**Assessing High-Resolution CubeSat Imagery and Machine Learning for Detailed, High Resolution Snow-Covered Area.** Anthony F. Cannistra, Nicoleta Cristea.

**Section X:** Methods and Data

We focus our efforts in this study on developing and evaluating a statistical modeling-based approach to enable the identification of snow-covered area in high-resolution remote sensing data. We chose machine learning as our methodological domain after preliminary experiments demonstrated high potential for these flexible statistical methods to be well-suited to the challenges presented by these high-resolution satellite data. Furthermore, the domain of snow-cover identification is particularly suited to a machine learning-driven analysis due to the ready availability of high-resolution airborne lidar-derived snow-cover data, explained further in Sections X.X and X.X. We construct our model following a standard machine learning paradigm (described xxx), and choose well-studied focal areas (described below) as benchmark locations to assess both the absolute performance of our model compared to ground truth (Section XXX) and relative model performance across particular variables of interest.

**X.X**: Study Region

The two focal regions for this study are the Upper Tuolumne Basin in the Sierra Nevada mountains of California, USA, and the Gunnison/East River Basin in the Central Rocky Mountains of Colorado, USA. These regions were selected for their robust temporal catalog of high-resolution airborne lidar data from the NASA Airborne Snow Observatory (See Section X.X), and to create an opportunity for model comparison across climatological zones.

**X.X:** Snow Cover Data

To allow our chosen statistical modeling approach to identify snow-covered regions in these high-resolution imagery data, our methods require the use of separate high-resolution snow covered area data to serve as “ground truth” (see Section X.X). For this we use the NASA/JPL Airborne Snow Observatory (ASO) data product (Painter et al., 2016). These data are gridded 3 meter snow depth derived from airborne lidar collected in chosen watersheds across the western United States. Data are collected on a weekly basis from mid-winter through complete snowmelt (February – June) in ASO target basins. We acquired these data from the National Snow and Ice Data Center’s Distributed Data Access Center (<https://nsidc.org/data/ASO_3M_SD>). We generate binary snow masks from these data by applying a threshold of 10cm to the snow depth field.

**X.X:** Satellite Imagery

Planet Labs, Inc. (“Planet”) is a commercial satellite imagery company that operates the “PlanetScope” constellation of approximately 130 small (10x10x30cm) satellites in sun-synchronous orbit. This constellation collects approximately 200 million km2 day-1 of optical (red, green, blue, and NIR) land-surface imagery at 3.7m GSD (at nadir) between ±81.5degrees latitude, with daily nadir revisit times. Two sensor types (“instruments”) are present in the PlanetScope constellation (“PS2” and “PS2.SD”) resulting from ongoing sensor development and satellite launches. These instruments have comparable (but not identical) spectral band centers and bandwidths (Table X).

|  |  |  |
| --- | --- | --- |
|  | PS2 | PS2.SD |
| Blue | 455 - 515 nm | 464 - 517 nm |
| Green | 500 - 590 nm | 547 - 585 nm |
| Red | 590 - 670 nm | 650 - 682 nm |
| NIR | 780 - 860 nm | 846 - 888 nm |

**Table X:** Spectral bandwidth of PS2 and PS2.SD instruments within the PlanetScope constellation, (Planet Labs, Inc., 2019a).

For the purposes of this study we consider these instruments to be identical, and acquire data from each interchangeably. This choice was motivated by our intent to construct a method that can leverage the entire temporal extent of the Planet imagery catalog.

Planet provides several levels of processing of collected imagery data. For this study we used the Level 3B PlanetScope “Analytic Ortho Scene,” an orthorectified multispectral surface reflectance data product. Planet’s data processing procedure converts top of atmosphere radiance (derived by applying sensor darkfield and flat field corrections to raw image sensor data) to surface reflectance using near-real-time MODIS data inputs and the 6SV2.1 radiative transfer code (Kotchenova et al., 2008; Planet Labs, Inc., 2019a). This data type is recommended by Planet for analytic applications.

Scene Selection and Acquisition

To pair relatively cloud-free imagery with Airborne Snow Observatory (ASO) collections (see Sections XXX and XXX), scenes with spatial and temporal overlap with individual ASO collect spatial footprints were selected from the Planet imagery catalog. We applied a 3-6 day temporal buffer to the ASO collection date of interest to ensure a higher probability of cloud-free imagery acquisition. (Temporal image density in the Planet catalog has increased dramatically over time, but regions of ASO collects prior to 2017 often required a larger (5-7 day) buffer to find spatially-overlapping imagery.) Image candidates were manually inspected for relative cloud fraction, and images with fewer clouds were selected for inclusion into our analysis. We used porder version 0.5.7, an open-source tool (Roy, 2019) for the Planet Orders v2 Application Programming Interface (API) (Planet Labs, Inc., 2019b), to both query the Planet catalog for imagery data and submit imagery orders. Analytic Ortho Scene assets were queried via the “PSScene4Band” identifier and the “analytic\_sr” bundle identifier. We used the Planet Clips API to acquire only those pixels overlapping our areas of interest (e.g. areas covered by ASO collect footprints) both to conserve our imagery quota and reduce data volume. Image assets were delivered from Planet directly to Amazon Web Services Simple Storage Service (S3) buckets for further processing.

**X.X:** Machine Learning Methodology

“Machine learning” is a set of statistical techniques to build predictive models of an outcome variable from data. Models are “trained” or “fit” to data by selecting a “training” subset of examples from the population of data. These examples are used to derive a predictive relationship with the response variable, and each machine learning technique varies on the precise methodology used to derive this relationship. Once fit, models are assessed for their ability to accurately predict response variables given “unseen” samples of data (the “test” subset) which is disjoint from the training set. In this study we employed a “supervised learning” approach, wherein the presence of the response variable in the data (known as a data “label”, in this case lidar-derived snow presence) guides the search for a statistical relationship between the input data (“features”) and the response (“label”). Once a supervised learning model is fit using data that contains the response variable, the resulting statistical relationship can be employed to predict the response variable from unlabeled data.

Identifying the spatial extent and categorical classification of regions within images is known as “image segmentation” or “instance segmentation” (CITE?). Classification of snow in satellite imagery fits well within this task definition, and this allowed us to use machine learning techniques specific to image segmentation. In our version of the task, the four bands of PlanetScope imagery in each pixel (red, green, blue, NIR; see Section XXX) represent the input data to our model (the “features”), and airborne lidar-derived binary snow masks (see Section XXX) represent the response variable (“labels”).

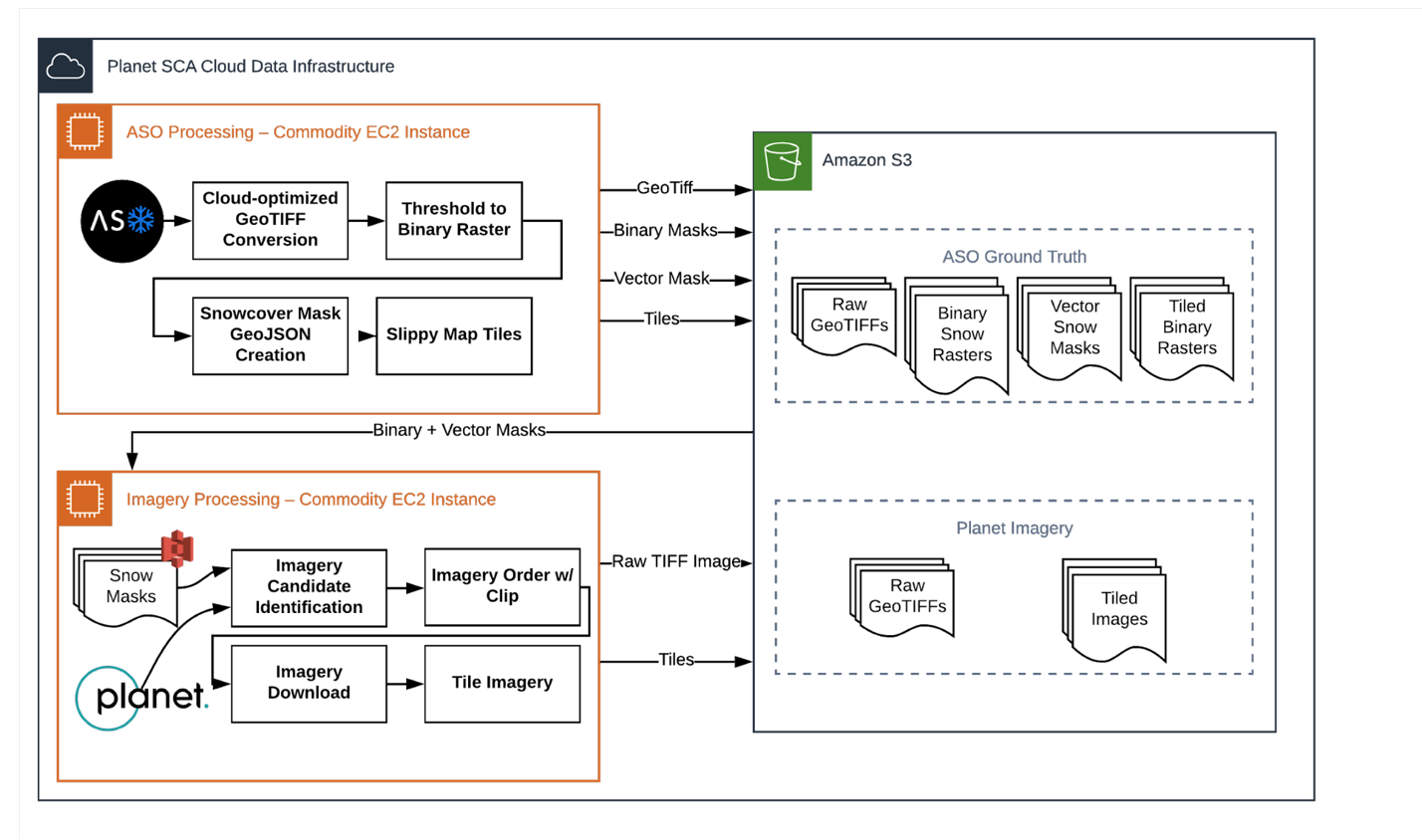
We employ a neural network to accomplish this image segmentation task. Neural networks are specific types of machine learning methods designed to extract statistically meaningful linear combinations of input features from data (PlanetScope bands) and model a dependent variable (snow presence/absence) as a nonlinear function of these derived linear combinations (Hastie et al., 2009, Section 11.1). We chose a method based on a network architecture demonstrated to perform very well in biomedical image segmentation (Known as the "U-Net" architecture; Ronneberger et al., 2015) and modified to perform well with satellite remote sensing imagery. The resulting network, known as TernausNetV2 and developed by Iglovikov et al., provides state-of-the-art satellite image segmentation when applied to the task of building detection in satellite imagery (Iglovikov et al., 2018). To our knowledge this method has not been applied to the segmentation of snow in satellite imagery.

**X.X:** Cyberinfrastructure & Model Training

Training a neural network is a computationally-demanding task requiring access to large quantities of data and specialized hardware. In particular, computers with access to large memory and graphics processing units (GPUs) greatly shorten the training time for our model and enable quicker experimentation. In addition, the large volume of both airborne lidar and satellite imagery data co-located with our study sites required us to have access to large data storage facilities. For these reasons we chose the compute and storage resources provided by Amazon Web Services (AWS), a commercial cloud service provider, to enable our training procedure.

Data Preprocessing

Once acquired, the imagery and airborne lidar-derived snow masks are stored as single or multiband GeoTIFF files (Open Geospatial Consortium, 2019) in AWS Simple Storage Service (S3) “buckets” to enable access by further processing tools. To enable co-registration of the snow mask data with imagery data and produce standardized “data units” required by neural network training, we then divide the raw imagery and snow mask data into 512 by 512 pixel images, or “tiles,” derived from a standardized global grid. We use the Spherical Web Mercator Spherical Tile standard (sometimes referred to as the “slippy map” tile standard due to their employment in interactive mapping applications) to define the grid of tiles (OpenStreetMap, 2019), and use the “mercantile,” “rasterio,” and “rio-tiler” open-source software packages to enable gridding and storage of these images (Mapbox, Inc., 2019; Vincent, 2019). The spherical Web Mercator tile standard assigns a unique spatially-explicit identifier to each 512x512 pixel image tile, which can then be used to align imagery tiles and snow mask tiles (e.g. the tiles “snow/1/2/3.tif” and “image/1/2/3.tif” have identical spatial extent). These tiles are stored as GeoTIFF files in AWS S3 buckets tagged with their image or ASO collection identifiers and dates of collection. This preprocessing effort is completed via Jupyter notebooks (Kluyver et al., 2016) on AWS Elastic Compute Cloud (EC2) compute instances (see Figure X).



**Figure X:** Schematic of data preprocessing procedure for co-located Planet Labs Inc. satellite imagery and Airborne Snow Observatory snow mask data via Amazon Web Services cloud infrastructure.

Model Training

Our implementation of the training procedure is based in the Python programming language (v.3.5; Python Software Foundation, <https://www.python.org>), and is a heavily modified version of the “robosat.pink” software, an open-source set of command-line tools to enable machine learning with satellite imagery via the TernausNetV2 image segmentation network (Courtin and Hofmann, 2019; Iglovikov et al., 2018). The original software in this package was developed for three band remote sensing imagery and as such was not able to leverage multispectral data. We modified the package to enable the use of any N-band multispectral imagery data product and to allow for the use of cloud-based data storage and computation infrastructure.

To allow for quicker experimentation and simpler reproducibility, we packaged the training code, dependencies, and other software for our neural network implementation into a platform-agnostic computational working environment (or “container”) via Docker (Merkel, 2014). We used the AWS “SageMaker” service to manage the training of our network, which greatly simplified experimentation with different network parameterizations and datasets. We chose the “p2\_xlarge” AWS EC2 instance type for our training, as it afforded sufficient memory and graphics processing units for the task at hand.

We produced several different models for this study in order to assess the effects of training procedure on the final predictive accuracy. Each of our models was trained using data from a single Airborne Snow Observatory collection site. We pair a given set of binary snow mask tiles corresponding to a single ASO collection with the corresponding set of imagery tiles (with potentially some duplicates due to multiple Planet imagery collections within the temporal imagery search window), and divide this set of image-mask pairs into training and testing subsets via a 70%/30% split. This technique ensures that only Planet images that spatially and temporally overlap the ASO data are included in the training. Each training effort undergoes 50 epochs with a batch size of 7 and a learning rate of 2.5 x 10-5. The resulting model parameter weights are saved into an AWS S3 bucket. If there are multiple ASO collect dates for a single site, we repeat the above procedure for any additional ASO collections, but we initialize the model training procedure with the weights derived from the previous model training run. This allows the training process to build upon previous training runs.

**X.X:** Performance Evaluation

To assess the relative ability of our trained models to identify snow in Planet imagery, we designed an assessment scheme which allowed us to compare the predictions of our models to several other co-located remotely-sensed snow-covered area data products. For each comparable snow-covered area dataset, described below, we computed several metrics of pixel classification performance in accordance with standard practice. The metrics we computed are precision, which computes the percentage of snow classifications predicted by our model that are also snow classifications in in the compared dataset:

Recall, which computes the percentage of true snow classifications which are also true snow classifications predicted by our model:

F-Score or F1 score, which is the harmonic mean of precision and recall:

And balanced accuracy, which normalizes the true positive and true negative predictions by the number of true positive and true negative samples to allow for a less biased assessment of accuracy given the relative accuracy of each prediction type:

Using these metrics we assessed the relative performance of our model predictions compared to several other snow covered area datasets. These datasets are described in Table X.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data | Observation Type | Reference | Spatial Resolution | Temporal Resolution |
| ASO Snow Depth | Airborne lidar | Painter et al., 2016 | 3 meters | Weekly, during ablation season |
| MODIS Daily Snow Cover | Satellite | Hall, 2015 | 500 meters | Daily |
| Sentinel 2 NDSI | Satellite | Drusch et al., 2012 | 10 meters | 2-5 Days |
| Landsat 8 fSCA | Satellite | U.S. Geological Survey, Earth Resources Observation And Science Center, 2018 | 30 meters | 16 Days |

**Table X**: Snow covered area datasets used for comparison to model predictions.