applied sciences

Article

Vickers Hardness Value Test via Multi-Task Learning Convolutional Neural Networks and Image Augmentation

Wan-Shu Cheng [1], Guan-Ying Chen [2], Xin-Yen Shih [3], Mahmoud Elsisi [4,5], Meng-Hsiu Tsai [3,6,*] and Hong-Jie Dai [2,6,7,8,*]

[1] Department of Computer Science and Information Management, Providence University, Taichung 43301, Taiwan
[2] Intelligent System Laboratory, Department of Electrical Engineering, College of Electrical Engineering and Computer Science, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan
[3] Department of Mold and Die Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan
[4] Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan
[5] Department of Electrical Engineering, Faculty of Engineering at Shoubra, Benha University, Cairo 11629, Egypt
[6] School of Dentistry, College of Dental Medicine, Kaohsiung Medical University, Kaohsiung 80708, Taiwan
[7] School of Post-Baccalaureate Medicine, Kaohsiung Medical University, Kaohsiung 80708, Taiwan
[8] National Institute of Cancer Research, National Health Research Institutes, Tainan 70456, Taiwan
* Correspondence: tmh@nkust.edu.tw (M.-H.T.); hjdai@nkust.edu.tw (H.-J.D.);
Tel.: +886-07-3814526 (ext. 5410) (M.-H.T.); +886-07-3814526 (ext. 15510) (H.-J.D.)

Featured Application: Feature Applications: This paper proposes a data-driven approach based on convolutional neural networks to measure the Vickers hardness value directly from the image of the specimen to get rid of the requirement of the manually generation of indentations for measurement.

Abstract: Hardness testing is an essential test in the metal manufacturing industry, and Vickers hardness is one of the most widely used hardness measurements today. The computer-assisted Vickers hardness test requires manually generating indentations for measurement, but the process is tedious and the measured results may depend on the operator's experience. In light of this, this paper proposes a data-driven approach based on convolutional neural networks to measure the Vickers hardness value directly from the image of the specimen to get rid of the aforementioned limitations. Multi-task learning is introduced in the proposed network to improve the accuracy of Vickers hardness measurement. The metal material used in this paper is medium-carbon chromium-molybdenum alloy steel (SCM 440), which is commonly utilized in automotive industries because of its corrosion resistance, high temperature, and tensile strength. However, the limited samples of SCM 440 and the tedious manual measurement procedure represent the main challenge to collect sufficient data for training and evaluation of the proposed methods. In this regard, this study introduces a new image mixing method to augment the dataset. The experimental results show that the mean absolute error between the Vickers hardness value output by the proposed network architecture can be 10.2 and the value can be further improved to 7.6 if the multi-task learning method is applied. Furthermore, the robustness of the proposed method is confirmed by evaluating the developed models with an additional 59 unseen images provided by specialists for testing, and the experimental results provide evidence to support the reliability and usability of the proposed methods.

Keywords: convolutional neural network; multi-task learning; Vickers hardness test

Citation: Cheng, W.-S.; Chen, G.-Y.; Shih, X.-Y.; Elsisi, M.; Tsai, M.-H.; Dai, H.-J. Vickers Hardness Value Test via Multi-Task Learning Convolutional Neural Networks and Image Augmentation. *Appl. Sci.* **2022**, *12*, 10820. https://doi.org/10.3390/app122110820

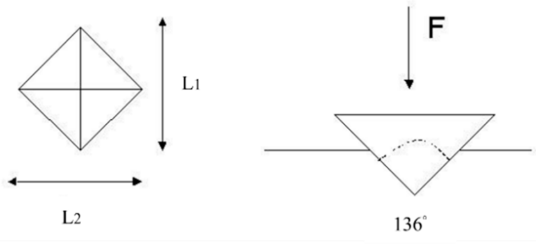
Academic Editor: Christian W. Dawson

Received: 9 September 2022
Accepted: 18 October 2022
Published: 25 October 2022

1. Introduction

Chromium-molybdenum (Cr-Mo) alloy steel has been widely used in fastener, machine parts, shafts, and screw products because of its high strength and hardness properties [1,2]. Different types of heat treatments result in different microstructures of Cr-Mo steel, which affects the mechanical properties, such as hardness, tensile strength, etc. Various researchers have attempted to process Cr-Mo steel via heat treatment in different quench temperatures and cooling methods to obtain the desired microstructure and mechanical properties [3–6]. Cr-Mo alloy has been reported with different compositions and heat treatments, and a different phase has been observed consistent with a perlite and bainitic structure via being normalized and ferrite with a bainitic structure via being tempered in 1Cr-0.5Mo-0.1C alloy and tempered martensite structure via being normalized and tempered in 9Cr-2Mo-Nb-V-0.1C alloy [3,7]. Further, 42Cr-4Mo steel via quenching tempering and step quenching heat treatment produced tempered martensite and ferrite-bainite-martensite of two different microstructures, resulting in different tensile strengths (940 MPa and 633 MPa) [1]. Chen et al. [4] compared the hardness of Cr-Mo (SCM 435) alloy with austempering temperature 830 °C for 25 min and 15 min and salt bath at 290 °C. The specimen with austempering at 830 °C for 25 min and salt bath at 290 °C for 30 min had a bigger grain size of austenite and retained γ phase. It also had higher hardness and strength. The Vickers hardness test [8] is commonly used to evaluate key mechanical properties of materials. Evaluation of the hardness value (HV) of a metallic material enables determination of its wear resistance and the approximate value of its ductility, cut, and flow tension [9].

For industrial use, both metallography and hardness/micro-hardness are used to evaluate the microstructure and mechanical properties after heat treatment. However, because the mass effect [10] and different parameter settings [1] during heat treatment could result in different mechanical properties, analysis of the complex microstructure is challenging [11]. In HV testing, the conventional manual measurement relied on the operator's experience to interpret the indentation on the surface of the material made by a pyramidal diamond indenter through microscopy, as illustrated in Figure 1b. Because of the high number of tests performed by the operator, tiredness may also lead to reading of incorrect values and, consequently, obtaining incorrect HVs.



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

Figure 1. (**a**) The Vickers hardness indentation generated by the tester; (**b**) the side view of the diamond indenter.

Computer-assisted hardness testing systems have been developed to provide more accurate measurements of micro-hardness based on the Vickers hardness test. The conventional automatic methods relied on determination of the diagonal length of the indentation from the edge lines, as illustrated in Figure 1a. Araki et al. [12] proposed to ignore any point deviating from the edge lines of the indentation image to obtain the diagonal length of the indentation. However, the method is sensitive to the surface properties of the specimen and limited with respect to the types of specimens to which they could be applied. In order to address this issue, many methods have been proposed. Sugimoto et al. [11] developed an algorithm to determine the threshold level for generation of binary image data of the indentation based on the brightness histogram profile of the indentation and applied the least-square method to estimate the vertex coordinates of the indentation to estimate the

HV. With advancement in image processing technology, mathematical morphology methods [13], such as erosion, extension, skeleton extraction, and edge detection, have also started to be applied in engineering inspection. Maier et al. [14] investigated several image sharpness methods for improving Vickers hardness measurement and suggested that the Brenner autofocus function provides the best compromise between computational effort and accuracy.

However, typical image processing approaches often fail to detect indentations in materials with poor surface uniformity, rough surfaces, surfaces with noise patterns, distorted indentation shapes, or cracks, leading to accurate detection and segmentation of indentations in Vickers images remaining challenging. In order to address the issue, data-driven and machine learning approaches have recently been proposed. For example, Hu et al. [15] proposed a microstructure-informatic strategy with random forests and linear LASSO models to predict HVs. Swetlana et al. [9] extracted microstructure features through image processing tools and developed a machine learning model to measure hardness. Jalilian et al. [16], Tanaka et al. [17], Seino et al. [18], Chen et al. [19], and Li et al. [20] proposed to leverage the convolutional neural network (CNN) [21] to obtain location information in the convolutional layer, representing the edges or corners of feature vectors for accurate localization and segmentation of Vickers indentations (see Figure 1a). However, these methods still relied on the indentations generated by forcing a pyramidal diamond indenter, which may result in noisy surfaces containing cracks, sparkles, and other distortions in the created images.

The purpose of this study is to use deep-learning-based methods to measure Vickers hardness by directly analyzing optical images of the SCM 440 steel. Considering the indentation procedure illustrated in Figure 1, it is time-consuming and requires trained operators, noisy surfaces, distorted indentation shapes, and cracks of the created Vickers images affect the accuracy and robustness of automated methods developed in previous works. In this regard, we propose a data-driven approach to develop a neural network model that can directly analyze the pixels in an image, capturing geometrical patterns in microstructures to measure Vickers hardness. The feasibility of the proposed method is evaluated on a real dataset specifically collected for this study, along with a report of the performance comparison among several state-of-the-art CNN models applied for the task.

2. Materials and Methods

2.1. The Material Used in This Study and the Dataset Generation Process

The SCM 440 supplied by the China Steel company was used as the experimental material studied in this paper. SCM 440 is a Japanese grade of structural steel for machinery manufacturing. According to JIS (Japanese industrial standard), the alphabet SCM indicates that the major alloying elements of the steel are Cr-Mo. The following numeric codes represent the main alloying element content and average carbon content. The material is widely used in a variety of devices, such as bolts, shafts, spindle barrels, gears, motorcycle components, etc. Metallographic microscopy is mainly used to identify and analyze the internal structure and organization of metals such as SCM 440. In study, the optical image samples were etched with 3% Nital solution. Vickers micro-hardness (500 g, 15 s hold) of 15 individual measurements in each optical image were performed on the metallographic specimens using an Akashi MVK-H100 (Osaka, Japan) machine. Figure 2 shows example results of triangular pyramidal indentations obtained by the Vickers hardness test [8].

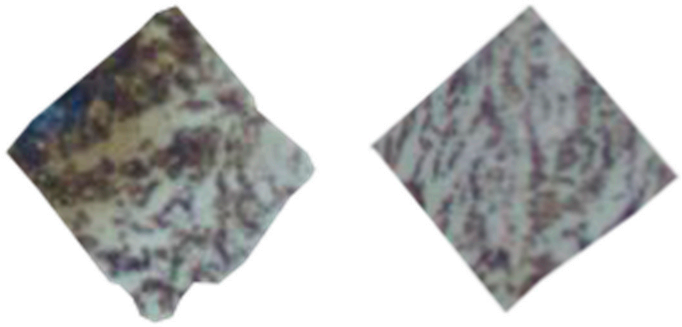
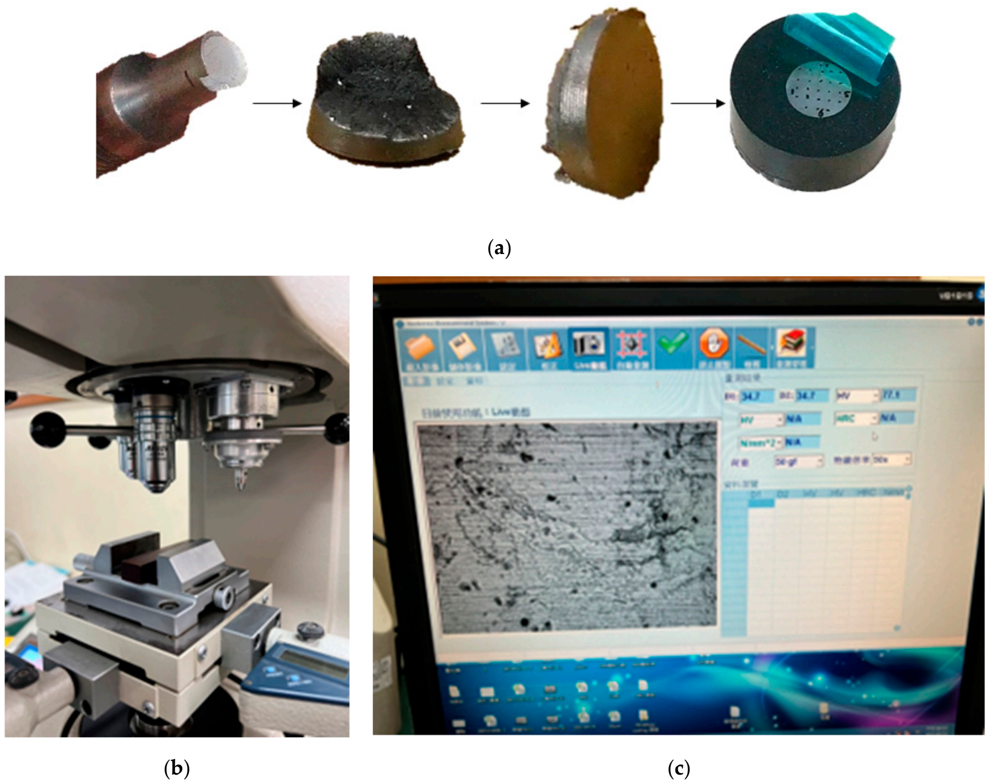


Figure 2. Example results obtained by the Vickers hardness assessment method for SCM 440.

Figure 3 shows the data generation process applied in this paper, in which the Vickers hardness tester (MVK-H100) is used for the Vickers hardness testing manually to generate the training and test datasets consisting of figures as shown in Figure 2, which were used by the computer-assisted method for calculating the HVs. In general, the harder the specimen is, the smaller the size of the indentation is and vice versa. The entire process is elaborated as follows.



(**a**)

(**b**)

(**c**)

Figure 3. The data generation process applied by this study. (**a**) The sample used for tensile testing; (**b**) the Vickers hardness tester used in this study; (**c**) the usage of the Vickers hardness tester software.

The sample was generated by tensile testing for the material under test. After etching with a 3% Nital solution, the Vickers hardness tester was used to produce the specimen images of the sample shown in Figure 2 by pressing the indenter into the surface of the sample with a certain load and holding the load for a certain period to form an indentation, as illustrated in Figure 1a, where L_1 and L_2 are the diagonal lengths of the indentation. Figure 1b shows the side view of the diamond indenter, which has a relative angle of 136°. The image of the indentation displayed on the screen of Figure 3c was photographed by an optical microscope. Finally, the micro-hardness (the HV) was obtained by measuring the diagonal lengths of the square indentations (L_1 and L_2) on the surface of the captured image, as shown in Figures 1a and 3c, calculating the average value and applying it to Equation (1):

$$\mathrm{HV} = \frac{F}{S} = \frac{2F * \sin\left(\frac{136^\circ}{2}\right)}{g * d^2} \cong 0.1891 * \frac{F}{d^2}\left[\mathrm{N/mm^2}\right], \tag{1}$$

where F represents the load (in Newton force) and S represents the calculated area. The indenter angle is 136°, d is the average diagonal length of the indentation, and g is the acceleration of gravity.

Because the samples of the SCM 440 material are limited and the aforementioned manual generation process of the indentation image for the specimen and its corresponding HV is extremely tedious, only 105 indentations were obtained in this study. Figure 4 illustrates an example of the original specimen; the microstructure that consists of ferrite, banite, pearlite, and carbide was shown in different contrast in Figure 4a and the same specimen after the Vickers hardness test in Figure 4b. Dark and grey contrast areas present pearlite and banite microstructure, while lighter regions include a mixture of microstructures, mainly ferrite and spherical cementite, as shown in Figure 4a. Each microstructure has different hardness; ferrite is soft and banite is hard. It is difficult to test only one microstructure due to the size limitation of the indentation.

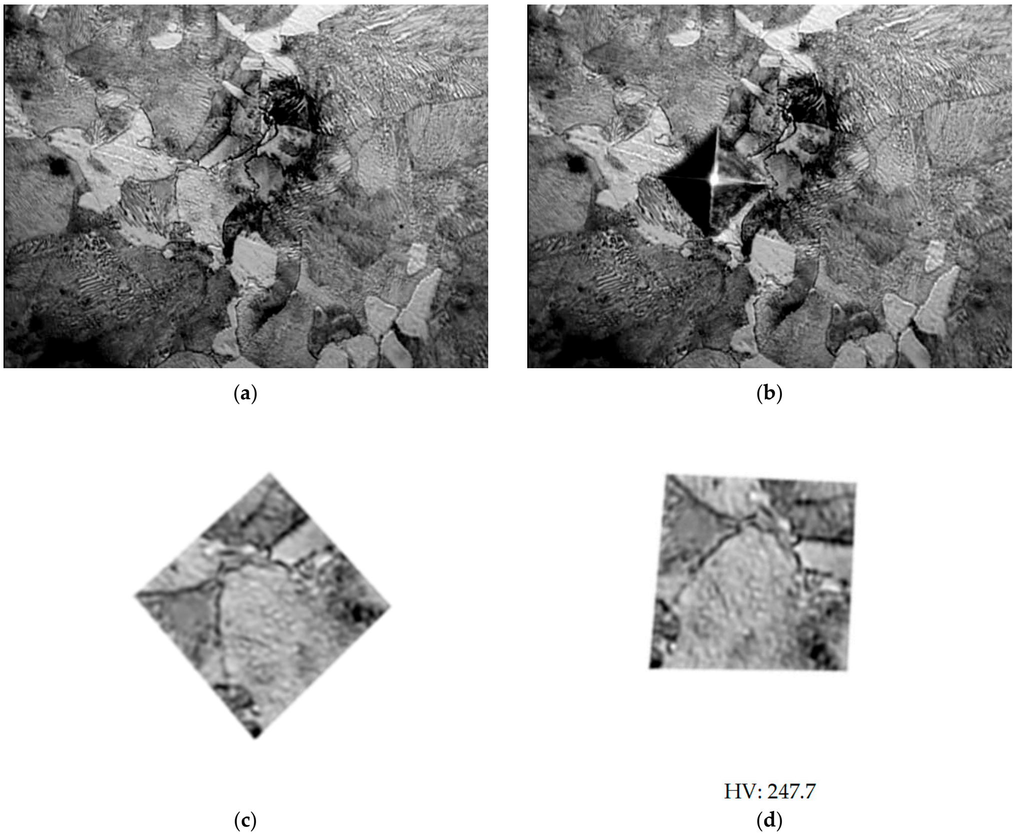


Figure 4. The training sample generation process applied by this study. (**a**) The original image of a specimen captured by a microscope; (**b**) the image of the same specimen after the Vickers hardness test; (**c**) the indentation extracted by the image processing technology; (**d**) the final rotated image and its corresponding HV.

The data collected for the study were divided into two stages owing to the difficulty of gathering the samples. In the first stage, only 46 images were collected. The data augmentation method described in the next subsection was applied to synthesize the training, validation, and test sets consisting of 5000, 500, and 100 mixed images, respectively. More detail is described in Section 2.6.1 "Experimental Dataset". In the second stage, we received 59 new images, which were considered an additional unseen test set. After collecting a total of 105 images, as in Figure 4a,b, the image of the indentation illustrated in Figure 4c was extracted from the original image by using a computerized image processing technology, and then the diamond was rotated to create a square as shown in Figure 4d. For each image, we recorded the corresponding HV, and the resulting image with the corresponding HV was collected as the training data for this paper.

2.2. Image Data Augmentation

To enable effective model training, we applied the image mixing data augmentation method to artificially augment the size of the training data. The square indentation image shown in Figure 4d was merged into an image of similar size to the original 50× optical microscope image. Because of the different sizes of the diamond indentation images, it was found after analysis that a random combination of about 16 images would produce an image with a similar size to that originally taken with the 50× optical microscope. In our implementation, the images were randomly selected to form the images depicted in Figure 5. The original HVs of the 16 images were averaged to form the final HV of the image. In addition to the random selection of images, we also rotated the images randomly to make the combined images different so as to generate a large number of example images used in subsequent experiments.

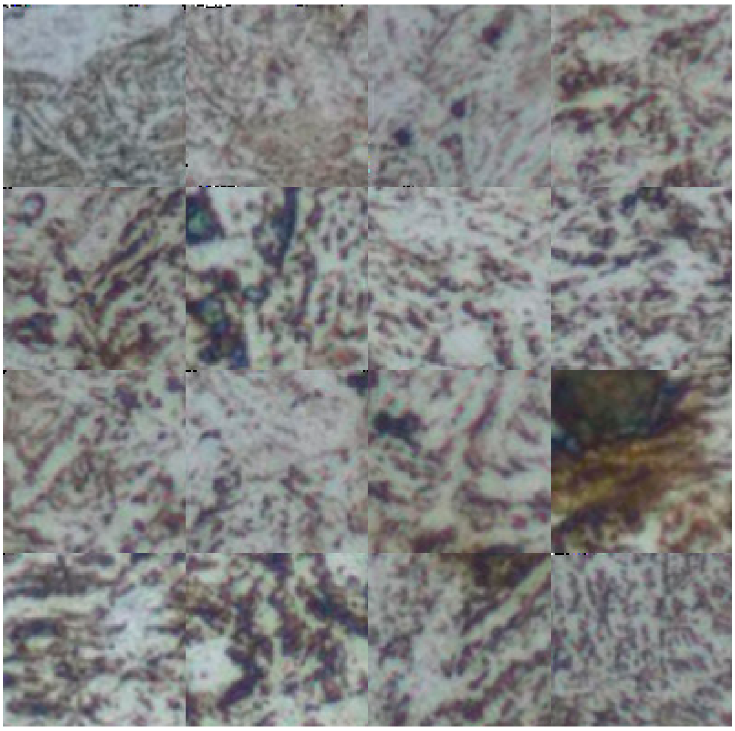


Figure 5. An example of the training image after merging with 16 images. The HV of this synthesized image is 359.4.

2.3. Convolutional Neural Network Backbones Used in This Work

In the field of deep learning, CNN is one of the best-known and most frequently used neural network architectures in various fields [22–24], including image recognition, computer vision, and so on. CNNs usually consist of a convolutional layer, a pooling layer, and a fully connected layer. A series of convolutional and pooling layers are the main building blocks used to construct the main structure of CNNs. The convolutional layer is the most important component of a CNN, mainly responsible for extracting regional features. The pooling layer, after the convolutional layer, decreases the spatial size of feature maps, which also enhances the performance by identifying the most important features. The filter steps through all the data at a constant rate and performs the weighted sum of the input data and the corresponding elements of the filter at the same time. The output feature maps will be imported into the following hidden layers, which consist of similar building blocks to extract higher-level features based on the feature maps from the previous layers. Finally, the feature maps are flattened and classified or regressed by fully connected layers (FCLs) to generate predictions.

The aforementioned CNNs represent the backbone of the proposed network, which was used for extracting features. In the following subsections, we briefly introduced the selected CNN backbones, which were not selected randomly but according to their popularity and performance.

2.3.1. AlexNet

Krizhevsky et al. [25] proposed AlexNet, which was trained by using graphics processing units (GPUs) and won the ImageNet challenge [26] in 2012. AlexNet employed the rectified linear units (ReLu) activation function instead of the Sigmoid to improve the training efficiency. Because of the success of AlexNet, since 2012, many advanced deep neural networks, such as VGG [27], ResNet [28], GoogLeNet [29], and SqueezeNet [30], have been subsequently proposed.

2.3.2. VGGNet

In 2014, VGG [27], a name derived from the Visual Geometry Group from Oxford, was released. Compared to AlexNet, the VGG model uses smaller convolutional filters of sizes 2 × 2 and 3 × 3, allowing it to reduce the complexity of the computation but achieve similar performance as larger filters. The authors released a series of VGG models with different numbers of layers from 11 to 19. Since some previous studies [31] have reported higher performance achieved by VGG with 16 layers, this paper adopts the VGG-16 model.

2.3.3. GoogLeNet

The GoogLeNet architecture was introduced in 2015 by Google researchers Szegedy et al. [29], which achieved good classification and detection results in the ImageNet large-scale visual recognition challenge (ILSVRC) 2014. Compared with AlexNet, the error rate is significantly reduced by introduction of the inception module, which simultaneously applies a variety of filter sizes 1 × 1, 3 × 3, and 5 × 5, and max-poling.

2.3.4. ResNet

ResNet [28] won 1st place in the ILSVRC 2015 classification task. ResNet solved the issue of the vanishing gradient problems in training of very deep neural networks by skipping connections or residuals, which takes the input of one layer and adds it to the output of the next linear function. Such structures are called residual blocks. By stacking them, ResNet is constructed.

2.3.5. SqueezeNet

In terms of model complexity, SqueezeNet [30] is the smallest CNN model employed in this study. It was developed based on AlexNet with 50 fewer parameters by reducing the size of the filter and still maintaining high accuracy.

2.4. Baseline Network Architecture—Single Task Learning

The architecture of the model is an obvious factor to improve the performance of downstream applications. In this paper, we focused on the above CNN backbones in our proposed network architecture to extract features for HV prediction. As mentioned before, the conventional CNN is not fully convolutional because it often contains FCLs as the last layers, leading to the requirement of a fixed-sized input. In order to enable the model to adapt to various sizes of images, a baseline network architecture based on fully CNN (FCN) [23] is exhibited in Figure 6, which replaces FCLs with convolutions and subsampling operations. We dropped the FCL of the backbone and the extracted feature map is adaptively convolved into a 1 × 1 vector and finally generates the HV as its output. The advantage of FCN is that it is not limited by the input size of the image, so we can accept the arbitrary size of the input image after developing the model.

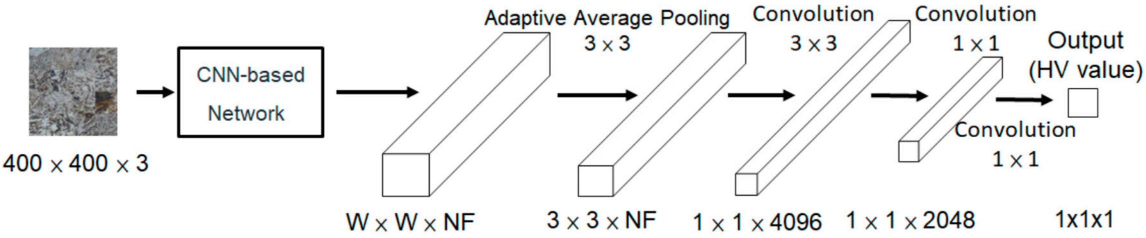


Figure 6. A fully convolutional neural network model architecture for HV. Here, NF refers to the number of filters of the underline CNN backbone.

The loss function is used to identify whether the neural regression model has learned important features to predict the HV. Therefore, the lower the loss function value, the better the model is. For training of the baseline model, we used the absolute error loss function, L1 loss, to estimate the HV errors predicted by the developed models. The principle is to take the average of the absolute error of the predicted HV $f(x_i)$ and the real HV (y_i), as in Equation (2).

$$L1Loss(\mathbf{x},\mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - f(x_i)| \tag{2}$$

2.5. Network Architecture with Multi-Task Learning

Multi-task learning is a training paradigm focused on performing multiple tasks simultaneously through shared representation to learn the common ideas between a collection of related tasks. Figure 7 illustrates the proposed multi-task learning network architecture in which the backbone with FCN is shared between two tasks: (1) the original regression task of predicting HV; similar to the baseline model, the L1 loss described previously was used to estimate the loss; (2) the binning task, which predicts the value of the image to the bin it falls into. Herein, we divided the range of the HVs into intervals to create bins, so the objective of the model is to predict the target bin based on the given image. The loss function used for the second task is negative log likelihood (NLL) loss, which is commonly used for multi-classification tasks. The NLL loss formula is given in Equation (3). Both the L1 loss and NLL loss were added to the overall training objective when training the multi-task model depicted in Figure 7.

$$\text{NLL Loss} = -\frac{1}{N}\sum_{k=1}^{N} log(y_k) \tag{3}$$

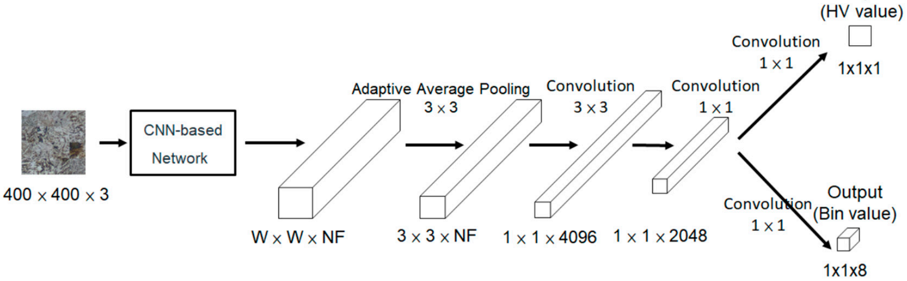


Figure 7. The proposed multitask model architecture for prediction of HVs. Here, NF refers to the number of filters of the CNN backbone.

2.6. Experiment Design and Configuration

2.6.1. Experimental Dataset

As described in the previous subsection, we received the data in two stages. Because only 46 images were collected in the initial stage, we randomly split the 46 images into training and test sets with a ratio of 8:1 and generated three synthesized datasets by using the image mixing data augmentation method described in Section 2.2. In the experiments of this study, the training, validation, and test sets consist of 5000, 500, and 100 mixed images, respectively. Note that, during generation of the training and validation sets, the six images reserved for the test were untouched. However, for the test set, in order to generate a wide range of HVs, the synthetization process will randomly select one to ten images from the 40 training samples combined with the samples from the test set to generate a test image with a size of 400 × 400. The distribution of HVs for these datasets is shown in Figure 8. As can be seen from these distributions, the distribution of the HVs can be divided into 8 intervals from 350 to 430. Therefore, the number of bins was set to 8 in the proposed model based on multi-task learning. The 59 new images received later were considered as an additional unseen test set to evaluate the generalization of the developed model. In our evaluation, we did not combine these images with any other image collected in the first stage.

2.6.2. Experimental Configuration

Based on the generated datasets, we conducted experiments to compare the performance of the baseline and multi-task models along with AlexNet, VGG-16, ResNet, SqueezeNet, and GoogLeNet as the backbones. In addition to training the models directly from the training set, we developed the models by using the pre-trained parameters available for each backbone to initialize the weights. The models pre-trained on the ImageNet dataset were employed in this study.

On the validation set, we observed that the relatively shallow models (e.g., AlexNet, VGG16, and SqueezeNet) with higher learning rates suffered from the overfitting problem, while the others, including ResNet18 and GoogLeNet, with smaller learning rates result in the underfitting problem. Therefore, we applied the learning rate range test method [32] to find the best initial learning rate. Table 1 summarizes the hyperparameters of the developed models with different backbones applied to both the baseline and multitask learning models. All the developed models were implemented by using PyTorch, and the experiments were conducted on the Windows 10 environment with a Nvidia GPU.

Table 1. Hyperparameters of the developed models.

Backbone	Learning Rate	Batch Size
AlexNet	10^{-5}	32
VGG-16	10^{-5}	32
ResNet-18	10^{-3}	32
SqueezeNet	10^{-5}	32
GoogLeNet	10^{-3}	32

2.6.3. Evaluation Metrics

The mean absolute error (MAE) defined in Equation (4) was used in our experiment to compare the model performance.

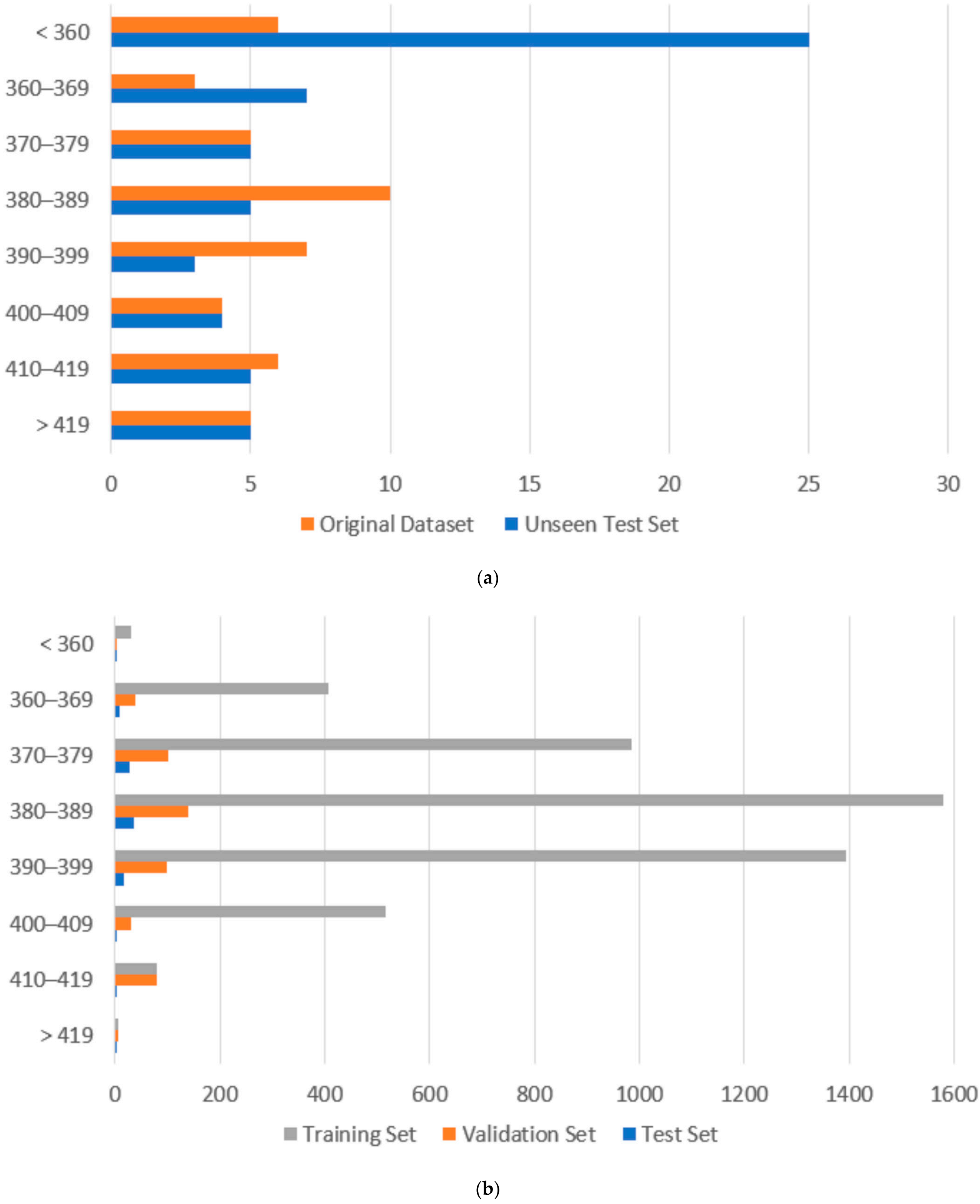


Figure 8. (**a**) Distribution of the original and unseen dataset; (**b**) distribution of the mixed datasets for training and testing the developed models.

$$\mathrm{MAE}(\mathbf{y},\ \mathbf{x}) = \sum_{i=1}^{n} |y_i - \mathrm{f}(x_i)| \tag{4}$$

where n is the number of test samples, y_i is the expected value for the ith sample, x_i is the i^{th} sample image in the given list of images ($\mathbf{x}$), and $f(x_i)$ is the output of a model for x_i. In addition to reporting the overall MAE estimated based on all test samples, we report the mean absolute error over bin (MAEB), which is defined in Equation (5), in order to examine the model's predictive power across different HV ranges.

$$\text{MAEB}(\mathbf{y},\ \mathbf{x}) = \sum_{i=1}^{b} MAE(\mathbf{y}_b, \mathbf{x}_b) \tag{5}$$

In Equation (5), $\mathbf{y}_b$ refers to the expected HVs for all the samples that fall into bin b and $\mathbf{x}_b$ refers to the corresponding list of images. Because the distribution of HVs in our dataset is a normal distribution, MAEB should be a better indicator of the model's predictive power than MAE.

3. Results

This section is divided into two subsections: in the first subsection, we compare the performance of the baseline models with the proposed multi-task models with and without using pre-trained backbones on the synthesized test set. In the second subsection, we evaluate the performance of the developed models on 59 unseen images to examine their generalization.

3.1. Performance Comparison of the Baseline and Multitask Learning Models with Different Backbones

As shown in Table 2, regardless of inclusion of the pre-trained models, the MAEs of most backbones were improved after introducing multi-task learning, except AlexNet. The model with VGG-16 as the backbone achieved the greatest improvement in MAE after exploiting multi-task learning and outperformed the others without loading pre-trained parameters. Furthermore, we observed that, although the domain of the ImageNet dataset used for training the pre-trained models is quite different from the task considered in this work, the results shown in Table 2 exhibited that initializing the backbone with the pre-trained parameters led to better results in terms of both metrics. The best MAE and MAEB evaluated on the test set were 7.6 and 5.92, respectively, which were achieved by the proposed multi-task learning model with the pre-trained VGG-16 as the backbone.

Table 2. Performance comparison between the baseline (single-task, abbreviated as S) and the proposed multitask learning models (abbreviated as M) on the synthesized test set. The column "I" indicates the improvement achieved by applying multi-task learning.

Backbone	**W/O Pre-Trained**						**W/Pre-Trained**					
	MAE			**MAEB**			**MAE**			**MAEB**		
	S	**M**	**I**	**S**	**M**	**I**	**S**	**M**	**I**	**S**	**M**	**I**
AlexNet	15.2	20.4	−5.2	13.3	13.4	−0.02	10.2	15.1	−4.9	11.3	11.66	−0.34
VGG-16	16.4	**7.8**	+**8.6**	13.2	11.2	+2	9.3	**7.6**	+**1.7**	**6.54**	**5.92**	+0.62
ResNet-18	**10.2**	9.3	+0.9	**11.6**	11.6	+0.06	**8.1**	7.8	+0.3	11.6	11.64	0
SqueezeNet	11.8	10.7	+1.1	13.6	10.6	+**3.06**	8.5	8.1	+0.4	9.94	7.26	+2.68
GoogLeNet	9.5	9.3	+0.2	11.7	**10.5**	+1.22	**8.1**	7.9	+0.2	12.4	9.32	+**3.08**

The bold highlight is the best-performed configuration. The best MAE and MAEB are underlined.

3.2. Performance Comparison of the Unseen Test Set

In order to test the model's ability to adapt properly to new or images of arbitrary size to support the requirements of the domain experts and real applications, 59 original images provided by the experts were used in this experiment. More specifically, two experiments were conducted on the additional unseen test set: (1) in the first experiment, the original image with a size of about 100 × 100 was directly used as the input for the developed models; (2) in the second experiment, before the test was conducted, we followed a similar mixing procedure to generate an image as the input for the model that is the same size as the training data (400 × 400) but comes from the same source image.

The results for the first and second experiments are shown in Tables 3 and 4, respectively. Indeed, the best MAE and MAEB were achieved by the model initialized with the pre-trained parameters. Among them, the model with GoogLeNet backbone achieved the best MAE if the original image was used, while the model with VGG as a backbone generally outperformed the others in terms of both MAE and MAEB.

Table 3. Performance comparison on the original unseen test set. In the table, the baseline model is denoted as S and the proposed multitask learning model is denoted as M. The column "I" indicates the improvement achieved by applying multi-task learning.

Backbone	W/O Pre-Trained						W/Pre-Trained					
	MAE			MAEB			MAE			MAEB		
	S	M	I	S	M	I	S	M	I	S	M	I
AlexNet	38.9	38.4	+0.5	40.6	39.8	+0.8	38.3	34.4	+3.9	36.3	40.3	−4
VGG-16	**28.6**	27.1	+1.5	32.4	31.6	+0.8	**27.4**	26.8	+0.6	**22.6**	**20.8**	+1.8
ResNet-18	33.6	33.3	+0.3	**30.4**	30.5	−0.1	33.3	33.2	+0.1	26.8	25.3	+1.5
SqueezeNet	45.2	39.0	+6.2	35.5	**23.9**	**+11.6**	40.3	38.5	+1.8	49.1	28.2	**+20.9**
GoogLeNet	31.9	**25.5**	+6.4	31.2	24.5	+6.7	**27.4**	**25.3**	+2.1	30.3	25.8	+4.5

The bold highlight is the best-performed configuration. The best MAE and MAEB are underlined.

Table 4. Performance comparison on the mixed unseen test set. In the table, the baseline model is denoted as S and the proposed multitask learning model is denoted as M. The column "I" indicates the improvement achieved by applying multi-task learning.

Backbone	W/O Pre-Trained						W/Pre-Trained					
	MAE			MAEB			MAE			MAEB		
	S	M	I	S	M	I	S	M	I	S	M	I
AlexNet	36.9	36.8	+0.1	34.7	34.0	+0.7	30.1	29.9	+0.2	28.0	28.8	−0.8
VGG-16	32.0	30.7	+1.5	28.6	28.3	+0.3	**20.1**	**19.7**	+0.4	**20.1**	**20.0**	+0.1
ResNet-18	30.1	30.0	+0.1	27.6	27.6	0	30.2	30.1	+0.1	26.3	26.3	0
SqueezeNet	37.1	27.0	**+10.1**	34.4	23.4	**+11.0**	22.3	21.2	+1.1	22.1	21.1	+1.0
GoogLeNet	**27.6**	**25.6**	+2.0	**22.3**	**25.1**	+2.8	32.0	25.7	**+6.3**	28.5	24.1	**+4.4**

The bold highlight is the best-performed configuration. The best MAE and MAEB are underlined.

Furthermore, considering both the results shown in this and previous subsections, we can observe that the proposed model with multi-task learning indeed reduces regression errors, regardless of whether the test is conducted by using mixed images or original images. However, comparing the experimental results shown in Table 4 with Table 3 illustrates that the regression errors for the original images are higher than those of the mixed images. We believe that is for the following reasons: the first is that, during our training process, all the developed models were trained by using images with a fixed size of 400 × 400, so the models should perform better on the input images with similar dimensions and aspect ratios. The second reason is that, as shown in Figure 8a, the HV distribution of the unseen test set is different from the training set. The lowest/highest HVs of the unseen test set are 300.9/570.0 but 323.4/494.0 in the original dataset. Therefore, the MAE and MAEB are large in these ranges.

Even though the regression errors seem to be large on the unseen test set, we believe that the proposed method still provides a feasible solution to estimate the HV; one piece of evidence to support that statement is that the acceptable HV error tolerance value is considered to be below 30 HV after consulting with the domain experts. Therefore, the results shown in Tables 3 and 4 are still confirmed to be within the acceptable range.

4. Discussion

The empirical results reported herein should be considered in light of some limitations. In our study, the presented results were estimated on the samples of the SCM 440 steel only. The presented performance of the proposed network architectures may be biased, and the effectiveness of the best backbone (VGG-16) and the transferred pre-trained weights may not be generalized to other materials. Furthermore, we only demonstrate the strength of the proposed multi-task learning architecture on one material; while it has potential beneficial effects on other materials, we did not have the opportunity to collect data and conduct experiments to highlight the potential benefits.

The mechanical properties in steels are strongly governed by the presented microstructures. In our study, several factors—the mechanical properties of the particular material, mixture microstructure, and grain size—are required for further discussion. For microhardness in different studied microstructures, steels with pearlite, bainite, and martensite microstructures were cold-rolled to maximum equivalent strains of 3.0, 2.5, and 2.0, respectively [33]. The hardness evolution as a function of the equivalent von Mises strain for the above microstructures is shown to be 300–500 HV for pearlite, 450–700 HV for banite, and 700–1000 HV for martensite [33]. Mixed microstructures combined particular properties. For example, the soft ferrite phase provides this steel grade its high ductility. The martensite phase—which has extremely high strength, such as regarding the quenched grades that are used for springs and cutting tools—is responsible for the high tensile strength. Combining the advantages of the two phases, called dual phase steels [34], the soft ferrite phase surrounds the hard islands of the martensite phase, which creates the very high initial work hardening rate exhibited by these steels. Banitic steels have a more complex microstructure, which contains upper and lower bainite in combination with martensite and carbides. The lower the bainitic transformation temperature, the finer the bainitic structure is, and it results in higher hardness, toughness, and wear resistance [35,36]. Furthermore, minimizing the martensite and cementite fractions in the microstructure contributes to improved damage resistance [35]. Grain size [37] and precipitates strength [38] effect should be considered. Liu et al. [37] reported the mechanical properties and grain size in steels after warm-rolling and annealing process. The original grain size of ferrite was ~0.53 µm, the yield strength and tensile strength were 951 MPa and 968 MPa, respectively, and the total elongation rate was 11.5% after warm-rolling at 600 °C. Additionally, after the next 4 h of annealing, the grain size of ferrite and particle size of cementite increased to ~1.35 µm and ~360 nm, and the yield strength and tensile strength decreased to 600 MPa and 645 MPa, respectively, with a total elongation increase of 20.9%. This implied a finer grain with better mechanical properties. We investigated the microstructure analysis using an optical microscope at magnification up to 500×, and it is difficult to analyze the above effect at this magnification.

5. Conclusions

In recent years, due to the rapid development of neural networks, they have been successfully applied in many fields. In this paper, we proposed a unique and novel CNN-based approach for Vickers hardness test. First of all, we proposed to apply an image data augmentation method to address the issue of insufficient training examples. Then, we developed a neural network architecture with state-of-the-art CNN architecture as the backbone to address the HV regression problem. The experimental results clarified that the network trained with two objectives—one is to predict the HV of an image, and the other is to predict the possible HV range for an image—outperforms the one that learns to predict HV only.

We also studied the effectiveness of adopting pre-training parameters trained on a totally different domain for the HV regression task. We observed that, although the pre-trained backbone was trained by using images of, for example, cats, dogs, etc., which are very different from the image domain used in this study, the network initialized with the pre-trained parameters still performed better than that trained from scratch. The best MAE of 6.54 and MAEB of 5.92 were achieved by the multi-task learning network with pre-trained VGG-16 as the backbone. The same model can achieve an MAE of 19.7 and an MAEB of 20.0 on an unseen dataset whose HV distribution is quite different from the original training set. The results were reliable after confirming with domain experts; thus, the usability of this proposed method is verified. In this paper, we have demonstrated the feasibility of developing a model that can predict HVs by using CNN and the data augmentation method. We plan to apply the multi-scale training method and other advanced image augmentation methods as future work to enhance the proposed model's capability in dealing with arbitrary input images. In addition, we will continue to cooperate with domain experts to create more training examples and apply self-supervised learning methods to generate in-domain pre-trained backbones to improve the performance of the developed models.

Author Contributions: Conceptualization, H.-J.D. and M.-H.T.; methodology, H.-J.D. and G.-Y.C.; software, G.-Y.C.; validation, G.-Y.C. and W.-S.C.; formal analysis, G.-Y.C. and X.-Y.S.; investigation, G.-Y.C. and W.-S.C.; resources, H.-J.D. and M.-H.T.; data curation, X.-Y.S.; writing—original draft preparation, G.-Y.C., W.-S.C., H.-J.D., M.E. and M.-H.T.; writing—review and editing, W.-S.C., H.-J.D., M.E. and M.-H.T.; visualization, G.-Y.C., H.-J.D. and M.-H.T.; supervision, H.-J.D. and M.-H.T.; project administration, H.-J.D.; funding acquisition, H.-J.D. and M.-H.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministry of Science and Technology, grant number 108-2218-E-992-320-MY2 and 109-2221-E-992-074-MY3.

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

References

1. Abdollah-Zadeh, A.; Salemi, A.; Assadi, H. Mechanical behavior of CrMo steel with tempered martensite and ferrite–bainite–martensite microstructure. *Mater. Sci. Eng. A* **2008**, *483–484*, 325–328. [CrossRef]
2. Itoh, H.; Ochi, M.; Fujiwara, I.; Momoo, T. SCC Life Estimation Based on Cracks Initiated from the Corrosion Pits of Bolting Material SCM435 Used in Steam Turbine. *JSME Int. J. Ser. B Fluids Therm. Eng.* **2004**, *47*, 316–322. [CrossRef]
3. Saroja, S.; Parameswaran, P.; Vijayalakshmi, M.; Raghunathan, V.S. Prediction of microstructural states in Cr-Mo steels using phase evolution diagrams. *Acta Metall. Mater.* **1995**, *43*, 2985–3000. [CrossRef]
4. Chen, C.-Y.; Hung, F.Y.; Lui, T.S.; Chen, L.H. Microstructures and Mechanical Properties of Austempering Cr–Mo (SCM 435) Alloy Steel. *Mater. Trans.* **2013**, *54*, 56–60. [CrossRef]
5. Zheng, Y.; Wang, F.; Li, C.; Lin, Y.; Cao, R. Effect of Martensite Structure and Carbide Precipitates on Mechanical Properties of Cr-Mo Alloy Steel with Different Cooling Rate. *High Temp. Mater. Process.* **2019**, *38*, 113–124. [CrossRef]
6. Thakare, A.S.; Butee, S.P.; Dhanorkar, R.; Kambale, K.R. Phase transformations and mechanical properties of thermomechanically processed 34CrMo4 steel. *Heliyon* **2019**, *5*, e01610. [CrossRef]
7. Pugh, S.F.; Little, E.A. Ferritic steels for fast reactor steam generators. *Nucl. Energy* **1978**, *17*, 179–183.
8. Hardness, A.B. *Standard Test Method for Microindentation Hardness of Materials*; ASTM Committee: West Conshohocken, PA, USA, 1999; Volume 384, p. 399.
9. Swetlana, S.; Khatavkar, N.; Singh, A.K. Development of Vickers hardness prediction models via microstructural analysis and machine learning. *J. Mater. Sci.* **2020**, *55*, 15845–15856. [CrossRef]
10. Nam, K.-S.; Hyun, Y.K.; Jo, C.Y.; Cho, Y.J. Mass Effect on the Heat Treated Mechanical Properties of SCM440(H) and SNCM439 Steel. *J. Korean Soc. Heat Treat.* **2011**, *24*, 10–15.
11. Sugimoto, T.; Kawaguchi, T. Development of an automatic Vickers hardness testing system using image processing technology. *IEEE Trans. Ind. Electron.* **1997**, *44*, 696–702. [CrossRef]
12. Araki, I.; Suzuki, K. Automatic measurement of Vickers hardness by microcomputer (I). *Bull. Fac. Educ.* **1982**, *1*, 77–83.
13. Serra, J. *Image Processing and Mathematical Morphology*; Academic Press: Now York, NY, USA, 1982.
14. Maier, A.; Niederbrucker, G.; Stenger, S.; Uhl, A. Efficient focus assessment for a computer vision-based Vickers hardness measurement system. *J. Electron. Imaging* **2012**, *21*, 021114. [CrossRef]
15. Hu, X.; Li, J.; Wang, Z.; Wang, J. A microstructure-informatic strategy for Vickers hardness forecast of austenitic steels from experimental data. *Mater. Des.* **2021**, *201*, 109497. [CrossRef]
16. Jalilian, E.; Uhl, A. Deep Learning Based Automated Vickers Hardness Measurement. In *International Conference on Computer Analysis of Images and Patterns*; Springer: Cham, Germany, 2021.
17. Tanaka, Y.; Seino, Y.; Hattori, K. Automated Vickers hardness measurement using convolutional neural networks. *Int. J. Adv. Manuf. Technol.* **2020**, *109*, 1345–1355. [CrossRef]
18. Tanaka, Y.; Seino, Y.; Hattori, K. Measuring Brinell hardness indentation by using a convolutional neural network. *Meas. Sci. Technol.* **2019**, *30*, 065012. [CrossRef]
19. Chen, Y.; Fang, Q.; Tian, H.; Li, S.; Song, Z.; Li, J. Automatic Measurement Algorithm for Brinell Indentations Based on Convolutional Neural Network. *Sens. Mater.* **2022**, *34*, 1043–1056. [CrossRef]
20. Li, Z.; Yin, F. Automated measurement of Vickers hardness using image segmentation with neural networks. *Measurement* **2021**, *186*, 110200. [CrossRef]

21. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
22. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [CrossRef]
23. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
24. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.
25. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
26. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpanthy, A.; Khosla, A.; Berstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
27. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
28. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.
29. 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.
30. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. *arXiv* **2016**, arXiv:1602.07360.
31. Lathuilière, S.; Mesejo, P.; Alameda-Pineda, X.; Horaud, R. A comprehensive analysis of deep regression. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 2065–2081. [CrossRef] [PubMed]
32. Smith, L.N. Cyclical learning rates for training neural networks. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 24–31 March 2017; IEEE: Manhattan, NY, USA, 2017.
33. Wang, X.; Zurob, H.S.; Xu, G.; Ye, Q.; Bouaziz, O.; Embury, D. Influence of Microstructural Length Scale on the Strength and Annealing Behavior of Pearlite, Bainite, and Martensite. *Metall. Mater. Trans. A* **2012**, *44*, 1454–1461. [CrossRef]
34. Azuma, M.; Goutianos, S.; Hansen, N.; Winther, G.; Huang, X. Effect of hardness of martensite and ferrite on void formation in dual phase steel. *Mater. Sci. Technol.* **2012**, *28*, 1092–1100. [CrossRef]
35. Shipway, P.H.; Wood, S.J.; Dent, A.H. The hardness and sliding wear behaviour of a bainitic steel. *Wear* **1997**, *203–204*, 196–205. [CrossRef]
36. Hajizad, O.; Kumar, A.; Li, Z.; Petrov, R.H.; Sietsma, J.; Dollevoet, R. Influence of Microstructure on Mechanical Properties of Bainitic Steels in Railway Applications. *Metals* **2019**, *9*, 778. [CrossRef]
37. Liu, G.; Xia, C. Microstructure Evolution and Mechanical Properties of Medium Carbon Martensitic Steel during Warm Rolling and Annealing Process. *Materials* **2021**, *14*, 6900. [CrossRef]
38. Hättestrand, M.; Nilsson, J.O.; Stiller, K.; Liu, P.; Andersson, M. Precipitation hardening in a 12%Cr–9%Ni–4%Mo–2%Cu stainless steel. *Acta Mater.* **2004**, *52*, 1023–1037. [CrossRef]