Using computer vision and drill core photography to automate geological logging workflows and improve orebody knowledge

Nathanael Pittaway
Datarock Pty Ltd
nathanaelpittaway@datarock.com.au

Harvey Nguyen
Datarock Pty Ltd
harveynguyen@datarock.com.au

Rian Dutch
Datarock Pty Ltd
riandutch@datarock.com.au

SUMMARY

Core logging is often carried out by many geologists with different levels of skill and experience who collectively do not interpret and record their observations in the same way over the life of a project. This has traditionally led to logged datasets often being viewed as inconsistent and lacking any real auditability. Computer vision techniques address some of these issues by extracting geological information from drill core imagery with significantly improved consistency and detail compared to current manual logging. Veins are one of the most neglected and poorly logged datasets due to their complexity, scale, and volume throughout an ore body. Consequently, logged vein data often comprises subjective estimates or averages and sub-sampled detail that is inconsistently logged. Using computer vision-based techniques, we analyse traditional RGB core photography to generate new types of high-resolution vein data including vein morphology segmentations at the pixel scale as well as vein area and percentage. These novel vein data are then interrogated in 3D to demonstrate how these high-resolution vein characteristics can provide new geological insights and improve ore body knowledge. This work demonstrates the advantages of computer vision logging techniques based on their ability to create new types of logging data with improved consistency. Computer vision logging also benefits geologists by allowing them to move beyond routine and repetitive work and focus more on higher level technical tasks.

Key words: geological data, drill core, artificial intelligence, computer vision, machine learning

INTRODUCTION

Traditional drill core logging datasets have been a significant source of geological understanding and thus value for exploration and mining companies but are resource intensive in their collection (Olson et al, 2015) and often impacted by bias (Fowler, 2013). The information obtained from the collection of these datasets is dependent on the skills and expertise of the logger and due to a shortage of experienced core loggers, these logs are often left to inexperienced or junior core loggers. Additionally, the time between drilling the core and obtaining the relevant geological information could be shortened, as backlogs develop, and geological information isn't being collected in time for subsequent models. Furthermore, a lack in accuracy or timely data collection impacts ore body knowledge which can be delayed, overlooked, or poorly understood.

Automating the collection of logging data from drill core imagery has historically been seen as an achievable method to address traditional logging shortcomings. As early as the start of the century, companies have focussed on collecting logging data from controlled source core scanning workflows that can either produce a more consistent and accurate RGB imagery (Lemy, 2001; Saricam and Oztruk, 2018) or provide a 3D laser scan dataset (Olson et al, 2015; Harraden et al. 2016). These have primarily focussed on gathering geotechnical data as these datasets are less reliant on the RGB aspect of the core imagery. The costs associated with the automation methods employed to extract the required data from these bespoke imaging systems, however, are often significant and cannot be applied to the ever-growing volume of historic drilling due to sampling, degradation or even disposal of core post-drilling.

For geological modelling, traditional RGB core photography taken with a digital single-lens reflex (DSLR) camera is still the most common method of capturing core imagery. However, image quality variation due to multiple imaging workflows as DSLR technology and resolution improved has produced inconsistent datasets, meaning these images are underutilised as a data resource for modelling.

Recent developments in robust supervised and unsupervised computer vision image analysis workflows have improved the extraction of higher quality geological data from historic RGB core imagery, allowing for improvements in the efficiency, quantity and importantly, standard of geological data collected throughout the exploration and mining industries (Ashraf et al, 2021; Caté and Cevik, 2021). Additionally, these new workflows, when utilised with new drilling programs, can assist in augmenting the logging process and relieve the issues concerning skilled workforce recruitment.

This paper presents a robust computer vision image analysis workflow that can detect veining to generate both area and percentage of veining throughout a hole when compared to the host-rock. Historically, this type of geological data is generally not logged at scale. This methodology was deployed across 14 drill holes from the Mineral Systems Drilling Program, a precompetitive drilling campaign managed by the Geological Survey of South Australia (GSSA) and the Deep Exploration Technologies Cooperative Research Centre (DET CRC). These automated datasets are compared against the program's geological logs to illustrate the benefits this automation provides to data consistency, auditability and speed compared to manual logging.

THE MINERAL SYSTEMS DRILLING PROGRAM

The Mineral Systems Drilling Program (MSDP) was a drill program managed by the GSSA in collaboration with the DET CRC (Fabris et al., 2017). The program consisted of 14 diamond drill holes, drilled over 8 months along the southern margin of the Gawler Range Volcanics (GRV). The GRV are a suite of volcanics and volcaniclastics formed during the Mesoproterozoic between 1595-1587 Ma (Allen et al, 2008; Allen et al, 2003) and are associated with the Gawler IOCG metallogenic event. The aim of the program was to collect new data to improve the geological and metallogenic understanding along the southern margin of the GRV). The drilling was split into 3 phases to target the different basement rocks beneath the GRV.

Whilst the drilling did not uncover any new or significant metal concentrations, colloform banded quartz-sulphide veins were intersected. The number and density of this veining was not logged and was only recorded as major and minor lithologies or within the comments of the lithology logs. To extract quantitative information about these veins, the 14 MSDP drill holes were processed through a computer vision-based workflow described below.

VEIN SEGMENTATION MODELLING METHODOLOGY

Overview

The automated vein segmentation modelling workflow includes several steps to quantify the veining area and percentage from core imagery. A general quartz vein detection model (trained on other datasets) was used as an initial model that will be adapted and re-trained for this processing application using the steps summarised in Figure 1.

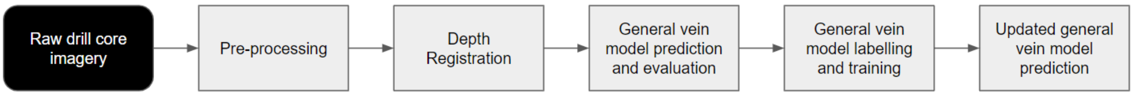

Figure 1. Overview of the steps to predict and extract vein segmentation area and percentage from drill core imagery.

Pre-processing

The first step once the raw imagery has been collated is to process the core photos on the Datarock Core platform to turn them into analysis-ready images prior to depth registration. These pre-processing steps include de-warping of images and the cropping of trays and core rows.

Depth Registration

Depth registration occurs once the images have been processed and refers to the assignment of correct depths to drill core imagery. This step is imperative, as generating geological data without accurate depths has little value whilst assigning correct depths to images allows for reliable comparisons of logging data or other geological analysis.

There are several steps to assigning depths to each pixel on a drill core image and these include identification of coherent and incoherent rock within core rows, quantification of compaction of incoherent rock, accounting for core-loss and metadata entry errors, and optical character recognition of depth annotations. This paper utilises the Datarock Core platform to generate depth registered and analysis ready core row images.

Vein Segmentation

This study will create a vein segmentation model by adapting an existing general vein segmentation model (train on data outside of this project). This will involve adding additional training data from these 14 drill holes a selection of to ensure the segmentation model will perform accurately in this application.

The steps involved for adapting an initial generalised vein segmentation model include (i) running the general model across the 14 drill holes to evaluate examples of model confusion with respect to a validation set, (ii) selection of additional training data for labelling, and (iii) retraining and re-evaluating the model until acceptable results were produced.

For this study, one round of retraining was completed by selecting 60 depth registered core row imagery containing veins that were poorly detected by the initial model. These row images were added to Computer Vision Annotation Tool (CVAT), an open-source web-based image annotation tool. The row images were then labelled by Datarock geologists who manually created polygons around individual vein examples. The labelling focussed primarily on maintaining a general vein model and so included quartz, quartz-carbonate, quartz-calcite, and calcite veining. The false predictions that were being re-trained were primarily due to the model not having seen many intrusive host-rocks and confusion was greatest throughout the gneissic rock units (Figure 2). Other lithologies which caused confusion included the lighter shale bands in the sediments of Phase 1 holes (Figure 2). The entire relabelling process took 2-3 hours for two Datarock geologists to complete.

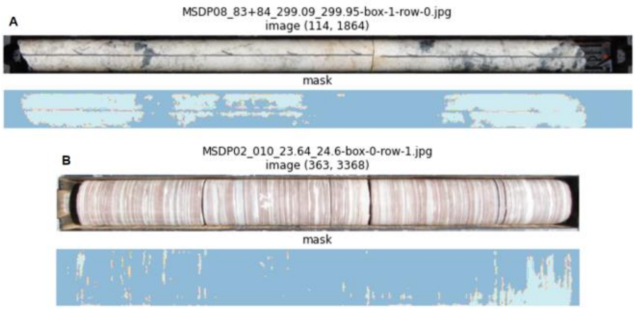

Figure 2. Example of model confusion from the vein segmentation model when encountering gneiss (a) and banded shale (b) lithologies as indicated by the model masking sections of the core with no veining observed.

The new labelled core rows were then added to the existing labelled data and a new model was trained. This new model was then applied to all 14 MSDP holes with the new predictions again being evaluated by Datarock geologists.

Calculating Vein Area and Percentage

The final process in the vein segmentation model workflow is the calculation of vein area compared to rock for a particular core row. The vein area is calculated by summing the pixels within the predicted vein segments and those pixels which fall within coherent and incoherent rock predictions as pre-determined by the first step of the depth registration stage discussed above. This is then converted into an area as seen in Figure 3 (in this case centimetre squared) and from the difference in area between the two vein and non-vein polygons, a percentage of vein for each row is produced. These two calculations, vein/rock area and vein/rock percentage, create a linear dataset which was not captured with the original logging provided (Figure 4).

Figure 3. Example of vein segmentation prediction masks superimposed on two core rows with calculated vein and rock area. (a) MSDP11 between 203.03m - 203.93m; and (b) MSDP04 between 583.6m - 584.47m.

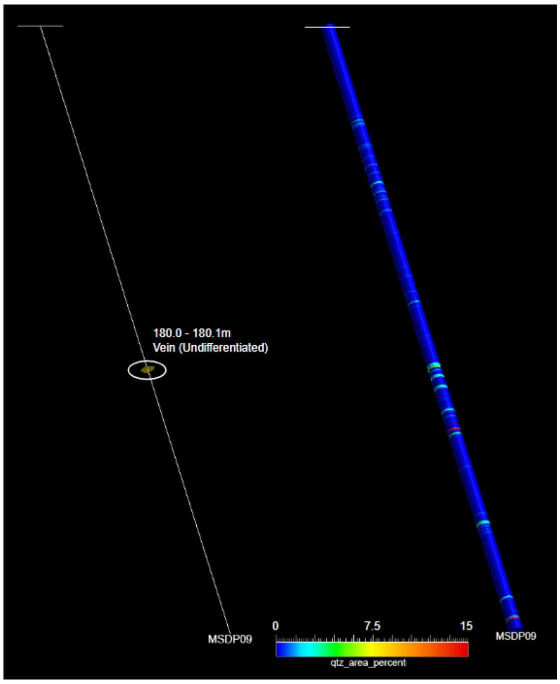

Figure 4. Example of comparison between logged veining and the vein segmentation percentage dataset from MSDP09. Coverage in the vein segmentation dataset not only includes the logged veining at 180m but also captures zones of greater veining downhole.

RESULTS AND DISCUSSION

Analysis of Vein Segmentation Before and After Training

The consistency and accuracy of the data generated pre- and post-training of the general vein segmentation model has been evaluated below. Table 1 illustrates the over-predicting of vein material pre-training where there was on average almost 5x the vein percentage per row being predicted. This resulted in a much higher percentage of total vein material per hole being reported, 1.34% compared to 0.27% post-training.

General Vein Segmentation Model	***Row Count >0.00cm² Vein***	***Average Vein % per Row***	***Average Total Vein % per hole***
Pre-training	5,375	1.42	1.34
Post-training	5,479	0.30	0.27

Table 1. Breakdown of model performance pre- and post-training comparing the row count of each model where veining was predicted (i.e. >0.00cm²), the average vein percentage predicted against each row, and the average total vein percentage predicted across the 14 MSDP holes.

The dominant drilling phase that influenced the overall statistics of the model pre-training was Phase 3 where there were 120% more rows being counted as containing some veining (2,553 vs 2,117) and almost 10x the amount of total vein material being predicted on average (2.87% per hole vs 0.30% per hole) when compared to the model post-training. This over predicting of vein material can be seen in Figure 5 where MSDP09 has intersected the Sleaford Gneiss but then with the new training, the predictions throughout the Sleaford Gneiss decrease in volume and overall, the model accuracy and precision increases.

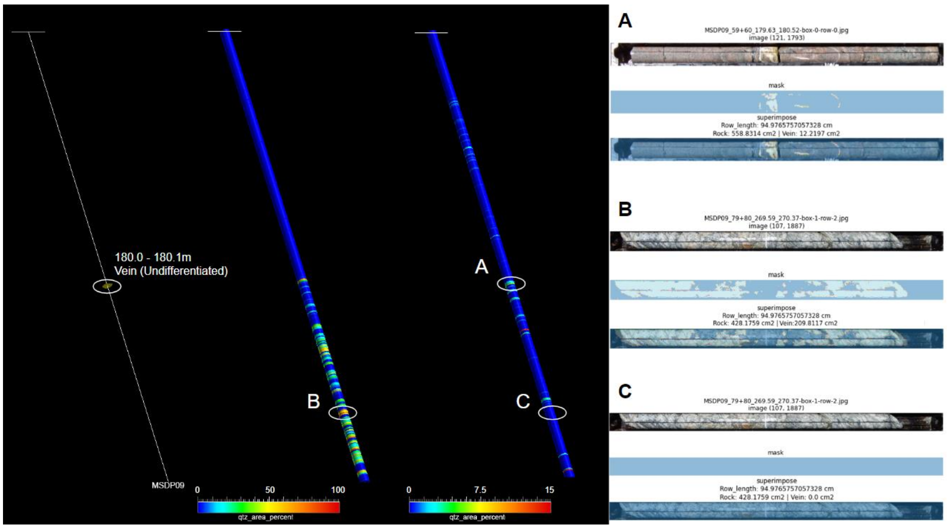

Figure 5. Impact of retraining the model with simple labelling. (a) Logged veining at 180m remains in the post-training model and is more precise upon visual inspection; (b) Main confusion in vein model pre-training throughout Phase 3 drilling was with the Sleaford Gneiss; and (c) This is not the case throughout the same section of drill hole post-training. Image is looking north.

Overall, the re-training of the general vein segmentation model was needed to improve its accuracy for this application, The general vein segmentation model was particularly confused by the gneissic lithologies logged throughout Phase 3 drill holes.

Vein Segmentation Model Evaluation and Confusion

Now that the newly trained vein segmentation model's accuracy and precision has improved, its performance with respect to a manually labelled dataset was evaluated. The vein segmentation model was evaluated based on the area of vein detected versus a manually labelled polygon by trained geologist.

Evaluation metrics were calculated using a validation set of data that the model was not trained on. This data is selected from images that are sampled to represent the textural variation in the MSDP dataset. The distribution of the difference of length between ground truth and inferred results as well as the distribution of difference between Labelled and Modelled length of vein per depth-registered row for the validation set can be seen in Figure 6, leading to a R-squared value of 0.778 and a gradient of best fit of 0.932.

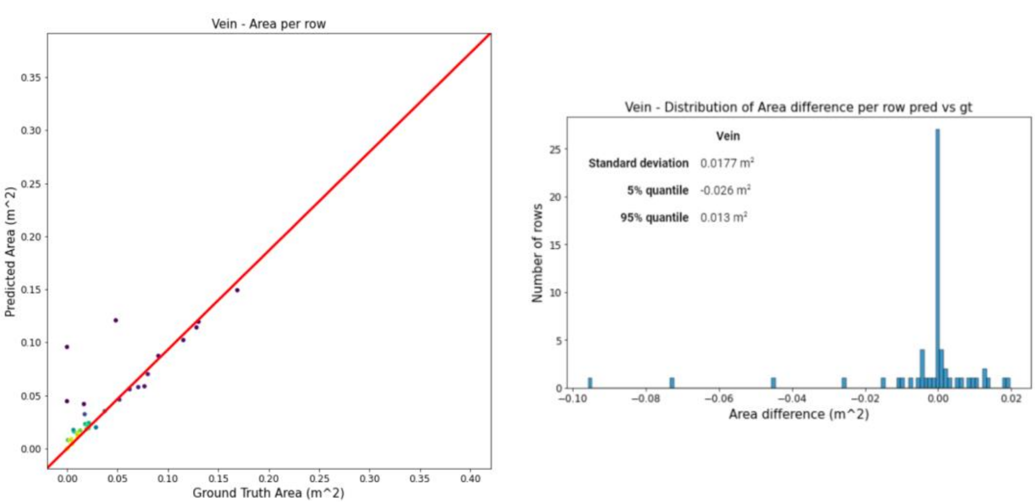

Figure 6. The distribution of difference between Labelled and Modelled lengths of vein per depth-registered row for the validation set (left - warmer colours relate to higher point density and the red line is the line of best fit) and distribution of the difference of length between ground truth and inferred results (right).

Based on these evaluation plots it is clear there is still model confusion after re-training. Model confusion can occur either as a false negative or a false positive. False positives occur when the model predicts a vein where none exist, and a false negative occurs where the model does not detect a vein. Examples of these incorrect predictions are shown in Figure 7 illustrating both a false positive (Insert A - veining predicted across a white core block) and a false negative (Insert B - majority of vein missed by model) within MSDP01. The figure also highlights a correct prediction when compared to the Vein (Undifferentiated) lithology logged between 763.9m and 771.0m.

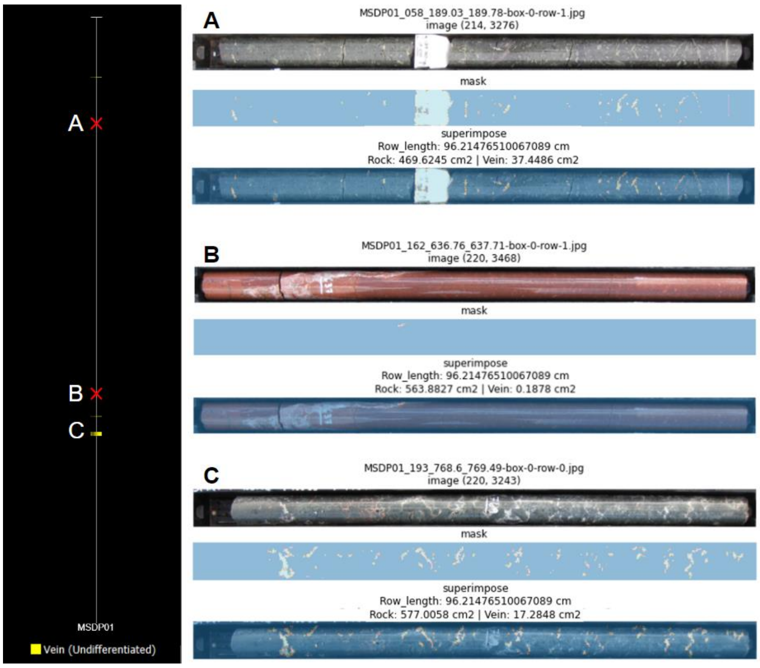

Figure 7. False predictions observed in MSDP01. (a) False positive with the model being confused with a white core block; (b) False negative with the model not fully predicting the vein at 637m; and (c) correct prediction compared to logging of Vein (Undifferentiated) between 763.9m - 771.0m.

Comparison to logged datasets

The comparison with the MSDP logged geology will be split based on the phases of drilling as there are distinct lithological differences between each phase which influence veining volume. It should be noted most of the logged veining within the manual logs have been added to the comments/description section instead of as major or minor lithology or a separate vein log. A breakdown of these logging statistics can be found in Table 2 where veining is rarely logged as a lithological unit, and thus able to be used for future modelling.

Phase	*Veins logged as Major Lithology (Count)*	*Veins logged as Major Lithology (Sum % of Core)*	*Veins logged as Minor Lithology (Count)*	*Veins logged as Minor Lithology (Sum % of Core)*	*Lithology Units with Veins noted in Description (Count)*	*Lithology Units with Veins noted in Description (Sum % of Core)*	*Dominant Host Rock/s*
1	9	0.37	91	24.9	157	73.5	Basalt, Shale, Rhyolite
2	1	0.01	0	0.00	27	47.8	Dacite
3	3	0.04	7	0.32	120	35.0	Gneiss, Dolomite, Rhyolite

Table 2. Count, Sum % of Core and Dominant Host Rocks across all three phases of MSDP drilling when veins have been manually logged as either major or minor lithologies or not logged but instead added to the description for a particular lithology interval.

Phase 1

Phase 1 drilling includes the largest number of veins manually logged as either a major or minor lithology with the majority being basalt hosted. The vein segmentation model correlates well with the manual logging within the dominant basalt host rock (Figure 8). The key differentiation between the two datasets is the quantitative linear data the vein segmentation model provides as a more robust way to visualise and model how veining is distributed throughout the holes. The use of vein percentage also illustrates how veining is dispersed across less dominant lithologies such as sandstone (MSDP02) and dolerite (MSDP04). This provides a greater insight into the distribution of potential fluid pathways as opposed to the point data originally logged from the core.

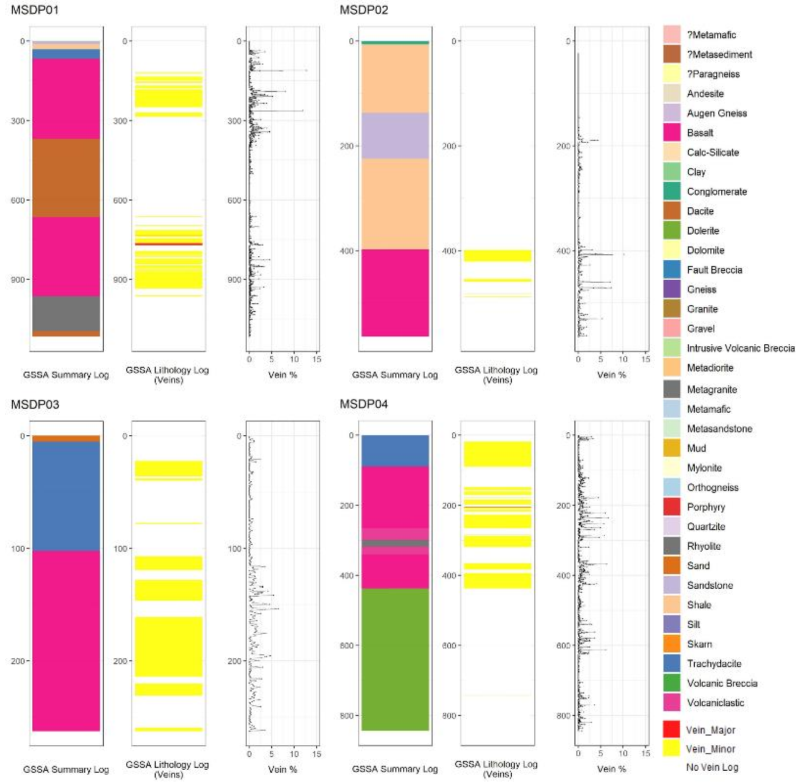

Figure 8. Phase 1 comparison between manual summary and vein logging from GSSA, and vein percentage from the vein segmentation model.

Phase 2

Phase 2 drilling contains the least amount of manually logged veining and yet, the vein segmentation model illustrates that all three holes from this phase of drilling contain some degree of veining (Figure 9). Much like Phase 1, the vein segmentation model correlates well with the dominant host rock (dacite in this case) which again is only query able via the description section of the GSSA interval logs.

MSDP05 is a good example of the usefulness of this vein percentage dataset as it highlights how veining is distributed differently between the three main dacite sections. The bulk of the veining is hosted within the second dacite intersected between 136.38m - 183.43m whilst the data also captures a vein-poor core within the largest and deepest intersected dacite between 300m - 400m. Information which is lost when only referring to the manual logging.

The volume and distribution of veining throughout the rhyolite in MSDP07 is also lost when referring to the manual logging, including the fact that the highest percentage of veining occurs beneath the contact with the volcaniclastics at 231.59m and again ~300m.

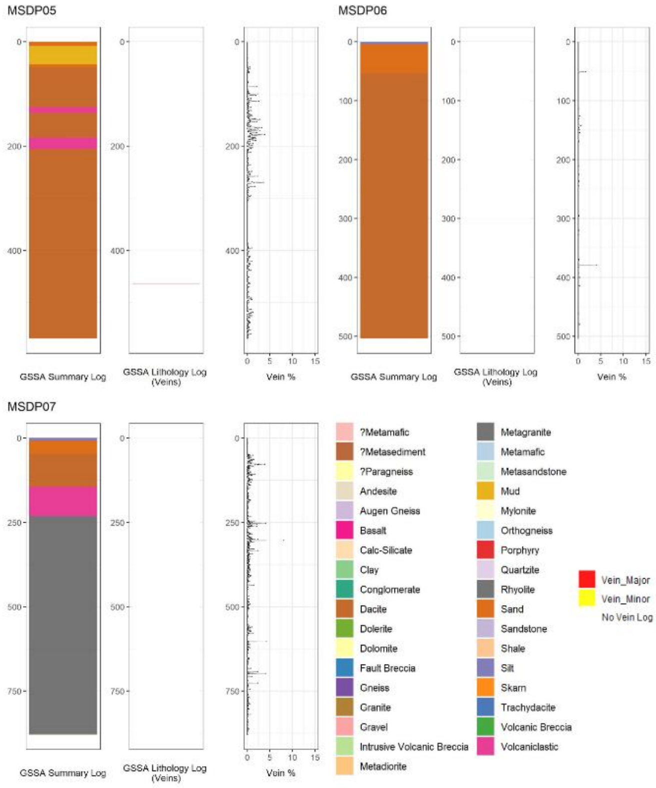

Figure 9. Phase 2 comparison between manual summary and vein logging from GSSA, and vein percentage from the vein segmentation model.

Phase 3

Phase 3 offers the most diverse range of host rocks (Figure 10), however, the volume of gneiss in MSDP11 and MSDP14 is a good indication of why the general vein segmentation model was initially confused throughout these holes. The re-training, whilst still predicting veining throughout the gneiss, now more accurately reflects the manual logging where, especially in MSDP14, the gneiss can be vein-poor.

Throughout these holes is that the highest modelled vein percentage doesn't often correlate with the major and to a lesser extent, minor logged undifferentiated veining by GSSA geologists. Other than the veining logged at ~400m in MSDP11, which under-predicted veining dominated by chlorite and epidote alteration, these log intervals over represent the veining present compared to similar veining throughout the holes. The vein segmentation model removes this bias as it reports an auditable and consistent dataset which more accurately reflects the nature and distribution of veining compared to the manual logging.

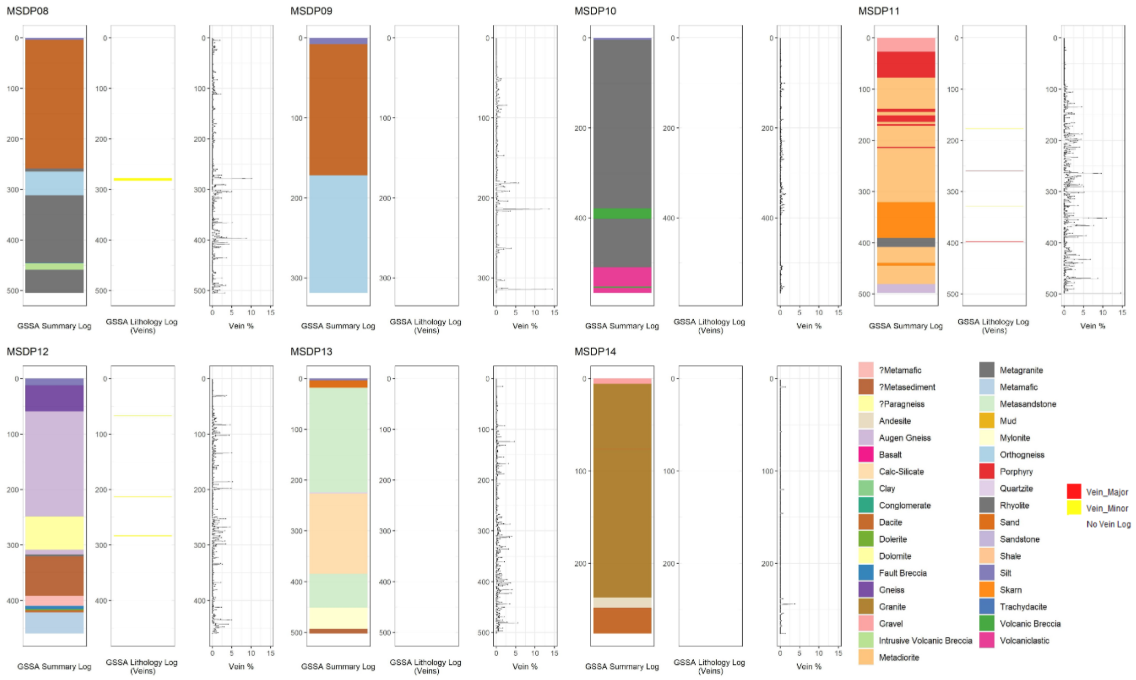

Figure 10. Phase 3 comparison between manual summary and vein logging from GSSA, and vein percentage from the vein segmentation model.

Discussion

These results demonstrate that the vein area and percentage datasets generated from the vein segmentation model compare well against, and in some places exceeded, the traditionally logged vein data from the MSDP holes. Continual development and re-training of this segmentation model would result in further refinement of results.

The data generated from this automated method can be inserted into existing geological logging workflows to add additional value to traditional logging during the discovery or grade control phases of a drill program. The key benefits of this automated vein segmentation workflow include:

- Consistency: drill program-wide datasets can be developed with exact repeatability and not be impacted by logging bias (experience of personnel, time of day, day in roster etc) or information which was not identified as being needed to be logged at the beginning of a campaign.
- Auditability: all predictions can be visualised on the imagery, meaning results can be reviewed if required. This is also the case for the training data which is not only visible but also editable with either updates to the training labels or addition of new labels.
- Speed: not only with the training process (a few hours) but once this training is complete, data can be generated within a matter of minutes and hours pending scale of drilling.
- Scale: vein volume would not be practical to collect with a human logger given the very detailed nature of the measurement.

CONCLUSIONS

We have presented an automated method to detect, classify and extract vein information from drill core photography using state of the art computer vision algorithms. The proposed vein segmentation method was successfully applied on core photography from 14 Mineral Systems Drilling Program holes in South Australia and the data produced is not only consistent with, but also more suitable for geological modelling, when compared to the traditionally logged datasets. This study illustrates the significant benefits of computer vision-based core logging workflows can provide compared to traditional logging with respect to consistency, auditability, and speed.

ACKNOWLEDGMENTS

The authors would like to recognise the efforts of Datarock geologists (Morgan O'Neill, Caitlin Brown, and Heather Watson) who collected, formatted, and labelled the original core photos and logging data from the GSSA. Adrian Fabris from the GSSA is thanked for providing additional datasets associated with the MSDP drilling program. We also acknowledge Datarock Pty Ltd for providing the computing resources required to perform the described computer vision methods.

REFERENCES

Allen, S.R., McPhie, J., Ferris, G. and Simpson, C., 2008, Evolution and architecture of a large felsic Igneous Province in western Laurentia: The 1.6 Ga Gawler Range Volcanics, South Australia, Journal of Volcanology and Geothermal Research, 172, 132-147.

Allen, S.R., Simpson, C., McPhie, J. and Daly, S., 2003, Stratigraphy, distribution and geochemistry of widespread volcanic units in the Mesoproterozoic Gawler Range Volcanics, South Australia, Australian Journal of Earth Sciences, 50, 97-112.

Ashraf, U., Zhang, H., Anees, A., Mangi, H., Ali, M., Zhang, X., Imraz, M., Abbasi, S.S., Abbas, A., Ullah, Z., & Ullah, J., and Tan, S., 2021, A Core Logging, Machine Learning and Geostatistical Modeling Interactive Approach for Subsurface Imaging of Lenticular Geobodies in a Clastic Depositional System, SE Pakistan, Natural Resources Research 2021, 10.

Caté, A., and Cevik, I., 2021, Automated Logging Using Core Photographs: Recent Developments and Implications for Core Logging. Available at https://www.srk.com/en/publications. Access October 14, 2022 SRK Consulting.

Fabris, A.J., Tylkowski, L., Brennan, J., Flint, R.B., Ogilvie, A., McAvaney, S., Werner, M., Pawley, M., Krapf, C., Burtt, A.C., Rowe, R., Henschke, C., Chalmers, N.C., Rechner, S., Hardwick, I., and Keeling, J., 2017, Mineral Systems Drilling Program in the southern Gawler Ranges, South Australia, Report Book 2016/00030, Department of the Premier and Cabinet, South Australia, Adelaide.

Fowler, M.J., 2013, Structural data bias in the digital age, Slope Stability 2013.

Harraden, C., Berry, R.F., and Lett, J., 2016, Proposed methodology for using automated core logging technology to extract geotechnical index parameters. In: Proceedings of 3rd international geometallurgy conference.

Lemy. F., Hadjigeorgiou, J., Côté, P., and Maldague, X., 2001, Image analysis of drill core, Min Technol, 110, 172–177.

Olson, L., Samson, C., and McKinnon, S.D., 2015, 3-D laser imaging of drill core for fracture detection and rock quality designation, Int J Rock Mech Min Sci, 73, 156–164.

Saricam, T., and Ozturk, H., 2018, Estimation of RQD by digital image analysis using a shadow-based method, Int J Rock Mech Min Sci, 112, 253–265.