Meth-AIN'T

Locating oil & gas infrastructure from aerial images

Sarah Daniels

December 7th, 2022

Background

What? Methane is responsible for 25% of global warming

Where? Global methane emissions from the energy sector are 70% more than reported

Why? Satellite imagery shows 1,800 methane gas leaks largely due to oil and gas infrastructure (ref)

When? New satellites dedicated to higher res methane plume detection will be launched soon (ref)

Who? U.S. Environmental Protection Agency (EPA) clamping down on methane leaks

Objective: Locate U.S. oil & gas industry point sources with focus on minimizing false negative results

Target metric = Recall (with reasonable precision)

The Washington Post
Democracy Dies in Darkness

EPA to regulate methane leaks from oil and gas to fight climate change

By Sarah Kaplan and Dino Grandoni
Updated November 11, 2022 at 6:23 a.m. EST | Published November 11, 2022 at 5:30 a.m. EST

A flare burns off methane and other hydrocarbons as oil pumpjacks operate in the Permian Basin in Midland, Tex. (David Goldman/AP)

Methods

Data Collection

SOURCE
Aerial Imagery U.S. National Agriculture Imagery Program (NAIP)

CURATOR
Stanford's ML group

DATA
Training: 5,525 (-) | 127 (+)
Validation: 693 (-) | 13 (+)
Test: 697 (-) | 9 (+)

Data Preparation

Size Reduction

Augmentation

Preprocessing

Transfer Learning

CNN MODELS
Xception

ResNet

DenseNet
(similar to ResNet)

Model Comparisons

Report
Recall, Precision and predicted values from models

Tools

Data Collection
google.colabs, OS, shutil

SOURCE
Aerial Imagery U.S. National Agriculture Imagery Program (NAIP)

CURATOR
Stanford's ML group

DATA
Training: 5,525 (-) | 127 (+)
Validation: 693 (-) | 13 (+)
Test set: 697 (-) | 9 (+)

Data Preparation
Tensorflow (Keras): Applications

Size Reduction
Augmentation
Preprocessing

Transfer Learning
Tensorflow (Keras): Applications, layers

CNN MODELS
Xception
ResNet
DenseNet

Model Comparison
sklearn metrics, matplotlib, numpy, Tensorflow (Keras)

Report
Recall, Precision and predicted values from models

Performed in Google CoLabs

Sample Data

Baseline Models

Model	Val Precision	Val Recall
Xception	0.00	0.00
ResNet50	0.24	0.33
ResNet152	0.18	0.33
DenseNet121	0.00	0.00

Data Imbalance Strategies Tested

Random Undersampling (Majority: Minority, 9:1) N=1102

Random Oversampling (Majority: Minority, 2.4:1) N= 7845

Increase Batch Size (2-3x)

Augmentation

Random:

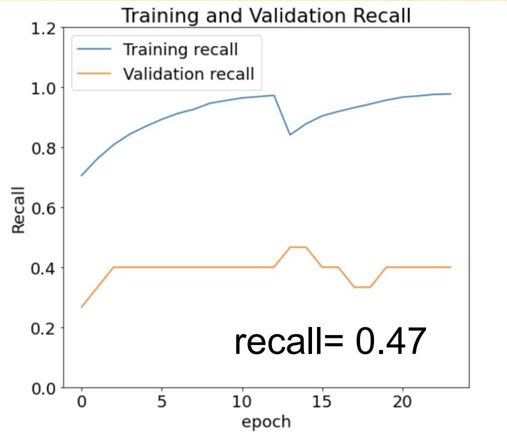
- Flip
- Rotation
- Contrast

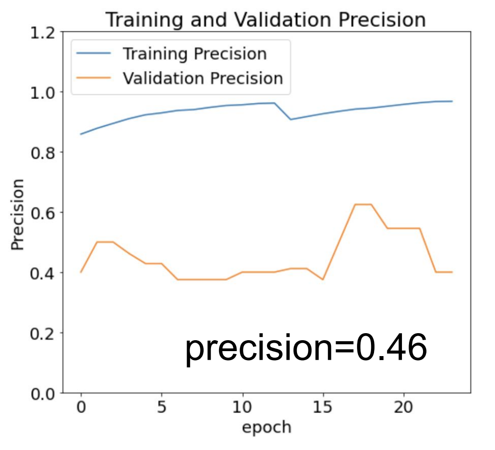
Reduce Parameters with Transfer Learning Layer

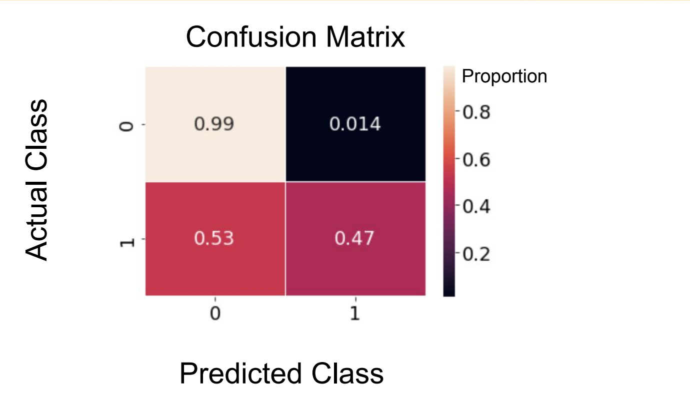
Random Over vs. Undersampling Models

	Baseline		Random Undersampling		Random Oversampling	
Model	Val Precision	Val Recall	Val Precision	Val Recall	Val Precision	Val Recall
Xception	0.00	0.00	0.33	0.07	0.02	1.00
ResNet50	0.24	0.33	0.16	0.33	0.40	0.14
ResNet152	0.18	0.33	0.06	0.13	0.00	0.00
DenseNet121	0.00	0.00	0.50	0.13	0.46	0.47

DenseNet121 Transfer Learning Model







Conclusions

- Severe imbalance requires random oversampling approach
- DenseNet is most optimal for aerial images
- Models are sensitive to aggressive augmentation and parameter reduction (via pooling)

On the Horizon

- **Employ advanced augmentation techniques**
  (i.e., TensorFlow Models Vision Libraries)

- **Try balancing input batches with proportion of each class**
  (i.e., with Keras' BalancedBatchGenerator)

- **Look into alternative domain specific-weights from satellite data**
  (in substitute of ImageNet)

Thank you