サトラスプレトレインリモートセンシング画像理解のための大規模データセット

Favyen Bastani Wolters Gupta Joe Ferdinando Aniruddha Kembhavi
Allen Institute for AI
{favyenb,piperw,ritwikg,joef,anik3}@allenai.org

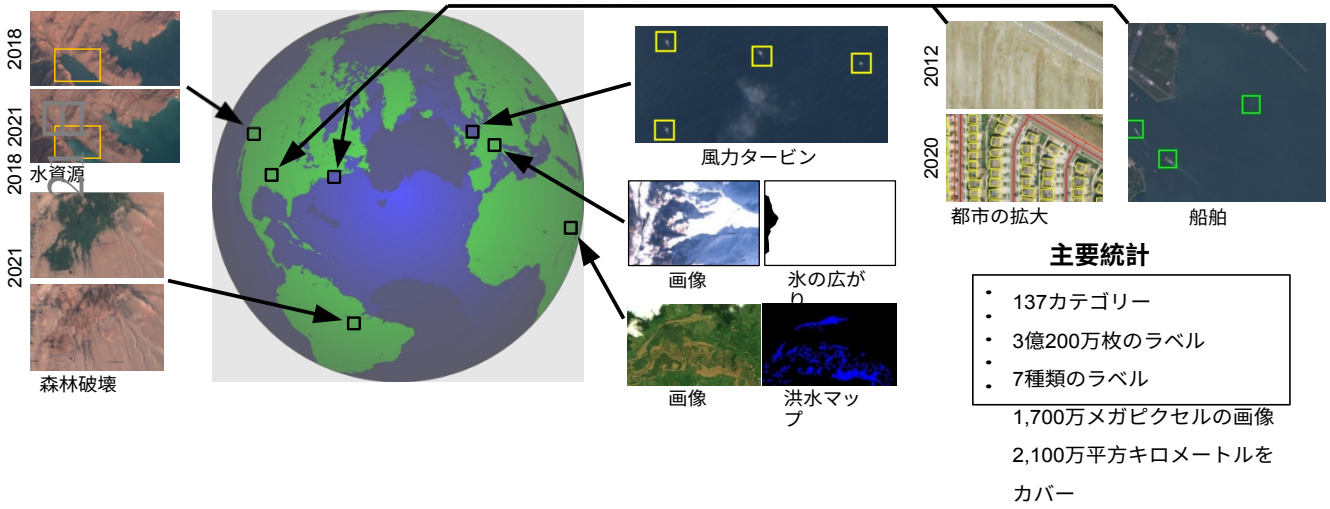

図1：SATLASPRETRAINは大規模なリモートセンシングデータセットである。SATLASPRETRAINのラベルは、水資源のモニタリング、森林伐採の追跡、インフラマッピングのための風力タービンの検出、氷河の損失の追跡、洪水の検出、都市の拡大の追跡、違法漁業に取り組むための船舶の検出など、多くの重要な惑星監視アプリケーションに関連している。

要旨

リモートセンシング画像は、森林伐採の追跡から違法漁業への取り組みまで、さまざまな地球監視アプリケーションに役立っている。地球は非常に多様であり、リモートセンシング画像に含まれる潜在的なタスクの量は膨大であり、特徴の大きさは数キロメートルからわずか数十センチメートルまで幅広い。しかし、一般化可能なコンピュータビジョン手法を作成することは、多くのタスクに対してこれらの多様な特徴を捉える大規模なデータセットがないこともあり、困難である。本論文では、Sentinel-2 とNAIPの画像を組み合わせ、137のカテゴリと7つのラベルタイプの下に302Mのラベルを持つ、広さと規模の両方で大規模なリモートセンシングデータセットであるSatlasPretrainを紹介する。SatlasPretrainを用いて、8つのベースラインと提案手法を評価し、非常に異なるタイプのセンサーからの画像からなる画像時系列の処理や、長距離の空間的コンテキストの活用など、リモートセンシング特有の再探索の課題に対処する

*上で、かなりの改善の余地があることを発見する。さらに、*SATLASPRETRAIN*で事前学習することで、下流タスクの性能が大幅に向上し、平均精度がImageNetより18%、次善のベースラインより6%向上することがわかった。その結果*

データセット、訓練済みモデルの重み、コードは *https://satlas-pretrain.allen.ai/。*

1. はじめに

衛星画像や航空画像は、物理的世界に関する多様な情報を提供する。都市部の画像では、マッピングされていない道路や建物を特定し、デジタル地図データセットに取り込むことができる。工業地帯の画像では、太陽光発電所や風力タービンを記録し、再生可能エネルギー導入の進捗状況を追跡することができる。氷河や森林の画像では、氷河の減少や森林伐採のようなゆっくりとした自然の変化を監視することができる。EUのSentinelミッション[4]のような、グローバルで、定期的に更新され、パブリックドメインのリモートセンシング画像ソースが利用できるようになったことで、私たちは、月単位、あるいは週単位で、地球規模で、これらのアプリケーションのすべて、あるいはそれ以上のアプリケーションのために地球を監視することができる。

地球は巨大なスケールを持っているため、リモートセンシング画像の全地球的な手動解析はコスト的に不可能である。これまでの研究では、リモートセンシング画像の位置情報を自動的に推測するために、コンピュータ・ビジョンを応用することが提案されている。

森林伐採や都市拡大などの土地被覆や土地利用の変化の監視 [46、47]、違法漁業に取り組むための船舶の位置や種類の予測 [42]、洪水、山火事、竜巻などの自然災害の進行状況や程度を追跡する [8、23、44]。しかし、実際には、展開されているアプリケーションのほとんどは、2つの理由から、再モータ・センシング画像の完全自動解析ではなく、手動または半自動解析に依存し続けている[1]。第一に、道路交通情報[12]のような主要なアプリケーションでさえ、精度が依然として障壁となっており、完全自動化は現実的ではない。第二に、専門家によるアノテーションが必要だが、ラベル付けされた例がほとんどないリモートセンシング・アプリケーションのロングテールが存在する（例えば、最近のニューヨーク・タイムズ紙の研究では、衛星画像を使用してブラジルの違法滑走路を手作業で記録した[9]）。

我々は、超大規模なマルチタスクリモートセンシングデータセットの欠如が、今日のリモートセンシングタスクの自動化手法の進歩の大きな障害になっていると考えている。第一に、ViT[26]やCLIP[43]のような最先端のアーキテクチャは、最高の性能を達成するために巨大なデータセットを必要とする。しかし、DOTA [55]、iSAID [58]、Deep Globe [24]のような物体検出、インスタンスセグメンテーション、およびセマンティックセグメンテーションのための既存のリモートセンシングデータセットは、COCOの328K画像やCLIPの学習に使用された数百万画像と比較して、それぞれ10K画像未満しか含まれていません。第二に、既存のリモートセンシングベンチマークは断片的であり、道路[41]、船舶[42]、作物の種類[28]などのカテゴリに個別のベンチマークがあるが、多くのカテゴリにまたがるベンチマークはない。大規模で、一元化され、アクセス可能なベンチマークがないため、タスク間での学習機会の移転ができず、コンピュータビジョンの研究者がこの領域で研究することが困難である。

SATLASPRETRAINは、リモートセンシング画像理解モデルを改善するための大規模なデータセットである。SATLASPRETRAINの目的は、*衛星画像に見える全てのものにラベルを付けることである*。この目的のために、SATLASPRETRAINはSentinel-2とNAIPの画像を組み合わせ、137の多様なカテゴリと7つのラベルタイプの下に302Mの明確なラベルを付けました。ラベルタイプは、風車や給水塔のような**点**、建物や航空港のような**ポリゴン**、道路や河川のような**ポリライン**、土地被覆カテゴリや水深（水深）のようなセグメンテーションと**回帰**ラベル、風力タービンのローター直径のような物体の**特性**、画像内の煙の存在のような**パッチ分類**ラベルです。図1は、SATLASPRETRAINの幅広いカテゴリと、それらが役立つ多様なアプリケーションを示しています。

我々は、SATLASPRETRAINの巨大なスケールにより、事前学習がダウンストリーム性能を大幅に向上させることを発見した。我々は、SATLASPRETRAINによる事前学習を、他のデータセットによる事前学習と比較した。

SATLASPRETRAINは、ImageNetと比較して7つのダウンストリームタスクの平均性能を18%向上させ、次善のベースラインと比較して6%向上させた。これらの結果は、SATLASPRETRAINが、コストのかかる専門家のアノテーションを必要とする多くのニッチなリモートセンシングタスクの精度を容易に改善できることを示している。

さらに、SATLASPRETRAINは、リモートセンシング領域特有の研究課題に取り組むコンピュータビジョン手法の研究に勇気を与えると信じている。汎用のコンピュータビジョン手法と比較して、リモートセンシングモデルでは、長距離の空間的コンテクストを考慮し、マルチスペクトル画像や合成開口レーダー（SAR）のような多様なセンサーによって撮影された経時的な画像間のインフォーメーションを合成し、何kmにも及ぶ森林（[2]）から街灯に至るまで、大きさが大きく異なる物体を予測するような特殊な技術が必要とされる。我々はSATLASPRETRAIN上で8つのコンピュータビジョンベースラインを評価し、既存のどの手法もSATLASPRETRAINの全てのラベルタイプをサポートしていないことを発見した。したがって、タスク固有の出力ヘッダを統合した再センタの研究[21, 29, 35, 36]に触発され、我々はSATLASNETと呼ばれる統一モデルを開発し、そのような7つのヘッダを組み込むことで、データセットの全てのカテゴリから学習できるようにした。各ラベルタイプで別々に学習する場合と比較して、全てのカテゴリでSATLASNETを共同学習し、その後各ラベルタイプで微調整することで、平均性能が7.1%向上することがわかった。

まとめると、我々の貢献は以下の通りである：

1. SATLASPRETRAINは、7つのラベルタイプの下に137のカテゴリーを持つ大規模なリモートセンシングデータセットである。
2. SATLASPRETRAINでの事前学習により、7つのダウンストリームデータセットの平均性能が6%向上することを示す。
3. SATLASNETは、SATLASPRETRAINのすべてのラベルタイプの予測をサポートする統一モデルです。

我々はデータセットとコードを https://satlas-pretrain.allen.ai/に公開した。また、SATLASPRETRAINで事前に訓練されたモデルの重みを再リースしています。

2. 関連作品

大規模リモートセンシングデータセット。汎用のリモートセンシングコンピュータビジョンデータセットがいくつかリリースされている。UC Merced Land Use (UCM)[57]とBigEarthNet[51]データセットは、それぞれ21と43のカテゴリからなる土地被覆分類を含み、AID[56]、Million-AID[39]、RESISC45[19]、Functional Map of the World (FMoW)[22]データセットは、さらに以下のような人工構造物に対応するカテゴリを含む。

	種類	クラス	ラベル	ピクセル	km2
サトラスプレトレイン	7	137	302222K	17003B	21320K
UCM【57	1	21	2K	1B	1K
ビッグアースネット[51]	1	43	1750K	9B	850K
AID【56	1	30	10K	4B	14K
ミリオンエイド[39]	1	51	37K	4B	18K
RESISC45 [19]	1	45	32K	2B	10K
FMoW [22]	1	63	417K	437B	1748K
DOTA [55]	1	19	99K	9B	38K
iSAID【58	1	15	355K	9B	38K

表1：既存のリモートセンシングデータセットに対するSATLASPRETRAINの比較（K=千、B=億）。種類はラベルの種類数、km^2。

橋や鉄道駅など、最大63のカテゴリーがある。シーン分類以外のタスクに焦点を当てたデータセットもいくつかある。DOTA[55]は、ヘリコプターからラウンドアバウトまでの18のカテゴリの物体を検出する。

これらのデータセットはすべて、単一のラベルタイプについて予測を行うものであり、ほとんどのデータセットは単一の画像から予測を行うものである。従って、これらのデータセットは3つの点で 制限されている：物体カテゴリの数、ラベルの多様性、そして画像時系列を横断して特徴 を合成することを学習するアプローチの機会である。対照的に、SATLASPRETRAINは、7つのラベルタイプ（表1の完全な比較を参照）の下に137のカテゴリを組み込み、予測精度を向上させるために手法が活用できる画像時系列を提供する。

xView3[42]は、SAR画像における船舶の位置（物体検出）と、船舶のタイプや長さといった船舶の属性（物体ごとの分類と回帰）を予測する。PASTIS-R[28]は、Sentinel-1およびSentinel-2コンステレーションによって撮影されたSARおよび光学衛星画像の時系列を使用して、作物畑の作物タイプの全球セグメンテーションを行う。IEEEのデータフュージョンデータセットは、土地被覆のセグメンテーションのようなタスクのために、様々な航空画像と衛星画像を含んでいる[48]。

リモートセンシングのための自己教師学習とマルチタスク学習。我々の研究と同様に、これらのアプローチは、ラベルが少ない下流アプリケーションの精度を向上させるという目標を共有している。いくつかの手法[7, 11, 40, 49, 50, 54]は、異なる時間に撮影された同じ場所の画像は、異なる場所の画像よりも近い表現になるように、時間的補強を対比学習のフレームワークに組み込んでいる。このモデルは、照明や直下角の条件の違い、季節の変化など、同じ場所の画像間の一時的な差異に対する不変性を学習することで、下流の性能を向上させることを示している。GPNAは、自己教師あり学習と、多様なタスクに対す る教師あり学習を組み合わせることを提案する[45]。

3. サトラスプレトレイン

SATLASPRETRAINは、リモートセンシング画像理解のための超大規模データセットであり、3つの重要な点において既存のリモートセンシングデータセットを凌駕している：

1. **スケール：** SATLASPRETRAINは、既存の最大データセットよりも40倍の画像ピクセルと150倍のラベルを含んでいる。
2. **ラベルの多様性：** 表1の既存のデータセットは、例えば分類のみなど、ラベルが一様である。SATLASPRETRAINのラベルは*7つのラベルタイプにまたがって*おり、さらに137のカテゴリーから構成され、既存の最大のデータセットより2倍多い。
3. **時空間画像とラベル：** 個々のリモートセンシング画像に関連付けるのではなく、我々のラベルは地理的座標（すなわち緯度経度位置）と時間範囲に関連付けられている。これにより、時間軸を超えた複数の画像からの予測や、隣接する画像からの長距離の空間的コンテキストの活用が可能になる。これらの特徴は、解決すればモデルの性能を大幅に向上させることができる、新たな再探索の課題を提示している。

まず、SATLASPRETRAINの構造の概要を説明し、その中に含まれるイメージを以下に詳述する。続いて、ラベルとそのレクチャー方法について説明する。

3.1. 構造とイメージ

SATLASPRETRAINは856Kの*タイル*で構成されている。これらのタイルは、ズームレベル13のWeb-Mercatorタイルに対応する。つまり、世界は2D平面に投影され、$2^{13} \times 2^{13}$のグリッドに分割され、各タイルはグリッドセルに対応する。したがって、SATLASPRETRAINの各タイルは、最大25km^{2}に及ぶ不連続な空間領域をカバーする。各タイルには、(1) そのタイルのリモートセンシング画像の時系列、(2) 137のSATLASPRETRAINカテゴリから抽出されたラベルが含まれる。図2はこのデータセットの概要で、図3はその全球の地理的範囲を示している。

既存のデータセットは通常、高解像度の画像（0.5〜2m/ピクセル）[19, 22, 39, 58]か、低解像度の画像（0.5〜2m/ピクセル）を使用している。
の画像（10 m/ピクセル）[27, 51]を使用している。高解像度の画像はより高い予測精度を実現する が、低解像度の画 像 は より頻繁に（毎週、毎年）、より広範 に（世界的に、限られた国で）入手可能であるため、実用 的なアプリケーションではしばしば採用される。したがって、SATLASPRETRAINでは、低解像度画像と高解像度*画像*の両方を組み込む。それぞれの画像モードについて、訓練とテストの分割を定義し、それぞれのモードについて独立に手法を比較する。

すべての856Kタイル（828K訓練と28Kテスト）において、低解像度の512x512画像を提供する。具体的には

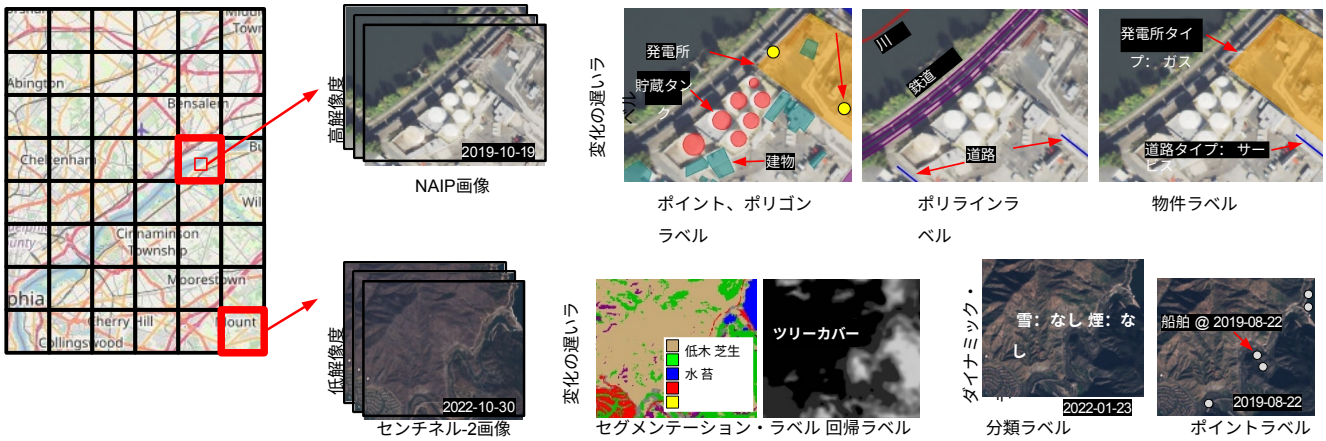

図2：SATLASPRETRAINデータセットの概要。SATLASPRETRAINは、ズーム13（左）の856K Web-Mercatorタイルの画像時系列とラベルから構成される。高解像度のNAIP画像（上）と低解像度のSentinel-2画像（下）。ラベルはゆっくりと変化するもの（タイルの最新の画像に対応）、あるいは動的なもの（特定の画像と時刻を参照）がある。

図3: SATLASPRETRAINの地理的範囲。明るいピクセルは、データセット内の画像とラベリングでカバーされている場所を示す。SATLASPRETRAINは南極大陸を除く全ての大陸をカバーしている。

2022年に撮影された8-12センチネル-2画像; これは、予測精度を向上させるために、場所の複数の空間的に整列された画像を活用する方法を可能にする。洪水や船の位置のような動的なラベルに関連する2016年から2021年の過去の画像も含む。Sentinel-2は10m/ピクセルのマルチスペクトル画像をキャプチャしており、欧州宇宙機関（ESA）はこれらの画像をオープンに公開している。いくつかのカテゴリーは低解像度画像では見えないので、このモードでは137カテゴリー中122カテゴリーについてのみ手法を評価する。

46K タイル（45.5K 訓練と 512 テスト）で、8192x8192の高解像度画像を提供する。2011年から2020年の間に米国国立農業画像プログラム(NAIP)から公開された3-5枚の1m/pixelの航空画像を含む。これらの画像は米国内でしか入手できないため、高解像度モードの訓練用およびテスト用のタイルは米国内に限定される。

ESAとUSGSから画像をダウンロードし、GDAL [6]を使って画像をWeb-Mercatorタイルに加工する。

SATLASPRETRAINの構造は、メソッドが空間的・時間的コンテクストを活用することを可能にする。メソッドは

同様に、予測精度を向上させるために、データセットの各タイルに含まれる画像時系列全体の特徴を合成することを学習することができる。同様に、予測精度を向上させるために、データセットの各タイルに含まれる画像時系列全体の特徴を合成することを学習することができる。例えば、ある作物畑で栽培されている作物の種類を予測する場合、農業サイクルの異なる段階における作物畑の観察から、そこで栽培されている作物の種類に関する異なる手がかりを得ることができる。対照的に、既存のデータセット（表1のFMoW以外を含む）は、通常、各ラベルを1つの画像に関連付け、その1つの画像のみでラベルを予測する手法を必要とする。

3.2. ラベル

SATLASPRETRAINのラベルは137のカテゴリーにまたがり、7つのラベルタイプがある（図2の例を参照）：

1. セマンティック・セグメンテーション-例えば、ピクセルごとの土地被覆の予測（水、森林、開発地域など）。
2. 回帰-例えば、ピクセルごとの水深（波の深さ）や樹木の被覆率を予測する。
3. ポイント（物体検出）-風力タービン、油井、船舶の予測など。
4. ポリゴン（インスタンス分割）- ビル、ダム、養殖場の予測など。
5. ポリライン-道路、河川、鉄道の予測など。
6. 点、多角形、多角形の特性-例：風力タービンのローター直径。
7. 分類-例えば、ある画像が山火事の煙の濃度を無視できるか、低いか、高いか。

ほとんどのカテゴリーは、道路や風力タービンのような変化の遅い物体を表している。データセット作成時に、これらのカテゴリのラベルが、各タイルで利用可能な最新の画像に対応することを目指す。したがって、推論中、もしこれらの

オブジェクトは、タイルで利用可能な画像時系列にわたって変化するため、モデル予測は時系列の最後の画像を反映する必要があります。いくつかのカテゴリは、船舶や洪水のような動的オブジェクトを表す。これらのカテゴリのラベルでは、オブジェクトの位置を指定することに加えて、ラベルは対応する画像のタイムスタンプを指定する。動的なカテゴリでは、時系列中の各画像に対して、モデルは別々の予測を行う必要がある。我々はSATLASPRETRAINのラベルを7つの情報源から導出する：ドメインの専門家による新しいアノテーション、Ama-zon Mechanical Turk (AMT)のワーカーによる新しいアノテーション、そして5つの既存データセット（OpenStreetMap [30]、NOAAライダースキャン）を処理する、WorldCover [53]、Microsoft Buildings [3]、C2S [5]。

各カテゴリは、タイルのサブセットでのみ注釈（有効）である。したがって、あるタイルでは、与えられたカテゴリは無効であることがあり、これはそのタイルにそのカテゴリのグランドトゥルースがないことを意味する。他のタイルでは、カテゴリは有効でもラベルがゼロの場合があり、これはタイルにそのカテゴリのインスタンスがないことを意味する。補足セクションA.1では、各カテゴリについて、そのカテゴリが有効であるタイルの数、そのカテゴリが少なくとも1つのラベルを持つタイルの数、およびそのカテゴリの下のラベルの数を詳述する。

SATLASPRETRAINのラベルは、多くの惑星や環境モニタリングのアプリケーションに関連しており、これについては補足セクションA.2で説明する。

各データソースのデータ収集プロセスを以下に要約する。

専門家によるアノテーション。2人の専門家が12のカテゴリー（沖合風力タービン、沖合プラットフォーム、砂州、6つの樹木被覆カテゴリー（低いか高いかなど）、3つの積雪カテゴリー（なし、部分的、完全）にアノテーションを行った。このプロセスを容易にするため、Sivと呼ばれる専用のアノテーションツールを構築した。例えば、海洋物体にアノテーションを行う場合、船舶と固定インフラを正確に区別するためには、同じ海洋位置の異なる時刻の画像を表示することが重要であることが分かった（一般に、船舶はいずれかの画像にのみ表示されるが、風力タービンやプラットフォームはすべての画像に表示される）。同様に、樹木の被度については、NAIPやSentinel-2の画像で樹木の被度が明確でない場合、Google MapsやOpenStreetMapのような外部ソースを参照することで精度が向上することがわかった。

AMT。AMT作業員は9つのカテゴリー（海岸の土地、海岸の水、防火剤の投下、野火で焼かれた地域、飛行機、屋上のソーラーパネル、3つの煙の存在カテゴリー（なし、低い、高い））にアノテーションを行った。AMTのアノテーションには、Sivのアノテーションツールを再利用し、必要に応じてカテゴリーごとにカスタマイズを加えた（詳細についてはSup-Sup-Supに記載）。

セクションA.3.1参照）。

アノテーションの質を最大化するために、各カテゴリーについて、まず適格性評価タスクを通してAMTワーカーを選択した：ドメインエキスパートが100-400タイルの間でアノテーションを行い、各候補AMTワーカーに同じタイルのアノテーションを依頼した。まず1人のワーカーに各タイルに注釈をつけてもらい、ラベルの品質を分析することで、タイルごとに必要なワーカー数を決定した（セクションA.3.2参照）。例えば、飛行機は1人の作業員で十分であることがわかったが、山火事で燃やされた地域については3人の作業員に各タイルにラベル付けをしてもらった。

OpenStreetMap（OSM）。OSMは、ユーザーによる編集によって構築された共同地図データセットである。OSMの対象は、道路から変電所まで幅広いカテゴリーにわたる。我々は2022年7月9日にOSM PBFファイルとしてGeofabrikからOSMデータを入手し、101のカテゴリーを抽出するためにGoのosmpbfライブラリを使って処理した。

リコールはOSMから得られたラベルの重要な問題である。初期の定性分析では、OSMのオブジェクトは精度は高いが、リコールはまちまちであることが一貫して観察された。大半のオブジェクトは正しいが、カテゴリーによっては、多くのオブジェクトが衛星画像では見えるが、OSMではマッピングされていないものもあった。この問題を軽減するために、タイル内のラベルと明確なカテゴリの数に基づいて、想起率が低い可能性が高いタイルを自動的にプルーンするヒューリスティックを採用した。例えば、道路は多いが建物はないタイルは、サイロや給水塔のような他のカテゴリのオブジェクトが欠落している可能性が高いことがわかった。これらのヒューリスティックは補足セクションA.4で詳述する。

これらのヒューリスティクスは、ほとんどのカテゴリーで高品質のラベルを得るのに十分であることがわかった。しかし、ガスステーション、ヘリポート、油井など、回収率の低い13のカテゴリーを特定した。1300のタイルの分析から、これらのカテゴリーではまだ少なくとも80%の想起率があると判断した。しかし、これらの13のカテゴリは、テストセットには十分な想起を持たないと判断した。したがって、高度に正確なテスト・セットを確保するために、これら13のカテゴリのそれぞれについて、OSMラベルで初期モデルをトレーニングし、高リコール低精度検出のために信頼閾値をチューニングした。セクションA.4では、これらのカテゴリとテストセットで確認された欠落ラベルの数について詳述する。

NOAAライダースキャン。ライダースキャンから作成されたNOAA沿岸地形図には、陸地の標高データと水深データが含まれている。このようなマップをNOAAの様々な調査から5,868枚ダウンロードし、それらを処理して、海抜と水深を算出した。

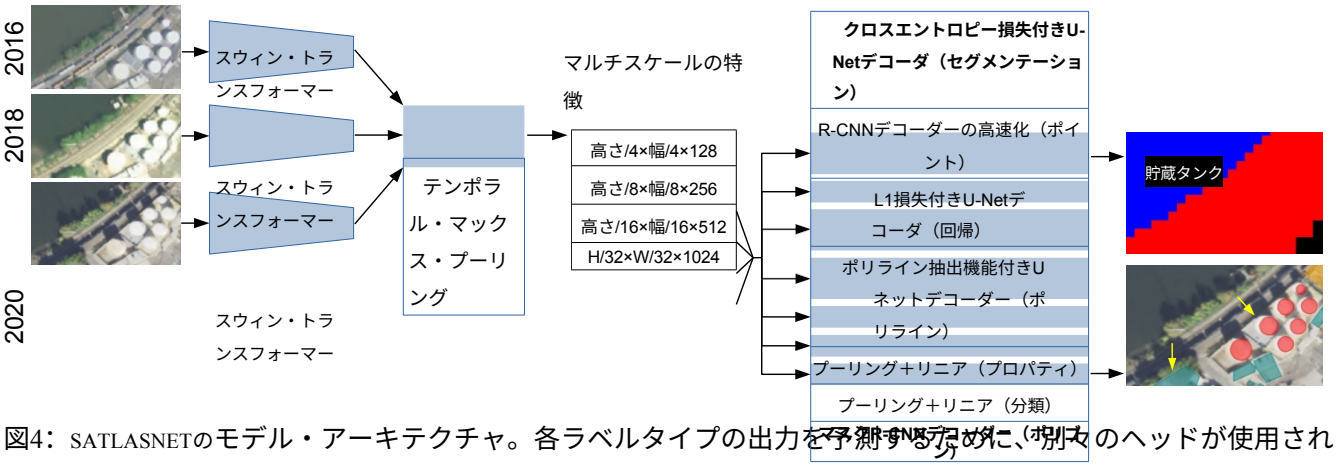

図4：SATLASNETのモデル・アーキテクチャ。各ラベルタイプの出力を予測するために、別々のヘッドが使用される。このような2つのヘッド（セグメンテーションとポリゴン）からの出力例を可視化する。

5,123個のSatlasPretrainタイルのピクセル深度と標高ラベル。

ワールドカバーWorldCover [53]は、欧州宇宙機関（ESA）によって開発されたグローバルな土地被覆マップである。我々はこの地図を加工して、不毛の土地から開発された地域まで、11の土地被覆と土地利用のカテゴリーを導き出した。

Microsoft Buildings。 SATLASPRETRAINでは、Microsoft Buildingsの様々なデータセット[3]から70のGeoJSONファイルを処理し、建物のポリゴンを導出しています。データはODbLの下で公開されています。

C2S。C2S [5]は、Sentinel- 2の画像に含まれる洪水と雲のラベルで構成されており、CC-BY-4.0の下で公開されている。我々はこのラベルをWeb-Mercatorにワープし、SATLASPRETRAINに含める。また、SATLASPRETRAINで他のSentinel-2画像と同じ処理を共有するために、C2Sで使用された画像に正確に対応するSentinel-2画像をダウンロードして処理します。

ラベルの規模とラベルの質のバランスを取ることは、新しいアノテーションを管理し、処理するデータソースを選択する上で重要な検討事項であった。データセットを開発する際、収集したラベルの精度と想起度を評価するために繰り返し分析を行い、この情報に基づいて後のアノテーションを改善し、データソースの処理を改良した。補足的なセクションA.5では、最終的なデータセットの各カテゴリーにおける不正確なラベルと欠落したラベルの分析が含まれている。90-95%の精度、2つは80-90%の精度である。

4. サトラスネット

例えば、Mask2Former [18]はセマンティックセグメンテーションとインスタンスセグメンテーションを同時に行うことができるが、ポリゴンの特性を予測したり画像を分類したりするようには設計されていない。このため、これらのモデルは、SATLASPRETRAINに存在する伝達学習の機会をフルに活用することができません。例えば、建物のポリゴンを検出することは、土地被覆や土地利用について画像をセグメンテーションするのに有用であると考えられます。

人間によって開発されたカテゴリーである。我々は、7つのラベルタイプ全てから学習可能な統一モデルSATLASNETを開発する。

図4は我々のモデルの概略図である。SATLASNETは、タスクに特化した出力ヘッド[21, 29, 35]を採用した最近の研究や、リモートセンシング画像の時系列を横断して特徴を合成する手法[22,27]にインスパイアされている。空間的に整列された画像の時系列を入力し、各画像（3つ以上のバンドを含むことがある）をSwin Transformer [38]バックボーン（Swin-Base）を通して処理し、4つのスケールで各画像の特徴マップを出力する。各スケールで最大時間プーリングを適用し、1セットのマルチスケール特徴を導き出す。この特徴量を7つの出力ヘッド（各ラベルタイプに1つずつ）に渡して出力を計算する。ポリラインについては、特殊なポリライン抽出アーキテクチャーが精度を向上させることが示されているが[10, 33, 52]、我々はより単純なセグメンテーションアプローチ[60]を採用することを選択する。このアプローチでは、UNetヘッドを適用してポリラインカテゴリのために画像をセグメンテーションし、セグメンテーション確率をバイナリ閾値処理、形態学的間引き、ラインフォローと単純化[20]で後処理してポリラインを抽出する。

5. 評価

まずセクション5.1では、SATLASPRETRAINのテストスプリットで、我々の手法と8つの分類、セグメンテーション、インスタンスセグメンテーションのベースラインを評価する。次に、SATLASPRETRAINでの事前学習と、他のリモートセンシングデータセットでの事前学習、およびリモートセンシングに特化した自己教師付き学習技術を比較する。

5.1. SatlasPretrainでの結果

方法SATLASPRETRAIN上でSATLASNETと8つのベースラインを比較する。SATLASPRETRAINのラベルタイプの部分集合に対して、標準的なモデル、または最先端の性能を提供するモデルのいずれかをベースラインとして選択する。どのベースラインもSATLASPRETRAINの7つのラベルタイプ全てを扱うことはできない。特性予測と分類については、ResNet [32]、ViT [26]、および

方法	NAIPの高解像度画像						センチネル2低解像度画像						
	セグ	レジ	白金	Pgon↑上位	ライン	*提案*	セグ	レジ	白金	Pgon↑上位	ライン	*提案*	Cls↑。
PSPネット（ResNext-101）【59	77.8	15.0	-	-	53.2	-	62.1	16.2	-	-	30.7	-	-
リンクネット（ResNext-101）[15]	77.3	12.9	-	-	61.0	-	61.1	14.1	-	-	41.4	-	-
ディープラボv3（ResNext-101）[16]	80.1	10.6	-	-	59.8	-	61.8	13.9	-	-	44.7	-	-
ResNet-50 [32]	-	-	-	-	-	87.6	-	-	-	-	-	70.3	97
ViT-ラージ[26]	-	-	-	-	-	78.1	-	-	-	-	-	66.9	99
スウィン・ベース[38]	-	-	-	-	-	87.1	-	-	-	-	-	69.4	99
マスクR-CNN（ResNet-50）[31]	-	-	27.6	30.4	-	-	-	-	22.0	12.3	-	-	-
マスクR-CNN（スウィンベース）[31]	-	-	30.4	31.5	-	-	-	-	25.6	15.2	-	-	-
ISTR [34]	-	-	2.0	4.9	-	-	-	-	1.2	1.4	-	-	-
サトラスネット（シングルイメージ、タイプ毎）	79.4	8.3	28.0	30.4	61.5	86.6	64.8	9.3	25.7	14.8	42.5	67.5	99
サトラスネット（シングルイメージ、ジョイント）	74.5	7.4	28.0	31.1	60.9	87.3	55.8	10.6	22.0	10.3	45.5	73.8	99
サトラスネット（シングルイメージ、微調整済み）	79.8	**7.2**	32.3	33.0	**62.4**	**89.5**	65.3	9.0	27.4	16.3	45.9	**80.0**	99
サトラスネット（マルチイメージ、タイプ毎）	79.4	8.2	25.8	27.5	59.2	77.3	67.2	10.5	31.9	19.0	48.1	67.1	99
サトラスネット（マルチイメージ、共同）	79.2	7.8	31.2	33.8	53.6	87.8	66.7	8.5	31.5	19.5	41.9	78.8	99
サトラスネット（マルチイメージ、微調整可能）	**81.0**	7.6	**33.2**	**34.1**	61.1	89.2	**69.7**	**7.8**	**32.0**	**20.2**	**50.4**	**80.0**	99

表2：SATLASPRETRAINテストセットの高解像度画像モードと低解像度画像モードでの結果。セグメンテーション(Seg)、回帰(Reg)、ポイント(Pt)、ポリゴン(Pgon)、ポリライン(Pline)、プロパティ(Prop)、分類(Cls)。 Regの絶対誤差（低い方が良い）と、その他の精度を示す（高い方が良い）。

Swin Transformer [38]。セグメンテーション、回帰、ポリラインについては、PSPNet [59]、LinkNet [15]、DeepLabv3 [16]を比較。点とポリゴンについては、Mask R-CNN [31]と ISTR [34]を比較する。

SATLASNETの3つのバリエーションを訓練する：

- タイプごと：ラベルのタイプごとに別々にトレーニングする。
- ジョイント：全カテゴリーを合同でトレーニングする。
- Fine-tuned：各ラベルタイプで共同学習したパラメータを微調整する。

すべてのベースラインは、ラベルタイプごとに（扱えるラベルタイプのサブセットで共同学習した後に）微調整され、最高の性能を発揮する。

各SATLASNETのバリエーションについて、単一画像モードと複数画像モードでの評価も行った。すべてのベースラインと単一画像SATLASNETについて、以下のどちらかで学習例をサンプリングする。

(a)タイルをサンプリングし、そのタイルの最新の画像とゆっくり変化するラベル（ダイナミックおよび他の有効なカテゴリはマスクされている）をペアリングする、または、(b)タイルと画像をサンプリングし、画像と対応するダイナミックラベルをペアリングする。低解像度モードでは8枚のSentinel-2画像の時系列を、高解像度モードでは4枚のNAIP画像の時系列を入力として提供する。ゆっくりと変化するラベルの場合、画像はタイムスタンプ順に並べられるが、動的ラベルの場合は、常に時系列入力の最後にサンプリングされた画像を並べる。すべての場合において、例中に現れるカテゴリー（cat-egories）の最大逆頻度に基づいて例をサンプリングする。ここではRGBバンドのみを用いるが、9つのSentinel-2バンドを用いた単一画像SATLASNETの結果を補足のセクションCに含める。

高解像度の推論では、タイルを覆う画像が8K x 8Kであるため、256個の512x512ウィンドウを個別に処理し、モデルの出力をマージする。ランダムなトリミング、水平方向と垂直方向のトリミングを採用している。

学習中に、垂直方向の反転、ランダムなサイズ変更の拡張を行う。ImageNetで事前学習した重みでモデルを初期化する。Adamオプティマイザを使用し、学習率を10^{-4}に初期化し、学習損失がプラトーに達した時点で半分の10^{-6}まで減衰させる。バッチサイズ32で100Kバッチを学習する。

メトリクス。分類には精度、セグメンテーションにはF1スコア、回帰には平均絶対誤差、点とポリゴンにはmAP精度、ポリラインにはGEO精度[14]というように、ラベルタイプごとに標準的なメトリクスを使用する。メトリックスをカテゴリーごとに計算し、各ラベル・タイプのカテゴリー間の平均を報告する。

結果SATLASPRETRAINの結果を表に示す。

2.SATLASNETは、7つのラベルタイプにおいて、タイプごとに学習させた場合、最新のベースラインの性能に匹敵するかそれを上回る。SATLASNETの1つのパラメータセットを全てのカテゴリに対して共同学習させると、いくつかのラベルタイプにおける平均性能は低下するが、SATLASNETはほとんどのケースで競争力を維持する。この学習モードは、推論時に、ラベルタイプごとに1回ではなく、各画像に対して1回だけバックボーン特徴を計算する必要があるため、大きな効率向上をもたらす。SATLASNETを各ラベルタイプで共同学習から得られたパラメータを用いて微調整した場合、ラベルタイプと画像モード全体で、タイプごとの学習と比較して平均7.1%の相対的な改善が得られた。これは、ラベルタイプ間に伝達学習の可能性があるという我々の仮説を支持するものであり、性能向上のための統一モデルの有用性を検証するものである。多画像SATLASNETは、平均性能においてさらに5.6%の相対的改善をもたらし、より良い予

測を生み出すために画像時系列間の情報を効果的に
合成できることを示している。

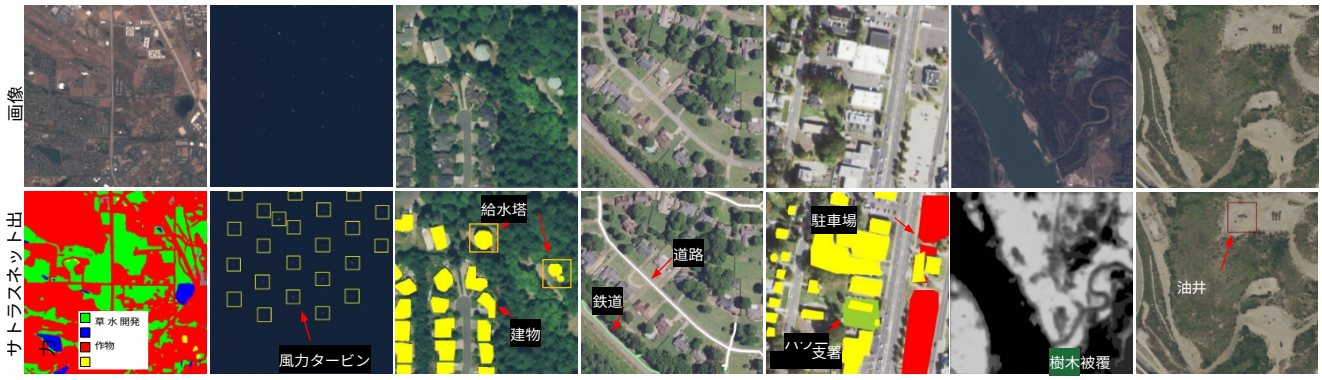

図5：SATLASPRETRAINの定性的結果。右端：SATLASNETが1/5の油井しか検出できなかった失敗例。

方法	ユーシーエム		RESISC45		エイド		FMoW		マス・ロード		マス・ビルディング		エアバス船舶		**平均**	
	50	すべて	50	すべて	50	すべて	50	すべて	50	すべて	50	すべて	50	すべて	50	すべて
ランダム初期化	0.26	0.86	0.15	0.77	0.18	0.68	0.03	0.17	0.69	0.80	0.68	0.77	0.31	0.53	0.33	0.65
イメージネット [25]	0.35	0.92	0.17	0.95	0.20	0.81	0.03	0.21	0.77	0.85	0.78	0.83	0.37	0.65	0.38	0.75
ビッグアースネット [51]	0.35	0.95	0.20	0.94	0.23	0.78	0.03	0.27	0.78	0.85	0.81	0.85	0.40	0.68	0.40	0.76
ミリオンエイド[39]	0.72	0.97	0.30	0.96	0.30	0.82	0.04	0.35	0.78	0.84	0.82	0.85	0.46	0.67	0.49	0.78
DOTA [55]	0.56	**0.99**	0.28	0.95	0.33	0.83	0.03	0.30	**0.82**	0.86	0.84	0.87	**0.62**	0.75	0.50	0.79
iSAID【58	0.60	0.97	0.29	0.97	0.34	0.86	0.04	0.30	**0.82**	0.86	0.84	0.86	0.55	0.73	0.50	0.79
MoCo【17	0.14	0.14	0.07	0.09	0.05	0.12	0.02	0.03	0.56	0.69	0.62	0.63	0.01	0.21	0.21	0.27
SeCo【40	0.48	0.95	0.20	0.90	0.27	0.74	0.03	0.26	0.70	0.81	0.71	0.77	0.27	0.54	0.38	0.71
サトラスプレトレイン	**0.83**	**0.99**	**0.36**	**0.98**	**0.42**	**0.88**	**0.06**	**0.44**	**0.82**	**0.87**	**0.87**	**0.88**	0.56	**0.80**	**0.56**	**0.83**

表3：50例(50)またはダウンストリームデータセット全体(All)で微調整した場合の7つのダウンストリームタスクの結果。精度はUCM、RESISC45、AIDについて、F1スコアはFMoW、Mass Roads、Mass Buildings、Airbus Shipsについて報告されている。SATLASPRETRAINの事前トレーニングにより、タスク全体の平均精度が、次に良いベースラインよりも6%向上しました。

リモートセンシング画像時系列の作成方法には、さらなる改善の余地がある。

図5に定性的な結果を示し、補足のセクションEに追加例を示す。風力タービンやタワーなど、いくつかのカテゴリーでは高い精度を達成した。しかし、油井については、1つの油井は検出されるが、他のいくつかの油井は検出されない。同様に、道路や鉄道のようなポリラインの特徴では、これらのカテゴリの十分なトレーニングデータにもかかわらず、モデルは短いノイズの多いセグメントを生成します。ポリライン[33, 52]のような特殊な出力タイプに合わせたモデルを組み込み、改善することで、精度が向上する可能性があると考えます。

5.2. 川下パフォーマンス

次に、SATLASPRETRAINで事前学習した場合の7つのダウンストリームタスクの精度を、既存の4つのリモートセンシングデータセットで事前学習した場合と比較して評価する。各ダウンストリームタスクについて、わずか50例で学習した場合と、データセット全体で学習した場合の精度を評価する。これは、専門家のアノテーションを必要とし、したがってラベル付き例が

少ないニッチなリモートセンシングアプリケーションの性能を向上させるという課題に焦点を当てるためである。

方法高解像度画像での事前学習と、高解像度画像での事前学習を比較する。

SATLASPRETRAINでは、既存の4つのリモートセンシングデータセットで事前学習を行っている：BigEarthNet [51]、Million- AID [39]、DOTA [55]、iSAID [58]。 我々はすべてのケースでSATLASNETを使用し、事前に訓練されたSwinバックボーンを各ダウンストリームデータセットで微調整した。

また、2つの自己教師付き学習法、Momentum Contrast v2 (MoCo) [17] と Seasonal Con- trast (SeCo) [40]を比較する。後者はリモートセンシングに特化した手法で、同じ場所の複数の画像キャプチャを活用し、季節変化に対する不変性を学習する。MoCoでは、SATLASNETモデルを使用し、SATLASPRETRAIN画像で学習する。SeCoについては、彼らのデータセットで訓練したオリジナルモデルを評価する。ダウンストリームタスクの自己教師を通して学習した重みを微調整する。補足のセクションB.3では、その他のバリエーションについての結果を示す。

まずバックボーンを凍結し、32K例について予測ヘッドのみを訓練し、その後モデル全体を微調整することで、事前訓練と自己教師あり学習の両方の方法を微調整する。補足セクションB.1に実験の詳細を示す。

下流のデータセット。ダウンストリームタスクは、21から63のカテゴリーで分類された、既存の4つの大規模リモートセンシングデータセットで構成されている：

UCM[57]、AID[56]、RESISC45[19]、FMoW[22]。他の3つは、Massachusts BuildingsとMas- sachusts Roadsデータセット[41]で、これらはセマンティックセグメンテーションを伴う。

結果表3はトレーニングセットのサイズを変化させた場合のダウンストリームのパフォーマンスを示している。SATLASPRETRAINは一貫してベースラインを凌駕しており、50例で学習した場合、ImageNetの事前学習よりもタスク全体の平均精度を18%向上させ、次善のベースラインよりも6%向上させた。SATLASPRETRAINの事前学習から得られる表現が一般化可能であること、そしてSATLASPRETRAINが多くのニッチなリモートセンシングアプリケーションの性能を向上させる可能性があることを明確に示している。補足B.2に、より多様な訓練例を用いた結果を掲載する。

6. AIが生成する地理空間データへの利用

私たちはSATLASPRETRAINを導入し、衛星画像からAIが生成するグローバルな地理空間データのプラットフォームであるSatlas（https://satlas.allen. ai/）の高精度モデルを開発している。風力タービンや太陽光発電所の位置のようなタイムリーな地理空間データは、排出削減、災害救援、都市計画などの意思決定に不可欠である。しかし、手作業によるキュレーションはコスト高になることが多いため、高品質なグローバル地理空間データ製品を入手するのは困難である。Satlasはその代わりに、風力タービンの検出などのタスク用に微調整されたモデルを適用し、衛星画像から地理空間データを毎月自動的に抽出する。Satlasは現在、風力タービン、太陽光発電所、海上プラットフォーム、樹木被覆の4つの地理空間データ製品で構成されている。

7. 結論

SATLASPRETRAINは、既存のデータセットの規模とラベルの多様性の両方を改善することで、リモートセンシング手法のための効果的な超大規模データセットとして機能する。SATLASPRETRAINで事前学習することで、ImageNetと比較して18%、既存のリモートセンシングデータセットと比較して6%もダウンストリームの平均精度が向上し、ラベル付きサンプルが少ないロングテールのリモートセンシングタスクに容易に適用できることを示している。我々はすでにSATLASPRETRAINで事前に訓練したモデルを活用し、風力タービン、太陽光発電所、海上プラットフォーム、Satlasプラットフォーム（https://satlas.allen.ai/）の樹木被覆を正確に検出している。

A. 補足資料

補足資料は https://github.com/allenai/satlas/blob/main/SatlasPretrain.mdからアクセスできる。

参考文献

[1] 地球の衛星画像 をより良く解析するためのマシンビジョンへの挑戦。*MITテクノロジーレビュー*

[2] Airbus ship detection challenge. https://www.kaggle.com/c/airbus-ship-detection, 2018.Airbus.

[3] Microsoft Building Footprints Datasets, 2021.マイクロソフト

[4] Copernicus Sentinel Missions. https://sentinel.esa.int/web/sentinel/home, 2022.European Space Agency.

[5] 世界の洪水事象と雲量データセット（バージョン1.0）、 2022。Cloud to Street, Microsoft, Radiant Earth Foundation.

[6] GDAL、2023年オープンソース地理空間財団。

[7] ペリ・アキバ、マシュー・プリ、マシュー・レオッタ。リモートセンシングタスクのための自己教師付き素材・テクスチャ表現学習.*コンピュータビジョンとパターン認識に関する IEEE/CVF Conference (CVPR)*, pages 8203-8215, 2022.

[8] ロバート・S・アリソン、ジョシュア・M・ジョンストン、グレゴリー・クレイグ、シオン・ジェニングス。Airborne Optical and Thermal Remote Sensing for Wildfire Detection and Monitoring.*Sensors*, 16(8):1310, 2016.

[9] マヌエラ・アンドレオーニ、ブラッキ・ミリオッツィ、パブロ・ロブレス、デニス・ルー。The Illegal Airstrips Bringing Toxic Mining to ブラジル先住民の土地。*The New York Times*.

[10] Favyen Bastani 、 Songtao He 、 Sofiane Abbar 、 Mohammad Alizadeh 、 Hari Balakrishnan 、 Sanjay Chawla、Sam Mad-den、David DeWitt。RoadTracer：航空画像からの道路網の自動抽出。*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4720-4728, 2018.

[11] Favyen Bastani, Songtao He, Satvat Jagwani, Mohammad Alizadeh, Hari Balakrishnan, Sanjay Chawla, Sam Madden, and Mohammad Amin Sadeghi.衛星画像で検出された変化を利用したストリートマップの更新。このような場合、このような情報を利用することで、より多くの情報を得ることができる。

[12] ファビエン・バスタニ、サミュエル・マデン道路抽出を超えて：空中画像を用いた地図更新のためのデータセット。*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11905-11914, 2021.

[13] Anil Batra, Suriya Singh, Guan Pang, Saikat Basu, CV Jawa- har, and Manohar Paluri.方位とセグメンテーションの共同学習による道路連結性の改善。In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10385-10393, 2019.

[14] ジェームズ・ビアジオーニとヤコブ・エリクソン。地図推論はノイズと格差に直面する。*第20回地理情報システムの進歩に関する国際会議（In-ternational Conference on Advances in Geographic Informa- tion Systems）予稿集*、79-88ページ、2012年。

[15] アビシェック・チャウラシアとエウジェニオ・カルルチェッロLinkNet： Ex- ploiting encoder representations for efficient semantic seg- mentation.*IEEE Visual Communications and Image Pro- cessing (VCIP)*, pages 1-4, 2017.

[16] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam.Rethinking Atrous Convolution for Semantic Image Segmentation.*ArXiv*, abs/1706.05587, 2017.

[17] (訳 注 : 邦 訳 は 「 邦 訳 邦 題 」) 。 *arXiv preprint arXiv:2003.04297*, 2020.

[18] Bowen Cheng, Ishan Misra, Alexander G. Schwing, Alexan- der Kirillov, and Rohit Girdhar.Masked-attention Mask Transformer for Universal Image Segmentation.2022.

[19] Gong Cheng, Junwei Han, and Xiaoqiang Lu.Remote Sensing Image Scene Classification： Benchmark and State of the Art.*Proceedings of the IEEE*, volume 105, pages 1865-1883, 2017.

[20] Guangliang Cheng, Ying Wang, Shibiao Xu, Hongzhen Wang, Shiming Xiang, and Chunhong Pan.Cascaded End-to-end Convolutional Neural Network による自動道路検出と中心線抽出。*IEEE Transactions on Geoscience and Remote Sensing*, 55(6):3322-3337, 2017.

[21] Jaemin Cho, Jie Lei, Hao Tan, and Mohit Bansal.テキスト生成による視覚と言語タスクの統合。In *Interna- tional Conference on Machine Learning*, pages 1931-1942.PMLR, 2021.

[22] ゴードン・クリスティ、ニール・フェンドリー、ジェームズ・ウィルソン、ライアン・ムカルジー。Functional Map of the World.*Proceedings of the IEEE/CVF Conference on Computer Vision and Pat- tern Recognition (CVPR)*, 2018.

[23] (1)SAR画像と付随データを組み合わせた洪水検知のためのベイジアンネットワーク。SAR画像と補助データを組み合わせた洪水検出のためのベイジアンネットワーク。*IEEE Transactions on Geoscience and Remote Sensing*, 54(6):3612-3625, 2016.

[24] Ilke Demir、 Krzysztof Koperski、 David Lindenbaum、 Guan Pang、Jing Huang、Saikat Basu、Forest Hughes、 Devis Tuia、Ramesh Raskar。ディープグローブ2018：衛星画像から地球を解析する挑戦。*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 172-181, 2018.

[25] J.Deng, W. Dong, R. Socher, L. J. Li, Kai Li, and Li Fei-Fei.ImageNet：大規模階層型画像データベース。*IEEE Conference on Computer Vision and Pattern Recogni- tion (CVPR)*, 2009.

[26] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 画 像 は 16x16語の価値がある：(1)画像は16x16語の価値がある : Transformers for Image Recognition at Scale.In *International Conference on Learning Representa- tions*, 2021.

[27] Vivien Sainte Fare Garnot and Loic Landrieu.Convo-lutional Temporal Attention Networks を用いた衛星画像時系列のパノプティックセグメンテーション。*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4872-4881, 2021.

[28] ヴィヴィアン・サントファール・ガルノ、ロイック・ランドリュー、ネスリーヌ・チェハタ。衛星時系列からの作物マッピングのためのマルチモーダル時間アテンションモデル。*ISPRS Journal of Pho- togrammetry and Remote Sensing*, pages 294-305, 2022.

[29] タンメイ・グプタ、アミタ・カマス、アニルッダ・ケンバヴィ、デレク・ホイエム。汎用ビジョンシステムに向けて：エンド・ツー・エンドでタスクに依存しないビジョン言語アーキテクチャ。

テュア。*Proceedings of the IEEE/CVF Conference on Com-puter Vision and Pattern Recognition (CVPR)*, pages 16399- 16409, 2022.

[30] モルデカイ・ハクレイとパトリック・ウェーバー OpenStreetMap：ユーザー生成ストリートマップ。*IEEE Pervasive computing*, 7(4):12-18, 2008.

[31] Kaiming He, Georgia Gkioxari, Piotr Dolla´r, and Ross Gir- shick.Mask r-cnn. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2980-2988, 2017.

[32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun.Deep Residual Learning for Image Recognition.In *Proceed- ings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770-778, 2016.

[33] Songtao へ、 Favyen Bastani 、 Satvat Jagwani 、 Mohammad Alizadeh 、 Hari Balakrishnan 、 Sanjay Chawla 、 Mohamed M Elshrif 、 Samuel Madden 、 Mohammad Amin Sadeghi 。Sat2Graph： Graph-Tensor Encoding による道路グラフ抽出。*European Conference on Computer Vision*, pages 51-67, 2020.

[34] ISTR: End-to-End Instance Segment with Transform- ers.ISTR: End-to-End Instance Segmentation with Transform- *ers.ArXiv*, abs/2105.00637, 2021.

[35] ロンハン・フー、アマンプリート・シンユニット： Unified Transformerによるマルチモーダル・マルチタスク学習。 In *Proceedings of the IEEE/CVF International Conference on Computer Vi- sion (ICCV)*, pages 1439–1449, 2021.

[36] (1)視覚的な視覚のための、ウェブ的な教師付き概念拡張,(2)視覚的な視覚のための、ウェブ的な教師付き概念拡張,(3)視覚的な視覚のための、ウェブ的な教師付き概念拡張.Webly Supervised Concept Expansion for General Purpose Vision Models. *arXiv preprint arXiv:2202.02317*, 2022.

[37] Zuoyue Li, Jan Dirk Wegner, and Aure´lien Lucchi.Topolog- ical Map Extraction from Overhead Images.*Proceedings of the IEEE/CVF International Conference on Computer Vi- sion (ICCV)*, pages 1715-1724, 2019.

[38] Z.リウ、リン、カオ、フー、ウェイ、チャン、リン、S. B. 郭 。 Swin Transformer: Hierarchical Vision Transformer using Shifted Windows.In *IEEE/CVF International Confer- ence on Computer Vision (ICCV)*, pages 9992-10002, 2021.

[39] ヤン・ロン、グイ・ソン・シャ、シェンヤン・リー、ウェン・ヤン、マイケル・イン・ヤン、シャオ・シアン・ズー、リャンペイ・チャン、デレン・リー。空中画像解釈のためのベンチマークデータセットの作成について： Review, Guidances, and Million-AID.*IEEE Journal of Selected Topics in Applied Earth Observa- tions and Remote Sensing*, pages 4205-4230, 2021.

[40] オスカー・マナス、アレクサンドル・ラコスト、グザビエ・ジロ・イ・ニエト、ダビド・バスケス、パウ・ロドリゲス。季節のコントラスト：未修正のリモートセンシングデータからの事前学習。*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[41] ヴォロディミル・ムニ*Machine Learning for Aerial Image La- beling*.博士論文、トロント大学、2013年。

[42] xView3-SAR: Detecting Dark Fishing Ac- tivity Using Synthetic Aperture Radar Imagery.*Neural In- formation Processing Systems*, 2022.

[43] このような場合、「自然言語スーパーヴィジョン」から「伝達可能な視覚モデル」を学習することになる。 *International Conference on Machine Learning*, pages 8748-8763, 2021.

[44] スダ・ラディカ、田村幸雄、松井正弘.Appli- cation of Remote Sensing Images for Natural Disaster Mit-igation using Wavelet based Pattern Recognition Analysis.In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 84-87, 2016.

[45] 地理空間システムのための汎用ニューラル・アーキテクチャ。 A General Purpose Neural Architecture for Geospatial Systems.*HADR Work- shop at NeurIPS 2022*, 2022.

[46] Caleb Robinson, Le Hou, Kolya Malkin, Rachel Soobitsky, Jacob Czawlytko, Bistra Dilkina, and Nebojsa Jojic.多解像度データによる大規模高解像度土地被覆マッピング。 *Proceedings of the IEEE/CVF Confer- ence on Computer Vision and Pattern Recognition (CVPR)*, pages 12726-12735, 2019.

[47] Caleb Robinson, Anthony Ortiz, Kolya Malkin, Blake Elias, Andi Peng, Dan Morris, Bistra Dilkina, and Nebojsa Jojic.Human-Machine Collaboration for Fast Land Cover Map- ping. pages 2509-2517, 2020.

[48] Ronny Ha¨nsch; Claudio Persello; Gemine Vivone; Javiera Castillo Navarro; Alexandre Boulch; Sebastien Lefevre; Bertrand Le Saux.データフュージョンコンテスト2022（dfc2022）、2022年。

[49] Linus Scheibenreif, Joe¨lle Hanna, Michael Mommert, and Damian Borth.Linus Scheibenif, Joe¨le Hanna, Michael Momert, and Damianorth.このような場合、「自己教師型視覚変換器」による土地被覆のセグメンテーションと分類 .*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1422-1431, 2022.

[50] Linus Scheibenreif, Michael Mommert, and Damian Borth.Contrastive Self-Supervised Data Fusion for Satellite Im- agery.*ISPRS Annals of the Photogrammetry, Remote Sens- ing and Spatial Information Sciences*, 3:705-711, 2022.

[51] Gencer Sumbul、 Marcela Charfuelan、 Begum Demir、 Volker Markl。BigEarthNet：リモートセンシング画像理解のための大規模ベンチマークアーカイブ。In *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019.

[52] Yong-Qiang Tan, Shang-Hua Gao, Xuan-i Li, Ming-Ming Cheng, and Bo Ren.VecRoad：道路グラフ抽出のための点ベースの反復的グラフ探索。このような、道路グラフを抽出するためのポイントベースの反復的なグラフ探索を、 VecRoad: Point-based-terterative graph exploreoration for Road Graph Extraction.

[53] Ruben Van De Kerchove, Daniele Zanaga, Wanda Keersmaecker, Niels Souverijns, Jan Wevers, Carsten Brockmann, Alex Grosu, Audrey Paccini, Oliver Cartus, Maurizio Santoro, et al. ESA WorldCover：センチネル1号と2号のデータに基づく2020年の10m分解能での全球土地被覆マッピング。 In *AGU Fall Meeting Abstracts*, volume 2021, pages GC45I- 0915, 2021.

[54] Yi Wang, Nassim Ait Ali Braham, Zhitong Xiong, Chenying Liu, Conrad M. Albrecht, and Xiao Xiang Zhu.SSL4EO-S12：大規模マルチモーダル、マルチタイムデータセット

地球観測における自己教師付き学習のための。*ArXiv*, abs/2211.07044, 2022.

[55] DOTA: Large-scale Dataset for Object Deection of Airer Images.DOTA: A Large-scale Dataset for Object Detection in Aerial Images.*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[56] グイ・ソン・シャ、ジングウェン・フー、ファン・フー、バオグァン・シー、シャン・バイ、ヤンフェイ・ジョン、チャン・リャンペイ。AID：空中シーン分類の性能評価のためのベンチマークデータセット。*IEEE Journal of Transactions on Geoscience and Remote Sensing*, 55(7):3965-3981, 2017.

[57] イー・ヤン、ショーン・ニューサム。Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification.*ACM Con- ference on Spatial Information (SIGSPATIAL)*, 2010.

[58] Syed Waqas Zamir, Aditya Arora, Akshita Gupta, Salman Khan, Guolei Sun, Fahad Shahbaz Khan, Fan Zhu, Ling Shao, Gui-Song Xia, and Xiang Bai：空中画像のインスタンス分割のための大規模データセット。In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019.

[59] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia.ピラミッドシーンパーシングネットワーク。*Proceedings of the IEEE/CVF Conference on Computer Vi- sion and Pattern Recognition (CVPR)*, pages 6230-6239, 2017.

[60] Lichen Zhou、Chuang Zhang、Ming Wu。D-LinkNet：高解像度衛星画像の道路抽出のための事前訓練されたエンコーダーと拡張畳み込みを用いたLinkNet.*Proceedings of the IEEE/CVF Conference on Computer Vi- sion and Pattern Recognition (CVPR) Workshops*, pages 182-186, 2018.

[61] Stefano Zorzi, Shabab Bazrafkan, Stefan Habenschuss, and Friedrich Fraundorfer.PolyWorld：衛星画像におけるグラフ・ニューラル・ネットワークによる多角形建物抽出。*コンピュータビジョンとパターン認識に関するIEEE/CVF会議(CVPR)予稿集*, ページ1848-1857, 2022.