RESEARCH METHODS VOLCANICA

Case study of deep learning image segmentation for the purposes of rapid 2D petrographic analysis in volcanic rocks

Brenna A. Halverson*[α,β], Matthew Loewen[γ], Hannah Dietterich[γ], and Alan Whittington[α]

[α]The University of Texas at San Antonio, 1 UTSA Circle, San Antonio, TX, 78249, USA.
[β]Durham University, Stockton Road, Durham, DH1 3LE, UK.
[γ]U.S. Geological Survey, Alaska Volcano Observatory, Anchorage, AK, 99508, USA.

Abstract

Automation using deep learning methods is a useful alternative to manual methods of petrographic segmentation, but often requires user familiarity with coding and/or algorithms. We examine the Dragonfly™ program's deep learning tools for application by users with a variety of skill levels as a method for petrographic image segmentation. An image processing methodology, bimodal image stacking, was created for low-input-data, high-efficacy training of models which can then be applied to varied samples. Using backscatter electron images we show that the resulting model segmentations agree with manual segmentation total and modal crystallinity values within 5 %, and calculated plagioclase crystal size distribution (CSD) values within 2σ, despite limitations in discriminating mafic phases. Model creation and training takes <24 hours, 1–3 hours of which are supervised, and the resultant model can then be applied to new uncharacterized samples in <15 minutes per image. This allows for non-experts to create and utilize deep learning models to segment images of variable brightness and texture, at low user-time cost and resulting in size and shape data which are within uncertainty of manual segmentation. While some limitations are noted (for example, sieve-textured phases may need manual correction, and different minerals with similar BSE intensity may not be resolved as separate phases), this methodology can be utilized for general application of models to wide ranges of volcanic crystalline and bubble textures, and to create a library of models for rapid petrological analysis during volcanic eruptions.

Keywords: Comet Dragonfly™; Deep learning segmentation; Bogoslof volcano; CSD; Basalts.

Submitted: 2024-07-29 | Accepted: 2025-06-05 | Published: 2025-10-05

Editor: A. Hornby | Typesetter: J. Farquharson

1 Introduction

The characterization of bubbles and crystals in volcanic products provides important information about pre-, syn-, and post-eruptive magmatic systems and processes. Of particular importance are crystal size and aspect ratio, which record eruption dynamics [e.g. Wright et al. 2007; Mujin and Nakamura 2014; Cashman 2020], thermal and degassing history, and disequilibrium in the volcanic system [e.g. Wallace and Carmichael 1994]. These parameters are traditionally calculated from segmentation of crystal and bubble phases in two- and three-dimensional images [Cashman 1988; Innocenti et al. 2013; Valdivia et al. 2022; Preece et al. 2023] captured from reflected/transmitted light microscopy, backscattered electron imaging (BSE), compositional mapping by energy or wavelength dispersive spectroscopy (EDS/WDS), or computed tomography (CT). Segmentation can be readily conducted on chemical maps obtained by EDS or WDS [e.g. Higgins et al. 2021; Sheldrake and Higgins 2021; Halverson et al. 2024]; however, these methods require significant instrument time and can be financially costly, resulting in smaller datasets which provide less robust statistics on the overall sample. In most cases, the segmentation of these image types is usually done by image thresholding or manual delineation.

Image thresholding of BSE or reflected light images is a quick and effective way of segmenting crystal phases in crystal-poor samples with large amounts of grayscale contrast between crystal phases, glass, and bubbles. However, many samples contain phases which have a similar mean atomic number to the surrounding glass, or to other phases, or may show internal chemical changes within the crystals themselves. These factors can result in poor separation by grayscale thresholding and commonly require manual segmentation (outlining and infilling of crystals using tracing or drawing tools in image editing software), and a significant time investment. Phase segmentation in volcanic materials is thus an ideal area to apply deep learning techniques to automate time-consuming manual tasks.

While deep learning applications for phase segmentation in geologic samples are becoming more common, applications to volcanic samples are rare, often limited to individual phases [e.g. Leichter et al. 2022], segmentation using grayscale values [e.g. Lormand et al. 2018], or encompassing whole ash grains [Benet et al. 2024]. A methodology which provides users of varying skill-levels the ability to apply models to wide ranges of volcanic crystalline and bubble textures, utilizing smaller training datasets which are obtainable from limited scanning electron microscope (SEM) time, or reflected light microscopy, would be highly valuable for both general use and rapid textural analysis in volcanic eruption response settings [e.g. Ganseki et al. 2019].

Here we investigate the utility of a free-for-non-commercial-use graphical user interface tool for deep learning segmentation of rapidly acquired BSE images in Dragonfly™. We apply this to a case study from the 2016–2017 eruption of Bogoslof volcano, Alaska, which produced an extended series of violent explosions and effusive periods with a variety of crystalline eruptive products from basaltic to trachytic composition.

*halversonbrenna@gmail.com

1.1 Imaging and machine learning background

Image segmentation for crystal and bubble morphological analyses has been utilized in geology for decades [e.g. Cashman 1988; Hammer et al. 1999; Blundy and Cashman 2008; Shea et al. 2010; Iezzi et al. 2011; Lanzafame et al. 2013; McClinton et al. 2014; Yasuda and Hokanishi 2022; Preece et al. 2023]. This is done primarily using microscope images, either from transmitted/polarizing/reflected light or SEM. Both techniques result in images where crystal phases can be distinguished from one another and any surrounding glass by a variety of factors, including brightness, contrast at crystal edges, and internal and external morphology. In SEM BSE imaging of samples containing crystal and glass phases with large differences in mean atomic number, grayscale thresholding segmentation is a fast and easy method of crystal delineation. In this scenario, crystal and bubble phases are easily separated from any interstitial glass, and one another, allowing for determination of their size and shape parameters [e.g. Shea et al. 2010].

However, many samples contain crystals which are complexly zoned or sieved, and/or phases which have similar atomic-number grayscale values to other phases, including surrounding glass. This complexity is exacerbated in reflected light imaging, as the difference in contrast is not determined by chemical characteristics of the mineral phases, but by their color and reflectivity. As such, in both SEM and reflected light imaging, it is common to have pixel values which crossover between phases, causing overlapping segmentation. Simple thresholding will then result in large errors in phase selection, as certain areas are assigned to the wrong phase based on grayscale values alone.

While manual segmentation is widely used to combat this issue, it is highly time intensive, resulting in either low robustness in statistics due to small amounts of data collected, or to a large amount of time designated to this task. In these cases, deep learning can be very useful.

The term "deep learning" applies to a subset of machine-learning algorithms which utilize multi-layered neural networks to enhance the ability of computers to learn from data inputs [Sarker 2021]. For the purposes of this paper, we focus on semantic segmentation models, a subset of deep learning models. These are built specifically for image classification and have been used to segment portions of geological samples based on chemistry, morphology, and spatial relationships [e.g. Iglesias et al. 2019; Chen et al. 2020; Koh et al. 2021; Liu et al. 2021; Di Martino et al. 2023; Jutzeler et al. 2024]. This is an automated segmentation method using supervised classification to train a neural network to identify objects, such as for medical imaging or computer vision. This allows discrimination of phases of interest both spectrally (color, chemistry) and spatially (shape, spatial relationships). Segmentation is usually kept to 2–3 phases, and often requires the use of hundreds to thousands of training frames (subsets within images or entire images within larger datasets) for robust training. This requires large amounts of user time to manually label training data, as well as time and computational expense in the training and application of the models [e.g. Chen et al. 2020; Koh et al. 2021; Liu et al. 2021; Malik et al. 2022]. For rapid volcanic textural characterization, however, even with the potential for higher uncertainty, it would be useful to be able to train and apply deep learning models quickly from limited, diverse datasets.

1.2 Geological background

We use the Bogoslof volcano 2016–2017 eruption as a case study because of the diversity of eruptive products and the critical nature of magma textures in the components of the eruption and regulating eruption style. Bogoslof volcano is an emergent back-arc volcano in the Aleutian Islands whose nine-month long eruption in 2016–2017 consisted of a series of explosive episodes along with occasional short-lived periods of dome growth [Waythomas et al. 2020a; b]. The complex nature of the eruptive sequences, as well as the explosive nature and predominantly basaltic composition [Loewen et al. 2019], provides a series of interesting textural components with which to test our models.

Samples from throughout the eruption were collected afterward by Alaska Volcano Observatory scientists in 2018 and 2019. These were composed of distinct chemical and morphological endmembers. Loewen et al. (2019) described three different components of large lapilli/block size: 1) basaltic scoria, 2) light colored trachyte composition pumice, and 3) dense crystalline older lavas of intermediate trachyandesite composition.

Here we focus on the dense and scoriaceous basaltic large lapilli and blocks, juvenile samples whose textural characteristics may record information about storage and eruption conditions. Four distinct samples were chosen which covered a range of textural characteristics, whilst containing the same crystal phases (Table 1; Figure 2). These contain plagioclase, clinopyroxene, orthopyroxene, amphibole, and oxide phenocrysts, as well as vesicles and glass bearing microlites of all phases with additional trace rounded olivine microlites Coombs et al. [2019] and Loewen et al. [2019]. The variety of size and shape for each phase across the four samples creates a framework for determining best practices in developing a segmentation methodology, which will allow for training and application of low-data, high efficacy models for rapid image segmentation and petrological characterization.

2 Methodology

The workflow presented here is a multi-step procedure which consists of the image acquisition and the bimodal image stacking pre-processing methodology, set-up and training of models in `Dragonfly`™, selection of models, and analysis of model segmentations (Figure 1). A detailed description and standard operating procedure (SOP) can be found in the supplemental files for those who wish to deploy these techniques (Supplemental Material 1a,b – SOP).

2.1 Image acquisition

2.1.1 *Sample choice and preparation*

We chose four samples to use as training data for this study, 18CW100-10, 18CW100-16, 18CW100-22, and 18CW100-30 (Table 1). These samples are representative of the basaltic textures produced by this eruption. The samples all have the

Table 1: List of samples from the Bogoslof volcano 2016–2017 eruption used in this study. The first four were used for training and validation of subsequent modal statistics of crystals, while the last two were used to determine model validity in entirely new samples. See Supplementary Material 3 for schematics of where each image subset is positioned in original mosaic.

Sample name	Sample type	Subset	
		used for modal abundances, aspect ratio, and area	used for CSD calculations
18CW100-10	Basaltic scoria with entrained clasts	Image 1a	Image 1b
18CW100-16	Basaltic scoria	Image 2a	Image 2b
18CW100-22	Dense scoria with intermediate enclaves	Image 3	Image 3
18CW100-30	Dense basalt	Image 4	Image 4
18CW100-11	Dense basalt with vesicular enclave	N/A - Samples used to test model on images dissimilar to those used to train the model	
18CW100-13	Dense basalt with dixytactic enclave		

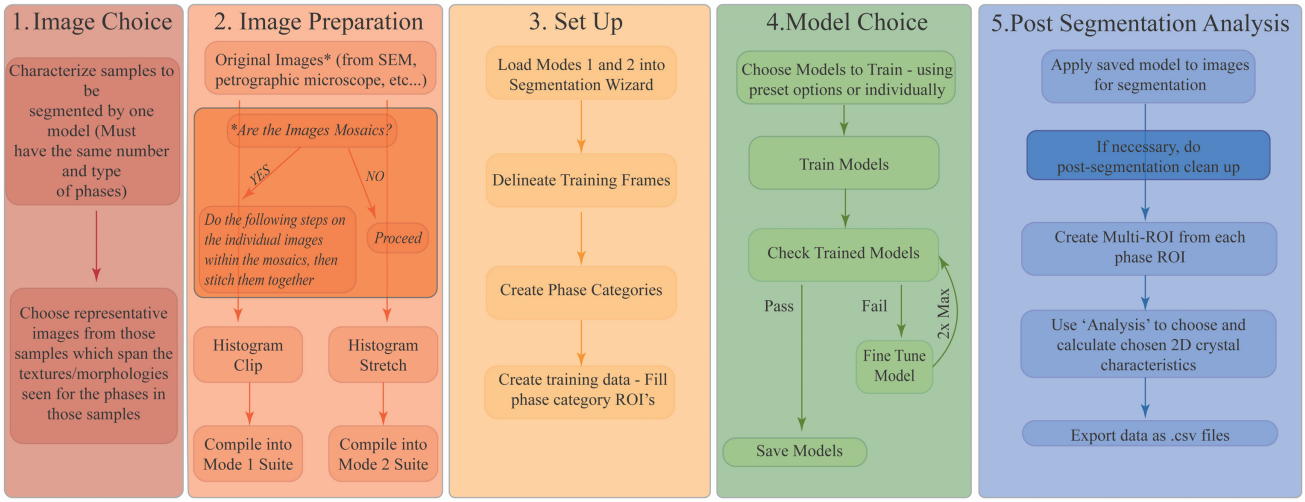


Figure 1: Schematic workflow from sample choice to data analysis of the methodology presented herein.

same assemblage (six phases in BSE images: vesicles, glass, and four distinguishable crystalline phases), but span the varied textures of those phases seen in the samples (intergrowths, change in size, etc.). This sample set provides a robust suite of training imagery for the model, which will enable it to be applied to a wider range of samples with the same phase assemblage.

Cylindrical cores of representative portions of the samples, 1.5 cm wide by 1–3 cm tall, were mounted in epoxy, then cut into 1.5 cm × 8 mm disks. These disks were then polished and a 22 nm carbon coat applied for SEM use. While disks were utilized during this work, any polished sample (thin section, grain mount, etc.) which can be imaged at sufficient resolution can be used.

2.1.2 *Imaging criteria*

We collected images using the BSE detector on a JEOL JSM-6510 LV Scanning Electron Microscope at the U.S. Geological Survey Alaska Volcano Observatory's Tephra Lab. Each image was taken with an accelerating voltage of 20 kV, a working distance of 12 mm, spot size (~beam current) of 65, and a resolution of 0.21 µm per pixel. Individual image frames collected were 4096 by 3072 pixels in size, and we merged 25–35 images using the `Automated Photomerge` tool in `Adobe Photoshop`® to create a mosaic with an average total image area of ~9.5 mm^2.

Where possible, brightness and contrast were adjusted during imaging to provide either the highest contrast between phases at the time of imaging without saturation of any phases, or skewing the image toward darker values, which could be more easily modified post-processing. Fluctuation in beam current over long imaging sessions resulted in small shifts in the contrast and brightness between images, requiring additional post-processing grayscale normalization for mosaics (detailed in Supplementary Material 1a,b – SOP).

While BSE images were used for this work, images obtained by any microscopy imaging technique can be used, including polarized and reflected light microscopy [Halverson and Whittington 2025].

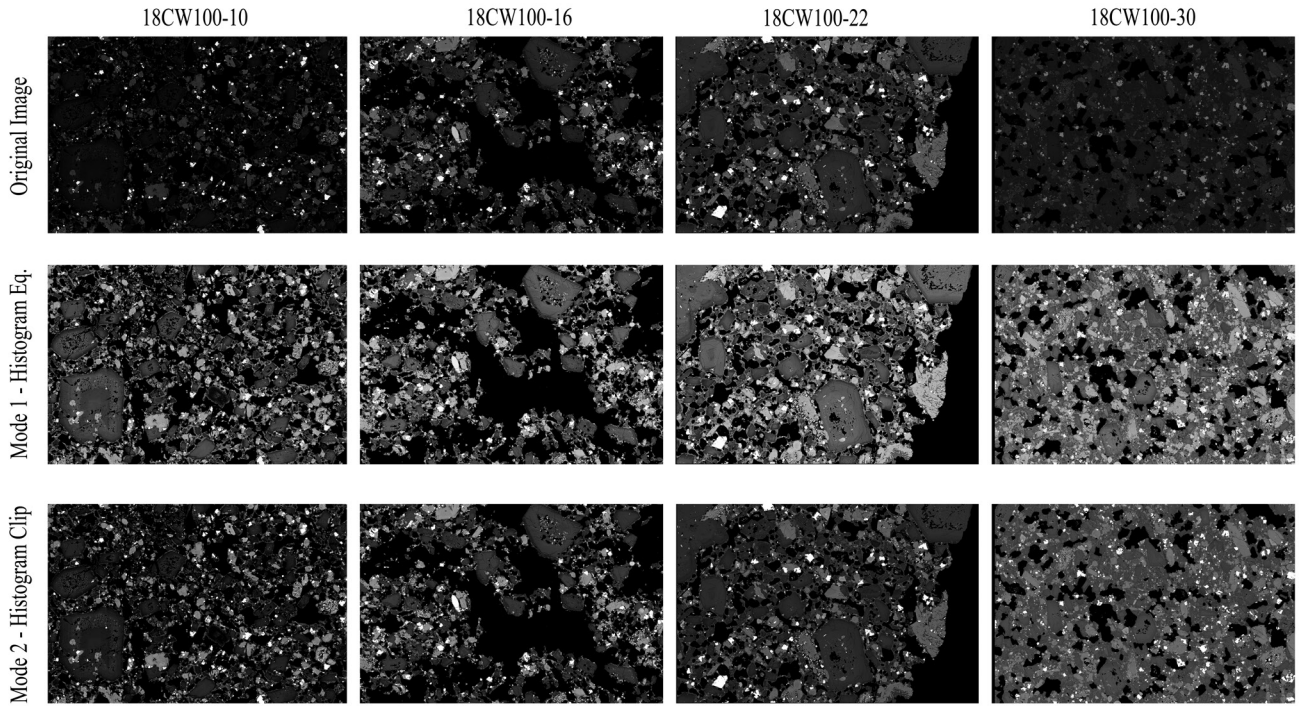


Figure 2: Example of the Bimodal Image Stack architecture. Each image is 2.7 × 2.5 mm in size, for an area of ~7 mm^2.

2.2 Image pre-processing

Due to possible variation between individual images in brightness, contrast, and grayscale values, whether from different instruments or users, or instrumental fluctuations, pre-training processing of images is necessary to ensure models can be applied to a wide range of images. We developed a methodology of image selection and pre-processing the bimodal image stack, which allows for the use of non-ideal images in segmentation by rapidly trained models (Full SOP of the following methodology is located in Supplemental Material 1a,b).

2.2.1 *Image selection*

A suite of images (hereafter referred to as the 'original' images) should be chosen for training which encompass the full range of phases, grayscale values, and textures seen in the samples and images the user wishes to segment with the subsequent model (Figure 2, first row). More training images will result in more robust models; however, the user should weigh the time and computational constraints prior to creation. Larger datasets require more time to create training images, which is done manually, and result in longer training times for the models.

We selected four images, one from each sample, which spanned large enough areas to provide statistical information on crystal characteristics for model training and validation purposes, and encompassed all textures seen throughout the samples. Using these, we created the bimodal image stack images using the following steps carried out using the `FIJI` ® software [Schindelin et al. 2012, `ImageJ 1.54f`].

2.2.2 *Histogram equalization*

Histogram equalization is the task by which the brightness histogram of an image is modified to have an equal distribution across the whole range of grayscale values. Because most images of the same material will have similar relationships between the grayscale values of each phase, regardless of the brightness of the image, this increases the contrast between phases.

Using the original image suite, we performed a histogram equalization on each image in the suite separately. For those images which were mosaics, we applied the histogram equalization separately on each frame within the mosaic before stitching them together. The suite of images is then compiled into the Mode 1 suite for model training (Figure 2, second row).

2.2.3 *Histogram clipping*

Histogram equalization is very reproducible, and can enhance contrast, but often results in aberrations in mosaicked images where slight changes in brightness and contrast during collection causes variable brightness and contrast across the image. Additionally, this equalization process shifts more pixels into the higher and lower grayscale values, which can reduce contrast at the extremities of the grayscale. To counter this shift, a second suite of images is created using manual histogram stretching to clip maximum (brightest) and minimum (darkest) pixel values. In our samples, the brightest pixel values belong to Fe-Ti oxides, whereas the darkest pixel values belong to vesicles. We ensured that neither of these phases were over- nor undersaturated during imaging. This provides

a reproducible process to apply to other images that enhances contrast between phases near the grayscale extremities.

As with the histogram equalization, histogram clipping is performed on each image from the original suite separately, and to frames within mosaics prior to stitching into the mosaic. The newly clipped images are then compiled into the Mode 2 suite for model training (Figure 2, third row).

The two sets of images described above are compiled to become the two inputs for model training, which doubles the training data created in essentially the same amount of time.

2.3 Development of image segmentation models

Deep learning often requires significant pre- and post-processing of data, and a working knowledge of the vocabulary of machine learning applications, and their associated code. The `Dragonfly`™ software (version 2022.2 used here; only available for Windows and Linux; link in Supplementary Material 1a,b – SOP to the newest 2024.1 version) implements multiple tools which allow for easy utilization of multiple types of machine- and deep learning models [Makovetsky et al. 2018]. In particular, the `Segmentation Wizard` feature allows users to create, train, and validate models with minimal user knowledge or input. For non-expert users, sets of models can be chosen by user-specified characteristics, such as 'sparse training data' or 'high accuracy.' Each option includes certain models which are presented to the user as a package, which can be trained directly as a set unless otherwise specified.

Here, we use the `UNET` and `UNET++` model architectures [Ronneberger et al. 2015; Zhou et al. 2018]. These models, which were written for use in medical imagery, are particularly useful in image segmentation, as their U-shaped construction allows for convolution of the original image to determine 'occurrence' of the assigned objects, but then deconvolution back to the same resolution as the original, effectively providing a 'where' to the objects, rather than just an 'if.' Additionally, this model augments image frames during training with variations in elastic deformation (distortion of lines in the images) and scaling (changing in scale of the images) to increase robustness of the output models [Ronneberger et al. 2015; Zhou et al. 2018], which is extremely useful for small datasets. While the elastic deformation aspect may be considered problematic for segmentation of crystals with particular morphologies, we did not find this to be an issue.

2.3.1 *Creation of training frames*

In the `Segmentation Wizard`, training frames are created from the input images, and all pixels within those frames are assigned to a 'class' (discernable crystal types, vesicles, glass, etc.) chosen by the user. For the best outcomes, each training frame should contain all phases to be segmented. Three training frames (each ~0.1 to 0.25 mm^2) per image was the minimum number that produced acceptable results from training the models. This may vary depending on size of training frames, feature size, and the size of the images themselves, but for the images used here, three was determined to be the most efficient number of frames for the desired outcome (see Supplementary Material 1a,b – SOP for more details).

In each sample used here, the images displayed four discernable crystal phases: a) a dark tabular crystal, b) a bright white crystal of relatively equant shape, c) a lighter medium-gray phase, and d) a darker medium-gray phase. To determine exact phases present, we used point EDS data in each of these phases. The dark tabular phase was determined to be plagioclase and the bright white phase Fe-Ti oxide.

However, the two medium-gray phases are more ambiguous. While the spot checks of three random crystals from each indicated the lighter of the two was orthopyroxene and the darker of the two was clinopyroxene, cleavage and habit of some crystals within the darker phase indicated they were amphibole. This is expected from these samples, as past research has determined 3–10 % amphibole in basalts from this eruption [Loewen et al. 2019]. Additionally, Loewen et al. [2019] indicated that some olivine microlites were possible, which would group under the lighter medium-grey phase. For the purposes of creating training data, it was impossible to distinguish between all the amphibole and clinopyroxene crystals where neither distinct habit nor cleavage were visible, as well as between olivine or orthopyroxene microlites. As such, the medium-gray phases are simply referred to as Mafic 1 (the darker medium-gray phase containing both amphibole and clinopyroxene) and Mafic 2 (the lighter gray phase containing orthopyroxene and possible olivine microlites).

These four discernible crystal phases (Plagioclase, Oxides, Mafic 1, and Mafic 2) were segmented manually in all training frames, along with vesicles and glass, to ensure all pixels within the frames were allocated to one of the crystal phases, glass, or vesicles.

2.3.2 *Training*

Once the frames are finalized, model architectures can be chosen. We trained `UNET` and `UNET++` models of 32–128 initial filter count and 5–7 levels and found that filter sizes of 32–64 and 6–7 levels were optimal for both the segmentation and training time (~24 hours on a workstation with 98 GB RAM and AMD Ryzen Threadripper 3960X, for a total of 576 CPU-hours). The filter size dictates how large, by pixels, the individual portions of the training data filtered through the architecture are, while the level number indicates how 'deep' the architecture is and equates loosely to how complex the segmentation criteria are, with more layers used for more complex segmentation. Both can greatly affect training time, especially filter size.

Optimization of these parameters was done qualitatively by applying trained models of varying architectures to different sections of the images than those upon which they were trained, and to new images, and visually determining whether they sufficiently segmented the appropriate phases in a 'pass'/'fail' system. In models with fewer levels, the models did not capture the crystal edges, nor variation in crystal type well between images, while higher filter sizes resulted in no appreciable difference in segmentation, but much longer training timeframes, and larger model size.

The only quantifiable metric given by `Dragonfly`™ during the `Segmentation Wizard` training to determine validity of models is the Dice score (also known as an F1-score). This is calculated as:

$$\text{Dice score} = \frac{2\,TP}{2\,TP + FP + FN} \quad (1)$$

where TP is the number of true positives (properly assigned pixels), FP is the number of false positives (model segmentation pixels assigned incorrectly to a phase), and FN is the number of false negatives the model finds (model segmentation pixels assigned to a different phase than manual segmentation had identified. For more details, see Segmentation Wizard Interface*). This score acts as an indication of the precision of the model compared to validation data. While it is a useful first step metric if comparing differing model architectures (`UNET` vs. `UNET++` vs. Random Forest, etc.), it should not be the only criterion used to determine model viability, as the Dice score will inevitably increase with overfitting to the data offered. Qualitative determination by model segmentation on new frames (i.e. visual inspection of the segmentation and judgment of goodness-of-fit), as described above, is a better method for determination of model validity whilst in a training/fine-tuning phase. We acknowledge this may not be the most accurate metric by which to determine model robustness; however, it is the only rapid way to do so in this off-the-shelf manner without the need to export the data and generate confusion matrices by hand.

It can be tempting to train models until they fit the training data perfectly. However, this can result in overfitting, where the model becomes too specialized during fine-tuning (training iterations after initial creation and training of the model), reducing its generalizability, and thus performing worse on new, unseen data. Consequently, one training session and one fine-tuning session are suggested. Here, a 'training session' refers to the first time that the model is allowed to iterate for 100 epochs or until a certain 'loss criterion,' which is the difference between training and validation data (areas of the training frames set aside for comparison), is reached. Any additional 're-training' steps after the initial creation and training of the model are considered to be 'fine-tuning' sessions.

2.4 Utilization of segmentation data

Of particular interest for this study is the comparison between data derived from model and manual segmentation. Textural information of crystal size and shape is utilized in several ways, including (i) volume fraction for use in rheological calculations [e.g. Costa 2005; Cimarelli et al. 2011], (ii) microlite and bubble number density, and crystal size distribution (CSD) graphs for the purpose of crystal/bubble residence/growth time calculations [e.g. Shea et al. 2010; Cashman 2020], and (iii) stratigraphic correlations [Jutzeler et al. 2024].

To test model efficacy for such calculations, four new subsets (each 0.26 mm^2) of the larger mosaics taken from each sample are segmented both by the model, and manually through a combination of tracing crystals and thresholding. Bulk volume fraction is approximated by area fraction and calculated by combining all segmented crystal phases. When internal sieving was seen in plagioclase, we applied the `'fill holes'` function so that the entire phenocryst was treated as a single object in the final model segmentation. The area fraction of each phase is reported both before and after this function is applied (Table 2).

For more detailed crystal characteristics, it is necessary to separate all non-touching groups of pixels in each class into separate objects, turning a Region of Interest (ROI; all pixels defined as a certain class/phase) into a Multi-ROI (where all touching pixels in a particular class are grouped individually, resulting in separate crystals or groups of touching crystals). For each image segmented, the size (measured as the area (mm^2)), long axis (mm) and short axis (mm), and aspect ratio of the Multi-ROI-divided crystals for each phase were calculated in `Dragonfly`™ and output as a .csv file for analysis. We additionally utilize these data to calculate CSD of plagioclase crystals, which are well suited to growth time calculations due to well-constrained growth rates [e.g. Cashman 1988; Shea and Hammer 2013; Arzilli et al. 2015; 2019; Moshrefzadeh et al. 2023].

To calculate CSD, separate larger subsets from the mosaics of 18CW100-10 and 18CW100-16 (1.6 and 2.8 mm^2, respectively) were used for CSD calculations to ensure inclusion of the largest phases seen. As these were so large that manual segmentation would have required tens of hours to segment all phases, we only used these larger Image 1b and Image 2b frames for plagioclase CSD calculations of those samples. Image 1a and Image 2a refer to the original 0.26 mm^2 subset frames upon which the rest of the crystal data were calculated (See Supplementary Material 3). All objects of ten or fewer pixels were removed [Hammer et al. 1999], to ensure that no erroneous pixels bias the CSD. In addition, we removed any object cut off by the boundary of the image, so that no incomplete values were included.

The long and short axes data for individual plagioclase crystals were exported from `Dragonfly`™ and used to estimate the third dimension in `HabitEst3D` software [Li et al. 2022]. These dimensions were fed into the `CSDCorrections` application [Higgins 2000], which generates CSD graphs. Different size populations within the samples are delineated by changes in slope of the CSD curves. Through each of these regions a linear regression is fitted, the slopes of which can be used to calculate the crystallization/residence time (τ) of those size populations:

$$\tau = \frac{-1}{\text{slope} \times G_T} \quad (2)$$

where the slope is the slope of linear regression on the CSD graph for a particular size fraction and G_T is the growth rate ($mm\,s^{-1}$) of the phase in question (in this case, plagioclase) [Marsh 1988; Higgins 2006]. We used four different growth rates of plagioclase in these calculations as upper and lower bounds for deep-conduit growth for the larger crystals: 10^{-6} $mm\,s^{-1}$ [Shea and Hammer 2013] and 10^{-7} $mm\,s^{-1}$ [Arzilli et al. 2015], respectively; and upper and lower bounds for upper conduit/syn-eruptive growth for the smaller crystal populations: 10^{-4} $mm\,s^{-1}$ [Arzilli et al. 2019] and 2×10^{-5} $mm\,s^{-1}$ [Arzilli et al. 2015], respectively.

*https://www.theobjects.com/dragonfly/dfhelp/2024-1/Content/Artificial%20Intelligence/Segmentation%20Wizard/Segmentation%20Wizard.htm

Table 2: Comparison of model versus manual segmentation phase and total crystallinity % by area for the four 0.26 mm^2 subset images of each sample.

Sample	Phase area %							Area % normalized on a vesicle-free basis				
	Glass	Vesicles	Plagioclase	Mafic			Oxides	Total crystallinity	Plagioclase	Mafic			Oxides	Total crystallinity*
				1	2	1 + 2				1	2	1 + 2		
Image 1a - Model	22.86	47.47	20.17	8.45	0.49	8.94	0.55	29.66	38.41	16.09	0.93	17.02	1.04	56.48
Image 1a - Model†	16.97	42.41	31.15	8.44	0.49	8.93	0.55	40.62	54.08	14.66	0.84	15.5	0.95	70.54
Image 1a - Manual	14.81	40.81	32.96	9.25	1.62	10.86	0.55	44.38	55.69	15.62	2.73	18.35	0.93	74.97
Image 2a - Model	14.73	46.97	28.78	6.27	1.31	7.58	1.95	38.30	54.27	11.82	2.48	14.30	3.67	72.23
Image 2a - Model†	11.89	46.66	31.95	6.26	1.30	7.55	1.95	41.45	59.90	11.73	2.43	14.16	3.65	77.71
Image 2a - Manual	12.22	46.31	32.05	6.58	1.05	7.63	1.80	41.48	59.69	12.25	1.96	14.20	3.36	77.25
Image 3 - Model	24.77	30.75	27.59	10.75	1.66	12.41	4.48	44.48	39.84	15.52	2.40	17.92	6.48	64.23
Image 3 - Model†	23.74	30.56	28.84	10.71	1.66	12.38	4.48	45.70	41.53	15.43	2.39	17.82	6.46	65.81
Image 3 - Manual	22.71	31.32	28.99	11.12	1.52	12.64	4.34	45.97	42.21	16.19	2.21	18.41	6.32	66.93
Image 4 - Model	34.84	10.61	39.67	11.09	1.87	12.96	1.85	54.48	44.42	12.42	2.09	14.51	2.07	61.00
Image 4 - Model†	33.48	10.61	41.06	11.07	1.93	13.00	1.85	55.91	43.72	12.22	2.06	14.28	2.04	60.03
Image 4 - Manual	31.69	10.29	42.46	11.46	2.31	13.77	1.79	58.02	42.73	11.94	2.01	13.96	1.99	58.67

* On a vesicle-free basis. Vesicle-normalized values are abundances for phases when bubbles are removed, i.e. the dense rock equivalent areal % of those phases.
† Filled plagioclase: the tool `'fill holes'` was used on the plagioclase to fill the internal sieved and zoned areas which may not have been properly assigned.

We created CSD plots and used these to calculate crystallization time for both model and manual image segmentation on areas at least twice that of the largest crystals seen for each sample. While detailed petrological interpretation of the samples would require analysis of larger areas, these images are large enough to gain information on crystal shape and size differences between manual and model segmentation, especially when carried out across four separate samples, which all contain different textural characteristics (Figure 3).

3 Results

A `UNET` model with 64 filter size and 7 layers was the best fitting for the training and validation data, with a Dice score of ~0.95 out of 1 (Equation 1). Training was halted after one fine-tuning round, with the total time for training and fine-tuning lasting an average of 12 hours each on the same workstation described previously.

Two model records are given in Table 2, one directly from the `Dragonfly`™ output, and one labeled 'filled plagioclase' where the `'fill holes'` tool was applied. This simple post-model step resulted in an increased similarity in total crystallinity with the manual segmentation, most strongly seen in Image 1a. Comparison of the 'filled plagioclase' model to manual segmentation shows a very similar crystal area fraction within 0.5–4 % (Table 2) for vesicle normalized calculations. Additionally, the difference in modal area calculated for each phase was <5 % for all phases except Mafic 2, which has an average difference of ~25 % (Table 2). If Mafic 1 and Mafic 2 are combined into one phase, their average difference in modal area becomes ~6 %.

The ability of the model to provide reasonable shape and size characteristics for individual crystals in each phase is equally important. For each phase in each image, the 2D aspect ratio, minimum and maximum axis length, and total 2D area was recorded, and subsequently compared with the same values from manual segmentation. Figure 3 showcases frequency distributions for Mafic 1 crystals in all four images by their aspect ratio (Figure 3A) and 2D area (Figure 3B). All the images have similarly increased frequency of low-aspect ratio (more equant) Mafic 1 crystals segmented by the model than by manual segmentation, on average ~3 times more by model segmentation. However, the decay in detections with increasing aspect ratio appears comparable between manual and model segmentations per image. This results in very similar cumulative aspect ratio distributions, with a strong increase at low aspect ratios and tapering off at high aspect ratios. Images 2 and 3 have slight deviations, due to gaps in aspect ratio distributions which differ between model and manual segmentation. All other phases show similar patterns to Image 1a and Image 4 Mafic 1 aspect ratios, with very similar cumulative frequency curves across all images (Supplementary Material 6).

Crystal area for Mafic 1 shows similar patterns across all four images, with the model reporting larger amounts of small crystals (Figure 3; Supplemental Material 6). On average, the model finds ~5 times more crystals of sizes 1–50 μm^2 than the manual segmentation, resulting in an average of 1.8 times increase in cumulative area over those sizes. This results in

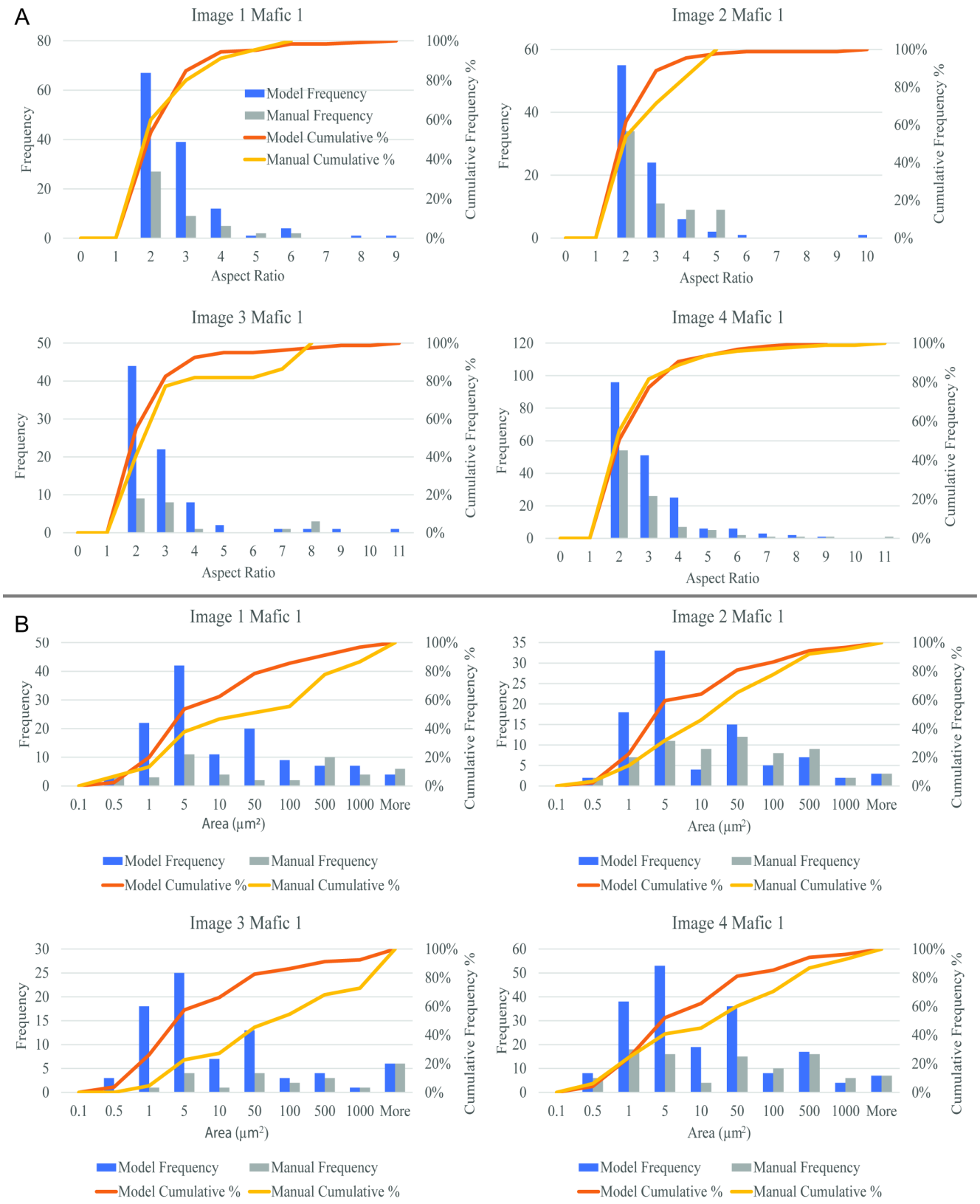
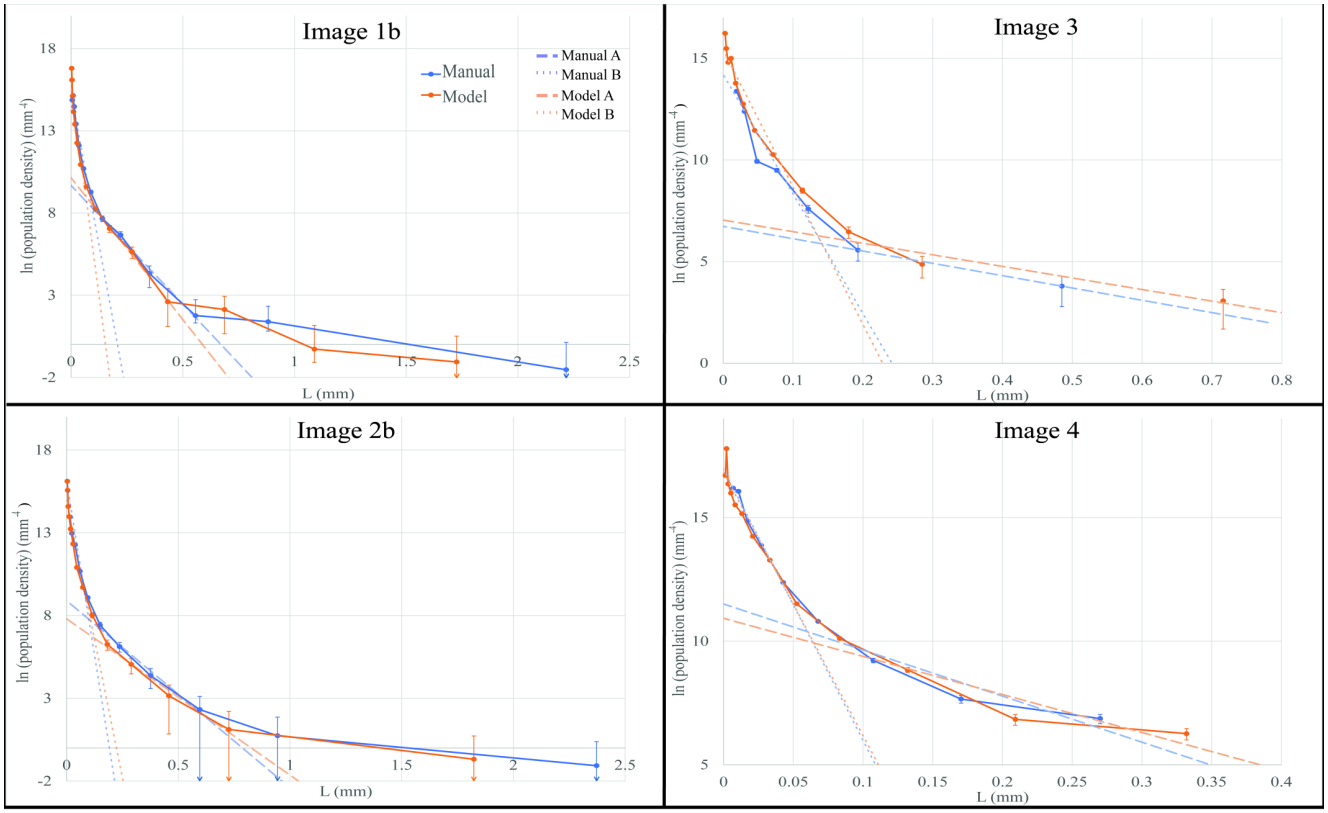


Figure 3: Aspect ratio [A] and 2D area [B] histograms of Mafic 1 for Images 1–4. Comparison of model vs. manual segmentation of this phase indicate that aspect ratios are similar, while many more objects of small areas are segmented by the model than by manual segmentation. The same graphs for all other crystal phases are available in Supplementary Material 6.



Linear Regression Slopes

	Image 1b	Image 2b	Image 3	Image 4
Manual A	-14.41	-22.29	-6.06	-18.63
Model A	-17.24	-19.26	-5.69	-15.40
Manual B	-73.42	-109.56	-58.77	-109.72
Model B	-103.20	-135.76	-67.44	-106.57

Figure 4: Plagioclase CSD graphs made from application of the model to new sections of the samples used in model training. *L* is the stereologically determined longest dimension of the crystals, the population density is the number of crystals of that size per volume [Higgins 2000]. Manual segmentation values are in blue, while those from the model are in orange. The dashed and dotted lines indicate the linear regressions for the larger crystal size populations and smaller crystal size populations, respectively. The slope of each linear regression is used to calculate growth times for each population and shown in the table at the base of the figure.

Table 3: Comparison of model versus manual segmentation plagioclase CSD calculated growth times for the bimodal populations of crystals seen in each image.

		Crystallization/residence time			
		Days*	Days**	Minutes†	Minutes‡
Image 1	A - Deeper conduit (Model, Manual)	0.7, 0.8	6.7, 8.0		
	B - Shallow conduit/Syn-eruptive (Model, Manual)			1.6, 2.3	8.1, 11.4
Image 2	A - Deeper conduit (Model, Manual)	1.2, 1.0	12.3, 10.2		
	B - Shallow conduit/Syn-eruptive (Model, Manual)			2.5, 2.0	12.5, 9.9
Image 3	A - Deeper conduit (Model, Manual)	2.0, 1.9	20.3, 19.1		
	B - Shallow conduit/Syn-eruptive (Model, Manual)			2.5, 2.8	12.4, 14.2
Image 4	A - Deeper conduit (Model, Manual)	0.8, 0.6	7.5, 6.2		
	B - Shallow conduit/Syn-eruptive (Model, Manual)			1.6, 1.5	7.8, 7.6

* Time calculated using growth rate of 10^{-6} mm s^{-1};
** Time calculated using growth rate of 10^{-7} mm s^{-1};
† Time calculated using growth rate of 10^{-4} mm s^{-1};
‡ Time calculated using growth rate of 2×10^{-5} mm s^{-1}.

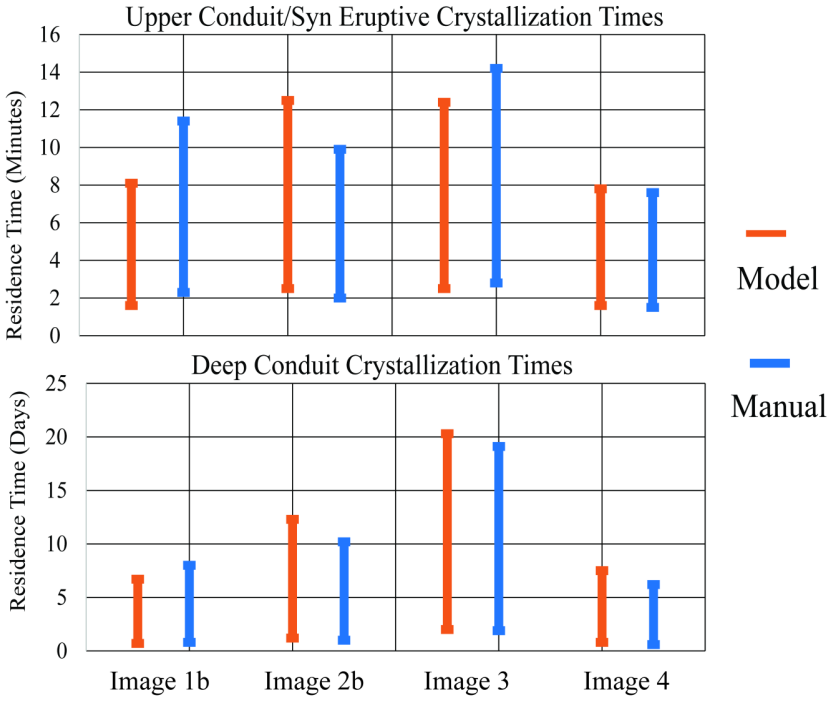


Figure 5: Graphical illustration of model and manual segmentation calculated crystallization times. Upper and lower bounds on the bars indicate the upper and lower growth rates for each proposed crystal growth depth.

a slightly steeper cumulative distribution of crystals by size, due to the larger uptick in small crystals found by the model. These values are strongly weighted in Image 3, where there is a much larger uptick in detections in these size brackets by the model over the manual segmentation than the other sample images.

In the larger crystal sizes, 100 to 1000 μm^2, the increase in detection frequency by the model over the manual segmentation is just ~1.2 times. This results in a decrease in slope for model cumulative distribution and a slight steepening in the cumulative curve for manual segmentation. The change in slope results in overall similarly sloped cumulative size curves, though less equivalent cumulative aspect ratio distributions. This pattern of a higher number of small crystals detected by the model over the manual segmentation is also seen in Mafic 2 and plagioclase (at different rates of increase), though the opposite is true for oxides, where the manual segmentation reports higher numbers of smaller size populations (see Supplemental Material 6 for graphs for all phases).

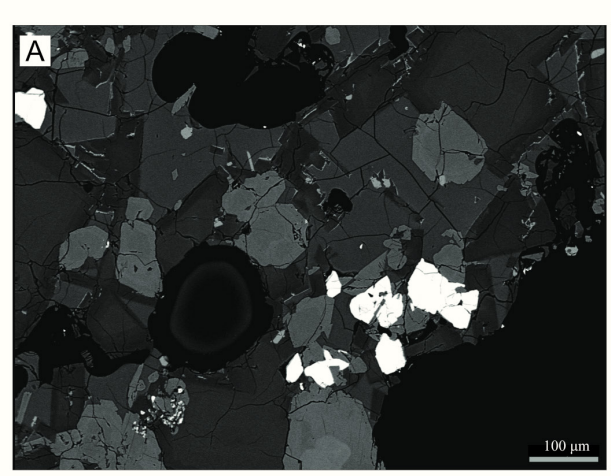
Using Images 1b, 2b, 3, and 4, calculations of the CSD comparisons between manual and model segmented plagioclase are shown in Figure 4. Model values are calculated directly from the model ROIs, without any post-segmentation manual modification such as filling of plagioclase, or manual segmentation of missed or overestimated crystals. The variance for each size interval was calculated using the approximation $2\sigma = 2\sqrt{N}$ [Higgins 2000; Gualda 2006], where N is the crystal number reported for each size interval from `CSDCorrections`. All calculated values for the model and manual segmentations agree within these uncertainties for Images 1b, 2b, and 4. Image 3 shows a slightly increased population density in the smaller phases obtained by the model relative to manual segmentation. The crystallization times calculated from the slopes of regression lines drawn through the CSD plots are seen in Table 2. The choice of which points to include in the linear fits, from which CSDs are calculated, introduces greater uncertainty into the timescale calculation than any difference between model and manual segmentation, which is always less than 20 % relative. The largest uncertainty, of about one order of magnitude, comes from the choice of growth rate utilized in the calculations (Table 3, Figure 5).

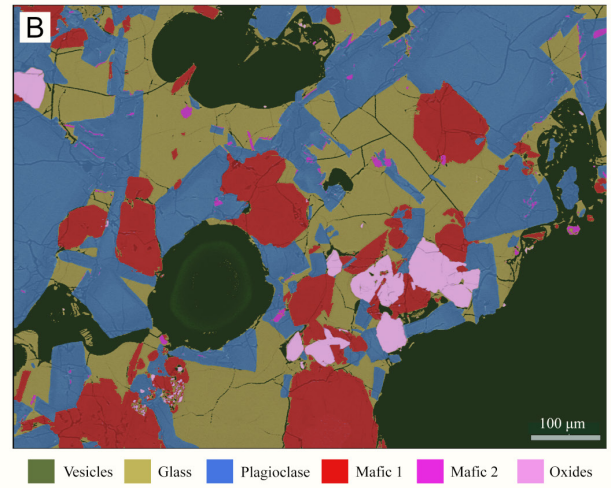
Finally, application of the model to samples not used in training provided reasonable results (Figure 5 and Figure 6). Confusion matrices are calculated unidirectionally by:

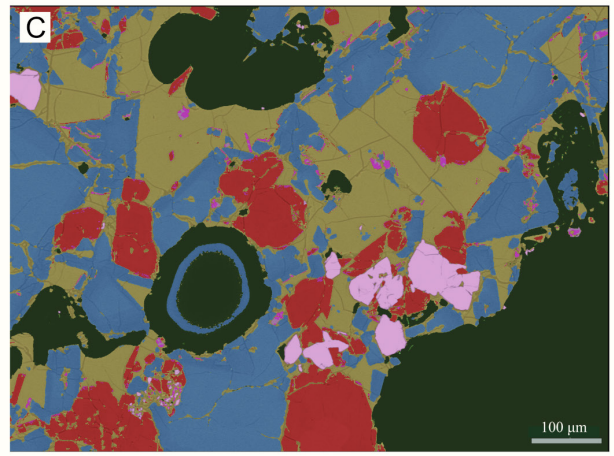
$$\text{overlap of phase}_x = \frac{\left(\begin{array}{l}\text{\# of pixels of model phase}_x \\ \quad \text{intersection with manual phase}_{x,y,z\ldots}\end{array}\right)}{\text{\# of pixels in model phase}_x} \tag{3}$$

where the $\text{phase}_{x,y,z,\ldots}$ in this case refers to the number of pixels the model or manual segmentation assigned to plagioclase, oxides, Mafic 1, Mafic 2, glass, or vesicles. These matrices provide quantification of the fidelity between model and manual segmentation in samples upon which the model was not trained.

In sample 18CW100-13, oxide and Mafic 1 phases have a true-positive rate of >91 %, while that of plagioclase is 85 % with the zoned and sieved centers assigned to glass. Mafic 2 is much more unreliable, with a true-positive rate of 43 %, with 41 % assigned as false positives to Mafic 1 (Figure 6).







D

Model Segmented Phase	Manually Segmented Phases: Plagioclase	Mafic 1	Mafic 2	Oxides	Glass	Vesicles
Plagioclase	84.8%	0.0%	0.0%	0.0%	13.0%	0.4%
Mafic 1	0.3%	98.5%	0.0%	0.3%	0.9%	0.0%
Mafic 2	7.0%	42.1%	40.7%	4.7%	5.8%	0.0%
Oxides	2.2%	1.5%	3.4%	91.6%	2.5%	0.1%
Glass	18.9%	5.3%	0.0%	0.0%	72.2%	3.5%
Vesicles	1.1%	0.1%	0.0%	0.0%	1.1%	96.8%

Figure 6: Sample 18CW100-13, a basaltic scoria sample which was not used in training. [A] is the original BSE image, [B] is the manual segmentation of the image, and [C] is the model segmentation. [D] A confusion matrix for model and manual segmentation, where % is the % overlap (Equation 3). For example, 92 % of the pixels the model designated as oxides are true positive, while 3 % of the pixels designated as oxides are actually Mafic 1, 1 % are Mafic 2, and 2 % are plagioclase.)

Sample 18CW100-11 has a similar pattern in the confusion matrix (Figure 7). Plagioclase has a true positive value of 96.3 %, Mafic 1 of 97 %, and Oxides of 72.6 %. The high precision of the plagioclase, but the obvious omission of the interiors of the large plagioclase crystals is due to 96.3 % of model assigned plagioclase overlapping with the manual segmentation plagioclase, but as none of those pixels assigned to plagioclase by the model overlap with the manually segmented glass, the confusion matrix indicates high fidelity, even with the large omissions. Instead, the confusion matrix reports this mismatch in the glass segmentation, where approximately 55 % of the glass assigned by the model overlaps with the plagioclase values from manual segmentation.

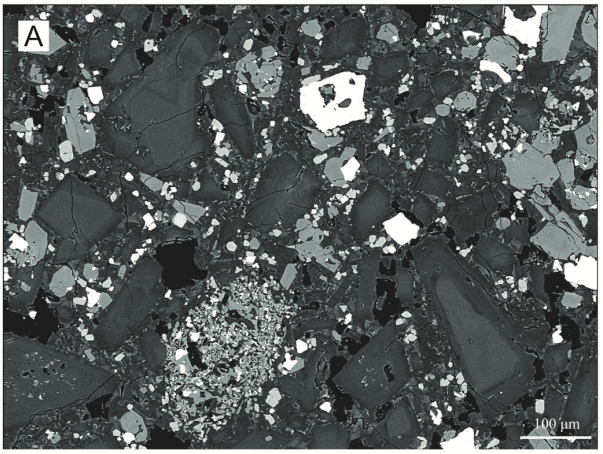
Mafic 2 has a similar true to false positive rate as 18CW100-11, where 50.2 % are properly assigned, and 44.6 % are assigned to Mafic 1. This pattern echoes that from the training images, where oxide and Mafic 1 phases are accurately assigned, plagioclase is well segmented around the edges but may require some small amount of post-processing to fill holes, and Mafic 2 is the most falsely assigned.

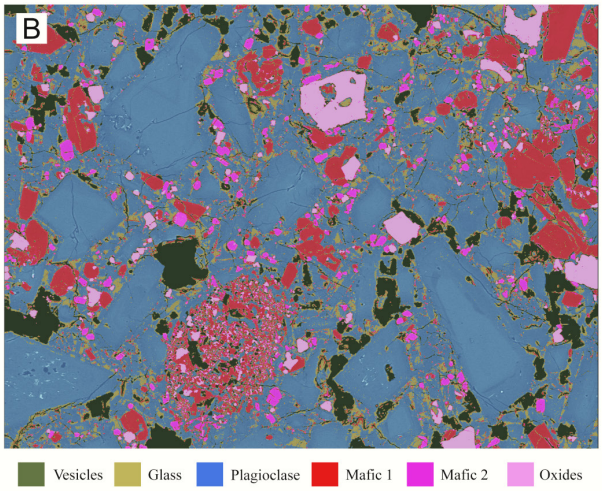
No post-model-application manual segmentation was applied to these images, to ensure that the resultant values showcase model efficacy only.

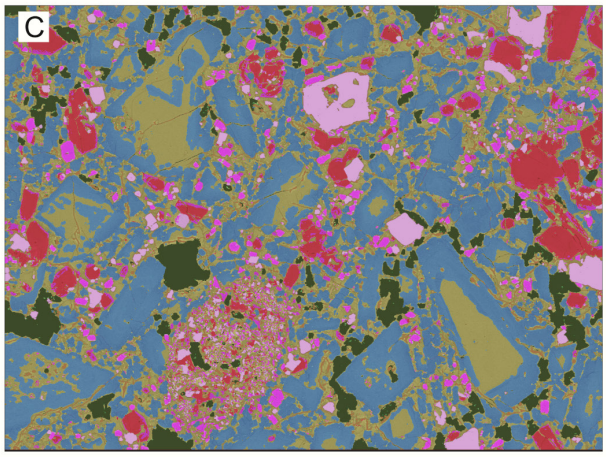
4 Discussion

The automated segmentation model trained on these samples shows good agreement with manual segmentation when applied to different areas of the same samples on which it was trained (Table 1 and 2; Figure 3 and 4). Total crystallinity shows an average difference of ~3 % from manual segmentation, while individual phases are within <6 % of manual segmentation. Shape comparison of individual crystals of each phase indicates similar distributions of aspect ratios of model and manual segmentation (Figure 2, Supplementary Material 6), although the model reports higher detections for each aspect ratio value than manual segmentation, especially in the lower aspect ratios. This may be due to anomalous pixel detections, which predominantly add to the small, equant crystal populations; or loss of edges of crystals by the model in samples, shortening elongate crystals and resulting in more equant geometries reported.

A comparison of the 2D area of individual crystals results in a general trend of a higher frequency of small crystal detections by the model for all phases, except for oxides, where this is reversed. The former trend is likely due to some amount of anomalous pixel detections, although some objects may be truly segmented due to a heightened sensitivity of the model to







D

Model Segmented Phase	Manually Segmented Phases: Plagioclase	Mafic 1	Mafic 2	Oxides	Glass	Vesicles
Plagioclase	96.3%	0.1%	0.0%	0.0%	3.2%	0.4%
Mafic 1	0.2%	97.0%	0.9%	0.0%	1.9%	0.1%
Mafic 2	0.9%	44.6%	50.2%	3.1%	2.0%	0.0%
Oxides	1.0%	8.1%	14.1%	72.6%	4.3%	0.4%
Glass	55.6%	7.5%	0.1%	0.0%	29.2%	8.3%
Vesicles	2.1%	0.3%	0.2%	0.3%	4.5%	90.7%

Figure 7: Sample 18CW100-11 a basaltic scoria sample which was not used in training. [A] is the original BSE image, [B] is the manual segmentation of the image, and [C] is the model segmentation. [D] A confusion matrix for model and manual segmentation, where % is the % overlap (Equation 3).

pixel-level changes when compared to manual segmentation. Conversely, the reversal of the trend for oxides may be due to the thresholding used in manual segmentation, which likely resulted in the inclusion of non-associated pixels as oxides by brightness value alone.

Plagioclase CSD calculations using values from the model result in very clear similarities with the manual segmentation values, even without post-processing to fill sieved crystals. For Images 1b, 2b, and 4, values from each size range fall within error of one another, rendering the graphs statistically identical. Image 3 shows some separation between the manual and model segmentation CSD values, predominantly in crystals <0.1 mm. As with the crystal area, this discrepancy may be due to the two factors of a) anomalous pixel detection by the model, or b) the sensitivity of the model to pixel-scale objects. As this sample shows very diffuse edges of the plagioclase crystals, the latter would result in the model detecting smaller objects (e.g. microlites) which could be missed entirely in manual segmentation. When the CSD is propagated to calculate residence time, there is less than 20 % difference in results from crystal segmentation between model and manual segmentation in this specific case (Table 3), which is less than uncertainties from choices in data selected for line fitting and the assumed crystal growth rate (Figure 5).

Sample 18CW100-13, on which the model was not trained, shows distinct segmentation of oxides, Mafic 1, and plagioclase (Figure 6). However, Mafic 2 is equally likely to be assigned as Mafic 1 as it is a true positive. This is echoed in sample 18CW100-11, where Mafic 2 has a true-positive rate of 50.2 % and a false positive rate with Mafic 1 of 44.6 %. The three other crystal phases show similar true positive rates to those in sample 18CW10-11, with the added complexity of the higher amount of intergrowth between plagioclase. In this case, while much of the manually segmented plagioclase is designated as glass by the automated model, inspection of the images indicates that the model may be stronger at delineating the edges of the intergrown plagioclase crystals around small glass pockets than the manual thresholding segmentation. As such, the model trained on these samples may in fact be able to detect crystal edges in samples with diffuse crystal boundaries and/or intergrown textures better than manual segmentation can.

We find that models trained on only a few images, in the bimodal image stack structure, produce results in close agreement with manual segmentation for total crystallinity, and raw size and shape statistics. These would also propagate into analyses using this data, such as calculating the impacts of crystal textures on viscosity [e.g. Costa 2005].

4.1 Limitations and considerations

There are a few main limitations and considerations for using this automated approach (for a more detailed list, see Supplementary Material 1a,b – SOP).

1. Samples with different crystal assemblages must have their own individual model for detailed image segmentation. If the model has not been trained on a certain phase, it will assign pixels based on phases it has been trained on, resulting in a mismatch for that phase. We recommend that a maximum of one training session and one to two fine-tuning sessions be carried out to avoid overtraining the models (Supplementary Material 1a,b – SOP, 2).

2. CNN models are usually trained with significantly more input data (on the order of 100s to 1000s of images [Zhou et al. 2018]) than we have provided our models. The more the input data showcases different views of the same phases (shapes, grayscale values, sizes, context, etc.), the more robust the resultant models will be when applied to new images.

3. Anomalous pixel detections by the model often result in an increased population of small, equant crystal detections. These are often found as slight variations in greyscale of interstitial glass or interiors of complexly zoned crystals. However, in some samples, the model may be detecting small, crystals that were missed during manual segmentation. Both possibilities were seen in the samples used here. This issue may be alleviated by an increase in image resolution or training data, or the removal of crystals less than a certain areal threshold (often <10 pixels—for further explanation, see SOP).

4. The models often failed to properly allocate sieved or zoned interiors of crystals to their appropriate phase. While this may prove positive for those users who wish to delineate zonation patterns or melt inclusions, it is non-ideal for total-crystal-area calculations. This can be relatively easily fixed using the `'fill holes'` algorithm included in `Dragonfly` segmentation tools. Some user supervision may be necessary, however, to ensure that (i) no areas bounded by crystals becomes filled in and allocated to the phase of the bounding crystals, or that (ii) no zoned/sieved crystals are not properly filled, due to truncation of external zonation in 2D from crystal breakage or angling into/out of the image plane.

5. As noted in the Methodology section, the automated segmentation model developed herein using SEM BSE imagery resulted in segmentation which was unable to differentiate between amphibole and clinopyroxene, resulting in the binning of those phases as Mafic 1. Likewise, the orthopyroxene and possible olivine microlites are binned under Mafic 2. This is the result of the training data used, as those phases were unable to be differentiated using the BSE images when creating training data using manual segmentation. For the purposes outlined for this particular model, this was an acceptable and useful outcome. However, this may not be the case for all users, and so more training data or the inclusion of an EDS Mode may be useful in future use of this model creation methodology.

While we have identified a reasonable approach for low-time, low-input data semantic segmentation, the results are imperfect, and post-modeling manual segmentation may be necessary to achieve the desired results, such as correcting for sieved or zoned plagioclase cores (Figure 4).

4.2 Future work

The crystal-rich basaltic samples from Bogoslof volcano were used to test the ability of the bimodal image stack methodology to aid in low-input data and rapid-return sample segmentation of complex samples. Future work on this methodology should include testing its utility when applied to more silicic systems. In these systems, differentiation between crystal phases and glass may be more difficult, due to similar mean atomic numbers.

The ability to utilize images from reflected and transmitted light microscopy would also prove useful, as these techniques require less expensive equipment than electron microscopy. Some preliminary work using basaltic samples from other localities indicates that the proposed methodology combined with the `UNET` or `UNET++` [Ronneberger et al. 2015; Zhou et al. 2018] works equally well in segmentation of reflected light imaging as the BSE images used here [Halverson and Whittington 2025].

These expansions in the utility of the methodology would have wider impacts for both academic and governmental scientific institutions, including volcano observatories. One application of this method is rapid determination of crystal size and shape distributions from volcanic rocks in much less user-required time (<24 hours for model preparation, which can be done in advance; then 2–20 minutes to apply the model, compared to 2–12 or more hours of manual segmentation, dependent on image size). This approach can facilitate rapid petrographic monitoring of volcanic materials in eruption scenarios at volcano observatories. To date, rapid petrological monitoring has focused on crystallinity and chemical composition, for example at Kīlauea, Hawai'i, in 2018 [Gansecki et al. 2019], at Cumbre Vieja, Canary Islands, in 2021 [e.g. Pankhurst et al. 2022], and on the Reykjanes peninsula, Iceland, from 2021–2024 [e.g. Halldórsson et al. 2022]. Eruptive style of lava can be more sensitive to changes in texture than to changes in chemical composition [e.g. Loewen et al. 2021], so the technique presented here represents a substantial improvement in real-time petrological monitoring capability for volcanic eruption response. It can also save user time in less time-sensitive situations. In many cases, volcanoes erupt material with similar phase assemblages to prior eruptions, and models can be trained during quiescent periods, in anticipation of future activity.

5 Conclusion

We have tested a new pre- and post-processing methodology for the use of `UNET` and `UNET++` models to segment basaltic volcanic samples using the free `Dragonfly`™ software. This software and application allow for the easy utilization of deep learning techniques in image segmentation for users who have no prior knowledge of deep learning or coding techniques. Our methodology works within this framework to allow for

the segmentation of a variety of samples with an average modal abundance agreement of 92 % per phase, with one pre-trained model.

We have shown that this methodology, combined with UNET and UNET++ architectures, allows for the creation of models which can segment multiple mineral phases and generate total crystallinity values along with phase-specific size and shape data that are broadly applicable to studies of magmatic and volcanic processes and rheology. Furthermore, utilizing CSD calculations, we have shown that the segmentation done by the model is able to match calculated values from manual segmentation within uncertainty. It could be used to create a library of deep learning models trained for different magma compositions and textures in anticipation of rapidly analyzing a variety of sample types for time-sensitive applications such as volcanic crises.

Author Contributions

Brenna A Halverson: Primary researcher and writer. Matthew Loewen: Internship advisor, provided technical help with imaging and analysis, as well as assistance with writing. Hannah Dietterich: Internship advisor provided technical help with imaging and analysis, as well as assistance with writing. Alan Whittington: Doctoral advisor provided technical and writing assistance.

Acknowledgements

The authors would like to acknowledge the NSF INTERN program, award EAR 2309154, which allowed for B.A. Halverson to complete this research at the U.S. Geological Survey Alaska Volcano Observatory. This work was also supported by NSF EAR 1928923 to AW. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Data Availability

Supplementary Information:

- Contact: halversonbrenna@gmail.com.

- Please see the Supplementary Material at DOI: 10.5281/zenodo.15574184 for the SOP and other supplemental images and data.

- Below are the requirements for Dragonfly™ download and usage of Deep Learning software within it:

– General Dragonfly™ info:

To download Dragonfly™ for non-Commercial Use: https://dragonfly.comet.tech/en/non-commercial-licensing. This will open and download a 30-day trial, after the end of 30 days, apply for a years' license and the codes should be sent to the email within 30 hours (usually faster).

– Hardware requirements:

From: Getting Started - Comet Dragonfly

* AVX compliant CPU
* GPU with 1+ GB of VRAM (NVIDIA or AMD) or Intel integrated graphics processor (IGP) with UHD Graphics[†]
* Support for Open GL 4.3 or higher
* 8 GB of RAM
* 10+ GB of local hard disk space for download and installation

[†]Performance will likely not be comparable to running Dragonfly 3D World on a system with a dedicated GPU.

– Program language:

* N/A – has python source code that can be modified but is not necessary for any part of the above process.

– Software required:

* Windows 7/8/8.1/10/11 (64-bit)[‡] or Linux (64-bit), starting from Ubuntu 18.04

[‡]Windows 10/11 is required for deep learning.

Copyright notice

© The Author(s) 2025. 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

Arzilli, F., C. Agostini, P. Landi, A. Fortunati, L. Mancini, and M. R. Carroll (2015). "Plagioclase nucleation and growth kinetics in a hydrous basaltic melt by decompression experiments". *Contributions to Mineralogy and Petrology* 170(5–6). DOI: 10.1007/s00410-015-1205-9.

Arzilli, F., G. La Spina, M. R. Burton, M. Polacci, N. Le Gall, M. E. Hartley, D. Di Genova, B. Cai, N. T. Vo, E. C. Bamber, S. Nonni, R. Atwood, E. W. Llewellin, R. A. Brooker, H. M. Mader, and P. D. Lee (2019). "Magma fragmentation in highly explosive basaltic eruptions induced by rapid crystallization". *Nature Geoscience* 12(12), pages 1023–1028. DOI: 10.1038/s41561-019-0468-6.

Benet, D., F. Costa, C. Widiwijayanti, J. Pallister, G. Pedreros, P. Allard, H. Humaida, Y. Aoki, and F. Maeno (2024). "VolcAshDB: a Volcanic Ash DataBase of classified particle images and features". *Bulletin of Volcanology* 86(1). DOI: 10.1007/s00445-023-01695-4.

Blundy, J. and K. V. Cashman (2008). "Petrologic Reconstruction of Magmatic System Variables and Processes". *Reviews in Mineralogy and Geochemistry* 69(1), pages 179–239. DOI: 10.2138/rmg.2008.69.6.

Cashman, K. V. (1988). "Crystallization of Mount St. Helens 1980–1986 dacite: A quantitative textural approach". *Bulletin of Volcanology* 50(3), pages 194–209. DOI: 10.1007/bf01079682.

– (2020). "Crystal Size Distribution (CSD) Analysis of Volcanic Samples: Advances and Challenges". *Frontiers in Earth Science* 8. DOI: 10.3389/feart.2020.00291.

Chen, Z., X. Liu, J. Yang, E. Little, and Y. Zhou (2020). "Deep learning-based method for SEM image segmentation in mineral characterization, an example from Duvernay Shale samples in Western Canada Sedimentary Basin". *Computers & Geosciences* 138, page 104450. DOI: 10.1016/j.cageo.2020.104450.

Cimarelli, C., A. Costa, S. Mueller, and H. M. Mader (2011). "Rheology of magmas with bimodal crystal size and shape distributions: Insights from analog experiments: Rheology of porphyritic magmas". *Geochemistry, Geophysics, Geosystems* 12(7). DOI: 10.1029/2011gc003606.

Coombs, M., K. Wallace, C. Cameron, J. Lyons, A. Wech, K. Angeli, and P. Cervelli (2019). "Overview, chronology, and impacts of the 2016–2017 eruption of Bogoslof volcano, Alaska". *Bulletin of Volcanology* 81(11). DOI: 10.1007/s00445-019-1322-9.

Costa, A. (2005). "Viscosity of high crystal content melts: Dependence on solid fraction". *Geophysical Research Letters* 32(22). DOI: 10.1029/2005gl024303.

Di Martino, A., G. Carlini, G. Castellani, D. Remondini, and A. Amorosi (2023). "Sediment core analysis using artificial intelligence". *Scientific Reports* 13(1). DOI: 10.1038/s41598-023-47546-2.

Gansecki, C., R. L. Lee, T. Shea, S. P. Lundblad, K. Hon, and C. Parcheta (2019). "The tangled tale of Kīlauea's 2018 eruption as told by geochemical monitoring". *Science* 366(6470). DOI: 10.1126/science.aaz0147.

Gualda, G. A. R. (2006). "Crystal Size Distributions Derived from 3D Datasets: Sample Size Versus Uncertainties". *Journal of Petrology* 47(6), pages 1245–1254. DOI: 10.1093/petrology/egl010.

Halldórsson, S. A., E. W. Marshall, A. Caracciolo, S. Matthews, E. Bali, M. B. Rasmussen, E. Ranta, J. G. Robin, G. H. Guðfinnsson, O. Sigmarsson, J. Maclennan, M. G. Jackson, M. J. Whitehouse, H. Jeon, Q. H. A. van der Meer, G. K. Mibei, M. H. Kalliokoski, M. M. Repczynska, R. H. Rúnarsdóttir, G. Sigurðsson, M. A. Pfeffer, S. W. Scott, R. Kjartansdóttir, B. I. Kleine, C. Oppenheimer, A. Aiuppa, E. Ilyinskaya, M. Bitetto, G. Giudice, and A. Stefánsson (2022). "Rapid shifting of a deep magmatic source at Fagradalsfjall volcano, Iceland". *Nature* 609(7927), pages 529–534. DOI: 10.1038/s41586-022-04981-x.

Halverson, B. A., A. Emerson, J. Hammer, J. Lira, and A. Whittington (2024). "Estimates of Crystallinity Utilizing Differential Scanning Calorimetry: Application to the Kīlauea 2018 Lower East Rift Zone Eruption". *Journal of Petrology* 65(5). DOI: 10.1093/petrology/egae010.

Halverson, B. A. and A. Whittington (2025). "From flow to furnace: Low viscosity of three-phase lavas measured at Kīlauea 2018 eruption conditions". *Geology* 53(2), pages 135–139. DOI: 10.1130/g52679.1.

Hammer, J. E., K. V. Cashman, R. P. Hoblitt, and S. Newman (1999). "Degassing and microlite crystallization during pre-climactic events of the 1991 eruption of Mt. Pinatubo, Philippines". *Bulletin of Volcanology* 60(5), pages 355–380. DOI: 10.1007/s004450050238.

Higgins, M. D. (2000). "Measurement of crystal size distributions". *American Mineralogist* 85(9), pages 1105–1116. DOI: 10.2138/am-2000-8-901.

– (2006). "Grain and crystal sizes". *Quantitative textural measurements in igneous and metamorphic petrology*. Cambridge: Cambridge University Press. Chapter 3, pages 39–134. ISBN: 9780511535574. DOI: 10.1017/CBO9780511535574.003.

Higgins, O., T. Sheldrake, and L. Caricchi (2021). "Quantitative chemical mapping of plagioclase as a tool for the interpretation of volcanic stratigraphy: an example from Saint Kitts, Lesser Antilles". *Bulletin of Volcanology* 83(8). DOI: 10.1007/s00445-021-01476-x.

Iezzi, G., S. Mollo, G. Torresi, G. Ventura, A. Cavallo, and P. Scarlato (2011). "Experimental solidification of an andesitic melt by cooling". *Chemical Geology* 283(3–4), pages 261–273. DOI: 10.1016/j.chemgeo.2011.01.024.

Iglesias, J. C. Á., R. B. M. Santos, and S. Paciornik (2019). "Deep learning discrimination of quartz and resin in optical microscopy images of minerals". *Minerals Engineering* 138, pages 79–85. DOI: 10.1016/j.mineng.2019.04.032.

Innocenti, S., S. Andreastuti, T. Furman, M.-A. del Marmol, and B. Voight (2013). "The pre-eruption conditions for explosive eruptions at Merapi volcano as revealed by crystal texture and mineralogy". *Journal of Volcanology and Geothermal Research* 261, pages 69–86. DOI: 10.1016/j.jvolgeores.2012.12.028.

Jutzeler, M., R. J. Carey, Y. Dagasan, A. McNeill, and R. A. F. Cas (2024). "Machine-learning crystal size distribution for volcanic stratigraphy correlation". *Scientific Reports* 14(1). DOI: 10.1038/s41598-024-82847-0.

Koh, E. J., E. Amini, G. J. McLachlan, and N. Beaton (2021). "Utilising convolutional neural networks to perform fast automated modal mineralogy analysis for thin-section optical microscopy". *Minerals Engineering* 173, page 107230. DOI: 10.1016/j.mineng.2021.107230.

Lanzafame, G., S. Mollo, G. Iezzi, C. Ferlito, and G. Ventura (2013). "Unraveling the solidification path of a pahoehoe "cicirara" lava from Mount Etna volcano". *Bulletin of Volcanology* 75(4). DOI: 10.1007/s00445-013-0703-8.

Leichter, A., R. R. Almeev, D. Wittich, P. Beckmann, F. Rottensteiner, F. Holtz, and M. Sester (2022). "Automated Segmentation of Olivine Phenocrysts in a Volcanic Rock Thin Section Using a Fully Convolutional Neural Network". *Frontiers in Earth Science* 10. DOI: 10.3389/feart.2022.740638.

Li, J., Z.-F. Yang, and Y. Wang (2022). "HabitEst3D: A User-Friendly Software for Estimating Mixed Crystal Habits from Two-Dimensional Sections in Igneous Rocks". *Minerals* 12(8), page 1001. DOI: 10.3390/min12081001.

Liu, Y., Z. Zhang, X. Liu, L. Wang, and X. Xia (2021). "Efficient image segmentation based on deep learning for mineral image classification". *Advanced Powder Technology* 32(10), pages 3885–3903. DOI: 10.1016/j.apt.2021.08.038.

Loewen, M. W., H. R. Dietterich, N. Graham, and P. Izbekov (2021). "Evolution in eruptive style of the 2018 eruption of Veniaminof volcano, Alaska, reflected in groundmass tex-

tures and remote sensing". *Bulletin of Volcanology* 83(11). DOI: 10.1007/s00445-021-01489-6.

Loewen, M. W., P. Izbekov, J. Moshrefzadeh, M. Coombs, J. Larsen, N. Graham, M. Harbin, C. Waythomas, and K. Wallace (2019). "Petrology of the 2016–2017 eruption of Bogoslof Island, Alaska". *Bulletin of Volcanology* 81(12). DOI: 10.1007/s00445-019-1333-6.

Lormand, C., G. F. Zellmer, K. Németh, G. Kilgour, S. Mead, A. S. Palmer, N. Sakamoto, H. Yurimoto, and A. Moebis (2018). "Weka Trainable Segmentation Plugin in ImageJ: A Semi-Automatic Tool Applied to Crystal Size Distributions of Microlites in Volcanic Rocks". *Microscopy and Microanalysis* 24(6), pages 667–675. DOI: 10.1017/s1431927618015428.

Makovetsky, R., N. Piche, and M. Marsh (2018). "Dragonfly as a Platform for Easy Image-based Deep Learning Applications". *Microscopy and Microanalysis* 24(S1), pages 532–533. DOI: 10.1017/s143192761800315x.

Malik, O. A., I. Puasa, and D. T. C. Lai (2022). "Segmentation for Multi-Rock Types on Digital Outcrop Photographs Using Deep Learning Techniques". *Sensors* 22(21), page 8086. DOI: 10.3390/s22218086.

Marsh, B. D. (1988). "Crystal size distribution (CSD) in rocks and the kinetics and dynamics of crystallization: I. Theory". *Contributions to Mineralogy and Petrology* 99(3), pages 277–291. DOI: 10.1007/bf00375362.

McClinton, J. T., S. M. White, A. Colman, K. H. Rubin, and J. M. Sinton (2014). "The role of crystallinity and viscosity in the formation of submarine lava flow morphology". *Bulletin of Volcanology* 76(9). DOI: 10.1007/s00445-014-0854-2.

Moshrefzadeh, J., P. Izbekov, M. Loewen, J. Larsen, and S. Regan (2023). "Dating individual zones in phenocrysts from the 2016–2017 eruption of Bogoslof volcano provides constraints on timescales of magmatic processes". *Journal of Volcanology and Geothermal Research* 435, page 107741. DOI: 10.1016/j.jvolgeores.2022.107741.

Mujin, M. and M. Nakamura (2014). "A nanolite record of eruption style transition". *Geology* 42(7), pages 611–614. DOI: 10.1130/g35553.1.

Pankhurst, M. J., J. H. Scarrow, O. A. Barbee, J. Hickey, B. C. Coldwell, G. K. Rollinson, J. A. Rodríguez-Losada, A. Martín Lorenzo, F. Rodríguez, W. Hernández, D. Calvo Fernández, P. A. Hernández, and N. M. Pérez (2022). "Rapid response petrology for the opening eruptive phase of the 2021 Cumbre Vieja eruption, La Palma, Canary Islands". *Volcanica* 5(1), pages 1–10. DOI: 10.30909/vol.05.01.0110.

Preece, K., F. van der Zwan, J. Hammer, and R. Gertisser (2023). "A Textural Perspective on the Magmatic System and Eruptive Behaviour of Merapi Volcano". *Merapi Volcano*. Springer International Publishing, pages 265–289. ISBN: 9783031150401. DOI: 10.1007/978-3-031-15040-1_9.

Ronneberger, O., P. Fischer, and T. Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, pages 234–241. ISBN: 9783319245744. DOI: 10.1007/978-3-319-24574-4_28.

Sarker, I. H. (2021). "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions". *SN Computer Science* 2(6). DOI: 10.1007/s42979-021-00815-1.

Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona (2012). "Fiji: an open-source platform for biological-image analysis". *Nature Methods* 9(7), pages 676–682. DOI: 10.1038/nmeth.2019.

Shea, T. and J. E. Hammer (2013). "Kinetics of cooling- and decompression-induced crystallization in hydrous mafic-intermediate magmas". *Journal of Volcanology and Geothermal Research* 260, pages 127–145. DOI: 10.1016/j.jvolgeores.2013.04.018.

Shea, T., B. F. Houghton, L. Gurioli, K. V. Cashman, J. E. Hammer, and B. J. Hobden (2010). "Textural studies of vesicles in volcanic rocks: An integrated methodology". *Journal of Volcanology and Geothermal Research* 190(3–4), pages 271–289. DOI: 10.1016/j.jvolgeores.2009.12.003.

Sheldrake, T. and O. Higgins (2021). "Classification, segmentation and correlation of zoned minerals". *Computers & Geosciences* 156, page 104876. DOI: 10.1016/j.cageo.2021.104876.

Valdivia, P., A. A. Marshall, B. D. Brand, M. Manga, and C. Huber (2022). "Mafic explosive volcanism at Llaima Volcano: 3D x-ray microtomography reconstruction of pyroclasts to constrain shallow conduit processes". *Bulletin of Volcanology* 84(1). DOI: 10.1007/s00445-021-01514-8.

Wallace, P. J. and I. S. E. Carmichael (1994). "Petrology of Volcán Tequila, Jalisco, Mexico: disequilibrium phenocryst assemblages and evolution of the subvolcanic magma system". *Contributions to Mineralogy and Petrology* 117(4), pages 345–361. DOI: 10.1007/bf00307270.

Waythomas, C. F., K. Angeli, and R. L. Wessels (2020a). "Evolution of the submarine–subaerial edifice of Bogoslof volcano, Alaska, during its 2016–2017 eruption based on analysis of satellite imagery". *Bulletin of Volcanology* 82(2). DOI: 10.1007/s00445-020-1363-0.

Waythomas, C. F., M. Loewen, K. L. Wallace, C. E. Cameron, and J. F. Larsen (2020b). "Geology and eruptive history of Bogoslof volcano". *Bulletin of Volcanology* 82(2). DOI: 10.1007/s00445-019-1352-3.

Wright, H. M. N., K. V. Cashman, M. Rosi, and R. Cioni (2007). "Breadcrust bombs as indicators of Vulcanian eruption dynamics at Guagua Pichincha volcano, Ecuador". *Bulletin of Volcanology* 69(3), pages 281–300. DOI: 10.1007/s00445-006-0073-6.

Yasuda, A. and N. Hokanishi (2022). "Introduction to Automated Tools for the Analysis of Volcanic Ejecta Built on an Analysis Platform Developed in the INeVRH Project". *Journal of Disaster Research* 17(5), pages 683–693. DOI: 10.20965/jdr.2022.p0683.

Zhou, Z., M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang (2018). "UNet++: A Nested U-Net Architecture for Medical Image Segmentation". *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer International Publishing, pages 3–11. ISBN: 9783030008895. DOI: 10.1007/978-3-030-00889-5_1.