



Methane Leak Quantification on Edge Devices Using Deep Learning

Abhinav Kohar^a, Prapti Miglani^{b*}, Ramakant Yadav^c, Anuj Shetty^d

^{a,b}SLB, 5599 San Felipe St, Houston, TX 77056, United States of America

^cSLB, III Floor, Tidel Park, Coimbatore, TN 641014, India

^dWindBorne Systems, 910 San Antonio Rd, Palo Alto, CA 94303, United States of America

^aEmail:fabhinav@slb.com,^bEmail:pmiglani@slb.com

^cEmail:ryadav32@slb.com,^dEmail:anuj@windbornesystems.com

Abstract

Oil and gas domain deals with a varied set of problems ranging from methane leaks to prognostics and health management. This study demonstrates a solution to identify whether there is a methane leak or not and classify on what level the leak occurs based on different flow rates (5.3 g/h – 2051.6 g/h) at 5 distances (4.6 m - 15.6 m), using various spatial and temporal preprocessing techniques and deep learning models. For this study, we are using the GasVid methane leak dataset which consists of videos taken from infrared cameras on 2 different separators with a frame rate of 15 frames per second. Firstly, we applied a series of preprocessing steps, including contrast enhancement, Gaussian blur, three different background subtraction methods, namely moving average background subtraction method, KNN Gaussian mixture model, and Gunnar Farneback optical flow analysis. To save computational resources, temporal down-sampling was applied on the video frames. Thereafter, experiments were conducted using the 3D CNN model by modifying hyperparameters. It was found that on comparing the Adam and Lion (EvoLved Sign Momentum) optimizers, the Lion Optimizer increased accuracy by more than 34% and achieved state-of-the-art accuracy of 41.36% on 4.6 m videos. Further, the moving average background subtraction method's performance surpassed other background subtraction techniques. In addition, applying spatial preprocessing and down-sampling raw videos were compressed from 2.3GB to ~200MB which is a reduction factor of 11.

Keywords: methane emissions; deep learning; artificial intelligence; spatial and temporal preprocessing.

Received: 8/30/2025

Accepted: 10/30/2025

Published: 11/8/2025

* Corresponding author.

1. Introduction

Different greenhouse gases and their emissions contribute to global warming and understanding the level of impact they can have will significantly change the measures employed to address climate change. Although carbon dioxide (CO₂) with its centuries-long persistence and huge quantities is a primary contributor. Methane, though less abundant, is responsible for 17.3% of warming on a century-long timeline. However, reducing methane emissions will have more immediate benefits for climate change owing to its 20-year lifespan and quicker effect in reducing global warming rather than efforts aimed at CO₂.

Approximately 33% of methane emissions in the United States are from oil and gas systems. Natural gas is endorsed as an environmentally friendly alternative to coal because it reduces greenhouse gas emitted on burning by half, however, a more nuanced picture emerges on considering full lifecycle emissions. Natural gas operations occur in several phases such as extraction, processing and distribution, all of which may have methane leakage. With a minimal leakage rate of 0.2% the climate impact of natural gas may supersede coal, thus casting doubts on its clean energy status (Wang and his colleagues [2]). Estimates suggest that actual rates may be substantially higher, 1.4% as per EPA and 11% by other studies. Furthermore, the top 5% of major leaks account for nearly half the emissions, which highlights the urgency for targeted detection and intervention.

Our research proposes developing an autonomous, edge-based methane leak detection and monitoring system for oil and gas systems. Using infrared cameras in place of expensive sensors can help deploy such a system across thousands of sites in the US. This can also combat logistic barriers prevalent at remote, bandwidth-limited operation sites. The EPA and DoE announced \$1 billion to reduce methane emissions from the oil and gas sector, along with fining excess emissions from certain facilities, providing significant financial incentives for such emission monitoring systems.

2. Materials and methods

The GasVid dataset proposed by Wang and his colleagues [1] has 31 recordings nearly 24 minutes each amounting to 669,600 frames in total. A wide range of leaks from 5.3 gCH₄/h to 2051.6 gCH₄/h at distances between 4.6 m to 15.6 m are covered using FLIR GF-320 infrared camera enhanced by including different ambient variables. The videos were collected at Colorado State University's METEC laboratory, home to over 50 metered leak sources, offers lab-controlled settings to reproduce actual gas leaks. To maintain uniformity for model validation, the first 15 and last 5 seconds of each recording are omitted. This research study utilizes convolutional neural networks (CNNs), optical gas imaging (OGI) and machine vision captured with infrared cameras to detect methane emissions from natural gas. GasNet, the CNN model built by training, validating and testing on the GasVid dataset, outperforms the baseline change detection system based on Gunnar Farneback optical-flow algorithm. Other background subtraction methods like fixed background subtraction, moving average background subtraction and Mixture of Gaussian-based (MOG) background subtraction were assessed in parallel leading to the conclusion that moving average background subtraction surpasses the rest in methane plume isolation. It was also observed that for minor leaks, accuracy is inversely proportional to distance. The GasNet models had detection accuracies between 95% and 99% for leaks of all sizes and at varying distances.

Notably, binary detection accuracy for major leaks at smaller distances surpassed 97% and the GasNet-2 model had optimal performance across varying distances.

VideoGasNet [2] by Wang and his colleagues uses their previously introduced dataset GasVid and further subdivides each video into 3-minute sections comprising 2700 frames to minimize the cost of computation. Variable duration datasets with 15 to 100 frames are generated from the video subsections using a sliding window approach. Image preprocessing methodologies such as moving average background subtraction and pixel value normalization are applied on the input dataset along with analysis of different segmentation lengths to maintain consistent training dataset sizes and improve model accuracy. On comparing 2D CNN, 3D CNN, and ConvLSTM models, 3D CNN provided the best performance and is referred as VideoGasNet. On applying the model for binary detection, its accuracy is about 100%, for multi-class leak classification for three-class categorization it achieved 78.2% accuracy and for the complex eight-class categorization, 39.1%. These results show considerable improvement over random chance accuracy of 12.5%.

Ravikumar and his colleagues [3] developed an open-source model that simulates passive infrared (IR) imaging in optical gas imaging. They conducted experiments near Sacramento, California to record 800 methane frames using a FLIR GF320 camera at rates between 1 g/s and 29 g/s and distances between 10 to 60 meters. To avoid strong winds or sudden direction changes, 50 frames were selected from each for model validation. For image processing Lucas-Kanade optical flow technique and direct image enhancement using filtering and thresholding were used. Accuracy was determined using pixels of plume in a frame and comparing it against the simulation results. The greatest factor influencing accuracy of detection was identified to be distance, at 10 meters 80% emission was detected. Another factor is the background of the image, aerial observation is less efficient in contrast to land-based detection against the sky or low-emission backgrounds. Temperature of the scenes does not impact efficiency of OGI-based leak detection. The minimum detectable leak rate (MDLR) for distances from 10 to 200 meters is increased nearly 25 times. It is observed that detection efficiency is enhanced by super emitters, therefore the focus should be on larger breaches to mitigate the effects of methane emissions.

Another deep learning framework to detect point-source methane emissions is introduced by Rouet-Leduc & Hulbert [9]. It uses a Vision Transformer (ViT) encoder combined with a U-Net decoder architecture for emission detection in multi-spectral satellite images (Sentinel-2). Synthetic Gaussian plumes at different emission rates and wind conditions were embedded into Sentinel-2 scenes that were collected over varied land covers, topographies as well as climates. The input consists of two time-steps ($t-1$ and t) for ten Sentinel-2 bands (resampled to 20 m where needed). By comparing bands over a span of time transient signatures can be detected. On synthetic test data, the model achieves high F1-scores (harmonic mean of precision and recall) down to signal-to-noise ratios of ~5%, substantially outperforming the state-of-the-art Multi-Band Multi-Pass (MBMP) thresholding methods used on the same data. When applied to real methane plumes cataloged by airborne surveys (Carbon Mapper), the model reliably detects plumes of approximately 0.01 km² area, corresponding to emission rates of 200-300 kg methane/h under typical wind conditions. False positive rates are low: e.g. <0.03% at the pixel level for an ~85% true positive rate at that threshold. The authors discuss that detections decline sharply for much smaller leak rates/plume extents below those thresholds.

Daugela and his colleagues [10] to study methane emissions in Lithuania from the Kariotiskes landfill. The UADS integrated GIS-derived products with multispectral imaging (RGB, near-infrared, thermal infrared) and collected data between 2015-2018 flying DJI Matrice 200 UAVs with Zenmuse X4S and Trimble UX5 with Sony NEX-5R sensors. Digital elevation models (DEMs), NDVI maps and high spatial resolution orthoimages developed were used to determine areas of localized methane emissions, Further, a lightweight gas sensor array and thermal infrared imagery of the identified areas enabled mapping of surface temperature anomalies and validate methane presence. The approach enables rapid localization of emission hotspots and monitoring of landfill topography with minimal structural changes observed over time, though methane influence is inferred rather than quantified precisely for many areas.

A number of optical techniques namely tunable diode laser spectroscopy (TDLS), lidar, cavity-enhanced absorption spectroscopy (CEAS) for methane detection were reviewed by Kwasny & Bombalska [11]. These included various domains of application such as atmospheric monitoring, industrial leak detection, etc. and ranged from low-ppm/ppb levels to nearly 100% concentrations. Selectivity of absorption lines, field deployability, interference of absorption bands from CO₂ and water vapor, cost, trade-offs between compactness and sensitivity refined by employing high-quality cavities or long optical path lengths emerged as key areas of concern. They also present Designs of laser methane analyzers (e.g. differential absorption lidar, TDLS, NIR instruments) were presented and their performance parameters such as detection limits, optical path lengths, and spectral regions used were compared. In summary, though remote sensing and laboratory methods have achieved very low detection limits, practical systems still struggle with robustness, cost, and interference in “real world” outdoor conditions.

During the California Methane (COMEX) campaign, Airborne Visible/Infrared Imaging Spectrometer-Next Generation (AVIRIS-NG) were used by Thompson and his colleagues [12] for methane detection in real-time. A matched-filter algorithm was used to extract methane absorption features from the shortwave infrared region from hyperspectral data streams. For plume concentrations 1% above surroundings, column enhancements are estimated at noise-equivalent levels of ~141 ppm·m which retained sensitivity while reducing computational cost. Concurrent in-situ sensors and non-imaging spectroscopic measurements were validated which maintained cross-platform consistency. Not only does this support adaptive sampling strategies but also allows for real time redirection of airborne measurements toward active emission hotspots. As one of the first demonstrations of real-time for airborne hyperspectral methane detection, its results show reliable plume localization irrespective of atmospheric conditions.

2.1. Experimental Setup/Dataset

This study has been performed on the GasVid dataset proposed by Wang and his colleagues [1] created at METEC, Colorado State University in a controlled setting recorded between July 10 to July 14, 2017. It has 31 videos, each 24 minutes long with a total of over a million frames with recordings from production separators and tanks using a FLIR GF-320 IR camera. The METEC facility simulated varied real gas leaks from over 50 metered leak sources at varying distances and size of leaks ranging between 5.3 to 2051.6 g CH₄/h. Since it was a controlled setting external factors such as moving vegetation and steam were excluded. Although there are

challenges like environmental factors(wind) and low video resolution, the dataset helps in developing and testing methane leak detection systems. Of the 31 videos, 5 videos were excluded as they were at a distance of 18.6m which is too far for an operator to detect leaks, 1 video at a distance of 8.8m was excluded as it would result in imbalance of the dataset. Among the 25 videos being considered, 10 were recorded at separator 1 and the rest on separator 2 at different times of the day and varying distances. Separator 2 videos were used for training and Separator 1 videos were used for validation to ensure broad applicability and robustness of the model.



Figure 1: Training focus on 4.6-meter proximity

To set up a baseline for leak detection under controlled conditions, videos recorded at 4.6m which is closest is taken into consideration. For validation, videos taken at longer distances like 6.9 and 9.8 meters to account for leaks differing in distance and visual appearance due to atmospheric conditions and camera angles to simulate real-world scenarios. However, the GasVid dataset does have some shortcomings. It does not simulate large leaks, since it uses standardized 5mm tubing, which are rare but important due to their excessive contribution to total methane emissions. It does not include real-world interference from moving objects, water vapor or steam which are important obstacles to consider while detecting leaks. Another critical shortcoming is the uncontrolled wind factor as wind speed and direction massively change the appearance of methane plumes reinforcing the requirement for another dataset that can address these concerns.



Figure 2: Testing focus on 6.9-meter proximity

The test dataset consists of methane leaks recorded from a raised vent having 2.5cm diameter but seven varying rate of emissions between 1 g/s (equivalent to 5 mscf/day) to around 29 g/s (135 mscf/day). The leaks were

captured at distances between 10 to 60 meters using a FLIR GF320 camera near Sacramento, CA. The videos were captured at 16 fps resulting in nearly 800 frames for each flow rate. To extract high-quality uncompressed videos, the ResearchIR software was utilized. The magnitude of the leak changed the plume's velocity, both horizontally and vertically and the plume's dispersion was affected by surrounding winds and vertical velocity. To make the test dataset more complex, several environmental components such as airplane contrails in the distance (observed in specific video segments such 7 as LeakVid71) and airborne insects were included. Further, the plume source was intentionally placed at different places (corner of video frame) to ensure model robustness.

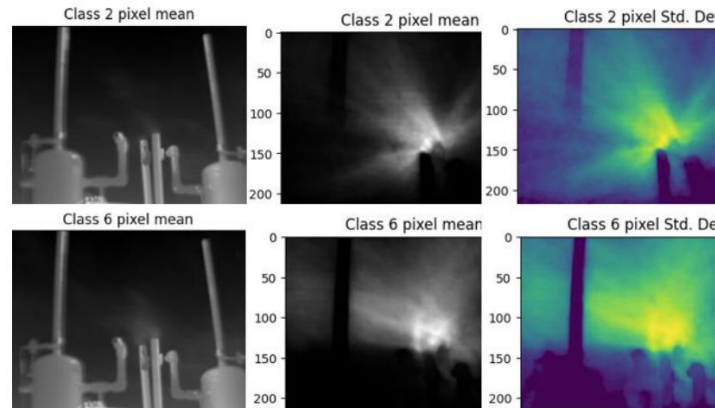


Figure 3: Comparing pixel, grayscale pixel mean, and pixel standard deviation of Class 2 and 6 images

Figure 3 contrasts methane leakage for two classes: Class 2 (lower methane leak) and Class 6 (greater methane leak). Although the average pixel images seem visually comparable, the grayscale pixel mean images clearly reveal Class 6 with a brighter and more extensive emission source, suggesting higher methane leakage. While Class 2 shows a smaller, more focused emission pattern with less variability, the pixel standard deviation images highlight this difference by showing more variability and dispersion for Class 6 commensurate with larger leak intensity.

2.2 Preprocessing

2.2.1 Spatial preprocessing

Contrast enhancement: A series of contrast enhancement techniques were applied to the input frames beginning with top-hat transformation, followed by contrast limited adaptive equalization (CLAHE). As the input frames are gray scale and represent white methane plumes against a dark background, we utilized top-hat filter with a kernel size of 15, which extracts brighter and lighter objects of interest contrasting with a dark setting. It enhances the minor details of the input image and extracts subtle variations in pixels in front of a dim backdrop. Following that, we executed CLAHE (Contrast Limited Adaptive Histogram Equalization) to improve the contrast of the input frames while managing the over-amplification of noise. We used the default tile size of 8x8 and applied histogram equalization to smaller independent parts of the images to capture local contrast.

Gaussian blur: In gaussian blur, every pixel is substituted by the weighted average of the neighboring pixel values which smooths the image by reducing noise. In this paper, we used a kernel size of 5, which defines the dimensions of the square for the calculation of the mean of the weights of the surrounding pixels.

Background subtraction: In this section, we will introduce 2 different background subtraction methods to detect dynamically moving entities. For the dataset, we need to extract foreground masks of dynamically moving objects against the static dark background. We adopted the moving average background subtraction method by Wang and his colleagues [1], as it has demonstrated superior performance against the fixed background subtraction and Mixture of Gaussians-based (MOG) background subtraction. Also, we implemented KNN Mixture of Gaussians-based method and Gunnar Farneback optical flow algorithm that extracts foreground mask by subtracting current frame and background model.

- **Moving average:** Wang and his colleagues [1] proposed this approach where a background model is created by a series of initial frames by taking the running average of pixels of the previously processed frames and the current frame. Subsequently, the absolute difference between the present frame and the background model is computed to detect the new entities. For this paper, we have used a learning rate of 0.01 which determines the impact of the current frame while calculating the background model. An accuracy threshold parameter of 30 is chosen for this research paper, which is important to contrast stationary, dark backdrop from foreground entities.

- **Optical flow analysis:** Gunnar Farneback A dense optical flow algorithm introduced by Gunnar Farneback is applied to detect the relative motion of objects between two subsequent frames by computing intensity gradient of vectors along both horizontal and vertical axis. In our study, we used a threshold value of 5 for optical flow estimation, defining the minimum numerical value of the flow of objects in subsequent frames to filter out small and large motions of objects.

Binarization: To identify specific features and extract foreground elements from the background, we converted the grayscale images to binary images, where each pixel denotes either white (1) or black (0) color. White and black are assigned based on pixel intensity values. If the value is above a pre-determined threshold, it is assigned white denoting foreground entity, otherwise it is assigned black denoting a background element. In our study, we experimented with two threshold values, i.e. 5 and 20 to identify the region of interest and the magnitude of noise we intend to keep in the preprocessing steps.

2.2.2 Temporal Segmentation

Down-sampling: The dataset consists of 25 videos of approximately 24 minutes each with a frame rate of 15 frames per second of 67 MB. In this research, we applied a down sampling technique, which reduces the number of frames in a video by eliminating selected frames while maintaining essential visual information for efficient storage and video compression. Here, we deployed down sampling technique with a factor of 0.5, which removes every alternate frame from the original video to create a new video by reducing the frame rate by half. After completing the preprocessing and down sampling steps the final size was reduced to 24 MB, signifying a significant decrease of 65%.

Sliding window: After performing preprocessing steps and down sampling, the 24-minute videos are reduced to ~12 minutes set, which brings it down to ~1400 frames per video. Due to computational limitations while training deep learning models (Kohar and his colleagues [7]), we further break down a 1.5-minute video to multiple videos of 15 frames each using sliding window of size 15 (Wang and his colleagues [1]) and 5 frame strides, thus reducing parameters fed into the model. A 5-size stride is indicative of number of frames skipped to interpret the next window and 15 size window indicates the number of consecutive video frames taken into consideration while performing processing.

2.2.3 Data augmentation

As methane gas leak videos could be affected by factors like camera angles, distances, lighting condition and other external environmental interference, it is pertinent to simulate these variations by augmenting the data, making the model more robust to real-world scenarios. In this work, we use 10 data augmentation techniques like uniform zoom, asymmetric zoom, combined affine transformations (translation, scaling, rotation), 90° angle rotation, 45° angle rotation, random rotation, horizontal flipping, vertical flipping, diagonal flipping and sequential augmentation of the raw data described below.

Zoom: We conduct various zooming parameters, in combination with flipping and rotation of the data. We find that zooming in or scaling doesn't expose the model to variations in the data set as the training and test datasets already cover a similar range of scales, leading to redundant information.

Flip: In the GasVid dataset, plumes appear on either side of the separator because of the wind direction. Thus, images have been flipped horizontally to diminish methane plumes' horizontal alignment which increases model robustness and ensures invariance to wind among other environmental factors.

Rotate: To reproduce real-world situations, images are rotated to account for multiple orientations of the object and camera. Our study uses 20% rotation, which implies random rotation by an angle of +72 degree to -72 degree.



Figure 4: Data Augmentation Techniques for Methane Gas Leak Videos

2.2.4 Deep learning architectures

In this section, the methodology of deep learning frameworks used is detailed, including the selection criteria of hyper-parameters used. As processing all data at once requires more computational and memory resources, we

trained all data into batches to enable parallel processing of data.

3D CNN model: We use the modification of 3D convolution structure presented by Wang and his colleagues Reference [2] comprising of three units: 3 convolutional units, single dense layer and an output dense layer with 8 units. First, the input data is rescaled from a range of [0,255] to [-1, 1] for better model performance. In the convolutional structure each convolutional layer has increasing kernel sizes (4,8,16), coupled with a batch normalization layer, ReLU activation layer, a dropout layer 11 for regularization and a max pooling layer. Additionally, it has a flattening layer plus a 256 unit dense layer and an 8-unit (leak0 - leak7) softmax output layer for multi-classification. To prevent overfitting while training the dropout layer is set at 0.25 and sets a subsection of the input for each iteration to be 0. Each layer's weight is initialized based on the previous layer using the He Normal initializer to converge to the global minimum of the loss function fast.

ResNet model: First the 2D images are augmented by horizontal flipping and a rotation between +72 deg and -72 deg after which it is fed into the ResNet-18 model with 18 layers and a “skip connection”, which means that the input given to a layer can be fed to the output. This solves the vanishing gradient problem while backpropagating. The residual network's output is flattened and processed by a dense layer with He normalization, L2 regularization and ReLU activation. Two more layers for regularization, batch normalization and a dropout layer followed by a final dense layer having softmax activation and 8 units used for multi-class problems form the complete model.

Octave Convolution model: An image could be divided into 2 categories based on spatial frequency: low spatial frequency component, which captures gradual variations in structure and high spatial frequency components, which illustrate fine detailed structural changes. We use Octave Convolution for the division of low- and high-level spatial frequency. Our Octave convolution model is divided into 3 building blocks: an initial block, which divides the input data into low and high frequency components, the core block for processing both frequency components and finally the concluding component to combine the high and low frequency segments. High-level features are processed like traditional convolution neural networks, while low frequency units are sampled and processed separately, reducing spatial resolution, thus leading to a reduction of computational resources.

2.2.5 Hyper-parameter tuning optimizer

This research study uses two optimizers namely Adam optimization which has stochastic gradient descent adaptive estimation method based on calculating the first and second order moments. The other one is Lion which uses the sign function to estimate only the 12 first-order momentum and is also a memory-efficient optimizer. It is observed in Chen and his colleagues (2024) that with larger batch sizes Lion's performance improves gradually. This paper utilizes a batch size of 64 for Lion and 8 for Adam.

Batch size: The training batch size depends on the rate of convergence, duration of each epoch, and computational efficiency. Although small batch sizes are less memory-intensive they have a noisy gradient and large batch sizes provide more accurate gradient calculations. Experimenting with a wide range of batch sizes

(8, 16, 32, 64, 128, and 256) it was determined that the Adam optimizer performs better with smaller batch sizes (8) while Lion provided superior results for large batch sizes (64) especially while training the Octave Convolution.

Learning rate: Learning rate is a critical hyper parameter which governs the step size to adjust model's weights during each iteration to minimize the loss function. While a higher learning rate often leads to overshooting the minima of the loss function, a smaller learning rate takes longer to converge and might get stuck into local minima. In this study, we use a smaller learning rate of $1e-4$ for Lion optimizer as it results in larger norm due to sign operator used in calculating the momentum. For Adam optimizer, we use a high learning rate of $1e-3$. Both these optimizers Lion and Adam have inherent adaptive learning rate capabilities.

Weight initialization He: Random weight assignment is done based on He norm initialization with the range of the initialization governed by incoming connections from previous layers. To get a more effective gradient descent, a Gaussian distribution is used to choose weights with standard deviation depending on previous layer sizes and mean of 0.

2.2.6 Student-teacher distillation

Our research delves deep into understanding knowledge distillation and its application in Octave convolution. Our study uses the Octave convolution model as a 'teacher' model. After training the teacher model, we utilize its acquired weights to transfer its knowledge to a 'student' model that has half the number of filters. In addition, we started the training of the teacher model from its initial state to establish a standard method for comparison. A distillation loss function is used to measure the difference between the softened predictions made by the student model and the softened labels provided by the teacher model, built at a specific temperature. An alpha coefficient for balancing student and distillation losses. During the computation loss step, we coordinate a forward pass of both the teacher and the student models. We then determine the loss by multiplying the student loss and distillation loss by alpha and $(1 - \alpha)$, respectively. In our studies, we applied a value of 0.5 to the alpha parameter and a value of 2 to the temperature parameter. We perform training on the base Octave CNN model. The teacher model had about 4,920,124 parameters, equivalent to almost 19MB. The student model on the other hand, has a reduced complexity, with 1,230,906 parameters, occupying nearly 4MB of space. On multiplying the student model's size by two to take into consideration 4, 8, 16, and 128 filters, its total parameters went up to 2,460,860, equivalent to 9.39 MB. This demonstrates a large decrease in size compared to the teacher model, while still maintaining substantial power for learning. By using half the filters, the student model achieved a validation accuracy consistent with the teacher model's accuracy. This student model is deployed on Agora edge devices for methane detection and methane leak prediction.

3. Constraints and Limitations

Despite the controlled design and careful methodology, this study faces several important constraints that limit the scope and generalizability of its findings. These are discussed below in terms of dataset design, preprocessing and modeling, and eventual real-world deployment.

3.1 Dataset Constraints

3.1.1 Restricted Leak Types and Sizes

The GasVid dataset is constructed using standardized 5 mm tubing, which only allows for small to medium-scale leaks to be simulated. While these leaks are important for calibration and controlled experimentation, they fail to capture the large or catastrophic leak events that, although less frequent, account for a disproportionate share of global methane emissions. As a result, the dataset underrepresents scenarios that are most impactful for climate monitoring and industrial safety.

3.1.2 Controlled Environment Bias

Among the videos collected at METEC, Colorado State University's artificial testbed, videos with environmental disruptions were deliberately omitted. This meant real-world operational complexities prevalent in oil and gas fields or industrial facilities were not considered. Consequently, GasVid has data consistency “clean” laboratory-like conditions but might result in model overfitting.

3.1.3 Distance and Angle Limitations

In the training dataset, proximity (4.6 m) recordings are used while validation is at intermediate distances (6.9 - 9.8 m), which does not cover the entire operational extent sensors need to cover. Although plume appearance variations are simulated, in the real world, leaks may need to be detected from tens to hundreds of meters away or at unfavorable camera angles. Thus, the scalability of the trained models in real industrial inspection scenarios is constrained.

3.2 Preprocessing and Modeling Constraints

3.2.1 Augmentation Coverage Gaps

Although zooming, rotation, flipping, introduce several variations in orientation of the plume and wind direction, certain convoluted environmental conditions could not be replicated. Fog, rain, snow, dust, or changes in thermal background were not taken into consideration which are likely to introduce ambiguity to methane plumes in practice. Their absence could result in poor generalization by the trained models in adverse field conditions.

3.2.2 Down-Sampling Trade-offs

Frames were down sampled by a factor of 0.5 which cut the frame count nearly in half. This results in eliminating temporal dynamics such as faint turbulence patterns in the plume motion. While the computational overhead is reduced, thereby improving training efficiency, it may be critical for accurate leak classification.

3.3 Generalizability and Deployment Constraints

3.3.1 Lack of Multi-Plume Scenarios

The Sacramento dataset and GasVid simulate single-source plumes, however, real-world scenarios have numerous leaks with overlaying plumes. Although methane leak detection and its classification become easier, the model's relevance to scenarios where the leak source is not isolated such as industries becomes limited.

3.3.2 Edge Dependence and Transferability

While the FLIR GF-320 camera is an extensively used infrared sensor, its availability and suitability for all deployments may not occur. This constrains the universality of the suggested methods since portability to different camera types offering various resolutions, sensitivities, or spectral ranges was not tested.

4. Results and Discussion

4.1 Comparing model architectures

Table 1: The runs were conducted on a dataset of 4.6 million videos, utilizing preprocessing techniques including contrast enhancement, Gaussian blur, moving average background subtraction, down-sampling, and binarization

Model	Training accuracy	Validation accuracy
OctConv	37.69%	41.36%
ResNet	52.84%	35.05%
3D CNN	58.75%	37.62%

Table 1 depicts that Octave Convolution (OctConv) has the highest validation accuracy of 41.36%. The 3D CNN showed strong overfitting indicated by a higher training accuracy (58.75%) and a sharp drop in its validation accuracy 37.62%. ResNet shows poor generalization (35.05%), despite having decent training accuracy (52.84%).

Low frequency plumes appear as smooth, slowly varying structures while sharper edges and boundaries signify high frequencies. OctConv separates and processes high and low-frequency components separately while also being computationally lighter than conventional CNNs or frame-based classification models making it the most accurate and practical choice for methane plume detection in real time.

4.2 Comparing background subtraction methods

Table 2 depicts the accuracies of moving average background subtraction, Gunnar-Farneback optical flow and KNN-based subtraction. Moving average background subtraction performs substantially better than the other methods, nearly 6% which further strengthens the belief that, for static backgrounds, simple temporal averaging

techniques perform better.

Uniform, low-texture infrared backgrounds prove difficult for KNN-based background subtraction while motion-based methods like Gunnar-Farneback optical flow (validation 28.57%) are not effective for optical flow resulting in incorrect detections for subtle plume motion due to its sensitivity to small intensity variations and noise. Therefore, moving average is the clear choice for its computational efficiency and robustness.

Table 2: The runs were conducted on a dataset of 4.6 million videos, utilizing preprocessing techniques including contrast enhancement, Gaussian blur, moving average background subtraction, down-sampling, and binarization

Method	Training accuracy	Validation accuracy
Moving average	37.69%	41.36%
Gunnar-Farneback	58.52%	28.57%
K-nearest neighbours	45.21%	34.79%

4.3 Comparing training and validation datasets with varying distances

Table 3: All runs were conducted with preprocessing techniques including contrast enhancement, Gaussian blur, moving average background subtraction, down-sampling, and binarization

Training dataset	Validation dataset	Validation accuracy
4.6m	4.6m	41.36%
	6.9m	27.4%
	6.9m(cropped to 4.6m)	31.18%
	9.8m	24.18%
	9.8m(cropped to 4.6m)	28.09%
6.9m	6.9m	32.89%
9.8m	9.8m	24.16%

Table 3 shows accuracies for different distances, where closer distances resulted in better accuracies. At 4.6 m the models yield the best accuracy (41.36%), and performance gradually declines as validation distance is increased to 6.9 m and 9.8 m. This leads to the conclusion that distance directly impacts plume detection. This indicates that plume detectability decreases as distance increases, owing to atmospheric dispersion and reduced infrared contrast.

Camera positioning and calibration prove to be critical, as observed by an improvement in accuracy on cropping and resizing validation data to emulate closer distances(e.g., 27.4% $\rightarrow$ 31.18% for 6.9 m). Pre-deployment calibration along with software-based adjustments can significantly boost performance when close distance data is unattainable.

4.4 Comparing optimizers

Table 4: All runs were done on 4.6m videos, with preprocessing methods of contrast enhancement, Gaussian blur, moving average background subtraction, down sampling and binarization, using an OctConv model

Optimizer	Training accuracy	Validation accuracy
Adam	38.67%	39.45%
Lion	37.69%	41.36%

Table 4 shows that the Lion optimizer shows less overfitting and slightly better validation accuracy (41.36%) while Adam (39.45%) tends to overfit small datasets showing higher training accuracy but worse generalization. Usage of sign-based momentum estimation by Lion improves regularization, making it more suitable for larger batch training in this study.

4.5 Comparing different augmentation parameters

We experimented with different sets of augmentation parameters in the training dataset to find the optimal values that will lead to highest validation accuracy. We started with different combinations of rotation and zoom values.

4.5.1 Changing rotation with fixed zoom level

Table 5 shows the accuracy on rotation to simulate wind direction and turbulence on the plume. Moderate rotation (0.3) provided the best validation accuracy (41.36%) however, excessive rotation (>0.3) degraded accuracy as it introduced noise in place of useful variability.

Table 5: All runs used an OctConv model on 4.6m videos, with preprocessing methods of contrast enhancement, Gaussian blur, down sampling and binarization

Rotation	Zoom	Training accuracy	Validation accuracy
0.1	0	42.79%	39.62%
0.2	0	39.07%	40.69%
0.3	0	37.69%	41.36%
0.4	0	36.63%	40.95%

4.5.2 Changing zoom with fixed rotation level

Table 6 shows the impact of zoom on accuracy. The best results were achieved at zoom = 0 indicating that introducing artificial magnification distorts plume morphology and has no improvements on accuracy.

Table 6: All runs used an OctConv model on 4.6m videos, with preprocessing methods of contrast enhancement, Gaussian blur, down sampling and binarization

Rotation	Zoom	Training accuracy	Validation accuracy
0.3	0	37.69%	41.36%
0.3	0.1	37.07%	38.53%
0.3	0.2	35.87%	37.73%
0.3	0.3	36.09%	35.82%

Overall, rotation-based augmentation proved beneficial, while zooming offers little to no advantage.

4.6 Grad-CAM visualizations

Grad-CAM visualizations (Figure 5) depicted that the OctConv model activated along plume boundaries and dense regions of emissions implying no overfitting to background patterns. The deployment is trustable as it helps in visualizing the model's field of view during classification.

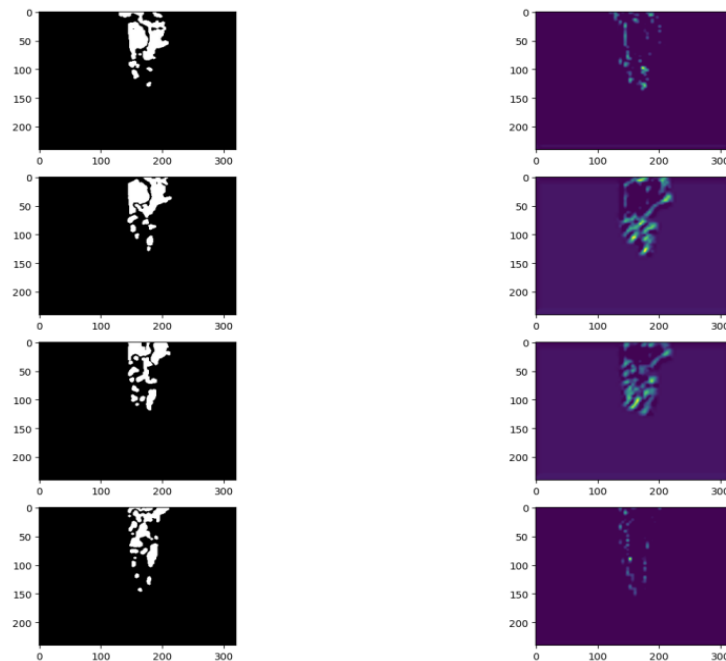


Figure 5: Grad-CAM visualizations of various preprocessed frames

4.7 Test data: Finding the best zoom multiplier

On applying the OctConv model to independent test videos at 0.65 zoom multiplier it resulted in the best accuracy (34.14%) (Figure 6), thus proving that deployment time calibration (cropping/zooming) significantly improves accuracy without retraining. Further, this one-time low-cost calibration strategy adapts the model to various camera placement distances and site setups.

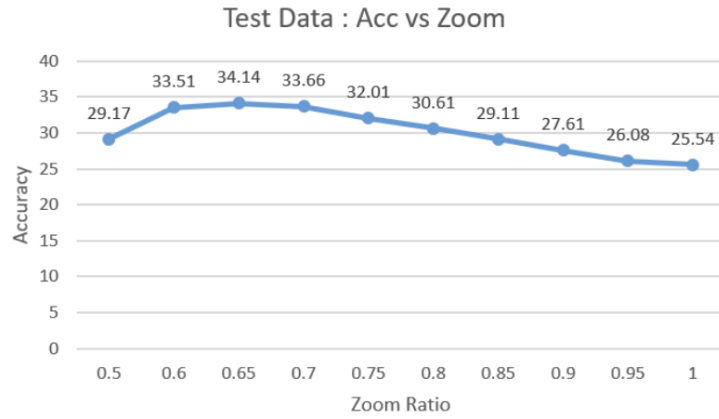


Figure 6: Comparing test accuracy for different zoom factors

All runs used an OctConv model on 4.6m videos, on different zoom test data

4.8 Eight-class classification problem

Figure 7 shows the full eight-class classification results in an overall accuracy of 41.36%, outperforming state-of-the-art 39.1% benchmark. Class 0 “no leak” cases were identified with 100% accuracy, however larger leaks (Classes 4–7) have high intensity plumes which show varying shapes and were less distinguishable by the model. Thus, the model minimizes false positives but is less reliable for larger leaks reinforcing the trade-off between granularity and reliability. Although fine-grained classification is thorough it is challenged by label confusion, especially for higher leak classes.

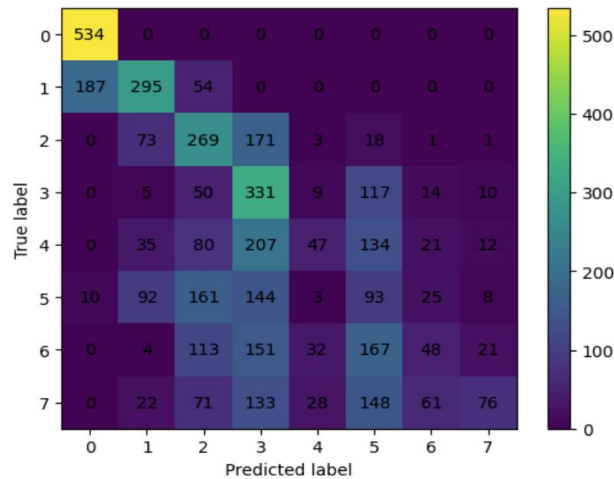


Figure 7: Confusion Matrix with Eight-class classification

4.9. Leak and non-leak binary classification problem

During preprocessing, binarization helps in detecting the plume which makes binary classification possible without model inference. Not only does this reduce computational load by avoiding frames with no detected plumes but also enhances field deployment efficiency as leaks may not be present in most frames.

4.10. Three-class classification problem

Figure 8 shows that combining eight leak classes into small, medium, large leaks categories significantly improved accuracy to 66%. Reduction in granularity of classification improves the reliability of the model. Moreover, identifying “large leaks” is more pertinent to operations than being able to distinguish between adjacent leak classes.

Overall, the accuracy, comprehensibility, and actionable insights of the three-class system makes it more suitable for practical deployment.

- Small leaks: Class 0-1
- Medium leaks: Class 2-4
- Large leaks: Class 5-7

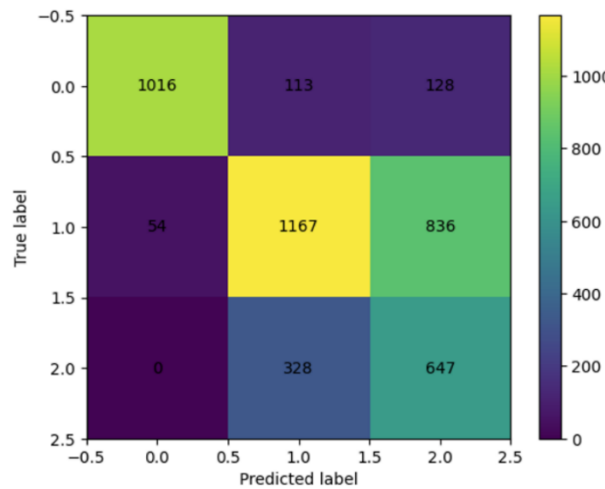


Figure 8: Confusion Matrix with Three-class classification

By preprocessing, we compress the videos by 68%. Using Student-Teacher distillation, we are reducing the size of our model by 79%, reducing it to 4MB from 19MB, making it very lightweight for edge devices. For a 1-minute video, 20 preprocessing takes ~6 seconds and classification takes ~4 seconds. Since we can run this in parallel, we can get real time classification and quantification of methane leaks.

Significance - We have achieved a better 41.36% accuracy which is better than state of the art accuracy of 39.1%. We minimized memory requirements by using a smaller model and significant compression of raw data in preprocessing. This optimizes for constraints on bandwidth in remote sites and hence are ideal for deploying on edge devices. Processing data in batches resulted in reduced memory requirements and faster processing and classification speeds. Low-cost real-time classification and quantization of methane leaks is achieved by running analysis in parallel to smaller chunks of a video. This can be easily generalized to various locations, distances of the camera from the plume, backgrounds, etc.

5. Conclusion and future work

Our paper illustrates a unique application of methane leak detection by early detection. Using the GasVid dataset, we have built a model whose overall size has been reduced by 79% (from 19MB to 4 MB) which is helpful for its use in Internet of Things (IoT) applications because the devices have less processing power and small storage capabilities. The research has used many different optimizers like Lion, Adam and on comparing it found that Lion Optimizers provides the highest accuracy. Data augmentation, the process of adding to the existing data to improve model performance, by maintaining original image size (zoom 0), horizontal flip followed by slight rotation (0.3) drastically improved results. The research also showed that prepping the background of images using the moving average method increased accuracy by 6%. Overall, the Octave Convolutional Model is the most effective with an accuracy score of 41.36%.

Estimating flow rate of the leakage and using that as an added feature in our current methodology and utilizing segmentation analysis to gain thorough understanding of leak related data. As future scope, applying this model to relevant, related datasets especially drone data and satellite data will deliver even better conclusions. In conclusion, developing economical, efficient and accurate models for methane leak detection will have real-world applications, in IoT and edge computing, beyond academic research.

References

- [1] J. Wang, L. P. Tchapmi, A. P. Ravikumar, M. McGuire, C. S. Bell, D. Zimmerle, S. Savarese, and A. R. Brandt, "Machine vision for natural gas methane emissions detection using an infrared camera." *Applied Energy*, vol. 257, 113998, 2020.
- [2] J. Wang, J. Ji, A. P. Ravikumar, S. Savarese, and A. R. Brandt, "Videogasnet: Deep learning for natural gas methane leak classification using an infrared camera." *Energy*, vol. 238, 121516, 2022.
- [3] A. P. Ravikumar, J. Wang, and A. R. Brandt. (2017, Jan.). "Are optical gas imaging technologies effective for methane leak detection?." *Environmental Science & Technology*. [Online]. 51(1), pp. 718–724. Available: <https://doi.org/10.1021/acs.est.6b04512> [Apr. 17, 2025].
- [4] Couto-Silva, C. M., Shetty, S., Olid-Gonzalez, A., Wallez, G., Chatelet, V., and A. Kohar. "Mitigating Nonproductive Time: A Novel Algorithm for Dsl Fault Detection." Paper presented at the International Petroleum Technology Conference, Dhahran, Saudi Arabia, February 2024. doi: <https://doi.org/10.2523/IPTC-24515-MS>
- [5] Duarte, J. J., Vaithianathan, S., Mohan, D. K., Nakamura, M., Soin, M., Tilakpure, N, Tan, D., and A. Kohar. "Condition-Based Maintenance of Oilfield Cement Pumper: A Data-Driven Approach." Paper presented at the SPE Western Regional Meeting, Palo Alto, California, USA, April 2024. doi: <https://doi.org/10.2118/218870-MS>
- [6] Rodriguez, A., Rey-Torres, C., Kohar, A., and A. Malik. "MWD Tools' Electronic Components Data-

Driven Failure Detection." Paper presented at the SPE Western Regional Meeting, Palo Alto, California, USA, April 2024. doi: <https://doi.org/10.2118/218834-MS>

- [7] A. Kohar., J. Sirignano, J. Peng, "Deep learning models for high-frequency financial data." Ph.D. dissertation, University of Illinois at Urbana-Champaign, Urbana, IL, USA, 2019.
- [8] Agora IoT, "AgoraGateway 403 product-sheet." Internet: agoraiot.com/Agora/media/Agora/AgoraGateway_403_Product-Sheet_2022_v8.pdf, 2022.
- [9] N. Rouet-Leduc and C. G. Hulbert, "Automatic detection of methane emissions in multispectral satellite imagery using a vision transformer." *Nature Communications*, vol. 15, 47754, 2024.
- [10] I. Daugėla, J. Sužiedelytė Visockienė, and J. Kumpiene, "Detection and analysis of methane emissions from a landfill using unmanned aerial drone systems and semiconductor sensors." *Detritus*, vol. 10, pp. 127-138, 2020.
- [11] M. Kwaśny and A. Bombalska, "Optical Methods of Methane Detection." *Sensors*, vol. 23, no. 5, 2834, 2023.
- [12] D. R. Thompson, I. Leifer, H. Bovensmann, M. Eastwood, M. Fladeland, C. Frankenberg, K. Gerilowski, R. O. Green, S. Kratwurst, T. Krings, B. Luna, and A. K. Thorpe, "Real-time remote detection and measurement for airborne imaging spectroscopy: a case study with methane." *Atmospheric Measurement Techniques*, vol. 8, pp. 4383-4397, 2015.