

# Mask defect detection with hybrid deep learning network

Peter Evanschitzky,<sup>a,\*</sup> Nicole Auth,<sup>b</sup> Tilmann Heil<sup>©,b</sup>,  
Christian Felix Hermanns,<sup>b</sup> and Andreas Erdmann<sup>©,a</sup>

<sup>a</sup>Fraunhofer Institute for Integrated Systems and Device Technology IISB,  
Erlangen, Germany

<sup>b</sup>Carl Zeiss SMT GmbH, Rossdorf, Germany

## Abstract

**Background:** Deep learning is a very fast-growing field in the area of artificial intelligence with remarkable results in recent years. Many works in the lithography and photomask field have shown progress of technology in this application area and the potential to improve lithography by the automation of processes. Despite this progress, the use of machine learning techniques in the field of mask repair still seems to be at the beginning.

**Aim:** We show that deep learning-based methods can successfully be applied to mask repair applications.

**Approach:** The presented system is a hybrid and modular approach based on a combination of several deep learning networks and analytical methods, enabling the detection of mask pattern defects and the determination of the exact defect shapes from SEM images. In the current version, the system is trained for line/space patterns and contact patterns with typical defect types. The modularity allows for the extensibility to new use cases. The issue of an insufficient amount of training data is addressed using purely computer-generated simplified SEM data in combination with a specific network architecture.

**Results:** The very good functionality and defect detection accuracy of the system are demonstrated with a set of real SEM images with line/space patterns and contact patterns with numerous defects. In particular, a 100% true defect detection rate could be obtained.

**Conclusions:** The presented machine learning approach demonstrates the successful defect identification, location, and shape determination from real mask SEM images.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMM.20.4.041205](https://doi.org/10.1117/1.JMM.20.4.041205)]

**Keywords:** deep learning; mask repair; SEM image; mask defect.

Paper 21049SS received May 28, 2021; accepted for publication Sep. 23, 2021; published online Oct. 5, 2021.

## 1 Introduction

The availability of continuously increasing computing capabilities enabled by more powerful chips continues to improve our lives in many aspects. Tremendous advances in chip manufacturing technology underlie this development. As early as 1965, Intel cofounder Gordon Moore observed that the number of transistors in an integrated circuit on a chip doubles every 2 years.<sup>1</sup> This exponential growth has now continued for more than 50 years. The key technology for realizing such continuously shrinking feature sizes on chip is optical nanolithography.<sup>2</sup> Its basic process steps are described as follows. The electronics design of the chip to be produced is broken down into individual layers. The information of each layer is then physically transferred to a pattern on a so-called photo mask. The mask is then optically imaged onto a silicon wafer that has previously been coated with light-sensitive resist. In this way, the mask pattern

---

\*Address all correspondence to Peter Evanschitzky, [peter.evanschitzky@iisb.fraunhofer.de](mailto:peter.evanschitzky@iisb.fraunhofer.de)

is selectively transferred into a pattern of developed and nondeveloped resist that can be etched, thus allowing the wafer to be structured. Repeating this process many times for all layers of the electronics layout, a three-dimensional (3D) structure builds up creating the integrated circuit. Thus, optical nanolithography enables a contact-free multiplication of the mask patterns containing the chip layer information with the highest possible productivity. Obviously, all masks used in this process must be free from defects to avoid the patterning of unintended structures, which may scrap the entire chip. Therefore, the manufacturing of perfect, defect-free masks is a prerequisite for the successful implementation of optical nanolithography and, hence, the continuation of Moore's law.

Due to the high complexity of the mask making process, a first-time right production of defect-free masks is, at least economically, impossible. This makes mask repair a key step in the mask making value chain. For today's high-end chips, gas-assisted focused electron-beam-induced etching and material deposition is the mask repair process of record. In a MeRiT<sup>®</sup> mask repair tool, qualification and repair preparation of a defect site are done using scanning electron microscope (SEM) images.<sup>3,4</sup> Thus, highly accurate defect detection, classification, and shape determination from SEM images are essential. In this paper, we present an approach for this task based on deep learning methods.

Deep learning is one of most exciting and fast-growing fields in the area of artificial intelligence with remarkable results in recent years.<sup>5,6</sup> However, in the lithography and photomask industry, the technology is still in an early phase, and the introduction into production is still at the beginning. In recent years, many works have shown progress in this field and the potential of the technology to improve lithography by the automation of processes. In several works, the defect classification from wafer and mask data with deep learning methods has been demonstrated.<sup>7-13</sup> Such methods also have been used to perform pattern matching, contour extraction, and 3D profile reconstruction from SEM images.<sup>14-16</sup> A general property of SEM images, the noise, has been addressed in further works, with goal of reducing the noise to obtain higher accuracy in the image analysis.<sup>17-19</sup> Other application fields of machine learning methods operating on wafer and layout data are hot spot detection, layout classification, and pattern similarity detection.<sup>20-22</sup> The methods are also used for tasks that are not targeted at the processing and use of image data. For example, the computation of OPC and SRAFs,<sup>23-25</sup> the efficient computation of 3D mask near fields,<sup>26</sup> and the surface defect detection from scattering data<sup>27</sup> have been demonstrated in corresponding works.

When deep learning methods are applied in lithography, one general problem occurs consistently. Deep learning systems need a large amount of training data, which are typically not available. This opens the field for another kind of deep learning application: the creation of digital twins for the efficient generation of such data.<sup>28,29</sup> One example is the generation of SEM images with defects from SEM images without defects.<sup>30</sup> In this work, the digital twin approach is used to generate simplified SEM images with defects from mask pattern data. In contrast to the previous example, the approach is not based on deep learning but on simplified physical considerations. The details are explained in Sec. 3.2. Another problem in this field is the availability of large amounts of unlabeled data in combination with only a small number of labeled data. Labeled data are, for example, SEM images with additional information about the existence of defects and the defect locations and types. In such a situation, unsupervised pretraining with an autoencoder is a general method for initializing the network variables.<sup>31</sup> This leads to better accuracy of the actual training. Despite all of the progress in using machine learning methods in lithography, their application in the field of mask repair seems to be at the beginning.

This paper shows the application of deep learning in this field. Defective mask patterns are detected from SEM images, and the exact defect shapes are determined. Such shapes are required for the defect assessment and as a starting point for repair. The problem of training big data is solved using purely artificial data generated with a simplified SEM data approach. The problem of an exact defect shape determination is solved by a hybrid approach combining the deep learning network with analytical image processing. The network is characterized by a modular design, which enables the extension to new use cases that have not been trained so far. The network is able to process larger SEM images with multiple patterns, for example, with a size of  $4000 \times 4000$  pixels and  $6 \mu\text{m} \times 6 \mu\text{m}$ , respectively, and to localize multiple defects and their shapes on such images.

The basic idea of the network and of deep learning in general is to use multiple neural network layers that extract increasingly complex features of the input data from layer to layer to carry out a transformation of the data. In image processing, which is the basis of the networks in this paper, the lower layers (closer to the input) extract, for example, more simple structures such as basic shapes in the images. The following higher layers combine these structures into more complex shapes such as specific combinations and arrangements of basic structures relevant to solving the specific task of the network. During the training, the network learns to recognize the best structures and their locations inside the network. However, this only takes place within a given network structure that has a strong influence on the success of the training. A summary of deep learning with many further readings is given in Ref. 32. In the different networks of this work, such structures are used to perform classifications of the input images and to recompose new images showing special characteristics of the input images. Specifically, the pattern types and pitches in mask SEM images are classified, and defective areas of the patterns in the images are detected. Depending on the availability of labeled and unlabeled training data, deep learning networks can be trained by different learning strategies such as supervised learning, semisupervised learning, active learning, and unsupervised learning. The first technique requires training data that are all labeled, whereas the last option uses only unlabeled data. The two techniques in between make use of partly labeled data. In the case of semisupervised learning, a smaller amount of labeled data is combined with a larger amount of unlabeled data. In active learning, the algorithm actively queries for labels with the result that the number of labels can be much lower than the number required in supervised learning. In this work, artificially generated training data were used, which allows for the computation of the labels for all data. Therefore, purely supervised trainings were performed.

The paper is structured as follows. Section 2 gives an introduction to mask repair as a specific field of application of the developed deep learning system. In Sec. 3, the deep learning networks for defect detection, the generation of artificial training data, and the combination of the networks with an analytical method for defect shape determination are explained. Section 4 presents the application of the networks to real SEM images and demonstrates the very good accuracy. The planned next steps after defect detection, which are the defect and repair shape assessment based on lithography simulations, are sketched in Sec. 5. The paper ends with a conclusion in Sec. 6.

## 2 Mask Repair

The manufacturing of lithographic masks is a demanding high-tech task. Because of its complexity and economic reasons, the fabrication of such masks cannot be perfect, leaving a certain number of defects. Thus, mask repair is practically a prerequisite for profitable mask making. Today, the process of record for the repair of the most advanced mask features is gas-assisted focused electron-beam-induced etching and material deposition. In this process, well-selected precursor molecules are directed to and adsorbed on the mask surface to be repaired. Using a focused electron beam, these molecules are very locally activated such that they induce an etching or deposition process to either remove unwanted mask material or to add missing material, respectively.<sup>33</sup> The precursor chemistry and process parameters have to be carefully selected such that the process residuals are volatile and can be pumped away to yield a clean, repaired site on mask.

Recently, optical nanolithography using EUV light has been introduced into high-volume chip manufacturing. Due to its small wavelength of only 13.5 nm, EUV lithography is enabling further shrinking feature sizes on wafers and hence a continuation of Moore's law throughout this decade.<sup>34</sup> Accordingly, the minimum feature size on mask will shrink as well, and new mask materials optimized for EUV light are expected to be introduced. From the mask repair point of view, this leads to three major challenges. First, the decreasing feature size on mask calls for a better repair resolution, i.e., a decreasing minimum repair size. Second, also driven by the feature shrink on the wafer, the edge placement accuracy of the repaired site needs to be improved with every new node. Third, new mask materials require the development of specifically tailored precursor chemistry and repair processes. Figure 1 shows an impression of the next-generation



**Fig. 1** The next-generation mask repair tool MeRiT<sup>®</sup> LE.

MeRiT LE<sup>®</sup> mask repair tool that is currently introduced into the market to master these challenges for the upcoming EUV nodes.

The MeRiT LE<sup>®</sup> is equipped with a new high-resolution electron-beam column operating at very low landing energy of only 400 eV and features a highly stabilized tool platform that minimizes tool jitter. Thus, by combining high-resolution, low landing energy, and high stability, the MeRiT LE<sup>®</sup> significantly improves minimum repair size and edge placement performance as compared with its predecessor tools.<sup>35</sup> Furthermore, the MeRiT LE<sup>®</sup> offers a broad range of precursor chemistries, enabling the repair of the upcoming EUV materials.

To achieve excellent repair results, the electron beam needs to scan precisely determined defect shapes following sophisticated scanning strategies. Thus, excellent defect detection and defect shape determination performance are essential to take full advantage of the improved tool capability. The noise level of the SEM images from the MeRiT<sup>®</sup> LE, taken as input for the defect detection and shape determination, is already at a low level. To even further suppress the contribution of image noise to the error of the defect shape detection and generation, deep learning-based tools have proven to be quite successful.<sup>17-19</sup> Therefore, the question of whether deep learning-based methods can successfully be applied to mask repair applications is of high relevance.

### 3 Hybrid Deep Learning Network for Defect Detection

This section presents an AI-system based on deep learning networks detecting defects on mask patterns in SEM images. Furthermore, the system is able to determine the defect shapes and makes recommendations for the repair of the detected defects. Due to the basic nature of a deep learning network, a large amount of training data is required. With such training data, the network is forced to learn a relation between input and output data with the goal of enabling the network to produce correct output data by itself after the training phase. In the case of defect detection from SEM images, the input data are SEM images with potentially defective patterns, and the output data are images masking the defective areas. It is important to note that the training data must be composed in such a way that all cases of interest will be trained. This means that all relevant pattern types and sizes as well as defect types, sizes, and positions must be learned by the network. Furthermore, the relevant image variations (e.g., brightness, contrast, and gray value distribution) and possible image sections resulting from different image recording conditions must be learned as well. By providing sufficient training data set coverage and

well-balanced variation within that coverage, the network should finally be able to detect defects within the specified ranges. One practical issue is the fact that it is nearly impossible to generate a sufficient number of real SEM images with programmed defects. A proper training requires typically several thousand images. Therefore, the generation and usage of artificial training data is an important point in this work. Section 3.1 describes the properties of real SEM images, mask patterns, and defects used in this work and to be learned by the network. In Sec. 3.2, a special kind of training data with a limited variation range is presented. To enable the network to recognize all cases of interest in real SEM images, the training data do not have to include all real image properties and variations. The limitation to some special properties and reduced variation ranges is sufficient for successful network training. Furthermore, the network can be trained exclusively with such artificial data. The overall task of defect detection and defect shape determination is split into subtasks. Individual subnetworks for the respective subtasks are combined in a single AI-system. The details are explained in Sec. 3.3. Another weak point of the used deep learning system is the computation of accurate data in terms of exact numbers. In the considered case, these are exact defect shapes. For this specific purpose, an analytical method supporting the AI-system was developed and combined with the network. This combination is described in Sec. 3.4.

The presented system is implemented in Python. For the implementation of the network parts of the system, the deep learning interface of the Fraunhofer IISB lithography simulator Dr. LiTHO is used. This interface is adapted to the specific needs of this work and allows for the implementation of different kinds of deep learning networks. The interface is also based on Python and uses TensorFlow. All trainings are performed on GPUs.

### 3.1 Real SEM Images, Mask Patterns, and Defects

In a MeRiT<sup>®</sup> mask repair system, SEM images are generated by scanning a focused electron beam across a certain area of the structured mask surface. Most common masks have an absorber layer on a quartz blank or on a MoSi multilayer, which is designed to reflect EUV light. In both cases, the absorber material is very smooth, usually amorphous, and structured down to the quartz or multilayer, respectively, with a sidewall angle close to 90 deg. In the EUV case, which is the main application field of the presented network in this paper, absorber material has a typical thickness in the range of 60 nm for standard TaBN absorbers to 30 nm for more advanced materials. As a result, these masks have only two distinct height levels.

The image signal is recorded by detecting the backscattered (BSE) and secondary electrons (SE) emitted from the sample surface. The SE and BSE yields vary depending on sample material and incident energy of the electrons, resulting in a material contrast between the absorber and the mask blank material. Due to the 3D absorber profile, more electrons are emitted in the vicinity of the absorber edge.<sup>36</sup> Therefore, the SEM images show a characteristic edge brightness enhancement separating the absorber from the blank regions.

To minimize possible electron-induced mask degeneration processes and to reduce charging artifacts, a low beam current of several 10 pA and scan speeds in the  $\mu$ s-range are used. Therefore, the resulting images exhibit a high noise level, which is typical for SEM images and has been widely discussed in the literature.<sup>37,38</sup>

In this work, we focus on a subset of mask patterns and defect shapes according to their high relevance for the mask industry. Images of repetitive lines and spaces patterns as well as contact hole arrays with different half pitches ranging from 70 to 250 nm (mask scale) are recorded and analyzed. These patterns also contain a variety of programmed defects, such as extrusions, mousebites, pinholes, bridges, broken lines, oversize, undersize, and missing holes with different dimensions and locations. Furthermore, most of the images include multiple defects. Such features are typically used for evaluating the repair performance and therefore are good candidates for evaluating the deep learning network.

### 3.2 Artificial Training Data

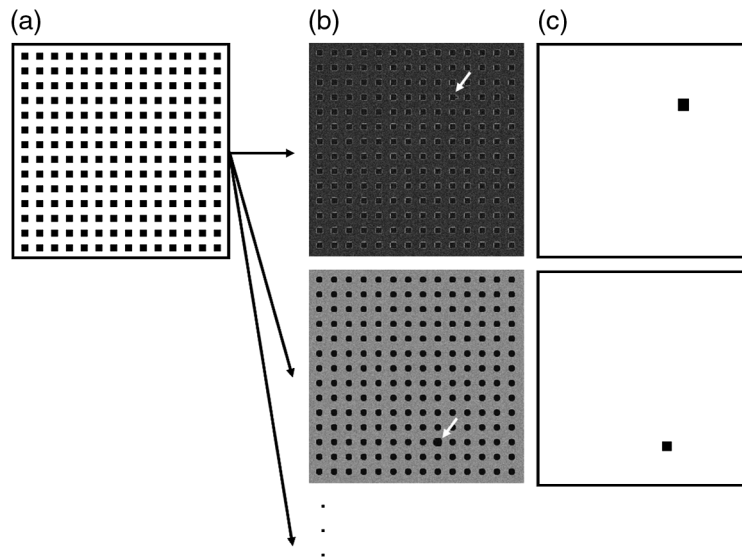
Deep learning networks typically need a tremendous amount of training data. In the case of defect detection from SEM images, this would require thousands of SEM images. The defective

areas must be marked in each image, and for the specific work in this paper, the pattern types and pitches must be labeled. All of this information is required for the different network trainings and would have to be done by hand. It is almost impossible to perform such a task in practice. Therefore, the concept of artificial training data was investigated and applied. Artificial training data can be generated with digital twins of real systems.<sup>39</sup> There are different ways to obtain such data. One approach is the physical simulation of the data. Depending on the complexity of the underlying physics, this can be very challenging and the simulation times can be long, even too long for generating a sufficiently large amount of data. In consequence of the continuous development of AI-techniques and specifically deep learning-techniques, another powerful approach is increasingly applied: the representation of digital twins with generative adversarial networks.<sup>28</sup> In the case of SEM images, the idea of this approach is to generate output data in the style of an SEM image from arbitrary input data. To do so, the network has to learn how SEM images look. Once the network is trained, it can be used to generate realistic looking SEM images from mask layouts. But this approach needs, of course, the development and training of a corresponding network in advance. In this work, the digital twin approach is used to generate simplified SEM images with defects from mask pattern data. The approach is based on simplified physical considerations and does not use AI methods. The idea of such simplified data is not to compute realistic or realistic looking SEM images. The data encode only properties that are relevant for the defect detection and neglect other properties of real SEM images. Specifically, the training data do not have to include all real SEM image properties and variations to enable the network to learn all cases of interest and to process such real images correctly. Furthermore, the training can be carried out exclusively with the artificial data. The required properties and data variation ranges are described in the following list below. Tests have shown that the concept works well if a specific network architecture is used. This architecture forces the network to learn only the defect-relevant properties. Consequently, other real SEM image properties are less important and do not have to be represented in a realistic manner in the training data. The corresponding architecture is explained in Sec. 3.3.1.

As a first result of the work presented in this paper, the following rules for the simplified SEM image data have been derived:

- The data have to include the desired mask pattern types with all pitch/duty cases.
- The data have to include all defect types.
- Randomly selected image gray levels for absorber, blank, and edge regions (see Sec. 3.1) within certain gray level ranges are sufficient. The ranges do not have to cover the ranges of the real images.
- Randomly selected defect sizes and defect positions within certain ranges are sufficient. The ranges do not have to cover the ranges of the real images.
- For feature edge and feature width roughness, it is sufficient to vary the entire edge of the pattern within the desired range. A realistic description of the roughness is not required.
- A slight random image noise of about 10%, which does not have to correspond to the real image noise (see Sec. 3.1), is used.
- A random feature corner rounding, which does not have to match the real image corner rounding, is used.
- The number of defects inside one image is not critical. Even a training with only one defect per image is sufficient and is therefore used in all training data. The resulting network is still able to detect multiple defects inside an image (this is tested with up to 10 defects per image).

The described rules are implemented in a software algorithm that generates simplified SEM images. This enables the computation of larger numbers of images within a short period of time. For example, the computation of 10,000 images takes about 3 h on a state-of-the-art personal computer (Intel Core i5, 3.8 GHz, 16 GB memory, Windows 10). In addition to the SEM images, the training data also must include the desired output images required for the defect detection and, furthermore, the pattern type labels and the pitch labels to be learned by the network. The output images mask the defect areas on the input images. Therefore, the output images are in general white with black pixels in the defect areas. The black areas form rectangular bounding



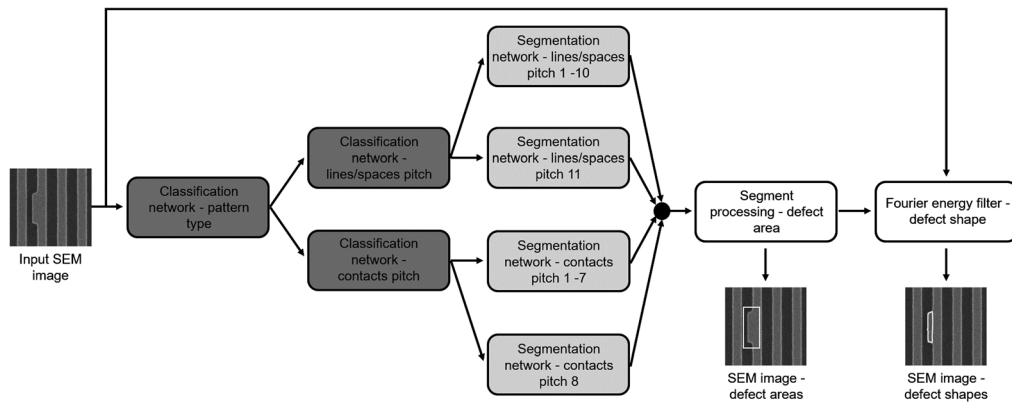
**Fig. 2** Example of artificial training data. (a) Basic mask patterns defining the feature type and pitch/duty. (b) Two examples of artificial input training data derived from the left picture according to the rules described in the text. The white arrows indicate the defective patterns. (c) Output training data corresponding to the middle pictures. The output images mask the defective features. For each basic mask pattern image, a set of multiple training data pairs is generated.

boxes around the defect areas in the case of line/space patterns and rectangular bounding boxes around the entire defective patterns in the case of contacts. This is indicated in the training data examples given in Fig. 2. Since all parameters of the generated SEM images are known, the corresponding output images and labels can be generated in parallel to the SEM images. One SEM image and the corresponding output image or one SEM image and the corresponding label form one training data pair. Training tests with the artificial data have shown another important point. Well balanced data sets must be used for the training of the networks, i.e., the training data have to include all feature types, pitches, and defect types with identical numbers and variations of parameters. The parameter variations are computed from a discrete uniform distribution. Figure 2 shows an example of artificial training data.

A basic mask pattern image defining the feature type and pitch/duty [Fig. 2(a)] is transformed into a set of training data pairs. Two examples from the set are shown in Figs. 2(b) and 2(c). Figure 2(b) shows the input data derived from the basic mask pattern image by applying the rules described above. The white arrows indicate the defective patterns with an extrusion defect in the upper image and an oversize defect in the lower image. Figure 2(c) shows the corresponding output images masking the defective structures on the input images. In this work, for each pattern type and pitch/duty combination, between 5400 and 6300 training data pairs are generated. This leads to an artificial SEM training data base with about 100,000 training data pairs covering the line/space patterns and contact patterns with all pitch/duty combinations according to Sec. 3.1. Due to the modularity of the developed AI-system, further use cases can be added without touching the already trained parts. Details are provided in the following sections.

### 3.3 General Network Structure

Many different network architectures for a wide range of applications have been proposed in the literature. Examples from lithography are given in Sec. 1. Basic tests and investigations with different deep learning network architectures have been performed to identify appropriate network architectures and to specify an appropriate overall structure of the AI-system. The result of the tests suggested splitting the tasks of defect detection and defect shape determination into subtasks, using subnetworks for the individual subtasks, and combining the subnetworks into one overall system. Furthermore, several analytical methods are integrated to support the networks. Concerning the individual subtasks and related network architectures, a pyramidal



**Fig. 3** Overall system architecture showing the coupling of classification networks (dark gray boxes), segmentation networks (light gray boxes), and analytical processing steps (white boxes). From the input SEM image, the defect areas are detected, and the defect shapes are determined.

encoder–decoder structure is used for the defect detection task. A modified U-net structure<sup>40</sup> with a relatively small number of layers and with relatively large filter sizes has been identified as the appropriate solution for this task. This network type outperformed all other investigated architectures. For the classification tasks, a more classically designed convolutional neural network from Ref. 41 with one specific extension is used. Figure 3 shows the implemented overall system structure.

The basic idea of the network structure is to separate data into classes with different properties and to process the classes with individual specialized networks. This means that the input SEM images are split, according to their pattern type and pitch, into subnetworks that are specialized to cover the respective cases. As specified in Sec. 3.1, this is currently implemented for line/space patterns and contact patterns for a larger range of pitches. The modular approach allows for adding further use cases without changing the already trained specialized networks. Consequently, in the first step, a basic classification of the pattern type and pitch or pitch range is done (dark gray boxes in Fig. 3). Then, depending on the detected use case, the image goes into the corresponding specialized segmentation network (light gray boxes in Fig. 3). This network detects the defective area of any kind of defect for any SEM image condition within the specifications, but only for the given use case. Further use cases can be trained and added to the network. This is realized either by a retraining of the corresponding already existing segmentation network or by training and adding a new segmentation network. The second option needs an additional network, but it speeds up the training time since the already existing segmentation networks do not have to be touched, and the training of the new network is limited to the new pitch/duty combinations. For instance, the number of training data of the segmentation network for lines/spaces, pitches 1 to 10 (first light gray box in Fig. 3) are 10 times larger compared with the segmentation network for lines/spaces, pitch 11 (second light gray box in Fig. 3) because of the required balanced training data according to Sec. 3.2. Consequently, the training time of one epoch of the first network is 10 times longer compared with the second network. The modularity is demonstrated for both the contact patterns and the line/space patterns for a single pitch, which is trained separately (light gray boxes in Fig. 3 with “pitch 11” for the lines/spaces and with “pitch 8” for the contacts). The other pitches are trained together in one network (light gray boxes in Fig. 3 with “pitches 1 to 10” for the lines/spaces and with “pitches 1 to 7” for the contacts). The single pitch training shows the extensibility of the approach, and the multiple pitch training demonstrates the universal usability since the network is also able to process images containing more than one pitch inside the image. At this point of the image processing flow, the use case and the defect areas as returned by the networks are known. To get the final defect areas, an analytical image processing step follows (left white box in Fig. 3). The result after this step is rectangular areas defining all detected defect areas on the original SEM image (“SEM image—defect areas” with the white rectangle in Fig. 3). The final part is the determination of the defect shape. The defect shape serves as a first repair structure that is used to assess the lithographic impact of the defect and of the repaired mask pattern. Lithography simulations

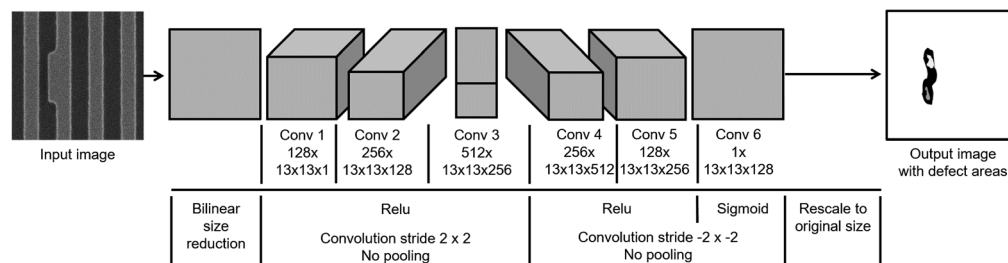


based on rigorous mask diffraction computations are used for such investigations. Therefore, the mask simulation requires the exact geometry of the investigated features, which means that the defect shapes returned by the networks must be as close as possible to the real defect shapes. Different networks have been trained and investigated to directly return the shape of a detected defect. However, it was not possible to obtain results with acceptable accuracy. Therefore, the implemented solution involves an analytical method that operates on the defective areas that are returned by the segmentation networks. As indicated in the right white box in Fig. 3, the method is based on a Fourier energy filtering. The input of the method is given by original SEM images and the defect areas as provided by the defect detection networks. The method returns the polygons of the detected defect shapes. Details of the Fourier energy filtering are explained in Sec. 3.4.

### 3.3.1 Segmentation network for defect detection

The most important part of the overall network structure is the segmentation networks (light gray boxes in Fig. 3) to identify the defective areas in the SEM image. The basic idea of this network type is to transform the input SEM image into an output image that is masking the defect areas. The output image consists of black pixels tagging a defect area on a white background for the defect-free area of the mask. Furthermore, the network is trained in such a way that it generates rectangular black areas around a defect (see Sec. 3.2). At this point of the image processing flow, the goal is to generate bounding boxes around the defects but not to reconstruct the defect shapes. Due to accuracy reasons, this is done in an analytical processing step, which is described in Sec. 3.4. The architecture of the segmentation networks (light gray boxes in Fig. 3) was driven by the decision to use exclusively artificial training data and, furthermore, by the relatively large size of the SEM images, and the corresponding mask areas to be processed in one step (e.g.,  $4000 \times 4000$  pixels and  $6 \mu\text{m} \times 6 \mu\text{m}$ , respectively). The goal of the training data is to encode all SEM image properties relevant for defect detection but not to provide SEM data, which are as realistic as possible. As described in Sec. 3.2, tests have demonstrated that such simplified artificial SEM data are sufficient for network training if a specific network architecture is used. A network performing an image-to-image transformation might learn direct pixel-to-pixel transformations to a certain extent. This could lead to misleading behavior since the pixel level information of the artificial training data does not describe real SEM images with sufficient accuracy. In contrast to that, a pyramidal encoder/decoder architecture, which is used here, reduces lateral image information during the encoding step and increases the image depth. This allows the network to find those basic structures that are relevant for defect detection and to neglect other nonrelevant SEM image properties. The following decoder structure recomposes the desired output image from the basic structures. Therefore, with an appropriate architecture, such simplified SEM data can be used to train a network that is supposed to operate on real SEM images. Figure 4 shows the details of the corresponding network setup.

In the first step, the input SEM images are resized with a bilinear interpolation to a uniform size of  $1024 \times 1024$  pixels (bilinear size reduction in Fig. 4). The original images are much larger, but to achieve acceptable training and analysis times, a reduction to the given size is done. An example of the training time for different image sizes is given in the text below Fig. 5. A potentially slight



**Fig. 4** Detailed view of a segmentation network. Convolutional network with pyramidal encoder/decoder architecture to compute the defective areas of a SEM image. The specific architecture detailed in the picture results from the exclusive usage of artificial training data and from the large input image size.

modification of the defect shape caused by the size reduction is not a problem at this point of the input image processing since the network returns a defect area around the defective feature but not the defect shape. The exact defect shape is computed inside this area in the next step that uses the original SEM image. This is explained in Sec. 3.4. In the next step, the image is encoded into basic structures relevant for defect detection by reducing the lateral size down to  $128 \times 128$  pixels and increasing the depth to 512 filters (“Conv 1,” “Conv 2,” and “Conv 3” in Fig. 4). This is realized with three consecutive convolutions with a filter size of  $13 \times 13$  pixels and a stride of  $2 \times 2$  pixels. The relatively large filter size is the result of tests that have shown that this specific dimension leads to the best results compared with smaller as well as larger filters. The tests have further demonstrated that the filter size is determined by the largest pitch to be considered by the network. In the real SEM images used in this work, this pitch is about 86 pixels on the resized input image, resulting in the  $13 \times 13$  pixels filter size. The next part of the network, the decoding part, is a mirror of the encoding part. The output image is reconstructed from the basic features in the middle layer. This is realized by an inverse convolution with a stride of  $-2 \times -2$  pixels and the same filter size of  $13 \times 13$  pixels. The negative stride numbers indicate an increase of the lateral image size. Furthermore, the image depth is reduced back to one layer. The activation function of the last convolutional layer is a sigmoid-function. All other layers use the ReLu-function. Finally, the output image is resized to the original input image size. During the training, the cross-entropy of input and output image is minimized with an Adam-optimizer.<sup>42</sup> The variables are initialized with a Glorot-uniform-initializer (also referred to as Xavier-uniform-initializer),<sup>43</sup> and a mixed precision of 32 bit for the variables and 64 bit for the error function is used. For all use cases according to Sec. 3.1, a training with 50 epochs was performed.

Several other network architectures with different numbers of convolutional layers, filter sizes, and strides as well as pooling, batch normalization, and residual connections have been investigated, with the result that the presented structure has demonstrated the best performance for this specific application task. As an alternative, the original U-net structure<sup>40</sup> developed for image segmentation was tested but could not reach the accuracy of the structure presented here. For example, the important true defect detection rate was only 93.9% compared with 100% reached by the network structure in this work. More details on the performance are given in Sec. 4. The investigation of a specific mask defect classification task in a previous work<sup>41</sup> has shown that other networks from the literature such as DenseNet and ResNet<sup>44,45</sup> also did not reach the same performance as the specifically developed architecture in our work. However, it is important to note the relatively large size of the images to be processed by the network. Due to the GPU and memory hardware available for our work, the feasible network size was limited. Larger residual networks with more layers (e.g., typically 30 and more) could not be realized or tested in our investigations.

As indicated in the output image in Fig. 4, the detected defect area bounding boxes are usually neither perfectly rectangular nor completely black. This deviation from the trained ideal case is used to finally decide whether a structure in the output image really belongs to a defect area on the SEM image. To do that, some image processing steps follow (left white box in Fig. 3). First, a certain area at the image borders is excluded from the analysis. Tests have shown that in a small image border region some artifacts can occur, which means that the network can detect defects by mistake. For the presented network, an area of 3% of the image size is sufficient for suppressing this effect. Then, a thresholding at 50% of the maximum output image value, contouring, and the contour bounding boxes are computed. Boxes smaller than the minimum defect size to be considered are removed, and overlapping boxes are combined. This eliminates areas not belonging to a defect and returns the bounding boxes around all detected defects.

In summary, the encoder–decoder structure with lateral size reduction and depth increase from layer to layer in the encoding part and lateral size increase and depth reduction in the decoding part are important for the functionality of the network and enable the use of simplified training data. A network with constant layer sizes was also trained with the same simplified data, but it could not generate acceptable results. With this network, artificial test SEM data were well processed, i.e., all defects in the test images were detected correctly and no artifacts appeared. In contrast to that, the results obtained from the real SEM test images were not acceptable. The defects were detected, but a large number of additional artifacts (e.g., 100 and even more per image) were returned by network.

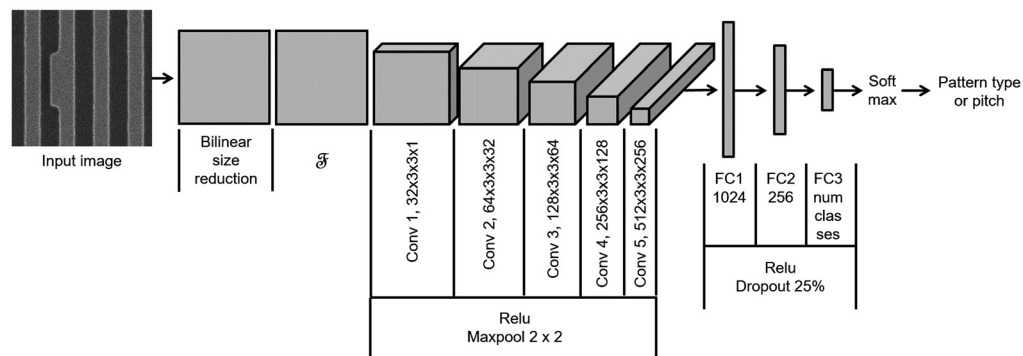
### 3.3.2 Classification network for pattern type and pitch

For the two classification tasks, during the SEM image processing (dark gray boxes in Fig. 3), a convolutional neural network structure is employed. The first classification determines the pattern type. As outlined in Sec. 3.1, horizontal line/space patterns, vertical line/space patterns, and contact patterns are currently considered. A network structure that was already used for other tasks in the field of defect classification<sup>41</sup> were applied successfully in this work as well. The architecture is shown in Fig. 5. For the second classification task, the determination of the pitch, the same network structure with one extension is used. In the second layer of the network, a Fourier transform is applied to the input image. The Fourier transform of the SEM image can be regarded as the simplified diffraction spectrum of the corresponding lithography mask patterns. Since the pitch is encoded in the diffraction spectrum, the analysis of the diffraction orders can be used to determine the pitch. Tests with and without Fourier transform have shown that, in the first case, the network reached 100% accuracy after one training epoch, i.e., all pitches are detected correctly. 100% accuracy is required for the correct processing of the SEM images. In the second case, the network did not reach 100% even after five epochs. Therefore, no further analysis was carried out. The Fourier transform is computed on the same GPU used for the network training and did not lead to a noticeable increase of the training time.

As for the segmentation networks, the input SEM images are resized in the first layer with a bilinear interpolation to a uniform size of  $1024 \times 1024$  pixels (bilinear size reduction in Fig. 5) to obtain acceptable training and analysis times. The increase of the input image size by a factor of 2 in both directions leads to a training time that is about four times longer. This rule of thumb was derived from tests performed on the specific hardware used in this work and applies to both networks in Figs. 4 and 5. The original image size of  $4000 \times 4000$  pixels would lead to an increase of the training time by a factor of about 16 compared with the  $1024 \times 1024$  pixel images. The training time of the segmentation networks (Fig. 4) with  $1024 \times 1024$  pixel images is around 3 days. A theoretical training time of about 48 days must be expected with  $4000 \times 4000$  pixel input images. Such a long time would make the network development very challenging. The next layer is the described Fourier transformation layer (“F” in Fig. 5), which is only used for the pitch classification. The convolutional part of the network performing a feature extraction with five convolutional layers (“Conv 1” to “Conv 5” in Fig. 5) is followed by the classification part with three fully connected layers (“FC1,” “FC2,” and “FC3” in Fig. 5). The output is the pattern type or pitch. Further network details are given in the figure.

### 3.4 Combination of Deep Learning and Analytical Image Processing

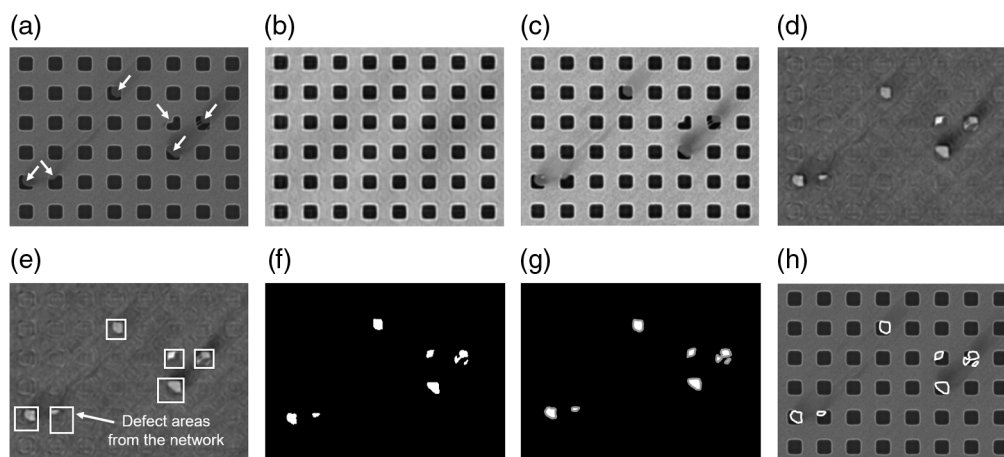
One particular challenge of the defect detection is the determination of the defect shape that is required for the lithographic assessment of the defect and as a starting point for the repair. The segmentation network explained in Sec. 3.3.1 is able to detect the defective areas inside a SEM image, but it is not capable of determining the exact shapes of the defects inside those



**Fig. 5** Convolutional neural network architecture for SEM image classification. The network is used for pattern type and pitch classification. The Fourier transformation layer (“F” in the figure) is only used for pitch classification.

areas. As mentioned, several other networks have been investigated to solve this task. However, no appropriate architecture has been found so far. Therefore, an analytical method was developed for this specific part. The basic idea of the method is to make use of the fact that a defect is a rare object inside the pattern structures of a SEM image. This means, from a theoretical point of view, that such a defect object has only a low amount of energy compared with the other objects of the image. To obtain the energy, the image is transformed into the spatial frequency domain by a two-dimensional spatial Fourier transformation. In this domain, all Fourier orders are sorted by their amount of energy. A following high-pass filtering removes all energies below a certain threshold, forcing the removal of all objects with corresponding low energies. Hence, defects are removed. Tests have shown that defects will be removed if all energies below 2% of the  $\pm$ first-order energies are cut off. The first-order energies are used as the reference since the zero-order is only an offset with no relevant information. This works for all tested pattern types, pitches, and defects as described in Sec. 3.1. A histogram matching between the original image and filtered image and the resulting difference provides the defect shapes. Since the method considers the energies of the diffraction orders only, it is not limited to periodic features. However, in practice, there is one important point to be considered. Defects are removed by the method. However, due to the threshold filtering of image energies, other modifications of the original image also may occur and cannot be avoided. This results in a difference image that clearly exhibits not only the defect shapes but also other objects that cannot be distinguished from the defects. Therefore, it is not possible to get the defect shapes solely from this difference image. At this point, the segmentation networks come into play. From the networks, all bounding boxes around the real defects are known. The difference image is only evaluated inside the boxes. In these areas, the defect shapes can be well detected due to the clear contrast of the defects. By applying an additional segmentation method based on the Chan–Vese algorithm,<sup>46</sup> the defect shapes are determined. The algorithm identifies objects that (clearly) differ from the environment without the requirement of well-defined object borders. This scenario applies to the defects. An algorithm for filling the remaining holes and contouring finally leads to the polygons describing the defect shapes. Another optional step can be applied at this point. If the network detects a defect area by mistake, the contrast of the potential objects in this area is typically much smaller than the contrast of a real defect. This property can be employed to exclude network artifacts, but the method still needs more investigations. Figure 6 shows the method.

Figure 6(a) shows a cut-out of an original SEM image with several intrusion defects indicated by the white arrows. The first processing step is the Fourier energy filtering, which is shown



**Fig. 6** Extraction of defect shapes by combination of Fourier energy filtering and output of the segmentation network. Sequence of image processing from (a) to (h): (a) cut-out of original SEM image, (b) Fourier energy-filtered image, (c) histogram-matched original image, (d) difference image between (b) and (c), (e) difference image with defect areas detected by the network, (f) segmentation of the defect areas in (e), (g) contouring of (f), and (h) original SEM image with determined defect shapes.

Fig. 6(b). The defects are removed, but other modifications caused by the filter occur (e.g., slight modifications around the feature edges). The next step is histogram matching between the original image and filtered image shown in Fig. 6(c). The gray values of the original image are adapted to the filtered image. The difference of the matched images, which is the following step, is shown in Fig. 6(d). The defect shapes can be seen, but other objects also appear in the image. At this point, the defect areas detected by the network come into play and are used for further processing. Figure 6(e) shows the difference image with the detected defect areas. In the next step, a segmentation of the difference image inside the defect areas only is carried out. This is shown in Fig. 6(f). Then, contouring determines the defect shapes [light gray polygons in Fig. 6(g)]. Finally, the determined defect polygons are displayed in the original SEM image, which is shown in Fig. 6(h). The important point is that the defect areas must be known for a successful application of the described procedure [see Fig. 6(e)]. This requires AI networks as an essential part of the defect shape determination.

#### 4 Defect Detection in Real SEM Images

Although the networks are supposed to be used for the analysis of real SEM images, they are exclusively trained with artificial data. To validate operation of the networks, a set of real SEM images was processed. However, for the monitoring of the training, only the artificial training data have been used due to the complex processing of the real SEM images. For all networks, the training ran until an average accuracy of 99% for randomly selected training images was reached. This was the case for all networks after about 50 training epochs. To compute the accuracy, the output images of the segmentation networks are compared with the corresponding training data. Tests show that this procedure leads to a very good accuracy for the processing of real SEM images, as will be demonstrated in the following.

The real SEM images used for validation include horizontal line/space patterns, vertical line/space patterns, and contact patterns with different pitches and duty ratios. All typical defect types appear in the images. Furthermore, for each pattern type, a large variety of different images is available. Details are given in Sec. 3.1. Therefore, the data are well suited for validating the operation of the network. For a single test, a SEM image is passed on the network, and the returned defect areas and defect shapes are assessed visually by comparing them with the original image. The goal is to monitor whether all defects on the SEM image were detected correctly and whether the determined defect shapes correspond to the real shapes without clear outliers and failures (denoted as “true defect shapes with acceptable quality” in Tables 1 and 2). A more detailed evaluation of the defect shapes is not done at this point since they will be calibrated and optimized with lithography simulations in a follow-up step. Table 1 shows the overall results for the line/space patterns.

The table clearly demonstrates the good functionality of the network. Most importantly, all real defects are detected by the network (see detected true defects versus real defects in the table). Furthermore, only two false defects that do not exist in the real images are detected (see detected false defects in the table). This leads to the following overall result for the investigated line/space patterns:

**Table 1** Network defect detection performance for real SEM images with line/space patterns. The SEM images, mask patterns, and defects are described in Sec. 3.1.

SEM images with line/space patterns	67
Real defects in all images	453
Detected true defects in all images	453
Detected false defects in all images	2
True defect shapes with acceptable quality	453

**Table 2** Network defect detection performance for real SEM images with contact patterns. The SEM images, mask patterns, and defects are described in Sec. 3.1.

SEM images with contact patterns	64
Real defects in all images	179
Detected true defects in all images	179
Detected false defects in all images	8
True defect shapes with acceptable quality	179

- True defect detection rate = 100%
- False defect detection rate = 0.44%
- True defect detection rate = 99.56%

The same investigation was done for the contact patterns. Table 2 summarizes the obtained results.

The good functionality of the network for the contact patterns can be demonstrated as well. As for the lines/spaces, the most important point is again that all real defects are detected by the network (see detected true defects versus real defects in the table). Furthermore, eight false defects that do not exist in the real images are detected (see detected false defects in the table). This leads to the following overall result for the investigated contact patterns:

- True defect detection rate = 100%
- False defect detection rate = 4.28%
- True defect detection rate = 95.72%

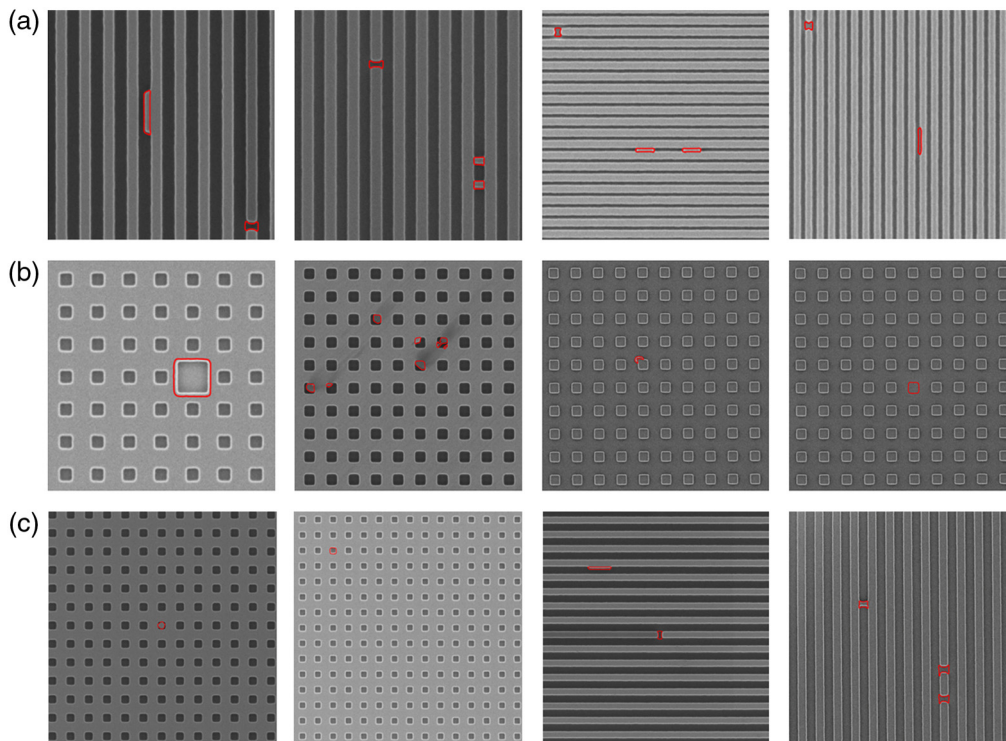
Figure 7 shows some selected examples of detected defects and their shapes. For better visibility, only a cut-out of the entire processed SEM image is shown.

The analysis result returned by the network after the processing of a SEM image is the polygons of all detected defect shapes with their exact coordinates. In the pictures in Fig. 7, the determined polygons are drawn in red on top of the respective original SEM image. It can be clearly seen that the defects are detected and that the determined defect shapes correspond well to the real defect shapes. Even more advanced defect shapes such as in the middle-left picture of the middle row are well determined. Finally, one example of detected false defects (see “detected false defects” in Tables 1 and 2) is shown in Fig. 8.

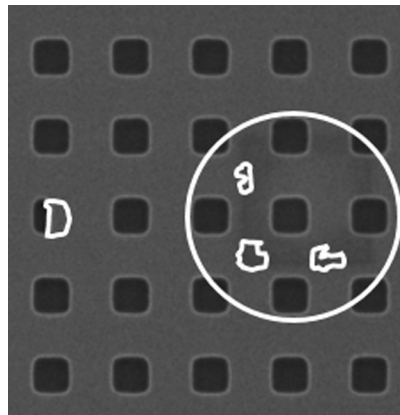
The figure shows three false defects inside the white circle. There are obviously no defects in the SEM image, but the network identifies the three shown structures (false defects). On the left side of the image, there is another real defect that is correctly detected. To further reduce the number of false defects, different methods are currently under investigation. Two approaches show promising results. False defects can be excluded by the classification of the defect type with a network. Alternatively, false defects can be identified by the determination of the defect contrast, which is typically much lower than a real defect contrast. In all investigated contact images, a total of eight false defects are returned by the network, and in all investigated line/space images, a total of two false defects are returned by the network. Nevertheless, the goal is to further reduce the false defects to the feasible minimum while keeping the true defect detection rate at 100%.

## 5 Further Work and Outlook

The presented networks are trained for periodic structures. The extension to more complex mask geometries such as logic structures is one of the next steps to be investigated. The current networks also have some potential for improvements. In the case of contact patterns, a closer



**Fig. 7** Selected defect detection examples. (a) Horizontal and vertical line/space patterns with extrusion, intrusion, broken line, and bridge defects are shown. (b) Contact patterns with extrusion, intrusion, oversize, and missing hole defects are shown. (c) Two contact patterns with a hole defect and an undersized defect and two line patterns with further extrusion, bridge, and broken line defects are shown. The defect shapes determined by the network are shown in red.



**Fig. 8** Example of three false defects (inside the white circle) detected by the network.

boundary around the defects can improve the defect shape determination. This requires an optimized training. A use case specific adaptation of the convolution filter sizes can reduce the training time. The Fourier energy filtering method in combination with the defect segmentation can be further optimized to be more stable in the case of low-quality SEM images with, for example, strong noise or blurred feature edges.

The next step in the SEM image processing flow after defect detection is the assessment of the lithographic impact of the defect. For such investigations, the results of the SEM image analysis have to be coupled with lithography simulations. To this end, the lithography simulator Dr. LiTHO of Fraunhofer IISB<sup>47</sup> is used. The lithography simulations typically compute the

through focus aerial images of mask patterns and evaluate the images in terms of relevant lithography metrics such as critical dimension, process window, normalized image log slope, non-tecentricity, and others. The numerical values of these metrics characterize the lithographic impact of a defect. The core element of the simulations is the rigorous computation of the mask diffraction spectrum, which requires the exact 3D geometry of the mask. Consequently, the next task to be solved is the transformation of the detected defective patterns in the SEM images into 3D mask structures and furthermore into a specific Manhattanized format required by the simulator.<sup>48</sup> This transformation needs advanced image processing in combination with additional information about the mask and mask materials to allow for the 3D extension of the data. After that, the determined defect shape is used to mimic a repair, which requires further processing of geometry data. The defect shape determined by the deep learning network serves as the starting point toward the final repair shape. The quality of this shape will be assessed with corresponding lithography simulations in the same way as the defects. Finally, further lithography simulations in combination with the optimization features of Dr. LiTHO will be used for the repair shape optimization.

## 6 Conclusions

We present a hybrid machine learning approach, which is able to detect mask pattern defects and their shapes in SEM images. In particular, a 100% true defect detection rate from real mask SEM images has been demonstrated. The system is based on a combination of several deep learning networks and analytical methods. In the current version, the networks are trained for line/space patterns and contact patterns with typical defects and a larger number of pitches and duty ratios. Further use cases can be added to the networks. The need for a large amount of labeled SEM images with programmed defects for the network training is addressed using purely computer-generated artificial training data in combination with a specific network architecture. The task of the exact defect shape determination is solved by a hybrid approach combining the deep learning networks with analytical image processing. The entire system structure is characterized by a modular design allowing for the extensibility to new use cases. The very good functionality and defect detection accuracy of the system is demonstrated with a set of real SEM images with line/space patterns and contact patterns with numerous defects.

## Acknowledgments

This project received funding from the Federal Ministry of Education and Research (Germany) in the framework of the BMBF project 16ESE0377K “PIN3S” and from the ECSEL Joint Undertaking (JU) under Grant Agreement No. 826422. JU receives support from the European Union’s Horizon 2020 research and innovation program and Netherlands, Belgium, Germany, France, Romania, Israel.

## References

1. G. E. Moore, “Cramming more components on electronic circuits,” *Electronics* **38**(8), 114 (1965).
2. H. J. Levinson, *Principles of Lithography*, 3rd ed., SPIE Press, Bellingham, Washington, DC (2011).
3. T. Liang et al., “Advanced photolithographic mask repair using electron beams,” *J. Vac. Sci. Technol. B* **23**(6), 3101–3105 (2005).
4. K. Edinger et al., “Electron-beam-based photomask repair,” *J. Vac. Sci. Technol. B* **22**(6), 2902–2906 (2004).
5. F. Emmert-Streib et al., “An introductory review of deep learning for prediction models with big data,” *Front. Artif. Intell.* **3**, 4 (2020).
6. A. Voulodimos et al., “Deep learning for computer vision: a brief review,” *Comput. Intell. Neurosci.* **2018**, 7068349 (2018).



7. K. Imoto et al., "A CNN-based transfer learning method for defect classification in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.* **32**(4), 455–459 (2019).
8. X. Chen et al., "A light-weighted CNN model for wafer structural defect detection," *IEEE Access* **8**, 24006–24018 (2020).
9. T. Schlosser et al., "A novel visual fault detection and classification system for semiconductor manufacturing using stacked hybrid convolutional neural networks," in *24th IEEE Int. Conf. Emerg. Technol. and Factory Autom.*, pp. 1511–1514 (2019).
10. K. Maksim et al., "Classification of wafer maps defect based on deep learning methods with small amount of data," in *Int. Conf. Eng. and Telecommun.*, pp. 1–5 (2019).
11. U. Batool et al., "Convolutional neural network for imbalanced data classification of silicon wafer defects," in *16th IEEE Int. Colloquium Signal Process. and Its Appl.*, pp. 230–235 (2020).
12. I. Graur et al., "New method of detection and classification of yield-impacting EUV mask defects," *Proc. SPIE* **9635**, 96350M (2015).
13. Y. Ma et al., "Machine learning based wafer defect detection," *Proc. SPIE* **10962**, 1096208 (2019).
14. J. Okude et al., "In-line schematic failure analysis technique by defect SEM images," *Proc. SPIE* **11611**, 116110M (2021).
15. Y. Okamoto et al., "Improvement of EPE measurement accuracy on ADI wafer, the method of using machine learning trained with CAD," *Proc. SPIE* **11611**, 116111W (2021).
16. W. Sun et al., "Evaluation of deep learning model for 3D profiling of HAR features using high-voltage CD-SEM," *Proc. SPIE* **11611**, 116110X (2021).
17. N. Chaudhary, S. A. Savari, and S. S. Yeddulapalli, "Line roughness estimation and Poisson denoising in scanning electron microscope images using deep learning," *J. Micro/Nanolith. MEMS MOEMS* **18**(2) 024001 (2019).
18. H. Lei et al., "Denoising sample-limited SEM images without clean data," *Proc. SPIE* **11611**, 116111A (2021).
19. B. Dey et al., "SEM image denoising with unsupervised machine learning for better defect inspection and metrology," *Proc. SPIE* **11611**, 1161115 (2021).
20. J. W. Park, A. Torres, and X. Song, "Litho-aware machine learning for hotspot detection," *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **37**(7), 1510–1514 (2018).
21. J. Zhang et al., "Lithography layout classification based on graph convolution network," *Proc. SPIE* **11613**, 116130U (2021).
22. F. Batarseh et al., "Pattern similarity profiling using semi-supervised learning algorithm," *Proc. SPIE* **11614**, 1161409 (2021).
23. X. Xu et al., "Subresolution assist feature generation with supervised data learning," *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **37**(6), 1225–1236 (2018).
24. W. Ciou et al., "SRAF placement with generative adversarial network," *Proc. SPIE* **11613**, 1161305 (2021).
25. S. Zhang et al., "Fast optical proximity correction based on graph convolution network," *Proc. SPIE* **11613**, 116130V (2021).
26. J. Lin et al., "Fast extreme ultraviolet lithography mask near-field calculation method based on machine learning," *Appl. Opt.* **59**, 2829–2838 (2020).
27. M. Liu et al., "On-machine surface defect detection using light scattering and deep learning," *J. Opt. Soc. Am. A* **37**, B53–B59 (2020).
28. L. Pang et al., "Making digital twins using the deep learning kit (DLK)," *Proc. SPIE* **11148**, 111480B (2019).
29. A. K. Baranwal et al., "A deep learning mask analysis toolset using mask SEM digital twins," *Proc. SPIE* **11518**, 1151814 (2020).
30. Z. Wang, L. Yu, and L. Pu, "Defect simulation in SEM images using generative adversarial networks," *Proc. SPIE* **11611**, 116110P (2021).
31. M. Kohlbrenner et al., *Pre-Training CNNs Using Convolutional Autoencoders*, TU Berlin, Berlin, Germany (2017).
32. Wikipedia, "Deep learning," [en.wikipedia.org/wiki/Deep\\_learning](https://en.wikipedia.org/wiki/Deep_learning).
33. S. J. Randolph, J. D. Fowlkes, and P. D. Rack, "Focused, nanoscale electron-beam-induced deposition and etching," *Crit. Rev. Solid State Mater. Sci.* **31**(3), 55–89 (2006).

34. V. Bakshi, *EUV Lithography*, 2nd revised ed., SPIE Press, Bellingham, Washington, DC (2018).
35. H. Schneider et al., “High-end EUV photomask repairs for 5 nm technology and beyond,” *Proc. SPIE* **11518**, 1151808 (2020).
36. D. C. Joy and D. E. Newbury, “Advanced SEM imaging,” *AIP Conf. Proc.* **449**, 653, (1998)
37. L. Reimer, *Scanning Electron Microscopy: Physics of Image Formation and Microanalysis*, Springer, Berlin, Germany (1998).
38. N. Marturi, S. Dembélé, and N. Piat, “Scanning electron microscope image signal-to-noise ratio monitoring for micro-nanomanipulation,” *Scanning* **36**, 419–429 (2014).
39. Wikipedia, “Digital twin,” [en.wikipedia.org/wiki/Digital\\_twin](https://en.wikipedia.org/wiki/Digital_twin).
40. O. Ronneberger, P. Fischer, and T. Brox, “U-Net: convolutional networks for biomedical image segmentation,” arXiv:1505.04597v1 (2015).
41. P. Evanschitzky, “Mask defect assessment from SEM images aided by deep learning methods,” in *EMLC Dresden*, VDE (2019).
42. D. Kingma and J. Ba, “Adam: a method for stochastic optimization,” in *Int. Conf. Learn. Represent.* (2014).
43. X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Proc. 13th Int. Conf. Artif. Intell. and Stat.*, Vol. 9, JMLR (2010).
44. G. Huang et al., “Densely connected convolutional networks,” in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 2261–2269 (2017).
45. K. He et al., “Deep residual learning for image recognition,” in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 770–778 (2016).
46. T. Chan et al., “Active contours without edges for vector-valued images,” *J. Vis. Commun. Image Represent.* **11**, 130–141 (2000).
47. T. Fuehner et al., “Dr. LiTHO: a development and research lithography simulator,” *Proc. SPIE* **6520**, 65203F (2007).
48. P. Evanschitzky and A. Erdmann, “Advanced EUV mask and imaging modeling,” *J. Micro/Nanolithgr. MEMS MOEMS* **16**(4), 041005 (2017).

Biographies of the authors are not available.