Marine and Petroleum Geology 132 (2021) 105159

Contents lists available at ScienceDirect

Marine and Petroleum Geology

journal homepage: www.elsevier.com/locate/marpetgeo

Detection of framboidal pyrite size distributions using convolutional neural networks

Artur Davletshin[a], Lucy Tingwei Ko[b], Kitty Milliken[b], Priyanka Periwal[b], Chung-Che Wang[b], Wen Song[a,*]

[a] Department of Petroleum and Geosystems Engineering, University of Texas at Austin, Austin, TX, 78712, USA
[b] Bureau of Economic Geology, University of Texas at Austin, Austin, TX, 10611, USA

ARTICLE INFO

Keywords:
Pyrite framboids
Marine anoxia
Machine learning
Convolutional neural networks
Feature extraction
Object detection
Instance segmentation

ABSTRACT

Pyrite (FeS_2) framboids, spheroidal groups of discrete equant pyrite microcrysts, are found in sediments of all geological ages. The size of a pyrite framboid is established during early diagenesis and preserved through time. Framboid size distributions are hence useful for the evaluation of depositional conditions. In this work, we present machine learning approaches to characterize the size distributions of pyrite framboids to understand the intensity and duration of anoxia and euxinia during the Middle Devonian of the Appalachian foreland basin by analyzing framboid size distributions of the Marcellus Shale from Lycoming County, Pennsylvania. Importantly, we overcome the time-consuming and laborious nature of current manual tracing methods to enable the processing of high volumes of micrograph data. Specifically, we implement convolutional neural networks (CNNs) to characterize framboids from 14 samples across depths in the Marcellus Shale. We show that CNNs enable the precise and fast measurement of framboid size distributions from scanning electron micrographs. CNN architectures including Inception, ResNet, Inception-Resnet, and Mask R-CNN were trained and tested on a total of ~6,800 framboids from 128 grayscale and 32 colored scanning electron micrographs. Kolmogorov-Smirnov tests on the framboidal equivalent diameter distributions measured from CNNs and manual tracing show that the CNN algorithms detected framboids with up to 99% precision. Importantly, once trained, the CNNs were ~100 times faster than current manual tracing. A straightforward extension of this work includes the application of CNNs to characterize pores, fractures, organic matter, and/or mineral grains in geological materials.

1. Introduction

Marine anoxia, a condition where waters are depleted of oxygen, is one of the “Big Five” mass extinction mechanisms proposed. Importantly, marine anoxia is believed to have caused the End-Permian Extinction (~252 Ma) that eliminated ~ 95% of marine species (Raup and Sepkoski, 1982; Sepkoski, 1996; Wignall and Twitchett, 1996; Payne et al., 2004; DiChristina et al., 2006; Meyer et al., 2008; Brennecka et al., 2011; Lau et al., 2016). To understand the paleo-redox change in geologic history, one approach is to analyze the characteristics of pyrite (FeS_2) framboids in mudstones. Specifically, the sizes of pyrite framboids are dictated by local anoxia and euxinia (i.e., low oxygen and high hydrogen sulfide, H_2S, level) conditions during their formation and are preserved in time thereafter (e.g., Wilkin et al., 1996, 1997; Wignall and Newton, 1998; Bond and Wignall, 2010; Blood and Lash, 2015; Huang et al., 2017). The use of pyrite framboids as a geochemical proxy for anoxia/euxinia, therefore, is advantageous due to its relative independence from grain sizes and subsequent geochemical activity (Wilkin et al., 1996, 1997), however pyrite framboid sizes may also be controlled by other factors including sedimentation rates that influence the nucleation and growth times of framboids formed in the sediment (e.g., Gallego-Torres et al., 2015).

Pyrite framboids are clusters of discrete equant pyrite microcrysts that are arranged into a spherical or spheroidal/ellipsoidal external structure (Love, 1966) Rickard, 1970). Framboidal pyrites are distinct from other textures of pyrite (i.e., fibrous, radiating, nodular, and euhedral pore-filling or replacement forms) by their equant microcrysts, external symmetry, and geometrical regularity (Rickard, 1970). The existence of equant microcrysts suggests simultaneous nucleation and constant growth rates prior to aggregation.

Sedimentary pyrite formation is controlled by the concentration of dissolved sulfate, the rate of organic matter decomposition (i.e., degree

* Corresponding author.
E-mail address: wensong@utexas.edu (W. Song).

https://doi.org/10.1016/j.marpetgeo.2021.105159
Received 3 November 2020; Received in revised form 5 May 2021; Accepted 24 May 2021
Available online 29 May 2021

of anoxia/euxinia), and the supply of reactive detrital iron minerals (Berner, 1970, 1984). Specifically, pyrite formation is enabled by sulfate-reducing micro-organisms that reduce sulfate (SO_4^{2-}, electron acceptor) into hydrogen sulfide/bisulfide (H_2S/HS^-) in anoxic and/or euxinic sediments. Characterization of pyrite framboid size distributions, therefore, provides information about the degree of anoxia and euxinia in the depositional environment. In oxygen-deprived depositional environments (anoxic or euxinic conditions), framboid formation is transport limited (i.e., framboids are formed in the water column and/or sediments and sink below the iron reduction zone where growth is stopped) and the framboids are limited to small diameters ($\lesssim$ 5–6 μm) (Bond and Wignall, 2010; Blood and Lash, 2015). On the other hand, in oxygen-available environments (e.g., oxic or dysoxic conditions), the formation of pyrite framboids is reaction limited (i.e., formed near the water-sediment interface where their sizes are governed by locally available reactants) and the resulting pyrite framboids are larger and more variable in size (Wilkin et al., 1996; Wilkin and Barnes, 1997; Suits and Wilkin, 1998; Wei et al., 2012).

To delineate changes in depositional conditions, variations in framboid size distributions (e.g., mean, minimum, maximum, standard deviation, and skewness of the framboid diameters) are characterized in sedimentary succession (Wilkin et al., 1996, 1997; Wignall and Newton, 1998; Huang et al., 2017; Bond and Wignall, 2010; Blood and Lash, 2015; Zhou and Jiang, 2009). Current methods to quantify pyrite framboid size distributions in sedimentary materials include (i) centrifugal separation of pyrites from the sample, and (ii) manual tracing of pyrite framboids from micrographs. Centrifugation yields only ~50–70% of the total pyrite in the sample when benchmarked to results from sequential extraction and therefore does not provide a representative sample for analysis (Wilkin et al., 1996; Shen et al., 2007). Image analyses of micrographs from optical light microscopy and scanning electron microscopy (SEM) yield up to 100% of the framboids available in the cross-section (Wilkin et al., 1996, 1997; Wang et al., 2013; Karduck, 2015; Blood and Lash, 2015; Wignall and Newton, 1998; Bond and Wignall, 2010; Shen et al., 2007; Zhou and Jiang, 2009; Huang et al., 2017). While framboid diameters traced from micrographs are subject to underestimates of the true diameter due to uncertainty in the cutting plane of the framboid spheroid, deviations are limited to $\lesssim$ 10% if more than 100 framboids are measured (Wilkin et al. (1996)). As a result, image analysis of pyrite framboids is the predominant method that is used to characterize depositional conditions (e.g., Permian–Triassic boundary in equatorial eastern Paleotethys of China (Shen et al., 2007), eastern Greenland (Wignall and Twitchett, 2002), the Boreal shelf seas of Spitsbergen, Greenland and mid-latitude Neotethyan oceans of Western Australia (Bond and Wignall, 2010), the eastern Panthalassan margin of Idaho, United States (Bond and Wignall, 2010), southwestern Japan (Isozaki, 1997), Sosio valley (Wignall and Twitchett, 2002), Italy (Wignall and Twitchett, 1996), Slovenia (Wignall and Twitchett, 1996), Kashmir (Wignall et al., 2005), the Devonian shale (e.g., Schieber and Baird, 2001; Bond et al., 2004; Formolo and Lyons, 2007; Marynowski et al., 2007; Blood and Lash, 2015)).

Identification of pyrite from other minerals in SEM is enabled by its high back-scattered electron (BSE) intensity. Current methods to identify framboid sizes from SEM images, however, are limited to manual and semi-automated tracing that is labor- and time-intensive, and results in a small sample of framboids analyzed. Limitations in the size of the dataset (i.e., number of framboids analyzed) reduces the accuracy and the volume of information generated and thus limits the ability to determine the sedimentary redox conditions. In addition, human biasing (e.g., bias towards "better" samples) also influences data collection.

To generate larger datasets of framboid sizes that are immune to human biasing in a timely manner, automated image processing is favorable. Current semi-automated image processing methods identify pyrite from SEM images by leveraging its high BSE intensity through techniques including histogram thresholding, K-means clustering, and watershed segmentation (Roduit,), but are limited in their applicability. Specifically, variations in instrument settings (e.g., magnification, electron energy, and imaging aperture) introduce differences in background image intensities and, therefore, differences in the intensities of framboids. Variation in framboid intensities requires modifications in thresholding parameters, and therefore weakens the ability to automate simple threshold-based image processing workflows. Currently, thresholding parameters are selected by human operators that are prone to subjective bias and inconsistent parameter selection. Human-based image processing of geological materials are therefore impractical, and a more robust and automatic method is required to characterize geological features.

Machine learning (ML) is a state-of-the-art method that enables the automated extraction of geological features from micrographs. Favorably, ML methods are fast, robust, and optimized to process high volumes of data. Both supervised and unsupervised learning techniques have been applied to estimate porosity from thin sections (Richa et al., 2006). Koeshidayatullah et al. (2020) measured porosity of carbonate thin section materials with a fully automated DCNN based object detection method. Tang and Spikes (2017) presented a workflow for the semantic segmentation of SEM and EDS elemental maps, and Tang et al. (2020) extended the application of machine learning to point counting and segmentation of arenite in thin sections. Convolutional neural networks have been applied to SEM images at the microscale (Anderson et al., 2020; Ikeda et al., 2019). Supervised machine learning techniques including Classification and Regression Trees (CART), k-Nearest Neighbor (k-NN), and Random Forest (RF) have been applied to segment geological materials, including sand grain recognition (Maitre et al., 2019). Wu et al. (2019) presented a machine-learning-assisted workflow that involved feature extraction to identify five components from grayscale images: pores/cracks, kerogen, calcite, pyrite, and rock matrix. Further, Tian and Daigle (2019) characterized shale microstructures grayscale BSE images and EDS elemental maps using machine learning. Deep learning-based methods have also been applied by Chen et al. (2020) where a U-net architecture was used for semantic segmentation of clay aggregates and for mineral classification.

In this work, we propose a robust ML method to extract framboid features and to estimate their statistical characteristics using two classes of convolutional neural networks: object detection and instance segmentation. Pyrite framboids make an excellent test case for ML-based image processing due to its intensity in SEM images and its composition (Fe, S) in EDS images. Specifically, elemental distributions derived from EDS images are useful in distinguishing mineral features with similar degrees of intensity in SEM images. Characterization of framboid size distributions allows the delineation of depositional redox history in geological records. We compared the performance of several object detection models to estimate the accuracy of each method using metrics including precision, recall, and F1 scores. The presented methods extract the size distributions of the framboids with up to 0.99 precision and 0.78 recall. To extract the size of identified framboids, we used a post-processing Otsu binarization method. Lastly, to achieve additional flexibility for post processing, we present a comparison with an instance segmentation method and find a similar precision in its identification and measurement of pyrite framboids.

2. Methods

Pyrite framboids in shales from the Middle Devonian Marcellus Formation in the Appalachian foreland basin were assessed to delineate marine redox conditions at the time of deposition. Image acquisition, image processing by manual tracing, and the application of convolutional neural networks (i.e., object detection and instance segmentation architectures) to extract pyrite framboid size distributions automatically are described. For the purpose of understanding local paleo-redox histories, we traced only those pyrite framboids that were intact and otherwise unaltered (i.e., spherical and ellipsoidal). All other fragmented, dissolved, or altered framboids were not included in the present study. The method presented here is general, however, and can be modified to characterize additional classes of framboid geometries including fragmented and dissolved framboids.

2.1. SEM image acquisition and analyses

This study examined core samples from the Hamilton #1H Marcellus Shale of Lycoming County, Pennsylvania. Sixteen samples were selected on the basis of different lithofacies for micro-petrographic studies on texture, fabric, types of organic matter, mineral assemblage, and diagenesis (Ko, 2019). Specifically, core vertical heterogeneities were characterized using integrated wireline logs, high-resolution X-ray fluorescence (every 2 inches), and 69 thin section samples. Out of the 69 thin sections, 16 samples were selected for framboid size distribution characterization. All samples were cut normal to the bedding plane. The image data contain ~6,800 pyrite framboids of interest that were used to train and test the CNNs. To minimize the influence of geologic heterogeneity, analysis of a larger volume of pyrite framboid data is useful and well-suited for machine learning approaches. Each sample was Ar-ion milled (Leica TIC 020 Triple Ion beam) for 10 h using an accelerating voltage of 8 keV and a current of 2.8 mA. The Ar-ion milled samples were imaged using a high-resolution field emission scanning electron microscope (FE-SEM, FEI Nova NanoSEM 430). The FESEM is equipped with two 30-mm^2 Bruker XFlash silicon-drift EDS detectors for elemental identification.

Two SEM datasets were obtained for framboid size characterization (Fig. 1a). The first dataset (Fig. 1a, Dataset 1: BSE/SE/TLD) contained 128 grayscale BSE, secondary electron (SE), and SE through-the-lens detector (TLD) images. The grayscale BSE/SE/TLD images were collected at instrument magnifications of 459X to 70,000X (horizontal field width HFW = 4.26 μm to 651 μm), accelerating voltages of 2 to 15 keV, and spatial resolutions of 1024 × 880 pixels (0.004–0.45 μm per pixel). The magnifications were chosen to determine the influence of pixel size on framboid size characterization (e.g., skew toward larger framboids due to pixelation at low magnifications). The second dataset (Fig. 1a, Dataset 2: BSE/EDS) comprised 32 colored images of iron (Fe) and sulfur (S) EDS maps overlaid on BSE images. The colored BSE/EDS maps were acquired at an accelerating voltage of 15 keV with spot sizes between 4.0 and 5.0 and total count times of t > 800 s.

2.2. Ground truth labelling

To serve as a testing benchmark for the convolutional neural networks (CNNs), a total of ~6,800 framboids from the BSE/SE/TLD and BSE/EDS micrographs were traced manually. The manually traced framboids served as ground truths to train and test the CNNs. Framboids were traced by two different geologists to minimize human biases. Statistics on human error are included in the supplementary materials (see SM Appendix A).

2.3. Automated image processing workflow

Two different CNN workflows were used to measure framboid size distributions (Fig. 1b): (i) an object detection-based method that proposes small regions of interest that contain target features (i.e., framboids), and (ii) an instance segmentation method that identifies each instance of the object of interest (i.e., framboids). The two workflows were trained and tested using the two datasets with a train/test split of ~75%/25% for the grayscale BSE/SE/TLD images (95 images in training set, 33 images in testing set) and a train/test split of 81%/19% for the colored BSE/EDS images (26 images in training set, 6 images in testing set). We chose the train/test split based on the number of images available and the number of framboids contained within the images to achieve a split that is close to the 75/25% split that is used classically. Specifically, we chose this split to ensure a high enough training and testing data volume to minimize data variances (Perez and Wang, 2017). TensorFlow API (v 1.13.1) was run on a Windows-based computer (Intel Core i7, 4.20 GHz, 32 GB RAM and Nvidia GeForce GPU P4000 with 8

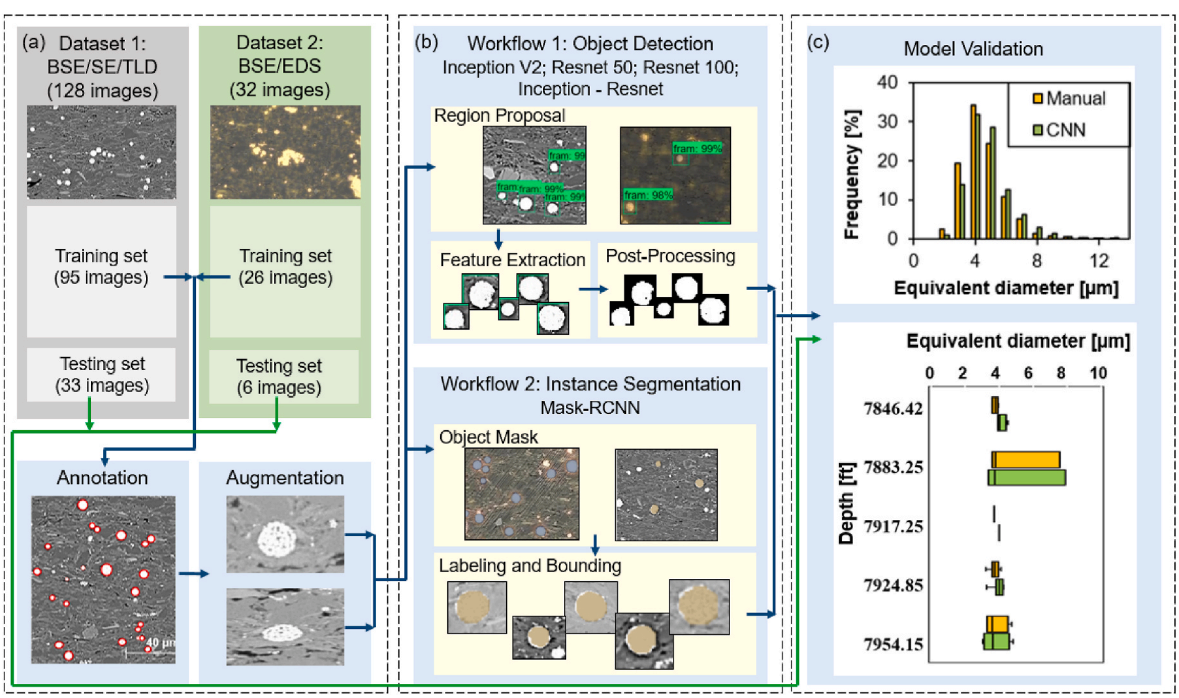

Fig. 1. Overview of the two workflows used in this work. (a) Two datasets are prepared and used. Dataset 1 comprises 128 grayscale BSE/SE/TLD images and is divided into a training set (95 images) and a testing set (26 images). Dataset 2 comprises 32 colored BSE/EDS images and is divided into a training set (26 images) and a testing set (6 images). The data are annotated and augmented (see Table 1) prior to training. (b) ML-based workflows used to measure framboid size distributions. In Workflow 1, object detection proposes likely regions of framboids, extracts the objects of interest, and post-processes the objects into binary images. In Workflow 2, instance segmentation labels, bounds, and masks each instance of a framboid. (c) Model is validated by comparing CNN results with manually traced framboid size distributions.

GB memory) to apply CNN models to process the images.

To improve the performance of the workflows, variability was introduced to the training dataset by (i) varying the imaging conditions (e.g., magnification, energy, etc.) of the BSE/SE/TLD and the BSE/EDS images, and (ii) augmenting the training sets to generate additional data for training (1408 BSE/SE/TLD and 352 BSE/EDS images in total). Augmentation strategies were applied to ~ half of the training datasets, and used methods including random cropping, flipping, contrast normalization, addition of Gaussian noise, multiplying, and perspective transformation (Table 1) to increase the volume of training data available. Importantly, high volumes of training data enable increased precision and recall of the CNN architectures (Perez and Wang, 2017). The image augmentation strategy here produced over 700,000 framboid instances for training and testing from an original data of ~6800 framboid instances.

2.3.1. Faster R–CNN object detection

The object detection model used in this work, Faster R–CNN, is comprised of two separate modules. The first module is a deep fully convolutional network (FCN) that proposes regions that may contain framboids (i.e., region proposal network, RPN), and the second is a Faster R–CNN detector that uses the FCN proposed regions to determine the likelihood of a framboid in the proposed region. The two modules work together as a single system and provides advantages in performance time and in improvements to the accuracy of proposed region within the image (Ren et al., 2016). Although other approaches are possible, including YOLO and Detectron2, Mask R–CNN achieves the highest precision and recall for complex systems such as geologic media. In particular, YOLO is excellent for real-time detection (e.g., in self-driving vehicles), but geologic materials do not generally evolve quickly in time.

2.3.1.1. Region proposal network. In each image, possible framboid objects are suggested by an RPN. To generate a set of proposed regions, a small network translates and extracts possible regions of interest using a shared CNN. Redundant candidate bounding boxes were eliminated using non-maximum suppression. Each RPN sliding network uses size-varying spatial windows of the input image. All windows are down-sized to lower-dimensional features that are fed to fully connected layers (i.e., box-regression and box classification layers). The candidate bounding boxes are then separated by a Region of Interest (RoI) layer. The final output is acquired by the Softmax multi-classification function and associated with a category label (Ren et al., 2016).

2.3.1.2. Feature extraction. Accurate feature extraction of the target regions with framboids is the most important step in the automation process. CNN models including Inception V2, ResNet-50, ResNet-101, and Inception-ResNet-v2 were tested as candidate feature extraction networks for deep learning. The specific features of each architecture are introduced and explained below.

ResNet was proposed by He et al. (2016) to ease the training of deep networks. In ResNet, residual learning frameworks use reformulate layers as learning residual functions with reference to layer inputs, and improves the capacity to optimize and improve the accuracy of the architecture by increasing its depth.

Inception was introduced by Ioffe and Szegedy (2015) and aims to leverage increased computational costs that are associated with processing high volumes of data. Inception makes the convolution network wider rather than deeper. Inception V2 factorizes convolutions into smaller convolutions, and thus increases computational power and speed of the feature extraction process.

Inception-ResNet is a hybrid module (Fig. 2) where training decreases significantly with residual connections (Szegedy et al., 2016a,b). Single frame recognition is improved on the ILSVRC 2012 classification task using Inception-ResNet. During training, images are propagated through the CNN architecture, and visual data are transformed into intermediate feature maps (Fig. 3). The intermediate feature maps are used to understand and improve the performance of CNN architectures.

2.3.1.3. Loss function for object detection. To monitor the training process and to minimize method error, the total loss function of the RNP, L ({p_i},{t_i}), was calculated by summing the classification and localization losses (Ren et al., 2016):

$$L\left(\{p_i\},\ \{t_i\}\right)=\frac{1}{N_{cls}}\sum_i L_{cls}\left(p_i,p_i^*\right)+\lambda\frac{1}{N_{loc}}\sum_i p_i^* L_{loc}\left(t_i,t_i^*\right) \tag{1}$$

where *i* is the index of each reference box (i.e., anchor), *p* is the predicted probability that the reference box is a framboid object, and p^* is a ground truth label where $p^* = 0$ for positive anchors *i* and $p^* = 1$ for negative *i*. Further, *t* is the coordinate vector of the predicted bounding box, t^* is the coordinate vector of the ground truth box, λ is a balancing parameter, N_{cls} is the size of the mini batch, and N_{loc} is the number of anchor locations. The total classification loss, $L_{cls,T} = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*)$, shows how accurately the CNN can predict the labels (i.e., whether the instance is a framboid or background), and the total localization loss, $L_{loc,\ T} = \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{loc}(t_i, t_i^*)$, shows the accuracy of the location estimate of the framboid instances.

2.3.1.4. Post-processing. To convert proposed instances into binary masks for quantification, Otsu's binarization was implemented as a post-processing algorithm (Otsu, 1979). Specifically, a bimodal intensity distribution was achieved such that the images consisted of bright framboids and a dark matrix. Otsu's algorithm was used to find a threshold value that minimizes the weighted variance of the two thresholded groups. In the resulting binarized image, the white pixels within the labeled objects were then used to calculate framboid areas and equivalent diameters. Specifically, for ellipsoidal framboids, an equivalent circular area diameter (ECD) was defined such that the cross-sectional area of the framboid was equal to that of a circle with a diameter of the ECD.

2.3.2. Instance segmentation and mask R–CNN based detection

The second workflow (Fig. 1b, Workflow 2) used in this work was an instance segmentation architecture. In this framework, training datasets were fed to the Mask R–CNN framework to calculate framboid diameter distributions. The Mask R–CNN framework uses a ResNet101+FPN (feature pyramid networks) backbone to extract framboid instances by outputting a label, a bounding box, and an object mask for each candidate (He et al.,

Table 1
Data augmentation strategy to increase variability and decrease overfitting.

Filter	Range	
Crop	0–10%	Images are cropped from 0 to 10% on each side.
Flip	N/A	Vertical flip.
Contrast Normalization	−25% to +50% absolute value	Changing the contrast in images by moving pixel values away from or closer to 128.
Gaussian Noise	0.025 × white value for one channel	Add Gaussian noise to images.
Multiply	−20 to +20%	Multiply pixel channel values by a random number between 0.8 and 1.2.
Perspective transformation	1–10%	Distort images locally.

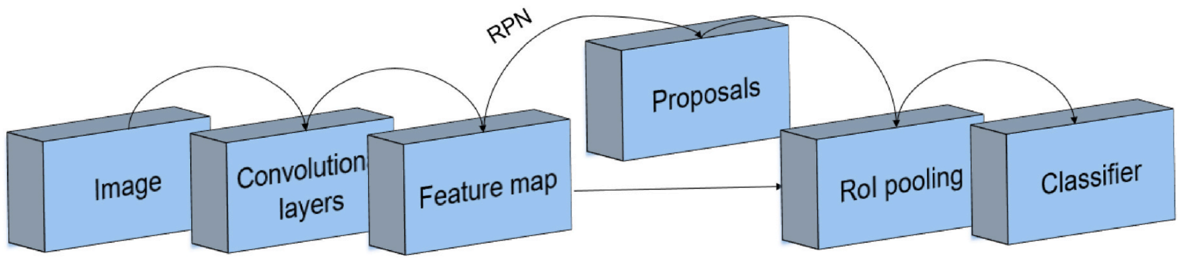

Fig. 2. Block diagram of the Inception-ResNet Faster R–CNN architecture.

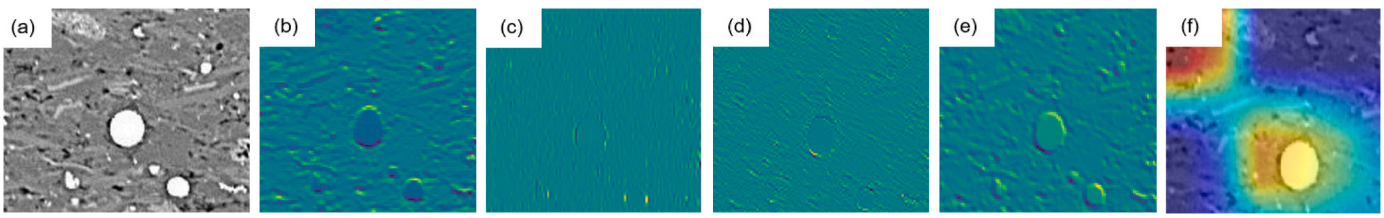

Fig. 3. Visual reproduction of the outputs of the convolutional layers showing the transformation of the original image (a) into feature maps (b to e) of the first convolution layer of the ResNet CNN. Each of the four images (b to e) represent each of the four channels. (f) A heat map representing the class probability score.

2017). Importantly, Mask R–CNN enables pixel-to-pixel alignment and therefore does not require post-processing as in object detection.

2.3.2.1. Loss function for instance segmentation. To monitor the training process for instance segmentation, a multi-task loss function, L, was calculated for each sampled RoI:

$$L = L_{cls,T} + L_{loc,T} + L_{mask,T} \tag{2}$$

where $L_{cls,T}$ is the classification loss, $L_{loc,T}$ is the localization loss, and $L_{mask,T}$ is an average cross-entropy loss. Recall that the sum of $L_{cls,T}$ and $L_{loc,T}$ is equal to the total loss from the object detection method (see Eq. (1)). The average cross-entropy loss, $L_{mask,T}$, is defined as:

$$L_{mask,T} = -\frac{1}{output}\sum_{i=1}^{output} y_i \cdot log\ \hat{y}_i + (1-y_i)\cdot \log\left(1-\hat{y}_i\right) \tag{3}$$

where $\hat{y}_i$ is the scalar value of model output, y_i is the corresponding target value, and *output* is the number of scalar values in the output.

2.4. Object detection task metrics

The performance of each model was evaluated based on its recall, precision, and F1 metrics with respect to the “ground truth” labeling that was achieved by manual tracing. Recall, R, is the ratio of detected true positive (TP) pyrite framboids to the total number of framboids of the sample (Powers, 2011). The total number of framboids of the sample is the sum of the detected TP and the missed false negatives (FN). The recall of the model, therefore, is given by

$$R = \frac{TP}{TP + FN} \tag{4}$$

A second performance metric is measured by the precision, P, of the model, given by the ratio of TPs to the total number of framboids traced automatically. The total number of framboids traced automatically includes both the true positives and false positives (FP, i.e., the number of misidentified framboids). The precision of the model is:

$$P = \frac{TP}{TP + FP} \tag{5}$$

It is worthy to note that precision varies inversely with recall. That is, the lower the recall value, the higher the precision. To find a set of generalized metrics to evaluate the overall performance of the model, we use an F1 score that is found as a harmonic mean of the recall and precision. The F1 parameter is defined by:

$$F1 = \frac{2 \times P \times R}{P + R} \tag{6}$$

Using the recall, precision, and F1 scores, the performances of the CNN methods for both BSE/SE/TLD and the BSE/EDS testing sets were evaluated. To calculate performance metrics of framboid detection (i.e., precision, recall, and F1), all proposed framboid objects with intersection over union (IoU) greater than 10% were retained (Jaccard, 1912). IoU is defined by:

$$IoU = \frac{Area\ of\ overlap}{Area\ of\ Union} \tag{7}$$

where the area of overlap is the area that *both* the predicted bounding box *and* the ground-truth bounding box exist, and the area of union is the total area that is covered by *either* the predicted bounding box *or* the ground-truth bounding box.

2.5. Distribution fit evaluation

Comparison of framboid size distributions from ML methods to manual tracing is not straightforward. In this work, three methods were applied to compare one-dimensional probability distributions for framboid sizes: (i) a graphical method, (ii) a Kolmogorov-Smirnov (K–S) test, and (iii) box-and-whisker plots. Graphical methods compare visually framboid size distributions from ML methods and manual (ground-truth) labeling. K–S tests measures the fit between two distributions (i.e., framboid size distributions from ML-based methods and from manual tracing). A K–S test provides two values: a D-statistic and a p-value that describes the probability of obtaining those test results (Stephens, 1974). The D-statistic is found by $D_{n,\ m} = sup_x|F_{1,n}(x) - F_{2,m}(x)|$, where $F_{1,n}$ and $F_{2,m}$ are the empirical distribution functions of manual tracing and Faster R–CNN/Mask R–CNN tracing respectively, and *sup* is the supremum function (Stephens, 1974). If the p-value corresponding to the D-statistic value is higher than 0.05, then the null hypothesis cannot be rejected and therefore the two distributions agree. On the contrary, if the p-value is lower than 0.05, then the null hypothesis is rejected. Further, we use the K–S test to determine D-statistic values that show the absolute maximum distance between the empirical cumulative distribution functions (ECDFs) of the two samples.

3. Results and discussion

Two machine learning models were trained and tested to calculate framboid diameter distributions. An object detection model and an instance segmentation model were trained and tested on 128 grayscale BSE/SE/TLD images and 32 colored BSE/EDS images. To prevent overfitting of the models, the training dataset was divided into training and validation subsets. The models were trained using 95 BSE/SE/TLD images and 26 BSE/EDS images and were tested on 33 BSE/SE/TLD images and 6 BSE/EDS images. Results from the ML methods were compared with manually traced results to evaluate the performance of the models.

3.1. Model training

In order to trace the framboid instances accurately, stable minimum total loss function values are required. Using the grayscale BSE/SE/TLD training images, total loss functions for all models reached a stable value (Fig. 4). We note that the loss drops significantly after ~100 epochs here due to hyperparameter fine tuning and a large number of proposed regions of interest. Similar results were obtained for the colored BSE/EDS training images (see SM Fig. 1). The training process was finished when the total loss function reached a stable minimum. Loss function curves were calculated for both training and validation datasets. During training, we used a classical 80/20 split ratio for training and validation data. Initial training datasets were split using a classical 80/20 ratio. During training, model loss functions converged and achieved a stable minimum by fine-tuning hyperparameters such as dropout, batch size, learning rate, and maximum number of proposals. Specifically, we optimized the CNNs to remove proposed regions of interest with low-quality instances (i.e., dropout), the number of images being processed (i.e., batch size), the step size of each iteration when pushing a batch through the CNN (i.e., learning rate), and the maximum number of proposed regions of interest to identify instances of pyrite framboids (i.e., maximum number of proposals). The original SEM images were used to demonstrate the reliability of applying CNN models to the raw data. To achieve highest recall value, we increased the number of proposals at the first stage to track as many proposals as possible and decreased the IoU value. To achieve minimum loss function, we used dynamic learning rate that gradually decreased from 0.0003 to 0.00003 with step 0.00003 after 5000 epochs.

3.2. Framboid detection metrics

To quantify framboid size distributions, framboid occurrences were first detected from the SEM datasets. The detection of framboids was evaluated for object detection and instance segmentation using 33 grayscale BSE/SE/TLD and 6 colored BSE/EDS images (Fig. 5). Both object detection and instance segmentation identified only those instances of spherical and/or elliptical pyrite framboids from the BSE/SE/TLD and BSE/EDS images and no irregularly shaped pyrites. Performance of the machine learning methods on the BSE/SE/TLD (Fig. 5a,c) and the BSE/EDS (Fig. 5b,d) datasets were comparable. The comparable performance on the two datasets, even though the colored BSE/EDS dataset is ~4 times smaller than the grayscale BSE/SE/TLD dataset, is due to the higher volume of data contained in the color data (i.e., red, green, and blue data as opposed to a single grayscale dataset).

Results of framboid detection metrics for object detection and for instance segmentation are shown in Tables 2 and 3, respectively. The performance of the ResNet architecture applied in this work is comparable with results reported in the literature (e.g., Wang et al., 2018). In order to capture framboid sizes accurately, we used an IoU threshold value of 90% and recall value of ~0.55. This combination of proposed recall and precision values maximizes the number of framboids captured and minimizes the errors associated with the missed (FN) and misidentified (FP) framboids.

Overall, all CNN models tested in this work captured framboids with precisions up to 0.98 and recalls up to 0.78. A recall value of 0.78 means that the CNN detects 78% of the total number of spherical framboids contained within the tested image. The comparatively lower recall value here is due to the use of raw, unprocessed SEM images taken at various settings (e.g., magnification, voltage, etc.), the large number of framboids in each image, and the number of proposed regions of interest. In this work, hyperparameters were tuned to calculate a statistical distribution of pyrite framboid sizes to inform redox histories, and capturing all possible instances of framboids (i.e., recall approaching 1) was not the main goal. The CNNs here were trained and optimized to capture representative size distribution data from raw, unprocessed data to simplify the characterization process for geoscientists. Increased recall values are achievable by pre-processing the data or by increasing the number of proposed regions of interest.

Notably, ResNet 100 and Inception-ResNet achieved the highest F1 scores because of the presence of residual connections and batch normalization that help improve the training process. Because the trained and tested SEM images were acquired using a wide range of operating conditions (e.g., magnifications, energy, etc.), we expect to observe lower values for recall scores than those from training and testing datasets obtained using uniform operating conditions. Further, model performances on grayscale BSE/SE/TLD images and colored BSE/EDS images cannot be compared directly because BSE/EDS dataset is ~4 times smaller than the BSE/SE/TLD dataset. Similar precision, however, was achieved for the Inception-ResNet model using both datasets. A straightforward extension of this work includes the application of Mask R–CNN methods to identify pores, fractures, organic matter, and/or mineral grains in geological materials.

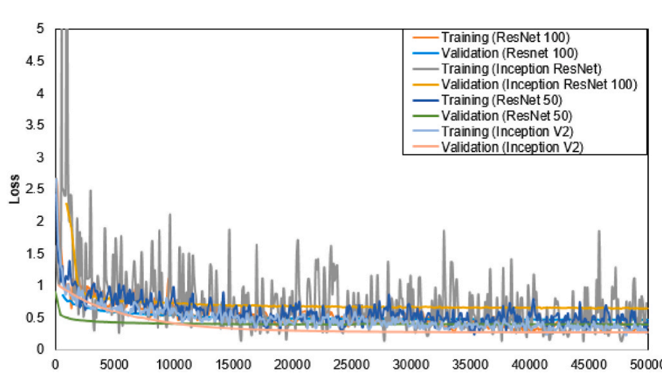

Fig. 4. Loss functions of the BSE/SE/TLD dataset. Exponential smoothing was applied with a smoothing constant of 0.6.

3.3. Comparison of framboid size distribution results

Framboid size distributions from BSE/SE/TLD grayscale images and for BSE/EDS images using object detection and instance segmentation agreed well with ground truth labeling (Figs. 6 and 7). To calculate framboid sizes, we (i) binarized the occurrences of framboids proposed by object detection, and (ii) used the instance segmentation output masks. Framboid sizes and their occurrence frequencies obtained using each method were compared with manually traced data, and frequency distributions of normalized equivalent circular diameter (ECD) of all framboids were recorded. In comparison to the manually traced ground truth, the framboid size distributions extracted by object detection are characterized by the same ranges and similar variability (Fig. 6). For instance segmentation, the extracted framboid size distributions are also comparable with the manually traced data (Fig. 6).

Visual inspection of CNN traced and extracted framboid size

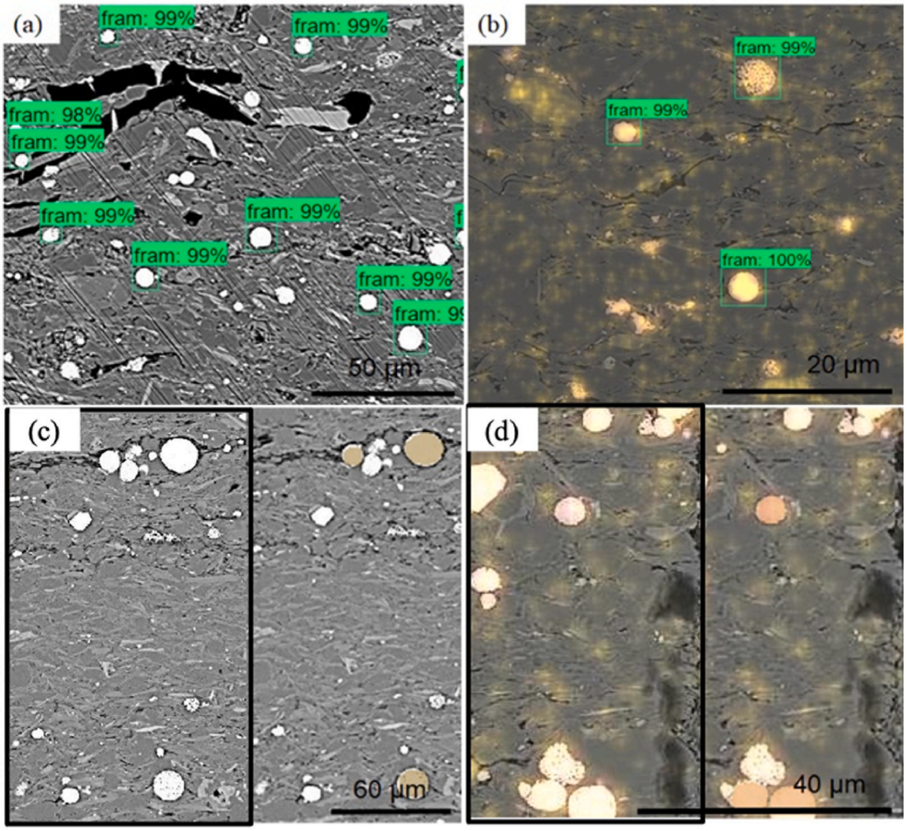

Fig. 5. Examples of circular framboid detection on test images. (a) Identification of framboids using object detection, i.e., Faster R–CNN models, from BSE/SE/TLD grayscale images. All traced instances containing circular framboids are labeled with IoU values (e.g., 99%). (b) Object detection of BSE/EDS maps. (c) Instance segmentation model of grayscale BSE/SE/TLD images with the original raw SEM image (left) and the processed image with traced spherical framboids on BSE image using Mask R–CNN (right). (d) Instance segmentation of BSE/EDS maps images with the original raw SEM image (left) and the processed image with traced spherical framboids on BSE image using Mask R–CNN (right).

Table 2
Performance of the object detection Faster R–CNN models for BSE/TLD and BSE/EDS dataset.

	BSE/SE/TLD images			BSE/EDS images		
	Precision	Recall	F1	Precision	Recall	F1
Inception V2	0.98	0.62	0.74	0.95	0.48	0.59
ResNet 50	0.96	0.61	0.73	0.92	0.52	0.60
ResNet 100	0.96	0.76	0.83	1.00	0.44	0.56
Inception-ResNet	0.98	0.78	0.87	0.98	0.53	0.64

Table 3
Performance of the instance segmentation Mask R–CNN models for BSE/SE/TLD and BSE/EDS datasets.

	BSE/SE/TLD images			BSE/EDS images		
	Precision	Recall	F1	Precision	Recall	F1
Mask R–CNN	0.96	0.60	0.72	0.98	0.48	0.64

distributions shows agreement between the manual and R–CNN traced distributions, especially in the case of the relatively large BSE/SE/TLD dataset. Although most framboid sizes were extracted accurately, the models are not perfect. For example, very small framboids (diameter ~ 2 to 3 μm) were not identified correctly in low magnification images due to limitations in image resolution. Further, application of noise reduction filtering during post-processing may erode and eliminate very small framboids. In this work, the parameters for the post-processing algorithm were kept uniform to compare the performance of the four different Faster R–CNN models. D-statistic and P-values calculated for framboid size distributions from both the object detection and instance segmentation models show agreement with manually traced results (Tables 4 and 5).

P-values computed using D-statistic values show that the frequency distribution functions are comparable to manually traced frequency distribution functions with insignificant deviations. K–S tests were applied on normalized distributions that were acquired from raw distributions after assigning the size of the framboid to the specific range. Among the CNN models, Inception-ResNet showed the best performance with D-statistic value of 0.14 and P-value of 0.99.

The best performing models, ResNet 100 and Inception-ResNet, solve the object detection task most accurately. Subsequent post-processing with Otsu thresholding gives the most precise model for capturing of framboid size distributions. Overall, all models trace and extract framboids that capture the manually traced ground-truth distribution. Importantly, once trained, the ML methods require only $\lesssim 6$ s to process a BSE/SE/TLD image that contains ~50 framboid objects, a ~100-fold improvement on current manual tracing (~10 min/image).

3.4. *Framboid size distributions informing changes in ocean oxygen level*

Common tools used to evaluate local paleo-redox conditions include Fe speciation, δ^{56}Fe, macrofauna, biomarkers, ichnofacies index, δ^{98}Mo, δ^{53}Cr, Ce/Ce* ratio and REEs profiles. For the present study area, the maturity of the Marcellus Shale limits the use of Fe speciation, δ^{56}Fe, biomarkers, δ^{98}Mo, δ^{53}Cr, Ce/Ce* ratio and REEs. Specifically, the concentration of trace elements is affected significantly by sedimentation rate. In the foreland basin setting, the sedimentation rate can vary during the formation of the foreland and it is therefore challenging to delineate the redox history using the concentration of redox-sensitive trace elements such as Mo, V, and U alone. In the deep marine settings during the Marcellus deposition, bioturbation is rare and limited. Ichnofacies data, as a result, are to elucidate the local redox conditions.

To understand the sedimentary redox conditions of the Marcellus Shale in the Middle Devonian foreland basin, we integrated our pyrite framboid size distribution data derived from the CNN architectures with

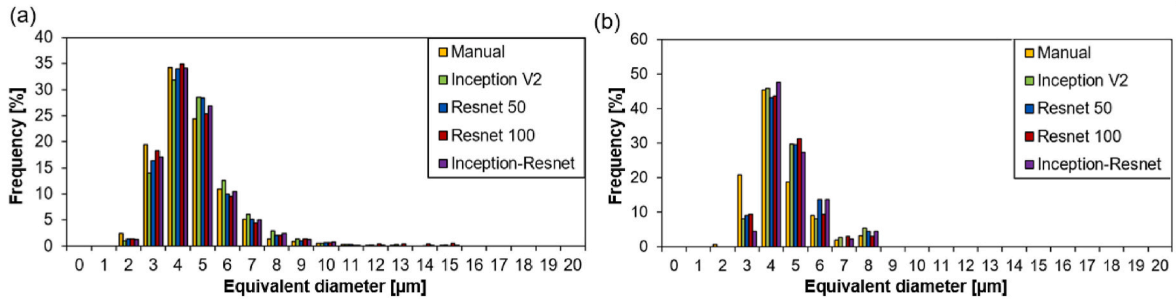

Fig. 6. Framboid size distributions measured using object detection in agreement with results from manual tracing for (a) the grayscale BSE/SE/TLD dataset and (b) the colored BSE/EDS dataset.

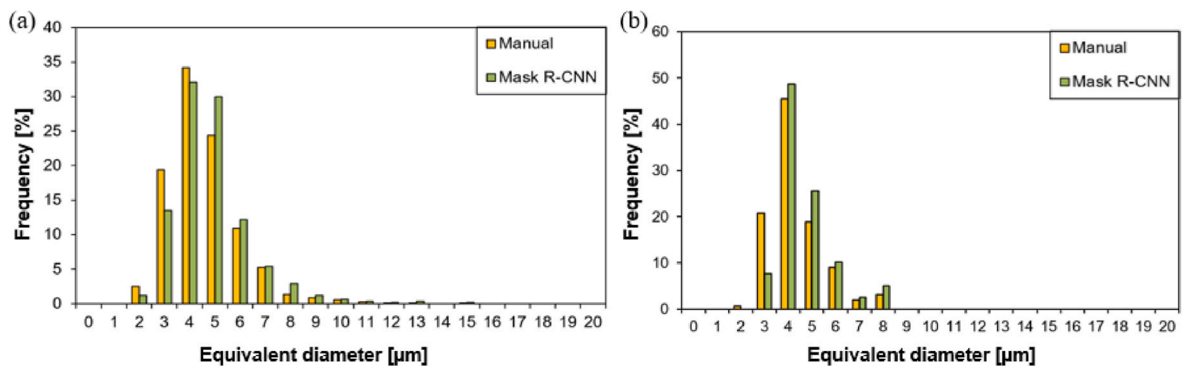

Fig. 7. Framboid size distribution measured using instance segmentation in agreement with results from manual tracing for (a) the grayscale BSE/SE/TLD dataset and (b) the colored BSE/EDS dataset.

Table 4
D-statistic and P-values for distribution fitness evaluation of framboid size distributions obtained using object detection.

	BSE/SE/TLD images		BSE/EDS images	
	D-statistic	P-value	D-statistic	P-value
Inception V2	0.14	0.98	0.095	0.99
ResNet 50	0.19	0.85	0.095	0.99
ResNet 100	0.17	0.83	0.047	0.99
Inception-ResNet	0.14	0.99	0.047	0.99

Table 5
D-statistic and P-values for distribution fitness evaluation of framboid size distributions obtained using instance segmentation.

	BSE/SE/TLD images		BSE/EDS images	
	D-statistic	P-value	D-statistic	P-value
Mask R–CNN	0.14	0.98	0.095	0.99

data on lithofacies, fauna assemblage, and major and trace elemental profiles. Specifically, we incorporated a high-resolution (~2-inch interval) major and trace elemental profile of the Marcellus cored interval and a detailed core description, including thin-section and SEM petrographical studies, of biota assemblage and variation, sedimentary structures, extent of bioturbation, organic matter type and content (Ko et al., 2019).

The framboid size distributions computed from the ML methods were plotted as a function of sample depth and compared to the distributions calculated from manual tracing (Table 6). The data are presented as box-and-whisker plots in sedimentary succession (Fig. 8). Variations in the burial redox conditions manifest as changes in the mean and standard deviation (SD) of the framboid size distributions across depths. Recall that oxygen-deprived euxinic and anoxic bottom water conditions produce abundant small framboids with a narrow size distribution (transport limited pyrite formation) whereas oxygen-available oxic/dysoxic bottom waters produce larger framboids and with large size distributions (reaction limited pyrite formation). The framboid size distributions

Table 6
Quantitative data to estimate framboid size distribution for manual and Faster R–CNN Inception-ResNet tracing.

	Manual					Inception-ResNet				
Depth [ft]	Count	Mean	SD	Min	Max	Count	Mean	SD	Min	Max
7846.42	713	4.07	1.29	1.56	10.37	261	4.29	1.42	1.59	13.89
7869	178	4.43	1.58	2.00	10.98	99	4.57	1.62	2.47	9.83
7883.25	328	4.27	1.85	1.20	14.47	169	4.36	1.80	1.79	14.41
7917.25	224	3.76	1.28	1.68	11.77	112	4.12	1.44	2.12	11.95
7924.85	249	3.80	1.35	1.49	10.56	143	3.92	1.34	1.62	10.20
7954.15	115	3.98	1.26	1.77	8.53	85	3.62	1.18	1.61	8.27

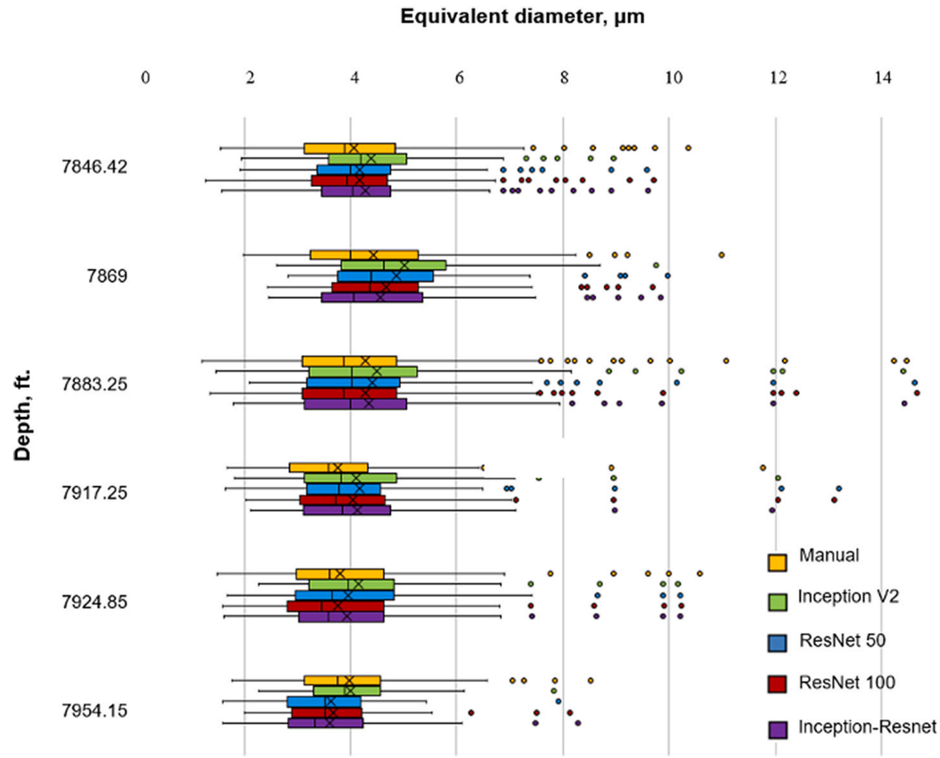

Fig. 8. Box-and whisker plots for the tested images associated with depth in the Marcellus Shale. BSE/SE/TLD dataset.

calculated using the CNNs show that for burial depths of ~7840–7960 ft in the Marcellus Shale, the majority of the framboid sizes were within 5 μm (Fig. 8) and thereby indicating a persistent euxinic and/or anoxic bottom water condition in the deep region of the foreland basin during the Middle Devonian. Characterization data show the existence of larger framboids (~6–15 μm) in each sample and suggest the availability of relatively oxygenated conditions in the deep waters of the foreland basin.

Framboid size distributions measured from this cored well were compared to those from two others in Upshur Co. and Greene Co. (Blood and Lash, 2015). Overall, the sizes of framboids in our locality are smaller due to the proximity of the well to deeper regions of the foreland basin. Further, the biota in thin sections suggest variation in redox conditions in the upper Marcellus.

3.5. Challenges in CNN image processing of framboids

Although the results of framboid size distributions computed using object detection and instance segmentation agreed with the manually traced data, some challenges were encountered. Specifically, data availability, single-channel grayscale images, and inconsistent imaging magnifications limit the accuracy of extracted data significantly. For example, low-magnification images have the advantage of capturing a significant number of framboids to improve the volume of data, but the accuracy of the size detection method was reduced due to insufficient spatial resolution of the images. Specifically, accuracy of the framboid size distribution measured by CNN methods varied inversely with magnification (Fig. 9). To estimate the effect of magnification, 1806 framboids traced manually and 869 framboids traced by the Inception-ResNet model were compared. For each image, the mean framboid diameter computed by the CNN models were compared with results from manual tracing. We find that images captured at pixel resolutions greater than 0.25 μm/pixel track a relatively high number of framboids but the mean framboid diameter deviates

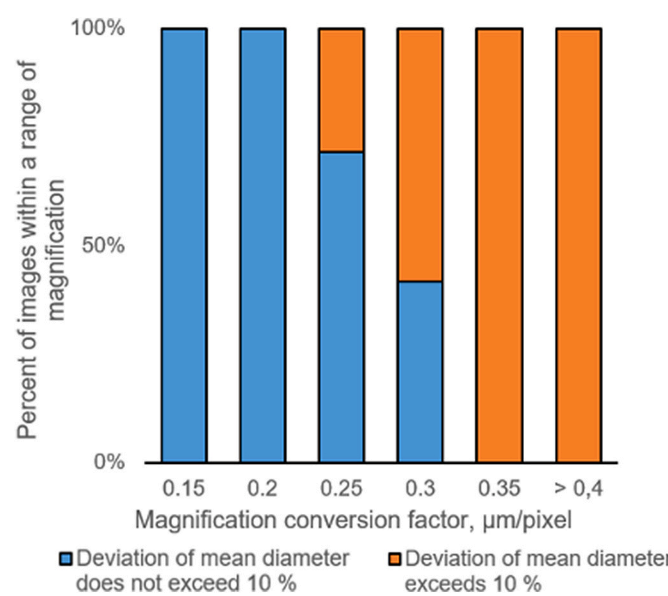

Fig. 9. Inverse effects of SEM magnification on the accuracy of framboid size detection using CNN image processing methods.

significantly from the manually traced results. A possible cause for this deviation could be due to the inability of the R–CNN method to trace small framboid instances due to a limited number of framboid pixels.

A second challenge in using CNN to quantify framboid size distributions stem from the associated computational (Table 7). Using the Windows-based computer (Intel Core i7, 4.20 GHz, 32 GB RAM and Nvidia GeForce GPU P4000 with 8 GB memory), the method with the best performance, Inception-ResNet, required 26 h of computing time to train the

Table 7
Estimated computational cost to train the object detection models for grayscale BSE/SE/TLD images and for colored BSE/EDS images. The number of steps were chosen based on the task that achieved minimum stable loss function behavior.

	BSE/SE/TLD images			BSE/EDS images		
	Epochs	Time per epoch [s]	Total Time [h]	Epochs	Time per epoch [s]	Total time [h]
Inception V2	50000	0.14	10	30000	0.17	7
ResNet 50	50000	0.25	17	30000	0.27	8
ResNet 100	50000	0.35	24	30000	0.418	21
Inception-ResNet	50000	1.1	26	30000	1.15	23

grayscale BSE/SE/TLD dataset (95 original images and ~1044 augmented images). Similarly, 23 h was required to train the colored BSE/EDS dataset (26 original images and ~286 augmented images). This challenge, favorably, can be overcome by the use of computing clusters with parallel computing capabilities, high performance cloud computing, higher power GPUs, and/or multiple cores.

4. Conclusion

This work presents machine-learning as an automated tool to analyze micrographs of geological samples. Full automation of image processing removes human biasing, is fast and robust, and, importantly, enables the processing of high volumes of visual data. In this work, we demonstrated the utility of ML-based methods by characterizing pyrite framboid size distributions to inform the depositional redox conditions of geological samples. Object detection models, including Inception V2, ResNet 50, ResNet 100, and Inception-ResNet, and an instance segmentation model, Mask R–CNN, demonstrate the applicability of CNN-based image analyses techniques toward geological micrograph processing. Notably, the Inception-ResNet object detection model achieved the highest performance metrics (precision 0.99, recall 0.78). Framboid size distributions obtained from manual and ML-methods were compared using a K–S test, visual inspection, and box-and-whisker plots and show agreement between the methods. Favorably, for the Inception-ResNet model, the K–S test yielded a P-value of 0.99 and a D-statistic of 0.13.

In this work, grayscale BSE/SE/TLD and color BSE/EDS images were used as two separate datasets to demonstrate ML-based model performance. We find that in analyzing a smaller colored BSE/EDS dataset, both the object detection and instance segmentation models leveraged the multi-channel nature of the colored image data to achieve similar performance with higher volume grayscale BSE/SE/TLD data. Further, we find that images with magnification lower than ~1000 X (~0.3 μm/pixel) resulted in decrease precision of framboid size distributions despite capturing a greater number of framboids. Importantly, once trained, the ML methods are ~100 times faster than current manual tracing. Automated imaging provides a path toward capturing increased numbers of high magnification images to resolve issues related to small datasets. Broadly, the ML-based approach demonstrated in this work can be extended to quantify grayscale BSE/SE/TLD and colored EDS/EDS micrographs of other geological features, including pores/cracks, organic matter, carbonates, and siliciclastic minerals, to enable further geochemical and geomechanical understanding of the Earth. The use of ML in geologic image characterization improves both the quantity and quality of image analyses that are currently performed by hand tracing. In addition to confirming manual identification, once trained, the CNNs here enable the fast and reliable processing of a huge volume of data, something that was not previously possible. Improved data volumes eliminate challenges in analyses including biasing due to visual selection and inadequate sampling.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge funding from the American Chemical Society.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.marpetgeo.2021.105159.

References

Anderson, T.I., Vega, B., Kovscek, A.R., 2020. Multimodal imaging and machine learning to enhance microscope images of shale. Comput. Geosci. 145.
Berner, R.A., 1970. Sedimentary pyrite formation. Am. J. Sci. 268 (1), 1–23.
Berner, R.A., 1984. Sedimentary pyrite formation: an update. Geochem. Cosmochim. Acta 48 (4), 605–615.
Blood, D.R., Lash, G.G., 2015. Dynamic redox conditions in the Marcellus Shale as recorded by pyrite framboid size distributions. In: Larsen, D., Egenhoff, S.O., Fishman, N.S. (Eds.), Paying Attention to Mudrocks: Priceless!, vol. 515. The Geological Society of America Special Papers, pp. 153–168.
Bond, D.P., Wignall, P.B., 2010. Pyrite framboid study of marine Permian–Triassic boundary sections: a complex anoxic event and its relationship to contemporaneous mass extinction. GSA Bulletin 122 (7–8), 1265–1279.
Bond, D., Wignall, P.B., Racki, G., 2004. Extent and duration of marine anoxia during the Frasnian–Famennian (Late Devonian) mass extinction in Poland, Germany, Austria and France. Geol. Mag. 141 (2), 173–193.
Brennecka, G.A., Hermann, A.D., Algeo, T.J., Anbar, A.D., 2011. Rapid expansion of oceanic anoxia immediately before the end-Permian mass extinction. Proc. Natl. Acad. Sci. U. S. A 108 (43), 17631–17634.
Chen, Z., Liu, X., Yang, J., Little, E., Zhou, Y., 2020. Deep learning-based method for SEM image segmentation in mineral characterization, an example from Duvernay Shale samples in Western Canada Sedimentary Basin. Comput. Geosci. 138, 104450.
DiChristina, T.J., Bates, D.J., Burns, J.L., Dale, J.R., Payne, A.N., 2006. Shewanella: novel strategies for anaerobic respiration. Past and Present Water Column Anoxia. Springer, Dordrecht, pp. 443–469.
Formolo, M.J., Lyons, T.W., 2007. Accumulation and preservation of reworked marine pyrite beneath an oxygen-rich Devonian atmosphere: constraints from sulfur isotopes and framboid textures. J. Sediment. Res. 77 (8), 623–633.
Gallego-Torres, D., Reolid, M., Nieto-Moreno, V., Martínez-Casado, F.J., 2015. Pyrite framboid size distribution as a record for relative variations in sedimentation rate: an example on the Toarcian Oceanic Anoxic Event in Southiberian Palaeomargin. Sediment. Geol. 330, 59–73.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. Proceedings of the IEEE international conference on computer vision 2961–2969.
Huang, Y., Chen, Z.Q., Wignall, P.B., Zhao, L., 2017. Latest Permian to Middle Triassic redox condition variations in ramp settings, South China: pyrite framboid evidence. GSA Bulletin 129 (1–2), 229–243.
Ikeda, K., Goldfarb, E., Tisato, N., 2019. Convolutional neural networks for semantic segmentation of micro-pores in SEM based images of shales. In: 2019 AAPG Annual Convention and Exhibition.
Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning (pp. 448-456). PMLR.
Isozaki, Y., 1997. Permo-Triassic boundary superanoxia and stratified superocean: records from lost deep sea. Science 276 (5310), 235–238.
Jaccard, P., 1912. The distribution of the flora in the alpine zone. 1. New Phytol. 11 (2), 37–50.
Karduck, S.M., 2015. Pyrite Framboid Size Distributions in the Union Springs Member, Marcellus Formation and Implications for Paleowater Column Euxinia. Master Thesis. The Pennsylvania State University, p. 96.
Ko, L.T., 2019. Facies, depositional environments, chemostratigraphy, and reservoir quality of the Middle devonian Marcellus Formation, appalachian basin, northeastern Pennsylvania. Houston Geological Society Bulletin 62, 25.
Koeshidayatullah, A., Morsilli, M., Lehrmann, D.J., Al-Ramadan, K., Payne, J.L., 2020. Fully automated carbonate petrography using deep convolutional neural networks. Mar. Petrol. Geol. 122, 104687.
Lau, K.V., Maher, K., Altiner, D., Kelley, B.M., Kump, L.R., Lehrmann, D.J., Silva-Tamayo, J.C., Weaver, K.L., Yu, M., Payne, J.L., 2016. Proc. Natl. Acad. Sci. U. S. A 113 (9), 2360–2365.

Love, L.G., 1966. Review of microscopic pyrite from the devonian chattanooga shale and rammelserg banderz. Fortschr. Mineral. 43, 277–309.
Maitre, J., Bouchard, K., Bédard, L.P., 2019. Mineral grains recognition using computer vision and machine learning. Comput. Geosci. 130, 84–93.
Marynowski, L., Rakocinski, M., Zaton, M., 2007. Middle Famennian (Late Devonian) interval with pyritized fauna from the Holy Cross Mountains (Poland): organic geochemistry and pyrite framboid diameter study. Geochem. J. 41 (3), 187–200.
Meyer, K.M., Kump, L.R., Ridgwell, A., 2008. Biogeochemical controls on photic-zone euxinia during the end-Permian mass extinction. Geology 36 (9), 747–750.
Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics 9 (1), 62–66.
Payne, J.L., Lehrmann, D.J., Wei, J., Orchard, M.J., Schrag, D.P., Knoll, A.H., 2004. Large perturbation of the carbon cycle during recovery from the end-Permian extinction. Science 305 (5683), 506–509.
Perez, L., Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning. arXiv:1712.04621.
Powers, D.M., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. J. Mach. Learn. Technol. 2 (1), 37–63.
Raup, D.M., Sepkoski, J.J., 1982. Mass extinction in the marine fossil record. Science 215, 1501–1503. https://doi.org/10.1126/science.215.4539.1501.
Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster r-cnn: towards real-time object detection with region proposal networks. Adv. Neural Inf. Process. Syst. 91–99.
Richa, Mukerji, T., Mavko, G., Keehm, Y., 2006. Image analysis and pattern recognition for porosity estimation from thin sections. In: SEG Technical Program Expanded Abstracts 2006. Society of Exploration Geophysicists, pp. 1968–1972.
Rickard, D.T., 1970. The origin of framboids. Lithos 3 (3), 269–293.
Roduit, N. JMicroVision: Image analysis toolbox for measuring and quantifying components of high-definition images. Version 1.3.1. https://jmicrovision.github.io (accessed April 5th, 2019).
Schieber, J., Baird, G., 2001. On the origin and significance of pyrite spheres in Devonian black shales of North America. J. Sediment. Res. 71 (1), 155–166.
Sepkoski, J.J., 1996. Patterns of Phanerozoic extinction: a perspective from global data bases. Global Events and Event Stratigraphy in the Phanerozoic. Springer, Berlin, Heidelberg, pp. 35–51.
Shen, W., Lin, Y., Xu, L., Li, J., Wu, Y., Sun, Y., 2007. Pyrite framboids in the Permian–Triassic boundary section at Meishan, China: evidence for dysoxic deposition. Palaeogeogr. Palaeoclimatol. Palaeoecol. 253 (3–4), 323–331.
Stephens, M.A., 1974. EDF statistics for goodness of fit and some comparisons. J. Am. Stat. Assoc. 69 (347), 730–737.
Suits, N.S., Wilkin, R.T., 1998. Pyrite formation in the water column and sediments of a meromictic lake. Geology 26 (12), 1099–1102.
Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A., 2016a. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning arXiv preprint arXiv: 1602.07261.
Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016b. Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826.
Tang, D., Spikes, K., 2017. Segmentation of shale SEM images using machine learning. SEG Technical Program Expanded Abstracts 2017. Society of Exploration Geophysicists, pp. 3898–3902.
Tian, X., Daigle, H., 2019. Preferential mineral-microfracture association in intact and deformed shales detected by machine learning object detection. J. Nat. Gas Sci. Eng. 63, 27–37.
Wang, P., Huang, Y., Wang, C., Feng, Z., Huang, Q., 2013. Pyrite morphology in the first member of the late cretaceous qingshankou formation, songliao basin, northeast China. Palaeogeogr. Palaeoclimatol. Palaeoecol. 385, 125–136.
Wang, Z., Hu, M., Zhai, G., 2018. Application of deep learning architectures for accurate and rapid detection of internal mechanical damage of blueberry using hyperspectral transmittance data. Sensors 18 (4), 1126.
Wei, H., Chen, D., Wang, J., Yu, H., Tucker, M.E., 2012. Organic accumulation in the lower Chihsia Formation (Middle Permian) of South China: constraints from pyrite morphology and multiple geochemical proxies. Palaeogeogr. Palaeoclimatol. Palaeoecol. 353–355, 73–86.
Wignall, P.B., Newton, R., 1998. Pyrite framboid diameter as a measure of oxygen deficiency in ancient mudrocks. Am. J. Sci. 298 (7), 537–552. https://doi.org/10.2475/ajs.298.7.537.
Wignall, P.B., Twitchett, R.J., 1996. Oceanic anoxia and the end Permian mass extinction. Science 272 (5265), 1155–1158.
Wilkin, R.T., Barnes, H.L., Brantley, S.L., 1996. The size distribution of framboidal pyrite in modern sediments: an indicator of redox conditions. Geochem. Cosmochim. Acta 60 (20), 3897–3912. https://doi.org/10.1016/0016-7037(96)00209-8.
Wilkin, R.T., Arthur, M.A., Dean, W.E., 1997. History of water-column anoxia in the Black Sea indicated by pyrite framboid size distributions. Earth Planet Sci. Lett. 148 (3–4), 517–525. https://doi.org/10.1016/S0012-821X(97)00053-8.
Wu, Y., Misra, S., Sondergeld, C., Curtis, M., Jernigen, J., 2019. Machine learning for locating organic matter and pores in scanning electron microscopy images of organic-rich shales. Fuel 253, 662–676.