A CNN-Based Transfer Learning Method for Defect Classification in Semiconductor Manufacturing

Kazunori Imoto, Tomohiro Nakai, Tsukasa Ike, Kosuke Haruki, and Yoshiyuki Sato

***Abstract*—In this paper, we focus on a defect analysis task that requires engineers to identify the causes of yield reduction from defect classification results. We organize the analysis work into three phases: defect classification, defect trend monitoring and detailed classification. To support the first and third engineer's analytical work, we use a convolutional neural network based on the transfer learning method for automatic defect classification. We evaluated our proposed methods on real semiconductor fabrication data sets by performing a defect classification task using a scanning electron microscope image and thoroughly examining its performance. We concluded that the proposed method can classify defect images with high accuracy while lowering labor costs equivalent to one-third the labor required for manual inspection work.**

***Index Terms*—Machine learning, deep learning, transfer learning, defect classification, semiconductor manufacturing.**

I. Introduction

DEFECT inspection and detect trend monitoring, which provide useful information for engineers endeavoring to identify root causes of process failures, are crucially important for yield quality control. Inline inspection systems, usually comprising optical wafer inspection tools and scanning electron microscope (SEM)-based review tools, are deployed at semiconductor wafer production sites for process monitoring [1], [2]. However, as shown in Figure 1, defects in semiconductor device fabrication have a wide range of shapes and textures due to the sophistication of manufacturing process and as a consequence, the accuracy of manual defect classification depends greatly on the expertise of inspectors. Automatic defect classification (ADC) is a function that automatically classifies defect images into pre-determined defect classes based on their appearance [3]. Several methods have been proposed for ADC systems: rule-based classifiers [4], learning-based classifiers [5], [6], and hybrid-type classifiers [7]. However, poor data, and a deceptive environment in the manufacturing process where the classification problem itself varies over time, renders the ADC task difficult to solve.

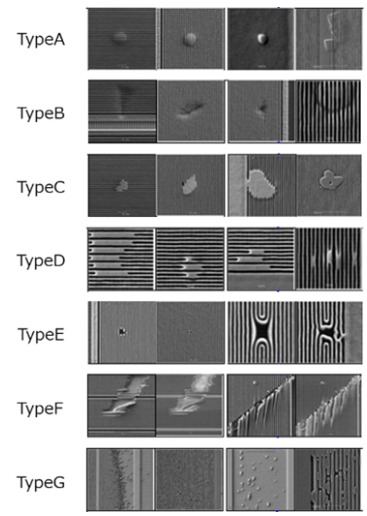

Fig. 1. Example of Defects.

Recent advances in deep learning technology have achieved human-level classification performance [8], and provided advanced analytical tools for analyzing big data from manufacturing [9]. Deep learning-based techniques typically require ground-truth labels for a large training data set. For many tasks, however, the data-labeling process is expensive, making it difficult to obtain strong supervision information [10]. Additionally, the given labels are not always ground-truth due to the sophistication of the process. To overcome inconsistent manual classification and other costly problems, we present a convolutional neural network(CNN)-based transfer learning method of automatic defect classification [11]. We evaluated our proposed methods on real semiconductor fabrication data sets using an SEM-image classification task.

The remainder of this paper is organized as follows. In Section II, we examine the defect analysis task. In Section III, we introduce works related to our method. In Section IV, a CNN method is adopted for automatic defect classification. In Section V, we introduce a transfer learning approach to reduce labeled data for training. In Section VI, we discuss about the acceleration of model training. Finally, Section VII presents the conclusion of this paper.

Manuscript received May 30, 2019; revised July 27, 2019 and August 31, 2019; accepted September 5, 2019. Date of publication September 16, 2019; date of current version October 29, 2019. *(Corresponding author: Yoshiyuki Sato.)*

K. Imoto, T. Nakai, T. Ike, and K. Haruki are with the Analytics AI Laboratory, Corporate Research and Development Center, Toshiba Corporation, Kawasaki 212-8582, Japan (e-mail: kazunori.imoto@toshiba.co.jp; tomohiro.nakai@toshiba.co.jp; tsukasa.ike@toshiba.co.jp; kosuke.haruki@toshiba.co.jp).

Y. Sato is with Toshiba Memory Corporation, Minato, Japan (e-mail: yoshiyuki.sato@toshiba.co.jp).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSM.2019.2941752

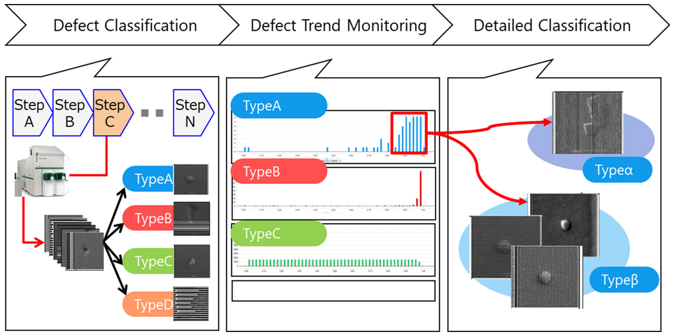

Fig. 2. Overview of Defect Quality Control.

TABLE I
CNN CONFIGURATION

CNN Parameters	Value
Number of convolution layers	33 layers
Module	Inception v2
Activation function	ReLU
Regularization method	Dropout
Optimizer	SGD
Weight decay	0.005
Momentum	0.9
Learning rate	0.001
Batch Size	64
Mini batch iteration (pre training)	50,000
Mini batch iteration (fine tuning)	20,000

II. DEFECT ANALYSIS TASK

In this paper, we focus on a defect analysis task that requires engineers to identify the causes of yield reduction from defect classification results. During inspection processes of semiconductor manufacturing, defect images are classified according to types of defects, in order to find process malfunctions and suppress yield reduction. As shown in Figure 2, the analysis falls into three phases: defect classification, defect trend monitoring and detailed classification. First, defect images captured by an inspection system using a scanning electron microscope (SEM) are classified into several dozens of defect types. Using the classification results, the frequency of each defect type occurring is monitored. If an increase in the frequency of a certain defect type is detected, images classified as the detected type will be further classified into more specific sub-categories in order to identify the root causes of the process failure(s). While such trend monitoring can be automated based on classification results, the first and third phase of the analysis require costly manual inspection or reconfirmation. We therefore use deep learning technologies in the first and third phases of the analysis to assist the engineers' work.

III. RELATED WORK

Recent advances in deep learning technology (e.g., CNN) have achieved human-level classification performance [8], and provided advanced analytical tools for analyzing big data from manufacturing [9]. Deep learning technologies, such as surface defect classification of steel sheets [12] and fabric defect classification [13], have been introduced in the manufacturing sector and automatic inspection techniques have been widely applied in manufacturing processes to ensure the high quality and performance of products [14]. The semiconductor industry has also shown interest in deep learning applications: Nakazawa and Kulkarni applied a CNN for wafer-map classification [15] and Nakata *et al.* applied a CNN for failure recurrence monitoring by classifying wafer-map patterns [16]. However, CNN models for wafer-surface SEM defect classification have not been addressed. Kim *et al.* developed a CNN-based defect image classification model for through-silicon via processes [17]. Cheon *et al.* proposed a single CNN model that can extract effective features for defect classification [18]. Yang *et al.* proposed a transfer learning based online Mura defect classification method [19]. A drawback of these methods is that they require more than several thousand training data points with accurate ground-truth labels.

One approach to lower the cost of data-labeling is the use of weakly supervised learning. Zhou divide week supervision into three types incomplete, inexact, and inaccurate they describe as follow [10]. In incomplete supervision, only a (small) subset of training data is labeled while the rest of the data remains unlabeled. In inexact supervision, only coarse-grained labels are used. In inaccurate supervision, the labels given are not always ground-truth, due to worker fatigue or the difficulty of categorizing certain images.

We adopted inaccurate supervision because we already have a large set of data that was manually and inconsistently labeled. We also used the transfer learning method to reduce the required amount of training data with ground-truth labels. Our inaccurate supervision approach and transfer learning method is explained in Sections IV and V, respectively.

IV. DEFECT CLASSIFICATION BY DEEP LEARNING

A. Network Structure

Table I shows our CNN configuration. The input SEM image size was resized to 128 x 128. We adopted the Inception model developed by Szegedy *et al.* [20]. Each module is composed of 3 different-sized of filters (1×1, 3×3, 5×5) and the max pooling and concatenated outputs are sent to the subsequent inception module. To lower cost, the number of input channels were limited by adding an extra 1x1 convolution before the 3x3 and 5x5 convolutions. This method, called "convolution factorization," decreases the number of parameters in each inception module in order to reduce the computational cost. We adopted 10 inception modules comprising 33 convolutional layers. Rectified linear activation was used for each convolutional layer. The fully connected (FC) layer with a size of 256 was added after the convolutional layers with sigmoid activation. After dropout [21], another FC layer with the size of the defect class was added. The final layer is a softmax layer for outputting class probability calculation.

B. Learning Strategy

Our proposed method comprises two stages, pre-training and fine-tuning, as shown in Fig. 3. In the first training

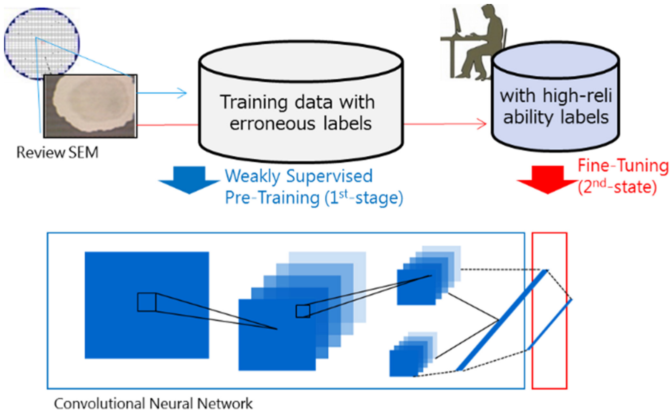

Fig. 3. Two-Stage Transfer Training Strategy.

TABLE II
EVALUATION DATASET

	Set A	Set B	Set C	Set D
Pre-training data (by non-expert)	36,034	32,339	16,289	13,218
Fine-tuning data (by expert)	2,400	2,400	5,386	1,800
Evaluation dataset (by expert)	2,400	2,400	5,388	1,800
Defect type	21	21	12	14

stage, parameters of all layers in the CNN model are trained from tens of thousands stacked image data points that contain numerous incorrect labels attributable to weakly supervised training. The ratio of overlap in defect classification labels between operators is < 0.9. Inconsistent labels can worsen the performance of the classification model. In the second training stage, the output layer of the pre-trained CNN model is extended with randomly initialized weights and a small learning rate is used to tune all parameters from their original values to minimize loss on the target task with the few data points that contain highly reliable labels. The mini-batch stochastic gradient descent method, which was the most used optimization algorithms for deep learning, was used for CNN parameter learning. Other training parameters are shown in Table I.

C. Evaluation Setup

Wafer-surface-defect SEM images were sampled from an actual manufacturing facility to evaluate the performance of the automatic defect classification (ADC) method and the proposed method. For the experiment, all defect images were normalized to a uniform size of 128×128. We prepared four defect image data sets. The number of image data points and defect types are summarized in Table II. Each data set has noisy data labeled by non-experts and pure data labeled by experts. There is no overlap between noisy data and pure data. Pure data is used for fine-tuning and evaluation. Each dataset contains several thousand images composed with several dozen defect classes. Sample images of these defects are shown in Fig. 1.

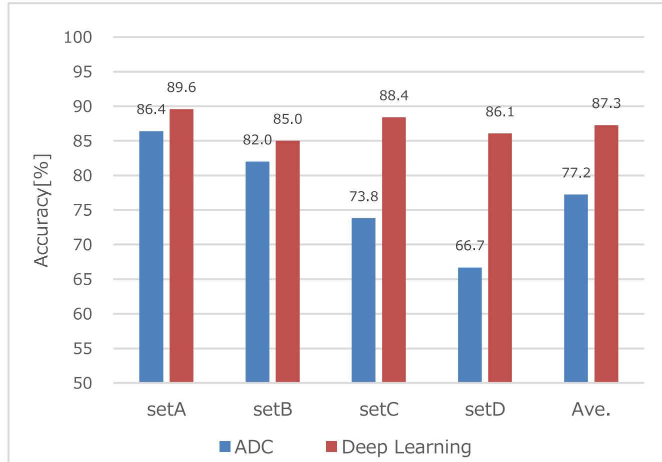

Fig. 4. Comparison of ADC and the Proposed Method.

TABLE III
DETAILED RESULT OF SET C

Defect	No of samples	Precision(%)	Recall(%)
A	7	0.500	0.286
B	49	0.773	0.694
C	7	1.000	0.857
D	684	0.877	0.810
E	895	0.817	0.765
F	20	0.833	0.250
G	226	0.771	0.566
H	7	0.750	0.429
I	60	0.872	0.683
J	95	0.596	0.589
K	3311	0.917	0.982
L	27	0.000	0.000
Total	5388	0.884	0.884

D. Result

Figure 4 shows the classification performance of the commercially available traditional ADC system and our proposed method, the average accuracies of which were 77.23% and 87.26%, respectively. Our proposed method showed high performance on all evaluation data sets (from set A to set D). This result indicates that our proposed method outperformed the traditional ADC system. To further our examination of the proposed method, we analyzed classification results of set C for each defect type in greater detail. Table III shows the per-class classification precision (referred to as "purity"), recall (referred to as "accuracy") and the number of samples. For confidentiality, only the defect symbol name is shown Data set C contains 12 classes. Class K accounted for a very large portion of the data set (61.45%), whereas classes A and H accounted for only a small fraction (0.13%). Class performance generally degrades as the number of instances becomes small. Accordingly, classes D and K had very high performance due to their high frequency of appearance.

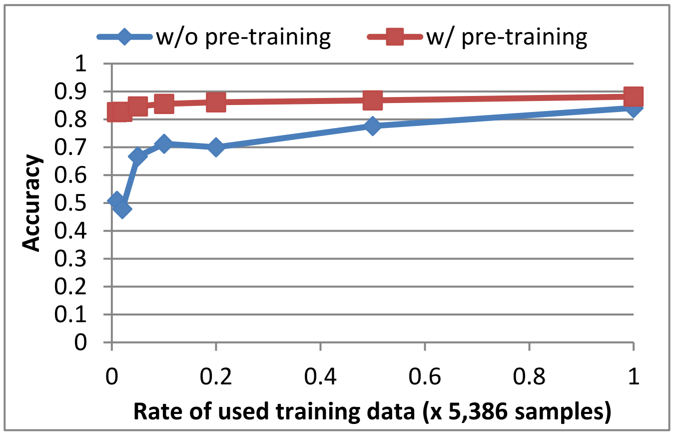

Fig. 5. Performance of the Proposed Method.

V. Transfer Learning for Fine-Grained Classification

A. Learning Strategy

In the third analysis, there were few data points with highly reliable labels available for training because fine-grained defects occurred less frequently than expected and manual classification was more difficult. Such problems can be addressed via transfer learning. Approaches to transfer learning can be divided into four types based on what to transfer: (1) instance-based, (2) feature-representation, (3) parameter and (4) the relational knowledge [22]. We adopted the feature-representation transfer approach in order to reuse the "good" feature representation of the source domain, because the domain used for fine-grained defect classification in the third phase is the same as domain one used for rough defect classification in first phase. In concrete terms, the output layer of the pre-trained CNN model for the rough defect classification in the first phase is extended with randomly initialized weights for the third phase, and a small learning rate is used to tune all parameters from their original values to minimize loss on the new task with small number of labeled data points.

B. Evaluation Setup

For the experiment of third phase of the analysis phase, data set C was prepared by relabeling 5386 fine-grained image-label sets for fine-tuning training and 5388 image sets for evaluation; there were 29 classes. To evaluate the relationship between the ratio of fine-grained image and the accuracy, six conditions were set from 0.01 to 1. (0.01 is equivalent to 54 images for fine-tuning)

C. Results

Figure 5 shows the classification accuracies of fine-grained defects for data set C, which were obtained by varying the number of training data points. The classification accuracy initially increased as the number of training data points increased and the proposed method with pre-training outperformed the method without pre-training under all conditions. Notably, the proposed method had high classification accuracy even when the number of training data points was small. The results indicate that the proposed method has higher robustness against a lack of labeled data.

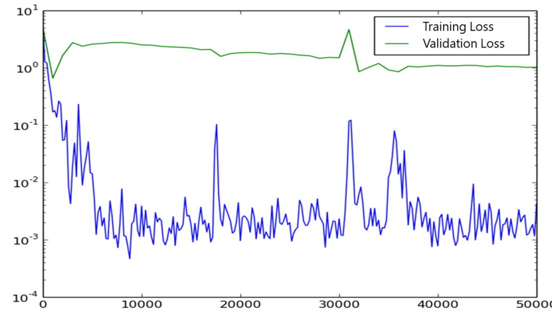

Fig. 6. Learning Curve of Mini-Batch Training.

VI. Acceleration of Model Training

For deep learning technology to be of practical use, it is necessary to lower computational cost [23]. Early stopping, in which training is stopped based on a validation loss is a simple but powerful solution to reduce training time. Figure 6 shows a learning curve derived from plotting model learning performance during mini-batch training. Monitoring the learning curves of models during training can be used to diagnose problems with learning, such as the model being underfit or overfit. As can be seen in Fig. 6, the model is sufficiently trained by the 5000the iteration. In this work, we reduced the runtime of training with hundreds of thousands of images from 40 h to 4 h, increasing the speed tenfold by combining use of early stopping and GPU parallel computing.

VII. Conclusion

In this paper, we focused on a defect analysis task that requires engineers to identify the causes of yield reduction from defect classification results. To overcome inconsistent manual classification and other costly problems, a CNN-based transfer learning method of automatic defect classification was presented. Since deep learning requires a large amount of labeled training data, classification performance sometimes deteriorates when sufficient reliable labeled data are not available. We introduced transfer learning which exploits unreliable labeled data or labeled data of different tasks. From experimental results using real semiconductor fabrication data, we have confirmed that the proposed method outperforms the conventional system and high classification accuracy is realized using a limited number of reliable labeled data. At the manufacturing site, defect types that exceed a precision standard of automation are excluded from manual inspection work. Because our proposed method can classify frequent defect types with high accuracy, the labor required for manual inspection work decreases nearly 2/3 compared to commercially available traditional ADC system.

References

[1] Y. Lee and J. Lee, "Accurate automatic defect detection method using quadtree decomposition on SEM images," *IEEE Trans. Semicond. Manuf.*, vol. 27, no. 2, pp. 223–231, May 2014.

[2] M. Harada, Y. Minekawa, and K. Nakamae, "Defect detection techniques robust to process variation in semiconductor inspection," *Meas. Sci. Technol.*, vol. 30, no. 3, 2019, Art. no. 035402.

[3] R. Nakagaki, T. Honda, and K. Nakamae, "Automatic recognition of defect areas on a semiconductor wafer using multiple scanning electron microscope images," *Meas. Sci. Technol.*, vol. 20, no. 7, 2009, Art. no. 075503.

[4] P. B. Chou, A. R. Rao, M. C. Sturzenbecker, F. Y. Wu, and V. H. Brecher, "Automatic defect classification for semiconductor manufacturing," *Mach. Vis. Appl.*, vol. 9, no. 4, pp. 201–214, Feb. 1997.

[5] K. Watanabe, Y. Takagi, K. Obara, H. Okuda, R. Nakagaki, and T. Kurosaki, "Efficient killer-defect control using reliable high-throughput SEM-ADC," in *Proc. ASMC Conf.*, 2000, pp. 219–222.

[6] S. Chen, T. Hu, G. Liu, Z. Pu, M. Li, and L. Du, "Defect classification algorithm for IC photomask based on PCA and SVM," in *Proc. Congr. Image Signal Process.*, vol. 1, 2008, pp. 491–496.

[7] J. Ritchison, A. Ben-Porath, and E. Molasay, "SEM based ADC evaluation and integration in an advanced process fab," in *Proc. SPIE*, vol. 3998, 2000, pp. 258–268.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[9] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: Methods and applications," *J. Manuf. Syst.*, vol. 48, pp. 144–156, Jul. 2018.

[10] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 44–53, Jan. 2018.

[11] K. Imoto, T. Nakai, T. Ike, K. Haruki, and Y. Sato, "A CNN-based transfer learning method for defect classification in semiconductor manufacturing," in *Proc. Int. Symp. Semicond. Manuf.*, Dec. 2018, pp. 1–3.

[12] S. Zhou, Y. Chen, D. Zhang, J. Xie, and Y. Zhou, "Classification of surface defects on steel sheet using convolutional neural networks," *Materiali Tehnologije*, vol. 51, no. 1, pp. 123–131, 2017.

[13] A. Şeker, K. A. Peker, A. G. Yüksek, and E. Delibaş, "Fabric defect detection using deep learning," in *Proc. SIU*, 2016, pp. 1437–1440. doi: 10.1109/SIU.2016.7496020.

[14] S.-H. Huang and Y.-C. Pan, "Automated visual inspection in the semiconductor industry: A survey," *Comput. Ind.*, vol. 66, pp. 1–10, Jan. 2014.

[15] T. Nakazawa and D. V. Kulkarni, "Wafer map defect pattern classification and image retrieval using convolutional neural network," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 2, pp. 309–314, May 2018.

[16] K. Nakata, R. Orihara, Y. Mizuoka, and K. Takagi, "A comprehensive big-data-based monitoring system for yield enhancement in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 4, pp. 339–344, Nov. 2017.

[17] J. Kim, S. Kim, N. Kwon, H. Kang, Y. Kim, and C. Lee, "Deep learning based automatic defect classification in through-silicon via process: FA: Factory automation," in *Proc. 29th Annu. SEMI Adv. Semicond. Manuf. Conf. (ASMC)*, Saratoga Springs, NY, USA, 2018, pp. 35–39.

[18] S. Cheon, H. Lee, C. O. Kim, and S. H. Lee, "Convolutional neural network for wafer surface defect classification and the detection of unknown defect class," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 2, pp. 163–170, May 2019.

[19] H. Yang, S. Mei, K. Song, B. Tao, and Z. Yin, "Transfer-learning-based online Mura defect classification," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 1, pp. 116–123, Feb. 2018.

[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2818–2826.

[21] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014.

[22] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[23] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, U.K.: MIT Press, 2016.

[24] R. Ren, T. Hung, and K. C. Chen, "A generic deep-learning-based approach for automated surface inspection," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 929–940, Mar. 2018.

[25] S. Mei, H. Yang, and Z. Yin, "Unsupervised-learning-based feature-level fusion method for Mura defect recognition," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 1, pp. 105–113, Feb. 2017.