

Article

A Lightweight Image-Based Decision Support Model for Marine Cylinder Lubrication Based on CNN-ViT Fusion

Qiuyu Li ¹, Guichen Zhang ^{1,*} and Enrui Zhao ^{1,2} ¹ Merchant Marine College, Shanghai Maritime University, Shanghai 201306, China; liqiuyu@stu.shmtu.edu.cn (Q.L.); zhaoenrui@stu.shmtu.edu.cn (E.Z.)² School of Navigation and Shipping, Shandong Jiaotong University, Weihai 264209, China

* Correspondence: gc Zhang@shmtu.edu.cn

Abstract

Under the context of “Energy Conservation and Emission Reduction,” low-sulfur fuel has become widely adopted in maritime operations, posing significant challenges to cylinder lubrication systems. Traditional oil injection strategies, heavily reliant on manual experience, suffer from instability and high costs. To address this, a lightweight image retrieval model for cylinder lubrication is proposed, leveraging deep learning and computer vision to support oiling decisions based on visual features. The model comprises three components: a backbone network, a feature enhancement module, and a similarity retrieval module. Specifically, EfficientNetB0 serves as the backbone for efficient feature extraction under low computational overhead. MobileViT Blocks are integrated to combine local feature perception of Convolutional Neural Networks (CNNs) with the global modeling capacity of Transformers. To further improve receptive field and multi-scale representation, Receptive Field Blocks (RFB) are introduced between the components. Additionally, the Convolutional Block Attention Module (CBAM) attention mechanism enhances focus on salient regions, improving feature discrimination. A high-quality image dataset was constructed using WINNING’s large bulk carriers under various sea conditions. The experimental results demonstrate that the EfficientNetB0 + RFB + MobileViT + CBAM model achieves excellent performance with minimal computational cost: 99.71% Precision, 99.69% Recall, and 99.70% F1-score—improvements of 11.81%, 15.36%, and 13.62%, respectively, over the baseline EfficientNetB0. With only a 0.3 GFLOP and 8.3 MB increase in model size, the approach balances accuracy and inference efficiency. The model also demonstrates good robustness and application stability in real-world ship testing, with potential for further adoption in the field of intelligent ship maintenance.

Keywords: marine main engine; cylinder lubrication; lightweight; image retrieval; CNN; vision transformer



Academic Editor: Weicheng Cui

Received: 15 September 2025

Revised: 2 October 2025

Accepted: 11 October 2025

Published: 13 October 2025

Citation: Li, Q.; Zhang, G.; Zhao, E. A Lightweight Image-Based Decision Support Model for Marine Cylinder Lubrication Based on CNN-ViT Fusion. *J. Mar. Sci. Eng.* **2025**, *13*, 1956. <https://doi.org/10.3390/jmse13101956>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Amid growing global emphasis on energy conservation and emission reduction, the maritime industry faces mounting environmental challenges. As the primary regulatory body for international shipping, the International Maritime Organization (IMO) has introduced a series of regulations to promote sustainable transformation. In response, major shipping companies have increasingly adopted Low-Sulfur Fuel Oil (LSFO) [1,2].

Although low-sulfur fuel oil (LSFO) can significantly reduce sulfur oxide emissions, its complex composition, poor stability, and batch-to-batch variability also introduce considerable challenges [3,4]. These characteristics pose new threats to the operational stability

and reliability of marine diesel engines, particularly affecting cylinder lubrication systems [5,6]. Under conventional fuels, engine manufacturers usually provide recommended lubrication feed rates based on fuel type and operating conditions, typically ranging from 0.8 to 1.4 g/(kWh) [7]. However, with the adoption of LSFO, these reference values have become increasingly unsuitable [8–10]. Especially under fluctuating sea conditions, adjusting the lubrication feed rate (FR) and base number (BN) has become substantially more complex [11–13]. Islam and Martin [14] performed a reliability assessment of marine main propulsion diesel engines, demonstrating that variations in operating conditions and maintenance strategies are critical determinants of engine longevity and operational stability.

Due to the lack of standardized guidelines and model-based support, current lubrication strategies onboard still largely depend on crew members’ subjective judgment and trial-and-error operations. Such experience-driven approaches suffer from poor repeatability, increased maintenance costs, and unstable lubrication performance. An appropriate feed rate is critical to the stable operation of two-stroke marine engines: adequate lubrication enables cooling, cleaning, sealing, and anti-corrosion functions while maintaining a stable oil film to reduce friction and wear. Conversely, over-lubrication may cause excessive deposits and waste, whereas under-lubrication can lead to scuffing, seizure, and even premature failures is illustrated in Figure 1. Under LSFO conditions, variations in fuel composition and operational environments further amplify these uncertainties.

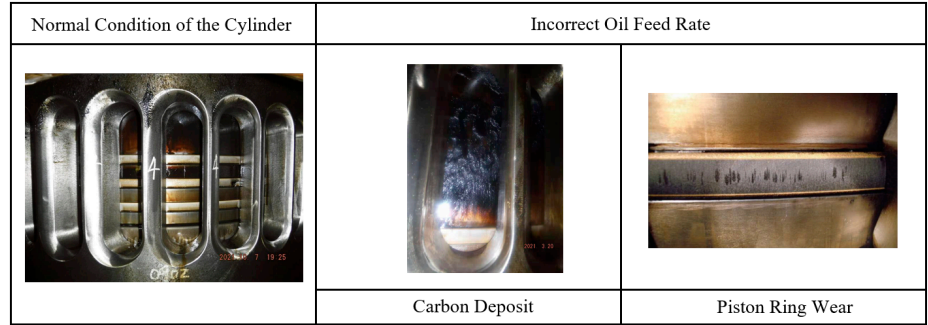


Figure 1. Comparison of cylinder conditions under different lubrication settings.

In summary, the application of LSFO has complicated lubrication feed rate strategies, making it necessary to draw insights from existing studies. However, most related research has focused on fuel properties and tribological mechanisms, with a noticeable gap in addressing intelligent lubrication decision-making. For instance, Cihan et al. [15] combined test-bench experiments and numerical simulations to evaluate the combustion and emission characteristics of biodiesel blends, observing reduced CO and soot but increased NOx and fuel consumption; however, their work was limited to test-bench engines and specific biodiesel formulations, with limited applicability to low-speed two-stroke LSFO scenarios. Geng et al. [16] investigated ternary diesel/ethanol/n-butanol blends in marine engines and introduced a data-driven diagnostic framework, revealing the high sensitivity of combustion and emissions to fuel ratio and injection parameters, but without direct linkage to cylinder surface conditions or lubrication settings. Meanwhile, the development of green shipping has accelerated the adoption of alternative fuels. Sagin et al. [17] emphasized the impact of various alternative fuels on the emissions and performance of marine diesel engines, yet such fuel diversity further increases the uncertainty of lubrication parameter settings. On the other hand, physics-based modeling approaches have also been applied in engine condition monitoring. Fu et al. [18] proposed a physics-driven framework for online condition monitoring of marine engine systems, aiming to support health management. However, such models often rely on simplified assumptions and lack cylinder-liner surface image evidence, thereby limiting their applicability to lubrication decision-making. In

contrast, Jiao et al. [19] developed a PR–CL lubrication model based on modified radial tension, quantifying its effects on oil film distribution and blow-by, yet relying on idealized boundary assumptions and limited validation data. Lyu et al. [20,21] analyzed tribofilm formation and lubrication failure mechanisms under temperature and thermo-mechanical effects, proposing a coupled thermal–frictional framework, but with limited connection to LSFO lubrication strategies. From an engineering perspective, Kamiński [9] validated a feed-rate prediction and optimization method under real operational conditions, demonstrating practical applicability; however, his work did not account for deposit or scuffing risk prediction and lacked image-based validation under diverse sea states and complex disturbances, limiting its external generalizability. Overall, although valuable progress has been made in fuel, tribology, and engineering domains, studies focusing on image-based interpretable retrieval and feed-rate decision support remain scarce. Meanwhile, the cyclical nature of shipping operations and high crew turnover [13] hinder knowledge retention, further exacerbating uncertainties in lubrication management.

Against this backdrop, developing a scientific, standardized, and intelligent lubrication decision system based on objective data has become an urgent requirement in ship operation and engine maintenance. Recent advances in deep learning and computer vision provide promising solutions. As a non-contact, low-cost, and efficient approach, computer vision enables high-resolution image acquisition and intelligent feature analysis for surface condition recognition [5], and has been widely applied in industrial inspection and fault diagnosis. Foundational research has also laid methodological and engineering groundwork: Liu et al. [22] systematically reviewed computer vision and deep learning workflows, providing reproducible practices from preprocessing to deployment for lightweight modeling and edge applications; Kim et al. [23] developed a deep hybrid diagnostic model for ship engines, demonstrating the applicability of deep learning in maritime machinery health monitoring; and Yang et al. [24] integrated visual and sensor data via hybrid deep learning, showing that multimodal fusion enhances detection accuracy and robustness, directly inspiring extensions to cylinder image retrieval frameworks.

In practice, traditional visual inspection of deposits through scavenging ports is subjective and inefficient. By leveraging computer vision, high-resolution liner surface images can be captured and processed with deep learning models to extract discriminative features and link them to historical operating conditions, enabling intelligent matching between images and lubrication feed rates. In particular, retrieval-based approaches allow the system to automatically identify similar cases from historical databases and recommend corresponding lubrication parameters, thereby improving the scientific rigor and stability of lubrication management while reducing reliance on manual expertise and supporting interpretable decision-making for intelligent ship maintenance.

Nevertheless, image-driven research in marine cylinder lubrication remains at an exploratory stage. Zhang Guichen et al. [25] applied Hu invariant moments to extract features from cylinder images, performing normalization by rotation, scaling, and translation to analyze lubrication states. However, this method is highly sensitive to illumination variations and local blurring, often causing feature distortion and reduced retrieval accuracy. In real maritime environments, cylinder images are subject to diverse interferences such as angle deviations, illumination changes, and contaminant effects, leading to high complexity and variability. These challenges expose the limitations of shallow, hand-crafted feature methods [26], highlighting the urgent need for lightweight deep learning models with enhanced robustness and deployability to support intelligent lubrication management in practice.

To address these limitations, this study focuses on typical marine engine cylinders and constructs a high-quality RGB image dataset encompassing diverse wear patterns

and lubrication states under various operating conditions. Although modules such as EfficientNet, Vision Transformers (ViT), RFB, and CBAM have been extensively studied in general vision tasks, their direct application to marine cylinder lubrication image retrieval remains unexplored. The complex maritime environment, with illumination disturbances, oil contamination, and constrained onboard computational resources, introduces unique challenges that have not yet been sufficiently resolved. To address this gap, this study proposes a novel lightweight deep learning architecture, integrating CNNs and ViT to jointly capture local texture details and global structural information. By improving and tailoring these architectures for the marine lubrication scenario, the proposed model aims to maintain retrieval accuracy while reducing computational complexity and enhancing deployment efficiency. This research not only contributes to the intelligent operation and maintenance of marine diesel engine lubrication systems but also provides technological support for achieving emission reduction goals in the maritime industry, highlighting its engineering significance and research value.

The main contributions of this study are as follows:

- (1) A high-quality and representative image dataset of marine engine cylinder lubrication states was constructed using operational data from WINNING's large bulk carriers under various sea and engine conditions.
- (2) A lightweight image retrieval model for marine engine cylinders was proposed, with the following innovations:
 - (A) A novel feature extraction network was designed, integrating EfficientNetB0 as the backbone with MobileViT modules. This hybrid structure enhances the model's sensitivity to complex backgrounds and subtle wear patterns, improving retrieval robustness and accuracy across variable conditions.
 - (B) To further optimize performance, the model incorporates a Receptive Field Block (RFB) to expand the receptive field and improve multi-scale feature modeling, along with the CBAM to strengthen attention to key regions.
- (3) The developed computer vision-based lubrication system was applied on the WINNING UNIVERSE, effectively addressing issues such as insufficient engineering expertise and uncertainty in cylinder oil adjustment.

2. Materials and Methods

2.1. Study Subject

This study focuses on the cylinder systems of marine main engines. The data were collected from two large bulk carriers operated by WINNING (Singapore)—WINNING UNIVERSE and SUNNY KANKAN. Each vessel has a deadweight of approximately 200,000 tons and operates on a fixed route between Yantai Port, China, and Boké Port, Guinea, characterized by complex and variable sea conditions, making them highly representative. Both vessels are equipped with MAN B&W 6S60MC low-speed, two-stroke diesel engines featuring six-cylinder configurations, which are widely used in large bulk carriers and are thus considered representative models for this study.

2.2. Data Acquisition and Preprocessing

2.2.1. Cylinder Image Acquisition from Marine Main Engines

This study aims to develop an image retrieval model for marine engine cylinders, utilizing an image-based matching mechanism to optimize current lubrication strategies based on historical cylinder conditions and corresponding oil feed rate data. The goal is to reduce the risk of failures such as carbon buildup and scuffing, thereby ensuring the safe and stable operation of main engines.

For image acquisition, a digital camera was employed to capture fine details of wear and carbon deposits on the inner cylinder wall. The process involved manual operation during engine shutdown: crew members entered the engine room and photographed the inner walls of each cylinder through the scavenging ports from multiple angles. This ensured comprehensive visual coverage of critical features such as deposit formation and surface wear, providing high-quality data for model training and lubrication optimization. The data collection workflow is illustrated in Figure 2.



Figure 2. Data acquisition workflow.

To enhance representativeness and applicability, image acquisition covered typical operating scenarios such as engine shutdowns and port stays, under various sea conditions. Since 13 February 2020, a total of 2838 cylinder images have been collected, forming a dataset with broad temporal and operational diversity.

2.2.2. Preprocessing of Cylinder Images from Marine Main Engines

Given the absence of publicly available datasets specifically for marine engine cylinder image retrieval, and the impracticality of capturing scavenging port images during normal ship operation, the image samples in this study are primarily sourced from the bulk carriers WINNING UNIVERSE and SUNNY KANKAN. Due to the limited number of raw images, data augmentation was applied to enhance structural diversity, improve model generalization, and mitigate overfitting. Augmentation techniques included image rotation, random cropping, brightness adjustment, sharpness variation, and random occlusion—simulating variations caused by lighting, equipment, and operator differences. Examples of augmented images are shown in Figure 3.

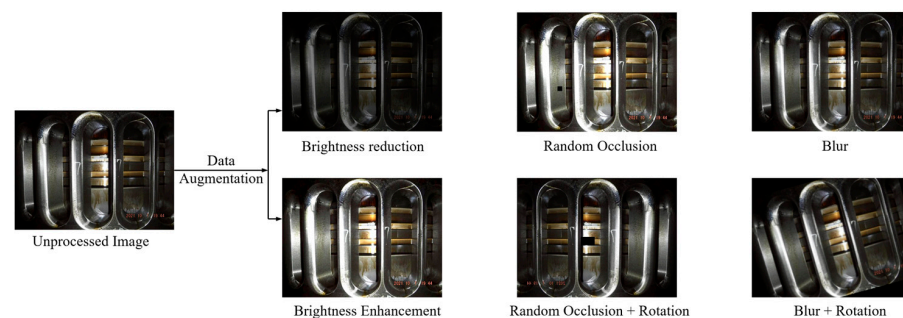


Figure 3. Illustration of data augmentation techniques for typical scavenging port images.

Following augmentation, the dataset was expanded to 31,218 images. To meet memory constraints during deep learning training, all images were resized to 224×224 pixels prior

To enhance the model's multi-scale perception and improve local structure representation, a Receptive Field Block (RFB) is introduced following the backbone. RFB employs parallel convolutional paths with varying receptive field sizes to effectively capture features at different spatial scales, improving the network's ability to represent fine structural details such as scratches and carbon deposit edges [28].

To address the limitations of convolutional networks in modeling global dependencies, a MobileViT module is incorporated to perform global feature modeling. By leveraging the capabilities of Vision Transformers, this module captures long-range dependencies and complements the CNN's local focus, enhancing the completeness and expressiveness of feature representations [29].

To further improve the model's discriminative capacity for key regions, a CBAM is integrated [30]. CBAM combines channel and spatial attention mechanisms to adaptively highlight critical regions and suppress background noise and redundant information, enhancing sensitivity to wear and deposit patterns on cylinder surfaces.

In the output stage, Global Average Pooling is applied to reduce feature dimensionality and improve generalization, while an L2 Normalization layer standardizes embedding vectors as unit-length outputs. This ensures consistent and stable similarity computation during retrieval. The coordinated design of these modules significantly enhances the model's performance in marine cylinder image retrieval tasks.

2.4.1. Backbone Construction Based on EfficientNetB0

The backbone network in this study integrates EfficientNetB0 with RFB and MobileViT modules, achieving a balance between computational efficiency and feature representation. By combining receptive field expansion and global modeling, the network strengthens its capability to capture image semantics. EfficientNetB0, based on Mobile Inverted Bottleneck Convolution (MBConv) and a compound scaling strategy, adjusts depth, width, and resolution jointly to optimize the trade-off between compactness and performance. Its overall architecture is shown in Figure 6. The MBConv module expands channels through a 1×1 pointwise convolution, followed by batch normalization (BN) and the Swish activation function, as defined in Equation (1):

$$\text{Swish}(x) = x \cdot \sigma(x) = x \cdot \frac{1}{1 + e^{-x}} \quad (1)$$

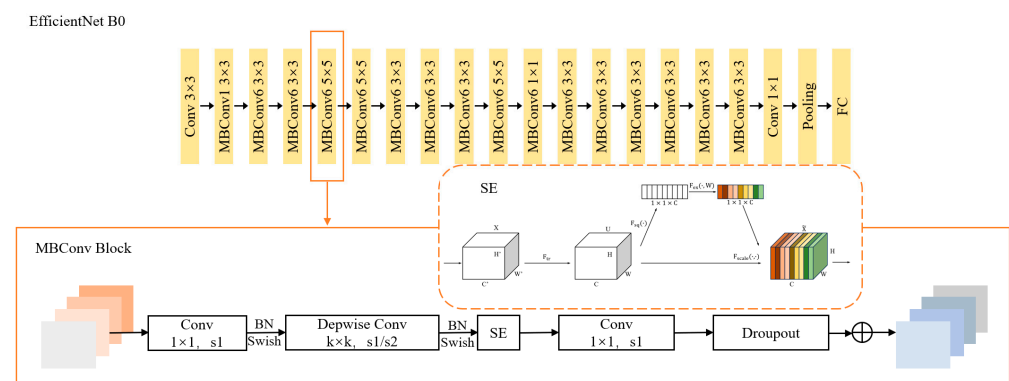


Figure 6. Backbone architecture based on EfficientNetB0.

In this activation, x denotes the input to the nonlinear unit (the output of the preceding layer). The function $\sigma(x)$ is the logistic sigmoid, $\sigma(x) = \frac{1}{1 + e^{-x}}$, which maps each input to $(0, 1)$. The Swish function is then defined as $\text{Swish}(x) = x \cdot \sigma(x)$, applied element-wise for vectors or tensors.

This is followed by depthwise separable convolution for spatial feature extraction and another 1×1 convolution for channel compression. Finally, a Squeeze-and-Excitation (SE) attention mechanism is integrated to model and reweight inter-channel importance, thereby enhancing the model's responsiveness to critical features. The corresponding operations are shown in Equations (2)–(5):

$$U_c = V_c \times X = \sum_{s=1}^{C'} V_c^s \times X^s \quad (2)$$

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (3)$$

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (4)$$

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c u_c \quad (5)$$

In these equations, U_c represents the transformed feature representation, which is obtained by weighting the input features X with the weight matrix V_c . z_c is the compressed result obtained by applying the Squeeze operation to the input feature u_c , representing the importance of the channel features. Next, s is the scalar value calculated using the activation function σ , which adjusts the weight of the features. Finally, the weighted feature $\tilde{x}_c$ is generated by scaling the input feature u_c with the channel weight s_c .

MBConv applies depthwise convolution with stride 1 and residual connections when maintaining resolution, and uses stride 2 with increased channels for downsampling to enhance high-level feature representation.

To enlarge the receptive field and improve anomaly detection, an RFB module is integrated into the backbone. Using a multi-branch design with 1×1 , 3×3 , and 5×5 kernels, each followed by dilated convolutions with varying rates, RFB simulates multi-scale receptive fields while maintaining low complexity. The concatenated outputs form a rich representation that complements EfficientNetB0 in capturing complex wear patterns and detecting subtle anomalies in cylinder lubrication images, as illustrated in Figure 7.

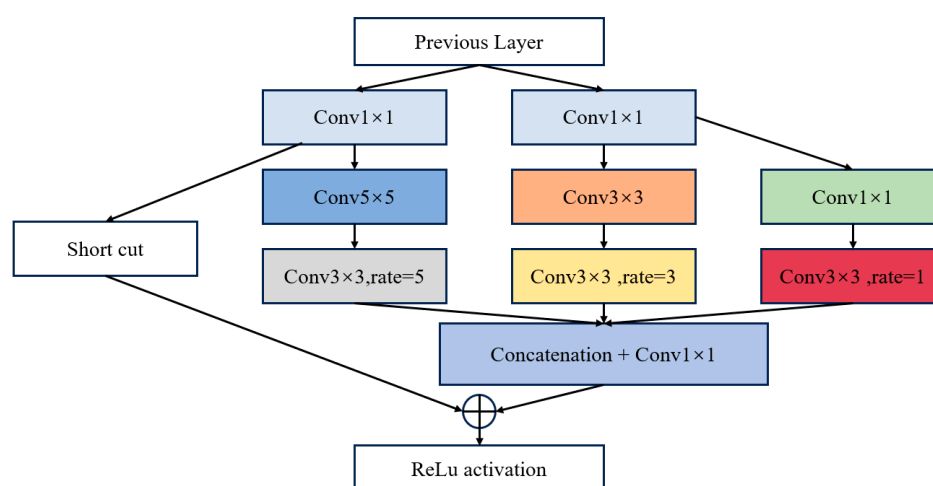


Figure 7. Architecture of the Receptive Field Block (RFB).

To overcome the limitations of convolutional structures in modeling long-range dependencies, this study integrates the MobileViT module into the backbone network. By embedding self-attention mechanisms within local windows, the module enables global context modeling while preserving the local receptive capabilities of convolution, thereby

enhancing structural representation completeness and robustness under complex operating conditions.

The MobileViT module consists of three components: a local feature encoding module, a global feature encoding module, and a feature fusion module. Its overall architecture is shown in Figure 8. First, the local encoding module receives an input tensor $X \in \mathbb{R}^{H \times W \times C}$ and applies a standard $n \times n$ convolution to extract local spatial features. This is followed by a 1×1 pointwise convolution to project the features into a higher-dimensional space, yielding an intermediate representation of size $\mathbb{R}^{H \times W \times d}$.

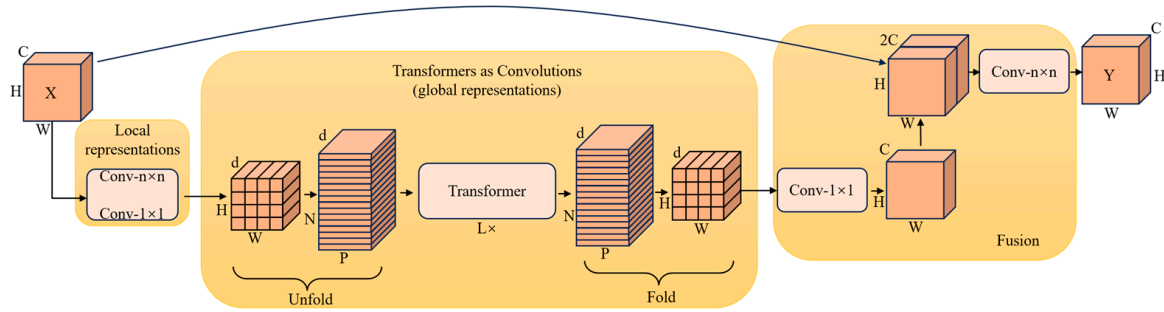


Figure 8. Illustration of the convolutional fusion mechanism in the MobileViT Block.

Next, the tensor is partitioned into N non-overlapping flattened patches and reshaped into $\mathbb{R}^{d \times N \times P}$, where $P = h \times w$ represents the number of pixels per patch. These patches are then passed through a Transformer encoder in the global encoding module to model long-range dependencies between them, producing a global feature tensor of the same dimensionality $\mathbb{R}^{d \times N \times P}$. This process effectively maps local features to global semantics. The encoding procedure is shown in Equation (6):

$$X_G(p) = \text{Transformer}(X_U(p)), 1 \leq p \leq P \quad (6)$$

Here, $X_U(p)$ is the locally encoded feature of the p -th patch, $\text{Transformer}(\cdot)$ denotes global encoding, and $X_G(p)$ is the globally encoded output, with $p \in [1, P]$ and $P = h \times w$.

The feature fusion module then reshapes the output of the global encoding module back to the original spatial dimensions, yielding a tensor of size $\mathbb{R}^{H \times W \times d}$. A 1×1 convolution is applied to project this tensor into a lower-dimensional space, which is then concatenated with the original input feature X to form a new tensor $Y \in \mathbb{R}^{H \times W \times 2C}$. Finally, an $n \times n$ convolution is used to fuse the concatenated features, producing the final output tensor $Y \in \mathbb{R}^{H \times W \times C}$.

Through this three-stage structure, the MobileViT module effectively integrates local and global information, providing richer feature representations for cylinder lubrication image retrieval. The improved backbone network, rebuilt upon the EfficientNetB0 structure, is shown in Figure 9.

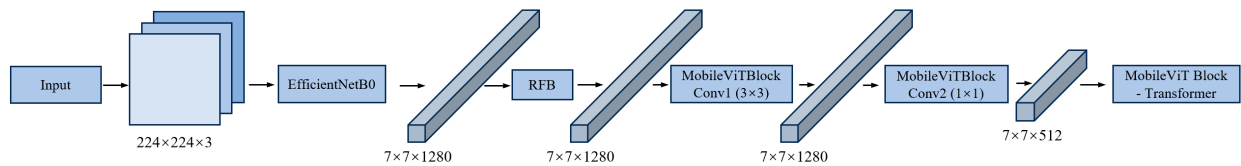


Figure 9. Workflow of the improved backbone network.

The structural parameters are shown in Table 1.

Table 1. Structural parameters of the backbone network.

Stage	N	Params	Module Name	Feature Map Size
Image Input			—	$224 \times 224 \times 3$
Feature Extraction	1	896	Conv3 $\times$ 3	$224 \times 224 \times 32$
	1	304	MBConv1, $k3 \times 3$	$112 \times 112 \times 16$
	1	3192	MBConv6, $k3 \times 3$	$112 \times 112 \times 24$
	1	9400	MBConv6, $k5 \times 5$	$56 \times 56 \times 40$
	1	21,440	MBConv6, $k3 \times 3$	$28 \times 28 \times 80$
	2	131,744	MBConv1, $k5 \times 5$	$14 \times 14 \times 192$
	1	135,264	MBConv6, $k3 \times 3$	$7 \times 7 \times 320$
Receptive Field	1	247,040	Conv1 $\times$ 1	$7 \times 7 \times 1280$
	1	1,474,560	RFB Module	$7 \times 7 \times 1280$
Local Modeling	1	14,746,880	MobileViTBlock.Conv1	$7 \times 7 \times 1280$
	1	655,872	MobileViTBlock.Conv2	$7 \times 7 \times 512$
	2	5,767,168	Transformer Encoder	$7 \times 7 \times 512$
	1	656,640	MobileViTBlock.Conv3	$7 \times 7 \times 1280$
	1	29,492,480	MobileViTBlock.Conv4	$7 \times 7 \times 1280$

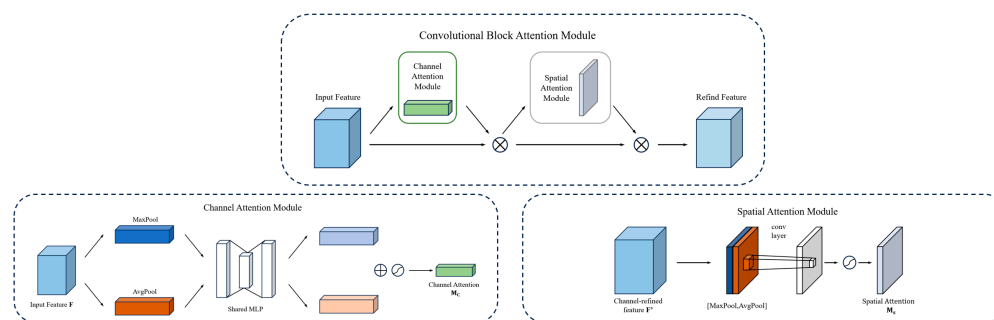
2.4.2. CBAM Attention Enhancement Mechanism

Images from scavenging ports often contain interferences such as oil stains that visually resemble target regions, increasing retrieval difficulty. To address this, the CBAM module is integrated into the backbone, where channel and spatial attention guide the network to focus on key regions (e.g., carbon deposits, piston rings, pistons) while suppressing irrelevant noise. This enables adaptive feature refinement with minimal additional complexity.

As shown in Figure 10, CBAM sequentially applies channel and spatial attention to refine input features. Channel attention generates descriptors via global average and max pooling, which are passed through a shared multilayer perceptron (MLP) to compute channel weights. The reweighted features are then processed by spatial attention, where average and max pooling along the channel dimension produce spatial maps. After concatenation, a 7×7 convolution and Sigmoid activation yield the spatial attention map, which is multiplied with the input features to produce the final output. This mechanism adds negligible computational cost while significantly enhancing the model's focus on informative regions and its feature representation capacity. The attention computation is shown as follows:

$$M_c(F) = \sigma \{ M_{LP} [P_{avg}/ + M_{LP} [P_{max}(F)]] \} \\ = \sigma \{ W_1 [W_0 (F_{avg}^c)] + W_1 [W_0 (F_{max}^c)] \} \quad (7)$$

$$M_s(F) = \sigma \{ f^{7 \times 7} [P_{avg}(F); [P_{max}(F)]] \} \\ = \sigma \{ f^{7 \times 7} (F_{avg}^s; F_{max}^s) \} \quad (8)$$

**Figure 10.** Schematic diagram of the CBAM attention mechanism module.

Here, $M_c(F)$ denotes the output of the channel attention module, and $M_s(F)$ denotes the output of the spatial attention module. F is the input feature map. AvgPool and MaxPool represent global average pooling and global max pooling operations, respectively; MLP refers to a multilayer perceptron; σ denotes the Sigmoid activation function; $f^{7 \times 7}$ indicates a 7×7 convolution operation; and $[\text{AvgPool}(F); \text{MaxPool}(F)]$ denotes the concatenation of the pooled feature maps along the channel axis.

2.4.3. Design of the Lightweight Retrieval Module

After feature enhancement by the CBAM module, the lightweight image retrieval module receives a high-dimensional feature map of size $7 \times 7 \times 1280$. First, global average pooling (GAP) is applied to compress the feature map into a $1 \times 1 \times 1280$ vector, effectively aggregating semantic information across the entire image. This operation reduces feature dimensionality while preserving global responses across channels, facilitating the construction of a unified and compact image representation for subsequent retrieval tasks.

Next, the resulting feature vector is passed through an L2 normalization layer, projecting it onto a unit hypersphere by normalizing its norm to 1. This step eliminates magnitude disparities among features and ensures that similarity comparisons focus on directional differences rather than absolute values, thereby improving the stability and comparability of distance calculations. The normalized features exhibit stronger discriminative capability, contributing to improved retrieval accuracy.

In practical applications, the same feature extraction and normalization processes are applied to both the query image and all images in the database. Image similarity is then measured using cosine similarity, and the top-ranked matches with the highest similarity scores are selected as retrieval outputs. The cosine similarity is shown in Equation (9):

$$\text{Sim}(f_1, f_2) = \frac{f_{1i} \cdot f_{2i}}{|f_{1i}| |f_{2i}|} = \frac{\sum_{i=1}^n f_{1i} \cdot f_{2i}}{\sqrt{\sum_{i=1}^n f_{1i}^2} \cdot \sqrt{\sum_{i=1}^n f_{2i}^2}} \quad (9)$$

In this equation, f_1 and f_2 represent the feature vectors of the query image and a database image, respectively.

This retrieval mechanism provides marine engineering personnel with an efficient and intuitive basis for image matching, facilitating rapid assessment of cylinder lubrication conditions and identification of carbon deposits or abnormal wear. It offers reliable support for oil feed rate adjustment, maintenance decision-making, and fault prediction.

2.5. Model Training

During the experimental phase, a rigorously controlled environment was employed to ensure the reproducibility and reliability of results. Python 3.12 was used as the primary programming language in conjunction with the Windows 11 operating system, providing a stable foundation for experimentation. The deep learning framework PyTorch 2.7 served as the core platform for model training and evaluation. Additionally, computational efficiency was enhanced through the use of an NVIDIA GeForce RTX 4060 GPU and a 13th-generation Intel Core i9-13980HX CPU. The key hyperparameter settings used during model training are summarized in Table 2.

Table 2. Hyperparameter settings for model training.

Hyperparameter	Value
Weights	Pretrained EfficientNetB0
Epochs	150
Batch-size	32

Table 2. *Cont.*

Hyperparameter	Value
Image-size	224×224
LR	0.00001
Optimizer	Adam
Others	Remain unchanged

2.6. Model Evaluation

To comprehensively evaluate the model's accuracy in image retrieval tasks, this study adopts three commonly used performance metrics: True Positive (TP), False Positive (FP), and False Negative (FN). TP refers to the number of correct target samples successfully retrieved by the model that match the query image. FP denotes the number of non-target samples incorrectly identified as matches, while FN represents the number of actual target samples the model fails to retrieve. These metrics collectively reflect the model's correctness and error rate from different perspectives. The corresponding calculation formulas are as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (10)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (11)$$

$$\text{F1-Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (12)$$

To more comprehensively assess the practical deployment potential of the image retrieval model, this study introduces three lightweight evaluation metrics—model size, GFLOPs, and FPS—as effective complements to accuracy-based indicators. From a computational resource perspective, model size reflects the parameter space occupied by the model, indicating its demand for memory and storage. A smaller model size implies lower deployment requirements, particularly suitable for resource-constrained environments such as edge devices and embedded systems.

GFLOPs (Giga Floating Point Operations) measures the total number of floating-point operations required for a single forward pass, serving as an indicator of computational complexity. Higher GFLOPs values suggest greater computational load and potentially longer inference time. FPS (Frames Per Second) quantifies the number of image frames the model can process per second, providing a key measure of real-time performance. A higher FPS indicates faster response and better suitability for real-time image retrieval tasks.

Together, these three metrics constitute essential dimensions for evaluating the model's lightweight characteristics, facilitating deployment optimization and runtime efficiency without compromising retrieval accuracy.

3. Results

3.1. Comparison of Backbone Network Architectures

This study aims to develop a lightweight yet high-performance image retrieval model for cylinder lubrication, targeting practical deployment in resource-constrained environments. To this end, a systematic comparison was conducted across several mainstream convolutional neural network architectures, including AlexNet [31], VGG16 [32], GoogLeNet (Inception v1) [33], ResNet50 [34], and EfficientNetB0.

The evaluation was performed using four metrics: Top-1 Accuracy, GFLOPs, FPS, and Model Size. Among these, model size was considered the primary criterion, reflecting deployment feasibility under storage constraints. The other three metrics served as comple-

mentary indicators to assess the trade-off between accuracy and computational efficiency. The detailed results are presented in Table 3.

Table 3. Comparison of parameters across mainstream network architectures.

Model Name	Top-1 Accuracy (%)	GFLOPs	FPS	Size (MB)
AlexNet	57.2	0.72	239.96	240
VGG16	71.5	15.5	8.78	528
GoogLeNet	69.8	1.5	47.68	27
ResNet50	76.0	4.1	17.46	102
EfficientNetB0	77.1	0.39	130.14	21

The results indicate that among the five evaluated CNN architectures, EfficientNet-B0 demonstrates a significant advantage in model compactness, with a parameter size of only 5.3 MB—substantially smaller than the other models. It also outperforms the alternatives in both Top-1 classification accuracy (77.1%) and GFLOPs (0.39), exhibiting a well-balanced trade-off between accuracy and computational efficiency. Although its inference speed is slightly lower than that of AlexNet, EfficientNet-B0 achieves a considerably smaller model size while maintaining high accuracy and low computational cost, making it more suitable for real-world deployment scenarios. Accordingly, this study adopts EfficientNet-B0 as the baseline model for subsequent exploration of lightweight structural design and further optimization.

3.2. Comparison of Lightweight Model Improvements

To further enhance the performance of the image retrieval model while meeting lightweight deployment requirements, this study introduces architectural improvements based on EfficientNet-B0. By integrating the RFB, MobileViT Block, and CBAM attention mechanism into the EfficientNet-B0 framework, the model achieves significant gains in feature extraction and representation capability without substantially increasing its size.

The objective of these enhancements is to improve retrieval accuracy for cylinder lubrication images while preserving a lightweight structure suitable for real-world deployment. To comprehensively evaluate the effectiveness of the proposed modifications, this study compares the original and improved models using the pre-divided dataset across six dimensions: Recall, Precision, F1-score, Model Size, GFLOPs, and FPS. The detailed performance differences and improvements are presented in Table 4.

Table 4. Performance comparison of different networks after optimization.

Model	Recall (%)	Precision (%)	F1-Score (%)	Model Size (MB)	GFLOPs	FPS
EfficientNetB0	84.33	87.90	86.08	21	0.39	133
EfficientNetB0 + MobileVit block	93.04	93.50	93.27	27	0.60	108
EfficientNetB0 + MobileVit block + ECA	93.15	93.56	93.35	27.5	0.61	104
EfficientNetB0 + MobileVit block + CBAM	93.18	93.56	93.37	27.8	0.63	101
EfficientNetB0 + RFB + MobileVit block + CBAM	99.69	99.71	99.70	29.3	0.69	98

As shown in Table 4, all improved models exhibit significantly better retrieval performance than the original EfficientNet-B0, while maintaining low computational overhead. Notably, the incorporation of the MobileViT Block increases the recall to 93.04% and the F1-score to 93.27%, representing a substantial improvement over the baseline. Meanwhile, the model size increases by only approximately 6 MB, and GFLOPs rise from 0.39 to 0.60—still within the lightweight range.

In the comparative experiments involving attention mechanisms, both ECA and CBAM modules contributed to improvements in retrieval performance. Overall, the performance

advantage of CBAM over ECA was relatively limited. Both achieved identical precision scores (93.56%), with a slight increase in recall from 93.15% to 93.18% and a marginal rise in F1-score from 93.35% to 93.37%. Given the minimal differences in quantitative metrics, this study further employed heatmap visualizations to intuitively compare the two mechanisms in terms of feature extraction and spatial attention. The visualization results are shown in Figure 11.

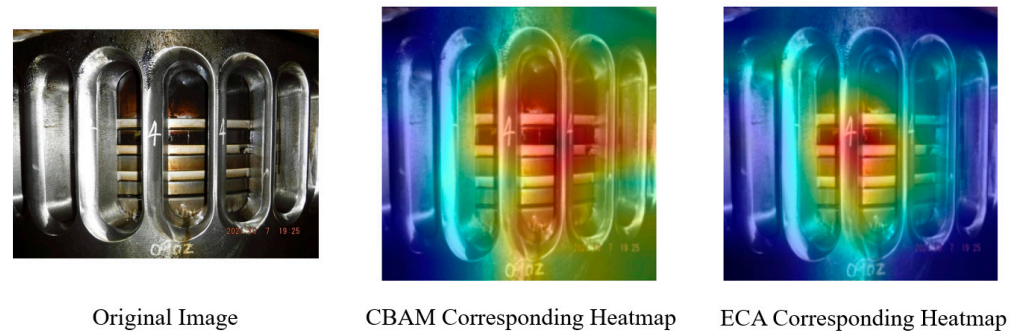


Figure 11. Comparison of effects among different attention mechanisms.

As shown in the heatmap comparison, CBAM (left) yields more comprehensive coverage of the cylinder region, effectively capturing the entire area of interest. In contrast, the ECA heatmap (right) exhibits narrower focus, with a risk of missing critical structures, indicating weaker spatial awareness. The color scale reflects the relative attention intensity, where warmer colors (e.g., red and yellow) indicate regions with higher attention weights, while cooler colors (e.g., blue) represent lower responses. Based on this observation, CBAM was ultimately selected as the preferred attention mechanism for this study.

To further enhance retrieval accuracy, the RFB was introduced alongside the attention mechanism to expand the effective receptive field and strengthen the representation of local features. The integration of RFB significantly improved the model's ability to capture multi-scale semantic information, thereby enhancing feature matching robustness. Experimental results confirm that the model achieved optimal performance with the addition of RFB: Recall reached 99.69%, Precision 99.71%, and F1-score 99.70%. Meanwhile, the model maintained a compact size of 29.3 MB, a computational complexity of 0.69 GFLOPs, and an inference speed of 98 FPS, striking a favorable balance between accuracy and efficiency. These results underscore its strong engineering applicability and deployment potential.

Taking into account retrieval performance, model size, and inference speed, this study identifies the EfficientNetB0 + RFB + MobileViT Block + CBAM configuration as the optimal lightweight solution for cylinder lubrication image retrieval tasks.

3.3. Modeling Results of the Optimal Lightweight Strategy

Model convergence is also a critical indicator for evaluating reliability. Therefore, this study visualizes the training dynamics of the optimal lightweight cylinder lubrication image retrieval model—EfficientNetB0 + RFB + MobileViT Block + CBAM. The plotted curves include training loss (Train Loss), validation loss (Val Loss), Precision, Recall, and F1-score over the course of training. The corresponding results are shown in Figures 12–15.

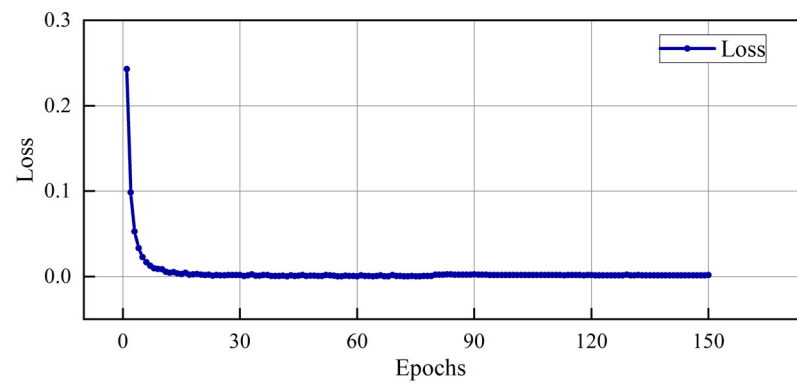


Figure 12. Training loss curve.

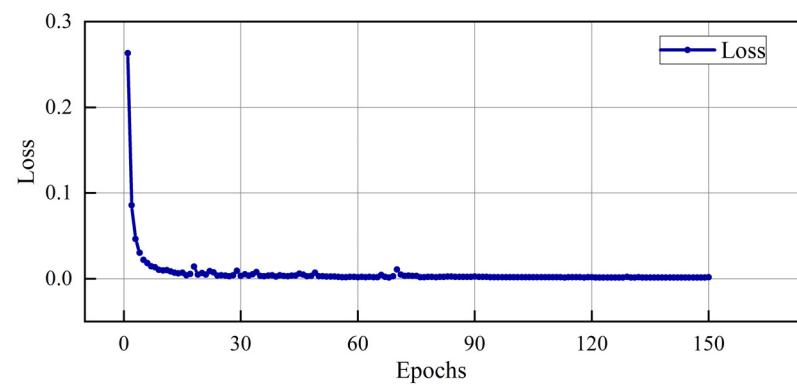


Figure 13. Validation loss curve.

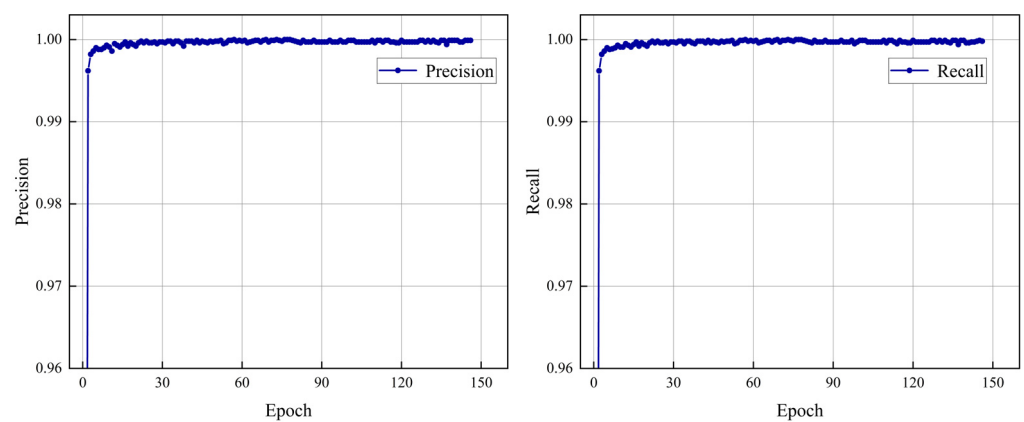


Figure 14. Precision curve and Recall curve.

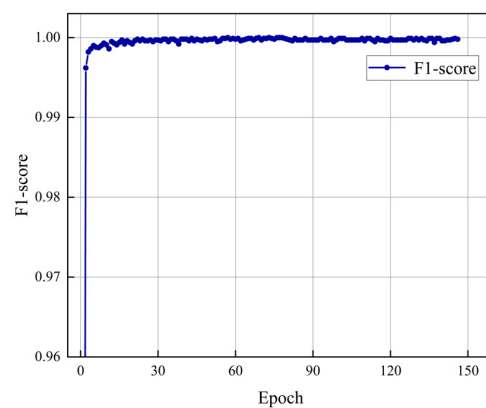


Figure 15. F1-Score curve.

The results demonstrate that the loss curves of the EfficientNetB0 + RFB + MobileViT Block + CBAM model exhibit effective convergence during both training and validation, indicating the model's reliability. Moreover, the training curves of Precision, Recall, and F1-score also show clear convergence and optimization. These findings confirm the robustness and reliability of the proposed lightweight image retrieval model for cylinder lubrication.

3.4. Retrieval Results of Cylinder Lubrication Images

This study aims to develop an image retrieval model for analyzing the lubrication state of marine engine cylinders. To ensure practical applicability and generalization, the collected dataset encompasses typical variations such as carbon buildup and wear under diverse sea conditions. To evaluate the model's retrieval performance across different operating scenarios, a subset of image samples was randomly selected during the testing phase to conduct real-world retrieval tasks. The objective was to assess matching accuracy and robustness under multi-condition backgrounds. The detailed results are presented in Figures 16 and 17 and Table 5.

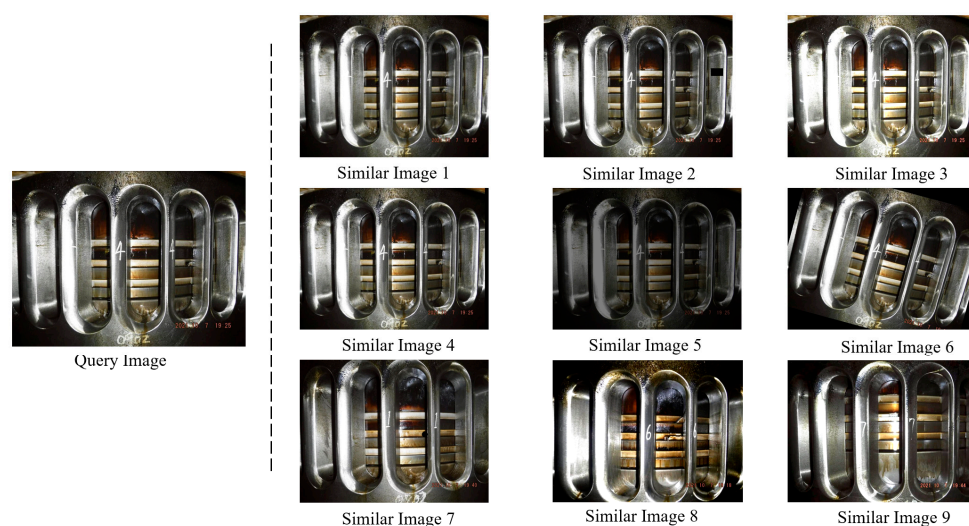


Figure 16. Example retrieval results for typical scavenging port images.

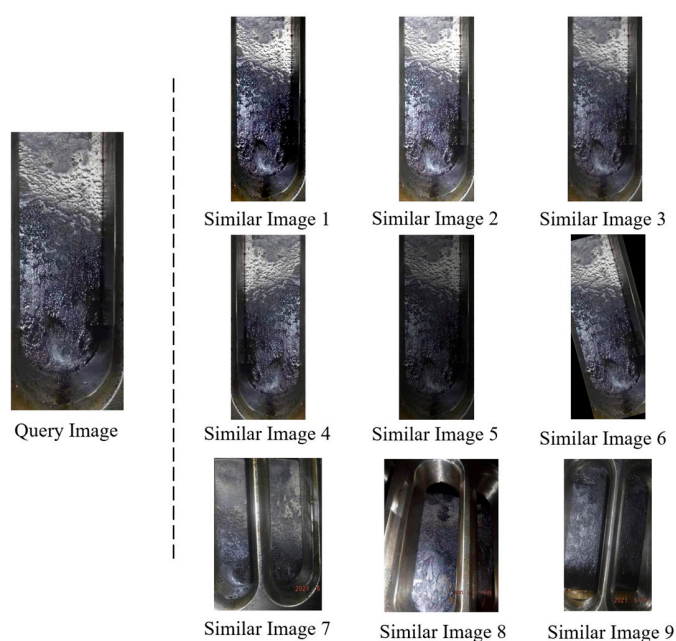


Figure 17. Example retrieval results for typical cylinder liner carbon deposit images.

Table 5. Retrieval results.

Similar Image	Cosine Similarity (%)	Similar Image	Cosine Similarity (%)
1	99.95	1	99.65
2	99.72	2	99.59
3	99.66	3	99.39
4	99.06	4	98.63
5	98.97	5	98.21
6	98.88	6	97.90
7	98.12	7	91.25
8	97.61	8	91.07
9	97.20	9	90.42

Figure 17 Cosine similarities

Figure 18 Cosine similarities

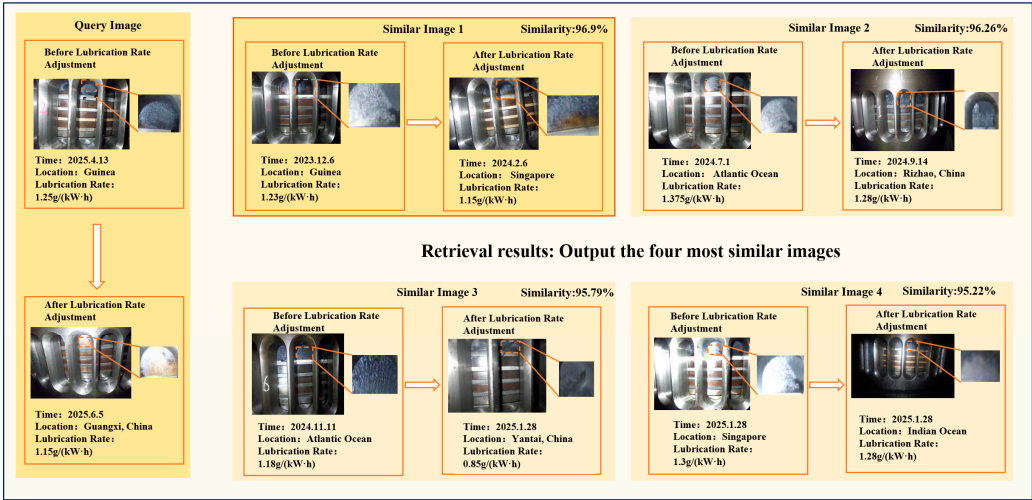


Figure 18. Image Retrieval Results for Lubrication Rate Adjustment: Comparison of Query and Similar Cases.

Application and validation experiments were conducted using the proposed lightweight cylinder lubrication image retrieval model on both full-cylinder images and localized carbon deposit regions. The results demonstrate that the model achieves high retrieval accuracy and exhibits strong stability and generalization across diverse input conditions. Specifically, it is capable of efficiently and accurately identifying the most similar historical image samples from the augmented dataset, thereby providing reliable decision support and a model foundation for optimizing cylinder oil feed rates.

3.5. Application Workflow of the Cylinder Lubrication Image Retrieval Model

During port stays, engine crew typically conduct routine inspections of the main engine cylinders via scavenging ports. To enable intelligent evaluation of lubrication conditions and optimization of oil feed rates, this study integrates image acquisition into the inspection process and inputs the captured scavenging port images into the proposed lightweight image retrieval system.

During retrieval, the system compares the input image with samples in the historical image database and returns the most visually similar image, along with its associated operational data—such as oil feed rate, operating time, and wear condition—as reference for current lubrication decisions. It is important to note that, to ensure the effectiveness of strategy adjustment, the system must prioritize similar samples that demonstrated improved operational outcomes during subsequent inspection cycles. This avoids incorrect recommendations based on degraded historical samples and enhances the reliability and practical applicability of the model in real-world engine environments.

3.6. Real-Ship Application

On 13 April 2025, a shutdown inspection was conducted on the main engine at Boké Port, Guinea, during which cylinder liner images were captured through the scavenge port. The images revealed significant carbon deposits on the piston crown, with the recorded lubrication rate for that voyage segment being 1.25 g/(kW·h) (see Figure 18, top left). In the context of operational constraints and limited maintenance window, determining the appropriate adjustment of the lubrication rate based on historical data became a key decision-making factor for optimizing cylinder lubrication performance and preventing further damage.

To support this decision, the captured image was processed using the lightweight cylinder lubrication image retrieval model developed in this study. The system's database stores metadata for each historical image, including the date, location, operational context, and lubrication rate, along with follow-up inspection images and their respective metadata. The system also retains samples with adverse outcomes (e.g., cases where carbon buildup increased or scuffing occurred after adjustments) to provide both positive and negative reference cases for decision-making.

The retrieval module returned the top-10 most similar cases to the query image, and for each case, an “adjustment before—adjustment after” comparison was made. Due to space limitations, only four representative cases with high similarity scores (96.9%, 96.26%, 95.79%, and 95.22%, see Figure 18, right side and bottom) are presented. Among these cases, only Similar Image 1 demonstrated a clear reduction in carbon deposits following lubrication rate adjustment.

Specifically, Similar Image 1 was captured on December 6, 2023, at Boké Port, Guinea, and the follow-up inspection took place on February 6, 2024, in Singapore. In this case, the lubrication rate was adjusted from 1.25 g/(kW·h) to 1.15 g/(kW·h), resulting in a noticeable reduction in carbon buildup on the piston crown without any increase in scuffing (see Figure 18, top right, “Similar Image 1” before-and-after comparison). This adjustment outcome closely aligned with the current scenario, making it a valuable reference for the present lubrication rate adjustment.

In contrast, the remaining three similar images (Similar Images 2, 3, and 4) showed lubrication rate changes, but their follow-up images did not indicate a significant reduction in carbon deposits. In some instances, carbon buildup either remained unchanged or worsened. Therefore, these images were not considered valid references for this adjustment, as they did not exhibit similar improvements.

Based on the positive results from Similar Image 1, a conservative lubrication rate reduction strategy was adopted for the current voyage, adjusting the rate from the previous range of 1.23–1.25 g/(kW·h) to 1.15 g/(kW·h). This decision was made after a thorough analysis of historical cases, taking into account factors such as fuel quality, load conditions, and other operational parameters, selecting a strategy that would both reduce carbon buildup and avoid risks of inadequate lubrication.

After the adjustment, the vessel completed its return voyage and docked at Fangchenggang Port, Guangxi, on 5 June 2025, where a follow-up inspection was conducted on the same day. New images were captured through the scavenge port, and after automatic evaluation by the retrieval model and manual verification, the results showed a significant reduction in carbon deposits compared to the baseline images from April, without any increase in scuffing (see Figure 18, bottom left, “After Lubrication Rate Adjustment”). This outcome validates the effectiveness of the image retrieval-assisted lubrication rate adjustment strategy.

4. Discussion

This study aims to develop an efficient and lightweight image retrieval model for cylinder lubrication in marine engines, supporting intelligent oil feed rate optimization. In the context of frequent crew rotation and limited experience transfer, the proposed system leverages visual recognition to automatically match historical lubrication strategies, thereby mitigating fluctuations in lubrication management caused by human factors and demonstrating strong practical value.

In this research, we first reviewed the relevant literature to verify the feasibility and effectiveness of image retrieval-based lubrication adjustment. However, challenges such as lighting variation, shadow interference, and structural occlusion in scavenging port image acquisition significantly increased the complexity of image processing and model design. While higher algorithmic complexity may improve retrieval accuracy, it often leads to increased model size and slower inference—factors incompatible with real-world deployment on resource-constrained marine platforms.

To address these constraints, this study proposes a lightweight deep learning model that integrates CNNs with ViT modules for efficient image retrieval. EfficientNetB0 was selected as the backbone network after a comparative analysis of mainstream lightweight retrieval architectures. With a model size of only 21 MB and a Top-1 accuracy of 77.1%, EfficientNetB0 strikes a strong balance between accuracy and efficiency.

To enhance global structural modeling capabilities, the MobileViT module was integrated into the EfficientNetB0 backbone. While CNNs excel at local feature extraction, they are limited in capturing global contextual relationships. The addition of MobileViT improved Precision without significantly increasing model size (an approximate 6 MB increase).

To strengthen the connection between CNN and Transformer components, a RFB was introduced between EfficientNetB0 and MobileViT. The RFB, composed of multi-branch dilated convolutions, simulates features across varying receptive fields, enhancing the model's ability to capture multi-scale structures and contextual dependencies—thus providing richer semantic input to the Transformer layers.

Additionally, two lightweight attention mechanisms, ECA and CBAM, were evaluated to further improve focus on salient image regions. CBAM was ultimately selected for its superior performance in retrieval accuracy, achieved without significantly increasing model parameters.

In summary, this study presents a lightweight feature extraction model comprising EfficientNetB0, RFB, MobileViT Block, and CBAM. The final model achieves an effective balance between accuracy, size (29.3 MB), and inference efficiency (GFLOPs within acceptable bounds), meeting the requirements for embedded deployment. The method has also undergone preliminary validation in real-ship operating environments, demonstrating its practical applicability.

5. Conclusions

This study proposes a lightweight image retrieval model for cylinder lubrication management by introducing targeted structural modifications to the EfficientNetB0 backbone. The designed network achieves a favorable balance between recognition accuracy and deployment efficiency, exhibiting clear advantages over baseline architectures. Quantitative evaluation on real-ship operational datasets demonstrates excellent retrieval performance, with a Precision of 99.71%, Recall of 99.69%, and F1-score of 99.70%. The model size is constrained to 29.3 MB with only a 0.3 GFLOPs increase in computational complexity, making it well suited for deployment in resource-constrained shipboard environments while satisfying the requirements of real-time inference. Compared with the lightweight backbone model EfficientNetB0, the proposed method achieves a more advantageous

trade-off between accuracy and complexity, delivering superior retrieval performance with comparable or lower computational overhead, thereby underscoring its potential for edge deployment.

The practical value of the model is reflected in its “retrieval-driven decision support” mechanism. Under the widespread application of low-sulfur fuels, the system can retrieve case samples and corresponding empirical adjustment strategies from historical databases according to the currently acquired images. In doing so, it provides traceable and interpretable references for engineers in oil-feed regulation. By leveraging similarity metrics and Top-K nearest neighbor retrieval, the model reduces reliance on subjective experience, enhances transparency and explainability in decision-making, and contributes to improving the operational quality and cost-effectiveness of lubrication management. This, in turn, helps mitigate the risks of accelerated liner wear and increased lubricant consumption caused by excessive or insufficient lubrication.

Nevertheless, several aspects of this study remain to be improved. First, the training data are mainly derived from specific vessel types, fixed engine series, and limited operating conditions, which constrains the generalization capability and applicability of the model. Second, the shipboard validation was conducted on a single bulk carrier, serving primarily as a preliminary application trial. While these results provide an important foundation for subsequent research, they do not yet cover systematic validation across multiple vessel types and operating scenarios. In addition, significant differences in computational resources, interface protocols, and control logic across embedded platforms pose practical challenges for large-scale integration and cross-platform deployment.

Future work will focus on three directions: (1) expanding data collection and validation to include a wider variety of vessel types and operating conditions, thereby enhancing model generalization and robustness; (2) further optimizing the network architecture to reduce inference latency and energy consumption, while systematically benchmarking alternative lightweight designs across multiple performance dimensions; and (3) exploring multi-source data fusion by integrating sensor readings, engine operating parameters, and maintenance records into a unified decision framework, complemented with uncertainty estimation and anomaly detection mechanisms to improve diagnostic reliability and adaptability under complex conditions. These improvements are expected to substantially enhance diagnostic accuracy and operational robustness, extend the applicability of the model across diverse vessel types and scenarios, and provide a more reliable technical foundation for the development of intelligent ship maintenance and decision support systems.

In summary, this study validates the feasibility and application value of a lightweight image retrieval model in cylinder lubrication management. However, to achieve widespread applicability across platforms and operating conditions, continuous efforts are still required in generalization validation, resource adaptation, and multi-source data fusion. Only through such advancements can the model be better aligned with the practical demands of intelligent ship maintenance and provide more stable technical support for cylinder lubrication management in the era of low-sulfur fuels.

Author Contributions: Q.L.: conceptualization, methodology, software, investigation, formal analysis, data curation, visualization, and writing—original draft preparation. E.Z.: experiments, resources, data processing, validation, and writing—review and editing. G.Z.: conceptualization, methodology, supervision, project administration, funding acquisition, and writing—review and editing (corresponding author). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (NSFC51779136).

Data Availability Statement: The original contributions presented in this study are included in this article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors are grateful to the anonymous reviewers for their constructive comments that improved the presentation of this paper. We also thank colleagues for valuable discussions, and the technical staff of Shanghai Maritime University for their assistance with data acquisition.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rao, X.; Sheng, C.; Guo, Z.; Xu, C.; Dai, L.; Yuan, C. Corrosion behaviors of cylinder liner in marine diesel engines burning low-sulfur fuel oil: An experimental and molecular dynamics simulation study. *Tribol. Int.* **2022**, *171*, 107575. [\[CrossRef\]](#)
2. Wei, C.; Chen, X.-X.; Chen, L.; Li, P.; Liu, M. Effects of Pilot Oil Injection Timing on Combustion, Covariance and Knocking of a Natural Gas–Diesel Dual-Fuel Low-Speed Engine. *Pol. Marit. Res.* **2024**, *31*, 69–75.
3. Kamiński, W. Study to develop prediction models to determine the most suitable cylinder lubricating oil feed rates for marine slow-speed engines running on low sulphur fuel oil. *Measurement* **2025**, *247*, 116838. [\[CrossRef\]](#)
4. Temizer, İ. The combustion analysis and wear effect of biodiesel fuel used in a diesel engine. *Fuel* **2020**, *270*, 117571.
5. Hashmi, A.W.; Mali, H.S.; Meena, A.; Hashmi, M.F.; Bokde, N.D. Surface Characteristics Measurement Using Computer Vision: A Review. *Comput. Model. Eng. Sci.* **2023**, *135*, 917–1005. [\[CrossRef\]](#)
6. Zhao, E.; Zhang, G.; Li, Q.; Zhu, S. Impact of High-Concentration Biofuels on Cylinder Lubricating Oil Performance in Low-Speed Two-Stroke Marine Diesel Engines. *J. Mar. Sci. Eng.* **2025**, *13*, 1189. [\[CrossRef\]](#)
7. Operation Data and Maintenance Instruction Manual, MC Series Engines Edition, 2014. Editor B&W Man.
8. Dueholm, B.C.; Christiansen, J.D.C.; Endelt, B.; Kristensen, N. Injection rate of cylinder lubrication oil in large two-stroke marine diesel engines using a common rail lubrication system. *arXiv* **2023**, arXiv:2307.03408. [\[CrossRef\]](#)
9. Kamiński, W. Marine Slow-Speed Engines' Cylinder Oil Lubrication Feed Rate Optimization in Real Operational Conditions. *Energies* **2022**, *15*, 8378. [\[CrossRef\]](#)
10. Li, T.; Jiao, B.; Ma, X.; Lu, X.; Qiao, Z. Optimization study of oil-feed parameters to improve lubrication performance of two-stroke marine diesel engines. *Eng. Fail. Anal.* **2022**, *137*, 106356.
11. Bernhaed, F.; Berhard, S. Comprehensive multi-scale cylinder lubrication model for reciprocating piston compressors: From rigid-body dynamics to lubricant-flow simulation. *Tribol. Int.* **2023**, *178*, 108028.
12. He, Y.; Zhou, P.; Xie, L.; Zhang, J. Design and experimental development of a new electronically controlled cylinder lubrication system for the large two-stroke crosshead diesel engines. *Int. J. Engine Res.* **2019**, *20*, 986–1000. [\[CrossRef\]](#)
13. Zhao, B.; Hu, X.; Li, H.; Si, X.; Dong, Q.; Zhang, Z.; Zhang, B. A new approach for modeling and analysis of the lubricated piston skirt-cylinder system with multi-physics coupling. *Tribol. Int.* **2022**, *167*, 107381. [\[CrossRef\]](#)
14. Islam, R.; Martin, S. A Reliability Assessment of a Vessel's Main Propulsion Engine. *J. Mar. Sci. Eng.* **2025**, *13*, 1278. [\[CrossRef\]](#)
15. Cihan, Ö. Experimental and numerical investigation of the effect of fig seed oil methyl ester biodiesel blends on combustion characteristics and performance in a diesel engine. *Energy Rep.* **2021**, *7*, 5846–5856. [\[CrossRef\]](#)
16. Geng, P.; Hu, X.; Chang, X. Research on Combustion, Emissions, and Fault Diagnosis of Ternary Mixed Fuel Marine Diesel Engine. *J. Mar. Sci. Eng.* **2025**, *13*, 1561. [\[CrossRef\]](#)
17. Sagin, S.; Haichenia, O.; Karianskyi, S.; Kuropyatnyk, O.; Razinkin, R.; Sagin, A.; Volkov, O. Improving Green Shipping by Using Alternative Fuels in Ship Diesel Engines. *J. Mar. Sci. Eng.* **2025**, *13*, 589. [\[CrossRef\]](#)
18. Fu, C.; Lu, K.; Li, Q.; Xu, Y.; Gu, F.; Ball, A.D.; Zheng, Z. Physics-Based Modelling for On-Line Condition Monitoring of a Marine Engine System. *J. Mar. Sci. Eng.* **2023**, *11*, 1241. [\[CrossRef\]](#)
19. Jiao, B.; Ma, X.; Wang, Y.; Liu, Z.; Li, T.; Lu, X. Lubrication analysis of the piston ring for the two-stroke marine diesel engine based on modified radial tension. *Int. J. Engine Res.* **2023**, *25*, 200–213. [\[CrossRef\]](#)
20. Lyu, X.; Hu, J.; Wang, Y.; Sheng, J.; Ma, X.; Li, T.; Ge, C.; Lu, X. Effect of temperature on tribofilm growth and the lubrication of the piston ring-cylinder liner system in two-stroke marine engines. *Friction* **2024**, *12*, 1858–1881. [\[CrossRef\]](#)
21. Lyu, B.; Meng, X.; Zhang, R.; Wen, C. A Deterministic Contact Evolution and Scuffing Failure Analysis Considering Lubrication Deterioration Due to Temperature Rise under Heavy Loads. *Eng. Fail. Anal.* **2021**, *123*, 105276. [\[CrossRef\]](#)
22. Liu, Y.; Chan, F.; Wang, D. *Computer Vision and Deep Learning: Using MATLAB and Python as Tools*; Publishing House of Electronics Industry: Beijing, China, 2019.
23. Kim, S.-H.; Kim, T.-G.; Lee, J.; Song, H.-K.; Moon, H.; Chun, C.-J. Deep hybrid model for fault diagnosis of ship's main engine. *J. Mar. Sci. Eng.* **2025**, *13*, 1398. [\[CrossRef\]](#)
24. Yang, A.-M.; Zhi, J.-M.; Yang, K.; Wang, J.-H.; Xue, T. Computer Vision Technology Based on Sensor Data and Hybrid Deep Learning for Security Detection of Blast Furnace Bearing. *IEEE Sens. J.* **2021**, *21*, 24982–24992. [\[CrossRef\]](#)

25. Zhang, G.C.; Gao, R.H.; Zhang, C.C.; Li, Y.F. Computer vision-based lubrication management of ship main engine cylinders. *Navig. China* **2024**, *47*, 40–47.
26. Gao, X.; Qian, Y.; Loomes, M.; Barn, B.; Comley, R.; Chapman, A.; Rix, J.; Hui, R.; Tian, Z. Retrieval of 3D medical images via their texture features. *Int. J. Adv. Softw.* **2011**, *4*, 499–509.
27. Tan, M.; Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv* **2019**. [[CrossRef](#)]
28. Liu, S.; Huang, D.; Wang, Y. *Receptive Field Block Net for Accurate and Fast Object Detection*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Computer Vision—ECCV 2018; Springer: Cham, Switzerland, 2018; Volume 11215.
29. Mehta, S.; Rastegari, M. MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv* **2021**. [[CrossRef](#)]
30. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: *Convolutional Block Attention Module*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Computer Vision—ECCV 2018; Springer: Cham, Switzerland, 2018; Volume 11211.
31. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems (NeurIPS)*. 2012, pp. 1–9. Available online: <https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html> (accessed on 10 October 2025).
32. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *Computer Science. arXiv* **2014**. [[CrossRef](#)]
33. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A.; Liu, W.; et al. Going deeper with convolutions. In *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
34. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.