

A COMPLETE ARTIFICIAL INTELLIGENCE (AI) SOLUTION IN SEMICONDUCTOR MANUFACTURING INDUSTRY

1. TECHNICAL FIELD OF THE INVENTION

The present invention provides a machine learning operation system for
5 performing vision inspection on semiconductor device, comprising of a plurality of
semiconductor vision inspection equipment configured to communicate with a
centralized workstation through at least one network switch. The present invention
also provides a method for performing vision inspection on semiconductor device,
using said system, whereby said method implements continuous machine learning
10 model training using collected datasets and continuous deployment of trained
machine learning model to said semiconductor vision inspection equipment.

2. BACKGROUND OF THE INVENTION

The conventional procedure for implementing artificial intelligence (AI) in
vision inspection of the prior art follows a series of structured phases, namely manual
15 image collection, offline model training, and subsequent manual deployment of the
trained model to the vision inspection machine. In the initial phase, the collection of
images depicting defect samples slated for inspection involves manual efforts on the
production line. Upon accumulating an ample dataset of images, the AI model
undergoes offline training. This stage entails data preparation, algorithm selection,
20 and meticulous model fine-tuning to align its performance with predetermined
objectives. The intricacies of this process often require specialized expertise in

machine learning techniques. After the successful model training, the AI model is manually deployed into the production environment. This deployment procedure entails physically relocating the model to the production system, configuring the requisite software components, and harmonizing its integration within the existing inspection workflow.

While this conventional approach endeavors to establish AI-powered inspection capabilities in the market, it is not without inherent limitations that can compromise operational efficiency, accuracy, and the model's capacity to adapt. The following are the limitations. The first limitation is resource and time intensity, whereby the conventional process of deploying the AI model for vision inspection, with its distinct phases of image collection, offline model training, and manual deployment, is characterized by a notable demand for resources and time. This resource and time intensity can ultimately result in operational inefficiencies, impacting the overall effectiveness and productivity of the AI implementation process.

The second limitation is stagnation in learning, whereby the lack of continuous learning limits the model's ability to improve based on real-world feedback. The AI model is trained on a static dataset, often compiled from historical data. This snapshot of data is unable to capture the dynamic nature of the production environment. Consequently, once the model is deployed, it lacks the mechanism to incorporate new experiences or learn from the data it encounters during real-time operation.

Besides that, another limitation is the dependence on expertise, whereby effective model training requires the involvement of domain experts who possess a comprehensive understanding of the intricacies involved. They are required to make informed decisions regarding data cleaning, feature engineering, and algorithmic choices to ensure that the resulting model meets the performance expectations. This kind of specialized nature of model training expertise can lead to a bottleneck in the development cycle.

Hence, it would be advantageous to alleviate the shortcomings by having a machine learning operation system and method for performing vision inspection on semiconductor device using machine learning operation system, which is able to perform automated retrieval of real-time data, coupled with user-friendly user interface (UI) for model development and remote deployment.

3. SUMMARY OF THE INVENTION

Accordingly, it is the primary aim of the present invention to provide a machine learning operation system and method for performing vision inspection on semiconductor device using machine learning operation system which enables automated retrieval of real-time data, coupled with user-friendly user interface for model development and remote deployment which creates a dynamic ecosystem that promotes accuracy, adaptability and operational efficiency.

It is yet another objective of the present invention to provide a machine learning operation system and method for performing inspection on semiconductor device

using machine learning operation system whereby the automated retrieval of production images through local network connections fosters real-time and accurate data acquisition, eliminating the need for manual interventions. This ensures a continuous flow of up-to-date information and enabling the continuous learning of AI
5 model to enhance the inspection accuracy.

It is yet another objective of the present invention to provide a machine learning operation system and method for performing inspection on semiconductor device using machine learning operation system whereby the incorporation of a user-friendly UI for guided dataset labeling and model optimization democratizes AI model
10 creation and accelerates model development through streamlined iterations.

It is yet another objective of the present invention to provide a machine learning operation system and method for performing inspection on semiconductor device using machine learning operation system whereby the remote deployment of machine learning model expedites machine learning model integration into the production
15 line, eliminating manual transfer processes and minimizing downtimes.

Additional objects of the invention will become apparent with an understanding of the following detailed description of the invention or upon employment of the invention in actual practice.

According to the preferred embodiment of the present invention the following
20 is provided:

A machine learning operation system for performing vision inspection on semiconductor device, comprising of:

a plurality of semiconductor vision inspection equipment;

characterized in that

5 said system further comprises of a centralized workstation configured to communicate with said semiconductor vision inspection equipment through at least one network switch;

 wherein said centralized workstation is configured to continuously generate a plurality of training dataset of image and image data sent
10 from said plurality of semiconductor vision inspection equipment, train at least one machine learning model using said training dataset, validate said trained machine learning model and deploy the trained machine learning model to its intended semiconductor vision inspection equipment to ensure the latest trained machine learning model being
15 deployed is able to detect acceptable defects at said semiconductor device at a predetermine tolerance level.

In another embodiment of the invention there is provided:

A method for performing vision inspection on semiconductor device, comprising the steps of:

- 5 (i) capturing at least one image of at least one semiconductor device by at least one image capturing device in at least one semiconductor vision inspection equipment, before obtaining at least one image data from said image and storing said image and said image data in at least one data storage device in said semiconductor vision inspection equipment;
- (ii) transferring said image and image data from said semiconductor vision inspection equipment to a centralized workstation through at least one network switch;
- 10 (iii) generating a plurality of training dataset;
- (iv) training at least one machine learning model using said training dataset to develop at least one trained machine learning model;
- (v) evaluating and validating said trained machine learning model;
- 15 (vi) deploying at least one trained machine learning model by said centralized workstation to its intended semiconductor vision inspection equipment through said network switch;

wherein steps (i) to (vi) are repeated continuously to ensure the latest trained machine learning model is able to detect acceptable defects at a predetermined tolerance level.

4. BRIEF DESCRIPTION OF THE DRAWINGS

Other aspect of the present invention and their advantages will be discerned after studying the Detailed Description in conjunction with the accompanying drawings in which:

5 Fig. 1 is a block diagram of the machine learning operation system of the present invention.

Fig. 2 is a flowchart of the method of the present invention.

5. DETAILED DESCRIPTION OF THE DRAWINGS

10

In the following detailed description, numerous specific details are set forth in order to provide a thorough understanding of the invention. However, it will be understood by the person having ordinary skill in the art that the invention may be practised without these specific details. In other instances, well known methods, 15 procedures and/or components have not been described in detail so as not to obscure the invention.

The invention will be more clearly understood from the following description of the embodiments thereof, given by way of example only with reference to the accompanying drawings, which are not drawn to scale.

20 The invention is a machine learning operation system 101 for performing vision inspection on semiconductor device, comprising of a plurality of semiconductor

vision inspection equipment 105 and a centralized workstation 107; wherein said centralized workstation 107 is configured to communicate with said semiconductor vision inspection equipment 105 through at least one network switch 109, as shown in Fig. 1. The centralized workstation 107 is configured to continuously generate a plurality of training dataset of image and image data sent from said plurality of semiconductor vision inspection equipment 105, train at least one machine learning model using said training dataset, validate said trained machine learning model and deploy the trained machine learning model to its intended semiconductor vision inspection equipment 105 to ensure the latest trained machine learning model being deployed is able to detect acceptable defects at said semiconductor device at a predetermine tolerance level.

The centralized workstation 107 is connected to said network switch 109 through at least one local area networks (LAN) cable and said semiconductor vision inspection equipment 105 is connected to said network switch 109 through at least one fibre optic cable. Fibre optic cable is generally used to enlarge the data bandwidth.

The present invention is also a method for performing vision inspection on semiconductor device as shown in Fig. 2, comprising the following steps. In step (i), at least one image of at least one semiconductor device is captured by at least one image capturing device in at least one semiconductor vision inspection equipment 105, before obtaining at least one image data from said image and storing said image and said image data in at least one data storage device in said semiconductor vision inspection equipment 105. The image data may comprise of information of said image

such as lot number, wafer number, die coordinate, presence of defect, type of defect or combination thereof.

Optionally, said method for performing vision inspection on semiconductor device of the present invention further comprises of the step of monitoring status of connection between a plurality of semiconductor vision inspection equipment 105 in a localised network before step (i). Localised network refers typically to the wired network connection between the plurality of semiconductor vision inspection equipment 105 and with a centralized workstation 107 through at least one network switch 109. Wireless connections are not preferred between said plurality of semiconductor vision inspection equipment 105 due to security purpose, whereby data transferred through wireless connection may be forcefully accessed by unauthorized personal. It is essential to check the connection status with said plurality of semiconductor vision inspection equipment 105 to ensure the ability to perform the following steps in the method of the present invention. At times, the internet protocol (IP) address of at least one of semiconductor vision inspection equipment may be changed by the user which prevents any meaningful data to be transferred between said centralised workstation 107 and said semiconductor vision inspection equipment 105. The monitoring can be done by a centralized workstation 107 by sending at least one status request signal to said semiconductor vision inspection equipment 105 and observing connection status of said semiconductor vision inspection equipment 105. The centralised workstation 107 is also configured to provide at least one alarm event if connection with said semiconductor vision inspection equipment 105 is broken.

The step of monitoring of status connection before step (i) can be done automatically or manually triggered. The monitoring of status connection can be done automatically by said centralized workstation 107 for a predetermined frequency or can be done manually by the user for a predetermined frequency.

5 The centralized workstation 107 is disposed at the location of said semiconductor vision inspection equipment 105 or away from the location of said semiconductor vision inspection equipment 105 using a web-based software application.

10 In step (ii), said image and image data is transferred from said semiconductor vision inspection equipment 105 to a centralized workstation 107 through at least one network switch 109. The transferring is done in batches or continuously transferred upon said semiconductor vision inspection equipment 105 obtaining said image and image data.

15 In step (iii), a plurality of training dataset is generated. said step (iii) comprises of the following sub-steps. In sub-step (iii)(a), at least one relevant image with its image data that is suitable to be used for training of at least one machine learning model is selected and said relevant image data is managed by at least one data processing means in a centralized workstation 107. The following is an example of how said data processing means manage said relevant image data. Of the many
20 images being transferred from said semiconductor vision inspection equipment 105 to said centralised workstation 107, if majority of said images are with no defect, said data processing means in said centralized workstation 107 will create a workspace to

import underkill or overkill images to be used as training dataset. In sub-step (iii)(b), said image data is pre-processed. An example of said pre-processing of image data is augmentation to increase the amount of training data from existing images. In sub-step (iii)(c), at least one training dataset is prepared to ensure all the defect that is
5 required to detect is correctly labelled.

In step (iv), at least one machine learning model is trained using said training dataset to develop at least one trained machine learning model. Generally, any suitable training method can be used, such as the same model algorithm as the previously used machine learning model or any suitable general pre-trained machine
10 learning model.

In step (v), said trained machine learning model is evaluated and validated. Validation ensures the overall accuracy of all the testing data is within a predetermined acceptable level while evaluation is done automatically and set by the user based on the tolerance level of said user with reference to amount of false rejects
15 and not escapees.

In step (vi), at least one trained machine learning model is deployed by said centralized workstation 107 to its intended semiconductor vision inspection equipment 105 through said network switch 109. Step (vi) further comprises of monitoring said trained machine learning to ensure the latest version of trained
20 machine learning is being deployed to its intended semiconductor vision inspection equipment 105. Each of the plurality of semiconductor vision inspection equipment 105 in said localised network only works with a specific type of machine learning

model because each semiconductor vision inspection equipment 105 may be inspecting different types of semiconductor device. In other words, each semiconductor vision inspection equipment is only suitable to be deployed with a particular trained machine learning model.

5 Steps (i) to (vi) are repeated continuously to ensure the latest trained machine learning model is able to detect acceptable defects at a predetermined tolerance level. Steps (i) to (vi) are done using one user interface at said centralized workstation 107.

 While the present invention has been shown and described herein in what are considered to be the preferred embodiments thereof, illustrating the results and
10 advantages over the prior art obtained through the present invention, the invention is not limited to those specific embodiments. Thus, the forms of the invention shown and described herein are to be taken as illustrative only and other embodiments may be selected without departing from the scope of the present invention, as set forth in the claims appended hereto.

WHAT IS CLAIMED IS:

1. A machine learning operation system (101) for performing vision inspection on semiconductor device, comprising of:

a plurality of semiconductor vision inspection equipment (105);

5 characterized in that

said system (101) further comprises of a centralized workstation (107) configured to communicate with said semiconductor vision inspection equipment (105) through at least one network switch (109);

10 wherein said centralized workstation (107) is configured to continuously generate a plurality of training dataset of image and image data sent from said plurality of semiconductor vision inspection equipment (105), train at least one machine learning model using said training dataset, validate said trained machine learning model and deploy the trained machine learning model to its intended semiconductor vision inspection
15 equipment (105) to ensure the latest trained machine learning model being deployed is able to detect acceptable defects at said semiconductor device at a predetermine tolerance level.

2. The machine learning operation system (101) for performing vision inspection on semiconductor device as claimed in Claim 1, wherein said
20 centralized workstation (107) is connected to said network switch (109)

through at least one local area networks (LAN) cable and said semiconductor vision inspection equipment (105) is connected to said network switch (109) through at least one fibre optic cable.

3. A method for performing vision inspection on semiconductor device
5 using machine learning operation system, comprising the steps of:

(i) capturing at least one image of at least one semiconductor device
by at least one image capturing device in at least one
semiconductor vision inspection equipment (105), before
obtaining at least one image data from said image and storing
10 said image and said image data in at least one data storage device
in said semiconductor vision inspection equipment (105);

(ii) transferring said image and image data from said semiconductor
vision inspection equipment (105) to a centralized workstation
(107) through at least one network switch (109);

15 (iii) generating a plurality of training dataset;

(iv) training at least one machine learning model using said training
dataset to develop at least one trained machine learning model;

(v) evaluating and validating said trained machine learning model;

(vi) deploying at least one trained machine learning model by said centralized workstation (107) to its intended semiconductor vision inspection equipment (105) through said network switch (109);

5 wherein steps (i) to (vi) are repeated continuously to ensure the latest trained machine learning model is able to detect acceptable defects at a predetermined tolerance level.

4. The method for performing vision inspection on semiconductor device as claimed in Claim 3, further comprising the step of monitoring status
10 of connection between a plurality of semiconductor vision inspection equipment (105) in a localised network before step (i).

5. The method for performing vision inspection on semiconductor device as claimed in Claim 4, wherein said monitoring before step (i) is done by a centralized workstation (107) by sending at least one status request
15 signal to said semiconductor vision inspection equipment (105) and observing connection status of said semiconductor vision inspection equipment (105).

6. The method for performing vision inspection on semiconductor device as claimed in Claim 4 or 5, wherein said centralised workstation (107)
20 provides at least one alarm event if connection with said semiconductor vision inspection equipment (105) is broken.

7. The method for performing vision inspection on semiconductor device as claimed in Claim 4, wherein said monitoring of status connection before step (i) is done automatically or manually triggered.
- 5 8. The method for performing vision inspection on semiconductor device as claimed in Claim 4, wherein said centralized workstation (107) is disposed at the location of said semiconductor vision inspection equipment (105) or away from the location of said semiconductor vision inspection equipment (105) using a web-based software application.
- 10 9. The method for performing vision inspection on semiconductor device as claimed in Claim 3, wherein said image data comprises of information of said image such as lot number, wafer number, die coordinate, presence of defect, type of defect or combination thereof.
- 15 10. The method for performing vision inspection on semiconductor device as claimed in Claim 3, wherein said transferring in step (ii) is done in batches or continuously transferred upon said semiconductor vision inspection equipment (105) obtaining said image and image data.
11. The method for performing vision inspection on semiconductor device as claimed in Claim 3, wherein said step (iii) comprises of the following sub-steps:

(a) selecting at least one relevant image with its image data that is suitable to be used for training of at least one machine learning model and managing said relevant image data by at least one data processing means in a centralized workstation (107);

5 (b) pre-processing of said image data;

(c) preparing at least one training dataset.

12. The method for performing vision inspection on semiconductor device as claimed in Claim 11, wherein pre-processing of image data in step (iii)(b) is augmentation to increase the amount of training data from
10 existing images.

13. The method for performing vision inspection on semiconductor device as claimed in Claim 3, wherein step (vi) further comprises of monitoring said trained machine learning to ensure the latest version of trained machine learning is being deployed to its intended semiconductor vision
15 inspection equipment (105).

14. The method for performing vision inspection on semiconductor device as claimed in Claim 3, wherein said steps are done using one user interface at said centralized workstation (107).