Article

From Segmentation to Classification: A Deep Learning Scheme for Sintered Surface Images Processing

Yi Yang [1,2,†] , Tengtuo Chen [1,†] and Liang Zhao [3,*]

1   School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China;
     yiyang@buaa.edu.cn (Y.Y.); tengtuo@buaa.edu.cn (T.C.)
2   Peng Cheng Laboratory, Shenzhen 518000, China
3   School of Software Technology, Dalian University of Technology, Dalian 116024, China
*   Correspondence: liangzhao@dlut.edu.cn
†   These authors contributed equally to this work.

Abstract: Effectively managing the quality of iron ore is critical to iron and steel metallurgy. Although quality inspection is crucial, the perspective of sintered surface identification remains largely unexplored. To bridge this gap, we propose a deep learning scheme for mining the necessary information in sintered images processing to replace manual labor and realize intelligent inspection, consisting of segmentation and classification. Specifically, we first employ a DeepLabv3+ semantic segmentation algorithm to extract the effective material surface features. Unlike the original model, which includes a high number of computational parameters, we use SqueezeNet as the backbone to improve model efficiency. Based on the initial annotation of the processed images, the sintered surface dataset is constructed. Then, considering the scarcity of labeled data, a semi-supervised deep learning scheme for sintered surface classification is developed, which is based on pseudo-labels. Experiments show that the improved semantic segmentation model can effectively segment the sintered surface, achieving 98.01% segmentation accuracy with only a 5.71 MB size. In addition, the effectiveness of the adopted semi-supervised learning classification method based on pseudo-labels is validated in six state-of-the-art models. Among them, the ResNet-101 model has the best classification performance, with 94.73% accuracy for the semi-supervised strategy while only using 30% labeled data, which is an improvement of 1.66% compared with the fully supervised strategy.

Keywords: deep learning; semantic segmentation; semi-supervised classification; sintered surface; iron ore sintering

1. Introduction

Iron and steel metallurgy, as a typical process industry in the traditional manufacturing sector, contributes significantly to both infrastructure development and economic growth [1]. In order to achieve the objectives of high-quality, high-yield, and low-consumption iron and steel manufacturing, the iron ore sintering process offers raw materials for the subsequent blast furnace ironmaking process, which is a crucial preliminary step for contemporary iron and steel smelting [2]. A poor and inconsistent quality of sintered ore can severely impact the blast furnace smelting process, resulting in, for example, unstable working conditions, poor iron quality, and other smelting product quality losses. Furthermore, the entire sintering process is fueled mostly by coal and coke, which produces significant amounts of carbon monoxide and carbon dioxide, resulting in latent environmental hazards [3,4]. In the future, the intelligent sintering process and measurement are poised to be a prominent research area in both academia and industry, aiming for quality improvement, enhanced productivity, energy conservation, environmental protection, and sustainable development.

Take the example of iron ore sintering. Due to the complexity of the sintering process and the temperature conditions, the quality of the sintered ore is influenced by the

composition and the original state of the material. In addition, the heat treatment process and the level of process parameter control also affect sintering results in many different ways. Consequently, even minor oversights regarding the sintered material's surface can lead to intricate defects, compromising the quality. Among the common sintering defects, the types of crack defects are the most complex, as shown in Figure 1. At present, mostly manual estimation and control are used to regulate the regulated physical quantities in the sintering process, such as temperature, ventilation, and electric pressure. However, manual categorization is both subjective and ineffective, which is why it frequently produces subpar sintered products [5]. Additionally, the sintering workplace is frequently polluted with toxic vapors and gases, making it unsuitable for prolonged manual labor. Hence, the classification and measurement of defects on the sintered surface are important for the effective control of the sintering process.
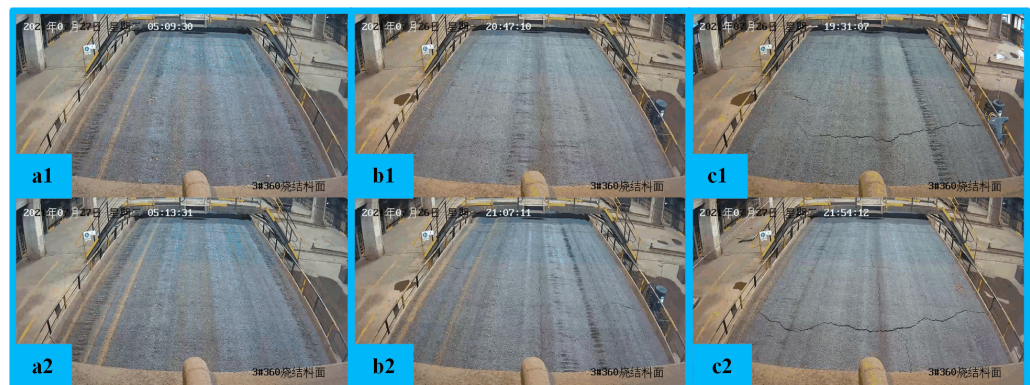
Figure 1. Three typical sintered surfaces. (**a1,a2**) are qualified sintered surfaces without obvious cracks. (**b1,b2**) are defective sintered surfaces with short cracks. (**c1,c2**) are serious defective sintered surfaces with long cracks.

In early studies, several image processing methods have been proposed and utilized for automated sintering analysis. For instance, Li et al. [6] utilize the feature image decomposition technique to extract the combustion state of sintering flame. Based on morphological image processing, Coster et al. [7] extract the morphological parameters in the microstructure of cerium dioxide during sintering. Nellros et al. [8] present a method for automatically calculating the geometrical properties and curvature of particles in microscopic sintering images to obtain finally the degree of sintering. Donskoi et al. [9] combine the Mineral4/Recognition4 software from the Commonwealth Scientific Industrial Research Organization (CSIRO) and structural texturing methods to obtain common sinter phases. Despite the fact that these image processing-based methods obtain good results with the designed descriptors, they do not focus on sintered surfaces classification. Moreover, images of sintered surfaces are complex in terms of multi-feature representation of textural details, and parameters such as particle density and reflectance do not accurately represent the severity of sintered cracks. Therefore, it is difficult to classify sintered images by manual feature extraction.

In recent years, deep learning has demonstrated powerful visual feature learning abilities [10–14]. A large number of deep learning methods are applied in object detection [15], semantic segmentation [16], image classification [17], recommender systems [18] and medical image analysis [19]. However, the application of deep learning methods for downstream tasks in sintering industrial scenarios have some challenges. For example, the high temperature inside the sintering workroom places demands on the high-temperature-resistant real-time operation of the camera capturing the images. Second, the acquisition of sintering surface images is often obstructed by the workers' field operations. In addition, there is a lack of publicly available datasets based on different degrees of sintered surface cracks. In addition, it is difficult for existing deep neural network models [20–23] with supervised learning schemes to classify sintered surfaces accurately while there is a lack of labeled

data. The classification results must, in particular, deliver accurate and timely information to the igniter and controlled devices, while reserving adequate computational resources to implement subsequent control algorithms.

To address the aforementioned issues, we use a DeepLabv3+ semantic segmentation algorithm based on the lightweight SqueezeNet [24] backbone to extract the effective material surface information required for classification, so as to construct the sintered surface dataset. Meanwhile, we design a semi-supervised deep learning method for sintered surface classification, which is a self-learning model based on pseudo-labels.

The deep learning scheme for sintered surfaces proposed in this paper consists of the following parts: sintered surface image acquisition, semantic segmentation based on lightweight networks, and semi-supervised classification based on self-learning, as shown in Figure 2. First, the images are captured in the sintering workshop by high-temperature-resistant camera equipment. Then, the acquired images are semantically segmented to obtain the sintered surface, with occlusions and extraneous objects removed. Image cropping is required after semantic segmentation in order to obtain the optimal material surface region used for prediction. Finally, a network model trained based on semi-supervised learning is applied to determine the crack severity of the sintered surface. Results from the classifier may be forwarded to the sintering process controller, where the necessary parameters can be modified to improve sintering.
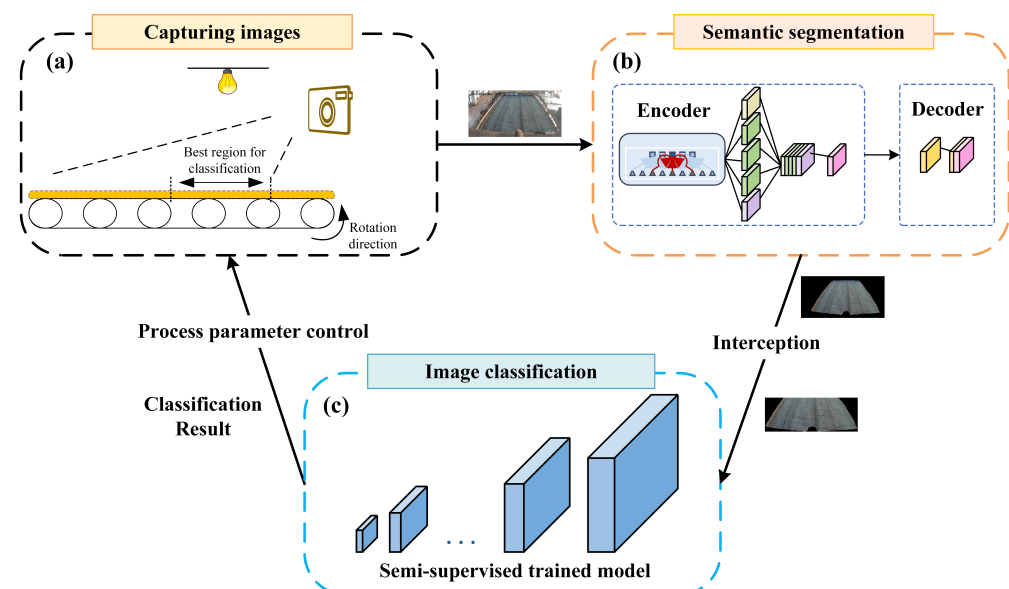
Figure 2. Deep learning scheme for sintered surface images processing. (**a**) Sintered surface image acquisition. The acquired image contains a time-lagged material surface and occlusions. (**b**) Semantic segmentation. The images obtained in step (**a**) are segmented using the modified DeepLabV3+. (**c**) Image classification. A net model obtained by training based on the semi-supervised method is used to classify the sintered images. The obtained classification results are provided to the process controller to improve the sintering.

Overall, in this paper, we develop a deep learning scheme for sintered surface image processing, including segmentation and classification. The following three aspects describe the primary contributions of this study.

(1) We design a scheme that pioneers lightweight semantic segmentation preprocessing of sintered surfaces. Considering the lightweight requirement of the semantic segmentation model, we change the backbone network of DeepLabV3+ to SqueezeNet. The newly proposed semantic segmentation model dramatically improves the segmentation efficiency and saves computing resources. Therefore, our proposed semantic segmentation method has significant efficiency improvement compared with traditional image preprocessing methods. Specifically, the proposed semantic segmentation

method can achieve 97.08% mean IOU and 98.03% mean accuracy with only 3.7% labeled images.
(2) We create a multiclassified dataset for sintered surface crack classification. For the first time, we have labeled the sintered crack types, incorporating insights from authoritative experts. A total of 1334 samples are labeled, with a resolution of 800×330, which can be used to study the classification method of sintered surface cracks. It is worth mentioning that the images in the dataset are acquired in real high-temperature sintering scenarios, which is a challenging process. In addition, tagging thousands of images is a labor-intensive task due to the similarity between images.
(3) We innovatively apply a semi-supervised self-learning method based on pseudo-labels to crack classification of sintered surfaces. The training strategy using pseudo-labels can significantly reduce the need for training samples. Therefore, this method is suitable for the task of sintered surface classification where it is difficult to obtain labeled data. Experimental results demonstrate that our method can rival the classification accuracy of supervised learning with only 20% of training samples and can exceed the classification effect of supervised learning with 30% of training samples.

2. Methods

2.1. Data Acquisition and Annotation

The sintering data studied in this paper are obtained from the iron sintering plant in Guangxi, China. The initial image data are captured by an industrial-grade professional video camera with a resolution of 1920×1080, and the shooting time is from 26 July 2021 to 3 January 2022. The expert criteria for image labeling are shown in Table 1. The width of the sintered surface is denoted as S and the length of the crack is denoted as x. The width of the sintered surface is denoted as S and the length of the crack is denoted as x. Due to the angle at which the photographs are taken, there is a visual difference between the sintered surfaces in the images, which are near large and far small. Therefore, we use the numerical reduction method to approximate the crack length by considering it as an isosceles trapezoid to obtain the approximate crack length.

Table 1. Labeling criteria and quantity of sintered images.

Class	Crack Severity	Basis for Annotation	Number
LC	low	Tiny cracks or undetectable	809
MC	moderate	$x < 0.5\,S$	348
HC	high	$x > 0.5\,S$	177

2.2. Semantic Segmentation

Initial sintered images are difficult to use directly for classification because of the complex disturbances that affect imaging at industrial sites. Specifically, in addition to the sintered surfaces required for classification, the images contain igniter ends, sintering platform edges, self-contained textual information, and possible construction worker obstructions. The above disturbances can directly affect the classification results. Considering that image processing methods cannot directly obtain the complete material surface, this paper adopts the idea of semantic segmentation to solve this problem. After completing the semantic segmentation, in order to be used to obtain the best area for classification, we crop half of the sintered surface at the proximal end of the industrial camera.

DeepLabV3+, as a mature semantic segmentation model, is widely used in various tasks [25–29]. However, it suffers from high complexity and computational resource consumption. With the higher resolution of the image used, the computational and parametric quantities of the model will explode accordingly, which has a great impact on the efficiency of the model. Therefore, for real industrial scenarios, the model is required to guarantee the efficiency of the algorithm and limit the model size to a certain extent, while improving

the accuracy. For practical deployment in the sintering industry, it is most practical and effective to apply an improved lightweight module. Therefore, we redesign the backbone network of DeepLabV3+, choosing the lightweight network SqueezeNet as a substitution.

SqueezeNet is a classical lightweight network. The model uses fire modules to reduce the dimension of feature maps and has been used in many industrial scenarios [30,31]. The model has outstanding lightweight performance, dramatically reducing the number of parameters with guaranteed performance, and can be deployed on microcontrollers and FPGAs. Hence, adopting SqueezeNet as the backbone of the semantic segmentation model can save computational and storage resources for subsequent calculation and control.

The semantic segmentation model we have designed is shown in Figure 3. The model consists of two parts, the encoder and the decoder.

Encoder: used to extract shallow and deep features. Firstly, the sintered image undergoes a modified backbone to generate two valid features. The shallow features go directly into the decoder, while the deeper features need to be downsampled. Atrous spatial pyramid pooling (ASPP) is employed in the downsampling structure. ASSP uses atrous convolution with expansions of 6, 12, and 18 for feature extraction. Atrous convolution expands the receptive field while preserving information, allowing each convolution output to hold multi-scale data. After stacking the feature layers, channel counting is performed by 1×1 convolution to obtain the pink feature layer in Figure 3, which has high semantic information.

Decoder: used to obtain the segmented image. The decoder can fuse the shallow features generated by SqueezeNet with the deep features generated by the encoder. First, the deep features are fed into the decoder for up-sampling. Then, the features are fused with the results obtained from the shallow features after 1×1 convolution. Finally, feature extraction is performed using 3×3 convolution. Up-sampling guarantees that the output picture is the same size as the input image, and the prediction results are achieved, allowing the sintered area extraction to be realized.
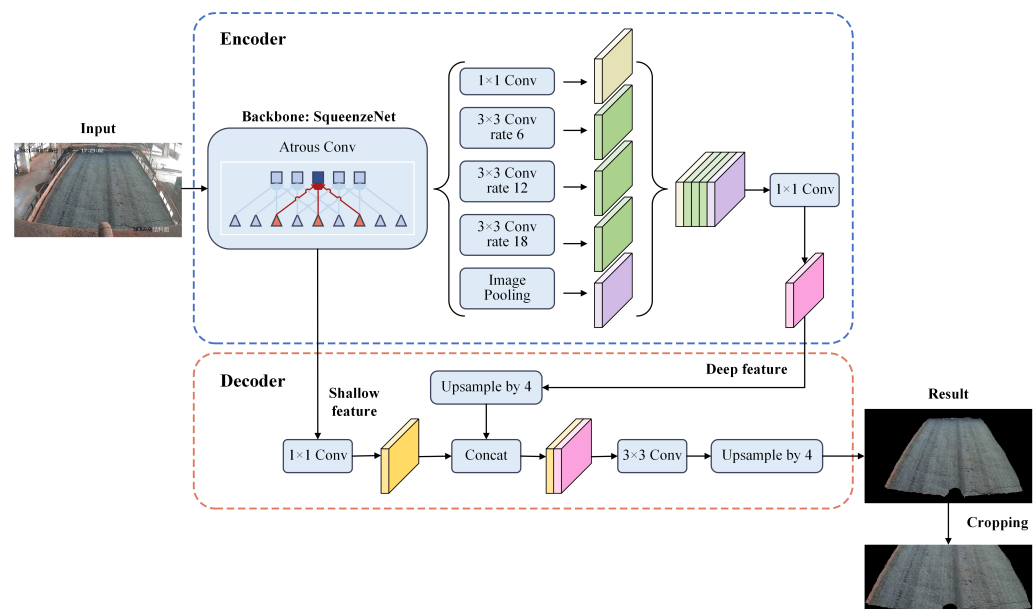
Figure 3. Improved DeepLabV3+ semantic segmentation model using SqueezeNet as the backbone.

The proposed model can segment sintered surfaces accurately and efficiently with low requirements for the number of manual annotations, and these conclusions will be demonstrated in experiments. After obtaining the results, cropping is performed to obtain the most suitable sintered surface region for classification. The final image of one-half of the region close to the camera is obtained, with a resolution of 800×330.

2.3. Semi-Supervised Classification Based on Self-Learning

The method of semantic segmentation is not effective for segmentation of small samples such as sintered cracks, and the associated labeling cost is too high. Therefore, in this paper, the strategy of classification is adopted to give the degree of defects on the sintered surface visually.

Image classification models based on deep neural networks have developed rapidly in recent years, and include ResNet [32], Vision Transformer [33], MobileNetV3 [34], EfficientNetV2 [35], and EdgeNeXt [36]. These methods are able to achieve promising classification results with sufficient training data. However, when training on small amounts of labeled data, they are unable to produce accurate classification results. Therefore, supervised learning-based classification models cannot be applied on scarce labeled data such as sintered surface images. In addition, the computational cost of these network models is high. To solve the above problems, the idea of semi-supervised learning is adopted. We have designed a concise yet effective semi-supervised classification structure, which is based on pseudo-labels and self-learning.

Semi-supervised learning is a learning approach that involves developing models using both labeled and unlabeled data. Semi-supervised approaches, as opposed to supervised learning algorithms, can increase learning performance by employing more unlabeled instances [37]. The self-training approach we utilize generates pseudo-labels for unlabeled data by using the model's own confidence predictions. The technique employs an entropy regularization strategy to avoid decision boundaries from passing across densely packed data point areas. Encouraging the model to produce low entropy predictions for unlabeled data and adding pseudo-labeled samples to the labeled dataset for a typical supervised learning setup enables semi-supervised classification. The pseudo-label presents a straightforward and efficient method for semi-supervised training of neural networks, where the network undergoes supervised trained using both labeled and unlabeled input. As seen in Figure 4, the model is trained through cross-entropy loss in the typical supervised method on labeled data. The same model is applied to unlabeled data to obtain predictions for a batch of unlabeled samples. Pseudo-labeling, the forecast with the highest degree of confidence, also has the highest likelihood of being correct. That is, the pseudo-labeling model trains a neural network with loss function $\mathcal{L}$.

$$\mathcal{L} = \frac{1}{n} \sum_{m=1}^{n} \sum_{i=1}^{K} \mathcal{R}(y_i^m, f_i^m) + \alpha(t) \frac{1}{n'} \sum_{m=1}^{n'} \sum_{i=1}^{K} \mathcal{R}(y'^m_i, f'^m_i). \tag{1}$$

where n represents the number of small batches in the labeled data, n' represents the unlabeled data, y_i^m is the output unit of the m-th sample in the labeled data, y_i^m is its label, f'^m_i is the output unit of the m-th sample in the unlabeled data, y'^m_i is the pseudo-label of the unlabeled data, and $\alpha(t)$ is the coefficient responsible for balancing the supervised and unsupervised loss terms in the t-th training epoch.

A reasonable $\alpha(t)$ is important for network performance. If $\alpha(t)$ is too high, even labeled data will interfere with training. Whereas if $\alpha(t)$ is too small, we cannot benefit from unlabeled data. Therefore, a deterministic annealing process in Equation (2) is used in our experiments. By slowly increasing $\alpha(t)$, it is expected to help the optimization process to avoid worse local minima, thus enabling the pseudo-labels of unlabeled data to be as close as possible to the authentic labels.

$$\alpha(t) = \begin{cases} 0 & t < T_1 \\ \frac{t-T_1}{T_2-T_1} \alpha_f & T_1 \le t < T_2 \\ \alpha_f & T_2 \le t \end{cases} \tag{2}$$

In Equation (2), t is the current epoch, and T_1, T_2, and α_f are manually selected values. In the subsequent experiments in this paper, consistent values are used in training process, i.e., $\alpha_f = 2$, and $T_1 = 50$, $T_2 = 150$.
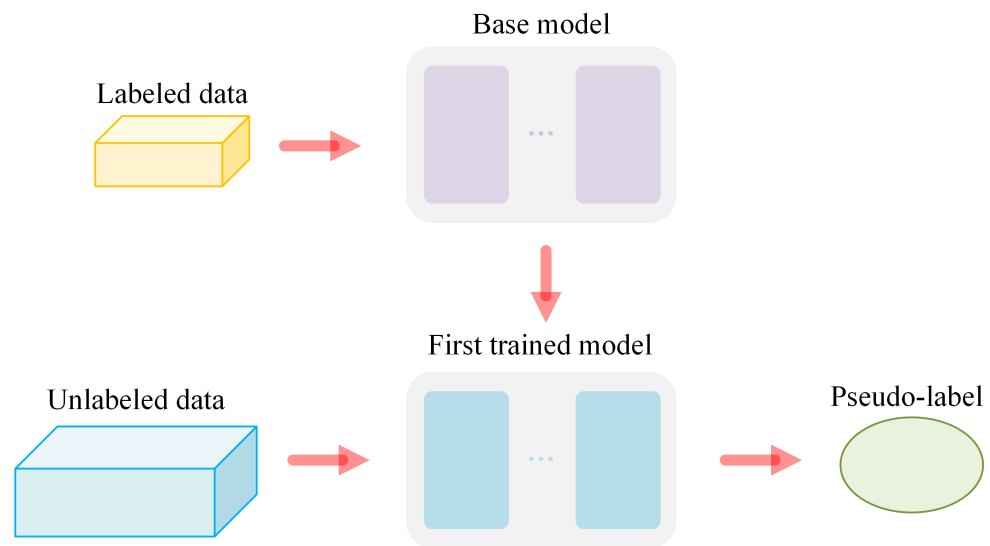
Figure 4. Pseudo-label generation. First, the labeled data are used to obtain a pre-trained classification model. Then, the obtained classification model is employed to classify the unlabeled data, and the results with higher confidence are selected as pseudo-labels.

After obtaining the pseudo-label, we can really proceed to the training of the final model. Based on the semi-supervised learning theory and loss design, the semi-supervised self-learning model training process we constructed is shown in Figure 5. Specifically, the original labeling data and the most confident pseudo-labels are integrated into a new dataset as the training set. Then, we use the new dataset for the training of the final model. It is worth noting that, in the self-learning process we designed, the network structure of the base model and final model are the same, which avoids the additional adverse effects of model inconsistency.
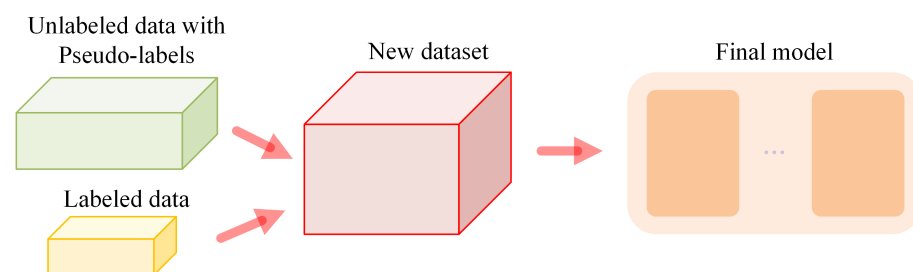
Figure 5. Final model generation.

3. Experiments

In this part, we look at the development and performance of the suggested framework for analyzing sintered surface images, which includes surface segmentation and semi-supervised classification. The experimental setup and the dataset for sintered surfaces are first discussed. The effectiveness of the suggested approaches is then assessed twice. In the first section, we compare the segmentation performance and runtime of the enhanced semantic segmentation approach employing SqueezeNet as the backbone with those of five different backbone networks. The comparative studies demonstrate the suggested method's accuracy and portability. In the second section, we run tests with the created semi-supervised training framework on six different network topologies and evaluate how well it performs supervised learning. The outcomes show how well the framework for semi-supervised training on sintered surfaces classification performs.

3.1. Sintered Surface Dataset

The sintered surface dataset is made up of 1334 photos with a size of 1920 × 1080 pixels that have been obtained from a sintering plant in Guangxi Province, China. A high-temperature-resistant picture-capture platform acquires the dataset, which can depict the state of the sintered surface over time. The source photos are semantically segmented to provide sintered surface images at a resolution of 800 × 330 pixels of the best classified region.

As shown in Table 1, based on the severity of the cracks on the sintered surface, the experts categorized the acquired images into low crack (LC), moderate crack (MC), and high crack (HC). Typical images for each category are displayed in Figure 1.

3.2. Application Specifics

The designed improved lightweight DeepLabV3+ semantic segmentation model is implemented in the PyTorch framework. Labeling for segmentation is performed using Labelme. The Adam optimizer is used to update the network parameters during the training process. The maximum training epoch is 20 and the start-up learning rate is 0.0001, which reduces by a rate of 0.1 after 10 epochs with an 8-batch size. In the experiments on semantic segmentation, only 50 labeled samples are used as a training set to demonstrate the generalization ability of the model. The same training strategy and parameter settings are applied to the other backbone networks used for comparison.

In the experiments on semi-supervised classification, the scale of the training set is set to 0.1, 0.2, 0.3, and 0.4, and the rest of the data are put into the first trained model as unlabeled data to obtain the pseudo-labels. In the fully supervised (F-S) classification experiments, for comparison, the scale of the training set is 0.8. The initial learning rate is set to 0.001, using the same learning rate decay strategy as for semantic segmentation. The training epoch for classification is 200. All experiments are performed on standard workstations equipped with NVIDIA GeForce GTX 3070 GPUs, 12th Gen Intel(R) Core (TM) i7-12700K CPUs, and 64GB of RAM on standard workstations.

3.3. Performance Measures

The comparison experiment of improved lightweight semantic segmentation for sintered surface segmentation is shown in Table 2. In order to compare the performance of different backbones on semantic segmentation of sintered surfaces, accuracy (ACC), mean accuracy (m-ACC), global-ACC, intersection-over-union (IoU), mean intersection-over-union (m-IoU), and model size are selected as evaluation metrics.

The performance of our improved model in sintered surface segmentation can be clearly shown by comparison. On the segmentation metrics of the background of the sintered image, the SqueezeNet-based backbone designed by us performs the best. In addition, our model achieves the best results on both m-ACC and global-ACC. Despite the slight advantage in the ACC and IoU metrics of the sintered surface segmentation with the model using ResNet-101 as the backbone, the size of our model is only 5.71 MB, which is much smaller than ResNet-101's size of 144.53 MB. Although the model size is only one-twenty-fifth of ResNet-101, our model still achieves satisfactory results. In addition, Figure 6 shows the effect of sintered surface splitting under different backbones. It can be seen that our method achieves better results in the completeness of the sintered surface segmentation and the treatment of the edges.

Image	Xception	ResNet-101	MobileNetV3	ShuffleNetV2	Ours
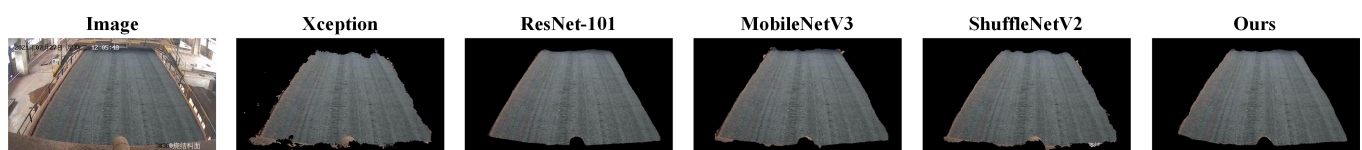
Figure 6. Visual comparison of different backbones for sintered surface segmentation.

Table 2. Comparison results (%) of semantic segmentation performance. The best result in each classification is in bolded red, and the second is in bolded black.

| Backbone | Background | | Surface | | m-ACC | Global-ACC | m-IoU | Model Size |
	ACC	IoU	ACC	IoU				(MB)
Xception [23]	89.04	85.72	86.13	85.1	87.59	88.15	85.41	97.15
ResNet-101 [30]	**98.07**	**97.92**	**97.22**	**96.47**	**97.85**	**97.98**	**97.2**	144.53
MobileNetV3 [32]	95.51	92.11	95.07	91.59	95.29	95.09	91.85	22.57
ShuffleNetV2 [36]	93.76	92.35	91.43	88.74	92.6	93.02	90.55	**9.46**
Ours	**98.84**	**98.18**	**97.18**	95.98	**98.01**	**98.03**	97.08	**5.71**

In experiments on semi-supervised sintered surfaces classification, we apply AlexNet [38], ResNet-101 [32], VGG16 [39], EfficientNetV2 [35], MobileNetV3 [34], and ShuffleNetV2 [40] as models to a pseudo label-based self-learning approach. ACC and F1-scores are used as metrics to assess classification performance. As can be obtained from Figure 7, Tables 3 and 4, the proposed semi-supervised strategy can achieve similar results as fully supervised (F-S) classification with only 20% labeled data. The corresponding accuracy and F1-score achieve satisfactory performance. It is worth noting that a subset of algorithms such as AlexNet, ResNet-101, MobileNetV3, and ShuffleNetV2 can outperform the fully supervised strategy in ACC when the amount of labeled data reaches 30%. We hypothesize that this is because using the first trained model can avoid a portion of the data noise, resulting in better classification results. This is also verified in the subsequent semi-supervised experiments using 40% labeled data. As can be seen from Tables 3 and 4, except for EfficientNetV2 and MobileNetV3, the prediction results using 40% labeled data are conversely not as good as those using 30% labeled data.

Among the metrics for classification, ResNet-101, which uses 30% labeled data for semi-supervised learning, performs best. Its ACC is 94.73% and the F1-score is 0.9415, which reflects that a deeper and larger network structure tends to have a better fitting effect in complex sintered crack classification tasks. However, the results in Table 2 tell us that a bigger backbone for semantic segmentation does not tend to be better. Therefore, it is important to choose the right network model for the appropriate downstream task so that not only the task requirements can be met, but also the waste of computational and storage resources can be avoided.
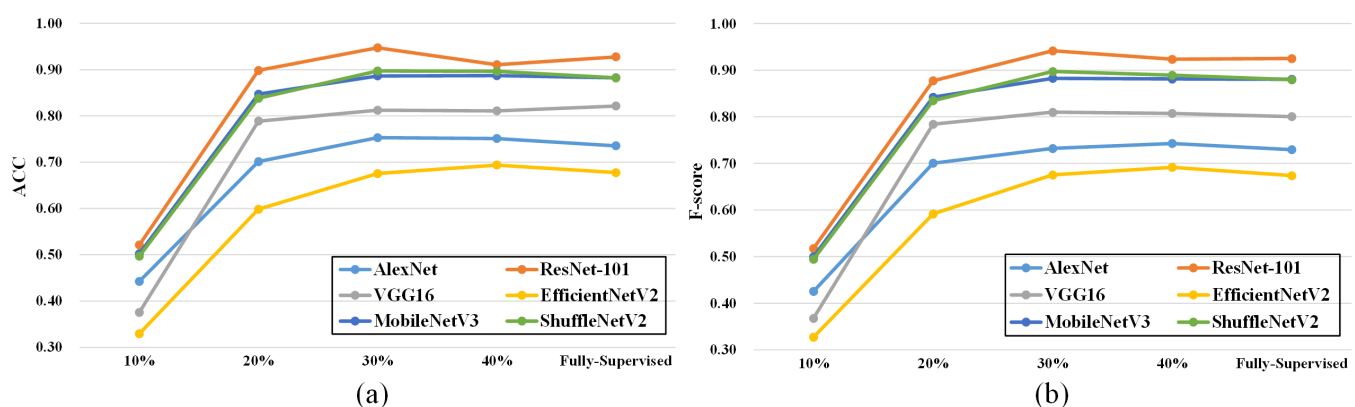
Figure 7. Comparison results of classification performance. (**a**) ACC with the semi-supervised scheme in different proportions of labeled data and the fully supervised method. (**b**) F1-score with the semi-supervised scheme in different proportions of labeled data and the fully supervised method.

To summarize, the improved DeepLabV3+ semantic segmentation model can be used for sintering surface extraction. And the pseudo labeling-trained ResNet-101 can be applied to real working conditions to provide feedback for sintering process control.

Table 3. Comparison results of classification ACC (%) in detail.

Model	10%	20%	30%	40%	F-S
AlexNet	44.26	70.14	75.31	75.12	73.54
ResNet-101	52.12	89.83	**94.73**	91.07	92.76
VGG16	37.54	78.87	81.24	81.09	82.15
EfficientNetV2	32.95	59.83	67.55	69.39	67.75
MobileNetV3	50.23	84.71	88.64	88.70	88.25
ShuffleNetV2	49.71	83.83	89.72	89.64	88.22

Table 4. Comparison results of classification F1-score in detail.

Model	10%	20%	30%	40%	F-S
AlexNet	0.4254	0.7004	0.7321	0.7427	0.7294
ResNet-101	0.5174	0.8772	**0.9415**	0.9233	0.9249
VGG16	0.3677	0.7841	0.8098	0.8072	0.8004
EfficientNetV2	0.3273	0.5916	0.6752	0.6916	0.6737
MobileNetV3	0.4998	0.8419	0.8824	0.881	0.8805
ShuffleNetV2	0.4944	0.8346	0.8972	0.8891	0.8795

4. Conclusions

In this paper, we propose a deep learning scheme for sintered surface images processing. This scheme consists of a semantic segmentation method for building the dataset and a semi-supervised training strategy for sintered surface crack classification. Specifically, we first collect and label the sintered surface images at a real sintering site. Then, DeepLabV3+ based on the SqueezeNet lightweight network is designed for semantic segmentation of sintered images. After obtaining the optimal classification region of the sintered surface, a semi-supervised classification method based on pseudo-labels is used to classify the dataset.

Experiments reveal that the enhanced semantic segmentation model is able to achieve 98.01% segmentation accuracy with only 5.71 MB size, and can effectively segment the sintered surface. The designed semi-supervised strategy is validated on six state-of-the-art web models. The accuracy is comparable to supervised learning when using only 20% labeled data. Not only that, the classification accuracy obtained by semi-supervision can outperform supervised learning when using 30% labeled data. Among them, ResNet-101 performs best in semi-supervised classification. Therefore, the proposed methods can solve the defect classification problem effectively under the difficulty of obtaining sintering image data, and provide reliable defect information for sintering process control.

In the future, we will concentrate on two specific issues and further consider the migratability, replicability, and scalability of the model. On the one hand, we will attempt to apply picture enhancement techniques to boost classification accuracy for sintered surfaces. To increase the automation of the sintering process and the sintering quality successfully, on the other hand, we intend to develop a feedback control model based on the projected sintering crack types and process parameters. In order to ensure the portable, replicable, and scalable operation of the proposed algorithms, validation in more industrial scenarios and real data will also be a part of the follow-up research.

References

1. Kwon, W.H.; Kim, Y.H.; Lee, S.J.; Paek, K.N. Event-based modeling and control for the burnthrough point in sintering processes. *IEEE Trans. Control. Syst. Technol.* **1999**, *7*, 31–41. [CrossRef]
2. du Preez, S.P.; van Kaam, T.P.M.; Ringdalen, E.; Tangstad, M.; Morita, K.; Bessarabov, D.G.; van Zyl, P.G.; Beukes, J.P. An Overview of Currently Applied Ferrochrome Production Processes and Their Waste Management Practices. *Minerals* **2023**, *13*, 809. [CrossRef]
3. Chen, R.; Shi, L.; Huang, H.; Yuan, J. Extraction of Iron and Alumina from Red Mud with a Non-Harmful Magnetization Sintering Process. *Minerals* **2023**, *13*, 452. [CrossRef]
4. Chen, S.; Li, J.; You, Q.; Wang, Z.; Shan, W.; Bo, X.; Zhu, R. Improving the Air Quality Management: The Air Pollutant and Carbon Emission and Air Quality Model for Air Pollutant and Carbon Emission Reduction in the Iron and Steel Industries of Tangshan, Hebei Province, China. *Atmosphere* **2023**, *14*, 1747. [CrossRef]
5. Fan, J.; Liu, M.; Wang, X.; Wang, J.; Wen, H.; Wang, Y. A novel automatic classification method based on the hybrid lightweight shunt network for sintered surfaces. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–11. [CrossRef]
6. Li, W.; Wang, D.; Chai, T. Flame image-based burning state recognition for sintering process of rotary kiln using heterogeneous features and fuzzy integral. *IEEE Trans. Ind. Inform.* **2012**, *8*, 780–790. [CrossRef]
7. Coster, M.; Arnould, X.; Chermant, J.; Chermant, L.; Chartier, T. The use of image analysis for sintering investigations: The example of CeO_2 doped with TiO_2. *J. Eur. Ceram. Soc.* **2005**, *25*, 3427–3435. [CrossRef]
8. Nellros, F.; Thurley, M.J.; Jonsson, H.; Andersson, C.; Forsmo, S.P. Automated measurement of sintering degree in optical microscopy through image analysis of particle joins. *Pattern Recognit.* **2015**, *48*, 3451–3465. [CrossRef]
9. 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]
10. Nosratabadi, S.; Mosavi, A.; Duan, P.; Ghamisi, P.; Filip, F.; Band, S.S.; Reuter, U.; Gama, J.; Gandomi, A.H. Data Science in Economics: Comprehensive Review of Advanced Machine Learning and Deep Learning Methods. *Mathematics* **2020**, *8*, 1799. [CrossRef]
11. Wang, C.; Zhang, Q.; Tian, Q.; Li, S.; Wang, X.; Lane, D.; Petillot, Y.; Wang, S. Learning Mobile Manipulation through Deep Reinforcement Learning. *Sensors* **2020**, *20*, 939. [CrossRef]
12. Xu, J.; Xi, X.; Chen, J.; Sheng, V.S.; Ma, J.; Cui, Z. A Survey of Deep Learning for Electronic Health Records. *Appl. Sci.* **2022**, *12*, 11709. [CrossRef]
13. Vithayathil Varghese, N.; Mahmoud, Q.H. A Survey of Multi-Task Deep Reinforcement Learning. *Electronics* **2020**, *9*, 1363. [CrossRef]
14. Yang, Y.; Chen, T.; Zhao, L.; Gu, J.; Tang, X.; Zhang, Y. Defects Clustering for Mineral Sintering Surface Based on Multi-source Data Fusion. In Proceedings of the 2023 2nd Conference on Fully Actuated System Theory and Applications (CFASTA), Qingdao, China, 14–16 July 2023; pp. 670–674. [CrossRef]
15. Liu, Z.; Wang, L.; Liu, Z.; Wang, X.; Hu, C.; Xing, J. Detection of Cotton Seed Damage Based on Improved YOLOv5. *Processes* **2023**, *11*, 2682. [CrossRef]
16. Chen, Y.; Yan, Q.; Huang, W. MFTSC: A Semantically Constrained Method for Urban Building Height Estimation Using Multiple Source Images. *Remote Sens.* **2023**, *15*, 5552. [CrossRef]
17. Ong, W.; Liu, R.W.; Makmur, A.; Low, X.Z.; Sng, W.J.; Tan, J.H.; Kumar, N.; Hallinan, J.T.P.D. Artificial Intelligence Applications for Osteoporosis Classification Using Computed Tomography. *Bioengineering* **2023**, *10*, 1364. [CrossRef]
18. El Youbi El Idrissi, L.; Akharraz, I.; Ahaitouf, A. Personalized E-Learning Recommender System Based on Autoencoders. *Appl. Syst. Innov.* **2023**, *6*, 102. [CrossRef]

19. Alom, M.Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S.; Hasan, M.; Van Essen, B.C.; Awwal, A.A.S.; Asari, V.K. A State-of-the-Art Survey on Deep Learning Theory and Architectures. *Electronics* **2019**, *8*, 292. [CrossRef]

20. Wu, Z.; Shen, C.; Van Den Hengel, A. Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recognit.* **2019**, *90*, 119–133. [CrossRef]

21. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

22. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.

23. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.

24. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and <0.5MB Model Size, 2016. Available online: http://xxx.lanl.gov/abs/1602.07360 (accessed on 4 November 2016).

25. Chen, M.; Jin, C.; Ni, Y.; Xu, J.; Yang, T. Online Detection System for Wheat Machine Harvesting Impurity Rate Based on DeepLabV3+. *Sensors* **2022**, *22*, 7627. [CrossRef]

26. Chen, Y.; He, G.; Yin, R.; Zheng, K.; Wang, G. Comparative Study of Marine Ranching Recognition in Multi-Temporal High-Resolution Remote Sensing Images Based on DeepLab-v3+ and U-Net. *Remote Sens.* **2022**, *14*, 5654. [CrossRef]

27. Hu, S.; Liu, J.; Kang, Z. DeepLabV3+/Efficientnet Hybrid Network-Based Scene Area Judgment for the Mars Unmanned Vehicle System. *Sensors* **2021**, *21*, 8136. [CrossRef] [PubMed]

28. Emek Soylu, B.; Guzel, M.S.; Bostanci, G.E.; Ekinci, F.; Asuroglu, T.; Acici, K. Deep-Learning-Based Approaches for Semantic Segmentation of Natural Scene Images: A Review. *Electronics* **2023**, *12*, 2730. [CrossRef]

29. Antonelli, L.; De Simone, V.; di Serafino, D. A view of computational models for image segmentation. *Ann. Dell'Universita' Ferrara* **2022**, *68*, 277–294. [CrossRef]

30. Ciaburro, G.; Padmanabhan, S.; Maleh, Y.; Puyana-Romero, V. Fan Fault Diagnosis Using Acoustic Emission and Deep Learning Methods. *Informatics* **2023**, *10*, 24. [CrossRef]

31. Fu, G.; Le, W.; Zhang, Z.; Li, J.; Zhu, Q.; Niu, F.; Chen, H.; Sun, F.; Shen, Y. A Surface Defect Inspection Model via Rich Feature Extraction and Residual-Based Progressive Integration CNN. *Machines* **2023**, *11*, 124. [CrossRef]

32. Shafiq, M.; Gu, Z. Deep Residual Learning for Image Recognition: A Survey. *Appl. Sci.* **2022**, *12*, 8972. [CrossRef]

33. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In Proceedings of the International Conference on Learning Representations, Virtual Event, 3–7 May 2021.

34. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.

35. Tan, M.; Le, Q. Efficientnetv2: Smaller models and faster training. In Proceedings of the International Conference on Machine Learning. PMLR, Virtual, 18–24 July 2021; pp. 10096–10106.

36. Maaz, M.; Shaker, A.; Cholakkal, H.; Khan, S.; Zamir, S.W.; Anwer, R.M.; Shahbaz Khan, F. Edgenext: efficiently amalgamated cnn-transformer architecture for mobile vision applications. In Proceedings of the European Conference on Computer Vision. Springer, Tel Aviv, Israel, 23–27 October 2022; pp. 3–20.

37. Yang, X.; Song, Z.; King, I.; Xu, Z. A Survey on Deep Semi-Supervised Learning. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 8934–8954. [CrossRef]

38. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 84–90. [CrossRef]

39. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.

40. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.