Interactive Image Annotation and AI-assisted Segmentation of TEM Images for Automatic CD Measurement

Dongok Kim*†, Wonhee Lee†, Yeny Yim†, Byeongkyu Cha, Hansaem Park, Subong Shon, Myungjun Lee
Mechatronics Research, Samsung Electronics Co., Ltd., 1-1 Samsungjeonja-ro, Hwaseong-si, Gyeonggi-do, 18448, Republic of Korea

ABSTRACT

As the semiconductor structures become increasingly miniaturized and complex, the process of measuring and analyzing the structure using a microscope becomes crucial. High-resolution transmission electron microscope (TEM) images are widely used, but they are expensive to acquire and analyze. If the region and boundary of the material in the TEM image can be automatically segmented, the measurement cost will be reduced. We proposed a method to generate a segmentation label for a TEM image using a deep learning model that performs segmentation based on weak supervision and active learning. The proposed method achieved an accuracy of 98% in 10% of the time compared to the manual method. This approach will reduce the cost of high-resolution TEM image analysis and accelerate the semiconductor device development process.

Keywords: transmission electron microscope, automatic metrology, image segmentation, generation of ground truth label, image annotation, weak labeling, active learning

1. INTRODUCTION

As transistor dimensions continue to shrink and the aspect ratio of structures increases, the needs for sophisticated metrology and inspection (MI) technologies become increasing in the semiconductor manufacturing process. Destructive analysis becomes more important due to the difficulties in manufacturing process.

Two key factors in MI are throughput and accuracy. To accelerate MI and improve throughput, new technologies such as multi-beam inspection [1] and massive overlay measurement [2-3] have been developed. For accuracy enhancement, EUV imaging [4], multi-angle ellipsometry [5], fluorescence imaging [6], ptychography [7], and microsphere-assisted hyperspectral imaging [8] have been developed.

In addition to above mentioned in-line MI technologies, out-fab destructive analysis is important to have deeper understanding of the root causes and complex phenomena occurring within the semiconductor manufacturing process. Destructive analysis consists of sample preparation, data acquisition, and data analysis [9-10]. Transmission electron microscope (TEM) [11-12] is a vital tool for acquiring data from destructive analysis sample in the process. Angström-resolution TEM images are widely used to measure the critical dimensions (CDs) of the structure in the device.

However, it is difficult to analyze TEM images due to ambiguous boundaries, image quality variability, and equipment variability. Traditionally, skilled engineers (users) have manually measured the CDs of the structures in images, because distinguishing between material boundaries can be challenging as shown in Fig. 1 [9-10]. Manual measurement is extremely time-consuming and has clear limitations in terms of amount of analysis, statistical accuracy and data quality control. To improve metrology reliability during the device development and accelerate high volume manufacturing, automatic CD measurement in TEM images is essential.

*dongok10.kim@samsung.com, †These authors contributed equally.

Metrology, Inspection, and Process Control XXXVIII, edited by Matthew J. Sendelbach, Nivea G. Schuch, Proc. of SPIE Vol. 12955, 129550Y · © 2024 SPIE
0277-786X · doi: 10.1117/12.3009407

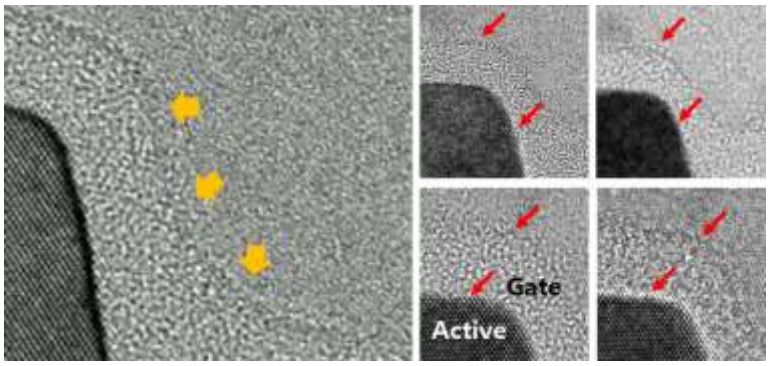

Figure 1. Ambiguous boundaries of materials in a TEM image (left) and image quality variability due to data acquisition condition (right). These characteristics of TEM images pose a challenge for engineers to accurately distinguish between material boundaries.

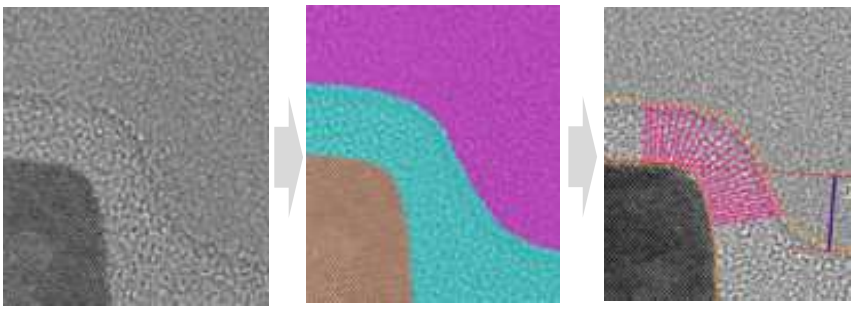

Figure 2. Automatic CD measurement process of the structures in the image. The image (left) can be segmented to distinguish the boundaries (center). CDs of the structure can be automatically measured using a segmentation label (right).

Recently, numerous semiconductor equipment companies such as FEI, Hitachi, and AMAT have been working for automatic measurement of the structures in devices [13-15]. These works include the metrology in electron microscope images. For automatic measurement of structures in electron microscope images, it is crucial to segment the structures separately according to their materials. This technique is well-known as image segmentation [16-17]. The automatic CD measurement process using image segmentation is described in Fig. 2 [9-10].

State-of-the-art researches for image segmentation of electron microscope images are based on a supervised learning [18-20]. A large number of images and their corresponding ground truth (GT) labels are required to learn the features of the images and segmented regions [21]. The GT label refers to the accurate annotation of regions or pixels in an image. To generate the GT label, the image must be fully annotated in pixel-level according to the materials. It requires a lot of manual works and elaborated labeling. Manual image annotation is tedious, time-consuming, and has variabilities among users. It takes about 30 minutes to process a certain TEM image from a DRAM device. An efficient annotation method is necessary to reduce these costs from manually generated labels.

We propose a method to generate the GT label quickly and accurately in order to segment a TEM image into structures and measure the CDs automatically. The proposed method is based on weak labeling [22-23] and active learning [24]. In aspects of weak labeling, the proposed method generates GT labels with a minimal amount of user interactions compared to full annotation. Though the initial label could be partial or noisy, the accuracy of the label can be improved by iterating labeling and training the model. It has an advantage that can reduce time and cost for labeling. With an active learning, proposed method can improve the accuracy of the model by recommending the users the regions with the highest uncertainty values in the image and updating the model with the weak labels added on the regions by users. The uncertainty here is a probabilistic estimate of the uncertainty associated with the predictions made by a deep learning model. It is generated by the model and provides a measure of the confidence that the model has in its predictions [25].

The model utilized for generating GT labels can also be trained on the TEM images and their corresponding GT labels through supervised learning. Even with a limited dataset, the model can be fine-tuned to accurately capture the materials and structures present in the TEM images. This domain-specific fine-tuning improves the performance of the GT label generation process, allowing for the generation of more GT labels within the same timeframe.

The proposed method has the following contributions. It can generate the GT label quickly within less than 10% of the time required for the manual approach, while achieving comparable segmentation accuracy with a minimal number of user interactions. We were able to improve the segmentation accuracy of the GT label and reduce uncertainties by adding weak labels to the most uncertain spots. Furthermore, the utilization of supervised learning for the model enabled faster and accurate segmentation of TEM images. This method can be extended to other images, such as scanning electron microscope (SEM) [26] or optical images obtained from most equipment in the field of metrology and inspection. This can accelerate the feedback of the semiconductor manufacturing process by automatic CD measurement, visualization, and storing the result to the database.

2. METHODS

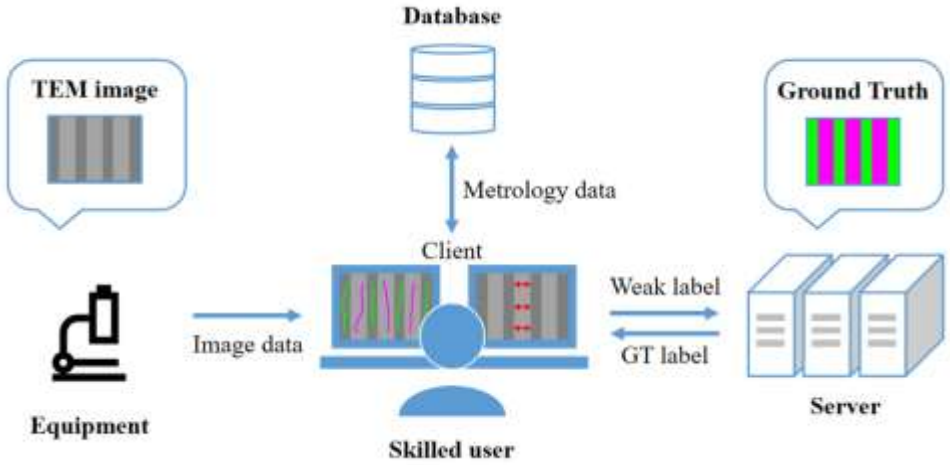

Figure 3. A schematic diagram of the proposed method. It consists of four components: TEM equipment, client, server, and database. The TEM equipment provides TEM image data to a skilled user at the client, who can add weak labels to the image. The server learns from weak labels received from the client and generates segmentation results. This feedback loop continues until the GT label is generated. Metrology data is automatically acquired from the segmentation result and stored in the database.

A schematic diagram of the proposed method is illustrated in Fig. 3. It consists of 4 components: TEM equipment, client, server and database. The TEM equipment digitizes a sample section as a two-dimensional image. At the client, a skilled user annotates weak labels on the TEM image to modify the image segmentation results. The weak annotations are sent to the server, where fully annotated images, also known as GT labels, are generated. The cycle of the user feedback at the client and GT label generation at the server is repeated. A collection of images and their paired GT labels is stored in the server, and a segmentation model is trained using these datasets. The trained model can segment a new image of the client. The CDs of the structures in the image can be measured automatically, and the measurement results are stored in the database.

In this paper, we propose a method for generating GT labels of TEM images. The method employs two deep learning models, which we refer to as 'guide model' and 'segmentation model'. A *guide model* assists users in generating GT labels, which are then used to train the *segmentation model*. Although the guide model and the segmentation model share the same architecture and functions of outputting a segmentation map from an image, they have different objectives. The purpose of the guide model is to generate a GT label for a given image. It updates itself by learning from the weak labels annotated by the user and is intentionally overfitted to the given image. Conversely, the segmentation model aims to generate a segmentation result for a new image. It updates itself by learning from the GT labels of the images produced by the guide model. It is important to enhance the generalizability of the model so that it can produce accurate segmentation result for any new images. Techniques such as data augmentation, cross-validation, and regularization can be employed to broaden the generalizability of the segmentation model [27].

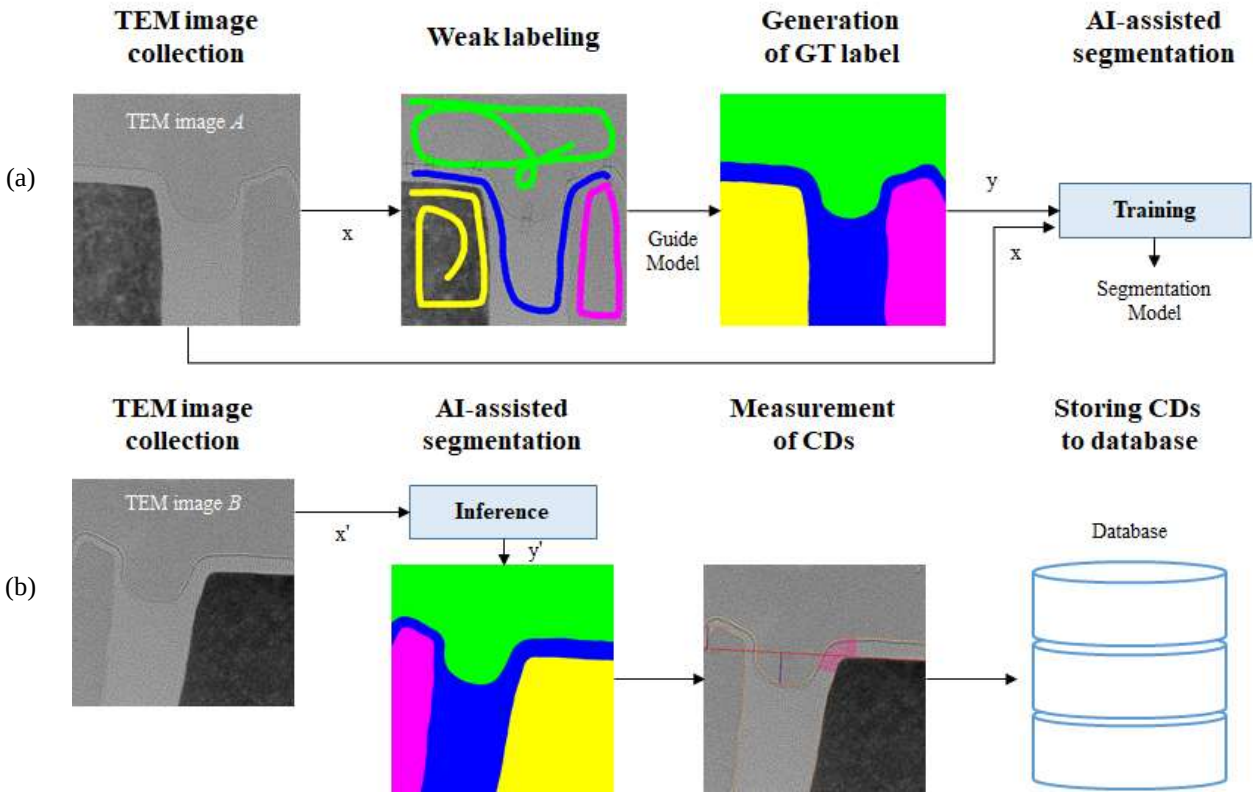

Figure 4. The overall process flow for segmentation and automatic CD measurement. To explain, two distinct TEM images (*A*, *B*) are introduced; (a) Training: the guide model generates a GT label of the TEM image *A* using weak labels annotated by a user. This labeled pair of the TEM image *A* and its corresponding GT label, (*x*, *y*), is utilized to train the segmentation model; (b) Inference: the trained segmentation model generates a segmentation for the image *B* that has not been annotated. Metrology data associated with the TEM image is automatically acquired from segmentation result and stored in the database.

The process flow of the proposed method is illustrated in Fig. 4. The flow includes two main stages; (a) training and (b) inference. In the training stage, a user and guide model collaboratively generate GT labels for a given set of TEM image collection. The user may annotate weak labels on the TEM image to train the guide model until the GT label is generated properly. Then a segmentation model is trained on the pairs of images and their corresponding GT labels. The second stage is inference. The trained model generates a segmentation result for the image which has not been annotated. Note that this stage does not contain any weak labeling nor training processes, instead the inference result is utilized to measure CDs in the image. The measured CD values are stored in the database.

Let there be two TEM images acquired at the same device and step. They may be acquired by distinct equipment or condition, but they need to have the same species of materials. Without any GT labels of TEM images, the detail process flow is as follow.

For a given image *A*, a user can annotate specific materials with weak labels of different colors. The guide model suggests an appropriate position to annotate weak labels based on the uncertainty value u_i at a pixel i,

$$u_i = -\sum_c p_i^c \log(p_i^c), \tag{1}$$

where p_i^c represents the predicted probability of a segmentation class (material) c at pixel i. The uncertainty here is a probabilistic estimate of the uncertainty associated with the predictions made by a deep learning model [25]. The weak label added by the user to that spot is used to update the guide model. This approach can be considered as active learning, which can reduce the annotation work required to achieve the desired accuracy. The guide model learns weak labels and propose a new segmentation map to the user. The accuracy of the map can be improved by repeating the process of weak labeling and updating the guide model. After several iterations of these steps, a GT label can be produced.

The total loss $\mathcal{L}$ is calculated by summing up the cross-entropy of each pixel with weights:

$$\mathcal{L} = \sum_i -w_i \log(p_i^c), \tag{2}$$

where w_i is the weight at pixel i. This formula was first introduced in [28].

We propose an overfitted and balanced active learning (OBAL) algorithm in order to generate overfitted GT label for each TEM image and assign balanced weight by considering features of the TEM image. The weight w_i is calculated using the following two pieces of information and six conditions:

Information

1. Pixels on the weak labels are considered as foreground and the other parts of the pixels are considered as background.

2. Inference results consists of the segmentation map, accuracy, and uncertainty map.

Conditions

1. Weight is inversely proportional to accuracy in the foreground.

2. Weight is inversely proportional to uncertainty in the background.

3. The range of uncertainty values in the background is clamped to limit too large or too small values.

4. To avoid bias towards one class, weight at each class is equally contributed to the total weight.

5. The ratio for the sum of weights in the background and foreground can be manually adjusted.

6. The weights always have positive values, and the sum of weights over the entire image is 1.

By repeatedly collecting pairs of images and their corresponding GT labels, the segmentation model can be iteratively trained and evaluated. After updating the segmentation model, it can be used as pre-trained weights for the guide model. This helps the guide model converge quickly and accelerate a generation of GT labels. The guide model provides GT labels to the segmentation model, and the segmentation model provides better initial weights to the guide model. They are learned in a complementary manner. Image and GT label pairs can be collected from various users of multiple clients and used to train a segmentation model. The trained segmentation model infers segmentation results from the new image *B* in Fig.4. From the segmentation result of the image *B*, the CDs can be measured automatically.

3. RESULT

The performance of the proposed method was evaluated using TEM images of DRAM devices in terms of accuracy and processing time. The images were manually segmented by skilled users, and their corresponding GT labels were used to assess the accuracy of the model. For quantitative evaluation, we evaluated the accuracy of the segmentation map generated by guide model at each iteration of training. Nine images acquired at DRAM gate module were used to generate the GT label. To assess the accuracy, the segmentation maps were compared with the GT label pixel-by-pixel and mean values of the all-pixel accuracies at each training iteration were evaluated. We also compared the map generation time for two different cases: segmentation model #1 and #2 which utilize pre-trained weight from natural and TEM images as an initial weight. For qualitative evaluation, we demonstrated the segmentation maps and uncertainty maps calculated at different iterations and the most uncertain spots at the maps.

Figure 5 shows the accuracy of two models with respect to the number of training iterations. The figure shows that all-pixel accuracy from the segmentation model #1 was 24.5%$\pm$2.7% on average at the initial inference result, which meets the expected random accuracy for 4 materials, 25%. A standard deviation is represented as an error bar. This deviation was originated from the different patterns of material boundaries in the image. As each guide model is trained from the weak labels, the average accuracy increases and the standard deviation decreases. The average accuracy became 97.4% at the 9th iteration. The average elapsed time until that moment was 182 seconds (about 3 minutes).

Segmentation model #2, pre-trained with TEM images of the same device and step, showed better performance than segmentation model #1, which was pre-trained with only natural images. The accuracy was more than 98% and the guide

models were trained within less than 4 times of weak label annotation. Their average accuracy and elapsed time at the last iteration were 98.2% and 20 seconds, respectively.

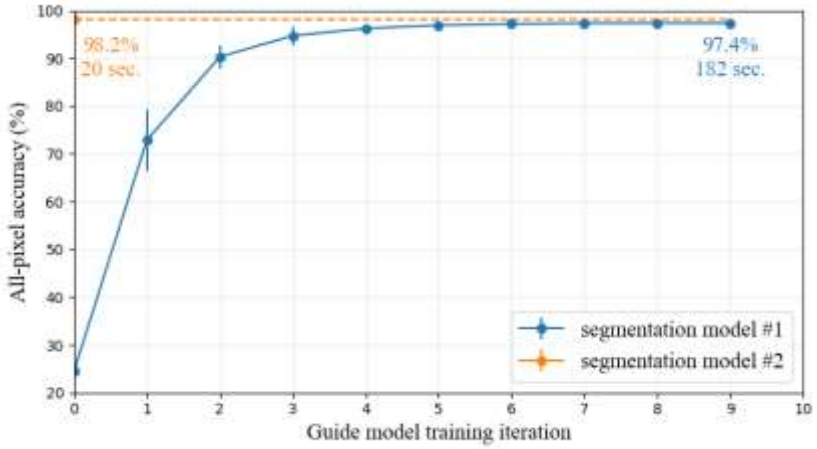

Figure 5. The average all-pixel accuracy of the segmentation maps with respect to the number of training iteration count for two segmentation models. A standard deviation is represented as an error bar. Guide models were trained by annotating weak labels at each iteration count. At the 9th iteration of a guide model from the segmentation model #1, the average elapsed time was about 3 minutes and the average all-pixel accuracy was measured as 97.4%. In the case of a guide model from segmentation model #2, these values were improved to 20 seconds and 98.2%. Since some of segmentation maps showed the accuracy of more than 98.0% at the first iteration count, guide models were trained at most 4 times and their last accuracy and elapsed time were averaged.

This result implies a positive feedback cycle between the guide model and segmentation model. The trained segmentation model provides better initial weights to the guide model, and the time taken to generate GT label is reduced. The guide model from a trained segmentation model can generate more GT labels at the same time.

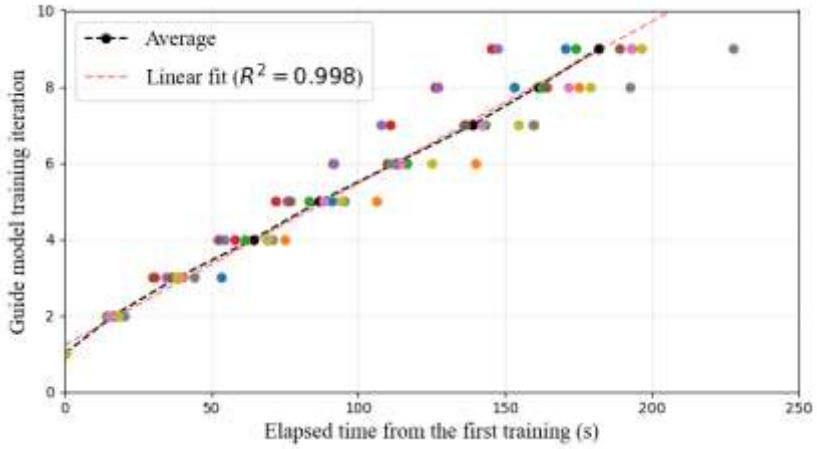

Figure 6. Elapsed time to generate 9 GT labels for each training iteration count. The black and red dashed lines represent the average elapsed time and its linear fitted line, respectively. Other colors indicate each TEM image case. Although the dataset was acquired from the same device and step, the difficulty of GT label generation can vary depending on the image quality, equipment condition, and engineer who annotates weak labels.

Figure 6 illustrates the elapsed time to generate GT labels for each training iteration with guide models from segmentation model #1. The black and red dashed line represents the average elapsed time for each training iteration and its linear fitted line, respectively. The other colors indicate the elapsed time for nine images. The GT label generation process depends on the image quality, equipment condition, and weak labels annotated by the engineer. Therefore, the time taken to annotate weak labels can also depend on the image, and the distributions of the elapsed time at each iteration are shown. However, as a segmentation map converges, so does the accuracy as shown in Fig. 5. The standard deviations after the 4th iteration

almost vanish. The average elapsed time shows a linear behavior ($R^2 = 0.998$), hence a TEM image variability can be averaged out during the GT label generation process.

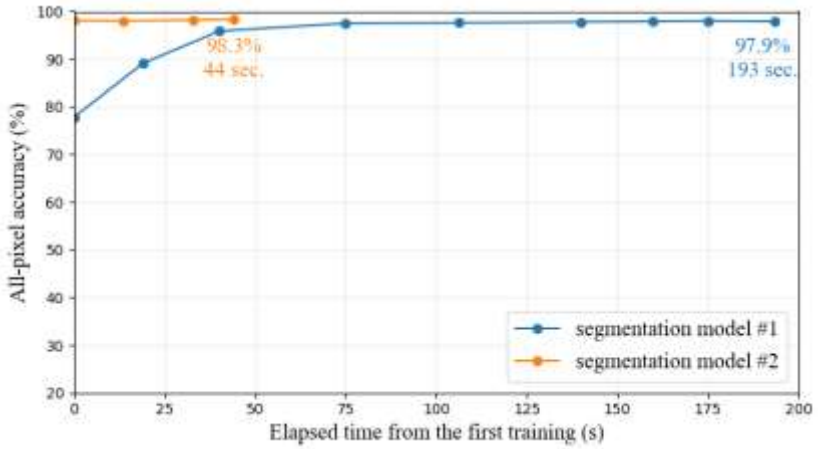

Figure 7. The all-pixel accuracy of the segmentation maps with respect to the elapsed time from the first training for two segmentation models. Each dot represents a training iteration, where the leftmost dot is the first training iteration. A guide model from segmentation model #1 showed an accuracy of 97.9% at the 9th training iteration and took 193 seconds, while that of segmentation model #2 had an accuracy of 98.3% within 1 minute at the 4th training iteration. Time intervals between training iterations were reduced by training the segmentation model.

The all-pixel accuracy of the segmentation maps during one of GT label generation processes is shown in Fig. 7. Each dot represents a training iteration for segmentation model #1 and #2, where the leftmost dot for each model is the first training iteration. The dataset was chosen due to the need for additional trainings to generate GT label with a guide model from segmentation model #2. Nevertheless, a segmentation model of the guide model at the 4th training iteration reached 98.3% accuracy in 44 seconds. The guide model from segmentation model #1 had 97.9% accuracy in 193 seconds. The mean time intervals between training iterations for segmentation model #1 and #2 are calculated as about 24 and 15 seconds, respectively, so the time taken to annotate weak labels is reduced by more than 30%.

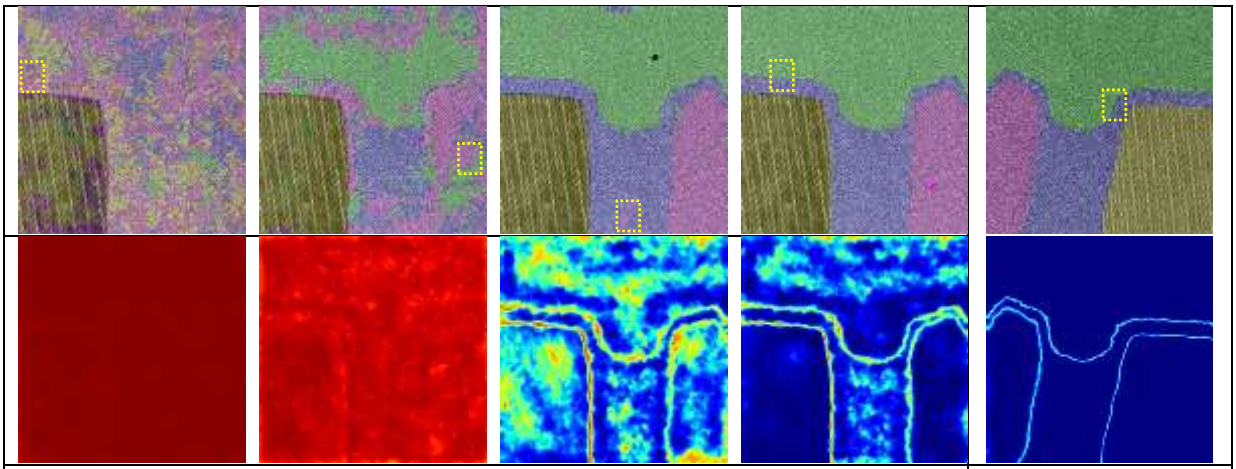

Figure 8. Segmentation maps and uncertainty maps of the guide model during the training of the first image of DRAM Gate module. As the number of iterations increases, the uncertainty (blue for zero, red for maximum) of bulk regions converges to almost zero. The regions marked by yellow boxes represent the most uncertain spots in the image. The segmentation maps and uncertainty maps of the second image is compared (rightmost). The performance of the second image was much better than those of the first image in terms of accuracy and uncertainty by updating the initial weight of the guide model with the weights of the segmentation model pre-trained using TEM images.

Figure 8 illustrates the uncertainty maps calculated at various iterations when generating the GT labels with the segmentation model #1. The uncertainty value for each pixel is computed using Eq. (1). Yellow boxes in the segmentation maps indicate the most uncertain spots in each image. This recommendation makes user to add weak labels to the spot and can accelerate to reduce the uncertainty and improve the accuracy. The uncertainties were high for most pixels until the 4-th iteration. They were reduced to almost zero at the 9-th iteration except the regions around boundaries. The last column in Figure 8 illustrates the segmentation results of the guide model during the training with the model #2. The segmentation map at the initial inference shows each label was correctly assigned to the corresponding materials. As the guide model was fine-tuned with the pre-trained weights using the segmentation model #1, the accuracy of the initial segmentation map with the model #2 was 98.2% which was higher than the accuracy at the 8th iteration of 97.9%.

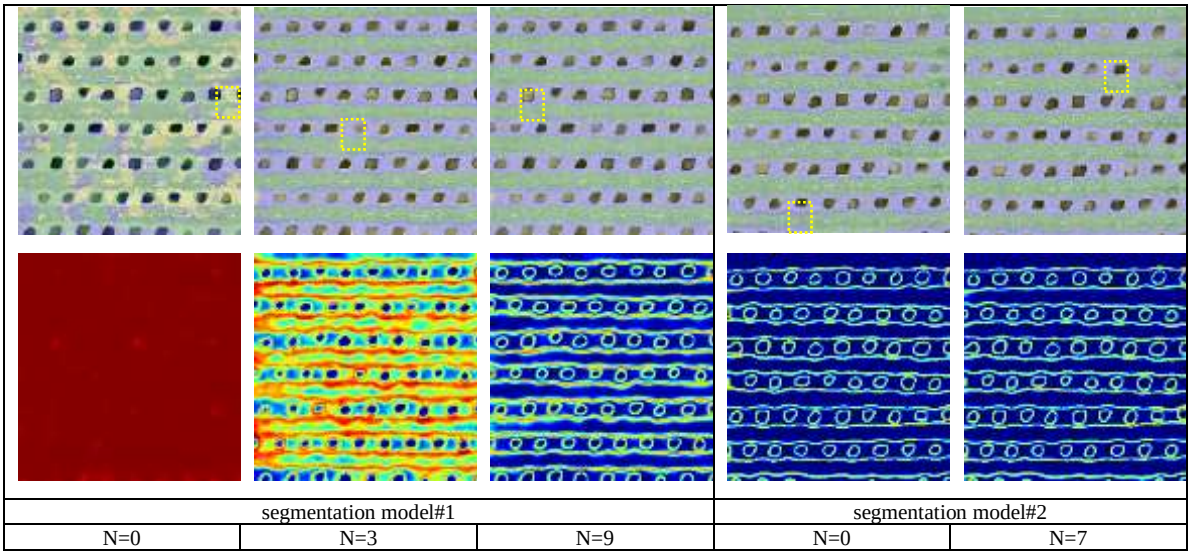

Figure 9. Segmentation maps and uncertainty maps of the guide model during the training of the first image of DRAM Active module. The regions marked by yellow boxes represent the most uncertain spots in the image. The segmentation maps and uncertainty maps of the second image is compared (the rightmost two columns). The performance of the second image was much better than those of the first image in terms of accuracy and uncertainty by updating the initial weight of the guide model with the weights of the segmentation model pre-trained using TEM images.

Figure 9 illustrates the segmentation maps and uncertainty maps calculated at various iterations when generating the GT labels for the image of DRAM Active module. The accuracy of the model #1 at iteration 9 was similar to that of the model #2 at iteration 7. However, the uncertainty of the latter was 0.16 compared to that of the former of 0.29.

4. CONCLUSIONS

In this work, we proposed a method to generate the GT labels with a limited amount of user interactions and segment a TEM image into structures for automatically measuring the CDs. Based on weak labeling and active learning, the proposed method significantly reduced the time required to generate GT labels corresponding to TEM images of semiconductor devices. The performance of the guide model can be enhanced as its weights are updated. It will be continued as the improved guide models generate more GT labels and the segmentation model is trained.

In the proposed method, we achieved an accuracy of 98% within less than 10% of the time required by the manual method. The strategy of using pre-trained weights as initial weights allowed the accuracy to converge faster. Additionally, we were able to improve the segmentation accuracy of the GT label and reduce uncertainties by adding weak labels to the most uncertain spots.

This method will reduce segmentation costs in TEM image analysis for semiconductor device development and enable automatic measurement of CDs. Furthermore, it can be applied to other images, such as SEM or optical images obtained from most equipment in the field of metrology and inspection.

ACKNOWLEDGEMENTS

The authors would like to thank to the members of the MI Equipment R&D Team at Samsung Electronics Co., Ltd. for their insightful discussions.

REFERENCES

[1] E. Ma, K. Chou, et al., "Multi-beam inspection (MBI) for the 7nm node and beyond: technologies and applications", Proc. of SPIE 10959, Metrology, Inspection, and Process Control for Microlithography XXXIII, (2019).
[2] Yoon, C., et al., "Toward realization of high-throughput hyperspectral imaging technique for semiconductor device metrology," Journal of Micro/Nanopatterning, Materials, and Metrology 21(2), 021209 (2022).
[3] Son, J., et al., "Massive overlay metrology solution by realizing imaging Mueller matrix spectroscopic ellipsometry," Proc. SPIE 12496, Metrology, Inspection, and Process Control XXXVII, (2023).
[4] Esashi, Y., et al., "Tabletop extreme ultraviolet reflectometer for quantitative nanoscale reflectometry, scatterometry, and imaging," Rev. Sci. Instrum. 94, 123705 (2023).
[5] Jung, J. et al., "Multi spectral holographic ellipsometry for a complex 3D nanostructure," Opt. Express 30(26) 46956-46971 (2022).
[6] Jeong, U. et al., "Development of highly dense material-specific fluorophore labeling method on Silicon-based semiconductor materials for three-dimensional multicolor super-resolution fluorescence imaging," Chem. Mater. 2023, 35, 14, 5572-5581 (2023).
[7] Wang, H. et al., "Fourier ptychographic topography," Opt. Express 31(7) 11007-11018 (2023).
[8] Kwon, S., et al., "Microsphere-assisted, nanospot, non-destructive metrology for semiconductor devices," Light Sci. Appl. 11, 32 (2022).
[9] Park, H. and Shin, Y. "AI technology in semiconductor metrology and inspection," Next Generation Lithography Conference (2023).
[10] Park, H. and Jo, T. "MI technology enabling next-generation semiconductor development," SEMICON Korea (2024).
[11] Williams, D. B. and Carter, C. B. "Transmission electron microscopy: a textbook for materials science," Springer Science & Business Media (2009).
[12] Kirkland, E. J. "Advanced computing in electron microscopy," Springer Science & Business Media (2010).
[13] Schoenmakers, R., Peeman, M., Boughorbel, F. and Potocek, P "Method, device and system for remote deep learning for microscopic image reconstruction and segmentation," U.S. Patent No. US10903043(B2) (2021).
[14] Yumiba, R., Sakai, K. and Yamaguchi, S. "Measurement system, method for generating learning model to be used when performing image measurement of semiconductor including predetermined structure, and recording medium for storing program for causing computer to execute processing for generating learning model to be used when performing image measurement of semiconductor including predetermined structure," U.S. Patent No. US20220277434(A1) (2022).
[15] Baruch, E. B., Elkayam, S., Cohen, S. and Ben-Shlomo, T. "Segmentation of an image of a semiconductor specimen," U.S. Patent No. US20210407093(A1) (2023).
[16] Gonzalez, R. C. and Woods, R. E. "Digital image processing," Pearson Education India (2017).
[17] Pal, N. R. and Pal, S. K. "A review on image segmentation techniques," Pattern recognition, 26(9), 1277-1294 (1993).
[18] Long, J., Shelhamer, E. and Darrell, T. "Fully convolutional networks for semantic segmentation," In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440) (2015).
[19] Ronneberger, O., Fischer, P. and Brox, T. "U-net: Convolutional networks for biomedical image segmentation," In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241) (2015).
[20] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A. L. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848 (2018).
[21] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., et al. "Segment anything," arXiv e-print (2023).

[22] Zhang, Y., Song, H. and Liu, Y. “Weakly supervised semantic segmentation with a deep convolutional neural network,” Journal of Visual Communication and Image Representation, 58, 357-364 (2019).
[23] Sun, S., Xian, M., Xu, F., Yao, T. and Capriotti, L. “CFR-ICL: Cascade-forward refinement with iterative click loss for interactive image segmentation,” arXiv e-prints (2023).
[24] Zhan, X., Wang, Q., Huang, K.-. hao., Xiong, H., Dou, D. and Chan, A. B. “A comparative survey of deep active learning,” Human in the Loop Learning (HiLL) Workshop at NeurIPS 2022 (2022).
[25] Kullback, S., Leibler, R. A. “On information and sufficiency,” Ann. Math. Statist. 22(1): 79-86 (1951).
[26] Goldstein, J., Newbury, D. E., Echlin, P., Joy, D. C., Lyman, C. E., Fiori, C., and Lifshin, E. “Scanning electron microscopy and X-ray microanalysis,” Springer (2018)
[27] Goodfellow, I., Bengio, Y., Courville, A., “Deep learning: adaptive computation and machine learning,” MIT Press, Cambridge, MA, USA, 2016.
[28] Shannon, C., “A mathematical theory of communication,” The Bell System Technical Journal, 27(3): 379-423 (1948).