The International Journal of Advanced Manufacturing Technology (2021) 113:35–58
https://doi.org/10.1007/s00170-021-06592-8

SURVEY PAPER

Recent advances in surface defect inspection of industrial products using deep learning techniques

Xiaoqing Zheng [1] · Song Zheng [1] · Yaguang Kong [1] · Jie Chen [1]

Received: 12 March 2020 / Accepted: 3 January 2021 / Published online: 25 January 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd. part of Springer Nature 2021

Abstract
Manual surface inspection methods performed by quality inspectors do not satisfy the continuously increasing quality standards of industrial manufacturing processes. Machine vision provides a solution by using an automated visual inspection (AVI) system to perform quality inspection and remove defective products. Numerous studies and works have been conducted on surface inspection algorithms. With the advent of deep learning, a number of new algorithms have been developed for better inspection. In this paper, the state-of-the-art in surface defect inspection using deep learning is presented. In particular, we focus on the inspection of industrial products in semiconductor, steel, and fabric manufacturing processes. This work makes three contributions. First, we present the prior literature reviews on vision-based surface defect inspection and analyze the recent AVI-related hardware and software. Second, we review traditional surface defect inspection algorithms including statistical methods, spectral methods, model-based methods, and learning-based methods. Third, we investigate recent advances in deep learning-based inspection algorithms and present their applications in the steel, fabric, and semiconductor industries. Furthermore, we provide information on publicly available datasets containing surface image samples to facilitate the research on deep learning-based surface inspection.

Keywords Machine vision · Automated visual inspection · Automated surface inspection · Defect inspection · Defect detection · Deep learning

1 Introduction

Surface defect inspection refers to surface inspection of a finished product to identify defects such as scratches, pits, protrusions, and stains. Manual surface inspection methods performed by quality inspectors have the disadvantages of low efficiency, high labor intensity, low accuracy, low real-time performance, etc. They cannot satisfy the continuously increasing quality standards of industrial manufacturing processes. As one of the key technologies in manufacturing, machine vision provides a solution to fulfill the increasing demands on the documentation of quality and the traceability of products, by using engineering systems to perform quality inspection and remove defective products from production lines [1]. The machine vision system has the advantages of high precision, high efficiency, high speed and continuous detection, non-contact measurement, etc. Thereby, a large variety of solutions and applications has been inspired and utilized in this field since the 1980s, and their number has continued to grow. European and North American market data reveal that the growth of machine vision applications generally outperform the overall economic growth. Moreover, China has also become a major market for machine vision in recent years [2]. According to [3], the size of the global machine vision market was approximately USD $7.2 billion in 2017, growing 6.8% year-on-year.

Golnabi and Asadpour [4] classified the applications of machine vision into four categories: visual inspection, process control, part identification, and robotic guidance and control mechanisms. Among these, automated visual inspection (AVI) is the most significant and widely used application. Numerous studies and works have been performed on the research of AVI algorithms. Traditional AVI algorithms can be classified into statistical methods, spectral methods, model-based methods, and learning-based methods [5], which generally comprise two stages (feature extraction and defect

✉ Xiaoqing Zheng
zhengxiaoqing@hdu.edu.cn

[1] School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China

identification). Evidently, these depend heavily on human expert-designed features and are sensitive to variations in the application conditions. In recent years, deep learning has achieved remarkable performance in face recognition, speech recognition, natural language processing, etc. However, it has relatively few applications in the field of AVI. The probable reason is that deep learning relies strongly on a large amount of training data, whereas surface defect datasets are generally small and challenging to be collected or labeled. Nonetheless, compared to traditional defect detection methods, deep learning-based methods can learn high-level features automatically from training data without the design of manual features. They are more versatile in detecting different types of defects and less sensitive to variations in application conditions. In our work, recent advances and applications of deep learning-based AVI algorithms are investigated. In particular, we focus on AVI of industrial products in semiconductor, steel, and fabric manufacturing processes. Our literature survey indicates that a large number of AVI methods and applications have been studied in these fields. We believe that they are the most important application areas of AVI for the following reasons. The complexity and miniaturization of printed circuit boards (PCBs) and integrated circuits (ICs) may make inspection feasible only through AVI systems. Presently, many steps in semiconductor production can be performed reliably only through the use of machine vision [6]. AVI is also essential for quality control in the steel manufacturing process, because traditional manual surface inspection procedures are inadequate for guaranteeing quality surfaces [7]. In the fabric manufacturing process, AVI is definitely an important means to replace manual inspection, although it continues to be a challenging task owing to the variability in texture and diversity of defects [8]. This motivates many studies to identify a better solution.

The remainder of this paper is organized as follows: Prior literature reviews are presented in Section 2. The hardware, software, and algorithms of AVI are described in Section 3. In Section 4, traditional surface defect detection approaches are reviewed, including statistical methods, spectral methods, model-based methods, and learning-based methods. In Section 5, our analysis of deep learning networks and public datasets for surface defect inspection is presented. Our investigation of the deep learning-based inspection approaches and their applications in steel, fabric, and semiconductor industry is described. The challenges and solutions in this field are also discussed. This work is concluded in Section 6.

2 Prior literature review

A number of reviews and surveys of AVI methods have been conducted since the 1980s. The reviews published in the early years are available in [9], whereas recent review papers are presented chronologically in this section and illustrated in Table 1.

Malamas et al. [10] presented a review on industrial vision systems, applications, and tools in 2003 and discussed the important issues and directions for designing and developing industrial vision systems. In 2008, Xie [11] systematically reviewed the advances in surface inspection using computer vision and image processing techniques, particularly based on texture analysis methods under four categories: statistical approaches, structural approaches, filter-based methods, and model-based approaches. Kumar [12] surveyed the computer vision-based fabric defect detection methods, in 2008. He divided the methods into three categories: statistical, spectral, and model-based approaches. The paper also indicated that the combination of statistical, spectral, and model-based approaches could yield better results than any individual approach. Mahajan et al. [13] reviewed and described the fabric defect detection methods for visual inspection. They characterized the feature extraction and decision-making methods into three categories: statistical, spectral, and model-based methods. Hani et al. [14] presented a literature review of the pattern recognition algorithms for automated visual inspection of surface mount device printed circuit board (SMD-PCB). The review focused on segmentation algorithms, feature extraction algorithms, and performance evaluation of different types of classifiers. Ngan et al. [15] offered a survey of fabric defect detection methods with description of their characteristics, strengths, and weaknesses in 2011. They divided the methods into seven approaches (statistical, spectral, model based, learning, structural, hybrid, and motif based). Neogi et al. [7] presented a comprehensive review of vision-based steel surface inspection systems, in 2014. The review covered overall aspects of steel surface inspection and classified steel surfaces into six types: slab, billet, plate, hot strip, cold strip, and rod/bar. In 2015, Huang and Pan [9] studied AVI systems and reviewed their applications in the surface inspection of semiconductor products including wafer, TFT-LCD, and light-emitting diode (LED). They classified the inspection algorithms to projection methods, filter-based approaches, learning-based approaches, and hybrid methods. Hanbay et al. [16] presented a comprehensive literature review of fabric defect detection methods in 2016. Defect detection methods were divided into structural approaches, statistical approaches, spectral approaches, model-based approaches, learning approaches, and hybrid approaches. The main concepts underlying these approaches as well as with their strengths and weaknesses were discussed. Anitha and Rao [17] reviewed the defect detection methods for various categories of PCB such as single layer, double layer, and multilayer bare PCB and assembled PCB, in 2017. In 2018, Sun et al. [3] studied the research status and trends of steel inspection from the perspectives of detected object, hardware, and software. In addition, the detection algorithms were divided

Table 1 Recent literature review papers

Reference	Year	Content of the review	Object detected
[10]	2003	Reviewed the industrial vision systems, applications, and tools	Industrial products
[11]	2008	Analyzed texture analysis methods under four categories: statistical approaches, structural approaches, filter-based methods, and model-based approaches	Texture
[12]	2008	Studied fabric defect detection methods including statistical, spectral, and model-based approaches	Fabric
[13]	2009	Reviewed fabric defect detection methods under three categories: statistical, spectral, and model-based methods	Fabric
[15]	2011	Reviewed fabric defect detection methods and divided them into seven approaches (statistical, spectral, model based, learning, structural, hybrid, and motif based)	Fabric
[16]	2016	Reviewed fabric defect detection methods and divided them into six categories (structural, statistical, spectral, model based, learning based, and hybrid approaches)	Fabric
[14]	2011	Reviewed pattern recognition algorithms for automated visual inspection of SMD-PCB	Semiconductor
[9]	2015	Studied AVI systems in semiconductor industry and classified the inspection algorithms into projection methods, filter-based approaches, learning-based approaches, and hybrid methods	Semiconductor
[17]	2017	Reviewed defect detection methods for PCB categories including single layer, double layer, and multilayer bare PCB and assembled PCB	Semiconductor
[7]	2014	Reviewed the overall aspects of steel surface inspection and classified the steel surface types as slab, billet, plate, hot strip, cold strip, and rod/bar	Steel
[3]	2018	Studied steel surface inspection methods and divided the detection algorithms into four categories (statistical, filtering-based, model-based, and machine learning-based methods)	Steel

into statistical method, filtering-based methods, model-based methods, and machine learning-based methods.

3 AVI system

The principle of designing an automated visual inspection system is to replace the manual inspection process completely [18], as shown in Fig. 1. AVI is composed mainly of the following processes: image acquisition, defect detection, and quality control. An image acquisition process is aimed at measuring and acquiring images of the object to be inspected, using an optical system. The optical system consists of a digital camera or analog camera with a CCD or CMOS sensor as well as lighting system. The defect detection process refers to defect detection and recognition using image-processing techniques such as image preprocessing, feature extraction, and classification. The detection results are output to a quality control system to serve as a guide for defective product rejection. The detection results may include information on whether a sensed image is defective or defect free, the severity of the defects, and the category of the defects.

3.1 Camera and lighting

The sensor is the most important part of a camera. It is used to generate the image. There are two main sensors: CCD sensor and CMOS sensor. Compared with CCD, CMOS image

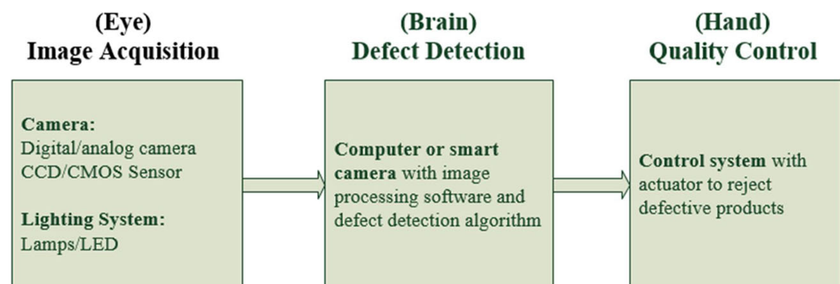

Fig. 1 Fundamental principle of automated visual inspection system

sensors are advanced technologies and are predominantly used in digital circuits. It is convenient for CMOS to incorporate the functions, such as analog-to-digital conversion, addressing, windowing, gain and offset adjustments, and smart preprocessing, on the chip for smart use. It is considered that CMOS will become the dominant sensor technology for machine vision in the future [2]. The trends of AVI also include a smart camera consisting of a sensor and a processing core that performs major image processing operations in situ and transmits only necessary information to the computer workstation [19].

Cameras can be categorized as analog or digital cameras depending on whether they produce an analog or digital video signal after acquiring an image. The transmission of an analog signal requires a special interface card called frame grabber, whereas a digital camera performs an analog-to-digital conversion internally and transmits the digital video signal to a computer. Analog video transmission has been the dominant technology in the machine vision industry for a long time. However, because analog video transmission may cause image quality degradation, its applications have declined in recent years. Advanced AVI systems typically use digital video transmission. Apart from higher image quality, digital cameras offer the advantages of significantly higher resolutions and frame rates, significantly smaller size, and less power requirement than those of analog cameras [1].

A suitable lighting system makes the entire vision inspection system more efficient and accurate. The types of light sources that are commonly used in machine vision include incandescent lamps, xenon lamps, fluorescent lamps, and light-emitting diode (LED) [1]. Presently, LED is the primary illumination method for machine vision. LED has a long life cycle, and its lifetime is commonly longer than 100,000 h. Its brightness can be controlled conveniently with low power consumption and low heat production. It can be designed into different sizes and shapes and can irradiate at different angles. Guidelines for setting LED light sources include the achievement of good contrast between the foreground and background for reliable measurement and of good contrast among the internal features [20].

3.2 Software and algorithms

The most frequently used imaging processing software for AVI includes OpenCV, Halcon, VisionPro, etc. OpenCV [21] is an open source image processing library with algorithms: smoothing images, morphology transformations, image pyramids, image moments, thresholding operations, histogram calculation, histogram comparison, template matching, etc. In addition, machine learning algorithms (such as support vector machines) and deep neural networks (such as GoogLeNet network) are included. Halcon [22] contains the image processing library used for blob analysis, morphology, matching, measuring, and identification. In addition, it provides 3D vision using shape-based 3D matching and surface-based 3D matching, as well as deep learning algorithms based on CNN. VisionPro [23] is a library with pattern matching, blob, caliper, line location, and image filtering algorithms. It also offers deep learning-based image analysis.

The frequently used online defect detection algorithms of industrial AVI are reference-based approaches and rule-based approaches. Reference-based approaches consist primarily of image subtraction and template matching. These measure the difference between a sensed image of the object to be inspected and a predefined reference pattern [9]. Image subtraction conducts pixel-by-pixel subtraction of a sensed image and a reference ideal image. The defects of the object are displayed in the subtracted images. Image subtraction is simple and can be implemented directly. However, it is excessively sensitive to image variation and may cause a lot of false positives. Template matching is feature-level comparison of the extracted object features and the predefined ideal templates, which are composed of feature patterns or models. The fundamental form of template matching is to move an image of the object to be detected across the template image and compute a similarity measure at each position [24]. The reference-based approach is intuitive, convenient for practical application, and reliable for detecting possible defects. However, it exhibits problems including inflexibility to variation and the need to store and maintain a large number of reference patterns.

A rule-based approach involves the extraction of features from the sensed object and comparison of those features to a list of rules that describes an ideal model. It can circumvent the need for an extensive database of templates by examining the sensed object with respect to a list of design rules or against the features that can be extracted from design rules [6]. The rules can utilize attributes such as surface area, perimeter, ratio of perimeter to area, number of holes, area of holes, minimum enclosing bounding box area, maximum radius, and minimum radius [25]. For PCB, the design rules can be [26] (1) the minimum and maximum trace widths for all the traces used, (2) the minimum and maximum circular pad diameters, (3) the minimum and maximum hole diameters, (4) the minimum conductor clearance, and (5) the minimum annular rings and trace termination rules. The disadvantage of the rule-based approach is that it may omit the flaws that do not violate the rules [27] or may require complicated schemes to eliminate false alarms [28].

In the early years, most of the industrial inspection systems utilized the template matching approach and rule-based comparison schemes [6]. However, these have been evolving into intelligent classifiers that have the capability to learn complex and subtle classification strategies [19]. With the advent of the state-of-the-art deep learning techniques, a number of new

algorithms have also been developed for better surface defect detection. As these algorithms mature, they will eventually promote the development of industrial surface defect detection algorithms.

3.3 Evaluation metrics

The evaluation metrics include error escape rate, false alarm rate, accuracy, precision, recall, and F1-score. Escape rate and false alarm rate are frequently used in addition to accuracy, for the performance evaluation of defect detection algorithms. Whereas error escape rate is the ratio of the number of defective samples detected as defect free to the total number of defective samples, false alarm rate is the ratio of the number of defect-free samples detected as defective to the total number of defect-free samples.

$$Error\ Escape\ Rate = FP/(TN + FP)$$
$$False\ Alarm\ Rate = FN/(TP + FN)$$
$$Accuracy = (TP + TN)/(TP + FN + TN + FP)$$
$$Precision = TP/(TP + FP)$$
$$Recall = TP/(TP + FN)$$
$$F1{-}Score = (2 \times Precision \times Recall)/(Precision + Recall)$$

where TP represents the numbers of true positives, FN represents the numbers of false negatives, TN represents the numbers of true negatives, and FP represents the numbers of false positives.

4 Traditional AVI algorithms

Traditional methods for defect detection proceed in two stages: feature extraction and defect identification. Features could be in the spatial domain, such as histogram, local binary pattern (LBP), and co-occurrence matrix, or in the transform domain, such as Fourier transform, wavelet transform, and Gabor transform [29]. Following feature extraction, defect identification can be performed by using common pattern classifiers such as SVM, K-nearest neighbor, random forest, and K-means. From the perspective of feature extraction and identification, surface defect detection can be categorized mainly into four general approaches: statistical methods, spectral methods, model-based methods, and learning-based methods [5]. The comparative studies are available in [3, 5, 9, 11–13, 15, 16]. Traditional defect inspection methods and their applications are illustrated in Table 2.

4.1 Statistical methods

Statistical methods measure the spatial distribution of pixel values with the assumption that the statistics of defect-free regions are stationary [13]. The defects are detected using the first-order statistics such as mean-, variance-, and histogram-based computations, in conjunction with the second-order statistics based on the co-occurrence matrix [30]. Popular statistical methods include histogram properties, co-occurrence matrix, mathematical morphology, and local binary pattern (LBP). Commonly used histogram statistics include range, mean, geometric mean, harmonic mean, standard deviation, variance, and median, as well as histogram comparison statistics such as L1 norm, L2 norm, Bhattacharyya distance, and Matusita distance [11]. Histogram properties have been successfully used in real-world applications as they are convenient to implement and invariant to rotation and translation [31]. The spatial gray level co-occurrence matrix (GLCM) introduced by Haralick [32] is widely used for texture defect detection. It describes the spatial distribution of texture by calculating the gray correlation between two pixels. Commonly used GLCM features include contrast, correlation, energy, entropy, and uniformity. Mathematical morphology is based on lattice theory and topology. It includes operations such as corrosion and expansion, open and closed operations, skeleton extraction, limit corrosion, hit-and-miss transformation, morphological gradient, top-hat transformation, particle analysis, and watershed transformation [3]. Mathematical morphology is highly suitable for defect detection of random or natural textures [16]. LBP, introduced by Ojala [33], considers the neighborhood of an image and compares the gray value of the pixel in the center with those of the other pixels in the neighborhood [34]. LBP is widely used in surface defect detection as they are robust to grayscale level variance such as illumination [35].

Several recent applications of statistical methods are available in [36–39]. Ashour et al. [36] presented a method based on gray-level co-occurrence matrix and discrete shearlet transform in 2018. Luo et al. [37] proposed a generalized completed local binary pattern framework with two variants for steel surface defect classification, in 2018. Li et al. [38] presented a fabric defect detection algorithm based on saliency histogram features, in 2019. Luo et al. [39] investigated the LBP method and proposed a selectively dominant LBP to quantitatively exploit the functional information from non-uniform patterns, in 2019.

4.2 Spectral methods

Spectral methods are also called filter-based methods. These transform signals from the spatial domain to the frequency domain by mathematical transformation, for feature extraction. Examples are Fourier transform, Gabor filter, and wavelet transform. There are a number of applications of these filter-based methods. Fourier transform is an important frequency-based analysis method for defect detection. It provides global information through an analysis of the frequency of signal over an entire time period. However, it cannot analyze local details of an image [35]. A Gabor filter is a type of

Table 2 Traditional defect inspection methods and their applications

Methods	Characteristics	Applications
Statistical methods	Histogram statistics include range, mean, geometric mean, and harmonic mean. Histogram methods are widely used owing to their computational simplicity and invariance to rotation, etc.	[38]
	GLCM calculates the gray correlation between two pixels. GLCM is widely used in texture defect detection with high detection accuracy	[36]
	LBP considers the neighborhood of an image. LBP is widely used in surface defect detection owing to their robustness to grayscale level variance such as illumination	[37, 39]
	Mathematical morphology is based on lattice theory and topology. It is suitable for random or natural textures	/
Spectral methods	Fourier transform is an important frequency-based analysis method for defect detection. It provides global information through an analysis of the frequency of signal over an entire time period. However, it cannot analyze local details of an image	[41, 45]
	A Gabor filter is a type of short-term Fourier transform. It can be customized with different scale and angle values based on different image texture structures. It is used extensively in texture defect detection.	[42, 46–50]
	Wavelet transform is based on the multi-resolution signal decomposition theory. It offers localized information from the horizontal, vertical, and diagonal directions on an input image	[40, 43, 44]
Model-based methods	Markov random field (MRF) used two random fields named the labeling field and feature field to describe an image	[55]
	An auto-regressive model describes the linear dependence between different pixels of an image by using linear equation systems. These incur low computational effort and cost	[56]
Learning-based methods	Support vector machine (SVM) is one of the most widely used machine learning and pattern recognition algorithms for traditional surface defect detection	[60–66]
	Artificial neural network (ANN) is also a frequently used machine learning and pattern recognition algorithm	[67–72]
	Other learning-based algorithms include random forest, clustering methods, and generic algorithms	[73–77]
Combination methods	Combining different categories of the aforementioned methods is effective for achieving optimal performance	[78–80]

short-term Fourier transform and applies a function of the Gaussian distribution. It is used extensively in texture defect detection because it can be customized with different scale and angle values based on different texture structures [16]. Wavelet transform is based on multi-resolution signal decomposition theory. It offers localized information from the horizontal, vertical, and diagonal directions on an input image [15]. Li and Tsai [40] presented a wavelet-based defect detection in solar wafer images with inhomogeneous texture, in 2012. Malek et al. [41] optimized the automated online fabric inspection by fast Fourier transform and cross-correlation, in 2013. Bissi et al. [42] adopted a Gabor filter for automated defect detection of uniform and structured fabrics, in 2013. Hu et al. [43] presented automated defect detection in textured materials using wavelet-domain hidden Markov models, in 2014. Wen et al. [44] developed a new fabric defect detection method based on adaptive wavelet by designing appropriate wavelet bases for different fabric images, in 2014. Hu et al. [45] presented an unsupervised defect detection method in textiles based on Fourier analysis and wavelet shrinkage, in 2015. Bi et al. [46] presented a defect detection method for LCD using Gabor filters, in 2015. Hu [47] presented an approach that addressed defect detection in textured surface by using an optimized elliptical Gabor filter, in 2015. Tong et al. [48] established a defect detection model using an optimized Gabor filter to address the woven fabric inspection problem in the textile industry, in 2016. Chol et al. [49] presented an algorithm for detecting pinholes in steel slabs by using a Gabor filter combination and morphological features, in 2017. Ma et al. [50] presented a surface defect detection method based on improved Gabor filters for scratch identification in industrial pipeline, in 2018.

4.3 Model-based methods

Model-based methods construct representations of images by modeling multiple properties of defects [51]. The popular model-based methods are the Markov random field (MRF) [52] and auto-regressive model [53]. In MRF, two random fields named the labeling field and feature field are used to describe the image, and a distribution function is used to describe the distribution of feature vectors under the condition of

the labeling field [3]. The application of MRFs for surface inspection can be traced to the 1990s [54]. Recently, Xu and Huang [55] developed a Gaussian Markov random field model for automatic pattern extraction and defect detection in nanomaterials, in 2012. An auto-regressive model describes the linear dependence between different pixels of an image by using linear equation systems, which incur less computational effort and cost than nonlinear systems [16]. Recently, Kulkarni et al. [56] presented an automatic surface defect detection algorithm using a two-dimensional auto-regressive model for fringe-projected-surface images, in 2019.

4.4 Learning-based methods

Learning-based methods are developed with machine learning and pattern recognition algorithms [9]. The highly popular pattern recognition algorithms such as support vector machine (SVM) [57], artificial neural network (ANN), k-nearest neighbor (k-NN) [58], random forest [59], generic algorithms, and clustering methods are applied frequently for defect classification. Among these, SVM is one of the most widely used classifiers for traditional surface defect detection.

4.4.1 SVM for surface inspection

Jia et al. [60] presented a real-time visual inspection system that used SVM to automatically learn complicated defect patterns for steel surface inspection, in 2004. Gao et al. [61] presented an algorithm for fabric defect detection based on dimensional histogram statistic and SVM, in 2006. Kang et al. [62] proposed an automated defect classification algorithm based on machine learning and the SVM classifier for TFT-LCD panel inspection, in 2009. Baly and Hajj [63] applied SVM for wafer classification and illustrated the selection of the values of SVM parameters, in 2012. Huang and Lu [64] proposed an automatic defect classification algorithm for TFT-LCD by using a linear SVM based on features including shape, histogram, and color, in 2013. Xie et al. [65] presented a defect detection and classification approach for PCBs and wafers, using SVM with a combination of median filter, background removal, morphological operation, and segmentation, in 2013. Zhang et al. [66] introduced an automated defect detection method for PCB, in 2018. In this method, detection was achieved by obtaining the defect region based on template matching, extracting the histogram and geometric features of the defect region, and using SVM classifier for recognition and classification.

4.4.2 ANN for surface inspection

Kumar et al. [67] proposed an approach for segmenting local textile defects using a feed-forward neural network, in 2003. Herein, principal component analysis (PCA) for dimension reduction of feature vectors was applied. Kang and Liu [68] introduced a method for detecting local defects in cold rolled strips, in 2005. PCA using singular value decomposition was also employed to reduce the dimension of the extracted feature vector. The feed-forward neural network was then adopted to detect the defects in the steel strips. Yang et al. [69] recommended a hybrid defect recognition method for steel surface inspection, in 2007. They used neural networks for identification and morphology processing for noise filtering. Ashour et al. [70] proposed a supervised texture classification method based on the feed forward ANN and the multi-class SVM, in 2008. Chen et al. [71] adopted four neural networks, namely, backpropagation, radial basis function, and two learning vector quantization networks, for TFT-LCD defect identification, in 2009. Tseng et al. [72] proposed an automatic defect classification scheme for color-filter production through three stages, namely, defect extraction, feature description, and defect-type classification using a neural network decision tree classifier, in 2011.

4.4.3 Other learning-based algorithms

Other learning-based algorithms include random forest, clustering methods, and generic algorithms. Several application examples are presented here. Kwon and Kang [73] proposed a defect detection algorithm based on random forest to determine the irregularity of the variety surface, in 2011. Tseng et al. [74] proposed an automatic detection method for multicrystalline solar cells, using binary clustering of features, in 2015. Hu et al. recommended a hybrid chromosome genetic algorithm for surface defect classification of a large-scale strip steel image collection, in 2016 [75]. Tian and Xu [76] developed an algorithm for identifying surface defects in hot rolled steel plates, based on a genetic algorithm and an extreme learning machine, in 2017. Piao et al. [77] proposed a decision tree ensemble learning-based method for wafer map failure pattern recognition, in 2018.

4.5 Combination methods

In different literature, defect detection methods are divided into different categories. It generally includes statistical, spectral, and model-based methods and occasionally also includes learning-based methods, structural methods, or other methods not described in this paper. The literature survey in our work reveals that regardless of how these methods are classified, the combinations of these methods can achieve optimal performance.

Several representative applications of combinations of methods are presented. Celik et al. [78] developed a system for fabric inspection through feature extraction based on wavelet transform, double thresholding binarization, and morphological operations and for defect classification using the

gray level co-occurrence matrix and a feed forward neural network. Wang et al. [79] proposed an online diagnosis system based on clustering techniques to identify spatial defect patterns for semiconductor manufacturing. Specifically, a spatial filter was used to assess whether the input data contained systematic cluster and to extract it from the noisy input. Then, an integrated clustering scheme combining fuzzy C means was adopted to separate the defect patterns. Furthermore, a decision tree was applied for decision-making. Nguyen et al. [80] proposed an automatic defect detection system for organic light-emitting diode (OLED) panels by combining three learning-based algorithms: SVM, random forest, and k-NN. Possible features were designed, and feature selection using PCA and random forest was adopted. Then, a hierarchical structure of classifiers (SVM, random forest, k-NN) was applied for defect identification.

5 Deep learning-based AVI algorithms

5.1 Deep learning networks and defects database

5.1.1 Deep learning networks

In its initial year (2006) [81], deep learning's application was focused on the MNIST digit image classification problem, thereby breaking the supremacy of SVMs [82]. Then, the breakthrough was achieved on the ImageNet [83] dataset and in ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Recently, Feng et al. [84] discussed how the deep neural network algorithms accomplish the computer vision tasks such as image classification, object detection, and image segmentation. The survey covered image classification networks including AlexNet, VGGNet, GoogLeNet, ResNet, and DenseNet and object detection algorithms including Faster-RCNN, YOLO, and SSD.

As a deep convolutional neural network, AlexNet [85] consists of five convolutional layers, three max-pooling layers, and three fully connected layers, having 60 million parameters and 650,000 neurons. AlexNet is known as the foundation work of modern deep CNN [84]. VGGNet [86] is a significantly deeper CNN network achieved by stacking convolutional layers and using an architecture with very small (3×3) convolution filters. It is capable of pushing the depth to 16–19 weight layers. In VGGNet, a stack of convolutional layers is followed by three fully connected layers and a final soft-max layer. GoogLeNet [87] modifies the convolution layers by using the Inception module to extend the depth and width of the networks. It has 22 layers. In the Inception module, 1×1 convolutions are used before 3×3 and 5×5 convolutions to reduce the computation cost. Google Inception-v3 [88] saves computation cost further by factorizing convolutions into smaller convolutions or asymmetric convolutions. For example, the 5×5 convolution is decomposed into two 3×3 convolution operations, and the convolution of kernel size $n \times n$ is decomposed into two convolutions of sizes $1 \times n$ and $n \times 1$. Although increasing the depth of networks aids in obtaining higher accuracy, as the number of network layers increase to a certain extent, the training accuracy saturates and then declines rapidly. ResNet [89] proposes residual building blocks to address this degradation problem. This involves the addition of parameter-free identity shortcut connections to feed-forward neural networks. The residual building blocks with short connections fully utilize features from previous layers to alleviate the degradation problem. Therefore, the network performance can be improved by stacking more residual blocks. This enables ResNet to have up to 152 layers. To further strengthen feature reuse and propagation, DenseNet [90] connects each layer to every other layer in a feed-forward fashion. A DenseNet network with L layers has $L \times (L+1) / 2$ direct connections. Owing to this dense connection structure, DenseNets can scale up to hundreds of layers without optimization challenges.

Apart from the aforementioned deeper and larger CNNs, a set of lightweight CNNs has been developed to reduce computation complexity while maintaining high accuracy. They are suitable for mobile or real-time applications that have limited computation resources or high computation speed requirements, such as the online AVI applications discussed in this paper.

The classical lightweight neural networks include SqueezeNet, MobileNet, and ShuffleNet. SqueezeNet [91] is a deep CNN network using the fire module, which comprises a squeeze layer and an expand layer. The squeeze layer is used to decrease the number of input channels to the expand layer and thereby reduce the quantity of parameters. Furthermore, the majority of the 3×3 filters are replaced by 1×1 filters to reduce the number of parameters. Although SqueezeNet has a minimal number of parameters, it achieves an accuracy level similar to that of AlexNet, on ImageNet with 50× fewer parameters. MobileNet [92] is also a lightweight neural network adapted for mobile and embedded vision applications with high accuracy. It utilizes depthwise separable convolution [93] and factorizes a standard convolution into a depthwise convolution and pointwise convolution to reduce computation and model size substantially. The depthwise convolution applies a filter to each input channel. Then, the pointwise convolution applies 1×1 convolution to combine the outputs of depthwise convolution. The cost of 3*3 depthwise separable convolution is $3*3*M*D*D + M*N*D*D$, and the cost of 3*3 standard convolution is $3*3*M*N*D*D$, where M is the number of input channels, N is the number of output channels, and $D*D$ is the size of output feature map. Therefore, compared with the standard convolution, the 3*3 depthwise separable convolution can save 8 to 9 times the amount of calculation at only a small reduction in accuracy [92]. MobileNet-v2 [94]

introduces a novel inverted residual layer to decrease the number of operations further. ShuffleNet [95] is also a computationally efficient CNN model designed for mobile devices. It has two novel operations: pointwise group convolution and channel shuffle. Pointwise group convolution is used to reduce computation complexity, whereas channel shuffle is used to aid the information flow across feature maps. ShuffleNet-V2 [96] introduces an operation called channel split to further improve the performance of its first version.

An object detection task is occasionally a part of a defect detection process. It is aimed at identifying the location of the object of interest. The most popular deep learning-based object detection algorithms are Faster-RCNN, YOLO, and SSD. Faster-RCNN [97] introduces a region proposal network (RPN), which is a fully convolutional network for proposal generation. It integrates RPN and Fast-RCNN [98] to share convolutional features and achieve high object detection accuracy. However, the Faster R-CNN still has a low detection speed, because it is a two-stage method that detects objects through region proposal and region classification. YOLO [99] and SSD [100] are one-stage object detection methods that detect objects using regression. In YOLO, a neural network predicts bounding boxes and class probabilities directly from full images in an evaluation. Although YOLO can achieve real-time speed, it is less accurate than the two-stage Faster-RCNN. The single shot multibox detector (SSD) outperforms YOLO in accuracy owing to two major improvements. First, SSD extracts important features from multi-scale CNN feature maps. Second, it adopts a number of default bounding boxes by following the concept of anchor proposed by Faster R-CNN [84].

Another set of deep learning algorithm suitable for surface defect inspection is unsupervised or semi-supervised learning methods. The representative methods are auto-encoder and generative adversarial network (GAN). Auto-encoder is a typical unsupervised learning algorithm based on two neural networks called encoder and decoder. It was introduced by Rumelhart et al. [101] in 1986 and extended to deep auto-encoder by Hinton et al. [102] in 2006. To achieve a higher robustness than that of the deep auto-encoder, the denoising auto-encoder [103] (introduced in 2008) adopts an approach that combines corruption and denoising to make the learned representations robust to partial corruption of the input pattern. The denoising auto-encoder is one of the common options for surface defect detection when considering unsupervised deep learning algorithms. GAN is an unsupervised learning framework introduced by Goodfellow et al. [104] in 2014 and has since developed [105]. It contains a generative model G and discriminative model D. G captures the data distribution with the aim of maximizing the probability of D committing an error. D estimates the probability that a sample originated from the training data rather than the generative model. The framework corresponds to a minimax two-player game. In addition, GAN can be extended for semi-supervised learning [106], which combines supervised learning and unsupervised learning in a framework. The above-mentioned deep learning networks suitable for defect dection are illustrated in Table 3.

5.1.2 Defects database

We have conducted a survey on publicity available datasets containing surface image samples of steel, textile, and semiconductor products. The information on a few datasets is presented in Table 4. This database information is provided with the aim of facilitating researchers from the AVI community or deep learning community in initiating further innovation and applications of deep learning in solving traditional AVI problems.

The DAGM texture database [107] was provided by the open competition “Weakly supervised learning for industrial optical inspection” organized by DAGM (German chapter of the International Association for Pattern Recognition) and the GNSS (German Chapter of the European Neural Network Society). The DAGM dataset consists of six types of artificially generated texture images. Each type has 1000 non-defective images and 150 defective images with a labeled defect on the background texture.

WM-811K [108] is a large publicly accessible dataset of wafer maps, containing 811,457 real-world wafer maps. Among these, 696,599 images are unique wafer maps. Approximately 20% of the wafer maps are labeled from one of the nine types (54,356 in the training set and 118,595 in the test set), which include eight defective types (Center, Donut, Edge-local, Edge-ring, Local, Near-full, Random, and Scratch) and a normal type.

The Northeastern University (NEU) surface defect database [109] contains six types of typical surface defects observed on hot-rolled steel strips: rolled-in scale (RS), patches (Pa), crazing (Cr), pitted surface (PS), inclusion (In), and scratches (Sc). The database contains 1800 grayscale images, each having 300 samples.

The DeepPCB dataset [110] contains 1500 image pairs of PCBs. Each of these consists of a defect-free template image and an aligned tested image with annotations including positions. Six common types of PCB defects are provided: open, short, mousebite, spur, pin hole, and spurious copper.

The magnetic tile defect dataset [111] contains 2688 defect images of six common magnetic tile defects, with their pixel level ground-truth labeled. The solar cell dataset [112] contains 2624 samples of 300 × 300 pixel 8-bit grayscale images of functional and defective solar cells with varying degrees of degradations, extracted from 44 solar modules.

RSDDs [113] contain images of two types of rail surface defects. One type comprises images of express rails (67 images), whereas the other type comprises images of common/heavy haul rails (128 images). Every image contains at least one defect and has a complex background with substantial noise.

Table 3 Deep learning networks suitable for defect detection

Category	Reference	
Classical deep CNNs	AlexNet [85]	As the foundation work of modern deep CNN, AlexNet consists of five convolutional layers, three max-pooling layers, and three fully connected layers
	VGGNet [86]	VGGNet uses 3×3 convolution filters and has a network depth of 16 to 19 weight layers. In VGGNet, a stack of convolutional layers is followed by three fully connected layers and a final soft-max layer
	GoogLeNet [87]	GoogLeNet has 22 layers and introduces the inception module. In the Inception module, 1×1 convolutions are used before 3×3 and 5×5 convolutions to reduce the computation cost
	Google Inception-v3 [88]	Inception-v3 factorizes convolutions into smaller convolutions or asymmetric convolutions to save computation cost. For example, the 5×5 convolution is decomposed into two 3×3 convolution operations, and the convolution of kernel size $n\times n$ is decomposed into two convolutions of sizes $1\times n$ and $n\times 1$
	ResNet [89]	ResNet [89] proposes residual building blocks to address degradation problem. The network performance can be improved by stacking more residual blocks. This enables ResNet to have up to 152 layers
	DenseNet [90]	DenseNet connects each layer to every other layer in a feed-forward fashion. Owing to this dense connection structure, DenseNets can scale up to hundreds of layers without optimization challenges.
Lightweight CNNs	SqueezeNet [91]	SqueezeNet is a lightweight CNN network by using the fire module to decrease the quantity of parameters. Furthermore, the majority of the 3×3 filters are replaced by 1×1 filters to reduce the number of parameters
	MobileNet [92, 94]	MobileNet is also a lightweight neural network adapted for mobile and embedded vision applications with high accuracy. It utilizes depthwise separable convolution
	ShuffleNet [95, 96]	ShuffleNet is also a computationally efficient CNN model designed for mobile devices. It has two novel operations: pointwise group convolution to reduce computation complexity and channel shuffle to aid the information flow across feature maps
Object detection algorithms	Faster-RCNN [97, 98]	Faster-RCNN introduces a region proposal network for proposal generation. It is a two-stage method that detects objects through region proposal and region classification
	YOLO [99]	YOLO is a one-stage object detection method that detects objects using regression. Although YOLO can achieve real-time speed, it is less accurate than the two-stage Faster-RCNN
	SSD [100]	SSD is also a one-stage object detection method. SSD outperforms YOLO in accuracy owing to two major improvements
Unsupervised learning algorithms	Auto-encoder [101–103]	Auto-encoder is a typical unsupervised learning algorithm based on two neural networks called encoder and decoder. One of its improved versions is the denoising auto-encoder, which is commonly used in defect detection when considering unsupervised deep learning algorithms
	GAN [104–106]	GAN is an unsupervised learning framework, which contains a generative model and a discriminative model. GAN's framework corresponds to a minimax two-player game. In addition, GAN can also be extended for semi-supervised learning

TILDA [114] is a benchmark database for textile defect detection. It contains 3200 images of eight representative textile types. Each textile is classified into seven defective types and a defect-free type, and each type consists of 50 images (768 × 512 pixel, 8-bit, gray level image).

In addition to the above-mentioned defect datasets, a few datasets are available that contain fabric or texture images without defects. For example, Fabrics Dataset [116] consists of approximately 2000 images of garment and fabric samples. The Kylberg texture dataset [115] contains 28 texture classes with 160 unique texture patches per class. The texture patch size is 576 × 576 pixels, and all the patches are normalized. The KTH-TIPS database [117] presently contains images of ten types of texture materials: sandpaper, crumpled aluminum foil, Styrofoam, sponge, corduroy, linen, cotton, brown bread, orange peel, and cracker B. Although there are no defect images in these datasets, they can be used for image classification or defect detection by synthesizing defects on them.

5.2 Research status of deep learning-based AVI algorithms and applications

The advent of the aforementioned deep learning techniques has inspired a number of novel deep learning-based defect detection algorithms. They integrate the two phases of the traditional detection methods, i.e., feature extraction and defect identification, into one phase. They extract features and classify defects simultaneously by learning from the training samples. They are not required to design a set of human features such as statistical or spectral features, as in traditional

Table 4 Publicly available surface image datasets

Dataset	Data type	Description	Download
DAGM texture dataset [107]	Texture images	Six types, each has 1000 non-defective images and 150 defective images	Link
WM-811 K [108]	Wafer maps	172,951 labeled images of wafer map (54,356 in training set and 118,595 in test set) with eight failure patterns and a normal pattern	Link
NEU surface database [109]	Defects images of hot-rolled steel strip	1800 grayscale images: six types, each has 300 samples	Link
DeepPCB [110]	PCB images	1500 image pairs, each consists of a defect-free template image and an image with six common defects	Link
Magnetic tile defect dataset [111]	Magnetic tile defect images	2688 images: six types of common magnetic tile defects	Link
Solar cell dataset [112]	Solar cell images	2624 grayscale images of functional and defective solar cells	Link
RSDDs dataset [113]	Rail surface defect images	Two types of rail surface images (67 images and 128 images)	Link
TILDA [114]	Textile images	3200 images of eight representative textile types. Each textile is classified into seven defective types and a defect-free type, and each type consists of 50 images	Link
Kylberg texture dataset [115]	Texture images	28 texture classes with 160 unique texture patches per class	Link
The Fabric dataset [116]	Fabric images	2000 samples of garments and fabrics	Link
KTH-TIPS database [117]	Texture images	Ten types of texture materials including sandpaper and crumpled aluminum foil	Link

methods. Without using an expert-designed feature set, deep learning-based detection algorithms can automatically generate distinct features from the training set and enable users to circumvent manual identification of rules for feature extraction or classification. Furthermore, they are generally capable of achieving higher detection accuracy [118].

A literature survey indicates that most of deep learning-based surface defect detection approaches employ deep CNN-based supervised learning for defect recognition. CNN is the most popular and used group of deep learning algorithms because of their wide application potential in pattern recognition [119]. It is a deep neural network architecture specialized in image processing and pattern recognition and whose hierarchical structure enables the extraction of multi-level image features to achieve accurate pattern identification [120]. CNN consists of three types of layers: convolutional layer, pooling layer, and fully connected layer. The convolutional layer learns feature representation of the input and outputs a feature map. The pooling layer is used for dimensionality reduction of the feature map. The fully connected layer performs the mapping of input data to a feature vector for final classification. As CNN exhibits a unique feature-learning capability, wherein it learns features from image samples automatically and exhibits strong reliability, it is generally the preferred option for surface quality inspection using deep learning.

The deep CNN-based approaches and applications of surface defect detection in the semiconductor industry are described in Table 5. The deep CNN-based approaches and applications for surface defect inspection of fabrics are illustrated in Table 6. The deep CNN-based approaches and applications for surface inspection of steel and other products are presented in Table 7. Semi-supervised learning methods and unsupervised learning methods are demonstrated in Table 8. These employ auto-encoder-based methods, Faster-RCNN, YOLO, SSD, and GAN for unsupervised or semi-supervised learning of surface defects.

5.2.1 CNN-based supervised learning methods for semiconductors

Yang et al. [121] presented an online detection method for Mura defects by combining a deep convolutional feature extractor and a sequential extreme learning machine classifier. It is capable of learning and recognizing a Mura defect image within 1.5 ms. Kim et al. [122] proposed a CNN network for surface mount technology (SMT) defect detection by modifying AlexNet and adopting the ResNet structure. Additional input image transformation was conducted by histogram stretching and chip region extraction to improve the detection accuracy. Kim et al. [123] proposed a CNN-based defect image classification model based on residual network for through-silicon via (TSV) process. They achieved classification performance of up to 97.2% accuracy. Jang et al. [124] proposed a defect inspection method by using deep CNN and defect probability images obtained from traditional inspection techniques. It outperforms a conventional CNN model using RGB or grayscale image. Zhang et al. [125] proposed a multi-task CNN model to handle the multi-label PCB classification problem by defining each label learning as a binary

Table 5 Deep CNN-based supervised learning approaches for surface inspection in semiconductor industry

Reference	Year	Deep learning methods	Dataset[a]	Accuracy[b]	Object detected
[121]	2017	CNN (AlexNet) and sequential extreme learning machine	14,491 images (TN: 11,500; TS: the rest)	94%	Flat panel displays
[122]	2018	CNN (AlexNet and ResNet) and image transform	Not supplied	91.1%	SMT
[123]	2018	CNN (ResNet)	TN: 37,400; TS: 9444 and 220,281, respectively	97.2%	Through-silicon via (TSV)
[124]	2018	CNN using defect probability image	1500 defect and 1796 good images (TN 50%; TS 50%)	98.813%	Chip-size package (CSP)
[125]	2018	CNN	1350 images (TN: 1200; TS: 150)	Multi-categories: 89.89%; two categories: 92.86%	PCB
[126]	2018	CNN	4526 images of normal and defect images	99.7%	PCB
[127]	2018	CNN (Inception v3)	7543 images (TN 6167; VS 576; TS 800)	91.125%	PCB
[128]	2018	CNN	1818 images for six types of PCB defects (TN 1069; VS 323; TS 426)	95.7%	PCB
[129]	2018	CNN	28,600 images (TN 15,400; VS 6600; TS 6600)	98.2%	Wafer
[130]	2019	CNN and extreme gradient boosting	11,730 images (TN 9384; TS 2346)	99.2%	Wafer
[131]	2019	CNN (VGG)	WM-811K (TN 54,355 and TS 118,595)	94.7% for multiple failure pattern	Wafer map
[132]	2019	CNN and k-NN	2123 images of five defect classes (TN70%; VS 15%; TS 15%)	96.2%	Wafer
[133]	2018	CNNs (LeNet, CifarCNN, and GoogleNet)	658 images (TN: 600; VS: 48; TS: 10)	98%	Photovoltaic cell
[112]	2019	SVM and CNN(VGG-19)	2624 images (TN 1968; TS 656)	88.42%	Photovoltaic cell
[134]	2019	CNN combined with class activation mapping technique	LEDC-GD-C1: TN 24,000; TS 6000 LEDC-GD-C2: TN 10,400; TS 2600	LEDC-GD-C1: 94.96%; LEDC-GD-C2: 94.49%	LED chip

[a] Number of image samples for the experiments. *TN* training set, *TS* testing set, *VS* validation set

[b] The best or average accuracy achieved

classification task. They achieved good performance on the PCB defect dataset. Deng et al. [126] proposed an automatic defect verification system by fast circuit comparison and deep CNN-based defect classification to decrease the false alarm rate of AVI for the PCB industry. Ghosh et al. [127] proposed a transfer learning-based method to classify PCB defects

Table 6 Deep CNN-based supervised learning approaches for fabric inspection

Reference	Year	Deep learning methods	Dataset[a]	Accuracy[b]	Object detected
[135]	2016	CNN	12 datasets, each has 2000 images of six types of surfaces	98%	Fabric, stone, wafer, wood, and solid/pearl color paint
[136]	2016	CNN	DAGM dataset (6 classes, each class consists of 1000 defect-free images and 150 defective images)	99.2%	Texture surfaces
[137]	2018	CNN	DAGM dataset (6 classes, each class consists of 1000 defect-free images and 150 defective images)	99.8%	Texture surfaces
[138]	2018	CCN with a multi-scaling averaging scheme	TILDA public dataset (1500 defective images containing six defects; 350 defect-free images)	96.55%	Fabric
[8]	2019	CNN	1200 images (50% defect; 50% defect free)	96.52%	Fabric
[118]	2019	CNN and multilayer perceptron	27,200 images (four subsets, each subset: TN 5400 TS 1400)	97.82%	Fabric

[a] Number of image samples for the experiments. TN for training set, TS for testing set, VS for validation set

[b] The best or average accuracy achieved

Table 7 Deep CNN-based supervised learning approaches for inspecting steel and other products

Reference	Year	Deep learning methods	Dataset[a]	Accuracy[b]	Object detected
[139]	2018	Preprocessing and CNN (AlexNet, GoogleNet, ResNet)	NEU public dataset (six types of defects and 300 samples for each class)	99.95%	Steel
[140]	2018	CNN (integrating ResNet-32 and WRN-28-10 and WRN-28-20)	NEU public dataset (six types of defects and 300 samples for each class)	99.889%	Steel
[141]	2019	CNN (Google Inception dual network)	4000 images of steel strips images	99.47%	Steel
[142]	2019	CNN (VGG, Inception, ResNet, DenseNet)	4806 images of steel strips (TN+VS: 3938; TS:868)	About 90%	Steel
[143]	2018	CNN	The training set consists of 3000 image samples expanded based on 230 defect-free screws, 287 stripped screws, 205 surface-damaged screws, and 256 surface dirty screws	98.4%	Metal screw
[144]	2018	CNN (Alexnet) and SVM	6283 images (TN 70%; TS 30%)	98.248%	Industrial products
[145]	2014	CNN	2532 images	99.44%	Rail
[51]	2018	CNN, thresholding, and segmenting	Three public dataset and an industrial dataset	99.27%, 90.50%, and 94.29% on the three public datasets	Weld, wood, steel, titanium fan blades

[a] Number of image samples for the experiments. TN for training set, TS for testing set, VS for validation set

[b] The best or average accuracy achieved

without reference images or the need to locate the defects in the images. An adaptation network was trained by extracting mid-level representations of PCB images from an intermediate layer of a pre-trained Inception-v3 network. Wei et al. [128] studied the method of extracting defect areas using morphology and deep CNN for PCB defect classification. They achieved significantly better results than a traditional classification algorithm based on digital image processing, on a

Table 8 Semi-supervised learning and unsupervised learning-based approaches for surface defect detection

Reference	Year	Deep learning methods	Dataset[a]	Accuracy[b]	Object detected
[29]	2017	Denoising auto-encoder	2600 samples of jacquard pattern fabric (TN 2000; TS 600)	98~99.47%	Fabric
[146]	2018	Convolutional denoising auto-encoder network	Four texture datasets	83.8%, 85.2%, 80.3%, and 84.0% on the four datasets	Fabric
[147]	2018	Auto-encoder network and data augmentation	Scratch images from NEU dataset, solder defect images, and industrial defect images for evaluation	True positive rate and false positive rate are given rather than accuracy	Surface defects
[148]	2018	CNN (VGG and ResNet), RPN, and Faster-RCNN	2132 images (TN 1549; VS 297; TS 286)	95.48%	Fabric
[149]	2018	YOLO	6 classes of 4655 steel strip surface images (TN 3455; TS 1200)	97.55%	Steel
[150]	2018	SSD and CNN(MobileNet)	Data augmentation based on 400 images obtained from the real-world	96.73%	Sealing surface
[151]	2019	SSD and speed model	3000 images (TN: 2140; TS: 860)	98.00%, 99.00%, 97.80%, and 79.40% for the four defect types	Darning needle
[152]	2019	Convolutional auto-encoder and semi-supervised GAN	(1) Hot rolled plates: TN 5000, TS 1400; (2) hot rolled steel strips: TN 9000, TS 1800; (3) cold rolled strips: TN 4000, TS 1800	97.2% for hot rolled plates, 98.2% for hot rolled steel strips, and 96.7% for cold rolled strips	Steel
[153]	2020	CNN with Pseudo-label	NEU dataset (TN 1500, TS 300)	90.7%	Steel

[a] Number of image samples for the experiments. TN for training set, TS for testing set, VS for validation set

[b] The best or average accuracy achieved

dataset containing 1818 images. Nakazawa et al. [129] presented a CNN-based wafer map defect pattern classification method for synthetic wafer maps, containing 22 defect classes generated theoretically. They achieved an overall classification accuracy of 98.2%. Yuan-Fu [130] employed CNN and extreme gradient boosting for wafer map retrieval tasks and defect pattern classification. They observed CNN to be more applicable for wafer map image classification because it is capable of learning relevant features from the input image. Ishida et al. [131] proposed a deep CNN network based on VGG to recognize wafer map failure patterns. A data augmentation technique with noise reduction was used for data processing. Experimental results on a benchmark dataset demonstrated the high accuracy of the method. Cheon et al. [132] proposed a wafer surface defect detection method by combining CNN and k-NN. It can extract effective features for defect classification without additional feature extraction algorithms and achieve high defect classification performance in wafer surface defect. Banda et al. [133] used deep learning for identifying defective photovoltaic cells automatically, based on CNNs including LeNet, CifarCNN, and GoogleNet architecture. The method successfully distinguished between a defective and a normal photovoltaic cell. Deitsch [112] investigated two approaches for automatic defect detection in solar photovoltaic cells: an approach based on hand-crafted features that are classified in an SVM and an end-to-end deep CNN approach. Experiments revealed the CNN-based approach to be more accurate than the SVM-based approach. Lin et al. [134] proposed an application of CNN in LED chip defect inspection. In the CNN, a class activation mapping technique was introduced to localize defect regions exactly. They achieved an accuracy of 94.96% for LED chip defects inspection.

5.2.2 CNN-based supervised learning methods for fabric

Park et al. [135] proposed a new surface defect inspection method for automatic visual inspection of dirties, scratches, burrs, and wears on surface parts. CNNs with different depths and layer nodes were tested to select an adequate structure for defect inspection. Weimer et al. [136] proposed a CNN method for texture surface defect recognition. They utilized 70% of 1,299,200 samples obtained after data augmentation for training and achieved a classification accuracy of 99.2%. Wang et al. [137] proposed a deep CNN for defect detection with less prior knowledge on the images and robust to noise. They achieved fast detection as well as high accuracy on a benchmark database. Jeyaraj et al. [138] proposed a multi-scaling CNN algorithm for fabric defect detection. They achieved an average accuracy of 96.55% on six different fabric materials. This is higher than that of conventional fabric defect detection. Gao et al. [8] investigated the problem of woven fabric defect detection using a CNN with multi-convolution and max-pooling layers. They obtained an overall detection accuracy of 96.52%. Furthermore, the authors constructed a high-quality database that includes images of common defects in woven fabric with solid color. Li et al. [118] proposed a compact CNN architecture with multilayer perceptron for detecting a few common fabric defects. In addition, multi-scale analysis, filter factorization, multiple locations pooling, and parameter reduction were used to improve the detection accuracy.

5.2.3 CNN-based supervised learning methods for steel

Saiz et al. [139] proposed an automatic defect classifier method for steel surfaces with two independent stages: preprocessing and CNN. They achieved a classification rate of 99.95%, outperforming other traditional detection methods on a publicly available dataset. Chen et al. [140] proposed an ensemble approach that integrates three deep CNNs for steel surface defect recognition: ResNet-32 and wide residual networks (WRNs) WRN-28-10 and WRN-28-20. Liu et al. [141] proposed a new neural network by utilizing Google Inception architecture and residual structure for steel defect detection. They achieved an accuracy of over 99.47%. Vannocci et al. [142] proposed an application of CNN in classifying steel strip images and a comparison with classical machine learning approaches. Thereby, they established the effectiveness and general validity of deep learning. Song et al. [143] developed a deep CNN-based detection method for micro defects on metal screw surfaces. A comparison with traditional template matching-based techniques and LeNet-5 has demonstrated the superiority of the proposed deep CNN-based method. Chun and Zhao [144] combined CNN and SVM to inspect industrial products more effectively. Here, CNN was used for feature extraction and SVM for decision-making. Soukup et al. [145] trained classical deep CNN on a database of photometric stereo images of rail surfaces in a purely supervised manner. They achieved significantly higher performance than that of a traditional model-based approach. Ren et al. [51] presented a deep learning-based approach requiring small training data for automated surface inspection on three public and an industrial datasets. It was realized by extracting patch feature using deep CNN, generating the defect heat map based on patch features, and predicting the defect area by thresholding and segmenting the heat map.

5.2.4 Semi-supervised learning and unsupervised learning methods

Li et al. [29] proposed a Fisher criterion-based stacked denoising auto-encoder with the objective of learning more discriminative features for patterned fabric defect detection when limited defective samples are available. Mei et al. [146] proposed an unsupervised learning-based automated approach by using a multi-scale convolutional denoising auto-encoder network and Gaussian

pyramid to detect and localize fabric defects. They achieved good overall performance. Mujeeb et al. [147] proposed an unsupervised learning algorithm to detect surface level defects by using a deep auto-encoder network and training input reference images. During training, various copies were generated automatically through data augmentation. The unsupervised algorithm does not rely on the availability of defect samples for training. Siegmund et al. [148] presented a comprehensive defect detection method for two common fabric defects groups. The proposed method employed VGG and ResNet for feature extraction. Then, a regional proposal network (RPN) and Faster-RCNN were used to generate the region proposal and detect objects. Li et al. [149] provided an end-to-end solution for the surface defects detection in steel strips. They improved YOLO by making it completely convolutional and capable of simultaneously predicting the class, location, and size information of defect regions. Li et al. [150] proposed a surface defect detection method by adopting an SSD network combined with MobileNet to identify the types and locations of surface defects. The method can automatically detect surface defects more accurately and rapidly than traditional machine learning methods. Yang et al. [151] proposed a real-time defect detection algorithm for tiny parts, based on an SSD and the speed model. The values of accuracy of detection of defect types 1, 2, 3, and 4 are 98.00%, 99.00%, 97.80%, and 79.40%, respectively. Di et al. [152] proposed a semi-supervised learning method based on convolutional auto-encoder and semi-supervised GAN to classify surface defects on steel. It has yielded remarkable performances. Gao et al. [153] proposed a semi-supervised deep learning method called PLCNN for steel surface defect recognition. PLCNN is a convolutional neural network improved by Pseudo-label that unlabeled data can be used in the training process. Comparative analysis with other conventional methods demonstrated that the proposed method has a significant improvement with the help of the unlabeled samples.

6 Discussions

6.1 Analysis of the deep learning-based defect dection algorithms

There are three paradigms for deep-learning defect detection: supervised learning, unsupervised learning, and semi-supervised learning. Supervised learning is most widely used and capable of achieving high detection accuracy and is reliable for online industrial application given sufficient training data. This is illustrated in Table 5, Table 6, and Table 7. We can conclude from the three tables that supervised learning-based defect detection methods generally utilize convolutional neural networks by adopting one of three approaches: the transfer learning approaches, the approach of constructing a CNN based on classical network structures such as ResNet, and the approach of constructing a CNN from scratch by stacking convolutional layers, pooling layers, and fully connected layers.

The transfer learning approach is to utilize transfer-learning technology to transfer a pre-trained model that has already been trained in a large dataset, to the target detection problem (which generally has a small training dataset). Transfer learning relaxes the hypothesis that the training data must be independent and identically distributed with the test data [154]. This enables a deep learning model to utilize the knowledge (including the model structure and pre-trained parameters) of a trained model from other fields. The common transfer learning approach is to leverage a pre-trained network and alter the final layers to fine tune the weight parameters on the target dataset [155]. For example, Ghosh et al. [127] proposed a transfer learning-based method to classify PCB defects by utilizing a pre-trained Inception-v3 network.

The second approach is to utilize the classical network structures such as AlexNet, VGGNet, Google Inception networks, and ResNet and perform a few modifications to make them adaptive for solving the target detection problems. This is the common way of utilizing CNN for defect detection. The aforementioned papers including Yang et al. [121], Kim et al. [122], Kim et al. [123], Ishida et al. [131], Banda et al. [133], Saiz et al. [139], Chen et al. [140], Liu et al. [141], Vannocci et al. [142], and Chun and Zhao [144] can be assigned into this category.

The third approach is to propose a novel CNN structure with different depth and width by stacking three types of layers (convolutional layer, pooling layer, and fully connected layer) in different ways. The convolutional layer learns the feature representation of the input and outputs a feature map. The pooling layer is used for dimensionality reduction of the feature map. The fully connected layer performs the mapping of input data to a feature vector for final classification. The best depth or width of a proposed CNN network can be obtained through comparison experiments and testing. The aforementioned papers including Zhang et al. [125], Deng et al. [126], Wei et al. [128], Nakazawa et al. [129], Cheon et al. [132], Park et al. [135], Wang et al. [137], Jeyaraj et al. [138], Gao et al. [8], Li et al. [118], Song et al. [143], and Soukup et al. [145] can be assigned to this category.

Another conclusion from Table 5–Table 7 is that a few of these CNN-based supervised learning methods introduce and combine other image processing technologies or pattern recognition methods. For example, Yang et al. [121] utilized CNN as well as sequential extreme learning machine for Mura defect detection. Cheon et al. [132] proposed a wafer defect detection method by combining CNN and k-NN. Li et al. [118] proposed a compact CNN architecture with multilayer perceptron for fabric defect detection. This is a feasible and occasionally effective approach to combining different methods to achieve the objective.

Moreover, we plotted a graph to illustrate the number of total images of the datasets used by the CNN-based methods listed in Table 5, Table 6, and Table 7 and the accuracy of these methods. There are 29 application examples listed in the three tables. We omit four of them: [122], the one which has not supplied image information; [123, 131], the two which have maximum number of images; and [51], the one which uses three datasets. The data number of the remaining 25 application examples are analyzed and illustrated in Fig. 2. As shown in Fig. 2, most of these use approximately 5000 image samples for training and testing. They can achieve high defect detection accuracy with an average accuracy of 96.82%. Therefore, for defect detection, CNN-based methods generally use thousands of original image samples to obtain high detection accuracy. Its demand for the number of training data is not as high as anticipated. Furthermore, as shown in Fig. 2, the trend of the accuracy curve does not match that of the curve of the number of images, which indicates that a higher amount of training images does not guarantee a higher accuracy. This contradicts our usual belief that a higher amount of training data can improve the performance of the model and alleviate the overfitting problem. Overfitting problem is a highly common problem that occurs when a large deep CNN model is applied to a small dataset. We may explain it this way. Obtaining sufficient training data is an important means to enable the training set generalize well to the test set and avoid overfitting. Alternative means can be to use regularization technology or to modify the neural network architecture. In addition to improving the generalization capability, we also need to minimize the gap between training error and human-level error by using larger neural networks, adopting appropriate hyperparameters or trying better optimization algorithms during training. Once we have obtained both good generalization capability and high training accuracy, the accuracy of a specific defect detection problem depends on the difficulty of the detection problem itself. For instance, if it is difficult for humans to recognize tiny defects or distinguish defects with similar features, then the supervised learning-based methods will face the same dilemma. That may explain why PCB and solar cell defects are not easy to be recognized in the following methods: [125], its accuracy is 89.89% for multi-categories; [127], its accuracy is 91.125%; and [112], its accuracy is 88.42%. The samples of these defects are described in Fig. 3.

As shown in Fig. 3, it is not easy for humans to clearly identify and classify these defects, let alone supervised learning methods. In this case, it is particularly important to set up a suitable lighting system to achieve good contrast between the foreground and background; thereby, AVI system can clearly capture these defective images without introducing additional noise. If there is noise, it may make supervised learning methods more confusing in defect recognition. Furthermore, it is also significant to correctly label these defects; otherwise, it may cause supervised learning methods more prone to make incorrect judgments. On the contrary, as illustrated in Fig. 4, the defects in NEU dataset are larger and more obvious, and the features of various defects are also significantly different; thereby, they are easier to be recognized by humans, and the same is true for supervised learning methods. As shown in Table 7, [139, 140] have achieved much higher accuracy, up to 99.95% and 99.889%, respectively.

Compared to above CNN-based supervised learning approaches, very few studies on unsupervised learning or semi-supervised learning-based defect detection approaches have been conducted. The frequently used unsupervised learning frameworks are auto-encoder and GAN. As presented in Table 8, Li et al. [29] and Mei et al. [146] proposed a denoising auto-encoder for fabric defect detection. Mujeeb et al. [147] utilized a deep auto-encoder network for surface defect detection. However, unsupervised learning is less reliable than the supervised learning method. Therefore, it has few online industrial AVI applications. Semi-supervised learning provides an alternative solution when insufficient labeled data are provided. It can achieve similar precision as supervised learning albeit using fewer labeling samples. However, the state-of-the-art semi-supervised learning technology innovated by the deep learning community has rarely been employed for defect detection.

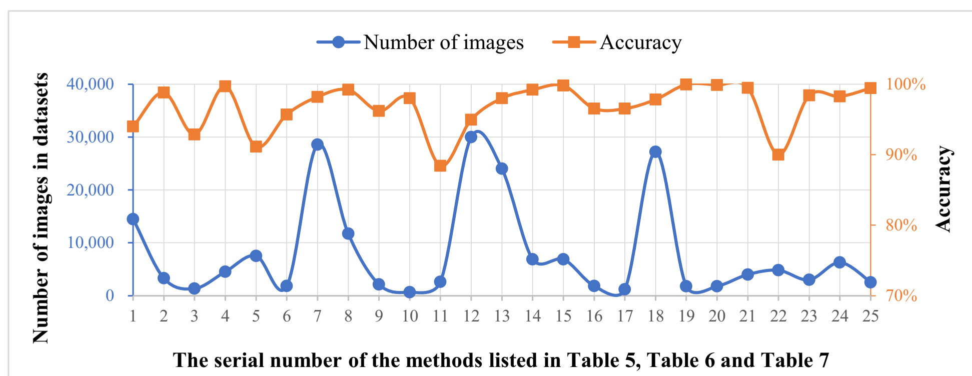

Fig. 2 Number of images and accuracy of the CNN-based methods listed in Tables 5, 6, and 7

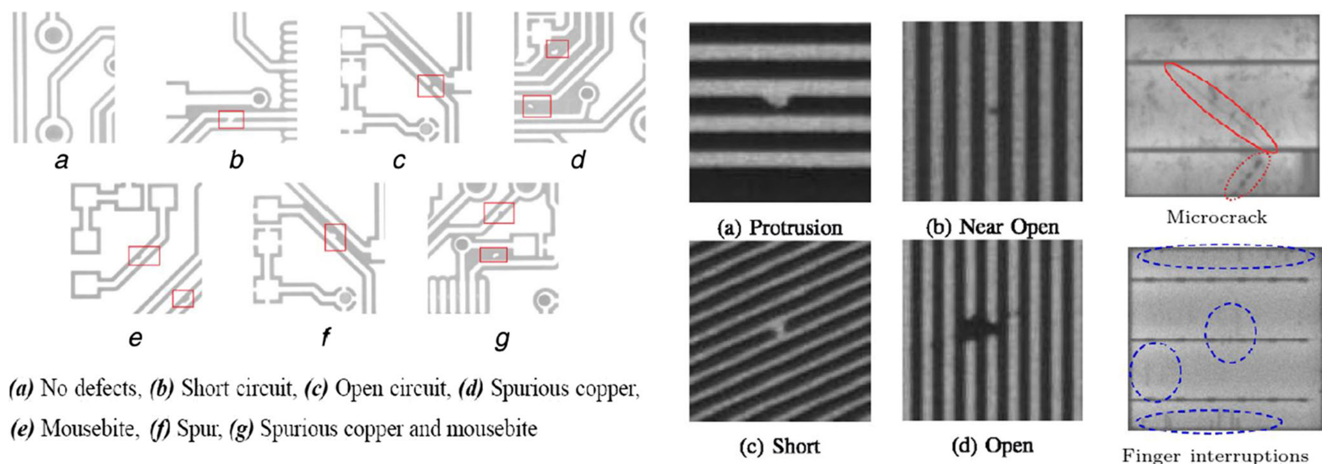

Fig. 3 Samples of PCB defects in [125] (left) and [127] (middle) and the solar cell defects in [112] (right)

6.2 Challenges and solutions

As demonstrated in the previous section, most of deep learning-based surface defect detection approaches employ deep CNN-based supervised learning for defect recognition. And they are frequently implemented in three ways. The first way is to use the transfer learning method, which utilizes the knowledge (including model structure and pre-trained parameters) of a trained model from other fields and fine tunes them on the target datasets in order to reduce the amount of training data or training time. The second way is to adopt classic convolutional neural network structures, such as Inception-v3 and ResNet, and modify them to a certain extent to make them suitable for the target defect detection problems. The third way is to construct a convolutional neural network from scratch by stacking convolutional layers, pooling layers, and fully connected layers together and train them to achieve the desired accuracy. However, these methods mainly consider the accuracy of defect recognition and classification and less consider how to achieve high efficiency and low computational cost. In order to improve the detection accuracy, these methods generally tend to deepen or expand network scale, which consumes a lot of computing time and requires high-performance computing resources. They are less able to meet the millisecond-level real-time detection requirements in industrial AVI applications, thereby limiting their application in industrial fields. Therefore, how to build a deep CNN-based defect detection model that meets both high precision and real-time requirements is a challenge for deep-learning-based AVI applications.

A probable solution is to directly use lightweight networks such as SqueezeNet [91], MobileNet [92], and ShuffleNet [95] as the main networks of defect detection, because they are tailored for mobile applications, or they are aimed at achieving a balance between lowest computation cost and highest accuracy. The details of these networks have been described in the previous section. An alternative solution is to utilize effective convolutional algorithms, such as depthwise separable convolution [93] applied in MobileNet and the fire module introduced by SqueezeNet. When considering saving the computational cost of convolution, depthwise separable convolution should always be the first choice, because a 3*3 depthwise separable convolution can save 8 to 9 times the amount of calculation at only a small reduction in accuracy. It is realized by decomposing standard convolution into depthwise convolution (each input channel is convoluted

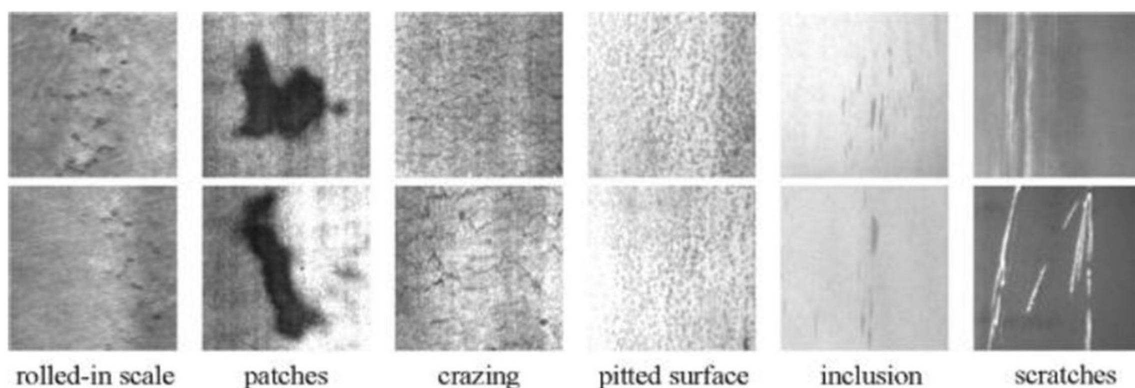

Fig. 4 Samples of defects in Neu dataset [140]

by applying a filter) and pointwise convolution (1*1 convolution to combine the outputs of depthwise convolution). The fire module has a squeeze convolution layer (which has only 1*1 filters), feeding into an expand layer that has a mix of 1*1 and 3*3 convolution filters. It sets the number of filters in the squeezer layer (all 1*1 convolutions) to be less than the filters in the expander layer (1*1 and 3*3 convolutions), so the squeeze layer helps to limit the number of input channels to the 3*3 filters, thereby reducing the calculation amount [91]. These convolution algorithms can greatly help achieve a high detection speed while maintaining a high detection accuracy. For instance, we have proposed a lightweight deep convolutional neural network based on the depthwise separable convolution and a squeeze-and-expand mechanism to detect the surface defects of the copper clad laminate (CCL) images obtained from the industrial CCL production line, and high computation speed has been achieved while maintaining good detection accuracy [156]. In general, by developing more lightweight networks or more efficient convolution algorithms, we can strike a balance between lowest computation cost and highest accuracy and finally realize rapid and accurate defect detection in industrial online applications.

Another challenge faced by AVI applications based on deep learning is that deep neural networks usually require a large amount of labeled data as training samples, but the preparation of labeled data incurs significant labor and time costs. Moreover, it is occasionally highly challenging or unfeasible to label or collect sufficient training data. At the same time, industrial high-speed production lines often produce defects that have never appeared before. The new defects are not included in the training samples, and this might also impede the application of deep learning in industrial AVI. Therefore, when a large amount of labeled data cannot be provided, how to use deep learning for defect detection is still a challenge.

Data augmentation technology can alleviate the problem of insufficient training samples to some extent. It preprocesses the original images by performing image transformation (such as flipping, random cropping, re-scaling, and color shifting) to expand the original dataset. The transformed image samples will be added to the original dataset to form an expanded dataset, which is fed to the network for training. Data augmentation can also be performed automatically during training [147]. But it cannot completely address the problem of insufficient data. Unsupervised learning can address the deficiency of training data, but it is less reliable than the supervised learning method and thereby unfeasible for online industrial AVI applications. An alternative solution could be to use semi-supervised learning paradigm. Semi-supervised learning can achieve similar precision as supervised learning albeit using fewer labeling samples. It uses both labeled and unlabeled data for training, which contrasts supervised learning (data all labeled) or unsupervised leaning (data all unlabeled) [157]. It can maximize the use of unlabeled data that are relatively easy to obtain. Traditional semi-supervised learning includes generative modeling and graph-based methods, etc. The details of these methods and more comprehensive overviews are provided in [157–159]. The newly proposed GAN can be attributed to the category of generative modeling and is one of the research hotspots in semi-supervised deep learning [160, 161]. However, it may suffer from unstable training and are too complicated to use in online AVI application. Many recent approaches for semi-supervised learning add a loss term which is computed on unlabeled data and encourages the model to generalize better to unseen data by using the following methods: entropy minimization, which encourages the model to output confident predictions on unlabeled data, and regularization, which encourages the model to produce the same output distribution when its inputs are perturbed and avoid overfitting the training data [162]. For instance, Berthelot et al. [162] from Google Research proposed a holistic semi-supervised learning algorithm named MixMatch, which introduces a unified loss term for unlabeled data that seamlessly reduces entropy. MixMatch has obtained state-of-the-art results across many datasets. Zheng et al. [163] have proposed a sophisticated algorithm based on MixMatch for automated surface inspection and revealed that it is effective for two public defect datasets (DAGM and NEU) and one industrial dataset (CCL).

In general, the challenge of achieving accurate and fast detection and the lack of sufficient training samples hinder the application of deep learning in industrial AVI. Probable solutions might be to utilize lightweight neural networks, efficient convolution algorithms, automatic data augmentation, semi-supervised deep learning paradigm, and other deep learning technologies that are still under development. Although extensive research has been conducted on deep learning-based defect detection, there is still consideration room for improvement in accuracy and computation speed. The state-of-the-art in deep learning should still be comprehensively studied to make online AVI applications more applicable.

7 Conclusion

Traditional defect detection algorithms generally conduct detection in two stages: feature extraction and defect identification. They have to design a set of human features, which are heavily dependent on extensive domain knowledge. Furthermore, these methods tend to work effectively only under specified conditions and are sensitive to input variations. Once the application condition varies, the algorithm needs to be adjusted substantially.

The recent advancement in deep learning provides generic tools that conduct detection in one stage. It learns features and identifies defects simultaneously. It is capable of learning

high-level features from training data automatically without requiring additional feature extractor or domain expert knowledge. A deep network-based detection approach is applicable to different objects and defect types as long as it is trained based on corresponding data. Moreover, it is insensitive to the variations in the input or application conditions when the training data has not varied substantially. In general, compared to the traditional defect detection methods, deep learning-based detection approaches are more automatic, more generic, and more robust because they do not have to design feature manually, are applicable to different types of objects and defect types, and insensitive to variations.

Notwithstanding these advantages, deep learning-based defect identification has been rarely used in practical industrial applications. It remains an unsolved problem given insufficient image samples. There are three paradigms for deep-learning defect detection: supervised learning, unsupervised learning, and semi-supervised learning. Supervised learning is the most widely used and capable of achieving high detection accuracy. However, it has disadvantage of being strongly dependent on a large amount of labeled training data. The preparation of labeled training data incurs significant labor and time costs. Moreover, it is occasionally highly challenging or unfeasible to label or collect sufficient training data. Unsupervised learning can address the deficiency of training data. However, it is less reliable than the supervised learning method and, therefore, unfeasible for online industrial AVI applications. Semi-supervised learning may provide a solution that can achieve similar precision as supervised learning albeit using fewer labeling samples. It uses both labeled and unlabeled data for training and maximizes the use of unlabeled data that are relatively easy to obtain. Many recent approaches for semi-supervised learning add a loss term which is computed on unlabeled data and encourages the model to generalize better to unlabeled data by using the following methods: entropy minimization, which encourages the model to output confident predictions on unlabeled data, and regularization, which encourages the model to produce the same output distribution when its inputs are perturbed and avoid overfitting the training data.

In addition, the absence of large amount of training samples for supervised learning-based defect detection can be alleviated through data augmentation technology. There are two approaches to conducting data augmentation. One is to preprocess the original images to expand the original dataset. This is implemented by performing image transformation such as flipping, random cropping, re-scaling, and color shifting. The transformed image samples are added to the original dataset to form an expanded dataset, which is fed to the network for training. An alternative method is to generate images automatically through data augmentation during training. This method can be utilized also in semi-supervised learning framework.

Another challenge faced by deep learning-based defect detection is to meet the millisecond-level real-time detection requirements in industrial applications while maintaining high accuracy. By developing lightweight neural networks or efficient convolution algorithms, we can strike a balance between lowest computation cost and highest accuracy and finally realize rapid and accurate deep learning-based defect detection in industrial online applications.

As semi-supervised learning and data augmentation can be used to alleviate or address the absence of large amount of training samples, and lightweight neural network and efficient convolution algorithms can be employed to improve the computation speed, we consider that deep learning exhibits the potential to gradually replace the traditional defect detection algorithms. The future development direction of deep learning-based defect detection approaches may be the utilization of automated data augmentation during training, the development of semi-supervised learning approaches to alleviate the problem of insufficient training data, and the innovation of efficient convolution algorithms and lightweight neural networks to meet real-time computation requirement. And we believe that with the continuous development of deep learning, surface defect inspection using deep learning has a promising future.

Funding This work was supported in part by the National Natural Science Foundation of China under grant number U1609212, Zhejiang Provincial Science and Technology Plan under grant number 2019C04021, and Zhejiang Province Public Technology Research Project under grant number LGG20F030002.

Data availability Not applicable.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent to publish Not applicable.

Code availability Not applicable.

References

1. Steger C, Ulrich M, Wiedemann C (2018) Machine vision algorithms and applications: second completely revised and Enlarged Edition. Wiley-VCH, Hoboken
2. Hornberg A (2017) Handbook of machine and computer vision: the guide for developers and users, Second edn. Wiley-VCH. https://doi.org/10.1002/9783527413409

3. Sun XH, Gu JA, Tang SX, Li J (2018) Research progress of visual inspection technology of steel products-a review. Appl Sci-Basel 8(11). https://doi.org/10.3390/app8112195
4. Golnabi H, Asadpour A (2007) Design and application of industrial machine vision systems. Robot Comput Integr Manuf 23(6): 630–637. https://doi.org/10.1016/j.rcim.2007.02.005
5. Ozseven T (2019) Surface defect detection and quantification with image processing methods. In: Ozseven T (ed) Theoretical investigations and applied studies in engineering. Ekin Publishing House, pp 63–98
6. Newman TS, Jain AK (1995) A survey of automated visual inspection. Comput Vis Image Underst 61(2):231–262
7. Neogi N, Mohanta DK, Dutta PK (2014) Review of vision-based steel surface inspection systems. EURASIP J Image Vide:1–19. https://doi.org/10.1186/1687-5281-2014-50
8. Gao C, Zhou J, Wong WK, Gao T Woven fabric defect detection based on convolutional neural network for binary classification. In: Artificial Intelligence on Fashion and Textiles Conference, AIFT 2018, June 27, 2018 - June 29, 2018, Hong Kong, China, 2019. Advances in intelligent systems and computing. Springer Verlag, pp 307–313. https://doi.org/10.1007/978-3-319-99695-0_37
9. Huang SH, Pan YC (2015) Automated visual inspection in the semiconductor industry: a survey. Comput Ind 66:1–10
10. Malamas EN, Petrakis EGM, Zervakis M, Petit L, Legat JD (2003) A survey on industrial vision systems, applications and tools. Image Vis Comput 21(2):171–188. https://doi.org/10.1016/S0262-8856(02)00152-X
11. Xie X (2008) A review of recent advances in surface defect detection using texture analysis techniques. Electron Lett Comput Vis Image Anal 7(3):1–22
12. Kumar A (2008) Computer-vision-based fabric defect detection: a survey. IEEE Trans Ind Electron 55(1):348–363. https://doi.org/10.1109/Tie.2007.896476
13. Mahajan PM, Kolhe SR, Patil PM (2009) A review of automatic fabric defect detection techniques. Adv Comput Res 1(2):18–29
14. Hani AFM, Malik AS, Kamil R, Thong CM (2012) A review of SMD-PCB defects and detection algorithms. Proc SPIE 8350. https://doi.org/10.1117/12.920531
15. Ngan HYT, Pang GKH, Yung NHC (2011) Automated fabric defect detection-a review. Image Vis Comput 29(7):442–458. https://doi.org/10.1016/j.imavis.2011.02.002
16. Hanbay K, Talu MF, Ozguven OF (2016) Fabric defect detection systems and methods-a systematic literature review. Optik 127(24):11960–11973. https://doi.org/10.1016/j.ijleo.2016.09.110
17. Anitha DB, Rao M (2017) A survey on defect detection in bare PCB and assembled PCB using image processing techniques. In: 2017 2nd Ieee international conference on wireless communications, signal processing and networking (Wispnet), pp 39–43. https://doi.org/10.1109/WiSPNET.2017.8299715
18. Lu R, Wu A, Zhang T, Wang Y (2018) Review on automated optical (visual) inspection and its application in defect detection. Acta Opt Sin 38(437 (8)):15–50
19. Shirvaikar M (2006) Trends in automated visual inspection. J Real Time Image Process 1(1):41–43. https://doi.org/10.1007/s11554-006-0009-6
20. Shreya SR, Priya CS, Rajeshware GS (2017) Design of machine vision system for high speed manufacturing environments. In: India Conference, 2017
21. OpenCV Tutorials. https://docs.opencv.org/master/d9/df8/tutorial_root.html. Accessed Oct. 2019
22. HALCON_18.11_brochure. https://www.mvtec.com. Accessed Oct. 2019
23. VisionPro. https://www.cognex.com. Accessed Oct. 2019
24. Demant C, Streicher-Abel B, Garnica C (2013) Industrial image processing: visual quality control in manufacturing, 2nd edn. Springer. https://doi.org/10.1007/978-3-642-33905-9
25. Van Gool L, Wambacq P, Oosterlinck A (1991) Intelligent robotic vision systems. Marcel Dekker Inc, New York
26. Bible RE (1984) Automated optical inspection of printed circuit boards. Test Meas World Oct.:208–213
27. Moganti M, Ercal F, Dagli CH, Shou T (1996) Automatic PCB inspection algorithms: a survey. Comput Vis Image Underst 63(2):287–313
28. Silven O, Virtanen I, Pietikainen M (1985) Cad data-based comparison method for printed wiring board (PWB) inspection. In: Society of Photo-optical Instrumentation Engineers Conference Series, 17 January 1985. https://doi.org/10.1117/12.946210
29. Li YD, Zhao WG, Pan JH (2017) Deformable patterned fabric defect detection with fisher criterion-based deep learning. IEEE Trans Autom Sci Eng 14(2):1256–1264. https://doi.org/10.1109/Tase.2016.2520955
30. Liu K, Wang H, Chen H, Qu E, Sun H (2017) Steel surface defect detection using a new Haar-Weibull-variance model in unsupervised manner. IEEE Trans Instrum Meas 99:1–12
31. Huangpeng Q, Zhang H, Zeng XR, Huang WW (2018) Automatic visual defect detection using texture prior and low-rank representation. IEEE Access 6:37965–37976. https://doi.org/10.1109/Access.2018.2852663
32. Haralick RM, Shanmugam K, Dinstein I'H (1973) Textural features for image classification. IEEE Trans Syst Man Cybern SMC-3(6):610–621. https://doi.org/10.1109/TSMC.1973.4309314
33. Ojala T, Harwood I (1996) A comparative study of texture measures with classification based on feature distributions. Pattern Recogn 29(1):51–59
34. Tajeripour F, Kabir E, Sheikhi A (2008) Fabric defect detection using modified local binary patterns. EURASIP J Adv Sig Process 2008. https://doi.org/10.1155/2008/783898
35. Tang B, Kong J, Wu S (2017) Review of surface defect detection based on machine vision. J Chin Image Graph 22(12):1640–1663. https://doi.org/10.11834/jig.160623
36. Ashour MW, Khalid F, Halin AA, Abdullah LN, Darwish SH (2018) Surface defects classification of hot-rolled steel strips using multi-directional shearlet features. Arabian Journal for Science & Engineering 44:2925–2932. https://doi.org/10.1007/s13369-018-3329-5
37. Luo Q, Sun Y, Li P, Simpson O, He Y (2018) Generalized completed local binary patterns for time-efficient steel surface defect classification. IEEE Trans Instrum Meas 99:1–13
38. Li M, Wan SH, Deng ZM, Wang YJ (2019) Fabric defect detection based on saliency histogram features. Comput Intell-Us 35(3): 517–534. https://doi.org/10.1111/coin.12206
39. Luo Q, Fang X, Sun Y, Liu L, Simpson O (2019) Surface defect classification for hot-rolled steel strips by selectively dominant local binary patterns. IEEE Access 99:1–1
40. Li WC, Tsai DM (2012) Wavelet-based defect detection in solar wafer images with inhomogeneous texture. Pattern Recogn 45(2): 742–756. https://doi.org/10.1016/j.patcog.2011.07.025
41. Malek AS, Drean JY, Bigue L, Osselin JF (2013) Optimization of automated online fabric inspection by fast Fourier transform (FFT) and cross-correlation. Text Res J 83(3):256–268. https://doi.org/10.1177/0040517512458340
42. Bissi L, Baruffa G, Placidi P, Ricci E, Scorzoni A, Valigi P (2013) Automated defect detection in uniform and structured fabrics using Gabor filters and PCA. J Vis Commun Image Represent 24(7):838–845
43. Hu GH, Zhang GH, Wang QH (2014) Automated defect detection in textured materials using wavelet-domain hidden Markov models. Opt Eng 53(9):093107

44. Wen ZJ, Cao JJ, Liu XP, Ying SH (2014) Fabric defects detection using adaptive wavelets. Int J Cloth Sci Technol 26(3):202–211. https://doi.org/10.1108/Ijcst-03-2013-0031
45. Hu GH, Wang QH, Zhang GH (2015) Unsupervised defect detection in textiles based on Fourier analysis and wavelet shrinkage. Appl Opt 54(10):2963–2980. https://doi.org/10.1364/Ao.54.002963
46. Bi X, Xu XP, Shen JH (2015) An automatic detection method of Mura defects for liquid crystal display using real Gabor filters. In: 2015 8th International Congress on Image and Signal Processing (Cisp), pp 871–875. https://doi.org/10.1109/CISP.2015.7408000
47. Hu GH (2015) Automated defect detection in textured surfaces using optimal elliptical Gabor filters. Optik 126(14):1331–1340. https://doi.org/10.1016/j.ijleo.2015.04.017
48. Tong L, Wong WK, Kwong CK (2016) Differential evolution-based optimal Gabor filter model for fabric inspection. Neurocomputing 173:1386–1401. https://doi.org/10.1016/j.neucom.2015.09.011
49. Chol DC, Jeon YJ, Kim SH, Moon S, Yun JP, Kim SW (2017) Detection of pinholes in steel slabs using Gabor filter combination and morphological features. ISIJ Int 57(6):1045–1053. https://doi.org/10.2355/isijinternational.ISIJINT-2016-160
50. Ma JX, Wang YX, Shi C, Lu CW (2018) Fast surface defect detection using improved Gabor filters. In: 2018 25th Ieee International Conference on Image Processing (Icip), pp 1508–1512. https://doi.org/10.1109/ICIP.2018.8451351
51. Ren RX, Hung T, Tan KC (2018) A generic deep-learning-based approach for automated surface inspection. IEEE Trans Cybern 48(3):929–940. https://doi.org/10.1109/Tcyb.2017.2668395
52. Kindermann R, Snell JL (1980) Markov random fields and their applications
53. Comer ML, Delp EJ (1999) Segmentation of textured images using a multiresolution Gaussian autoregressive model. IEEE Trans Image Process 8(3):408–420
54. Cohen FS, Fan Z, Attali S (1991) Automated inspection of textile fabrics using textural models. IEEE Trans Pattern Anal Mach Intell 13(8):803–808
55. Xu LJ, Huang Q (2012) Modeling the interactions among neighboring nanostructures for local feature characterization and defect detection. IEEE Trans Autom Sci Eng 9(4):745–754. https://doi.org/10.1109/Tase.2012.2209417
56. Kulkarni R, Banoth E, Pal P (2019) Automated surface feature detection using fringe projection: an autoregressive modeling-based approach. Opt Lasers Eng 121:506–511. https://doi.org/10.1016/j.optlaseng.2019.05.014
57. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297
58. Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 46(3):175–185
59. Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and regression trees. CRC Press, Boca Raton
60. Jia HB, Murphey YL, Shi JJ, Chang TS (2004) An intelligent real-time vision system for surface defect detection. Int C Patt Recog: 239–242. doi: https://doi.org/10.1109/Icpr.2004.1334512
61. Gao XD, Gao B, He Z, Xin WH (2006) Fabric defect detection based on support vector machine. J Text Res 27(5):26–28
62. Kang SB, Lee JH, Song KY, Pahk HJ (2009) Automatic defect classification of TFT-LCD panels using machine learning. In: 2009 IEEE International Symposium on Industrial Electronics, pp 2175–2177. https://doi.org/10.1109/ISIE.2009.5213760
63. Baly R, Hajj H (2012) Wafer classification using support vector machines. IEEE Trans Semicond Manuf 25(3):373–383. https://doi.org/10.1109/Tsm.2012.2196058
64. Huang W, Lu H (2013) Automatic defect classification of TFT-LCD panels with shape, histogram and color features. Int J Image Graph 13(03):1350011. https://doi.org/10.1142/S0219467813500113
65. Xie LJ, Huang R, Cao ZQ (2013) Detection and classification of defect patterns in optical inspection using support vector machines. Lect Notes Comput Sci 7995:376–384
66. Zhang ZQ, Wang XD, Liu S, Sun L, Sun LY, Guo YM (2018) An automatic recognition method for PCB visual defects. In: 2018 International Conference on Sensing, Diagnostics, Prognostics, and Control (Sdpc), pp 138–142. https://doi.org/10.1109/Sdpc.2018.00034
67. Kumar A (2003) Neural network based detection of local textile defects. Pattern Recogn 36(7):1645–1659. https://doi.org/10.1016/S0031-3203(03)00005-0
68. Kang GW, Liu HB (2005) Surface defects inspection of cold rolled strips based on neural network. In: 2005 International Conference on Machine Learning and Cybernetics 8:5034–5037. https://doi.org/10.1109/ICMLC.2005.1527830
69. Yang CH, Zhang JX, Ji G, Fu YJ, Hong X (2007) Recognition of defects in steel surface image based on neural networks and morphology. In: Second Workshop on Digital Media and Its Application in Museum & Heritage, Proceedings, pp 72–75. https://doi.org/10.1109/Dmamh.2007.56
70. Ashour MW, Hussin MF, Mahar KM (2008) Supervised texture classification using several features extraction techniques based on ANN and SVM. I C Comput Syst Appl:567–574. https://doi.org/10.1109/Aiccsa.2008.4493588
71. Chen LF, Su CT, Chen MH (2009) A neural-network approach for defect recognition in TFT-LCD photolithography process. IEEE Trans Electron Packag Manuf 32(1):1–8
72. Tseng DC, Chung IL, Tsai PL, Chou CM (2011) Defect classification for Lcd color filters using neural-network decision tree classifier. Int J Innov Comput I 7(7a):3695–3707
73. Kwon BG, Kang DJ (2011) Fast defect detection algorithm on the variety surface with random forest using GPUs. In: 2011 11th International Conference on Control, Automation and Systems (Iccas), pp 1135–1136
74. Tseng DC, Liu YS, Chou CM (2015) Automatic finger interruption detection in electroluminescence images of multicrystalline solar cells. Math Probl Eng 2015:1–12. https://doi.org/10.1155/2015/879675
75. Hu H, Liu Y, Liu M, Nie L (2016) Surface defect classification in large-scale strip steel image collection via hybrid chromosome genetic algorithm. Neurocomputing 181:86–95
76. Tian SY, Xu K (2017) An algorithm for surface defect identification of steel plates based on genetic algorithm and extreme learning machine. Metals-Basel 7(8). https://doi.org/10.3390/met7080311
77. Piao M, Jin CH, Lee JY, Byun JY (2018) Decision tree ensemble-based wafer map failure pattern recognition based on radon transform-based features. IEEE Trans Semicond Manuf 31(2):250–257. https://doi.org/10.1109/Tsm.2018.2806931
78. Celik HI, Dulger LU, Topalbekiroglu M (2014) Development of a machine vision system: real-time fabric defect detection and classification with neural networks. J Text Inst 105(6):575–585. https://doi.org/10.1080/00405000.2013.827393
79. Wang CH, Wang SJ, Lee WD (2006) Automatic identification of spatial defect patterns for semiconductor manufacturing. Int J Prod Res 44(23):5169–5185. https://doi.org/10.1080/02772240600610822
80. Nguyen VH, Pham VH, Cui X, Ma M, Kim H (2017) Design and evaluation of features and classifiers for OLED panel defect recognition in machine vision. J Inf Telecommun:334–350. https://doi.org/10.1080/24751839.2017.1355717
81. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. https://doi.org/10.1038/nature14539

82. Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal 35(8): 1798–1828. https://doi.org/10.1109/Tpami.2013.50
83. Deng J, Dong W, Socher R, Li LJ, Li K, Li FF (2009) ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp 248–255. https://doi.org/10.1109/CVPR.2009.5206848
84. Feng X, Jiang Y, Yang X, Du M, Li X (2019) Computer vision algorithms and hardware implementations: a survey. Integration. 69:309–320. https://doi.org/10.1016/j.vlsi.2019.07.005
85. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: 2012 25th International Conference on Neural Information Processing Systems 1:1097–1105. https://doi.org/10.5555/2999134.2999257
86. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556. https://arxiv.org/abs/1409.1556
87. Szegedy C, Liu W, Jia YQ, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: 2015 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr), pp 1–9. https://doi.org/10.1109/CVPR.2015.7298594
88. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: 2016 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr), pp 2818–2826. https://doi.org/10.1109/Cvpr.2016.308
89. He KM, Zhang XY, Ren SQ, Sun J (2016) Deep residual learning for image recognition. In: 2016 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr), pp 770–778. https://doi.org/10.1109/Cvpr.2016.90
90. Huang G, Liu Z, van der Maaten L, Weinberger K (2017) Densely connected convolutional networks. In: Conference on Computer Vision and Pattern Recognition, 2017. https://doi.org/10.1109/CVPR.2017.243
91. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K (2016) SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. arXiv:1602.07360. https://arxiv.org/abs/1602.07360
92. Howard AG, Zhu M, Bo C, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861. https://arxiv.org/abs/1704.04861
93. Sifre L (2014) Rigid-motion scattering for image classification. Ecole Polytechnique, Paris
94. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C (2018) MobileNetV2: inverted residuals and linear bottlenecks. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 4510–4520. https://doi.org/10.1109/CVPR.2018.00474
95. Zhang X, Zhou XY, Lin MX, Sun R (2018) ShuffleNet: an extremely efficient convolutional neural network for mobile devices. In: 2018 Ieee/Cvf Conference on Computer Vision and Pattern Recognition (Cvpr), pp 6848–6856. https://doi.org/10.1109/Cvpr.2018.00716
96. Ma N, Zhang X, Zheng H-T, Jian S (2018) ShuffleNet V2: practical guidelines for efficient CNN architecture design. In: 2018 European Conference on Computer Vision (ECCV). https://doi.org/10.1007/978-3-030-01264-9_8
97. Ren SQ, He KM, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 39(6):1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031
98. Girshick R (2015) Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp 1440–1448. https://doi.org/10.1109/ICCV.2015.169
99. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: 2016 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr), pp 779–788. https://doi.org/10.1109/Cvpr.2016.91
100. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) SSD: Single Shot MultiBox Detector. Computer vision - Eccv 2016, Pt I 9905:21–37. https://doi.org/10.1007/978-3-319-46448-0_2
101. Rumelhart DE (1986) Learning representations by back-propagating errors. Nature. https://doi.org/10.1016/B978-1-4832-1446-7.50035-2
102. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507. https://doi.org/10.1126/science.1127647
103. Vincent P, Larochelle H, Bengio Y, Manzagol PA (2008) Extracting and composing robust features with denoising autoencoders. In: the 25th International Conference on Machine Learning (ICML 2008), pp 1096–1103. https://doi.org/10.1145/1390156.1390294
104. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. Adv Neural Inf Proces Syst 27(nips 2014):27
105. Schlegl T, Seebock P, Waldstein SM, Langs G, Schmidt-Erfurth U (2019) f-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks. Med Image Anal 54:30–44. https://doi.org/10.1016/j.media.2019.01.010
106. Akcay S, Atapour-Abarghouei A, Breckon TP (2019) GANomaly: semi-supervised anomaly detection via adversarial training. Computer vision - Accv 2018, Pt Iii 11363:622–637. https://doi.org/10.1007/978-3-030-20893-6_39
107. DAGM texture dataset. https://hci.iwr.uni-heidelberg.de/node/3616. Accessed Oct. 2019
108. Wu MJ, Jang JSR, Chen JL (2015) Wafer map failure pattern recognition and similarity ranking for large-scale data sets. IEEE Trans Semicond Manuf 28(1):1–12. https://doi.org/10.1109/Tsm.2014.2364237
109. Song KC, Yan YH (2013) A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. Appl Surf Sci 285:858–864. https://doi.org/10.1016/j.apsusc.2013.09.002
110. Tang S, He F, Huang X, Yang J (2019) Online PCB defect detector on a new PCB defect dataset
111. Huang YB, Qiu CY, Guo Y, Wang XN, Yuan K (2018) Surface defect saliency of magnetic tile. Ieee Int Con Auto Sc:612–617
112. Deitsch S, Christlein V, Berger S, Buerhop-Lutz C, Maier A, Gallwitz F, Riess C (2019) Automatic classification of defective photovoltaic module cells in electroluminescence images. Sol Energy 185:455–468
113. Gan JR, Li QT, Wang JZ, Yu HM (2017) A hierarchical extractor-based visual rail surface inspection system. IEEE Sensors J 17(23):7935–7944. https://doi.org/10.1109/Jsen.2017.2761858
114. TILDA Textile Texture-Database (1996). https://lmb.informatik.uni-freiburg.de/resources/datasets/tilda.en.html. Accessed Oct. 2019 2019
115. Kylberg G (2011) The Kylberg Texture Dataset v. 1.0. External report (Blue series) vol No. 35. Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University. http://www.cb.uu.se/~gustaf/texture/. Accessed 19 Jan 2021
116. Kampouris C, Zafeiriou S, Ghosh A, Malassiotis S (2016) Fine-grained material classification using micro-geometry and reflectance. Computer vision - Eccv 2016, Pt V 9909:778–792. https://doi.org/10.1007/978-3-319-46454-1_47

117. Fritz M, Hayman E, Caputo B, Eklundh J-O (2019) The KTH-TIPS database. Accessed Oct. 2019
118. Li YY, Zhang D, Lee DJ (2019) Automatic fabric defect detection with a wide-and-compact network. Neurocomputing 329:329–338. https://doi.org/10.1016/j.neucom.2018.10.070
119. Michalski P, Ruszczak B, Tomaszewski M (2018) Convolutional neural networks implementations for computer vision. Adv Intell Syst 720:98–110. https://doi.org/10.1007/978-3-319-75025-5_10
120. Caggiano A, Zhang JJ, Alfieri V, Caiazzo F, Gao R, Teti R (2019) Machine learning-based image processing for on-line defect recognition in additive manufacturing. Cirp Ann Manuf Technol 68(1):451–454. https://doi.org/10.1016/j.cirp.2019.03.021
121. Yang H, Mei S, Song K, Tao B, Yin Z (2018) Transfer-learning-based online Mura defect classification. IEEE Trans Semicond Manuf 31(1):116–123. https://doi.org/10.1109/TSM.2017.2777499
122. Kim Y-G, Lim D-U, Ryu J-H, Park T-H SMD Defect classification by convolution neural network and PCB image transform. In: 3rd IEEE International Conference on Computing, Communication and Security, ICCCS 2018, October 25, 2018 - October 27, 2018, Kathmandu, Nepal, 2018. Proceedings on 2018 IEEE 3rd International Conference on Computing, Communication and Security, ICCCS 2018. Institute of Electrical and Electronics Engineers Inc, pp 180–183. https://doi.org/10.1109/CCCS.2018.8586818
123. Kim J, Kim S, Kwon N, Kang H, Kim Y, Lee C Deep learning based automatic defect classification in through-silicon Via process: FA: Factory automation. In: 29th Annual SEMI Advanced Semiconductor Manufacturing Conference, Saratoga Springs, NY, United states, 2018 2018. 2018 29th Annual SEMI Advanced Semiconductor Manufacturing Conference, ASMC 2018. Institute of Electrical and Electronics Engineers Inc, pp 35–39. https://doi.org/10.1109/ASMC.2018.8373144
124. Jang C, Yun S, Hwang H, Shin H, Kim SS, Park Y (2018) A defect inspection method for machine vision using defect probability image with deep convolutional neural network. In: 2018 Asian Conference on Computer Vision (ACCV), pp 142–154. https://doi.org/10.1007/978-3-030-20887-5_9
125. Zhang L, Jin Y, Yang X, Li X, Duan X, Sun Y, Liu H (2018) Convolutional neural network-based multi-label classification of PCB defects. J Eng 16:1612–1616. https://doi.org/10.1049/joe.2018.8279
126. Deng Y-S, Luo A-C, Dai M-J Building an automatic defect verification system using deep neural network for PCB defect classification. In: 4th International Conference on Frontiers of Signal Processing, ICFSP 2018, September 24, 2018 - September 27, 2018, Poitiers, France, 2018. 2018 4th International Conference on Frontiers of Signal Processing, ICFSP 2018. Institute of Electrical and Electronics Engineers Inc, pp 145–149. https://doi.org/10.1109/ICFSP.2018.8552045
127. Ghosh B, Bhuyan MK, Sasmal P, Iwahori Y, Gadde P Defect classification of printed circuit boards based on transfer learning. In: 2018 IEEE Applied Signal Processing Conference, ASPCON 2018, December 7, 2018 - December 9, 2018, Kolkata, India, 2018. Proceedings of 2018 IEEE Applied Signal Processing Conference, ASPCON 2018. Institute of Electrical and Electronics Engineers Inc, pp 245–248. https://doi.org/10.1109/ASPCON.2018.8748670
128. Wei P, Liu C, Liu M, Gao Y, Liu H (2018) CNN based reference comparison method for classifying bare PCB defects. J Eng 2018(16):1528–1533. https://doi.org/10.1049/joe.2018.8271
129. Nakazawa T, Kulkarni DV (2018) Wafer map defect pattern classification and image retrieval using convolutional neural network. IEEE Trans Semicond Manuf 31(2):309–314. https://doi.org/10.1109/TSM.2018.2795466
130. Yuan-Fu Y (2019) A deep learning model for identification of defect patterns in semiconductor wafer map. In: 30th Annual SEMI Advanced Semiconductor Manufacturing Conference, ASMC 2019. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ASMC.2019.8791815
131. Ishida T, Nitta I, Fukuda D, Kanazawa Y (2019) Deep learning-based wafer-map failure pattern recognition framework. In: 2019 20th International Symposium on Quality Electronic Design (Isqed), pp 291–297. https://doi.org/10.1109/ISQED.2019.8697407
132. Cheon S, Lee H, Kim CO, Lee SH (2019) Convolutional neural network for wafer surface defect classification and the detection of unknown defect class. IEEE Trans Semicond Manuf 32(2):163–170. https://doi.org/10.1109/Tsm.2019.2902657
133. Banda P, Barnard L A deep learning approach to photovoltaic cell defect classification. In: 2018 Annual Conference of the South African Institute of Computer Scientists and Information Technologists: Technology for Change, Port Elizabeth, South Africa, 2018 2018. ACM International Conference Proceeding Series. Association for Computing Machinery, pp 215–221. https://doi.org/10.1145/3278681.3278707
134. Lin H, Li B, Wang XG, Shu YF, Niu SL (2019) Automated defect inspection of LED chip using deep convolutional neural network. J Intell Manuf 30(6):2525–2534. https://doi.org/10.1007/s10845-018-1415-x
135. Park JK, Kwon BK, Park JH, Kang DJ (2016) Machine learning-based imaging system for surface defect inspection. Int J Precis Eng Manuf Green Technol 3(3):303–310
136. Weimer D, Scholz-Reiter B, Shpitalni M (2016) Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. Cirp Ann Manuf Technol 65(1):417–420. https://doi.org/10.1016/j.cirp.2016.04.072
137. Wang T, Chen Y, Qiao MN, Snoussi H (2018) A fast and robust convolutional neural network-based defect detection model in product quality control. Int J Adv Manuf Technol 94(9–12):3465–3471. https://doi.org/10.1007/s00170-017-0882-0
138. Jeyaraj PR, Samuel Nadar ER (2019) Computer vision for automatic detection and classification of fabric defect employing deep learning algorithm. Int J Cloth Sci Technol 31(4):510–521. https://doi.org/10.1108/IJCST-11-2018-0135
139. Saiz FA, Serrano I, Barandiaran I, Sanchez JR A robust and fast deep learning-based method for defect classification in steel surfaces. In: 9th International Conference on Intelligent Systems, IS 2018, September 25, 2018 - September 27, 2018, Funchal - Madeira, Portugal, 2018. 9th International Conference on Intelligent Systems 2018: Theory, Research and Innovation in Applications, IS 2018 - Proceedings. Institute of Electrical and Electronics Engineers Inc, pp 455–460. https://doi.org/10.1109/IS.2018.8710501
140. Chen W, Gao Y, Gao L, Li XA (2018) New ensemble approach based on deep convolutional neural networks for steel surface defect classification. In: 51st CIRP Conference on Manufacturing Systems, CIRP CMS 2018, May 16, 2018 - May 18, 2018, Stockholm, Sweden. Elsevier B.V, pp 1069–1072. https://doi.org/10.1016/j.procir.2018.03.264
141. Liu Z, Wang X, Chen X Inception dual network for steel strip defect detection. In: 16th IEEE International Conference on Networking, Sensing and Control, ICNSC 2019, May 9, 2019 - May 11, 2019, Banff, AB, Canada, 2019. Proceedings of the 2019 IEEE 16th International Conference on Networking, Sensing and Control, ICNSC 2019. Institute of Electrical and Electronics Engineers Inc, pp 409–414. https://doi.org/10.1109/ICNSC.2019.8743190
142. Vannocci M, Ritacco A, Castellano A, Galli F, Vannucci M, Iannino V, Colla V Flatness defect detection and classification in hot rolled steel strips using convolutional neural networks. In:

15th International Work-Conference on Artificial Neural Networks, IWANN 2019, June 12, 2019 - June 14, 2019, Gran Canaria, Spain, 2019. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp 220–234. https://doi.org/10.1007/978-3-030-20518-8_19

143. Song LM, Li XY, Yang YG, Zhu XJ, Guo QH, Yang HD (2018) Detection of micro-defects on metal screw surfaces based on deep convolutional neural networks. Sensors-Basel 18(11). https://doi.org/10.3390/s18113709
144. Chun LP, Zhao QF (2018) Product surface defect detection based on deep learning. In: 2018 16th Ieee Int Conf on Dependable, Autonom and Secure Comp, pp 250–255. https://doi.org/10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00051
145. Soukup D, Huber-Mork R (2014) Convolutional neural networks for steel surface defect detection from photometric stereo images. Advances in visual computing (Isvc 2014), Pt 1 8887:668–677
146. Mei S, Wang YD, Wen GJ (2018) Automatic fabric defect detection with a multi-scale convolutional denoising autoencoder network model. Sensors-Basel 18(4). https://doi.org/10.3390/s18041064
147. Mujeeb A, Dai WT, Erdt M, Sourin A (2018) Unsupervised surface defect detection using deep autoencoders and data augmentation. In: 2018 International Conference on Cyberworlds (Cw), pp 391–398. https://doi.org/10.1109/Cw.2018.00076
148. Siegmund D, Prajapati A, Kirchbuchner F, Kuijper A (2018) An integrated deep neural network for defect detection in dynamic textile textures. In: Progress in Artificial Intelligence and Pattern Recognition, Iwaipr 2018, vol 11047, pp 77–84. https://doi.org/10.1007/978-3-030-01132-1_9
149. Li JY, Su ZF, Geng JH, Yin YX (2018) Real-time detection of steel strip surface defects based on improved YOLO detection network. IFAC-PapersOnLine 51(21):76–81. https://doi.org/10.1016/j.ifacol.2018.09.412
150. Li YT, Huang HS, Xie QS, Yao LG, Chen QP (2018) Research on a surface defect detection algorithm based on MobileNet-SSD. Appl Sci-Basel 8(9). https://doi.org/10.3390/app8091678
151. Yang J, Li S, Wang Z, Yang G (2019) Real-time tiny part defect detection system in manufacturing using deep learning. IEEE Access 7:89278–89291. https://doi.org/10.1109/ACCESS.2019.2925561
152. Di H, Ke X, Peng Z, Zhou D (2019) Surface defect classification of steels with a new semi-supervised learning method. Opt Lasers Eng 117(1):40–48
153. Gao YP, Gao L, Li XY, Yan XG (2020) A semi-supervised convolutional neural network-based method for steel surface defect recognition. Robot Comput Integr Manuf 61:101825. https://doi.org/10.1016/j.rcim.2019.101825
154. Tan CQ, Sun FC, Kong T, Zhang WC, Yang C, Liu CF (2018) A survey on deep transfer learning. Artificial neural networks and machine learning - Icann 2018, Pt Iii 11141:270–279. https://doi.org/10.1007/978-3-030-01424-7_27
155. Liu SP, Tian GH, Xu Y (2019) A novel scene classification model combining ResNet based transfer learning and data augmentation with a filter. Neurocomputing 338:191–206. https://doi.org/10.1016/j.neucom.2019.01.090
156. Zheng X, Chen J, Wang H, Zheng S, Kong Y (2020) A deep learning-based approach for the automated surface inspection of copper clad laminate images. Applied Intelligence. https://doi.org/10.1007/s10489-020-01877-z
157. Zhu XJ (2005) Semi-supervised learning literature survey. University of Wisconsin-Madison Department of Computer Sciences. http://digital.library.wisc.edu/1793/60444. Accessed 19 Jan 2021
158. Chapelle O, Schölkopf B, Zien A (2006) Semi-supervised learning, vol 2. MIT Press Cortes, Cambridge
159. Cortes C, Mohri M (2014) Domain adaptation and sample bias correction theory and algorithm for regression. Theor Comput Sci 519:103126
160. Odena A (2016) Semi-supervised learning with generative adversarial networks. arXiv:1606.01583. https://arxiv.org/abs/1606.01583
161. Li W, Wang Z, Li J, Polson J, Speier W, Arnold CW (2019) Semi-supervised learning based on generative adversarial network: a comparison between good GAN and bad GAN approach. In: 2019 Computer Vision and Pattern Recognition (CVPR) Workshops. arXiv:1905.06484. https://arxiv.org/abs/1905.06484
162. Berthelot D, Carlini N, Goodfellow I, Papernot N, Oliver A, Raffel CA (2019) Mixmatch: a holistic approach to semi-supervised learning. In: Advances in Neural Information Processing Systems, pp 5049–5059
163. Zheng X, Wang H, Chen J, Kong Y, Zheng S (2020) A generic semi-supervised deep learning-based approach for automated surface inspection. IEEE Access 8:114088–114099

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.