Check for updates

#### **OPEN ACCESS**

EDITED BY Paul Liu, North Carolina State University, United States

REVIEWED BY Han Du, Tsinghua University, China Mingwei Wang, Ministry of Natural Resources, China

\*CORRESPONDENCE Xing Du, ⊠ duxing@fio.org.cn

RECEIVED 19 March 2025 ACCEPTED 19 May 2025 PUBLISHED 18 June 2025

#### CITATION

Wei M, Yu Y, Du X, Song Y, Dong L, Zhou Q, Wang L, Zhang L and Wang Y (2025) Automated detection of submarine pipelines in the Yellow River Estuary: a deep learning approach for side-scan sonar data in dynamic deltaic systems. *Front. Earth Sci.* 13:1596238. doi: 10.3389/feart.2025.1596238

#### COPYRIGHT

© 2025 Wei, Yu, Du, Song, Dong, Zhou, Wang, Zhang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Automated detection of submarine pipelines in the Yellow River Estuary: a deep learning approach for side-scan sonar data in dynamic deltaic systems

Min Wei<sup>1,2</sup>, Yongqing Yu<sup>2</sup>, Xing Du<sup>3</sup>\*, Yupeng Song<sup>3</sup>, Lifeng Dong<sup>3</sup>, Qikun Zhou<sup>3</sup>, Linfeng Wang<sup>2</sup>, Longying Zhang<sup>2</sup> and Yamei Wang<sup>2</sup>

<sup>1</sup>College of Marine Geosciences, Ocean University of China, Qingdao, China, <sup>2</sup>Marine Oil Production Plant, Shengli Oilfield Company, SINOPEC, Dongying, China, <sup>3</sup>First Institute of Oceanography, Ministry of Natural Resources of the People's Republic of China, Qingdao, China

The integrity of submarine pipelines and cables is crucial for safeguarding marine oil, gas, and information transmission, as well as ecological security. Employing automated identification of side-scan sonar (SSS) images can enhance marine geophysical survey efficiency, enabling high-frequency assessment of seabed anthropogenic footprints. However, there is a notable gap in research regarding the comparative performance of different models and the impact of data expansion. This study presents an in-depth comparison of various convolutional neural network (CNN) models-specifically, AlexNet, GoogleNet, and VGG-16-focusing on their prediction accuracy and computational efficiency in analyzing SSS datasets. Our findings reveal that GoogleNet outperforms the others, offering superior prediction accuracy with balanced computational demands. While AlexNet is less accurate, it is beneficial for scenarios with limited computational resources. Conversely, VGG-16 shows comparatively weaker performance, making it less suitable for SSS image analysis. Notably, data expansion significantly influences model accuracy, although its impact varies across different models. This research contributes critical insights into model selection for marine geological applications, demonstrating the potential of intelligent interpretation systems in modern marine geology.

KEYWORDS

marine geophysical monitoring, seabed anthropogenic features, intelligent earth observation, sonar image interpretation, coastal zone management

## **1** Introduction

Marine infrastructure, particularly submarine pipelines and cables, is essential for exploiting oceanic oil and natural gas resources, playing a critical role in maintaining economic and ecological stability. These anthropogenic structures also interact dynamically with marine sedimentary systems, potentially altering local seabed geomorphology through scouring effects and sediment redistribution processes. However, these structures

are vulnerable to damage from sediment erosion and movement, potentially causing significant economic and environmental harm. Such pipeline failures may trigger localized geohazards, including seabed subsidence and slope instability, underscoring the importance of high-resolution monitoring in coastal geological surveys. The study focuses on the Yellow River subaqueous delta, an area known for high sediment discharge and rapid morphological change. Wave action significantly influences seabed dynamics, driving erosion and redistribution processes (Liu et al., 2020). These conditions foster seabed instability, including documented sediment failures linked to liquefaction (Zhang et al., 2023) and storm-wave-induced deformation (Wang et al., 2018). Geohazards such as submarine landslides, for which susceptibility modeling exists (Meng et al., 2024), present risks to subsea infrastructure. Notably, seabed slides can exert significant lateral forces on buried pipelines (Guo et al., 2023). Therefore, reliable monitoring of pipeline integrity in this dynamic environment is essential but challenging using traditional methods. Thus, accurate detection of submarine pipelines is crucial. Traditional detection methods primarily use side-scan sonar imaging, which, while producing high-resolution seabed images, requires labor-intensive and errorprone manual interpretation. This bottleneck limits the temporal resolution of geomorphological change detection in dynamic marine environments. This highlights the need for more automated, efficient approaches.

In the field of earth sciences, Artificial Intelligence (AI) has made breakthrough progress, covering aspects such as remote sensing (Ji et al., 2020; Pouyan et al., 2021; Xu et al., 2022; Casagli et al., 2023; Li et al., 2024; Kamal Basha and Nambiar, 2025; Yang et al., 2025), prediction of geological disasters (Choubin et al., 2019; Mousavi et al., 2020; Stanley et al., 2020; Jones et al., 2021; Ma and Mei, 2021; Zennaro et al., 2021), exploration (Fuentes et al., 2020; Fan et al., 2022; Jin et al., 2022; Liu et al., 2024), and energy development (Kim and Ji, 2022; Li et al., 2022). Recent advances in intelligent earth observation have demonstrated the potential of deep learning in quantifying seabed sediment transport patterns and mapping submarine anthropogenic footprints. However, the applicability and effectiveness of these technologies in the specific field of Pipeline or Cable (POC) detection have not been thoroughly explored, particularly in the context of integrated coastal zone management and marine geological risk assessment within dynamic settings like the Yellow River Delta. In particular, Convolutional Neural Networks (CNN) have shown great potential in processing underwater data, suggesting the possibility of using CNNs to address complex POC detection tasks (Gašparović et al., 2022). Although preliminary applications of CNNs in underwater data processing have covered areas like fish identification and seabed mapping (Jin et al., 2019; Huo et al., 2020; Ge et al., 2021), these domains possess much lower data complexity compared to POC detection. Notably, the discrimination of linear anthropogenic features from natural geomorphic structures remains a key challenge in marine geophysical image interpretation. With technological advancements, researchers are beginning to apply CNNs to more complex scenarios such as detection of sunken ships (Du et al., 2023b), real-time processing of side-scan sonar data (Yan et al., 2020; Li et al., 2023), and developing novel SSS image recognition models like U-Net (Dong et al., 2022) and VIT (Sun et al., 2022). While initial studies have attempted POC recognition by establishing single models (Du et al., 2023a), comprehensive comparisons across multiple models and exploration of model differences remain insufficiently explored.

Despite these developments, comprehensive model comparisons and explorations of model differences in POC detection remain limited. Deep learning has shown promise in predicting sidescan sonar images, but challenges in data acquisition and the limitations of existing datasets are pressing issues. Research typically focuses on algorithm comparison using single public datasets, while demonstrating CNN potential in POC detection (Du et al., 2023a), do not fully tackle fundamental dataset challenges. Acquiring sufficient marine data is difficult, and the narrow applicability of existing datasets restricts their utility across different regions. Hence, investigating effective data expansion methods and model performance under these conditions is vital for developing optimal predictive models from limited data.

Our study builds on existing work by employing classic CNN models, including AlexNet, GoogleNet, and VGG-16, for seabed pipeline SSS image recognition within the Yellow River Delta. We extend previous engineering-focused analyses by incorporating geophysical interpretability metrics, evaluating model performance against known sediment disturbance patterns. We aim to evaluate the effectiveness of these CNN methods by comparing predictive accuracy, accuracy variation post data expansion, and computational efficiency. This research not only advances AI application in marine engineering geology but also establishes a benchmark for intelligent interpretation of anthropogenic features in geophysical surveys. Through this comprehensive model comparison, we aspire to contribute new perspectives and methodologies to marine geoscientific monitoring and infrastructure impact assessment.

## 2 Data and method

## 2.1 Data description

The study focuses on SSS datasets acquired from the Yellow River Estuary, a typical high-sediment-load coastal environment characterized by rapid bedform migration and complex hydrodynamic conditions. The Yellow River Estuary is a sedimentdominated delta with highly dynamic geomorphology, where submarine pipelines are vulnerable to burial or exposure due to rapid sediment transport. We used the Marine-PULSE dataset (Du et al., 2023a) for model training. This dataset assembled using sophisticated instruments like the EdgeTech4200FS, Benthos SIS-1624, EdgeTech4200 MP, Klein-2000, and Klein-3000, provides a comprehensive view of underwater engineering structures. It introduces a classification system encompassing four distinct categories of objects found in marine geology. The dataset was enhanced with seabed surface imagery for a more comprehensive scope. Named Marine-PULSE, it contains 323 images of pipelines or cables (POCs), 134 of underwater residual mounds (URMs), 180 of the seabed surface (SS), and 82 of engineering platforms (EPs), aptly capturing the variety of data that side-scan sonar technology can uncover in marine environments.

Figure 1 presents a sample from the Marine-PULSE dataset, exhibiting the varied morphologies detected in SSS images. The



FIGURE 1

Samples from the Marine-PULSE dataset (Du et al., 2023a). Samples in rows (a–d) are pipelines or cables, underwater residual mounds, seabed surface, and engineering platforms legs, respectively.

assortment stems from several elements such as the object's inherent characteristics, the sonar's angle and distance to the object, the type of instrument, its settings, and the prevailing marine conditions.

In SSS imagery, submarine pipelines or cables (POCs) typically manifest as prominent linear patterns, though determining their diameters can be intricate. Underwater residual mounds, which are formations from sediment with greater strength than the surrounding matrix, exhibit unique morphological structures due to erosion. The seabed surface depicts variations from flat to rugged terrains, adding to the morphological diversity observed in SSS images. Engineering platforms, characterized by multiple piles, disrupt acoustic signals, creating an absence of linear signals in a band pattern, which adds to the morphological variety in SSS images.

The principal objective of this investigation was the automated detection of submarine pipelines or cables in SSS imagery. Hence,

we categorized the dataset into "POC" for images of pipelines or cables, and "Non-POC," which encompasses the remaining image types from the collection.

## 2.2 Applied CNN models

In this study, we employ three types of Convolutional Neural Networks (CNN) (Rumelhart et al., 1986) to model and recognize SSS image types: AlexNet, Vgg-16, and GoogleNet. These classic models were selected as well-established benchmarks representing varied architectural complexities, providing a suitable basis for foundational comparative analysis on SSS data. Investigation of other architectures like U-Net or ViT is considered beyond the scope of this initial comparison and is left for future work. Here, we provide a brief introduction to the three CNN models.

## 2.2.1 AlexNet

In 2012, Alex Krizhevsky introduced Krizhevsky. (2012), a groundbreaking deep learning architecture that clinched the top spot in that year's ILSVRC challenge. This network boasted a complexity unprecedented for its time, consisting of five convolutional layers followed by three fully connected layers. It processed input images resized to  $227 \times 227$  pixels, learning to classify them through its vast network comprising 630 million connections, 60 million parameters, and 650,000 neurons. AlexNet distinguished itself by being the first to apply the Rectified Linear Unit (ReLU) activation function after convolutional and fully-connected layers, incorporating Dropout regularization (typically 0.5 probability) before the first two fully connected layers to curb overfitting, and utilizing CUDA to hasten training - collectively enhancing its processing power and precision, thereby accelerating deep learning development.

AlexNet's revolutionary entry into deep learning was marked by its intricacy and depth, which allowed it to process highresolution images effectively. The convolutional layers utilized filters of varying sizes  $(11 \times 11)$  in the first layer,  $5 \times 5$  in the second, and  $3 \times 3$  in subsequent layers), followed by overlapping maxpooling layers after the first, second, and fifth convolutional layers. One of the notable features of AlexNet is its introduction of the ReLU activation function, which helped it to solve the vanishing gradient problem common in deep networks, leading to faster convergence. The network also mitigated overfitting through the use of Dropout, a technique that randomly deactivates neurons during training to prevent complex co-adaptations on training data. Additionally, the use of CUDA for accelerating training processes allowed AlexNet to train much larger networks in a reasonable time. However, its large number of parameters necessitated significant computational resources, and the network's size made it prone to overfitting without the proper regularization techniques like Dropout.

## 2.3 VGG

The VGG model, developed by the Visual Geometry Group from Oxford, emerged during the ILSVRC in 2014 (Simonyan and Zisserman, 2015). It underscored the network depth's influence on performance. Renowned for its depth, the model, particularly the VGG-16 variant used in this study, became a benchmark in CNN designs, featuring 13 convolutional layers complemented by three fully connected layers. Despite its large number of parameters (approximately 138 million) making it challenging to train, VGG's depth, achieved by consistently using small  $3 \times 3$  convolutional filters stacked multiple times (two or three layers deep) between pooling operations, enhanced its computational prowess compared to its predecessor AlexNet. Max-pooling layers ( $2 \times 2$  with stride 2) followed blocks of convolutional layers.

VGG, known for its simplicity and depth, leverages a homogeneous architecture that uses only  $3 \times 3$  convolutional layers throughout the network, which helps in reducing the complexity of hyperparameter tuning. The model is particularly noted for showing that depth is a crucial component for the successful training of neural networks. VGG's uniform structure also makes it easy to understand and implement, which has contributed to

its widespread use in the computer vision community. However, the model is quite heavy, with a significant number of parameters, resulting in high memory usage and computational costs. Training VGG from scratch requires extensive computational resources, and it is also relatively slow in inference compared to more modern architectures.

## 2.4 GoogleNet

Google's Christian Szegedy introduced GoogleNet (Szegedy et al., 2014), which incorporated the first inception architecture. This innovation streamlined deep neural networks by altering the traditional sequential CNN layout into a parallel internal connection through the inception module, as depicted in Figure 2. Data traverses four distinct pathways concurrently within each module: branches with  $1 \times 1$ ,  $3 \times 3$ , and  $5 \times 5$  convolutional filters, plus a  $3 \times 3$  max-pooling branch. The outputs of these branches are then concatenated depth-wise. The inception's hallmark is its dual benefit: multi-scale convolution captures diverse-scale features enhancing classification accuracy, while strategic use of  $1 \times 1$ convolutions before the  $3 \times 3$  and  $5 \times 5$  convolutions and after the pooling layer serves as dimensionality reduction, cutting down on computational load. With 22 layers (including pooling layers when counting layers with parameters), GoogleNet surpasses AlexNet's and VGGNet's depths but achieves superior accuracy with only about 6.8 million parameters-significantly fewer than both AlexNet and VGG-16.

GoogleNet, with its inception architecture, represents a shift towards more efficient designs in network architectures. Its inception modules perform convolutions at multiple scales, allowing the network to capture complex features at various resolutions. This multi-scale processing capability is one of its standout features, providing significant improvements in accuracy. The  $1 \times 1$  convolutions used for dimensionality reduction not only preserve important features but also reduce the computational burden, making GoogleNet efficient and faster in terms of computation. The network also reduces the number of parameters dramatically compared to its predecessors, which helps prevent overfitting to some extent.

## 2.5 Utilization of established CNN models

The overall process is illustrated in Figure 3, where we utilize AlexNet, GoogleNet, and VGG-16 to construct recognition models for SSS images of underwater pipelines, aimed at investigating the accuracy of different models in identifying subsea pipelines. The modeling process is broadly divided into several stages: data preprocessing, data expansion, model establishment, and model evaluation.

### 2.5.1 Data preprocessing

Data preprocessing primarily consists of two parts: dataset partitioning and normalization. For the Marine-PULSE dataset, which comprises 719 images, we randomly divided it into a training set with 431 images, and validation and test sets, each containing 144 images, following a 60%:20%:20% split ratio.





Normalization was applied to equalize the pixel intensity levels across the three channels of the SSS imagery, scaling the values to fit within a [-1, 1] interval. This step ensures uniform data distribution, which is crucial for consistent training performance, as large variations could lead to less than optimal neural network training results.

Following this initial data handling, the SSS images were primed for the deep learning phase. The preparation set the stage for in-depth learning and interpretation of the pipeline or cable (POC) visuals contained in the data collection. Once labeled and normalized, the images were ready to enter the convolutional neural network (CNN) training process, laying the groundwork for the detailed evaluation of subaquatic infrastructure.

### 2.5.2 Data expansion

Alongside the preprocessing measures previously described, we expanded the original training dataset from 431 to 1500 images through data expansion methods such as rotations, flips, and contrast adjustments. Unlike traditional data expansion, where transformations are applied randomly during training without altering the dataset size, this approach involved saving the altered images, thus physically enlarging the training set. This technique, validated by prior research (Du et al., 2023a), has been shown to enhance model accuracy.

In each training cycle, the images were systematically modified using these data expansion strategies, introducing a broader spectrum of variation and complexity to the dataset. These controlled modifications bolstered the dataset, aiding the neural networks in acquiring more generalized and invariant feature recognition capabilities. By incorporating altered perspectives and varied orientations of the input data, the networks developed resilience to overfitting and an improved ability to decipher complex real-world scenarios.

This data expansion not only bolstered the volume and heterogeneity of the training pool but also fortified the model's proficiency in navigating actual environmental fluctuations. Consequently, these strategies were instrumental in refining the model's feature discernment, significantly elevating the precision and dependability of the resultant model.

## 2.5.3 Model establishment

Upon dataset preparation, the training commenced using established CNN architectures. This study utilized AlexNet, VGG-16, and GoogleNet for model training. Additionally, we employed transfer learning, leveraging pre-trained parameters from the ImageNet dataset to refine our models. Both the original dataset of 431 images and the augmented set of 1500 images were used to train the AlexNet, GoogleNet, and VGG-16 models, assessing the impact of dataset size on model performance. The models' predictive accuracies were then verified using a separate validation set containing 144 images. The overarching goal was to compare the performance of the various CNN architectures, with a specific focus on the influence of dataset size and model depth on the accuracy of subsea pipeline detection.

#### 2.5.4 Evaluation methods

To evaluate the effectiveness of the implemented CNN models (AlexNet, GoogleNet, and VGG-16) in the autonomous identification of underwater pipeline entities within SSS imagery, four evaluative metrics were utilized: accuracy (Equation 1), precision (Equation 2), recall (Equation 3), and the F1 score (Equation 4). The overall accuracy represents the model's correctness by reflecting the ratio of accurately identified items. Precision is the measure of the model's exactness in classifying positive instances, calculated by the fraction of true positives among all positive predictions. Recall, on the other hand, gauges the model's efficiency in identifying all positive instances, expressed as the proportion of true positives to the entire count of actual positives. The F1 score synthesizes precision and recall into a singular metric, offering a harmonious evaluation that considers both elements. Together, these parameters furnish a thorough appraisal of the model's performance in detecting underwater pipeline entities in SSS imagery. The computation of these evaluative metrics is encapsulated in the subsequent formulas (Table 1):

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$
(1)

$$Precision = TP / (TP + FP)$$
(2)

$$Recall = TP / (TP + FN)$$
(3)

$$F1 \text{ score} = 2 \times (Precision \times Recall) / (Precision + Recall)$$
(4)

Here, a true positive (TP) denotes a POC correctly identified as such, while a true negative (TN) corresponds to a non-POC accurately recognized. Conversely, a false positive (FP) indicates a non-POC erroneously classified as POC, and a false negative (FN) represents a POC incorrectly labeled as non-POC.

In addition to accuracy, the total training and prediction time for each model on an identical dataset were tallied, along with the number of model parameters, to provide a holistic evaluation of the computational efficiency and resource expenditure across different CNN architectures.

#### 2.5.5 Experimental settings

This research is structured to meticulously evaluate the performance of convolutional neural networks (CNNs) in the context of subsea pipeline identification from SSS images. The core aim is to discern the efficacy of CNN models under varying training conditions and to understand the influence of data volume and pre-training on model accuracy and computational efficiency.

In the first set of experiments (Table 2), we embark on a comparative analysis using the original training dataset comprising 413 images. The objective is to train the AlexNet, GoogleNet, and VGG-16 models from scratch, utilizing their distinctive architectural strengths to identify subsea pipelines. By comparing the results, we aim to pinpoint which model most accurately classifies the SSS images. This experiment will shed light on the inherent capabilities of each CNN structure when dealing with unamplified datasets.

The second experiment aims to measure the impact of data expansion on model performance. By training the same CNN models on an augmented dataset expanded to 1500 images, we examine whether the increased data volume translates into higher prediction accuracy. The experiment will compare each model's performance with the original and expanded datasets, providing insight into the value of data expansion in deep learning for underwater image analysis.

Lastly, the third experiment focuses on evaluating the computational aspects of the CNN models. By analyzing the training and prediction times of models trained on the expanded dataset, we assess which architecture delivers not just the highest accuracy but also operates with optimal computational efficiency. This approach ensures that the models are not only powerful in terms of performance but also practical for real-world application where resources and time are often limited.

#### 2.5.6 Implementation

For the training process, we utilized the Adam optimizer with an initial learning rate of 0.0001 and a batch size of 32 for all three CNN models to ensure a consistent comparison basis. The cutting-edge models were developed leveraging PyTorch, an esteemed deep learning platform known for its comprehensive suite of libraries and tools that facilitate the creation of neural networks. These models were computed on a high-performance Dell 3660 workstation, which is equipped with an i9-12900k CPU, 128 GB of RAM, and an RTX 4090 GPU, ensuring swift and efficient processing capabilities.

TABLE 1 Confusion matrix for binary classification of model POCs and non-POCs.

Predict label/True label	Positive sample (POC)	Negative sample (non-Poc)	
Positive Sample (POC)	ТР	FN	
Negative Sample (non-POC)	FP	$TN^{a}$	

<sup>a</sup>Referencing Table 1, a confusion matrix aids in binary classification, designating POC as positive samples and non-POC as negative ones.

#### TABLE 2 Experimental design overview.

NO.	Train	Models	Validation <sup>a</sup>	Test
1	Original train dataset (413)	AlexNet	Validation dataset (144)	Test dataset (144)
2	Original train dataset (413)	GoogleNet	Validation dataset (144)	Test dataset (144)
3	Original train dataset (413)	Vgg-16	Validation dataset (144)	Test dataset (144)
4	Expanded train dataset (1500)	AlexNet	Validation dataset (144)	Test dataset (144)
5	Expanded train dataset (1500)	GoogleNet	Validation dataset (144)	Test dataset (144)
6	Expanded train dataset (1500)	Vgg-16	Validation dataset (144)	Test dataset (144)

<sup>a</sup>The number in parentheses indicates the number of images within the specified dataset (Validation dataset or Test dataset).

## **3** Results and discussions

# 3.1 Accuracy comparison of different models

In our comprehensive comparison of CNN models-AlexNet, GoogleNet, and VGG-16-we analyzed their performance over 100 epochs using train and validation datasets. GoogleNet emerged as the superior model in several key aspects. As illustrated in Figure 4a, GoogleNet's accuracy on the training set rapidly ascended from approximately 3.5%, and on the validation set, it dramatically increased from near 0% to over 86%. This showcases its remarkable learning efficiency and generalization capabilities. In contrast, AlexNet exhibited a steady yet slower increase in accuracy for both training and validation datasets, plateauing around 75% on the validation set-significantly lower than GoogleNet's 86%. VGG-16, starting with an initial training set accuracy of around 50%, showed a gradual increase to about 85%. However, its oscillating accuracy on the validation set suggested a potential overfitting issue. The loss trends across epochs for each model, depicted in Figure 4b, reveal that GoogleNet experienced a swift decline in loss for both training and validation sets, with a notably lower loss on the validation set, underscoring its superior generalization. In comparison, AlexNet's training loss decreased steadily but at a slower rate than GoogleNet. VGG-16's loss trajectory was more erratic and decreased at a slower pace compared to the other models.

Further analysis of the models' predictive performance on the test set, as detailed in Figure 5 and Table 3, reaffirmed GoogleNet's dominance. GoogleNet achieved an impressive accuracy of 88.19%, significantly outperforming AlexNet's 81.25% and VGG-16's 79.17%. In terms of precision, GoogleNet led with 85.07%, exceeding AlexNet's 76.81% and VGG-16's 72.97%. This suggests

that GoogleNet had a higher proportion of true positive predictions. Regarding recall, GoogleNet attained 89.06%, surpassing AlexNet's 82.81% and VGG-16's 84.35%, indicating its higher efficiency in identifying true positives among all positive samples. The F1 score, a metric combining precision and recall, saw GoogleNet scoring 87.02%, eclipsing AlexNet's 79.7% and VGG-16's 78.26%.

These results conclusively demonstrate that GoogleNet excels in accurately categorizing samples (accuracy), minimizing false positives (precision), effectively identifying true positives (recall), and achieving a balance between precision and recall (F1 score). Conversely, although AlexNet displayed a commendable recall rate, it lagged behind GoogleNet in other metrics. VGG-16, comparable to AlexNet in recall, fell short in both accuracy and precision. Therefore, in overall performance, GoogleNet exhibits strong potential and practical applicability for SSS image classification tasks.

# 3.2 Impact of data expansion on model performance

To quantitatively assess the impact of data expansion, we compared the performance of AlexNet, GoogleNet, and VGG-16 trained on the original dataset (431 images) *versus* the expanded dataset (1500 images). The expanded dataset was generated using techniques such as image flipping, rotation, and contrast adjustment. Performance metrics, including accuracy, precision, recall, and F1 score, were evaluated on the independent test set (144 images) for models trained with and without data expansion. Given the inherent challenges of sonar images, such as high noise levels and occasionally blurred target boundaries, data expansion aims to enhance the models' ability to discern complex features.



#### FIGURE 4

Comparison of prediction accuracies of AlexNet, GoogleNet, and VGG-16. (a) is the accuracy of train and validation dataset for the AlexNet, GoogleNet, and VGG-16 models with respect to the number of epochs; (b) is the loss of train and validation dataset for the three CNN models across epochs.



Figure 6 illustrates the positive impact of data expansion on the training, validation, and test accuracies of AlexNet, GoogleNet, and VGG-16. With the expanded dataset, the AlexNet model showed an increase of 1.45% in training accuracy, 3.88% in validation accuracy, and 5.13% in test accuracy. GoogleNet experienced

smaller yet notable improvements of 0.23%, 2.21%, and 3.95% in these respective metrics. VGG-16 demonstrated the most substantial gains in training (12.53%) and validation accuracies (6.61%), but interestingly, a 3.51% decrease in test accuracy was observed.

Model/Metric	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
AlexNet	81.25	76.81	82.81	79.7
GoogleNet 88.19		85.07	89.06	87.02
Vgg-16	79.17	72.97	84.35	78.26

TABLE 3 Statistical analysis of test dataset predictions by different CNN models.



TABLE 4 Impa	et of data	expansion	on model	performance	(test set metrics).
--------------	------------	-----------	----------	-------------	---------------------

CNN models	Train accuracy difference <sup>a</sup> (%)	Validate accuracy difference (%)	Test accuracy difference (%)	Train increment <sup>b</sup> (%)	Validate increment (%)	Test increment (%)
AlexNet	1.42	3.47	4.17	1.45	3.88	5.13
GoogleNet	0.23	2.08	3.48	0.23	2.21	3.95
Vgg-16	11.14	5.56	-2.78	12.53	6.61	-3.51

<sup>a</sup>Training accuracy difference is the difference between the best accuracy before and after data expand method.

<sup>b</sup>Training increment is the ratio of training accuracy difference to the training accuracy of the original CNN, model.

The unique characteristics of SSS images significantly influenced the enhanced performance of these models. The diverse disturbances and artifacts present in these images necessitate robust feature representation learning by the models, thereby improving their predictive capabilities in various real-world scenarios. According to Table 4, both AlexNet and GoogleNet exhibited increases in accuracy, particularly on the test dataset, post data expansion, indicating improved adaptability to new data. However, the decline in test accuracy for VGG-16 raises concerns, suggesting that data expansion might need to be complemented with other strategies to prevent overfitting.

In summary, while data expansion effectively increased accuracy for AlexNet and GoogleNet in identifying seabed pipeline SSS images, it resulted in decreased performance for VGG-16 due to overfitting issues. This underscores that while data expansion can positively impact the generalization abilities of deep learning models, its effectiveness is contingent upon specific model architectures and dataset characteristics. Beside, evaluating only one level of data expansion (from 431 to 1500 images) provides initial insights but may not fully capture the complex relationship between dataset size and model performance for this specific task, particularly given the varied responses observed across different architectures.

## 3.3 Calculation efficiency and difficulty

Computational efficiency and complexity are pivotal in determining the practical utility of deep learning models, encompassing aspects like computation time, number of layers, parameters, and floating-point operations (FLOPs). We assessed these elements for AlexNet, GoogleNet, and VGG-16, using the seabed pipeline side-scan sonar (SSS) image dataset.

CNN models	Layers	Number of parameters (M)	FLOPs <sup>a</sup> (M)	Test dataset Accuracy (%)	Calculation time (s)
AlexNet	11	60	727	85.42	872
GoogleNet	87	6.8	2000	91.67	1071
VGG-16	16	138	16,000	76.39	2564

TABLE 5 Comparison of the calculation efficiency of different CNN models.

<sup>a</sup>FLOPs (Floating point Operations) represent a number of floating point calculations, reflecting the difficulty of the calculation.

Our analysis focused on key metrics of computational difficulty (Table 5): layer count, parameter number, and FLOPs. AlexNet, with 11 layers and 60 million parameters, requires 727 million FLOPs and achieved 85.42% accuracy on the test dataset within 872 s. GoogleNet, comprising 87 layers, is more efficient in terms of parameter count (6.8 million) and FLOPs (2 billion), achieving 91.67% accuracy in 1071 s. VGG-16, with 16 layers and the highest parameter count (138 million), demands a significant 16 billion FLOPs and reached only 76.39% accuracy in 2564 s.

GoogleNet's Inception architecture, despite its higher layer count, effectively reduces the number of parameters and enhances information processing complexity. This design not only minimizes memory and computational resource demands but also reduces training time, making GoogleNet an efficient performer in processing complex seabed pipeline SSS image datasets. Although GoogleNet's computation time slightly exceeds AlexNet's, its substantial accuracy improvement justifies this trade-off in practical applications. VGG-16, while theoretically capturing more detailed features, is constrained by its high parameter and FLOP demands, leading to prolonged computation times and limited practicality in resource-constrained scenarios.

In summary, GoogleNet's structural optimization and computational efficiency make it the preferred model for seabed pipeline SSS image dataset processing. It exemplifies balancing a high number of layers with controlled parameters and computational resource use, offering insights for designing deep learning models. Task-specific requirements and resource availability should guide model selection, aiming for an optimal balance between performance and efficiency.

# 4 Outlook

In this study, we employed three classic CNN models-AlexNet, GoogleNet, and VGG-16-for seabed pipeline image recognition using side-scan sonar, enhanced by data expansion techniques. This approach provided accurate model recognition in complex marine environments. We utilized a multi-model analysis approach, focusing on GoogleNet, to deepen our understanding of submarine pipeline identification challenges (Du et al., 2023a). The methodology demonstrates particular promise for monitoring sediment-pipeline interaction processes critical to coastal geohazard assessment, including scour development and pipeline free-span evolution.

However, our research, while leveraging established CNN architectures, has not fully exploited the unique attributes of sidescan sonar images, such as the characteristic acoustic shadows, linear target geometries, and influence of seabed texture. Future endeavors should aim to develop specialized network structures tailored to SSS image characteristics. For instance, incorporating attention mechanisms specifically designed to enhance linear feature detection or developing custom convolutional kernels that are sensitive to typical SSS textural patterns and acoustic scattering effects could improve recognition accuracy. Furthermore, integrating geophysical domain knowledge more directly, perhaps by designing multi-modal architectures that fuse SSS imagery with bathymetric data or sediment classifications, could lead to more robust "geology-aware" models. This could be particularly valuable for distinguishing pipelines from similar-looking natural features (e.g., bedrock outcrops, sand ridges) and for quantitatively analyzing sediment-pipeline interactions, thereby improving predictive capabilities for geohazard assessment in dynamic deltaic systems and potentially reducing computational resource consumption.

This work not only progresses marine engineering geology but also significantly advances deep learning model development and optimization for complex image recognition tasks in challenging environments like the ocean floor. The established framework provides a novel pathway for integrating anthropogenic infrastructure monitoring with coastal zone geological surveys, particularly in assessing human-induced modifications to submarine geomorphology.

# **5** Conclusion

This study establishes a deep learning framework for submarine pipeline recognition in side-scan sonar (SSS) images, with critical implications for coastal geological monitoring and infrastructure risk assessment. Our comparative analysis of AlexNet, GoogleNet, and VGG-16 models under varying training regimes yields four principal findings:

- (1) All three CNN models demonstrated the ability to accurately predict SSS seabed pipeline images, with GoogleNet showing the most outstanding performance in terms of accuracy and learning efficiency. This highlights CNN's capability in resolving pipeline signatures within complex submarine geological settings characterized by sediment interference and bedform variations.
- (2) Data expansion techniques significantly improved the predictive accuracy of the models, though the extent of improvement varied across different models. AlexNet and

GoogleNet showed enhanced performance, while the accuracy of VGG-16 decreased. This emphasizes the importance of considering model-specific characteristics and data compatibility when applying data expansion.

- (3) GoogleNet is the optimal choice, offering a well-balanced mix of high accuracy and computational efficiency. AlexNet, while less accuracy, is suitable for achieving acceptable accuracy with minimal computational demands. Conversely, VGG-16's performance in SSS image recognition is suboptimal, leading us to recommend against its use for this specific task.
- (4) While GoogleNet holds an advantage in overall performance, the performance of each model may vary in specific marine environments. Future work should explore customized model adjustments tailored to specific seabed characteristics or introduce more marine-environment-related transformations in data expansion strategies, to simulate the variable conditions that might be encountered in real-world applications.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

# Author contributions

MW: Data curation, Methodology, Validation, Writing – original draft, Software. YY: Validation, Writing – original draft. XD: Software, Conceptualization, Investigation, Writing – original draft, Funding acquisition, Validation, Writing – review and editing, Data curation, Project administration. YS: Supervision, Writing – review and editing, Project administration. LD: Validation, Writing – review and editing. QZ: Formal Analysis, Writing – original draft, Writing – review and editing. LW: Writing – review and editing, Software. LZ: Methodology, Writing – original draft. YW: Software, Writing – original draft.

# References

Casagli, N., Intrieri, E., Tofani, V., Gigli, G., and Raspini, F. (2023). Landslide detection, monitoring and prediction with remote-sensing techniques. *Nat. Rev. Earth Environ.* 4, 51–64. doi:10.1038/s43017-022-00373-x

Choubin, B., Borji, M., Mosavi, A., Sajedi-Hosseini, F., Singh, V. P., and Shamshirband, S. (2019). Snow avalanche hazard prediction using machine learning methods. *J. Hydrology* 577, 123929. doi:10.1016/j.jhydrol.2019.123929

Dong, Z., An, S., Zhang, J., Yu, J., Li, J., and Xu, D. (2022). L-unet: a landslide extraction model using multi-scale feature fusion and attention mechanism. *Remote Sens.* 14, 2552. doi:10.3390/rs14112552

Du, X., Sun, Y., Song, Y., Dong, L., and Zhao, X. (2023a). Revealing the potential of deep learning for detecting submarine pipelines in side-scan sonar images: an investigation of pre-training datasets. *Remote Sens.* 15, 4873. doi:10.3390/rs15194873

Du, X., Sun, Y., Song, Y., Sun, H., and Yang, L. (2023b). A comparative study of different CNN models and transfer learning effect for underwater object classification in side-scan sonar images. *Remote Sens.* 15, 593. doi:10.3390/rs15030593

# Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by the Foundation item: The National Natural Science Foundation of China under contract NO. 42102326; the Basic Scientific Fund for National Public Research Institutes of China under contract NO. 2022Q05.

# **Acknowledgments**

The authors would like to thank to the developers who proposed the Pytorch deep learning package (https://pytorch.org/), which support the CNN modeling in this paper.

# **Conflict of interest**

Authors MW, YY, LW, LZ, and YW were employed by Marine Oil Production Plant, Shengli Oilfield Company, SINOPEC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# **Generative AI statement**

The author(s) declare that no Generative AI was used in the creation of this manuscript.

# Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Fan, T., Li, P., Qi, Z., Zhao, Z., Fang, X., Yan, B., et al. (2022). Borehole transient electromagnetic stereo imaging method based on horizontal component anomaly feature clustering. *J. Appl. Geophys.* 197, 104537. doi:10.1016/j.jappgeo.2022.104537

Fuentes, I., Padarian, J., Iwanaga, T., and Willem Vervoort, R. (2020). 3D lithological mapping of borehole descriptions using word embeddings. *Comput. and Geosciences* 141, 104516. doi:10.1016/j.cageo.2020.104516

Gašparović, B., Lerga, J., Mauša, G., and Ivašić-Kos, M. (2022). Deep learning approach for objects detection in underwater pipeline images. *Appl. Artif. Intell.* 36, 2146853. doi:10.1080/08839514.2022.2146853

Ge, Q., Ruan, F., Qiao, B., Zhang, Q., Zuo, X., and Dang, L. (2021). Side-scan sonar image classification based on style transfer and pre-trained convolutional neural networks. *Electronics* 10, 1823. doi:10.3390/electronics10151823

Guo, X., Liu, X., Li, M., and Lu, Y. (2023). Lateral force on buried pipelines caused by seabed slides using a CFD method with a shear interface weakening model. *Ocean. Eng.* 280, 114663. doi:10.1016/j.oceaneng.2023.114663

Huo, G., Wu, Z., and Li, J. (2020). Underwater object classification in sidescan sonar images using deep transfer learning and semisynthetic training data. *IEEE Access* 8, 47407–47418. doi:10.1109/ACCESS.2020.2978880

Ji, S., Dawen, Y., Shen, C., Li, W., and Xu, Q. (2020). Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. *Landslides* 17, 1337–1352. doi:10.1007/s10346-020-01353-2

Jin, C., Wang, K., Han, T., Lu, Y., Liu, A., and Liu, D. (2022). Segmentation of ore and waste rocks in borehole images using the multi-module densely connected U-net. *Comput. and Geosciences* 159, 105018. doi:10.1016/j.cageo.2021. 105018

Jin, L., Liang, H., and Yang, C. (2019). Accurate underwater ATR in forwardlooking sonar imagery using deep convolutional neural networks. *IEEE Access* 7, 125522–125531. doi:10.1109/ACCESS.2019.2939005

Jones, S., Kasthurba, A. K., Bhagyanathan, A., and Binoy, B. V. (2021). Landslide susceptibility investigation for Idukki district of Kerala using regression analysis and machine learning. *Arab. J. Geosci.* 14, 838. doi:10.1007/s12517-021-07156-6

Kamal Basha, S., and Nambiar, A. (2025). "S3Simulator: a benchmarking side scan sonar simulator dataset for underwater image analysis," in *Pattern recognition*. Editors A. Antonacopoulos, S. Chaudhuri, R. Chellappa, C.-L. Liu, S. Bhattacharya, and U. Pal (Cham: Springer Nature Switzerland), 219–235. doi:10.1007/978-3-031-78444-6\_15

Kim, H.-S., and Ji, Y. (2022). Three-dimensional geotechnical-layer mapping in Seoul using borehole database and deep neural network-based model. *Eng. Geol.* 297, 106489. doi:10.1016/j.enggeo.2021.106489

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "ImageNet classification with deep convolutional neural networks," in *Advances in neural information processing systems* (Red Hook, NY: Curran Associates, Inc.). Available online at: https://proceedings. neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.

Li, L., Li, Y., Wang, H., Yue, C., Gao, P., Wang, Y., et al. (2024). Side-scan sonar image generation under zero and few samples for underwater target detection. *Remote Sens.* 16, 4134. doi:10.3390/rs16224134

Li, L., Li, Y., Yue, C., Xu, G., Wang, H., and Feng, X. (2023). Real-time underwater target detection for AUV using side scan sonar images based on deep learning. *Appl. Ocean Res.* 138, 103630. doi:10.1016/j.apor.2023.103630

Li, Y., Peng, J., Zhang, L., Zhou, J., Huang, C., and Lian, M. (2022). Quantitative evaluation of impact cracks near the borehole based on 2D image analysis and fractal theory. *Geothermics* 100, 102335. doi:10.1016/j.geothermics. 2021.102335

Liu, X., Zhang, H., Zheng, J., Guo, L., Jia, Y., Bian, C., et al. (2020). Critical role of wave-seabed interactions in the extensive erosion of Yellow River estuarine sediments. *Mar. Geol.* 426, 106208. doi:10.1016/j.margeo. 2020.106208

Liu, Y., Du, K., Shan, L., Zhu, L., Jiang, H., Wang, Y., et al. (2024). Segmentation of seabed sediment images based on convolutional neural network. *JMEE* 11, 173–189. doi:10.32908/JMEE.v11.2024082601

Ma, Z., and Mei, G. (2021). Deep learning for geological hazards analysis: data, models, applications, and opportunities. *Earth-Science Rev.* 223, 103858. doi:10.1016/j.earscirev.2021.103858

Meng, X., Liu, X., Wang, Y., Zhang, H., and Guo, X. (2024). Submarine landslide susceptibility assessment integrating frequency ratio with supervised machine learning approach. *Appl. Ocean Res.* 153, 104237. doi:10.1016/j.apor.2024.104237

Mousavi, S. M., Ellsworth, W., Weiqiang, Z., Chuang, L., and Beroza, G. (2020). Earthquake transformer-an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat. Commun.* 11, 3952. doi:10.1038/s41467-020-17591-w

Pouyan, S., Pourghasemi, H. R., Bordbar, M., Rahmanian, S., and Clague, J. J. (2021). A multi-hazard map-based flooding, gully erosion, forest fires, and earthquakes in Iran. *Sci. Rep.* 11, 14889. doi:10.1038/s41598-021-94266-6

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature* 323, 533–536. doi:10.1038/323533a0

Simonyan, K., and Zisserman, A. (2015). Very deep convolutional networks for largescale image recognition. Available online at: http://arxiv.org/abs/1409.1556 (Accessed June 1, 2022).

Stanley, T. A., Kirschbaum, D. B., Sobieszczyk, S., Jasinski, M. F., Borak, J. S., and Slaughter, S. L. (2020). Building a landslide hazard indicator with machine learning and land surface models. *Environ. Model. and Softw.* 129, 104692. doi:10.1016/j.envsoft.2020.104692

Sun, Y., Zheng, H., Zhang, G., Ren, J., Xu, H., and Xu, C. (2022). DP-ViT: a dualpath vision transformer for real-time sonar target detection. *Remote Sens.* 14, 5807. doi:10.3390/rs14225807

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2014). Going deeper with convolutions. Available online at: http://arxiv.org/abs/1409.4842 (Accessed June 1, 2022).

Wang, Z., Jia, Y., Liu, X., Wang, D., Shan, H., Guo, L., et al. (2018). *In situ* observation of storm-wave-induced seabed deformation with a submarine landslide monitoring system. *Bull. Eng. Geol. Environ.* 77, 1091–1102. doi:10.1007/s10064-017-1130-4

Xu, S., Dimasaka, J., Wald, D. J., and Noh, H. Y. (2022). Seismic multi-hazard and impact estimation via causal inference from satellite imagery. *Nat. Commun.* 13, 7793. doi:10.1038/s41467-022-35418-8

Yan, J., Meng, J., and Zhao, J. (2020). Real-time bottom tracking using side scan sonar data through one-dimensional convolutional neural networks. *Remote Sens.* 12, 37. doi:10.3390/rs12010037

Yang, N., Li, G., Wang, S., Wei, Z., Ren, H., Zhang, X., et al. (2025). SS-YOLO: a lightweight deep learning model focused on side-scan sonar target detection. *J. Mar. Sci. Eng.* 13, 66. doi:10.3390/jmse13010066

Zennaro, F., Furlan, E., Simeoni, C., Torresan, S., Aslan, S., Critto, A., et al. (2021). Exploring machine learning potential for climate change risk assessment. *Earth-Science Rev.* 220, 103752. doi:10.1016/j.earscirev.2021.103752

Zhang, H., Lu, Y., Liu, X., Li, X., Wang, Z., Ji, C., et al. (2023). Morphology and origin of liquefaction-related sediment failures on the Yellow River subaqueous delta. *Mar. Petroleum Geol.* 153, 106262. doi:10.1016/j.marpetgeo.2023.106262