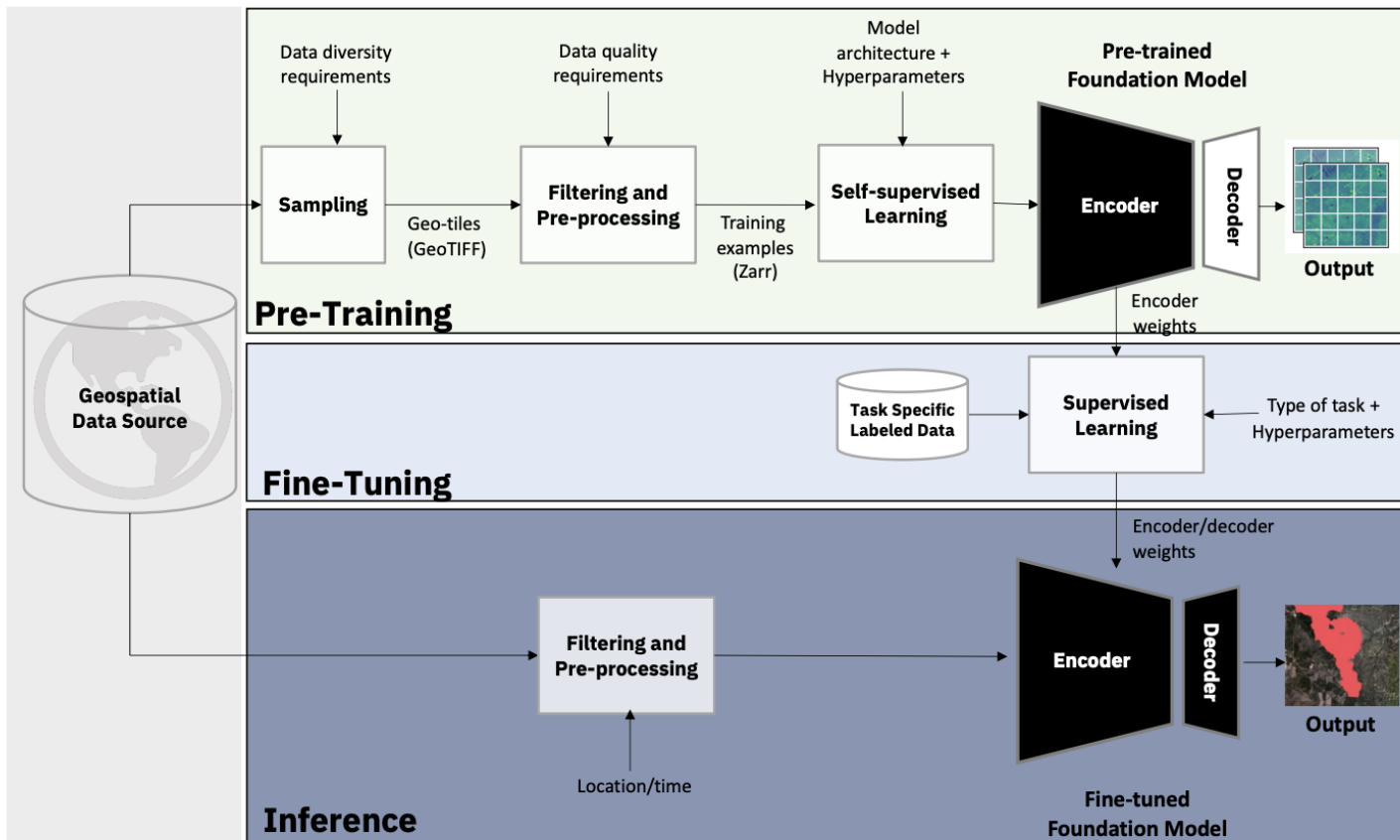


応用例)

※ <https://research.ibm.com/blog/nasa-hugging-face-ibm>

※ <https://www.ibm.com/blogs/solutions/jp-ja/nasa-hugging-face-ibm/> (日本語)

Prithvi 1.0 概要図



HLS as global remote sensing data to train GFM

Harmonized Landsat Sentinel-2:

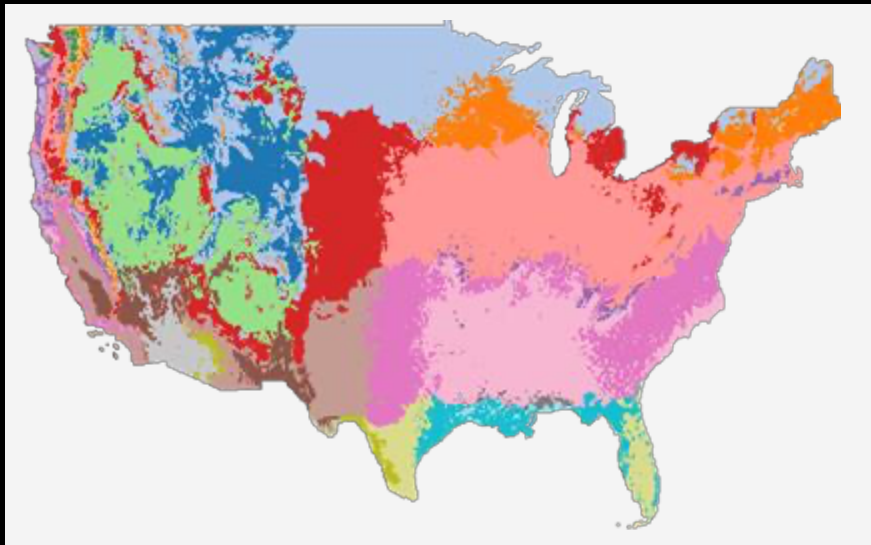
Geospatial foundation models
focused on *remote sensing* data.

Harmonized Landsat Sentinel-2
(*HLS*) provides consistent global
observations of the land.

- Data available in *tiles*, aligned with
the Military Grid Reference
System (MGRS).
- 30m resolution
- Each tile has **3660 x 3660 pixels**,
corresponding to ~110 x 110 km.



Diversified sampling to address unbalanced data for pre-training



Selecting pre-training data

Requirement → diversified pre-training dataset.

- For a given region, images can look similar across time.
- Random sampling → can bias towards most common landscapes.

Intelligent sampling scheme based on geospatial statistics.

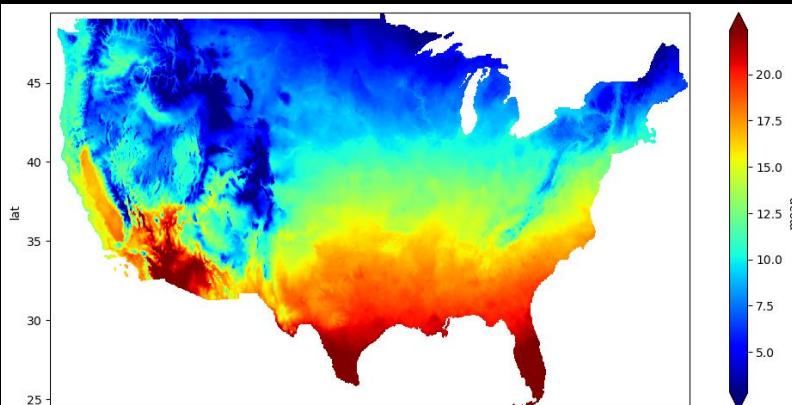
Sampling data from the United States

Sampling scheme

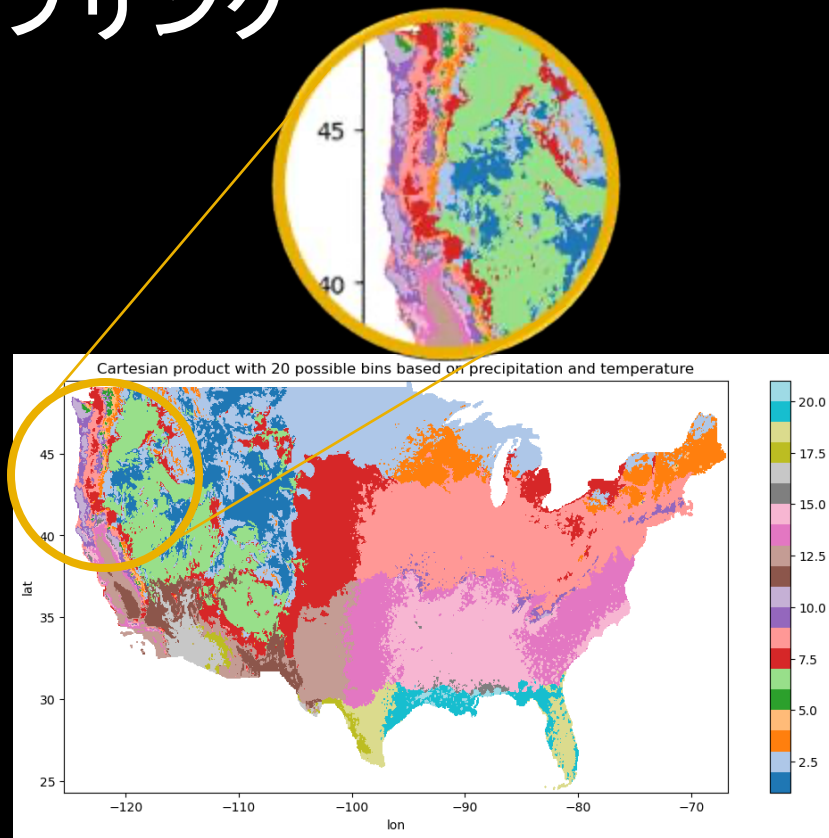
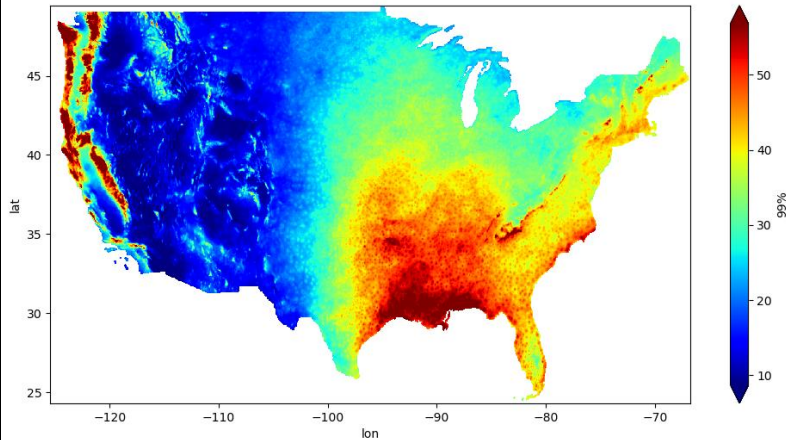
1. Aggregate various geospatial statistics (temperature and precipitation).
2. Divide the region into groups based on these statistics.
3. Sample HLS tiles as equally as possible from each group.

US領域（アラスカ除く）サンプリング

気温

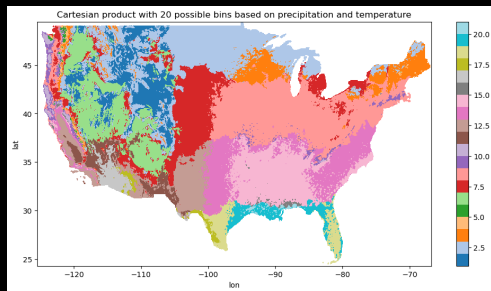


降水量

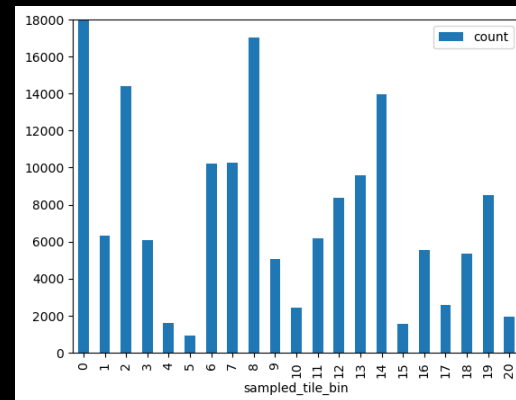
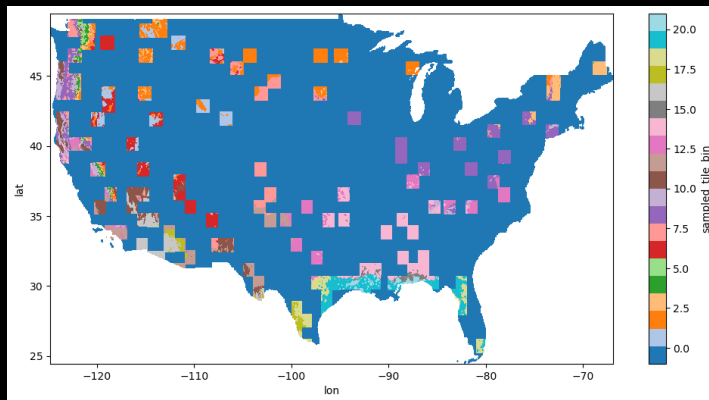


20特徴領域に分類

US領域（アラスカ除く）サンプリング



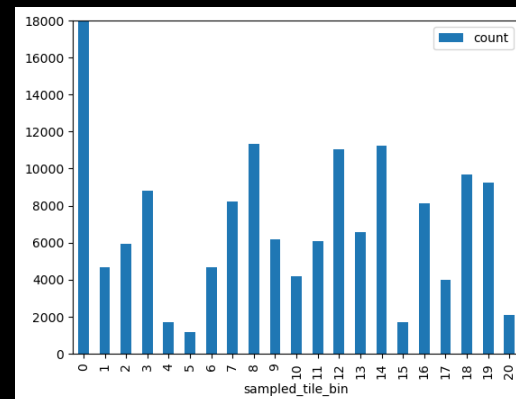
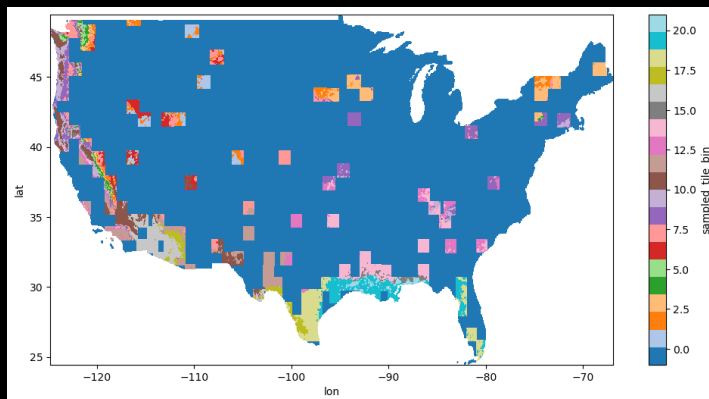
20特徴領域に分類



entropy: 4.005, (min: 0.000, max: 4.322)



サンプリングの効果:
985 tiles → 130 tiles
(~40TB → 5TB)



entropy: 4.114, (min: 0.000, max: 4.322)

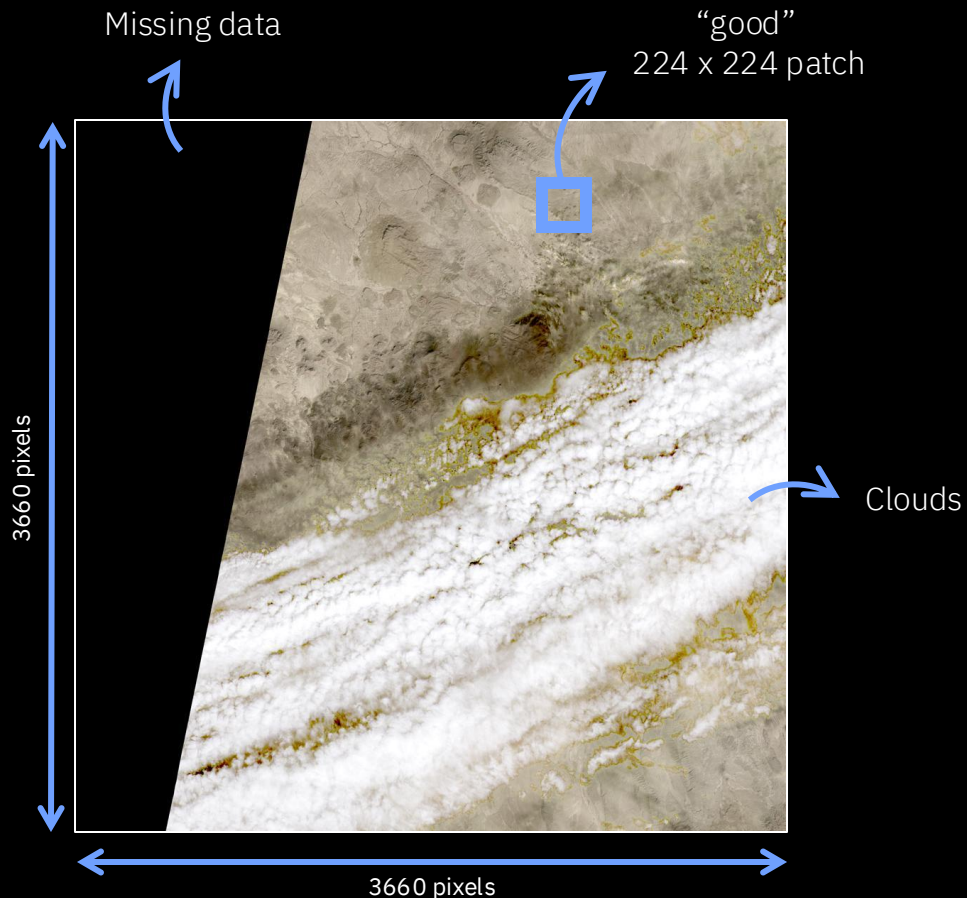
Filtering out “missing” data

Satellite images can contain significant parts filled with clouds or missing values.

For each tile, HLS provides a *cloud mask*.

The mask allows us to compute the percentage of clouds and missing data.

Only sub-regions with low percentage of clouds or missing data are considered.



⇒ US: 250k patches



3D ViT Model Arch. self-supervised via masking

MAE → Masked AutoEncoder

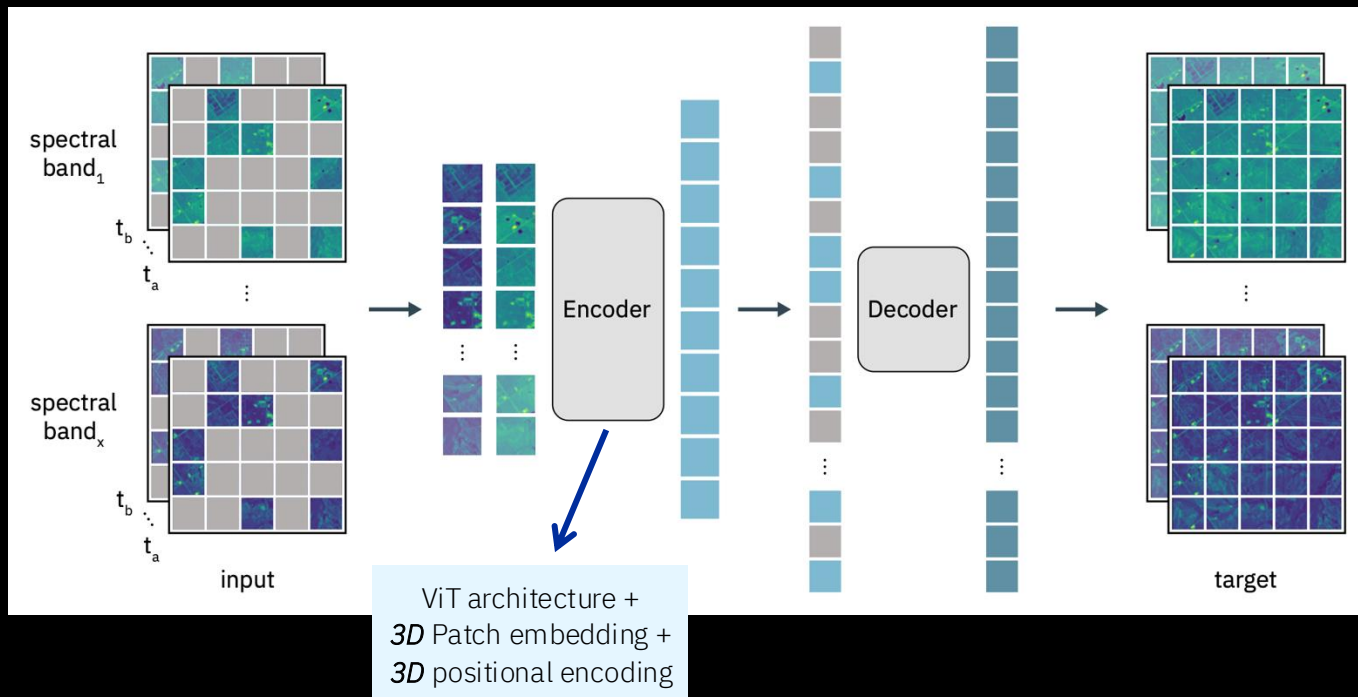
– Pre-training task: reconstruct
masked patches → target =
original data.

– MSE loss on *masked* patches.

Encoder → Vision transformer
(ViT) for multispectral **3D data**.

– 3D patch embeddings
– 3D positional encoding

Decoder → Transformer blocks
+ linear projection layer to
match the target patch size.



Input

number of frames	3
image size	224 x 224
bands	B02, B03, B04 (RGB) B05, B06, B07 (infrared)

Model

patch size (T, H, W)	(1, 16, 16)
masking ratio	0.75
number of parameters	100M

Training

machine	vela cluster
number of GPUs	64
number of CPUs	¹⁹ 512
time to train	~4.5 days

基盤モデルの精度



マスク画像

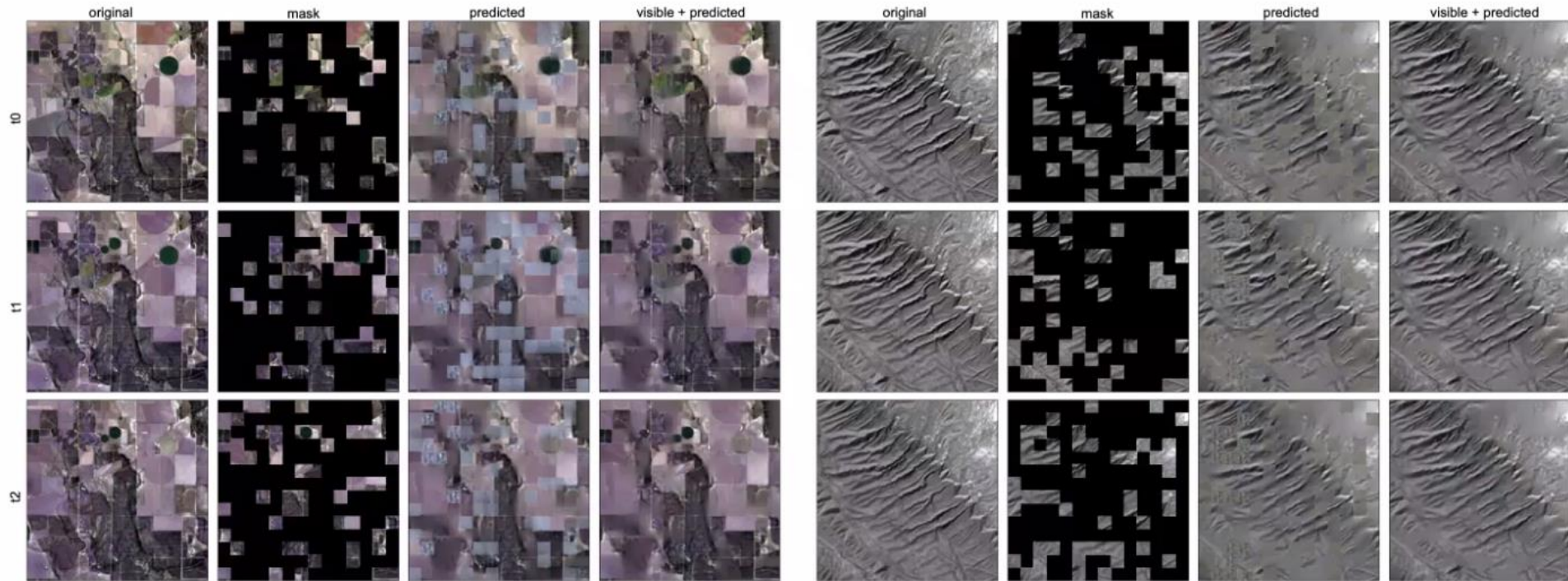


生成画像

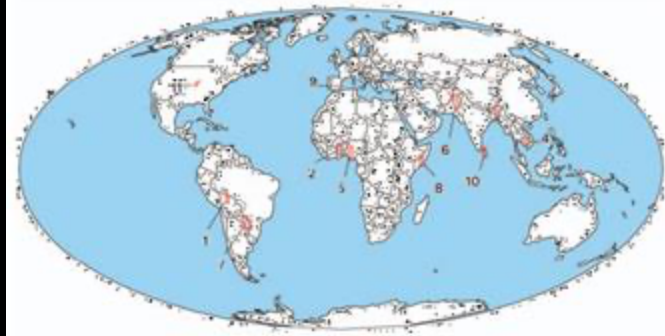


正解画像

基盤モデルの精度（時系列データ）

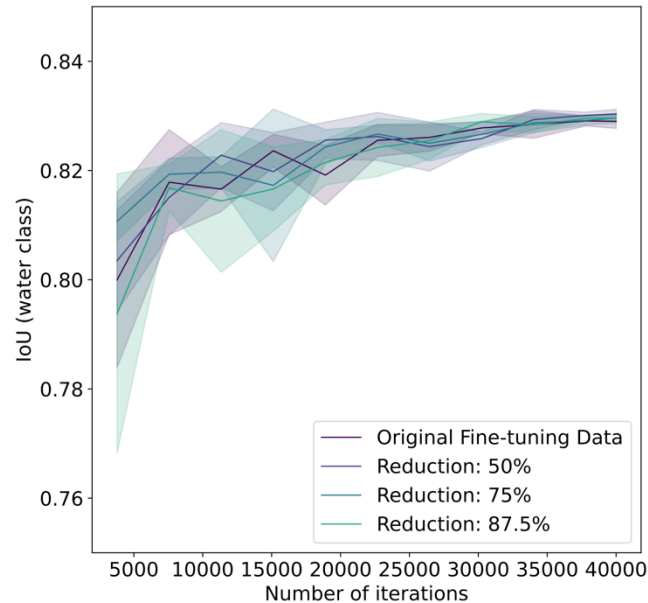


下流タスク：洪水領域可視化

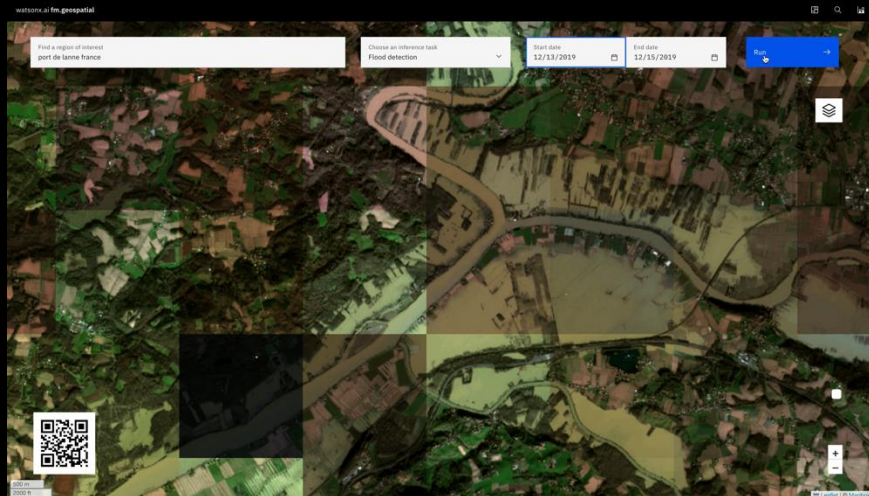


	IoU (water)	F1 (water)	mIoU (both classes)	mF1-score (both classes)	mAcc (both classes)
Baseline [55]	24.21	—	—	—	—
ViT-base [19]	67.58	80.65	81.06	88.92	88.82
Swin [60]	79.43	88.54	87.48	93.13	90.63
Swin [†] [60]	80.58	89.24	87.98	93.44	92.02
AFTER 50 EPOCHS					
Prithvi (not pretrained)	80.67	89.30	88.76	93.85	94.79
→ Prithvi (pretrained)	81.26	89.66	89.10	94.05	95.07
AFTER 500 EPOCHS					
Prithvi (not pretrained)	82.97	90.69	90.14	94.66	94.82
→ Prithvi (pretrained)	82.99	90.71	90.16	94.68	94.60

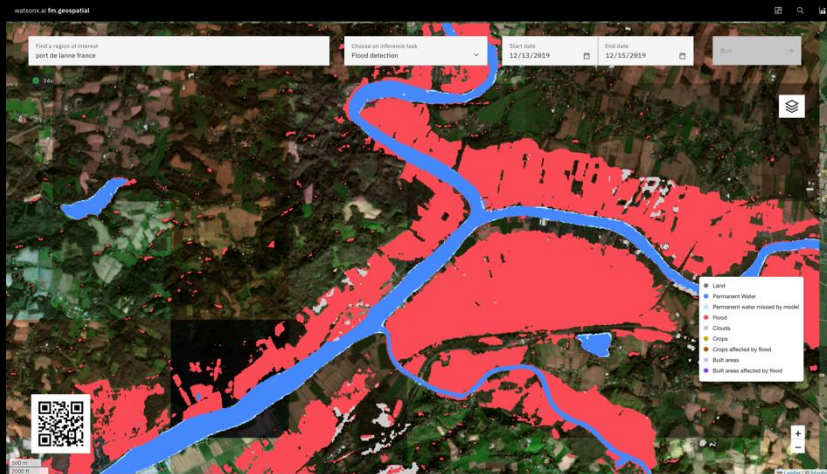
Table 2: Prithvi performance compared to baseline of Sen1Floods11 [55], as well as recent vision transformer architecture baselines used off-the-shelf with standard hyperparameters from [56]. Performance is calculated pixel-wise over the test set, accounting for class imbalance. [†] Swin pretrained on ADE20K.



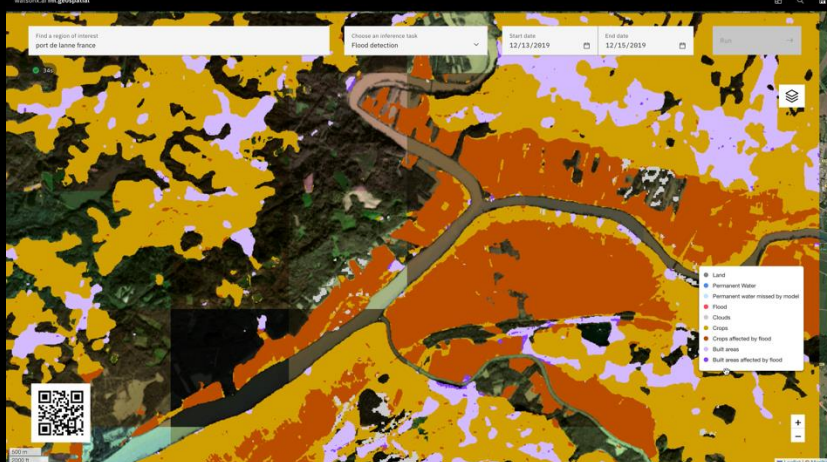
下流タスク：洪水領域可視化



洪水
領域
推定



洪水
被害
推定



Prithvi EO 2.0

- Prithvi 1.0からの進化点

- データ

- サンプル : 25万→420万 (17倍)
 - 領域 : US→世界規模
 - サンプリング : 気候→土地被覆、気候、生態系ブロックを活用
 - 期間 : 2017-2018→2014-2023年のデータ

- 時間

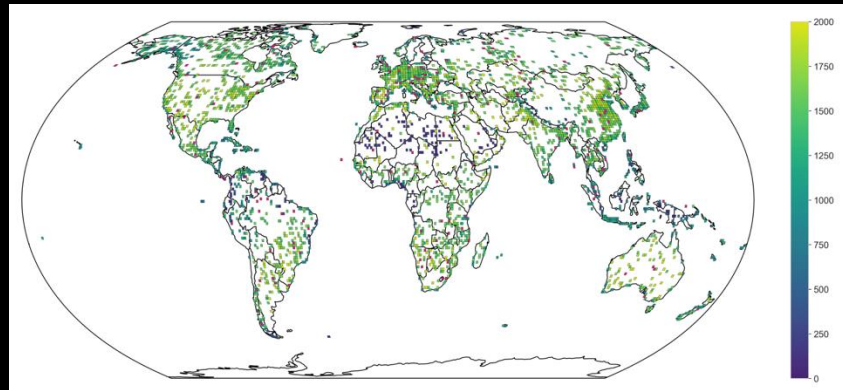
- フレーム数 : 3 フレーム→4 フレーム
 - 間隔 : 連続するフレーム→1 ヶ月から6 ヶ月

- モデルサイズ

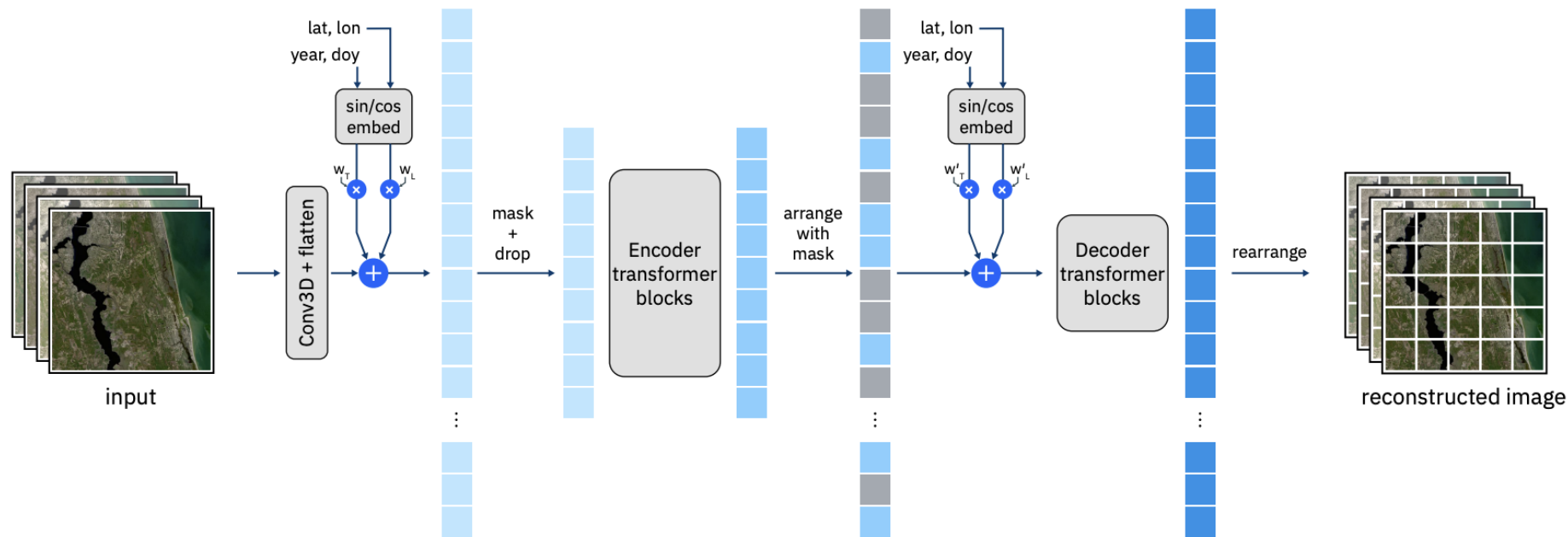
- パラメーター数 : 100M→300M、600M

- アーキテクチャ

- 位置と時間を入力



Prithvi EO 2.0のアーキテクチャ

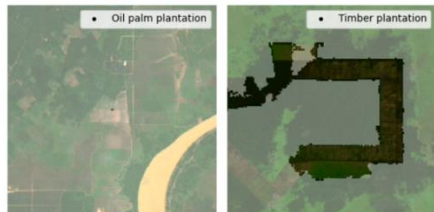


- Year, Day情報、緯度(latitude)と経度(longitude)は絶対的な値
- これらのメタデータもDropoutさせる

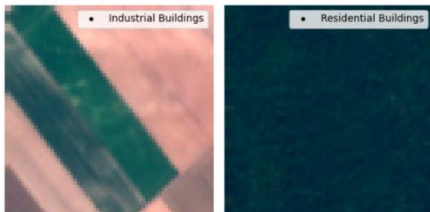
Prithvi EO 2.0の結果

- GEO-Bench datasets [<https://github.com/ServiceNow/geo-bench>]

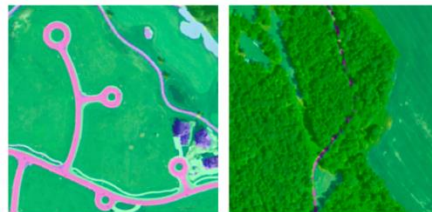
m-forestnet



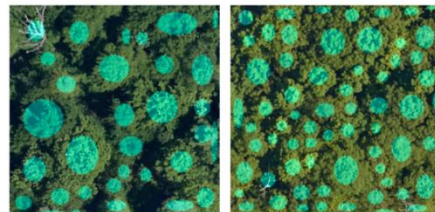
m-eurosat



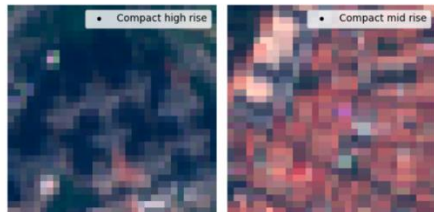
m-chesapeake



m-NeonTree



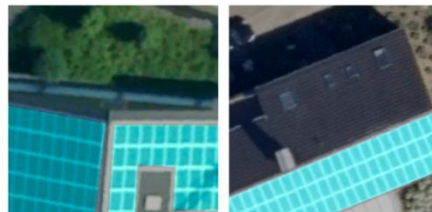
m-so2sat



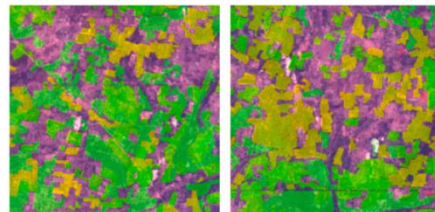
m-pv4ger



m-pv4ger-seg



m-cashew-plantation

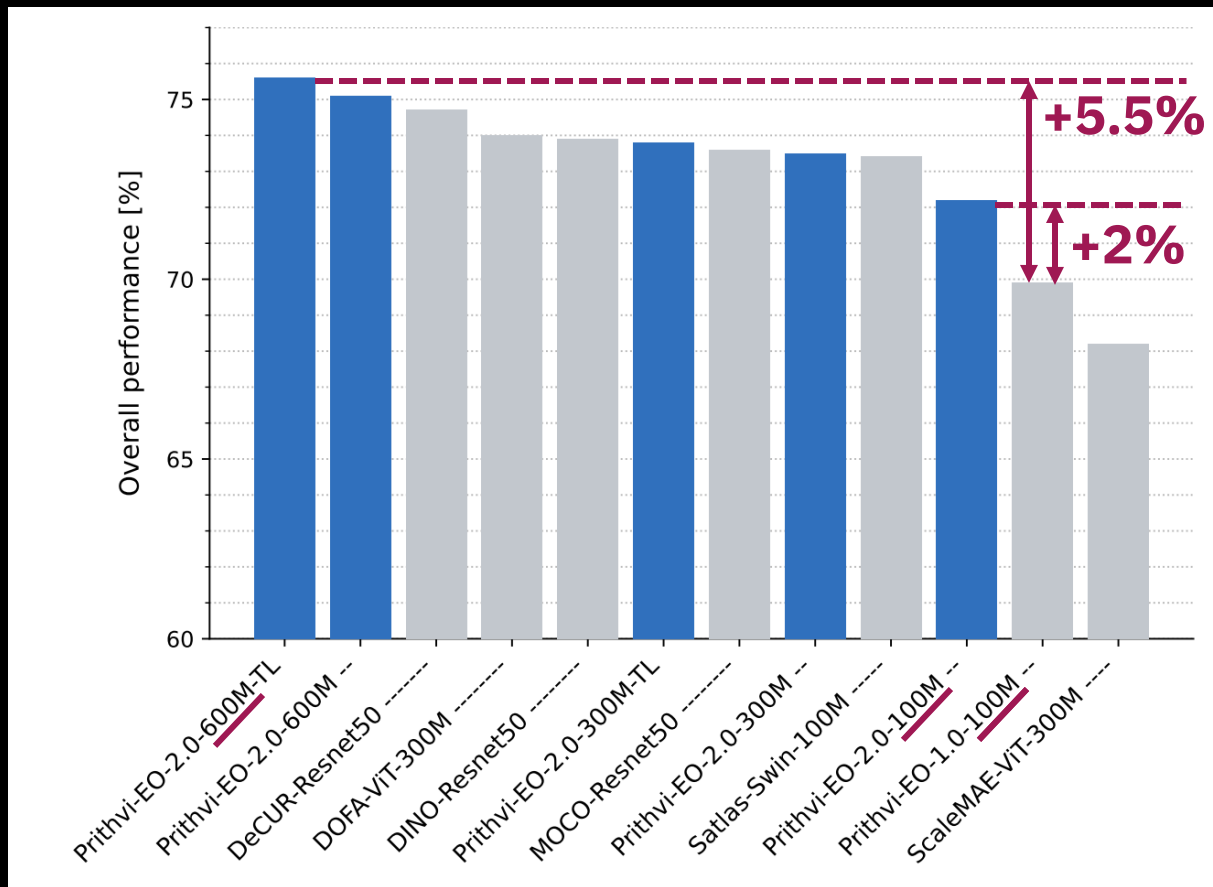


識別タスク

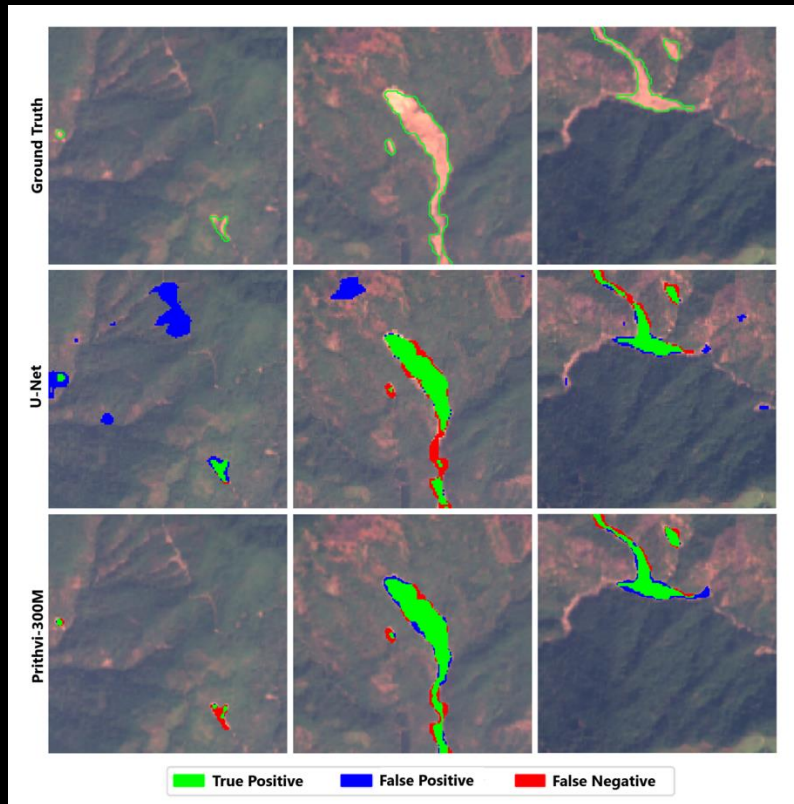
セグメンテーションタスク

※こちらは一例です

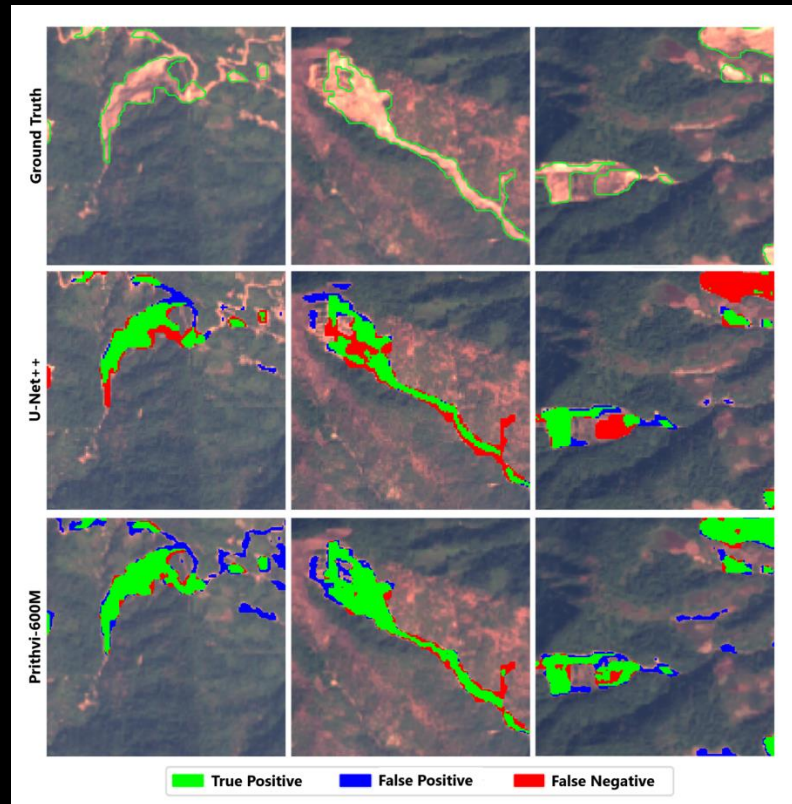
Prithvi EO 2.0の結果



Prithvi EO 2.0の結果（土砂崩れ）



フルデータセット（3,799画像）
[IoU] U-Net: 70.4, Prithvi: **71.3**



少量データセット（100画像, 2.5%）
[IoU] U-Net: 65.1, Prithvi: **67.1**

Prithvi Tinyモデル (2025年10月11日発表)

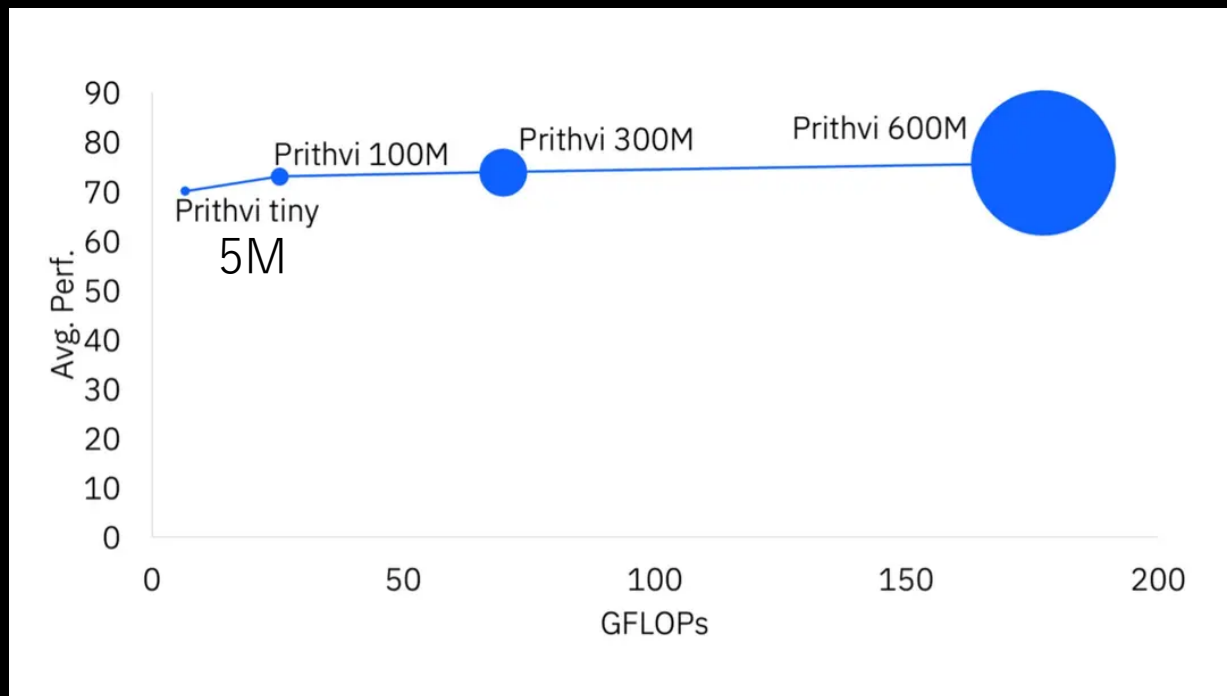
- より軽量モデルも最近発表

- Encoder Memory

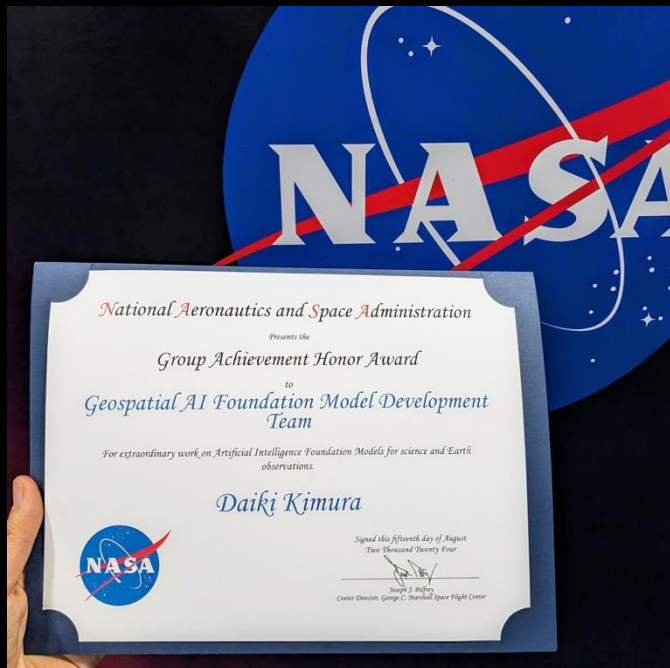
- Tiny: 22Mb
- 100M: 328Mb

- Frames/sec on iX10:

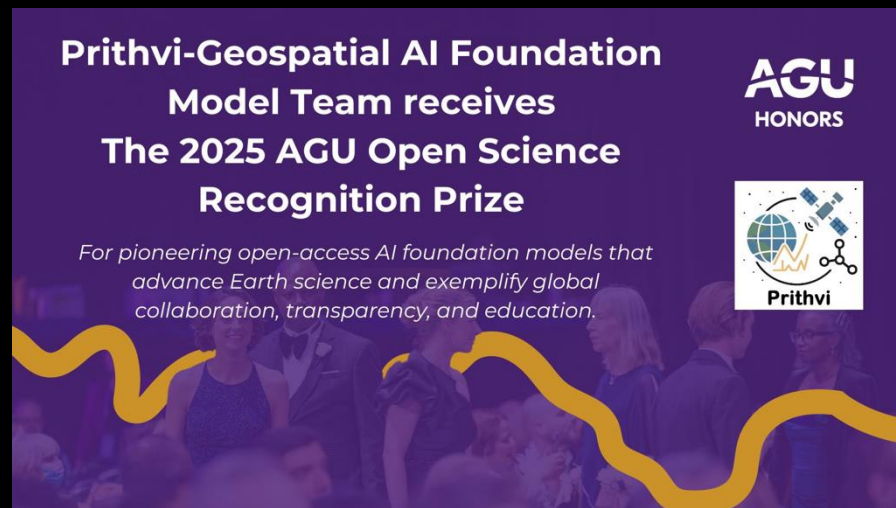
- 329 frams



Prithvi-EOシリーズの外部からの評価



NASAより
Group Achievement Honor Award



AGU（アメリカ地球物理学連合）より
Open Science Recognition Prize

気象・気候の基盤モデル (Prithvi WxC)

NASAが保有する40年間の地球観測データを使用した基盤モデル

プレスリリース (経緯)

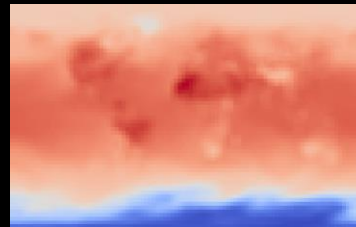
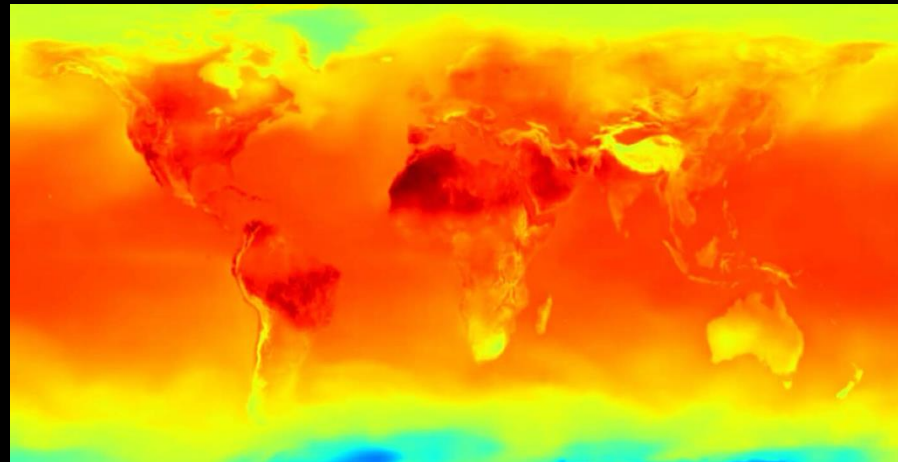
- 2024年9月
IBMとNASA、気象・気候への応用に適したオープンソースのAIモデルをHugging Faceで公開

応用先

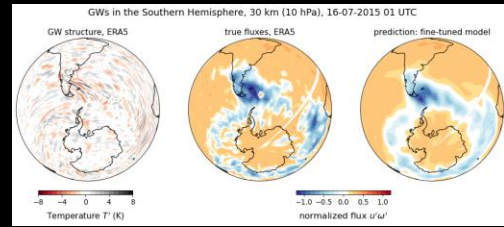
- 局地的な観測に基づいたターゲットを絞った予報の作成
- 異常気象パターンの検出と予測
- 地球規模の気候シミュレーションの空間解像度の向上
- 数値気象・気候モデルでの物理プロセスの表現方法の改善

現在公開済みのファイン・チューニング・モデル

- 気候と気象データのダウンスケーリング：**
低解像度から高解像度を推測する気象学の一般的な実装の一つで、気温、降水量、地上風などで実験
- 重力波パラメタリゼーション：**
このモデルは、重力波の発生をより正確に推定し、数値気象・気候モデルの精度を向上させ、将来の気象・気候現象をシミュレーションする際の不確実性を抑制するのに役立つ



気温に関する
6倍ダウンスケーリング



重力場パラメタリゼーション

Prithvi WxCの学習データ

- 160変数 = $10 \times 14 + 20$
- 学習：1980 – 2019（約40年）
- テスト：2020 – 2023

Table 3: List of Native Vertical Level Variables

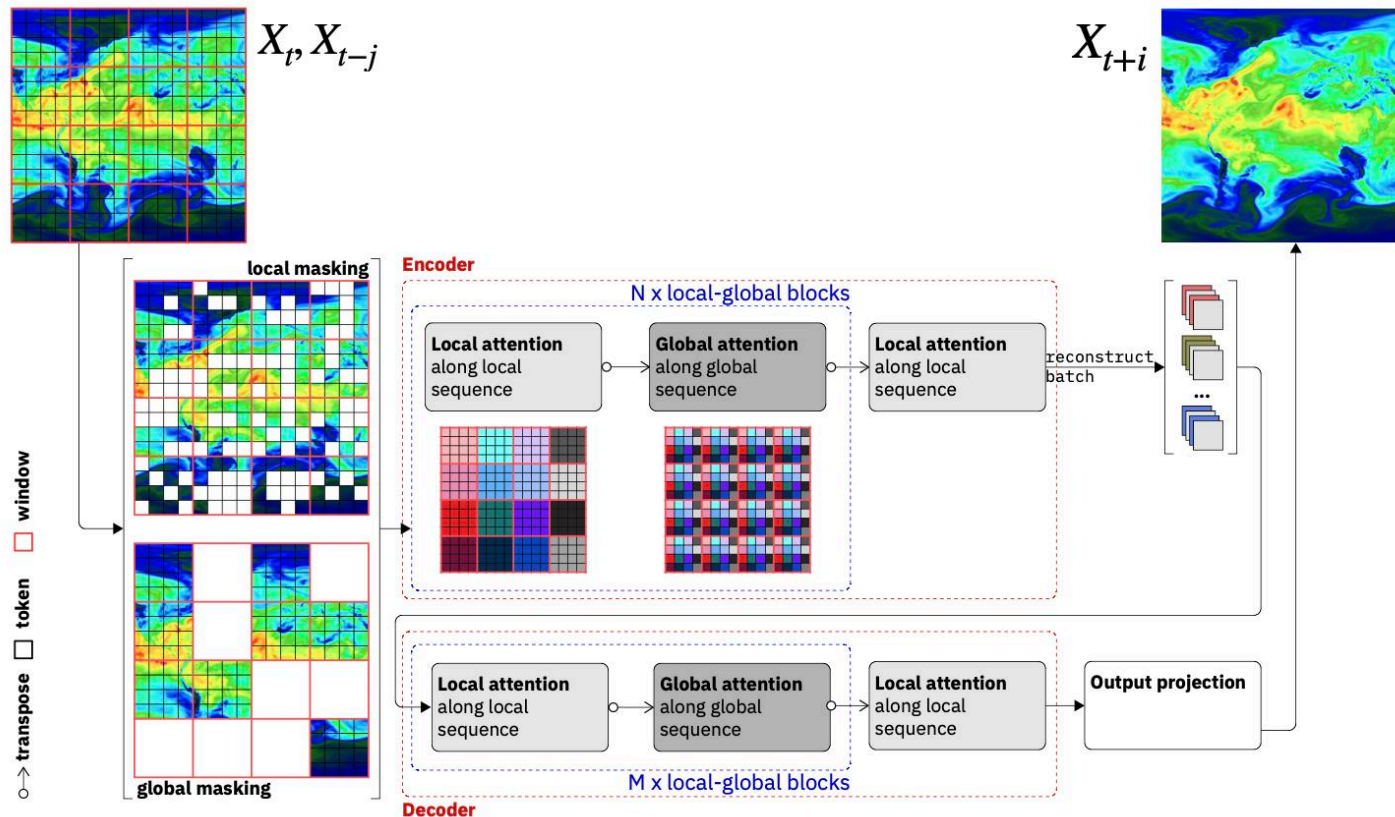
Variable	Collection	Description
U	M2I3NVASM	Wind speed/direction
V	M2I3NVASM	Wind speed/direction
OMEGA	M2I3NVASM	Vertical motions
T	M2I3NVASM	Air temperature
QV	M2I3NVASM	Specific humidity
PL	M2I3NVASM	Actual mid-level pressure
H	M2I3NVASM	Mid-layer height (equivalent to the geopotential height)
CLOUD	M2I3NVASM	Cloud fraction at this layer for radiation
QI	M2I3NVASM	Cloud mass fraction that is ice
QL	M2I3NVASM	Cloud mass fraction that is water
Nominal Pressure (hPa)		
985 970 925 850 700 600 525 412 288 245 208 150 109 48		

Table 2: List of Surface Variables

Variable	Collection	Description
U10	M2I1NXASM	10 m zonal wind
V10	M2I1NXASM	10 m meridional wind
T2M	M2I1NXASM	2 m surface temperature
QV2M	M2I1NXASM	2 m specific humidity
PS	M2I1NXASM	Surface Pressure
SLP	M2I1NXASM	Sea Level Pressure
TS	M2I1NXASM	Skin Temperature
TQI	M2I1NXASM	Column-total ice
TQL	M2I1NXASM	Column-total liquid water
TQV	M2I1NXASM	Column-total water vapor
GWETROOT	M2T1NXLND	Rootzone soil wetness relative to soil holding capacity
LAI	M2T1NXLND	Leaf area index
EFLUX	M2T1NXFLX	Surface latent heat flux
HFLUX	M2T1NXFLX	Surface sensible heat flux
ZOM	M2T1NXFLX	Surface roughness
LWGEM	M2T1NXRAD	Longwave radiation emitted by the surface
LWGAB	M2T1NXRAD	Longwave radiation absorbed by the surface
LWTUP	M2T1NXRAD	Upward longwave at the top of atmosphere
SWGNT	M2T1NXRAD	Net downward shortwave radiation at the surface
SWTNT	M2T1NXRAD	Net shortwave at top of atmosphere

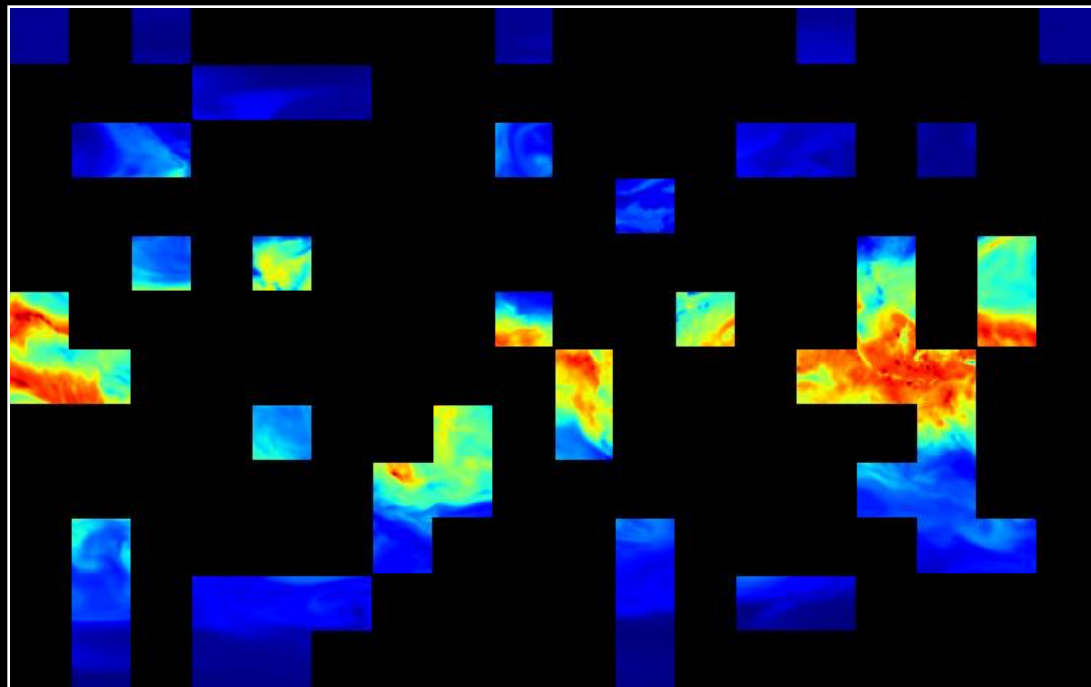
Prithvi WxCのモデル

- Encoder
 - 13 Local
 - 12 Global
- Decoder
 - 3 Local
 - 2 Global
- 6時間後の予測等を実施
- モデル: 2.3B



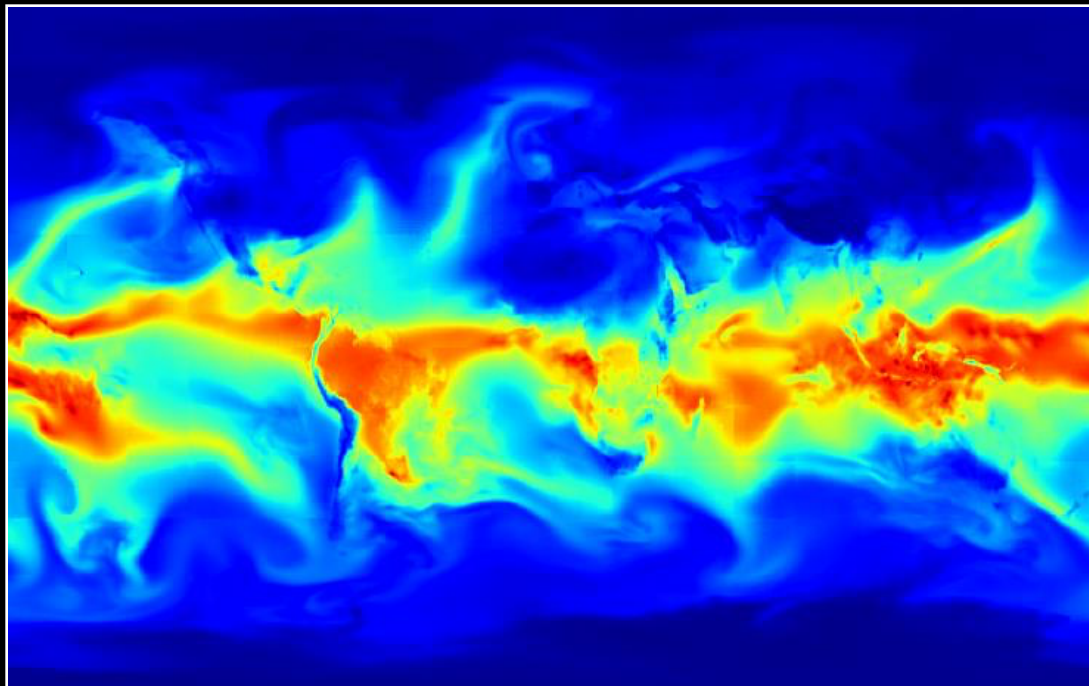
Prithvi WxCの事前学習の結果

- 水蒸気量 (75%マスク)



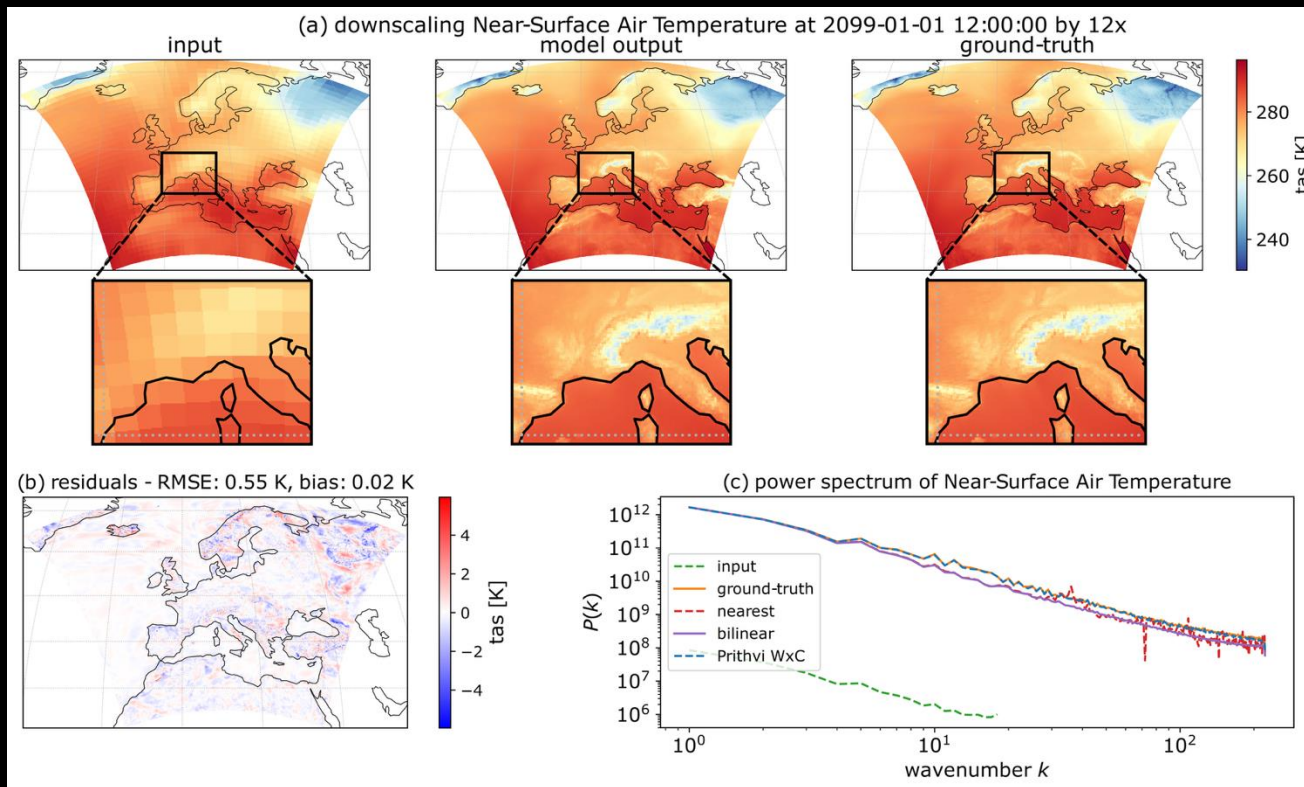
Prithvi WxCの事前学習の結果

- 水蒸気量（0時間後の推定結果）

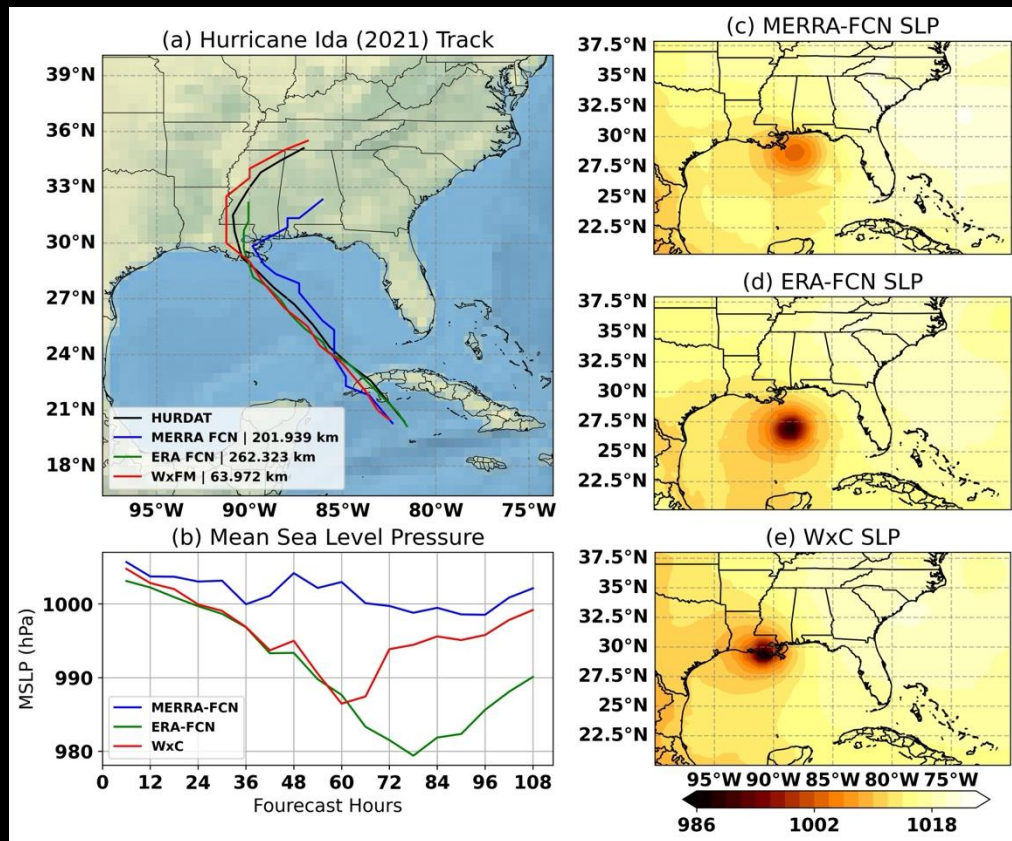


Prithvi WxCの超解像タスク

- 気温の12倍超解像

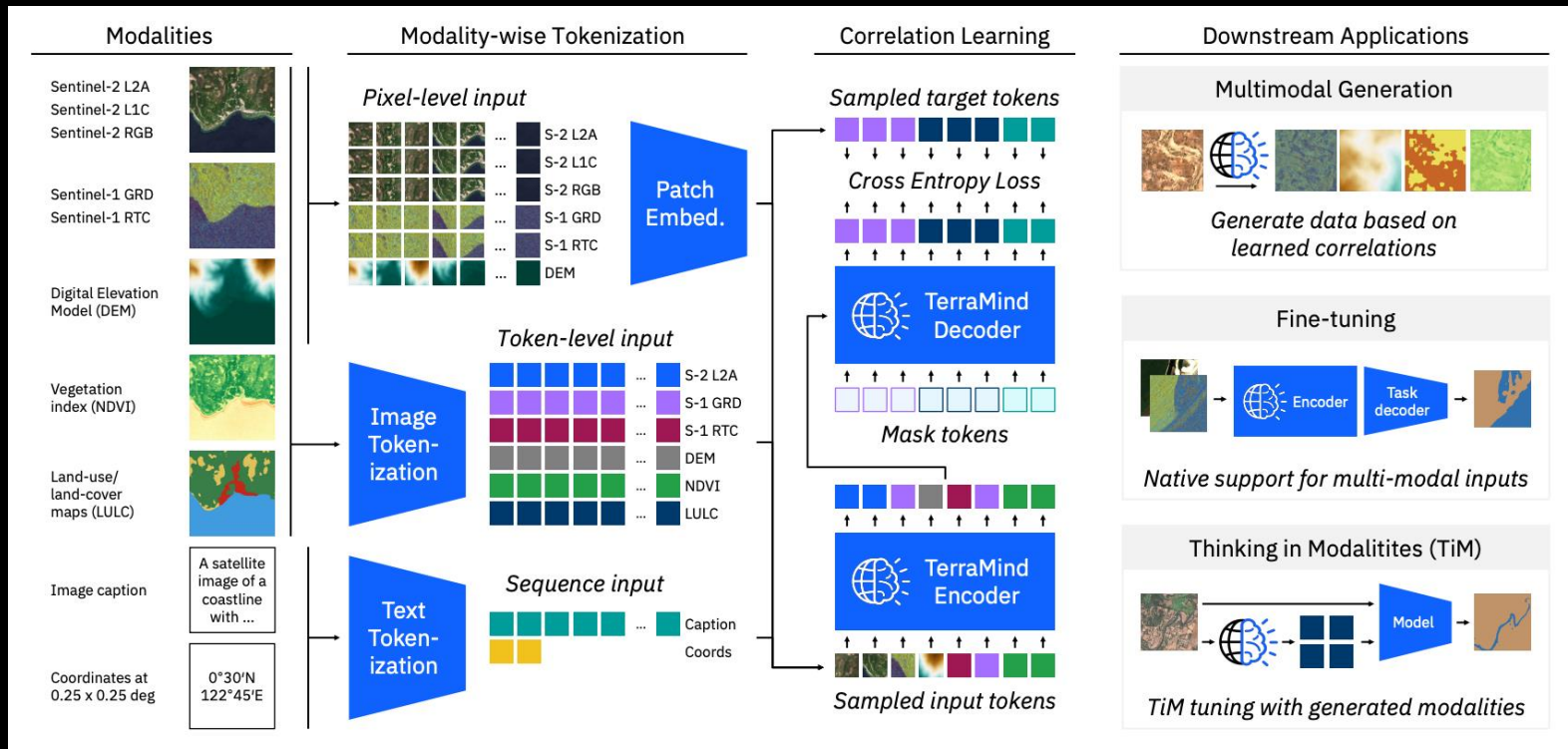


Prithvi WxCのハリケーン予測タスク



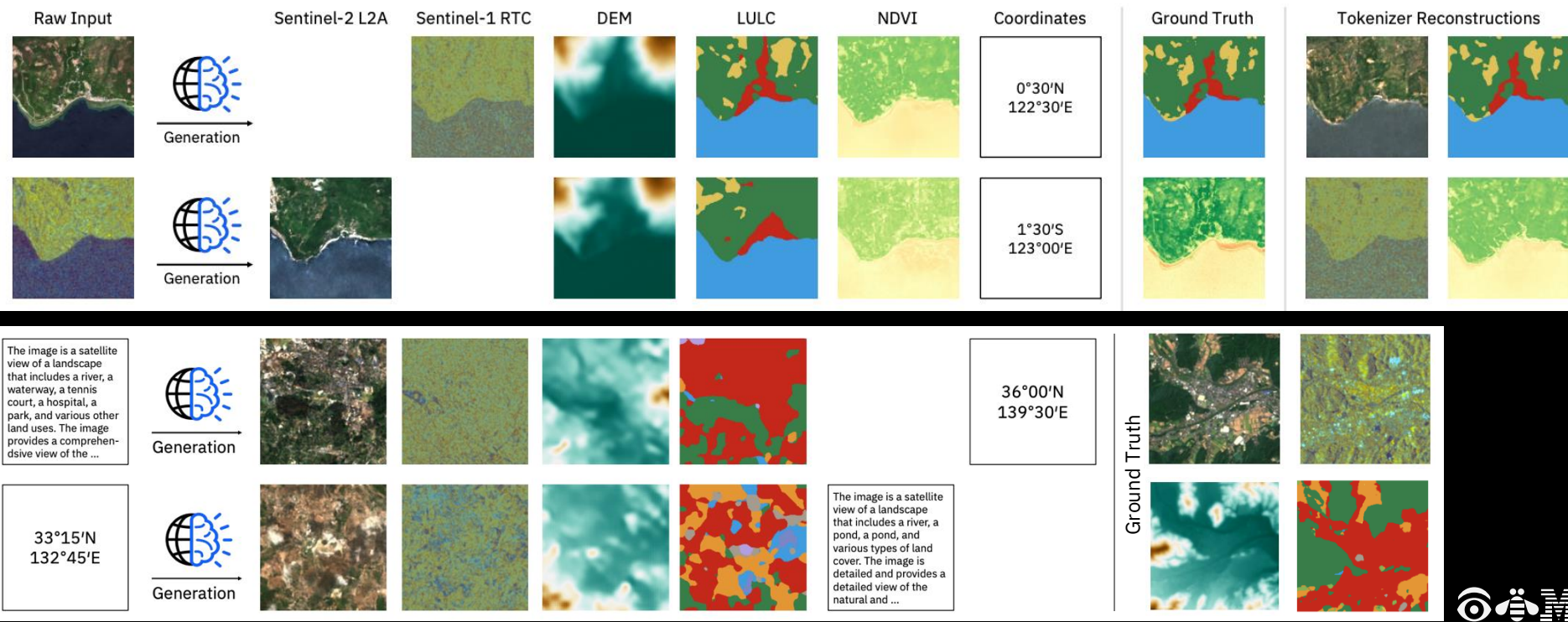
TerraMind: マルチモーダル生成

- 世界初のAny-to-Any生成が可能なモデル



TerraMind: Any-to-Any生成

- 一つの入力から、複数のモダリティの情報を生成する
 - 例えば、光学衛星画像からSAR画像、標高、土地利用などを生成



TerraMind: Any-to-Any生成の精度

- 画像生成に対する精度
 - 光学画像は、値自身の精度も構造も生成できている
 - 標高は、値の絶対値は誤りが多いが、構造（相関）は生成できている
 - SAR画像は、構造は苦手だが、値自身は高精度に生成できている

Modalities	MAE ↓	RMSE ↓	SSIM ↑	PSNR ↑
S-1 GRD → S-2 L2A	0.074	<u>0.116</u>	<u>0.750</u>	26.210
S-1 GRD → DEM	163.0	320.8	<u>0.878</u>	20.694
S-1 GRD → NDVI	0.180	0.225	0.438	18.990
S-1 RTC → S-2 L2A	0.113	0.194	0.695	24.251
S-1 RTC → DEM	298.8	<u>799.2</u>	<u>0.873</u>	20.009
S-1 RTC → NDVI	0.172	0.211	<u>0.465</u>	19.529
S-2 L2A → S-1 GRD	2.942	3.877	<u>0.531</u>	<u>28.678</u>
S-2 L2A → S-1 RTC	2.636	3.391	0.430	<u>28.993</u>
S-2 L2A → DEM	215.8	745.5	0.942	20.616

TerraMind: ダウンストリームタスクでの評価

- PANGAEAベンチマークでの評価

Model	BurnSr*	MADOS*	PASTIS	Sen1F11	FBP*	DEN*	CTM-SS	SN7*	AI4Farms*	Avg. mIoU	Avg. Rank
CROMA	82.42	67.55	32.32	<u>90.89</u>	51.83	38.29	49.38	59.28	25.65	55.29	6.61
DOFA	80.63	59.58	30.02	89.37	43.18	<u>39.29</u>	51.33	61.84	27.07	53.59	8.22
GFM-Swin	76.90	64.71	21.24	72.60	67.18	34.09	46.98	60.89	27.19	52.42	10.00
Prithvi 1.0 100M	83.62	49.98	33.93	90.37	46.81	27.86	43.07	56.54	26.86	51.00	11.00
RemoteCLIP	76.59	60.00	18.23	74.26	69.19	31.78	52.05	57.76	25.12	51.66	11.22
SatlasNet	79.96	55.86	17.51	90.30	50.97	36.31	46.97	61.88	25.13	51.65	10.67
Scale-MAE	76.68	57.32	24.55	74.13	<u>67.19</u>	35.11	25.42	62.96	21.47	49.43	11.44
SpectralGPT	80.47	57.99	35.44	89.07	33.42	37.85	46.95	58.86	26.75	51.87	10.11
S.-S12-MoCo	81.58	51.76	34.49	89.26	53.02	35.44	48.58	57.64	25.38	53.02	10.06
S.-S12-DINO	81.72	49.37	36.18	88.61	51.15	34.81	48.66	56.47	25.62	52.51	10.89
S.-S12-MAE	81.91	49.90	32.03	87.79	51.92	34.08	45.80	57.13	24.69	51.69	12.39
S.-S12-Data2Vec	81.91	44.36	34.32	88.15	48.82	35.90	54.03	58.23	24.23	52.22	10.72
UNet Baseline	84.51	54.79	31.60	91.42	60.47	39.46	47.57	<u>62.09</u>	46.34	57.58	4.89
ViT Baseline	81.58	48.19	38.53	87.66	59.32	36.83	44.08	52.57	<u>38.37</u>	54.13	10.28
TerraMindv1-B	82.42	<u>69.52</u>	40.51	90.62	59.72	37.87	55.80	60.61	28.12	<u>58.35</u>	<u>3.94</u>
TerraMindv1-L	82.93	75.57	43.13	90.78	63.38	37.89	<u>55.04</u>	59.98	27.47	59.57	3.44

TerraMindv1-B: 500B tokens for 6 days on 32 NVIDIA A100 GPUs

TerraMindv1-L: 500B tokens for 10 days on 32 NVIDIA A100GPUs

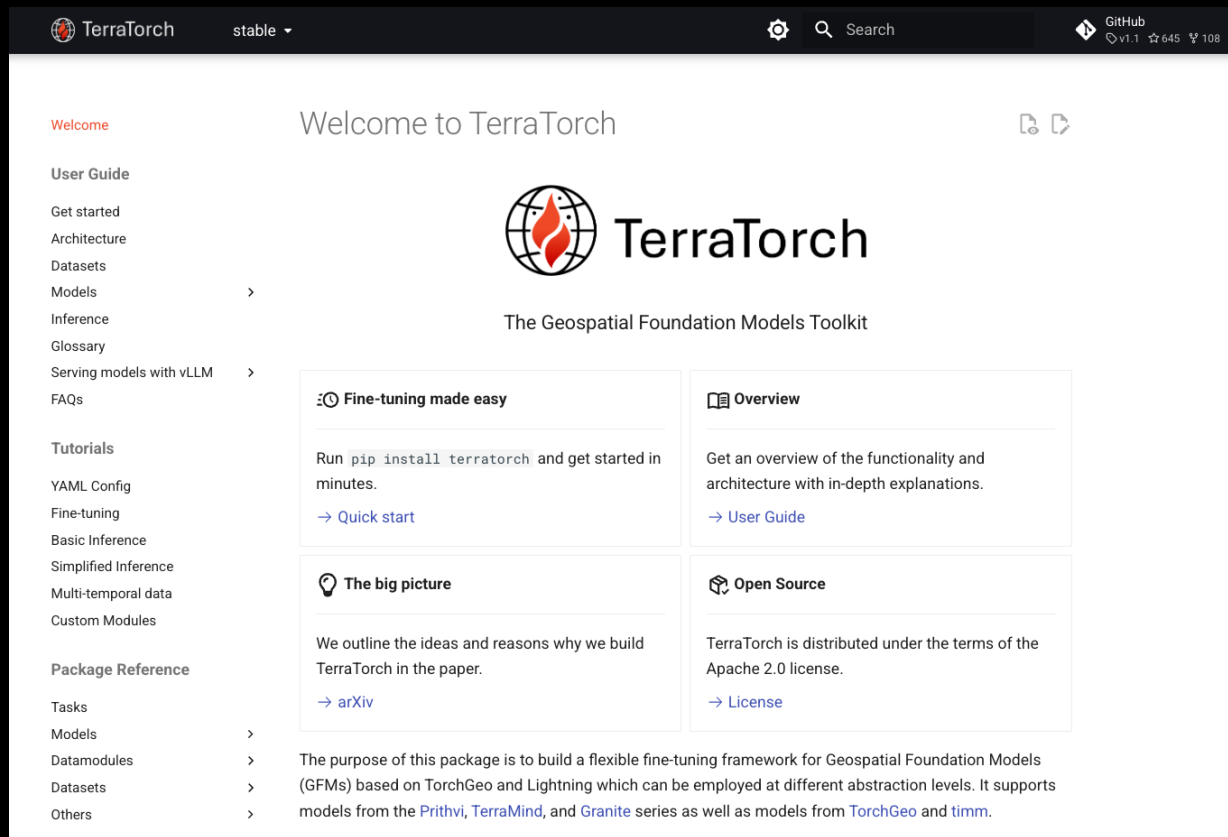


TerraMind: Thinking in Modalities

- LLMにおけるchain-of-thoughtの考え方を応用した方式
- 洪水領域推定を行う際に、土地被覆マップを生成し、それを入力することで、洪水領域推定の精度向上が行われるかを実験

Fine-Tuning	Input	$\text{IoU}_{\text{Water}}$	mIoU
TerraMindv1-B	S-1	68.00	81.06
TerraMindv1-B	S-2	82.26	89.70
TerraMindv1-B TiM	S-1 + <i>gen. LULC</i>	72.25	83.65
TerraMindv1-B TiM	S-2 + <i>gen. LULC</i>	84.75	91.14

地理空間基盤モデルを活用するプラットフォーム



The screenshot shows the TerraTorch website interface. At the top, there's a navigation bar with the TerraTorch logo, a version dropdown set to 'stable', a search icon, and a GitHub link showing 645 stars and 108 forks. The main content area has a 'Welcome to TerraTorch' message with a globe icon. Below this, the text 'The Geospatial Foundation Models Toolkit' is displayed. A left sidebar contains a 'User Guide' section with links to 'Get started', 'Architecture', 'Datasets', 'Models', 'Inference', 'Glossary', 'Serving models with vLLM', and 'FAQs'. Below that is a 'Tutorials' section with links to 'YAML Config', 'Fine-tuning', 'Basic Inference', 'Simplified Inference', 'Multi-temporal data', and 'Custom Modules'. At the bottom of the sidebar is a 'Package Reference' section with links to 'Tasks', 'Models', 'Datamodules', 'Datasets', and 'Others'. The main content area features four cards: 'Fine-tuning made easy' with a 'Quick start' link, 'Overview' with a 'User Guide' link, 'The big picture' with an 'arXiv' link, and 'Open Source' with a 'License' link. The bottom card also contains a paragraph about the package's purpose and supported models.

Welcome to TerraTorch

The Geospatial Foundation Models Toolkit

Fine-tuning made easy

Run `pip install terratorch` and get started in minutes.

[→ Quick start](#)

Overview

Get an overview of the functionality and architecture with in-depth explanations.

[→ User Guide](#)

The big picture

We outline the ideas and reasons why we build TerraTorch in the paper.

[→ arXiv](#)

Open Source

TerraTorch is distributed under the terms of the Apache 2.0 license.

[→ License](#)

The purpose of this package is to build a flexible fine-tuning framework for Geospatial Foundation Models (GFM) based on TorchGeo and Lightning which can be employed at different abstraction levels. It supports models from the [Prithvi](#), [TerraMind](#), and [Granite](#) series as well as models from [TorchGeo](#) and [timm](#).

<https://ibm.github.io/terratorch/stable/>



まとめ：Prithviシリーズ

- 光学衛星を活用した基盤モデル
 - Prithvi EO 1.0: [arXiv paper](#), [Hugging Face page](#)
 - Prithvi EO 2.0: [arXiv paper](#), [Hugging Face page](#)
 - Prithvi EO 2.0 Tiny: [Hugging Face page](#)
- 気象・気候の基盤モデル
 - Prithvi WxC: [arXiv paper](#), [Hugging Face page](#)
- マルチモーダル生成基盤モデル
 - TerraMind: [arXiv paper](#), [Hugging Face page](#)
- 地理空間基盤モデルのファインチューニング環境
 - TerraTorch: [User guide](#)

まとめ

- IBM Research（IBM基礎研究所）の簡単な紹介
- 基盤モデルの利点（少数データで高精度モデル）の紹介
- NASAとESA、IBMの衛星画像基盤モデル Prithviシリーズ
 - 概要
 - Prithvi 1.0: 基盤モデルとしての学習と、下流タスク、デモ
 - Prithvi 2.0: 1.0との違い
 - Prithvi WxC: 気象・気候
 - TerraMind: マルチモーダル生成
 - TerraTorch: プラットフォーム
- デジタルツインと、Prithviの将来