Discovery of Undocumented Oil and Gas Wells

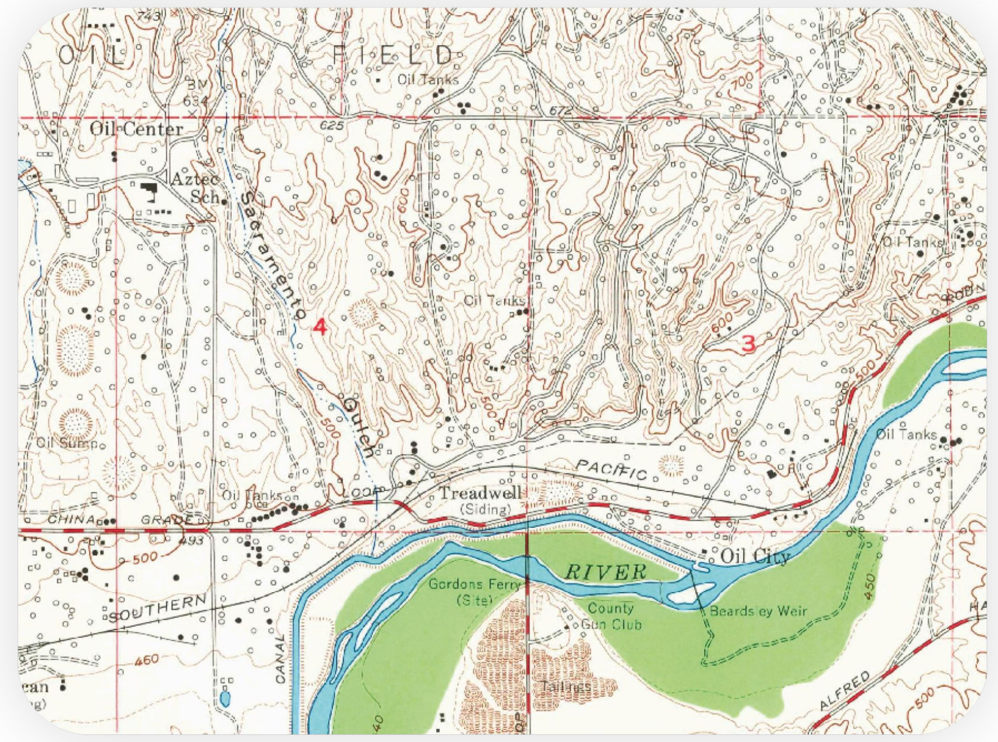
Ryan Beam   Armand Bidault   Nick Lemoff   Grace Zhang   Supervised by Fabio Ciulla

BERKELEY LAB

example of a leaking well

The United States has a long history of oil extraction with about **3.7 million oil and gas wells** drilled starting from the 1850s. Besides those, The Interstate Oil and Gas Compact Commission estimates that there are up to **800,000 wells that are undocumented**, namely whose location and status is unknown. These wells pose a **risk to the environment and human health** in that they can leak methane to the surface and contaminate freshwater aquifers underground.

We contributed to the development of a method to **automatically identify the locations** of these undocumented wells using **historical topographic maps**. These maps display natural features like mountains and rivers, but also anthropogenic structures such as roads, buildings and, notably, oil and gas wells. Our contribution helped enhance the detection of oil and gas well symbols from these maps and started extending this capability to non-georeferenced ones.



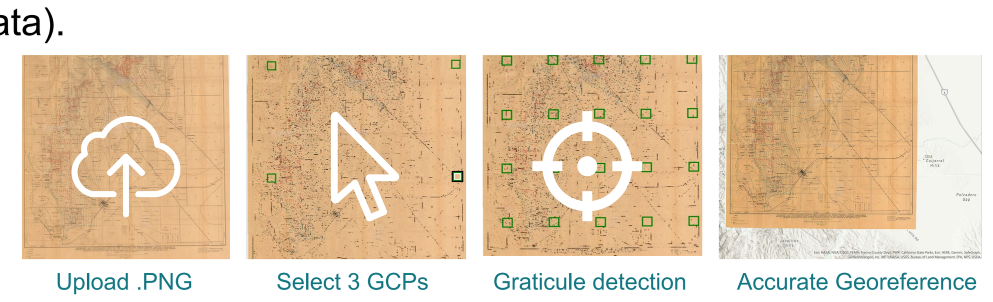
Deep neural networks for computer vision are the main tool to identify well symbols. We investigated the performance of models like the **Segment Anything Model (SAM)** both out-of-the-box and after **fine tuning**. Additionally, the popular architecture **U-Net** has been improved via **image augmentation** and exploration of **ensemble techniques**. Traditional computer vision methods like **edge detection** and **color partitioning** are also analyzed, as well as post-processing steps to compare the location of detected symbols with respect to the position of documented wells in official databases.

Finally, we expanded the methodology to **non-georeferenced maps**, allowing to extract information from maps dated as early as the **beginning of the 20th century**. Semi-automatic georeferencing tools have been researched and GIS software used to isolate the location of old wells unavailable in common topographic maps.

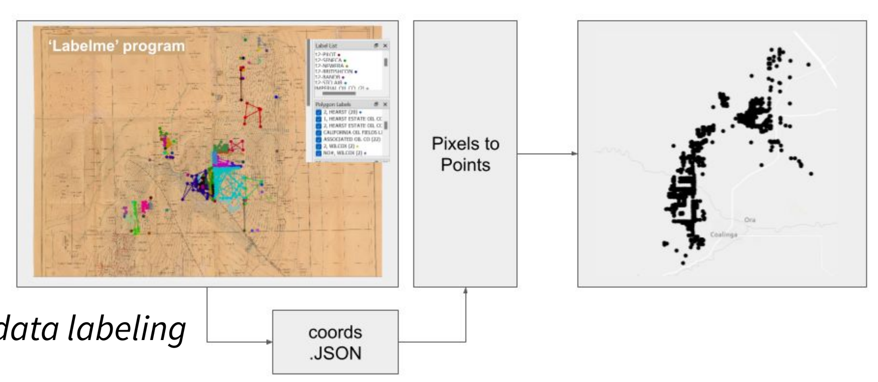
Consortium Advancing Technology for Assessment of Lost Oil & Gas Wells.

Georeferencing, labeling, and cross-referencing

LBNL's current program to identify orphan wells **only** accepts *geo--referenced 'quadrangle' maps* as input. However, there is **valuable data** in maps that **are not georeferenced** (do not carry geographical metadata).



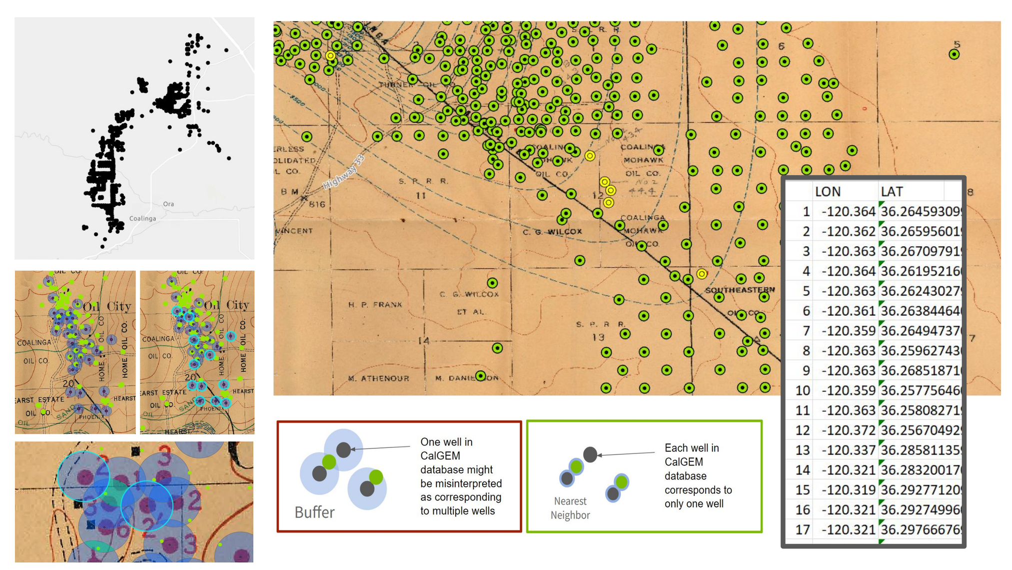
We use **semi-automatic georeferencing** tools to quickly generate geographical metadata for these generic maps, **preparing** them for **labeling**.



Manual data labeling

Post-feature-extraction workflow:

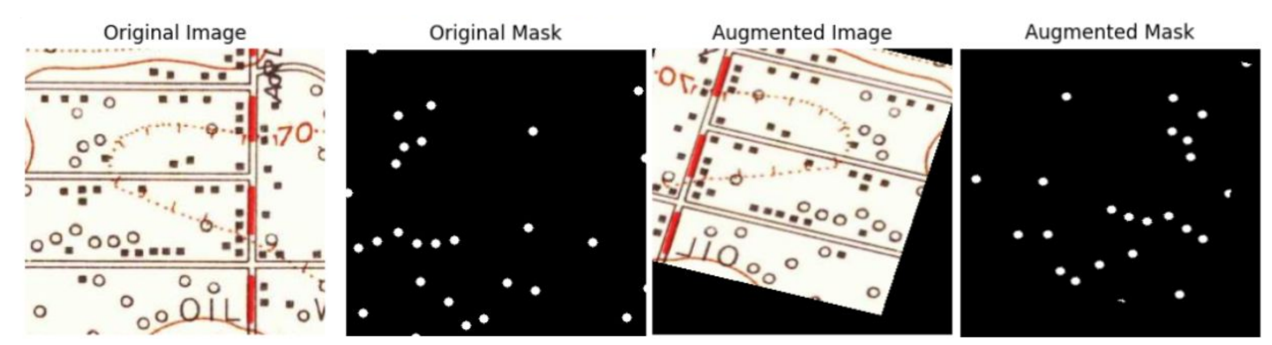
- **Convert pixel coordinates** on geoTIFF map to latitude, longitude (osgeo Python library)
- Cross-reference **extracted** well **feature locations** with California government (CalGEM) well database
- **Generate map** and **.csv** of suspected orphan well coordinates in ArcGIS, forward to field team



Optimizing the Computer Vision Process

We enhanced the precision of the computer vision neural network algorithm while ensuring its runtime efficiency.

The first step involved implementing **on-the-fly augmentation**, which dynamically modifies the input data during the training phase, enriching the dataset without the need for additional storage.This helps the model to generalize better to new, unseen data by **simulating a more comprehensive range** of possible scenarios.



In the second phase, we applied **bagging and boosting ensemble techniques**, each offering distinct benefits. Bagging, involves training multiple models on different subsets of the data and then averaging their predictions. This method is particularly effective in **reducing variance and overfitting.** Boosting sequentially trains models, where each model attempts to correct the errors of the previous ones. This approach is instrumental in **increasing the model's accuracy.** This is an ongoing effort to with the goal of combining advantages from both methods.

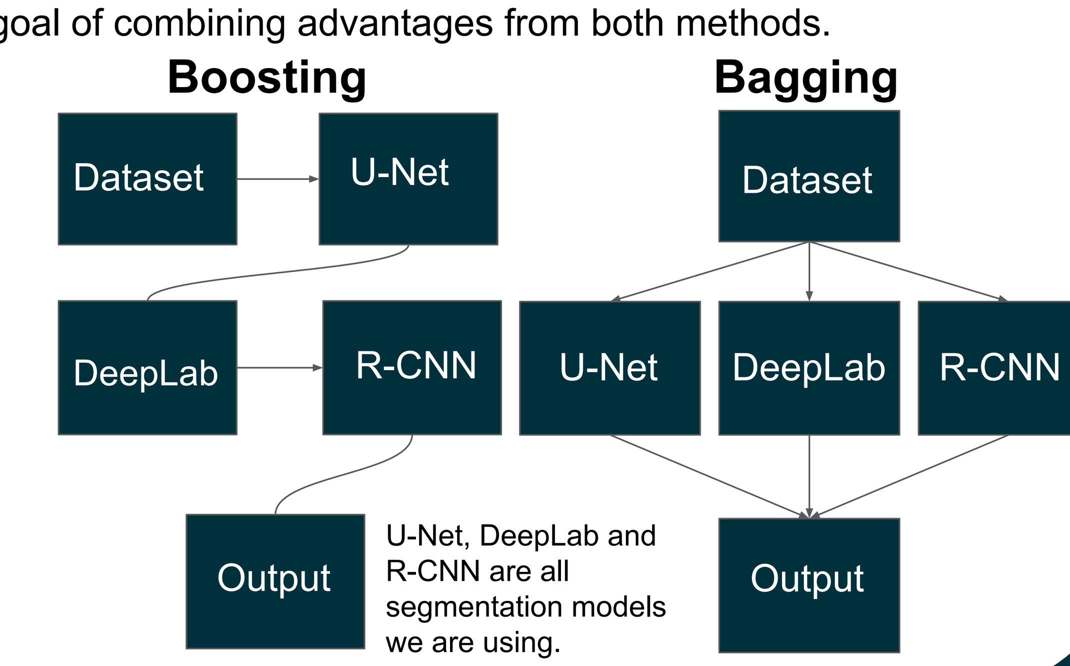
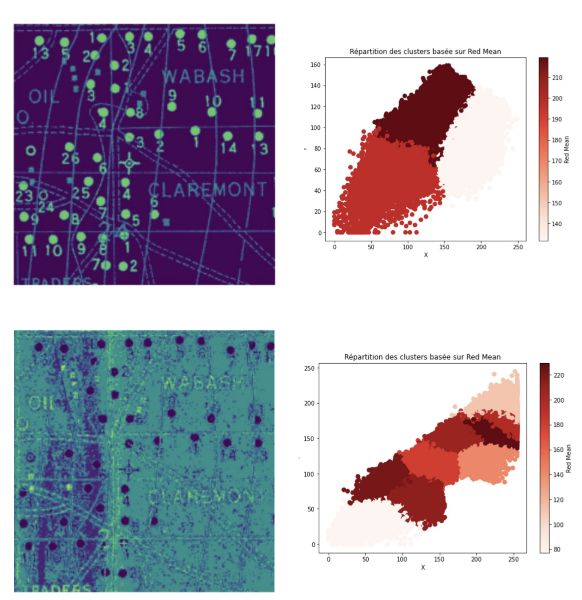


Image Processing & Coordinate correction

Working with **old topographic maps** is unusual for an artificial intelligence model. We therefore tried to **highlight key information** using different methods.

We applied different **color filters**, different **effects** (blur, erode, etc.) and different types of **binarization** to our maps to highlight the wells.

** Here we subtract two binary masks obtained by filtering red and dark colors to obtain another one*

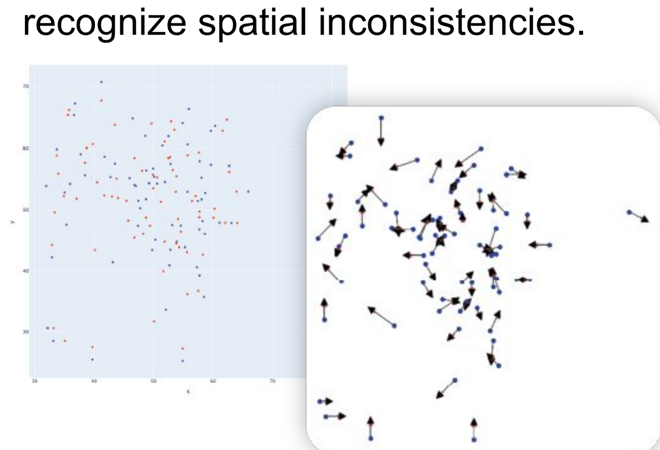


We used a **clustering** technique (K-means) on the color spectrum of the maps to extract different groups of pixels.

K = 3 is not precise enough in all cases.
K ≈ 9 is a suitable value for differentiating points from wells.

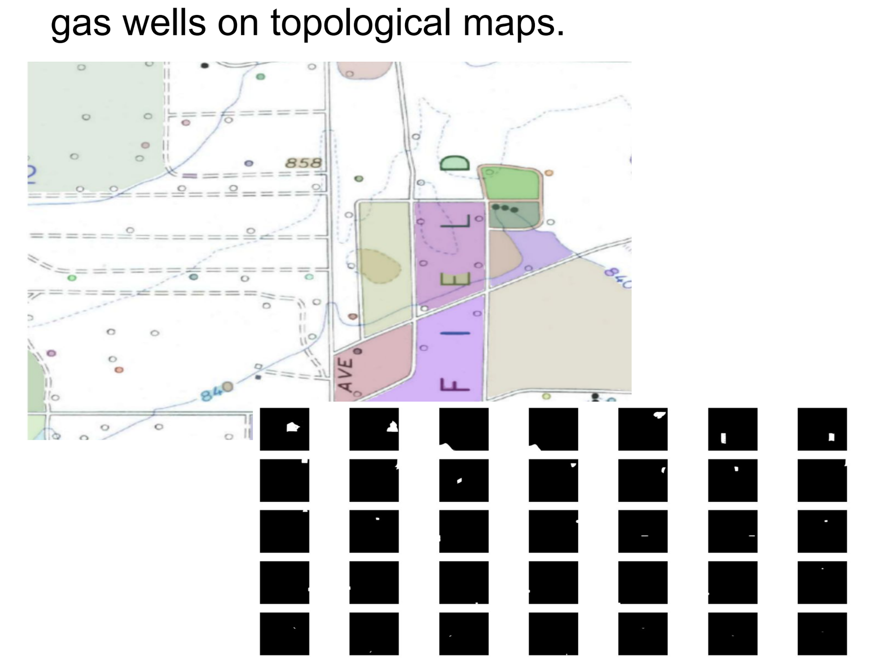
Documented well location do not always exactly match those in the maps. This is due to systematic errors in official databases coordinates, limitation of spatial accuracy or other reasons. We investigated methods that accounts for spatial shifts to **improve geographical correspondence.**

Lots of room for improvement: use a robust **Iterative Closest Point** algorithm and consider training a model that recognize spatial inconsistencies.



SAM for Oil and Gas Well Segmentation

We investigated SAM (Segment Anything Model) to segment and identify undocumented gas wells on topological maps.



Current Progress:

- **Training the Original SAM:** The original SAM model was initially trained and its performance was evaluated. The key observation from this phase was that the model struggled to segment out small gas wells, indicating a need for further refinement.
- **Fine-Tuning on Non-Augmented Dataset:** The model underwent fine-tuning on a non-augmented dataset following the initial training. This process included training with 5 epochs using sam-vit-base weights. The focus here was to enhance the model's capability in accurately identifying and segmenting the gas wells.
- **Challenges Encountered:** Despite the fine-tuning efforts, SAM still currently faces challenges in correctly identifying and segmenting out oil wells.

Potential Improvements: Incorporating different weights for fine-tuning