

cvpaper.challenge

Remote sensing × Computer vision meta survey

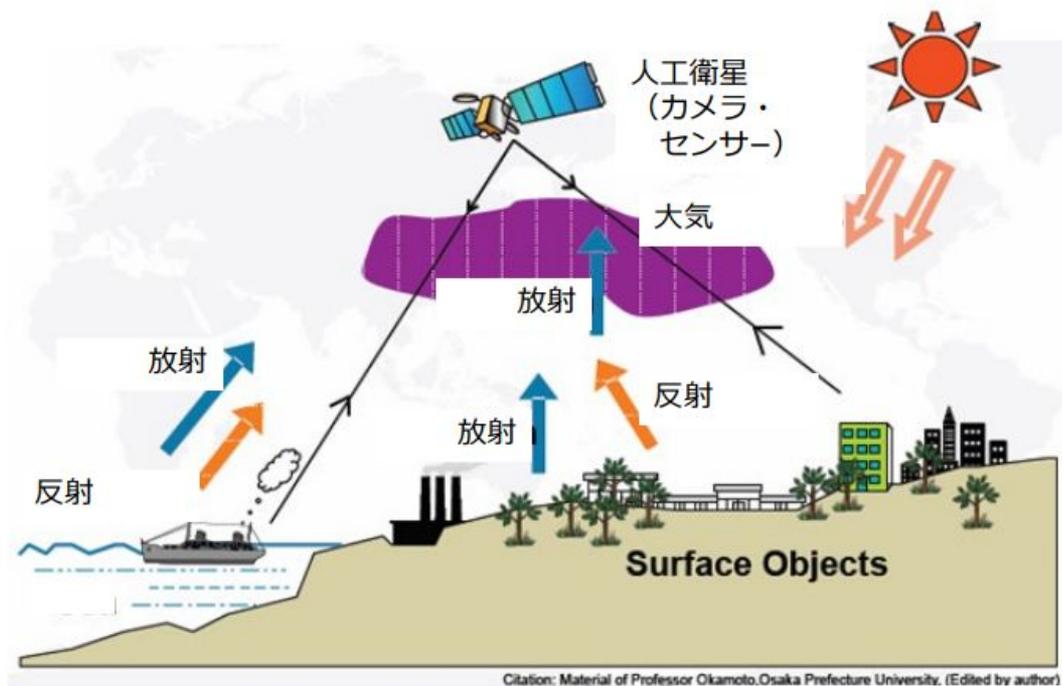
中村 凌、篠原 崇之、修 浩毅、二村 忠宏
天野 凌我、神山 徹、中田 和真、中原 龍一

<http://xpaperchallenge.org/cv>

- 衛星リモートセンシングについて
- センサーから取得されるデータについて
- リモートセンシング分野におけるタスク
- 衛星データを使った機械学習用データセット(タスク)論文紹介
- サーベイを通しての気づき
- CV分野では多くのモデルの研究が行われている
- なぜ日本のリモセン界隈ではCVで開発された最先端モデルが利用されていない？
- 日本のリモセン分野はどう発展していくと良い？

離れた場所から観測する技術

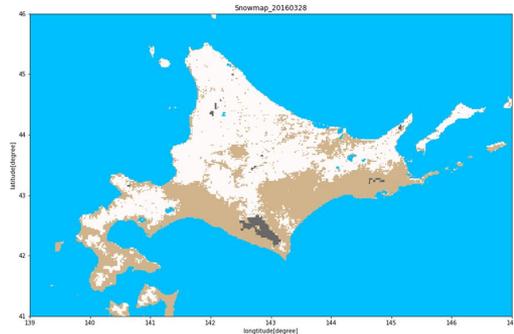
- **衛星リモートセンシング**とは、人工衛星に搭載された「センサー」を用いて地球の太陽光/電波の反射・放射等を計測すること



Citation: Material of Professor Okamoto, Osaka Prefecture University. (Edited by author)

衛星リモートセンシングによる観測は様々な分野目的達成のために活用されている

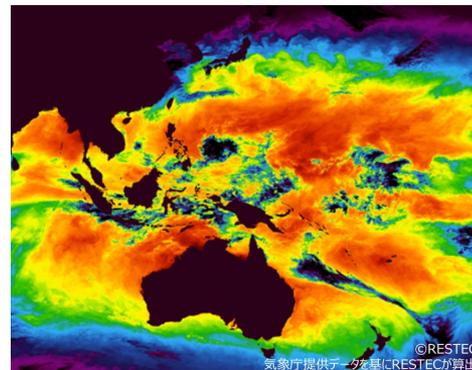
気象(気候変動を把握)



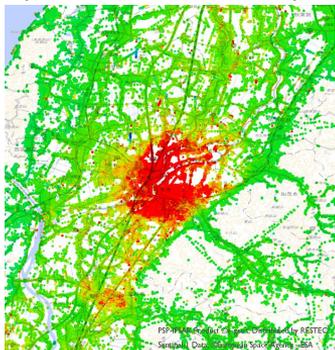
防災(被災域を把握)



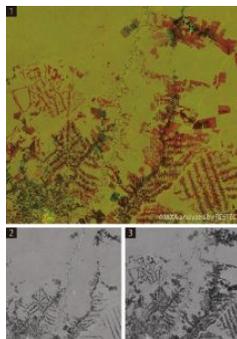
漁業(魚のいそうな海域を把握)



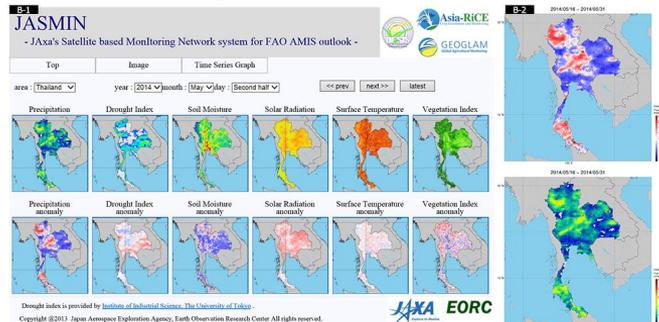
地盤沈下
(インフラ被害を把握)



森林監視
(CO2吸収量を把握)



農業(農作物の収量を把握)



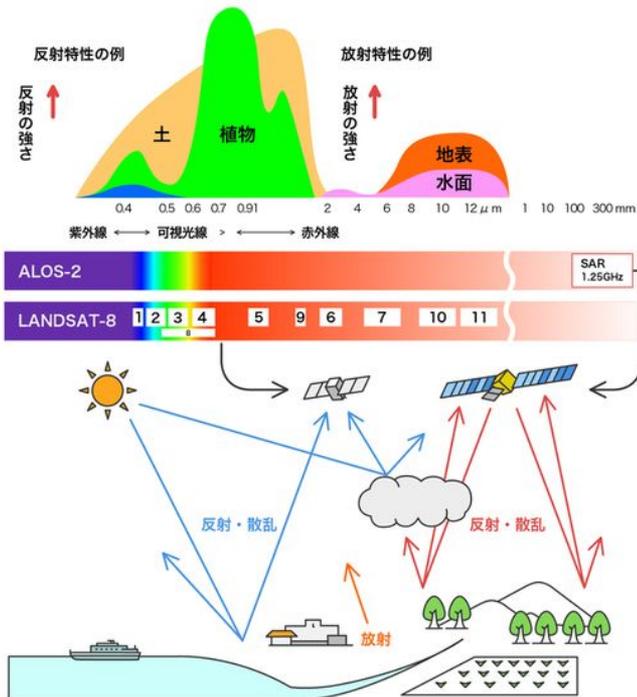
衛星リモートセンシングに用いられるセンサにはいくつかの種類がある

● 光学センサ

- 太陽光の反射や放射を測る
- 主な観測対象
 - 植物の有無
 - 地表の温度
 - 海面の温度
 - 地表の高さ
 - 雲の状態
 - 水の有無

● マイクロ波センサ

- センサからマイクロ波を放射し、対象から反射されたマイクロ波を測る
- **合成開口レーダ(SAR)** など
- 主な観測対象
 - 植物の有無
 - 地表の高さ・変化
 - 雲の状態
 - 水の有無



マイクロ波センサ



光学センサ



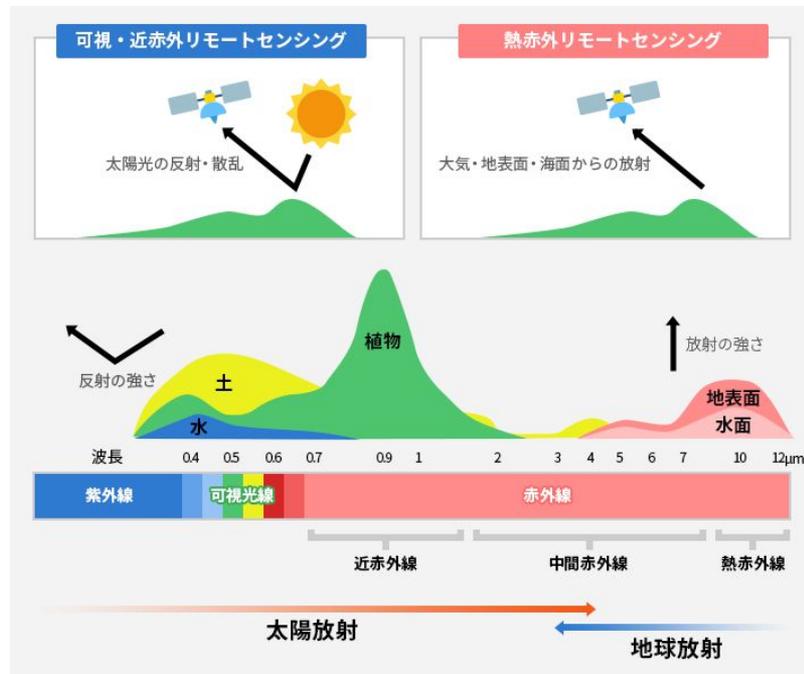
光学衛星センサは可視から近赤外である太陽光の反射・熱赤外では地球放射を観測

● 主な衛星

- Landsat-8,9(米,USGS/NASA)
- Sentinel-2, 3(欧州,ESA)
- SPOT-6,7(仏,AIRBUS Defence & Space)
- WorldView-3(米,Maxar Technologies Inc.)
- HISUI (国際宇宙ステーションに設置)(日本, 経産省)
- Dove(米, Planet labs)
- ひまわり(日本, 気象庁)

● 特性: 反射特性が地表面に依存する

- 可視光・赤外線は地表面の状態、波長依存の吸収の有無によって反射される電磁波の強度が変化する
 - この特性を利用することで地表面の状態を衛星データから推定できる
 - 逆に言うと反射特性が異なる場合、可視光だけでは判別が難しい地物も判別できる
- 可視域では夜間光、熱赤外では熱放射など地球からの放射も利用される



観測されたデータは一般ユーザーが使用できるように公開が進んでいる

ESA系の衛星データのオープンフリー化
Sentinel1(SAR), Sentinel2(光学)

Landsat program



[LANDSAT MISSIONS](#) | [DATA](#)

Landsat Data Access

By [Landsat Missions](#)

HOME

LANDSAT MISSIONS

SCIENCE

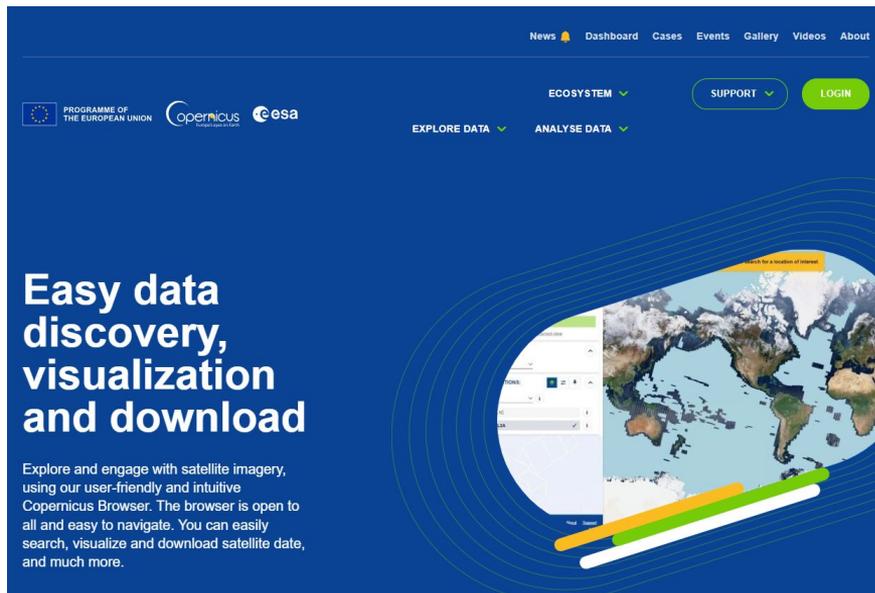
PRODUCT
INFORMATION

DATA

Since 2008, Landsat Level-1 data, as well as Level-2 and Level-3 science products held in the USGS archive, have been available for download at no charge from a variety of data portals. This page provides information about searching and downloading Landsat data and science products.

2008 USGS Technical Announcement

USGS <https://www.usgs.gov/landsat-missions/landsat-data-access> より引用



ESA <https://dataspace.copernicus.eu/> より引用

物理量の計測から計測したデータを用いた土地利用分類まで様々なタスクが存在

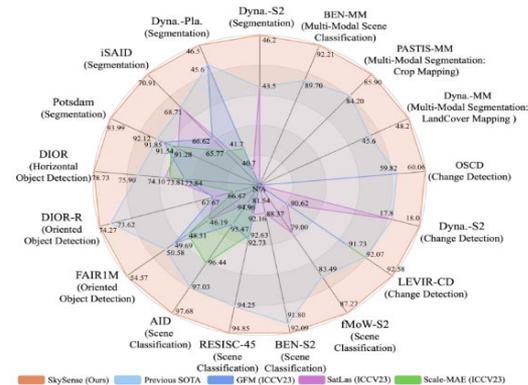
- **物理量計測 (非ML手法)**
 - 計測対象の例
 - エアロゾル, 温室効果ガス, 海面水温, 積雪, 海氷, 農業, 森林, 砂漠化
 - 物理量は観測結果から一意に決まる
- **地球科学**
 - 地形判読、火山モニタリング、海洋モニタリング、気象観測
- **都市管理**
 - 土地利用分類、土地被覆分類
- **付加価値創出 (CV技術が取り入れられつつある)**
 - 地物の判読
 - 高解像度衛星画像から建物・道路抽出など
 - 超解像
 - 低解像度の衛星画像を擬似的に高解像度に
 - 疑似着色
 - 古い白黒画像や、SARの強度画像をカラー画像に
 - 単画像高さ推定
 - ステレオじゃない衛星画像から建物高さなどを推定



目的タスクの精度向上のためにML向け多様なデータセットが公開されてきている

- CNNの性能が世の中に広まって以降、リモートセンシング分野でも深層学習技術が取り入れられている
 - 例) 衛星画像からの地物(建物や道路など)の自動抽出
- 深層学習技術の普及に伴い、リモートセンシングデータのデータセットが公開されている
- 有名なベンチマークデータセット(以下の図を参照)

	Dataset	Year	#Image	#Class	#Box	Avg. Res.	
retrieval	RET-3	RSICD [25]	2017	8483	-	-	224×224
		RSITMD [24]	2021	3603	-	-	256×256
		UCMerced [26]	2018	1676	-	-	256×256
detection	DET-10	AUAIR [86]	2020	32,823	8	132,031	1920×1080
		CARPK [87]	2017	1,568	1	106,690	1280×720
		DIOR [77]	2019	23,463	20	192,472	800×800
		DOTA [76]	2017	1,409	15	98,990	1504×1395
		HRRSD [78]	2019	21,761	13	57,137	1406×1264
		HRSC [81]	2017	1,055	1	1,055	1105×791
		LEVIR [80]	2020	37.91	3	11,028	800×600
segmentation	SEG-4	RSOD [79]	2021	936	4	7,400	1051×900
		Stanford [88]	2016	17,351	6	355,443	1424×1088
		Visdrone [89]	2018	6,471	11	77,547	1509×849
		iSAID [84]	2019	30,821	15	987,239	896×896
		loveDA [85]	2021	4,187	6	97,989	1024×1024
Potsdam [83]	2012	5,421	4	92,161	512×512		
Vaihingen [82]	2012	742	4	16,875	512×512		



Liu et al., “RemoteCLIP: A Vision Language Foundation Model for Remote Sensing”, in IEEE TGRS.

Guo et al., “SkySense: A Multi-Modal Remote Sensing Foundation Model Towards Universal Interpretation for Earth Observation Imagery”, in CVPR 2024

以降のスライドでリモセン特有のタスクをいくつか紹介

衛星から取得した画像データをもとに山火事の領域を推定する

- CalFire(2022): Sentinel-2をもとに作成。イベントを一つ一つアノテーションするコストをかけず教師なし変化検知で解決。 詳細 p56
- FireRisk(2023): NAIP (National Agriculture Imagery Program)をもとに作成。事前学習済みモデルによる予測性能評価を行った。 詳細 p69
- Sen2Fire(2024): Sentinel-2, Sentinel-5をもとに作成。データセットの大規模化などを実現するために、マルチスペクトルデータおよびスペクトル指標と野火時に発生するエアゾルを含んだSen2Fireを提案した。 詳細 p72
- 表現学習: SimCLR, 3DConv-AE, Tile2Vec, DINO, MAE
- モデル: ResNet, ViT

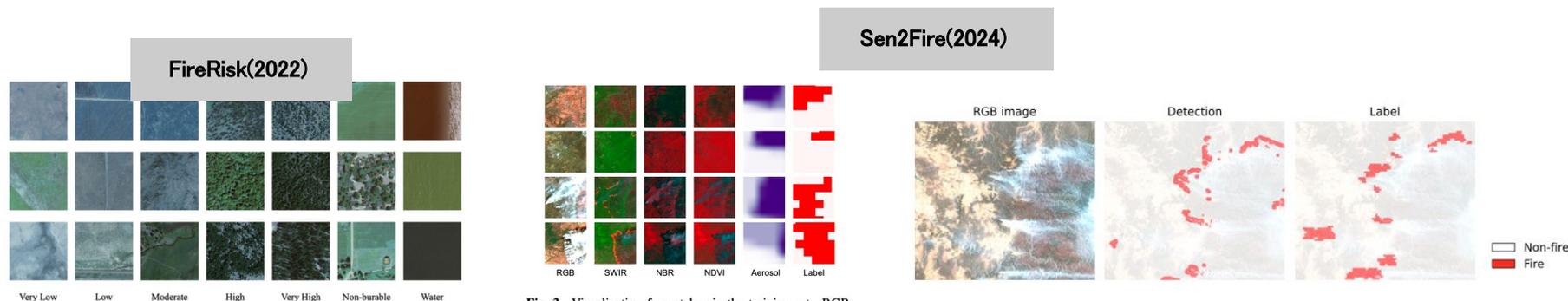


Fig. 3. Visualization for patches in the training set. RGB, SWIR, NBR, and NDVI denote different color composites.

船舶をモニタリングするために船舶に関する特徴を様々なデータを活用

- HRSID (2020): SAR、Object detection (夜間でも撮影可能、雲の影響を受けないのが特徴)
- SSDD (2021): SAR、Object detection、instance segmentation
- TISD (2022): 熱赤外画像、Object detection (薄雲や霧、夜間でも撮影可能)
- SynthWakeSAR (2022): SAR、Classification (船舶が通った後の波を合成データで作成)
- 使用されるモデル: AlexNet、R-CNN、SSD

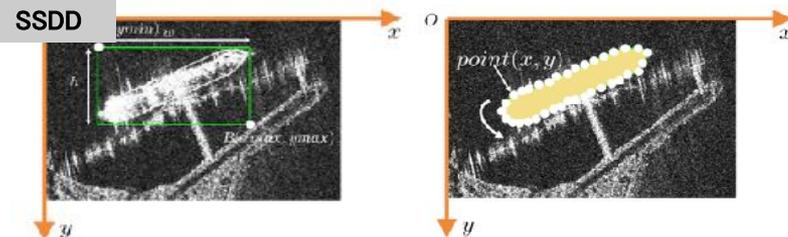
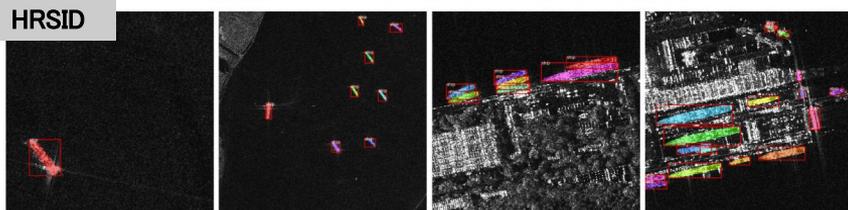
詳細 p50

詳細 p55

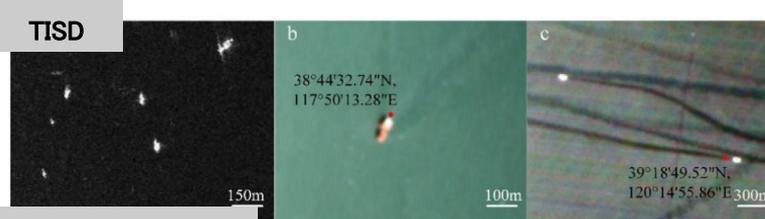
詳細 p59

詳細 p60

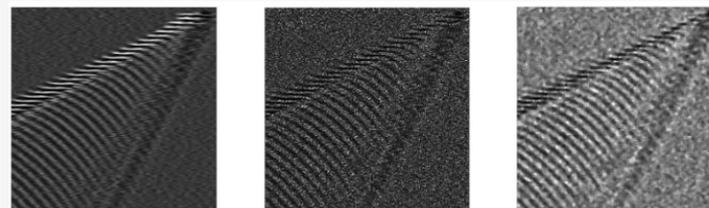
SARを用いて船舶の形状検出



熱や波を用いて船舶を検出



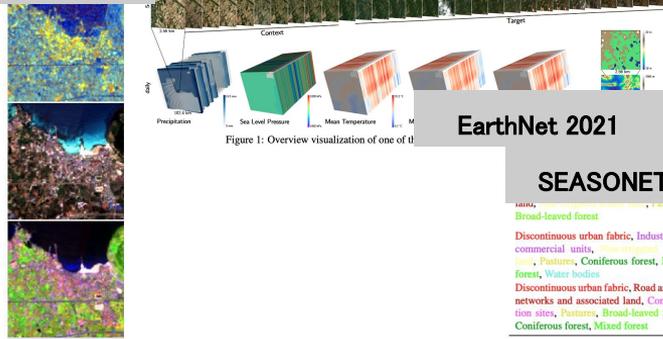
SynthWakeSAR images ($\theta_r = 20^{\circ}$) of ship wake (Passenger Vessel I) with $V_s = 8$ m/s, $D_s = 45^{\circ}$ and $V_{w0} = 3$ m/s: (a) noise-free I ; (b) with noise I_n ; (c) denoised I_d .



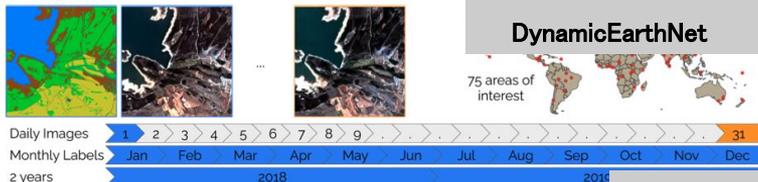
地物のモニタリングや予報を行うため衛星データを用いて分類・予測

- SEN12MS(2019): 解像度500m Sentinel1/2とMODISのLand coverを用いて作成したデータセット 詳細 p48
- EarthNet2021: Sentinel2を用いた未来の土地被覆を予測するデータセット 詳細 p52
- DynamicEarthNet(2022): 地球上に分布する75の選択された関心領域の毎日のマルチスペクトル衛星画像とそのセグメンテーションラベルを持ったデータセット 詳細 p57
- SEASONET(2022): 季節のデータとセグメンテーションラベルを増やしたLULCデータセット
- MultiSenGE(2022): 解像度10mのSentinel 1 (SAR), Sentinel 2(光学)のセットに、解像度50m相当のセグメンテーションマップ(土地利用・土地被覆図: LULC)を組み合わせたデータセット 詳細 p65
- AIR-PolSAR-Seg(2022): 合成開口レーダ(SAR)を用いて地表の解析を行う際には、照射した電磁波の偏波特性を使用して地物の判読を行うデータセットを作成 詳細 p62

SEN12MS



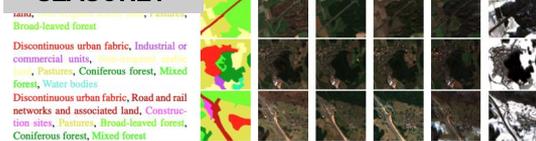
EarthNet 2021



DynamicEarthNet



SEASONET

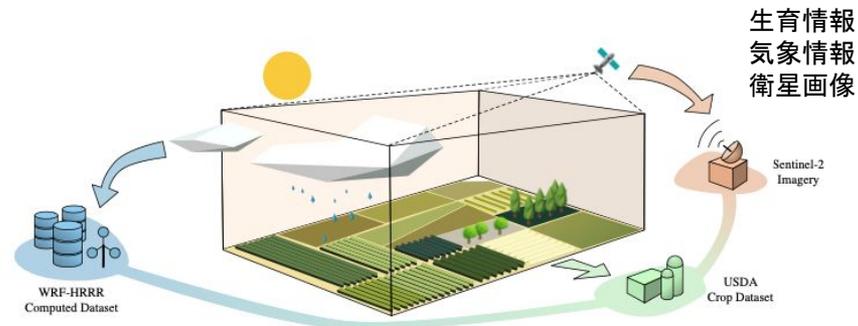
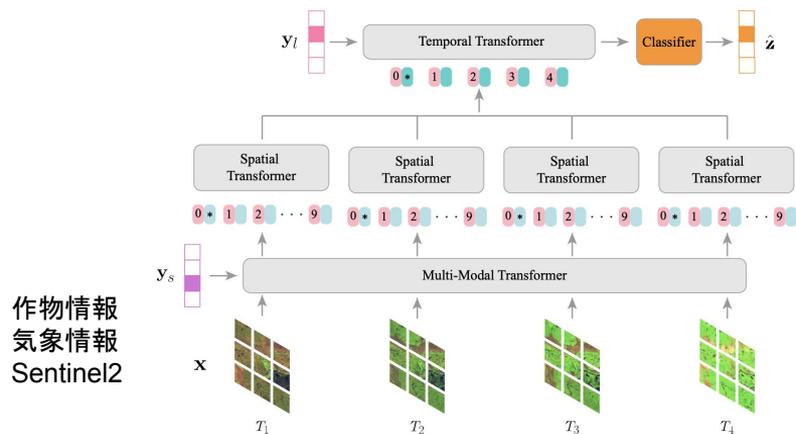


AIR-PolSAR-Seg



穀物の育成状況を衛星画像で把握するためのデータセットが提案

- 穀物の収量予測を行い食糧不足に陥るかを事前に把握するのは国家戦略として重要であり、収量を予測するために衛星画像が用いられつつある。
- 衛星画像から小麦の収量予測を行う際には、時系列の衛星画像を使用して,CNN+LSTMによる今後の収量を予測する手法が一般的であった。
- しかし、2023年くらいから、気象データなどと衛星画像を組み合わせたマルチモーダルデータセットが提案されており、高精度に収量予測が可能になってきた。

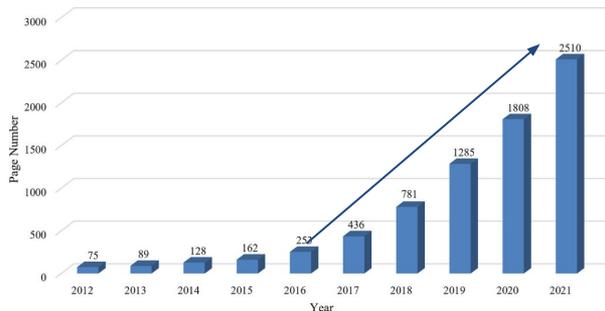


(国内では)深層学習技術があまり使用されていない

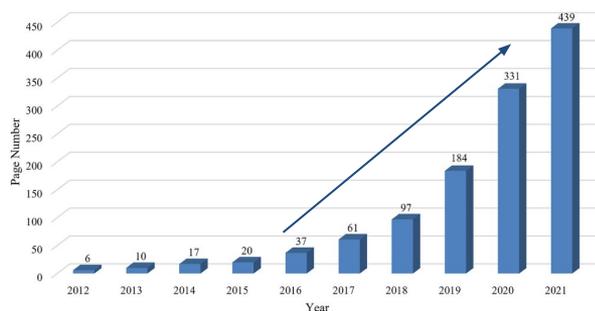
- リモセン分野の国際雑誌では、DL/CV技術を利用した研究の投稿数は増え続けている
 - リモセン系の国際雑誌以外でも、CV分野の会議論文で提案手法の評価として衛星画像も使用される
- 一方、国内の学会発表数はそこまで増えていない
- 海外と比べてコンテンツやソリューションの民間需要が少ない？

リモセン系の国際雑誌における論文数

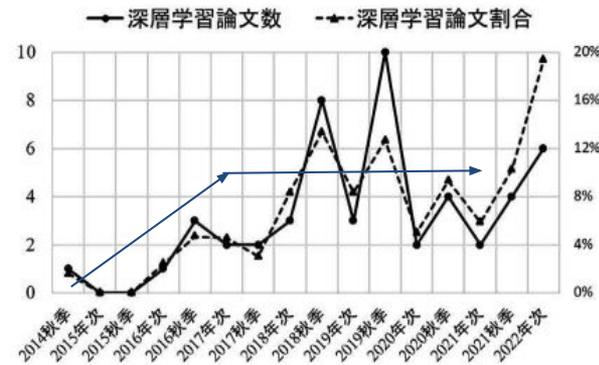
(a) Remote sensing AND (machine learning OR deep learning)



(b) Remote sensing AND (machine learning OR deep learning) AND (lithology OR surface water OR soil OR mineral OR geological structure OR glacier OR disaster)



国内学会の発表数



Wei Han. (2023), "[A survey of machine learning and deep learning in remote sensing of geological environment: Challenges, advances, and opportunities](#)", ISPRS Journal of Photogrammetry and Remote Sensing, 2022 より引用

李 勇 鶴. (2022), "[日本写真測量学会学術講演会の発表論文からみ/写真測量・リモートセンシング分野における深層学習の活用状況](#)写真測量とリモートセンシング, 2022より引用

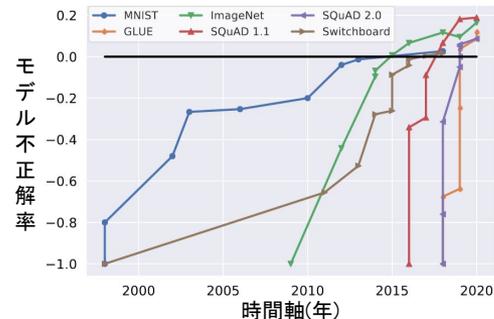
(国内では)最新のCV技術ではなく古いモデルをベースに手法を改良するが多い

- 色々な人に検証された深層学習モデルが主に使用される
 - 分類タスクだとVGG, ResNetが今でも現役な場合がある
 - セグメンテーションではUnetが人気※最新手法をベースラインにして、それを改良する日本の研究グループ・企業もいる
- 一般画像で性能がでてリモセン画像に対して高性能が出ないことが多い
 - 衛星毎に分解能が異なるため汎化的なマルチスケール性に対応できない
 - 通常の画像では無視される小さなオブジェクトが重視される
 - RGB以外にも赤外領域のチャンネルやハイパースペクトル、SARなどが存在
- 新しい分野との人材交流の機会が少ない？
 - 狭い国土から伝統的な現地調査に繋ぐことが比較的容易(国内の場合)
- 常に目的が先にあるので、手法が優秀かどうかより目的を達成できるかどうか判断基準？

以降のスライドでCVで行われているモデル研究について紹介

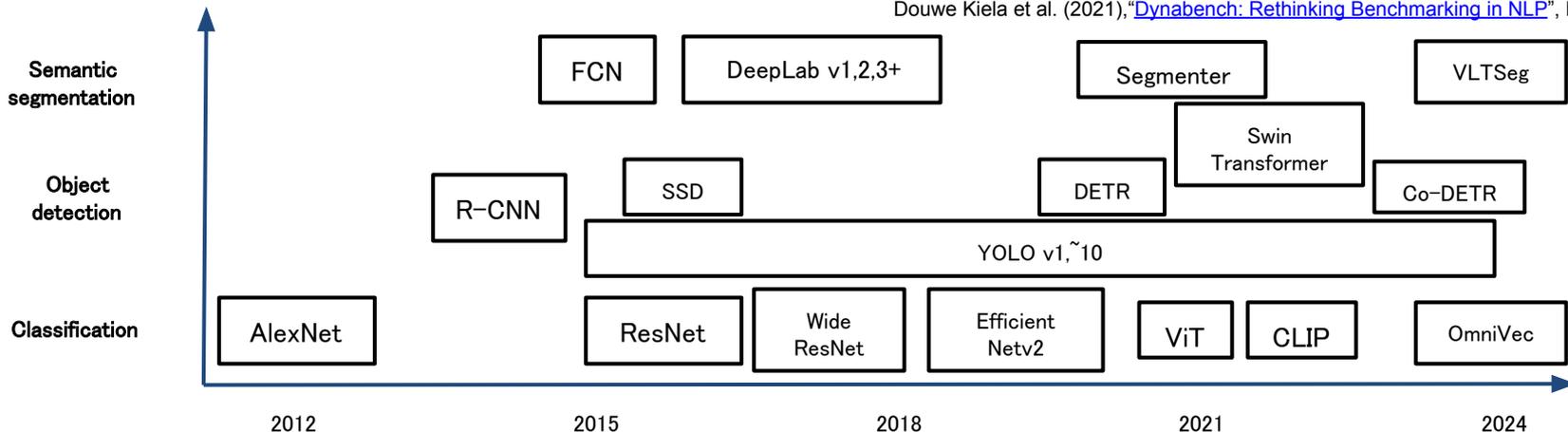
ベンチマークデータセットを評価基準とした深層学習モデルのupdate(時系列に沿って紹介)

- モデル開発の目的は、分類、物体認識、領域分割の各タスクでベンチマークに対して精度・計算効率を向上させること
- モデルの大規模化・精度向上に応じて、よりデータ数の大きいベンチマークが登場
- タスクによっては人間の正答率を上回るモデルが出現



(初期モデル不正解率を -1, 人間の不正解率を 0として正規化)

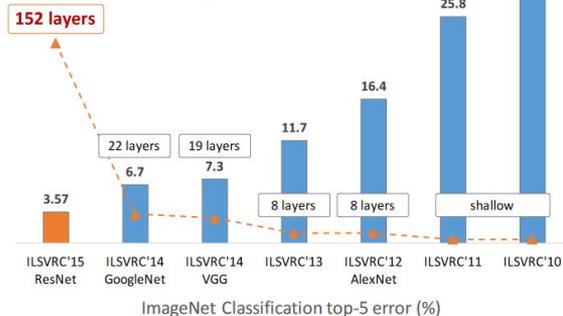
Douwe Kiela et al. (2021), "[Dynabench: Rethinking Benchmarking in NLP](#)", NAACL 2021. より引用



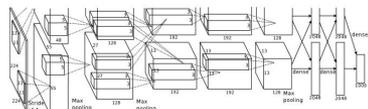
深層学習モデルの初期の研究は層を深くするための研究が盛ん(2014~2016)

- 深層学習が登場以降、層を深くすることで複雑な画像認識モデルを近似できることから深層化が行われた
- 代表的な研究はAlexNet、VGGNet、GoogLeNet
- ResNetは残差接続を導入することで深層化をさらに促進する流れとなった

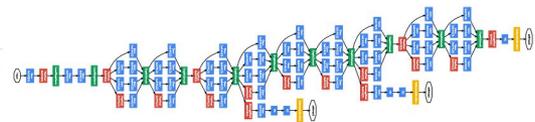
Revolution of Depth



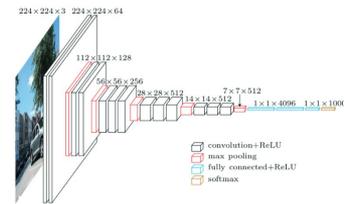
[Deep Residual Networks](#)
[Deep Learning Gets Way Deeper](#) より引用



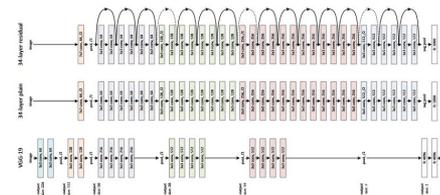
AlexNet [Krizhevsky+, ILSVRC2012]
 ILSVRC2012 winner, 深層学習と畳み込み層が主流となるきっかけ



GoogLeNet [Szegedy+, ILSVRC2014/CVPR2015]
 ILSVRC2014 winner, 22層モデル,
 計算コスト削減とフィルタマルチスケール化を実現



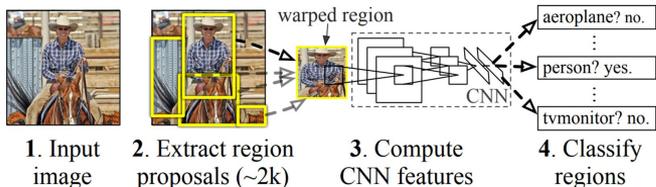
VGGNet [Simonyan+, ILSVRC2014]
 (図は [Singunum+, 2019])
 16/19層モデル, 深層化による精度向上を実現



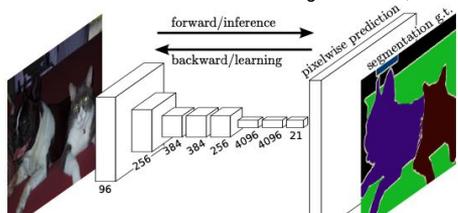
ResNet [He+, ILSVRC2015/CVPR2016]
 ILSVRC2015 winner, 152層モデル,
 残差接続による勾配消失の解決と深層化

CNNを応用し物体検出や領域分割タスクに特化する流れに(2014 ~)

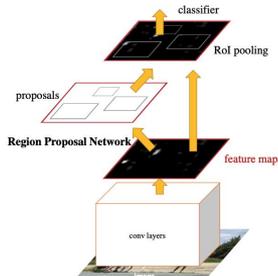
- 物体検出: R-CNN(R-CNN)、Fast/Faster R-CNN、YOLO、SSD (2014~)
 - それ以降は高速化を目指した手法が複数提案
 - その後、精度向上を目指した手法の提案
- 領域分割: FCN、SegNet、U-Net、PSPNet、DeepLab (2015~2017年)
 - その後、ピラミッド構造のPooling 構造を用いてグローバル情報を利用いたPSPNetやDeepLabが提案されて精度向上が行われた(2017年)



Ross Girshick et al. (2014), "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR. より引用

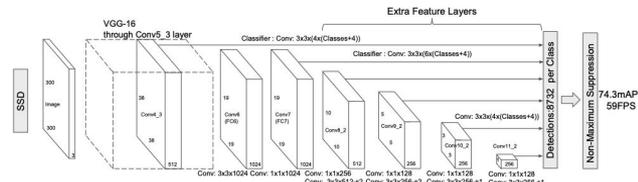


Jonathan Long et al. (2015), "Fully Convolutional Networks for Semantic Segmentation", CVPR. より引用

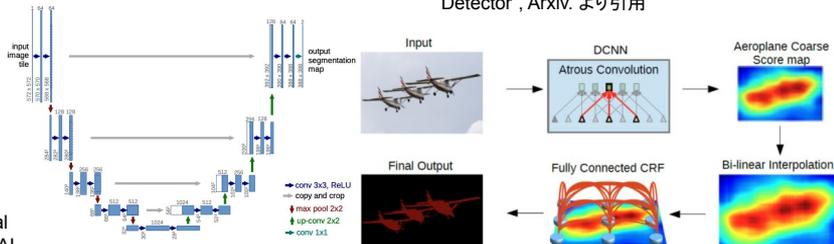


Shaoqing Ren et al. (2015), "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", NeurIPS. より引用

Olaf Ronneberger et al. (2015), "U-Net: Convolutional Networks for Biomedical Image Segmentation", MICCAI. より引用



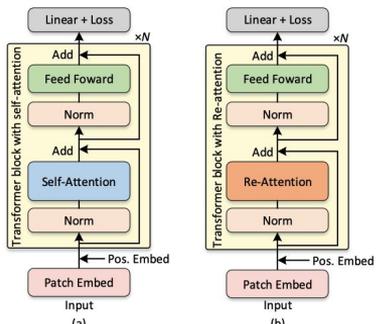
Wei Liu et al. (2016), "SSD: Single Shot MultiBox Detector", Arxiv. より引用



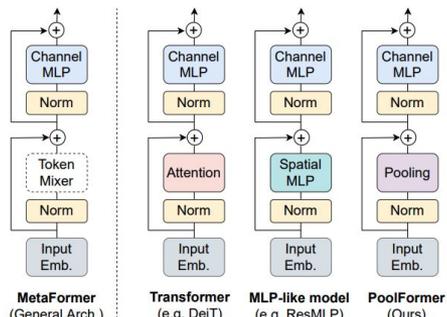
Liang-Chieh Chen et al. (2017), "U-Net: Convolutional Networks for Biomedical Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs", TPAMI. より引用

Vision Transformerの登場によりViTの理解・改善が活発化 (2020 ~ 2021)

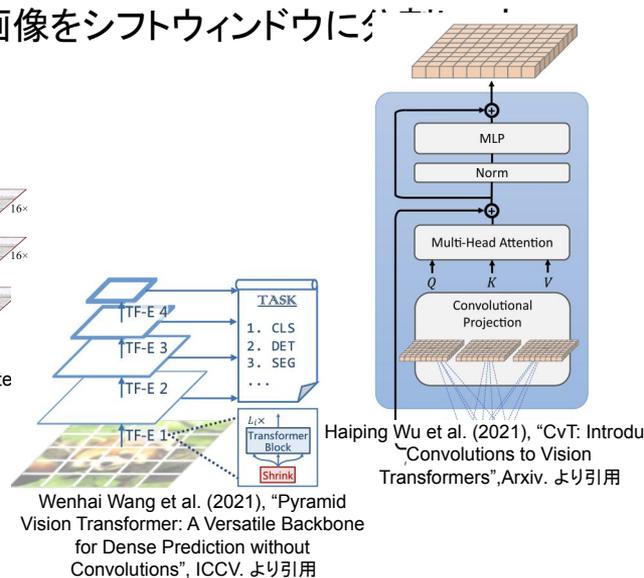
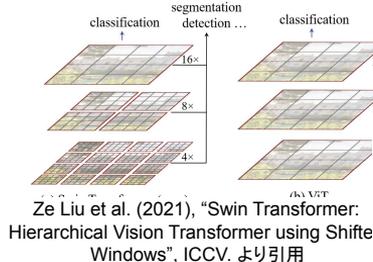
- ViTの処理を理解し改善する取り組み (2021)
 - DeepViT (ViTの深層化のために異なるヘッド間で特徴量を混合するRe-attentionを導入)
 - gMLP、MetaFormer、PoolFormer (Attention必要ないのでは? → Token information mixingが大事 → 情報さえ共有さえできればPooling層のみで十分)
- CNNでの知見をViTに組み込む取り組み
 - [Swin-Transformer](#): 画像の局所的な特徴を効率的に捉えるため、画像をシフトウィンドウに分割し、ドウ内の自己注意を計算
 - PVT、CvT ... (ViTの構造に畳み込みをうまく使うことで精度向上)



Daquan Zhou et al. (2021), "DeepViT: Towards Deeper Vision Transformer", Arxiv. より引用



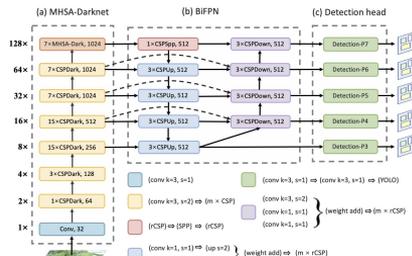
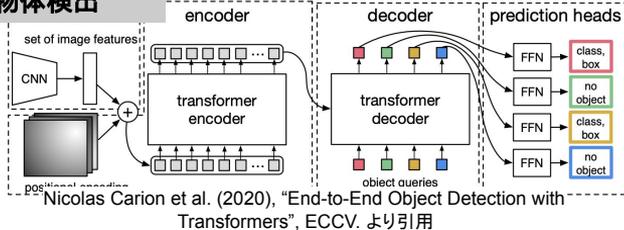
Daquan Zhou et al. (2021), "DeepViT: Towards Deeper Vision Transformer", Arxiv. より引用



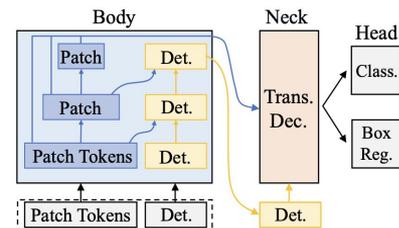
CNNの流れと同様にTransformerは多分野タスクへの応用が行われる (2020~)

- 物体検出: (Co-)DETR(Transformerと物体検出の融合)、[ViT-YOLO](#)(TransformerとYOLOを組み合わせさせて精度と速度向上)、ViDT(Swin-TransformerとDETRを計算効率が良い形で組み合わせ精度向上を行った)
- 領域分割: [Segmenter](#)、[SegFormer](#) (2021)、Swin-Unet (2021)、TransSegNet

物体検出

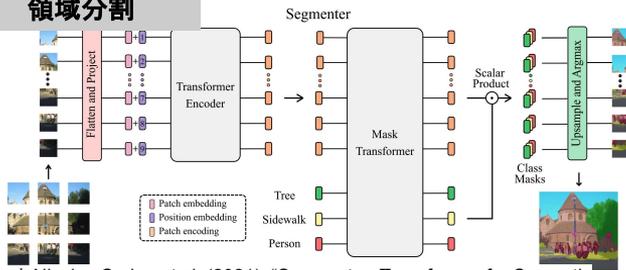


Zixiao Zhang et al. (2021), "ViT-YOLO: Transformer-Based YOLO for Object Detection", ICCV workshop. より引用

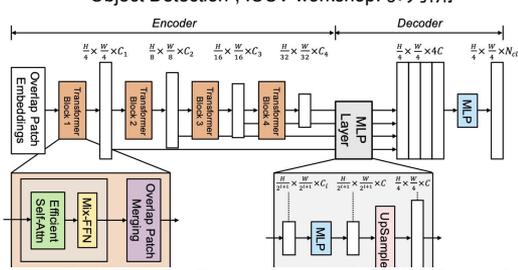


Hwanjun Song et al. (2022), "ViDT: An Efficient and Effective Fully Transformer-based Object Detector", ICLR. より引用

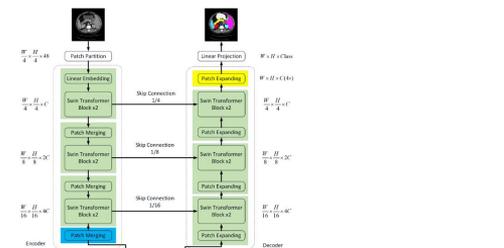
領域分割



Nicolas Carion et al. (2021), "Segmenter: Transformer for Semantic Segmentation", ICCV. より引用



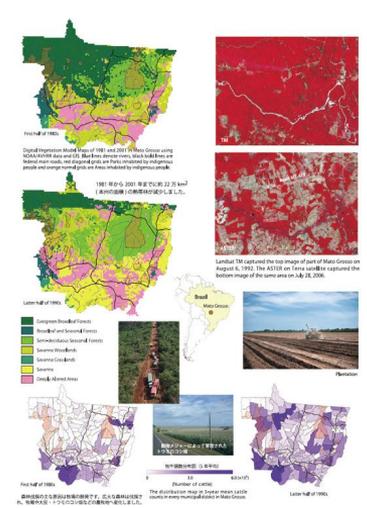
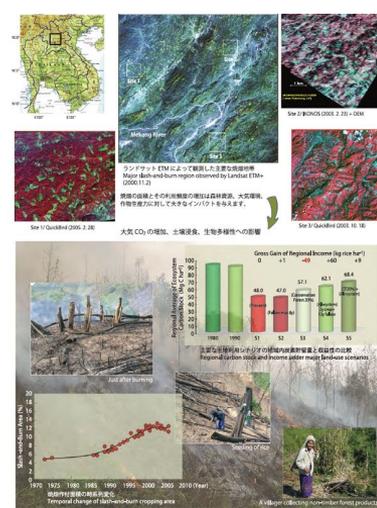
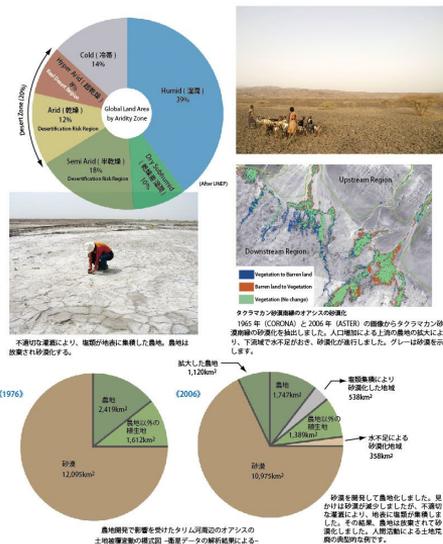
Enze Xie et al. (2021), "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers", NeurIPS. より引用



Hu Cao et al. (2021), "Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation", ECCVW. より引用

理由①: リモートセンシングではミッション目的の達成が観測の最優先事項

- 衛星リモートセンシングは地球環境問題の解明 に使われることが多いが、観測種別・高次処理データ処理アルゴリズムは打ち上げ前に確定していることが多い
- 試しにデータを取ってみるといふ観測は少なく、PI(研究主宰者)からの要求をもとに調整し観測
- 研究的な観測提案機会であっても成果の見通しを述べる必要があり、従来技術による見積もりをベースにする場合が多い

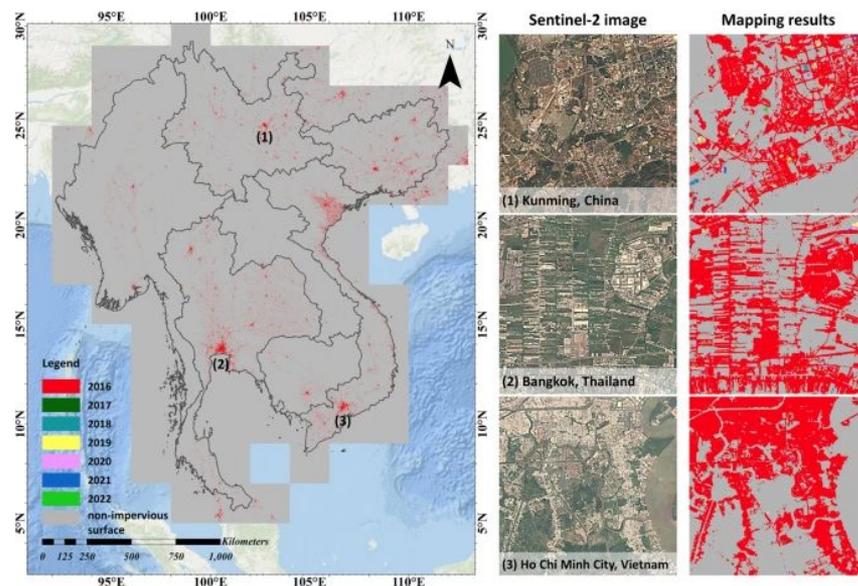


補足: ミッション目的・搭載機器・機器スペックは早い段階で決定される

- 衛星開発のタイムライン
 - ミッション目的の決定(宇宙からの観測で何がしたいか・何ができるか) → 人工衛星開発(設計から構築) → 打ち上げ → 運用
- ミッション目的の決定はすべてに先行して行われる
 - 地球観測衛星の構成は、宇宙からの観測で何がしたいか(目的)から決定される(e.g. 陸域監視)。
 - さらにその目的から貢献できる課題は何か、どこまで達成できるのか、ということが議論される(e.g. 災害監視、農業支援、資源探査、環境モニタリング、etc...)。
 - 並行して人工衛星に何を載せるのか(光学センサかSARか、波長帯やそのバンド幅、分解能、太陽同期/傾斜軌道)、投入軌道などが決定する。
 - 研究者はミッション要求の設定、機器開発・校正精度評価、データ処理の検討などハードウェア・ソフトウェア両面からミッション目的達成に必要な事項を明らかにする役割を担う
- 搭載機器が決まったのち、要求性能をハードウェアのみで達成できるように機器スペックが決まる(要求性能自体HWで達成しうるギリギリを調査し決定される)
 - 機器の性能評価はすべて説明できる(スペックシートに記述できる)こと
 - 超解像など機械学習的な力を借りる後処理まで込みで性能とすることは少ない
 - CVの関係者がミッションの組成段階から携わると、現行と異なる人工衛星開発になる可能性がある

理由②: リモセン分野では衛星画像で観測したい事象をモニタリングできればOK

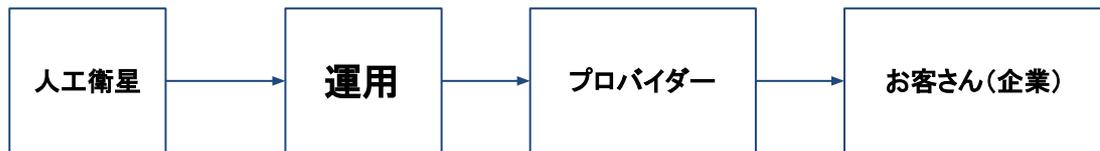
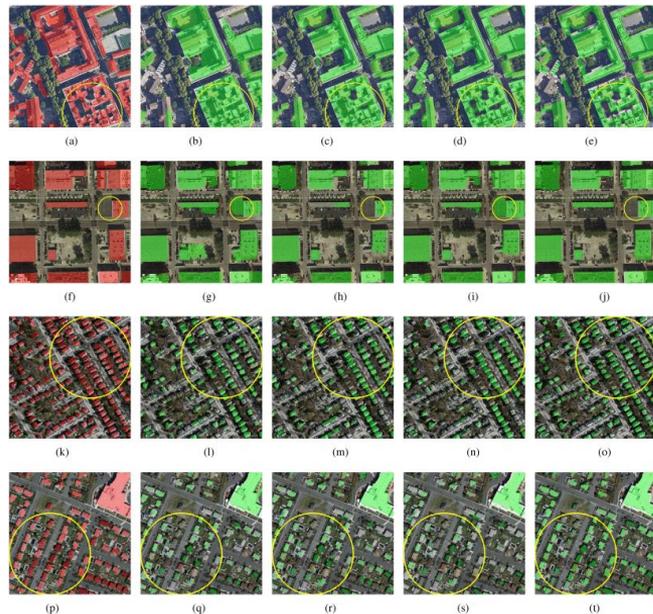
- 例えば地球環境問題としてヒートアイランド現象を予測したいとする
 - ヒートアイランドに寄与する指標は何かを考える
 - 建物や舗装道路が寄与する
 - それに関連する指標を衛星画像から計算する
 - 不透水面(建物・舗装道路)など
- 目的を達成できる指標が計算できたらOK
 - ミッションで決めた目的に対して自動的に全球規模で定期的にモニタリングするのが重要



Sun et al., “[A 10-m resolution impervious surface area map for the greater Mekong subregion from remote sensing images](#)”, Scientific Data 2023 より引用

理由③: 衛星画像に付加価値を与えるタスクは重きを置かれてないから？

- 物理量のモニタリング以外にも衛星画像を使用したサービスを展開したい場合を考える
 - 衛星画像から建物の位置を推定できたら、固定資産税向けの材料になる
 - セグメンテーションで建物の場所を推定
 - 多時期であれば新築建物の自動抽出
 - 低解像度の衛星画像を擬似的に高解像度にできれば航空写真のように判読しやすい画像になる
 - 超解像
 - ※実際にはないものができていない保証が併せて必要 (Artifactの抑制手法が必要)



元データ

付加価値

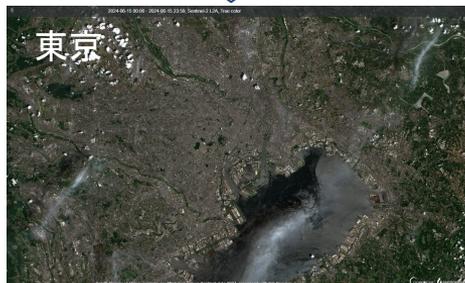
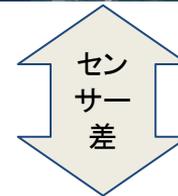
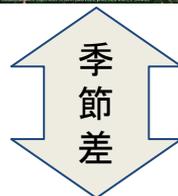
← リモセン重要度大

理由④:ドメインギャップがあるからモデル側よりデータ側が重要視されている？

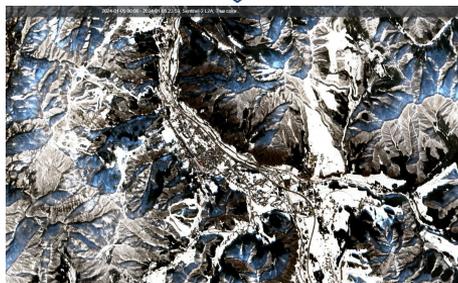
- 地域差・季節差・センサー差の影響を受けるので、汎用的なモデルを作るよりデータセット作成側が重要になる



夏

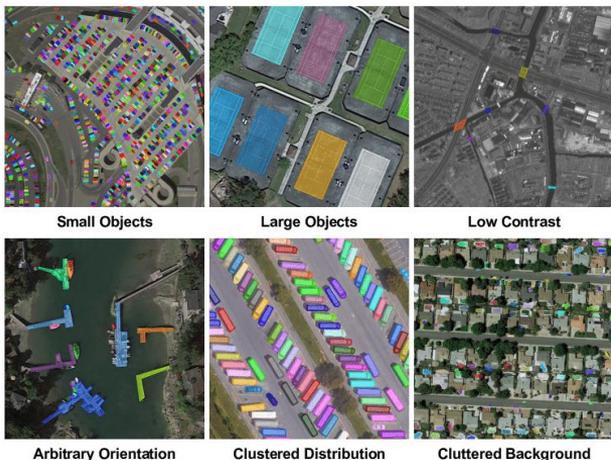


冬

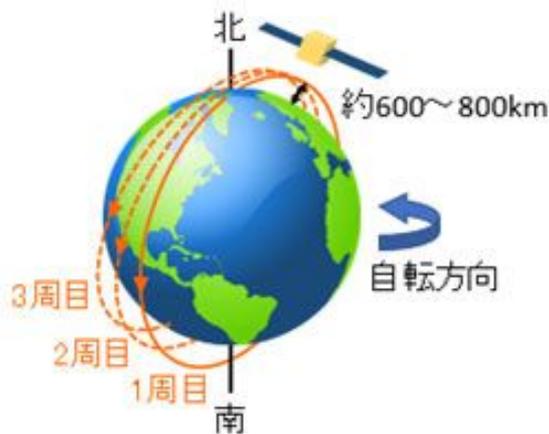


展望1: 情報付加価値のタスクはCVの人たちが衛星画像特有の難しさを解決する

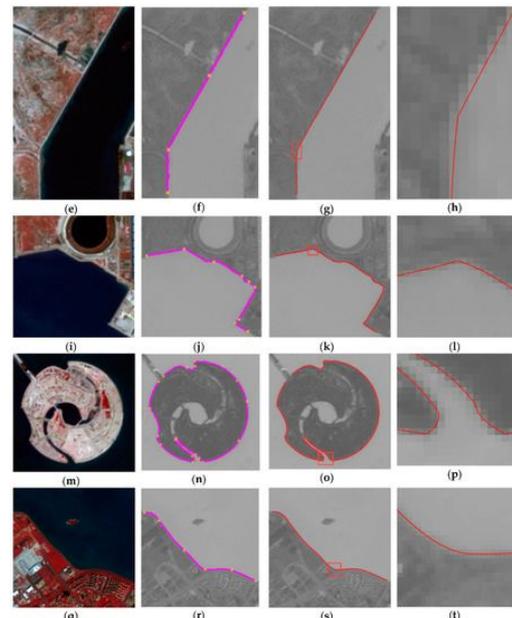
画像の特性
=> マルチスケールや小物体



高頻度なグローバルデータ
=> 超大量のデータ



サブピクセルレベルの判読
=> 輪郭の要求性能が高い



Ye Liu. (2023), "Learning to Aggregate Multi-Scale Context for Instance Segmentation in Remote Sensing Images", IEEE Transactions on Neural Networks and Learning Systems (TNNLS), 2023 より引用

国立環境研究所
<https://adaptation-platform.nies.go.jp/local/communication/collaboration/satellite-image-data/index.html> より引用

Song, Y. (2019), "Automatic Semi-Global Artificial Shoreline Subpixel Localization Algorithm for Landsat Imagery", Remote Sens. 2019 より引用

海外の事例①: 衛星画像にマルチスケール表現を活用する

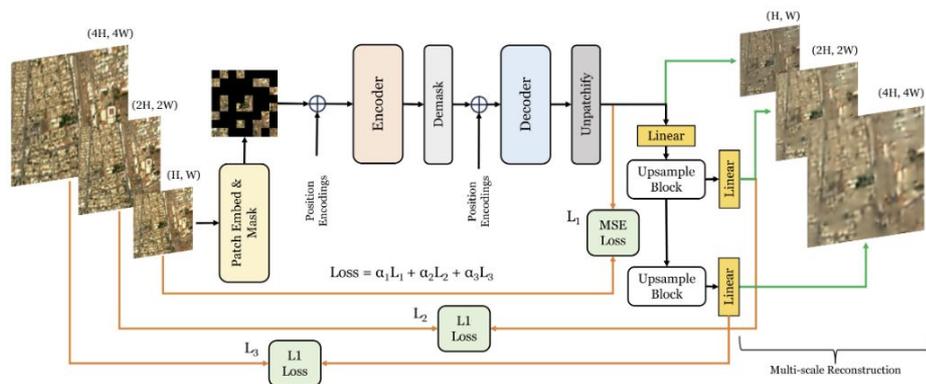
- スケール: 自然画像と違い、衛星画像の隣り合った画素間の距離はセンサーや観測条件によっては0.3m ~ 1.0kmと広範囲にわたって変動する(たった1画素でも距離が遠い場合がある)。
 センサを跨いだ良い汎化性能を得るにはマルチスケールな表現を獲得することが重要である。

スケールを考慮することでより高い性能を獲得



Reed et al., “[Scale-MAE: A Scale-Aware Masked Autoencoder for Multiscale Geospatial Representation Learning](#)”, in ICCV 2023 より引用

明示的にマルチスケール表現を獲得する

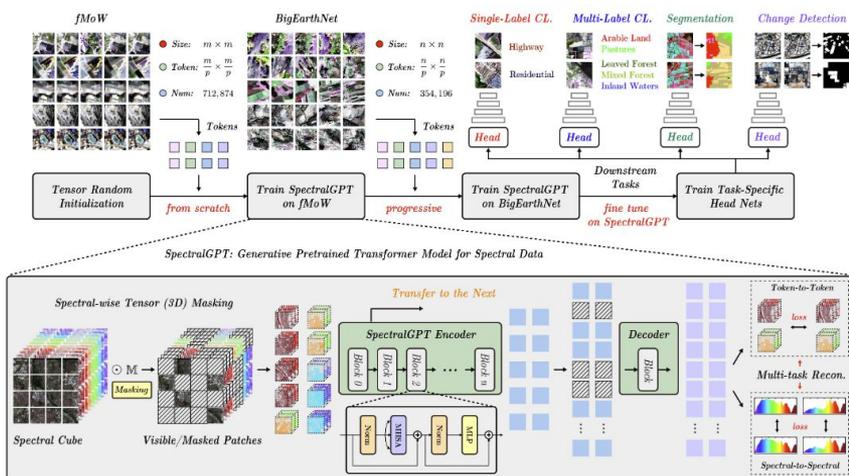


Noman et al., “[Rethinking Transformers Pre-training for Multi-Spectral Satellite Imagery](#)”, in CVPR 2024 より引用

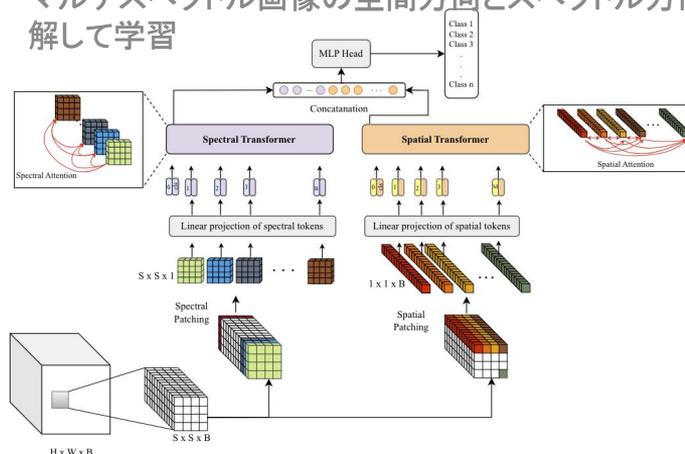
海外の事例②: マルチスペクトルとMAEの融合

- マルチスペクトル: 衛星画像は自然画像 (RGB画像) に比べて使用できるチャンネル数が多い。MAEなどの自己教師あり学習をするときは、空間とスペクトル(チャンネル)方向のマスキングの仕方を分けるなどの工夫によって、より良い表現が得られる。

token だけではなくスペクトルも復元



マルチスペクトル画像の空間方向とスペクトル方向を分解して学習



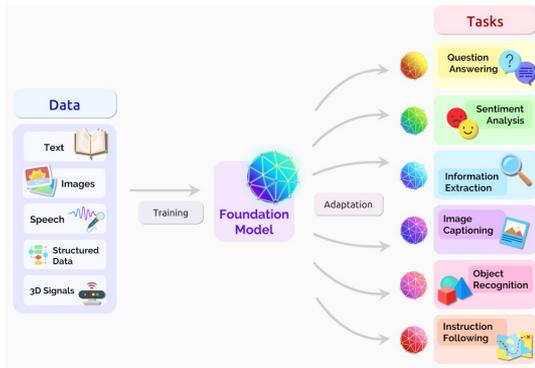
Mohamed et al., "[FactoFormer: Factorized Hyperspectral Transformers with Self-Supervised Pretraining](#)", in IEEE TGRS より引用

Hong et al., "[SpectralGPT: Spectral Remote Sensing Foundation Model](#)", in TPAMI より引用

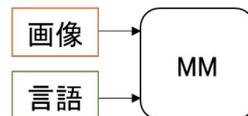
展望2: ML系の最先端である大規模言語モデル(LLM)の進化をキャッチアップ

- 基盤モデルの出現
 - 基盤モデルは、言語だけでなく画像や音声などの様々な入力に対して様々なタスクが実行可能
 - 一つのモデルで様々なタスクに対応できるためユーザーにとって利点が多い
- 基盤モデルの発展
 - CLIPに代表される、マルチモーダルモデル(MM) ⇒ 異なるモダリティを同時に扱う
 - ChatGPTに代表される大規模言語モデル(LLM) ⇒ 言葉(プロンプト)で様々なタスクに対応
 - LLaVAに代表されるマルチモーダル大規模言語モデル(MLLM) ⇒ 画像+言語(プロンプト)

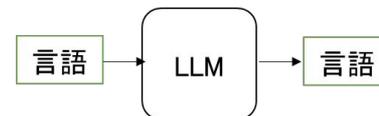
基盤モデル



Multimodal Model



Large Language Model

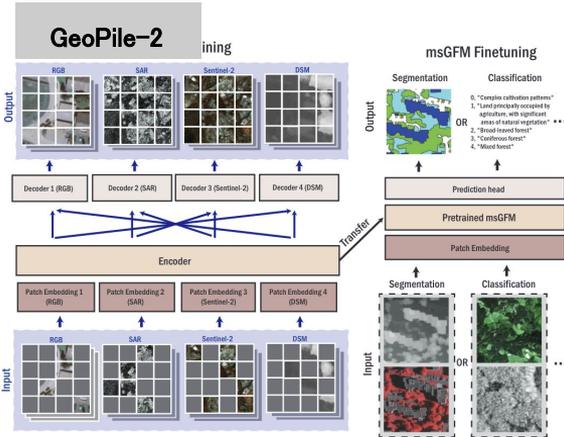


Multimodal Large Language Model



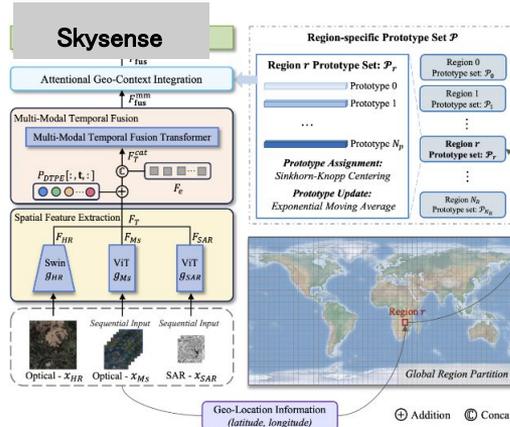
事例: 海外のリモートセンシングでも基盤モデル用のデータセットが提案

- 複数種類の衛星で撮られた地球各地の大規模データセットで学習し、強い汎化性能を獲得
- 光学、マルチスペクトルとSAR画像のマルチモーダルフュージョンにより大局的な地表の理解を実現
- 単・多時期画像同時学習により、一つのモデルでより多くのタスクに対応
- リモセン特有の地理的文脈やスケール表現をモデルデザインに取り入れることで、さらに汎化性能が向上
- 自然言語モデルと組み合わせてラベルにとらわれないオープンボキャブラリーな衛星画像の理解が可能に



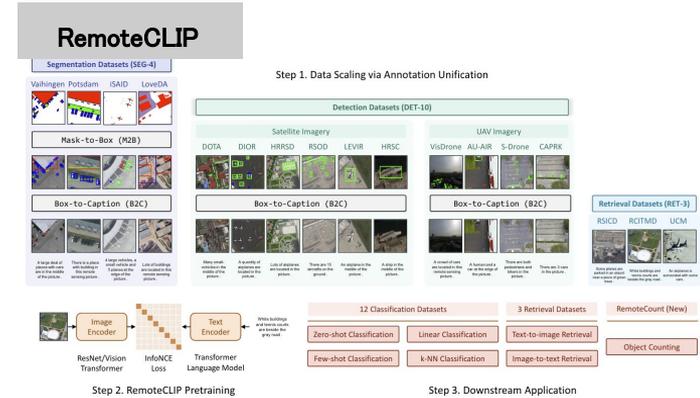
光学+マルチスペクトル+SAR

“[Bridging Remote Sensors with Multisensor Geospatial Foundation Models](#)”, CVPR 2024. より引用



単・多時期画像学習 + 地理的文脈学習

“[SkySense: A Multi-Modal Remote Sensing Foundation Model Towards Universal Interpretation for Earth Observation Imagery](#)”, CVPR 2024. より引用

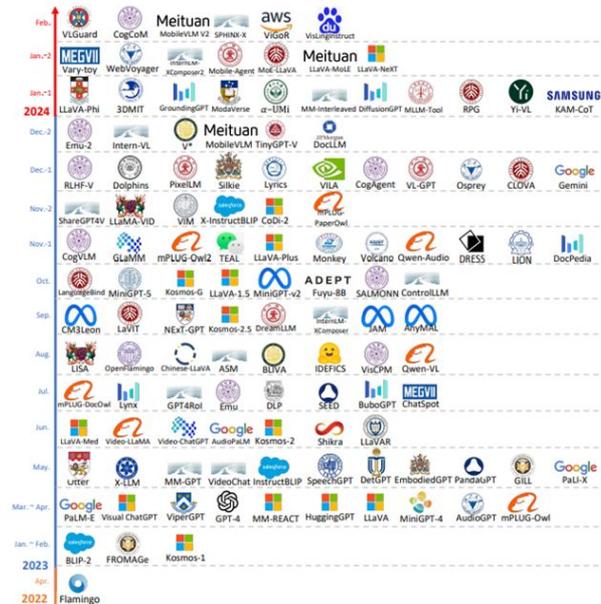


光学衛星画像と自然言語との接続

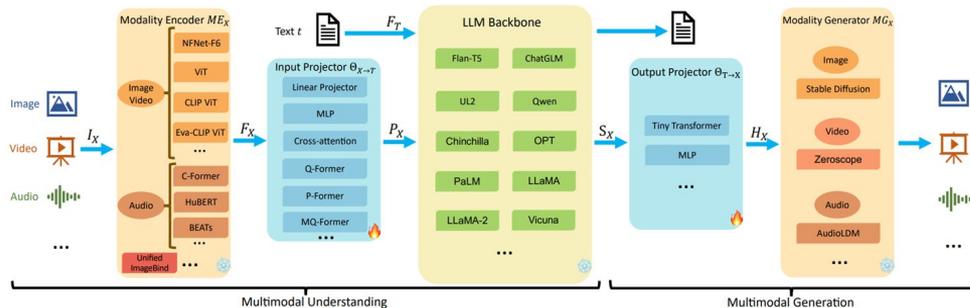
“[RemoteCLIP: A Vision Language Foundation Model for Remote Sensing](#)”, IEEE TGRS. より引用

LLMの応用: マルチモーダル大規模言語モデル(MMLLM)の発展

- 言語と画像を同時に扱えるマルチモーダル大規模言語モデル(MMLLM)は急速に増えており、2024年4月時点でも122種類のモデルが確認されている

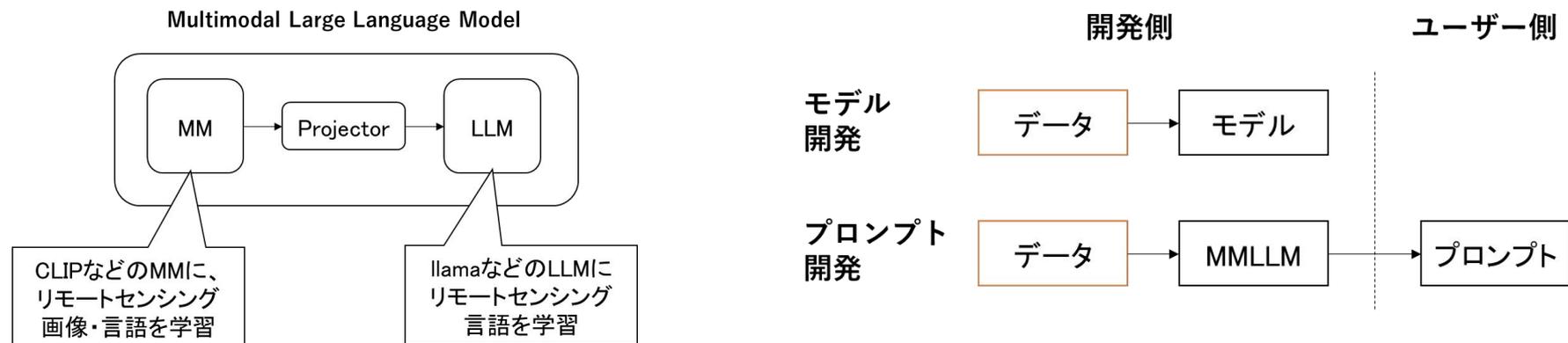


- MMLLMは既存モデル(オレンジ)を接続(青)することで構成できるため、学習コストが相対的に低い



MMLLMとリモートセンシング

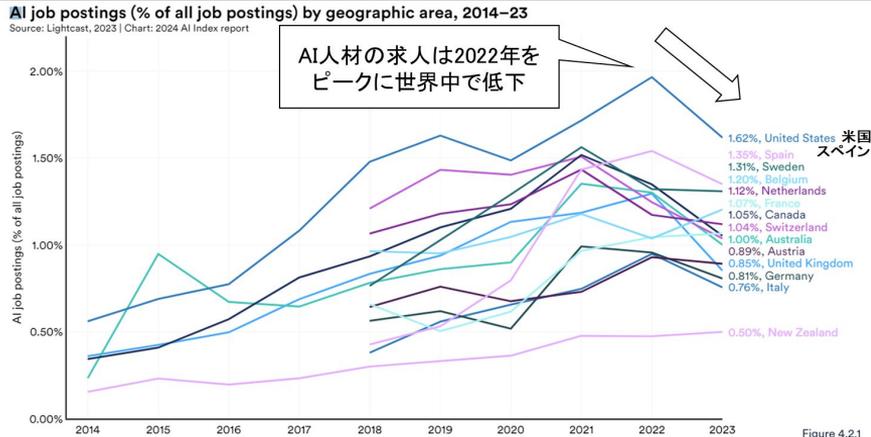
- 現在のマルチモーダル大規模言語モデル(MMLLM)はリモートセンシング領域とドメインが遠いため、ゼロショットの利用は難しい
- MMLLMはマルチモーダルモデル(MM)と大規模言語モデル(LLM)の接続で構成されているため、両者をリモートセンシングデータで学習すれば、ドメイン適応できてしまう。
- その文脈でRemoteCLIP等が開発されているため、データ開発が今後は重要となる可能性が高い。
- またMMLLMが一般化すれば、ユーザー側によるプロンプト開発も可能になる。



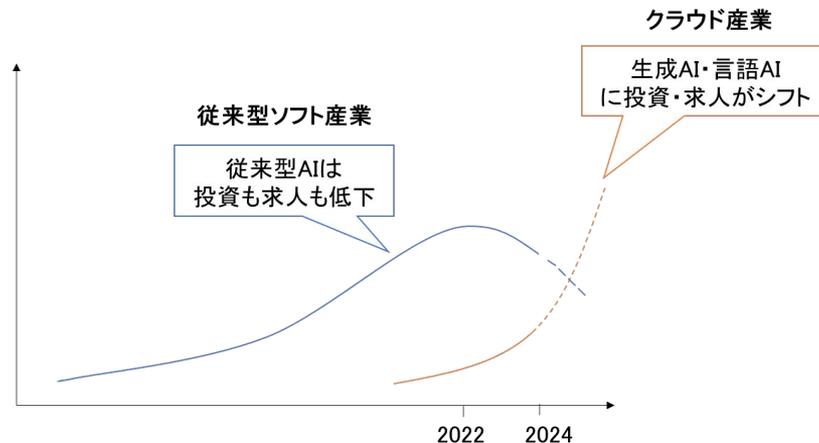
クラウド産業によるAI投資から見える未来

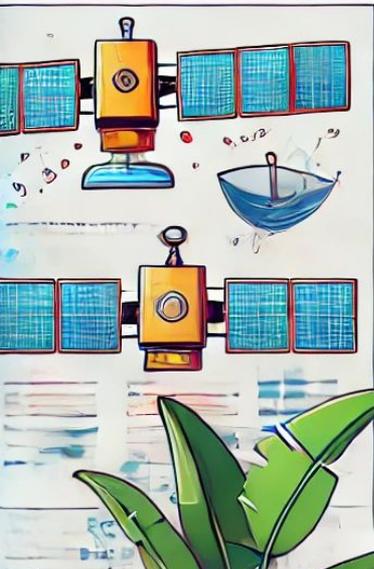
- 従来AI型の投資から基盤モデル(生成AI・言語AI)へと世界のAI投資トレンドも変化している
- これらの投資はクラウド産業の資金が主体になってきているため、**今後の大規模言語モデル開発はクラウドで動くモデルと、ローカルで動くモデルに二分される可能性が高い**
- データの秘密度に応じて、クラウド運用とローカル運用を分けて考える必要がある

世界各国のAI人材求人も2022年をピークに低下



世界の投資・求人トレンドが言語AIにシフト





メタサーベイ著者紹介



自己紹介 (Meta-survey project manager)



中村 凌

株式会社天地人 / cvpaper.challenge HQ (福岡大学出身)

cvpaper.challenge業績

- 研究効率化Tips (ViEW2021講演)
- CCCS, W2021/2022運営(登録者800名超え)
- IROS / ICCV 2023, ICASSP2024採択
- SSII2023オーディエンス賞受賞

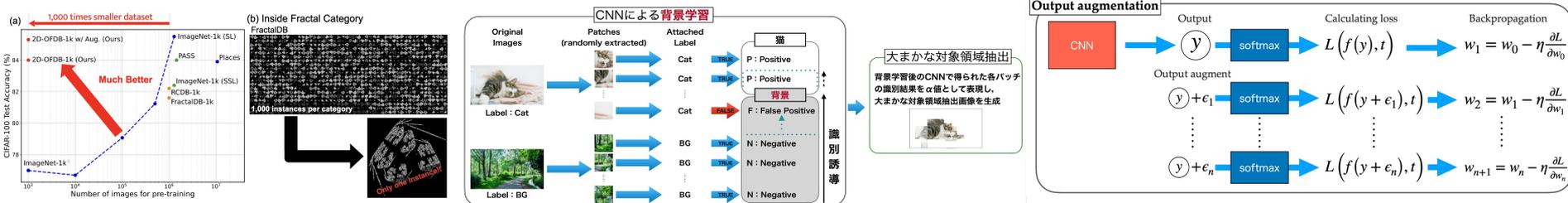
Twitter



LinkedIn



研究テーマ : Formula-driven Supervised Learning, Weakly supervised object localization, Output Augmentation



自己紹介



篠原 崇之

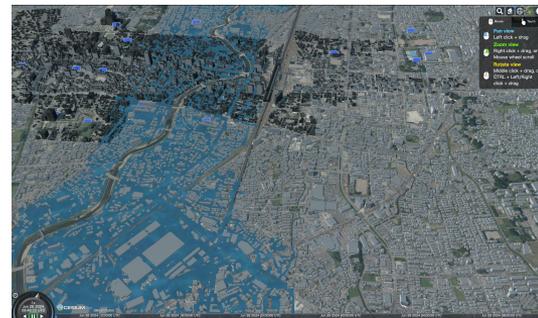
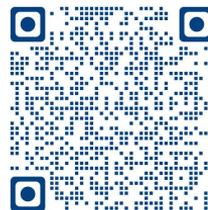
産総研 GSJ

- デジタルツイン構築と利活用
- 物理シミュレーションの近似を行うAI
- (産総研のサッカー部)

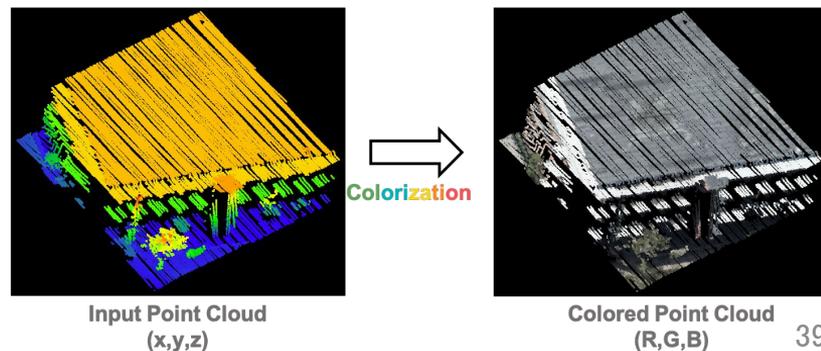
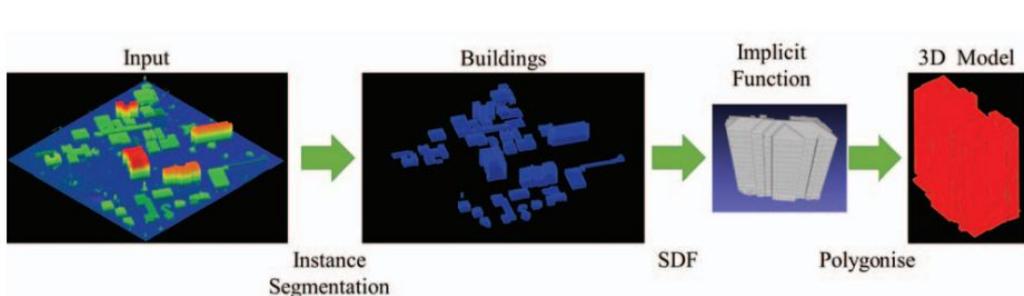
X(旧 Twitter)



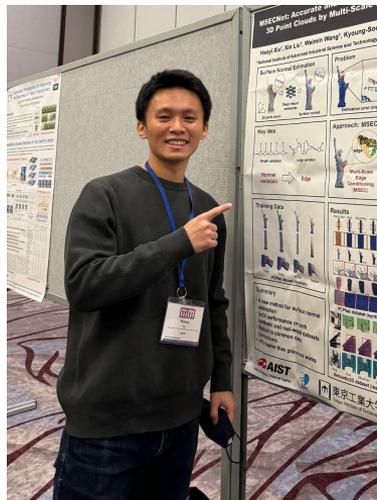
LinkedIn



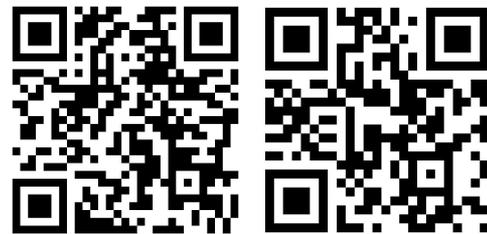
研究テーマ : データ基盤、点群深層学習、人間中心AI、3次元モデリング、サロゲートモデル



自己紹介



修 浩 毅

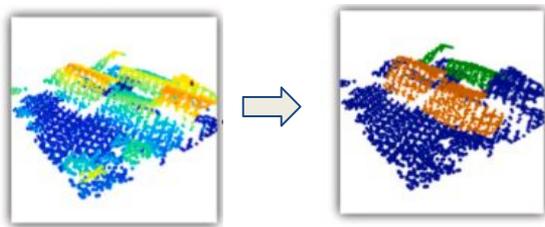


Linkedin

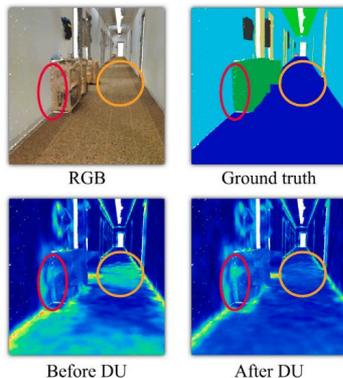
GitHub

産総研 データプラットフォーム研究チーム

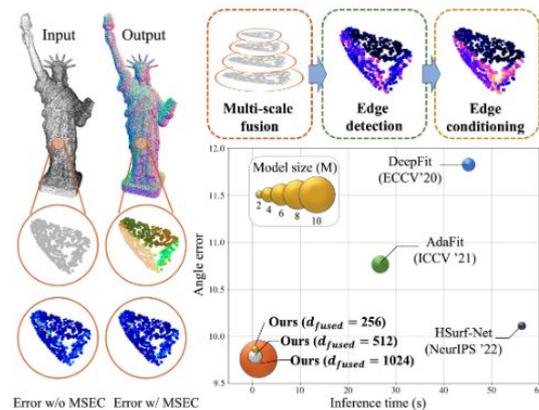
- 3次元点群解析
- コンピュータ・グラフィックス
- 建物被害検知



航空ライダーからの建物被害検知



点群セグメンテーション



点群からの法線推定

自己紹介

二村 忠宏



LinkedIn



スカパーJSAT株式会社スペースインテリジェンス事業部

過去の仕事: 災害対応用衛星データ提供システム内自動解析アルゴリズム開発 (CVベース、前々職)

損害保険自動支払い用AI開発PoC (技術DD中心、前職)

小型SAR衛星コンステレーションの利用拡大に向けた実証 (CV+AI、現職)

研究(業務)テーマ: 実業務に向けた衛星解析手法の探求

SAR画像の鮮明化 (SARの原理から計算)

SAR画像単画像からの土砂崩れ箇所抽出

自己紹介



天野 凌我

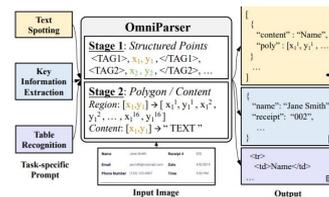
所属:

日立製作所 横浜研究所

研究テーマ:

Multimodal LLMを用いた特徴情報抽出

COBOLマイグレーションに向けたRAGシステムの構築



CVPR2024に聴講で参加したので

今度は発表者として参加できるように頑張りたい

自己紹介

神山 徹

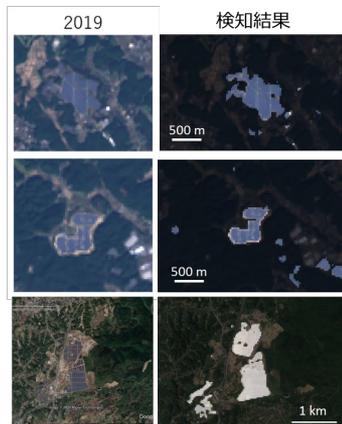
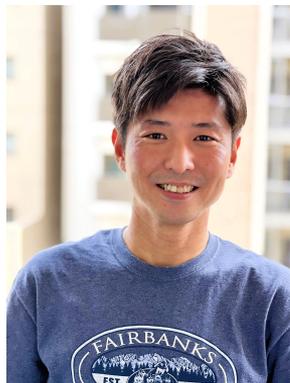
産業技術総合研究所 デジタルアーキテクチャ研究センター
地理空間サービス研究チーム

専門:(出身)惑星探査学:はやぶさ2・あかつき
衛星リモートセンシング x AI
センサキャリブレーション・衛星の運用

衛星運用



はやぶさ2撮影



[Koyama+, 2020 IGARSS]



自己紹介

中田 和真



Neural Network Libraries cvpaper.challenge コラボ

ベイズ誤差推定による
機械学習タスク難易度の定量化と
SOTA モデル評価

AI論文解説 ICLR2023



所属: 山口大学 情報認識工学研究室 D3

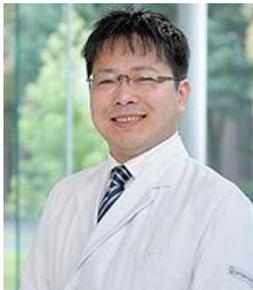
研究テーマ: 機械学習による超音波画像を用いた肝硬変の検出

その他活動:

- 2024年度 山口大学 SPRINGスカラシップ研究学生
- 第23回 IEEE広島支部学生シンポジウム 運営
- cvpaper.challenge コラボ AI論文解説

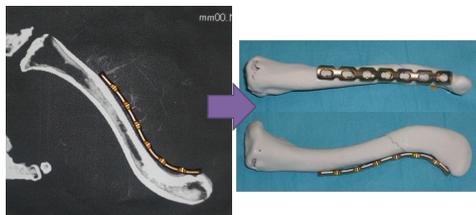
自己紹介

中原龍一

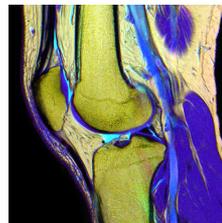


所属: 岡山大学整形外科

研究テーマ: AIを用いた言語(電子カルテ)・画像解析(MRI/CT/Xp/超音波)



手術と3Dプリンター



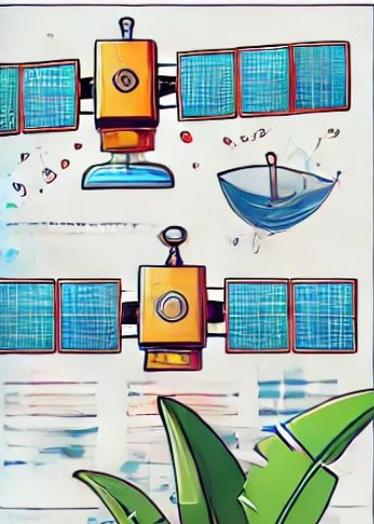
MRI 多次元解析



Doc2Chart: 文章解析
(NECとの研究)

感想:

- 医療領域では独自のモダリティーやタスクがあるため、AI開発にはドメイン特有の工夫が必要です。リモートセンシングにおいてもドメイン特有の工夫が行われている点が医療と似ていると感じました。
- 大規模言語モデルに代表される基盤モデルの出現により、ドメイン側の開発はモデル開発からデータ開発に移行しつつあります。この変化がリモートセンシングの領域でも生じていることに、とても興味をそそられました。
- リモートセンシングに関して素人である私に親切に教えていただいた皆さんに、厚く感謝申し上げます。

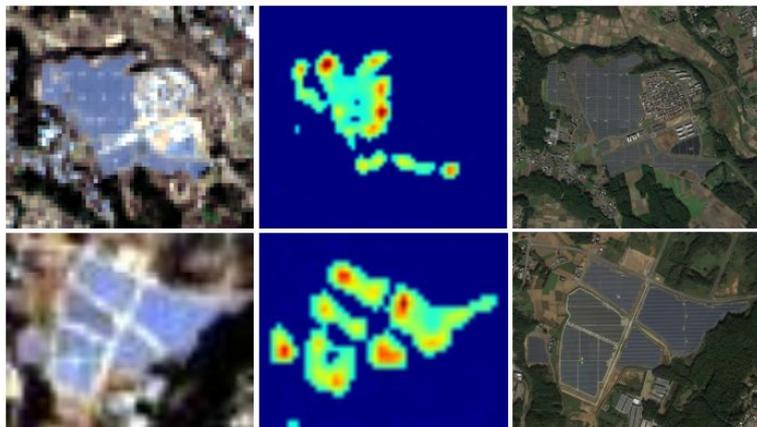


付録: 1ページサマリ



太陽パネルを光学衛星画像から検出するためのデータセットを公開

- 地球温暖化対策のために太陽光発電所が世界中で増加しているが、無許可に建造される太陽光発電所も問題になっており、太陽光発電所のマップを作る必要性が出てきている
- Landsat 8 と呼ばれる中解像度の光学衛星で取得されたマルチバンド画像を用いて、太陽光発電所のある画像か否かのパッチと太陽パネルのセグメンテーションマスクで構成されるデータセットを構築した
- ベースアルゴリズムとしてCNNを用いたパッチ画像の分類のほか、CAMなどの顕著マップによる弱教師付き物体検出手法の実験を行った



高解像度のGoogle Earthとの比較結果

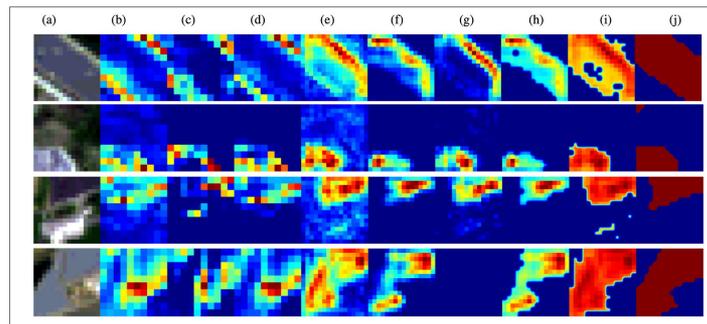


Figure 6: Solar power plant detection results of (a) sample images (RGB visualization of multi-spectral data) (b) I-Net [2] feature averaging, (c) I-Net+gap, (d) I-Net using Grad-CAM in [18], (e) feature averaging on FB-Net w/o gap, (f) FB-Net using CAM in [17], (g) FB-Net w/o gap using Grad-CAM [18], (h) Proposed FB-Net with mPCNN based CAM, (i) FB-Net w/o gap layer with mPCNN based Grad-CAM, (j) ground-truth

太陽パネルのパッチに対する顕著マップの可視化結果

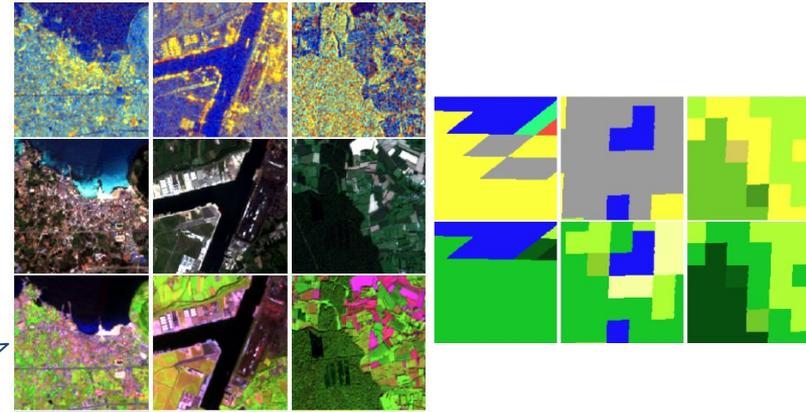
Sentinel-1,2画像とMODIS土地被覆マップを使ってSEN1-2を改良

- SEB12MSは従来のSentinel1/2から構成されたデータセットSEN1-2を改良したデータセットを提案
- SEN1-2は、データセットを単純化のため、dBスケールで垂直偏光(VV)画像とSentinel 2のマルチスペクトルデータのヒストグラムを調整したRGB画像に縮小していた
- 欧州宇宙機関(ESA)のCopernicusプログラムのSentinel衛星によって取得された自由に利用可能なデータとGoogle Earth Engineのクラウドコンピューティング機能を利用し、SAR、画像パッチ、マルチスペクトルSentinel-2画像パッチ、およびMODIS土地被覆マップの180,662トリプレットからなるデータセットを提供した

Dataset	Nr. of Images	Image Size	Data Source	Description
SEN12MS	541,986	256 × 256	Sentinel-1, Sentinel-2, MODIS Land Cover	globally distributed; MODIS Land Cover maps can either be used as labels or auxiliary data

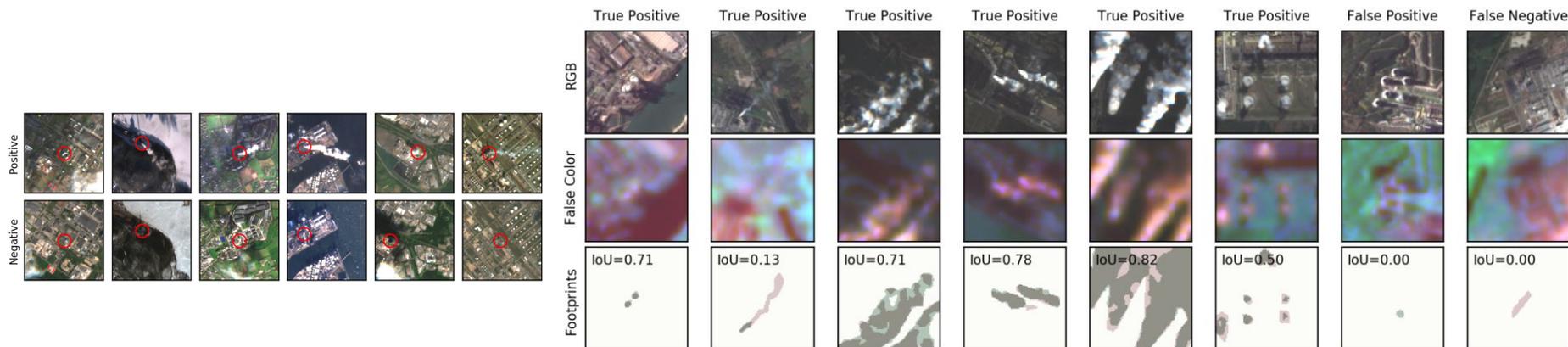
上から順に

Sentinel-1 SAR (R: VV, G: VH, B: VV/VH), Sentinel-2 RGB, Sentinel-2 SWIR, IGBP Land cover, LCCS Land cover



工場からの排煙を光学衛星画像から検出するためのデータセットを公開

- 地球温暖化と温室効果ガスの関係を定量的にモニタリングするためにリモートセンシング技術が必要
- Sentinel-2と呼ばれる欧州宇宙機関の衛星で取得された可視光以外の波長も観測可能なマルチバンド画像を用いて、産業由来の排煙の検出を行うデータセットを構築した
- ベースアルゴリズムとして、Unetを用いて排煙領域のセグメンテーションを実行した結果、同じような色を持つ雲と排煙の区別が可能で、IoU 0.608の性能を示した



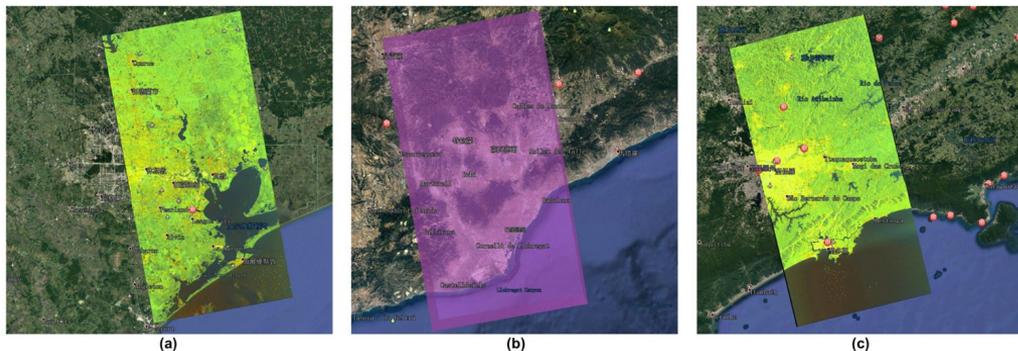
データセットの例

ベンチマークによるセグメンテーション結果の例

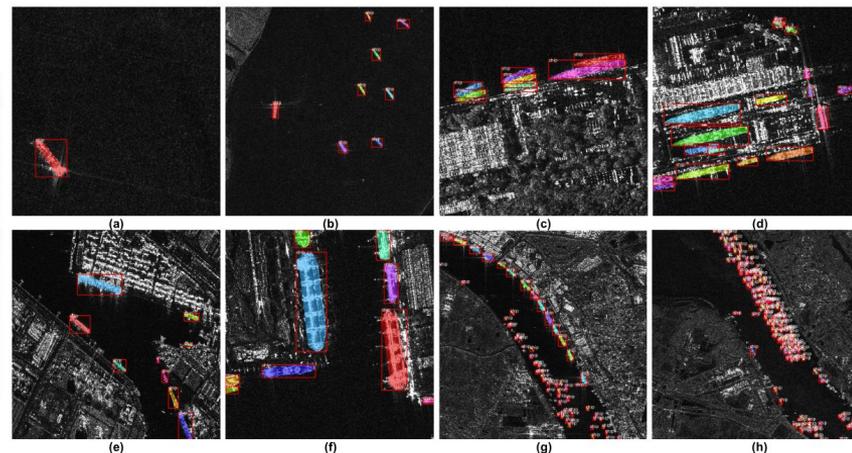
船舶を対象とした高解像度、SAR画像データセット

- 船舶を対象とした高解像度、SAR画像データセット
- 1mから5mまでの解像度を持つ136枚のパノラマSAR画像をもとに、800x800ピクセルのSAR画像を切り出して作成
- オリジナルのSAR画像は、99枚のSentinel-1B画像、36枚のTerraSAR-X画像、1枚のTanDEM-X画像
- 最終的に5,604枚のSAR画像と16,951隻の船舶のラベルを作成

切り出し前の大域画像

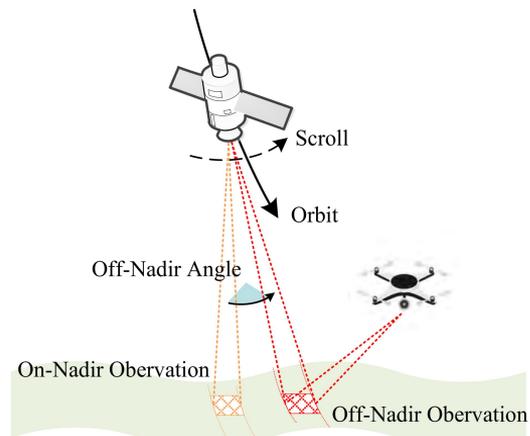


切り出し後の船舶ラベル

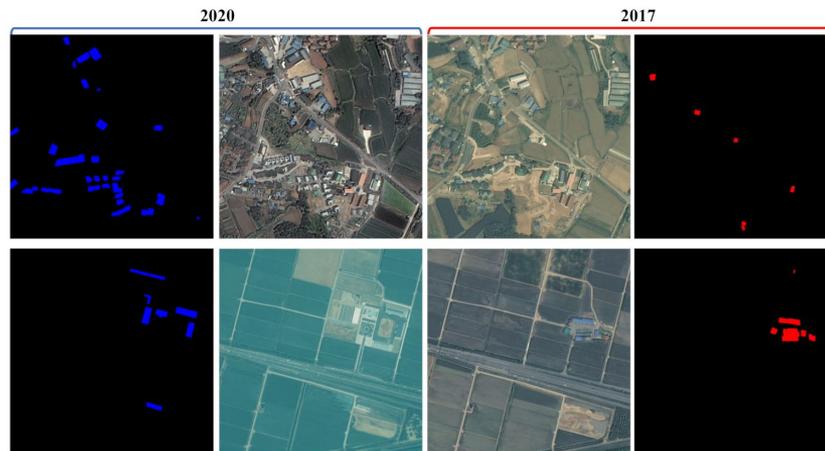


衛星画像特有の傾きに対応するための手法開発向けのデータセットを公開

- 衛星画像からの建物抽出は軍事や危機管理の分野で重要であるが、衛星画像のような広範囲を撮影する条件下では建物と衛星の角度が大きく、真上から見た建物ではなくなるため変化検出が難しい
- これは、サイドルッキングと呼ばれており、このサイドルッキング画像における建物変化検出を評価するためのデータセット(S2Looking)を公開した
- S2Lookingには、世界中の農村地域における5,000組の二時期画像が存在しており、既存の建物変化検出データセットよりも視野角が大きく、照度の変動、農村画像特有の複雑さが難しい点である



衛星画像における角度の問題



データセットの例 ※GaoFen (GF), SuperView (SV) and BeiJing-2 (BJ-2)

地表面予報という新しいタスクのための最初のデータセットを作成

- EarthNet2021という地表面予測のためのデータセットを提案
- 地表面予測のためには高時間・高空間分解能の衛星画像時系列と追加の気候予測因子が必要だが、これを実現するために5日おきに再訪する高解像度の光学画像のSentinel2をしようする
- 気象予測因子では海上天気予報が過去研究では用いられているが、気象モデルから予測値を得るためには計算上困難なためE-OBS、そして地形データEU-DEMを用いている

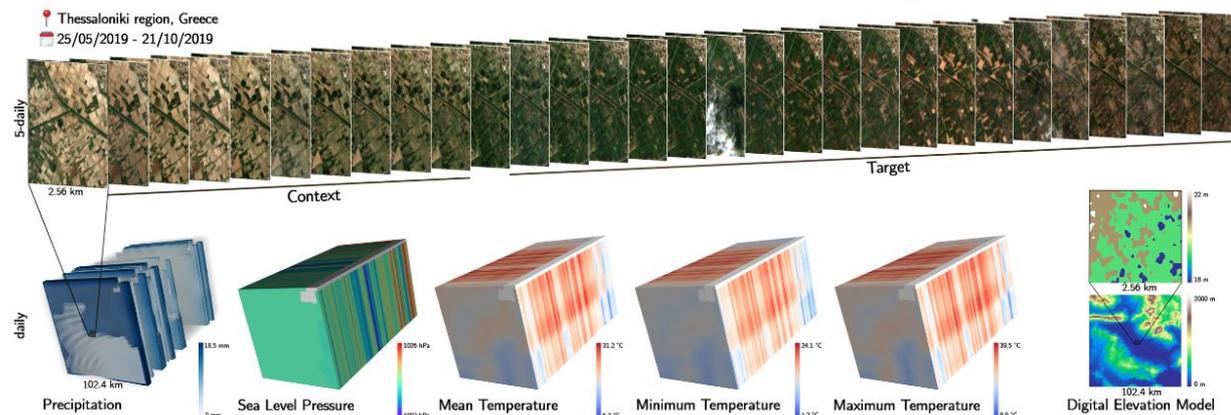
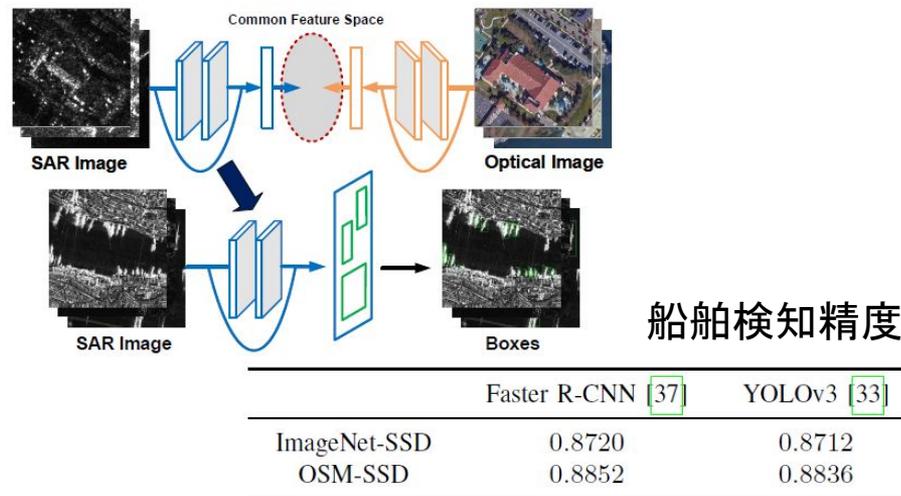
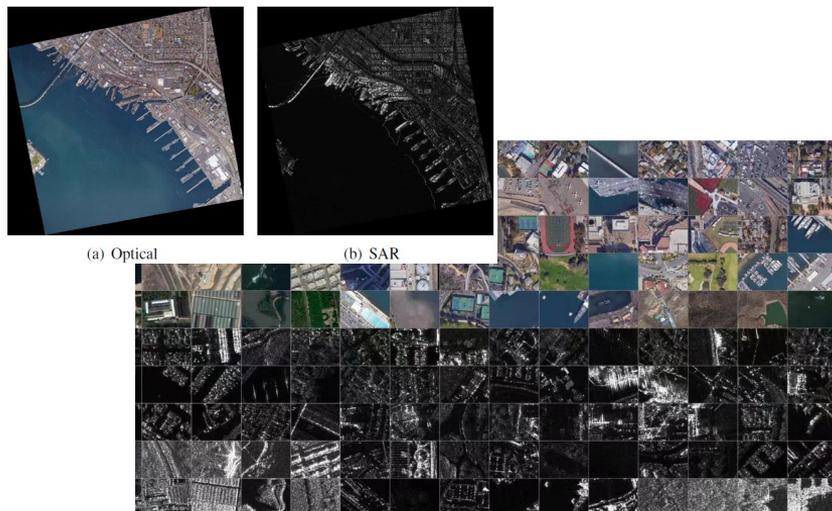


Figure 1: Overview visualization of one of the over 32000 samples in EarthNet2021.

位置合わせを重視した高解像度SAR画像とGoogle Earth画像のペアデータセットを作成

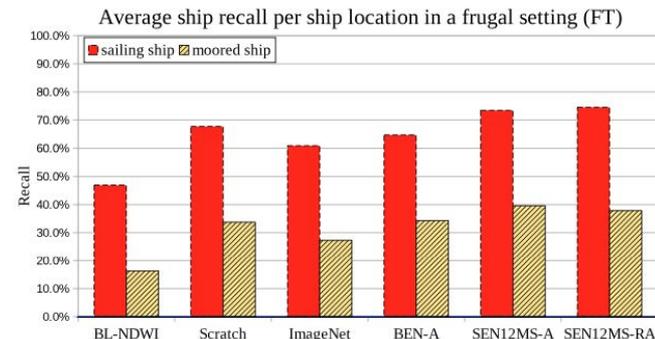
- 高い解像度(1m x 1m)をもつGaofen-3 SAR画像と位置がマッチするGoogle Earth画像を切り出し、20,000の画像ペアを作成。この解像度で256x256サイズ、多様なシーンを含むデータセットは無かった
- ベンチマークとして「SAR-光学ペアの対応画像検索」と「SARによるShip detection」を実施。後者では光学画像をガイドにweightを獲得、SAR単独で高精度なShip detectionができることをデモンストレーション



ラベルを持たないマルチスペクトル画像を用いた船舶の自動検出のためのSSLパイプラインを提案

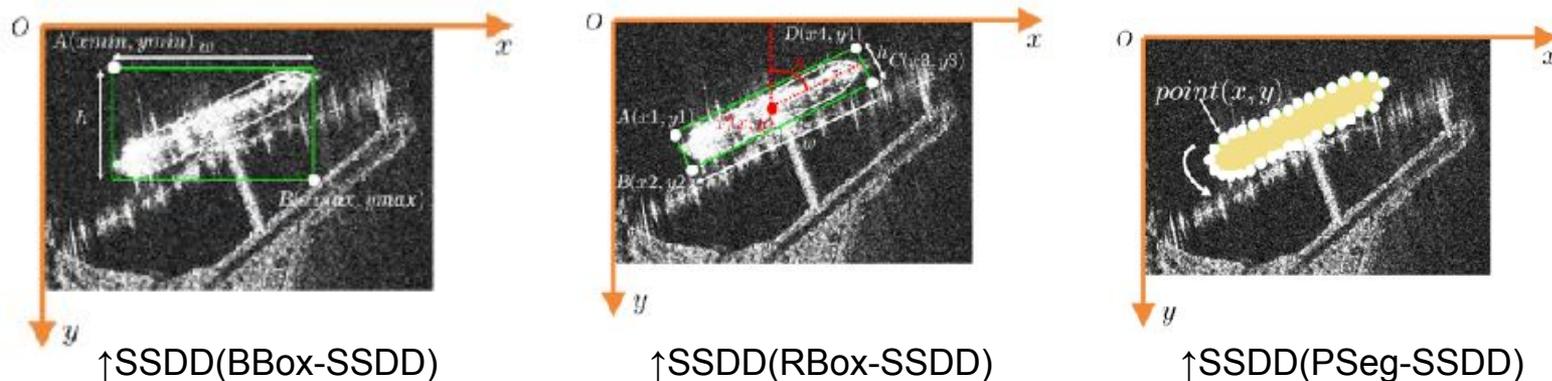
- 船舶検知は、海賊や違法漁業の検知、物流チェーンの監視など、経済情報と海上保安における重要な課題
- マルチスペクトル画像の船舶の微小な違いを区別できる点や光学画像ベースの 公海と港湾のコンテキストが扱える良い点があるが、ラベル付された画像がほとんどない
- そこで本研究ではSelf-supervised learningに焦点をあて、少ない教師で学習することができるかどうかを検討
- SSL手法はMoCoを用いており、Pretext taskとしてデータ拡張のみ(A)や地域ごとの類似性とデータ拡張を用いた方法が検証されている。
- 地域ごとの類似性は対象学習に用いられる画像との地理的距離が近い画像(RA)や遠い画像(RA-lo)で計算される。
- 数ショット学習設定において、自己教師付き事前学習がImageNet事前学習よりも良い結果を達成

Pre-Training Dataset	SSL Pretext Task	Transfer Learning			Fine-Tuning		
		Precision	Recall	F1	Precision	Recall	F1
EuroSAT	A	17.0	80.1	24.7	-	-	-
BigEarthNet	A	19.3	81.5	27.1	18.4	78.1	26.0
SEN12MS	A	21.3	76.7	29.1	22.5	74.3	29.5
SEN12MS	RA	21.9	76.9	29.1	25.2	76.4	33.0
SEN12MS	RA-lo	21.0	77.1	28.4	25.7	77.4	33.0



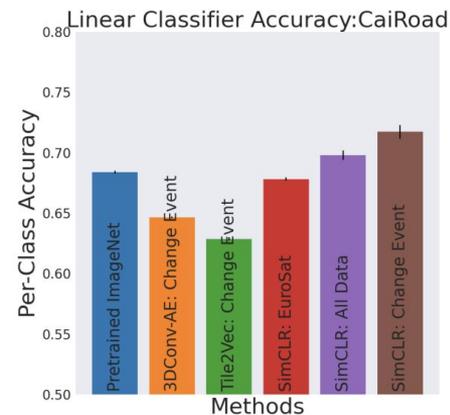
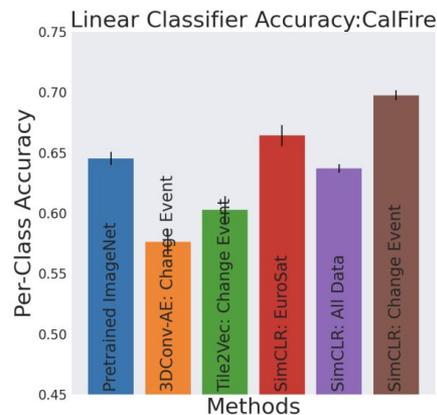
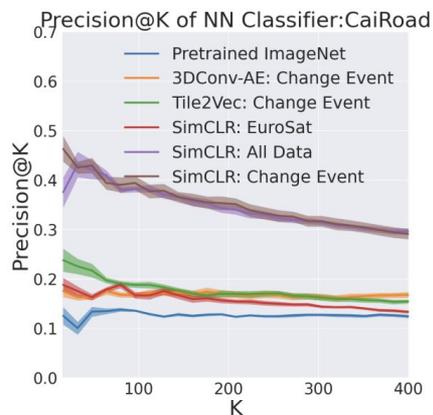
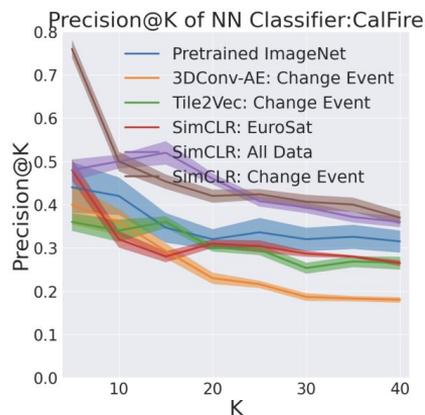
従来のSSDDを修正しDLベースの初めてのSSDDのデータセットを公開

- これまでにSARを用いた船舶検出はいくつか提案されていたが、全ての論文で訓練データの提供がされていなかった。
- さらに、従来の教師ラベルはBB_{ox}やRBox、ピクセルレベルレベルのタスクに対応できない
- そこで本研究では、SSDD(BBox-SSDD)、SSDD(RBox-SSDD)、SSDD(PSeg-SSDD)を提案
- SSDDでは多数の分析により(1)訓練-試験区分の決定、(2)陸上-海上プロトコルの決定、(3)船舶サイズの合理的な定義、(4)密に分布する小型船舶サンプルの決定、(5)港湾に密に並列停泊する船舶サンプルの決定をおこなっている



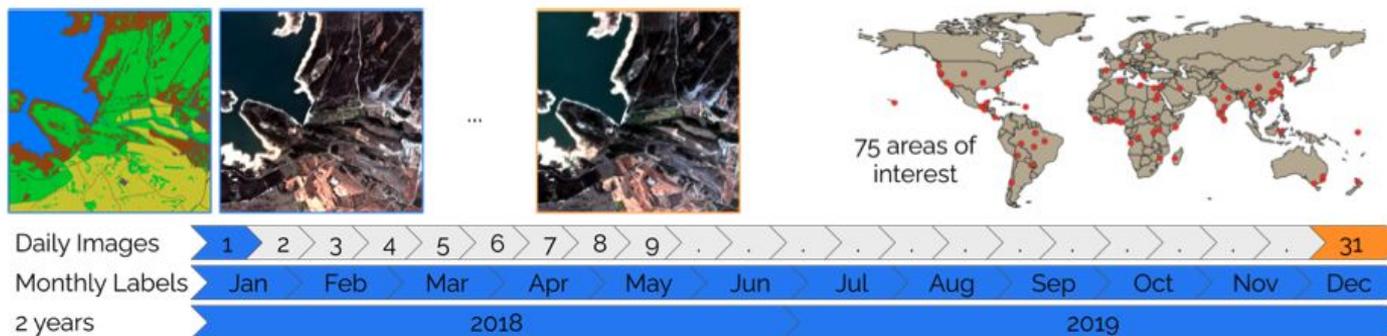
道路建設と森林火災イベントの変化を検出するデータセットを教師なしで作成

- ーSentinel-2衛星画像(空間分解能10m、時間分解能1ヶ月)を用いて作成 **CaiRoad** と**CalFire** と呼ばれるデータセット作成・ベンチマークを計算した。
- CaiRoadデータセットには合計28015個の変化イベントを含む道路建設イベントから構成
- CalFireデータセットには2172個の変化イベントを含む森林火災イベント
- データセット作成ではアノテーションを用いず、衛星画像から取得した時空間データを変化イベントを見つける教師有り学習手法を用いている。



日々更新されるマルチスペクトル画像と月毎にピクセルラベルが更新されるデータセットの提案

- 地球観測は、特定の関心領域における土地利用の変遷を監視するための基本的なツールで、このツールを実現するためには時系列データとピクセル単位のセグメンテーション情報が必要となる
- このために、本研究ではPlanet Labsの画像による、地球上に分布する75の選択された関心領域の毎日のマルチスペクトル衛星観測からなるDynamicEarthNetデータセットを提案
- Dynamic EarthNetは衛星が取得したデータと月毎にセグメンテーションラベルを提供する最初のデータセット
- 実験では日々の観測データを追加学習データとして利用する(半教師付き学習)か、一度に複数の観測データを利用する(時空間学習)、いくつかの確立されたベースラインを提供している

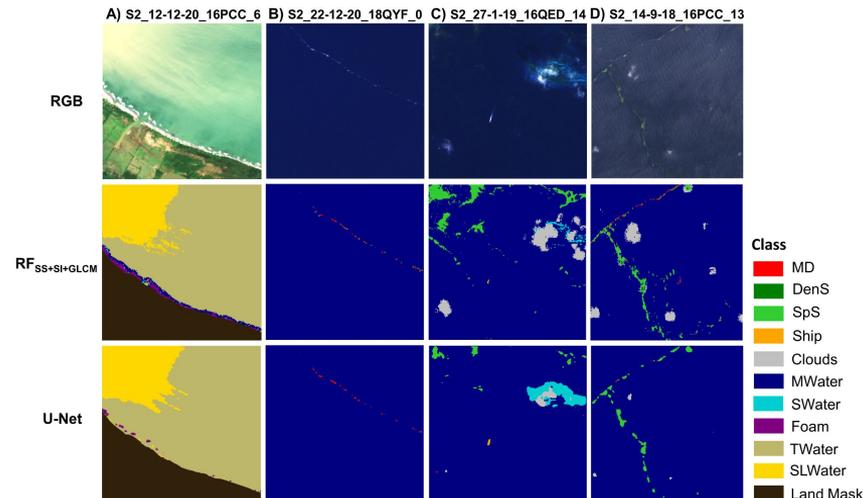


海洋ゴミを光学衛星画像から検出するためのデータセットを公開

- 衛星画像の分光反射特性から海洋ゴミを検出することはリモートセンシング分野で盛んに研究されている
- この海洋ごみを検出するタスクを深層学習モデルで学習し、そのモデルの評価を実行するベンチマークデータセット(MARIDA)を作成し公開した
- 世界中の異なる地域・季節・年・海の条件のSeintinel-2と呼ばれる中解像度の光学衛星画像から、プラスチックごみを検出するタスクである

Class Name	Acronym	Description	Number of Patches
Marine Debris	MD	Floating plastics or other polymers, mixed anthropogenic debris	373
Dense <i>Sargassum</i>	DenS	Dense floating <i>Sargassum</i> macroalgae	49
Sparse <i>Sargassum</i>	SpS	Sparse floating <i>Sargassum</i> macroalgae	106
Natural Organic Material	NatM	Vegetation & Wood	71
Ship	Ship	Sailing & Anchored Vessels	182
Clouds	Cloud	Clouds including thin Clouds	181
Marine Water	MWater	Clear Water	870
Sediment-Laden Water	SLWater	High-Sediment river discharges with brown colour	51
Foam	Foam	Foam recorded at river fronts or coastal wave breaking area	59
Turbid Water	TWater	Turbid waters close to coastal areas	220
Shallow Water	SWater	Coastal waters, including coral reefs and submerged vegetation	64
Waves	Waves	Waves	54
Cloud Shadows	CloudS	Cloud Shadows	71
Wakes	Wakes	Wakes & Waves from a sailing vessel	106
Mixed Water	MixWater	Water near floating materials	140
Total			1381

<https://doi.org/10.1371/journal.pone.0262247.t002>



データセットにおけるクラスの定義

データセットの例

熱赤外バンドで船舶検出用のデータセットを構築

- 熱赤外バンドのデータは薄雲や霧、夜間などにおいてもデータが得られるため、全天候型の船舶の情報収集手段として重要であるが、今までオープンなデータセットは存在していなかった
- SDGSAT-1(中国)の撮影データを利用し、データセットを構築した
- その際センサー間の位置ずれを元解像度90mから30mにアップサンプリングし、修正している
- Faster R-CNN、SSD、Yolo5sなどのFine tuningの試行で所定の精度で検出できており、他にはない熱赤外バンドのデータセットとして有用なものとなっている

※論文ではオープンデータセットとしているがデータセット自体は現在公開されていない



Figure 1. (a) The SAR image with the resolution of 15 m in SSDD dataset [31], The (b) true color image of 460–520 nm, 520–600 nm and 630–690 nm with resolution of 10 m in 38°44′32.74″N,117°50′13.28″ from SDGSAT-1 multispectral imager, and (c) pseudo-color image of 8~10.5 μm , 10.3~11.3 μm 11.5~12.5 μm with resolution of 30 m in 39°18′49.52″N,120°14′55.86″E from SDGSAT-1 TIS.

Table 7. The evaluation criteria of different methods by using the TISD (The bold data are the best and second-best results of different models).

Methods	Bands	Scene	Image Size	Precision	Recall	mAP@0.5	Val_loss	GFLOPs	Parameters
Faster R-CNN [7]	B123	ALL	768 × 768	0.6617	0.6372	0.5998	0.0221	46.7	31.3 M
SSD [8]	B123	ALL	768 × 768	0.6915	0.6791	0.6572	0.0189	19.6	138.0 M
Yolov5s [20]	B123	ALL	640 × 640	0.7668	0.6308	0.6618	0.0096	17.1	7.3 M
The improved Yolov5s	B123	ALL	640 × 640	0.7485	0.6651	0.6378	0.0187	8.2	3.2 M
		Cloud		0.6925	0.4948	0.4983	0.0114		
		River		0.7637	0.5173	0.5444	0.0194		
		Sea		0.8150	0.7132	0.7163	0.0044		
		Day		0.7864	0.6424	0.6768	0.0097		
		Night		0.6165	0.4143	0.4246	0.0119		

船舶の移動に伴って発生する波についてシミュレーションによるSyntheticデータを作成

- 計46,080枚、10種の船舶モデルに対し16方向、6速度、8風向、2風速、3種のSARの入射角でシミュレーション
- 船の速度により発生する波の形状という実計測が難しいパラメータに対して、データセットを作成することで船本体が捉えられない低分解能衛星画像に対しても対応可
- スクラッチからトレーニングしたAlexnetにおいて船種を分類した結果90%以上の精度達成

Figure 1. Simulated SAR images ($\theta_r = 20^\circ$) of ship wake (Passenger Vessel I) with $V_s = 8$ m/s, $D_s = 45^\circ$ and $V_w = 3$ m/s: (a) noise-free I ; (b) with noise I_n ; (c) denoised I_d .

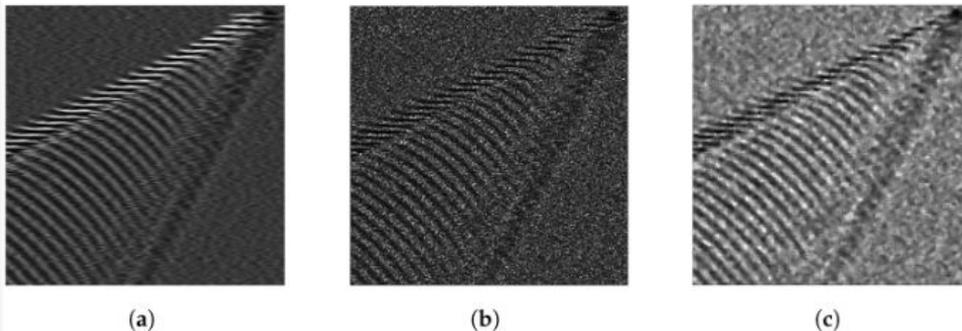


Table 3. Accuracies of the trained CNNs for different datasets including training and test sets.

Network	Training Set			Test Set		
	I	I_n	I_d	I	I_n	I_d
I -Net	98.68	77.35	74.21	96.16	75.90	73.18
I_n -Net	—	97.82	72.24	—	92.70	69.93
I_d -Net	—	50.21	99.16	—	48.79	93.59

モデルに4スケールの層を加えることで小対象物のDetection精度を向上

- It uses a four-scale fusion method to combine the shallow position information with the deep semantic information, effectively adapting to the multiscale target detection in SAR images, significantly improving the detection effect of small targets.
- Finally, to further improve the detection performance of the small-scale SAR aircraft dataset, we propose a domain adaptive transfer learning method
- Experiments on SADD show that this method can significantly improve the recall rate and F1 score. At the same time, we find that the transfer effect of using homologous but different types of targets as the source domain is better than those of heterologous

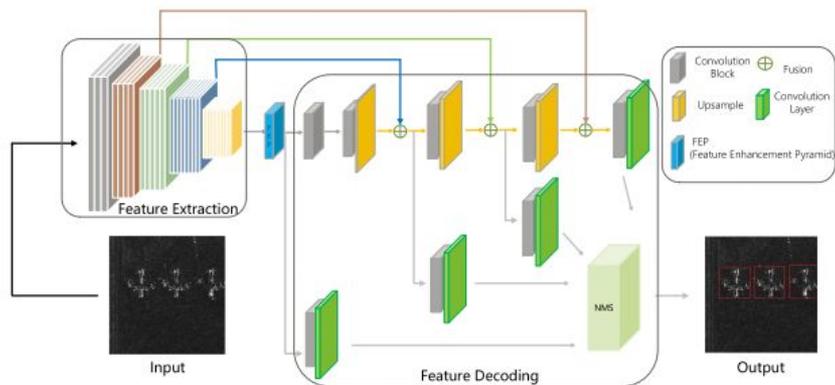


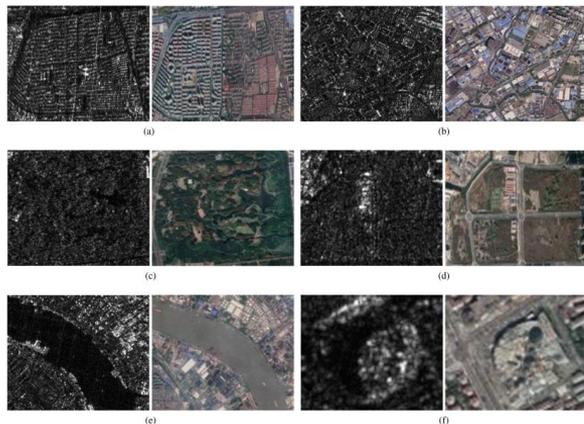
TABLE VI
COMPARISONS TO THE STATE-OF-THE-ARTS

Model	TP	FP	FN	P	R	F1
SSD	802	166	85	0.828	0.904	0.865
YOLOV3	848	148	39	0.851	0.956	0.901
Faster R-CNN	826	121	61	0.872	0.931	0.901
Cascade R-CNN	831	110	56	0.883	0.937	0.909
YOLOX	860	82	27	0.913	0.969	0.940
SEFEPNet	871	106	16	0.891	0.982	0.934

The significance of bold entities indicate best values.

航空機に搭載された偏波SAR向けのセグメンテーションデータセットを構築

- 合成開口レーダ(SAR)を用いて地表の解析を行う際には、照射した電磁波の偏波特性を使用して地物の判読を行っています
- カテゴリは工業エリア・自然エリア・土地利用・水面・家屋・その他の分類を行うセグメンテーション用のデータセットを作成
- 古典的機械学習手法や、深層学習を利用したセグメンテーション手法を適用し、性能比較を行った



データの例

Method	Terrain Category / IoU(%)						Evaluation Metric			
	Industrial Area	Natural Area	Land Use	Water	Housing	Other	$mIoU(\%)$	$OA(\%)$	$AA(\%)$	$Kappa(\%)$
SIFT [45]+SVM [48]	29.98	55.11	0.10	41.08	57.36	18.60	33.70	63.71	51.52	46.86
SIFT [45]+RF [49]	29.73	52.36	0.14	47.58	54.71	16.00	33.42	62.40	52.32	45.24
HOG [46]+SVM [48]	30.39	55.68	0.10	41.89	57.83	19.01	34.15	64.19	51.97	47.53
HOG [46]+RF [49]	33.89	57.98	0.29	47.13	59.98	19.65	36.49	67.45	55.83	51.89
LBP [47]+SVM [48]	34.39	57.50	0.17	34.46	59.57	19.05	34.19	66.14	53.77	50.38
LBP [47]+RF [49]	35.35	58.85	0.18	36.82	60.42	20.06	35.28	67.21	54.79	51.81
FCN [50]	37.78	71.58	1.24	72.76	67.69	39.05	48.35	76.28	57.84	63.43
ANN [53]	41.23	72.92	0.97	75.95	68.40	56.01	52.58	77.46	64.97	65.73
PSPNet [51]	33.99	72.31	0.93	76.51	68.07	57.07	51.48	76.55	63.81	63.88
DeepLab V3+ [52]	40.62	70.67	0.55	72.93	69.96	34.53	48.21	76.81	63.55	64.92
EmaNet [57]	36.92	72.49	1.22	69.72	67.98	37.05	47.56	76.36	60.30	63.87
CCNet [54]	32.54	72.27	0.49	72.58	66.75	34.12	46.46	75.53	55.83	62.06
DANet [55]	39.56	72.00	1.00	74.95	67.79	56.28	51.93	76.91	62.79	64.79
NonLocal [56]	35.51	72.12	2.47	70.60	68.39	16.31	44.23	76.05	53.90	63.23
EncNet [58]	32.95	71.59	1.89	75.66	67.16	37.24	47.75	75.67	57.51	62.42
PSANet [59]	40.70	69.46	1.33	69.46	68.75	32.68	47.14	76.21	62.92	63.95
GCNet [60]	38.19	72.64	2.43	74.48	68.37	29.24	47.56	76.75	57.10	64.30

セグメンテーションの結果

太陽光発電のソーラーパネルを抽出するデータセットを構築

- 持続可能な発電方法で太陽発電が用いられるが、住宅用の太陽光発電が普及するに連れて発電と送電を安定して供給するのが困難になりつつある
- 太陽光発電の台帳はあるはずだが、データにアクセスできないため、航空写真を用いて太陽光発電の設置状況を把握することが重要になる
- 航空写真から家屋の屋根において太陽光発電が設置されている箇所を特定するセグメンテーション用のデータセットを構築し、ベースライン評価を行った

Table 4: Segmentation baselines for the SolarDK testset. First group of runs were pre-trained using COCO train2017 and Pascal VOC, while the bottom section was pre-trained on German data [10]. The model with best mean performance(s) is marked with bold-face font.

Model	Recall	Precision	IoU
ResNet50-DeepLabV3Plus	0.81±0.03	0.86±0.01	0.72±0.02
ResNet101-DeepLabV3Plus	0.79±0.05	0.86±0.02	0.70±0.03
ResNet152-DeepLabV3Plus	0.79±0.04	0.88±0.03	0.71±0.02
ResNet50-FPN	0.80±0.03	0.87±0.03	0.72±0.01
ResNet101-FPN	0.79±0.02	0.87±0.02	0.71±0.01
ResNet152-FPN	0.81±0.06	0.87±0.05	0.72±0.01
ResNet50-PSPNet	0.75±0.04	0.85±0.03	0.64±0.04
ResNet101-PSPNet	0.66±0.13	0.88±0.05	0.61±0.07
ResNet152-PSPNet	0.72±0.05	0.85±0.04	0.63±0.02
DeepSolarDE (inference)	0.53	0.34	0.51
DeepSolarDK	0.85	0.75	0.62

太陽光発電のセグメンテーションの定量評価

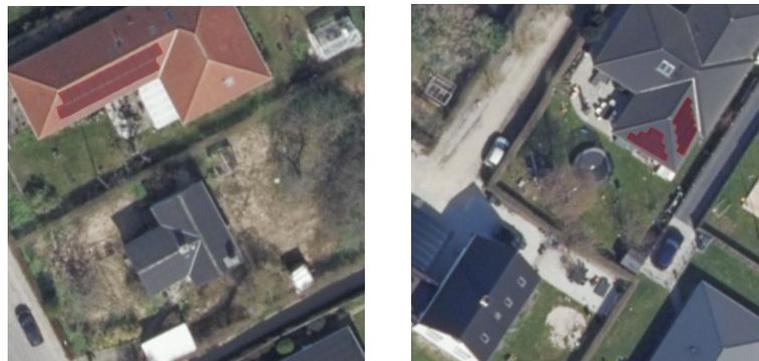
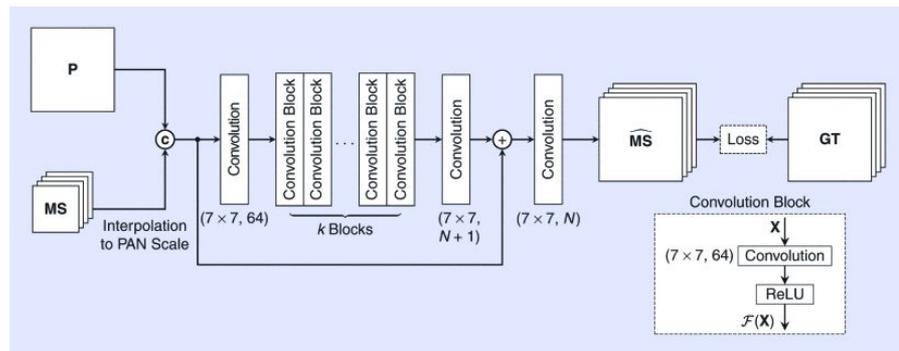


Figure 2: Examples of segmentation masks (indicated by red).

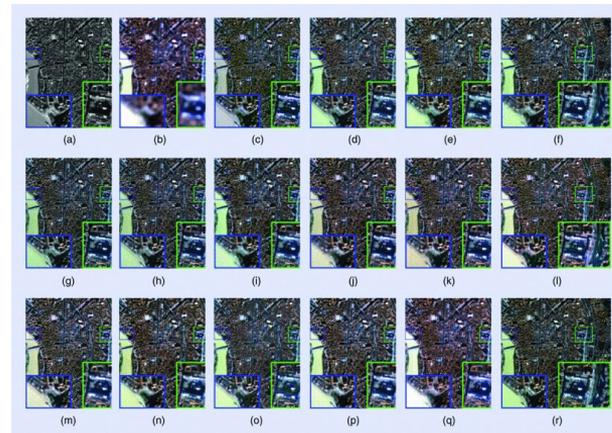
太陽光発電のセグメンテーションマスクの例

パンシャープニングを機械学習で実行・評価するためのデータセットを構築

- リモートセンシング分野では、高解像なパンクロマティック画像と、低解像な多バンド画像を組み合わせ、高解像なカラー画像を作成するパンシャープニング処理が行われている
- 深層学習を用いたパンシャープニング処理を行う場合、一般的にはパンクロマティック画像と多バンド画像を深層学習モデルを入力して画像変換タスクとして解きます
- 統一的に機械学習ベースのパンシャープニング処理を評価するためにデータセットを構築して公開した



深層学習によるパンシャープニング処理の例

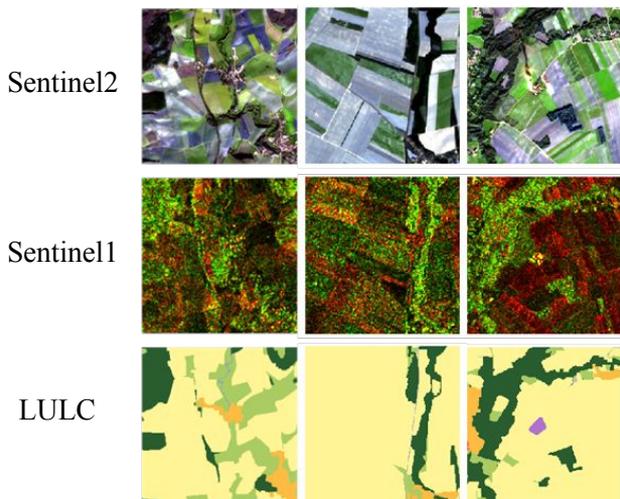


パンシャープニング処理

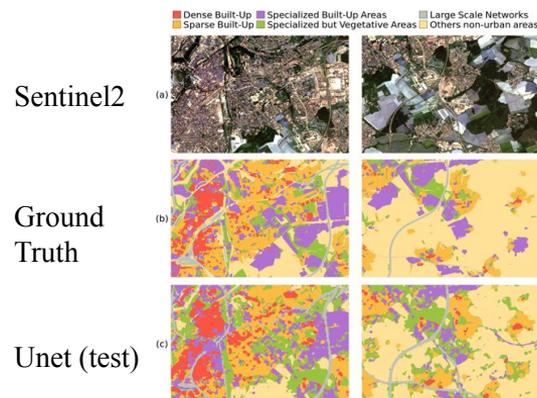
MultiModal(SAR+光学)かつMutliTemporalな衛星画像と 良好な解像度のセグメンテーションマップを組み合わせたデータセットを作成

- 解像度10mのSentinel 1 (SAR), Sentinel 2(光学)のセットに、解像度50m相当のセグメンテーションマップ(土地利用・土地被覆図: LULC)を組み合わせたデータセット。過去に提案されていたSEN12MSではLULCは解像度500mで衛星画像との直接的な比較は難しかった。
- NN?で解像度10mにそろえ、256 x 256 pixelサイズの8,157のパッチを整備
-

MultiSenGE Multi-temporal: Georeference 情報を付記しユーザーが自身で準備できるように



Unetで[IR,R, B] vs [IR,R, B+Index]
=> [IR, R, B]のほうが良い学習



源泉が異なる(SAR/光学, 異なる解像度)が、同じ場所で2時期の画像をそれぞれそろえ、新規建築を検知するタスクの学習に役立つデータセットを作成

- 高解像度画像に中分解能画像を組み合わせるとより高精度な変化検知が可能。このアイデアを基に「Build-up area」に着目した新規データセットを整備。
- 3つの異なる衛星(高分解能:GF-2, SAR: Sentinel1, マルチバンド: Sentinel2)の画像から場所をそろえて変化の前・後のパッチを切り出し、変化ラベルを付与したデータセット(MSBC、MSOSCD)を整備。
- 異なる画像種別の情報をFeatureで統合するU-Net+Siamese構造モデルにより変化検知タスクをテスト、単一の高解像度画像を用いた検知より複数衛星画像を組み合わせた方が高精度となることを確認。

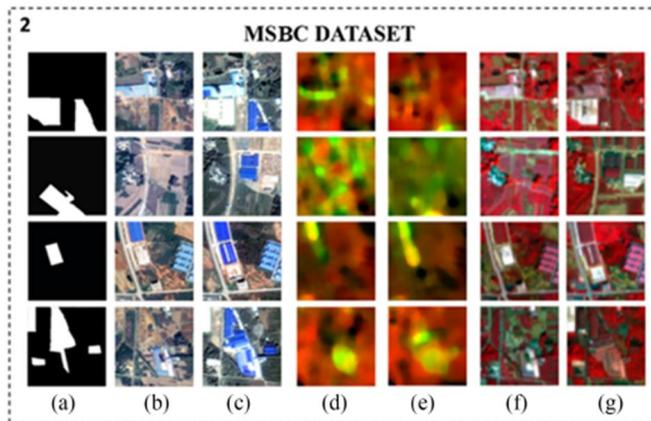


TABLE IV[†]
DIFFERENT FEATURE COMBINATION EXPERIMENT RESULTS WITH MSDCUNET
ON TWO DATASETS[†]

Input	MSOSCD(test)		MSBC(test)	
	IOU(%)	F1(%)	IOU(%)	F1(%)
RGB	57.15	69.85	34.12	55.32
RGB+S1	78.56	87.99	48.22	65.07
RGB+S2	83.47	90.99	47.35	64.27
RGB+S1*	85.51	91.25	44.49	61.58
RGB+S2*	84.30	91.31	44.29	61.39
RGB+S1S2	86.12	92.89	47.68	64.57
RGB+S1*S2*	87.05	93.1	49.00	65.77

Semantic segmentationと季節の変動に強いLULCデータセットの提案

- 衛星画像は日々取得されており土地利用や気候変動の監視、災害防止には衛星画像を用いて土地被覆分類することが重要
- しかし、深層学習を用いて高い性能を出そうと思うと大規模データセットが必要となり、従来のBigEarthやSEN12MSなどのデータセットは特にSemantic segmentationのラベル付きデータが少ない
- さらに、従来のデータセットは季節の変動の学習をするために様々季節の衛星画像を含むようなデータセットを作成しているが、同じ位置で季節が異なる衛星画像データセットは作成されていない
- そこで本研究では、これらを克服するためのデータセットを提案した

Dataset name	Image Type	Annotation Type	Number of Images
Agriculture-Vision [9]	Aerial Multispectral	Multi-Label + Pixel Label	94 986
MLRSNet [10]	Aerial/Sat RGB	Multi-Label	109 161
BigEarthNet [2]	Sat. Multispectral	Multi-Label	590 326
SEN12MS [3]	Sat. Multispectral	Multi-Label + Region Label	541 986
This Work	Sat. Multispectral	Multi-Label + Pixel-Label	1 759 830

Image-level labels	Pixel-level labels	Spring (460 104)	Summer (481 456)	Fall (500 149)	Winter (218 663)	Snow (99 458)
Road and rail networks and associated land, <i>(Non-irrigated) arable land</i> , Pastures, Broad-leaved forest						
Discontinuous urban fabric, Industrial or commercial units, <i>(Non-irrigated) arable land</i> , Pastures, Coniferous forest, Mixed forest, Water bodies						
Discontinuous urban fabric, Road and rail networks and associated land, Construction sites, Pastures, Broad-leaved forest, Coniferous forest, Mixed forest						

SAR、マルチスペクトル、ハイパースペクトル、DSM、GISデータを含むマルチモーダルデータセットMDASを作成

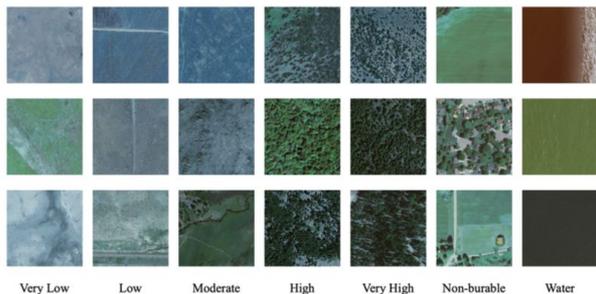
- エリア : Ausberg, Germany. 時期 : 2018年5月7日
- 5つのモダリティを提供
- super resolution、spectral unmixing、land cover classificationの三つのタスクで代表的な手法をMDAS上で検証

Data	Modality	Sensor
Sentinel-1	SAR	Sentinel-1 payload
Sentinel-2	Multispectral image	Sentinel-2 payload
DSM	DSM	DLR 3K
HySpex	Hyperspectral image	HySpex
S2eteS_S2	Multispectral image	S2eteS Spatial and spectral simulation
EeteS_EnMAP_10	Hyperspectral image	EeteS Spatial and spectral simulation
EeteS_EnMAP_30	Hyperspectral image	EeteS Spatial and spectral simulation
EnMAP	Hyperspectral image	EeteS
GIS	GIS	Open street map (OSM)
Endmember		Manual labeling
Land cover maps		Manual labeling



地球表面のFireRiskを計算するためのデータセットを作成

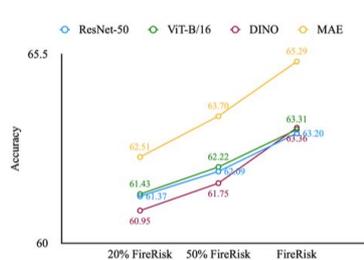
- 本研究の目的はリモートセンシング画像 と地球表面の火災リスク との間をマッピングを記述する簡単なスキームを開発すること
- FireRiskはWHP (Wildfire Hazard Potential) ラスターデータセットによって提供された火災リスククラスでラベリングされており、高解像度リモートセンシング画像のNAIP (National Agriculture Imagery Program) を用いて収集されている
- 実験では教師あり表現と自己教師あり表現のベンチマーク性能を示し、ImageNet1kで事前に訓練されたMasked Autoencoders (MAE) [16]が65.29%という最高の分類精度を達成している



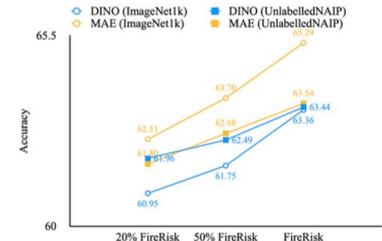
収集された画像セット森林被覆と火災危険度が相関があることが感じられる

Dataset	Model	pre-trained	Acc.	F1
FireRisk	ResNet [17]	ImageNet1k [8]	63.20	52.56
	VIT [10]	ImageNet1k [8]	63.31	52.18
	DINO [5]	ImageNet1k [8]	63.36	52.60
	DINO [5]	UnlabelledNAIP	63.44	52.37
	MAE [16]	ImageNet1k [8]	65.29	55.49
	MAE [16]	UnlabelledNAIP	63.54	52.04
50% FireRisk	ResNet [17]	ImageNet1k [8]	62.09	50.27
	VIT [10]	ImageNet1k [8]	62.22	50.07
	DINO [5]	ImageNet1k [8]	61.75	51.21
	DINO [5]	UnlabelledNAIP	62.49	51.35
	MAE [16]	ImageNet1k [8]	63.70	50.23
	MAE [16]	UnlabelledNAIP	62.68	52.05
20% FireRisk	ResNet [17]	ImageNet1k [8]	61.37	49.53
	VIT [10]	ImageNet1k [8]	61.43	48.80
	DINO [5]	ImageNet1k [8]	60.95	50.72
	DINO [5]	UnlabelledNAIP	61.96	50.83
	MAE [16]	ImageNet1k [8]	62.51	51.13
	MAE [16]	UnlabelledNAIP	61.80	50.07

ImageNet1k MAE事前学習、FireRiskFientuningが最高精度



(a) Performance of models on different sizes of FireRisk.

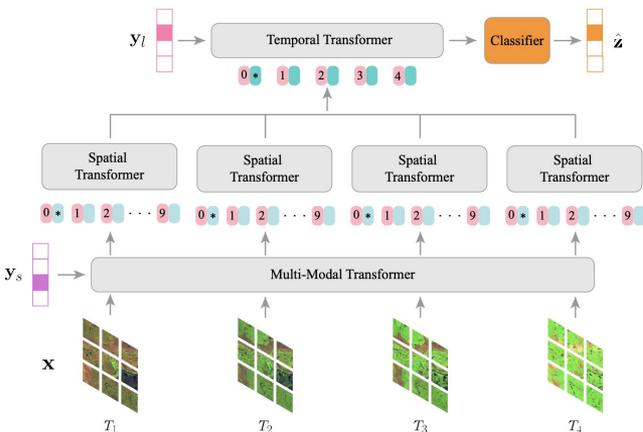


(b) Accuracy differences between different latent representations of the self-supervised benchmark.

modelとself-supervised learningに共通してデータセットが多い方が精度が高い傾向に

タイムリーな作物・気象データ・Sentinel2を用いて作物の収穫量を予測モデルを提案

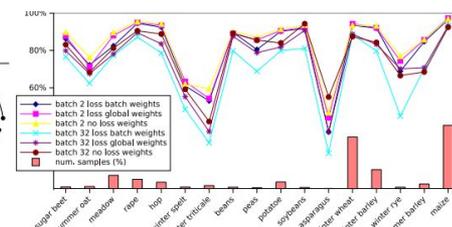
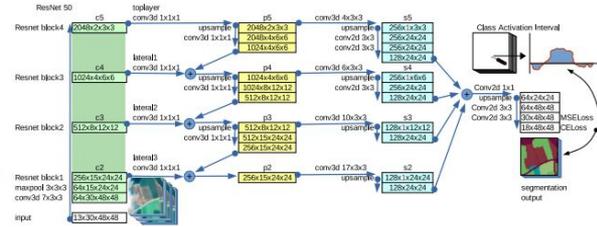
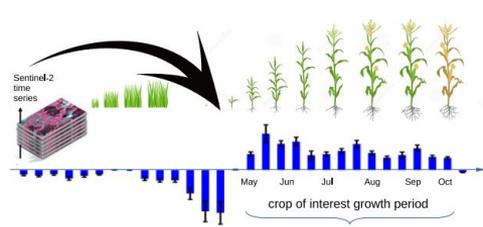
- 作物の生育は生育期の天候変動や気候変動の影響を受けやすいため、作物の収量をタイムリーに予測することは依然として困難
- そこで、DLベースの手法として Multi-Modal Spatial-Temporal Vision Transformer(MMST-ViT)を提案し、生育期の短期的な気象変動と長期的な気候変動が作物に及ぼす影響を考慮することで、全米の郡レベルの作物収量を事前予測を行う
- また、作物データ、タイムリーな気象データ、Sentinel2のデータを用いたTiny Cropを用いて学習
- 広範囲の実験の結果、提案法は関心のある3つの性能指標において、同等モデルを上回った。



Method	Corn			Cotton			Soybean			Winter Wheat		
	RMSE (↓)	R ² (↑)	Corr (↑)	RMSE (↓)	R ² (↑)	Corr (↑)	RMSE (↓)	R ² (↑)	Corr (↑)	RMSE (↓)	R ² (↑)	Corr (↑)
ConvLSTM	18.6	0.611	0.782	65.4	0.715	0.846	7.2	0.616	0.785	7.4	0.511	0.715
CNN-RNN	14.6	0.705	0.839	69.5	0.653	0.808	5.8	0.703	0.839	7.5	0.614	0.783
GNN-RNN	14.2	0.730	0.854	58.5	0.647	0.804	5.4	0.748	0.865	6.0	0.621	0.788
Ours	10.5	0.811	0.900	42.4	0.790	0.889	3.9	0.843	0.918	4.6	0.785	0.886

作物が「いつ、どこで」育てられたのかに関する情報を抽出可能なCNNモデルの開発

- Sentinel-2より得られた作物の画像の各ピクセルに対して, CNNによって分類の推論を解釈し, 入力タイムシリーズのどの時間間隔においてクラスが存在する可能性が高いかを決定することができるClass Activation Interval (CAI) というモデルを提案.
- 異なるスケールのオブジェクトを検出するのに有効なフィーチャピラミッドネットワークを, 入力時系列に3Dカーネルを使用しネットワーク全体に3Dカーネルを使用するモデルに拡張.
- このモデルによって得られた稲作データセットは, 不均一なデータセットを中和することを目的と曾田バッチサイズを変更する等の処理を加える必要がない, という結果を得た.



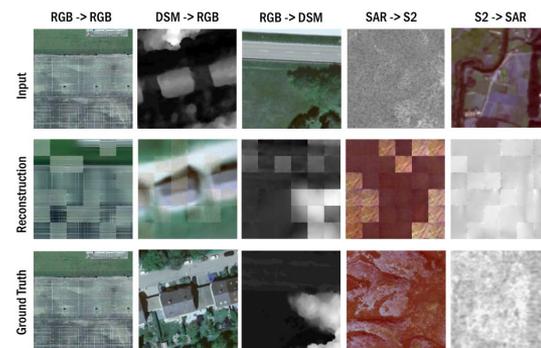
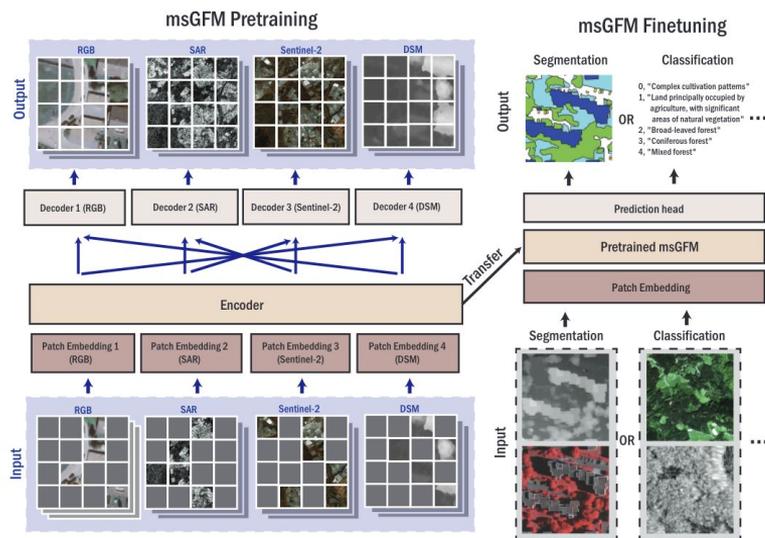
画像セグメンテーション用にトレーニングされた完全量込みCNNによって, 入力時系列のクラスの存在を決定する時間間隔を特定

ResNet50を利用したCovNetのフィードフォワード計算と各カーネルサイズ・クロスエントロピーと平均二乗誤差の和で定義される損失関数を最小にする.

ミュンヘンデータセットでバッチサイズが変化するにつれて3Dフィーチャピラミッドネットワークの適合率がどう変わるか検証.

マルチセンサー事前学習にを用いてより包括的な衛星画像表現を獲得

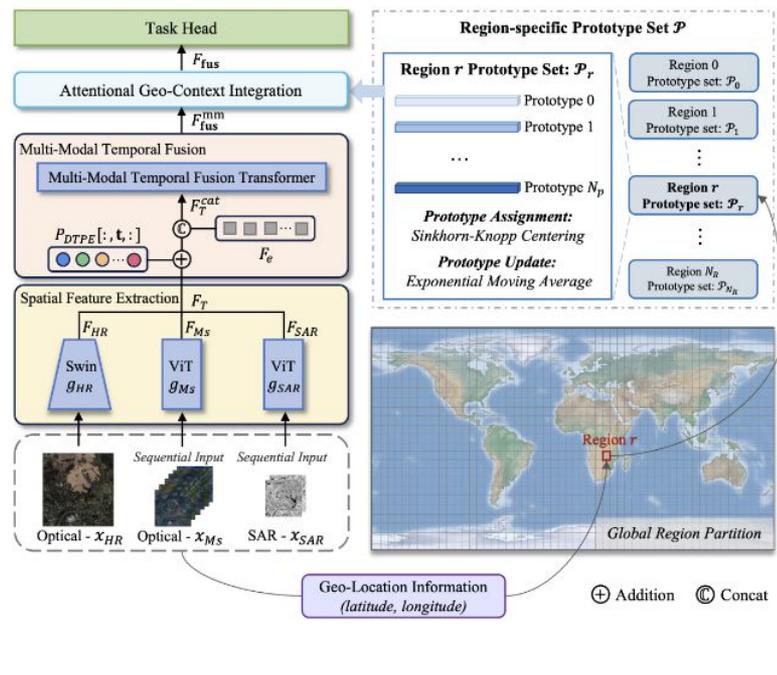
- MAEに基づいたマルチセンサー (RGB, SAR, Sentinel-2, DSM) 事前学習手法を提案。
- シングル・マルチセンサータスクに適用可能な基盤モデルを構築。
- データセットは既存のpaired/unpairedデータセットを組み合わせたハイブリッド (2M)
- マルチセンサー表現は自然画像から蒸留するよりもscratchでトレーニングした方が高精度



Dataset	# Images	GSD	Sensor modality	# Channels	paired sensors?
GeoPile [39]	600K	0.1m - 30m	RGB ^a	3	✗
MillionAID [37]	1M	0.5 - 153m	RGB ^a	3	✗
SEN12MS [52]	320K	10m	SAR / sentinel-2	2/14	✓
MDAS [29]	40K	0.1m - 10m	DSM/ RGB ^b	1/3	✓

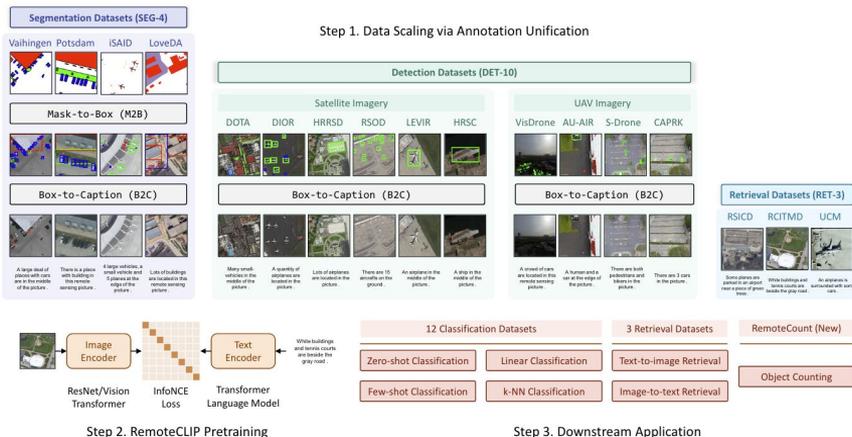
シングル・マルチモーダル、単・多時期、分類・ローライゼーション全てに適用できる基盤モデルを提案

- 高解像度光学+光学+SARのマルチモーダル・多時期学習を実現
- モーダル別エンコーダで多様な下流タスクに応用可能
- メタデータをGeo-contextとして学習に使用
- 21.5M多時期画像を使い2Bパラメータのモデルを構築
- 16データセット上の7つのタスクでSoTA達成



既存の衛星とUAV画像のデータセットを利用しリモセン版のCLIPを提案

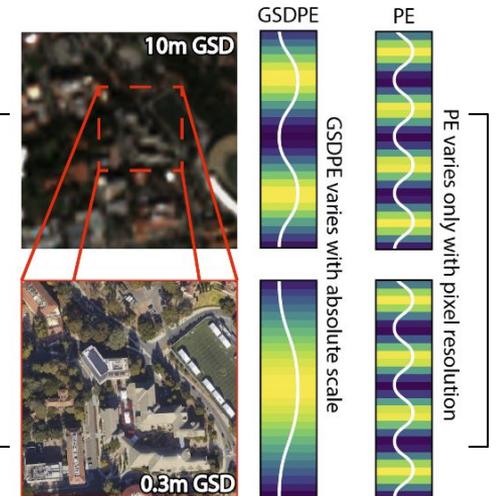
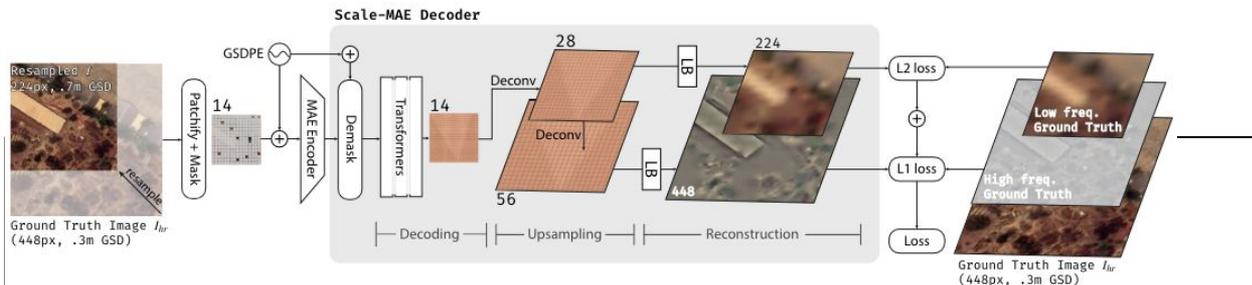
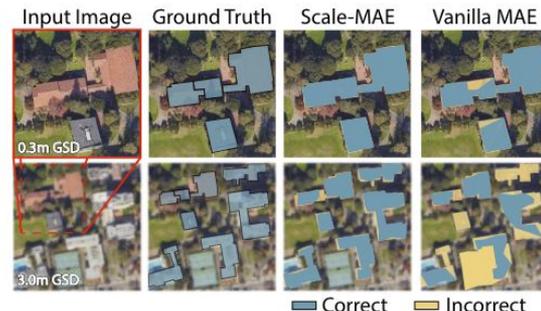
- 既存の17のデータセットを使用
- 種類の違うアノテーションをCLIPが学習できるimage-caption形式に変換し統一
- 新しい物体カウント用データセットRemoteCountを作成
- リモートセンシング用の基盤モデルRemoteCLIPを提案
- 16のデータセット上でベースラインの基盤モデルより高精度
- 12のタスクでオリジナルのCLIPより平均6.39%の性能向上



	Dataset	Year	#Image	#Class	#Box	Avg. Res.
RET-3	RSICD [29]	2017	8483	-	-	224×224
	RSITMD [28]	2021	3603	-	-	256×256
	UCMerced [30]	2018	1676	-	-	256×256
DET-10	AUAIR [85]	2020	32,823	8	132,031	1920×1080
	CARPCK [86]	2017	1,568	1	106,690	1280×720
	DIOR [76]	2019	23,463	20	192,472	800×800
	DOTA [75]	2017	1,409	15	98,990	1504×1395
	HRRSD [77]	2019	21,761	13	57,137	1406×1264
	HRSC [80]	2017	1,055	1	1,055	1105×791
	LEVIR [79]	2020	37.91	3	11,028	800×600
	RSOD [78]	2021	936	4	7,400	1051×900
	Stanford [87]	2016	17,351	6	355,443	1424×1088
	Visdrone [88]	2018	6,471	11	77,547	1509×849
SEG-4	iSAID [83]	2019	30,821	15	987,239	896×896
	loveDA [84]	2021	4,187	6	97,989	1024×1024
	Potsdam [82]	2012	5,421	4	92,161	512×512
	Vaihingen [81]	2012	742	4	16,875	512×512

地上解像度を考慮したMAEに基づく事前学習手法を提案

- 衛星の仕様によっては衛星画像の1ピクセルは0.3mから1kmまで変動するため、地上解像度(GSD)を考慮したpositional encodingをMAEのエンコーダーに導入
- マルチスケールの表現を獲得するため、Laplacian Pyramidに基づいたデコーダーをMAEに導入
- kNN分類評価において8のデータセットでSoTAを達成

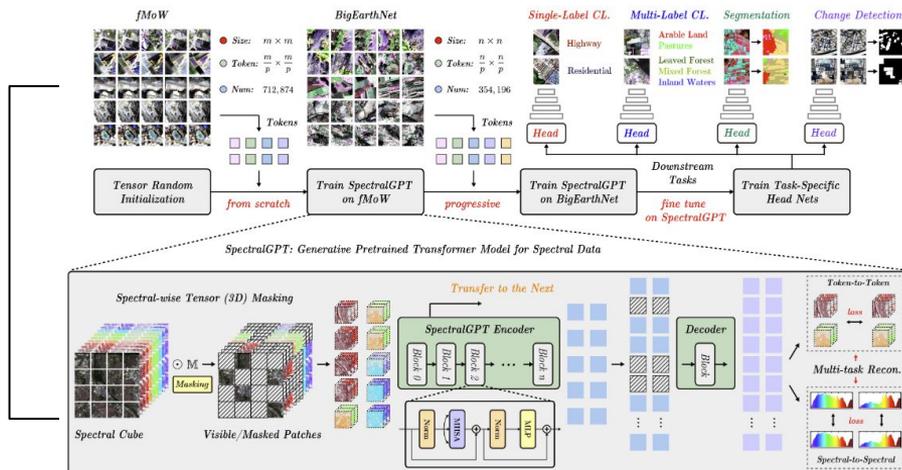
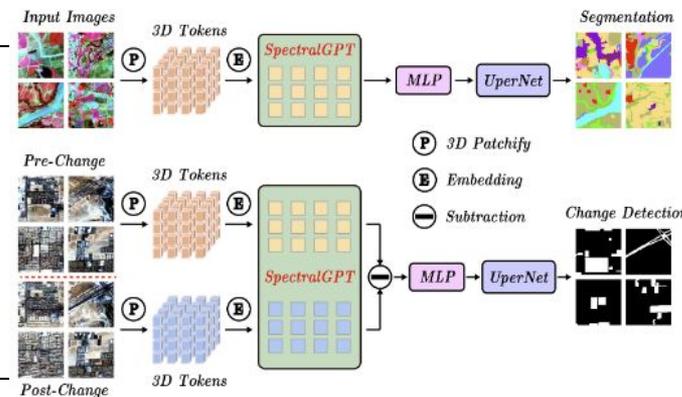
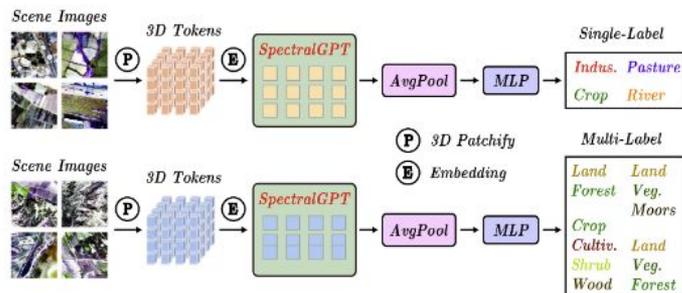


Method	GSDPE	KNN 50%	KNN 100%
Vanilla MAE		72.8	77.8
Vanilla MAE	✓	75.4	78.5
MAE + LP		75.3	79.6
Scale-MAE	✓	78.1	80.7

Dataset	Average Accuracy (%)		
	Scale-MAE	SatMAE	ConvMAE
AiRound	63.2	57.8	59.7
CV-BrCT	69.7	66.2	68.4
EuroSAT	86.7	84.4	88.8
MLRSNet	81.7	75.0	79.5
OPTIMAL-31	65.5	55.7	61.7
RESISC	70.0	61.0	67.0
UC Merced	75.0	69.8	70.0
WHU-RS19	79.5	78.5	77.0

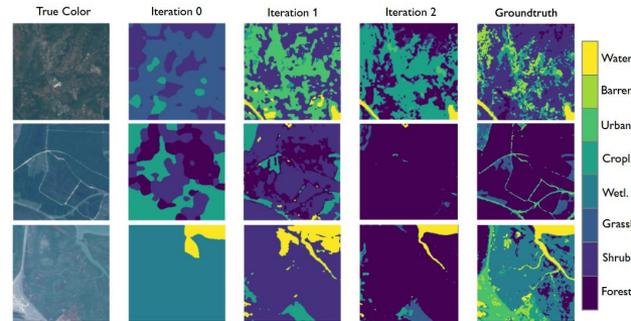
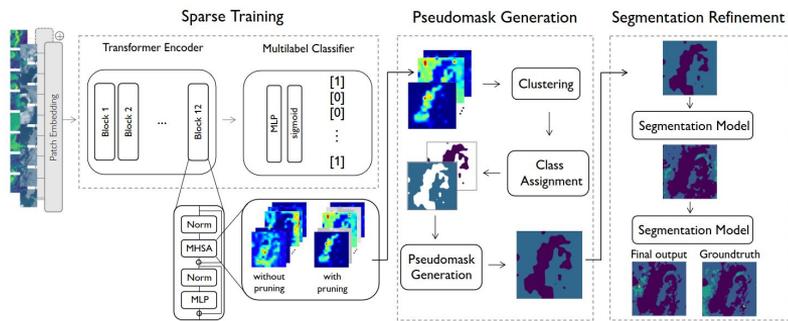
スペクトルデータに対応する事前学習手法と基盤モデルを提案

- スペクトルデータに対応した3D cube maskingとspectral-to-spectral reconstructionを提案
- progressive training により多種多様の画像サイズに対応
- 1M+ 衛星画像によって訓練した100M, 300M, 600Mの基盤モデルを構築
- Semantic Segmentationのための新しいデータセット(SegMunich)を提案
- single/multi-label scene classification (EuroSat, BigEarthNet), semantic segmentation (SegMunich), change detection (OSCD)のタスクにてSoTAを達成



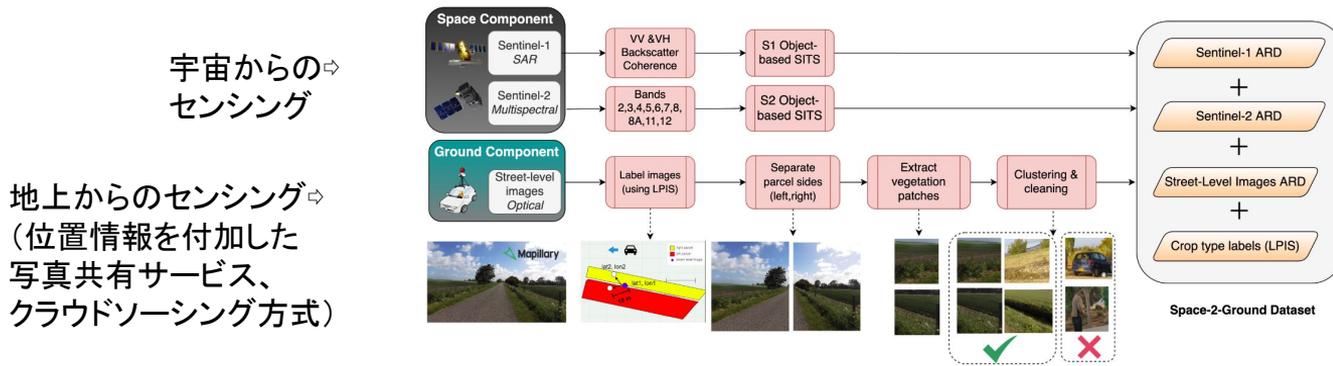
画像単位のラベル情報から、地表のセグメンテーションを行う弱教師あり学習手法

- ViTベースモデルで特徴抽出における各MHSA(Multi Head Self Attention)に対してGating unitsを乗算することで、ラベル分類に重要でないヘッドをpruning
これにより、パラメータのスパース性を確保し、分類に重要な特徴のみ抽出を可能にした
- 領域分割では、pruningされていないヘッドをk-meansでクラスタリング
各クラスタの平均画像を2値化し、入力画像をマスクする(pseudomask作成)
さらに、pseudomaskを教師とし、Unetを学習することで、領域分割を可能にした
- 2つのSAR画像データによるマルチモーダルな学習で、一般の教師あり学習Unet と同等の領域分割精度を達成



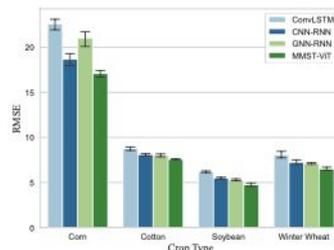
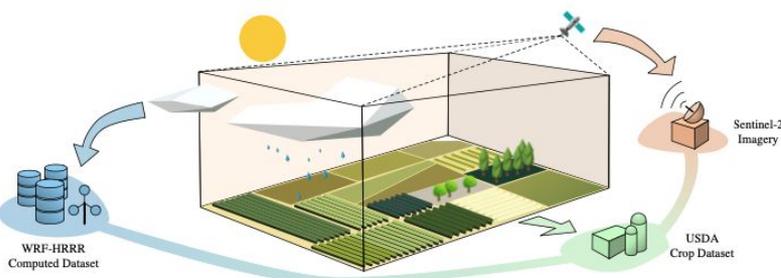
宇宙と地上の異なるモダリティで作成されたデータを統合した農業用データセット作成

- EU地域において、広大な土地での農業を安全・持続可能にするために、CAP補助金が導入されており、分配をデータに基づいて行う必要性があった
- Sentinel 画像(10-60m分解能)では精度が不十分なため、google streetやLPISデータセットなど地上で撮影した画像データセットを利用し、マルチモーダルによるラベル情報(草原や穀物の種類)の高精度化を提案
- データセット作成に使用したデータは地理情報が含まれているため、それらを統合
 - 地上での撮影画像はアングル情報をもとにSentinel 画像とのラベルを統合
 - クラスタリング・PCAにより、Sentinel や LPIS データセットの荒いラベル情報を多量の地上撮影画像へ高精度に伝搬させた

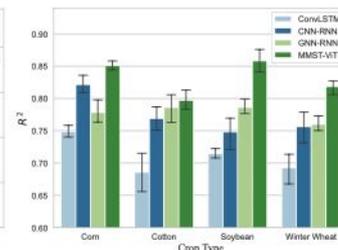


気象と育成状況の情報から作物量を推定するためのデータセットを提案

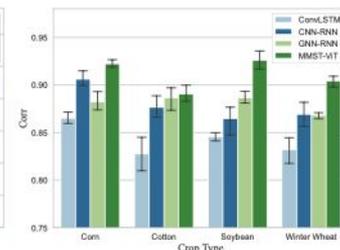
- 正確な農作物収穫量の予測は、食糧安全保障と持続可能な農業を確保するために国家的に重要
- しかし、これまでの深層学習データセットは(1)地上の作物収穫量の情報がないため作物収穫量の予測にてきていなかった。(2)作物収穫量には作物の成長の追跡や気象学的な天候変動の影響を同時に捉える必要があり複数のデータを用いる必要がある。
- 本研究では気象の変動と作物の育成状況、作物の収穫量の3つのモダリティを扱うCropNetと呼ぶデータセットを提案
- 実験結果により、CropNetの適用可能性および有効性が示された。
- PyPIでCropNetのパッケージも公開されている



(a) RMSE



(b) R^2



(c) Corr