

# 英語論文#4

2023/10/17

M1

建元 了

# 論文の概要

- タイトル
  - **satlasPretrain: A Large-Scale Dataset for Remote Sensing Image Understanding**
- 執筆者
  - Favyen Bastini, Piper Wolters, Ritwik Gupta, Joe Fedinando, Aniruddha Kenmbhavi, Allen Institute for AI
- 掲載
  - ICCV 2023
- 選択理由
  - リモートセンシングにおける高解像度と低解像度を含むデータセットの調査

# 背景

- 衛星画像や航空画像は、森林伐採や都市拡大などの土地被覆や土地利用の変化、自然災害の監視、多くの分野で利用されている
- EUのSentinel2のように定期的に更新され、パブリックドメインの月単位、週単位で衛星データを利用できるようになった

# 背景

- リモートセンシング画像は完全に自動解析ではなく、**手動または半自動解析に依存している**
  - **精度の低さ**が障壁となっている
- 専門家による**アノテーションが必要**であり、ラベリングの前例があまりないリモートセンシング・アプリケーションでは**ロングテールが存在する**

# 導入

- 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障害となっているのでは？
  - ViTやCLIPのような最先端アーキテクチャでは、Microsoft COCO（328K画像）などで学習しているが、**既存のリモートセンシングベンチマークは10K画像未満**である
  - ベンチマークが道路、船舶、作物などの個別のカテゴリで断片的であり、**多くのカテゴリにまたがるベンチマークはない**
  - 大規模かつ一元化され、アクセス可能なベンチマークがないため、**タスク間での転移学習が困難**

# 導入

- リモートセンシング画像モデルを改善するための大規模データセットの作成を目的とし、Sentinel-2とNAIPの画像を組み合わせた**SatlasPretrain**を提案
- 137の多様なカテゴリと7つのラベルタイプから成る、302Mのラベルを付与した
- ImageNetと比較して、下流タスクで平均性能を18%向上させた

# 既存のリモートセンシングデータセット

	Types	Classes	Labels	Pixels	km <sup>2</sup>
<b>SatlasPretrain</b>	7	137	302222K	17003B	21320K
UCM [57]	1	21	2K	1B	1K
BigEarthNet [51]	1	43	1750K	9B	850K
AID [56]	1	30	10K	4B	14K
Million-AID [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

Table 1: Comparison of SATLASPRETRAIN against existing remote sensing datasets (K=thousands, B=billions). Types is number of label types and km<sup>2</sup> is area covered.

# 提案手法

- **SatlasPretrain**

1. **Scale:** 既存のデータセットよりも40倍の画像ピクセル、150倍のレベル
2. **Label:** 7タイプ、137カテゴリ
3. **Spatio-temporal and Label:**  
個々の画像に関連づけるのではなく、地理的座標と時間範囲で関連づける



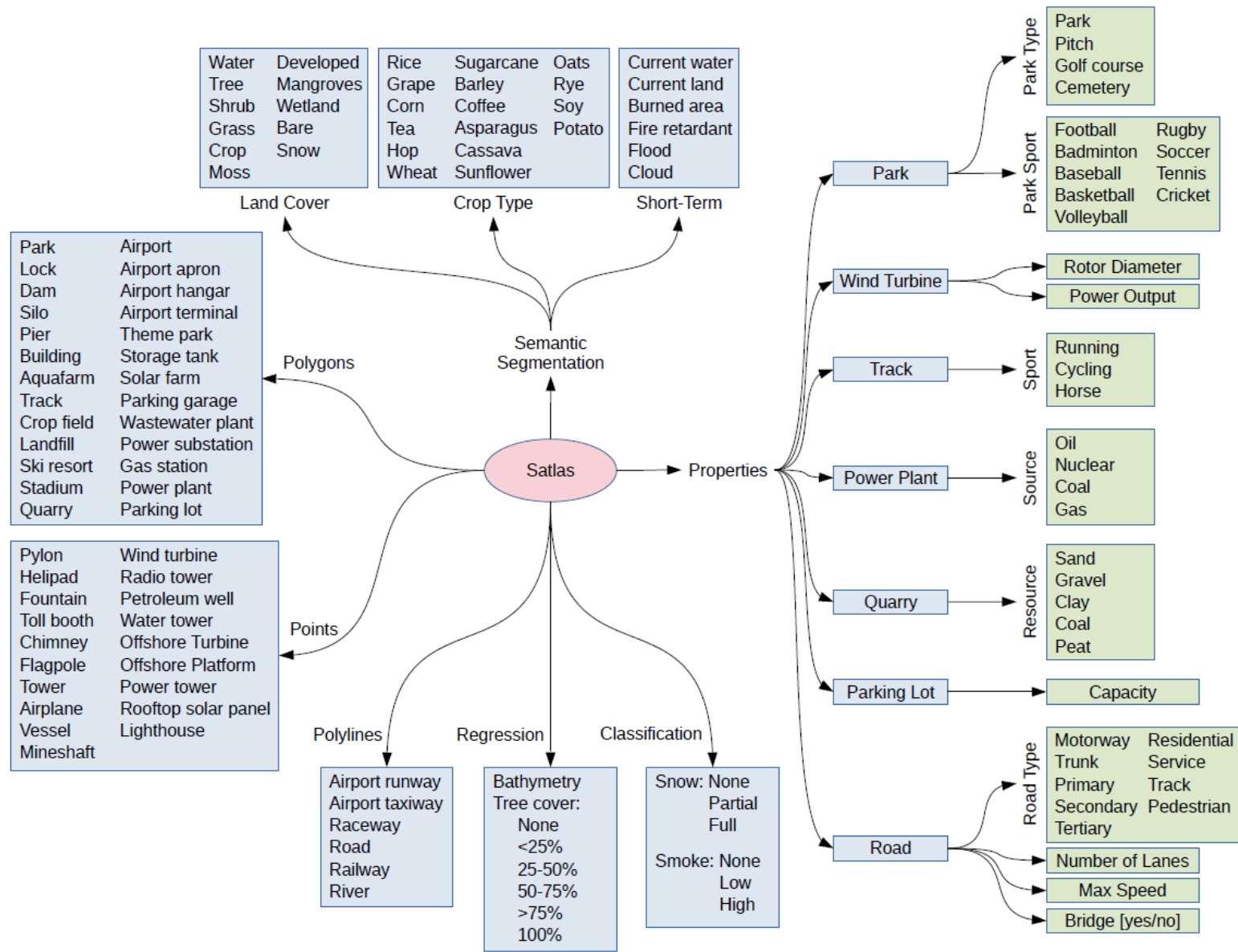


Figure 2: The 137 SATLASPRETRAIN categories, organized by the seven label types. Top: semantic segmentation. Left: polygons and points. Bottom: polylines, regression, and classification. Right: properties; for properties, we show the property category in green, and the point, polygon, or polyline category that the property is an attribute of in blue.

# 画像とデータ構造

- SatlasPretrainは856Kのタイルで構成
  - 25km<sup>2</sup>に及ぶ不連続な空間領域をカバー

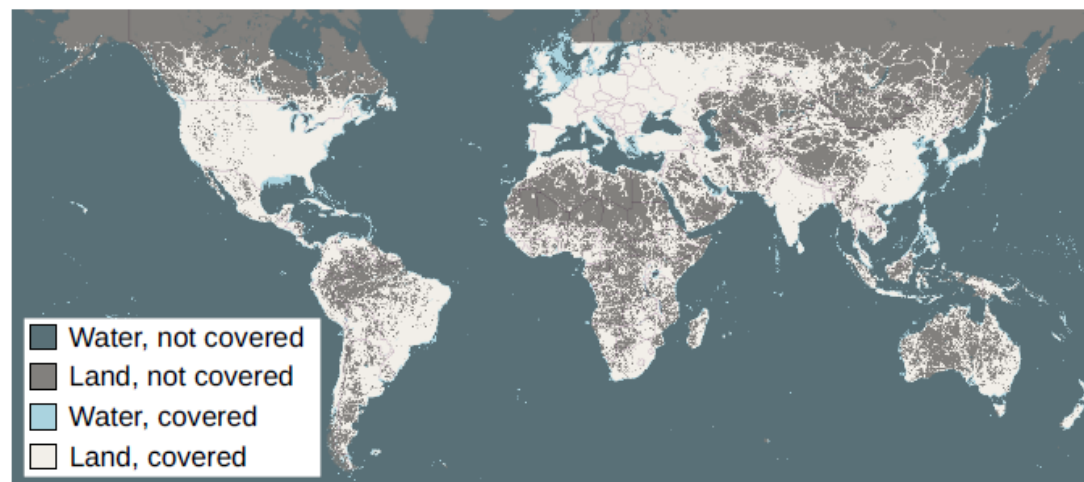


Figure 3: Geographic coverage of SATLASPRETRAIN, with bright pixels indicating locations covered by images and labels in the dataset. SATLASPRETRAIN spans all continents except Antarctica.

# 画像とデータ構造

- 既存のデータセット
  - 高解像度画像：0.5~2m/pixel
  - 低解像度画像：10m/pixel
- 高解像度画像は高い予測精度を実現できるが、低解像度画像は広範囲かつ頻繁に入手可能（実用的なアプリケーションではしばしば採用される）
  - SatlasPretrainでは両方を扱う

# 画像とデータ構造

## • SatlasPretrain

- 高解像度画像：8192×8192サイズ 46K (train: 45.5K, test: 512)
- 低解像度画像：512×512サイズ 856 K (train: 828K, test: 28K)

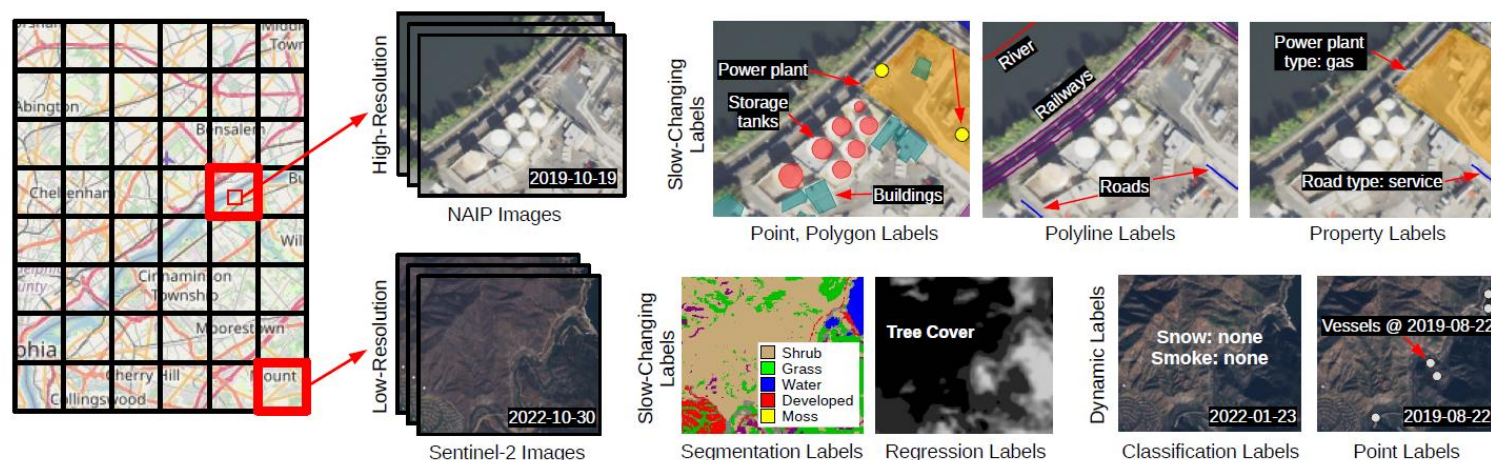


Figure 2: Overview of the SATLASPRETRAIN dataset. SATLASPRETRAIN consists of image time series and labels for 856K Web-Mercator tiles at zoom 13 (left). There are two image modes on which methods are trained and evaluated independently: high-resolution NAIP images (top) and low-resolution Sentinel-2 images (bottom). Labels may be slow-changing (corresponding to the most recent image at a tile) or dynamic (referencing a specific image and time).

# ラベリング

- ラベルタイプ

1. **Semantic Segmentation** 水、森林、開発地域など
2. **Regression** 水深、樹木の被覆率
3. **Points(Object Detection:OD)** 風力タービン、油井、船舶
4. **Polygons(Instance Segmentation)** ビル、ダム、養殖場
5. **Polylines** 道路、河川、鉄道
6. **Properties** 風力タービンのローター直径
7. **Classification** 煙の濃度、雪の有無

# ラベリング

- 既存データセットの活用
  - OpenStreetMap
  - NOAA lidar scans
  - WorldCover
  - Microsoft Buildings
  - C2S
- 専門家による新しいアノテーション

# SatlasNet

- **SatlasNet**

- 7つすべてのラベルタイプを学習可能とする統一モデル
- タスクに特化した出力ヘッドを利用した研究や、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を採用

# SatlasNet

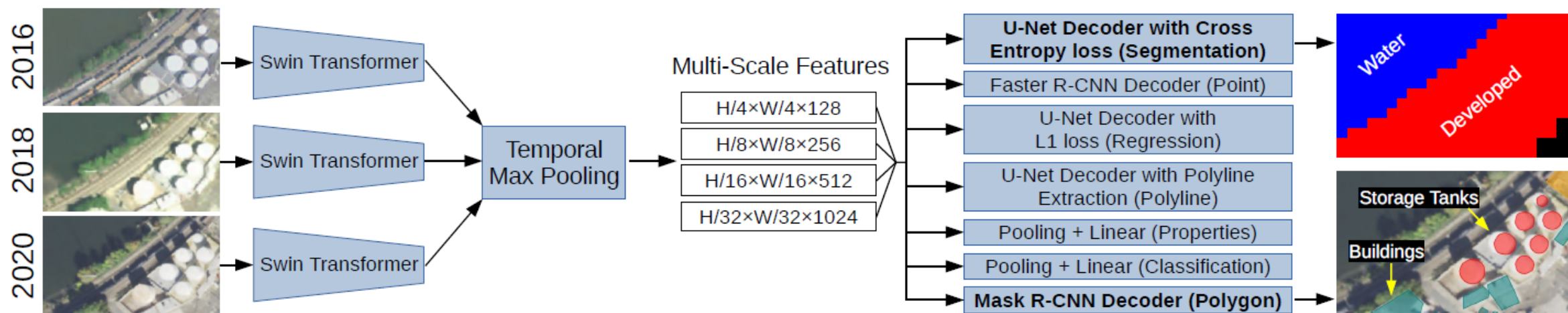


Figure 4: Model architecture of SATLASNET. A separate head is used to predict outputs for each label type. We visualize example outputs from two such heads (segmentation and polygon).



# SatlasNet (補足)

- Swin Transformer自体はViTでの問題点である多様な解像度や物体スケールの変化を改善したモデルである
- 画像を時系列に入力する手法は季節ごとによる農作地の変化に対応するために提案された衛星画像時系列(SITS)を用いた
  - 「**Panoptic Segmentation of Satellite Image Time Series with Convolutional Temporal Attention Networks**」

# 評価

- 学習時間（事前学習）
  - NVIDIA Titan X Pascal GPU / MoCo
  - 5 days(batch size 32)

# 評価

- SatlasNetと 8 つの従来手法を比較

Method	High-Resolution NAIP Images						Low-Resolution Sentinel-2 Images						
	Seg ↑	Reg ↓	Pt ↑	Pgon ↑	Pline ↑	Prop ↑	Seg ↑	Reg ↓	Pt ↑	Pgon ↑	Pline ↑	Prop ↑	Cls ↑
PSPNet (ResNext-101) [59]	77.8	15.0	-	-	53.2	-	62.1	16.2	-	-	30.7	-	-
LinkNet (ResNext-101) [15]	77.3	12.9	-	-	61.0	-	61.1	14.1	-	-	41.4	-	-
DeepLabv3 (ResNext-101) [16]	80.1	10.6	-	-	59.8	-	61.8	13.9	-	-	44.7	-	-
ResNet-50 [32]	-	-	-	-	-	87.6	-	-	-	-	-	70.3	97
ViT-Large [26]	-	-	-	-	-	78.1	-	-	-	-	-	66.9	99
Swin-Base [38]	-	-	-	-	-	87.1	-	-	-	-	-	69.4	99
Mask R-CNN (ResNet-50) [31]	-	-	27.6	30.4	-	-	-	-	22.0	12.3	-	-	-
Mask R-CNN (Swin-Base) [31]	-	-	30.4	31.5	-	-	-	-	25.6	15.2	-	-	-
ISTR [34]	-	-	2.0	4.9	-	-	-	-	1.2	1.4	-	-	-
SatlasNet (single-image, per-type)	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
SatlasNet (single-image, joint)	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
SatlasNet (single-image, fine-tuned)	79.8	<b>7.2</b>	32.3	33.0	<b>62.4</b>	<b>89.5</b>	65.3	9.0	27.4	16.3	45.9	<b>80.0</b>	99
SatlasNet (multi-image, per-type)	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
SatlasNet (multi-image, joint)	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
SatlasNet (multi-image, fine-tuned)	<b>81.0</b>	7.6	<b>33.2</b>	<b>34.1</b>	61.1	89.2	<b>69.7</b>	<b>7.8</b>	<b>32.0</b>	<b>20.2</b>	<b>50.4</b>	<b>80.0</b>	99

Table 2: Results on the SATLASPRETRAIN test set for the high- and low-resolution image modes. We break down results by label type: segmentation (Seg), regression (Reg), points (Pt), polygons (Pgon), polylines (Pline), properties (Prop), and classification (Cls). We show absolute error for Reg (lower is better), and accuracy for the others (higher is better).

# 評価

- タイプごとの従来手法の性能との比較

Method	UCM		RESISC45		AID		FMoW		Mass Roads		Mass Buildings		Airbus Ships		Average	
	50	All	50	All	50	All	50	All	50	All	50	All	50	All	50	All
Random Initialization	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
ImageNet [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
BigEarthNet [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
MillionAID [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	<b>0.99</b>	0.28	0.95	0.33	0.83	0.03	0.30	<b>0.82</b>	0.86	0.84	0.87	<b>0.62</b>	0.75	0.50	0.79
iSAID [58]	0.60	0.97	0.29	0.97	0.34	0.86	0.04	0.30	<b>0.82</b>	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
SatlasPretrain	<b>0.83</b>	<b>0.99</b>	<b>0.36</b>	<b>0.98</b>	<b>0.42</b>	<b>0.88</b>	<b>0.06</b>	<b>0.44</b>	<b>0.82</b>	<b>0.87</b>	<b>0.87</b>	<b>0.88</b>	0.56	<b>0.80</b>	<b>0.56</b>	<b>0.83</b>

Table 3: Results on seven downstream tasks when fine-tuned with 50 examples (50) or the entire downstream dataset (All). Accuracy is reported for UCM, RESISC45, and AID while F1 Score is reported for FMoW, Mass Roads, Mass Buildings, and Airbus Ships. SATLASPRETRAIN pre-training improves average accuracy across the tasks by 6% over the next best baseline.

# 評価



Figure 5: Qualitative results on SATLASPRETRAIN. Rightmost: a failure case where SATLASNET detects only 1/5 oil wells.

# 結論

- 既存のデータセットの規模とラベルの多様性を改善することで、効果的な大規模データセットを作成できた
- SatlasPretrainの事前学習で、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%も平均精度が向上
- ラベ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も容易に適用できることが示された