An Active Deep Learning Method for the Detection of Defects in Power Semiconductors

Marco Bellini
Hitachi ABB Power Grids,
Semiconductors
Lenzburg, Switzerland
marco.bellini@hitachi-powergrids.com

Georges Pantalos
Department of Mechanical Engineering
ETH
Zürich, Switzerland
georgespantalos@gmail.com

Peter Kaspar
Hitachi ABB Power Grids,
Semiconductors
Lenzburg, Switzerland
peter.kaspar@hitachi-powergrids.com

Lars Knoll
Hitachi ABB Power Grids,
Semiconductors
Lenzburg, Switzerland
lars.knoll@hitachi-powergrids.com

Luca De-Michielis
Hitachi ABB Power Grids,
Semiconductors
Lenzburg, Switzerland
luca.de-michielis@hitachi-powergrids.com

***Abstract*—Accurate detection of semiconductor defects is crucial to ensure reliability of operation and to improve yield by understanding and eradicating yield detractors. Recent advances in computer vision driven by Deep Convolutional Neural Networks (DCNN) and transfer learning have enabled novel techniques for defect detection and classification [1-7]. However, training neural networks requires very large datasets, even with transfer learning. This paper addresses this shortcoming by introducing for the first time the active learning approach for semiconductor devices. The proposed neural network can accurately identify defective dies with modest efforts in terms of annotating the image set. Finally, the feature maps of the DCNN are used to generate an unsupervised taxonomy of the semiconductor die defects, supporting further investigations to address yield detractors**

Keywords—Silicon defects, deep learning, active learning, convolutional neural network

I. Introduction

The impressively high yield of semiconductor fabrication is a result of the combined efforts in semiconductor processing, inspection, metrology, characterization and statistical analysis. Optical inspection of semiconductors relies on visually identifying rare inhomogeneities and abnormal structures that can result from wafer non-uniformity, lithography misalignment, contamination, particles, deviations in processing parameters. Although automatic optical inspection (AOI) tools are routinely used to detect defects, the variability in the shape, size, and appearance of defects requires final assessment and verification steps by human experts [1], which are extremely expensive and prone to subjectivity [2]. In fact, agreement between experienced operators can score as low as 43% and long-term repeatability less than 93% [3]. Reliable automated techniques able to distinguish real defects, shown in Fig. 1 from false positives, illustrated in Fig. 2, reduce the need for human review of AOI tools. Furthermore, such techniques allow further downstream data analysis, such as computation of similarity metrics and unsupervised clustering, providing insight in the origin of defects.

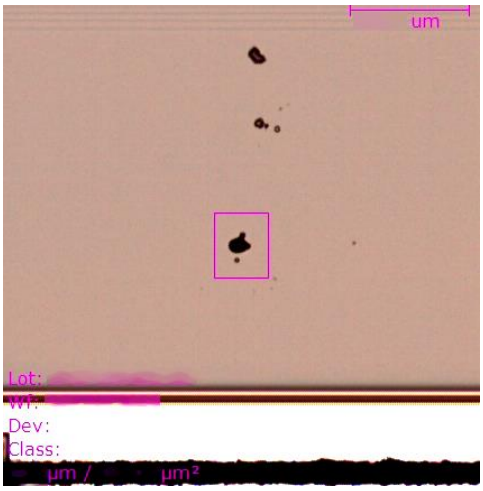

Fig. 1 Example of a true positive image. This image of a power semiconductor device shows 4 defects but only one is correctly identified by the automated optical inspection tool (AOI). The markings and bounding boxes displayed in the magenta color are automatically added by the AOI tool and cannot be removed. Sensitive details such as dimensions and identifiers have been redacted.

II. State of the Art

A. Defect classification techniques

Conventional AOI tools rely on classical computer vision (CV) algorithms to compare product images with golden reference images. Any discrepancy is marked as a defect depending typically on ad-hoc rules on defect dimensions and area. Techniques such as multi-threshold contrast-adjustment,

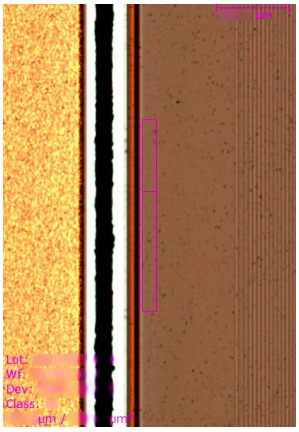

Fig. 2 Example of a false positive image. This image of a power semiconductor device contains no defects but the automated inspection tool wrongly proposes two adjacent defective regions.

morphological merging and segmentation can be used to improve the level of detection accuracy, as presented in [4].

Recent advances in data science are also applied to defect detection: [5] presents a fuzzy-rule inference algorithm detecting spatial patterns and [6] proposes a technique based on convolutional neural networks (CNN), singular value decomposition (SVD) for dimensionality reduction combined with extreme gradient boosting tree classifier (XGboost).

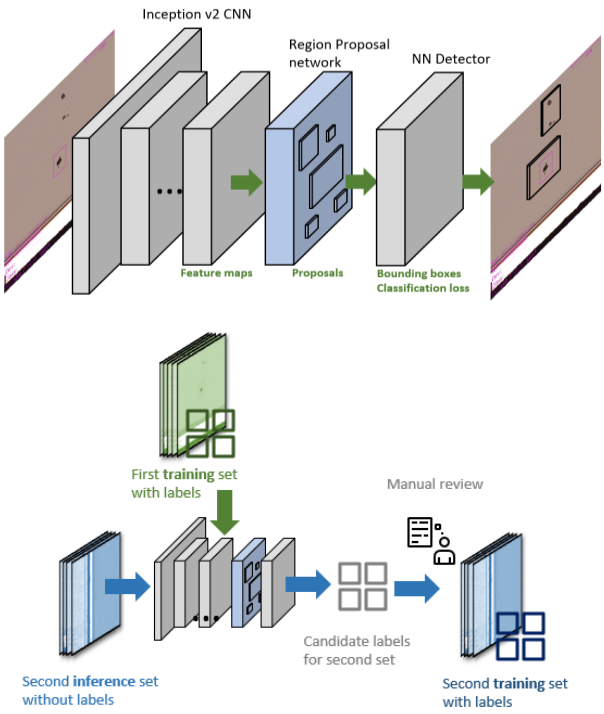

Fig. 3 The architecture of the Faster RCNN network is shown on the top. The image is fed first into an Inception v.2 Convolutional Neural Network (CNN) then to a region proposal network (RPN). The bottom illustration exemplifies schematically the active learning approach.

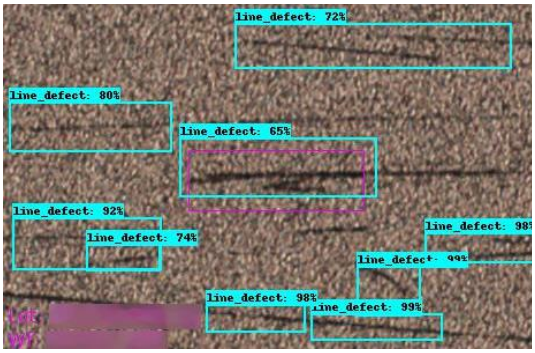

Fig. 4 Example of line defects. The annotation of the lines is broken into several bounding boxes to avoid including too much background.

Finally [7] introduces a defect classification approach based on the Inception version 2 DCNN architecture with weakly supervised initial training sets of 13,000-36,000 examples of cropped images of defects followed by finetuning training sets of 1,800-2,400 examples. Conversely, this work proposes a novel approach combining Faster Region-based Convolutional Neural Networks (Faster RCNN) with active learning to detect small defects with a training set of only 500-2,500 examples.

B. Advances in Deep Convolutional Neural Networks

Recent advances in GPU computing power and CNN architectures enable greater accuracy in classification tasks. Object detection is a much more challenging task, as the details to be detected may be small (in this work as small as 0.01% to 0.1% of the total image area). The Faster R-CNN architecture, proposed by [8] and available as a TensorFlow API from [9] combines high accuracy and fast training and inference times by processing the full image with a CNN to obtain the feature maps,

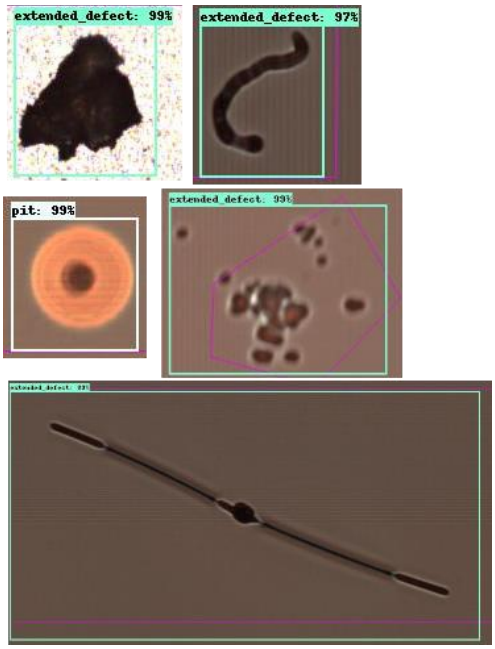

Fig. 5 Example of defect morphology and classification taxonomy. While several defects are very frequent, others are quite rare and exhibit large variations in shape, color or features.

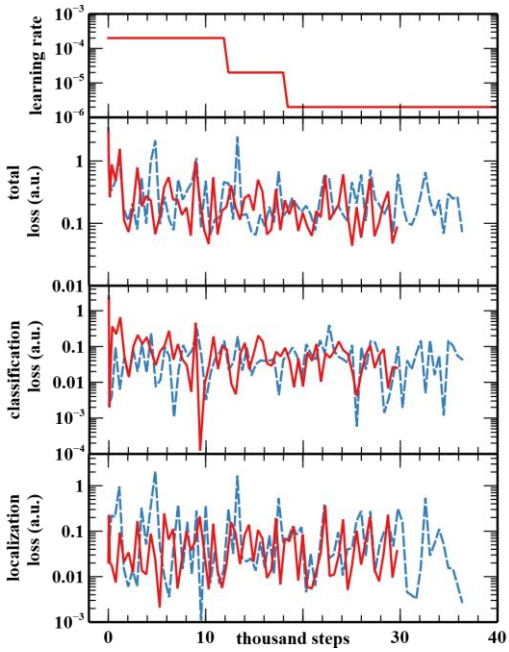

Fig. 6 The figure shows the learning rate and total loss, classification loss, and localization loss for the 1st (blue, dashed line) and 5th (red, solid line) learning cycle iterations.

then by extracting regions-of-interest (ROI) with a region proposition network (RPN), as illustrated in Fig. 3 (top). Several pre-trained networks are available, optimized for object detection on the MS COCO database, comprising 120,000 annotated images and 80 classes of daily objects [10].

III. Goals and Approach

Fig. 1 shows the characteristic output of the AOI tool for a true positive: the defect is identified by a pink bounding box. The image is annotated with the lot, wafer and product identifiers as well as with an image scale and dimensions. However roughly 13% of the time the AOI tool produces a false positive, as shown in Fig. 2, and a non-defective chip region is erroneously marked as a defect. The goal of the present work is discriminating the images provided by the AOI tool into a negative class ("defect") positive class ("false detection").

However, since many defects are minuscule and the decision hinges on very small area of the image, the goal is reframed as

TABLE I. Classification Results

Learning cycle	Active learning sets and results			
	training examples	*Annotation time*	*Precision*	*Recall*
1	314	8h	0.99	0.62
2	491	5h	0.97	.0.63
3	868	3h	0.99	0.71
4	1305	3h	0.98	0.74
5	1737	2h	0.99	0.79
6	2431	1.5h	0.98	0.81

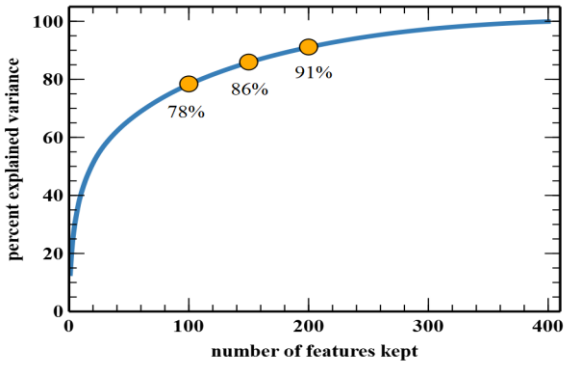

Fig. 7 The figure shows that Principal Component Analysis (PCA) reduces the dimensionality of the defects' feature maps from 16'348 to 200, retaining 91% of the variability (e.g. the information content).

an object detection problem, as indicated in Fig. 4 and 5: the defects are classified and located by a rectangular bounding box. Since often defects occur concurrently, as also shown in Fig. 4, the probability of correct image classification is greater than the probabilities of single defect detection as expressed by Eq. 1:

$$p_{image} = 1 - \prod_i (1 - p_{defect,i}) \quad (1)$$

Since the background is learned in every detection example, it is reasonable to expect a very low number of false positives.

A. Image Characteristics

This study analyzes microscopic, wafer-level images obtained from 5 different power semiconductor devices (IGBTs and power diodes of different voltage classes in the 1.2 kV to 6.5 kV range) after fabrication is completed. The features of

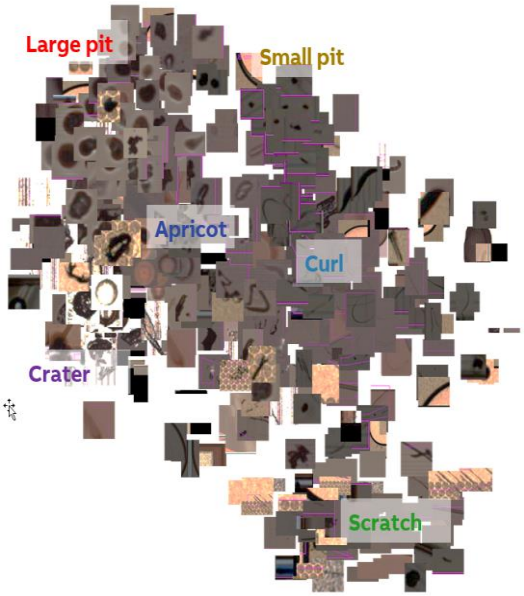

Fig. 8 Example of unsupervised defect taxonomy obtained through t-Distributed Stochastic Neighbor Embedding (t-SNE). The 200-dimension feature maps associated to the defects are mapped into a 2-dimentional space while preserving the relative entropy between similar objects.

devices photographed can be considered morphologically similar, even if there are differences in the designs and processing steps.

Moreover, magnification and focus of the images is automatically adjusted by the AOI tool. Image scale, size and to some extent brightness vary considerably between the images. Additionally, a small fraction of the images (<0.5%) may exhibit blurring or focusing issues.

B. Image Annotation

The disadvantage of object detection compared to classification is the effort needed to label examples: the defects are classified and annotated with bounding boxes.

Semiconductor defects occur in many sizes and shapes. Given that the AOI tool does not maintain fixed magnification and scale between the images, defect size in this context is to be understood as the size in pixels. Defects analyzed in this study span from a size of roughly ten to several hundred thousand pixels. The largest defects can also exceed the image size and are then cropped by the AOI tool.

The initial defect classes include pits, lines, curvilinear and extended defects. As illustrated in Fig. 4, lines need to be annotated in smaller segments, to avoid including too much background, which would confuse the network. Also, as shown in Fig. 5, some defects are rare and not easy to classify in clearly defined classes.

C. Active Learning

An active learning approach is used to reduce the effort of annotating the images. As shown in Fig. 3 (bottom) the network is initially trained on 300 images (roughly 10% of the training set). The model is run in inference mode, producing a candidate XML annotation file, on an equally sized unannotated image set, comprising example images with and without defects. The outcome is split into groups according to class confidence score: defects detected with confidence > 90%, defects not detected and the remainder. Then the output of the model is reviewed and corrected, producing a new labelled training set. As the network is trained with several active learning cycles, the review activity becomes less onerous: many images need only to be approved or necessitate only a minor correction.

D. Training

The training set is augmented with random brightness, color and contrast adjustments, vertical and horizontal flipping and scaling and jittering of the bounding boxes. Both the CNN (Inception v.2) and the RPN are trained simultaneously starting from the MS COCO checkpoint with decreasing learning rate, as shown in Fig. 6 (top) with a batch size of 1, given that the GPU memory limitation. The loss stabilizes after 30,000-40,000 trainings that require 3-4 hours on an entry-level NVIDIA Quadro P400 GPU, as shown in Fig 6 (bottom). The oscillations shown in the training losses are likely due to the very small batch size and can be reduced with larger batch sizes, if more GPU memory is available.

IV. Results

The use of active learning is extremely effective in reducing the annotation effort, as indicated in Table I, which shows a steady decrease of the annotation time in later learning cycles. After a few cycles of learning typically more than two thirds of the images only require approval. The remainder mostly need adjustment of the bounding boxes or correction of the class attribute. However, it should be noted that the time needed to approve an image is roughly fixed. Therefore, if most images require only approval, the annotation time for additional active learning sets reaches a plateau as the network performance improves.

Some defects are quite hard also for human experts to classify as they fall between categories, suggesting that the defect taxonomy should be refined.

A. Precision and Recall

Since the AOI tool generates false positives only 13% of the time, accuracy (defined as the percentage of correct predictions) is a poor metric to evaluate a classifier. In fact, a classifier marking every single image as containing defects would be by definition 87% accurate. In such situations, precision (the ratio of true positives over predicted positive) and recall (the ratio of true positives over real positives) are much more meaningful metrics. The results, shown in Table I, demonstrate that the proposed approach achieves a strong classification performance with a very small training set. The tradeoff between precision and recall can be adjusted by finetuning the detection threshold, to balance the competing business objectives (i.e. fewer defective chips undetected versus fewer working chips incorrectly discarded).

Obviously, part of the success can be ascribed to the use of transfer learning from the MS COCO database. Even if the classes and images from the MS COCO database are very different from Silicon chips and defects, many low-level filters for color, line and edge detection can be transferred to the new task. Furthermore, the background of chip images has a limited degree of variability, simplifying the detection task.

On the other hand, some defects and some features are very small or very rare, increasing the level of complexity of the problem.

B. Automatic defect taxonomy generation

The initial defect classification was proposed empirically after observing only a few hundred images. However, the trained network can also be used to for an additional goal: extracting a better taxonomy of defects. The last layer of the CNN shown in Fig. 3 condenses the image in a vector of 16,348 floating point elements, referred to as feature maps. The classification of defects learned by the network can be visualized by obtaining the feature maps of the cropped defects and clustering them according to a similarity metric. To reduce the computational burden, the dimensionality of the feature maps can be reduced to 200 with Principal Component Analysis (PCA), while maintaining 91% of the variance (i.e. the information content), as shown in Fig. 7. Finally, the 200 element vectors corresponding to each defect can be mapped to 2D space through t-Distributed Stochastic Neighbor Embedding (t-SNE). The 200-dimension feature maps associated to the defects are mapped into a 2-dimentional space (for ease of visualization) while preserving the relative entropy between similar objects. The resulting unsupervised clustering of defects, shown in Fig.

8, suggests additional classes with clearly distinct morphologies. Feature maps clustering can be used to refine the example labels without costly manual re-annotation.

C. Conclusions and Recommendations

This work demonstrates competitive performance in detection of semiconductor defects with reduced image annotation effort by combining for the first time an active learning approach with modern object detection deep neural networks.

Careful and consistent labelling is necessary to provide high quality training data necessary to improve the performance of the detector. Therefore, the proposed strategy used to reduce the cost of labelling is a key finding demonstrated by this work.

The trained neural network generalizes well to images of different semiconductor products if the morphology resulting from design and processing is similar.

Interestingly, the network has also shown the capability to generalize well to new kinds of defects not present in the training set, especially for large size defects. Conversely, novel morphological features such as a new shape of alignment marks not present in the training set tend to be misclassified as defects.

Further developments focus on further reducing image annotation time and increasing training set size for rare defects.

REFERENCES

[1] N. G. Shankar, Z. W. Zhong, "Defect detection on semiconductor wafer surfaces", Microelect. Eng., vol. 77, no. 3, pp. 337-346, 2005.

[2] F. L. Chen, S. F. Liu, "A neural-network approach to recognize defect spatial pattern in semiconductor fabrication", IEEE Trans. Semiconduct. Manuf., vol. 13, no. 3, pp. 366-373, Aug. 2000.

[3] A. Freeman, M. McIntyre, M. Retersdorf, C. Wooten, X. Song, A. Hesse, "The application and use of an automated spatial pattern recognition (SPR) system in the identification and solving of yield issues in semiconductor manufacturing", Proc. IEEE/SEMI Adv. Semiconduct. Manuf. Conf., pp. 302-305, 2007-Jun..

[4] A. F. Said and N. S. Patel, "Die level defects detection in semiconductor units," ASMC 2013 SEMI Advanced Semiconductor Manufacturing Conference, Saratoga Springs, NY, 2013, pp. 130-133.

[5] H. W. Hsieh, F. L. Chen, "Recognition of defect spatial patterns in semiconductor fabrication", lnt. J. Production Res., vol. 42, no. 19, pp. 4153-4172, 2004

[6] Y. Yuan-Fu, "A Deep Learning Model for Identification of Defect Patterns in Semiconductor Wafer Map," 2019 30th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC), Saratoga Springs, NY, USA, 2019, pp. 1-6.

[7] K. Imoto, T. Nakai, T. Ike, K. Haruki and Y. Sato, "A CNN-Based Transfer Learning Method for Defect Classification in Semiconductor Manufacturing," in IEEE Transactions on Semiconductor Manufacturing, vol. 32, no. 4, pp. 455-459, Nov. 2019.

[8] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017.

[9] Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., et al. "Speed/accuracy trade-offs for modern convolutional object detectors", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7310-7311.

[10] Lin, T. Y., Maire, M., Belongie, S., et al., "Microsoft COCO: Common Objects in COntext", Proceedings of the European Conference on Computer Vision, 2014, pp. 740-755