



# Mapping and characterizing Arctic beaded streams through high resolution satellite imagery

Merritt E. Harlan<sup>a,\*,1</sup>, Colin J. Gleason<sup>a</sup>, Jonathan A. Flores<sup>a</sup>, Theodore M. Langhorst<sup>b</sup>, Samapriya Roy<sup>c</sup>

<sup>a</sup> Department of Civil & Environmental Engineering, University of Massachusetts, Amherst, MA, United States

<sup>b</sup> Department of Geological Sciences, University of North Carolina, Chapel Hill, NC, United States

<sup>c</sup> Planet Labs Inc., San Francisco, CA, United States

## ARTICLE INFO

Edited by Dr. Menghua Wang

### Keywords:

Beaded streams  
Object detection  
YOLO  
Planet imagery

## ABSTRACT

Arctic beaded streams provide unique ecosystem functions and serve as important tundra habitats. Their unique 'beads-on-a-string' morphology is thought to form from thermokarst erosion, and they are densely represented in permafrost-ridden landscapes. Despite their ubiquity in high latitude regions, beaded stream formation and occurrence is not well studied, and beaded streams are not globally mapped. Access to these streams is challenging in their remote, dynamic environment, and up until recently, monitoring these streams through satellite imagery was difficult given their relatively small size with channel widths of a few meters. The availability of high-resolution imagery from Planet data now makes it possible to detect and map these streams over large areas. Here we observe and predict the location of beaded stream catchments throughout the pan-Arctic domain by combining the location of known beaded streams with recent advances in computer vision and high-resolution (3 m) satellite imagery. Specifically, we use the location of known existing beaded streams to classify potential river catchments as beaded or non-beaded, then download high resolution imagery across those regions, and use the latest You-Only-Look-Once (YOLO) object detection algorithm to identify beaded streams throughout the pan-Arctic, estimating  $138,500 \pm 43,700$  beaded catchments globally, occurring in an estimated one third of all pan-Arctic catchments. In the largest dataset of beaded streams to date (Arp et al., 2015), only 375 catchments that contain beaded streams were identified, thus our estimate significantly expands our current understanding of the location and prevalence of Arctic beaded streams.

## 1. Introduction

Beaded streams are a common yet understudied thermokarst landform in permafrost ridden landscapes (Arp et al., 2015). These streams are characterized by their unique morphology of deep low-energy pools followed by high-energy chutes which appear as 'beads-on-a-string' (Oswood et al., 1989). Beaded streams are important sources of freshwater storage and serve as ecological habitats for species such as Arctic grayling (Heim et al., 2016; McFarland et al., 2018). Beaded streams are classified as a type of tundra stream (Craig and McCart, 1975) which typically flow from the foothills across the coastal plain in catchments composed of thin layers of peat underlain by permafrost (Hobbie and Kling, 2014). The pools, or beads, are often circular or elliptical, relatively deep (up to 2 m) and wide (1–35 square meters), and form from

thermokarst erosion at the intersection of ice wedge polygons (Péwé, 1966). The connecting chutes typically follow ice wedges, and thus tend to be shallower and narrower with straight, steep sides (Hopkins and Karlstrom, 1955).

Several field studies within the North Slope of Alaska have further characterized beaded stream water chemistry, temperature profiles, hyporheic exchange, as well as discharge and electrical conductivity profiles. From these studies we have learned that these streams may be sensitive to increased inputs of nutrients from road dust and fertilization of roadsides (Benstead et al., 2005), but seem to be resilient to increases in dissolved organic carbon (Larouche et al., 2015). Their unique morphology and thermal stratification allows them to react slowly to seasonal solar input and maintain thaw thicknesses longer (Brosten et al., 2006), and this thermal stratification also effects transient storage

\* Corresponding author.

E-mail address: [mharlan@umass.edu](mailto:mharlan@umass.edu) (M.E. Harlan).

<sup>1</sup> Present address: U.S. Geological Survey, Denver, CO, United States

of solutes, which in turn effects hyporheic exchange (Zarnetske et al., 2007, 2008). The water chemistry of these streams and their interaction with the subsurface and nearby lakes and broader streams can have broad implications for watershed nutrients (Wollheim et al., 2001).

Although these field studies provide important insights into the physical and biogeochemical makeup and processes of these streams, these findings came from a relatively small number of sample streams, and thus results may not be generalizable across all beaded streams. For example, a study conducted in a Russian province found the existence of beaded channels in the permafrost zone to be unstable (Tarbeeva and Surkov, 2013), which contrasts to relatively stable beaded channels in Alaska (Arp et al., 2015). Additionally, Tarbeeva and Surkov (2013) found beaded streams outside of the continuous permafrost region that were formed from other factors such as anthropogenic changes to flow regimes, unevenness of geologic structures, and ongoing karst processes. These divergent findings motivate the need to continue studying beaded streams across more representative regions of the Arctic.

In 2015, Arp et al. conducted an expansive survey of beaded networks from high resolution aerial imagery, identifying 445 beaded networks north of 66° latitude. However, the lack of availability of high-resolution snow free imagery significantly reduced the number of identifiable channels in Siberia and Canada. Their estimates suggest that there are over 1900 individual beaded networks throughout the world, with an approximate 13% in northern Canada, 18% in Alaska, and 69% in northern Russia (Arp et al., 2015). Mapping the approximately 1500 (1900 estimated total - 445 identified) additional beaded networks estimated by Arp et al. (2015) may help us better understand the formation, morphology, and occurrence of beaded streams. More accurate beaded stream maps may also help to better contextualize and broaden the impact of previous field studies that have shown the important role of beaded streams in Arctic ecology, hydrology, and biogeochemistry.

Assessing these streams on a global scale poses some challenges, as to date, they are not globally mapped. Beaded streams tend to be too small for global hydrography networks such as MERIT Hydro (Yamazaki et al., 2019) or HydroSHEDS (Lehner et al., 2008), while higher resolution stream networks such as the National Elevation Dataset (U.S. Geological Survey) or the Sentinel-ArcticDEM merged river network SARN (Lu et al., 2020) are not globally available. Because of their small size, with an average bead pool width of 6 m from field surveys (Arp et al., 2015), beaded streams are also hard to detect and identify from publicly available satellite imagery, as the resolution of 30 m with Landsat or 10 m with Sentinel-2 makes it difficult to resolve streams with widths < 10 m. The generally unknown location and logistical impossibility of a pan-Arctic field survey necessitates remote sensing, and yet these traditional sensors are ill-suited to the task.

Recent advances in the private remote sensing sector are promising, such as the PlanetScope optical satellite image dataset, which has 3 m, 4-band imagery with approximately daily near-global coverage. This extensive spatial and temporal coverage results in enough available imagery that eventually within the summer months these satellites can capture cloud-free, snow-free imagery throughout the pan-Arctic domain. Planet imagery has successfully been used to map surface water in high latitude and mountainous regions, from large rivers (Feng et al., 2019; Kaab et al., 2019; Strick et al., 2019) to wetlands (Cooley et al., 2017a) and lakes (Cooley et al., 2019; Lezine et al., 2021; Qayyum et al., 2020). Here we made use of the dense spatiotemporal resolution and availability of PlanetScope imagery to map and locate beaded streams globally to better understand their occurrence and formation.

To map beaded streams we 1) identified river catchments that potentially contain beaded streams from literature thresholds and global hydrography datasets, 2) downloaded tens of thousands of high-resolution PlanetScope satellite images within those catchments, and 3) employed a state-of-the-art computer vision algorithm to detect beaded stream reaches within each image. We therefore provide the first global assessment of these common high latitude streams. The remainder of this paper is organized as follows: first, we provide an

overview of the data sources and methods, including the classification scheme, and beaded stream detection through the YOLO-v5 computer vision algorithm (Jocher et al., 2021). Next, we present the results of both the catchment classification, beaded stream detection training, validation, and testing, and the global properties of beaded stream occurrence and catchments. Finally, we provide a comparison to previous studies including the Arp et al. (2015) survey, and discuss limitations of this approach, as well as future directions for both object detection of beaded streams and geospatial surveys more broadly.

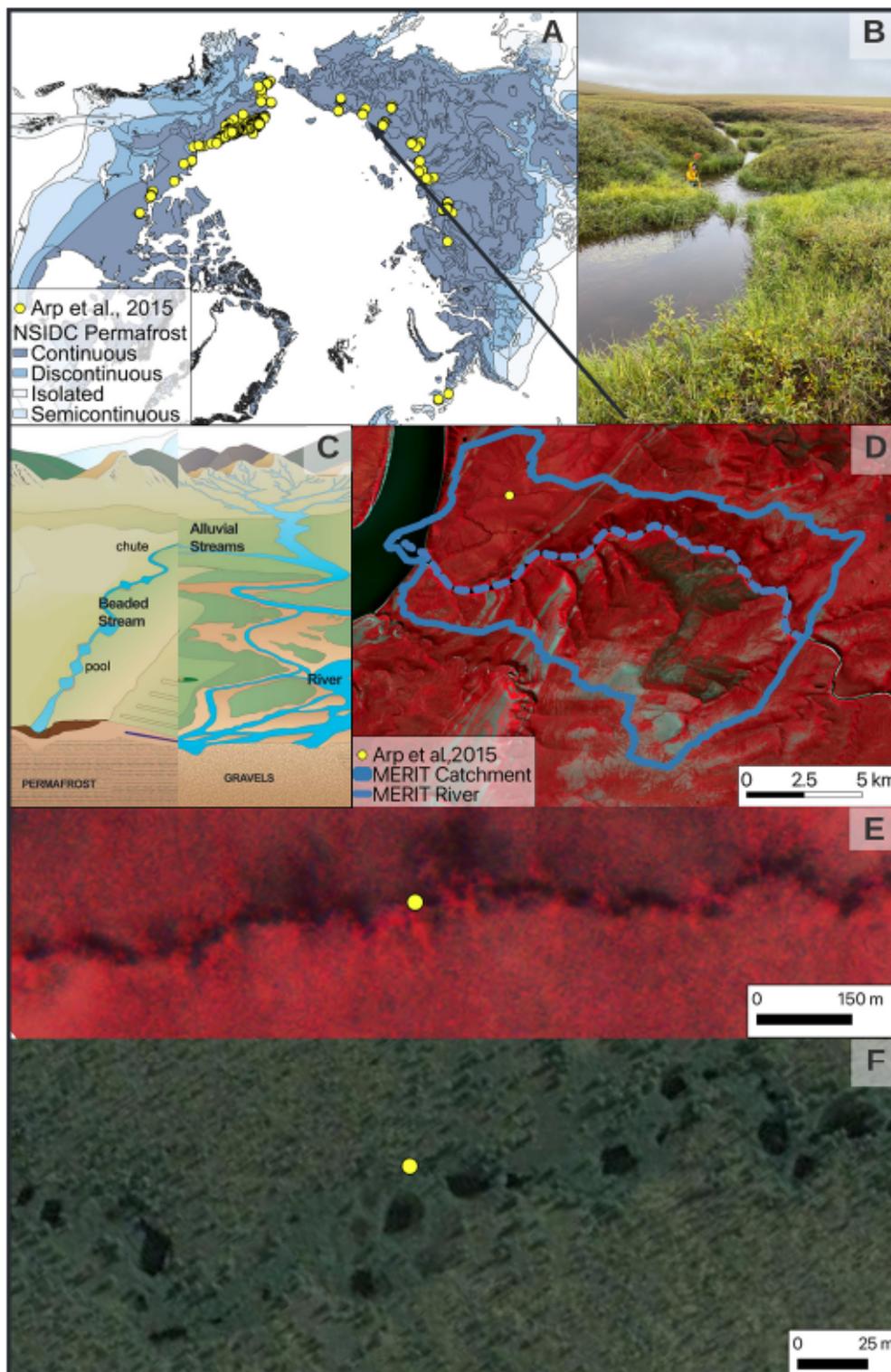
## 2. Data

Data acquired for this study consists of 3 m, 4-band PlanetScope imagery (Planet Team, 2021), a global hydrography dataset from MERIT Hydro (Yamazaki et al., 2019), a circum-Arctic permafrost and ground ice map from the National Snow and Ice Data Center (NSIDC; Heginbottom et al., 2002), and an a priori map of known beaded stream locations (Arp et al., 2015). Fig. 1 displays the locations of the 445 beaded streams mapped by Arp et al. (2015) underlain by the NSIDC permafrost coverage map, as well as a diagram of a beaded and alluvial stream adapted from Trochim et al. (2016), and an example beaded stream at varying resolutions in false-colour imagery. By design, Arp et al. (2015) narrowed their study to high latitudes, thus a high percentage of their streams are in continuous permafrost coverage. The conceptual diagram in Fig. 1 captures defining differences between beaded and alluvial streams, namely the presence of the characteristic pool-chute structure, underlying permafrost, and comparatively low slope regions.

To map river catchments that may contain beaded streams, we relied on global MERIT Hydro datasets. MERIT Hydro is based on the high-resolution (3 arc-second, ~90 m) high-accuracy MERIT (Multi-Error-Removed Improved-Terrain) DEM (Yamazaki et al., 2017) and several inland water maps. The data contain flow direction, flow accumulation, elevation, and river channel width. Here we relied on a new vector-based version of MERIT Hydro (Lin et al., 2021) that delineates MERIT Hydro into flowlines, watersheds, and basins using TauDEM software. Within this dataset, basin boundaries are redefined using the updated MERIT DEM resolution, then split with Pfafstetter coding to individual watersheds with a median global size of 461 km<sup>2</sup>, approximately similar to HydroBASINS level-08 classification (Lehner and Grill, 2013).

As shown in Fig. 1, Panel D, the MERIT Hydro river network in its vectorized form does not reliably extend to streams as small as beaded streams, with a 1 km<sup>2</sup> threshold for the river network. However, MERIT Hydro watersheds, referred to as catchments here, cover continuous areas that invariably contain beaded streams, and have relevant attributes associated with each catchment, including catchment area (km<sup>2</sup>), slope (m/m), upstream drainage area (km<sup>2</sup>), and channel width (m) from the Global River Width from Landsat (GRWL; Allen and Pavelsky, 2018) database. Within each catchment, channel width is represented as the average width of rivers wide enough to be observed by Landsat (> 30 m) within the catchment, thus does not include beaded channels. For Arctic rivers, MERIT-Hydro derived flowlines are considerably more accurate than HydroSHEDS (Lin et al., 2021), and thus this catchment delineation was chosen here for both its accuracy and available attributes. Since we scan images of each entire CART-selected MERIT catchment with YOLOv5 to detect streams therein, if we used more refined catchments (e.g., Lu et al., 2020) we would generate different catchment properties and have searched less images, but we would still have likely found the same beaded streams without needing additional geoprocessing required from SARN.

In addition to the catchment-level data, base maps such as the NSIDC Circum-Arctic Map of Permafrost and Ground Ice Conditions provide permafrost extent and ground ice content information. Permafrost is defined as ground that remains at or below 0 °C for at least two years, and permafrost extent is categorized into continuous (90–100% coverage), discontinuous (50–90%), sporadic (10–50%), and isolated



**Fig. 1.** Known beaded streams. Shown in yellow points are the locations of beaded streams found in the Arp et al. (2015) survey, used here for training purposes, underlain by the NSIDC permafrost coverage map. Permafrost is categorized as C: continuous, D: discontinuous, I: isolated, and S: semi-continuous. Panel B displays a photograph of beaded streams, and Panel C shows a conceptual diagram of beaded and alluvial streams adapted from (Trochim et al., 2016), highlighting the characteristic pool chute structure of beaded streams. Panel D shows a representative MERIT Hydro catchment and network delineation in blue, as well as the location of a beaded stream within the catchment. Panels E and F display closeups of the beaded stream shown in Panel D, underlain by PlanetScope imagery (Panel E), and aerial imagery (Panel F). Panels D and E display false-colour satellite imagery (Landsat and PlanetScope respectively), while Panel F displays true-colour imagery (Maxar Technologies, Map data ©2022 Google). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

patches (0–10%). Land type is broadly categorized as either *f*: “lowlands, highlands, and intra- and intermontane depressions characterized by thick overburden cover (>5–10m)” or *r*: “mountains, highlands ridges, and plateaus characterized by thin overburden cover (>5–10m) and exposed bedrock”, and has additional categories for glaciers, relict permafrost, inland lakes, ocean/inland seas, and land (Heginbottom et al., 2002). Although permafrost coverage has changed since 2002 (e.g. Biskaborn et al., 2019), given the Arp et al. (2015) findings of beaded stream stability over decades in time, beaded stream occurrence likely

has not changed dramatically in the past two decades.

Together, the Arp et al. (2015) stream locations, MERIT Hydro catchments, and permafrost base maps form the training data for the catchment level classification. These three datasets are spatially joined to overlay permafrost coverage on the catchments, and to find which catchments contain one or more beaded streams as identified in the Arp et al. (2015) paper.

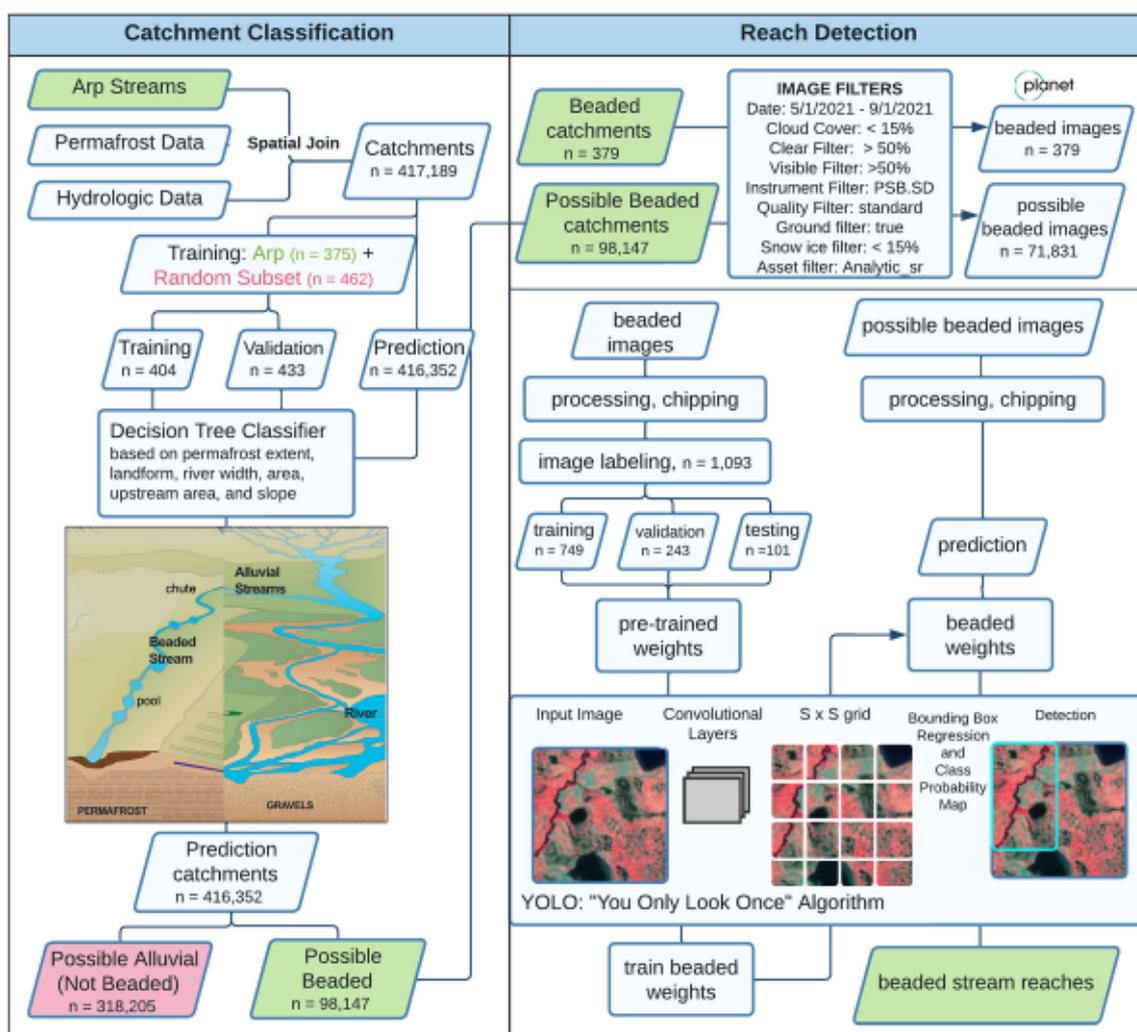
3. Methods

It is infeasible to manually scan every square kilometer of the Arctic with high resolution imagery to look for beaded streams. Fortunately, we can rely on previous literature to help narrow the range of where beaded streams could form. *Arp et al. (2015)* found that most beaded streams transition to alluvial channels with increasing drainage area and decreasing channel slope. In an earlier study which focused on three watersheds in Alaska, it was discovered that most beaded streams were in areas with marine silt deposits and that 61% of beaded streams initiated from lakes and 29% were from drained thermokarst lake basins (DTLB), with only a few initiating from hillslopes (*Arp et al., 2012*). Within the continuous permafrost zone in Alaska, *Farquharson et al. (2016)* confirmed these results, finding that beaded streams are less common in areas of aeolian sand and favor high ground ice content and greater topographic relief. Within the northeastern part of Yakutia, *Tarbeevea and Surkov (2013)* provided additional hydrologic bounds, finding that beaded streams occur in rivers with catchment areas ranging from 3 to 10 km<sup>2</sup>, gradients of up to 2 m/km, and with maximum flow rates of 0.5–1 m<sup>3</sup>/s. These findings limit likely areas with

beaded streams to regions with moderate to high permafrost coverage, in silty, relatively low-slope regions with high concentrations of lakes or drained lakes.

We applied these criteria to the MERIT Hydro catchments to identify potential regions that may contain beaded streams. To do so, we used a supervised classification approach to turn our a priori knowledge of beaded streams into a potential catchment identifier, shown in the left panel of *Fig. 2*, and described in further detail below. Our catchment classification enabled us to narrow the range of where beaded streams could be found. From this narrowed range of possible beaded catchments, we then downloaded PlanetScope imagery to detect beaded stream reaches (*Fig. 2* right panel).

A traditional geomorphology approach to detecting beaded streams might consist of generating a water mask from the images and combining that water mask with a digital elevation model (DEM) to determine stream locations within the catchment, for example following methods from *Lu et al., 2020* and *Lu et al., 2021*. From there, beaded streams could be differentiated from non-beaded (alluvial) streams by their size, width profile, slope, and/or drainage area, following methods such as *RivWidthCloud (Yang et al., 2020)*, *RivaMap (Isikdogan et al.,*



**Fig. 2.** Methodology outline. On the left panel, the beaded catchment classification is outlined, while on the right the beaded reach object detection workflow is outlined. Input to the classification involves the beaded streams found in the previous *Arp et al. (2015)* study, alongside permafrost and catchment data. The classifier is first trained on known beaded and non-beaded catchments, then applied globally to the remaining catchments to predict a binary beaded/non-beaded decision tree. An example of a typical alluvial and beaded stream is shown in the center of the panel; both are often found in near proximity to each other, thus a “beaded” catchment does not preclude alluvial streams within. On the right panel, the beaded reach detection workflow includes filtering imagery to the catchments and image quality filters, then proceeding with image processing, chipping, and labeling of the training data. Like the catchment classification, the object detection algorithm (YOLOv5) is trained on known beaded imagery that were manually digitized, and then applied globally to detect beaded stream reaches.

2017), and TauDEM (Tarboton, 1997). However, a challenge with this approach is generating an accurate water mask, along with the computational time associated with generating the river network across thousands of catchments. Beaded streams, unlike larger rivers and lakes, tend to have a different spectral signal, and can be difficult to mask with common indexes such as the modified normalized difference water index. Although thresholding can help solve this issue (Cooley et al., 2017b; Lu et al., 2021; Yang et al., 2014), it is still possible that some beaded streams might be confused for vegetation or small unconnected ponds.

Instead of detecting beaded streams from a pixel-based method such as index thresholding, an alternative approach is through object detection, a computer vision technique. Unlike a classification algorithm and object-based image analysis, object detection is a technology that detects instances of semantic objects, such as detecting floating objects in rivers (Kale and Chaczko, 2015), identifying invasive fish species (Zhang et al., 2016), or in this case mapping beaded stream reaches. Given the task at hand of detecting beaded streams globally, this approach was promising in that it avoided having to distinguish first water masks and river networks, and from there beaded networks within an entire catchment full of hydrologically diverse features, and focused instead on solely detecting beaded streams, providing an end-to-end methodology.

Combined, our catchment classification and beaded reach detection using computer vision enabled us to efficiently map beaded stream locations throughout the pan-Arctic. Below we provide further details on our classification methods, object detection algorithm, and metrics for evaluating these methods.

### 3.1. Catchment identification methods

To identify beaded catchments, we classified all MERIT Hydro catchments within the pan-Arctic domain as either beaded or non-beaded using the rpart R package which implements classification and regression trees (CART) following Breiman et al. (2017). Variables used in the classification include catchment properties (area, upstream area, slope, and river width) as well as permafrost properties from NSIDC (land type, permafrost extent). To train the classifier, we identified MERIT Hydro catchments that overlapped with Arp et al. (2015) beaded stream locations (375 in total), and randomly selected an additional 462 MERIT catchments, which were then manually classified as either beaded or non-beaded from visual inspection of Google Satellite imagery (2022 Landsat/ Copernicus, Maxar Technologies, Map data 2022 Google). Arp et al. (2015) identified beaded streams through a nested survey approach, scanning cloud-free, ice-free imagery in Google Earth, as well as aerial transects in the North Slope of Alaska, and high-resolution photography in the Fish Creek watershed in Alaska. Each of the Arp et al. (2015) stream locations were manually inspected using Google Satellite imagery to confirm the presence of a current beaded stream. From the combined 837 classified catchments, we randomly split approximately half of the data into training (433 catchments) and validation (404 catchments). Of these 837 catchments, 379 were beaded, 375 identified from Arp et al. (2015) locations, and 4 identified from manual inspection of the 462 catchments that were randomly selected (the remaining 458 were classified as non-beaded). From here, we trained the CART decision tree classifier to predict differences between beaded and non-beaded catchments across the pan-Arctic from our 433 training catchments, with the goal of reducing the total area of high-resolution imagery needed for object detection. We validated this approach on the 404 validation catchments.

There are two main processes in building the regression tree for the CART model. First, the single variable which best separates beaded and non-beaded streams is chosen, and the data is then recursively separated into subgroups, until the subgroups reach a minimum size, or no further improvement is made. The second process uses cross-validation to trim back the tree, reducing the complexity. Each split is based on the Gini impurity index (Breiman et al., 1984), which is calculated as the

probability of mislabeling an element (or catchment in this case) assuming that element is randomly labeled from the distribution of all classes in the set (in this case beaded and non-beaded).

This approach also provides confidence values for the degree to which catchments have been classified as beaded or non-beaded, allowing us the flexibility to further narrow or expand the range of possible beaded streams. For this analysis, a standard binary classification threshold of 0.50 for the confidence value was used to discriminate beaded from non-beaded catchments, where all confidence values > 0.50 indicate a CART-classified beaded stream. Thus, all catchments classified as beaded were included. Following the CART model training (results described below), we classified all pan-Arctic catchments as either possibly beaded (98,147 catchments) or non-beaded (318,205 catchments), approximately reducing the total area fourfold.

### 3.2. Image processing methods

We identified 71,831 PlanetScope images from May August 2021 that covered these possibly beaded catchments after filtering for cloud coverage, snow coverage, visibility, clearness, and the type of instrument, quality, ground, and asset filter. Each of these filters are defined in PlanetScope metadata (Planet Team, 2022). During these summer months, ice break-up in the river channels generally occurs by mid-June, generating peak flows, and stream temperatures typically start declining mid-August (Heim et al., 2016). These filtered images cover 90% of the potential catchments (88,250 out of 98,147), and there was no discernible spatial pattern to the catchments where beaded imagery of sufficient quality was not available, thus this omission is spatially non-biased. Data were limited to summer 2021 to take advantage of the latest Planet sensors and limit the quantity of this commercial product needed.

To identify beaded reaches within these images, we used object detection methods. Object detection methods can be broadly split into neural network approaches that rely on deep learning, and non-neural approaches. Deep neural networks constitute some of the latest advancements in the field and tend to perform better than traditional, non-neural algorithms (Erhan et al., 2014; Zhao et al., 2019). Zhao et al. (2019) classify generic object detection (locating and classifying objects in bounding boxes) into two main subcategories: regional and regression/classification. The regional method first generates regions to classify, then classifies each region into different categories, and includes popular methods such as R-CNN (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2015), and Mask R-CNN (He et al., 2017). The regression/classification approach adopts a unified framework to both categorize and locate objects at once, allowing for real-time detection, and include algorithms such as MultiBox (Erhan et al., 2014), SSD (Liu et al., 2016), and YOLO (Jocher et al., 2021; Redmon et al., 2016; Redmon and Farhadi, 2016, 2018; Wang et al., 2021).

YOLO (Jocher et al., 2021), or You-Only-Look-Once, uses an artificial neural network to predict the probability that the object it is trained to find (e.g., beaded streams) is within a bounding box that exactly bounds the feature. Thus, YOLO does not yield polygons or rasters of beaded stream outlines as in traditional object-oriented classifications, but instead gives a probability confidence that there is exactly one beaded stream reach within each YOLO bounding box. These boxes encompass the entirety of the detected feature, and therefore there is a 1:1 mapping of the number of bounding boxes to the number of beaded stream reaches.

YOLO makes its predictions after looking once at each image using non-max suppression, a technique to select the best bounding box out of potential overlapping bounding boxes to avoid objects detected more than once (Redmon and Farhadi, 2016). YOLO's confidence score is computed as the product of conditional class probability (in this case the probability that the box contains a beaded stream) and the box confidence score (or the accuracy of the area and location of the box).

Because the algorithm “only looks once” it can be optimized faster compared with other object detection algorithms, with similar or higher accuracy rates (Srivastava et al., 2021; Tan et al., 2022). We chose YOLOv5 (Jocher et al., 2021), one of the latest versions of YOLO, for this study based on these published results.

Our goal in this study is to detect the beaded streams within the catchments classified as possibly beaded in Section 3.1. Each PlanetScope image within the 88,250 catchments (71,831 images- some images covered multiple catchments) was cropped to the catchment file, equalized to ensure the image was visible with sufficient contrast by standardizing the histogram, converted to 8-bit imagery, and ‘chipped’ to smaller 512 × 512 pixel images. This resulted in a total of 2,188,531 individual chips for classification. This chipping produces standard images that run quickly with computer vision and is essentially inverse mosaicking.

PlanetScope images containing a known beaded stream derived from the Arp et al. (2015) study were manually inspected through the LabelMe Annotation tool (Wada, 2022) to confirm the presence of beaded streams within the training chips, yielding 1093 image chips with a beaded stream (Fig. 3). For each chip, we labeled (other common terms for this process include ‘digitally identified’ and ‘digitized’) beaded streams by drawing a unique box for each beaded reach. Here we define a beaded reach as a section of a beaded stream where the beaded pattern and width were consistent. For example, if both a tributary and a larger connecting stream were beaded, one bounding box would encompass each branch for a total of two boxes. Any streams that exhibited a pool-chute structure that appeared to be low order streams were included. Each of the annotated images from LabelMe were then cross-checked and converted to the YOLO format through the Roboflow

Annotation Tool (<https://roboflow.com/>).

These labeled boxes associated with individual PlanetScope image chips represent the training objects for YOLO. For this study, we used a single-class classification, only training on beaded streams, instead of creating multiple training classes of other features also present in Arctic catchments (e.g., alluvial streams, ponds, lakes, roads). This simplified the generation of manually digitized datasets and reduced the computational cost required for training and classification within the YOLOv5 algorithm, likely at the slight cost of commission error accuracy, discussed further in results and discussion.

Following image preprocessing, we then trained the YOLOv5 algorithm (Jocher et al., 2021) with the 1093 training chips. Hyperparameters needed to train the algorithm include the size of the training set (‘batch size’) and number of times the model will work through the training set (‘epochs’). To train YOLOv5, we relied on a standard batch size of 12, with 300 epochs, although the algorithm stopped improving after only 147 epochs. Training the model was relatively fast (approximately 1 h in total) with 1 graphics processing unit (GPU). GPU processing is necessary for this study, and therefore access to GPUs is a prerequisite for this work as CPU processing YOLOv5 on 71,831 images would be computationally extensive.

Within YOLOv5 training, the algorithm uses the manually labeled bounding boxes of known beaded streams to refine the prediction of beaded stream reaches for each trained image chip. Therefore, we do not input beaded stream objects or outlines into the training- we use a manually drawn bounding box around each beaded stream. To test the risk of overfitting, we divided the training chips using a 70/20/10 training/testing/validation split and compared results from 10 random samples of this split. From the 10 training samples, the training model

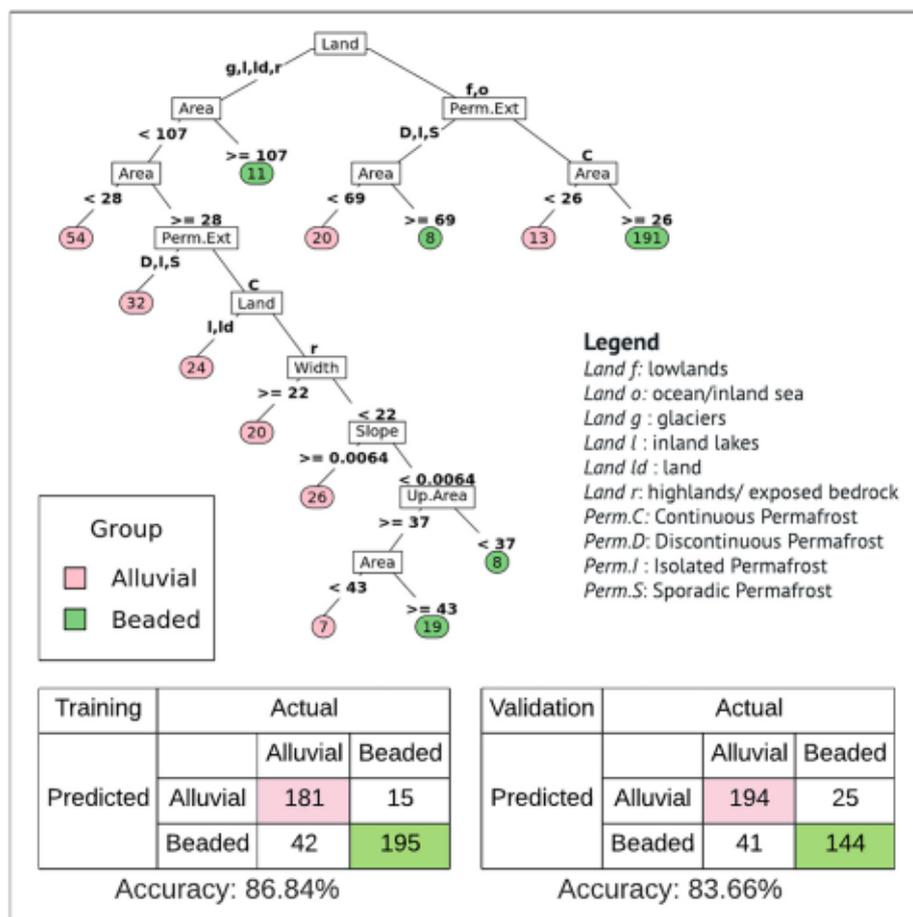


Fig. 3. Beaded catchment classification results. On top, the tree diagram shows the splits created from the training data (half of the dataset of beaded and non-beaded catchments). Variables are shown in boxes and ordered from most important (top) to least important (bottom). The values or attributes of each split are shown in bold, e.g. for the first split, data is categorized as either land type ‘g’, ‘l’, ‘ld’, or ‘r’, or as ‘f’ or ‘o’. The legend explains each of the land type (Land) and permafrost extent (Perm.Ext) acronyms. Slope is in units of m/m, width in m, and area and upstream area (Up.Area) in km<sup>2</sup>. The number of stream catchments correlated with each split are shown circled in numbers and are denoted as either beaded (green) or alluvial/non-beaded (pink). Shown in the bottom are the confusion matrices for the training (left) and validation (right) datasets, along with the overall accuracy. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with the median accuracy was chosen for the beaded reach detection so as not to overstate the results.

After training YOLOv5, we applied the computer vision model to the 71,831 PlanetScope images, chipped into 2 million discrete images, to identify beaded reaches, summarizing them into a total number of beaded stream catchments. We define a beaded stream catchment as a catchment that likely contains one or more beaded streams as identified from both our catchment classification AND computer vision object detection. Our analysis is focused on beaded catchments, as opposed to beaded stream reaches or bounding boxes, in order to be able to compare with pan-Arctic beaded catchments and the river networks identified in the [Arp et al. \(2015\)](#) study.

We estimate the total beaded catchment count from both the catchment classification and beaded reach detection from YOLOv5. To arrive at this count, we first compiled all the beaded catchments identified from the classification and predicted beaded reach boxes within those catchments. If a catchment contained a beaded stream with a YOLOv5 confidence score of 0.1 we identify it a possibly beaded catchment. From there, we account for commission and omission through the testing and training results, described in greater detail below.

### 3.3. Training metrics

To validate both our CART regression training and YOLO, we rely mainly on omission and commission errors, as well as an overall accuracy metric, defined below for CART:

For the CART analysis, accuracy refers to the proportion of correctly categorized beaded and non-beaded catchments over the total number of predicted catchments. We define our omission error as the proportion of beaded catchments misclassified as non-beaded out of the total number of beaded catchments, while our commission error is the proportion of non-beaded catchments misclassified as beaded out of the total number of all catchments classified as beaded.

Accuracy for YOLO refers to the number of correctly identified beaded reaches divided by the sum of total beaded reaches in the validation set and misclassified non-beaded reaches. A beaded stream was deemed to be correctly identified so long as a portion of the predicted and actual bounding box overlapped. Traditional object detection typically calculates error and accuracy in terms of the area of overlap,

but in this case any overlap was considered a positive outcome, since our goal is to map the locations of beaded streams, and not necessarily outline them with a high degree of precision.

This accuracy metric helps simplify our commission and omission errors into one value, but often a suite of error metrics and accuracies are used in computer vision work ([Padilla et al., 2020](#)), which we compute here:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{5}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{6}$$

Recall and precision are identical to our omission and commission error formulas, with the exception that now we are referring to beaded boxes encompassing beaded reaches, not catchments. From these metrics, we can also compute the average precision (mAP), or the mean precision over a range of recall values for a confidence threshold of 0.5. For our final estimate of beaded catchments, we combine the omission and commission errors from both the catchment classification, image download, and object detection:

$$\text{Accuracy} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives} + \text{False Negatives}} \tag{7}$$

$$\text{Omission Error} = \frac{\text{False Negatives}}{\text{True Positives} + \text{False Negatives}} \tag{8}$$

$$\text{Commission Error} = \frac{\text{False Positives}}{\text{True Positives} + \text{False Positives}} \tag{1}$$

where  $\frac{\text{Beaded Catchment}_{\text{pred beaded}}}{\text{Beaded Catchment}_{\text{true}}}$  represents the fraction of correctly identified beaded catchments divided by the true number of beaded catchments (or bounding boxes in the second half of Eq. 7). In Eq. 8, the formula is identical, except the denominator now represents the total number of predicted catchments or bounding boxes, including those incorrectly identified, and here we do not factor in the error from the image availability. Together, these errors represent the percentage of all beaded streams that were likely to be missed from the analysis, and the percentage of alluvial or non-beaded streams or regions misclassified as a beaded stream from the combination of both the catchment classification and object detection.

## 4. Results

### 4.1. Catchment classification accuracy and findings

We tested our CART classifier on our set of testing catchments, finding an accuracy of 83.66%, with 14.80% omission error and 22.16% commission error ([Fig. 3](#)). Here we are primarily concerned with omission, as we are identifying potential locations for further computer vision work to assess whether or not a beaded stream exists with a catchment. Commission error is therefore acceptable provided it does

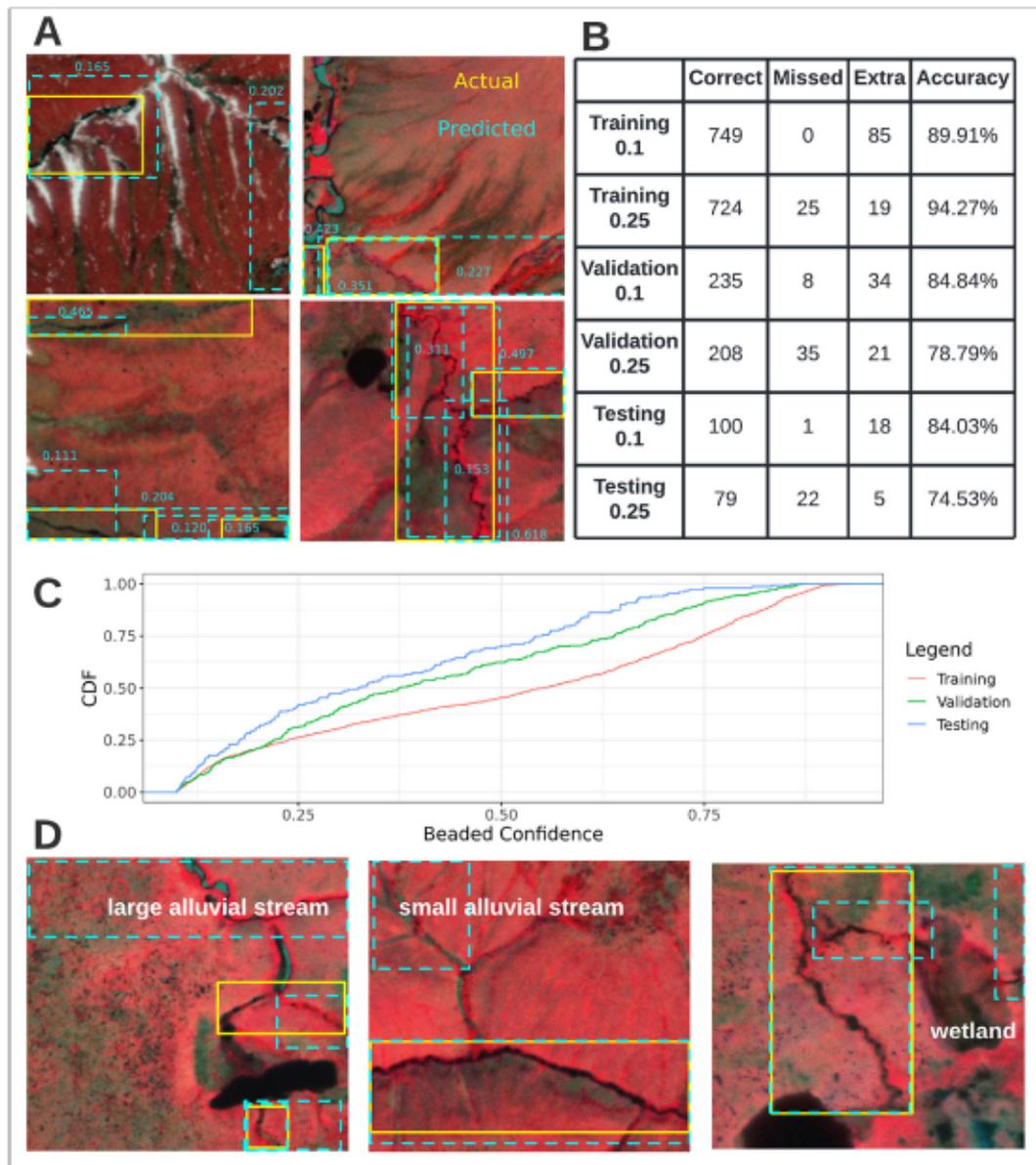
not produce too large a global surface, but omission error will propagate into our final analysis. Therefore, we begin our process with up to 15% underestimation of catchments containing beaded streams based on our omission accuracy. Applying this classifier on our global data identified 98,147 catchments, or roughly  $\frac{1}{4}$  of all pan-Arctic catchments, as “possibly beaded.”

As shown in Fig. 3, our classification showed that the variable that most distinguished beaded and alluvial streams in the training dataset was landcover, with lowlands (‘f’) and inland seas (‘o’) associated with beaded streams, and glaciers, land and inland lakes associated with alluvial streams. Within the lowland and inland sea category, beaded streams were further distinguished by continuous permafrost coverage and catchments with areas  $>26 \text{ km}^2$ , or if permafrost was discontinuous,

semicontinuous, or isolated, in catchments with areas  $>69 \text{ km}^2$ .

#### 4.2. Object detection accuracy and findings

In training YOLOv5 on our manually identified beaded stream objects, we found accuracies ranging between 76.2% and 99.2%, with a mean accuracy of 84.9% across our ten samples, indicating some sensitivity to the training sample. The training model mean average precision (mAP) is 0.716, while the precision (commission error) is 0.805, and the recall value (omission error) is 0.737. To put the mAP value in context, mAP values range from 0 to 1, with 1 representing a perfect detection. Representative YOLOv5 mAP values for classification on the Microsoft COCO (Common Objects in Context; Lin et al., 2014)



**Fig. 4.** Object detection training and validation. Panel A displays four sample PlanetScope image chips in near infrared false colour composite. The yellow boxes mark labeled bounding boxes, manually digitized for training, and the cyan dashed boxes represent the predicted bounding boxes from the YOLO algorithm. Depicted values are the probability that each box accurately encompasses a beaded stream. Panel B shows results from the training, validation, and testing sets. Predicted boxes are classified as “correct” if they overlap with a labeled box (top left image on the left), and “extra” if they do not (top left image on the right). If no predicted box overlaps with a labeled bounding box, then it is marked as “missed”. “0.1” represents values from all predicted boxes with a beaded confidence  $>0.1$  (10%); likewise, “0.25” represents all predicted boxes with a confidence value  $>25\%$ . Panel C shows the cumulative density function (CDF) of beaded stream confidence values from the testing, training, and validation datasets. Three example ‘extra’ bounding boxes showing misclassified non-beaded stream features are displayed below in Panel D, denoted by turquoise boxes that do not overlap with beaded stream boxes, shown in yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

benchmark dataset ranged between 0.457 and 0.727 at a 0.5 confidence threshold (Jocher et al., 2021). Thus, our mAP value of 0.716 is at the higher end of training results from the COCO dataset, implying adequate performance.

Fig. 4 displays results from YOLOv5 training on this median model across two confidence score thresholds, 0.1 and 0.25, which include all predicted bounding boxes with overall confidence scores greater than those thresholds. As the threshold shifts from 0.1 to 0.25, the percentage of omission (missed) and commission (extra) errors change accordingly, with omission increasing and commission decreasing. Unlike the catchment level classification, both errors propagate into the overall mapping of beaded streams, but the confidence scores help us better estimate the rate of commission and omission. In terms of the

distribution of the confidence scores, approximately half of all predicted beaded reach bounding boxes have confidence scores between 0.1 and  $0.35 \pm 0.05$  depending on the dataset, as shown in the cumulative density function plot in Fig. 4.

As the rate of omission is low using the 0.1 confidence threshold (0%, 3.4%, and 1% omission for training, validation, and testing, respectively), we rely on this value for the remaining images in the prediction set. Our commission errors at the 0.1 confidence threshold are higher (10.2%, 12.3%, and 15.1%), indicating that we could expect approximately 10–15% of our predicted boxes to misclassify a non-beaded object as a beaded stream, or in our case exactly 18 ‘extra’ boxes in our testing dataset. Of these 18 ‘commission’ boxes, with examples shown in Fig. 4 Panel D, misclassified non-beaded features include small alluvial

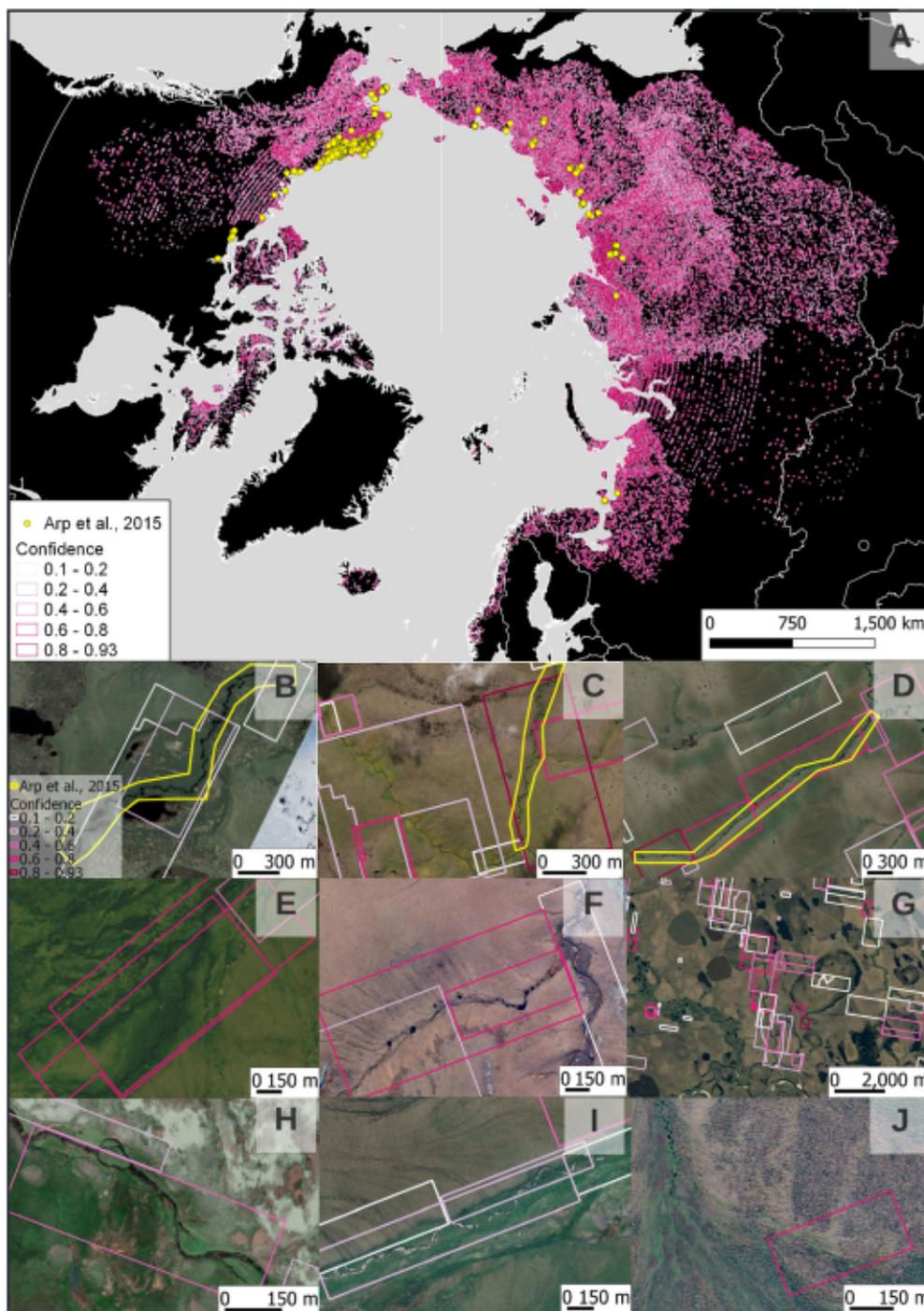


Fig. 5. Beaded reach predictions. Panel A displays a map of the pan-Arctic domain with the Arp et al. (2015) beaded stream locations shown in yellow, along with the bounding boxes of predicted streams colorized by the confidence value, ranging from 0.1 to 0.93. Panels B-D show three beaded stream locations identified by both the Arp et al. (2015) study (manually outlined in yellow) and this study, shown in predicted bounding boxes. Panels E-G display beaded streams newly identified in this study, while Panels H-J show misclassified streams. Panels B-J are all underlain by Google satellite imagery (©2022 Landsat/ Copernicus, Maxar Technologies, Map data ©2022 Google), which helps to capture some of the variety in appearance of beaded and high latitude small streams. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

streams with widths <5 m (7 boxes), larger alluvial streams (width > 10 m; 4 boxes), and wetland and floodplain areas (7 boxes). No lakes, ponds, or roads, were misclassified as beaded streams, justifying our single class approach.

From the remaining YOLOv5 predictions, aggregated to a total number of catchments, we can set informed bounds on the number of global beaded stream catchments, which is our 160,866 catchments +32,500 for probable commission error – 54,900 for probable omission error for a total of 138,500 ± 43,700 catchments. This estimate of 138,500 ± 43,700 beaded catchments is likely an overestimation, but we infer that the true number is nevertheless orders of magnitude greater than the Arp et al. (2015) initial estimate, who estimated 1900 beaded stream networks. These beaded networks may overlap several catchments, but likely on the order of 0.5–10 catchments per stream network, not 50–100 as would be needed here if only 1900 beaded networks existed. By design, commission is a greater source of error than omission because of our choices of thresholds and desire to locate all beaded streams, hence our likely overestimation. Fig. 5 below displays results from the object detection, a visual comparison between the Arp et al. (2015) streams and detection from this study, as well as some examples of correctly and incorrectly identified beaded streams.

The beaded catchments we identified are primarily in continuous permafrost (70.5%), but several beaded streams were identified in semi-continuous (5.5%), discontinuous (8.5%), and isolated permafrost (4.4%) catchments. In terms of land type, less than expected were found in lowlands (42.1%). This result is both a factor of the initial tree classification, which identified 57.9% as lowland, and the beaded reach object detection, which predicted beaded reaches in 74% of lowlands. Fig. 6 below contextualizes our catchment findings in regard to slope, stream order, permafrost coverage, and continental coverage compared with the Arp et al. (2015) study.

Error in our estimation of beaded stream locations arises from three main sources: 1) the catchment level classification, 2) image availability and 3) the beaded stream object detection. The majority of this error comes from catchment level classification, with omission and commission errors of 17.36% and 22.2% respectively, whereas the image

availability contributes an approximate 10%. For the YOLO object detection errors, for omission we find values of 1% and for commission we find 15.25%. From all three error sources combined, we expect an omission error of 20.2%, and a commission error of 34.1%.

### 5. Discussion

We present the first global map of Arctic beaded streams by combining a supervised classification and a recent advancement in object detection algorithms with high-resolution satellite data. From our findings, we estimate that beaded streams are more common than previously thought, occupying 138,500 ± 43,700 catchments globally. Further, we find that an estimated 30% of beaded streams reside outside of the continuous permafrost zone, confirming findings from Tarbeeva and Surkov (2013) who noted the presence of beaded streams outside of continuous permafrost. The range of catchment slopes and upstream areas is much larger than previously thought, as is catchment size, implying more heterogeneity within beaded stream distribution than previously identified within the Arp et al. (2015) study.

From our CART analysis, we learned that the attributes most closely attributed with the beaded and alluvial training data in order of importance are land type, permafrost coverage, and catchment area.

Encouragingly, our classification findings trained on the Arp et al. (2015) study generally match findings from other beaded stream literature, which note the high presence of beaded streams in continuous permafrost regions in lowlands and low-slope regions, and find that beaded streams tend to arise from larger catchments with drainage areas >1 km<sup>2</sup> compared to alluvial hillslope channels (Arp et al., 2012; Farquharson et al., 2016). In areas outside of lowlands or inland seas, beaded streams are again distinguished by a larger catchment area, as well as continuous permafrost, catchment river widths <22 m, slopes <0.0064 m/m, and upstream areas <37 km<sup>2</sup>, and if not, catchment areas >43 km<sup>2</sup>.

These findings are not representative of all beaded streams, and several of the beaded stream catchments found in the Arp et al. (2015) study used in this training set were not identified from our regression

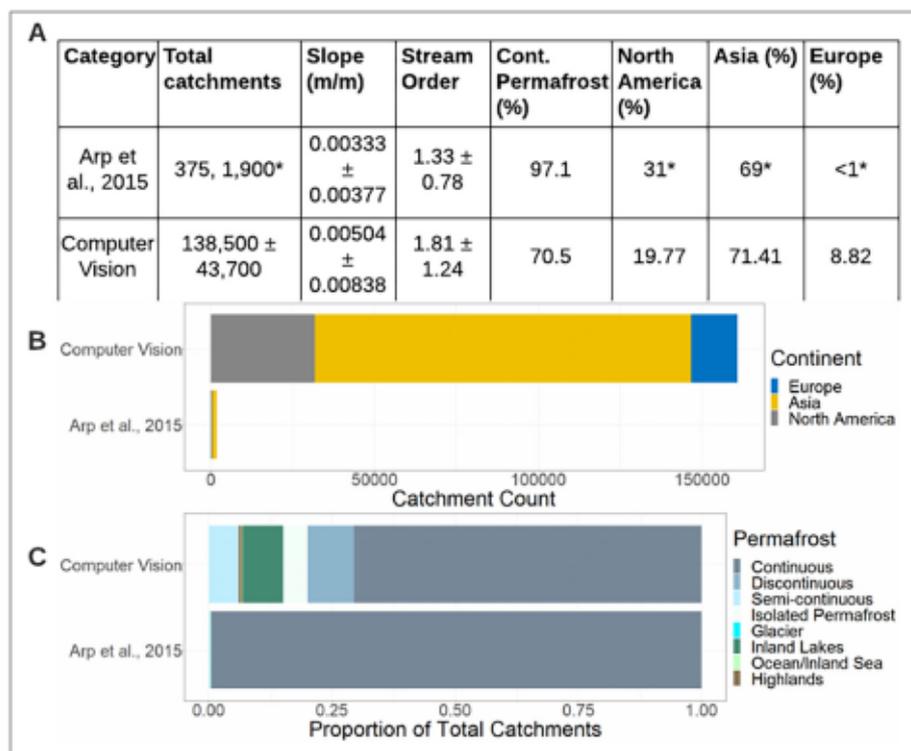


Fig. 6. Comparison of catchment properties. The table in Panel A displays the number of catchments, slope (m/m), stream order (1–6), percent of catchments in continuous permafrost regions (%), and percent of catchments within North America, Asia, and Europe respectively. Results are presented for the Arp et al. (2015) study and this study. The catchment distributions (slope and stream order) are signified by the mean ± the standard deviation. Panel B shows the distribution of each study in terms of the number of catchments in each continent, and Panel C shows the distribution of permafrost and landcover scaled across each study. \*For the Arp et al. (2015) row, we include the number of catchments that one or more identified streams (375), with the number of estimated stream networks (1900) and include their estimated values for the percentage of beaded streams across geographic regions.

trees (15 out of 210 catchments in the training set, and 25 out of 169 in the test set). Further, these categorizations may also be biased towards the set of beaded streams the [Arp et al. \(2015\)](#) study identified. To avoid some of this bias, several attributes or variables of the catchments were purposefully not included, including latitude, longitude, continent, distance to known beaded streams, or other variables that may have geospatially biased the end results. However, while this regression tree does not likely encompass all beaded stream catchments, it significantly helped to narrow the scope of pan-Arctic regions, and the choices within the regression tree are largely confirmed by trends from literature. Despite the credibility of these findings, the catchment level classification remains the largest source of error throughout this workflow, revealing that detecting beaded streams through catchment properties has its limitations.

On the other hand, detecting beaded streams through the YOLO algorithm provided relatively accurate results, when combined with multiple thresholds, although the accuracy of results was sensitive to the training data. In general, this approach was largely suitable to the task of identifying beaded reaches throughout the pan-Arctic, with fast training and prediction times of around 0.02 s per chipped image using 1 GPU with 40 cores and 60 GB of memory. When comparing catchments identified from the YOLOv5 algorithm with the catchments identified in the [Arp et al. \(2015\)](#) study, we find that catchments identified here have higher slope and stream order than the catchments identified in the [Arp et al. \(2015\)](#) study, and lower permafrost coverage and catchment area. The initial estimates of the distribution of beaded streams throughout North America, Russia (Asia) and Europe within the [Arp et al. \(2015\)](#) study are for the most part similar to our estimates here, and similar to the regional distribution of all catchments, but with slightly more beaded catchments identified in Europe than North America, as visualized in the Panel B of [Fig. 6](#). Possibly the largest difference is within the percentage of beaded streams within continuous permafrost regions shown in [Fig. 6](#) Panel C, indicating that our work implies a larger percentage of streams outside of continuous permafrost.

Although our computer vision work was both quick and efficient, there were some challenges associated with applying computer vision to the task at hand of identifying beaded streams. One limitation in particular was having to rely on boxes to outline the training data, as they encapsulated mostly land surrounding the beaded streams, and it was difficult to follow the morphology of the stream reaches with just a box. Further, the output of bounding boxes made it challenging to translate our findings into a succinct value of beaded streams, or beaded stream networks, hence we conducted most of our analysis at the beaded catchment level. At this catchment level, our ability to quantify beaded stream characteristics is thus limited to passing on catchment-wide characteristics such as land cover and permafrost cover from other data sources, and the strength of our analysis is limited by the assumptions between the relationship of MERIT Hydro-observable rivers and small beaded streams, and the uniformity of catchment land and permafrost cover.

To more accurately characterize and delineate beaded streams, we could have relied on a more traditional geomorphology approach of detecting streams from digital elevation model stream burning that may detect stream networks easier, but at the cost of computation time and potentially accuracy as well ([Lu et al., 2020](#)). Instead of object detection, another approach within computer vision is image segmentation, which would classify individual pixels within an image as beaded or non-beaded, enabling a closer look at properties such as the area and shape of beaded streams. Given the relative width of beaded streams (~6 m) compared to the pixel size of a PlanetScope image (3 m) sub-pixel classification would be needed to accurately map beaded streams. In this case, we chose the faster, more efficient method of object detection, with output of bounding boxes rather than pixel-level classification, but note that continued work, possibly with higher-resolution imagery, could explore these methods as well. Further access to GPU clusters and the continued translation of computational efficiency from

computer science to the sciences is also needed to make some of this work feasible.

In general, the suite of methods relied on within this study (classification, object detection, confidence thresholds, what denotes a beaded stream in training) is sensitive. Throughout this method, we made many informed choices regarding the selection of classification training data, permafrost and landcover variables, PlanetScope image thresholds, and computer vision algorithms, and each of these choices impacted and likely introduced errors to our outcome. Ultimately, our choice was to err on the side of commission, and thus we based our decisions on attempting to include as many beaded streams in our final output as possible. A potential downside of allowing higher commission error is reducing the computational efficiency of YOLO, but the reduction of speed was not an impediment to our goals, as the intent of this study was not near real-time detection, and we rely on GPU processing.

Other limitations of this study include the limitations on the number of high-resolution Planet images we were able to access, process, and download. If obtaining and downloading high resolution imagery everywhere throughout the pan-Arctic was possible, we could have relied solely on the computer vision method, instead of first classifying catchments via a CART classifier, which we introduced to limit the total area we downloaded imagery for. We were also limited by the amount of beaded stream training data we had access to, as identifying streams beyond the [Arp et al. \(2015\)](#) set was difficult, hence the need for automation in this study. Future studies and mapping of beaded streams could hopefully iterate from the data and locations we have provided here, to further improve our error metrics.

Other potential future directions for this work include increasing the number of classes used in the training data, for both the catchment level classification and/or the beaded object detection work. For example, we could have classified catchments as beaded, alluvial, and lake, or classified bounding boxes as containing beaded streams, alluvial streams, or wetlands to try and decrease our commission error. Although our likely commission is high at 34.1% with the 0.1 confidence threshold, a silver lining of this commission error is that we have also identified likely water features and rivers throughout the pan-Arctic domain that are also unmapped, even if they might not be technically beaded, as most of our misclassified predicted boxes in the testing set were other, relatively small Arctic streams. Using this methodology with a wider set of training data may be helpful for detecting and mapping various types of small Arctic streams and headwater streams beyond beaded streams.

Although our uncertainties are large, they vastly improve upon our current understanding of the locations of beaded streams, and by combining varying confidence thresholds, we can better understand the effects of our commission and omission errors on the total prediction. Further, these results demonstrate the potential of applying computer vision to high-resolution satellite imagery to map and identify surface water on a global scale at high spatial resolution.

## 6. Conclusion

Beaded streams are a common feature of permafrost-ridden landscapes, but to date they are not globally studied or extensively mapped. By relying on recent advances in computer vision and high-resolution satellite imagery, we predict the location of beaded streams with an estimated 20.2% omission rate, resulting in the first global mapping of beaded streams in the region. We characterize differences between beaded and non-beaded streams and confirm that a large majority of beaded streams are in continuous permafrost in lowland regions, but that more than previously thought are in discontinuous permafrost regions. In total we estimate that beaded streams can be found within one third of all pan-Arctic catchments (138,500 out of 417,189). Additionally, we provide further evidence towards the high prevalence of beaded streams within Russia and demonstrate the potential of computer vision for addressing gaps in hydrologic understanding of small Arctic rivers.

## CRediT authorship contribution statement

Merritt E. Harlan: Conceptualization, Methodology, Data curation, Formal analysis, Visualization, Software, Writing original draft, Writing review & editing, Funding acquisition. Colin J. Gleason: Conceptualization, Methodology, Writing review & editing, Supervision, Funding acquisition. Jonathan A. Flores: Methodology, Writing review & editing, Software. Theodore M. Langhorst: Conceptualization, Methodology, Writing review & editing. Samapriya Roy: Methodology, Writing review & editing, Software.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data and code can be accessed here: <https://doi.org/10.5281/zenodo.7223257>

## Acknowledgments

M.E. Harlan was funded by a NASA Future Investigators in Earth and Space Science Fellowship (80NSSC19K1328). C.J. Gleason's time was supported by NASA SWOT Science Team grant 80NSSC20K1141 and NSF CAREER grant 1748653. We thank our digitizing party (Elisa Friedmann, Taylor Rowley, and Nikki Tebaldi) for their time manually digitizing beaded stream outlines, as well as the Integrated Geoscience and Engineering program undergraduate students for their time and support in Alaskan beaded stream fieldwork that helped motivate this study.

## References

- Allen, G.H., Pavelaky, T.M., 2018. Global extent of rivers and streams. *Science* 361, 585–588. <https://doi.org/10.1126/science.aat0636>.
- Arp, C.D., Whitman, M.S., Jones, B.M., Grouse, G., Gaglioti, B.V., Heim, K.C., 2015. Distribution and biophysical processes of beaded streams in Arctic permafrost landscapes. *Biogeosciences* 12, 29–47. <https://doi.org/10.5194/bg-12-29-2015>.
- Arp, C.D., Whitman, M.S., Jones, B.M., Kemnitz, R., Grouse, G., Urban, P.E., 2012. Drainage network structure and hydrologic behavior of three lake-rich watersheds on the Arctic Coastal Plain, Alaska. *Arct. Antarct. Alp. Res.* 44, 385–398. <https://doi.org/10.1657/1938-4246-44.4.385>.
- Benstead, J.P., Deegan, L.A., Peterson, B.J., Huryx, A.D., Bowden, W.B., Suberkropp, K., Busby, K.M., Green, A.C., Vacca, J.A., 2005. Responses of a beaded Arctic stream to short-term N and P fertilisation. *Freshw. Biol.* 50, 277–290. <https://doi.org/10.1111/j.1365-2427.2004.01319.x>.
- Biskaborn, B.K., Smith, S.L., Noetzel, J., Matthes, H., Vieira, G., Streletskiy, D.A., Schoeneich, P., Romanovsky, V.E., Lewkowicz, A.G., Abramov, A., Allard, M., Boike, J., Cable, W.L., Christiansen, H.H., Delaloye, R., Diekmann, B., Drozdov, D., Eitzelmueller, B., Grouse, G., Ouglielmin, M., Ingeman-Nielsen, T., Isakson, K., Ishikawa, M., Johansson, M., Johannsson, H., Joo, A., Kaverin, D., Kholodov, A., Konstantinov, P., Kroger, T., Lambiel, C., Lanckman, J.-P., Luo, D., Malkova, G., Meiklejohn, I., Moskalenko, N., Oliva, M., Phillips, M., Ramos, M., Sannel, A.B.K., Sergeev, D., Seybold, C., Skryabin, P., Vasiliev, A., Wu, Q., Yoshikawa, K., Zheleznyak, M., Lantuit, H., 2019. Permafrost in warming at a global scale. *Nat. Commun.* 10, 264. <https://doi.org/10.1038/s41467-018-08240-4>.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. *Classification and Regression Trees*. Routledge, Boca Raton. <https://doi.org/10.1201/9781315139470>.
- Brosten, T.R., Bradford, J.H., McNamara, J.P., Zarnetske, J.P., Gooseff, M.N., Bowden, W.B., 2006. Profiles of temporal thaw depths beneath two arctic stream types using ground-penetrating radar. *Permafrost. Periglac. Process.* 17, 341–355. <https://doi.org/10.1002/ppp.566>.
- Cooley, S.W., Smith, L.C., Ryan, J.C., Pitzer, L.H., Pavelaky, T.M., 2019. Arctic-boreal Lake dynamics revealed using CubeSat imagery. *Geophys. Res. Lett.* 46, 2111–2120. <https://doi.org/10.1029/2018GL081584>.
- Cooley, S.W., Smith, L.C., Stepan, L., Mascaro, J., 2017a. Tracking dynamic northern surface water changes with high-frequency planet CubeSat imagery. *Remote Sens.* 9, 1306. <https://doi.org/10.3390/rs9121306>.
- Cooley, S.W., Smith, L.C., Stepan, L., Mascaro, J., 2017b. Tracking dynamic northern surface water changes with high-frequency planet CubeSat imagery. *Remote Sens.* 9, 1306. <https://doi.org/10.3390/rs9121306>.
- Craig, P.C., McCart, P.J., 1975. Classification of stream types in Beaufort Sea drainages between Prudhoe Bay, Alaska, and the Mackenzie Delta, N. W. T. Canada. *Arct. Alp. Res.* 7, 183–198. <https://doi.org/10.1080/00040851.1975.12003821>.
- Erhan, D., Szegedy, C., Toshev, A., Anguelov, D., 2014. Scalable object detection using deep neural networks. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. Presented at the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Columbus, OH, USA, pp. 2155–2162. <https://doi.org/10.1109/CVPR.2014.276>.
- Farquharson, L.M., Mann, D.H., Grouse, G., Jones, B.M., Romanovsky, V.E., 2016. Spatial distribution of thermokarst terrain in Arctic Alaska. *Geomorphology* 273, 116–133. <https://doi.org/10.1016/j.geomorph.2016.08.007>.
- Feng, D., Gleason, C.J., Yang, X., Pavelaky, T.M., 2019. Comparing discharge estimates made using the BAM algorithm in high-order Arctic Rivers derived solely from optical CubeSat, Landsat, and Sentinel-2 data. *Water Resour. Res.* 55, 7753–7771. <https://doi.org/10.1029/2019WR025599>.
- Girshick, R., 2015. Fast R-CNN. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. Presented at the 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>.
- Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. Presented at the 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587. <https://doi.org/10.1109/CVPR.2014.81>.
- He, K., Okioxari, G., Dollár, P., Girshick, R., 2017. In: *Mask R-CNN*. Presented at the 2017 IEEE International Conference on Computer Vision (ICCV), IEEE Computer Society, pp. 2980–2988. <https://doi.org/10.1109/ICCV.2017.322>.
- Heginbottom, J., Brown, J., Ferriana, O., Melnikov, E.S., 2002. Circum-Arctic Map of Permafrost and Ground-Ice Conditions, Version 2. <https://doi.org/10.7265/SK3G-KF16>.
- Heim, K.C., Wipfli, M.S., Whitman, M.S., Arp, C.D., Adams, J., Falke, J.A., 2016. Seasonal cues of Arctic grayling movement in a small Arctic stream: the importance of surface water connectivity. *Environ. Biol. Fish.* 99, 49–65. <https://doi.org/10.1007/s10641-015-0453-x>.
- Hobbie, J.E., Kling, G.W., 2014. *Alaska's Changing Arctic: Ecological Consequences for Tundra, Streams, and Lakes*. Oxford University Press.
- Hopkins, D.M., Karlstrom, T.N.V., 1955. *Permafrost and Ground Water in Alaska* 69.
- Isikdogan, F., Bovik, A., Pasmalacqua, P., 2017. RivaMap: an automated river analysis and mapping engine. *Remote Sens. Environ.* 202, 88–97. <https://doi.org/10.1016/j.rse.2017.03.044>.
- Jocher, G., Stoken, A., Chaurasia, A., Borovec, J., Kwon, Y., Michael, K., Changyu, L., Fang, J., TaoXie, Skalki, P., Hogan, A., Nadar, J., Mammanna, L., Fati, C., Montes, D., Hajek, J., Diaconu, L., Minh, M.T., Marc, FangVA, tkianai, yxNONG, yxNONG, AlexWang1900, albinoxavi, fatih, oleg, wanghaoyang0106, 2021. ultralytics/yolov5: v6.0 - YOLOv5n Nano models, Roboflow integration, TensorFlow export, OpenCV DNN support. <https://doi.org/10.5281/zenodo.5563715>.
- Kaas, A., Altena, B., Mascaro, J., 2019. River-ice and water velocities using the Planet optical cubesat constellation. *Hydro. Earth Syst. Sci.* 23, 4233–4247. <https://doi.org/10.5194/hess-23-4233-2019>.
- Kale, A., Chacsko, Z., 2015. Evolutionary feature optimization and classification for monitoring floating objects. In: Borowik, G., Chacsko, Z., Jacak, W., uba, T. (Eds.), *Computational Intelligence and Efficiency in Engineering Systems, Studies in Computational Intelligence*. Springer International Publishing, Cham, pp. 3–16. [https://doi.org/10.1007/978-3-319-15720-7\\_1](https://doi.org/10.1007/978-3-319-15720-7_1).
- Larouche, J.R., Abbott, B.W., Bowden, W.B., Jones, J.B., 2015. The role of watershed characteristics, permafrost thaw, and wildfire on dissolved organic carbon biodegradability and water chemistry in Arctic headwater streams. *Biogeosciences* 12, 4221–4233. <https://doi.org/10.5194/bg-12-4221-2015>.
- Lehner, B., Grill, G., 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydro. Process.* 27, 2171–2186. <https://doi.org/10.1002/hyp.9740>.
- Lehner, B., Verdin, K., Jarvis, A., 2008. New global hydrography derived from spaceborne elevation data. *Eos Trans. AGU* 89, 93–94. <https://doi.org/10.1029/2008EO100001>.
- Lesine, E.M.D., Kyzivat, E.D., Smith, L.C., 2021. Super-resolution surface water mapping on the Canadian shield using planet CubeSat images and a generative adversarial network. *Can. J. Remote. Sens.* 47, 261–275. <https://doi.org/10.1080/07038992.2021.1924646>.
- Lin, P., Pan, M., Wood, E.F., Yamazaki, D., Allen, G.H., 2021. A new vector-based global river network dataset accounting for variable drainage density. *Sci. Data* 8, 28. <https://doi.org/10.1038/s41597-021-00819-9>.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: Common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), *Computer Vision – ECCV 2014, Lecture Notes in Computer Science*. Springer International Publishing, Cham, pp. 740–755. [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48).
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016. SSD: Single shot MultiBox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), *Computer Vision – ECCV 2016, Lecture Notes in Computer Science*. Springer International Publishing, Cham, pp. 21–37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2).
- Lu, X., Yang, K., Bennett, M.M., Liu, C., Mao, W., Li, Y., Zhang, W., Li, M., 2021. High-resolution satellite-derived river network map reveals small Arctic river hydrography. *Environ. Res. Lett.* 16, 054015. <https://doi.org/10.1088/1748-9326/abf463>.

- Lu, X., Yang, K., Lu, Y., Gleason, G.J., Smith, L.C., Li, M., 2020. Small Arctic rivers mapped from Sentinel-2 satellite imagery and ArcticDEM. *J. Hydrol.* 584, 124689 <https://doi.org/10.1016/j.jhydrol.2020.124689>.
- McFarland, J.J., Wipfli, M.S., Whitman, M.S., 2018. Trophic pathways supporting Arctic grayling in a small stream on the Arctic coastal plain, Alaska. *Ecol. Freshw. Fish* 27, 184–197. <https://doi.org/10.1111/eff.12336>.
- Oswood, M.W., Everett, K.R., Schell, D.M., 1989. Some physical and chemical characteristics of an Arctic beaded stream. *Ecography* 12, 290–295. <https://doi.org/10.1111/j.1600-0587.1989.tb00848.x>.
- Padilla, R., Netto, S.L., da Silva, E.A.B., 2020. A survey on performance metrics for object-detection algorithms. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP). Presented at the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 237–242. <https://doi.org/10.1109/IWSSIP48289.2020.9145130>.
- Pewe, T.L., 1966. Ice-wedges in Alaska – classification, distribution, and climatic significance. In: Presented at the Permafrost International Conference, Lafayette, Indiana, pp. 76–81.
- Planet Team, 2022. PGScene4Band [WWW Document] accessed 11.14.22. <https://developers.planet.com/docs/data/pgscene4band/>.
- Planet Team, 2021. Satellite Imagery and Archive [WWW Document]. Planet. URL (accessed 5.8.22). <https://planet.com/products/planet-imagery/>.
- Qayyum, N., Ghuffar, S., Ahmad, H.M., Yousof, A., Shahid, I., 2020. Glacial Lakes mapping using multi satellite PlanetScope imagery and deep learning. *ISPRS Int. J. Geo-Inf.* 9, 560. <https://doi.org/10.3390/ijgi9100560>.
- Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, NV, USA, pp. 779–788. <https://doi.org/10.1109/CVPR.2016.91>.
- Redmon, J., Farhadi, A., 2018. YOLOv3: An Incremental Improvement. *ArXiv180402767* Co (accessed 5.8.22).
- Redmon, J., Farhadi, A., 2016. YOLO9000: Better, Faster, Stronger. *ArXiv161208242* Co.
- Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. In: *Advances in Neural Information Processing Systems*. Curran Associates Inc.
- Srivastava, S., Divekar, A.V., Anilkumar, C., Naik, I., Kulkarni, V., Pattabiraman, V., 2021. Comparative analysis of deep learning image detection algorithms. *J. Big Data* 8, 66. <https://doi.org/10.1186/s40537-021-00434-w>.
- Strick, R.J.P., Ashworth, P.J., Sambrook Smith, G.H., Nicholas, A.P., Best, J.L., Lane, S. N., Parsons, D.R., Simpson, C.J., Unsworth, C.A., Dale, J., 2019. Quantification of bedform dynamics and bedload sediment flux in sandy braided rivers from airborne and satellite imagery. *Earth Surf. Process. Landf.* 44, 953–972. <https://doi.org/10.1002/esp.4558>.
- Tan, L., Huangfu, T., Wu, L., Chen, W., 2022. Comparison of YOLO v3, Faster R-CNN, and SSD for Real-Time Pill Identification. <https://doi.org/10.21203/rs.3.rs-668895/v1>.
- Tarbeeve, A.M., Surkov, V.V., 2013. Beaded channels of small rivers in permafrost zones. *Geogr. Nat. Resour.* 34, 216–220. <https://doi.org/10.1134/S1875372813030049>.
- Tarboton, D.G., 1997. A new method for the determination of flow directions and uplope areas in grid digital elevation models. *Water Resour. Res.* 33, 309–319. <https://doi.org/10.1029/96WR03137>.
- Trochim, E.D., Jorgenson, M.T., Prakash, A., Kane, D.L., 2016. Geomorphic and biophysical factors affecting water tracks in northern Alaska. *Earth Space Sci.* 3, 123–141. <https://doi.org/10.1002/2015EA000111>.
- Wada, K., 2022. Labelme: Image Polygonal Annotation with Python. <https://doi.org/10.5281/zenodo.5711226>.
- Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y.M., 2021. Scaled-YOLOv4: Scaling Cross Stage Partial Network. *ArXiv201108036* Co.
- Wollheim, W.M., Peterson, B.J., Deegan, L.A., Hobbie, J.E., Hooker, B., Bowden, W.B., Edwardson, K.J., Arcott, D.B., Herzhey, A.E., Finlay, J., 2001. Influence of stream size on ammonium and suspended particulate nitrogen processing. *Limnol. Oceanogr.* 46, 1–13. <https://doi.org/10.4319/lo.2001.46.1.0001>.
- Yamazaki, D., Ikehima, D., Gosa, J., Bates, P.D., Allen, G.H., Pavelaky, T.M., 2019. MERIT hydro: a high-resolution global hydrography map based on latest topography dataset. *Water Resour. Res.* 55, 5063–5073. <https://doi.org/10.1029/2019WR024873>.
- Yamazaki, D., Ikehima, D., Tawatari, R., Yamaguchi, T., O Loughlin, P., Neal, J.C., Sampson, C.C., Kanai, S., Bates, P.D., 2017. A high-accuracy map of global terrain elevations. *Geophys. Res. Lett.* 44, 5844–5853. <https://doi.org/10.1002/2017GL072874>.
- Yang, K., Li, M., Liu, Y., Cheng, L., Duan, Y., Zhou, M., 2014. River delineation from remotely sensed imagery using a multi-scale classification approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 7, 4726–4737. <https://doi.org/10.1109/JSTARS.2014.2309707>.
- Yang, X., Pavelaky, T.M., Allen, G.H., Donchyts, G., 2020. RivWidthCloud: an automated Google Earth Engine algorithm for river width extraction from remotely sensed imagery. *IEEE Geosci. Remote Sens. Lett.* 17, 217–221. <https://doi.org/10.1109/LGRS.2019.2920225>.
- Zametke, J.P., Gooseff, M.N., Bowden, W.B., Greenwald, M.J., Brosten, T.R., Bradford, J.H., McNamara, J.P., 2008. Influence of morphology and permafrost dynamics on hyporheic exchange in Arctic headwater streams under warming climate conditions. *Geophys. Res. Lett.* 35 <https://doi.org/10.1029/2008GL033504>.
- Zametke, J.P., Gooseff, M.N., Brosten, T.R., Bradford, J.H., McNamara, J.P., Bowden, W.B., 2007. Transient storage as a function of geomorphology, discharge, and permafrost active layer conditions in Arctic tundra streams. *Water Resour. Res.* 43 <https://doi.org/10.1029/2005WR004816>.
- Zhang, D., Lee, D.-J., Zhang, M., Tippetts, B.J., Lillywhite, K.D., 2016. Object recognition algorithm for the automatic identification and removal of invasive fish. *Biosyst. Eng.* 145, 65–75. <https://doi.org/10.1016/j.biosystemseng.2016.02.013>.
- Zhao, Z.-Q., Zheng, P., Xu, S.-T., Wu, X., 2019. Object detection with deep learning: a review. *IEEE Trans. Neural Netw. Learn. Syst.* 30, 3212–3232. <https://doi.org/10.1109/TNNLS.2018.2876865>.