Detection and localization of fugitive emissions in industrial plants using surveillance cameras

Oscar D. Pedrayes [a,*], Darío G. Lema [a], Rubén Usamentiaga [a], Daniel F. García [a]

[a] Department of Computer Science and Engineering, University of Oviedo, Campus de Viesques, Gijon 33204 Asturias, Spain

ARTICLE INFO

ABSTRACT

Industrial plants commonly generate gas emissions that are not caused intentionally. These emissions are known as fugitive emissions. Early detection of fugitive emissions helps to find points of failure in the different processes and avoid sources of pollution, helping to reduce danger to the environment and to respect legislation. Despite the importance of the problem, there are no published solutions in the specialized literature about the location and automated detection of fugitive emissions in industrial plants. Therefore, this article proposes an effective approach based on convolutional neural networks for semantic segmentation. The proposed solution takes advantage of existing surveillance cameras to apply state-of-the-art image segmentation methods, in particular, the semantic segmentation network DeeplabV3 + . This work explores aspects such as the ability to differentiate gases like water vapor and clouds from fugitive emissions, the possibility of reusing models in different industrial plants, the differences between multi-class and binary classification, the importance of proportions in the number of images in each class, the use of weights to balance classes, the comparison of a standard size test versus a real use case test, and the feasibility of an area-based alarm system to warn of emissions. This paper describes a methodology to configure the proposed solution for a specific industrial facility.

1. Introduction

Fugitive emissions (Laconde, 2018) are greenhouse gas emissions that are not intentionally produced by a stack or vent. This type of emissions are much more complex, as they do not follow a stack trace, but are scattered, resulting in areas of low opacity and gaps. They are usually caused in industrial plants by the production, processing, transmission, storage and use of fuels. Other causes of fugitive emissions are uncontrolled elements such as wind stirring up accumulated dust or improperly stored products. In some cases, emissions accumulate and escape to the outside through openings other than chimneys. There are many causes of fugitive emissions, so detecting and locating these emissions is essential to finding the problem and correcting it.

Pollution prevention (Johnson, 1992; Freeman et al., 1992) is a priority if the environment is to be preserved. Fugitive emissions pollute the air, endangering the lives of people and animals living nearby. In addition, depending on the composition of the emissions, they can contribute to the greenhouse effect. For example, methane emissions from oil and gas industries have an impact 25 times greater than that of carbon dioxide (Solomon et al., 2007).

As pollution regulation laws become more strict (Lee, 2021; Condren and Dunning, 2021), companies need to find new and innovative ways to control and prevent pollution in the most cost-effective manner in order to remain competitive.

Some of the most common low-cost air pollution sensors include: .

- Electrochemical sensors (Bakker and Telting-Diaz, 2002): are based on a chemical reaction between gases in the air and the electrode in a liquid. These sensors are very sensitive to temperature and humidity variations.
- Photoionization detectors (Davenport and Adlard, 1984): ionize volatile organic compounds and measure the resulting electric current. Such sensors are more expensive and do not distinguish between gases.
- Optical particle counters (Liu et al., 1974): measure particulate matter by detecting the light scattered by the particles.

* Corresponding author.
E-mail address: UO251056@uniovi.es (O.D. Pedrayes).

- Optical sensors (Bogue, 2015): can detect gas by measuring the absorption of infrared light.

The use of non-optical emission detection devices (Williams et al., 2014), have notorious disadvantages. These include the need to install sensors throughout the industrial plant, modifying the existing infrastructure. In addition, the sensors must be in close proximity to emissions in order to detect them. There may be high pollution levels in one area, and a few meters away, very low concentrations. Thus, the results depend to a large extent on where sensors are installed: if they are not installed in the right locations, their results will not be representative. Non-optical sensors can be highly sensitive to weather conditions, such as wind speed, temperature and humidity. Surveillance cameras do not have these drawbacks, however they depend on daylight making them useless at night. One way to use cameras at night is by using infrared sensors. For example, optical gas imaging cameras use different infrared spectral ranges to visualize and detect different types of gas emission, such as methane or ethylene (Naranjo et al., 2010).

The use of surveillance cameras for vision-based detection is an appropriate and low-cost approach when compared to other sensors, especially because the emissions to be detected are in the visible spectrum. For this reason, this work studies the possibility of using video surveillance cameras and adapting the existing infrastructure of the industrial plants themselves. One of the disadvantages of the proposed method is that the chemicals found in the emissions cannot be recognized. Thus, it is not possible to analyze the concentration of different chemical compounds, only their location and the area they occupy. This method is able to detect emissions based on examples of previous emissions. If a new type of emission were to occur, it would probably not be detected properly and the network would have to be retrained.

Surveillance cameras are often low resolution, low quality devices (Rofeim, 2019), badly focused on irrelevant locations. They must be evaluated on an individual basis to determine their usefulness for emission detection. Cleaning and maintenance may be required.

Video surveillance cameras typically cover the majority of industrial plant areas, with coverage of almost 100 % in many cases. For this reason, they are much less intrusive than the installation of sensors, especially for areas such as the air space directly above the industrial plant. Although surveillance cameras cannot detect the chemical composition of emissions, they may be able to detect their location and size.

There are deep learning approaches to detect smoke in RGB imagery (Park and Song, 2019), some of which use ultraviolet (Osorio et al., 2017; Wang et al., 2020) or infrared (Wang et al., 2022) imagery. However, to the best of the authors' knowledge, there is no existing research about image segmentation for fugitive emission detection in industrial plants. This may be because companies are reluctant to make data showing levels of contamination public. In addition, private research may not be shared in order to have an advantage over competitors.

Creating a dataset for this purpose is complex as it might interfere with the normal operation of the company, labeling the images is costly and the images may be of a sensitive nature. The only dataset found that could be of use is a dataset from Project RISE (Hsu et al., 2020), which obtains its data from outside an industrial plant rather than using existing infrastructure from the plant. Furthermore, the dataset is not labeled for semantic segmentation.

At present, there is no research in this field and certainly none with state-of-the-art technologies based on fully convolutional neural networks for semantic segmentation. This paper focuses on fugitive emissions, using a solution based on state-of-the-art image segmentation methods. The use of surveillance cameras is proposed to gather these complex images with low opacity emissions and gaps.

From this proposal several questions arise:

1. The use of pixel-wise segmentation to obtain a mask detailing the location of the emission;
2. The distinction of fugitive emission from other gases such as water vapor and/or clouds despite not being able to obtain their composition;
3. The evaluation of the difference between a binary classification and a multi-class classification and its effects on the fugitive emission class;
4. The execution of trained models on a dataset with realistic emission/non-emission proportions in order to test their effectiveness if put into production;
5. The feasibility of transferring already trained models with images from an industrial plant for application in images from other industrial plants;
6. The study of the minimum number of images necessary for the training of a model.
7. The transfer of models from other industrial plants to reduce the number of images needed to train.
8. The use of detections as an alarm to warn of emissions.

These questions are addressed in Section 3. To answer these questions, three different datasets have been developed, each focusing on a single industrial plant. Each dataset is composed of 1000 images with emission and up to 11,500 images without emission. The non-emission images are simpler to obtain for binary classification because they do not require the creation of a ground truth mask.

As a predictor tool, the semantic segmentation network known as DeepLabV3 + (Chen et al., 2018) is used. This network, developed by Google for Tensorflow, is the state of the art in this field of research, but has never been used for fugitive emissions localization. It obtains excellent results in other areas such as autonomous driving (Sgibnev et al., 2020), land cover classification (Pedrayes et al., 2021), or brain tumor detection (Choudhury et al., 2018) among other fields for pixel-wise segmentation. To obtain the best possible results with this network, a thorough hyper-parametric tuning process is necessary. In this paper, the most important aspects to fine-tune the network for these specific datasets are detailed.

2. Methods and materials

2.1. Datasets

This paper shows images from three different industrial plants (Plant1, Plant2 and Plant3). Each industrial plant provides images from a single camera. These cameras were already existing surveillance cameras to monitor buildings with risk of fugitive emission. However, the monitoring process was manual. In the figures of this paper, the regions corresponding to the buildings are colored with the red corresponding to the building class. For the images of industrial plants 1 and 2, a crop is applied to avoid showing the silhouette of the building. This is to protect the anonymity of the company that provides this information. However, all experimentation has been done with the original images.

A dataset for each industrial plant is generated. Training and test sets are divided in 75 % and 25 % respectively. All the images are taken at random times on four different days, excluding night-time since the cameras do not have enough visibility. The minimum time separation between two images is 5 s. The cameras have a 4 K/8MP sensor, a framerate of 30fps and a horizontal angle of 95–10 degrees. Quality is degraded because zoom levels are adjusted to frame the

images to the corresponding buildings. The lens of video surveillance cameras are designed to have a larger field of view in order to cover as much ground as possible. The area of interest is only a specific area of the image itself. In addition, they are low cost and low maintenance cameras because they are located in inaccessible places, so it is common that there is blurring due to dirt on the lenses. All the images consist of the red, green, and blue (RGB) bands and have a resolution of 2048 × 1536 pixels or 1024 × 768 pixels. To alleviate the amount of memory needed in VRAM and to keep all image sizes constant, all images are scaled to 512 × 384 pixels using the bicubic interpolation.

A real use case has a proportion of 1:34 emission to non-emission images. During a normal day on average, only one image with fugitive emissions was captured for every 34 images with no emission. To test how models trained with different proportions performed in a real environment, a test set of 250 fugitive emission images, and 8500 non-fugitive emission images was created and evaluated. This real test is generated for Plant1 and Plant2.

To study the importance of the proportion used to train a model that will normally be used under other proportions (realistic proportions), multiple variations are designed for each dataset. Each variation uses a different number of non-emission images. In Table 1, the different proportions for the datasets for Plants 1 and 2 along with the number of images that correspond to emissions and non-emissions for training and test sets are shown. The fugitive emission images are maintained throughout all the variations to improve comparability. The dataset from Plant 3 has only the proportion 2:1 since it does not have enough images.

Plant1 has the following classes: building, cloud, sky, fire, water vapor chimneys, and fugitive emission. Plant2 has: building, cloud, sky, water vapor chimneys, and fugitive emission. Plant3 has: building, cloud, sky, and fugitive emission. The labeling was carried out by experts using software tools and then reviewed by the operators of the industrial plants.

Classes in the ground truth masks have the following colors associated: .

- ■ Building
- ■ Water vapor chimney
- ■ Cloud
- ■ Fire
- ■ Fugitive emission
- ■ Sky

Fugitive emissions are the target class for this evaluation study. This is the only class of interest, therefore, a comparison between multi-class classification and binary classification and its effects on the target class can be studied.

2.2. Network architecture

DeepLabv3 + (Chen et al., 2018) is the most recent version of DeepLab (Chen et al., 2014; Chen et al., 2017a; Chen et al., 2017b), a convolutional neural network architecture for semantic segmentation developed by Google. The architecture of this network is based on an encoder-decoder structure with an Atrous Spatial Pyramid Pooling (ASPP) module in the encoder part. This evaluation study

Table 1
Proportions for the different datasets. (emission:non-emission).

Proportion	Train	Test
2:1	750:375	250:125
1:1	750:750	250:250
1:2	750:1500	250:500
1:4	750:3000	250:1000

uses the official implementation from Google's Github, which can be accessed with the following link: https://github.com/tensorflow/models/tree/master/research/deeplab. This implementation is based on TensorFlow. A diagram of the DeepLabV3 + architecture can be seen in Fig. 1.

2.3. Metrics

One of the most common metrics to determine the quality of a prediction in semantic segmentation is the F_1-Score. As seen in Eq. (1), it is calculated as a combination of both Precision and Recall and is equivalent to the Dice Coefficient with two classes.

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{1}$$

Precision is calculated using Eq. (2), as the correctly classified pixels from the total predicted pixels. Recall is calculated using Eq. (3), as the pixels classified correctly from the pixels that correspond to that particular class in ground truth. If both metrics are very different (unbalanced) it usually means that the predictions will tend to over classify pixels from that particular class (low Precision and high Recall), or that it will only predict pixels that are too obvious (high Precision and low Recall).

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

Intersection-Over-Union or IoU is equivalent to the Jaccard Index and is used to measure the area of similarity of a prediction to its ground truth. This metric is common for measuring the quality of prediction in the context of image segmentation. It is calculated using Eq. (4), as the ratio between the true positives and the sum of all the pixels that are not true negatives.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} = \frac{TP}{TP + FN + FP} \tag{4}$$

2.4. Training

To get the best possible results, hyper-parameters must be tuned accordingly for the dataset used. Optimal hyper-parameters were studied for the three datasets separately. As they are similar, the optimal hyper-parameters coincide.

The following hyper-parameters were tuned to improve predictions: different input sizes (2048 × 1536, 1024 × 768, 512 × 384, and 256 × 192) and their effect on batch size; number of classes using multi-class and binary predictions; class weighting comparing the Median Frequency Weighting (MFW) (Eigen and Fergus, 2015) method against custom weights; learning rate and epochs; output strides of 8, 16, and 32; different backbone networks (Resnet50, Xception45, Xception65, Xception71, MobileNetV2, MobileNetV3Small, MobileNetV3Large); L2 regularization; and solver algorithms such as Adam or Stochastic Gradient Descent with Momentum (SGDM).

Data augmentation was used to obtain more training data in order to reduce overfitting, and to improve predictions. The augmentation process consists of zooms of the images with varying zoom values ranging from 0.5 to 2.0 at intervals of 0.25. The dataset was shuffled for epoch to prevent overfitting.

As testing all possible combinations of hyper-parameters with a single computer would be too time consuming, each hyper-parameter was adjusted individually. Batch size was re-evaluated for every change that involves VRAM usage. Architecture changes, such
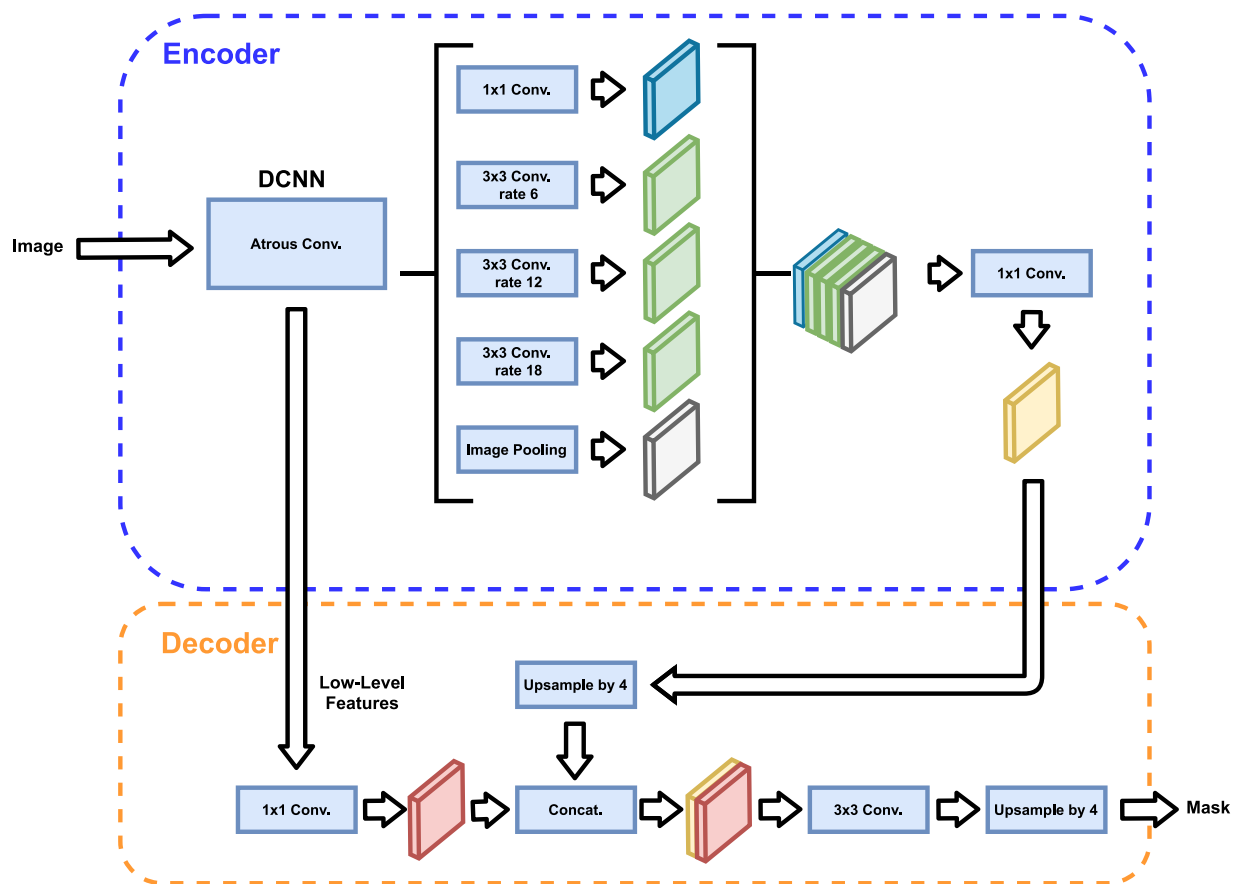
Fig. 1. DeepLabv3 + architecture.

as the backbone network, can reduce VRAM usage, allowing for larger batch sizes.

The best configuration found for binary classification for all three datasets can be seen in Table 2. As the three datasets are similar, the best configurations are the same, except for the value of the custom class weights, which may vary slightly. For multi-class classifications, hyperparameters are the same except for class number, and class weighting. In the multi-class classification no custom class weighting was performed, only MFW weighting was used. This is due to the added difficulty of the compound effect of the different classes affecting each other. Obtaining a configuration of custom weighting values would be too time consuming.

Models were trained using an NVIDIA RTX 2080 Ti GPU with 11 GB of VRAM. Training the DeepLabV3 + network from scratch can take weeks, mainly due to the backbone network. For this reason, pretrained models with the ImageNet dataset (Deng et al., 2009) are used for each of the backbone networks. From these pretrained models only the backbone network part of the architecture is loaded. The pretrained models are provided by Google in its original GitHub for DeepLab.

3. Results and discussions

3.1. Multi-class experimentation

First, to determine if an image segmentation of fugitive emissions in industrial plants can be obtained from surveillance camera images, multi-class experiments were carried out. This section shows the results for experiments using DeepLabV3 + for multi-class segmentation. These experiments use the MFW method for class weighting balancing.

Metrics from Table 3 show high F_1-Score values, surpassing the 80% barrier in every class for all three datasets proving that a functional prediction mask can be obtained. This solves question (1) from the "Introduction" in Section 1. Fugitive emission is one of the classes with the lowest F_1-Scores and its metrics are usually sightly biased towards Recall over Precision. These metrics are to be expected since the fugitive emission class is much harder to see and has greater variability in both area and color than other classes. This makes both the network and the ground truth prone to error. In the same way, since it is hard to distinguish between them, classes Cloud and Sky might increase their metrics by being merged into the same class.

From these results it can be established that it is possible to differentiate clouds, water vapor chimneys, and fugitive emissions, solving question (2).

Visualization for the experiments of Table 3 can be seen for Plant1 in Fig. 2, for Plant2 in Fig. 3, and Plant3 in Fig. 4. The first column of the three figures shows the input image; the second column, the original multi-class ground truth mask; the third column, the multi-class classification prediction from the model; the fourth column, a superposition of the ground truth mask with the input image; and the fifth column, the superposition of the predicted multi-class mask from the model over the input image.

Results for Plant1 (Fig. 2) are very good visually. Most predictions are indistinguishable from the ground truth, except the fourth image

Table 2
Training parameters for DeepLabV3 + .

Training parameters	
Input size	512 × 384 × 3
Classes	2
Backbone	Xception65
Output stride	16
Padding	Yes
Solver	Adam
Epochs	80
Batch size	6
Learning rate	0.00005
Class weighting	Non-Target: ~ 0.40 - Target: ~ 0.60
Gradient clipping	No
L2 regularization	0.0004
Data augmentation	Scale 0.5–2.0 with 0.25 steps
Shuffle	Yes

Table 3
Metrics for the multi-class classification experiments.

Dataset	Class	Precision	Recall	IoU	F_1
Plant1	Building	0.995	0.997	0.992	0.996
	Vapor	0.915	0.900	0.831	0.907
	Clouds	0.955	0.834	0.802	0.890
	Fire	0.794	0.837	0.688	0.815
	Emission	0.836	0.832	0.715	0.834
	Sky	0.926	0.951	0.884	0.938
Plant2	Building	0.997	0.988	0.985	0.992
	Vapor	0.744	0.934	0.707	0.847
	Clouds	0.800	0.931	0.755	0.861
	Emission	0.767	0.895	0.704	0.826
	Sky	0.984	0.935	0.921	0.959
Plant3	Building	0.992	0.994	0.986	0.993
	Clouds	0.779	0.710	0.591	0.743
	Emission	0.881	0.939	0.833	0.909
	Sky	0.954	0.936	0.895	0.945

which detects fire even when the ground truth shows none. In this case, it appears that the ground truth may be labeled incorrectly, as a small transparent fire can be seen in the original input image. This means that the model, for this particular case, is outperforming the ground truth. Obviously, this is impossible to detect just by looking at the metrics, but it does mean that the model may be performing better than the numbers show. Results from Plant2 (Fig. 3) are also visually excellent. Most predictions are indistinguishable from the ground truth, except the second image, which detects fugitive emissions when the ground truth does not. Finally, results from Plant3 (Fig. 4) are also very good visually. However, the second image shows that the emission is not fully detected.

3.2. Binary experimentation

Although multi-class segmentation can successfully localize fugitive emissions in industrial plants, in order to reduce the complexity of the ground truth masks, binary segmentation is evaluated and compared with multi-class segmentation. Binary segmentation has great advantages over multi-class segmentation. It is much cheaper and easier to build ground truth masks with a single class rather than multiple classes that require several different regions per image. In addition, the process required to adjust the network is much simpler, since parameters such as class weighting are easier to set. This is because there are only two classes, so modifying one only affects the other rather than affecting multiple classes at the same time. This allows for an in-depth study of class weighting for class balancing. Table 4 only show results for the emission class. The "non-emission" class is not detailed since it is not needed.

Results from the MFW class weighting experiments for binary segmentation shown in Table 4 are comparable to those of the multi-class experiment from Table 3. These experiments are more biased towards Recall, obtaining very high Recall but lower Precision. Even though there is a disparity between Recall and Precision, F_1-Score values are not far from those in the multi-class experiments (see Table 3). Using binary classification instead of multi-class makes the imbalance between Recall and Precision bigger, lowering the quality of the segmentations. Plant3 is the exception because multi-class and binary results are practically the same. This indicates that these differences are highly dependant on the dataset used and its complexity, given that Plant3 has no elements such as water vapor chimneys or fire. In answer to question (3) from the "Introduction" in Section 1, merging all non-target classes into a single class causes that class to be more unbalanced than the target class than when they are separated. In other words, the difference in the number of pixels between the two classes increases, causing Recall to remain high, but Precision to decrease.

3.3. Class weighting experimentation

The results of the binary segmentation experiments show that their metrics are biased towards Recall over Precision. This clearly limits the quality of the prediction: Recall and Precision must be balanced. Class weighting has a great impact on the training process of the model. Standard methods for class weighting, such as Inverse Frequency Weighting (IFW) (Cui et al., 2019) and Median Frequency Weighting (MFW) are often used. However, as there is a problem balancing Recall and Precision, custom weights are studied in order to reach a balance. Plant3 is not included in this study as its Precision and Recall are already acceptably balanced. Binary segmentation benefits from the fact that only the best weights for two classes need be found. This greatly simplifies the study of how class weighting affects the results obtained.

In Table 5 multiple values for class weighting are tested to study the behavior of the network under different weightings. To make the class weighting experimentation easier, all values are normalized between 0 and 1. In this way, if a value of 0.60 is used for the target class, is it assumed that the remaining 0.40 is used for the non-target class.

Metrics from Table 5 show that custom weights for binary classification can balance Recall and Precision and improve overall accuracy, increasing its F_1-Score. Values from the custom weights for binary classification achieve even better results than those of multi-class classification with MFW. Balancing both Recall and Precision has a great impact on the quality of the model. After balancing the Recall and Precision correctly, there is no benefit to adding classes other than the target classes to segment. In this case, only the fugitive emission class is of value. For this reason, and because of the advantages mentioned above, the rest of the study will be focused on binary segmentation for the fugitive emission class.

Visualization for the experiments with the highest F_1-Scores in Table 5 can be seen for Plant1 in Fig. 5, for Plant2 in Fig. 6, and Plant3 (from Table 4) in Fig. 7. The three figures show: in the first column, the input image; in the second column, the original multi-class ground truth mask; In the third column, the binary classification prediction from the model; in the fourth column, a superposition of the ground truth mask with the input image; and in the fifth column, the superposition of the predicted binary mask from the model over the input image.
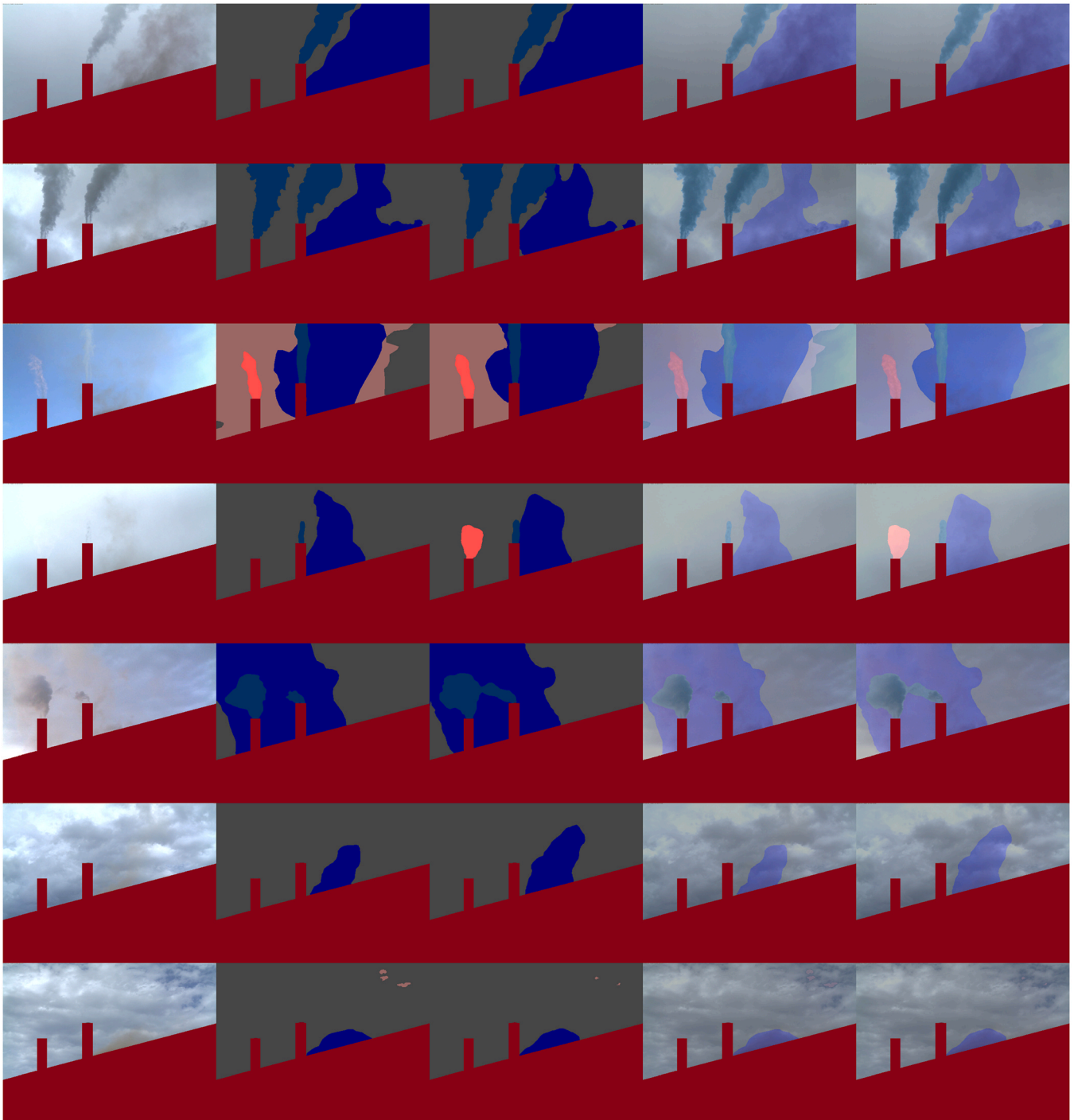
Fig. 2. Plant1 multi-class classification test visualization.

The results for Plant 1, 2 and 3 are visually very good, as can be seen in (Fig. 5, 6, and 7). In Plant1 most predictions are indistinguishable from the ground truth, except the second image, which detects fewer fugitive emissions than the ground truth. In Plant2 most predictions are also indistinguishable from the ground truth, except the second image, which detects fugitive emissions when the ground truth does not. Finally, Plant3 has nearly perfect visual results.

3.4. Training with different class proportions and a real test

Binary segmentation with custom class weighting has robust results. To further validate these results, a larger test set is evaluated. This test set aims to replicate the proportions in which images exhibit fugitive emissions over a full day. In this case, for every 34 images without emission, one image with emission is produced. For the sake of simplicity, this is referred to as a 1:34 proportion.
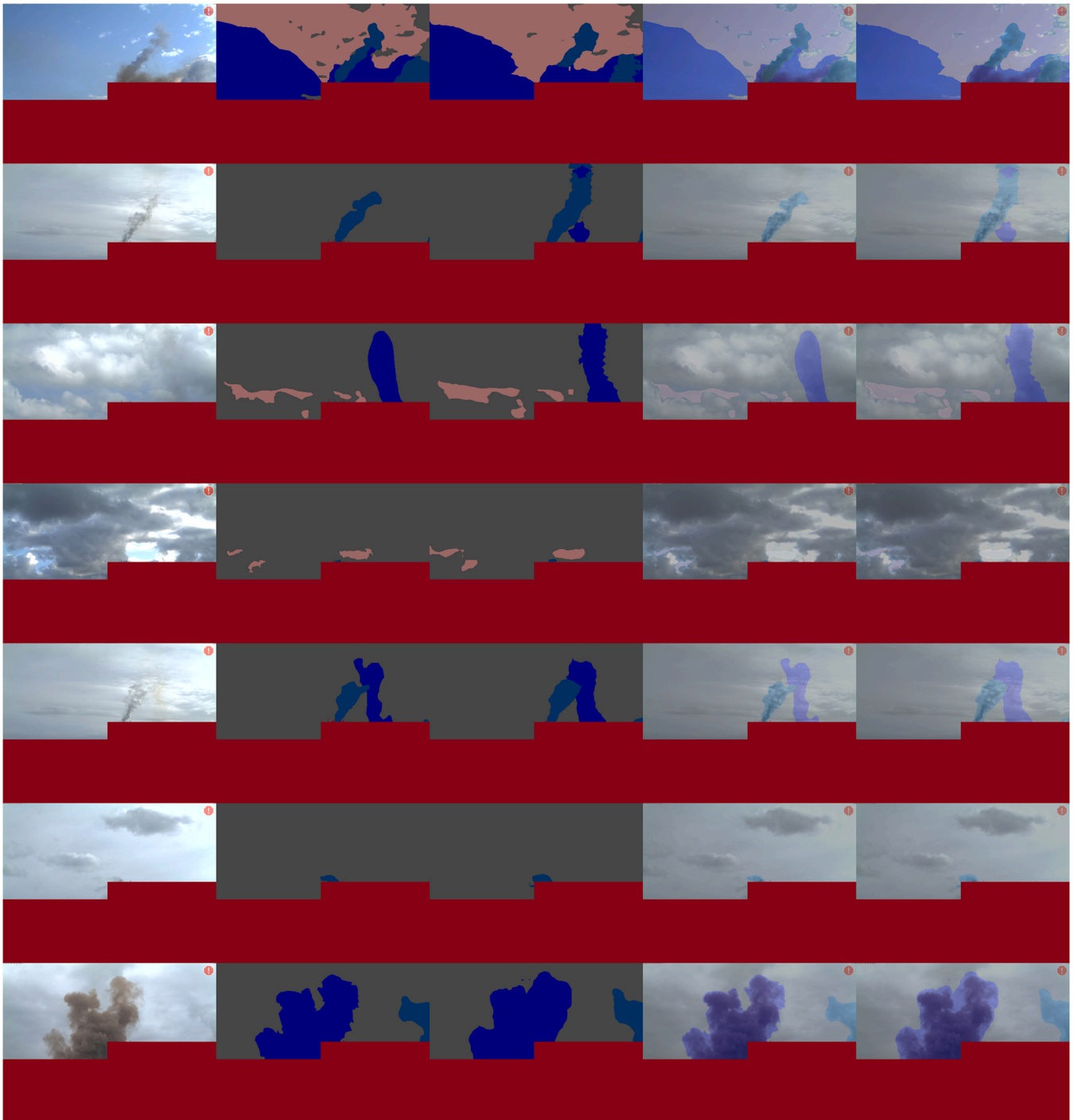
Fig. 3. Plant2 multi-class classification test visualization.

In the case of Plant3 there are not enough images to create this kind of test set. For Plant1 and Plant2, 250 images with emission and 8500 non-emission images are used.

When a network trained with a set of 2:1 proportion is tested with the real proportion of 1:34, Precision is greatly reduced due to new False Positives (see Table 6). However, the Recall remains the same since no new images with fugitive emission were added to the test. This effect can be seen in the first row for Plant1 and the fifth row for Plant2 in Table 6. In order to solve this problem, experiments with varying class proportions were carried out. The objective was to determine if a change in proportions during training affects the Precision of the predictions when adjusting the class weighting of the classes to balance Recall and Precision. Table 6 shows the best custom class weighting experiment for each proportion.
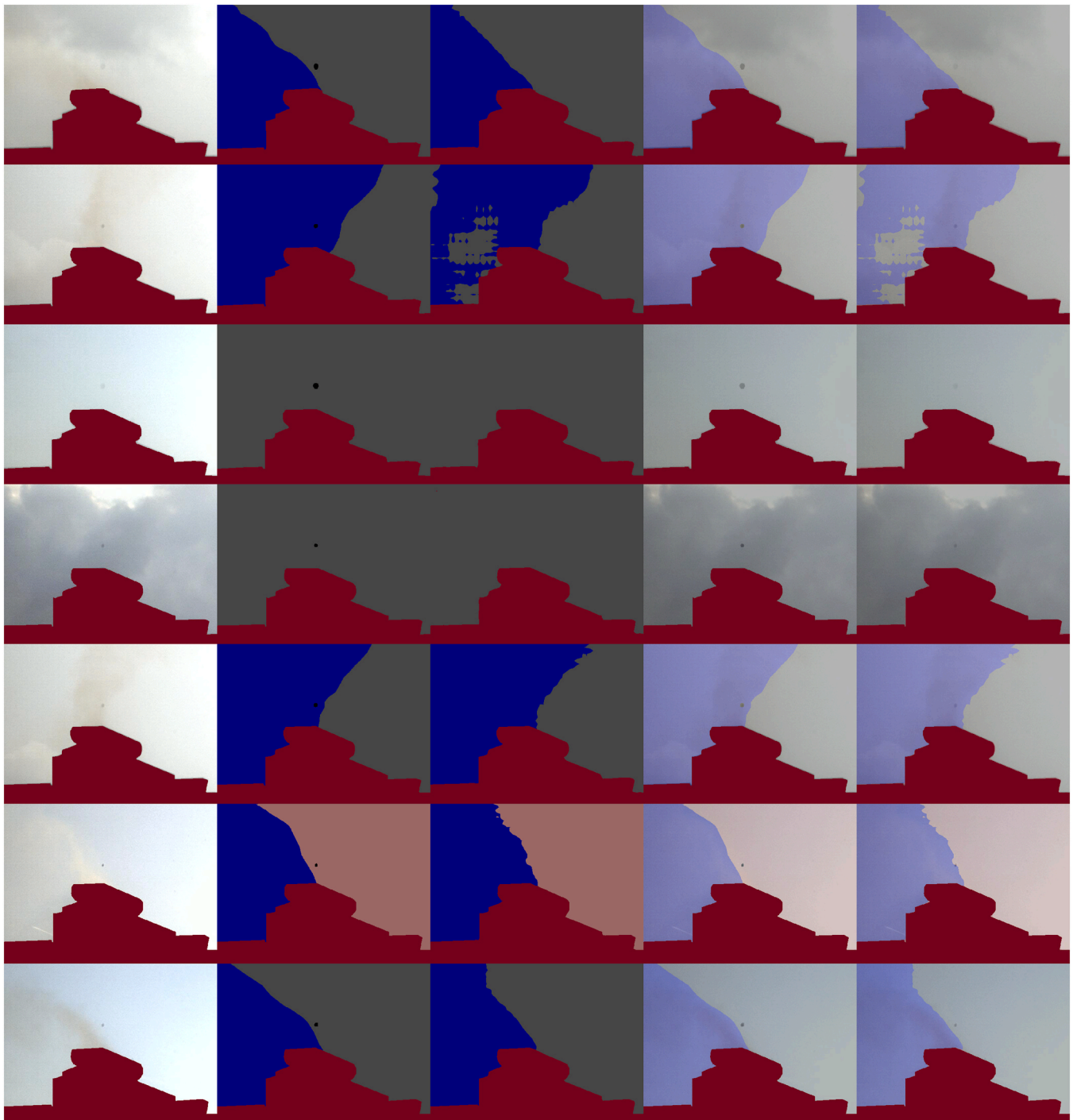
Fig. 4. Plant3 multi-class classification test visualization.

Table 4
Metrics for the binary classification experiments.

Dataset	Precision	Recall	IoU	F$_1$
Plant1	0.634	0.958	0.617	0.763
Plant2	0.648	0.965	0.633	0.775
Plant3	0.885	0.938	0.836	0.911

In Table 6 it is observed that a proportion of one emission image per two non-emission images (1:2) is the optimal proportion in both datasets.

A higher non-emission proportion gives the model far more stability and fewer false positives. This discovery is vital since the standard training tests do obtain excellent results. To answer question (4), if no real test were done, the erroneous assumption that the

Table 5
Metrics for the binary classification with custom class weighting experiments.

Dataset	Weight	Precision	Recall	IoU	F$_1$
Plant1	(MFW) 0.88	0.634	0.958	0.617	0.763
Plant1	0.70	0.765	0.884	0.695	0.820
Plant1	0.60	0.826	0.841	0.715	**0.833**
Plant2	(MFW) 0.923	0.648	0.965	0.633	0.775
Plant2	0.80	0.779	0.891	0.711	0.831
Plant2	0.70	0.842	0.842	0.727	**0.842**
Plant2	0.60	0.866	0.779	0.695	0.820

models could generalize as well as the 2:1 test in a real scenario could be made. Training using a realistic proportion of 1:34 is impractical, as the time required to set up and train such a model would be too time consuming; the number of images required is beyond the scope of this research. For this same reason, no experiments are performed with proportions higher than 1:4, even though such experiments would be interesting to confirm this statement or to find an upper limit.
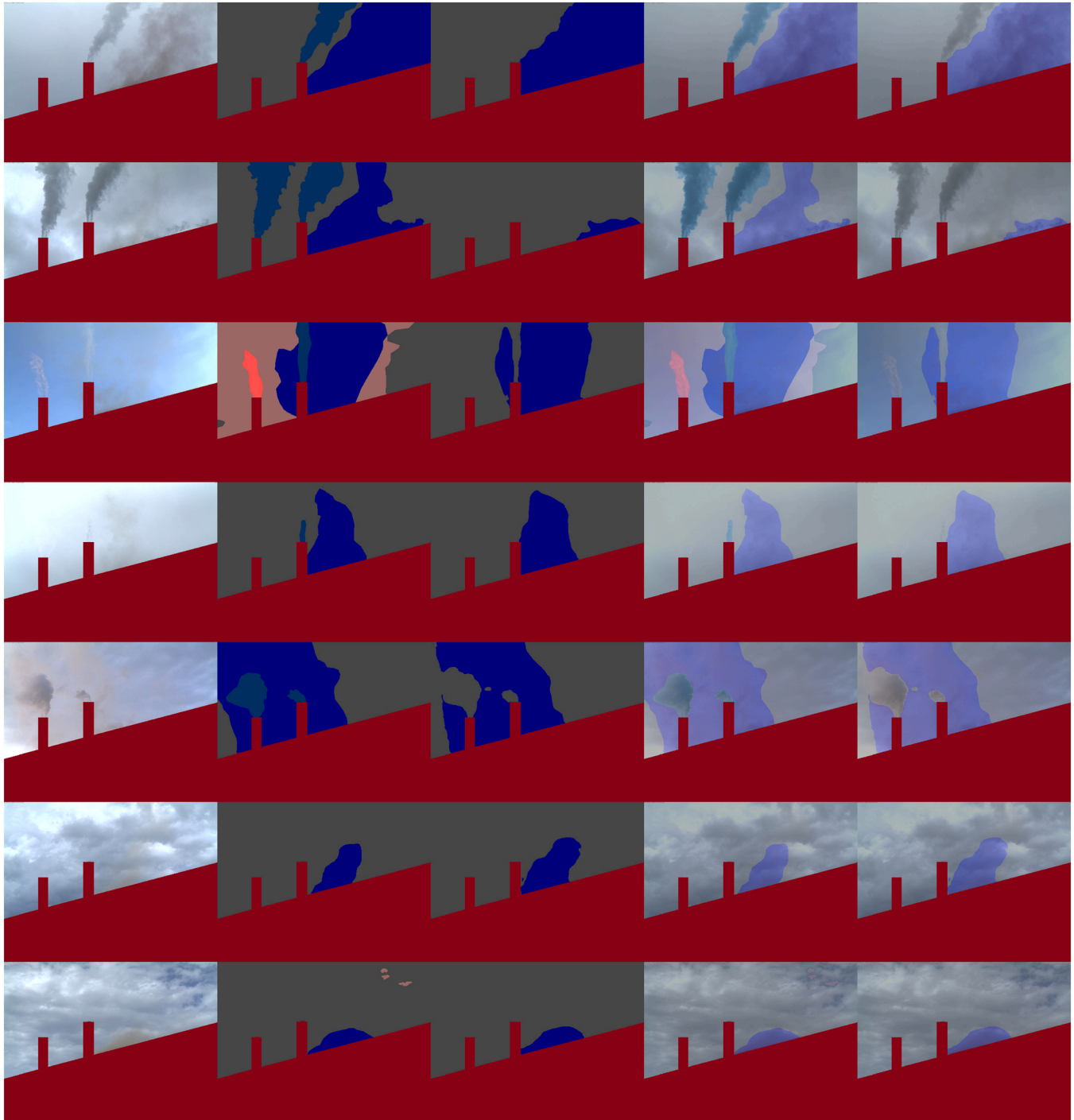

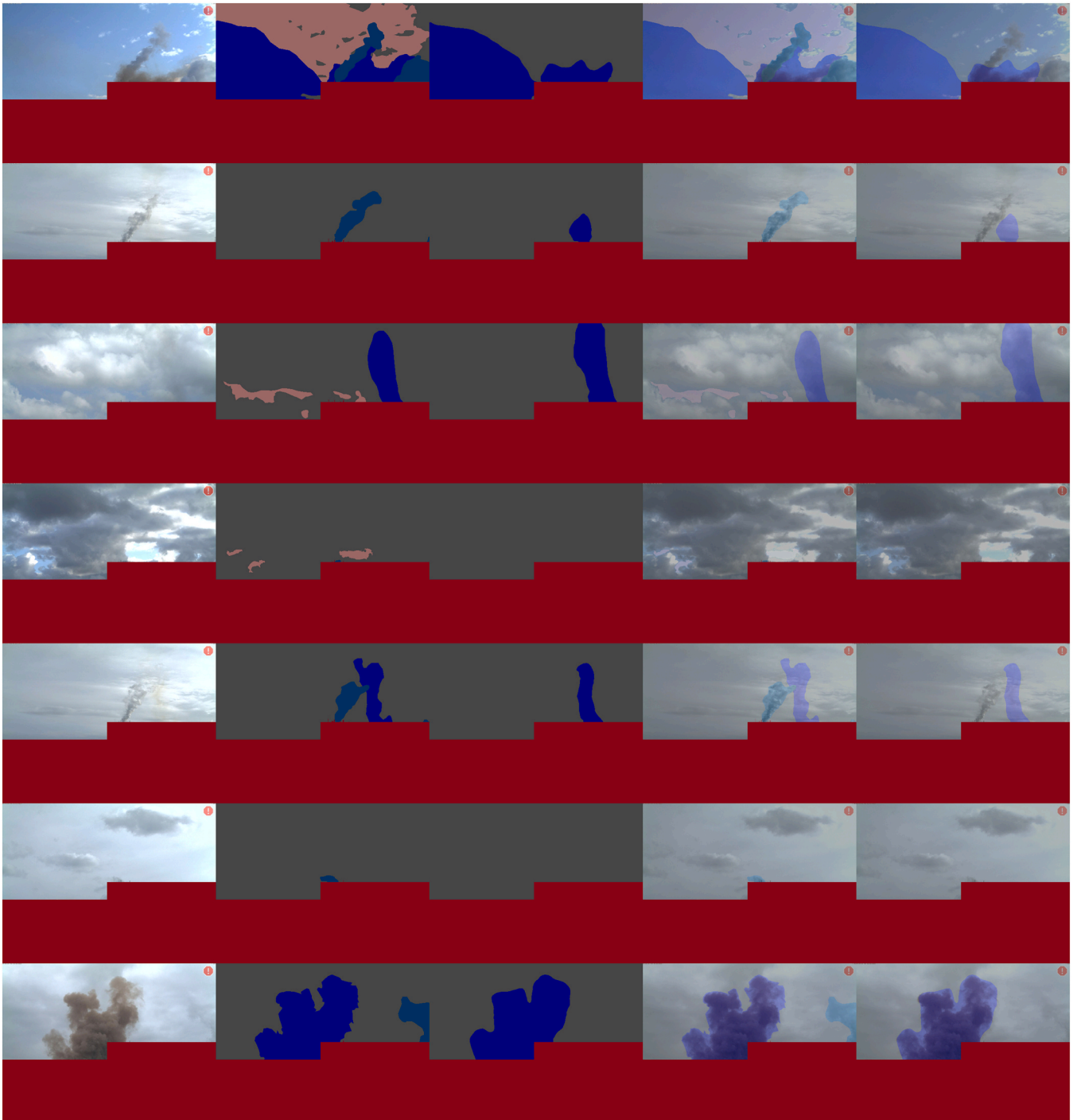**Fig. 5.** Plant1 binary classification test visualization.

Fig. 6. Plant2 binary classification test visualization.

3.5. Generalization between datasets

Since a large amount of parameter configuration is required, it is reasonable to try to take advantage of models trained in one industrial plant to be used in another. Experiments using a model trained with images from one industrial plant and testing with images from another are carried out and their results are presented in Table 7.

Metrics from Table 7 show that testing with different datasets does not produce good quality predictions. To answer question (5), models are not directly transferable between datasets. Clearly, models cannot be used indistinctly in different industrial plants.

3.6. Reduced training

As it is not possible to reuse a model already trained in another industrial plant, it is interesting to study the minimum number of images needed to train a model and obtain metrics similar to using a large dataset. This could greatly reduce time and costs. Experiments with fewer images for Plant2 and Plant3 are evaluated in Table 8. For these experiments, a proportion of 2:1 is maintained to facilitate comparisons. Testing sets are not modified to make results directly comparable.

To answer question (6), the reduced training shows that using only 100 images is enough to obtain acceptable results, as shown in
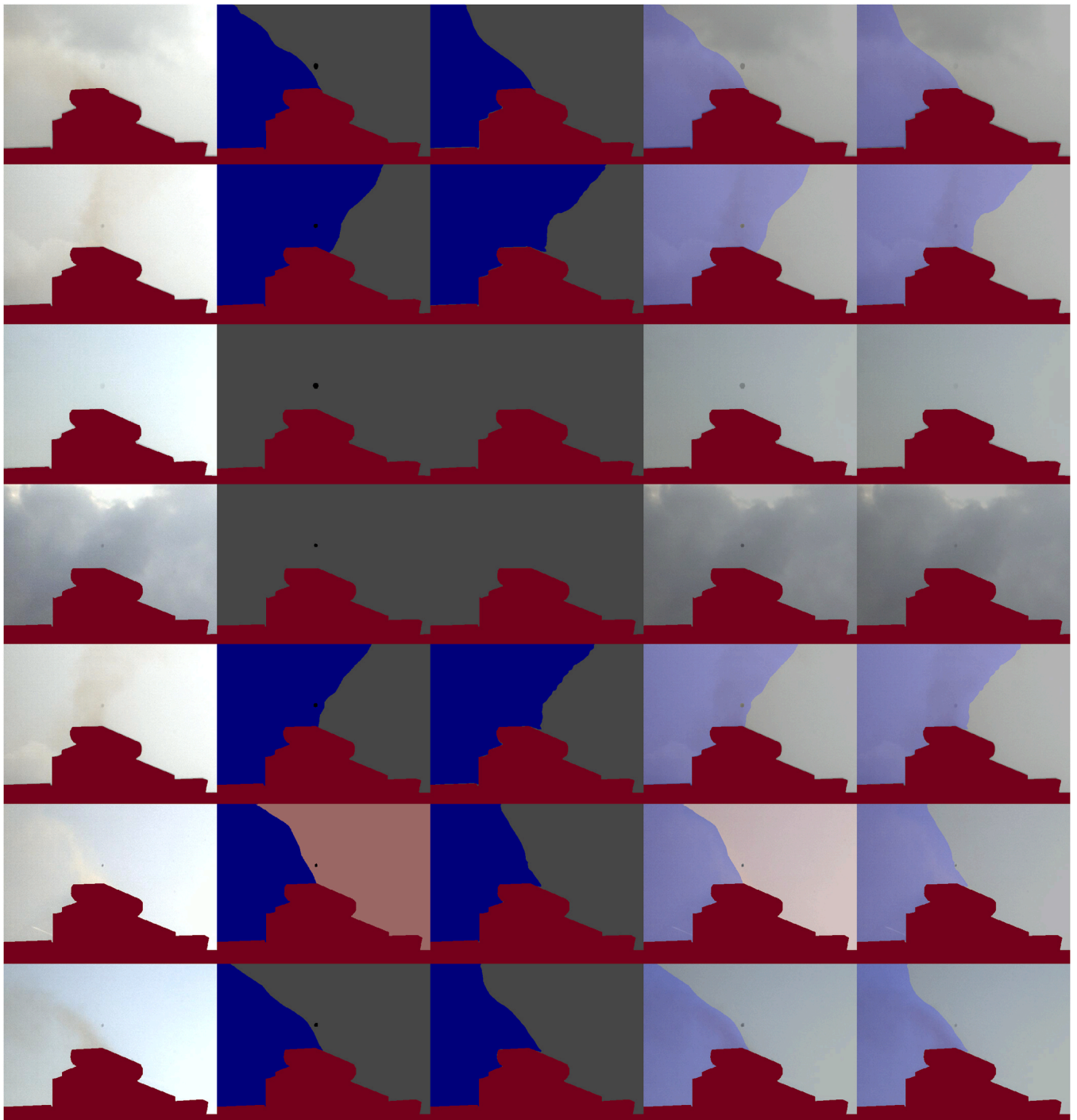
Fig. 7. Plant3 binary classification test visualization.

Table 6
Metrics for the proportion experiments.

Parameters			Tests with train proportions				Tests with real proportions (1:34)			
Dataset	Train prop.	Weight	Precision	Recall	IoU	F_1	Precision	Recall	IoU	F_1
Plant1	2:1	0.60	0.826	0.841	0.715	0.833	0.585	0.841	0.527	0.690
Plant1	1:1	0.51	0.831	0.787	0.678	0.808	0.678	0.787	0.573	0.728
Plant1	1:2	0.60	0.833	0.815	0.700	0.824	0.819	0.815	0.691	0.817
Plant1	1:4	0.50	0.783	0.850	0.688	0.815	0.540	0.850	0.493	0.660
Plant2	2:1	0.70	0.842	0.842	0.727	0.842	0.365	0.842	0.342	0.509
Plant2	1:1	0.55	0.852	0.826	0.723	0.839	0.786	0.826	0.675	0.806
Plant2	1:2	0.635	0.849	0.816	0.713	0.832	0.835	0.816	0.702	0.825
Plant2	1:4	0.57	0.843	0.823	0.714	0.833	0.815	0.823	0.694	0.819

Table 7
Metrics for the dataset cross testing experiments.

Dataset	Precision	Recall	IoU	F_1
Plant1 → Plant2	0.728	0.380	0.333	0.499
Plant1 → Plant3	0.446	0.129	0.111	0.200

Table 8
Metrics for reduced training set.

Dataset	Images	Precision	Recall	IoU	F_1
Plant2	100	0.668	0.834	0.590	0.742
Plant2	250	0.697	0.837	0.614	0.761
Plant3	100	0.847	0.895	0.770	0.870
Plant3	250	0.862	0.890	0.779	0.876

Table 9
Metrics for transfer learning experiments.

Dataset	Images	Precision	Recall	IoU	F_1
Plant1 → Plant2	10	0.698	0.620	0.489	0.657
Plant1 → Plant2	50	0.777	0.683	0.571	0.727
Plant1 → Plant2	100	0.788	0.757	0.629	0.772
Plant1 → Plant2	250	0.729	0.815	0.625	0.770
Plant1 → Plant3	10	0.801	0.801	0.668	0.801
Plant1 → Plant3	50	0.808	0.927	0.760	0.863
Plant1 → Plant3	100	0.837	0.871	0.745	0.854
Plant1 → Plant3	250	0.841	0.916	0.781	0.877

Table 8. These metrics are almost 5 % lower than the experiments from Table 4, which use 1125 images. However, a 5 % reduction using a dataset ten times smaller may be acceptable when time and cost restraints exist.

3.7. Transfer learning

Since at least 100 images are needed to achieve robust training, it is interesting to study whether this number can be further reduced by reusing a model already trained in another industrial plant. Even if it cannot be used directly, perhaps another model can serve as a checkpoint to reduce training time on other datasets. In this section, experiments are carried out to observe if a model trained for one industrial plant can be used as a base for training with a new dataset for another industrial plant (i.e., transfer learning). This would mean training with fewer images. In both cases, "Transfer learning" and "Reduced training" are trained with the exact same images to allow for comparisons.

Transfer learning results from Table 9 show an improvement of 3% in the F_1-Score for Plant2 when compared to the reduced training of 100 images from Table 8. However, in the case of Plant3 this does not improve its metrics, and even reduces its F_1-Score by 2 % in the 100-image experiment. This shows that the models might be too dependant on the dataset used and cannot be generalized for the application of emission segmentation. Since it is already possible to train with only 100 images, transfer learning may not be necessary. Thus, in answer to question (7), transfer learning cannot be used to reduce the number of images needed.
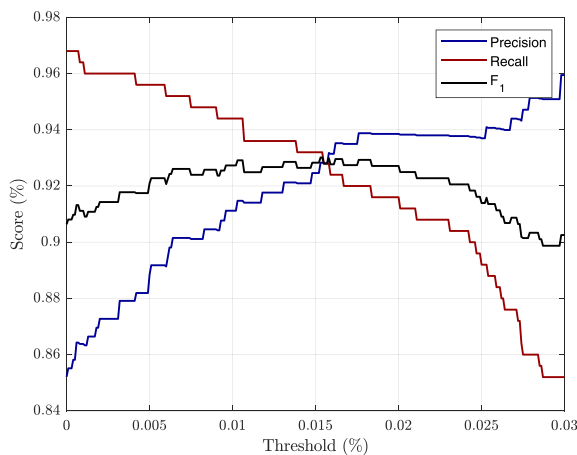
3.8. Emission detection

Once a robust model for fugitive emission segmentation has been obtained, it seems reasonable to use these segmentations to detect fugitive emissions. To raise an alarm when a fugitive emission is detected it is necessary to apply semantic segmentation to each frame of the video feed. This alarm is used to indicate whether or not there are fugitive emissions in the image (i.e., it is a binary alarm). In this way, the predictions are used for image classification. The realistic test set is used to study its behavior.

In order to avoid raising the alarm for false positives consisting of extremely small regions, an area threshold for the segmented areas is needed. This threshold is measured as the minimum percentage of pixels with emission out of the total pixels of the image. In this section, the optimal area thresholds are calculated using the models from the realistic test from Plant1 and Plant2.

For each threshold evaluated, using steps of 0.01 %, the images of the dataset that will be considered as true emissions are determined. Then the predicted images with and without emission are analyzed as a function of the threshold used, obtaining the metrics of Recall, Precision and F_1-Score. These metrics are represented in Figs. 8(a) and 8(b). It should be noted that although the metrics used are the same, in this case they are calculated per image and not per pixel.

Plant1 has an optimal threshold of 1.56 % of the area of the total image (Fig. 8(a)). Plant2 has an optimal threshold of 0.5 % of the area of the total image (Fig. 8(b)).

Metrics from Table 10 show that over 85 % and almost 93 % F_1-Score values are achieved for Plant1 and Plant2 respectively. These experiments demonstrate that this method can be successfully used for image classification, answering question (8). In addition, the detected segmentation regions can be used to determine different levels of severity based on area or shape. For example, a fugitive emission severity level from 1 to 10 could be declared using 10% area

(a) Plant1
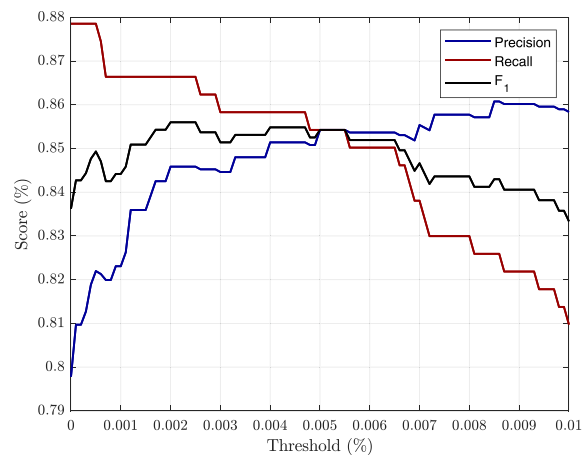
(b) Plant2

Fig. 8. Threshold study.

Table 10
Metrics for the emission detection alarm.

Dataset	Threshold	Precision	Recall	F_1
Plant1	0.0050	0.854	0.854	0.854
Plant2	0.0156	0.928	0.928	0.928

increments of the emissions. In other words, a severity level of 3 could be 30 % of the sky area occupied by emissions.

4. Conclusion

The method proposed in this paper to segment and detect fugitive emissions in images taken from surveillance cameras in industrial plants, reaches results above 80 % F_1-Score. Clouds, water vapor chimneys, and fugitive emissions can be distinguished with no problem. These results are achieved despite the low resolution and quality of the surveillance cameras used to acquire the images, some of which include images with dirt on the lenses. This means that the location and detection of fugitive emissions using surveillance cameras is possible. Overcoming other traditional methods such as sensors that measure volume concentrations or chemical compounds in the air, as these sensors require to be placed in areas where a large flow of emissions is expected, i.e. planned emissions such as emissions produced by stacks. This is a great advantage since a much larger area can be covered than with such sensors. Another advantage over non-optical sensors is that the result of the detection is a new 2D image with classification at pixel level. This type of image is very easy to interpret and verify visually.

Surveillance cameras are a type of optical sensor. These cameras only detect the visible spectrum. If another optical sensor were used that could detect a wider spectral range, results would most likely improve. However, this would also increase the cost dramatically. One of the advantages of this method is that surveillance cameras are low cost sensors, and they usually already exist in the industrial plant.

It is observed that there is no need to add more classes to the training process, since there is no significant improvement over using only the target classes, as long as the Precision and Recall are balanced. This can save labeling time and is far more cost efficient when creating new datasets.

This is relevant since a model trained for a particular camera of a particular industrial plant does not generalize well enough to be used for another industrial plant with the same type of fugitive emissions. The differences between datasets are far too great for the model to produce good results. Furthermore, transfer learning does not provide any improvement over training from a pre-trained model from Imagenet. Transfer learning between different industrial plants is highly dependant on the similarity of the datasets and cannot be used in a generalized manner. However, transfer learning is not needed since with only 100 training images results are close to a full training with a dataset over ten times larger.

It is common to use metrics obtained from a test with fewer images than the training set. It is usually a test that does not follow the class proportions of a real use case, as it commonly focuses on training the target classes. This paper confirms that, whenever possible, a test with as many images as possible, using a realistic proportion, should be used to ensure that the results of a model are fully validated.

Different class proportions during training have a significant effect in real tests even when class weighting is used. Regardless of the class weighting strategy used, the class proportion is of much greater importance. A proportion that is closer to the real scenario works better than training with all the classes in the same proportion. However, custom class weighting is still essential to balance Recall and Precision even if more realistic image proportions are

used. Results from the best model, in a real test, obtain over 80 % F_1-Score, with balanced Precision and Recall.

Finally, when applying an area threshold to the predictions and used as an image classification approach to serve as an alarm method for emission detection, the F_1-Score can be as high as 90 %. This proves that this method can be used to detect fugitive emissions and thus help to control pollution in industrial plants.

Given the great importance of pollution control in industrial plants to preserve the environment, the need to continue this research into new ways of detecting and locating fugitive emissions is essential. It becomes apparent that the approach of using semantic segmentation as a basis for further analysis is appropriate, especially considering the speed at which this field is improving. The main contribution of this paper is the study of different possibilities for detecting and locating fugitive emissions with the new state-of-the-art semantic segmentation technologies and their possible uses for emission alerts.

CRediT authorship contribution statement

Oscar D. Pedrayes: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization **Darío G. Lema:** Investigation, Resources, **Supervision Rubén Usamentiaga:** Investigation, Resources, Data Curation, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition, Funding acquisition Daniel **F. García:** Investigation, Resources, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

Bakker, E., Telting-Diaz, M., 2002. Electrochemical sensors. Anal. Chem. 74, 2781–2800.

Bogue, R., 2015. Detecting gases with light: a review of optical gas sensor technologies. Sens. Rev. 35, 133–140.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017a. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. 40, 834–848.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2014. Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv:1412.7062.

Chen, L.C., Papandreou, G., Schroff, F., Adam, H., 2017b. Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the ECCV. pp. 833–851.

Choudhury, A.R., Vanguri, R., Jambawalikar, S.R., Kumar, P., 2018. Segmentation of brain tumors using deeplabv3. In: Proceedings of the International MICCAI Brainlesion Workshop, Springer.154–167.

Condren, M., Dunning, H., 2021. Experts push for stricter air pollution standards in new environment legislation. http://www.imperial.ac.uk/news/231324/experts-push-stricter-pollution-standards-environment/.

Cui, Y., Jia, M., Lin, T.Y., Song, Y., Belongie, S., 2019. Class-balanced loss based on effective number of samples. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9268–9277.

Davenport, J., Adlard, E., 1984. Photoionization detectors for gas chromatography. J. Chromatogr. A 290, 13–32.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 248–255.

Eigen, D., Fergus, R., 2015. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2650–2658.

Freeman, H., Harten, T., Springer, J., Randall, P., Curran, M.A., Stone, K., 1992. Industrial pollution prevention! a critical review. J. Air Waste Manag. Assoc. 42, 618–656.

Hsu, Y.C., Huang, T.H., Hu, T.Y., Dille, P., Prendi, S., Hoffman, R., Tsuhlares, A., Pachuta, J., Sargent, R., Nourbakhsh, I., 2020. Project rise: recognizing industrial smoke emissions. arXiv:2005.06111.

Johnson, S.M., 1992. From reaction to proaction: The 1990 pollution prevention act. Columbia. J. Environ. Law 17, 153.

Laconde, T., 2018. Fugitive emissions: a blind spot in the fight against climate change. Fugitives emissions-sector profile. INIS 51.

Lee, H., 2021. Air quality watchdog oks tighter regulations for la county oil refineries. https://www.dailybreeze.com/2021/11/05/air-quality-watchdog-oks-tighter-regulations-for-la-county-oil-refineries/.

Liu, B.Y., Berglund, R.N., Agarwal, J.K., 1974. Experimental studies of optical particle counters. Atmos. Environ. 8 (1967), 717–732.

Naranjo, E., Baliga, S., Bernascolle, P., 2010. Ir gas imaging in an industrial setting. In: Proceedings of the Thermosense XXXII, SPIE. pp. 160–167.

Osorio, M., Casaballe, N., Belsterli, G., Barreto, M., Gómez, Á., Ferrari, J.A., Frins, E., 2017. Plume segmentation from uv camera images for so2 emission rate quantification on cloud days. Remote Sens. 9, 517.

Park, J.S., Song, J.K., et al., 2019. Fcn based gas leakage segmentation and improvement using transfer learning. Proceedings of the 2019 IEEE Student Conference on Electric Machines and Systems (SCEMS 2019). IEEE, pp. 1–4.

Pedrayes, O.D., Lema, D.G., García, D.F., Usamentiaga, R., Alonso, Á., 2021. Evaluation of semantic segmentation methods for land use with spectral imaging using sentinel-2 and pnoa imagery. Remote Sens. 13, 2292.

Rofeim, M., 2019. Why are your security cameras blurry or fuzzy? https://diysecuritytech.com/why-are-your-security-cameras-blurry-or-fuzzy.

Sgibnev, I., Sorokin, A., Vishnyakov, B., Vizilter, Y., 2020. Deep semantic segmentation for the off-road autonomous driving. The International Archives of Photogrammetry, Remote Sensing and Spatial. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 43, 617–622.

Solomon, S., Manning, M., Marquis, M., Qin, D., et al., 2007. Climate Change 2007-The Physical Science Basis: Working group I Contribution to the Fourth Assessment Report of the IPCC. volume 4. Cambridge University Press.

Wang, J., Tchapmi, L.P., Ravikumar, A.P., McGuire, M., Bell, C.S., Zimmerle, D., Savarese, S., Brandt, A.R., 2020. Machine vision for natural gas methane emissions detection using an infrared camera. Appl. Energy 257, 113998.

Wang, J., Ji, J., Ravikumar, A.P., Savarese, S., Brandt, A.R., 2022. Videogasnet: deep learning for natural gas methane leak classification using an infrared camera. Energy 238, 121516.

Williams, R., Kilaru, V., Snyder, E., Kaufman, A., Dye, T., Rutter, A., Russell, A., Hafner, H., 2014. Air Sensor Guidebook. US Environmental Protection Agency.