Deep Learning for Rare Energy Infrastructures in Satellite Imagery

Project Members: Tyler Feldman, Matt Robbins
Project Manager: Bohao Huang
Team Leaders: Kyle Bradbury, Jordan Malof
Team Contributors: Leslie Collins, Robert Fetter, Marc Jeuland, Luana Marangon Lima, Robyn Meeks
Lack of Training Data: Obstacle for Rare Objects

- Deep neural networks require **large amount of training samples**
- Many energy infrastructures are “rare”, which makes it difficult to acquire enough training data

![Power Plants](image1.png)  ![Wind Turbines](image2.png)  ![Oil Tanks](image3.png)
CityEngine: Create Synthetic Images

1. Virtual World Creation -- CityEngine
   - Virtual world generation
     - Import 3D models
     - Randomize size and placement

2. Synthetic Overhead Imagery Generation
   - Capture imagery

Output
Research Steps

1. Create and process a real dataset to train our model as a baseline

2. Create synthetic data using CityEngine and online 3D models of energy infrastructure

3. Train our network with a dataset supplemented by our synthetic data and compare with our baseline
Looking at our Data

- Our dataset is a set of satellite imagery containing power plants
- We focus specifically on images containing wind turbines
Data Preprocessing

• Labeled and localized each turbine
  • Labels are used as ground truths that can be compared with the model’s predictions to determine performance

• Raw satellite images split into patches
• Adapt labels for each patch
Creating Synthetic Data

• Generated models on top of images that did not contain wind turbines
• Used a script to generate new images and models randomly and then change the camera position and take photos of the scene
Experimental Setup

**Training Dataset A**
- 1239 real images of wind turbines

**Training Dataset B**
- 1239 real images of wind turbines
- 441 synthetic images

**Testing Data**
- 661 real images

Results with real training images
Results with real + syn. training images

Data is split between training and testing randomly
Performance Metrics

- **Precision and Recall**
- **Prioritize recall**
  - Easier to remove misclassified turbines than to find undetected wind turbines

![Diagram showing correct predictions, missed turbine, and incorrect predictions]

\[
\text{Precision} = \frac{2}{4} \\
\text{Recall} = \frac{2}{3}
\]
Adding Synthetic Data Improves Performance

<table>
<thead>
<tr>
<th>Training Data</th>
<th>Testing Data</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real</td>
<td>Real</td>
<td>0.813</td>
<td>0.825</td>
</tr>
<tr>
<td>Real + Synthetic</td>
<td>Real</td>
<td>0.831*</td>
<td>0.827</td>
</tr>
</tbody>
</table>
Performs well on large wind turbines
Inconsistent on small wind turbines
Future Work

- Make synthetic imagery more representative of the real imagery
- Observe performance as the amount of synthetic data is varied
- Apply this model on a large scale
- Apply these techniques to detect other types of energy infrastructure
Thank You for Listening

- Project Website: https://dataplus-2020.github.io/
- Project GitHub Repository: https://github.com/dataplus-2020/yolov3_wnd_code
- Energy Data Analytics Lab: https://energy.duke.edu/research/energy-data