A Decoupled Segmentation-Classification Strategy Based on Semantic-SAM for Precise Semantic Segmentation in Coal Mine Areas

Libing Wang, *Student Member, IEEE*, Jihong Dong, *Member, IEEE*, Lei Wang, Feng Liu, Qingke Wen, and Bekir Genc

***Abstract*—To address complex semantic segmentation in coal mine areas, this study proposes the SAM-SEF (SAM-based Semantic Enhancement Framework). It integrates Semantic-SAM's zero-shot segmentation capability with specialized deep learning models through a "segmentation-classification-update" strategy. Semantic-SAM initially provides category-agnostic multi-granularity segmentations; specialized deep learning models then perform semantic understanding of these regions, and a category mapping mechanism generates precise results. Evaluating 18 models, SAM-SEF-ENet-B0 achieved the highest IoUs for Open-pit Coal Mining Sites (OCMS) (93.68%) and Composite Coal-related Sites (CCS) (81.52%), with an mIoU of 87.60%, significantly outperforming traditional models. A Semantic-SAM-based semi-automatic annotation tool reduced manual annotation time by approximately 60%. A multi-scale evaluation system (based on 33% and 66% target area percentiles) revealed a positive correlation between target scale and segmentation performance, with small-scale target segmentation being the primary challenge (IoUs of 0.79 for OCMS and 0.67 for CCS in small scales). Systematic experiments validated Semantic-SAM's superiority over SAM (mIoU 54.03%) and SAM2 (mIoU 45.33%) in semantic understanding, architectural adaptability, and computational efficiency. The study also found that classification networks within the SAM-SEF framework outperformed specially designed segmentation networks. This demonstrates the effectiveness of a "classification-driven segmentation" paradigm, where lightweight classifiers operating on purified targets from a frozen foundation model can surpass traditional end-to-end segmentation networks in specific, complex domains, offering new technical solutions for remote sensing monitoring.**

***Index Terms*—Coal mine site segmentation, deep learning, multiscale evaluation, remote sensing image segmentation, semantic-SAM, semi-automatic annotation.**

Received 31 March 2025; revised 16 May 2025; accepted 29 May 2025. Date of publication 3 June 2025; date of current version 23 June 2025. This work was supported in part by the Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant KYCX24_2828, in part by the Deep Earth Probe and Mineral Resources Exploration-National Science and Technology Major Project under Grant National Key R&D Program of China 2024ZD1004104, in part by the Graduate Innovation Program of China University of Mining and Technology under Grant 2024WLKXJ177, and in part by the National Energy Group Ningxia Coal Industry Company, Ltd. Enterprise Project under Grant [2023]016. *(Corresponding author: Jihong Dong.)*

Libing Wang is with the School of Environment and Spatial Informatics, China University of Mining and Technology and Engineering Research Center of Mine Ecological Restoration, Ministry of Education, Xuzhou 221116, China (e-mail: tb22160013a41@cumt.edu.cn).

Jihong Dong is with the School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China (e-mail: dongjihong@cumt.edu.cn).

Lei Wang and Feng Liu are with China Coal Society, Beijing 100013, China (e-mail: lei9118@126.com; mtkjlf@sina.com).

Qingke Wen is with the National Engineering Research Center for Geomatics, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China (e-mail: wenqk@aircas.ac.cn).

Bekir Genc is with the School of Mining Engineering, University of the Witwatersrand, Johannesburg 2050, South Africa (e-mail: bekir.genc@wits.ac.za).

Digital Object Identifier 10.1109/JSTARS.2025.3576285

I. Introduction

WITH the deepening of the digital transformation in the global energy industry, the demand for refined monitoring and scientific management of coal mine areas has become increasingly urgent [1], [2], [3]. Coal mining activities play a crucial role in economic development but inevitably lead to geological environmental destruction, ecological degradation, and surface deformation, posing threats to the living environment and infrastructure safety of surrounding regions [4], [5]. In this context, remote sensing technology, with its advantages of wide coverage, flexible acquisition cycles, and relatively low monitoring costs, has gradually become an important means for environmental regulation and resource assessment in coal mine areas [6], [7], [8], [9]. The accessibility of multisource and multitemporal data from satellite and UAV remote sensing has significantly enhanced the capability for real-time dynamic monitoring of open-pit coal storage yards, underground mining areas, and related infrastructure [10], [11], [12].

However, performing accurate semantic segmentation on remote sensing imagery presents significant general challenges. These include large variations in the scale of objects (e.g., small auxiliary buildings versus vast open pits), high intraclass variance (different appearances within the same land cover type) combined with low interclass variance (similar appearances between different types), the presence of complex background clutter and frequent occlusions, and often ambiguous or transitional boundaries between adjacent land cover types [13], [14]. These challenges are particularly intensified in the context of coal mine areas. Specifically, coal mine environments exhibit highly dynamic landscapes due to continuous excavation, dumping, and reclamation activities, leading to rapid changes in topography and object morphology that necessitate frequent updates [15]. Furthermore, these areas typically feature a complex mixture of diverse natural and man-made elements, such as extraction pits, waste dumps, coal stockpiles, processing plants, transportation networks, vegetation patches, and water bodies, often co-existing in close spatial proximity [15]. A significant hurdle arises from the substantial spectral and textural similarities among key components like coal itself, dark overburden materials, shadows cast by terrain or structures, and certain types of bare soil or rock formations, complicating their accurate discrimination using standard image analysis techniques [16]. Consequently, there is a critical need for efficient, accurate, and regularly updated monitoring methods to effectively manage operational logistics, assess environmental impacts, and ensure regulatory compliance [17]. For mining areas requiring large-scale and high-precision monitoring, traditional manual interpretation and annotation are not only time-consuming and labor-intensive but also inadequate for fully capturing multiscale and multiangle scene changes [18]. Therefore, efficiently and accurately automating the semantic segmentation of multiple target categories in coal mine sites has become a key research focus in the fields of remote sensing and geosciences [19], [20], [21].

To meet the demands of large-scale automated processing of remote sensing images, deep learning models have demonstrated powerful learning and representation capabilities in tasks such as object detection and semantic segmentation [22], [23]. Early approaches predominantly employed convolutional neural networks (CNNs), such as ResNet, DenseNet, or UNet, which extract image features through multiple layers of convolution and pooling, achieving significant results in object detection and land cover classification in remote sensing images [24], [25]. Subsequently, to better address challenges such as large-scale variation of target objects and complex scene structures in high-resolution remote sensing images, researchers introduced transformer architectures, multiscale feature fusion, and attention mechanisms to comprehensively model multisource heterogeneous data [26], [27]. For example, Wang et al. [28] proposed YOLOFIV, which enables fast and accurate object detection in aerial remote sensing images, while Ren et al. [29] effectively enhanced the localization accuracy of remote sensing objects based on a multiscale feature aggregation network. Hou et al. [30] achieved high segmentation accuracy in building damage recognition by implementing multiscale feature fusion on the basis of U-Net. However, these mainstream deep learning-based methods generally rely heavily on large-scale, high-quality pixel-level annotated data, leading to high annotation costs and time investment [31]. This is particularly challenging in coal mine areas with variable terrain and complex visual information, making it difficult to construct comprehensive training datasets. Additionally, significant spectral and texture differences exist among different scales of coal storage yards, mining transportation channels, and related buildings, resulting in models often struggling to maintain stable segmentation performance when faced with cross-regional or multitemporal scenes [32], [33].

In recent years, the concept of "zero-shot segmentation" has gradually emerged, with new models represented by segment anything model (SAM) [34] showing considerable application prospects across various fields. In agricultural remote sensing, researchers have attempted to combine SAM with temporal analysis for intelligent monitoring of soybean growth processes and have achieved accurate classification of soybean populations using multimodal information [35]. In ship detection, SAM has demonstrated high robustness in detecting small targets against single and complex backgrounds through focused detection and fine segmentation [36], [37]. Some scholars have adapted SAM to specialized tasks such as building recognition and urban boundary extraction in high-resolution remote sensing images by employing finetuning strategies, introducing prior knowledge, or modifying the architecture [38], [39]. However, literature [40], [41] point out that the original SAM, primarily trained on natural image datasets, tends to produce coarse masks and overly smooth boundaries in land cover scenes with low contrast and high texture similarity. This issue is particularly pronounced in complex terrains like coal mine areas, where the loss of large-scale high-frequency information often leads to insufficient segmentation performance for targets such as surface subsidence and open-pit coal storage yards. Therefore, enhancing SAM's ability to maintain high adaptability in small-sample scenes while significantly improving its performance in handling the complexity of coal mine scenarios is a crucial issue in current remote sensing segmentation research.

To overcome the aforementioned challenges, this study proposes a precise semantic segmentation scheme for coal mine areas by combining Semantic-SAM [42] with specialized deep learning models, aiming to fully leverage the potential of zero-shot segmentation and conduct customized optimizations tailored to the characteristics of coal mine scenes. Specifically, we constructed the SAM-based semantic enhancement framework (SAM-SEF), whose core idea lies in a decoupled "segmentation-classification-update" strategy. This approach first leverages the powerful, category-agnostic, multigranularity segmentation capability of a pretrained and frozen Semantic-SAM to generate high-quality initial object boundary proposals without requiring task-specific segmentation annotations. Then, it utilizes specialized, potentially off-the-shelf deep learning classification models to perform accurate semantic discrimination on purified target regions derived from these initial masks. Finally, through an efficient mapping and update mechanism, it integrates the semantic labels with the precise boundaries to generate the final fine-grained segmentation map. This decoupled strategy offers flexibility, reduces dependency on extensive pixel-level annotations typically needed for training end-to-end segmentation models or fine-tuning foundation models [43], [44], and allows optimization of the segmentation and classification components somewhat independently.

The main contributions of this study are summarized as follows:

1) Development of the SAM-SEF. This framework introduces an efficient decoupled "segmentation-classification-update" strategy. It distinctively integrates the zero-shot, category-agnostic, multigranularity segmentation capability of a frozen Semantic-SAM for precise boundary delineation with the semantic understanding capability of specialized, often lightweight, deep learning classifiers operating on purified target samples derived from Semantic-SAM masks. This approach avoids the need for finetuning SAM and offers flexibility in choosing classifiers, aiming for high accuracy with reduced annotation effort in complex domains like coal mine areas.

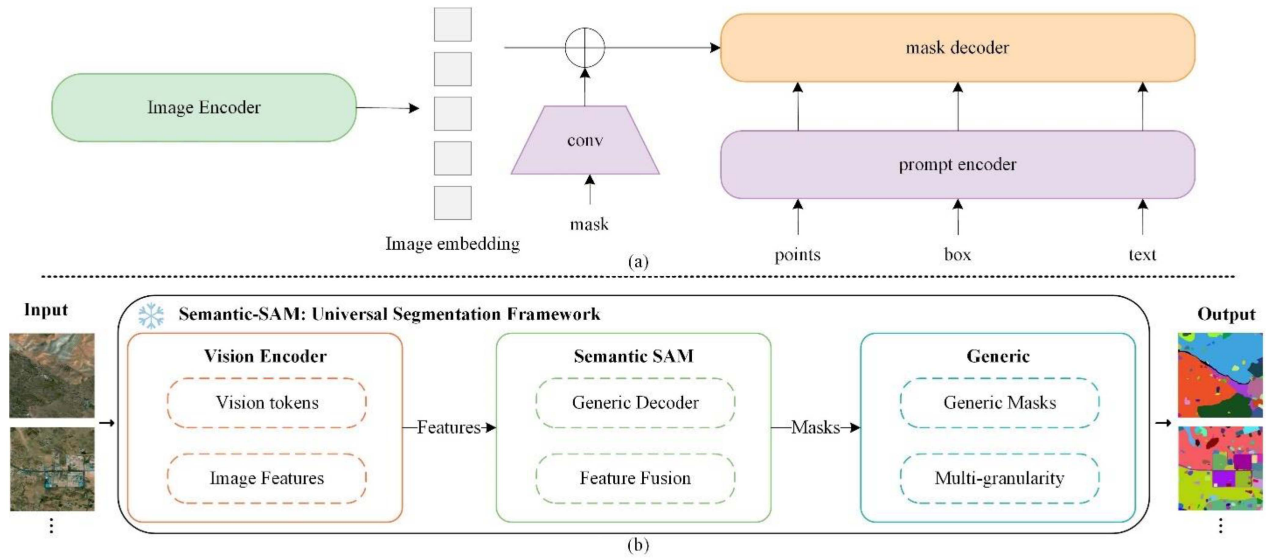

Fig. 1. SAM-based coal mine scene segmentation framework. (a) Basic architecture of SAM. (b) Multigranularity segmentation process based on Semantic-SAM in this study.

2) Construction of a semantic segmentation system for coal mine areas. Developed a semi-automatic annotation tool based on Semantic-SAM, enhancing annotation efficiency through presegmentation and interactive editing functions (reducing manual annotation time by approximately 60%); established a multiscale evaluation system based on the percentile of target area, analyzing the segmentation performance of targets at different scales. These tools and systems support the application and rigorous evaluation of the proposed SAM-SEF framework.
3) Validation of the effectiveness of the semantic segmentation method based on Semantic-SAM. Through comparative experiments with models such as SAM and SAM2, we analyzed the performance of Semantic-SAM in semantic understanding capability, architectural adaptability, and computational efficiency, justifying its selection. Crucially, the study demonstrates the effectiveness of the SAM-SEF framework and its decoupled strategy, showing that classification networks operating within this framework can achieve superior segmentation performance compared to traditional end-to-end segmentation networks in this specific task, highlighting the potential of this "classification-driven segmentation" paradigm when high-quality initial segment proposals can be obtained.

The remainder of this article is organized as follows: Section II introduces the theoretical foundations of the SAM model and its application progress in the remote sensing field; Section III elaborates on the precise recognition method for coal mine area scenes based on Semantic-SAM; Section IV validates the effectiveness of the proposed method through experiments and discussions; and finally, Section V presents the conclusions and future prospects.

II. Technical Theories and Methodological Foundations

A. Theoretical Foundations of the SAM Model

SAM [34] is a universal image segmentation foundational model based on the transformer architecture. As illustrated in Fig. 1(a), the core architecture of SAM comprises three key components: the image encoder, the prompt encoder, and the mask decoder. The image encoder adopts a pretrained vision transformer (ViT) structure, transforming the input image into a sequence of high-dimensional feature representations known as image embeddings through a self-attention mechanism. The prompt encoder is capable of flexibly handling various forms of input prompts, including points, boxes, and text descriptions, encoding these prompts into unified embedding representations. The mask decoder utilizes a bidirectional attention mechanism to integrate image features with prompt information, generating segmentation masks for target objects. Specifically, the convolutional (conv) module processes mask inputs, which are then fused with image features through additive operations to achieve precise segmentation of target regions.

However, in specialized applications such as remote sensing image analysis, SAM's interactive prompt mechanism has limitations. Coal-related sites in remote sensing images (e.g., open-pit coal mines and above-ground parts of underground coal mines) feature complex spatial structures and blurred boundaries, making it challenging to accurately capture target regions solely through interactive prompts like clicks or boxes. To address this issue, SAM introduces a global segmentation strategy, implemented through three key technical modules: grid prompt generation, multimask prediction, and pos-processing optimization.

During the grid prompt generation phase, the model first divides the input image into a regular 32×32 grid, generating foreground point prompts at each grid intersection. This dense grid sampling strategy ensures comprehensive coverage of image content, enabling the model to detect target objects of varying scales and locations. To handle nested objects and ambiguous cases within the image, SAM's mask decoder is designed to output multiple candidate masks (typically three) for a single prompt point. This multimask prediction mechanism is trained by minimizing the loss function, where only the loss of the prediction most matching the ground truth mask is calculated.

In the postprocessing optimization phase, SAM employs a complete quality control mechanism: first, high-quality masks are filtered by setting confidence thresholds; secondly, the stability of mask predictions is ensured by comparing mask consistency under different probability thresholds ($0.5\pm\varepsilon$); and finally, nonmaximum suppression algorithms are applied to remove duplicate mask predictions. To enhance the segmentation of small targets, the model also adopts a multiscale processing strategy, capturing more detailed features by analyzing overlapping enlarged regions. SAM is pretrained on the SA-1B dataset (comprising 11 million images and 1.1 billion masks) using an iterative data engine strategy. This training approach endows the model with strong zero-shot transfer capabilities, allowing it to adapt to various new scenes and tasks, demonstrating segmentation quality approaching that of manual annotations in evaluations. However, for specialized sites such as coal mine areas, relying solely on SAM's general segmentation capabilities is insufficient for achieving precise segmentation. Enhancing SAM with more robust semantic understanding and multigranularity analysis capabilities is necessary.

B. Theoretical Foundations of the Semantic-SAM Model and Global Segmentation Enhancement

Although SAM has achieved significant results in global segmentation tasks, its output lacks semantic understanding capabilities, particularly in specialized domain applications. Semantic-SAM [42] enhances the model's semantic understanding by introducing semantic awareness and multigranularity segmentation capabilities while retaining SAM's original segmentation performance. This model employs a unified framework to handle various segmentation tasks, providing foundational support for fine-grained segmentation in complex scenes.

In terms of model architecture, Semantic-SAM adopts a query-based Mask Decoder design that supports multiple prompt types. For point prompts, the model uniformly represents them in an anchor box format (x, y, w, h), where w and h are small values approximating the point's location. To capture masks at different granularities, each click is first encoded as a positional prompt and then combined with K different content prompts, where $K = 6$ corresponds to up to six hierarchical masks per click in the SA-1B dataset. Specifically, the representation of the ith query is as follows:

$$\mathbf{q}_i = \mathbf{q}_i^{\text{level}} + \mathbf{q}_i^{\text{type}} \tag{1}$$

where $\mathbf{q}_i^{\text{level}}$ is the embedding for the ith granularity level, and $\mathbf{q}_i^{\text{type}}$ is the embedding for the click or box type. The output of the mask decoder can be represented as follows:

$$O = \text{DeformDec}\,(Q, b, F) \;\text{ with } O = (o_1, \ldots, o_K) \tag{2}$$

where DeformDec is the deformable decoder, Q represents the query features, b denotes the reference boxes, and F stands for the image features. Each output $o_i = (c_i, m_i)$ includes the predicted semantic category c_i and the mask m_i.

In terms of semantic recognition, Semantic-SAM innovatively proposes a decoupled method for object and part classification. The model employs a shared text encoder to separately encode object and part information, enabling the model to independently perform object and part segmentation tasks. This design facilitates finegrained segmentation in specialized scenes but still requires optimization for specific applications in practice. Regarding multigranularity segmentation, Semantic-SAM overcomes SAM's one-to-many matching limitation through a multiselect learning scheme. This improvement allows the model to generate multiple levels of segmentation results from a single prompt point, providing rich candidate regions for subsequent specialized scene segmentation. Particularly in the segmentation of coal-related sites, this multigranularity segmentation capability aids in capturing target features at various scales.

To address the limitations of SAM in professional scene segmentation, this study employs the Semantic-SAM model for coal mine area scene segmentation, as shown in Fig. 1(b). In this application, input images of coal mine sites are first processed by a Vision Encoder to extract visual features, including vision tokens and image features. Subsequently, the pretrained and frozen Semantic-SAM model processes these features, where the generic decoder and feature fusion modules work in tandem to generate category-agnostic multigranularity segmentation results (generic masks). The multigranularity segmentation capability of Semantic-SAM is particularly crucial for segmenting coal-related sites, as it helps capture target features at different scales, providing rich candidate regions for subsequent finegrained segmentation.

C. Integration of General Image Segmentation Models and Conventional Neural Network Models

With the widespread application of SAM in computer vision, researchers have explored various approaches to integrating SAM with conventional neural network models to enhance performance on specialized tasks. Existing studies primarily focus on three aspects: serial processing architectures, feature fusion mechanisms, and complementary enhancement strategies.

In the context of remote sensing semantic segmentation, adapting foundation models like SAM has become a significant research direction. Recent efforts to leverage SAM in remote sensing often involve strategies like finetuning, using adapters, fusing features, or employing advanced prompt engineering. For instance, some studies focus on finetuning the entire SAM model or parts of it on specific remote sensing datasets to improve performance for tasks such as building extraction or detailed land cover mapping [43], [41]. While effective, this approach typically requires substantial labeled data for the target task and may potentially compromise the model's original generalization capabilities. Another strategy involves using lightweight adapter modules [44]. These modules are inserted into the SAM architecture and trained on the target task data, allowing adaptation with fewer trainable parameters compared to full finetuning, aiming to strike a balance between task-specific performance and model efficiency [45]. Feature fusion techniques represent another avenue, where features extracted by SAM's image encoder, or the segmentation masks generated by its decoder, are combined with features or outputs from other network architectures, such as YOLO or other specialized remote sensing models [35], [46]. The goal here is often to leverage SAM's powerful spatial localization abilities while incorporating complementary contextual or spectral information from other models [47]. Furthermore, research has explored various prompt engineering methods, utilizing points, bounding boxes, or even outputs from object detectors or other segmentation models as prompts to guide SAM more effectively towards segmenting specific objects of interest within complex remote sensing imagery [47], [48].

Our proposed SAM-SEF framework adopts a different approach compared to these methods. Instead of finetuning [43], [41] or using adapters [44], we utilize a pretrained and frozen Semantic-SAM [42] directly for its zero-shot, category-agnostic segmentation capabilities. Unlike direct feature fusion methods [35], [46], we employ a decoupled strategy where specialized classifiers operate on purified target regions generated based on SAM's masks. This differs from prompt engineering [47], [48] as it does not rely on interactive or predefined prompts for specific objects but rather processes the outputs of SAM's automatic mask generation. Beyond these SAM adaptation studies directly targeting segmentation, the field of remote sensing image analysis has seen other advanced representation and reasoning methodologies. For example, the dynamic spatio-temporal graph reasoning method for videoQA proposed by Nie et al. [49] excels in handling temporal changes and event recognition. Similarly, the cross-modal progressive perspective matching network for image-text retrieval developed by Zheng et al. [50] has achieved significant progress in multimodal information fusion and understanding. Although the direct application scenarios of these methods differ from our current study, their ideas on complex scene modeling, spatio-temporal information utilization, and multisource data fusion offer valuable insights for intelligent remote sensing image interpretation and point towards potential future research directions.

Despite the significant progress achieved thus far, multiple challenges remain in harnessing SAM's segmentation capabilities and achieving efficient feature fusion in complex industrial scenarios such as coal mine areas. First, it is necessary to maintain the model's architectural simplicity while fully leveraging SAM's advantages in multiscale object segmentation. Second, high segmentation accuracy should be achievable with only a small number of annotated samples, thereby reducing the reliance on large annotated datasets. Finally, it is crucial to preserve SAM's generalization capabilities in a simplified model structure to achieve precise segmentation of complex scenes. Addressing these issues is key to building an efficient and practical scene segmentation system for coal mine areas.

III. Precise Scene Recognition Method for Coal Mine Areas Based on Semantic-SAM

A. *Overview of the SAM-SEF Framework*

The SAM-SEF provides a systematic solution based on a "segmentation-classification-update" procedure, wherein SAM not only represents the Segment Anything Model but also encompasses models such as Semantic-SAM and other general image segmentation foundational models with functionalities akin to SAM. By combining the segmentation capabilities of these general segmentation models with the scene understanding capabilities of specialized deep learning models, this framework effectively tackles the challenges of segmenting complex scenes.

As shown in Fig. 2, the overall technical workflow of this framework consists of five key steps. Step 1 involves data preprocessing, where Sentinel-2 multispectral imagery is patched into 512 $\times$ 512 pixel segments to construct datasets for two types of sites, OCMS and CCS. In step 2, initial semantic segmentation is performed using Semantic-SAM for unsupervised, category-agnostic multi-granularity segmentation to generate initial segmentation results. Step 3 focuses on sample annotation and target purification, sample annotation using an annotation tool and standardizing target regions to 224 $\times$ 224 pixels. Step 4 encompasses deep learning classification, where models are trained and selected based on multiple backbone networks (e.g., ResNet18, EfficientNet-B0 to B7, MobileNetV2, etc.). Finally, step 5 involves result fusion and update, establishing a category-value mapping table based on model predictions and integrating the predicted information with the initial segmentation results to produce the final precise segmentation image.

The core of the SAM-SEF framework lies in its decoupled strategy. This design intentionally separates the challenge of precise boundary delineation from semantic labeling. It leverages the powerful zero-shot segmentation ability of a foundation model (Semantic-SAM) to handle the complex task of finding object boundaries accurately without needing task-specific training for this stage. Subsequently, it simplifies the semantic labeling task by feeding purified, object-focused patches to standard, potentially pre-trained, classification models. It is important to acknowledge that the performance of the downstream classification models inherently depends on the quality of the initial segmentation masks generated by Semantic-SAM; errors in initial segmentation can propagate and affect final accuracy. However, the selection of Semantic-SAM, known for its superior segmentation capabilities (as demonstrated in Section IV-F), and the target purification step aim to minimize such errors and provide robust inputs for classification. This strategic division aims to improve overall accuracy and efficiency, especially in data-scarce or complex environments, compared to training large end-to-end models or requiring extensive data for finetuning the foundation model itself. The "zero-shot segmentation" capability primarily refers to the initial mask generation by the frozen Semantic-SAM, while the subsequent specialized deep learning classifiers do require labeled data (patch-level

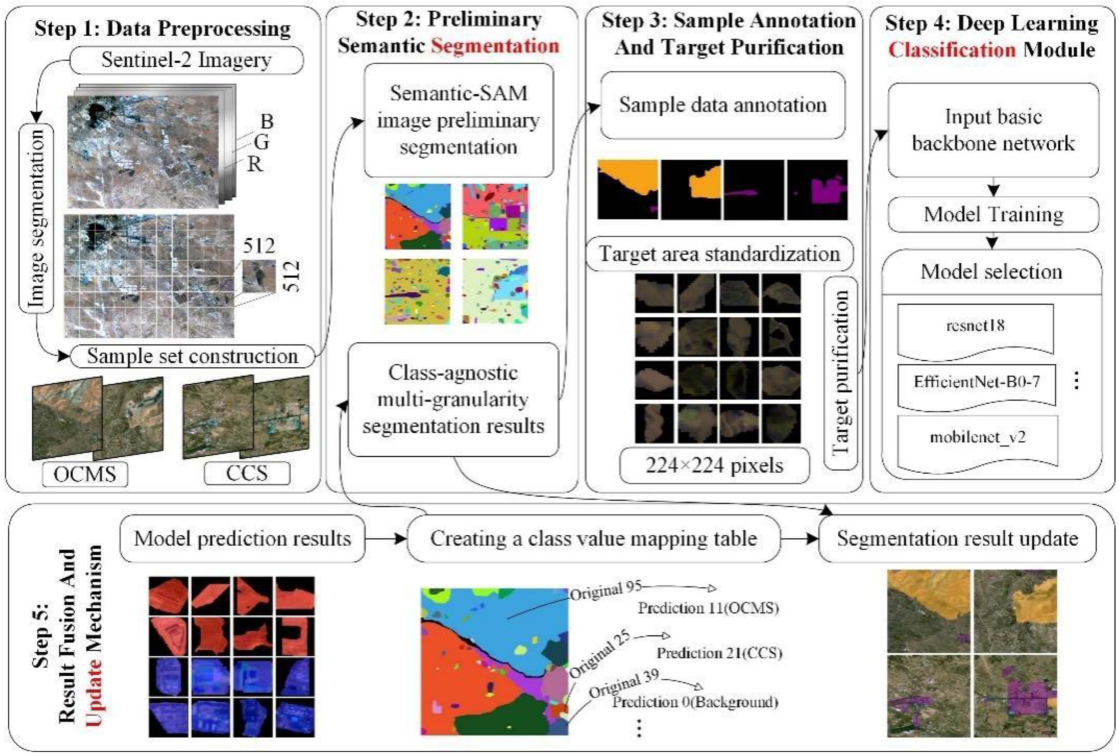

Fig. 2. Technical workflow of the SAM-SEF framework. The framework comprises five main steps: Data preprocessing, initial semantic segmentation, sample annotation and target purification, deep learning classification, and result fusion & update.

class labels, not pixel-level segmentation labels) for training, as detailed in Section III-B.

A specific instantiation of the SAM-SEF framework, SAM-SEF-ResNet18, can be mathematically expressed as follows:

$$\mathcal{F}_{\mathrm{SAM-SEF}}(I) = \Psi\left(\mathcal{S}_{\mathrm{sam}}(I), \mathcal{A}_{\mathrm{resnet18}}(I)\right) \tag{3}$$

where I is the input image (a 512×512-pixel Sentinel-2 patch in this study), $\mathcal{S}_{\mathrm{sam}}$ denotes the segmentation operation of SAM to generate category-agnostic multigranularity segmentation masks, and $\mathcal{A}_{\mathrm{resnet18}}$ represents the semantic analysis operation by ResNet18 (which can be substituted by other backbone architectures) acting on standardized 224×224-pixel samples. Ψ is the semantic enhancement function that fuses the segmentation results with the category information.

The framework's core innovations lie in three aspects: the target purification mechanism, the semantic analysis strategy, and the category mapping mechanism. The target purification mechanism (detailed in Section III-C-3) utilizes the masks from Semantic-SAM to extract and standardize target regions, effectively removing background clutter and allowing the classifier to focus solely on the object's features, thereby enhancing classification accuracy and efficiency. In the semantic analysis strategy, deep learning models classify the standardized samples using a cross-entropy loss function, as elaborated in Section III-D. The category mapping mechanism (detailed in Section III-E) ensures accurate integration of the classification results back into the original image space, leveraging unique identifiers for precise pixel-level updates. By leveraging its generality, efficiency, adaptability, and systematic nature, the SAM-SEF framework demonstrates remarkable advantages in semantic segmentation tasks involving complex industrial environments. Integrating SAM's fundamental segmentation capability with specialized models' semantic understanding, it is particularly suited for precise segmentation of open-pit coal mining sites (OCMS) and Composite coal-related sites (CCS). This systematic workflow ensures high segmentation granularity while enhancing scene segmentation accuracy, thereby effectively resolving the challenge of precise segmentation in the complex contexts of coal mine areas.

B. Data Preprocessing and Sample Construction

During the data preprocessing phase, this study selected Sentinel-2 multispectral satellite imagery as the primary data source. The data were obtained from the Google Earth Engine platform using the COPERNICUS/S2_SR_HARMONIZED dataset, covering the period from March to July 2024 and filtered for cloud cover below 10% . The dataset included three main bands—Blue (B), Green (G), and Red (R)—with a spatial resolution of 10 meters. For ease of processing and analysis, the raw imagery was patch-partitioned into 512×512 pixels using a sliding window approach, constructing typical sample sets of OCMS and CCS.

As shown in Fig. 3, the study area primarily encompasses two major coal production bases in China: the Ningdong coal base and the Shendong coal base. These regions were chosen for their representativeness, covering diverse mining conditions and geomorphological types. The Ningdong coal base is situated

TABLE I
STATISTICS OF THE COAL MINE SCENE DATASET

Scene Type	Total Samples (512 × 512)	Purified Samples (224 × 224)	Training and validation samples (224 × 224)	Test Samples (224 × 224)
Open-pit Coal Mining Site	201	3247	2946	301
Composite Coal-related Site	272	1531	1377	154

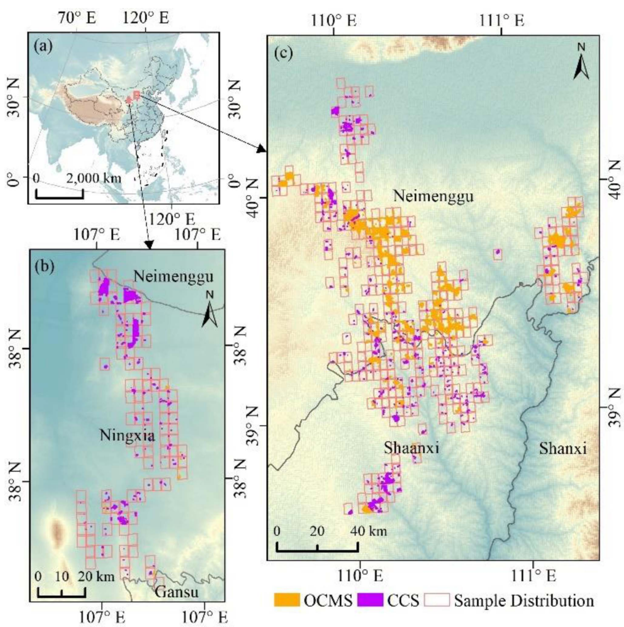

Fig. 3. Distribution of research area samples. (a) Geographic location of the study area; (b) distribution of samples in the Ningdong coal base located on the Ningxia-Gansu provincial boundary; (c) distribution of samples in the Shendong coal base spanning Shaanxi, Inner Mongolia, and Shanxi provinces. Orange areas represent OCMS (open-pit coal mining sites), purple areas represent CCS (composite coal-related sites), and the pink boxes indicate the spatial extent of the samples.

in eastern Ningxia near the Gansu provincial boundary, serving as one of China's critical coal production hubs. The Shendong coal base, China's largest coal production area, spans northern Shaanxi, southern Inner Mongolia, and northwestern Shanxi provinces. The samples in these two regions are highly representative, covering a variety of terrain conditions and mining methods.

As detailed in Table I, the constructed dataset comprises two representative site types. The OCMS include 201 original samples, which yielded 3247 standardized 224 × 224-pixel samples after target purification. The dataset is split 8:1:1 into training (2597 samples), validation (349 samples), and testing (301 samples). The CCS include 272 original samples, which yielded 1531 standardized samples, also divided 8:1:1 into training (1223), validation (154), and testing (154). The difference in sample size reflects the varying complexity of the two site types: OCMS sites contain more extractable target regions, whereas CCS sites are relatively scattered in spatial distribution. This dataset construction approach accounts for the unique characteristics of each site type, ensuring an adequately balanced and representative sample foundation for subsequent model training. The labels required for training the classifiers within the SAM-SEF framework are patch-level class labels assigned to the 224 × 224 purified samples, which is significantly less labor-intensive than generating pixel-level segmentation masks typically required for end-to-end segmentation models.

C. Semantic-SAM Segmentation and Sample Processing

1) Initial Segmentation via Semantic-SAM: After obtaining the preprocessed imagery, this study uses Semantic-SAM to perform an initial segmentation of the scenes. As a semantically enhanced version of SAM, Semantic-SAM not only inherits SAM's advantages in object segmentation but also incorporates semantic understanding capabilities. In the initial segmentation phase, no category information is introduced; the primary goal is to capture precise target boundaries. Semantic-SAM, as an advanced semantic segmentation model, can automatically identify and segment salient target regions in the image without requiring predefined category information. This unsupervised segmentation approach is particularly suited for the complex and varied nature of coal mine sites. In practice, Semantic-SAM processes each 512 × 512-pixel image patch to generate initial segmentation masks. These masks preserve accurate target boundary information and ensure spatial continuity through connected-component analysis. Preliminary experiments indicate that compared with traditional threshold- and edge-based segmentation methods or directly employing SAM, Semantic-SAM exhibits stronger robustness and adaptability in handling complex industrial scenes.

2) Development of an Annotation Tool: To enhance annotation efficiency and quality, this study developed an annotation tool based on Semantic-SAM. Leveraging

Semantic-SAM's semantic segmentation capabilities, the tool facilitates a "pre-segmentation–interactive annotation–post-processing" workflow for efficient annotation. Built with a layered architecture, the tool consists of a data management layer, a core processing layer, and an interactive presentation layer. The data management layer unifies the handling of raw images, pre-segmentation masks generated by Semantic-SAM, and annotation outcomes. At the core processing layer, the segmentation results from Semantic-SAM enable polygon contour editing, semantic category mapping, and region merging. The interactive presentation layer employs the PyQt5 framework to provide an intuitive user interface, supporting real-time visualization and annotation operations.

According to the framework expression $\mathcal{F}_{\mathrm{SAM-SEF}}(I)$ concerning the $\mathcal{S}_{\mathrm{sam}}(I)$ component, the core algorithm for annotation can be represented as follows: First, Semantic-SAM yields the segmentation outcome

$$\mathcal{S}_{\mathrm{sam}}(I) = M \tag{4}$$

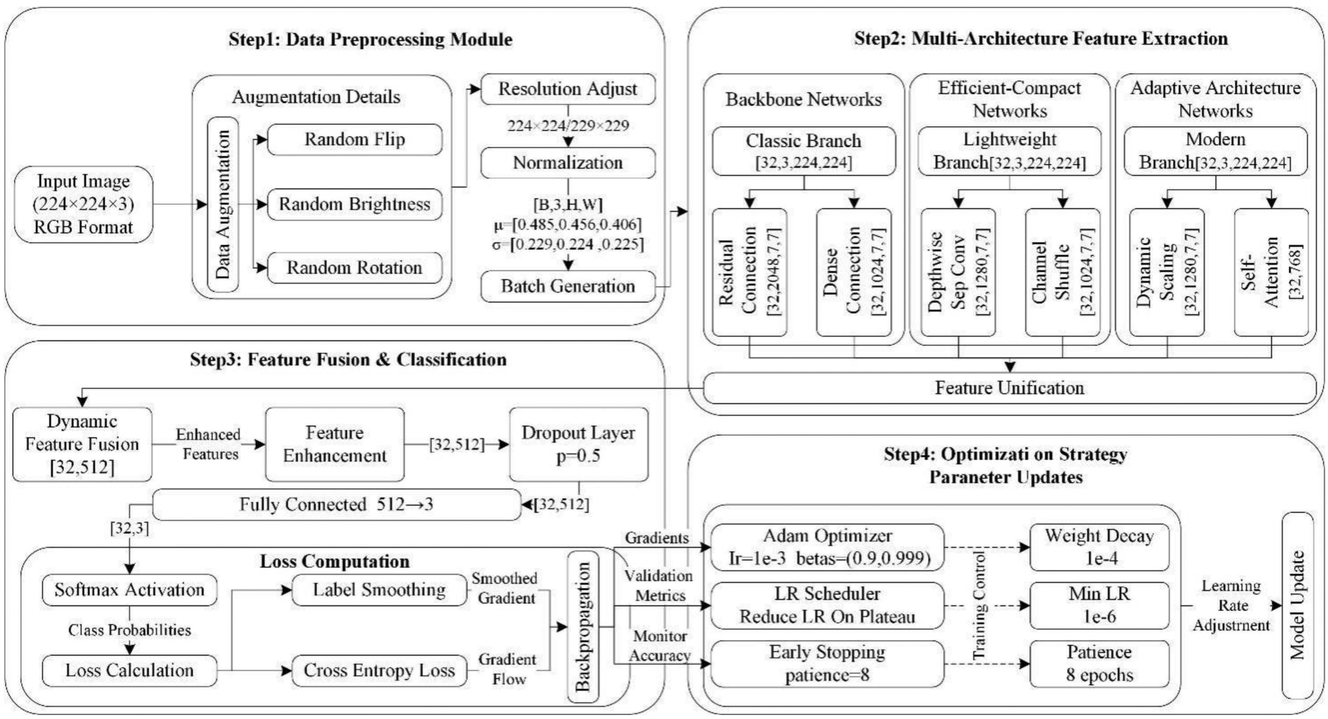

Fig. 4. Deep learning model training framework based on multiarchitecture feature extraction.

where M is the presegmentation mask. The annotated result L is then defined as follows:

$$L(x,y) = \begin{cases} c_k, & \text{if } (x,y) \in R_k \text{ and } M(x,y) = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Here, (x, y) denotes the pixel coordinates, R_k represents the kth connected region generated by Semantic-SAM, c_k is the user-assigned category label, and $M(x, y)$ is the value of the presegmentation mask at that location.

3) Target Purification Preprocessing: To further enhance the model's segmentation performance, this study proposes a target purification preprocessing approach for the sample set. This step is crucial within the SAM-SEF framework as it prepares focused inputs for the downstream classifier. According to the target purification mechanism in the framework expression $\mathcal{F}_{\text{SAM-SEF}}(I)$, for each target region I in the input image R_i, the standardized output O is computed as follows:

$$O(x,y) = \begin{cases} I'(sx, sy), & \text{if } (x,y) \in R'_i \text{ and } M(x,y) = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

In this formulation, I denotes the original input image (512×512 pixels), and I' is the interpolated image. The scaling factor is $s = \frac{W_i}{T}$, where (x, y) refers to pixel coordinates, R'_i is the scaled target region, and $M(x, y)$ is the segmentation mask from Semantic-SAM. This target purification process emphasizes preserving the essential spatial structural and textural features of the target object while significantly minimizing interference from irrelevant background pixels. Through interpolation algorithms, key texture information is retained during scaling, while eliminating background interference from nontarget regions significantly enhances feature purity. This preprocessing approach effectively improves recognition accuracy while reducing computational complexity, laying the groundwork for subsequent deployment of lightweight models.

D. Deep Learning Classification Module

1) Overall Architecture Design: This study designs a systematic deep learning model training framework, as illustrated in Fig. 4, comprising four main components: data preprocessing, feature extraction, classification, and optimization strategy. Such a modular design not only improves the maintainability of the model but also facilitates targeted optimizations and enhancements. According to the framework expression $\mathcal{F}_{\text{SAM-SEF}}(I)$ in the $\mathcal{A}_{\text{resnet18}}(I)$ section, these modules work in coordination to achieve accurate scene classification.

2) Data Preprocessing Module: The data preprocessing module implements a comprehensive data augmentation strategy. Random flips are applied horizontally and vertically to significantly increase the diversity of training data. A random brightness adjustment mechanism is introduced to enhance the model's ability to adapt to varying illumination conditions. Additionally, random rotations within a 15° range increase the model's robustness to changes in object orientation. Together, these data augmentation strategies substantially boost the model's generalization capability.

For image normalization, two resolution standards (224×224 pixels and 229×229 pixels) are used to resample images to accommodate the input requirements of different model architectures. The normalization process employs

ImageNet dataset statistics

$$I_{\text{norm}}(c,h,w) = \frac{I(c,h,w) - \mu_c}{\sigma_c} \tag{7}$$

where $I_{\text{norm}}(c,h,w)$ denotes the normalized image, $I(c,h,w)$ is the original image, (c,h,w) represent the channel, height, and width dimensions, respectively. The values $\mu_c = [0.485, 0.456, 0.406]$ and $\sigma_c = [0.229, 0.224, 0.225]$ are the mean and standard deviation for each channel, and $c \in \{R,G,B\}$ indicates the color channel. Finally, a batch generator is employed for efficient data organization and loading, providing a continuous and balanced data flow during model training.

3) Multiarchitecture Feature Extraction Module: This study compares the feature extraction capabilities of three typical categories of deep learning architectures. For an input feature map $X \in \mathbb{R}^{C\times H\times W}$, each architecture employs a distinct approach to feature extraction.

The backbone networks category encompasses classic architectures including the ResNet series (ResNet18/34/50), DenseNet121, and GoogleNet. According to the framework expression $\mathcal{F}_{\text{SAM-SEF}}(I)$ under the $\mathcal{A}_{\text{resnet18}}(I)$ component, ResNet18 serves as a representative with the following feature extraction process:

$$\mathcal{A}_{\text{resnet18}}(I) = X + \mathcal{F}(X). \tag{8}$$

For efficient-compact networks, which include lightweight architectures such as MobileNetV2, ShuffleNetV2, MNASNet, and Inception-v3, the framework's $\mathcal{A}_{\text{mobile}}(I)$ component is represented by MobileNetV2 that employs depthwise separable convolutions

$$\mathcal{A}_{\text{mobile}}(I) = \sigma(\text{BN}(\text{DW}(X))) \cdot \sigma(\text{BN}(\text{PW}(X))). \tag{9}$$

The adaptive architecture networks category features state-of-the-art architectures including the vision transformer series (ViT-B/16, ViT-B/32) and the EfficientNet series (B0–B7). Under the framework's $\mathcal{A}_{\text{vit}}(I)$ component, the vision transformer (ViT) leverages a self-attention mechanism:

$$\mathcal{A}_{\text{vit}}(I) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V. \tag{10}$$

Here, X is the input feature map, σ is the ReLU activation function, BN denotes batch normalization, DW and PW indicate depthwise and pointwise convolutions, respectively, and Q, K, and V are the query, key, and value matrices used in attention computations. The parameter d represents the feature dimensionality. These architectures strike different trade-offs between computational efficiency and representational power through their unique feature extraction mechanisms.

4) Feature Fusion and Classification Module: The classification module uses a standard classification approach to predict labels from extracted features. As the output component of $\mathcal{A}_{\text{resnet18}}(I)$, the classification process adopts cross-entropy loss with label smoothing:

$$\mathcal{L}_{\text{CE}} = -\sum_{k=1}^{K} y'^{k} \log(p_k) \tag{11}$$

where y'_k is the smoothed label

$$y'_k = \begin{cases} 1-\epsilon, & \text{if } k = y \\ \epsilon/(K-1), & \text{otherwise.} \end{cases} \tag{12}$$

In the above, $\mathcal{L}_{\text{CE}}$ is the cross-entropy loss value, k is the class index, K is the total number of classes, y'_k is the smoothed label value, p_k is the model-predicted probability for class k, y is the ground-truth class index, and ϵ is the label smoothing coefficient (set to 0.1 in this study).

5) Optimization Strategy Module: This study adopts the Adam optimizer for model training, with its core update steps given by the following:

$$\begin{aligned} m_t &= \beta_1\, m_{t-1} + (1-\beta_1)\, g_t \\ v_t &= \beta_2\, v_{t-1} + (1-\beta_2)\, g_t^2 \\ \theta_t &= \theta_{t-1} - \alpha \frac{m_t/\sqrt{v_t}}{\sqrt{1-\beta_1^t}/\sqrt{1-\beta_2^t}} \end{aligned} \tag{13}$$

where m_t and v_t denote the first- and second-order moment estimates, g_t is the current gradient, $\beta_1 = 0.9$ and $\beta_2 = 0.999$ are the momentum decay rates, α is the learning rate, and θ_t represents the model parameters.

A learning rate scheduler with early stopping is employed for adaptive learning rate adjustments

$$\alpha_t = \alpha_{t-1} \cdot \begin{cases} \gamma, & \text{if } n_{\text{plateau}} \geq p \\ 1, & \text{otherwise} \end{cases} \tag{14}$$

where $\gamma = 0.1$ is the decay factor, n_{plateau} is the number of consecutive epochs with no improvement in the validation metric, and $p = 8$ is the patience parameter.

E. Result Fusion and Updating Mechanism

In the final segmentation result update stage, the predictions from the deep learning model—applied to each purified 224 × 224-pixel target region—are associated with the Semantic-SAM segmentation results to generate a final 512 × 512-pixel segmentation image by building a category-value mapping table. This result fusion and updating mechanism constitutes a core part of the SAM-SEF framework, aiming to seamlessly integrate the precise segmentation capacity of Semantic-SAM with the semantic understanding provided by the deep learning model. The mechanism uses a "local-to-global" processing strategy: it first completes category prediction at the purified-sample level, then maps these predictions back onto the original segmented image, and finally refines boundaries for accurate segmentation.

According to the framework expression $\mathcal{F}_{\text{SAM-SEF}}(I)$, the semantic enhancement function Ψ drives the result fusion and updating process, represented as follows:

$$\Psi(\mathcal{S}_{\text{sam}}(I), \mathcal{A}_{\text{resnet18}}(I)) = M\left(I'_{i,j}\right). \tag{15}$$

Concretely, let $I \in \mathbb{R}^{h\times w}$ be the original segmented image matrix with $h = w = 512$. Define a category mapping function

$$M : C_{\text{orig}} \to C_{\text{new}} \tag{16}$$

where C_{orig} is the set of original categories (including the temporary labels generated by Semantic-SAM), and C_{new} is the updated set of categories (including two primary targets: OCMS and CCS). For each pixel $I_{i,j}$ in the image, the update rule is

$$I'_{i,j} = \begin{cases} M\left(I_{i,j}\right) & \text{if } I_{i,j} \in C_{\text{orig}} \\ I_{i,j} & \text{otherwise.} \end{cases} \tag{17}$$

The category mapping relationship is constructed based on

$$M\left(c\right) = \begin{cases} v & \text{if } v \in \mathcal{U}\left(I_c\right) \setminus \{0\} \\ 0 & \text{if } \mathcal{U}\left(I_c\right) \setminus \{0\} = \emptyset \end{cases} \tag{18}$$

where $\mathcal{U}(I_c)$ denotes the unique set of pixel values in the image corresponding to category c. This mapping guarantees a 100% accurate association between each classification result and its correct spatial location defined by the initial semantic SAM mask, preventing any misalignment or ambiguity.

Through this complete fusion and updating mechanism, the SAM-SEF framework achieves precise recognition of OCMS and CCS. Not only does the method retain Semantic-SAM's advantages in boundary segmentation, but the deep learning model also endows the segmentation results with accurate semantic information, providing reliable foundational data for subsequent scene analysis and monitoring.

IV. Experimental Results and Analysis

This section verifies the effectiveness of the Semantic-SAM-based precise semantic segmentation method for coal mining areas through a comprehensive experimental evaluation system. The evaluation system encompasses five interrelated dimensions: annotation tool application analysis, model benchmark performance comparison, boundary accuracy assessment, multiscale target segmentation capability analysis, and comparative validation with SAM series models. The efficiency of the annotation tool lays the foundation for constructing large-scale datasets, thereby supporting comprehensive performance comparisons with 18 deep learning models. Boundary accuracy and multiscale evaluations further validate the technical advantages of the model, while theoretical analysis and experimental results substantiate the rationale for selecting Semantic-SAM as the foundational model. This progressive evaluation system comprehensively demonstrates the superior performance of the proposed method in coal mining area semantic segmentation tasks.

A. Experimental Design and Evaluation System

To comprehensively evaluate model performance, this study employs multiple evaluation metrics for quantitative analysis. intersection over union (IoU) serves as the primary evaluation metric, effectively addressing class imbalance issues and being sensitive to segmentation boundary accuracy. IoUs are calculated separately for OCMS and CCS to assess model performance across different scene types. The F1 score measures the balance between precision and recall, particularly suitable for evaluating the segmentation of small industrial facilities. The mean intersection over union (mIoU) reflects the overall performance of the model in multiclass segmentation tasks. Pixel accuracy (PA) intuitively indicates the accuracy of pixel-level classification, especially useful for evaluating the segmentation of large homogeneous regions. The selection of these metrics considers the characteristics of coal mining area sites and provides a unified evaluation standard for in-depth analysis in subsequent sections.

The selection of comparison methods in this study was guided by a clear rationale aimed at thoroughly evaluating the proposed SAM-SEF framework. First, comparisons within SAM-SEF using different backbones (see Table II) were conducted to assess the framework's flexibility in accommodating various classification architectures and to determine the optimal configuration for the specific task at hand. Second, the performance of SAM-SEF was benchmarked against 17 diverse traditional semantic segmentation models (see Table III), widely used in remote sensing, to establish how the proposed strategy measures up against established end-to-end segmentation approaches. Finally, comparisons with the original SAM [34] and SAM2 [51] within the SAM-SEF framework (see Table IV) were performed to experimentally validate the choice of Semantic-SAM [42] as the most effective foundational segmentation component for the decoupled strategy, particularly in the context of coal mine site analysis.

Experiments were conducted on a workstation equipped with an NVIDIA Quadro RTX 6000 GPU (64 GB VRAM), an Intel(R) Core(TM) i9-10980XE CPU @ 3.00 GHz, and 128 GB RAM. The software environment included the Windows 10 Pro for Workstations (Version 22H2) operating system. The deep learning framework versions used were Torch 1.12.1+cu113, Torchaudio 0.12.1+cu113, TorchMetrics 1.5.1, and TorchVision 0.13.1+cu113, all compatible with CUDA 11.3. We utilized the official implementation of Semantic-SAM. Other key libraries included GDAL 3.4.1, OpenCV 4.5.5, scikit-learn 1.0.2, NumPy 1.21.5, Pandas 1.3.5, and Matplotlib 3.5.1. Fixed random seeds were used during model training to ensure reproducibility. Standard numerical stability techniques, such as monitoring loss functions during training with the Adam optimizer, were employed, and no significant numerical issues like NaN loss were observed. All computations were performed using single-precision floating-point (float32) numbers. To ensure comparability and reliability of results, all 18 deep learning models, including the ResNet [52] series (ResNet18, ResNet34, ResNet50), Vision Transformer [53] series (ViT-B/16, ViT-B/32), classical convolutional neural networks (GoogleNet [54], DenseNet121 [55], Inception-V3 [56]), lightweight networks (MobileNet-V2 [57], ShuffleNet-V2 [58]), and the EfficientNet [59] series (B0-B7), were trained and tested under identical experimental conditions.

The dataset was randomly split into training, validation, and testing sets in an 8:1:1 ratio. To avoid sample distribution bias, a stratified sampling strategy was employed, ensuring consistent proportions of OCMS and CCS scenes across subsets. The test set remained entirely independent and was used solely for final performance evaluation. All models were trained and tested under these identical conditions to ensure fair comparison. The training set was used for model parameter optimization, the validation set for hyperparameter tuning and early stopping, and

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT DEEP LEARNING MODELS IN COAL MINING AREA SEMANTIC SEGMENTATION TASKS

Framework	backbone	IoU(%)		F1(%)		mIoU(%)	PA(%)	F1(%)
		OCMS	CCS	OCMS	CCS			
SAM-SEF-ResNet18	ResNet18	85.40	76.32	92.12	86.57	80.86	84.52	89.35
SAM-SEF-ResNet34	ResNet34	82.89	75.46	90.65	86.01	79.18	82.26	88.33
SAM-SEF-ResNet50	ResNet50	86.62	75.75	92.83	86.20	81.18	85.44	89.52
SAM-SEF-ViT16	ViT-B/16	73.77	60.11	84.91	75.09	66.94	73.68	80.00
SAM-SEF-ViT32	ViT-B/32	73.95	58.72	85.02	73.99	66.33	73.69	79.51
SAM-SEF-GoogleNet	GoogleNet	84.28	78.37	91.47	87.87	81.32	84.08	89.67
SAM-SEF-DenseNet	DenseNet121	82.99	73.92	90.70	85.00	78.45	82.15	87.85
SAM-SEF-Inception	Inception-V3	80.19	75.86	89.01	86.27	78.03	79.87	87.64
SAM-SEF-MobileNet	MobileNet-V2	89.60	79.93	94.51	88.85	84.76	88.28	91.68
SAM-SEF-ShuffleNet	ShuffleNet-V2	90.52	68.72	95.02	81.46	79.62	88.60	88.24
SAM-SEF-ENet-B0	EfficientNet-B0	**93.68**	**81.52**	**96.74**	**89.82**	**87.60**	**92.04**	**93.28**
SAM-SEF-ENet-B1	EfficientNet-B1	91.42	74.69	95.52	85.51	83.05	90.13	90.51
SAM-SEF-ENet-B2	EfficientNet-B2	86.02	76.74	92.48	86.84	81.38	84.93	89.66
SAM-SEF-ENet-B3	EfficientNet-B3	87.91	75.92	93.57	86.31	81.92	86.48	89.94
SAM-SEF-ENet-B4	EfficientNet-B4	86.97	76.94	93.03	86.97	81.95	85.89	90.00
SAM-SEF-ENet-B5	EfficientNet-B5	86.41	72.18	92.71	83.84	79.29	85.80	88.27
SAM-SEF-ENet-B6	EfficientNet-B6	91.78	77.14	95.71	87.10	84.46	90.06	91.40
SAM-SEF-ENet-B7	EfficientNet-B7	85.30	77.11	92.07	87.08	81.21	84.54	89.57

Note: Bold numbers indicate the highest values for each metric.

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT SEMANTIC SEGMENTATION MODELS IN COAL MINING AREA SEMANTIC SEGMENTATION TASKS

Method	IoU(%)		F1(%)		mIoU(%)	PA(%)	F1(%)
	OCMS	CCS	OCMS	CCS			
UNet	73.38	74.91	84.65	85.66	74.14	74.08	85.15
ENet	84.63	76.02	91.67	86.37	80.32	84.13	89.02
ESNet	**85.97**	79.25	**92.46**	88.43	82.61	**85.84**	90.44
SQNet	68.11	75.37	81.03	85.96	71.74	70.51	83.49
CGNet	74.34	58.77	85.28	74.03	66.55	72.68	79.66
SegNet	87.12	75.03	93.12	85.73	81.07	86.43	89.42
LEDNet	77.68	**83.02**	87.44	**90.73**	**80.35**	79.21	**89.08**
ESPNet	69.74	74.63	82.17	85.47	72.18	71.87	83.82
FPENet	63.17	53.91	77.43	70.06	58.54	62.53	73.74
FSSNet	62.19	80.30	76.69	89.07	71.24	65.74	82.88
DABNet	75.01	61.23	85.72	75.95	68.12	73.31	80.84
EDANet	78.26	62.70	87.81	77.07	70.48	76.45	82.44
ERFNet	84.69	76.14	91.71	86.45	80.41	84.20	89.08
LinkNet	77.36	70.39	87.23	82.62	73.87	76.94	84.93
FastSCNN	60.75	53.08	75.58	69.35	56.92	61.26	72.47
ContextNet	58.92	63.17	74.15	77.43	61.05	61.34	75.79
DeepLabV3+	85.81	67.28	92.36	80.44	76.54	83.29	86.40

Note: Bold numbers indicate the highest values for each metric.

the test set solely for final performance evaluation. All datasets underwent a unified preprocessing pipeline, including image normalization and data augmentation.

B. Annotation Tool Application Analysis

To enhance the efficiency and quality of data annotation, we developed an interactive annotation tool based on Semantic-SAM. This tool leverages Semantic-SAM's presegmentation and semantic understanding capabilities to realize a semi- automated annotation workflow. Users can complete annotations by simply clicking on target regions presegmented by Semantic-SAM (indicated by white contours in Fig. 5). If partial annotations are unsatisfactory, the tool provides modification functionalities. As illustrated in Fig. 5, the annotation tool interface comprises an image display area, a category selection panel, and an operation control area, supporting batch image loading and saving/loading of annotation results.

The annotation tool utilizes Semantic-SAM's intelligent presegmentation to automatically generate initial segmentation masks, supporting automatic recognition and segmentation of multiscale sites. It offers mask transparency adjustment for better visualization of segmentation effects. In terms of interactive editing, the tool provides functionalities for adding, editing, and deleting categories, supporting real-time mask modifications and finetuning, along with undo/redo operations to enhance annotation flexibility. The tool also supports serialized loading and processing of multiple images, providing batch saving of annotation results for unified category management and mapping. For visualization, it dynamically displays segmentation

TABLE IV
PERFORMANCE COMPARISON BETWEEN SAM AND SAM2 MODELS

Framework	Method	IoU(%)		F1(%)		mIoU(%)	PA(%)	mAP(%)
		OCMS	CCS	OCMS	CCS			
SAM-SEF-ENet-B0	SAM	64.54	43.52	78.45	60.65	54.03	61.61	71.17
SAM-SEF-ENet-B6	SAM	68.55	38.32	81.34	55.41	53.43	65.01	66.20
SAM-SEF-MobileNet	SAM	65.46	43.12	79.13	60.25	54.29	62.34	71.03
SAM2-SEF-ENet-B0	SAM2	61.07	29.60	75.83	45.67	45.33	56.48	64.36
SAM2-SEF-ENet-B6	SAM2	64.05	27.51	78.08	43.16	45.78	59.08	61.64
SAM2-SEF-MobileNet	SAM2	68.18	28.63	81.08	44.52	48.40	62.72	63.81

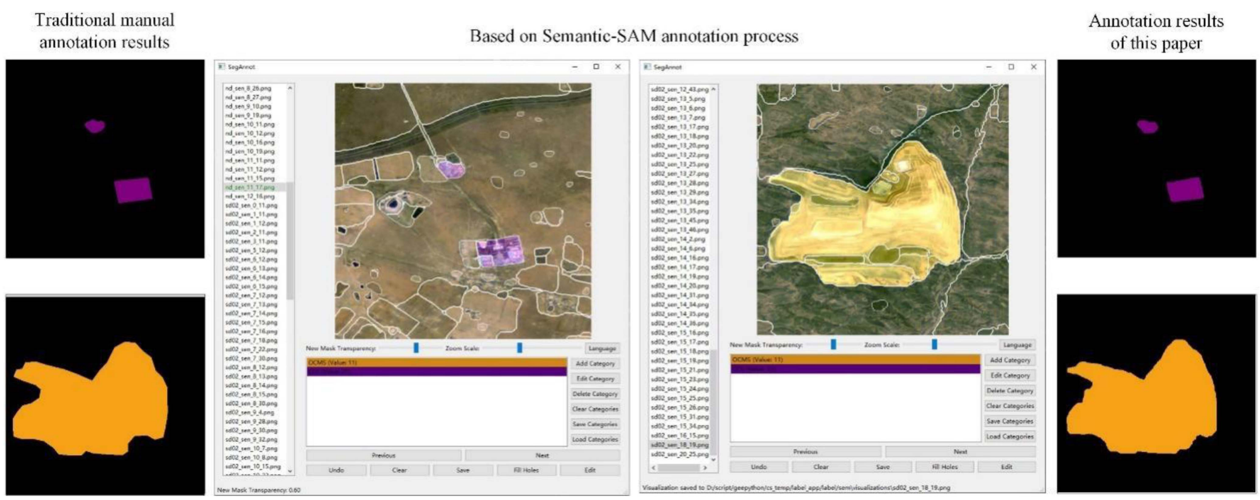

Fig. 5. Interface of the Semantic-SAM-based annotation tool. From left to right: traditional manual annotation results, the interactive interface of the annotation tool, and the annotation results of this study. Purple areas represent composite coal-related sites (CCS), and orange areas represent open-pit coal mining sites (OCMS).

boundaries and category information, supports zooming and panning, and offers multiple view mode options.

Compared to traditional manual annotation methods, this tool offers significant advantages in efficiency, quality, and professional adaptability. The pre-segmentation and batch processing capabilities substantially reduce manual annotation time and enhance efficiency while simplifying the annotation workflow and reducing operational complexity. In terms of quality assurance, Semantic-SAM ensures accurate boundary segmentation, unified category management prevents annotation inconsistencies, and interactive editing facilitates precise adjustments. Tailored to the characteristics of coal mining area sites, the tool supports precise annotation of OCMS and CCS sites and provides specialized annotation parameter configurations. The intuitive graphical interface design lowers the usage barrier, and comprehensive file management functions facilitate data organization, while flexible operation modes accommodate different user preferences.

Empirical results demonstrate that using this annotation tool reduces traditional manual annotation time by approximately 60% while maintaining high annotation quality. Particularly when handling large-scale datasets, the tool's batch processing capabilities significantly enhance annotation efficiency. Additionally, the interactive design provides necessary intervention and adjustment mechanisms for professionals, ensuring the accuracy and reliability of annotation results.

C. Model Benchmark Performance Evaluation

1) Overall Segmentation Accuracy Analysis: Among all compared deep learning models, the SAM-SEF-ENet-B0 model exhibits the most outstanding performance. As shown in Table II, this model achieves an mIoU of 87.60%, significantly outperforming other models. It also demonstrates excellent performance in segmenting both types of sites, with IoUs of 93.68% for OCMS and 81.52% for CCS, and F1 scores of 96.74% and 89.82%, respectively. These metrics fully demonstrate the exceptional performance of the SAM-SEF-ENet-B0 model in coal mining area semantic segmentation tasks.

Fig. 6 illustrates the performance of three high-performance models in segmenting typical coal mining area sites. Comparative analysis reveals that the SAM-SEF-ENet-B0 model exhibits significant advantages in several key aspects. In terms of target boundary delineation, this model provides the most precise contour mapping for open-pit coal mining areas, closely matching the manually annotated ground truth. In contrast, SAM-SEF-ENet-B6 and SAM-SEF-MobileNet show slight over-segmentation or under-segmentation in complex terrains such as pit edges and material stacking areas. Regarding segmentation stability, while all three models accurately capture the main industrial facility features within composite coal-related sites, SAM-SEF-ENet-B0 demonstrates unique advantages in handling transitional scene areas, particularly at the junctions between industrial buildings and open-pit mining sites. This model

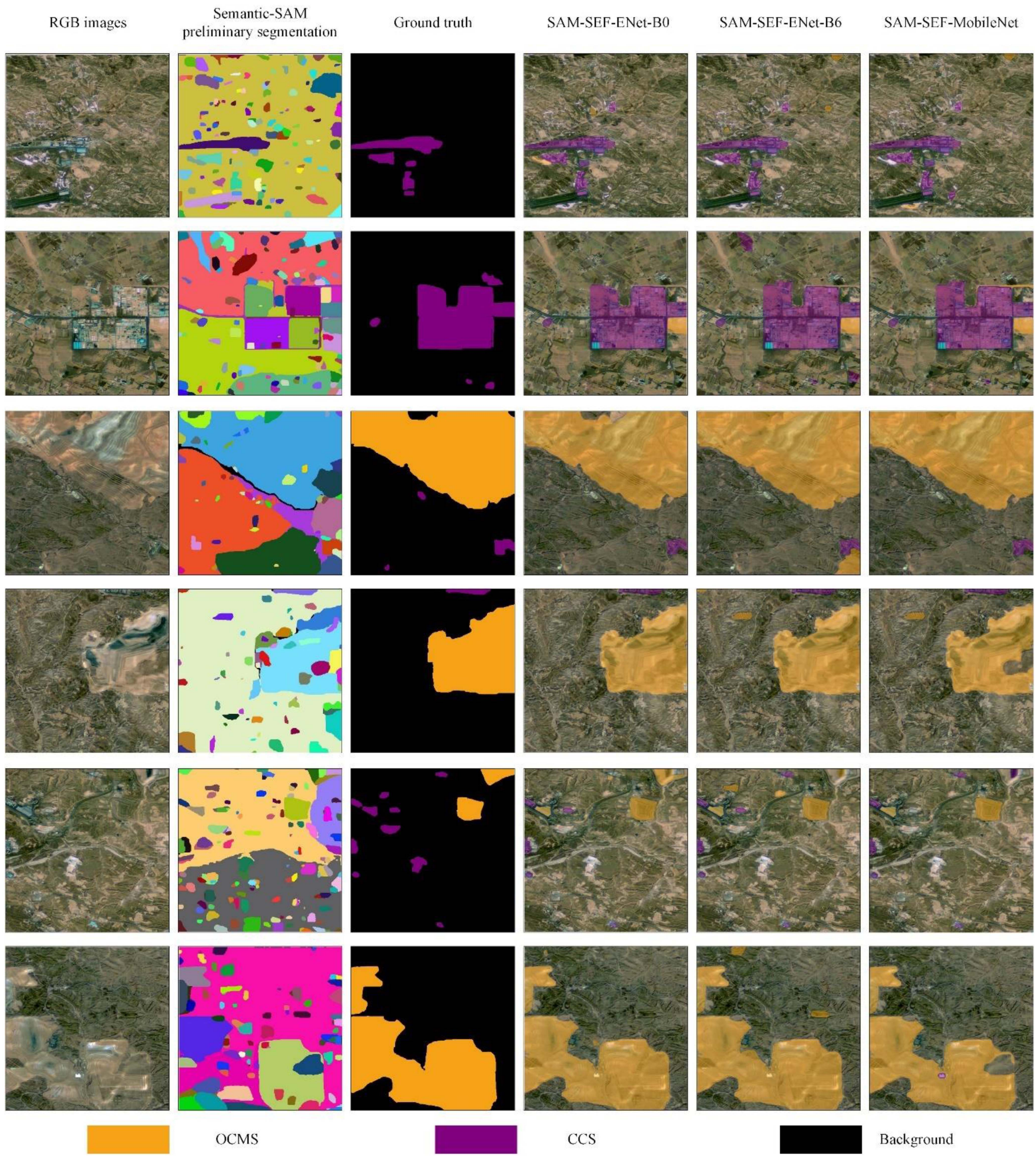

Fig. 6. Comparative segmentation effects of typical coal mining area scene features. From left to right: original RGB remote sensing image, initial segmentation results by semantic-SAM, manually annotated ground truth, recognition results by SAM-SEF-ENet-B0, SAM-SEF-ENet-B6, and SAM-SEF-MobileNet. Orange indicates open-pit coal mining sites (OCMS), purple indicates composite coal-related sites (CCS), and black represents background areas.

maintains site integrity more effectively, avoiding category confusion issues. From the initial segmentation results by Semantic-SAM, it is evident that the model preserves detailed site features well. During the subsequent classification process, SAM-SEF-ENet-B0 shows enhanced detail segmentation capabilities when handling small industrial facilities and scattered material stacks. In terms of adaptability to complex environments, SAM-SEF-ENet-B0 exhibits excellent interference resistance, accurately segmenting industrial facilities partially obscured by vegetation and shadows, especially within composite coal-related sites. From the perspective of spatial consistency, although all three models maintain good overall coherence, SAM-SEF-ENet-B0 performs more stably in handling large-scale sites, as demonstrated by the complete segmentation of large open-pit mining sites and the coherent handling of dispersed industrial facilities.

The deep learning models compared in this study exhibit distinct performance stratification. At the highest performance tier, SAM-SEF-ENet-B0, SAM-SEF-ENet-B6, and SAM- SEF-MobileNet stand out, with mIoUs exceeding 84% and F1 scores above 91%, showcasing excellent segmentation capabilities. In the medium performance tier, models such as SAM-SEF-ENet series (B1 to B5 and B7), SAM-SEF-ResNet series, SAM-SEF-GoogleNet, SAM-SEF-DenseNet, and SAM-SEF-ShuffleNet exhibit relatively stable performance, with mIoUs ranging from 78% to 84% and F1 scores between 87% and 91% . In the basic performance tier, models like SAM-SEF-ViT16, SAM-SEF-ViT32, and SAM-SEF-Inception show relatively weaker performance, with mIoUs below 78% and F1 scores below 87%, indicating a need for improved adaptability in this task. Systematic analysis of experimental results reveals that Semantic-SAM plays a crucial role in coal mining area semantic segmentation. This is attributed to its ability to provide high-quality initial segmentation masks which serve as reliable spatial priors, significantly enhancing the subsequent classification model's ability to discern boundary features accurately. This pre-segmentation-based approach effectively reduces the difficulty of direct semantic segmentation, laying a solid foundation for subsequent precise segmentation.

In terms of model selection, lightweight models demonstrate significant advantages. Models such as SAM-SEF-ENet-B0 and SAM-SEF-MobileNet achieve excellent segmentation performance with lower computational complexity. This finding suggests that, in coal mining area semantic segmentation tasks, model complexity and performance do not simply correlate positively; appropriate model architecture design is more critical than merely increasing network depth and parameter count. Additionally, the study found significant differences in segmentation difficulty across different site types, with OCMS site segmentation accuracy generally higher than CCS sites, with IoU disparities of approximately 15% –20% . This difference primarily stems from the more complex object composition and spatial distribution characteristics of CCS sites, which impose higher requirements on the model's feature extraction and scene understanding capabilities. Based on these findings, it is recommended to prioritize efficient models like SAM-SEF-ENet-B0 for practical system deployment, as these models can meet real-world application demands for computational efficiency while ensuring segmentation accuracy. While this study did not conduct exhaustive new efficiency benchmark tests (e.g., detailed FPS or parameter counts for all compared models), the selection of lightweight backbones like EfficientNet-B0 and MobileNetV2 for the classification stage within SAM-SEF, coupled with the use of a frozen Semantic-SAM, inherently addresses efficiency considerations. The complexity of training these lightweight classifiers is significantly lower than training large end-to-end segmentation models or finetuning entire foundation models. For different site types, differentiated segmentation strategies can be adopted, particularly for the more complex CCS sites, where specialized optimization methods can be employed to enhance segmentation performance. Implementing these recommendations will improve the practicality and reliability of coal mining area semantic segmentation systems.

Through an in-depth analysis of experimental results, this study discovered a phenomenon of significant theoretical importance: classification networks based on the SAM-SEF framework (e.g., SAM-SEF-ENet-B0) outperform specifically designed segmentation networks in segmentation tasks for this specific application. This result suggests that the proposed strategy of decoupling semantic segmentation into "general segmentation (via frozen Semantic-SAM) + specialized classification (on purified targets)" offers unique advantages in certain task domains. This "classification instead of segmentation" paradigm not only simplifies the model structure (by focusing the training effort on relatively smaller, specialized classifiers instead of large, monolithic segmentation networks) but also leverages Semantic-SAM's precise segmentation capabilities and the classification network's semantic understanding abilities to achieve superior overall segmentation performance. This finding provides a potentially valuable approach for addressing segmentation problems in specialized scenarios where high-quality initial segment proposals can be obtained, and offers important references for the deployment of deep learning models in professional fields, potentially favoring more lightweight classification backbones within such a decoupled framework.

2) Performance Comparison With Traditional Semantic Segmentation Models: To comprehensively evaluate the performance advantages of the proposed Semantic-SAM-based coal mining area semantic segmentation framework, we selected 17 representative semantic segmentation models for systematic comparison. These models can be categorized into four main technical groups: Lightweight Networks: including ENet [60], ERFNet [61], ESPNet [62], EDANet [63], FastSCNN [64], ContextNet [65], and SQNet [66]. These models optimize network structures and computational units to significantly reduce computational overhead while maintaining good performance, making them suitable for deployment in resource-constrained real-world applications. Multiscale feature fusion networks: including DeepLabV3+ [67], FPENet [68], CGNet [69], and LEDNet [70]. These models focus on the effective fusion of multiscale features, enhancing the understanding of complex scenes through mechanisms such as spatial pyramid pooling and context modules. Encoder–decoder architectures: including UNet [71], SegNet [72], LinkNet [73], and FSSNet [74]. These models adopt symmetrical downsampling and upsampling structures, achieving finegrained semantic segmentation through multilevel feature extraction and skip connection mechanisms. Feature enhancement networks: including DABNet [75] and ESNet [76]. These models emphasize improving feature extraction capabilities through innovative mechanisms such as deep aggregation blocks and enhanced separable convolutions to boost feature representation. These models, encompassing various technical paradigms, constitute a comprehensive performance evaluation benchmark, providing a reliable basis for in-depth analysis of the technical advantages of our proposed framework.

As can be seen from Fig. 7, the traditional segmentation model suffers from obvious boundary blurring and category confusion problems when dealing with complex coal mine area

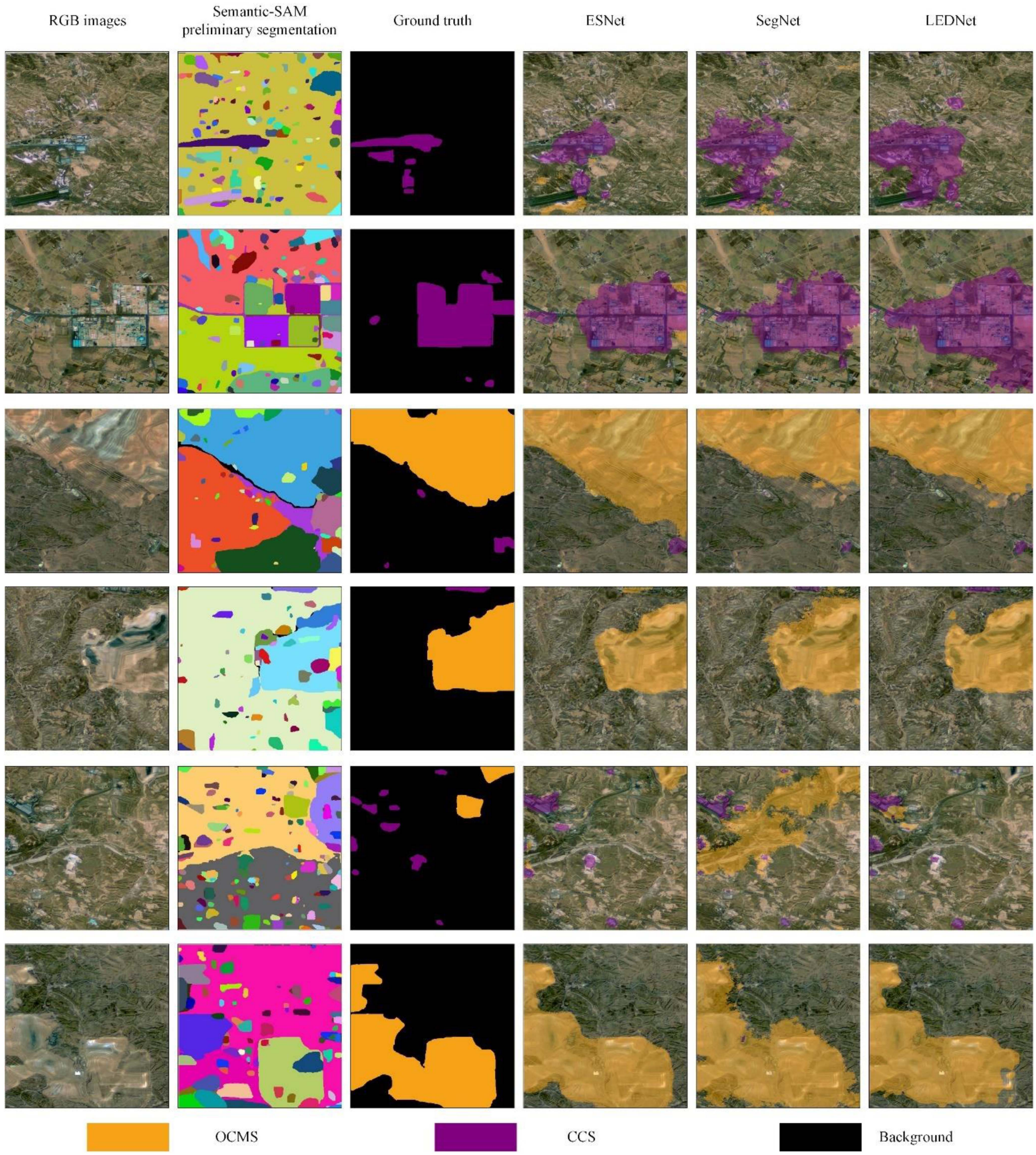

Fig. 7. Comparative segmentation results of traditional semantic segmentation models. From left to right: original RGB remote sensing image, UNet segmentation results, ESNet segmentation results, LEDNet segmentation results, and DeepLabV3+ segmentation results. Orange indicates open-pit coal mining sites (OCMS), purple indicates composite coal-related sites (CCS), and black represents background areas.

sites, and especially performs poorly in the transition region between OCMS and CCS sites. In contrast, the Semantic-SAM-based framework proposed in this study can better maintain boundary clarity and category consistency. In-depth analysis of experimental data reveals that our framework demonstrates significant advantages across three key dimensions: In terms of overall performance, the SAM- SEF-ENet-B0 model achieves an mIoU of 87.60%, compared to the best-performing traditional segmentation model, ESNet, which achieves an mIoU of 82.61%, marking an improvement of nearly 5 percentage points (see Table III). This performance enhancement not only reflects numerical superiority but also indicates the unique advantages of the SAM-SEF strategy. By leveraging Semantic-SAM's high-quality initial boundary proposals, the subsequent classification model can focus more effectively on semantic discrimination within well-defined regions, leading to improved overall segmentation accuracy while maintaining boundary precision.

In terms of scene adaptability, our framework exhibits robust cross-site generalization capabilities. For relatively regular OCMS sites, the framework achieves an IoU of 93.68%, significantly surpassing the highest traditional model's performance of 85.97% . In the structurally complex and feature-diverse CCS sites, although our framework's IoU is slightly lower than LEDNet's at 83.02%, it demonstrates better stability in transitional areas between industrial buildings and open-pit mining sites through precise guidance by the target purification strategy, effectively avoiding common classification confusion issues seen in traditional models.

In terms of boundary accuracy, our framework shows distinct advantages in boundary accuracy. Qualitative analysis indicates that traditional segmentation models often encounter oversegmentation or undersegmentation issues when handling complex boundaries, especially at object category intersections. Conversely, our framework leverages Semantic-SAM's excellent boundary awareness and precise guidance from the target purification strategy to more accurately delineate target contours (see Fig. 7). This enhanced boundary performance, further quantified in Section IV-D, is directly attributed to utilizing the high-fidelity masks from Semantic-SAM as the basis for the final segmentation output, refined by the classification labels.

These in-depth comparative analyses fully confirm the technical advantages of the Semantic-SAM-based semantic segmentation framework proposed in this study. The framework not only achieves significant improvements across various quantitative metrics but also demonstrates excellent generalization capabilities and operational stability in practical applications. These performance advantages are attributable to the framework's design, which thoroughly considers the specificity of coal mining area sites and integrates multiple innovative strategies to achieve precise segmentation of complex industrial scenes.

D. Boundary Accuracy and Segmentation Quality

Boundary accuracy is a critical metric for evaluating the performance of semantic segmentation models, directly impacting the effectiveness of downstream applications. In the complex environments of coal mining areas, precise boundary detection is especially important for distinguishing different functional zones, assessing mining extents, and conducting environmental monitoring.

In designing the evaluation methodology, we selected the F1 score, Precision, and Recall as the primary evaluation metrics. These three indicators reflect the quality of boundary detection from different perspectives: Precision measures the accuracy of predicted boundaries, Recall indicates the completeness of boundary detection, and the F1 score comprehensively assesses the model's overall performance. Specifically, we employed a 3-pixel tolerance band to evaluate boundary matching and utilized an adaptive thresholding method for boundary extraction, where the threshold dynamically adjusts based on the image gradient distribution. The evaluation targets included six representative models: SAM-SEF-ENet-B0, SAM-SEF-ENet-B6, SAM-SEF-MobileNet, ESNet, SegNet, and LEDNet, tested across both OCMS and CCS.

Quantitative evaluation results revealed significant performance differences. As shown in Fig. 8, models based on Semantic-SAM consistently outperformed traditional methods across all metrics. In the OCMS category, SAM-SEF-ENet-B0 achieved the highest performance with an F1 score of 0.376, Precision of 0.329, and Recall of 0.468. SAM-SEF-ENet-B6 followed closely with an F1 score of 0.343. Notably, all models exhibited higher Recall than Precision in this category, indicating a tendency to classify more pixels as boundary regions, thereby generating broader predicted boundaries than the actual ones. This phenomenon is particularly evident when handling large-scale targets such as open-pit mines, likely due to the complexity of terrain undulations and boundary transition zones.

In the CCS category, SAM-SEF-ENet-B0 maintained its leading advantage with an F1 score of 0.602, Precision of 0.671, and Recall of 0.592. SAM-SEF-MobileNet also demonstrated strong performance with an F1 score of 0.576. Contrary to the OCMS category, Precision exceeded Recall in CCS scenes, indicating a more conservative approach in predicting boundaries of industrial buildings and other composite scenes. This prediction strategy helps reduce false detections but may result in missing some ambiguous or complex boundary areas. As illustrated in Fig. 9, analysis of the F1 score distribution across different models clearly shows that Semantic-SAM-based models (SAM-SEF-ENet-B0, SAM-SEF-ENet-B6, SAM-SEF-MobileNet) have significant advantages over traditional models (ESNet, SegNet, LEDNet), especially in handling complex CCS sites with better stability and consistency. This superior boundary performance of the SAM-SEF models reinforces the benefit of leveraging Semantic-SAM's high-quality initial segmentation masks, which form the spatial basis for the final labeled output.

From a qualitative analysis perspective, different site types exhibit unique boundary characteristics. OCMS sites typically feature more regular geometric shapes and clear boundaries, with the primary challenge being the preservation of large-scale target integrity. In contrast, CCS sites possess more complex boundary features, often encompassing multiple interconnected industrial facilities with higher boundary ambiguity. Semantic-SAM-based models demonstrate clear advantages in handling both site types: accurately capturing the contours of pits and the boundaries of material stacking areas in OCMS sites, and effectively managing the complex boundaries of building clusters and transition zones between facilities in CCS sites (see Fig. 7).

These findings provide important insights for further model optimization. Firstly, model complexity does not correlate straightforwardly with boundary detection performance, as evidenced by SAM-SEF-ENet-B0 outperforming the more complex SAM-SEF-ENet-B6. Secondly, boundary detection difficulty varies significantly across different site categories, primarily due to inherent scene complexity and feature distribution characteristics. Lastly, the tradeoff between Precision and Recall varies by category, offering directions for targeted optimization.

Based on the above analysis, we recommend focusing on several key areas in future research. For addressing over-prediction issues in OCMS sites, improvements in loss function design or incorporation of boundary constraints can help mitigate excessive boundary predictions. To better handle complex boundaries

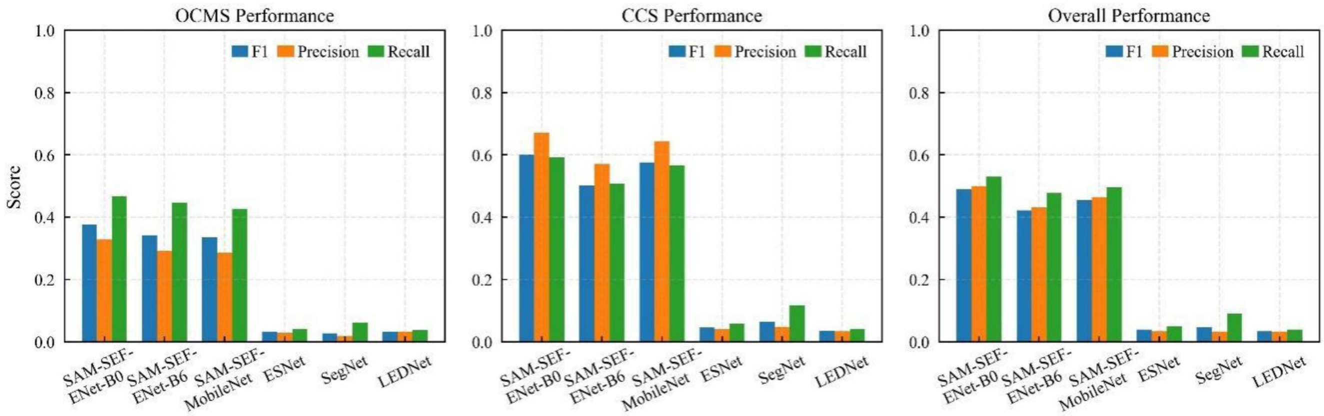

Fig. 8. Comparison of boundary detection performance across different models in OCMS and CCS sites. From left to right: OCMS performance, CCS performance, and overall performance. Each scene includes three metrics: F1 score (blue), Precision (orange), and Recall (green).

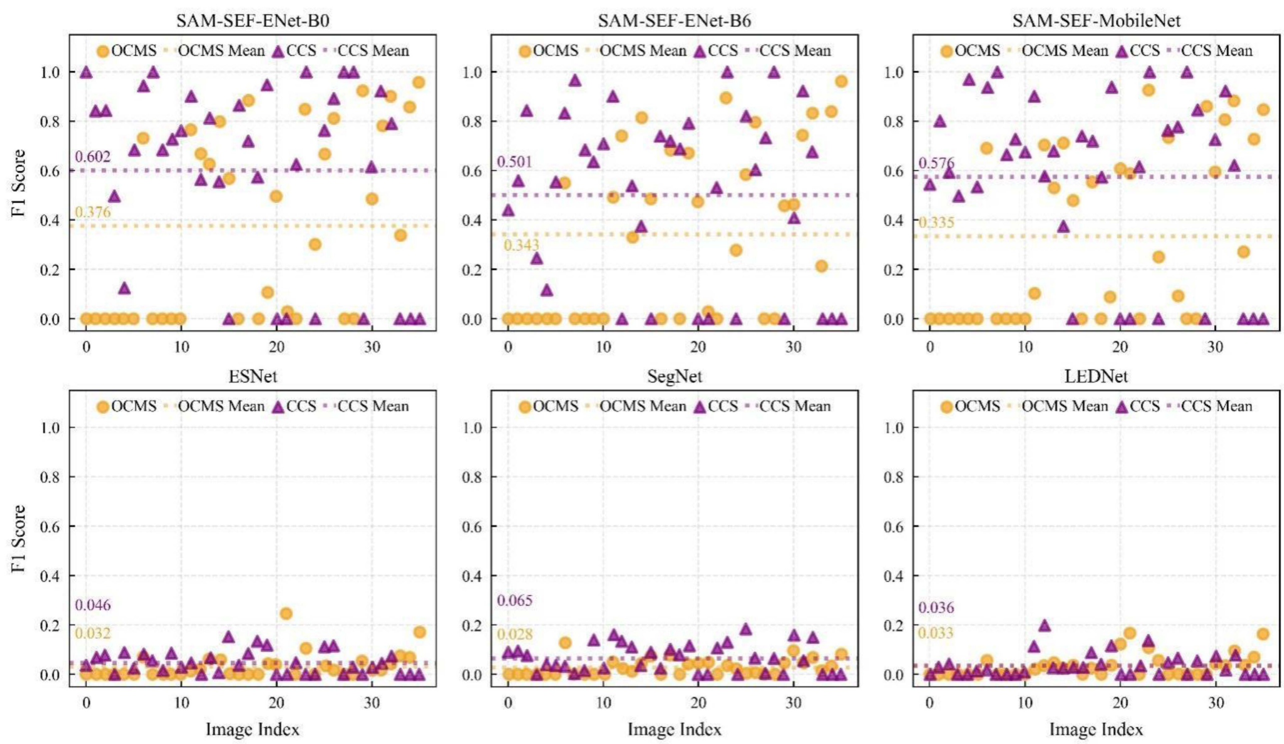

Fig. 9. Distribution of F1 scores across different models. The top three plots display the F1 score distributions for Semantic-SAM-based models (SAM-SEF-ENet-B0, SAM-SEF-ENet-B6, SAM-SEF-MobileNet), while the bottom three plots show distributions for traditional models (ESNet, SegNet, LEDNet). Yellow circles represent F1 scores for OCMS sites, purple triangles represent CCS sites, and dashed lines indicate the average F1 scores for each site type.

in CCS sites, introducing multiscale feature fusion mechanisms would enhance performance in these challenging scenarios. Additionally, selecting appropriate model configurations based on specific site types and considering ensemble strategies when necessary can achieve better results. These optimization measures will collectively help enhance the accuracy and robustness of boundary detection.

E. Multiscale Adaptability Analysis

To systematically evaluate the model's ability to segment targets of different scales, this section categorizes analysis objects into small, medium, and large scale levels based on the 33rd and 66th percentiles of target area within each sample in the dataset. This statistical-based categorization objectively reflects the scale distribution characteristics of the dataset, providing a reliable benchmark for model performance evaluation. As shown in Fig. 10, the sample target areas for OCMS sites are approximately 3000–4000 pixels (small), 12 000–13 000 pixels (medium), and 110 000–120 000 pixels (large). For CCS sites, the areas are approximately 500 pixels (small), 1300–1400 pixels (medium), and 8000–47 000 pixels (large). The evaluation targets included six representative models: SAM-SEF-ENet-B0, SAM-SEF-ENet-B6, SAM-SEF-MobileNet, ESNet, SegNet,

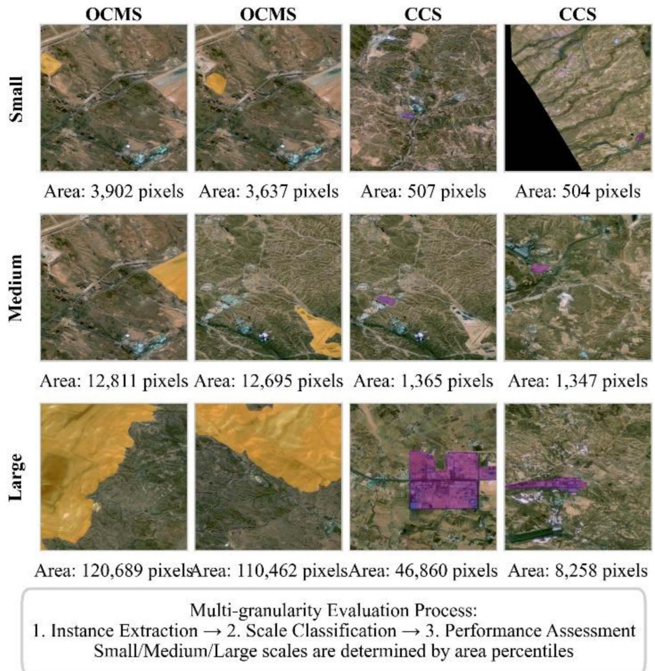

Fig. 10. Examples of multiscale samples in coal mining area scenes. From top to bottom: Small-scale, medium-scale, and large-scale samples for OCMS and CCS sites. Each scale level displays the pixel area of target regions.

and LEDNet, assessed using IoU, F1 score, and PA. As depicted in Fig. 11, in the analysis of small-scale targets, SAM-SEF-ENet-B0 performs the best for the OCMS category, achieving an IoU of 0.79,while other models exhibit relatively weaker performance, particularly lightweight models like LEDNet and SAM-SEF-MobileNet, which show noticeable performance declines. In the CCS category, although overall performance is lower than in OCMS, SAM-SEF-ENet-B0 still maintains a leading advantage with an IoU of approximately 0.67, and performance differences among models are pronounced. These performance disparities primarily arise from the difficulty in feature extraction for small-scale targets and the varying feature complexity of different site categories at small scales.

In the medium-scale target analysis, all models show significant performance improvements. For OCMS sites, SAM-SEF-ENet-B0, SAM-SEF-ENet-B6, and SAM-SEF-MobileNet perform excellently, with IoUs exceeding 0.95 and minimal performance differences among models. In the CCS category, performance generally improves, and the performance gap compared to OCMS narrows accordingly, with the EfficientNet-based model series continuing to hold a dominant position. This indicates that medium-scale targets contain richer feature information, facilitating accurate model segmentation.

In the large-scale target analysis, model performance stabilizes and remains at a high level. In the OCMS category, the SAM-SEF-ENet series and SAM-SEF-MobileNet continue to show clear advantages, with model performances converging. In the CCS category, performance further improves, with the performance gap with OCMS reaching its smallest and model performance differences no longer being significant. This phenomenon indicates that large-scale targets provide ample spatial features and semantic information, enabling different models to achieve good segmentation results.

Through multiscale evaluation analysis, we found a clear positive correlation between target scale and segmentation performance, with small-scale target segmentation being the main challenge. From the perspective of category differences, the performance of OCMS sites consistently surpasses that of CCS sites across all scales, with the most significant difference observed at small scales and minimal differences at large scales. This is closely related to the feature complexity and semantic hierarchy of each category. Regarding model characteristics, EfficientNet-based models performed the best in multiscale analysis, while lightweight models, despite showing significant disadvantages in small-scale analysis, exhibited consistent performance across different scales. The SAM-SEF framework generally demonstrates good adaptability across scales, though the challenge of small object segmentation persists, suggesting potential areas for improvement perhaps in the classification stage or the initial mask generation sensitivity of Semantic-SAM.

Based on these findings, we recommend focusing on both technical enhancements and application strategies. On the technical side, efforts should concentrate on enhancing models' small-scale feature extraction capabilities, developing more effective multiscale feature fusion mechanisms, and optimizing cross-scale feature learning strategies. For applications, it is important to establish unified multiscale evaluation standards, develop adaptive scale analysis methods, and optimize multi-scale data augmentation strategies. These optimization measures will help improve model performance in segmenting targets of various scales, particularly enhancing small-scale target segmentation capabilities and reducing performance discrepancies across different categories.

F. Verification of Model Selection Rationality

The selection of Semantic-SAM as the foundational model for this study is primarily based on its significant advantages in semantic understanding, architectural adaptability, and computational efficiency.

Semantic-SAM significantly enhances the model's ability to understand professional scenes in coal mining areas by introducing specialized semantic understanding modules. Its innovative decoupling method for object and component classification, achieved through a shared text encoder, allows independent execution of object and component segmentation tasks. This design is particularly suited to the complex segmentation needs of coal mining areas. Additionally, its unique semantic feature extraction mechanism effectively captures key features of specialized scenes such as open-pit mines and industrial buildings, laying a solid foundation for subsequent precise classification.

Semantic-SAM aligns closely with the target purification strategy proposed in this study. Its multigranularity segmentation capability enables the generation of multiple levels of segmentation results from a single prompt point, providing a reliable basis for precise classification. Particularly in handling complex coal mining scenes, its hierarchical feature extraction architecture can adaptively adjust the receptive field size, effectively

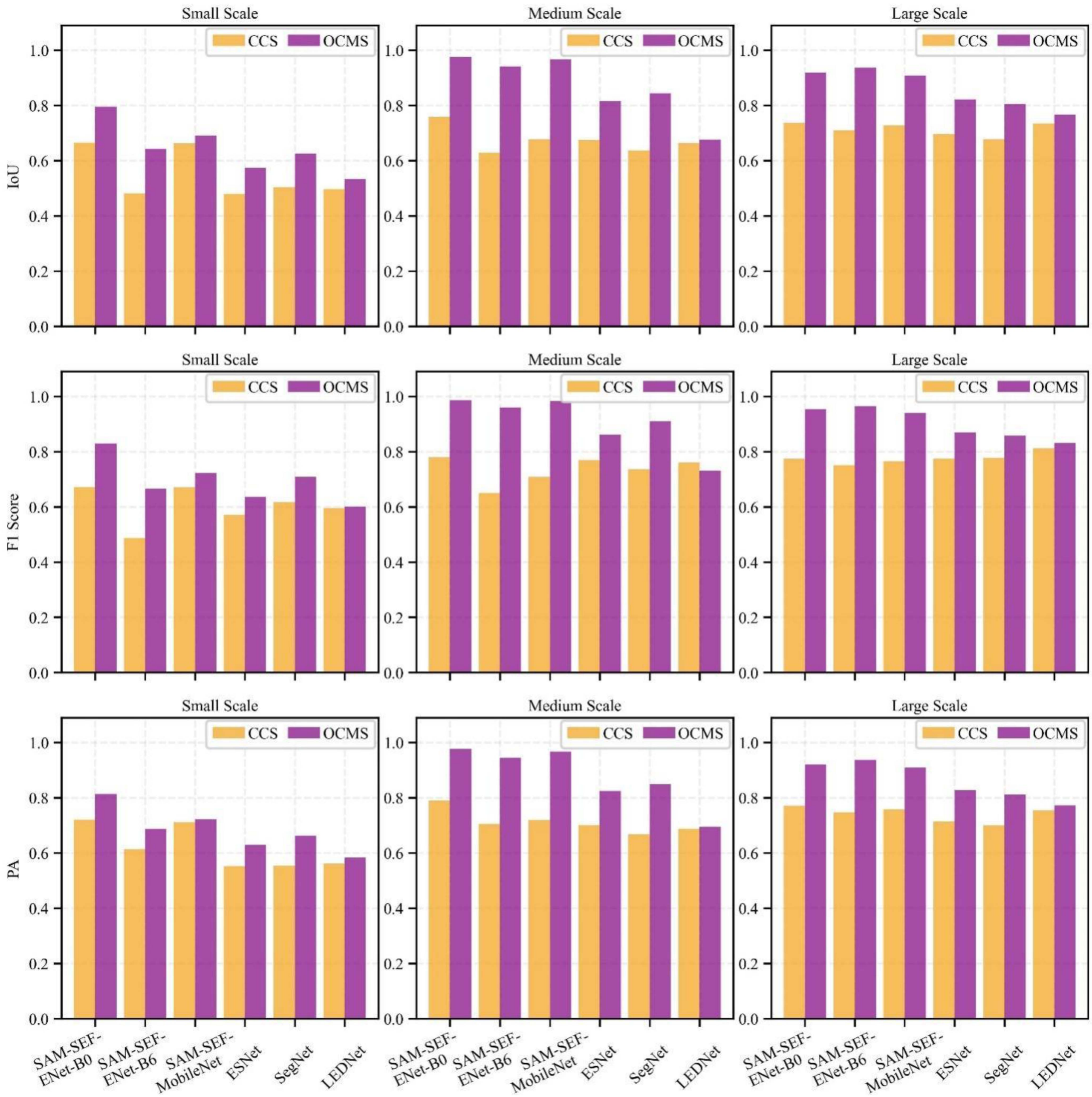

Fig. 11. Model performance comparison across different scales.

balancing the acquisition of local details and global semantic information. This architectural feature allows the model to better handle targets of different scales, thereby improving overall segmentation accuracy. As shown in Fig. 12, compared to SAM [34] and SAM2 [51], Semantic-SAM exhibits clear advantages in boundary detection accuracy and integrity. Specifically, in handling transition areas between industrial buildings and natural terrain, Semantic-SAM better maintains boundary continuity and accuracy, avoiding the oversegmentation and boundary blurring issues commonly seen in SAM and SAM2.

Semantic-SAM employs efficient attention mechanisms and feature fusion strategies, significantly reducing computational overhead while maintaining high performance. Its optimized inference process enables the model to rapidly process large-scale remote sensing image data, especially high-resolution images, demonstrating noticeable speed advantages. This efficient computational characteristic makes it particularly suitable for the demands of practical engineering applications.

Experimental data further validate the rationality of selecting Semantic-SAM (see Table IV). In the OCMS category, SAM-based models achieved an average IoU of 66.18%, while SAM2 models scored 64.43% . In the CCS category, SAM models averaged an IoU of 41.65%, significantly higher than SAM2 models' 28.58% . Overall, SAM models achieved an average mIoU of 53.92%, approximately 7 percentage points higher than SAM2 models. In terms of PA, SAM models

Fig. 12. Comparative performance of SAM, SAM2, and semantic-SAM in boundary detection within coal mining areas. From top to bottom: segmentation results of SAM, SAM2, and semantic-SAM, each row containing six typical scene examples. White contour lines indicate the detected boundaries by the models.

averaged 62.99%, outperforming SAM2 models at 59.43% . Additionally, SAM models achieved an average mAP of 69.47%, about 6 percentage points higher than SAM2 models. Notably, SAM models exhibited minimal performance fluctuations across different architectures (standard deviation of only 0.44%), whereas SAM2 models showed larger performance variability (standard deviation of 1.66%), indicating better architectural adaptability and performance stability in SAM models.

Although the newly released SAM2 has made improvements in model efficiency and boundary prediction accuracy, Semantic-SAM was chosen over SAM2 for this study based on the following reasons. Semantic-SAM exhibits enhanced semantic adaptability, demonstrating stronger scene adaptability in semantic understanding, especially when handling complex industrial scenes in coal mining areas. Its architectural design is more compatible with the target purification strategy proposed in this study, and it also demonstrates better computational efficiency and performance stability in practical applications. These advantages make Semantic-SAM the optimal choice for the framework proposed in this research.

G. Threats to Validity

Several potential threats to the validity of this study should be acknowledged. Regarding internal validity, instrumentation threats, such as potential bugs in our SAM-SEF framework implementation or evaluation scripts, were mitigated through careful code review and the use of well-tested libraries. Selection bias, although we chose two major and diverse coal production bases in China, means that the findings might not fully generalize to all global coal mine types with vastly different characteristics. Historical and maturation factors were considered to have minimal impact on this single-instance segmentation task, and the semi-automatic annotation tool helped maintain consistency.

Concerning external validity, the model's performance on other geographical regions, geological conditions, or significantly different remote sensing data sources (e.g., much higher/lower resolution imagery, SAR data) requires further investigation, as our study primarily focused on Sentinel-2 imagery. The applicability of the SAM-SEF decoupled strategy to other remote sensing tasks or non-remote sensing domains would also need specific validation for each new context.

For construct validity, our chosen evaluation metrics (IoU, F1-score, mIoU, PA) are standard in semantic segmentation and appropriately reflect the "precise semantic segmentation" construct. The definitions for "OCMS" and "CCS" were consistently applied during annotation.

Regarding conclusion validity, while we did not perform rigorous statistical significance testing, all model comparisons were conducted under strictly controlled and identical conditions, ensuring the relative fairness of the performance evaluation. The observed superiority of the SAM-SEF framework is attributed to its core design principles, supported by systematic experiments and ablation studies.

V. Conclusion and Future Work

The precise semantic segmentation method for coal mining areas based on Semantic-SAM proposed in this study demonstrates significant advantages across multiple dimensions. By innovatively integrating the zero-shot segmentation capabilities of Semantic-SAM with specialized deep learning models, we constructed an efficient SAM-SEF framework, achieving a systematic "segmentation-classification-update" processing workflow.

Experimental results indicate that the SAM-SEF-ENet-B0 model within this framework exhibits outstanding performance in recognizing segmenting OCMS and CCS, achieving IoUs of 93.68% and 81.52%, respectively, and a mIoU of 87.60% . Multiscale evaluations reveal a positive correlation between target scale and segmentation performance, with small-scale target segmentation remaining the primary challenge. The semi-automatic annotation tool developed in this study significantly enhanced annotation efficiency through presegmentation and interactive editing functionalities, reducing traditional manual annotation time by approximately 60% . Systematic experiments validated the necessity of selecting Semantic-SAM, as it demonstrates clear advantages over SAM and SAM2 in semantic understanding, architectural adaptability, and computational efficiency. Notably, this study found that classification networks based on the SAM-SEF framework achieve superior performance in segmentation tasks compared to specifically designed segmentation networks for this specific application. This strategy of decoupling semantic segmentation into "general segmentation + specialized classification" not only simplifies the model structure but also exhibits unique advantages in specific task domains. However, we acknowledge this advantage is demonstrated within our specific context (coal mine sites, Sentinel-2 data, specific target classes) and may depend on the quality of the initial segmentation provided by the foundation model and the characteristics of the target classes. This finding provides valuable reference for addressing complex visual tasks in specialized scenarios.

Future research will delve deeper into technical optimization, application expansion, evaluation systems, and data processing. On the technical front, efforts will focus on enhancing the segmentation capabilities for small-scale targets and exploring more effective multi-scale feature extraction and fusion mechanisms. Future work could also explore iterative refinement mechanisms or feedback loops where classification results might guide or correct initial segmentations to further mitigate error propagation from the initial segmentation stage. In terms of applications, the model can be extended to other industrial sites, developing automated monitoring systems tailored to practical engineering needs. Specifically, incorporating advanced reasoning mechanisms for dynamic spatio-temporal analysis could enable the monitoring of temporal changes and activities within coal mine areas. Furthermore, drawing inspiration from recent advancements in cross-modal learning, future efforts could investigate fusing multisource data (e.g., SAR, DEM, textual information) with the SAM-SEF framework to enhance semantic understanding and segmentation robustness. Regarding evaluation systems, there is a need to establish more comprehensive multiscale evaluation standards and develop automated performance assessment tools, potentially including more detailed efficiency benchmarks (e.g., model parameters, FPS) for deployment considerations. In data processing, emphasis will be placed on researching data augmentation and sample balancing strategies, as well as exploring semi-supervised and weakly supervised learning methods. These research directions will further enhance the accuracy and efficiency of semantic segmentation in coal mining areas, providing more reliable technical support for intelligent mine monitoring and offering valuable references for remote sensing image analysis in other specialized fields.

References

[1] T. T. Werner, G. M. Mudd, A. M. Schipper, M. A. J. Huijbregts, L. Taneja, and S. A. Northey, "Global-scale remote sensing of mine areas and analysis of factors explaining their extent," *Glob. Environ. Change*, vol. 60, 2020, Art. no. 102007.

[2] Y. Cai et al., "A review of monitoring, calculation, and simulation methods for ground subsidence induced by coal mining," *Int. J. Coal Sci. Technol.*, vol. 10, no. 1, 2023, Art. no. 32.

[3] M. Zhu et al., "High-precision monitoring and prediction of mining area surface subsidence using SBAS-InSAR and CNN-BiGRU-attention model," *Sci. Rep.*, vol. 14, no. 1, Nov. 2024, Art. no. 28968.

[4] Y. Liu et al., "A review of water pollution arising from agriculture and mining activities in Central Asia: Facts, causes and effects," *Environ. Pollut.*, vol. 291, 2021, Art. no. 118209.

[5] H. Zhang, F. Zhang, J. Song, M. L. Tan, H. Kung, and V. C. Johnson, "Pollutant source, ecological and human health risks assessment of heavy metals in soils from coal mining areas in Xinjiang, China," *Environ. Res.*, vol. 202, 2021, Art. no. 111702.

[6] T. Chen, X. Zheng, R. Niu, and A. Plaza, "Open-pit mine area mapping with Gaofen-2 satellite images using U-Net+," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3589–3599, 2022.

[7] S. Du, J. Xing, J. Li, S. Du, C. Zhang, and Y. Sun, "Open-pit mine extraction from very high-resolution remote sensing images using OM-DeepLab," *Natural Resour. Res.*, 2022.

[8] M. Couttenier, S. Di Rollo, L. Inguere, M. Mohand, and L. Schmidt, "Mapping artisanal and small-scale mines at large scale from space with deep learning," *PLoS One*, vol. 17, no. 9, p. e0267963, 2022, doi: 10.1371/journal.pone.0267963.

[9] J. Gallwey, C. Robiati, J. Coggan, D. Vogt, and M. Eyre, "A Sentinel-2 based multispectral convolutional neural network for detecting artisanal small-scale mining in Ghana: Applying deep learning to shallow mining," *Remote Sens. Environ.*, vol. 248, 2020, Art. no. 111970.

[10] H. Xie, Y. Pan, J. Luan, X. Yang, and Y. Xi, "Semantic segmentation of open pit mining area based on remote sensing shallow features and deep learning," in *Proc. Big Data Analytics Cyber-Phys. Syst. Smart City*, 2021, pp. 52–59.

[11] R. Guan, Z. Li, T. Li, X. Li, J. Yang, and W. Chen, "Classification of heterogeneous mining areas based on ResCapsNet and Gaofen-5 imagery," *Remote Sens.*, vol. 14, no. 13, 2022, Art. no. 3216.

[12] G. Zhou, J. Xu, W. Chen, X. Li, J. Li, and L. Wang, "Deep feature enhancement method for land cover with irregular and sparse spatial distribution features: A case study on open-pit mining," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4401220.

[13] L. Huang, B. Jiang, S. Lv, Y. Liu, and Y. Fu, "Deep-learning-based semantic segmentation of remote sensing images: A survey," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 8370–8396, 2024.

[14] L. Ding, J. Zhang, and L. Bruzzone, "Semantic segmentation of large-size VHR remote sensing images using a two-stage multiscale training architecture," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5367–5376, Aug. 2020.

[15] D. Basu and S. Mishra, "Mine reclamation practices and effects of stakeholder perception — A case study of Saoner mines, Maharashtra, India," *J. Eng. Appl. Sci.*, vol. 71, no. 1, Dec. 2024, Art. no. 62.

[16] C. S. Rathore† and R. Wright, "Monitoring environmental impacts of surface coal mining," *Int. J. Remote Sens.*, vol. 14, no. 6, pp. 1021–1042, Apr. 1993.

[17] J. Hu, B. Ye, Z. Bai, and Y. Feng, "Remote sensing monitoring of vegetation reclamation in the Antaibao open-pit mine," *Remote Sens.*, vol. 14, no. 22, Nov. 2022, Art. no. 5634.

[18] W. Chen, X. Li, H. He, and L. Wang, "A review of fine-scale land use and land cover classification in open-pit mining areas by remote sensing techniques," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 15.

[19] D. Zhang, J. Leng, X. Li, W. He, and W. Chen, "Three-stream and double attention-based DenseNet-BiLSTM for fine land cover classification of complex mining landscapes," *Sustainability*, vol. 14, no. 19, 2022, Art. no. 12465.

[20] J. Shi, D. Li, X. Chu, J. Yang, and C. Shen, "Intelligent classification of land cover types in open-pit mine area using object-oriented method and multitask learning," *J. Appl. Remote Sens.*, vol. 16, no. 3, 2022, Art. no. 038504.

[21] J. Li, X. Cai, and J. Qi, "AMFNet: An attention-based multi-level feature fusion network for ground objects extraction from mining area's UAV-based RGB images and digital surface model," *J. Appl. Remote Sens.*, vol. 15, no. 3, 2021, Art. no. 036506.

[22] D. Wang, J. Zhang, B. Du, G.-S. Xia, and D. Tao, "An empirical study of remote sensing pretraining," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5608020.

[23] A. A. Aleissaee et al., "Transformers in Remote Sensing: A survey," *Remote Sens.*, vol. 15, no. 7, 2023, Art. no. 1860.

[24] A. A. Adegun, S. Viriri, and J.-R. Tapamo, "Review of deep learning methods for remote sensing satellite images classification: Experimental survey and comparative analysis," *J. Big Data*, vol. 10, no. 1, Jun. 2023, Art. no. 93.

[25] Z. Su, W. Li, Z. Ma, and R. Gao, "An improved U-Net method for the semantic segmentation of remote sensing images," *Appl. Intell.*, vol. 52, no. 3, pp. 3276–3288, 2022.

[26] X. Zhang, Q. Liu, H. Chang, and H. Sun, "High-resolution network with transformer embedding parallel detection for small object detection in optical remote sensing images," *Remote Sens.*, vol. 15, no. 18, 2023, Art. no. 4497.

[27] Y. Bazi, L. Bashmal, M. M. A. Rahhal, R. A. Dayil, and N. A. Ajlan, "Vision transformers for remote sensing image classification," *Remote Sens.*, vol. 13, no. 3, 2021, Art. no. 516.

[28] H. Wang et al., "YOLOFIV: Object detection algorithm for around-the-clock aerial remote sensing images by fusing infrared and visible features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 15269–15287, 2024.

[29] H. Ren, M. Xia, L. Weng, K. Hu, and H. Lin, "Dual-attention-guided multiscale feature aggregation network for remote sensing image change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 4899–4916, 2024.

[30] Y. Hou, K. Liu, X. Zhai, and Z. Chen, "Mbda-net: A building damage assessment model based on a multi-scale fusion network," *Signal, Image Video Process.*, vol. 18, no. 12, pp. 9363–9374, 2024.

[31] G. Murdaca, F. Ricciuti, A. Rucci, B. L. Saux, A. Fumagalli, and C. Prati, "A semi-supervised deep learning framework for change detection in open-pit mines using SAR imagery," *Remote Sens.*, vol. 15, no. 24, Dec. 2023, Art. no. 5664.

[32] Z. Li, Y. Zhao, H. Ren, and Y. Sun, "A novel index for detecting bare coal in open-pit mining areas based on landsat imagery," *Remote Sens.*, vol. 16, no. 24, 2024, Art. no. 4648.

[33] D. Jung and Y. Choi, "Systematic review of machine learning applications in mining: Exploration, exploitation, and reclamation," *Minerals*, vol. 11, no. 2, 2021, Art. no. 148.

[34] A. Kirillov et al., "Segment anything," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 3992–4003.

[35] H. Yu et al., "Time-series field phenotyping of soybean growth analysis by combining multimodal deep learning and dynamic modeling," *Plant Phenomics*, vol. 6, 2024, Art. no. 0158.

[36] J. Zhang, Q. Zhang, Y. Gong, J. Zhang, L. Chen, and D. Zeng, "Weakly supervised semantic segmentation with consistency-constrained multiclass attention for remote sensing scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5621118.

[37] W. Zhang, M. Cai, T. Zhang, G. Lei, Y. Zhuang, and X. Mao, "Popeye: A unified visual-language model for multi-source ship detection from remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 20050–20063, 2024.

[38] X. Zhang, Y. Liu, Y. Lin, Q. Liao, and Y. Li, "UV-SAM: Adapting segment anything model for urban village identification," in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, no. 20, pp. 22520–22528, Mar. 2024.

[39] X. Zhao, Z. Wu, Y. Chen, W. Zhou, and M. Wei, "Fine-grained high-resolution remote sensing image change detection by SAM-UNet change detection model," *Remote Sens.*, vol. 16, no. 19, 2024, Art. no. 3620.

[40] W. Ji, J. Li, Q. Bi, T. Liu, W. Li, and L. Cheng, "Segment anything is not always perfect: An investigation of SAM on different real-world applications," *Mach. Intell. Res.*, vol. 21, no. 4, pp. 617–630, 2024.

[41] X. Zhou et al., "MeSAM: Multiscale enhanced segment anything model for optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5623515.

[42] F. Li et al., "Semantic-SAM: Segment and recognize anything at any granularity," 2023, *arXiv:2307.04767*.

[43] D. Wang et al., "SAMRS: Scaling-up remote sensing segmentation dataset with segment anything model," in *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds. San Jose, CA, USA: Curran Assoc., 2023, pp. 8815–8827.

[44] J. Wu et al., "Medical SAM adapter: Adapting segment anything model for medical image segmentation," *Med. Image Anal.*, vol. 102, 2025, Art. no. 103547.

[45] Y. Li, D. Wang, C. Yuan, H. Li, and J. Hu, "Enhancing agricultural image segmentation with an agricultural segment anything model adapter," *Sensors*, vol. 23, no. 18, Sep. 2023, Art. no. 7884.

[46] C. Lu, E. Nnadozie, M. Camenzind, Y. Hu, and K. Yu, "Maize plant detection using UAV-based RGB imaging and YOLOv5," *Front. Plant Sci.*, vol. 14, 2024, Art. no. 1274813.

[47] L. P. Osco et al., "The segment anything Model (SAM) for remote sensing applications: From zero to one shot," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 124, 2023, Art. no. 103540.

[48] M. Diab, P. Kolokoussis, and M. A. Brovelli, "Optimizing zero-shot text-based segmentation of remote sensing imagery using SAM and grounding DINO," *Artif. Intell. Geosci.*, vol. 6, no. 1, Jun. 2025, Art. no. 100105.

[49] J. Nie, X. Wang, R. Hou, G. Li, H. Chen, and W. Zhu, "Dynamic spatio-temporal graph reasoning for VideoQA with self-supervised event recognition," *IEEE Trans. Image Process.*, vol. 33, pp. 4145–4158, 2024.

[50] C. Zheng et al., "Cross-modal progressive perspective matching network for remote sensing image-text retrieval," *IEEE Trans. Multimedia*, pp. 1–13. to be published, doi: 10.1109/TMM.2025.3535365, 2025.

[51] N. Ravi et al., "Sam 2: Segment anything in images and videos," 2024, *arXiv:2408.00714*.

[52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[53] D. Alexey, "An image is worth 16 × 16 words: Transformers for image recognition at scale," 2020, *arXiv: 2010.11929*.

[54] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[55] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[56] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.

[57] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.

[58] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient cnn architecture design," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 116–131.

[59] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[60] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A deep neural network architecture for real-time semantic segmentation," 2016, *arXiv:1606.02147*.

[61] E. Romera, J. M. Álvarez, L. M. Bergasa, and R. Arroyo, "ERFNet: Efficient residual factorized ConvNet for real-time semantic segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 263–272, Jan. 2018.

[62] S. Mehta, M. Rastegari, A. Caspi, L. Shapiro, and H. Hajishirzi, "Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 552–568.

[63] S.-Y. Lo, H.-M. Hang, S.-W. Chan, and J.-J. Lin, "Efficient dense modules of asymmetric convolution for real-time semantic segmentation," in *Proc. 1st ACM Int. Conf. Multimedia Asia*, 2020, pp. 1–6.

[64] R. P. Poudel, S. Liwicki, and R. Cipolla, "Fast-scnn: Fast semantic segmentation network," 2019, *arXiv:1902.04502*.

[65] R. Poudel, U. Bonde, S. Liwicki, and C. Zach, "Contextnet: Exploring context and detail for semantic segmentation in real-time," 2018, *arXiv:1805.04554*.

[66] M. Treml et al., "Speeding up semantic segmentation for autonomous driving," in *NIPS MLITS, NIPS Workshop on Mach. Learn. for Intell. Transp. Syst.*, vol. 2, pp. 1–17, 2016.

[67] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.

[68] M. Liu and H. Yin, "Feature pyramid encoding network for real-time semantic segmentation," 2019, *arXiv:1909.08599*.

[69] T. Wu, S. Tang, R. Zhang, J. Cao, and Y. Zhang, "CGNet: A light-weight context guided network for semantic segmentation," *IEEE Trans. Image Process.*, vol. 30, pp. 1169–1179, 2021.

[70] Y. Wang et al., "Lednet: A lightweight encoder-decoder network for real-time semantic segmentation," in *Proc. IEEE Int. Conf. Image Process.*, 2019, pp. 1860–1864.

[71] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.

[72] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[73] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process.*, 2017, pp. 1–4.

[74] X. Zhang, Z. Chen, Q. M. J. Wu, L. Cai, D. Lu, and X. Li, "Fast semantic segmentation for scene perception," *IEEE Trans. Ind. Informat.*, vol. 15, no. 2, pp. 1183–1192, Feb. 2019.

[75] G. Li, I. Yun, J. Kim, and J. Kim, "DABNet: Depth-wise asymmetric bottleneck for real-time semantic segmentation," 2019, *arXiv:1907.11357*.

[76] Y. Wang, Q. Zhou, J. Xiong, X. Wu, and X. Jin, "ESNet: An efficient symmetric network for real-time semantic segmentation," in *Proc. 2nd Chin. Conf. Pattern Recognit. Comput. Vis.*, 2019, pp. 41–52.

Libing Wang (Student Member, IEEE) received the B.S. degree in geographical science and the M.S. degree in physical geography from Northwest Normal University, Lanzhou, China, in 2019 and 2022, respectively. He is currently working toward the Ph.D. degree in photogrammetry and remote sensing in the School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou, China.

He has published 12 academic papers, holds one patent, and three software copyrights. He has presided more than two scientific research projects and participated in seven scientific research projects. His research interests include mine environmental remote sensing and intelligent identification.

Jihong Dong (Member, IEEE) received the Ph.D. degree in environmental science from the China University of Mining and Technology, China, in 2008.

She is a Professor and Doctoral Supervisor with the School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou, China. She has held international visiting professorships and authored influential publications and patents. Her research interests include mine environment, ecological restoration, and geospatial technologies

Dr. Dong is a Member of the National Environmental Benchmark Expert Committee, an Academic Leader for the Ministry of Education's "111 Plan," and a Member of the Japan Resources Engineering Society. She was the recipient of prestigious awards for her contributions.

Lei Wang received the Ph.D. degree.

She is a Senior Engineer and Secretary-General with the China Coal Society, Beijing, China. Her work is dedicated to advancing the coal sector through technological innovation and sustainable practices. Her research interests include promoting the high-quality development of the coal industry and the deep integration of science, technology, and industry, driven by technological innovation, green and low-carbon transformation, and international development.

Feng Liu received the Ph.D. degree in management science and engineering from the China University of Mining and Technology, Beijing, China.

He is a Researcher with the China Coal Society, Beijing, China. He serves as President of the China Coal Society and Vice President of the China National Coal Association. He is also Editor-in-Chief of the *Journal of China Coal Society*. He has long been engaged in policy formulation for technological innovation in the coal sector and has published multiple academic papers. His research interests include mining engineering, industrial economy, and automation technology.

Qingke Wen received the Ph.D. degree in cartography and geographic information systems from the institute of remote sensing applications, Chinese Academy of Sciences, China, in 2009.

She is an Associate Researcher with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. She served as a Visiting Scholar at the European Commission Joint Research Centre (EC-JRC). She has authored academic works, holds patents, and her research interests include remote sensing applications for land resources and environment.

Dr. Wen was the recipient of notable awards for her contributions. She serves as a Specially Invited Procurator Assistant and is a main participant in Global Dam Watch.

Bekir Genc received the Ph.D. degree from The School of Mining Engineering, University of the Witwatersrand, South Africa, in 2015.

He is a Professor and Doctoral Supervisor with the School of Mining Engineering, University of the Witwatersrand, Johannesburg, South Africa. He is an NRF-rated Researcher, holds patents, and has published widely. His research interests include mine safety, coal technology, and energy.

Dr. Genc is a Registered Professional Engineer (ECSA) and a Fellow of the SAIMM. He serves on international journal editorial boards and is a Member of the Society of Mining Professors (SOMP). He was the recipient of notable awards for his work.