Project Sunroof

Pranjal Patil (ppatil@stanford.edu) Vedang Vadalkar (vedangv@stanford.edu)

Background and Motivation

Solar energy has become a vital part of the energy infrastructure and attempts are being made to achieve net zero energy buildings in California by 2020. Easy prediction of the solar energy potential of rooftops will lead to easy adoption of the solar energy systems. In this project we apply deep learning networks to aerial imagery to semantically classify roof and non roof pixels. We use convolutional Neural network based on U-Net as a starting point for our project. This architecture was originally developed for biomedical image segmentation for binary segmentation.

Dataset

- The dataset in the form of images was from Massachusetts Building Dataset
- Dataset consisted of 150 images of dimension 1500*1500
- Each image was also accompanied with the ground truth image demarcating the building footprint

Input Data Visualization

Input Image

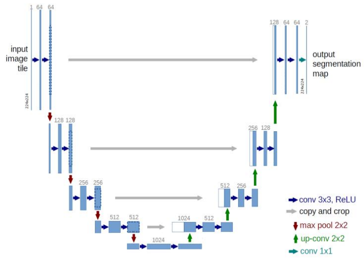
Ground Truth

Features

- 224*224 crops from each image were generated to match the input size of model
- Thus effective size of dataset is 5400
- Further the input images were converted to arrays to feed into the model
- Accuracy was used as the metric to compare various architectures

Model

- A Convolutional Neural Network model based on U-Net architecture was used
- A sequence of convolutions, ReLU activations and max poolings contracts the input to extract features of different levels, allowing to capture the context of each pixel
- Then upsampling is carried out to increase the resolution of the detected features
- The number of features is doubled at each level of downsampling



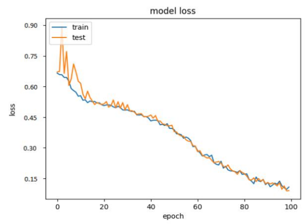
The loss metric in U-Net architecture was modified to binary cross entropy loss

$$BCE = -\frac{1}{N}\sum_{i=0}^{N} y_i \cdot log(\hat{y}_i) + (1-y_i) \cdot log(1-\hat{y}_i)$$

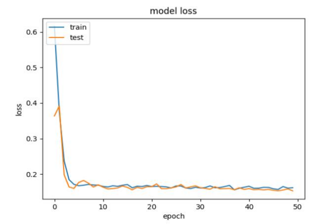
Results and Discussion

Model	Validation Loss	Validation Accuracy	Training Loss	Training Accuracy
11-Layers U-Net (Max filters 1024)	0.039	0.921	0.027	0.953
9-Layers U-Net (Max filters 1024)	0.192	0.881	0.189	0.896

Loss curve: 9 layer U-Net



Loss curve: 11 layer U-Net



- Two models with 9 layers and 11 layers were trained on the dataset with a train, validation and test split of 90% , 5%, 5% respectively
- The 11 layer model with maximum 1024 filters achieved accuracy comparable to the current state of the art models for semantic segmentation
- An extra layer of 1024 filters had to be added to increase accuracy of the model to be comparable to performance of U-Net models on other datasets
- On visualization it was found that the model was lacking in correctly classifying roof pixels, this was due to the class imbalance in the input data
- Adam Optimizer was used since it is known to converge faster than SGD

Future Work

- Deeper analysis of the errors and features leading to poorly classified pixels
- Adding further analysis after detecting the roof pixels that will help in detecting the size of the roof and the orientation of the roof
- Start differentiating between the roofs with solar panels vs those without any solar panels

References

1. Olaf Ronneberger et al: U-Net: Convolutional Networks for Biomedical Image Segmentation
2. Jonathon Long et al: Fully Convolutional Networks for Semantic Segmentation
3. Jiangye Yuan: Automatic Building Extraction in Aerial Scenes using Convolutional Networks