GOSPODARKA SUROWCAMI MINERALNYMI – MINERAL RESOURCES MANAGEMENT

2026 Volume 42 Issue 2 Pages 69–87

DOI: 10.24425/gsm.2026.1425

XIAOYAN LUO[1], HAOXIANG ZOU[2], MENGZHEN WANG[3], WENCHONG TANG[4]

RA-UNet++ based image segmentation of adherent ores

Introduction

Intelligent mining seeks to establish a data-driven, digital enterprise by leveraging artificial intelligence, in which automated ore image processing becomes a critical component for efficiently handling complex and hazardous tasks. Within this framework, Particle Size Measurement (PSM), a method for quantifying the dimensions of ore from images, is one of the most pivotal tasks in mineral processing analysis. The particle size distribution of the ore serves as a key performance indicator that directly reflects the operational status of the crusher. By monitoring the particle size within the fine crushing stage, it is possible to gain timely insights into the crusher's performance, optimize process parameters, and enhance

✉ Corresponding Author: Xiaoyan Luo; e-mail: 1104556648@qq.com

[1] School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology Ganzhou, Jiangxi Province, China; e-mail: 1104556648@qq.com

[2] School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology Ganzhou, Jiangxi Province, China; ORCID iD: 0009-0002-9770-2174; e-mail: 2591319931@qq.com

[3] School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology Ganzhou, Jiangxi Province, China; e-mail: 1564322148@qq.com

[4] School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology Ganzhou, Jiangxi Province, China; e-mail: 1976678028@qq.com

overall crushing efficiency. This ultimately improves the recovery rate of valuable minerals, thereby yielding significant economic benefits for the mining operation (Wang et al. 2023).

In the adverse operational environment of a mine, the fidelity of ore image processing is impeded by numerous factors. These include external interference from airborne dust and non-uniform illumination, as well as intrinsic material properties such as significant variations in particle volume, color, and texture. Further complications arise from indistinct particle boundaries and the strong adhesion of ore particles. Images acquired from conveyor belts during the final stages of crushing present a particularly formidable challenge, as the material is characterized by a finer particle size distribution, heightened inter-particle adhesion, and a high degree of agglomeration. Consequently, these conditions render the accurate segmentation and analysis of individual particles substantially more difficult (Shrivastava et al. 2023). Currently, mainstream ore image segmentation techniques include thresholding, edge detection, clustering, and deep learning-based approaches (Xiao et al. 2020; Jin et al. 2021; Sun et al. 2022). Zhang et al. (2017) proposed a dual-window threshold segmentation method based on Fisher's discriminant analysis. This technique applies local thresholding to each pixel's neighborhood using Fisher's criterion, thereby improving segmentation efficiency. However, it struggles to accurately segment ores with heterogeneous types and complex features. Salinas et al. (2005) employed a threshold-based method to identify target ores before performing segmentation. While this approach effectively delineates ore contours and mitigates over-segmentation, its algorithmic complexity and low efficiency limit practical applications. Huang et al. (2022) applied the Canny edge detection operator for ore image segmentation. Although effective for smooth-surfaced ores, this method is highly sensitive to noise and texture and lacks robustness under complex imaging conditions. In summary, traditional approaches such as thresholding, edge detection, and clustering perform inadequately when dealing with ore images characterized by complex morphology, multiple classes, and strong adhesion. In contrast, deep learning-based segmentation has emerged as a promising research direction, offering a powerful solution for handling the challenges of complex ore image analysis.

Deep convolutional networks (DCNs) have significantly advanced the field of image segmentation due to their remarkable efficiency and precision. Prominent models such as FCN (Long et al. 2015), U-Net (Ronneberger et al. 2015), SegNet (Badrinarayanan et al. 2017), and deformable convolutional networks (Dai et al. 2017; Mac et al. 2019) have driven notable progress by enabling end-to-end semantic segmentation. Among these architectures, U-Net has emerged as a foundational model widely applied in medical imaging, remote sensing, and industrial vision tasks, benefiting from its symmetrical encoder-decoder structure and skip connections. However, original U-Net structures exhibit limitations when addressing multi-scale targets, intricate boundaries, and limited data scenarios, which have spurred researchers to propose numerous structural variants and enhancement strategies.

Naylor et al. (2018) proposed SegAN, an enhanced U-Net framework integrating generative adversarial networks to improve segmentation accuracy for tumor boundaries. Oktay et al. (2018) developed Attention U-Net, employing attention gates to highlight

critical regions, thereby significantly improving performance in organ segmentation tasks. Dash et al. (2019) introduced PsLSNet, a deep 29-layer U-Net variant specifically designed for automated segmentation of psoriasis lesions. Tavakolian et al. (2025) proposed a rapid organ segmentation network for abdominal CT images, combining attention mechanisms and residual compression modules within a U-Net-based structure, effectively enhancing both the precision and speed of liver and kidney segmentation. Addressing challenges in brain tumor segmentation, Chaurasia et al. (2025) developed a framework integrating multi-channel preprocessing and improved U-Net architecture to effectively handle noise interference and tumor heterogeneity. Chen (2025) systematically evaluated various U-Net architectures in medical image segmentation tasks, proposing an advanced U-Net enhanced with deep attention gates and edge-guided loss, demonstrating substantial performance improvements across multiple medical datasets. Additionally, U-Net is extensively employed in ore image segmentation tasks, facilitating automated segmentation and granularity extraction (Krizhevsky et al. 2012; Touvron et al. 2021). Mustafa et al. (2020) introduced Residual Connections (RC) into the U-Net framework to develop the Res-U-Net model, which significantly improves segmentation accuracy and processing efficiency for ore targets in high-resolution remote sensing images. By enhancing feature gradient propagation, the model optimizes ore size estimation performance and outperforms the conventional U-Net in both detection speed and segmentation precision. Yuan and Duan (2018) proposed a segmentation approach based on the Holistically-Nested Edge Detection (HED) model built upon the VGG network. The segmented results are further refined using skeletonization, followed by region identification and color labeling to address ore adhesion and occlusion. While this method performs well in identifying large ore particles, its accuracy in segmenting smaller particles remains limited. For underwater ore imagery, Song et al. (2019) applied a U-Net-based architecture for seabed ore segmentation. However, seabed ores are typically scattered and easier to distinguish, whereas ore images obtained from mining operations exhibit severe adhesion, making segmentation significantly more challenging. Li et al. (2023) proposed the DDR-U-Net, a model that incorporates a dynamic feature reorganization mechanism to enhance the accuracy of particle size distribution analysis. This approach achieves highly robust segmentation, particularly along particle boundaries. Nonetheless, due to challenges such as strongly adhered edges, complex surface features, and high visual similarity among ore particles, traditional U-Net models still struggle to achieve accurate segmentation, often leading to the misclassification of prism-like structures on ore surfaces.

To address the challenges of low adaptability, unsatisfactory segmentation performance, and difficulty in segmenting adhered ore edges in existing ore image segmentation algorithms, we propose an adherent ore image segmentation method based on RA-UNet++. First, the original "convolution + pooling" feature extraction modules in the UNet++ architecture are replaced with residual modules to deepen the network and enhance feature representation. Second, a self-attention mechanism is incorporated into the network to enable the model to emphasize salient features within specific local regions, thereby improving model sensitivity and prediction accuracy. Finally, a multiscale atrous spatial pyramid pooling

(ASPP) structure is integrated to enrich semantic information by expanding the receptive field, further enhancing segmentation performance. Experimental results demonstrate that the proposed method exhibits robust and superior segmentation performance, significantly improving the accuracy of ore image segmentation.

1. RA-UNet++ image segmentation method for adherent ores

1.1. General structure of the RA-UNet++ network

To address issues such as unclear ore particles, residual edges, and incomplete segmentation in ore images, this paper proposes the RA-UNet++ network, which comprises four key components: an encoder-decoder framework, residual blocks, an attention module, and an atrous spatial pyramid pooling (ASPP) module. The overall architecture of RA-UNet++ is illustrated in Figure 1. Specifically, the peripheral U-shaped network is responsible for deep feature extraction and upsampling, with the encoder positioned on the left and the decoder on the right.

The main features of the RA-UNet++ network are as follows:

1. The UNet++ architecture replaces the traditional UNet backbone, leading to enhanced network performance.
2. Residual modules are integrated into the feature extraction process to capture deeper and more complex representations.
3. A self-attention mechanism is embedded within the hidden layers to selectively extract salient features across multiple scales and hierarchical levels.
4. An ASPP module is incorporated at the encoder end to augment the network's capacity for capturing rich semantic context.

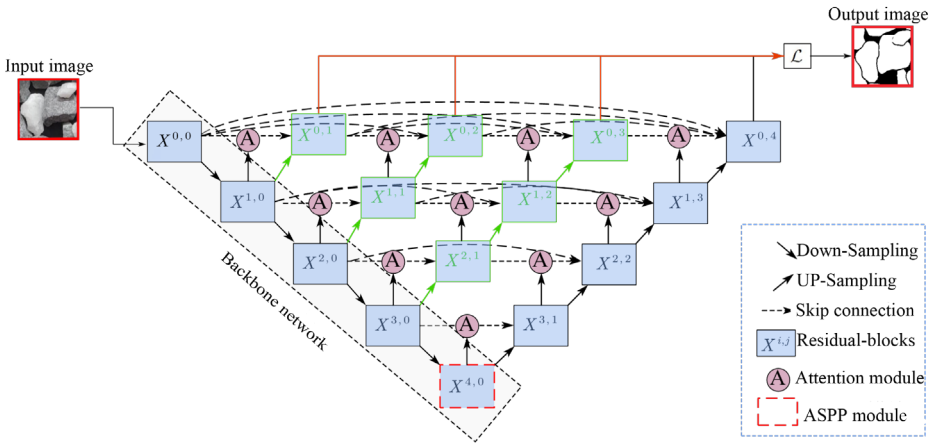

Fig. 1. RA-UNet++ network structure diagram

Rys. 1. Schemat struktury sieci RA-UNet++

1.2. UNet++ Networks

UNet++ network (Zhou et al. 2019) consists of three parts: encoder, jump connection, and decoder. The encoder extracts image features through a convolution operation, and its extracted high-resolution features can provide precise positioning of the target, and the jump connection enables the network to make better use of the feature information of different layers to improve the segmentation accuracy. The decoder downsizes the features through a pooling operation and maps the features onto the category labels by the fully-connected layer mapping to category labels. UNet++ realizes the spatial information fusion between different layers from shallow to deep by adjusting the long connections of U-Net to nested dense short connections, which enables each layer to preserve as much detail and global information as possible, i.e., connecting the high-resolution features to the up-sampling output features of the decoder. Meanwhile, the semantic information fusion process of the dense hopping connections between the modules of the same layer reduces the information loss in feature extraction and shortens the semantic gap between the encoder and the decoder, which enables the network model to fully and effectively capture the detailed features of the target. The UNet++ network structure is shown in Figure 2.

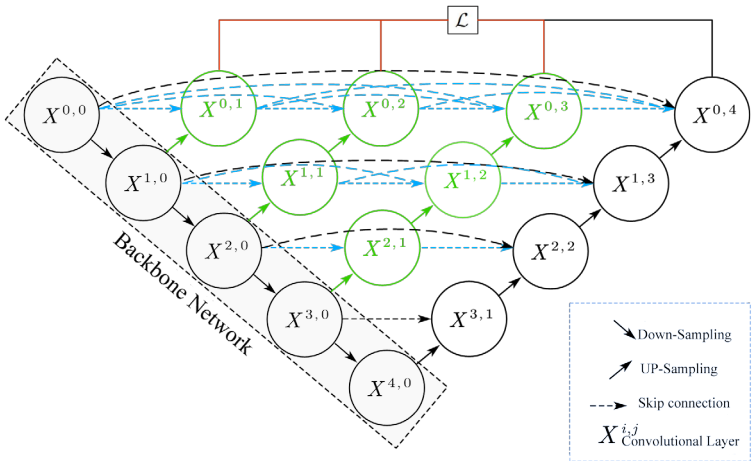

Fig. 2. UNet++ network structure diagram

Rys. 2. Schemat struktury sieci UNet++

1.3. Residual module

Ore images are characterized by high information density, irregular particle morphology, and complex backgrounds. Consequently, shallow network architectures, such as those comprising simple stacked convolutional and pooling layers, often struggle to extract

discriminative deep features. To address this limitation, it is essential to deepen the feature extraction network to capture richer hierarchical features and thereby improve model performance. However, significantly deeper networks are prone to issues like vanishing or exploding gradients, which can lead to performance degradation. Therefore, this paper adopts the Residual Network (ResNet) (He et al. 2016) as the feature extraction backbone.

As illustrated in Figure 3, a ResNet block learns a residual function $F(x)$ with respect to the block's input x, defining the output as $H(x) = F(x) + x$. The core idea is that if an identity mapping is optimal, the network can easily learn to set the residual $F(x)$ to zero. This ensures that the network's performance will not degrade with added depth. In practice, the stacked residual layers learn new features by adding non-zero residuals to the input features, thus enhancing the model's representational power. To reduce the parameter count and computational complexity, we employ the three-layer "bottleneck" residual block. This design utilizes 1×1 convolutions to first reduce and then restore channel dimensions, effectively minimizing computational load.

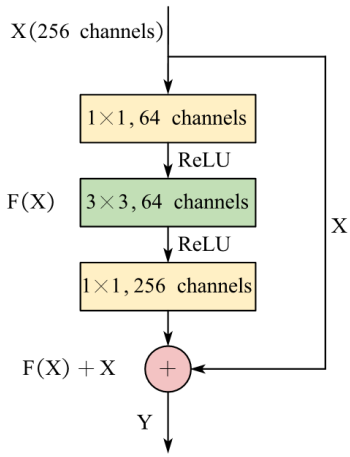

Fig. 3. Sketch of ResNet structure

Rys. 3. Schemat struktury sieci ResNet

1.4. Self-attention mechanism

The core idea of self-attention is to construct an "affinity matrix" of input features using learnable parameters, and to dynamically aggregate the features based on this matrix (Shaw 2018). This approach breaks through the limitations of the local receptive field in traditional convolutions, suppresses irrelevant regions in the input image, and highlights the important features from specific local regions. As a result, it enhances the model's sensitivity and prediction accuracy. In UNet++, the self-attention mechanism is incorporated by inserting a self-attention module before concatenating the features from each resolution of the encoder

with their corresponding features in the decoder. This module readjusts the encoder's output features, thus providing UNet++ with global context modeling capability. The self-attention mechanism used in this paper is shown in Figure 4.

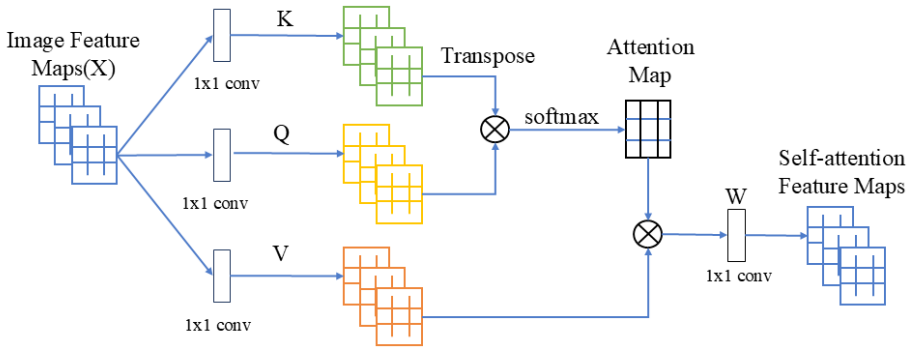

Fig. 4. Schematic diagram of the self-attention mechanism

Rys. 4. Schemat mechanizmu samo-uwagi

1.5. Atrous spatial pyramid pooling

The Atrous Spatial Pyramid Pooling (ASPP) (Yurtkulu et al. 2019) module is a multi-scale feature extraction block. As illustrated in Figure 5, it typically comprises several parallel atrous convolutional layers with different dilation rates, an image-level global average pooling layer, and a final 1×1 convolutional layer to fuse the resulting features. The primary function of the ASPP module is to probe an incoming feature map with multiple fields of view.

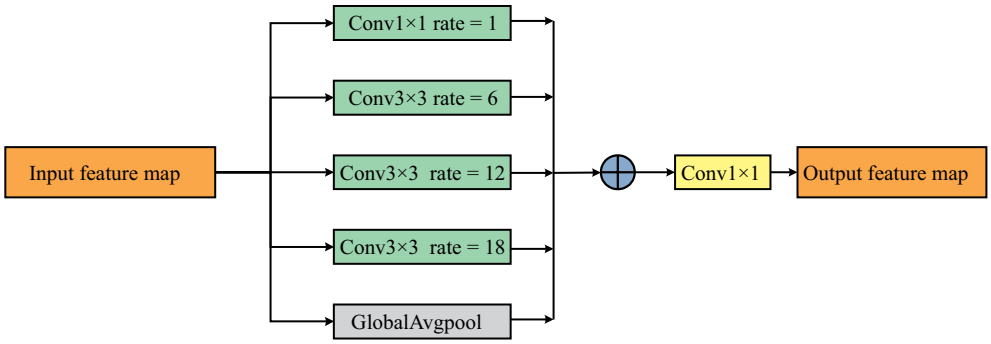

Fig. 5. ASPP module

Rys. 5. Moduł ASPP

By capturing contextual information at various scales and then merging these features, ASPP enhances the network's ability to understand global context and detailed semantic information simultaneously. In the context of ore image segmentation, the integration of an ASPP module allows the model to more robustly extract multi-scale semantic features, which is crucial for differentiating between complex ore textures and backgrounds.

1.6. Loss function

The loss function is used to measure how well the network predicts, and usually, the smaller its value, the more accurate the network's prediction and the better the fit. Therefore, it is very important to choose a suitable loss function in the experiment. In this paper, the loss function used in training is the Binary Cross Entropy loss function and the Dice loss function, using a combination of two different loss functions can alleviate the problems caused by the imbalance of segmentation categories, and also accelerate the learning speed of the network, and after testing, when α is taken to be 0.6, the training loss of the network is minimized. As shown in Equation (1).

$$L_{loss} = (1 - \alpha)L_{BCE} + \alpha L_{dice} \tag{1}$$

- L_{BCE} – dichotomous cross-entropy loss function,
- L_{dice} – Dice loss function,
- α – parameter that regulates the balance of weights between the two loss functions.

The Binary Cross Entropy (BCE) loss function ensures accurate segmentation of the background, thus improving segmentation accuracy. In addition, the use of the cross entropy loss function allows the neural network to converge more efficiently and accelerates the network training speed, the expression of which is:

$$L_{BCE} = \frac{1}{N}\sum_i L_i = \frac{1}{N}\sum_i -\left[y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)\right] \tag{2}$$

- L – loss,
- y_i – label of sample I,
- p_i – probability that sample i is predicted to be true.

The Dice loss function is derived from the Dice coefficient, which is a metric function used to measure the similarity of sets with a domain of values between [0,1] and is defined as:

$$Dice(X,Y)=\frac{2|X\cap Y|}{|X|+|Y|} \tag{3}$$

- X – prediction result,
- Y – manual labeling result,
- $|X\cap Y|$ – overlap between X and Y. The closer the value of Dice is to 1, the higher the segmentation accuracy and the better the segmentation effect of the algorithm.

In image segmentation, combining the Dice coefficients yields the definition of the Dice loss function as:

$$L_{dice}=1-\frac{2|X\cap Y|}{|X|+|Y|} \tag{4}$$

2. Datasets and preprocessing

The dataset employed in this study is a custom-collected collection of ore images sourced directly from mining production environments. It consists of 500 images, each with a resolution of 512 × 512 pixels. For each image, a corresponding pixel-wise segmentation mask was meticulously generated through manual annotation to serve as the ground truth. To prepare the data for training, we first applied bilateral filtering to each image. This step effectively reduces image noise while preserving the sharp edge details of the ore particles, which is critical for accurate boundary detection. Subsequently, we performed extensive data

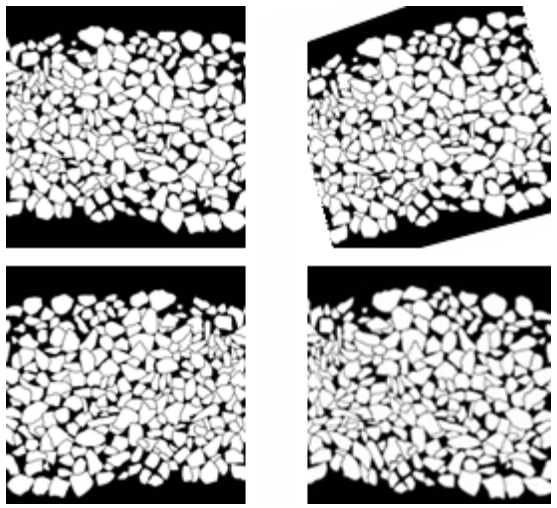

Fig. 6. Data enhancement results

Rys. 6. Wyniki wzbogacania danych

augmentation to increase the diversity of the training set and enhance the model's robustness against variations in appearance. The augmentation pipeline included several geometric transformations: horizontal flipping, vertical flipping, and random elastic deformations. As illustrated in Figure 6, these operations generate realistic variations of the original images. Through this process, the initial set of 500 images was expanded to a total of 2,000 images. This augmented dataset was then partitioned, with 90% (1,800 images) allocated for training and the remaining 10% (200 images) reserved for validation.

3. Experimentation and analysis

3.1. Training parameter settings

The model was trained using the RMSprop optimizer with an initial learning rate set to 1e-4. We trained the network for a total of 200 epochs with a batch size of 4. All experiments were conducted on a workstation equipped with the hardware specifications detailed in Table 1.

Table 1. Configuration of experimental hardware and software platform

Tabela 1. Konfiguracja eksperymentalnej platformy sprzętowej i programowej

Experimental platform configuration		Parameters
Hardware parameters	Processor (CPU)	Intel Core i7-10700@2.90GHz
	Random access memory (RAM)	16GB DDR4
	Graphics card (GPU)	NVIDIA GeForce RTX3060
Software system	Computer operating system	64-bit Windows 10 Professional operating system
	Framework environment	Pytorch-GPU2.4.0
	CUDA	CUDA Toolkit 10.2
	Programming language	Python

3.2. Evaluation indicators

In order to objectively evaluate the effect of ore image segmentation, four indexes, namely, pixel segmentation accuracy (*ACC*), mean intersection and merger ratio (*MIoU*), *F*1 score (*F*1), and ore segmentation accuracy, were selected as evaluation indexes, and the definitions of each index are shown in Equation (5) to Equation (10).

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

$$MIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{TP}{TP + FP + FN} \tag{6}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{7}$$

- TP – number of positive classes correctly predicted as positive classes,
- TN – number of negative classes correctly predicted as negative classes,
- FP – number of negative classes incorrectly predicted as positive classes,
- FN – number of positive classes incorrectly predicted as negative classes.

Precision stands for the precision rate and *Recall* stands for the recall rate, which are calculated by the formulas respectively:

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

In order to be able to compare the segmentation effect of different methods in a more targeted way, this paper introduces the ore segmentation accuracy, which is defined as shown in Equation (10):

$$\eta = \frac{N_1}{N_0} \tag{10}$$

- N_1 – number of ores correctly segmented ores during the test,
- N_0 – total number of ores in the original ore image.

3.3. Analysis of experimental results

In order to verify the effectiveness of the network structure, the ore image dataset is trained with UNet, and RA-UNet++ networks respectively, and the training results are tested and the results are compared on the test set, and the results are compared with the segmentation results of the traditional Otsu algorithm.

To validate the effectiveness of our proposed network, we conducted a comparative analysis against several baseline methods. We trained our model, hereafter referred to as RA-UNet++, and the standard U-Net on the same ore image dataset. Additionally, we benchmarked these deep learning models against a traditional image processing technique, Otsu's thresholding algorithm. Figure 7 presents a qualitative comparison of the segmentation results from these methods on sample test images. Each row displays the original input image, its corresponding ground-truth mask, and the segmentation outputs from Otsu's algorithm, U-Net, and our proposed RA-UNet++. To facilitate a more detailed examination of the performance differences, Figure 8 provides magnified views of specific

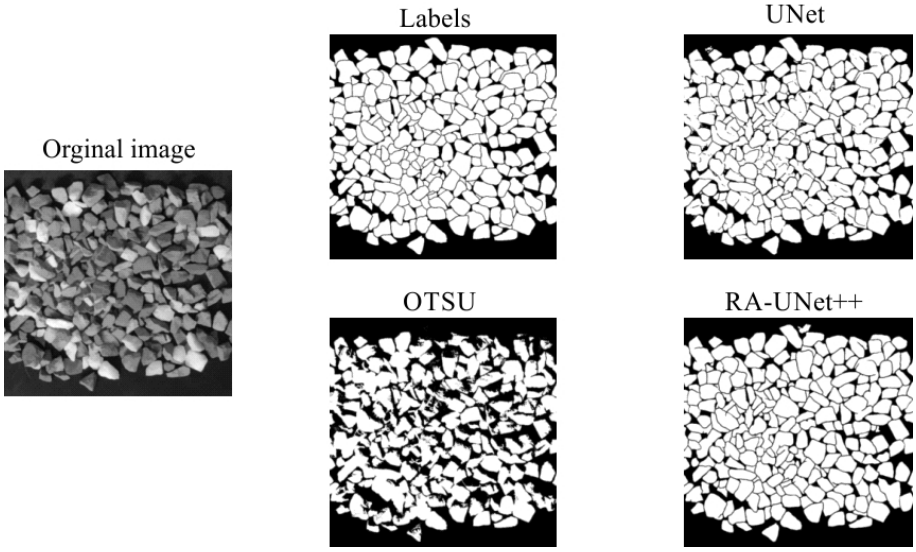

Fig. 7. Segmentation of image by different methods

Rys. 7. Segmentacja obrazu przy użyciu różnych metod

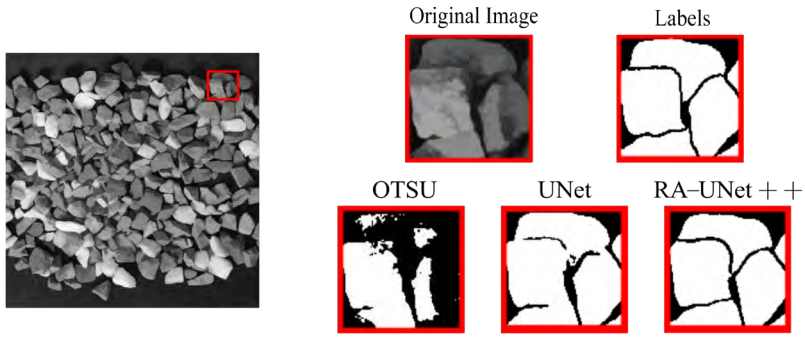

Fig. 8. Local enlargement

Rys. 8. Powiększenie lokalne

regions of interest. These zoomed-in images highlight the superior capability of our method in accurately delineating ore boundaries and preserving fine-grained details compared to the baseline approaches.

As observed from the detailed comparisons in Figure 8, the three methods exhibit markedly different levels of performance.

Otsu's algorithm: this traditional thresholding method yields the poorest results. It fails to delineate complete ore boundaries and introduces significant errors within the ore bodies themselves (i.e., under-segmentation). This is attributable to its reliance on a single global threshold, which is insufficient for images with complex ore types and variable illumination, leading to low adaptability.

U-Net: the baseline U-Net model demonstrates a substantial improvement over Otsu's algorithm, successfully segmenting the general shape of most ore particles. However, it struggles with a critical challenge: separating adjacent or touching ore particles. The model often incorrectly merges these "adherent" particles, failing to segment the fine boundaries between them.

Our Proposed RA-UNet++: in contrast, our method achieves the most accurate and robust segmentation among the three. It excels particularly in the challenging scenario of adherent ores, correctly identifying and separating particles that are in close contact. This superior performance indicates that our model, enhanced with attention and multi-scale features, can effectively capture the subtle edge cues necessary to distinguish individual ore instances, thereby preserving the integrity of each particle's boundary.

To further visualize the instance-level segmentation quality, we generated overlay images as shown in Figure 9. In these visualizations, each segmented region identified as

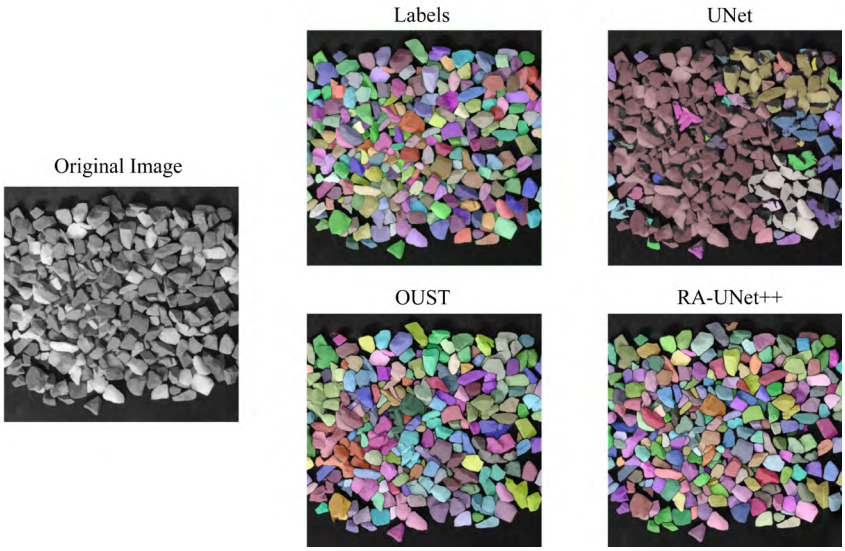

Fig. 9. Image overlay effect

Rys. 9. Efekt nakładania obrazu

an independent connected component is filled with a unique color and superimposed onto the original image. A distinctly colored particle thus represents a successfully identified, individual ore grain. This representation clearly demonstrates our method's ability to produce closed, complete contours for each ore particle.

The overlay maps in Figure 9 provide a clear visual assessment of instance segmentation performance. The map for Otsu's algorithm shows large, uniform color blocks, a result of its inability to form closed edges around most particles. This leads to the merging of numerous ores into a single connected domain, thus yielding a low count of unique colors. While the U-Net method identifies more particles than Otsu's, as shown by the increased number of colors, it still incorrectly merges many adherent ores. In contrast, our RA-UNet++ method successfully separates the vast majority of even closely packed particles into distinct, closed-contour instances. This is reflected in the significantly higher number of unique colors in its map, which corresponds to a more accurate count of individual ore grains and thus superior segmentation quality.

To move beyond qualitative assessment and to more rigorously evaluate the model's performance, particularly on challenging cases of adherent ores, we conducted a detailed quantitative analysis. Four representative ore images, designated as Image 1 through Image 4 and displayed in Figure 10, were selected for this purpose.

The segmentation performance on these images was measured using a suite of standard evaluation metrics: Accuracy (ACC), Mean Intersection over Union (MIoU), and F1-Score.

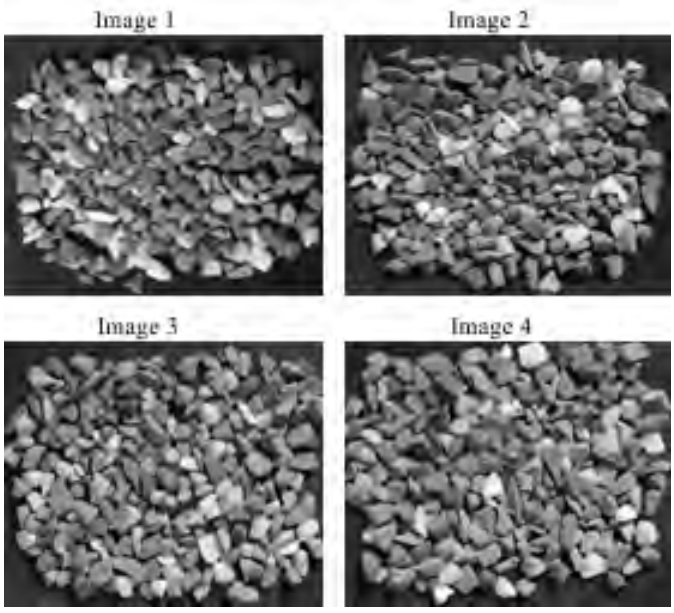

Fig. 10. Four different ore images

Rys. 10. Cztery różne zdjęcia rudy

In addition, we introduce a domain-specific metric, "Ore Adhesion Accuracy", to specifically assess the model's ability to separate touching particles. The comprehensive results for all evaluated methods are summarized in Table 2.

Table 2. Results of ACC, MIOU, F1 score, and ore segmentation accuracy evaluation metrics for the three algorithms

Tabela 2. Wyniki wskaźników oceny dokładności, takich jak ACC, MIOU, wskaźnik F1 oraz dokładność segmentacji rudy dla trzech algorytmów

Imagery	Algorithm	ACC (%)	MIOU (%)	F1 scores (%)	Ore segmentation accuracy (%)
Image 1	OTSU	81.03	67.69	83.27	6.14
	UNet	94.77	89.32	95.91	66.85
	RA-UNet++	96.13	92.04	96.94	93.13
Image 2	OTSU	77.94	63.30	80.81	9.15
	UNet	94.58	88.59	95.94	65.03
	RA-UNet++	95.41	90.31	96.53	92.85
Image 3	OTSU	78.22	62.35	82.89	10.37
	UNet	95.35	89.18	96.78	68.89
	RA-UNet++	95.99	90.61	97.21	94.30
Image 4	OTSU	76.33	59.25	81.86	6.46
	UNet	94.74	87.27	96.47	63.08
	RA-UNet++	95.50	89.05	96.97	93.63

The quantitative results presented in Table 2 provide a clear performance hierarchy among the evaluated methods.

Otsu's algorithm consistently yields the lowest scores across all metrics (ACC, MIoU, and F1-Score), with an average pixel accuracy below 90% and a particularly poor Ore Adhesion Accuracy under 15%. This confirms our qualitative findings and highlights its inadequacy for fine-grained segmentation tasks, as its global thresholding approach cannot handle the complex textures and color variations inherent in ore imagery.

The baseline U-Net model significantly outperforms Otsu's algorithm, demonstrating the power of deep learning for this task. However, its performance is still suboptimal. We attribute this to its relatively simple architecture, which may lead to insufficient feature representation, especially when fusing shallow, high-resolution features with deep, semantic features. This limitation manifests as an inability to resolve fine details, such as the boundaries of adherent particles.

In contrast, our proposed RA-UNet++ consistently achieves the highest scores across all evaluation metrics, demonstrating its superior segmentation capability. On average, our model surpasses 95% in both ACC and F1-Score and exceeds 90% in MIoU. This exceptional performance stems from several key architectural advantages. First, the ResNet backbone, with its residual connections, enables the training of a deeper network that captures more robust, hierarchical features without suffering from gradient vanishing. Second, the dense skip connections inherent in the UNet++ design facilitate more effective feature fusion across different semantic levels. Finally, the integrated self-attention modules allow the model to selectively focus on the most informative features and critical boundary regions. Collectively, these components empower our model to produce segmentation results that are closest to the truth label, validating its excellent performance and significant advantages for ore image segmentation.

Conclusion

In this paper, we addressed the challenging problem of accurately segmenting adherent ore images. To this end, we proposed a novel deep learning architecture, RA-UNet++, built upon the UNet++ framework. Our model integrates three key enhancements:

1) a ResNet backbone to enable deeper and more robust feature extraction;
2) a self-attention mechanism to focus on salient ore features and critical boundaries;
3) an Atrous Spatial Pyramid Pooling (ASPP) module to capture multi-scale context and improve global feature comprehension.

Extensive experiments demonstrate the superior performance of our proposed network. The RA-UNet++ model effectively segments individual ore particles, even in complex cases of adhesion, producing clear and complete boundaries. It achieved an average segmentation accuracy exceeding 93%, representing a significant improvement of 85.4% and 27.5% over the traditional Otsu's method and the baseline U-Net model, respectively.

These results validate the effectiveness of our approach and indicate its strong potential. The high-accuracy segmentation provided by RA-UNet++ can serve as a foundational component for automated ore particle size detection systems, thereby offering valuable technological support for the mining industry.

The Authors have no conflict of interest to declare.

REFERENCE

Badrinarayanan et al. 2017 – Badrinarayanan, V., Kendall, A. and Cipolla, R. 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 39(12), https://doi.org/10.1109/TPAMI.2016.2644615.

Chaurasia et al. 2025 – Chaurasia, A., Sharma, V. S., Srivastava, A., Abbas, Z. and Nagpal, R. 2025. *Insights into Brain Tumor Detection: Exploring U-Net Architecture*. [In:] *Mathematics and Logics in Computer Science*, pp. 173–183, https://doi.org/10.1007/978-981-96-3256-5_13.

Chen, H. 2025. Applications of deep learning in medical image segmentation: an evaluation of U-Net architecture and its enhancements. *International Conference on Physics, Photonics, and Optical Engineering (ICPPOE 2024)*, pp. 728–735, https://doi.org/10.1117/12.3060868.

Dai et al. 2017 – Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H. and Wei, Y. 2017. Deformable convolutional networks. *Proceedings of the IEEE international conference on computer vision*, pp. 764–773, https://doi.org/10.48550/arXiv.1703.06211.

Dash et al. 2017 – Dash, M., Londhe, N.D., Ghosh, S., Semwal, A. and Sonawane, R.S. 2019. PsLSNet: Automated psoriasis skin lesion segmentation using modified U-Net-based fully convolutional network. *Biomedical Signal Processing and Control* 52(12), pp. 226–237, https://doi.org/10.1016/j.bspc.2019.04.002.

He et al. 2016 – He, K., Zhang, X., Ren, S. and Sun, J. 2016. Identity mappings in deep residual networks. *Lecture Notes in Computer Science* 9908, pp. 630–645). https://doi.org/10.1007/978-3-319-46493-0_38.

Huang et al. 2022 – Huang, M., Liu, Y. and Yang, Y. 2022. Edge detection of ore and rock on the surface of explosion pile based on improved Canny operator. *Alexandria Engineering Journal* 61(12), pp. 10769–10777, https://doi.org/10.1016/j.aej.2022.04.019.

Jin et al. 2021 – Jin, F., Zhan, K., Chen, S., Huang, S. and Zhang, Y. 2021. Image segmentation method of mine pass soil and ore based on the fusion of the confidence edge detection algorithm and mean shift algorithm. *Gospodarka Surowcami Mineralnymi – Mineral Resources Management* 37(4), pp. 133–152, https://doi.org/10.24425/gsm.2021.139742.

Krizhevsky et al. 2012 – Krizhevsky, A., Sutskever, I. and Hinton, G.E. 2012. ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25.

Li et al. 2023 – Li, F., Liu, X., Yin, Y. and Li, Z. 2023. DDR-Unet: A high-accuracy and efficient ore image segmentation method. *IEEE Transactions on Instrumentation and Measurement* 99, https://doi.org/10.1109/TIM.2023.3317480.

Long et al. 2015 – Long, J., Shelhamer, E. and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. *Conference: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 3431–3440, https://doi.org/10.1109/CVPR.2015.7298965.

Mac et al. 2019 – Mac, K.N.C., Joshi, D., Yeh, R.A., Xiong, J., Feris, R.S. and Do, M.N. 2019. Learning motion in feature space: Locally-consistent deformable convolution networks for fine-grained action detection. *IEEE/CVF International Conference on Computer Vision*, pp. 6282–6291.

Mustafa et al. 2020 – Mustafa, N., Zhao, J., Liu, Z., Zhang, Z. and Yu, W. 2020. Iron ORE region segmentation using high-resolution remote sensing images based on Res-U-Net. *IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2563–2566, https://doi.org/10.1109/IGARSS39084.2020.9324218.

Naylor et al. 2018 – Naylor, P., Laé, M., Reyal, F. and Walter, T. 2018. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE transactions on medical imaging* 38(2), pp. 448–459, https://doi.org/10.1109/TMI.2018.2865709.

Oktay et al. 2018 – Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M.P., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B. and Rueckert, D. 2018. *Attention u-net: Learning where to look for the pancreas*, https://doi.org/10.48550/arXiv.1804.03999.

Ronneberger et al. 2015 – Ronneberger, O., Fischer, P. and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. *Conference: International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.

Salinas et al. 2005 – Salinas, R.A., Raff, U. and Farfan, C. 2005. Automated estimation of rock fragment distributions using computer vision and its application in mining. *IEE Proceedings-Vision, Image and Signal Processing* 152(1), https://doi.org/10.1049/ip-vis:20050810.

Shaw et al. 2018 – Shaw, P., Uszkoreit, J. and Vaswani, A. 2018. *Self-attention with relative position representations*, https://doi.org/10.48550/arXiv.1803.02155.

Shrivastava et al. 2023 – Shrivastava, S., Bhattacharjee, S. and Deb, D. 2023. Segmentation of mine overburden dump particles from images using Mask R CNN. *Scientific Reports* 13(1), https://doi.org/10.1038/s41598-023-28586-0.

Song et al. 2019 – Song, W., Zheng, N., Liu, X., Qiu, L. and Zheng, R. 2019. An improved u-net convolutional networks for seabed mineral image segmentation. *IEEE Access* 7, pp. 82744–82752, https://doi.org/10.1109/ACCESS.2019.2923753.

Sun et al. 2022 – Sun, G., Huang, D., Cheng, L., Jia, J., Xiong, C. and Zhang, Y. 2022. Efficient and lightweight framework for real-time ore image segmentation based on deep learning. *Minerals* 12(5), https://doi.org/10.3390/min12050526.

Tavakolian et al. 2025 – Tavakolian, N., Nazemi, A. and Suen, C.Y. 2025. Review of the AI-Based Analysis of Abdominal Organs from Routine CT Scans. *Applied Sciences* 15(5), https://doi.org/10.3390/app15052516.

Touvron et al. 2021 – Touvron, H., Cord, M., Sablayrolles, A., Synnaeve, G. and Jégou, H. 2021. Going deeper with image transformers. *Conference: 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 32–42, https://doi.org/10.1109/ICCV48922.2021.00010.

Wang et al. 2023 – Wang, W., Li, Q., Zhang, D. and Fu, J. 2023. Image segmentation of adhesive ores based on MSBA-Unet and convex-hull defect detection. *Engineering Applications of Artificial Intelligence* 123, https://doi.org/10.1016/j.engappai.2023.106185.

Xiao et al. 2020 – Xiao, D., Liu, X., Le, B. T., Ji, Z. and Sun, X. 2020. An ore image segmentation method based on RDU-Net model. *Sensors* 20(17), https://doi.org/10.3390/s20174979.

Yuan, L. and Duan, Y. 2018. *A method of ore image segmentation based on deep learning*. [In:] *Intelligent Computing Methodologies*, pp. 508–519, https://doi.org/10.1007/978-3-319-95957-3_53.

Yurtkulu et al. 2019 – Yurtkulu, S.C., Şahin, Y.H. and Unal, G. 2019. Semantic segmentation with extended DeepLabv3 architecture. *2019 27th signal processing and communications applications conference (SIU)*, https://doi.org/10.1109/SIU.2019.8806244.

Zhang et al. 2017 – Zhang, G., Li, M., Zhan, Y. and Shi, X. 2017. Ore image thresholding segmentation using double windows with fisher discrimination. *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pp. 2715-2719, https://doi.org/10.1109/FSKD.2017.8393208.

Zhou et al. 2019 – Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N. and Liang, J. 2019. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging* 39(6), pp. 1856–1867, https://doi.org/10.48550/arXiv.1912.05074.

RA-UNet++ BASED IMAGE SEGMENTATION OF ADHERENT ORES

Keywords

ore image segmentation, UNet++, residual module,
self-attention mechanism, Atrous Spatial Pyramid Pooling

Abstract

Ore particle size information is an important indicator for evaluating the operating status and production efficiency of crushers. However, in the actual industrial environment, adhesion phenomena often occur during the acquisition and transportation of ores, resulting in overlapping edges and blurred contours of ores in the images. Traditional image segmentation methods are difficult to achieve high-precision recognition and segmentation. To this end, this paper proposes an RA-UNet++ adhered ore image segmentation method based on the improved UNet++ structure to improve the segmentation performance in complex scenes. Based on the UNet++ coender-decoding architecture, the residual module and the self-attention mechanism are integrated to enhance the model's ability to extract and express the edge details of ores. Meanwhile, multi-scale atrous convolution is introduced

at the end of the encoder to construct the Atrous Spatial Pyramid Pooling (ASPP) structure, expand the receptive field, enhance the multi-scale perception ability for ores of different particle sizes, and thereby improve the overall segmentation effect. The experimental results show that RA-UNet++ performs excellently in the task of image segmentation of adhered ores, significantly improving the clarity and integrity of edge segmentation. Compared with the Otsu method and the standard UNet model, this method has advantages in terms of robustness, boundary preservation, and segmentation accuracy. The pixel-level and target-level segmentation accuracy rates of adhered ore images both exceed 93%, showing good potential for industrial applications.

SEGMENTACJA OBRAZÓW RUD PRZYLEGAJĄCYCH Z WYKORZYSTANIEM SIECI RA-UNet++

Słowa kluczowe

segmentacja obrazu, UNet++, moduł resztkowy,
mechanizm samo-uwagi, Atrous Spatial Pyramid Pooling

Streszczenie

Informacje dotyczące wielkości cząstek rudy stanowią ważny wskaźnik służący do oceny stanu pracy i wydajności produkcyjnej kruszarek. Jednak w rzeczywistych warunkach przemysłowych podczas pozyskiwania i transportu rud często dochodzi do zjawisk zlepiania się cząstek, co powoduje nakładanie się krawędzi i rozmycie konturów rudy na obrazach. Tradycyjne metody segmentacji obrazów mają trudności z osiągnięciem wysokiej precyzji rozpoznawania i segmentacji. W związku z tym w niniejszym artykule proponuje się metodę segmentacji obrazów rudy ze zjawiskiem przywierania RA-UNet++, opartą na ulepszonej strukturze UNet++, mającą na celu poprawę wydajności segmentacji w złożonych scenach. W oparciu o architekturę kodowania i dekodowania UNet++ zintegrowano moduł resztkowy oraz mechanizm samo-uwagi, aby wzmocnić zdolność modelu do wyodrębniania i wyrażania szczegółów krawędzi rudy. Jednocześnie na końcu kodera wprowadzono wieloskalową konwolucję atrous, aby zbudować strukturę Atrous Spatial Pyramid Pooling (ASPP), rozszerzyć pole receptywne, zwiększyć zdolność postrzegania w wielu skalach dla rud o różnych rozmiarach cząstek, a tym samym poprawić ogólny efekt segmentacji. Wyniki eksperymentalne pokazują, że RA-UNet++ doskonale radzi sobie z zadaniem segmentacji obrazów rud przylegających, znacznie poprawiając klarowność i integralność segmentacji krawędzi. W porównaniu z metodą Otsu i standardowym modelem UNet metoda ta ma przewagę pod względem odporności, zachowania granic i dokładności segmentacji. Wskaźniki dokładności segmentacji obrazów rudy przylegającej, zarówno na poziomie pikseli, jak i na poziomie obiektów docelowych, przekraczają 93%, co wskazuje na duży potencjał w zastosowaniach przemysłowych.