Grain Segmentation in Ultraviolet, Infrared, X-ray and Low-Quality Images

Reece van der Bank[a,*], Geoff Nitschke[a], Chi Ly[b]

[a]*Department of Computer Science, University of Cape Town, Cape Town, South Africa*
[b]*Bruker Nano Analytics, Houston, Texas, USA*

Abstract

The issue of segmenting every grain in an image is one of the most important forever ongoing problems in automated mineralogy. Ultraviolet (UV), infrared (IR) and X-ray images reveal grains that visible-light (VIS) images do not, yet the general-purpose "segment anything" models - widly used in industry - promise to segment them automatically. In this work we argue that non-visible images are worth segmenting in their own right, and report, at pixel and grain level, how existing one-shot and training-free models (more than 20 screened, four retained after tuning) behave under the imaging conditions we tested. On matched fields of view, the best tuned model carries over to one of three non-visible bands. On different imaging domains (X-ray sections, blurred and noisy images) the tuned settings still fail. A prompted alternative helps in some regimes, including held-out UV images, but does not yet generalise reliably. Our findings suggest the flaws in existing industry standard models, which limits the best possible exploitation of advanced imaging methods, and give some directions for further research.

Keywords: Automated Mineralogy, Grain Segmentation, Foundation Models, Multispectral Imaging, Machine Learning

1. Introduction

The two major steps in automated mineralogy are acquiring an image of the sample and segmenting that image into grains. Only then can modal mineralogy, liberation and deportment be computed (Gottlieb et al., 2000; Fandrich et al., 2007; Becker et al., 2016; Hrstka et al., 2018; Schulz et al., 2020). Whether every grain is outlined correctly greatly determines the quality of these figures, and differences in how segmentation is set up contribute to the spread seen between laboratories (Lastra and Paktunc, 2016).

Globally, we distinguish three forms of segmentation. By *classical* segmentation we mean thresholding (Otsu, 1979) followed by watershed splitting (Vincent and Soille, 1991), which needs a clean background and adjustment by hand. By *trained* segmentation we mean models learned from labelled examples, from Cellpose (Stringer et al., 2021) and StarDist (Schmidt et al., 2018) to mineral networks for SEM (Chen et al., 2020) and optical thin-section images (Koh et al., 2021). These need new labels for every new setup. By *training-free automatic* segmentation we mean foundation models, such as Segment Anything (SAM, Kirillov et al., 2023) and its efficient derivatives (Xiong et al., 2024; Zhang et al., 2024, 2023; Zhao et al., 2023), that segment an unseen image with no task-specific training and no human prompts. Their settings may still be tuned, and ours were.

We are aware of other work that combines modalities. Juránek et al. (2022) fuse SEM images with sparse EDS spectra, and Vellappally et al. (2024) run SAM at several scales on thin sections, in visible light only. MM-SAM (Xiao et al., 2024) adapts SAM to other sensors using paired training data. We did not test it, but that requirement sits poorly with the range of mineral images we expect. The accuracy of foundation models is known to vary across modalities (Mazurowski et al., 2023; Huang et al., 2024).

This study has a two-fold objective. First, we show what ultraviolet (UV) and X-ray images reveal that visible-light (VIS) images do not. Second, we report how existing training-free models behave. We ask whether tuned settings transfer *across bands* on the same field of view, and whether they transfer to *different imaging domains*, where specimen, preparation and resolution change as well.

2. Why grain segmentation beyond visible light is difficult

Foundation models learn what an object looks like from natural photographs, which are well lit, in focus and show objects many pixels across. Mineralogical images break each of these assumptions (Fig. 1).

- Non-visible illumination changes what separates a grain from its background. Under IR dark grains match the backing, and under UV2 the contrast collapses.
- Fast, low-magnification imaging gives blurred, noisy images with grains only 9–10 px across.
- Blur and dense packing merge neighbouring grains.

Moreover, scores at these resolutions are sensitive to boundaries. For a circular grain of diameter d, a uniform one-pixel boundary error gives IoU $= (d/(d+2))^2$ or $((d-2)/d)^2$ (Fig. 2). Dilating or eroding every grain in our masks of the low-quality images

*Corresponding author.
Email: vbnree001@myuct.ac.za (Reece van der Bank), gnitschke@cs.uct.ac.za (Geoff Nitschke), Chi.Ly@bruker.com (Chi Ly)

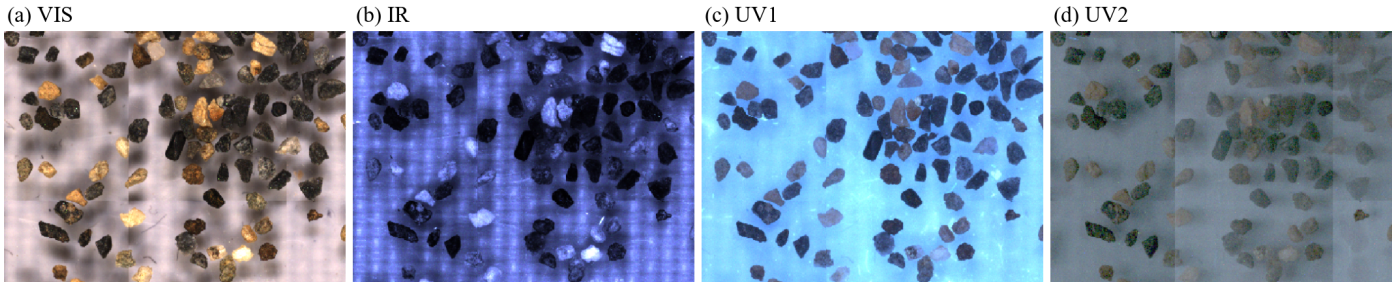

Figure 1: The same field of view under (a) visible (VIS), (b) infrared (IR) and (c, d) two ultraviolet (UV1, UV2) illuminations. VIS and IR show a dark halo around each grain and, in IR, a woven backing mesh. UV1 and UV2 are largely halo-free, but UV2 has little grain–background contrast.

by one pixel, scored as pooled foreground IoU, gives 0.66–0.76, and the two annotation sources agree at IoU 0.85 and PQ 0.70–0.82 (Table A.8). These figures measure how much error lies within a pixel of the boundary.

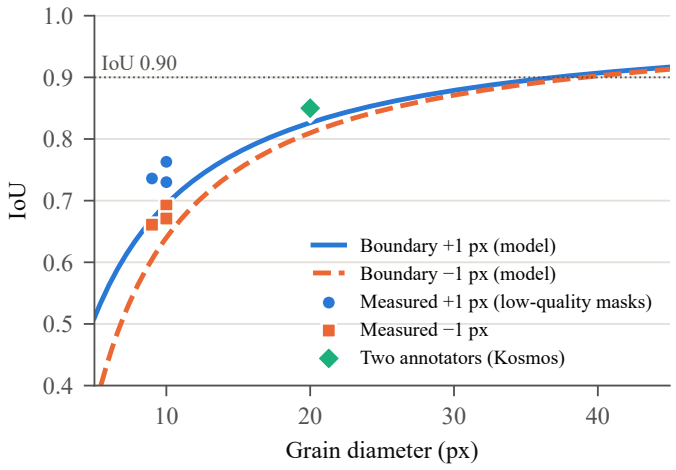

Figure 2: Boundary sensitivity. IoU of an otherwise perfect segmentation of circular grains after a uniform 1 px dilation or erosion, against grain diameter (lines), the same operation applied to our masks of the low-quality images and scored as pooled foreground IoU (points), and the agreement between the two annotation sources.

3. What value non-visible imaging adds

So, why segment non-visible images at all? Roughly speaking, because other bands record other physical properties. Under UV some minerals fluoresce while most groundmass stays dark, and X-rays respond to density and atomic number. Instruments such as the Bruker M4 TORNADO capture these views in register with one another. In a kimberlite section, grains that fluoresce strongly under UV1 have no visible counterpart in VIS (Fig. 3). In a volcanic sand section, small bright crystals that are clear in the X-ray image are absent from the optical image, and contrast enhancement (Zuiderveld, 1994) does not recover them (Fig. 4). Even on the X-ray image, tuned SAM takes the glass matrix as one grain (Fig. 4c), whereas the ideal mask separates the faults as distinct features (Fig. 4d). These bands are worth segmenting.

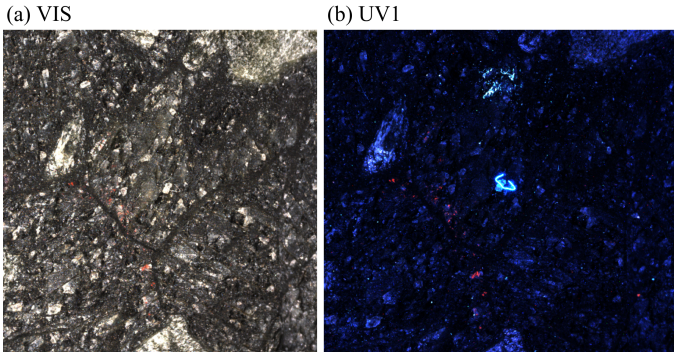

Figure 3: Sample of the same kimberlite section in (a) VIS and (b) UV.

4. Approach

4.1. Images and targets

Table 1 lists the images. The low-quality set matches Kosmos in pixel dimensions but is blurred and noisy, with grains half the size. A grain is a particle on the backing in loose sand, a crystal or vesicle against the glass matrix in volcanic sections, and a grain within the groundmass in kimberlite.

4.2. Models and tuning

Of more than 20 models screened, we tuned eight and retained four (Appendix Appendix A). SAM ViT-B, EfficientSAM ViT-Tiny and EfficientViT-SAM L0 prompt a grid of points, and FastSAM-s (Jocher et al., 2023) segments in one pass. Each hyperparameter was swept one at a time on three VIS views, and the tuned settings were applied unchanged elsewhere.

4.3. A prompted alternative

We also separate *finding* grains from *outlining* them (Fig. 5). Peaks of a Laplacian-of-Gaussian response (Lindeberg, 1998) give grain centres, each becomes a point prompt for EfficientSAM, and a background colour model (Mahalanobis, 1936) trims spillover and adds missed grain pixels.

4.4. Evaluation

Overall IoU compares all predicted with all reference grain pixels but cannot tell separated grains from merged ones, so we also report panoptic quality (PQ, Kirillov et al., 2019), grain precision and recall at IoU ≥ 0.5, masks per grain and the error

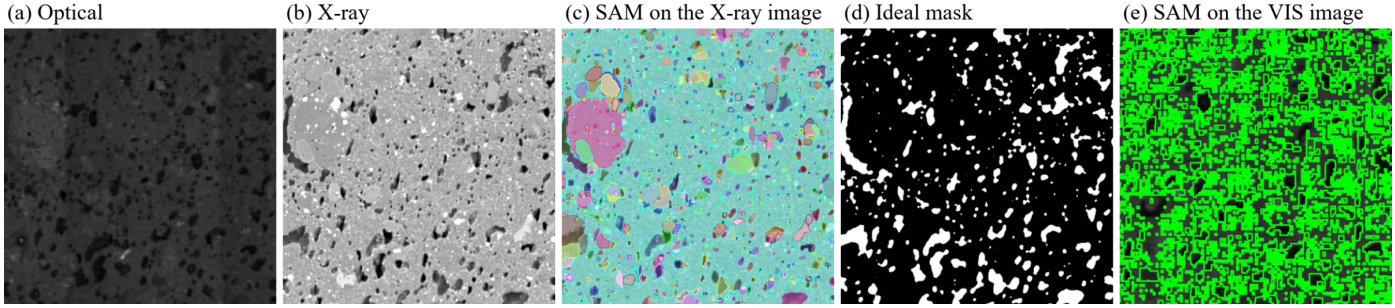

Figure 4: Showing the benefits and issues with existing methods of segmentation when applied to volcanic sands images. The same volcanic sand section in (a) an optical image and (b) an X-ray image. (c) Tuned SAM on the X-ray image. (d) An ideal mask for removing faults. (e) Same SAM on the VIS image.

Table 1: Images used. VIS, IR and UV images of a view show the same grains. AMICS masks are expert annotations. Ours were made with a point-prompt tool and corrected by hand.

Set	Views	Bands	Pixels	Grains	Masks	Role
Kosmos sand, 3 samples	15×3	VIS	752×648	9,601	AMICS	3/12 view tune $\times 3$ seeding averaged
Low-quality sand	10	VIS	512×512	3,204	AMICS	prompted-method development
Volcanic sand sections	18	VIS, X-ray, IR	1952×744	–	ours	visual test (Figs. 4, 7)
Kimberlite sections	9	VIS, UV1	6227×4923	–	ours	illustration (Fig. 3)

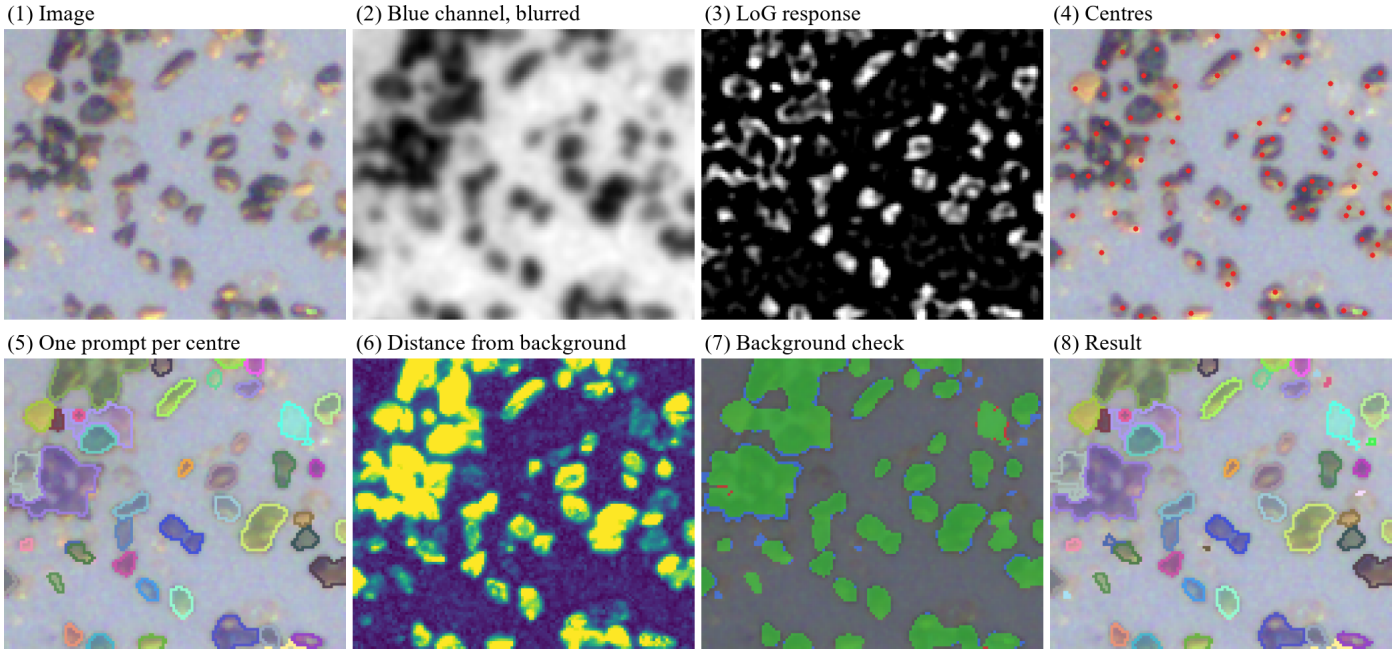

Figure 5: The prompted method on a crop of a low-quality image (enlarged 4×), showing (1) image, (2) blurred blue channel, (3) Laplacian-of-Gaussian response, (4) detected centres, (5) one EfficientSAM prompt per centre, (6) distance from the background colour model, (7) background check (red trimmed, blue added), (8) result.

in median grain size (D50). Scores are means over images, with ± one standard deviation (Appendix Appendix A).

5. Results & Discussion

5.1. Visible light

At default settings every model was precise but incomplete (Table 2), because the grid-prompted models ask 32×32 points and FastSAM returns at most 300 objects (Fig. 6a). Tuning lifted FastSAM to IoU 0.874 ± 0.041 and PQ 0.804, with median grain size within 0.3 %. For SAM and EfficientViT-SAM, tuning raised IoU but lowered PQ (SAM 0.710 to 0.429), as denser prompting split each grain into three to thirteen masks (Fig. 6b, c) and shrank median grain size by 55–71 %. Held-out PQ was within 0.003 of the full value for every model. Tuning one parameter at a time also misses interactions (Eiben et al., 1999), so the ranking reflects the tuning method too.

5.2. Across bands, on matched views

This is the controlled comparison, where only the band changes (Table 3). FastSAM held up on UV1 (IoU 0.880, PQ

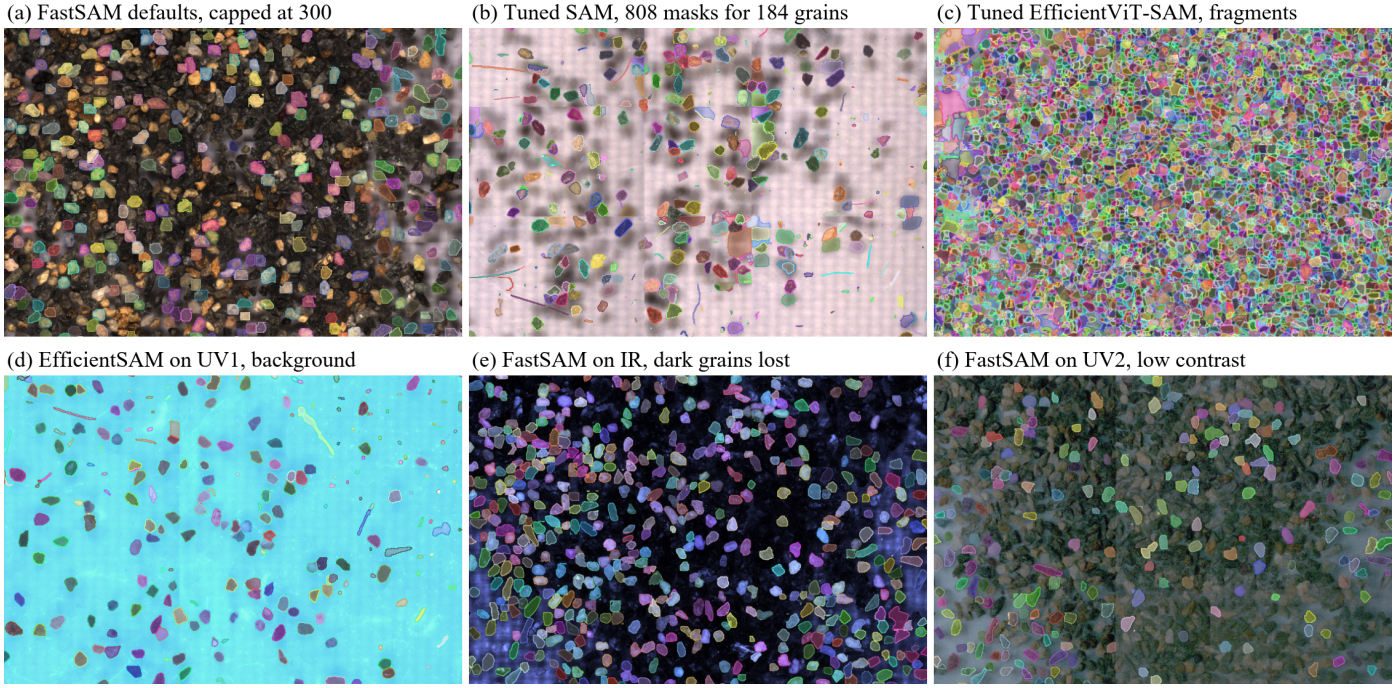

Figure 6: Typical failures of one-shot segmentation. (a) FastSAM at default settings stops at 300 masks on a dense image of 1,066 grains. (b) Tuned SAM on a sparse image gives 808 masks for 184 grains, including fibres, scratches and the dish rim. (c) Tuned EfficientViT-SAM gives over 10,000 fragments on a dense image. (d) EfficientSAM with VIS-tuned settings on UV1 outlines the background (overall IoU 0.097). (e) FastSAM on IR misses dark grains on the dark mesh. (f) FastSAM on UV2 gives 236 masks for 1,112 grains.

Table 2: Default and tuned settings on the VIS images, mean over the 15 views. Precision, recall, masks per grain and D50 error are for the tuned settings. Held-out PQ excludes the three tuning views. Time is inference and post-processing per image.

	IoU		PQ				Masks	D50	PQ	
Model	Default	Tuned	Default	Tuned	Precision	Recall	per grain	error	held out	s/image
FastSAM-s	0.532	**0.874**	0.564	**0.804**	**0.890**	0.947	1.06	−0.3 %	**0.807**	1.4
EfficientSAM ViT-T	0.637	0.825	0.701	0.720	0.687	**0.963**	1.45	−8.5 %	0.723	17.8
SAM ViT-B	0.664	0.729	0.710	0.429	0.318	0.958	3.19	−55.1 %	0.428	58.7
EfficientViT-SAM L0	0.103	0.418	0.152	0.057	0.040	0.504	12.64	−70.5 %	0.059	235.4

0.809, against 0.893 and 0.831 on VIS) and fell on IR and UV2. EfficientSAM's IoU fell on UV1 to 0.571 but its PQ did not (0.720 against 0.719), as thresholds loosened for VIS let background masks through (Fig. 6d), so the loss is at pixel level. Counting a grain as found if matched in either VIS or UV1, without fusing the segmentations, covers 97.5 % of grains against 95.1 % for VIS alone.

Table 3: IoU and PQ per band on the 11 matched views imaged in every band, with VIS-tuned settings.

	FastSAM		EfficientSAM		SAM	
Band	IoU	PQ	IoU	PQ	IoU	PQ
VIS	**0.893**	**0.831**	0.819	0.719	0.688	0.412
IR	**0.665**	**0.692**	0.645	0.607	0.597	0.368
UV1	**0.880**	**0.809**	0.571	0.720	0.488	0.398
UV2	**0.557**	0.589	0.408	**0.679**	0.539	0.433

5.3. *Different imaging domains*

These comparisons also change specimen, preparation, grain size and target, so they show that the tuned settings fail without isolating the cause.

On X-ray images of three volcanic sand sections (Fig. 7), which have no reference masks, FastSAM outlined only 4–5 % of each image, while in five of six runs EfficientSAM and SAM took the glass matrix as one grain covering 52–84 % of it. The prompted method, never tuned on X-ray images, did best by visual inspection, outlining more crystals and vesicles than the others, though it fragments parts of the matrix.

On the low-quality images, tuned FastSAM outlined only 2–12 % of each image (Fig. A.8). A brute-force search reached 0.85 on its two tuning images but returned square background masks on new ones (IoU 0.261). Blur followed by Otsu thresholding reached IoU 0.69 but PQ only 0.1–0.2, as it merges touching grains.

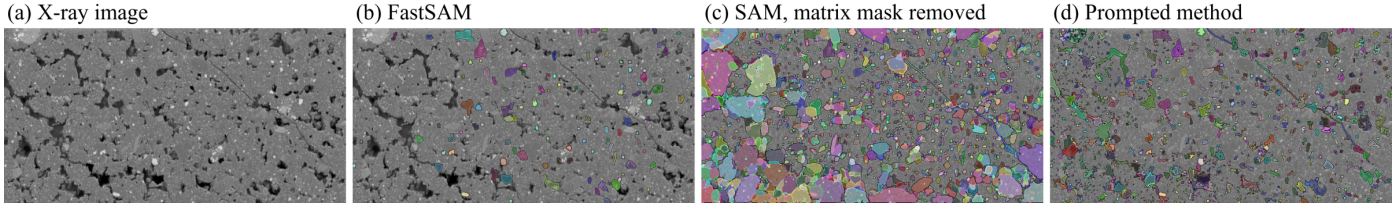

Figure 7: X-ray image of a volcanic sand section, showing (a) the image, (b) FastSAM and (c) SAM with VIS-tuned settings, and (d) the prompted method. SAM's largest mask, the glass matrix taken as one grain (52 % of the image), is removed in (c) so the rest can be seen. Each mask is drawn in its own colour. While the prompting outperforms SAM, it still underperforms from ideal as a one-shot alternative in this case.

5.4. The prompted alternative

On UV1 views neither method was tuned on, the prompted method reached IoU 0.917 against FastSAM's 0.878 (Table 4). On our five masks, where its settings were developed, it reaches 0.799 (0.719 without the background check) and PQ 0.426, while cross-validation with an earlier version of the check gave the fairer 0.713 (0.667–0.762). It fails on dense samples, because when grains cover 70 % of the image the background model learns the grain colour (Fig. A.9).

Table 4: Prompted method. The top half compares it with tuned FastSAM by overall IoU against the AMICS masks on the sparse views. UV1 (held out) covers only the UV1 views that neither method was tuned on, while UV1 covers all five. The bottom half gives development figures on our masks, which the prompted method was tuned on.

Band	Prompted	FastSAM
UV1 (held out)	**0.917**	0.878
UV1	**0.918**	0.880
VIS	0.623	**0.906**
IR	0.607	**0.718**
Our masks	IoU	PQ
Low-quality VIS	0.79–0.89	0.32–0.65
IR	0.631	0.327
UV1	0.890	0.866
All	0.799	0.426

5.5. Future Work

As these are still preliminary results, the reference set is rather small but varied, its band images repeat the same grains, our masks were sourced from various individuals with various opionions of 'ideal' segments. We tested no simpler baselines beyond thresholding, so the separate contributions of centre detection, EfficientSAM and the background check remain open. Under the conditions tested, these pipelines show measurable failure modes, and the prompted alternative helps in some regimes but does not yet generalise reliably.

Further more there are still several open research issues. Training band-specific segmenters from transferred VIS masks, with a separate check for band-only grains. Reference masks for X-ray images, and controlled degradations that isolate the band from the specimen. Baselines such as watershed or region growing from the detected centres, and ablations of each component.

6. Contributions

In this work we argued that non-visible images are worth segmenting, and found that, under the conditions we tested, existing training-free models transfer poorly. On matched views the best tuned model carries over to UV1 only, and on X-ray and low-quality images the tuned settings fail. Pixel-level scores also hide grain-level failures, which PQ and grain size expose. A prompted alternative helps on held-out UV images and, judged visually, on X-ray sections, but fails on dense samples. Methods that segment the band images directly, and are checked at grain level, are still needed.

Acknowledgements

We thank an experienced industry mineralogist, who wishes to remain anonymous, for supplying the AMICS reference masks, validating their accuracy and assisting with real world insights. We also wish to thank Andrew Menzies (of Bruker Nano Analytics) for supplying various multispectral images, and for his preliminary advice on volcanic sands and on the value of multispectral imaging for them.

References

Becker, M., Wightman, E.M., Evans, C.L. (Eds.), 2016. Process Mineralogy. Number 6 in JKMRC Monograph Series in Mining and Mineral Processing, Julius Kruttschnitt Mineral Research Centre, Brisbane.

Chen, Z., Liu, X., Yang, J., Little, E., Zhou, Y., 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, 104450.

Eiben, A.E., Hinterding, R., Michalewicz, Z., 1999. Parameter control in evolutionary algorithms. IEEE Transactions on Evolutionary Computation 3, 124–141.

Fandrich, R., Gu, Y., Burrows, D., Moeller, K., 2007. Modern SEM-based mineral liberation analysis. International Journal of Mineral Processing 84, 310–320. doi:10.1016/j.minpro.2006.07.018.

Gottlieb, P., Wilkie, G., Sutherland, D., Ho-Tun, E., Suthers, S., Perera, K., Jenkins, B., Spencer, S., Butcher, A., Rayner, J., 2000. Using quantitative electron microscopy for process mineralogy applications. JOM 52, 24–25. doi:10.1007/s11837-000-0126-9.

Hrstka, T., Gottlieb, P., Skála, R., Breiter, K., Motl, D., 2018. Automated mineralogy and petrology – applications of TESCAN integrated mineral analyzer (TIMA). Journal of Geosciences 63, 47–63. doi:10.3190/jgeosci.250.

Huang, Y., Yang, X., Liu, L., et al., 2024. Segment anything model for medical images? Medical Image Analysis 92, 103061.

Jocher, G., Chaurasia, A., Qiu, J., 2023. Ultralytics YOLOv8. https://github.com/ultralytics/ultralytics.

Juránek, R., Výravský, J., Kolář, M., Motl, D., Zemčík, P., 2022. Graph-based deep learning segmentation of EDS spectral images for automated mineral phase analysis. Computers & Geosciences 165, 105109. doi:10.1016/j.cageo.2022.105109.

Kirillov, A., He, K., Girshick, R., Rother, C., Dollár, P., 2019. Panoptic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9404–9413.

Kirillov, A., Mintun, E., Ravi, N., et al., 2023. Segment anything, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4015–4026.

Koh, E.J., Amini, E., McLachlan, G.J., Beaton, N., 2021. Utilising convolutional neural networks to perform fast automated modal mineralogy analysis for thin-section optical microscopy. Minerals Engineering 173, 107230.

Lastra, R., Paktunc, D., 2016. An estimation of the variability in automated quantitative mineralogy measurements through inter-laboratory testing. Minerals Engineering 95, 138–145. doi:10.1016/j.mineng.2016.06.025.

Lindeberg, T., 1998. Feature detection with automatic scale selection. International Journal of Computer Vision 30, 79–116.

Mahalanobis, P.C., 1936. On the generalised distance in statistics. Proceedings of the National Institute of Sciences of India 2, 49–55.

Mazurowski, M.A., Dong, H., Gu, H., Yang, J., Konz, N., Zhang, Y., 2023. Segment anything model for medical image analysis: An experimental study. Medical Image Analysis 89, 102918.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics 9, 62–66.

Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O.R., Jagersand, M., 2020. U^2-Net: Going deeper with nested U-structure for salient object detection. Pattern Recognition 106, 107404.

Ravi, N., Gabeur, V., Hu, Y.T., et al., 2025. SAM 2: Segment anything in images and videos, in: International Conference on Learning Representations (ICLR).

Schmidt, U., Weigert, M., Broaddus, C., Myers, G., 2018. Cell detection with star-convex polygons, in: Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 265–273.

Schulz, B., Sandmann, D., Gilbricht, S., 2020. SEM-based automated mineralogy and its application in geo- and material sciences. Minerals 10, 1004. doi:10.3390/min10111004.

Stringer, C., Wang, T., Michaelos, M., Pachitariu, M., 2021. Cellpose: A generalist algorithm for cellular segmentation. Nature Methods 18, 100–106.

Vellappally, A., Hou, S., Emmings, J., 2024. Automated grain segmentation and mineral classification in rock thin sections, in: 85th EAGE Annual Conference & Exhibition. doi:10.3997/2214-4609.2024101405.

Vincent, L., Soille, P., 1991. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence 13, 583–598.

Xiao, A., Xuan, W., Qi, H., Xing, Y., Yokoya, N., Lu, S., 2024. Segment anything with multiple modalities. arXiv preprint arXiv:2408.09085 .

Xiong, Y., Varadarajan, B., Wu, L., et al., 2024. EfficientSAM: Leveraged masked image pretraining for efficient segment anything, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 16111–16121.

Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., Sang, N., 2018. BiSeNet: Bilateral segmentation network for real-time semantic segmentation, in: Proceedings of the European Conference on Computer Vision (ECCV), pp. 325–341.

Zhang, C., Han, D., Qiao, Y., et al., 2023. Faster segment anything: Towards lightweight SAM for mobile applications. arXiv preprint arXiv:2306.14289 .

Zhang, Z., Cai, H., Han, S., 2024. EfficientViT-SAM: Accelerated segment anything model without accuracy loss. arXiv preprint arXiv:2402.05008 .

Zhao, X., Ding, W., An, Y., et al., 2023. Fast segment anything. arXiv preprint arXiv:2306.12156 .

Zuiderveld, K., 1994. Contrast limited adaptive histogram equalization, in: Heckbert, P.S. (Ed.), Graphics Gems IV. Academic Press, pp. 474–485.

Appendix A. Reproducibility and supplementary results

Hardware and software.. One desktop computer with an NVIDIA GeForce RTX 5060 Ti GPU, Python 3.12, PyTorch 2.11 (CUDA 12.8) and ONNX Runtime 1.31. Times are inference and post-processing per 752×648 image, after one warm-up image, excluding model loading. The GPU was at times shared, so they are indicative only. Scores are the mean over images of per-image values. Precision, recall and PQ use a one-to-one (Hungarian) matching at IoU ≥ 0.5. D50 is the number-weighted median equivalent-circle diameter at 41.31 μm per pixel. Figs. A.8 and A.9 show two further failure cases. Table A.5 gives the screening outcome, Table A.6 the search space and tuned settings, Table A.7 the per-view results and Table A.8 the agreement between annotation sources.

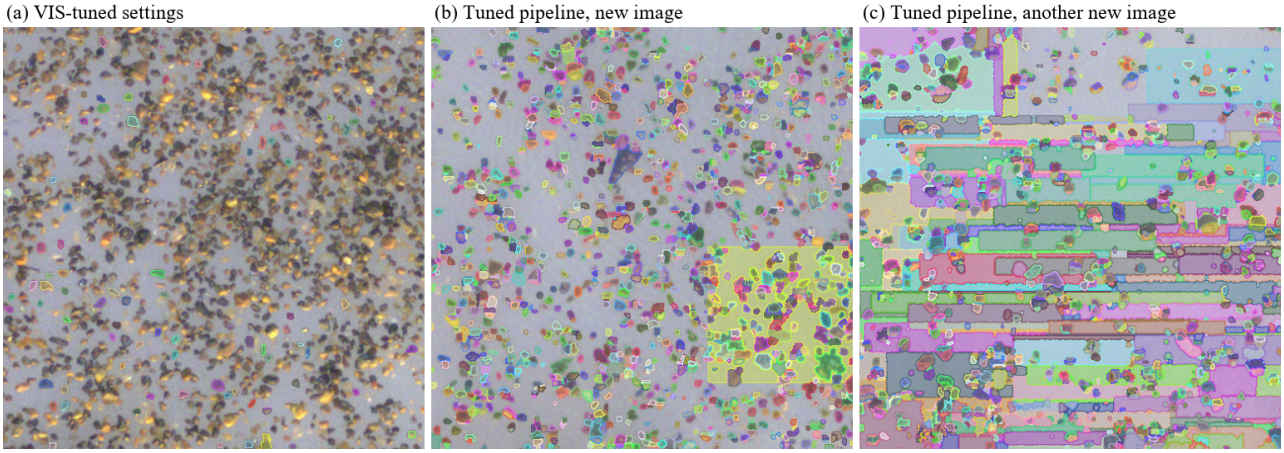

Figure A.8: FastSAM on low-quality images. (a) VIS-tuned settings outline 2.4 % of the image. (b, c) A pipeline tuned on two other images returns square background masks on new images, covering 92 % of the image in (c). Even with tuning the system still struggles, and although such masks are easy to remove in post-processing, they remain an issue for full automation.

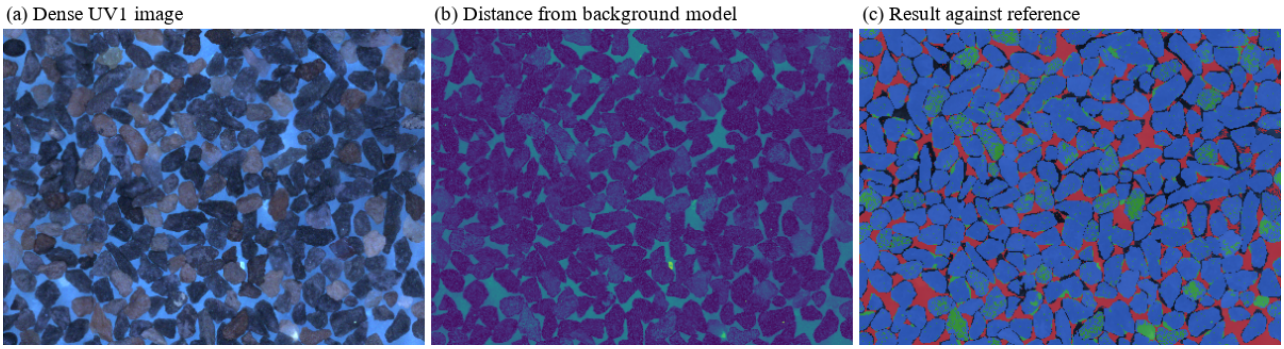

Figure A.9: Where the prompted method fails. (a) Dense UV1 image, where grains cover 70 % of the frame. (b) The fitted "background" is the grain colour, so grains are close to it. (c) Against the reference, grains are missed (blue) and the gaps are outlined (red). One of the easiest ways to boost accuracy in UV imaging is to account for lost colour channels, but high grain counts make it harder to separate background from foreground automatically, again an issue for full automation.

Table A.5: Screening. Each model was first run at default settings. Retained models are those reported in this study.

Stage	Models
Retained	SAM ViT-B, EfficientSAM ViT-Tiny EfficientViT-SAM L0, FastSAM-s
Screened	SAM 2.1, μSAM, TinySAM, EfficientTAM, EdgeSAM, SAM 2 (Ravi et al., 2025), MobileSAM, U^2-Net (Qin et al., 2020), BiSeNet (Yu et al., 2018), and other segmentation models

Table A.6: Search space. Each hyperparameter was swept from minimum to maximum in 10 % steps with the others at default, on the three tuning views, and the best values were then combined. Points per batch affects speed only.

Model	Hyperparameter	Range	Default	Tuned
FastSAM-s	image size (px)	512–2048	1024	2048
	confidence	0–1	0.4	0.5
	NMS IoU	0–1	0.9	0.9
	maximum detections	100–3000	300	2130
EfficientSAM ViT-T	points per side	16–96	32	96
	predicted-IoU threshold	0–1	0.7	0.1
	stability threshold	0–1	0.9	0.9
	box NMS	0–1	0.7	0.7
	minimum region area (px)	0–500	100	100
SAM ViT-B	points per side	16–96	32	80
	predicted-IoU threshold	0–1	0.88	0.3
	stability threshold	0–1	0.95	0.9
	box NMS	0–1	0.7	0.7
	crop layers	0–2	0	1
	crop overlap	0–0.5	0.34	0.2
	crop point downscale	1–4	1	1
	minimum region area (px)	0–500	0	0
EfficientViT-SAM L0	points per side	16–96	32	72
	predicted-IoU threshold	0–1	0.8	0.4
	stability threshold	0–1	0.85	0.4
	box NMS	0–1	0.7	0.7
	crop layers	0–2	0	2
	crop overlap	0–0.5	0.34	0.1
	crop point downscale	1–4	1	1
	minimum region area (px)	0–500	100	100

Table A.7: Per-view results on the VIS images with tuned settings (IoU, PQ). * Tuning view.

		FastSAM		EfficientSAM		SAM		EfficientViT-SAM	
View	Grains	IoU	PQ	IoU	PQ	IoU	PQ	IoU	PQ
Sparse 1*	200	0.910	0.841	0.766	0.656	0.700	0.372	0.132	0.047
Sparse 2	184	0.895	0.841	0.744	0.649	0.635	0.350	0.122	0.055
Sparse 3	156	0.901	0.857	0.768	0.646	0.405	0.331	0.097	0.072
Sparse 4	159	0.916	0.859	0.779	0.664	0.429	0.343	0.104	0.081
Sparse 5	166	0.908	0.861	0.726	0.664	0.417	0.342	0.104	0.060
Medium 1*	709	0.870	0.794	0.838	0.738	0.815	0.458	0.502	0.046
Medium 2	711	0.901	0.824	0.881	0.783	0.821	0.479	0.499	0.058
Medium 3	496	0.907	0.853	0.889	0.787	0.855	0.465	0.454	0.077
Medium 4	649	0.900	0.825	0.894	0.796	0.835	0.466	0.423	0.062
Medium 5	656	0.898	0.835	0.878	0.793	0.797	0.462	0.435	0.053
Dense 1*	1112	0.815	0.746	0.847	0.733	0.853	0.468	0.697	0.046
Dense 2	1087	0.808	0.714	0.831	0.723	0.847	0.461	0.669	0.054
Dense 3	1147	0.823	0.721	0.835	0.704	0.848	0.474	0.685	0.056
Dense 4	1066	0.835	0.751	0.848	0.726	0.840	0.470	0.661	0.038
Dense 5	1103	0.819	0.742	0.847	0.738	0.840	0.492	0.694	0.044

Table A.8: Agreement between the two annotation sources on the one view both labelled (AMICS on VIS, ours on IR and UV1), in the metrics used for the models.

Band	IoU	PQ	Precision	Recall	Grains (AMICS / ours)
IR	0.846	0.699	0.681	0.940	200 / 276
UV1	0.850	0.824	0.897	0.910	200 / 203