minerals

MDPI

Article

Res-UNet Ensemble Learning for Semantic Segmentation of Mineral Optical Microscopy Images

Chong Jiang [1,2], Alfian Abdul Halin [2,*], Baohua Yang [1,*], Lili Nurliyana Abdullah [2], Noridayu Manshor [2] and Thinagaran Perumal [2]

[1] School of Information Science and Engineering, HuNan Woman's University, Changsha 410004, China; jiangchong@hnwu.edu.cn
[2] Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Kuala Lumpur 43400, Malaysia; liyana@upm.edu.my (L.N.A.); ayu@upm.edu.my (N.M.); thinagaran@upm.edu.my (T.P.)
* Correspondence: alfian@upm.edu.my (A.A.H.); yangbaohua@hnwu.edu.cn (B.Y.)

Abstract: In geology and mineralogy, optical microscopic images have become a primary research focus for intelligent mineral recognition due to their low equipment cost, ease of use, and distinct mineral characteristics in imaging. However, due to their close reflectivity or transparency, some minerals are not easily distinguished from other minerals or background. Secondly, the number of background pixels often vastly exceeds the number of pixels for individual mineral particles, and the number of pixels of different mineral particles in the image also varies significantly. These have led to the issue of data imbalance. This imbalance results in lower recognition accuracy for categories with fewer samples. To address these issues, a flexible ensemble learning for semantic segmentation based on multiple optimized Res-UNet models is proposed, introducing dice loss and focal loss functions and incorporating a pre-positioned spatial transformer networks block. Twelve optimized Res-UNet models were used to construct multiple Res-UNet ensemble learnings using heterogeneous ensemble strategies. The results demonstrate that the system integrated with five learners using the weighted voting fusion method (RUEL-5-WV) achieved the best performance with a mean Intersection over Union (mIOU) of 91.65 across all nine categories and an IOU of 84.33 for the transparent mineral (gangue). The results indicate that this ensemble learning scheme outperforms individual optimized Res-UNet models. Compared to the classical Deeplabv3 and PSPNet, this scheme also exhibits significant advantages.

Keywords: optical microscopy images; deep learning; ensemble learning; semantic segmentation; mineralogy

Citation: Jiang, C.; Abdul Halin, A.; Yang, B.; Abdullah, L.N.; Manshor, N.; Perumal, T. Res-UNet Ensemble Learning for Semantic Segmentation of Mineral Optical Microscopy Images. *Minerals* **2024**, *14*, 1281. https://doi.org/10.3390/min14121281

Academic Editor: Sebastian Iwaszenko

Received: 13 November 2024
Revised: 10 December 2024
Accepted: 11 December 2024
Published: 17 December 2024

1. Introduction

In geology, phase analysis and mineral characterization are essential for dating stratigraphy and identifying significant minerals. In mineral processing, accurate mineral identification facilitates the efficient extraction of target minerals and reduces operational costs. However, traditional mineral identification methods heavily depend on experts, resulting in low efficiency, high error rates, and significant subjectivity [1–3]. In recent years, numerous scholars have investigated automatic mineral recognition using various types of mineral images. Due to its lower cost and ease of use compared to other devices such as electron scanning microscopes, CT scanners, etc., optical microscopes have been widely used in the study of minerals and geological materials [4–6]. Some minerals with similar chemical compositions that are difficult to distinguish using scanning electron microscopes can be identified accurately in optical microscopy images [1,7].

Computer vision-based devices and methods have emerged as mainstream technologies for mineral identification and characterization. This approach, known as automatic mineral identification (AMI), encompasses three primary steps [5,8]:

1. Feature extraction: Engineers can extract features based on their expertise and datasets, including color, shape, texture, granularity and others. Alternatively, features can be algorithmically derived from the original image information to generate a lower-dimensional feature map.
2. Pixel clustering or image segmentation: Following feature extraction, similar pixels within each image are clustered, visually segmenting the image into multiple non-overlapping regions. Each region potentially corresponds to a specific mineral or background, which will be identified in the subsequent step.
3. Classification: A specific algorithm assigns each segmented region in the image to a designated mineral or background category.

1.1. Related Studies

Early automatic mineral identification from optical microscopy images predominantly concentrated on image segmentation, necessitating manual intervention by experts for feature extraction and classification [8]. Initially, traditional image processing algorithms were employed for image segmentation, leveraging low-level visual information such as grayscale values, color, texture, and edges. These traditional algorithms included threshold-based algorithms [1,9–11], edge detection methods, region-based techniques and others [12–20].

With advancements in imaging technologies and computer science, machine learning-based computer vision techniques have been developed and applied to mineral identification, automating both image segmentation and the classification of segmented regions in AMI. These machine learning algorithms utilize image characteristics like color, texture, shape, and others for image segmentation [21–23], incorporating methods such as decision trees, random forests, naive Bayes, K-nearest neighbors, artificial neural networks, support vector machines, etc.

The study of various machine learning algorithms is common, and the integration of traditional methods with machine learning was also used. Ref. [24] proposed a maceral identification strategy based on image segmentation and classification, using K-means clustering to divide the image into regions with similar properties. Comprehensive features and random forest were then used to classify the binder and seven types of maceral components, achieving an accuracy of 90.44%. Ref. [25] developed a random forest-based model to classify different phases of coal macerals and minerals. The random forest classifier segmented macerals while ignoring the background, utilizing features related to microstructure for classification, resulting in an overall classification accuracy of over 90%. Ref. [26] proposed a complex method combining image processing and machine learning algorithms to analyze petrographic thin sections, integrating structural object segmentation and rock classification for images obtained in both non-polarized and polarized light. This method achieved approximately 90% accuracy in tests, providing results that included grain size class, rock type, and mineral composition. However, there are also some evident limitations in machine learning algorithms. Feature engineering before algorithm implementation generally requires manual intervention, which demands high levels of expertise and experience [3]. Additionally, steps like feature selection and dimensionality reduction increase the complexity of the model [27]. For pixel-level classification tasks in image segmentation, machine learning methods typically consider individual pixels rather than pixel regions, leading to suboptimal segmentation performance. Furthermore, bottlenecks exist in processing speed and operational parallelism [28].

Over the past decade, the field of deep learning, a pivotal branch of machine learning, has experienced remarkable growth, especially in Convolutional Neural Networks (CNNs)-based semantic segmentation models. Some scholars have conducted studies on the automated identification of transparent and opaque minerals in polished sections. Ref. [29] applied an enhanced UNet model for semantic segmentation on polished sections of 10 mineral types, attaining an average IoU of 0.813 and an average accuracy of 0.892. Ref. [5] presented a method for detecting mineral particles in reflected light microscope

images using a UNet variant. The results suggest this method detects all mineral particles in samples, although there were 34% under-modeled and 4.73% over-modeled areas. Ref. [4] utilized a U-Net-like structure to develop a deep learning model for the semantic segmentation of optical microscopic images of geological polished sections. Comparative experiments demonstrated that using both cross-polarized light (XPL) and plane-polarized light (PPL) images significantly improved segmentation results. Ref. [7] utilized an enhanced Deeplabv3 model for segmenting opaque and non-opaque minerals in reflected light microscopy images, achieving accuracy and F1 scores exceeding 90%. Ref. [30] introduced a supervised semantic segmentation model for the pixel-level classification of 2D RGB images of sandstones acquired via transmission light microscopy, distinguishing between pores and various minerals. Experimental comparisons revealed that the Deeplab V3+ Resnet-18 network produced optimal results. In mineralogy and geology, fully automated tasks in feature extraction, image segmentation, and classification have been realized for automatic mineral identification by deep learning.

1.2. Issues

Despite advancements in automatic mineral identification from optical microscope images, there are still some unresolved issues.

1.2.1. Transparent and Optically-Alike Minerals Identification

Certain minerals, due to their reflectance or transparency, exhibit colors similar to other minerals or the background, leading to lower accuracy rates in automated identification. This is evident with minerals such as sphalerite and magnetite as well as in differentiating transparent mineral from the background [5,7,31]. Similar challenges have been observed in distinguishing between pyrite and marcasite, or pyrite and arsenopyrite [1,4,32,33]. The limitations can lead to challenges in accurately quantifying mineral grades and may introduce biases in the analysis of the texture of the studied mineral particles, including assessments of mineral liberation degree.

1.2.2. Data Imbalance

The number of pixels in different categories, including background and mineral particles, may vary significantly in polished section images. It results in imbalanced sample data, directly lowering accuracy for categories with fewer samples and affecting the overall accuracy [4,34–36]. Data imbalance can lead to lower accuracy for classes with a limited number of pixels, resulting in unclear or inaccurate boundary segmentation. It may also cause significant fluctuations in the loss function, making it challenging for the model to converge to an optimal state.

1.3. Solution to the Issues

1.3.1. Feature Extracting

Improving the learning ability of the model and obtaining more effective semantic features can enhance the segmentation accuracy of transparent particles and similar regions. Introducing LeakyRelu, residual networks and spatial transformer networks (STNs) into the encoder are the possible approaches. The residual connections in ResNet architectures address the vanishing gradient problem commonly encountered during training [37–39]. Although ReLU is simple and fast, it may encounter the "neuron death" problem, where some neurons never activate during training, preventing their weights from being updated [40,41]. LeakyReLU mitigates the "neuron death" issue by not outputting zero for negative values through a simple function [42,43]. The capacity of CNNs for spatial transformation adaptation remains restricted. Spatial transformer networks (STNs) are trainable modules that execute spatial manipulations on data. STN empower a model to dynamically implement spatial transformations, such as translation, deformation, rotation, and scaling, which are contingent upon the features themselves [44]. Incorporating STN

into CNNs enhances performance on datasets exhibiting spatial transformations of the foreground and other positive samples [45].

1.3.2. Loss Function

Dice loss, focal loss, or their combinations with cross-entropy loss (CE Loss) are important techniques for addressing the problem of data imbalance. CE Loss is a widely utilized loss function in classification tasks, quantifying the cross-entropy between the predicted probability distribution and the true label distribution. As a widely used loss function in image segmentation, CE Loss tends to perform poorly in the presence of class imbalance [46,47]. The Dice coefficient is an overlap measure extensively utilized in image segmentation. Dice loss, derived from the Dice coefficient, effectively addresses the issue of imbalanced quantities of foreground and background pixels [47,48]. Focal loss is specifically designed to address class imbalance in one-stage object detection. As a modified version of the cross-entropy loss function, it reduces the weight of easy-to-distinguish examples, thereby directing the model's focus toward more challenging examples during training [47,49]. In natural language processing (NLP) tasks [50,51], weather forecasting [52], and medical image segmentation [53–55], some scholars have explored the role of the dice loss and focal loss function in handling data imbalance, and they have demonstrated a significant improvement in model performance.

1.3.3. Ensemble Learning

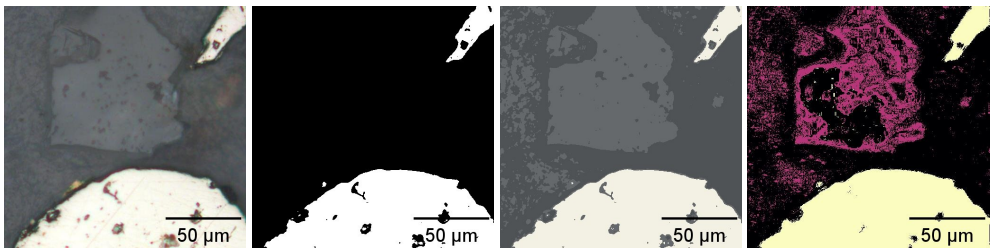
Ensemble learning denotes a methodology that amalgamates multiple weak learners to create a robust learner, thereby achieving a specified task [56]. In general, ensemble classifiers exhibit greater robustness and superior performance compared to individual models [57]. With the advancement of deep learning techniques, the base learners in ensemble learning have gradually evolved from traditional classifiers to deep learning models. Although ensemble learning based on deep learning models presents greater challenges compared to traditional classifiers, it demonstrates superior performance in various fields, such as computer vision, natural language processing (NLP), and others [58,59]. From some studies on medical image recognition, it is evident that ensemble learning based on deep learning models demonstrates superior performance compared to individual semantic segmentation models [60–67].

1.4. Preliminary Work

Comparative experiments were conducted in the preliminary work of this study, utilizing Otsu's thresholding, k-means clustering, and Random Forest algorithms for image segmentation. Figure 1 illustrates the results. As shown in Figure 1b, the transparent gangue was not detected using Otsu's thresholding due to its similar intensity to the background. In the result of the k-means algorithm (Figure 1c), some pixels of the left side of the image were incorrectly classified as background. In addition, the key parameter k, presenting the number of categories, must be predetermined before executing the k-means algorithm, which increases the algorithm's complexity. Similar to the results of the k-means algorithm, the Random Forest algorithm also classified some pixels on the left side of the image as the gangue category, as shown in Figure 1d. Moreover, for the segmented gangue particle, only the clear boundary is visible, while the majority of the pixels is identified as the background. In pixel-level classification tasks, such as image segmentation, it is essential not only to determine the category of individual pixels but also to consider the relationships between adjacent pixels and regions. The traditional algorithms have only achieved the former, which is the primary reason for their suboptimal results.

To address the issues, a flexible Res-UNet ensemble learning for semantic segmentation based on multiple optimized Res-UNet models is proposed. The system achieved better results than traditional models, and it also exhibited certain advantages compared to other studies in the same field. The structure of this paper is as follows: Section 2 covers the sampling, data acquisition and augmentation, Res-UNet and optimization, and the

workflow of the proposed ensemble learning system. Section 3 details image annotation, model training and evaluation metrics. Section 4 presents the Results and Discussion. The final section offers the conclusions.



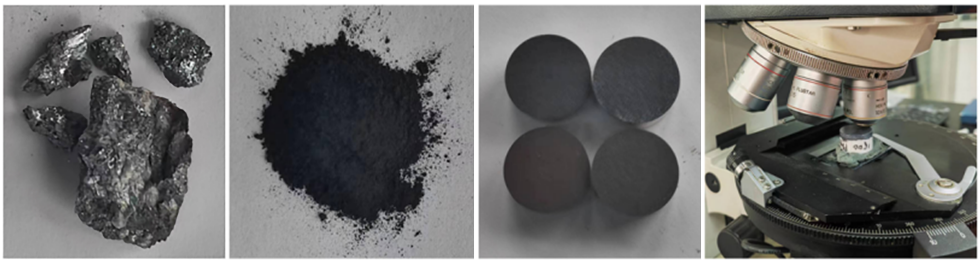
(**a**) (**b**) (**c**) (**d**)

Figure 1. Comparison between results and original image. (**a**) Original image; (**b**) the result of Otsu's threshold; (**c**) the result of k-means; (**d**) the result of Random Forest.

2. Dataset and Methodology

2.1. Mineral Sampling

In this experiment, four types of ore samples were selected, primarily composed of pyrite, galena, sphalerite, and magnetite, with minor components including pyrrhotite, bornite, chalcopyrite, and gangue. The pyrite and magnetite samples originated from the Inner Mongolia Autonomous Region, China, while the galena and sphalerite samples were sourced from Hunan Province, China. Each ore type was prepared at the Hunan Nonferrous Metals Research Institute. Sample preparation involved grinding the raw ore (Figure 2a) into powder (Figure 2b), then mixing with resin, pressing, and polishing to create the polished sections depicted in Figure 2c. The mineral powder was sieved to a mesh size of 200. These samples were subsequently examined under a polarized light microscope with imaging parameters configured to capture plane-polarized light images of the mineral particles, as shown in Figure 2d.



(**a**) (**b**) (**c**) (**d**)

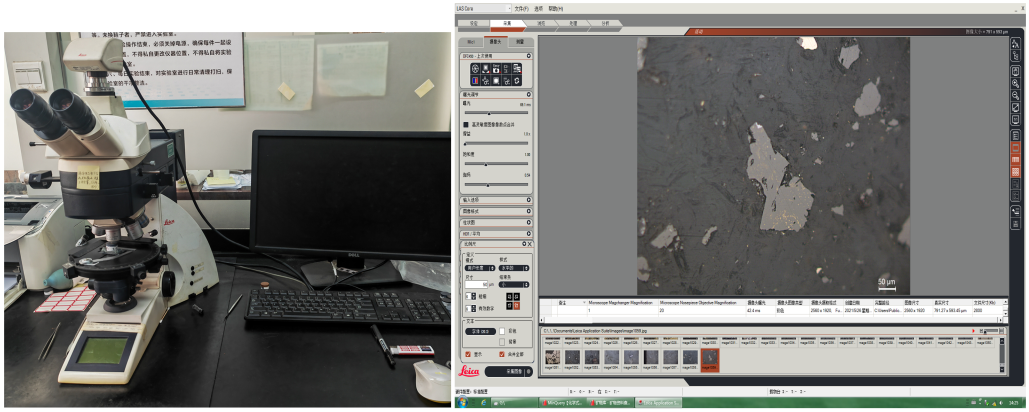
Figure 2. Sampling and image capturing. (**a**) Original ore; (**b**) ore powder; (**c**) polished sections; (**d**) the microscope for images capturing.

2.2. Data Acquisition and Augmentation

Original images were captured by polarized light microscope using the PPL mode. The optical microscope used in this study is an upright Leica DM4500P polarization model equipped with a Leica DFC450 camera. This digital microscope camera, featuring a C-mount interface, houses a high-quality 5-megapixel CCD sensor. The microscope and imaging equipment are depicted in Figure 3a. The integrated software system allows for the real-time transmission of captured images to a computer, as shown in the software interface in Figure 3b. The imaging parameters were set as follows: exposure time of 69.1 ms, gain control at 1.0×, saturation at 1.00, and gamma at 0.54. The image resolution was 2560 × 1920.

A total of 98 original images were cropped to 512 × 512 size. After screening, 1377 images were selected as the experimental dataset. Considering the quantity and the ratio of various mineral categories, 1146 images were designated for the training set

and 231 for the test set. Data augmentation was applied, including flips and rotations. The number of images in the training and test set reached 6876 and 1386 after augmentation.



(**a**) (**b**)

Figure 3. Software and hardware equipment for image acquisition. (**a**) Optical microscope and computer; (**b**) image acquisition interface.

2.3. Res-UNet and Optimization

2.3.1. Res-UNet

Res-UNet, a variant of the UNet architecture, is widely utilized in medical image segmentation, demonstrating remarkable performance [68–73]. The optimized Res-UNet serves as the learner in Res-UNet ensemble learning with the overall structure depicted in Figure 4. Layer1, Layer2, Layer3, and Layer4 each consist of one Downsampling Bottleneck (Figure 4b) and several Normal Bottlenecks (Figure 4c). The number of obttlenecks and the values of n in each layer are detailed in Table 1. The optimization of ResNet50 as the encoder of Res-UNet focuses on two main aspects: the activation function, Leakyrelu, and the incorporation of STN.

Table 1. Number of bottlenecks in Layers 1–4 and the values of n.

	Layer 1	**Layer 2**	**Layer 3**	**Layer 4**
Downsampling bottleneck	1	1	1	1
Normal bottleneck	2	3	5	2
n	1	2	3	4

2.3.2. Activation Function

According to comparative experiments, LeakyReLU(0.1) was selected in this study with its formula shown in Equation (1).

$$f(x) = \begin{cases} x, x > 0 \\ 0.1x, x \leq 0 \end{cases} \quad (1)$$

2.3.3. Cross-Entropy Loss

The CE Loss is calculated as shown in Equation (2).

$$L(y, p) = -\sum_{i=1}^{n} y_i \log(p_i) \quad (2)$$

Here, y_i is the true label and p_i is the predicted value.

2.3.4. Dice Loss

The formula for calculating dice loss is presented in Equation (3).

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{n} p_i y_i}{\sum_{i=1}^{n} p_i^2 + \sum_{i=1}^{n} y_i^2} \tag{3}$$

Here, y_i represents the actual pixel value and p_i denotes the predicted pixel value. For multi-class segmentation, the second term averages over all classes.

2.3.5. Focal Loss

The calculation formula of focal loss is provided in Equation (4).

$$\mathrm{FL}(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \tag{4}$$

For easy-to-distinguish samples, $(1 - p_t)^{\gamma}$ approaches 0, thereby reducing the loss value. Conversely, for hard-to-distinguish samples, $(1 - p_t)^{\gamma}$ approaches 1, maintaining the loss value. This mechanism increases the contribution of hard samples to the overall loss by reducing the loss from easy samples. The term α_t refers to an α-balanced form of focal loss, typically ranging between 0 and 1. The variable p_t denotes the predicted probability for a specific pixel.

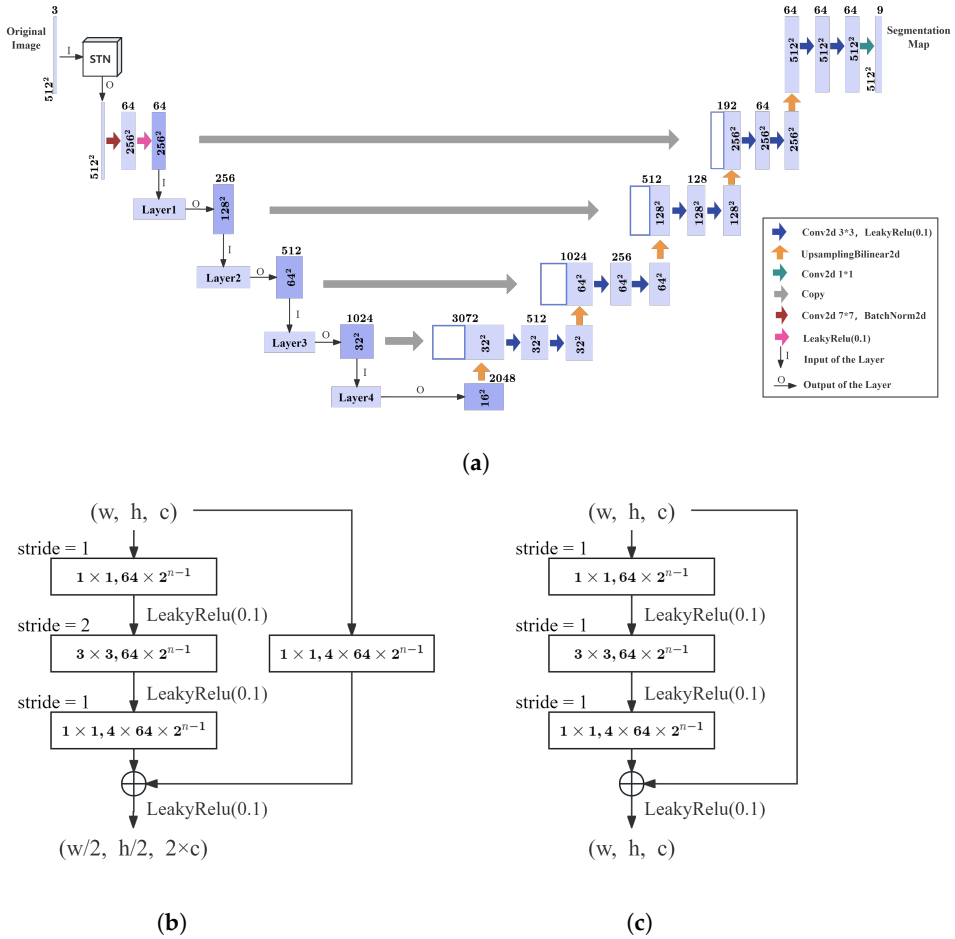


Figure 4. Improved Res-UNet for 512 × 512 Size Image. (**a**) Overall structure; (**b**) downsampling bottleneck; (**c**) normal bottleneck.

2.3.6. Spatial Transformer Networks (STNs)

Mineral particles within the same category can exhibit substantial variations in size and shape. To enhance the generalization performance of the learner, an STN is integrated

with the Res-UNet model. The STN represents an adaptive mechanism capable of performing spatial transformations on images or feature maps by generating customized transformations for individual input samples images and improving feature extraction efficiency [74]. Figure 5 illustrates the STN architecture, which is composed of a localization net, grid generator and sampler.

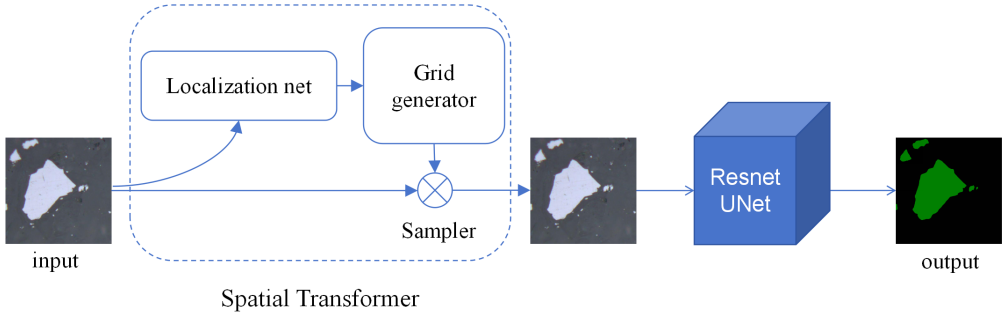


Figure 5. The structure and workflow of STN in the Res-UNet.

2.4. Res-UNet Ensemble Learning

The Res-UNet Ensemble Learning (RUEL) scheme incorporating multiple learners was proposed, offering the advantage of avoiding an increase in hyper-parameters and minimizing the complexity of weak learners. The architecture utilizes a heterogeneous ensemble strategy, integrating multiple learners based on the same dataset. The fusion method applied is the voting method, specifically plurality voting and weighted voting. All learners are arranged in descending order based on a selected evaluation metric with first the n learners selected to ensemble. Each ensemble assigns two sets of weights to the base models: one with equal weights (plurality voting) and another with decreasing weights (weighted voting) expressed by Equation (5).

$$w_i = \frac{n-i}{\sum_{j=1}^{n} n} \tag{5}$$

In the equation, $i \in [0, n-1]$.

The workflow of RUEL can be outlined as follows:

1. Identify n trained and ranked learners along with their respective weights, ensuring the sum of all weights equals 1. For a given pixel, each learner produces a prediction result $pred_i$, representing different mineral categories.
2. Define set S to gather all unique prediction values and their corresponding weighted sums. S is a collection of key–value pairs, where each pair represents a unique prediction value for the pixel and its weighted sum. The formula utilized for this set S is illustrated in Equation (6):

$$S = \left\{ \left(pred_i, \sum_{j=1}^{n} w_j | pred_j = pred_i \right) \right\} \tag{6}$$

3. Select the prediction value with the maximum weighted sum from S as the final output y, which denotes the final predicted class for the pixel by RUEL. The calculation formula is illustrated in Equation (7).

$$y = \arg \max_{(pred, \sum w)} \left(\sum w | (pred, \sum w) \in S \right) \tag{7}$$

The workflow of RUEL is depicted in Figure 6.

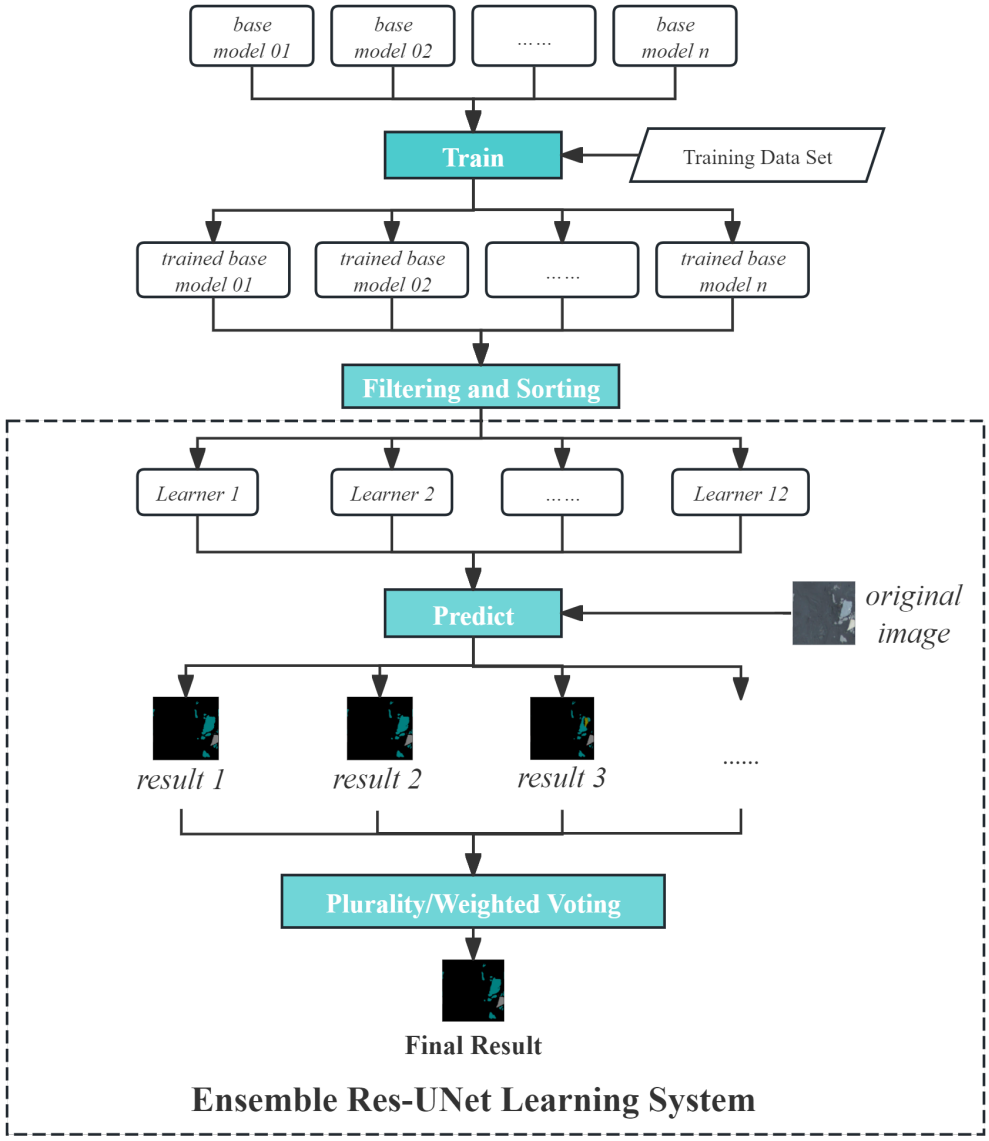


Figure 6. Workflow of the Res-UNet Ensemble Learning (RUEL).

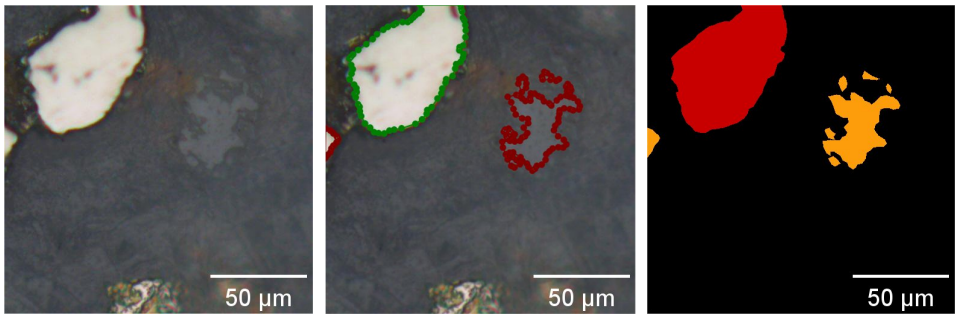
3. Experiments

3.1. Hardware and Software

This research is conducted on a workstation equipped with two Intel® Xeon® Silver 4210 2.2 GHz processors, 128 GB of memory, and two NVIDIA GeForce RTX 3090 GPUs. The Integrated Development Environment (IDE) utilized is Visual Studio Code (version: 1.75.1), configured with a Python+PyTorch environment, incorporating Python 3.9.13, torch 1.7.1+cu101, numpy 1.12.5, and matplotlib 3.5.1.

3.2. Image Annotation

The images in the training and testing sets were pre-annotated to create corresponding masks for the original images. Annotation was performed using Labelme 5.1.1. Initially, the mineral particles were labeled as different categories in Labelme (Figure 7b) [75,76], and the corresponding masks were subsequently generated (Figure 7c), serving as ground truth labels during training and testing. The original images and their corresponding labels in the training dataset are 512 × 512 in size with the original images being JPG format RGB images (Figure 7a) and the labels being single-channel PNG images (Figure 7c).



(**a**) (**b**) (**c**)

Figure 7. Three stages of image annotation. (**a**) Original sample image; (**b**) Two different classes annotated with different colors; (**c**) Ground truth for each class represented by different colors.

3.3. Model Training

Transfer learning was employed in model training. Initially, pre-trained weights from ImageNet were loaded into the encoder, while the weights of the decoder were initialized randomly. The training process was divided into two stages: the first stage involved freezing the backbone and training the decoder, while the second stage entailed unfreezing the backbone and training the entire network. This experiment implemented two sets of training epochs with frozen and unfrozen settings, specifically 40 + 80 and 50 + 100 epochs. The initial learning rate was set to 1×10^{-4} and decayed to 1×10^{-6} using a cosine annealing schedule. The optimizer was ADAM with a momentum of 0.9. Nearly 30 models were trained, incorporating different loss functions, activations, STN, epochs, and batch sizes. Based on the training and prediction results, 12 learners were selected to construct Res-UNet ensemble learning. The specific settings of each model are detailed in Table 2.

Table 2. Settings of all models.

ID	Backbone	Loss Function	Activation	STN	Epoches	Batchsize
b0	VGG	CE Loss	Relu()	0	40 + 80	32/16
b1	VGG	CE Loss	Relu()	0	50 + 100	32/16
1	ResNet	CE Loss + Dice Loss	LeakyRelu(0.1)	1	40 + 80	32/16
2	ResNet	CE Loss + Focal Loss	LeakyRelu(0.1)	1	40 + 80	32/16
3	ResNet	CE Loss	LeakyRelu(0.1)	1	40 + 80	32/16
4	ResNet	CE Loss + Dice Loss	LeakyRelu(0.1)	0	40 + 80	32/16
5	ResNet	CE Loss + Focal Loss	LeakyRelu(0.1)	0	40 + 80	32/16
6	ResNet	CE Loss	LeakyRelu(0.1)	0	40 + 80	32/16
7	ResNet	CE Loss + Dice Loss	LeakyRelu(0.1)	1	50+100	32/16
8	ResNet	CE Loss + Focal Loss	LeakyRelu(0.1)	1	50+100	32/16
9	ResNet	CE Loss	LeakyRelu(0.1)	1	50+100	32/16
10	ResNet	CE Loss + Dice Loss	LeakyRelu(0.1)	0	50+100	32/16
11	ResNet	CE Loss + Focal Loss	LeakyRelu(0.1)	0	50+100	32/16
12	ResNet	CE Loss	LeakyRelu(0.1)	0	50+100	32/16

3.4. Evaluation Metrics

3.4.1. Confusion Matrix

Semantic segmentation entails classifying each pixel in an image into categories, typically encompassing the background and multiple foregrounds. All evaluation metrics used are derived from the confusion matrix. It is an $n * n$ square matrix, where n denotes the number of categories. In the matrix, each column indicates the number of pixels predicted by the model as a certain category, while each row reflects the true number of pixels in that category. Specifically, in Table 3, num_{ij} represents the number of pixels that belong to the true category $class_i$ and are predicted as $class_j$.

Table 3. Confusion matrix for multi-class classification.

Confusion Matrix		Predicted Values				
		$\boldsymbol{class_0}$	.....	$\boldsymbol{class_j}$	.....	$\boldsymbol{class_{n-1}}$
True Values	$class_0$	num_{00}	.....	num_{0j}	.....	$num_{0(n-1)}$
	.....	.....	.....	.....	.....	.....
	$class_i$	num_{i0}	.....	num_{ij}	.....	$num_{i(n-1)}$
	.....	.....	.....	.....	.....	.....
	$class_{n-1}$	$num_{(n-1)0}$	.....	$num_{(n-1)j}$	.....	$num_{(n-1)(n-1)}$

3.4.2. Mean Pixel Accuracy

Mean Pixel Accuracy (mPA) indicates the average proportion of correctly predicted pixels for each category relative to the total number of pixels predicted for that category. Essentially, it quantifies the accuracy of pixel predictions for each category. The formula for mPA is presented in Equation (8):

$$\text{mPA} = \frac{1}{n}\sum_{i=0}^{n-1}\frac{num_{ii}}{\sum_{j=0}^{n-1} num_{ji}} \tag{8}$$

3.4.3. Mean Recall

Mean Recall (mRecall) represents the average proportion of correctly predicted pixels for each category relative to the total number of actual pixels in that category. This metric indicates the proportion of correctly identified pixels for each category. The calculation formula is presented in Equation (9):

$$\text{mRecall} = \frac{1}{n}\sum_{i=0}^{n-1}\frac{num_{ii}}{\sum_{j=0}^{n-1} num_{ij}} \tag{9}$$

3.4.4. F1 Score

In semantic segmentation, the F1 score integrates both mPA and Recall metrics with equal weighting. A value closer to 1 indicates that the predicted results closely match the ground truth. The formula for the F1 score is presented in Equation (10):

$$F1score = 2 \times \frac{mPA \times m\,\text{Re}\,call}{mPA + m\,\text{Re}\,call} \tag{10}$$

3.4.5. Mean Intersection over Union

Mean Intersection over Union (mIoU) is a standard metric in semantic segmentation, specifically measuring the proportion of the intersection to the union of the ground truth and predicted values. A higher IoU value indicates greater similarity between predicted and actual values. mIoU represents the average IoU across various categories in image segmentation. The calculation formula for mIoU is presented in Equation (11):

$$mIoU = \frac{1}{n}\sum_{i=0}^{n-1}\frac{num_{ii}}{\sum_{j=0}^{n-1} num_{ij} + \sum_{j=0}^{n-1} num_{ji} + num_{ii}} \tag{11}$$

4. Result and Discussion

4.1. Training Results

The 14 models, including 12 learners and 2 baselines, were categorized into two groups: Group T and Group F. The training epochs of the two groups were set to 120 and 150, respectively. The metric curves for all models are depicted in Figure 8. Columns (a) and (b) represent the training conditions for Group T and Group F. All metrics are averaged over

nine categories with mIoU measuring the mean Intersection over Union for all categories in the training set.

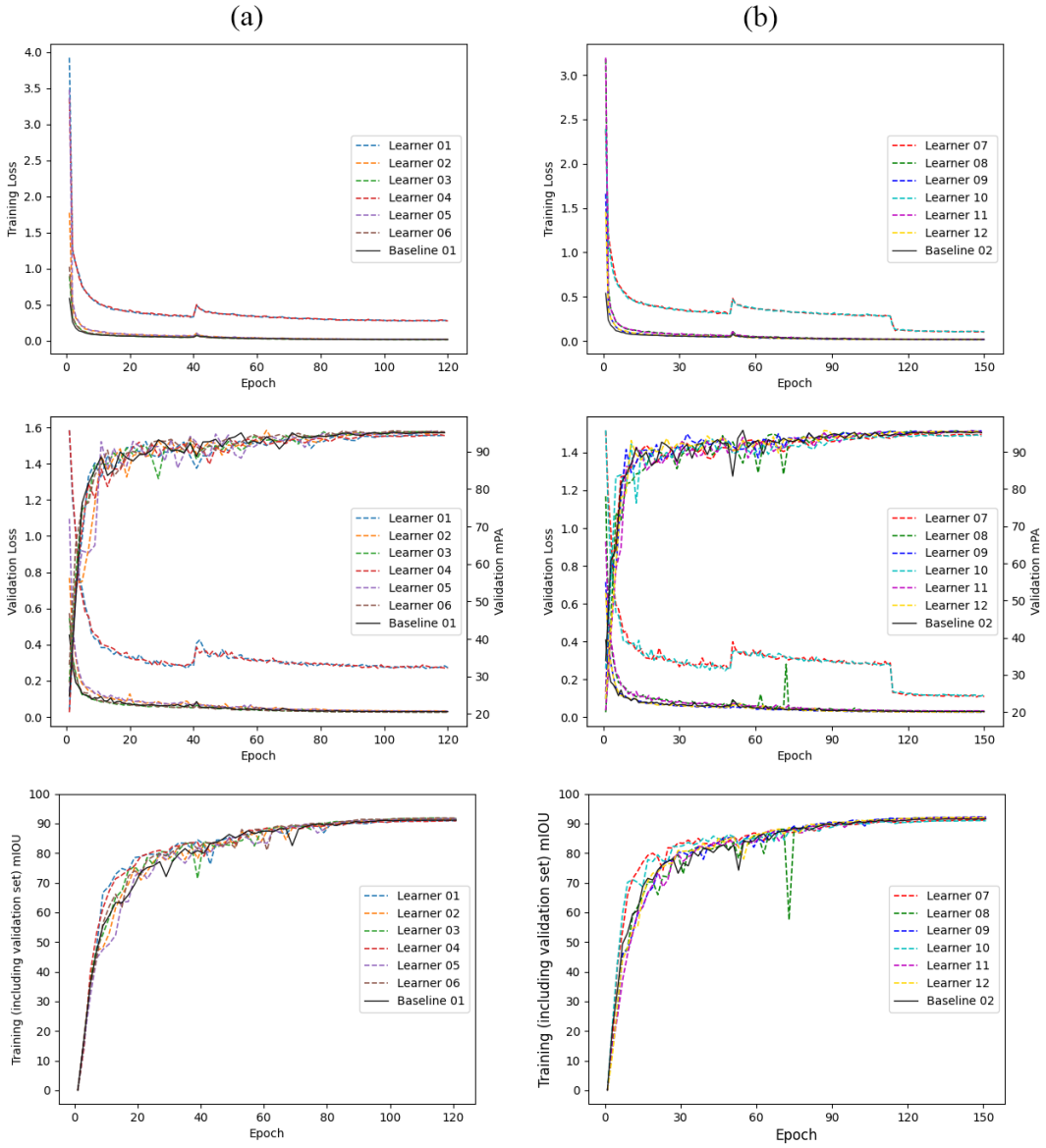


Figure 8. Training loss, validation loss, validation mPA and mean intersection over union results for all models based on (**a**) 120-epochs and (**b**) 150-epochs.

4.2. Test Set Results

All trained models conducted semantic segmentation on 1386 images in the test dataset. The test dataset underwent the same data augmentation operations as the training dataset. The quantitative metrics of the test results include mIoU, mPA, and mRecall for nine categories, from which the F1 score was calculated based on PA and Recall, along with the average values of these metrics for all categories. The 12 improved models demonstrated superior results compared to the two baselines, indicating the effectiveness of the model optimization, as detailed in Table 4.

Table 4 presents the metrics for each learner on the test set in descending order by the key of gangue's mIoU, encompassing the average of all minerals and the transparent mineral gangue. The improved 12 learners achieved higher IoU and accuracy for sphalerite and magnetite compared to the baselines. The results for gangue exhibit a substantial disparity from the average values. The main reason is the similarity between the transparent mineral gangue and the background, making identification more challenging than for other

mineral particles. Consequently, both the baseline and learners recorded the lowest IoU and accuracy for gangue particle segmentation among all categories. This context, along with the primary issues targeted by this study, was considered when constructing the ensemble learning and evaluating performance.

Table 4. Quantification summary of the test set for the evaluation of metrics ordered by mIOU of gangue.

Learner Id	All Categories		Gangue	
	mIoU (%)	F1 Score (%)	mIoU (%)	F1 Score (%)
08	91.34	95.43	83.85	91.21
06	91.23	95.37	83.26	90.87
09	91.38	95.45	83.06	90.75
02	91.08	95.29	83.06	90.74
11	90.96	95.22	82.51	90.42
05	90.98	95.23	81.84	90.01
10	90.97	95.22	81.49	89.8
03	90.82	95.13	81.38	89.73
12	91.07	95.27	81.09	89.56
01	91.36	95.43	80.99	89.5
04	90.93	95.2	80.93	89.46
07	90.94	95.2	79.95	88.86
Baseline 01	89.89	94.6	78.85	88.17
Baseline 02	89.87	94.59	78.46	87.93

Using the 12 learners, multiple Res-UNet ensemble learnings were constructed. The optimal scheme was identified by comparing IoU, F1 score, and TPI (Time Per Image). Following the guidelines in Section 2.4, the 12 base models were ranked in descending order based on the IoU and F1 score of gangue, as detailed in Table 4. Subsequently, $i(i = 3, 4 \ldots 12)$ top-ranked base learners were selected to construct the 20 ensemble learning systems. The Res-UNet ensemble learning schemes are denoted as RUEL-i-PV/WV, where RUEL represents Res-UNet Ensemble Learning, i denotes the top i learners from Table 4, PV stands for plurality voting, and WV stands for weighted voting. The semantic segmentation results for test set is presented in Table 5. Given that Baseline 01 performed slightly better than Baseline02, only Baseline01 was used in subsequent comparative experiments.

Table 5. Quantification summary of the test set for the metrics evaluation (gangue) of all ensemble learning schemes.

ID	Plurality Voting (PV)			Weighted Voting (WV)		
	F1 Score	IOU	TPI(s)	F1 Score	IOU	TPI(s)
RUEL-3	91.35	84.08	1.3	91.21	83.85	1.3
RUEL-4	91.47	84.28	1.5	91.41	84.19	1.4
RUEL-5	91.31	84.02	1.7	91.50	84.33	1.7
RUEL-6	91.34	84.06	2.0	91.44	84.23	1.9
RUEL-7	91.30	83.99	2.3	91.38	84.12	2.2
RUEL-8	91.32	84.03	2.5	91.36	84.09	2.5
RUEL-9	91.27	83.94	2.7	91.34	84.06	3.0
RUEL-10	91.29	83.97	3.1	91.33	84.05	3.1
RUEL-11	91.27	83.95	3.2	91.35	84.07	3.1
RUEL-12	91.26	83.92	3.9	91.32	84.03	3.9

As illustrated in Table 5, all ensemble learning systems exhibited varying degrees of improvement in IoU and F1 score for the gangue category compared to Learner 08, which had the best individual performance. RUEL-5-WV achieved the highest IoU value of

84.33, which was followed by RUEL-4-PV with 84.28. Compared to Baseline01, RUEL-5-WV's Gangue IoU increased by 5.48, and its F1 score increased by 3.33. When compared to Learner 08, the gangue IoU increased by 0.48 percent, and the F1 score increased by 0.29 percent. According to Table 5, the shortest processing time for a single image is 1.3 s, and the longest is 3.9 s. Considering both segmentation accuracy and time efficiency for the gangue category of transparent minerals, RUEL-5-WV performed the best among the constructed ensemble learning systems.

Table 6 presents the IoU and number of labeled particles for each category in the test set for Baseline 01, Learner 08, and RUEL-5-WV. The "Particle Nums" column in Table 6 is the total number of each mineral segments in the training set. This column highlights data imbalance in the training set, with magnetite having the highest number of particles (11,160) and pyrrhotite the fewest (216). The IoU results indicate that Learner 08 and RUEL-5-WV achieve the most significant enhancements in segmentation performance for categories with fewer particles and lower IoU. Learner 08 employs CE Loss + focal loss, suggesting that focal loss helps address data imbalance in mineral microscopic images, and ensemble learning systems further enhance semantic segmentation performance. For categories with fewer particles such as bornite and pyrrhotite, which feature distinct colors, regular shapes, and clear edges, Baseline 01 already achieves IoUs of 96.02% and 90.74%, respectively, leading to smaller IoU gains with ensemble learning systems. Compared to Learner 08, RUEL-5-WV achieves additional enhancements in most single-category and overall mean IoUs. However, the TPI is nearly eight times that of Learner 08, and the size of the test dataset needs to be considered when applying them. Compared to other two studies of this area [4,5], both RUEL-5-WV and Learner 08 achieved superior results.

Table 6. Comparison of IoU and number of particles for each category.

Category	Baseline 01 (IoU)	Learner 08 (IoU)	RUEL-5-WV (IoU)	Particle Nums
Background	98.40	98.69	98.72	900
Pyrite	90.64	93.57	93.77	4686
Galena	91.94	91.91	92.11	9444
Sphalerite	90.31	90.47	90.44	7344
Chalcopyrite	82.03	87.13	88.08	486
Bornite	96.02	95.75	96.06	390
Magnetite	90.12	89.65	90.17	11,160
Pyrrhotite	90.74	91.02	91.14	216
Gangue	78.85	83.85	84.33	1830
Mean	89.89	91.34	91.65	4051

4.3. Qualitative Results

Along with a comparison to the ground truth, the original images and the results of semantic segmentation by Baseline 01, Learner 08, and RUEL-5-WV are displayed in Figure 9. For the particles or regions indicated by the white arrows, RUEL-5-WV achieved superior results. In images 1 and 2, the primary categories of particles are gangue and pyrite. Images 3 and 4 feature monomeric and composite particles of sphalerite and magnetite. Image 5 mainly contains pyrrhotite and magnetite particles, with the magnetite particles being relatively dispersed. Image 6 includes various categories of minerals, such as sphalerite, bornite, galena, pyrite, and chalcopyrite. In the dataset, there are many images on which multiple different categories of mineral particles are present, just like image 1, image 2, image 3, image 5 and image 6 in Figure 9. Among these categories, gangue is a transparent mineral with particles exhibiting color characteristics similar to the background, some of which have clear edges. The semi-transparent sphalerite and opaque magnetite display similar colors.

The segmentation results of various models indicate that RUEL-5-WV and Learner 08 provide more accurate segmentation for the transparent mineral gangue. Compared to Baseline 01, RUEL-5-WV and Learner 08 distinguish sphalerite and magnetite in composite

particles more accurately. The segmentation results for small granularity particles across multiple images show that RUEL-5-WV achieves better outcomes. For images like image 6, which contain multiple types of mineral particles, RUEL-5-WV generally identifies them accurately and performs better than other models. However, some shortcomings were identified in the improved models and ensemble learning systems. The segmentation results for gangue particles with unclear edge contours are suboptimal. The micro or blurry particles are easily overlooked.

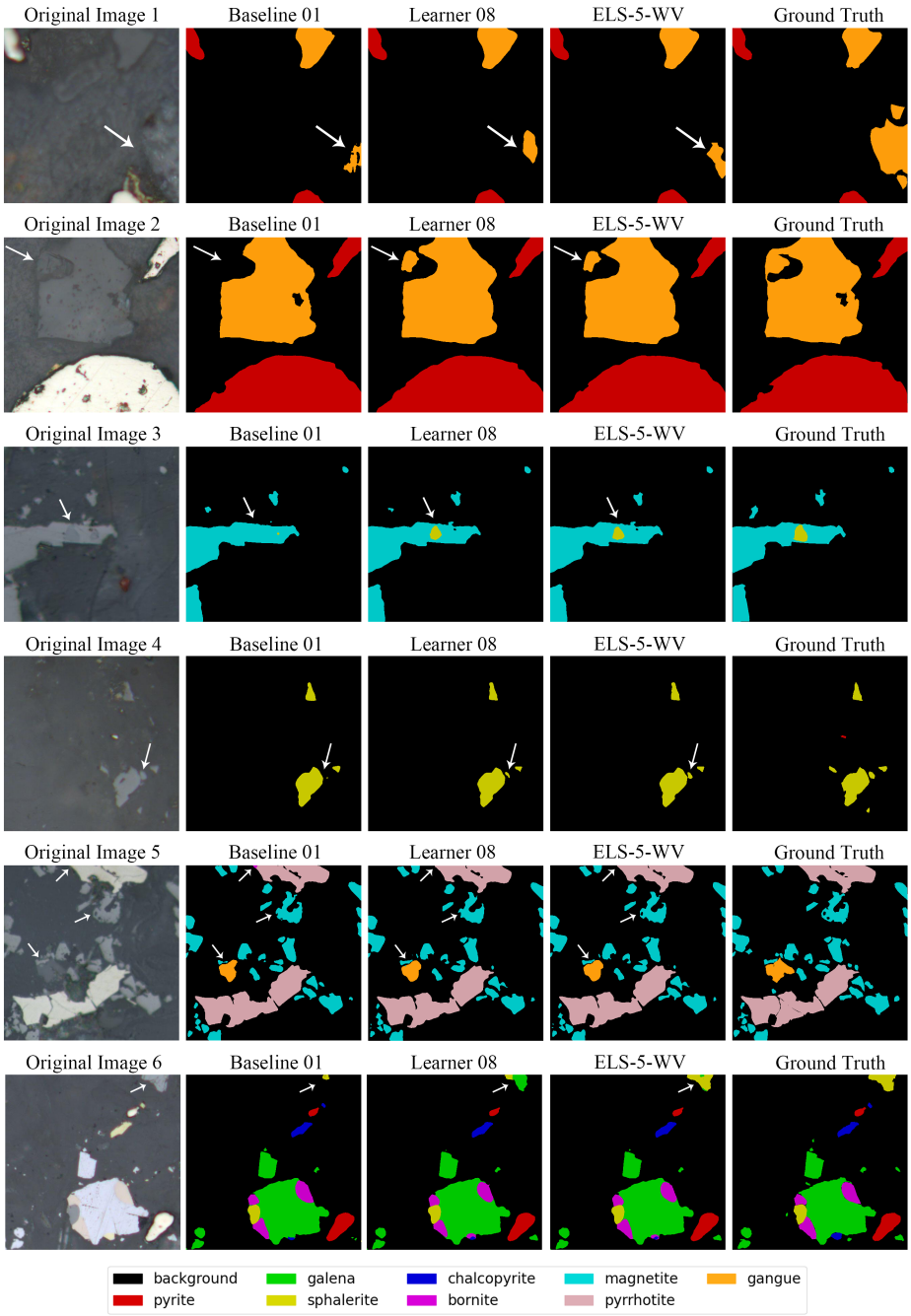


Figure 9. Comparison of segmentation results of sample images. For the particles or regions indicated by the white arrows, ELS-5-WV achieved superior segmentation results.

4.4. Comparative Experiment

Experiments were conducted on the dataset using several other existing algorithms, including Deeplab v3 and PSPNet, which are all outstanding semantic segmentation algorithms that have emerged in recent years. The comparative results of the quantification

for each model are presented in Table 7, which includes the Intersection over Union (IOU) and F1 score for all mineral categories, along with their means. The proposed approach (RUEL-WV-5) achieved higher IOU and F1 score compared to Deeplab v3 and PSPNet, particularly for minerals such as pyrrhotite, bornite, chalcopyrite, galena and gangue.

Table 7. Comparison of the proposed model with other algorithms.

	IOU			F1 Score		
	RUEL-WV-5	**Deeplabv3**	**PSPNet**	**RUEL-WV-5**	**Deeplabv3**	**PSPNet**
gangue	84.33	82.16	73.79	91.50	90.21	84.42
pyrrhotite	91.14	86.79	74.44	95.37	92.93	85.35
magnetite	90.17	88.39	67.25	94.83	93.84	80.42
bornite	96.06	90.91	82.84	97.99	95.24	90.61
chalcopyrite	88.08	82.71	70.77	93.66	90.54	82.88
sphalerite	90.44	89.99	77.76	94.98	94.73	87.49
galena	92.11	88.74	74.69	95.90	94.03	85.52
pyrite	93.77	92.48	83.39	96.78	96.58	90.94
background	98.72	98.51	96.25	99.35	99.25	98.09
mean	91.65	88.97	77.91	95.60	94.11	87.38

5. Conclusions

A flexible semantic segmentation ensemble learning system based on multiple optimized UNet models is proposed. The results indicate that this system outperforms traditional models in identifying transparent and similar minerals. For minority minerals in the dataset, the segmentation performance of the ensemble learning system is significantly enhanced. The proposed system is not a trainable neural network module but an integrated solution. It is compatible with other semantic segmentation models without altering their structures and offers good scalability.

Some issues in this field have also been reflected in the research process. Compared to accurately identified large, normal and small particles, some microparticles were not identified. The supervised learning approach used in this study is data-driven, requiring pre-annotation for mask generation, which is a time-consuming and labor-intensive task. Thus, despite achieving good results, the data processing stage remains cumbersome. One possible reason why some microparticles were not identified is that the particles of this size were not annotated in the training set.

Future research can focus on the following directions using deep learning-based computer vision technology:

1. Enhance the accuracy of identifying micro mineral grains by applying more efficient data annotation methods and multi-scale modules.
2. Implement faster intelligent identification in small-scale datasets using semi-supervised, weakly supervised, or unsupervised learning methods with minimal image annotation.
3. Employ instance segmentation for intergrown minerals to count mineral grains and automatically calculate the liberation degree.

Author Contributions: Conceptualization, C.J., A.A.H., L.N.A. and N.M.; Methodogy, C.J., B.Y., L.N.A. and T.P.; Software, C.J.; Visualization, C.J.; Writing—Original Draft, C.J.; Writing—Review and Editing, C.J.; Resources, A.A.H. and B.Y.; Supervision, A.A.H.; Data Curation, B.Y.; Funding Acquisition, B.Y.; Validation, B.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Scientific Research Fund of Hunan Provincial Education Department (No. 21A0603).

Data Availability Statement: Dataset inquiries can be directed to the corresponding author. The source codes are available for downloading at the link: https://github.com/jessiejch/eruls_mineral_segmentation (accessed on 12 December 2024).

Acknowledgments: We thank Scientific Research Fund of Hunan Provincial Education Department for providing funding. We sincerely thank Hunan Research Institute for Nonferrous Metals for granting access to their experimental facilities, and Senior Engineer Zhilian Lei and Senior Engineer JianXiong Wang for their professional technical assistance and valuable contributions to this study.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AMI	automatic mineral identification
CT	Computed Tomography
CNNs	Convolutional Neural Networks
XPL	cross-polarized light
PPL	plane-polarized light
STN	Spatial Transformer Networks
NLP	Natral Language Processing
CCD	Charge-Coupled Device
RUEL	Res-UNet Ensemble Learning
mPA	mean pixel accuracy
WV	weighted voting
PV	plurality voting

References

1. Donskoi, E.; Manuel, J.R.; Hapugoda, S.; Poliakov, A.; Raynlyn, T.; Austin, P.; Peterson, M. Automated optical image analysis of goethitic iron ores. *Miner. Process. Extr. Metall.* **2022**, *131*, 14–24. [CrossRef]
2. Álvarez Iglesias, J.C.; Santos, R.B.M.; Paciornik, S. Deep learning discrimination of quartz and resin in optical microscopy images of minerals. *Miner. Eng.* **2019**, *138*, 79–85. [CrossRef]
3. Zhang, Y.; Li, M.; Han, S.; Ren, Q.; Shi, J. Intelligent Identification for Rock-Mineral Microscopic Images Using Ensemble Machine Learning Algorithms. *Sensors* **2019**, *19*, 3914. [CrossRef] [PubMed]
4. Razzhivina, D.I.; Korshunov, D.M.; Boguslavsky, M.A.; Khvostikov, A.V.; Sorokin, D.V. Registration and segmentation of PPL and XPL images of geological polished sections containing anisotropic minerals. *Comput. Math. Model.* **2024**, *34*, 16–26. [CrossRef]
5. De Castro, B.; Benzaazoua, M.; Roychowdhury, S.; Chopard, A.; Quintal Lauzon, F.; Plante, B. Novel technique for the preparation and analysis of powder-based polished sections by automated optical mineralogy: Part 2—Use of deep learning approach for transparent mineral detection. *Miner. Eng.* **2024**, *206*, 108508. [CrossRef]
6. Amaral Pascarelli Ferreira, B.; Soares Augusto, K.; César Álvarez Iglesias, J.; Dias Pinheiro Caldas, T.; Bryan Magalhães Santos, R.; Paciornik, S. Instance segmentation of quartz in iron ore optical microscopy images by deep learning. *Miner. Eng.* **2024**, *211*, 108681. [CrossRef]
7. Filippo, M.P.; da Fonseca Martins Gomes, O.; da Costa, G.A.O.P.; Mota, G.L.A. Deep learning semantic segmentation of opaque and non-opaque minerals from epoxy resin in reflected light microscopy images. *Miner. Eng.* **2021**, *170*, 107007. [CrossRef]
8. Koh, E.J.; Amini, E.; McLachlan, G.J.; Beaton, N. Utilising convolutional neural networks to perform fast automated modal mineralogy analysis for thin-section optical microscopy. *Miner. Eng.* **2021**, *173*, 107230. [CrossRef]
9. Poliakov, A.; Donskoi, E. Automated relief-based discrimination of non-opaque minerals in optical image analysis. *Miner. Eng.* **2014**, *55*, 111–124. [CrossRef]
10. Donskoi, E.; Poliakov, A.; Manuel, J. 4—Automated Optical Image Analysis of Natural and Sintered Iron Ore. In *Iron Ore*; Lu, L., Ed.; Woodhead Publishing: Sawston, UK, 2015; pp. 101–159. [CrossRef]
11. Donskoi, E.; Hapugoda, S.; Manuel, J.R.; Poliakov, A.; Peterson, M.J.; Mali, H.; Bückner, B.; Honeyands, T.; Pownceby, M.I. Automated Optical Image Analysis of Iron Ore Sinter. *Minerals* **2021**, *11*, 562. [CrossRef]
12. Goodchild, J.; Fueten, F. Edge detection in petrographic images using the rotating polarizer stage. *Comput. Geosci.* **1998**, *24*, 745–751. [CrossRef]
13. Heilbronner, R. Automatic grain boundary detection and grain size analysis using polarization micrographs or orientation images. *J. Struct. Geol.* **2000**, *22*, 969–981. [CrossRef]
14. Zhou, Y.; Starkey, J.; Mansinha, L. Segmentation of petrographic images by integrating edge detection and region growing. *Comput. Geosci.* **2004**, *30*, 817–831. [CrossRef]
15. Barraud, J. The use of watershed segmentation and GIS software for textural analysis of thin sections. *J. Volcanol. Geotherm. Res.* **2006**, *154*, 17–33. [CrossRef]

16. Obara, B. A new algorithm using image colour system transformation for rock grain segmentation. *Mineral. Petrol.* **2007**, *91*, 271–285. [CrossRef]
17. Fueten, F.; Mason, J. An artificial neural net assisted approach to editing edges in petrographic images collected with the rotating polarizer stage. *Comput. Geosci.* **2007**, *33*, 1176–1188. [CrossRef]
18. Hoffmann, P.; Marschallinger, R.; Unterwurzacher, M.; Zobl, F. Marble provenance designation with Object Based Image Analysis: State-of-the-art rock fabric characterization from petrographic micrographs. *Austrian J. Earth Sci.* **2013**, 40–49.
19. Asmussen, P.; Conrad, O.; Günther, A.; Kirsch, M.; Riller, U. Semi-automatic segmentation of petrographic thin section images using a "seeded-region growing algorithm" with an application to characterize wheathered subarkose sandstone. *Comput. Geosci.* **2015**, *83*, 89–99. [CrossRef]
20. Izadi, H.; Sadri, J.; Mehran, N.A. A new intelligent method for minerals segmentation in thin sections based on a novel incremental color clustering. *Comput. Geosci.* **2015**, *81*, 38–52. [CrossRef]
21. Han, Z.; Li, J.; Zhang, B.; Hossain, M.M.; Xu, C. Prediction of combustion state through a semi-supervised learning model and flame imaging. *Fuel* **2021**, *289*, 119745. [CrossRef]
22. Lei, M.; Rao, Z.; Wang, H.; Chen, Y.; Zou, L.; Yu, H. Maceral groups analysis of coal based on semantic segmentation of photomicrographs via the improved U-net. *Fuel* **2021**, *294*, 120475. [CrossRef]
23. Wang, Y.; Bai, X.; Wu, L.; Zhang, Y.; Qu, S. Identification of maceral groups in Chinese bituminous coals based on semantic segmentation models. *Fuel* **2022**, *308*, 121844. [CrossRef]
24. Wang, H.; Lei, M.; Chen, Y.; Li, M.; Zou, L. Intelligent Identification of Maceral Components of Coal Based on Image Segmentation and Classification. *Appl. Sci.* **2019**, *9*, 3245. [CrossRef]
25. Tiwary, A.K.; Ghosh, S.; Singh, R.; Mukherjee, D.P.; Shankar, B.U.; Dash, P.S. Automated coal petrography using random forest. *Int. J. Coal Geol.* **2020**, *232*, 103629. [CrossRef]
26. Wang, B.; Han, G.; Ma, H.; Zhu, L.; Liang, X.; Lu, X. Rock thin sections identification under harsh conditions across regions based on online transfer method. *Comput. Geosci.* **2022**, *26*, 1425–1438. [CrossRef]
27. Shirmard, H.; Farahbakhsh, E.; Müller, R.D.; Chandra, R. A review of machine learning in processing remote sensing data for mineral exploration. *Remote Sens. Environ.* **2022**, *268*, 112750. [CrossRef]
28. Liu, Y.; Zhang, Z.; Liu, X.; Wang, L.; Xia, X. Efficient image segmentation based on deep learning for mineral image classification. *Adv. Powder Technol.* **2021**, *32*, 3885–3903. [CrossRef]
29. Khvostikov, A.V.; Korshunov, D.M.; Krylov, A.S.; Boguslavskiy, M.A. Automatic identification of minerals in images of polished sections. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *XLIV-2/W1-2021*, 113–118. [CrossRef]
30. Saxena, N.; Day-Stirrat, R.J.; Hows, A.; Hofmann, R. Application of deep learning for semantic segmentation of sandstone thin sections. *Comput. Geosci.* **2021**, *152*, 104778. [CrossRef]
31. Tang, H.; Wang, H.; Wang, L.; Cao, C.; Nie, Y.; Liu, S. An Improved Mineral Image Recognition Method Based on Deep Learning. *JOM* **2023**, *75*, 2590–2602. [CrossRef]
32. De Castro, B.; Benzaazoua, M.; Chopard, A.; Plante, B. Automated mineralogical characterization using optical microscopy: Review and recommendations. *Miner. Eng.* **2022**, *189*, 107896. [CrossRef]
33. De Castro, B.; Benzaazoua, M.; St-Jean, A.; Scortino, M.; Plante, B.; Bélisle, B.; Cloutier, R. Automated mineralogy using optical microscopy in a geometallurgical context: A comparative study on Dumont nickel project ores, Amos, Quebec. *Miner. Eng.* **2023**, *198*, 108089. [CrossRef]
34. Latif, G.; Bouchard, K.; Maitre, J.; Back, A.; Bédard, L.P. Deep-Learning-Based Automatic Mineral Grain Segmentation and Recognition. *Minerals* **2022**, *12*, 455. [CrossRef]
35. Zhang, Z.; Li, Y.; Wang, G.; Carranza, E.J.M.; Yang, S.; Sha, D.; Fan, J.; Zhang, X.; Dong, Y. Supervised Mineral Prospectivity Mapping via Class-Balanced Focal Loss Function on Imbalanced Geoscience Datasets. *Math. Geosci.* **2023**, *55*, 989–1010. [CrossRef]
36. Farahbakhsh, E.; Maughan, J.; Müller, R.D. Prospectivity modelling of critical mineral deposits using a generative adversarial network with oversampling and positive-unlabelled bagging. *Ore Geol. Rev.* **2023**, *162*, 105665. [CrossRef]
37. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016; pp. 770–778. [CrossRef]
38. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. *IEEE Access* **2021**, *9*, 82031–82057. [CrossRef]
39. Shafiq, M.; Gu, Z. Deep Residual Learning for Image Recognition: A Survey. *Appl. Sci.* **2022**, *12*, 8972. [CrossRef]
40. Ramachandran, P.; Zoph, B.; Le, Q.V. Searching for Activation Functions. *arXiv* **2017**, arXiv:1710.05941. [CrossRef]
41. Sharma, S.; Sharma, S.; Athaiya, A. Activation functions in neural networks. *Towards Data Sci.* **2017**, *6*, 310–316. [CrossRef]
42. Apicella, A.; Donnarumma, F.; Isgrò, F.; Prevete, R. A survey on modern trainable activation functions. *Neural Netw.* **2021**, *138*, 14–32. [CrossRef] [PubMed]
43. Dubey, S.R.; Singh, S.K.; Chaudhuri, B.B. Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing* **2022**, *503*, 92–108. [CrossRef]
44. Jaderberg, Simonyan, Zisserman, and kavukcuoglu. Spatial Transformer Networks. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2015; Volume 28.
45. Yu, H.; Xu, Z.; Zheng, K.; Hong, D.; Yang, H.; Song, M. MSTNet: A Multilevel Spectral—Spatial Transformer Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5532513. [CrossRef]

46. Rajput, V. Robustness of different loss functions and their impact on networks learning capability. *arXiv* **2021**, arXiv:2110.08322 .
47. Tian, Y.; Su, D.; Lauria, S.; Liu, X. Recent advances on loss functions in deep learning for computer vision. *Neurocomputing* **2022**, *497*, 129–158. [CrossRef]
48. Milletari, F.; Navab, N.; Ahmadi, S.A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571. [CrossRef]
49. Jadon, S. A survey of loss functions for semantic segmentation. In Proceedings of the 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Via del Mar, Chile, 27–29 October 2020; pp. 1–7. [CrossRef]
50. Li, X.; Sun, X.; Meng, Y.; Liang, J.; Wu, F.; Li, J. Dice Loss for Data-imbalanced NLP Tasks. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; Jurafsky, D., Chai, J., Schluter, N., Tetreault, J., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 465–476. [CrossRef]
51. Zhang, X.; Zang, L.; Liu, Q.; Wei, S.; Hu, S. MetaPETR: An Effective Model for Handling Class-Imbalanced Data About Event Temporal Relations. In *Proceedings of the Advanced Intelligent Computing Technology and Applications*; Springer Nature: Singapore, 2024; pp.390–401.
52. You, X.-X.; Liang, Z.-M.; Wang, Y.-Q.; Zhang, H. A study on loss function against data imbalance in deep learning correction of precipitation forecasts. *Atmos. Res.* **2023**, *281*, 106500. [CrossRef]
53. Hossain, M.S.; Betts, J.M.; Paplinski, A.P. Dual Focal Loss to address class imbalance in semantic segmentation. *Neurocomputing* **2021**, *462*, 69–87. [CrossRef]
54. Pasupa, K.; Vatathanavaro, S.; Tungjitnob, S. Convolutional neural networks based focal loss for class imbalance problem: A case study of canine red blood cells morphology classification. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 15259–15275. [CrossRef]
55. Büttner, M.; Schneider, L.; Krasowski, A.; Pitchika, V.; Krois, J.; Meyer-Lueckel, H.; Schwendicke, F. Conquering class imbalances in deep learning-based segmentation of dental radiographs with different loss functions. *J. Dent.* **2024**, *148*, 105063. [CrossRef]
56. Yang, Y.; Lv, H.; Chen, N. A Survey on ensemble learning under the era of deep learning. *Artif. Intell. Rev.* **2023**, *56*, 5545–5589. [CrossRef]
57. Mienye, I.D.; Sun, Y. A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access* **2022**, *10*, 99129–99149. [CrossRef]
58. Ganaie, M.; Hu, M.; Malik, A.; Tanveer, M.; Suganthan, P. Ensemble deep learning: A review. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105151. [CrossRef]
59. Mohammed, A.; Kora, R. A comprehensive review on ensemble deep learning: Opportunities and challenges. *J. King Saud Univ.-Comput. Inf. Sci.* **2023**, *35*, 757–774. [CrossRef]
60. Xue, D.; Zhou, X.; Li, C.; Yao, Y.; Rahaman, M.M.; Zhang, J.; Chen, H.; Zhang, J.; Qi, S.; Sun, H. An Application of Transfer Learning and Ensemble Learning Techniques for Cervical Histopathology Image Classification. *IEEE Access* **2020**, *8*, 104603–104618. [CrossRef]
61. Moon, W.K.; Lee, Y.W.; Ke, H.H.; Lee, S.H.; Huang, C.S.; Chang, R.F. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Comput. Methods Programs Biomed.* **2020**, *190*, 105361. [CrossRef] [PubMed]
62. Hu, J.; Gu, X.; Gu, X. Mutual ensemble learning for brain tumor segmentation. *Neurocomputing* **2022**, *504*, 68–81. [CrossRef]
63. Du, L.; Liu, H.; Zhang, L.; Lu, Y.; Li, M.; Hu, Y.; Zhang, Y. Deep ensemble learning for accurate retinal vessel segmentation. *Comput. Biol. Med.* **2023**, *158*, 106829. [CrossRef] [PubMed]
64. S, S.; G, M.; Sherly, E.; Mathew, R. M-Net: An encoder-decoder architecture for medical image analysis using ensemble learning. *Results Eng.* **2023**, *17*, 100927. [CrossRef]
65. Dang, T.; Nguyen, T.T.; McCall, J.; Elyan, E.; Moreno-García, C.F. Two-layer Ensemble of Deep Learning Models for Medical Image Segmentation. *Cogn. Comput.* **2024**, *16*, 1141–1160. [CrossRef]
66. Ennaji, A.; Khoukhi, H.E.; Sabri, M.A.; Aarab, A. Malignant melanoma detection using multi-scale image decomposition and a new ensemble-learning scheme. *Multimed. Tools Appl.* **2024**, *83*, 21213–21228. [CrossRef]
67. Singh, S.; Singh, B.; Kumar, A. Multi-organ segmentation of organ-at-risk (OAR's) of head and neck site using ensemble learning technique. *Radiography* **2024**, *30*, 673–680. [CrossRef]
68. Xiao, X.; Lian, S.; Luo, Z.; Li, S. Weighted Res-UNet for High-Quality Retina Vessel Segmentation. In Proceedings of the 2018 9th International Conference on Information Technology in Medicine and Education (ITME), Hangzhou, China, 19–21 October 2018; pp. 327–331. [CrossRef]
69. Liu, Z.; Yuan, H. An Res-Unet Method for Pulmonary Artery Segmentation of CT Images. *J. Phys. Conf. Ser.* **2021**, *1924*, 012018. [CrossRef]
70. Kesavan, S.M.; Al Naimi, I.; Al Attar, F.; Rajinikanth, V.; Kadry, S. Res-UNet Supported Segmentation and Evaluation of COVID-19 Lesion in Lung CT. In Proceedings of the 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), Puducherry, India, 30–31 July 2021; pp. 1–4. [CrossRef]
71. Maji, D.; Sigedar, P.; Singh, M. Attention Res-UNet with Guided Decoder for semantic segmentation of brain tumors. *Biomed. Signal Process. Control.* **2022**, *71*, 103077. [CrossRef]

72. Huang, L.; Miron, A.; Hone, K.; Li, Y. Segmenting Medical Images: From UNet to Res-UNet and nnUNet. In Proceedings of the 2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS), Guadalajara, Mexico, 26–28 June 2024; pp. 483–489. [CrossRef]
73. Li, X.; Fang, Z.; Zhao, R.; Mo, H. Brain Tumor MRI Segmentation Method Based on Improved Res-UNet. *IEEE J. Radio Freq. Identif.* **2024**, *8*, 652–657. [CrossRef]
74. Lee, M.C.H.; Oktay, O.; Schuh, A.; Schaap, M.; Glocker, B. Image-and-Spatial Transformer Networks for Structure-Guided Image Registration. In *Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2019*; Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A., Eds.; Springer: Cham, Switzerland, 2019; pp. 337–345.
75. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [CrossRef]
76. Aljabri, M.; AlAmir, M.; AlGhamdi, M.; Abdel-Mottaleb, M.; Collado-Mesa, F. Towards a better understanding of annotation tools for medical imaging: a survey. *Multimed. Tools Appl.* **2022**, *81*, 25877–25911. [CrossRef] [PubMed]