Enhancing mineral processing with deep learning: Automated quartz identification using thin section images

Gökhan Külekçi, Kemal Hacıefendiolu, and Hasan Basri Baaa

View the article online at SpringerLink or IJMMM Webpage.

Articles you may be interested in

Si-wei Wu, Jian Yang, and Guang-ming Cao, Prediction of the Charpy V-notch impact energy of low carbon steel using a shallow neural network and deep learning, *Int. J. Miner. Metall. Mater.*, 28(2021), No. 8, pp. 1309-1320. https://doi.org/10.1007/s12613-020-2168-z

Zheng-hua Deng, Hai-qing Yin, Xue Jiang, Cong Zhang, Guo-fei Zhang, Bin Xu, Guo-qiang Yang, Tong Zhang, Mao Wu, and Xuan-hui Qu, Machine-learning-assisted prediction of the mechanical properties of Cu–Al alloy, *Int. J. Miner. Metall. Mater.*, 27(2020), No. 3, pp. 362-373. https://doi.org/10.1007/s12613-019-1894-6

Xiao-liang Zhang, Jue Kou, Chun-bao Sun, Rui-yang Zhang, Min Su, and Shuo-fu Li, Mineralogical characterization of copper sulfide tailings using automated mineral liberation analysis: A case study of the Chambishi Copper Mine tailings, *Int. J. Miner. Metall. Mater.*, 28(2021), No. 6, pp. 944-955. https://doi.org/10.1007/s12613-020-2093-1

Bao-hua Yang, Ai-xiang Wu, Guillermo A. Narsilio, Xiu-xiu Miao, and Shu-yue Wu, Use of high-resolution X-ray computed tomography and 3D image analysis to quantify mineral dissemination and pore space in oxide copper ore particles, *Int. J. Miner. Metall. Mater.*, 24(2017), No. 9, pp. 965-973. https://doi.org/10.1007/s12613-017-1484-4

Sajjad Sattari and Amir Atrian, Effects of the deep rolling process on the surface roughness and properties of an Al-3vol%SiC nanoparticle nanocomposite fabricated by mechanical milling and hot extrusion, *Int. J. Miner. Metall. Mater.*, 24(2017), No. 7, pp. 814-825. https://doi.org/10.1007/s12613-017-1465-7

Peng Xing, Bao-zhong Ma, Peng Zeng, Cheng-yan Wang, Ling Wang, Yong-lu Zhang, Yong-qiang Chen, Shuo Wang, and Qiu-yin Wang, Deep cleaning of a metallurgical zinc leaching residue and recovery of valuable metals, *Int. J. Miner. Metall. Mater.*, 24(2017), No. 11, pp. 1217-1227. https://doi.org/10.1007/s12613-017-1514-2

IJMMM WeChat   QQ author group

Enhancing mineral processing with deep learning: Automated quartz identification using thin section images

Gökhan Külekçi[1],✉, *Kemal Hacıefendioğlu*[2,3,4], *and Hasan Basri Başağa*[2,3]

1) Mining Engineering Department, Faculty of Engineering and Natural Sciences, Gümüşhane University, Gümüşhane, Türkiye
2) Department of Civil Engineering, Faculty of Engineering, Karadeniz Technical University, Trabzon, Türkiye
3) Earthquake and Structural Health Application and Research Center, Karadeniz Technical University, Trabzon, Türkiye
4) Civil Engineering Academy R & D Software Consulting Limited Company, Samsun, Türkiye

Abstract: The precise identification of quartz minerals is crucial in mineralogy and geology due to their widespread occurrence and industrial significance. Traditional methods of quartz identification in thin sections are labor-intensive and require significant expertise, often complicated by the coexistence of other minerals. This study presents a novel approach leveraging deep learning techniques combined with hyperspectral imaging to automate the identification process of quartz minerals. The utilized four advanced deep learning models—PSPNet, U-Net, FPN, and LinkNet—has significant advancements in efficiency and accuracy. Among these models, PSPNet exhibited superior performance, achieving the highest intersection over union (IoU) scores and demonstrating exceptional reliability in segmenting quartz minerals, even in complex scenarios. The study involved a comprehensive dataset of 120 thin sections, encompassing 2470 hyperspectral images prepared from 20 rock samples. Expert-reviewed masks were used for model training, ensuring robust segmentation results. This automated approach not only expedites the recognition process but also enhances reliability, providing a valuable tool for geologists and advancing the field of mineralogical analysis.

Keywords: quartz mineral identification; deep learning; hyperspectral imaging; deep learning in geology

1. Introduction

The recognition of quartz minerals holds great significance in the fields of mineralogy and geology. Quartz, a widespread mineral in our natural environment and industrial applications, demands precise identification due to its prevalence [1–2]. Quartz possesses specific optical characteristics, yet it frequently coexists with other minerals, complicating the identification process. These challenges become apparent during the examination of thin sections, causing significant difficulties for geologists [3–4].The mineralogical analysis of thin sections is a commonly employed method in geological studies [5–6]. Thin sections are used to examine the structure of rocks and minerals. However, these analyses are often time-consuming and require expertise [6–10]. Geologists must individually inspect each part of these thin sections and identify the minerals. This process involves gathering a substantial amount of data and interpreting it, demanding time [11–12].

Quartz identification using traditional optical mineralogy, such as polarizing microscopy, has been widely used. This method requires expert interpretation of optical properties like birefringence and extinction angles. However, such manual approaches are time-consuming and prone to human error, especially in heterogeneous rock samples [3,9].

The presented method using deep learning automates the entire identification process. Deep learning, especially semantic segmentation models, such as PSPNet and U-Net, significantly reduces the time and effort required, providing more consistent and accurate results [8,13]. The automation helps minimize the challenges faced by traditional methods in distinguishing quartz from other minerals, especially when these minerals coexist.

Previous mineral identification studies have often relied on traditional machine learning models such as support vector machines (SVM) and random forest (RF). These models perform classification based on handcrafted features, which limits their ability to capture the complexity of mineral textures and spectral variations [8,14].

Although existing approaches using more traditional hyperspectral processing algorithms have been used, the deep learning method appears to be more applicable. For example, spectral angle matching (SAM), principal component analysis (PCA) or k-means clustering have been traditionally used for hyperspectral mineral identification [15]. These techniques rely on manual thresholding or unsupervised learning and are generally less effective than deep learning in cases where precise segmentation is required [14].

✉ Corresponding author: Gökhan Külekçi   E-mail: gokhankulekci@gmail.com; gkulekci@gumushane.edu.tr

SAM has been applied to hyperspectral mineral identification but has limitations in performance when identifying fine-grained minerals in complex datasets [16]. Agrawal *et al.* [3] applied random forest and support vector machines for mineral identification using hyperspectral data and noted that these methods, while effective for certain minerals, struggle with generalization across diverse mineral types. Including a comparison with these approaches could further emphasize the robustness of deep learning, especially for quartz identification. PCA has been a common method for dimensionality reduction in hyperspectral data analysis, but it often sacrifices the granularity needed for precise mineral identification [17].

In recent years, the advancement of deep learning techniques and hyperspectral imaging technology have presented new opportunities for the automatic recognition of quartz minerals [14,17]. Hyperspectral images can measure the reflections of objects across numerous spectral bands, providing a rich source for mineral recognition. Deep learning serves as an effective approach for analyzing this vast dataset and recognizing minerals [5,18–21]. Consequently, a deep learning-based approach for the automatic identification of quartz minerals offers an alternative to traditional methods that require both time and expertise.

This study introduces a deep learning-based approach developed to enhance the automatic recognition of quartz minerals from hyperspectral images. This approach leverages the advantages of using hyperspectral images and expedites the recognition of quartz minerals. Additionally, experimental results will be presented to evaluate the accuracy and reliability of this method [1,22–23]. Regarding the contributions of this study, foremost, it has the potential to expedite the automatic recognition of quartz minerals, enabling geologists to access more data in less time. Furthermore, the accuracy and reliability of this deep learning-based approach surpass traditional methods. Therefore, this study can be considered a significant step in the recognition of quartz minerals [15–16,24]. By integrating four advanced semantic segmentation models—PSPNet, U-Net, Feature Pyramid Network (FPN), and LinkNet—we systematically analyze and compare their performance in accurately recognizing quartz minerals. The preparation of geological thin sections, the microscopic examination, and the data processing techniques are meticulously detailed, ensuring a robust foundation for the deep learning models. Extensive experimental results are presented, highlighting the superior accuracy and reliability of these models in segmenting quartz from complex geological samples.

This research fills a critical gap in the literature by demonstrating how cutting-edge technologies can be creatively applied to overcome the limitations of traditional methods. By offering a scalable, accurate, and automated solution to mineral identification, the study paves the way for future advancements in geological research and automated analytical techniques.

2. Methodology

Accurate identification of quartz minerals is of paramount importance in mineralogy and geology, as it aids in understanding rock composition and petrogenetic processes. This study employs a deep learning framework to automatically detect quartz cross-sections within thin-section images, thereby addressing the time-consuming challenges associated with manual identification. Drawing on microscopic examinations of 120 thin sections prepared from 20 rock samples, the methodology utilizes four segmentation models—PSPNet, U-Net, FPN, and LinkNet—to classify the presence of quartz in optical micrographs.

All image processing and data manipulation steps were performed in Python using Numpy [25], while data visualization was managed through Matplotlib [26]. The deep learning architectures were implemented using well-known libraries such as Keras–TensorFlow [27].

2.1. Preparation of geological thin sections and identification of quartz minerals

The production of geological thin sections and the accurate identification of minerals within them are of paramount importance in the fields of mineralogy and geology. In this article, we will delve into the process of creating geological thin sections, the methods for mineral identification using a light microscope, and a comparative analysis of these processes. The production of geological thin sections commences with the collection of samples from a specific geological site. These samples are then prepared in a laboratory setting for subsequent examination. Rocks are first cut into specific dimensions and then sliced into thin layers. These thin sections are mounted onto prepared slides for optical analysis.

Geological thin sections are examined by using a light microscope, a vital tool for observing the optical properties of minerals. The identification of quartz minerals is particularly intriguing. Quartz is characteristically transparent and is not confined to a specific color. However, when viewed under specific light polarizations, quartz exhibits distinct shapes and colors. This feature serves as a key criterion for distinguishing quartz from other minerals when using a light microscope.

Nevertheless, quartz is often found intermingled with other minerals, complicating the identification process. Therefore, geologists must consider the presence of other minerals when making identifications. Another challenge faced by geologists is the time-consuming nature of traditional methods in making accurate mineral identifications. The manual examination of each part of a thin section and the manual identification of minerals require significant time and expertise.

In conclusion, the preparation of geological thin sections and the process of mineral identification using a light microscope hold great significance in the fields of mineralogy and geology. These processes are fundamental tools for examining and accurately identifying the structures of rocks and minerals. While quartz's distinct color and shape can be readily identified by using a light microscope, the potential for intermingling with other minerals necessitates careful consideration. Additionally, the time-intensive nature of

these processes has spurred the exploration of automated identification methods, such as deep learning.

2.1.1. Preparation of rock thin sections

For the identification of quartz minerals to be taught to the deep learning system, 120 thin sections were prepared from 20 rock samples collected from different regions. It was noted that the rock samples taken from 20 different rocks for deep learning were rich in quartz minerals and appeared in different shapes in different rocks. The rock sample numbers taken are given in Table 1.

Table 1.   Number of rocks and samples taken

Rock type	Number of samples
Quartzite	5
Granite	3
Sandstone	2
Gneiss	2
Rhyolite	3
Schis	1
Pegmatite	1
Basalt	1
Diorite	1
Gabbro	1

Thin slices, measuring 0.5 cm × 2.0 cm × 4.0 cm, were cut from the rocks, and after smoothing one of their surfaces, and they were attached to 2.5 cm × 5.0 cm glass slides using Canada balsam. The rock sample attached to the glass was then thinned down to a thickness of 0.025 mm using abrasives, making it ready for petrographic examination (Fig. 1).
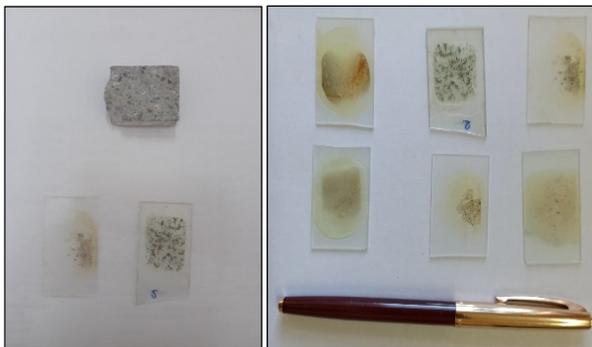
Fig. 1.   Thin section images.

2.1.2. Microscopic examinations

The 120 prepared thin section samples were individually examined by using Leica brand polarizing microscope. Through petrographic studies, samples which contained quartz minerals and suitable for introduction to deep learning were identified. The relationships between the minerals constituting the rocks, along with their optical properties, were observed. Subsequently, 1180 microphotographs were taken from the thin sections using a microscope-attached camera under both single and cross-nicol setups (Fig. 2).

2.1.3. Dataset

Techniques such as X-ray diffraction (XRD) and scanning electron microscopy (SEM) have been effective in
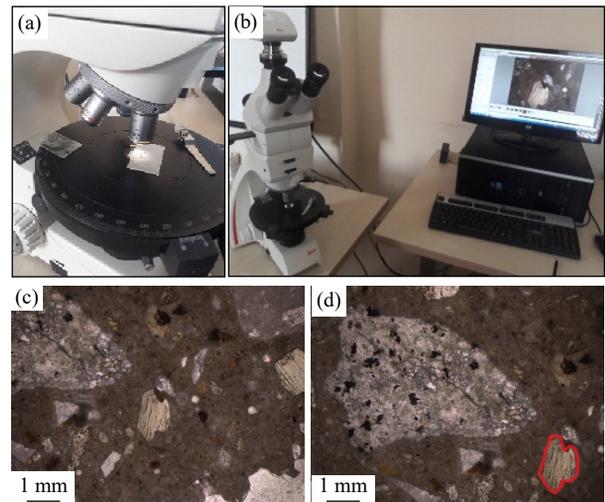
Fig. 2.   Thin sections obtained from different rock types: (a) pasting on slides, (b) viewing under a light microscope, (c) obtaining 1 mm scale microphotographs containing quartz minerals, and (d) creating raw data by marking quartz minerals.

identifying mineral compositions but lack the spectral resolution needed for precise image-based classification. Moreover, these methods are less suitable for automated large-scale mineral identification [17]. Hyperspectral imaging captures information across hundreds of spectral bands, which allows for finer distinction between minerals with subtle spectral differences. Study [17] has shown the advantages of hyperspectral imaging in identifying minerals with high spectral variance. This paper builds on that by introducing deep learning models to automate the extraction of this detailed information, resulting in more efficient mineral identification [13]. The integration of these two technologies has been rare in previous works but is crucial in achieving the claimed advances.

Quartz minerals were identified in these captured images and loaded into the deep learning program (Fig. 3). The training and test datasets were carefully balanced to ensure that both sets contained a representative distribution of images across conditions. In the initial stage, 100 random images of size 256 px × 192 px were extracted from 2560 px × 1920 px surface images, resulting in a total of 2470 images. The data was divided into training and testing sets in an 8:2 ratio. Additionally, techniques such as stratified sampling were used to ensure diversity in mineral appearances and contexts. This approach aims to increase the generalizability and robustness of the models in real-world applications. Images were carefully selected and preprocessed to ensure uniformity in lighting conditions and to minimize potential variation. Each image was captured under controlled illumination during microscopic examinations, and steps were taken to standardize image quality. This approach improves the reproducibility of our results by increasing the robustness of the models to illumination inconsistencies.

Consequently, 1976 and 494 images were used for training and test, respectively. All images were reviewed by experts to identify the areas containing quartz minerals, and masking operations were performed accordingly. The resulting data was organized into two folders named "images" and

Fig. 3.   Some of the photographs containing quartz taken from thin sections.
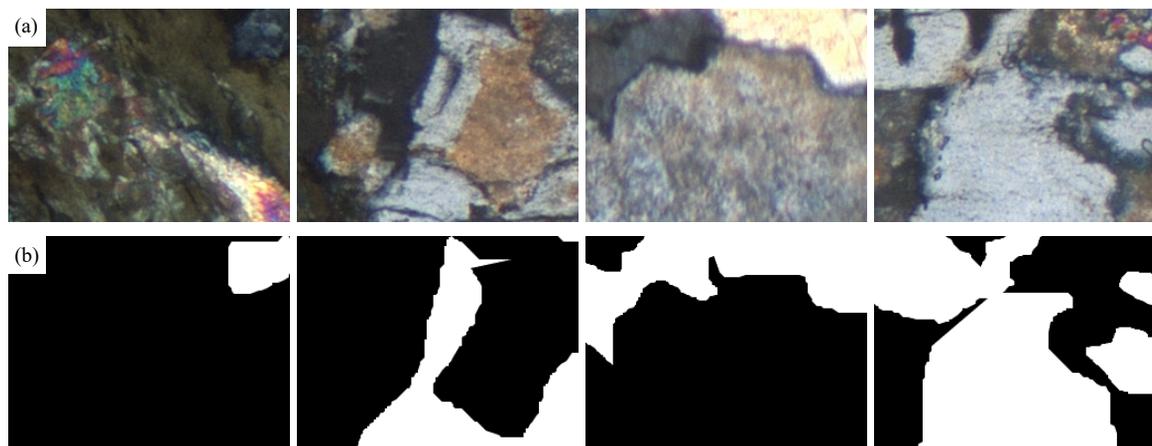
Fig. 4.   Some of the photographs (256 px × 192 px) containing quartz for (a) images and (b) masks.

"masks." Examples of the images and masks are shown in Fig. 4.

2.2. Semantic segmentation models

In recent years, deep learning (DL) techniques have increasingly been applied to image segmentation tasks, with semantic segmentation becoming crucial in fields such as medical imaging and disaster management. The objective of semantic segmentation is to label each pixel in an image according to the object class it belongs to. This task is particularly challenging due to variations in object shapes, sizes, orientations, and the potential for low-quality or occluded images in disaster scenarios.

To address these challenges, several DL architectures have been developed [28]. This study employs four encoder–decoder-based semantic segmentation models (SSMs) for segmenting collapsed buildings post-earthquake: PSPNet (pyramid scene parsing network) [29], U-Net (u-shaped network) [30], FPN (feature pyramid network) [31], and LinkNet (link network) [32]. U-Net features an encoder–decoder structure with skip connections that improve segment-

ation outcomes. LinkNet is similar with U-Net but utilizes residual blocks in its encoder and decoder. FPN, akin to U-Net, uses a 1×1 convolution layer and combines features differently. PSPNet incorporates a pyramid pooling module for global context aggregation and an auxiliary loss [33]. Various encoders were employed for feature extraction, chosen based on their performance in prior studies and suitability for this task, including different variations of networks [34].

2.2.1. PSPNet

PSPNet [29] is a deep learning architecture designed for semantic segmentation that excels in capturing context information at various scales. The key feature of PSPNet is its pyramid pooling module, which performs pooling operations at four different scales to enhance the global representation of features. As shown in Fig. 5, the pyramid pooling module processes feature maps at different scales, capturing global context information effectively. In PSPNet, the final feature map P is defined by Eq. (1):

$$P(x) = \mathrm{Concat}\left(\mathrm{Up}(x1), \mathrm{Up}(x2), \mathrm{Up}(x3), \mathrm{Up}(x4), x\right) \quad (1)$$

where x denotes the original feature map, $x1$, $x2$, $x3$, and $x4$

806

Int. J. Miner. Metall. Mater., Vol. 32, No. 4, Apr. 2025

(a) Input image   (b) Feature map   (c) Pyramid pooling module   (d) Final prediction
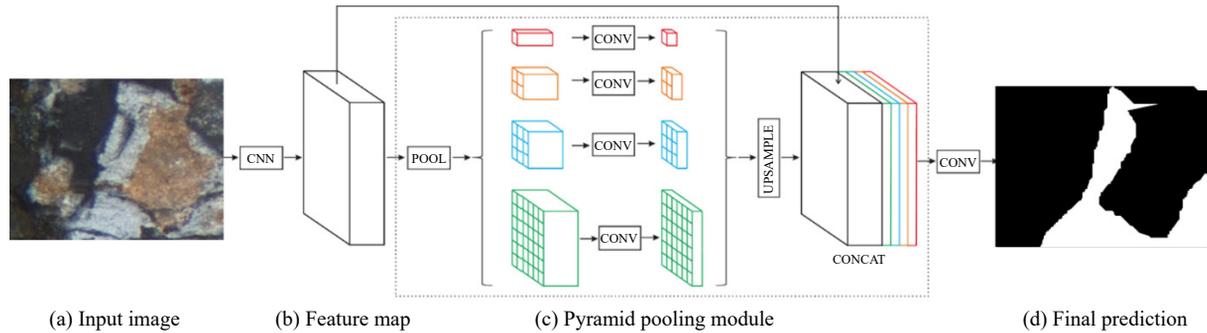
Fig. 5. Architecture overview of PSPNet.

are the feature maps pooled at varying scales, $\mathrm{Up}()$ is the upsampling function, and $\mathrm{Concat}()$ is the concatenation function. This design allows PSPNet to integrate contextual information effectively, making it highly suitable for complex scene parsing tasks.

In Fig. 5, CNN represents a convolutional neural network used for initial feature extraction, CONV denotes a convolution operation applied to refine feature maps, POOL stands for pooling, which reduces spatial dimensions to capture multi-scale features, and CONCAT refers to the concatenation operation that combines upsampled feature maps from different pooling scales into a unified representation. Considering Fig. 5, input images are typically greater than (256,256). Using transfer learning and dilated convolutions, the network constructs feature maps. Smaller kernels gather information over larger areas, with the number of feature maps N as a tunable hyperparameter. The pyramid pooling module performs average pooling at scales such as global average pooling and (2×2) to segment varying object sizes. For instance, $N = 512$ maps and $n = 4$ pooling sizes yield $N/n = 128$ feature maps per level. Module B contains three layers of residual blocks, outputting 256 feature maps, Module C implements pooling to reduce pooled feature maps to 64, totaling 512 maps and Module D, a convolution layer, outputs maps sized (256,256,3), flattened to 196608 for further processing.

2.2.2. U-Net

U-Net is a powerful convolutional neural network designed for biomedical image segmentation, characterized by its unique architecture that combines a contracting path for feature extraction and an expansive path for precise localization. The model operates on an input image and progressively reduces its dimensionality while capturing context through convolutional layers and max pooling. At the bottleneck, the network maintains critical information, and during the expansive phase, it upscales and concatenates feature maps from the contracting path. This architecture allows U-Net to produce high-quality segmentation results, effectively distinguishing fine details in complex images [30].

Let $X \in \mathbf{R}^{H \times W \times C}$, where H and W are height and width, respectively, and C is the number of channels. The contracting path consists of n convolutional layers defined as Eq. (2). Here, k indexes the convolutional layer, with $k = 1, 2, \ldots, n$.

$$F_k = \mathrm{ReLU}(Z_k * X + b_k) \tag{2}$$

where Z_k are the convolutional filters, b_k is the bias, and $*$ denotes convolution. Max pooling reduces dimensions by a factor of 2. At the bottleneck, the feature maps are processed as Eq. (3).

$$F_n = \mathrm{Conv}(F_{n-1}) \tag{3}$$

The expanding path upscales the feature maps using transposed convolution:

$$F' = \mathrm{Conv}_{\mathrm{transpose}}(F_n) \tag{4}$$

Followed by concatenation with corresponding feature maps from the contracting path:

$$F_{\mathrm{concat}} = \mathrm{concat}(F', F_k) \tag{5}$$

The final segmentation output is computed as:

$$Y = \mathrm{Softmax}(\mathrm{Conv}(F_{\mathrm{concat}})) \tag{6}$$

The architecture ensures that both high-level features and spatial context are preserved, allowing for precise segmentation of complex images.

Crucial to U-Net's effectiveness are its skip connections, which concatenate feature maps from the encoder directly to the decoder, enhancing feature propagation and enabling precise pixel classification. This architecture is particularly adept at capturing detailed spatial hierarchies necessary for high-accuracy segmentation, making it ideal for tasks requiring detailed localization such as medical imaging. Fig. 6 depicts the U-Net structure, consisting of two feature encoding and decoding steps.

2.2.3. FPN

FPN shown in Fig. 7 is a robust architecture for semantic segmentation that enhances multi-scale feature learning. It builds on a backbone network, such as ResNet, by constructing a pyramid of feature maps at various scales. The key strength of FPN lies in its top-down pathway, where high-level semantic features from deeper layers are upsampled and combined with corresponding lower-level features through lateral connections. This fusion of high-resolution spatial features with semantic-rich layers allows FPN to perform accurate segmentation, particularly for detecting objects of various sizes across an image [31].

FPN is a widely used architecture for multi-scale feature extraction, particularly in object detection and semantic segmentation tasks. Given an input image $I \in \mathbf{R}^{H \times W \times C}$, where H and W represent the height and width of the image, and C represents the number of channels in the input image. For in-
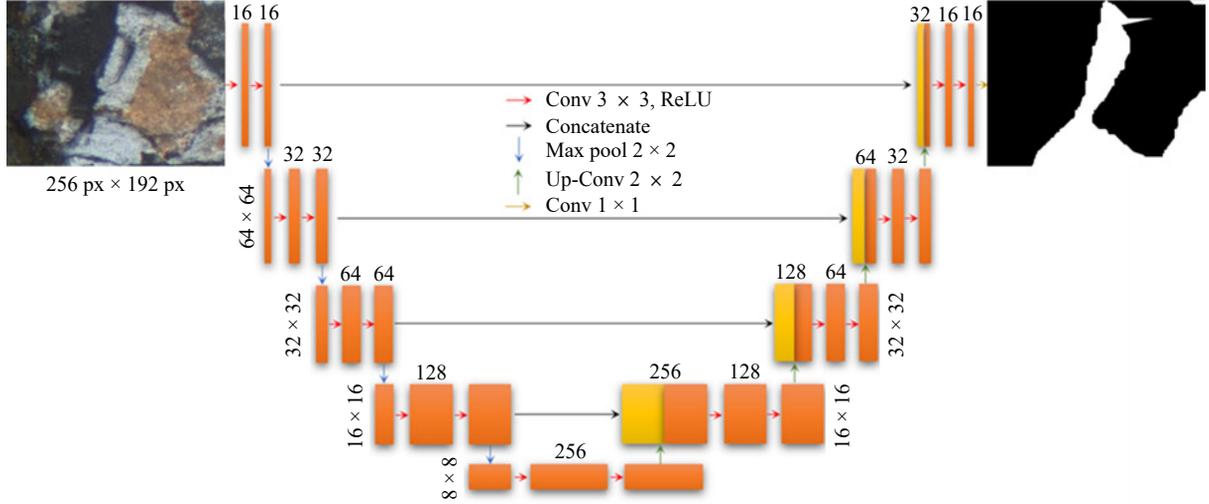
Fig. 6.　U-Net architecture.
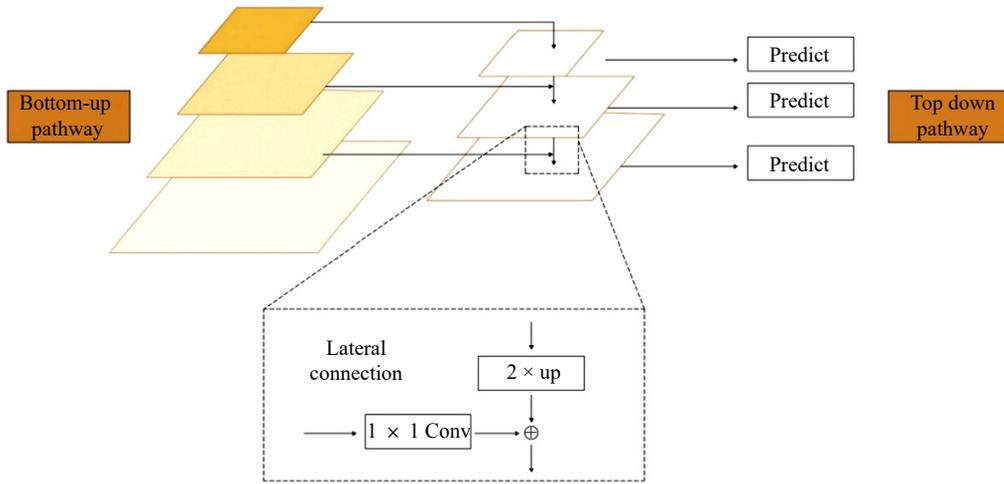
Fig. 7.　An illustration of the FPN architecture.

stance, for RGB images, there are 3 channels corresponding to the red (R), green (G), and blue (B) color channels. FPN typically uses a backbone such as ResNet to extract features at different stages, creating feature maps P_k at different scales. These feature maps are generated at varying resolutions, for example:

$$
\begin{cases}
P_2 \in \mathbf{R}^{\frac{H}{4} \times \frac{W}{4} \times D_2} \\
P_3 \in \mathbf{R}^{\frac{H}{8} \times \frac{W}{8} \times D_3} \\
P_4 \in \mathbf{R}^{\frac{H}{16} \times \frac{W}{16} \times D_4} \\
P_5 \in \mathbf{R}^{\frac{H}{32} \times \frac{W}{32} \times D_5}
\end{cases}
\tag{7}
$$

where D_k represents the number of channels (or depth) at each scale, usually increasing as the resolution decreases. Typically, $D_2 \approx D_3 \approx D_4 \approx D_5$, where $D_k = 256$.

In the FPN, the highest-resolution feature map P_5 is up-sampled using a factor of 2 to match the spatial dimensions of P_4. Mathematically, this is:

$$
P_4^{\text{up}} = \text{Upsample}(P_5)
\tag{8}
$$

This upsampling can be done using bilinear interpolation

or transposed convolutions. The upsampled feature map P_4^{up} is then added to P_4:

$$
P_4' = P_4^{\text{up}} + P_4
\tag{9}
$$

This process continues until all the feature maps from the top layers have been processed:

$$
\begin{cases}
P_3' = \text{Upsample}(P_4') + P_3 \\
P_2' = \text{Upsample}(P_3') + P_2
\end{cases}
\tag{10}
$$

where $\text{Upsample}(P_4')$ refers to the process of increasing the spatial resolution (height and width) of the feature map P_4 to match the resolution of the next lower-level feature map (P_3). Similarly, $\text{Upsample}(P_3')$ means increasing the spatial resolution of P_3 to match the resolution of P_2. Lateral connections provide a direct path from lower-resolution layers to higher-resolution layers to maintain fine-grained spatial information. This connection is established by 1×1 convolutions on each feature map before the addition:

$$
P_k = \text{Conv}1 \times 1(P_k)
\tag{11}
$$

where $\text{Conv}1 \times 1(P_k)$ refers to applying a 1×1 convolution on the feature map at level k. This ensures the dimensional consistency before merging feature maps. The final

feature maps are processed to predict pixel-wise segmentation. Each P'_k is either upsampled to match the dimensions of P_2, concatenated, and passed through a final convolutional layer for segmentation.

Let's assume we're classifying each pixel into N_{class} categories. The final prediction map would have dimensions $\mathbf{R}^{H \times W \times N_{class}}$. The softmax activation function is applied to produce probability distributions across all classes for each pixel. The model is typically trained using a loss function such as pixel-wise cross-entropy:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \lg(\hat{y}_{i,c}) \tag{12}$$

where $y_{i,c}$ is the true label for pixel i and class c, and $\hat{y}_{i,c}$ is the predicted probability.

By incorporating these multi-scale features and top-down pathway refinement, FPN achieves better segmentation performance, especially for detecting objects across different scales.

2.2.4. LinkNet

LinkNet is designed for efficient semantic segmentation using a streamlined encoder–decoder architecture. Its defining feature is the direct linkage between each encoder and decoder block through shortcut connections, which facilitate the retention and restoration of spatial and feature information lost during down-sampling. As shown in Fig. 8, it employs an encoder–decoder framework, uniquely integrated with link connections that facilitate the flow of feature maps from the encoder directly to the decoder.

Let $I \in \mathbf{R}^{H \times W \times C}$ be the input image. The encoder generates feature maps E_k, where k represents the level:

$$E_k = f_k \tag{13}$$

Each level consists of convolution, batch normalization, and activation operations. For the decoder, the feature maps from the encoder are gradually upsampled and combined via

Fig. 8.   Illustration of the LinkNet architecture.

skip connections. The decoder maps are denoted as:

$$D_k = \text{Upsample}(E_{k+1}) + E_k \tag{14}$$

where the feature map from the next level E_{k+1} is upsampled and added to E_k, ensuring spatial detail preservation.

The final output $O \in \mathbf{R}^{H \times W \times C_{out}}$ is achieved by refining the decoder's output through convolutional layers to predict segmentation masks. Where, O represents the final output feature map or segmentation mask produced by the network, C_{out} refers to the number of segmentation classes. LinkNet is optimized for computational efficiency, allowing real-time segmentation with fewer parameters than traditional encoder–decoder networks like U-Net.

2.2.5. Evaluation metrics

In image segmentation, evaluating the performance of models is crucial, and several metrics are commonly utilized for this purpose: accuracy, loss, specificity, sensitivity, precision, recall, F1 score, intersection over union (IoU), dice coefficient and area under the curve (AUC). These metrics assess the effectiveness of segmentation models in distinguishing between distinct image regions or objects, providing insights into the quality of segmentation.

Accuracy is the most intuitive performance measure, and it is simply a ratio of correctly predicted observations to the total observations. It is suitable for binary and multiclass classification problems. Accuracy is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative rates, respectively.

Specificity, also known as the true negative rate, measures the proportion of actual negatives that are correctly identified as such (e.g., the percentage of healthy people who are correctly identified as not having the condition):

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{16}$$

Sensitivity, also known as the true positive rate, measures the proportion of actual positives that are correctly identified as such (e.g., the percentage of sick people who are correctly diagnosed):

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{17}$$

Precision measures the model's accuracy in identifying relevant instances, avoiding false positives, and is defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{18}$$

Recall evaluates the model's capability to identify all relevant instances within the dataset:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{19}$$

The F1 score (Dice) combines precision and recall, offering a balanced measure of a model's performance, particularly useful in scenarios with imbalanced datasets:
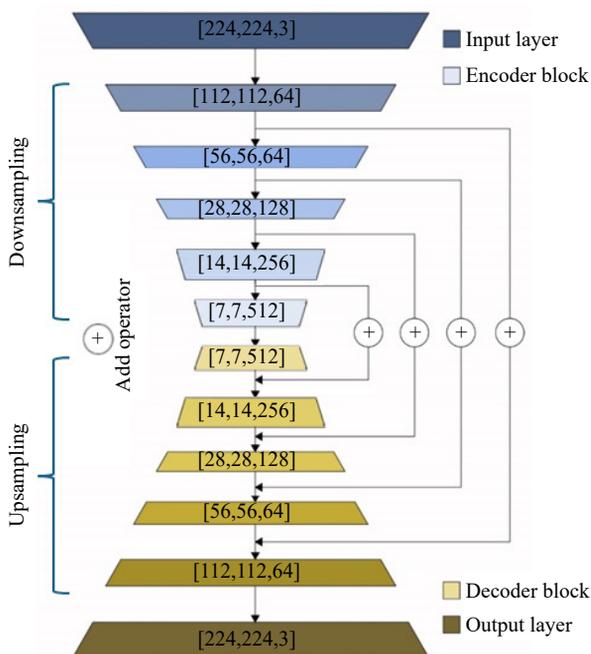
$$F1 = \text{Dice coefficient} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} =$$

$$\frac{2TP}{2TP + FP + FN} \tag{20}$$

IoU quantifies the overlap between the predicted and ground truth regions, serving as a measure of similarity:

$$IoU = \frac{TP}{TP + FP + FN} \tag{21}$$

Lastly, AUC measures the overall performance of binary classifiers across various thresholds by plotting the true positive rate against the false positive rate, providing a comprehensive assessment of classifier effectiveness.

These metrics are integral to selecting the most appropriate segmentation models for specific applications, as they highlight different aspects of model performance, from accuracy to the balance between precision and recall.

3. Results and discussion

3.1. Performance analyses

The comparative analysis of the four models—PSPNet, U-Net, FPN, and LinkNet—is shown in Figs. 9–12. In Fig. 9, the performance metrics for PSPNet show its robustness across all measured parameters. Both training and validation accuracy curves demonstrate strong performance, with training accuracy nearing 0.95 and validation accuracy slightly above 0.9. This suggests that PSPNet maintains high accuracy on both seen and unseen data. The training loss decreases to about 0.3, and validation loss converges around 0.4, indicating good generalization with minimal overfitting. The IoU scores are impressive, with training IoU close to 0.8 and validation IoU also around 0.75, indicating effective seg-

mentation performance. The Dice coefficient curves show similar trends, with training reaching 0.85 and validation around 0.8, further supporting the robust performance of the model.

The performance metrics for U-Net are illustrated in four plots in Fig.10. The accuracy for both training and validation sets improves steadily over the epochs, with the training accuracy approaching 0.98 and validation accuracy stabilizing around 0.85. This indicates a strong learning capability with some overfitting. The training loss decreases significantly, stabilizing around 0.2, while the validation loss levels off around 0.75. The discrepancy between training and validation loss suggests overfitting. The IoU score for training data approaches 0.9, whereas the validation IoU remains around 0.64, indicating that while the model performs well on training data, it generalizes less effectively on unseen data. The Dice coefficient follows a similar pattern, with training exceeding 0.9 and validation stabilizing around 0.75, further highlighting the overfitting issue.

In Fig. 11, the FPN model exhibits steady learning progress but shows some overfitting, similar to U-Net. The training accuracy approaches 0.98, while the validation accuracy stabilizes around 0.86. This significant gap suggests overfitting. Training loss decreases to below 0.2, while validation loss levels off around 0.6, reinforcing the overfitting observation. The IoU for training data reaches about 0.95, but validation IoU remains around 0.68, indicating less effective generalization. The Dice coefficient for training nears 0.98, while validation remains around 0.78, indicating similar overfitting issues.

In Fig. 12, LinkNet also shows a consistent performance with some overfitting, though it performs better in some as-
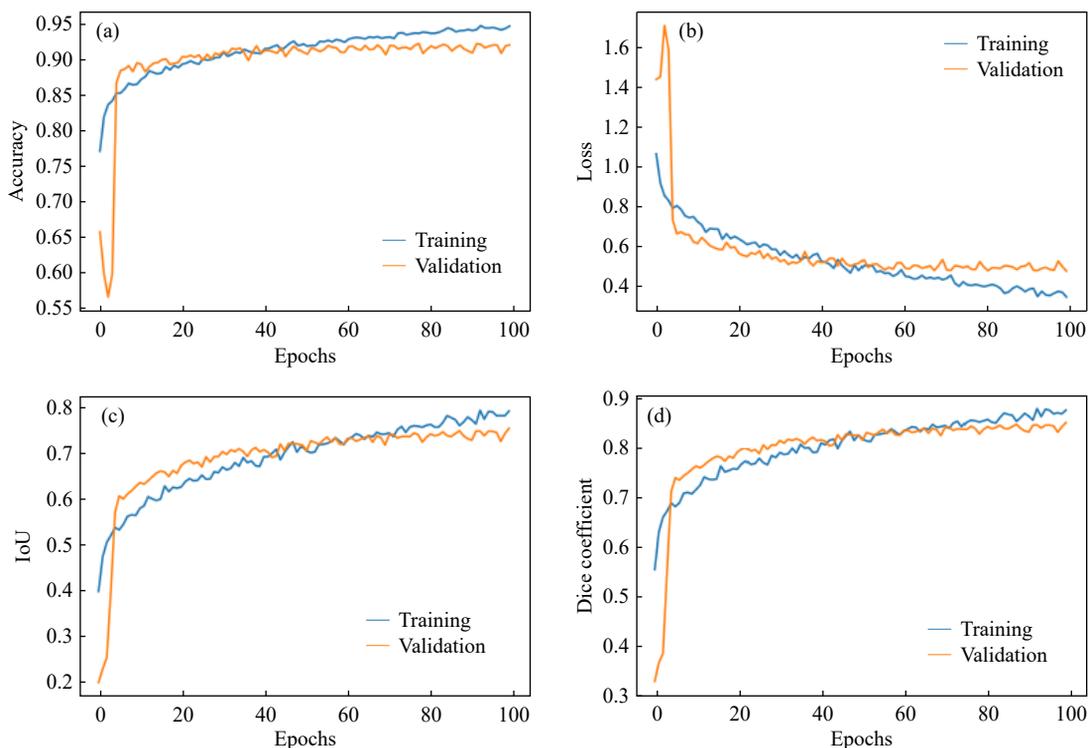
Fig. 9. Performance metrics for PSPNet: (a) accuracy; (b) loss; (c) IoU; (d) Dice coefficient.
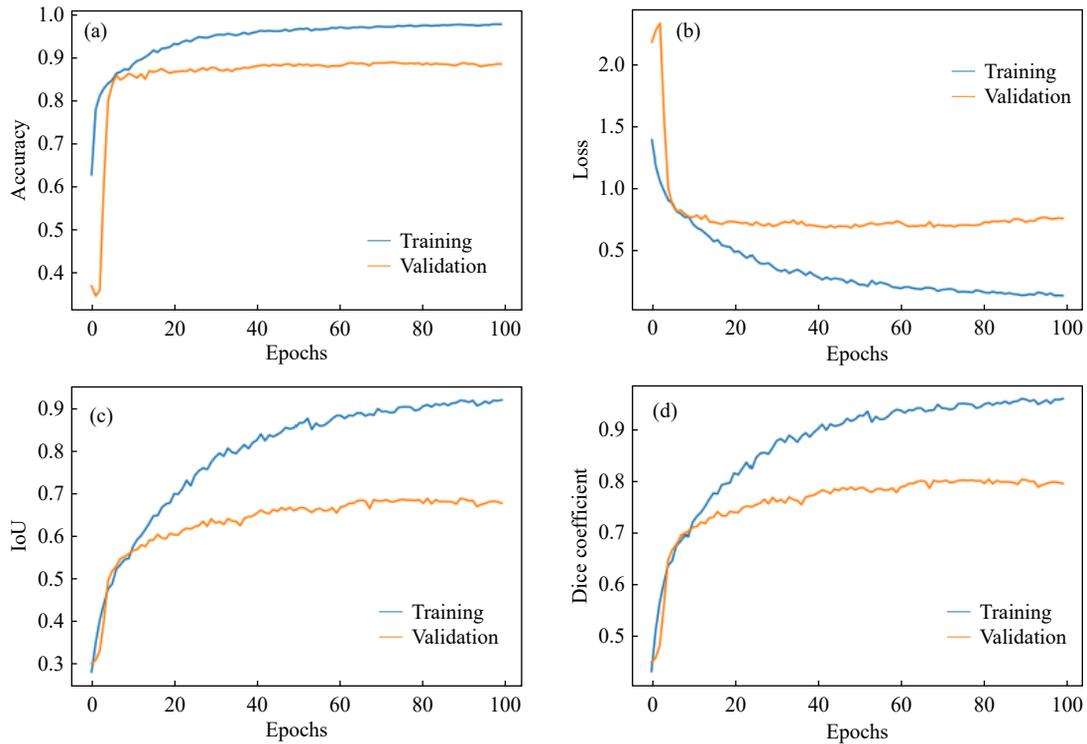
Fig. 10.   Performance metrics for U-Net: (a) accuracy; (b) loss; (c) IoU; (d) Dice coefficient.
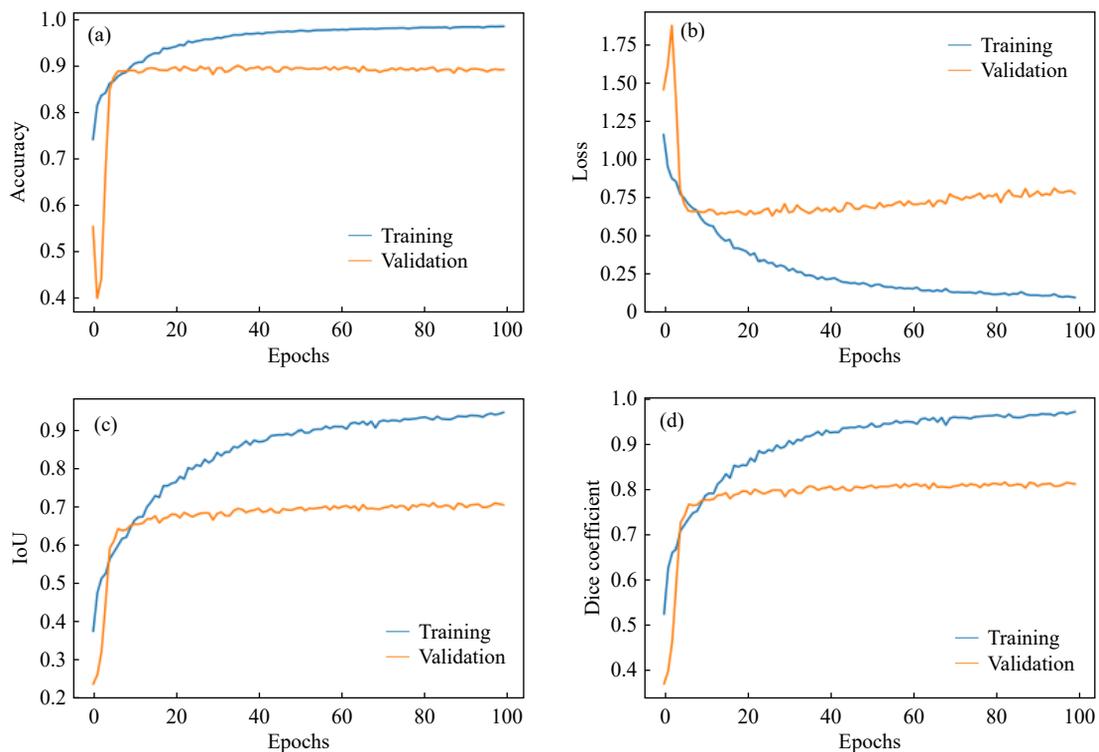
Fig. 11.   Performance metrics for FPN: (a) accuracy; (b) loss; (c) IoU; (d) dice coefficient.

pects compared to FPN and U-Net. Training accuracy approaches 0.95, while validation accuracy stabilizes around 0.85, showing better generalization compared to FPN. Training loss decreases to about 0.3, while validation loss levels off around 0.5, indicating moderate overfitting. Training IoU reaches about 0.90, with validation IoU around 0.65, showing better segmentation performance compared to U-Net and FPN. The Dice coefficient for training approaches 0.95,

while validation stabilizes around 0.78, suggesting reasonable generalization capability.

The comparative analysis of the four models—PSPNet, U-Net, FPN, and LinkNet—reveals that PSPNet consistently delivers the most robust performance across all metrics, indicating effective generalization and high accuracy. U-Net and FPN show significant overfitting, as evidenced by the large gaps between training and validation metrics. LinkNet,
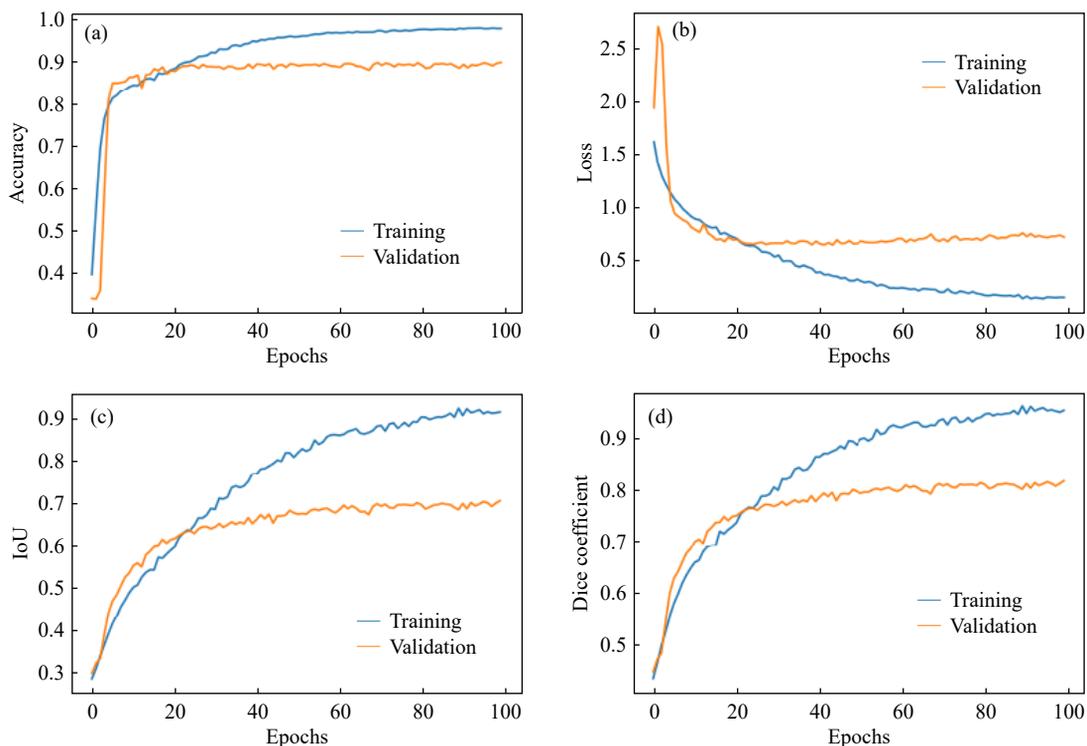
Fig. 12.   Performance metrics for LinkNet: (a) accuracy; (b) loss; (c) IoU; (d) Dice coefficient.

while also overfitting to some extent, performs better than U-Net and FPN in terms of generalization. Future work should focus on addressing overfitting issues, especially for U-Net and FPN, to enhance their performance on unseen data.

3.2. Performance evaluation criteria

In this study, four prominent segmentation models—PSPNet, U-Net, FPN, and LinkNet—are evaluated across a spectrum of performance metrics including accuracy, precision, recall, F1 score, sensitivity, specificity, AUC, and IoU. The analysis, presented in a radar chart format shown in Fig. 13, offers a comprehensive visual comparison of each model's capabilities.

PSPNet emerges as the superior model, excelling particularly in accuracy (0.924395), specificity (0.938460), and AUC (0.917368). The high specificity and AUC values suggest that PSPNet is highly effective at distinguishing both positive and negative classes, making it an optimal choice for applications demanding high precision in predictions. Additionally, PSPNet also shows robust performance in recall (0.896275) and IoU (0.798084), indicating a proficient handling of true positive identifications and significant overlap between the predicted and actual positive regions, respectively.

U-Net, while offering moderate performance across all metrics, exhibits relatively lower scores in precision (0.850823) and IoU (0.707615). The lower IoU score, in particular, suggests a reduced efficacy in capturing the overlap between predicted and true positive areas, which could limit its utility in applications where precise segmentation is critical.

FPN demonstrates commendable recall (0.878658) and sensitivity (0.878658), indicating its capability to effectively
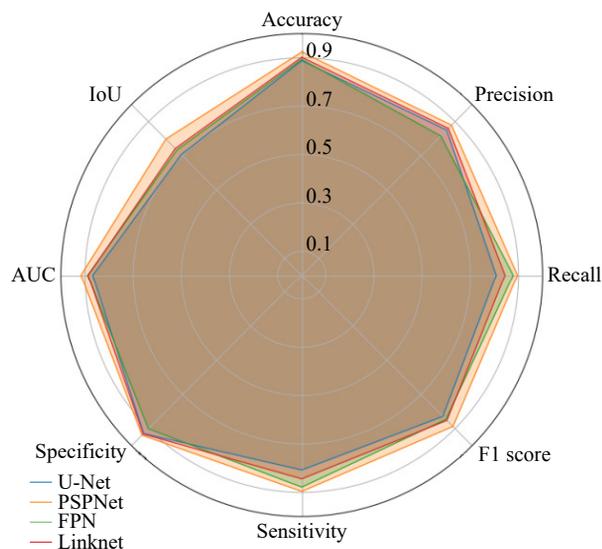
Fig. 13.   Comparison of four models using radar chart.

detect most positives. However, its precision (0.816231) and AUC (0.889833) scores are comparatively lower, which might lead to higher false positive rates. This trade-off highlights FPN's suitability in scenarios where high sensitivity is prioritized over precision.

LinkNet presents a balanced profile with no significant leads in any specific metric but consistent performance across the board. Its specificity (0.931725) and IoU (0.742880) are particularly noteworthy, suggesting a reliable capacity in specific identifications and adequate overlap in segmented regions. This model's balanced attributes render it versatile for diverse segmentation tasks.

In conclusion, our comparative analysis underscores PSPNet's overall dominance across multiple metrics, estab-

lishing it as the preferable model for high-stakes segmentation tasks. However, the choice of model should still be tailored to specific application needs, considering the trade-offs highlighted between sensitivity and precision among the models evaluated.

3.3. Experimental results

Figs. 14–17 comprehensively compare the performance of four different segmentation models—PSPNet, U-Net, FPN, and LinkNet—in detecting quartz minerals. IoU scores of these models were evaluated along with their prediction masks, real masks and segmentation images. The overall performance of each model in different scenarios and important observations are presented below.

In Fig.14, PSPNet achieved the highest IoU score of 0.736. The predicted mask closely aligns with the actual mask, demonstrating high accuracy in segmenting the quartz mineral region. This result highlights PSPNet's robustness in accurately identifying mineral boundaries. U-Net attained an IoU score of 0.669. The predicted mask, while reasonably accurate, shows slight discrepancies compared to PSPNet. This suggests that U-Net is effective but less precise in segmenting quartz minerals under certain conditions. FPN recorded an IoU score of 0.674, comparable to U-Net. The predicted mask effectively delineates the quartz mineral, indicating FPN's capability to perform accurate segmentation. LinkNet displayed an IoU score of 0.672, slightly lower than FPN but still demonstrating effective segmentation. The performance is robust, although marginally less accurate than PSPNet.

Fig. 15 shows that PSPNet delivered an exceptional IoU
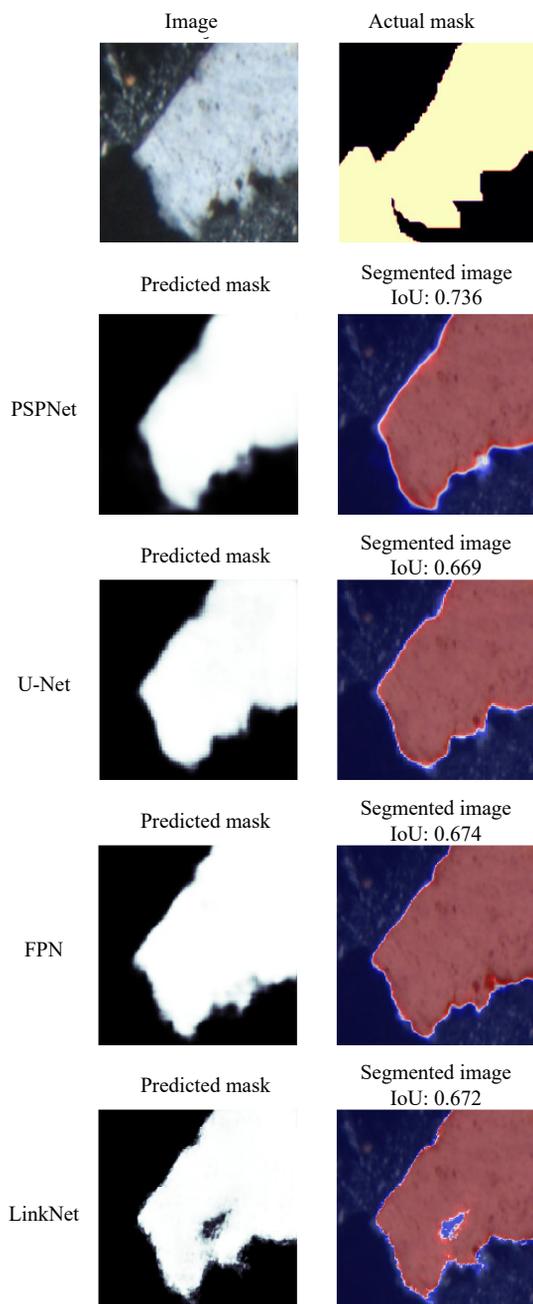
Fig. 14.   Automatic identification example 1 of quartz minerals.
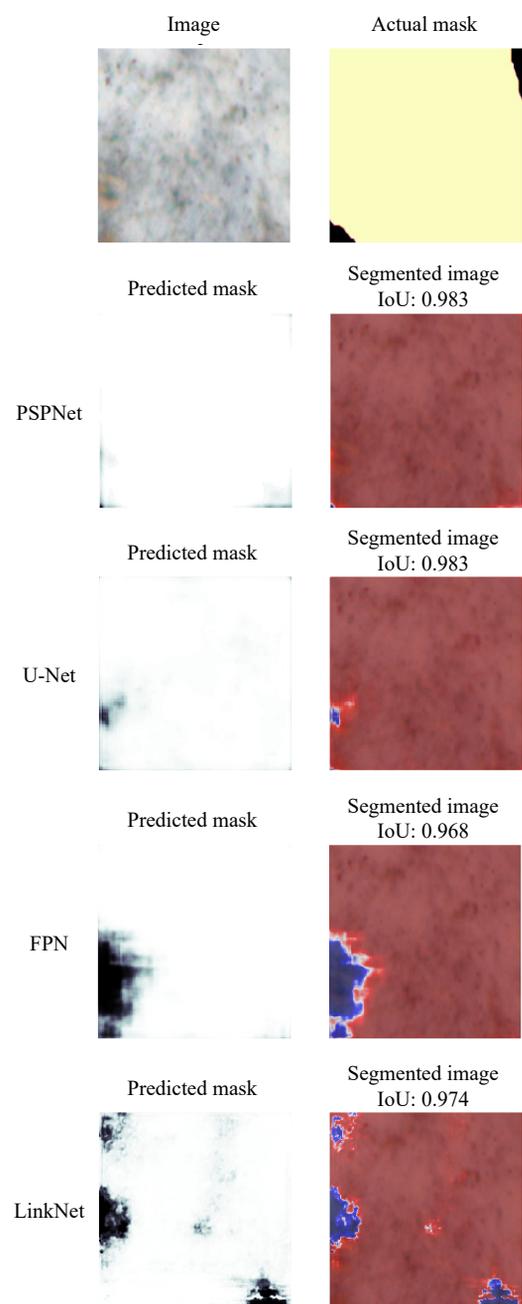
Fig. 15.   Automatic identification example 2 of quartz minerals.

score of 0.983, with the predicted mask nearly perfectly matching the actual mask. This underscores PSPNet's superior accuracy in quartz mineral segmentation. U-Net matched PSPNet with an IoU score of 0.983, indicating equally high performance in this specific example. This demonstrates U-Net's potential to achieve precise segmentation in optimal conditions. FPN achieved a slightly lower IoU score of 0.968, with minor inaccuracies in the predicted mask. Despite this, FPN shows strong performance in segmenting quartz minerals. LinkNet scored 0.974, demonstrating high accuracy in segmentation, comparable to PSPNet and U-Net. This indicates LinkNet's effectiveness in identifying mineral regions accurately.

In Fig. 16, PSPNet achieved an IoU score of 0.847, despite the generally high performance. The predicted mask shows some inaccuracies, indicating challenges in segment-

ing under less ideal conditions. U-Net exhibited a significantly lower IoU score of 0.435, struggling with accurate segmentation in this instance. The results suggest that U-Net may face difficulties in complex scenarios. FPN recorded an IoU score of 0.707, performing better than U-Net. The predicted mask identifies the quartz region, although with notable inaccuracies. LinkNet displayed an IoU score of 0.690, indicating moderate performance. The predicted mask is less accurate compared to PSPNet, reflecting the challenges in low-performance scenarios.

In Fig. 17, PSPNet demonstrated a high IoU score of 0.907, accurately identifying the quartz mineral despite the challenging conditions. This reaffirms PSPNet's robustness in diverse scenarios. U-Net attained an IoU score of 0.836, showing good performance but with some inaccuracies in the predicted mask. This indicates U-Net's potential, though it
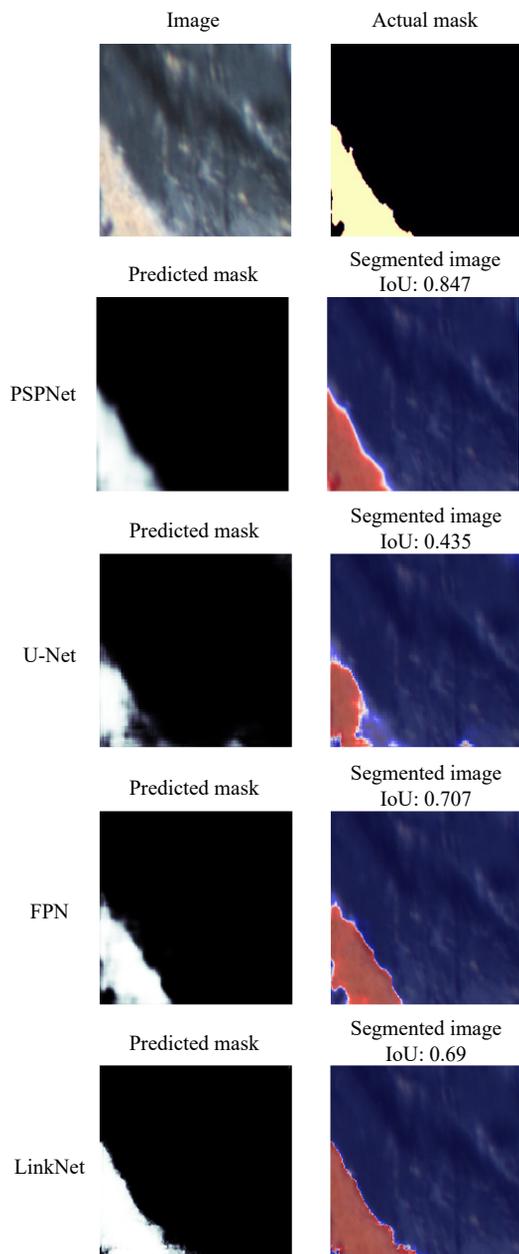
Fig. 16.   Automatic identification example 3 of quartz minerals.
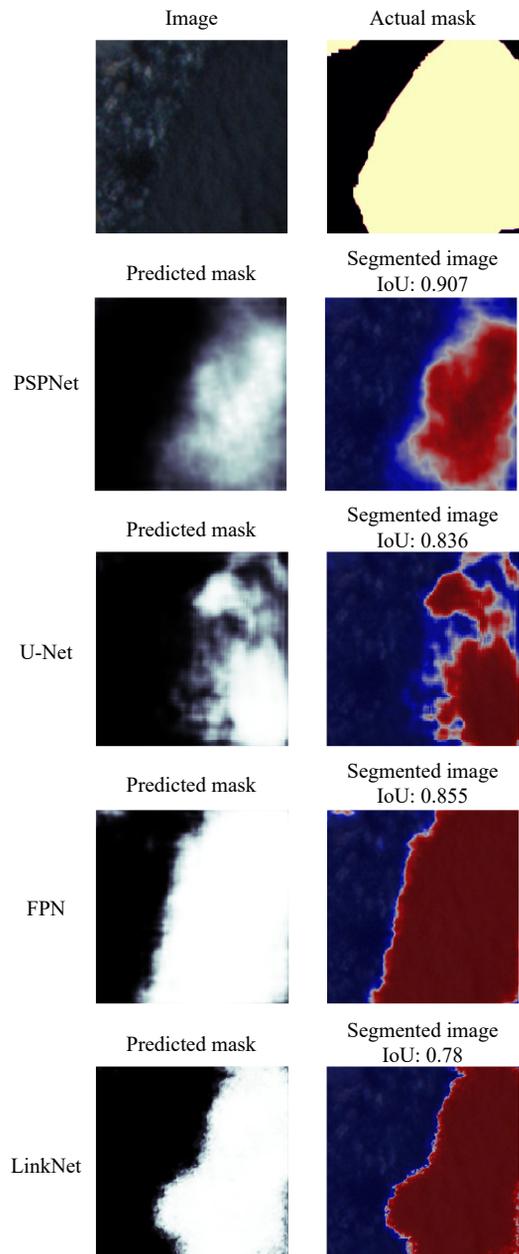
Fig. 17.   Automatic identification example 4 of quartz minerals..

814

Int. J. Miner. Metall. Mater., Vol. 32, No. 4, Apr. 2025

may require further refinement. FPN achieved an IoU score of 0.855, indicating effective segmentation with minor inaccuracies. This result underscores FPN's reliability in varying conditions. LinkNet recorded an IoU score of 0.780, demonstrating reasonable accuracy but lower performance compared to PSPNet and FPN. This suggests that while LinkNet is effective, there is room for improvement.

The comparative analysis of segmentation models for quartz mineral identification reveals PSPNet as the most robust and reliable model, consistently delivering high IoU scores and accurate segmentation masks. U-Net, FPN, and LinkNet also show potential, particularly in high-performance examples, but face challenges in more complex scenarios. Future work could explore hybrid approaches or model improvements to enhance segmentation accuracy across diverse conditions.

Fig. 18 presents examples where all four models—PSPNet, U-Net, FPN, and LinkNet—exhibited low performance, highlighting their inadequacies in challenging segmentation tasks. Although generally demonstrating high performance, PSPNet achieved a low IoU score in this instance, failing to accurately segment the quartz mineral. This indicates that PSPNet can struggle under certain conditions. U-Net encountered significant difficulties with complex and low-contrast images, resulting in the second lowest IoU scores observed. This highlights U-Net's limitations in challenging segmentation scenarios. FPN performed slightly better than the other models but still failed to accurately delineate mineral boundaries, as evidenced by its relatively low IoU score. This suggests that FPN has room for improvement in handling difficult cases. LinkNet recorded the lowest IoU score, making it the least effective model for mineral segmentation in this example. The predicted mask was highly inaccurate, demonstrating LinkNet's struggles with this task.

These results illustrate that the performance of segmentation models can vary significantly under different conditions and that there is a need for further improvement. The model's lower performance under these scenarios can be attributed to several factors. In low-contrast and complex images, PSPNet may struggle to differentiate subtle mineral boundaries, especially when the feature extraction process encounters noise or lacks distinguishing characteristics. This can result in misclassification or incomplete segmentation. Future research will investigate more advanced techniques to address these issues, such as incorporating attention mechanisms, improving multi-scale feature extraction, and using more diverse training datasets to increase robustness.

4. Conclusions

This study presents a novel approach for the automatic identification of quartz minerals using deep learning techniques combined with hyperspectral imaging. The results demonstrate the significant advancements in the efficiency and accuracy of mineral recognition brought about by this method. By integrating four advanced semantic segmenta-
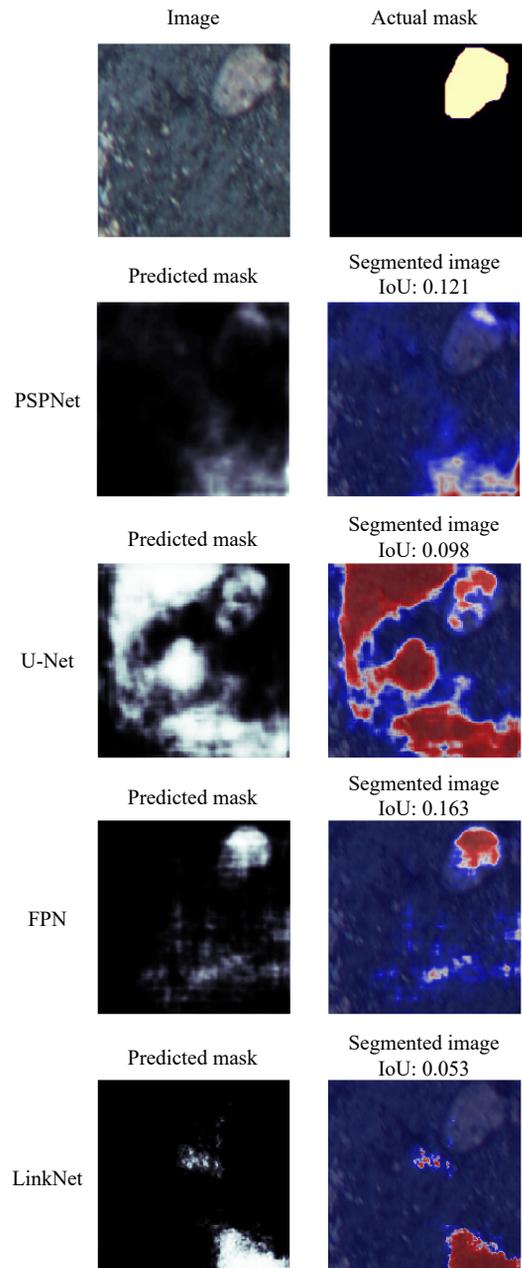
Fig. 18. Automatic identification examples of quartz minerals demonstrating low performance.

tion models—PSPNet, U-Net, FPN, and LinkNet—this research offers a comprehensive analysis and comparison of their performance in accurately recognizing quartz minerals.

The innovative aspect of this work lies in the application of deep learning and hyperspectral imaging to automate a traditionally manual and expertise-driven process. Unlike conventional optical methods, our approach leverages modern AI techniques to expedite and enhance mineral identification, setting a new standard in the field of mineralogical analysis. This study utilized a dataset comprising 120 thin sections prepared from 20 rock samples, producing 2470 images, which were divided into training and testing sets. Expert-reviewed images were masked and organized for model training, providing a robust foundation for deep learning applications.

The experimental results highlight the superior performance of the PSPNet model, which consistently outperformed the other models across multiple metrics, including accuracy, specificity, and IoU scores. PSPNet demonstrated exceptional accuracy and reliability in segmenting quartz from complex geological samples, achieving the highest IoU scores in various performance scenarios. The U-Net, FPN, and LinkNet models also showed potential, particularly in high-performance examples, but faced challenges in more complex scenarios, indicating the need for further refinement.

Despite the overall success of these models, certain conditions revealed limitations in their segmentation accuracy, especially in challenging and low-contrast images. This suggests that while the current models provide a substantial improvement over traditional methods, there is still room for enhancement to ensure more reliable and accurate outcomes under diverse conditions.

In conclusion, this study significantly contributes to the field of automated mineralogical analysis by demonstrating the practical utility and high performance of deep learning models. Future work should focus on addressing the observed limitations, exploring hybrid approaches, and refining model architectures to improve segmentation accuracy, particularly in challenging scenarios. The work could indeed focus on techniques such as data augmentation and regularization to improve the performance of these models. Data augmentation, including random transformations like cropping, flipping, and brightness adjustments, would introduce greater variability to the dataset, effectively reducing overfitting. Additionally, regularization techniques such as L2 regularization or dropout can further enhance model generalization. Another avenue to explore would be the implementation of early stopping to avoid overfitting during the training phase. Moreover, expanding the dataset with diverse and complex examples would improve robustness and lead to more generalized model performance, which can be critical in ensuring accurate segmentation across different applications.

The continued development and application of these advanced techniques will further enhance the efficiency and reliability of mineral identification, providing valuable tools for geologists and advancing the broader field of geology.

Conflict of Interest

All authors do not have competing interests to declare.

References

[1] W.T. Chen, X.J. Li, and L.Z. Wang, *Remote Sensing Intelligent Interpretation for Mine Geological Environment: From Land Use and Land Cover Perspective*, Springer Nature, Singapore, 2022.

[2] M. Tian, K. Ma, Z.H. Liu, Q.J. Qiu, Y.J. Tan, and Z. Xie, Recognition of geological legends on a geological profile via an improved deep learning method with augmented data using transfer learning strategies, *Ore Geol. Rev.*, 153(2023), art. No. 105270.

[3] N. Agrawal, H. Govil, S. Chatterjee, G. Mishra, and S. Mukherjee, Evaluation of machine learning techniques with AVIRIS–NG dataset in the identification and mapping of minerals, *Adv. Space Res.*, 73(2024), No. 2, p. 1517.

[4] H.J. Wang, Intelligent identification of logging cuttings based on deep learning, *Energy Rep.*, 8(2022), p. 1.

[5] N. Agrawal, H. Govil, G. Mishra, M. Gupta and P.K. Srivastava, Evaluating the performance of prisma shortwave infrared imaging sensor for mapping hydrothermally altered and weathered minerals using the machine learning paradigm, *Remote Sens.*, 15(2023), No. 12, p. 3133.

[6] A. Gomez-Flores, S. Ilyas, G.W. Heyes, and H. Kim, A critical review of artificial intelligence in mineral concentration, *Miner. Eng.*, 189(2022), art. No. 107884.

[7] X. Liu, V. Chandra, A.I. Ramdani, R. Zuhlke, and V. Vahrenkamp, Using deep-learning to predict Dunham textures and depositional facies of carbonate rocks from thin sections, *Geoenergy Sci. Eng.*, 227(2023), art. No. 211906.

[8] R. Pires de Lima, D. Duarte, C. Nicholson, R. Slatt, and K.J. Marfurt, Petrographic microfacies classification with deep convolutional neural networks, *Comput. Geosci.*, 142(2020), art. No. 104481.

[9] N. Saxena, R.J. Day-Stirrat, A. Hows, and R. Hofmann, Application of deep learning for semantic segmentation of sandstone thin sections, *Comput. Geosci.*, 152(2021), art. No. 104778.

[10] R.G. Zuo, Y.H. Xiong, J. Wang, and E.J.M. Carranza, Deep learning and its application in geochemical mapping, *Earth Sci. Rev.*, 192(2019), p. 1.

[11] W.W. Chen, D.Q. Tong, S.C. Zhang, X.L. Zhang, and H.M. Zhao, Local PM_{10} and $PM_{2.5}$ emission inventories from agricultural tillage and harvest in northeastern China, *J. Environ. Sci.*, 57(2017), p. 15.

[12] Z.H. Xu, W. Ma, P. Lin, and Y.L. Hua, Deep learning of rock microscopic images for intelligent lithology identification: Neural network comparison and selection, *J. Rock Mech. Geotech. Eng.*, 14(2022), No. 4, p. 1140.

[13] W.T. Chen, S.B. Ouyang, J.W. Yang, X.J. Li, G.D. Zhou, and L.Z. Wang, JAGAN: A framework for complex land cover classification using Gaofen-5 AHSI images, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.*, 15(2022), p. 1591.

[14] D. Ali and S. Frimpong, Artificial intelligence, machine learning and process automation: Existing knowledge frontier and way forward for mining sector, *Artif. Intell. Rev.*, 53(2020), No. 8, p. 6025.

[15] T. Long, Z.B. Zhou, G. Hancke, Y. Bai, and Q. Gao, A review of artificial intelligence technologies in mineral identification: Classification and visualization, *J. Sens. Actuator Network*, 11(2022), No. 3, art. No. 50.

[16] T. Sun, H. Li, K.X. Wu, F. Chen, Z. Zhu, and Z.J. Hu, Data-driven predictive modelling of mineral prospectivity using machine learning and deep learning methods: A case study from southern Jiangxi province, China, *Minerals*, 10(2020), No. 2, art. No. 102.

[17] H.J. Zhao, K.W. Deng, N. Li, Z.W. Wang, and W. Wei, Hierarchical spatial-spectral feature extraction with long short term memory (LSTM) for mineral identification using hyperspectral imagery, *Sensors*, 20(2020), No. 23, art. No. 6854.

[18] N. Agrawal and H. Govil, A deep residual convolutional neural network for mineral classification, *Adv. Space Res.*, 71(2023), No. 8, p. 3186.

[19] Y.S. Chen, Z.H. Lin, X. Zhao, G. Wang, and Y.F. Gu, Deep learning-based classification of hyperspectral data, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.*, 7(2014), No. 6, p. 2094.

[20] D.Y. Li, Z.D. Liu, Q.Q. Zhu, C.X. Zhang, P. Xiao, and J.Y. Ma, Quantitative identification of mesoscopic failure mechanism in granite by deep learning method based on SEM images, *Rock Mech. Rock Eng.*, 56(2023), No. 7, p. 4833.

816

Int. J. Miner. Metall. Mater., Vol. 32, No. 4, Apr. 2025

[21] Z.D. Liu, D.Y. Li, Q.Q. Zhu, C.X. Zhang, J.Y. Ma, and J.J. Zhao, Intelligent method to experimentally identify the fracture mechanism of red sandstone, *Int. J. Miner. Metall. Mater.*, 30(2023), No. 11, p. 2134.

[22] E.J.Y. Koh, E. Amini, G.J. McLachlan, and N. Beaton, Utilising convolutional neural networks to perform fast automated modal mineralogy analysis for thin-section optical microscopy, *Miner. Eng.*, 173(2021), art. No. 107230.

[23] H. Liu, Y.L. Ren, X. Li, *et al*., Rock thin-section analysis and identification based on artificial intelligent technique, *Pet. Sci.*, 19(2022), No. 4, p. 1605.

[24] W.L. Chen, C.N. Ji, D. Xu, and N. Srinil, Wake patterns of freely vibrating side-by-side circular cylinders in laminar flows, *J. Fluids Struct.*, 89(2019), p. 82.

[25] T.E. Oliphant, *Guide to NumPy*, [2024–08–20], https://csc.uc-davis.edu/~chaos/courses/nlp/Software/NumPyBook.pdf

[26] J.D. Hunter, Matplotlib: A 2D graphics environment, *Comput. Sci. Eng.*, 9(2007), No. 3, p. 90.

[27] *Keras-Resources*, GitHub [2024–08–20], https://github.com/fchollet/keras-resources.

[28] *Segmentation Models*, GitHub [2024–08–20], https://github.com/qubvel/segmentation_models.

[29] H.S. Zhao, J.P. Shi, X.J. Qi, X.G. Wang, and J.Y. Jia, Pyramid scene parsing network, [in] *2017 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 2017, p. 6230.

[30] O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation, [in] *Medical Image Computing and Computer-assisted Intervention—MICCAI 2015*: *18th International Conference*, Munich, 2015, p. 234.

[31] T.Y. Lin, P. Dollár, R. Girshick, K.M. He, B. Hariharan, and S. Belongie, Feature pyramid networks for object detection, [in] *2017 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 2017, p. 936.

[32] A. Chaurasia and E. Culurciello, LinkNet: Exploiting encoder representations for efficient semantic segmentation, [in] *2017 IEEE Visual Communications and Image Processing* (*VCIP*), St. Petersburg, 2017, p. 1.

[33] J.X. Hu, L. Li, Y.J. Lin, F.G. Wu, and J.S. Zhao, A comparison and strategy of semantic segmentation on remote sensing images, [in] *15th International Conference on Natural Computation*, *Fuzzy Systems and Knowledge Discovery*, Kunming, 2019, p. 21.

[34] K.M. He, X.Y. Zhang, S.Q. Ren, and J. Sun, Deep residual learning for image recognition, [in] *2016 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 2016, p. 770.