SatlasPretrain: A Large-Scale Dataset for Remote Sensing Image Understanding Supplementary Material

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

The supplementary material is organized as follows:

- In Section A, we detail the SATLASPRETRAIN categories and how labels for each category were annotated. We also perform a per-category accuracy analysis on the final labels in SATLASPRETRAIN; we find that, despite the large scale of the dataset, the labels in each category are highly accurate.
- In Section B, we provide additional details about the experiments on downstream tasks, including per-dataset results.
- In Section C, we discuss synthetic aperture radar images captured by Sentinel-1 that are present in the dataset (but not used in our experiments).
- In Section D, we present additional analysis about geographical accuracy biases of models trained on SATLASPRETRAIN and downstream accuracy when pre-training on individual label types.
- In Section E, we show examples of each category in SATLASPRETRAIN, along with corresponding SATLASNET outputs.

A. SATLASPRETRAIN Labels

Here, we provide additional details about the SATLASPRETRAIN categories and how labels for each category were annotated:

- In Section A.1, we show various statistics about the SATLASPRETRAIN categories and labels.
- In Section A.2, we discuss the potential use cases and practical applications of the SATLASPRETRAIN categories.
- In Section A.3, we detail the annotation tool that we used for the categories labeled with new annotation (12 categories with expert annotation and another 9 categories with Amazon Mechanical Turk), and discuss customizations that we made for individual categories to maximize annotation quality.
- In Section A.4, we detail OpenStreetMap processing, including heuristics that we used to select tiles and filter labels, and additional annotation we conducted to improve recall in the test set.
- In Section A.5, we perform an accuracy analysis for each of the 137 categories.
- In Section A.6, we detail how the test splits were selected.

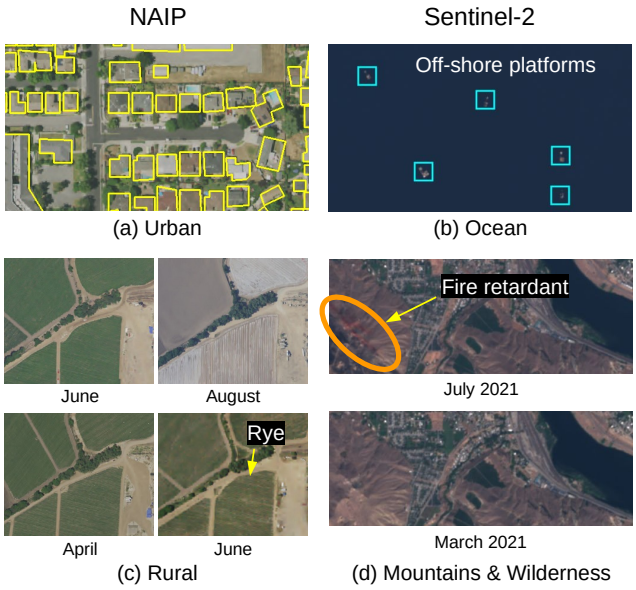


Figure 1: SATLASPRETRAIN incorporates both high-resolution NAIP and low-resolution Sentinel-2 images that exhibit very different characteristics. In (c) and (d), comparing multiple images helps distinguish rye crops and a fire retardant drop.

A.1. Statistics

Figure 1 shows examples of the low-resolution Sentinel-2 and high-resolution NAIP images in SATLASPRETRAIN.

Figure 2 summarizes the SATLASPRETRAIN categories, organized by the seven label types.

For each category, Table 1 details the label type, data source, the number of tiles where the category is valid (but may have zero labels), the number of tiles where the cate-

gory has at least one label, and the number of labels. Table 1 also specifies a "use case" for each category; we discuss each of these use cases further in the next section.

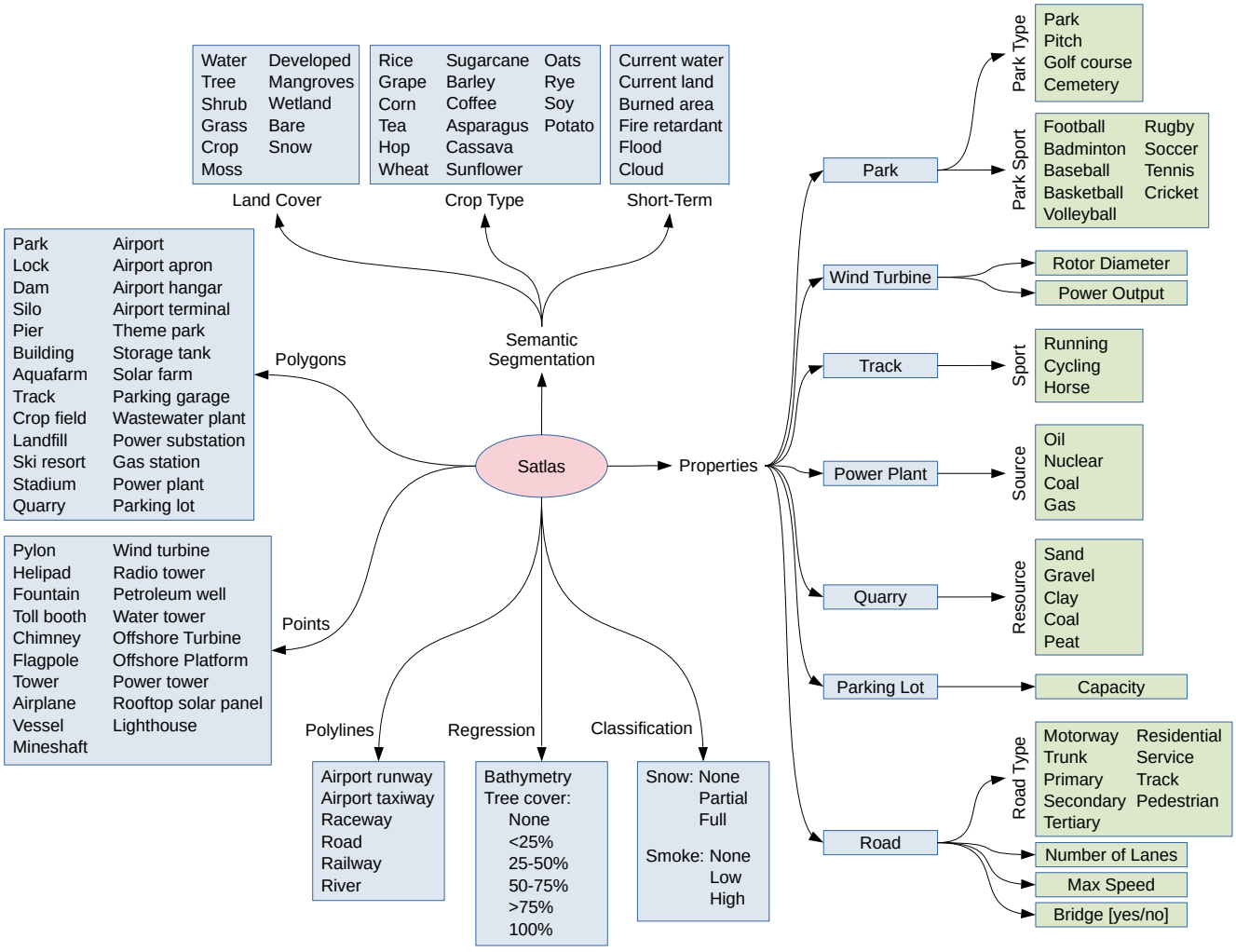


Figure 2: The 137 SATLASPRETRAIN categories, organized by the seven label types. Top: semantic segmentation. Left: polygons and points. Bottom: polylines, regression, and classification. Right: properties; for properties, we show the property category in green, and the point, polygon, or polyline category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

Category	Label Type	Dynamic?	Data Source	# Tiles Valid	# Tiles Nonzero	# Labels	Img Mode	Use Case
Aquafarm	Polygon	Slow-Changing	OpenStreetMap	740297	7555	66328	Both	Infrastructure
Lock	Polygon	Slow-Changing	OpenStreetMap	740297	2079	3281	Both	Infrastructure
Dam	Polygon	Slow-Changing	OpenStreetMap	740297	40016	53030	Both	Infrastructure
Solar Farm	Polygon	Slow-Changing	OpenStreetMap	740297	18278	32803	Both	Infrastructure
Power Plant	Polygon	Slow-Changing	OpenStreetMap	740297	8296	9538	Both	Infrastructure
Gas Station	Polygon	Slow-Changing	OpenStreetMap	740297	89443	190697	High-res	Digital Maps
Park	Polygon	Slow-Changing	OpenStreetMap	740297	385486	3317581	Both	Digital Maps
Parking Garage	Polygon	Slow-Changing	OpenStreetMap	740297	7674	18053	Both	Digital Maps
Parking Lot	Polygon	Slow-Changing	OpenStreetMap	740297	247031	3884224	Both	Digital Maps
Landfill	Polygon	Slow-Changing	OpenStreetMap	740297	31704	47478	Both	Infrastructure
Quarry	Polygon	Slow-Changing	OpenStreetMap	740297	122257	207694	Both	Infrastructure
Stadium	Polygon	Slow-Changing	OpenStreetMap	740297	36320	48305	Both	Digital Maps
Airport	Polygon	Slow-Changing	OpenStreetMap	740297	35773	36761	Both	Infrastructure
Airport Apron	Polygon	Slow-Changing	OpenStreetMap	740297	18484	41223	Both	Infrastructure
Airport Hangar	Polygon	Slow-Changing	OpenStreetMap	740297	10429	54280	High-res	Infrastructure
Airport Terminal	Polygon	Slow-Changing	OpenStreetMap	740297	6067	8918	Both	Infrastructure
Ski Resort	Polygon	Slow-Changing	OpenStreetMap	740297	4845	6196	Both	Digital Maps
Theme Park	Polygon	Slow-Changing	OpenStreetMap	740297	5631	6520	Both	Digital Maps
Storage Tank	Polygon	Slow-Changing	OpenStreetMap	740297	58560	411079	High-res	Infrastructure
Silo	Polygon	Slow-Changing	OpenStreetMap	740297	32737	194172	High-res	Infrastructure
Track	Polygon	Slow-Changing	OpenStreetMap	740297	54969	87748	Both	Digital Maps
Wastewater Plant	Polygon	Slow-Changing	OpenStreetMap	740297	53541	74155	Both	Infrastructure
Power Substation	Polygon	Slow-Changing	OpenStreetMap	740297	199395	451149	High-res	Infrastructure
Pier	Polygon	Slow-Changing	OpenStreetMap	740297	3611	24628	High-res	Digital Maps
Crop Field	Polygon	Slow-Changing	OpenStreetMap	740297	349482	6577038	Both	Infrastructure
Building	Polygon	Slow-Changing	Microsoft Buildings	33632	33632	48746983	High-res	Digital Maps
Wind Turbine	Point	Slow-Changing	OpenStreetMap	523337	46141	354120	Both	Infrastructure
Lighthouse	Point	Slow-Changing	OpenStreetMap	523337	7548	9541	Both	Infrastructure
Mineshaft	Point	Slow-Changing	OpenStreetMap	523337	3409	10703	Both	Infrastructure
Aerialway Pylon	Point	Slow-Changing	OpenStreetMap	523337	6416	121108	Both	Infrastructure
Helipad	Point	Slow-Changing	OpenStreetMap	523337	33661	53858	Both	Infrastructure
Fountain	Point	Slow-Changing	OpenStreetMap	523337	7452	28139	High-res	Digital Maps
Toll Booth	Point	Slow-Changing	OpenStreetMap	523337	1713	4257	High-res	Digital Maps
Chimney	Point	Slow-Changing	OpenStreetMap	523337	2065	6542	High-res	Infrastructure
Communications Tower	Point	Slow-Changing	OpenStreetMap	523337	2191	2948	Both	Infrastructure
Flagpole	Point	Slow-Changing	OpenStreetMap	523337	7350	26916	High-res	Digital Maps
Petroleum Well	Point	Slow-Changing	OpenStreetMap	523337	19271	158672	Both	Infrastructure
Water Tower	Point	Slow-Changing	OpenStreetMap	523337	87379	130089	Both	Infrastructure
Offshore Wind Turbine	Point	Slow-Changing	New Annotation - Expert	523337	1327	8726	Both	Infrastructure
Offshore Platform	Point	Slow-Changing	New Annotation - Expert	523337	1393	2275	Both	Infrastructure
Power Tower	Point	Slow-Changing	OpenStreetMap	523337	404398	8347197	Both	Infrastructure
Airplane	Point	Dynamic	New Annotation - AMT	374	374	26568	High-res	Airplane/Vessel
Rooftop Solar Panel	Point	Slow-Changing	New Annotation - AMT	27	27	25413	High-res	Infrastructure
Vessel	Point	Dynamic	New Annotation - Expert	2032	2032	5002	Low-res	Airplane/Vessel
Airport Runway	Polyline	Slow-Changing	OpenStreetMap	266037	65248	80987	Both	Infrastructure
Airport Taxiway	Polyline	Slow-Changing	OpenStreetMap	266037	24582	267064	Both	Infrastructure
Raceway	Polyline	Slow-Changing	OpenStreetMap	266037	12769	35333	Both	Digital Maps
Road	Polyline	Slow-Changing	OpenStreetMap	266037	204987	35880795	Both	Digital Maps
Railway	Polyline	Slow-Changing	OpenStreetMap	266037	119970	694271	Both	Digital Maps
River	Polyline	Slow-Changing	OpenStreetMap	266037	115120	246563	Both	Digital Maps
Land Cover - Water	Segment	Slow-Changing	WorldCover	59032	55152	2642496	Both	Land Cover
Land Cover - Developed	Segment	Slow-Changing	WorldCover	59032	58534	4284426	Both	Land Cover
Land Cover - Tree	Segment	Slow-Changing	WorldCover	59032	58267	16669429	Both	Land Cover
Land Cover - Shrub	Segment	Slow-Changing	WorldCover	59032	18075	2308571	Both	Land Cover
Land Cover - Grass	Segment	Slow-Changing	WorldCover	59032	58383	16954305	Both	Land Cover
Land Cover - Crop	Segment	Slow-Changing	WorldCover	59032	57172	7251955	Both	Land Cover
Land Cover - Bare	Segment	Slow-Changing	WorldCover	59032	58505	8770137	Both	Land Cover
Land Cover - Snow	Segment	Slow-Changing	WorldCover	59032	89	1460	Both	Land Cover
Land Cover - Wetland	Segment	Slow-Changing	WorldCover	59032	22836	547734	Both	Land Cover

Table 1: Details for each Satlas category. The number of valid tiles indicates tiles where the category is annotated; this includes tiles that contain zero instances of the category. The number of non-zero tiles includes only tiles where there is at least one label in the category. Img Mode indicates whether the category is used in both image modes or just one: (a) some slow-changing categories are excluded when using low-resolution images because they are not visible; and (b) dynamic categories are always labeled in only one image type. The use cases listed in the last column are discussed further in the next section.

Category	Label Type	Dynamic?	Data Source	# Tiles Valid	# Tiles Nonzero	# Labels	Img Mode	Use Case
Land Cover - Mangroves	Segment	Slow-Changing	WorldCover	59032	657	21645	Both	Land Cover
Land Cover - Moss	Segment	Slow-Changing	WorldCover	59032	542	53586	Both	Land Cover
Crop Type - Rice	Segment	Slow-Changing	OpenStreetMap	19215	10685	259013	Both	Crop Type
Crop Type - Grape	Segment	Slow-Changing	OpenStreetMap	19215	294	568	Both	Crop Type
Crop Type - Corn	Segment	Slow-Changing	OpenStreetMap	19215	2874	16199	Both	Crop Type
Crop Type - Sugarcane	Segment	Slow-Changing	OpenStreetMap	19215	2097	8399	Both	Crop Type
Crop Type - Tea	Segment	Slow-Changing	OpenStreetMap	19215	553	3060	Both	Crop Type
Crop Type - Hop	Segment	Slow-Changing	OpenStreetMap	19215	626	6859	Both	Crop Type
Crop Type - Wheat	Segment	Slow-Changing	OpenStreetMap	19215	1342	5281	Both	Crop Type
Crop Type - Soy	Segment	Slow-Changing	OpenStreetMap	19215	615	4981	Both	Crop Type
Crop Type - Barley	Segment	Slow-Changing	OpenStreetMap	19215	289	1875	Both	Crop Type
Crop Type - Oats	Segment	Slow-Changing	OpenStreetMap	19215	2	2	Both	Crop Type
Crop Type - Rye	Segment	Slow-Changing	OpenStreetMap	19215	10	15	Both	Crop Type
Crop Type - Cassava	Segment	Slow-Changing	OpenStreetMap	19215	64	134	Both	Crop Type
Crop Type - Potato	Segment	Slow-Changing	OpenStreetMap	19215	48	120	Both	Crop Type
Crop Type - Sunflower	Segment	Slow-Changing	OpenStreetMap	19215	157	332	Both	Crop Type
Crop Type - Asparagus	Segment	Slow-Changing	OpenStreetMap	19215	65	127	Both	Crop Type
Crop Type - Coffee	Segment	Slow-Changing	OpenStreetMap	19215	351	916	Both	Crop Type
Coast - Water	Segment	Dynamic	New Annotation - AMT	3507	3507	26201	High-res	Coastline
Coast - Land	Segment	Dynamic	New Annotation - AMT	3507	3507	26201	High-res	Coastline
Burned Area	Segment	Dynamic	New Annotation - AMT	399	319	883	Low-res	Wildfire
Fire Retardant	Segment	Dynamic	New Annotation - AMT	399	131	415	Low-res	Wildfire
Flood	Segment	Dynamic	C2S	2751	1829	45110	Low-res	Flood/Cloud
Cloud	Segment	Dynamic	C2S	2751	1227	8463	Low-res	Flood/Cloud
Bathymetry	Regress	Slow-Changing	NOAA Lidar Scans	5123	5123	21306701	Both	Bathymetry
Tree Cover - None	Regress	Slow-Changing	New Annotation - Expert	2831	2831	1296	Both	Tree Cover
Tree Cover - Sparse	Regress	Slow-Changing	New Annotation - Expert	2831	2831	2450	Both	Tree Cover
Tree Cover - Low	Regress	Slow-Changing	New Annotation - Expert	2831	2831	455	Both	Tree Cover
Tree Cover - Medium	Regress	Slow-Changing	New Annotation - Expert	2831	2831	284	Both	Tree Cover
Tree Cover - High	Regress	Slow-Changing	New Annotation - Expert	2831	2831	248	Both	Tree Cover
Tree Cover - Full	Regress	Slow-Changing	New Annotation - Expert	2831	2831	408	Both	Tree Cover
Wind Turb. - Rotor Diam.	Property	Slow-Changing	OpenStreetMap	4067	4067	27661	Both	Infrastructure
Wind Turbine - Power	Property	Slow-Changing	OpenStreetMap	17506	17506	122398	Both	Infrastructure
Parking Lot - Capacity	Property	Slow-Changing	OpenStreetMap	70661	70661	337542	Both	Digital Maps
Track - Running	Property	Slow-Changing	OpenStreetMap	32670	26277	41950	Both	Digital Maps
Track - Cycling	Property	Slow-Changing	OpenStreetMap	32670	2321	4612	Both	Digital Maps
Track - Horse	Property	Slow-Changing	OpenStreetMap	32670	5170	7415	Both	Digital Maps
Road Type - Motorway	Property	Slow-Changing	OpenStreetMap	238868	33340	577135	Both	Digital Maps
Road Type - Trunk	Property	Slow-Changing	OpenStreetMap	238868	35179	303723	Both	Digital Maps
Road Type - Primary	Property	Slow-Changing	OpenStreetMap	238868	79694	776820	Both	Digital Maps
Road Type - Secondary	Property	Slow-Changing	OpenStreetMap	238868	105240	1003712	Both	Digital Maps
Road Type - Tertiary	Property	Slow-Changing	OpenStreetMap	238868	135065	1065756	Both	Digital Maps
Road Type - Residential	Property	Slow-Changing	OpenStreetMap	238868	223834	9671420	Both	Digital Maps
Road Type - Service	Property	Slow-Changing	OpenStreetMap	238868	206058	17566909	Both	Digital Maps
Road Type - Track	Property	Slow-Changing	OpenStreetMap	238868	146859	1108582	Both	Digital Maps
Road Type - Pedestrian	Property	Slow-Changing	OpenStreetMap	238868	76743	5070309	Both	Digital Maps
Road - Number of Lanes	Property	Slow-Changing	OpenStreetMap	147367	147367	3331621	Both	Digital Maps
Road - Max Speed	Property	Slow-Changing	OpenStreetMap	110753	110753	2183903	Both	Digital Maps
Road - Bridge	Property	Slow-Changing	OpenStreetMap	266037	100205	518177	Both	Digital Maps
Road - Not Bridge	Property	Slow-Changing	OpenStreetMap	266037	266037	35362618	Both	Digital Maps
Power Plant - Oil	Property	Slow-Changing	OpenStreetMap	8484	756	867	Both	Infrastructure
Power Plant - Nuclear	Property	Slow-Changing	OpenStreetMap	8484	370	400	Both	Infrastructure
Power Plant - Coal	Property	Slow-Changing	OpenStreetMap	8484	3368	3723	Both	Infrastructure
Power Plant - Gas	Property	Slow-Changing	OpenStreetMap	8484	4117	4762	Both	Infrastructure
Quarry - Sand	Property	Slow-Changing	OpenStreetMap	7910	2226	2796	Both	Infrastructure
Quarry - Gravel	Property	Slow-Changing	OpenStreetMap	7910	1427	1855	Both	Infrastructure
Quarry - Clay	Property	Slow-Changing	OpenStreetMap	7910	877	2836	Both	Infrastructure
Quarry - Coal	Property	Slow-Changing	OpenStreetMap	7910	1640	2099	Both	Infrastructure
Quarry - Peat	Property	Slow-Changing	OpenStreetMap	7910	1803	3048	Both	Infrastructure

Table 1 (continued): Details for each Satlas category. The number of valid tiles indicates tiles where the category is annotated; this includes tiles that contain zero instances of the category. The number of non-zero tiles includes only tiles where there is at least one label in the category. Img Mode indicates whether the category is used in both image modes or just one: (a) some slow-changing categories are excluded when using low-resolution images because they are not visible; and (b) dynamic categories are always labeled in only one image type. The use cases listed in the last column are discussed further in the next section.

Category	Label Type	Dynamic?	Data Source	# Tiles Valid	# Tiles Nonzero	# Labels	Img Mode	Use Case
Park Type - City Park	Property	Slow-Changing	OpenStreetMap	578761	226794	1168319	Both	Digital Maps
Park Type - Pitch	Property	Slow-Changing	OpenStreetMap	578761	341244	2045889	Both	Digital Maps
Park Type - Golf Course	Property	Slow-Changing	OpenStreetMap	578761	44141	64919	Both	Digital Maps
Park Type - Cemetery	Property	Slow-Changing	OpenStreetMap	578761	325478	480511	Both	Digital Maps
Park Sport - American Football	Property	Slow-Changing	OpenStreetMap	273915	13197	19227	Both	Digital Maps
Park Sport - Badminton	Property	Slow-Changing	OpenStreetMap	273915	1717	4259	Both	Digital Maps
Park Sport - Baseball	Property	Slow-Changing	OpenStreetMap	273915	44641	151679	Both	Digital Maps
Park Sport - Basketball	Property	Slow-Changing	OpenStreetMap	273915	74755	208555	Both	Digital Maps
Park Sport - Cricket	Property	Slow-Changing	OpenStreetMap	273915	6697	12265	Both	Digital Maps
Park Sport - Rugby	Property	Slow-Changing	OpenStreetMap	273915	5151	9511	Both	Digital Maps
Park Sport - Soccer	Property	Slow-Changing	OpenStreetMap	273915	194269	504876	Both	Digital Maps
Park Sport - Tennis	Property	Slow-Changing	OpenStreetMap	273915	116611	419207	Both	Digital Maps
Park Sport - Volleyball	Property	Slow-Changing	OpenStreetMap	273915	14857	22681	Both	Digital Maps
Snow - None	Classify	Dynamic	New Annotation - Expert	6340	5691	5699	Low-res	Snow
Snow - Partial	Classify	Dynamic	New Annotation - Expert	6340	248	248	Low-res	Snow
Snow - Full	Classify	Dynamic	New Annotation - Expert	6340	401	401	Low-res	Snow
Smoke - None	Classify	Dynamic	New Annotation - AMT	197	177	440	Low-res	Wildfire
Smoke - Low	Classify	Dynamic	New Annotation - AMT	197	45	111	Low-res	Wildfire
Smoke - High	Classify	Dynamic	New Annotation - AMT	197	12	14	Low-res	Wildfire

Table 1 (continued): Details for each Satlas category. The number of valid tiles indicates tiles where the category is annotated; this includes tiles that contain zero instances of the category. The number of non-zero tiles includes only tiles where there is at least one label in the category. Img Mode indicates whether the category is used in both image modes or just one: (a) some slow-changing categories are excluded when using low-resolution images because they are not visible; and (b) dynamic categories are always labeled in only one image type. The use cases listed in the last column are discussed further in the next section.

A.2. Use Cases

We now detail the potential use cases and practical applications of SATLASPRETRAIN categories. We organize our discussion around the use cases that appear in the "use case" column in Table 1 on the preceding page.

Land Cover. Land cover and land use maps describe a combination of the earth's surface material (e.g., trees, water, asphalt, or ice) and how the land is used by humans (e.g., developed areas versus uninhabited areas). We process the WorldCover map [6] to derive 11 land cover and land use categories, ranging from barren land to developed areas. Tracking changes in land cover has numerous applications, such as monitoring urban expansion, monitoring deforestation, and identifying growing and receding water bodies.

Timestamped Coastline. We use Amazon Mechanical Turk to annotate 10K images for the current coastline contour visible in each image. While land cover maps include a category for water bodies, they focus on a long time horizon, labeling water based on the mean high water level (i.e., the average level of water at high tide). The actual water level can vary greatly both through the tide cycle and over seasonal changes. For the timestamped coastline, we instead label the water level apparent from NAIP images captured at particular times. Current water levels are of interest to earth and environmental scientists, while the contours themselves can improve accuracy for detecting coastal features like near-shore vessels and platforms.

Crop Type. We derive crop type labels from OpenStreetMap. In 20K images, we process 16 crop type categories: rice, grape, corn, sugarcane, tea, hop, wheat, soy, barley, oats, rye, cassava, potato, sunflower, asparagus, and coffee. Crop type mapping is useful for both environmental monitoring and food price prediction.

Wildfire. We use Amazon Mechanical Turk to annotate 10K images for three wildfire-related categories in Sentinel-2 images: fire retardant drops, burned areas, and smoke presence. Automated methods for these categories can help inform both real-time fire response efforts, and longer-term forest management practices. Fire retardant is a substance dropped along the boundaries of a wildfire to contain or slow its progression. We define burned areas as grass or tree covered areas recently burned within the last three months of the current image. Note that we pick windows for annotation by selecting ones containing polygons from the National Interagency Fire Center historic perimeters.

Flood and Cloud Labels. We derive flood and cloud labels from C2S [1]. Cloud labels can improve performance on other categories by helping the model learn which parts of the image are occluded by clouds or cloud shadows. Real-time flood mapping is critical for coordinating disaster relief efforts.

Bathymetry. We process coastal bathymetry lidar scans from NOAA to derive bathymetry (water depth) data. Prior work has shown that water depth (bathymetry) up to around 10 m can be measured from optical satellite images [2]. Coastal bathymetry has several applications, including near-shore ship navigation and for monitoring changes in water reservoir volumes.

Tree Cover. Two domain experts on our team annotate tree cover canopy percentage in 3K images. Tree cover data can capture forest growth and loss on a more fine-grained scale than land cover maps. For example, consider a scenario where forest density is reduced due to selective logging, where many trees are cut but many others are left intact. This would not constitute a change in land cover, since the region is still tree-covered. However, the density reduction would still be captured by a change in tree cover canopy percentage.

Digital Map Features. We derive labels for digital map features like buildings, roads, and parking lots from OpenStreetMap. Automatically detecting features that are missing from digital maps can reduce the cost of improving those maps, along with the cost of keeping those maps up-to-date with changes in the real world.

Infrastructure. We derive labels for various energy and water infrastructure like power substations, wind turbines, oil wells, wastewater plants from OpenStreetMap. We also use Amazon Mechanical Turk to label rooftop solar panels, which are not accurately mapped in existing datasets due to their recent rapid growth. Similarly, two domain experts label off-shore platforms and off-shore wind turbines. Monitoring these features is useful for tracking renewable energy deployment progress, or for evaluating the environmental and public health impacts of activities like oil drilling.

Airplanes and Vessels. We use Amazon Mechanical Turk to label airplanes in NAIP images, and two domain experts on our team label vessels in Sentinel-2 images. Vessel detection in satellite images has recently attracted interest because it is useful for tackling illegal fishing [5]: for example, systems can alert authorities if they detect vessels with visible fishing equipment in marine protected areas.

Properties. We derive labels for 12 properties from OpenStreetMap. Five are numerical: the rotor diameter and electricity output of a wind turbine, the capacity of a parking lot, and the number of lanes and maximum speed of a road. The other seven are categorical, and span 35 categories: the type of a sports track (e.g. running, cycling), the type of a road (e.g. motorway, residential street), the type of a park (e.g. city park, sports pitch), the type of a power plant (e.g. oil, nuclear), the material produced by a quarry (e.g. sand, gravel), and the presence of a bridge on a road (binary).

Snow. Two domain experts on our team label the degree of snow accumulation (none, some, and substantial) in

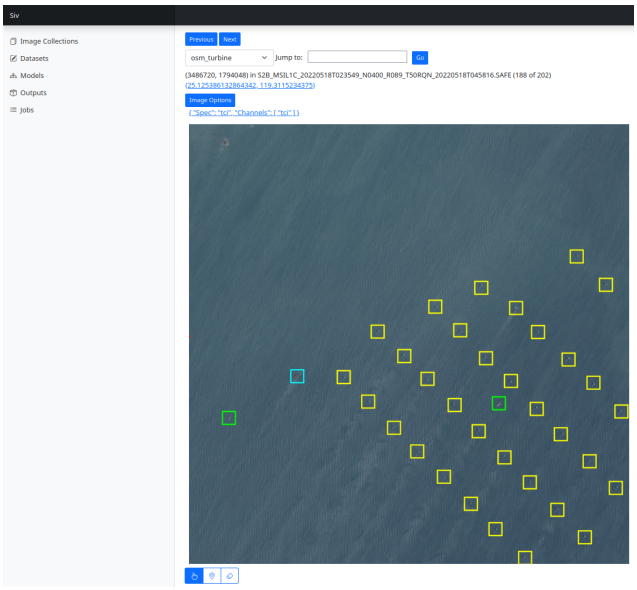


Figure 3: An overview of our annotation tool called Siv.

Sentinel-2 images. Tracking snow accumulation over time is useful for climate forecasting. In cities, monitoring snow accumulation and removal by plowing is useful for identifying and addresses biases in the deployment of snow plowing resources.

A.3. New Annotation

We label 12 categories with expert annotation and 9 categories with Amazon Mechanical Turk (AMT). To facilitate this annotation, we developed a dedicated annotation tool called Siv. A key feature of Siv is that it is highly customizable; this way, when a particular SATLASPRETRAIN category presents unique challenges that require a tailored interface to address, we can deploy Siv in an updated configuration for that category.

Below, we first detail Siv and the per-category customizations that we made to it for various categories. Then, for the categories that we annotated using AMT, we detail the number of workers we asked to label each tile.

A.3.1 Siv

We show a screenshot from Siv in Figure 3. The annotation interface displays a window of a remote sensing image, with current labels for the window overlayed on top (here, vessels are in green, offshore platforms in cyan, and offshore wind turbines in yellow). Metadata about the window, including its timestamp and longitude-latitude coordinates, are shown at the top. Users can press the longitude-latitude coordinates to view a list of images captured at the same location, or to view Google Maps or OpenStreetMap at that position. Users can also press Image Options to adjust the

Figure 4: We include a water mask derived from OpenStreetMap when labeling the timestamped coastline.

contrast of the image. The toolbar at the bottom allows toggling between selecting labels, adding labels, and removing labels. The type of label is configured outside of the web interface: Siv supports annotating points, polygons, polylines, and segmentation labels.

For AMT, we use a lighter and less cluttered version of Siv that lacks features like the per-window timestamp and coordinate metadata. We also include instructions that provide guidance on how the images should be labeled, and examples of different types of labels.

We make customizations to Siv for each category that we use it for, in order to maximize annotation quality. We detail these customizations below.

Timestamped Coastline. We found that, initially, AMT workers were occasionally not able to distinguish water bodies from other objects like marshes due to similar appearance in the NAIP imagery and a lack of experience with those images. Since our focus here was on labeling the coastline at a particular point in time, and not identifying previously unknown water bodies, we decided to include a water mask from OpenStreetMap that shows roughly where the water bodies are. This helped focus the workers on the water bodies intersecting the tile covered by the image. We show an example in Figure 4. Note that we pick windows for annotation by selecting ones that contain at least 20% water and at least 20% land, based on the OpenStreetMap-derived water mask.

Airplanes. We find that no customizations are needed for annotating airplanes accurately. We show an example in Figure 5. Note that we pick windows for annotation by selecting ones intersecting airports in OpenStreetMap.

Off-shore Platform, Off-shore Wind Turbines, and Vessels. For these categories, we find that it is useful to compare multiple images of the same location to distinguish moving objects like vessels from static objects like platforms and wind turbines. Thus, we customize Siv to enable

Figure 5: Interface for annotating airplanes. We include instructions telling AMT workers to label the point at which the wings connect with the fuselage.

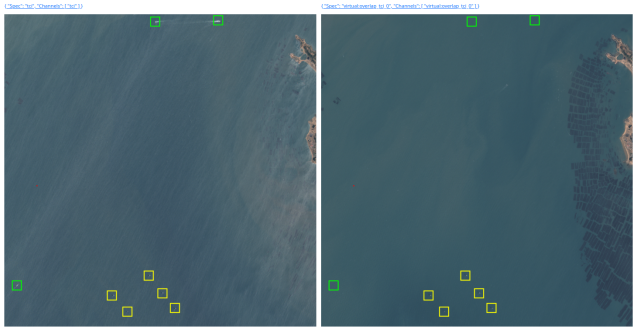


Figure 6: When labeling off-shore platforms and off-shore wind turbines, we customize Siv so that the arrow keys can be used to toggle between current (left) and historical (right) images.

domain experts to use the arrow keys to toggle between the current image that is being labeled and historical images of the same location. We show an example of this in Figure 6. Note that we pick windows for annotation by selecting ones containing wind turbines and platforms in OpenStreetMap, USGS Wind Turbine Database, or HIFLD Oil and Natural Gas Platforms Database.

Fire Retardant. Although fire retardant drops are visible in individual Sentinel-2 images, we find that including a historical image of the same location from a few months before a recorded wildfire is helpful to improve accuracy at localizing the fire retardant drops. We show an example in Figure 7.

Category	Workers
Airplane	1
Rooftop Solar Panel	3
Timestamped Coast Water	3
Timestamped Coast Land	3
Burned Area	3
Fire Retardant	3
Smoke (None)	3
Smoke (Low)	3
Smoke (High)	3

Table 2: The number of different AMT workers that annotated each tile for different SATLASPRETRAIN categories.

Smoke. We find that adding a historical image is helpful for annotating smoke presence as well. We show an example in Figure 8.

Burned Areas. We reuse the historical image display from fire retardant and smoke labeling discussed above. However, we find that burned areas are still often difficult to distinguish. Based on feedback from domain experts, we find that infrared channels captured by Sentinel-2 can provide a clearer indication of vegetation loss than the visible light RGB channels. Thus, we include an infrared image in the interface. Since we label burned areas using AMT, we include additional instructions in the interface discussing how the infrared image should be interpreted. We show an example in Figure 9.

Tree Cover. We originally tried to label contours of different tree cover densities in NAIP images. However, we determined that ambiguity in the exact position of the contour made it difficult to annotate the data in this way. Instead, we decided to have domain experts label the tree cover within a square overlayed on the NAIP image. We show an example in Figure 10.

A.3.2 Worker Parallelism

For each category in SATLASPRETRAIN that we annotated using Amazon Mechanical Turk workers, in Table 2, we show the number of workers that we asked to annotate each tile. When the number of workers is greater than one, we compute the final labels by performing majority voting over the multiple workers.

We decided the number of workers needed per tile on a per-category basis, by first having one worker annotate each tile, and then analyzing the label quality. For example, we found that airplanes were unambiguous enough that a

Figure 7: Interface for annotating fire retardant. It displays an image from roughly three months before a wildfire event on the left, and an image during or after the event in the middle.

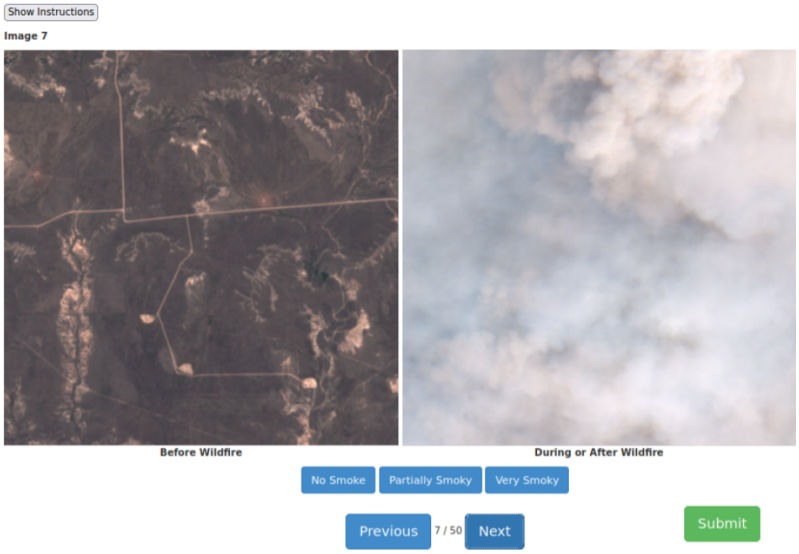


Figure 8: Interface for annotating smoke presence in Sentinel-2 images.

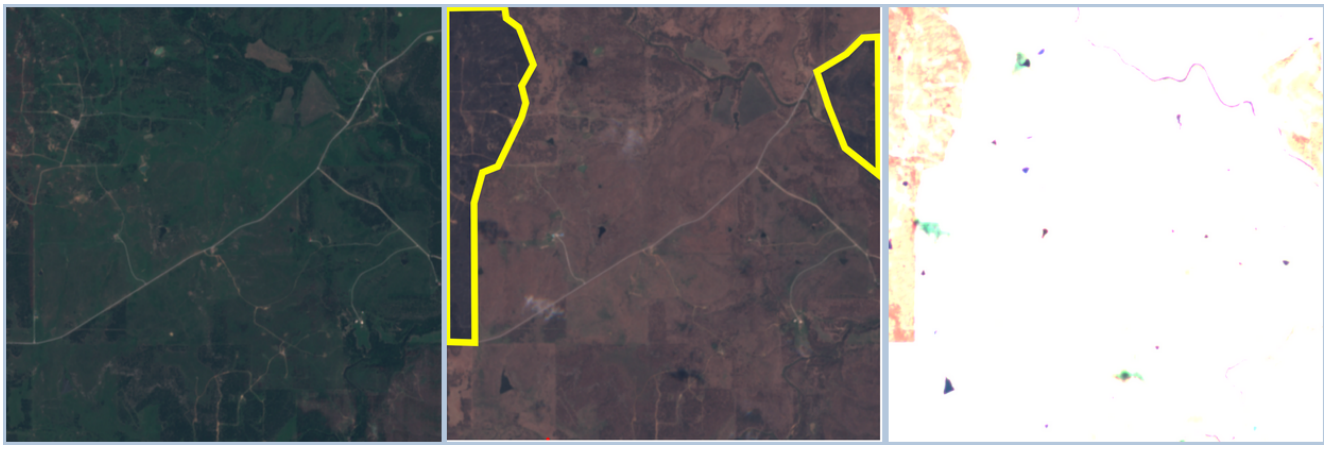


Figure 9: Interface for annotating burned areas. It displays a historical image on the left, an RGB image during or after the wildfire event in the middle, and an infrared image during or after the wildfire event on the right.

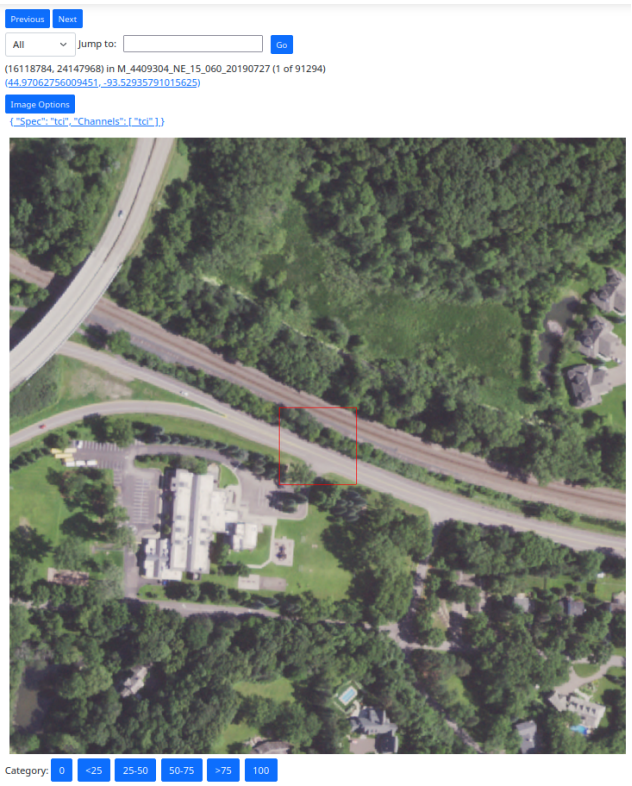


Figure 10: Interface for annotating tree cover. We show a red square at the center of the image, and ask domain experts to label the tree cover canopy percentage within the square. The portions of the image surrounding the square provide useful context about the region being visualized.

single worker was sufficient, while we had three workers label each tile for areas burned by wildfires.

A.4. OpenStreetMap Processing

A.4.1 Heuristics for Accurate Processing

OpenStreetMap is a collaborative map dataset. It consists of features that specify longitude and latitude coordinates. These features may be points (one longitude-latitude coordinate), ways (a sequence of coordinates specifying polylines or polygons), or relations (an ordered list of other features). Each feature can be tagged with arbitrary key-value pairs.

When initially processing OpenStreetMap features to extract labels for SATLASPRETRAIN, we found that the resulting labels exhibited high precision (i.e., there were very few incorrect labels), but at some tiles exhibited low recall. We found that these low-recall tiles typically had few distinct categories mapped in OpenStreetMap, e.g., tiles that only contained roads and buildings were more likely to be missing other features like water towers and power substations. Thus, we experimented with selecting subsets of tiles

crop
wind_turbine
aquafarm
lighthouse
lock
dam
solar_farm
power_plant
gas_station
park
parking_garage
parking_lot
landfill
quarry
mineshaft
stadium
airport
airport_runway
airport_taxiway
airport_apron
airport_hangar
airstrip
airport_terminal
ski_resort
theme_park
storage_tank
silo
track
raceway
wastewater_plant
water_park
aerialway_pylon
communications_tower
helipad
petroleum_well
water_tower
power_tower
power_substation

Table 3: List of "rarer" categories in OpenStreetMap that we use to select tiles for processing. We only collect labels from OpenStreetMap for a tile if there is at least one feature in OpenStreetMap in at least one of these categories contained in that tile.

by using various heuristics about the distinct categories appearing in a tile.

We found that the simple strategy of picking only tiles with "rarer" categories was highly effective: if a tile contains a wind turbine, it is most likely well mapped and has good coverage of all the other categories that we derive from OpenStreetMap. Thus, in our final processing, we only extract labels from OpenStreetMap for tiles with at least one

Category	Rule
storage_tank	No tag `location=underground`.
road	No tag `tunnel=yes`.
railway	No tag `tunnel=yes`.
railway	No tag `railway=abandoned`.

Table 4: List of rules we use to exclude OpenStreetMap features that do not correspond to visible physical objects. We organize the rules by category.

Category	# Detections	# New Labels
Chimney	498	0
Fountain	2626	882
Gas Station	137	96
Helipad	699	39
Parking Garage	149	55
Parking Lot	7220	101
Petroleum Well	1602	148
Pier	180	0
Power Substation	71	5
Silo	99	8
Storage Tank	167	43
Toll Booth	1449	0
Water Tower	447	3

Table 5: Remaining low-recall categories for which we performed additional annotation in the test set. We show the number of detections we went through, and the number of additional labels we added to SATLASPRETRAIN.

feature under one of the categories in Table 3.

We also incorporated several rules to exclude OpenStreetMap features that did not correspond to physical objects visible from remote sensing images. We enumerate these rules in Table 4.

A.4.2 Additional Test Set Annotation

Although the heuristics described above improve the recall of most categories in the dataset to higher than 95%, we find that 13 categories derived from OpenStreetMap still have low recall. We list the categories in Table 5. From an analysis of 1300 tiles, we determined that recall was still at least 80% for these categories, which we deemed sufficient for the training set: there are methods for learning from sparse labels, and large-scale training on noisy labels has produced models like CLIP that deliver state-of-the-art performance. We show the results of this analysis in Section A.5 below. However, we deemed that these 13 categories did not have sufficient recall for the test set.

Thus, to ensure a highly accurate test set, for each of these 13 categories, we trained an initial model on OSM labels and tuned its confidence threshold for high-recall low-precision detection; we then hand-labeled its predictions to add missing labels to the test set. For each of these categories, we list the number of detections we went through and the number that we added to the dataset in Table 5.

A.5. Accuracy Analysis

We perform an accuracy analysis of all 137 categories in the final SATLASPRETRAIN training set. For each category, we uniformly randomly sample 100 positive labels, and ask two domain experts to annotate the correctness of each label, i.e., (a) for slow-changing categories, whether the label is consistent with the most recent image at the tile; and (b) for dynamic categories, whether the label is consistent with the image that the label references. For the three categories with fewer than 100 labels (see Table 1), we use all the labels. For point, polygon, and polyline categories, we also uniformly randomly sample 100 images where the label is valid but not present, and ask domain experts to determine whether each image contains any missing labels under the category. For the other categories, tiles with missing labels are not well-defined; for example, for a group of classification categories like Smoke Presence, missing labels in "Smoke—None" are captured in the incorrect labels of the associated categories "Smoke—Low" and "Smoke—High".

We report the results of this analysis in Table 6. Overall, the analysis demonstrates that labels in SATLASPRETRAIN are highly accurate, with 116/137 categories having >99% precision and 131/137 categories having >95% precision. Labels like Offshore Wind Turbine and Tree Cover that were annotated by domain experts are consistently high-accuracy, showing that expert annotation is crucial for getting the best annotation quality. Labels derived from properties of OpenStreetMap features are also consistently high in quality, likely because contributors only tag the properties at features where they know the value; as a result, though, there may be geographical biases in where the property values are tagged.

Note that the missing tiles percentage is different from recall, which measures the fraction of actual labels that are covered by SATLASPRETRAIN labels. Instead, if a category is missing labels in $M\%$ of tiles with m missing labels per tile, and is present in SATLASPRETRAIN in $P\%$ of tiles with p existing labels per tile (with an assumption that there are no missing labels in the tiles where the category is labeled at least once), then the recall is $\frac{Pp}{Pp+Mm}$. Thus, one limitation of the analysis above is that the missing tiles percentage does not provide an accurate estimate of recall for categories that appear in fewer than 1% of tiles. However, obtaining a more accurate recall estimate for these categories would require analyzing far more than 100 negative tiles, which is cost-prohibitive with the number of categories in SATLASPRETRAIN.

Category	Precision	Missing Tiles
Aquafarm	100	0
Lock	100	0
Dam	100	0
Solar Farm	97	0
Power Plant	100	0
Gas Station	100	1
Park	100	0
Parking Garage	100	1
Parking Lot	100	5
Landfill	100	0
Quarry	100	0
Stadium	100	0
Airport	100	0
Airport Apron	100	0
Airport Hangar	100	0
Airport Terminal	100	0
Ski Resort	100	0
Theme Park	100	0
Storage Tank	100	0
Silo	100	0
Track	100	0
Wastewater Plant	99	0
Power Substation	100	0
Pier	99	0
Crop Field	100	0
Building	100	0
Wind Turbine	91	0
Lighthouse	98	0
Mineshaft	100	0
Aerialway Pylon	95	0
Helipad	99	0
Fountain	96	0
Toll Booth	99	0
Chimney	99	0
Communications Tower	80	0
Flagpole	100	0
Petroleum Well	99	0
Water Tower	100	0
Offshore Wind Turbine	100	0
Offshore Platform	100	0
Power Tower	98	0
Airplane	100	0
Rooftop Solar Panel	98	0
Vessel	98	0
Airport Runway	100	0
Airport Taxiway	100	0
Raceway	100	0
Road	100	1
Railway	100	0
River	100	0
Land Cover - Water	96	-
Land Cover - Developed	100	-
Land Cover - Tree	99	-
Land Cover - Shrub	96	-
Land Cover - Grass	100	-
Land Cover - Crop	96	-
Land Cover - Bare	94	-
Land Cover - Snow	96	-
Land Cover - Wetland	99	-

Table 6: Accuracy analysis of Satlas categories. Precision indicates the percentage of labels that are correct, as determined by two domain experts based on 100 positive labels. Missing Tiles indicates the percentage of tiles with missing labels, based on 100 negative tiles.

Category	Precision	Missing Tiles
Land Cover - Mangroves	100	-
Land Cover - Moss	100	-
Crop Type - Rice	100	-
Crop Type - Grape	100	-
Crop Type - Corn	100	-
Crop Type - Sugarcane	100	-
Crop Type - Tea	100	-
Crop Type - Hop	100	-
Crop Type - Wheat	100	-
Crop Type - Soy	100	-
Crop Type - Barley	100	-
Crop Type - Oats	100	-
Crop Type - Rye	100	-
Crop Type - Cassava	100	-
Crop Type - Potato	100	-
Crop Type - Sunflower	100	-
Crop Type - Asparagus	100	-
Crop Type - Coffee	100	-
Coast - Water	98	-
Coast - Land	100	-
Burned Area	96	-
Fire Retardant	86	-
Flood	100	-
Cloud	95	-
Bathymetry	100	-
Tree Cover - None	100	-
Tree Cover - Sparse	100	-
Tree Cover - Low	100	-
Tree Cover - Medium	100	-
Tree Cover - High	100	-
Tree Cover - Full	100	-
Wind Turbine - Rotor Diameter	100	-
Wind Turbine - Power	100	-
Parking Lot - Capacity	100	-
Track - Running	100	-
Track - Cycling	100	-
Track - Horse	100	-
Road Type - Motorway	100	-
Road Type - Trunk	100	-
Road Type - Primary	100	-
Road Type - Secondary	100	-
Road Type - Tertiary	100	-
Road Type - Residential	100	-
Road Type - Service	100	-
Road Type - Track	100	-
Road Type - Pedestrian	100	-
Road - Number of Lanes	100	-
Road - Max Speed	100	-
Road - Bridge	100	-
Road - Not Bridge	100	-
Power Plant - Oil	100	-
Power Plant - Nuclear	100	-
Power Plant - Coal	100	-
Power Plant - Gas	100	-
Quarry - Sand	100	-
Quarry - Gravel	100	-
Quarry - Clay	100	-
Quarry - Coal	100	-
Quarry - Peat	100	-

Table 6 (continued): Accuracy analysis of Satlas categories. Precision indicates the percentage of labels that are correct, as determined by two domain experts based on 100 positive labels. Missing Tiles indicates the percentage of tiles with missing labels, based on 100 negative tiles.

Category	Precision	Missing Tiles
Park Type - City Park	100	-
Park Type - Pitch	100	-
Park Type - Golf Course	100	-
Park Type - Cemetery	100	-
Park Sport - American Football	100	-
Park Sport - Badminton	100	-
Park Sport - Baseball	100	-
Park Sport - Basketball	100	-
Park Sport - Cricket	100	-
Park Sport - Rugby	100	-
Park Sport - Soccer	100	-
Park Sport - Tennis	100	-
Park Sport - Volleyball	100	-
Snow - None	100	-
Snow - Partial	100	-
Snow - Full	100	-
Smoke - None	97	-
Smoke - Low	94	-
Smoke - High	92	-

Table 6 (continued): Accuracy analysis of Satlas categories. Precision indicates the percentage of labels that are correct, as determined by two domain experts based on 100 positive labels. Missing Tiles indicates the percentage of tiles with missing labels, based on 100 negative tiles.

A.6. Test Split Selection

We first chose test split sizes so that refinement annotation would be feasible (see A.4.2). However, this restriction results in small test sizes, especially for the high-res image mode which has categories that require refinement. Then, due to the small test split size relative to total dataset size, randomly sampling the test split would yield too few examples in rare categories. Thus, for each rare category we randomly sampled 4 tiles for high-res and 128 tiles for low-res containing that category, and added these tiles to the test splits. This provided $\sim$1/4 of the desired test tiles; we randomly sampled the rest without constraints.

B. Downstream Experiments

B.1. Training Details

We provide additional details about training and fine-tuning for the downstream experiments here.

For all runs, when fine-tuning on downstream datasets, we first freeze the backbone and only train the prediction head for 32K examples, and then fine-tune the entire model. After unfreezing the backbone, we increase the learning rate linearly from 0 to 10^{-4} over another 32K examples. We then reduce the learning rate on plateaus in the training loss, down to a minimum of 10^{-5}.

For pre-training on SATLASPRETRAIN and the four dataset baselines, we employ random resizing, random cropping, and horizontal and vertical flipping augmentations. We select the batch size based on the maximum that can be used with NVIDIA A6000 GPU: we use a batch size of 16 for SATLASPRETRAIN DOTA, and iSAID; a batch size of 32 for Million-AID; and a batch size of 64 for BigEarthNet. We use the Adam optimizer, and initialize the learning rate to 10^{-4}, decaying via halving down to 10^{-6} upon plateaus in the training loss. We train until convergence in the training loss.

We use the SeCo weights provided by the authors. We train MoCo on SATLASPRETRAIN with a batch size of 32, for 5 days on 1 NVIDIA Titan X Pascal GPU. For MoCo training, 44x44 chunks are extracted from the original 512x512 images, to replicate the shapes used in the MoCo variant evaluated in the SeCo authors. We fine-tune the weights learned through self-supervision on the downstream tasks.

B.2. Per-Dataset Results

In Figures 11, 12, 13, and 14, we show performance for SATLASPRETRAIN compared to the four self-supervised and five supervised pre-training methods, on the seven downstream tasks with varying number of training samples. The sizes of the downstream datasets range from 137 to 363,572 training samples, and we choose seven few-shot scenarios in addition to training on the whole set for each task. In most cases, SATLASPRETRAIN outperforms the other baselines, especially so when fewer training examples are provided.

B.3. Additional SeCo Variants

In the main paper, we show results for SeCo trained with the authors' model on the authors' dataset, which performs the best. In Table 7, we show results for two additional variants of SeCo that do not perform as well: training their model on SATLASPRETRAIN, and training SeCo with SATLASNET on SATLASPRETRAIN.

Downstream results show that the default SeCo approach usually outperforms the SeCo variants using SATLASPRETRAIN. This is likely because, as stated in the SeCo paper, locations were carefully selected for inclusion in the SeCo dataset to minimize the number of uniform images that show monotonous areas such as oceans and deserts. Since SATLASPRETRAIN includes many such areas, we hypothesize that the model struggles when presented e.g. with images of different ocean areas that look identical to each other. We hypothesize that this also contributes to MoCo SATLASNET on SATLASPRETRAIN not performing well on the downstream tasks.

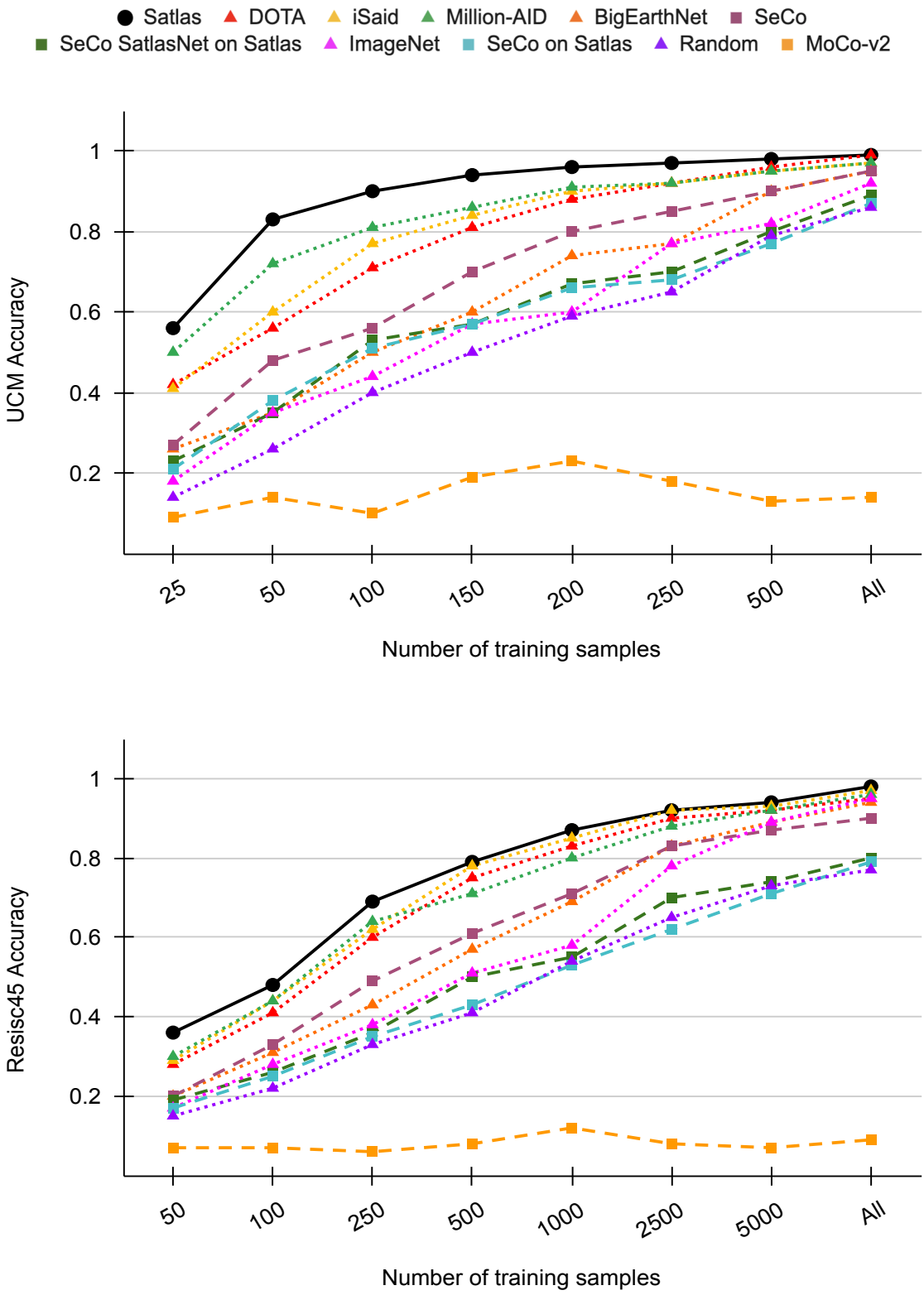

Satlas
DOTA
iSaid
Million-AID
BigEarthNet
SeCo
SeCo SatlasNet on Satlas
ImageNet
SeCo on Satlas
Random
MoCo-v2
UCM Accuracy
1
0.8
0.6
0.4
0.2
25
50
100
150
200
250
500
All
Number of training samples
Resisc45 Accuracy
1
0.8
0.6
0.4
0.2
50
100
250
500
1000
2500
5000
All
Number of training samples

Figure 11

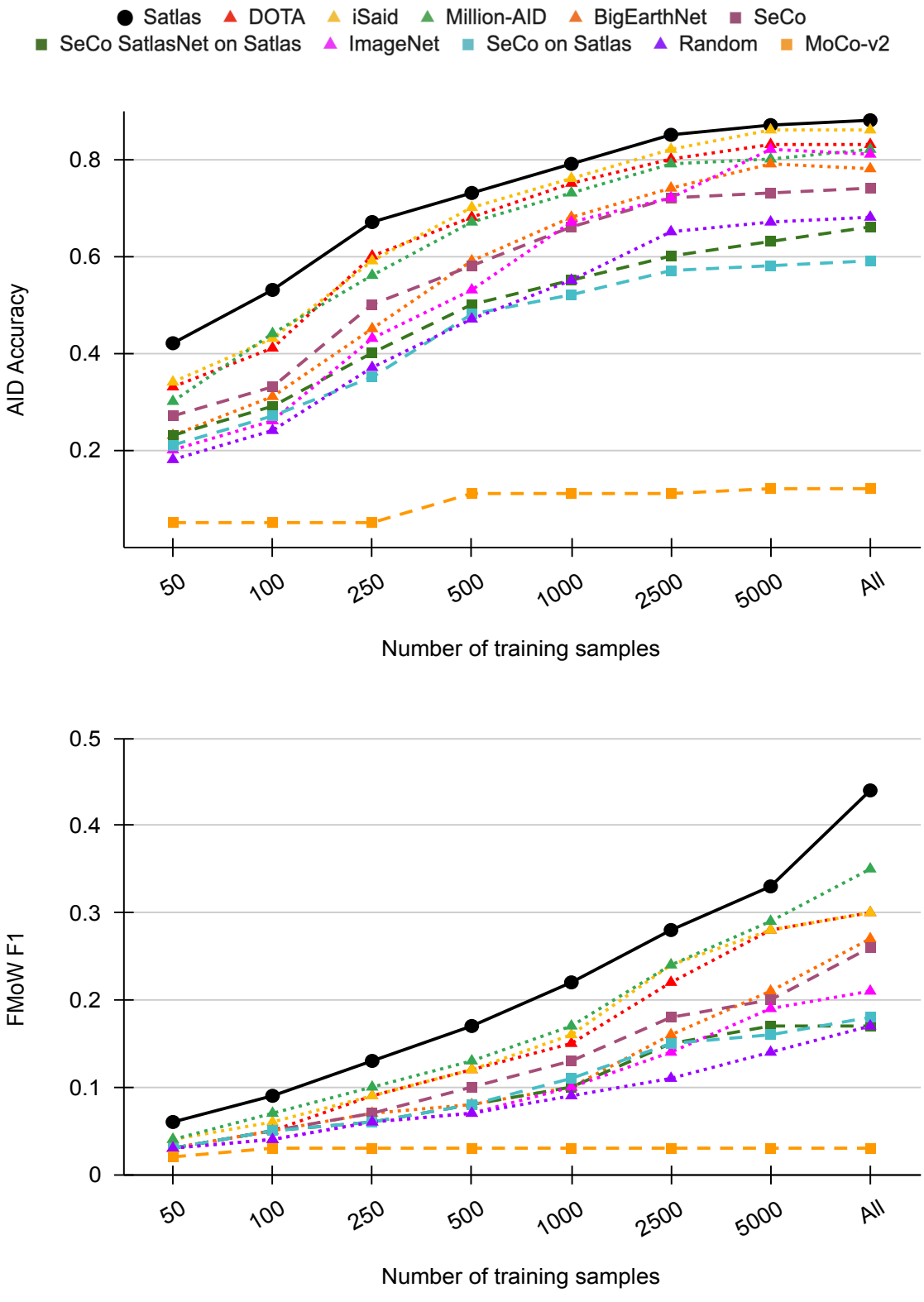

Satlas
DOTA
iSaid
Million-AID
BigEarthNet
SeCo
SeCo SatlasNet on Satlas
ImageNet
SeCo on Satlas
Random
MoCo-v2
AID Accuracy
0.2
0.4
0.6
0.8
50
100
250
500
1000
2500
5000
All
Number of training samples
FMoW F1
0
0.1
0.2
0.3
0.4
0.5
50
100
250
500
1000
2500
5000
All
Number of training samples

Figure 12

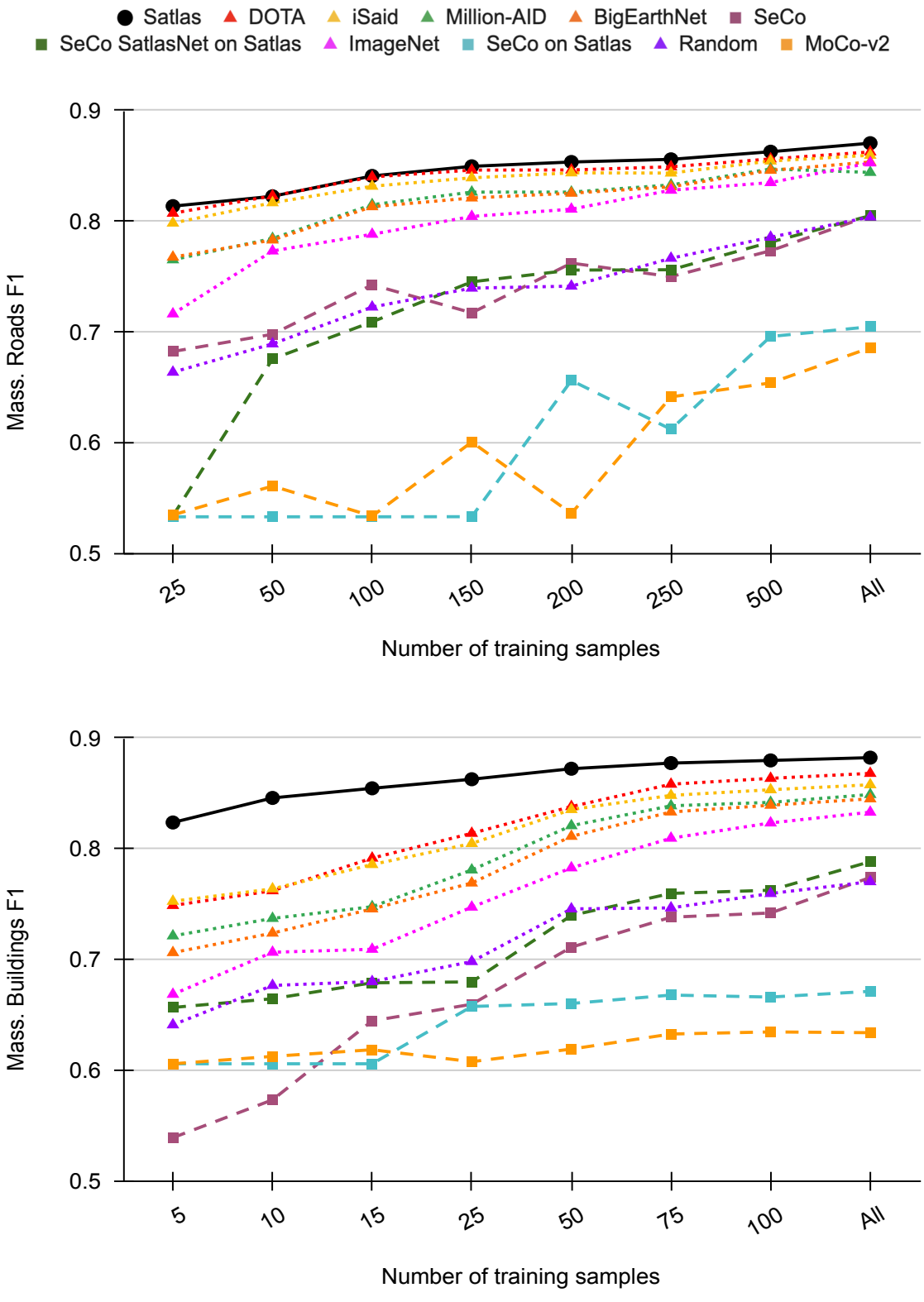

Satlas
DOTA
iSaid
Million-AID
BigEarthNet
SeCo
SeCo SatlasNet on Satlas
ImageNet
SeCo on Satlas
Random
MoCo-v2
Mass. Roads F1
0.9
0.8
0.7
0.6
0.5
25
50
100
150
200
250
500
All
Number of training samples
Mass. Buildings F1
0.9
0.8
0.7
0.6
0.5
5
10
15
25
50
75
100
All
Number of training samples

Figure 13

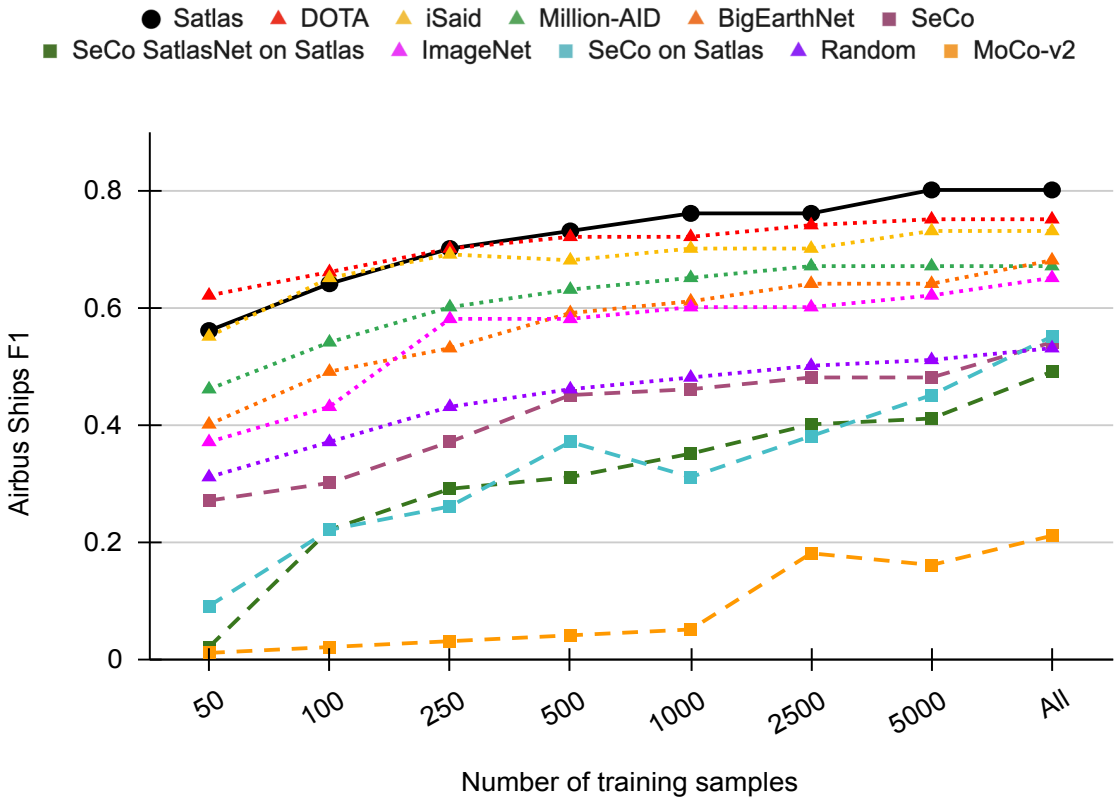


Figure 14

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
MoCo [3]	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 [4]	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
SeCo on SatlasPretrain	0.38	0.87	0.17	0.79	0.21	0.59	0.03	0.18	0.53	0.71	0.66	0.67	0.09	0.55	0.30	0.62
SeCo on SatlasPretrain with SatlasNet	0.35	0.89	0.19	0.80	0.23	0.66	0.03	0.17	0.68	0.81	0.74	0.79	0.02	0.49	0.32	0.66

Table 7: Results for SeCo variations 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.

C. Multispectral Bands and Sentinel-1 Images

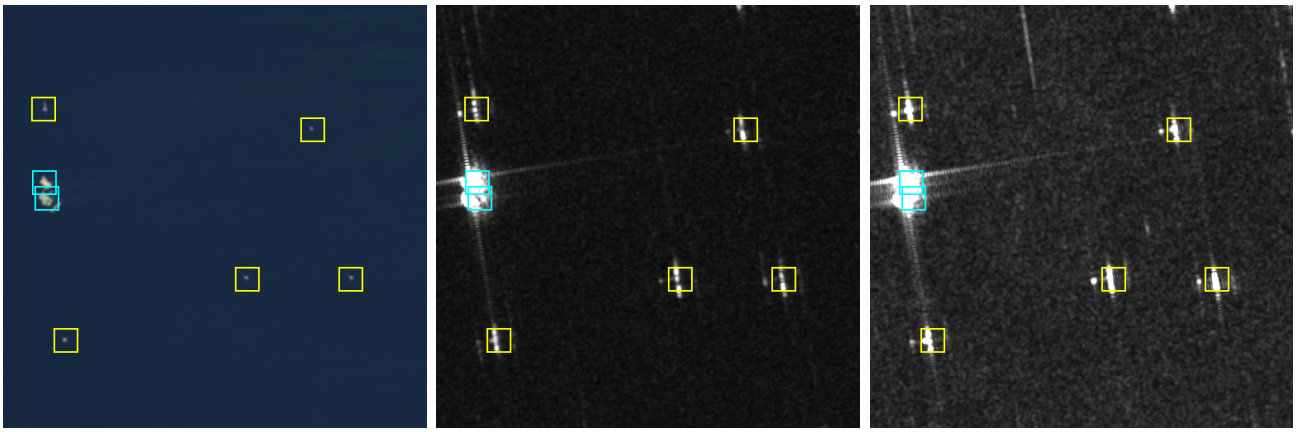
In the experiments in Section 5.1, we use RGB bands of Sentinel-2 only: B4, B3, and B2. In Table 8, we show results when training SATLASNET (single-image, joint) on (a) all 10 and 20 m/pixel bands, i.e., B2, B3, B4, B5, B6, B7, B8, B11, and B12; and (b) the 10 and 20 m/pixel bands other than RGB bands, i.e., B5, B6, B7, B8, B11, and B12. We normalize all values to between 0-1 by dividing the raw 16-bit band values by 8000, except B2, B3, and B4, which we obtain from the 8-bit true-color image product by dividing values by 255.

Results are mixed, with additional bands improving performance on points but actually reducing performance on segmentation, regression, and polygon categories. For categories annotated using the Sentinel-2 RGB bands, like offshore wind turbines and fire retardant, we suspect that because labeling was done based on RGB only, the extra information carried by the additional bands is not relevant to the labels and thus does not improve accuracy. For other categories, we hypothesize that the additional bands have the potential to improve accuracy, but that more training time and more data may be needed for the model to learn to effectively leverage the additional bands. We suspect that improved model architectures may make it possible to improve accuracy with the additional bands at lower training and/or annotation costs.

At all 856K tiles, in addition to providing Sentinel-2 images, we also include 4 synthetic aperture radar images captured by Sentinel-1 during 2022. Specifically, we include 10 m/pixel Sentinel-1 vh+vv IW GRD images. These images provide information that is complementary to the optical Sentinel-2 images, such as texture of the surface material, since SAR measures backscatter from the surface at each pixel. However, in our experiments we do not use the Sentinel-1 images due to cost (which is especially high for training multi-image SATLASNET). We show an example comparing Sentinel-1 and Sentinel-2 in Figure 15.

Method	Seg ↑	Reg ↓	Pt ↑	Pgon ↑	Pline ↑	Prop ↑	Cls ↑
SatlasNet (RGB only)	55.8	10.6	22.0	10.3	45.5	73.8	99
SatlasNet (all bands)	52.4	12.8	24.4	8.2	-	67.6	99
SatlasNet (all except RGB)	55.0	11.6	23.6	9.2	-	69.2	99

Table 8: Results on low-resolution SATLASPRETRAIN test set. 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). Polyline accuracy is missing due to a bug in the polyline data during the training of these models.



Sentinel-2 Sentinel-1 (vh) Sentinel-1 (vv)

Figure 15: Sentinel-2 and Sentinel-1 images of the same location, with off-shore wind turbines in yellow and off-shore platforms in cyan.

D. Further Analysis

D.1. Geographical Biases

In this section, we evaluate the accuracy of SATLASNET on a per-continent level for the point, polygon, segmentation, and property label types. We use single-image, low-res SATLASNET that is fine-tuned on each label type. Since many categories are not present on all continents, we evaluate on 5 point, 13 polygon, 9 segmentation, and 5 property categories, and report the average accuracy across categories for each label type.

We show results in Table 9. While performance varies widely for each label type, the model is more accurate in each geography on some label types and less accurate in others. The results are inconclusive especially because geographies with low coverage in the dataset are correlated with geographies where the task is easiest, since the most large-scale objects tend to be mapped first. For example, a model may be trained on 1K solar farms in A and only 100 in B; one would expect the model to exhibit higher performance in A, but it may instead perform comparably if the solar farms in B are larger on average with smaller solar farms left unlabeled in the dataset.

D.2. Per-Label-Type Downstream Performance

In this section, we compare downstream performance when training on all label types in SATLASPRETRAIN to training on each individual label type. We do not include classification labels since we pre-train on high-res images for downstream tasks, but SATLASPRETRAIN only contains classification labels for low-res images.

We show results in Table 10. Training on all label types outperforms training on most individual label types, but training on polygon labels performs comparably, indicating that improved pre-training strategies or larger models may further increase downstream performance.

Region	Point	Polygon	Segmentation	Property
Africa	35.5	16.9	61.5	76
Asia	21.2	17.5	60.1	79
Europe	22.8	17.6	56.4	75
North America	25.2	20.5	59.8	79
South America	30.0	16.9	61.1	75
Oceania	40.0	24.2	54.6	79

Table 9: Per-continent accuracy of single-image, low-res SatlasNet on four label types.

	Point		Polygon		Polyline		Property		Regression		Segmentation		All	
	50	100	50	100	50	100	50	100	50	100	50	100	50	100
UCM	0.664	0.985	0.822	0.995	0.769	0.994	0.637	0.979	0.348	0.870	0.708	0.988	0.831	0.994
AID	0.392	0.882	0.411	0.878	0.406	0.875	0.334	0.835	0.208	0.744	0.395	0.866	0.424	0.876
R45	0.307	0.969	0.323	0.969	0.356	0.972	0.296	0.950	0.157	0.866	0.295	0.962	0.358	0.976
FMoW	0.101	0.510	0.108	0.482	0.104	0.515	0.102	0.467	0.088	0.372	0.104	0.510	0.055	0.437
Mass Buildings	0.835	0.864	0.871	0.883	0.849	0.871	0.814	0.844	0.799	0.836	0.839	0.858	0.872	0.882
Mass Roads	0.817	0.858	0.834	0.864	0.857	0.872	0.807	0.847	0.750	0.841	0.799	0.862	0.823	0.871
Ships	0.503	0.805	0.530	0.826	0.496	0.814	0.4290	0.780	0.346	0.747	0.452	0.791	0.559	0.799

Table 10: Downstream performance with 50 and 100 downstream examples, comparing pre-training on individual label types in SATLASPRETRAIN against simultaneously pre-training on all label types.

E. Per-Category SATLASPRETRAIN Examples

On the following pages, we show example labels and SATLASNET predictions for each category in SATLASPRETRAIN, grouped by label type.

Points


Image
Ground Truth
Predictions
Image
Ground Truth
Predictions
wind_turbine
airplane
rooftop_solar_panel
fountain
toll_booth
chimney
radio_tower

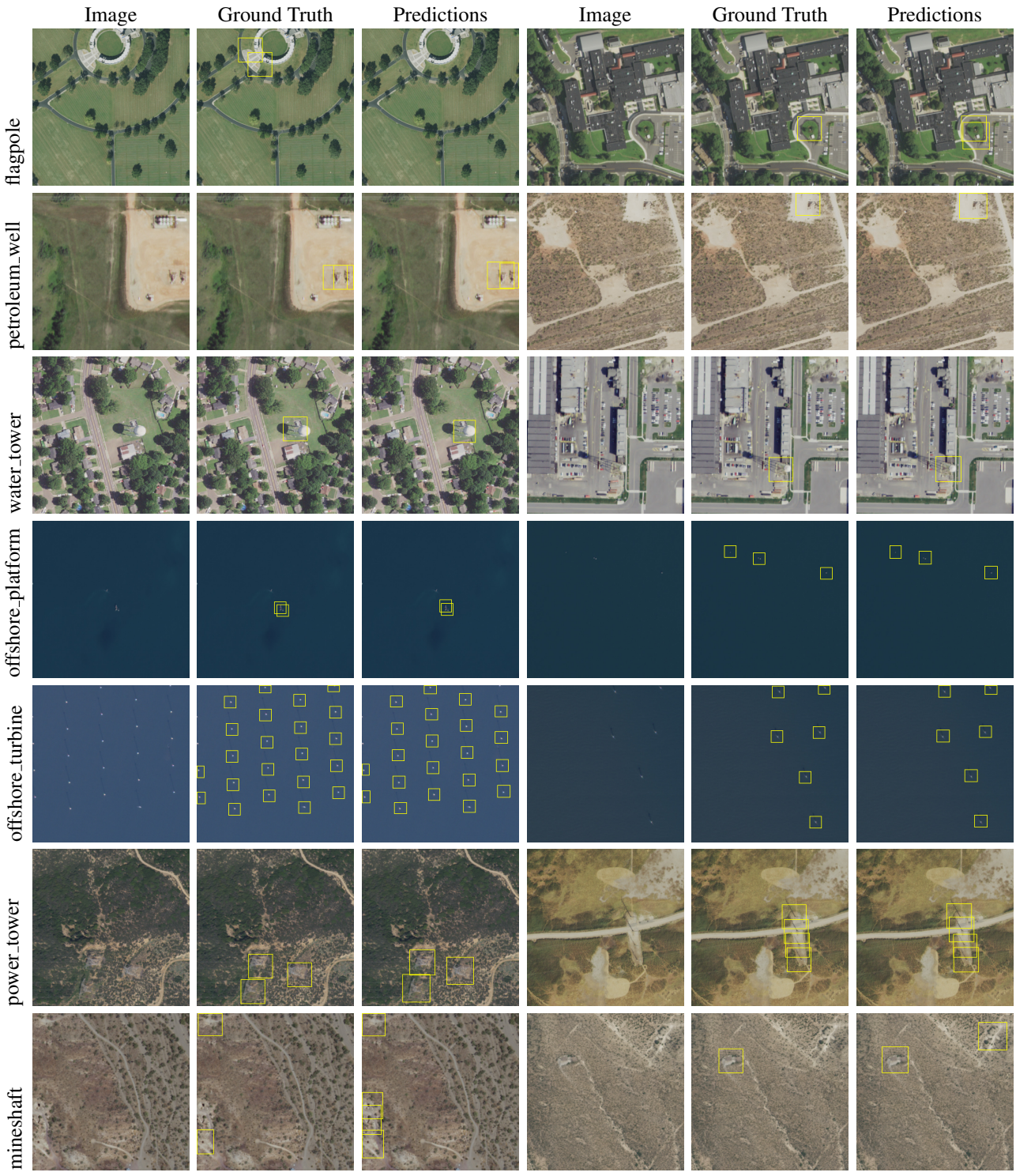


Image
Ground Truth
Predictions
Image
Ground Truth
Predictions
flagpole
petroleum_well
water_tower
offshore_platform
offshore_turbine
power_tower
mineshaft

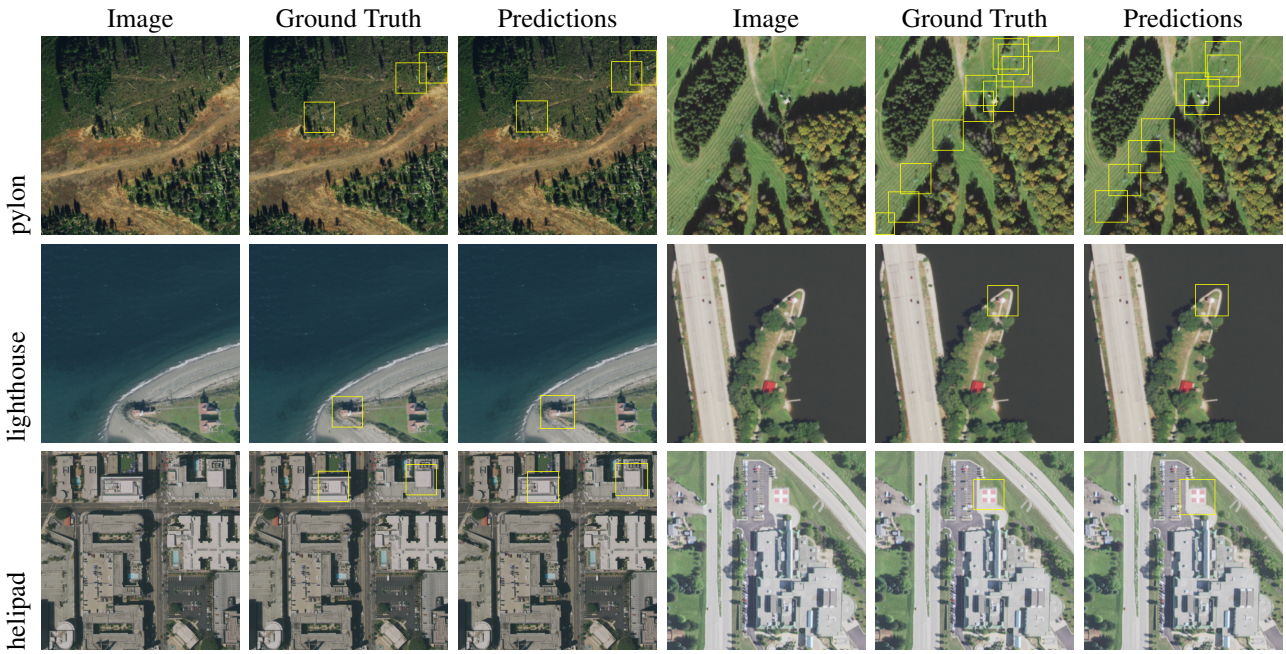


Image
Ground Truth
Predictions
Image
Ground Truth
Predictions
pylon
lighthouse
helipad

Polygons


Image
Ground Truth
Predictions
Image
Ground Truth
Predictions
building
dam
solar_farm
power_plant
gas_station
park
parking_garage


Image
Ground Truth
Predictions
Image
Ground Truth
Predictions
parking_lot
landfill
quarry
stadium
airport
airport_apron
airport_hangar

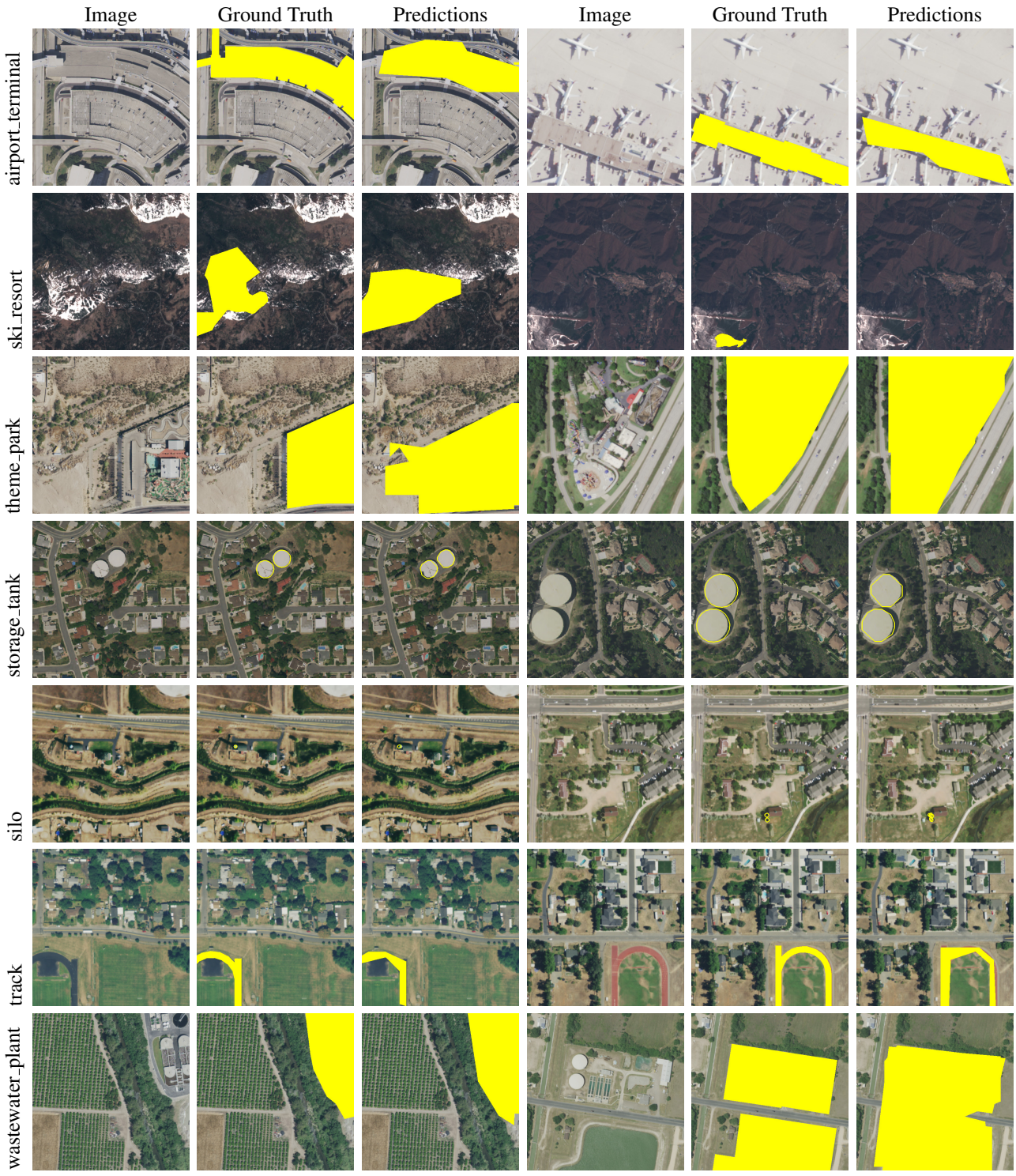


Image
Ground Truth
Predictions
Image
Ground Truth
Predictions
airport_terminal
ski_resort
theme_park
storage_tank
silo
track
wastewater_plant

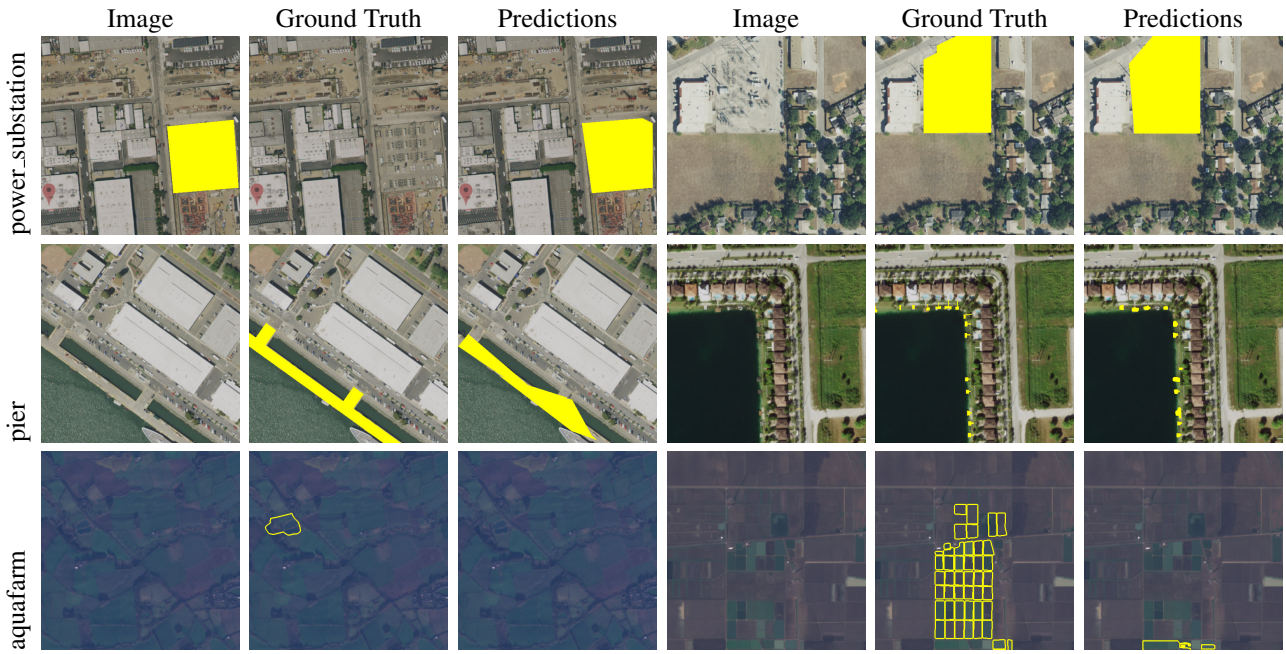


Image
Ground Truth
Predictions
Image
Ground Truth
Predictions
power_substation
pier
aquafarm

Polylines

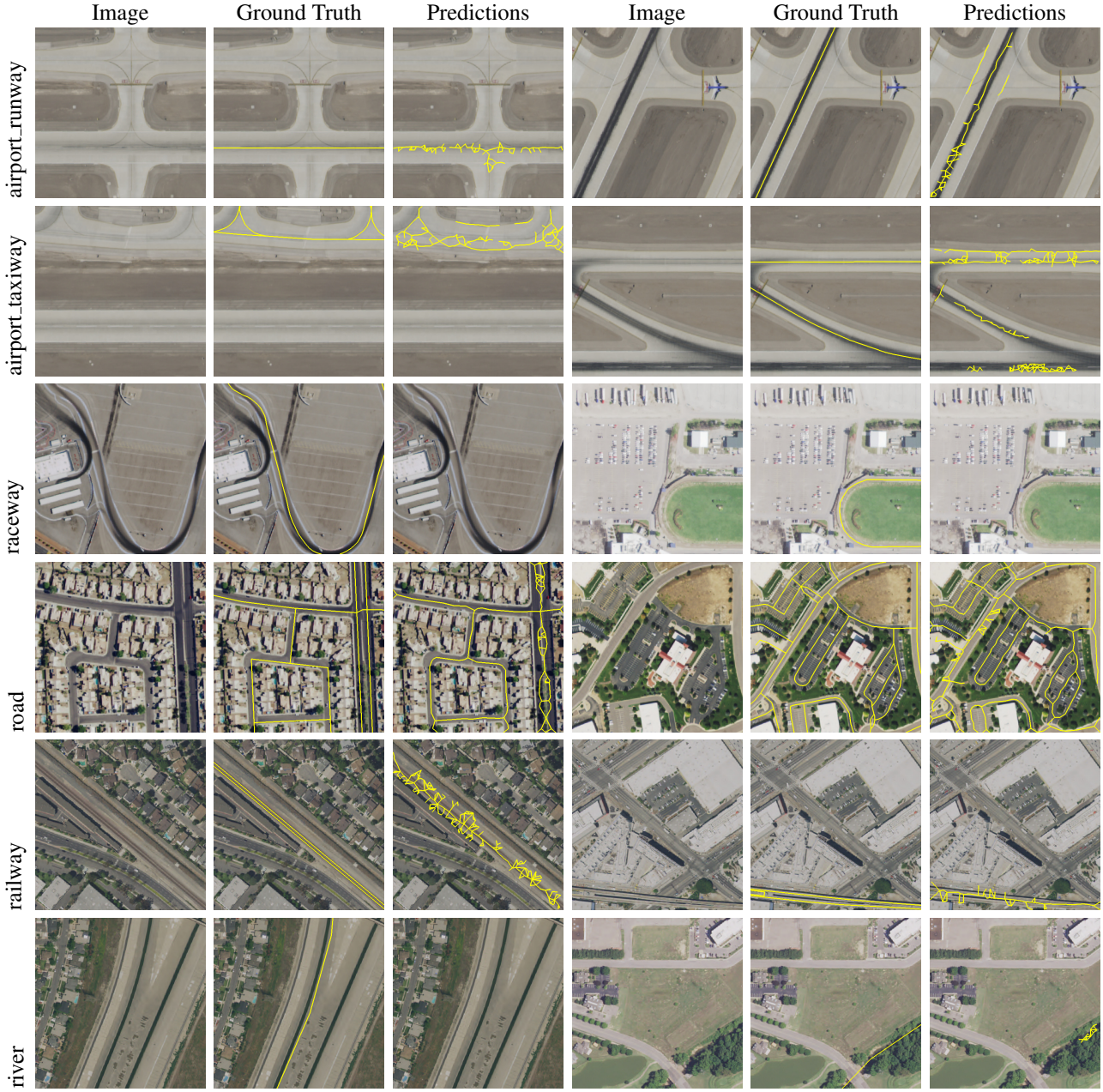
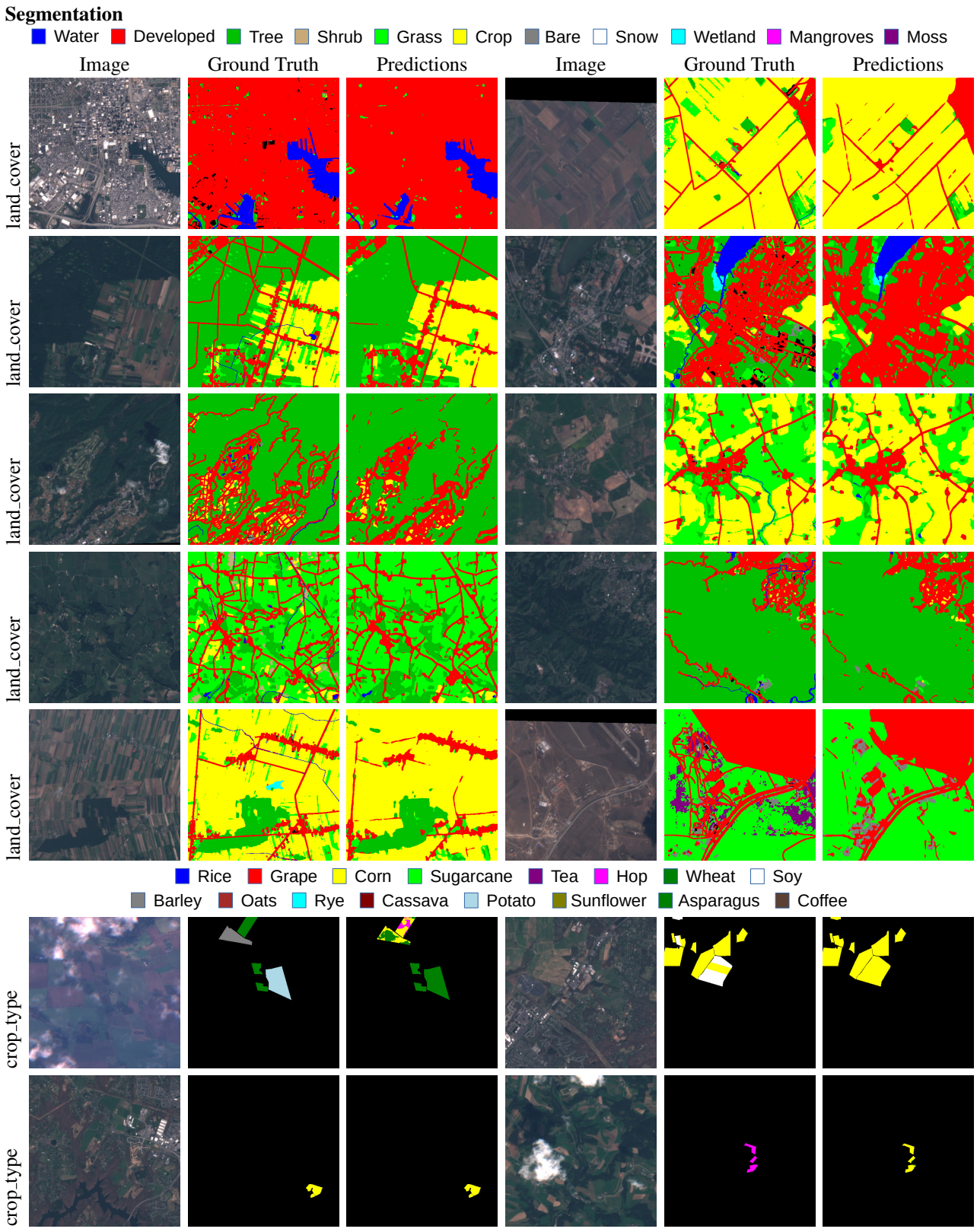


Image
Ground Truth
Predictions
Image
Ground Truth
Predictions
airport_runway
airport_taxiway
raceway
road
railway
river


Segmentation
Water
Developed
Tree
Shrub
Grass
Crop
Bare
Snow
Wetland
Mangroves
Moss
Image
Ground Truth
Predictions
Image
Ground Truth
Predictions
land_cover
land_cover
land_cover
land_cover
land_cover
Rice
Grape
Corn
Sugarcane
Tea
Hop
Wheat
Soy
Barley
Oats
Rye
Cassava
Potato
Sunflower
Asparagus
Coffee
crop_type
crop_type

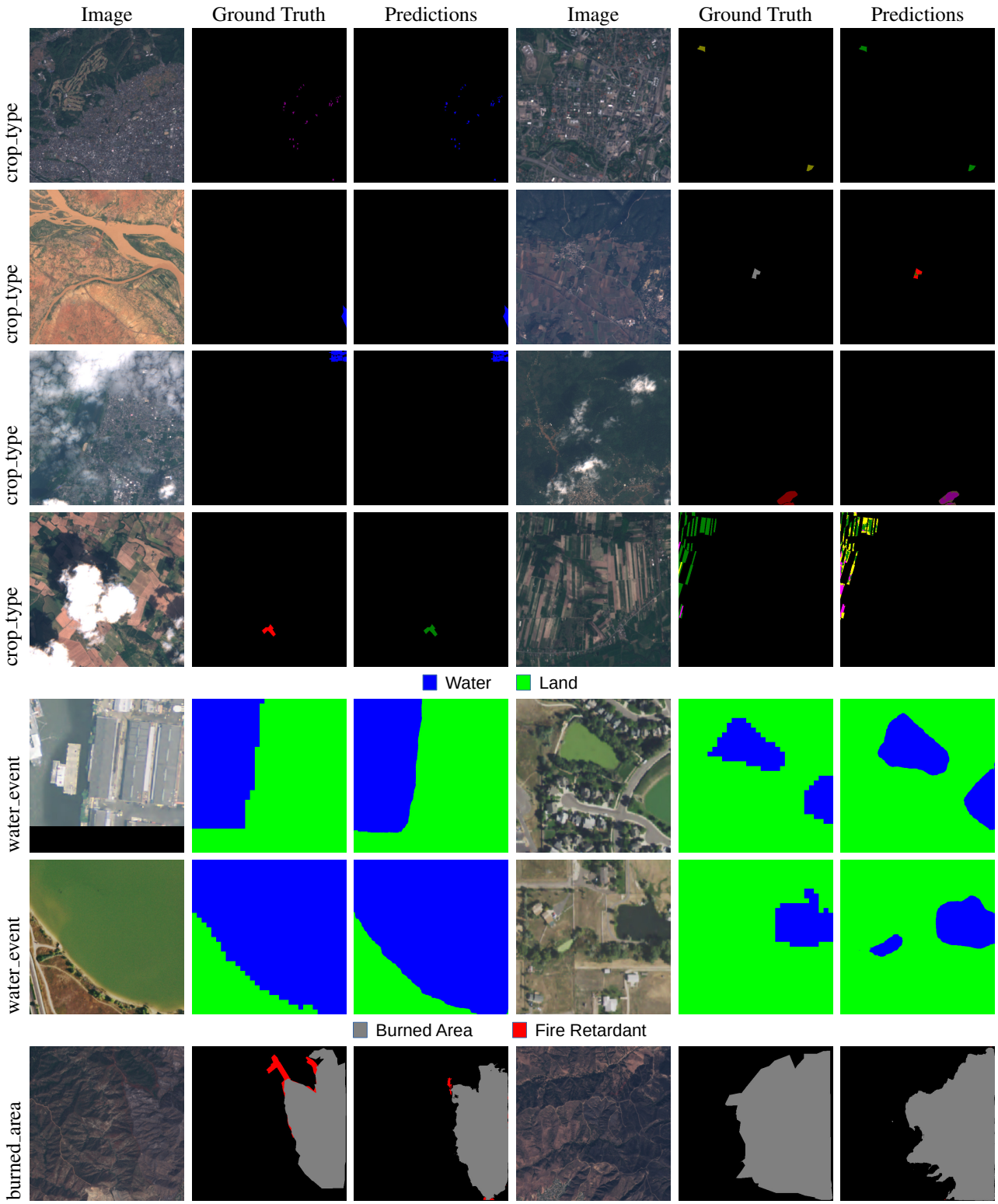
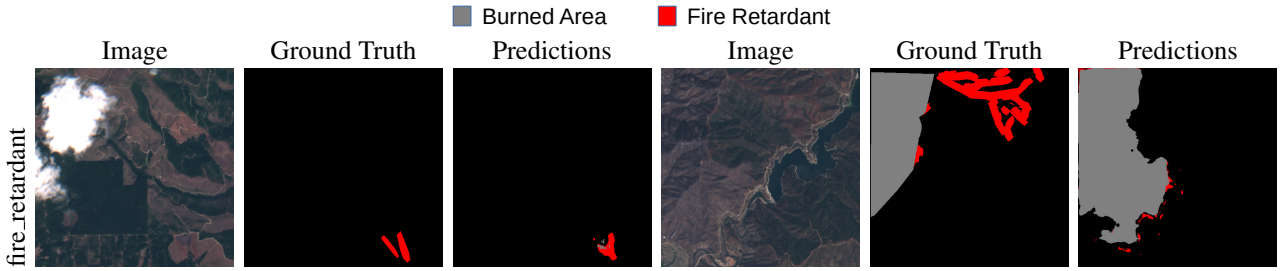


Image
Ground Truth
Predictions
Image
Ground Truth
Predictions
crop_type
crop_type
crop_type
crop_type
Water
Land
water_event
water_event
Burned Area
Fire Retardant
burned_area


Burned Area
Fire Retardant
Image
Ground Truth
Predictions
Image
Ground Truth
Predictions
fire_retardant

Regression

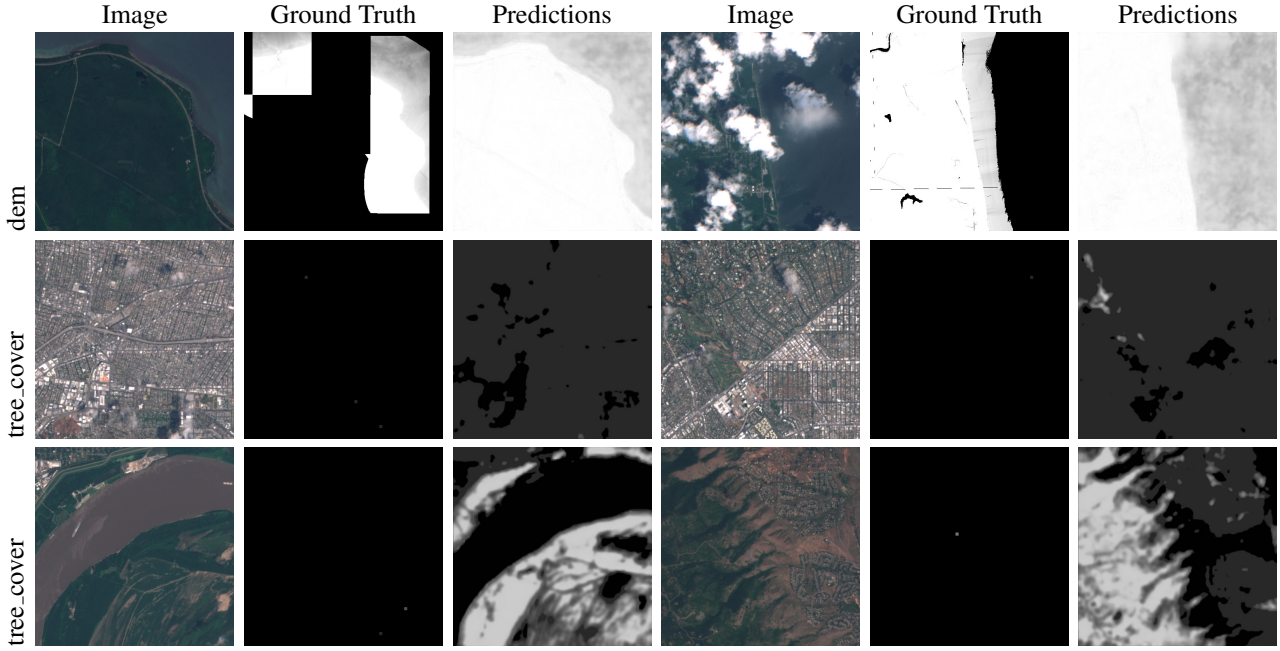
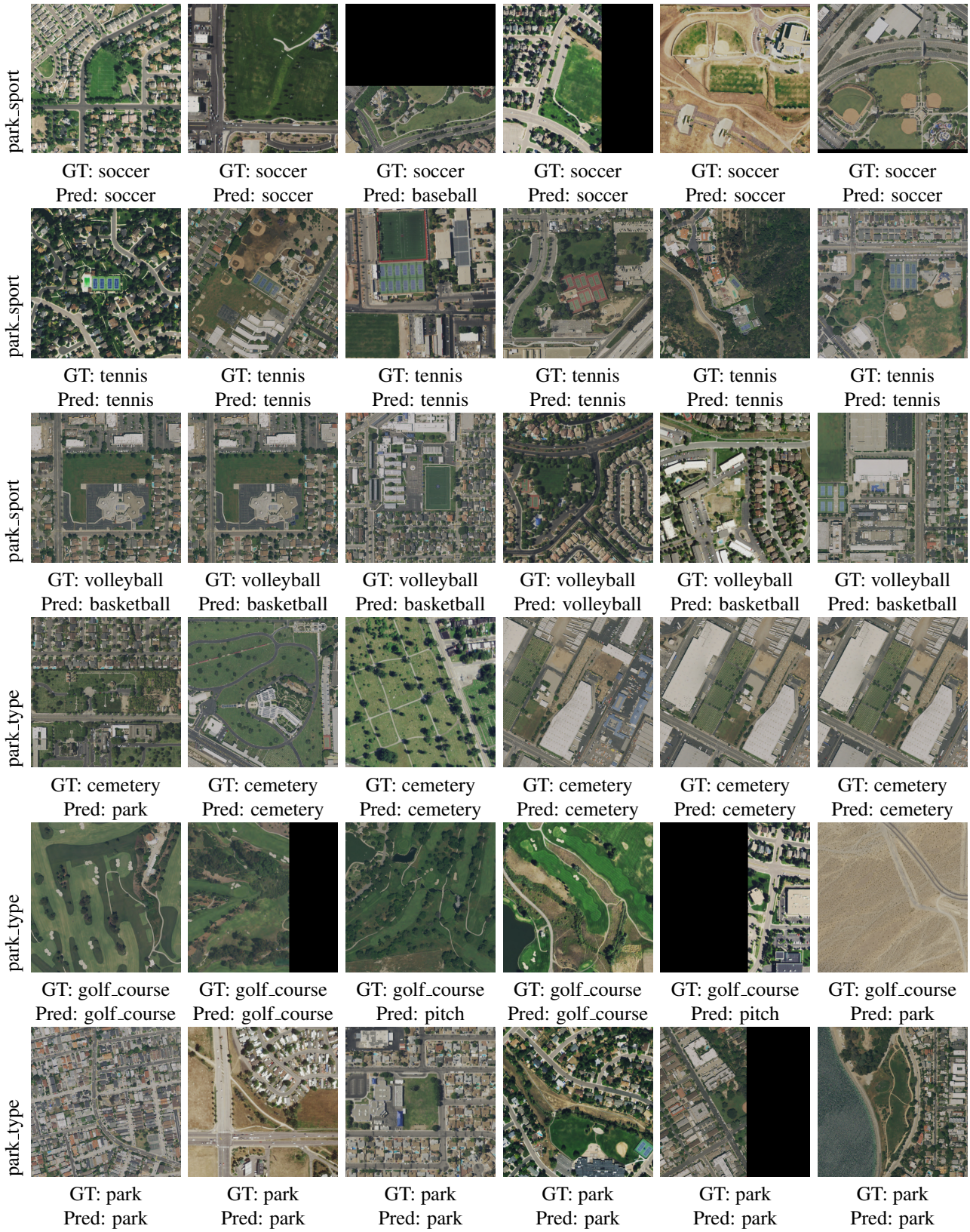
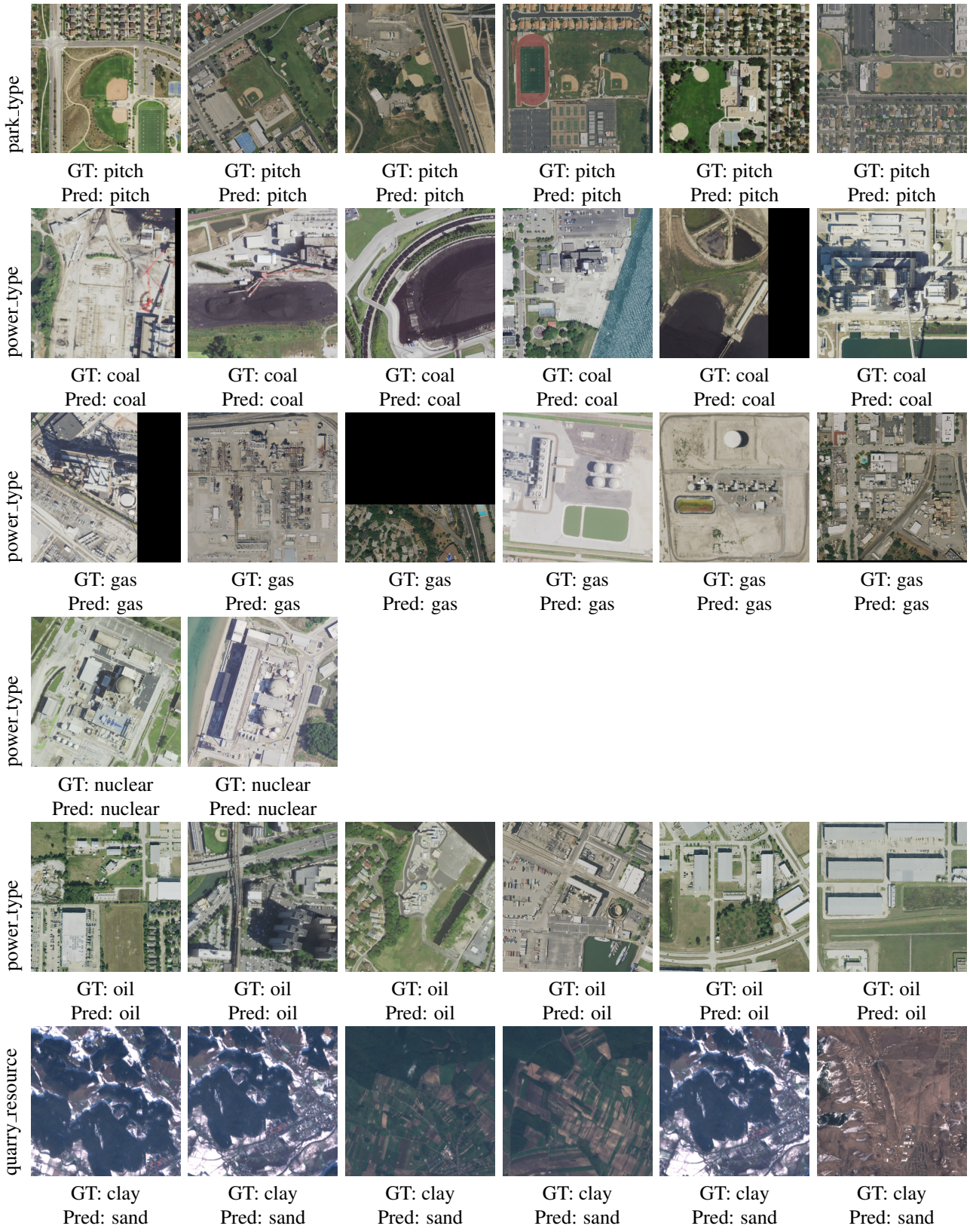


Image
Ground Truth
Predictions
Image
Ground Truth
Predictions
dem
tree_cover
tree_cover

Properties

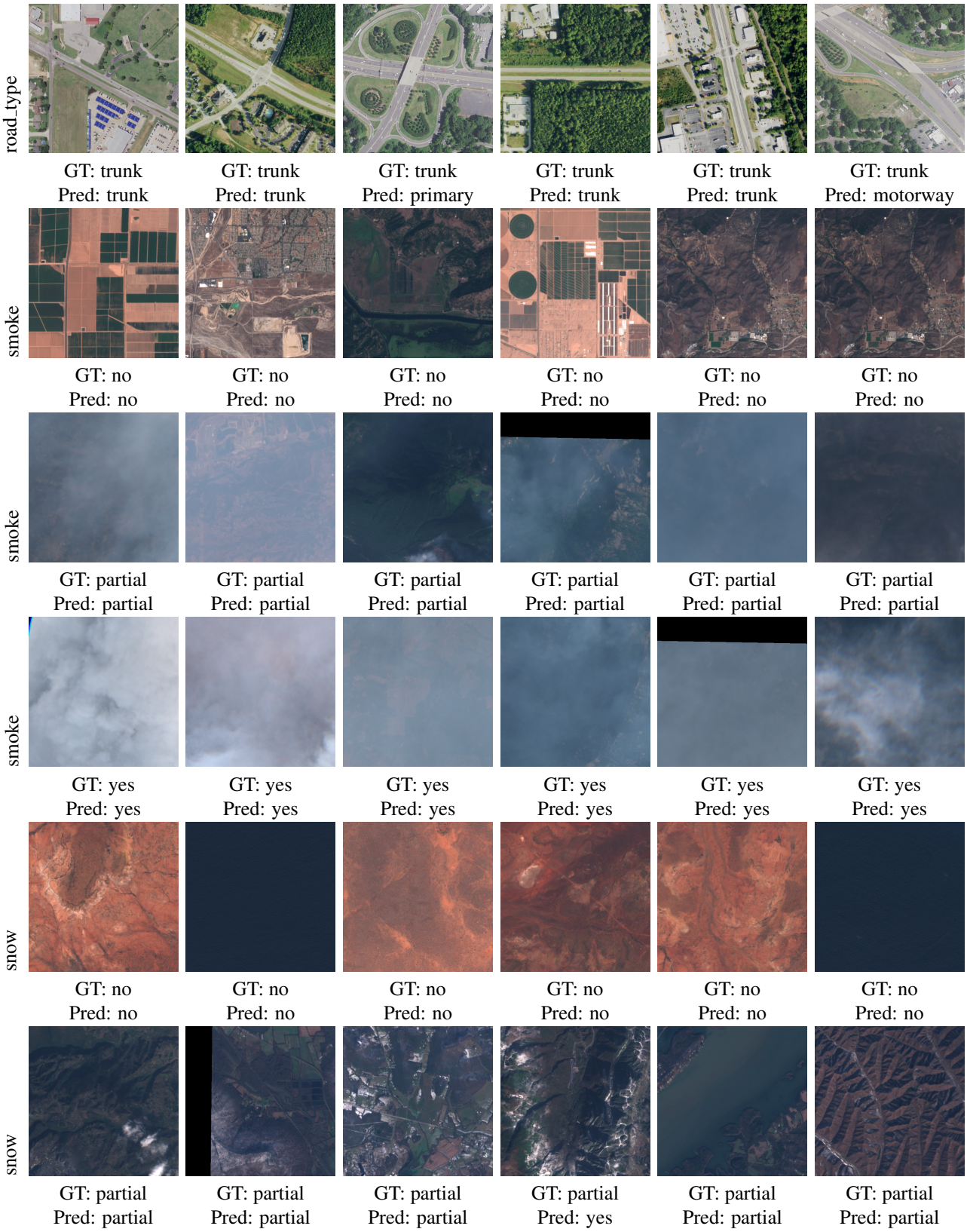

park_sport
GT: badminton
Pred: basketball
GT: badminton
Pred: volleyball
GT: badminton
Pred: basketball
GT: badminton
Pred: basketball
GT: badminton
Pred: basketball
GT: badminton
Pred: volleyball
park_sport
GT: baseball
Pred: baseball
GT: baseball
Pred: baseball
GT: baseball
Pred: baseball
GT: baseball
Pred: baseball
GT: baseball
Pred: baseball
GT: baseball
Pred: baseball
park_sport
GT: basketball
Pred: basketball
GT: basketball
Pred: basketball
GT: basketball
Pred: basketball
GT: basketball
Pred: basketball
GT: basketball
Pred: basketball
GT: basketball
Pred: basketball
park_sport
GT: cricket
Pred: cricket
GT: cricket
Pred: cricket
GT: cricket
Pred: cricket
GT: cricket
Pred: cricket
GT: cricket
Pred: cricket
GT: cricket
Pred: baseball
park_sport
GT: football
Pred: football
GT: football
Pred: football
GT: football
Pred: football
GT: football
Pred: football
GT: football
Pred: football
GT: football
Pred: football
park_sport
GT: rugby
Pred: soccer
GT: rugby
Pred: soccer
GT: rugby
Pred: soccer
GT: rugby
Pred: soccer
GT: rugby
Pred: soccer
GT: rugby
Pred: football

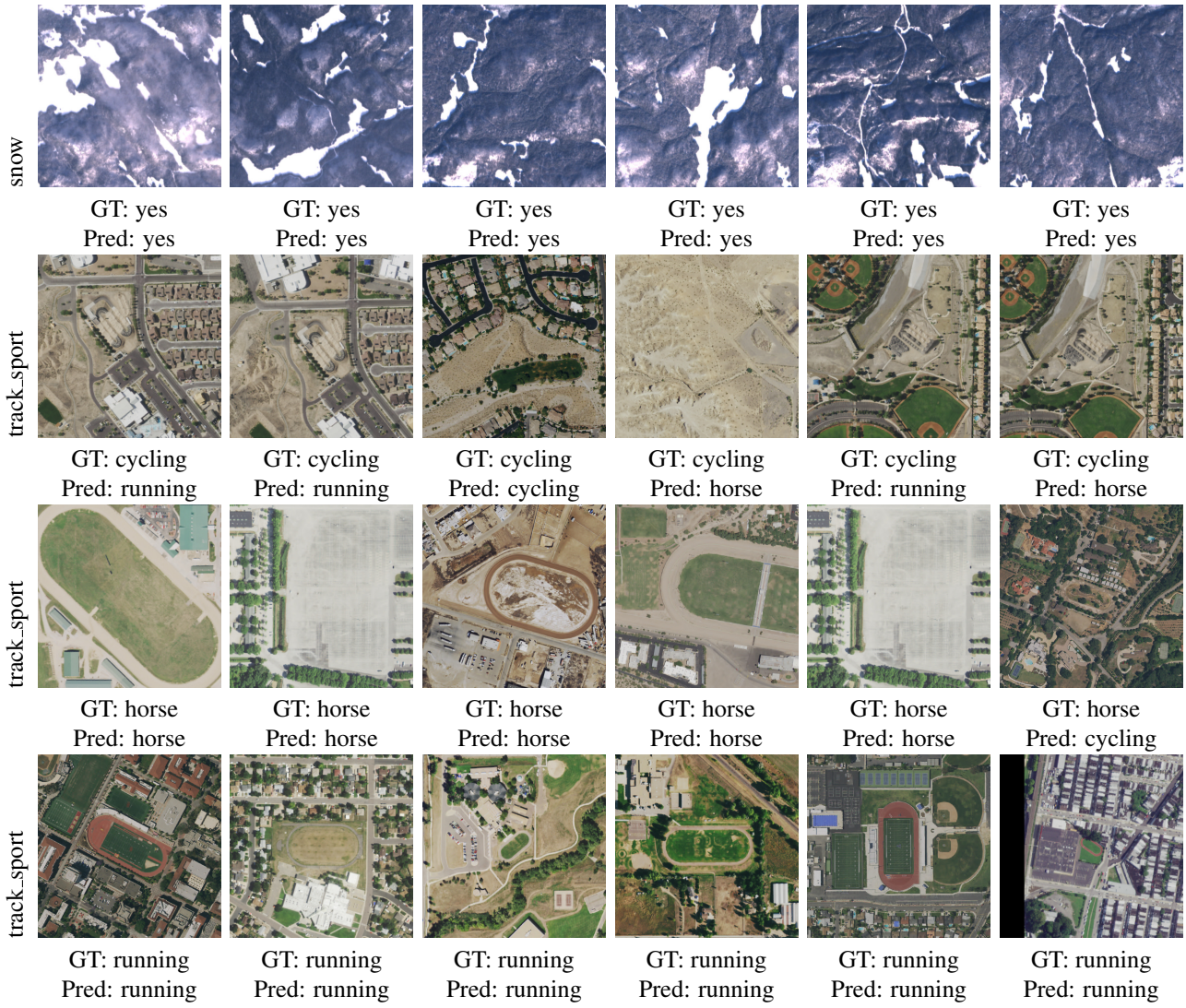

park_sport
GT: soccer
Pred: soccer
GT: soccer
Pred: soccer
GT: soccer
Pred: baseball
GT: soccer
Pred: soccer
GT: soccer
Pred: soccer
GT: soccer
Pred: soccer
park_sport
GT: tennis
Pred: tennis
GT: tennis
Pred: tennis
GT: tennis
Pred: tennis
GT: tennis
Pred: tennis
GT: tennis
Pred: tennis
GT: tennis
Pred: tennis
park_sport
GT: volleyball
Pred: basketball
GT: volleyball
Pred: basketball
GT: volleyball
Pred: basketball
GT: volleyball
Pred: volleyball
GT: volleyball
Pred: basketball
GT: volleyball
Pred: basketball
park_type
GT: cemetery
Pred: park
GT: cemetery
Pred: cemetery
GT: cemetery
Pred: cemetery
GT: cemetery
Pred: cemetery
GT: cemetery
Pred: cemetery
GT: cemetery
Pred: cemetery
park_type
GT: golf_course
Pred: golf_course
GT: golf_course
Pred: golf_course
GT: golf_course
Pred: pitch
GT: golf_course
Pred: golf_course
GT: golf_course
Pred: pitch
GT: golf_course
Pred: park
park_type
GT: park
Pred: park
GT: park
Pred: park
GT: park
Pred: park
GT: park
Pred: park
GT: park
Pred: park
GT: park
Pred: park


park_type
GT: pitch
Pred: pitch
power_type
GT: coal
Pred: coal
power_type
GT: gas
Pred: gas
power_type
GT: nuclear
Pred: nuclear
power_type
GT: oil
Pred: oil
quarry_resource
GT: clay
Pred: sand


quarry_resource
GT: coal
Pred: coal
GT: coal
Pred: sand
GT: coal
Pred: sand
GT: coal
Pred: coal
GT: coal
Pred: peat
GT: coal
Pred: coal
quarry_resource
GT: gravel
Pred: sand
GT: gravel
Pred: gravel
GT: gravel
Pred: sand
GT: gravel
Pred: sand
GT: gravel
Pred: sand
GT: gravel
Pred: sand
quarry_resource
GT: peat
Pred: sand
GT: peat
Pred: peat
GT: peat
Pred: peat
GT: peat
Pred: peat
GT: peat
Pred: sand
GT: peat
Pred: sand
quarry_resource
GT: sand
Pred: sand
GT: sand
Pred: gravel
GT: sand
Pred: sand
GT: sand
Pred: sand
GT: sand
Pred: coal
GT: sand
Pred: sand
road_type
GT: motorway
Pred: motorway
GT: motorway
Pred: motorway
GT: motorway
Pred: motorway
GT: motorway
Pred: motorway
GT: motorway
Pred: primary
GT: motorway
Pred: motorway
road_type
GT: pedestrian
Pred: pedestrian
GT: pedestrian
Pred: pedestrian
GT: pedestrian
Pred: pedestrian
GT: pedestrian
Pred: pedestrian
GT: pedestrian
Pred: residential
GT: pedestrian
Pred: service


road_type
GT: primary
Pred: primary
GT: primary
Pred: secondary
GT: primary
Pred: secondary
GT: primary
Pred: secondary
GT: primary
Pred: secondary
GT: primary
Pred: service
road_type
GT: residential
Pred: residential
GT: residential
Pred: residential
GT: residential
Pred: residential
GT: residential
Pred: residential
GT: residential
Pred: residential
GT: residential
Pred: residential
road_type
GT: secondary
Pred: secondary
GT: secondary
Pred: secondary
GT: secondary
Pred: secondary
GT: secondary
Pred: secondary
GT: secondary
Pred: service
GT: secondary
Pred: secondary
road_type
GT: service
Pred: service
GT: service
Pred: service
GT: service
Pred: service
GT: service
Pred: service
GT: service
Pred: service
GT: service
Pred: service
road_type
GT: tertiary
Pred: residential
GT: tertiary
Pred: service
GT: tertiary
Pred: secondary
GT: tertiary
Pred: service
GT: tertiary
Pred: service
GT: tertiary
Pred: residential
road_type
GT: track
Pred: track
GT: track
Pred: track
GT: track
Pred: track
GT: track
Pred: pedestrian
GT: track
Pred: track
GT: track
Pred: tertiary


road_type
GT: trunk
Pred: trunk
GT: trunk
Pred: trunk
GT: trunk
Pred: primary
GT: trunk
Pred: trunk
GT: trunk
Pred: trunk
GT: trunk
Pred: motorway
smoke
GT: no
Pred: no
GT: no
Pred: no
GT: no
Pred: no
GT: no
Pred: no
GT: no
Pred: no
GT: no
Pred: no
smoke
GT: partial
Pred: partial
GT: partial
Pred: partial
GT: partial
Pred: partial
GT: partial
Pred: partial
GT: partial
Pred: partial
GT: partial
Pred: partial
smoke
GT: yes
Pred: yes
GT: yes
Pred: yes
GT: yes
Pred: yes
GT: yes
Pred: yes
GT: yes
Pred: yes
GT: yes
Pred: yes
snow
GT: no
Pred: no
GT: no
Pred: no
GT: no
Pred: no
GT: no
Pred: no
GT: no
Pred: no
GT: no
Pred: no
snow
GT: partial
Pred: partial
GT: partial
Pred: partial
GT: partial
Pred: partial
GT: partial
Pred: yes
GT: partial
Pred: partial
GT: partial
Pred: partial


snow
GT: yes
Pred: yes
GT: yes
Pred: yes
GT: yes
Pred: yes
GT: yes
Pred: yes
GT: yes
Pred: yes
GT: yes
Pred: yes
track_sport
GT: cycling
Pred: running
GT: cycling
Pred: running
GT: cycling
Pred: cycling
GT: cycling
Pred: horse
GT: cycling
Pred: running
GT: cycling
Pred: horse
track_sport
GT: horse
Pred: horse
GT: horse
Pred: horse
GT: horse
Pred: horse
GT: horse
Pred: horse
GT: horse
Pred: horse
GT: horse
Pred: cycling
track_sport
GT: running
Pred: running
GT: running
Pred: running
GT: running
Pred: running
GT: running
Pred: running
GT: running
Pred: running
GT: running
Pred: running

References

[1] A global flood events and cloud cover dataset (version 1.0), 2022. Cloud to Street, Microsoft, Radiant Earth Foundation.

[2] Joe Avis, Henri Dolou, Jean Laporte, and François-Régis Martin-Lauzer. Sentinel Coastal Charting Worldwide. *ESA Report*, 2019.

[3] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved Baselines with Momentum Contrastive Learning. *arXiv preprint arXiv:2003.04297*, 2020.

[4] Oscar Manas, Alexandre Lacoste, Xavier Giro i Nieto, David Vazquez, and Pau Rodriguez. Seasonal Contrast: Unsupervised Pre-Training from Uncurated Remote Sensing Data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[5] Fernando Paolo, Tsu ting Tim Lin, Ritwik Gupta, Bryce Goodman, Nirav Patel, Daniel Kuster, David Kroodsma, and Jared Dunnmon. xView3-SAR: Detecting Dark Fishing Activity Using Synthetic Aperture Radar Imagery. *Neural Information Processing Systems*, 2022.

[6] 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: Global land cover mapping at 10 m resolution for 2020 based on Sentinel-1 and 2 data. In *AGU Fall Meeting Abstracts*, volume 2021, pages GC45I–0915, 2021.