PALAIOS
Emphasizing the impact of life on Earth's history

PALAIOS, 2020, v. 35, 391–402
Research Article
DOI: http://dx.doi.org/10.2110/palo.2019.102

CONVOLUTIONAL NEURAL NETWORKS AS AN AID TO BIOSTRATIGRAPHY AND MICROPALEONTOLOGY: A TEST ON LATE PALEOZOIC MICROFOSSILS

RAFAEL PIRES DE LIMA,[1,2] KATIE F. WELCH,[1] JAMES E. BARRICK,[3] KURT J. MARFURT,[1] ROGER BURKHALTER,[4] MURPHY CASSEL,[1] AND GERILYN S. SOREGHAN[1]

[1]*School of Geosciences, The University of Oklahoma, 100 East Boyd Street, RM 710, Norman, Oklahoma 73019, USA*
[2]*The Geological Survey of Brazil–CPRM, 55 Rua Costa, São Paulo, São Paulo, Brazil*
[3]*Department of Geosciences, Mail Stop 1053, Texas Tech University, Lubbock, Texas 79409, USA*
[4]*Sam Noble Museum, The University of Oklahoma, 2401 Chautauqua Ave., Norman, Oklahoma 73072, USA*
email: rlima@ou.edu; rafael.lima@cprm.gov.br

ABSTRACT: **Accurate taxonomic classification of microfossils in thin-sections is an important biostratigraphic procedure. As paleontological expertise is typically restricted to specific taxonomic groups and experts are not present in all institutions, geoscience researchers often suffer from lack of quick access to critical taxonomic knowledge for biostratigraphic analyses. Moreover, diminishing emphasis on education and training in systematics poses a major challenge for the future of biostratigraphy, and on associated endeavors reliant on systematics. Here we present a machine learning approach to classify and organize fusulinids—microscopic index fossils for the late Paleozoic. The technique we employ has the potential to use such important taxonomic knowledge in models that can be applied to recognize and categorize fossil specimens. Our results demonstrate that, given adequate images and training, convolutional neural network models can correctly identify fusulinids with high levels of accuracy. Continued efforts in digitization of biological and paleontological collections at numerous museums and adoption of machine learning by paleontologists can enable the development of highly accurate and easy-to-use classification tools and, thus, facilitate biostratigraphic analyses by non-experts as well as allow for cross-validation of disparate collections around the world. Automation of classification work would also enable expert paleontologists and others to focus efforts on exploration of more complex interpretations and concepts.**

INTRODUCTION

Biostratigraphy is a critical approach for dating and correlating sedimentary successions, particularly given the common absence of material appropriate for radioisotopic dating of sedimentary strata. Biostratigraphy hinges on detailed analysis of extracted fossils, or thin-sections of fossils, to identify specimens to the genus and species levels. In addition to their utility for dating, fossil assemblages shed light on paleoenvironmental conditions; foraminiferal assemblages, for example, can yield information critical for reconstructing histories of paleoclimatic and paleoceanographic conditions (e.g., Gooday 1994; Culver 2003; Kucera 2007; Roozpeykar and Moghaddam 2016). Commonly, biostratigraphy relies on species-level identification of well-established microfossil groups such as foraminifera, conodonts, and palynomorphs. The complex morphology of fossil organisms requires the use of specialists for reliable and correct systematics, especially for the generation of detailed and accurate biostratigraphic correlation. Unfortunately, education and training in the identification of fossil taxa is diminishing, greatly crippling future capacity in this area (Farley and Armentrout 2000, 2002).

Although the largest investment of resources, both time and financial, for biostratigraphic studies may be for data acquisition and sample preparation, not every institution possesses the requisite expertise enabling accurate species-level identification for biostratigraphic analyses. Both old and new paleontological collections of significant biostratigraphic value may be overlooked or ignored because no one is available to perform the necessary taxonomic identifications. This growing paucity of experts brings forth the need for a new approach to facilitate access to existing taxonomic knowledge to a broader audience.

The ongoing revolution in big data and statistical analysis enables the possibility of accelerating and standardizing fossil characterization and identification with machine learning techniques. In deep learning, machine learning models consisting of more than one artificial neural network layer, have the ability to learn representations of data with different levels of abstraction (LeCun et al. 2015). Recent advances in the architecture of deep learning convolutional neural networks (CNN) have greatly improved the fields of image classification and computer vision. LeCun et al. (2015) provided details on deep learning and showed some of the breakthoughs achieved by such technology. Dumoulin and Visin (2016) showed details on convolutions and arithmetic for deep learning procedures. We provide in Online Supplemental File 1 the very basics of artificial neural networks and CNNs. Our material is built on images and examples to provide the reader with an intuitive understanding of the mechanics of deep learning and CNNs without delving into the details of the mathematical computations.

CNN models increase the levels of accuracy in numerous image classification tasks. For example, current CNN models can differentiate not only the image of a leopard from that of a mite or a container ship (objects with significantly different characteristics), but can differentiate images of leopards from their biological cousins—jaguars, cheetahs, and snow leopards (objects with very similar characteristics; e.g., Krizhevsky et al. 2012). Szegedy et al.'s (2015) Inception V3 CNN architecture reached a 3.5% top-5 error (frequency in which the model cannot predict the correct class as one of the top five most probable guesses) and 17.3% top-1 error

Published Online: September 2020
 0883-1351/20/035-391

in the classification of the ImageNet Large-Scale Visual Recognition Challenge (Russakovsky et al. 2015). The training data for the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), is a subset of ImageNet containing the 1000 mixed-object categories and 1.2 million images.

A currently underutilized application of machine learning is fossil identification—a key component of biostratigraphy. Ranaweera et al. (2009) used a computer-aided approach in which they applied clustering techniques followed by expert labeling for identification of foraminifers. Recent work has focused on the generation of a foraminiferal identification pipeline that uses CNN and other machine learning methods and compares such results to classifications performed by human experts (Zhong et al. 2017). Hsiang et al. (2019) presented an extensive image library of modern planktonic foraminifera and showed results for the classification of such of planktonic foraminifera using CNN. Marchant et al. (2019) used the data from Hsiang et al. (2019) to help train CNN models to classify foraminifera from Holocene down-cores. Kong et al. (2016) showed a novel technique applicable for fossil pollen identification, by our understanding, the first application of CNN image classification techniques applied to fossil specimens. Kong et al. (2016) selected patches of fossil pollen grains in photomicrographs and used a pretrained CNN model to extract features for pollen species identification. Pollen researchers have been working on automated identification for some time and are experimenting with CNN models (e.g., Sevillano and Aznarte 2018). CNN has also been used for diatom and radiolarian classification (e.g., Keçeli et al. 2017; Pedraza et al. 2017). Given the current proliferation of efforts to digitize biological specimens, both modern and fossil (e.g., Blagoderov et al. 2012; Ellwood et al. 2015, Hsiang et al. 2019; Valan et al. 2019), successful application of CNN methods could greatly facilitate research that relies upon fossil identification and biostratigraphy.

We provide what we believe is a novel attempt to conduct automated fossil classification using CNN models, and the first attempt on a fossil group (late Paleozoic fusulinids) identified specifically through 2D thin-section analysis. This methodology does not depend on specialized bench work and can be applied to existing photomicrographs in legacy collections—indirectly capturing the knowledge of the researchers that performed the classification or labeling of these collections. Our test case analyses provide proof-of-concept verification, as we obtained highly accurate results with significantly smaller domain-specific training data relative to traditional CNN applications. Although researchers are working with CNN models to perform image recognition using only a few examples for training, in some cases a single example per class (e.g., Koch 2015; Lake et al. 2015; Santoro et al. 2016), most CNN applications use hundreds to hundreds of thousands of examples per class. With the additional imaging of the numerous specimens in the large legacy collections of fusulinids in North America and other paleobiogeographic realms (e.g., Ross 1967), automated classification can potentially organize a large volume of taxonomic and biostratigraphic information into a reliable and coherent system easily accessible to a variety of users, including both specialists and non-specialists. Our methodology uses data coming from traditional paleontological field and laboratory work and depends on specimen quality, but does not aim to diminish the importance of current paleontological techniques and expertise. Our objective is to both help accelerate and disseminate fossil classification knowledge, and to bring fossil classification expertise to research groups with the requisite data, but little or no access to expert paleontologists.

SHORT GLOSSARY

Because machine learning, and CNN in particular, may be unfamiliar to many paleontologists, we provide this simple glossary to define some of the technical terms used in the manuscript. More detailed machine learning definitions appears in the list of references as well as online under "Machine Learning Glossary | Google Developers" (https://developers.google.com/machine-learning/glossary; accessed 1/2019).

Accuracy.—The ratio between the number of correct classifications and the total number of classifications performed. Values range from 0.0 to 1.0 (equivalently, 0% to 100%). A perfect score of 1.0 means all classifications were correct whereas a score of 0.0 means all classifications were incorrect.

Class.—The name, or category assigned to each data sample. In this paper we use "class" in the machine learning sense rather than in the biological sense.

Classification.—The process of assigning data to a particular class.

Convolution.—A mathematical operation that uses two functions, one generally interpreted as the "input", and the other as a "filter". The filter can also be referred to as the "kernel". The kernel is applied on the input, producing an output image or signal. In machine learning applications, a convolutional layer uses the convolutional kernel and the input data to train the convolutional kernel weights.

Cross Entropy Loss.—A measure of the difference between the model's predictions are from the provided label; specifically, cross-entropy measures the difference between two probability distributions.

Convolutional Neural Network (CNN).—A neural network architecture in which at least one layer is a convolutional layer. Some authors also use ConvNets as a shorter term.

Deep Neural Network.—An artificial neural network that uses more than one hidden layer. The process of using deep neural networks is sometimes referred to as deep learning.

Epoch.—Generally used to depict a single pass through the full training set during the training stage. Not to be confused with a geological time epoch.

Fine Tuning.—A secondary training executed to further adjust the weights of an already trained model so the model can better achieve a secondary task.

Hyperparameter.—The available "options" a user can change for different attempts to train a model. Hyperparameters contrast with weights/parameters that are automatically updated by the model, following the model's algorithm. For example, the number of epochs used to train a model is a hyperparameter.

Labels.—Names applied to an instance, sample, or example (for image classification, an image) associating it with a given class. In this paper the labels are the names of the target genus analyzed.

Layer.—In (artificial) neural networks, refers to a set of neurons that processes the same set of input features.

Loss.—A measure of the model's performance, or how far the predictions are from the desired output.

Machine Learning.—A collection of approaches in which models improve their performance through automatic analysis of data.

Natural Images.—A term commonly used in computer vision literature and without a strict definition. In a broad sense, the resulting color photograph taken with an ordinary camera.

(Artificial) Neural Networks.—A model that is composed of neurons commonly organized into layers; a set of connected neurons, vaguely inspired in the biological brain neural network.

(Artificial) Neuron.—A system that reads one or more inputs and produces an output. A neuron applies a linear transformation on the inputs, often followed by a nonlinear transformation to produce an output.

Pooling.—A filter that reduces the size of the input data, for example, replacing the value of four adjacent pixels with its maximum or mean value.

Softmax.—A function that provides the probability a sample belongs to each possible class.

Test Data, Test Set.—Samples not used in training but held aside to test the performance of the trained model. Ideally, the test set is used to evaluate only the final model, unlike the validation set that can be used to tune the model during training.

Training.—An iterative process that determines the ideal parameters of a machine learning model.

Training Data, Training Set.—The subset of the data used for training.

Transfer Learning.—A technique that uses information learned in a primary machine learning task (e.g., bird classification) to perform a secondary machine learning task (e.g., fossil classification).

Validation Data, Validation Set.—The subset of the data used to evaluate the training model during model construction and help select hyperparameters.

Weights/Parameters.—The coefficients of a machine learning model. In a simple linear equation, the slope and intercept are the weights of the model. In CNNs, the weights are the convolutional kernel values. The training objective is to find the ideal weights of the machine learning model.

METHODS

In the realm of machine learning techniques, most of the tasks investigated can generally be divided into unsupervised or supervised learning. In unsupervised learning, the user provides data to the algorithm and the algorithm tries to identify patterns present in the data. In supervised learning, the user provides data and corresponding labels and the algorithm tries to learn a function or a relationship to map the data to the labels. In this paper we use supervised learning—we provide data (thin-section images) and labels for training and expect the CNN to provide a relationship between the data and the labels (i.e., the expert-defined fusulinid genus or species).

In general, the reliability of CNN results relates directly to the amount of labeled data used during training. The more examples provided to the CNN, the more improvement occurs in weights used by the model, generating higher-accuracy and more reliable results. The CNN requires examples to recognize the features of each class it attempts to differentiate. The work here focuses on assembling fusulinid thin-section data, and using transfer learning (Pan and Yang 2010) to generate a CNN model to classify fusulinids. Figure 1 shows a simple representation of the transfer learning process. We provide more details of transfer learning in the next subsection.

Fusulinids are an order of large benthic foraminifera abundant during the Pennsylvanian and Permian, and became extinct at the end of the Permian. Like other large benthic foraminifera, fusulinids developed a carbonate test that was internally divided into a series of complex chambers. As the fusulinids grew, new chambers were added along the longitudinal axis covering previous chambers in involute coiling. Fusulinids ranged in size from millimeters to centimeters and can have spindle-like, subcylindrical, or globose shapes. Genera and species are distinguished using a combination of morphological features such as test shape, test wall microstructure, and the arrangement and complexity of internal features (Thompson 1964).

Accurate identification of a fusulinid relies on attributes observable from an oriented section exposed along the long axis of the (prolate spheroid-shaped) fusulinid, bisecting the center. A transverse section is useful, but the longitudinal section is essential (Fig. 2). Both sections reduce the complex internal morphology of fusulinids to 2D views that can be easily imaged. Because fusulinid workers have used these oriented sections for years, an extremely large number of specimens oriented in the same manner exist in legacy collections in museums, although access to such images can be challenging. Thin-section collections, however, commonly consist not only of individual specimens of well-oriented longitudinal sections, but also thin-sections of fossil-bearing rocks in which cuts through specimens are randomly oriented and thus yield apparently different sizes and shapes. In this initial work, the training set contained only those thin-sections with well-oriented longitudinal cuts.

Our fusulinid dataset comprises original fusulinid thin-sections from Waddell (1966) housed at the Sam Noble Museum at the University of Oklahoma (OU) imaged through modern digital photography. The Waddell collection comprises four different Pennsylvanian fusulinid genera: *Beedeina (Fusulina*), *Wedekindellina*, *Triticites*, and *Fusulinella*. Samples from the American Museum of Natural History (AMNH) acquired through the iDigBio portal, an important initiative in digital access to biological collections, provided three additional Permian genera: *Parafusulina*, *Pseudofusulina*, and *Schwagerina*. Figures from Thompson (1954) and Wahlman (2019) provide additional samples of *Beedeina* (*Fusulina*), *Fusulinella*, *Pseudofusulina*, *Pseudoschwagerina*, *Schwagerina*, *Wedekindellina*, and *Triticites*. We also extracted data from Williams (1963), Stevens and Stone (2009), Kobayashi (2012), Kossovaya et al. (2016), Kobayashi and Furutani (2019), Barrick and Wahlman (2019), and Wahlman (2019). Data from the Waddell collection and AMNH were available as independent image files. Data from the other sources were extracted from the figures in the published papers. The images were prepared to be used as input for our models by simply selecting the desired specimen from the digital publication file and saving the images such that the specimen was the main object in the image. We also centralized the specimens and tried to maintain relative sizes, although the microscopic scales reported by the original publications were most likely lost in our cropping and centralization process. Although our data acquisition process was not exhaustive, the data acquired represents some of the more representative fusulinids in the literature—or at least those most readily accessible in our search. However, the dataset we assembled is biased towards fusulinids from the southwestern United States owing to the ready availability of these data in the museum collections we were able to access. We discarded collections of fusulinid genera represented by fewer than 10 different specimens. Differences in thin-section image properties (e.g., background color), and data quality (e.g., original image dimensions) increased the difficulties encountered for training.

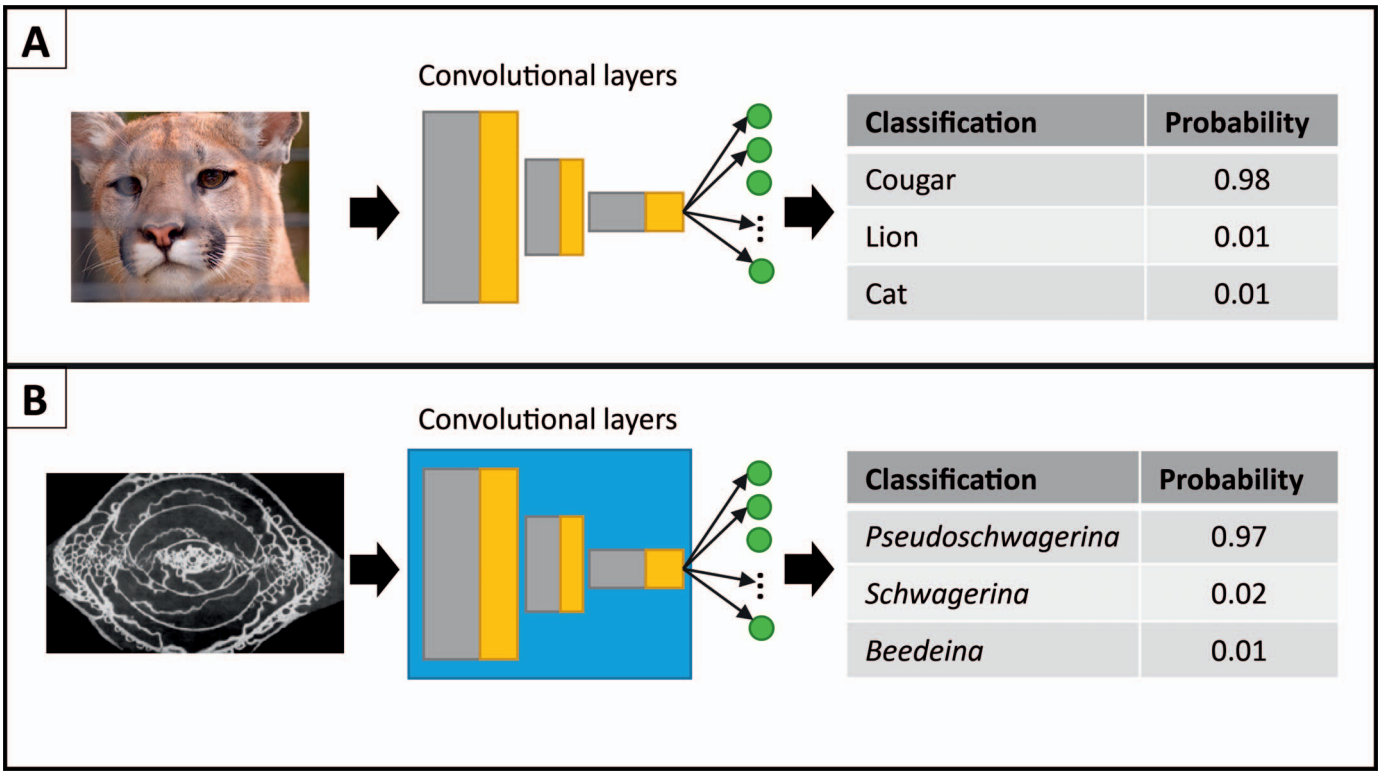

FIG. 1.—Visual representation of the transfer learning process. A CNN is trained on the primary task, generally containing many (millions) of samples. We generically represent convolutional and pooling layers with gray and golden rectangles whereas green circles represent densely connected neurons, commonly used in the classification layers. **A**) "Primary task" in this case represents an image from the ImageNet dataset going through a generic CNN model (convolutional layers and classification layers) trained on the same dataset. The CNN model then outputs the probability of the image belonging to one of the thousands of classes of the ImageNet. **B**) "Secondary task" uses the weights learned by the convolutional layers on primary task using the blue rectangle to represent weights learned on the primary task. We then train a new classification model. The *Pseudoschwagerina* on B is modified from Williams (1963, pl. 8, fig. 3).

Table 1 shows the data available for the fusulinid experiment. Table 2 shows the data split into train, validation, and test sets. More details on the references of the sources of the data used are provided in the Online Supplemental files.

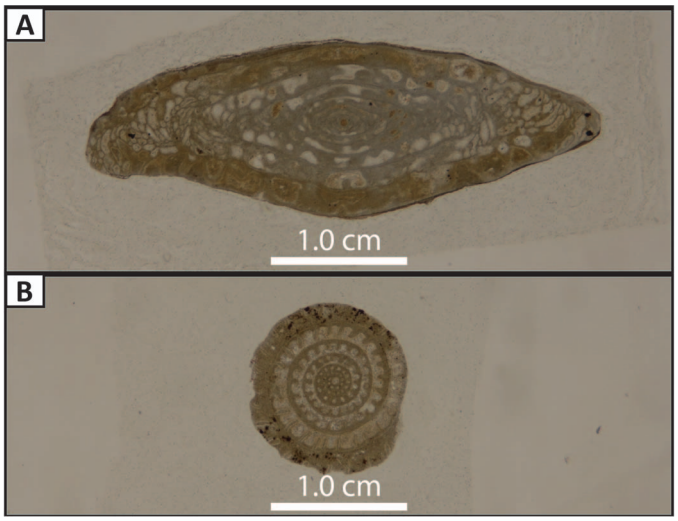

FIG. 2.—Thin-sections with different orientations from the analyzed collection: **A**) *Beedeina mutabilis* (OU 9286) with a longitudinal cut. **B**) *Beedeina mutabilis* (OU 9287) with a transverse cut.

Transfer Learning and Data Augmentation

Transfer learning can be used to address the shortage of sufficient domain-specific training data (Carranza-Rojas et al. 2017). In transfer learning, the learned characteristics of a base model trained on a primary dataset and task are reused for a secondary task (Yosinski et al. 2014). Therefore, layers previously trained with a substantial volume of labeled data can be used to address different objectives. Thus, a CNN model trained to identify the images of the ILSVRC can be used to classify photomicrographs of fusulinid by genera with the help of transfer learning (Fig. 1). In a study analyzing medical image data, Tajbakhsh et al. (2016) found that results achieved with transfer learning equaled or exceeded the quality of results from training a CNN model from scratch (with randomly initialized weights). Yosinski et al. (2014) concluded that using transfer learning on subsets of ILSVRC classes perform better than training CNN models from scratch. Examples of transfer learning include Carranza-Rojas et al. (2017) for herbarium specimens, Esteva et al. (2017) for skin cancer classification, Gomez Villa et al. (2017) for camera-trap images, Pires de Lima et al. (2019a) for lithofacies classification, Valan et al. (2019) for classification of insects, and Marchant et al. (2019) for down-core foraminifera. Pires de Lima et al. (2019b) showed examples of transfer learning using different geological images.

Deep neural networks exhibit a cascading pattern in which the output of one processing layer provides input to the next layer of the model. When trained on datasets of natural images, the first layers of CNN models learn features that resemble either color blobs or some variation of textures. This behavior is so common in CNN models that the analysis is reevaluated every time the initial layers learn any other image characteristics (other

TABLE 1.—*Genus and source for the images used in this experiment.*

Genus	Source
Beedeina	Alexander (1954), Waddell (1966), Barrick and Wahlman (2019), Wahlman (2019)
Fusulinella	Waddell (1966), Wahlman (2019)
Parafusulina	iDigBio, Stevens and Stone (2009), Kobayashi (2012)
Pseudofusulina	iDigBio, Kossovaya et al. (2016), Kobayashi and Furutani (2019), Thompson (1954)
Pseudoschwagerina	Thompson (1954), Williams (1963), Wahlman (2019)
Schwagerina	iDigBio, Williams (1963), Stevens and Stone (2009), Wahlman (2019)
Triticites	iDigBio, Williams (1963), Waddell (1966), Kobayashi and Furutani (2019), Wahlman (2019)
Wedekindellina	Waddell (1966), Barrick and Wahlman (2019), Wahlman (2019)

TABLE 2.—*Number of samples per class in each set. Note "class" here is used in the machine learning not the biological sense.*

Class	Training	Validation	Test	Total
Beedeina	61	9	18	88
Fusulinella	10	2	3	15
Parafusulina	14	3	5	22
Pseudofusulina	28	5	9	42
Pseudoschwagerina	17	3	5	25
Schwagerina	42	6	12	60
Triticites	49	7	14	70
Wedekindellina	14	2	4	20

than colors or texture) resulting in a transition from general to specific features learned by the model (Yosinski et al. 2014). This behavior explains why CNN with good performance on the ILSVRC (e.g., Krizhevsky et al. 2012; Simonyan and Zisserman 2014; Szegedy et al. 2014, 2015) can be successfully retrained for new, field-specific classification problems (e.g., Carranza-Rojas et al. 2017; Esteva et al. 2017; Gomez Villa et al. 2017; Norouzzadeh et al. 2018). As the layers become more specific to the dataset the deeper they are in the model (i.e., closer to the output of the CNN than the input), some workers find it useful to extract only these more general image features. Kong et al. (2016) studied examples of when the features of an ILSVRC pre-trained model can be used without modifications to extract features from pollen data. Another powerful approach is to fine tune the ILSVRC model weights by updating them with training data from the secondary task. Here, we apply three training modes: feature extraction, fine tune, and randomly initialized weights. Feature extraction "locks" (or "freezes") the pre-trained layers extracted from the primary models. Fine tuning begins as feature extraction, with the primary model frozen, but eventually allows all the layers of the model to learn. Randomly initialized weights mode starts the entire model with randomly initialized weights and all the weights are updated during training. Randomly initialized weights is not a transfer learning process, just ordinary training. For the sake of standardization, all modes train the model for 100 epochs. In fine tuning, where part of the model is frozen, we first train for 50 epochs. Then we allow all layers of the model to learn for another 50 epochs. We use five different well known CNN models: VGG19 (Simonyan and Zisserman 2014), InceptionV3 (Szegedy et al. 2015), MobileNetV2 (Sandler et al. 2018), ResNet50 (He et al. 2016), and DenseNet121 (Huang et al. 2016) originally trained on ILSVRC. We use complete CNN models, substituting their last fully connected classification layers with our "top model". The top model comprises an average pooling, followed by one fully connected layer with 512 neurons, a dropout layer (Srivastava et al. 2014) used during training, and a final fully connected layer with a softmax output where the number of neurons depends on the number of classes for the task (eight genera, eight neurons). The models are trained using Keras (Chollet et al. 2015), with TensorFlow as its backend (Abadi et al. 2016). When kernels are initialized, we use the Glorot uniform (Glorot and Bengio 2010) distribution of weights. The models are optimized using a stochastic gradient descent with a learning rate of $1.0e^{-3}$ and momentum of 0.9 to minimize the categorical cross entropy loss. Rather than trying to find the best accuracy possible, our objective is to show the behavior of different CNN models and training modes using fusulinid images from thin-section data. Therefore, we choose to keep the hyperparameters fixed, as described for all experiments. We use a NVIDIA GeForce RTX 2060 for the experiments. Even though transfer learning provides a powerful approach to address the problem of an insufficient amount of training data, and has been successfully implemented in different fields, the relatively small number of digitized thin-sections available for this work created challenges in assembling the training set. Recent examples using transfer learning for image classification employed training datasets of 10^5 images (Carranza-Rojas et al. 2017; Esteva et al. 2017; Gomez Villa et al. 2017). In contrast, we relied upon 10^2 original images of fusulinid specimens, three orders-of-magnitude smaller than other studies using transfer learning.

Owing to this limited dataset for training the CNN, we used a common bootstrap process to generate pseudo-samples using the available images. The population was augmented by simple data rotation. Each longitudinally aligned original image was rotated through a range of angles $\pm 5°$about the horizontal axis, as well as flipped about the horizontal and vertical axes to expand the training data set. Such approaches increase the number of images that could be used for training and help in the generalization of the model. Therefore, to facilitate training and reduce the chance of overfitting, we augment the training and validation data using Bloice et al's. (2019) Augmentor tool, as well as Keras' (Chollet et al. 2015) generators pipelines.

RESULTS

We fit five different models using three different training modes, effectively performing 15 experiments. Figure 3 shows an example of the loss and accuracy evolution during fine tune training of InceptionV3. Plots like Figure 3 are useful to investigate whether or not the models are overfitting the training data. Because of the many parameters CNN usually have, it is possible for the model to simply remember all the training data and have a poor generalization performance, i.e., poor performance when the model classifies new data. Ideally, the training and validation set curves should be close to each other, although in the great majority of cases the models exhibit a better performance on the training set than on the validation set. Figure 3 shows that the model starts to overfit in the first training stage, when part of the model is frozen, and validation accuracy begins to increase only on the second stage of fine tune training, when all the layers in the model are allowed to change their weights. The results (Fig. 3) show that the model is indeed overfitting the training set. The accuracy for the training set is 1.0 on the final epochs and the loss is very close to zero, whereas validation set metrics are not perfect. With the current implementation, more data would help prevent overfitting.

Because the validation set is used during training, a better evaluation of the model's generalization is obtained using the test set. The test set is never used during training and provides the expected performance of the model on data of the same quality, but that were never seen by the model. Figure 4 shows the test set accuracy of all the models and training modes analyzed. Finer details for the best performances for each one of the models are presented in Table 3. Although the models are overfitting the training data, their performance on the validation and test set is appropriate, producing high levels of accuracy.

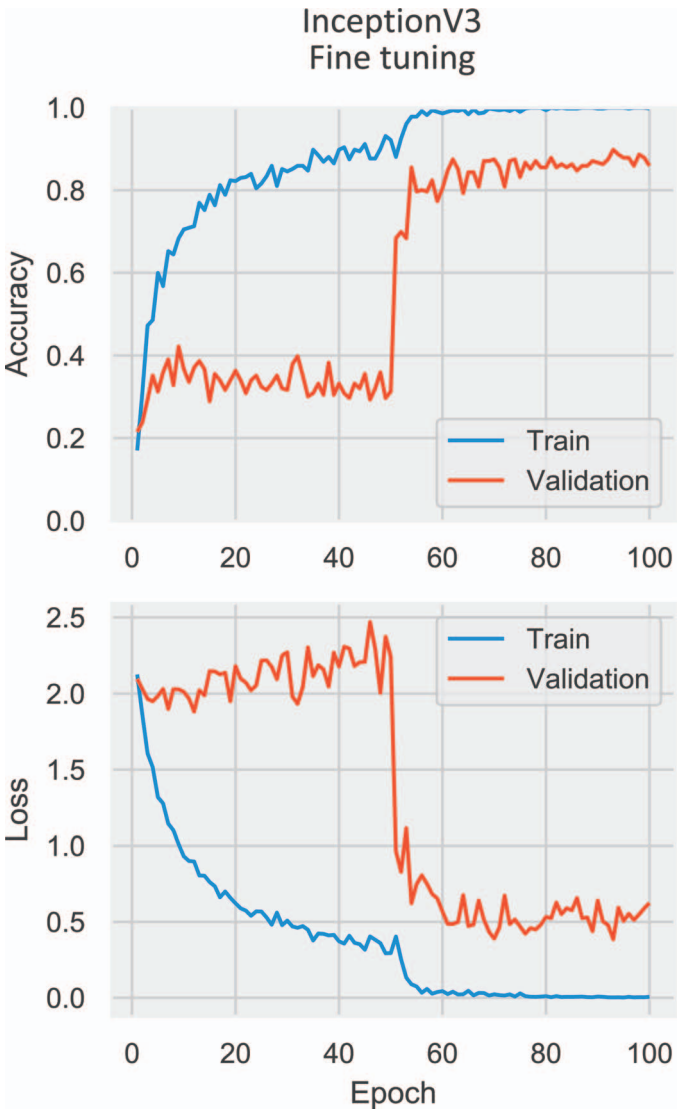

FIG. 3.—Loss and accuracy evolution of train and validation set during training InceptionV3 in the fine tune mode.

Figure 5 shows the confusion matrix computed on the test set using InceptionV3 trained on the fine tune mode. Confusion matrices are helpful to summarize the differences between the classifications performed by the model and the classification performed by the paleontologists. Figures 6 and 7 show examples of fine-tuned InceptionV3 classification of images in the test set and provide more details of the strengths and weaknesses of the mode. The loss and accuracy evolution during training, the confusion matrix, and the classification of the test set of all models and training modes are available in the Online Supplemental Files.

The results (Figs. 5–7; Online Supplemental Files) indicate that the misclassifications appear to be attributable to a combination of specimens' attributes, and photograph characteristics. For example, the Inception V3 fine-tuned model classified one *Beedeina eximia* as *Parafusulina*. The *Beedeina eximia* specimen is narrower perpendicular to the axis of coiling than most of the *Beedeina* specimens used for training. We also observed a tendency of models to misclassify *Triticites* into *Beedeina*, but not the opposite (*Beedeina* samples classified as *Triticites*). Particularly, the models commonly classify one specimen of *Triticites bensonensis*, one *Triticites ellipsoidalis*, and one *Triticites mcgrewensis* as *Beedeina*. The wall structure of *Beedeina* differs considerably from that of *Triticites*; *Beedeina* has four-layered simple walls, whereas *Triticites* has only two, a dark tectum layer and a thick complex keriothecal layer (Thompson 1964). This difference is sufficiently large that distinguishing between the Middle Pennsylvanian *Beedeina* and the Late Pennsylvanian *Triticites* is feasible using only fragments of fusulinid (e.g., from well cuttings). Although we are unsure why the models consistently misidentify these specimens, these results indicate that the model does not employ wall structure for the classification. Note that the wall microstructure occupies only a small area of the total image—potentially unresolved by the technique, which might explain the resultant confusion in classification. The models likely rely on more obvious features such as shape and chamber complexity than wall structure for the classification. A possible alternative to force the models to focus on wall structure would be to modify the training data so that the samples consist only of images of details of wall structure.

DISCUSSION

Despite the difference between the classification of ILSVRC's natural images and the thin-section fusulinid classification task, the CNN models trained on ILSVRC learned to extract features useful for fossil classification. Figure 4 and Table 3 show how different training modes can affect the model's performance and how fine tuning a model previously trained on a large and complex dataset such as the ILSVRC outperformed other training modes. Loss and accuracy for training and validation figures in the Online Supplemental material show that, in general, the feature extraction training mode seems to overfit the data very quickly. Randomly initialized weights validation metrics apparently starts plateauing with a high level of overfitting as well. Thus, updating the initial ILSVRC-based weights on the CNN models with fine tuning is more effective than training a model with randomly initialized weights, except for VGG19 in our case. As the ILSVRC is a complex task in which samples in the same class are very different from each other, the models need to learn very effective transformations to properly differentiate the classes. Because thin-sections have specific characteristics that differ from more common natural images, such transformations must be updated for a proper thin-section classification. Curiously, fine tuning VGG19 with our choice of hyperparameters led the model to a local minimum and degraded highly accurate results found using feature extraction for such a model. It is likely that both a hyperparameter search or more samples could help prevent degrading of results for the model, but such analysis is beyond the scope of this study.

To our knowledge this is the first study providing details on the use of thin-sections, commonly used in fusulinid biostratigraphy, as input for a CNN model useful for identification of microfossils. Our dataset comprises thin-section images from different sources and with differing qualities, with many of the samples simply extracted from published literature. In the approach we use here, the final user can simply provide an adequate image and the CNN model would output the probability of assignment of that specimen to a fusulinid genus—given that the model can choose only from the classes (biological genus in this case) on which it was trained. This study differs from Kong et al. (2016) because we use 2D thin-section images (not 3D stacked images), and our process uses the complete image during training and testing (not selected patches). We also achieved similar accuracy (89%) in our best performing results, differentiating from among more classes (eight genera) than Kong et al. (2016) (three species). Even lacking an extensive database of images, the methodology we applied achieved high levels of accuracy. Although we limited this study to classification of genera, the proposed methods and workflow are applicable to classification of fusulinid species. Groves and Reisdorph (2009) used multivariate morphometry to show that statistically significant separation of species of *Beedeina* is possible. With appropriate data and more samples per species, our CNN methodology should enable classification to high levels of accuracy because *Beedeina* has clearly morphologically distinct

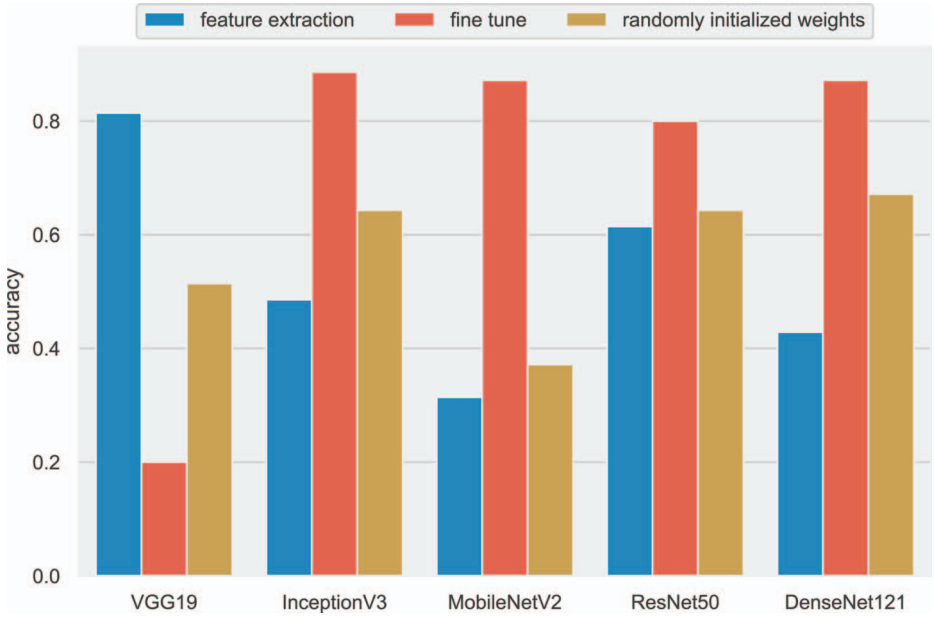

FIG. 4.—Test set accuracy for the five different models using three different training modes.

species. However, in fusulinid paleontological work to date, some disagreement exists in species classification owing to very slight morphological differences. If ambiguity in classification exists amongst the paleontologists, then the CNN will similarly have trouble providing discrete classification.

Unlike a human interpreter who relies upon not only specimen visual characteristics but also additional contextual information to perform taxonomic classifications, the CNN operates from no prior knowledge of specific attribute analysis, and performs the classification based on image characteristics alone. This observation also implies that a CNN model, at least with this current implementation, cannot be used to define a new taxonomic division (e.g., a new species), although it may aid separation of specimens that do fit into existing species. The set of transformations created by the CNN are abstract and do not rely on specific phylogenetic attributes; rather, the rules are akin to a cascading set of filters. But because the CNN models have many such filters, it is often difficult to discuss the interpretability of CNN models. CNN interpretability by itself is a topic in research (e.g., Olah et al. 2017, 2018). When analyzing a new image, the CNN model, as implemented in this study, will generate a set of probabilities that a particular image belongs to the CNN's learned classes, never declaring the image is none of the pre-defined classes. In contrast, some studies make use of CNN-generated features to decide whether or not a new image is sufficiently similar to the data used for training (e.g., Schroff et al. 2015; Staar et al. 2019), which perhaps could have applications for a quantitative measurement of difference between species. Nonetheless, the methodology we implement here can be readily generalized and will improve as new images are digitized and made available to the scientific community. Considering that different taxonomic divisions request different attribute analysis—e.g., during the interpretation of conodonts, specimen surface texture is not as important as caudal point and rostral point (e.g., Hogancamp and Manship 2016)—we envision that CNN techniques will develop more significant modifications as they are applied to other taxonomic groups.

TABLE 3.—*Accuracy for the highest performing training mode for each one of the models.*

Model	Mode	Accuracy
VGG19	feature extraction	0.81
InceptionV3	fine tuning	0.89
MobileNetV2	fine tuning	0.87
ResNet50	fine tuning	0.80
DenseNet121	fine tuning	0.87

Although our approach is similar to recent studies employing transfer learning in image classification (e.g., Carranza-Rojas et al. 2017; Esteva et al. 2017; Gomez Villa et al. 2017), the work we present achieves highly reliable fossil classification using a limited domain-specific dataset three orders-of-magnitude smaller than used in these referenced studies. Kong et al. (2016) used data acquired with a fluorescence microscope with a form of structured illumination to produce high-resolution, 3D image stacks. In contrast, the data we use here was acquired through photomicrographs (2D) taken with an inexpensive and easily available consumer camera and lens, as well as many samples simply obtained by searching published literature, meaning the input data set spanned images of varying vintage and quality. We assume standard preparation of the fusulinid samples in all original studies using methods that have been in use for decades by paleontologists. Because we used such standard image data, we predict the methodology used in this paper has potential for wide applicability and rapid deployment with minimal start-up costs. As more image data are digitized, the technique we use can be applied without the need for laboratory-specific tools and knowledge, which represents a significant improvement over previous approaches requiring specialty image acquisition for CNN (e.g., Zhong et al. 2017). In fact, many of our samples were simply extracted from online (commonly scanned analog) versions from the published literature, making our dataset somewhat irregular.

As digitization of legacy data accelerates, the approach presented here will improve with more detailed image processing. Image segmentation techniques can be used to clip the thin-sections containing significant presence of biotic or abiotic components (noise) besides the organism being analyzed; this will help both in the CNN training and in sequential sample classification. With more data available, object detection—the computer vision task to detect occurrences of objects of different classes (Szegedy et al. 2013; Agarwal et al. 2018; Zhao et al. 2018)—can be applied, increasing the potential of paleo-tailored CNNs in the identifica-

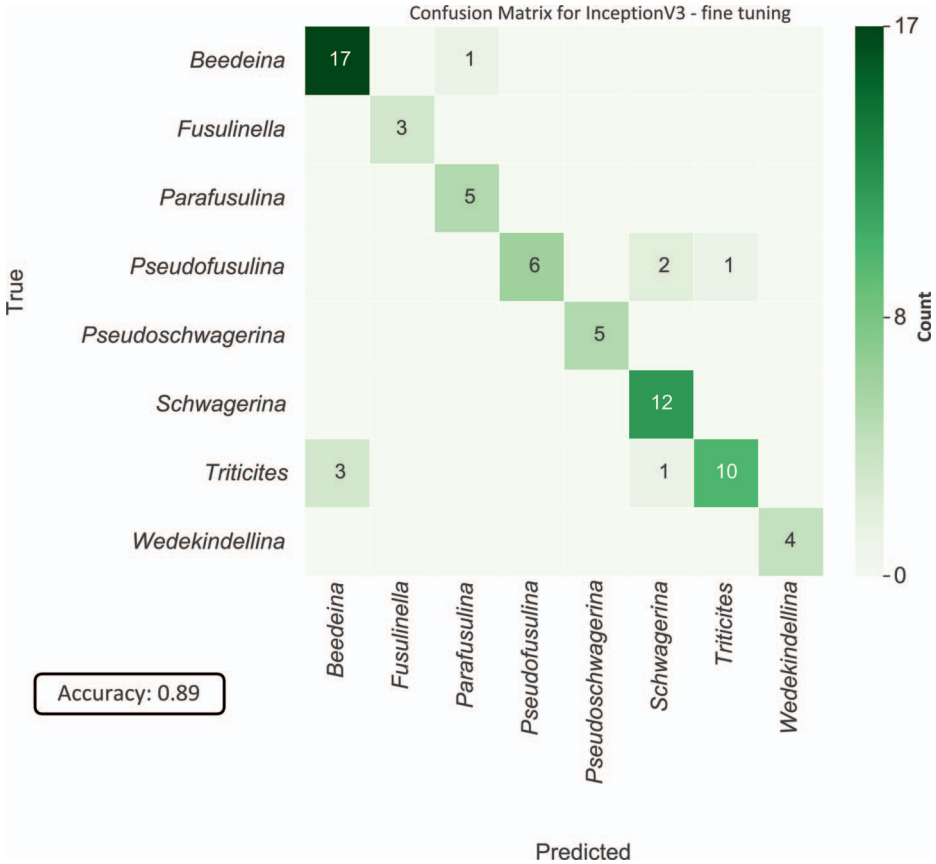

FIG. 5.—Confusion matrix for InceptionV3 trained in the fine tune mode. The confusion matrix shows the expert provided labels versus the model predicted labels. A perfect agreement between model and expert yields a matrix with values only on the main diagonal. Zero values are omitted.

tion of varying taxa captured in the same sample. The technique we demonstrated here is generalized and modified to suit the identification of different fossil groups, such as conodonts, ostracods, ammonites, and others, as long as the specimen can be classified with a 2D representation (thin-section or comparable digital image).

However, we acknowledge that CNNs may be harder to apply to other fossil groups or to provide more detailed boundaries for Foraminifera classification. For example, in trilobites it is unusual to find a complete specimen; in most cases the membrane connecting the thorax, cephalon, and pygidium will deteriorate, causing the exoskeleton to fall apart. This leads to the need to make identifications using fragmented and isolated sclerites. Using trilobite sclerite images as a means of training and classification would likely be more challenging for CNN projects than a dataset with complete specimens, such as those available for other fossil groups. Trilobite sclerites are problematic because they are single pieces of a trilobite; however, different orientations of thin sections of the Foraminifers pose a similar problem (e.g., Fig. 2). Different cross sections of a fusulinid specimen will produce different appearances and no cross section considers external characters. However, the results we obtained with CNN models trained with oriented sections demonstrate that the approach can be useful. Moreover, CNN performance for the classification of insects provides promising results using datasets composed of body parts (e.g., Valan et al. 2019)

A major problem with Foraminifers is the gradation of morphotypes between species. Usually such gradation is addressed by a paleontologist examining a series of samples spanning one form to the other, and choosing an appropriate boundary between species. The CNN model might miss identification of such a boundary using the methodology described here.

As the CNN models are trained with expert-labeled data, such expertise is captured in the model's weights in the deep neural network. Therefore, a CNN model, trained on different collections and having input from different paleontologist experts, provides a means of sharing collections and interpretations across great distances. A fusulinid expert working in the United States can help train a CNN fusulinid classifier and such a model (and an abstract form of the expert knowledge) can be used in Asia with no significant cost; in the meantime, researchers working with *Saccorhytus* fossils in China (Han et al. 2017) can train another CNN to classify their data. Such ease in the exchange of knowledge can help validate interpretations of data globally.

If we are able to capture and mix different paleontological expertise (training CNNs to identify a wide range of taxa), such models can be helpful to identify specimens that might have previously been misclassified. The combination of CNN as an easy-to-use but highly accurate tool and the digitization of stored paleontological samples can provide a rapid method to bring collections residing for decades in museum drawers into use. With easy access to these valuable data, the community can then apply modern statistics to better analyze spatial and temporal distributions, construct more precise assemblages, or simply track evolutionary trends more effectively. We should not discount the "discovery" component; museums commonly have a special exhibition of a fossil or bone that was collected decades ago and was only now identified as a new genus/species; this activity remains in the realm of the expert paleontologist.

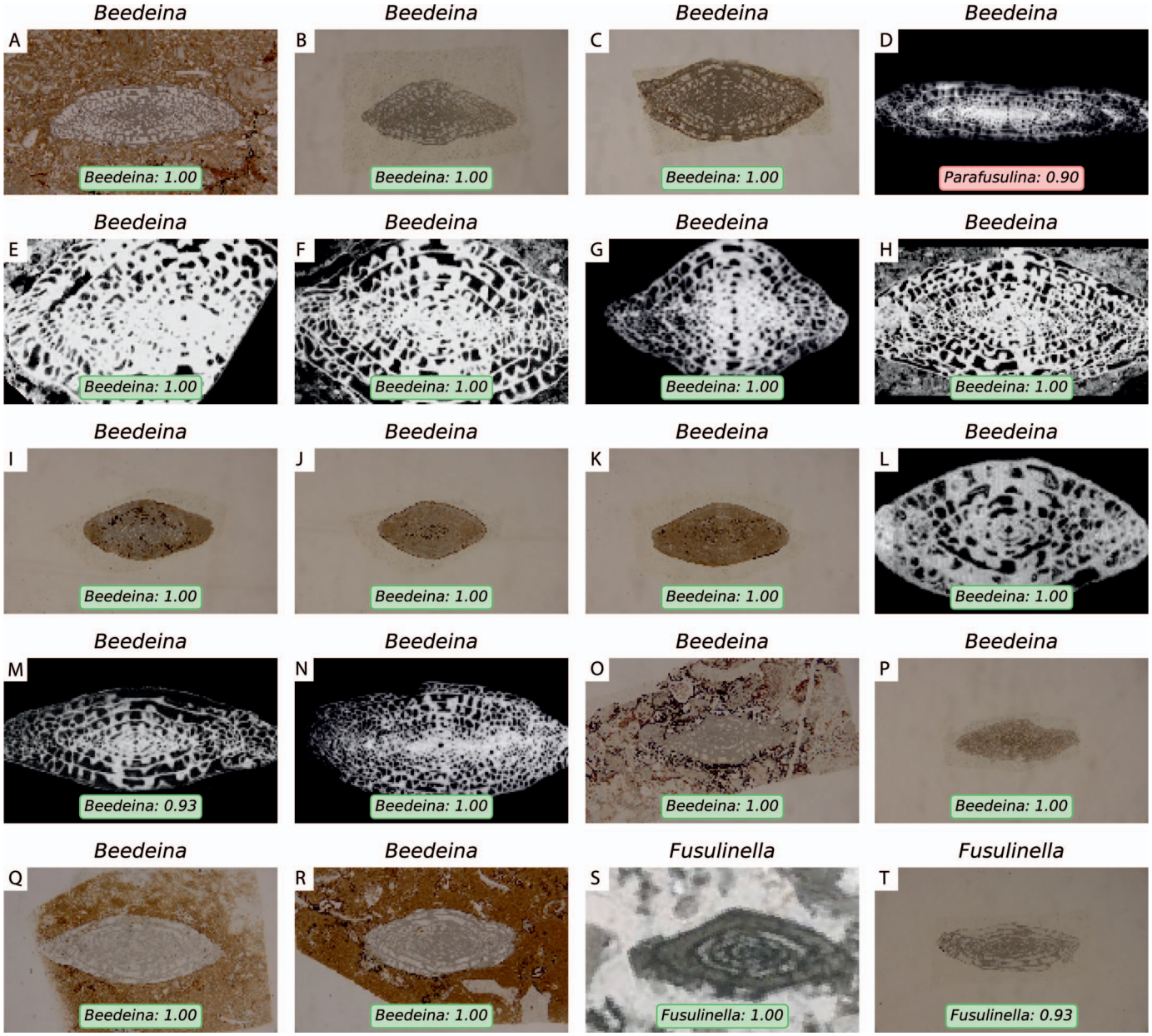

FIG. 6.—Examples of images in the test set classified by InceptionV3 trained on the fine tuning mode. The titles of each of the images are the classification provided by the paleontologists in their original publication, while the text boxes inside the thin-section images are the classification provided by the CNN model. The text box is green when the model assigned the same class (biological genus) as the paleontologist, and red otherwise. The value in the text box shows the probability assigned by the CNN model for that class. **A**) *Beedeina aff. F. whitakeri Stewart* (OU 9333). **B**) *Beedeina cf. F. novamexicana Needham* (OU 9300). **C**) *Beedeina euryteines Thompson* (OU 9306). **D**) *Beedeina eximia* (Wahlman 2019, fig. 12d). **E**) *Beedeina girtyi* (Barrick and Wahlman 2019, pl. 2, fig. 12). **F**) *Beedeina girtyi* (Barrick and Wahlman 2019, pl. 2, fig. 13). **G**) *Beedeina girtyi* (Wahlman 2019, fig. 12g). **H**) *Beedeina haworthi* (Barrick and Wahlman 2019, pl. 2, fig. 9). **I**) *Beedeina insolita Thompson* (OU 9279). **J**) *Beedeina insolita Thompson* (OU 9280). **K**) *Beedeina insolita Thompson* (OU 9281). **L**) *Beedeina insolita* (Wahlman 2019, fig. 12v). **M**) *Beedeina leei* (Wahlman 2019, fig. 12s). **N**) *Beedeina megista* (Wahlman 2019, fig. 12b). **O**) *Beedeina mutabilis Waddell* (OU 9283). **P**) *Beedeina plattensis Thompson* (OU 9297). **Q**) *Beedeina sp*. (OU 9363). **R**) *Beedeina sp*. (OU 9366). **S**) *Fusulinella aff. F. devexa* (OU 9970). **T**) *Fusulinella dakotensis Thompson* (OU 9276).

CONCLUSIONS

This paper outlines a highly effective and inexpensive approach to show that CNNs can correctly identify fusulinid specimens to at least genus level with a very high success rate. The most evident drawback of the methodology we apply here is its current dependency on a large quantities of data to generate robust classification models. The dataset we use was created with relatively few original images, even though it uses images from different sources. Notwithstanding, our tests show that CNNs can correctly identify fusulinid specimens to genus level with a significantly high probability when compared to the other taxonomic classification options. With access to more labeled data, training can be improved, enabling the generation of a model sufficiently robust to overcome complications such as the presence of more than one specimen, geologic noise (e.g., background rock matrix), and other issues. The move towards

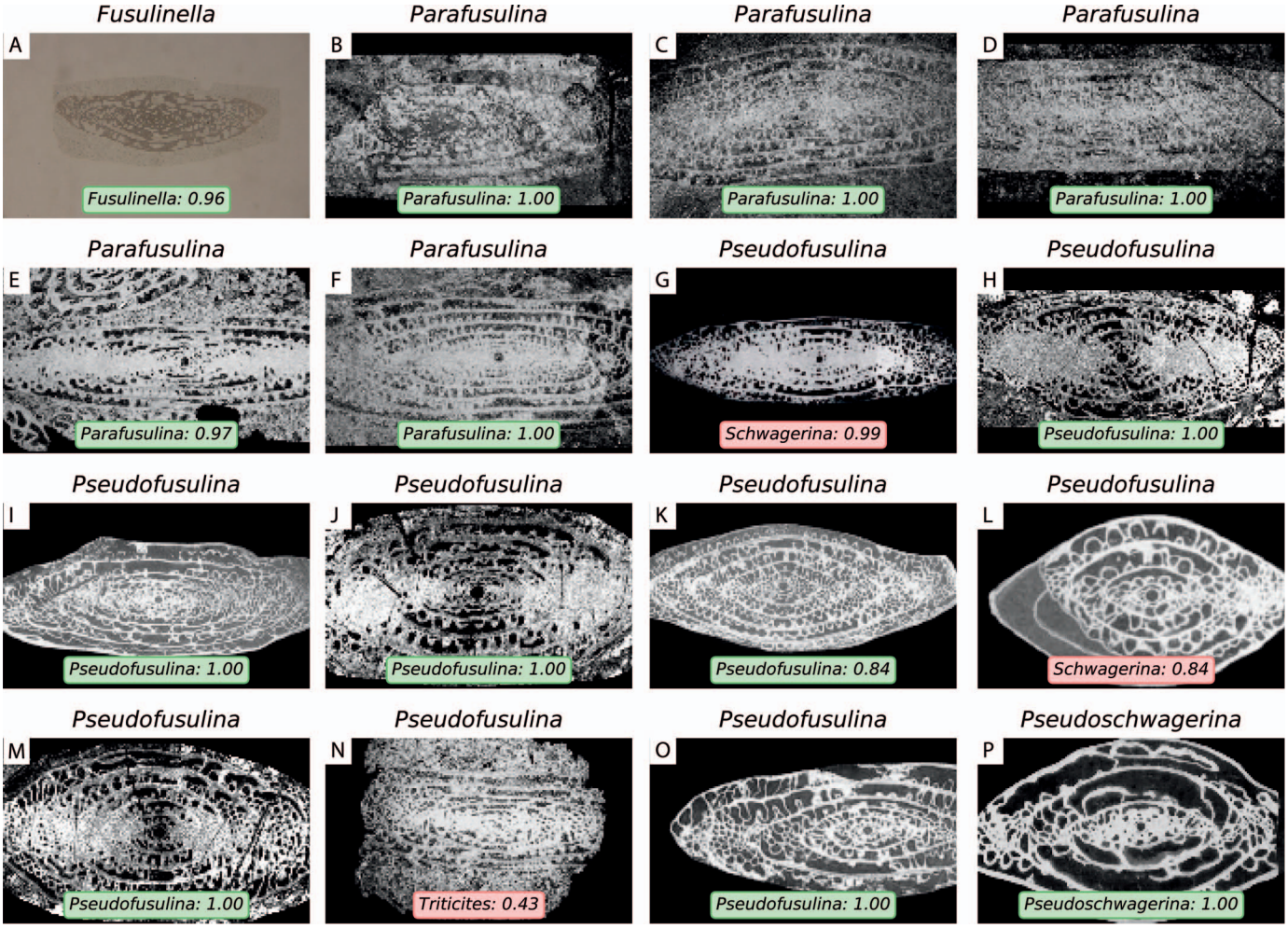

FIG. 7.—More examples of images in the test set classified by InceptionV3 trained on the fine tuning mode. The titles of each of the images are the classification provided by the paleontologists in their original publication, while the text boxes inside the thin-section images are the classification provided by the CNN model. The text box is green when the model assigned the same class (biological genus) as the paleontologist, and red otherwise. The value in the text box shows the probability assigned by the CNN model for that class. **A**) *Fusulinella vacua Waddell* (OU 9273). **B**) *Parafusulina australis* (AMNH-FI-28263). **C**) *Parafusulina erratoseptata* (AMNH-FI-28267). **D**) *Parafusulina guatemalaensis* (AMNH-FI-28269). **E**) *Parafusulina peruana* (AMNH-FI-27087). **F**) *Parafusulina subrectangularis* (AMNH-FI-28281). **G**) *Pseudofusulina huecoensis* (Wahlman 2019, fig. 23b). **H**) *Pseudofusulina isomie* (Kobayashi 2019, fig. 8, 1). **I**) *Pseudofusulina loringi* (Thompson 1954, pl. 41, 1). **J**) *Pseudofusulina nalivkini* (Kobayashi 2019, fig. 3, 15). **K**) *Pseudofusulina nelsoni nelsoni* (Thompson 1954, pl. 44, 6). **L**) *Pseudofusulina nelsoni opima* (Thompson 1954, pl. 44, 9). **M**) *Pseudofusulina norikurensis* (Kobayashi 2019, fig. 7, 5). **N**) *Pseudofusulina rasuvilcensis* (AMNH-FI-27083). **O**) *Pseudofusulina robleda* (Thompson 1954, pl. 42, 7). **P**) *Pseudoschwagerina beedei* (Williams 1963, pl. 3, 5).

digitization of biological and paleontological collections at numerous museums will provide the big data enhancement to enable assessment of the CNN methodology for examples of fossils from around the world, and ultimately identification to species level, revolutionizing the use of fossil identification in a host of studies.

Efforts in data digitization are important initiatives to protect scientific knowledge. The approach documented here contributes to such endeavors and aids the use of biostratigraphic data in the scientific community. Biological variation, differences in specimen size, different imaging techniques and other considerations will complicate the automation of the classification process, but can ultimately lead to deeper understanding, and significant enhancement for all work that relies upon fossil identification. Ultimately, such automation does not replace the expert paleontologist, but enables more rapid and efficient implementation of classification tasks, freeing up time and expertise for exploration of more complex interpretations and concepts.

ACKNOWLEDGMENTS

We thank the iDigBio initiative for providing access to the community for biodiversity collections data. Rafael acknowledges CNPq (grant 203589/2014-9) for the financial support and CPRM for granting the leave of absence allowing pursuit of his Ph.D. studies. We thank two anonymous reviewers for input that improved this manuscript.

SUPPLEMENTAL MATERIAL

Data are available from the PALAIOS Data Archive: https://www.sepm.org/supplemental-materials.

REFERENCES

ABADI, M., BARHAM, P., CHEN, J., CHEN, Z., DAVIS, A., DEAN, J., DEVIN, M., GHEMAWAT, S., IRVING, G., ISARD, M., KUDLUR, M., LEVENBERG, J., MONGA, R., MOORE, S., ET AL., 2016, TensorFlow: a system for large-scale machine learning, *in* 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16): USENIX The Advanced

Computing Systems Association, Savannah, GA, p. 265–283; https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf.

Agarwal, S., Ogier Du Terrail, J., and Jurie, F., 2018, Recent Advances in Object Detection in the Age of Deep Convolutional Neural Networks: ArXiv e-prints: arXiv:1809.03193 [cs.CV].

Alexander, R., 1954, Desmoinesian Fusulinids of Northeastern Oklahoma: Oklahoma Geological Survey, Norman, OK, Circular 31, 67 p., http://ogs.ou.edu/docs/circulars/C31.pdf.

Barrick, J.E. and Wahlman, G.P., 2019, Conodont and fusulinid biostratigraphy of the Strawn Group (Desmoinesian, Middle Pennsylvanian) and lower part of the" Wolfcamp Shale"(Missourian–Virgilian, Late Pennsylvanian) in the northern Midland Basin, West Texas: Stratigraphy, v. 16, p. 65–86.

Blagoderov, V., Kitching, I.J., Livermore, L., Simonsen, T.J., and Smith, V.S., 2012, No specimen left behind: industrial scale digitization of natural history collections: ZooKeys, p. 133–46, doi: 10.3897/zookeys.209.3178.

Bloice, M.D., Roth, P.M., and Holzinger, A., 2019, Biomedical image augmentation using augmentor: Bioinformatics, v. 35, p. 4522–4524, doi: 10.1093/bioinformatics/btz259; https://doi.org/10.1093/bioinformatics/btz259.

Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., and Joly, A., 2017, Going deeper in the automated identification of Herbarium specimens: BioMed Central (BMC) Evolutionary Biology, v. 17, p. 181, doi: 10.1186/s12862-017-1014-z.

Chollet, F. and others, 2015, Keras: https://keras.io.

Culver, S.J., 2003, Benthic foraminifera across the Cretaceous–Tertiary (K–T) boundary: a review: Marine Micropaleontology, v. 47, p. 177–226, doi: 10.1016/S0377-8398(02)00117-2.

Dumoulin, V. and Visin, F., 2016, A guide to convolution arithmetic for deep learning: ArXiv e-prints: arXiv:1603.07285 [stat.ML].

Ellwood, E.R., Dunckel, B.A., Flemons, P., Guralnick, R., Nelson, G., Newman, G., Newman, S., Paul, D., Riccardi, G., Rios, N., Seltmann, K.C., and Mast, A.R., 2015, Accelerating the digitization of biodiversity research specimens through online public participation: BioScience, v. 65, p. 383–396, doi: 10.1093/biosci/biv005.

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

Farley, M.B. and Armentrout, J.M., 2000, Fossils in the oil patch: Geotimes, October, p. 14–17.

Farley, M.B. and Armentrout, J.M., 2002, Biostratigraphy becoming lost art in rush to find new exploration tools: Offshore, February, p. 94–95.

Glorot, X. and Bengio, Y., 2010, Understanding the difficulty of training deep feedforward neural networks, *in* Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS'10): Society for Artificial Intelligence and Statistics, Sardinia, Italy, v. 9, p. 249–256.

Gomez Villa, A., Salazar, A., and Vargas, F., 2017, Towards automatic wild animal monitoring: identification of animal species in camera-trap images using very deep convolutional neural networks: Ecological Informatics, v. 41, p. 24–32, doi: 10.1016/J.ECOINF.2017.07.004.

Gooday, A.J., 1994, The biology of deep-sea foraminifera: a review of some advances and their applications in paleoceanography: PALAIOS, v. 9, p. 14, doi: 10.2307/3515075.

Han, J., Morris, S.C., Ou, Q., Shu, D., and Huang, H., 2017, Meiofaunal deuterostomes from the basal Cambrian of Shaanxi (China): Nature, v. 542, p. 228–231, doi: 10.1038/nature21072.

He, K., Zhang, X., Ren, S., and Sun, J., 2016, Identity mappings in deep residual networks, *in* B. Leibe, J. Matas, N. Sebe, and M. Welling (eds.), Computer Vision—ECCV 2016: Springer International Publishing, Cham, p. 630–645.

Hogancamp, N.J. and Manship, L.L., 2016, Comparison of morphometric techniques and the ability to accurately reconstruct the form and distinguish between species of the *Palmatolepis winchelli* group-Conodonta, Upper Devonian: Micropaleontology, v. 62, p. 439–451.

Hsiang, A. Y., Brombacher, A., Rillo, M.C., Mleneck-Vautravers, M.J., Conn, S., Lordsmith, S., Jentzen, A., Henehan, M.J., Metcalfe, B., Fenton, I.S., Wade, B.S., Fox, L., Meilland, J., Davis, C. V, Baranowski, U., Groeneveld, J., Edgar, K.M., Movellan, A., Aze, T., Dowsett, H.J., Miller, C.G., Rios, N., and Hull, P.M., 2019, Endless forams: >34,000 modern planktonic foraminiferal images for taxonomic training and automated species recognition using convolutional neural networks: Paleoceanography and Paleoclimatology, v. 34, p. 1157–1177, doi: 10.1029/2019PA003612.

Huang, G., Liu, Z., and Weinberger, K.Q., 2016, Densely connected convolutional networks, *in* IEEE Conference on Computer Vision and Pattern Recognition (CVPR): IEEE, Honolulu, HI, p. 2261–2269, doi: 10.1109/CVPR.2017.243.

Keçeli, A.S., Kaya, A. and Keçeli, S.U., 2017, Classification of radiolarian images with hand-crafted and deep features: Computers and Geosciences, v. 109, p. 67–74, doi: 10.1016/j.cageo.2017.08.011.

Kobayashi, F., 2012, Late Paleozoic Foraminifers from limestone blocks and fragments of the Permian Tsunemori Formation and their connection to the Akiyoshi Limestone Group, Southwest Japan: Paleontological Research, v. 16, p. 219–243, doi: 10.2517/1342-8144-16.3.219.

Kobayashi, F. and Furutani, H., 2019, Late Early Permian fusulines along Gongendani, South of Mt. Ryozen, Shiga Prefecture, Central Japan: Paleontological Research, v. 23, p. 131, doi: 10.2517/2018PR014.

Koch, G.R., 2015, Siamese Neural Networks for One-Shot Image Recognition: Unpublished M.S. thesis, University of Toronto, Toronto, Canada, 30 p.

Kong, S., Punyasena, S., and Fowlkes, C., 2016, Spatially aware dictionary learning and coding for fossil pollen identification, *in* 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW): IEEE, Las Vegas, NV, p. 1305–1314.

Kossovaya, O.L., Novak, M., and Weyer, D., 2016, Large-sized Early Permian "caninioid" corals from the Karavanke Mountains, Slovenia: Journal of Paleontology, v. 90, p. 1049–1067, doi: 10.1017/jpa.2016.105.

Krizhevsky, A., Sutskever, I., and Hinton, G.E., 2012, ImageNet classification with deep convolutional neural networks, *in* Proceedings of the 25th International Conference on Neural Information Processing Systems, Volume 1: Neural Information Processing Systems NIPS'12, Curran Associates Inc., Lake Tahoe, NV, p. 1097–1105, http://dl.acm.org/citation.cfm?id=2999134.2999257.

Kucera, M., 2007, Chapter Six, Planktonic Foraminifera as Tracers of Past Oceanic Environments, *in* C. Hillaire–Marcel and A. de Vernal (eds.), Proxies in Late Cenozoic Paleoceanography: Developments in Marine Geology, Elsevier, Amsterdam, Netherlands, p. 213–262.

Lake, B.M., Salakhutdinov, R., and Tenenbaum, J.B., 2015, Human-level concept learning through probabilistic program induction: Science, v. 350, p. 1332–1338, doi: 10.1126/science.aab3050.

LeCun, Y., Bengio, Y., and Hinton, G., 2015, Deep learning: Nature, v. 521, p. 436–444, doi: 10.1038/nature14539.

Machine Learning Glossary, 2019, Google Developers website: https://developers.google.com/machine-learning/glossary/#top_of_page. Checked January 2019.

Marchant, R., Tetard, M., Pratiwi, A., and de Garidel-Thoron, T., 2019, Classification of down-core foraminifera image sets using convolutional neural networks: bioRxiv, doi: 10.1101/840926.

Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C., and Clune, J., 2018, Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning: Proceedings of the National Academy of Sciences of the United States of America, v. 115, p. E5716–E5725, doi: 10.1073/pnas.1719367115.

Olah, C., Mordvintsev, A., and Schubert, L., 2017, Feature Visualization: Distill, doi: 10.23915/distill.00007. Checked August 2020.

Olah, C., Satyanarayan, A., Johnson, I., Carter, S., Schubert, L., Ye, K., and Mordvintsev, A., 2018, The Building Blocks of Interpretability: Distill, doi: 10.23915/distill.00010. Checked August 2020.

Pan, S.J. and Yang, Q., 2010, A survey on transfer learningIEEE Transactions on Knowledge and Data Engineering, v. 22, p. 1345–1359, doi: 10.1109/TKDE.2009.191.

Pedraza, A., Bueno, G., Deniz, O., Cristóbal, G., Blanco, S., and Borrego-Ramos, M., 2017, Automated diatom classification (part B): a deep learning approach: Applied Sciences, v. 7, p. 460, doi:10.3390/app7050460.

Pires de Lima, R., Bonar, A., Coronado, D.D., Marfurt, K., and Nicholson, C., 2019a, Deep convolutional neural networks as a geological image classification tool: The Sedimentary Record, v. 17, p. 4–9, doi: 10.210/sedred.2019.2.

Pires de Lima, R., Suriamin, F., Marfurt, K.J., and Pranter, M.J., 2019b, Convolutional neural networks as aid in core lithofacies classification: Interpretation, v. 7, p. SF27–SF40, doi: 10.1190/INT-2018-0245.1.

Ranaweera, K., Harrison, A.P., Bains, S., and Joseph, D., 2009, Feasibility of computer-aided identification of foraminiferal tests: Marine Micropaleontology, v. 72, p. 66–75, doi: 10.1016/J.MARMICRO.2009.03.005.

Ross, C.A., 1967, Development of Fusulinid (Foraminiferida) Faunal Realms: Journal of Paleontology, v. 41, p. 1341–1354.

Roozpeykar, A. and Moghaddam, I.M., 2016, Benthic foraminifera as biostratigraphical and paleoecological indicators: an example from Oligo–Miocene deposits in the SW of Zagros Basin, Iran: Geoscience Frontiers, v. 7, p. 125–140, doi: 10.1016/J.GSF.2015.03.005.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., and Fei-Fei, L., 2015, ImageNet large scale visual recognition challenge: International Journal of Computer Vision, v. 115, p. 211–252, doi: 10.1007/s11263-015-0816-y.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C., 2018, MobileNetV2: inverted residuals and linear bottlenecks, *in* IEEE/CVF Conference on Computer Vision and Pattern Recognition: IEEE, Salt Lake City, UT, p. 4510–4520, doi: 10.1109/CVPR.2018.00474

Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., and Lillicrap, T.P., 2016, One-shot learning with memory-augmented neural networks: ArXiv e-prints, arXiv:1605.06065 [cs.LG].

Schroff, F., D. Kalenichenko, and J. Philbin, 2015, FaceNet: a unified embedding for face recognition and clustering, *in* The IEEE Conference on Computer Vision and Pattern Recognition (CVPR): IEEE, Boston, MA, p. 815–823, doi: 10.1109/CVPR.2015.7298682.

Sevillano, V. and Aznarte, J.L., 2018, Improving classification of pollen grain images of the POLEN23E dataset through three different applications of deep learning convolutional neural networks, S. Rutherford (ed.): PLOS ONE, v. 13, p. e0201807, doi: 10.1371/journal.pone.0201807.

Simonyan, K. and Zisserman, A., 2014, Very deep convolutional networks for large-scale image recognition, *in* Third International Conference on Learning Representations

(ICLR): International Conference on Learning Representations (ICLR), San Diego, CA, http://arxiv.org/abs/1409.1556.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., 2014, Dropout: a simple way to prevent neural networks from overfitting: Journal of Machine Learning Research, v. 15, p. 1929–1958.

Staar, B., M. Lütjen, and M. Freitag, 2019, Anomaly detection with convolutional neural networks for industrial surface inspection: Procedia CIRP, v. 79, p. 484–489, doi: 10.1016/j.procir.2019.02.123.

Stevens, C.H. and Stone, P., 2009, New fusulinids from Lower Permian turbidites at Conglomerate Mesa, Southeastern Inyo Mountains, east-central California: Journal of Paleontology, v. 83, p. 399–404, doi: 10.1666/08-162.1.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., 2014, Going deeper with convolutions, *in* IEEE Conference on Computer Vision and Pattern Recognition (CVPR): IEEE, Boston, MA, p. 1–9, doi: 10.1109/CVPR.2015.7298594.

Szegedy, C., Toshev, A., and Erhan, D., 2013, Deep Neural Networks for Object Detection, *in* C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger (eds.), Advances in Neural Information Processing Systems 26: Curran Associates, Inc., Lake Tahoe, NV, p. 2553–2561, http://papers.nips.cc/paper/5207-deep-neural-networks-for-object-detection.pdf.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z., 2015, Rethinking the inception architecture for computer vision, *in* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR): IEEE, Las Vegas, NV, p. 2818–2826, doi: 10.1109/CVPR.2016.308.

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., and Liang, J., 2016, Convolutional neural networks for medical image analysis: full training or fine tuning?: IEEE Transactions on Medical Imaging, v. 35, p. 1299–1312, doi: 10.1109/TMI.2016.2535302.

Thompson, M. L., 1964, Part C, Protista 2: Treatise on Invertebrate Paleontology, University of Kansas and Geological Society of America, v. 1, p. C358–C436.

Valan, M., Makonyi, K., Maki, A., Vondráček, D., and Ronquist, F., 2019, Automated taxonomic identification of insects with expert-level accuracy using effective feature transfer from convolutional networks: Systematic Biology, v. 68, p. 876–895, doi: 10.1093/sysbio/syz014.

Waddell, D.E., 1966, Pennsylvanian fusulinids in the Ardmore Basin –Love and Carter counties, Oklahoma: Oklahoma Geological Survey Bulletin 113, 128 p.

Wahlman, G.P., 2019, Pennsylvanian and Lower Permian Fusulinid Biostratigraphy of the Permian Basin Region, Southwestern USA, *in* S.C. Ruppel (ed.), Anatomy of a Paleozoic Basin: The Permian Basin, USA, Volume 1: Bureau of Economic Geology Report of Investigations 285, AAPG Memoir 118, p. 167–227.

Williams, T.E., 1963, Fusulinidae of the Hueco Group (Lower Permian) Hueco Mountains, Texas: Bulletin 18, Peabody Museum, Yale University, New Haven, CT, 156 p.

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H., 2014, How transferable are features in deep neural networks?: Advances in Neural Information Processing Systems, v. 27, p. 3320–3328.

Zhao, Z.-Q., Zheng, P., Xu, S.-T., and Wu, X., 2018, Object detection with deep learning: a review: IEEE Transactions on Neural Networks and Learning Systems, v. 30, p. 3212–3232, doi: 10.1109/TNNLS.2018.2876865.

Zhong, B., Ge, Q., Kanakiya, B., Marchitto, R.M.T., and Lobaton, E., 2017, A comparative study of image classification algorithms for Foraminifera identification, *in* 2017 IEEE Symposium Series on Computational Intelligence (SSCI): IEEE, Honolulu, HI, p. 1–8, http://ieeexplore.ieee.org/document/8285164/.

Received 26 November 2019; accepted 1 July 2020.