Computers & Geosciences 207 (2026) 106059

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

Weakly supervised semantic segmentation of microscopic carbonates on marginal devices

Keran Li [a,*,1], Yujie Gao [b,1], Yingjie Ma [b,1], Chengkun Li [b], Junjie Ye [b], Hao Yu [b], Yiming Xu [b], Dongyu Zheng [c,*], Ardiansyah Koeshidayatullah [d,*]

[a] State Key Laboratory of Mineral Deposit Research, School of Earth Sciences and Engineering, Nanjing University, Nanjing 210023, China
[b] College of Computer Science and Cyber Security, Chengdu university of Technology, Chengdu, 610059, China
[c] State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation and Key Laboratory of Deep-time Geography and Environment Reconstruction and Applications, MNR and Institute of Sedimentary Geology, Chengdu University of Technology, Chengdu 610059, China
[d] Department of Geosciences, College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

ARTICLE INFO

Keywords:
Weakly supervised learning
Microscopic carbonates images
Marginal devices

ABSTRACT

Microscopic analysis is the cornerstone to uncover petrological and mineralogical characteristics of carbonate rocks. In addition, such information is critical for precise identification of carbonate microfacies and diagenetic evolution. This type of information is important, but relies too much on manual experience, which is time-consuming and laborious. Recently, several successful deep learning models showed great potential in the identification process. However, current deep learning models have typically complex model architectures greatly hinder the deployment-inference in practical and lightweight environments. To overcome the difficulty of deep learning models in reasoning in actual edge scenes, a three-stage segmentation method by weakly supervised learning was proposed. The approach embeds class activation mapping (CAM), grey level co-occurrence matrix (GLCM), and knowledge distillation (KD) modules to achieve attention transfer to the lightweight network (CamNet). Furthermore, based on the performance of the model algorithm and application requirements, a lightweight carbonate thin section image-assistant recognition system has been developed. Through ingenious control flow design, this system achieves an effective balance between runtime latency and resource consumption, demonstrating superior performance metrics. Experimental results indicate that CamNet's total parameter count is only 800k. When deployed in embedded systems, CamNet achieves an inference speed of 6.87 fps. Our successful development verifies the efficiency and practicality in marginal devices.

1. Introduction

Carbonate sediments, as fundamental components of the Earth's surface, encapsulate the evolutionary history of the Earth system. Sediments record pivotal events such as the origins and mass extinctions of life, climate fluctuations over the past billion years, and tectonic shifts (Bathurst, 1972). Beyond the scientific significance (Bosence, 2002), analyzing carbonates in both academic and industry are essential. Traditional carbonates analysis deeply relies on thin-section images observations with polarizing microscopes. The microscopic carbonate images analysis depends heavily on observer experience and cause ambiguity in observation quality. Meanwhile, traditional approach is time-intensive. The efficiency and accuracy of traditional empirical image analysis greatly hinders the accessibility and reliability of carbonate rocks analysis.

These inherent limitations of manual analysis are further exacerbated in practical scenarios, where carbonate thin-section analysis such as field geological surveys, well-site logging, and real-time teaching demonstrations imposes stringent requirements on device portability and processing speed. In fieldwork or drilling site, researchers and engineers often lack access to high-performance computing clusters and bulky equipment cannot be deployed near outcrops or drilling platforms. The resource-limited situation necessitates lightweight models that can run on marginal devices (e.g., embedded boards integrated with microscopes). Additionally, the volume of microscopic image data generated by continuous sampling (typically 10 to 50 GB per field trip) makes cloud-based processing impractical due to latency and bandwidth constraints. Moreover, real-time inference (at least 5 fps) is critical for on-site decision-making, such as identifying key lithofacies

* Corresponding authors.
E-mail addresses: keranli98@outlook.com (K. Li), zhengdongyu@cdut.edu.cn (D. Zheng), a.koeshidayatullah@kfupm.edu.sa (A. Koeshidayatullah).
[1] Keran Li, Yujie Gao and Yingjie Ma were equally contribution

https://doi.org/10.1016/j.cageo.2025.106059
Received 8 March 2025; Received in revised form 23 August 2025; Accepted 22 September 2025
Available online 27 September 2025

immediately during thin-section preparation, which directly influences subsequent sampling strategies. Thus, balancing segmentation accuracy with efficiency on marginal devices addresses a long-standing gap in petrological analysis workflows.

To address these challenges from both the inefficiency of manual analysis and the demands of practical deployment, researchers have turned to computer vision or deep learning for automation. In traditional microscopic carbonate images observation, experienced observers play the role of extracting and summarizing information such as texture, color, and morphology from carbonates. The observer simultaneously re-prioritized the importance of the image features. The speed and accuracy actually depend on human's completeness of feature extraction and the accuracy of feature mapping mechanisms. Theoretically, the complex process can be replaced by computer vision or deep learning methods. Recent advances in deep learning have significantly improved semantic segmentation techniques, which are possible and vital for analyzing microscopic carbonate images.

In the domain of petrological thin-section analysis, several computer vision (see the summary from Froute et al., 2025; da Rocha et al., 2025) or deep learning models have been explored while each faces distinct limitations. Initially, convolutional neural networks (CNNs) revolutionized feature extraction in petroligical image analysis. Yamada and Di Santo (2022) applied Mask R-CNN with Feature Pyramid Networks (FPN) to separate the foreground and background of petrological particles and achieve pixel level instance segmentation. The CNNs was also successfully adapted in the microscopic carbonate images analysis (Koeshidayatullah et al., 2020). However, CNNs would cause massive information redundancy and calculation time increasing. To improve the computing efficiency, the end-to-end segmenting U-Net (Ronneberger et al., 2015a; Malik et al., 2022; Chen et al., 2020a; Froute et al., 2025) model was adopted in sandstone and mudstone image analysis. Although U-Net has significantly improved the efficiency of segmentation, U-Net's weak perception of boundaries leads to poor performance in complex petrological image segmentation. To enhance the global information representation, the transformer-based models, had been utilized in petrological segmentation (Vaswani et al., 2017; Dosovitskiy et al., 2020; Kirillov et al., 2023a; Chen et al., 2021). Transformer-based model with attention mechanism indicated advances in for petrological images analysis (Koeshidayatullah et al., 2022; Ferreira-Chacua and Koeshidayatullah, 2023; Ma et al., 2023a,b; Koeshidayatullah, 2023). However, high model complexity requires large number of high-quality annotated images for training. Compared with the commonly used real-world annotated datasets such as VOC, COCO, and CIFAR in deep learning and the datasets of annotated thin-section images, the petrological image dataset is smaller in data size and poor semantic annotation. Pixel level image annotation is almost impossible to achieve in carbonate images, especially on non skeletal carbonates.

To overcome the bottleneck of scarce labeled data, weakly supervised semantic segmentation (WSSS) has emerged as a promising approach, though it too has limitations in handling complex carbonate textures. The WSSS based on class-level labels is one potential solution in overcoming the lack of labels (Ahn and Kwak, 2018). The usual augmentations in WSSS include class activation maps (CAMs) (Zhou et al., 2015) and random walk (LOV'ASZ et al., 1993) to propagate sparse localization information to recover target contours.

Beyond the challenges of model design and training, an even more critical gap exists in the deployment of these models for real-world petrological applications. Almost no research has been conducted involving model deployment in thin-section image segmentation. High-precision segmentation models often have large parameter sets and poses significant challenges for edge deployment. For general application tasks, two primary strategies are employed, containing hardware and software optimizations for embedded edge devices. The hardware methods include FPGA design frameworks for CNN sparsity and acceleration (Li et al., 2017), the OMNI framework for integrated hardware and software optimizations (Liang et al., 2021), and Cryptensor for accelerating CNNs and polynomial convolutions (See et al., 2023). The software methods include model compression techniques like pruning, quantization, and knowledge distillation. Liu et al. (2017) pioneered model pruning, which automatically trims unimportant channels during training. Model quantization has seen substantial development from the initial proposal by Jacob et al. (2017) to the adoption of TensorRT (Zhou et al., 2023a). Knowledge distillation has also been instrumental in model compression (Zhou et al., 2023b; Kanwal et al., 2023; Gou et al., 2023). While models for marginal devices have been widely applied in various fields, real-time deployment of deep learning models for petrological image segmentation on marginal devices remains understudied. The complexity of petrological images means that model compression could adversely affect accuracy and segmentation quality, hindering subsequent analysis. Even with compression the model size meets the deployment requirements, displaying on marginal devices is challenging. Because in practical scenarios, it is necessary to consider how to improve the display effect on hardware. The combination of hardware display delay and inference time delay determines the display effect of deep learning models on marginal devices. The traditional affinity matrix estimation methods in computer vision (gray-level co-occurrence matrix, GLCM and secondary information extraction) has already proven effective in complex segmentation extracts image texture (De Siqueira et al., 2013; Partio et al., 2002; Mohanaiah et al., 2013). Carbonate thin section images have more complex geometric features. The applicability of traditional lightweight model enhancement methods also needs to be verified.

To address these interconnected challenges—scarce labels, complex textures, and marginal deployment, we propose a novel framework called Feature Localization Knowledge Distillation (FLKD) for efficient generation of class activation maps (Gou et al. (2021), CAMs) using weakly supervised semantic pseudo-label generators (Fig. 1). This framework leverages CAMs generated by large deep learning models (like ResNet101 (He et al., 2015) in this study) as constraints, guiding lightweight deep learning models (like the novel light model in this study, CamNet) to produce sparse localization information thus allowing for direct use of pseudo-labels as segmentation results and supporting auxiliary recognition tasks. Additionally, a lightweight carbonate thin-section image auxiliary recognition system has been developed, which integrates hardware facilities such as a microscopic camera, high-definition display, and embedded development board. To enhance the processing efficiency of video streams, we propose a clever workflow combining the Video Frame Difference Algorithm (VFDA) (Huang and Aggarwal, 1981), Frame Rate Reduction and Frequency Interpolation (FRRFI) (Ling et al., 2008), and asynchronous processing techniques for a composite control flow design.

In summary, this work makes three key contributions that address the identified gaps: (1) A novel CAM–GLCM–KD framework with weakly supervised learning to handle high-complexity carbonate images; (2) CamNet, a lightweight network (800k parameters) achieving 6.87 fps inference on edge devices, five times faster than existing models; (3) The first marginal devices to smoothly segment carbonate thin-section images on marginal devices.

2. Related work

2.1. Petrological thin section image segmentation

Initial petrological thin-section image segmentation research focused on traditional computer vision methods such as threshold cutting and edge detection (Starkey and Samantaray, 1993; Zhou et al., 2004; Zhou and Ren, 2012; Andrä et al., 2013). These methods are restricted to specific rock types due to petrological complexity and often fail to generalize across diverse carbonate textures and morphologies. Advances in feature extraction, leveraging texture (Chen et al., 2020b), shape (Iwaszenko and Nurzynska, 2019), color (Ren et al.,

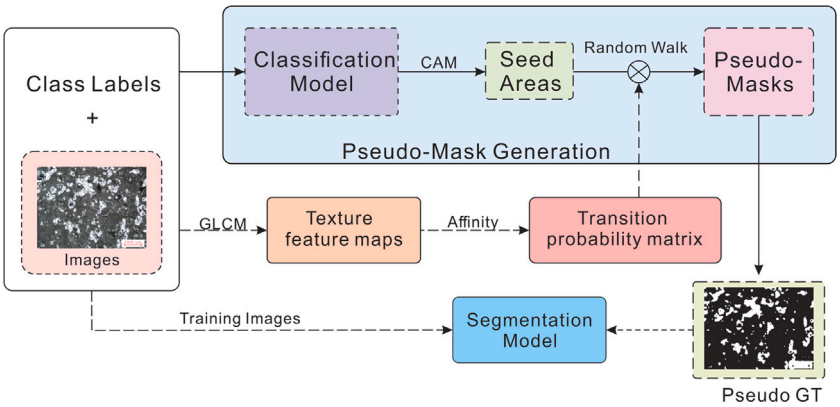

Fig. 1. Illustration for weakly supervised semantic segmentation to fuse the vision features and generate ground truth (GT).

2021), and hybrid features (Sun et al., 2019), have limitedly enhanced segmentation but remain constrained by handcrafted feature design.

The adoption of deep learning has revolutionized petrological image analysis, with deep learning-based computer vision methods becoming essential for rock image recognition (Chen et al., 2020b; Iwaszenko and Nurzynska, 2019; Ren et al., 2021; Sun et al., 2019; Manzoor et al., 2023). Convolutional neural networks (CNNs) have been pivotal for texture extraction in petrological segmentation (Niu et al., 2020; Koeshidayatullah et al., 2020; Manzoor et al., 2023), while explainable artificial intelligence (XAI) techniques (Mamalakis et al., 2021, 2022; Zheng et al., 2023) have improved model interpretability.

Despite fully supervised CNN-based methods (e.g., U-Net, Ronneberger et al., 2015b, DeepLabv3+, Chen et al., 2017) succeeded in general semantic segmentation, carbonate thin-section analysis as a more specific task encounters unique obstacles. Specifically, carbonate thin-section deep learning model reliance on dense pixel annotations clashes with the inherent properties of carbonate rocks: (1) The labor intensity of labeling is exacerbated by the need to distinguish subtle textural variations (e.g., between micrite and microspar), escalating annotation time to several hours per one image for expert geologists. (2) Ambiguous boundaries in non-skeletal carbonates (e.g., homogeneous micritic limestone) lead to irreconcilable inter-annotator discrepancies, undermining label reliability. (3) Lithofacies-specific features (e.g., oolite concentric layers vs. dolomite crystal cleavage) result in poor cross-class generalization, as models trained on one lithology fail to adapt to others without retraining.

To address data scarcity, recent studies have explored supervised methods with incomplete labeling (Li et al., 2023b) and data augmentation via diffusion models or the segment anything model (SAM) (Ma et al., 2023a,b). This underscores the shift toward minimal-labeling strategies, with weakly supervised (Zhou, 2018) and unsupervised learning (Xie et al., 2015) emerging as dominant trends to reduce reliance on manual annotation.

2.2. Weakly Supervised Semantic Segmentation (WSSS)

Weakly supervised semantic segmentation (WSSS) has emerged as a key solution to mitigate the shortage of labeled training data, but its application to carbonate images remains challenging, which is characterized by high textural complexity.

Early WSSS approaches relied on weak supervision signals such as bounding boxes (Khoreva et al., 2016), scribbles (Lin et al., 2016), or points (Bearman et al., 2015). While these reduced reliance on dense annotations, weak supervision signals still required manual input, which failed to simplify the annotation of complex petrological semantics in carbonate images. This limitation highlighted the need for more scalable supervision strategies.

Image-level class labels emerged as a more practical alternative, as they are easier to annotate at scale. However, initial attempts to train segmentation models directly from Image-level class labels (Pinheiro and Collobert, 2014; Papandreou et al., 2015) lost critical pixel-scale information, which is particularly problematic for carbonate segmentation (e.g. fine details like crystal boundaries or fossil fragments are essential). The gap between class-level labels and pixel-level segmentation demands a bridge between high-level categories and low-level spatial features.

Class activation maps (CAMs) (Zhou et al., 2015; Selvaraju et al., 2016; Desai and Ramaswamy, 2020) addressed the bridge between categories and spatial features by visualizing object regions and generating sparse localization cues and effectively translated image-level labels into pseudo-pixel annotations. This breakthrough enabled WSSS to retain spatial information while leveraging lightweight supervision.

Subsequent research further refined pseudo-label generation, including enriching contextual features (Xu et al., 2020a; Li et al., 2020), expanding receptive fields via dilated convolutions (Wei et al., 2018), integrating global contexts with graph convolution (Yao et al., 2021), and developing pixel clustering methods (Huang et al., 2018; Wang et al., 2019; Ahn and Kwak, 2018; Ru et al., 2022). These advances improved pseudo-label quality for general images but still struggle with carbonate-specific challenges.

Traditional CAMs and random walk techniques (LOV'ASZ et al., 1993) are core to many WSSS pipelines, which fail to capture the fine-grained features of carbonate images, such as subtle crystal edges, fragmented fossils, or complex textural variations. This mismatch underscores the need for WSSS frameworks tailored to the unique characteristics of petrological imagery.

2.3. Lightweight deployment of deep learning models

The deployment of deep learning models in petrology remains underexplored, with a critical challenge being the adaptation of complex models to resource-constrained edge devices.

A primary obstacle lies in the large parameter sets of high-precision segmentation models, which inherently hinder their deployment on edge devices. To address this, existing research has focused on two main directions: hardware-level optimizations and software-level compression techniques. Hardware solutions include FPGA frameworks tailored for CNN sparsity (Li et al., 2017), OMNI for integrated hardware–software co-optimization (Liang et al., 2021), and Cryptensor for accelerating convolution operations (See et al., 2023). On the software side, methods such as pruning (Liu et al., 2017), quantization (Jacob et al., 2017; Zhou et al., 2023a), and knowledge distillation (KD) (Hinton et al., 2015) have been widely adopted to reduce model size and computational demands.

Among these, knowledge distillation (KD) has proven particularly effective in compressing models while preserving performance. By training a smaller "student" network to mimic the behavior of a larger

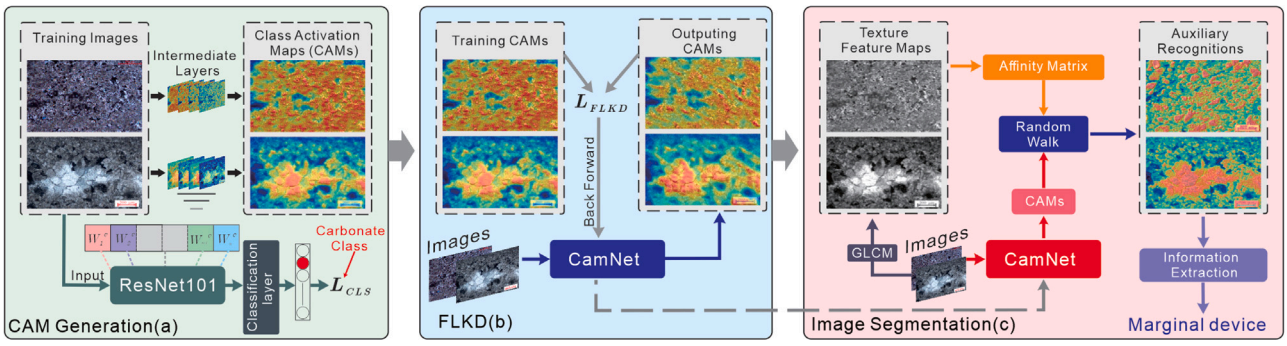

Fig. 2. The three-stages workflow. (a) CAM Generation; (b) framework of Feature Localization Knowledge Distillation (FLKD); (c) Image segmentation and deployment.

"teacher" network (Yim et al., 2017; Zagoruyko and Komodakis, 2016; Park et al., 2019), KD retains critical patterns, with strategies like attention map mimicry (Zagoruyko and Komodakis, 2016) further enhancing the student's accuracy.

However, general compression methods often struggle with complex petrological images. The loss of critical texture details such as subtle lithofacies boundaries, crystal textures, or fossil fragments during compression directly degrades segmentation accuracy. Because texture features are pivotal for petrological analysis. The mismatch between generic lightweight strategies and the unique rock image segmentation demands underscores the need for tailored deployment frameworks in petrology.

3. Methods

3.1. Overview of the CAM-GLCM-KD framework

To address the challenges of weakly supervised segmentation and lightweight deployment in carbonate image analysis, this study proposes a novel framework named CAM–GLCM–KD. This section details its core components, implementation steps, experimental setup, and deployment strategy, organized to clarify the logical flow from algorithm design to practical application.

The CAM–GLCM–KD framework is proposed to address the challenges of weakly supervised semantic segmentation of microscopic carbonate images on resource-constrained marginal devices. It integrates three core modules, Class Activation Mapping (CAM), Grey Level Co-occurrence Matrix (GLCM), and Knowledge Distillation (KD)—in a three-stage workflow to achieve accurate and efficient segmentation (Fig. 2).

As shown in Fig. 2, the first stage involves generating CAMs using a pre-trained deep neural network, which serves as pseudo-labels for weakly supervised learning. The second stage employs Feature Localization Knowledge Distillation (FLKD) to transfer feature localization capabilities from a complex teacher model to a lightweight student model (CamNet). The third stage enhances the segmentation results by fusing texture features extracted via GLCM with CAM-derived spatial information, followed by semantic propagation through random walk. This integrated framework ensures a balance between segmentation accuracy and model efficiency, enabling deployment on marginal devices (Wu et al., 2019).

3.2. Pseudo-label creation

CAMs are generated to convert image-level class labels into pixel-level spatial localization cues, providing the foundation for weakly supervised training. A well-trained model can fully and accurately represent the labels (carbonate classification information as labels in datasets) in original petrological images by heat maps (Gou et al., 2021). Fully represented heat maps (CAMs) from a well-trained model is the basis of consequent lightwise processing. After comparison, the ResNet101 deep architecture was adapted. The ResNet-series models address the vanishing gradient issue with residual blocks, enhance feature capture and retain robustness in various visional tasks (Wu et al., 2019).

3.2.1. Teacher model

ResNet101 (He et al., 2015) is selected as the teacher model for CAM generation due to its robust feature extraction capability by the residual block design that mitigates the vanishing gradient problem. The residual function in ResNet101 is defined as:

$$F(x, W_i) = x + F(x, W_i') \tag{1}$$

where x denotes the input carbonate thin section image, W_i represents the weights of the ith layer, and $F(x, W_i')$ is the residual mapping computed as:

$$F(x, W_i') = W_{i-1} * \sigma(W_i * x) \tag{2}$$

Here, σ denotes the sigmoid activation function, and $*$ represents the convolution operation.

3.2.2. Grad-CAM calculation

Grad-CAM is utilized to generate class-discriminative heatmaps without altering the network architecture (Selvaraju et al., 2016). For a given class c, the Grad-CAM output is calculated as:

$$L^c_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right) \tag{3}$$

where A^k is the kth activation heatmap, and α_k^c is the importance weight of the kth feature map for class c, derived from global average pooling on feature map gradients:

$$\alpha_k^c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \frac{\partial S_c}{\partial A_{ij}^k} \tag{4}$$

In Eq. (4), H and W are the height and width of the feature map A^k, and ∂S_c is the pre-softmax score for class c. These CAMs convert class label information into spatial distributions, serving as pseudo-labels for subsequent distillation.

3.3. Feature localization knowledge distillation in CamNet

Following (Zheng et al., 2022), FLKD is designed to transfer the feature localization ability of the teacher model (ResNet101) to a lightweight student model (CamNet), ensuring efficient deployment on marginal devices.

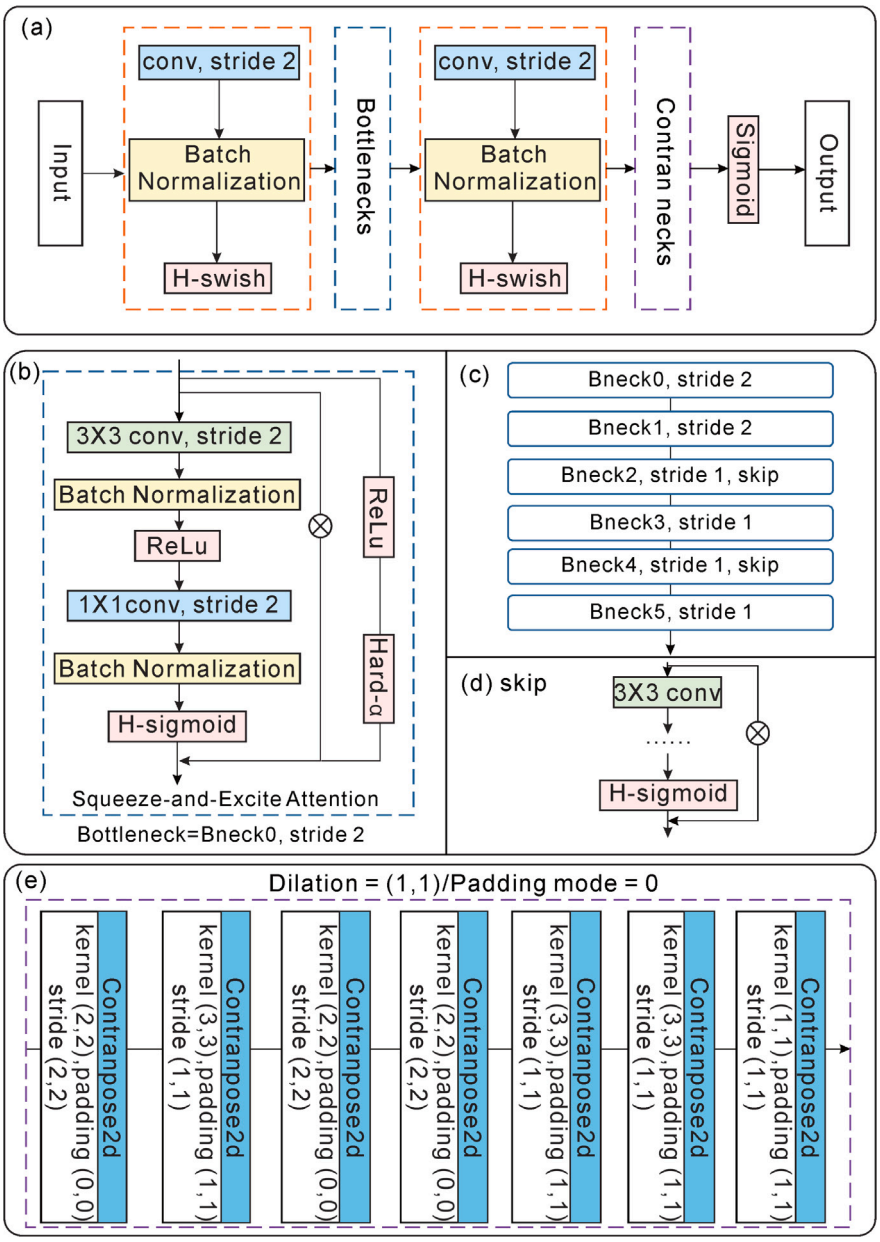

Fig. 3. CamNet Model Structure. (a) Overall architecture. (b) SE attention module. (c) and (d) BottleNecks. (e) Added deconvolution.

3.3.1. FLKD loss function

The FLKD loss function minimizes the discrepancy between the student and teacher models' outputs, enabling the student to approximate the teacher's feature localization (Zheng et al., 2022). It is defined as:

$$L_{\mathrm{FLKD}} = \frac{1}{N}\sum_{i=1}^{N} \|S(x_i) - T(x_i)\|^2 \tag{5}$$

where $S(x_i)$ and $T(x_i)$ are the outputs of the student and teacher models for the ith carbonate category, respectively, and N is the total number of categories.

3.3.2. CamNet architecture

As shown in Fig. 3, CamNet is developed based on MobileNetV3_small (Howard et al., 2017) with targeted optimizations for lightweight deployment. Firstly, CamNet aims to reduce bottleneck modules by approximately 30% to lower parameter count. Meanwhile, CamNet integrates deconvolution layers for precise CAM mapping onto the original image. Lastly, CamNet streamlines convolutional channels and attention modules. All improvements result a super-light model size of 800k parameters.

3.4. Lightweight model enhancement

Considering a neighborhood around each pixel in microscopic images represents the depositional fabrics, the segmentation performance can be improved by traditional computer-visional methods. The improvements would ease the deployment of models on marginal devices and enhance the fluidity in displaying. To refine segmentation performance, Grey Level Co-occurrence Matrix (GLCM, Eichkitz et al., 2013) is firstly employed to extract texture features. Then the secondary information metrics are introduced to characterize depositional fabrics. GLCM analyzes spatial pixel relationships to compute an affinity matrix, facilitating semantic propagation via random walk:

$$GLC_o^{(k+1)} = \Lambda \cdot GLC_o^{(k)} \tag{6}$$

where $GLC_o^{(k+1)}$ is the updated GLCM at iteration $k+1$, and Λ is the transfer matrix estimated from texture similarity. High-confidence regions in the diffused CAMs are selected using a threshold τ.

For GLCM implementation, geometry parameters are systematically optimized based on the scale and directional characteristics of carbonate textures. Specifically, we employed multiple distance metrics (d=10,100,500 pixels) and angular orientations ($\theta = 0, 45, 90, 135$ degrees) to capture multi-scale textural dependencies, as carbonate fabrics often exhibit layered or granular structures at sub-pixel to pixel-scale resolutions. The optimal parameter set was determined via cross-validation on a dataset of 500 carbonate thin-section images by prioritizing metrics that maximized texture discriminability (e.g., contrast, dissimilarity, homogeneity) for key lithofacies (e.g., oolitic, peloidal, and crystalline carbonates). This multi-directional approach ensures robust capture of both isotropic (e.g., granular) and anisotropic (e.g., laminated) textures prevalent in carbonate sediments.

The secondary information extraction contains two metrics are introduced to enhance segmentation performance (Riloff et al., 1993; Kreuzthaler et al., 2015; Wang et al., 2018). For the main carbonate deposits in displaying, composite rate (CR) is firstly calculated by:

$$CR = \frac{\mathrm{Pix}_{\mathrm{Mineral}}}{\mathrm{Pix}_{\mathrm{whole}}} \times 100\% \tag{7}$$

Then the image entropy ($H(I)$) for geological richness assessment is calculated to enhance the detail displaying by:

$$H(I) = -\sum_{i=0}^{L-1} p(i)\log(p(i)) \tag{8}$$

where $\mathrm{Pix}_{\mathrm{Mineral}}$ and $\mathrm{Pix}_{\mathrm{whole}}$ are the number of mineral pixels and total pixels, respectively. $p(i)$ is the probability of pixel value i, and L is the number of grey levels.

At last, the enhanced displaying ($Imp(I)$) can be defined as:

$$\mathrm{Imp}(I) = \sum_{i=I}^{j=J} H(I)\cdot CR \tag{9}$$

3.5. Evaluation metrics

To comprehensively assess the effectiveness of the CAM–GLCM–KD framework, a set of complementary metrics is designed to quantify three core aspects of performance: model efficiency, segmentation accuracy, and alignment between the lightweight student model and the complex teacher model. These metrics are tailored to the unique challenges of carbonate thin section segmentation—including limited annotation availability, high texture complexity, and the need for deployment on resource-constrained edge devices and draw on both traditional computer vision benchmarks and domain-specific considerations.

3.5.1. Classification and segmentation metrics

Accuracy metrics are crucial for evaluating how effectively the framework captures the fine-grained features of carbonate rocks, such as crystal boundaries and fossil fragments, and aligns with domain expertise. Given the challenge of generating dense pixel-level annotations for complex petrological images, a fuzzy sampling mechanism is employed to balance rigor and feasibility. This mechanism involves annotating key semantic points (e.g., mineral particles, background regions) across the dataset and using random uniform sampling to mitigate bias from non-stationary rock textures. For each sampled point, targets are labeled as positive (Ps, representing mineral features) or negative (Ns, representing non-mineral regions). Based on these labels, several core metrics are used to assess the model's performance:

Precision (P), recall (R), specificity (S), and the F1-score are fundamental for evaluating the model's performance in handling ambiguous boundaries and textural variations in carbonate thin sections.

$$P = \frac{TP}{TP+FP}, \quad R = \frac{TP}{TP+FN}, \quad S = \frac{TN}{TN+FP}, \quad F1 = 2\times\frac{P\times R}{P+R} \tag{10}$$

Here, precision measures the proportion of correctly identified positive samples among all positive predictions. Recall reflects the model's ability to capture all actual positive samples, while specificity evaluates its performance in correctly identifying negative samples. The F1-score balances precision and recall, avoiding overemphasis on either metric. TP (true positives) represents correctly segmented positive samples, FP (false positives) are non-target regions misclassified as positive, TN (true negatives) are correctly excluded negative samples, and FN (false negatives) are positive samples missed by the model.

The Top1 - error quantifies the proportion of samples where the model's top-ranked predicted class differs from the true class, calculated as Top1 - error $= 1 - \frac{\text{Correct Predictions}}{\text{Total Samples}}$. Accuracy, a fundamental metric, measures the overall proportion of correct predictions: Accuracy $= \frac{TP+TN}{TP+FP+FN+TN}$. The Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) evaluates the model's class discrimination across all possible thresholds. The ROC curve plots the true positive rate (recall) against the false positive rate ($FPR = \frac{FP}{FP+TN}$), and the AUC, ranging from 0 to 1, indicates model performance, with 1 representing a perfect classifier and 0.5 a random one, calculated as $\mathrm{AUC} = \int_0^1 \mathrm{TPR}(FPR)\, d(FPR)$.

Due to the fact that F1-score, AUC, Top1-error, and accuracy can cover the evaluation of precision, recall, and specificity, this study only selected F1-score, AUC, Top1-error, and accuracy as evaluation indicators when comparing teacher models. For CamNet based on key points annotation evaluation under WSSS, the potential annotation imbalance leads to abnormal outliers in AUC and accuracy. Therefore, F1-score, specificity, recall, and precision are used to evaluate CamNet under the WSSS framework to highlight the segmentation ability of the model in the absence of labels. The key points strategy concludes two steps. Let $\mathcal{I} \subset \mathbb{R}^{H\times W}$ denote a carbonate thin-section image. Semantic ambiguity and texture non-stationarity preclude dense ground-truth labeling. Instead, a sparse set of key points $\mathcal{K} = \{(x_i, y_i)\}_{i=1}^{N}$ is selected such that each point carries maximal petrological information. On the first, semantic saliency map would be sampled by a gradient-based saliency function

$$S(x,y) = \|\nabla \mathcal{I}(x,y)\|_2 + \lambda_{\mathrm{tex}}\, T_{\mathrm{LBP}}(x,y)$$

combines intensity gradient and local binary pattern texture T_{LBP}. Local maxima of S define candidate points $\mathcal{C} \subset \mathcal{I}$.

In the second, to counter spatial bias, a jittered grid is imposed. For grid node (x, y), the final key point is

$$(x', y') = \Big(\mathrm{clip}(x+\Delta_x,\, 0,\, W-1),\ \mathrm{clip}(y+\Delta_y,\, 0,\, H-1)\Big),$$

where

$$\Delta_x, \Delta_y \sim \mathcal{U}\Big(-\tfrac{I}{2}, \tfrac{I}{2}\Big), \quad I = \left\lfloor \sqrt{\frac{HW}{N}} \right\rfloor.$$

The set $\mathcal{K}$ consists of the top-N most salient jittered points.

In addition, in the zero-shot experiment, mean intersection over union (mIoU, Rezatofighi et al., 2019), mean pixel accuracy (MPA, Li et al., 2019) and Dice coefficient (Dice, Shamir et al., 2019) were also introduced to evaluate the mask generation effect.

These metrics collectively provide a comprehensive assessment of the model's performance in classifying and segmenting carbonate rock features, accounting for both prediction correctness and the ability to handle geological heterogeneity.

3.5.2. Model efficiency metrics

Efficiency is critical for ensuring the framework's practicality on marginal devices, where computational resources (memory, processing power) are limited. Two key metrics capture this dimension:

Model parameter size (MPS) quantifies the structural complexity of the model by summing the total number of parameters across all layers. Mathematically, it is defined as:

$$\mathrm{MPS} = \sum_{i=1}^{n} P_i \tag{11}$$

where n denotes the total number of layers in the model, and P_i represents the number of parameters in the ith layer. A smaller MPS indicates a more lightweight model, directly supporting deployment on edge devices with restricted storage and memory.

Inference time (IT) measures the average duration required for the model to process a single input image and generate a segmentation result. IT can be calculated as:

$$\mathrm{IT} = t_{\mathrm{end}} - t_{\mathrm{start}} \tag{12}$$

where t_{start} and t_{end} are the timestamps marking the initiation and completion of the inference process, respectively. For real-time applications such as on-site petrological analysis using polarizing microscopes, IT must be minimized to ensure smooth interaction.

3.5.3. Alignment metrics

To ensure the lightweight student model (CamNet) retains the feature localization capabilities of the teacher model (ResNet101). The structural similarity index (SSIM, Brunet et al., 2011) is employed to quantify spatial consistency between their outputs. SSIM is defined as:

$$\mathrm{SSIM} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{13}$$

where x and y represent two images (e.g., CAMs from the teacher and student models), μ_x and μ_y are their mean luminance values, σ_x^2 and σ_y^2 are their variances, σ_{xy} is their covariance, and c_1, c_2 are small constants to stabilize division

In this framework, SSIM is computed in two key contexts, including $SSIM\text{-}cam$ (comparing CAMs generated by the teacher and student models) and $SSIM\text{-}seg$ (comparing their final segmentation results). A high SSIM value (close to 1) indicates that CamNet successfully inherits the teacher model's ability to localize critical petrological features, validating the effectiveness of the FLKD distillation process.

Together, these metrics form a holistic evaluation system. Accuracy metrics validate its ability to capture meaningful geological features. Efficiency metrics ensure the model is lightweight enough for edge deployment and alignment metrics confirm that the compression does not come at the cost of losing the teacher model's discriminative power. This multi-faceted assessment ensures the CAM–GLCM–KD framework meets both computational and domain-specific requirements for carbonate thin section analysis.

3.6. Dataset and deployment

3.6.1. Dataset details

The experiment utilizes a series of open-source dataset of carbonate microscopic images (Li et al., 2023a; Xin and Chen, 2020; Qian et al., 2020; Chang et al., 2020; Xu et al., 2020b; Li and Hu, 2020; Han and Hu, 2020; Ma et al., 2020; Chai et al., 2020; Lai et al., 2020; Zhang and Hu, 2020; Qi et al., 2020) and the carbonate world (https://carbonateworld.com/). After carefully checking the images with corresponding descriptions in excel files, all microscopic images were re-labeled as classification tasks following the format of ImageNet dataset (Deng et al., 2009). The processed labeled carbonate dataset consists of twenty-two categories of carbonate sediments (Fig. 4). All images have a resolution of 1280 × 960 pixels with JPG, JPEG and PNG formats.

3.6.2. Embedded system

The deployment of deep learning models typically occurs on CPUs or GPUs, yet in real-world scientific research, educational, and industrial settings, such as when using polarizing microscopes (Blanco-Filgueira et al., 2019), high-performance computing resources are scarce. This necessitates the deployment of these models on edge devices for microscopic image applications, which in turn requires a meticulous optimization of model size, performance, and hardware conditions. In this study, we initiate the simulation of deep learning model deployment on edge devices to explore the performance of CamNet, laying the groundwork for future applications in earth science. The system workflow is designed to meet the stringent requirements of efficient real-time and fit video processing. Commencing with image capture by an electron microscope camera, the images are then processed on the core development board. Branching operations, as illustrated in Fig. 5a–b, are employed to adapt to dynamic changes, significantly enhancing both processing efficiency and accuracy. In dynamic scenarios, upon detecting scene changes, the system executes smoothing operations to reduce noise, followed by real-time image display. For static conditions, the system generates auxiliary thermal and segmentation images, conducts secondary information extraction and smoothing, and finally presents the final display. The hardware architecture of the embedded system, depicted in Fig. 5a–b, comprises a microscope electronic camera, high-definition monitor, embedded core development board, and TF card storage. The camera captures rock thin section images under polarized microscopy, while the monitor provides real-time display of images and processing results. The Nvidia Jetson Nano, selected as the core development board with its Linux operating system environment, supports deep learning model deployment. Leveraging GPU acceleration and optimized libraries, it enables efficient inference, thereby providing a solid foundation for high-definition image capture, real-time data processing, and large data storage. To further optimize the system's performance, the control flow design integrates the video frame difference algorithm (VFDA), frame rate reduction and frequency interpolation (FRRFI) algorithm, and asynchronous processing, as shown in Fig. 5c–d. The VFDA monitors video frame changes to accelerate the edge device's response to semantic segmentation, bypassing unnecessary steps when significant variations are detected. The FRRFI algorithm reduces processing frequency for static image segmentation, minimizing latency and enhancing video smoothness. The asynchronous processing mechanism allows concurrent execution of frame processing and model inference tasks, significantly improving the system's concurrent processing capabilities and real-time performance. Collectively, these components of the embedded system—configuration, workflow, hardware, and control flow—work in synergy to ensure efficient, accurate, and smooth operation, even when handling complex images.

4. Results and discussion

4.1. Teacher model training

A systematic benchmark of contemporary vision backbones was conducted for sediment–fabric classification. ResNet-101 achieved the highest scores across accuracy (ACC), area under the ROC curve (AUC), F1-score, and Top-1 error (Fig. 6a). The evaluation pool included foundational convolutional networks (Krizhevsky et al., 2012; Sermanet et al., 2013; Zeiler and Fergus, 2013; Simonyan and Zisserman, 2014), the Inception family (Szegedy et al., 2014; Ioffe and Szegedy, 2015; Szegedy et al., 2015, 2016), ResNet variants (He et al., 2015, 2016; Han et al., 2016; Xie et al., 2016; Zhang et al., 2016; Gao et al., 2019), DenseNet variants (Huang et al., 2016; Chen et al., 2017; Yang et al., 2018), and state-of-the-art Transformer backbones (Dosovitskiy et al., 2020; Touvron et al., 2020; Liu et al., 2021a,b).

Loss trajectories on the validation and test sets (Fig. 6b–c) confirm the advantage of ResNet-101. It converged to 0.05 and 0.12,

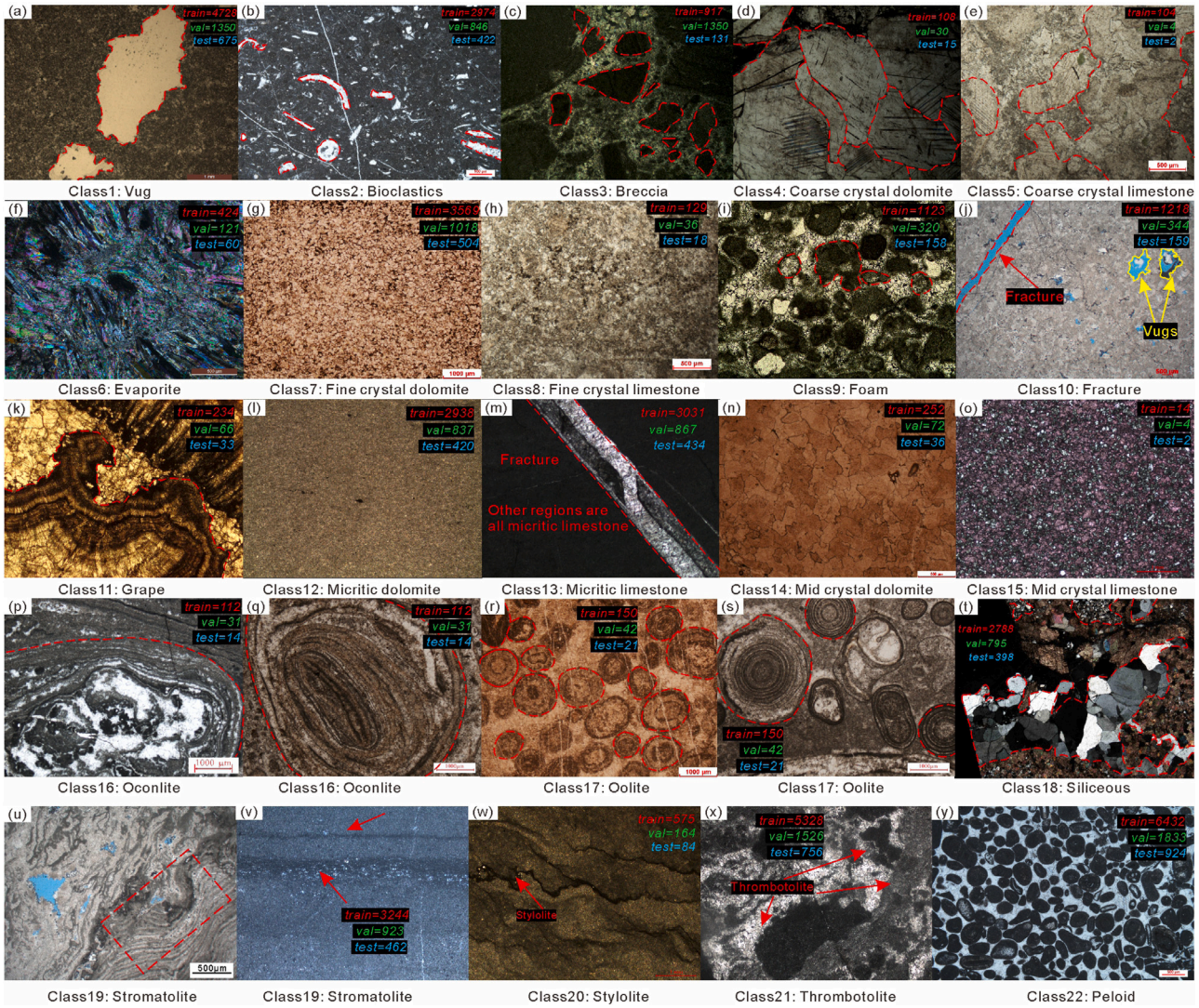

Fig. 4. Carbonate thin section images dataset. Totally 22 classes were collected. The representing images denoted the name of each class and numbers of the train, validation and test datasets were shown by red, green and blue colors.

respectively, while sustaining accuracies of 0.87 and 0.85. These values are the lowest among all ResNet depths. The absence of divergence indicates that ResNet-101 balances capacity and generalization.

Class-wise performance was examined with confusion matrices. Every fabric except peloid exceeded 80% accuracy (Fig. 7). Approximately 10% of vug samples were mis-classified as bioclastic. Breccia samples showed 20% confusion, split between bioclastic (13%) and evaporite (6%). Among crystalline carbonates, 5% were mis-assigned by crystal size, and 10% by lithology (calcite vs. dolomite) (Fig. 7b,d). Fabrics with distinct morphology were classified almost perfectly. The only exception was peloid, which was occasionally labeled as oolite (Fig. 7e). The overlap in circular-to-elliptical geometry explains this ambiguity. Probabilities for absent classes are zero because all categories were evaluated jointly.

Superior discrimination alone does not justify the use of ResNet-101 as a teacher model. The decisive requirement is the generation of high-fidelity class activation maps (CAMs) that delineate sediment boundaries in thin sections. In weakly supervised semantic segmentation (WSSS), the reliability of pseudo-labels determines downstream performance. Standard CAMs often suffer from gradient saturation and spatial diffusion (Wang et al., 2020). Such artifacts degrade boundary precision and reduce their value as segmentation labels.

CAMs produced by ResNet-101 were therefore inspected at pixel level (Fig. 8a–h). Fig. 8e–h shows that the maps from the teacher model align tightly with sediment outlines and coincide with expert geological interpretation ((Fig. 8a–d). The correspondence between CAM-derived masks and thin-section observations confirms that ResNet-101 supplies stable supervision. Within the FLKD framework, this quality allows a lightweight CamNet to be trained without dense manual annotations.

4.2. Performance and analysis of CamNet

Storage constraints on marginal devices limit the choice of feasible network backbones to lightweight architectures such as ShuffleNet-V2 (Ma et al., 2018) and MobileNet-V3-Small (Howard et al., 2019, 2017). To evaluate the effectiveness of our proposed framework, two established lightweight networks with improved CamNet trained under the FLKD framework are compared. After 30 epochs training, all light model candidates's loss curves converged smoothly (Fig. 9a), indicating stable training processes across architectures.

To comprehensively assess performance, three criteria consistent with the core goal of lightweight distillation for edge deployment have been introduced, including (i) alignment between the student model's CAMs and the teacher model's (ResNet-101) CAMs (to validate knowledge transfer), (ii) consistency of segmentation results (to ensure practical utility), and (iii) accuracy in sediment classification (Fig. 9b). CamNet matched or nearly matched the best-performing baseline across

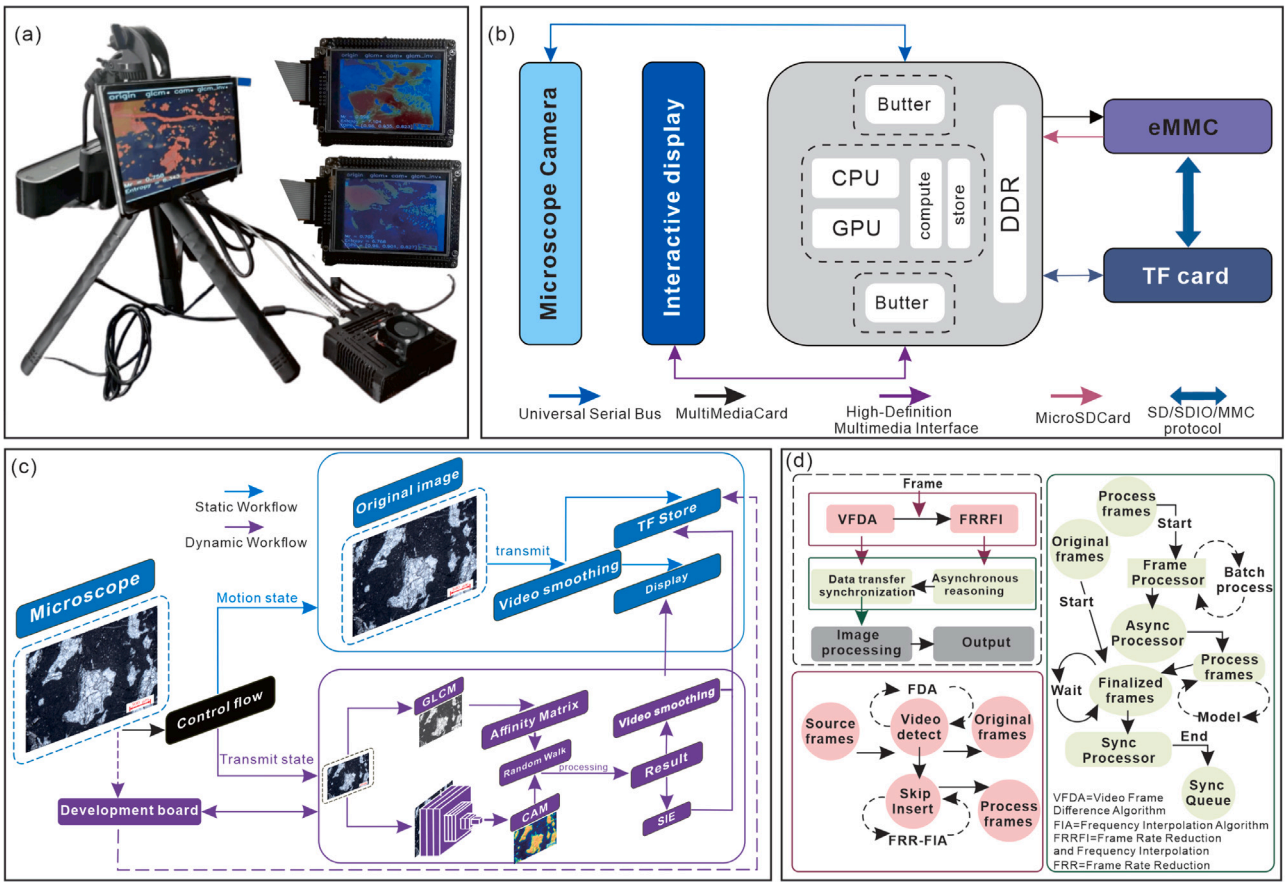

Fig. 5. The embedding system. (a) The overview of the embedding systems; (b) Workflow of the marginal devices; (c) The illustration of the inference process by control flow; (d) The illustration of displaying details.

all metrics. CamNet lagged by only 5% in recall and F1-score, 3% in precision, and approximately 1% in specificity and training loss (Fig. 9b). Most notably, CamNet outperformed all competitors in SSIM-CAM and SSIM-seg—metrics critical for validating the fidelity of feature localization, a key objective of the FLKD framework (Fig. 9b).

The parameter compression achieved by CamNet is particularly striking. While maintaining accuracy comparable to ResNet-101, CamNet uses only 0.6% of the teacher model's parameters—representing an 85%–90% reduction relative to ShuffleNet-V2 and MobileNet-V3-Small (Fig. 9a). This drastic reduction directly addresses the storage constraints of marginal devices and makes real-world deployment feasible without sacrificing core performance.

A natural concern with such aggressive compression is whether it distorts the model's receptive field or blurs semantic localization. However, visual inspection of CamNet's CAMs reveals spatial fidelity indistinguishable from ResNet-101, with no evidence of semantic leakage (Fig. 8i–l). The key distinction lies in granularity. CamNet emphasizes fine-scale sedimentary details, whereas ResNet-101 tends to highlight broad, contiguous regions. ResNet-101's activations often extend into the background and create wide green-to-yellow zones of mid-level attention, whereas CamNet's activations tightly focus on primary sedimentary features, minimizing such diffuse shading. This targeted attention aligns with the high SSIM-CAM and SSIM-seg scores observed earlier, confirming that compression does not compromise localization precision.

This edge-focused behavior is most evident in segmentation tasks involving particles and skeletal fossils (Fig. 10a–e). CamNet's CAMs show minimal mid-level activation within crystals or fossils (Fig. 10f–i), with one exception in mid-crystal dolomite (Fig. 10b). The absence of a dominant foreground object triggers broader mid-level attention, slightly increasing computational overhead and softening segmentation sharpness. Even in this case, the deviation remains minor, underscoring the model's robustness.

CamNet's ability to balance fine-grained precision with parsimony stems from architectural design of the integration of squeeze-and-excitation (SE) attention modules and cascaded 3 × 3 convolutions. SE modules amplify salient micro-features (e.g., crystal edges), while stacked small-kernel convolutions confine receptive fields to individual sediment grains. In sedimentary imagery where semantic information is inherently discontinuous, large kernels risk aggregating features across adjacent grains and spill activation into the background (as seen in ResNet-101's green-yellow haze). By contrast, CamNet's progressive convolution strategy attenuates inter-grain semantic interference and suppresses redundancy and concentrating computation on informative boundaries.

In summary, FLKD-trained CamNet achieves precise localization of sedimentary features while reducing the parameter count by two orders of magnitude. The selective attention mechanism eliminates extraneous background computation, yielding segmentation maps that are both accurate and efficient, which is ideally suited for resource-constrained deployment on marginal devices.

4.3. Performance and analysis of embedded systems

The performance metrics of the embedded system, summarized in Table 1, underscore the critical necessity of optimizing deep learning deployment for resource-constrained edge devices in geological applications. In practical scenarios such as field petrological analysis or on-site microscopic inspection of carbonate thin sections, equipment is often limited by compact form factors, constrained power supplies, and

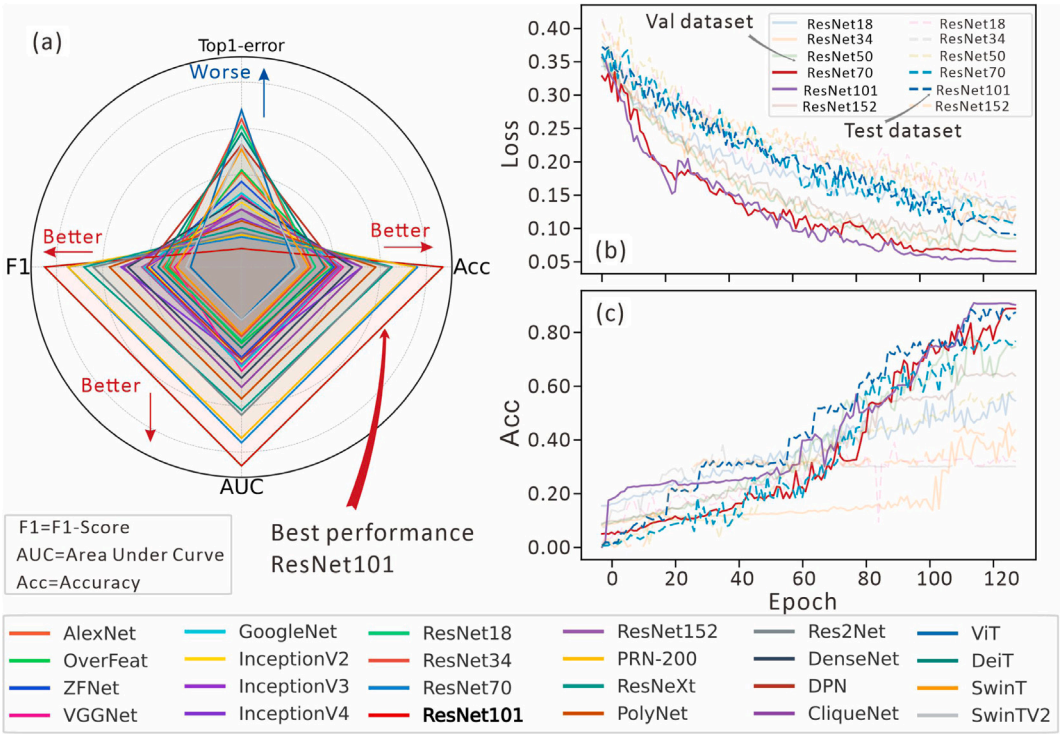

Fig. 6. Performance comparison in classification task. (a) Radar plots of different current image classification models (models consist of AlexNet (Krizhevsky et al., 2012), OverFeat (Sermanet et al., 2013), ZFNet (Zeiler and Fergus, 2013), VGGNet (Simonyan and Zisserman, 2014), GoogleNet (Szegedy et al., 2014), InceptionV2 (Ioffe and Szegedy, 2015), IceptionV3 (Szegedy et al., 2015), IceptionV4 (Szegedy et al., 2016), ResNet18/34/70 (He et al., 2015), ResNet101/152 (He et al., 2016), PRN-200 (Han et al., 2016), ResNeXt (Xie et al., 2016), PolyNet (Zhang et al., 2016), Res2Net (Gao et al., 2019), DenseNet (Huang et al., 2016), DPN (Chen et al., 2017), CluqueNet (Yang et al., 2018), ViT (Dosovitskiy et al., 2020), DeiT (Touvron et al., 2020), SwinT (Liu et al., 2021a) and SwinTV2 (Liu et al., 2021b)); (b) and (c) are training process of ResNet-101.

Table 1
Experimental results on performance of embedded systems.

Types	FPS	CPU	GPU	Memory utilization
Original system	1.15 fps	9.3%	99.4%	84.6%
Optimization system	6.87 fps	8.5%	99.6%	92.3%

restricted computational resources including low-end GPUs and limited RAM. Traditional deep learning models, despite their high accuracy, typically require substantial memory and processing power, making them incompatible with such environments. For instance, the original system, which lacked targeted optimizations, achieved a mere 1.15 fps, far below the threshold for real-time analysis generally considered 5–10 fps for smooth interaction. This sluggish throughput would severely hinder practical workflows, as geologists would face significant delays between image capture and segmentation results, undermining the efficiency of on-site decision-making.

Against this backdrop, the optimized system, powered by the lightweight CamNet and enhanced through the framework's control flow design, delivers transformative improvements. The six-fold increase in inference speed from 1.15 fps to 6.87 fps directly addresses the real-time requirement, enabling seamless integration with microscopic imaging equipment. This leap is not merely a technical achievement but a practical enabler: it allows geologists to obtain immediate feedback on sedimentary features such as crystal boundaries and fossil fragments during fieldwork, accelerating the analysis of carbonate microfacies and diagenetic processes.

The underlying efficiency gains stem largely from the Lightweight Model Enhancement strategies, which play a pivotal role in balancing performance and resource usage. First, the integration of Grey Level Co-occurrence Matrix (GLCM) ensures that CamNet—despite its reduced parameter count (800k, 0.6% of the teacher model's size)—retains the ability to capture fine-grained textural cues critical for carbonate segmentation. By modeling multi-scale and multi-directional texture relationships, GLCM compensates for the lightweight network's limited receptive field, reducing erroneous segmentation of visually similar features (e.g., peloids vs. ooids). This reduction in misclassification minimizes the need for iterative corrections, directly contributing to faster inference. Second, secondary information extraction via composite rate and image entropy further optimizes processing by prioritizing geologically significant regions (e.g., high-entropy zones with diverse fossils) and de-emphasizing redundant background. This targeted processing reduces unnecessary computations, allowing the system to allocate resources efficiently to critical features.

These enhancements synergize with the refined control flow, which incorporates the Video Frame Difference Algorithm (VFDA) for dynamic scenes and Frame Rate Reduction and Frequency Interpolation (FRRFI) for static scenarios, to optimize resource allocation. The impact is evident in hardware utilization metrics: CPU usage drops slightly from 9.3% to 8.5%, as optimized scheduling (aided by CR-driven region prioritization) minimizes redundant processing of low-importance areas. Concurrently, GPU utilization climbs to 99.6%, reflecting more efficient exploitation of the accelerator's capabilities to process GLCM-derived texture features, which are precomputed and streamlined to avoid bottlenecks.

The modest increase in memory utilization from 84.6% to 92.3% is a deliberate trade-off enabled by enhancement strategies. This uptick primarily reflects the storage of intermediate texture features from GLCM and secondary metrics (CR, H(I)), which are compactly encoded to

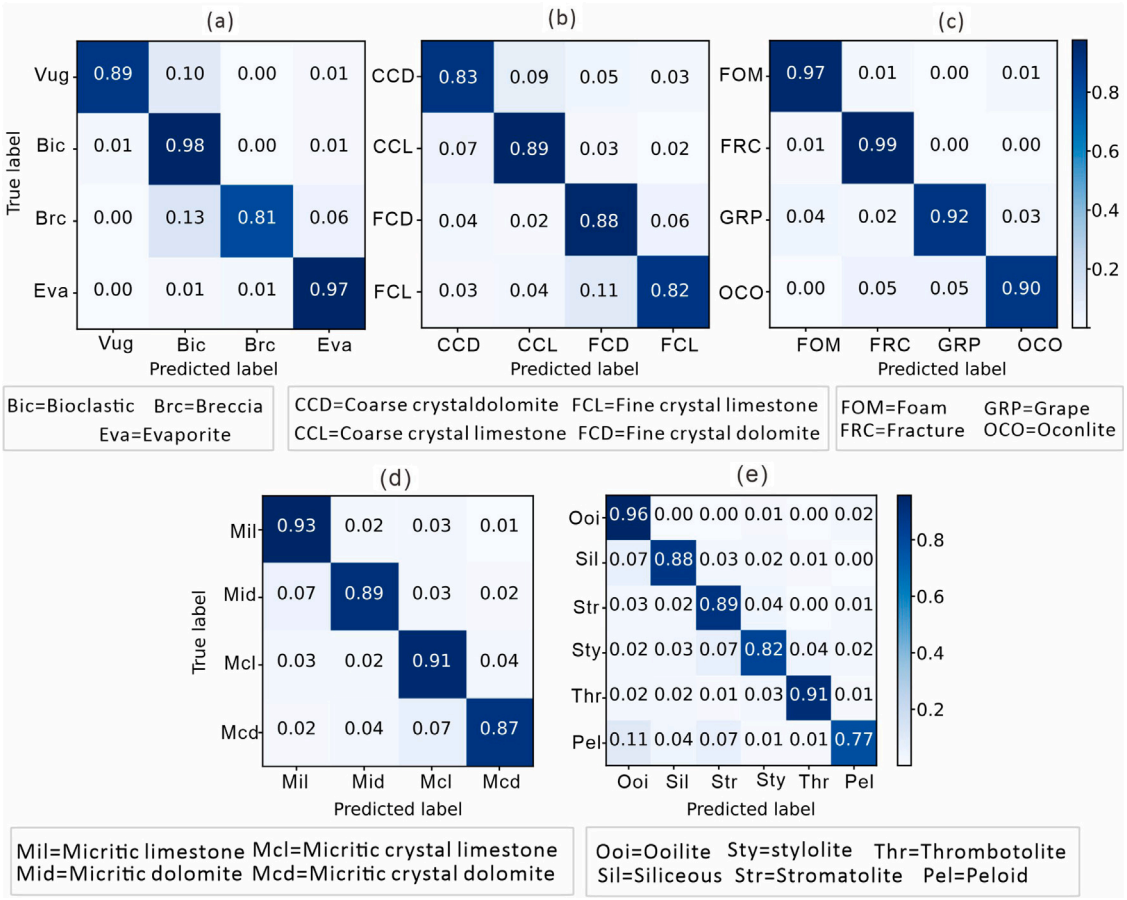

Fig. 7. Confusion matrices in classification task. The teacher model is ResNet-101 and the analyzed results were obtained from the test dataset.

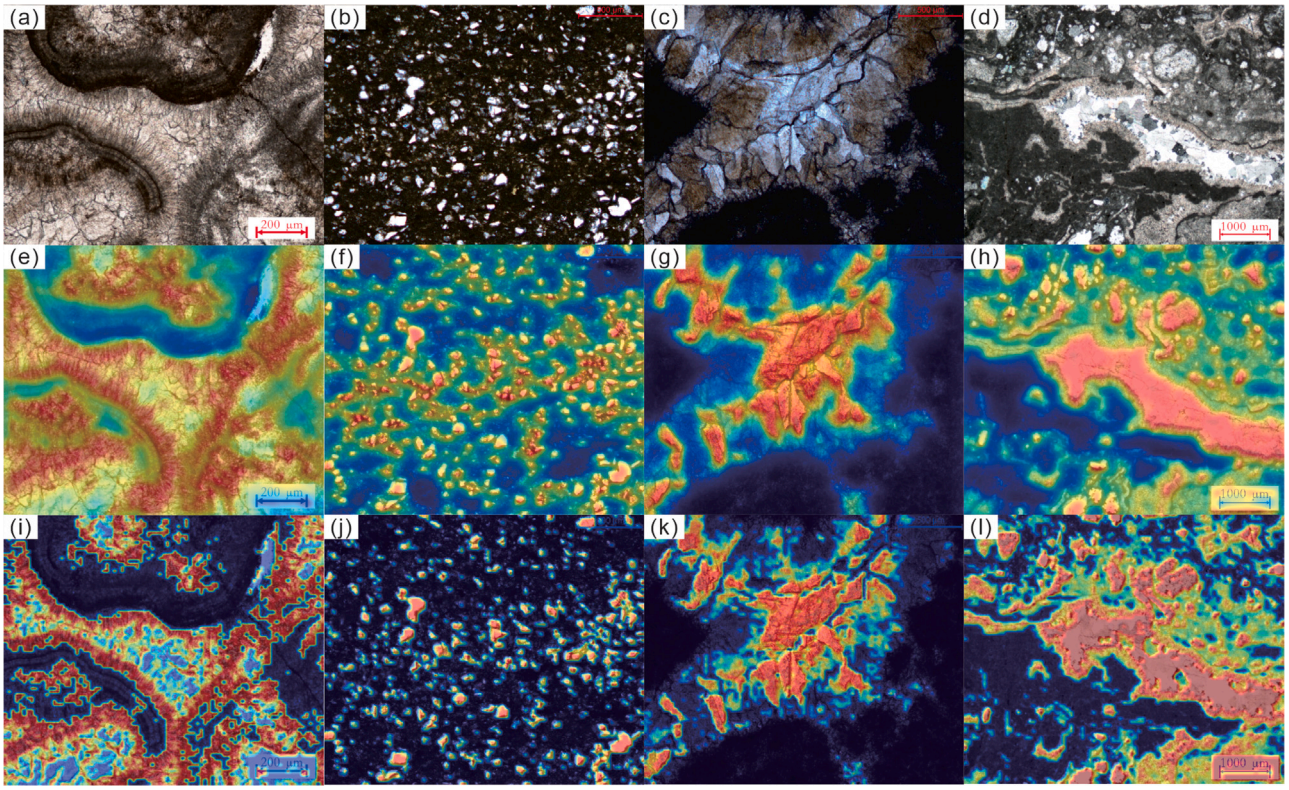

Fig. 8. Performs of CamNet. (a)–(d) Original images; (e)–(h) CAMs by ResNet101; (i)–(l) Outputs of Camnets.

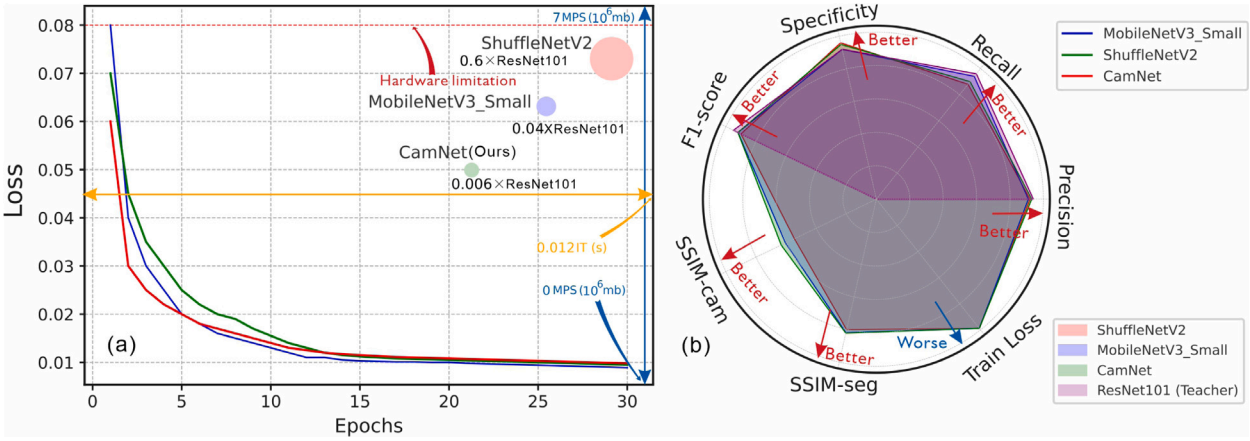

Fig. 9. The loss, inference time, model size and model performance. (a) Knowledge distillation loss and model parameters. The horizontal orange line is the inference time (s). The vertical blue line is the model size (10^6 mb). (b) Radar plot of different models.

Table 2
Quantitative comparison of segmentation models.

Model	MIoU	MPA	Precision	Dice
CNNs (Long et al., 2015)	57.4	72.1	75.6	72.5
U-Net (Ronneberger et al., 2015a)	61.8	75.8	77.7	76.2
DeepLabv3+ (Chen et al., 2018)	54.1	69.5	71.7	69.8
LR-ASPP (Howard et al., 2019)	60.4	74.5	77.8	75.0
SAM (Kirillov et al., 2023a)	68.3	80.9	81.1	81.0
Ours	62.1	76.3	78.1	75.8

avoid excessive overhead. Importantly, this memory usage supports the system's ability to sustain intensive frame-level computations without throttling—a capability critical for handling the high-texture complexity of carbonate images, where unoptimized lightweight models would otherwise falter.

Together, these results validate the effectiveness of lightweight model enhancement in bridging the gap between efficiency and accuracy. By explicitly addressing the limitations of compact networks through texture-aware refinement and geological context enrichment, the framework ensures that CamNet not only fits within edge device constraints but also delivers segmentation quality comparable to larger models. This not only makes deep learning-based carbonate segmentation feasible in resource-limited settings but also paves the way for broader adoption of AI-driven tools in geological fieldwork, where efficiency and portability are paramount.

4.4. Advantages and limitations of CamNet

CamNet is a lightweight model trained under weak supervision. It matches the segmentation quality of ResNet-101 (Fig. 8) yet stores only 800 KB, which is less than 1% of common segmentation networks. Class activation maps reveal that CamNet retains microscopic sediment features (Fig. 10f–i). Output masks delineate carbonate textures more sharply than far larger baselines (Fig. 10j–n).

Transfer learning was applied to the segment anything Model (SAM, Kirillov et al., 2023b). SAM masks capture only coarse outlines (Fig. 10o–s). Individual crystals remain undetected (Fig. 10o–q). CamNet resolves crystal boundaries and their spatial relationships (Fig. 10j–l). In skeletal carbonates, CamNet cleanly separates fossil fragments from matrix (Fig. 10m–n). SAM omits corals (Fig. 10r) and hallucinates deposits in barren areas (Fig. 10d). Spicules, ignored by SAM (Fig. 10s), are accurately segmented by CamNet (Fig. 10n). Thus, despite a three-order-of-magnitude parameter reduction (800 KB vs. 2.4 GB), CamNet exceeds SAM in geological detail and meets edge-device constraints.

Purely unsupervised methods (GMM, K-Means, MeanShift, and SLIC) were evaluated on the same data (Fig. 11). GMM yields jagged contours and blank masks on powder-crystal images. K-Means and MeanShift rely on intensity and lose semantic meaning. SLIC also fails to encode geological context. Weakly supervised CamNet therefore outperforms all unsupervised alternatives.

Zero-shot transfer to clastic-rock thin sections (the annotated dataset follows Zheng et al., 2022, 2024b,a) was tested without retraining (Table 2). CamNet attains 62.1% mIoU, 76.3% mPA, and 78.1% precision. The Dice coefficient is 75.8%, ranking among the top methods. Visual results (Fig. 12) of the CamNet's masks show crisp boundaries that follow ground-truth boundaries. In contrast, CNN baselines (U-Net, DeepLabv3+, LR-ASPP) have blur edges. SAM yields comparable fidelity but not superior detail. Therefore, CamNet models trained solely for image level classification still provide dense predictive capabilities without pixel level supervision.

Within the CAM–GLCM–KD framework, the system provides real-time, efficient, portable, and low-cost carbonate identification. Fabric analysis executes entirely on edge microscopes, eliminating cloud dependence. Several limitations remain. First, CamNet occasionally loses semantic information at image borders on fine-grained carbonates due to padding on variable-size inputs (Fig. 11). Second, the CamNet is brightness-sensitive where dark regions within coarse dolomite grains are under-segmented (row 3, left panel, Fig. 11). Thirdly, although mixed scale sediments have been successfully processed by CamNet, the representativeness of the largest nuclear allotropes is insufficient (row 3, right panel, Fig. 11). Future work will refine receptive-field calibration and illumination normalization to address these issues.

4.5. Implications for model development and feature improvement

Most recent work on deep learning for rock and microscope images has prioritized architectural novelty over domain-specific design. Mainstream vision models continue to scale on general-purpose datasets, and petrographic counterparts follow the same trend. The complexity of the model has significantly increased, but the unique morphology, structure, and multi-scale properties of sedimentary components have received limited attention. Whether these large feature extractors truly benefit thin-section analysis remains open. CamNet offers a

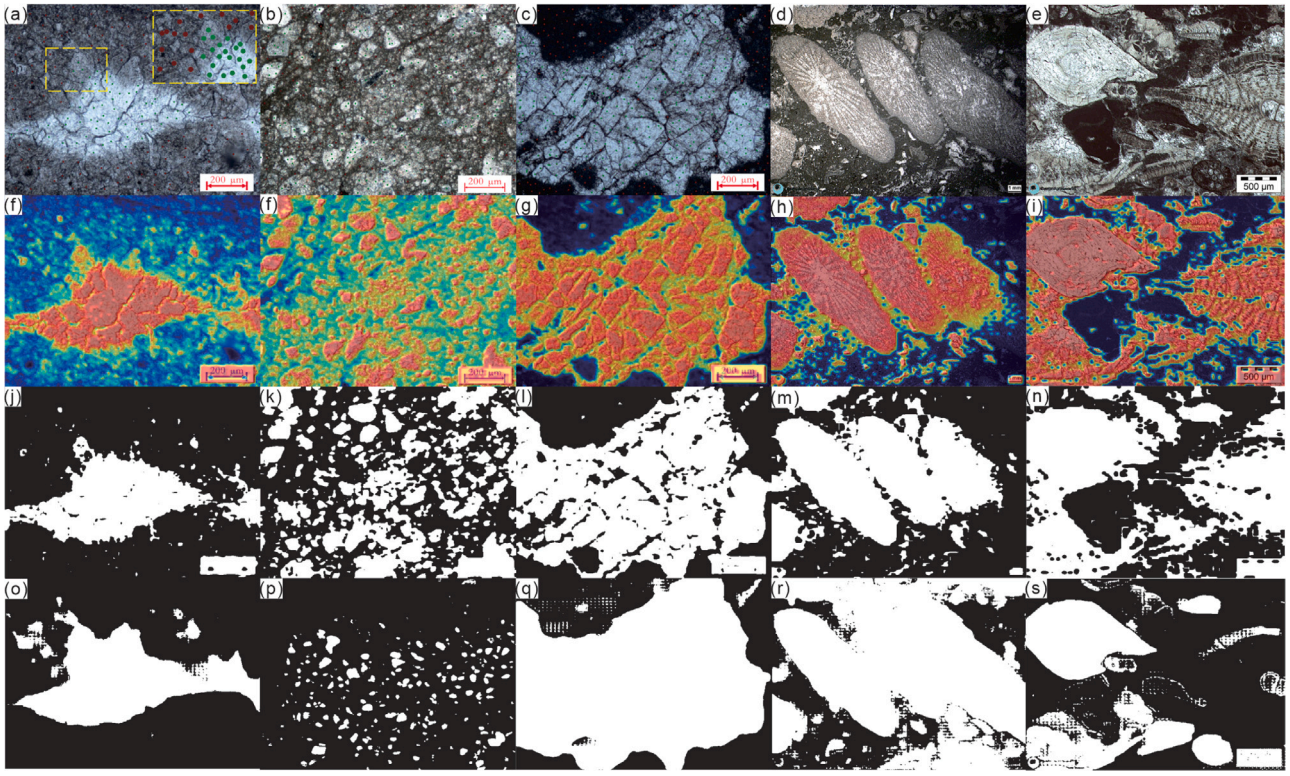

Fig. 10. Original thin section, CAMs and segmented masks. (a)–(e) original carbonate thin section images. The dashed yellow frame denotes the labeled key points to evaluate the performance in weakly supervised learning. The green color represents true positive (TP), while the red color represents false positive (FP). The details can be seen in Section Evaluation. (f)–(i) show the CAMs of CamNet. (j)–(n) display the segmented images by CamNet. (o)–(s) display the segmented images by a fine-tuned SAM model.

counterexample.CamNet replaces the conventional 7 × 7 input convolution (a standard in ResNet-style backbones)with cascades of 3 × 3 and 1 × 1 kernels linked by dense residuals. This choice removes redundant peripheral activations. Class activation maps confirm the difference. ResNet-101 produces diffuse attention beyond grain boundaries, whereas CamNet focuses on fine-grained sedimentary detail. Redundant extraction is especially pronounced for small targets, indicating that scaling laws derived from natural images do not transfer directly to petrographic domains. Explicit geological priors can therefore guide more efficient architectures. Deployment constraints are often ignored during model design. The FLKD pipeline compresses the teacher network by two orders of magnitude while incurring only a 1%–5% accuracy drop. This trade-off enables real-time inference on edge devices and closes the gap between laboratory-grade accuracy and field-deployable hardware.

The proposed WSSS framework also supports scalable training. Carbonate thin sections exhibit two regimes: (i) well-delimited grains and (ii) structures with indistinct boundaries, both spanning multiple scales. Pixel-level annotation is labor-intensive, and ambiguous boundaries hinder consistent labeling. In contrast, image class level labels are produced rapidly. Experimental results show that a well-trained classifier within the WSSS pipeline suffices to drive downstream segmentation. These findings motivate the creation of a foundational model trained on large-scale, weakly labeled petrographic datasets.

Future directions include the following:

1. **Multi-scale feature extraction.** Pyramid architectures or multi-scale fusion mechanisms can represent sedimentary structures across scales and improve both fine-detail recovery and contextual coherence.
2. **Enhanced WSSS framework.** Texture-aware pixel-affinity modules tailored to mineralogical classes can refine semantic diffusion and yield accurate pixel-level masks within an edge-compatible design.
3. **System-level optimization.** Joint algorithmic and hardware tuning—including parallelism profiling, memory-footprint reduction, and accelerator exploitation—can further elevate inference efficiency without sacrificing versatility.
4. **User-centric interface.** An intuitive graphical interface with interactive tutorials, real-time feedback, and collaborative tools can facilitate adoption in educational and research environments.

5. Conclusion

The research has successfully developed a three-stage segmentation method that integrates weakly supervised learning techniques, including class activation mapping (CAM), grey level co-occurrence matrix (GLCM), and knowledge distillation (KD), to enhance the performance of a lightweight network, CamNet, for the analysis of carbonate thin section images. This innovative approach addresses the challenges associated with complex deep learning models in practical and edge computing environments by focusing on a lightweight and efficient solution.

The CamNet model, with its significantly reduced parameter count of only 800k, demonstrates a remarkable balance between computational efficiency and accuracy. It achieves an inference speed of 6.87 fps when deployed in embedded systems, making it suitable for real-time applications.

In conclusion, the lightweight carbonate thin section image-assistant recognition system not only meets the demands of modern petrological and mineralogical analysis. This research contributes significantly to the field by offering a scalable and deployable solution that leverages the power of deep learning while maintaining the simplicity required for edge deployment. The system's success in both academic and practical settings highlights its potential for widespread adoption in geological studies.

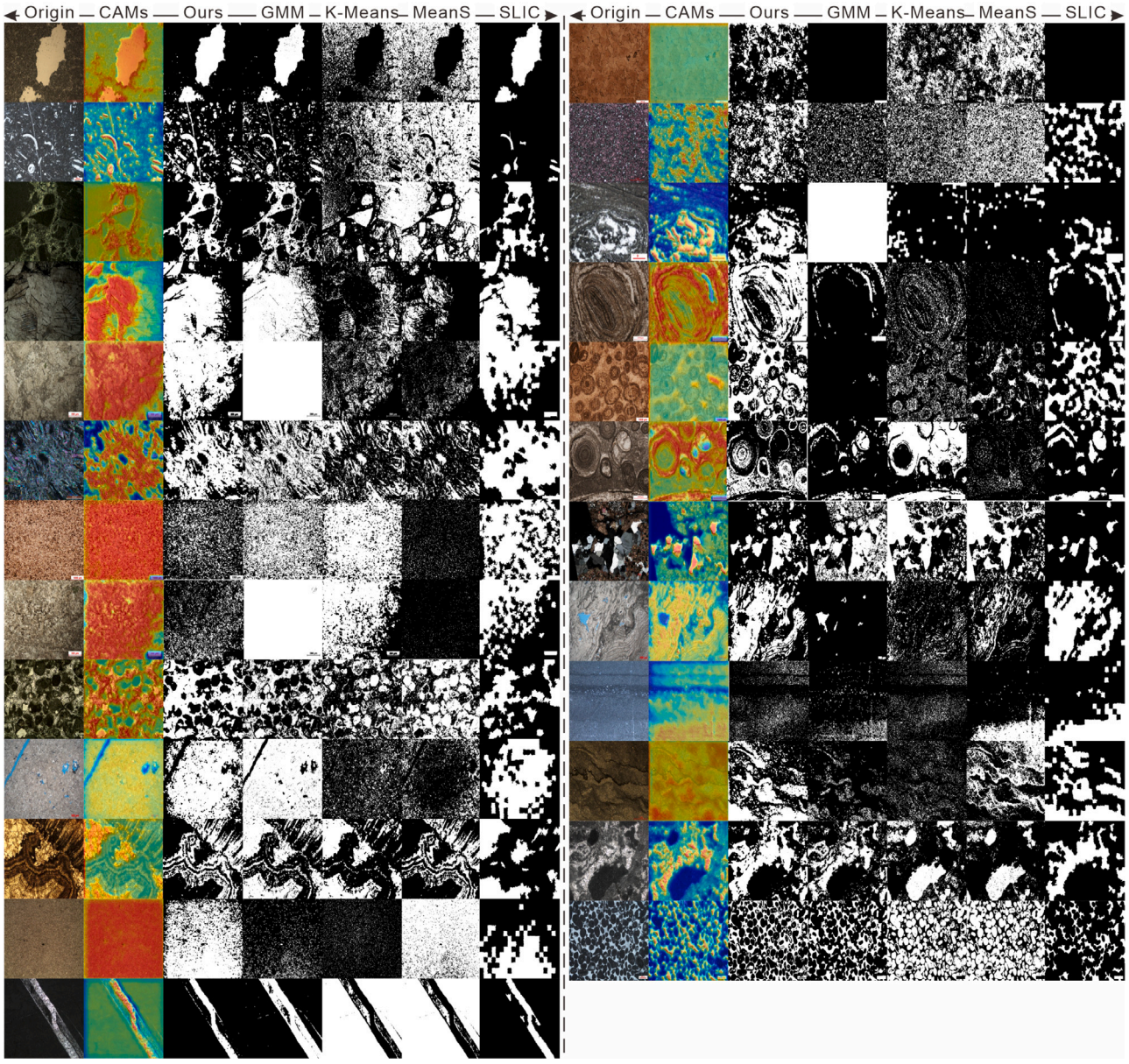

Fig. 11. Comparisons of ours weakly supervised CamNet and unsupervised Gaussian Mixture Model (GMM, Reynolds, 2015), K-means (Dhanachandra et al., 2015), Mean Shift (shorten as MeanS, Comaniciu and Meer, 1999) and SLIC (Achanta et al., 2012).

Code availability

Name of the code/library
Contact: keranli98@outlook.com
Hardware requirements: No
Program language: Python
Software required: Python
Program size: 500KB
The source codes are available for downloading at the link: https://github.com/ai4geology/CamNet

CRediT authorship contribution statement

Keran Li: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yujie Gao:** Writing – original draft, Software, Methodology, Investigation, Formal analysis. **Yingjie Ma:** Writing – original draft, Software, Methodology, Investigation, Formal analysis. **Chengkun Li:** Software. **Junjie Ye:** Software. **Hao Yu:** Software. **Yiming Xu:** Software. **Dongyu Zheng:** Writing – review & editing, Supervision, Project administration. **Ardiansyah Koeshidayatullah:** Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

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

Acknowledgments

This work was inspired and facilitated by the contributions of numerous colleagues and students in AI4Geoscience research group. We extend our gratitude to Liao Mu for insightful discussions in deep learning methods and hardware that greatly improved these results. This research did not receive any financial support. All software and hardware expenses were self-funded by AI4Geoscience research group members.

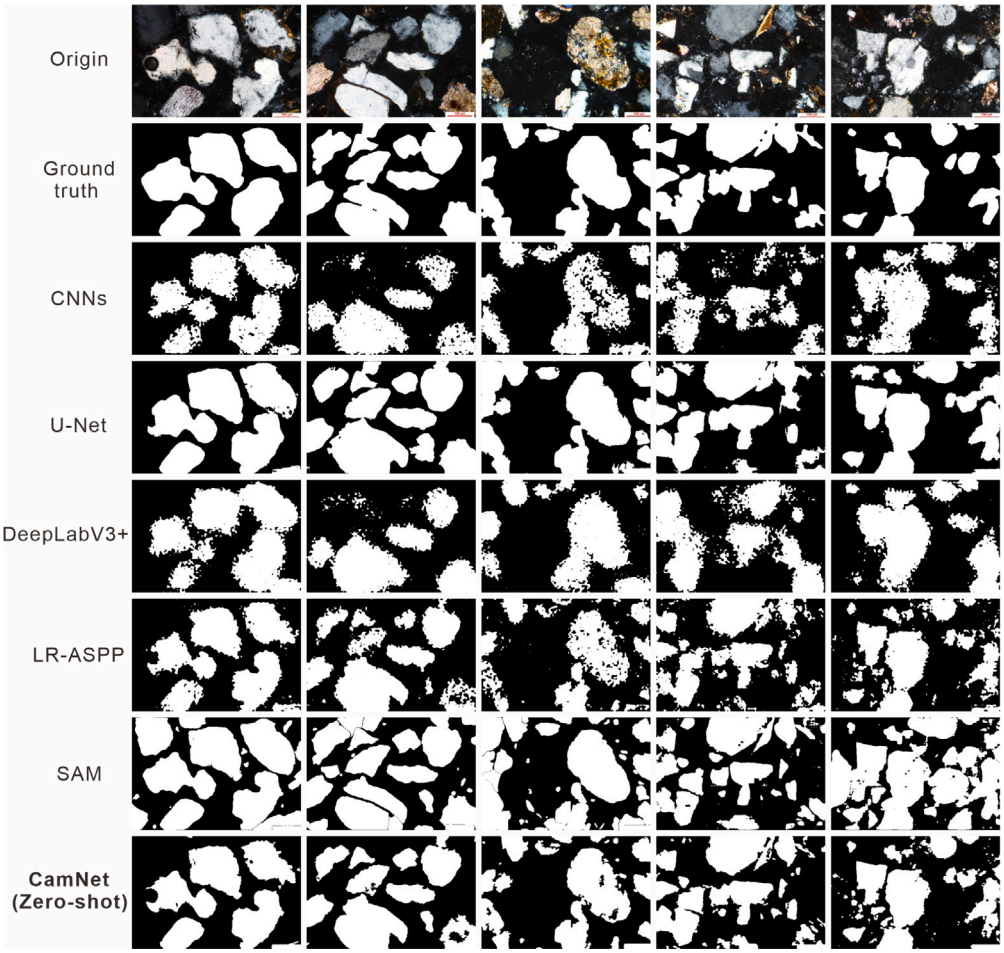

Fig. 12. Binary mask generation of Zero-shot CamNet and fully trained CNN model (Ronneberger et al., 2015b), U-Net model (Ronneberger et al., 2015a), DeepLabV3+ model (Chen et al., 2016), LR-ASSP model (Canziani et al., 2016) and SAM model (Kirillov et al., 2023b).

Data availability

Data will be made available on request.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. 34 (11), 2274–2282.

Ahn, J., Kwak, S., 2018. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4981–4990, URL: https://api.semanticscholar.org/CorpusID:4702470.

Andrä, H., Combaret, N., Dvorkin, J.P., Glatt, E., Han, J., Kabel, M., Keehm, Y., Krzikalla, F., Lee, M., Madonna, C., Marsh, M., Mukerji, T., Saenger, E.H., Sain, R., Saxena, N., Ricker, S., Wiegmann, A., Zhan, X., 2013. Digital rock physics benchmarks - part I: Imaging and segmentation. Comput. Geosci. 50, 25–32, URL: https://api.semanticscholar.org/CorpusID:5722082.

Bathurst, R.G.C., 1972. Carbonate sediments and their diagenesis. URL: https://api.semanticscholar.org/CorpusID:128640568.

Bearman, A., Russakovsky, O., Ferrari, V., Fei-Fei, L., 2015. What's the point: Semantic segmentation with point supervision. In: European Conference on Computer Vision. URL: https://api.semanticscholar.org/CorpusID:1356654.

Blanco-Filgueira, B., Garcia-Lesta, D., Fernández-Sanjurjo, M., Brea, V.M., López, P., 2019. Deep learning-based multiple object visual tracking on embedded system for IoT and mobile edge computing applications. IEEE Internet Things J. 6 (3), 5423–5431.

Bosence, D., 2002. Carbonate reservoirs porosity evolution and diagenesis in a sequence stratigraphic framework. Mar. Pet. Geol. 19, 1295–1296, URL: https://api.semanticscholar.org/CorpusID:129503702.

Brunet, D., Vrscay, E.R., Wang, Z., 2011. On the mathematical properties of the structural similarity index. IEEE Trans. Image Process. 21 (4), 1488–1499.

Canziani, A., Paszke, A., Culurciello, E., 2016. An analysis of deep neural network models for practical applications. arXiv preprint arXiv:1605.07678.

Chai, H., Xing, F., Gu, Q., Chen, X., Zhou, S., 2020. A carbonate micrograph dataset of feixianguan formation in northwestern margin of upper yangtze. China Sci. Data 5, 1–10. http://dx.doi.org/10.11922/csdata.2020.0007.zh, URL: http://csdata.org/p/7/414/.

Chang, X., Hou, M., Liu, X., Fan, T., 2020. A micrograph dataset of late ordovician carbonate rocks (including bioclasts) in Northwest Tarim and South China. China Sci. Data 5, 1–10. http://dx.doi.org/10.11922/csdata.2020.0059.zh, URL: http://csdata.org/p/7/463/.

Chen, Y., Li, J., Xiao, H., Jin, X., Yan, S., Feng, J., 2017. Dual path networks. In: Neural Information Processing Systems. URL: https://api.semanticscholar.org/CorpusID:35602767.

Chen, Z., Liu, X., Yang, J., Little, E., Zhou, Y., 2020a. Deep learning-based method for SEM image segmentation in mineral characterization, an example from Duvernay Shale samples in western Canada Sedimentary Basin. Comput. Geosci. 138, 104450.

Chen, Z., Liu, X., Yang, J., Little, E., Zhou, Y., 2020b. Deep learning-based method for SEM image segmentation in mineral characterization, an example from Duvernay Shale samples in western Canada Sedimentary Basin. Comput. Geosci. 138, 104450, URL: https://api.semanticscholar.org/CorpusID:214543469.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y., 2021. TransUNet: Transformers make strong encoders for medical image segmentation. ArXiv, arXiv:2102.04306, URL: https://api.semanticscholar.org/CorpusID:231847326.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K.P., Yuille, A.L., 2016. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40, 834–848, URL: https://api.semanticscholar.org/CorpusID:3429309.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 801–818.

Comaniciu, D., Meer, P., 1999. Mean shift analysis and applications. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, IEEE, pp. 1197–1203.

De Siqueira, F.R., Schwartz, W.R., Pedrini, H., 2013. Multi-scale gray level co-occurrence matrices for texture description. Neurocomputing 120, 336–345.

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

Desai, S.S., Ramaswamy, H.G., 2020. Ablation-CAM: Visual explanations for deep convolutional network via gradient-free localization. In: 2020 IEEE Winter Conference on Applications of Computer Vision. WACV, pp. 972–980, URL: https://api.semanticscholar.org/CorpusID:214604773.

Dhanachandra, N., Manglem, K., Chanu, Y.J., 2015. Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. Procedia Comput. Sci. 54, 764–771.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. ArXiv, arXiv:2010.11929, URL: https://api.semanticscholar.org/CorpusID:225039882.

Eichkitz, C.G., Amtmann, J., Schreilechner, M.G., 2013. Calculation of grey level co-occurrence matrix-based seismic attributes in three dimensions. Comput. Geosci. 60, 176–183.

Ferreira-Chacua, I., Koeshidayatullah, A., 2023. ForamViT-GAN: Exploring new paradigms in deep learning for micropaleontological image analysis. IEEE Access 11, 67298–67307, URL: https://api.semanticscholar.org/CorpusID:258048518.

Froute, L., Nazarova, M., Jolivet, I.C., Creux, P., Chaput, E., Kovscek, A.R., 2025. Identification of vaca muerta shale microlithofacies using convolutional neural networks with characterization by electron microscopy. Gas Sci. Eng. 134, 205519.

Gao, S., Cheng, M.-M., Zhao, K., Zhang, X., Yang, M.-H., Torr, P.H.S., 2019. Res2Net: A new multi-scale backbone architecture. IEEE Trans. Pattern Anal. Mach. Intell. 43, 652–662, URL: https://api.semanticscholar.org/CorpusID:91184391.

Gou, J., Sun, L., Yu, B., Wan, S., Ou, W., Yi, Z., 2023. Multilevel attention-based sample correlations for knowledge distillation. IEEE Trans. Ind. Inform. 19, 7099–7109, URL: https://api.semanticscholar.org/CorpusID:252552769.

Gou, J., Yu, B., Maybank, S.J., Tao, D., 2021. Knowledge distillation: A survey. Int. J. Comput. Vis. 129 (6), 1789–1819.

Han, Z., Hu, X., 2020. A photomicrograph dataset of early-middle Jurassic rocks in the Tibetan Tethys Himalaya. China Sci. Data 5, 1–10. http://dx.doi.org/10.11922/csdata.2020.0048.zh, URL: http://csdata.org/p/7/452/.

Han, D., Kim, J., Kim, J., 2016. Deep pyramidal residual networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 6307–6315, URL: https://api.semanticscholar.org/CorpusID:5398883.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 770–778, URL: https://api.semanticscholar.org/CorpusID:206594692.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity mappings in deep residual networks. In: European Conference on Computer Vision. URL: https://api.semanticscholar.org/CorpusID:6447277.

Hinton, G.E., Vinyals, O., Dean, J., 2015. Distilling the knowledge in a neural network. ArXiv, arXiv:1503.02531, URL: https://api.semanticscholar.org/CorpusID:7200347.

Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al., 2019. Searching for mobilenetv3. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1314–1324.

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. MobileNets: Efficient convolutional neural networks for mobile vision applications. ArXiv, arXiv:1704.04861, URL: https://api.semanticscholar.org/CorpusID:12670695.

Huang, T.S., Aggarwal, J.K., 1981. Image sequence analysis. URL: https://api.semanticscholar.org/CorpusID:117971514.

Huang, G., Liu, Z., Weinberger, K.Q., 2016. Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 2261–2269, URL: https://api.semanticscholar.org/CorpusID:9433631.

Huang, Z., Wang, X., Wang, J., Liu, W., Wang, J., 2018. Weakly-supervised semantic segmentation network with Deep Seeded Region growing. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7014–7023, URL: https://api.semanticscholar.org/CorpusID:51690586.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. ArXiv, arXiv:1502.03167, URL: https://api.semanticscholar.org/CorpusID:5808102.

Iwaszenko, S., Nurzynska, K.O., 2019. Rock grains segmentation using curvilinear structures based features. In: Defense + Commercial Sensing. URL: https://api.semanticscholar.org/CorpusID:182769609.

Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A.G., Adam, H., Kalenichenko, D., 2017. Quantization and training of neural networks for efficient integer-arithmetic-only inference. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2704–2713, URL: https://api.semanticscholar.org/CorpusID:39867659.

Kanwal, N., Eftestøl, T., Khoraminia, F., Zuiverloon, T.C.M., Engan, K., 2023. Vision transformers for small histological datasets learned through knowledge distillation. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. URL: https://api.semanticscholar.org/CorpusID:258959193.

Khoreva, A., Benenson, R., Hosang, J.H., Hein, M., Schiele, B., 2016. Simple does it: Weakly supervised instance and semantic segmentation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 1665–1674, URL: https://api.semanticscholar.org/CorpusID:12124389.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., Girshick, R.B., 2023a. Segment anything. In: 2023 IEEE/CVF International Conference on Computer Vision. ICCV, pp. 3992–4003, URL: https://api.semanticscholar.org/CorpusID:257952310.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., et al., 2023b. Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026.

Koeshidayatullah, A., 2023. Riding the wave: One-touch automatic salt segmentation by coupling sam and seggpt. Day 2 Tue, October 03, 2023 URL: https://api.semanticscholar.org/CorpusID:265758449.

Koeshidayatullah, A., Al-Azani, S., Baraboshkin, E.E., Alfarraj, M., 2022. FaciesViT: Vision transformer for an improved core lithofacies prediction. In: Frontiers in Earth Science. URL: https://api.semanticscholar.org/CorpusID:252686243.

Koeshidayatullah, A., Morsilli, M., Lehrmann, D.J., Al-Ramadan, K., Payne, J.L., 2020. Fully automated carbonate petrography using deep convolutional neural networks. Mar. Pet. Geol. 122, 104687.

Kreuzthaler, M., Schulz, S., Berghold, A., 2015. Secondary use of electronic health records for building cohort studies through top-down information extraction. J. Biomed. Inform. 53, 188–195.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90, URL: https://api.semanticscholar.org/CorpusID:195908774.

Lai, W., Jiang, J., Qiu, J., Yu, J., Hu, X., 2020. A photomicrograph dataset of rocks for petrology teaching at Nanjing University. China Sci. Data 5, 1–10. http://dx.doi.org/10.11922/csdata.2020.0071.zh, URL: http://csdata.org/p/7/474/.

Li, J., Hu, X., 2020. A photomicrograph dataset of late cretaceous to early paleogene carbonate rocks in tibetan himalaya. China Sci. Data 5, 1–10. http://dx.doi.org/10.11922/csdata.2020.0072.zh, URL: http://csdata.org/p/7/475/.

Li, K., Song, J., Liu, S., Feng, Y., Li, Z., Jin, X., Fan, J., Chen, W., 2023a. A dataset of microscopic images of carbonate rocks of leikoupo formation of middle triassic in the central sichuan. China Sci. Data 8, 1–10. http://dx.doi.org/10.11922/11-6035.csd.2021.0054.zh, URL: http://csdata.org/p/7/621/.

Li, S., Wen, W., Wang, Y., Han, S., Chen, Y., Li, H.H., 2017. An FPGA design framework for CNN sparsification and acceleration. In: 2017 IEEE 25th Annual International Symposium on Field-Programmable Custom Computing Machines. FCCM, 28–28, URL: https://api.semanticscholar.org/CorpusID:29321204.

Li, M., Zhang, P., Hai, T., 2023b. Pore extraction method of rock thin section based on attention U-net. J. Phys.: Conf. Ser. 2467, URL: https://api.semanticscholar.org/CorpusID:258739978.

Li, S., Zhao, X., Zhou, G., 2019. Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. Computer-Aided Civ. Infrastruct. Eng. 34 (7), 616–634.

Li, X., Zhou, T., Li, J., Zhou, Y., Zhang, Z., 2020. Group-wise semantic mining for weakly supervised semantic segmentation. In: AAAI Conference on Artificial Intelligence. URL: https://api.semanticscholar.org/CorpusID:228064085.

Liang, Y., Lu, L., Xie, J., 2021. OMNI: A framework for integrating hardware and software optimizations for sparse CNNs. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. 40, 1648–1661, URL: https://api.semanticscholar.org/CorpusID:226587136.

Lin, D., Dai, J., Jia, J., He, K., Sun, J., 2016. ScribbleSup: Scribble-supervised convolutional networks for semantic segmentation. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 3159–3167, URL: https://api.semanticscholar.org/CorpusID:3121011.

Ling, Y., Wang, J., Liu, Y., Zhang, W., 2008. A novel spatial and temporal correlation integrated based motion-compensated interpolation for frame rate up-conversion. IEEE Trans. Consum. Electron. 54, URL: https://api.semanticscholar.org/CorpusID:8815987.

Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L., Wei, F., Guo, B., 2021a. Swin transformer V2: Scaling up capacity and resolution. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 11999–12009, URL: https://api.semanticscholar.org/CorpusID:244346076.

Liu, Z., Li, J., Shen, Z., Huang, G., Yan, S., Zhang, C., 2017. Learning efficient convolutional networks through network slimming. In: 2017 IEEE International Conference on Computer Vision. ICCV, pp. 2755–2763, URL: https://api.semanticscholar.org/CorpusID:5993328.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021b. Swin transformer: Hierarchical vision transformer using shifted windows. In: 2021 IEEE/CVF International Conference on Computer Vision. ICCV, pp. 9992–10002, URL: https://api.semanticscholar.org/CorpusID:232352874.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440.

LOV'ASZ, L., Winkler, P., Luk'acs, A., Kotlov, A., 1993. Random walks on graphs: A survey. URL: https://api.semanticscholar.org/CorpusID:10655982.

Ma, Z., He, X., Kwak, H., Gao, J., Sun, S., Yan, B., 2023a. Enhancing rock image segmentation in digital rock physics: A fusion of generative AI and state-of-the-art neural networks. ArXiv, arXiv:2311.06079, URL: https://api.semanticscholar.org/CorpusID:265128753.

Ma, Z., He, X., Sun, S., Yan, B., Kwak, H., Gao, J., 2023b. Zero-shot digital rock image segmentation with a fine-tuned segment anything model. ArXiv, arXiv:2311.10865, URL: https://api.semanticscholar.org/CorpusID:265295035.

Ma, R., Liu, C., Yang, J., Wang, Y., Liu, J., 2020. A carbonate microscopic image dataset of the permo-carboniferous taiyuan formation in the southern margin of north China block. China Sci. Data 5, 1–10. http://dx.doi.org/10.11922/csdata.2020.0052.zh, URL: http://csdata.org/p/7/456/.

Ma, N., Zhang, X., Zheng, H.-T., Sun, J., 2018. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 116–131.

Malik, O.A., Puasa, I., Lai, D.T.C., 2022. Segmentation for multi-rock types on digital outcrop photographs using deep learning techniques. Sensors (Basel, Switzerland) 22, URL: https://api.semanticscholar.org/CorpusID:253146458.

Mamalakis, A., Barnes, E.A., Ebert-Uphoff, I., 2022. Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience. ArXiv, arXiv:2202.03407, URL: https://api.semanticscholar.org/CorpusID:246634041.

Mamalakis, A., Ebert-Uphoff, I., Barnes, E.A., 2021. Neural network attribution methods for problems in geoscience: A novel synthetic benchmark dataset. Environ. Data Sci. 1, URL: https://api.semanticscholar.org/CorpusID:232270084.

Manzoor, S., Qasim, T., Bhatti, N., Zia, M., 2023. Segmentation of digital rock images using texture analysis and deep network. Arab. J. Geosci. 16, URL: https://api.semanticscholar.org/CorpusID:259254764.

Mohanaiah, P., Sathyanarayana, P., GuruKumar, L., of E.C.E NEC-Narasaraopeta Professor, D., Nellore, I., 2013. Image texture feature extraction using GLCM approach. URL: https://api.semanticscholar.org/CorpusID:9132178.

Niu, Y., Mostaghimi, P., Shabaninejad, M., Swietojanski, P., Armstrong, R.T., 2020. Digital rock segmentation for petrophysical analysis with reduced user bias using convolutional neural networks. Water Resour. Res. 56, URL: https://api.semanticscholar.org/CorpusID:212996425.

Papandreou, G., Chen, L.-C., Murphy, K.P., Yuille, A.L., 2015. Weakly- and semi-supervised learning of a DCNN for semantic image segmentation. ArXiv, arXiv:1502.02734, URL: https://api.semanticscholar.org/CorpusID:3035960.

Park, W., Kim, D., Lu, Y., Cho, M., 2019. Relational knowledge distillation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 3962–3971, URL: https://api.semanticscholar.org/CorpusID:131765296.

Partio, M., Cramariuc, B., Gabbouj, M., Visa, A., 2002. Rock texture retrieval using gray level co-occurrence matrix. In: Proc. of 5th Nordic Signal Processing Symposium, vol. 75.

Pinheiro, P.H.O., Collobert, R., 2014. From image-level to pixel-level labeling with convolutional networks. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 1713–1721, URL: https://api.semanticscholar.org/CorpusID:7656505.

Qi, Z., Hou, M., Xu, S., et al., 2020. A Carbonate Microscopic Image Dataset of Sinian Dengying Period in Northwestern Margin of Sichuan Basin. Science Data Bank, http://dx.doi.org/10.11922/sciencedb.j00001.00105.

Qian, H., Xing, F., Zhang, C., Gu, Q., Chen, X., Fu, H., 2020. A dataset of microscope images of middle cambrian rock sections from xuzhuang formation in ordos basin. China Sci. Data 5, 1–10. http://dx.doi.org/10.11922/csdata.2020.0070.zh, URL: http://csdata.org/p/7/473/.

Ren, Y., Liang, J., Luo, S., 2021. Two extraction methods for carbonate rock oolites based on image segmentation algorithm. In: 2021 IEEE 4th International Conference on Information Systems and Computer Aided Education. ICISCAE, pp. 137–141, URL: https://api.semanticscholar.org/CorpusID:244044632.

Reynolds, D., 2015. Gaussian mixture models. In: Encyclopedia of Biometrics. Springer, pp. 827–832.

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S., 2019. Generalized intersection over union: A metric and a loss for bounding box regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 658–666.

Riloff, E., et al., 1993. Automatically constructing a dictionary for information extraction tasks. In: AAAI. 1, (1), pp. 811–816.

da Rocha, H.O., Moraes, P.E.F., Carrasquilla, A.A.G., de Figueiredo, J.J., 2025. Petrography image segmentation for analyzing the connectivity of pore networks in Brazilian pre-salt carbonate rocks of the barra velha formation. Carbonates Evaporites 40 (2), 56.

Ronneberger, O., Fischer, P., Brox, T., 2015a. U-net: Convolutional networks for biomedical image segmentation. ArXiv, arXiv:1505.04597, URL: https://api.semanticscholar.org/CorpusID:3719281.

Ronneberger, O., Fischer, P., Brox, T., 2015b. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, pp. 234–241.

Ru, L., Zhan, Y., Yu, B., Du, B., 2022. Learning affinity from attention: End-to-end weakly-supervised semantic segmentation with transformers. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 16825–16834, URL: https://api.semanticscholar.org/CorpusID:247292515.

See, J.-C., Ng, H.-F., Tan, H.K., Chang, J.-J., Mok, K.M., Lee, W.-K., Lin, C.-Y., 2023. Cryptensor: A resource-shared co-processor to accelerate convolutional neural network and polynomial convolution. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. 42, 4735–4748, URL: https://api.semanticscholar.org/CorpusID:260009675.

Selvaraju, R.R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., Batra, D., 2016. Grad-CAM: Visual explanations from deep networks via gradient-based localization. Int. J. Comput. Vis. 128, 336–359, URL: https://api.semanticscholar.org/CorpusID:15019293.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., 2013. OverFeat: Integrated recognition, localization and detection using convolutional networks. CoRR, arXiv:1312.6229, URL: https://api.semanticscholar.org/CorpusID:4071727.

Shamir, R.R., Duchin, Y., Kim, J., Sapiro, G., Harel, N., 2019. Continuous dice coefficient: a method for evaluating probabilistic segmentations. arXiv preprint arXiv:1906.11031.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. CoRR, arXiv:1409.1556, URL: https://api.semanticscholar.org/CorpusID:14124313.

Starkey, J., Samantaray, A.K., 1993. Edge detection in petrographic images. J. Microsc. 172, URL: https://api.semanticscholar.org/CorpusID:98250515.

Sun, Z., Xuan, P., Song, Z., Li, H., Jia, R., 2019. A texture fused superpixel algorithm for coal mine waste rock image segmentation. Int. J. Coal Prep. Util. 42, 1222–1233, URL: https://api.semanticscholar.org/CorpusID:212802959.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2016. Inception-v4, inception-ResNet and the impact of residual connections on learning. ArXiv, arXiv:1602.07261, URL: https://api.semanticscholar.org/CorpusID:1023605.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2014. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 1–9, URL: https://api.semanticscholar.org/CorpusID:206592484.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2015. Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 2818–2826, URL: https://api.semanticscholar.org/CorpusID:206593880.

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., J'egou, H., 2020. Training data-efficient image transformers & distillation through attention. In: International Conference on Machine Learning. URL: https://api.semanticscholar.org/CorpusID:229363322.

Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. In: Neural Information Processing Systems. URL: https://api.semanticscholar.org/CorpusID:13756489.

Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., Hu, X., 2020. Score-CAM: Score-weighted visual explanations for convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 24–25.

Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F., Afzal, N., Liu, S., Zeng, Y., Mehrabi, S., Sohn, S., et al., 2018. Clinical information extraction applications: a literature review. J. Biomed. Inform. 77, 34–49.

Wang, G., Wang, G., Zhang, X., Lai, J., Yu, Z., Lin, L., 2019. Weakly supervised person re-ID: Differentiable graphical learning and a new benchmark. IEEE Trans. Neural Netw. Learn. Syst. 32, 2142–2156, URL: https://api.semanticscholar.org/CorpusID:195886226.

Wei, Y., Xiao, H., Shi, H., Jie, Z., Feng, J., Huang, T.S., 2018. Revisiting dilated convolution: A simple approach for weakly- and semi-supervised semantic segmentation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7268–7277, URL: https://api.semanticscholar.org/CorpusID:44097132.

Wu, Z., Shen, C., Van Den Hengel, A., 2019. Wider or deeper: Revisiting the resnet model for visual recognition. Pattern Recognit. 90, 119–133.

Xie, S., Girshick, R.B., Dollár, P., Tu, Z., He, K., 2016. Aggregated residual transformations for deep neural networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 5987–5995, URL: https://api.semanticscholar.org/CorpusID:8485068.

Xie, J., Girshick, R.B., Farhadi, A., 2015. Unsupervised deep embedding for clustering analysis. ArXiv, arXiv:1511.06335, URL: https://api.semanticscholar.org/CorpusID:6779105.

Xin, H., Chen, J., 2020. A photomicrograph dataset of cambrian miaolingian–furongian carbonates at the jiulongshan section in western shandong. China Sci. Data 5, 1–9. http://dx.doi.org/10.11922/csdata.2020.0066.zh, URL: http://csdata.org/p/7/469/.

Xu, L., Bennamoun, Boussaid, F., Sohel, F., 2020a. Scale-aware feature network for weakly supervised semantic segmentation. IEEE Access 8, 75957–75967, URL: https://api.semanticscholar.org/CorpusID:218494680.

Xu, Y., Hu, X., Sun, G., 2020b. A photomicrograph dataset of mid-cretaceous rocks from langshan formation in the northern lhasa terrane, tibet. China Sci. Data 5, 1–8. http://dx.doi.org/10.11922/csdata.2020.0049.zh, URL: http://csdata.org/p/7/453/.

Yamada, T., Di Santo, S., 2022. Instance segmentation of piled rock particles based on mask R-CNN. In: IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium. pp. 163–166, URL: https://api.semanticscholar.org/CorpusID:252590474.

Yang, Y., Zhong, Z., Shen, T., Lin, Z., 2018. Convolutional neural networks with alternately updated clique. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2413–2422, URL: https://api.semanticscholar.org/CorpusID:4623288.

Yao, Y., Chen, T., Xie, G., Zhang, C., Shen, F., Wu, Q., Tang, Z., Zhang, J., 2021. Non-salient region object mining for weakly supervised semantic segmentation. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 2623–2632, URL: https://api.semanticscholar.org/CorpusID:232380138.

Yim, J., Joo, D., Bae, J.-H., Kim, J., 2017. A gift from knowledge distillation: Fast optimization, network minimization and transfer learning. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 7130–7138, URL: https://api.semanticscholar.org/CorpusID:206596723.

Zagoruyko, S., Komodakis, N., 2016. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. ArXiv, arXiv:1612.03928, URL: https://api.semanticscholar.org/CorpusID:829159.

Zeiler, M.D., Fergus, R., 2013. Visualizing and understanding convolutional networks. ArXiv, arXiv:1311.2901, URL: https://api.semanticscholar.org/CorpusID:3960646.

Zhang, S., Hu, X., 2020. Polarized light micrograph dataset of late cretaceous-eocene rock thin sections from western tarim basin, xinjiang. China Sci. Data 5, 1–10. http://dx.doi.org/10.11922/csdata.2020.0010.zh, URL: http://csdata.org/p/7/417/.

Zhang, X., Li, Z., Loy, C.C., Lin, D., 2016. PolyNet: A pursuit of structural diversity in very deep networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 3900–3908, URL: https://api.semanticscholar.org/CorpusID:17265670.

Zheng, D., Hou, L., Hu, X., Hou, M., Dong, K., Hu, S., Teng, R., Ma, C., 2024a. Sediment grain segmentation in thin-section images using dual-modal vision transformer. Comput. Geosci. 191, 105664.

Zheng, Z., Ye, R., Wang, P., Ren, D., Zuo, W., Hou, Q., Cheng, M.-M., 2022. Localization distillation for dense object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9407–9416.

Zheng, D.-Y., Zhong, H., Camps-Valls, G., Cao, Z., Ma, X., Mills, B.J.W., Hu, X., Hou, M., Ma, C., 2023. Explainable deep learning for automatic rock classification. Comput. Geosci. 184, 105511, URL: https://api.semanticscholar.org/CorpusID:266473104.

Zheng, D., Zhong, H., Camps-Valls, G., Cao, Z., Ma, X., Mills, B., Hu, X., Hou, M., Ma, C., 2024b. Explainable deep learning for automatic rock classification. Comput. Geosci. 184, 105511.

Zhou, Z.-H., 2018. A brief introduction to weakly supervised learning. Natl. Sci. Rev. 5, 44–53, URL: https://api.semanticscholar.org/CorpusID:44192968.

Zhou, Y., Guo, Z., Dong, Z., Yang, K., 2023a. Tensorrt implementations of model quantization on edge SoC. In: 2023 IEEE 16th International Symposium on Embedded Multicore/Many-Core Systems-on-Chip. MCSoC, pp. 486–493, URL: https://api.semanticscholar.org/CorpusID:262093141.

Zhou, B., Khosla, A., Lapedriza, À., Oliva, A., Torralba, A., 2015. Learning deep features for discriminative localization. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 2921–2929, URL: https://api.semanticscholar.org/CorpusID:6789015.

Zhou, Y., Lyu, K., Rawat, A.S., Menon, A.K., Rostamizadeh, A., Kumar, S., Kagy, J.-F., Agarwal, R., 2023b. DistillSpec: Improving speculative decoding via knowledge distillation. ArXiv, arXiv:2310.08461, URL: https://api.semanticscholar.org/CorpusID:263909387.

Zhou, Y., Ren, H., 2012. Segmentation method for rock particles image based on improved watershed algorithm. In: 2012 International Conference on Computer Science and Service System. pp. 347–349, URL: https://api.semanticscholar.org/CorpusID:6815846.

Zhou, Y., Starkey, J., Mansinha, L., 2004. Segmentation of petrographic images by integrating edge detection and region growing. Comput. Geosci. 30, 817–831, URL: https://api.semanticscholar.org/CorpusID:31706697.