

Article

Study on the Recognition of Metallurgical Graphs Based on Deep Learning

Qichao Zhao ¹, Jinwu Kang ^{1,*} and Kai Wu ²

¹ Key Laboratory for Advanced Materials Processing Technology, School of Materials Science and Engineering, Tsinghua University, Beijing 100084, China; zhaoqc22@mails.tsinghua.edu.cn

² School of Weiyang, Tsinghua University, Beijing 100084, China; wu-k21@mails.tsinghua.edu.cn

* Correspondence: kangjw@tsinghua.edu.cn

Abstract: Artificial intelligence has been widely applied in image recognition and segmentation, achieving significant results. However, its application in the field of materials science is relatively limited. Metallography is an important technique for characterizing the macroscopic and microscopic structures of metals and alloys. It plays a crucial role in correlating material properties. Therefore, this study investigates the utilization of deep learning techniques for the recognition of metallographic images. This study selected microscopic images of three typical cast irons, including ductile, gray, and white ones, and another alloy, cast aluminum alloy, from the ASM database for recognition investigation. These images were cut and enhanced for training. In addition to coarse classification of material type, fine classification of material type, composition, and the conditions of image acquisition such as microscope, magnification, and etchant was performed. The MobileNetV2 network was adopted as the model for training and prediction, and ImageNet was used as the dataset for pre-training to improve the accuracy. The metallographic images could be classified into 15 categories by the trained neural networks. The accuracy of validation and prediction for fine classification reached 94.44% and 93.87%, respectively. This indicates that neural networks have the potential to identify types of materials with details of microscope, magnification, etchants, etc., supplemental to compositions for metallographic images.



Citation: Zhao, Q.; Kang, J.; Wu, K.

Study on the Recognition of Metallurgical Graphs Based on Deep Learning. *Metals* **2024**, *14*, 732.

<https://doi.org/10.3390/met14060732>

Academic Editors: Crystopher Cardoso de Brito and Marta Oliveira

Received: 11 May 2024

Revised: 31 May 2024

Accepted: 17 June 2024

Published: 20 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: artificial intelligence; deep learning; metallurgical graph; microstructure; recognition

1. Introduction

For metal materials, their physical and mechanical properties are directly determined by their microstructure. Therefore, it is necessary to perform metallographic analysis by examining their microstructure using different microscopes and devices [1]. Optical microscopy, electron microscopy, X-ray microscopy, etc., are typical methods for macroscopic analysis. Microscopic images require professional knowledge for identification, which is challenging and subject to the knowledge and practice of professionals. To reduce human recognition errors and improve efficiency, metallographic analysis, mainly driven by computer technology, is a growing trend.

With the rapid development of computer technology, new technologies such as machine learning have emerged. Machine learning is a subset of artificial intelligence that optimizes models by learning the inherent distribution patterns of data and can make various predictions. Deep learning is a form of multi-layer model that can better extract deep features from data, making it widely used in fields such as natural language processing [2], medical applications [3], image segmentation [4], and face recognition [5].

In the 1990s, the combination of neural networks and support vector machines greatly promoted the development of image recognition. Subsequently, more and more models were used for image processing. The convolutional neural network (CNN) model, which can extract features layer by layer and is very suitable for problems such as image recognition, consists of convolutional layers, pooling layers, and fully connected layers with input

and output layers. VGGNet [6], proposed by the Oxford Visual Geometry Group, is a CNN neural network with 16 to 19 layers based on modifications to AlexNet. The DBN model [7] proposed by Geoffrey consists of a multi-layer constrained Boltzmann machine and a certain type of classifier. The GAN model [8] is composed of a generator and a discriminator which learn together through adversarial means in order to continuously improve the generator's predictive ability. Neural network models have been widely applied in various fields. In 2017, Zhong et al. [9] used an improved DBN model to classify remote sensing images and achieved high accuracy. Zhang et al. [10] proposed a weighted DenseNet network with multiple features and achieved significant results in object and image recognition by adaptively recalibrating feature responses and establishing dependencies between different convolutional layers. Wang Junmin et al. [11] accurately predicted different texture maps with small training samples by adopting CNN and transfer learning methods.

Deep learning techniques can be used for the recognition of microscopic images for metal materials. Chowdhury et al. [12] separated microscopic images with or without dendritic microstructures using models such as support vector machines, with an accuracy over 90%. The dataset used in their study consisted of micrographs with and without dendritic morphology, including 528 images, each with a size of 227×227 pixels. This demonstrates that deep learning is effective in recognizing microscopic images containing microstructural details. Zhang et al. [13] utilized a CNN neural network to detect and categorize four types of heat-resistant steel structures in images. The database they used consisted of 2717 heat-resistant steel images, using the following four categories: austenite, bainite, tempered martensite, ferrite, and pearlite. The data were preprocessed by numerical normalization and size standardization, a convolutional neural network was trained, and positive prediction results were achieved. Kesireddy et al. [14] achieved positive results by training a radial basis neural network to recognize different steel phases, such as pearlite, ferrite, martensite, and cementite. The dataset they used categorized steel alloys into three groups: ASTM 1038 steel for pearlite, carbon steel for ferrite, and damascus steel for martensite and cementite.

Indeed, metallography is used for a wide range of materials. In the process of metallographic analysis, it is essential to establish a comprehensive database containing all of these materials. However, many studies suffer from a limited variety of training samples. Additionally, different microscopic images of the same material can exhibit significant variations in their characteristics. Therefore, further classifying microscopic images into basic material types and the conditions of the acquirement of the metallography comes to be an important task. For the observation of metallographic images, various types of microscopes can be used, including optical microscopes, SEMs (scanning electron microscopes), TEMs (transmission electron microscopes), etc. These microscopes offer a wide range of magnification options, from 1 to 10^7 . Unlike optical microscopes, SEMs and TEMs use electron beam to observe images separately. The imaging mechanisms of these two are different, resulting in significant differences in microstructure in the obtained microscopic images. The depth of field differs greatly between optical microscopy and scanning electron microscopy [15]. The preparation of metallographic specimens requires steps such as polishing, sectioning, and corrosion, and different etchants [16] can also affect the profiles and contrast of microscopic images. Microscopic images exhibit substantial variations in features at different magnifications. At high magnification, the field of view may become smaller, limiting the comprehensive information on microstructure. However, the increased resolution allows for more details to be exposed, such as fine phases [17]. In addition, different compositions may also alter the microstructure of materials, resulting in significant differences in microscopic images. The same material may still have different phases under different chemical compositions. Liao et al. [18] investigated the effect of Cu content on the microstructure and properties of 7XXX series alloys and found that as the copper content increased, the number of large second-phase particles increased. Therefore, various compositions, data sources, microscope magnification, and the use of etchants play important roles in image recognition.

In this paper, images from different microscopes and magnifications, materials, and corrosion etchants were meticulously classified based on their features. It aims to achieve the fine recognition of materials, thereby enabling the prediction not only of material types but also of factors related to metallographic images.

2. Modified MobileNetV2 Model

2.1. Network Structure

The network designed in this paper is illustrated in Figure 1. The model is mainly based on MobileNetV2 [19], with pooling dropout and dense layers affiliated to it before output. This MobileNetV2 networks are first pre-trained using ImgeNet [20], a public image database aimed at enhancing the training of image recognition software, with over 14 million assorted images such as animals, plants, and objects which are not related to microstructure. The model in this study, including MobileNetV2, ImageNet, Softmax, ReLU, etc, is not an independent software, it is included in the TensorFlow library, and the version number and the information of the library is listed in Table 1. This pre-training is widely used in many professional applications and has proved to be useful, although the images in the database are not professional [21]. The whole model is trained with metallographic images as inputs, with the network parameters of MobileNetV2 kept unchanged; i.e., only the parameters of the dense layer are trained.

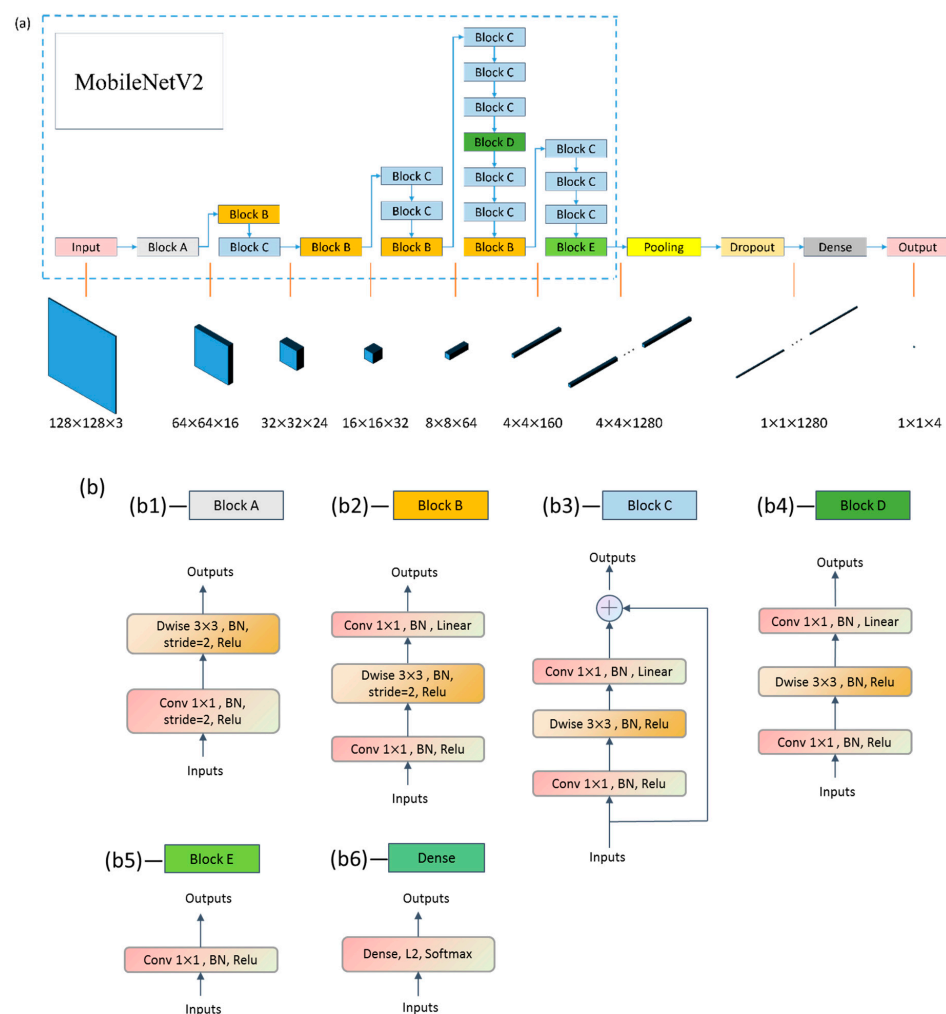


Figure 1. Network structure in this paper: (a) classification network structure based on MobileNetV2; (b) block and dense structure. (b1–b6) represent Block A, Block B, Block C, Block D, Block E, and Dense module in a, respectively. The different colors have no special meaning, only for differentiation and eye-catching purposes.

Table 1. Experimental environment and parameter configuration.

Items	Contents
CPU	Intel® Core™i7-10700 CPU
Operation system	Windows 11
Libraries	TensorFlow 2.4
Python version	Python 3.7

The MobileNetV2 model, proposed by M. Sandler et al. [19], is a convolutional neural network designed specifically for image inputs and outputs. It can be cited directly from the TensorFlow library. The MobileNetV2 model comprises 18 modules, as shown in Figure 1a,b. Each module essentially consists of two initial convolutional layers and a depthwise separable convolutional layer in the middle. The MobileNetV2 model is followed by an average pooling module, a dropout regularization operation, and a flattened layer. Finally, an output layer is connected. The flattened layer uses the Softmax activation function, and L2 regularization is also applied.

2.2. Network Operation

MobileNetV2 utilizes convolution operations to extract input features. The core of convolution operation is to use the convolution kernel and the matrix corresponding to the position of an image for point product operation. Generally speaking, the convolutional kernel moves along the length and width dimensions while maintaining consistency with the size of the feature map in the depth dimension. Depthwise separable convolution involves extracting the features of each channel in the depth dimension using a convolution kernel according to a specific pattern and then integrating the results using a 1×1 convolution kernel.

In deep learning, overfitting can occur when there are too many training parameters but insufficient training samples. This leads to high training accuracy but low prediction accuracy. To address this issue, regularization operations are introduced [22]. The goal of regularization is to enhance model complexity by adding information in order to prevent overfitting. In this paper, dropout [23], L2 regularization [24], and BN (batch normalization) [25] are employed as regularization techniques. Dropout [23] is a regularization operation that randomly drops a certain proportion of connections between neurons in each training iteration. Each convolutional layer of the MobileNetV2 network is connected with a BN operation, a dropout layer is connected behind the MobileNetV2 network, and L2 operation is added to the Dense layer. The specific operation involves inserting the following equation between two consecutive computations.

$$\tilde{y} = y^i \times r^i \quad (1)$$

where, y^i represents the i -th neuron in the next layer, and r^i is a randomly generated binary value (0 or 1). By using this equation, each neuron in the specific layer that needs regularization is processed one by one. When r is 0, this indicates that the neuron is dropped, and when r is 1, it indicates that the neuron is kept.

L2 regularization [24] is achieved by introducing a regularization term with a factor α in the objective function to reduce the impact of less significant features. The equation below represents L2 regularization.

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m (\theta^T \cdot x^{(i)} - y^{(i)})^2 + \alpha \sum_{i=1}^n \theta_i^2 \quad (2)$$

where m is the number of samples, and $\theta^T \cdot x^{(i)}$ and $y^{(i)}$ represent the network's computation result and the true value, respectively. α is the regularization factor, which can be set manually. θ_i represents the linear weights used for the summation of the inputs applied

during the addition process. The goal of training is to minimize the loss function $J(\theta)$ as much as possible.

BN (batch normalization) [25] regularization involves normalizing the samples within the same batch across three dimensions: the dimensions of the sample itself and the batch dimension, which means that the data of each dimension of a three-dimensional object is composed of feature maps with length and width dimensions within the same batch. It performs the following computation on all values within the same dimension. Each dimension is operated independently, and the operation form is shown in the following equation:

$$x_i = \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \varepsilon}} \quad (3)$$

where x_i represents each value, μ_β and σ_β^2 represent the mean and variance of all values in that dimension, and ε is a small constant.

Dense layers typically consist of fully connected layers and activation functions, and they may also include regularization operations. In a fully connected layer, each neuron is connected to all the neurons in the previous layer, transforming high-dimensional data into low-dimensional data. In the convolutional or fully connected stage, activation functions are used to selectively activate neurons, enabling the modeling of non-linear relationships. The main choices for activation functions in this paper are Softmax and ReLU. In this study, the activation function used for each convolutional layer is ReLU, while the activation function used for the dense layer is Softmax.

Softmax is a normalization function represented by the following equation:

$$\text{Softmax}(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

where x represents the input to function Softmax.

ReLU (rectified linear unit) is an activation function commonly used in neural networks to introduce non-linearity. Which is represented by the following equation:

$$\text{ReLU}(x) = \begin{cases} 0 & (x \leq 0) \\ x & (x > 0) \end{cases} \quad (5)$$

where x represents the input to function ReLU.

A set of metallographic microscopic images with uniformly processed pixel sizes is chosen as the input. The output of the network is a probability distribution representing the different types of materials.

With the probability distribution, the operator can not only obtain the results but also gain an understanding of the confidence level associated with that prediction. This information can be valuable for decision-making or further analysis based on the predicted results.

2.3. Parameter Settings

The TensorFlow library, a Python programming tool, is used for implementation. The TensorFlow library is developed by Google, which is in Mountain View, CA, USA. The device information and parameter settings are shown in Tables 1 and 2, respectively.

Table 2. Parameters used in the modified MobileNetV2 model.

Parameters	Epoch	Batch Size	Validation Frequency	Learning Rate of L2 Regularization	Probability of Dropout Regularization
Value	25	64	1	0.001	0.5

The network is trained using the Adam optimizer and the SparseCategoricalCrossentropy loss function for parameter settings. The SparseCategoricalCrossentropy function is used to calculate the cross-entropy for multi-class classification problems as follows:

$$Q = -\frac{1}{m} \sum_{i=1}^m (y_i \log(f(x_i)) + (1 - y_i) \log(1 - f(x_i))) \quad (6)$$

where m represents the total number of samples, $f(x_i)$ denotes the probability distribution calculated by the neural network's computations, and y_i represents the true probability distribution.

The evaluation metric is `sparse_categorical_accuracy`. During validation, the model provides probability distributions for each input, and the output is determined by selecting the class with the highest probability. The accuracy is computed by comparing the predicted outputs with the true results.

The hardware of the computer is listed in Table 1. The parameters of neural network training and validation are listed in Table 2. The training and validation are alternatively performed, and the frequency of validation is the number of training epochs between two consecutive validations.

3. Training Set Preparation, Training, and Validation

3.1. Data Classification and Preprocessing

Three typical cast irons, namely ductile, gray, and white, were selected for study considering their wide application and their different exhibition of carbon into different profiles of graphite or carbide. To expand the generalization of the deep learning model for totally different kinds of alloys, aluminum alloys were selected for the expansion of the training and prediction sets.

Most of the micrographs are sourced from the ASM database [26], while other data were obtained from research results found on websites. These original images vary from 239×360 to 524×360 . To ensure consistency, they were cropped into multiple images of 128×128 pixels, as shown in Figure 2a. The final dataset consists of 87 images of ductile cast iron, 226 images of gray cast iron, 146 images of white cast iron, and 66 images of cast aluminum, 525 images in total. The typical images of these four categories after cropping are shown in Figure 2b. The dataset was divided into three parts: 360 images in the training set, 90 images in the validation set, and 75 images in the test set.

This dataset is referred to as dataset 1, corresponding to training scheme 1, and is used for training and validation of neural networks.

3.2. Data Augmentation

In order to improve the training accuracy, image augmentation techniques such as rotations, Gaussian noise addition, and mirroring were primarily employed to expand the training dataset. Gaussian noise [27], a random signal that conforms to a Gaussian distribution, was applied with a percentage of 0.1. The black dots added by the Gaussian noise are always far smaller than the graphite nodules so as to avoid big changes in the image. Rotations of 90 degrees, vertical flipping, and horizontal flipping were adopted into the original images. An example of data augmentation is shown in Figure 3. Thus, the training set was expanded five times to a total of 2625 images. This dataset is referred to as dataset 2 corresponding to training scheme 2. The data ratio of the training, validation, and test sets is the same as that of dataset 1.

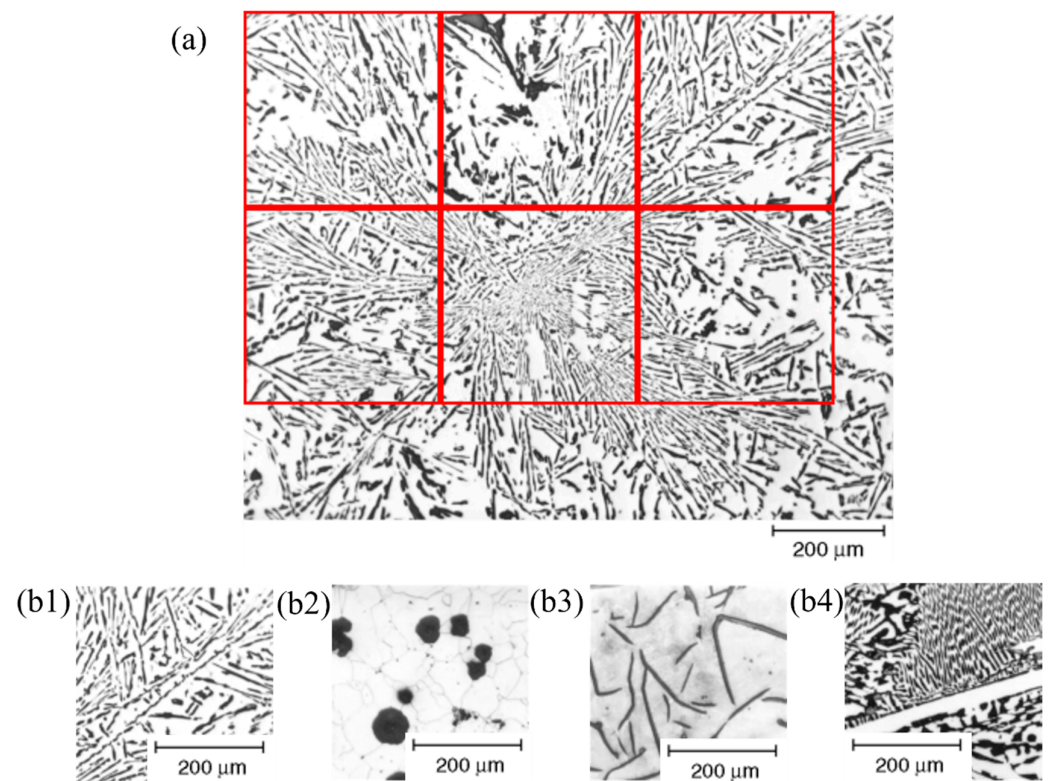


Figure 2. Microscopic images of four classes of materials. (a) Cropping of an original image. (b1–b4) Images of 128×128 . (b1) Cast aluminum. (b2) Ductile cast iron. (b3) Gray cast iron. (b4) White cast iron. The red color in the figure represents the cropping boundary.

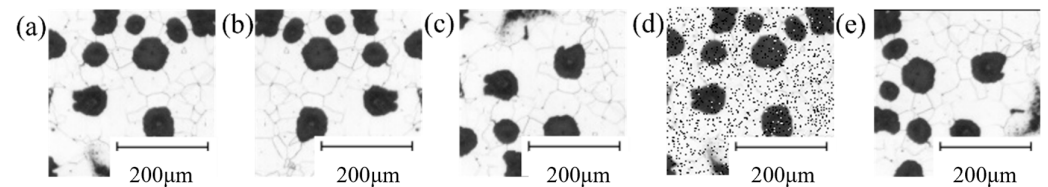


Figure 3. Data augmentation by image rotation, flipping, and Gaussian noise addition. (a) Original image. (b) Counterclockwise rotation of 90° . (c) Horizontal flipping. (d) Gaussian noise addition with a percentage of 0.1. (e) Vertical flipping.

3.3. Fine Classification

In the introduction section, we discussed the impact of various factors on the microstructure of different materials. Therefore, in order to accurately classify the training dataset, it is necessary to carefully identify the compositions, microscopes, magnification, and use of etchants through microscopic images of different materials. The microstructure of a material can be classified into different categories based on these conditions. Meanwhile, the ASM database [26] contains the above information. Thus, based on composition, microscope type, magnification, and etchants, the dataset was further classified into 3, 5, 4, and 3 subcategories for ductile cast iron, gray cast iron, white cast iron, and cast aluminum, respectively. The etchants include nital, picral, etc. The compositions include the contents of carbon and other alloy elements. The cast aluminum was classified into 1xx, 2xx, etc., based on composition. The microscope types primarily consisted of optical and scanning electron microscopes (SEMs). The microscope magnification roughly falls into two groups, $100\times$ and $500\times$. The specific classification scheme is illustrated in Figure 4, referred to as dataset 3 corresponding to training scheme 3. Table 3 lists the classifications and the number of images for each dataset. The data ratio of the training set, validation set, and test set are the same as those of dataset 1.

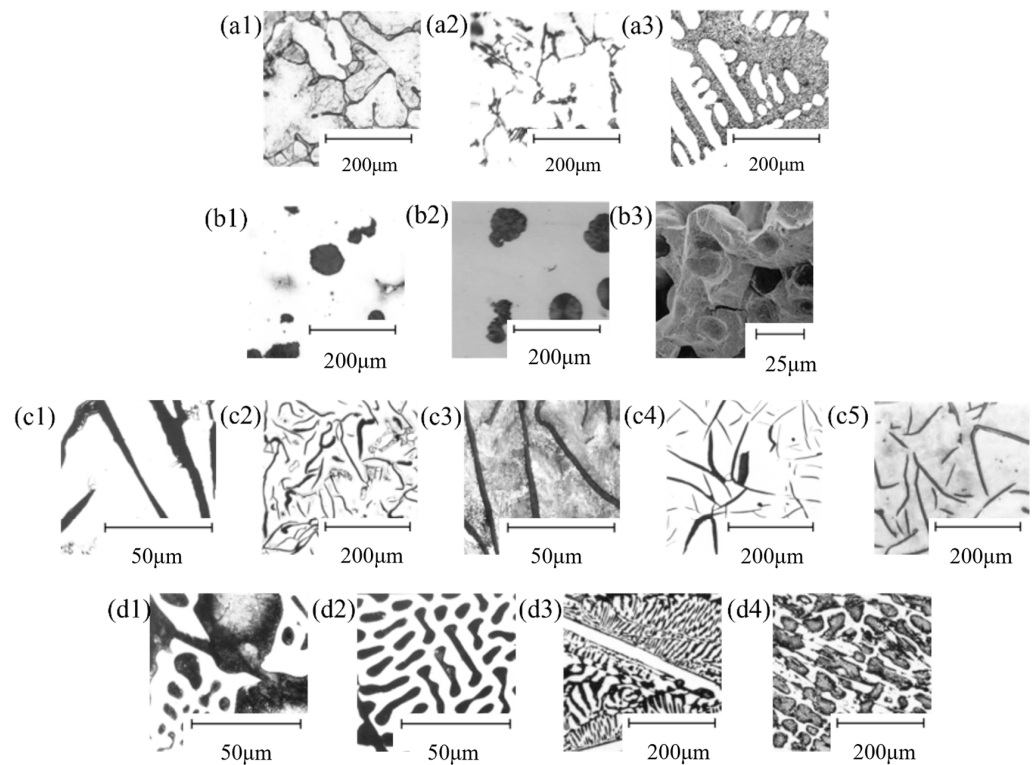


Figure 4. Microscopic images of fine classification of materials. **(a1–a3)** Classification of cast aluminum: **(a1)** 2xx (Al-Cu alloys), 100 \times , OM (optical microscope); **(a2)** 3xx (Al-Mn alloys), 100 \times , OM; **(a3)** 4xx (Al-Si alloys), 100 \times , OM. **(b1–b3)** Classification of ductile cast iron: **(b1)** etched, nital, 100 \times , OM; **(b2)** unetched, 100 \times , OM; **(b3)** 500 \times , SEM. **(c1–c5)** Classification of gray cast iron: **(c1)** etched, nital, 500 \times , OM; **(c2)** etched, nital, 100 \times , OM; **(c3)** etched, picral, 500 \times , OM; **(c4)** etched, picral, 100 \times , OM; **(c5)** unetched, 100 \times , OM. **(d1–d4)** Classification of white cast iron: **(d1)** HCC (high carbon content: C% > 4%), 500 \times , OM; **(d2)** HCC, 100 \times , OM; **(d3)** LCC (low carbon content: C% < 4%), 500 \times , OM; **(d4)** LCC, 100 \times , OM.

Table 3. List of training datasets.

Training Dataset 1 Rough Classification		Training Dataset 2 Rough Classification with Data Augmentation		Training Dataset 3 Fine Classification	
Class	Number of Images	Class	Number of Images	Class	Number of Images
Ductile cast iron	71	Ductile cast iron	360	Etched, nital, 100 \times , OM	228
				Unetched, 100 \times , OM	83
				500 \times , SEM	65
Gray cast iron	192	Gray cast iron	968	Etched, nital, 500 \times , OM	126
				Etched, nital, 100 \times , OM	177
				Etched, picral, 500 \times , OM	204
White cast iron	129	White cast iron	642	Etched, picral, 100 \times , OM	262
				Unetched, 100 \times , OM	194
				HCC, 500 \times , OM	266
				HCC, 100 \times , OM	79
				LCC, 500 \times , OM	80
Cast aluminum	58	Cast aluminum	280	LCC, 100 \times , OM	202
				2xx, 100 \times , OM	74
				3xx, 100 \times , OM	82
				4xx, 100 \times , OM	128

Note: OM—optical microscope, SEM—scanning electron microscopes, HCC—high carbon content: C% > 4%, LCC—low carbon content: C% < 4%, 2xx—Al-Cu alloy, 3xx—Al-Mn alloy, 4xx—Al-Si alloy.

3.4. Training, Validation and Prediction

For the three training schemes, the training parameters remain consistent, as mentioned in Table 2. The trained neural network is utilized for prediction. The networks in three training datasets are used to predict the data that were not included in the training and validation sets. The training, validation, and prediction datasets are listed in Table 4.

Table 4. The number of images in datasets using different training schemes.

The Number of Images in Datasets	Training Dataset 1 Rough Classification	Training Dataset 2 Rough Classification with Data Augmentation	Training Dataset 3 Fine Classification
Training set	360	1800	1800
Validation set	90	450	450
Test set	75	375	375

4. Results

4.1. Training, Validation, and Prediction Results for Established Dataset

The training effects obtained using different training schemes are shown in Figure 5 and Table 5. As the number of epochs increases, the loss rapidly decreases and then decreases gradually. The training accuracy of training scheme 1, training scheme 2, and training scheme 3, respectively, reaches 92.50%, 94.67%, and 95.11%. The validation accuracy of the three training schemes reaches 91.11%, 94.44%, and 94.44%, respectively. Among them, training scheme 3 has the highest accuracy. Scheme 2 shows a significant improvement in accuracy compared to scheme 1, but the difference between scheme 2 and scheme 3 is not significant. The loss of scheme 3 is close to that of scheme 1, but the accuracy is higher than scheme 1. This is because the loss function is SparseCategoricalCrossentropy, which calculates the probability distribution. Scheme 3 has more categories than scheme 1, and there are differences in the probability distribution.

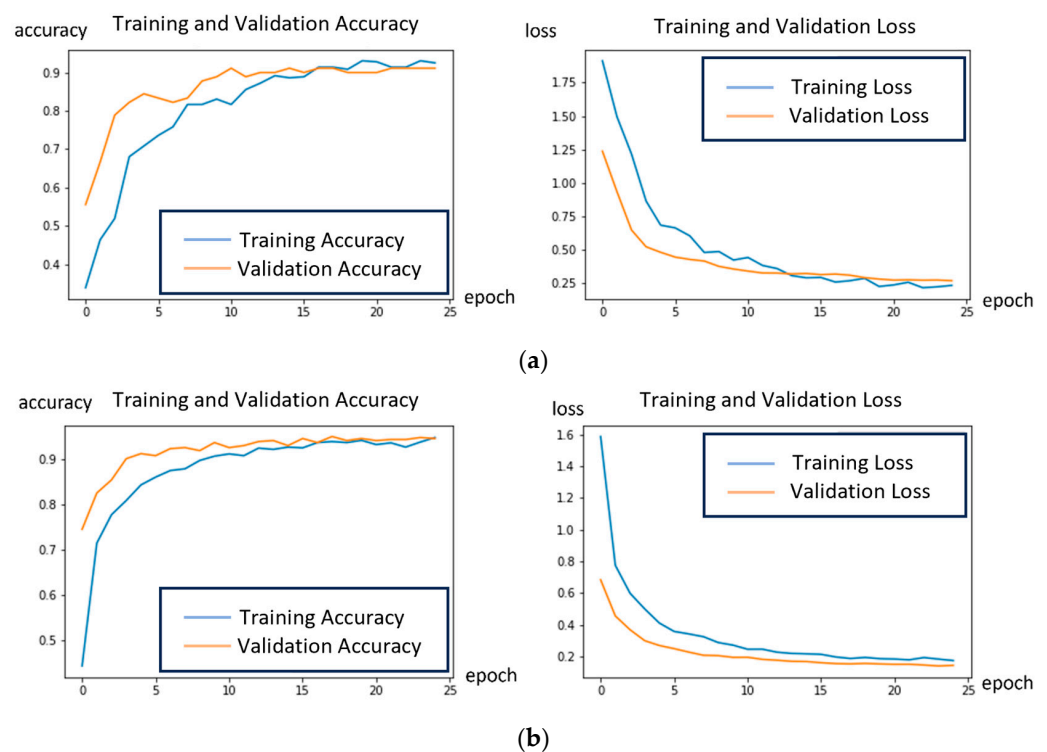


Figure 5. Cont.

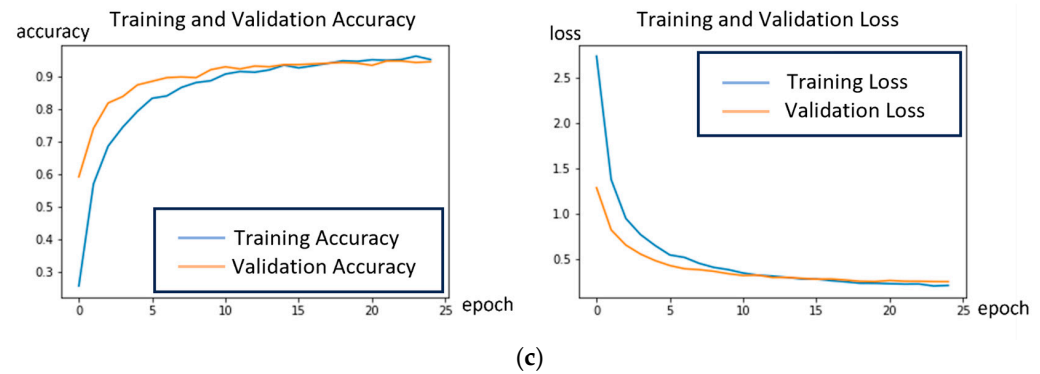


Figure 5. Loss and accuracy curves of 3 training schemes. (a) Loss and accuracy curves of training and validation of scheme 1. (b) Loss and accuracy curves of training and validation of scheme 2. (c) Loss and accuracy curves of training and validation of scheme 3.

Table 5. The training effects using different training schemes.

	Training		Validation		Prediction
	Accuracy	Loss	Accuracy	Loss	Accuracy
Scheme 1 (rough classification)	92.50%	0.2319	91.11%	0.2670	86.67%
Scheme 2 (rough classification with data augmentation)	94.67%	0.1754	94.44%	0.1453	94.44%
Scheme 3 (fine classification)	95.11%	0.2094	94.44%	0.2504	93.87%

The prediction accuracy of all three schemes is higher than 85%, proving the excellent prediction ability of the MobileNetV2 network for microscopic image classification problems. The prediction accuracy of training schemes 2 and 3 is similar, and both of them are higher than the prediction accuracy of training scheme 1.

4.2. Results of Particular Classes

The prediction accuracies of microscopic images for each category corresponding to training scheme 3 are listed in Table 6. The accuracy of all categories of images is higher than 80%, and there are 4 categories with a prediction accuracy of 100%. Among them, the prediction accuracy of unetched, 100 \times , OM ductile cast iron is the lowest, at 80%. From the perspective of material types, ductile cast iron has the lowest prediction accuracy at 86.44%, while white cast iron has the highest prediction accuracy at 98.06%.

Table 6. The prediction accuracy in different classifications.

Class of Rough Classification	Class of Fine Classification	Number of Images	Prediction Accuracy
Ductile cast iron	Etched, nital, 100 \times , OM,	37	86.49%
	Unetched, 100 \times , OM	10	80%
	500 \times , SEM	12	91.67%
Gray cast iron	Etched, nital, 500 \times , OM	24	95.83%
	Etched, nital, 100 \times , OM	38	89.47%
	Etched, picral, 500 \times , OM	36	94.44%
	Etched, picral, 100 \times , OM	38	94.74%
	Unetched, 100 \times , OM	31	93.55%

Table 6. Cont.

Class of Rough Classification	Class of Fine Classification	Number of Images	Prediction Accuracy
White cast iron	HCC, 500 \times , OM	11	90.91%
	HCC, 100 \times , OM	10	100%
	LCC, 500 \times , OM	44	100%
	LCC, 100 \times , OM	38	97.37%
Cast aluminum	2xx, 100 \times , OM	6	83.33%
	3xx, 100 \times , OM	18	100%
	4xx, 100 \times , OM	22	100%

4.3. Application

Images selected from the literature [28,29] were cropped to 128×128 , as shown in Figure 6. They were predicted by the trained models. The prediction accuracies are listed in Table 7. Except for the two images at the bottom of Figure 6a, 20 images participated in the prediction, with a total prediction accuracy of 90%. The prediction of fine classification of cast aluminum is incorrect, mistaking it as another type of aluminum alloy. This may be due to categories of aluminum in the training set being insufficient. Further detailed classification optimization of the prediction results may be needed. All other microscopic images were accurately predicted, indicating that the model can be applied.

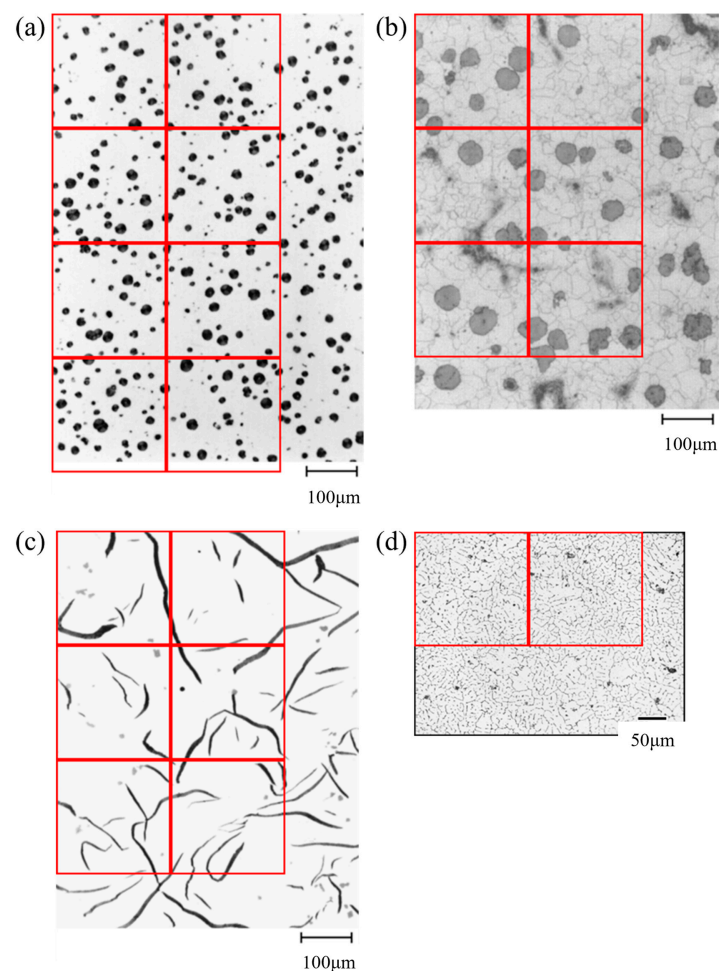


Figure 6. Examples of classification criteria. (a) The original image of unetched 100 \times OM ductile cast iron. (b) The original image of etched nital 100 \times OM ductile cast iron. (c) The original image of unetched 100 \times OM gray cast iron. (d) The original image of 3xx 100 \times OM cast aluminum. The red color in the figure represents the cropping boundary.

Table 7. The prediction accuracy under different conditions.

Class	Material	Microscope Type	Etchant	Magnification	Component
Unetched, 100×, OM ductile cast iron	100%	100%	100%	100%	100%
Etched, nital, 100×, OM ductile cast iron	100%	100%	100%	100%	100%
Unetched, 100×, OM gray cast iron	100%	100%	66.67%	66.67%	100%
3xx, 100×, OM cast aluminum	100%	100%	-	100%	0%

5. Discussion

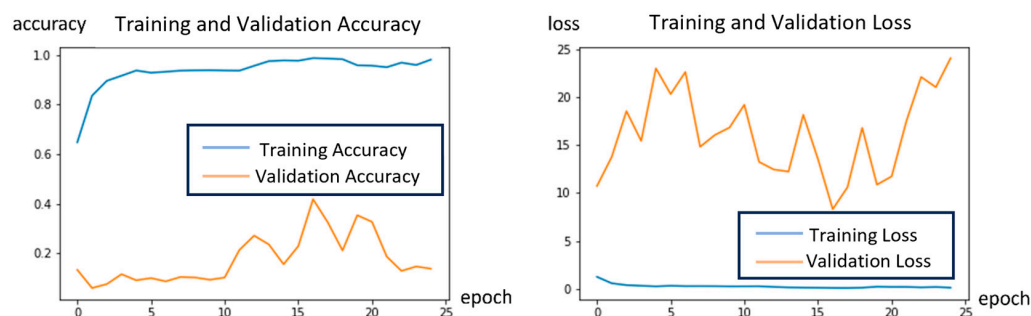
5.1. Effect of MobileNetV2 and Pre-Training

In this study, MobileNetV2 is used as a pre-trained network. Pre-training is the process of training a model through unsupervised learning on a large training set. This study used the ImageNet dataset for pre-training. After training, we fixed the parameters of MobileNetV2 and performed targeted supervised learning, that is, learning through the microscopic image training set of this study. In situations where there are not many known datasets, this training method can enable the model to converge quickly while avoiding overfitting.

Table 8 and Figure 7 compare the training results of MobileNetV2 as a pre-trained model and MobileNetV2 as part of the model for training scheme 3, and the latter shows a significant overfitting phenomenon. This is because MobileNetV2 has 18 blocks, and the number of data for each category in training scheme 3 is not enough. Meanwhile, ImageNet includes various recognizable images, which can be used to transfer and learn the features of microscopic images, thereby improving prediction accuracy. Therefore, it is more appropriate to train MobileNetV2 using a known larger dataset and use it as a pre-training model.

Table 8. The training effect with and without the MobileNetV2 network.

	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
MobileNetV2 as a pre-trained model	95.11%	0.2094	94.44%	0.2504
MobileNetV2 as part of the model	97.78%	0.0913	14.44%	12.3712

**Figure 7.** Loss and accuracy curve of MobileNetV2 as part of the model.

In order to find pre-trained models with better predictive performance, this study selected VGG19 [6], MobileNetV2, and Xception [30] models for comparison. These network models are all recorded in the Tensorflow library. Through comparison, the characteristics of MobileNetV2 can be better discovered. The results are shown in Figure 8 and Table 9. Compared to VGG19, MobileNetV2, and Xception have more complex structures. Both use depthwisely separable convolution and BN regularization operations, while VGG19 only uses general convolution operations. Thus, MobileNetV2 and Xception have better prediction performance. In Xception, residual modules are extensively used instead of activation functions. MobileNetV2 utilizes the ReLU activation function to preserve low dimensional input information as much as possible, resulting in better prediction performance.

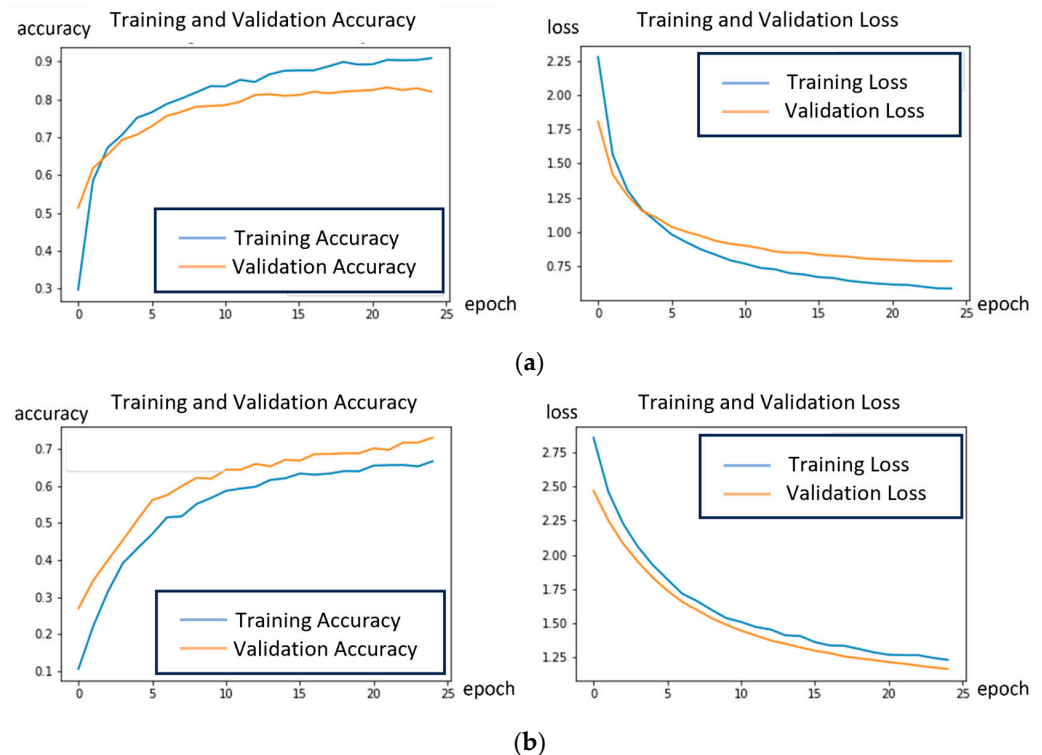


Figure 8. Loss and accuracy curves of different networks. (a) Loss and accuracy curve of Xception network. (b) Loss and accuracy curve of VGG19 network.

Table 9. Comparison of the training effect with the MobileNetV2 network, the Xception network, and the VGG19 network.

	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss	Prediction Accuracy	Number of Parameters	Training Time (s)
MobileNetV2	94.76%	0.2373	94.44%	0.2590	93.87%	2,263,108	284.69
Xception	90.83%	0.5863	82.00%	0.7864	81.33%	20,892,215	905.07
VGG19	66.72%	1.2328	73.11%	1.1666	70.93%	20,032,079	1761.58

5.2. Visual Interpretation of the Model

Grad CAM, proposed by Selvaraju et al. [31], explains how neural networks perform during image recognition. Its main calculation method is to obtain gradients in various positions of a convolutional layer feature map of the trained model. In this study, the convolutional layers are all from MobileNetV2, and their parameters remain unchanged during the training process. Therefore, for each class in the test set of training scheme 3, one image was selected, and the Grad CAM values were calculated under the pre-trained MobileNetV2 network. The results are shown in Figure 9. The red color indicates a high CAM value, and the red areas indicate areas that the neural network can clearly recognize. It can be seen that when the features are large and continuous, the neural network can easily recognize them, as with ductile iron. When features are dispersed, the main features are easily mistaken for noise, as with gray cast iron.

For each category in Section 4.2, Figure 9 provides some reasonable explanations. For various materials with different magnifications, the features are more prominent at high magnifications, and the features are extracted at low magnifications, so this information can be effectively recognized. For white cast iron, there is a significant difference in the microscopic features extracted by neural networks for images with different carbon contents, which are similar to fishbone and thick segments, respectively. Therefore, the prediction accuracy of this material is high. For ductile cast iron, the features extracted from images with different etchants are all circular in size, so the prediction accuracy is low.

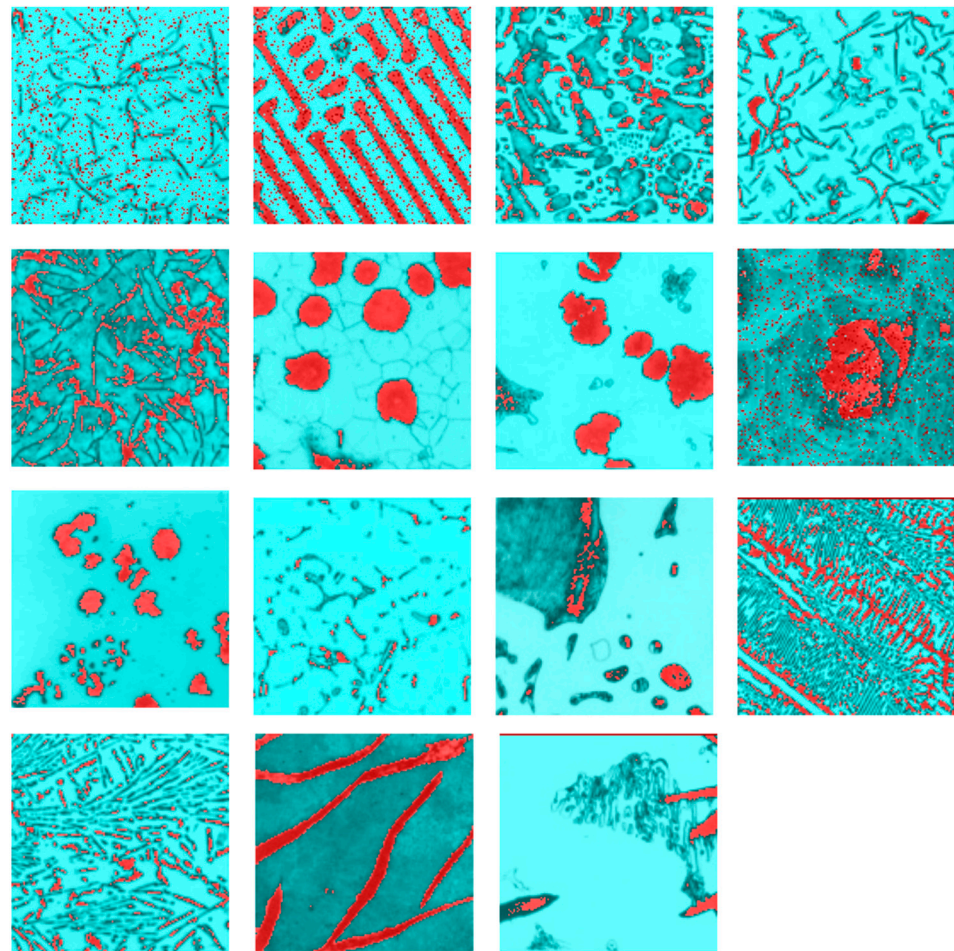


Figure 9. Grad CAM images of different materials. The red position in the figure has a higher value for Grad CAM.

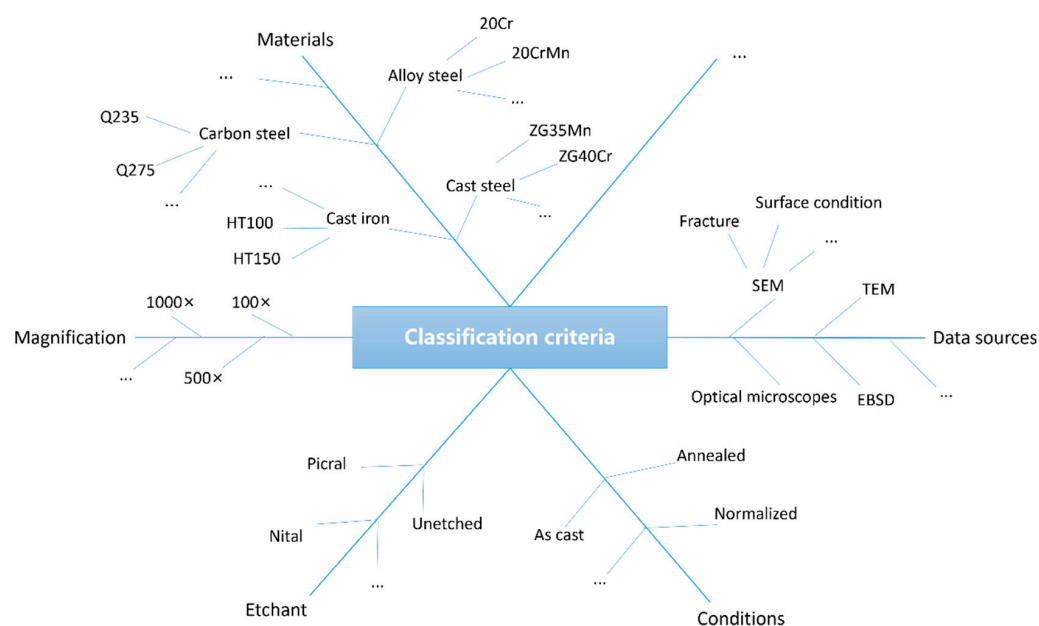
5.3. Satisfaction of the Actual Requirement

Through fine classification, training accuracy slightly increases and prediction accuracy remains high. This preliminary evidence demonstrates that the network can effectively predict more detailed information for different microstructure images. In future research, the classification ability of the network can be enhanced by adding more categories that include various material information in the training set. The classification can be based on material or other information related to microstructural features in the process of metallographic identification.

The prediction accuracy of MobileNetV2 on the training set of 53 kinds of fruits and vegetables can reach 96.23%, indicating its high prediction performance in dozens of categories [32]. Due to pre-training on the ImageNet dataset, MobileNetV2 can achieve high prediction accuracy even when a small number of images in each category are used during training. When the category of the training set exceeds that of the ImageNet dataset, its accuracy may decrease significantly. To solve this problem, different methods can be chosen, for example, using network structures with more parameters for pre-training, increasing the number of training sets, using other training sets for transfer learning, and modifying network models used outside of pre-training. Additionally, changing the network structure can also improve the prediction performance for multiple categories. Liu et al. summarized different network structures used for multi-label classification, such as embedding methods, which compress label vectors, and tree-based methods, which hierarchically divide categories [33].

Metal materials commonly have various types of microstructures or phases, and different materials possess different microstructures under different heat treatment processes. They are often categorized based on crystal phases and morphological features. The data sources are also diverse, including observations from different devices such as optical microscopes or scanning electron microscopes. Various parameters of the devices can differ as well. Observation modes include bright-field images and dark-field images, which are totally different in their foreground and background. Magnification ranges from several times to tens of millions of times, such as of coarse cast grains bigger than 10 mm, subgrains at the micron level, and precipitated metallic compounds around the nanometer level, and the observed features include grains, grain boundaries, dendrites, and dislocation at different magnifications.

During classification, multiple pieces of information can be utilized together to describe a category. A training set based on microscopic image information can be established. By training on this dataset, more detailed information about the materials can be directly obtained. Alternatively, metallographic images can be classified hierarchically, with each level representing a specific piece of information. For example, the parent category can represent the overall structure, and the subcategories can represent observation methods. This approach provides clearer identification results, and the neural network structure can be modified to output multiple levels of information step by step. In addition, different classification criteria can be presented in parallel, for example, the predicted results based on observation patterns and magnification levels can be output simultaneously. Figure 10 presents a chart of different classification criteria and specific examples. Some types of data may have minimal differences in features, while others may suffer from insufficient images. These factors contribute to the increased difficulty in training and predicting with deep learning. Further exploration is needed to delve into these aspects in the future.



Note: The material brand is based on Chinese standards.

Figure 10. The chart of classification criteria.

6. Conclusions

A modified MobileNetV2 model was constructed with the following regularization techniques and a flattening layer. Datasets of three main types of cast irons including ductile, gray, and white cast irons and aluminum alloy were established with an image size of 128×128 pixels. These four materials were further categorized into 15 subcategories

based on their composition, data sources, microscope magnification, and use of etchants for materials.

- (1) The network exhibits high accuracy during both training and prediction and relatively successfully predicts detailed information including composition, microscope, magnification, and etchant. The accuracies of training, validation, and prediction for fine classification with data augmentation respectively reach 95.11%, 94.44%, and 93.87%. In each category of the fine classification, the prediction accuracy is higher than 80%.
- (2) The training effect can be influenced by the classification level and the data augmentation method. The method of data augmentation can significantly increase training accuracy. Under the conditions of the pre-trained MobileNetV2 network and a dataset with a relatively small quantity and category used in this study, the accuracy of prediction is acceptable.
- (3) ImageNet, as a pre-trained dataset, greatly improves the training and prediction performance of MobileNetV2 networks. Under the same condition, MobileNetV2 has better prediction accuracy than Xception and VGG19 networks.
- (4) This study shows that neural networks can be used to recognize metallographic images not only by their material type but also by more details, such as composition, microscope, magnification, the use of etchants, etc.

Author Contributions: Conceptualization, J.K.; methodology, Q.Z. and K.W.; validation, Q.Z.; formal analysis, J.K. and Q.Z.; investigation, Q.Z. and K.W.; resources, K.W.; data curation, K.W. and Q.Z.; writing—original draft preparation, Q.Z.; writing—review and editing, J.K.; visualization, Q.Z.; supervision, J.K.; project administration, J.K.; funding acquisition, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: The research is sponsored by the Tsinghua-Toyota Joint Research Fund and Key Technologies R&D Program of Guangdong Province (2022B0909070001).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: Shichang Cheng, Mengqi Jiao, Xinyi Li, Chao Li: software; Yanrui Tang, Chiyuan Wang: investigation.

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

References

1. Smallman, R.E.; Ashbee, K.H. *Modern Metallography: The Commonwealth and International Library: Metallurgy Division*; Elsevier: Amsterdam, The Netherlands, 2013.
2. Bharadiya, J.P. A Comprehensive survey of deep learning techniques natural language processing. *Eur. J. Technol.* **2023**, *7*, 58–66. [\[CrossRef\]](#)
3. Thiéry, A.H.; Braeu, F.; Tun, T.A.; Aung, T.; Girard, M.J.A. Medical application of geometric deep learning for the diagnosis of glaucoma. *Transl. Vis. Sci. Technol.* **2023**, *12*, 23. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Conze, P.-H.; Andrade-Miranda, G.; Singh, V.K.; Jaouen, V.; Visvikis, D. Current and emerging trends in medical image segmentation with deep learning. *IEEE Trans. Radiat. Plasma Med. Sci.* **2023**, *7*, 545–569. [\[CrossRef\]](#)
5. Liu, F.; Chen, D.; Wang, F.; Li, Z.; Xu, F. Deep learning based single sample face recognition: A survey. *Artif. Intell. Rev.* **2022**, *56*, 2723–2748. [\[CrossRef\]](#)
6. Zhen, X.; Chen, J.; Zhong, Z.; Hrycushko, B.; Zhou, L.; Jiang, S.; Albuquerque, K.; Gu, X. Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: A feasibility study. *Phys. Med. Biol.* **2017**, *62*, 8246. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Hinton, G.E.; Osindero, S.; Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *2*, 2672–2680.
9. Zhong, P.; Gong, Z.; Li, S.; Schonlieb, C.-B. Learning to diversify deep belief networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3516–3530. [\[CrossRef\]](#)
10. Zhang, K.; Guo, Y.; Wang, X.; Yuan, J.; Ding, Q. Multiple feature reweight densenet for image classification. *IEEE Access* **2019**, *7*, 9872–9880. [\[CrossRef\]](#)

11. Wang, J.; Fan, Y.; Li, Z. Texture Image Recognition Based on Deep Convolutional Neural Network and Transfer Learning. *J. Comput.-Aided Des. Comput. Graph./Jisuanji Fuzhu Sheji Yu Tuxingxue Xuebao* **2022**, *34*, 701–710. [CrossRef]
12. Chowdhury, A.; Kautz, E.; Yener, B.; Lewis, D. Image driven machine learning methods for microstructure recognition. *Comput. Mater. Sci.* **2016**, *123*, 176–187. [CrossRef]
13. Zhao, Q.; Kang, J.; Wu, K.; Cheng, S.; Jiao, M.; Tang, Y.; Wang, C.; Li, X.; Li, C. Research on Self-organized CNN Modeling to Identify Metallographic Structure of Heat-resistant Steel. *Mater. Rep.* **2022**, *36*, 21030032-1–21030032-6.
14. Kesireddy, A.; McCaslin, S. Application of image processing techniques to the identification of phases in steel metallographic specimens. In *New Trends in Networking, Computing, E-Learning, Systems Sciences, and Engineering*; Springer International Publishing: New York, NY, USA, 2015.
15. Vernon-Parry, K. Scanning electron microscopy: An introduction. *III-Vs Rev.* **2000**, *13*, 40–44. [CrossRef]
16. Sun, H.; Dong, J.; Liu, F.; Ding, F. Etching of two-dimensional materials. *Mater. Today* **2021**, *42*, 192–213. [CrossRef]
17. Sellaro, T.L.; Filkins, R.; Hoffman, C.; Fine, J.L.; Ho, J.; Parwani, A.V.; Pantanowitz, L.; Montalto, M.; JI, F.J.H. Relationship between magnification and resolution in digital pathology systems. *J. Pathol. Inform.* **2013**, *4*, 21. [CrossRef]
18. Liao, Y.-G.; Han, X.-Q.; Zeng, M.-X.; Jin, M. Influence of Cu on microstructure and tensile properties of 7XXX series aluminum alloy. *Mater. Des.* **2015**, *66*, 581–586. [CrossRef]
19. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
20. Stanford Vision Lab; Stanford University; Princeton University. ImageNet. Available online: <https://www.image-net.org/> (accessed on 22 November 2023).
21. Li, Z.; Zhu, Y.; Yang, F.; Li, W.; Zhao, C.; Chen, Y.; Chen, Z.; Xie, J.; Wu, L.; Zhao, R.; et al. Univip: A unified framework for self-supervised visual pre-training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022.
22. Friedrich, S.; Groll, A.; Ickstadt, K.; Kneib, T.; Pauly, M.; Rahnenführer, J.; Friede, T. Regularization approaches in clinical biostatistics: A review of methods and their applications. *Stat. Methods Med. Res.* **2023**, *32*, 425–440. [CrossRef]
23. Nitish, S. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
24. Jaiswal, S.; Mehta, A.; Nandi, G. Investigation on the effect of L1 and L2 regularization on image features extracted using restricted boltzmann machine. In Proceedings of the 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 14–15 June 2018.
25. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning PMLR 2015, Lille, France, 7–9 July 2015.
26. ASM INTERNATIONAL. ASM Materials Information ASM Micrograph Database. Available online: <https://matdata.asminternational.org/mgd/index.aspx> (accessed on 22 November 2023).
27. Wang, J.; Shi, W.; Zeng, X. Optimal wavelet estimators of the heteroscedastic pointspread effects and Gauss white noises model. *Commun. Stat.—Theory Methods* **2022**, *51*, 1133–1154. [CrossRef]
28. Radzikowska, J.M. Metallography and microstructures of cast iron. In *ASM Handbook, Volume 9: Metallography and Microstructures*; ASM International, Ed.; Vander Voort, G.F.: Materials Park, OH, USA, 2004; pp. 565–587. [CrossRef]
29. Geng, Y. Microstructure Evolution during Extrusion of AA3xxx Aluminum Alloys. Doctoral Dissertation, University of British Columbia, New York, NY, USA, 2011.
30. 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.
31. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
32. Shahi, T.B.; Sitaula, C.; Neupane, A.; Guo, W. Fruit classification using attention-based MobileNetV2 for industrial applications. *PLoS ONE* **2022**, *17*, e0264586. [CrossRef]
33. Liu, W.; Wang, H.; Shen, X.; Tsang, I.W. The emerging trends of multi-label learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 7955–7974. [CrossRef] [PubMed]

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.