Classification of Natural Gas Leaks

Jingfan Wang, Sindhu Sreedhara
Department of Energy Resources Engineering, Stanford University, Stanford CA

Motivation and Problem Statement

- Methane: high global warming potential (a major component of natural gas)
- Problems of current Leak Detection and Repair Technology (LDAR):
  - Labor costs for IR surveys are high
  - Continuous monitoring with IR is infeasible
  - IR surveys cannot tell the operator how big a leak is in the real time
- Importance of methane quantification:
  - Classifying leaks into 'small', 'medium', and 'large' categories is sufficient for cost-effective emission reduction
  - Quantification measurements with great precision are not necessary to achieve benefits of focusing on large leaks
- Our interdisciplinary project expands upon EPA-approved IR imaging and will harness the potential for deep learning advances to allow for rapid automatic classification of methane leaks.

Dataset and Methods

Data Collection and Datasets

- GasImagery: a large dataset for training the CNN network, which includes labeled videos of volume-known methane leaks from various leakage sources covering wide range of leak sizes.
- Half a million of frames taken at METEC (Fort Collins, Colorado)

3-min-video per leak size-distance × 8 leak sizes per distance × 31 24-min-videos × Around 15 frames/s = 669,600 frames

Videos were collected from 2 pieces of equipment: separator on pad 1 (13 videos), separator on pad 2 (18 videos)

GasImagery: Separator Leaks at METEC

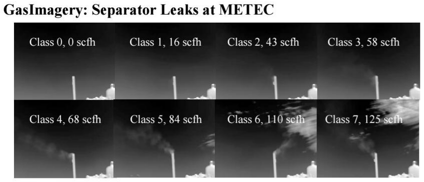
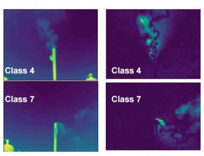


Image Preprocessing: Moving Average Background Subtraction



Before Subtraction   After Subtraction

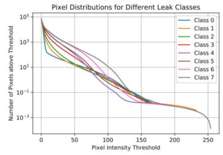
- We generate a unique moving average background for every frame in the video. We generate the background image by calculating the median over the previous 210 images.
- The images on the left represent heat maps of image pixel intensities before and after background subtraction.

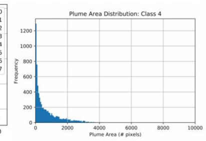
Baseline Method: Image Processing

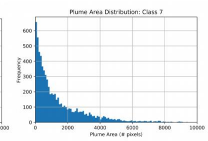
Perform background subtraction on all frames in the training set → Choose a pixel intensity threshold based on the average pixel intensity distributions for all classes → Calculate average number of pixels above the threshold for each of the leak classes → Using the averages from the previous step as breakpoints, classify the test set

Models and Results

Baseline Model







Three Deep Learning Models

1. CNN on Frames

Conv-Pool Structure1
3*3*4, stride 2, zero padding
ReLU
Droput (0.5)
maxpool 2*2, stride 2, zero padding
Batch Normalization

↓

Conv-Pool Structure 2
3*3*8, stride 2, zero padding
ReLU
Dropout (0.5)
maxpool 2*2, stride 2, zero padding
Batch Normalization

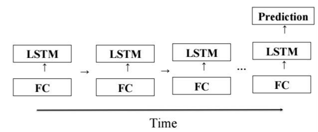
↓

Fully Connected Layer 1
2400-200
Dropout (0.5)

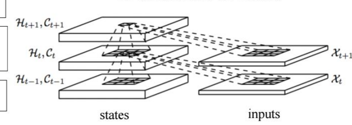
↓

Fully Connected Layer 2
200-8

2. FC+LSTM on Videos



3. ConvLSTM on Videos



Training, Validation and Test Dataset

- 80% of the data from separator on pad 2 is used as training data, and the remaining 20% is used as validation data.
- Data from another piece of equipment (separator on pad 1) is used as test data and never introduced into the model during training or validation.
- We only consider the classification problem on the videos taken in the shortest distance (5 feet).
- Frames: train set (45474), validation set (11368) and test set (37864)
- Videos: if frame rate is chosen to be 5 frames/second, video contains 5 seconds of images, train set (606), validation set (151) and test set (504)

Results of 8-Class Classification Problem 0/1/2/3/4/5/6/7

No.	Model	Random Guess Accuracy	Increase Rate
1	Random Guess	12.50%	0
2	Baseline Model	25.67%	105%
3	CNN on Frames	33.90%	171%
4	FC+LSTM on Videos	39.22%	213%
5	ConvLSTM on Videos	43.34%	246%

Results of 2-Class Classification Problem 0,1,2,3/4,5,6,7

No.	Model	Random Guess Accuracy	Increase Rate
1	Random Guess	50.00%	0
2	Baseline Model	69.21%	38.42%
3	CNN on Frames	76.90%	53.80%
4	FC+LSTM on Videos	80.12%	60.24%
5	ConvLSTM on Videos	83.45%	66.90%

Accuracy with Hyperparameter Change

Model	Method 1	Method 2	Method 3
Parameter	Dropout number	Size of FC	Video Length
Accuracy	1 – 30.23%	50 – 37.31%	5 seconds – 43.34%
	2 – 33.27%	100 – 37.43%	10 seconds – 35.01%
	3 – 33.90%	200 – 39.22%	2 seconds – 37.02%

Confusion Matrix (Y-Lable, X-Prediction)

Based on the highest accuracy in 8-class problem

	Leak 0	Leak 1	Leak 2	Leak 3	Leak 4	Leak 5	Leak 6	Leak 7
Leak 0	43	17	3	0	0	0	0	0
Leak 1	17	33	10	3	0	0	0	0
Leak 2	4	19	21	9	6	3	1	0
Leak 3	2	5	12	18	10	11	4	1
Leak 4	0	3	8	13	25	8	4	2
Leak 5	1	1	2	9	10	28	9	3
Leak 6	0	2	1	5	4	15	23	13
Leak 7	0	1	2	6	5	8	13	28

Discussion & Conclusion

- From the baseline model, we know that plume area is a good indicator to classify different leaks. In general, as the leak size goes up, the plume area increases.
- CNN model is better than the baseline model which simply calculates the plume area. CNN model can extract more spatial information from the images.
- In this problem, the temporal information of plume motion, which tells us how the plume changes over time is significantly useful as we see that LSTM based method works better than CNN model.
- The major drawback of FC-LSTM in handling spatiotemporal data is its usage of full connections in input-to-state and state-to-state transitions in which no spatial information is encoded. ConvLSTM makes use of both the spatial information from the full image and the temporal information.
- 2-class classification is much easier than 8-class classification.
- Wind orientation changed a lot when the data was recorded, which makes one leak look like a leak of different leak size when the wind blows the plume away. This adds noise to the dataset.
- 5 seconds is a good video length as it includes enough plume motion information and ensures that the test set has a good amount of data.

Future Work

- Collect more video footage and test the algorithm on different videos.
- Modify the architecture in order to decrease the false positive rate and false negative rate
- Instead of treating the problem as classification problem, due to the small range of the leak size in the dataset, solve the problem as a regression problem in order to treat the mislabeled data differently and increase the accuracy.
- Integrate the wind speed and orientation into the model.
- Use FEAST model (The Fugitive Emissions Abatement Simulation Toolkit) to estimate the cost and the emission reduction associated with our classification technology.

References

[1] R. W. Howarth, R. Santoro, and A. Ingraffea, "Methane and the greenhouse-gas footprint of natural gas from shale formations," Clim. Change, vol. 106, no. 4, pp. 679–690, 2011.
[2] Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. In Advances in neural information processing systems (pp. 568-576).
[3] Hochreiter, S., Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation.
[4] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
[5] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
[6] Hsu, Y. C., Dille, P., Sargent, R., & Nourbakhsh, I. (2016). Industrial Smoke Detection and Visualization. Technical Report Carnegie Mellon University-RI-TR-16--55. Robotics Institute, Pittsburgh, PA.
[7] Xingjian, S. H. I., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Advances in neural information processing systems (pp. 802-810).