Self-supervised learning for timeseries from multi-spectral satellite imagery

Chris Bethune
Uncharted Software
cbethune@uncharted.software

Philippe Horne
Uncharted Software
phorne@uncharted.software

Zachary Hills
Uncharted Software
zhills@uncharted.software

Steve Kramer
Kung Fu AI
steve.kramer@kungfu.ai

Jeffrey Gleason
Kung Fu AI
jeff.gleason@kungfu.ai

Ben Johnson
jataware.com
ben@jataware.com

Ezekial Barnett
jataware.com
zeek@jataware.com

Richard Brath
Uncharted Software
0000-0001-6006-2092

Scott Langevin
Uncharted Software
0000-0002-3309-8935

***Abstract*— Distil is a system for constructing point-and-click machine learning models, here extended for multi-spectral satellite imagery for timeseries data leveraging an autoML pipeline, adding embedding model trained using self-supervised learning; rapid data labeling facilitated with image query; hierarchical geospatial timeseries modeling; and sub-image feature extraction using weakly-supervised segmentation.**

Keywords*— *geospatial timeseries, multi-spectral imagery

I. Introduction

Analysts working with geo-spatial data confront many new and different formats of spatial data. The analytic task may require the data to be labelled, i.e. appropriately tagged, so that a viewer can find and analyze the appropriate features of interest. Standard machine learning based approaches typically rely on labeled data to use as training data – but with new datasets and/or new ways to use datasets, those labels may not exist (i.e. there may be noground truth labels). With spatial data this is further complicated by massive data (terabytes); multispectral data (beyond RGB); noisy data (e.g. clouds, bad data); incomplete coverage; resolution, frequency and projection issues; and the combination of multiple data sources.

We address the above with our system Distil. Our primary contribution in this paper is self-supervised learning on new geospatial datasets for building predictive models from the vast amount of readily available multispectral satellite imagery to identify conditions of interest. We leverage past approaches in self-supervised learning, multi-datatype auto-ML pipeline; imagery analysis using convolutional neural networks on multispectral satellite imagery; and extensions to hierarchical time-series analysis and sub-image feature extraction.

This work was supported, in part, by the Defense Advanced Research Projects Agency (DARPA) (contract number D3M (FA8750-17-C-0094)).

II. Background

A. AutoML pipeline

Auto-machine learning (AutoML) makes machine learning accessible to domain experts with machine-guided, point-and-click data discovery, iterative model definition and analysis [1]. Issues with AutoML systems often include initial data wrangling challenges and require visual exploration of the data relative to the analytic objective [2]. One approach is to combine the ML pipeline operations with a corresponding generated visualization-oriented interface to interactively curate, define and refine machine inferences and resulting models e.g. [3].

B. Computer vision and geo-spatial imagery

State of the art approaches in computer vision use embedding models to create feature vectors from source images, which are then used in downstream machine learning tasks. The embedding models are generated by pre-training on large, labelled image datasets[4]. However, creating these models are challenging for geo-spatial imagery, as publicly labeled datasets may be constrained to limited channels (such as visible light) or geographic areas (e.g. Europe only). Thus, they are not generalizable to other regions or images with different channels.

C. Self-supervised learning

To overcome a lack of embedding models for multi-spectral imagery, our organization had prior success with self-supervised learning which constructs a model using a "pre-text" task that relies only on unlabeled data [5]. These pre-trained models act as a "backbone" for a number of downstream tasks such as image retrieval or to construct classifiers.

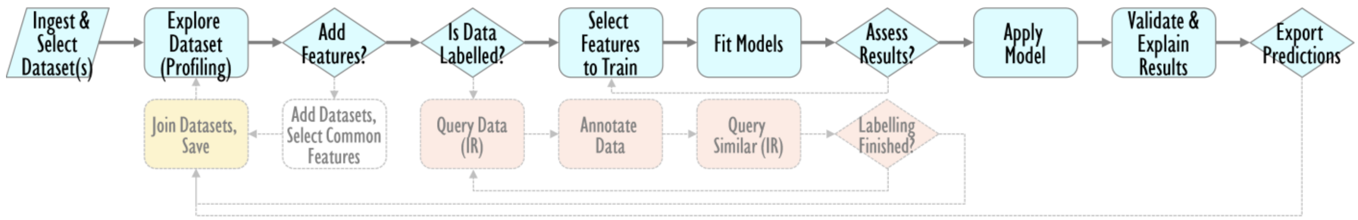

Figure 1: Distil's step-by-step model-building workflow: primary workflow on top in blue, with optional steps and iterations below.

D. *Image Query & Rapid Data Labeling*

We have prior expertise creating applications where we want to classify data, but do not have labeled training data. As users' time is valuable, we want to leverage their expertise but minimize effort. We created interfaces for experts to find and label a small set of positive examples; then use information retrieval techniques to find similar examples. Users label more images as needed, then create a machine learning classifier to label data based on those examples. After a few iterations, we have been able to approach benchmarks associated with well-labeled datasets [6,7]. This rapidly labeled data can then be used to construct predictive machine learning models.

E. *Spatio-Temporal Visual Analytics*

Many approaches to spatio-temporal visual analytics have focused on either time (e.g. variants of timelines [8]) or space (e.g. variants of maps [9]); or multiple panels (e.g. primary map, side panel with a timeline). Thus the approaches tend to favor one or the other representation.

III. SYSTEM

Our solution, Distil [10], uses a mixed initiative approach to maximize the value of the domain expert in guiding the machine learning processes. Many existing data science tools assume users are: a) familiar with the data; b) familiar with statistical analysis, features and models; c) able to prepare their data and select appropriate features; and d) able to drive the software to the goal. These tools may not provide workflow to support the analytic goal, limited assistance with model development, limited support for iterative analysis, and limited ability to introspect the features most utilized by the model.

Distil's design goals include: a) easy step-by-step workflow of the model-building process with consistent affordances; b) perform data and model analysis as background tasks and surface with the same affordances; c) reduce cognitive load with consistent simple visual representations; and d) build trust with introspection of the resulting model.

A. *Workflow Overview*

Fig. 1 outlines the primary workflow left to right across the top line in blue: After *ingesting data*, the domain expert can interactively preview, profile and *explore data*, and optionally enrich the data by *adding features* or *labels*, by diverting into additional tasks shown on the second line. Proceeding with model-building, the domain expert *selects features*, and *fits the model*, thereafter allowing the user to *assess results*, including iterating on training the model and comparing to previous models. The trained model can then be *applied* to the full dataset, and *explained* with model metrics (e.g. accuracy and recall) or ML feature explanation. Lastly, results can be *exported* as new data, including use in composing new multi-models by iterating through the workflow.

B. *System Architecture*

Distil consists of: a) an Auto ML Server, which leverages a common data process and model primitives composable into an ML pipeline to generate, validate and execute models; b) a Dataset/Model repository, which provides storage and indexing to support user-driven queries at interactive speed; and c) an Application Server, including a front-end to provide user-driven workflows and delegate compute intensive tasks to the Auto ML Server and Dataset/Model repository.

C. *Modeling Overview*

As per Fig. 1, Distil provides a general purpose, human-in-the-loop workflow for solving a range of machine learning problems, and supports datasets such as tabular data, timeseries or images. For spatial imagery, we have extended Distil to support multi-spectral image classification, such as satellite imagery and associated spatial datasets such as rainfall and temperature (which can be incorporated in the "Query data" and "Select features" steps).

To support the data exploration step

beyond profiling and information retrieval, we cluster images (e.g. using K-means for HDBSCAN) and flag outlier images using Isolation Forest. Models are created using our AutoML service. For spatial data, the pipeline relies on a transfer learning approach, in which feature vectors are first generated from input images and then shallow learners (e.g. SVM, Random Forest, XGBoost) are applied to the outputs.

D. *Spatial Data Overview*

When working with multi-spectral satellite imagery, data challenges include: a) scale of data is massive; b) heterogeneous data, as it varies in spatial and temporal resolution and available channels; c) data is inherently noisy (e.g. clouds, atmosphere); and d) there are very few labeled training sets.

Our work leverages Sentinel-2 [11] data, which produces images with 12 separate channels, 3 of which are standard visible wavelengths (RGB). We wanted to use all available channels, as those beyond RGB provide important indicators (e.g. moisture or vegetation) for model training and problem solving. Reusing standard pre-trained RGB models would sacrifice the value in these additional channels. To address this, we use a modified CNN and a self-supervised method [12] to pre-train an embedding model that extends feature vectors across all image channel values. We tested the quality of the produced vectors against a land-use classification task and achieved 10x label efficiency compared to a fully supervised CNN using only RGB image weights.

For *scalability* and performance within Distil, we apply this model to an input dataset as part of a one-time ingest step. This creates image vectors that can be stored and used for downstream modeling, with the added benefit of compressing the input samples, facilitating more efficient use of system memory during model building tasks. When training a land-use classification model, Distil will fit a lightweight learner, such as an SVM to the produced image vectors. Image retrieval is performed based on distance metrics within the vector space.

The previously mentioned classification task requires labeled data, which is often not available in remote-sensing. To create *labels* for a data subset to train the model, we provide a human-in-the-loop workflow (Fig. 1, "Is data labelled" subtask on the lower line). Users interactively find and tag examples of positive and negative images, which are used as a baseline query. Then the system ranks all images based on similarity (weighted average search). This list can then be used to group and label larger sets of images. This approach is iterated upon until sufficient samples are labelled. Distil's AutoML will automatically generate models based on this training data to label the data. Once labeled, models can be built, as indicated in the "Fit Models" step. Depending on the target, models could be classifiers, or timeseries or segmentation (next sections).

E. *Hierarchical Geospatial Timeseries Analysis*

In previous work we developed a hierarchical time-series forecasting approach using a regularized embedding space [13] to make predictions at different levels of aggregation simultaneously by leveraging the inherent hierarchical relationships to produce better forecasts. Applying this technique to geo-spatial data is a natural application for making predictions at different spatial and temporal resolutions. This approach can overcome gaps or noise present at different granularity [14][15].

A classifier trained using the method described in section *D. Spatial Data Overview* and applied to multi-spectral satellite imagery to extract features over time can yield a multivariate timeseries for each area, thereby converting the data into a hierarchical timeseries analysis (Fig. 2). Distil has workflows to enable analysts to perform these operations to conduct complex multi-variate timeseries analysis by also combining diverse multi-variate geo-temporal datasets.

Figure 2: Geo-spatial hierarchical time-series analysis.

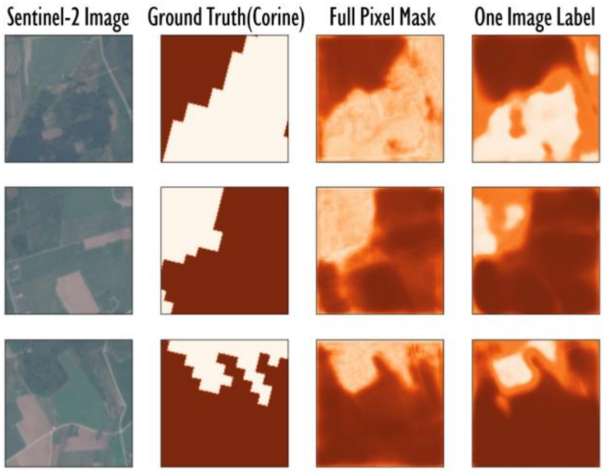

Figure 3: Results of sub-image segmentation using a self-supervised feature extractor for agriculture from 10 provided training examples (every image pixel labeled vs. one label per image)

F. Sub-image feature extraction using weakly-supervised segmentation

Extraction of image-level features such as the presence of farmland can be extended to sub-image features using image segmentation to group pixels in an image for finer grained analysis. Our work on weakly supervised segmentation investigated the performance of segmentation models that had access to a small number of image-level labels rather than each individual pixel as training data. This paradigm aligns with our user-driven labelling approach, as the same process that generates labels for image-level classification can be applied to generating labels for the weakly-supervised segmentation task. We explored a variety of different methods [16][17] to segment images based on pixel-level and image level labels in a weakly-supervised context. We found we could retain the performance of the segmentation model when substituting our self-supervised embedding model for a fully-supervised encoder, eliminating the need for a pre-training model on large amounts of labeled data which fits our geo-spatial modeling needs (Fig. 3).

IV. Example

A pilot with an agency investigated assessing insect infestation of crop land via remote sensing. Using Distil, we can: a) run the pipeline to classify imagery for indicators of insect infestation, using local ground truth reports of infestations to label the initial training data; b) run the pipeline with iterative user labeling and sub-feature extraction to identify important crop land; and then c) run the timeseries model across sequences of output from the prior models to create timeseries of propensity for potential infestation across a region down to plots of land. These timeseries can then be used to predictively assess fields at risk, e.g., differentiating fields with a high propensity still trending upwards, versus fields with a high propensity but trending downwards.

V. Conclusion

Distil is a work in progress with promising results for domain experts to create timeseries from unlabelled multi-spectral satellite imagery fused with other geospatial datasets. Labelling effort is reduced and output quality is significantly improved over naïve labelling approaches.

References

[1] Endert, A., Hossain, M.S., Ramakrishnan, N., North, C., Fiaux, P. and Andrews, C., 2014. The human is the loop: new directions for visual analytics. *Journal of intelligent information systems*, *43*(3), pp.411-435.

[2] Kandel, S., Paepcke, A., Hellerstein, J.M. and Heer, J., 2012. Enterprise data analysis and visualization: An interview study. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), pp.2917-2926.

[3] Sumbul, G., Charfuelan, M., Demir, B. and Markl, V., 2019, July. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium* (pp. 5901-5904). IEEE.

[4] Deng, J., Dong, W., Socher, R., Li, L., Li, K., and FeiFei, L., 2009 Imagenet: A large-scale hierarchical image database, *IEEE Conference on Computer Vision and Pattern Recognition* (2009).

[5] Langevin, S., Bethune, C., Horne, P., Kramer, S., Gleason, J., Johnson, B., Barnett, E., Husain, F. and Bradley, A., 2021. Useable machine learning for Sentinel-2 multispectral satellite imagery. In *Image & Signal Processing for Remote Sensing XXVII* (Vol. 11862, pp. 97-114). SPIE.

[6] Li, X. and Liu, B., 2003, August. Learning to classify texts using positive and unlabeled data. In *IJCAI* (Vol. 3, No. 2003, pp. 587-592).

[7] Hagerman, C., Brath, R. and Langevin, S., 2019. Visual Analytic System for Subject Matter Expert Document Tagging using Information Retrieval and Semi-Supervised Machine Learning. In *2019 23rd International Conference Information Visualisation (IV)* (pp. 234-240). IEEE.

[8] Xu, P., Wu, Y., Wei, E., Peng, T.-Q., Liu, S., Zhu, J. J. and Qu, H. 2013. Visual analysis of topic competition on social media. *IEEE Transactions on Visualization and Computer Graphics* (TVCG), 19(12):2012–2021.

[9] Andrienko, G., Andrienko, N., Bak, P., Kisilevich, S. and Keim, D. 2009. Analysis of community-contributed space- and time-referenced data. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST),* pp. 213–214.

[10] Langevin, S., Jonker, D., Bethune, C., Coppersmith, G., Hilland, C., Morgan, J., Azunre, P. and Gawrilow, J., 2018. Distil: A Mixed-Initiative Model Discovery System for Subject Matter Experts. In *International Conference on Machine Learning AutoML Workshop*.

[11] (ESA) https://sentinel.esa.int/web/sentinel/missions/sentinel-2

[12] Bachman, P., Hjelm, R. D., & Buchwalter, W. 2019. Learning representations by maximizing mutual information across views. *Advances in neural information processing systems*, *32*.

[13] Gleason, J. L. "Forecasting Hierarchical Timeseries with a Regularized Embedding Space," 6th KDD Workshop on Mining and Learning from Timeseries, San Diego, 2020.

[14] Rangapuram, M. W. Seeger, J. Gasthaus, L. Stella, Y. Wang, and T. Januschowski, Deep State Space Models for Timeseries Forecasting, 2018. *NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems*.

[15] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, N-BEATS: Neural basis expansion analysis for interpretable timeseries forecasting, *ICLR* 2020.

[16] Wang, W. Chen, S. M. Xie, G. Azzari, and D. B. Lobell, Weakly Supervised Deep Learning for Segmentation of Remote Sensing Imagery, *Remote Sensing*, vol. 12, no. 2, Art. no. 2, Jan. 2020, doi: 10.3390/rs12020207.

[17] P. Ulmas and I. Liiv, Segmentation of Satellite Imagery using U-Net Models for Land Cover Classification, *arXiv:2003.02899 [cs]*, Mar. 2020, Accessed: Apr. 12, 2021. http://arxiv.org/abs/2003.02899.

CANADA