Introduction

The overarching project idea involves the detection, monitoring, and characterization of uranium enrichment and plutonium reprocessing facilities using satellite imagery. This includes utilizing various types of satellite data such as thermal, optical, synthetic aperture radar (SAR) imagery. SAR imagery is obtained by sensors emitting microwave signals towards the Earth's surface and then recording the reflected energy. This process creates detailed images, regardless of weather conditions and lighting.

By combining these diverse data sources, the aim is to identify and monitor the specific signatures of operational nuclear facilities. The ultimate goal is to develop a comprehensive methodology for non-intrusive surveillance that can provide continuous updates on the status and activity levels of these sites.

This application domain is crucial for several reasons. First, satellite-based surveillance provides a non-intrusive method for monitoring nuclear facilities, eliminating the need for physical inspections, which are often politically sensitive and logistically challenging. Second, regular monitoring through satellite imagery ensures compliance with international treaties, such as the Non-Proliferation Treaty (NPT) (Niemeyer, Dreicer, & Stein, 2020). Third, continuous updates from satellite data enable the implementation of early warning systems, which can detect changes in reactor activity level and facilitate preventative diplomatic measures (Göttsche & Glaser, 2021).

Datasets

The dataset used for training and evaluation is the Power Plant Satellite Imagery Dataset (Bradbury, Brigman, & Chandrasekar, 2017). It contains satellite images from the Landsat program, specifically from 4,454 power plants across the United States. Additionally, there are binary annotations for each image and metadata for some observations included (Figure 2).

The original US Power Plant Dataset project (Brigman, Chandrasekar, Hossain, Li, & Nagenalli, 2017) employed an LDA classifier, which could be run using either the available metadata file containing about 144 instances or by generating it through provided code on the raw data file. This process is not only time-consuming but also demands significant storage space. To address these challenges, I utilized the already available folders of preprocessed images and binary masks and developed an alternative method to match images and masks, bypassing the need for the metadata file. This approach allows the use of all 4,454 image-mask pairs, enhancing the dataset's utility for training and evaluation with CNNs.

Previous work on this dataset has therefore focused on traditional machine learning techniques. The proposed use of CNNs represents a significant advancement, leveraging the power of deep learning to enhance detection accuracy and efficiency. Binary cross-entropy was used as the loss function, and the Adam optimizer was employed for training and evaluation.

Models

Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a statistical technique used for classification. It finds a linear combination of features that best separates two or more classes. The classifier used performs binary classification, i.e., distinguishing between two categories, such as plant vs. non-plant. A 'loadObservation' method extracts features from each image (Li, 2017). The features include the mean and variance of the RGB channels computed within a window around each pixel. The features describe the local texture and color information. After training, the classifier can predict the probability of each pixel belonging to the positive class. I used this example in order to understand how to work with the dataset and to do classification with CNNs.

```
Feature Properties
  egrid_ID: 300
  plant_name: Black Mountain Generating Station
  state_name: AZ
  county_name: MOHAVE
  primary_fuel: NG
  fossil_fuel: GAS
  capacity_factor: 0.049
  nameplate_cap_MW: 121.0
  co2_emission: 30564.6130614826
  availability: NAIP&LANDSAT
```

Figure 2: Metadata example

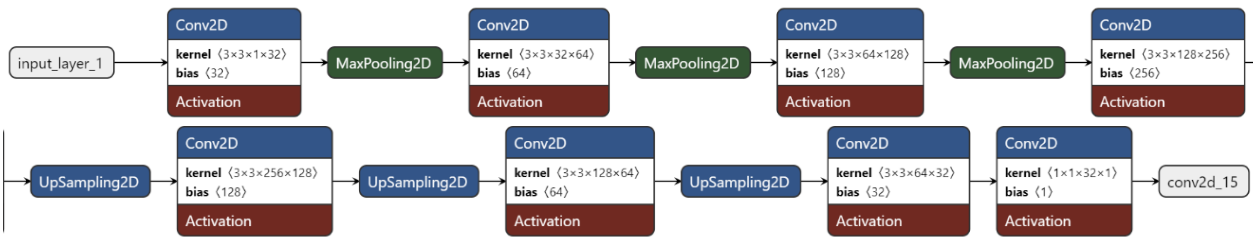

Figure 1: Custom CNN Architecture

Custom CNN

Before using larger pre-trained models, I decided to use a simple custom model (Figure 1) to understand how to design the image processing and dimension adjustments. First, I had to address data loading since the metadata file only contained 144 instances, instead of all 4,454.

1. Directory Setup: Defines paths to directories containing the images and masks
2. Data Extraction: Collects and organizes file paths of images and masks based on their IDs, which do not match this dataset. This ensures that each image and its corresponding mask are correctly paired.
3. Loading and Resizing: Loads each image and mask, resizes them to a standard size of 2256x256 (if not already the case), and stores them in separate lists.
4. Error Handling: Includes checks to handle cases where images or masks fail to load to ensure that the data is handled correctly before training the CNN. I used the sigmoid activation function (binary classification) and the Adam optimizer.

I used the sigmoid activation function (for binary classification) and the Adam optimizer.

Pretrained U-Net

First introduced in "U-Net: Convolutional Networks for Biomedical Image Segmentation" (Ronnenberger, Fischer, & Brox, 2015), U-Net (Figure 11) is a convolutional neural network architecture designed for image segmentation tasks. It features an encoder-decoder structure with the encoder capturing important features of the image and the decoder using the feature to create the final output.

In my project, I used a version of U-Net that comes with a VGG16 backbone. The VGG16 part enhances the U-Net model's performance on the dataset as it has already learned patterns from training on other images. By using this pretrained architecture, I aimed to leverage VGG16's learned features and improve the accuracy of segmentation on the power plant imager dataset.

Pretrained U-Net with Augmentation (Figure 3)

```
image_datagen = ImageDataGenerator(
    rotation_range=20,
    width_shift_range=0.2,
    height_shift_range=0.2,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    fill_mode='nearest'
)
```

Figure 3: Augmentation Techniques

- Rotation_range=20: Randomly rotates the image within the specified range of degrees (± 20 degrees). This helps the model become invariant to the orientation of object in the image.
- Width_shift_range=0.2: Randomly shifts the image horizontally by up to 20% of the total width. This helps the model handle variations in object placement.
- Height_shift_range=0.2: Randomly shifts the image vertically by up to 20% of the total height. Similar to width shift, this helps the model handle vertical variations in object placement.
- Shear_range=0.2: Applies random shearing transformations to the image. This helps the model handle distortions and variations in object shapes by simulating skewed perspectives.
- Zoom_range=0.2: Randomly zooms in or out of the image by up to 20%. This simulates variations in object scales and improves the model's ability to generalize across different zoom levels.
- Horizontal_flip=True: Randomly flips the image horizontally. This improves the model's robustness to left-right variations in object orientation.
- Fill_mode='nearest': Determines how to fill in pixels that are created by the aforementioned transformations. The 'nearest' mode uses the nearest pixel value to fill these areas, ensuring the augmented images are visually coherent.

<u>Results</u>

Linear Discriminant Analysis (LDA)

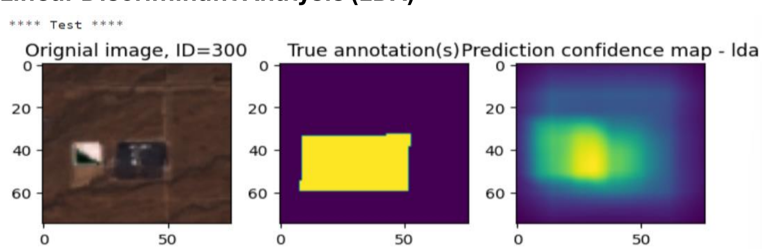

Figure 5: Prediction Example LDA

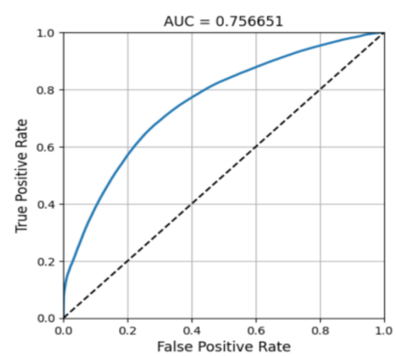

Figure 4: ROC Curve

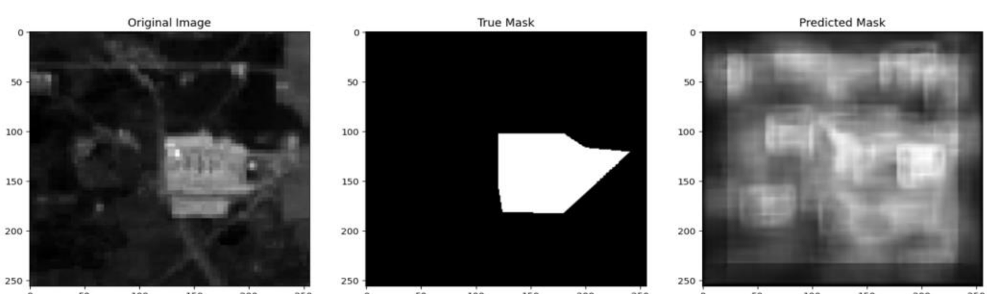

Figure 6: Prediction Example Custom CNN

Custom CNN

As expected, the less complex custom CNN model's prediction accuracy is notably lower, however its performance is superior compared to the LDA classifier. This discrepancy is evident in the accuracy comparison table (Table 1), which compares the results of all models used. The less sophisticated architecture of the custom CNN model nevertheless struggles with capturing the patterns necessary for accurate predictions, thereby leading to worse performance compared to the U-Net models. This outcome underscores the limitations of simpler models in complex classification tasks, where more advanced classifiers can provide better results.

Table 1: Accuracy Comparison Table

LDA	Custom CNN	Pretrained U-Net	Pretrained U-Net with Augmentation
77%	79%	82%	81%

Pretrained U-Net

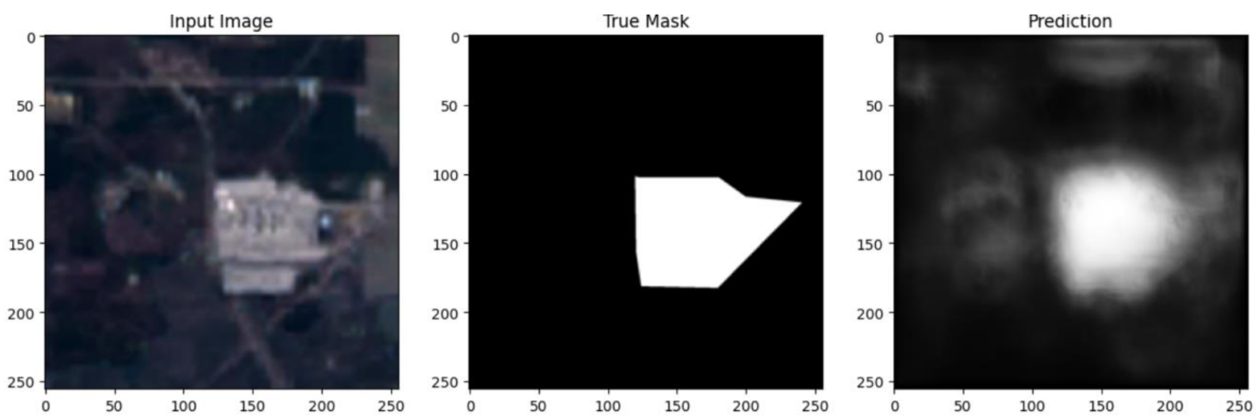

Figure 7: Prediction Example Pretrained U-Net

From Figure 7, it is evident that the pretrained U-Net model demonstrates a significantly improved performance in capturing the most critical parts of the power plant building. The model effectively distinguishes the power plant from the background, with discrepancies occurring in areas where other buildings are situated nearby. This indicates that the pretrained U-Net model is able to identify and delineate key features of buildings from the background.

Pretrained U-Net with Augmentation

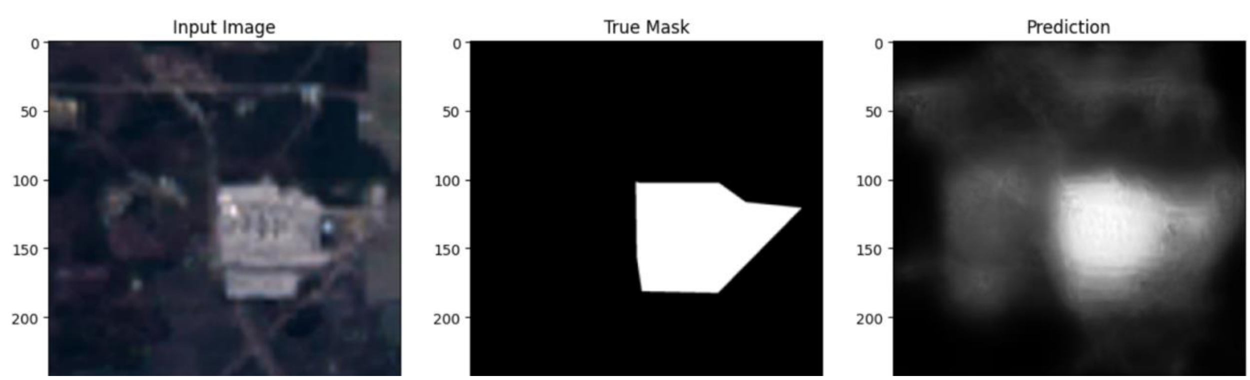

Figure 8: Prediction Example Pretrained U-Net with Augmentation

The pretrained U-Net model with augmentation exhibits similar performance, with the added benefit of slightly clearer delineation of the mask shape (Figure 8). Augmentation seems to enhance the model's ability to refine the boundaries of the power plant building, leading to more precise segmentation. However, it still faces the same challenges of differentiating the power plant from other buildings in the vicinity.

Figures 9 and 10 illustrate the accuracy and loss curves for both the custom CNN and pretrained U-Net models. The relatively small dataset contributes to noticeable overfitting in the U-Net model, as evidenced by the divergence between the training and validation loss curves.

Overall, both models perform well in segmenting the power plant, but the augmented version offers a marginal improvement in clarity while continuing to grapple with the issue of nearby structures.

Discussion

These results underscore the feasibility of using CNNs for detecting power plants in satellite imagery. The performance of the pretrained U-Net model demonstrates that CNNs are capable of effectively identifying and segmenting power plant buildings from satellite images.

Future work will include the analysis of thermal and SAR imagery. Thermal imagery can provide valuable insight into the activity levels of power plants. Active power plants typically exhibit thermal changes not only in the buildings themselves but also in nearby water bodies used for cooling. This thermal signature can be indicative of operational status. For example, imagery analysis at Stanford have utilized thermal satellite imagery to assess the activity levels of nuclear power plants, demonstrating the practical applications of this technology (Park, Puccioni, & Verville, 2024).

SAR imagery has the potential to detect increased activity through changes in radar backscatter. Although there is currently only limited research on using SAR imagery specifically for monitoring nuclear power plants (Shin, Flower, & Pasquali, 2022), its ability to capture detailed surface changes regardless of weather conditions and lighting makes it a promising tool for future studies.

Combining these different types of satellite data with CNNs will likely lead to even more robust and precise detection methods, advancing the field of satellite-based power plant monitoring and detection.

Ethical Concerns

1. Privacy Implications: Monitoring power plants, including nuclear facilities, involves capturing detailed imagery and information about sensitive infrastructure, which rases privacy concerns. Unauthorized surveillance or espionage could occur if the data is misused.
2. Security Risk: Detailed analysis of nuclear facilities and power plants could inadvertently reveal sensitive information about their capabilities and vulnerabilities. This poses security risks if such information is accessed by malicious actors, potentially leading to targeted attacks or breaches.
3. Training Bias and Generalization Issues: Machine and deep learning models trained primarily on data from specific region or types of power plants may develop biases, affecting their accuracy in other contexts. For example, models trained on Western nuclear facilities might perform poorly on facilities in different regions due to variations in design and appearance. Incorrect classification of facilities or misjudging activity levels can lead to serious consequences, such as unnecessary investigations or military actions.
4. Potential Harm from Wrong Classifications: Misclassification of facility's activities as harmful could result in unwarranted escalations, such as military interventions or sanctions. This could cause substantial harm to humans and the environment, including accidental damage to infrastructure, disruption of power supply, or public panic.

Note: The software to run the code was Jupyter Notebook (Kluyver, Ragan-Kelley, & Pérez, 2016) and the visualization tool used to visualize the convolutional neural network was Netron (Roeder, 2019).

GitHub Repository: https://github.com/jasminsteinbrecher/AI4EO-final.git

<u>References</u>

Bradbury, K., Brigman, B., & Chandrasekar, G. (2017). Retrieved from https://figshare.com/articles/dataset/Power_Plant_Satellite_Imagery_Dataset/5307364?file=9104302

Brigman, B., Chandrasekar, G., Hossain, S., Li, B., & Nagenalli, T. (2017). *USPowerPlantDataset*. Retrieved from https://github.com/bl166/USPowerPlantDataset

Fuhrmann, M. (2015). Retrieved from http://www.matthewfuhrmann.com/datasets.html#:~:text=The%20NL%20dataset%20provides%20information,globally%20during%20the%20nuclear%20age.

Göttsche, M., & Glaser, A. (2021). Toward Nuclear Disarmament: Building Up Transparency and Verification. *German Foreign Office*.

Kluyver, T., Ragan-Kelley, B., & Pérez, F. (2016). Jupyter Notebooks - a publishing format for reproducible computational workflows. *Positioning and Power in Academic Publishing: Players, Agents and Agendas*.

Li, B. (2017). Retrieved from https://github.com/bl166/USPowerPlantDataset

Niemeyer, I., Dreicer, M., & Stein, G. (2020). *Nuclear Non-Proliferation and Arms Control Verification*. Springer Cham.

Park, S., Puccioni, A., & Verville, F. (2024). Retrieved from https://hai.stanford.edu/events/thermal-satellite-imagery-detection-and-monitoring-nuclear-weapons-early-approaches

Roeder, L. (2019). *Netron: Visualizer for neural network, deep learning and machine learning models*. Retrieved from https://netron.app/

Ronnenberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention* .

Shin, J., Flower, M., & Pasquali, P. (2022). Usage of Synthetic Aperture Radar Data For Nuclear Fuel Cycle Monitoring. *IAEA's Symposium on International Safeguards: Reflecting on the Past and Anticipating the Future*.

Appendix

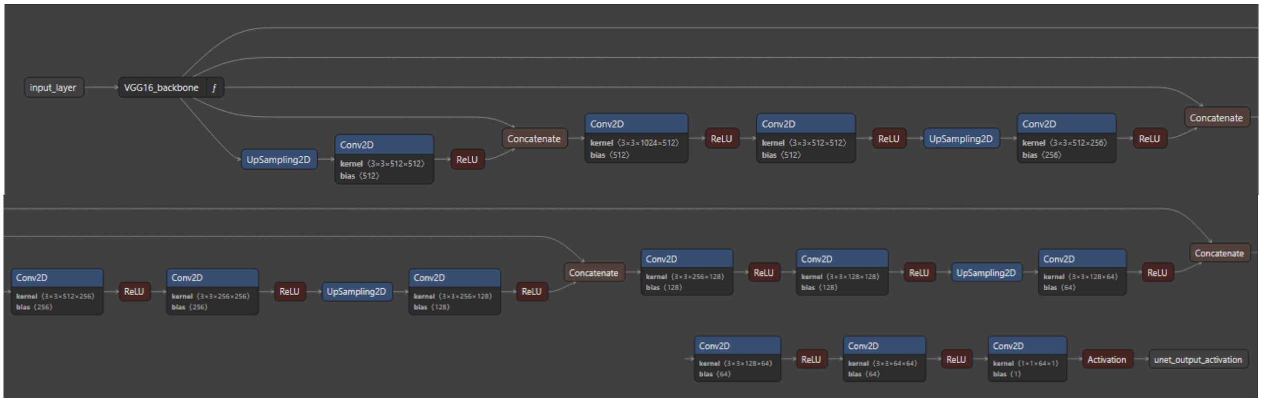

Figure 11: Pretrained U-Net Architecture

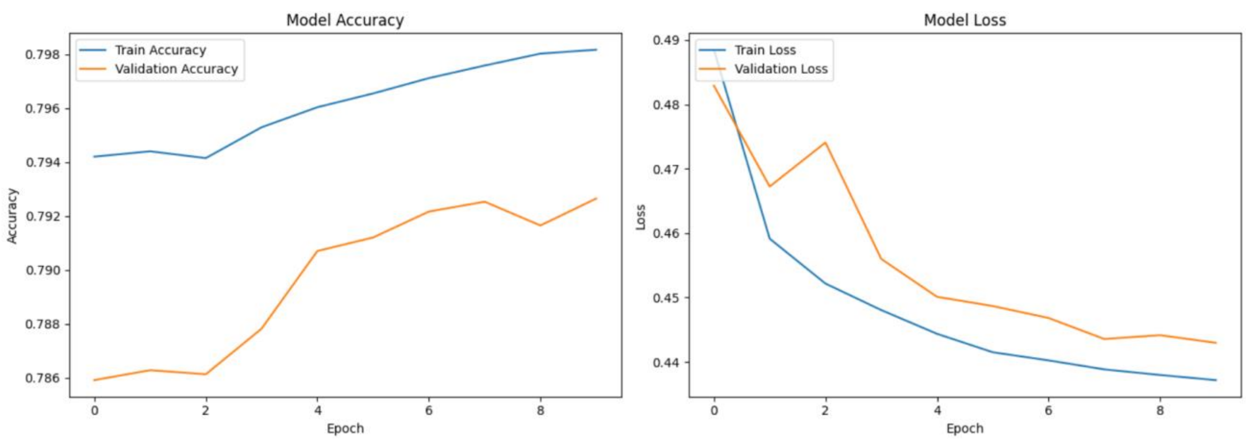

Figure 10: Custom CNN Accuracy & Loss Curves

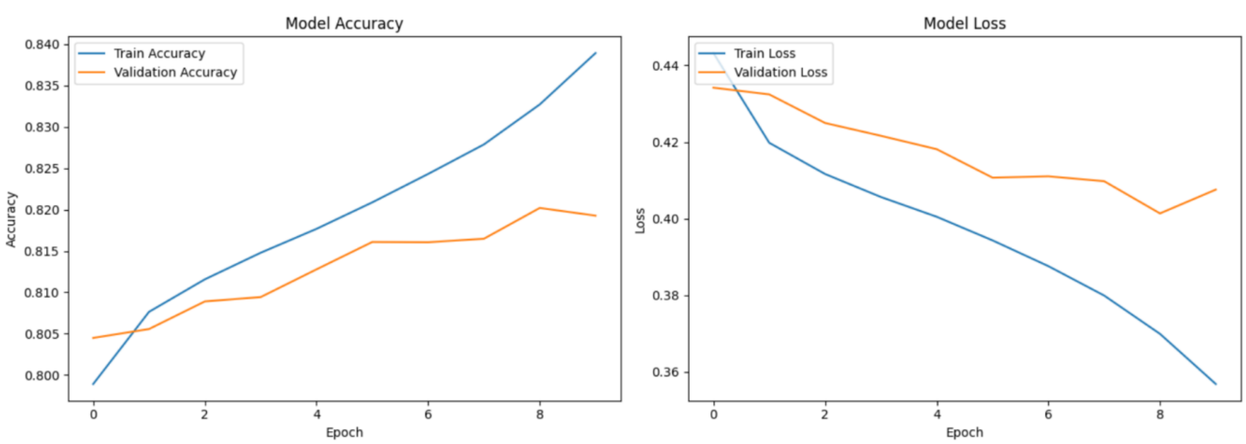

Figure 9: Pretrained U-Net Model Accuracy & Loss Curves