UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

PETROGRAPHIC ANALYSIS WITH DEEP CONVOLUTIONAL NEURAL NETWORKS

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

RAFAEL AUGUSTO PIRES DE LIMA
Norman, Oklahoma
2019

PETROGRAPHIC ANALYSIS WITH DEEP CONVOLUTIONAL NEURAL NETWORKS

A THESIS APPROVED FOR THE
GALLOGLY COLLEGE OF ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Charles Nicholson, Chair

Dr. Randa Shehab

Dr. Roger Slatt

Table of Contents

List of Tables

List of Figures

Preface and acknowledgements

During my years as a graduate student at the University of Oklahoma, I had the opportunity to work with specialists in different fields of geoscience as well as data science. As I was eager to learn how to integrate tools and knowledge from both fields, I concurrently worked towards a PhD in Geophysics as well as a Masters in Data Science and Analytics. Due to the multidisciplinarity of my work, I had the opportunity to work with the Los Alamos National Laboratory for most of 2019, where I conducted research on various applications of machine learning to geoscience problems. This manuscript, presented as a thesis for the Masters in Data Science and Analytics, shows two of the studies I conducted at the University of Oklahoma incorporating elements of geoscience and data science. In Chapter 1, I show a paper as published by The Sedimentary Record entitled "Deep convolutional neural networks as a geological image classification tool" (Pires de Lima et al., 2019a[1]). Chapter 1 serves as an introduction to the use of convolutional neural networks, an important tool used for computer vision tasks, which can aid on geoscience problems. Then, in Chapter 2, I delve deeper into the classification of microfacies using thin-section images, showing both the highlights and drawbacks of the methodology we applied. Chapter 2, that names this thesis as "Petrographic analysis with deep convolutional neural networks" (Pires de Lima et al., 2019b[2]), is currently under review for publication. As the chapters are presented as published/in review, and because the manuscripts were written in collaboration, I maintain plural verbs and subjects thorough most of this thesis.

[1] Pires de Lima, R., Bonar, A., Coronado, D.D., Marfurt, K., Nicholson, C., 2019a. Deep convolutional neural networks as a geological image classification tool. Sediment. Rec. 17, 4–9. https://doi.org/10.210/sedred.2019.2

[2] Pires de Lima, R., Duarte, D., Nicholson, C., Slatt, R., Marfurt, K., 2019b. Petrographic analysis with deep convolutional neural networks. *In Review*.

As mentioned, the work I present here was conducted in collaboration with colleagues and experts in different fields. I would like to acknowledge my friends David Duarte and Alicia Bonar, as well as my geophysics PhD advisor Dr. Kurt Marfurt, for their help in the execution of the projects presented in this thesis. I would also like to acknowledge the help and support of Dr. Charles Nicholson (advisor), Dr. Randa Shehab, and Dr. Roger Slatt, respectively, chair and committee members of my Master in Data Science and Analytics, for their help with the study presented here.

Chapter 1: Deep convolutional neural networks as a geological image classification tool

Rafael Pires de Lima[1,2], Alicia Bonar[1], David Duarte Coronado[1], Kurt Marfurt[1], Charles Nicholson[3]

[1]School of Geology and Geophysics, The University of Oklahoma, 100 East Boyd Street, RM 710, Norman, Oklahoma, 73019, USA

[2]The Geological Survey of Brazil – CPRM, 55 Rua Costa, São Paulo, São Paulo, Brazil

[3]School of Industrial and Systems Engineering, The University of Oklahoma, 202 West Boyd Street, RM 124, Norman, Oklahoma, 73019, USA

Abstract

A convolutional neural network (CNN) is a deep learning (DL) method that has been widely and successfully applied to computer vision tasks including object localization, detection, and image classification. DL for supervised learning tasks is a method that uses the raw data to determine the classification features, in contrast to other machine learning (ML) techniques that require pre-selection of the input features (or attributes). In the geosciences, we hypothesize that deep learning will facilitate the analysis of uninterpreted images that have been neglected due to a limited number of experts, such as fossil images, slabbed cores, or petrographic thin sections. We use transfer learning, which employs previously trained models to shorten the development time for subsequent models, to address a suite of geologic interpretation tasks that may benefit from ML. Using two different base models, MobileNet V2 and Inception V3, we illustrate the successful classification of microfossils, core images, petrographic photomicrographs, and rock and mineral hand sample images. ML does not replace the expert geoscientist. The expert defines the labels (interpretations) needed to train the algorithm and also monitors the results to address

incorrect or ambiguous classifications. ML techniques provide a means to apply the expertise of skilled geoscientists to much larger volumes of data

Authorship statement: RPL developed the conception and design of study, wrote the necessary scripts, performed analysis, and wrote the manuscript. DD acquired the data, participated in the analysis, and helped write the manuscript. CN guided the analysis, helped writing the manuscript, and revised the manuscript critically for important intellectual content. RS helped writing the manuscript, and revised the manuscript critically for important intellectual content. KJM helped in the conception and design of study, helped write the manuscript, and revised the manuscript critically for important intellectual content.

Introduction

Machine learning (ML) techniques have been successfully applied, with considerable success, in the geosciences for almost two decades. Applications of ML by the geoscientific community include many examples such as seismic-facies classification (Meldahl et al., 2001; West et al., 2002; de Matos et al., 2011; Roy et al., 2014; Qi et al., 2016; Hu et al., 2017; Zhao et al., 2017), electrofacies classification (Allen and Pranter, 2016), and analysis of seismicity (Kortström et al., 2016; DeVries et al., 2018; Perol et al., 2018; Sinha et al., 2018), and classification of volcanic ash (Shoji et al., 2018), among others. Conventionally, ML applications rely on a set of attributes (or features) selected or designed by an expert. Features are specific characteristics of an object that can be used to study patterns or predict outcomes. In classification modeling, these features are chosen with the goal of distinguishing one object from another.

Typically, feature selection is problem dependent. For example, a clastic sedimentary rock is most broadly classified by its grain size; therefore, a general classification for a rock sample (data) is sandstone if its grain sizes (features) lie from 0.06 mm to 2.0 mm following the Wentworth size class. In this example, a single feature is used to classify the sample, but more complex and/or detailed classification often requires analysis of multiple features exhibited by the sample. An inefficiency of traditional ML approaches is that many features may be constructed while only a subset of them are actually needed for the classification.

The use of explicitly designed features to classify data was the traditional approach in ML applications within the geosciences as in many other research areas. This classification approach works well when human interpreters know and can quantify the features that distinguish one object from another. However, sometimes an interpreter will subconsciously

classify features and have difficulty describing what the distinguishing features might be, relying on “I’ll know what the object is when I see it”. In contrast to feature-driven ML classification algorithms, deep learning (DL) models extract information directly from the raw unstructured data rather than the data being manually transformed.

Because of their greater complexity (and resulting flexibility and power) convolutional neural networks (CNN) usually requires more training data than traditional ML processes. However, when expert-labeled data are provided, non-experts can use the CNN models to generate highly accurate results (e.g. TGS Salt Identification Challenge | Kaggle, 2019).

DL applications in the geosciences require experts to first define the labels used to construct the necessary data sets as well as identify and address any ambiguous results and anomalies. In order to bring awareness and provide basic information regarding CNN models, DL techniques, and the necessity of expert-level knowledge needed to utilize these advancements, we applied these methods to four different geologic tasks. Figure 1 shows samples of different types of data that can be interpreted and labeled by experienced geologists. We use such interpretations to train our models. In this manuscript, we show how CNN can aid geoscientists with microfossil identification, core descriptions, petrographic analyses, and as a potential tool for education and outreach by creating a simple hand specimen identification application.

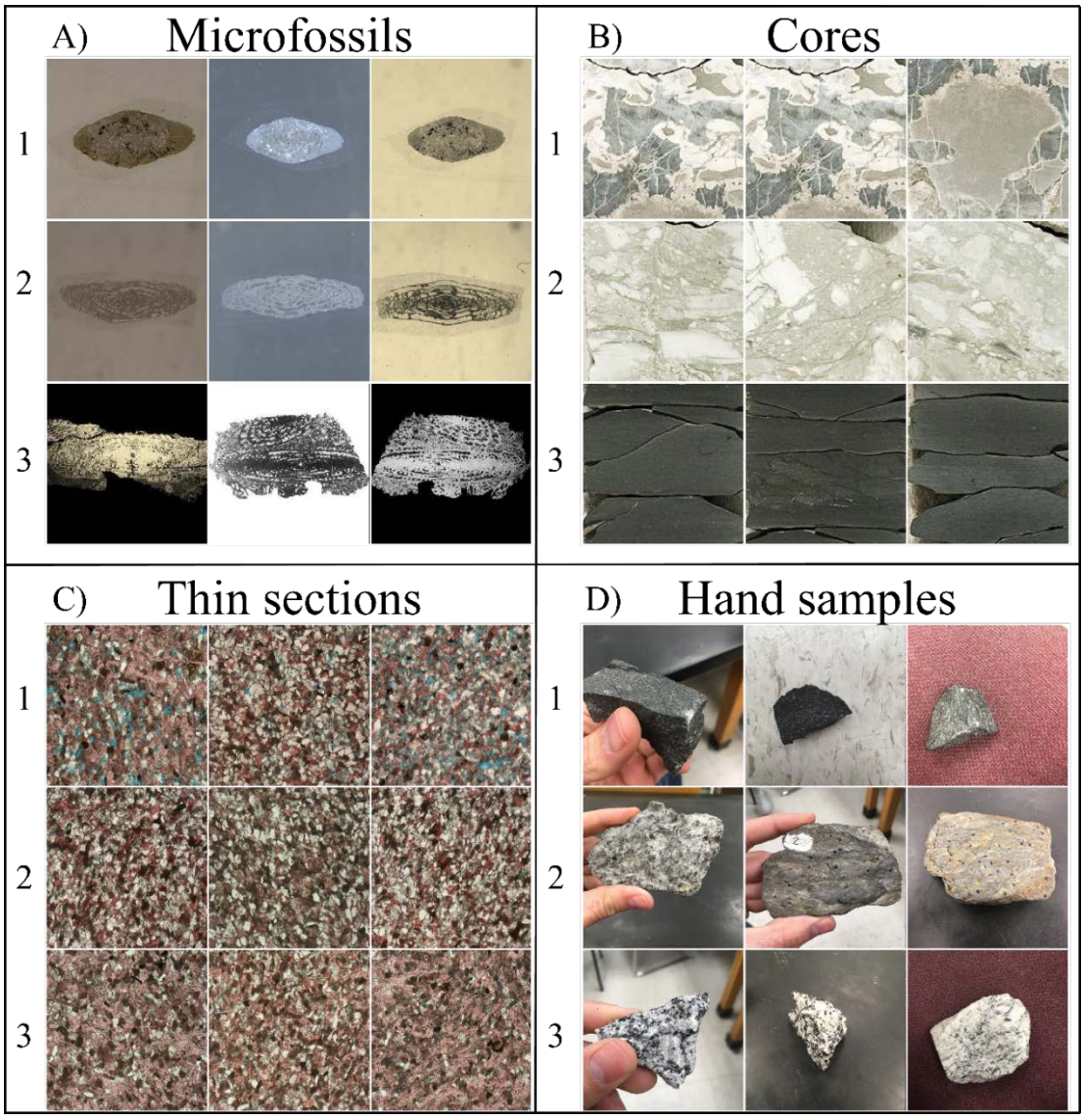

Figure 1: Examples of the data used in this study. A) Three of the seven Fusulinids groups (*Beedeina* (1), *Fusulinella* (2), and *Parafusulina* (3)). B) Three of the five lithofacies (bioturbated mudstone-wackestone (1), chert breccia (2), and shale (3)). C) Reservoir quality classes (high (1), intermediate (2), and low (3)) D) Three of the six rock sample groups (basalt (1), garnet schist (2), and granite (3)). Samples were interpreted by professionals working with each separate dataset.

Convolutional neural networks and transfer learning

Recent CNN research has yielded significant improvements and unprecedented accuracy (the ratio between correct classifications and the total number of samples classified) in image classification and are recognized as leading methods for large-scale visual recognition problems, such as the annual ImageNet Large Scale Visual Recognition Challenge (ILSVRC, Russakovsky

et al. (2015)). Specific CNN architectures have been the leading approach for several years now (e.g., Szegedy et al., 2014; Chollet, 2016; He et al., 2016; Huang et al., 2016; Sandler et al., 2018). Researchers noted that the parameters learned by the layers in many CNN models trained on images exhibit a common behavior – layers closer to the input data tend to learn general features, such as edge detecting/enhancing filters or color blobs, then there is a transition to more specific dataset features, such as faces, feathers, or object parts (Yosinski et al., 2014; Yin et al., 2017). These general-specific CNN layer properties are important points to be considered for the implementation of transfer learning (Caruana, 1995; Bengio, 2012; Yosinski et al., 2014). In transfer learning, first a CNN model is trained on a base dataset for a specific task. The learned features (model parameters) are repurposed, or transferred, to a second target CNN to be trained on a different dataset and task (Yosinski et al., 2014).

New DL applications often require large volumes of data, however the combination of CNNs and transfer learning allows the reuse of existing DL models to novel classification problems with limited data, as has been demonstrated in diverse fields, such as botany (Carranza-Rojas et al., 2017), cancer classification (Esteva et al., 2017), and aircraft detection (Chen et al., 2018). Analyzing medical image data, Tajbakhsh et al. (2016) and Qayyum et al. (2017) found that transfer learning achieved comparable or better results than training a CNN model with randomly initialized parameters. As an example, training the entire InceptionV3 (Szegedy et al., 2015) with 1000 images (five classes, 50 original images for each class, four copies of each original image) with randomly initialized parameters can be 10 times slower than the transfer learning process (11 minutes vs 1 minute on average for five executions) using a Nvidia Quadro M2000 (768 CUDA Cores). On a CPU (3.60 GHz clock speed), training the entire model can take up to 2 hours whereas transfer learning can be completed within a few minutes. We also

noticed that transfer learning is easier to train. During the speed comparison test, transfer learning achieved high accuracies (close to 1.0) within 5 epochs (note the dataset is very simple with most of the samples being copies of each other). Successful applications of computer vision technologies in different fields suggest that ML models could be extremely beneficial for geologic applications, especially those in the category of image classification problems.

For the examples we present in this paper (Figure 1), we rely on the use of transfer learning (Yosinski et al., 2014) using the MobileNetV2 (Sandler et al., 2018) and InceptionV3 as our base CNN models. Both MobileNetV2 and InceptionV3 were trained on ILSVRC. Therefore, the CNN models we used were constructed based on inputs of 3-channels (RGB) of 2D photographic images. We randomly select part of the data to be used as a test set maintaining the same proportion of samples per class as in the training set. The data in the test set is not used during the computational process for model training; rather, it is used to evaluate the quality and robustness of the final model. Due to limited space, we refrained showing the CNN mistakes and many of the steps necessary for data preparation.

CNN-Assisted fossil analysis

Biostratigraphy has become a less common focus of study in the discipline of paleontology (Farley and Armentrout, 2000, 2002), but the applications of biostratigraphy are necessary for understanding age-constraints for rocks that cannot be radiometrically dated. Access to a specific taxonomic expert to accurately analyze fossils at the species-level can be as challenging as data acquisition and preparation. Using labeled data from the University of Oklahoma Sam Noble Museum and iDigBio portal, we found that Fusulinids (index fossils for the Late Paleozoic) can be accurately classified with the use of transfer learning. Accurate

identification of a Fusulinid depends on characteristics that must be observed and exposed along the long axis of the (prolate spheroid-shaped) Fusulinid. We used a dataset of 1850 qualified images including seven different Fusulinid genera. After retraining the CNN model, we obtained an accuracy for the test set (10% of the data) of 1.0 for both retrained MobileNetV2 and InceptionV3 (Table 1). Figure 2 shows a schematic view of the classification process.

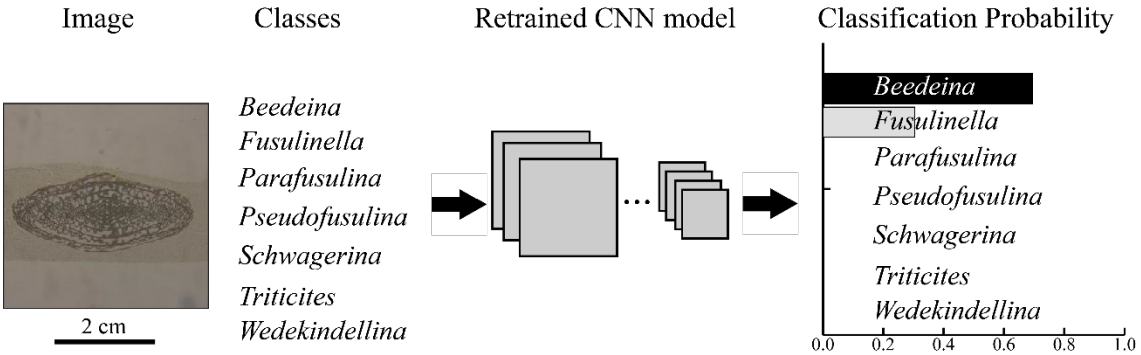

Figure 2: An example of the classification process. In this example, a thin-section image that should fit one of the seven Fusulinid genera is analyzed by the model. The model outputs the probability assigned to each of the possible classes (all probabilities summing to 1.0). The term "classes" here is used in the ML sense rather than the biological one. In the example provided, our model provided a high probability for the same class as the human expert. Note that in the implementation we use the model will classify any image as one of the seven learned classes – even if the image is clearly not a fossil. This highlights the importance of a domain expert intervention.

Table 1: Summary of test accuracy for the examples in this study.

Dataset	**Number of training samples**	**Number of test samples**	**Number of output classes**	**MobileNetV2 Accuracy**	**InceptionV3 Accuracy**
Microfossils (Fusulinids)	1480	184	7	1.00	1.00
Core	227	28	5	1.00	0.97
Petrographic thin-sections	194	31	3	0.81	0.81
Rock samples	1218	151	6	0.98	0.97

CNN-Assisted core description

Miles of drilled cores are stored in boxes in enormous warehouses, many of which have either been neglected for years or never digitally described. Core-based rock-type descriptions are important for understanding the lithology and structure of subsurface geology. Using several hundred feet of labeled core from a Mississippian limestone in Oklahoma (data from Suriamin and Pranter, 2018 and Pires de Lima et al., 2019), we selected a small sample of 285 images from five distinct lithofacies to be classified by the retrained CNN models. Pires de Lima et al. (2019) describes how a sliding window is used to generate CNN input data, cropping small sections from a standard core image. We used 10% of the data as the test set and achieved an accuracy of 1.0 using the retrained MobileNetV2 and an accuracy of 0.97 using the retrained InceptionV3 (Table 1).

CNN-Assisted reservoir quality classification using petrographic thin sections

Petrography focuses on the microscopic description and classification of rocks and is one of the most important techniques in sedimentary and diagenetic studies. Potential information gained from thin section analysis compared to hand specimen descriptions include mineral distribution and percentage, pore space analysis, and cement composition. Petrographic analyses can be laborious even for experienced geologists. Using a total of 161 photomicrographs of parallel Nicol polarization of thin sections from the Sycamore Formation shale resource play in Oklahoma, we classified these images as representatives of high, intermediate, and low reservoir quality depending on the percent of calcite cement and pore space. We used 20% of the images in the test set and obtained a test set accuracy of 0.81 for both the retrained MobileNetV2 and the retrained InceptionV3 (Table 1).

CNN-Assisted rock sample analysis

By creating a simple website, the general population could have immediate access to a rock identification tool using transfer learning technology. For this work in progress, we used smartphones to acquire 1521 pictures of six different rock types, using five different hand samples for each one of the rock types. We took pictures with different backgrounds, as visually depicted in Figure 1, however all pictures were taken in the same classroom. After retraining the CNN models, we obtained an accuracy for the test set (10% of original data) of 0.98 using the retrained MobileNetV2 and 0.97 using the retrained InceptionV3 (Table 1). We note that our model does not perform well with no-background images (i.e., pictures in which the rock sample is edited and seems to be within a white or black canvas) as such images were not used in training.

Conclusions and future work

Although gaining popularity and becoming established as robust technologies in other scientific fields, transfer learning and CNN models are still novel with respect to application within the geoscience community. In this paper, we used CNN and transfer learning to address four potential applications that could improve data management, organization, and interpretation in different segments of our community. We predict that the versatile transfer learning and deep learning technologies will play a role in public education and community outreach, allowing the public to identify rock samples much as they currently can use smart phone apps to identify visitors to their bird feeder. Such public engagement will increase geological awareness and provide learning opportunities for elementary schools, outdoor organizations, and families.

For all of our examples, we were able to achieve high levels of accuracy (greater than 0.81) by repurposing two different CNN models originally assembled for generic computer vision tasks. We note that the examples and applications demonstrated here are curated, and therefore we expected highly accurate results. We presented demonstrations with limited classes and relatively well-controlled input images, so near perfect accuracies cannot necessarily be expected in an open, free-range deployment scenario. Regardless, the ability to create distinctive models for specific sets of images allows for a versatile application.

The techniques we have shown could greatly improve the speed of monotonous tasks such as describing miles of core data with very similar characteristics or looking at hundreds of thin sections from the same geologic formation. While the tasks are performed by the computer, the geoscience expert is still the most important element in every analysis in order to create the necessary datasets and provide quality control of the generated results. In the end, the expert validates the correctness of the results and looks for anomalies that are poorly represented by the target classes. We believe ML can help maintain consistency in interpretations and even provide a resource for less common observations and data variations, such as previously overlooked fossil subspecies and unique mineralogical assemblages in small communities and private collections, thereby building and reconciling a more complete international database. By combing expert knowledge and time efficient technology, ML methods can accelerate many data analysis processes for geologic research.

Acknowledgements

We thank the iDigBio initiative for providing access to the community for biodiversity collections data. Rafael acknowledges CNPq (grant 203589/2014-9) for the financial support and CPRM for granting the leave of absence allowing the pursuit of his Ph.D. studies. We thank Roger J. Burkhalter from the University of Oklahoma Sam Noble Museum of Natural History for providing the Fusulinids images used in this manuscript.

References

Allen, D.B., Pranter, M.J., 2016. Geologically constrained electrofacies classification of fluvial deposits: An example from the Cretaceous Mesaverde Group, Uinta and Piceance Basins. Am. Assoc. Pet. Geol. Bull. 100, 1775–1801. https://doi.org/10.1306/05131614229

Bengio, Y., 2012. Deep Learning of Representations for Unsupervised and Transfer Learning, in: Guyon, I., Dror, G., Lemaire, V., Taylor, G., Silver, D. (Eds.), Proceedings of ICML Workshop on Unsupervised and Transfer Learning, Proceedings of Machine Learning Research. PMLR, Bellevue, Washington, USA, pp. 17–36.

Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., Joly, A., 2017. Going deeper in the automated identification of Herbarium specimens. BMC Evol. Biol. 17, 181. https://doi.org/10.1186/s12862-017-1014-z

Caruana, R., 1995. Learning Many Related Tasks at the Same Time with Backpropagation, in: Tesauro, G., Touretzky, D.S., Leen, T.K. (Eds.), Advances in Neural Information Processing Systems 7. MIT Press, pp. 657–664.

Chen, Z., Zhang, T., Ouyang, C., Chen, Z., Zhang, T., Ouyang, C., 2018. End-to-End Airplane Detection Using Transfer Learning in Remote Sensing Images. Remote Sens. 10, 139. https://doi.org/10.3390/rs10010139

Chollet, F., 2016. Xception: Deep Learning with Depthwise Separable Convolutions. CoRR abs/1610.0.

de Matos, M.C., Yenugu, M. (Moe), Angelo, S.M., Marfurt, K.J., 2011. Integrated seismic texture segmentation and cluster analysis applied to channel delineation and chert reservoir characterization. Geophysics 76, P11–P21. https://doi.org/10.1190/geo2010-0150.1

DeVries, P.M.R., Viégas, F., Wattenberg, M., Meade, B.J., 2018. Deep learning of aftershock patterns following large earthquakes. Nature 560, 632–634. https://doi.org/10.1038/s41586-018-0438-y

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S., 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118. https://doi.org/10.1038/nature21056

Farley, M.B., Armentrout, J.M., 2002. Tools, Biostratigraphy becoming lost art in rush to find new exploration. Offshore 94–95.

Farley, M.B., Armentrout, J.M., 2000. Fossils in the Oil Patch. Geotimes 14–17.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90
Hu, S., Zhao, W., Xu, Z., Zeng, H., Fu, Q., Jiang, L., Shi, S., Wang, Z., Liu, W., 2017. Applying principal component analysis to seismic attributes for interpretation of evaporite facies: Lower Triassic Jialingjiang Formation, Sichuan Basin, China. Interpretation 5, T461–T475. https://doi.org/10.1190/INT-2017-0004.1
Huang, G., Liu, Z., Weinberger, K.Q., 2016. Densely Connected Convolutional Networks. CoRR abs/1608.0.
Kortström, J., Uski, M., Tiira, T., 2016. Automatic classification of seismic events within a regional seismograph network. Comput. Geosci. 87, 22–30. https://doi.org/10.1016/J.CAGEO.2015.11.006
Meldahl, P., Heggland, R., Bril, B., de Groot, P., 2001. Identifying faults and gas chimneys using multiattributes and neural networks. Lead. Edge 20, 474–482. https://doi.org/10.1190/1.1438976
Perol, T., Gharbi, M., Denolle, M., 2018. Convolutional neural network for earthquake detection and location. Sci. Adv. 4, e1700578. https://doi.org/10.1126/sciadv.1700578
Pires de Lima, R., Suriamin, F., Marfurt, K.J., Pranter, M.J., 2019. Convolutional neural networks as aid in core lithofacies classification. Interpretation 7, SF27–SF40. https://doi.org/10.1190/INT-2018-0245.1
Qayyum, A., Anwar, S.M., Awais, M., Majid, M., 2017. Medical image retrieval using deep convolutional neural network. Neurocomputing 266, 8–20. https://doi.org/10.1016/J.NEUCOM.2017.05.025
Qi, J., Lin, T., Zhao, T., Li, F., Marfurt, K., 2016. Semisupervised multiattribute seismic facies analysis. Interpretation 4, SB91–SB106. https://doi.org/10.1190/INT-2015-0098.1
Roy, A., Romero-Peláez, A.S., Kwiatkowski, T.J., Marfurt, K.J., 2014. Generative topographic mapping for seismic facies estimation of a carbonate wash, Veracruz Basin, southern Mexico. Interpretation 2, SA31–SA47. https://doi.org/10.1190/INT-2013-0077.1
Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y
Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. ArXiv e-prints.
Shoji, D., Noguchi, R., Otsuki, S., Hino, H., 2018. Classification of volcanic ash particles using a convolutional neural network and probability. Sci. Rep. 8, 8111. https://doi.org/10.1038/s41598-018-26200-2
Sinha, S., Wen, Y., Pires de Lima, R.A., Marfurt, K., 2018. Statistical controls on induced seismicity. Unconventional Resources Technology Conference. https://doi.org/10.15530/urtec-2018-2897507-MS
Suriamin, F., Pranter, M.J., 2018. Stratigraphic and lithofacies control on pore characteristics of Mississippian limestone and chert reservoirs of north-central Oklahoma. Interpretation 1–66. https://doi.org/10.1190/int-2017-0204.1
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2014. Going Deeper with Convolutions. CoRR abs/1409.4.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2015. Rethinking the Inception Architecture for Computer Vision. CoRR abs/1512.0.

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J., 2016. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? IEEE Trans. Med. Imaging 35, 1299–1312. https://doi.org/10.1109/TMI.2016.2535302

TGS Salt Identification Challenge | Kaggle [WWW Document], n.d. URL https://www.kaggle.com/c/tgs-salt-identification-challenge (accessed 1.10.19).

West, B.P., May, S.R., Eastwood, J.E., Rossen, C., 2002. Interactive seismic facies classification using textural attributes and neural networks. Lead. Edge 21, 1042–1049. https://doi.org/10.1190/1.1518444

Yin, X., Chen, W., Wu, X., Yue, H., 2017. Fine-tuning and visualization of convolutional neural networks, in: 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, pp. 1310–1315. https://doi.org/10.1109/ICIEA.2017.8283041

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? Adv. Neural Inf. Process. Syst. 27, 3320–3328.

Zhao, T., Li, F., Marfurt, K.J., 2017. Constraining self-organizing map facies analysis with stratigraphy: An approach to increase the credibility in automatic seismic facies classification. Interpretation 5, T163–T171. https://doi.org/10.1190/INT-2016-0132.1

Chapter 2: Deep convolutional neural networks as a geological image classification tool

Rafael Pires de Lima[1,2], David Duarte Coronado[1], Kurt Marfurt[1], Charles Nicholson[3], Roger Slatt[1], Kurt J. Marfurt[1]

[1]School of Geology and Geophysics, The University of Oklahoma, 100 East Boyd Street, RM 710, Norman, Oklahoma, 73019, USA

[2]The Geological Survey of Brazil – CPRM, 55 Rua Costa, São Paulo, São Paulo, Brazil

[3]School of Industrial and Systems Engineering, The University of Oklahoma, 202 West Boyd Street, RM 124, Norman, Oklahoma, 73019, USA

Abstract

Petrographic analysis is based on the microscopic description and classification of rocks and is a crucial technique for sedimentary and diagenetic studies. When compared to hand specimens, thin sections of rocks provide better and more accurate means for analysis of mineral distribution and percentage, pore space analysis, and cement composition. Because of the rich information they contain, thin section data are commonly used not only by the mining and petroleum industry, but by the academic community as well. Most petrographic analysis relies on visual inspection of rock thin sections under a microscope, a task that is laborious even for experienced geologists. Large projects with a tight time frame requiring the analysis of a large amount of thin sections may require multiple petrographers, thereby risking the introduction of inconsistency in the analysis. To address this challenge, we explore the use of deep convolutional neural networks (CNN) as a tool that can allow the petrographer to analyze and classify more samples in a consistent manner. Unlike previous studies using deep learning models trained on large volumes of thin section data, we make use of transfer learning based on robust and reliable

CNN models trained with a large amount of non-geological images. With a much smaller number of labeled thin sections used in training followed by “fine-tuning” we are able to construct convolutional neural networks that achieve low error levels (<5% when images of same quality are used for training and testing) in thin section classification. While becoming widely accepted as a useful tool in the biological and manufacturing disciplines, CNN is currently underutilized in the geoscience community; we foresee an increase of use of such techniques to help accelerate and quantify a wide variety of geological tasks.

Glossary

We provide a simple glossary with common denominations in machine learning applications and used throughout the manuscript. For a more comprehensive list we refer the reader to Google's machine learning glossary ("Machine Learning Glossary | Google Developers," accessed August 2019).

- **Accuracy**: the fraction of total objects correctly classified. Values range from 0.0 to 1.0 (equivalently, 0% to 100%). A perfect score of 1.0 means all classifications were correct whereas a score of 0.0 means all classifications were incorrect.
- **Convolution**: a mathematical operation that combines two functions producing an output. In machine learning applications, a convolutional layer uses two discrete functions, the input data and a convolutional kernel, to train the convolutional kernel weights.
- **Convolution Neural Networks (CNN)**: a neuron network architecture in which at least one layer is a convolutional layer.
- **Deep Learning (DL)**: an artificial neural network architecture that contains many hidden layers.
- **Fine Tuning**: the process of adjusting machine learning model parameters of a pre-trained model to improve performance for a specific problem type
- **Label:** the names applied to an instance, sample, or example (for image classification, an image) associating it with a given class.
- **Layer:** a group of neurons in a machine learning model that process a set of input features.

- **Machine Learning (ML)**: a collection of approaches in which systems improve their performance through automatic analysis of data.
- **Neural Networks (NN)**: a machine learning model that combines linear and nonlinear transformations, loosely inspired in the behavior of brain neurons. It is typically organized in layers where each layer contains a number of nodes (or neurons).
- **Neuron**: A node in a neural network, typically taking in multiple input values and generating one output value. The neuron calculates the output value by applying an activation function (nonlinear transformation) to a weighted sum of input values.
- **Training**: the process of finding the most appropriate weights of a machine learning model.
- **Transfer Learning**: a technique that uses information learned in a primary machine learning task to perform a secondary machine learning task.
- **Top-X error**: a measure of model accuracy. A classification is considered correct as long as the correct label is in one of the top X guessed labels. Top-1 error is the ratio of the incorrect classifications over the total number of classifications (1.0 minus accuracy).
- **Weights**: the coefficients of a machine learning model. In a simple linear equation, the slope and intercept are the weights of the model. In CNNs, the weights are the convolutional kernel values. The training objective is to find the ideal weights of the machine learning model.

Introduction

Petrography focuses on the microscopic description and classification of rocks and remains one of the most used techniques in geoscience studies. The essential tool in petrographic

analysis is the microscope that uses plane and polarized transmitted light to capture the optical properties of minerals. The geologist or petrographer uses such a microscope to examine a rock thin section, which is a flat rock sample usually 30 μm thick, mounted on a glass slide. The goal is to observe and describe the characteristics of the rock such as grain geometry, structure, mineralogy and texture.

One of the most important uses of petrographic studies is to define microfacies based on the aforementioned thin section characteristics. However, hundreds of thin sections need to be described when classifying microfacies and such description is a very time-consuming process. Although the point-count method provides accurate and arguably undisputable classification for a thin section, point counts are often discarded as a classification option as it is considered a draining task. In our experience, a qualified geologist can take up to 20 minutes to count 300 points (the average number of points necessary for classification) in a single thin section when the petrographer is familiar with the mineralogical composition of the rock. Due to the long time required for the analysis of a single sample, the mechanical thin section point-count is often replaced by an interpretative approach. A single thin section interpretation then can take less than a minute, in cases in which the petrographer is familiar with the microfacies, or up to tens of minutes, in cases in which the thin section presents elements that are unfamiliar to the petrographer. The interpretation process can be subjective, thereby running the risk of inconsistent labeling. Cheng et al. (2018) observed that new thin sections are continuously produced, growing the number of samples that needs to be analyzed and archived by the geoscience community. With the development of new geological concepts, the application of new interpretation frameworks, the acquisition of new acreage, or the integration of multiple collections analyzed by different work teams, vast amounts of data constantly need to be re-

interpreted. Our goal is to generate machine learning (ML) models with the ability to produce reliable results in significantly shorter times and to provide more quantitative decision-making required in the oil and gas industry to organize data to allow the evaluation of new geological concepts.

The microfacies description obtained through images thin sections are analogous to image classification problems. Datta et al. (2008) reported that image classification is one of the tasks in which machines have excelled, often obtaining faster and more accurate results than humans. Because ML models have been successful in a wide variety of image classification problems, we hypothesize that deep convolutional neural network (CNN) holds similar promise in the microfacies classification of thin section photographs.

We begin our paper with a brief review of recent advances in using CNN as image classification in other fields, as well as some of the limited CNN applications using rock thin section data. Next, we describe the thin section preparation and data. We then describe the processing and analysis performed on the data and summarize our results. We conclude our manuscript with a summary of the advantages and limitations of the technology.

A short review of image processing using machine learning

Customary ML methods are limited in their ability to process raw data (such as the pixel values of an image). Due to such limitations, for many years the construction of a pattern-recognition model demanded carefully detailed feature engineering (e.g. the analysis of the wings of an insect or the leaves of a tree) performed by domain experts (LeCun et al., 2015; Yin et al., 2017). Yang et al. (2018) observed that one of the reasons deep learning (DL) models attracted the attention of the research community is DL's capacity to discover an effective

feature transformation for a specific task. Current progress in DL models, specifically CNN architectures, have improved the state-of-the-art in visual object recognition and detection, speech recognition and many other fields of study (LeCun et al., 2015). The model described by Krizhevsky et al. (2012), frequently referenced to as AlexNet, is considered a breakthrough and influenced the rapid adoption of DL in the computer vision field (LeCun et al., 2015). A variant of AlexNet won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC, Russakovsky et al., 2015) in 2012 achieving a top-5 test error rate (how often a true label is not one of the top 5 labels assigned by the model, a common metric for the ILSVRC) of 15%. The second-best entry for ILSVRC in 2012 had a top-5 error rate of 26%. AlexNet, with only five convolutional layers, has 60 million parameters to be trained. At first glance, such a large number of parameters might seem like a drawback for the implementation of DL models. However, with the advances of graphics processing units (GPUs), the previously prohibitive long training time has been significantly reduced (Mou et al., 2017; Yang et al., 2018).

In 2012 AlexNet used a five-layer deep CNN model; today many models competing in the ILSVRC use twenty to hundreds of layers. Huang et al. (2016) has even proposed models with thousands of layers. Due to the vast number of operations performed in deep CNN models, it is often difficult to discuss the interpretability, or the degree to which a decision taken by a model can be rationalized. For this reason, many workers consider CNN to be a black box, with CNN interpretability itself a research topic (e. g. Simonyan et al., 2013; Olah et al., 2017, 2018; Yin et al., 2017).

Recent CNN developments include several model architectures that achieved top-5 error rates under 10% in the ILSVRC dataset (e.g. Szegedy et al., 2014; Chollet, 2016; He et al., 2016a; Huang, Liu, et al., 2016; Sandler et al., 2018). Yosinski et al. (2014) and Yin et al. (2017)

also reported that the parameters learned by the layers in many CNN models trained on images exhibit a very common behavior. The layers closer to the input data tend to learn general features, such as edge detection/enhancement filters or color blobs. Then there is a transition to more specific dataset features, such as faces, feathers, or object parts. These general-specific CNN layers feature properties led to the development of transfer learning (e.g. Caruana, 1995; Bengio, 2012; Yosinski et al., 2014).

In transfer learning, first a CNN model is trained on a dataset for a primary task using large amounts of data. After training, the weights of the model are then repurposed or transferred to a second CNN that can be trained using a smaller dataset, generally domain-specific, for a secondary task (Yosinski et al., 2014).

The domain-specific characteristics of a CNN being used for a new task are often addressed through fine-tuning. We provide a brief explanation of the fine-tuning process in the Methods section. Carranza-Rojas et al. (2017) observed that the processes of transfer learning and fine-tuning are important tools that can be used to address the shortage of sufficient domain-specific training data.

Even though large datasets help the performance of DL models, the combination of these technologies (CNNs, transfer learning, and fine-tuning) facilitated the application of DL techniques to other scientific fields. Carranza-Rojas et al. (2017) used transfer learning for herbarium specimens classification, Esteva et al. (2017) for dermatologist-level classification of skin cancer, Gomez Villa et al. (2017) for camera-trap images, Hong et al. (2018) for soccer video scene and event classification, Chen et al. (2018) for airplane detection using remote sensing images, and Pires de Lima et al. (2019) for oil field drill core images. In a study analyzing medical image data, Qayyum et al. (2017) found that transfer learning achieved results

comparable to or better than results from training a CNN model with randomly initialized parameters. Given this record of success to diverse applications, we propose that ML models will also be beneficial for petrographic analysis.

Petrographic Analysis and Thin Sections

Petrographic studies, based on microscopy and image analysis, are essential components of geological analysis, ranging from academic studies of mid-ocean ridges to petroleum-industry exploration and development of shale resource plays. Launeau and Robin (1996), Přikryl (2001), and Nasseri and Mohanty (2008) reported that the progress of computer-aided image analysis techniques has facilitated the characterization of the microscopic properties of the rock through analysis of digital thin section images. The need to partially automate this process has resulted in the proposal of several DL and ML methods.

Cheng and Guo (2017) used CNN models with five, four, and three layers to perform image classification based on granularity analysis from thin-section images. The authors successfully differentiated between three feldspar sandstone classes based solely on grain size: coarse-grained, medium-grained, and fine-grained rocks, achieving an accuracy of 98.5%. With high-resolution micro-computed tomography images or rock samples, Karimpouli and Tahmasebi (2019) used CNN to perform the segmentation of minerals in images mainly composed of quartz. Cheng et al. (2018) used CNN for the image retrieval of rock thin sections. The CNN is used to extract features from the thin-section images which are then stored in a feature database. The images can then be retrieved based on estimates of the similarity between different images, those thin section images stored in the database and the new thin section image

to be classified. De Lima et al. (2019) presented some preliminary results of geoscientific images classifications, including thin section images.

Huang et al. (2016) noted that when crafting CNN models, researchers are uncertain whether to choose from shorter or deeper networks. Shorter networks have a more efficient forward and backward information flow; however, they might not be expressive enough to represent the image features properly. Deeper networks can generate more complex models, helping in feature extraction, but are more difficult to train in practice. We avoid the challenges of model architecture development making use of well-established and robust CNN models previously trained on the ILSVRC.

Data

In our study, we analyze 98 thin sections under plane polarized light (PPL) to identify microfacies. Based on the structure, composition, and porosity five microfacies were identified: argillaceous siltstone, bioturbated siltstones, calcareous siltstone, porous calcareous siltstones, and massive calcite-cemented siltstones. All these microfacies can be identified using plane polarized light and a 10X magnification zoom. We take three photographs for every thin section. The stage where the thin section is placed was rotated randomly for every photograph to simulate different orientations for the same lithofacies. Table 1 summarizes the number of thin sections and respective photographs taken for each one of the five microfacies.

To determine whether the models generated from the data in Table 1 have more general applicability in classifying thin sections coming from different sources, we use thin section images from the public domain (referred to as public data) coming from diverse geological formations stored at the Oklahoma Petroleum Information Center (OPIC) (Table 3).

Table 2: Original data used in this study. The thin sections are from the Mississippian Strata in the Ardmore basin, Oklahoma.

Microfacies	Number of thin sections	Number of photographs
Argillaceous siltstone	16	48
Bioturbated siltstone	29	87
Massive calcareous siltstone	15	45
Massive calcite-cemented siltstone	25	75
Porous calcareous siltstone	13	39

Table 3: Public data used as final test for this study.

Microfacies	Number of thin sections
Argillaceous siltstone	0
Bioturbated siltstone	25
Massive calcareous siltstone	19
Massive calcite-cemented siltstone	18
Porous calcareous siltstone	19
Lithofacies not present in training data (referenced as “Unknown”)	20

Methods

As we make use of robust CNN architectures developed by computer vision specialists and previously put to test on a data-rich problem, we mainly focus on the adaptation of such CNN models to our domain-specific task: the petrographic thin section analysis problem.

Although such an approach (transfer learning and fine-tuning) does not exempt the researcher of common DL use-complexities, it greatly facilitates and accelerates the process of adopting these successful techniques in different fields.

The methodology we follow in this study can be summarized with the flowchart shown in Figure 3. Because grain size plays a crucial role in the petrographic analysis, we use images with a consistent 10x magnification zoom. To compensate for the relatively low resolution of most CNN models used to construct the ILSVRC dataset (usually ranging between 200 by 200 to 400 by 400 pixels), we crop the original thin section photographs (1292 by 968 pixels) into a suite of smaller 644 by 644 pixels, overlapping square images (subimages, Figure 4) thereby augmenting the number of training images. Data augmentation increases the diversity of training samples thereby reducing overfitting (Cireşan et al., 2011; Takahashi et al., 2018). We eliminate the bottom right cropped images because many of them contain an alphanumeric scale bar (Figure 4). The choice for the size of the smaller images is justified as they will have enough resolution to be used for transfer learning, there is some overlap between the subimages helping to show that grain position is not important, and that the size is sufficiently large to avoid isolating spurious bigger grains that could negatively impact the training.

The image cropping processes also increases the reliability in our final test data evaluation. Somewhat similar as for how a petrographer classifies a thin section (or photograph of a thin section) based on an average of the visual aspect of the grains in the complete sample being analyzed, our model provides the classification based on the arguments of the maxima of the smaller subimages. We call such an approach “voting” as the photograph of the thin section will be classified based on the microfacies with the most numbers of “votes”. In that manner, if a thin section image has most of its smaller subimages labeled as argillaceous siltstone and fewer

of those smaller image crops labeled as bioturbated siltstone, the final lithofacies assigned by our model will be of argillaceous siltstone. In the cases in which there is not a single absolute maximum (e.g. five total votes, two votes for massive calcite-cemented siltstone, two votes for porous calcareous siltstone, one vote for massive calcareous siltstone), we declare the model assigned a “tie” for the thin section image.

During initial training, we observed that most of the incorrect CNN prediction labeling was due to a poor color balance in the photographs within the same microfacies, with some images having a color shift to red or yellow. Such color shift occurs due to the difference in color temperature when light passes through the thin section and it goes through the objective lens. Bianco et al. (2017) studied the effects of color balancing and found that suitable color balancing yields a significant improvement in the accuracy for many CNN architectures. We follow Limare et al.'s (2011) methodology and compensate for the color shift assuming that the highest values of red, green, and blue observed in a photograph correspond to white, and the lowest values to black. Figure 5 shows the effect of the color balancing on a representative thin section.

After balancing colors of each image, we subdivide our thin section data into training, validation, and test data sets. The training set goes through another simple step of data augmentation in which we simply rotate the subimages in 90, 180, and 270 degrees; then we flip the initial smaller cropped image around the horizontal axis and rotate it in 90, 180, and 270 degrees again. Therefore, we are able to generate seven variations from a single subimage. Unlike other computer vision tasks in which the orientation or the relative position of an element is important for the overall performance, position and rotation of grains in a thin section are irrelevant. Table 4 shows the training, validation, and testing data set count after the pre-

processing steps. These datasets are based on the subimages and are available to download along with the original parallel polarized thin section photographs.

Table 4: Original data separated in training, validation, and test sets.

Lithofacies	Training set	Validation set	Test set
Argillaceous siltstone	880	55	90
Bioturbated siltstone	1200	110	190
Massive calcareous siltstone	680	70	80
Massive calcite-cemented siltstone	1160	120	125
Porous calcareous siltstone	640	30	85

With the data prepared, we fine-tune four different CNN models: VGG19 (Simonyan and Zisserman, 2014), MobileNetV2 (Sandler et al., 2018), InceptionV3 (Szegedy et al., 2015), and ResNet50 (He et al., 2016). The fine-tuning technique we use is very similar to the one implemented by Yin et al. (2017):

1. Remove the top layers of the CNN model with ILSVRC parameters, and use the CNN model as fixed feature extractor (traditional transfer learning, Yin et al., 2017). With the features extracted by the convolutional layers, we train a small classification network with five outputs (according to our number of classes/microfacies) by using Stochastic Gradient Descent (SGD) optimization.
2. Combine the newly trained small classification network on the top of the CNN model. We again use SGD with a small learning rate (le-4, reducing by a factor of 10 on plateaus), to update the parameters for the complete CNN model.

We use cross-entropy $H(\boldsymbol{p}, \boldsymbol{q})$ during training:

$$H(\boldsymbol{p}, \boldsymbol{q}) = -\sum_{c=1}^{C} \boldsymbol{p}_c \log(\boldsymbol{q}_c) \quad (1)$$

where C is the number of classes, logis the natural logarithm, $\boldsymbol{p}$ represents the true labels, and $\boldsymbol{q}$ the output of the last classification layer in the network. $H(\boldsymbol{p}, \boldsymbol{q})$ represents the cost of a single sample and we minimize the loss, sum of costs of all samples, over all training samples. When we minimize the cross-entropy, we incentivize the CNN to increase the probability the analyzed image to be assigned to the class *c* when the image true label belongs to the class *c*.

We evaluate the performance of the fine-tuned models based on the test data separated from our original data set. We then select the best model and perform a final evaluation based on the classification our model provides to the public data. To perform the final evaluation, we use the six subimages crops (Figure 4) and three extra randomly centered crops with the same dimensions as the subimages as shown in Figure 4. These three extra subimages help in the voting process to reduce the chances of ties.

Results

Table 5 shows the test set accuracy of the four fine-tuned CNN models. We present both the accuracy for the subimages and the accuracy for the resulting thin section photograph voting. All models reach an accuracy higher than 90%. The fine-tuned InceptionV3 and ResNet50 tied with the best accuracy (0.96) in the test set data. We select the model ResNet50 and provide a detailed analysis of its results.

Figure 6 shows the resulting classification assigned by the fine-tuned ResNet50 to different subimages of each one of the five classes present in the training data. For each one of the five classes, we select thin section photographs of subimages in which the fine-tuned

ResNet50 assigned the same classification as the petrographer. The examples we select did not have all of the possible voting subimages agreeing on the assigned class (i.e. most, but not all, of the subimages voted for the same classification as the final/petrographer provided class). Therefore Figure 6 shows examples in which the class provided by the fine-tuned ResNet50 for the photographs subimages agrees with the classification provided by the petrographer as well as examples in which the classification is different.

We compare the performance of the fine-tuned ResNet50 against the petrographer-provided classification for both the smaller cropped images as well as the thin section photographs making use of confusion matrices. Figure 7 shows the confusion matrix for the subimages test set and Figure 8 shows the confusion matrix for the thin section photographs test set.

For our final analysis, we used the fine-tuned ResNet50 to classify public data from the OPIC. Figure 9 shows the confusion matrix for the public data thin section photographs. This evaluation of our model using public data is an important comparison for this project as it serves as initial evaluation of a possible multi-basin thin section CNN classifier. As we continue to add more training data and further tailor our CNN models, we anticipate further acceleration and accuracy of thin section analysis.

Table 5: Test set accuracy of smaller crop images and thin section photographs provided by fine-tuned models. The thin section receives the label according to the winning vote of its labeled smaller image crops.

Fine-tuned model	Accuracy – subimages	Accuracy – thin section photograph voting
VGG19	0.93	0.95
MobileNetV2	0.91	0.93
InceptionV3	0.91	0.96
ResNet50	0.91	0.96

Discussion

Based on our bibliographic research, this is the first study conducted using rock thin sections, a crucial source for sedimentary and diagenetic analysis, as input for a CNN model that can be used to classify different microfacies. In the methodology we implement, a user can take multiple photographs of a single thin section, and obtain its classification as predicted by the model. Based on our tests, the accuracy of the procedure we described is comparable to accuracies of a petrographer, as long as the lithofacies being analyzed were present in the training data. Our study is different than Cheng and Guo (2017) because we differentiate between five different lithofacies, whereas Cheng and Guo (2017) differentiate between three granulometric classifications. Neither Cheng and Guo (2017) or Karimpouli and Tahmasebi (2019) provide the metrics of their model when tested with significantly different data, as we present in our public data evaluation.

Unlike a human interpreter who relies upon a defined set of morphological measurements to perform microfacies classifications, the CNN operates from no knowledge of specific attribute analysis and performs the classification based on image characteristics. In this manner, CNN

labeled datasets have the potential to reduce petrographer bias, yielding a reduced inconsistency on thin sections classification. This also implies that a CNN model, with this current implementation, will always assign a microfacies for any image analyzed. When analyzing a new image, the CNN model (as implemented in this study) will always generate a set of probabilities that such image belongs to the CNN's learned microfacies. For that reason, Figure 9 shows that the CNN provides classifications for all the thin sections classified as unknown by the petrographer. The number of unknowns can be reduced when more examples of microfacies are provided to the CNN models.

Figure 6 indicates that the CNN misclassifications are in fact similar to the description a petrographer would assign to a particular section of a thin section photograph. Therefore, our voting scheme then is helpful as it reduces possible misconceptions. Due to thin section heterogeneities, the CNN classification maybe is correct for the particular subimage in analysis.

One of the explanations for the misclassification is the criteria that the petrographer used during the interpretation of the thin sections. There are two main groups of rock types: (1) structureless or massive, and (2) structured. To divide the microfacies within these two main groups, the petrographers uses a qualitative-visual criterion. For example, the massive siltstones can be calcareous, porous, and calcite-cemented. However, the criteria used to divide between them was the visual content of calcite cement and porosity and no statistical method was used to quantify the proportion of cement or porosity. We suggest including other data to quantify the amount of cement, mineralogy and porosity. With a more quantitative interpretation, we can reduce the interpretation bias.

Another explanation for the label bias is the use of thin sections with different characteristics. The model was built with thin sections stained with red alizarin for calcite

identification, and blue alizarin for porosity identification. However, public data thin sections do not always have these features. Therefore, thin sections with high calcite content could be re-labeled as microfacies without calcite. In fact, most of the confusion between massive calcite-cemented siltstones and the calcareous siltstones could be explained by the lack of alizarin stains.

Finally, the photograph by itself plays an important role in the model and so can contribute to the label bias. The original labels resulted from the observation of the actual thin sections under the microscope and not based on the photographs. Dozens of different photographs without any overlap can be taken from the same thin section with 10X objective magnification. The photographs we captured for this study were taken randomly in different locations of the thin section. However, what differentiates between argillaceous and bioturbated siltstones are the bioturbation patterns. Some photographs of bioturbated siltstones do not show evident bioturbation, but there is evidence of bioturbation in the thin section that can be observed under the microscope. Thus, to avoid misclassification the photographs should depict the criteria used by the petrographer for the original classification. This difficulty in capturing complete characteristics of the entirety of the thin sections with random photographs indicates that most of the misclassification is the result of the preparation and labeling of the data used to train the model rather than the CNN model by itself. This misclassification pattern also shows a potential improvement that the use of CNN models can provide. If the thin section is captured in its entirety, the CNN can quickly provide classifications for all its sections. A petrographer can then quality control the CNN results as well as easily note outliers that could either be mistakes or important features that can be further analyzed.

As the digitization of legacy data accelerates, and thin section preparation and data storage methodologies are standardized the approach presented here can improve with more

detailed and directed image processing. Image segmentation techniques can be used to differentiate between different minerals, which can be a powerful tool for microfacies classification. The technique we demonstrate in this manuscript is very general and can easily be modified to suit the identification of thin sections coming from different formations.

Conclusions

In this paper, we propose the use of transfer learning and fine-tuning of robust CNN models for petrographic thin section classification, achieving accuracies above 90% for all the models tested. Furthermore, with our test with public data, we investigate how CNN models can be used to classify petrographic samples acquired with significantly different parameters. In the future, further experiments shall be conducted to increase the number of lithofacies that can be identified by the CNN.

We focus on the use of parallel polarized petrographic thin section images as they are sufficient to differentiate between the classes/microfacies present in our dataset. Cross-polarized images shall be included for the cases in which such imaging technique is crucial for proper lithofacies classification, for example to differentiate between a rock enriched in quartz grains and a rock enriched in feldspars grains. In addition, this paper mostly concentrates on the use of CNN models at a specific 10x magnification level. As different lithological and diagenetic properties can only be analyzed in different scales, many other studies can be conducted with a similar technique. Our manuscript targets petrographic thin section classification, but other geoscience tasks can be accelerated with the use of ML. Seismic denoising and interpretation, wireline well logging, and remote sensing classification are some of many fields that are implementing analysis driven by ML models. We believe that the implementation of the

methodology we discuss here has the potential to further improve petrographic thin section classification speed and help geoscientists make use of such invaluable data.

Figures and figure captions

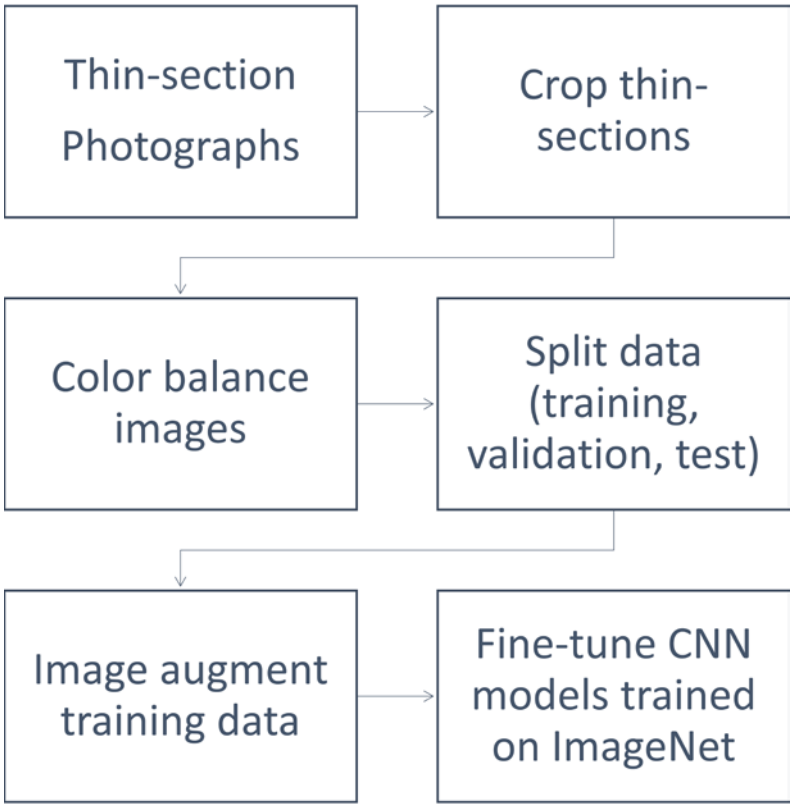

Figure 3: Methodology flowchart. Data preparation is an important part of the procedures for the work we present in this paper. We first take multiple pictures of each one of the 98 thin sections available. These photographs are then cropped in multiple ways, helping us increase the dataset for training and to remove unwanted image features (the scale bar). We then color balance the cropped images and split the data in training, validation, and test set. The training set data is augmented using simple image rotations. We then have appropriate data to be used for fine-tuning the CNN models.

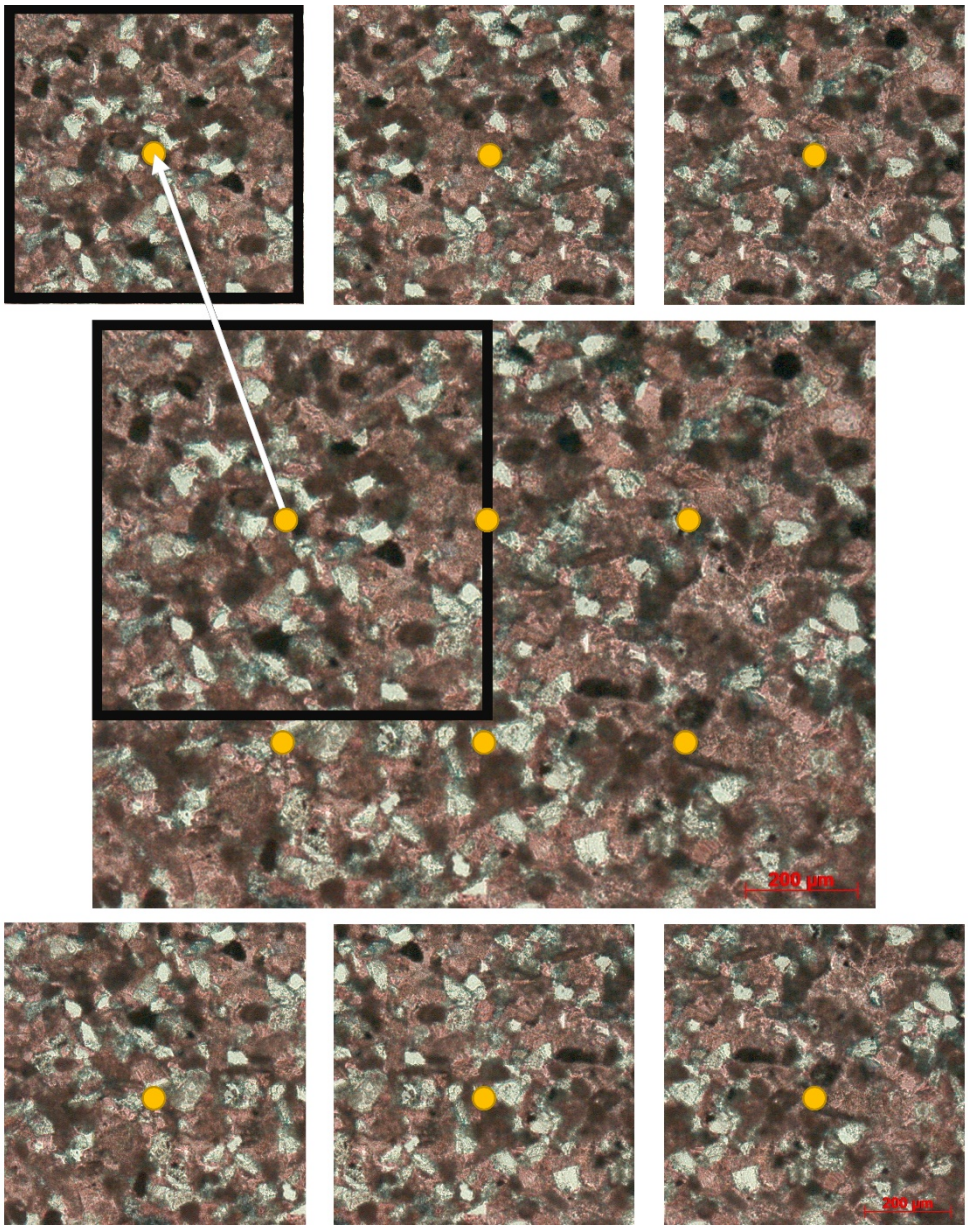

Figure 4: An original photograph of a Massive calcareous siltstone thin section (center, bigger) taken with 10x objective magnification and the subimages used for training and testing (top and bottom rows, smaller). The subimage a indicates with a black outline the boundaries and the center of the cropped image with a golden circle with the respective letter, the other subimages are only represented by their center letters. The subimage f is discarded in the training and validation set as some original photographs will be marked with a scale bar.

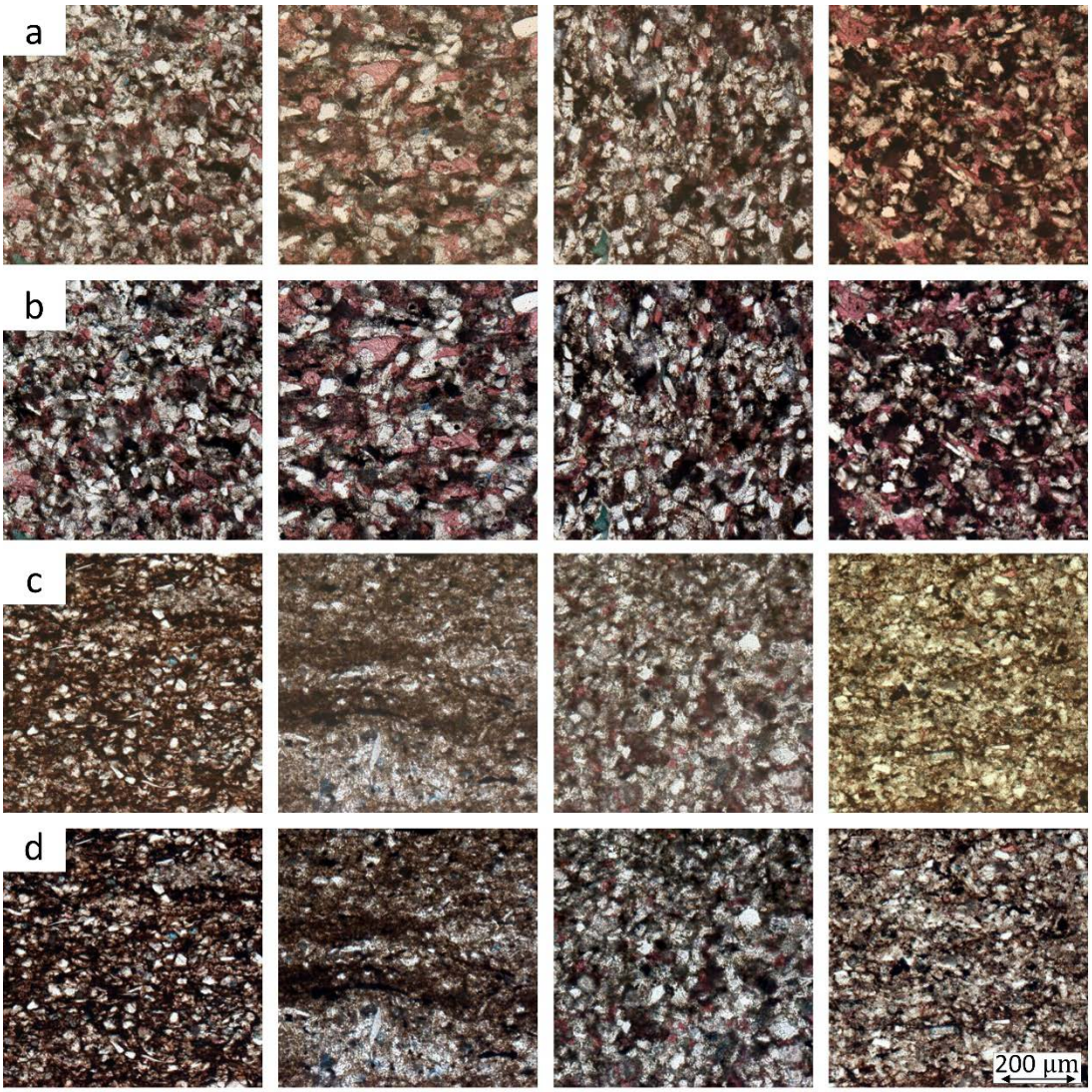

Figure 5: Effects of color balancing. Row (a) examples of cropped photographs of massive calcareous siltstone before and row (b) after color balancing. Row (c) bioturbated siltstone before and (d) after color balancing. Note the examples in the last column. Sometimes photographs tend to be yellow, red or blue. The color balancing process helps to merge these images with the rest of the dataset.

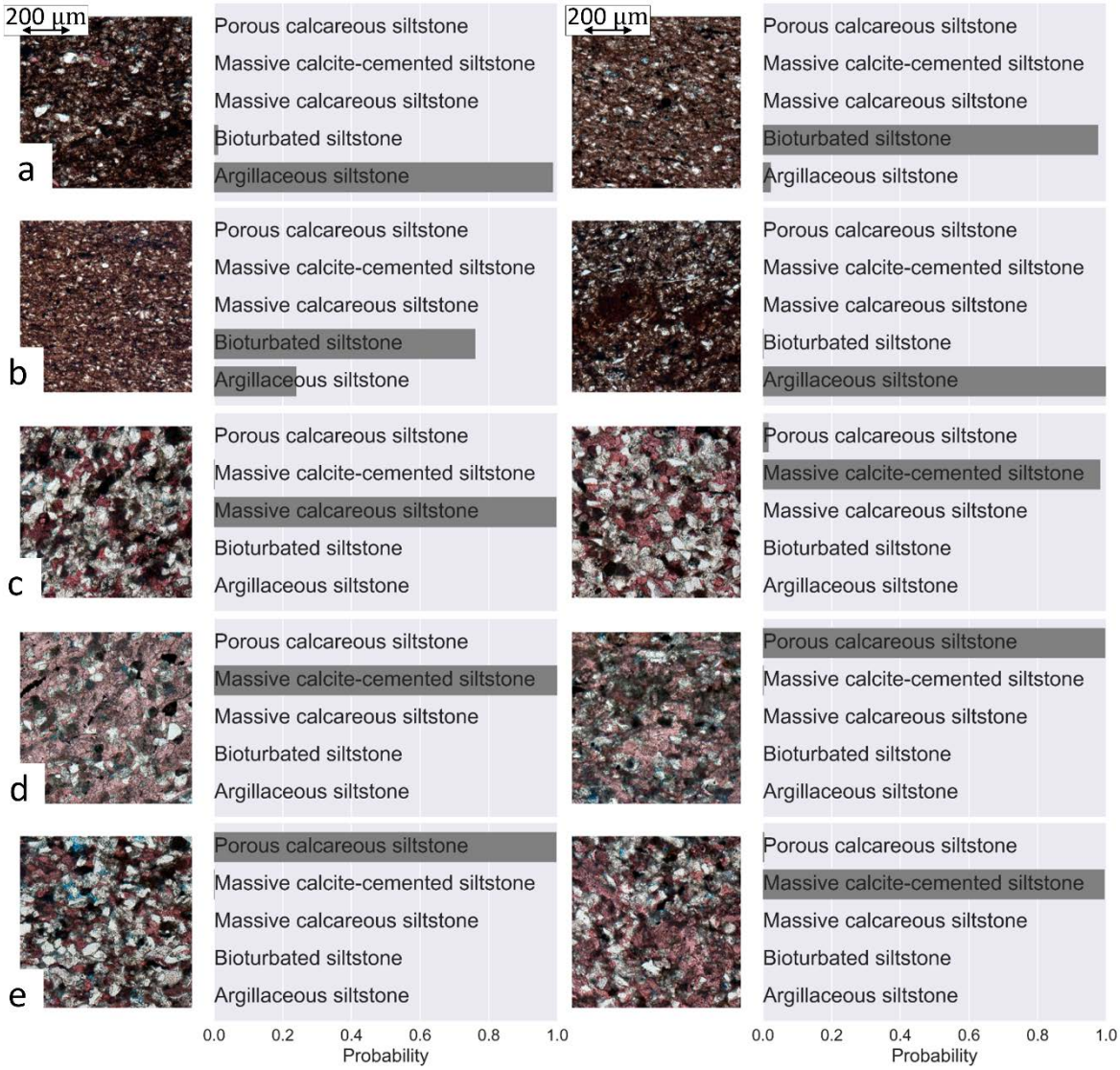

Figure 6: Examples of classification provided by fine-tuned ResNet50 for the smaller cropped images in the test set. Images in the same row were extracted from the same microfacies as labeled by the interpreter. The left column shows examples of smaller cropped images in which the classification provided by the CNN model is the same as the classification provided by the petrographer. In contrast, the right column shows examples of smaller cropped images in which the classification provided by the CNN is not the same as the classification provided by the petrographer. Row (a) shows smaller crops extracted from a photograph classified as argillaceous siltstone by the petrographer, row (b) was classified as bioturbated siltstone, (c) as massive calcareous siltstone, (d) massive calcite-cemented siltstone, and (e) porous calcareous siltstone.

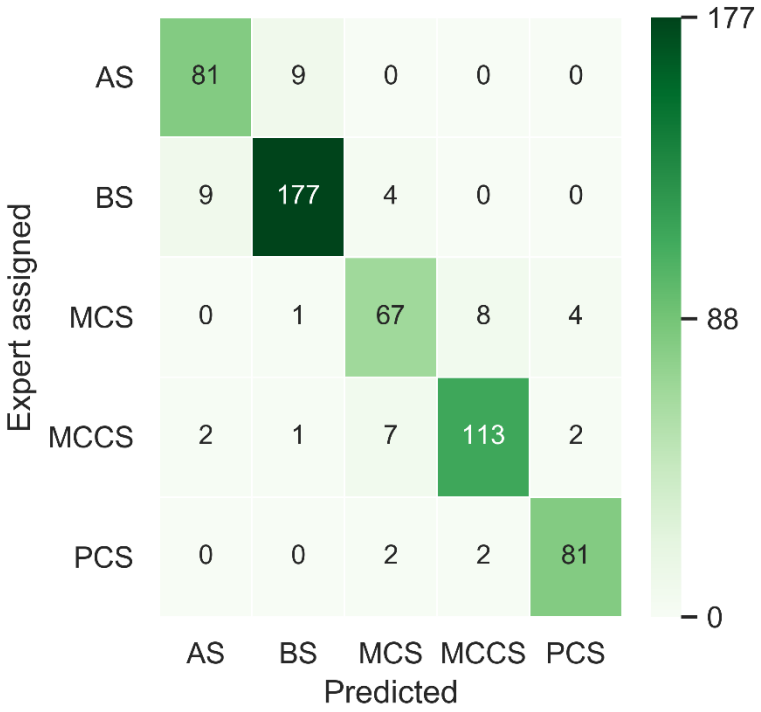

Figure 7: Confusion matrix comparing the classification provided by the petrographer expert and the classification obtained with the fine-tuned ResNet50 for the test set smaller image crops. The class names are abbreviated: Argillaceous siltstone (AS), Bioturbated siltstone (BS), Massive calcareous siltstone (MCS), Massive calcite-cemented siltstone (MCCS), and Porous calcareous siltstones (PCS).

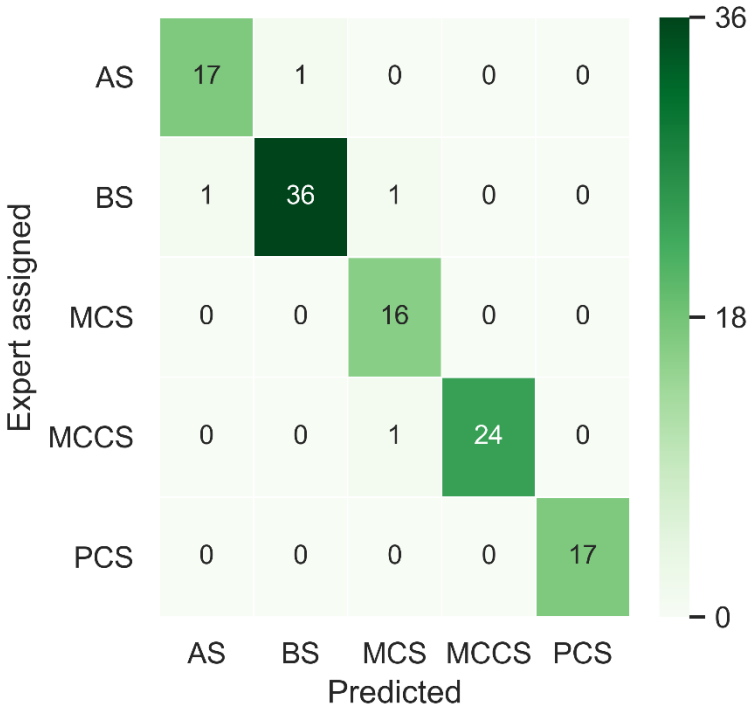

Figure 8: Confusion matrix comparing the classification provided by the petrographer expert and the classification obtained with the fine-tuned ResNet50 for the test set thin section photographs. The class names are abbreviated: Argillaceous siltstone (AS), Bioturbated siltstone (BS), Massive calcareous siltstone (MCS), Massive calcite-cemented siltstone (MCCS), and Porous calcareous siltstones (PCS).

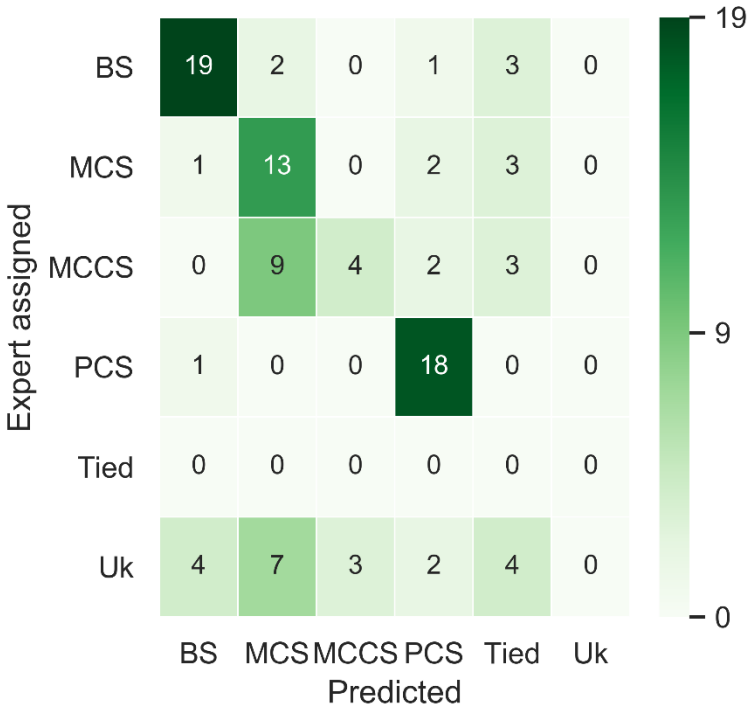

Figure 9: Confusion matrix comparing the classification provided by the petrographer expert and the classification obtained with the fine-tuned ResNet50 for the final public data test set of thin section photographs. The class names are abbreviated: Bioturbated siltstone (BS), Massive calcareous siltstone (MCS), Massive calcite-cemented siltstone (MCCS), Porous calcareous siltstones (PCS), Tied, and Unknown (Uk).

Acknowledgments

Rafael acknowledges CNPq (grant 203589/2014-9) for the financial support and CPRM for granting the leave of absence allowing the pursuit of his Ph.D. studies. The data used to build the CNN model was funded through University of Oklahoma's Institute of Reservoir Characterization Consortium.

References

Bengio, Y., 2012. Deep Learning of Representations for Unsupervised and Transfer Learning, in: Guyon, I., Dror, G., Lemaire, V., Taylor, G., Silver, D. (Eds.), Proceedings of ICML Workshop on Unsupervised and Transfer Learning, Proceedings of Machine Learning Research. PMLR, Bellevue, Washington, USA, pp. 17–36.

Bianco, S., Cusano, C., Napoletano, P., Schettini, R., Bianco, S., Cusano, C., Napoletano, P., Schettini, R., 2017. Improving CNN-Based Texture Classification by Color Balancing. J. Imaging 3, 33. https://doi.org/10.3390/jimaging3030033

Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., Joly, A., 2017. Going deeper in the automated identification of Herbarium specimens. BMC Evol. Biol. 17, 181. https://doi.org/10.1186/s12862-017-1014-z

Caruana, R., 1995. Learning Many Related Tasks at the Same Time with Backpropagation, in: Tesauro, G., Touretzky, D.S., Leen, T.K. (Eds.), Advances in Neural Information Processing Systems 7. MIT Press, pp. 657–664.

Chen, Z., Zhang, T., Ouyang, C., Chen, Z., Zhang, T., Ouyang, C., 2018. End-to-End Airplane Detection Using Transfer Learning in Remote Sensing Images. Remote Sens. 10, 139. https://doi.org/10.3390/rs10010139

Cheng, G., Guo, W., 2017. Rock images classification by using deep convolution neural network. J. Phys. Conf. Ser. 887, 012089. https://doi.org/10.1088/1742-6596/887/1/012089

Cheng, G., Yue, Q., Qiang, X., 2018. Research on Feasibility of Convolution Neural Networks for Rock Thin Sections Image Retrieval. 2018 2nd IEEE Adv. Inf. Manag. Autom. Control Conf. 2539–2542.

Chollet, F., 2016. Xception: Deep Learning with Depthwise Separable Convolutions. CoRR abs/1610.0.

Cireşan, D.C., Meier, U., Masci, J., Gambardella, L.M., Schmidhuber, J., 2011. Flexible, High Performance Convolutional Neural Networks for Image Classification, in: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two, IJCAI'11. AAAI Press, pp. 1237–1242. https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-210

Datta, R., Joshi, D., Li, J., Wang, J.Z., 2008. Image Retrieval: Ideas, Influences, and Trends of the New Age. ACM Comput. Surv. 40.

de Lima, R.P., Bonar, A., Coronado, D.D., Marfurt, K., Nicholson, C., 2019. Deep convolutional

neural networks as a geological image classification tool. Sediment. Rec. 17, 4–9. https://doi.org/10.210/sedred.2019.2
Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S., 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118. https://doi.org/10.1038/nature21056
Gomez Villa, A., Salazar, A., Vargas, F., 2017. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. Ecol. Inform. 41, 24–32. https://doi.org/10.1016/J.ECOINF.2017.07.004
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90
Hong, Y., Ling, C., Ye, Z., 2018. End-to-end soccer video scene and event classification with deep transfer learning, in: 2018 International Conference on Intelligent Systems and Computer Vision (ISCV). IEEE, pp. 1–4. https://doi.org/10.1109/ISACV.2018.8369043
Huang, G., Liu, Z., Weinberger, K.Q., 2016a. Densely Connected Convolutional Networks. CoRR abs/1608.0.
Huang, G., Sun, Y., Liu, Z., Sedra, D., Weinberger, K., 2016b. Deep Networks with Stochastic Depth.
Karimpouli, S., Tahmasebi, P., 2019. Segmentation of digital rock images using deep convolutional autoencoder networks. Comput. Geosci. 126, 142–150. https://doi.org/10.1016/J.CAGEO.2019.02.003
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, NIPS'12. Curran Associates Inc., USA, pp. 1097–1105.
Launeau, P., Robin, P.-Y.F., 1996. Fabric analysis using the intercept method. Tectonophysics 267, 91–119. https://doi.org/10.1016/S0040-1951(96)00091-1
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539
Limare, N., Lisani, J.-L., Morel, J.-M., Petro, A.B., Sbert, C., 2011. Simplest Color Balance. Image Process. Line 1, 297–315. https://doi.org/10.5201/ipol.2011.llmps-scb
Machine Learning Glossary | Google Developers [WWW Document], n.d. URL https://developers.google.com/machine-learning/glossary/#top_of_page (accessed 1.21.19).
Mou, L., Ghamisi, P., Zhu, X.X., 2017. Deep Recurrent Neural Networks for Hyperspectral Image Classification. IEEE Trans. Geosci. Remote Sens. 55, 3639–3655. https://doi.org/10.1109/TGRS.2016.2636241
Nasseri, M.H.B., Mohanty, B., 2008. Fracture toughness anisotropy in granitic rocks. Int. J. Rock Mech. Min. Sci. 45, 167–193. https://doi.org/10.1016/J.IJRMMS.2007.04.005
Olah, C., Mordvintsev, A., Schubert, L., 2017. Feature Visualization. Distill. https://doi.org/10.23915/distill.00007
Olah, C., Satyanarayan, A., Johnson, I., Carter, S., Schubert, L., Ye, K., Mordvintsev, A., 2018. The Building Blocks of Interpretability. Distill. https://doi.org/10.23915/distill.00010
Pires de Lima, R., Suriamin, F., Marfurt, K.J., Pranter, M.J., 2019. Convolutional neural networks

as aid in core lithofacies classification. Interpretation 7, SF27–SF40. https://doi.org/10.1190/INT-2018-0245.1
Přikryl, R., 2001. Some microstructural aspects of strength variation in rocks. Int. J. Rock Mech. Min. Sci. 38, 671–682. https://doi.org/10.1016/S1365-1609(01)00031-4
Qayyum, A., Anwar, S.M., Awais, M., Majid, M., 2017. Medical image retrieval using deep convolutional neural network. Neurocomputing 266, 8–20. https://doi.org/10.1016/J.NEUCOM.2017.05.025
Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y
Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. ArXiv e-prints.
Simonyan, K., Vedaldi, A., Zisserman, A., 2013. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. CoRR abs/1312.6.
Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv e–prints.
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2014. Going Deeper with Convolutions. CoRR abs/1409.4.
Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2015. Rethinking the Inception Architecture for Computer Vision. CoRR abs/1512.0.
Takahashi, R., Matsubara, T., Uehara, K., 2018. Data Augmentation using Random Image Cropping and Patching for Deep CNNs. CoRR abs/1811.0.
Yang, X., Ye, Y., Li, X., Lau, R.Y.K., Zhang, X., Huang, X., 2018. Hyperspectral Image Classification With Deep Learning Models. IEEE Trans. Geosci. Remote Sens. 56, 5408–5423. https://doi.org/10.1109/TGRS.2018.2815613
Yin, X., Chen, W., Wu, X., Yue, H., 2017. Fine-tuning and visualization of convolutional neural networks, in: 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, pp. 1310–1315. https://doi.org/10.1109/ICIEA.2017.8283041
Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? Adv. Neural Inf. Process. Syst. 27, 3320–3328.

Final Remarks

In this thesis, I show general applications of transfer learning for geoscience images and provide details on the application so convolutional neural networks for microfacies identification. In Chapter 1 I show a generic application of transfer learning for many different geoscience images. In Chapter 2, the main contribution of this thesis, I provide more details of microfacies identification using convolutional neural networks.

As Chapter 2 reports, the methodology I use is greatly dependent on image acquisition standardization. Color plays a significant role in mineral classification, and samples with and without alizarin-red or blue-epoxy stains present significantly different colors for the same minerals under plane parallel light. Such shortcoming, however, can easily be addressed by data owners, either saving images of both clean and stained thin-section, or maintaining a standard data preparation. Nonetheless, as discussed in Chapter 2, my implementation has the potential to be automatically coupled with imaging tools for fast microfacies identification. Oil and gas companies already use the necessary tools to completely store the image of thin-sections in their entirety, a process that can provide data with much more details than the one I use in this thesis. Thus, the identification of particular properties important for the classification of the thin-section, e.g. bioturbation, could be addressed with a greater amount of detail and accuracy.

Not explored in this thesis due to resources constraints, the use of images of thin-sections under cross parallel light might facilitate thin-section classification as they bring more features helpful for mineral identification. Another topic that should be explored is the segmentation of the minerals in the imaged thin-section instead of the classification of the thin-section as a whole. For a segmentation task, a dataset of thin-section images with masks labels, i.e. an image with each one of the minerals, pores, and cement identified, is crucial, but such dataset will demand

significant resources to be generated. Such dataset could also help identification of porosity. Wang et al. (2018), Alqahtani et al. (2019), and Duarte-Coronado et al., (2019) presented examples of the use of convolutional neural networks for the estimation of porosity using micro X-ray computed tomography images, high-resolution scanning electron microscope images, and thin-section images. Unlike classification problems, in which we aim to predict a categorical data, research on the use of machine learning for prediction of continuous properperties, like porosity or modal composition, are still underexplored. An expert-created thin-section dataset could help members of the community as a benchmark for different tasks. Many geoscientific data seem to remain inexplored by machine learning, yet the advances and popularization of deep learning models can lead to the creation of helpful tools. For example, vitrinite reflectance interpretation is heavily dependent on interpreter experience, yet such experience could likely be modeled by convolutional neural networks. With the current increase on the applications of machine learning in different fields of geoscience ranging from seismology studies (e.g. Perol et al., 2018; Ren et al., 2019; Sinha et al., 2018), geological mapping (e.g. Cracknell and Reading, 2014; Pires de Lima and Marfurt, 2018), seismic facies classification (e.g. Lubo-Robles and Marfurt, 2019; Qi et al., 2016; Zhao, 2018; Zhao et al., 2016), and CO_2 sequestration (e.g. Pires de Lima et al., 2019; Pires de Lima and Lin, 2019; Sun et al., 2019; Zhong et al., 2019), it is likely we will experience an increase of applications of machine learning methods using thin-section data in the near future.

References

Alqahtani, N., Alzubaidi, F., Armstrong, R.T., Swietojanski, P., Mostaghimi, P., 2019. Machine learning for predicting properties of porous media from 2d X-ray images. J. Pet. Sci. Eng. 106514. https://doi.org/https://doi.org/10.1016/j.petrol.2019.106514

Cracknell, M.J., Reading, A.M., 2014. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial

distribution of training data and the use of explicit spatial information. Comput. Geosci. 63, 22–33. https://doi.org/10.1016/J.CAGEO.2013.10.008
Duarte-Coronado, D., Tellez-Rodriguez, J., Pires de Lima, R., Marfurt, K., Slatt, R., 2019. Deep convolutional neural networks as an estimator of porosity in thin-section images for unconventional reservoirs, in: SEG Technical Program Expanded Abstracts 2019. Society of Exploration Geophysicists, pp. 3181–3184. https://doi.org/10.1190/segam2019-3216898.1
Lubo-Robles, D., Marfurt, K.J., 2019. Independent Component Analysis for reservoir geomorphology and unsupervised seismic facies classification in the Taranaki Basin, New Zealand. Interpretation 1–76. https://doi.org/10.1190/int-2018-0109.1
Perol, T., Gharbi, M., Denolle, M., 2018. Convolutional neural network for earthquake detection and location. Sci. Adv. 4, e1700578. https://doi.org/10.1126/sciadv.1700578
Pires de Lima, R., Lin, Y., 2019. Geophysical data integration and machine learning for multi-target leakage estimation in geologic carbon sequestration, in: SEG Technical Program Expanded Abstracts 2019. pp. 2333–2337. https://doi.org/10.1190/segam2019-3215405.1
Pires de Lima, R., Lin, Y., Marfurt, K.J., 2019. Transforming seismic data into pseudo-RGB images to predict CO2 leakage using pre-learned convolutional neural networks weights, in: SEG Technical Program Expanded Abstracts 2019. Society of Exploration Geophysicists, pp. 2368–2372. https://doi.org/10.1190/segam2019-3215401.1
Pires de Lima, R., Marfurt, K.J., 2018. Principal component analysis and K-means analysis of airborne gamma-ray spectrometry surveys. https://doi.org/10.1190/segam2018-2996506.1
Qi, J., Lin, T., Zhao, T., Li, F., Marfurt, K., 2016. Semisupervised multiattribute seismic facies analysis. Interpretation 4, SB91–SB106. https://doi.org/10.1190/INT-2015-0098.1
Ren, C.X., Dorostkar, O., Rouet-Leduc, B., Hulbert, C., Strebel, D., Guyer, R.A., Johnson, P.A., Carmeliet, J., 2019. Machine Learning Reveals the State of Intermittent Frictional Dynamics in a Sheared Granular Fault. Geophys. Res. Lett. 46, 7395–7403. https://doi.org/10.1029/2019GL082706
Sinha, S., Wen, Y., Pires de Lima, R.A., Marfurt, K., 2018. Statistical controls on induced seismicity. Unconventional Resources Technology Conference. https://doi.org/10.15530/urtec-2018-2897507-MS
Sun, A.Y., Zhong, Z., Jeong, H., Yang, Q., 2019. Building complex event processing capability for intelligent environmental monitoring. Environ. Model. Softw. 116, 1–6. https://doi.org/https://doi.org/10.1016/j.envsoft.2019.02.015
Wang, Y., Arns, C.H., Rahman, S.S., Arns, J.-Y., 2018. Porous Structure Reconstruction Using Convolutional Neural Networks. Math. Geosci. 50, 781–799. https://doi.org/10.1007/s11004-018-9743-0
Zhao, T., 2018. Seismic facies classification using different deep convolutional neural networks, in: SEG Technical Program Expanded Abstracts 2018. Society of Exploration Geophysicists, pp. 2046–2050. https://doi.org/10.1190/segam2018-2997085.1
Zhao, T., Zhang, J., Li, F., Marfurt, K.J., 2016. Characterizing a turbidite system in Canterbury Basin, New Zealand, using seismic attributes and distance-preserving self-organizing maps. Interpretation 4, SB79–SB89. https://doi.org/10.1190/INT-2015-0094.1
Zhong, Z., Sun, A.Y., Yang, Q., Ouyang, Q., 2019. A deep learning approach to anomaly detection in geological carbon sequestration sites using pressure measurements. J. Hydrol. 573, 885–894. https://doi.org/https://doi.org/10.1016/j.jhydrol.2019.04.015