Classification of Natural Gas Leaks
Project Category: Computer Vision

Jingfan Wang (SUNET: jingfan)
Sindhu Sreedhara (SUNET: sindhus)

1. Introduction

Natural gas plays an increasingly significant role in transforming the energy system towards a low-carbon path. However, there is a fierce debate over the use of natural gas in a low-carbon future. Upon combustion, natural gas emits half as much CO2 as coal per unit of primary energy. But leaks from the natural gas supply chain have high global warming impacts due to the release of methane, which has a global warming potential over 30 times larger than CO2 over 100 years. Moreover, natural gas leaks waste money, reduce energy availability, and affect local air quality.

Therefore, it is crucial to mitigate natural gas methane emissions by means of cost-effective reduction methods. In this project, we propose an interdisciplinary project that expands upon infrared camera (IR) imaging, an Environmental Protection Agency-approved leak detection technique, and will harness the potential for computer vision and deep learning advances to allow for rapid and automatic classification of methane leaks. This aims to solve multiple problems: labor costs for IR surveys are high, continuous monitoring with IR is infeasible, and IR surveys cannot tell the operator how big a leak is in the real time. There are few researchers working on this problem because: (1) deep learning hasn't been introduced to field of leak detection and quantification (2) a natural gas leak is not a rigid body and it is hard to classify it using only computer vision techniques because various weather and wind conditions affect the appearance of the plume in images.

2. Datasets

We conducted controlled-release experiments at the Methane Emissions Technology Evaluation Center (METEC) in Fort Collins from July 10 - July 14 2017. Jingfan Wang collected a total of 31 3-min-videos (nearly 697,598 frames in total) at two leak locations: separator on pad 1 (13 videos) and separator on pad 2 (18 videos). Each video was taken with a unique combination of distance and field of view. Within the 3-min-video, eight levels of leaks were recorded: 0 scfh (Class 0), 16 scfh (Class 1), 43 scfh (Class 2), 58 scfh (Class 3), 68 scfh (Class 4), 84 scfh (Class 5), 110 scfh (Class 6) and 125 scfh (Class 7) where scfh is a standard cubic foot per hour. The leak type is point-source leak from both two separators.

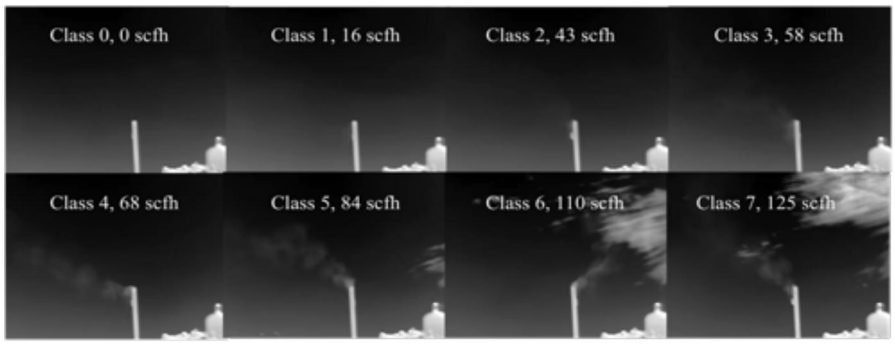

Fig 1. Representative frames of Video 2567 show the eight leaks of separator on pad 2. (1,1) Frame 00380, Class 0. (1,2) Frame 3044, Class 1. (1,3) Frame 05708, Class 2. (1,4) Frame 08470, Class 3. (2,1) Frame 11457, Class 4. (2,2) Frame 13826, Class 5. (2,3) Frame 17100, Class 6. (2,4) Frame 20190, Class 7.

3. Method

3.1 Image Preprocessing Method

We perform a background subtraction on the images in order to remove all non-plume details from the image (such as pipes and clouds). We implement this by calculating a moving average of the

210 images prior to the current image in the video and subtracting this average from the current image. The final step is to take an absolute value of all the pixel intensities to account for parts of the plume that may have been below the average intensity. The resulting image contains only plume pixels. The following images show heat maps of pixel intensities before and after background subtraction for two leaks: one from Class 4 and one from Class 7.

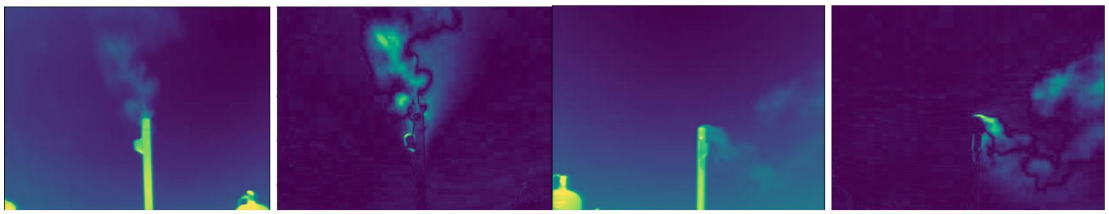

Fig 2. (Left to right) Heatmaps of class 4 leak before background subtraction, class 4 leak after background subtraction, class 7 leak before background subtraction, class 7 leak after background subtraction

3.2 Baseline Method

To get baseline test set accuracies, we first perform background subtraction on all frames in the training set. Then by observing the average pixel intensity distributions across all the training set frames, we choose a pixel intensity threshold of 30. With 30 as the threshold, we count the average number of pixels with intensity above the threshold for each leak class. Using these averages as breakpoints, we classify frames in the test set based on the number of pixels above the threshold in each frame.

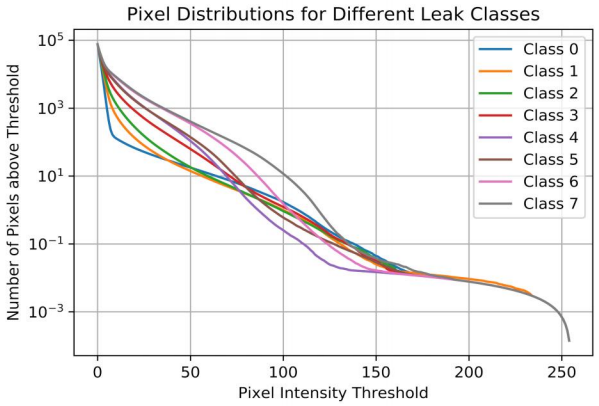

Fig 3. Cumulative distribution of pixel intensity for each leak class

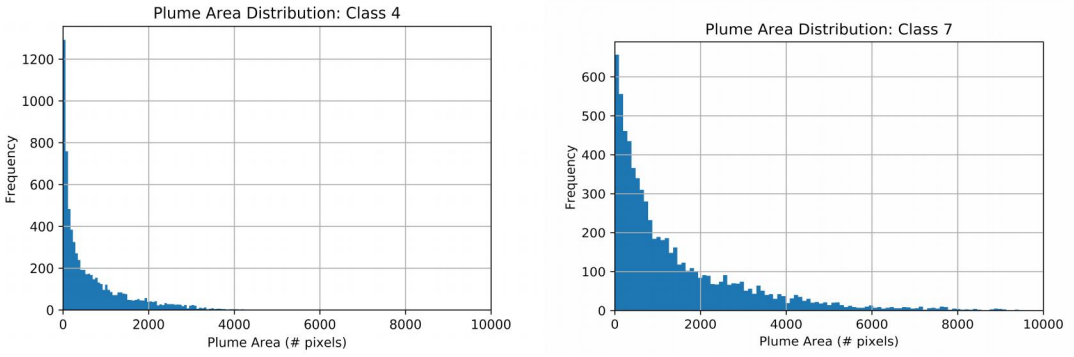

Fig 4. Cumulative distribution of pixel intensity for each leak class

3.3 Deep Learning Method

In this project, we will perform three deep learning algorithms. Firstly, we develop our own CNN model which has two Conv-Pool structures and two fully connected layers. This model attempts to classify the leak based on images after background subtraction. Secondly, because the temporal information of the plume is necessary to be used in the model, FC-LSTM is built, which uses the output of the fully-connected layer 1 in the CNN model as input for the LSTM structure. Thirdly, in order to make use of the full image, which has the greatest amount of the spatial information, ConvLSTM is developed to have the image to be the input of the LSTM structure. All the parameters inside the LSTM, such as gate and cell output, all possess spatial information.

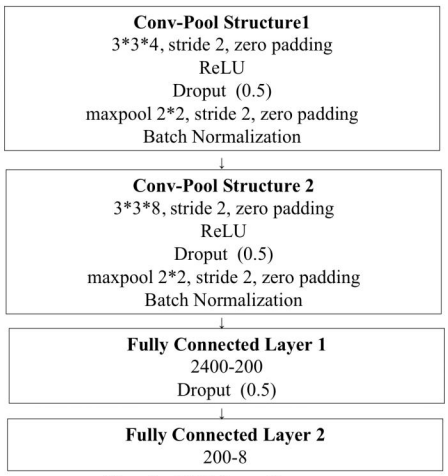

Fig 5. CNN on frames

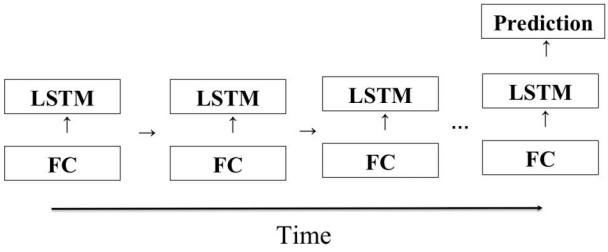

Fig 6. FC+LSTM on Videos

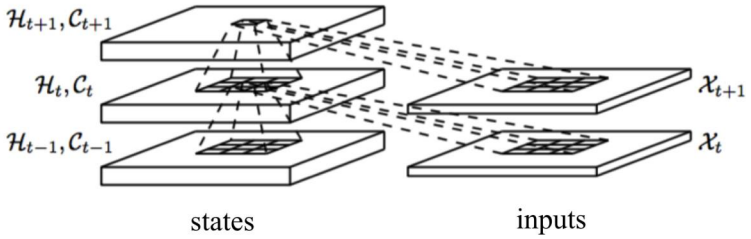

Fig 7. ConvLSTM on Videos

3.4 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). Model 1 is works on frames. For the frame dataset, the number of frames in training set, validation set and test set are 45474, 11368 and 37864 respectively. Model 2 and Model 3 deals with video input. If frame rate is chosen to be 5 frames/second, video contains 5 seconds of images, and number of frames in the training set is 606 while that in the validation set is 151 and test set is 504.

4. Results

In the section 3, we introduced the baseline model and three deep learning models. In this section, we will compare the four models.

4.1 Results of 8-Class Classification Problem

Firstly, we want to classify 8 leak classes as stated in the section of dataset. Table 1 illustrates the accuracy results. Random guess accuracy, the probability of correct classification by a random guess, is 12.5% (i.e. 1/8). Increase rate is calculated based on the random guess accuracy. It is clear that CNN model is better than the baseline model, LSTM based models have better performance and ConvLSTM is the best model in this problem.

Table 1. The results of 8-class classification problem

No.	Model	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%

We also try to change the hyperparameters in the three deep learning models. For model 1, we add different number of dropout layers. For model 2, number of neurons in the fully connected layer changes. For model 3, the video length ranges from 2 seconds to 10 second. We find that the three models work the best in the following cases, where dropout layer number is 3, neuron number in fully connected layer is 200, video length is 5 seconds, respectively.

Table 2. 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%

The confusion matrix is generated based on the best accuracy (ConvLSTM) in this problem. We infer that highest score in each class is in the diagonal of the matrix. A lot of the leaks are misclassified as smaller leak sizes.

Table 3. Confusion Matrix (Y-Lable, X-Prediction)

	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

4.2 Results of 2-Class Classification Problem

Secondly, we treat the problem as 2-class classification problem and group leak sizes 0,1,2,3 to be new leak size 0, and leak sizes 4,5,6,7 to be new leak size 1. Table 4 lists results similar to those listed in the 8-class classification problem.

Table 4. The results of 8-class classification problem

No.	Model	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%

5. 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, which includes the plume area, the plume intensity change, the concentration of the plume and so on.

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 disadvantage of FC-LSTM is that the full connections in input-to-state and state-to-state transitions don't possess enough spatial information. ConvLSTM makes use of both the spatial information from the full image and the temporal information. This is why ConvLSTM works the best in both 2-class classification problem and 8-class classification problem.

Apparently 2-class classification problem is much easier than 8-class classification problem. In practice, we may not need to classify all the leaks. It is more important to fix the larger leaks which contribute to the majority of natural gas leaks. In this case, the 2-class classification problem is meaningful to explore and the result of the problem can guide real-life practice.

Wind orientation changed a lot when the data was recorded, which often 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. If the video length is 10 seconds, the total size of training and test is not enough to train a good model. If the video length is 2 seconds, the video doesn't have enough temporal information inside.

6. Future Work

We will continue to enlarge the dataset so that it can represent the diversity of leaks that are observed in the real world and make the algorithm be more generalizable to unseen leaks.

A new metric needs to be found to penalise the cases where a leak is misclassified to other classes and treat the mislabeled data differently. It is apparent that the case where leak size 2 is classified as leak size 7 is worse than the case where leak size 2 is classified as leak size 3.

The architecture needs to be improved to make full use of the spatial and temporal information in order decrease the false positive rate and false negative rate.

It is worth trying to integrate the wind speed and orientation into the model.

We can 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] T. M. L. Wigley, "Coal to gas: The influence of methane leakage," Clim. Change, vol. 108, no. 3, pp. 601–608, 2011.
[3] R. A. Alvarez, S. W. Pacala, J. J. Winebrake, W. L. Chameides, and S. P. Hamburg, "Greater focus needed on methane leakage from natural gas infrastructure," Proc. Natl. Acad. Sci., vol. 109, no. 17, pp. 6435–6440, 2012.
[4] A. Burnham, J. Han, C. E. Clark, M. Wang, J. B. Dunn, and I. Palou-Rivera, "Life-cycle greenhouse gas emissions of shale gas, natural gas, coal, and petroleum," Environ. Sci. Technol., vol. 46, no. 2, pp. 619–627, 2012.
[5] R. B. Jackson et al., "Natural gas pipeline leaks across Washington, DC," Environ. Sci. Technol., vol. 48, no. 3, pp. 2051–2058, 2014.
[6] United States Environmental Protection Agency, "Understanding Global Warming Potentials." .
[7] Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. In Advances in neural information processing systems (pp. 568-576).
[8] Hochreiter, S., Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation.
[9] 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.
[10] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
[11] 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.
[12] 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).