

Enhancing the automatic facies classification of Brazilian presalt acoustic image logs with SwinV2-Unet: Leveraging transfer learning and confident learning

Nan You¹ and Yunyue Elita Li²

ABSTRACT

Facies classification of image logs plays a vital role in reservoir characterization, especially in the heterogeneous and anisotropic carbonate formations of the Brazilian presalt region. Although manual classification remains the industry standard for handling the complexity and diversity of image logs, it has the notable disadvantages of being time consuming, labor intensive, subjective, and nonrepeatable. Recent advancements in machine learning offer promising solutions for automation and acceleration. However, previous attempts to train deep neural networks for facies identification have struggled to generalize to new data due to insufficient labeled data and the inherent intricacy of image logs. In addition, human errors in manual labels further hinder the performance of trained models. To overcome these challenges, we develop adopting the

state-of-the-art SwinV2-Unet to provide depthwise facies classification for Brazilian presalt acoustic image logs. The training process incorporates transfer learning to mitigate overfitting and confident learning to address label errors. Through a k -fold cross-validation experiment, with each fold spanning more than 350 m, we achieve an impressive macro $F1$ score of 0.90 for out-of-sample predictions. This significantly surpasses the previous model modified from the widely recognized U-Net, which provides a macro $F1$ score of 0.68. These findings highlight the effectiveness of the used enhancements, including the adoption of an improved neural network and an enhanced training strategy. Moreover, our SwinV2-Unet enables a highly efficient and accurate facies analysis of the complex yet informative image logs, significantly advancing our understanding of hydrocarbon reservoirs, saving human effort, and improving productivity.

INTRODUCTION

Borehole imaging is a rapidly developing well-logging technique that offers unrolled, high-resolution images of the borehole walls, known as image logs. These logs provide detailed visual information about the structural, textural, and lithologic properties of subsurface formations, serving various purposes such as fracture identification, in situ stress regime and stratigraphy analysis, and borehole stability control (Prensky, 1999). Ultimately, they contribute to enhanced reservoir characterization, particularly in assessing permeability, porosity, and lithology, which are crucial factors associated with hydrocarbon abundance and extraction feasibility within the reservoir (Lai et al., 2018).

According to Akbar et al. (2000), carbonate reservoirs, which comprise more than 60% of the global oil reserves and 40% of the global gas reserves, stand out as one of the most abundant hydrocarbon reservoirs worldwide. This study focuses on the carbonate reservoirs located in the Brazilian presalt region, which account for 74.88% of the Brazilian national oil equivalent as of April 2023 (ANP, 2023). These reservoirs are situated within a highly complex geologic context, buried beneath a thick salt layer, which presents significant challenges for large-scale production (Branco and de Sant'Anna Pizarro, 2012; da Costa Fraga et al., 2015). Consequently, accurate and detailed characterization of these reservoirs is of the utmost importance for oil and gas production in this region. However, the carbonate reservoirs in this area exhibit notable

Manuscript received by the Editor 8 August 2023; revised manuscript received 14 February 2024; published ahead of production 18 March 2024.

¹Aramco Americas — Houston Research Center, Houston, Texas, USA. E-mail: nan.you@aramcoamericas.com.

²Purdue University, Department of Earth, Atmospheric, and Planetary Sciences, Sustainability Geophysics Project, West Lafayette, Indiana, USA. E-mail: li4017@purdue.edu (corresponding author).

© 2024 Society of Exploration Geophysicists. All rights reserved.

heterogeneity and anisotropy, posing exceptional challenges in accurately estimating rock properties.

With recent advancements in borehole imaging, image logs have emerged as a vital and promising technique for characterizing the sophisticated nature of the Brazilian presalt carbonate reservoirs. Specifically, the establishment of image log facies, based on characteristics such as dip type, dip pattern, and color scheme, allows for the translation of image logs into various rock types. These rock types exhibit distinct sedimentary and structural features, including depositional microfacies, sedimentary structures, bedding sequences, vugs, faults, fractures, and paleocurrent directions (Donselaar and Schmidt, 2005; Wilson et al., 2013; Muniz and Bosence, 2015). Through calibration with cores and conventional logs, the image log facies and their stacking patterns can be further interpreted as lithofacies associations closely tied to specific petrophysical properties, such as permeability, stiffness, and anisotropy (Lai et al., 2018). Therefore, the classification of the image log facies provides a detailed depiction of the distribution of the rock types and their unique depositional, structural, and petrophysical properties along the well trajectory. This sheds light on the heterogeneity and anisotropy of the carbonate reservoirs in close proximity to the wellbore, contributing significantly to our understanding of these complex geologic formations.

Despite the widespread belief among geologists in the oil and gas industry that borehole image logs hold valuable information about crucial petrophysical properties, such as permeability, porosity, and lithofacies, there is currently no established automated procedure to objectively, reliably, and quantitatively predict these properties from image logs. The prevailing industry practice still relies on manual interpretation, wherein geologists subjectively categorize image logs into different facies and offer qualitative facies descriptions encompassing details about lithology, sedimentary textures, paleoflow directions, and the processes of sedimentation and diagenesis (Muniz and Bosence, 2015; Lai et al., 2018). However, manual interpretation has significant limitations due to the large amount of time and workforce it requires, as well as the subjectivity and lack of repeatability in its results. Therefore, the automation of the image log interpretation process is urgent and critical.

Inspired by the remarkable self-learning capacity of machine-learning (ML) models, numerous researchers have used ML techniques for automatic image log interpretation. One approach is to automate the classification process using unsupervised learning. The general workflow involves extracting representative features from the image logs using conventional image processing techniques and then using unsupervised learning algorithms, such as the mean-shift algorithm and self-organizing map, to separate the extracted features into different classes (Hall et al., 1996; Ye et al., 1998; Al-Sit et al., 2015; Yang et al., 2020). Notably, Lima et al. (2019) propose an unsupervised feature extraction approach in which an autoencoder network is trained to automatically encode image data into low-dimensional high-level representations. Although these methods eliminate the need for laborious and time-consuming manual labeling, the resulting class separation may be vague or not align with specific requirements due to the absence of guidance in clustering. To address this issue and incorporate domain-specific knowledge, various deep neural networks (DNNs) have been trained with manual or simulated labels to perform various classification tasks, such as lithology detection (Valentín et al., 2019), fracture identification (Gupta et al., 2019), breakout

detection (Dias et al., 2020), and vuggy facies recognition (Jiang et al., 2021) from image logs. However, a common limitation of the published DNN models is their reliance on synthetic or high-quality image logs with minimal artifacts and noise, raising concerns about their generalizability to new field data. Furthermore, the performance of the DNNs, particularly in terms of generalizability and accuracy, is significantly constrained by the quantity and quality of the available labels.

As mentioned previously, automatic classification approaches have been developed for image logs using unsupervised or supervised learning. The separate classes represent facies in a broad sense, which are distinct rock bodies exhibiting a unique appearance, composition, and texture (Parker, 1984). In addition to facies identification from borehole image logs, extensive studies have been conducted to automatically detect facies from conventional petrophysical logs, such as the porosity, density, and gamma-ray logs, using various ML techniques (Dubois et al., 2007; Hall, 2016; Imamverdiyev and Sukhostat, 2019). Because borehole image logs provide detailed textural information at a millimeter resolution, they have proven to be a valuable complementary data set for facies analysis, especially in complex geologic settings (Basu et al., 2002; Chai et al., 2009). A significant breakthrough in this field was achieved by You et al. (2023), who introduce the first DNN model, specifically a modified U-Net (Ronneberger et al., 2015), for depthwise facies classification of the acoustic image logs from the Brazilian presalt region. For convenience, we will refer to this modified U-Net as Facies-Unet in the following discussion. The researchers carefully defined the facies, taking into account the lamination characteristics, the presence of bioclasts, and the impact of artifacts. Following the definition, a set of manually labeled field data from eight wells and a substantial number of synthetic image patches were prepared for training. The continuous field data were divided into overlapping patches using a 1.3 m sliding window along the depth direction, with a step size of 0.325 m. To avoid data leakage, two 3.6 m continuous sections were extracted from each well, forming an independent 57.6 m test set. The remaining image patches from the data set were split into training and validation sets with a ratio of 8:2. Remarkably, the Facies-Unet achieved an accuracy of 77% for the test set, demonstrating its effectiveness in handling the complex and diverse data set from the investigated region. Trained with manual labels, the Facies-Unet outperformed manual labeling by achieving higher levels of efficiency, resolution, and consistency. However, the researchers observed a slight overfitting issue due to the limited availability of labeled field data and the complexity and diversity of the image logs from that region. They commented that the 57.6 m test set was not sufficient to confirm the model's generalizability. Moreover, the presence of inevitable human errors and inconsistencies in the manual labels hampered further improvements in the model's performance.

To address the limitations identified in the work by You et al. (2023), we propose a series of enhancements to improve the classification performance. First, we replace the U-Net architecture with a state-of-the-art neural network called SwinV2-Unet (Liu et al., 2021, 2022; Cao et al., 2022). The SwinV2-Unet is a pure transformer-based U-shaped model that integrates the advanced multi-head self-attention mechanisms (Vaswani et al., 2017) with the powerful U-shaped structure and skip-connection operations of the U-Net. This combination enables the SwinV2-Unet to learn local-global semantic features efficiently and has demonstrated

exceptional performance in multiorgan and cardiac image segmentation tasks (Cao et al., 2022). The effectiveness of transformers extends to various geophysical problems as well. For instance, Yang et al. (2023) create a multitask model based on foundational transformer and convolutional blocks, producing excellent results in simultaneously predicting the relative geologic time, horizons, and faults from seismic data. Second, we propose a refined training procedure that combines transfer learning and confident learning (CL) algorithms (Northcutt et al., 2021). Transfer learning is used to mitigate the overfitting issue by leveraging the pretrained weights of the Swin Transformer V2 model on a large image database, the ImageNet-1K (Deng et al., 2009). In addition, the CL algorithm is used to identify and prune label errors by estimating the joint distribution between the noisy manual labels and actual labels. To assess the generalizability of our model in an unbiased manner, we conduct a k -fold cross-validation experiment, in which the labeled data set is divided into fivefold, and each fold is used as the test set in turn during the five iterations. The lengths of the test folds vary between 366.6 and 374.4 m. To ensure a fair comparison, we also perform the same k -fold cross-validation experiment on the Facies-Unet. Our new model achieves a high macro $F1$ score of 0.90 for the out-of-sample predictions, surpassing the macro $F1$ score of 0.68 obtained by the Facies-Unet. This notable improvement highlights the enhanced accuracy and generalizability of our new model, which greatly contributes to efficient and accurate facies analysis in the geologically complex Brazilian presalt region.

IMAGE LOG DATA AND MANUAL LABELS

Acoustic imaging tools use a rotating transducer positioned at the center of the well to acquire unrolled high-resolution images of the entire borehole wall. The transducer emits ultrasonic pulses continuously and records the pulses reflected by the borehole wall. The amplitudes of the reflected pulses are then visualized as two types of borehole images. The first type is the static image, wherein the pixel values are normalized over the entire logged interval, effectively portraying the large-scale variations associated with lithologic changes and geologic events. The second type is the dynamic image, wherein the pixel values are normalized within a sliding window, revealing finer texture and fabric details with enhanced color contrast. Overall, both types of image logs are essential in providing a comprehensive representation of the downhole formations.

This study builds upon the previous work by You et al. (2023) on the deep-learning-assisted classification of acoustic image logs, with a continued focus on the same data set obtained from the Brazilian presalt oil fields. These offshore oil reserves predominantly comprise carbonate reservoirs. The presalt layer spans from the coast of Espírito Santo to the coast of Santa Catarina, lying beneath a salt layer that is approximately 2000 m thick. Above the salt layer, there are postsalt sediments exceeding 2000 m in thickness. Since Petrobras' initial exploration in 2006, the presalt oil reserves have attracted substantial investments from major oil companies due to their abundant and high-quality nature. However, drilling through the extensive postsalt sediments and the salt layer under the deep sea is a highly expensive endeavor. Hence, it is essential to enhance our understanding of the geologically complex carbonate reservoirs in this region.

High-resolution borehole image logs play a critical role in improving reservoir characterization as they provide substantial geologic information. One prominent feature observed in image logs is the presence of horizontal sinusoidal curves of a single period,

indicating planar geologic structures such as fractures and beddings intersecting the wellbore. Carbonate reservoirs from the Brazilian presalt region, in particular, exhibit a wider range of features associated with karst processes and biological activities. For example, vugs are visible as varying sized voids in the image logs, whereas stromatolites, layered deposits of limestone, display concentric layering patterns. In addition, shrub imprints formed during the deposition process are widespread in carbonate reservoirs, appearing as large-scale dendritic patterns encompassing multiple layers or small-scale v-shape patterns within layers. Apart from these geologic features, image logs also contain drilling-related artifacts, such as tool scratching and spiraling. Among the various artifacts, borehole breakouts, characterized as pairs of near-vertical irregularities spaced 180° apart in azimuth, are of particular interest because they are valuable for in situ stress regime analysis.

You et al. (2023) define five mutually exclusive facies for image logs from the Brazilian presalt region based on the most representative features observed, including the geometric attributes of the fine strata (mainly parallelism and continuity) and the presence of shrub imprints. The definition of the five facies is listed as follows:

- Facies 0: parallel thin laminations with branching shrub imprints,
- Facies 1: parallel continuous or discontinuous laminations across the wellbore without shrubs,
- Facies 2: faint incomplete laminations or transparent beds,
- Facies 3: chaotic fabrics with no clear laminations,
- Facies 4: strong artifacts covering all underlying geologic features.

According to the well-established classification system for carbonate rocks by Dunham (1962), the dominant rock types shift from boundstone to grainstone to mudstone from facies 0 to 2. Specifically, boundstone refers to rocks where the original sediments are tightly bounded during deposition, grainstone has a grain-supported fabric and lacks mud content, and mudstone is mud supported and lacks grain contents. Moreover, facies 3 is a broad chaotic class comprised of stromatolites, breccia, conglomerates, and reworked sediments. Because they are affected by widespread drilling artifacts, some image logs may exhibit characteristic features of facies 0, 1, or 2, but with lower certainty. They are manually classified as class 5, 6, or 7, respectively, and are assigned a lower weight during training. However, when the artifacts become too prominent, concealing all underlying geologic features, these segments are classified as facies 4. There are instances when the image logs contain intertwined features from multiple facies or have very low resolution, making it difficult to classify them accurately. Such images are identified as the uncertain class that is not classified as any facies (class 8), excluded from the training data set, and used for blind test after the training phase. A detailed comparison of the nine classes is shown in Table 1.

The data set investigated in this study comprises 15 wells originating from different oil fields in the Brazilian presalt region, including the Sapinhoa, Tupi, and Iracema. These wells were imaged by different tools, including Schlumberger's ultrasonic borehole imager (UBI), the ultrasonic imager tool (USIT), Halliburton's circumferential acoustic scanning tool (CAST), and Baker Hughes' circumferential borehole imaging log (CBIL). The image quality varies from "good" to "not bad," as assessed by an experienced geologist. The WellCAD software was used to visualize and manually classify the image logs into nine distinct classes. The manual labeling process

for the image logs of wells 0–7 was an extensive undertaking that spanned over a period of two months. Because manual labeling is too time consuming, the remaining wells are left unlabeled. More details about the wells can be found in Appendix A.

METHODOLOGY

In this paper, we propose an advanced methodology for acoustic image log classification using deep learning. Built upon previous

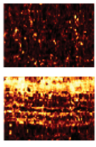
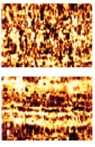
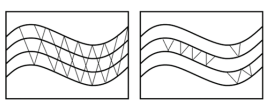
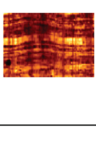
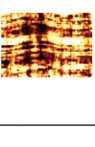
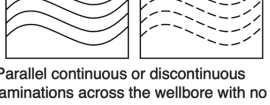
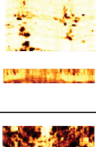
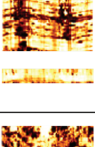



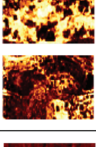
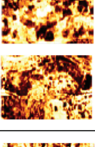
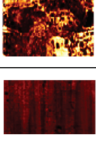
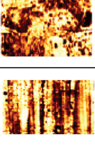

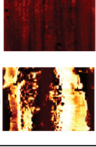
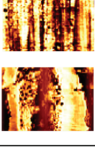
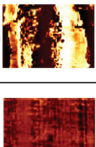
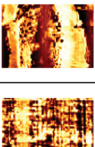
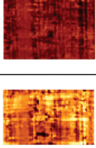
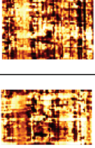
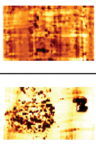
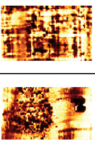
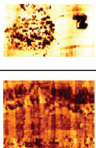
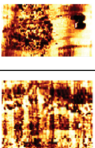
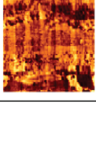
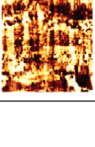
work by You et al. (2023), our approach achieves significant improvements by leveraging a superior pure transformer model called the Swin-Unet (Cao et al., 2022) and a more effective training strategy that combines multiple advanced techniques, including transfer learning and CL (Northcutt et al., 2021). In this section, we provide an introduction to the architecture of our neural network and then delve into the details of the used training strategy.

SwinV2-Unet

The Facies-Unet, introduced by You et al. (2023), used the popular U-Net structure (Ronneberger et al., 2015) for facies classification, which is distinguished by its U-shaped encoder-decoder structure and skip connections. Although Facies-Unet performed well in facies classification, it faces challenges in capturing long-range semantic interactions due to the intrinsic locality of convolution operations, resulting in oversegmentation issues and the misinterpretation of large-scale features, particularly stromatolites. In contrast, transformers (Vaswani et al., 2017), which have revolutionized the field of natural language processing, can perfectly model long-range dependencies through the multihead self-attention mechanism. To address the inadequacy of convolutional neural networks (CNNs) in handling long-range dependencies, researchers have explored the adaptation of transformers for computer vision tasks, wherein the Swin Transformer (Liu et al., 2021) has emerged as the state-of-the-art in numerous computer vision benchmarks. The Swin Transformer improves computation efficiency by restricting self-attention computation to shifted windows (for additional details, refer to Appendix B) while also enabling hierarchical learning with multiple patch merging layers. Subsequently, Liu et al. (2022) propose the Swin Transformer V2 that enhances the scaling-up capacity of the original Swin Transformer. Building upon the widely used U-Net and the cutting-edge Swin Transformer (V1), Cao et al. (2022) propose the Swin-Unet, which inherits the U-Net’s structure, with all basic convolutional blocks replaced by the Swin Transformer blocks. With its robust local-global semantic feature learning capability, the Swin-Unet has demonstrated state-of-the-art performance in multiorgan and cardiac segmentation tasks.

In this study, we customize the Swin-Unet (Cao et al., 2022) for our purpose of depthwise facies classification for acoustic image logs. We name our model the SwinV2-Unet because all the Swin Transformer blocks are upgraded to the Swin Transformer V2 blocks. As shown in Figure 1, the encoding path of our model is configured as the concatenation of the convolutional channel adapter and the tiny-size Swin

Table 1. A detailed description of the nine classes defined for carbonate rocks.

| Class | Examples Static image Dynamic image 0 127.5 255 | Stick figure and description | Rock type |
|-------|---|---|--|
| 0 | 0.5 m   |  Parallel thin laminations with branching shrubs. | Boundstone or grainstone |
| 1 | 0.5 m   |  Parallel continuous or discontinuous laminations across the wellbore with no shrubs. | Interbedded grainstone and mudstone |
| 2 | 0.5 m   0.12 m   |  Faint incomplete laminations or transparent beds. | Mudstone or silicified rocks |
| 3 | 0.5 m   0.5 m   |  Chaotic, discontinuous, and irregular, including stromatolites. | Breccia, conglomerates, reworked sediments, or stromatolites |
| 4 | 0.45 m   0.5 m   | Strong artifacts that conceal all underlying geologic structures | |
| 5 | 0.5 m   | Less certain facies 0 | |
| 6 | 0.45 m   | Less certain facies 1 | |
| 7 | 0.45 m   | Less certain facies 2 | |
| 8 | 0.75 m   | Uncertain section (entangled features or low resolution) | |

Classes 0–4 represent the mutually exclusive facies, classes 5–7 denote the facies with lower certainty, and class 8 corresponds to the uncertain sections.

Transformer V2 model (SwinV2-T) (Liu et al., 2022), whereas the decoding path mirrors the encoding path except for the last layer. Our model takes static and dynamic images as input, represented by two channels. The channel adapter expands the input data to three channels to accommodate the three-channel input required by the SwinV2-T model. Then, the three-channel data are partitioned into nonoverlapping 4×4 pixel patches, which are subsequently transformed into 1D vectors of dimension C using a linear embedding layer. These C -dimensional features are treated as tokens, namely the inputs to the transformer blocks. Each encoding block concludes with a patch merging layer that reduces the height and width of the feature maps by a factor of two while doubling the number of channels. Conversely, each decoding block begins with a patch-expanding layer, which performs the inverse operation of the patch-merging layer. The integration of the patch-merging and expanding layers enables hierarchical analysis of the input data, facilitating efficient local-global feature learning. In addition, the skip connections between the encoding and decoding paths help maintain fine-scale details that may be lost during the encoding process. After the last decoding block, the tokens are mapped back to the image domain using the reverse patch embedding layer. To incorporate azimuth information into the facies probability, we add a 2D convolutional layer as the final layer of the neural network. This layer has a kernel size of $1 \times W$, a stride size of one, and zero padding. The input and output channels are set to the number of classes (five in this study), and they are independently connected, given that the number of convolution groups also equals the number of classes. This layer generates a facies probability for each depth, presenting as a vector with a

length equal to the number of classes. The maximum element of the estimated facies probability represents the predicted probability, whereas the corresponding index indicates the predicted facies.

Training strategy

Transfer learning

As discussed by You et al. (2023), the image logs from the Brazilian presalt region are highly complex and diverse, making

Table 3. Well information of the investigated data set.

| Well index | Field | Vendor | Imaging tool | Image quality | Label |
|------------|----------|--------------|--------------|-------------------|-------|
| 0 | Sapinhoa | Schlumberger | UBI | Good | T |
| 1 | Sapinhoa | Schlumberger | UBI | Not bad | T |
| 2 | Sapinhoa | Schlumberger | UBI | Decent | T |
| 3 | Tupi | Schlumberger | UBI | Good | T |
| 4 | Sapinhoa | Halliburton | CAST | Decent | T |
| 5 | Sapinhoa | Schlumberger | UBI | Decent | T |
| 6 | Iracema | Schlumberger | USIT | Decent to not bad | T |
| 7 | Tupi | Halliburton | CAST | Not bad | T |
| 8 | Tupi | Halliburton | CAST | Decent | F |
| 9 | Tupi | Halliburton | CAST | Not bad | F |
| 10 | Sapinhoa | Halliburton | CAST | Not bad | F |
| 11 | Iracema | Baker Hughes | CBIL | Not bad | F |
| 12 | Sapinhoa | Halliburton | CAST | Not bad | F |
| 13 | Tupi | Halliburton | CAST | Decent | F |
| 14 | Tupi | Baker Hughes | CBIL | Not bad | F |

The image quality degrades from good to “decent” to not bad. In the last column, “T” denotes the manually labeled data, whereas “F” indicates the unlabeled data.

Table 2. The results of the ablation study examining the impact of transfer learning, CL, and the mean teacher method.

| Test | Model | Transfer learning | Freeze | GR | Trainable weights (million) | Label | Mean teacher method | Initial learning rate | Batch size | $p_{\text{less certain}}$ | Training speed (min/epoch) | Macro F1 score |
|------|-------------|-------------------|--------|---------|-----------------------------|-------|---------------------|-----------------------|------------|---------------------------|----------------------------|----------------|
| 1 | SwinV2-Unet | Y | Y | N | 14.16 | Raw | N | 0.0005 | 32 | 0.5 | 4.1 | 0.78 |
| 2 | SwinV2-Unet | Y | N | N | 41.34 | Raw | N | 0.0005 | 32 | 0.5 | 4.8 | 0.74 |
| 3 | SwinV2-Unet | N | N | N | 41.34 | Raw | N | 0.0005 | 32 | 0.5 | 4.8 | 0.70 |
| 4 | Facies-Unet | N | N | Channel | 2.18 | Raw | N | 0.0001 | 32 | 0.5 | 1.3 | 0.68 |
| 5 | SwinV2-Unet | Y | Y | N | 14.16 | Clean | N | 0.0005 | 32 | 1 | 4.1 | 0.90 |
| 6 | Facies-Unet | N | N | Channel | 2.18 | Clean | N | 0.0001 | 32 | 1 | 1.1 | 0.86 |
| 7 | Facies-Unet | N | N | N | 2.18 | Clean | N | 0.0001 | 32 | 1 | 1.1 | 0.84 |
| 8 | SwinV2-Unet | Y | Y | Channel | 14.16 | Raw | N | 0.0005 | 32 | 0.5 | 4.1 | 0.77 |
| 9 | SwinV2-Unet | Y | Y | 1D-CNN | 14.18 | Raw | N | 0.0005 | 32 | 0.5 | 4.2 | 0.78 |
| 10 | SwinV2-Unet | Y | Y | N | 14.16 | Raw | Y | 0.0005 | 64 | 0.5 | 7.7 | 0.78 |

“Y” and “N” represent yes and no, respectively. As for the “gamma ray (GR)” column, N represents that GR logs are not used, “channel” represents that the GR logs are appended to the inputs as a third channel, and “1D-CNN” represents that a model composed of four 1D convolutional layers is used to process GR logs in parallel to the SwinV2-Unet. The clean labels used in test 5 are obtained by applying CL to the initial out-of-sample predictions given by test 1, whereas tests 6 and 7 use clean labels from test 4 predictions. When using the mean teacher method, each batch is composed of 32 labeled samples and 32 unlabeled samples. The macro F1 score is the average F1 score across five classes in out-of-sample predictions during k -fold cross validation.

the manual labeling of only eight wells inadequate for effectively training a DNN with robust generalizability. Furthermore, the SwinV2-Unet, being a large model with approximately 42 million trainable parameters, is prone to overfitting when trained on small annotated data sets comprising only a few hundred thousand images

(Oquab et al., 2014). Therefore, it is imperative to improve the training strategy to address the overfitting challenge.

Transfer learning has emerged as a prominent technique to overcome the scarcity of labeled data and accelerate training. Extensive research has demonstrated the effectiveness of pretraining models on large data sets such as the ImageNet (Deng et al., 2009) and subsequently fine tuning them on smaller task-specific data sets. Despite the variations in data statistics and tasks, these studies have consistently shown substantial performance improvements (Oquab et al., 2014; Yosinski et al., 2014). In view of the proven success of pretraining in diverse computer vision tasks, we adopt this technique for our model as well. Specifically, we initialize the encoding blocks and the bottleneck of our model using the pretrained SwinV2-T model on the ImageNet-1K published by Liu et al. (2022), whereas the other blocks are initialized randomly. The weights loaded from the pretrained model are kept frozen during training. The reason to freeze them will be explained in the ‘‘Discussion’’ section.

Confident learning

In addition to the scarcity of labeled data, the quality of the labels plays a crucial role in ML performance. Our manual labels contain inevitable human errors and inconsistencies, in spite of the extensive time and effort we have invested in them. Northcutt et al. (2021) propose a generalized CL framework to identify and eliminate label errors, ultimately revealing the ‘‘true’’ labels from noisy data. This CL framework has been demonstrated to be effective through theoretical analysis and experimental validation. It is performed in four steps, which are described as follows.

The first step involves training the initial DNN on the noisy original data set using the k -fold cross-validation algorithm (Hastie et al., 2009) to generate out-of-sample predictions for all samples. To use the k -fold cross-validation algorithm, the entire data set is evenly split into k -folds. Then, the same DNN is trained k times, with each fold used as the test set in rotation, whereas the remaining folds serve as the training set. The DNNs predictions for the test set are the out-of-sample predictions. This approach also provides an unbiased evaluation of the model’s generalizability, which is crucial when dealing with small and highly heterogeneous data sets such as the Brazilian presalt data used in this study.

The second step is to estimate the joint distribution between the observed noisy labels (or the manual labels), denoted as \tilde{y} , and the true labels, denoted as y^* , based on out-of-sample predictions. We use $p(y = j; \mathbf{x})$ to represent the out-of-sample predictions, specifically the probability of a sample \mathbf{x} belonging to class j as predicted by the DNN. Moreover, $\hat{\mathbf{X}}_{\tilde{y}=i, y^*=j}$ denotes the

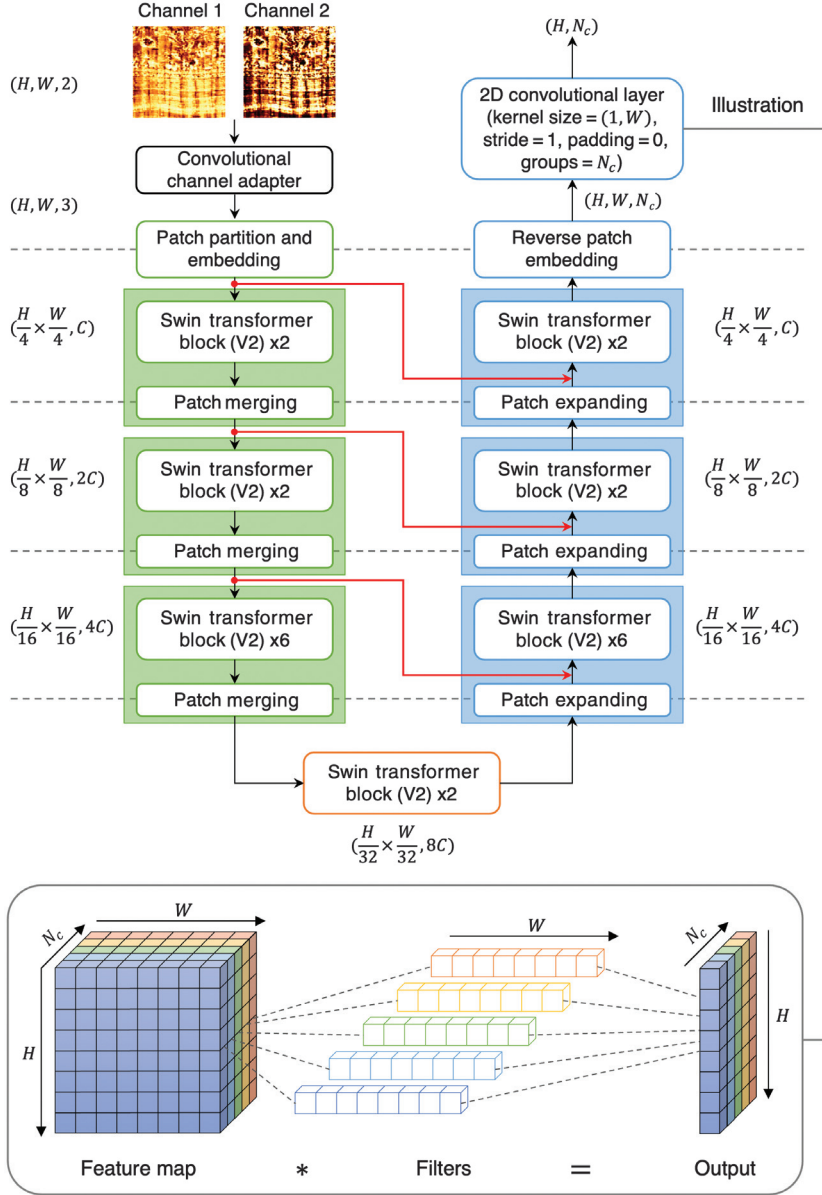


Figure 1. The architecture of our SwinV2-Unet, which is composed of the channel adapter, green encoding blocks, orange bottleneck blocks, and red skip connection operations. The representation shape of each level is put between the corresponding dashed gray lines. The image height, width, and embedding dimension of the patch embedding layer are denoted as H , W , and C , respectively. The number of classes is represented with N_c . The convolutional channel adapter is composed of two 2D convolutional layers: the first layer has a kernel size of one, a stride size of one, zero padding, and 32 output channels; the second layer has a kernel size of one, a stride size of one, zero padding, and three output channels. The configuration of the final output layer is illustrated at the bottom, with features and filters for different classes represented in distinct colors. Each input channel is convolved with its own filter of size $1 \times W$, generating a logit at each row. The softmax function is then applied along the channel dimension (N_c) to obtain the probability for each class at every row.

estimated subset of the samples that are labeled as class i but with a sufficiently high probability ($p(y = j; \mathbf{x})$) to be considered as belonging to class j . Specifically, $\hat{\mathbf{X}}_{\tilde{y}=i, y^*=j}$ is estimated under the constraint of a per-class threshold t_l :

$$\hat{\mathbf{X}}_{\tilde{y}=i, y^*=j} = \left\{ \mathbf{x} \in \mathbf{X}_{\tilde{y}=i} : j = \underset{l \in [m]: p(y=l; \mathbf{x}) \geq t_l}{\operatorname{argmax}} p(y = l; \mathbf{x}) \right\}, \quad (1)$$

where $\mathbf{X}_{\tilde{y}=i}$ denotes the subset of samples labeled as class i and $[m]$ represents the list of all classes. The threshold t_l is designed as the average self-confidence for each class:

$$t_l = \frac{1}{|\mathbf{X}_{\tilde{y}=l}|} \sum_{\mathbf{x} \in \mathbf{X}_{\tilde{y}=l}} p(y = l; \mathbf{x}). \quad (2)$$

Essentially, the threshold t_l is proportional to the confidence of the DNN for each class, thereby enhancing CL robustness to the class imbalance and uneven class probability distribution. Next, a confident joint matrix $\mathbf{C}_{\tilde{y}, y^*}$ can be derived, where the entry $\mathbf{C}_{\tilde{y}=i, y^*=j}$ represents the number of samples labeled as class i but estimated as class j ($|\hat{\mathbf{X}}_{\tilde{y}=i, y^*=j}|$). The joint distribution $\mathbf{Q}_{\tilde{y}, y^*}$ is derived by normalizing $\mathbf{C}_{\tilde{y}, y^*}$ to match the observed marginal distribution of \tilde{y} and ensure $\mathbf{Q}_{\tilde{y}, y^*}$ sums to one. The diagonal entries of $\mathbf{Q}_{\tilde{y}, y^*}$ present the correct rates of the manual labels, whereas the off-diagonal entries denote their asymmetric noise rates.

The third step is to rank and prune the label errors or data cleaning. Northcutt et al. (2021) introduce five rank-and-prune approaches. We selected the most robust one in our experiments, which is to prune by noise rate. For each off-diagonal entry in $\mathbf{Q}_{\tilde{y}=i, y^*=j} (i \neq j)$, we prune $n \cdot \mathbf{Q}_{\tilde{y}=i, y^*=j}$ samples that are labeled as class i with max margin $p(y = j; \mathbf{x}) - p(y = i; \mathbf{x})$, with n denoting the total number of samples in the data set. The preserved samples are considered to be correctly labeled, forming a clean data set.

The last step is to train the DNN using the cleaned data set. The k -fold cross-validation algorithm is used in this step to assess the generalizability of the ML model more accurately.

Training workflow

Based on the mentioned training strategies, we propose a training workflow for the complex Brazilian presalt data with limited noisy manual labels. The first step is to prepare the training data. As described previously, we have labeled eight wells by hand. The labeled wells contain uncertain sections marked as class 8, which are excluded from the labeled data set. Because image logs are usually a few hundred meters long, we need to split them into small patches to let the neural network analyze them locally. On the one hand, we expect that the image patches partially overlap each other to avoid boundary effects as well as augment the training data. On the other hand, we need to avoid data leakage from the test set to the training set in the presence of overlapping patches. To balance the two aspects, we propose a two-step image log splitting workflow for the labeled data set from wells 0 to 7. First, the continuous image logs are split into 3.9 m nonoverlapping segments, which are split into fivefold evenly and randomly for the k -fold cross-validation algorithms. Specifically, folds 0–3 consist of 96 segments each, measuring 374.4 m in length, whereas fold 4 is slightly smaller, containing 94 segments with a length of 366.6 m. It is worth noting

that the facies distribution varies slightly among the folds, with facies 4 being the least represented. Further information about the fivefold can be found in Table 4. In the second step, each segment is split into overlapping patches with a 1.3 m sliding window moving in the depth direction, whose step size is set to be one fourth of its height (0.325 m). The discretization of the extracted patches is reconciled to 0.00508 m per pixel in depth and 1.4° per pixel in azimuth via linear interpolation, generating 256×256 image log patches. You et al. (2023) synthesize facies 0, 1, 2, and 4 as the superposition of the convolution between various 2D features (sinusoids, voids, etc.) and their feature maps to augment the training data, which improved the test accuracy by approximately 4%. Hence, we generate 13,824 synthetic image patches (evenly composed of facies 0, 1, 2, and 4) to augment the training set in this work as well.

After data preparation, we proceed with CL to address the presence of noisy labels. First, the k -fold cross-validation algorithm is performed to obtain the out-of-sample predictions. To conduct transfer learning, the encoding path of the SwinV2-Unet, excluding the channel adapter, is initialized with the ImageNet-1K pretrained SwinV2-T weights and remains fixed throughout the training, whereas the rest of the SwinV2-Unet is initialized randomly and optimized during the training process. The same SwinV2-Unet is trained five times from this initial status. Specifically, in the k th training, the k th fold is taken as the test set; the remaining folds combined with the synthetic data are taken as the training set. Following the approach of You et al. (2023), we employ weighted cross-entropy as the training loss, which has been widely used to measure the difference between the true and predicted probability distributions. Classes 5–7, which exhibit lower certainty, are assigned a weight less than one ($p_{\text{less certain}}$) to reflect the reduced confidence in their manual labels. Next, the less certain classes are converted to the corresponding certain classes (classes 0–2) for training purposes. The loss function for a batch of data can be expressed as

$$L = - \frac{\sum_{i=0}^{N-1} \sum_{d=0}^{255} w_{id} y_{id} \cdot \log p_{id}}{\sum_{i=0}^{N-1} \sum_{d=0}^{255} w_{id}}, \quad (3)$$

where the subscript id represents the parameters at the d th row of the i th sample; w denotes the label certainty, which is $p_{\text{less certain}}$ for the less certain classes and one for the certain classes; y and p represent the label and prediction, respectively; and N is the batch size. The SwinV2-Unet is optimized to minimize the loss function equation 3 with a widely used gradient descent algorithm, the AdamW

Table 4. The statistics of the field data from the labeled wells.

| Fold | Number of segments | Length (m) | Facies ratio |
|-----------------------------|--------------------|------------|--------------|
| 0 | 96 | 374.4 | 4:3:3:3:1 |
| 1 | 96 | 374.4 | 3:5:4:5:1 |
| 2 | 96 | 374.4 | 2:3:2:2:1 |
| 3 | 96 | 374.4 | 4:3:2:2:1 |
| 4 | 94 | 366.6 | 2:4:2:4:1 |
| Uncertain section (class 8) | — | 466.1 | — |

The facies ratio is the ratio of facies 0–4 with classes 5–7 combined to facies 0–2.

optimizer (Loshchilov and Hutter, 2017). To accelerate training and prevent overfitting, the loss of the test set is taken as the metric to adjust the learning rate and to determine the early stopping point. Specifically, the learning rate is halved when the test loss stops decreasing in three consecutive epochs; the training is halted when the test loss fails to decrease for six consecutive epochs after the model has been trained for 42 epochs. As the training stops, the model achieving the lowest test loss is retained to give out-of-sample predictions on the test set or the k th fold. Next, based on the out-of-sample predictions on all labeled data and the manual labels, we detect and prune label errors using the prune-by-noise-rate approach.

After performing data cleaning, we apply the k -fold cross-validation algorithm to the cleaned data. We use the same transfer-learning methods mentioned previously in the process. This approach ensures that the models are trained and evaluated on multiple folds of the data, improving the robustness and generalizability of the results.

RESULTS

In this section, we first present the results obtained from CL, followed by the final training results using the cleaned data set.

In CL, the batch size and initial learning rate are determined through trial and error. A batch size of 32 and an initial learning rate of 0.0005 are chosen, as this combination results in a stable training process. The less certain weight $p_{\text{lesscertain}}$ is set to 0.5, which has been determined to be the optimal value in the previous work (You et al., 2023). The SwinV2-Unet model is trained on one NVIDIA-A100-40 GB GPU, with an average training speed of approximately 4 min per epoch. The total training time for one iteration of the k -fold cross validation is approximately 3 h with early stopping. The learning curves of one iteration are shown in Appendix C, which offers a detailed illustration of the model’s evolution. The model’s performance is assessed using classification accuracy, which represents the percentage of correct predictions out of all the predictions made. In k -fold cross-validation, the accuracy achieved in each iteration is aggregated to calculate the mean and standard deviation, providing an overall performance measure. The average

training accuracy is 88%, with a standard deviation of 3%. As for the test accuracy, the mean is 78%, with a standard deviation of 1%.

To better assess the generalizability of the model for this multi-class classification task with imbalanced data, we aggregate all the out-of-sample predictions across the five iterations and plot their confusion matrices in Figure 2: the first matrix is normalized by rows, and the second matrix is normalized by columns. The diagonal values of Figure 2a and 2b are the recall and precision of each class, respectively. Particularly, the recall represents the proportion of each actual facies to be accurately identified, whereas the precision represents the proportion of each predicted facies to be correct. The $F1$ score, which is the geometric mean of the recall and precision, is another widely used metric that provides a balanced measure of the model’s accuracy by considering its ability to correctly classify the positive instances (precision) and its ability to capture all the positive instances (recall). The $F1$ scores for facies 0–4 are 0.79, 0.75, 0.75, 0.82, and 0.80, respectively. These scores suggest that the model performs slightly better for facies 0, 3, and 4 than for facies 1 and 2. The average $F1$ score over all classes is called the macro $F1$ score, which quantifies the overall classification performance of the model with one value. For the initial SwinV2-Unet trained with noisy labels, the macro $F1$ score is 0.78.

Based on the out-of-sample predictions, the estimated joint distribution between the noisy labels and uncorrupted/true labels is shown in Figure 3a. The prune-by-noise-rate approach prunes 11% of the manual labels from the training data set. The pruning rates for each class are shown in Figure 3b. In general, the less certain classes 5–7 have significantly higher pruning rates than the more certain classes 0–3, which aligns with our expectations. Class 4 exhibits a pruning rate of 20%, primarily due to the inherent challenge of detecting the artifacts with confidence, as they often coexist with other facies. Furthermore, in Appendix D, we show six randomly selected mislabeled patches, demonstrating the proficiency of CL in detecting inaccurate labels. Specifically, the identified false labels mainly pertain to low-quality images that lack distinct features complying with the facies definition.

After data cleaning, the SwinV2-Unet is trained with the cleaned data set using the k -fold cross-validation algorithm. In CL, the remaining labels are considered to be true. Hence, the remaining less certain classes are weighted equally with the other more certain classes by setting the parameter $p_{\text{lesscertain}}$ to one. The batch size and initial learning rate used in the CL phase are maintained for this training. The training speed remains 4 min per epoch using one NVIDIA-A100-40 GB GPU. Each iteration of the k -fold cross-validation process takes approximately 3–4 h to complete. The learning curves for one iteration are available in Appendix C. Across the five iterations, the mean training accuracy is measured to be 96% with a standard deviation of 1%. On average, the test accuracy reaches an impressive value of 89% with a standard deviation of 1%. Notably, the data cleaning process has led to substantial improvements in training and test accuracies, with gains of 8% and 11%, respectively. Figure 4 shows the normalized confusion matrices obtained from the out-of-sample predictions. The achieved recalls

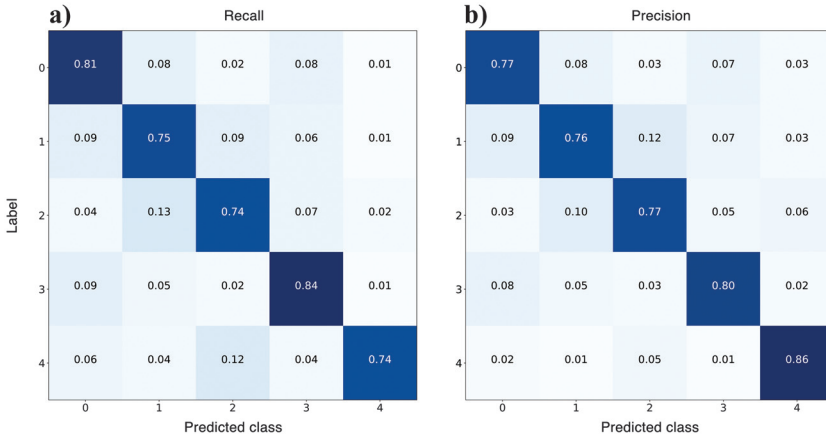


Figure 2. The normalized confusion matrices of all out-of-sample predictions obtained in CL. (a) Row normalization: Each row of the matrix is independently normalized, and the diagonal values indicate the recall for each class. (b) Column normalization: Each column of the matrix is independently normalized, and the diagonal values represent the precision of each class.

and precisions are consistently high, exceeding 0.87 for all facies. The corresponding $F1$ scores for facies 0–4 are 0.90, 0.87, 0.88, 0.91, and 0.91, respectively. The model’s macro $F1$ score averages 0.90, surpassing the initial model’s macro $F1$ score (trained with noisy labels) by a significant margin of 0.12. Overall, the final SwinV2-Unet model demonstrates exceptional generalizability, delivering highly accurate classification results for the test data with high recall and precision across all facies.

The classification capability of SwinV2-Unet is further analyzed by visualizing its prediction results for multiple test sections. Because SwinV2-Unet is specifically designed for depthwise facies classification, its predictions may include ultrathin segments. To avoid oversegmentation, segments with a thickness of fewer than eight centimeters and an average probability of below 0.75 are considered unreliable. These segments are replaced with the nearest facies predictions, yielding the final predictions. As shown in Figure 5, the final predictions coincide well with the manual labels, highlighting the high accuracy and strong generalizability of our model. As described previously, the implementation of the shifted-window-based self-attention mechanism in the SwinV2-Unet enables the efficient learning of local-global representations. This capability is particularly evident in the correct classification of the large-scale stromatolite below line 2. Our model accurately classifies it as the chaotic facies, taking into account the localized shrubby features and the larger concentric layering structure, in contrast to the previous Facies-Unet model (You et al., 2023) that solely focuses on local features and misinterprets it as facies 0. Moreover, the SwinV2-Unet accurately identifies some fine-scale facies that have been overlooked in the manual labeling process. For instance, the section between lines 4 and 5 appears transparent without any visible layering. Our model correctly predicts it as facies 2, whereas the human interpreter annotated it as facies 0 or 1. Fortunately, this fake label (marked with purple) is detected and pruned by the CL approach. In addition to addressing false labels, the CL process detects segments that are challenging to classify due to low image quality. Above line 1, the section displays faint, thin beddings and shrubs typical of facies 0, but the strong noise makes it difficult to classify with certainty. As a result, the manual labels of facies 4 and our model’s predictions of facies 0 seem plausible. Similarly, the section above line 3 suffers from low resolution. It exhibits vuggy and rough characteristics with traces of laminations, resembling the predicted facies 0 in the same subplot. Therefore, our model’s prediction is highly likely to be correct.

The uncertain sections, which may exhibit high levels of noise, entangled features, or low resolution, are used for the blind testing of our model. In Figure 6, we present the classification results of our model for three uncertain segments obtained from different wells. The interpretation of the first segment is hindered by its low resolution. However, our model’s predictions show a strong correlation with the image textures. The smooth sections with clear continuous laminations across the wellbore are predicted as facies 1, whereas the smooth sections lacking clear laminations are classified as facies 2. Moreover, the rough, vuggy, and weakly laminated section between lines 1 and 2 is appropriately classified

as facies 0. The second segment suffers from low-resolution and vertical artifacts. It predominantly exhibits rough and vuggy textures, which our model correctly assigns to facies 0. In addition,

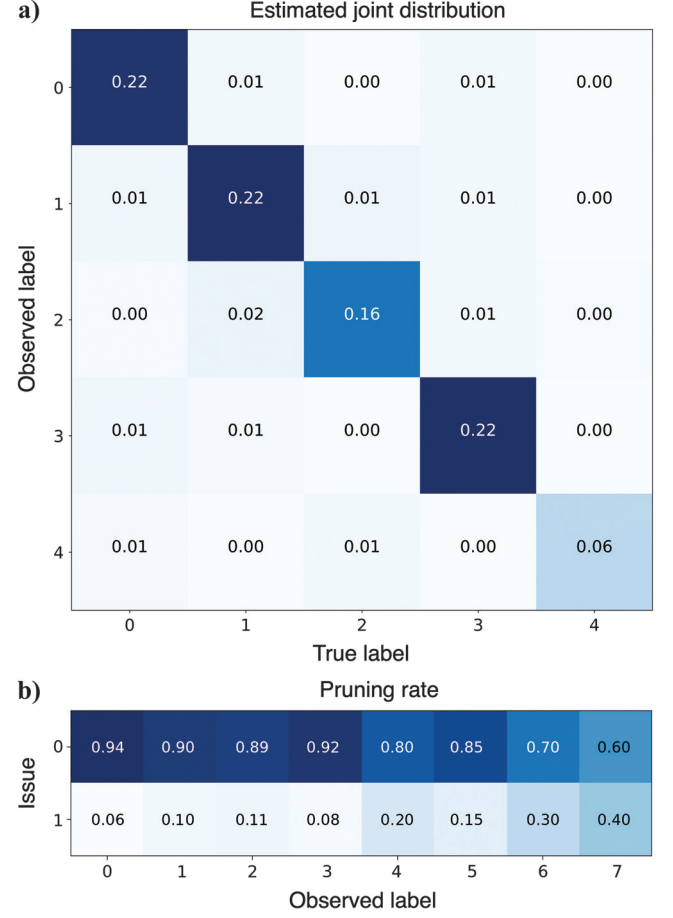


Figure 3. (a) The joint distribution between the observed noisy labels and the true uncorrupted labels and (b) the percentage of the pruned samples of each class. Here, “0” represents the preserved samples, whereas “1” represents the pruned samples.

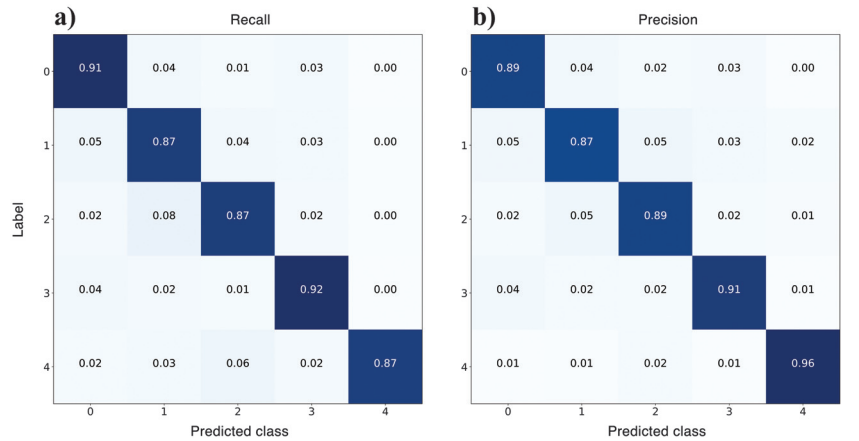


Figure 4. The normalized confusion matrices of all the out-of-sample predictions from models trained with the cleaned data set: (a) row normalization and (b) column normalization.

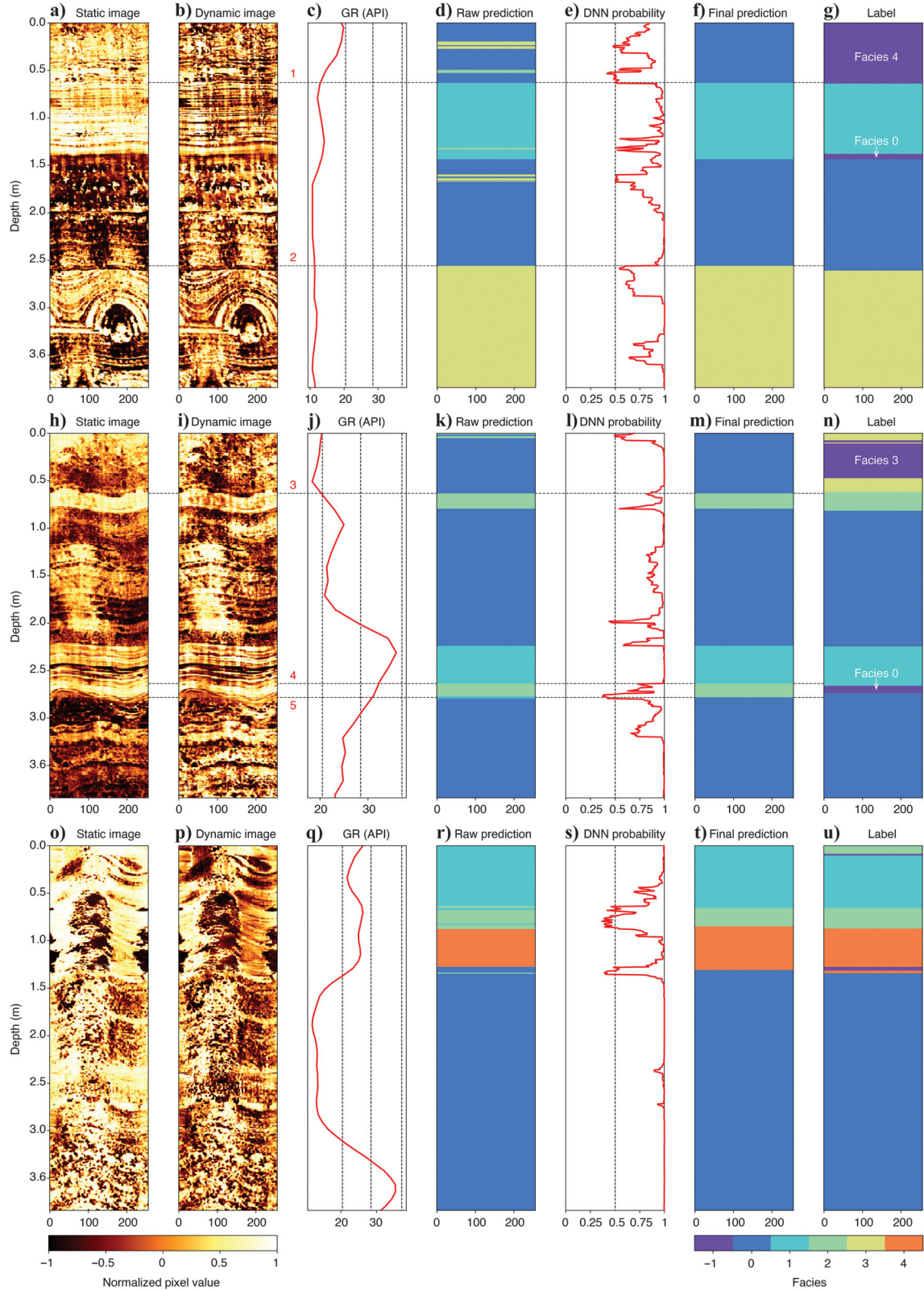


Figure 5. The predictions given by the well-trained SwinV2-Unet for three test sections. (a–u) The static image, dynamic image, gamma-ray log, raw DNN prediction, DNN probability, final prediction, and manual label. The vertical axis represents relative depth. The mean gamma-ray log values for the manually labeled facies 0–2 are 20, 28, and 37, respectively, which are marked with the dashed black lines in the gamma-ray log. The manual labels are updated with classes 5–7 added to classes 0–2. The label errors identified by the CL are annotated as –1 and displayed in purple, with their original labels written next to them. The final prediction is obtained by replacing sections with an average DNN probability below 0.75 and a thickness below 8 cm with the nearby facies.

the chaotic portion with bright, irregular silicified strips between lines 3 and 4 is reasonably predicted as facies 3. The third segment is affected by the artifacts. The section below line 7 exhibits high-angle fractures with a dip direction opposite to the local strata, along with evident borehole breakout, both identified as artifact facies. Furthermore, the section between lines 5 and 6 contains strong imprints of the logging tool, overlapping the high-angle fracture parallel to the one below. Hence, this section is also classified as facies 4 by our model. Overall, our model provides rational predictions for the low-quality uncertain sections, serving as a valuable reference for human interpreters to make better judgments about the facies.

DISCUSSION

In this study, we propose training a pure transformer-based model, the SwinV2-Unet, to provide depthwise facies classification for the acoustic image logs from the Brazilian presalt region. We introduce a specially designed training strategy that combines transfer learning and CL to overcome the limitations of previous studies (e.g., You et al., 2023), mainly limited generalizability due to insufficient labeled data and the presence of erroneous labels. Assessed with the k -fold cross-validation algorithm, our final SwinV2-Unet achieves a high classification accuracy of 89% and an impressive macro $F1$ score of 0.90 for out-of-sample predictions, demonstrating the superior generalizability of our model. Apart from the carefully labeled sections, our model also offers reliable predictions for the low-quality uncertain well sections from the labeled wells. Therefore, it can serve as a robust reference for human interpreters in deciphering the facies of the complex carbonate reservoirs. Notably, our model outperforms the tedious and laborious conventional manual classification approach, offering higher levels of efficiency, consistency, and spatial resolution. In addition, it provides depthwise facies predictions along with corresponding probabilities, which are valuable for uncertainty analysis. With access to more training data, our model holds great potential to evolve into a real-time facies analysis tool for complex image log data in the geologically intricate Brazilian presalt region.

As mentioned previously, You et al. (2023) achieve a classification accuracy of 77% for a 57.6 m test set based on the same Brazilian presalt data. Building upon the advanced SwinV2-Unet architecture and an improved training strategy, we have significantly improved the test accuracy for much larger test folds of 366.6 or 374.4 m in length, achieving an impressive average accuracy of 89%. It is important to explore the individual effect of each operation, including the use of the SwinV2-Unet, transfer learning, and CL. The optimal experiment setting for

the raw data set, referred to as test 1 in Table 2, involves initializing the encoding path of our model with pretrained SwinV2-T weights from the ImageNet-1K data set and keeping them frozen during training. Taking this optimal setting as the base model, we conduct an ablation study by systematically disabling specific operations while using the same k -fold cross-validation algorithm to ensure a fair and unbiased comparison. The macro $F1$ score of the aggregated out-of-sample predictions for the entire data set is used to

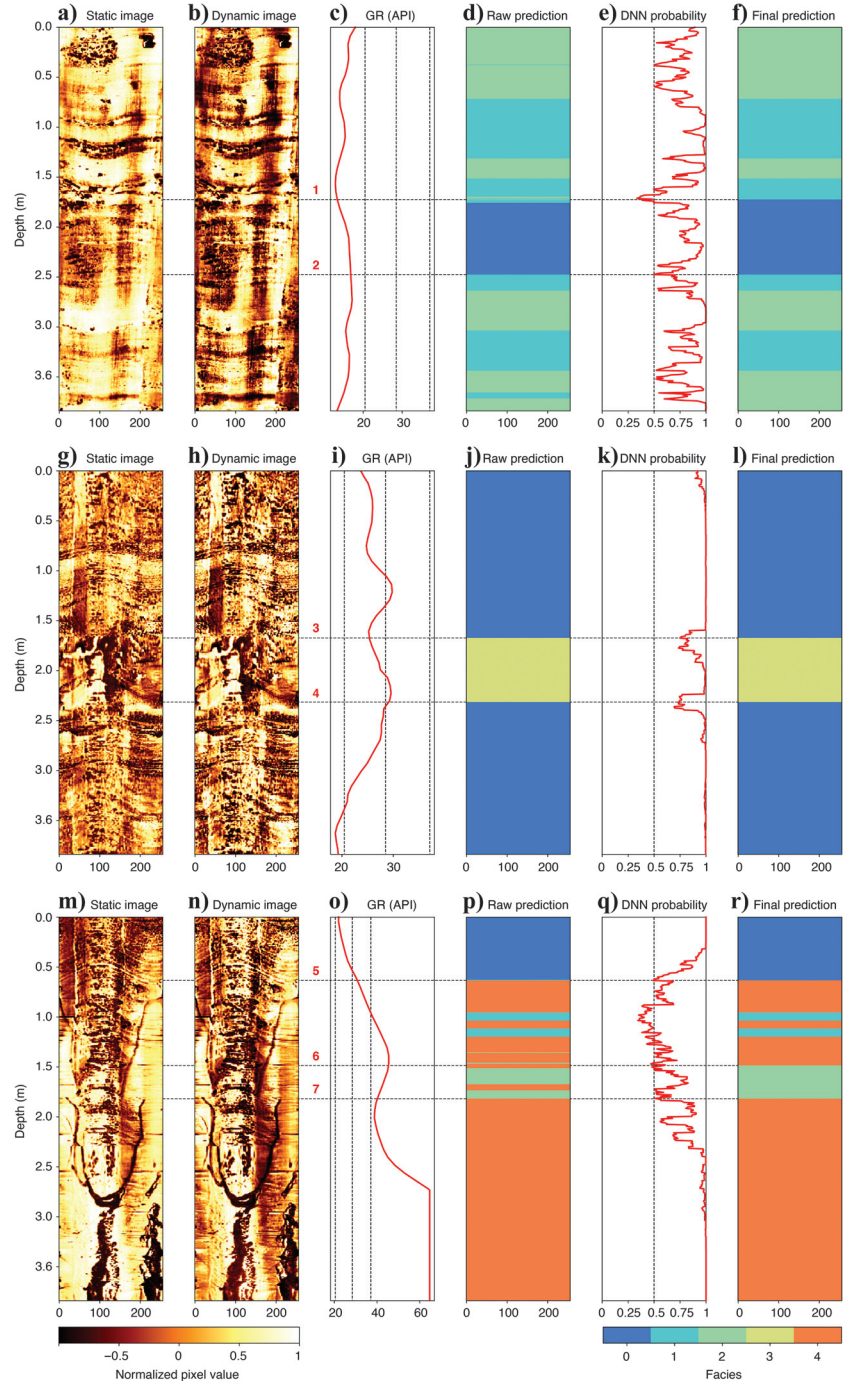


Figure 6. (a–r) The predictions given by the well-trained SwinV2-Unet for three uncertain segments.

measure the classification performance. If we allow the loaded pre-trained weights to be updated during training, the number of trainable parameters increases significantly from 14.16 to 41.34 million, whereas the macro $F1$ score decreases slightly from 0.78 to 0.74. This indicates that the pretrained weights can extract representative features of the input images effectively, making it more efficient to focus on training the channel adapter, the decoding path, and the skip connections. By comparing test 3 with test 1, we observe a notable increase in the macro $F1$ score of 0.08 through transfer learning on the pretrained SwinV2-T weights from the ImageNet-1K data set. Test 4 assesses the performance of the Facies-Unet (You et al., 2023) on raw labels. Because the Facies-Unet is a small model, its training speed is as high as 1.3 min/epoch. By substituting the Facies-Unet with the more advanced SwinV2-Unet, the macro $F1$ score improves by 0.02, as demonstrated by tests 3 and 4. Following data cleaning based on the out-of-sample predictions obtained in test 1, the subsequent round of training, denoted as test 5, uses the clean data set. This CL strategy leads to a prominent increase in the macro $F1$ score by 0.12 for the SwinV2-Unet. Likewise, the same CL algorithm is applied to the Facies-Unet. Based on model predictions from test 4, 23% of the manual labels are identified as false labels, which is more than twice the number pruned by the SwinV2-Unet (11%). This implies that the SwinV2-Unet is more confident about its predictions than the Facies-Unet. Postdata cleaning, the macro $F1$ score of the Facies-Unet rises to 0.86 with the inclusion of the gamma-ray log data as an input channel and to 0.84 without it. This performance falls behind that of the SwinV2-Unet model trained with clean data by 0.04 and 0.06, respectively. Considering the more substantial removal of the data for the Facies-Unet compared with the SwinV2-Unet, the Facies-Unet is anticipated to yield a lower score when trained with data cleaned by the SwinV2-Unet. Thus, the superiority of the SwinV2-Unet over the Facies-Unet is further amplified when coupled with CL. Overall, the combination of an advanced neural network architecture and a superior training strategy that incorporates transfer learning and CL has led to a substantial improvement in our neural network's performance for facies classification.

The success of our training strategy holds profound implications. First, the notable improvement achieved through transfer learning and freezing the pretrained weights indicates the existence of underlying similarities between the borehole images and natural images, despite their apparent dissimilarity to human eyes. This finding encourages us to leverage the wealth of publicly available pretrained weights of cutting-edge deep-learning models trained on large-scale benchmark data sets such as ImageNet for our geophysical applications, particularly rock image analysis. This approach not only saves training time but also addresses a critical challenge faced by ML studies in our field, namely the scarcity of labeled data. Furthermore, the success of the CL method in our study suggests its potential for human error estimation, offering an objective approach to evaluate the labels made by human interpreters. This allows the interpreters to meticulously examine and rectify any identified erroneous labels themselves. Through this iterative process, the accuracy and reliability of facies interpretations are significantly enhanced, contributing to more precise and dependable results.

Facies-Unet (You et al., 2023) integrates gamma-ray logs as a third input channel to supplement the static and dynamic image patches, which is proven to increase the test accuracy by approximately 2%. This enhancement in performance is consistently

observed for the Facies-Unet after data cleaning, as evidenced by tests 6 and 7, presented in Table 2. In contrast, our SwinV2-Unet exclusively uses image logs as inputs. As shown in Table 2, the macro $F1$ score experiences a marginal decrease from 0.78 (test 1) to 0.77 (test 8) for the noisy raw labels when gamma-ray logs are introduced as an additional input channel to our optimal experiment settings. We also explore an alternative approach to incorporate gamma rays, which is to construct a simple four-layer 1D-CNN to process the gamma-ray data and then fuse the outputs from the 1D-CNN and the SwinV2-Unet using a 1D convolutional layer to generate the final predictions. However, as demonstrated by test 9 in Table 2, the macro $F1$ score remains 0.78 with the use of a parallel 1D-CNN for gamma-ray analysis. Therefore, there is no observed improvement by introducing the gamma-ray log as an additional input to the SwinV2-Unet. Several factors may contribute to this outcome. First, the SwinV2-Unet may already possess sufficient power to infer facies solely from the major features of the facies, specifically the visual patterns in the acoustic image logs, rendering the use of gamma-ray logs redundant. Second, gamma-ray logs inherently possess lower spatial resolution than borehole image logs, limiting their impact on image log segmentation. Furthermore, the presence of measurement failures in the gamma-ray data, coupled with potential depth misalignment between the gamma-ray logs and image logs, underscores the importance of rectifying inaccuracies and aligning their depths before integrating them into the model. Finally, 1D CNN is limited in its capacity to handle long-range dependence; hence, exploring the potential of long short-term memory (Hochreiter and Schmidhuber, 1997) or transformer neural networks for gamma-ray log analysis could be a promising avenue for future research.

Apart from transfer learning, semisupervised learning is also widely used to mitigate overfitting and enhance model performance by using the labeled data to ground the predictions and using the plentiful, low-cost, unlabeled data to learn the shape of the larger data distribution (Zhu, 2005). As shown in Tables 3 and 4, we have seven unlabeled wells along with 466.1 m uncertain sections from the labeled wells that have not been used for the training of our neural network. Hence, it is worthwhile to explore the use of the unlabeled data set for our task. Among various semisupervised learning approaches, the mean teacher method (Tarvainen and Valpola, 2017) is a well-established method that involves the use of two neural networks, one acting as a teacher and the other acting as a student. Inspired by the finding that model weights averaged over the training steps tend to be more accurate than the weights at the last step (Polyak and Juditsky, 1992), Tarvainen and Valpola (2017) construct the teacher model as the exponential moving average (EMA) of the student models over previous training steps. In each iteration, the same batch of samples is augmented randomly and then fed to the two neural networks. In addition to the classification loss calculated for the labeled samples, a consistency loss (mostly the mean-squared error) measuring the distance between the predictions of the two models is computed for all input samples. Then, the loss function can be expressed as the weighted sum of the classification and consistency losses. The weights of the student model are updated with gradient descent algorithms to minimize the training loss, whereas the teacher weights are updated as the EMA of the student weights. In essence, the inclusion of a consistency loss enforces that similar inputs are categorized into the same class, thereby allowing the label of one example to help identify the

labels of other similar examples. As shown in Table 2, we test the effect of the mean teacher method in test 10. The inclusion of the unlabeled data set slows down the training speed to 7.7 min/epoch, whereas the macro $F1$ score stays the same as test 1, where the mean teacher method is not enabled. Therefore, the mean teacher method does not bring obvious improvement to the test accuracy on the labeled wells. This could be attributed to the inferior image quality of the remaining unlabeled wells compared with the labeled data. Moreover, the image logs from different wells always present distinct features on top of the basic characteristics of the five facies. The combined factors of lower image quality in the unlabeled wells and variations in image characteristics pose significant challenges in improving the classification accuracy on the labeled wells with the help of the unlabeled well logs. Therefore, the effectiveness of the mean teacher method heavily depends on the quality and representativeness of the unlabeled data in relation to the labeled data.

Table 2 demonstrates that test 5 achieves the highest accuracy for out-of-sample predictions on the labeled wells. However, it is also crucial to assess the performance of different models on the unlabeled wells to evaluate their generalizability to new wells with slightly different geologic and operational conditions. In general, models trained on larger and more diverse data tend to exhibit higher generalizability. Thus, we hypothesize that models trained with the mean teacher method possess better generalizability to new wells compared with models trained without it because the mean teacher method allows the model to use the unlabeled data set during training. To evaluate this hypothesis, we compare the prediction results of tests 1 and 10 for the unlabeled wells, using the results obtained from the optimal experiment, test 5, as the reference. Due to the length limit, we present the comparison of the three models' performance on two sample sections from the unlabeled wells in Appendix E. After carefully reviewing the predictions of the three models for all the unlabeled wells, we observe that tests 1 and 10 provide comparable predictions for the unlabeled wells, whereas test 5 consistently performs the best overall. Therefore, our initial conjecture that the mean teacher method enhances the generalizability of our model is incorrect. This discrepancy primarily stems from the inferior image quality of the seven unlabeled wells, which contain numerous artifacts and noise and have low resolution. Consequently, the teacher model cannot provide reliable targets or pseudolabels for the unlabeled wells to guide the student model, making it ineffective to improve the student model's performance on the unlabeled wells by enforcing the alignment of the teacher and student model predictions with the consistency loss. Furthermore, because the models' performance on the unlabeled wells correlates well with their performance on the labeled wells, we can conclude that the model's performance on test folds from the labeled wells accurately reflects its generalizability to new wells.

Although our model has shown significant improvement in accuracy and generalizability, it is important to acknowledge the challenges posed by the quality of the acoustic image logs. The acoustic image logs contain widespread artifacts that would interfere with the existing geologic features, hindering an accurate interpretation of the acoustic image logs. In addition, the resolution of the image logs may not always be sufficient to distinguish between different facies, further complicating the task of determining the underlying facies. Given the limited labeled data set and the high variability in image quality, it is almost impossible to train a single end-to-end neural network to distinguish the facies precisely and accurately

under the disturbance of the artifacts and inadequate resolution. Although our model is capable of predicting the probability of each facies at each depth, its assistance is limited to uncertainty analysis. Therefore, we think it is essential to remove the artifacts beforehand using either conventional image processing approaches or ML methods. As for the low-resolution image logs, it can be beneficial to train a superresolution neural network for the image logs, which has been extensively studied in the computer vision community and aimed to improve image resolution (Anwar et al., 2020). In our future work, we plan to integrate these preprocessing steps, including artifact removal and image resolution enhancement, into the interpretation workflow and finally eliminate the impedance brought by poor image quality.

Although our model has demonstrated strong generalization to labeled wells and satisfactory performance on unlabeled wells, we have to admit that its predictions still require further assessment by experienced geologists, especially for low-quality data from new wells. Nonetheless, our model significantly enhances and accelerates the facies analysis process by providing real time, reasonable initial facies predictions along with quantified probabilities, thereby alleviating the heavy workload on geologists. Following the first pass of automatic facies classification using our model, geologists can rectify a subsection of the new well with reference to the initial predictions. The corrected subsection, combined with sections showing sufficiently high probabilities, can be used to update our model through transfer learning. The updated model is poised to provide better predictions for the new well as it has been exposed to portions of the new data during training. With the ongoing accumulation of labeled data during the model application phase, we can improve our model's performance through iterative updates using transfer learning on the consistently expanding labeled data set. This iterative process enhances the accuracy and applicability of our model to Brazilian presalt data, gradually reducing human workloads, enhancing productivity, and improving work efficiency over time.

CONCLUSION

In this work, we propose to enhance the automatic facies classification performance for the acoustic image logs from the Brazilian presalt region in two key aspects. First, we adopt the advanced U-shaped transformer, SwinV2-Unet, which integrates the strengths of the Swin Transformer and U-Net architectures. Second, we enhance the training strategy by incorporating transfer learning and CL techniques to mitigate overfitting and address label errors. Our training workflow involves two rounds of training using the k -fold cross-validation algorithm. In the initial round, ML models are trained with the original noisy labels, generating out-of-sample predictions used for CL. In the second round, we train the final model with the cleaned data set and assess its performance in an unbiased manner. The labeled data set is randomly split into fivefold, with each fold spanning 374.4 or 366.6 m, serving as the test set iteratively. Remarkably, our approach achieves an impressive average accuracy of 89% and an unprecedentedly high macro $F1$ score of 0.90 for the test folds. This surpasses the 68% classification accuracy and 0.68 macro $F1$ score obtained by the previous Facies-Unet when evaluated under the same conditions (i.e., undergoing the same data preparation and k -fold cross-validation processes but without transfer or CL). Through an ablation study, we investigate the contribution of different operations, with the macro $F1$ score serving as the evaluation metric. The results show that the substitution of U-Net

with SwinV2-Unet, the inclusion of transfer learning, and the use of CL techniques improve the macro $F1$ score by 0.02, 0.08, and 0.12, respectively. Therefore, our modifications made in this study are demonstrated to be highly effective. Furthermore, compared with the manual labeling approach, our final facies classification model exhibits superior performance in terms of efficiency, consistency, and resolution. It also provides reasonable predictions for low-quality uncertain sections, offering valuable guidance for geologists to make better judgments about the facies. In summary, we have developed a robust end-to-end facies classification model that exhibits high accuracy, efficiency, and generalizability when applied to image logs from the Brazilian presalt region, contributing significantly to the field of automatic facies classification.

ACKNOWLEDGMENTS

The authors thank ExxonMobil Technology and Engineering Company (EMTEC) for providing financial support and data. We extend special thanks to S.-J. Ye, A. Nolting, H. Denli, A. Usadi, A. Scribner, D. Kalita, N. Vento, A. Baumstein, and F. Song from EMTEC for their valuable insights and engaging discussions. We thank Advanced Logic Technology (ALT) for the noncommercial license for WellCAD. The original codes for Swin-Unet are available at <https://github.com/HuCaoFighting/Swin-Unet>.

DATA AND MATERIALS AVAILABILITY

Data associated with this research are confidential and cannot be released.

APPENDIX A

WELL INFORMATION AND FIELD DATA STATISTICS

Table 3 shows the general information of the well-log data provided by EMTEC. To conduct the k -fold cross-validation experiment, the field data are divided into fivefold. Detailed statistics of these fivefold, along with the uncertain section from the labeled wells, are presented in Table 4.

APPENDIX B

SWIN TRANSFORMER BLOCKS

The basic mechanism constituting a Swin Transformer block is the well-known self-attention mechanism proposed by Vaswani et al. (2017). The first block of Figure B-1 shows the self-attention mechanism within nonoverlapping windows. Given a sequence of 1D tokens as input, linear layers are used to generate the query, key, and value vectors for each token. The correlation between the different

tokens is then computed to produce updated tokens. The query, key, and value vectors are represented as matrices $Q, K, V \in \mathbb{R}^{M^2 \times d}$, where M^2 is the number of patches in a window and d is the query/key dimension. The mathematical operation is expressed as

$$\text{Attention}(Q, K, V) = \text{SoftMax}(QK^T / \sqrt{d} + B)V, \quad (\text{B-1})$$

where $B \in \mathbb{R}^{M^2 \times M^2}$ represents the relative distance matrix between any two patches in a window.

The second and third blocks of Figure B-1 depict two consecutive Swin Transformer blocks that use different window partitioning strategies. The first block uses a regular window partitioning strategy, starting from the top-left pixel, with 4×4 patches evenly split into 2×2 windows of size 2×2 patches. In the subsequent block, the windows are shifted by half the window size (i.e., one patch) in both dimensions, resulting in nine local windows of varied sizes. The self-attention mechanism is consistently computed within the local windows (marked with red frames) in both Swin Transformer blocks. This window-shifting scheme introduces connections between neighboring nonoverlapping windows in the preceding layer.

APPENDIX C

LEARNING CURVES IN TRANSFER LEARNING

In our proposed training workflow, two rounds of k -fold cross-validation experiments are con-

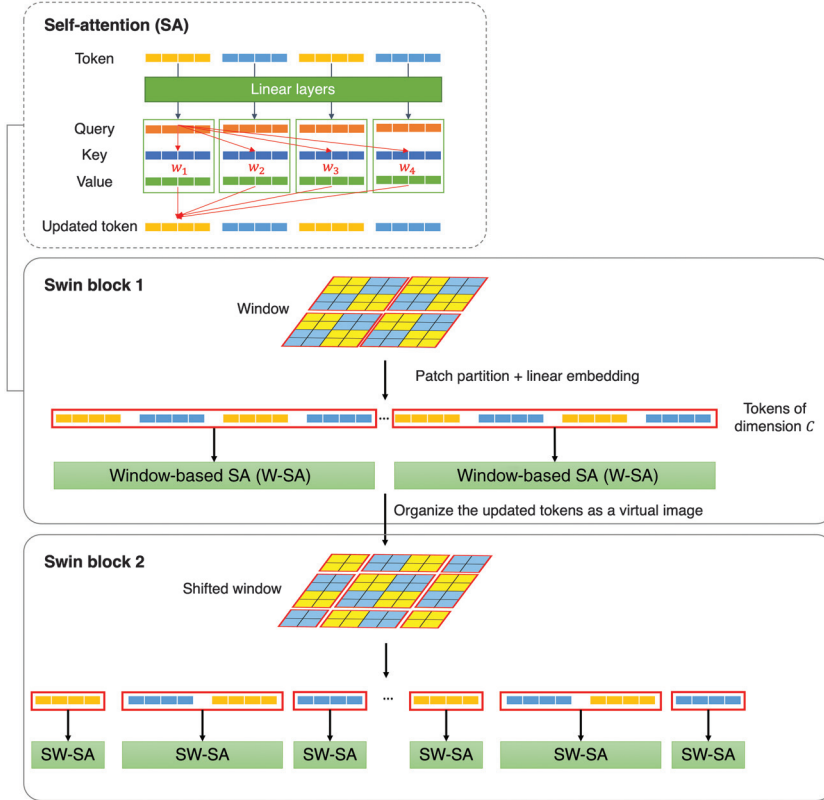


Figure B-1. Illustration of the Swin Transformer blocks (Liu et al., 2021, 2022), featuring the self-attention mechanism (SA) within a local window, window-based self-attention mechanism (W-SA), and shifted-window-based self-attention mechanism (SW-SA) arranged from top to bottom. Each grid in the image corresponds to a pixel, and the overall image is partitioned into 16 patches, each comprising 2×2 pixels. The shifting window, highlighted by the red frames, is of size 2×2 patches.

ducted on the prepared data. The initial training phase uses the original noisy manual labels. The learning curves for the second iteration, where fold 1 is designated as the test set, and the remaining folds form the training set, are illustrated with dashed curves in Figure C-1. Leveraging transfer learning, the training and test accuracies reach 88% and 74%, separately, after just one epoch. The

optimal model is chosen based on the lowest test loss. Prior to correcting label errors, a significant (more than 15%) difference exists between the training and validation accuracies for the best model, underscoring the need for further improvements in model generalizability. After data cleaning via CL, the learning curves for the same model structure, initialized with the same pretrained weights

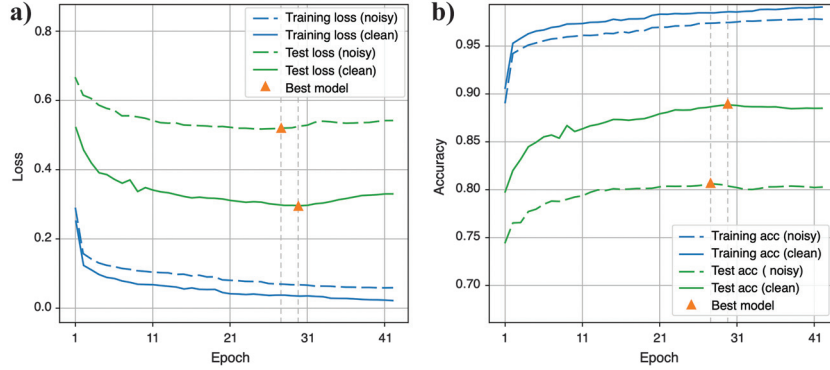


Figure C-1. Learning curves for the second iteration of the k -fold cross-validation process before and after data cleaning, wherein fold 1 serves as the test set and the remaining folds constitute the training set. (a and b) The evolution of loss and accuracy across different epochs, respectively. The dashed curves represent the learning curves before data cleaning, whereas the solid curves depict the curves after the cleaning process.

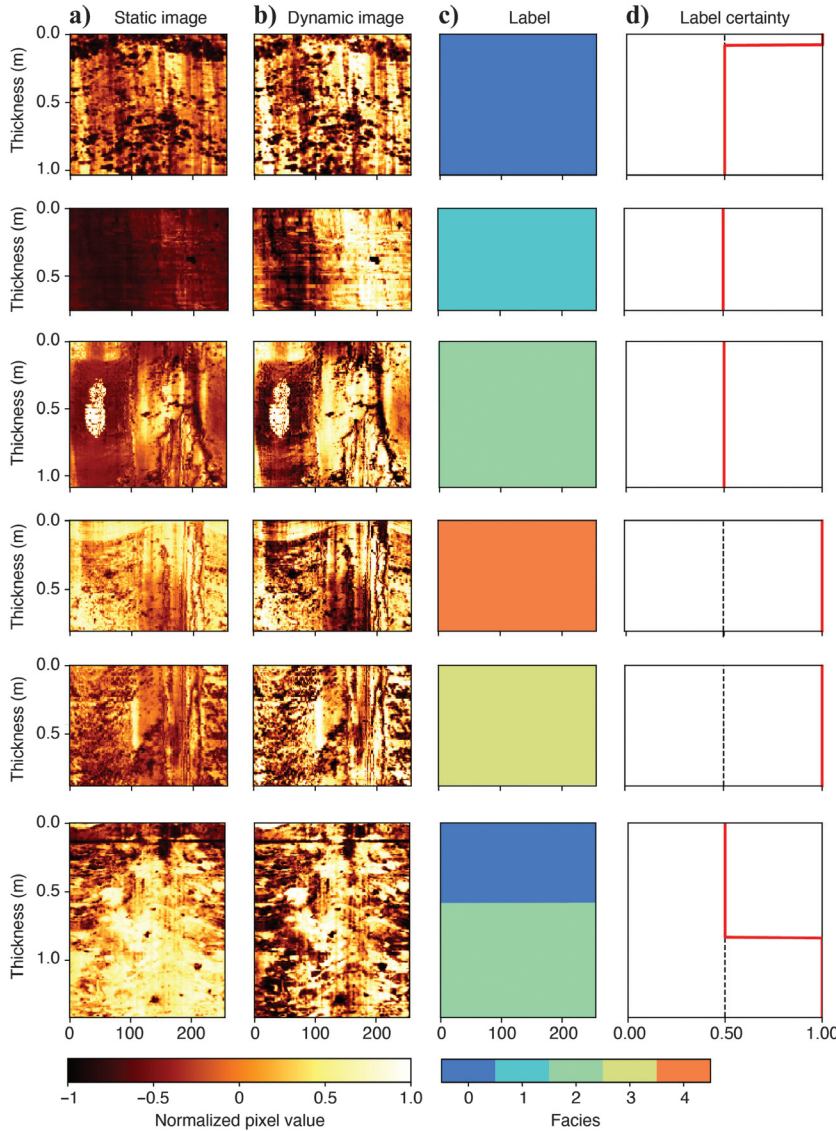


Figure D-1. (a–d) Examples of the mislabeled image log patches identified through CL. The label certainty panel reflects the human interpreter's confidence in assigned labels. A certainty value of 0.5 pertains to the less certain classes (5–7), whereas a value of 1.0 corresponds to the certain classes (0–4).

(the ImageNet-1K-pretrained SwinV2-T weights), are presented as solid curves in Figure C-1. The learning curves for the cleaned data exhibit trends similar to those trained with noisy labels, with the respective loss curves shifting downward and the accuracy curves shifting upward, indicating better classification performance. Moreover, the gap between the training and validation accuracies of the best model narrows to approximately 10%, emphasizing the enhanced generalizability achieved through CL.

APPENDIX D

CL DETECTED MISLABELED IMAGE LOG SAMPLES

Figure D-1 shows six randomly selected instances of the image log patches with false manual labels, as identified through CL. In general, these patches exhibit contamination from artifacts or low resolution, resulting in ambiguous feature patterns that do not align clearly with the defined facies. Consequently, most of these patches are assigned to a class with low certainty.

The first example contains widespread vugs that follow a sinusoidal trend but with no distinct laminations or shrubby features, making the manual label of facies 0 unsupported. The second patch does not contain complete laminations across the wellbore, confirming that the classification of facies 1 is inaccurate. The third patch contains various artifacts against a bright background, making it unconvincing to assign it to facies 2, which typically displays faint laminations or transparent beds. The fourth and fifth patches exhibit similar characteristics but are assigned to different classes, reflecting the inevitable inconsistency in manual labels. In both patches, traces of laminations and shrubs coexist with vertical artifacts, suggesting that they may belong to facies 0 in addition to facies 3 and 4. Consequently, it is challenging to classify both patches into any facies with sufficient confidence, and they might be better suited for the uncertain section category. The last patch presents scattered vugs and bright spots without clear layering or shrubs, diminishing the credibility of the corresponding labels, facies 0 and 2.

In conclusion, CL proves effective in detecting the inaccurate labels in our experiment, as evidenced by the randomly chosen samples. The detected false labels predominantly relate to low-quality images lacking the distinct features of the defined facies. Although CL successfully pinpoints these inaccuracies, the true labels for the identified false ones remain uncertain. Therefore, rather than correcting them, we opt to prune the false labels from the training data.

APPENDIX E

MODEL PERFORMANCE ON UNLABELED WELLS

We compare the facies predictions given by the SwinV2-Unet models obtained in tests 1, 5, and 10 for two sections from the unlabeled wells.

As for the first section shown in Figure E-1, the predictions given by the three models primarily vary between lines 1 and 5. The key difference between facies 1 and facies 2 lies in the presence or absence of clear, complete laminations across the wellbore. We observe clear, continuous laminations between lines 1 and 2 as well as between lines 3 and 4, indicating that these two sections belong to facies 1. Tests 5 and 10 correctly classify these two sections as facies 1, whereas test 1 incorrectly predicts them as facies 2. In general, for the section between lines 1 and 5, test 1 tends to underestimate facies 1, test 10 tends to overestimate facies 1, and test 5 consistently provides the most reliable results. In addition, test 10 identifies a possible presence of facies 0 between lines 4 and 5, but this seems unreliable as we do not observe visible shrubby features. The associated probability assigned by the neural network is also very low, at approximately 0.5. Overall, test 10 performs slightly better than test 1 in this example, whereas test 5 achieves the best performance.

For the second example shown in Figure E-2, test 1 performs the best in identifying the borehole breakouts, whereas test 10 fails to detect any of them. Test 5 identifies most of the borehole breakouts, although it is slightly less accurate than test 1. Specifically, the section between lines 1 and 2 is incorrectly predicted as facies 0 in test 5 due to the influence of the ambiguous small-scale v-shape patterns.

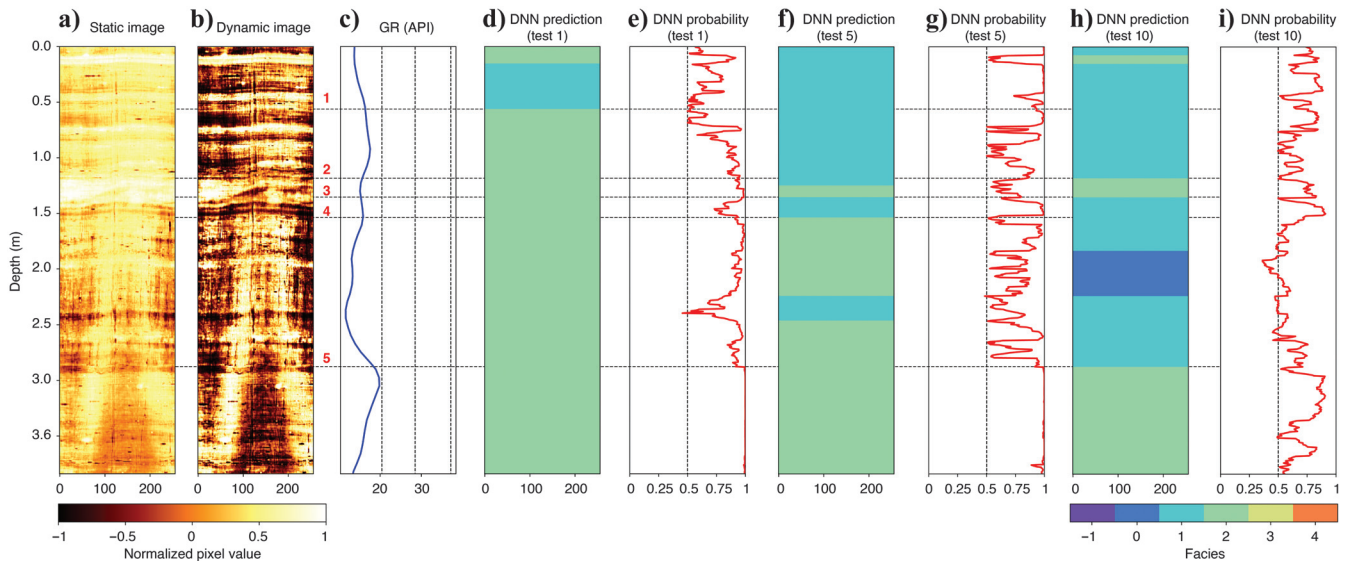
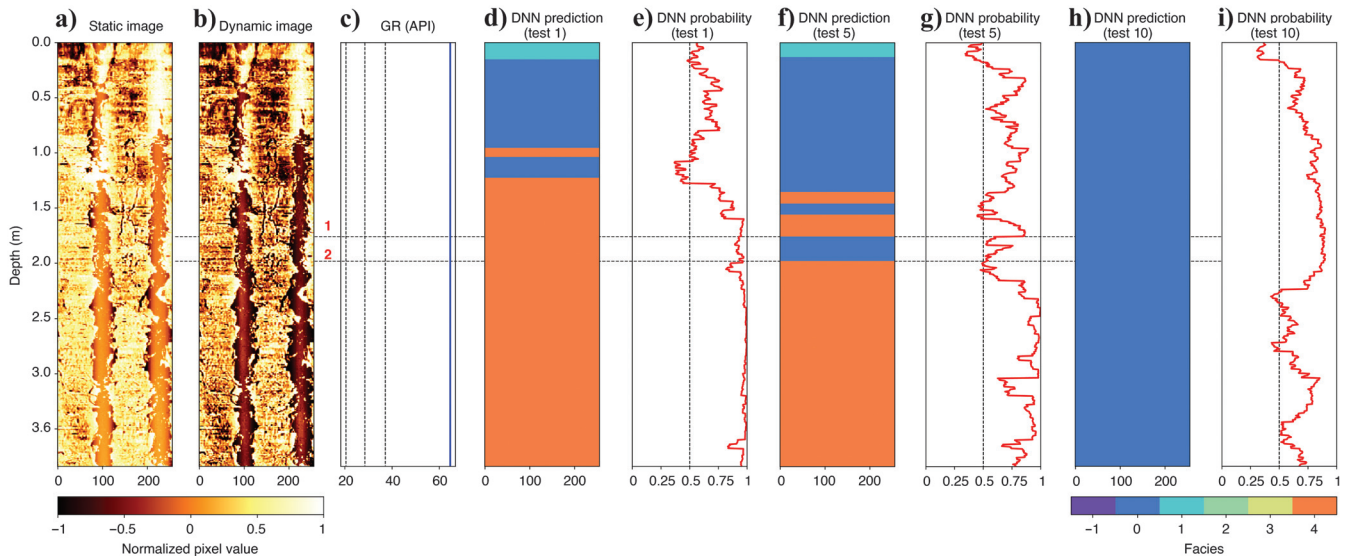


Figure E-1. (a-i) Comparison of the predictions given by tests 1, 5, and 10 for a section from the unlabeled wells.



Alkbar, M., B. Vissapragada, A. H. Alghamdi, D. Allen, M. Herron, A. Carnegie, D. Dutta, J.-R. Olesen, R. Chourasiya, D. Logan, and D. Stief, 2000, A snapshot of carbonate reservoir evaluation: *Oilfield Review*, **12**, 20–21.

Al-Sit, W., W. Al-Nuaimy, M. Marelli, and A. Al-Ataby, 2015, Visual texture for automated characterisation of geological features in borehole televiewer imagery: *Journal of Applied Geophysics*, **119**, 139–146, doi: [10.1016/j.jappgeo.2015.05.015](https://doi.org/10.1016/j.jappgeo.2015.05.015).

ANP, 2023, Painel dinâmico de produção de petróleo e gás natural, <https://app.powerbi.com/view?r=eyJrJoiNzVmNmZlMzQyNTYlNC00ZGVhLTk5N2ItNzBkMDNhY2IxZTlxliwidCI6IjQ0OTlmNGZmLTI0YTtYnG10Mi1lN2VmLTExYNGFmY2FkYzkyZkYzMyJ9>, accessed 6 June 2023.

Anwar, S., S. Khan, and N. Barnes, 2020, A deep journey into super-resolution: A survey: *ACM Computing Surveys*, **53**, 1–34, doi: [10.1145/3390462](https://doi.org/10.1145/3390462).

Basu, T., R. Dennis, B. Al-Khobar, W. Al Awadi, S. Isby, E. Vervest, and R. Mukherjee, 2002, Automated facies estimation from integration of core, petrophysical logs, and borehole images: Presented at the Annual Convention, AAPG.

Branco, C. C., and J. O. de Sant'Anna Pizarro, 2012, Challenges in implementing an EOR project in the pre-salt province in deep offshore Brasil: EOR Conference at Oil and Gas West Asia, SPE, doi: [10.2118/155665-MS](https://doi.org/10.2118/155665-MS).

Cao, H., Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, 2022, Swin-Unet: Unet-like pure transformer for medical image segmentation: European Conference on Computer Vision, Springer, 205–218.

Chai, H., N. Li, C. Xiao, X. Liu, D. Li, C. Wang, and D. Wu, 2009, Automatic discrimination of sedimentary facies and lithologies in reef-bank reservoirs using borehole image logs: *Applied Geophysics*, **6**, 17–29, doi: [10.1007/s11770-009-0011-4](https://doi.org/10.1007/s11770-009-0011-4).

da Costa Fraga, C. T., A. C. Capeleiro Pinto, C. C. M. Branco, J. O. de Sant'Anna Pizarro, and C. A. da Silva Paulo, 2015, Brazilian pre-salt: An impressive journey from plans and challenges to concrete results: Offshore Technology Conference, doi: [10.4043/25710-MS](https://doi.org/10.4043/25710-MS).

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, 2009, ImageNet: A large-scale hierarchical image database: IEEE Conference on Computer Vision and Pattern Recognition, 248–255.

Dias, L. O., C. R. Bom, E. L. Faria, M. B. Valentin, M. D. Correia, P. Márcio, P. Marcelo, and J. M. Coelho, 2020, Automatic detection of fractures and breakout patterns in acoustic borehole image logs using fast-region convolutional neural networks: *Journal of Petroleum Science and Engineering*, **191**, 107099, doi: [10.1016/j.petrol.2020.107099](https://doi.org/10.1016/j.petrol.2020.107099).

Donselaar, M. E., and J. M. Schmidt, 2005, Integration of outcrop and borehole image logs for high-resolution facies interpretation: Example from a fluvial fan in the Ebro Basin, Spain: *Sedimentology*, **52**, 1021–1042, doi: [10.1111/j.1365-3091.2005.00737.x](https://doi.org/10.1111/j.1365-3091.2005.00737.x).

Ennes, S., 1999–017, doi: [10.1017/cage.2008.08.011](https://doi.org/10.1017/cage.2008.08.011).

1962, Classification of carbonate rocks according to depositional textures: AAPG.

Gupta, K. D., V. Vallega, H. Maniar, P. Marza, H. Xie, K. Ito, and A. Abubakar, 2019, A deep-learning approach for borehole image interpretation: 60th Annual Logging Symposium, SPWLA, Extended Abstracts, doi: [10.30632/T60ALS-2019_BB](https://doi.org/10.30632/T60ALS-2019_BB).

Hall, B., 2016, Facies classification using machine learning: The Leading Edge, **35**, 906–909, doi: [10.1190/le35100906.1](https://doi.org/10.1190/le35100906.1).

Hall, J., M. Ponzi, M. Gonfalini, and G. Maletti, 1996, Automatic extraction and characterisation of geological features and textures from borehole images and core photographs: 37th Annual Logging Symposium, SPWLA, Extended Abstracts, SPWLA-1996-CCC.

Hastie, T., R. Tibshirani, and J. Friedman, 2009, The elements of statistical learning: Data mining, inference, and prediction: Springer.

Hochreiter, S., and J. Schmidhuber, 1997, Long short-term memory: *Neural Computation*, **9**, 1735–1780, doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).

Imamverdiyev, Y., and L. Sukhostat, 2019, Lithological facies classification using deep convolutional neural network: *Journal of Petroleum Science and Engineering*, **174**, 216–228, doi: [10.1016/j.petrol.2018.11.023](https://doi.org/10.1016/j.petrol.2018.11.023).

Jiang, J., R. Xu, S. C. James, and C. Xu, 2021, Deep-learning-based vuggy facies identification from borehole images: SPE Reservoir Evaluation & Engineering, **24**, 250–261, doi: [10.2118/204216-PA](https://doi.org/10.2118/204216-PA).

Lai, J., G. Wang, S. Wang, J. Cao, M. Li, X. Pang, C. Han, X. Fan, L. Yang, Z. He, and Z. Qin, 2018, A review on the applications of image logs in structural analysis and sedimentary characterization: *Marine and Petroleum Geology*, **95**, 139–166, doi: [10.1016/j.marpetgeo.2018.04.020](https://doi.org/10.1016/j.marpetgeo.2018.04.020).

Lima, L., N. Bize-Forest, A. Eysukoff, and R. Leonhardt, 2019, Unsupervised deep learning for facies pattern recognition on borehole images: Offshore Technology Conference.

Liu, Z., H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, and B. Guo, 2022, Swin transformer V2: Scaling up capacity and resolution: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 12009–12019.

Liu, Z., Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, 2021, Swin transformer: Hierarchical vision transformer using shifted windows: Proceedings of the IEEE/CVF International Conference on Computer Vision, 10012–10022.

Loschilov, I., and F. Hutter, 2017, Decoupled weight decay regularization: arXiv preprint, doi: [10.48550/arXiv.1711.05101](https://doi.org/10.48550/arXiv.1711.05101).

Muniz, M., and D. Bosence, 2015, Pre-salt microbialites from the Campos Basin (offshore Brazil): Image log facies, facies model and cyclicity in lacustrine carbonates: Geological Society, London, Special Publications, 221–242.

Northcutt, C. G., L. Jiang, and I. L. Chuang, 2021, Confident learning: Estimating uncertainty in dataset labels: *Journal of Artificial Intelligence Research*, **70**, 1373–1411, doi: [10.1613/jair.1.12125](https://doi.org/10.1613/jair.1.12125).

Oquab, M., L. Bottou, I. Laptev, and J. Sivic, 2014, Learning and transferring mid-level image representations using convolutional neural

- networks: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1717–1724.
- Parker, S. P., 1984, McGraw-Hill concise encyclopedia of science & technology: McGraw-Hill.
- Polyak, B. T., and A. B. Juditsky, 1992, Acceleration of stochastic approximation by averaging: *SIAM Journal on Control and Optimization*, **30**, 838–855, doi: [10.1137/0330046](https://doi.org/10.1137/0330046).
- Prensky, S. E., 1999, Advances in borehole imaging technology and applications: Geological Society, London, Special Publications, 1–43.
- Ronneberger, O., P. Fischer, and T. Brox, 2015, U-Net: Convolutional networks for biomedical image segmentation: 18th International Conference on Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015, Springer, 234–241.
- Tarvainen, A., and H. Valpola, 2017, Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results: *Advances in Neural Information Processing Systems*.
- Valentín, M. B., C. R. Bom, J. M. Coelho, M. D. Correia, P. Márcio, P. Marcelo, and E. L. Faria, 2019, A deep residual convolutional neural network for automatic lithological facies identification in Brazilian pre-salt oilfield wellbore image logs: *Journal of Petroleum Science and Engineering*, **179**, 474–503, doi: [10.1016/j.petrol.2019.04.030](https://doi.org/10.1016/j.petrol.2019.04.030).
- Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, 2017, Attention is all you need: *Advances in Neural Information Processing Systems*.
- Wilson, M. E., D. Lewis, D. Holland, L. Hombo, and A. Goldberg, 2013, Development of a Papua New Guinean onshore carbonate reservoir: A comparative borehole image (FMI) and petrographic evaluation: *Marine and Petroleum Geology*, **44**, 164–195, doi: [10.1016/j.marpetgeo.2013.02.018](https://doi.org/10.1016/j.marpetgeo.2013.02.018).
- Yang, J., X. Wu, Z. Bi, and Z. Geng, 2023, A multi-task learning method for relative geologic time, horizons, and faults with prior information and transformer: *IEEE Transactions on Geoscience and Remote Sensing*, **61**, 5907720, doi: [10.1109/TGRS.2023.3264593](https://doi.org/10.1109/TGRS.2023.3264593).
- Yang, S., Y. Wang, I. Le Nir, and A. He, 2020, Ai-boosted geological facies analysis from high-resolution borehole images: 61st Annual Logging Symposium, SPWLA, Extended Abstracts.
- Ye, S.-J., P. Rabiller, and N. Keskes, 1998, Automatic high resolution texture analysis on borehole imagery: 39th Annual Logging Symposium, SPWLA, Extended Abstracts.
- Yosinski, J., J. Clune, Y. Bengio, and H. Lipson, 2014, How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*.
- You, N., E. Li, and A. Cheng, 2023, Automatic facies classification from acoustic image logs using deep neural networks: *Interpretation*, **11**, no. 2, T441–T456, doi: [10.1190/INT-2022-0069.1](https://doi.org/10.1190/INT-2022-0069.1).
- Zhu, X. J., 2005, Semi-supervised learning literature survey: Technical report, Computer Sciences, University of Wisconsin-Madison.

Biographies and photographs of the authors are not available.