Data Collection Infrastructure for Feature Similarity Analysis in Satellite Images of Artic Sea Ice

C.J. Moore, Gavin Moore, Dylan Berndt   Advisor: Xin Miao

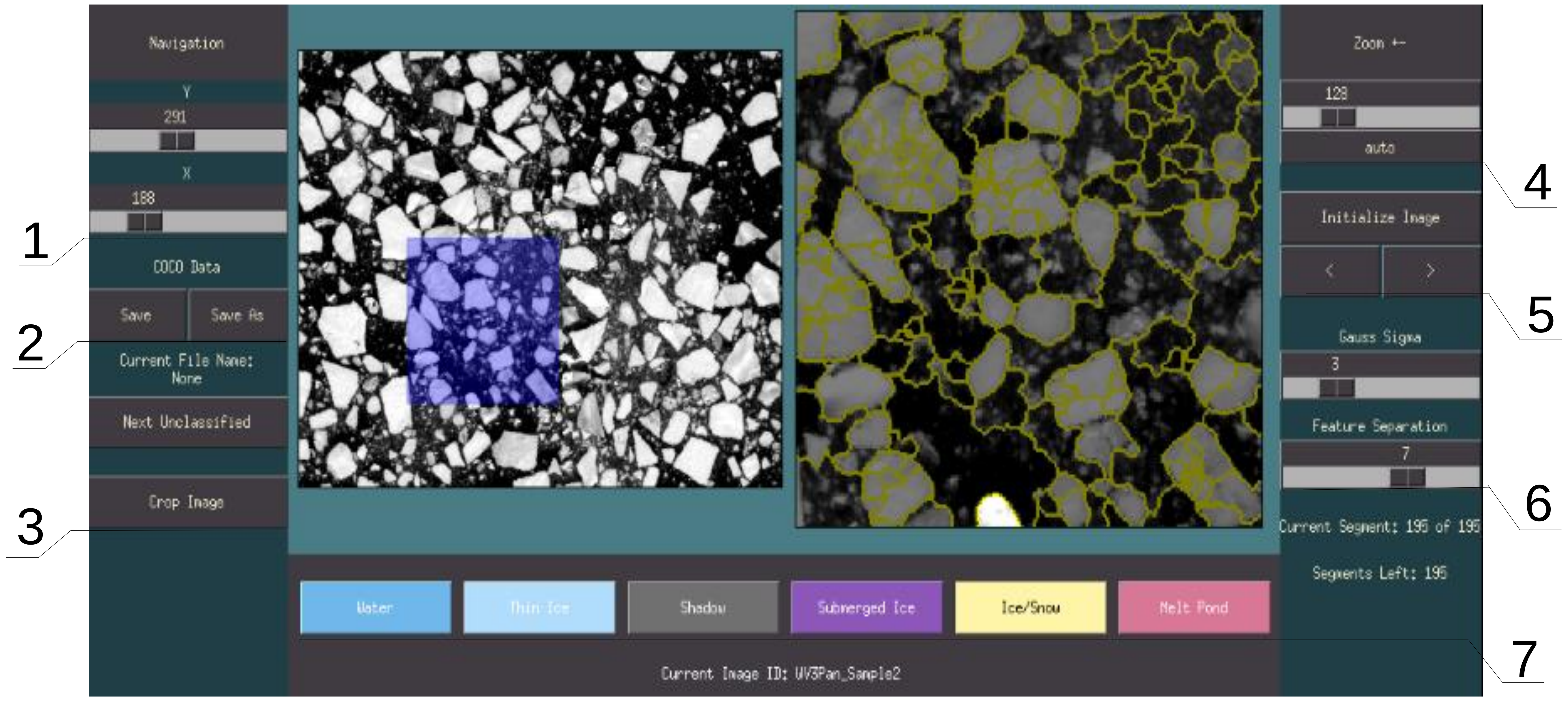
Sponsored by NSF Collaborative Research: Elements: Data: HDR: Developing On-Demand Service Module for Mining Geophysical Properties of Sea Ice from High Spatial Resolution Imagery, $127,966.00 (Jan 1, 2019 - Dec 31, 2021)

Abstract

Remote detection of sea ice features present in Arctic regions can be highly informative to design and verify global climate models. Satellite imagery can provide instrumental data in uncovering spatial patterns of sea ice by way of novel deep learning (DL) models that are trained to perceive feature similarities from labeled images. Training a DL model to classify previously unencountered sea-ice images requires a comprehensive dataset of labeled sea-ice samples. Creating an effective and accurate dataset requires a high degree of domain knowledge and carries major time costs even for multiple parties. To solve this problem, we develop a suite of open-source data collection tools to allow researchers to efficiently create a specialized training dataset. Using common Python libraries including Matplotlib and Tkinter we develop a platform that employs an automated segmentation algorithm alongside segmentation labeling, image navigation, and data management functionalities to streamline the process of training data collection. We find that moderate degrees of accuracy can be achieved by segmentation-based DL classification models trained on small amounts of data created, and we provide an easily adaptable framework to develop, analyze, augment, and consolidate similar datasets for other computer-vision tasks within and beyond geospatial sciences.

Introduction

Conventional methods of collecting data for sea ice classification include image annotation software that almost exclusively limit the user to drawing manual annotations for present features. This is highly time consuming, and the high degree of spatial feature variance within many sea ice images makes this task almost impossible using traditional means. The problem of an unrealistic time-cost for the expansive dataset necessary to create an accurate sea ice classifier is best solved by eliminating various time intensive tasks and consolidating functionalities needed for dataset creation into one system. This system should have a rigid process to ensure uniformity, and it should include data management capabilities to allow for changes to be made, and for datasets to be combined. As part of a multi-institutional effort funded by the NSF [1], we introduce the Arc-CI training GUI. The Arc-CI training GUI cuts a tremendous amount of time from the annotation process by automatically drawing segmentation lines around features within the target image. This combined with many other functionalities provides a set of tools that aim to make the process of training data creation significantly quicker and easier.



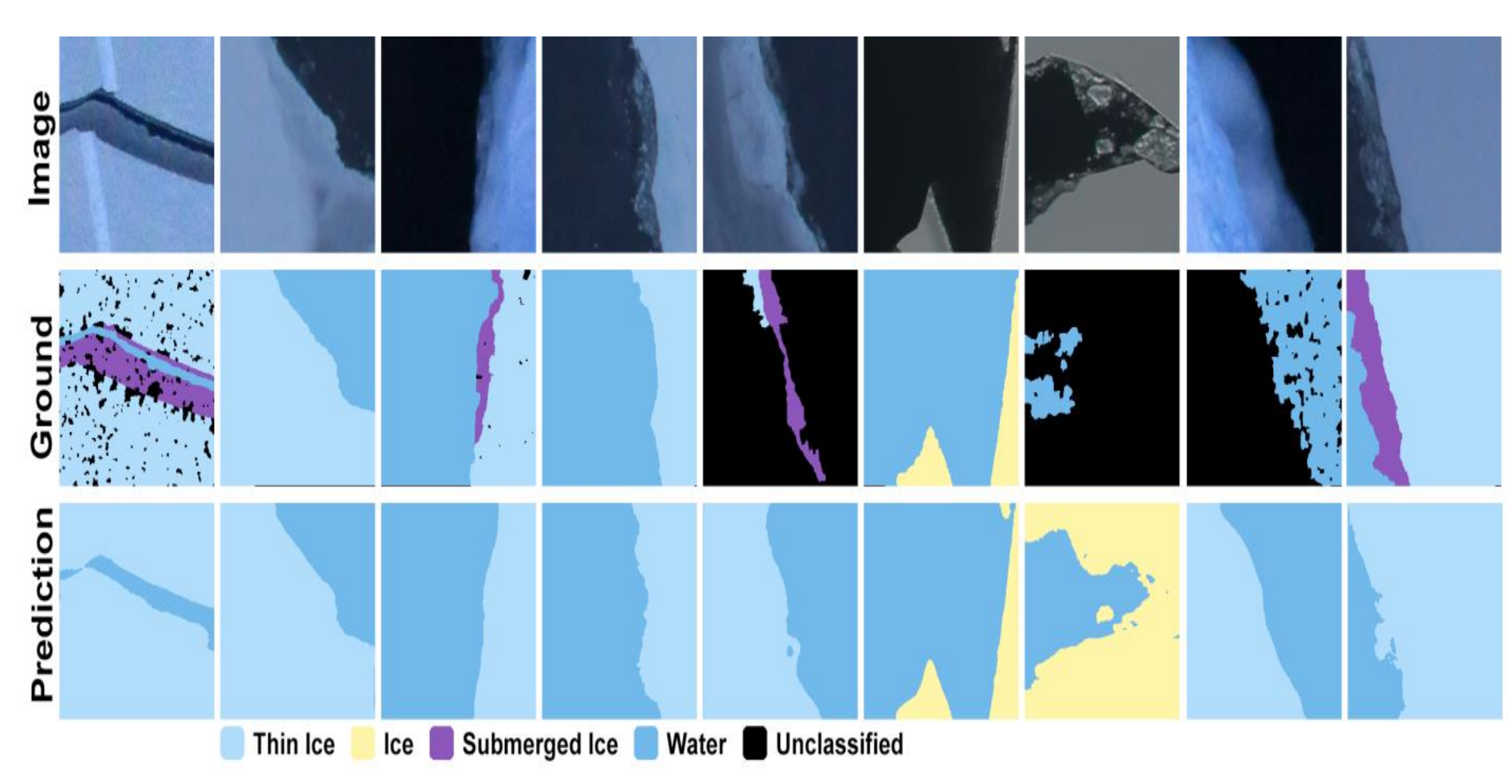
The Arc-CI Training GUI: 1) Navigation Sliders 2) Dataset Management 3) Crop Image Button 4) Zoom Slider 5) Image Upload Buttons 6) Segmentation Algorithm Parameters 7) Image Classification Buttons

Methods

Our automated segmentation algorithm starts by accentuating the edges of a feature within an image: providing the groundwork for where it draws its segment lines[2][3] . The algorithm draws segments around features present in the image being analyzed by the program, and the user can manipulate these segments using parameter sliders present in the training GUI's user interface. The user will attempt to find the most effective combination of parameters for each individual image. The best heuristic for determining the ideal set of parameters is: "does it look accurate?" requiring no understanding of the algorithm for effective utilization. After the user has tuned the parameters to display an accurate segmentation map, they click on a segment and press the classification buttons on the bottom of the GUI to pass a classification label to that segment. When the user saves their data, a copy of the labeled image including feature masks will be returned, and corresponding labeled segment data will be appended to the working dataset. The data is stored in .JSON format using COCO naming conventions popular with many large training datasets. Data integrity can then be analyzed using the viewCOCO module included with this software, and its effectiveness proven with training on various DL models.

Results

We find that with a small dataset (<100 images) we can illicit feature similarity learning within multiple DL frameworks, suggesting that data from our program is functional and effective at providing the necessary statistical insights for feature extraction. Below we show results of a small dataset of around 30 images where a custom fully convolutional network (FCN) is clearly capable of delineating between features, suggesting that our data is effective in a training scenario



This figure shows images of artic sea ice in comparison with ground truth data and actual model predictions.

Conclusions

The functional nature of data collected with our system coupled with the efficient pace in which it can be created is promising due to the wide range of applications within and beyond geospatial analysis. Many computer vision tasks require that large, crowded, and complex images be annotated en masse by informed users. The reduction in time cost as well as the precise machine drawn annotations our program allows for can greatly enhance the training data collection process in many different projects using similar means. The emergence of accelerated throughput of accurate training data is also promising considering new DL frameworks that specialize in semantic segmentation with higher degrees of prediction accuracy than before possible

References

[1] Dexuan Sha, Xin Miao, Mengchao Xu, Chaowei Yang, Hongjie Xie, Alberto M. Mestas-Nuñez, Yun Li, Qian Liu and Jingchao Yang, "An On-Demand Service for Managing and AnalyzingArctic Sea Ice High Spatial Resolution Imagery", *Data*, 5:39, 2020. doi:10.3390/

[2] Xin Miao, Hongjie Xie, Stephen F. Ackley, Songfeng Zheng, "Object-Based Arctic Sea Ice Ridge Detection From High-Spatial-Resolution Imagery", IEEE Geoscience and Remote Sensing Letters, 13(6): 787-791, 2016.

[3] Xin Miao, Hongjie Xie, Stephen F. Ackley, Donald K. Perovich, and Changqing Ke, "Object-Based Detection of Arctic Sea Ice and Melt Ponds Using High Resolution Aerial Photographs", Cold Regions Science and Technology, 119: 211-222, 2015.