SPE-197255-MS

Carbonate Lithology Identification with Machine Learning

Takashi Nanjo, Japan Oil, Gas and Metals National Corporation; Satoru Tanaka, Japan Oil, Gas and Metals National Corporation; GRID INC.

This paper was prepared for presentation at the Abu Dhabi International Petroleum Exhibition & Conference held in Abu Dhabi, UAE, 11-14 November 2019.

Abstract

Machine learning has attracted the attention of geoscientists over the years. In particular, image analysis via machine learning has promise for application to exploration and production technologies. Demands have grown for the automation of carbonate lithology identification to shorten the delivery time of work and to enable unspecialized engineers to conduct it. The image analysis of carbonate thin sections is time consuming and requires expert knowledge of carbonate sedimentology. In this study, the authors propose an image analysis technique based on deep neural network for carbonate lithology identification of a thin section, which is an important image analysis process required for oil and gas exploration. In addition, the authors consider that porosity and permeability variations in the same facies are controlled by the grain, cement, pore, and limemud contents. If the contents are accurately measured, the porosity and permeability can be determined more accurately than by using traditional methods such as point counting. The elucidation of the complex relation of porosity and permeability is the objective of automation of carbonate lithology identification. To perform image analysis of the thin section, the authors prepared a data set mainly comprising pictures of the Pleistocene Ryukyu Group, which were composed of reef complex deposits distributed in southern Japan. The data set contains 306 thin section pictures and annotation data labeled by a carbonate sedimentologist. The rock components was divided into four types (grain, cement, pore, and limemud). A convolution neural network (CNN) was utilized to train the model. After training the neural network, each of the four categories was interpreted by the trained model automatically. Resultantly, the accuracy of automatic Dunham classification was 90.6% and the mean average test accuracy of category identification was 83.9%. The interpretation seems highly consistent between human vision and machine vision in both the overview and pixelwise scales. This result indicates that it has sufficient potential to assist geologists and become a basic tool for practical applications. However, the accuracy of category identification is still insufficient. The authors believe that the model requires higher quality supervised data and a greater number of supervised data.

Introduction

Carbonate is a key component for understanding oil and gas reservoirs. The reservoirs of the largest oil and gas fields in the world such as the Zakum field in Abu Dhabi, the Ghawar field in the Kingdom of

Saudi Arabia, the North Dome field in the State of Qatar, and South Pars field in the Islamic Republic of Iran consist of carbonates. Carbonate sedimentary rocks are composed of various kinds of grains, and it has complex sedimentary textures. Primary carbonate sediments consist of carbonate minerals such as aragonite and calcite. These minerals easily suffer from diagenesis after deposition and consequently make the carbonate sedimentary rock texture more complex. Therefore, a detailed carbonate facies analysis requires high expertise. Additionally, traditional thin section analyses approaches such as the point counting method require excessive time to conduct a large number of thin section analyses.

Machine learning, including deep learning, is attracting significant attention. In particular, image analysis using convolutional neural networks (CNNs) has been dramatically developed since the emergence of AlexNet in 2012 (Krizhevsky et al. 2012). CNNs achieved superhuman image recognition capability by utilizing a deep layer structure that consists of a convolutional layer, activation function, etc. In the field of petroleum exploration and production, several studies on image analysis using CNNs have been performed by petroleum exploration and production companies and universities (e.g. Alqahtani et al. 2018; Du et al. 2018; Saporetti et al. 2018; Alaudah et al. 2019; Zeng et al. 2019). However, few studies have focused on carbonate lithology interpretation at the pixel scale by using machine learning.

Currently, pixel wise segmentation using machine learning is a common approach for image analysis. It predicts a suitable category or value for each pixel based on the non-linear relations of pixels around the target pixel. The applications of pixel wise segmentation are widespread in automotive and medical image analyses (e.g. Liu et al. 2017; Siam et al. 2017; Fernández et al. 2018; Moor et al. 2018).

The porosity and permeability of carbonate reservoirs have a significant impact of productivity of reservoirs. They exhibit some correlation with facies and the depositional environment (e.g. Mahdi et al. 2013). However, porosity and permeability usually vary even in the same facies and depositional environment. That porosity and permeability variations may be controlled by the proportion of grain, cement, pore, and limemud content. The elucidation of this complex relation of porosity and permeability is the final objective of the automation of carbonate lithology identification.

In this study, pixel wise segmentation utilizing CNN was performed for the automation of carbonate lithology identification.

Dataset and methodology

The authors collected 306 thin section pictures from published data and JOGMEC owned data. The samples were mainly composed of the Plesitocene Ryukyu Group. However, it included not only the Ryukyu Group but also the other carbonates. These thin section pictures cover grainstone, packstone, wackestone, and limemudstone, and the thin section pictures were taken to include different variations within the same facies. These thin section pictures were interpreted and annotated by a carbonate sedimentologist. Then, pixels of these pictures were divided into four categories (grain, cement, pore, and limemud). The limemud area was defined as the area not containing grain, cement, or pore.

To reshape and augment the data set, the annotated images were split into left and right side images. The total number of thin section pictures and annotated images were 612. From all the data, 80% was used as the training data set and 20% were used as the test data set. Therefore, the total number training data and training label data was 490 and that for the test data was 122. These data were randomly separated into the training and test data sets.

The dimensions of the training data and label data were 512×512 pixels, and the sizes of the output images were identical. The hyper parameters were selected to minimize learning loss and to reach the sufficient convergence. The average accuracy, precision, recall, f1 score, mean IoU, limemud accuracy, cement accuracy, grain accuracy, and pore accuracy were calculated based on the test results.

The area percentages of grain, cement, pore, and limemud were calculated by counting the pixels of each category. Carbonates are categorized into six lithologies (crystalline carbonate, boundstone, grainstone,

packstone, wackestone, limemudstone) in Dunham classification (Dunham 1962). These lithologies are defined by depositional texture, grain content, and limemud content. The data set of this study covers grainstone, packstone, wackestone, and limemudstone.

Results

The mean average test accuracy of the trained model was 83.9% (Fig. 1). The average values for precision, recall, F1 score, and mean IoU were 85.7%, 83.9%, 83.7%, and 64.9%, respectively. The mean average test accuracy of the trained model achieved stable equilibrium in approximately 300 epochs. The average loss gradually decreased with increasing epochs (Fig.1). This indicates that the trained model converges well.

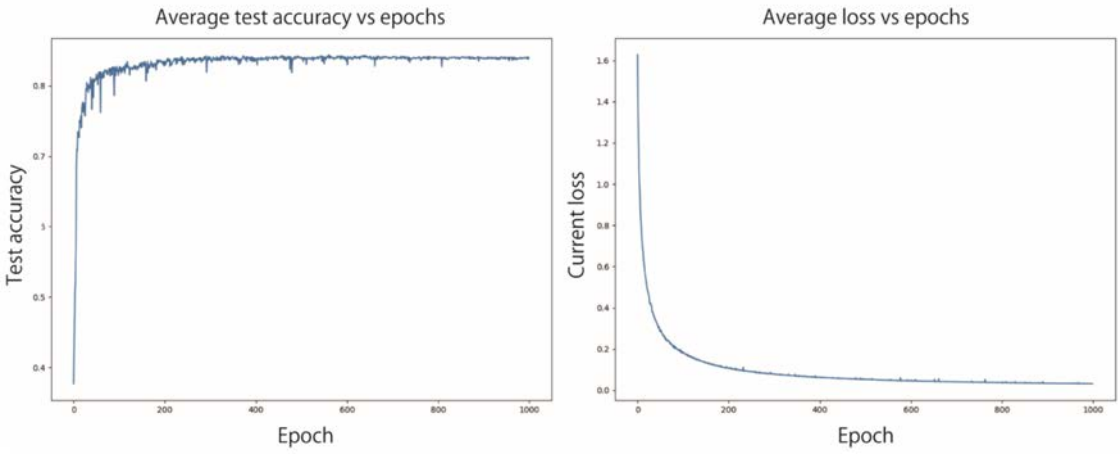

Figure 1—Average test accuracy vs epoch and average loss vs epochs.

The average accuracy of each category was as follows: grain was 80.9%, cement was 74.1%, pore was 85.0%, and limemud was 76.4%. Relatively speaking, grain and pore had a high accuracy whereas cement and limemud had a low accuracy.

The highest average accuracy sample is shown in Fig. 2. The average accuracy, precision, recall, and mean IoU were 98.6%, 98.6%, 98.6%, and 96.2%, respectively. The ground truth and the trained model prediction perfectly match and exhibit a high accuracy in all categories (grain, cement, and pore). In this thin section, the boundaries of each area are clear. The lithology was grainstone. This thin section strongly suffered from meteoric diagenesis, and some grains exhibit intermediate characteristics between grain and cement. Nonetheless, the trained model could correctly identify each category.

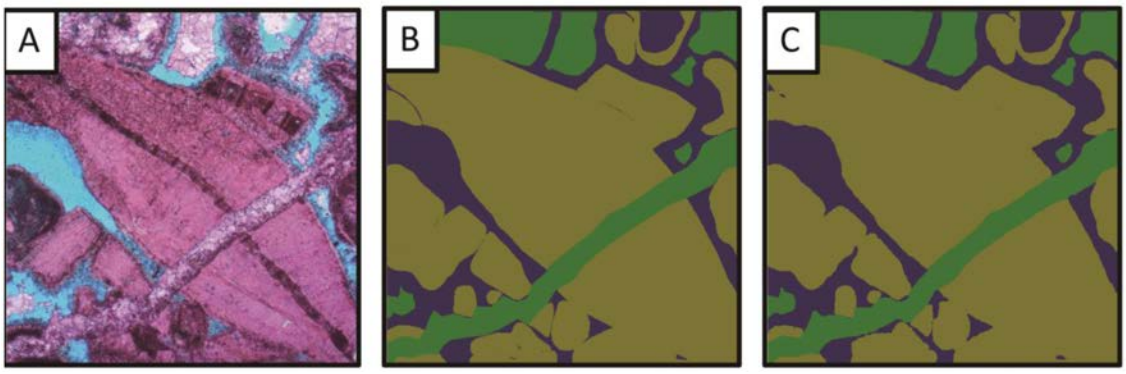

Figure 2—Trained model prediction of the highest sample. Green is cement, yellow is grain, and blue is pore. (A) Thin section, (B) ground truth, (C) trained model prediction.

The lowest average accuracy sample is shown in Fig. 3. The average accuracy, precision, recall, and mean IoU were 53.1%, 69.9%, 53.1%, and 40.5%, respectively. In this thin section, the grain occupies more than half of the thin section. The cement has clear boundaries with the grain and the pore. On the other hand, the grain has complicated internal textures. The accuracy of grain was 43.5 %, and the accuracy of cement and pore were 72.1% and 98.2%, respectively. Clearly, cement and pore have relatively higher accuracy than grain.

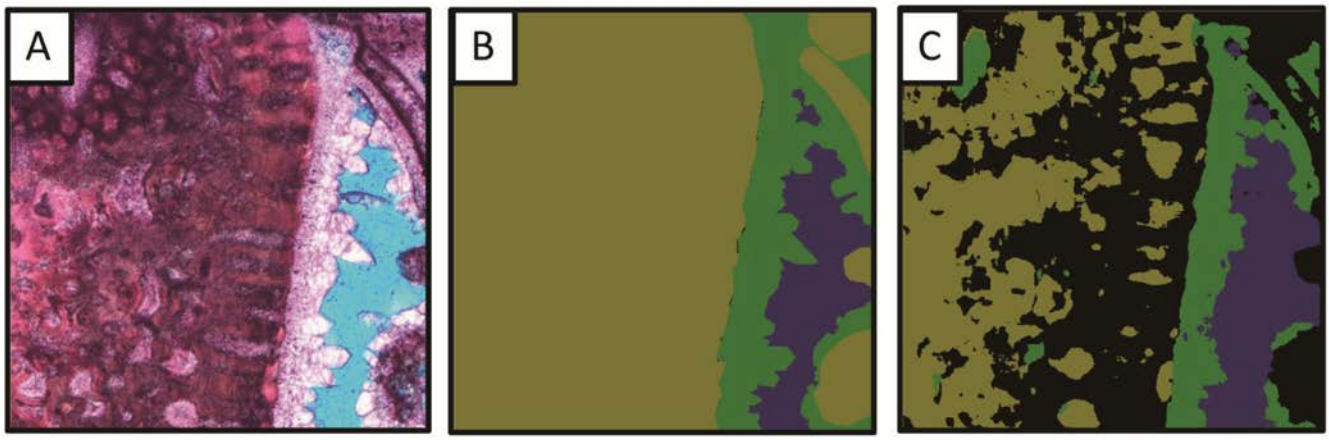

Figure 3—Trained model prediction of the lowest sample. Green is cement, yellow is grain, blue is pore, and black is limemud. (A) Thin section, (B) ground truth, (C) trained model prediction.

Fig. 4 and Fig. 5 depict examples of positive results. Fig. 4 shows the example of grainstone prediction by the trained model. The average accuracy, precision, recall, and mean IoU were 88.9%, 88.8%, 88.9%, and 78.3%, respectively. The ground truth and the trained model prediction are very similar. The boundaries of each area is predicted as clear as Fig. 2. Fig. 5 shows an example of packstone prediction by the trained model. The average accuracy, precision, recall, and mean IoU were 92.5%, 92.5%, 92.5%, and 82.9%, respectively. The skeletal frameworks were well preserved, and clear boundaries with other areas were also observed. The intraparticle pores of the grain were filled with limemud. The sedimentologist interpreted this as grain including fills of intraparticle; however, the trained model predicted the fill sediments in the intraparticle pores as limemud.

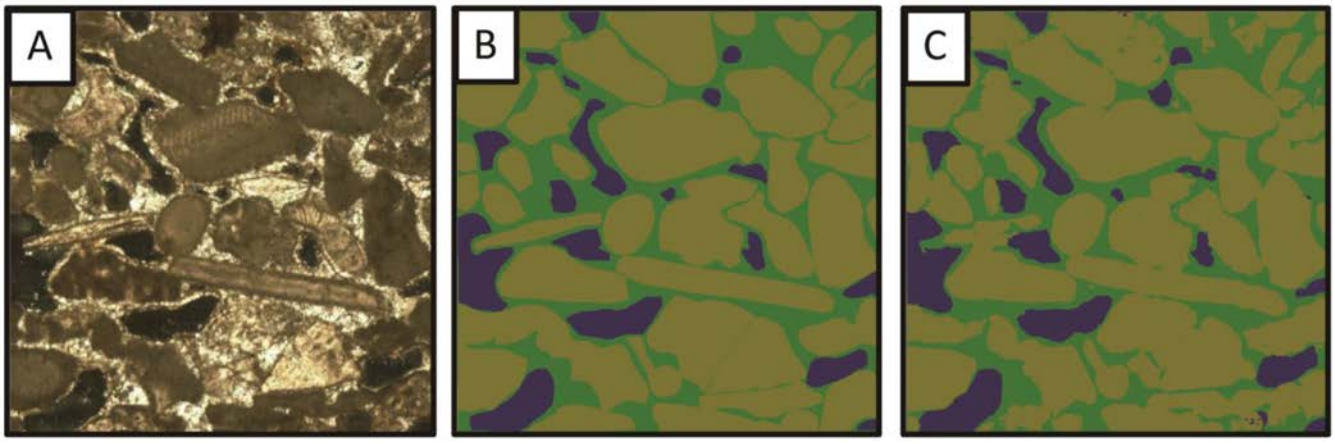

Figure 4—Trained model prediction of example of positive result. Green is cement, yellow is grain, and blue is pore. (A) Thin section, (B) ground truth, (C) trained model prediction.

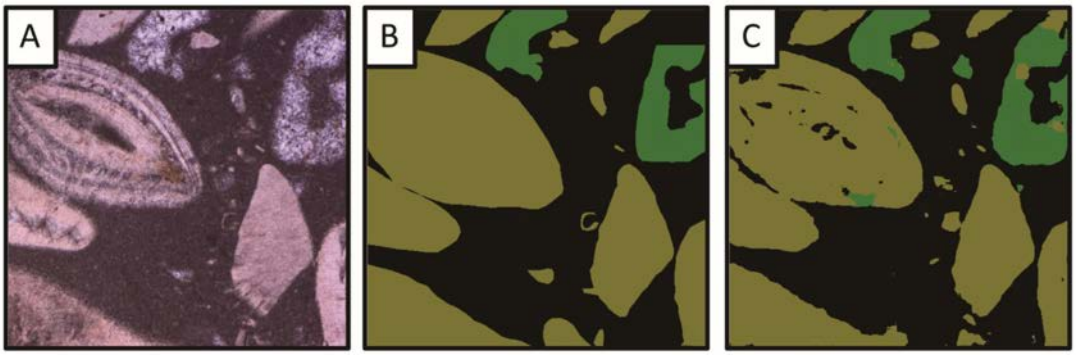

Figure 5—Trained model prediction of example of positive result. Green is cement, yellow is grain, and black is limemud. (A) Thin section, (B) ground truth, (C) trained model prediction.

Fig. 6 and Fig. 7 shows examples of poor results. In Fig. 6, the average accuracy, precision, recall, and mean IoU were 71.4%, 71.7%, 71.4%, and 49.1%, respectively. This thin section has uncertain boundaries and low contrast with each area. The two objects on the upper left are probably the fragments of grain; however, it was difficult to correctly identify these two grains even for the carbonate sedimentologist. The boundaries between the fragments of grain and limemud were ambiguous. The grain (benthic foraminifera) was observed around the center, and these boundaries between the grain and limemud was not clear. The trained model was also unable to detect this grain. In Fig. 7, the average accuracy, precision, recall, and mean IoU were 60.7%, 63.0%, 60.7%, and 43.1%, respectively. The grain occupied more than half of the thin section and the boundary between the grain and limemud was uncertain. In addition, the internal texture of this grain was similar to the texture of limemud.

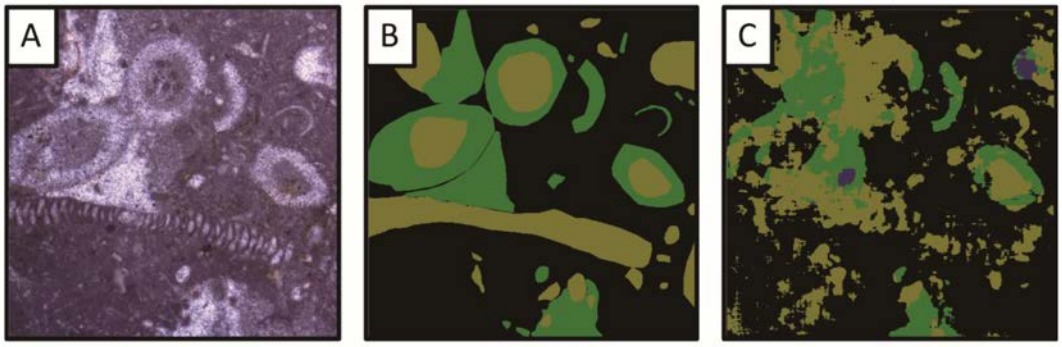

Figure 6—Trained model prediction of example of poor result. Green is cement, yellow is grain, blue is pore, and black is limemud. (A)Thin section, (B) ground truth, (C) trained model prediction.

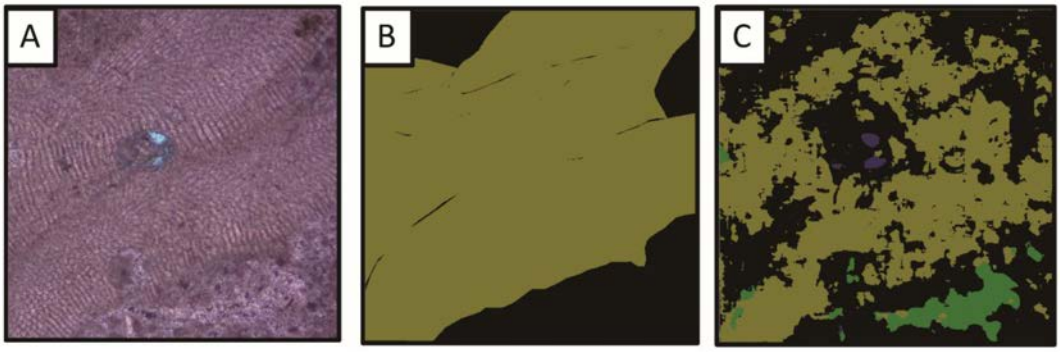

Figure 7—Trained model prediction of example of poor result. Green is cement, yellow is grain, blue is pore, and black is limemud. (A) Thin section, (B) ground truth, (C) trained model prediction.

Discussions

Accuracy of each category

The mean average accuracy of the entire test data was 83.9% (Fig. 1). The accuracy for each category also exhibited a high accuracy. The accuracies were as follows: grain was 80.9%, cement was 74.1%, pore was 85.0%, and limemud was 76.4%. The carbonate sedimentologist interpretation and the trained model prediction have a strongly positive correlation (Fig. 8). Therefore, the trained model of this study has a high degree of completion. However, the correlation of low limemud content samples is weak. There are some reasons for this. The first reason is the characteristic features of carbonate lithology. Low limemud content samples are usually rare, because grainstone essentially has a lack of limemud and packstone, wackestone, and mudstone have basically high limemud content. Hence, the reasons is a lack of supervised data. The second reason is the high difficulty level of interpretation. For example, consider the sample where a single grain covers most of picture (Fig. 7). This sample is low limemud content sample. The interpretation of this sample is difficult even for a carbonate sedimentologist, and these images usually need to be interpreted in an integrated manner with the other areas of the same thin section sample. The third reason is the definition of each category. In this study, limemud was defined as background. Limemud includes all areas except for grain, pore, and cement. Tiny limemud pixels sometimes appear in the grain area as noise, even without including limemud (Fig. 9).

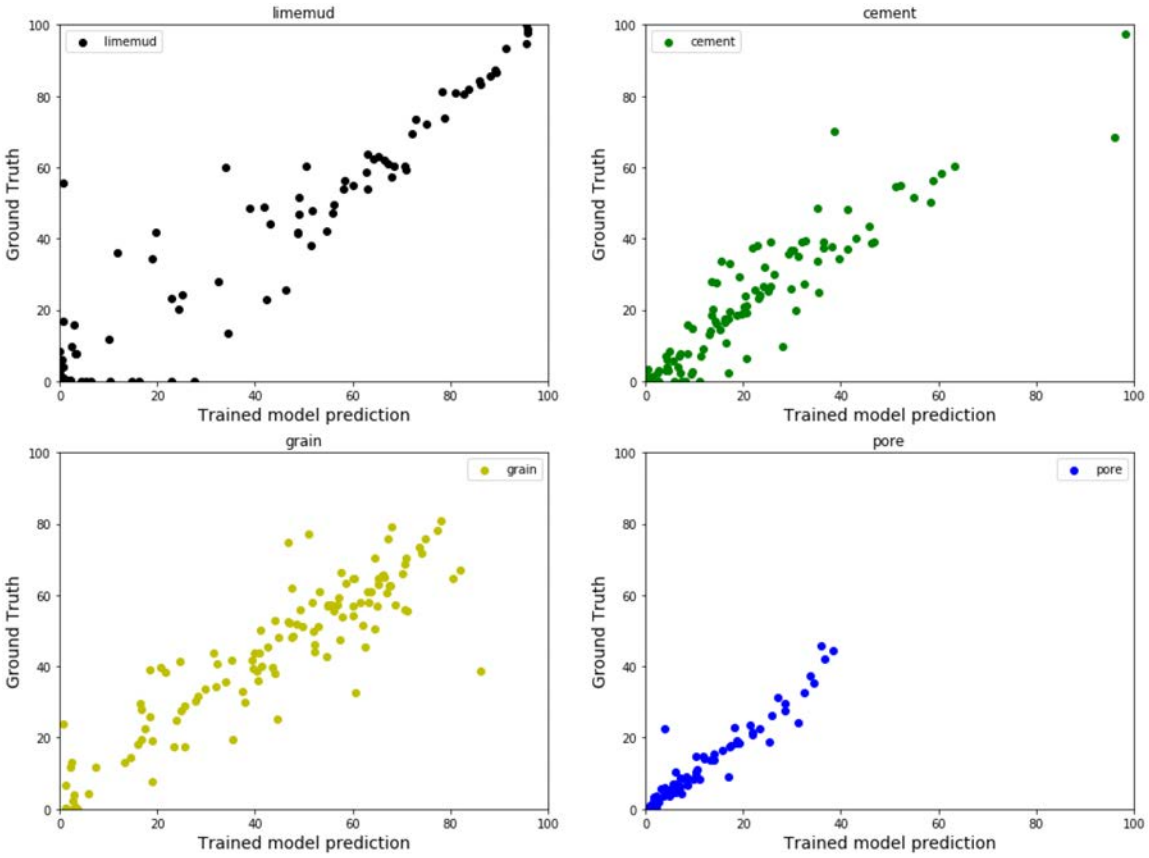

Figure 8—Cross plots of each category between the trained model and the ground truth.

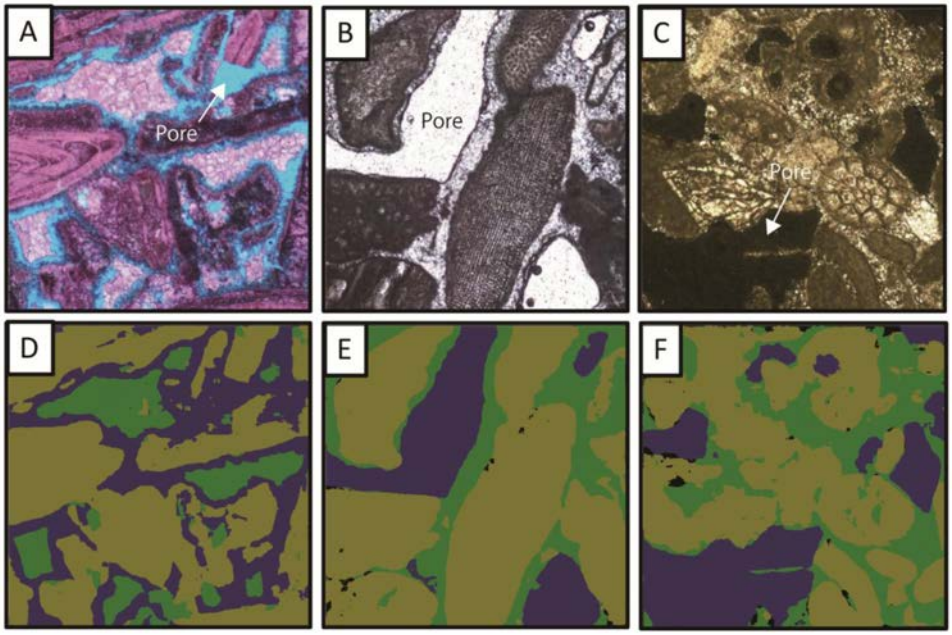

Figure 9—Variation of pores. Green is cement, yellow is grain, blue is pore, and black is limemud. (A) Filled by blue resin in open nichol, (B) open nichol, (C) cross nichol, (D, E, F) trained model prediction of the above thin section picture.

Grain and pore have higher accuracies than limemud and cement. Pores usually have high contrast compared with the other categories. In other words, the boundaries with the other areas are usually clear. Pores have three patterns: filled by blue resin, open nichol, and cross nichol (Fig. 9). The trained model correctly identifies these three pore patterns (Fig. 9). Therefore, the color has a low influence on category identification. A large number of skeletal grains and fragments of skeletal grains contain in the thin section pictures in the data set. They tend to form a clear boundary with the other areas, similar to pores. The main samples of this study were collected from the Ryukyu Group. The Ryukyu Group suffered from meteoric diagenesis after deposition (e.g. Matsuda et al. 1995; Nanjo et al. 2013). Therefore, the current textures of some samples are sometimes strongly influenced by cementation and dissolution. In particular, the boundaries between cement and limemud were often uncertain (Fig. 10). The average accuracy of the trained model prediction of the samples that have uncertain area boundaries were poor compared with the samples that have clear area boundaries (Figs. 2, 4, and 5). This indicates that clear area boundaries are the most important factor for correct category identification by using machine learning.

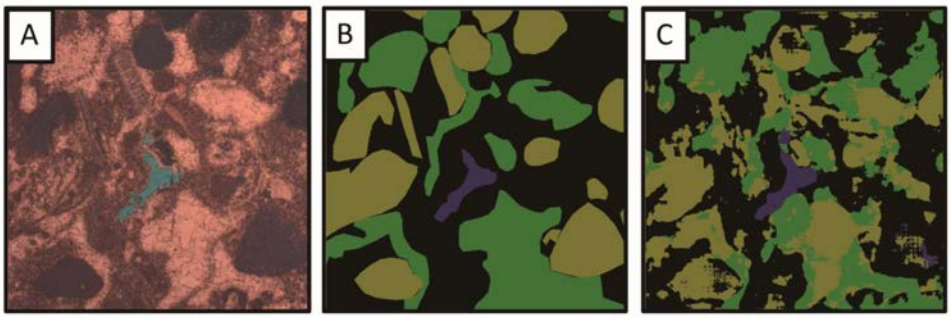

Figure 10—Uncertain boundary sample. Green is cement, yellow is grain, blue is pore, and black is limemud. (A) Thin section picture, (B) ground truth, (C) trained model prediction. The boundary of each area is uncertain. The trained model prediction was not acceptable.

Differences and similarities between a carbonate sedimentologist and the trained model

The significant difference between the interpretations of the trained model and a carbonate sedimentologist is in the inference method. A carbonate sedimentologisit interpret the thin section pictures, deductively. Therefore, the carbonate sedimentologist can interpret the thin section pictures even a little number of thin section pictures. On the other hand, machine learning predict the thin section pictures, statistically. Therefore, a large number of thin section pictures and supervised images are needed to achieve high accuracy prediction by using machine learning.

The detailed definitions of interpretation differ subtly among carbonate sedimentologists. For example, the grain in Fig. 5 was interpreted as a grain that included the fill sediments (limemud) in the intraparticle pore. However, another carbonate sedimentologist may interpret the fill sediments (limemud) in the intraparticle pore as limemud. In contrast, the trained model interprets the carbonates in the thin section pictures at the pixel scale. The grain in Fig. 5 was separately interpreted as grain and limemud, because the machine provides the interpretation at the pixel level. Hence, the prediction of the trained model is more consistent and accurate than the interpretation of the carbonate sedimentologists.

High accuracy and low accuracy samples in the trained model exhibit a clear trend. The samples that have clear boundaries with the other area have a high accuracy. In contrast, the complicated samples (e.g. those that have suffered from heavy diagenesis) have low accuracy. In other words, the samples that the carbonate sedimentologist can easily interpret are also easily interpreted by the trained model. The carbonate sedimentologist cannot provide a 100% accurate interpretation for a thin section picture at the pixel level. Thus, the trained model prediction outperforms that of the carbonate sedimentologist at the pixel level interpretation. This fact clearly indicates that the interpretation level of the trained model will surpass that of human specialist in the future. However, the trained model is not a copy of a carbonate sedimentologist, but acts as another carbonate specialist.

Improving the accuracy of the trained model

The mean average accuracy of the trained model was 83.9%. There are two approaches to improve the accuracy apart from network improvement: increase the number of supervised data and develop higher quality supervised data. Both approaches are significant, but generating higher quality samples is more important than increasing the number of supervised data. A deep neural network usually learns based on the supervised data. Therefore, regardless of increases in the number of supervised data, if the quality is low, the neural network will learn the wrong answer. In this study, the carbonate sedimentologist interpreted the fill sediment in the intraparticle pore as grain; in contrast, the trained model predicted it as limemud. Dunham (1962) doesn't specify whether intraparticle pores and fill sediments are included in grain. Therefore, the detailed definition of interpretation of carbonate differ subtly among carbonate sedimentologists. To achieve high accuracy automatic carbonate identification with machine learning, the detailed Dunham standard of carbonate classification at the pixel level is required.

Auto Interpretation

The lithology of Dunham classification is defined by depositional texture, grain content and lime mud content. Therefore, the auto interpretation were conducted based on the definition of Dunham classification. Fig. 11 and Fig. 12 shows the result of auto interpretation. The average accuracy of the trained model prediction of Fig. 11 was 81.8%, and the trained model and carbonate sedimentologist both interpreted the sample as packstone. Note that the trained model perfectly interpreted the thin section.

Figure 11—Interpretation of carbonate sedimentologist vs the trained model prediction. Green is cement, yellow is grain, blue is pore, and black is limemud. (A) Thin section picture, (B) interpretation of carbonate sedimentologist, (C) trained model prediction.

The same process was performed for all predicted samples. The accuracy of Dunham classification prediction by the trained model compared with the carbonate sedimentologist's interpretation was 90.6%. The confusion matrix shows the correct and incorrect number of prediction (Fig. 13). Compared with the carbonate sedimentologist interpretation, the main failure in lithology interpretation by the trained model were for packstone and wackestone. For example, in Fig. 12, the carbonate sedimentologist interpreted the sample as packstone, and the trained model predicted the sample as wackestone. Comparing the carbonate sedimentologist's interpretation and the trained model prediction for each category, both interpretations were similar; however, the results of lithology interpretation were different. The carbonate sedimentologist interpreted the fill sediments in the intraparticle pore as grain; however, the trained model predicted them as pore and limemud. Therefore, the trained model tends to overestimate limemud compared with the carbonate sedimentologist. In addition, this thin section is located on the border of packstone and wackestone. It is difficult to distinguish samples on the border, even for the carbonate sedimentologist. This image usually need to be interpreted in an integrated manner with the other areas of the same thin section sample. Therefore, the trained model can correctly predict the carbonate lithology based on the Dunham classification with higher accuracy.

Figure 12—Interpretation of carbonate sedimentologist vs the trained model prediction. Green is cement, yellow is grain, blue is pore, and black is limemud. (A) Thin section picture, (B) interpretation of carbonate sedimentologist, (C) trained model prediction.

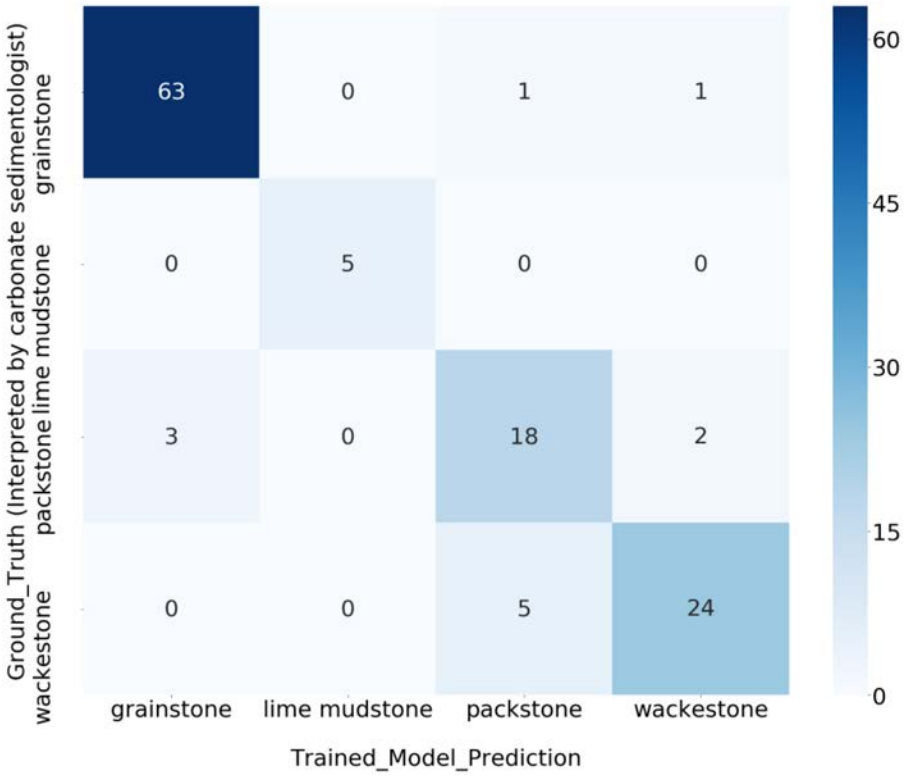

Figure 13—Confusion matrix.

Summary

This study developed automatic carbonate lithology identification with a deep neural network. The results of this study indicated that the accuracy of automatic Dunham classification exceeded 90%, and the mean accuracy of category identification was approximately 85%. In addition, the trained model partly surpassed the human interpretation. This indicates that it has sufficient accuracy to assist geologists and become a basic model for field applications. Therefore, this trained model has potential applications in carbonate projects as a geologist assist tool.

Some problems exist to be solved for completely automatic identification, including reservoir quality. The most important aspect is the accuracy of detection for grain, pore, cement, and limemud. As mentioned previously, our final objective is not automatic carbonate lithology identification, but understanding the correlation between porosity, permeability, and each category's area percentage using a deep neural network. The accuracy of category identification is still insufficient to achieve this objective. The authors believe that the model requires higher quality supervised data and a greater number of supervised data. To develop higher quality supervised data, the authors also need to consider the detailed standard of carbonate analysis (e.g. intraparticle pore, fill sediments). The authors see massive potential in carbonate lithology identification via machine learning. The architecture has been progressing continuously, and the authors also consider using the latest architectures such as generative adversarial network (GAN) in automatic carbonate lithology identification.

Carbonate records the information of the paleo depositional environment through its lithology and grain types. In addition to lithology identification, the authors will tackle grain type identification via machine learning in the future.

Acknowledgments

The authors gratefully acknowledge the work of the past and present members of the Geology and Geophysics division and Digital technology team of JOGMEC. The authors would also like to thank

Hirokazu Tada and Yoshikazu Yaguchi of the Information Center for Petroleum Exploration and Production for technical assistance with the labeling of supervised data. The authors are grateful to Takashi Shimazu for helpful discussions. The authors would like to thank Editage (www.editage.com) for English language editing.

Reference

Alqahtani, N., Armstrong, R. T., and Mostaghimi, P. 2018. Deep Learning Convolutional Neural Networks to Predict Porous Media Properties. SPE Asia Pacific Oil and Gas Conference and Exhibition, Brisbane, Australia, 23-25 October. SPE-191906-MS. https://doi.org/10.2118/191906-MS.

Alaudah, Y., Michalowicz, P., Alfarraj, M. et al 2019. A Machine Learning Benchmark for Facies Classification. *Interpretation* **7**(3): 1–51. http://dx.doi.org/10.1190/int-2018-0249.1.

Dunham, R. J., 1962. Classification of carbonate rocks according to depositional texture. In Classification of carbonate rocks: A Symposium. *American Association of Petroleum Geologists Memoir 1*, ed. Ham, W. E., 108–121. Tulsa, Oklahoma: AAPG.

Du, M., Yin, H., Chen, X. et al 2018. Oil-in-Water Two-Phase Flow Pattern Identification From Experimental Snapshots Using Convolutional Neural Network. *IEEE* Access **7**: 6219–6225. http://dx.doi.org/10.1109/ACCESS.2018.2888733.

Fernández, G., Bulnes, J. M., Llorca, D. F. et al 2018. High-Level Interpretation of Urban Road Maps Fusing Deep Learning-Based Pixelwise Scene Segmentation and Digital Navigation Maps. *Journal of Advanced Transportation* **2018**: 1–15. https://doi.org/10.1155/2018/2096970.

Liu, Y., Gadepalli, K., Norouzi, M. et al 2017. *Detecting Cancer Metastases on Gigapixel Pathology Images*. arXiv:1703.02442 [cs.CV].

Krizhevsky, A., Sutskever, I., and Hinton, G. E. 2012. ImageNet Classification with Deep Convolutional Neural Networks. NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, 3-6 December. **1**: 1097-1105. http://dx.doi.org/10.1145/3065386.

Matsuda, H., Tsuji, Y., Honda, N. et al 1995. Early Diagenesis of Pleistocene Carbonates from a Hydrogeochemical Point of View, Irabu Island, Ryukyu Islands: Porosity Changes Related to Early Carbonate Diagenesis. In Unconformities and Porosity in Carbonate Strata. *American Association of Petroleum Geologists Memoir 63*, ed. Budd, A. B., Saller, H. A., Harris, M. P., 35–54. Tulsa, Oklahoma: AAPG. https://doi.org/10.1306/M63592.

Mahdi, T.A., Aqrawi, A.A.M., and Horbury, A.D. et al 2013. Sedimentological characterization of the mid-Cretaceous Mishrif reservoir in southern Mesopotamian Basin, Iraq. *GeoArabia* **18** (1): 139–174.

Moor, T., Ruiz, A. R., Mérida, A. G. et al 2018. *Automated soft tissue lesion detection and segmentation in digital mammography using a u-net deep learning network*. arXiv:1802.06865v2 [cs.CV].

Nanjo, T., Sasaki, K., and Matsuda, H. 2013. Sea-level changes of approximately 62-52 ka reconstructed from subaerial exposure surface based on carbon and oxygen isotopic compositions. *Jour. Geol, Soc*, Japan **119** (3): 155–170. https://doi.org/10.5575/geosoc.2012.0079.

Saporetti, C. M., Fonseca, L. G., Pereira, E. et al 2018. Machine learning approaches for petrographic classification of carbonate-siliciclastic rocks using well logs and textural information. *Journal of Applied Geophysics* **155**: 217–225.

Siam, M., Elkerdawy, S., Jagersand, M. et al 2017. Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges. 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16-19 October. http://dx.doi.org/10.1109/ITSC.2017.8317714.

Zeng, Y., Jiang, K., and Chen, J. 2019. Automatic Seismic Salt Interpretation with Deep Convolutional Neural Networks. ICISDM 2019 Proceedings of the 2019 3rd International Conference on Information System and Data Mining, Houston, TX, USA, 6-8 April. http://dx.doi.org/10.1145/3325917.3325926.