Detecting Abandoned Oil And Gas Wells Using Machine Learning And Semantic Segmentation

Michelle Lin
School of Computer Science
McGill University
Montreal, QC
michelle.lin2@mail.mcgill.ca

David Rolnick
School of Computer Science
McGill University
Montreal, QC
drolnick@cs.mcgill.ca

Abstract

Around the world, there are millions of unplugged abandoned oil and gas wells, leaking methane into the atmosphere. The locations of many of these wells, as well as their greenhouse gas emissions impacts, are unknown. Machine learning methods in computer vision and remote sensing, such as semantic segmentation, have made it possible to quickly analyze large amounts of satellite imagery to detect salient information. This project aims to automatically identify undocumented oil and gas wells in the province of Alberta, Canada to aid in documentation, estimation of emissions and maintenance of high-emitting wells.

1 Problem and Motivation

Around the world, millions of abandoned oil and gas wells exist in a kind of limbo, often the creation of companies that are now defunct. Decades later, such wells continue to be a major environmental hazard - by contaminating surrounding ecosystems, the groundwater used by the communities around them and contributing greenhouse gases to the atmosphere equivalent to millions of tons of carbon dioxide every year [12].

The number of abandoned wells continues to grow each year. There are approximately 400,000 abandoned wells in Canada alone – with the estimate being ten times higher in the United States [12]. While databases exist for the locations of some abandoned wells, the locations of the majority of such wells remains unknown. For example, the number of wells recorded by the Pennsylvania Department of Environmental Protection is only about a tenth of the total number of wells estimated to exist in the state [5]. An understanding of their environmental impacts is similarly incomplete, with these undocumented wells described as the most uncertain source of methane emissions in Canada [12].

This project aims to leverage machine learning to (1) identify the existence and locations of previously undocumented oil and gas wells in Alberta, and (2) precisely localize and correct inaccurate locations of known abandoned oil and gas wells. The geospatial information we obtain will aid experts in efforts to monitor, assess, and plug such wells, *plugging* being the process wherein a well bore is fitted with a cement plug to prevent contamination and further methane leakage. In future work, we aim to automatically monitor and create more precise methane inventories from abandoned oil and gas wells and accelerate the identification process of especially high methane-emitting wells.

2 Background and Related Work

Semantic segmentation is a fundamental and well-established task in computer vision. This pixel-wise classification technique has been used in a variety of data-abundant remote sensing problems, including tasks using multi-band hyperspectral satellite imagery, such as tree and vegetation classification

Submitted to Tackling Climate Change with Machine Learning: workshop at NeurIPS 2021.

Figure 1: Ground level images of abandoned oil and gas wells from Alberta's Site Rehabilitation Program [3]

[11], crop cover and type analysis [14] and environmental monitoring [1]. In addition, segmentation techniques have been used in geolocalization tasks such as improving localization and mapping on slums and small-scale urban structures [13].

The U-Net [8] is a fully convolutional neural network (FCN) with a symmetric encoder-decoder architecture. This particular architecture contains an expanding decoder path to enable precise localization and recovery of object details [8, 10]. Originally developed for medical image segmentation, the U-Net has been used in a variety of other problems, such as road extraction [15] and greenhouse detection [4], thanks to its success at performing image segmentation with minimal training data.

However, limited work has been done to semantically identify and localize oil and gas infrastructure wells. To date, such efforts have been purely applied on the detection of *active* oil and gas wells with large spatial features, including large identifiable machinery and infrastructure that span up to kilometers, using low to medium resolution satellite imagery [6, 9]. Conversely, abandoned oil and gas wells are only a few meters large at most, requiring high resolution satellite imagery for detection.

3 Proposed Approach

Our machine learning methodology is intended to localize abandoned wells from satellite imagery. Specifically, we will use U-Net-inspired neural networks, which allow fully convolutional implementations that can rapidly process large areas in parallel [8]. These methods will be trained using partial data on over 200,000 well locations available from the AER-ST37 database provided by the Alberta Energy Regulator and high-resolution, multi-band, geospatial Skysat satellite imagery from Planet Labs, on the scale of 0.5m per pixel, to detect features of a small spatial size. Training images will consist of satellite imagery around each datapoint representing a well (from the AER-ST37 dataset). Our neural networks will be trained to output binary masks with each pixel labelled as belonging to the *well* class or *not well* class – with every pixel within a fixed radius of a well's point location labeled as "well" and every pixel outside labeled as "not well" (see Figure 2).

The immediate output of the classifier will be a prediction mask of probabilities. From this, the relevant information (locations of the predicted wells) can be distilled in one of two ways: (1) by clustering pixels classified as "well" and outputting as a well any cluster that exceeds a given number of pixels, or (2) for every location, summing the probabilities of "well' at neighboring pixels within a given radius and outputting as a well any location that exceeds a given threshold.

A methodological issue we anticipate is in the dataset containing imbalanced data – many more negative examples than positive ones, since the majority of pixels in satellite imagery are clearly not abandoned wells. To mitigate this, we plan on enforcing a relatively balanced training dataset, then optimally selecting thresholds to compensate for imbalanced data during test time.

Our methods will not only be used to detect previously unknown abandoned well locations, but to give much more accurate locations for known wells that are already present in the database; these locations are currently known only very imprecisely, with errors of up to kilometers, thereby complicating on-the-ground assessments, monitoring, maintenance, and plugging of such wells. We anticipate that our algorithm will pinpoint some currently active wells along with abandoned wells, due to visual similarity; these can be filtered out in post-processing, given that information on active wells is more complete than that for abandoned wells.

It is worth noting that since some of the locations given in the database for abandoned wells are incorrect, some of the labels given to the neural network will be inaccurate to varying degrees. However, we anticipate that the neural network will able to ignore certain amounts of label noise (see, e.g., [7]), and that there is a sufficient amount of fully accurate labels for effective training.

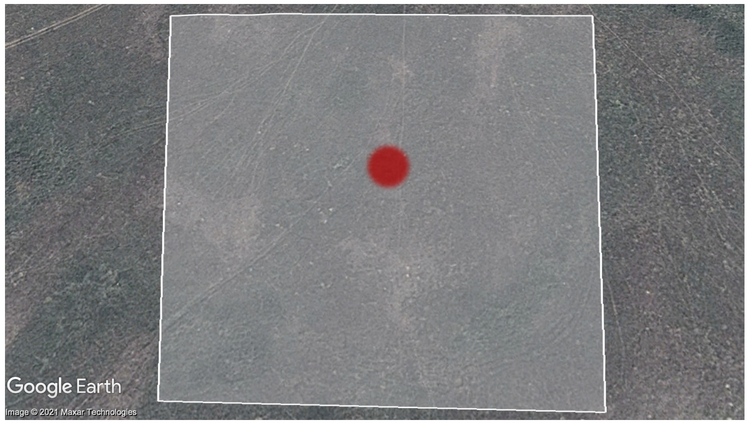

Figure 2: An example aerial image of a labelled well ("Well" pixels) shown in red. The area inside the white polygon ("Not Well" pixels) includes negative examples for the classifier.

4 Future Work

In a future stage of this project, the focus will be on the ability of the algorithm to generalize to various geographical regions. While this model will be developed using data from Alberta, there are undocumented wells in other Canadian provinces and countries, notably in the United States and many countries of the former Soviet Union. We will use meta-learning methodologies such as model-agnostic meta-learning (MAML) [2] to generalize effectively between multiple regions with minimal additional data.

Another future stage of this project moves beyond localization of abandoned oil and gas wells to quantification of methane leakage. Exact estimates of methane leakage from wells is an exceptionally hard problem without specialized measurements, such as hyperspectral imagery (which is difficult in the case of abandoned wells since individual wells tend to yield diffused plumes). Active learning techniques will be used to sending expert ground-truth teams to specific wells to determine methane concentration levels. While such measurements are time-intensive, far fewer field measurements will need to be taken thanks to this technique.

We collaborate closely with a team of civil engineers in the Subsurface Hydrology and Geochemistry Research Group at McGill University, who specialize in assessing methane emissions from abandoned oil and gas wells and have monitored such wells extensively across the globe. The process of automatic well identification, and future steps for methane emissions quantification, will be used by these collaborators to better understand these wells, their impacts, the creation of more complete database records – in addition to maintaining, plugging, and conducting field assessments on these wells.

References

[1] Thomas Blaschke, Stefan Lang, Eric Lorup, Josef Strobl, and Peter Zeil. Object-oriented image processing in an integrated GIS/remote sensing environment and perspectives for environmental applications. *Environmental information for planning, politics and the public*, 2:555–570, 2000.

[2] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pages 1126–1135, 2017.

[3] Government of Alberta. Landowner and indigenous community site nomination. `https://www.alberta.ca/landowner-and-indigenous-community-site-nomination.aspx`, 2021.

[4] Government of Canada, Statistics Canada. Greenhouse detection with remote sensing and machine learning: Phase one. `https://www.statcan.gc.ca/eng/data-science/network/greenhouse-detection`, Sep 2021.

[5] Mary Kang, Shanna Christian, Michael A. Celia, Denise L. Mauzerall, Markus Bill, Alana R. Miller, Yuheng Chen, Mark E. Conrad, Thomas H. Darrah, and Robert B. Jackson. Identification and characterization of high methane-emitting abandoned oil and gas wells. *Proceedings of the National Academy of Sciences*, 113(48):13636–13641, 2016.

[6] Jakaria Rabbi, Nilanjan Ray, Matthias Schubert, Subir Chowdhury, and Dennis Chao. Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network, 2020.

[7] David Rolnick, Andreas Veit, Serge J. Belongie, and Nir Shavit. Deep learning is robust to massive label noise. *CoRR*, abs/1705.10694, 2017.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015.

[9] Ling Tang and Sangeet Matthew. Using machine learning and deep learning with imagery in ArcGIS. `https://www.esri.com/conteacnt/dam/esrisites/en-us/about/events/media/UC-2019/technical-workshops/tw-5748-894.pdf`.

[10] Irem Ülkü and Erdem Akagündüz. A survey on deep learning-based architectures for semantic segmentation on 2d images. *arXiv preprint arXiv:1912.10230*, 2019.

[11] Irem Ulku, Panagiotis Barmpoutis, Tania Stathaki, and Erdem Akagunduz. Comparison of single channel indices for U-Net based segmentation of vegetation in satellite images. In *Twelfth International Conference on Machine Vision (ICMV 2019)*, volume 11433, pages 338 – 345, 2020.

[12] James P. Williams, Amara Regehr, and Mary Kang. Methane emissions from abandoned oil and gas wells in Canada and the United States. *Environmental Science & Technology*, 55(1):563–570, 2021.

[13] Michael Wurm, Thomas Stark, Xiao Xiang Zhu, Matthias Weigand, and Hannes Taubenböck. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150:59–69, 2019.

[14] Shuai Yang, Qihao Chen, Xiaohui Yuan, and Xiuguo Liu. Adaptive coherency matrix estimation for polarimetric SAR imagery based on local heterogeneity coefficients. *IEEE Transactions on Geoscience and Remote Sensing*, 54(11):6732–6745, 2016.

[15] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.