공학석사학위논문

반도체 제조 공정을 위한 GAN 기반 이종 이미지 정렬 체계

GAN based Multimodal Image Alignment Framework for Semiconductor Manufacturing

2019년 8월

서울대학교 대학원

기계항공공학부

주 성 호

반도체 제조 공정을 위한 GAN 기반 이종 이미지 정렬 체계

GAN based Multimodal Image Alignment Framework for Semiconductor Manufacturing

지도교수 김 도 년

이 논문을 공학석사 학위논문으로 제출함

2019년 4월

서울대학교 대학원

기계항공공학부

주 성 호

주성호의 공학석사 학위논문을 인준함

2019년 6월

위 원 장 ______________________________

부위원장 ______________________________

위 원 ______________________________

Abstract

In semiconductor manufacturing process, visual inspection on wafer using template-based detection is widely researched topic. As a prerequisite of detection process, new demand for aligning multimodal image has emerged. To address this issue, this paper proposes a framework with GAN based image translation followed by NCC based template matching algorithm. Different from previous function based approaches, our deep learning based framework effectively transforms an image to another domain where template matching is much easier. Also, for practical usage, we propose a new training data generation strategy which allows our model to train from only 20 pre-aligned images. Experimental results on 4 types of manually aligned data, consisted of 400 pairs of images, demonstrate that our method successfully transforms image regardless of the presence of defect or noise. Also, using transformed image, alignment process with NCC based template matching achieved almost 100% accuracy on every types of image. Moreover, our framework shows great efficiency as it takes only 15 minutes for training and 0.25 seconds per image in test time.

Keywords : Multimodal Image, Alignment, GAN, NCC, Template Matching

Student Number : 2017-27667

Table of Contents

List of Figures

List of Tables

1. Introduction

In semiconductor manufacturing process, visual inspection on wafer is widely researched topic. According to their purpose, wafers are scanned by various apparatus such as scanning electronic microscopes (SEM), cameras, and other optical instruments. In the smallest scale, SEM image is used to find problematic spots(hotspots) in patterns. In inspecting SEM image, one of commonly used method is template-based detection which employs comparison between defect-free template and test image. [1]–[7] This is also prevailing method in other industries. [8]–[11]

Although template-based detection is intuitive and robust, this approach is prone to geometric misalignment such as rotation, scaling, and translation (RST). In real world data, not only RST but also variations in light condition and noise in image make alignment task more burdensome. [12] Specifically, for wafer hotspot detection, image registration task is even different from other industrial cases. In most of works from other industries, image alignment task is between images from same domain(mono-modal). On the other hand, in semiconductor manufacturing process, demand for aligning multimodal images has emerged. [13] As shown in Figure 1, design layout image (CAD) and SEM image are similar to some extent as CAD is supposed to be target image of SEM image. However, due to SEM's noisy nature and sensitive manufacturing procedure, they show significant difference in texture of image and detailed shape of figures. While comparing CAD and SEM image for inspection has clear benefit, image registration became more difficult task than mono-modal case. Yet, only few works address this issue relying on function-based methods. Morokuma *et al.* [14] , Hibino *et al.* [15], and Bisschop *et al.* [16] proposed contour-based method. They first extract contours of figures in SEM image and align them with simulated contour which is derived from CAD. In other words, this method suggests some type of mapping from one image domain to another for ease of alignment. However, it takes time for simulation and relies heavily on quality of contour extraction. Recently, Tseng *et al.* [17] proposed pattern density-based searching method which involves heuristic contour extraction from SEM image. Yet, heuristic contour extraction process is easily affected by SEM image's texture and this leads to inaccurate matching result.

As mentioned above, mapping between template and test image is one way of solving multimodal image registration problem. To overcome problems stated above, we propose a framework with generative adversarial net (GAN) based image translation followed by template matching algorithm. GAN [18] is a type of generative network which targets to produce realistic output. Compared to traditional methods such as autoencoder and variational autoencoder, it showed great performance especially in image domain. After image translation, template matching is performed in CAD domain for several reasons. CAD is binary and free of noise. Also, it shows clear distinction between pattern and background. Therefore, our GAN model transforms SEM image to CAD. Moreover, to deal with shortage of training data, which is one of common obstacle of deep learning application, new training set generation strategy is introduced.

We would like to inform that all SEM images and CAD shown in paper are not real images but generated images. Also, we use 'template' as target image to be found and 'domain' as search image which contains template. The remainder of paper is organized as follows. In Section 2, proposed framework is described in detail. In Section 3, experimental results of quality of generated image and accuracy of template matching is presented. Lastly, conclusion is in Section 4.

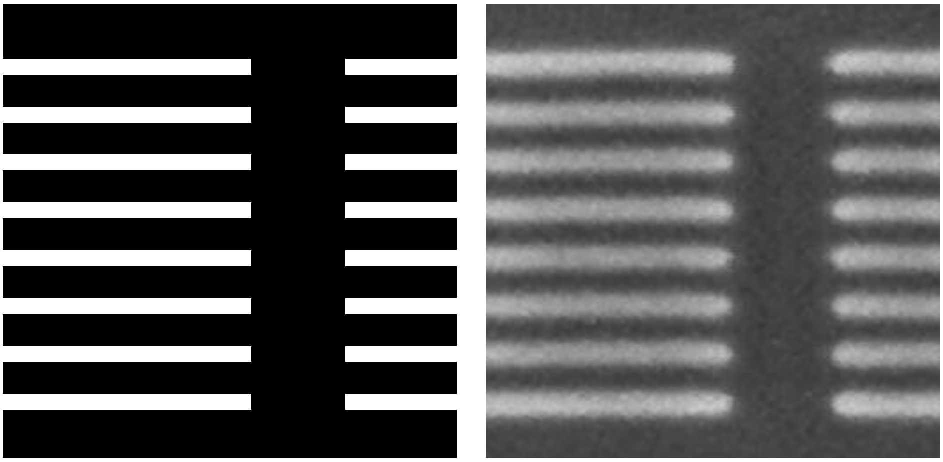


Figure 1. Samples of CAD (left) and SEM image (right)

2. Proposed Framework

In this section, we describe alignment procedure employing conditional GAN and NCC based template matching. Our framework includes image preprocessing, training data generation, conditional GAN training, and template matching. (Figure 2) Following section describes the detailed procedure.

2.1 Training image generation and image preprocessing

For training image generation, minimum number of aligned image set is required and they are prepared manually. This is necessary procedure as texture of SEM image and configuration of pattern differ for every manufacturing condition.

From the set of aligned images, training data for GAN are obtained after preprocessing steps. Aligned image refers to 2-layer image (Figure 3.A) where each layer contains CAD and SEM image respectively. First, wafer background area is specified and removed. This area of image is irrelevant to training but randomly appears in real data. Also, this area becomes noise in training procedure. As wafer background area always has same pixel value, we can specify them in image. However, as SEM image is gray-scale, other spots with same pixel value happens to exist in normal area like a salt and pepper noise. (Figure 3.B) Therefore, wrong spots are removed using median filter.

After removing wafer background area, aligned pairs of SEM image and CAD are cropped from remainder of image to be a training data. Size of cropped image depends on the size of original image.

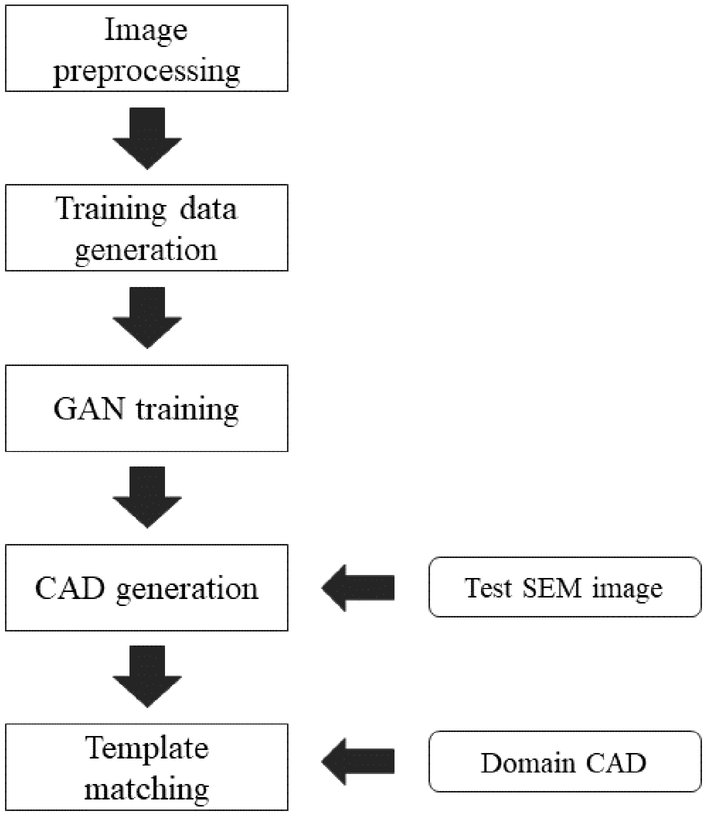
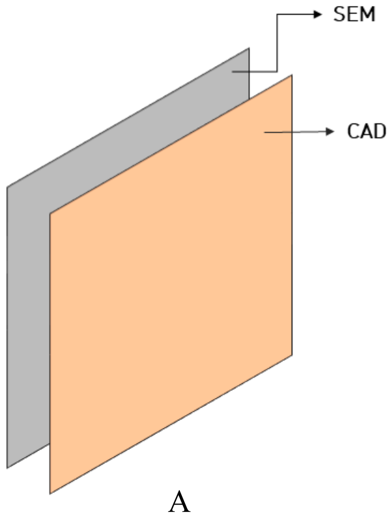
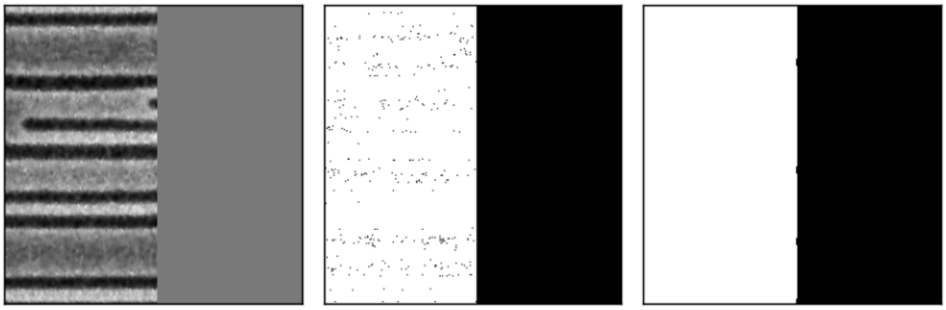


Figure 2. The framework of proposed alignment method.



A



B

Figure 3. (A) Illustration of 2-layer image. (B) Removal of wafer background. From original image(left), pixel value of wafer background is specified(center) and applied median filter to obtain clear distinction(right).

2.2 GAN based image translation and template matching

Our GAN model adapts the architecture from Isola *et al.* [19] who have shown very stable and general performance for many applications. One of advantage of this network for our task is that our data are paired. Even though this network is only capable of learning with paired dataset, it showed great accuracy over other networks given right data.

Unlike original Pix2Pix model, single channel image is used as input and output for generator. Accordingly, 2-layer image is fed into discriminator. The generator is based on U-net [20] and discriminator follows the concept of patchGAN. Also, we modified the architecture depending on the resolution of input image. When we train GAN with smaller size image (128 ×128 pixels), 1 convolution layer and 1 deconvolution layer are removed from each of encoder and decoder compared to bigger size 256 × 256 pixels. Also, 1 convolution layer is removed from discriminator as well. Consequently, receptive field of discriminator is 34 × 34 pixels and 70 × 70 pixels for 128 ×128 and 256 × 256 image respectively. However, we could not find significant difference from change of receptive field for our task.

To explain training procedure briefly, from the 2-layer training image, we separate SEM image and CAD which are fed into generator as input and target respectively. Then, generated CAD and corresponding input SEM image are given to discriminator to be judged as fake sample. Similarly, real CAD and corresponding SEM image are given to discriminator to be judged as true sample. (Figure 4) Training details are further discussed in Section 3.

Once GAN is trained, generator is separated from the model. Then, test images are cropped from SEM image of interest and fed into generator. Test images are periodically cropped from SEM image that they can be recollected to make original size image again. (Figure 5) Generated CAD are reconnected to be full size image and pass through median filter to eliminate minor noise. Next, filtered output and CAD of interest are supplied to NCC based template matching method as template and domain respectively. As a threshold, a value is set as 97% of highest NCC score within image. As the result shows block of coordinates, we calculate the centroid of proposed region [21]. This enables our model to acquire subpixel accuracy. Using calculated centroid, original SEM image matches with domain CAD.

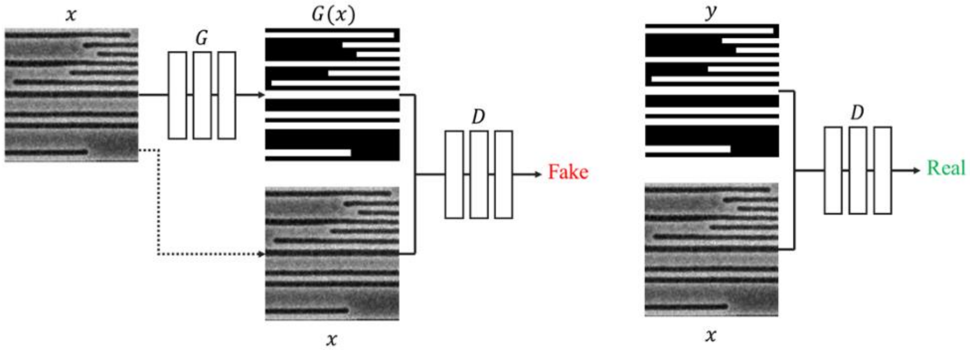


Figure 4. Illustration of training procedure of GAN model. G and D implies generator and discriminator. x, G(x), and y indicate input SEM image, generated CAD, and ground truth CAD respectively.

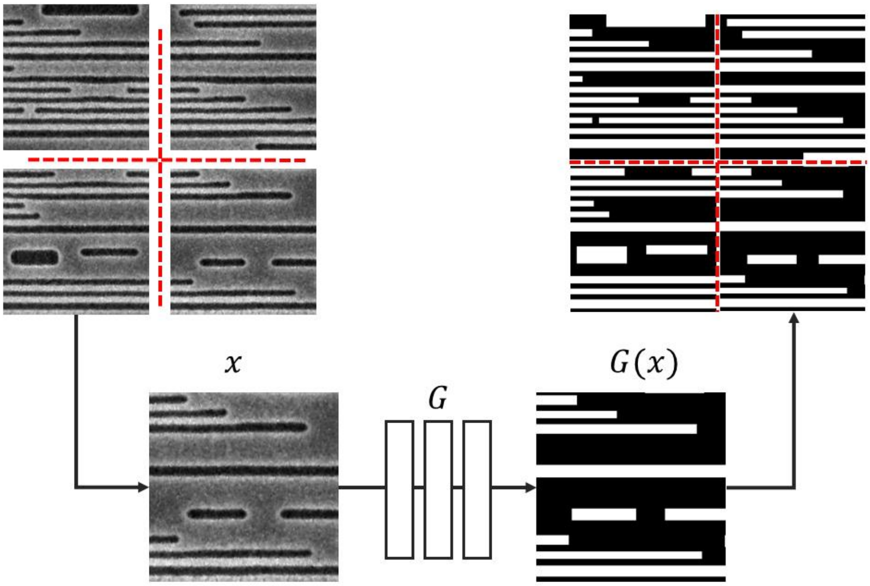


Figure 5. Illustration of test image generation process. First, full size image (512 × 512) is divided into 4 pieces (256 × 256) or 16 pieces (128 × 128). Each image passes through generator and generated image is recollected to be full size image again.

3. Experimental Results

In this section, our framework is evaluated using 400 image pairs from 4 types of real-world data. Each type consists of 100 pairs of CAD and SEM images which are manually aligned. We tried to select images with diverse pattern and texture. Samples of 4 types of images are shown in Figure 6. All implementations are coded in Python 3.7.3 and Tensorflow 1.13.1 with single Pascal Titan XP GPU and Intel Xeon Silver 4114 CPU.

In following section, each step of alignment is evaluated. Firstly, quality of generated image is tested. Next, accuracy of NCC based template matching is checked. Finally, running time of each process is calculated.

3.1 Performance of image generation

To check the quality of generated image, we used NCC score as a metric. Although it is not common metric for evaluating GAN performance [22], it is reasonable choice since we perform template matching based on NCC algorithm. Thereby, we evaluated the NCC score between target CAD and generated CAD.

For GAN training, 20 pairs and 80 pairs of aligned images are used as training and test image set respectively for every types of data. Training and test image set are chosen randomly for every case. For initial 20 images, resolution of image is 512 × 512 pixels. To test the effect of size of training data, we randomly cropped 120 images and 480 images from each 512 × 512 images for case A (256 × 256 pixels) and case B (128 × 128 pixels) respectively. Also, for fair comparison, case A, and B shares same training and test images.

For training details, we set hyperparameters following [19] as reference. It was especially important to find appropriate batch size and total epoch because these hyper parameters are proportional to training time. However, too large batch or small total epoch ruins generation quality. We found that batch size 4 and 17 epoch generate satisfying images for our case which are same as day and night transformation from [19].

After the training, test images pass through generator and are recollected to be original resolution of 512 × 512 pixels as described in Figure 5. Then, NCC score between target CAD and recollected output is evaluated. In evaluation procedure, samples with NCC score below 0.7 are treated as outliers. (Figure 8)

Figure 7 and Figure 8 illustrate generated images and achieved NCC score for each type respectively. As shown in Figure 7, GAN model generates accurate CAD regardless of image type or size of training image. Also, generated images show characteristics of CAD such as sharp edge. Quality of generated image along with high NCC score denote that our model properly transforms SEM image into CAD.

Figure 9 shows that distribution of NCC score is similar between case A, and B for every types of images. This implies that our modification on architecture of neural net does not have significant effect. Also, change of the number of training data (Case B has 4 times more training data than Case A) does not have significant effect too. However, this seems to be an obvious outcome because total information from which model can learn is limited by the amount of information initial training images (512 × 512 pixels) possess. Therefore, we can assume that the constitution of initial training data is important. Here, constitution of data is the diversity of pattern that data contains.

This argument is also verified with analysis of outliers. Red bars in Figure 8 show outliers. A sample of outlier is presented in Figure 10. Case A and Case B from Type 1 have 3 outliers each and they are originated from same test images. Outliers from Type 1 have one common point that they contain patterns that are not in training image set. Images which are not outlier but have relatively low NCC score also contain unseen patterns but with less portion. However, unseen pattern rarely presents in image and occupies only small parts of image. Therefore, even outliers are aligned accurately by template matching algorithm as majority of images are transformed well.

In Figure 9, difference of mean NCC score between each type of image is illustrated. Mean NCC score is highest in Type 1 and lowest in Type 3. We presume that this is because Type 1's CAD is more deterministic than Type 3. Deterministic CAD means that only one CAD can be derived from corresponding SEM image. However, in Type 3, patterns in SEM image and CAD are hardly one-to-one correspondence. Patterns in SEM image show huge variance as figures shrink differently even from same pattern. Also, Type 3 image has more defects than other types. Frequency of defects and the variance of SEM image is almost impossible to predict because of complex manufacturing procedure. These factors negatively affect rule-based mapping algorithms such as intensity-based functions. However, our model generates satisfactory CAD regardless of presence of defects or non-deterministic characteristic of SEM image. As a result, any type of image is aligned precisely by template matching algorithm.

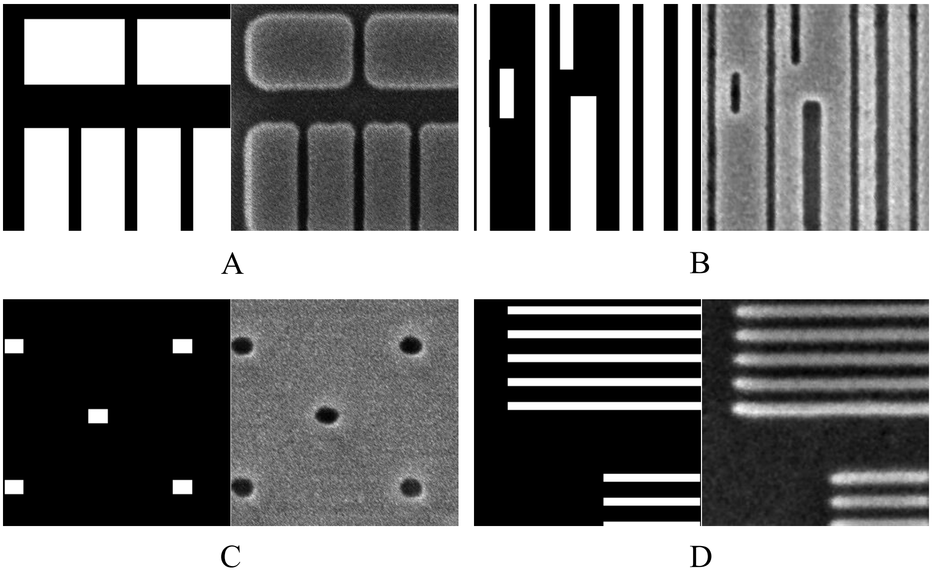


Figure 6. Samples of 4 types of images. (A) Type 1. (B) Type 2. (C) Type 3. (D) Type 4.

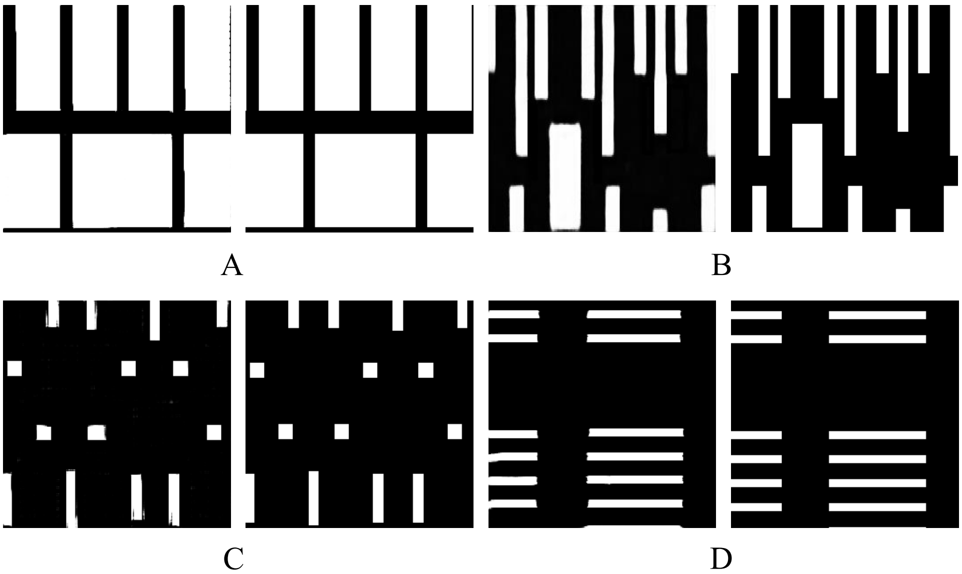


Figure 7. Generated CAD (left) and target CAD (right) for each type.

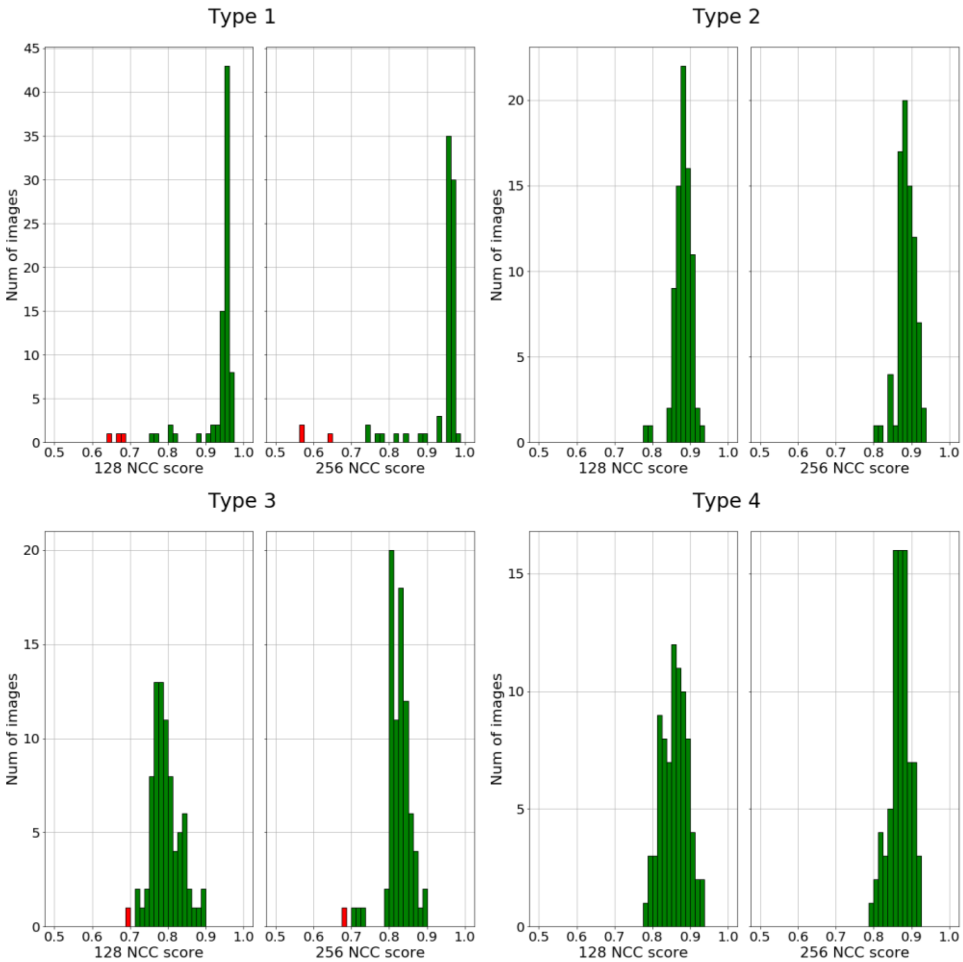


Figure 8. Histograms of NCC score between generated CAD and target CAD for case A (256 × 256) and case B (128 × 128) of each type.

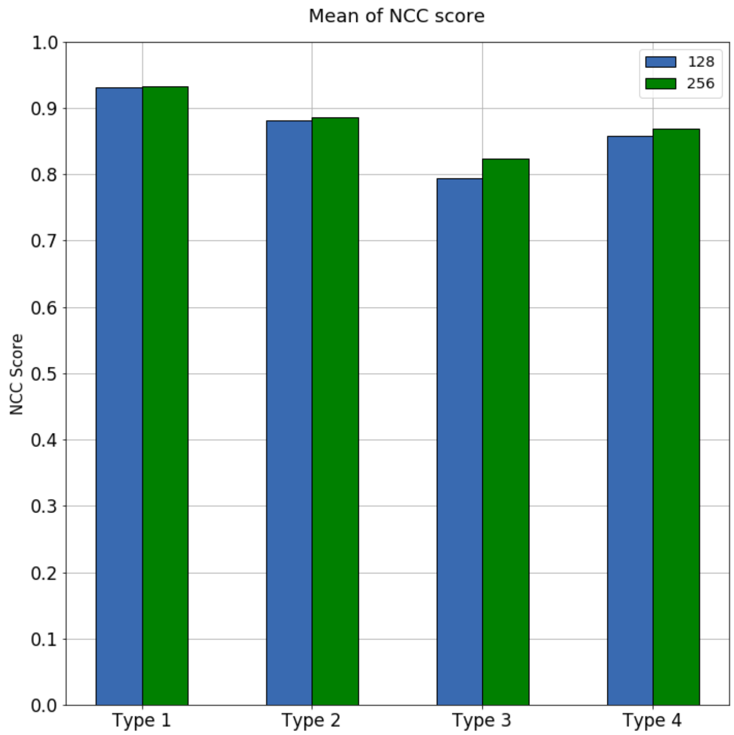


Figure 9. Histogram of mean NCC score between generated CAD and target CAD.

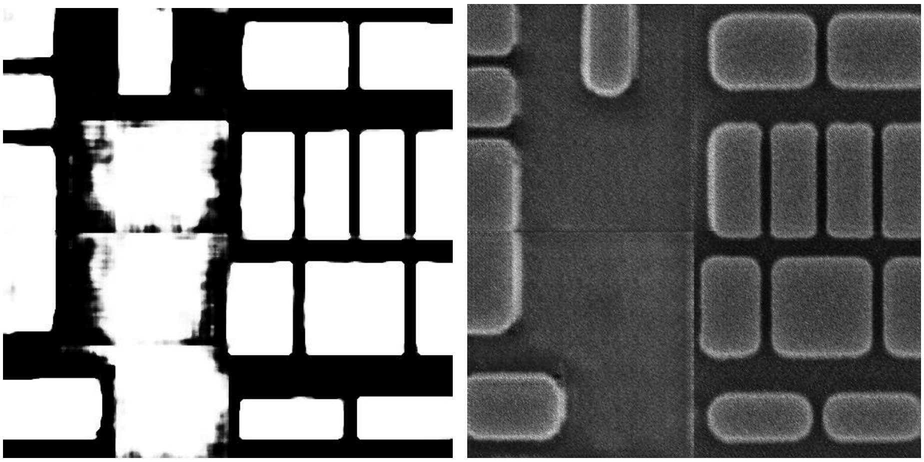


Figure 10. Example of outlier for Type 1 case A (256 × 256) and input SEM image.

3.2 Accuracy of template matching

To perform the template matching experiment, we defined the characteristic of alignment task based on two assumptions: 1) The amount of misalignment does not exceed 100 pixels for both x and y direction. and 2) There is only one unique match within domain. Although our framework is capable of finding match regardless of the size of domain, we limited the amount of misalignment to decide the size of domain. We also presume that 100-pixel difference is practical value for registration task. Also, the model cannot decide which one is the right match if there are multiple possibilities. However, our framework can find multiple matches if necessary by picking multiple high score coordinates.

Based on first assumption, resolution of domain is 712 $\times$ 712 pixels which centers 512 $\times$ 512 target CAD. Consequently, the target coordinate is (100,100) which is originally aligned manually. In evaluating the result, coordinate difference up to 3 pixels is considered correct match considering uncertain boundary of patterns in SEM image. The result is summarized in Table 1. Our framework showed almost 100% accuracy on every type of images. Also, even wrong matches hardly show misalignment.

		X coordinate accuracy (%)	Y coordinate accuracy (%)
Type 1	128	100	100
	256	100	100
Type2	128	100	98.75 (79/80)
	256	100	98.75 (79/80)
Type 3	128	100	100
	256	100	100
Type 4	128	100	100
	256	98.75 (79/80)	100

Table 1. Accuracy of template matching result. Difference up to 3 pixels is considered correct.

3.3 Running time of framework

Similar to other deep learning algorithms, most of time is spent in training session. For training time, it takes 15 minutes for 256 × 256 case (2400 training images) and 1 hour for 128 × 128 case (9600 training images). In test time, it takes 10 seconds for both 256 × 256 case (320 test images from 80 images) and 128 × 128 case (1280 test image from 80 images). We used biggest batch size possible as we found that batch size in test time does not affect quality of output data. After generation process, template matching process takes 10 seconds for both cases. However, our framework can speed up in test time with more data as it takes 5 seconds to restore checkpoint. Except for initial restoration process, it takes 0.0625 seconds for four 256 × 256 images which is one original 512 × 512 image. Same applies to 128 × 128 case. The result is summarized in Table 2.

	Training time(s)	Average test time (s/image)
128	3600	0.25
256	900	0.25

Table 2. Summary of running time of framework. Average test time is sum of generation and template matching time.

4. Conclusion

Multimodal image registration became important for multiple tasks in semiconductor industry. Especially, accurate alignment can lead to successful inspection. In this paper, we propose multimodal image registration framework based on GAN and NCC based template matching. This framework is tested with manually aligned data which is similar to real images. To cope with small number of training data, we proposed new strategy which divides original aligned data. From only 20 pairs of original images, image translation process showed great accuracy regardless of the texture of input image. Also, we found that our model successfully generates target images from input with defects and non-deterministic property. With generated image, template matching result shows almost 100 % accuracy for every test image cases. Even wrongly predicted results are in fact not distinguishable from correct results. Moreover, our framework demonstrates great efficiency that it takes around 0.25 second per image in test time and this can be improved with larger test set. In conclusion, our learning based framework have shown great potential to be an excellent solution for practical usage in semiconductor manufacturing industry.

References

[1] P. Xie and S. U. Guan, “A golden-template self-generating method for patterned wafer inspection”, *Machine Vision and Applications*, vol 12, no. 3, pp 149–156, 10 2000.
[2] D. Tsai, I. Chiang and Y. H. Tsai, “A Shift-Tolerant Dissimilarity Measure for Surface Defect Detection”, *IEEE Transactions on Industrial Informatics*, vol 8, no. 1, pp 128–137, 2 2012.
[3] M. Zontak and I. Cohen, “Defect detection in patterned wafers using multichannel Scanning Electron Microscope”, *Signal Processing*, vol 89, no. 8, pp 1511–1520, 8 2009.
[4] N. G. Shankar and Z. W. Zhong, “Defect detection on semiconductor wafer surfaces”, *Microelectronic Engineering*, vol 77, no. 3, pp 337–346, 4 2005.
[5] H. W. Kim and S. I. Yoo, “Defect detection using feature point matching for non-repetitive patterned images”, *Pattern Anal Applic*, vol 17, no. 2, pp 415–429, 5 2014.
[6] C. Wang, B. C. Jiang, J. Lin and C. Chu, “Machine Vision-Based Defect Detection in IC Images Using the Partial Information Correlation Coefficient”, *IEEE Transactions on Semiconductor Manufacturing*, vol 26, no. 3, pp 378–384, 8 2013.
[7] X. Bai, Y. Fang, W. Lin, L. Wang and B. Ju, “Saliency-Based Defect Detection in Industrial Images by Using Phase Spectrum”, *IEEE Transactions on Industrial Informatics*, vol 10, no. 4, pp 2135–2145, 11 2014.
[8] H. Wang, J. Zhang, Y. Tian, H. Chen, H. Sun and K. Liu, “A Simple Guidance Template-Based Defect Detection Method for Strip Steel Surfaces”, *IEEE Transactions on Industrial Informatics*, vol 15, no. 5, pp 2798–2809, 5 2019.
[9] J. Kwak, K. B. Lee, J. Jang, K. S. Chang and C. O. Kim, “Automatic inspection of salt-and-pepper defects in OLED panels using image processing and control chart techniques”, *J Intell Manuf*, vol 30, no. 3, pp 1047–1055, 3 2019.
[10] C.-F. J. Kuo, C.-H. Tsai, W.-R. Wang and H.-C. Wu, “Automatic marking point positioning of printed circuit boards based on template matching technique”, *J Intell Manuf*, vol 30, no. 2, pp 671–685, 2 2019.
[11] C. Jian, J. Gao and Y. Ao, “Automatic surface defect detection for mobile phone screen glass based on machine vision”, *Applied Soft Computing*, vol 52, pp 348–358, 3 2017.
[12] H. Kong, J. Yang and Z. Chen, “Accurate and Efficient Inspection of Speckle and Scratch Defects on Surfaces of Planar Products”, *IEEE Transactions on Industrial Informatics*, vol 13, no. 4, pp 1855–1865, 8 2017.
[13] C. Tabery and L. Page, “Use of design pattern layout for automatic metrology recipe generation”, in *Metrology, Inspection, and Process Control for Microlithography XIX*, 2005, vol 5752, pp 1424–1434.
[14] H. Morokuma, A. Sugiyama, Y. Toyoda, W. Nagatomo, T. Sutani and R. Matsuoka, “A new matching engine between design layout and SEM image of semiconductor device”, in *Metrology, Inspection, and Process Control for Microlithography XIX*, 2005, vol 5752, pp 546–558.
[15] D. Hibino et al., “High-accuracy OPC-modeling by using advanced CD-SEM based contours in the next-generation lithography”, presented at the SPIE Advanced Lithography, San Jose, California, 2010, p 76381X.
[16] P. D. Bisschop and J. V. de Kerkhove, “Alignment and averaging of scanning electron microscope image contours for optical proximity correction modeling purposes”, *JM3*, vol 9, no. 4, p 041302, 10 2010.

[17] S. S.-E. Tseng, W.-C. Chang, I. H.-R. Jiang, J. Zhu and J. P. Shiely, “Efficient search of layout hotspot patterns for matching SEM images using multilevel pixelation”, in *Optical Microlithography XXXII*, 2019, vol 10961, p 109610B.
[18] I. J. Goodfellow et al., “Generative Adversarial Networks”, *arXiv:1406.2661 [cs, stat]*, 6 2014.
[18] P. Isola, J.-Y. Zhu, T. Zhou and A. A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks”, *arXiv:1611.07004 [cs]*, 11 2016.
[19] O. Ronneberger, P. Fischer and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, *arXiv:1505.04597 [cs]*, 5 2015.
[20] W.-Y. Chang, C.-C. Chiu and J.-H. Yang, “Block-Based Connected-Component Labeling Algorithm Using Binary Decision Trees”, *Sensors (Basel)*, vol 15, no. 9, pp 23763–23787, 9 2015.
[22] A. Borji, “Pros and Cons of GAN Evaluation Measures”, *arXiv:1802.03446 [cs]*, 2 2018.
[23] H. Shindo기타, “High-precision contouring from SEM image in 32-nm lithography and beyond”, in *Design for Manufacturability through Design-Process Integration III*, 2009, vol 7275, p 72751F.
[24] M. Abadi기타, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems”, *arXiv:1603.04467 [cs]*, 3 2016.

Abstract (Korean)

반도체 공정에서 템플릿을 이용한 비전 기반의 웨이퍼 검사는 널리 연구되는 분야이다. 이러한 검사 과정의 전제 조건으로 멀티모달 이미지 정렬에 대한 새로운 요구가 대두되었다. 이 문제를 해결하기 위해 본 논문은 GAN을 활용한 이미지 변환과 NCC 기반의 템플릿 정렬 알고리즘을 이용한 프레임워크를 제안한다. 이전의 함수 기반 접근법과 달리 딥러닝 기반 프레임워크는 이미지를 템플릿 정렬이 훨씬 용이한 도메인으로 효과적으로 변환한다. 또한 실용적인 관점에서 고안한 새로운 학습 데이터 생성 방법을 통해 오직 20개의 정렬된 초기 데이터를 통해서 딥러닝 모델을 성공적으로 학습할 수 있다. 각각 100쌍의 이미지로 이루어진 4가지 종류의 수작업으로 정렬한 데이터를 사용한 실험 결과를 통해 고안한 방법이 결함이나 노이즈의 존재여부와 상관없이 효과적으로 이미지를 변환한다는 것을 확인할 수 있다. 또한 변환된 이미지를 사용한 NCC 기반의 템플릿 정렬 알고리즘은 이미지 정렬에서 100%에 가까운 정확도를 보인다. 마지막으로 소요 시간에서 프레임워크는 학습에 15분, 테스트 시 이미지당 0.25 초 만을 소모하며 높은 효율을 보인다.