234 (2024) 212662

Contents lists available at ScienceDirect

Geoenergy Science and Engineering

journal homepage: www.sciencedirect.com/journal/geoenergy-science-and-engineering

DDViT: Advancing lithology identification on FMI image logs through a dual modal transformer model with less information drop

Li Hou [a,b], Chao Ma [a,b,*], Wenqiang Tang [a,b], Yuxuan Zhou [a,b], Shan Ye [c], Xiaodong Chen [d], Xingxing Zhang [d], Congyu Yu [a,b], Anqing Chen [a,b], Dongyu Zheng [a,b], Zhisong Cao [a,b], Yan Zhang [e], Mingcai Hou [a,b]

[a] State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation & Institute of Sedimentary Geology & College of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu 610059, China
[b] Key Laboratory of Deep-time Geography and Environment Reconstruction and Applications of Ministry of Natural Resources, Chengdu University of Technology, Chengdu 610059, China
[c] School of Information Engineering, China University of Geosciences (Beijing), Beijing, 100083, China
[d] Qinghai Oilfield Company, PetroChina, Dunhuang 736202, China
[e] Key Laboratory of Gold Mineralization Processes and Resource Utilization, Ministry of Natural Resources, Shandong Provincial Key Laboratory of Metallogenic Geological Process and Resource Utilization, Shandong Institute of Geological Sciences, Jinan 250013, China

ARTICLE INFO

Keywords:
Lithology identification
FMI images
Dual modal
Drop less information

ABSTRACT

Lithology is an essential topic in oil and gas reservoir studies. Lithological observation lays the foundation for assessing oil and gas prospects and guides future exploration and development. Currently, the prevailing approach in lithological observation heavily depends on the manual analysis of core samples. However, such an approach is highly subjective and time-consuming. With the development of deep learning, some automated deep learning-based methods have been proposed for lithology interpretation from logging curves. However, the Fullbore Formation MicroImager (FMI) image logging, while widely used in the oil field exploration and development process, is occasionally deployed and utilized for lithology identification. In this work, we proposed a **D**ual-modal **D**rop-less-information **Vi**sion **T**ransformer (DDViT), an FMI image lithology identification model based on transformer architecture. DDViT uses a dual-modal architecture to identify lithology using two different image modalities, namely dynamic FMI images (FMI_DYN) and static FMI images (FMI_STAT). These modalities reflect the local and overall characteristics, respectively. Furthermore, DDViT uses a less-information dropping module to drop the blank band information inherent in the FMI images to make our model more rational and stable. DDViT achieved a 90.81% lithology identification accuracy on Fengxi Well A of the western Qaidam Basin, providing a new approach to lithology identification and demonstrating the great potential of deep learning in geological images.

1. Introduction

Lithology is an essential aspect in describing rock properties (Tucker and Jones, 2023) and it contains information reflecting the Earth's structure and evolutionary processes (Haldar, 2020; Roberts, 2021). Rocks of the Earth are categorized into sedimentary, metamorphic, and igneous rocks (Philpotts and Ague, 2022), with sedimentary rocks covering 75% of the Earth's land surface. The research of sedimentary rocks is not only of academic interest but also of significant economic value. By studying the different combinations and alignments of lithology in sedimentary rocks, one can determine the different types of sedimentary facies (Zou and Qiu, 2021), which provide insights into the geological histories and processes of environmental evolution. Moreover, the diverse lithology types can serve as indicators of distinct climates and sedimentary conditions, enabling the estimation of paleoclimatic changes (Hou et al., 2023). Apart from its role in scientific research, lithology also holds immense importance in the industry (Wang and He, 2020; Li, J. et al., 2021). Various lithologies possess

* Corresponding author. State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation & Institute of Sedimentary Geology & College of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu 610059, China.
E-mail address: machao@cdut.edu.cn (C. Ma).

https://doi.org/10.1016/j.geoen.2024.212662
Received 5 August 2023; Received in revised form 25 December 2023; Accepted 10 January 2024
Available online 14 January 2024

distinct physical and chemical characteristics (Tian et al., 2016; Anees et al., 2022), which have a significant influence on the origin of oil and gas reservoirs (Hu et al., 2020; Jin et al., 2020; Wu, X. et al., 2021).

Lithology identification is commonly achieved through core sampling and the utilization of various observation techniques such as thin sections, scanning electron microscopy, and optical microscopy (Fu et al., 2017). These methods enable a comprehensive examination of the core's grain size, color, and composition, leading to a well-informed and conclusive lithological interpretation (Hall et al., 1996). Although reliable core analysis results can be obtained by these methods, these manual observation methods are intensely subjective and require a professional level of skillset to make robust analyses. In addition, these manual methods often demand a significant amount of time and labor effort. Therefore, in recent times, geologists have been actively exploring the possibilities of automated lithology identification techniques to enhance efficiency and reduce the need for repetitive labor (Imamverdiyev and Sukhostat, 2019; Valentín et al., 2019; Alzubaidi et al., 2021; Shehata et al., 2021; Santos et al., 2022; Zheng et al., 2022).

In recent years, the appearance of AlexNet (Krizhevsky et al., 2012) and its excellent performance in image classification and speech recognition have paved the way for subsequent advancements in neural network architectures and techniques. These advancements have resulted in significant breakthroughs in various domains of artificial intelligence. ResNet (He et al., 2016) was then proposed to make deeper and larger models possible, significantly boosting deep learning (DL). In the years that followed, deep learning began to thrive for various tasks in computer vision (Rawat and Wang, 2017; Minaee et al., 2022; Zou et al., 2023). Inspired by the great success of transformer architecture in Natural Language Processing (NLP) (Vaswani et al., 2017), the Vision Transformer model (ViT) (Dosovitskiy et al., 2020) first adopted the transformer architecture for computer vision and achieved success. The study of the transformer architecture extends to the input patch token, as DynamicViT (Rao et al., 2021) proposed a dynamic strategy for discarding useless tokens. SPViT (Kong et al., 2021) notes differences between each head in ViT's multi-head attention mechanism, deciding on individual heads for each patch token before their aggregation. Evit (Liang et al., 2022) proposed a parameter-free decision strategy that decides whether to discard a token based on the attention score calculated by the self-attention module and reuses the discarded token by aggregating it. In addition, deep learning is not limited to studying a single image data source as the processing of multiple image data sources is also booming. For example, CMX (Liu et al., 2022) can use many different data modal pairs, such as RGB and depth, RGB and thermal, and RGB and LiDAR, to work on the same task. Using these heterogeneous data sources to analyze the same target often results in a reliable and comprehensive interpretation (Zhang, Y. et al., 2021).

The development of artificial intelligence (AI), especially DL techniques (He et al., 2016; Krizhevsky et al., 2017; Vaswani et al., 2017; Dosovitskiy et al., 2021; Rao et al., 2021), has led to an increasing number of interdisciplinary applications, such as ChatGPT (Radford et al., 2018, 2019; Brown et al., 2020; OpenAI, 2023), autonomous driving (Chen et al., 2016), medical diagnosis (Rajpurkar et al., 2022), and industrial fault detection (Roth et al., 2022). In the field of geology, ongoing research and collaborations between geologists and data scientists is progressively refining and adapting DL techniques to tackle geological challenges (Maitre et al., 2019; Karimpouli and Tahmasebi, 2019; Sun et al., 2020; Li, S. et al., 2021; Chen et al., 2021; Saxena et al., 2021). Regarding lithology identification in oil and gas exploration, Xie et al. (2018) employed five different supervised learning models to identify lithologies in two gas fields, and subsequently evaluated the performance of these models. Valentín et al. (2019) used ultrasonic images and micro resistivity borehole image logs to identify the lithology of the São Francisco basin. Imamverdiyev and Sukhostat (2019) used logging curve information to identify the lithology and compared the performance of various optimizer algorithms. Shehata et al. (2021) combined different types of data, including log curves and FMI images, to predict petrography, permeability, and rock types of Beni Suef Basin. Alzubaidi et al. (2021) developed a CNN-based method to classify lithology of core images automatically and achieved excellent results. Xie et al. (2021) proposed a coarse-to-fine approach for addressing outliers in the dataset to further enhance the accuracy of lithologic classification. Santos et al. (2022) proposed a computational system based on a deep recurrent neural network (RNN) to judge the lithology of the Rio Bonito Formation and thoroughly compared their proposed method with various machine learning (ML) methods. Xie et al. (2023) proposed a semi-supervised, coarse-to-fine approach for lithology identification, designed to enhance lithologic classification accuracy with limited labeled data.

However, the emphasis on the methods mentioned above is often placed on investigating logging curves, while the study of FMI images is still in its progressive stages, holding significant potential for further development (Imamverdiyev and Sukhostat, 2019; Shehata et al., 2021; Santos et al., 2022). Since different lithologies present different characteristics under electric current, the FMI image can also reflect the lithology (Cui et al., 2013). Moreover, the FMI images consist of two distinct modalities, which are dynamic FMI images and static FMI images, each exhibiting its distinctive features that contribute to a more comprehensive interpretation of the lithology. Additionally, a significant portion of the research typically relies solely on traditional ML methods, such as Random Forest and support vector machine (SVM), as well as on convolutional neural network (CNN) methods for the identification of lithology. Despite the successful application of ML and CNN methods in lithology identification, ML is suitable for simpler data problems, whereas CNN is limited to operating on input data with local windows. In contrast, the transformer architecture employs a self-attention mechanism to capture long-range dependencies in images. Additionally, the transformer architecture can handle data of variable size. Thus, considering the superior performance of ViT architectures in diverse domains of computer vision, the application of ViT architectures to lithology identification is imperative. Nonetheless, FMI images often contain unavoidable blank bands, which are caused by the electrical imaging logger used in the FMI imaging process. These blank bands cover a substantial portion of the FMI image, necessitating processing to assist geologists in their research endeavors. Various popular existing approaches use deep learning networks for filling the blank bands in FMI images (Zhang, H. et al., 2021; Du et al., 2022; Sun et al., 2023), and all aiming to reduce the impact of blank bands on geologists by filling them. Diverging from prior studies, this paper's focus lies in constructing a deep learning model that autonomously identifies lithology through FMI images, hence our focus is on reducing the disturbance caused by blank bands for the model. To sum up, in this work, we proposed a model called DDViT for lithology identification based on FMI images. DDViT simultaneously processes the dynamic FMI images (FMI_DYN) and static FMI images (FMI_STAT) (see details in section 2) to obtain two types of feature information. It then uses these two types of feature information to determine the lithology type. Furthermore, DDViT employs a less-information dropping module to automatically remove blank bands, reducing distractions for the model and enhancing its stability. This setup allows DDViT to selectively concentrate on critical areas of the FMI image for precise lithological identification, demonstrating the rationale of the model.

2. Data

2.1. Data source

This paper utilizes image data from Fengxi Well A of the western Qaidam Basin in Qinghai Province, China, and these images are obtained from Schlumberger's MAXIS-500 logging series. The depth range of those images in the drilling well is 2780–4480 m. Based on drilling, logging, and core descriptions, the lithology of the studied well section predominantly contains mudstones, non-algal limestones, and algal

limestones, with minor sandstones.

FMI images were produced by a full borehole micro-resistivity scanning electrical imaging logger (Fu et al., 2023). FMI is an electrical imaging logging device that records the formation's characteristics as it changes with different current values by emitting currents around the well wall during logging (Yin et al., 2009). FMI takes these characteristics and converts them into the resistivity of the formation around the well wall and generates the FMI images. During the imaging process, different imaging techniques can separate the static FMI images (FMI_STAT) and the dynamic FMI images (FMI_DYN). The static FMI image uses a uniform standard for the whole well section in the imaging so that the static FMI image reflects the overall macroscopic characteristics of the formation (Shafiezadeh et al., 2015). In contrast, dynamic FMI images use a window-based imaging setting. Although it causes the final image to lose the macroscopic characteristics of the entire well, it does enhance the local features at each level, allowing the FMI image to show subtle characteristics (Goodall et al., 1998). FMI images can be employed to assess lithology by capturing differences in rock structure. For instance, mudstone typically exhibits well-defined layering on FMI images, displaying distinct horizontal bedding and occasional calcareous nodule development. Sandstone layers appear thick and fine-grained, with noticeable massive bedding or cross-bedding patterns. In the case of non-algal and algal limestone, FMI images often reveal massive bedding and floc-like or cloud-like textures. Furthermore, the two FMI image modalities exhibit distinct characteristics due to variations in imaging features. For example, within the same layer, FMI_DYN displays higher contrast and more vibrant colors compared to FMI_STAT, highlighting localized changes in resistivity within the well section. In this study, as shown in Fig. 1, the FMI_DYN image and the FMI_STAT image were used to identify the lithology.

2.2. Data preprocessing and data distribution

Due to their high resolution, FMI images pose challenges when directly used for model training. As lithology descriptions typically follow a layer-based approach, data preprocessing was performed to appropriately align the FMI images with the classification task. The FMI image, extracted from the logging report, has dimensions of 4995 pixels in height and 1037 pixels in width. To facilitate lithology analysis, we divided the original FMI image into smaller, non-overlapping images. Each of these has a standard size of 256 pixels in height and 1024 pixels in width, dimensions commonly used in computer vision. We defined each smaller image as a unit. This division retains most of the original data and simplifies subsequent processing. Subsequently, 1125 units were selected from the divided units to create the dataset for this paper. For the selection strategy, low-quality FMI images excluded, such as over-white or blurred images, which make up a significant proportion of the low-quality images, and we then calculated the lithology percentage within each unit and designated the lithology with the highest percentage as the working category for subsequent steps. Subsequently, algal limestone, non-algal limestone, and sandstone were picked from the filtered units based on the working category. For mudstone, random sampling was performed to ensure a similar quantity as the maximum number among the first three lithologies to prevent dataset over imbalance. Then, a random sample of 100 units was selected from 1125 as the testing dataset. Finally, the remaining units were divided into sample data with an image size of 16 × 1024, and these sample data were further split into training and validation datasets in a proportion of 7:3. The distribution of the dataset after processing is shown in Fig. 2.

3. Method

In this section, we introduce our proposed DDViT model and show the details of the design of each module, and finally, we show the loss functions we use.

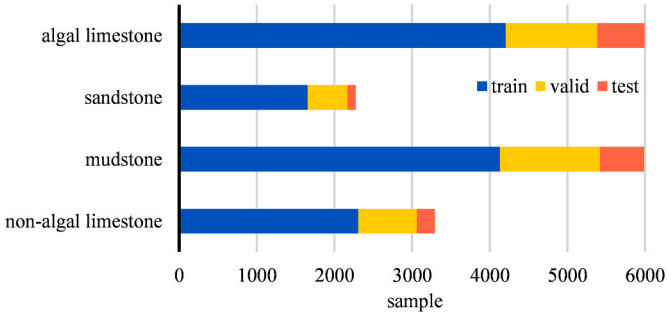

Fig. 2. The distribution of our dataset. The blue is represented as the training dataset, yellow as the validation dataset and red as the testing dataset.

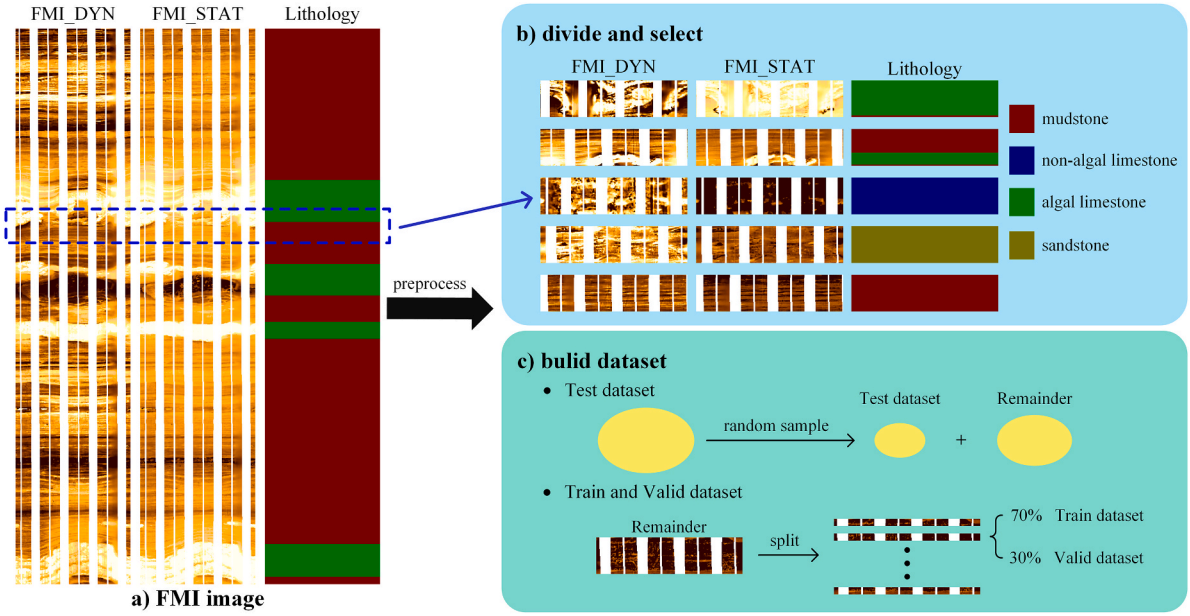

Fig. 1. a) Example data used in this study. b) and c) The workflow of our dataset-building method.

3.1. Overview

The overall framework of the proposed DDViT model is depicted in Fig. 3. The DDViT model consists of two main stages: 1) the parallel primary stage of feature information extraction, and 2) the base backbone network encoding stage. The parallel feature information extraction aims to extract the feature information of the FMI images of both modalities. After obtaining the feature information of both modalities, the extracted feature information is aggregated and given to the backbone network, which consists of multi-transformer blocks for encoding to obtain the lithological classification results. In addition, during the preliminary stages of parallel information extraction, DDViT employs the less-information dropping (LID) module to discard disruptive and redundant information (e.g., blank band information in FMI images), enhancing the model's stability and credibility in lithological classification.

3.1.1. Dual-modal architecture

Before accessing the dual-modal parallel module, it is necessary to preprocess the input images. The FMI_DYN images and the FMI_STAT images are defined as $X_{DYN}, X_{STAT} \in \mathbb{R}^{C\times H\times W}$. First, X_{DYN} and X_{STAT} are divided into N small patch tokens that $\mathbb{R}^{C\times H\times W} \Rightarrow \mathbb{R}^{C\times P^2\times N}$, where $N = HW/P^2$ and P is the size of the patch token. Then they are flattened to $\mathbb{R}^{N\times D}$ where $D = P^2 \times C$. The next step utilizes the patch embedding module to map the features of X_{DYN} and X_{STAT} while preserving these mapped features as X_{DYN} and X_{STAT}. Following this mapping, they are concatenated with their respective class token (cls) and accompanied by the positional embedding (PE). The class token serves as a representation of the category, and it aggregates all the feature information during both training and inference, ultimately facilitating the prediction of the final classification. The positional embedding plays a crucial role in allowing the network to capture the positional information of each patch, as the attention module itself is unable to inherently capture the positional information of the input sequence (Vaswani et al., 2017).

$$X_{DYN},\ X_{STAT} = (\mathrm{Cat}(cls_{DYN}, X_{DYN}), \mathrm{Cat}(cls_{STAT}, X_{STAT})) + PE \tag{1}$$

Finally, X_{DYN}, X_{STAT} are forwarded to the dual-modal parallel processing module (Dual) to calculate the respective feature mapping outputs for each modality. The parallel architecture is employed to extract features from both FMI_DYN and FMI_STAT patterns, and subsequently, the information from these two patterns is aggregated into the next module to derive the lithology classification results. Due to the distinct characteristics and distributions of the two modal FMI images, employing a single module for feature extraction from these differing FMI images would not be appropriate. Hence, a parallel structure has been devised, wherein each FMI image is directed to its dedicated feature information extraction process. This approach enables the model to comprehensively extract the distinct feature information inherent to both modalities, consequently enhancing its capacity to discern lithology. Then the output of the two modalities is concatenated and sent to the remainder module (Backbone) for encoding and feature mapping. In the final output, after multiple layers of encoding and feature mapping, all the class tokens within it are extracted and averaged (Mean). Finally, the lithology recognition results are categorized based on the class head (Class). This process is shown in Fig. 3(a).

$$\begin{aligned} X_{DYN}, X_{STAT} &= \mathrm{Dual}(X_{DYN}, X_{STAT}) \\ X &= \mathrm{Backbone}(\mathrm{Cat}(X_{DYN}, X_{STAT})) \\ Facies &= \mathrm{Class}(\mathrm{Mean}(cls_{DYN}, cls_{STAT})) \end{aligned} \tag{2}$$

3.1.2. Transformer block

The transformer block is the primary encoding block in DDViT, used for efficient feature extraction from the input FMI image and providing strong support for final lithology identification. The transformer block is the main component of the vision transformer (Dosovitskiy et al., 2021), as shown in Fig. 3(c). It is composed of a multi-head self-attention block (MHSA), an MLP block, and a LayerNorm layer (LN). Residual connections are incorporated to facilitate gradient preservation, thus aiding the training process. The core of the transformer block is its self-attention layer (Vaswani et al., 2017) within MHSA, an adaptive control mechanism that makes the network pay attention to features that have a substantial impact on the prediction results. Precisely, self-attention captures the importance of each patch concerning other patches through a set of learnable parameters. The input of the transformer block is defined as $X \in \mathbb{R}^{n\times d}$, where n is the number of patches and d is the feature dimension of each patch. First, obtain $Q \in \mathbb{R}^{n\times d_q}$, $K \in \mathbb{R}^{n\times d_k}$, $V \in \mathbb{R}^{n\times d_v}$ by a simple MLP block, where $d_q = d_k$.

$$Q, K, V = \mathrm{MLP}(X) \tag{3}$$

Subsequently, the self-attention layer (*Attention*) computes the attention matrix by performing a dot operation between Q and K, followed by normalizing the attention matrix using a *Softmax* process to yield the attention scores. The output of the *Attention* is finally obtained by multiplying the attention score with V, with the inclusion of a scaling factor of $\frac{1}{\sqrt{d_k}}$ to prevent the dot product between Q and K from yielding excessively large values.

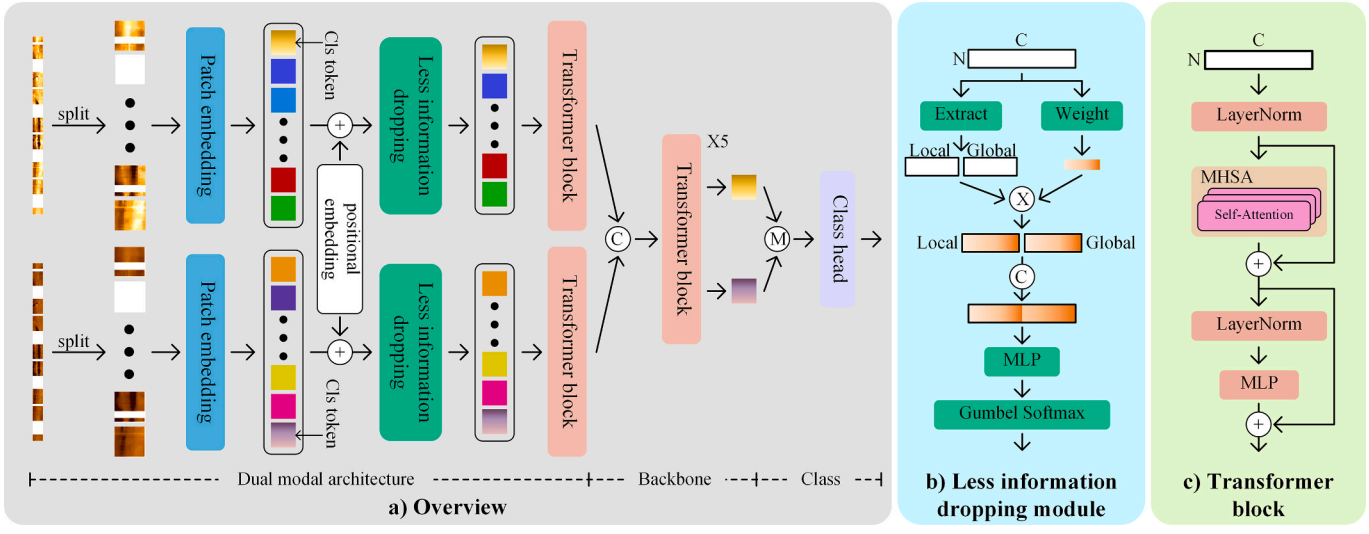

Fig. 3. a) Overview of our proposed model DDViT. b) The detail of our less-information dropping module. c) The detail of the transformer block.

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

A multi-head self-attention block (MHSA) is a parallel composition of h self-attention layers, wherein X is divided into h parts along the d dimension. Each part is processed by a dedicated self-attention layer. The results of these self-attention layers are subsequently concatenated together along the d dimension to obtain the final output of the multi-head self-attention block.

$$\mathrm{MHSA} = \mathrm{Cat}(Attention_i), i \in [1,h] \tag{5}$$

In addition to the MHSA block described previously, the transformer block also includes an MLP layer, which serves to project the feature information into the final result space. Furthermore, it incorporates an LN (LayerNorm) layer, which normalizes the feature information to enhance network stability and facilitate ease of training.

3.1.3. Less-information dropping module

In the context of FMI images, it is important to acknowledge that regular blank bands may appear as a result of the imaging process. These blank bands have the potential to impede and disrupt the analysis of FMI images. A proficient identification model for FMI images is expected to possess the capability to prioritize significant information within the FMI image, while simultaneously disregarding distracting information such as blank band information. Moreover, it is assumed that the model possesses the ability to autonomously learn the relevance of each piece of information for precise lithology identification. Building upon these assumptions, we capture intermediate information during the model's processing, subject it to mapping and assessment to ascertain its significance, and subsequently make decisions regarding its keep or drop. To fulfill this purpose, DDViT implements a less-information dropping (LID) module, which empowers the model to identify and eliminate blank bands. Please refer to Fig. 3(b) for a detailed illustration of this process.

Inspired by DynamicViT (Rao et al., 2021), $X \in \mathbb{R}^{n\times d}$ is defined as the input, and a binary decision encoding $M \in \mathbb{R}^n$ is defined to represent each token's state ('keep' or 'drop'). All values in M are initialized to 1, which means keeping every token used in the model process. Similarly, we first use the Extract block to extract the global and local feature information $I_{global}, I_{local} \in \mathbb{R}^{n\times\frac{d}{2}}$ of X and combine them to obtain the complete feature information $I \in \mathbb{R}^{n\times d}$. The Extract block primarily employs linear layers for information extraction.

$$\begin{aligned} I_{global}, I_{local} &= \mathrm{Extract}(X) \\ I &= \mathrm{Cat}\left(I_{global}, I_{local}\right) \end{aligned} \tag{6}$$

Furthermore, since each head in ViT's multi-head self-attention focuses on distinct content within the image, it follows that the importance of each head varies (Kong et al., 2021). Consequently, relying solely on I is insufficient for determining the weight of each token and deciding keep or drop. Thus, a learnable head weighting strategy is adopted to rescale the extracted information I. This operation shares similarities with SPViT (Kong et al., 2021). However, in the case of SPViT, weight calculations are performed within each head, and it does not consider the interaction information between heads. In contrast to SPViT, weights of LID are computed from each head within X through a *Weight* block. In the *Weight* block, the shape of X is first transformed following $\mathbb{R}^{n\times d}\Rightarrow\mathbb{R}^{h\times n\times\frac{d}{h}}$, where h represents the number of heads used in the network. Subsequently, the *Weight* block employs a multilayer convolutional layer to model the importance of each head in X, resulting in $W_{head} \in \mathbb{R}^h$. During the operation of *Weight* block, the information from each head in X is involved in all computations. This ensures that interactions among all heads are fully considered, a distinct approach from that of SPViT, which restricts its computation to the information of a single head when determining head weights. Finally, when assigning W_{head} to I, the information I undergoes a shape transformation according to $\mathbb{R}^{n\times d}\Rightarrow\mathbb{R}^{h\times n\times\frac{d}{h}}$. However, once I successfully learns the significance of each head, it reverts to its original shape.

$$\begin{aligned} W_{head} &= Weight(X) \\ I &= I * W_{head} \end{aligned} \tag{7}$$

Lastly, the LID module utilizes the derived information from I to decide whether to keep or drop each token and subsequently updates this decision in M. To make a drop decision, LID employs Gumbel-Softmax (Jang et al., 2016) to sample from the information in I based on the content therein. However, before executing this, the LID must conduct a feature mapping on the information I using the MLP block, ensuring the precision of the LID's decision regarding the token.

$$M = M * GumbelSoftmax(MLP(I)) \tag{8}$$

Dropped tokens don't participate in any subsequent model computations. In the training phase, to maintain the model's differentiability, LID employs the Gumbel-Softmax technique and an attention masking strategy (Rao et al., 2021). These approaches are designed to eliminate the impact of dropped tokens during attention computation. The defined binary decision encoding, M, indicates the location of the dropped token. This facilitates the attention masking strategy, effectively ignoring the dropped token in computations and ensuring stable model training. During the inference phase, differentiability is not a concern, allowing LID to directly drop tokens based on their importance in the information, I. The reduced data is then computed using the standard attention computation mechanism. As the LID's token drop operation is anchored in the transformer architecture's patch token mechanism, LID might also be compatible with other models based on the transformer architecture.

3.2. Loss function

A standard cross-entropy function is employed as the classification loss in this study, and cross-entropy functions are commonly utilized in neural networks for addressing classification problems (He et al., 2016; Dosovitskiy et al., 2020). Furthermore, an additional loss function, the Mean Squared Error (MSE), is utilized to control the dropping rate within the LID module. There are two inputs to the MSE loss function (Equation (9)), namely the predefined target keeping rate, denoted as r, and the model's actual keep rate during training, which can be obtained by averaging over the binary decision encoding M. The model's overall loss function is obtained by weighting these two loss functions, where η is the scaling factor. The value of η is set to 2 following DynamicViT (Rao et al., 2021). For the choice of the keeping rate r, it was observed that the blank band typically constituted approximately 50% of the FMI images. Therefore, we conducted experiments with r values of 0.4, 0.5, 0.6, and 0.7, respectively. Subsequent to experimental comparisons, it was determined that the model achieved optimal performance when r was set to 0.6.

$$\mathscr{F}_{loss} = \mathscr{F}_{ce}\left(y_{pred}, y_{label}\right) + \eta\mathscr{F}_{mse}(r, \mathrm{Mean}(M)) \tag{9}$$

4. Results

The experiments were conducted on two NVIDIA GeForce 3080 GPUs, and the code for training and testing was implemented using Python 3.8.15 within the PyTorch 1.10.0 framework. During the training phase, a learning rate of 0.00005 and a batch size of 128 were utilized. The AdamW optimizer was employed, and the model parameters were updated over 100 epochs.

4.1. The performance of our proposed model

DDViT was trained on the constructed dataset. Similarly, ResNet50 (He et al., 2016), ViT (Dosovitskiy et al., 2021), CvT (Wu, H. et al., 2021), and MViTv2 (Li et al., 2022) were trained for making

comparisons. For the single input models, namely ResNet50, ViT, CvT, and MViTv2, we assessed their performance using each of the three distinct input methods: DYN-only, STAT-only, and DYN-STAT mixed. In the DYN-only and STAT-only, the models were trained exclusively on FMI_DYN and FMI_STAT data, respectively. For the DYN-STAT mixed method, a combination of FMI_DYN and FMI_STAT data was utilized. Each three input methods feeding only one image to the model at a time. Regarding the dual-modal architecture of DDViT, we evaluated the DYN-STAT parallel input method, where FMI_DYN and FMI_STAT are input into the model simultaneously. For a detailed overview of the input methods, please refer to Fig. 4. Additionally, DDViT differs from other models in terms of loss functions. DDViT incorporates the cross-entropy function and MSE as its loss function, while other models solely utilize the cross-entropy function. Except for the differences in the data input and the loss function used, the remaining training settings, such as optimizer setting, learning rate, and number of training epochs, are the same.

The results of DDViT and comparison models are shown in Table 1. DDViT performed best in lithology identification, achieving a classification accuracy of 90.81%. Compared to DDViT, the accuracies of other models are lower in all three different input scenarios. As shown in Fig. 5, we visualize the prediction results of some test samples on these models, and the visualization also demonstrates the excellent performance achieved by DDViT.

4.2. The performance of dual-modal architecture

To show the effect of our proposed dual-modal architecture, we test the DDViT (without the LID module) compared to the original ViT (Dosovitskiy et al., 2021). For ViT, they it also uses the three input scenarios mentioned in the previous section. In the DDViT model without the LID module (w/o LID), the use of the MSE loss function to control the drop rate r is unnecessary. Instead, DDViT employs the same loss function as utilized in the ViT. All other settings remain unchanged; in this scenario, the sole distinction between the two models lies in the dual-modal architecture. The results are shown in Table 1. Compared with the three input scenarios of ViT, the dual-modal architecture (DYN-STAT parallel) achieves the best performance with an accuracy of 89.31%. Additionally, it's noteworthy that the STAT-only scenario consistently outperformed both the DYN-only and DYN-STAT mixed input scenarios.

4.3. The performance of the less-information dropping module

To further explore the performance of the LID module, the dropping results were visualized in Fig. 6. Fig. 6 shows that the LID module can learn and locate the blank bands and redundant information in the input image and remove most of these unnecessary elements. It keeps those elements of color and morphological information that indicate lithology.

The experiment of using SPViT's (Kong et al., 2021) multi-head selector directly (w/SPViT) compared with our LID module is also conducted, and the results are shown in Table 1. Utilizing the LID module yields an accuracy of 90.81%, outperforming the 88.94% accuracy achieved using the strategy mentioned in SPViT. LID proves to be superior, since it can fully capture the internal interaction information of each head when computing the information I.

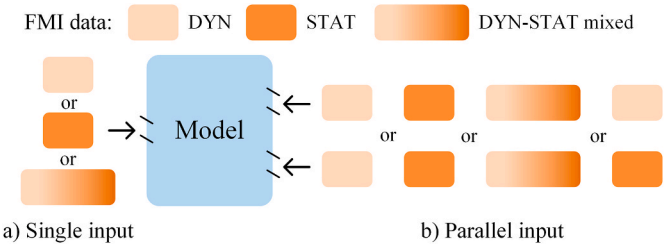

Fig. 4. Demonstration of various input methods. a) A single input, with components arranged from top to bottom corresponding to DYN-only, STAT-only, and DYN-STAT mixed, respectively. b) A parallel input, with components arranged from left to right corresponding to DYN parallel, STAT parallel, DYN-STAT mixed parallel, and DYN-STAT parallel.

Table 1
The result of our proposed model with other models.

	input method	accuracy	precision	recall	F1-score
ResNet50	DYN-only	76.50	75.82	67.59	70.47
	STAT-only	84.13	82.87	81.68	82.06
	DYN-STAT mixed	80.44	79.29	71.37	74.18
ViT	DYN-only	83.63	81.48	79.41	80.37
	STAT-only	86.19	83.81	84.93	85.21
	DYN-STAT mixed	85.94	85.51	85.42	85.44
CvT	DYN-only	79.25	74.25	74.17	74.09
	STAT-only	84.62	82.17	79.59	80.70
	DYN-STAT mixed	83.25	81.25	81.86	81.50
MViTv2	DYN-only	84.62	83.74	78.89	80.98
	STAT-only	88.31	88.32	86.75	87.40
	DYN-STAT mixed	86.03	86.42	83.94	85.01
DDViT (w/o LID)	DYN-STAT parallel	89.31	88.74	88.32	88.51
DDViT (SPViT)		88.94	87.84	87.96	87.84
DDViT		**90.81**	**90.22**	**90.06**	**90.14**

5. Discussion

DDViT is specifically crafted within the context of geological prior knowledge, incorporating FMI imaging and image features. Models founded on geological prior knowledge usually demonstrate increased effectiveness and rationality.

On the one hand, for the FMI imaging feature, the identification of lithology based on only one kind of FMI image (DYN-only or STAT-only) may lead to a restricted condition for the lithology identification, which is a waste of data due to the lack or inadequate utilizing the diversity of the FMI imaging process. To validate this concept, we assessed the performance of the DDViT model using three additional distinct input methods: DYN parallel, STAT parallel, and DYN-STAT mixed parallel. The first two are defined such that both inputs are either FMI_DYN or FMI_STAT, respectively. The third is designed wherein both inputs comprise a mixture of FMI_DYN and FMI_STAT. Details can be found in Fig. 4(b). As shown in Table 2, the results indicated that the model achieved an accuracy of 80.62% with DYN parallel, and 87.25% with STAT parallel. These accuracies are lower compared to the 90.81% achieved when employing the DYN-STAT parallel. In addition, when employing the DYN-STAT mixed, the model has difficulty learning a fixed pattern to match the two different modalities. As shown in Table 1, the accuracies of ResNet, ViT, CvT, and MViTv2 are 80.44%, 85.94%, 83.25%, and 86.03%, respectively, when using the DYN-STAT mixed input method. However, both accuracies are lower than the best performance of ResNet50 (84.13%), ViT (86.19%), CvT (84.62%), and MViTv2 (88.31%). Similarly, as evidenced in Table 2, the DDViT model achieves an accuracy of only 83.53% with mixed inputs, which is lower than the highest recorded accuracy of 90.81%. Given the distinct data characteristics and distributions of the two modalities, where FMI_DYN captures local window features and FMI_STAT reflects global features, using the same model to jointly process may impede the learning of each modality's unique features. Consequently, parallel processing of FMI_DYN and FMI_STAT images can minimize the interference caused by

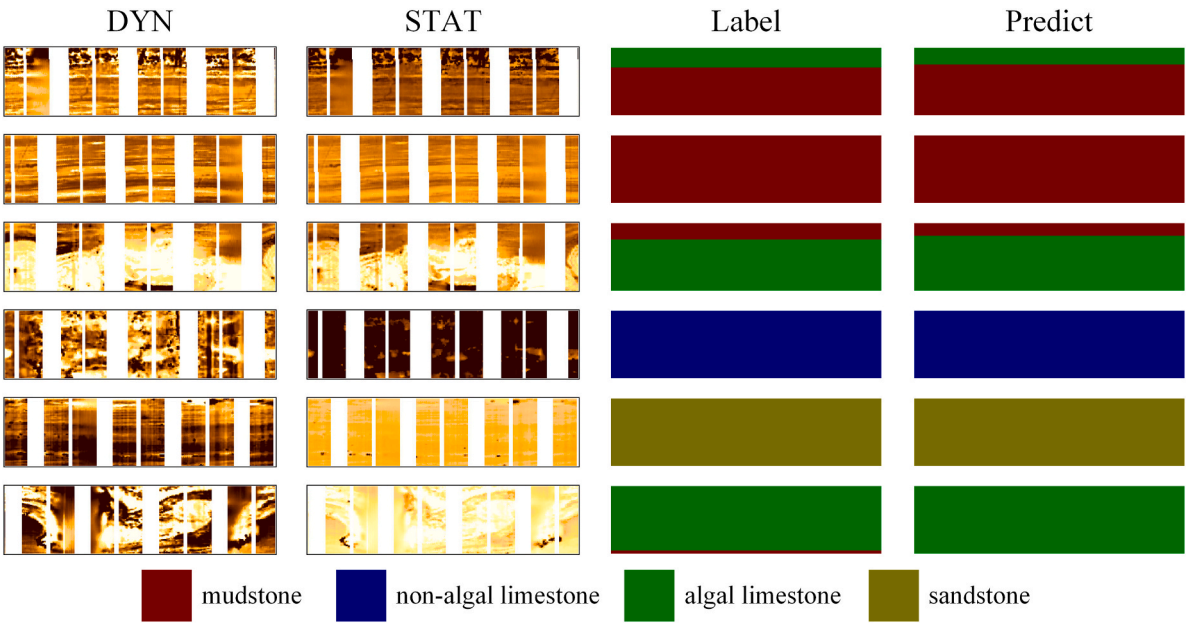

Fig. 5. Visualize the prediction of our model. FMI_DYN images and FMI_STAT images are represented by DYN and STAT, respectively. Label denotes the actual lithology that corresponds to the data, while Predict refers to the lithology predicted by our model.

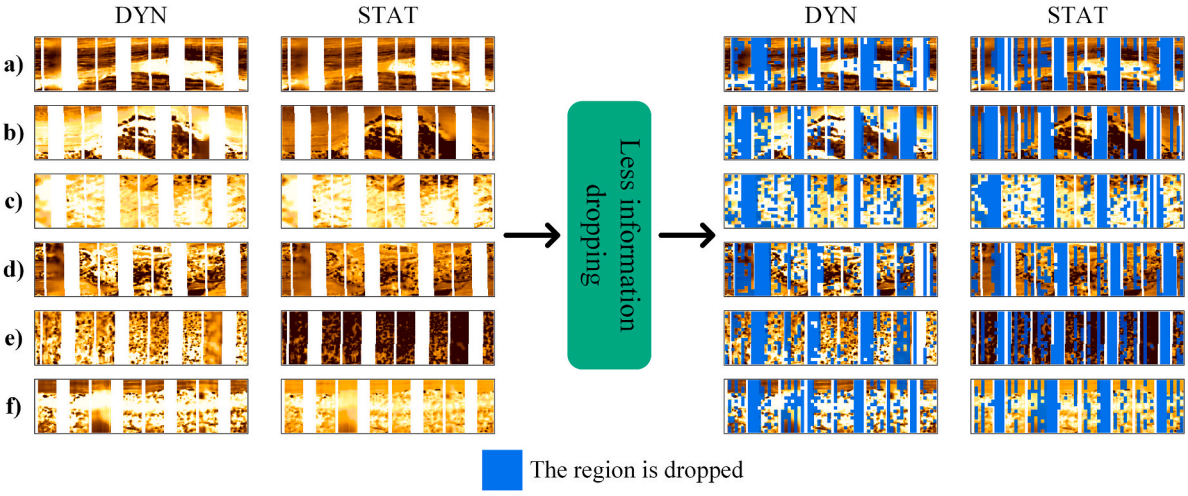

Fig. 6. Visualize the output of the FMI images after the LID module process. The blue region represents dropping.

Table 2
Comparison of using DYN parallel, STAT parallel, and DYN-STAT mixed parallel methods with the DYN-STAT parallel method.

	input method	accuracy
DDViT	DYN parallel	80.62
	STAT parallel	87.25
	DYN-STAT mixed parallel	83.53
	DYN-STAT parallel	**90.81**

their differing data characteristics and distributions. This approach not only enhances the model's performance but also aligns better with the characteristics of FMI imaging. Experimental results (Table 1) demonstrate the superiority of the parallel approach over other methods. It was also observed that using only FMI_STAT images was better than using only FMI_DYN images, and in ResNet50, ViT, CvT, and MViTv2 with STAT-only (84.13%, 86.19%, 84.62% and 88.31%) achieved superior performance over the DYN-only input method (76.50%, 83.63%, 79.25% and 84.62%). As observed in Table 2, a similar phenomenon is evident in the DDViT model, where the performance of the STAT parallel, at 87.25%, surpasses that of the DYN parallel, which stands at 80.62%. This may be because the FMI_STAT image reflects global information, which makes the FMI image use a uniform imaging standard for any case of imaging. This imaging setting allows all FMI_STAT images to have the same data distribution, which makes it easier for the model to learn the feature information of the FMI_STAT image.

Alongside the focused on the accuracy of lithology identification using FMI images, this study statistically analyzes other variables associated with each model, including model size, training time, and inference time. The comprehensive results are presented in Table 3. All data are included from the results of each model when operating at its lithological identification peak performance. Notably, DDViT possesses the largest model size, attributable to its utilization of a dual-modal architecture to process different modalities in FMI. This parallel architecture needs additional parameter branching relative to other models, resulting in a larger model size for DDViT. Despite this, the dual-modal architecture can enhance the exploitation of the FMI image's different

Table 3
Additional variables involved in the experiments for all models include model parameters, inference time, and training time.

	Params (M)	inference time (ms)	training time (min)
ResNet50	25.56	2.55	26.88
ViT	38.60	1.08	25.23
CvT	31.21	6.66	56.12
MViTv2	33.99	6.24	65.33
DDViT	49.11	2.45	59.62

modalities, leading to more precise lithology identification. Despite its size, DDViT does not lag behind other models in terms of training and inference time. It should be noted that DDViT needs to process two types of data, whereas other models handle only one. This efficiency stems from the LID module, which drops out certain redundant and disturbing data during modeling, consequently reducing computational demands and accelerating calculations.

On the other hand, due to the inherent imaging deficiencies of the FMI imaging technique, the FMI images contain a high proportion of blank bands and redundant information that don't contribute to lithology identification. Such cases may even disturb and interfere with the model's identification of the lithology. Thus, DDViT uses a LID module to remove the blank band information generated during the FMI imaging technique, as well as to remove some redundant information from the FMI images. The LID module includes an additional hyperparameter, r. Initially, we conducted an empirical analysis of the FMI image, which revealed that the blank band in the FMI image approximately occupied 50% of the total image area. Hence, we conducted experiments around the suggested value of $r = 0.5$. The test results, presented in Table 4, demonstrated that the model achieved optimal performance at $r = 0.6$. Furthermore, an ablation experiment on the LID module was performed, which showed that the LID module achieved the best performance of 90.81%, outperforming the performance without the LID module (89.31%). In addition, LID uses different head weights to rescale the extracted information and optimize the algorithm in SPViT to maintain the interaction information within each token. Additionally, we verified the improvement in accuracy (90.81%) compared to the original SPViT, which achieved an accuracy of only 88.94%. Fig. 6 depicts the visualized processing effect of the LID module. Fig. 6 illustrates the significant reduction of blank band information achieved by the LID module. Additionally, it was observed that small portion of non-blank band regions were also eliminated. It may indicate that these regions correspond to background information within the current layer. This observation is particularly evident in Fig. 6(a) and (b), where crucial details such as flocculent structures and prominent features are retained while less significant information, which would have a minor impact on lithology identification, is selectively discarded by the LID module. The visualization demonstrates that DDViT can perform lithology identification by focusing on the informative regions within the image, without relying on blank bands of information. This approach emphasizes the utilization of the image's relevant regions for identification, enhancing the interpretability of DDViT, similar to human actions of observation for FMI images.

Given that features along the wellbore may be autocorrelated, this work also explored the use of stratified sampling for dataset construction. The cleaned set of 1125 units served as the full dataset, with each unit designated as a stratum. The training, validation, and testing datasets were subsequently divided within each stratum according to the stratified sampling method, following a 7:3:1 ratio. The performance of DDViT was assessed on this newly constructed dataset under the same experimental conditions. The findings are detailed in Table 5. From the presented results, the accuracy achieved through random sampling is marginally superior to that of stratified sampling. However, the minor difference in accuracy is not the primary reason for preferring random sampling. Owing to the stratified nature of logging FMI images, there is a frequent observation that FMI images from the same or neighboring strata exhibit significant similarity. Consequently, if both training and testing datasets are formed using stratified sampling, their data distributions might be strikingly similar, leading to only minimal distinctions between them. Under such scenarios, the testing dataset might not be adequate to accurately assess the model's performance. On further examination of the random sampling method employed in this study, it becomes apparent that this method diverges from traditional stratified sampling primarily during the construction of the testing dataset. In this paper's method, a certain number of strata (units) are randomly chosen from all available strata to comprise the testing dataset. This contrasts with stratified sampling, which randomly extracts a portion from each stratum for the testing dataset. This approach offers an advantage over stratified sampling, as it provides better control over the differences between the testing and training datasets.

Overall, this work has achieved promising results in lithology identification by incorporating the imaging characteristics and properties of the FMI images into our model. Furthermore, the visualization of the LID module demonstrates the robustness of DDViT and strengthens confidence in its ability to accurately identify lithology based on the FMI images. This is because the model effectively filters out distribution information in the FMI images. These findings highlight the significant potential and promise of integrating geological a priori knowledge into DL. However, DDViT exhibited incorrect predictions for certain lithological identifications, as depicted in Fig. 7. For example, the model misidentifies a small portion of mudstone as algal limestone, possibly due to the presence of minimal continuous porosity in an incorrect location (Fig. 7(a)). Fig. 7(b) demonstrates the model's failure in accurately delineating the boundary between algal limestone and non-algal limestone, primarily attributed to the challenge of precisely identifying the flocculent structure boundary within the algal limestone. Prediction errors in Fig. 7(c) and (d) occur within regions characterized by feature disarray. Furthermore, Fig. 7(d), predominantly composed of sandstone, presents challenges in model learning due to the limited representation of sandstone samples in the source dataset. Additionally, the characteristics of sandstone distribution in the formation (mainly located below, thin, and little) and the higher geothermal temperature during sandstone sample collection contribute to lower image quality compared to the other three lithologies.

However, we acknowledge that our work has limitations; although this study demonstrates the effectiveness of the method for identifying lithology in FMI imaging logs, the energy picture output from different FMI imaging logs exhibits variations, requiring appropriate model tuning. Furthermore, this work is only experimenting with FMI images from a single area and needs to use more data to train and test our model in future work. In addition, this work only discusses and uses FMI images for lithology identification, while many other data are also essential for lithology identification, such as GR (Gamma Ray Logging), DEN (Density Logging), and AC (Acoustic Logging). In the future, we will use as much relevant data as possible to combine them and make a

Table 4
The results of the model for different values of r.

	r	accuracy
DDViT	0.4	90.44
	0.5	90.31
	0.6	**90.81**
	0.7	89.81

Table 5
Assess the performance of DDViT under both stratified and random sampling methodologies.

	accuracy	precision	recall	F1-score
stratified sampling	90.75	**90.30**	89.90	90.09
random sampling	**90.81**	90.22	**90.06**	**90.14**

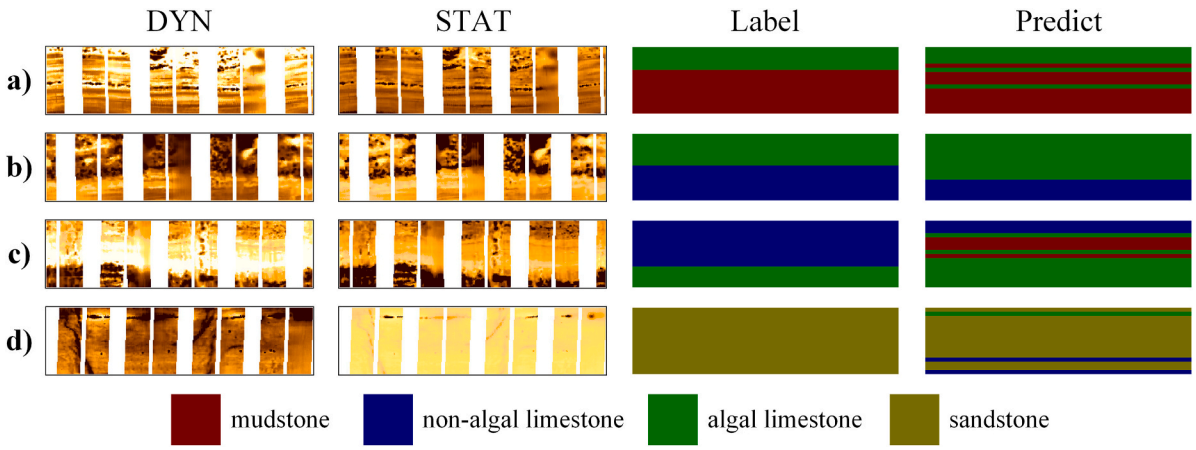

Fig. 7. Here are a few examples of incorrect predictions.

comprehensive identification of lithology based on their respective geological characteristics.

6. Conclusion

This paper introduces DDViT, a lithology identification model that utilizes information from FMI images during logging to interpret lithology. DDViT is built upon the image and imaging characteristics of the FMI. DDViT achieves a lithology identification accuracy of 90.81% on data from the Fengxi Well A of the western Qaidam Basin in Qinghai Province, China. DDViT utilizes a dual-modal architecture, simultaneously processing the FMI_DYN and FMI_STAT images to extract their respective features for lithology interpretation. To address the presence of unavoidable blank bands in FMI images, DDViT implemented a less-information dropping module that reduces the influence of blank bands on lithology identification. This module automatically learns the location of blank bands and excludes them during model training and prediction. DDViT utilization of a dual-modal architecture and less-information dropping module has yielded promising results in terms of model accuracy and rationality, highlighting the potential of DDViT for lithology identification in FMI images. DDViT is specifically designed to utilize FMI images alone for accurate lithology identification and it offers an alternative solution for future DL applications in the field of lithology identification. The design concept of DDViT is rooted in the inherent characteristics of geological data, and it offers valuable insights to other geoscientists exploring the applications of DL in their research.

Coda and data availability

Code and data will be made available on request.

CRediT authorship contribution statement

Li Hou: Data curation, Investigation, Methodology, Validation, Visualization, Writing – original draft. **Chao Ma:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Wenqiang Tang:** Formal analysis, Investigation, Writing – review & editing. **Yuxuan Zhou:** Writing – review & editing. **Shan Ye:** Writing – review & editing. **Xiaodong Chen:** Data curation, Resources. **Xingxing Zhang:** Data curation, Resources. **Congyu Yu:** Writing – review & editing. **Anqing Chen:** Writing – review & editing. **Dongyu Zheng:** Writing – review & editing. **Zhisong Cao:** Writing – review & editing. **Yan Zhang:** Writing – review & editing. **Mingcai Hou:** Writing – review & editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

We would like to acknowledge students Shengjian Zhou and Longgang Ye for their passionate and helpful discussions. The editors and reviewers are warmly thanked for their diligent work and invaluable comments on this work. This work was financially supported by the National Natural Science Foundation of China (No. 42050102, 42050104, 41888101, and 42172137), the National Key R&D Program of China (Grant No. 2023YFF0804000), and Sichuan Provincial Youth Science & Technology Innovative Research Group Fund (No.2022JDTD0004). This study is a contribution to the Deep-time Digital Earth (DDE) Big Science Program and IGCP 739.

References

Alzubaidi, F., Mostaghimi, P., Swietojanski, P., Clark, S.R., Armstrong, R.T., 2021. Automated lithology classification from drill core images using convolutional neural networks. J. Petrol. Sci. Eng. 197, 107933 https://doi.org/10.1016/j.petrol.2020.107933.

Anees, A., Zhang, H., Ashraf, U., Wang, R., Liu, K., Mangi, H.N., Jiang, R., Zhang, X., Liu, Q., Tan, S., others, 2022. Identification of favorable zones of gas accumulation via fault distribution and sedimentary facies: insights from Hangjinqi Area, Northern Ordos Basin. Front. Earth Sci. 9, 1375.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., others, 2020. Language models are few-shot learners. Adv. Neural Inf. Process. Syst. 33, 1877–1901.

Chen, L., Lin, W., Chen, P., Jiang, S., Liu, L., Hu, H., 2021. Porosity prediction from well logs using back propagation neural network optimized by genetic algorithm in one heterogeneous oil reservoirs of Ordos Basin, China. J. Earth Sci. 32, 828–838.

Chen, X., Kundu, K., Zhang, Z., Ma, H., Fidler, S., Urtasun, R., 2016. Monocular 3D object detection for autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Cui, J., Zhang, J., Zhang, H., 2013. Features of the carboniferous volcanic rocks fracture reservoirs in Hongshanzui oilfield, Junggar Basin. J. Earth Sci. 24, 997–1007.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., others, 2020. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale arXiv preprint arXiv: 2010.11929.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. ICLR.

Du, C., Xing, Q., Zhang, J., Wang, J., Liu, B., Wang, Y., 2022. Blank strips filling for electrical logging images based on attention-constrained deep generative network. Prog. Geophys. 37, 1548–1558.

Fu, G., Yan, J., Zhang, K., Hu, H., Luo, F., 2017. Current status and progress of lithology identification technology. Prog. Geophys. 32, 26–40.

Fu, Z., Lu, S., Wang, H., He, T., 2023. Natural fracture occurrence model based on FMI imaging logging. ACS Omega 8, 2034–2045.

Goodall, T., Møller, N., Rønningsland, T., 1998. The integration of electrical image logs with core data for improved sedimentological interpretation. Geological Society, London, Special Publications 136, 237–248.
Haldar, S.K., 2020. Introduction to Mineralogy and Petrology. Elsevier.
Hall, J., Ponzi, M., Gonfalini, M., Maletti, G., 1996. Automatic extraction and Characterisation of geological features and textures front borehole images and core photographs. Presented at the SPWLA 37th Annual Logging Symposium. SPWLA-1996-CCC.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Hou, H., Shao, L., Tang, Y., Li, Y., Liang, G., Xin, Y., Zhang, J., 2023. Coal seam correlation in terrestrial basins by sequence stratigraphy and its implications for paleoclimate and paleoenvironment evolution. J. Earth Sci. 34, 556–570.
Hu, Y., Peng, X., Li, Q., Li, L., Hu, D., 2020. Progress and development direction of technologies for deep marine carbonate gas reservoirs in the Sichuan Basin. Nat. Gas. Ind. B 7, 149–159.
Imamverdiyev, Y., Sukhostat, L., 2019. Lithological facies classification using deep convolutional neural network. J. Petrol. Sci. Eng. 174, 216–228. https://doi.org/10.1016/j.petrol.2018.11.023.
Jang, E., Gu, S., Poole, B., 2016. Categorical Reparameterization with Gumbel-Softmax. International Conference on Learning Representations,International Conference on Learning Representations.
Jin, F., Huang, J., Pu, X., Ma, C., Fu, L., Leng, C., Lou, D., Qin, M., 2020. Characteristics of the Cretaceous magmatism in Huanghua Depression and their relationships with hydrocarbon enrichment. J. Earth Sci. 31, 1273–1292.
Karimpouli, S., Tahmasebi, P., 2019. Segmentation of digital rock images using deep convolutional autoencoder networks. Comput. Geosci. 126, 142–150. https://doi.org/10.1016/j.cageo.2019.02.003.
Kong, Z., Dong, P., Ma, X., Meng, X., Sun, M., Niu, W., Shen, X., Yuan, G., Ren, B., Qin, M., others, 2021. Spvit: Enabling Faster Vision Transformers via Soft Token Pruning arXiv preprint arXiv:2112.13890.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90. https://doi.org/10.1145/3065386.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc.
Li, J., Tao, X., Bai, B., Huang, S., Jiang, Q., Zhao, Z., Chen, Y., Ma, D., Zhang, L., Li, N., others, 2021. Geological conditions, reservoir evolution and favorable exploration directions of marine ultra-deep oil and gas in China. Petrol. Explor. Dev. 48, 60–79.
Li, S., Chen, J., Liu, C., Wang, Y., 2021. Mineral prospectivity prediction via convolutional neural networks based on geological big data. J. Earth Sci. 32, 327–347.
Li, Y., Wu, C.-Y., Fan, H., Mangalam, K., Xiong, B., Malik, J., Feichtenhofer, C., 2022. MViTv2: improved multiscale vision transformers for classification and detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4804–4814.
Liang, Y., Ge, C., Tong, Z., Song, Y., Wang, J., Xie, P., 2022. Not All Patches Are what You Need: Expediting Vision Transformers via Token Reorganizations arXiv preprint arXiv:2202.07800.
Liu, H., Zhang, J., Yang, K., Hu, X., Stiefelhagen, R., 2022. CMX: Cross-Modal Fusion for RGB-X Semantic Segmentation with Transformers arXiv preprint arXiv:2203.04838.
Maitre, J., Bouchard, K., Bédard, L.P., 2019. Mineral grains recognition using computer vision and machine learning. Comput. Geosci. 130, 84–93.
Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D., 2022. Image segmentation using deep learning: a survey. IEEE Trans. Pattern Anal. Mach. Intell. 44, 3523–3542. https://doi.org/10.1109/TPAMI.2021.3059968.
OpenAI, 2023. GPT-4 Technical Report.
Philpotts, A.R., Ague, J.J., 2022. Principles of Igneous and Metamorphic Petrology. Cambridge University Press.
Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., others, 2018. Improving Language Understanding by Generative Pre-training.
Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., others, 2019. Language models are unsupervised multitask learners. OpenAI blog 1, 9.
Rajpurkar, P., Chen, E., Banerjee, O., Topol, E.J., 2022. AI in health and medicine. Nat. Med. 28, 31–38.
Rao, Y., Zhao, W., Liu, B., Lu, J., Zhou, J., Hsieh, C.-J., 2021. DynamicViT: efficient vision transformers with dynamic token sparsification. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P.S., Vaughan, J.W. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 13937–13949
Rawat, W., Wang, Z., 2017. Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput. 29, 2352–2449. https://doi.org/10.1162/neco_a_00990.
Roberts, N.M., 2021. An Introduction to Metamorphic Petrology.
Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P., 2022. Towards total recall in industrial anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14318–14328.
Santos, D.T. dos, Roisenberg, M., Nascimento, M., dos, S., 2022. Deep recurrent neural networks approach to sedimentary facies classification using well logs. Geosci. Rem. Sens. Lett. IEEE 19, 1–5. https://doi.org/10.1109/LGRS.2021.3053383.
Saxena, N., Day-Stirrat, R.J., Hows, A., Hofmann, R., 2021. Application of deep learning for semantic segmentation of sandstone thin sections. Comput. Geosci. 152, 104778.
Shafiezadeh, M., Ziaee, M., Tokhmechi, B., 2015. A new approach towards precise planar feature characterization using image analysis of FMI image: case study of gachsaran oil field well no. 245, south west of Iran. Journal of Petroleum Science and Technology 5, 51–58.
Shehata, A.A., Osman, O.A., Nabawy, B.S., 2021. Neural network application to petrophysical and lithofacies analysis based on multi-scale data: an integrated study using conventional well log, core and borehole image data. J. Nat. Gas Sci. Eng. 93, 104015 https://doi.org/10.1016/j.jngse.2021.104015.
Sun, D., Xu, J., Wen, H., Wang, Y., 2020. An optimized random forest model and its generalization ability in landslide susceptibility mapping: application in two areas of three gorges reservoir, China. J. Earth Sci. 31, 1068–1086.
Sun, Q., Su, N., Gong, F., Du, Q., 2023. Blank strip filling for logging electrical imaging based on multiscale generative adversarial network. Processes 11, 1709.
Tian, H., Ma, Z., Chen, X., Zhang, H., Bao, Z., Wei, C., Xie, S., Wu, S., 2016. Geochemical characteristics of selenium and its correlation to other elements and minerals in selenium-enriched rocks in Ziyang County, Shaanxi Province, China. J. Earth Sci. 27, 763–776.
Tucker, M.E., Jones, S.J., 2023. Sedimentary Petrology. John Wiley & Sons.
Valentín, M.B., Bom, C.R., Coelho, J.M., Correia, M.D., de Albuquerque, Márcio, P., de Albuquerque, Marcelo, P., Faria, E.L., 2019. A deep residual convolutional neural network for automatic lithological facies identification in Brazilian pre-salt oilfield wellbore image logs. J. Petrol. Sci. Eng. 179, 474–503. https://doi.org/10.1016/j.petrol.2019.04.030.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc.
Wang, J., He, Z., 2020. Responses of stream geomorphic indices to piedmont fault activity in the Daqingshan area of China. J. Earth Sci. 31, 978–987.
Wu, H., Xiao, B., Codella, N., Liu, M., Dai, X., Yuan, L., Zhang, L., 2021. CvT: introducing convolutions to vision transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 22–31.
Wu, X., Guo, Q., Zhang, W., He, D., Qi, X., Li, D., 2021. Characteristics of volcanic reservoirs and hydrocarbon accumulation of carboniferous system in junggar basin, China. J. Earth Sci. 32, 972–985.
Xie, Y., Jin, L., Zhu, C., Wu, S., 2023. A semi-supervised coarse-to-fine approach with bayesian optimization for lithology identification. Earth Science Informatics 16, 2285–2305. https://doi.org/10.1007/s12145-023-01014-7.
Xie, Y., Zhu, C., Hu, R., Zhu, Z., 2021. A coarse-to-fine approach for intelligent logging lithology identification with extremely randomized trees. Math. Geosci. 53, 859–876. https://doi.org/10.1007/s11004-020-09885-y.
Xie, Y., Zhu, C., Zhou, W., Li, Z., Liu, X., Tu, M., 2018. Evaluation of machine learning methods for formation lithology identification: a comparison of tuning processes and model performances. J. Petrol. Sci. Eng. 160, 182–193. https://doi.org/10.1016/j.petrol.2017.10.028.
Yin, X.-C., Liu, Q., Hao, H.-W., Wang, Z.-B., Huang, K., 2009. A rock structure recognition system using FMI images. In: Neural Information Processing: 16th International Conference, ICONIP 2009, Bangkok, Thailand, December 1-5, 2009, Proceedings, Part I 16. Springer, pp. 838–845.
Zhang, H., Sima, L., Wang, L., GuoQiong, C., Guo, Y., Yang, Q., 2021a. Blank strip filling method for resistivity imaging image based on convolution neural network. Prog. Geophys. 36, 2136–2142.
Zhang, Y., Sidibé, D., Morel, O., Mériaudeau, F., 2021b. Deep multimodal fusion for semantic image segmentation: a survey. Image Vis Comput. 105, 104042.
Zheng, D., Hou, M., Chen, A., Zhong, H., Qi, Z., Ren, Q., You, J., Wang, H., Ma, C., 2022. Application of machine learning in the identification of fluvial-lacustrine lithofacies from well logs: a case study from Sichuan Basin, China. J. Petrol. Sci. Eng. 215, 110610.
Zou, C., Qiu, Z., 2021. Preface: new advances in unconventional petroleum sedimentology in China. Acta Sedimentol. Sin. 39, 1–9.
Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J., 2023. Object detection in 20 years: a survey. Proc. IEEE 111, 257–276. https://doi.org/10.1109/JPROC.2023.3238524.