electronics

MDPI

Article

A Classification and Segmentation Model for Diamond Abrasive Grains Based on Improved Swin-Unet-SAM

Yanfen Lin [1], Tinghao Fan [1] and Congfu Fang [2,*]

[1] School of Data and Computer Science, Xiamen Institute of Technology, Xiamen 361021, China; linyanfen@xit.edu.cn (Y.L.); 2006120205@stumail.xit.edu.cn (T.F.)
[2] College of Mechanical Engineering and Automation, Huaqiao University, Xiamen 361021, China
* Correspondence: cffang@hqu.edu.cn

Abstract: The detection of abrasive grain images in diamond tools serves as the foundation for assessing the overall condition of the tools, encompassing crucial aspects of diamond abrasive grains like the quantity, size, morphology, and distribution. Given the intricate background textures and reflective characteristics exhibited by diamond images, diamond detection and segmentation pose a significant challenge. Recently, numerous defect detection methods based on machine learning and deep learning have emerged. However, several issues persist, such as detection accuracy and the interference caused by intricate background textures. The present work demonstrates an efficient classification and segmentation network algorithm that combines Swin-Unet with SAM (Segment Anything Model) to alleviate the existing problems. Specifically, four embedding structures were devised to bridge the two models for iterative training. The transformer blocks within the Swin-Unet model were enhanced to facilitate classification and coarse segmentation, and the mask structure in SAM was refined to enable fine segmentation. The experimental results show that under a small sample dataset with complex background textures, the average index values of *ACC* (accuracy), *SE* (Sensitivity), and *DSC* (Dice Similarity Coefficient) for the classification and segmentation of diamond abrasive grains reached 98.7%, 92.5%, and 85.9%, respectively. Compared with the model before improvement, its *ACC*, *SE* and *DSC* increased by 1.2%, 15.9%, and 7.6%, respectively. The test results, based on four different datasets, consistently indicated that this model has excellent segmentation performance and robustness and has great application potential in the industrial field.

Keywords: classification and segmentation; diamond detection; Swin-Unet; SAM; deep learning

check for updates

Citation: Lin, Y.; Fan, T.; Fang, C. A Classification and Segmentation Model for Diamond Abrasive Grains Based on Improved Swin-Unet-SAM. *Electronics* **2024**, *13*, 4213. https://doi.org/10.3390/electronics13214213

Academic Editor: Beiwen Li

Received: 7 October 2024
Revised: 22 October 2024
Accepted: 26 October 2024
Published: 27 October 2024

1. Introduction

In the manufacturing industry, diamond abrasive grains are widely used in grinding, lapping, polishing, and other fields due to their high hardness and high wear resistance, and they have become the main tools for processing hard and brittle materials [1,2]. Diamond abrasive grains fixed in these tools act on the workpiece material through scratching, plowing, and cutting to achieve processing control of the shape accuracy and surface quality of the workpiece. During the processing, both the binder holding the abrasive grains and the abrasive grains will be eroded by the workpiece material and coolant, resulting in situations such as the breakage and wear of the diamond abrasive grains, and severe wear of the binder will also cause the abrasive grains to fall off directly [3–5]. Therefore, the surface quality of the workpiece is profoundly affected by the morphology state of the abrasive grains on the tool surface, and the study of abrasive grain wear is of great significance for tool condition detection [6,7]. Due to the complexity of the environment, the multiplicity of external parameters, and the unique geometry of the abrasive grains, the identification and detection of abrasive grains still have great difficulties. Many scholars establish models by actually measuring the quantity of static abrasive grains and their degree of wear [8–10] in order to establish more accurate simulation and prediction models, but this requires accurate characteristic data of diamond abrasive grains.

In recent years, intelligent means such as machine vision and pattern recognition have replaced manual labor and gradually become the mainstream trend in the field of identification and detection. The processes of feature extraction and recognition classification are combined and a multi-layer neural network structure is utilized to automatically combine the simple features extracted at the lower layer and send them layer by layer to the top layer network to learn advanced features that can better represent the object, which has a strong model expression ability. This technology has been widely applied in image recognition tasks in different scenarios, such as face recognition, license plate recognition, speech recognition and defect detection, foreign object detection on railroad transmission lines, and mineral zoning detection [11–15]. Therefore, applying deep learning to the detection of diamond abrasive grain tools and completing the detection of diamond abrasive grain images through automatic learning is an effective approach.

Taking advantage of the superiority of deep learning, this paper presents a novel model based on Swin-Unet and SAM (Segment Anything Model) for the classification and segmentation of diamond abrasive grains. The main contributions of this paper are summarized as follows:

(1) A Swin-Unet-SAM network structure is designed based on Swin-Unet and SAM for the classification and segmentation of diamond abrasive grains.
(2) Four embedding structures are devised to bridge the two models of Swin-Unet and SAM for iterative training.
(3) The transformer blocks within the Swin-Unet model are enhanced to facilitate classification and coarse segmentation.
(4) The mask structure in SAM is refined to enable the fine segmentation performance of diamond abrasive grains.

This paper is structured into six sections. Section 1 describes the background of diamond abrasive grain detection and the motivation for writing this paper. Section 2 narrates and discusses the related work on diamond abrasive grain detection and segmentation and summarizes the existing problems and the contributions of this article. Section 3 provides the theories and knowledge required for this paper and presents the model and methods proposed in this paper. Section 4 describes the experimental details of this paper, the source characteristics of the dataset, and the model performance evaluation method. Section 5 analyzes and discusses the model performance and ablation experiments. Finally, Section 6 provides a summary of this paper.

2. Related Work

For different identification and detection problems and targets, different processing methods are selected, and there is no unified algorithm that is universally applicable. For the segmentation of diamond abrasive grains and similar small sample semantic segmentation, they can mainly be divided into feature extraction-based traditional methods and deep learning-based intelligent methods.

2.1. Feature Extraction-Based Traditional Method

Traditional methods based on feature extraction extract representative feature attributes from the image, perform a set transformation, and then use a certain threshold for segmentation. Yu et al. [16] proposed an abrasive grain image segmentation method based on background color recognition. After pre-segmentation using the gray threshold method, the gray average and standard deviation of the three-color components were calculated in the RGB channels and a certain component range was set to extract the abrasive grains. This has a certain improvement compared to using only the gray threshold method. In order to address the defect of mis-segmentation in the region division by the threshold segmentation method of the gray histogram of abrasive grains on diamond grinding wheel surfaces, Wu et al. [17] proposed an approach for segmenting abrasive grain images based on the connected region marking of the quadratic gray-level histogram. This method has a better segmentation effect, but is not applicable to abrasive grain images with complex textures. In view of the influence of complex background textures on the segmentation of abrasive grains, Gong et al. [18] generated a fusion image and a height matrix of diamond grains through the fusion method of focusing and applied the Snake model combined with the height index image to generate a new image energy function suitable for segmentation, but it was impossible to quantitatively analyze its wear degree. Although this method can effectively segment abrasive grain images, it has high requirements for the focused fusion images and is inefficient. Subsequently, Lin et al. [19] proposed an invalid edge elimination method to solve irregular background textures and obtained good segmentation results in some specific scenarios. Moreover, an abrasive grain segmentation method based on morphological reconstruction and bit-planes was further proposed in order to improve the segmentation accuracy based on the combined concept of coarse segmentation and fine segmentation [20]. Besides directly focusing on the segmentation of abrasive grains, some scholars have also conducted studies on the distribution of abrasive grains [21,22]. In these studies, the threshold segmentation method is mainly used. Since only the position distribution of abrasive grains is concerned, the requirements for the threshold segmentation results are not high.

In summary, feature extraction-based traditional methods have the advantages of high computational efficiency, strong interpretability, and no need for a large amount of training data, and still have a place in the segmentation field of small samples such as abrasive grains. However, due to the complexity of diamond images and the similar features of the abrasive grain targets and the image background, this leads to the limited feature parameters selected by traditional methods, resulting in inflexible models, low segmentation accuracy and poor robustness, which is difficult to obtain satisfactory results.

2.2. Deep Learning-Based Intelligent Method

In view of the certain limitations of methods based on traditional machine vision in terms of detection efficiency, detection accuracy and detection applicability in complex scenarios. In recent years, methods based on deep learning have developed rapidly; in particular, those using CNNs (convolutional neural networks) have the ability to handle complex scenarios, and the model accuracy and generalization ability in handling complex scenarios have been greatly improved. In the field of abrasive particle detection of ferrography, Wang et al. [23] combined technology in the field of image recognition, innovatively linked an image recognition model based on a CNN with abrasive particle analysis, and adopted a two-level abrasive particle recognition model in order to classify abrasive particles with similarity. Then, Peng et al. [24] proposed a hybrid convolution neural network with transfer learning and a support vector machine which was used to automatically identify wear particles. When the number of classes of particle wear increases, the previous automatic recognition model needs to be reconstructed and retrained. To solve this problem, Zhang et al. [25] proposed a CNN model based on class center vectors and distance comparison and constructed a feature extraction module, a class center vector extraction module, and a prediction module which could realize the classification of new types of wear grains. On the basis of the classification of wear particles, Yang et al. [26] proposed a digital characterization method of abrasive particles based on a Mask R-CNN. This method uses transfer-learning training to achieve the recognition and instance segmentation of abrasive particles in images. Although the importance of the CNN method in the intelligent recognition of wear particles was affirmed, they point out that the models of wear particle recognition in the aspects of the datasets still have a high labor cost and non-complete objective [27].

Compared with the abrasive wear detection of ferrography, the research at the level of diamond abrasive particles is much less and much delayed. Recently, in the field of diamond industrial production, Yang et al. [28] proposed a detection method of yellow industrial diamond grains based on deep learning: the purpose was to solve the problems of the slow manual detection speed, high labor intensity, and low-quality consistency of yellow industrial diamonds in industrial production. After constructing three base classifiers using three network structures, an integrated fusion method was adopted to realize the information fusion of multiple base classifiers and the classification decisions for yellow industrial diamond grains, and the class recognition *ACC* (accuracy) reached more than 85%. In the field of diamond application tools, Hu et al. [29] proposed a Mask R-CNN model with the deep learning framework TensorFlow. Multiple targets in the surface image of a single fixed abrasive pad could be effectively segmented and recognized, and the average *ACC* of the main evaluation indicators reached 78.9%. Recently, Suo et al. [30] conducted research on the segmentation of abrasive grains and pores on the surface of diamond grinding pads based on an improved Mask R-CNN. The segmentation *ACC* of diamond abrasive grains and pores were 82.1% and 93.4%, respectively. In comparison with the pore segmentation accuracy, the main cause for the relatively lower segmentation accuracy of diamond abrasive grains lies in the significant difference in diamond abrasive grain wear and the extremely complex background texture. As far as the current research on diamond abrasive grain detection based on deep learning is concerned, only a small number of reports have conducted classification and recognition studies, and few relevant studies have been reported at the diamond segmentation level. This is mainly because diamond abrasive grains will undergo dual wear of the abrasive grains and binders during the machining process, resulting in extremely complex abrasive grain morphology and binder background. Due to the difficulty of collecting various images of diamond wear morphologies, it is difficult to form a large-scale dataset, which poses great challenges to the research on diamond classification and segmentation due to the limited data.

In summary, deep learning has a high flexibility, good generalization ability, and strong expression ability, but it relies on a large amount of training data. For small sample datasets, some innovative methods need to be used to reduce the model's dependence on the dataset and prevent overfitting, so as to ensure the application effect of small sample semantic segmentation in related fields. With the proposal of Transformer [31], the field of deep learning has witnessed explosive growth; models based on the multi-head self-attention mechanism have emerged in an endless stream, and their effects are all good. In 2020, Dosovitskiy et al. [32] proposed Vision Transformer, introducing Transformer into the CV field for the first time to perform image segmentation tasks; in 2021, Chen et al. [33] proposed TransUnet, replacing the Encoder part of Unet with Transformer, and achieved a *DSC* (Dice Similarity Coefficient) index of 77.48% on the Synapse dataset, injecting new power into small sample semantic segmentation.

Based on the above research status, it is difficult to solve the problems in achieving high-precision and high-efficiency classification and segmentation caused by reasons such as the complex background of abrasive grain images, diverse abrasive grain morphologies, and a small number of data samples by simply using classical models or large models. Based on the advantages that the Swin-Unet model is not prone to overfitting when training with small amounts of sample data and the advantage that the SAM large model can perform fine segmentation under prompt information, these models are suitable for solving the problems and challenges faced in the classification and segmentation of diamond abrasive grains. Therefore, based on the Swin-Unet model and the SAM large model, a classification and segmentation network for diamond abrasive grains has been designed. This network model has a more obvious improvement effect in terms of accuracy and robustness.

3. Proposed Method

3.1. Swin-Unet

Swin-Unet [34], shown in Figure 1, is a network structure that combines the Swin-Transformer [35] with a sliding window self-attention mechanism and U-Net [36] with up-sampling and down-sampling capabilities, and it has symmetrical encoders and decoders. Among them, the encoder contains Linear Embedding, Patch Emerging, and the Swin Transformer Block layers in the Swin-Transformer, and has the down-sampling ability to extract the feature information of different receptive fields; the decoder contains the Patch Expanding layer and has the up-sampling ability to restore the resolution of the feature map to the input resolution. Between the corresponding layers of the encoder and decoder, the same residual connection structure as U-Net is also retained to prevent gradient vanishing and increase the information scale. Different from traditional convolutional neural networks, it can capture richer contextual information in the image, thereby effectively extracting semantic features and improving segmentation accuracy. However, when dealing with complex scenarios, it is still insufficient to capture enough relevant information. When dealing with datasets with imbalanced categories, the segmentation performance of the minority classes is poor, alongside other problems.

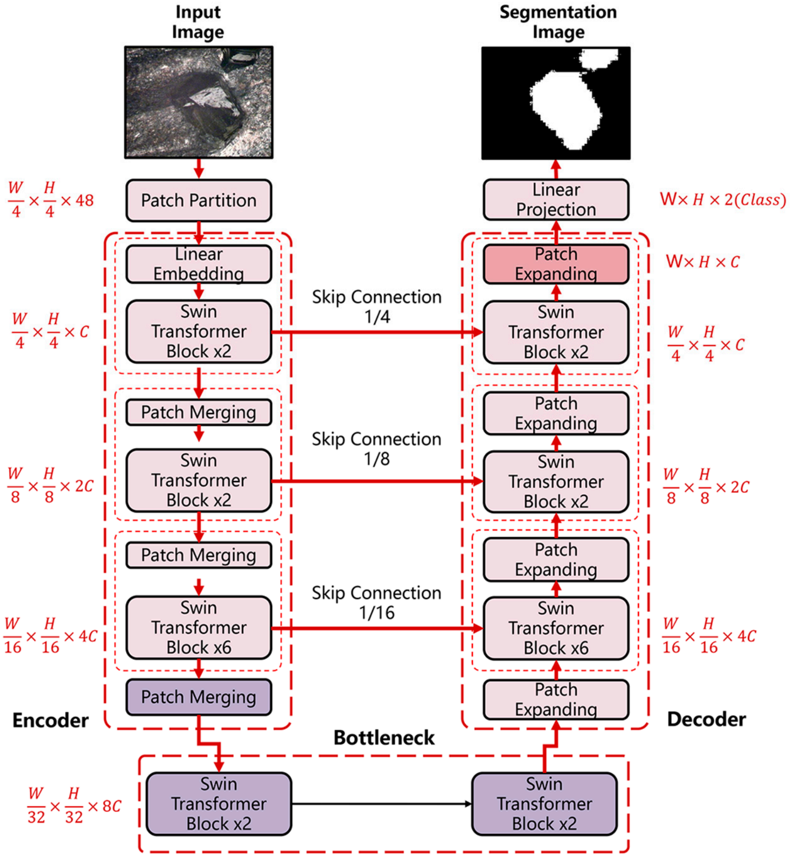
Input Image
Segmentation Image
$\frac{W}{4} \times \frac{H}{4} \times 48$
Patch Partition
Linear Projection
W× $H \times 2(Class)$
Linear Embedding
Patch Expanding
W× $H \times C$
Swin Transformer Block x2
Skip Connection 1/4
$\frac{W}{4} \times \frac{H}{4} \times C$
Swin Transformer Block x2
$\frac{W}{4} \times \frac{H}{4} \times C$
Patch Merging
Patch Expanding
Swin Transformer Block x2
Skip Connection 1/8
$\frac{W}{8} \times \frac{H}{8} \times 2C$
Swin Transformer Block x2
$\frac{W}{8} \times \frac{H}{8} \times 2C$
Patch Merging
Patch Expanding
Swin Transformer Block x6
Skip Connection 1/16
$\frac{W}{16} \times \frac{H}{16} \times 4C$
Swin Transformer Block x6
$\frac{W}{16} \times \frac{H}{16} \times 4C$
Patch Merging
Patch Expanding
Encoder
Bottleneck
Decoder
$\frac{W}{32} \times \frac{H}{32} \times 8C$
Swin Transformer Block x2
Swin Transformer Block x2

Figure 1. The Swin-Unet network structure.

3.2. SAM

The neural network structure of SAM [37] is a large image segmentation model proposed by Meta AI in 2023. The SAM structure mainly consists of three major parts: a powerful image encoder, a flexible prompt encoder, and an efficient Mask Decoder, as shown in Figure 2. The image encoder based on Vision Transformer can effectively capture the contextual information in an image and extract high-level features from the original image; the prompt encoder allows users to provide segmentation cues to the model through simple interactions such as clicking or drawing bounding boxes, which greatly improves the universality and practicality of the model. Additionally, the Mask Decoder combines the image features and user prompts to generate accurate segmentation masks. This structural design of SAM not only achieves efficient zero-shot learning ability, but also enables the model to easily adapt to various complex segmentation tasks. Through training on large-scale datasets, the SAM structure shows excellent performance and wide applicability. Deng et al. [38] used SAM to perform segmentation of tumor treatment completion, non-tumor tissue segmentation, and cell nucleus segmentation, but SAM failed to achieve satisfactory performance in dense instance object segmentation with 20 prompts per image. Ji et al. [39] explored that SAM performed well in various natural scenes, but in three types of concealed scenes, camouflaged animals, industrial defects, and medical lesions, SAM did not perform well and some segmentation failures occurred.

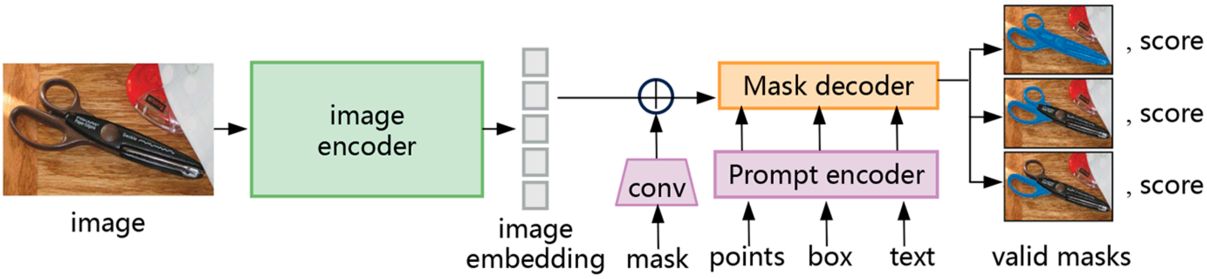

Figure 2. The SAM network structure [37].

SAM is a network structure trained with the extremely large dataset SA-1B containing 1.1 billion masks. It can perform segmentation on both trained and untrained target types and has a high degree of fitting for feature patterns. However, in professional fields with small sample datasets, such as medical images and industrial images, SAM cannot well understand the category and scope of the target, which will lead to huge segmentation errors due to multiple target categories.

3.3. Model Design

Affected by the workpiece materials and coolant in the processing, the quantity, distribution, size, morphology state of the abrasive grains on the surface of the diamond tool, and the morphology of the binder are constantly changing. In order to be able to solve the problem of abrasive grain detection and segmentation under such a complex background texture, the Swin-Unet-SAM network structure is designed, as shown in Figure 3, in which the red part is the improved and embedded structure. After the image is standardized, the Swin-Unet model, which has the advantage that it is not prone to overfitting when training with a small sample, is utilized to conduct targeted classification and coarse segmentation, and the design embedding is used to generate prompt information; then, the SAM large model is utilized to perform fine segmentation based on these information prompts to improve the segmentation accuracy. This model structure integrates the advantages of the small sample coarse segmentation of Swin-Unet and the prompt information fine segmentation of SAM, complementing each other and achieving high-precision image segmentation even with a small number of training samples. Specific improvements are as follows.

3.3.1. Improved Swin Transformer Block

After the standardized image undergoes Patch Partition and Linear Embedding expansion, it goes through three iterations of Patch Merging down-sampling and three times of Patch Expanding up-sampling. To better utilize the Swin-Unet network for the coarse segmentation of abrasive grain images, the original Swin Transformer Block with a low-level small window is modified to a high-level large-window structure to extract more abundant semantic information of the diamond morphology. For this, 2, 2, 12, 3, 12, 2, and 2 are used as the number of Swin Transformer Blocks in each module, and the Bottleneck structure of each is shown in Figure 4. The number of attention heads in the Swin Transformer Block is adjusted from 3, 6, 12, and 24 to 4, 8, 16, and 32. Different numbers of Swin Transformer Blocks are used for processing at four different scales, and the features of the same scale are concatenated. Finally, the coarse-segmented image is obtained using Linear Projection.

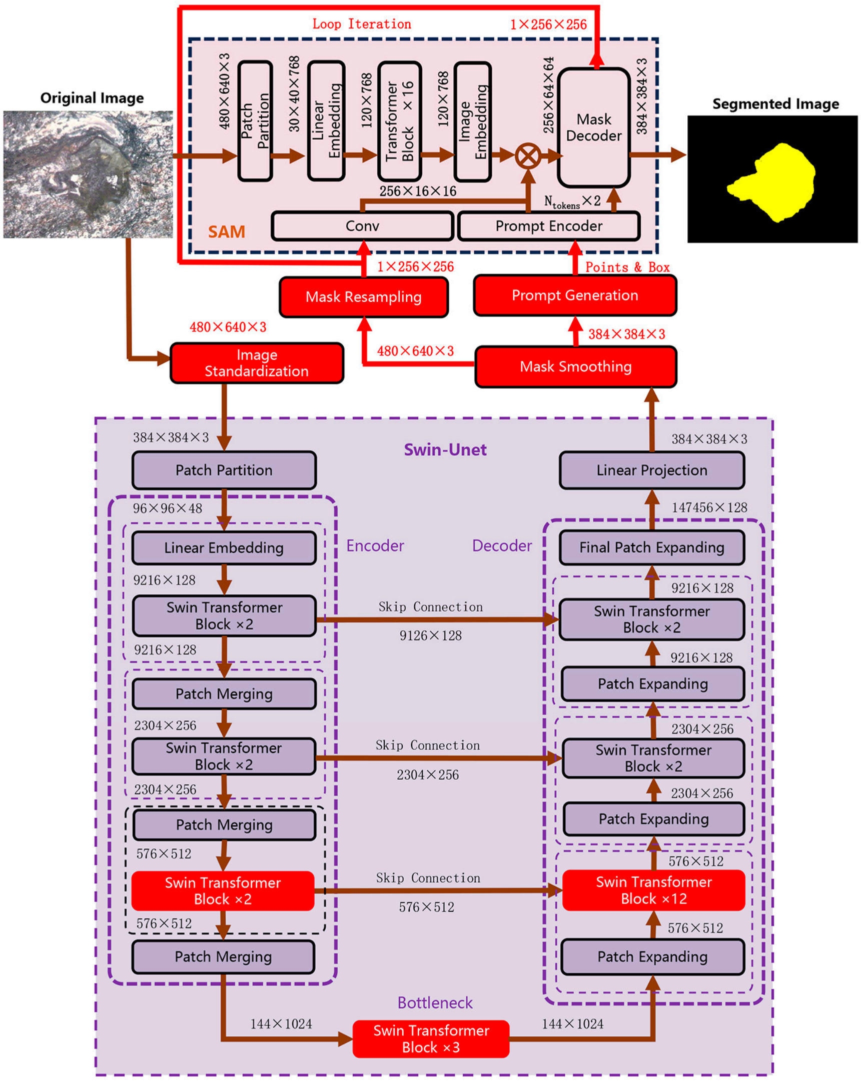

Figure 3. The architecture of proposed Swin-Unet-SAM network model.

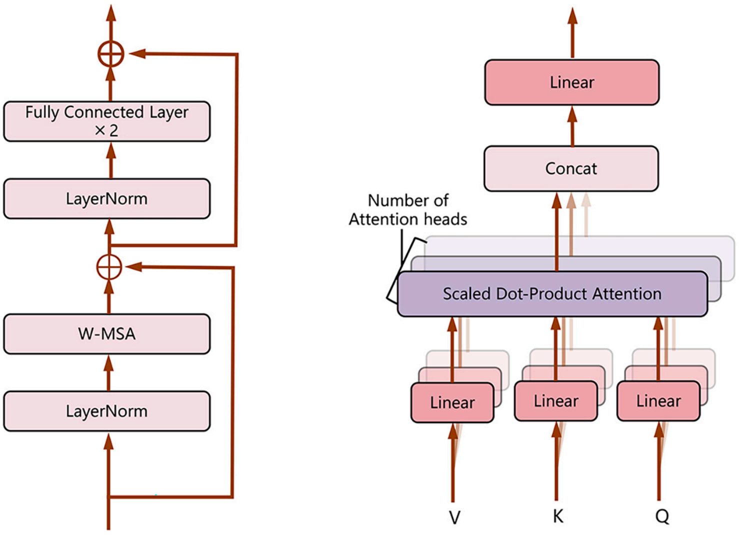
Fully Connected Layer × 2
LayerNorm
W-MSA
LayerNorm
Linear
Concat
Number of Attention heads
Scaled Dot-Product Attention
Linear
Linear
Linear
V
K
Q

Figure 4. The single Swin Transformer Block in the Bottleneck (**left**) and the meaning of the number of attention heads (**right**).

3.3.2. Improved SAM Structure

In the SAM structure, the abrasive grain image passes through the Image Encoder; that is, after the processing of the Patch Partition and Image Embedding, it is concatenated with the input Mask, and the prompt information is used in the Mask Decoder for operation to obtain the fine segmentation result. This article designs a function of loop iteration, which concatenates the result output from the Mask Decoder with the image processed by the Image Encoder, and then inputs it into the Mask Decoder for operation again. This multiple looping is to obtain a more accurate result as shown in Figure 5. The red line is the route of loop iteration. After a series of processing of the image and mask input, three masks of 1 × 256 × 256 and the corresponding score values are output from the top of the Mask Decoder, and the one with the highest score value is selected as the input mask. After convolution, it is concatenated with the processed image and then input into the Mask Decoder together with other types of prompt information in order to use the loop iteration function to achieve subsequent multiple loop operations and thereby obtain the fine segmentation result.

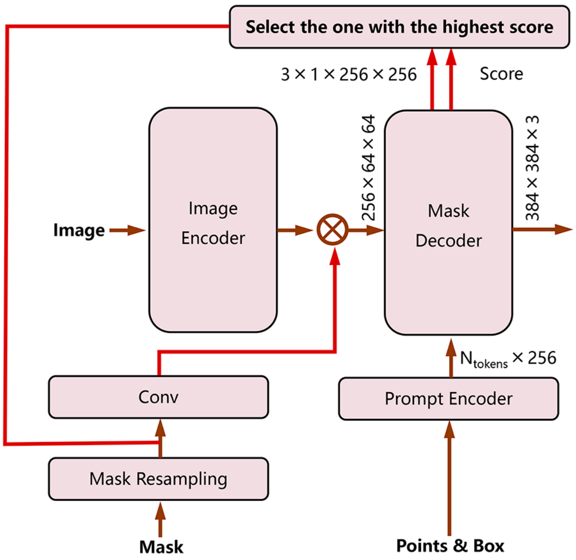

Figure 5. The structure of loop iteration.

3.3.3. Embedded Structure Design

In order to better connect the data coordination work between the two neural network structures of SAM and Swin-Unet, the following embedded structure is designed:

(1) Image Standardization

Considering that Swin-Unet mainly guarantees the accuracy of coarse segmentation and does not require overly high-resolution image information, And given that the resolution of the input image is relatively large, in order to maximize the transfer advantage of the pre-trained weights of the Swin-Transformer module, ensure that the attention window can cover sufficient image information, and at the same time avoid excessive consumption of computing resources, the size of the input image needs to be adjusted to a resolution that matches the size of the attention window. Compared with other image scaling methods, such as nearest neighbor interpolation and bilinear interpolation, cubic spline interpolation performs excellently in maintaining image smoothness and details. It can retain more image information during the scaling process.

Therefore, in order to make better use of the pre-trained weights and at the same time match the size of the attention window, cubic spline interpolation is performed on the input image and it is uniformly adjusted to 384 × 384 to retain image information as much as possible to avoid the occurrence of jagged phenomena.

(2) Mask Smoothing

Given that the original coarsely segmented image output by Swin-Unet alone has many phenomena such as sawtooth and noise points and an unsmooth edge, which will have a significant impact on the subsequent fine segmentation input into the SAM model, an appropriate smoothing operation needs to be performed on the original coarsely segmented image output by Swin-Unet. However, when additional convolutional layers are added and training parameters are modified, a satisfactory smoothing effect cannot be obtained. Considering the advantages of frequency-domain Gaussian filtering enhancement in smoothing images, this paper finally chooses to use a Gaussian low-pass filter [40] to smooth the mask image and uses the Canny operator to detect the edge and fill it to obtain a mask with a smooth edge, thereby achieving the edge smoothness of the original coarsely segmented image output by Swin-Unet. The transfer function formula of the Gaussian low-pass filter is as follows:

$$H(u,v) = e^{-\frac{D^2(u,v)}{2D_0^2}} \tag{1}$$

where $D(u,\ v)$ is the distance function in the frequency domain and D_0 is the cut-off frequency.

Mask Smoothing is used to smooth the image obtained by the coarse segmentation of Swin-Unet, filtering out small and useless targets. Since SAM cannot classify when using prompt information for segmentation, Mask Smoothing also undertakes the function of transmitting classification information. By designing Mask Resampling to resample the smoothed mask image, it is possible to input the mask as a kind of prompt information into SAM. In order to make full use of the Prompt Encoder of SAM and enable SAM to obtain as diverse inputs as possible, Prompt Generation is also designed to generate Point and Box prompt information.

(3) Mask Resampling

Since the pre-trained model SAM used in this paper has an image size format that does not match the image size format output by Swin-Unet, it is necessary to resample the coarsely segmented image output by Swin-Unet to adapt to the SAM model and achieve splicing in SAM. For the SAM model, its training image size format is 1 × 256 × 256, where 1 is the predicted confidence value and 256 × 256 is the image size.

To achieve this goal, first, the mask read from the output of Swin-Unet is converted from a 384 × 384 × 3 color image to a 384 × 384 × 1 grayscale image. Then, binarization is used to set the background pixels to 0 and the target pixels to 1, and the feature dimension is shifted to become 1 × 384 × 384. The mask is a binary image and needs to remain binary after scaling to be spliced in SAM. Nearest neighbor interpolation is used to scale the image to 1 × 256 × 256, thereby providing better mask data for the fine segmentation of SAM.

(4) Prompt Generation

In the SAM large model, the execution of its segmentation task depends on the prompt indicator. The quality of the prompt information provided by the prompt indicator will directly affect the segmentation performance of the SAM large model. Considering that diamond abrasive grains have multiple types of characteristics, in order to improve the segmentation accuracy of the SAM large model in multi-classification, this paper takes the Swin-Unet coarse segmentation mask separated by class as the prompt information, and also uses Points and Box to provide more prompt information for the segmentation target. Among them, Points provides segmentation clues for SAM segmentation by specifying the feature points of the diamond abrasive grain segmentation target. In order to improve the quality of the feature information provided by Points, this paper fine-tunes Points. It uses over-erosion and dilation, Canny operator, and the Bradenham algorithm to accurately calculate the central position feature information of the target diamond abrasive grain, thus forming prompt information containing the edge and center of the diamond abrasive grain. Box can use its flexibility to provide a suitable size and shape of Box for the segmentation target, providing space for the SAM model for fine segmentation. In order to improve the quality of the information provided by Box, this paper conducts area detection on the Swin-Unet coarse segmentation mask to enable Box to obtain a more accurate target area of diamond abrasive grains.

Through the above design improvements, the combined use of masking and Points and Box provides richer inputs for SAM, thereby achieving the purpose of accurate segmentation using SAM.

3.4. Loss Function

In the diamond abrasive grain images, abrasive grains are usually classified into four types: whole, micro-broken, macro-broken, and fall-off [41]. Due to the randomness and complexity in the processing of diamond tools, there is a significant difference in the number of abrasive grains among these types. Therefore, there exists a serious issue of unbalanced positive and negative samples in the collected image sample data of diamond abrasive grains. Additionally, due to the size variations of the diamond itself and the different depths of the binder coating, the abrasive grain sizes in the collected diamond abrasive grain images also have a large difference, with small target abrasive grains present in the images. In view of this, a CE-Dice loss [42] is adopted in the present paper. This loss is the combination of the Dice loss [43] and the CE (cross-entropy) loss [44]. The Dice loss is used for a global perspective investigation in scenarios with unbalanced positive and negative samples, while the CE loss is used for a micro-perspective investigation at the pixel level. By combining Dice loss and CE loss, Dice Loss is employed to calculate the disparity in Dice coefficients between the predicted segmentation result and the true segmentation result to guide model training. This enables the model to pay greater attention to the excavation of the foreground area and automatically balance the weights of different categories. Meanwhile, CE loss is utilized to calculate the results of different categories at the pixel level, effectively addressing the problem of the model's reduced ability to recognize minority categories caused by unbalanced positive and negative samples. In brief, the advantage of CE-Dice loss lies in its ability to flexibly adjust the weights on different tasks and datasets to achieve the optimal segmentation effect. The CE-Dice loss function adopted in this paper is as follows:

$$Loss = L_{dice} \times w_f + L_{cross} \times \left(1 - w_f\right) \tag{2}$$

$$L_{dice} = 1 - \frac{2\sum\limits_{i=1}^{n} p_i g_i}{\sum\limits_{i=1}^{n} p_i + \sum\limits_{i=1}^{n} g_i + \varepsilon} \tag{3}$$

$$L_{cross} = -\sum_{i=1}^{n} p_i \log(q_i) \tag{4}$$

where L_{dice} and L_{cross} denote Dice loss and Cross-Entropy loss, respectively. w_f is the weight value that is used to adjust the proportion between the Dice loss and cross-entropy loss. n represents the quantity of pixels in the diamond image, p represents the probability of the true label, q represents the probability of the predicted value, and ε represents the initiate term that is set to prevent the denominator from being zero.

In order to obtain the appropriate weight value w_f to improve the performance and stability of the model, ablation experiments were conducted on different weight values ranged from zero to one, and in this article, it was finally determined that $w_f = 0.75$.

4. Experiments

In this section, we detail the dataset experiments, the performance of the proposed method on different datasets, and the effectiveness of the proposed method, which was verified by a series of comparisons and verification ablation experiments through designing.

Corresponding experiments were conducted on a CPU AMD EPYC 7543, GPU NVIDIA A40 (40 GB), and the Pytorch (1.11.0) platform with Python 3.8.

4.1. Dataset Acquisition

Datasets are the foundation for training deep learning models. The quality of a dataset will directly have a profound impact on the accuracy and generalization ability of the model. In this paper, the image acquisition device Hirox KH-1000 (HIROX, Tokyo, Japan) video microscope system produced by HIROX Company in Japan is used for collection. This device has a magnification range of 50 times to 1000 times. The collected data are diamond tools at different wear stages. The diameters of the diamond abrasive grains on the tool surfaces are between 425 μm and 600 μm. The resolution of the collected images is 640 × 480. In order to construct a high-quality dataset, by tracking the changes in the wear state of diamond abrasive during different stages of processing, 17 groups of diamond abrasive grain areas on the tool surfaces are designed and collected. Each area contains about 137 images, and the entire process from whole, micro-broken, and macro-broken to fall-off is covered. By eliminating invalid and duplicate data, the original instance diamond abrasive grain images are finally obtained.

4.2. Dataset

The datasets of diamond abrasive grains in four categories, namely the whole dataset, micro-broken dataset, macro-broken dataset, and fall-off dataset, will be employed to verify the effectiveness of the proposed model. The representative images of different diamond wear categories are presented in Figure 6.

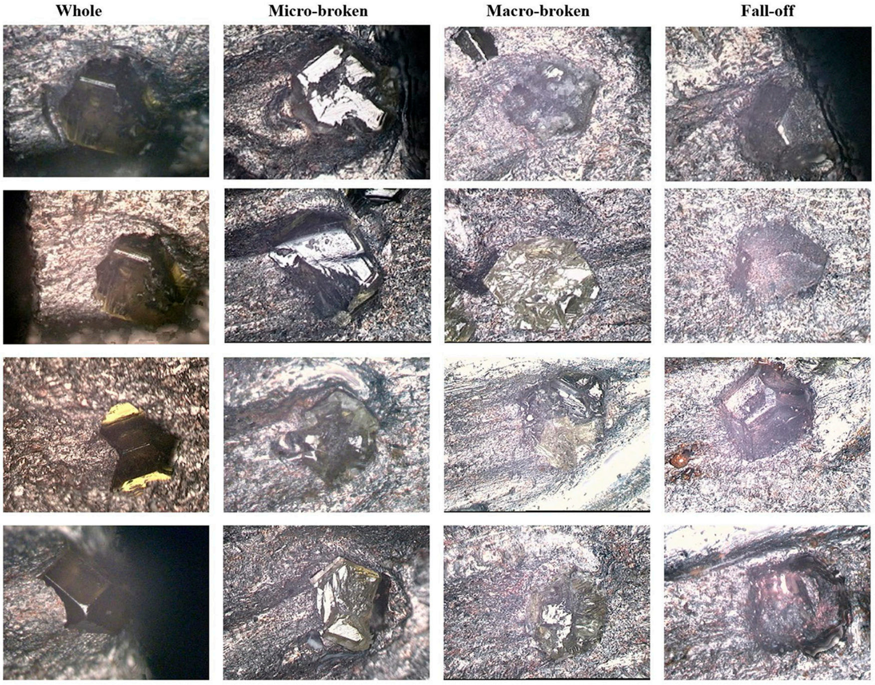

Figure 6. Typical abrasive grain categories in the abrasive grain dataset. Explanatory descriptions of abrasive grains in the dataset: (**first column**), whole category; (**second column**), micro-broken category; (**third column**), macro-broken category; (**fourth column**), fall-off category.

The annotation of the dataset is carried out in JSON format using Labelme. The storage data retrieval format includes information such as the name, information, size, mask type, and endpoint coordinates of the image. It should be noted that each sample in the diamond abrasive grain dataset contains one or more abrasive grains and category information. The obtained annotation results are shown in Figure 7. In the figure, the green mark indicates the whole category; the yellow mark indicates the micro-broken category; the red mark indicates the macro-broken category; and the white mark indicates the fall-off category. To ensure a balance between model training and performance evaluation, the dataset is divided into a training set and a testing set in a division ratio of 7:3. An overview of the key characteristics of the diamond abrasive grain wear category dataset is shown in Table 1.

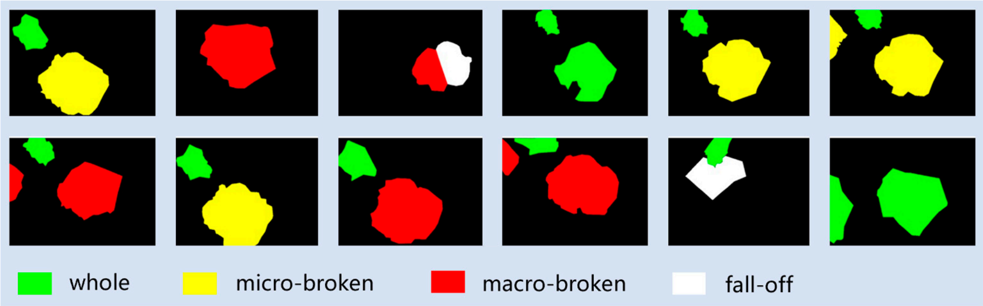

Figure 7. Typical annotation results of different categories of abrasive grains in dataset.

Table 1. Distribution of labeled samples.

Dataset	Total Labels	Total Images	Resolution	Division ratio
Whole	1642	657	640 × 480	7:3
Micro-broken	969	421	640 × 480	7:3
Macro-broken	1018	503	640 × 480	7:3
Fall-off	188	124	640 × 480	7:3

Note: An online enhancement method was utilized to expand the samples, which includes geometric transformation, noise processing, linear transformation, random color jitter, etc.

4.3. Evaluation Index

In order to better quantitatively evaluate the performance of the proposed model, this paper introduces four evaluation indexes: *ACC*, *SE* (Sensitivity), *DSC*, and *HD*95 (the 95th Percentile Hausdorff Distance). *ACC* represents the proportion of correctly predicted sample categories in the classification process, *SE* represents the proportion of positive samples classified correctly among the true positive samples, *DSC* represents the similarity between the predicted segmentation area and the true segmentation area, and *HD*95 represents the number of pixels of the boundary error between the predicted segmentation result and the true label segmentation result. The specific formulas are as follows:

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

$$SE = \frac{TP}{TP + FN} \tag{6}$$

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{7}$$

$$HD95 = \max_{t2}\left(\max_{t1}\left(\sqrt{T^2 - P^2}\right)\right) \times 0.95 \tag{8}$$

where *TP* (True Positive) is the number of pixels correctly classified as target pixels; *FP* (False Positive) is the number of pixels that are classified as the segmentation target but are actually the background; *TN* (True Negative) represents the number of background point pixels correctly classified; *FN* (False Negative) represents the number of pixels that actually belong to the segmentation target but are classified as the background; *T* represents the boundary of the true label segmentation, *P* represents the boundary of the segmentation area predicted by the model, *t*1 and *t*2 are both boundary pixels, *t*1 belongs to *P*, and *t*2 belongs to *T*.

5. Results and Discussion

5.1. Performance Comparison

To verify the superiority of the model proposed in this paper, the overall performance of the proposed model is compared with the original Unet, TransUnet, Swin-Unet, and other networks in the task of abrasive grain image segmentation. The hyperparameters for the model training process are chosen as follows: conduct uniform training for 2500 epochs. Each epoch is divided into five batches with a batch size of 24. A path dropout rate of 0.1 is uniformly employed to randomly discard neurons in the network to prevent overfitting. To enable the model to learn more detailed information about abrasive grains, the initial learning rate for training is set at 0.001 and the decay exponent is 0.9. This ensures that in the early stage of training, the model approaches the optimal solution more rapidly with a larger learning rate, and in the later stage, it adjusts parameters more precisely with a smaller learning rate and converges more accurately to the optimal solution. Meanwhile, to make the parameter update smoother and reduce computational cost, the Stochastic Gradient Descent with a momentum optimizer is utilized. In the optimizer, the momentum parameter is set to 0.9, the momentum damping term is set to 0.05, and the weight decay coefficient is set to 0.0001 for L2 regularization of the parameters to prevent overfitting. Based on the training and testing experiments, the time consumption in training and testing is presented in Table 2.

Table 2. Time consumption of different methods in training and testing.

Time Consumption	Unet	TransUet	Swin-Unet	Swin-T	Swin-S	Swin-B	DINOv2	Swin-Unet-SAM (Ours)
Training (h)	3.20	7.82	4.17	4.05	4.82	7.20	0.86	6.62
Testing (ms)	535	422	201	198	234	423	1412	2425

Note: The training time is the time taken for 2000 epochs where the method reaches stable results. The testing time is the average time consumed for testing each image.

As can be seen from Table 2, after 2000 epochs of training, almost all different models require several hours in the training stage, which is a fairly normal time consumption in deep learning. It should be noted that the training time of the DINOv2 model only calculates the training time consumption of the task head of the pre-trained model, not the complete training time consumption of the DINOv2 model. This is also the reason why it seems that the training time consumption of the DINOv2 model is significantly shorter than that of other models. Based on the training time consumption of each model, although the time consumption of Swin-Unet has increased compared to Unet, it is relatively less compared to TransUnet, Swin-S, and Swin-B. Our Swin-Unet-SAM model takes about 2.45 h more than using Swin-Unet alone. This is because the Imagecoder module in the SAM large model is time-consuming and needs to fine-tune the coarse segmentation result of Swin-Unet to achieve fine segmentation. In terms of the average time consumption for testing a single sample, models such as Unet, TransUnet, Swin-Unet, Swin-T, Swin-S, and Swin-B take about several hundred milliseconds, while the time consumption of the DINOv2 model and our Swin-Unet-SAM model is about 1.4 s and 2.4 s, respectively. The testing time is significantly longer than that of models such as Unet, TransUnet, Swin-Unet, and other models. Considering that the abrasive grain wear detection of diamond tools requires in situ or even offline detection, the real-time requirement for detection is not high. In addition, in industrial detection, more attention is paid to the accuracy and segmentation accuracy of diamond abrasive grain wear classification and segmentation. Therefore, in this context, the time consumption of the model proposed in this paper is acceptable.

Table 3 shows the evaluation performance of the proposed model and other models on the four types of diamond abrasive grain datasets. Judging from the values of various indices, the *ACC* index values and *SE* index values of the traditional Unet network structure for the classification of abrasive grains with different degrees of wear are generally low, indicating that the accuracy of its classification results is not high; while the lower *DSC* index values and the higher *HD95* index values reflect the poor segmentation effect. Compared with Unet, TransUnet has certain improvements in the performance of indices such as *ACC*, *SE*, *DSC*, and *HD95*. This is mainly due to the addition of a multi-head attention mechanism and the Unet encoder-decoder architecture in the TransUnet network structure, thereby enhancing the model's performance in complex image segmentation tasks. Compared with TransUnet, the performance of indices such as the *ACC*, *SE*, and *HD95* of the Swin series models has been improved to different degrees. This is because the Swin series models completely remove the CNN encoder and instead use a pure Transformer structure, giving full play to the advantages of the Transformer. However, the performance of the Swin series models in terms of the *DSC* index is not good, indicating that the segmentation effect has degraded. Comparing the DINOv2 model with the Swin series models, the performance of indices such as *ACC*, *SE*, *DSC*, and *HD95* has decreased to different degrees. In particular, the *HD95* index has increased significantly, indicating that the degradation of the segmentation effect is relatively serious. By comparing the Swin-Unet-SAM model proposed in this paper with other models, it can be found that although the *HD95* index value is not the best for two types of abrasive grains, macro-broken and fall-off, compared with the results of Unet, Swin-Unet, etc., it can be seen that the *HD95* index performance of the Swin-Unet-SAM model still has a significant improvement. It is worth mentioning that the abrasive grains that play the main cutting role in actual processing are the two types of the whole and micro-broken abrasive grains, while the two types of abrasive grains of macro-broken and fall-off have mostly lost or even completely lost their cutting function. Therefore, the wear types of the whole and micro-broken abrasive grains are the focus of attention, and having a relatively low *HD95* index value for these two types of abrasive grains is acceptable. Hence, the overall results of various evaluation indices indicate that the model proposed in this paper has a high robustness and good segmentation effect.

Table 3. Results of different methods compared with the results of the proposed method on four different datasets.

Model	**Whole**				**Micro-Broken**				**Macro-Broken**				**Fall-Off**			
	ACC	***SE***	***DSC***	***HD95***	***ACC***	***SE***	***DSC***	***HD95***	***ACC***	***SE***	***DSC***	***HD95***	***ACC***	***SE***	***DSC***	***HD95***
Unet [36]	0.9297	0.1985	0.3757	111.82	0.9488	0.4961	0.6349	83.94	0.9459	0.2409	0.3546	76.26	0.9713	0.5126	0.4159	88.02
TransUet [33]	0.9513	0.6586	0.6274	73.34	0.9470	0.6981	0.7181	45.89	0.9842	0.8314	0.8635	22.37	0.9909	0.5374	0.6802	19.97
Swin-Unet [34]	0.9728	0.6859	0.7773	72.23	0.9602	0.7804	0.7788	44.46	0.9745	0.6631	0.7382	89.96	0.9928	0.9350	0.8378	31.25
Swin-T [31]	0.9712	0.7878	0.7730	43.27	0.9623	0.8233	0.7843	39.45	0.9815	0.9063	0.8327	76.39	0.9958	0.8552	0.8871	31.95
Swin-S [31]	0.9650	0.7117	0.7235	86.45	0.9683	0.7117	0.7235	24.63	0.9746	0.9168	0.7492	47.45	0.9715	0.7611	0.7504	43.99
Swin-B [31]	0.9695	0.9168	0.6966	56.58	0.9536	0.6886	0.7582	31.73	0.9753	0.6469	0.7582	112.43	0.9931	0.7518	0.8275	37.06
DINOv2 [45]	0.9361	0.5331	0.5684	164.37	0.9483	0.4931	0.6316	98.04	0.9594	0.2627	0.4118	172.23	0.9669	0.8184	0.4956	85.33
Swin-Unet-SAM (Ours)	0.9864	0.9045	0.8909	36.49	0.9840	0.8629	0.9064	26.11	0.9856	0.9824	0.8491	39.26	0.9898	0.9515	0.7883	25.62

Note: A higher value is better for *ACC*, *SE*, and *DSC*. However, for *HD*95, a lower value is better. The results tell that the model of ours has overall better performance than other models in terms of *ACC*, *SE*, *DSC*, and *HD*95 under four different datasets that cover various wear degrees of abrasive grains.

According to Table 3, it can be found that under the small sample dataset with complex background textures, the average index values of *ACC*, *SE*, and *DSC* for the classification and segmentation of diamond abrasive grains have reached 98.7%, 92.5%, and 85.9%, respectively, which have increased by 1.2%, 15.9%, and 7.6%, respectively, compared to the model of Swin-Unet before improvement, which proves the effectiveness of the method proposed in this paper. Currently, there are extremely few studies on applying deep learning models for diamond abrasive grain segmentation. The segmentation *ACC* of related research generally remains at around 80% [28–30], which is not high. The main cause is that the morphology and size of abrasive grains vary significantly after wear, and the complex background texture also poses a great challenge to accurate segmentation. By employing the method proposed in this paper, the segmentation *ACC* of diamond abrasive grains is increased to 98.7%, which represents a substantial improvement compared to previous ones. This further substantiates that the proposed model has excellent segmentation performance and robustness.

Figure 8 shows the actual classification and segmentation results of the model proposed in this paper and other models on the four types of diamond datasets. Judging from the various segmentation results, the Unet network has over-segmentation and classification errors for the four types of abrasive grain wear classifications during the classification process. The main reason is that the diamond abrasive grains are relatively close to the background, and the up-sampling of the Unet network cannot extract the features of abrasive grains with different degrees of wear precisely, resulting in poor performance and a large gap between the boundary part and the true label. The segmentation effect of the TransUnet network structure is relatively smooth. By extracting features in a way that reduces the segmentation target, a large amount of edge information will be lost, and there are also some classification errors. Compared with the Unet network structure and the TransUnet network structure, the network structure of the Swin series performs relatively well in the classification of abrasive grain wear, but there are still a small number of classification errors. Compared with the true label, the edge information of the segmentation result is inaccurate and there are obvious missing parts, which is the reason for the improvement in this paper. On the latest open-source visual large model DINOv2, its segmentation and classification effects are relatively unsatisfactory, with problems such as classification errors and a small amount of target missing. In view of the problems such as classification errors and poor segmentation effects of the above models, in this paper, in the Swin-Unet, the number of Swin Transformer Blocks is improved from the original low-level small-window Swin Transformer Blocks to the high-level large-window structure to extract complex and rich edge semantic information, solving the problem that it is difficult to distinguish the information of diamond abrasive grains close to the complex background, and achieving accurate classification of abrasive grain wear. The embedded structure is designed to achieve the purpose of smoothing and filtering useless information, making the segmentation boundary smoother while ensuring the integrity of edge information, obtaining the coarse segmentation result. Combined with the improved fine-tuning and iterative loop of the SAM model to obtain the optimal result, the fine-segmented abrasive grain target is obtained.

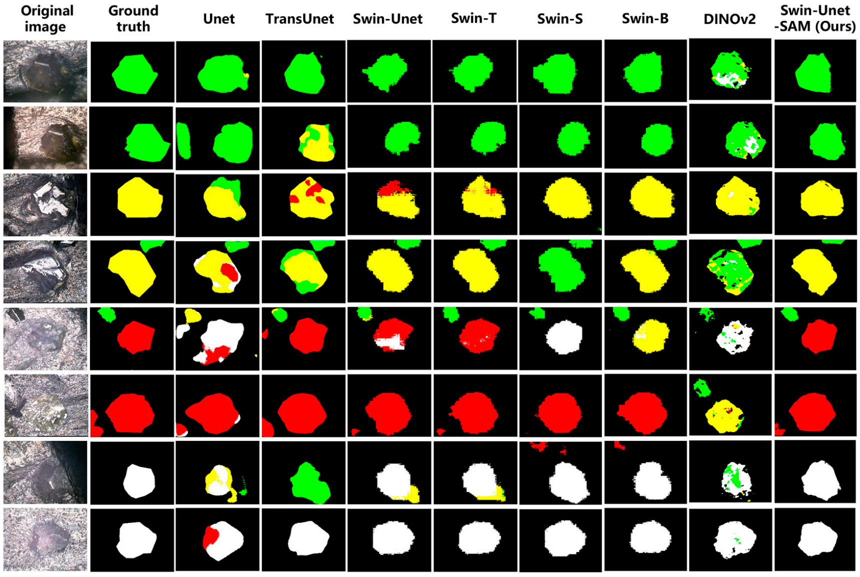

Figure 8. The proposed method compared with other state of the art methods. Analysis of detection results: (**first column**), original image; (**second column**), ground truth; (**third column**), Unet detection

results; (**fourth column**), TransUnet detection results; (**fifth column**), Swin-Unet detection results; (**sixth** to **eighth column**), Swin series detection results; (**ninth column**), detection results of the proposed method in this paper. In the figure, the green mark is for the whole category, the yellow mark for the micro-broken category, the red mark for the macro-broken category, and the white mark for the fall-off category. The results indicate that our method is superior to others in classifying and segmenting abrasive grains with different degrees of wear under a complex background.

Based on the analysis and discussion of Table 3 and Figure 9, the model in this paper is biased towards higher values of the three indicators *ACC*, *SE*, and *DSC*, which are significantly higher than or not weaker than other models. This reflects that the predicted categories of abrasive grains with different wear conditions are consistent with the true labels and can be classified correctly. It has a highly accurate segmentation result in target segmentation, and its coincidence degree with the label is also high. This confirms that the method proposed in this paper has good classification, segmentation performance, and robustness, and is suitable for the detection of multi-class wear targets with complex texture backgrounds.

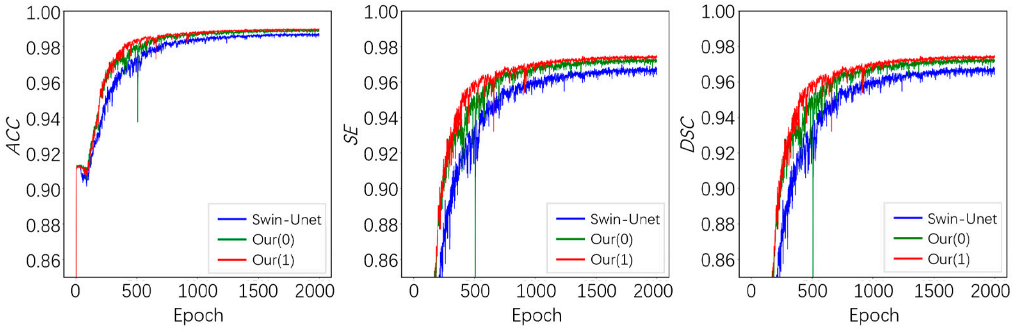

Figure 9. Evaluation of different models. Swin-Unet refers to the original Swin-Unet; Our(0) represents our proposed model with the improved structure without the CE-Dice loss; Our(1) refers to the proposed model with the improved structure without the CE-Dice loss. The results show that our model outperforms Swin-Unet in terms of *ACC*, *SE*, and *DSC*. This is because after adding relevant structures and loss function, the model has a faster convergence speed and more abundant learning ability of wear features.

5.2. Ablation Experiments

To better illustrate the superiority of the proposed model, this paper conducts the following ablation experiments in four different abrasive grain wear datasets to verify the role of the improved Swin-Unet structure, the CE-Dice loss function, and the improved SAM loop structure. After adding the improved Swin-Unet structure (Our(0)), comparing the blue and green curves in Figure 9, it can be seen that the training convergence speed of the improved model with the number of Swin Transformer Blocks becomes faster and the model learns more abundant wear feature information. The *ACC* and *SE* index values of abrasive particle wear classification have both significantly improved, and the *DSC* index value reflecting the abrasive grain segmentation effect performs well. For the model after using the CE-Dice loss function (Our(1)), by comparing the green and red curves, it can be found that the CE-Dice loss function makes the model converge faster and focus more on the feature analysis of the target abrasive grains, which has a very obvious advantage in the segmentation of worn abrasive grains.

Table 4 shows the index evaluation results of the four indicators *ACC*, *SE*, *DSC*, and *HD95* for abrasive grains with four different degrees of wear (whole, micro-broken, macro-broken, and fall-off). It can be seen from Table 4 that compared to Swin-Unet, the improved Swin-Unet structure Our(0), the CE-Dice loss function Our(1), and the addition of the SAM structure Our(2) have all achieved significant improvements in the performance of each indicator. This is because the output result of Our(1) is used as an input (mask), and the prompts of Points, Box, and mask, as shown in Figure 10, are input into the improved SAM large model. The fine-tuning function of the SAM model is utilized to achieve the rapid positioning of abrasive grain targets and regions and achieve fine segmentation, thereby improving the model performance.

Table 4. Ablation experiment results for three methods on the dataset.

Model	*ACC*	*SE*	*DSC*	*HD95*
Swin-Unet	0.9751	0.7382	0.7754	61.08
Our(0)	0.9791	0.8263	0.8192	33.90
Our(1)	0.9841	0.8762	0.8614	40.27
Our(2)	0.9865	0.9334	0.8767	39.69

Note: A higher value is better for *ACC*, *SE*, and *DSC*. However, for *HD95*, a lower value is better.

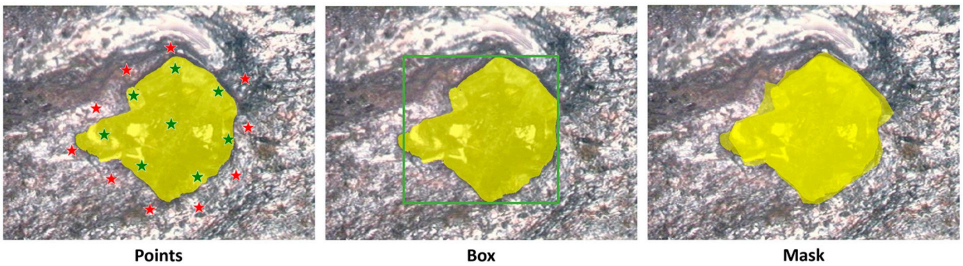

Figure 10. Prompts of the points, box and mask.

6. Conclusions

This paper proposes an effective Swin-Unet-SAM neural network for diamond abrasive grain image detection. Specifically, the proposed model addresses the issues of abrasive grain wear classification and segmentation. The proposed method can be roughly divided into three stages. Firstly, the Swin-Unet network is improved by adjusting the Swin Transformer Block, aiming to obtain richer semantic information for abrasive grain classification and coarse segmentation. In the second stage of the network, the SAM model structure is improved by using the output result of Swin-Unet as mask information to enhance the richness of network information and increase the loop iteration to obtain higher-precision fine segmentation results. Finally, four embedded structures are designed to combine the two models to achieve the input of abrasive grain images and the output of classification and segmentation. Additionally, we adopt the CE-Dice loss to improve the convergence speed and detection accuracy of the model. Extensive experiments have been conducted on four types of abrasive grain wear datasets to evaluate the effectiveness of our model. The proposed method not only significantly solves the rapid classification and segmentation tasks of diamond tool abrasive grains but also has good performance and potential application prospects in the industrial field.

In the future, we plan to further optimize the embedded structure and design an intermediate layer to achieve rapid model connection, thereby accelerating our model to use fewer resources. Regarding the classification and segmentation results, we will further analyze the wear characteristics of abrasive grains to provide more in-depth guidance for the processing process.

Author Contributions: Conceptualization and methodology, Y.L. and C.F.; software and validation, Y.L., T.F. and C.F.; resources and data collection, C.F.; writing—original draft preparation, Y.L. and T.F.; writing—review and editing, C.F.; project administration, Y.L.; funding acquisition, Y.L. and C.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the projects from Xiamen Natural Science Foundation (Grant number 3502Z20227330), the National Natural Science Foundation of China (Grant number 52275426), and the research innovation team of Intelligent Image Processing and Application at Xiamen Institute of Technology (Grant number KYTD202101).

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to them containing information that could compromise research participant privacy/consent.

Acknowledgments: We thank the editor and anonymous reviewers for their valuable suggestions to improve the quality of this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lavrinenko, V.I.; Poltoratskyi, V.G.; Pasichnyi, O.O.; Solod, V.Y. Using diamond grinding powders with combined coatings on diamond grain surfaces in abrasive tools. *J. Superhard Mater.* **2024**, *46*, 238–243. [CrossRef]
2. Li, S.; Cai, W.; Wang, K.; Ou, M.; Li, J.; Xiao, Q.; Yang, Z. Modeling and validation of the grinding morphology of ordered diamond grinding wheel. *Int. J. Adv. Manuf. Technol.* **2024**, *132*, 3267–3284. [CrossRef]
3. Hong, Q.; Zhou, R.; Guo, X.; Wang, Z.; Yin, S. A novel strategy for improving the wear resistance of electrodeposited Ni-diamond composite coatings by diamond surface morphology modification. *Diam. Relat. Mater.* **2023**, *137*, 110093. [CrossRef]
4. Wang, K.; Zhang, J.; Kang, J.; Zhang, H. Analysis of diamond wear morphology and segment wear evolution during the process of hard granite sawing. *Int. J. Refract. Met. Hard Mater.* **2023**, *110*, 106040. [CrossRef]
5. Buyuksagis, J.Y.S. Development of models for estimating specific energy and specific wear rate of circular diamond saw blades based on properties of carbonate rocks. *Int. J. Rock Mech. Min.* **2020**, *135*, 104497. [CrossRef]
6. Zhou, L.; Cui, C.; Huang, C.; Huang, H.; Ye, R. Grain edge detection of diamond grinding wheel. *Proc. SPIE* **2013**, *8759*, 378–384.
7. Xue, W.; Zhao, C.; Fu, W.; Du, J.; Yao, Y. Micro Vision-based Sharpening Quality Detection of Diamond Tools. In Proceedings of the International Conference on Intelligent Robotics and Applications, Harbin, China, 1–3 August 2022; Springer: Cham, Switzerland, 2022.
8. Chen, J.; Cui, C.; Huang, G.; Huang, H.; Xu, X. A new strategy for measuring the particles height uniformity of a grinding wheel. *Measurement* **2020**, *151*, 107250. [CrossRef]
9. Tang, J.; Qiu, Z.; Li, T. A novel measurement method and application for grinding wheel surface topography based on shape from focus. *Measurement* **2019**, *133*, 495–507. [CrossRef]
10. Deng, Z.H.; Zhao, X.Y.; Liu, W.; Wang, L.L. Modeling and Measurement of Grinding Wheel Based on Spherical Polyhedron and Light Density. *J. Mech. Eng.* **2016**, *52*, 190–197. [CrossRef]
11. Tabernik, D.; Šela, S.; Skvaŕc, J.; Skŏcaj, D. Segmentation-based deep-learning approach for surface-defect detection. *J. Intell. Manuf.* **2020**, *31*, 759–776. [CrossRef]
12. Huang, Y.; Jing, J.; Wang, Z. Fabric defect segmentation method based on deep learning. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–15. [CrossRef]
13. Hemamalini, S.; Sharma, O.P. Hybrid structures-based face recognition method using artificial neural network. *Int. J. Appl. Res. Inf. Technol. Comput.* **2022**, *13*, 12–23. [CrossRef]
14. Chen, Z.; Yang, J.; Feng, Z.; Zhu, H. RailFOD23: A dataset for foreign object detection on railroad transmission lines. *Sci. Data* **2024**, *11*, 72. [CrossRef]
15. You, K.; Liu, H. Feature detection of mineral zoning in spiral slope flow under complex conditions based on improved YOLOv5 algorithm. *Phys. Scr.* **2024**, *99*, 016001. [CrossRef]
16. Yu, S.Q.; Dai, X.J. Wear particle image segmentation method based on the recognition of background color. *Tribology* **2007**, *27*, 467–471.
17. Wu, W.Y.; Cui, C.C.; Ye, R.F.; Zhang, Y.Z.; Yu, Q. Image segmentation method using second time gray level histogram of connected component labeling of grinding wheel abrasives grains. *J. Huaqiao Univ.* **2016**, *37*, 422–426.
18. Gong, J.F.; Xu, X.P. A contour extraction of abrasive grain in diamond tools. *Tool Eng.* **2007**, *41*, 44–47.
19. Lin, Y.F.; Fang, C.F. Study on the segmentation of abrasive grains in diamond tools. *Int. J. Abras. Technol.* **2018**, *8*, 203–217. [CrossRef]
20. Lin, Y.F.; Wu, L.X. Improved abrasive image segmentation method based on bit-plane and morphological reconstruction. *Multimed. Tools Appl.* **2019**, *78*, 29197–29210. [CrossRef]
21. Su, L.L.; Huang, H.; Xu, X.P. Quantitative measurement of grit distribution of diamond abrasive tools. *China Mech. Eng.* **2014**, *25*, 1290–1294.
22. Li, H.Y.; Fang, C.F. Segmentation and evaluation of diamond abrasive grains based on K-Means clustering and convex hull detection. *Diam. Abras. Eng.* **2023**, *43*, 188–195.
23. Wang, S.; Wu, T.H.; Shao, T.; Peng, Z. Integrated model of BP neural network and CNN algorithm for automatic wear debris classification. *Wear* **2019**, *426*, 1761–1770. [CrossRef]
24. Peng, Y.; Cai, J.; Wu, T.; Gao, G.; Peng, Z. A hybrid convolutional neural network for intelligent wear particle classification. *Tribol. Int.* **2019**, *138*, 166–173. [CrossRef]

25. Zhang, T.; Hu, J.; Fan, S.; Yu, Y. CDCNN: A model based on class center vectors and distance comparison for wear particle recognition. *IEEE Access* **2020**, *8*, 113262–113270. [CrossRef]
26. Yang, Z.H.; He, S.Z.; Feng, W.; Li, Q.Q.; He, W.C. Intelligent identification of wear particles based on Mask R-CNN network and application. *Tribology* **2021**, *41*, 105–114.
27. Guan, H.; He, S.; Li, Q.; Yang, Z.; Qin, C.; He, W. A review of convolutional neural networks in equipment wear particle recognition. *Tribology* **2022**, *42*, 426–445.
28. Yang, J.X.; Lan, X.P.; Wang, B.; Yan, L.; Zhao, Z.; Feng, Y.D. Detection method based on deep learning for yellow industrial diamond. *Diam. Abras. Eng.* **2020**, *40*, 13–19.
29. Hu, W.D.; Wang, Z.K.; Dong, Y.H.; Zhang, Z.; Zhu, Y.W. Surface morphology characterization of fixed abrasive lapping pad based on deep learning. *Diam. Abras. Eng.* **2022**, *42*, 186–192.
30. Suo, W.L.; Lin, Y.F.; Fang, C.F. Surface morphology segmentation and evaluation of diamond lapping pad based on improved Mask R-CNN. *Diam. Abras. Eng.* **2024**. [CrossRef]
31. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762. [CrossRef]
32. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Houlsby, N. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
33. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Zhou, Y. TransUnet: Transformers make strong encoders for medical image segmentation. *arXiv* **2021**, arXiv:2102.04306.
34. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-Unet: Unet-like pure transformer for medical image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Tel Aviv, Israel, 23–28 October 2022; pp. 205–218.
35. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
36. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), Munich, Germany, 5–9 October 2015; pp. 234–241.
37. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y. Segment Anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2023; pp. 4015–4026.
38. Deng, G.; Zou, K.; Ren, K.; Wang, M.; Yuan, X.; Ying, S.; Fu, H. SAM-U: Multi-box prompts triggered uncertainty estimation for reliable SAM in medical image. *arXiv* **2023**, arXiv:2307.04973. [CrossRef]
39. Ji, G.P.; Fan, D.P.; Xu, P.; Zhou, B.; Cheng, M.M.; Gool, L.V. SAM struggles in concealed scenes—Empirical study on "Segment Anything". *Sci. China Inf. Sci.* **2023**, *66*, 226101. [CrossRef]
40. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 2nd ed.; Pearson Education: Upper Saddle River, NJ, USA, 2007.
41. Liao, Y.S.; Luo, S.Y. Wear characteristics of sintered diamond composite during circular sawing. *Wear* **1992**, *157*, 325–337. [CrossRef]
42. Galdran, A.; Carneiro, G.; Ballester, M.Á.G. On the optimal combination of Cross-Entropy and Soft Dice Losses for Lesion Segmentation with Out-of-Distribution Robustness. *arXiv* **2022**, arXiv:2209.06078.
43. Li, X.; Sun, X.; Meng, Y.; Liang, J.; Wu, F.; Li, J. Dice loss for data-imbalanced NLP tasks. *arXiv* **2019**, arXiv:1911.02855.
44. Xu, H.; Yang, M.; Deng, L.; Qian, Y.; Wang, C. Neutral cross-entropy loss based unsupervised domain adaptation for semantic segmentation. *IEEE Trans. Image Process.* **2021**, *30*, 4516–4525. [CrossRef]
45. Oquab, M.; Darcet, T.; Moutakanni, T.; Vo, H.; Szafraniec, M.; Khalidov, V.; Fernandez, P.; Haziza, D.; Massa, F.; El-Nouby, A.; et al. DINOv2: Learning Robust Visual Features without Supervision. *arXiv* **2024**, arXiv:2304.07193.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.