P1-1
17th International Conference on Machine Vision Applications (MVA)
Fully Online, July 25–27, 2021.

Machine-learning-based Quality-level-estimation System for Inspecting Steel Microstructures

Hiromi Nishiura*, Atsushi Miyamoto*, Akira Ito*
Shogo Suzuki†, Kouhei Fujii†, Hiroshi Morifuji†, and Hiroyuki Takatsuka†

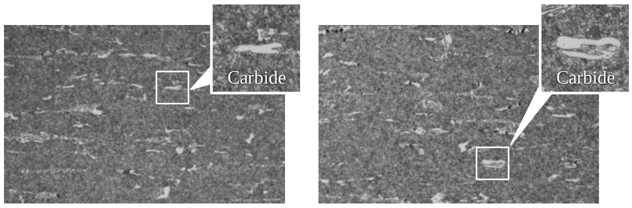
*Hitachi, Ltd., Yokohama, Kanagawa, Japan
†Hitachi Metals, Ltd., Yasugi, Shimane, Japan

Abstract

For quality control of special steels, the microstructure of the steel is visually inspected on the basis of microscopic images. In this study, aiming to eliminate the effect of personal differences between inspectors and reduce inspection costs, a system for automatically estimating quality level (hereafter, "automatic-quality-level-estimation system") based on machine learning is proposed and evaluated. Collecting the images is a manual task performed by the inspector, and it is difficult to prepare multiple training samples in advance. As for the proposed method, overfitting, which is a problem in training with few samples, is suppressed by data expansion based on variation distribution of correct-answer values. The correct-answer rate for judging quality level by an inspector was about 90%, while the proposed method achieved a rate of 90%, which is sufficient to render the method practically applicable.

1. Introduction

Special steel has excellent heat resistance and corrosion resistance, and it is used in a wide range of fields such as tool steel, electronic materials, industrial-equipment materials, and aircraft-related materials. Special steel is manufactured through the processes of casting, rolling, forging, and heat treatment, and inspecting the metallographic structure of the manufactured steel is essential to guarantee high quality. One of the inspection methods used is a "microstructure test"—in which the inspector visually judges the quality of the metal structure by examining inspection images taken with a microscope [1]. Examples of such images are shown in Figure 1. The white bands and grains in the images are carbides, and it is known that the shape of these carbides is highly correlated with the physical characteristics of special steel. The inspector compares a sample photograph (prepared in advance) for each quality level with the inspection image, and judges the quality level from the similarity. The quality level of sample 1 is higher than that of sample 2 in Figure 1. The quality level is



(a) Sample 1 (b) Sample 2

Fig. 1: Sample of microstructure image

described in detail in Section 2. In the meantime, different pattern variations in the shape of the microstructure occur, and several quality levels may be mixed up in the image. Accordingly, judging the quality level is a sensory evaluation that depends on the skill of the inspector, and the judgment results even between skilled inspectors may vary.

In this study, a system for automatically estimating the quality level (hereafter, "automatic quality-level estimation system") by machine learning—which aims to eliminate the influence of personal differences between inspectors on microstructure tests and reduce inspection costs—is proposed and evaluated. In recent years, a proposal called "a deep network model," represented by a convolutional neural network (CNN), has considerably improved the performance of machine learning [2][3]. On the contrary, as for machine learning, so-called "overfitting," namely, generalization performance deteriorates owing to over-optimization in regard to specific training data, is problematic. As mentioned above, while many variations exist in an inspection image, it is difficult for the inspector to (i) manually collect images and assign a quality level for each image and (ii) prepare a lot of training data. Moreover, the correct-answer value for the quality level taught includes fluctuations in judgment due to individual differences between inspectors and incorrect training. Under that condition, overfitting is considered to be particularly likely to occur. Data augmentation is known as common method of suppressing overfitting [4]. As for the proposed method as well, the aim is to achieve the same or better estimation performance as that achieved by an inspector by incorporating algorithms that apply data augmentation.

2. Quality-level-estimation system

The proposed automatic quality-level estimation system, which consists of a learning phase and an estimation phase, is shown schematically in Figure 2. In the learning phase, multiple inspection images for learning $\{f_i\}$ $(i = 1, \ldots, N_f)$ are acquired by microscope. For each inspection image, the inspector assigns the value of the correct answer for the quality level (g_i) in nine steps from 1.0 to 5.0 in increments of 0.5, and those levels are taken as training data $\{(f_i, g_i)\}$, the smaller the value of which, the higher the quality. With the inspection images as the input and the quality level as the output, the parameters of the estimator are optimized by using the training data. In the estimation phase, the trained estimator is used to estimate the quality levels of actual inspection images and create a "mill sheet" (i.e., an inspection certificate for steel materials).

As for the proposed method, each training sample (f_i, g_i) is subjected to the following data augmentation for suppressing overfitting. The training samples are "extended." Under the assumption that the value of the correct answer for the quality level (g_i) given by the inspector includes variation, multiple "extended correct-answer values" are generated and learned. This process prevents the estimator from being over-optimized for correct-answer values that may contain errors. Details of data expansion are described in Sections 3.1. The network model used for the estimator is described in detail in Section 3.2.

3. Algorithms

3.1. Data expansion based on variation distribution of correct-answer values

The correct-answer values for the quality level given by the inspectors vary. This variation includes fluctuations in judgments in the sensory evaluation and erroneous training. As a practical matter, improving the quality of training data is difficult through the efforts of the inspectors alone. Accordingly, a mechanism for maintaining judgment performance, even if the quality of the training data deteriorates, is important.

In consideration of the above-described issues, with the proposed method, the training data is subjected to "data expansion" on the basis of the distribution of the statistical variation of the correct-answer values. That is, as shown in Figure 3, multiple extended correct-answer values $\{g^*_{ij}\}$ $(j = 1, \ldots, N_{Si})$ are generated according to variation distribution (d) from the correct-answer value g_i for the quality level given by the inspector. These values are learned as an "extended training-sample group," namely, $\{(f_i, g^*_{ij})\}$. Varying the correct-answer values makes it possible to suppress excessive optimization for each training sample,

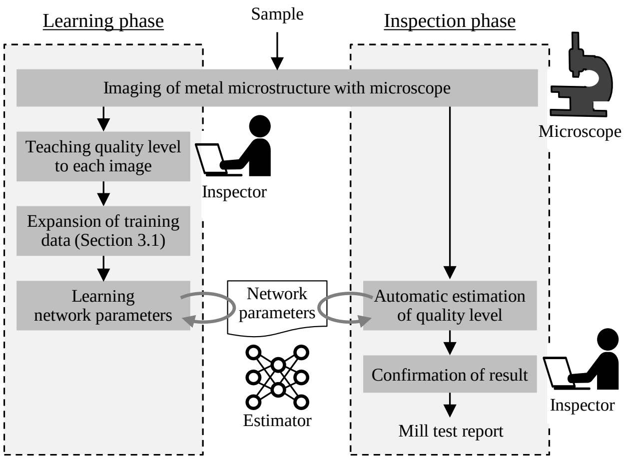


Fig. 2: Quality-level-estimation system

even if the trained correct-answer values are inaccurate to some extent, and generalization performance can be improved.

The key question is how to properly present the variability distribution. As for the proposed method, it was decided to switch variation distribution d according to the value of the correct answer, g_i. Specifically, as shown in Figure 4, the variation distribution for each correct-answer value is given as a probability distribution $d(g^*_i; g_i)$, were the extended correct-answer value g^*_i is given as a variable, and the correct-answer value g_i is given as a parameter. The variation distribution is due to mistakes that are made by the inspector according to their on-site experience, and it can be said to be a kind of "domain knowledge." For example, a discontinuous change in appearance occurs between quality levels 1.5 [Fig. 4(b)] and 2.0 [Fig. 4(c)], and judgment error spanning this change tends not to occur.

If the training data is expanded on the basis of an erroneous variation distribution of correct-answer values, a large amount of false training data will be generated, and it may conversely reduce judgment performance. Appropriate data expansion is possible by explicitly reflecting domain knowledge in the variation distribution.

3.2. Network model

The network model used for the discriminator is shown in Figure 5. It is based on a network called "VGG," which consists of a convolutional layer and a pooling layer [3]. In the first stage of the network, an inspection image is input, and features are extracted by the convolutional layer and the pooling layer. After the convolutional layer, the extracted features are subjected to normalization. Normalization prevents large fluctuations in the distribution of the input and suppresses overfitting. It has been reported that the commonly used batch normalization becomes unstable when batch size is small; accordingly, the

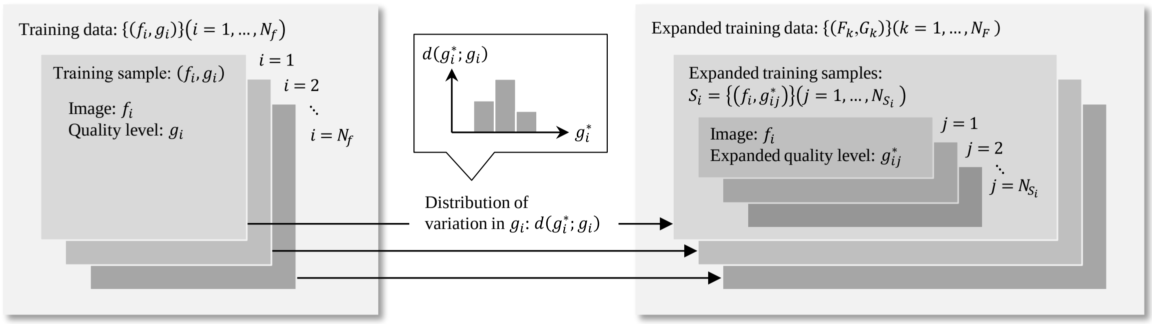


Fig. 3: Extension of training data

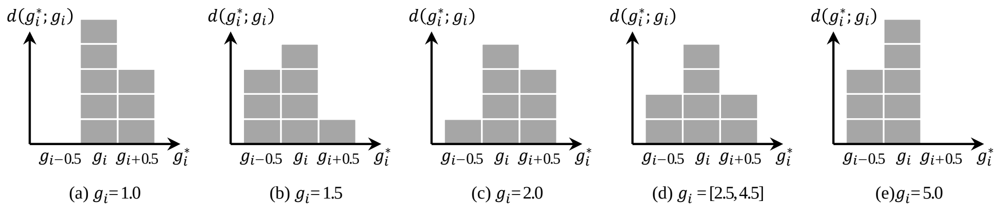


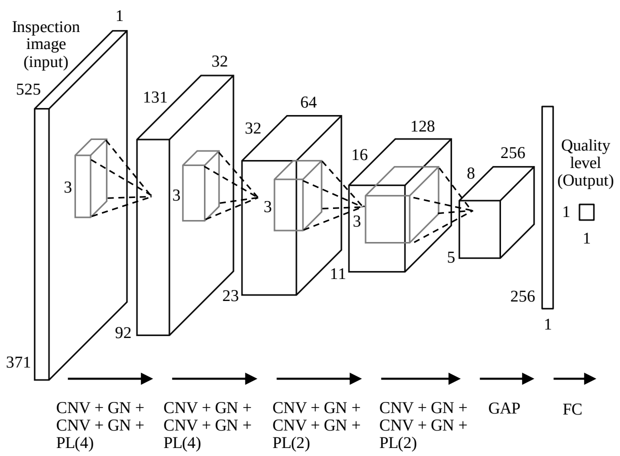
Fig. 4: Distribution of variation of correct-answer values for quality level

proposed method adopts a derivative of it, namely, group normalization [5] [6].

In the latter stages of the network, the extracted features are fully combined, and the quality level is output. To estimate the quality level as a regression problem, a ReLU function is used as the activation function of the output layer. The ε-insensitive loss method was used to learn the network parameters [7]. This method is said to have suppress overfitting by preventing excessive minimization of the loss function. A loss function in the k-th mini-batch learning step is given as follows:

$$\text{Loss}_k = \sum_{i \in \Phi_k} \sum_{j \in \Psi_{ik}} \max\left(\left|\hat{g}_{ij} - g_{ij}^*\right| - \varepsilon, 0\right), \quad (1)$$

where $\hat{g}_{ij}$ is estimated quality level of extended training-sample (f_i, g^*_{ij}); ε is the insensitive error; Φk and Ψik are groups of ID numbers of training samples and extended



CNV: Convolutional layer, GN: Group normalization layer,
PL(4): Pooling layer (pooling size=4), PL(2): Pooling layer (pooling size=2),
GAP: Global average pooling layer, FC: Fully connected layer.

Fig. 5: Network model of proposed estimator

training-samples included in the k-th mini-batch. Every extended training-sample is trained in one epoch.

4. Experimental results

4.1. Experimental conditions

The effectiveness of the proposed method was experimentally verified as follows. As the input data used in the experiment, 362 inspection images of alloy tool steel, SKD11 (JIS standard G4404 [8]), were used. This image set included various quality levels for performance evaluation of the proposed method. The images were evaluated by four-fold cross-validation taking the ratio of training, validation, and testing samples as 6.0:1.5:2.5.

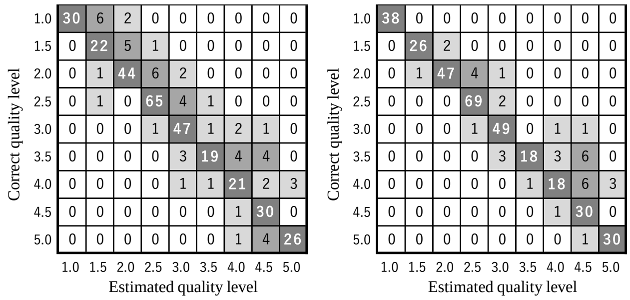
To independently verify the effects of data expansion described in Sections 3.1, a method that is similar to the proposed method but excludes data expansion (hereinafter, referred to as "comparative method") was also evaluated.

4.2. Estimation accuracy and learning curves

The estimated quality levels for the testing data are shown as confusion matrices in Figure 6. The horizontal axis of each matrix is the value of quality level estimated by the judgment device, and the vertical axis is the taught correct-answer value. It can be concluded from comparing matrix (a) with matrix (b) that the more the samples are lined up diagonally, the better the estimation result gets. As for the comparative method, shown in Fig. 6(a), the sample distribution tends to spread around the diagonal center line, and estimation accuracy is 84%. On the contrary, as for the proposed method, shown in Fig. 6(b), the spread of the sample distribution is small, and estimation correct-answer rate is 90%. The correct-answer rate of the proposed method is significantly improved compared to that of the comparative method. It can thus be said that data expansion effectively improved performance. It is noteworthy that the correct-answer rate is about 90% even in the case of visual judgment of the quality level by the inspector, so a correct-answer rate of 90% was taken as a guideline for the target performance required for practical use. The proposed method achieved a correct-answer rate equal to that of the inspector.

The learning curve is shown in Figure 7. The horizontal axis shows "epoch" (i.e., number of learning times), and the vertical axis shows "loss-function value." Since learning solves a minimization problem concerning this loss value, the loss value tends to decrease as the epoch progresses. As for the training data, it can be seen that the loss value in the case of the conventional method decreases sharply compared to that in the case of the proposed method. On the contrary, as for the validation data, the loss value in the case of the proposed method is smaller than that in the case of the conventional method. It is clear from these results that in the case of the proposed method, overfitting of the training data observed in the case of the conventional method was suppressed, and high generalization performance in regard to the validation data was obtained.

Even though it is not in an overfitting state, in the case of general learning, the loss value of the training data used for optimization tends to be smaller than the loss value of the validation data. However, it is interesting that in the case of the proposed method, the loss value of the validation data is noticeably smaller. This is because the correct-answer values are scattered by the data expansion; however, since the magnitude of the loss value is slightly reversed, it is conceivable that the variation distribution of correct-answer values in the data expansion differs slightly from the actual distribution. Of course, it is difficult to give an accurate variation distribution; even so, to further improve the judgment accuracy, it may be worth considering approaches such as changing the variation distribution of correct-answer values for each training sample according to the loss value.



(a) Comparison method (b) Proposed method

Fig. 6: Confusion matrix

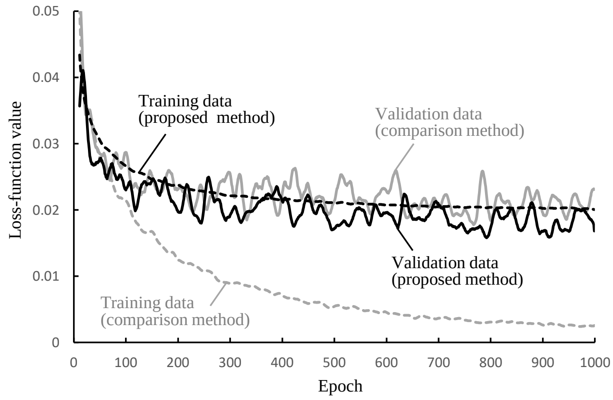


Fig. 7: Learning curves

5. Concluding remarks

In this study, an "automatic-quality-level-estimation system" based on machine learning was proposed and evaluated experimentally. By data expansion, overfitting was suppressed, and judgment accuracy of 90% was achieved. This accuracy is considered to be equivalent to the performance of visual judgment by an inspector.

As for future work, this system will be implemented at production sites with the aim of verifying its applicability. Implementing the system will eliminate the influence of personal differences between inspectors and reduce inspection costs. Moreover, applying the system to inspection of metals other than the alloy tool steel verified in the present study will be investigated, and the scope of automation based on this system will be expanded.

References

[1] D. S. Bulgarevich, S. Tsukamoto, T. Kasuya, M. Demura, and M. Watanabe: "Pattern Recognition with Machine Learning on Optical Microscopy Images of Typical Metallurgical Microstructures," Scientific Reports 8, pp. 2078-1-8, 2018.

[2] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio: "Object Recognition with Gradient-Based Learning," Shape, Contour and Grouping in Computer Vision, pp. 319-345, 1999.

[3] K. Simonyan and A. Zisserman: "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv: 1409.1556v6, pp. 1-14, 2015.

[4] B. Gao, C. Xing, C. Xie, J. Wu, and X. Geng: "Deep Label Distribution Learning With Label Ambiguity," IEEE Transactions on Image Processing, vol.26, no.6, pp. 2825-2838, 2017.

[5] S. Ioffe and C. Szegedy: "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," International Conference on Machine Learning, PMLR 37, pp. 448-456, 2015.

[6] Y. Wu and K. He: "Group Normalization," European Conference on Computer Vision, pp. 3-19, 2018.

[7] V. N. Vapnik: "The Nature of Statistical Learning Theory," Springer, 1995.

[8] "JIS G4404:1956. Alloy tool steels," 1956.