RESEARCH ARTICLE

VOLCANICA

# VIGIA–PlumeNet and VIGIA-PlumeData: Open-source AI segmenting model and database for volcanic plumes and emissions from optical cameras

Sophie Giffard-Roisin *[α], Yves Moussallam[β], Sébastien Valade[γ], Emily Ramos[δ], Thomas C. Wilkes[ε], Freddy Vásconez[δ], and Robin Campion[γ]

[α] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre, 38000 Grenoble, France.
[β] Lamont-Doherty Earth Observatory, Columbia University, New York, USA.
[γ] Universidad Nacional Autónoma de México, Instituto de Geofísica, Mexico City, Mexico.
[δ] Instituto Geofísico, Escuela Politécnica Nacional, Ap. 17-01-2759, Quito, Ecuador.
[ε] Department of Geography, University of Sheffield, Sheffield, United Kingdom.

Submitted: 2025-10-28 | Accepted: 2026-03-05 | Published: 2026-04-27

Editor: E. Nicholson | Typesetter: J. Farquharson

## Abstract

Continuous monitoring of volcanoes is essential for advancing scientific understanding and issuing alerts to at-risk populations. Many volcanoes are equipped with ground-based RGB cameras, whose recordings are analyzed by experts or semi-automatic models. Automatically distinguishing clouds from volcanic columns and other emissions remains a major challenge. To date, no open-source labeled database of volcanic images exists, and no study has attempted to localize multiple emissions beyond ash columns. We introduce VIGIA–PlumeNet, capable of operating across diverse environments a,nd sky conditions. It performs multi-class segmentation, identifying the volcano and distinguishing plumes, gases, and lava, with 88 % accuracy on the 58 test images when excluding the background class. This is achieved by using the versatile DINO–v2 AI model as a general visual processor, paired with a specialized component that we specifically trained to pinpoint and outline volcanic features; and constructing VIGIA-PlumeData, a dataset of 250 manually annotated images from over 60 volcanoes worldwide, collected through community effort. VIGIA-PlumeData and VIGIA–PlumeNet are released publicly, establishing the first benchmark dataset for volcanic image segmentation.

## Non-technical summary

Volcanoes can change quickly, and monitoring them in real time is important for keeping nearby communities safe. Many volcanoes are watched by regular cameras that take pictures all the time. Experts usually look at these images, but it is time-consuming and it can be hard to tell apart normal clouds from volcanic ash, gases, or other emissions. Until now, there has been no public database of labeled volcano images to help train computer models, and most existing tools could only detect volcanic ash columns. In this study, we created the first open database of more than 250 images from over 60 volcanoes around the world, with detailed labels showing different volcanic features. Using this dataset, we trained an artificial intelligence model called VIGIA–PlumeNet, which can automatically identify six categories, including the volcano itself, the column, ash emissions, gases, and lava. Applied to a test set of 58 images, the model reaches about 88 % accuracy when excluding the background class and works well even in difficult conditions. Both the database and the model are freely available, so that other researchers can improve them and develop new tools. This is an important first step toward better, open, and automatic monitoring of volcanoes using cameras.

Keywords: Volcanic monitoring; Machine learning; Plume; Segmentation; Optical Camera; Foundation models.

## 1 Introduction

Active volcanoes frequently release volcanic ash and gases in the form of eruption plumes. These plumes represent major volcanic hazards, ranging from local pyroclastic flows and ash deposits to ash clouds transported over large distances by wind. In addition to ash plumes, volcanoes also emit large amounts of water vapor and other gases (such as $SO_2$); these emissions are common and can be challenging to distinguish from meteorological clouds in visual imagery. Nowadays, imaging devices have become essential tools to monitor active volcanoes. Imaging is crucial for assessing the timing, size, and dispersion direction of volcanic plumes. It contributes both to a better scientific understanding of these processes, by enabling quantitative measurements of plumes and their relation to other signals such as seismic or geodetic data, and to more effective alert systems for populations at-risk.

Volcanic imaging is commonly achieved through satellite-based remote sensing and ground-based cameras, each modality providing complementary information. Cameras operating in different wavelength ranges serve distinct purposes: optical (RGB) cameras capture the visual appearance of the volcanic emissions and surrounding atmosphere, infrared (IR) cameras measure the thermal properties of pyroclastic emissions, and ultraviolet (UV) cameras detect $SO_2$ gas emissions. Ground-based cameras are particularly valuable for identifying the onset of explosive activity and supporting rapid alerts. Such cameras are typically operated by local observatories and often provide continuous or near-continuous monitoring

* sophie.giffard@univ-grenoble-alpes.fr

–Article in Press–

at different frame rates, with a portion of the data streams freely available (e.g. the Alaska Volcano Observatory* or the INGV–Osservatorio Etneo†). Integrated systems such as VIGIA [Vásconez et al. 2022] combine thermal and visible imagery to better track explosive activity.

Ground-based imagery is analyzed daily by experts across many active volcanoes worldwide. The monitoring process can be divided into three key tasks: detection, segmentation, and parameter extraction. Detection aims to separate inactive or obstructed frames from those containing visible plumes. Segmentation involves delineating the plume within the image, often producing a binary mask. Parameter extraction then quantifies physical properties such as plume height, direction, composition, velocity, or volume. While these tasks can be performed manually, several studies have explored automating them, though existing tools are often tailored to a single camera and a fixed viewpoint.

When automated, the first task (detection) is either performed using external information such as seismic signals [Centeno et al. 2024], or by thresholding the summit region of the volcano in thermal (IR) acquisitions [Vásconez et al. 2022] previously calibrated for a specific viewpoint. A 6-class classification of IR images was also performed by convolutional neural networks on a single IR camera of Etna [Nunnari and Calvari 2024]. To our knowledge, Witsil and Johnson [2020] was the first study to develop an automatic detection method for emissions in RGB ground-based images, using an artificial neural network to classify each frame into one of five categories (clouds, nocturnal glow, inactivity, dark emissions, and light emissions). However, this approach did not include segmentation (one class for the whole image) and was trained on images from a single camera at Villarrica volcano, Chile.

For the segmentation task, a variety of algorithms have been developed for thermal cameras, typically based on improved thresholding techniques [Valade et al. 2014; Bombrun et al. 2018; Vásconez et al. 2022]. A related approach was adapted for optical videos under clear-sky conditions in [Simionato et al. 2022].

Joint detection and segmentation of volcanic emissions in optical imagery became feasible with the advent of convolutional neural networks, particularly `U-Net` architectures [Ronneberger et al. 2015]. These networks, with their characteristic "U-shaped" design, are specifically made to evaluate an entire image and then output a highly detailed, pixel-by-pixel map of the target objects. For their training, a database with pairs of images and their manually labelled masks is needed. Notably, Guerrero Tello et al. [2022] developed a `U-Net` trained on data from three cameras at Mount Etna, Italy, while Centeno et al. [2024] applied a similar approach on images of Sabancaya, Peru. To date, only Wilkes et al. [2022] have attempted to develop a generic `U-Net` segmentation model, designed to operate across multiple volcanoes without retraining. For this, the authors collected 130 images worldwide, manually delineating the visible emissions and including inactive scenes. While the database was not released, the trained `U-Net` was made publicly available. The results are encouraging, yet the authors note that cloudy scenes are sometimes misclassified, and the segmentation remains limited to a binary distinction (emission versus no emission), without separating ash plumes from gas plume, nor the volcanic edifice itself.

Finally, parameter extraction—the third monitoring task—can be performed with or without prior segmentation. For instance, Aravena et al. [2023] estimated plume height directly from optical video frames using thresholding and geometric operations, after pre-calibration to the specific camera conditions. However, parameter extraction is more readily automated when applied to segmented masks. A number of studies have focused on extraction of plume characteristics, such as height, velocity, acceleration, shape, volume, ash loading, and entrainment coefficients [e.g. Valade et al. 2014; Bombrun et al. 2018; Dürig et al. 2018; Vásconez et al. 2022; Wilkes et al. 2022]. Most methods restrict their estimates to the camera field of view, with the exception of Barnie et al. [2023], who demonstrated how combining multiple cameras with 3D geometric constraints can provide precise estimates of the absolute height of $SO_2$ plumes.

The large majority of previous studies focus primarily on volcanic ash plumes, i.e. tephra columns. However, monitoring and analyzing active volcanoes also requires attention to other volcanic emissions and gases, which can be broadly classified into ash clouds and non-ash volcanic emissions. Ash clouds are less dense than eruption columns and often represent previous column emissions dispersing into the atmosphere (see Figure 1, images 4 and 47). Non-ash volcanic emissions include gases, aerosols, and thermal radiation, which can be further divided into high-temperature and low-temperature components. High-temperature emissions are gas mixtures that last equilibrated with magma at high temperature, typically above 500 °C, they contain $SO_2$, are less prone to condensation and may glow at night (see images 71, 166). Low-temperature emissions refers to gas mixtures that have interacted with a hydrothermal system, they are more water-rich, $SO_2$-poor or absent, always condensed and never reflect incandescence (see image 149). These can include fumaroles on the volcano's flanks or persistent degassing plumes that appear as steam. Optical images may also capture pyroclastic flows and lava (images 4 and 166).

Although experts can differentiate most of these emissions, relying solely on optical cameras is challenging due to missing spectral or thermal information. IR (thermal) cameras can sometimes help in separating the different emissions based on temperature, though in practice they may be at the same temperature by the time we image them. UV cameras are good at distinguishing low from high temperature gases because they see $SO_2$, but UV imaging is noisy and requires careful calibration. Visible cameras have been used to analyze $SO_2$ emissions but the automatization is only performed at the parameter extraction stage [Barnie et al. 2023]. To date, no automatic method has been proposed to segment more than the ash column in ground-based optical imagery.

In this work, we present a new freely available database and a deep learning algorithm, capable, for the first time, of multi-class segmentation of ground-based optical images. The network was trained to distinguish six classes: ash plumes

*https://avo.alaska.edu/
†https://www.ct.ingv.it/

–Article in Press–

(separated in ash emissions and eruption column), gas emissions, lava flows, land and background. We collected more than 250 images from over 60 volcanoes, including challenging conditions such as fully or partially cloudy skies, mixed emissions, varying illumination, and near- and far-field perspectives. Each pixel was manually labeled by experts into the different classes. The first contribution of this work is the release of this benchmark, organized into training, validation, and test sets. Second, we leveraged a generic AI foundation model, `DINO-v2-reg` [Darcet et al. 2023; Oquab et al. 2023], trained originally on 142 million natural images, and adapted it to our task by training a small convolutional segmentation head- few specialized neural network layers that translates the model's general visual features into precise, pixel-by-pixel boundaries of the volcanic emissions. This approach allows the model to disentangle complex clouds and volcanic emissions while using a relatively small dataset. The resulting model is also publicly released.

## 2 Data and Methodology

### 2.1 VIGIA-PlumeData: 256 open-source labelled RGB images of volcanic eruptions

From the original database of Wilkes et al. [2022], we extracted all images that were free for non-commercial use. This formed our initial 90-image batch, primarily consisting of photos taken with standard hand-held cameras, along with a few sequences from monitoring devices. We then expanded this dataset with images from various sources, including publicly available monitoring cameras[*†], personal acquisitions from manual and monitoring devices, and open-access images from the internet, such as Wikipedia. The final database comprises 256 images from over 60 volcanoes—including Stromboli (Italy), Yasur (Vanuatu), Cleveland (USA), Cotopaxi (Ecuador), Bezymianny (Russia), Karymsky (Russia), Popocatépetl (Mexico), Etna, Sakurajima (Japan), and Semisopochnoi (USA)—most captured from multiple viewpoints. The full list is available in the data repository. We included a wide range of conditions: fully obstructed and night views, partially visible volcanoes and emissions, various cloud and sky conditions, inactive volcanoes, non-ash emissions, lava, different illumination levels, diverse backgrounds and foregrounds, coarse to fine resolution images, multiple orientations, near- and far-field views, and columns of varying size. While no dataset of this size can be exhaustive, we aimed to cover the largest possible range of conditions so that new acquisitions would likely fall within the training distribution. Example images are shown in Figure 1.

Each image was manually labeled (using the `labelme` tool[‡]) into eight classes: *Background*, *Eruption column*, *Land*, *Eruption (ash) cloud*, *Pyroclastic flow*, *Low-temperature gas*, *High-temperature gas*, and *Lava*. While all eight labels are provided in the Vigia-PlumeData dataset released with this study, the `Vigia-PlumeNet` model was trained using only six of them (see Section 2.2). This labeling scheme was chosen for several reasons. First, identifying the volcano is crucial, as it allows us to quickly separate fully obstructed frames from the rest (see image 93 from Figure 4, or partial obstruction by clouds as in image 98 from Figure 1). However, a "volcano" label is not always easy to distinguish from surrounding land, since its base may not be visible, and close-up views may not reveal clear volcanic geometry (see images 47 and 148 in Figure 1). Therefore, we used a single *Land* label to encompass volcanoes, mountains, foreground terrain, buildings, and vegetation. Second, atmospheric clouds are an important consideration, as they can be difficult to distinguish from volcanic emissions. While labeling sky and clouds separately is possible in theory, in practice it is tedious and often ambiguous: for example, the boundary between a cloudy sky and a uniformly white or gray sky can be unclear (see image 149). Since our main goal is detecting volcanic emissions, we instead labeled all non-land areas—including sky, clouds, and water bodies such as lakes—as *Background*.

We also distinguished between the *Eruption column* and *Eruption (ash) clouds*. We define the class *Eruption column* as actively ascending ash plumes driven by eruptive momentum, typically exhibiting a vertical orientation (see early works, like Wilson et al. [1978]) and clear convective rolls, whereas *Eruption (ash) clouds* represent the passively drifting or dispersed ash that lacks these features. Although this is a continuous transition and classification cannot be perfect, the distinction is useful: the column label resembles the conventional "ash plume", typically easier to detect as it is opaque with clear volutes, while the more diffuse ash cloud is harder to identify as visibility fades (see image 4) and is somewhat less critical for monitoring. Moreover, we added non-ash gases and attempted, as much as possible, to separate them into high- and low-temperature categories (see images 47, 149, 150, 166). Lastly, the class *Lava* was included, and refers to both lava flows and lava fountains.

We manually labelled the data using the tool `labelme`[§], with the full procedure available in our code repository[¶]. All labeling choices were reviewed by volcanic gas experts and, where possible, supported with contextual information about the volcano, the acquisition date and with surrounding images when available. Nevertheless, we acknowledge that errors remain, as visible cameras—and often single frames—cannot perfectly separate these classes. We always took the 'most likely' class based on a volcanologist expertise. For training, we reduced the problem to six classes by merging *Low-* and *High-temperature gases* into *Non-ash emissions*, and combining *Eruption (ash) cloud* with *Pyroclastic flow* into a single (non-column) *Ash emissions* class. However, we kept all eight categories in the released database, since we expect this benchmark to be useful beyond our model and anticipate that future approaches may succeed in separating them.

Finally, the 256 images were divided into three sets. The training set, used to train the segmentation head, contains 139 images. The validation set, comprising 55 images, was used for hyperparameter tuning, while the test set of 58 images was

[*] https://avo.alaska.edu/
[†] https://www.ct.ingv.it/
[‡] freely available at https://labelme.io/
[§] Available at https://github.com/wkentaro/labelme
[¶] Available at https://gricad-gitlab.univ-grenoble-alpes.fr/giffards/vigia-plumenet/-/blob/main/labelling_directory/readme.md?ref_type=heads

–Article in Press–

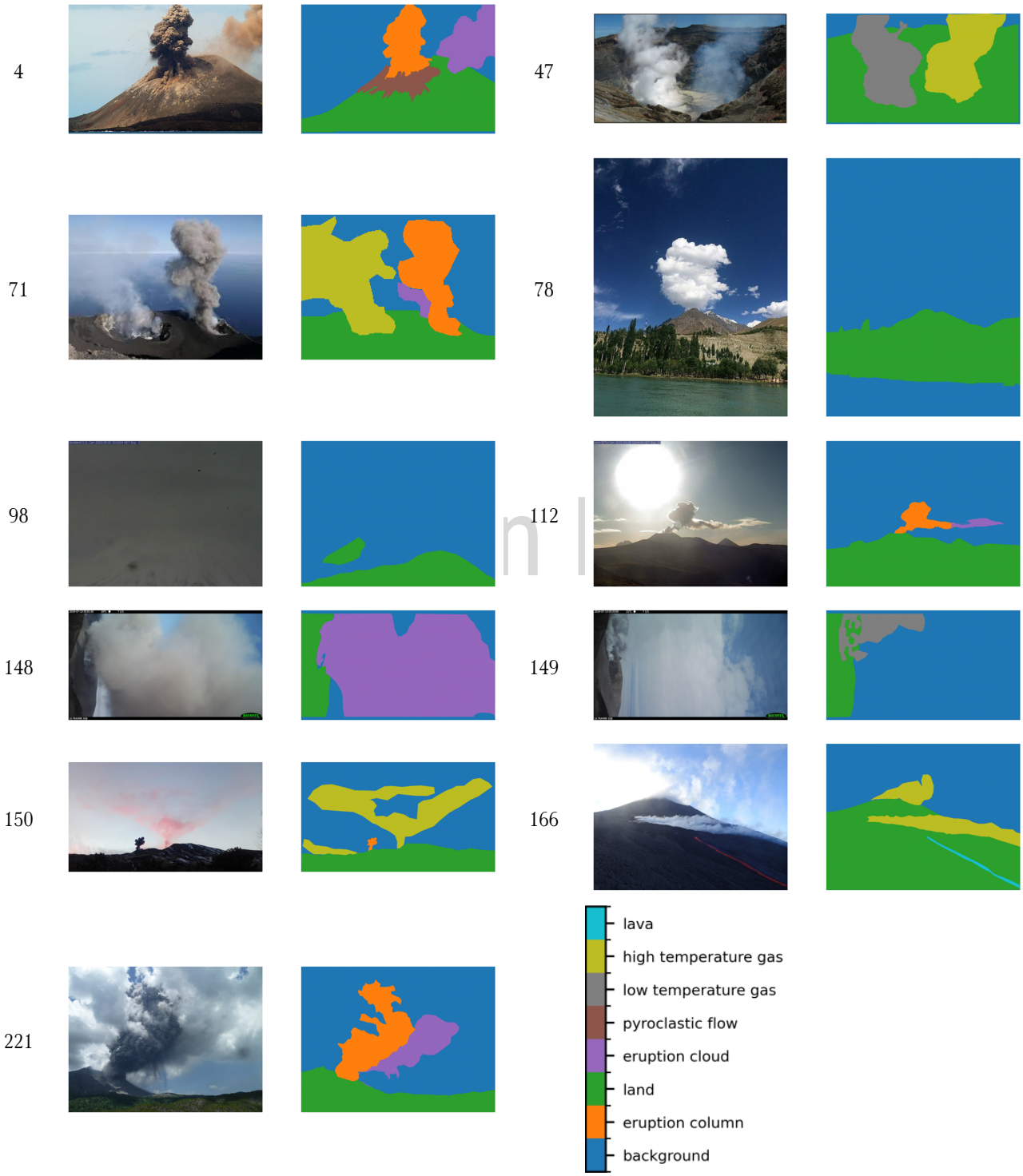


Figure 1: Examples of images and their manual expert labels from the VIGIA-PlumeData training set. The eight classes are: Background (sky, clouds, sea), Land, Eruption column, Eruption cloud, Pyroclastic flow, Low temperature gas (non-ash), High temperature gas (non-ash), and Lava. For the training of the model, we fused the classes Eruption cloud and Pyroclastic Flow into Ash emissions, and Low temperature gas and High temperature gas into Non-ash emissions.

reserved for final evaluation to minimize overfitting. The split was performed randomly but ensured that images from the same camera and viewpoint were not distributed across different sets, allowing us to evaluate generalization across cameras rather than repeated viewpoints. Additionally, three Stromboli monitoring sequences (4–10 frames each) were deliberately included in the test set to prevent the training set from being biased by highly similar images.

The database, as well as the table listing all the image characteristics, can be found in the EasyData repository*.

### 2.2 VIGIA–PlumeNet: Fine-tuning the foundation model DINO–v2–reg for multi-label plume segmentation

Computer vision and image processing have been profoundly transformed by deep learning. The first revolution came in 2012 with the breakthrough performance of convolutional neural networks (CNNs) on the ImageNet challenge, particularly with the AlexNet model [Krizhevsky et al. 2012]. More recently, two new paradigm shifts have reshaped the field:

1. Transformer architectures, originally introduced in natural language processing [Vaswani et al. 2017], later adapted to images via the Vision Transformer (ViT) [Dosovitskiy et al. 2020], which outperformed CNNs on large-scale tasks;

2. Diffusion models, which have demonstrated remarkable capabilities for high-quality image generation [Rombach et al. 2022], but which is not what we are interested in here.

The ViT architecture leverages the attention mechanism to capture relationships between features across an image. However, ViTs typically require substantially more parameters and large-scale datasets, making them unsuitable for training a plume segmentation model from scratch. Nevertheless, ViTs have sparked the rise of foundation models, which can generalize across diverse tasks and domains by learning on massive datasets through self-supervised learning [Kirillov et al. 2023; Oquab et al. 2023].

In this work, we focus on *encoder-based foundation models*. These act as a digital 'eye' that breaks an image into a grid of small squares (called tokens), translating the visual textures of each square into a rich mathematical description of the scene. These descriptions—which capture features like plume density or cloud edges—are then passed to a smaller decoder model. We can think of this decoder as a specialist that interprets the eye's findings to perform a specific task, such as drawing the boundaries of a volcanic plume. While universal models like the Segment Anything Model (SAM [Kirillov et al. 2023]) exist, they are designed to identify every possible object in the world. For our needs, SAM is too broad; we require a decoder specifically trained to distinguish between very similar-looking features, such as volcanic ash and meteorological clouds.

Among current open-source AI models specialized in images, we chose the widely used DINO–v2 [Oquab et al. 2023]. This model is highly robust and available in 'small' versions that require less computing power, making it ideal for real-time use in volcanic observatories. We specifically used a version that includes 'registers'—a technical enhancement that helps the model ignore visual noise and focus more accurately on the main objects in a photo [Darcet et al. 2023]. We used DINO–v2 as a 'frozen' model, meaning we kept its original knowledge (gained from 143 million images) completely intact and unchanged. Instead, we only trained a 'segmentation head': a lightweight, two-layer component that acts as a translator. It takes the general visual patterns recognized by DINO–v2 and converts them into the specific maps we need to identify volcanic plumes (Figure 2). Owing to its modest parameter count (less than 250,000), it can be effectively trained on our limited dataset. While adding additional up-convolutional layers would increase spatial resolution, this comes at the cost of higher parameter complexity and reduced class-level accuracy. For our application, we prioritize accurate class distinction over finer spatial resolution, as precise boundary delineation is less critical than correct class assignment.

We applied data augmentation using the transformations available in the torchvision library, including random rotations (up to 20°), horizontal and vertical flips, color jittering, and random resized cropping with a scaling factor between 0.75 and 1.1. All RGB images and corresponding masks were resized to 518 × 518 pixels. After hyperparameter tuning based on validation performance (manual search from standard parameter values), the final configuration was set to a batch size of 12, a learning rate of $1 \times 10^{-5}$, the AdamW optimizer [Loshchilov and Hutter 2019], and a weighted cross-entropy loss, trained for 75 epochs with early stopping. To address class imbalance, we assigned higher weights to pixels belonging to the *Eruption column* class, due to its importance, and to the *Lava* class, due to its under-representation in the training dataset.

The evolution of the (fine-tuning) training can be analyzed in Appendix A Figure A1 The training of the segmentation head was performed on a single GPU A100 for less than one hour, and the inference takes less than one second for each image, on a laptop with a GPU NVIDIA GeForce RTX 2050. The code repository is available at gricad-gitlab.univ-grenoble-alpes.fr/giffards/vigia-plumenet, the trained model at huggingface.co/SophieGif/VIGIA-PlumeNet. An readily available inference-only version, for CPU or GPU, is also proposed for an easy deployment.

## 3 Results

### 3.1 Quantitative Evaluation

We now evaluate VIGIA–PlumeNet on all 58 images of the test set, which were unseen during training and not used for hyperparameter tuning. Given the limited size of the test set, these metrics should be interpreted as indicative of performance trends rather than definitive statistical benchmarks. First, we perform a quantitative evaluation using a normalized confusion matrix (Figure 3), which summarizes all pixels of the test set by comparing the predicted class for each pixel against the true label. The matrix is expressed in percentages, with each row summing to 100 %. In a perfect model, all values would be 100 % on the diagonal and 0 % elsewhere. We also present in Table 1 the precision, recall, F1-score and Intersection over Union (IoU) on the test set for each class, together with their total number of pixels. These metrics, all between 0 and 1, respectively evaluate the correctness of the generated masks (precision), the avoidance of missed pixels

*waiting for approval, in the mean time, see https://drive.google.com/drive/folders/1Wewtq6ttV6omruL3gbF3iFpoiF-7FJYm?usp=sharing

–Article in Press–

–Article in Press–

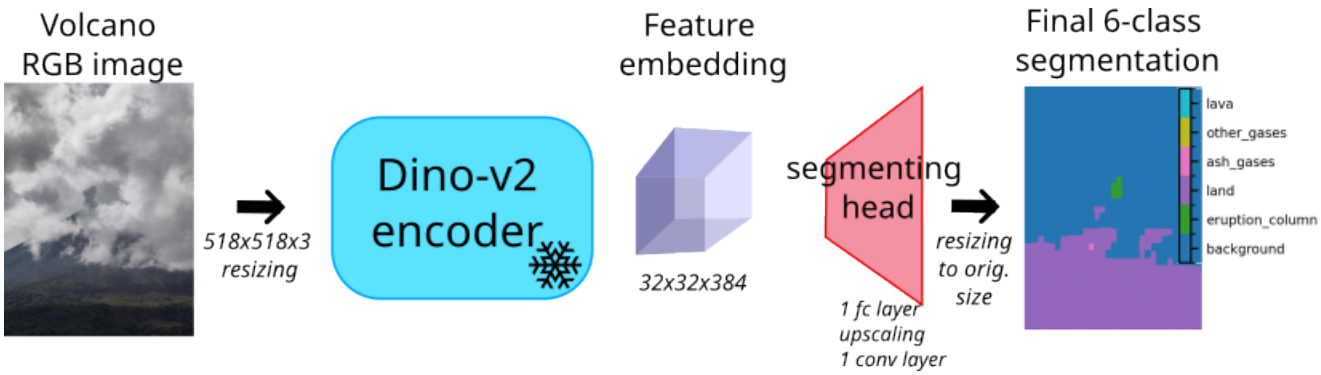


Figure 2: `VIGIA-PlumeNet` model: The foundation model `DINO-v2-reg` small is used as foundation model backbone to extract the feature embedding. During our training, only the segmentation head is optimized while the `DINO-v2` encoder is frozen. The input is any RGB ground-based image, and the output is a six-class segmentation mask of the same size of the original image.

Table 1: Per-class segmentation metrics (on the 58 images of the test set). #GT pixels: total number of test ground truth pixels. IoU: Intersection over Union.

| Class | # GT pixels | Precision | Recall | F1-score | IoU |
|---|---|---|---|---|---|
| Background | 87 339 920 | 0.937 | 0.973 | 0.955 | 0.914 |
| Land | 78 005 643 | 0.973 | 0.987 | 0.980 | 0.960 |
| Eruption column | 3 397 139 | 0.414 | 0.865 | 0.560 | 0.389 |
| Ash emissions | 13 985 150 | 0.861 | 0.351 | 0.499 | 0.332 |
| Non-ash emissions | 2 471 336 | 0.638 | 0.646 | 0.642 | 0.473 |
| Lava | 10 826 | 0.119 | 0.833 | 0.209 | 0.117 |

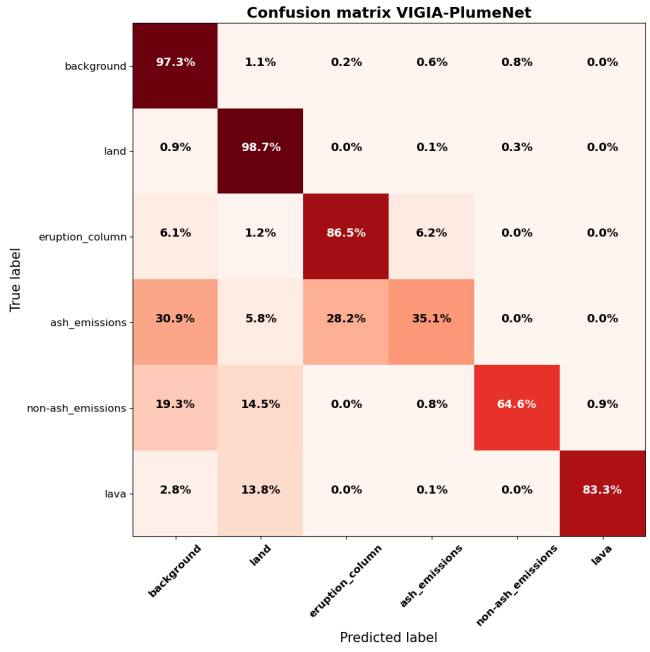


Figure 3: Normalized confusion matrix summarizing all pixels of the 58 images of the test set.

(recall), the overall balance of these two indicators (F1-score), and the area of exact overlap with human annotations (IoU).

Most pixels are correctly classified, with the majority appearing along the diagonal: the total pixel accuracy is 91 %,and it is 88.3 % without taking into account the *Background* class. This is particularly true for *Background* pixels (sky and clouds, 97 % correctly classified), *Land* (99 %), and *Eruption columns* (87 %). Only 7 % of *Eruption column* pixels were misclassified as *Background* or *Land*, mostly corresponding to boundary mismatches. Notably, all eruption columns were detected except in one image (Figure 5, image 180). This can also be seen in Table 1, where *Eruption column* Recall is high (0.86) while Precision is lower (0.41), meaning that the model successfully detects most of the actual eruption column pixels, even if it tends to over-predict (often over the *Ash-emissions* class).

*Ash emissions* (combining eruption ash clouds and pyroclastic flows) and *Non-ash emissions* (combining high and low temperature gases) are less accurately recovered, which is expected given the inherent difficulty and ambiguity in distinguishing these classes and their boundaries. Specifically, 37 % of *Ash emissions* pixels and 33 % of *Non-ash emissions* pixels were misclassified as *Background* or *Land*. While these numbers may seem high, they reflect the fact that the non-column emissions are harder to detect and can be confused for example with clouds or between ash and non-ash emissions. As a result, almost no *Background* or *Land* pixels are wrongly classified as emissions (<1 % each), and the Precision of both *Ash emissions* and *Non-ash emissions* classes is high (0.86 and 0.64) while Recall is smaller (0.35 and 0.64). In practical continuous monitoring, obstructed views and cloud cover occur frequently, and minimizing false alarms over these two classes is crucial for reliable operation.

Additionally, some ash emission pixels are misclassified as eruption columns, and vice versa, although to a lesser extent.

–Article in Press–

–Article in Press–

This is expected, as the boundary between these two classes is continuous, making the labeling inherently difficult. On a positive note, non-ash emissions are less confused with ash emissions or eruption columns, and vice versa, indicating that the model effectively distinguishes between these categories.

*Lava* is also well detected, with 83 % of pixels correctly classified. In Table 1, we can see that the Recall is high (0.83). However, the Precision is low (0.119), meaning that many *Lava* seems wrongly detected: this can be explained by two factors, first, the manual labelling missed some *Lava* pixels, yet we also have some false detected pixels in one image (see following section for more details). However, it is important to note that the total number of lava pixels is relatively small (see Table 1), so these metrics should be interpreted with caution.

### 3.2 Qualitative Evaluation

Figure 4 and Figure 5 present the predictions of VIGIA–PlumeNet on a subset of test images, including particularly challenging examples and error cases. The colors are the same as in Figure 1, except for the classes combining two labels, with a mixture of the corresponding colors. The full set of test results is available online*. The model demonstrates robust performance across a wide variety of conditions and orientations. For instance, in images with obstructed views such as image 93 (Figure 4), where even the land surface is not visible, the presence of volcanic emissions can be easily ruled out. Importantly, difficult cases where eruption columns are partially hidden within meteorological clouds are well detected, as illustrated by images 125 and 161 (Figure 5).

The distinction between (non-column) *Ash* and *Non-ash emissions* is another important aspect of performance. Examples such as images 34, 68, 159, and 169 show that the model is generally effective at this separation. However, the confusion matrix highlights that the precise spatial extent of these emissions remains the most challenging task. This is illustrated by images 23 and 53 (Figure 4), where the boundary between *Eruption column* and (non-column) *Ash emissions* is highly complex. Furthermore, labeling ambiguities contribute to apparent inconsistencies: in image 68, the left-hand cloud was not manually annotated as volcanic emission, although this classification could be debated. Similarly, in image 163, the right-hand region was annotated as *Ash emissions* based on the (hidden) temporal image sequence, although on a single frame it is difficult to rule out that it could be a cloud.

Some limitations of the current model are also evident. The only eruption column not detected, in image 180, is relatively small and with a very horizontal ash cloud. This suggests that the training dataset may contain an under-representation of such plumes, leading to under-detection in such cases. Yet this case is also probably due to the fact that the column as well as the volcano is small, and we mainly see an ash cloud in a background. With respect to *Lava* detection, the model performs strongly in image 255 (Figure 5) and even outperforms manual labeling in image 169 (Figure 5), correctly identifying lava signatures overlooked by human annotators. Conversely, image 100 (Figure 5) reveals a tendency toward over-detection of lava in regions with locally warm colors, such as those observed near sunrise or sunset. This was not apparent in the confusion matrix due to the large number of land pixels over lava pixels. This issue likely arises from the limited number of lava-containing images in the training set and from the consequent intentional weighting of lava pixels in the loss function. This limitation could be mitigated by incorporating a greater number of lava-bearing and low-light examples in training.

Finally, image 255 (Figure 5) also reveals an instance where *non-ash emissions* were not detected despite being clearly visible. This discrepancy may arise from the fact that the image is a long-exposure photograph, a type of input largely absent from the training dataset. The resulting texture of emissions differs substantially from standard imagery, which may have hindered correct identification.

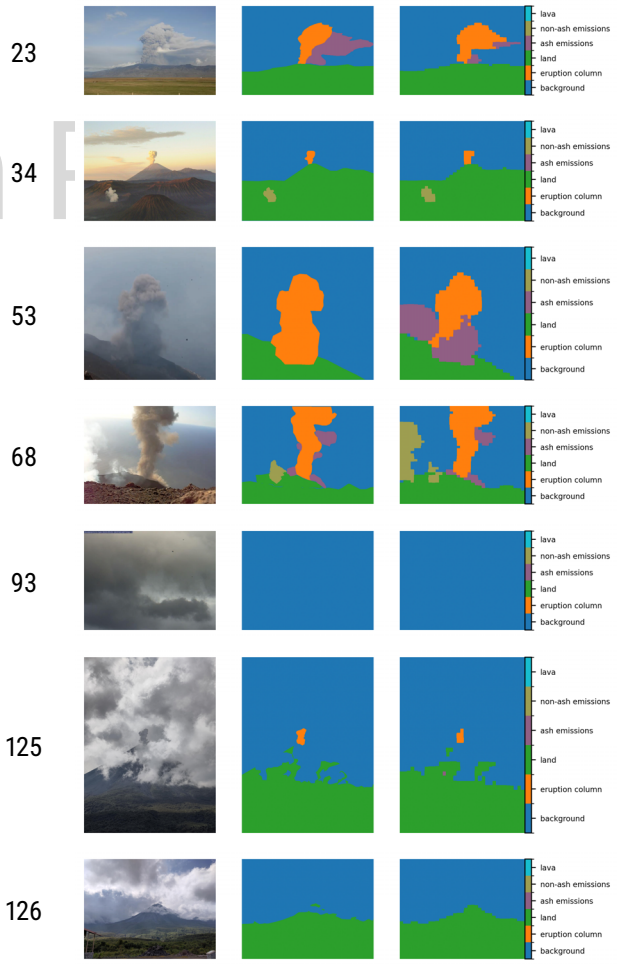


Figure 4: Examples of images of the test set. Left: original image, Middle: ground truth expert labelling, Right: VIGIA–PlumeNet prediction.

### 3.3 Discussion and future work

Overall, while VIGIA–PlumeNet demonstrates strong performance across diverse cases, the examples presented highlight both its strengths and its current limitations, pointing toward future improvements in dataset diversity and extension, as well as annotation consistency. Although the model is de-

*https://huggingface.co/SophieGif/VIGIA-PlumeNet

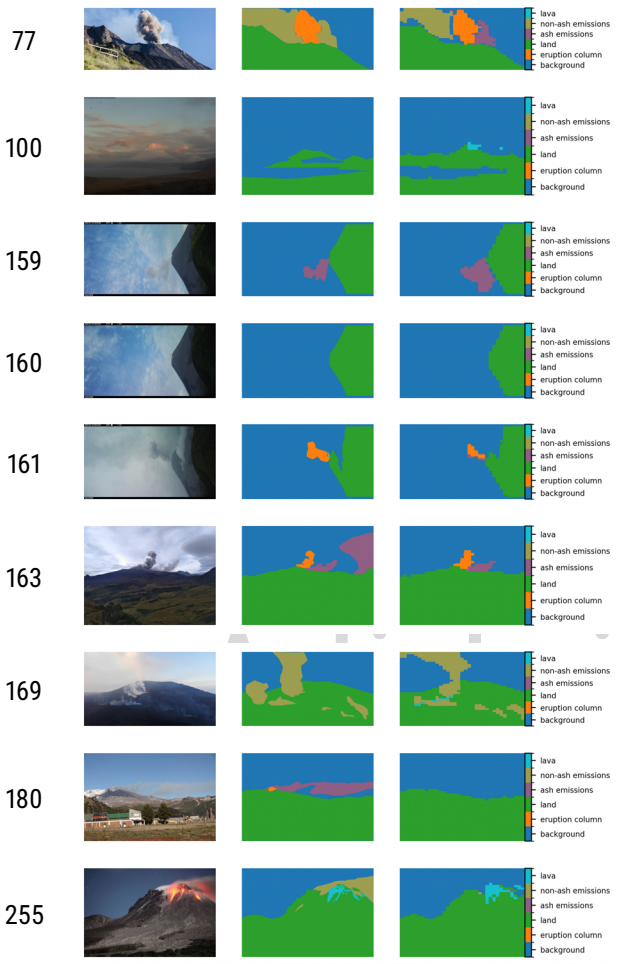


Figure 5: Second set of examples of images of the test set. Left: original image, Middle: ground truth expert labelling, Right: `VIGIA-PlumeNet` prediction.

signed as a generic solution, fine-tuning the segmentation head on camera-specific images—even with only a small number of samples—would likely yield substantial improvements for that particular camera. The code to perform additional training is also available open-access in the git repository. We additionally provide the codes and explanation on how to manually label new images to create additional data.

As this is the first labeled database to be freely available online, we encourage its extension to address current limitations. In particular, adding more examples of lava and pyroclastic flows, as well as images acquired at night or under low-light conditions, would be especially valuable, along with increasing the dataset size more generally. We believe that releasing this database will also foster the development of new models by providing a solid benchmark. Changing the class grouping, such as for example merging *Eruption column* and *Ash clouds*, could be tested in the future. A particularly promising direction would be to enhance the capacity to distinguish between low- and high-temperature emissions, which could bring significant benefits to the community. It could also be very valuable to have several experts labelling the same set of images, in order to compute an inter-expert accuracy metrics, as in Mitchinson et al. [2025] for volcano-seismicity. The rise of new foundation models (such as the recent `Dino-v3` [Siméoni et al. 2025]), as well as text–image multimodal approaches, may offer exciting advances in this area. However, many recent foundation models often come at the cost of computational efficiency and accessibility, potentially limiting their deployment in local observatories. It is also important to note that, while the use of a foundation model supports generalization to unseen environments, `VIGIA-PlumeNet` has been fine-tuned exclusively on volcanic imagery. For this reason, emissions from other sources such as fires or explosions might still be classified as volcanic activity.

It is important to note that each pixel is assigned the class with the highest probability score. The original output of `VIGIA-PlumeNet` is therefore a probability map for all six classes. This allows users to apply custom thresholds for class assignment or to visualize the full probability distribution, enabling a more comprehensive uncertainty analysis. Saving each class confidence as additional output is made available in the open-source code.

More broadly, this work represents only a first step, and several natural extensions remain. First, we did not include low light or near infrared acquisitions, that are able to capture incandescent emissions (such as ballistic) at night, because the acquired image has a very different signature than an RGB image, and would necessitate a specific training. This would be a natural extension on this work. Creating similar labeled datasets for infrared (IR) and ultraviolet (UV) imagery, followed by the development of corresponding models, would also be highly valuable. Moreover, incorporating time-series information could dramatically improve segmentation, as the dynamics of clouds and volcanic emissions differ significantly. Finally, combining multiple modalities (visual, IR, and UV) acquired at the same location—such as in the VIGIA system [Vásconez et al. 2022]—appears to be a promising direction, since these data sources provide complementary perspectives.

Importantly, the multi-class segmentation provided by `VIGIA-PlumeNet` will be beneficial for improving the processing of multi-sensor imagery operating in different wavelengths. In particular, pixels classified as *eruption cloud* could be used improve automated recovery of ash mass eruption rates from infrared images [Cerminara et al. 2015], while pixels classified as *(non-ash) gas plume* could be used to improve automated recovery of $SO_2$ gas flux [e.g. Delle Donne et al. 2019].

## 4 Conclusions

In this work, we presented `VIGIA-PlumeNet`, a deep learning model capable of segmenting multiple volcanic emission classes from ground-based optical images. To support this development, we compiled and manually annotated a diverse dataset VIGIA-PlumeData. Our quantitative and qualitative evaluations demonstrate that `VIGIA-PlumeNet` achieves high pixel-level accuracy (88 % when excluding the background class) and effectively distinguishes between eruption columns, ash emissions, non-ash emissions, lava, land and background (accounting clouds and sky). The model performs particularly well in detecting and separating eruption column pixels from

–Article in Press–

clouds, while the extents of ash and non-ash emissions remain the most challenging, reflecting the inherent ambiguity in labeling these classes. The probability-based output of the model allows for flexible thresholding and uncertainty analysis, making it suitable for continuous monitoring applications where false alarms must be minimized.

Overall, VIGIA–PlumeNet provides a first step toward automated, multi-class segmentation of volcanic emissions from optical imagery. Future work could expand the database size, in particular with additional challenging conditions, integrate multi-modal observations (thermal, UV, and visible imagery), and extend the model to track plume dynamics over time. We believe that this open-source resource will facilitate the development of more robust, generalizable tools for volcano monitoring, ultimately supporting hazard assessment and early warning systems worldwide.

## Author Contributions

Conceptualization: SG; Data Collection: TW, SV, RC, SG, FV; Data manual labelling: SG, YM, ER, TW; Methodology: SG; Software: SG; Validation: YM, ER; Writing - orginal draft: SG; Writing - proofreading and editing: SV, YM, TW; Project Administration: SG; Funding Acquisition: SG, YM.

## Acknowledgements

We would like to thank all the people that allowed their pictures to be part of this database, in particular the ones allowing a large quantity: Thomas R. Walter, Tom Pering and Richard Roscoe (http://www.photovolcanica.com). All the image credits can be found in the metadata spreadsheet that goes along with VIGIA-PlumeData at www.easydata.earth. S. G. gratefully acknowledges financial support for this research by the Fulbright U.S. Scholars Program, which is sponsored by the U.S. Department of State, the Franco-American Fulbright Commission and Université Grenoble Alpes. Its contents are solely the responsibility of the author and do not necessarily represent the official views of the Fulbright Program, the Government of the United States, or the Franco-American Commission. Y.M. acknowledges support from the US National Science Foundation under Award No. EAR-2240650. This work has been partially supported by MIAI @ Grenoble Alpes (ANR-19-P3IA-0003 and ANR-23-IACL-0006). S.V. acknowledges support from UNAM PAPIIT grant No. IN114625.

## Data Availability

Data, codes and results are publicly shared. VIGIA-PlumeData is stored in the www.easydata.earth repository (sent for review, waiting to be approved, in the mean time it is accessible at https://drive.google.com/drive/folders/1Wewtq6ttV6omruL3gbF3iFpoiF-7FJYm?usp=sharing), the VIGIA–PlumeNet codes at https://gricad-gitlab.univ-grenoble-alpes.fr/giffards/vigia-plumenet, the trained model and the test results at https://huggingface.co/SophieGif/VIGIA-PlumeNet.

## Copyright Notice

© The Author(s) 2026. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

Aravena, A., G. Carparelli, R. Cioni, M. Prestifilippo, and S. Scollo (2023). "Toward a Real-Time Analysis of Column Height by Visible Cameras: An Example from Mt. Etna, in Italy". *Remote Sensing* 15(10), page 2595. DOI: 10.3390/rs15102595.

Barnie, T., T. Hjörvar, M. Titos, E. M. Sigurðsson, S. K. Pálsson, B. Bergsson, Þ. Ingvarsson, M. A. Pfeffer, S. Barsotti, Þ. Arason, et al. (2023). "Volcanic plume height monitoring using calibrated web cameras at the Icelandic Meteorological Office: system overview and first application during the 2021 Fagradalsfjall eruption". *Journal of Applied Volcanology* 12(1), page 4. DOI: 10.1186/s13617-023-00130-9.

Bombrun, M., D. Jessop, A. Harris, and V. Barra (2018). "An algorithm for the detection and characterisation of volcanic plumes using thermal camera imagery". *Journal of Volcanology and Geothermal Research* 352, pages 26–37. DOI: 10.1016/j.jvolgeores.2018.01.006.

Centeno, R., V. Gómez-Salcedo, I. Lazarte, J. Vilca-Nina, S. Osores, and E. Mayhua-Lopez (2024). "Near-real-time multiparametric seismic and visual monitoring of explosive activity at Sabancaya volcano, Peru". *Journal of Volcanology and Geothermal Research* 451, page 108097. DOI: 10.1016/j.jvolgeores.2024.108097.

Cerminara, M., T. Esposti Ongaro, S. Valade, and A. J. L. Harris (2015). "Volcanic Plume Vent Conditions Retrieved from Infrared Images: A Forward and Inverse Modeling Approach". *Journal of Volcanology and Geothermal Research* 300, pages 129–147. DOI: 10.1016/j.jvolgeores.2014.12.015.

Darcet, T., M. Oquab, J. Mairal, and P. Bojanowski (2023). "Vision transformers need registers". *arXiv preprint arXiv:2309.16588*.

Delle Donne, D., A. Aiuppa, M. Bitetto, R. D'Aleo, M. Coltelli, D. Coppola, E. Pecora, M. Ripepe, and G. Tamburello (2019). "Changes in SO2 Flux Regime at Mt. Etna Captured by Automatically Processed Ultraviolet Camera Data". *Remote Sensing* 11(10), page 1201. DOI: 10.3390/rs11101201.

Dosovitskiy, A., L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. (2020). "An image is worth 16x16 words: Transformers for image recognition at scale". *arXiv preprint arXiv:2010.11929*.

Dürig, T., M. T. Gudmundsson, F. Dioguardi, M. Woodhouse, H. Björnsson, S. Barsotti, T. Witt, and T. R. Walter (2018). "REFIR- A Multi-Parameter System for near Real-Time Estimates of Plume-Height and Mass Eruption Rate during Explosive Eruptions". *Journal of Volcanology and*

–Article in Press–

Article in Press

*Geothermal Research* 360, pages 61–83. DOI: 10.1016/j.jvolgeores.2018.07.003.

Guerrero Tello, J. F., M. Coltelli, M. Marsella, A. Celauro, and J. A. Palenzuela Baena (2022). "Convolutional neural network algorithms for semantic segmentation of volcanic ash plumes using visible camera imagery". *Remote Sensing* 14(18), page 4477. DOI: 10.3390/rs14184477.

Kirillov, A., E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, et al. (2023). "Segment anything". *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026. DOI: 10.1109/ICCV51070.2023.00390.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). "Imagenet classification with deep convolutional neural networks". *Advances in neural information processing systems* 25. DOI: 10.1145/3065386.

Loshchilov, I. and F. Hutter (2019). "Decoupled Weight Decay Regularization". *International Conference on Learning Representations (ICLR)*.

Mitchinson, S., J. H. Johnson, B. Milner, O. Lamb, and Y. Behr (2025). "Capturing expert uncertainty: ICC-informed soft labelling for volcano-seismicity". *Bulletin of Volcanology* 87(10), page 84.

Nunnari, G. and S. Calvari (2024). "Exploring convolutional neural networks for the thermal image classification of volcanic activity". *Geomatics* 4(2), pages 124–137. DOI: 10.3390/geomatics4020007.

Oquab, M., T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, et al. (2023). "Dinov2: Learning robust visual features without supervision". *arXiv preprint arXiv:2304.07193*.

Rombach, R., A. Blattmann, D. Lorenz, P. Esser, and B. Ommer (2022). "High-resolution image synthesis with latent diffusion models". *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695. DOI: 10.1109/CVPR52688.2022.01046.

Ronneberger, O., P. Fischer, and T. Brox (2015). "U-net: Convolutional networks for biomedical image segmentation". *International Conference on Medical image computing and computer-assisted intervention*. Springer, pages 234–241. DOI: 10.1007/978-3-319-24574-8_28.

Siméoni, O., H. V. Vo, M. Seitzer, F. Baldassarre, M. Oquab, C. Jose, V. Khalidov, M. Szafraniec, S. Yi, M. Ramamonjisoa, et al. (2025). "Dinov3". *arXiv preprint arXiv:2508.10104*.

Simionato, R., P. A. Jarvis, E. Rossi, and C. Bonadonna (2022). "PlumeTraP: A new MATLAB-based algorithm to detect and parametrize volcanic plumes from visible-wavelength images". *Remote Sensing* 14(7), page 1766. DOI: 10.3390/rs14071766.

Valade, S. A., A. J. Harris, and M. Cerminara (2014). "Plume Ascent Tracker: Interactive Matlab Software for Analysis of Ascending Plumes in Image Data". *Computers and Geosciences* 66, pages 132–144. DOI: 10.1016/j.cageo.2013.12.015.

Vásconez, F., Y. Moussallam, A. J. Harris, T. Latchimy, K. Kelfoun, M. Bontemps, C. Macıas, S. Hidalgo, J. Córdova, J. Battaglia, et al. (2022). "VIGIA: A thermal and visible imagery system to track volcanic explosions". *Remote Sensing* 14(14), page 3355. DOI: 10.3390/rs14143355.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin (2017). "Attention is all you need". *Advances in neural information processing systems* 30. DOI: 10.48550/arXiv.1706.03762.

Wilkes, T. C., T. D. Pering, and A. J. McGonigle (2022). "Semantic segmentation of explosive volcanic plumes through deep learning". *Computers & Geosciences* 168, page 105216. DOI: 10.1016/j.cageo.2022.105216.

Wilson, L., R. Sparks, T. Huang, and N. Watkins (1978). "The control of volcanic column heights by eruption energetics and dynamics". *Journal of Geophysical Research: Solid Earth* 83(B4), pages 1829–1836. DOI: 10.1029/JB083iB04p01829.

Witsil, A. J. and J. B. Johnson (2020). "Volcano video data characterized and classified using computer vision and machine learning algorithms". *Geoscience Frontiers* 11(5), pages 1789–1803. DOI: 10.1016/j.gsf.2020.01.016.

## Appendix A

We can see in Figure A1 the evolution of the training loss (weighted cross-entropy), the validation accuracy (except background pixels) and the mean Intersection over Union (IoU) over the training steps of the segmentation head (fine-tuning `Dino-v2`). We can see that the validation metrics do improve during the 75 epochs.

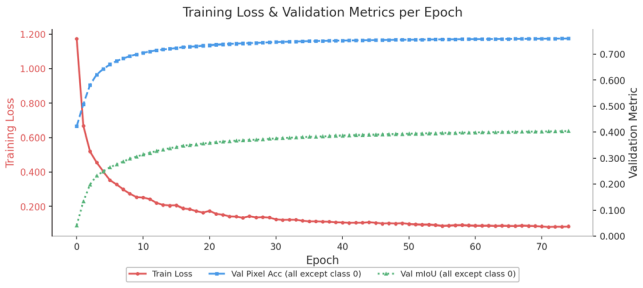


Figure A1: Evolution of the training loss, the validation accuracy (except the background pixels) and the mean Intersection over Union (IoU) over the (fine-tuning) training steps.

–Article in Press–